Blindness, hearing loss and brown crumbly teeth: determining the molecular basis of Heimler and Heimler plus syndromes and other related conditions

By

Claire Elizabeth Leigh Smith,

B.Sc. (Hons)

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Medicine

August 2016

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated overleaf. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Jointly authored publications statement

Chapter 3 includes data from two jointly authored publications, Ratbi et al. (2015) and Smith et al. (2016b):

Ratbi I, Falkenberg KD, Sommen M, Al-Sheqaih N, Guaoua S, Vandeweyer G, Urquhart JE, Chandler KE, Williams SG, Roberts NA, El Alloussi M, Black GC, Ferdinandusse S, Ramdi H, Heimler A, Fryer A, Lynch SA, Cooper N, Ong KR, Smith CEL, Inglehearn CF, Mighell AJ, Elcock C, Poulter JA, Tischkowitz M, Davies SJ, Sefiani A, Mironov AA, Newman WG, Waterham HR and Van Camp G. (2015) *Heimler syndrome is Caused by Hypomorphic Mutations in the Peroxisome-Biogenesis Genes PEX1 and PEX6.* ***American Journal of Human Genetics*** 97(4):535-545.

The author contributed to the publication by carrying out whole exome sequencing library preparation and sequencing analysis of DNA from individual 4505, family AI-202 (labelled as family 4 in the paper) as part of a larger cohort of Heimler syndrome patients. The author submitted findings from sequencing analysis and quality control metrics for the paper and commented critically on the manuscript prior to submission. The other authors provided all other data included in the paper.

Smith CEL, Poulter JA, Levin AV, Capasso JE, Price S, Ben-Yosef T, Sharony R, Newman WG, Shore RC, Brookes SJ, Mighell AJ, Inglehearn CF. (2016) *Spectrum of PEX1 and PEX6 variants in Heimler syndrome.* ***European Journal of Human Genetics*** 24 (11) 1565-1571.

The author wrote the paper and contributed all data except the details of the phenotype of the patients, which was contributed by the other authors. This paper includes families AI-10 (individuals 1492, 1499 and 1500), AI-146 (3000), AI-186 (4716), AI-16 (2113), AI-152 (3466 and 4922) and AI-71 (2645) which were labelled as families 1, 2, 3, 4, 5 and 6 respectively in the paper).

Chapter 4 includes data from one jointly authored publication, Smith et al. (2016a):

Smith CEL, Murillo G, Brookes SJ, Poulter JA, Silva S, Kirkham J, Inglehearn CF and Mighell AJ. (2016) *Deletion of amelotin exons 3-6 is associated with amelogenesis imperfecta.* ***Human Molecular Genetics*** doi: 10.1093/hmg/ddw203

The author wrote the paper and contributed all data except the details of the phenotype of the patients, which was contributed by the other authors. Micro-computerised tomography was carried out in collaboration with Dr. Steven Brookes (University of Leeds). This paper includes family AI-154 (individuals II:2, III:3 and IV:1).

Acknowledgements

Whilst a PhD is an individual achievement, it is a reflection of the valuable support and advice that others provide. Therefore I would like to express my thanks to my supervisors, who have each contributed something different, but equally valued, to my PhD experience. James Poulter has been an ever-present sounding board and has been available to answer my constant questions with a patience and understanding I have rarely observed in anyone else. Chris Inglehearn has been ever encouraging, in the execution of the project, my confidence and in my career. He has also provided an ever-critical eye to all outputs, but has delivered any criticism constructively and in the nicest possible way. Alan Mighell has been invaluable in recruitment of families and in providing phenotypic knowledge. In some cases, he has even been able to guess the causative gene before sequencing. He has also been ever driven to translate the results that I've obtained into patient benefit and his perspective helps to reveal the wider impact of genetic diagnosis. His thorough note keeping and clear prioritisation of tasks throughout my PhD have also been of immense value. In conclusion, I appreciate all of my supervisors' efforts and know how lucky I am to have had such people to guide me.

Thanks are also extended to Steve Brookes and Roger Shore, whose respective knowledge of enamel and teeth generally is astounding and has been a very useful resource for me to tap. I would also like to thank everyone who makes level 8 such a nice place to work.

My thanks would not be complete without also extending them to my parents. Firstly, thanks to my dad, who encouraged my interest in science and nature, especially meteorology and agriculture, from an early age. Also to my mum who has empathy without bounds and is always enthusiastic, supportive and interested in whatever I am doing. Their pride of my every achievement, however small, has always driven me to try to achieve more. Both have consistently encouraged me to do whatever I think will make me happy and without them, I would not have applied for, nor accepted, my PhD place.

Last but not least, I would like to thank Seb Tawn. Other than his cookery, holiday booking and foreign language skills, which are all fabulous and make my life far more enjoyable and run more smoothly than it otherwise would, he is also my favourite person and I look forward to being his wife.

Abstract

Enamel is the body's hardest, most mineralised tissue and protects the underlying tissues of the tooth from the forces exerted during mastication and from bacterial attack. Amelogenesis imperfecta (AI) is a heterogeneous group of genetic conditions that result in defective dental enamel formation. AI can present as either isolated disease or as part of a syndrome. This thesis documents the identification of the genes involved in Heimler syndrome (HS) and in hypomineralised AI.

HS is an autosomal recessively inherited combination of sensorineural hearing loss, AI and variable nail abnormalities, with or without visual defects. Biallelic mutations in the peroxisomal biogenesis factor genes, *PEX1* and *PEX6* were identified through whole exome sequencing (WES) of ten HS patients from seven families. Both proteins were found to be present throughout the retina. Mutations in *PEX1* and *PEX6* were already known to cause the Zellweger syndrome spectrum disorders (ZSSD), a group of conditions of varying severity. HS represents a mild ZSSD and results from combinations of mutations that include at least one hypomorphic variant, leaving patients with some residual peroxisomal function. For two additional families, mutations in the Usher syndrome (USH) gene, *USH2A*, were identified, highlighting the phenotypic overlap of HS with the more common USH.

One family with autosomal dominantly inherited hypomineralised AI was recruited and DNA from three affected individuals was subjected to WES. DNA copy number variant analysis identified a heterozygous in-frame deletion of exons 3-6 of amelotin. Exfoliated primary teeth from an affected family member had enamel that was of a lower mineral density compared to control enamel and exhibited structural defects. Some of this appeared to be associated with organic material as evidenced by elemental analysis.

Both cases in this study highlight the heterogeneity of AI and the importance of a genetic diagnosis for the clinical management of patients.

Abbreviations

AAA+	ATPases associated with diverse cellular activities
ACTB	actin, beta
AD	autosomal dominant
AI	amelogenesis imperfecta
AMBN	ameloblastin
AMEL	amelogenin, refers to both copies present on chromosomes X and Y in humans
AMELX	amelogenin, X linked
AMELY	amelogenin, Y linked
AMTN	amelotin
APS	ammonium persulphate
AR	autosomal recessive
ATAD3B	ATPase family, AAA domain containing 3B
ATAD3C	ATPase family, AAA domain containing 3C
ATP	adenosine triphosphate
BAM	binary alignment/map
BCA	bicinchoninic acid
BLAST	Basic Local Alignment Search Tool
BLAT	BLAST-like alignment tool
BMP	bone morphogenetic protein
bp	base pair
BWA	Burrows-Wheeler aligner
CADD	Combined Annotation Dependent Depletion
CAPN5	calpain 5
cDNA	complementary deoxyribonucleic acid
cM	centiMorgan
COL17A1	collagen, type XVII, alpha-1
CNV	copy number variant
CSNK1A1	casein kinase 1 alpha 1
C4orf26	chromosome 4 open reading frame 26
D	domain
DAB	3, 3′-diaminobenzidine

dbSNP	database of single nucleotide polymorphisms and multiple small-scale variations
ddNTPs	dideoxynucleoside triphosphates
DEJ	dentino-enamel junction
DLX3	distal-less homeobox 3
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
ddNTPs	dideoxynucleotide triphosphates
DSPP	dentin sialophosphoprotein
EDTA	ethylenediaminetetraacetic acid
EDX	energy dispersive X-ray spectroscopy
EMP	enamel matrix protein
ENAM	enamelin
ER	endoplasmic reticulum
ERS	enamel renal syndrome
EVS	Exome Variant Server
ExAC	Exome Aggregation Consortium
FAM20A	family with sequence similarity 20, member A
FAM20C	family with sequence similarity 20, member C
FAM83H	family with sequence similarity 83, member H
FGF	fibroblast growth factor
FS	Fisher strand
GATK	Genome Analysis Toolkit
GCL	ganglion cell layer
GPR68	G protein-coupled receptor 68
GRCh37	Genome Reference Consortium human 37
HA	hydroxyapatite
H&E	haematoxylin and eosin
HS	Heimler syndrome
IEE	inner enamel epithelium
IGV	Integrated Genomics Viewer
IPL	inner plexiform layer
iPSC	induced pluripotent stem cells

IRD	infantile Refsum disease
IS	inner segment
ITGB6	integrin, beta 6
JEB	junctional epidermolysis bullosa
Kb	kilobase
kDa	kilodalton
KLK	kallikrein related peptidase
KLK4	kallikrein related peptidase 4
KLK5	kallikrein related peptidase 5
LAMA3	laminin, alpha 3
LAMB3	laminin, beta 3
LCA	Leber congenital amaurosis
LOVD	Leiden Open Variant Database
MAF	minor allele frequency
Mb	megabase
MD	macular dystrophy
MEM	maximum entropy model
microCT	micro computerised tomography
MIH	molar incisor hypomineralisation
MIM	Mendelian Inheritance in Man
MMP20	matrix metalloproteinase 20
MQ	mapping quality
MSX1	msh homeobox 1
MT	mutation taster
MYO7A	myosin VIIA
MYO7ADB	LOVD retinal and hearing impairment database myosin VIIA (*MYO7A*)
mRNA	messenger ribonucleic acid
NALD	neonatal adrenoleukodystrophy
NCC	neural crest cells
NGS	next generation sequencing
NMD	nonsense mediated decay
NP	no primary antibody control
NS	non-syndromic

ODAM	odontogenic, ameloblast associated
OMIM	Online Mendelian Inheritance in Man
OPL	outer plexiform layer
PAGE	polyacrylamide gel electrophoresis
PAX9	paired homeobox 9
PBD	peroxisome biogenesis disorder
PCR	polymerase chain reaction
PEX1	peroxisomal biogenesis factor 1
PEX6	peroxisomal biogenesis factor 6
PEX26	peroxisomal biogenesis factor 26
PL	papillary cell layer
PS	phenotypic series
PSSM	position specific scoring matrix
PVDF	polyvinylidene fluoride
QD	quality by depth
RA	ruffle ended ameloblasts
RIPA	radio immunoprecipitation assay
RNA	ribonucleic acid
RP	retinitis pigmentosa
RPKM	reads per kilobase per million mapped reads
SAM	sequence alignment/map
SCPP	secretory calcium-binding phosphoproteins
SA	smooth ended ameloblasts
SDS	sodium dodecyl sulphate
SEM	scanning electron microscopy
SHH	sonic hedgehog
SI	stratum intermedium
SIBLING	small integrin binding ligand N-linked glycoproteins
SIFT	Sorting Intolerant From Tolerant
SLC24A4	solute carrier family 24 (sodium/potassium/calcium exchanger), member 4
SMART	Simple Modular Architecture Research Tool
SNHL	sensorineural hearing loss
SNP	single nucleotide polymorphism

SNV	single nucleotide variant
SRH	second region of homology
SROOGLE	Splicing Regulation Online Graphical Engine
STR	short terminal repeat
TAE	tris acetate EDTA
TBS	tris buffered saline
TBST	tris buffer saline-Tween 20
Taq	*Thermus aquaticus*
TE	tris EDTA
TEMED	tetramethylethylenediamine
UCSC	University of California, Santa Cruz
US	United States of America
USH	Usher syndrome
USH1	Usher syndrome type I
USH2	Usher syndrome type II
USH2A	usherin
USH2ADB	LOVD retinal and hearing impairment genetic Usher syndrome 2A (*USH2A*) mutation database
UV	ultraviolet
VCF	variant call format
VEP	variant effect predictor
WDR72	WD repeat-containing protein 72
WES	whole exome sequencing
WGS	whole genome sequencing
WT	wild-type
ZS	Zellweger syndrome
ZSSD	Zellweger syndrome spectrum disorder

Contents

List of Figures

List of Tables

1 Introduction

1.1 Structure, function and origins of teeth

Human teeth are formed from a collection of mineralised tissues, namely enamel, dentine and cementum, and the non-mineralised pulp tissue (Figure 1.1).

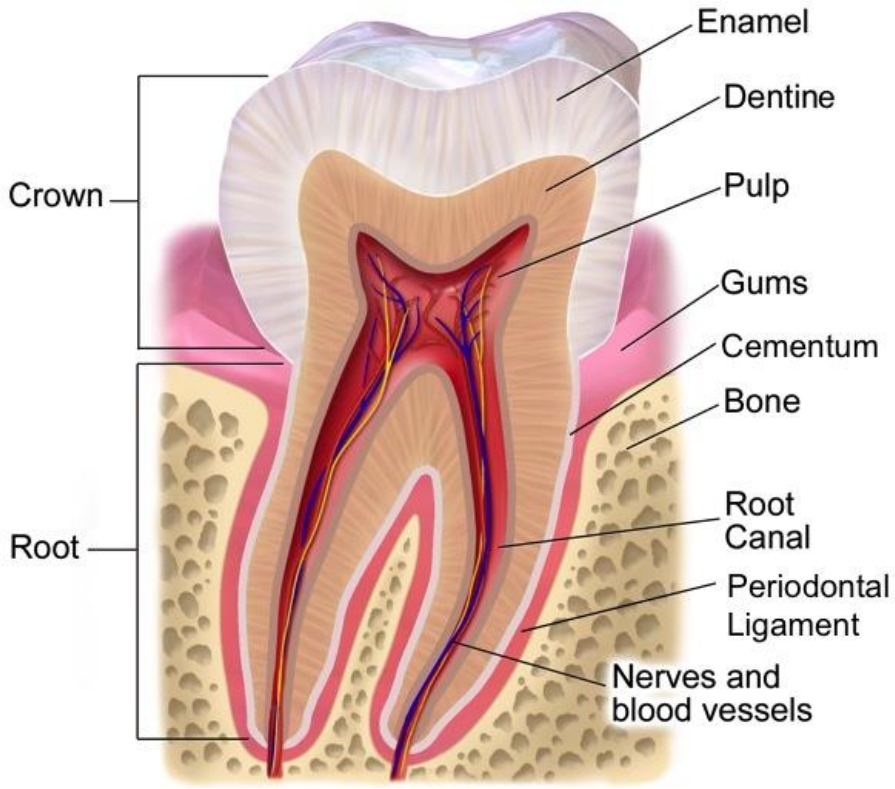


Figure 1.1: Structure of a human molar tooth.

Image reproduced with permission from Blausen.com staff. “Blausen gallery 2014”. *Wikiversity Journal of Medicine*. DOI:10.15347/wjm/2014.010. ISSN 200187762

The basic tooth structure consists of the crown and the root, which are composed of four layers of material. The hard enamel covering the crown overlays the softer and more elastic dentine of both crown and root. The enamel provides a surface for mastication and defence against acid attack and pathogens. The dentine is supported by soft pulp tissue that provides nutrients and contains cells that allow dentine to maintain and repair itself. The cementum links to the dentine layer of the root and covers the root of the tooth, linking it to the periodontal ligament and ultimately, to the bone.

Teeth provide a scaffold for the shape of the lower face and are important for proper speech. They are also essential for mastication of food and are involved in emotional and social cues and perceptions (Beall, 2007).

Teeth are found in jawed vertebrates (gnathostomes) and originated approximately 450 million years ago (Davit-Beal et al., 2009), although for several vertebrate taxa, either teeth or their enamel has been lost (Davit-Beal et al., 2009). Whilst tooth morphology has diversified during evolution, the underlying structure remains similar (Huysseune and Sire, 1998). Humans are heterodont (have teeth of more than one type of morphology) and diphyodont (have a deciduous and permanent dentition).

1.2 Biomineralisation

Biomineralisation is the formation of mineralised tissues by living organisms. It involves the production of organic signalling molecules and in some cases the establishment of a scaffold on which mineral is to form, followed by the process of mineral deposition. Many different tissues undergo biomineralisation, but the process of biomineralisation in enamel is unique since it involves a specific proteinaceous extracellular matrix instead of the collagenous scaffold used in bone and dentine and it cannot be repaired.

Biomineralised tissues are mainly formed from calcium phosphate in the form of hydroxy(l)apatite (HA), with varying amounts of proteins. The molecular formula of HA, $Ca_{10}(PO_4)_6(OH)_2$, is conventionally written as two molecules to represent the repeating unit of the crystal. However, HA has a crystal lattice structure in which particular ions are often substituted. For example, the mineral can be deficient in calcium or hydroxyl, and either carbonate, fluoride, sodium or magnesium may substitute for them, which in turn can affect the lattice's solubility (Robinson et al., 2000).

Biomineralisation results from the controlled deposition of calcium and phosphate from supersaturated compartments. Control is elicited by energy barriers to mineral seeding and by specific proteins that, depending upon the

context, can either stabilise the ions or promote their organised deposition on to extracellular matrices. These proteins include the secretory calcium-binding phosphoproteins (SCPPs) that include milk/salivary proteins, enamel proteins and dentine/bone proteins. The SCPPs are encoded by genes believed to have evolved via gene duplication events (Kawasaki and Weiss, 2003). In humans, they mainly reside within a conserved gene cluster on chromosome 4 (Figure 1.2) (Kawasaki and Weiss, 2003, Kawasaki and Weiss, 2006), although *AMELX* is also thought to have evolved via a gene duplication event but has been translocated to the X chromosome (with another copy present on the Y chromosome, *AMELY*) (Iwase et al., 2007). Other genes that are important for biomineralisation are distributed throughout the genome.

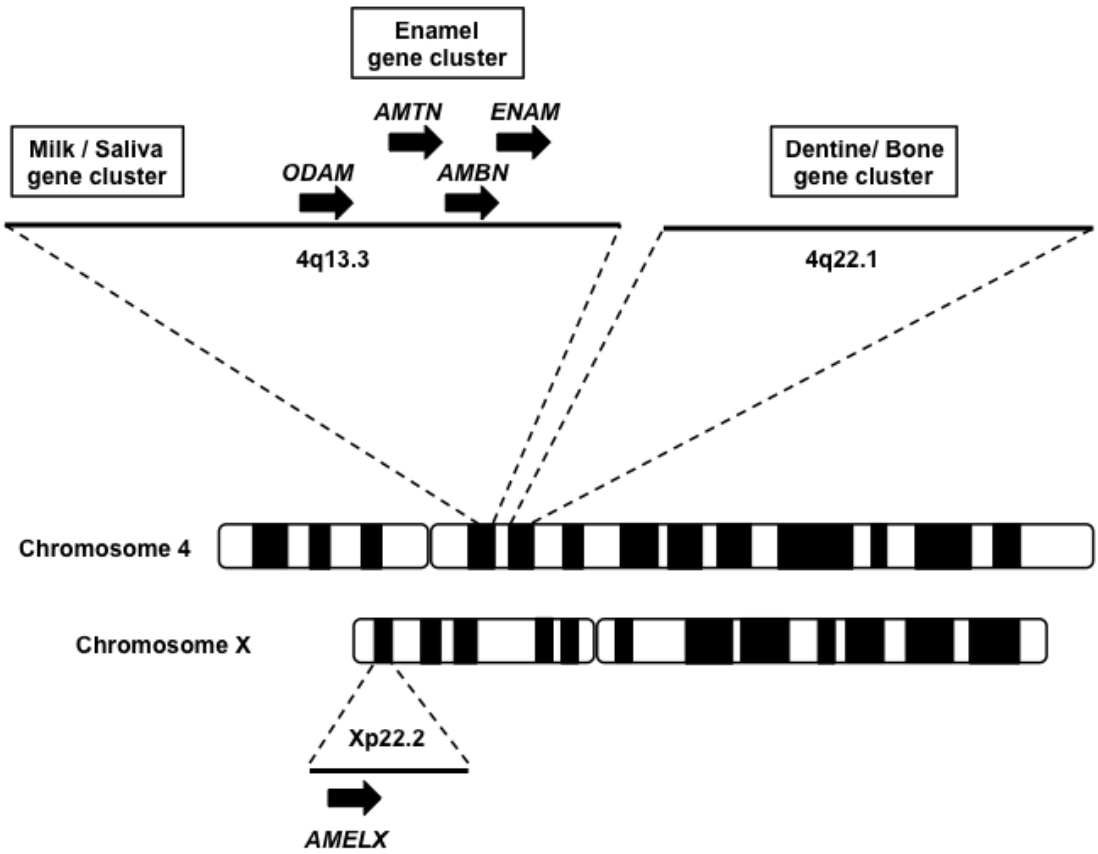


Figure 1.2: Chromosomal locations of the conserved SCPP genes in humans.
The main cluster of enamel genes are located on chromosome 4 distal to the milk / saliva gene cluster and proximal to the dentine / bone gene cluster. *AMELX* is located separately on chromosome X. Image based on information presented by Yoshizaki and Yamada (2013).

1.3 Tooth formation

Odontogenesis, the process of tooth formation, involves a number of processes that can be broadly defined as initiation, morphogenesis, cell differentiation and matrix secretion. These processes occur over a number of contiguous stages that are defined by their morphological features.

Human teeth begin to form from the seventh week of gestation (Nanci, 2012). They are derived from the ectodermal cranial neural crest-derived mesenchyme and epithelium of the first pharyngeal arch and a part of the frontonasal process (Figure 1.3) (Mitsiadis and Luder, 2011). Tooth formation is initiated by the oral epithelium, but subsequently the process is maintained by the ectomesenchyme, although transient epithelial signalling centres are established later in tooth morphogenesis to specify the shape of each tooth's crown (Kollar and Baird, 1970, Mina and Kollar, 1987, Thesleff and Mikkola, 2002). A "homeobox code" is established to provide instructions as to the position and type of teeth to be formed (Sharpe, 1995).

The majority of studies on tooth morphogenesis have been undertaken in rodents because of the difficulty in obtaining appropriate embryological human tissues and the ethical questions that such experimental studies would raise. Therefore the majority of current literature is based on findings in rodent model systems, more specifically on the odontogenesis of the mandibular arch.

Tooth development consists of five major stages: initiation / placode stage, bud stage, cap stage, bell stage and root formation followed later by eruption (Figure 1.4).

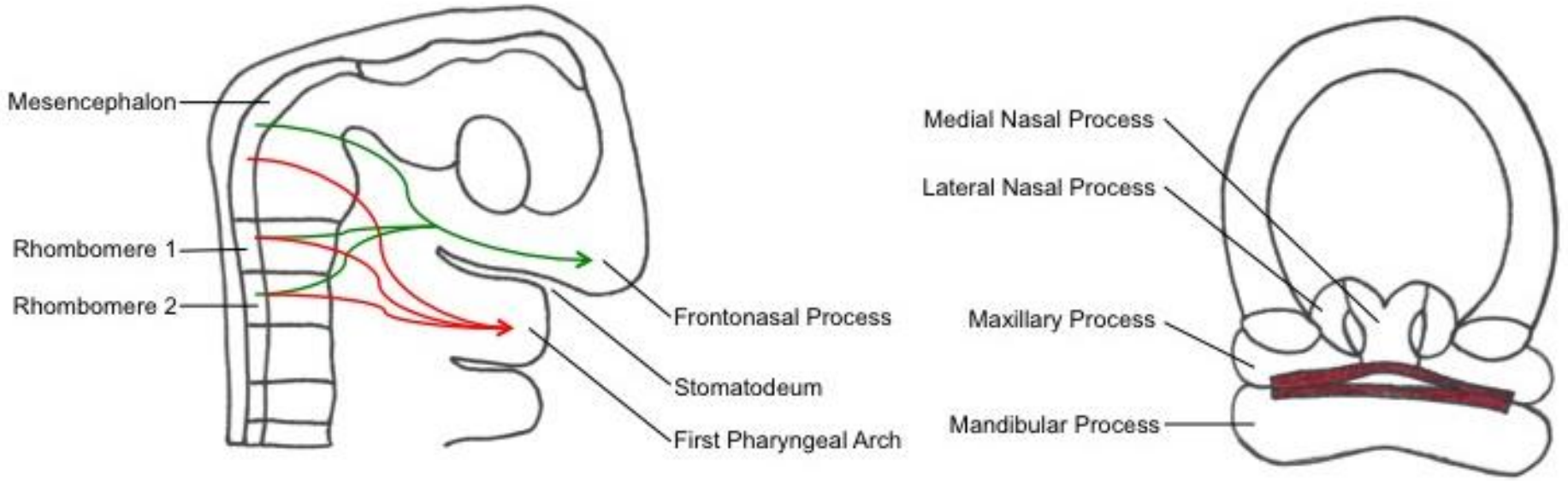


Figure 1.3: Early embryonic development of the head, face and jaws.

Left: Schematic diagram depicting the side view of a human embryo at around 4 weeks gestation. The arrows represent the patterns of migration of mesencephalic, and rhombencephalic neural crest cells (NCCs) into the frontonasal process and the first pharyngeal arch. The green arrows represent the first stream of migration of NCCs from the mesencephalon and rhobomeres 1 and 2, which contributes to the skull, forehead and nasal process. The red arrows represent the second stream of migration of NCCs from the mesencephalon and rhobomeres 1 and 2, which go on to contribute to the formation of the jaws. The stomatodeum, the depression between the frontonasal process and the first pharyngeal arch, is the precursor of the oral cavity. Right: Schematic diagram to depict the front view of a human embryo at around 6 weeks gestation. The maxillary and mandibular facial process and the medial nasal process contribute to a region of epithelium with the potential to form teeth (red shading).

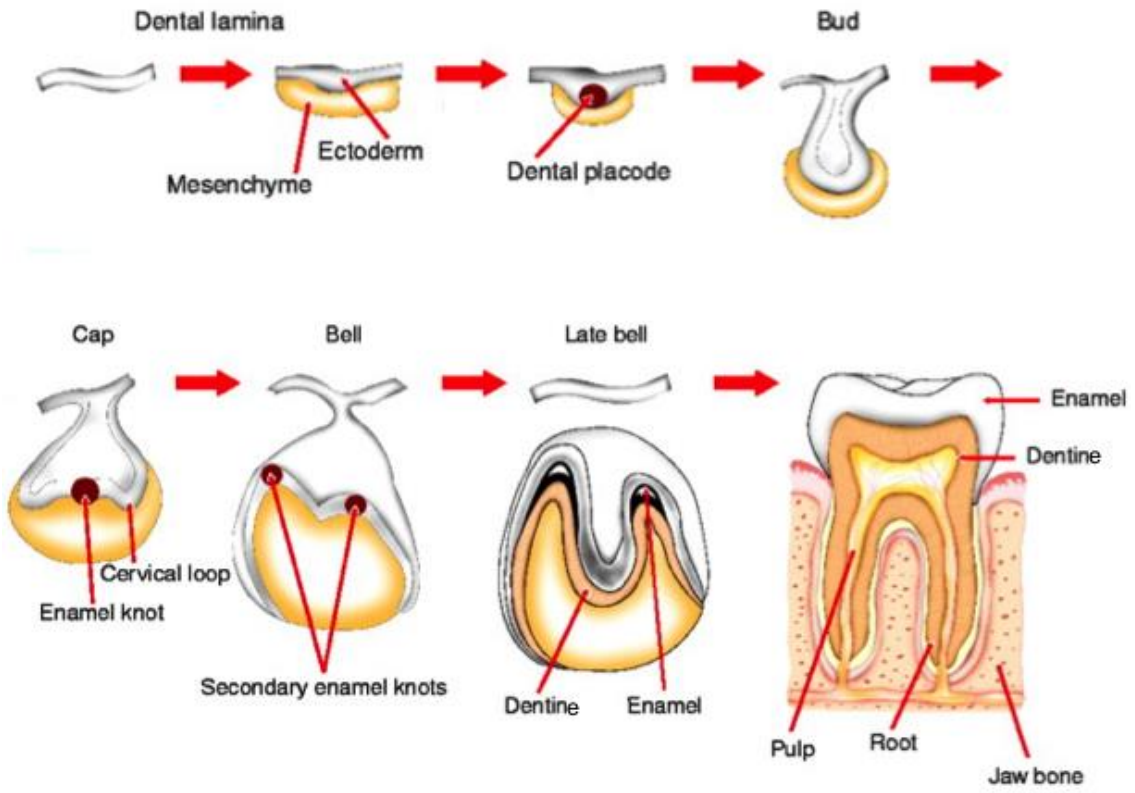


Figure 1.4: Diagram of the stages of tooth development.

The process of tooth development is described in Section 1.3. Red dots represent the positions of signalling centres that direct development of the tooth at that particular stage. Figure reproduced and adapted with permission from Thesleff (2003)

1.3.1 Initiation / placode stage (Mouse E11.5)

During initiation, patterns of expression of homeobox-containing genes are established in the ectomesenchyme by the oral epithelium (Tucker et al., 1999). These induce localised proliferation of the dental lamina to form dental placodes that act as transient epithelial signalling centres (Biggs and Mikkola, 2014). The exact tooth position and type is determined by a complex interplay and cascades of signalling molecules and transcription factors (Neubuser et al., 1997, Bei and Maas, 1998, Tucker et al., 1998, Trumpp et al., 1999, Ferguson et al., 2000). In particular, molecules of the bone morphogenetic protein (BMP), wingless, notch, sonic hedgehog (SHH) and fibroblast growth factor (FGF) families are involved in tooth positioning, formation and type patterning (Li et al., 2013). SHH, BMP4, FGF8, Msh homeobox1 (MSX1) and paired box 9 (PAX9) are the signalling molecules involved in the initial stages of tooth development, although some also have roles at later stages (Caton and Tucker, 2009).

1.3.2 Bud stage (Mouse E13.5)

The bud stage involves the invagination of the dental epithelium into the underlying condensed ectomesenchyme and corresponds to a shift in the odontogenic potential from the oral epithelium to the mesenchyme (Chen et al., 1996, Peters et al., 1998).

1.3.3 Cap stage (Mouse E14.5)

The cap stage is so-called since it results in the formation of an epithelially-derived “cap”, now termed the enamel organ, on top of a condensed ectomesenchyme “ball”, now called the dental papilla, for each tooth germ. Within the enamel organ, the stellate reticulum, a collection of star shaped epithelial cells, forms. Surrounding the entire structure is the dental follicle, which marks the limit of the condensed ectomesenchyme. A cluster of non-dividing epithelial cells, termed an enamel knot, establishes itself as a transient epithelial signalling centre as a result of signalling from the mesenchyme (Caton and Tucker, 2009) and contributes to the formation of each tooth’s distinctive architecture (Pispa et al., 1999). The cell cluster that will form the enamel knot is in fact present earlier than bud stage but is not technically referred to as an enamel knot until it undergoes morphological changes at the cap stage. In molar teeth secondary and tertiary enamel knots are also established (Jernvall et al., 1994).

1.3.4 Bell stage (Mouse E16.5-20)

During the bell stage, the epithelial “cap” proliferates more deeply into the dental papilla and the dental organ is separated from the oral epithelium as the dental lamina breaks down (Figure 1.5). The section of the epithelial layer that lies over the dental papilla, and that is bounded by the cervical loops, is termed the inner enamel epithelium (IEE). Some of the cells of the IEE differentiate to form the stratum intermedium (SI) (Harada et al., 2006). Waves of differentiation of the IEE and the dental papilla produce both preodontoblasts and preameloblasts that will go on to form the hard tissues of the crown; dentine and enamel (Tompkins, 2006).

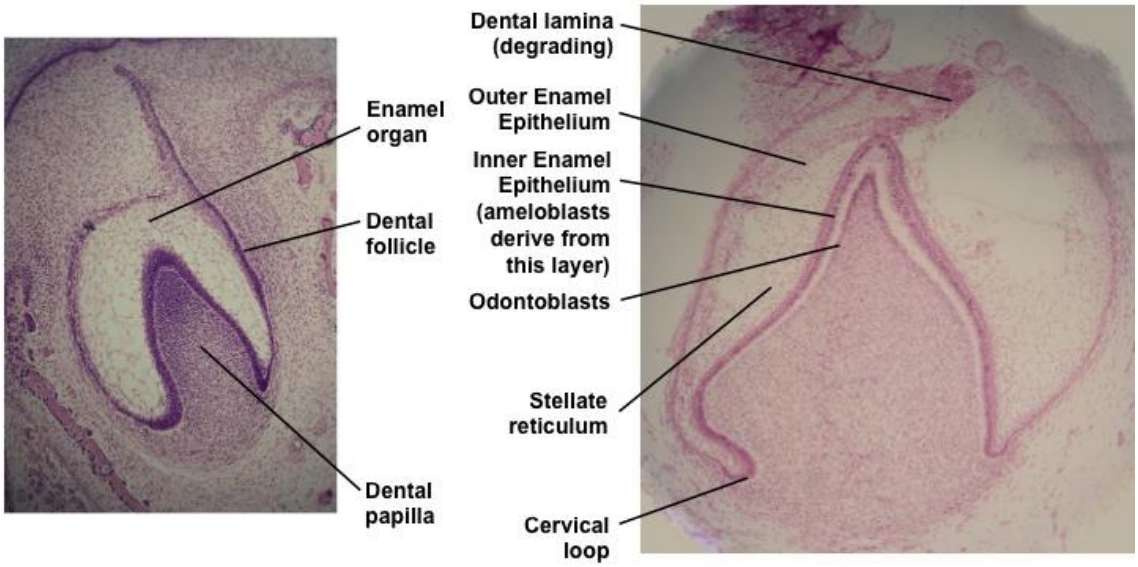


Figure 1.5: Haemotoxylin and eosin stained bell stage human tooth germs.

Left: early bell stage, right: late bell stage.

Images reproduced with permission from Dozenist, Wikipedia, accessed 24th June 2016.

Both dentinogenesis, the process of dentine formation (Section 1.5) and amelogenesis, the process of enamel formation (Section 1.4), take place in the bell stage. Signalling from the preameloblasts across the basement membrane to the preodontoblasts initiates the reciprocal signalling to initiate the final stages of each cell type's differentiation into ameloblasts and odonotoblasts (Ruch et al., 1995, Tompkins, 2006). The preodontoblasts polarise and undergo internal reorganisation to become odonotoblasts. The odontoblasts secrete a pre-dentine matrix, which when it begins to mineralise, prompts the preameloblasts to elongate and change their polarity so that their apical face is adjacent to the papilla (Nanci, 2012).

The odontoblasts move towards the centre of the dental papilla as they secrete an extracellular matrix, mainly composed of collagen I, that mineralises to form dentine (Nanci, 2012). At the same time the ameloblasts move away from the mineralising dentine while secreting an extracellular matrix of enamel proteins, that will mineralise to form enamel. The boundary between the layers of dentine and enamel is termed the dentine-enamel junction.

1.3.5 Root formation and eruption

Root formation initiates after coronal dentine formation and the root is itself formed from dentine covered by cementum (Nanci, 2012). Tooth eruption is prompted by the formation of the root. The remnants of the ameloblasts and other cells, form the reduced enamel epithelium that fuses with the oral epithelium and degenerates to form an epithelial canal through which the tooth erupts (Nanci, 2012). The reduced enamel epithelium and the oral epithelium then form the junctional epithelium (Bosshardt and Lang, 2005).

Once teeth erupt, they are subject to physical and chemical insult from mastication, physical injury and bacterial attack that can result in post eruptive change. The high mineral content of saliva, its pH buffering activity, antibody content, and action as a lubricant, all help to protect teeth (Lagerlof, 1983, Dawes et al., 2015).

1.3.6 Diphyodont dentitions

Humans, unlike mice, have a diphyodont dentition with deciduous teeth preceding a permanent dentition (Section 1.1). The formation of teeth in both dentitions is similar. In humans, the permanent successional teeth (teeth that have deciduous equivalents) such as incisors, canines and premolars, form from dormant tooth buds formed by proliferation of the dental lamina (Jarvinen et al., 2009). Non-successional teeth form from new tooth buds formed by posterior outgrowths of the dental lamina (Nanci, 2012).

1.4 Amelogenesis

Amelogenesis is the process of enamel formation. Mature enamel is the body's hardest, most mineralised tissue (at greater than 95% mineral by weight (Smith, 1998)) and forms from a secreted extracellular matrix that is progressively mineralised in a series of repetitive steps. It results in the formation of a highly organised and distinctive architecture, consisting of enamel prisms made up of hydroxyapatite crystals (Smith, 1998).

Since the cells that form enamel, the ameloblasts, are lost when teeth erupt (Abiko et al., 1996), enamel is formed only once for each dentition and cannot undergo cellular repair. Therefore enamel is required to maintain function over a lifetime. Its ability to achieve this is due to its high mineral content and complex structure (Section 1.6), both of which are acquired during its highly orchestrated formation.

Amelogenesis itself has three basic contiguous stages, namely secretory, transition and maturation, but the morphogenesis, cytodifferentiation, pre-secretory stage and eventual fate of the ameloblasts are also important to the formation and properties of enamel. The events of transition and maturation stages are difficult to separate as most maturation stage processes initiate in the transition stage. Histological sections of the murine incisor of various stages of amelogenesis are shown in Figure 1.6.

1.4.1 Pre-secretory stage

The ameloblasts are derived from cells of the IEE and differentiate in response to reciprocal signalling between the IEE and the dental papilla (Morotomi et al., 2005, Mitsiadis et al., 2008). The basal lamina then breaks down between the dental papilla and the IEE so that the cells are in contact with the pre-dentine (Figure 1.7A) (Reith, 1967, Kjoelby et al., 1994). Preameloblasts are known to secrete proteins even at the very early stages of their differentiation, including prior to the breakdown of the basal lamina (Nanci, 2012). Transient secretion of dentin sialoprotein (DSP) (a protein more associated with odontoblast secretion (Section 1.5)) has been noted during ameloblast differentiation (Ritchie et al., 1997).

Next the ameloblasts elongate and their nuclei shift proximally to the side of the cell nearest the SI, resulting in a reverse in their polarity (Matsuo et al., 1992). The Golgi apparatus and rough endoplasmic reticulum (ER) increase in size to increase ameloblast capacity for protein production, post-translational modification and secretion (Smith and Nanci, 1995). These develop at the distal end of the cell that faces the odontoblasts. The non-dividing cell becomes

polarised further as, just after a junctional complex, it forms a distal extension that will become the Tomes' process. Each ameloblast develops and maintains anchoring junctions that hold the cells in alignment and control what passes between them (Bartlett and Smith, 2013).

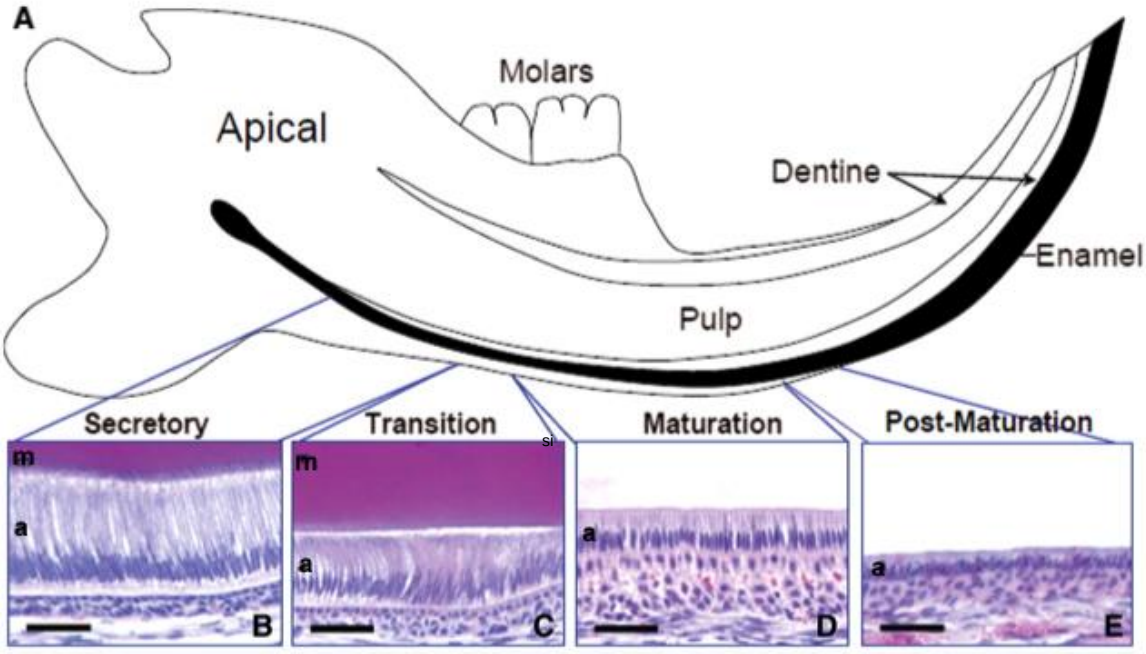


Figure 1.6: Histology of the ameloblasts in the murine mandibular incisor.

A: Representation of the basic structure of the murine mandible. B-E: Haemotoxylin and eosin stained sections of decalcified developing murine incisor. B: Secretory stage ameloblasts (a) with apical Tomes' processes and the developing enamel matrix (m). C: Transition stage ameloblasts: note the reduction in ameloblast length. D: Maturation stage ameloblasts with the stratum intermedium and papillary cell layer below. E: Post-maturation stage, reduced ameloblasts. The white space in panels D and E represents an artefactual space resulting from removal of the mineralised enamel. The mineralised enamel is removed by the demineralising treatment necessary for the histological analysis of hard tissues. Scale bars represent 50μm. Image reproduced with permission from Barron et al. (2010).

1.4.2 Secretory stage

During the secretory stage, protein secretion by ameloblasts is at its maximum and the future enamel is created to its full thickness. A proteinaceous extracellular matrix is secreted from the ameloblast Tomes' process, as the ameloblasts retreat from the dentine layer (Skobe, 1976). To achieve this, ameloblasts produce large amounts of membrane bound secretory granules

containing enamel matrix proteins (EMPs). EMPs are constitutively secreted by exocytosis (Kallenbach, 1977, Sasaki, 1984), into the extracellular space at the distal end of the cell, against the newly formed dentine. HA mineral immediately forms in this initial enamel matrix and forms a close association with the dentine mineral (Diekwisch et al., 1995, Fang et al., 2011). This will form aprismatic enamel (Figure 1.7B). The ameloblasts then begin to move away from the dentine and further develop their Tomes' processes at their distal end (Reith, 1967). To form prismatic enamel (Figure 1.7C, D), EMPs are secreted from two aspects of the ameloblasts. Enamel matrix that will form the interprismatic enamel is secreted from the proximal part of the Tomes' process and the enamel matrix that will form the prismatic enamel is secreted from the distal part (Nanci, 2012).

As secretion progresses, the Tomes' process of the ameloblasts lengthens and thins. The distal portion secreting the prism-forming enamel matrix is reduced before secretion ceases, meaning that the final enamel formed is aprismatic like the initial enamel.

EMPs that are secreted include amelogenin (AMEL, from both chromosome X and chromosome Y copies, Section 1.9.1), enamelin (ENAM, Section 1.9.3) and ameloblastin (AMBN, Section 1.9.2). AMEL, ENAM and AMBN proteins are cleaved at specific positions by the secretory stage proteinase, matrix metalloproteinase 20 (MMP20, Section 1.9.4) (Ryu et al., 1999, Moradian-Oldak et al., 2001, Yamakoshi et al., 2006, Iwata et al., 2007). Cleavage products from all three EMPs are believed to occupy particular compartments in the developing enamel architecture and to have specific roles in the formation and organisation of the developing matrix (Uchida et al., 1991, Murakami et al., 1997, Yamakoshi, 2011).

1.4.3 Transition stage

The transition stage is characterised by reduced enamel matrix protein secretion and internal reorganisation of the ameloblasts (Reith, 1970). The ameloblasts shorten to around half their original height and reduce in volume

(Figure 1.7E), the nuclei become more central and the ER is reduced in size. The Tomes' process is completely lost (Reith, 1970) and an atypical basal lamina is formed against the enamel matrix. Ameloblasts adhere to this via hemidesmosomal structures (Section 1.9.6).

During the transition stage, the cells of the SI, stellate reticulum and the outer enamel epithelium form the papillary cell layer (PL) (Josephsen et al., 2010). Capillaries invaginate into the PL and overlay the ameloblasts (Robinson et al., 1979). The PL and the ameloblasts have been proposed to constitute a functional unit since the cells of the PL may assist ameloblasts in the maturation stage by participating in ion transport and removal of enamel protein products and water from the developing enamel (Ohshima et al., 1998, Josephsen et al., 2010).

The ameloblast population reduces at this stage, and continues to reduce in the maturation stage, due to apoptosis. Overall, approximately 50% of the ameloblast cell population is lost (Smith and Warshawsky, 1977).

1.4.4 Maturation stage

During the maturation stage, the partially mineralised enamel matrix becomes fully mineralised by the breakdown and removal of residual enamel proteins, the removal of water and the expansion of existing HA crystals (Robinson et al., 1995). The maturation stage proteinase, kallikrein-related peptidase 4 (KLK4, Section 1.9.5), is responsible for the complete degradation of the EMP peptides that have been produced in the secretory stage by MMP20. This liberates a fluid filled space in which the HA crystals can grow in width and thickness.

Ameloblasts can take a long period of time to mature the enamel. Human enamel can take 4-5 years to complete, although the enamel of primary human teeth and that of rodents forms much faster than this (Smith, 1998). This process is the result of very small-scale changes achieved through repeated cyclical processes. The ameloblasts act as a gated, cell-lined thoroughfare for the movement of ions and degraded protein between the SI and the developing

enamel and vice versa. To this end, the ameloblast membrane facing the developing enamel repeatedly alternates in its structure between a ruffle-ended and a smooth-ended form (Figure 1.7F, G) (Warshawsky and Smith, 1974). This is achieved in coordinated groups of ameloblasts across the developing enamel (Takano et al., 1982). The ruffle-ended ameloblasts (RA) form membrane invaginations and tight junctions at the apical end, near the enamel surface, whereas the smooth-ended ameloblasts (SA) are more leaky (Josephsen and Fejerskov, 1977).

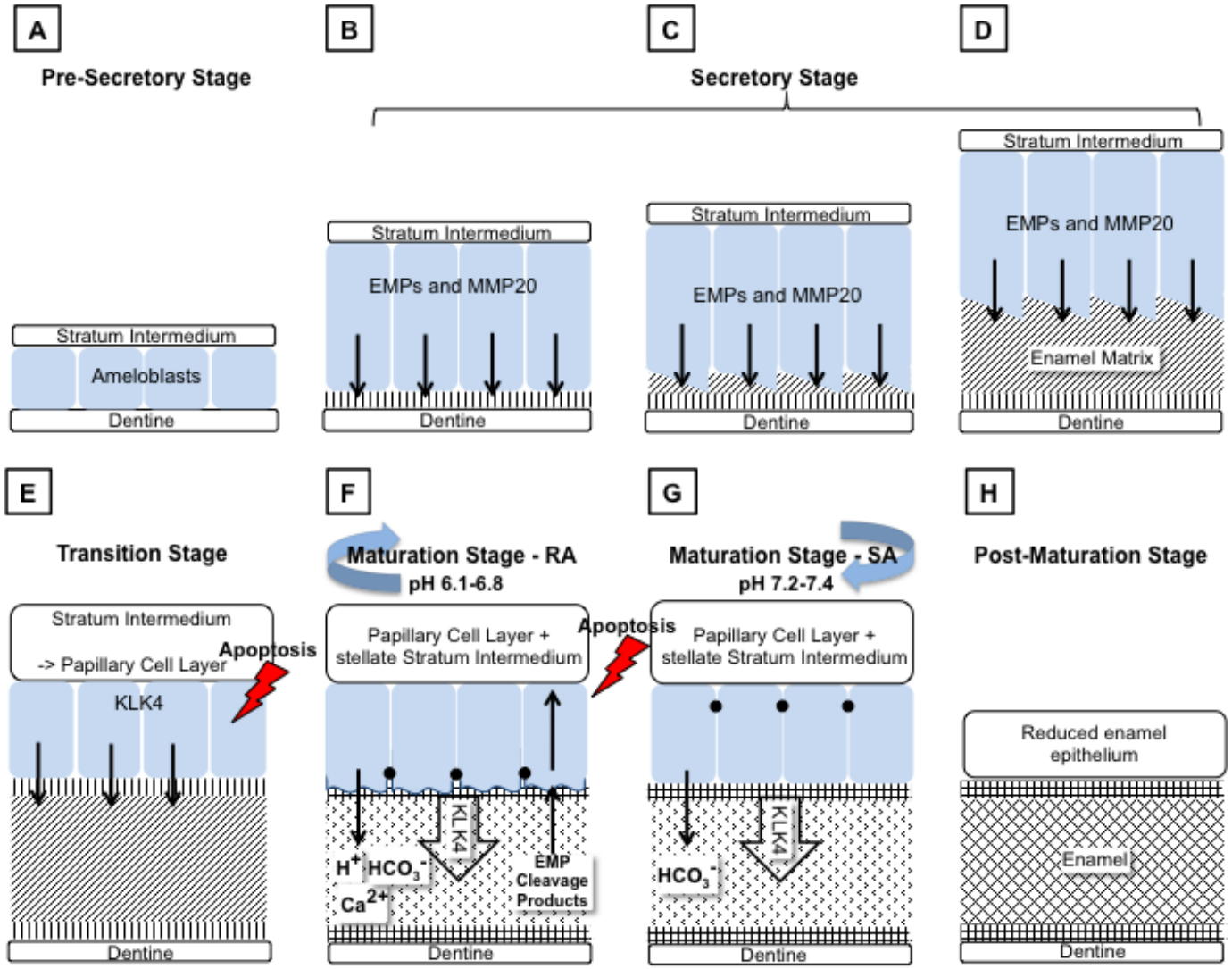
Growth of enamel crystals generates large amounts of protons (Simmer and Fincham, 1995), but it has also been shown that protons are pumped into the enamel by RA (Josephsen et al., 2010). Both RA and SA release bicarbonate ions in to the enamel that act as a buffer to increase pH (Josephsen et al., 2010). A mildly acidic pH is found in enamel at RA regions and a more neutral pH in SA regions (Sasaki et al., 1991, Smith et al., 1996).

1.4.5 Post-maturation stage

The ameloblasts, and other cells of the enamel organ, form the reduced enamel epithelium, which eventually contributes to the junctional epithelium of mature teeth (Bosshardt and Lang, 2005) (Figure 1.7H, Section 1.3.5). However, many of the ameloblasts apoptose before the formation of the junctional epithelium is completed (Abiko et al., 1996).

Figure 1.7: Schematic of amelogenesis.

See next page. Only the key features of each stage are shown and are described in Sections 1.4.1 to 1.4.5. Ameloblasts are represented in blue. Tight junctions are represented by black circles.


A
Pre-Secretory Stage
B
C
Secretory Stage
D
Stratum Intermedium
Ameloblasts
Dentine
Stratum Intermedium
EMPs and MMP20
Dentine
Stratum Intermedium
EMPs and MMP20
Dentine
Stratum Intermedium
EMPs and MMP20
Enamel Matrix
Dentine
E
Transition Stage
F
Maturation Stage - RA
pH 6.1-6.8
G
Maturation Stage - SA
pH 7.2-7.4
H
Post-Maturation Stage
Stratum Intermedium
-> Papillary Cell Layer
KLK4
Apoptosis
Dentine
Papillary Cell Layer +
stellate Stratum Intermedium
Apoptosis
KLK4
H+ HCO3- Ca2+
EMP
Cleavage
Products
Dentine
Papillary Cell Layer +
stellate Stratum Intermedium
KLK4
HCO3-
Dentine
Reduced enamel
epithelium
Enamel
Dentine

1.5 Dentinogenesis

Dentinogenesis is the process of dentine formation. Dentine is less mineralised than enamel, but more mineralised than bone, with 70% mineral content by weight (Nanci, 2012). It consists of type I collagen, which forms a lattice for mineral deposition (Prescott et al., 2008), and various non-collagenous proteins, including Small Integrin Binding Ligand N-Linked Glycoproteins (SIBLINGs). The SIBLINGs are a group of SCPPs that include dentin matrix phosphoprotein 1 (DMP1; MIM *600980) and dentin sialophosphoprotein (DSPP; MIM *125485). DSPP is cleaved by proteases to form dentin sialoprotein (DSP), dentin glycoprotein (DGP) and dentin phosphoprotein (DPP) (MacDougall et al., 1997). These proteins are thought to regulate dentinogenesis (Boskey et al., 1990)

The initiation of dentinogenesis is a pre-requisite for amelogenesis and the processes have many similarities. Pre-secretory odontoblasts form during the bell stage of tooth development (Nanci, 2012). They form from differentiation of the mesenchyme as a result of epithelial-mesenchymal interactions and signalling. Much like amelogenesis, dentinogenesis involves the secretion of an extracellular matrix followed by its mineralisation. Differentiated odontoblasts polarise to form cell bodies and a membrane process (Ruch et al., 1995). At the same time, odontoblasts move towards the centre of the dental papilla whilst secreting an initial collagenous predentine layer. Mineral crystals are seeded in the predentine and various non-collageneous proteins control mineral formation to form primary dentine. Other types of dentine, such as secondary dentine (Dean, 1998) and tertiary dentine (Smith et al., 1995) are formed throughout life to maintain the dentine layer. Therefore, unlike enamel, dentine can undergo cellular repair.

1.6 Enamel structure

Mature dental enamel consists of approximately 96% HA mineral by weight, 3% water and 1% protein (Avery et al., 2002). Enamel has a hierarchical structure that consists of crystallites as the most basic unit. These make up enamel

ribbons that form both the prismatic and interprismatic enamel created by secretion from two faces of each Tomes' process.

Rodent incisor enamel has a decussating (intersecting) pattern (Figure 1.8). Human enamel has less distinction between the prismatic and interprismatic enamel (Figure 1.9).

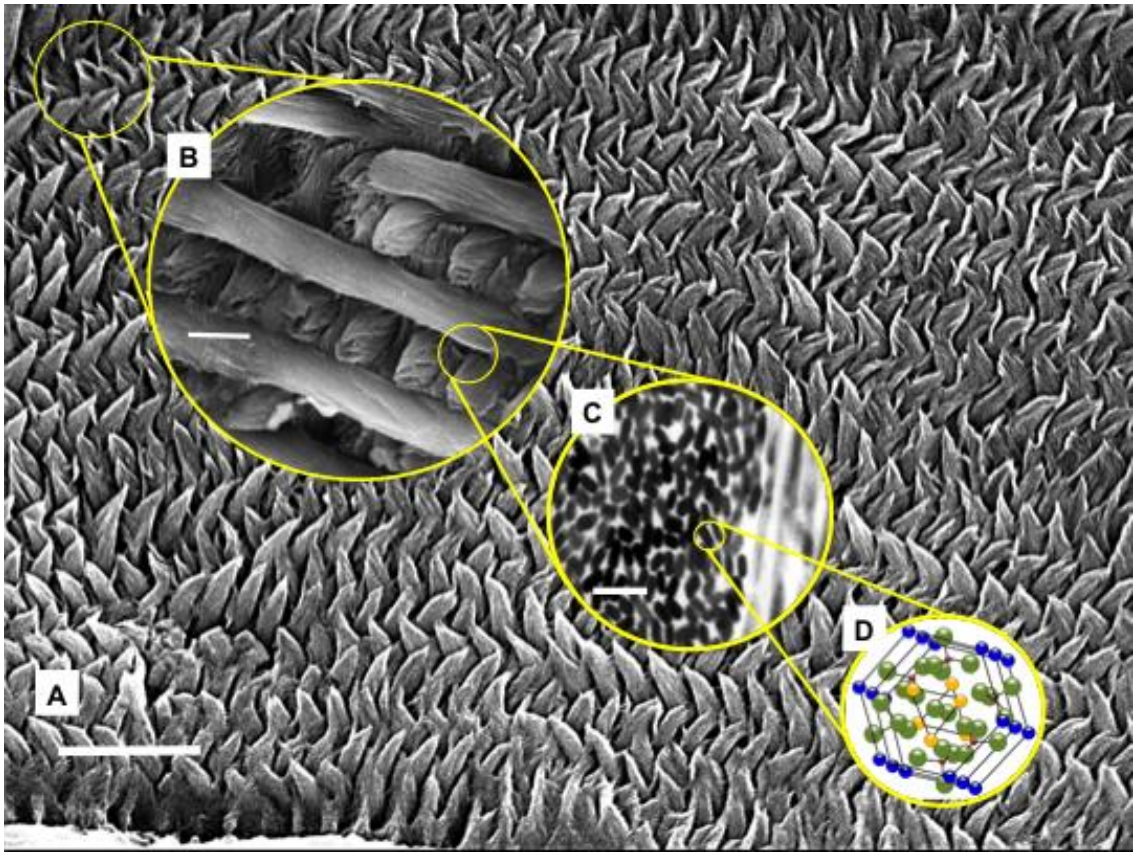


Figure 1.8: Electron microscopy (EM) images of rodent enamel.

A, B: Scanning EM images. C: Transmission EM image. Scale bars represent 40μm (A), 5μm (B) and 100nm (C). Note prism decussation shown in A. D: Crystal lattice structure of calcium hydroxy(l)apatite. Each colour represents the following: phosphorus (red), oxygen (green), calcium (screw-axis: yellow; columnar: blue). Adapted from images kindly provided by Dr. Steven Brookes, (University of Leeds). Inset image D reproduced with permission from Brunton et al. (2013).

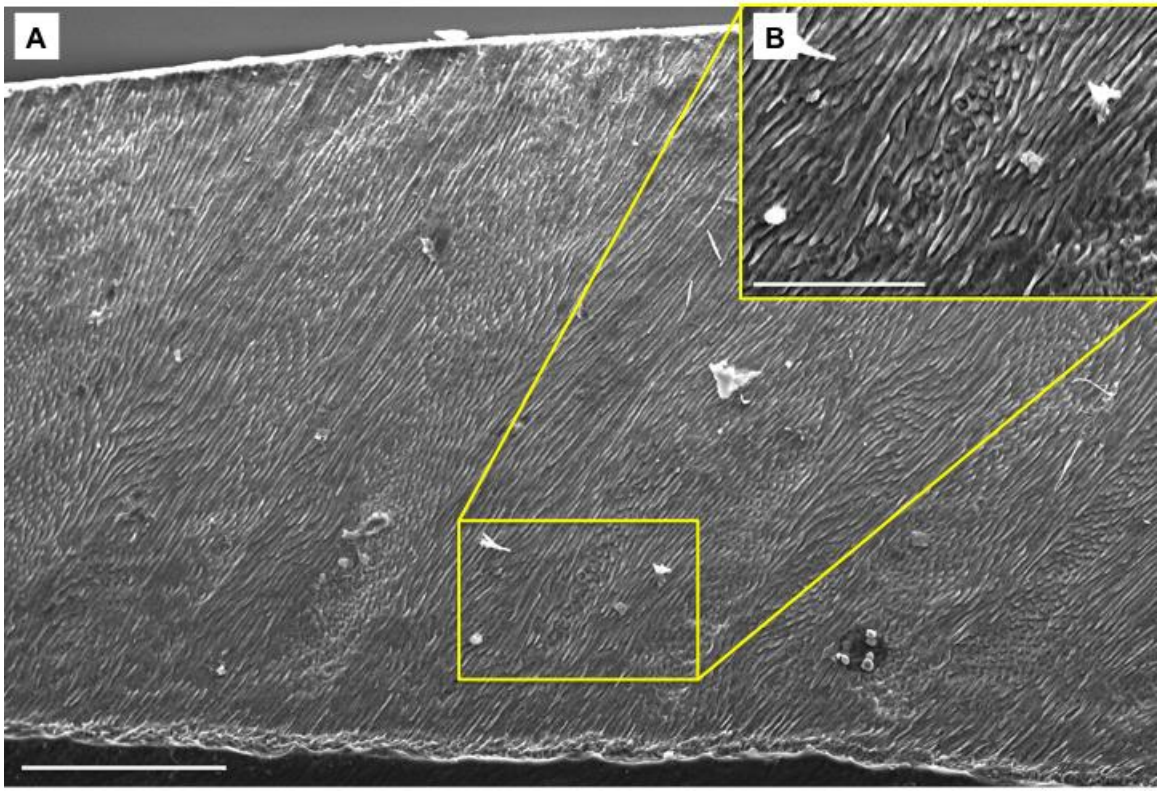


Figure 1.9: Scanning EM images of human enamel.

Scale bars represent 150μm (A) and 50μm (B).

1.7 Rodents as models for amelogenesis

Rodents have been used as experimental models for amelogenesis due to the utility of their constantly erupting incisor teeth. This characteristic means that dissected incisors can be used to study the histology of amelogenesis at all developmental stages. Since rodents and humans share evolutionarily conserved EMPs, proteinases and other proteins associated with amelogenesis, models to determine the effects of perturbations of gene expression have been engineered. These models either have genes “knocked out” by transgene insertion, specific mutations of the coding sequence inserted or have genes overexpressed or expressed in inappropriate compartments or at different developmental points via insertion of alternative promoters.

The use of rodent models to study the effects of changes in gene expression on amelogenesis is enabled by the extensive and detailed previous study of amelogenesis in rodent species. Importantly, the position of the secretory-

transition-maturation stage border has been histologically well defined in the mandibular incisor (Smith and Nanci, 1989).

Rodents also have molar teeth that, like human teeth, are formed once and for which the ameloblasts are lost upon tooth eruption. Amelogenesis has also been studied in these teeth, but unless animals at different developmental stages are studied, the teeth only provide a snapshot of amelogenesis.

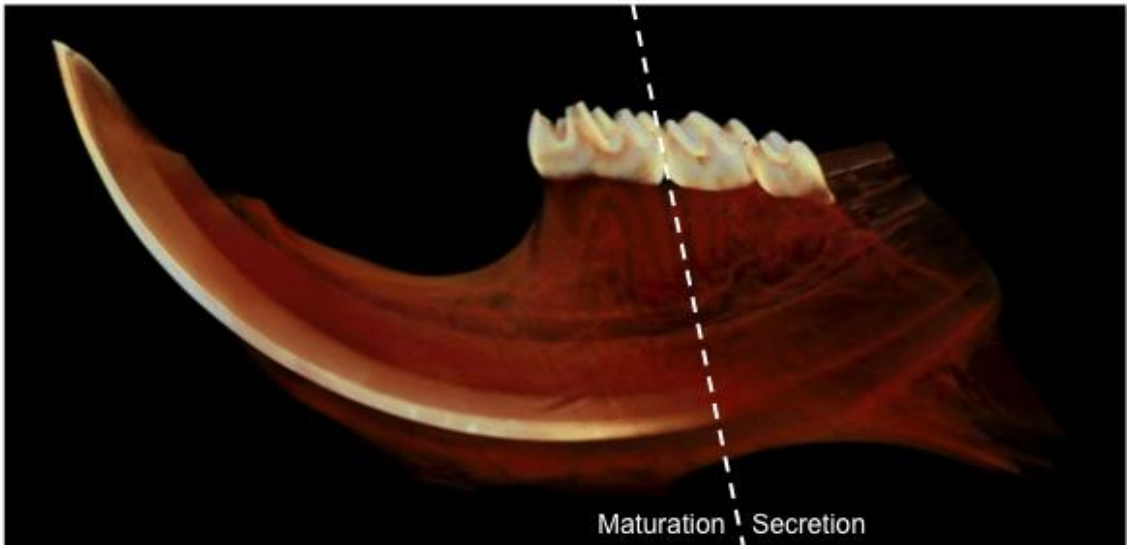


Figure 1.10: Micro computerised tomography scan of a murine lower mandible.
Section is shown through the mesial-distal axis. False colour has been used to highlight both the radiodense enamel of the molar teeth and of the incisor incisal edge. The dashed line shows the position of the molar line, used to define the position of the secretory-transition-maturation stage ameloblast border by some researchers. Adapted from an image kindly provided by Dr. Steven Brookes (University of Leeds).

1.8 Defects of enamel

1.8.1 Amelogenesis imperfecta

Amelogenesis imperfecta (AI; MIM PS104500) is a heterogeneous group of conditions characterised by defects in the formation of enamel of all teeth in both dentitions. The reported prevalence of AI differs depending upon the population studied. For example, in an isolated Swedish population, the incidence of AI was reported to be 1 in 700 (Backman and Holm, 1986), whereas it was reported to be 1 in 14,000 for a US population (Witkop and Sauk, 1976).

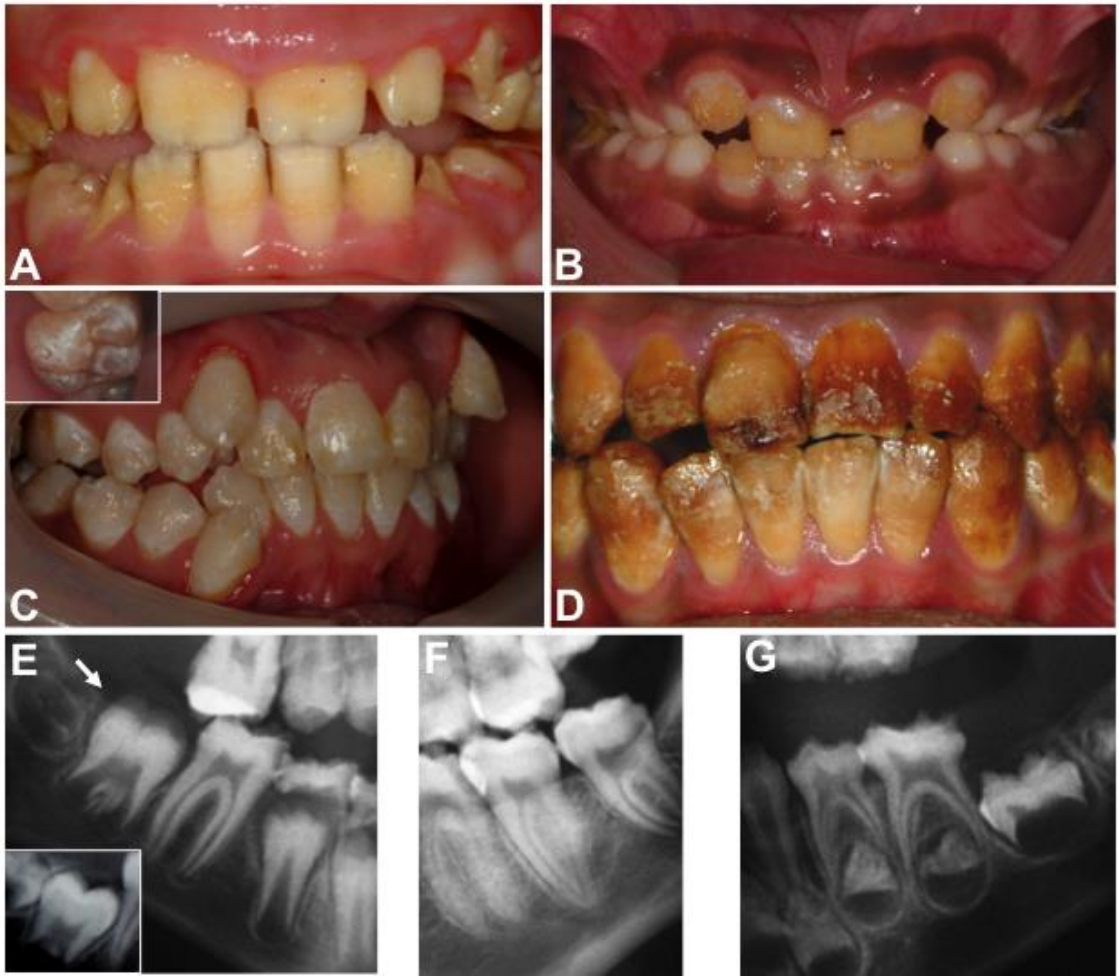


Figure 1.11: Clinical images and dental radiographs of dentitions affected by AI.

A-D: Clinical images that illustrate the phenotypic variability of AI. A: Hypoplastic AI is characterised by teeth without the curves associated with a normal enamel volume. B: AI enamel can be apparently normal, as illustrated by the posterior teeth, but structural weaknesses result in rapid post-eruptive enamel loss with exposure of the underlying dentine (B has features of hypomaturation AI). C: The quality of the enamel is abnormal, with variable colouration and enamel volume. Focal pitting is evident on the inset image. C depicts a mixed AI phenotype. D: Brown discolouration and early post-eruptive enamel loss is typical of poorly mineralised forms of AI. E-G: Dental radiographs from three unrelated cases of AI give insight into the phenotype with respect to enamel volume formed and differences in radiodensity compared to adjacent mineralised tissues. Comparison between unerupted and erupted teeth can highlight post-eruptive changes. E: The unerupted molar tooth (arrow) has a near normal enamel volume but without the difference in radiodensity between enamel and dentine that is observed in health (inset image). By contrast, the adjacent lower molar tooth that is in occlusion is characterised by minimal remaining enamel, reflecting post-eruptive hard tissue loss. F: The molar teeth have some mild changes to the enamel volume, but this is near-normal. There is a clear difference in radiodensity between enamel and dentine. Despite the radiographic appearances on clinical examination, there is obvious AI. G: The developing teeth are unable to

form normal enamel prior to eruption. Images kindly provided by Dr. Alan Mighell (University of Leeds).

AI can occur in both syndromic and non-syndromic forms. Although mixed phenotypes occur, AI can broadly be divided into hypoplastic and hypomineralised AI. Hypoplastic AI results from failure during the secretory stage and produces thin and variably mineralised enamel. Hypomineralised AI is caused by failure during the maturation stage, giving rise to enamel that is of full thickness but is weak and fails prematurely. The hypomineralised phenotype can be further subdivided into hypomaturation and hypocalicified AI. The former produces brittle enamel, while the latter produces soft enamel. The distinction between these two subdivisions is poor.

A combination of visual assessment of the physical appearance of the enamel and dental radiography is still used to classify AI. The enamel may be discoloured to yellow or brown, have a chalky or cheesy appearance and may have pits or grooves. In many cases, post-eruptive changes will alter the pre-eruption phenotype. This can make classification of AI via clinical features difficult, making the literature hard to interpret. Examples of the varied clinical presentation of AI are shown in Figure 1.11.

AI was first described in 1890 but it was not described as a separate clinical entity from dentinogenesis imperfecta (MIM #125490, #125500) until 1938 (Finn, 1938, Weinmann et al., 1945). Attempts have been made to classify AI based upon phenotype since 1945, with more recent systems focusing on the mode of inheritance or even the molecular defect as the first dichotomy in classification (Weinmann et al., 1945, Rao and Witkop, 1971, Winter and Brook, 1975, Witkop and Sauk, 1976, Sundell and Koch, 1985, Sundell and Valentin, 1986, Witkop, 1988, Aldred and Crawford, 1995, Aldred et al., 2003), for review see Aldred et al. (2003) and Crawford et al. (2007). Since the last review of classification in 2007, the number of genes in which mutations are known to cause AI has substantially increased (Section 1.9). Therefore it seems likely that this and the more frequent use of genetic diagnostics in medicine will lead to new classification based solely on the genetic cause.

AI can be inherited in an autosomal recessive (AR), dominant (AD) or X-linked manner, can occur in isolation or as a component of various syndromes, and is highly genetically heterogeneous. Despite mutations in a particular gene often resulting in one particular typical phenotype, studies have shown that the phenotype is sometimes substantially affected by the type and position of the mutation within each gene (Hart et al., 2002). The phenotype has also been reported to show variation between affected individuals within the same family (Aldred et al., 2002), suggesting that genetic background, the environment and life history factors may also contribute to the phenotype.

AI is also associated with other oral anomalies such as anterior open bite (Rowley et al., 1982, Pavlic et al., 2011, Alachioti et al., 2014), defined as a lack of vertical overlap between the maxillary and the mandibular dentition during occlusion, and taurodontism (Price et al., 1999), defined as molar teeth with enlarged pulp chambers. It can also occur as part of a syndromic condition which includes extra-oral phenotypes, such as Jalili syndrome (MIM #217080), junctional epidermolysis bullosa (JEB; MIM #226700, #226650), Kohlschütter-Tönz syndrome (MIM #226750) and trichodentoosseous syndrome (MIM #190320). Older classifications did not include syndromic AI with extra-oral symptoms but it has been proposed to include such syndromes within the AI classification due to the similarity of treatment options (Aldred et al., 2003). Therefore the definition of AI proposed by Aldred et al. is, “A group of conditions, genomic in origin, which affect the structure and clinical appearance of all or nearly all teeth, and which may be associated with morphologic or biochemical changes elsewhere in the body”.

Patients with AI experience dental pain and sensitivity, poor aesthetics and are reported to experience a lower quality of life due to social anxiety (Coffield et al., 2005, Hashem et al., 2013). Treatment of AI is expensive since it requires continuous lifetime assessment and management to maintain function and aesthetics.

1.8.2 Other developmental abnormalities of enamel

The long developmental period for enamel means that life history and environmental factors acting at this time, as well as genetics, may affect the eventual phenotype. Specific non-genetic causes include factors that lead to ameloblast stress or altered differentiation or function such as premature birth (Cruvinel et al., 2012), viral infection (Chaves et al., 2007), factors that lead to the incorporation within or diffusion into the enamel of extraneous substances such fluorosis (Robinson et al., 2004) and tetracycline exposure (Sanchez et al., 2004), or that affect calcium availability such as vitamin D deficiency (Foster et al., 2014).

Molar Incisor Hypomineralisation (MIH) is a common developmental enamel defect that has been found to affect between 2.8% (Cho et al., 2011) and 44% (Soviero et al., 2009) of people depending upon the definition of MIH used and the population and age of the cohort studied. MIH describes asymmetric enamel opacities of systemic origin that affect at least one of the permanent first molars and may also affect the incisors (Weerheijm et al., 2001). The enamel of these teeth is mineralising from around the 8th month *in utero* to around age 5 (Logan and Kronfeld, 1933). This suggests that ameloblasts are either only sensitive to the causative agent during this specific window of time or they are subject to an acute exposure at this time.

The aetiology of MIH is much debated and is thought to be multi-factorial. Both genetic and environmental factors have been suggested to be involved but all remain unproven (Whatling and Fearne, 2008, Crombie et al., 2009). Associated factors reported in the literature include mid to high socio-economic status (Balmer et al., 2012), use of certain drugs (chemotherapy (Bagattoni et al., 2014), antibiotics (Hong et al., 2011), anti-epileptic drugs (Jacobsen et al., 2013)), asthma (Guergolette et al., 2009), those whose mothers had experienced problems during pregnancy (Whatling and Fearne, 2008), early childhood infections (Lygidakis et al., 2008) environmental pollutants (Alaluusua et al., 2004, Jedeon et al., 2013) and a positive family history of MIH (Whatling and Fearne, 2008). However, the associations reported are often weak,

perhaps due to the involvement of multiple factors in aetiology. Also, the majority of epidemiological MIH studies were small and failed to control for potentially confounding factors such as water fluoridation.

One recent study assessed the degree of association of 63 single nucleotide polymorphisms (SNPs) in 21 candidate genes with MIH status in 101 Brazilian families (Jeremias et al., 2016). This study identified significant associations with SNPs positioned in many genes associated with tooth and enamel development, however the most significant was for a SNP within *AMELX*, suggesting that variants in *AMELX* may not only be associated with AI, but also with MIH. Additional studies involving different ethnicities and larger patient cohorts will be required to confirm these findings.

1.9 Genetics of non-syndromic amelogenesis imperfecta

Mutations in many genes are known to cause non-syndromic (NS) AI. Some of the encoded proteins have well documented roles in amelogenesis, acting as EMPs or the proteases that degrade them, HA nucleating factors, cell adhesion molecules or regulators of calcium homeostasis. However for others, their function is less clear, and further research is needed to understand the pathways and processes essential for the development of healthy enamel.

The first AI-causing mutations were identified in the genes encoding the EMPs. The EMPs are known to make up the bulk of the secreted enamel matrix. The proteins have a distinctive architecture since they all have a signal peptide and a conserved casein kinase 2 phosphorylation domain likely to be targeted by family with sequence similarity 20, member C (FAM20C) (Yang et al., 2016). Although the EMPs originate from a common ancestral gene, over the hundreds of millions of years since, their sequences have diverged and they have acquired distinct functions (Sire et al., 2007). Another group are the enamel matrix proteinases, which process the EMPs by cleavage and eventually degrade the protein matrix of the forming enamel so that it may be removed to accommodate subsequent mineral growth.

Several other genes, for which mutations are known to cause AI, have also been identified, however, for reasons of brevity only the *AMELX*, *AMBN* and *ENAM* (the EMPs) and *MMP20* and *KLK4* (the enamel matrix proteases) will be detailed here. A summary of the other genes known to be mutated in NS AI will be given in Section 1.9.6. Further details as to the human phenotypes and mutations associated with these genes can be found in Appendix 1 and the Leeds amelogenesis LOVD resource (Section 1.9.7). Mouse models used to study the effect of altered expression of genes known to cause NS AI are detailed in Appendix 2.

1.9.1 Amelogenin, X linked (*AMELX*)

Amelogenin, X linked (AMELX, MIM *300391) comprises around 90% of the enamel matrix proteins secreted by ameloblasts (Hart et al., 2000). In males, a highly homologous copy of *AMELX* exists as *AMELY* (MIM *410000) on the Y chromosome, but *AMELY* transcription is present at around 10% of that of *AMELX* in males (Aldred et al., 1992, Salido et al., 1992) and hence cannot compensate for loss of *AMELX* expression.

AMELX is regarded as a tooth specific protein since it has not been detected elsewhere in human tissues (Chan et al., 2011), however, in murine tissues, it has been detected in dentine-forming, cementum-forming and bone-forming cells, as well as in developing eye and brain (Fong and Hammarstrom, 2000, Janones et al., 2005, Haze et al., 2007). AMELX is thought to act as an enamel matrix buffer and as a scaffold for the spacing and growth of enamel crystallites (Chen et al., 2011). *AMELX* transcripts are subject to extensive alternative splicing (Brookes et al., 1995) that generates molecules with specific roles in enamel mineralisation and signalling (Veis et al., 2000, Veis, 2003). The most abundant human transcript lacks exon 4, which when spliced out has been proposed to act as a microRNA that may control ameloblast and odontoblast differentiation (Le et al., 2016). AMELX proteins are also subject to proteolysis to produce a variety of peptides (Ryu et al., 1999, Moradian-Oldak et al., 2001, Moradian-Oldak, 2012), and amelogenin peptides have been used

commercially to stimulate periodontal tissue regrowth in dental patients (Sculean et al., 2007, Olivares-Navarette, 2013).

Evolutionary analysis shows that *AMELX* (/*AMELY*) is most closely related to *AMBN* (Sire et al., 2007), although analysis of the amelogenin protein sequence shows that it consists of a signal peptide and an amelogenin domain, meaning that it is unlike other proteins. Amelogenin is rich in proline, glutamine, leucine and histidine and is therefore hydrophobic. This hydrophobic nature induces the assembly of AMELX protein into nanospheres *in vitro* (Du et al., 2005). However, the C-terminal end is hydrophilic and has been shown to enhance the binding of the peptide to HA *in vitro* (Aoba et al., 1987) and may be a mechanism by which AMELX orientates itself along the forming enamel prisms (Kirkham et al., 2000).

AMELX mutations cause X-linked AI (Lagerstrom et al., 1991). To date, twenty-six variants, including gene deletions have been reported. Heterozygous mutations tend to present in female AI patients as stripes of normal and AI affected enamel due to lyonisation (Berkman and Singer, 1971). The AI phenotype in males is determined by the type and position of the mutation. Large deletions and N-terminal variants cause a hypomaturation AI defect with variable hypoplasia while mutations in the signal peptide and C-terminus cause smooth hypoplastic AI (Hart et al., 2002).

1.9.2 Ameloblastin (*AMBN*)

Ameloblastin (AMBN; MIM *610259) is the second most abundant enamel matrix protein secreted by ameloblasts (Smith, 1998). AMBN is rich in glycine, leucine and proline and localises to the Tomes' processes and the dentino-enamel junction (DEJ) (Krebsbach et al., 1996, MacDougall et al., 2000), but has also been found in pre-odontoblasts, developing tooth roots and craniofacial bone (Fong et al., 1996, Fong et al., 1998, Spahr et al., 2006). AMBN is expressed throughout amelogenesis (Lee et al., 1996) but peaks during the secretory stage (Fukumoto et al., 2004). *AMBN* undergoes alternative splicing to form two isoforms (Krebsbach et al., 1996, Hu et al.,

1997). Porcine ameloblastin is extensively modified and is cleaved upon secretion by MMP20 to form a number of protein products that accumulate within different compartments (Bartlett and Simmer, 1999). Although it is known that AMBN can influence the differentiation and proliferation of ameloblasts (Fukumoto et al., 2004), it is also important in extracellular signalling to induce osteoblast differentiation (Iizuka et al., 2011) and cell adhesion via heparin and fibronectin (Beyeler et al., 2010), as well as in mineralisation (Yamakoshi et al., 2001, Zhang et al., 2011). A 16 amino acid synthetic N-terminal AMBN peptide has been shown to stimulate differentiation of human periodontal ligament cells (Kitagawa et al., 2011) and to promote bone formation *in vitro* (Kitagawa et al., 2016). Evolutionary analysis shows that *AMBN* is most closely related to *AMELX* (*lAMELY*) (Sire et al., 2007), although analysis of the AMBN sequence shows that it consists of a signal peptide and an amelin domain, meaning that it is unlike other proteins.

Two *AMBN* mutations have been reported in AI patients. A large in-frame deletion encompassing exon 6 segregated with recessive hypoplastic AI in a consanguineous family (Poulter et al., 2014c). Scanning electron microscopy (SEM) showed both reduced mineral density and enamel thickness, mirroring the murine *Ambn*$^{-5,6/-5,6}$ model (Poulter et al., 2014c). A second homozygous mutation, thought to alter splicing, was identified in a large cohort of patients with oro-dental disease, using a targeted next-generation sequencing assay (Prasad et al., 2016).

1.9.3 Enamelin (*ENAM*)

Enamelin (ENAM; MIM *606585) is a tooth specific glycoprotein expressed primarily by secretory stage ameloblasts (Hu and Yamakoshi, 2003, Hu et al., 2008). It is successively cleaved from its C-terminus, resulting in numerous products (Fukae et al., 1993, Hu and Yamakoshi, 2003, Lu et al., 2008). The uncleaved protein is found only at the mineralisation front and is known to be involved in enamel crystal extension (Hu et al., 2008). Some ENAM cleavage products, such as the 32kDa fragment identified in pigs, have the ability to bind HA crystals and accumulate within and between the enamel prisms (Hu and

Yamakoshi, 2003). More N-terminal products cannot bind HA. Its exact roles in enamel formation are not yet well defined. Like *AMELX* and *AMBN*, *ENAM* is believed to have diverged from a common ancestral SCPP gene, however *ENAM* is less closely related to *AMELX* and *AMBN* (Sire et al., 2007). Analysis of the RNA sequence shows that the untranslated regions of *ENAM* are much larger than for *AMELX* or *AMBN* and that exon 3 encodes the signal peptide (in *AMELX* and *AMBN*, the signal peptide is encoded by exon 2). The protein sequence does not show similarity to other proteins and the protein itself is much larger than AMELX or AMBN.

The first human mutation identified in *ENAM* caused autosomal dominant AI with a severe smooth hypoplastic phenotype as a result of a dominant-negative effect of aberrant splicing (Rajpar et al., 2001, Mardh et al., 2002). Subsequently, a milder local hypoplastic phenotype was identified in a family with a dominantly inherited nonsense mutation (Mardh et al., 2002). Autosomal recessive inheritance has also been documented for *ENAM* mutations (Hart et al., 2003, Ozdemir et al., 2005a, Chan et al., 2010); homozygous and heterozygous patients present with a severe and a milder, local form respectively (Ozdemir et al., 2005a). To date, twenty-six *ENAM* variants have been reported to cause AI. In addition to these variants, a non-synonymous SNP within *ENAM* has been associated with thinner enamel in the African American population (Daubert et al., 2016).

1.9.4 Matrix metalloproteinase 20 (*MMP20*)

Matrix metalloproteinases (MMPs) influence cell motility by regulating cell interactions and matrix degradation, crucial processes in many aspects of development (VanSaun and Matrisian, 2006). Like other MMPs, MMP20 (MIM *604629) is a zinc-dependent endopeptidase that is secreted in a form that must be cleaved to activate it (Llano et al., 1997). MMP20 is secreted by ameloblasts throughout the secretory stage, concurrent with the enamel matrix proteins, and is responsible for their cleavage, at specific residues, shortly after secretion (Simmer and Hu, 2002). This generates EMP cleavage products with specific and diverse roles in amelogenesis. Therefore, MMP20 has been shown

to be necessary for controlling HA crystal morphology (Prajapati et al., 2016). MMP20 is also capable of cleaving the extracellular domains of cadherins that mediate cell-cell interactions as part of adherens junctions to allow ameloblast cell movement (Guan and Bartlett, 2013, Guan et al., 2016). This may affect amelogenesis since ameloblasts must move in synchronous groups in order to form the typical enamel architecture. Since cadherins are linked to the actin cytoskeleton via catenins, cadherin cleavage releases β-catenin, which can act as a transcription factor and may be important for ameloblast differentiation (Bartlett et al., 2011, Guan et al., 2016).

Many MMPs have been identified throughout the genome, although MMP20 resides within a cluster of nine MMP genes on chromosome 11 (Kawasaki and Suzuki, 2011). Whilst other genes within this cluster encode MMPs that are grouped as collagenases or stromelysins, *MMP20* is not part of either group and its expression has only been detected outside of the tooth at very low levels (Turk et al., 2006). MMP20 consists of a signal peptide, a propeptide, a catalytic domain, a linker and a hemopexin domain, consisting of hemopexin repeats (Ryu et al., 1999). The hemopexin domain is of unknown function, but through homology, it may influence substrate specificity, bind to inhibitors or activation of the pro-enzyme, although it is not essential for enzymatic activity *per se* (Kim et al., 2005). Whilst the protein has low sequence similarity to other MMPs overall (45%), the catalytic domain is known to have a higher level of similarity with other MMPs (59%). This domain is known to bind both calcium and zinc ions at particular residues. Cleavage of MMP20 by KLK4 has been shown to inactivate MMP20 *in vitro* and may remove some or all of the metal ion coordinating residues (Yamakoshi et al., 2013).

Mutations in *MMP20* lead to autosomal recessive hypomaturation AI (Kim et al., 2005). To date, eight variants have been reported that include a variety of missense, nonsense, frameshift and splice site mutations, all resulting in a similar phenotype (Ozdemir et al., 2005b, Papagerakis et al., 2008, Lee et al., 2010, Gasse et al., 2013, Wang et al., 2013b).

1.9.5 Kallikrein-related peptidase 4 (*KLK4*)

Kallikrein-related peptidase 4 (*KLK4*; MIM *603767) encodes a serine protease that is expressed and secreted by ameloblasts in both transition and maturation stages (Hu et al., 2000, Hu et al., 2002, Simmer et al., 2009). Like MMP20, newly secreted KLK4 must be cleaved for activity. *In vitro* experiments have shown that MMP20 can activate newly secreted KLK4 and that KLK4 can inactivate MMP20, potentially explaining the shift in proteinase activity during the transition stage (Yamakoshi et al., 2013).

KLK4 acts to further degrade the enamel proteins cleaved by MMP20 during secretion (Bartlett, 2013). Such activity aids removal of protein from the developing enamel by ameloblasts, allowing the enamel prisms to grow in width (Simmer et al., 2009, Bartlett, 2013). KLK4 is also capable of functioning over the wide pH range that occurs during the maturation stage (Smith, 1998).

In humans, *KLK4* lies within a cluster of fifteen *KLK* genes on chromosome 19 and is believed to have arisen via duplication of *KLK5* (Lundwall, 2013). Hence the KLK4 protein shows the greatest sequence homology with KLK5 (45%). Unlike MMP20, KLK4 is expressed in tissues other than the tooth including the prostate (Nelson et al., 1999), testis, ovary, mammary and salivary gland (Yousef et al., 1999) and skin (Komatsu et al., 2003). The protein consists of a signal peptide, a propeptide and a trypsin-like serine protease domain. Like all KLKs, it contains a catalytic triad of residues critical to its function and a series of disulphide bonds are required to maintain its structure in an active conformation (Debela et al., 2006, Debela et al., 2008).

Human *KLK4* mutations were first identified in autosomal recessive hypomaturation AI (Hart et al., 2004). *KLK4* variants appear to be a rare cause of AI since only three mutations have been reported in a total of four families (Hart et al., 2004, Wright et al., 2011, Wang et al., 2013b, Seymen et al., 2015). The mutations include a frameshift and a stop mutation, both assumed to be null mutations (Hart et al., 2004, Wright et al., 2011, Wang et al., 2013b). The most recently reported variant is a frameshift mutation affecting a codon in the

final exon, which, whilst not subject to nonsense mediated decay (NMD), has been shown to result in greatly reduced protein expression and proteolytic function (Seymen et al., 2015). Despite the expression of KLK4 outside of the tooth, no additional phenotypes have been associated with mutations in *KLK4*.

1.9.6 Other non-syndromic amelogenesis imperfecta genes and pathways

The great variety of other genes known to be mutated in NS AI give snapshots of various important processes and structural features that are essential for normal enamel development. In some cases, the delineation between NS and syndromic AI is unclear, since several genes at first reported as causing NS AI have later been reported to be associated with additional phenotypes.

Mutations in genes encoding proteins that stabilise or are components of hemidesmosomes, such as laminin, alpha 3 (*LAMA3*; MIM *600805), laminin, beta 3 (*LAMB3*; MIM *150310) and collagen type XVII, alpha-1 (*COL17A1*; MIM *113811) are known to cause hypoplastic AI (Murrell et al., 2007, Pasmooij et al., 2007, Yuen et al., 2012, Kim et al., 2013, Poulter et al., 2014b). Hemidesmosomes link epithelial cells, such as ameloblasts, to the basement membrane of extracellular matrices such as the enamel matrix (Nievers et al., 1999). While homozygous mutations in these genes lead to JEB (Aberdam et al., 1994, Pulkkinen et al., 1994a, Pulkkinen et al., 1994b, Kivirikko et al., 1995, McGrath et al., 1995), a syndrome that includes AI, heterozygous carriers of particular mutations are known to have NS hypoplastic AI (Murrell et al., 2007, Pasmooij et al., 2007, Yuen et al., 2012, Kim et al., 2013, Poulter et al., 2014b). More recently, AI-causing mutations in another gene, family with sequence similarity 83, member H (*FAM83H*; MIM *611927), have been shown to affect the keratin cytoskeleton and to impair the formation of desmosomes via altered binding to casein kinase 1 alpha 1 (CSNK1A1; MIM *600505) (Kuga et al., 2016, Wang et al., 2016). Interestingly, all twenty-five of the hypocalcified AI causing truncating mutations identified so far in *FAM83H* are predicted to escape NMD and to and to maintain interaction with CSK1A1 but may instead

affect the normal cellular trafficking of CSK1A1 and phosphorylation of its target proteins within the cell (Wang et al., 2016).

Mutations in an integrin component, integrin, beta 6 (ITGB6; MIM *147558), have also been identified as causes of hypomineralised / hypoplastic AI (Poulter et al., 2014a, Wang et al., 2014b). Integrins mediate cell-cell and cell-extracellular matrix interactions by facilitating interaction with the cytoskeleton (Alberts, 2002). Ameloblast adhesion to the developing enamel matrix may therefore be perturbed in these patients.

Hypocalcified AI-causing mutations have been identified in the gene encoding the WD repeat domain 72 (WDR72; MIM *613214) protein. WDR72 is predicted to be an intracellular vesicle coat protein involved in endocytosis (El-Sayed et al., 2009, Katsura et al., 2014). A *Wdr72*$^{-/-}$ mouse exhibited mislocalisation of solute carrier family 24 (sodium/potassium/calcium exchanger), member 4 (SLC24A4; MIM *609840) protein away from the basement membrane (the ameloblast-enamel matrix interface) (Wang et al., 2015). Mutations in *SLC24A4* are also known to cause AI (Parry et al., 2013) and it is has been suggested that the encoded protein is responsible for calcium transport from the ameloblast in to the enamel matrix during the maturation stage, mainly in RA (Wang et al., 2014a, Bronckers et al., 2015). The switching of ameloblast cellular morphology between the SA and RA forms may be due to pH sensing mechanisms. Mutations in G-protein coupled receptor 68 (*GPR68*; MIM *601404), a pH sensing protein, have recently been shown to cause hypomineralised AI in three families (Parry et al., 2016).

Mutations in chromosome 4 open reading frame 26 (*C4orf26*; MIM *614829), encoding a member of the acidic phosphoprotein family, are also known to cause hypomineralised AI (Parry et al., 2012). Such proteins are known to promote HA crystallisation, and a C4orf26 peptide with a phosphorylated C-terminus was shown to promote nucleation and crystallisation *in vitro* (Parry et al., 2012). However, both the gene and the protein are relatively poorly studied.

Finally, family with sequence similarity 20, member A (FAM20A; MIM *611062)

is one of a family of three human homologues of the *Drosophila* four jointed protein kinase (Ishikawa et al., 2012). Additional family member, FAM20C (MIM *611061), is a Golgi casein kinase responsible for phosphorylating many secreted proteins involved in biomineralisation (Tagliabracci et al., 2012). Mutations in *FAM20C* lead to a severe, syndromic form of AI, Raine syndrome (MIM #259775). *In vitro* expression of FAM20A has shown that it is also located in the Golgi (Ishikawa et al., 2012, Wang et al., 2013a). FAM20A has been shown to potentiate the action of FAM20C *in vitro* and it has been shown that FAM20A promotes the secretion of FAM20C (Ohyama et al., 2016). It is currently debated whether FAM20A is itself a pseudokinase (Cui et al., 2015).

Mutations in *FAM20A* have been shown to cause autosomal recessive AI and gingival fibromatosis syndrome (O'Sullivan et al., 2011) and Enamel Renal Syndrome (ERS) (Jaureguiberry et al., 2012, Wang et al., 2013a). Both syndromes have since been recognised as phenotypic variants of the same condition and are now collectively termed ERS (MIM #204690) (de la Dure-Molla et al., 2014). Patients with *FAM20A* mutations have hypoplastic AI which, in extreme cases, may present as a complete absence of enamel (de la Dure-Molla et al., 2014). The symptoms associated with ERS (de la Dure-Molla et al., 2014, Poulter et al., 2015) suggest that FAM20A, perhaps in combination with FAM20C, phosphorylates proteins involved in enamel formation and other aspects of tooth development, as well as those critical to Ca^{2+} regulation within the kidney.

It is evident that some of the genes implicated in NS AI encode either proteins secreted into the enamel matrix or proteins that aid post translational modification or trafficking of secreted proteins. Successful amelogenesis hinges on the secretion of large amounts of protein, appropriate processing of the proteins by proteases and the endocytosis of the protein products, concurrent with mineralisation of the enamel matrix. Therefore the secretory pathway is of critical importance in amelogenesis. Perturbations of the secretion of mutated proteins, for example, by aberrant intracellular aggregation, as proposed in the *Amelx* Y64H mouse model (Barron et al., 2010), has been shown to result in ER stress and ameloblast cell death (Brookes et al., 2014). Such a mechanism of

disease for AI may be more important than previously realised and may partially explain the cause of disease of AI-causing mutations with dominant inheritance patterns.

1.9.7 Leiden Open Variant Database of NS AI mutations

In order to consolidate the large number of NS AI-causing mutations reported in the literature into one easily accessed, fully referenced resource for benefit of the scientific community, a Leiden Open Variant Database (LOVD) has been established. This resource is accessible at http://dna2.leeds.ac.uk/LOVD/ and includes over 160 variants in a total of 14 genes (Date accessed 1st July 2016). The webpage and LOVD hosting was created and is managed by Miss Agne Antanaviciute (University of Leeds) and has been curated by the author, for variants in all of the genes included therein, except for *FAM20A*, which is curated by Dr. James Poulter (University of Leeds).

1.10 Heimler syndrome

Heimler syndrome (HS; MIM #234580 and #616617) was first described as a combination of sensory neural hearing loss (SNHL), enamel hypoplasia and nail abnormalities (Heimler et al., 1991). Two siblings were originally documented and at the start of this study, three further reports of three families (five cases) had been described (Tischkowitz et al., 1999, Pollak et al., 2003, Ong et al., 2006). These reports are summarised in Table 1.1. Due to reports of HS in siblings and in both sexes, born from unaffected parents, HS was thought to be an AR genetic condition.

Descriptions of the phenotype suggest that the enamel defects affect the permanent dentition but not the deciduous. Some reports also detailed that only particular teeth were affected (Figure 1.12) (Heimler et al., 1991, Ong et al., 2006). The SNHL described was moderate to profound in severity, was generally bilateral and appeared to develop early in childhood rather than to be present from birth. Reports of nail abnormalities, such as leukonychia and Beau's lines, were not a consistent finding. Both leukonychia, the presence of

white spots on the nail and Beau's lines, nail ridging, are common abnormalities often related to trauma and infection, respectively (Fawcett et al., 2004). Therefore the presence of these symptoms may be coincidental.

Of the HS patients described in the literature, case 1-2 (Table 1.1) was later reported to have developed macular dystrophy (MD) at the age of 29 (Lima et al., 2011). This description of an eye phenotype was the only one reported in the seven HS patients, and the brother of case 1-2 (case 1-1), who also has HS, was not reported to have MD. However, all of the other individuals reported at that time were under 16 years old and therefore the eye phenotype reported only in case 1-2 may suggest that HS is a progressive disease that includes MD as a later onset symptom.

At the start of this study, no genetic cause for HS had been reported. Speculation by case report authors as to the cause of the HS led to the suggestion of a ciliary defect, since the tissues affected have a common ectodermal origin (Lima et al., 2011).

Despite the paucity of HS case reports, a literature search revealed a number of reports of patients diagnosed with Usher syndrome (USH; Section 1.11) who also have enamel defects (summarised in Table 1.1). These reports highlight the phenotypic overlap of the two syndromes and the possibility that HS cases with an eye phenotype may be misdiagnosed as USH. Alternatively, as only one report on HS detailed an enamel phenotype in the primary dentition, while other HS reports stated that only particular teeth were affected, the possibility that the enamel phenotype was not AI could not be ruled out. If this was the case, HS cases may have been USH cases with a coincidental enamel defect due to MIH or an environmental / life history cause (Section 1.8.2).

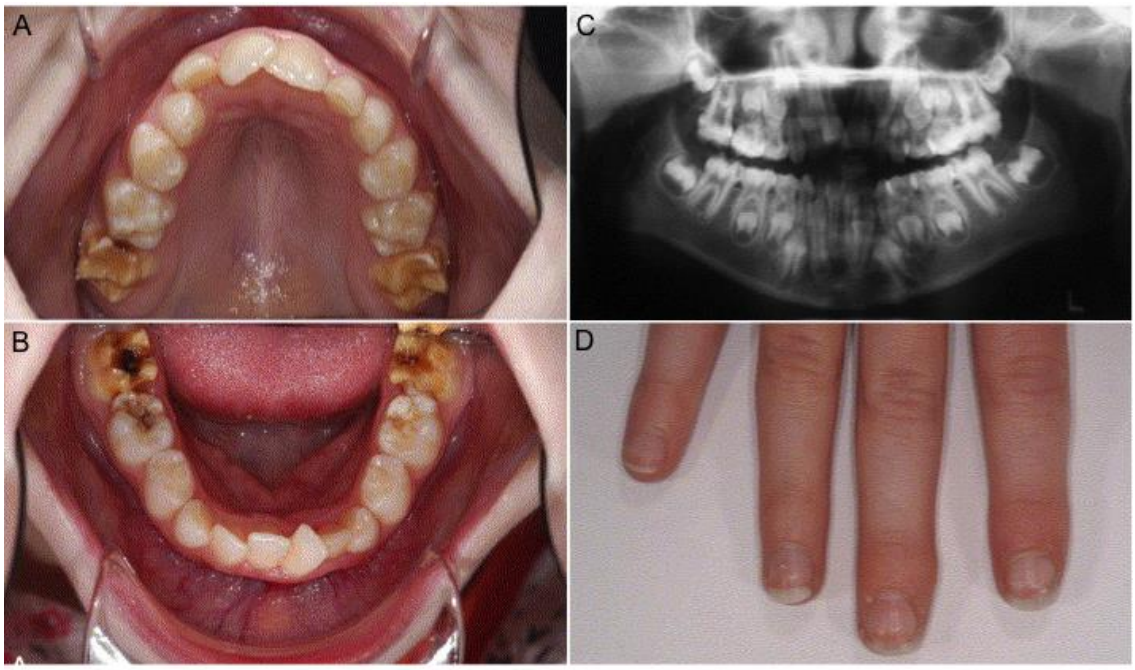


Figure 1.12: Clinical details of an HS patient.

Patient is labelled 4-1 in Table 1.1. A and B: Photographs of maxillary (A) and mandibular (B) dentitions of an HS patient at 11 years old showing hypoplastic enamel confined to the first molars and the lower second pre-molar teeth. C: Dental radiographs showing the developing dentition and the radiographic appearance of the hypoplastic first molar teeth. D: Leukonychia of the fingernails. Images reproduced with permission from Ong et al. (2006).

Case no.	Sex	Hearing loss diagnosis		Dentition		Vision	Nail abnormalities		Parental origin
		Age	Type; severity	Primary	Permanent		Beau's lines	Leukonychia	
1-1	M	18 mo	B; S to P	Normal	Hypoplastic	Normal	Yes - T	Yes - F	Mixed European
1-2	F	30 mo	B; P	Normal	Hypoplastic	Adult on-set macular dystophy diagnosed at 29	Yes - T	Yes - F	Mixed European
2	F	7 yrs	U-L; P	Normal	Hypomineralised	Normal	Yes - F+T	No	N/R; Caucasian
3-1	F	30 mo	B; Mo to S	Normal	Hypoplastic	Normal	Yes - F+T	No	Mixed European
3-2	F	14 mo	B; S to P	Normal	Hypoplastic	Normal	No	No	Mixed European
4-1	F	3 yrs	B; Mo to S	Normal	Hypoplastic / hypomineralised	Normal	Yes - T	Yes - F	N/R; Caucasian
4-2	F	3 yrs	B; Mo to S	Normal	Hypoplastic / hypomineralised	Normal	Yes - T	Yes - F	N/R; Caucasian
5	M	<2 yrs	B; S	N/R	Hypoplastic	RP, bone spicule formation	N/R	N/R	N/R
6	F	<5 yrs	B; Mo to S	N/R	Hypoplastic	Maculopathy, pigmentary clumping	N/R	N/R	N/R
7-1	M	4 yrs	B; Mi to S	N/R	Unerupted	RPE granularity, esp in macula; white lesions on RPE	N/R	N/R	N/R
7-2	F	3 yrs	B; Mi to S	N/R	Unerupted	RPE granularity, esp in macula; white lesions on RPE	N/R	N/R	N/R
7-3	M	26 mo	B; S	N/R	Unerupted	Abnormal heavy RPE granularity	N/R	N/R	N/R
8	N/R	N/R	B; S to P	Enamel hypoplasia of unknown dentition		Nystagmus since early infancy; RP diagnosed at 3	N/R	N/R	Mixed European
9	N/R	N/R	B; S to P	Enamel hypoplasia of unknown dentition		Nystagmus since early infancy; RP diagnosed at 5	N/R	N/R	Mixed European
10	F	N/R	B; P	Normal	Hypoplastic	Diagnosis: Usher syndrome type 1	N/R	N/R	European
11	M	N/R	"Deaf"	Hypoplastic	Hypoplastic	Diagnosis: Usher syndrome type 2	N/R	N/R	N/R

Table 1.1: Summary of case reports published as Heimler syndrome or as Usher syndrome and enamel defects.

Cases 1-1 to 4-2 were reported as Heimler syndrome. Cases 5 to 11 were reported as Usher syndrome and enamel defects. Familial cases are indicated by numbering, i.e. cases 1-1 and 1-2 are siblings, cases 2 and 3-1 are unrelated. Cases 4-1 and 4-2 are monozygotic twins. Abbreviations used: B bilateral, esp especially, F female, Fi fingernails, L left, M male, Mi mild, Mo moderate, mo months, N/R not reported, RPE retinal pigment epithelium, S severe, P Profound, T toenails, U unilateral, yrs years. References: cases 1-1 and 1-2: Heimler et al. (1991), Lima et al. (2011); case 2: Tischkowitz et al. (1999); cases 3-1 and 3-2: Pollak et al. (2003); cases 4-1 and 4-2: Ong et al. (2006); cases 5 and 6: Bateman et al. (1980); cases 7-1, 7-2, 7-3 Innis et al. (1998); cases 8 and 9: Pieke-Dahl et al. (2000); case 10: Balmer and Fayle (2007); case 11: de la Pena and Valea (2011).

1.11 Usher syndrome

Usher syndrome (USH; MIM PS276900) is a combination of SNHL and retinitis pigmentosa (RP) with a reported prevalence ranging from 1 in 6,000 to 1 in 50,000 (Keats and Corey, 1999, Kimberling et al., 2010). USH is thought to account for more than 50% of cases of deaf blindness in adults (Boughman et al., 1983) and is inherited in an AR manner. There are three distinct types of the syndrome. USH type I (USH1) is characterised by congenital, bilateral, profound SNHL and vestibular dysfunction. Progressive RP typically develops in childhood. USH type II (USH2) is the most common form and is characterised by prelingual moderate to severe SNHL and normal vestibular function. Progressive RP usually develops within the second decade. USH type III (USH3) is rare except in particular populations such as the Finnish and Ashkenazi Jewish. USH3 is characterised by variable onset and progression of the SNHL phenotype, variable vestibular dysfunction and variable RP onset. Atypical USH is also used to describe atypical phenotypes resulting from mutations in genes that are more often associated with USH1 or USH2. Mutations in eleven genes have been identified (Appendix 3) and other modifiers and genetic loci are known to exist (Mathur and Yang, 2015).

USH genes encode a wide variety of proteins essential for the correct development of the hair bundle in the cells of the cochlea and in the vestibular organ of the ear (Ahmed et al., 2013). The proteins are also important for the maintenance of the transport links within the retina (Millan et al., 2011). Since both the kinocilium, the structure that directs the development of the hair cell bundles, and the photoreceptors of the retina are modified cilia, USH is regarded as a ciliopathy (Goetz and Anderson, 2010, Wheway et al., 2014). The proteins implicated in USH1 and those implicated in USH2, are each known to form USH1 and USH2 protein complexes *in vivo* (Michalski et al., 2007, Grillet et al., 2009). These complexes are thought to bind to each other to form one large complex and to interact in the inner ear and retina to form a large network (Reviewed by Mathur and Yang (2015)).

The gene most commonly found to be mutated in USH is usherin (*USH2A*; MIM *608400) which leads to an USH2 phenotype (Aller et al., 2006, Baux et al., 2007, Dreyer et al., 2008). The most commonly mutated USH1 gene is myosin VIIA (*MYO7A*; MIM *276903) (Jaijo et al., 2006, Roux et al., 2011).

1.12 DNA sequencing

The development of the chain termination method of DNA sequencing (Sanger et al., 1977) and further improvement through the use of fluorescent dyes (Smith et al., 1986), capillary electrophoresis and laser detection of sequencing reads (Ruiz-Martinez et al., 1993) has meant that automated identification of 500-1000bp of DNA sequence in a few hours is now routine. This and the development of PCR (Saiki et al., 1988) have combined to allow the specific amplification of a DNA region of choice which can then undergo incorporation of four different, fluorescently labelled dideoxyribonucleotide tri-phosphates (ddNTPs) to provide chain terminated products for sequence analysis. Modern sequencers separate fragments using a capillary column filled with a replaceable polymer matrix, and detect the light emitted by the fluorescent ddNTPs upon laser excitation as they exit the capillary column. These four different emissions are then plotted graphically and the DNA sequence can be determined.

1.13 Next generation sequencing technologies

Chain termination capillary sequencing was used to complete sequencing of the human genome over thirteen years and at a cost of $2.7 billion, in 2001 (Lander et al., 2001, McPherson et al., 2001, Sachidanandam et al., 2001). Today a genome can be sequenced in a few days and costs around $1500. This is due to the development of massively parallel sequencing, also called next generation sequencing (NGS). NGS technologies include both whole genome sequencing (WGS, Section 1.13.1) and whole exome sequencing (WES, Section 1.13.2).

Although many sequencing platforms have been developed, all simultaneously sequence multiple DNA molecules within a single slide or flow cell. Leeds Next Generation Sequencing Facility uses both the Illumina HiSeq 2500 and 3000, which use a clonal bridge amplification and “sequencing by synthesis” reversible dye termination method (Figure 1.13) (Bentley et al., 2008).

Illumina sequencing is achieved by shearing target DNA into fragments, ligating adapters to each end, then adding these to a flow cell precoated with oligonucleotides complementary to the adaptors. Each DNA fragment, attached to the flow cell at both ends, forms a “bridge” which is then PCR amplified. Next the resulting strands are denatured and allowed to attach to adjacent oligonucleotides to form two attached copy strands in close proximity. Further rounds of the same process form millions of clusters of identical molecules, attached as spots to the flow cell. Sequencing proceeds by the use of four different fluorescently labelled reversible terminators. At each round of synthesis, one new nucleotide is incorporated, and molecules within each cluster, being identical copies, all fluoresce with the same colour. Thus each spot fluoresces with a colour corresponding to the next base of its sequence at each round of extension. An image of the flow cell surface logs which base has been incorporated in each spot. Afterwards, each incorporated nucleotide’s terminating 3′ end and fluorescence is chemically inactivated, so that a further round of sequencing can take place. The resulting sequence for each molecule is termed a read. These reads can be aligned to a reference sequence.

For studies investigating the genetic causes of Mendelian diseases, variations from the reference sequence are identified. These variants are then filtered against databases of known variation with reference to the frequency of the disease and the mode of inheritance. Refinement of the remaining variants can be achieved by the sequencing of multiple family members, use of mutation prediction software and by the individual Sanger sequencing and segregation analysis of candidates. The first Mendelian disease mutations identified by this method were reported in 2009 (Ng et al., 2010). NGS has shifted the rate limited step and challenge from sequence acquisition to the processing,

interpretation and storage of the vast amounts of data generated. It has also raised ethical considerations around secondary and incidental findings.

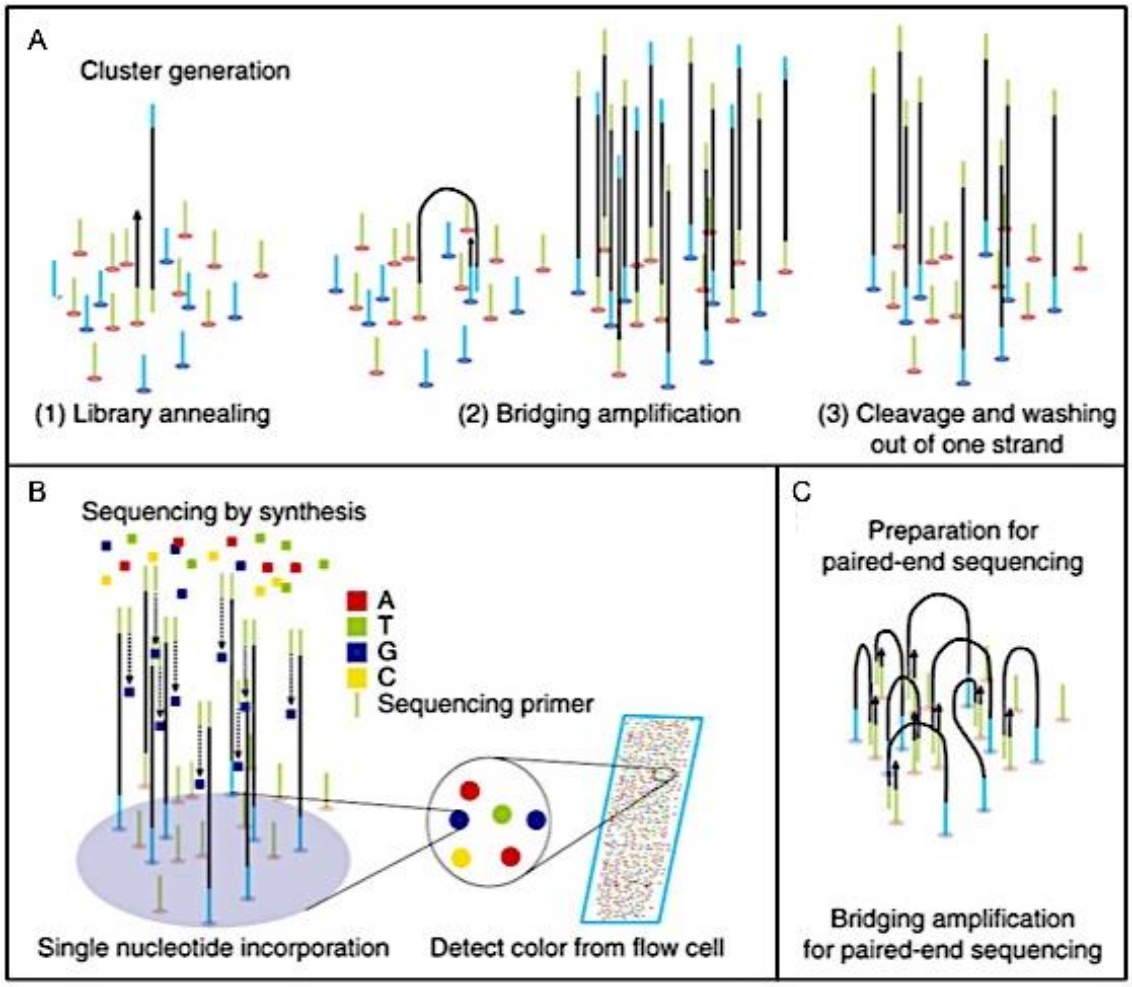


Figure 1.13: The Illumina clonal bridge amplification and sequencing by synthesis method.

A: Cluster generation on the surface of the flow cell. (1) Denatured libraries are annealed to the short oligonucleotides on the surface of the flow cell. (2) Bridging amplification generates clusters. (3) One strand from the double-strand DNA library is cleaved and washed out for unidirectional sequencing. B: Primers for inserts are annealed for the sequencing of the insert DNA. In each sequencing cycle, protected and fluorescently labelled A, T, G and C bases are applied. After the addition of each nucleotide, the sequencing reaction is stopped, and the image is taken. Because the newly added nucleotides within each cluster are identical, the signal is high enough to be detected by a light sensor. After the image is taken, the protection group and the fluorescent molecules are removed. C: When the first-strand sequencing reaction is finished, the synthesised strand is removed and the process is repeated for the opposite strand. Image reproduced with permission from Shin et al. (2014).

1.13.1 Whole Genome Sequencing (WGS)

WGS is the sequencing of the entire DNA sequence of an organism, the genome. Recently, use of WGS has increased due to falling costs. In addition, the absence of any capture step means that this approach may provide better coverage of the entire exome than WES (Belkadi et al., 2015). Regions of low or high GC content and complex regions, such as repeats, centromeres and telomeres, may not be well covered (Aird et al., 2011). At present, the scant annotation of the majority of non-protein coding regions limits the interpretation and prioritisation of non-coding changes (Kircher et al., 2014, Ritchie et al., 2014, Ionita-Laza et al., 2016). However, even low coverage WGS is useful for the assessment of copy number variants (CNVs) and other structural variants (Hayes et al., 2013).

1.13.2 Whole Exome Sequencing (WES)

WES traditionally sequences only the protein coding regions of the genome, the exome. However some more recently developed reagents also target functional non-protein coding sequences such as microRNAs. WES focuses sequence acquisition and therefore increases coverage of the approximately 1.5% of the genome that represents the exome. This is achieved by specifically capturing only these regions using biotinylated cRNA baits, which hybridise to target sequences by complementary base pairing (Gnirke et al., 2009). WES has become the preferred method for rare disease gene discovery since it is less costly to undertake, produces a greater read depth for targeted regions and produces a smaller volume of, and less complex, data than WGS.

Since pathogenic variants are found most commonly in the most conserved regions of the genome, it seems likely that the exome will contain the majority of the variants with functional consequences for proteins. As of 2008, over 99% of the mutations identified as causing a Mendelian disease were identified within the exome (Ng et al., 2008), although this must reflect, in part, the sole focus on the exome in the majority in studies. WES has been successfully used to identify genes in which mutations cause AI, with or without prior use of linkage

analysis, in a number of studies (Poulter et al., 2014a, Poulter et al., 2014b, Poulter et al., 2014c, Huckert et al., 2015).

1.14 The aims of this study

The first aim of this study was to identify the gene(s) mutated in HS and to determine the localisation of the encoded protein in the tissues affected.

WES of ten families segregating HS or HS-like phenotypes in an AR manner, identified biallelic mutations in *PEX1* and *PEX6* as the cause of disease in a total of seven families. Both PEX1 and PEX6 are well-characterised proteins with roles in peroxisomal function. Two of the other families had biallelic mutations in *USH2A* and were assumed to have USH, not HS. For the one remaining family, a heterozygous deletion within *MYO7A* was identified but remains unconfirmed. The localisation of PEX1 and PEX6 proteins within the murine retina was determined and attempts were made to determine the proteins' localisation within the murine incisor, however technical difficulties meant that this was not possible.

The second aim of this study was to identify gene(s) mutated in NS AI and to characterise the enamel defect in the patient(s).

WES of one family segregating NS AD hypomineralised AI revealed a heterozygous deletion of *AMTN* exons 3-6. Characterisation of the enamel phenotype revealed differences between the human phenotype and that of the *Amtn* mouse models. A dominant gain of function was predicted to be the mode of disease.

2 Materials and Methods

All reagents in this section were obtained from Sigma (St. Louis, MO, USA) unless otherwise stated.

2.1 Solutions

2.1.1 Red cell lysis solution

155mM	Ammonium chloride
10mM	Potassium bicarbonate
1mM	Ethylenediaminetetraacetic acid (EDTA; pH 8.0)

2.1.2 White cell lysis solution

25mM	EDTA (pH 8.0)
2% (w/v)	Sodium dodecyl sulphate (SDS)

2.1.3 1x Tris-EDTA (TE) buffer (pH 8.0)

10mM	Tris HCl (pH 8.0)
1mM	EDTA (pH 8.0)

2.1.4 Cell lysis buffer for DNA extraction

10mM	Tris HCl (pH 8.0)
100mM	EDTA (pH 8.0)
20μg/ml	RNAase A
0.25% (w/v)	Sodium dodecyl sulphate (SDS)

2.1.5 50x Tris-Acetate-EDTA (TAE) electrophoresis buffer

2M	Tris HCl
50mM	EDTA (pH 8.0)
0.97M	Glacial acetic acid

2.1.6 10x Gel loading dye

6x	TAE
20% (w/v)	Ficoll 400
0.1% (w/v)	Bromophenol blue
0.2% (w/v)	Xylene cyanol

2.1.7 Radio Immunoprecipitation Assay (RIPA) buffer

50mM	Tris HCl (pH 8.0)
150mM	Sodium chloride
1% (v/v)	IGEPAL® CA-630
0.5% (w/v)	Sodium deoxycholate
0.1% (w/v)	SDS

2.1.8 1x Tris Buffered Saline (TBS)

150mM	Sodium chloride
20mM	Tris HCl (pH 7.4)

2.1.9 1x TBS Tween (TBST)

1x	TBS
0.05%	Tween 20

2.1.10 Scott's tap water substitute

0.2% (w/v)	Sodium hydrogen carbonate
2% (w/v)	Magnesium sulphate

2.2 Patients

All samples were obtained with informed consent for research use under ethical approval from the Yorkshire & The Humber - South Yorkshire Research Ethics Committee Leeds (reference 13/YH/0028). The diagnostic criteria for Heimler syndrome and amelogenesis imperfecta are detailed in Sections 2.2.1 and 2.2.2 respectively.

2.2.1 Heimler syndrome patients

DNA samples from patients diagnosed with Heimler syndrome (HS; MIM #234580, #616617), along with samples from other family members, were obtained from a group of local, national and international collaborators. HS was suspected in patients who were reported to have both SNHL and AI. Since published reports of HS at the commencement of this study included only one individual who had developed retinal dystrophy in adulthood (Lima et al., 2011), this was not regarded as a diagnostic criterion.

Due possibly to the rarity of HS, and perhaps also in part because tooth abnormalities were diagnosed by a different set of healthcare professionals, in most cases patients had been diagnosed initially with Usher syndrome (USH; MIM PS276900) and amelogenesis imperfecta (AI; MIM PS104500).

2.2.2 Amelogenesis imperfecta patients

DNA samples from patients diagnosed with AI, along with samples from other family members, were obtained from a group of local, national and international collaborators. AI was diagnosed by dental practitioners based on the physical appearance of the dentition. Dental X-ray assisted in the diagnosis of the specific subtype of AI.

2.3 DNA samples

DNA from any blood sample received was extracted by Yorkshire Regional Genetics (St James's University Hospital, Leeds, UK) using a salt precipitation protocol (Section 2.3.1) or if blood had been previously frozen, a phenol-chloroform extraction was performed (Section 2.3.2). DNA from saliva samples obtained using Oragene® DNA collection kits (DNA Genotek Inc., Ottawa, Canada) was extracted according to the manufacturer's instructions but with an increased input volume of 750μl and 1.5 times the volumes of other reagents used. Pellets were dissolved in 1x TE buffer (Section 2.1.3).

2.3.1 Salt precipitation

A salt precipitation technique was performed to extract DNA from fresh blood samples. Briefly, 3ml of whole blood was aliquoted into a polypropylene tube and 9ml of red cell lysis solution (Section 2.1.1) added. The samples were shaken for 10 minutes and then centrifuged at 2,000 x g for 10 minutes. The supernatant was removed to leave the white cell pellet containing DNA. The pellet was resuspended in 3ml white cell lysis solution (Section 2.1.2) and the cells lysed by vigorous pipetting. To remove any contaminating protein, 1ml of protein precipitation solution (10M ammonium acetate) was added, the samples mixed for 20 seconds and then centrifuged for 10 minutes at 2000 x g. The supernatant was transferred to a new tube and the DNA was precipitated using 0.7x volume of isopropanol. Following centrifugation for 10 minutes at 2000 x g, the pellet was washed twice in 70% ethanol. The pellet was air dried and dissolved in 1x TE buffer (Section 2.1.3).

2.3.2 Phenol-chloroform extraction

A phenol-chloroform extraction was performed to extract DNA from blood that had been frozen. Briefly, 3ml of whole blood was aliquoted into polypropylene tubes and 9ml red cell lysis solultion (Section 2.1.1) was added. After shaking for 10 minutes, samples were centrifuged at 2000 x g for 10 minutes. The supernatant was removed to leave the white cell pellet to which 500μl of cell lysis buffer for DNA extraction (Section 2.1.4) was added. After an hour of incubation at 37^{o}C, proteinase K was added to a final concentration of 100μg/ml followed by incubation at 55^{o}C for an hour. 500μl of 1:1 phenol (pH 8):chloroform was then added and the tube inverted several times prior to centrifugation at 3100 x g for 10 minutes. The aqueous phase was removed to a fresh tube and an equal volume of chloroform added to it. The tube was again inverted several times and centrifuged at 3100 x g for 5 minutes. After transferring the aqueous phase to a fresh tube, sodium chloride was added to a final concentration of 0.2M and 2 volumes of ethanol were added. The sample was mixed and centrifuged at 3100 x g for 5 minutes. The supernatant was removed and 1ml 75% ethanol was added, followed by centrifugation at 3100 x

g for 5 minutes. After removal of the supernatant, the pellet was dried and dissolved in 1x TE buffer (Section 2.1.3)

2.3.3 DNA size fractionation using agarose gel electrophoresis

DNA size fractionation using agarose gel electrophoresis was carried out according to the protocol outlined in Sambrook and Russell (2000). Molecular biology grade agarose powder (Bioline, London, UK) was dissolved in 1x TAE (Section 2.1.5) to a concentration of 1.5% (w/v). Ethidium bromide was added to a final concentration of 500ng/ml. DNA samples were mixed with 10x gel loading dye to a final concentration of 1x. The samples were loaded into the wells at the cathode alongside an EasyLadder I (Bioline) or another appropriate molecular size standard. An electric field was applied for 40 minutes with a potential difference of 120V The resulting ethidium bromide labelled bands within the gel were visualised on a Bio-Rad molecular imager gel documentation system with a UV transilluminator and displayed using Image Lab (v.4.0) analysis software (Bio-Rad, Hemel Hempstead, UK).

2.4 Next generation (massively parallel) sequencing

2.4.1 Whole Exome Sequencing library preparation

DNA samples considered for WES were quantified using the Qubit® dsDNA BR assay kit (Life Technologies, Carlsbad, CA, USA). Two library preparation protocols were followed, one with a 3μg input and one with a 200ng input. Both protocols follow the original SureSelectXT method (Agilent, Santa Clara, CA, USA) based on the use of solution hybrid selection technology developed by Gnirke et al. (2009). Figure 2.1 shows the main steps in the SureSelectXT protocol.

Samples were then sequenced using the sequencing by synthesis method, using fluorescent reversible terminator deoxyribonucleotides, developed by Bentley et al. (2008) (Figure 1.13).

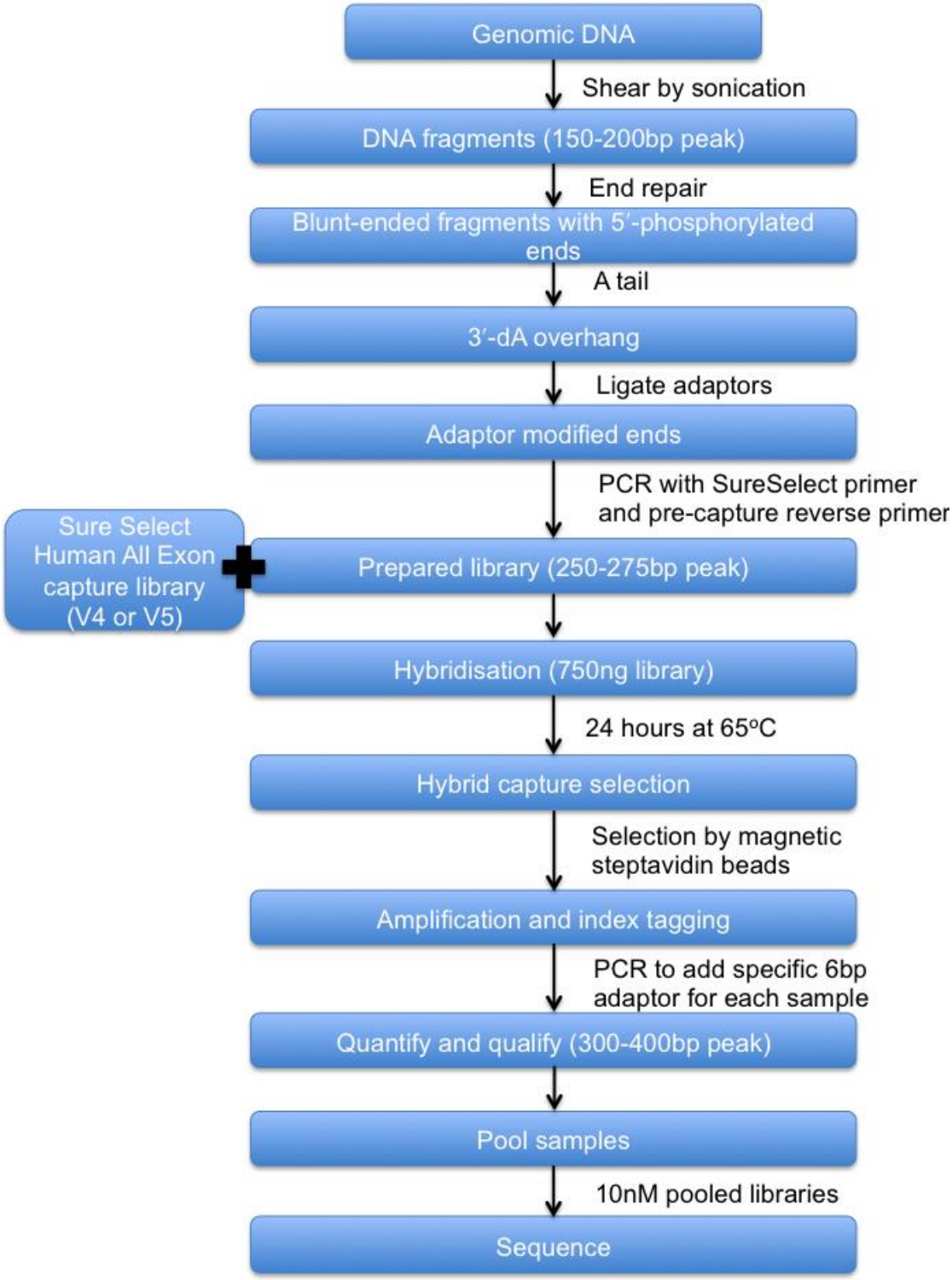


Figure 2.1: The Agilent SureSelectXT method.

2.4.1.1 3μg protocol

The manufacturer's guidelines were followed with a few modifications, although the protocol is also briefly described here. 3μg of DNA was diluted to 250μl in

1x TE buffer (Section 2.1.3) and sheared using the Covaris S2 Adaptive Focused Acoustics sonicator (Covaris, Woburn, MA, USA) at a standard shear setting of 19.9% duty cycle, intensity 8, 1000 cycles/burst, 25 repetitions at 20ºC to form fragments with a peak size of between 150-200bp. Sheared samples were purified and size selected with 1.8x volume of AMPure beads (Beckman Coulter, Brea, California, USA) as described in Agilent's SureSelectXT kit instructions. The following steps use reagents from the SureSelectXT kit (Agilent) unless otherwise stated. End repair was performed using T4 DNA polymerase, Klenow polymerase and T4 polynucleotide kinase. 3′ dA tailing was then performed using dATP and Exo(-) Klenow polymerase. Adaptor ligation was performed using T4 DNA ligase. After each step, samples were treated with AMPure beads as described in Agilent's SureSelectXT kit instructions. A pre-capture PCR amplification step was performed using Herculase II Fusion polymerase (Agilent) using the cycling conditions specified by the protocol to a total of 6 cycles. Samples were then treated with AMPure beads as described previously. 750ng of the prepared DNA was dissolved in 3.4μl distilled nuclease free water by pipetting. If necessary, prepared libraries were concentrated by dehydration using a Savant SpeedVac 110 vacuum concentrator (ThermoFisher Scientific, Waltham, MA, USA). Hybridisation was carried out with SureSelect Human All Exon v4 or v5 capture reagent (SSV4 or SSV5; Agilent) at 65ºC for 24 hours. Hybridised DNA was captured by use of Dynabeads® MyOne™ Streptavidin T1 (Life Technologies) streptavidin coated beads. Each sample was uniquely indexed using 6bp indexes in a post capture PCR amplification step carried out using the cycling conditions specified by the protocol to a total of 12 cycles. Samples were then treated with AMPure beads as described previously. Equal molar quantities of each captured library were pooled and diluted to 10nM in nuclease free water. Pooled libraries were stored at -20ºC until they were prepared for sequencing and sequenced by the Leeds St James' University Hospital Next Generation Sequencing Facility (Leeds, UK, http://dna.leeds.ac.uk/genomics/) using the Illumina HiSeq2500 platform (Illumina, San Diego, CA, USA) in 100bp paired end read mode.

For quality control and for information necessary to continue the protocol, sizing and quantification of DNA was carried out throughout the library preparation using the DNA1000 or DNAHiSensitivity assay kit (Agilent) using the 2100 Bioanalyzer (Agilent). Quantification was carried out using the 2100 Bioanalyzer Expert software (Agilent) and was achieved by integrating the area under the peak.

2.4.1.2 200ng protocol

Subsequent to the release of the SureSelectXT kit B4 protocol (Agilent), the 200ng protocol was used and the manufacturer's guidelines were followed with one modification; post capture PCR was carried out using the cycling conditions specified by the protocol to a total of 10 cycles. Pooled libraries were prepared for sequencing and sequenced by the Leeds St James' University Hospital Next Generation Sequencing Facility using the Illumina HiSeq3000 platform (Illumina) in 150bp paired end read mode.

2.4.2 Whole exome sequencing alignment and variant calling

Examples of simplified alignment and variant calling commands can be found in Appendix 4.

Fastq files were aligned to the human reference genome (Genome Reference Consortium human 37 (GRCh37)) using Novoalign software (Novocraft Technologies, Selangor, Malaysia) or BWA (Li and Durbin, 2009). The resulting alignment was processed in the SAM/BAM format using the SAMtools, Picard (http://picard.sourceforge.net) and GATK programs to correct alignments around indel sites and to mark potential PCR duplicates (Li et al., 2009, McKenna et al., 2010, DePristo et al., 2011).

Indel and single-nucleotide variants (SNVs) were called in the variant call format (VCF) using the Haplotype Caller function of the GATK program. Recalibration of indel and SNVs was performed separately with hard filtering using the following parameters to filter variants:

Quality by depth (QD) <2.0, Fisher Strand (FS) >60.0, Mapping quality (MQ) < 40.0, Mapping Quality Rank Sum <-12.5 (For SNVs)
Quality by depth (QD) <2.0, Fisher Strand (FS) >200.0 (For indels)
The outputs were then combined to produce one VCF file (Danecek et al., 2011).

2.4.3 Whole exome sequencing variant filtering

Examples of simplified variant filtering commands can be found in Appendix 4.

VCF files resulting from alignment and variant calling were filtered as standard using the following databases of known variation:

The National Center for Biotechnology Information's (NCBI's) database of single nucleotide polymorphisms and multiple small-scale variations (dbSNP) version 142 (Available at http://www.ncbi.nlm.nih.gov/SNP/) (Sherry et al., 2001).

The NHLBI Exome Sequencing Project's Exome Variant Server (EVS; available at http://evs.gs.washington.edu/EVS/).

The Broad Institute's Exome Aggregation Consortium (ExAc; available at http://exac.broadinstitute.org/) version 0.3 (Exome Aggregation Consortium et al., 2016).

The VCF hacks package (available at http://sourceforge.net/projects/vcfhacks) was used to carry out filtering and annotation of variant files. The cut off values for variant filtering were set according to the mode of disease inheritance. For rare recessive disease, variants with a minor allele frequency (MAF) of ≥1% were filtered. For rare dominant disease, variants with a MAF of ≥0.1% were filtered.

Synonymous, non-coding, intronic and intergenic variants, other than those potentially affecting splice donor or acceptor sites were filtered. Variants potentially affecting splice donor or acceptor sites were classed as those

residing within the intron up to 8bp from the splice junction or within the exon up to 3bp from the splice junction. Variants affecting known transcription factor binding sites or regulatory regions were also retained. The variants retained by using this filtering strategy are referred to in Chapters 3 and 4 as “functional variants”.

Variant filtration was also based on mode of inheritance, i.e. for recessive disease, only biallelic variants were retained, whereas for dominant disease, homozygous variants were removed. In cases where sequence data was available from multiple family members, variants were filtered based on their presence in affected family members and for dominant disease, their absence in unaffected family members. Figure 2.2 details possible strategies for variant filtering.

2.4.4 Whole exome sequencing variant annotation

Examples of simplified variant annotation commands can be found in Appendix 4.

VCF files were annotated using Ensembl’s Variant Effect Predictor version 79 (VEP; available at http://www.ensembl.org/info/docs/tools/vep/index.html).
Files were ranked according to their Combined Annotation Dependent Depletion (CADD) version 1.3 scaled C-score (Section 2.8.4.4) (Kircher et al., 2014).

2.4.5 Whole exome sequencing depth of coverage

An example of a command to obtain depth of coverage statistics can be found in Appendix 4.

Indel realigned and base recalibrated bam files were used to assess the depth of coverage using GATK’s Depth Of Coverage tool. Assessment of coverage was based on the use of the appropriate SureSelect Human All Exon captured regions bed file as the interval file (made available from Agilent).

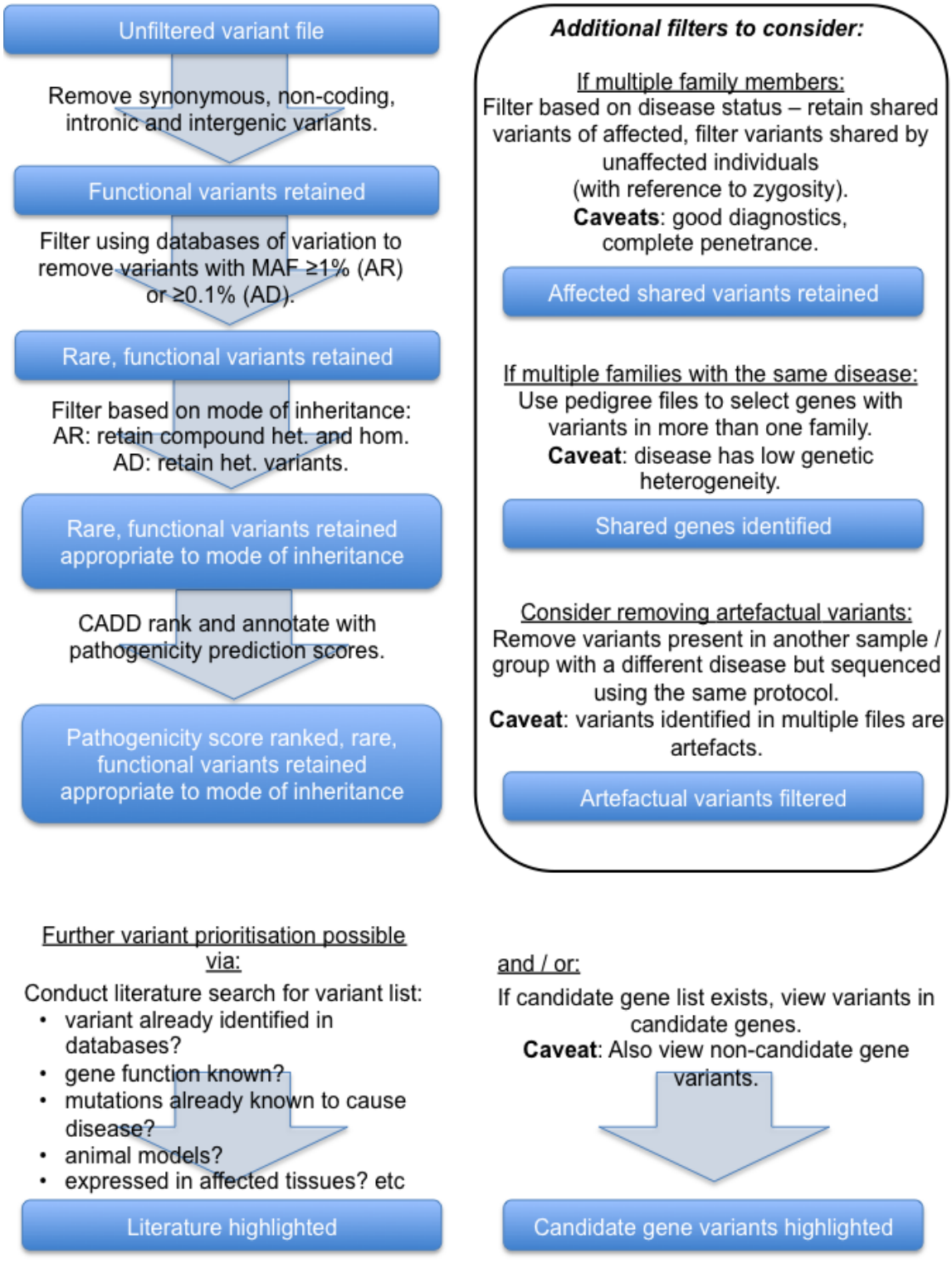


Figure 2.2: WES variant filtration and prioritisation strategies.

The pipeline on the left details the basic strategy for variant filtration, however other possible filters are detailed in the panel to the right. Variant prioritisation strategies are shown at the bottom of the figure. Prioritisation is based initially on the predicted pathogenicity of the variants identified, but can be modified by both literature searches and the existence of candidate genes.

Mean coverage, as well as the percentage of bases covered at minimum depths of 5, 10, 15, 20, 25 and 30x, were assessed.

2.4.6 Whole exome sequencing based detection of copy number variation

In order to identify copy number variation too large to be detected by the variant calling protocol described above (Section 2.4.2), both ExomeDepth (Plagnol et al., 2012) and FishingCNV (Shi and Majewski, 2013) were utilised, depending upon the availability of appropriate control data.

For sequencing batches consisting of six or more unrelated samples, aligned sequences, processed to produce an indel realigned, indexed bam file, were analysed using the R package, ExomeDepth (available at https://cran.r-project.org/web/packages/ExomeDepth/index.html) (Plagnol et al., 2012). ExomeDepth compares the read depths of all captured exons of the test exome with the read depths of a reference set of 5-10 exomes. The reference exomes were from unrelated individuals whose DNA had been processed in the same WES batches for each exome analysed, using identical conditions as for the test exome. The resulting CNV call .csv file output was annotated to identify common CNVs (Conrad et al., 2010). CNV calls were ranked, with calls not annotated as a common CNV prioritised and the highest rank assigned to the calls with the highest Bayes factor; the $\log_{10}$ of the likelihood ratio of data for the CNV call divided by that of the normal copy number call (Plagnol et al., 2012).

Graphical representations of ExomeDepth CNV calls were obtained by plotting the observed: expected read ratio against the chromosomal position of each captured exon. A region denoting the 99% confidence interval for the ratio in the absence of any CNV call was included in each plot. Read ratios of around 1.5 suggested a heterozygous duplication of the region, while those of around 2.0, 0.5 and 0 suggested a homozygous duplication, a heterozygous deletion and a homozygous deletion respectively.

For sequencing batches consisting of five or fewer unrelated families,

FishingCNV was used (Shi and Majewski, 2013). Fishing CNV compares coverage depth in the test sample against the distribution of coverage in a batch of control samples that need not have been sequenced in the same sequencing run. Aligned, indel realigned recalibrated bam files, without duplicates marked, were used as input for GATK's Depth Of Coverage command (Section 2.4.5), with a minimum mapping quality of 15 and minimum base quality of 10 specified. Reads per kilobase per million mapped reads (RPKM) files were generated using the sample_interval_summary output file for test and control samples. A minimum of sixty control files were pooled and used to analyse the test file for CNVs using the R package FishingCNV (available at http://sourceforge.net/projects/fishingcnv/) with principal component analysis normalisation specified to remove batch effects. Comma separated values file output was sorted on increasing Holm-Bonferroni adjusted p values, with smaller values indicative of more confident calls. Graphical representations of CNVs were obtained from the segmented data (.seg) file using the Broad Institute's Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

2.4.7 Low-coverage whole genome sequencing library preparation

DNA samples considered for WGS were quantified using the Qubit dsDNA BR assay kit. Libraries were prepared by the Leeds St James' University Hospital Next Generation Sequencing Facility using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) as described in the manufacturer's instructions with a few modifications. 100ng of DNA was diluted to 50μl in 1x TE buffer (Section 2.1.3) and sheared using the Covaris S2 Adaptive Focused Acoustics sonicator at a standard shear setting of 20% duty cycle, with an intensity setting of 5, and 200 cycles/burst for 140 seconds at 4°C to form fragments with a peak size of around 185bp. No size selection was performed due to the low input and the PCR amplification step was carried out using the cycling conditions specified by the protocol to a total of 12 cycles. Ten libraries were pooled in equal molar quantities and diluted to 10nM in nuclease free water and sequenced by the Leeds St James' University Hospital Next Generation Sequencing Facility using the Illumina HiSeq2500 platform (Illumina) in 50bp single end read rapid mode.

2.4.8 Low-coverage whole genome sequencing analysis

Analysis of low-coverage whole genome sequencing data was carried out by the Leeds St James's University Hospital Next Generation Sequencing Facility. The protocol for analysis of copy number variation was based on the methods detailed by Watson et al. (2014). Briefly, fastq files were aligned to the human reference genome (GRCh37) using the Burrows-Wheeler aligner (BWA; available at http://bio-bwa.sourceforge.net/) (Li and Durbin, 2009). Duplicate reads were removed using Picard (Section 2.4.2) and uniquely mapped reads from these and three control samples were counted into genomic windows containing equal numbers of reference reads with adjustment for GC% variation. The R module, DNAcopy (available from https://bioconductor.org/packages/release/bioc/html/DNAcopy.html) was then used to segment the data into regions of equal copy number (Venkatraman and Olshen, 2007). CNVs identified with copy numbers less than 0.683 or more than 1.18 were reported in an html file format with UCSC genome browser (Section 2.8.1) links embedded for both window midpoint and window boundaries, as well as the $\log_2$ ratio of reads in the test sample versus the control and the equivalent relative copy number. Cut off values were determined empirically by the Leeds Genetics Laboratory (http://www.leedsth.nhs.uk/a-z-of-services/the-leeds-genetics-laboratory/) in order to remove false positives from the dataset.

2.5 Polymerase Chain Reaction

2.5.1 Primer design

Primer design software tools (Section 2.8.2) were used to design primers. Optimum design parameters were set to produce primers with annealing temperatures of around 60°C, with products ideally ranging between 250bp - 600bp. For genomic templates, the appropriate mispriming library was used in order to reduce the design of primer sequences that included rare (MAF <1%) SNPs and to eliminate those that included common (MAF ≥1%) SNPs.

For screening genes, primers were designed at least 50bp from the intron-exon boundary. Large exons (>750bp) were split into overlapping PCR products of

approximately 500bp, with an overlap of at least 50bp. Primers were designed to be a minimum of 17bp to ensure specificity and a maximum of 27bp, to maintain a melting temperature of around 60°C. Runs of four or more mononucleotides and complimentary 5' and 3' ends were avoided to ensure secondary structures did not form. A GC% of 50, calculated by dividing the sum of Gs and Cs by the total primer length and multiplying by 100, was considered optimal, although 20-80% was accepted if necessary. Primer sequences were checked using UCSC's BLAST-like alignment tool (BLAT), through the UCSC Genome Browser (Section 2.8.1) to ensure primers only bound where expected and that no known SNPs were present at the binding sites.

2.5.2 Standard PCR

All reagents in this section were obtained from Invitrogen, Carlsbad, CA, USA, unless otherwise stated.

Reactions were carried out in a 12.5μl volume with 25ng of genomic DNA with the following reagents at the following concentrations: 400nM for each primer, 200μM for each of dATP, dCTP, dGTP, dTTP, 1x PCR buffer (20mM Tris-HCl (pH 8.4) and 50mM potassium chloride (KCl)), 1.5mM magnesium chloride ($MgCl_2$) and 1 unit *Taq* DNA polymerase. After the initial denaturation step at 95ºC for 2 minutes, the samples were processed through 35 cycles of 95ºC for 30 seconds, the appropriate annealing temperature for 30 seconds and 72ºC for 30 seconds per 500bp of expected product. A final extension step was performed at 72ºC for 5 minutes. PCR products were visualised by agarose gel electrophoresis (Section 2.3.3)

2.5.3 HotShot Master Mix (HSMM) PCR

Reactions were carried out in a 10μl volume with 25ng genomic DNA, with the following reagents and concentrations: 250nM of each primer and 1x PCR buffer / HotShot Diamond PCR Master Mix (Clent Life Science, Stourbridge, UK) with optional 1M betaine and / or 10% dimethyl sulphoxide (DMSO). DNA was amplified according to the manufacturer's instructions using a hot start

PCR cycle. Initial denaturation was at 95ºC for 10 minutes, followed by 35 cycles of 95ºC for 30 seconds, annealing temperature for 30 seconds and 72ºC for 30 seconds per 500bp of expected product. A final extension step was performed at 72ºC for 5 minutes. PCR products were visualised by agarose gel electrophoresis (Section 2.3.3).

2.6 Microsatellite Marker Genotyping

Genotyping of poly-CA microsatellites and other short terminal repeat (STR) variants in genomic DNA was carried out by PCR using fluorescently labelled (HEX or FAM) primers. Markers were selected and their genetic locations determined using the UCSC Genome Browser (Section 2.8.1). PCR was carried out as previously described (Section 2.5). Amplified DNA was subsequently diluted between 5-20x with Hi-Di Formamide (Applied Biosystems, Foster City, CA, USA) and 1μl of the dilution added to 8μl Hi-Di Formamide and 1μl 500 ROX size standard (Applied Biosystems). Fragments were resolved on an ABI3130xl sequencer using a 36cm array, POP7 polymer and 3730 buffer with the FragmentAnalysis36_pop7_1 module (Applied Biosystems). Resulting data were analysed on GeneMapper v4.0 (Applied Biosystems).

2.7 Sanger Sequencing

PCR template DNA to be sequenced underwent USB® ExoSAP-IT® (Affymetrix, Santa Clara, CA, USA) treatment in a 5:2 ratio followed by incubations of 15 minutes at 37°C and 15 minutes at 85°C. This was performed on PCR amplified DNA to remove unincorporated primers and dNTPs, which could interfere with subsequent sequencing reactions. 1μl of USB® ExoSAP-IT® treated DNA was added to 5.5μl of de-ionised H_2O, 1μl of BigDye® Terminator v3.1 (Applied Biosystems), 1.5μl of BigDye® Terminator v3.1 Sequencing Buffer (Applied Biosystems) and 1μl of primer of 1.6μM concentration. After an initial denaturation step at 96ºC for 1 minute, the samples underwent 25 cycles of 96ºC for 10 seconds, 50ºC for 5 seconds and 60ºC for 4 minutes. All temperatures were ramped at 1ºC/second. After cycling, reactions were precipitated initially using 8mM EDTA and 80% ethanol and

spun at 3061 x g for 30 minutes at 22ºC. After removal of supernatant, 70% ethanol was added and samples spun at 805 x g for 15 minutes at 4ºC. Precipitates were dissolved in 10μl Hi-Di Formamide (Applied Biosystems) and resolved at 60°C using a 36cm array on an ABI3130xl Genetic Analyzer (Applied Biosystems). POP7 polymer, 3730 sequencing buffer and FragmentAnalysis36_pop7_1 module (Applied Biosystems) were used for all runs. Resulting data were analysed on either Sequencing Analysis v5.2 or Seqscape v2.5 (Applied Biosystems).

2.8 Bioinformatics

Electronic resources, available online, that have been used in the design, analysis and interpretation of various assays in this project, are listed below.

2.8.1 UCSC Genome Browser

The UCSC genome browser for the human genome (available at http://genome.ucsc.edu/) brings together the vast wealth of genomic data into a single database (Kent et al., 2002). It comprises various builds of the genome sequence alongside an increasing number of custom tracks that can be added as required by the user. Custom tracks include databases of variation such as dbSNP, EVS and ExAc (Section 2.4.3) as well as the compendium of human genes and phenotypes, Online Mendelian Inheritance in Man (available at http://www.omim.org/). The intron-exon structures, genomic sequence, protein sequence and the locations and frequencies of polymorphisms were all determined using this tool. UCSC genome browser also hosts tools such as BLAT and *in silico* PCR that were used to assess the specificity of primer sequences.

2.8.2 Primer Design Software Tools

Primer design software is used to produce primer sequences that meet all of the criteria for successful, specific amplification more rapidly than selecting sequences manually. The program AutoPrimer3 (available at https://github.com/gantzgraf/autoprimer3) utilises the primer3 software tool

v4.0.0 (http://primer3.ut.ee/) to design and name primers for all coding exons of a specified gene automatically. The following settings were used:

Primer-target buffer	50bp
Maximal target (exon) size	500bp
Overlap (for large exons)	50bp
Annealing Temperature	60°C
Primer size (minimum/optimum/maximum)	18/20/27bp

AutoPrimer3 was utilised in preference to primer3 alone since it is able to automatically download the latest reference sequence file, can design primers for all exons simultaneously and can automatically avoid design of primers that lie over common SNPs. All primers were checked using BLAT, via the UCSC genome browser (Section 2.8.1) to avoid sequences that anneal over a known SNP.

If redesign of a primer was required, primer3 was used manually with sites of known SNPs marked to exclude them from the design template. The parameters were set as follows: a primer length of 20bp (range 18-27bp), a primer melting temperature of 60°C (range 57-63°C) and a GC% between 20% and 80%. Primer pairs were designed to have a melting temperature within 2°C of each other. If these conditions could not be met for a primer, then the melting temperature was raised to 65°C.

2.8.3 Literature Searches

2.8.3.1 PubMed

PubMed (http://www.ncbi.nlm.nih.gov/pubmed/), an online database of journal articles, was used to find relevant previously published information. Online Mendelian Inheritance in Man was used to obtain further information about disease phenotypes, loci and known genes.

2.8.3.2 PubMatrix

PubMatrix is an automated literature search tool. The user may enter up to 100 search terms and 10 modifier terms. The tool then creates up to 1000 different PubMed (Section 2.8.3.1) searches and reports articles matching the search terms as html links.

2.8.4 Mutation Prediction Software

The advent of Next Generation Sequencing and the wealth of variants of unknown significance identified in data from it has meant that there is a need for software to grade variants based on their disease-causing potential. This grading allows the most promising sequence variants to be prioritised for further analysis. Several online tools have been developed to assess the effects of variants and those used in this study are described below.

2.8.4.1 PROVEAN and Sorting Intolerant From Tolerant (SIFT)

PROVEAN (available at http://provean.jcvi.org/genome_submit.php)
produces a score based upon the amino acid variation within the context of the surrounding sequence. Therefore, PROVEAN is not limited to solely non-synonymous SNPs, but can predict the effect of other types of sequence variation, such as indels and multiple amino acid substitutions (Choi et al., 2012). Scores are defined as follows:

>-2.5	neutral
≤-2.5	deleterious

SIFT (available at http://provean.jcvi.org/genome_submit.php) classifies amino acid substitutions based upon the evolutionary conservation of the residue within the relevant protein family (Kumar et al., 2009). Scores are defined as follows:

<0.05	damaging
≥0.05	tolerated

2.8.4.2 Polymorphism Phenotyping (PolyPhen)-2

PolyPhen-2 (available at http://genetics.bwh.harvard.edu/pph2/) gives a score based upon its prediction of the possible impact of an amino acid substitution (Adzhubei et al., 2010). This is based upon the sequence and 3D structural features of the protein and its homologues and function of a protein and the likelihood of the substitution at that site. The HumVar trained PolyPhen-2 model is used to appraise variants in the case of Mendelian disease. Scores are defined as follows:

<0.2	benign
0.2-0.85	possibly damaging
>0.85	probably damaging
No score	unknown

2.8.4.3 MutationTaster

MutationTaster (MT) (available at http://www.mutationtaster.org/) utilises and presents information about evolutionary conservation, splice-site changes, protein features and potential alteration to the level of transcription (Schwarz et al., 2010). The tool can assess the impact of synonymous, selected intronic and complex changes as well as single amino acid changes using three different prediction models. Classes are as follows:

- Disease causing automatic
- Disease causing
- Polymorphism automatic
- Polymorphism

Automatic classification is assigned when the variant has been assigned disease causing / polymorphic variant status in a database of variation, including dbSNP (Section 2.4.3), 1000 Genomes Project (Abecasis et al., 2010), ClinVar (available at http://www.ncbi.nlm.nih.gov/clinvar/) (Landrum et al., 2016) and Human Gene Mutation Database (available at http://www.hgmd.cf.ac.uk/ac/index.php) (Stenson et al., 2014). Classifications are qualified by a certainty of prediction score, ranging from 0 to 1, with 1 being most certain.

2.8.4.4 Combined Annotation Dependent Depletion

Combined Annotation Dependent Depletion (CADD; available at http://cadd.gs.washington.edu/) annotation uses a model developed by scoring naturally occurring human variation and simulated variation for a range of measures including conservation, exon-intron geography, expression studies and protein level scores such as SIFT and PolyPhen-2 (Kircher et al., 2014). This information was used to develop a model capable of scoring all possible small indel and SNV variants to produce a scaled C score measure of deleteriousness that is scaled relative to all other possible alterations. Scaled C scores range from 1 to 99 with the top 10% of the most deleterious variants assigned scores of 10 or higher, and the top 1% and 0.1% assigned scores of 20 or higher and 30 or higher respectively.

2.8.4.5 Splicing RegulatiOn Online Graphical Engine (SROOGLE)

In order to assess whether intronic variants were likely to alter splicing, sequences were queried using SROOGLE. SROOGLE (available at http://sroogle.tau.ac.il/) is a platform that scores different types of splicing signals using different splicing prediction tools and incorporates a ranked percentile score of these scores based on a dataset of scores of a large number of exons (Schwartz et al., 2009).

For scoring of splice sites, SROOGLE utilises two models of splicing prediction: the Position Specific Scoring Matrix (PSSM) (Shapiro and Senapathy, 1987) and the Maximum Entropy Model (MEM) (Yeo and Burge, 2004). PSSM uses a position weight matrix, where each nucleotide at each position in a splice region is given an independent score based on its likelihood to occur at that position (Shapiro and Senapathy, 1987). Individual scores are then used to produce an overall score, with higher scores more likely to represent true splice sites. MEM compares the query sequence as a whole with a dataset of known splice sites, therefore taking into account the interdependence between each position in the splice site (Yeo and Burge, 2004). Higher MEM scores are more likely to represent true splice sites.

SROOGLE ranks the output of both PSSM and MEM scores to produce percentile ranks (0-1) (Schwartz et al., 2009). Scores indicate the percentage of exons SROOGLE's dataset with lower scores, for example, a score of 0.7 would indicate that 70% of exons in the dataset have lower scores than the query sequence and therefore 30% have higher scores.

2.8.5 Clustal Omega

Clustal Omega (available at http://www.ebi.ac.uk/Tools/msa/clustalo/) is a multiple sequence alignment program for DNA, RNA or proteins. Alignment output can be ordered by sequence similarity to highlight highly conserved sequences. Clustal Omega was used to analyse conservation at the positions of missense variants and the surrounding amino acid residues. Sequences for proteins of interest were obtained from the UCSC genome browser (Section 2.8.1). Paralogous sequences were retrieved using the Basic Local Alignment Search Tool (BLAST; available at http://blast.ncbi.nlm.nih.gov/Blast.cgi) and were used as input for Clustal Omega.

2.9 Protein extraction from murine tissues

Kidney and brain tissue was collected from adult mice immediately after euthanasia by cervical dislocation. The tissues were washed in sterile phosphate buffered saline and placed in RIPA buffer (Section 2.1.7) containing EDTA free complete protease inhibitor cocktail (Roche, Basel, Switzerland). Tissues were disrupted using an Ultra-Turrax™ T25 basic homogeniser (IKA, Staufen im Breisgau, Germany). Disrupted tissue was then pelleted by centrifugation at 4^{o}C for 20 minutes at 16000 x g. Supernatants were aliquotted and frozen at -80^{o}C until use. Protein was quantified using the Pierce™ bicinchoninic acid (BCA) protein assay kit (ThermoFisher Scientific).

2.10 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein was extracted from tissues as detailed in Section 2.9. 50 μg of protein was incubated at 95°C for 10 minutes with 1x Laemmli sample buffer (Bio-Rad). For reducing conditions, 10% 2-mercaptoethanol was added to the 4x Laemmli buffer stock prior to addition to the protein sample and its subsequent dilution to 1x. Samples and a Precision Plus™ Protein All Blue Prestained Protein Standards (Bio-Rad) were then loaded onto a 10% (for effective separation of 70-150kDa proteins) or 12% (for effective separation of 10-70kDa proteins) SDS PAGE gel (Section 2.10.1.1). The gel was run in a Mini-PROTEAN Tetra cell gel tank using 1x Tris glycine SDS PAGE running buffer (National Diagnostics, Madison, NJ, USA) at 200V for 60-90 minutes, depending upon the expected size of the product of interest.

If required, protein gels were stained by incubating with InstantBlue™ stain (Expedeon, Swavensey, UK) for one hour with gentle agitation.

2.10.1 SDS-PAGE gel compositions

2.10.1.1 Resolving gel

10% / 12%	Acrylamide Bis-acrylamide stock solution (Severn Biotech Ltd, Kidderminster, UK)
375mM	Tris-HCl pH 8.8
0.1%	SDS
21.25%	Glycerol
0.05%	Ammonium persulphate (APS)
0.1%	Tetramethylethylenediamine (TEMED)

2.10.1.2 Stacking gel

4%	Acrylamide Bis-acrylamide stock solution
125mM	Tris-HCl pH 6.8
0.1%	SDS
0.05%	APS
0.1%	TEMED

2.11 Western blotting

Proteins separated by SDS-PAGE (Section 2.10) were transferred from the gel to the polyvinylidene fluoride (PVDF) membrane in a Mini PROTEAN Tetra blotting module (Bio-Rad) at 30V for 60-90 minutes (depending upon the expected size of the product of interest). Immediately prior to transfer, the PVDF membrane (Life Technologies) was pre-wet in 100% methanol for 30 seconds and transferred to 1x Tris-glycine electroblotting buffer (National Diagnostics), containing 20% methanol, for 5 minutes. After transfer, the membrane was blocked in 5% Marvel original dried skimmed milk powder (Premier International Foods, Spalding, UK) in TBS (Section 2.1.8) overnight at 4^{o}C. The membrane was incubated with the primary antibody in 5% Marvel original dried skimmed milk powder in TBS for one hour at room temperature. After two washes of ten minutes each with TBST (Section 2.1.9) followed by one wash of ten minutes with TBS, the membrane was incubated with the appropriate secondary antibody for one hour at room temperature and then washed again as before.

Blots were visualised on a Bio-Rad molecular imager gel documentation system with chemiluminescence displayed using Image Lab (v.4.0) analysis software (Bio-Rad, Hemel Hempstead, UK). SuperSignal™ West Pico Chemiluminescent Substrate (ThermoFisher Scientific) or SuperSignal™ West Femto Maximum Sensitivity Substrate (ThermoFisher Scientific) was used to visualise any bands present on the membrane by chemiluminescence.

If required, blots were stripped of bound antibodies using Restore™ Western Blot Stripping Buffer (ThermoFisher Scientific) by incubating the blot in the buffer for 15 minutes and following the manufacturer's instructions. Blots were

probed with chemiluminescent substrate as before, to confirm that the stripping treatment had been successful prior to washing with TBS and reblocking with milk powder as before.

2.12 Histology

2.12.1 Tissues

2.12.1.1 Mouse incisor

Incisor tissue was obtained from seven C57Bl/6 mice, aged eight to fifteen weeks, kindly made available by Dr. Zakia Abdelhamed (University of Leeds). The mice were euthanised by cervical dislocation and the jaws removed by careful dissection. Each mandible and maxilla was separated in the mid-sagittal line. After fixation in 10% neutral buffered formalin for 72 hours, the incisors were demineralised at 4ºC by immersion in 0.5M EDTA pH 7.5 with gentle agitation. The buffer was refreshed twice each week and after 3 months, X-ray images showed that the incisors were completely demineralised (Appendix 18).

Incisors were embedded in paraffin wax using an automatic processor (Hypercentre XP™, Thermo Shandon Ltd, Runcorn, UK) and sectioned at a thickness of 5μm using a microtome (Leica Biosystems, Nussloch, Germany) with a disposable stainless steel blade (Leica Biosystems). Sections were mounted on SuperFrost Plus slides (Menzel-Glaser, Braunschweig, Germany) and dried overnight, in an oven at 37ºC. Haematoxylin and eosin staining (Section 2.12.3) aided identification of the sections that included all stages of amelogenesis.

2.12.1.2 Mouse eye

Formalin fixed, paraffin embedded mouse eyes were kindly made available by Mr. Mike Shires (University of Leeds). The eyes had been obtained from three adult C57Bl/6 mice and had been embedded and mounted as previously

described for incisors (Section 2.12.1.1) except that no demineralisation was carried out.

2.12.2 Dewaxing and rehydration

Slides were dewaxed by submersion in three staining baths of xylene (Acros Organics, Loughborough, UK) for 5 minutes each. Slides were drained on absorbent paper and then submerged in three baths of absolute ethanol for 1 minute each to remove any remaining xylene and to rehydrate the slides. Slides were rinsed in running cold tap water for a further 5 minutes before proceeding to haematoxylin and eosin staining (Section 2.12.3) or antigen retrieval (Section 2.12.5).

2.12.3 Haematoxylin and eosin staining

Dewaxed and rehydrated slides were stained for 30 seconds in Harris's haematoxylin (Solmedia, Shrewsbury, UK) and then rinsed for 1 minute under cold running tap water. Next the slides were blued in Scott's tap water substitute (Section 2.1.10) for 1 minute and then rinsed again in cold running tap water for 1 minute. Slides were then stained with eosin (Solmedia) for 2 minutes and 30 seconds and rinsed under cold running tap water for 1 minute. Finally the slides were dehydrated and mounted (Section 2.12.4).

2.12.4 Dehydration and mounting

Slides were dehydrated after staining by submersion in a bath of absolute ethanol for 2 minutes, followed by submersion in a further two baths of absolute ethanol for 1 minute each. The slides were thoroughly drained on absorbent paper and submerged in three baths of xylene for 1 minute each. Slides remained in the third bath until mounting. Slides were mounted using DPX (distyrene plasticiser and xylene) mountant and coverslips (ThermoFisher Scientific). The mountant was left to cure overnight.

2.12.5 Immunohistochemistry

Following dewaxing and dehydration (Section 2.12.2), antigen retrieval was carried out by microwave treatment at full power for 10 minutes with 10mM citrate buffer pH 6.0. For some tissues this was unsuitable as it led to tissue detachment from the slide. In such cases, antigen retrieval was carried out instead by enzymatic treatment with 0.1% chymotrypsin in 0.1% calcium chloride pH 7.8 at 37ºC for 25 minutes in a Coplin jar.

Blocking of endogenous peroxidases was achieved by incubating slides with 0.3% hydrogen peroxide in methanol for 10 minutes in a staining bath. The dilution of primary antibody to be used was determined by trial of a range of antibody dilutions (1:50, 1:100, 1:200, 1:500, 1:1000) with the lowest dilution (1:50) also trialled without antigen retrieval. All dilutions were prepared in antibody diluent (ThermoFisher Scientific). The primary antibody was applied for 1 hour. The secondary antibody to be used was determined by the species in which the primary antibody was raised. The duration of treatment and dilution also varied. Slides were washed twice with TBST and once with TBS after each reagent for 5 minutes each. Staining was achieved with DAB+ reagent from the EnVision+ System-HRP (DAB) for use with rabbit primary antibodies (Dako) as described in the manufacturer's instructions.

Counter staining was with haematoxylin and bluing with Scott's tap water substitute (Section 2.1.10) as previously described (Section 2.12.3). The sections were then dehydrated and mounted (Section 2.12.4).

2.13 Tooth phenotyping

Deciduous and permanent teeth were obtained by natural exfoliation or by extraction for clinical reasons. The teeth were photographed by the Medical and Dental Illustration department (University of Leeds) and were then subjected to ultrastructural analysis by micro computerised tomography (microCT) and surface electron microscopy (SEM). The chemical composition of the enamel was analysed by energy-dispersive X-ray spectroscopy (EDX).

2.13.1 Micro Computerised Tomography

Teeth were analysed by high resolution microCT using a Skyscan 1172 (Bruker, Billerica, MA, USA), operated at 100 kV with a source current of 100μA and an aluminium / copper filter to reduce beam hardening. CT slices were reconstructed using Skyscan Recon software (Bruker). The CT images were calibrated using a 3 point standard of hydroxyapatite mineral of densities 0.25, 0.75 (Bruker) and 2.9g/cm^3 (Himed, Old Bethpage, NY, USA). Regions of enamel in every tenth CT slice were automatically identified using ImageJ (available at http://imagej.nih.gov/ij/) and the Trainable Weka Segmentation plugin (available at http://fiji.sc/Trainable_Weka_Segmentation). This enabled the mean mineral density of the enamel to be calculated with reference to the standards for each scan.

Calibrated colour contour maps of mineral density were generated using ImageJ 3D interactive surface plot plugin (available at http://rsb.info.nih.gov/ij/plugins/surface-plot-3d.html).

2.13.2 Sectioning of human teeth

Teeth were cut in half longitudinally in the labial/bucco-lingual direction using an Accutom-5 cutter (Struers, Ballerup, Denmark) with a peripheral diamond cutting disc, cooled with minimal water to minimise the loss of organic matter from within the enamel.

One half of the tooth was polished plano-parallel with 600 grade carborundum paper (3M, Maplewood, MN, USA), then acid etched for 20 seconds with 30% phosphoric acid to remove the smear layer. The section was then washed with excess distilled water for 2 hours to stop the acid etching and to remove any debris on the surface. The tooth section was dried overnight with a vacuum applied.

2.13.3 Scanning Electron Microscopy

Sections were mounted on aluminium stubs and sputter coated with gold using an AGAR auto sputter coater (Agar Scientific, Elektron Technology, Stansted, UK). Microstructural analysis was undertaken using a Hitachi S-3400N scanning electron microscope (Hitachi, Tokyo, Japan), fitted with a 123eV Bruker Nano XFlash® Detector 5010 (Bruker) and operated at an accelerating voltage of 20kV using secondary electron detection.

2.13.4 Energy-dispersive X-ray spectroscopy

Elemental analysis was performed on selected regions of the enamel by EDX using a detector fitted with an ultrathin window using Bruker Quantax Espirit software version 1.9.4 (Bruker). A minimum of nine measurements were obtained for each enamel region from each tooth. The mean composition of each region was then calculated.

3 The identification of *PEX1* and *PEX6* mutations in Heimler syndrome

3.1 Introduction

The aim of the research presented in this chapter was to identify the genetic cause or causes of Heimler syndrome (HS). To achieve this, WES of affected individuals from a total of ten families with HS-like phenotypes was undertaken. The second aim was to identify the localisation of the affected encoded protein(s) within tissues affected by the syndrome. To achieve this, murine incisor and retina was used for immunohistochemical study.

At the start of this study, HS had been described as a combination of SNHL and enamel defects, with the presence of nail abnormalities and macular dystrophy reported in some cases. No genetic cause for HS was known and since only seven cases of HS had been reported in a total of four families, it remained possible that the majority of diagnosing clinicians remained unaware of the existence of HS and consequentially defined patients' symptoms as a combination of SNHL and RP (Usher syndrome; USH), and amelogenesis imperfecta (AI). Recruitment therefore focused on ascertaining patients with a combination of the constituent phenotypes. Since the nail abnormalities variably reported in published cases of HS are common in the general population and could have been caused by physical trauma or infection, their presence was not considered to be part of the diagnostic criteria for HS.

Ten unrelated families segregating an AR syndrome, diagnosed as either HS or USH and AI, were recruited to this study (Figure 3.1). At the time of recruitment, no genetic cause for HS had been identified. Blood or saliva DNA samples were therefore obtained from affected individuals and their unaffected relatives, with the aim of identifying the gene(s) and mutations underlying HS.

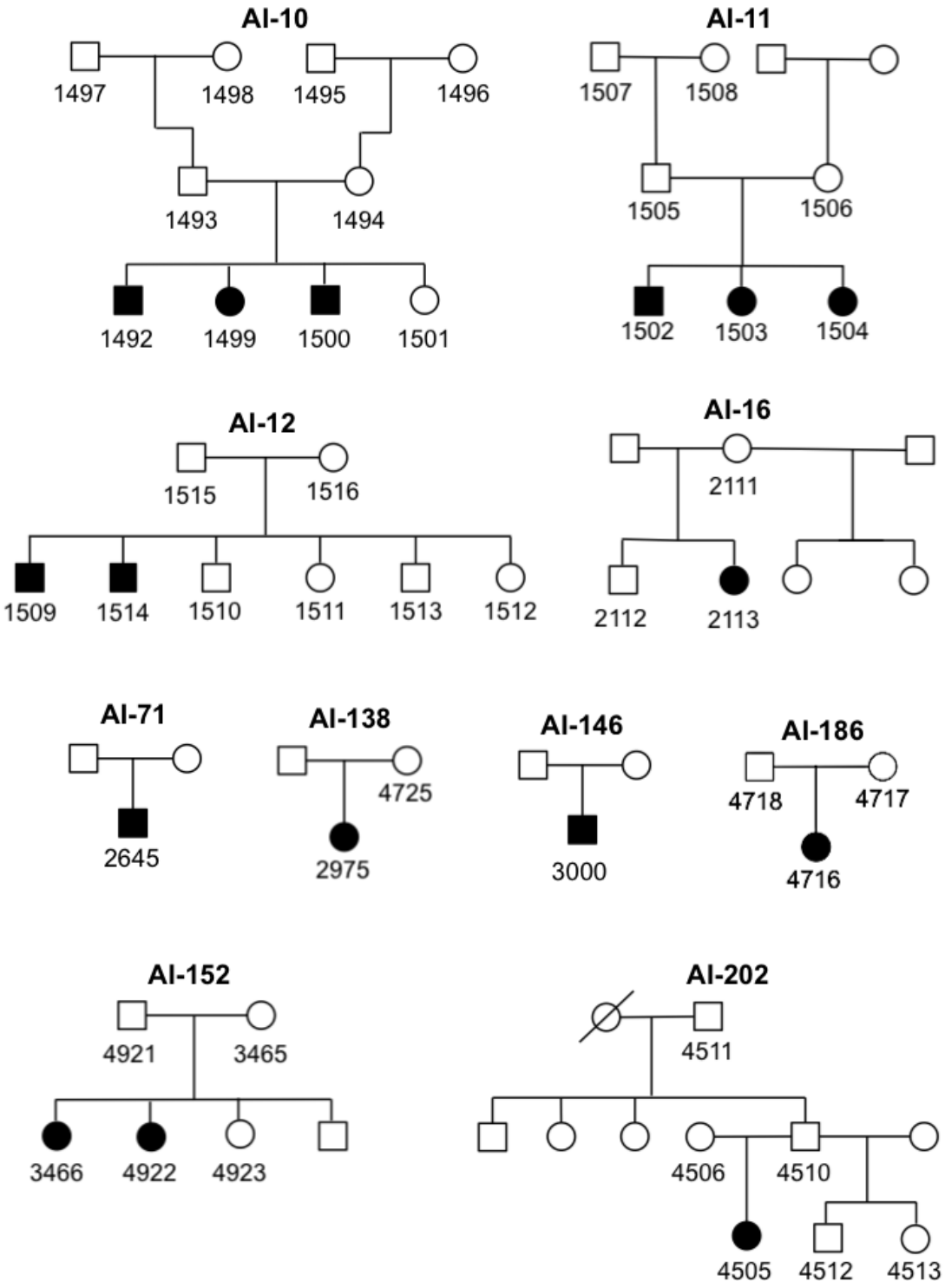


Figure 3.1: Pedigrees of families included in the HS study.

3.2 Results

Details of the primers used in the following section are listed in Appendix 5.

3.2.1 Clinical phenotype

A summary of the clinical phenotypes of the patients recruited to this study is detailed in Table 3.1 and Table 3.2. Families were recruited via a network of clinical collaborators working with the Leeds Vision and Amelogenesis Research groups, both before and during the course of this PhD project. The phenotype of 4716 (AI-186) is detailed in Figure 3.2. The phenotypes of 2975 (AI-138), 3000 (AI-146) and 2645 (AI-71) are detailed in Appendix 6. Pedigrees for each family are shown in Figure 3.1.

Individuals had all been diagnosed with SNHL, retinal dystrophy and AI or enamel defects. All patients had been diagnosed with a combination of USH and AI, except for individual 3000 (family AI-146), who had been diagnosed with HS. The phenotype of individual 2113 (family AI-16) had been reported previously (Balmer and Fayle, 2007). Individual 2645 (family AI-71) presented with additional features including schizophrenia, mild learning disability and skin abnormalities over his hands and lower legs. Individuals 3466 and 4922 (family AI-152) also presented with mild learning disability as well as facial skin lesions. Families were of US (families AI-10, AI-11, AI-12, AI-146 and AI-186), UK (AI-16, AI-138 and AI-202), Chinese (AI-71) and Israeli (AI-152) heritage. There were no reports of consanguinity in any of the families.

Family	AI-10			AI-11			AI-12		AI-16	AI-71	AI-138
Sample number	1492	1499	1500	1502	1503	1504	1509	1514	2113	2645	2975
Origin	USA	USA	USA	USA	USA	USA	USA	USA	UK	UK born Chinese	UK
Sex	M	F	M	M	F	F	M	M	F	M	F
Age (years) at last assessment	NA	NA	NA	NA	NA	NA	NA	NA	11	45	17
Amelogenesis imperfecta	+	+	+	+	+	+	+	+	+	+	? Permanent first molars and incisors affected only
Intellect, known psychiatric disorder	NA	NA	NA	NA	NA	NA	NA	NA	N	Mild LD; paranoid schizophrenia	N
Skin	NA	NA	NA	NA	NA	NA	NA	NA	NA	Dry split skin on hands; ichthyosis over limbs	N
SNHL											
Bilateral or unilateral	B	B	B	B	B	B	B	B	B	B	B
Age (years) of diagnosis	NA	NA	NA	NA	NA	NA	NA	NA	0	Early childhood	Early childhood
Degree of hearing loss	NA	NA	NA	NA	NA	NA	NA	NA	P, some VD	S	S to P
Nail Abnormalities											
Beau's lines	NA	NA	NA	NA	NA	NA	NA	NA	NA	+	-
Other nail changes	NA	NA	NA	NA	NA	NA	NA	NA	NA	-	-
Ocular Features											
Retinal dystrophy	+	+	+	+	+	+	+	+	+	+	+

Table 3.1: Clinical details of individuals with HS (part 1).

Abbreviations are as follows: B bilateral; F female; LD learning disability; M male; Mi mild; Mo moderate; N normal; NA not ascertained; P profound; S severe; SNHL sensory neural hearing loss; VD vestibular dysfunction.

Family	AI-146	AI-152		AI-186	AI-202
Sample number	3000	3466	4922	4716	4505
Origin	USA	Israel	Israel	USA	UK
Sex	M	F	F	F	F
Age (years) at last assessment	12	35	22	7	24
Amelogenesis imperfecta	+	+	+	+	+
Intellect, known psychiatric disorder	N	Mild LD	Mild LD	N	N
Skin	Hyper-pigmentation on left arm and shoulder	Dryness; light facial lesions	Light facial lesions	N	N
SNHL					
Bilateral or unilateral	B	B	B	NA	B
Age (years) of diagnosis	1.5	0	0	3	1.5
Degree of hearing loss	Mo to S	Mo to P	Mi to Mo	Mo to S	P
Nail Abnormalities					
Beau's lines	+	-	-	-	-
Other nail changes	-	-	Longitudinal stripes	-	-
Ocular Features					
Retinal dystrophy	+	+	+	+	+

Table 3.2: Clinical details of individuals with HS (part 2).

Abbreviations are as follows: B bilateral; F female; LD learning disability; M male; Mi mild; Mo moderate; N normal; NA not ascertained; P profound; S severe; SNHL sensory neural hearing loss; VD vestibular dysfunction.

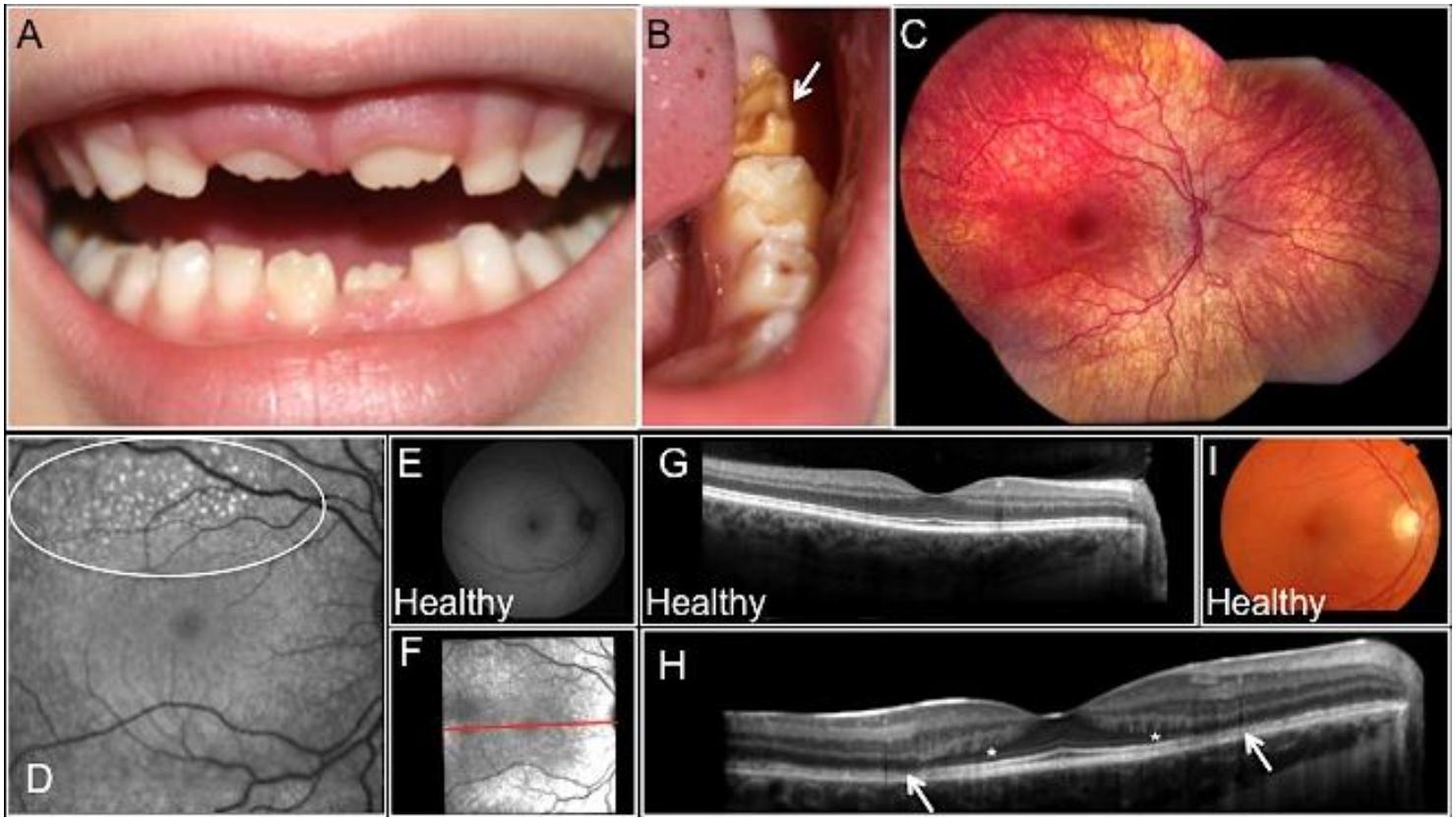


Figure 3.2: Clinical details of the phenotype of individual 4716 (family AI-186).

A and B: AI affecting the primary and secondary dentitions with generalised reduced enamel volume (hypoplasia) and variable hypomineralisation (particularly evident in the lower left permanent first molar tooth in B (arrow)). C: Fundus image showing pigmentary maculopathy and mild retina vascular attenuation in individual 4716. D: Fundus autofluorescence (FAF) showing hyperfluorescence at the peripheral macula (circled) in individual 4716. E: FAF image of a healthy retina for comparison. F: Infrared reflectance image to show position of optical coherence tomography (OCT) image in H for individual 4716. G: OCT of a healthy retina for comparison with H, note the defined retinal layers. H: OCT showing depletion of photoreceptors in the perifovea (arrows) and disruption of the outer nuclear layer (marked with *) for individual 4716. I: Fundus image of a healthy retina for comparison with C.

3.2.2 Whole exome sequencing of Heimler syndrome patient DNA

In order to identify the gene or genes mutated in HS, WES was undertaken. Whole exome DNA libraries were prepared from fourteen affected individuals from ten unrelated families (Section 2.4.1.1 and Table 3.1 and Table 3.2). Representative Bioanalyser traces obtained for sample 4716 (family AI-186) at various stages in the WES library preparation protocol are shown in Figure 3.3. Libraries were prepared and sequenced in six batches over the course of 18 months, during which time the SureSelect capture reagent was reformulated to include a greater proportion of coding exons. The version of the capture reagent used for each library preparation is detailed in Appendix 8.

Initially, WES libraries were prepared from two affected individuals each from families AI-10 (Individuals 1499 and 1500), AI-11 (1503 and 1504), AI-12 (1509 and 1514) and one affected individual from families AI-16 (2113), AI-71 (2645), AI-138 (2975) and AI-152 (3466) using the 3μg protocol (Section 2.4.1.1). Subsequently WES libraries were prepared for a further four individuals from families AI-146 (3000), AI-152 (4922), AI-186 (4716) and AI-202 (4505). In some cases, there was insufficient DNA stock to use the 3μg input specified by the manufacturer's instructions. Details of the DNA inputs used can be found in Appendix 8.

The resulting Fastq sequence files were aligned using Novoalign and the variants were called and filtered, using an analysis pathway appropriate for rare recessive diseases, described previously (Sections 2.4.2 and 2.4.3). The filtering protocol used the vcfhacks (v.0.1.11) perlscript, AnnotateSnps.pl, to exclude all variants present in dbSNP129 unless they had been marked as pathogenic in ClinVar (release 29/09/2014), indicated by the CLNSIG flag. This filter is later described as dbSNP129/pathogenic filter for brevity. The depth of coverage was also analysed (Section 2.4.5 and Appendix 8). The resulting filtered variant files were further filtered by excluding any identical variants present in two WES datasets from unrelated patients unaffected with, and without a family history of HS, USH or AI, processed using the same alignment and variant calling pipeline. This was to aid the exclusion of artefactual variants

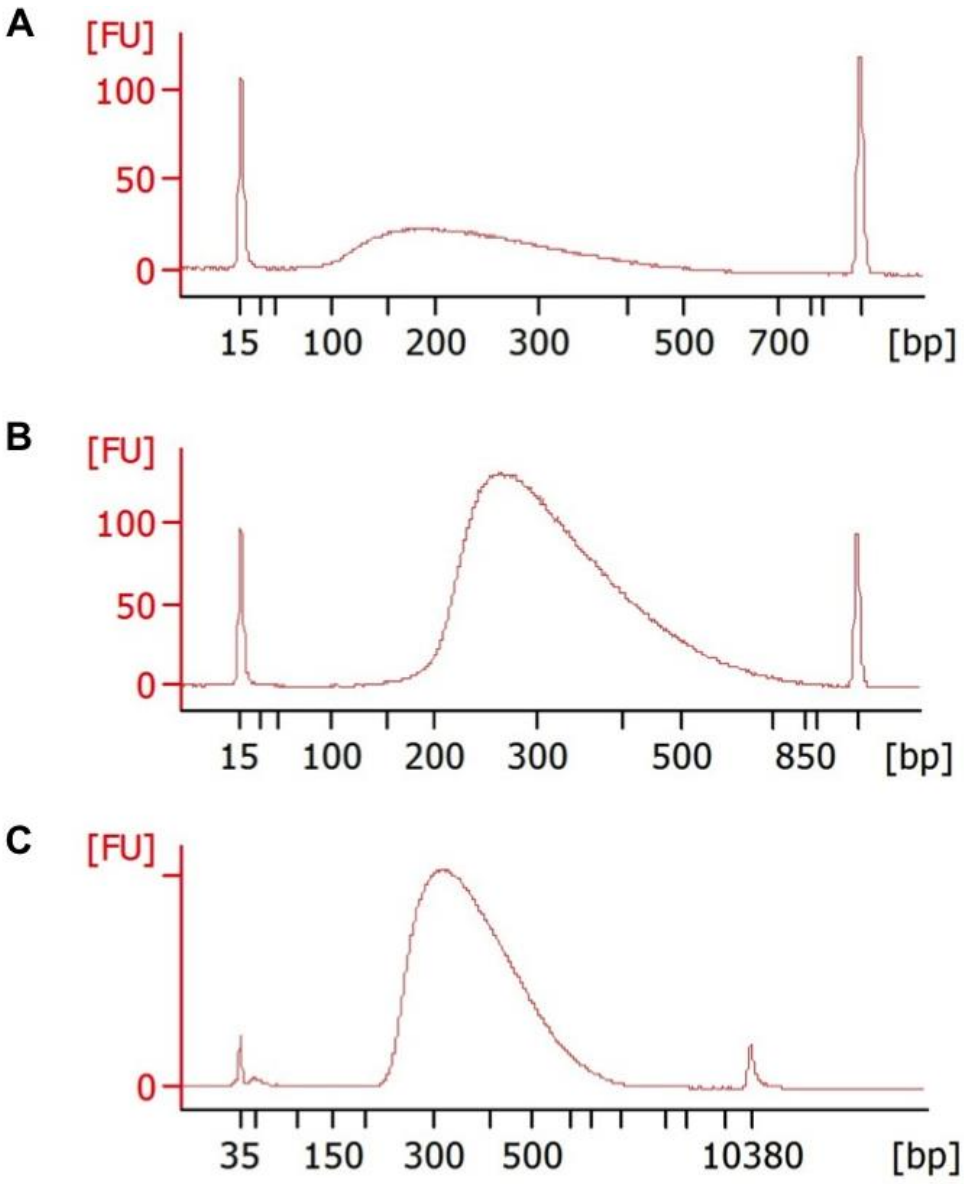


Figure 3.3: Representative bioanalyser traces.

Traces shown were obtained using the Agilent DNA1000 assay kit (A and B) and the Agilent DNAHiSensitivity (C) assay kit (Section 2.4.1.1) for sample 4716 (family AI-186). Abbreviations: bp base pairs; FU fluorescence units.

A: After shearing and AMPure bead purification, 1µl DNA (50µl total yield) from 4716 was analysed to ensure the peak DNA fragment size was between 150-200bp.

B: After end repair, A tailing, adaptor ligation and pre-hybridisation amplification with appropriate AMPure bead purification steps, 1µl DNA (30µl total yield) from 4716 was analysed to ensure the peak DNA fragment size was between 250-275bp and to quantify the DNA prior to hybridisation.

C: After hybridisation, streptavidin bead capture, post-hybridisation addition of index tags and amplification with appropriate AMPure bead purification steps, 1µl DNA (30µl total yield) from 4716 was analysed to ensure the peak DNA fragment size was between 300-400bp and to quantify the DNA, by integration of the area under the peak, prior to pooling.

inherent to the library preparation and sequencing platform. Variants were annotated as previously described (Section 2.4.4).

3.2.3 Identification of *USH2A* variants in families AI-12 and AI-138

Both filtered variant lists for 1509 and 1514 (family AI-12; Appendix 9) and 2975 (family AI-138; Appendix 10) included biallelic variants in *USH2A*, a known cause of USH2 and non-syndromic recessive retinitis pigmentosa (RP) (Section 1.11) (Eudy et al., 1998, Rivolta et al., 2000). The *USH2A* variants (annotation based on NM_206933; Table 3.3) were verified by Sanger sequencing of DNA from all available family members (Figure 3.4).

In individuals 1509 and 1514 (family AI-12), two heterozygous deletions, c.2299delG and c.11875_11876delCA, as well as an SNV, c.4714C>T, were identified after filtering of the WES data. Analysis of the segregation of these alleles revealed that the c.2299delG and c.4714C>T variant were on the same maternal allele in *cis*, with the c.11875_118876delCA variant on the paternal allele in *trans* (Figure 3.4A).

In individual 2975 (family AI-138), four heterozygous variants in *USH2A* were identified after filtering of WES data. Two of these variants were identical to those identified in 1509 and 1514 (family AI-12), c.2299delG and c.4714C>T. The SNVs c.11927C>T and c.11549-1G>A were also identified. Segregation of the four variants was performed but was limited by the availability of only one additional family member. As for family AI-12, analysis of the segregation of these alleles revealed that the c.2299delG and c.4714C>T variant were on the same paternal allele (this was inferred, due to its absence in 4725 and its co-segregation, with the same c.2299delG variant that had been identified in 1509 and 1514, family AI-12) in *cis*, with the c.11927C>T and c.11549-1G>A variants in *cis* on the maternal allele (Figure 3.4).

Family	Genomic variant (GRCh37)	Transcript variant	Amino acid change	Exon / Intron (i)	dbSNP142 reference	CADD v1.3	Provean	SIFT	Polyphen2 (HumVar)	Mutation Taster
AI-12 and AI-138	1:216420437 delC	c.2299delG	p.E767Sfs*21	13	rs80338903	33	N/A	N/A	N/A	Disease causing (1)
AI-12 and AI-138	1:216270469 G>A	c.4714C>T	p.L1572F	22	rs111033333	26.8	Neutral (-2.39)	Damaging (0.011)	Probably damaging (0.885)	Disease causing (0.97944)
AI-138	1:215914880 C>T	c.11549-1G>A	p.?	59(i)	N/A	24.1	N/A	N/A	N/A	Disease causing (1)
AI-12	1:215901562_ 215901563 delTG	c.11875_ 11876 delCA	p.Q3959Nfs*53	61	N/A	35	N/A	N/A	N/A	Disease causing (1)
AI-138	1:215901511 G>A	c.11927C>T	p.T3976M	61	rs142381713	25.5	Neutral (-2.43)	Tolerated (0.056)	Probably damaging (0.982)	Disease causing (0.99914)

Table 3.3: Details of the five heterozygous genomic *USH2A* variants identified by WES in families AI-12 and AI-138 after filtering.

The list is restricted to variants scoring 15 or more with CADD v1.3.

Provean, SIFT and Mutation Taster annotations were based on Ensembl transcript ENST00000307340.

Polyphen-2 annotations were based on RefSeq protein NP_996816.

Annotation of exon and intron numbers follows the Leiden Open Variant Database style of nomenclature.

All scores were obtained on the 11th February 2016.

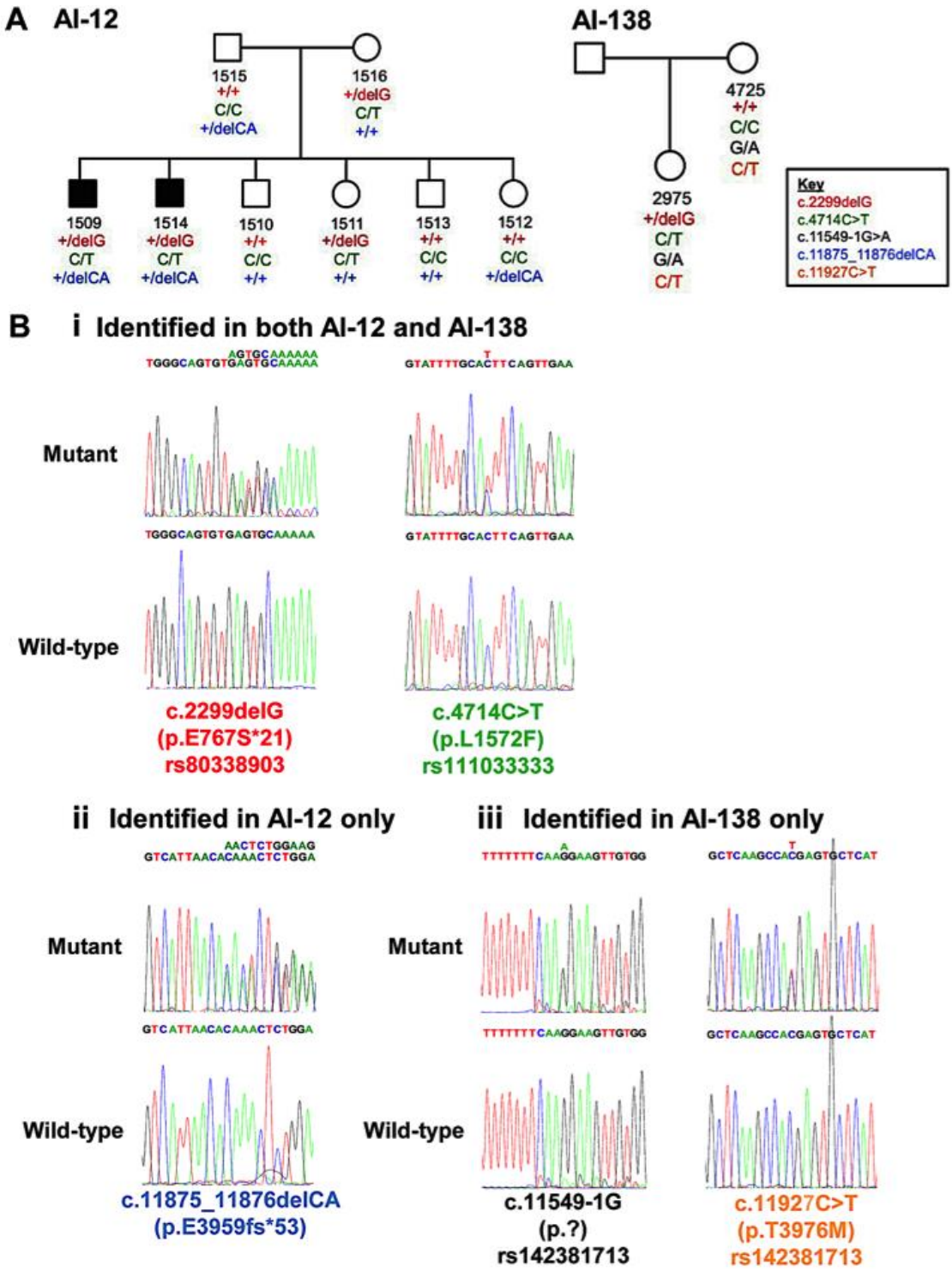


Figure 3.4: Segregation analysis and Sanger sequencing traces of *USH2A* variants identified in 1509 and 1514 (family AI-12) and 2975 (family AI-138).

Annotations based on RefSeq transcript NM_206933 and protein NP_996816.

A: Pedigrees for AI-12 and AI-138. Family members for whom DNA was available are labelled with a number. Affected individuals are indicated by shading. The genotype of each individual is labelled based on Sanger sequencing results for the *USH2A* variants identified by WES.

B: Sanger sequencing traces for mutant and wild-type sequences for the *USH2A* variants identified in i) both AI-12 and AI-138; ii) AI-12 only; and iii) AI-138 only.

Genomic variant (GRCh37)	dbSNP142 reference	ClinVar (release 07/01/2016)	ExAC (v0.3) number of variant alleles (total alleles at this position); allele frequency (all reported individuals heterozygous)	EVS number of variant alleles (total alleles at this position); allele frequency (all reported individuals heterozygous)	LOVD[a]; number of reports of the variant
1:216420437delC	rs80338903	Allele ID 17390	96 (121284); 7.92 x10^{-4}	12 (12518); 9.59 x10^{-4}	535
1:216270469G>A	rs111033333	N/P	95 (120984); 7.85 x10^{-4}	14 (13006); 1.08 x10^{-3}	115
1:215914880C>T	N/P	N/P	N/P	N/P	1
1:215901562_215901563delTG	N/P	N/P	2 (121344); 1.65 x 10^{-5}	N/P	7
1:215901511G>A	rs142381713	N/P	63 (121330); 5.19 x10^{-4}	13 (13006); 1.00 x10^{-3}	2

Table 3.4: Frequencies in public databases for the variants in *USH2A* identified in 1509 and 1514 (family AI-12) and 2975 (family AI-138).

[a]LOVD Retinal and hearing impairment genetic mutation database Usher syndrome 2A (USH2A) curated by David Baux (Roux et al., 2011).

Abbreviation is as follows: N/P Not present.

All databases accessed 27th January 2016.

Analysis of each *USH2A* variant identified by scoring with pathogenicity prediction software and investigation of each variant's frequency in a variety of publically available databases of known variation is shown in Table 3.4. The variant identified in 2975 (family AI-138) within an intronic splice site, c.11549-1G>A, was analysed using SROOGLE (2.8.4.5).

The single nucleotide variants identified in 2975 (family AI-138) included both missenses c.4714C>T and c.11927C>T and an intronic mutation affecting a splice site c.11549-1G>A. Since the missense variant (c.4714C>T) was found to be on the same allele as the paternal (inferred) frameshift variant (c.2299delG), the pathogenicity of the missense mutation is likely to be of no consequence to the overall effect on the protein since the frameshift variant will lead to production of a transcript that would be subject to nonsense mediated decay (NMD). For the missense variant on the maternal allele, c.11927C>T, pathogenicity prediction software, including Polyphen-2 and Mutation taster, predicted that the variant was pathogenic, with borderline neutral and tolerated scores obtained using Provean and SIFT. The splice site variant, c.11549-1G>A would be expected to alter the 3' acceptor site within intron 59 due to its location. Analysis and comparison of scores obtained from reference and variant sequences using SROOGLE, indicated that there was indeed likely to be altered function of the 3' acceptor site (Table 3.5).

Sequence (GRCh37)	**Max. entropy model score (MEMS)**	**MEMS constant exons percentile**	**MEMS alternative exons percentile**	**Position specific scoring matrix (PSSM) score**	**PSSM constant exons percentile**	**PSSM alternative exons percentile**
Reference sequence	8.57	0.47	0.54	91.05	0.94	0.94
1:215914880C>T	-0.18	0.01	0.02	67.61	0	0.01

Table 3.5: *In silico* analysis of the effect of *USH2A* variant c.11549-1G>A, identified in 2975 (family AI-138) on splicing.

Maximum entropy model and position specific scoring matrix scores and their associated percentile ranks were obtained from SROOGLE.

A literature search of the variants identified in families AI-12 and AI-138 revealed that all had been identified previously in USH patients (Eudy et al.,

1998, Baux et al., 2007, Dreyer et al., 2008, Lenassi et al., 2015). The identification of biallelic variants in *USH2A,* a gene previously implicated in USH2, in two patients from the HS cohort, suggested that these cases may have USH2 rather than HS. Despite a large body of literature detailing the phenotype of patients with *USH2A* mutations, there have been no reports of an associated enamel phenotype. It therefore seems likely that the tooth defect in these cases is due either to a separate mutation or to environmental enamel defects. The filtered WES data did not highlight any biallelic variants in genes other than *USH2A* for family AI-12. For 2975 (family AI-138), biallelic variants in *GXYLT1*, *WDR89*, *TMEM200C*, *TSC2*, *MUC6*, *MUC16*, *ZNF880*, *KRT25* and *HTT* remained after filtering. To determine whether *USH2A* or any of the other genes with biallelic variants are differentially expressed during amelogenesis and might therefore be candidate AI genes, the presence or absence of each gene in RNA sequencing studies of secretory stage and maturation stage amelogenesis in rats was checked (Yin et al., 2014b). *USH2A* was not identified as differentially expressed. From the remaining candidate genes, *TSC2*, *MUC6* and *KRT25* were found to be differentially expressed in rodent enamel organs during amelogenesis, although the fold change ratios for both were all within ±1.35 and the changes were not significant. Of these, mutations in *TSC2* are known to cause tuberous sclerosis 2 (OMIM #613254), a condition that can encompass a distinctive form of AI (Schwartz et al., 2007, Harutunian et al., 2011). However no evidence of tubular sclerosis was observed in individual 2975. Both *MUC6* and *KRT25* belong to extended gene families that consist of multiple paralogous genes. Alignment and subsequent variant calling may therefore be error prone for reads falling within these genes. In addition, variants within these genes are commonly observed in filtered variant lists generated by WES, however, the variants remain unconfirmed and of are of unknown significance.

Further investigation of the enamel defect in 2975 (family AI-138) revealed that it was refined to only the first permanent molars, the upper incisors, one upper canine tooth and one lower incisor, with no record of it having been found in the deciduous dentition. The dentist thought that the phenotype did not resemble AI, but was more suggestive of an environmental or life history associated

enamel defect. The clinical notes also detailed that 2975 (family 138) had suffered from perinatal fevers. This further suggests that the phenotype may be due to a combination of USH2 and an enamel defect other than AI such as MIH. For family AI-12, it was not possible to obtain further enamel phenotype details since contact had been lost with the recruiting clinician.

In summary, both families AI-12 (1509 and 1514) and AI-138 (2975) carry multiple variants in *USH2A*. Of the variants carried by 1509 and 1514 (family AI-12), the biallelic framshift variants, c.2299delG (p.E767Sfs*21) and c.11875_11876delCA (p.Q3959Nfs*53) are likely to cause USH2. Of the variants carried by 2975 (family AI-138), biallelic frameshift and splicing variants, c.2299delG (p.E767Sfs*21) and c.11549-1G>A (p.?) are likely to cause USH2. Therefore both AI-12 and AI-138 were excluded from further analysis in the search to identify a causative gene for HS.

3.2.4 Low coverage whole genome sequencing of Heimler syndrome patient DNA

DNA from individuals 1492 (AI-10), 1502 (AI-11), 2113 (AI-16), 2645 (AI-71), 3466 (AI-152) and 4716 (AI-186) was submitted to Leeds St James' University Hospital Next Generation Sequencing Facility for low coverage WGS, to screen for CNVs and rearrangements that might not be detected in WES data (Section 2.4.7). Analysis was undertaken by the same facility and results were returned in the form of html files (Section 2.4.8). Depth of coverage was calculated separately as previously described for WES (Section 2.4.5) but using the GRCh37 reference sequence as the interval.

Coverage was calculated to be 0.61x on average (Appendix 11). The resulting html files detailing the positions of CNVs detected were manually checked to identify CNVs that encompassed gene exons and that were not reported in the Database of Genomic Variants (MacDonald et al., 2014). DNA from all samples had exonic CNVs of unknown significance but no two families had CNVs over the same gene. With the exception of data from family AI-11, these results were

not analysed further but were archived as a data set to be revisited should further WES reveal a candidate gene for HS.

3.2.4.1 Identification of a putative CNV involving *MYO7A* and *CAPN5* in family AI-11

Although the CNVs detected for the majority of samples were of unknown significance, results for sample 1502 (AI-11) included a heterozygous deletion of the first two exons of myosin VIIA (*MYO7A*; Section 1.11) and the 3′UTR of calpain 5 (*CAPN5*; MIM *602537). A copy number plot for chromosome 11 for sample 1502 that includes the region of the heterozygous deletion is shown in Figure 3.5. Window boundaries and midpoints for the deletion were listed as chr11:76,835,000-76,845,306 and chr11:76,837,355-76,842,508 respectively, denoting a deletion of size 5,154bp minimum and 10,307bp maximum (Table 3.6 and Figure 3.6). The relative copy number for this region was calculated as 0.53. Overall resolution, based on 4 windows, each with an average size of 5,831bp, was 23kb for this sample.

The CNVs identified by low coverage WGS were then compared with WES data-based CNV analysis for the other individuals from family AI-11, 1503 and 1504. The batch of WES analysed on the sequencer lane that included AI-11 libraries 1503 and 1504 also included 1492 (AI-10), 1509 and 1514 (both AI-12) and another unrelated sample with AI. Having only three other families in the same batch was considered insufficient to serve as control data for detection of CNVs by ExomeDepth (Section 2.4.6). Instead, around sixty other sets of WES data, obtained from libraries prepared and sequenced using identical conditions but in different batches to AI-11 samples 1503 and 1504, were used as controls for FishingCNV (Section 2.4.6). The resulting prioritised CNV calls were then used to filter the low coverage WGS CNV calls obtained for 1502. After filtering, only two CNVs remained that were unreported by Conrad et al. (2010), exonic and detected in all three family members (Table 3.6). These included a heterozygous deletion of ATPase family, AAA+ domain containing 3B (*ATAD3B*; MIM *612317) and *ATAD3C* and the heterozygous partial deletion of

CAPN5 and *MYO7A* observed on low coverage genome sequencing of individual 1502.

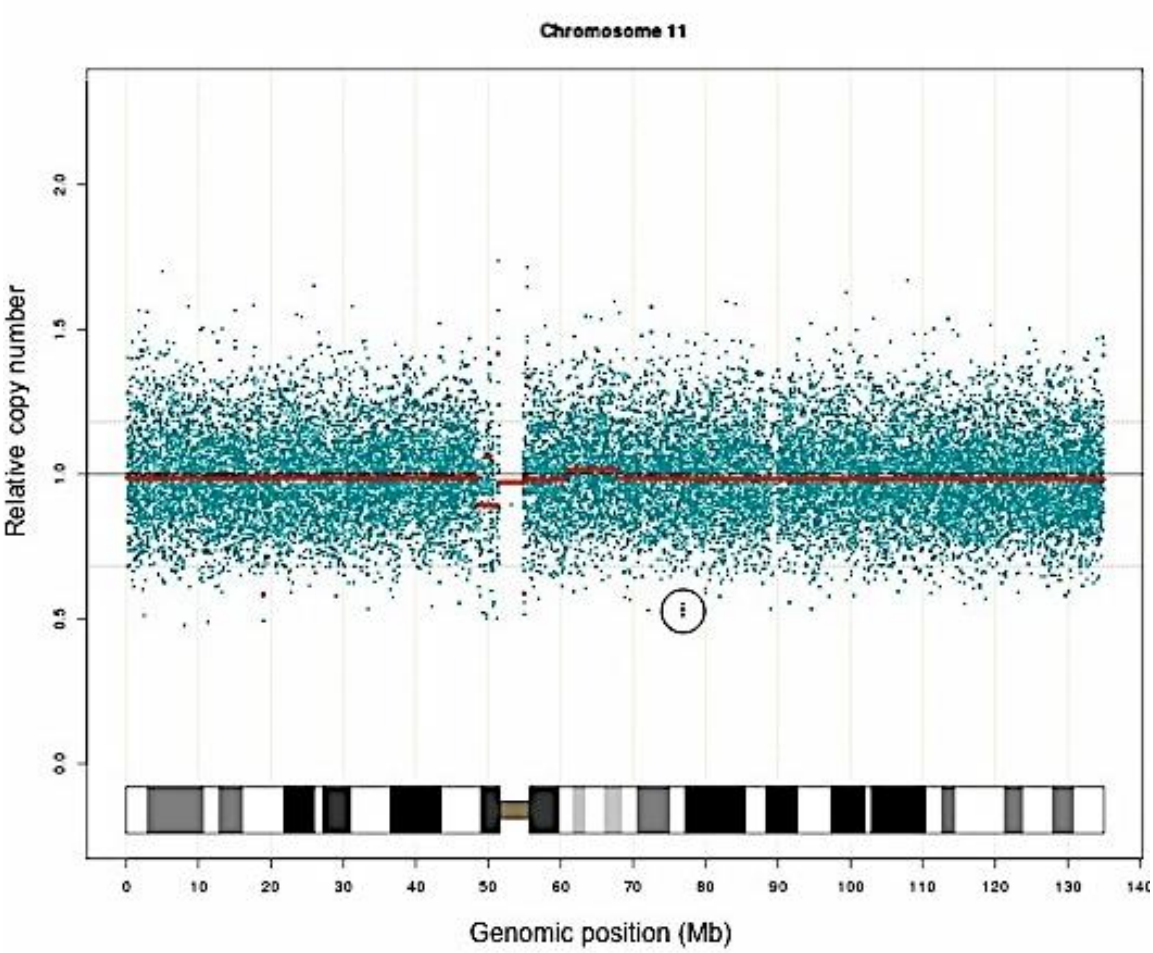


Figure 3.5: Relative copy number plot of chromosome 11 in individual 1502 (family AI-11). Plot details the copy number at the position of each genomic "window" assigned by DNA copy. The black line represents a normalised average copy number based on calls for aggregated control sequences. Each blue dot represents the relative copy number for each genomic window for 1502. The red line represents the relative copy number over chromosome 11 for 1502, based on groups of genomic windows. The circle highlights the heterozygous deletion of *MYO7A* exons 1 and 2 (point shown in red).

The heterozygous deletion of *ATAD3B* and *ATAD3C* was detected in 1503 and 1504, although the 3′ limit of the call differed for the two individuals. The FishingCNV call for 1503 indicated that the heterozygous deletion may be limited to exons 1-4 (inclusive) for *ATAD3B* (NM_031921), although calls for 1502 and 1504 suggested that the entire gene may be deleted. The segmentation mean values were -0.890 and -0.645 with Holm adjusted p values

of 6.90 x10^{-15} and 2.81 x10^{-15} respectively for individuals 1503 and 1504. The segmentation mean values equate to a relative copy number of 0.540 and 0.639 for samples 1503 and 1504 respectively for the region.

Reanalysis of WES data for 1503 and 1504 failed to identify any heterozygous variants in *ATAD3B* and *ATAD3C* that would be expected for recessive disease. Reanalysis was necessary since the variants could have been removed during the original filtering strategy due to their zygosity. Analysis of coverage of the *ATAD3B* and *ATAD3C* region revealed that several exons, in both genes, were covered by fewer than ten reads for both 1503 and 1504. Therefore, since these regions were not Sanger sequenced, it remains possible that a second, unidentified variant may be present in either *ATAD3B* or *ATAD3C*. Mutations in *ATAD3B* and *ATAD3C* have not been reported to result in disease in humans.

The heterozygous deletion over *MYO7A* was detected in 1503 and 1504 as spanning chr11:76,833,544-76,841,805 (The final two 3′ exons of *CAPN5* and the first two 5′ exons of *MYO7A*), with segmentation mean values of -0.8076 and -0.6836 with Holm adjusted p values of 1.41 x10^{-6} and 1.73 x10^{-9} respectively. The segmentation mean values equate to a relative copy number of 0.57 and 0.62 for samples 1503 and 1504 respectively for this region.

Sample	**Genomic position (GRCh37)**	**Size (bp)**	**Relative copy number (diploid=1; to 3 s.f)**	**Genes**	**Holm adjusted p value (FishingCNV only)**
1502	1:1387895-1430620	42,725	0.514	*ATAD3B*, *ATAD3C*	N/A
1503	1:1387417-1416232	28,815	0.540	*ATAD3B*, *ATAD3C*	1.37 x10^{-11}
1504	1:1387417-1431087	43,670	0.639	*ATAD3B*, *ATAD3C*	2.81 x10^{-12}
1502	11:76837355-76842508	5,153	0.534	*MYO7A*	N/A
1503	11:76833544-76841805	8,261	0.571	*CAPN5*, *MYO7A*	2.92 x10^{-9}
1504	11:76833544-76841805	8,261	0.623	*CAPN5*, *MYO7A*	3.21 x10^{-6}

Table 3.6: CNV calls shared by AI-11 family members 1502, 1503 and 1504.

Calls for 1502 were obtained from low coverage WGS data using DNAcopy. Genomic positions indicate DNAcopy window midpoints. Calls for 1503 and 1504 were obtained from WES data using FishingCNV.

Mutations in *MYO7A* are a known cause of USH type IB (USH1B) (Weil et al., 1995) and isolated SNHL with both recessive and dominant inheritance (Liu et al., 1997a, Liu et al., 1997b). Since USH1B has an AR mode of inheritance, the WES data for 1503 and 1504 were checked for heterozygous variants within *MYO7A* that would have been filtered out by the strategy employed in the original analysis due to their zygosity. Upon reanalysis, no further variants within *MYO7A* were identified. Since poor coverage over an exon containing a causative variant could potentially prevent its detection, coverage over *MYO7A* was assessed. It was found that exons 2, 13, 16, 32, 35, 38 and 42 contained regions of poor coverage (fewer than ten reads). Therefore these regions were Sanger sequenced, but no further variants were identified.

Since digenic inheritance has been suggested for USH, the possibility of a second USH variant in another USH gene was investigated. Upon reanalysis, no variants with a CADD score of greater than 2.7 were identified in both 1503 and 1504 (AI-11) within *USH1C*, *CDH23*, *PCDH15*, *USH1G*, *CIB2*, *USH2A*, *ADGRV1*, *DFNB31*, *CLRN1*, *PDZD7* and *HARS.*

Both predicted heterozygous deletions remain unconfirmed by Sanger sequencing and are of unknown significance.

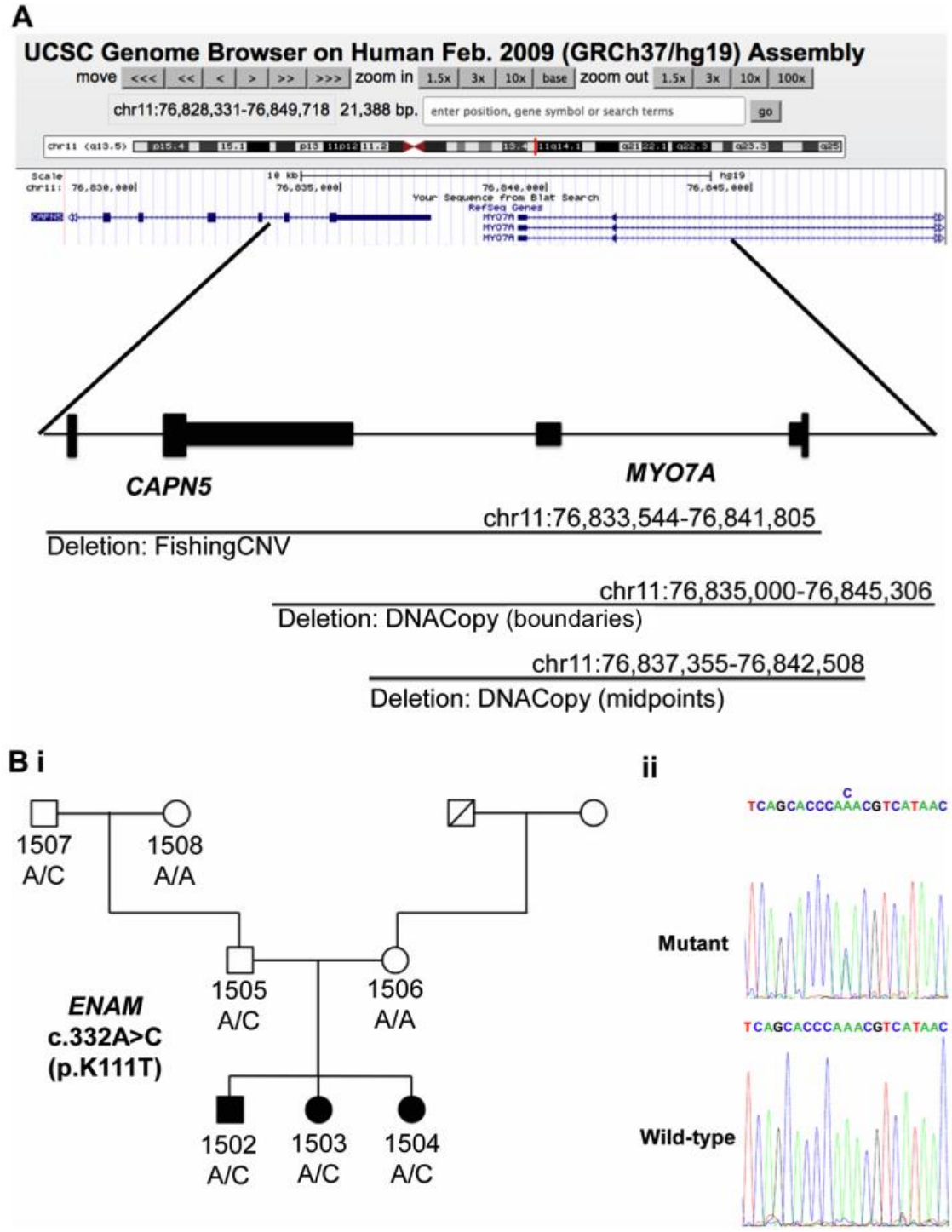


Figure 3.6: Schematic of the location of a predicted deletion of *MYO7A* exons 1 and 2 and segregation analysis of an *ENAM* SNV in family AI-11.

A: Schematic to show the locations of selected *CAPN5* and *MYO7A* exons relative to a predicted deletion with unconfirmed breakpoints. The deletion was predicted to span *MYO7A* exons 1 and 2 and was detected in 1502 using DNACopy analysis of WGS data (window boundaries and midpoints are shown). The deletion was also detected in 1503 and 1504 by FishingCNV analysis of WES data.

B: Segregation analysis of a heterozygous *ENAM* variant, c.332A>C (p.K111T); i: pedigree of Family AI-11; ii: Sanger sequencing trace of mutant and wild-type sequences.

3.2.5 Identification of an *ENAM* variant in family AI-11

After filtering WES data for AI-11 based on an AR mode of inheritance (Section 2.4.3), eleven variants remained (Appendix 12), of which, only *WDR89*, *ABCC11* and *PKHD1L1* had biallelic variants with CADD scores >15. These are not known AI, USH, retinal dystrophy or SNHL genes, and an examination of the available literature did not appear to support their involvement in AI. They were therefore archived for later study in the event that no better candidates were found.

In order to screen for potentially causative mutations for the AI phenotype, data was filtered to select variants within known AI genes without filtering for biallelic variants. A rare SNV in *ENAM* (Section 1.9.3), c.332A>C, p.K111T, was identified in WES data for both 1503 and 1504 with a CADD score of 16.01 (Appendix 13). The variant was confirmed by Sanger sequencing and its presence checked for the remaining family members Figure 3.6B. All affected AI-11 family members (1502, 1503, 1504) as well as the father (1505) and paternal grandfather (1507) carried this mutation. It therefore seems unlikely that this mutation is causative unless an AI phenotype in these two individuals had been missed. Loss of contact with the recruiting clinician meant that the phenotype of this family could not be reanalysed in more detail to determine whether the father or the paternal grandfather had AI. However analysis with mutation prediction software and investigation of public databases of variation would suggest that this variant is benign (Appendix 13).

3.2.6 Identification of *PEX1* variants in family AI-202

After filtering of WES data (Section 3.2.2) from individual 4505 (family AI-202), twenty variants remained (Appendix 14). Among these were biallelic variants in *PEX1*, including a single nucleotide duplication leading to a frameshift c.2097dupT, p.I700Yfs*42 and a missense variant, c.1742G>C, p.R581P (annotation based on NM_000466.2).

The *PEX1* variants were verified by Sanger sequencing of DNA from all available family members (Figure 3.7 and Appendix 7). The variants segregated

with the disease phenotype. Biallelic mutations in *PEX1* are known to cause the ZSSDs (Section 3.3.1). The maternally inherited, single nucleotide duplication variant identified in individual 4505, c.2097dupT, had previously been reported in individuals with ZSSDs (Collins and Gould, 1999). The allele inherited from the father that carried the SNV had not previously been reported in ZSSD patients. To check the conservation of the mutated residue within the orthologues of PEX1, the human sequence (NP_000457) was used to obtain orthologous sequences using BLAST. Subsequent alignment using Clustal Omega showed that the substitution altered a highly conserved residue within the PEX1 protein (Figure 3.8A).

The diagnosis of milder ZSSDs, such as IRD, especially for disease that presents as isolated sensory deficits, can be delayed until the teenage years (Waterham and Ebberink, 2012). ZSSDs are also known to have been misdiagnosed as USH or Leber congenital amaurosis (LCA) (Raas-Rothschild et al., 2002, Michelakakis et al., 2004, Majewski et al., 2011). Therefore, it was considered possible that individual 4505 had sensory defects due to *PEX1* mutations and that a separate mutation was responsible for the AI phenotype.

In order to check for potential dominantly inherited variants in genes associated with AI to account for the dental phenotype, data was filtered to select variants within known AI genes without filtering for biallelic variants. A heterozygous mutation in *LAMA3*, c.6389C>T, p.S2130F was identified, confirmed and shown to be maternally inherited (Appendix 15). The mother was re-examined but did not have AI.

A literature search was conducted to see whether a tooth phenotype had been previously reported in ZSSD patients. It identified four reports of such defects, including tooth formation and eruption defects as well as enamel hypoplasia, in ZS and IRD patients (Bader et al., 2000, Poll-The et al., 2004, Tran et al., 2011, Lertsirivorakul et al., 2014). In one cohort of 24 patients aged four years or over, dental defects were reported in one third (Poll-The et al., 2004). Therefore, it was concluded that the combination of sensory defects and AI in 4505 could be due to the biallelic mutations in *PEX1*.

3.2.7 The search for *PEX1* and *PEX6* variants in other Heimler syndrome individuals

After identification of biallllelic *PEX1* variants in AI-202, we revealed our findings to Prof. W. Newman (University of Manchester). He confirmed that he had also identified biallelic mutations in *PEX1*, and in the functionally-related gene *PEX6,* in HS patients.

The results of the original filtering strategy (Section 3.2.2), which aimed to identify rare, biallelic variants in each patient from the HS cohort, had failed to identify biallelic mutations in *PEX1* or *PEX6* in any family other than AI-202. In order to check for the possibility that causative *PEX1* variants might have been filtered out by the original strategy, the functional variants (Section 2.4.3) within the *PEX1* gene were selected from the unfiltered variant lists. *PEX1* is known to be the most common cause of PBDs but other genes, including *PEX2*, *PEX3*, *PEX5*, *PEX6*, *PEX7*, *PEX10*, *PEX11B*, *PEX12*, *PEX13*, *PEX14*, *PEX16*, *PEX19* and *PEX26* are also known to cause PBDs (Shimozawa et al., 1992, Dodt et al., 1995, Yahraus et al., 1996, Braverman et al., 1997, Chang et al., 1997, Motley et al., 1997, Purdue et al., 1997, Reuber et al., 1997, Honsho et al., 1998, Warren et al., 1998, Matsuzono et al., 1999, Shimozawa et al., 1999, Shimozawa et al., 2000, Matsumoto et al., 2003, Shimozawa et al., 2004, Ebberink et al., 2012). Therefore, functional variants within these genes were also selected, as well as functional variants within two other *PEX* genes not identified as a cause of PBDs, *PEX11A* and *PEX11G*.

Using this method, biallelic variants in *PEX1* were identified in 2645 (family AI-71) (Figure 3.7 and Table 3.7). The two variants were a single nucleotide deletion, c.1792delA, p.Q598Tfs*11 and an SNV, c.2966T>C. The frameshift variant had not been identified previously in ZSSD patients, whereas the missense variant had been reported previously (Maxwell et al., 2005). Variants were confirmed by Sanger sequencing but no additional family members were available for segregation analysis (Appendix 7).

Biallelic variants were also identified in *PEX6* for three other families; AI-10 (1499 and 1500), AI-146 (3000) and AI-152 (3466 and 4922) (Figure 3.7 and Table 3.7). All variants were confirmed by Sanger sequencing and segregated with the disease phenotype in all available family members (Appendix 7).

Affected individuals 1492, 1499 and 1500 from family AI-10 and 3000 from family AI-146 shared an identical heterozygous SNV, c.1802G>A, p.R601Q, that has been reported previously in ZSSD patients (Yik et al., 2009). From the maternal allele, affected individuals (1492, 1499 and 1500) in AI-10 carried a second heterozygous SNV, c.654C>G, p.F218L which was previously unreported. Individual 3000 (AI-146) also carried a previously unreported heterozygous variant, c.275T>G, p.V92G. No other samples were available for segregation analysis for AI-146.

Individuals 3466 and 4922 (AI-152) were homozygous for an SNV previously reported in ZSSD patients, c.1715C>T, p.T572I (Raas-Rothschild et al., 2002) (Figure 3.7 and Table 3.7).

For the individuals in the remaining families, AI-16 and AI-186, only one heterozygous variant in *PEX6* was identified from the WES data in each. In 2113 (AI-16), the previously unreported SNV, c.2714G>T, p.C905F was identified. In 4716 (AI-186), the *PEX6* SNV, c.1802G>A, p.R601Q, also found in 1499 and 1500 (AI-10) and in 3000 (AI-146), was identified as well as a heterozygous SNV in *PEX26*, c.728C>T, p.A243V. The presence of the *PEX6* variant, previously identified in two other HS families, in 4716 (AI-186), together with the low CADD score and other benign mutation prediction scores for the *PEX26* variant identified, led to the suspicion that 4716 (AI-186) may have an undetected *PEX6* variant. It was suspected that regions of low coverage might harbour additional heterozygous *PEX6* variants for both 2113 (AI-16) and 4716 (AI-186).

Individuals from families AI-11 (1503 and 1504), AI-12 (1509 and 1514) and AI-138 (2975) were not found to carry variants in any *PEX* genes.

3.2.8 WES depth of coverage analysis of *PEX1*, *PEX6* and *PEX26* for 2113 (family AI-16) and 4716 (family AI-186)

In order to identify regions of low coverage in WES data from 2113 (AI-16) and 4716 (AI-186) that may harbour additional PEX gene variants, the depth of coverage over genes *PEX1*, *PEX6* and *PEX26* was assessed. These genes were selected because heterozygous variants had already been identified in *PEX6* and *PEX26* in 2113 (AI-16) and 4716 (AI-186), and biallelic mutations in both *PEX1* and *PEX6* had been identified in the remainder of the HS cohort and Prof. Newman's cohort.

Bases covered by fewer that ten reads, and positioned within an exon or a splice site, were identified and screened by Sanger sequencing. For 2113 (AI-16), *PEX1* exon 7 and *PEX6* exons 1, 5 and 8, and for 4716 (AI-186), *PEX6* exons 1, 8 and 12, were not covered in their entirety by ten or more reads. This targetted sequencing revealed additional mutations in *PEX6* in both families. In 2113, an 8bp deletion was identified, c.1314_1321delGGAGGCCT, p.E439Gfs*3, which had been previously reported in ZSSD patients (Krause et al., 2009). In 4716 (AI-186), an SNV previously unreported in ZSSD patients, c.296G>T, p.R99L was identified. The compound heterozygous variants co-segregated with the disease phenotype in both AI-16 and AI-186 for all available family members. A summary of the *PEX1* and *PEX6* mutations identified in the HS cohort can be found in Table 3.7.

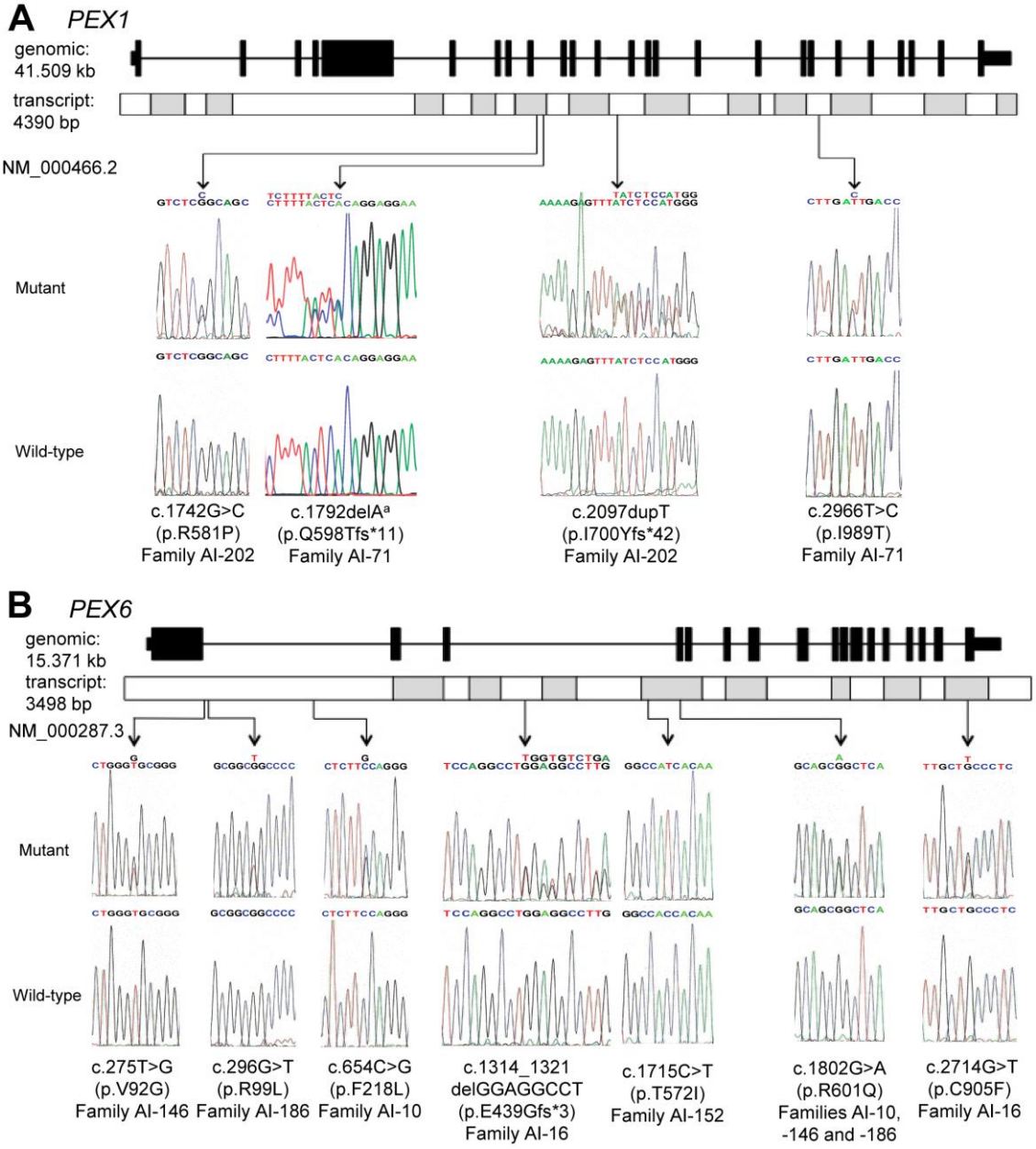


Figure 3.7: Sanger sequencing and genomic locations of the *PEX1* and *PEX6* mutations identified in this study.

A: A schematic diagram of *PEX1* genomic structure and transcript (NM_000466) shows the location and sequence traces of four mutations identified in this study.

B: A schematic diagram of *PEX6* genomic structure and transcript (NM_000287) shows the location and sequence traces of seven mutations identified in this study.

[a] The reverse sequence trace is shown for the *PEX1* c.1792delA variant.

Reproduced with permission from Smith et al. (2016b).

3.2.9 Bioinformatic analysis of the *PEX1* and *PEX6* variants in HS families

The *PEX1* and *PEX6* variants identified in the HS families AI-10, AI-16, AI-71, AI-146, AI-152, AI-186 and AI-202 were investigated by use of mutation prediction software, and their frequency in publically available databases was determined. Conservation analysis was undertaken for all missense variants using ClustalOmega.

Mutation prediction software including SIFT, Polyphen-2 (HumVar) and MutationTaster, generally concluded that all identified variants were pathogenic. Provean recorded neutral scores for three of the *PEX6* variants, p.V92G, p.R99L and p.F218L, identified in families AI-146 (3000), AI-186 (4716) and AI-10 (1499 and 1500) respectively.

Investigation of the variants' frequencies in publically available databases revealed that the majority are extremely rare or absent (Table 3.8). Although only one of the four *PEX1* variants was completely unreported, in contrast, four of the seven *PEX6* variants were unreported. The *PEX1* c.2097dupT variant has multiple records in the LOVD dbPEX database (available at http://www.dbpex.org/home.php) indicating that the variant has been previously identified in 87 ZSSD patients. dbPEX contains variant information for 52 and 115 unique variants for *PEX1* and *PEX6* respectively (Date accessed 19th February 2016) and therefore provides a unique resource of probable pathogenic variants that have been previously identified in ZSSD patients. In contrast, the *PEX6* c.1802G>A variant has been reported at a relatively high frequency in the ExAC and EVS databases but relatively few times in dbPEX.

Conservation analysis for missense variants that have not been previously reported (PEX1 p.R581P and PEX6 p.V92G, p.R99L, p.F218L and p.C905F) revealed that PEX1 p.R581 and PEX6 p.F218 and p.C905 residues are conserved through evolution in all mammalian species tested and zebrafish (Figure 3.8). Of these residues, only p.C905 was conserved in thale cress and none were conserved in fruit fly.

Gene	Genomic variant (GRCh37)	Exon	Transcript variant	Amino acid change	dbSNP142 reference	CADD v1.3	Provean	SIFT	Polyphen2 (HumVar)	Mutation Taster
PEX1	7:92136369C>G	10	c.1742G>C	p.R581P	N/A	21.8	Deleterious (-3.16)	Damaging (0.006)	Probably damaging (0.923)	Disease causing (0.926)
PEX1	7:92136319delT	10	c.1792delA	p.Q598Tfs*11	N/A	35	N/A	N/A	N/A	Disease causing (1)
PEX1	7:92132483dupT	13	c.2097dupT	p.I700Yfs*42	rs61750415	27.6	N/A	N/A	N/A	Disease causing (1)
PEX1	7:92123671A>G	19	c.2966T>C	p.I989T	rs61750427	29.6	Deleterious (-4.82)	Damaging (0.001)	Probably damaging (0.995)	Disease causing (1)
PEX6	6:42946614A>C	1	c.275T>G	p.V92G	N/A	22.9	Neutral (-1.5)	Damaging (0)	Probably damaging (0.987)	Disease causing (1)
PEX6	6:42946593C>A	1	c.296G>T	p.R99L	N/A	29.4	Neutral (-1.83)	Damaging (0)	Probably damaging (0.982)	Disease causing (0.787)
PEX6	6:42946235G>C	1	c.654C>G	p.F218L	N/A	23.8	Neutral (-2.06)	Damaging (0.005)	Possibly damaging (0.755)	Disease causing (0.982)
PEX6	6: 42937452-42937459 delAGGCCTCC	5	c.1314_1321 delGGAGGCCT	p.E439Gfs*3	rs267608216	33	N/A	N/A	N/A	Disease causing (1)
PEX6	6:42935275G>A	8	c.1715C>T	p.T572I	rs61753224	23.5	Deleterious (-4.54)	Damaging (0.001)	Probably damaging (0.963)	Disease causing (1)
PEX6	6:42935188C>T	8	c.1802G>A	p.R601Q	rs34324426	35	Deleterious (-3.25)	Damaging (0.001)	Probably damaging (0.964)	Disease causing (1)
PEX6	6:42932620C>A	16	c.2714G>T	p.C905F	N/A	34	Deleterious (-9.64)	Damaging (0.01)	Probably damaging (0.961)	Disease causing (1)

Table 3.7: Details of the four *PEX1* variants identified in families AI-71 and AI-202 and the seven *PEX6* variants identified in families AI-10, AI-16, AI-146, AI-152 and AI-186 by WES.

The list is restricted to variants scoring 15 or more when scored with CADD v1.3.

Provean, SIFT and Mutation Taster annotations were based on the relevant Ensembl transcript, Polyphen-2 annotations were based on the relevant RefSeq protein.

PEX1: Ensembl transcript: ENST00000248633, RefSeq protein NP_000457.

PEX6: Ensembl transcript: ENST00000304611, RefSeq protein NP_000287.

All scores were obtained 16th February 2016.

Genomic variant (GRCh37)	dbSNP142 reference number of variant alleles (total alleles at this position); allele frequency (all reported individuals heterozygous)	ClinVar (release 07/01/2016)	ExAC (v0.3) number of variant alleles (total alleles at this position); allele frequency (all reported individuals heterozygous unless otherwise stated)	EVS number of variant alleles (total alleles at this position); allele frequency (all reported individuals heterozygous)	dbPEX number of records
7:92136369C>G	rs370483961	ID 214068[a]	1 (121398) 8.24 x 10^{-6}	N/A	N/A
7:92136319delT	N/A	N/A	N/A	N/A	N/A
7:92132483dupT	rs61750415 no allele frequency stated	ID 7519	77 (120218) 6.41 x10^{-4}	10 (12518); 7.99 x 10^{-4}	87
7:92123671A>G	rs61750427; 2 (10016); 2.00 x 10^{-4}	N/A	6 (120686); 4.97 x 10^{-5}	N/A	1
6:42946614A>C	N/A	N/A	N/A	N/A	N/A
6:42946593C>A	N/A	N/A	N/A	N/A	N/A
6:42946235G>C	N/A	N/A	N/A	N/A	N/A
6:42937452-42937459 delAGGCCTCC	rs267608216; no allele frequency stated	N/A	4 (121224); 3.30 x 10^{-5}	1 (12518); 7.99 x 10^{-5}	N/A
6:42935275G>A	rs61753224; no allele frequency stated	N/A	N/A	N/A	3
6:42935188C>T	rs34324426; 5 (5008); 9.98 x 10^{-4}	ID 198709 (uncertain significance)	316 (98988); 3.19 x 10^{-3} (Includes 4 homozygotes)	42 (13006); 3.23 x10^{-3}	2
6:42932620C>A	N/A	N/A	N/A	N/A	N/A

Table 3.8: Details of the known variation in public databases for the variants in *PEX1* identified in families AI-71 and AI-202 and in *PEX6* identified in families AI-10, AI-16, AI-146, AI-152 and AI-186.

All databases accessed 17/02/2016.

[a] Variant identified in AI-202 and another HS family published by Ratbi et al. (2015), variant submitted to ClinVar 1st October 2015.

PEX1 p.R581P

```
Human        571 THSLLGR------PLSRQLMSLVAGLR 591
Chimpanzee   571 THSLLGR------PLSRQLMSLVAGLR 591
Marmoset     571 THSLLGR------PLSRQLMSLVAGLR 591
Guinea pig   559 TQSLLGR------PLSRQLTSLLAGLR 579
Mouse        572 THSLLGR------PLSRQLMALVAGLR 592
Rat          571 THSLLGR------PLSRQLMALVAGLR 591
Elephant     572 IHSLLGR------PLSRQLMSLVAGLR 592
Dog          555 THSLLGR------PLSRQLMSLVAGLR 575
Cow          570 THSLLGR------PLSRQLVSLVAGLR 590
Sheep        583 THSLLGR------PLSRQLVSLVAGLR 603
Cat          509 THSLLGR------PLSRQLMSLVAGLR 529
Sperm whale  559 THSLLGR------PLSRQLMALVAGLR 579
Horse        571 THSLLGR------PLSRQLMSLVAGLR 591
Zebrafish    535 SHALLGG------SLSRELISTGRGLR 555
Fruit fly    455 VDQVVQEL-----RMNLCLSADNSVMR 475
Thale cress  565 VSDVIKRMTVLLSPAAGMWFSKFKIPS 590
```

PEX1 p.I989T

```
Human        979 VLAATSRPDLIDPALLRPGRL 999
Chimpanzee   979 VLAATSRPDLIDPALLRPGRL 999
Marmoset     944 VLAATSRPDLIDPALLRPGRL 964
Guinea pig   968 VLAATSRPDLIDPALLRPGRL 988
Mouse        980 VLAATSRPDLIDPALLRPGRL 1000
Rat          979 VLAATSRPDLIDPALLRPGRL 999
Elephant     980 VLAATSRPDLIDPALLRPGRL 1000
Dog          963 VLAATSRPDLIDPALLRPGRL 983
Cow          978 VLAATSRPDLIDPALLRPGRL 998
Sheep        991 VLAATSRPDLIDPALLRPGRL 1011
Cat          917 VLAATSRPDLIDPALLRPGRL 937
Sperm whale  967 VLAATSRPDLIDPALLRPGRL 987
Horse        979 VLAATSRPDLIDPALLRPGRL 999
Zebrafish    935 VLAASSRPDLIDPALLRPGRL 955
Fruit fly    859 VIAATSRPELLDPALLRSGRI 879
Thale cress  983 VFAATRYSHPNKS-IL---HI 999
```

PEX6 p.V92G p.R99L

```
Human         82 RLLALGSGAWV----------RARAVRRPPALGWALLG 109
Chimpanzee    82 RLLALGSGAWV----------RARAVRRPPALGWALLG 109
Marmoset      82 RLLALGSGAWV----------RARPVRRPPALGWALLG 109
Guinea pig    82 QLLALGSGAQV----------RARPVRRPPALGWALLG 109
Mouse         82 RVLALGPGARV----------RARLVRRPPALGWALLA 109
Rat           82 RVLALSPGARV----------RARPVRRPPALGWALLG 109
Elephant      82 RLLALDSGAWV----------RARPVRRPPALGWALLG 109
Dog           82 RLLALGSGAWV----------RARPVRRPPALGWALLG 109
Cow           82 QLLALGSGAWV----------RARPVRRPPSLGWALLG 109
Sheep         82 QLVALGSGAWV----------RARPVRRPPSLGWALLG 109
Cat           82 RLLAL-------------------------------- 86
Sperm whale   37 ---------------------------------WALLG 41
Horse          0 ---------------------------------------- 0
Zebrafish     89 KHYGLREQSWGTLRPQSLLPLKKIVIGARTKPSFKWASSE 128
Fruit fly   1045 KHYGLREQSWGTLRPQSLLPLKKIVIGARTKPSFKWASSE 1085
Thale cress    0 ---------------------------------------- 0
```

PEX6 p.F218L


```
Human          208 SRSCLRGLGLFQGEWVWVAQA 228
Chimpanzee     208 SRSCLRGLGLFQGEWVWVARA 228
Marmoset       208 SRRCLRGLGLFQGEWVWVARA 228
Guinea pig     210 SRTCLRGLGLFQGEWVWVARA 230
Mouse          208 SRSCLRSLGLFQGEWVWVAQV 228
Rat            208 SRSCLRSLGLFQGEWVWVARV 228
Elephant       207 SRRCLRGIGLFQGEWVWVARA 227
Dog            208 SRSCLRSLGLFQGEWVWVTRA 228
Cow            208 SRSCLRSLSLFQGEWVWVTRA 228
Sheep          190 SRSCLRSLGLFQGEWVWVTRA 210
Cat            172 SRSCLRSLGLFQGEWVWVTRA 192
Sperm whale    140 SRSCLRSLSLFQGEWVWVTRA 160
Horse           57 SRSCLRSLGLFQGEWVWVTRA 77
Zebrafish      287 SRSLLVKLGVFNGEWVIASVP 307
Fruit fly     1244 SRSLLVKLGVFNGEWVIASVP 1264
Thale cress     79 STQLLKRLSINSGS LVVVKN 99
```

PEX6 p.T572I


```
Human          562 SCPPLMVVATTSRAQDLPADV 582
Chimpanzee     562 SCPPLMVVATTSRAQDLPADV 582
Marmoset       562 SCPPLLVVATTSRAQDLPAEV 582
Guinea pig     564 SCLPLMVVATTSRAQDLPADV 584
Mouse          563 RCPPLMVVATTSRVQDLPTDV 583
Rat            562 RCPPLMVVATTSRVQDLPTDV 582
Elephant       560 SCPPLMVVATTSRARDLPADV 580
Dog            562 SCPPLMVVATTSQAQDLPADV 582
Cow            562 SCPPLMVVATTSRAQDLPADV 582
Sheep          544 RXHPALSGESTERLQDLPADV 564
Cat            526 SCPPLITVATTSKAQDLPADV 546
Sperm whale    494 SCPPLMVVATTSRARDLPVDV 514
Horse          412 SCPPLMVVATTSRAQDLPADV 432
Zebrafish      655 --TSVVVVGSVSSQHELPSDV 673
Fruit fly     1521 ---------------------- 1521
Thale cress    502 RGHQVLLIASAESTEGISPTI 522
```

PEX6 p.R601Q


```
Human          591 LEVPALSEGQ--RLS--ILRALTAH 611
Chimpanzee     591 LEVPALSEGQ--RLS--ILRALTAH 611
Marmoset       591 LEVPALSEGQ--RLS--ILQALTAH 611
Guinea pig     593 LEMPVLSEGQ--RLS--ILRALTAH 613
Mouse          592 LEVPVLSEAQ--RLS--ILQALTAH 612
Rat            591 LEVPVLSESQ--RLS--VLQALTAH 611
Elephant       589 LEVPVLTEGQ--RLS--ILQALTAH 609
Dog            591 LEVPVLSEGQ--RLS--VLQALTAH 611
Cow            591 LEVPVLAEAQ--RLS--VLRALTAH 611
Sheep          573 LEGPVLAGAQGLRLS--VLRALTAH 595
Cat            555 LEVPVLSEGQ--RLS--VLRALTAH 575
Sperm whale    523 LEVPVLSEGQ--RLS--ILRALTAH 543
Horse          441 LEVPVLSEGQ--RLS--ILRALTAH 461
Zebrafish      682 VAIESLSEDQ--RRM--VLSSLSED 702
Fruit fly     1521 ------------------------ 1521
Thale cress    531 IRMGSLNDEQRSEMLSQSLQGVSQF 555
```

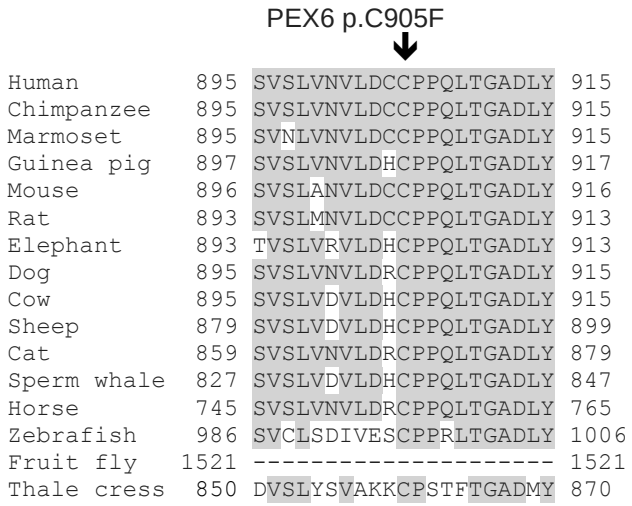


Figure 3.8: Clustal Omega multiple sequence alignment of homologous protein sequences for the residues surrounding each PEX1 and PEX6 missense variant identified in HS individuals.

Shaded residues indicate those that are identical to the human residue at the same position. Arrows indicate the position of the residues of interest.

PEX1 sequences used: Human, *Homo sapiens* NP_000457.1; Chimpanzee, *Pan troglodytes* XP_519198.2; Marmoset, *Callithrix jacchus* XP_009000794.1; Guinea pig, *Cavia porcellus* XP_003475146.1, Mouse, *Mus musculus* NP_001280735.1; Rat, *Rattus norvegicus* NP_001102690.1; Elephant, *Loxodonta africana* XP_003407197.1; Dog, *Canis lupus familiaris* XP_532459.1; Cow, *Bos taurus* NP_001179000.1; Sheep, Ovis aries XP_004008246.1; Cat, *Felis Catus* XP_011278778.1; Sperm whale, *Physeter catodon* XP_007103410.1; Horse, *Equus caballus* XP_001493415.2; Zebrafish, *Danio rerio* NP_001164306.1; Fruit fly, *Drosophila melanogaster* NP_652016.1; Thale cress, *Arabidopsis thaliana* NP_001331459.1.

PEX6 sequences used: Human, *Homo sapiens* NP_000278.3; Chimpanzee, *Pan troglodytes* XP_003950894.1; Marmoset, *Callithrix jacchus* XP_002746596.1; Guinea pig, *Cavia porcellus* XP_003473901.1, Mouse, *Mus musculus* NP_663463.1; Rat, *Rattus norvegicus* NP_476466.1; Elephant, *Loxodonta africana* XP_003403954.1; Dog, *Canis lupus familiaris* XP_538926.2; Cow, *Bos taurus* NP_001179876.1; Sheep, Ovis aries XP_004018878.1; Cat, *Felis Catus* XP_011280631.1; Sperm whale, *Physeter catodon* XP_007107679.1; Horse, *Equus caballus* XP_005614975.1; Zebrafish, *Danio rerio* XP_009294800.1; Fruit fly, *Drosophila melanogaster* NP_001163114.1; Thale cress, *Arabidopsis thaliana* NP_171799.2.

However the region encompassing PEX6 p.V92 and p.R99 residues was not present in three of the mammalian species tested and neither residue was conserved in zebrafish, fruit fly or thale cress.

3.2.10 Haplotyping of the region surrounding the common *PEX6* c.1802G>A variant

Three of the families in this study (AI-10, AI-146 and AI-186) and family 6, reported by Ratbi et al. (2015), were found to share the same *PEX6* c.1802G>A variant (rs34324426). All four families were all of either US or UK origin. This variant had also been reported in seven ZSSD patients for whom the ethnicities were not given but the studies themselves were of US origin (Yik et al., 2009, Ebberink et al., 2010). In addition, ExAC reports the allele's frequency in the European population as 0.416% and EVS reports the allele frequency in the European American population as 0.418% (Table 3.8). Therefore, evidence suggests that the c.1802G>A variant appears to be relatively common in populations of European origin.

In order to investigate whether the c.1802G>A variant represents a common founder mutation, analysis of the haplotype surrounding the variant was carried out by genotyping nearby microsatellite markers and SNPs (Appendix 16). A set of representative microsatellite genotyping electropherograms for marker D6S1650 is shown in Figure 3.9. For families AI-10 and AI-186, results from multiple family members enabled both haplotypes to be determined (Figure 3.10). For AI-146, only DNA from the affected individual (3000) was available and therefore the haplotype of the c.1802G>A variant could not be defined.

Results showed that the p.R601Q mutation was associated with a haplotype of two SNPs and one microsatellite spanning a region of at least 779kb (Table 3.9). However, for two of the families, AI-10 and AI-186, the shared haplotype extended over a region of at least 1,279kb.

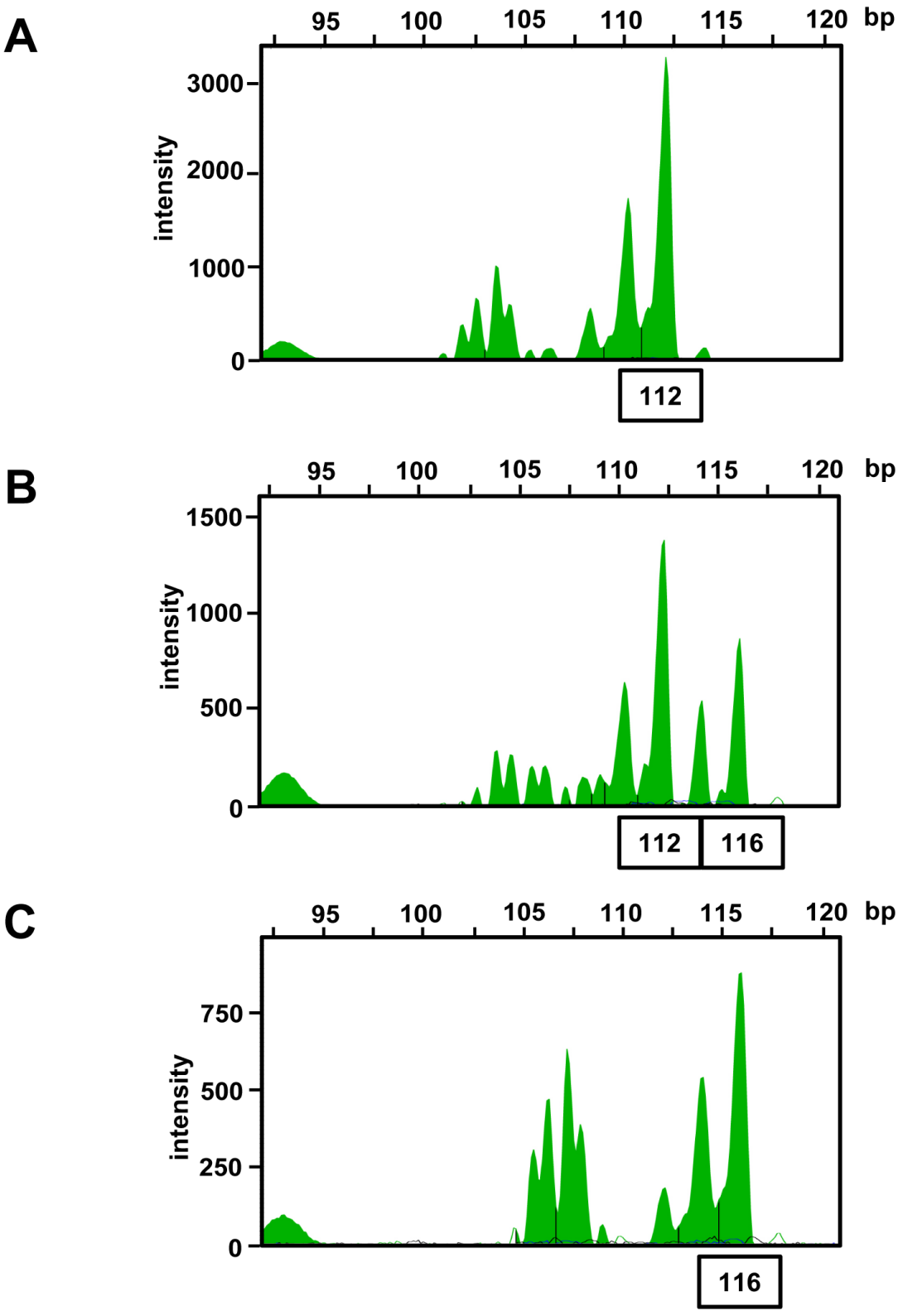


Figure 3.9: Examples of genotyping electropherograms.

Electropherograms showing two alleles of marker D6S1650 in AI-10 family members (A) 1492, (B) 1493 and (C) 1497 as visualised on GeneMapper.

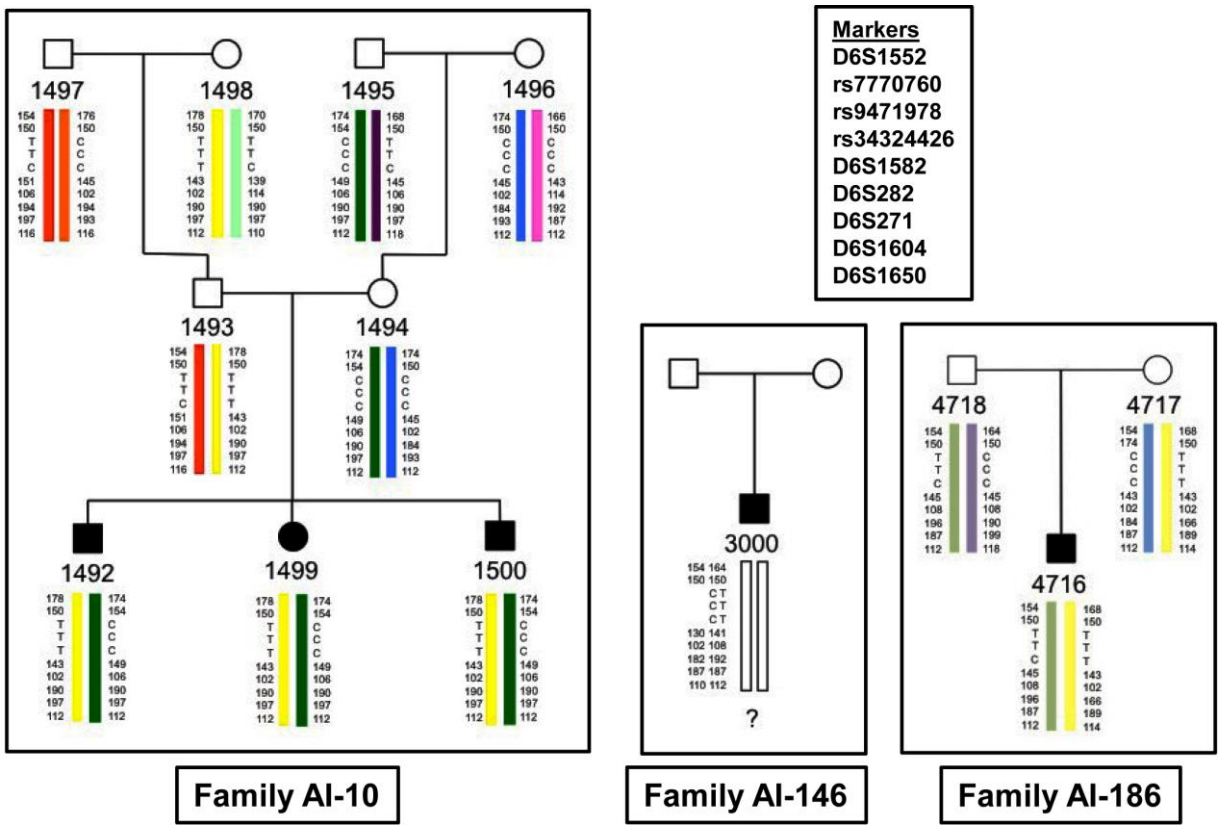


Figure 3.10: Analysis of the haplotype of the region surrounding the *PEX6* c.1802G>A variant for families AI-10, AI-146 and AI-186.

Results are shown for eight microsatellite markers and three SNPs, including the *PEX6* c.1802G>A variant. Markers used are described in Appendix 16. A summary of the results can be found in Table 3.9. Haplotypes are represented by coloured bars (each colour represents a different haplotype in each family). The numbers alongside each haplotype bar correspond to the size of each allele amplified by the microsatellite markers, in base pairs. The markers are displayed in order of their physical location. In families AI-10 and AI-186, the yellow haplotype carries the c.1802G>A variant and the green haplotypes carry the second *PEX6* variant.

Marker	Physical location (hg19 / GRCh37)	Genetic location (Marshfield) / cM	Result - family AI-10	Result - family AI-146	Result - family AI-186
D6S1552	6:41537757-41538086	62.28	178	154 / 164	168
D6S400	6:41955110-41955373	63.28	150	150 / 150	150
rs7770760	6:42915403	N/A	T	C / T	T
rs9471978	6:42929533	N/A	T	C / T	T
rs34324426	6:42935188	N/A	T	C / T	T
D6S1582	6:43098617-43098961	65.14	143	130 / 141	143
D6S282	6:43234575-43234945	66.37	102	102 / 108	102
D6S271	6:43500818-43501193	66.37	190	182 / 192	166
D6S1604	6:43679700-43680083	66.37	197	187 / 187	189
D6S1650	6:44130031-44130426	68.65	112	110 / 112	114

Table 3.9: Summary of genotypes comprising the haplotype segregating with the shared *PEX6* c.1802G>A variant (rs34324426) identified in families AI-10, AI-146 and AI-186.
Dark grey shading indicates the markers and SNPs with identical results for all three families, while the light grey shading indicates the markers and SNPs with identical results for families AI-10 and AI-186. For family AI-146, DNA was not available from additional family members, therefore it was not possible to determine the haplotype associated with the c.1802G>A variant.

3.2.11 Immunohistochemical staining of murine tissues

The pathological changes in the enamel and retinas of individuals with HS and other ZSSDs suggest that peroxisomes are crucial to the development of the dental enamel and the maintenance of a functional retina. However, at the start of the study there had been no previous reports detailing the location of PEX1 and PEX6 in these tissues.

3.2.11.1 Western blot

In order to carry out immunohistochemistry (IHC) on murine tissues, commercially available antibodies against PEX1 and PEX6 were obtained. An antibody against PEX14 was also obtained since PEX14 has been shown to be an optimal marker for identification and localisation of peroxisomes in a variety of cell types (Grant et al., 2013). Therefore the primary antibodies obtained were rabbit anti-human PEX14 polyclonal antibody (10594-1-AP, Proteintech, Rosemount, IL, USA), rabbit anti-human PEX1 polyclonal antibody (13669-1-AP, Proteintech, Chicago, IL, USA) and goat anti-rat PEX6 polyclonal antibody (ab175064, Abcam, Cambridge, UK). Although none of the antibodies obtained were raised against the mouse antigen, they were predicted to cross react with the corresponding murine antigen. In order to verify this and to attempt to

assess the specificity of the antibodies, western blots were carried out using 50μg total protein from murine kidney and brain tissues (Sections 0 and 2.11). Mouse kidney and brain were chosen since both tissues have been shown to express PEX1, PEX6 and PEX14 in various studies listed in the EMBL-EBI Expression Atlas (available at http://www.ebi.ac.uk/gxa/home). Samples were run under denaturing conditions, in both the presence and absence of a reducing agent. Both gel electrophoresis and transfer were run for 60 minutes (for gels and blots to be used for detection of PEX14) or 90 minutes (for gels and blots to be used for detection of PEX1 or PEX6). A Coomassie blue stained SDS PAGE gel is shown in Appendix 17.

Primary antibodies were applied at the following dilutions: anti-PEX14 1:250, anti-PEX1 1:500 and anti-PEX6 1:250. The secondary antibodies and dilutions used were: polyclonal goat anti-rabbit immunoglobulins/HRP 1:1000 (P0448; Dako, Ely, UK) for the blots incubated with the anti-PEX1 and anti-PEX14 antibodies and polyclonal rabbit anti-goat immunoglobulins/HRP 1:1000 (P0449; Dako) for the blots incubated with the anti-PEX6 antibody. Blots were stripped and reprobed with monoclonal mouse anti-β-actin antibody (ACTB; A5441, Sigma) to assess both protein loading and transfer to the membrane. The secondary antibody used was polyclonal rabbit anti-mouse 1:1000 (P0260; Dako). The results of the western blots are shown in Figure 3.11.

Western blot using the anti-PEX1 antibody (Figure 3.11A) showed that a band of approximately the expected size of 141.4kDa was present under reducing conditions (solid black arrows). Another band of approximately 90kDa was also present for both kidney and brain samples (dot filled arrows), a protein BLAST search, using the recombinant protein sequence used to raise the anti-PEX1 antibody, revealed that the antigen has some homology to nuclear valosin-containing protein-like protein isoforms, which have a predicted protein mass of around 94.5kDa. Upon reprobing the blot with the anti-β-actin antibody, it was evident that the protein loading and transfer had only been effective for the samples treated with reducing agent (Figure 3.11D).

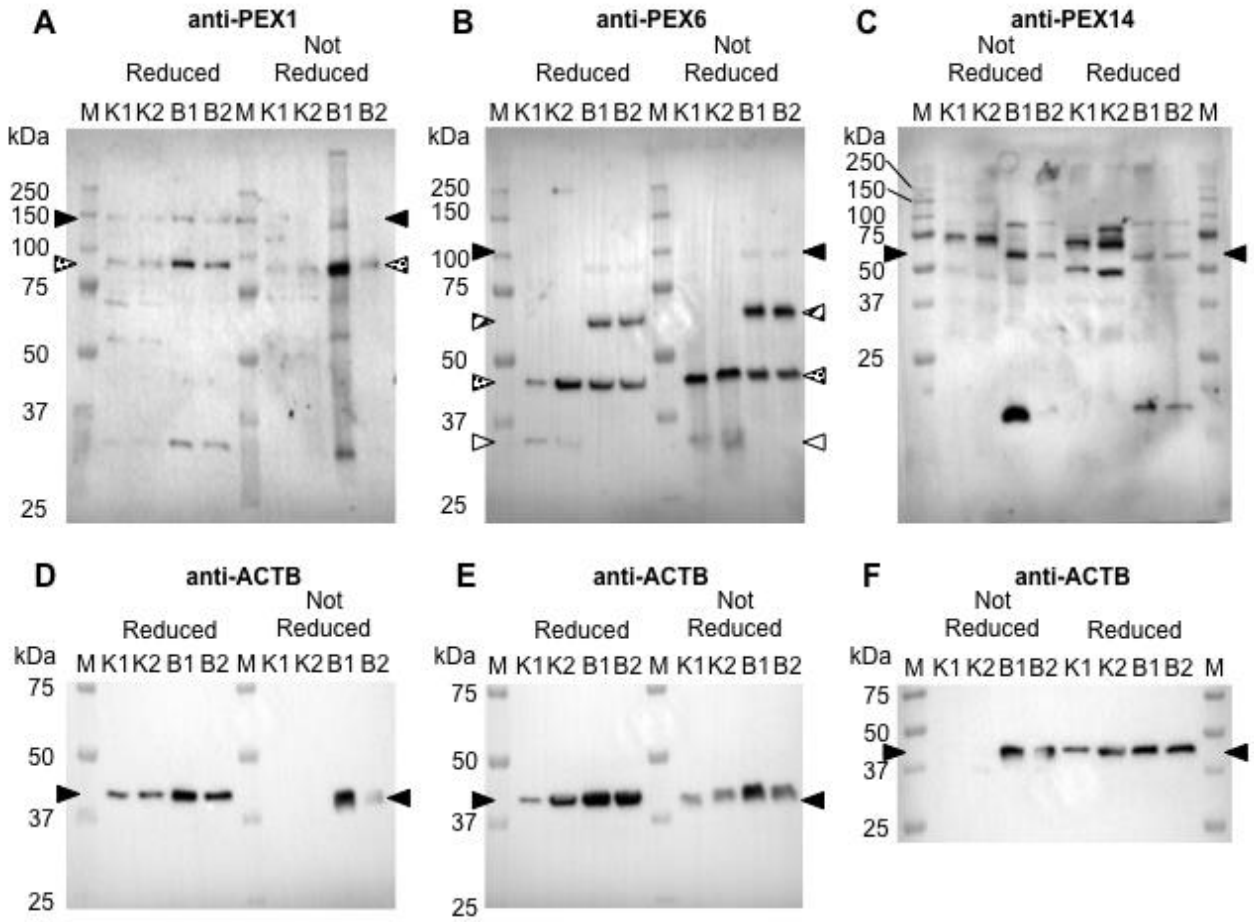


Figure 3.11: Western blots using anti-PEX1, -PEX6 and -PEX14 antibodies on murine tissues.

Western blots of 50µg kidney (K1, K2) and brain (B1, B2) total protein samples obtained from two mice. Blots were probed with the following primary antibodies: A: anti-PEX1, B: anti-PEX6, C: anti-PEX14, D-F: anti-ACTB (loading/transfer control for blots A-C respectively). Solid black arrows refer to bands corresponding to expected protein sizes. Other arrows refer to bands discussed in the text.

For the blot probed with anti-PEX6 antibody (Figure 3.11B), a faint signal of around 100kDa was present (solid black arrows). This was most evident for the brain samples treated without the reducing agent. This may represent PEX6 since the predicted protein weight is 104.5kDa. A protein BLAST search for the peptide sequence against which the anti-PEX6 antibody was raised revealed that it was also 100% identical to part of the PEX2 (MIM *170993) protein. PEX2 is predicted to be 34.9kDa and may represent the faint bands present around this size for the kidney samples (open arrows). Two other bands of around 45kDa (present for all samples; dotted arrows) and 60kDa (present only for brain; striped arrows) were evident. Amongst other proteins, the sequence used as the antigen also shares identity with FAM134A (61% coverage, 81% identity) which is predicted to be 57.5kDa and may represent the band present in brain samples (striped arrows). Reprobing the blot with the anti-β-actin antibody revealed that the protein loading and transfer had been effective for all samples (Figure 3.11E).

For the blot probed with anti-PEX14 antibody (Figure 3.11C), a signal at around 57kDa (based on the human protein) was expected. PEX14 protein is predicted to be 41.2kDa, but the antibody datasheet suggests that the protein does not migrate at the position that would be expected for its size, but at around 57kDa for the human protein. Bands around 55kDa were identified in brain samples (solid black arrows). A protein BLAST search of the peptide sequence used as the antigen to raise the antibody did not reveal any other sequences that matched other than the target. Reprobing the blot with the anti-β-actin antibody revealed that protein loading and transfer had only been effective for the samples treated with reducing agent and for the untreated brain samples (Figure 3.11F).

Therefore, on each blot, there were bands that corresponded to the respective sizes of each antigen for each antibody, suggesting that detection of the respective murine antigens with the antibodies raised against the human or rat antigens was possible. However other bands were present and therefore the western blots were inconclusive. The conditions present during SDS PAGE do

not reflect those present during immunohistochemistry, therefore it is not possible to make conclusions as to the efficacy of the antibodies.

3.2.11.2 Immunohistochemical staining of incisor tissues

Murine incisor provides an excellent model for the development of enamel since it grows continuously throughout life and therefore incisors display all stages of amelogenesis along their lengths (Section 1.7).

Therefore, in order to determine the localisation of PEX proteins throughout amelogenesis, immunohistochemical staining was undertaken. Mouse incisor tissue was collected, fixed, demineralised and sectioned as described (Section 2.12.1.1). Mandibular incisors were embedded planar to the mesial-distal axis and sectioned planar to the buccal-lingual axis. Staining of every tenth section with haematoxylin and eosin (H&E) enabled sections containing secretory, transition and maturation stage ameloblasts to be identified (Figure 3.12).

Due to the orientation during embedding and the natural shape of the murine incisors, the classic morphological changes in the ameloblast cells during amelogenesis became masked. The “molar line” has been used by other researchers as a morphological reference point to signify the secretion-transition-maturation stage boundary (Smith and Nanci, 1989). Ameloblasts immediately below the more distal root (of the two visible on a mid-line section) of the first molar are designated as residing at the transition between secretory and maturation stage, although the bodyweight of the animal can alter this position (Smith and Nanci, 1989). Ameloblasts that lay more mesial to the line are designated as secretory stage, while those that lay more distally are designated as maturation stage. In the H&E sections obtained here, the proximal root was not visible and the staining of the proteinaceous enamel matrix was lost mesial to the position of the molar line (Figure 3.12A). Therefore, this marker was not used as an approximate marker for the secretion-transition-maturation stage boundary since it may not have been accurate.

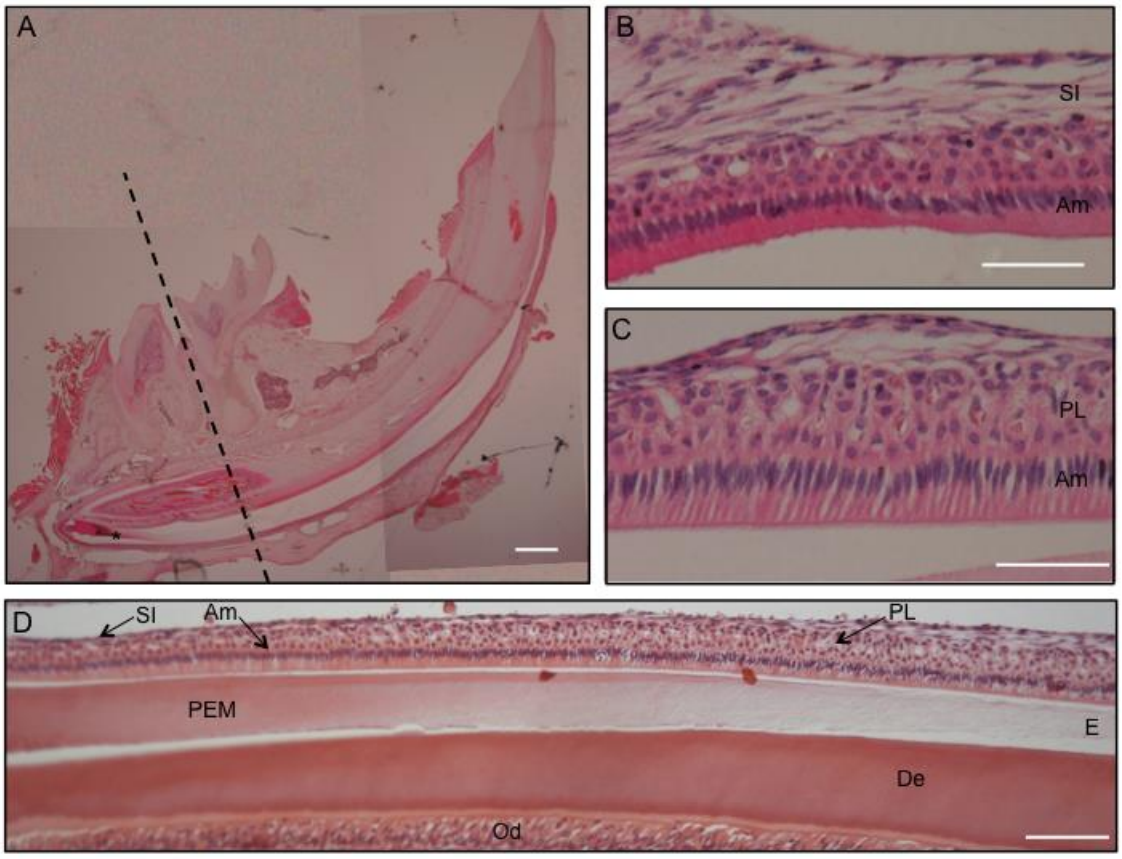


Figure 3.12: H&E staining of a mandibular incisor section.

A: Complete murine incisor section, the position of the molar line is marked with the dashed line. * indicates the proteinaceous enamel matrix (PEM).

B: Secretory stage ameloblasts (Am) and overlying stratum intermedium (SI).

C: Maturation stage ameloblasts with overlying papillary layer (PL).

D: A PEM is most visible adjacent to the secretory ameloblasts. An enamel space (E) reflects the presence of mineralised tissue that has been removed during demineralising treatment with EDTA. A layer of dentine (De) can also be seen that has been formed by odontoblasts (Od). Scale bars represent 600μm (A), 50μm (B and C) and 100μm (D).

Commercially available antibodies against PEX1, PEX6 and PEX14 were obtained as previously described. Although none of the antibodies obtained were raised against the mouse antigen, they were predicted to cross react with the murine antigen and bands of the predicted size of the murine antigen had been obtained by western blot (Section 3.2.11.1).

In order to optimise the conditions for IHC with each antibody, a range of antibody concentrations were tested with citrate-based antigen retrieval (Section 2.12.5). The highest antibody concentration was also tested without

any form of antigen retrieval. It was found that the incisor sections without any antigen retrieval did not stain. Those processed using citrate-based retrieval did stain, but the sections detached from the slides due to the microwave based boiling during antigen retrieval. Trials of different types of slides, including those treated with silane (CellPath, Newtown, Wales, UK and lab prepared silane slides (amino-propyl-tri-ethoxy-silane)) as well as heating using a lower intensity or for a shorter period of time were undertaken. Treatment of slides by baking at 70^{o}C in an oven for an hour prior to antigen retrieval was also tried. All combinations of treatments failed to provide conditions that resulted in both antigen retrieval and maintained the sections' adherence to the slide for imaging.

Therefore enzyme-based antigen retrieval using chymotrypsin, that did not involve a boiling treatment, was tested. A trial of antibody dilutions for the enzyme-based retrieval method found that optimal staining was achieved by using the anti-PEX1 antibody at a 1:300 dilution and anti-PEX14 at a 1:150 dilution. The anti-PEX6 antibody was used at a dilution of 1:80. Trials of higher dilutions resulted in loss of signal. For the anti-PEX1 and anti-PEX14 antibodies raised in rabbit, the secondary antibody used was the labelled polymer-HRP anti-rabbit reagent from the EnVision+ System-HRP (DAB), for use with rabbit primary antibodies (Dako) as described in the manufacturer's instructions. For the anti-PEX6 antibody, raised in goat, some modification to the standard method was required (Section 2.12.5). Avidin and biotin blocking (Vector Laboratories, Peterborough, UK) was used as described in the manufacturer's instructions, prior to primary antibody application for the sections treated with the anti-PEX6 antibody. The secondary antibody and dilution used for these sections was rabbit polyclonal anti-goat immunoglobulins/biotinylated (E0466, Dako) at 1:200. These sections were then treated with the Vectastain Elite ABC Kit (Vector Laboratories, Burlingame, CA, USA) as described in the manufacturer's instructions, prior to staining with DAB reagent (Dako) as previously described.

IHC using chymotrypsin-based antigen retrieval resulted in the no primary control treated slides also staining, although the staining of the no primary

antibody slide for the anti-goat secondary antibody was stronger than that for the anti-rabbit secondary antibody. An alternative anti-goat secondary antibody was sought but was unavailable for trial. Use of the anti-goat secondary antibody resulted in staining in secretory stage ameloblasts. Use of both the anti-PEX6 antibody and the secondary antibody resulted in staining of maturation stage ameloblasts (Figure 3.13).

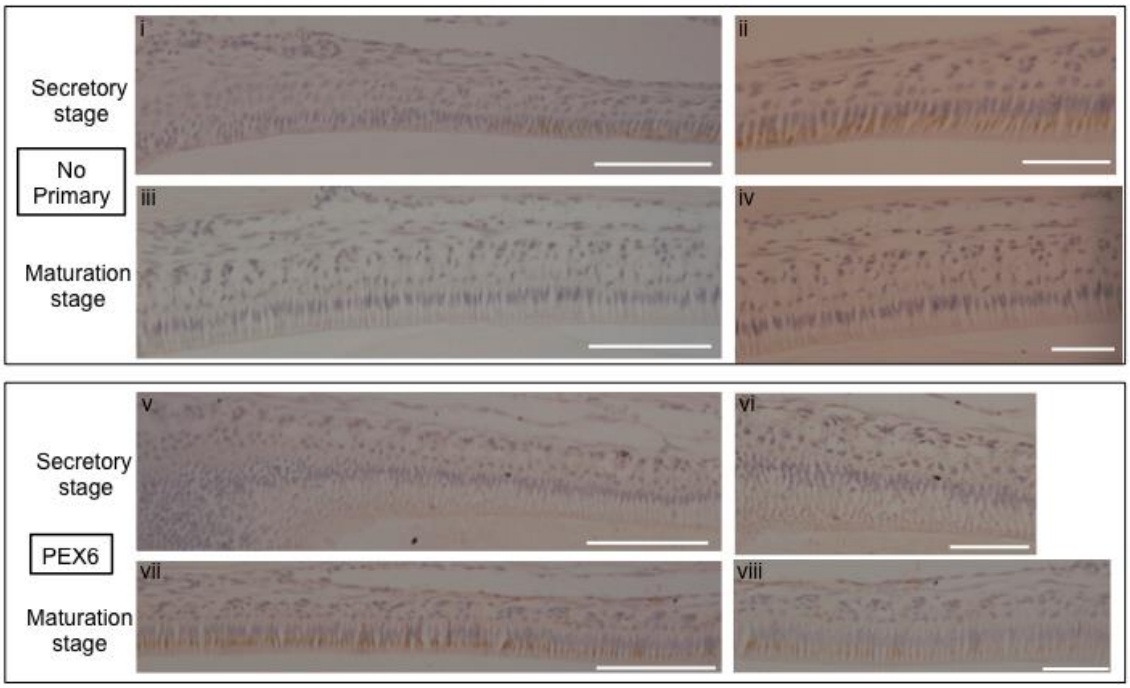


Figure 3.13: Immunohistochemical staining of murine ameloblasts for PEX6.

i-iv: No primary antibody control: secondary antibody (rabbit anti-goat) only. v-viii: anti-PEX6 antibody followed by secondary antibody. Secretory stage ameloblasts are shown in panels i, ii, v and vi. Maturation stage ameloblasts are shown in panels iii, iv, vii and viii. Scale bars for panels i, iii, v, vii represent 100μm, those for panels ii, iv, vi and viii represent 50μm.

Use of the anti-PEX1 antibody resulted in staining only in the proteinaceous enamel matrix, since PEX1 is a peroxisomal membrane protein and the enamel matrix is an acellular compartment, the staining could have been non-specific. The corresponding no primary control, treated only with the anti-rabbit secondary antibody, had staining that was restricted to late maturation stage ameloblasts (Figure 3.14iii), indicating that the staining with the anti-PEX1 antibody may not have been specific. No staining was present in the proteinaceous enamel matrix for the no primary control.

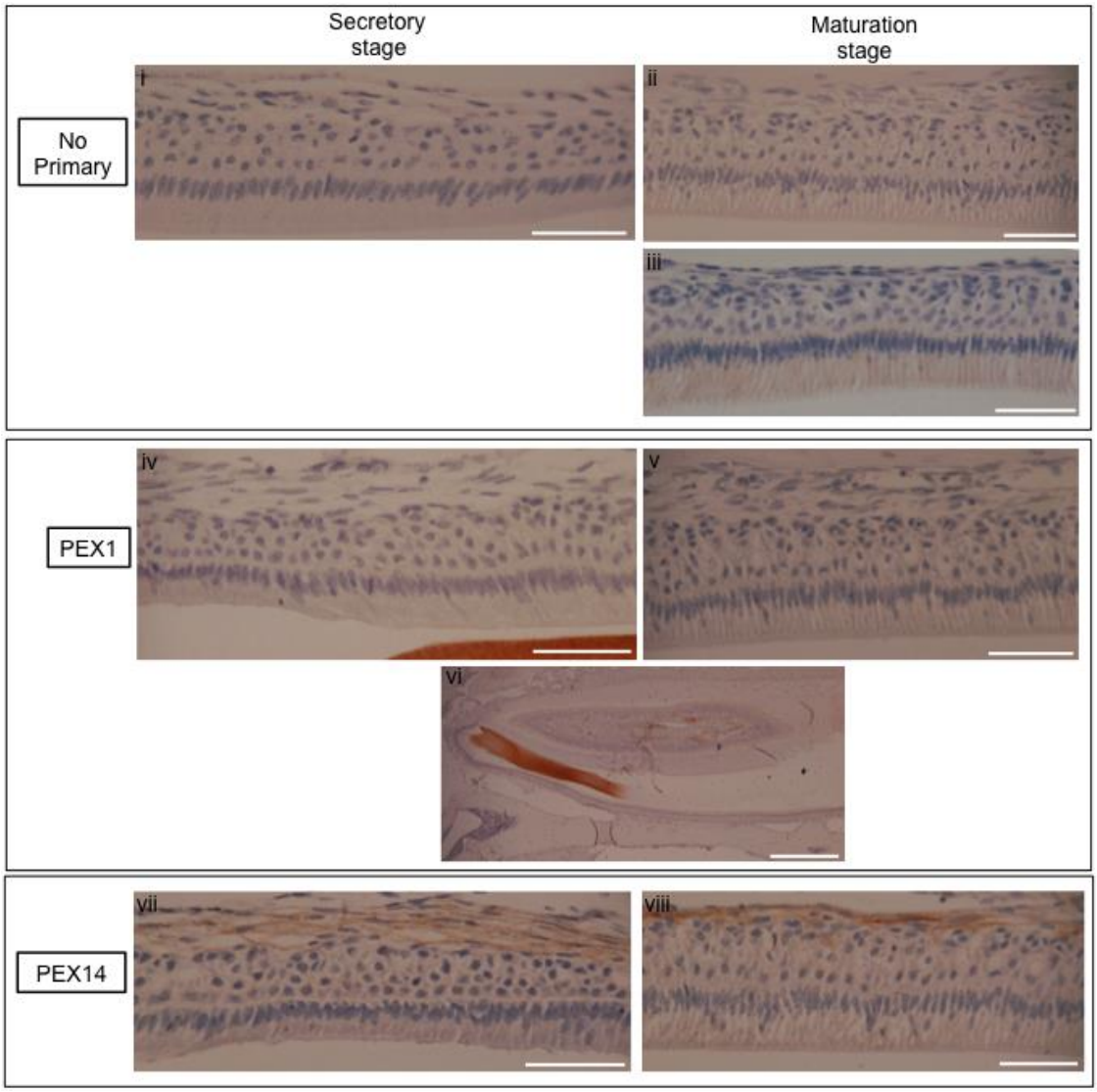


Figure 3.14: Immunohistochemical staining of murine ameloblasts for PEX1 and PEX14.
i-iii: No primary antibody control: secondary antibody (anti-rabbit reagent) only. iv-vi: anti-PEX1 antibody followed by secondary antibody. vii-viii anti-PEX14 antibody followed by secondary antibody. Secretory stage ameloblasts are shown in panels i, iv and vii. Maturation stage ameloblasts are shown in panels ii, iii (late maturation stage), v and viii. Image vi shows the staining present in the proteinaceous enamel matrix (and is shown in a different orientation to panels i-v). All scale bars represent 50μm, except in panel vi, which represents 500μm.

In summary, the use of the anti-PEX1 and anti-PEX6 antibodies resulted in staining in the no primary antibody controls for both antibodies. Therefore it was concluded that the staining was not specific for the PEX1 and PEX6 proteins using chymotrypsin based antigen retrieval and analysis was abandoned for both proteins in mouse incisor sections.

For the anti-PEX14 antibody, chymotrypsin based antigen retrieval, using a primary antibody dilution of 1:150 and the secondary antibody as instructed by the manufacturer, resulted in staining that was specific to the slides treated with the primary antibody (Figure 3.15). Immunoreactivity was present in the tissue overlying the ameloblast cell layer, the stratum intermedium, in secretory stage and within the papillary layer of the maturation stage regions of incisor sections. The staining was faint in the secretory stage region and much stronger in that of the maturation stage. For sections incubated with the secondary antibody only, immunoreactivity in this layer was absent in all stages of amelogenesis. Together, this suggests that peroxisomes are present in large quantities in the maturation stage papillary layer.

Staining of maturation stage ameloblasts was evident for sections treated with the anti-PEX14 antibody, but faint staining was also present in the no primary control sections, specifically in late maturation stage ameloblasts. This meant that it was impossible to conclude that the staining in the anti-PEX14 antibody treated sections was specific.

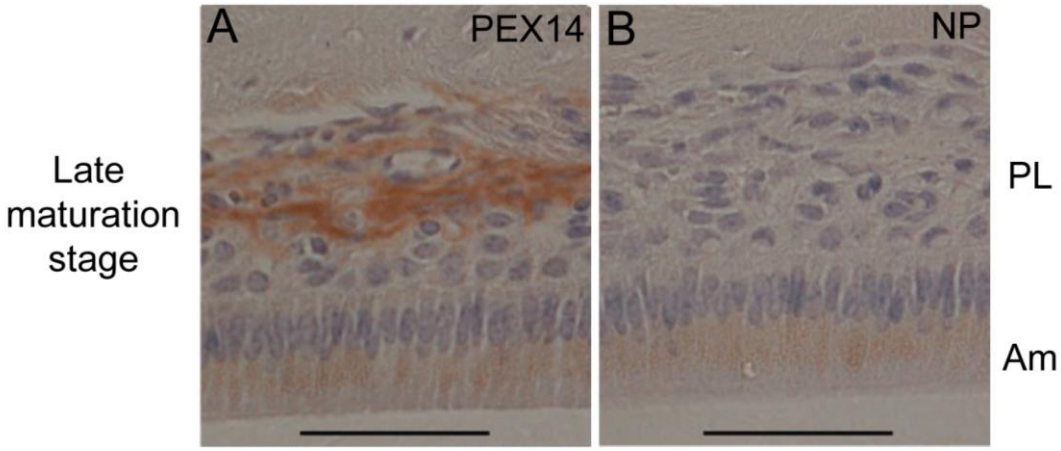


Figure 3.15: Immunohistochemical staining of late maturation ameloblasts for PEX14.
A: anti-PEX14 antibody followed by labelled polymer-HRP anti-rabbit reagent.
B: No primary antibody control: secondary antibody (labelled polymer-HRP anti-rabbit reagent) only. Scale bars represent 50 μm. Abbreviations are as follows: Am ameloblasts, NP no primary control, PL papillary layer.

3.2.11.3 Immunohistochemical staining of retinal tissues

Immunohistochemical staining of eye tissues was carried out using anti-PEX1, anti-PEX6 and anti-PEX14 antibodies. In order to optimise the conditions for IHC with each antibody, a range of antibody concentrations were tested with citrate-based antigen retrieval (Section 2.12.5). The highest antibody concentration was also tested without any form of antigen retrieval. It was found that the eye sections without any antigen retrieval did not stain. Sections processed using citrate buffer antigen retrieval did stain but those incubated without primary antibody did not, suggesting that staining was specific to each primary antibody. The final dilutions used were 1:450 anti-PEX1, 1:50 anti-PEX6 and 1:750 anti-PEX14. Higher dilutions resulted in the loss of staining. Immunohistochemical staining was carried out as previously described (Sections 2.12.5 and 3.2.11.2).

Use of anti-PEX1, anti-PEX6 and anti-PEX14 antibodies revealed staining throughout the retina with the exception of the photoreceptor outer segment in all cases (Figure 3.16). The ganglion cell layer (GCL) and the photoreceptor inner segment (IS) layer showed the most intense PEX14 immunoreactivity. For PEX1, the outer plexiform layer (OPL) stained the most intensely with strong staining also present in the GCL and the inner plexiform layer (IPL). For PEX6, the most intense staining was present in the GCL.

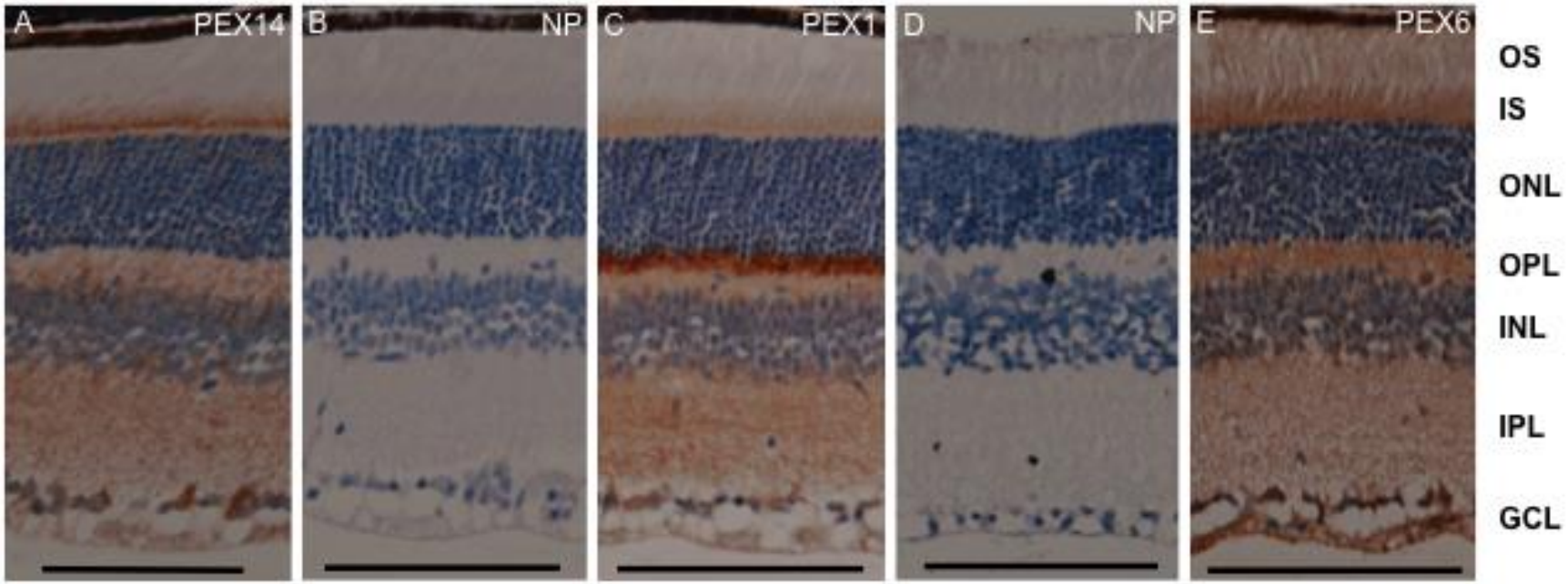


Figure 3.16: Immunohistochemical staining of murine retina.

A: anti-PEX14 antibody followed by labelled polymer-HRP anti-rabbit reagent.

B: No primary antibody control: secondary antibody (labelled polymer-HRP anti-rabbit reagent) only.

C: anti-PEX1 antibody followed by labelled polymer-HRP anti-rabbit reagent

D: No primary antibody control: secondary antibody (rabbit anti-goat antibody) only.

E: anti-PEX6 antibody followed by rabbit anti-goat antibody.

Scale bars represent 100 μm.

Abbreviations are as follows: GCL ganglion cells layer; INL inner nuclear layer; IPL inner plexiform layer; IS inner segment; NP no primary antibody control; ONL outer nuclear layer; OPL outer plexiform layer; OS outer segment.

Adapted and reproduced with permission from Smith et al. (2016b).

3.3 Discussion

In this study, ten families (Families AI-10, -11, -12, -16, -71, -138, -146, -152, -186 and -202) segregating a combination of SNHL, retinal dystrophy and enamel defects in a pattern consistent with autosomal recessive inheritance were analysed in order to identify the gene(s) and mutation(s) involved. DNA from at least one affected individual from each family underwent WES and analysis initially revealed that two of the families (AI-12 and -138) carried compound heterozygous mutations in the USH2 gene, *USH2A* (Section 3.2.3). WGS revealed that another family (AI-11) carried a heterozygous deletion that included the *MYO7A* start codon (Section 3.2.4). WES revealed that the same family also carried a heterozygous *ENAM* variant (Section 3.2.5).

Analysis of WES data from the remaining families revealed that all carried biallelic mutations in either *PEX1* or *PEX6*. Mutations in *PEX1* and *PEX6* were already known to cause the ZSSDs, whose phenotypes include craniofacial dysmorphism, neurological abnormalities, sensory defects and liver and kidney pathology (Section 3.3.1). Collaboration with another research group confirmed that *PEX1* and *PEX6* mutations had been identified in other HS patients. These results therefore showed that HS is a mild presentation of a ZSSD, with sensory defects as the major phenotype, resulting from hypomorphic mutations in *PEX1* or *PEX6*. Both PEX1 and PEX6 are well characterised, with known roles in the peroxisome matrix protein import and peroxisome biogenesis, although little was known about their distribution within the retina or during amelogenesis.

Immunohistochemical analysis of murine retina revealed that both PEX1 and PEX6 are present throughout many layers of the retina. Investigations of the localisation of PEX1 and PEX6 in murine incisors were inconclusive since the staining did not appear to be specific.

Family AI-202 (Labelled as family 4 by Ratbi et al.) was reported as one of six families detailed in the first paper to report the identification of *PEX1* and *PEX6* as the causative genes for HS (Ratbi et al., 2015). Families AI-10, -16, -71, -146, -152 and -186 were reported in a subsequent paper supporting the original

publication, as additional HS cases associated with *PEX1* and *PEX6* mutations (Smith et al., 2016b). These were labelled as families 1, 4, 6, 2, 5 and 3 respectively. This article also included details of the localisation of PEX1 and PEX6 within the murine retina.

3.3.1 Zellweger Syndrome Spectrum Disorders (ZSSDs)

The Zellweger Syndrome Spectrum Disorders (ZSSDs) are peroxisome biogenesis disorders (PBDs; MIM PS214100) and are inherited in an AR manner. PBDs are characterised by a wide range of phenotypes including craniofacial dysmorphism, neurological abnormalities, sensory defects and liver, kidney and bone abnormalities (Waterham and Ebberink, 2012). The ZSSDs include Zellweger syndrome (ZS), neonatal adrenoleukodystrophy (NALD) and infantile Refsum disease (IRD), which represent overlapping clinical phenotypes that vary in severity, with ZS being the most severe and IRD the least. Whilst ZS patients often present with serious disease at birth and live only a few weeks or months, patients with IRD and NALD generally present later in childhood, primarily with SNHL and retinal dystrophy, but also with multiple organ dysfunction and psychomotor impairments (Yik et al., 2009, Waterham and Ebberink, 2012). The reported incidence of ZSSDs in the US population is 1 in 50,000 (Gould et al., 2001). However, this may be an underestimate since the milder phenotypes overlap with and have been misdiagnosed as both USH (Raas-Rothschild et al., 2002) and Leber congenital amaurosis (LCA; MIM PS204000) (Ek et al., 1986, Michelakakis et al., 2004).

The ZSSDs are a result of impaired peroxisomal function. Peroxisomes are ubiquitous cellular organelles that perform numerous diverse vital functions, including β-oxidation of very long chain fatty acids, the synthesis of myelin precursors and detoxification of hydrogen peroxide (Wanders and Waterham, 2006). Defective peroxisomal function can result in changes in neuronal migration, proliferation, differentiation and survival (Powers and Moser, 1998, Faust et al., 2005, Krysko et al., 2007).

The genes most commonly mutated in ZSSDs are peroxisomal biogenesis factor 1 (*PEX1*; MIM *602136) and peroxisomal biogenesis factor 6 (*PEX6*; MIM *601498) (Ebberink et al., 2011). The phenotype variation seen in ZSSDs is related to the severity of the underlying *PEX* gene mutations. Biallelic loss of function or null alleles, caused by frameshift and nonsense mutations, often lead to a total absence of peroxisomes, resulting in ZS, whilst genotypes which include an allele with a minor import defect, caused by a missense mutation, may lead to NALD or IRD (Moser, 1999, Maxwell et al., 2002, Preuss et al., 2002). As a result of this study and data from other centres, the ZSSD phenotypic spectrum was expanded to include HS (Ratbi et al., 2015).

3.3.2 Heimler syndrome is the mildest presentation of a ZSSD

In this study, WES identified biallelic mutations in *PEX1* and *PEX6* in HS patients. At the beginning of this study, mutations in *PEX1* and *PEX6* were already known to cause the ZSSDs (Section 3.3.1) that encompass a complex phenotype that includes sensory defects. In rare cases, more mildly affected individuals have been reported to have enamel defects (Bader et al., 2000, Tran et al., 2011, Acharya et al., 2012), and it has been suggested that this may be a common but underreported aspect of the phenotype (Poll-The et al., 2004, Tran et al., 2011, Steinberg et al., 2012).

Mutations in *PEX1* and *PEX6* are the most common causes of ZSSD and result in impaired peroxisomal function (Ebberink et al., 2011). Peroxisomes are ubiquitous cellular organelles that consist of a single lipid bilayer membrane surrounding a dense matrix of proteins (Smith and Aitchison, 2013). Cellular peroxisomal content is controlled both by their production, mainly by fission, and their degradation, by pexophagy (Hutchins et al., 1999, Motley and Hettema, 2007). Before fission can occur, new matrix proteins, synthesised in the cytosol must be imported into the existing peroxisome. This is achieved via peroxisomal import machinery that specifically imports proteins containing a peroxisomal targeting signal.

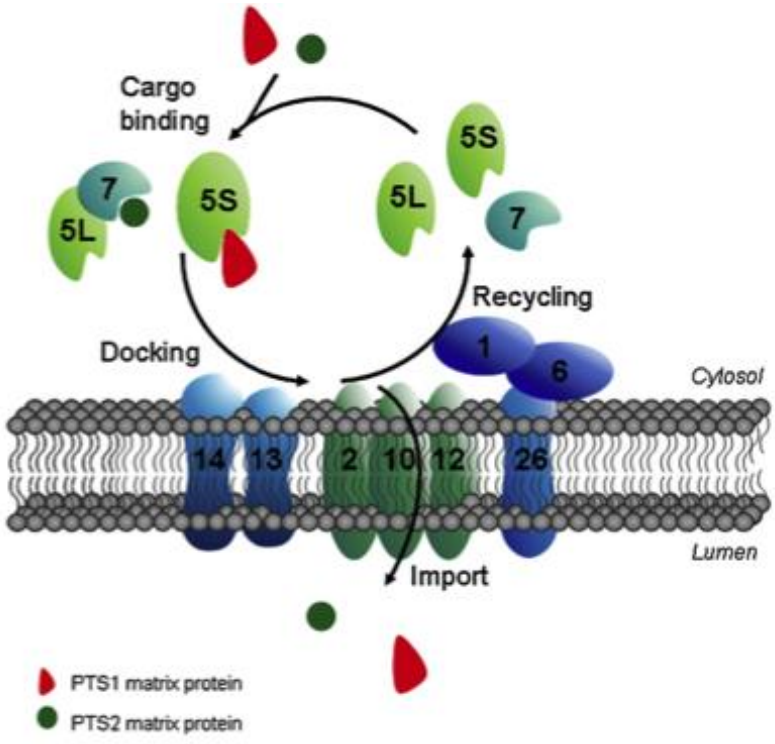


Figure 3.17: Schematic diagram of peroxisomal matrix protein import.
Proteins with either of the peroxisomal targeting sequences become bound to cytosolic cargo binding proteins PEX5 (5L and 5S represent the Long and Short protein isoforms respectively) and/or PEX7. The bound proteins then dock at the peroxisomal membrane complex consisting of PEX13 and PEX14. The bound matrix protein is then translocated across the membrane to the lumen and the receptor proteins are targetted for recycling via ubiquitination by PEX2, PEX10 and PEX12. PEX1, PEX6 and PEX26 then release the ubiquitinated cargo binding proteins PEX5 and/or PEX7 back to the cytosol where they are deubiquitinated, ready for additional rounds of matrix protein import. Adapted with permission from Waterham and Ebberink (2012).

PEX1 and PEX6 are type 2 AAA+ (ATPases Associated with diverse cellular Activities) ATPases. ATPases use energy from ATP hydrolysis to effect conformational changes on targets (Hanson and Whiteheart, 2005). PEX1 and PEX6 form a heterohexameric protein complex *in vivo* (Gardner et al., 2015). The complex is part of the peroxisomal import machinery that recycles the peroxisome-targeting signal receptor protein, PEX5, back to the cytosol after release of its protein cargo within the peroxisomal lumen (Figure 3.17) (Dodt and Gould, 1996). PEX1 and PEX6 provide the energy required to remove PEX5 from the luminal membrane (Platta et al., 2005) and their function is

dependent upon interaction with the membrane protein PEX26 (Fujiki et al., 2008). Mutations in the genes encoding these proteins therefore give rise to a peroxisomal protein import defect, which in turn leads to a deficiency of peroxisomal biogenesis. Evidence for this includes reports that PEX1 or PEX6 deficient cells contain so-called “peroxisomal ghosts”, enlarged structures that fail to import peroxisomal matrix proteins that are not present in wild-type cells (Santos et al., 1988, Wendland and Subramani, 1993). There are also fewer of these structures in affected cells than there are true peroxisomes in wild type cells. In yeast cells deficient in Pex1 or Pex6, an increase in pexophagy was observed compared to wild type cells (Nuttall et al., 2014).

Peroxisomes perform numerous diverse vital functions, including β-oxidation of very long chain fatty acids, the synthesis of myelin precursors and detoxification of hydrogen peroxide (Wanders and Waterham, 2006). Therefore defects in peroxisomal biogenesis affect many processes and the severity of the defect can alter the phenotype.

The phenotypic variation seen in ZSSD is related to the severity of the underlying *PEX* gene mutations. Biallelic loss of function or null alleles, caused by frameshift and nonsense mutations often lead to a total absence of peroxisomes, resulting in ZS, whilst genotypes which include an allele with a minor import defect, caused by a missense mutation, may lead to NALD or IRD (Moser, 1999, Maxwell et al., 2002, Preuss et al., 2002). During this study, HS was also shown to result from hypomorphic mutations in *PEX* genes (Ratbi et al., 2015). Functional complementation studies of *PEX1* and *PEX6* variants in peroxisome-deficient cells, carried out by the collaborating group, revealed that at least one of the HS alleles in each patient retained significant activity, including the c.1742G>C, p.R581P variant identified in AI-202 (Ratbi et al., 2015). Ratbi et al. (2015) reported a total of six families; four with biallelic *PEX1* mutations (including family AI-202, individual 4505, reported here) and two with biallelic *PEX6* mutations. Similarly, Zaki et al. (2016) reported one family with biallelic *PEX6* mutations. This study reports an additional family with biallelic *PEX1* mutations and five families with biallelic *PEX6* mutations (Smith et al.,

2016b), bringing the total number of HS families with mutations in either *PEX1* or *PEX6* to thirteen.

The report by Ratbi et al. (2015) included two families in which *PEX1* or *PEX6* variants were not identified. These cases may carry variants that are difficult to detect with the variant calling pipeline used by the authors. Alternatively the variants may lie in a promoter or in deep intronic regions not covered by WES. Further investigation using CNV calling algorithms and WGS may yet identify mutations within and/or affecting expression of *PEX1* or *PEX6.* Alternatively, these families may carry mutations in a different gene(s).

The report by Zaki et al. (2016), published after the initial report by Ratbi et al. (2015) and after the work described here was carried out, identified two siblings, born of consanguineous parents, with SNHL, RP, AI, developmental delay, liver dysfunction, cerebral white matter disease and microcephaly, with a homozygous *PEX6* mutation, c.1238G>T, p.G413V. The phenotype reported is distinct to that reported in other ZSSD and HS cases, since it includes microcephaly. Consequently, the authors do not assign HS as the diagnosis in this case, despite the clinical overlap. The microcephaly phenotype may represent further phenotypic variation on the ZSSD spectrum, since mutations in other genes have been known to cause both peroxisomal defects and microcephaly (Waterham et al., 2007). Alternatively, it may be due to a co-segregating variant at another locus, given the family's consanguineous history.

Immunohistochemical staining of murine incisor and retina was undertaken to determine whether PEX1 and PEX6 are normally expressed in these tissues. Whilst staining of murine incisor failed for both PEX1 and PEX6 due to detection of similar patterns of immunoreactivity for the control slide incubated without a primary antibody, PEX14 staining showed that there was strong expression in the stratum intermedium of maturation stage ameloblasts. Due to difficulties orientating the tissue during embedding and sectioning, the normal morphology of ameloblasts in the secretory or maturation stage was masked. Therefore the secretion-transition-maturation stage boundary was identified as the ameloblasts immediately below the more distal root (of the two visible on a

mid-line section) of the first molar (Smith and Nanci, 1989). Ameloblasts that lay more mesially were designated as secretory stage, while those that lay more distally, were designated as maturation stage. However, subsequent analysis of wild-type murine incisors by microCT found that mineralisation begins at a more mesial position than this (data not shown). It was therefore concluded that the anti-PEX1 and anti-PEX6 antibodies used could not be interpreted rigorously as they were not specific to the target proteins, while the data obtained with PEX14 could not be satisfactorily interpreted as the stages of amelogenesis could not be distinguished and the early secretory stage may not have been stained at all due to the orientation of the sections and the possibility that this region was excluded from the section.

Immunofluorescence in post-natal day 2 dental tissues by Zaki et al. (2016) detected PEX6 at the apical ends of both ameloblasts and odontoblasts, suggesting that it may be involved in protection of these cells from reactive oxygen species generated in mineralising tissues. These findings agree with those previously reported for PEX14 in developing murine molar teeth (Stelzig et al., 2013). In addition to this protective role, lipid metabolism by peroxisomes may influence the secretion of enamel matrix proteins by affecting membrane fluidity and may also affect signalling molecules involved in ameloblast differentiation. Tooth development in general may be affected since other oral phenotypes, such as delayed tooth eruption and microdontia, are associated with ZSSD (Poll-The et al., 2004, Lertsirivorakul et al., 2014). Lipids have been detected in both developing and mature enamel and dentine, although their source remains unknown (Goldberg and Septier, 2002). In enamel, the source has been suggested to be remnants of Tomes' processes (Goldberg and Septier, 2002).

Immunohistochemical staining of murine retinal tissue in this study revealed staining throughout the retina with the exception of the photoreceptor outer segment in all cases (Figure 3.16). The GCL and the IS showed the most intense PEX14 immunoreactivity. For PEX1, the OPL stained the most intensely, with strong staining also present in the GCL and IPL. For PEX6, the most intense staining was present in the GCL. Zaki et al. (2016) used

immunofluorescence to show that PEX6 localised throughout the retina, including the IPL and OPL but with the most intense staining at the IS and the connecting cilium. Zaki et al. (2016) suggested that this might reflect a role for PEX6 in outer segment maintenance and function. The results reported in this study overlap those of Zaki et al. (2016) but are not identical. The differences in the staining pattern between this and the previous study could be due to the use of different detection methods, antibodies or to differences in the ages of the mice studied. Whilst the staining of PEX1, PEX6 and PEX14 overlapped, the relative intensities across retinal layers differed. This is surprising since PEX1 and PEX6 are known to form a complex (Gardner et al., 2015) and, along with PEX14, all three proteins form part of the peroxisomal protein import machinery. A previous report of the immunolocalisation of PEX13 and PEX14, two peroxisomal membrane proteins that interact as part of a membrane complex (Schell-Steven et al., 2005), showed that the proteins had distinct localisations within mouse testis (Grant et al., 2013). PEX1 and PEX6 are known to require ATP to interact (Saffian et al., 2012), therefore, the proteins may also be present separately and this may explain their overlapping localisations with different relative intensities.

In this study, the staining for all three PEX proteins suggests that peroxisomes are particularly abundant within the OPL and GCL. Such layers contain features likely to provoke metabolic stress, such as synapses, and therefore may require higher numbers of peroxisomes in order to provide efficient means of detoxification for cell survival.

In conclusion, biallelic *PEX1* or *PEX6* variants have been identified in individuals with HS from a total of thirteen families, seven of them in this study (Ratbi et al., 2015, Smith et al., 2016b, Zaki et al., 2016). At least one allele in each family is likely to result in protein with some residual function. Therefore the aetiological basis of the HS phenotype is believed to be a mild peroxisomal protein import defect (Ratbi et al., 2015). A reduction in the number or efficient function of peroxisomes may compromise the function or survival of the ameloblasts, ganglion cells and the cells of the OPL.

3.3.3 Likely functional consequences of biallelic *PEX1* and *PEX6* variants in individuals with HS

The variants identified in *PEX1* and *PEX6* in the HS individuals in this study included those predicted to lead to missense and frameshift proteins changes. In all cases, there were combinations of either two hypomorphic alleles in *trans* or one hypomorphic allele and one null allele. Five of the variants identified had been previously reported in ZSSD patients and six were unreported. In two of the families, AI-71 (2645) and AI-202 (4505), variants in *PEX1* were identified. In the remaining five families, AI-10 (1499 and 1500), AI-16 (2113), AI-146 (3000), AI-152 (3466 and 4922) and AI-186 (4716), variants in *PEX6* were identified. All changes were shown to be biallelic except in families AI-71 (2645) and 146 (3000), where segregation of alleles was not possible due to lack of samples from additional family members. All variants discussed in the following section are detailed in Table 3.7. Their frequencies in databases of variation are reported in Table 3.8. Sanger sequencing electropherograms are shown in Figure 3.7 and segregation analysis is depicted in Appendix 7. Alignments of homologous protein sequences surrounding each previously unreported protein variant are depicted in Figure 3.8. The relative positions of important domains and structural features of PEX1 and PEX6 and of the variants identified in HS patients are shown in Figure 3.18.

As AAA+ ATPases, PEX1 and PEX6 contain conserved domains important to their function. Both proteins contain tandem arrangements of AAA+ domains, labelled Domain 1 (D1) and Domain 2 (D2). D1 and D2 each contain phosphate-binding Walker A and Walker B motifs (Walker et al., 1982, Hanson and Whiteheart, 2005). The D2 domains are highly conserved and the Walker A and B motifs are important for ATP binding and hydrolysis respectively (Ciniawsky et al., 2015). The D1 domains are less conserved and are thought not to be essential to ATP hydrolysis, but may be involved in PEX1/PEX6 complex formation (Tamura et al., 2006, Nashiro et al., 2011). It is noteworthy that none of the hypomorphic mutations identified so far in HS patients affect residues within the important D2 Walker motifs (Figure 3.18). This supports the hypothesis that only relatively mild mutations in *PEX1* and *PEX6* result in HS.

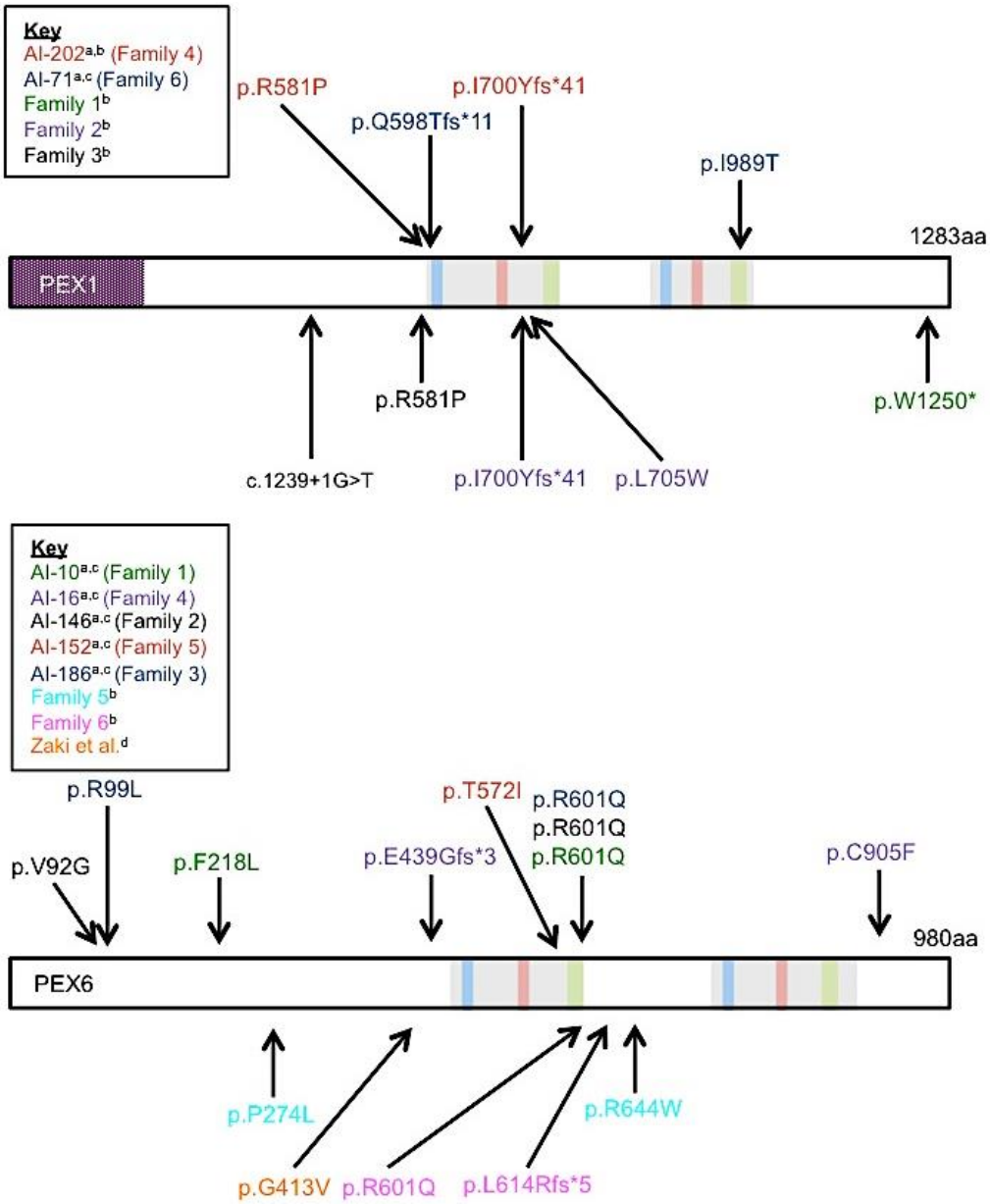


Figure 3.18: Schematic of PEX1 and PEX6 proteins.

PEX1 N terminal domain is shown by purple shading. For both PEX1 and PEX6, the AAA+ domains, D1 and D2, are shown in grey, conserved Walker A1 and A2 motifs are shown sequentially in blue, conserved Walker B1 and B2 motifs are shown in red and the second regions of homology (SRH) are shown in green. The position of the N terminal domain of PEX1 is as reported by Shiozawa et al. (2006). The AAA+ domains were determined using SMART (Schultz et al., 1998, Letunic et al., 2015). The positions of Walker motifs and SRH are as reported by Ciniawsky et al. (2015) and Ghenea (2001) respectively for both PEX1 and PEX6. Annotations show the positions of the residues predicted to be affected by the *PEX1* and *PEX6* variants identified in HS patients. Variants with labels above each schematic protein are those identified in this study. Variants with labels below each schematic protein were identified by Ratbi et al. (2015) or Zaki et al. (2016). Different colours denote the combinations of mutations identified in patients from different families.

[a] Families reported in this study

[b] Families reported by Ratbi et al. (2015).

[c] Families reported by Smith et al. (2016b).

[d] Family reported by Zaki et al. (2016).

The family numbers refer to those used in the relevant publications.

PEX1 variants c.2097dupT (p.I700Yfs*42) and c.1742G>C (p.R581P) were identified in *trans* in individual 4505 (family AI-202). The c.2097dupT variant has been reported in over 80 individuals in the LOVD dbPEX database and represents the second most commonly reported variant in ZSSD patients (Ebberink et al., 2011). Analysis of ZSSD patient cohorts has shown that of the approximate 58% with *PEX1* mutations (Ebberink et al., 2011), around 30-32% of these carry the c.2097dupT variant (Collins and Gould, 1999, Maxwell et al., 1999, Walter et al., 2001). c.2097dupT (p.I700Yfs*42) is associated with the more serious ZS phenotype when homozygous or in combination with other null alleles (Walter et al., 2001). The frameshift variant would be predicted to result in degradation of transcripts via NMD and analysis of RNA from a fibroblast cell line prepared from a patient homozygous for c.2097dupT showed the presence of around 8% of WT levels of *PEX1* mRNA (Maxwell et al., 2002). An assay to determine the extent to which transfection of a *PEX1* c.2097dupT construct could allow *PEX1* deficient fibroblasts without peroxisomes to produce peroxisome positive cells, showed that no peroxisomes formed following transfection (Ratbi et al., 2015). This suggests that the c.2097dupT mutation leads to a complete absence of peroxisomes *in vivo*.

The c.1742G>C (p.R581P) variant had not been previously reported in ZSSD patients. The mutation is predicted to alter a residue just proximal to the D1 AAA+ domain of *PEX1* at residues 591-741 (determined by Simple Modular Architecture Research Tool (SMART), available at http://smart.embl.de/ (Schultz et al., 1998)). Both mutations identified in individual 4505 (family AI-202) were reported separately in combination with other *PEX1* variants in two other HS families: c.[2097dupT];[2114T>G] in one of the index cases from the original HS family and as c.[1742G>C];[1239+1G>T] in an additional family (Ratbi et al., 2015).

The *PEX1* mutations identified in individual 2645 (family AI-71) included a previously unreported deletion, c.1792delA (p.Q598Tfs*11) and an SNV, c.2966T>C (p.I989T). The lack of additional family members for AI-71 has meant that it is not possible to determine whether these variants are in *cis* or in *trans*. If the mutations exist in *trans*, the combination of a frameshift allele, predicted to result in NMD of the mutant transcript, with a missense mutation mirrors the combination of mutations identified in 4505 (family AI-202). The c.2966T>C variant had been previously reported (Maxwell et al., 2005) and was predicted to alter a residue within the D2 Second Region of Homology (SRH; residues 983-1000 (Ghenea, 2001)). The SRH is reported to be a highly conserved motif within D1 and D2 (Lupas and Martin, 2002). The D1 SRH contains a conserved sensor motif (residues 384 and 385 (Ogura et al., 2004)) in which mutations in another AAA+ family protein have been shown to abolish ATPase activity and nucleotide binding (Zhao et al., 2010). The D2 SRH also contains two arginine residues (at positions 995 and 998) that are known to contribute to ATPase activity and stability of other functional AAA+ family oligomers (Ogura et al., 2004). Therefore the region surrounding the I989 residue appears to be functionally important. Additionally, I989 is conserved as a residue designated “hydrophobic tendency” by ClustalW (Thompson et al., 1994) in both PEX1 and PEX6 and their homologous and orthologous proteins. Its replacement with a hydrophilic threonine residue may alter the structure, and therefore also the function, of the region.

For the remaining HS families, biallelic *PEX6* mutations were identified. For each of families AI-10 (1499 and 1500), AI-146 (3000) and AI-186 (4716), a heterozygous c.1802G>A (p.R601Q) variant was detected. This missense variant has been previously reported in seven ZSSD patients although no phenotype details were provided (Yik et al., 2009, Ebberink et al., 2010). It was identified in *trans* with missense variants in six patients and with a seven base pair deletion; c.1314_1321delGGAGGCCT (a variant also identified in AI-16) in one patient (Yik et al., 2009, Ebberink et al., 2010). The variant is also reported in two entries in dbPEX (Accessed 19th February 2016) in *trans* with other missense mutations, one with c.2579G>A reported to have IRD and one with c.2434C>T reported to have NALD. In addition to these families and the three

HS families reported here, the c.1802G>A (p.R601Q) variant has also been identified in one other UK family with HS, suggesting that it is associated with milder presentations of ZSSD (Ong et al., 2006, Ratbi et al., 2015). Therefore, from the thirteen HS families reported to have mutations in *PEX1* or *PEX6*, four have the c.1802G>A variant. An assay to determine the extent to which transfection of a *PEX6* c.1802G>A (p.R601Q) construct could allow *PEX6* deficient fibroblasts without peroxisomes to produce peroxisome positive cells showed that after transfection, peroxisomes were present in around 35% of cells (Ratbi et al., 2015), suggesting that the c.1802G>A (p.R601Q) variant results in protein with a reduced, but not abrogated, function.

The c.1802G>A (p.R601Q) variant has been reported at an allele frequency of 0.4% in European (Non Finnish) populations (ExAC v0.3). In comparison to the most commonly identified ZSSD-causing *PEX1* mutation, rs61750420, c.2528G>A (p.G843D), that had been identified in 25-40% of ZSSD patients (Reuber et al., 1997, Imamura et al., 1998a, Maxwell et al., 1999, Walter et al., 2001, Preuss et al., 2002, Steinberg et al., 2004), the *PEX6* c.1802G>A variant is over nine times more frequent in the European population according to ExAC. Despite four reports in ExAC of homozygous individuals for the c.1802G>A (p.R601Q) variant, it has not been reported in a homozygous state in ZSSD patients. This contrasts with the *PEX1* c.2528G>A (p.G843D) variant, for which reports of homozygotes in cohorts of ZSSD patients are common (Collins and Gould, 1999, Walter et al., 2001, Maxwell et al., 2002, Preuss et al., 2002, Steinberg et al., 2004). This suggests that either the c.1802G>A allele is only able to cause disease when combined with other particular pathogenic mutations in *trans* or that disease due to c.1802G>A (p.R601Q) may be underreported. This, perhaps, is due to a failure to recognise the phenotype as part of the ZSSDs or a failure to identify the causative variant via targetted Sanger sequencing in clinic and research studies. Since the frequency of USH in the US population has been estimated to be 1 in 6,000 (Kimberling et al., 2010), it seems more likely that homozygous c.1802G>A is not sufficient to cause disease.

Analysis of the haplotype surrounding the c.1802G>A variant showed that it was associated with a haplotype of two SNPs and one microsatellite spanning a region of at least 779kb for families AI-10, AI-146 and AI-186 (Table 3.9). However, for two of the families, AI-10 and AI-186, the shared haplotype extended over a region of at least 1,279kb. These findings suggest that the variant may represent a founder mutation, though the relatively short distance over which the conserved haplotype extends makes this difficult to prove unequivocally, and may imply that it arose many generations ago.

The c.1802G>A (p.R601Q) variant in *PEX6* is predicted to alter a conserved arginine residue, that is within the consensus sequence of the SRH of D1, to a glutamine residue (Ghenea, 2001). While the first AAA+ domain is less conserved than the second (Shiozawa et al., 2004), it has been shown to be critical for interaction with PEX1 (Tamura et al., 2006). Within this domain, the R601 residue forms an arginine finger that is known to be conserved in the majority of AAA+ family members and is known to interact with ATP during PEX6 function (Ogura et al., 2004). Therefore the replacement of the positively charged R601 residue with an uncharged glutamine residue is likely to affect the function of PEX6.

The combinations of *PEX6* alleles identified in AI-10, AI-146 and AI-186 are unique to each family and the mutations in *trans* to the c.1802G>A allele have not been reported previously in the literature. This again suggests that particular combinations of hypomorphic variants cause HS as opposed to the other ZSSDs. AI-10 (1492, 1499, 1500) carries c.654C>G, p.F218L, AI-146 (3000) carries c.275T>G, p.V92G and AI-186 (4716) carries c.296G>T, p.R99L. The variants have not been reported in databases of human variation, suggesting that they are rare. All three mutations affect residues encoded by the first exon, which one study found to harbour around one third of all ZSSD-causing *PEX6* mutations (Ebberink et al., 2010). Conservation analysis of p.F218 showed that the residue is conserved in all mammalian species analysed and in zebrafish but not in thale cress or in fruit fly. Conservation analysis for p.V92 and p.R99 showed that whilst the residues are conserved in the majority of the mammalian species analysed, the region containing both residues is entirely absent for cat,

sperm whale and horse as well as thale cress sequences. The region is present in fruit fly and zebrafish sequences but the p.V92 and p.R99 residues are not conserved in these sequences and the mutant G residue at position 92 is present in both species. The fruit fly, zebrafish, cat, sperm whale, horse and thale cress sequence results in isolation suggest that the region may not be functionally important and that the mutation might be tolerated. However, the conservation of the residue in other species suggests that the region may have evolved to develop a novel function. Pathogenicity prediction algorithms generally predict the variants to be disease causing, however for all three of the variants, Provean assigns a neutral score. This may reflect the hypomorphic nature of the mutations.

The c.275T>G (p.V92G), c.296G>T (p.R99L) and c.654C>G (p.F218L) variants are predicted to affect residues in the N terminal portion of PEX6 and lie outside of the conserved D1 AAA+ domain (residues 462-597, as determined by SMART (Schultz et al., 1998)). The role of the N terminal domain in PEX6 is not well studied. In PEX1, the domain is known to bind phospholipids and is predicted to direct interaction with the peroxisomal membrane (Shiozawa et al., 2006). However, the N terminal domain of PEX6 has been reported to be very different from that of PEX1 when compared to the similarity displayed by the conserved AAA+ domains of these two related proteins (Shiozawa et al., 2004). In yeast, the N terminal portion of PEX6 is known to interact with PEX26 (Birschmann et al., 2003) and the difference between the N terminal domains of PEX1 and PEX6 have been hypothesised to delineate their distinct functions from each other (Shiozawa et al., 2004). A study that deleted the N terminus of PEX6 (residues 1-384) showed that while binding with PEX1 was detected via a yeast-two-hybrid assay, the construct was unable to complement PEX6 null cells with impaired import of peroxisomal proteins (Tamura et al., 2006). This suggests that the N terminal domain is essential for the proper function of PEX6.

In 2113 (AI-16), a combination of an eight bp deletion, c.1314_1321delGGAGGCCT (p.E439Gfs*3) and a missense mutation c.2714G>T (p.C905F) were identified. The deletion has been previously

reported in a number of ZSSD patients (Krause et al., 2009, Ebberink et al., 2010) and is described as the most commonly reported ZSSD *PEX6* mutation (Waterham and Ebberink, 2012). It is unclear from the Krause et al. (2009) study in what combination of alleles the deletion was detected or the severity of the phenotype. However, the report by Ebberink et al. (2010) that studied 75 ZSSD patients in the PEX6 complementation group, detected the variant in a homozygous state in five patients and in a heterozygous state in eight patients in *trans* with both null and missense alleles. This suggests that the allele may be relatively common among ZSSD patients and may be a mutational hotspot or founder mutation. The deletion would be expected to lead to NMD (Isken and Maquat, 2007).

The c.2714G>T (p.C905F) missense mutation has not been previously reported and is predicted to affect a residue distal to the 2nd AAA+ domain. Little is known about this C-terminal region, but when residues 883-980 of PEX6 were deleted, the resulting constructs were unable to complement PEX6 null cells with impaired import of peroxisomal proteins despite detection of the mutant protein binding with PEX1 by yeast-two-hybrid assay (Tamura et al., 2006). This suggested that the C-terminal domain is necessary for the normal function of PEX6. The c.2714G>T variant identified in 2113 (AI-16) was consistently predicted by pathogenicity prediction programs to be disease causing and CADD assigned a scaled C score of 34. This means that only 0.04% of all variants scored by CADD v1.3 are ranked as more deleterious.

In 3466 and 4922 (AI-152), a homozygous c.1715C>T (p.T572I) variant was identified. This variant has been previously reported in three families (Raas-Rothschild et al., 2002, Ebberink et al., 2010). In one report, a patient with the same homozygous genotype and same Yemenite Jewish ethnic origin as family AI-152 had initially been diagnosed with USH (Raas-Rothschild et al., 2002). His phenotype was described as “even milder than IRD” (Raas-Rothschild et al., 2002). Biochemical analysis of the patient, undertaken only after his child exhibited a phenotype intermediate between NALD and ZS, revealed mild peroxisomal biochemical dysfunction. Analysis of the patient’s fibroblasts revealed biochemical features associated with milder ZSSD phenotypes such

as peroxisomal mosaicism (Giros et al., 1996) and temperature sensitivity (Imamura et al., 1998b). Culture of the cells at a lower temperature showed that this increased the number of peroxisome positive cells from 7% to 82%. This has also been shown for a selection of other PEX mutants, including PEX1 p.G843D (Walter et al., 2001) and PEX6 p.L57P and p.I845T (Imamura et al., 2000). A temperature sensitive variation in the number of peroxisome positive cells suggests that these mutations cause instability of or misfolding of the protein. Such mutations have been shown to be amenable to therapy by use of non-specific chemical chaperones such as arginine, proline, DMSO, 4-phenylbutyrate, glycerol, betaine and trimethyline N-oxide (Wei et al., 2000, Zhang et al., 2010, Berendse et al., 2013), with a small molecule screen to identify more specific chemical chaperones already undertaken (Zhang et al., 2010).

The utility of therapy for HS patients will be determined by whether the syndrome is a progressive or stable phenotype. The majority of patients that have been identified so far have been assessed at ages ranging from early childhood to young adulthood, with only one individual identified so far over 31. It has already been reported that one of the index cases (Heimler et al., 1991) developed macular dystrophy at the age of 29 (Lima et al., 2011) and family 1 reported by Ratbi et al. (2015) were assessed at 12-16 to not have retinal pigmentation or macular dystrophy. Early recognition of the phenotype and molecular diagnosis would enable treatment to begin at a time when the majority of the permanent dentition would be mineralising and before the retinal dystrophy had manifested, so that it may be able to ameliorate the retinal and AI phenotypes to some degree.

Additionally, some HS patients have been shown to have intellectual disability (2645 (AI-71) and 3466 and 4922 (AI-152)) and early use of chaperone therapies may improve this outcome. It is possible that once parents are identified as carriers of missense mutations in *PEX1* or *PEX6*, drugs could be given *in utero* to prevent or ameliorate the sensory, intellectual and enamel phenotypes.

3.3.4 WES filtering strategy for variant identification

The original strategy used to filter WES variants led to the identification of biallelic *PEX1* mutations in individual 4505 (family AI-202), but failed to identify mutations in *PEX1* and *PEX6* for the remaining HS cases. This was for a number of reasons, that when combined, meant that the remaining mutations were not originally identified in any of the other six HS families.

Firstly, the original filtering strategy filtered all variants present in dbSNP129 unless they were marked as pathogenic (hereafter named as dbSNP129/pathogenic filter). Pathogenic variants were indicated only by values of 5 in the SCS flag in dbSNP and the CLNSIG flag in ClinVar. This annotation relies on submission of the variant to ClinVar by clinicians / scientists and evaluation of the evidence for pathogenicity by ClinVar. Such strict criteria meant that variants marked only as “clinically associated” in dbSNP (for example due to inclusion in locus specific databases such as dbPEX) were still filtered. In addition, all variants present in subsequent dbSNP builds, up to and including dbSNP142 at a minor allele frequency of 1% or more were filtered. dbSNP129, containing 11 million SNPs and 3 million indels (Abecasis et al., 2010), is widely regarded as the final dbSNP build to be free of unannotated pathogenic mutations. Subsequent dbSNP builds included significant deposits from large scale sequencing projects, for example new deposits to dbSNP130 included over 21 million SNPs from the 1000 Genomes Project, James Watson Genome, J. Craig Venter Genome, the Individual Chinese Genome and the Individual Korean Genome (National Centre Biotechnology Information, 2009). The removal of all SNPs identified in dbSNP129 but not marked as pathogenic has previously been used successfully as a filtering strategy in sequencing projects to identify causative variants in genetic diseases, including AI (Parry et al., 2012, Parry et al., 2013, Poulter et al., 2014a). Since then, more detailed annotation, including the addition of minor allele frequency (MAF) data to a greater number of entries in subsequent dbSNP builds, has meant that the use of dbSNP129 to filter WES data has become obsolete. The use of this filter resulted in the removal of one variant in *PEX1* or *PEX6* for each of individuals

1499 and 1500 (AI-10), 2645 (AI-71), 3000 (AI-146), 3466 and 4922 (AI-152) and 4716 (AI-186) (Table 3.10).

However, even if WES data had not been filtered using the dbSNP129/pathogenic filter, the majority of variants would still have been absent in the filtered file due to the use of the annotateSnps perl script (Section 3.2.2). The annotateSnps perl script filtered variants based on options specified by the user, for example using -b to specify a dbSNP build version (e.g. dbSNP129) to filter variants from in their entirely, -d to provide a dbSNP reference file with SNP and MAF data, -p to keep variants flagged as pathogenic or -f to specify a minor allele frequency cut off at and above which, variants are filtered. In-built to version 0.1.10, but unmentioned in the user manual, was a filter based on a variant's frequency, not only in dbSNP, but also in the data submitted from the major populations of the 1000 Genomes Project. dbSNP variants are considered "common" and are therefore, filtered out, if they are present in at least one 1000 Genomes Project major population with a minor allele frequency of 1% or higher and for which two or more founders contribute to that minor allele frequency (National Centre Biotechnology Information, 2016). A major population is defined as one of the twenty-six ethnic groups that were sequenced as part of the project. Such filtering may be relevant in some situations, for example, if investigating a rare disease in one of the particular populations studied in the 1000 Genomes Project. However, in this study, the variants that were filtered were present only in one or two populations from which the individuals recruited to this study did not originate. In isolation, this filter would have removed one variant from each of 1499 and 1500 (AI-10), 2645 (AI-71), 3000 (AI-146) and 4716 (AI-186) (Table 3.10). A newer version of annotateSnps (v0.2.0) includes an option to disable the use of 1000 Genomes Project data to filter SNPs from WES data, in part, due to feedback from users.

Finally, poor coverage of particular regions of the *PEX6* gene meant that a variant in each of 2113 (AI-16) and 4716 (AI-186) was not identified (Table 3.10). The regions that were poorly covered included exons 1 and 8 for both families. These exons are the most GC rich of the gene (70.4% (protein coding

region only) and 65.3% respectively). GC content has been shown to affect the library amplification step of WES, but also the other steps that include denaturation, such as cluster amplification (Aird et al., 2011). Regions of higher GC content often have poorer coverage. Typically, the first exon of a gene is affected, due to the presence of CpG islands in the proximity of the promoter region (Majewski and Ott, 2002). The poor coverage over these exons may also be due to sub-optimal bait design or capture conditions. Therefore, the poor coverage of these two exons may be typical of many other genes and the use of coverage statistics that report the mean coverage and even the percentage of bases covered by a specified number of reads may be misleading. A score of the evenness of coverage, such as that calculated by Mokry et al. (2010), may be more informative.

The use of the dbSNP129/pathogenic filter, annotateSnps 1000 Genomes Project filter and the low coverage of particular exons meant that only one variant in *PEX1* or *PEX6* remained in each of 1499 and 1500 (AI-10), 2113 (AI-16), 2645 (AI-71), 3000 (AI-146) filtered WES data and none at all for 3466 and 4922 (AI-152) and 4716 (AI-186). The use of a perl script to filter the remaining variants for those that were potentially biallelic resulted in the exclusion of the remaining *PEX1* and *PEX6* variants from the final list of filtered variants (Table 3.10).

Although the use of this variant filtering strategy delayed the identification of the causative genes for HS, a stringent filtering strategy was deemed necessary at the time since the number of variants reported from WES studies of multiple unrelated individuals is enormous and identification of the causative mutation from an unfiltered list would have been impossible. Individual exomes have been reported to contain around 40,000 SNVs and indels in sequences covered using the Agilent SureSelect v4 capture reagent (Chilamakuri et al., 2014) and this is likely to be even more with the v5 capture reagent that has a target size 39% larger than for v4.

However, with the benefit of hindsight, without the updated version of AnnotateSnps, a global analysis of all unfiltered HS WES data would have been

Gene	Variant	dbSNP reference, presence in dbSNP129 and pathogenicity classification in ClinVar release 29/09/2014	annotateSnps - common filter number of alleles identified in specified population(s) of the 1000 Genomes Project[a] (allele frequency in specified population(s))	WES coverage (OK: >10 reads Low coverage: ≤10 reads)	Biallelic filter	Families affected	Variant detected using filtering strategy?
PEX1	7:92136369C>G	N/P	N/P	OK	Pass	AI-202	Yes
PEX1	7:92136319delT	N/P	N/P	OK	**Filter**	AI-71	No
PEX1	7:92132483dupT	rs61750415 Present in dbSNP129 Marked as pathogenic	N/P	OK	Pass	AI-202	Yes
PEX1	7:92123671A>G	rs61750427 Present in dbSNP129 **Not marked as pathogenic**	KHV 2/198 (0.010)	OK	N/P	AI-71	No
PEX6	6:42946614A>C	N/P	N/P	OK	**Filter**	AI-146	No
PEX6	6:42946593C>A	N/P	N/P	**Low coverage**	N/P	AI-186	No
PEX6	6:42946235G>C	N/P	N/P	OK	**Filter**	AI-10	No
PEX6	6:42937452-42937459 delAGGCCTCC	rs267608216 Present in dbSNP137 Not marked as pathogenic	N/P	**Low coverage**	N/P	AI-16	No
PEX6	6:42935275G>A	rs61753224 Present in dbSNP129 **Not marked as pathogenic**	N/P	OK	N/P	AI-152	No
PEX6	6:42935188C>T	rs34324426 Present in dbSNP126 **Not marked as pathogenic**	CEU 2/198 (0.010) GIH 3/203 (0.0148)	OK	N/P	AI-10 AI-146 AI-186	No
PEX6	6:42932620C>A	N/P	N/P	OK	**Filter**	AI-16	No

Table 3.10: Details of the stages in the original filtering strategy when each *PEX1* and *PEX6* variant was retained or filtered.

The table shows that all *PEX1* and *PEX6* variants except those present in AI-202 were filtered out by the original variant filtering strategy. In each case a bold entry highlights the point at which a variant was lost in the filtering pipeline.

[a]1000 Genomes Project Phase 3 data, based on release 17, November 2015. Accessed 3rd March 2016.

Abbreviations: CEU Utah residents (CEPH) with Northern and Western ancestry (A 1000 Genomes Project major population; GIH Gujarati Indian from Houston, Texas (A 1000 Genomes Project major population); KHV Kinh in Ho Chi Minh City, Vietnam (A 1000 Genomes Project major population); N/P not present.

a more powerful strategy. Filtering variants that were present at a frequency of one percent or more in a large database of variation such as ExAC and retaining biallelic variants occurring in the same gene in two or more families, would have identified *PEX1* and *PEX6* variants in five of the seven HS families.

3.3.5 Identification of *USH2A* mutations in families AI-12 and AI-138

In two families, AI-12 (1509 and 1514) and AI-138 (2975), mutations in *USH2A* were identified from WES data (Table 3.3) and confirmed as biallelic by Sanger sequencing and segregation analysis (Figure 3.4). Biallelic mutations in *USH2A* are known to cause USH2 and non-syndromic RP (Eudy et al., 1998, Rivolta et al., 2000). In both families, it seems more likely that the resulting phenotype of these mutations is USH2, since the patients have both visual and hearing defects and the variants identified have already been reported in patients with USH2. *USH2A* variants associated with the non-syndromic RP phenotype (Lenassi et al., 2015) are not present in either family.

For 1509 and 1514 (AI-12), only five variants remained after filtering, and of these, the only gene with potentially biallelic variants with a CADD score of over 15 was *USH2A*, suggesting that the possibility that biallelic variants in another gene might be causing the SNHL was unlikely (Appendix 9). For 2975 (AI-138), ten genes with potentially biallelic variants remained after filtering, but none other than *USH2A* were known to lead to a SNHL phenotype (Appendix 10). Therefore it seems likely that the phenotypic overlap of the USH2 and HS had led to the inclusion of two families with USH2 in this study of HS.

USH is a combination of SNHL and RP without an associated enamel phenotype (Section 1.11), whereas, at the time of the commencement of this study, HS was described as a combination of SNHL and AI with or without MD (Lima et al., 2011). The timing of onset of the eye phenotype and the lack of reported vestibular dysfunction in HS most closely mirrors the phenotype of USH2, which is characterised by congenital moderate to severe hearing loss and onset of RP after puberty.

In general, USH is not associated with an enamel phenotype, although five reports of eight families detail individuals diagnosed with a combination of USH and an enamel phenotype (Table 1.1), including a report of one of the individuals included in this study, individual 2113 (AI-16) (Bateman et al., 1980, Innis et al., 1998, Pieke-Dahl et al., 2000, Balmer and Fayle, 2007, de la Pena and Valea, 2011). Of these individuals, none had mutations in known USH genes identified, so therefore none could be molecularly confirmed as USH patients. Should these patients be found to carry mutations in known USH genes, they may represent a new combination of phenotypes that includes an enamel defect, and may be the result of particular combinations of mutations in the USH gene affected or within an associated, but as yet unknown modifier of the USH phenotype. Alternatively, if the reported cases do not carry mutations in USH genes, they may be a mis-diagnosis of HS as USH (as is the case for individual 2113, family AI-16) or may consist of USH combined with an enamel defect that is not AI but is due to an environmental factor or life history exposure or event (Section 1.8.2).

The enamel phenotype described in individual 2975 (AI-138) consisted of defective enamel restricted to particular teeth (first permanent molars, the upper incisors, one upper canine tooth and one lower incisor). The dentist concerned did not diagnose AI since the phenotype was more suggestive of an environmental or life history associated enamel defect. Patient 2975 had experienced perinatal fevers and the restriction of the enamel phenotype to particular teeth may reflect the developmental time period over which enamel development was compromised. For individual 2975, the time period would be birth to around one year, since the development of the pre-molars was not affected (Nanci, 2012). This suggests that the phenotype may be due to a combination of USH2 and a non-genetic enamel defect. For 1509 and 1514 (AI-12), it was not possible to obtain a detailed description of the enamel phenotype and therefore it is not possible to determine whether it could be due to non-genetic causes.

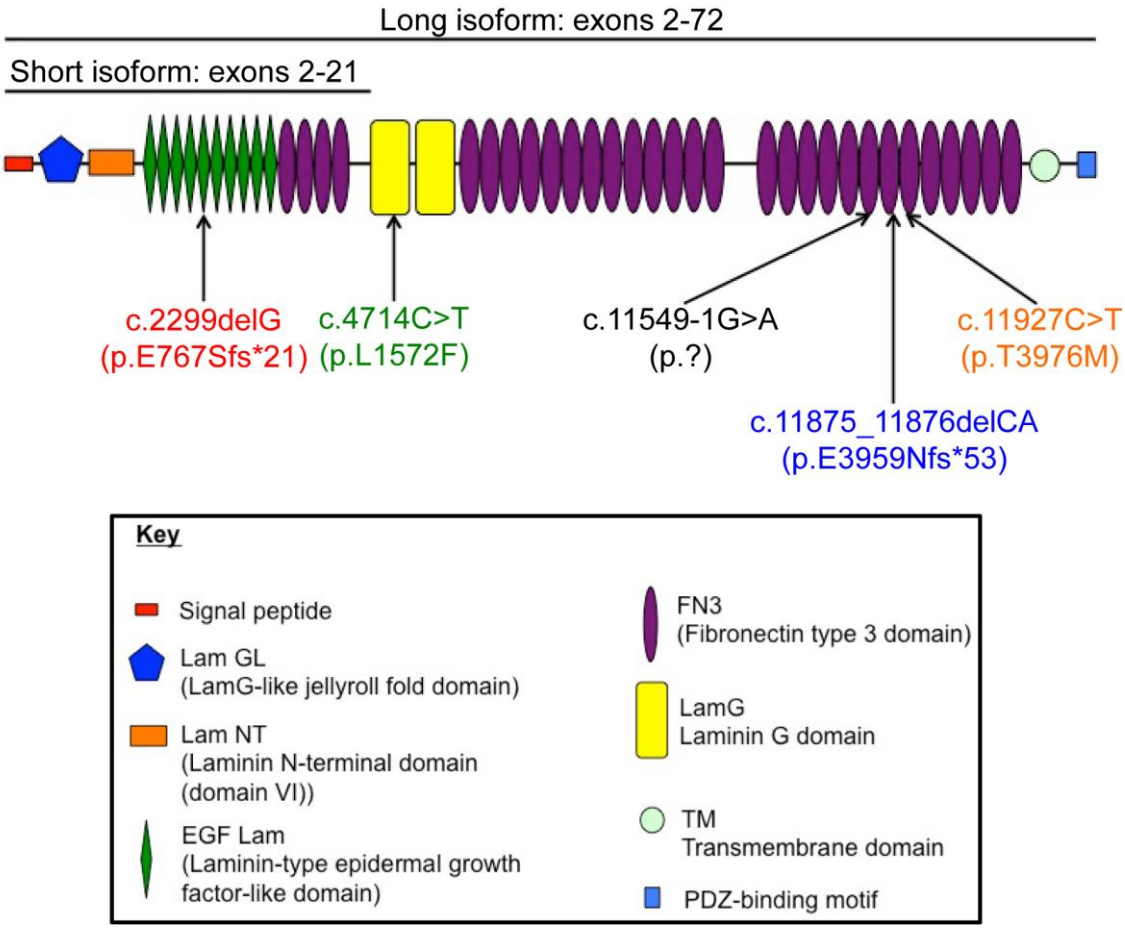


Figure 3.19: Schematic of USH2A protein.

The positions of the residues affected by the variants identified in AI-12 and AI-238 are shown. Reference transcript, protein and length for each isoform are as follows:

Long isoform: NM_206933, NP_996816, 5202 amino acids.

Short isoform: NM_007123, NP_009054 1546 amino acids.

USH2A spans 790kb and the full length transcript (NM_206933) includes 72 exons, encoding two alternatively spliced isoforms of the protein usherin (Figure 3.19). While the long isoform is a transmembrane protein encoded by all 72 exons, the short isoform is an extracellular protein encoded by exons 2 to 21 (Eudy et al., 1998, Weston et al., 2000, van Wijk et al., 2004). The two isoforms are known to have different expression patterns, structures, cellular localisations and may have different functions (Liu et al., 2007). The C-terminal PDZ binding motif present only in the long isoform has been shown to be crucial to interactions with other USH proteins that are necessary for its function (Adato et al., 2005, van Wijk et al., 2006, Ebermann et al., 2010). USH2A is expressed both in the inner ear and in the retina (Bhattacharya et al., 2002). In the

developing cochlea, the long isoform of usherin is a component of the ankle links, temporary structures crucial to the development of the stereocilia hair bundles and to the synapses (Reiners et al., 2005, van Wijk et al., 2006, Michalski et al., 2007, Zou et al., 2011). In the retina, usherin localises, along with other USH proteins, to the connecting cilium (the region between the inner and outer segments) and the synaptic region as well as to the periciliary membrane region (Maerker et al., 2008).

Evidence that mutations in *USH2A* cause USH has been strengthened by the generation of a homozygous knockout *Ush2a* (*Ush2a*$^{-/-}$) mouse model (Liu et al., 2007). *Ush2a*$^{-/-}$ mice initially had vision that was indistinguishable from that of wild-type (WT) mice, but over time signs of photoreceptor degeneration and death were detected and vision deteriorated as the photoreceptor nuclear layer and the photoreceptors' outer and inner segments thinned (Liu et al., 2007). This suggests that USH2A is required for the maintenance, but not the establishment of the photoreceptor layer of the retina. In the cochlea, hearing assessments at four months showed loss of sensitivity to higher frequency sounds, but the level of hearing loss remained stable upon subsequent reassessment. The hearing loss phenotype was therefore moderate and non-progressive, mirroring the phenotype seen in individuals with USH2 (Liu et al., 2007). Investigation of the cochlear hair bundles revealed that there was loss of the outer hair cells (Liu et al., 2007).

Tooth development, like that of the cochlear hair bundles, requires ciliated cells to coordinate important developmental signalling pathways involved in its formation (Liu et al., 2014). USH2A is temporally and spatially localised to the developing stereocilia bundle at the time the kinocilium is lost in the developing cochlea, and to the connecting cilium in photoreceptor cells (Liu et al., 2007). Since both the kinocilium and the connecting cilium are modified primary cilia (Goetz and Anderson, 2010, Wheway et al., 2014), this suggests that USH2A is involved in signalling from, or to, modified primary cilia. Cilia have also been found to be present during ameloblast development in rats (Sasano, 1986). Therefore it is possible that mutations in *USH2A* may affect amelogenesis,

although the absence of reports of USH patients with *USH2A* mutations and enamel defects suggests that this is not the case.

The mutations identified in 1509 and 1514 (AI-12) and 2975 (AI-138) had all been previously identified in USH patients. The single nucleotide deletion c.2299delG (p.E767Sfs*21), identified in both families, was amongst the first mutations reported in *USH2A* (Eudy et al., 1998) and is the most commonly identified mutation in USH patients, having been reported 535 times in the LOVD retinal and hearing impairment genetic Usher syndrome 2A (*USH2A*) mutation database (hereafter called USH2ADB) (Roux et al., 2011). It has been estimated that this mutation accounts for between 10 to 47.5% of all mutations identified in *USH2A* depending upon the population studied (Dreyer et al., 2008, Aller et al., 2010). The c.2299delG (p.E767Sfs*21) variant is believed to have originated from an ancestral haplotype that arose in the European population between 2700-7400 years ago and has been found to be commonly inherited with the c.4714C>T (p.L1572F) variant, also identified in both AI-12 and AI-138, with which it is in linkage disequilibrium (Aller et al., 2006, Baux et al., 2007, Aller et al., 2010, McGee et al., 2010). A study of the haplotypes surrounding both SNPs found that the wild-type c.2299G allele is always associated with the c.4714C allele and eight out of ten different mutant c.2299delG haplotypes identified in Southern European populations (in around 85% of individuals) were associated with the c.4714T allele (Aller et al., 2010). These ten haplotypes have been defined based on variants at nine variant SNPs within *USH2A* (Aller et al., 2010). Investigation of the WES data for AI-12 and AI-138 showed that haplotypes for each family are likely to be either M6 or M7 for AI-12 and M3 for AI-138 (Appendix 20), although additional Sanger sequencing would be required to confirm this. Due to the c.4714C>T variant's constant association with the c.2299delG frameshift allele, it is regarded as a non-pathogenic polymorphism, although the majority of scores obtained from mutation prediction software classify it as a pathogenic change (Table 3.3) (Baux et al., 2007, McGee et al., 2010). The c.2299delG (p.E767Sfs*21) mutation in exon 13 would be expected to lead to NMD of the transcript and therefore result in a null allele, but RT PCR analysis of RNA from homozygous individuals revealed that exon skipping of exon 13, or both exons 12 and 13, may also occur (Lenassi et

al., 2014). The skipping of exon 13 would potentially generate an in-frame protein product, although this may not function, could aggregate, or may be degraded due to misfolding (Waters, 2001, Lenassi et al., 2014).

For 1509 and 1514 (AI-12), a second frameshift variant, c.11875_11876delCA (p.E3959Nfs*53), was identified and was found to lie in *trans* to the c.2299delG (p.E767Sfs*21) variant. This mutation has been reported seven times in USH2ADB in European individuals (Roux et al., 2011). The combination of the c.2299delG and c.11875_11876delCA alleles identified in the two affected individuals of family AI-12 has been identified in at least two other European families (Le Quesne Stabej et al., 2012, Baux et al., 2014), although for one of these families, segregation analysis was not performed.

For individual 2975 (AI-138), a splicing mutation c.11549-1G>A (p.?) was identified and was found to be shared by the mother (4725). Since DNA from additional family members was not available for segregation analysis of the variants, it is possible either than the c.2299delG (p.E767Sfs*21) mutation is in *trans* on the paternal allele or that it is a *de novo* mutation and could be present on either allele. However, since the c.2299delG (p.E767Sfs*21) variant is regarded as a common founder mutation (Dreyer et al., 2001), it seems more likely that the mutation is present in *trans* on the paternal allele. The splicing variant has been reported only once in the USH2ADB in a Canadian individual (Roux et al., 2011), not in combination with the c.2299delG mutation (Lenassi et al., 2015). Analysis of the c.11549-1G>A mutation using both the PSSM and MEM algorithms via Sroogle (Section 2.8.4.5) suggested that it is expected to alter splicing by reducing the efficacy of the 3’ splice acceptor site in intron 59, since both scores for the variant were lower in comparison to the reference sequence. In addition, the percentile rank of each score (0.01 and 0 respectively for constant exons) indicated that nearly all other sites were more likely as 3’ acceptor splice sites (Table 3.5). Ideally, the effect of the mutation on splicing would be analysed by RT-PCR, using RNA obtained from nasal epithelial cells (Vache et al., 2010), but no RNA from individual 2975 was available. One possibility is that, since exon 59 is in phase zero (Kolkman and Stemmer, 2001), exon 60 could be omitted from the final spliced transcript with

a consequential loss of 76 amino acids but without the consequence of an amino acid sequence frameshift or premature stop signal. Investigation of Ensembl transcripts and a literature search did not identify a natural, alternatively spliced transcript that lacked exon 60, although in-frame alternative splicing to exclude exons 50-52 and 62 has been documented (van Wijk et al., 2004). The deletion of 76 amino acids resulting from the loss of exon 60 would disrupt fibronectin type III domains 23 and 24 and therefore the protein might be unstable. Whilst this scenario is one possibility, it also remains possible that alternative 3' splice acceptor sites could be used or that the intron could be retained. The effect of the mutation may even be different in different tissues and at different stages of development.

A missense variant, c.11927C>T (p.T3976M), was also identified in the proband and the mother. This variant has been reported once in a French USH patient (USH2ADB lists two family members who both carry the variant (Table 3.4)) that also carried a second mutation predicted to affect splicing on the same allele (Baux et al., 2007). This missense variant is therefore considered as unlikely to be pathogenic. Consistent with that view, mutation prediction outcomes (Table 3.3) varied between the four algorithms used, with two that were borderline neutral (Provean and SIFT) and another two predicting the variant to be damaging (Polyphen-2 and MutationTaster).

The frameshift variants identified in both AI-12 (1509 and 1514) and AI-138 (2975) would be expected to produce a transcript that would be subject to NMD. However the USH2A protein has both a short (encoded by exons 2-21) and a long isoform (encoded by exons 2-72). Therefore the short isoform protein, translated from the allele with the frameshift mutation in exon 61 (c.11875_118876delCA) in 1509 and 1514 (AI-12) and the allele with the splicing mutation in intron 59 (c.11549-1G>A) in 2975 (AI-138), may still function.

In conclusion, individuals from both families AI-12 and AI-138 carry biallelic variants in *USH2A* that would be expected to produce transcripts that were subject to NMD, or could result in the protein being misfolded and degraded.

The combinations of the c.2299delG (p.E767Sfs*21) and the c.11875_11876delCA (p.E3959Nfs*53) variants in 1509 and 1514 (AI-12) and the c.2299delG (p.E767Sfs*21) and the c.11549-1G>A (p.?) variants in 2975 (AI-138) are likely to be causing the USH phenotype. It therefore seems likely that the affected individuals in these two families have USH2 and not HS. The enamel defect in these families remains unexplained, but may result from a non-genetic cause, and for 2975 (AI-138), was not thought to be AI.

3.3.6 Identification of a *MYO7A* deletion and an *ENAM* SNV in family AI-11

In family AI-11, a heterozygous deletion that included the first two exons of *MYO7A* was detected by low coverage WGS of DNA from individual 1502 (Figure 3.6). Analysis of WES data from individuals 1503 and 1504, affected members of the same family, revealed that a deletion could also be detected in them by the use of FishingCNV, but that it was predicted to also include the final two 3′ exons of *CAPN5* (Figure 3.6 and Table 3.6). The deletion, and therefore also its breakpoints, remains unconfirmed by Sanger sequencing.

The predicted heterozygous deletion of *MYO7A* exons 1 and 2 was one of many CNVs identified by DNAcopy and FishingCNV, but was investigated due to overlapping calls from both programs using sequence data from three individuals, together with the known association of *MYO7A* with deafness and blindness phenotypes (Weil et al., 1995, Liu et al., 1997a). Both DNAcopy and FishingCNV identify large numbers of CNVs that must be prioritised in order to identify variants likely to cause disease. For DNAcopy, a conservative cut off of a copy number of ≤0.68 or ≥1.18 x diploid was used to minimise the possibility that the use of low coverage resulted in heterozygous CNVs being filtered out. Calls identified by DNAcopy were then compared with calls identified by Fishing CNV for both 1503 and 1504 and any non-overlapping calls were discarded. Unlike DNAcopy, FishingCNV assigns a Holm adjusted p value (Holm, 1979) to each CNV call, which allows the variants to be prioritised. A cut off value of $p=1 \times 10^{-5}$ was assigned to limit analysis to the CNVs most likely to represent true calls.

By definition, the CNV calls produced from WES by FishingCNV are limited to variants that include targetted regions i.e. they are protein coding. Therefore the strategy of prioritisation excluded DNAcopy CNV calls that did not affect regions that included protein coding exons. While it is possible that intergenic and intronic regions can have regulatory functions, current annotation of these regions limits prioritisation of CNV calls that include such regions. Also, in this case, the lack of data for such regions for two family members would have meant that prioritisation of variants could not be based on segregation with the disease phenotype in the family.

The CNV filtering strategy left only two CNV calls (Table 3.6). The calls were investigated by checking for known variation in the region via the Database of Genomic Variants (MacDonald et al., 2014) and by Pubmed search (Section 2.8.3.1) for information as to the genes' functions and expression profile.

One CNV call was a heterozygous deletion of *ATAD3B* and *ATAD3C*, genes not previously associated with disease and about which little is known. ATAD3B is known to be a mitochondrial membrane protein that contributes to the stability of nucleoids, structures consisting of mitochondrial DNA and proteins (He et al., 2007). An *Atad3* genetrap insertion model was shown to die at embryonic day 7.5 due to the impaired mitochondrial biogenesis and its impact on trophoblast development (Goller et al., 2013). The related protein ATAD1 (MIM *614452) is known to be associated with peroxisomes and has been suggested to be involved in peroxisomal homeostasis (Wiese et al., 2007). Therefore, it is possible that the heterozygous deletion of both *ATAD3B* and *ATAD3C* may, like variants in *PEX1* and *PEX6*, affect peroxisomal function. However a second variant, which would be expected in a recessively inherited disease, was not identified in either gene.

The other CNV call included a partial deletion of *MYO7A*. Mutations in *MYO7A* are a known cause of USH1 (USH1B) (Weil et al., 1995) and isolated SNHL with both recessive and dominant inheritance (Liu et al., 1997a, Liu et al., 1997b). The phenotype and mode of inheritance in AI-11 led to the suspicion that a second *MYO7A* variant may be present in WES data but had been

filtered out through the use of the original variant filtering strategy. A second variant was not identified in the unfiltered WES data for 1503 and 1504, even after exons covered by five or fewer reads were Sanger sequenced. Low coverage of particular exons, particularly those in GC rich regions, is a known problem for WES (discussed in Section 3.3.4).

Since the predicted heterozygous partial deletion of *MYO7A* identified in AI-11 remains unconfirmed by Sanger sequencing, it is not known which members of the family carry the variant. It seems likely that the deletion is present in affected individuals 1502, 1503 and 1504 since overlapping regions of 0.53, 0.57 and 0.62x diploid copy number could be detected in three affected family members either by analysis of WGS data using DNAcopy or by FishingCNV analysis of WES data. However, no data is available for unaffected family members. It is expected that the deletion breakpoints will be more accurately predicted by the WGS analysis since data from WES is limited to only the protein coding exons and so breakpoints that lie within non-coding regions cannot be accurately located. This suggests that the deletion encompasses the first two exons of *MYO7A* and some flanking intronic sequence but does not include exons 12 and 13 of *CAPN5* as suggested by the WES data. However, the accuracy of the breakpoint prediction from the low coverage WGS will be highly dependent on the local coverage at that region, with higher coverage providing more accurate breakpoint detection.

Mutations in *CAPN5* cause AD neovascular inflammatory vitreoretinopathy (Mahajan et al., 2012). This autoimmune condition may present at different stages of disease progression with different phenotypes, including uveitis, RP, proliferative diabetic retinopathy and proliferative vitreoretinopathy (Bennett et al., 1990). Even though the breakpoints of the putative heterozygous deletion remain unknown, and therefore it is uncertain to what extent the 3′ region of *CAPN5* is deleted, given the autosomal recessive mode of inheritance of disease in AI-11, it seems unlikely that any heterozygous deletion of *CAPN5* is causing the blindness phenotype in this family. Ideally, further investigation of the family's phenotype would have been undertaken, but this was not possible

due to loss of contact with the referring clinician.

Low-coverage WGS detection of CNVs has been shown to be comparable in efficacy to the use of array comparative genomic hybridisation (aCHG) profiling (Hayes et al., 2013). The resolution obtained for 1502 was 23kb in comparison to the 76kb quoted by Hayes et al. (2013) for low-coverage WGS. The heterozygous *MYO7A* deletion detected was predicted to span 5-10kb, suggesting that the resolution across this locus was higher than average.

Although mutations in *MYO7A* are known to be the most common cause (47.5-63.3%) of USH1 in European populations (Jaijo et al., 2006, Roux et al., 2011), with 470 unique mutations recorded in the LOVD retinal and hearing impairment database myosin VIIA (*MYO7A*) (hereafter called MYO7ADB) (Roux et al., 2011), very few of the mutations identified in patients are deletions of one or more exons. This may reflect the previous methodological bias towards identification of SNVs over CNVs, when Sanger sequencing was the method of choice for mutational screening and initial identification. Heterozygous whole exon deletions are “invisible” when sequencing primers do not anneal at positions flanking the deletion breakpoints. It may also result from a continued bias in the use of WES to identify only SNVs and small indels, with separate global analysis of WES datasets required to identify potential CNVs.

Investigation of MYO7ADB (Roux et al., 2011) revealed that only one deletion spanning the region of the predicted deletion in AI-11 has been reported (Accessed 29th February 2016). The heterozygous deletion extended from within the second intron of *CAPN5* (NM_004055) to intron 37 of *MYO7A* (Spanning chr11:76,799,656-76,913,682; 114,027bp) and was reported in *trans* with a heterozygous *MYO7A* pathogenic missense mutation (c.999T>G, p.Y333*) in an USH1 patient of French origin (Roux et al., 2011). Investigation of the Database of Genomic Variants (MacDonald et al., 2014), a database of structural and indel variation in healthy individuals, revealed two deletions over the *CAPN5* and *MYO7A* genes (rsv779225 and rsv779224) spanning chr11:76,832,011-76,891,661 (59,651bp) and chr11:76,832,011-76,873,620 (41,610bp) respectively (Cooper et al., 2011). While these deletions include the

region potentially deleted in AI-11, none of the deletions reported appear to match the approximate breakpoints. Since no additional variant in *trans* has been identified, it remains possible that the CNV may simply be part of the variation that each individual carries without pathological consequence (McCarroll et al., 2008, Conrad et al., 2010, Handsaker et al., 2015).

Mutations in *MYO7A* are most commonly associated with USH1, but reports of individuals carrying *MYO7A* mutations with an USH2 phenotype, as reported in the case here, also exist (Bonnet et al., 2011, Aparisi et al., 2014, Rong et al., 2014). Digenic inheritance of USH has also been suggested in an USH1 patient that carried single heterozygous *MYO7A* and *PCDH15* mutations, although this remains to be proven since only the exonic sequence was obtained in that study (Rong et al., 2014). Additional evidence to support the possibility that a combination of heterozygous mutations in different USH genes can contribute to a disease phenotype is provided by increased levels of hearing loss in *Myo7a*$^{+/-}$ double heterozygous *Ush1g*$^{+/-}$, *Cdh23*$^{+/-}$ and *Pcdh15*$^{+/-}$ mice (Zheng et al., 2012), although these results may be mutation and genetic background dependent, since an earlier report found no additive effect for *Myo7a*$^{+/-}$ *Cdh23*$^{+/-}$ mice (Holme and Steel, 2002). The possibility of digenic inheritance for AI-11 was investigated but no additional variants, predicted to be pathogenic, were identified.

Studies reporting on the sequencing of cohorts of USH patients have identified many patients for which only one mutation in *MYO7A* can be identified (Jaijo et al., 2006, Aparisi et al., 2014). If these cases are not caused by mutations in another gene, this could suggest that a second mutation lies either in the promoter, the UTRs or deep in intronic sequences that are not covered by conventional exonic Sanger sequencing or WES. Similarly it is also possible that WES and low coverage WGA failed to identify the second causative mutation in 1502, 1503 and 1504 (AI-11) because either the position of the variant or the coverage achieved meant that it was not recorded or was filtered out in the strategies utilised. High coverage WGS of two affected family members would allow a more thorough analysis through more even coverage of the regions targetted by WES, and greater overall coverage for more sensitive

detection of small to medium CNVs by reducing the window size of the DNAcopy analysis.

For AI-11, it is possible that the AI phenotype may also be caused by the *MYO7A* variants, although if mutations in *MYO7A* caused AI, it would be likely that this would have been reported, given the number of mutations already identified in USH patients. Additionally, expression studies did not identify *MYO7A* as being differentially expressed during amelogenesis in rodent enamel organs (Lacruz et al., 2011, Lacruz et al., 2012b, Simmer et al., 2014).

Alternatively, since it was possible that AI-11 was segregating mutations in two genes that cause components of the phenotype that has been diagnosed as a syndrome, variants in genes known to cause AI were investigated. This revealed a heterozygous variant in *ENAM*, c.332A>C, p.(K111T) shared by both 1503 and 1504 in WES data and by 1502 upon confirmation of the variant by Sanger sequencing. The variant was also carried by two family members not known to have AI. Loss of contact with the recruiting physician meant that it was not possible to verify that these two family members did not have AI. Analysis of the variant with mutation pathogenicity prediction programs (Appendix 13) suggested that the c.332A>C variant is likely to be benign. Publications of *ENAM* mutations that cause AI have detailed only two missense variants, with the remainder of reports detailing fourteen null mutations and one further mutation that leads to an in frame insertion (Appendix 19). Comparison of the variant's mutation pathogenicity prediction program scores with those for the *ENAM* variants previously reported showed that the c.332A>C variant was amongst the least pathogenic. Therefore it seems unlikely that the *ENAM* variant identified here is responsible for the phenotype of AI-11 affected individuals 1502, 1503 and 1504. As for 1509 and 1514 (AI-12) and 2975 (AI-138), it remains possible that the enamel phenotype of AI-11 is not genetic in origin.

4 Association of an *AMTN* deletion with AI

4.1 Introduction

At the time of this study, AI was known to be caused by mutations in a number of genes (Section 1.9). Of these, only mutations in *ENAM* (Rajpar et al., 2001), *FAM83H* (Kim et al., 2008), *LAMA3* (Yuen et al., 2012), *LAMB3* (Kim et al., 2013, Poulter et al., 2014b) and *COL17A1* (McGrath et al., 1996) were known to cause non-syndromic autosomal dominant (AD) AI. Mutation screening in cohorts of AI patients had shown that mutations in known AI genes accounted for between 28-49% of AI cases (Chan et al., 2011, Wright et al., 2011, Prasad et al., 2016), suggesting that more genes were yet to be identified.

A number of strong candidate genes for involvement in AI, implicated through function, pattern of expression and / or animal model studies, were also suggested in the literature. Foremost amongst these, was the gene encoding amelotin (Iwasaki et al., 2005).

The aim of the work presented in this chapter was to identify novel AI gene(s). The second aim was to characterise the enamel phenotype and to compare this to the phenotype of any reported mouse models.

Therefore, families with AI were recruited and their DNA subjected to WES. Teeth were also obtained and the enamel phenotype was analysed by microCT, SEM and EDX.

In collaboration with Dr. Gina Murillo (University of Costa Rica, San Pedro, Costa Rica), one family (family AI-154) with non-syndromic, AD hypomineralised AI was identified and recruited to the study. This family is the focus of this chapter.

4.2 Results

Details of the primers used in the following section are listed in Appendix 21.

4.2.1 Clinical phenotype

Five individuals were recruited to the study from family AI-154; three affected (II:2, III:3 and IV:1) and two unaffected (II:3 and IV:2). Affected individuals presented with hypomineralised AI in the absence of any co-segregating health problems. The pedigree for family AI-154 and the clinical phenotype of individual IV:1 are shown in Figure 4.1. The phenotype of the unaffected individual IV:2 is shown in the Appendix 22. Individuals were recruited in two batches, with initial recruitment including only the three affected individuals.

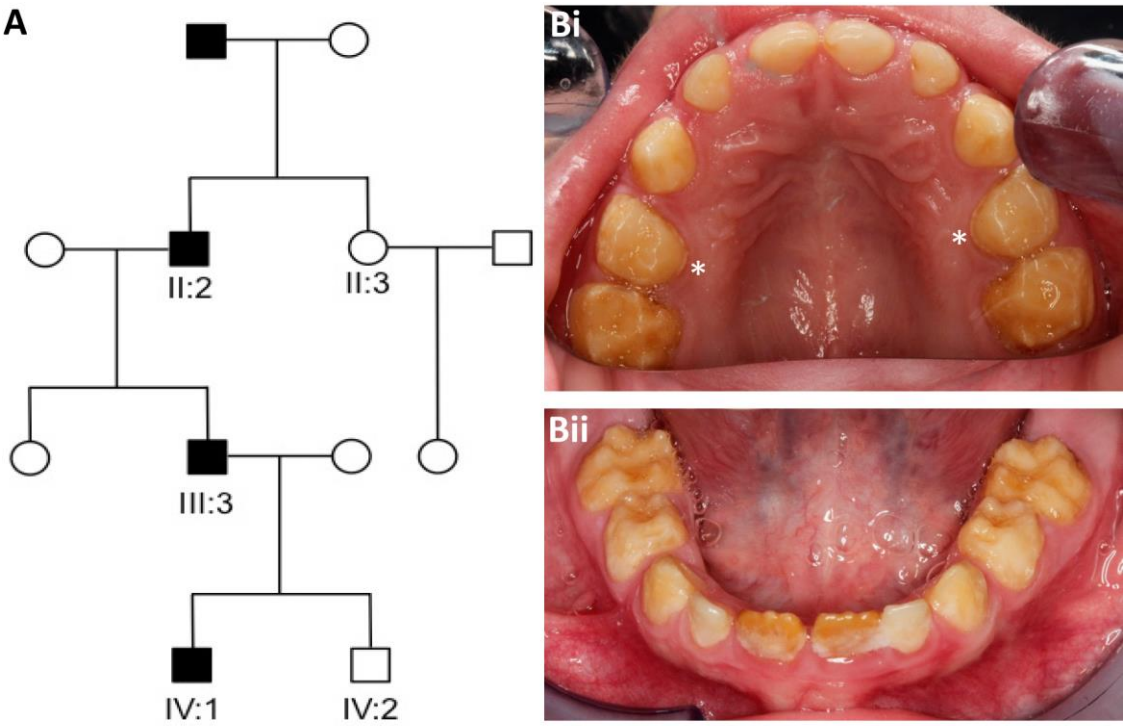


Figure 4.1: Pedigree and dental phenotype of AI-154.

A: Pedigree of the Costa Rican family investigated. Affected family members are shaded.

B: The mixed maxilliary (i) and mandibular (ii) dentition of the index case, IV:1, aged 5 years, was characterised as generalised hypomineralised AI with some hypoplasia involving all teeth. There was post-eruptive enamel loss with retention of a thin band of enamel at the cervical margin (examples marked with ∗). Reproduced with permission from Smith et al. (2016a).

4.2.2 Whole exome sequencing of two affected individuals from family AI-154

In an attempt to identify the genetic cause of AI in family AI-154, WES libraries were prepared from genomic DNA from two affected individuals, III:3 and II:2, using the 3μg protocol (Section 2.4.1.1). DNA was extracted from saliva DNA as previously described (Section 2.3). Ideally, DNA from IV:1 would have been used to prepare a WES library instead of III:3, however DNA from IV:1 was of insufficient quantity and quality for the 3μg protocol in use at the time. Therefore, libraries from III:3 and II:2 were prepared using the SSV5 capture reagent, each library was pooled with other samples and each pool was sequenced in a separate run.

The resulting Fastq sequence files were aligned using Novoalign and the variants called, filtered and annotated as previously described for rare dominant disease (Sections 2.4.2, 2.4.3 and 2.4.4). The depth of coverage across the exome was also analysed (Section 2.4.5) and the mean was found to be 66.4x and 60.1x for samples III:3 and II:2 respectively (Appendix 23). The resulting variant files were further filtered by selecting only the heterozygous variants that were shared by both II:2 and III:3. In addition, any identical variants that were present in an additional WES dataset from an unrelated patient affected with autosomal recessively inherited AI caused by a homozygous mutation in *MMP20*, processed using the same alignment and variant calling pipeline, were also filtered. This was to aid the exclusion of artefactual variants inherent to the library preparation and sequencing platform.

The resulting variant list contained 63 variants with a CADD score of 15 or more (data not shown). Investigation of the genes affected by each of the variants by PubMed searches did not present any clear candidates based on known function. Since a large number of candidate variants remained, no analysis of their segregation was attempted.

4.2.3 Identification of CNVs using WES data

As an extension of the standard pipeline of analysis of WES data, CNVs were identified using the ExomeDepth software (Section 2.4.6). This was used since adequate numbers of controls were available for each data set. Five and eleven unrelated samples were sequenced in the same batches as II:2 and III:3 respectively, although ideally ten reference exomes are required for optimal analysis (Plagnol et al., 2012).

Analysis of ExomeDepth data for individuals III:3 and II:2 revealed a total of twenty-five and twenty-two CNVs respectively that were not identified as a common CNV in the study by Conrad and co-workers (Conrad et al., 2010). Of these, four were identified in both individuals, none of these included a gene for which mutations are known to cause AI nor any known AI candidate gene. Only one CNV included a gene found to be expressed during amelogenesis (Yin et al., 2014b), tripartite motif 16 (*TRIM16*; MIM *609505).

Since CNV detection may not have identified all variants in both individuals, separate analysis of the CNVs identified in each individual was undertaken. The CNVs identified in III:3 included a predicted heterozygous deletion encompassing the region chr4:71388473-71394475 and therefore predicted to remove exons 3 to 6 of *AMTN* (Table 4.2, Figure 4.2B). No other variants were identified in known AI candidate genes in either individual. To determine whether a CNV within the *AMTN* gene was detectable using ExomeDepth for individual II:2 with less stringent parameters for calling variants, the transition probability setting was increased to 10^{-2} from the standard setting of 10^{-4}, however, again there were no CNVs located over *AMTN*. If present, the deletion may not have been detected due to fewer controls being available for analysis using ExomeDepth for individual II:2 compared with individual III:3.

Therefore, to determine whether the deletion over *AMTN* exons 3 to 6 was present in III:3 and to check whether it was also present in II:2 despite not being identified by ExomeDepth, the breakpoints of the predicted deletion in III:3 were investigated (Figure 4.2). Since ExomeDepth can only determine the extent of a

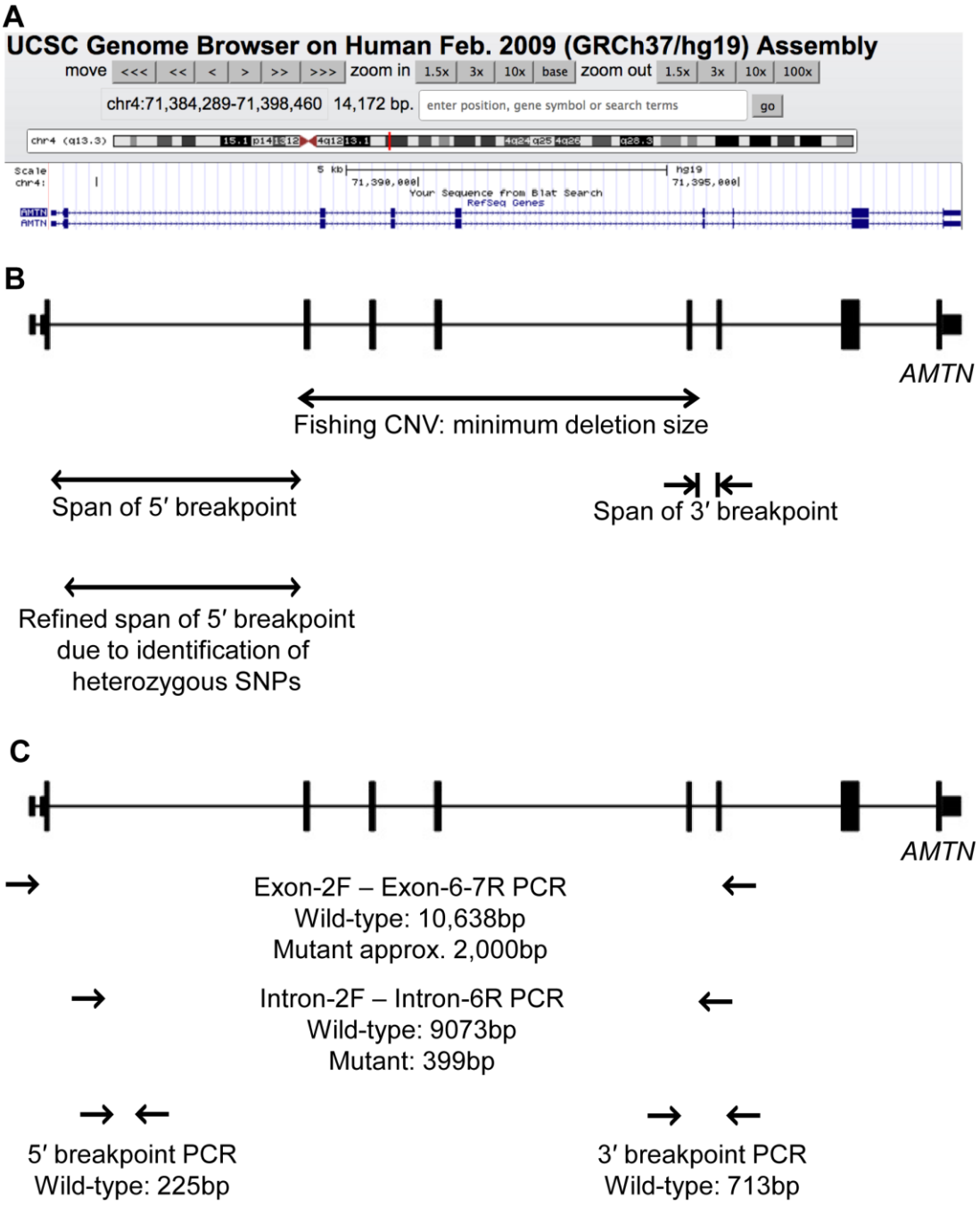


Figure 4.2: Identification of the breakpoints of the *AMTN* exons 3-6 deletion identified by ExomeDepth.

A: UCSC Genome Browser view of *AMTN* RefSeq transcripts NM_212557 and NM_001286731, showing their chromosomal position at chr4:71,384,289-71,398,460 (GRCh37) and the relative positions of each exon and intron. B: Schematic of *AMTN,* comparing the relative position and span of the deletion detected by ExomeDepth in patient III:3 and the potential location of the breakpoints based on the positions of the flanking regions captured for WES and the positions of heterozygous SNPs. Boxes of smaller and larger height indicate the position of non-coding and coding exons respectively. C: Schematic of *AMTN* labelled with the positions of the primers used to identify and characterise the breakpoint.

CNV based on the boundaries of the capture reagent's target for each exon, the possible position of the 5′ breakpoint spanned from the 3′ end of the capture reagents' *AMTN* exon 2 target at chr4:71,384,656 to the 5′ end of the *AMTN* exon 3 captured region at chr4:71,388,407, a region encompassing 3,751bp (Figure 4.2B). In order to further refine the 5′ breakpoint region, the position of the most 3′ heterozygous SNV was obtained from the unfiltered variant list for individual III:3. This refined the region of the 5′ breakpoint to between chr4:71,384,840 and chr4:71,388,407 (3,567bp; Figure 4.2B). For the 3′ breakpoint, the close proximity of exon 7 meant that its possible location was refined to between chr4:71,394,541-71,394,809 (268bp; the 3′ end of the captured region of exon 6 to the 5′ end of the captured region of exon 7; Figure 4.2B). Therefore primers that had previously been used to sequence exons 2 and exons 6-7 were used in an attempt to amplify DNA spanning the breakpoints (Figure 4.2C). A PCR using *AMTN* primers ex2F and ex6-7R produced a product of around 2,000bp from DNA templates from affected family members (Figure 4.3). For comparison, a product of size 10,638bp would be expected from wild-type DNA. The wild-type fragment was not amplified, likely due to the enzyme and PCR conditions used for the reaction being inappropriate for such a large product size.

The presence of the approximately 2,000bp product and the relative spans of the predicted 5′ and 3′ breakpoints suggested that sequencing of the product using the reverse (exon 6-7) primer would enable the sequence across the deletion breakpoints to be determined. Sanger sequencing confirmed the presence of the deletion and revealed that the 5′ and 3′ breakpoints were positioned at chr4:71,385,895 and chr4:71,394,574 respectively. In addition to the deletion, a 4bp insertion of CTCA between these points had occurred.

Based on the identification of the exact breakpoints, primers were designed to amplify a product spanning the deletion breakpoints with flanking sequence of only a few hundred base pairs each side (Figure 4.2C). This would produce a shorter product that spanned introns 2 to 6 that could be sequenced from both directions to confirm the breakpoints of the deletion. These primers (intron 2F-

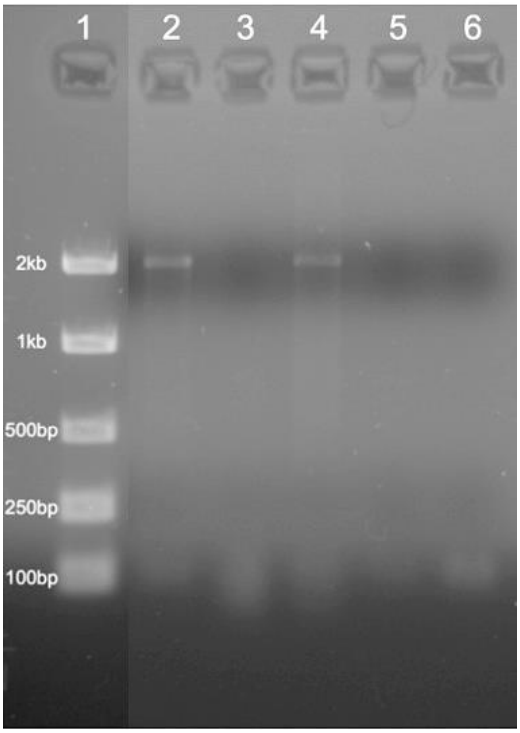


Figure 4.3: Amplification of the *AMTN* gene region, using primers designed to amplify exons 2 to 7, for AI-154 family members.

The PCR products generated by the amplification of DNA from AI-154 family members, III:3, IV:1 and II:2 (lanes 2-4 respectively) and unrelated DNA (lane 5). A negative control is shown in lane 6. A 100bp marker is shown in lane 1.

No product was generated from IV:1 due to the poor quality and low quantity of DNA.

intron 6R) were also used to analyse the segregation of the deletion in all available AI-154 family members. Saliva samples were obtained from additional family members (III:3 and IV:2) and a second sample was obtained for IV:1. DNA extraction (Section 2.3), PCR amplification (Section 2.5) and subsequent size fractionation by agarose gel electrophoresis (Section 2.3.3) revealed that the deletion was present in all affected family members (II:2, III:3 and IV:1) corresponding to a band of approximately 400bp (Figure 4.4A). The predicted wild-type band of 9,073bp was not observed in controls, nor in patients, likely due to its large size and the PCR conditions, which were designed to be optimal for amplification of the 400bp band. In contrast, a PCR amplification using primers flanking exons 6 and 7, which would amplify the wild-type allele, produced a band of the expected size (713bp) in all individuals of family AI-154 (Figure 4.4A).

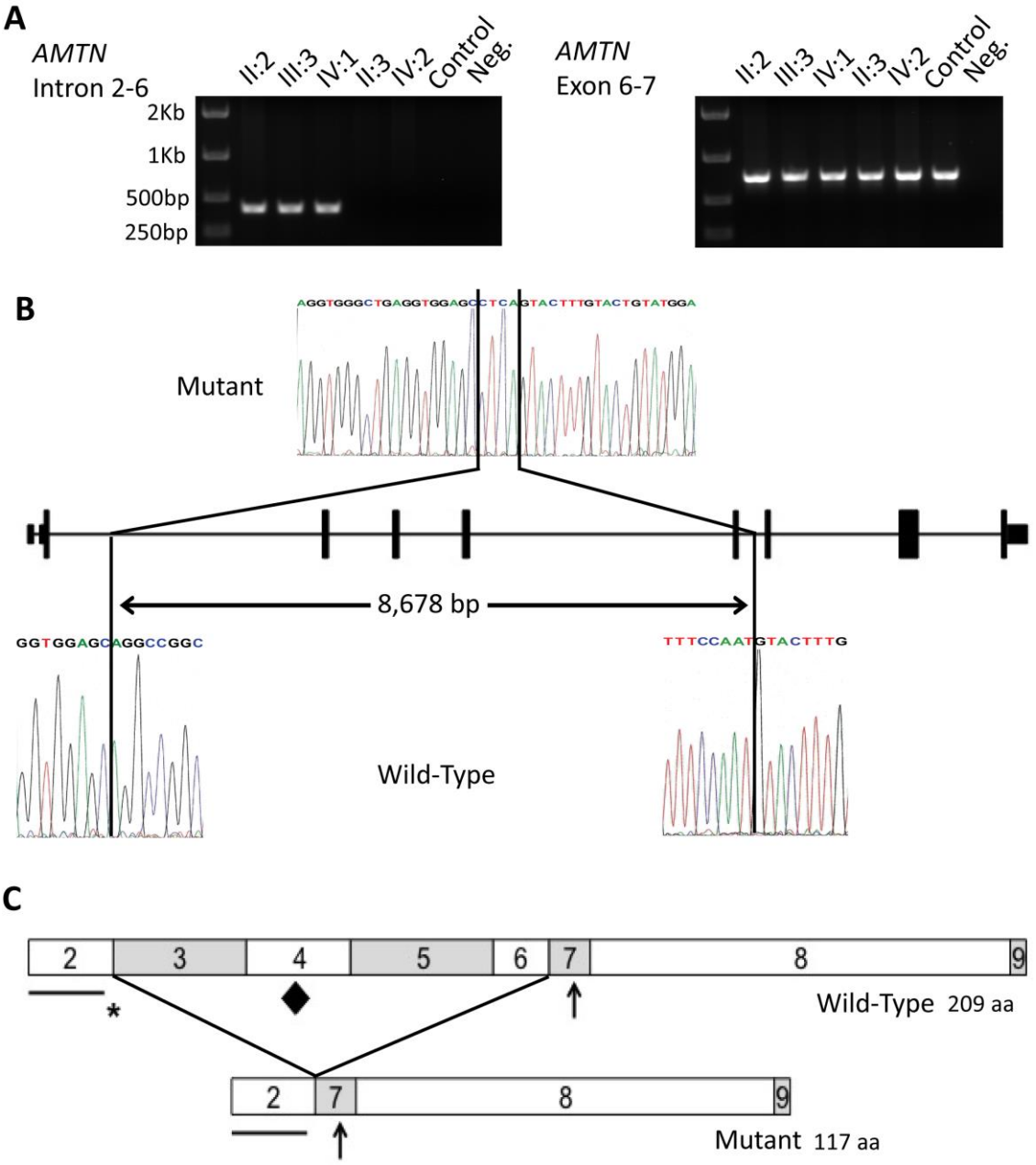


Figure 4.4: Genotyping of the *AMTN* deletion in family AI-154, identification of the deletion breakpoints and a schematic diagram of the predicted effect of the deletion on the AMTN protein.

A: PCR analysis of the heterozygous deletion of exons 3 to 6 of *AMTN* in the members of the family investigated. Amplification using primers spanning exons 6 to 7 produced a band of 713bp in all individuals. In contrast, amplification of introns 2 to 6, designed to produce a product of 9,073bp for the wild-type allele, produced a product of 399bp in affected individuals (II:2, III:3 and IV:1) confirming a deletion encompassing exons 3 to 6. No product of size 399bp was amplified in unaffected individuals (II:3 and IV:2) or a control sample.

B: Sanger sequencing electropherograms of mutant and wild-type alleles to identify the 5′ and 3′ breakpoints of the deletion. Comparison of the mutant sequence with a control sequence revealed a deletion of 8,678bp spanning exons 3 to 6 of *AMTN* and an insertion of 4bp.

C: Schematic diagrams of the wild-type and the predicted mutant AMTN protein structures.
The predicted mutant AMTN protein lacks the amino acid sequence encoded by exons 3 to 6, resulting in the loss of 92 amino acids. The contribution of each exon to the protein is signified by the boxes labelled with the number corresponding to the exon number. Important motifs are labelled with symbols as follows: the line shows the position of the putative signal peptide, encoded by exon 2, the asterisk shows the position of the LPQ motif, encoded by exons 2 and 3, and the diamond shows the position of the IPLT motif, encoded by exon 4, predicted to be an O-glycosylation site. The arrow shows the position of the SXE phosphorylation motif, encoded by exon 7. Reproduced with permission from Smith et al. (2016a).

The combination of both the intron 2-6 and the exon 6-7 PCR assays revealed that the *AMTN* heterozygous deletion spanning exons 3-6 was present for affected individuals II:2, III:3 and IV:1. Sequencing of the intron 2-F to intron 6-R product from affected individuals II:2, III:3 and IV:1 confirmed the breakpoints of the heterozygous deletion within *AMTN* spanning a total of 8,678bp, accompanied by a 4bp insertion (chr4:71,385,896-71,394,573delinsCTCA; GRCh37). The deletion encompasses all 276bp of exons 3-6 plus 2,576bp of 5′ genomic sequence and 98bp of 3′ sequence (c.54+1347_330+98delinsCTCA (NM_212557, Figure 4.4B). The heterozygous deletion was present in all available affected family members (II:2, III:3 and IV:1) and absent from all unaffected family members tested (II:3 and IV:2). Since both intron 2 and intron 6 are in phase zero (Sharp, 1981), this deletion is predicted to create an in-frame deletion of 92 amino acids (p.Q19_Q110del, NP_997722), shortening the protein from 209 to 117 amino acids (Figure 4.4C).

4.2.4 cDNA analysis

In an attempt to determine the consequences of loss of exons 3-6 on transcription of *AMTN* mRNA, leukocyte cDNA was prepared by Dr. Murillo after obtaining whole blood from affected individual III:3. While cDNA amplification with control primers (*P53*) was successful (data not shown), amplification of *AMTN* products failed in both patient and control cDNA. This is consistent with the reported tissue specific expression pattern of *Amtn* (Iwasaki et al., 2005). As a consequence there is no detectable expression of *AMTN* mRNA in blood, a known issue for enamel-specific transcripts (Poulter et al., 2014c).

4.2.5 Whole exome sequencing of additional family members

While the deletion in *AMTN* remained the prime candidate mutation in AI-154, there were still additional variants that were present in both affected individuals sequenced by WES. In order to exclude some of these SNVs, indels and CNVs, DNA from additional family members IV:1 and II:3 was subjected to WES using the 200ng protocol (Section 2.4.1.2).

The resulting Fastq sequence files were aligned using BWA and the variants called and filtered as previously described for a rare dominant disease. The depth of coverage was also analysed and the mean depth was found to be 110.5x and 89.8x for IV:1 and II:3 respectively (Appendix 23). The resulting filtered variant files were further filtered by selecting only the heterozygous SNV and indel variants that were shared by II:2, III:3 and IV:1 and absent in unaffected individual II:3. In addition, any identical variants that were present in an additional WES dataset from an unrelated patient affected with autosomal recessively inherited AI caused by a homozygous mutation in *MMP20*, processed using the same alignment and variant calling pipeline, were also filtered. This was to aid the exclusion of artefactual variants inherent to the library preparation and the sequencing platform.

The 26 genomic variants in a total of 25 genes that remained after filtering are listed in Appendix 24. In order to identify any genes for which an association with an oral phenotype in any species had been previously reported, a PubMatrix search was carried out (Table 4.2). Search terms used were “tooth”, “teeth”, “amelogenesis”, “enamel” and “ameloblast”. Using the search term “amelogenesis”, negative results were obtained for all 25 genes. The search term “enamel” returned positive results for two genes, epidermal growth factor receptor (*EGFR*; MIM *131550) and calcium channel, voltage-dependent, T type, alpha-1G subunit (*CACNA1G*; MIM *604065). *EGFR* expression is associated with tooth eruption, but the gene is also widely expressed during foetal development (Wise et al., 1992, Wakeling et al., 1998). Therefore, an *EGFR* mutation would be expected to result in more complex phenotypic effects than isolated AI as in the family presented here. In the case of *CACNA1G*, its

chromosomal position is in the vicinity of distal-less homeobox 3 (*DLX3*; MIM *600525) and *COL1A1* (MIM +120150), genes for which mutations result in tricho-odonto-osseous syndrome (MIM #190320; Section 1.8.1) and AI (MIM #104510) and osteogenesis imperfecta (MIM PS166200) respectively (Pope et al., 1985, Price et al., 1998). The reference identified by the PubMatrix search detailed an individual with an enamel phenotype with a deletion of chromosomal segments 17q21.33-q22 and therefore also of *DLX3*, *COL1A1* and *CACNA1G* (Harbuz et al., 2013). Therefore, it seems likely that deletion of *DLX3* or *COL1A1* is the cause of the enamel phenotype in this case and not deletion of *CACNA1G*. The search terms “teeth” and/or “tooth” returned positive results for the two genes already mentioned above, as well as protein-tyrosine phosphate, receptor-type, F polypeptide-interacting protein alpha-1 (*PPFIA1*; MIM *611054). Similarly, the reference identified by PubMatrix for *PPFIA1* reflects the gene’s chromosomal position near to the fibroblast growth factor (*FGF3*; MIM *164950) gene, in which mutations cause AR occulo-oto-dental syndrome (MIM #610706) (Gregory-Evans et al., 2007, Tekin et al., 2007). For comparison, *AMTN* was also included in the PubMatrix search and resulted in 42 and 54 hits for the search terms “enamel” and “teeth” and/or “tooth” respectively.

Gene	Predicted effect on the protein	CADD v1.3[a]	Pathogenicity scores[b]	Human phenotype[c]	Mouse model[d]	Pubmatrix search score[e]
AMTN	heterozygous deletion of exons 3-6, in-frame deletion of 92 amino acids	N/A	N/A	Not reported	Enamel defects	144
NAT8	heterozygous deletion of exon 2, containing the entire coding region, deletion of 227 amino acids	N/A	N/A	Not reported	Not reported	0
HECTD3	p.R583G	32	3/3	Not reported	Not reported	0
ADAM15	p.R288C	32	3/3	Not reported	No tooth phenotype reported	0
C20orf141	p.L84R	28.5	3/3	Not reported	Not reported	0
CLRN3	p.P126L	27.1	3/3	Not reported	Not reported	0
UBA7	p.I849V	26.4	2/3	Not reported	No tooth phenotype reported	0
ZNF563	p.G345C	26.0	2/3	Not reported	Not reported	0
EGFR	p.M908T	25.5	3/3	Neonatal inflammatory skin and bowel disease 2	No tooth phenotype reported; delayed ossification	70
IGSF9	p.V735M	25.1	2/3	Not reported	No tooth phenotype reported	0
CDH20	p.V722A	24.9	3/3	Not reported	Not reported	0
MLXIPL	p.Q68H	23.9	2/3	Not reported	Not reported	0
POU2F1	p.T138M	23.8	2/3	Not reported	Not reported	0
SFPQ	p.R663C	23.4	2/3	Not reported	Not reported	0
MOCOS	p.S361Ffs*18	23.2	N/S	Not reported	Not reported	0
LZTS3	p.K612T	23.3	2/3	Not reported	Not reported	0
LZTS3	p.K612E	23.1	2/3	Not reported	Not reported	0
BRINP3	p.T176M	22.9	1/3	Not reported	Not reported	0
MDGA1	p.A952T	22.9	1/3	Not reported	Not reported	0
ERICH2	p.P59L	22.8	1/3	Not reported	Not reported	0
ZNF717	p.A195V	22.6	1/3	Not reported	Not reported	0
SAMD14	p.E59_S64del	22.1	N/S	Not reported	Not reported	0
CACNA1G	p.A1582T	22.0	0/3	Spinocerebellar ataxia 42	No tooth phenotype reported	3

Gene	Predicted effect on the protein	CADD v1.3[a]	Pathogenicity scores[b]	Human phenotype[c]	Mouse model[d]	Pubmatrix search score[e]
SYTL2	p.E1305K	21.6	1/3	Not reported	Not reported	0
CST8	p.S41P	21.5	3/3	Not reported	Not reported	0
AMPD1	p.Q189H	18.62	1/3	Myopathy due to myoadenylate deaminase deficiency	Not reported	0
SH3BGRL2	p.V14E	17.56	1/3	Not reported	Not reported	0
PPFIA1	p.S224R	15.91	0/3	Not reported	Not reported	1

Table 4.1: Assessment of CNVs, SNVs and indels that remain after filtering AI-154 WES data.

List of the two heterozygous CNVs and twenty-six SNVs and indels identified by WES that remained after selecting only those shared by II:2, III:3 and IV:1, but not present in II:3, and excluding variants with a MAF of 0.1% or more in dbSNP142, EVS or ExAC, variants that are homozygous and those located on the X chromosome. Assessment included CADD v1.3 score[a], assessment of pathogenicity scores[b], known human phenotypes[c], known mouse phenotypes[d] and a score based on a Pubmatrix search[e].

[a] The list is restricted to variants scoring 15 or more when scored with CADD v1.3.

[b] Pathogenicity scores included were generated using Provean, SIFT and PolyPhen2. Scores were classified as pathogenic if they were not classified as neutral, tolerated or benign by the programs (Sections 2.9.4.1 and 2.9.4.2). The fraction here denotes the number of pathogenic scores obtained using the three different programs.

[c] A human phenotype was recorded if it was specified in OMIM (Section 2.9.1).

[d] A mouse phenotype was recorded if a mouse model was specified to exist in the gene's OMIM page. A dental phenotype was recorded if the teeth were reported to be affected in any way in the journal articles concerning the generation and characterisation of the phenotype of the mouse model(s) referenced on OMIM. Any phenotype reported for other mineralised tissues was also recorded.

[e] A Pubmatrix search (Section 2.9.3.2) was conducted with the search terms, "tooth", "teeth", "amelogenesis", "enamel" and "ameloblast". The number of hits were recorded as a combined score.

All pathogenicity scores were obtained, and the Pubmatrix search was conducted, on 4th April 2016.

4.2.6 CNV analysis using ExomeDepth of additional family members

For both IV:1 and II:3, CNVs were identified using ExomeDepth as before (Section 2.4.6). The sequencing batch included nine other unrelated exomes that were used as controls for ExomeDepth analysis.

Analysis of ExomeDepth data for IV:1 revealed a total of twenty-six CNVs not identified as a common CNV (Conrad et al., 2010), including the same heterozygous deletion spanning *AMTN* exons 3 to 6 identified in III:3 (Table 4.2). Analysis of ExomeDepth data for unaffected individual II:3 revealed a total of twenty-three CNVs not identified as a common CNV. The heterozygous deletion spanning *AMTN* exons 3 to 6 identified in individuals III:3 and IV:1 was not identified in II:3.

Sample	Type	Span (GRCh37)	Bayes Factor	Reads		
				Expected	Observed	Ratio
III:3	Deletion	4:71388473 -71394475	13.6	344	185	0.538
IV:1	Deletion	4:71388473 -71394475	24.4	401	208	0.519

Table 4.2: ExomeDepth output detailing the *AMTN* deletion identified in individuals III:3 and IV:1.

The Bayes Factor is the $\log_{10}$ of the likelihood ratio of data for the CNV call divided by that of the normal copy number call, i.e. higher scores suggest that the CNV is more like to be real.
The expected number of reads is based on the reads identified in the batch of control exomes.

Selection of the CNV calls shared by individuals III:3 and IV:1 but absent from II:3 resulted in two calls remaining after filtering (Appendix 25). Filtering did not take into account the CNV calls from individual II:2 due to the low number of controls available. One of the CNVs that remained after filtering was the heterozygous deletion of *AMTN* exons 3-6 and the other was a heterozygous deletion of exon 2 of the N-acetyltransferase 8 (GCN5-related, putative) gene (*NAT8;* OMIM *606716). This deletion is predicted to remove the entire coding sequence. However a PubMatrix search for *NAT8*, using the search terms

“tooth”, “teeth”, “amelogenesis”, “enamel” and “ameloblast”, did not identify any articles (Table 4.2). By comparison, the same search returned a total of 214 positive results for *AMTN*.

Graphical representations of ExomeDepth CNV calls in the region spanning chr4:71,339,000-71,480,000, which includes *AMTN* and its flanking genes mucin 7, secreted (*MUC7*) and *AMBN*, were plotted for each family member (Section 2.4.6) and are represented in Figure 4.5.

4.2.7 Screening of *AMTN* in a panel of autosomal dominant AI patients

In an attempt to identify additional *AMTN* mutations, all coding exons and flanking introns of *AMTN* in DNA samples obtained from an additional thirty-five families with suspected AD AI were sequenced (Appendix 26). No SNVs or small indels were identified. The cohort was also screened for the deletion identified in family AI-154, using the primer pair designed to amplify introns 2 to 6, but no other family was identified that carried the same deletion. Subsequently, WES libraries were prepared from genomic DNA, using the 200ng protocol (Section 2.4.1.2), from twelve of these families. Analysis of SNVs, indels and CNVs was undertaken as previously described but no mutations in *AMTN* were identified.

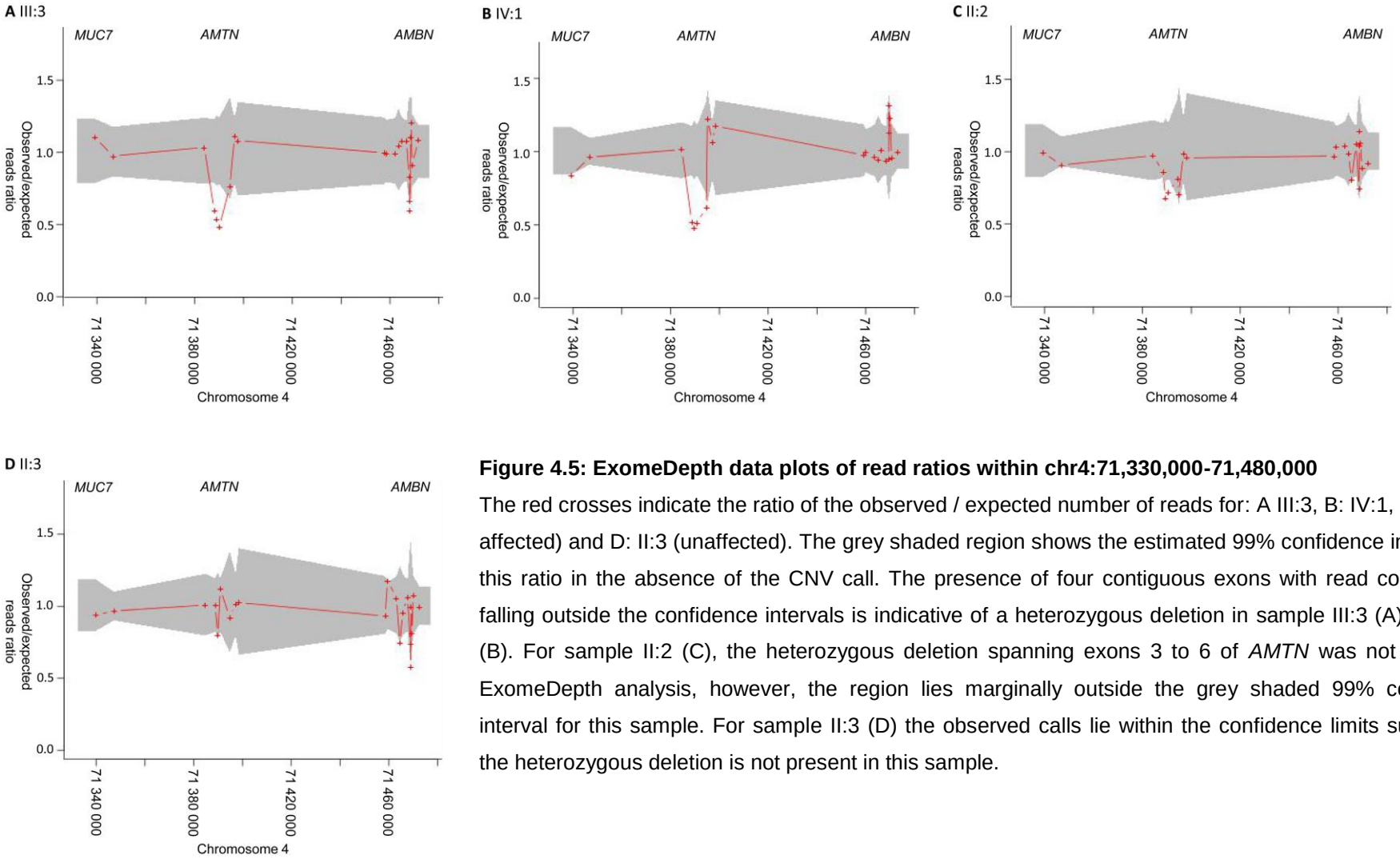


Figure 4.5: ExomeDepth data plots of read ratios within chr4:71,330,000-71,480,000

The red crosses indicate the ratio of the observed / expected number of reads for: A III:3, B: IV:1, C: II:2 (all affected) and D: II:3 (unaffected). The grey shaded region shows the estimated 99% confidence interval for this ratio in the absence of the CNV call. The presence of four contiguous exons with read count ratios falling outside the confidence intervals is indicative of a heterozygous deletion in sample III:3 (A) and IV:1 (B). For sample II:2 (C), the heterozygous deletion spanning exons 3 to 6 of *AMTN* was not found by ExomeDepth analysis, however, the region lies marginally outside the grey shaded 99% confidence interval for this sample. For sample II:3 (D) the observed calls lie within the confidence limits suggesting the heterozygous deletion is not present in this sample.

4.2.8 Investigation of the enamel phenotype

Based on the genetic analysis of AI-154, the most likely cause of the phenotype was a large deletion in *AMTN*. Localisation of murine AMTN within the basal lamina of maturation stage ameloblasts and at the internal basal lamina within the junctional epithelium of erupted rodent teeth had been previously reported (Iwasaki et al., 2005, Moffatt et al., 2006). Therefore, the family under investigation was revisited to check for any notable periodontal clinical phenotype that might have been overlooked previously but none was identified.

In order to determine whether the phenotype in the AI-154 affected individuals was comparable with the *Amtn*$^{-/-}$ rodent model, two deciduous incisor teeth from IV:1 were obtained after natural exfoliation. Photographs of the exfoliated teeth can be seen in Appendix 27. In order to investigate the phenotype, μCT, SEM and EDX (Section 2.13) were used to compare the teeth to two age and cuspid matched healthy deciduous human teeth obtained from the Skeletal Tissues Research Bank (School of Dentistry, University of Leeds) with ethical approval and patient consent.

4.2.8.1 Micro Computerised Tomography (micro CT)

In order to reveal the underlying structure of the teeth and to calculate the mineral density of the enamel in the teeth from IV:1, high resolution μCT was carried out (Section 2.13.1). Two control teeth (A and B) were also scanned for comparison with the teeth from IV:1 (C and D). Representative longitudinal sections through each tooth are presented in Figure 4.6. One of the teeth (tooth D) had undergone some restoration.

Mean enamel mineral density was calculated (Section 2.13.1) for each tooth by comparison to three HA standards of known mineral density. The three standards were scanned alongside each tooth and used to calibrate the mineral density measurement for each tooth separately, based on the readings of the standard obtained only during the same scan. For the tooth that had undergone restoration (Tooth D), a density threshold was established to identify the regions

of restoration. These regions were then excluded by the automated protocol run by the ImageJ software, which is used to identify the enamel regions using the Trainable Weka Segmentation plugin (Section 2.13.1). The regions in each section identified by the plugin as enamel were checked manually to ensure that the training had been successful. Selected regions were altered manually only if necessary, although this was rare. Mean enamel density across the whole tooth was then calculated and is presented in graphical form in Figure 4.6.

The enamel mineral densities of teeth C and D were calculated as 2.05 g/cm^3 +/-0.02 g/cm^3 (2x standard error of the mean) and 2.16 g/cm^3 +/-0.02 g/cm^3 respectively, compared to 2.66 g/cm^3 +/-0.02 g/cm^3 and 2.72 +/-0.06 g/cm^3 for the two control teeth, A and B. For reference, a range of 2.69-2.92 g/cm^3 has been previously reported for deciduous enamel (Wong et al., 2004).

The values obtained equate to a 18.8-24.6% reduction in enamel mineral density for the teeth from Individual IV:1 compared to teeth from the controls. While both teeth C and D had lower enamel mineral density measurements than control teeth A and B, statistical analysis was not appropriate due to the small number of non-independent samples investigated and the nested nature of the data.

High resolution-CT scanning (Figure 4.6) of control teeth A and B and affected teeth C and D from IV:I, showed that teeth C and D exhibited reduced enamel mineral density in comparison to control teeth A and B. Enamel thickness also appeared to be reduced compared to the control tooth though this was difficult to confirm due to the possibility of post-eruptive tissue loss and the normal variation of the enamel thickness from the cemento-enamel junction to the cuspal edge (Gillings and Buonocore, 1961).

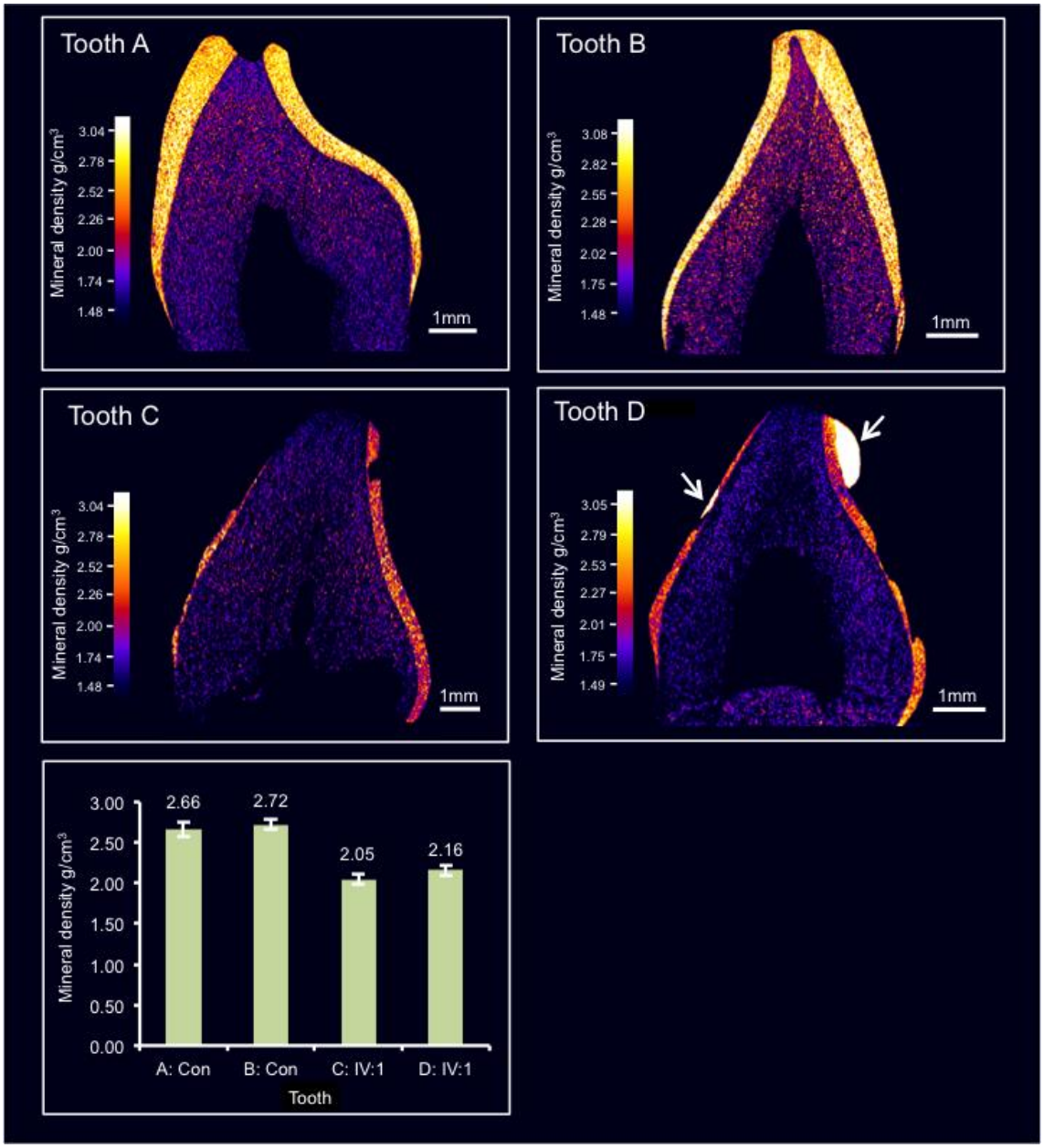


Figure 4.6: High resolution X-ray CT analysis of exfoliated teeth from two control individuals and from individual IV:1.

Typical CT sections through the teeth are presented using false colour calibrated with respect to mineral density to generate mineral density maps. Mean enamel mineral density for each tooth is also shown graphically, with error bars representing twice the standard error of the mean. The teeth from the control individuals (A and B) exhibit an enamel layer apparently normal in structure and density. Teeth C and D from IV:1 exhibit an obvious enamel covering though it is thinner, chipped, absent in places and is significantly reduced in mineral density compared with teeth A and B ($P < 0.05$). Note that tooth D has undergone restoration (arrows).

Figure reproduced with permission from Smith et al. (2016a).

4.2.8.2 Scanning Electron Microscopy (SEM)

In order to examine the enamel phenotype further, one of the teeth from individual IV:1 (Tooth D) was prepared for SEM (Section 2.13.2). A control tooth (Tooth A) was also prepared for comparison. Photographs of the sections from the tooth from IV:1 (Tooth D) as well as a control tooth (Tooth A) can be seen in Figure 4.7.

SEM (Section 2.13.3) revealed that the control tooth (Figure 4.7, A and B) exhibited the prismatic structure characteristic of normal human enamel. The tooth from individual IV:1 (Tooth D), also exhibited areas of prismatic enamel though the fine detail of the ultrastructure was obscured by overlying material, giving a “smooth” appearance (Figure 4.7C and D). The enamel also displayed distinct regions where the enamel structure was particularly disturbed; the prismatic structure having a “gnarled” appearance (Figure 4.7E and F). These regions spanned the entire enamel and were not confined to the outermost enamel layers. In some areas of the tooth, enamel was missing, perhaps indicative of post-eruptive failure under masticatory stress**.**

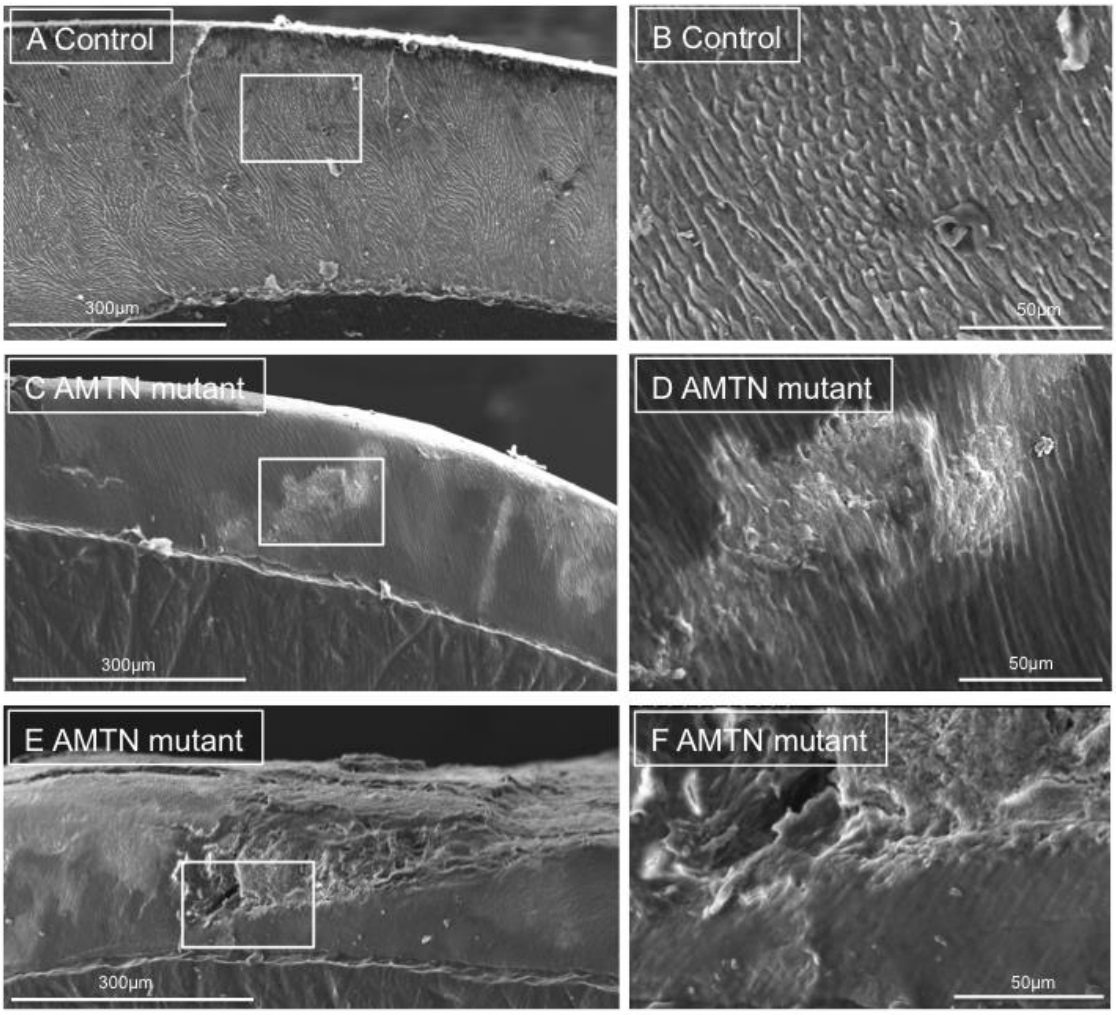


Figure 4.7: SEM of representative exfoliated teeth.

A and B: SEM of tooth A from the control individual. C-F: SEM of tooth D from individual IV:1. Boxed regions on pictures A, C and E reflect the boundaries of the photographs taken at higher power in these regions, labelled B, D and F. Tooth A exhibits normal, typical enamel architecture comprising prisms (rods) of individual enamel crystallites. Tooth D exhibits both regions of relatively normal enamel and disturbed structure. The cross sectional surface of Tooth D has a "smooth" appearance that may reflect the presence of organic material.

Figure reproduced with permission from Smith et al. (2016a).

4.2.8.3 Energy-dispersive X-ray spectroscopy (EDX)

In order to determine whether the enamel of the tooth from IV:1 (Tooth D) differed in its elemental composition to that of the control tooth (Tooth A), EDX was undertaken on the tooth sections prepared for SEM.

Since the EDX software (Section 2.13.4) utilises the peak to background atomic number, absorbance and fluorescence (PB-ZAF) correction (Small et al., 1978,

Statham and Pawley, 1978, Heckel and Jugelt, 1984, Trincavelli and Van Grieken, 1994), accurate elemental analysis does not require that results are compared with those obtained for standards. However, in order to verify that the detector and the software could detect the elements normally present in tooth enamel, a standard of HA mineral (Himed) was first analysed. Oxygen, calcium, carbon and phosphorus were detected, suggesting that the standard may also contain a carbonated form of HA or that another carbon containing compound may be present in addition to the HA. The atomic mass percent (AM%) calcium:phosphorus ratio of the standard was determined to be 2.11. The ratio for pure HA was calculated as 2.15, based on an atomic weight for HA of 1004 and taking into account that EDX is unable to analyse the percentage content due to hydrogen. Therefore the result of 2.11 appeared to be close to that expected for HA.

The limit of detection for EDX analysis is determined by the bremsstrahlung background radiation that is used as a calibration point in the PB-ZAF method. For the equipment and software used for detection and analysis in this study, this is specified as around 0.1% mass concentration (Bruker Quantax User Manual) which would equate to approximately 0.05-0.15% for elements with relative atomic mass of between 12 (C) and 40 (Ca), hence results are presented to one decimal place. In addition, the percentage error for measurements has been calculated by the manufacturer as between 3-5% (Bruker Quantax User Manual).

For the control tooth (Tooth A), twenty-three readings were taken throughout the enamel region and the elemental percentage contributions were recorded (Figure 4.9C). The control tooth A consisted primarily of elements expected to be found in enamel, which is predominantly comprised of a substituted calcium hydroxyapatite (Robinson et al., 1971, Robinson et al., 2000). Atomic mass percentage contributions showed that tooth A enamel consisted of 57.2% oxygen, 19.9% calcium, 12.3% carbon and 10.1% phosphorus, with the remaining 0.45% being sodium (Table 4.1). A representative EDX spectrum for control tooth A is presented in Figure 4.8A.

SEM analysis of tooth D had revealed several distinct regions of enamel with diverse architectures (Section 4.2.8.2). Therefore, EDX readings were taken from four distinct regions in order to try to identify if these regions of diverse architecture also had distinctive elemental percentage contributions. Separate analysis of each of these regions also avoided aggregating readings from regions with obvious structural differences, leading to the reporting of averaged results that may not have been representative. Again, the areas of enamel with restoration were not included in the analysis, but separate analysis of the spectra resulting from the restored region confirmed that it presented a very different spectral profile to the enamel which would be immediately recognisable as it contained elements such as Al, Si, Zr and Ba (Figure 4.8C).

Analysis of these four areas of tooth D enamel (individual IV:1) showed that they did consist of varying elemental percentage content, although areas D1 and D2 were similar. A representative EDX spectrum for tooth D area 2 is presented in Figure 4.8B. All areas of tooth D had higher carbon and lower oxygen percentage content than tooth A (Figure 4.9 and Table 4.3). Three of the four enamel areas analysed for tooth D also had a lower phosphorus percentage content than tooth A. Furthermore, three areas of tooth D enamel contained nitrogen (Area 1: 11.1%, area 2: 11.6% and area 4: 7.2%) which was not detected in enamel from tooth A.

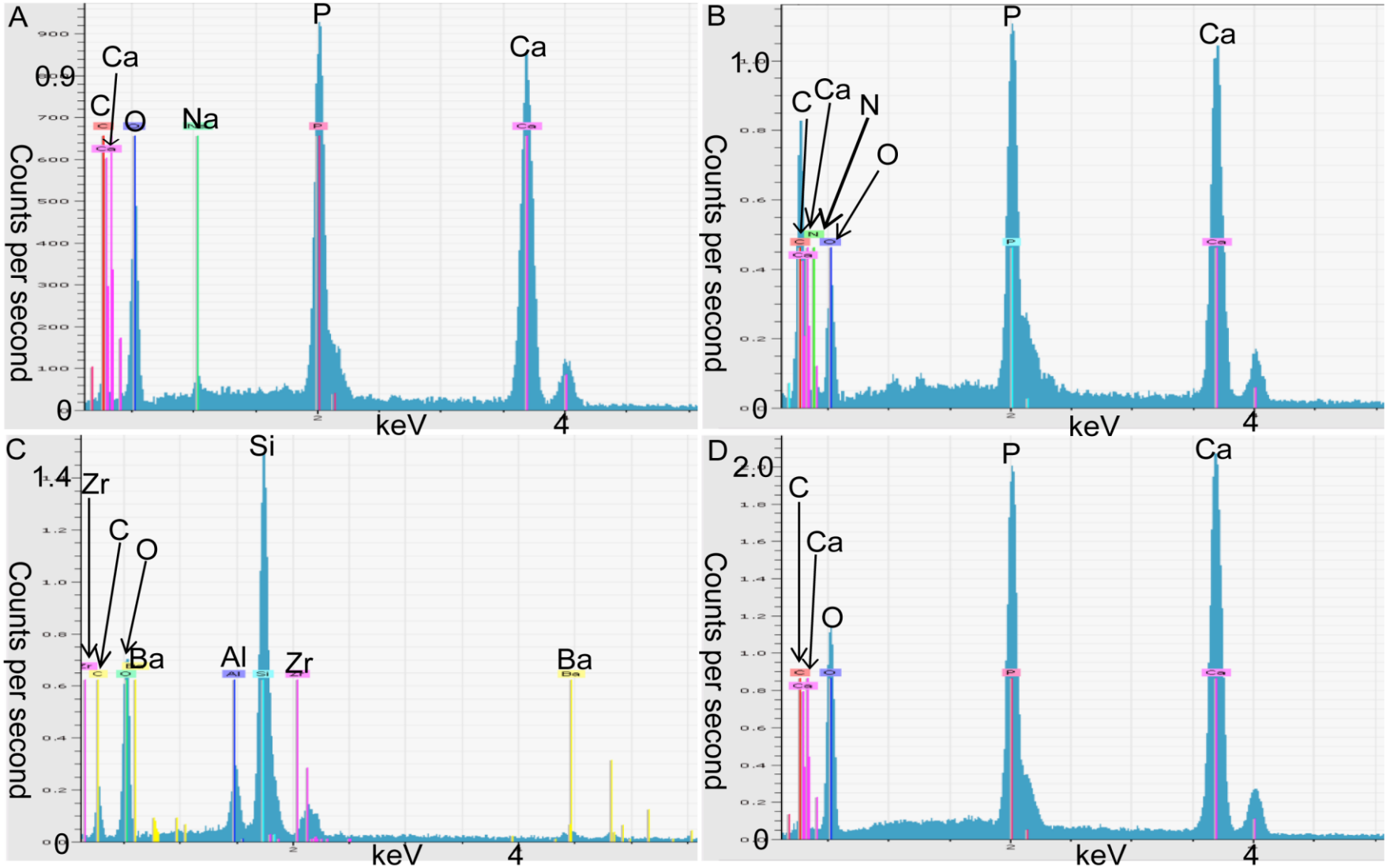

A
Ca
C
O
Na
P
Ca
Counts per second
keV
4
B
P
Ca
C
Ca
N
O
Counts per second
keV
4
C
Si
Zr
C
O
Ba
Al
Zr
Ba
Counts per second
keV
4
D
P
Ca
C
Ca
O
Counts per second
keV
4

Figure 4.8: Representative EDX spectra.

See previous page. A: Wild-type enamel (control tooth A) contained the elements typically found in HA substituted with CO_3^{2-} and Na. B: Enamel affected by the heterozygous deletion of *AMTN* exons 3-6 (tooth D from individual IV:1) shows N is also present in addition to the elements typically found in enamel (Area D - 2 in Table 4.3 and Figure 4.9). C: Analysis of the restored region of tooth D confirmed that it produced a distinct spectrum from that of typical enamel, since it contained Al, Zr, Si and Ba.

D: Analysis of a standard of HA mineral showed a similar spectral profile to the enamel of the wild-type tooth (control tooth A, panel A).

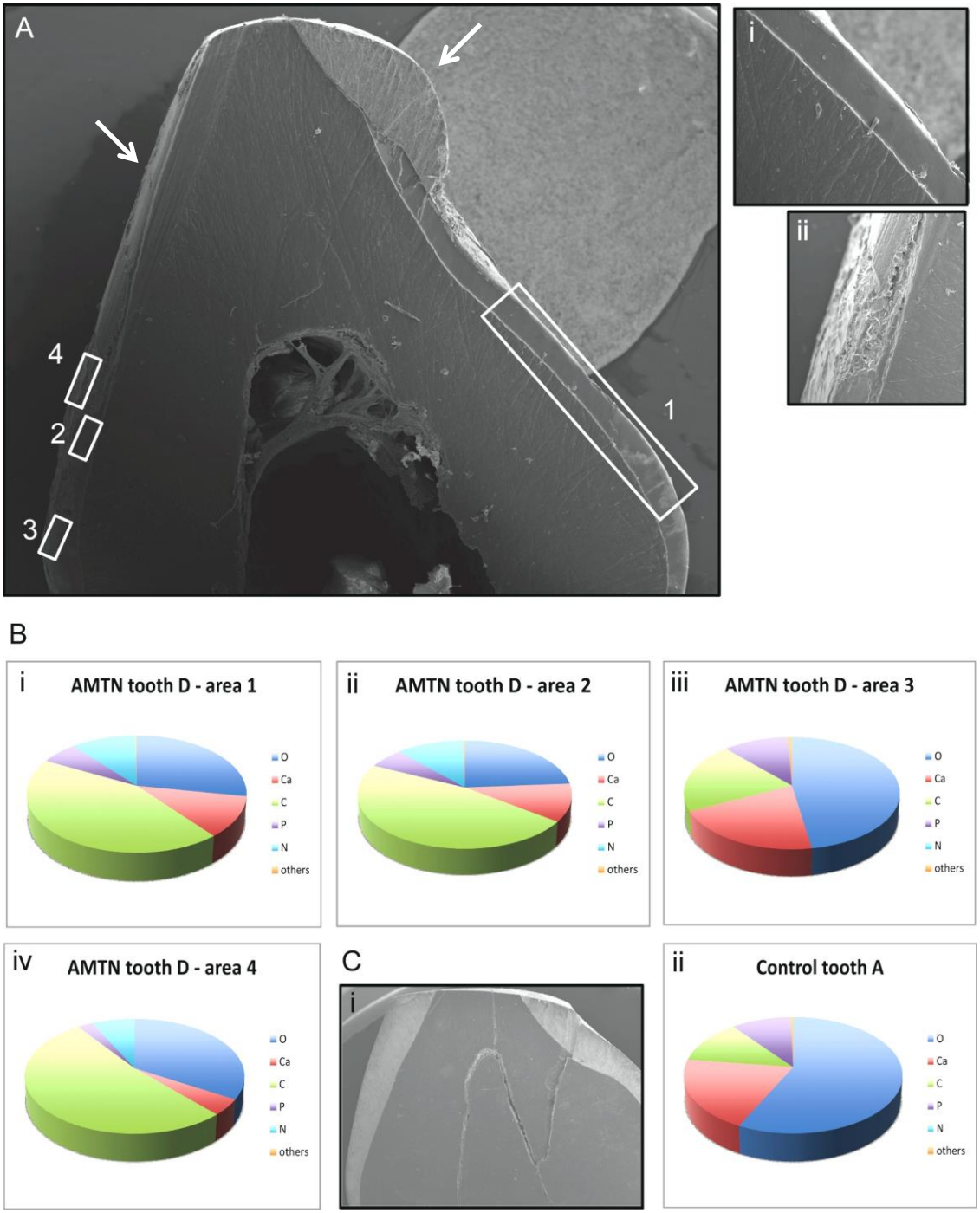


Figure 4.9: EDX elemental analysis of tooth D from individual IV and tooth A from a control individual.

A: Measurements for tooth D were taken from four separate areas of enamel, highlighted by the white boxes. White arrows indicate the areas of enamel that had undergone restoration and therefore where elemental analysis was specifically not undertaken. Panels i and ii detail the appearance of enamel areas 1 and 4 for tooth D.

B: Elemental analysis results are plotted as pie charts, with the percentage atomic contribution of each element to the enamel displayed in separate charts for each region. The number of

measurements was as follows: i: Area 1: N=30, ii: Area 2: N=11, iii: Area 3: N=10, iv: Area 4: N=9.

C For comparison, control tooth A was also analysed at points throughout the area of enamel shown in panel i. Panel ii shows elemental analysis results for control tooth A plotted as before, with N=23 measurements.

In all cases, the “others” section for each elemental analysis pie chart refers to the elements Na and/or Mg that were found to contribute <1% collectively to the enamel.

Detailed percentage contributions of each element to each of the tooth regions analysed can be seen in Table 4.3.

Standard / tooth and area analysed	Composition (Atomic Mass %[a])						Number of readings
	Oxygen	Calcium	Carbon	Phosphorus	Nitrogen	Other[b] (Specified)	
HA standard	59.9 ± 0.6	23.1 ± 0.4	6.02 ± 0.3	10.95 ± 0.7	0.0 ± 0.0	N/A	3
A	57.2 ± 0.5	19.9 ± 0.4	12.4 ± 0.9	10.1 ± 0.3	0.0 ± 0.0	0.5± 0.1 (Sodium)	23
D - 1	28.0 ± 2.3	11.3 ± 1.0	43.6 ± 2.8	6.2 ± 0.5	11.1 ± 1.9	0.2 ± 0.1 (Sodium)	30
D - 2	23.8 ± 2.4	11.3 ± 1.5	46.9 ± 4.9	6.1 ± 1.0	11.6 ± 2.6	0.3 ± 0.2 (Sodium)	11
D - 3	47.4 ± 1.1	20.3 ± 0.4	20.4 ± 1.7	11.2 ± 0.6	0.0 ± 0.0	0.7 ± 0.2 (Sodium)	10
D - 4	33.6 ± 1.6	4.5 ± 1.9	52.1 ± 3.1	2.6 ± 1.0	7.2 ± 3.6	N/A	9

Table 4.3: EDX elemental analysis of an HA standard, tooth A from a control individual and tooth D from individual IV:1.

Values shown are mean values calculated from the number of measurements specified for each area. The error was calculated as 2 x the standard error of the mean. This data has been displayed graphically and in relation to the areas measured in Figure 4.9.

[a] Values may not always sum to 100% since values are shown to 1 decimal place.

[b] For tooth D, areas 1 and 3, magnesium was also variably present, but the mean composition was less than 0.1% and so it is not included in the table above.

4.3 Discussion

A family from Costa Rica segregating AD hypomineralised AI was identified. DNA from three affected and one unaffected individuals was subjected to WES. Analysis of rare, heterozygous, functional variants that segregated with the disease phenotype *in silico* identified a total of twenty-six SNVs and small indels, none of which were in known or candidate AI genes. Analysis of rare CNVs using ExomeDepth identified two heterozygous deletions, one of which was predicted to include exons 3 to 6 of the AI candidate gene *AMTN*. Further investigation confirmed this to be an 8,678bp deletion encompassing *AMTN* exons 3-6 and flanking intronic sequence. PCR and electrophoresis showed that the deletion segregated with the disease phenotype in all available family members.

Attempts to study the effect of the deletion on leukocyte RNA splicing failed due to the tooth specific expression of *AMTN*. An attempt to identify additional mutations in *AMTN* by screening 35 additional AI families revealed no further mutations in the coding exons, or any recurrence of the same large deletion identified in AI-154.

Phenotypic analysis of the enamel of affected teeth using microCT revealed that it was less mineralised and appeared to be thinner in comparison to the enamel of control teeth. SEM revealed that the enamel of affected teeth contained regions with the prismatic architecture typical of enamel but that this appeared to be obscured by overlying material. There were also regions of more disturbed structure that spanned the entire enamel layer.

Elemental analysis showed that the affected teeth contained areas of enamel with elemental percentage contributions that were different, both in comparison to different enamel areas within the same tooth and in comparison to the enamel of the control. In general, compared to the control enamel, there were higher carbon and lower oxygen percentage contributions in affected enamel and, in three of the four regions analysed, nitrogen was present, but this was below the limit of detection for the control.

The research presented in this chapter is published in Smith et al. (2016a).

4.3.1 Whole exome sequencing analysis

WES was used to try to identify the genetic cause of AI in this family since all previously published AI-causing mutations reside in regions of DNA captured by WES (http://dna2.leeds.ac.uk/LOVD/). Although it is likely that AI-causing mutations in regions not targetted by WES exist (Maurano et al., 2012, Poulter et al., 2015), interpretation of non-coding variants is currently difficult.

This family was particularly informative for mutation discovery, since DNA from three affected individuals spanning three non-consanguineous generations was available. The pedigree, which demonstrated male-to-male transmission, also suggested that the disease had an AD pattern of inheritance. This knowledge informed the filtering strategy used, as it allowed the list of variants to be filtered for single shared heterozygous variants.

Filtering of WES data from three affected individuals and one unaffected individual based on an AD mode of inheritance resulted in twenty-six variants remaining with CADD scores over 15. Whilst it was possible that one of these variants could be responsible for the AI phenotype, none were implicated for further analysis by a known role in amelogenesis or by the existence of a mouse model with a tooth phenotype. WES of individual IV:2 would be expected to approximately halve the number of variants segregating with the disease phenotype and would be less labour intensive and potentially also cheaper than individual segregation analysis of each variant by Sanger sequencing. However, Sanger sequencing would also confirm the presence of each variant since alignment and variant calling of WES data can result in spurious variant calls, although the inclusion of the unaffected individual II:3 may have already removed such variants from the filtered list.

CNV analysis was performed separately since genetic variation encompasses changes that exist over a broad size range, but the standard variant calling pipeline using GATK is only capable of detecting smaller variations such as

SNVs and small indels. Studies of the degree of variation of genomes from reference sequence have shown that whilst differences of around 0.1% are due to SNVs, around 1.2% are due to indels and CNVs (Pang et al., 2010). Studies of the span of CNV regions have found that they equate to around 12% of the human genome (Redon et al., 2006) and are associated with changes in gene expression (Stranger et al., 2007). This suggests that such variation is also likely to encompass a significant proportion of the variants that cause disease, and that analysis of CNVs should be undertaken alongside SNV and indel variant analysis. Analysis of eleven WES data sets generated from patients with AI, for which an SNV or indel mutation in an AI candidate gene had not been identified, revealed that two carried CNVs that deleted one or more exons of an AI candidate gene. This equates to 18.2% and suggests that CNVs may be the cause of a substantial proportion of AI cases without a SNV or indel in a known AI gene.

Many algorithms exist that utilise meta-analysis of WES datasets to detect variation of a larger scale such as CNVs, for example, CoNIFER (Krumm et al., 2012), eXome Hidden Markov Model (Krumm et al., 2012), ExomeDepth (Plagnol et al., 2012), FishingCNV (Shi and Majewski, 2013) and CopyWriteR (Kuilman et al., 2015). Such algorithms compare read depths between different samples for the same targetted regions and use the information to infer changes in copy number. Despite the number of tools to carry out such analysis, the accurate detection of CNVs using WES data alone remains problematic. In comparison to WGS data, where paired end mapping (Korbel et al., 2007), split read mapping (Ye et al., 2009), assembly based- (Ye et al., 2009) or combined methodologies can also be used to identify CNVs, analysis of WES data is mostly limited to methodologies based on read depth analysis. For WES data, read depth over individual exons is affected by PCR bias and capture efficiency, whereas the read depth profile is more homogeneous for WGS data, meaning that it is easier to identify CNVs (Meynert et al., 2014). Successful identification of a true CNV using read depth analysis requires that the regions containing the CNV are normally well covered by WES and are not commonly variable. In addition, the nature of the exome, which represents a series of small, non-contiguous regions that are unevenly distributed across the

genome, means that analysis of WES data often fails to accurately map the breakpoints of CNVs, as statistically, they are more likely to reside within uncovered deep intronic and intergenic regions, as for the deletion identified here. Assessments of the accuracy of WES-based CNV discovery have shown that algorithms often produce erroneous calls (Samarakoon et al., 2014, Tan et al., 2014) and may miss true CNVs, especially those that are small, since each exon effectively represents a single data point (Krumm et al., 2012, de Ligt et al., 2013). Therefore, WGS would be a better choice of method for detection of CNVs and their breakpoints, but the additional cost and burdens of data volume and analysis associated with WGS are prohibitive to its use at present.

In this study, the ExomeDepth R package (Plagnol et al., 2012) was used to screen for CNVs from WES data. ExomeDepth builds a reference data set of read count data from, ideally ten, control samples sequenced in the same batch as the test sample. It then compares the test sample with the reference set to generate a likelihood for each copy number scenario, i.e. deletion, diploid or duplication. ExomeDepth has been reported to perform well in identifying clinically relevant CNVs (de Ligt et al., 2013) but has reported rates of true CNV detection ranging from 16-75% (Plagnol et al., 2012, de Ligt et al., 2013, Guo et al., 2013) depending upon the read depth and correlation of the samples analysed and the method of confirmation for comparison used. It is therefore difficult to assess its effectiveness.

The variability of CNV detection using ExomeDepth is apparent for the data presented here in that the heterozygous deletion of *AMTN* exons 3-6 was not detected from analysis of WES data from individual II:2. This may be due to the smaller number of reference sequences used for analysis of II:2 data (five) in comparison to III:3 (eleven) and IV:2 (nine) data. Therefore, if analysis using ExomeDepth is to be undertaken, a minimum of ten unrelated individuals should be sequenced per batch to increase the chances of the identification of all relevant CNVs. In the case of the heterozygous deletion identified here, its position within such a well known AI candidate gene as *AMTN* (Santos et al., 2007, Gasse et al., 2012, Abbarin et al., 2015) meant that it was prioritised for investigation. However, it is still possible that the remaining variants that

segregate with the phenotype in this family are contributing to the phenotype. WES of IV:2 would reduce the number of remaining variants by 50% on average, and further recruitment of additional family members may also help to exclude those present in unaffected individuals, although this may not be possible. Prior knowledge of the role of *Amtn* in rodent models and its status as an AI candidate gene may therefore have biased the decision to follow up on this variant above others, which is an inherent weakness in this study.

4.3.2 Amelotin (*AMTN*) is a candidate gene for non-syndromic amelogenesis imperfecta

Amelotin (*Amtn*) was first cloned in rodents during studies to identify proteins secreted by the enamel organ (Iwasaki et al., 2005, Moffatt et al., 2006). Since then, *AMTN* (MIM *610912) has become a strong candidate gene for AI due to its chromosomal position, structure, evolution, protein localisation and *in vitro* activity, as well as due to evidence from mouse models.

AMTN is found within the SCPP conserved enamel gene cluster and is thought to have evolved with the EMP genes from a common ancestral gene (Section 1.2). The SCPP EMP cluster includes the genes encoding enamelin (Section 1.9.3) and ameloblastin (Section 1.9.2), for which mutations are already known to cause AI (Rajpar et al., 2001, Poulter et al., 2014c). However AMTN is not regarded as an EMP due to its absence from the enamel matrix and lack of expression during the secretory stage (Sire et al., 2007). Pseudogenisation of *AMTN* in enamel-less species has meant that AMTN is regarded as an enamel-specific protein (Gasse et al., 2012). It has also been proposed that prismatic enamel evolved in mammals concurrently with AMTN expression becoming restricted to the maturation stage and with changes to *AMTN* exon boundaries and splicing (Gasse et al., 2012, Gasse et al., 2015a, Gasse et al., 2015b).

AMTN encodes a 209 amino-acid protein (NP_997722) rich in proline, leucine, threonine, glutamine, and glycine, with an N-terminal signal sequence that, once cleaved, gives rise to a mature 20.4kDa secreted protein (Moffatt et al., 2006, Ganss and Abbarin, 2014). AMTN has been shown to be restricted to the

basal lamina (the structure that links ameloblasts with the developing enamel) and to maturation, but not secretory, stage ameloblasts in rodents (Iwasaki et al., 2005, Moffatt et al., 2006, Dos Santos Neves et al., 2012, Somogyi-Ganss et al., 2012). AMTN has also been detected at the internal basal lamina within the junctional epithelium of erupted rodent teeth suggesting a possible function in dento-gingival attachment (Iwasaki et al., 2005, Moffatt et al., 2006).

The exact function of AMTN is unknown, although its localisation, experiments using AMTN peptides and observations from mouse models have inspired hypotheses. A functional study by Abbarin and colleagues found that AMTN promotes HA precipitation *in vitro*, leading the authors to suggest that AMTN may serve a critical role in the formation of the final compact aprismatic enamel surface layer during the maturation stage of amelogenesis (Abbarin et al., 2015). Consistent with these findings, both an *Amtn* overexpression mouse model in which AMTN was over-expressed under the control of an *Amel* promoter (p*Amel*:*Amtn*$^{+/+}$), and a knockout (*Amtn*$^{-/-}$) mouse model, have defective enamel (Lacruz et al., 2012a, Nakayama et al., 2015). The p*Amel*:*Amtn*$^{+/+}$ mouse model has brittle enamel that is thinner than wild-type (WT), with a defective surface layer (Lacruz et al., 2012a). By contrast, the *Amtn*$^{-/-}$ mouse model has mandibular incisors with a chalky appearance and rough, irregular surface enamel that is easily chipped away, though maxillary incisors and molars were unaffected, as were *Amtn*$^{+/-}$ mice (Nakayama et al., 2015).

These observations have made *AMTN* a strong candidate gene for involvement in human AI, as noted by a number of research groups (Santos et al., 2007, Gasse et al., 2012, Abbarin et al., 2015, Bartlett and Simmer, 2015). Núñez et al. speculated that AMTN loss of function mutations would result in AR AI (Nunez et al., 2015) but, at the start of this study, no human mutations had been reported.

4.3.3 The potential consequences of the heterozygous deletion of *AMTN* exons 3-6

AMTN consists of nine exons, of which eight (exons 2 to 9) are protein coding. The heterozygous deletion identified by ExomeDepth was confirmed by Sanger sequencing and found to be an 8,678bp deletion and 4bp insertion predicted to remove exons 3-6 of *AMTN*, c.54+1347_330+98delinsCTCA (NM_212557). No *AMTN* transcript could be amplified from patient blood, so it was not possible to determine the precise effect of the loss of exons 3-6 on the mRNA produced. One hypothesis is that the deletion leads to the production of a mature transcript containing exons 1 (5' UTR), 2, 7, 8 and 9. This is a strong possibility as the relevant splice consensus sequences remain intact and the introns flanking the deleted region are in phase zero, meaning that the deletion would leave any translated protein in frame (Kolkman and Stemmer, 2001). For this reason the transcript is unlikely to be subject to nonsense-mediated decay (NMD) (Isken and Maquat, 2008).

The deletion of *AMTN* exons 3-6 would delete an IPLT motif in exon 4 (Gasse et al., 2012), which is predicted to be an O-glycosylation site and is therefore likely to be important for AMTN function (Sire et al., 2007). Glycosylation plays an important role in influencing the physiochemical properties of proteins such as conformation, protection from proteolysis and interactions with other proteins and minerals (Brooks et al., 2008). Proteins encoded by known AI genes such as *KLK4*, *ENAM* and *AMBN* are all glycosylated at key stages of amelogenesis, suggesting correct post-translational modification of enamel proteins is required for enamel to form correctly (Kobayashi et al., 2007, Al-Hashimi et al., 2009, Yamakoshi et al., 2011).

Previous studies have found that *AMTN* exon 2 encodes the putative signal peptide to target the protein to the endoplasmic reticulum (ER) and exon 7 encodes a consensus SXE phosphorylation motif targeted by the casein kinase FAM20C (Moffatt et al., 2006, Gasse et al., 2012). Assuming that a mutant transcript and protein are produced, the deletion identified in the family described here would not remove either of these highly conserved regions

(Gasse et al., 2012). However, deletion of exons 3-6 would disrupt a conserved, amino terminal, IPV-like, tripeptide motif, LPQ, encoded by the last six bases of exon 2 and the first three bases of exon 3 respectively. This motif is thought to be responsible for the efficient trafficking of acidic proteins (the calculated pI of unmodified human AMTN is 5.29) out of the ER (Nam et al., 2014). The deletion identified in this family, if expressed, will replace the polar glutamine residue encoded by the first three bases of exon 3 with non polar glycine encoded by the first three bases of exon 7.

IPV-like motifs are found in many of the secretory calcium-binding phosphoproteins (SCPPs) and in other secreted proteins such as pro-hormones and those found in blood plasma. They are thought to function as a signal to, as yet unknown, receptor proteins that are responsible for the packaging of proteins into vesicles for transport from the ER to the Golgi. Analysis of the IPV-like motifs found in a selection of secreted proteins noted that they are positioned immediately carboxy to the signal peptide for secretion (Appendix 28). The consensus sequence for such a motif is hydrophobic / polar residue-proline-hydrophobic / polar residue (Nam et al., 2014). For the human secreted protein sequences analysed, proline was always present at position 2 of the motif and in only one case was a non-polar residue present at either of the flanking positions (glycine at position 1 of the motif for decorin (DCN; OMIM *125255)). More specifically, the motifs of the enamel proteins generally contain only hydrophobic residues at positions 1 and 3, but AMTN contains a polar glutamine residue at position 3. The variation of the residues used in the motifs for different secretory proteins may reflect targeting by multiple, as yet unknown receptors for ER export. Investigation of the residues present in the motif in AMTN in a variety of mammalian species found that while the residues at positions 1 and 2 are always leucine and proline respectively, the residue at position 3 varies (lysine, arginine, glutamine, leucine, methionine, threonine, alanine are present for the species analysed) and is sometimes positively charged (Appendix 29). This variation could suggest that substitution of the glutamine residue for glycine, as in the deletion identified in family AI-154, may be tolerated or may not, in isolation, completely abolish export of the protein from the ER. For both the human enamel and dentine/bone matrix proteins

encoded within gene clusters at 4q13.3 and 4q22.1 respectively and for AMELX, the motifs are interrupted by phase zero exon boundaries between the second and third residues, whereas for the other proteins studied, this is not the case (Appendix 28). It is unclear whether alternative splicing, such as is observed for AMELX (Brookes et al., 1995), could alter Golgi targeting for some functional reason in amelogenesis.

Mutations affecting the IPV motif of the dentine matrix protein, dentin sialophosphoprotein (DSPP, OMIM *125485), are known to result in dominant negative dentinogenesis imperfecta I (DGI-I, formerly DGI Shields type II; OMIM #125490) due to a failure to traffic the protein out of the ER and the associated retention of wild-type protein (Shields et al., 1973, von Marschall et al., 2012). Mutations identified in patients with DGI-I include those directly affecting the IPV residues at positions 16-18 (NM_014208; NP_055023), for example c.52G>T, p.V18F (Xiao et al., 2001) or those likely to result in the skipping of exon 3, such as the intronic mutation, c.52-3C>G (Kim et al., 2004). This mutation would be predicted to result in the substitution of the third residue in the motif, V, for the first residue encoded by exon 4, D. These *DSPP* mutations are similar to the *AMTN* exon 3-6 deletion identified in family AI-154 since they also result in substitution of the third residue in the IPV-like motif. Modelling of mutation of the third residue of the IPV-like motif for two other SCPPs, dentin matrix protein 1 (DMP1; OMIM *600980) and secreted phosphoprotein 1 (SPP1; OMIM *166490), as well as for the neuroendocrine secretory protein, chromogranin A (CHGA; OMIM *118910) also resulted in the retention of the mutant proteins within the ER (von Marschall et al., 2012). Mutation of the DMP1 IPV-like motif, LPV, at the second residue or at both the second and the third residues resulted in retention of the protein within the ER, although some secretion of the mutant protein into the media was detected (Liang et al., 2016). The authors suggested that this might reflect the existence of a transport mechanism independent of the IPV-like motif (Liang et al., 2016). This is possible, although the use of plasmid expression constructs with elevated expression profiles *ex vivo* may not accurately reflect what would occur *in vivo*.

The consequences of mutation of IPV-like domains are thought to be influenced

by the sequence of the rest of the protein (von Marschall et al., 2012, Nam et al., 2014, Liang et al., 2016). In the case of a *DSPP* mutation that delayed or prevented trafficking of the protein from the ER to the Golgi, the large number of acidic, negatively charged residues present within DSPP would interact with the high concentrations of Ca^{2+} found within the ER. This is thought to lead to aggregation (von Marschall et al., 2012) that may in turn lead to ER stress and autophagy. Unlike DSPP, AMTN does not contain large tracts of acidic residues. However it does contain an SXE motif that when phosphorylated, binds Ca^{2+} (Sorensen et al., 1995, Kawasaki et al., 2004). The deletion of *AMTN* exons 3-6 would leave this motif intact and therefore ER accumulation due to mutation of the IPV domain and aggregation of both wild-type and mutant protein due to aberrant Ca^{2+} binding via the SXE motif within the ER is a possible disease mechanism. However, it is likely that FAM20C acts to phosphorylate proteins only once they reach the Golgi, not in the ER. Therefore, it may be more likely that the protein is simply misfolded, forms aggregates and interacts with the wild-type protein resulting in sequestration of the wild-type protein and prevention of its normal function, thereby causing a dominant negative effect.

Overall, the deletion of a significant proportion of a protein or defective covalent modification, is likely to result in misfolding, leading to endoplasmic reticulum stress driven cell pathologies (Hetz et al., 2011). Rat *Amtn* transcript variants lacking exons 3 to 7 have been identified (Moffatt et al., 2006) suggesting that amelotin lacking the IPV-like tripeptide motif and the O-glycosylation site may still have some functional role and that the protein is tolerated during secretory trafficking. However, the expression level of this truncated *Amtn* transcript is unknown and it may be that it is expressed at very low levels only. It is important to note that, in addition to the exons deleted in the human mutation identified here, the alternatively spliced rat transcript also lacks exon 7. When phosphorylated, the SXE motif within exon 7 has been shown to promote HA precipitation *in vitro* (Abbarin et al., 2015) and is known to bind Ca^{2+} (Sorensen et al., 1995, Kawasaki et al., 2004). Therefore, in contrast to the alternatively spliced rat transcript, it is possible that the unique combination of the absence of exons 3-6 in the mutant transcript, along with the presence of the SXE motif

in exon 7, may be playing a role in the enamel pathology reported for this family.

4.3.4 *Amtn* mouse models and comparison of the enamel phenotype with that described for Individual IV:2 of family AI-154

Both an *Amtn* overexpression mouse model in which AMTN was over-expressed under the control of an *Amel* promoter (p*Amel*:*Amtn*$^{+/+}$), and a knockout (*Amtn*$^{-/-}$) mouse model, have defective enamel (Lacruz et al., 2012a, Nakayama et al., 2015). The p*Amel*:*Amtn*$^{+/+}$ mouse model has brittle enamel that is thinner than wild-type (WT), with a defective surface layer (Lacruz et al., 2012a). By contrast the *Amtn*$^{-/-}$ mouse model has mandibular incisors with a chalky appearance and rough, irregular surface enamel that is easily chipped away, though maxillary incisors and molars were unaffected, as were *Amtn*$^{+/-}$ mice (Nakayama et al., 2015).

The variation in the effect of loss of *Amtn* expression in different teeth in *Amtn*$^{-/-}$ mice may be related to the rate of formation/eruption of the different teeth. Comparatively, the enamel of the continually erupting mandibular incisors was most affected, while molars were apparently unaffected and maxillary incisors appeared stained (Nakayama et al., 2015). Mandibular incisor enamel has less time in which to fully mineralise prior to eruption and AMTN may be required to ensure that the finished enamel surface is fully mineralised prior to enamel eruption. Enamel thickness (at eruption) and the characteristic decussating prismatic architecture were also apparently unaffected in *Amtn*$^{-/-}$ teeth. This may reflect the fact that in WT animals AMTN is not expressed during the secretory stage (Moffatt et al., 2006, Somogyi-Ganss et al., 2012), during which enamel thickness and rod architecture are determined (Risnes, 1998). In the ensuing transition/maturation stages, where AMTN is maximally expressed and the enamel accrues the majority of its final mineral content (Hiller et al., 1975), secondary mineralisation was delayed in *Amtn*$^{-/-}$ mice, especially in the deeper enamel layers (Nakayama et al., 2015). The final mean enamel mineral density achieved across the full enamel thickness in *Amtn*$^{-/-}$ teeth was not significantly

different to WT (Nakayama et al., 2015, Nunez et al., 2015). Microhardness values for the inner and middle enamel layers in null mice were also no different to WT but the microhardness of the surface layer in null mice was significantly reduced by approximately 23% (Nunez et al., 2015). It was concluded that AMTN plays a crucial role in the production of a normally mineralised enamel surface in rodent mandibular incisors.

Pseudogenisation of *AMTN* in enamel-less species has meant that AMTN is regarded as an enamel-specific protein (Gasse et al., 2012). It has been proposed that prismatic enamel evolved in mammals concurrently with AMTN expression becoming restricted to the maturation stage and with changes to *AMTN* exon boundaries and splicing (Gasse et al., 2012, Gasse et al., 2015a, Gasse et al., 2015b). The exact function of AMTN in mammals remains unknown, but its expression within the basal lamina of maturation stage ameloblasts and at the internal basal lamina within the junctional epithelium of erupted rodent teeth suggests functions in both enamel mineralisation and in dento-gingival attachment (Iwasaki et al., 2005, Moffatt et al., 2006, Dos Santos Neves et al., 2012). However, dento-gingival attachment appeared unaffected in both amelotin null (*Amtn*$^{-/-}$) mice (Nakayama et al., 2015) and the AI affected individuals carrying the heterozygous *AMTN* exon 3-6 deletion reported here.

Analysis of the teeth from the individual with the heterozygous deletion of *AMTN* exons 3-6 found that the mean mineral density was 18.8-24.6% lower in comparison to matched control teeth. In contrast, the final mean enamel density achieved across the full enamel thickness in *Amtn*$^{-/-}$ teeth was reduced by only 2.1% and was not significantly different to WT (Nakayama et al., 2015, Nunez et al., 2015). Microhardness values for the inner and middle enamel layers in null mice were also no different to WT but the microhardness of the surface layer in null mice was significantly reduced by approximately 23% (Nunez et al., 2015). It was concluded that AMTN plays a crucial role in the production of a normally mineralised enamel surface in rodent mandibular incisors. Given that the human teeth characterised in this study were erupted and subject to wear in the mouth, it was not appropriate to assess the integrity of the outer aprismatic enamel (the enamel that was most severely disrupted in the *Amtn*$^{-/-}$ mouse) in the context of

the presence of the *AMTN* mutation. However, the difference in the percentage reduction in mean enamel density represents a clear phenotypic difference between the *Amtn*$^{-/-}$ mouse model and the human phenotype characterised in this study.

The differences observed between the mouse and human enamel phenotypes may reflect the type of mutation present. The human patients have an in-frame deletion, predicted to produce a mutant transcript and possibly also a mutant protein, whereas the mice have been shown not to produce any protein and to exhibit a 300-fold reduction in transcript levels (Nakayama et al., 2015). Núñez et al. reported that heterozygous *Amtn*$^{+/-}$ mice had no enamel defects and suggested that murine amelogenesis is unaffected even in the presence of half the normal amount of amelotin (Nunez et al., 2015). Altogether, this suggests that the *AMTN* deletion identified in this study may have a dominant-negative (toxic gain of function) effect due to the presence of a mutated protein, rather than acting through haploinsufficiency.

Whether any toxic gain of function associated with the 8,678bp heterozygous genomic deletion reported here is manifest in the extracellular or intracellular compartments (or both) is unclear. AMTN has been shown to interact with itself and with ODAM, leading to the hypothesis that it forms large extracellular aggregates that are integral to the structure and function of the basal lamina-like layer between the maturation stage ameloblasts and the enamel surface (Holcroft and Ganss, 2011, Dos Santos Neves et al., 2012, Bartlett and Simmer, 2015). Therefore the mutant AMTN protein may fail to form aggregates in appropriate extracellular compartments, such as between the ameloblasts and the enamel surface but may instead accumulate intracellularly in the ER and sequester some of the wild-type protein (as described in Section 4.3.3). The AMELX Y64H mouse model is known to have enamel defects as a result of a failure to traffic aggregated AMELX out of the ER (Barron et al., 2010, Brookes et al., 2014). The phenotype shows that the initial enamel layers form normally, but ER stress leads to production of enamel with fewer prisms and a disorganised structure that leaves the enamel weak (Brookes et al., 2014). The phenotype seen in the teeth from the individual carrying the deletion of *AMTN*

exons 3-6 does not show such defined layers of sound and disorganised enamel, but instead shows small regions of disturbed structure distributed throughout the enamel layers. Therefore the phenotype seen in the teeth of the patients in this study is not similar to that seen in the mouse. This could reflect the very different expression profiles of AMELX (secretory stage) and AMTN (transition and maturation stage) during amelogenesis. It could also reflect the difference in the rate of enamel formation between humans and mice or may suggest that ER stress is not the main pathological mechanism in the enamel of the patients identified in this study.

Elemental analysis of one of the teeth from a human patient carrying the *AMTN* mutation revealed consistently higher carbon and lower oxygen percentage content than control teeth. Also, nitrogen was present in three out of four regions analysed for the teeth from the human patient carrying the *AMTN* mutation but was undetected in the control tooth. These data suggest that organic material was present within the enamel and are consistent with the SEM appearance of the affected teeth, which showed the normal architecture of enamel crystals was apparently obscured by a smooth covering. Unusually high levels of organic material of developmental origin have been previously reported to be present in hypomaturation forms of AI (Wright et al., 1997). Organic material of external origin (for example, salivary protein and protein from dietary sources) have been detected in carious enamel (Shore et al., 2000) suggesting that disruption of the enamel surface may allow organic material to enter the enamel layer. In the affected tooth described here, the organic material was not confined to the surface layers but was present throughout the enamel. However, it is not possible to conclude whether the organic material detected entered the tissue post-eruption due to disturbances in the enamel structure or whether it is of developmental origin, representing retained matrix protein. Treatment of the tooth with sodium hypochlorite solution could allow the nature of the material to be determined. If the material was removed upon hypochlorite treatment, this would indicate that it was protein. The degree of removal could be determined by a combination of SEM and EDX analysis.

In summary, in patients with a heterozygous deletion of exons 3-6 of *AMTN*, the whole enamel thickness is hypomineralised, suggesting that the mutant AMTN may:

1. Disrupt the normal function of the basal lamina-like layer, compromising mineral ion ingress into the deeper enamel; and/or
2. Impede the egress of degraded enamel matrix components out of the enamel, impacting on the mineralisation of the whole enamel layer (the "smooth" appearance of affected enamel (Figure 4.7C-F), as suggested by the presence of organic material overlying the crystals and the nitrogen identified in EDX elemental analysis (Table 4.3 and Figure 4.9); and/or
3. Mis-fold, impairing trafficking through the ER, impacting directly on ameloblast function or causing cell death through ER stress in a manner similar to that described for the *Amelx* c.249T>C (p.Y64H) mutation (Barron et al., 2010, Brookes et al., 2014).

4.3.5 Future investigations

It is important to reiterate that the *AMTN* deletion identified is simply associated with the AI phenotype of affected individuals of the family studied here. No proof of causation has been obtained. The discovery of additional *AMTN* mutations in other individuals with AI would provide further evidence that the mutations are the cause of disease for this family. It may be more appropriate not to limit analysis to families segregating AD AI (as in this study; Section 4.2.7) since it is possible that AI caused by *AMTN* mutations may have more than one aetiological mechanism. For example, biallelic null mutations may also lead to disease but, counterintuitively, this may in fact be milder like the knockout mouse phenotype. This would result in an AR pattern of inheritance.

Neither the *Amtn*$^{-/-}$ nor the p*Amel*:*Amtn*$^{+/+}$ model are accurate models for the effect of the in-frame deletion identified in family AI-154. The development of a mouse model with deletion of *Amtn* exons 3-6, correctly replicating the human mutation, could be achieved since a BLASTp search revealed that murine AMTN (NP_082069) is highly similar to human AMTN (NP_97722.1) (Identities 59%, positives 68%, gaps 2%). Importantly, like the human protein, murine

AMTN contains an IPLT domain in exon 4 and an SXE domain in exon 7. Deletion of exons 3-6 would result in the same amino acid residues as for the mutant human protein sequence (LPG), (though the mouse IPT-like motif is actually LPK, not LPQ as in humans), because like the human protein, the motif spans exons 2-3, and the intron-exon architecture is similar since both species have introns that are in phase zero for the deleted exons. This enables the potential production of an in-frame protein product. Study of such a mouse model would reveal whether heterozygous animals had AI (as would be expected for a dominant gain of function model of disease). This would also enable the consequences of the deletion on the transcripts produced to be defined and would allow examination of how the development of enamel is perturbed in both incisors and molar teeth.

4.3.6 *AMTN* mutations in human variant databases

The lack of reported mutations in *AMTN* causing AI is surprising given its long standing status as a candidate gene for AI. It is possible that in Sanger sequencing and WES-based studies, the type of mutation identified here may have been overlooked due to biases inherent in these approaches. Furthermore, the large exonic deletion reported here is also unusual in that it leaves the protein in frame. It may therefore be the case that the majority of heterozygous *AMTN* mutations in human patients do not result in a toxic gain of function, but rather are functional *AMTN* knockouts, causing either no effect or a clinical phenotype so mild that they remain undiagnosed.

In an attempt to investigate the frequency of homozygous loss of function mutations, databases of known variation (Section 2.4.3) were searched for possible loss of function mutations. Mutations affecting the start codon, stop codon, splice acceptor / donor sites and those leading to frameshifts were identified and their frequencies, in the populations sampled, were recorded (Table 4.4).

Whilst this data shows that loss of function mutations are present in the databases, it is difficult to estimate an overall frequency. It is also likely that the c.466del1 frameshift mutation may represent an erroneous entry since six homozygotes are recorded but no heterozygotes have been reported. In addition, this variant is only identified in one database (EVS) despite its relatively high frequency. EVS specifies that indel calls have not been validated and have higher false positive rates in comparison to SNVs, making frequency data for indels more unreliable than for SNVs. Therefore, excluding the c.466del1 variant, the frequency of loss of function mutations can be estimated to be between 1.57×10^{-4} - 2.31×10^{-4} (depending upon the database used, Table 4.5). If heterozygous *AMTN* loss of function mutations led to AI, such frequencies would equate to around 1 in 4,300 to 1 in 6,400 people having AI due to *AMTN* mutations. Since the estimates of the frequency of AI due to all genetic causes range from 1 in 700 in an isolated Swedish population (Backman and Holm, 1986) to 1 in 14,000 in the outbred US population (Witkop and Sauk, 1976), it seems unlikely that heterozygous loss of function mutations in *AMTN* lead to clinically recognisable disease. Rather it seems more likely that it is only very rare gain of function mutations (possibly such as the deletion identified here) or homozygous loss of function mutations that cause AI.

Genomic variant (GRCh37)	Transcript variant	Predicted protein variant	Type	Alleles							Homozygotes (source)	CADD v1.3
				ExAC		EVS		dbSNP144				
				Variant (Total)	Frequency	Variant (Total)	Frequency	Variant (Total)	Frequency	Reference		
4:71384477T>C	c.-15-3T>C	p.?	splice acceptor	3 (121168)	2.5E-05	1 (13006)	7.69E-05	1 (5008)	2.00E-04	rs201193624		0.067
4:71384496T>A	c.2T>A	p.M1?	start loss	2 (121324	1.6E-05	1 (13006)	7.69E-05	2 (121406)	1.65E-05	rs370334023		25.7
4:71384497G>A	c.3G>A	p.M1?	start loss	1 (121320)	8.2E-06	N/A	N/A	1 (121406)	8.24E-06	rs755228148		25.5
4:71384544T>G	c.50T>G	p.L17*	stop gain	1 (121260)	8.2E-06	N/A	N/A	1 (121406)	8.24E-06	rs769401306		36
4:71384549G>A	c.54+1G>A	p.?	splice donor	N/A	N/A	1 (13006)	7.69E-05	1 (4542)	2.20E-04	rs141338494		24.6
4:71388470A/C	c.55-2A>C	p.?	splice acceptor	1 (121232)	8.2E-06	N/A	N/A	1 (121412)	8.24E-06	rs779389513		9.741
4:71388471G/A	c.55-1G>A	p.?	splice acceptor	1 (121226)	8.2E-06	N/A	N/A	1 (121412)	8.24E-06	rs748991501		22.2
4:71388556G/C	c.138+1G>C	p.?	splice donor	2 (121044)	1.7E-05	N/A	N/A	2 (121412)	1.65E-05	rs767938897		25.2
4:71389582C/T	c.139-3C>T	p.?	splice acceptor	1 (121154)	8.3E-06	1 (13006)	7.69E-05	1 (121412)	8.24E-06	rs367591948		10.43
4:71389583A/G	c.139-2A>G	p.?	splice acceptor	N/A	N/A	N/A	N/A	1 (1318)	7.59E-04	rs201144112		23.5
4:71389585G/T	c.139G>T	p.V47F	splice acceptor	6 (121184)*	5.0E-05	N/A	N/A	4 (5008)	7.99E-04	rs192563032	1 (ExAC)	23.3
4:71389585G/A	c.139G>A	p.V47I	splice acceptor	1 (121184)	8.3E-06	N/A	N/A	1 (5008)	2.00E-04	rs192563032		25.3
4:71389587C/G	c.141C>G	p.V47=	splice acceptor	13 (121206)	1.1E-04	1 (13006)	7.69E-05	1 (4552)	2.20E-04	rs140471397		12.36
4:71389618C/T	c.172C>T	p.Q58*	stop gain	1 (121270)	8.2E-06	N/A	N/A	1 (242822)	4.12E-06	rs752376041		39
4:71390678 T/C	c.294T>C	p.H98=	splice donor	2 (120746)	1.7E-05	N/A	N/A	2 (121404)	1.65E-05	rs749175667		0.35
4:71394447-71394460 CAATTTTTGTCACA/-	c.302_315del CAATTTTTGTCACA	p.I102Tfs*39	frameshift	N/A	N/A	N/A	N/A	1 (121390)	8.24E-06	rs778954160		35
4:71394461C/T	c.316C>T	p.Q106*	stop gain	1 (114686)	8.7E-06	N/A	N/A	1 (121390)	8.24E-06	rs746033781		40
4:71394928G/T	c.357+1G>T	p.?	splice donor	1 (63288)	1.6E-05	1 (12884)	7.76E-05	1 (3546)	2.82E-04	rs142798658		23.8
4:71396754A/T	c.358-2A>T	p.?	splice acceptor	1 (119496)	8.4E-06	N/A	N/A	1 (121410)	8.24E-06	rs770675733		21.7
4:71396864G/-	c.466del1	p.G156E*5	framehift	N/A	N/A	12 (12520)*	9.58E-04	N/A	N/A	N/A	6 (EVS)	27.8
4:71397018G/A	c.619+1G>A	p.?	splice donor	1 (119550)	8.4E-06	N/A	N/A	1 (121410)	8.24E-06	rs752143599		23.8

Table 4.4: Potential *AMTN* loss of function mutations identified in publically available databases of variation.

Mutations in ExACv0.3, EVS and dbSNP144 (Methods section ref) were recorded if they resulted in the loss of a start codon, the gain of a stop codon, a frameshift or affected splice acceptors or splice donors (variants positioned within introns up to 8bp away from the splice junction or within the exon, those up to 3bp away from the splice junction). Note that dbSNP144 includes some data from the ExAC database. All databases were accessed 16/04/2016.

The variant highlighted in yellow was excluded from subsequent analysis (Table 4.5) due to number of reported homozygotes relative to heterozygotes.

The variants highlighted in grey (CADD scores <15) were excluded from subsequent analysis.

Database	ExAC	EVS	dbSNP144
Total number of variant alleles	19	3	20
Largest number of variant alleles sampled at one position	121,324	13,006	121,412*
Overall allele frequency	1.57×10^{-4}	2.31×10^{-4}	1.65×10^{-4}
Alleles per 10,000	1.6	2.3	1.6

Table 4.5: Allele frequency of potential *AMTN* loss of function mutations.

Individual frequencies were calculated separately for information from each database. Mutations with CADD scores of <15 were excluded from the analysis.

* The most frequent number of alleles sampled is used instead of the largest number of alleles sampled for dbSNP144 since the largest number is around double that of the number of alleles at the majority of the other sites sampled. Information compiled 5th April 2016.

5 General Discussion

5.1 Summary of the key findings

This thesis describes a study investigating the causes of amelogenesis imperfecta (AI), a defect of tooth enamel. One section studied a syndromic form of AI and the other researched a cause of isolated non-syndromic AI.

In chapter 3, the aim was to identify the genetic cause of HS, an autosomal recessively inherited combination of SNHL, AI and variable retinal dystrophy. DNA from fourteen individuals from a total of ten families with SNHL, AI and retinal dystrophy was subjected to WES. Biallelic variants in *PEX1* and *PEX6* were identified in two and five families respectively. Biallelic variants in *PEX1* and *PEX6* were already known to cause ZSSDs. The ZSSDs include ZS, NALD and IRD and vary in severity, with ZS being the most severe and more often the result of null mutations, and IRD being previously thought to be the least severe, more often the result of at least one missense mutation. The combinations of *PEX1* and *PEX6* variants identified in the HS individuals always included at least one hypomorphic variant. This results in a phenotype restricted to fewer organs, meaning lower disease severity overall by comparison with other ZSSDs, and suggests that HS represents the mildest presentation of ZSSDs. This study therefore revealed a further extension to the phenotype spectrum associated with mutations in these genes, with HS being the mildest form of ZSSD. Investigation of the localisation of PEX1, PEX6 and PEX14 within the murine retina revealed that the proteins are found throughout multiple retinal layers, suggesting that they may contribute to the maintenance and function of the retina.

Biallelic mutations in *USH2A* were identified in two other families and therefore, these families were assumed to have USH and not HS. No genetic cause for the AI phenotype was identified and investigation of the enamel phenotype in one family suggested that the enamel defect was not AI.

In the remaining family, CNV analysis of WES data predicted a heterozygous deletion spanning the start codon of *MYO7A*. A subsequent search for heterozygous SNVs within *MYO7A* failed to reveal a second potentially pathogenic variant. Therefore the genetic cause of disease remains unknown in this family.

In chapter 4, the aim was to identify novel AI gene(s). DNA from a family segregating autosomal dominant hypomineralised AI underwent WES. Analysis of SNVs and small indel variants failed to highlight a strong candidate gene. However, CNV analysis using ExomeDepth identified a heterozygous deletion spanning exons 3 to 6 of *AMTN*, a gene implicated in AI through its chromosomal location, expression during amelogenesis and the phenotypes of the *Amtn* mouse models. Sanger sequencing revealed that the heterozygous CNV segregated with disease in all family members for which DNA was available. Screening of an additional thirty-five autosomal dominant AI families failed to identify any other mutations in *AMTN*, or any further occurrence of the same deletion.

The enamel phenotype in teeth from an individual with the heterozygous deletion of *AMTN* exons 3 to 6 was investigated using μCT, SEM and EDX. Analysis of the enamel mineral density by μCT found that the enamel of teeth from an individual with the heterozygous deletion of *AMTN* exons 3 to 6 was 18.8-24.6% less dense than control teeth. SEM revealed that the enamel appeared to have regions of very disturbed structure as well as regions with a more typical structure, although in all cases, the fine structure of the enamel was obscured with an apparently overlying material. Unlike the phenotype seen in the *Amtn*$^{-/-}$ mouse, the regions of disturbed enamel spanned the entire enamel layer. EDX analysis identified nitrogen within the enamel of the individual affected by AI, suggesting that organic material was present.

These findings, together with the evidence already present in the literature, strongly suggest that mutations in *AMTN* cause AI. However, identification of a second family in which an *AMTN* variant segregates with AI, or the development

of a more appropriate mouse model to assess the effects of the CNV identified in this study, would further strengthen this conclusion.

5.2 The impact of identification of the genetic cause of disease

Both of the studies undertaken to complete this thesis involved the identification of the genetic cause of diseases within families. In the past, disease diagnosis in inherited diseases was based solely on symptoms, but diagnosis can now often be confirmed by identification and segregation of mutations with disease within families. Advances in sequencing methodologies (Bentley et al., 2008), combined with the sequencing and detailed annotation of the human genome (Lander et al., 2001, McPherson et al., 2001) and its variation (Sherry et al., 2001, Abecasis et al., 2010), have facilitated this. In some cases, where symptoms of different syndromes overlap, such as with USH and HS, this can alter the diagnosis from that which was given initially and may prompt the implementation of a different lifestyle and/or treatment.

Knowledge of the cause of genetic disease can have both positive and negative effects on patients and their families. In cases of positive findings, it has the potential to inform life decisions regarding genetic testing, family planning, diet or whether to risk certain activities. In cases of negative findings, it can eliminate anxiety and the need for health surveillance where a family history had previously indicated the possibility of disease. Genetic diagnosis can also enable suitable treatments to be prescribed or can be the starting point for their development by recruitment to clinical trials of a specific disease cohort. However, such knowledge of an individual's genetic status can also affect life aspirations and can change people's perceptions of themselves and others.

Since the majority of cases of HS with mutations in *PEX1* or *PEX6* had originally been diagnosed as USH, an adjustment was inevitably required for families to adapt to the new health information that a correct genetic diagnosis brought. USH, whilst potentially progressing to total blindness in the long term, does not encompass symptoms other than sensory defects. In contrast, HS as

part of the ZSSD spectrum, whilst initially documented solely as a sensory and enamel disorder, could involve the development of symptoms in liver, kidney, brain or other organs in later life. Such symptoms may even affect life expectancy. The majority of currently identified HS patients are in adolescence and early adulthood. A few older HS patients exist, but a cohort needs to be studied over the longer term in order to determine whether symptoms do progress to include defects in other organs. IRD and NALD have already been reported to be progressive conditions (Berendse et al., 2016). When first described, HS did not encompass an eye phenotype (Heimler et al., 1991), but one of the index cases developed visual defects in early adulthood (Lima et al., 2011). The cases described here all have sight problems and whilst this could reflect a recruitment bias, it also might suggest that there is significant heterogeneity in the HS phenotype both between individuals and at different ages.

Reporting of variants identified in WES and WGS studies to healthcare practitioners, and ultimately to patients, is complex. Guidelines have been written to direct the classification of the variants identified but include many caveats (Richards et al., 2015). Equally, recommendations on the reporting of secondary and incidental findings have stressed reporting only variants involved in clinically actionable disease or those causing disease likely to have serious consequences for future generations (for example, diseases whose onset of symptoms is after child-bearing years) (Green et al., 2013). However such recommendations were criticised by some for removing individual patient choice as to what is reported (Holtzman, 2013, Townsend et al., 2013). Variants identified by sequencing may be previously unreported or may be in genes about which little is known. Alternatively, variants may reside in genes associated with serious disease but may actually be benign. Even variants identified in very well studied genes such as breast cancer 1, early onset (*BRCA1*; MIM *113705) and *BRCA2* (MIM *600185) known to be associated with breast, ovarian and other cancers, are often of unknown significance (Cheon et al., 2014). Therefore reporting must consider the certainty of the diagnosis and the level of clinical and social benefits that their reporting might realise.

5.3 Massively parallel sequencing in a clinical setting

In August 2014, former UK Prime Minister David Cameron announced funding for the Genomics England project, which aims to sequence 100,000 genomes, in collaboration with Illumina and the Wellcome Trust, by 2017. The project aims to sequence the genomes of cancer patients and patients with rare diseases in order to identify causative and disease-associated variants. Patients with rare diseases will be sequenced in trios, i.e. ideally with their mother and father to enable the identification of *de novo* variants and to allow segregation analysis to be performed *in silico*. At present, AI is not on the 100,000 genomes project's rare disease list, but the list does include PBDs (30th June 2016). 100,000 genomes and other examples of such large scale sequencing projects around the world (Genome of the Netherlands Consortium, 2014, Auton et al., 2015, Gudbjartsson et al., 2015) are indicative of the falling cost and greater infrastructure to support the use of gene panel / exome / genome sequencing. Use of diagnostic massively parallel sequencing seems likely to become routine in a healthcare setting in future (Ellingford et al., 2016a, Ellingford et al., 2016b, Prasad et al., 2016). Such implementation brings ethical, logistical and economic challenges and questions as to best practice.

Currently, genetics services within the UK NHS mainly utilise diagnostic gene panels in order to focus NGS on disease genes for particular phenotypes, for example, deafness or visual defects. This increases throughput whilst decreasing the cost (both for the sequencing and for verification of findings), complexity of analysis, data storage requirements and the chance of incidental findings. However, it is apparent in many examples of genetic disease that Sanger sequencing of exonic regions, WES or sequencing of diagnostic gene panels may not identify the causative mutation(s) (Braun et al., 2013, Small et al., 2016). Deep intronic, regulatory or splicing variants may be identified by WGS or RNA sequencing. However, their recognition as the cause of disease amongst thousands of other variants may be difficult, and additional proof of causation through experimental assays with patient-derived cell lines or other cell biology approaches, such as cell lines engineered with the patient's

mutation, will be required. For example, *USH2A* mutations have been analysed using patient nasal epithelial cells (Vache et al., 2010). However, analysis of splicing for genes with tooth specific expression will not be possible or, if minigene assays (Gaildrat et al., 2010) are utilised, analysis may be confounded by the use of a different promoter or cell type for expression or by the small proportion of the gene included in the construct.

5.4 The future of AI genetic diagnostics

Recently, Leeds Teaching Hospital's Yorkshire Regional Genetics Service designed and carried out a trial of an AI gene diagnostic panel. The panel targets a total of twenty-two known non-syndromic, syndromic and candidate AI genes. A focused clinical exome reagent is used to capture the sequence of a panel of around 6000 genes associated with clinically relevant diseases, including the 22 target AI genes. However, analysis is restricted to only DNA sequence within AI gene regions. In the trial, ten AI cases were sequenced and, of these, mutations in known AI genes were identified in six cases. The detection rate is likely to increase with the inclusion of a greater number of known and candidate AI genes. Mass screening of UK patients via a diagnostic panel will allow estimates of the frequency of AI in the various UK regions to be made, to allow for appropriate allocation of treatment and funding. Retrospective analysis of the efficacy of AI restorative treatments for groups of patients with known AI genotypes may lead to the development of best practice guidelines for AI treatments based on the causative mutation identified in each case. It will also facilitate the inclusion of AI variants into the Leeds AI LOVD and may aid classification of variants as pathogenic or benign through observed frequency in patients compared to well phenotyped controls.

Genetic diagnosis of AI will also be important for early implementation of appropriate dental treatment and stratification of patients for clinical trials. The paradigm that tooth enamel cannot undergo cellular repair has meant that current AI treatments have the aim of restoring function by use of artificial coverings to maintain crown function (Patel et al., 2013). The recent development and clinical safety trial of the use biomimetic self-assembly

peptides to treat small pre-carious enamel lesions has shown that such lesions are ameliorated by treatment (Brunton et al., 2013). Whilst such treatment is promising for minor lesions in otherwise sound enamel, its use in patients with very thin or soft enamel, as in AI, would not be effective, even if used from an early age.

However, genetic diagnosis of AI may also bring wider health benefits. AI can be part of a more complex phenotype, in which other symptoms only present later on in development. For example, *CNNM4* mutations cause AI that may be obvious from a young age, but cone rod dystrophy may only be identified at an early stage by specific ophthalmological investigations in some cases (Jalili, 2010). Also, whilst children with *FAM20A* mutations present with AI (O'Sullivan et al., 2011), it may not be until adulthood that pathology associated with calcification in other organs, such as nephrocalcinosis, is detected (Poulter et al., 2015). For genes more recently identified as causal of AI, longer term study of molecularly defined patient cohorts will help to determine whether other observed phenotypes are typical and indicative of syndromic disease or specific to particular patients or families. Early diagnosis of such syndromes may allow dietary or pharmaceutical interventions to delay or even to prevent the onset of later presenting symptoms.

When panel based NGS, such as that proposed as an NHS diagnostic screen, is used in collaboration with research, it allows researchers to focus on sequencing only samples without mutations in known AI genes, thus expediting new AI gene discovery and maximising the efficient use of research effort and funds. Leeds AI genetics research is well placed since it is a recruitment hub for cases across much of the UK as well as from other countries including Pakistan, Costa Rica and Oman. Such a range of populations show great genetic diversity, which facilitates the discovery of new genes for AI. Conversely, the identification of multiple patients with a particular mutation or mutations in a particular gene will allow for more detailed phenotyping of affected enamel, teeth and investigation of the association of other oral abnormalities such as anterior open bite with AI. Such information may enable more effective, tailored dental treatment for AI patients in future.

5.5 New AI genes remain to be identified

Analysis of AI cohort studies have indicated that there remains a large percentage of cases for which exonic mutations in known AI genes are not identified. In such AI cohorts, around 51-72% of cases remain genetically undiagnosed (Chan et al., 2011, Wright et al., 2011, Prasad et al., 2016). There are a number of possible explanations for this observation and it is likely that the unexplained cases/families fall into several different categories.

In a proportion of unexplained cases, the variants may affect regulatory regions outside of coding exons and therefore the causative variant will not be captured or annotated by current sequencing and analysis pipelines. For example, a report of biallelic *FAM20A* variants included an untranscribed allele in one family and an allele including a large duplication in another family (Poulter et al., 2015). In addition, causative mutations within exonic splicing enhancers or silencers, may be overlooked even if they are positioned within regions that are covered by WES due to the assignment of pathogenicity scores that do not reflect their important function.

Large intragenic heterozygous deletions such as the one identified in *AMTN* in this study have not previously been identified in AI patients. There are two reports of an identical deletion in *SLC24A4* (Seymen et al., 2014, Prasad et al., 2016), one of a deletion in *WDR72* (Hentschel et al., 2016), one of a deletion in *AMBN* (Poulter et al., 2014d) and one deletion in *AMELX* (Lagerstrom et al., 1991), identified through arrayCGH, gene panel NGS, WES-based depth of coverage analysis and candidate gene PCR respectively. Whilst these kinds of mutations were identified in these cases, each was homozygous or hemizygous, simplifying detection. The *AMTN* deletion reported herein is the first large heterozygous deletion detected in AI, which is likely to represent the technological limits of the available detection methods rather than a deficit of this kind of mutation. It is likely that the shift in research methodology from WES to WGS will make CNVs, including heterozygous variants, easier to identify. However, the nature of the sequence around such variants may mean that it is difficult to report them. For example, during the course of this study, a

homozygous deletion of *C4orf26* exon 2 was identified in one family, but due to the repetitive nature of the sequence surrounding the deletion, the breakpoints of the deletion could not be identified due to problems in amplifying the region using PCR. Such difficulty may be commonly experienced due to the repetitive nature of the types of sequences most likely to elicit structural variation (van Binsbergen, 2011).

Studies of MIH, that in part phenocopies AI, albeit only in particular teeth, have attempted to correlate SNPs in AI genes with MIH risk (Jeremias et al., 2016). Similarly, larger AI cohorts in particular populations may reveal that seemingly common variants in AI genes and other loci have an additive effect on the enamel phenotype and act as polygenic susceptibility loci. However clear Mendelian inheritance patterns over a number of generations in the great majority of identified families may argue that this is not the case.

In addition to the variants that are not identified due to their non-exonic location or type, the large proportion of AI cases without genetic diagnosis after candidate gene screening (Chan et al., 2011, Wright et al., 2011, Prasad et al., 2016) suggests that a number of AI genes remain to be identified. Furthermore, it is evident that AI is a genetically heterogeneous disorder and that the lines between non-syndromic and syndromic disease are indistinct. The heterogeneity of the genetic causes of AI may reflect the complexity of the processes involved in amelogenesis and their functional separation from other processes necessary to sustain life. The genes identified so far include both those with tooth or even ameloblast specific expression and those with more general expression, suggesting that mutations may influence the phenotype outside of the enamel, much like mutations in *PEX1* and *PEX6*.

5.6 Treatments for mutations leading to protein misfolding or causing premature termination codons

Some of the HS *PEX1* and *PEX6* variants and the exonic *AMTN* deletion identified in this study lead or may lead to protein misfolding. Protein misfolding can induce ER stress, which may lead to apoptosis (Wang and Kaufman, 2016).

The persistent state of oxidative stress induced by the reduced function of PEX1 or PEX6 in HS may also make cells more susceptible to ER stress (Redza-Dutordoir and Averill-Bates, 2016). Chemical chaperones have been sought and have undergone clinical trials for ZSSDs (as discussed in Section 3.3.3). They have also been used to prevent ER stress pathology in the *Amelx*$^{Y/Y64H}$ mouse model (Brookes et al., 2014). Systematic study of mutations that are predicted to lead to protein misfolding via a cell culture system would allow treatments to be targeted to individuals with mutations receptive to such treatments. Greater knowledge of which mutations lead to protein misfolding could allow a reliable algorithm, specific to each gene to be developed to predict the outcome of mutations *in silico*. *in utero* administration of treatments alleviating protein misfolding may be able to prevent the development of AI and other pathology, or to reduce severity, in the case of syndromic disease such as HS.

Premature termination codons (PTCs) due to nonsense mutations are common in ZS and have been identified in combination with hypomorphic alleles in HS. Normal translation termination is not entirely efficient and can result in readthrough, although the particular stop codon and downstream sequence influences the amount of readthrough (Peltz et al., 2013). In cases of PTCs, readthrough can also have the effect of suppressing NMD via removal of the exon junctional complex. The natural phenomenon of readthrough has been exploited by the design and screening of drugs that target ribosomes to enhance readthrough. Therapeutics such as aminoglycoside agents and the lower toxicity, non-aminoglycoside agents such as PTC124, also known as Ataluren, have been trialled as treatments for Duchenne muscular dystrophy and cystic fibrosis, although results are less promising than those obtained in animal models (Peltz et al., 2013). These reagents, like natural readthrough, act most efficiently on the UGA stop codon and least efficiently on the UAA stop codon (Welch et al., 2007). Therefore readthrough therapies would need to be prescribed on the basis of prior knowledge of each patient's genotype.

5.7 Correction of genetic defects in AI

A number of different therapeutic approaches are available for the correction of genetic defects, each of which could be applied to the defects identified during the course of this project.

At present, carriers of serious genetic diseases are able to choose to attempt to circumvent the possibility of passing on an inherited disease to their offspring by undergoing *in vitro* fertilisation and pre-implantation genetic diagnosis of embryos. Selection and implantation of only healthy embryos can therefore eliminate the inheritance of such diseases by future generations, although such selection will not prevent the occurrence of *de novo* mutations. Whilst ZSSDs are included in the Human Fertilisation and Embryology Authority's list of conditions deemed sufficiently serious to allow for such embryo selection, unsurprisingly, isolated AI is not. In theory though, not only could embryos be selected to avoid genetic disease, but faulty genes could be edited by clustered regularly interspaced short palindromic repeats (CRISPR) technologies. CRISPR is only currently in use in humans for research, for the purpose of studying the genes that control the early development of the embryo (Callaway, 2016). Current guidelines prohibit the culture of CRISPR-modified human embryos beyond seven days post fertilisation. It remains to be seen to what extent the advent of CRISPR technologies will result in genetic editing in human embryos, or even of adults, in future. Correction of genetic defects in murine germ cell progenitor cells and in adult mice has already been demonstrated (Yin et al., 2014a, Wu et al., 2015).

Such advances in genetic selection and manipulation must be tempered with respect for people's views on what warrants "genetic correction" and therefore also what constitutes disability. For example, deaf individuals have been known to identify themselves as belonging to a separate culture with their own identity rather than regarding themselves as disabled in any way (Nunes, 2006). Similarly, some individuals with primordial dwarfism have been reported to favour having a child with the same condition for cultural reasons (Sanghavi, 2006). Whilst the worldwide moratorium on reproductive human cloning and

therapeutic human cloning for medical practice has stood for nearly twenty years, it remains to be seen if genome editing will also be subject to such a long-standing embargo. Ethical arguments against gene editing are complex due to the clear benefit of correcting serious genetic defects with significant impact on quality of life for patients. However, even if gene editing was morally acceptable, it is not yet refined enough for use in humans due to the high chances of off-target effects, low frequency of successful targeting and unknown clinical outcome. Moral arguments against gene editing are much the same as for any other genetic intervention without defined boundaries: for which diseases should it be allowed and should parents be allowed to make such decisions for their future child? Such questions have different answers for different individuals, but society as a whole will decide what is permissible by law.

Viral delivery of genes has been used to treat human disease for over 25 years (Samiy, 2014). The use of adenovirus and recombinant adenovirus associated vectors (AAV) to deliver gene therapy has been developed in animal models and in some cases, has entered clinical trials for treatment of a number of diseases, including eye disorders such as LCA (Weleber et al., 2016). Unlike some other viral vectors, adenovirus and AAV do not integrate into the host genome and therefore insertional mutagenesis is not a safety concern. More recently, non-viral vectors such as nanoparticles have been the subject of proof of principle studies to show delivery of gene therapy to mouse retina is possible (Kachi et al., 2005).

Eye diseases seem suited to such treatments due to the small, enclosed, accessible, immune-privileged nature of the eye, the post-mitotic status of the cell types targetted and the fact that eye diseases are often only slowly progressive therefore providing time for therapies to be given. However, such treatment of enamel affected by AI is less practical. Once teeth are formed and have erupted to reveal an AI phenotype, such treatment is too late. The ameloblasts are lost upon tooth eruption, therefore the developing enamel, present *in utero* for the primary teeth or from around birth for many of the secondary teeth, would need to be targetted. Accurate delivery of the therapy

would be difficult and the encoded protein would be required to be active over a number of years, since the enamel can take years to fully develop.

Development of AI therapies is also hampered by the lack of a robust ameloblast cell model in which to screen and study treatments. Primary ameloblast cells have been isolated from the developing tooth germ but lack the potential for proliferation (Den Besten et al., 1998, DenBesten et al., 2005). Ameloblast cell lines have been derived from cells of the developing tooth germ in mouse (Chen et al., 1992, Nakata et al., 2003), rat (Kawano et al., 2002) and pig (DenBesten et al., 1999) but analysis of some of their expression profiles has shown that some cell lines may be more appropriate than others depending upon the stage of amelogenesis to be studied (Sarkar et al., 2014). Production of ameloblast like cells from induced pluripotent stem cells has been shown to be possible in rodents (Arakaki et al., 2012, Yoshida et al., 2015) and could be a starting point for regeneration of teeth or autologous transfer of functional ameloblasts to tooth buds in some cases. Rodent epithelial skin cells have also been shown to have the potential to be induced to become a functional, ameloblast-like cell (Liu et al., 2013). However, the question of to what extent ameloblasts can function in an environment where cells that participate in reciprocal signalling with ameloblasts or that function to support them still possess a genetic defect, is unknown.

Organ culture of tooth buds has been achieved and represents an *in vitro* system in which AI and novel therapies could be studied and tested respectively. Thin or soft enamel in AI results in both the functional failure of the tooth surface and the exposure of underlying dentine and ultimately also the innervated pulp tissue to the oral cavity resulting in pain. In theory, such teeth could be replaced by tooth buds generated from induced pluripotent stem cell (iPSC) populations grown in culture. Although this would require underlying tissues to be sufficiently healthy in order to support them. In rodents, tooth germs and units consisting of the tooth, periodontal ligament and alveolar bone have been successfully transplanted and erupted or engrafted respectively (Ikeda et al., 2009, Oshima et al., 2011).

Isolation of AI patient ameloblast like cells derived from iPSCs would provide opportunities to test gene and drug therapies and could also allow study of the effects of the AI causing mutations on ameloblast cell function. However, as already stated, ameloblasts function in concert with both the odontoblasts and the overlying stratum intermedium, and later, the papillary cell layer. Therefore studying ameloblasts in isolation may not accurately reflect their behaviour *in vivo*. One study established a primary AI ameloblast-like cell line from an AI patient's developing third molar tooth and found that it expressed ameloblast cell markers and could be transduced by adenovirus, albeit at a low efficiency (Borovjagin et al., 2011). However the technical difficulties compared with administration of gene therapy or genome editing reagents to the eye mean that it is unlikely that gene therapies for AI will be widely adopted in the foreseeable future.

In conclusion, research into AI is already translating into genetic diagnosis within a clinical setting. Wider application of such diagnosis is likely to increase the proportion of patients recruited to research studies without mutations in known AI genes. This will lead to expedited detection of new AI genes. Increased annotation of non-exonic regions of the genome and the wider application of WGS in research studies may also speed discovery of AI causing mutations. Identification of new AI genes will prompt studies to identify AI gene functions and may include the development of new animal models. This will in turn increase the understanding of amelogenesis and of which parts of the process are perturbed in AI. Such studies may lead to the identification of common pathways to target in the design of treatments. Endoplasmic reticulum stress induced activation of the unfolded protein response is already emerging as such a common pathway in the pathogenesis of AI. In addition, the identification of causative mutations in a large patient cohort will allow stratification of patients for clinical trials of restorative treatments and, after further research, potentially also genetic or chemical treatments to attenuate the AI phenotype or to correct AI gene defects.

References

Abbarin, N., S. San Miguel, J. Holcroft, K. Iwasaki and B. Ganss (2015). "The enamel protein amelotin is a promoter of hydroxyapatite mineralization." J Bone Miner Res **30**(5): 775-785.

Abecasis, G. R., D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles and G. A. McVean (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

Aberdam, D., M. F. Galliano, J. Vailly, L. Pulkkinen, J. Bonifas, A. M. Christiano, K. Tryggvason, J. Uitto, E. H. Epstein, Jr., J. P. Ortonne and et al. (1994). "Herlitz's junctional epidermolysis bullosa is linked to mutations in the gene (LAMC2) for the gamma 2 subunit of nicein/kalinin (LAMININ-5)." Nat Genet **6**(3): 299-304.

Abiko, Y., M. Nishimura, J. Arai, J. Kuraguchi, M. Saitoh and T. Kaku (1996). "Apoptosis in the reduced enamel epithelium just after tooth emergence in rats." Medical Electron Microscopy **29**(2): 84-89.

Acharya, B. S., P. Ritwik, G. M. Velasquez and S. J. Fenton (2012). "Medical-dental findings and management of a child with infantile Refsum disease: a case report." Spec Care Dentist **32**(3): 112-117.

Adato, A., G. Lefevre, B. Delprat, V. Michel, N. Michalski, S. Chardenoux, D. Weil, A. El-Amraoui and C. Petit (2005). "Usherin, the defective protein in Usher syndrome type IIA, is likely to be a component of interstereocilia ankle links in the inner ear sensory cells." Hum Mol Genet **14**(24): 3921-3932.

Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov and S. R. Sunyaev (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.

Ahmed, Z. M., G. I. Frolenkov and S. Riazuddin (2013). "Usher proteins in inner ear structure and function." Physiol Genomics **45**(21): 987-989.

Aird, D., M. G. Ross, W. S. Chen, M. Danielsson, T. Fennell, C. Russ, D. B. Jaffe, C. Nusbaum and A. Gnirke (2011). "Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries." Genome Biol **12**(2): R18.

Al-Hashimi, N., J. Y. Sire and S. Delgado (2009). "Evolutionary analysis of mammalian enamelin, the largest enamel protein, supports a crucial role for

the 32-kDa peptide and reveals selective adaptation in rodents and primates." J Mol Evol **69**(6): 635-656.

Alachioti, X. S., E. Dimopoulou, A. Vlasakidou and A. E. Athanasiou (2014). "Amelogenesis imperfecta and anterior open bite: Etiological, classification, clinical and management interrelationships." J Orthod Sci **3**(1): 1-6.

Alaluusua, S., P. Calderara, P. M. Gerthoux, P. L. Lukinmaa, O. Kovero, L. Needham, D. G. Patterson, Jr., J. Tuomisto and P. Mocarelli (2004). "Developmental dental aberrations after the dioxin accident in Seveso." Environ Health Perspect **112**(13): 1313-1318.

Alberts, B. J., A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. (2002). Molecular biology of the cell. New York, USA, Garland Science.

Aldred, M. J. and P. J. Crawford (1995). "Amelogenesis imperfecta--towards a new classification." Oral Dis **1**(1): 2-5.

Aldred, M. J., P. J. Crawford, E. Roberts, C. M. Gillespie, N. S. Thomas, I. Fenton, L. A. Sandkuijl and P. S. Harper (1992). "Genetic heterogeneity in X-linked amelogenesis imperfecta." Genomics **14**(3): 567-573.

Aldred, M. J., R. Savarirayan and P. J. Crawford (2003). "Amelogenesis imperfecta: a classification and catalogue for the 21st century." Oral Dis **9**(1): 19-23.

Aldred, M. J., R. Savarirayan, S. R. Lamande and P. J. Crawford (2002). "Clinical and radiographic features of a family with autosomal dominant amelogenesis imperfecta with taurodontism." Oral Dis **8**(1): 62-68.

Aller, E., T. Jaijo, M. Beneyto, C. Najera, S. Oltra, C. Ayuso, M. Baiget, M. Carballo, G. Antinolo, D. Valverde, F. Moreno, C. Vilela, D. Collado, H. Perez-Garrigues, A. Navea and J. M. Millan (2006). "Identification of 14 novel mutations in the long isoform of USH2A in Spanish patients with Usher syndrome type II." J Med Genet **43**(11): e55.

Aller, E., L. Larrieu, T. Jaijo, D. Baux, C. Espinos, F. Gonzalez-Candelas, C. Najera, F. Palau, M. Claustres, A. F. Roux and J. M. Millan (2010). "The USH2A c.2299delG mutation: dating its common origin in a Southern European population." Eur J Hum Genet **18**(7): 788-793.

Aoba, T., T. Tanabe and E. C. Moreno (1987). "Proteins in the enamel fluid of immature porcine teeth." J Dent Res **66**(12): 1721-1726.

Aparisi, M. J., E. Aller, C. Fuster-Garcia, G. Garcia-Garcia, R. Rodrigo, R. P. Vazquez-Manrique, F. Blanco-Kelly, C. Ayuso, A. F. Roux, T. Jaijo and J. M. Millan (2014). "Targeted next generation sequencing for molecular diagnosis of Usher syndrome." Orphanet J Rare Dis **9**: 168.

Arakaki, M., M. Ishikawa, T. Nakamura, T. Iwamoto, A. Yamada, E. Fukumoto, M. Saito, K. Otsu, H. Harada, Y. Yamada and S. Fukumoto (2012). "Role of epithelial-stem cell interactions during dental cell differentiation." J Biol Chem **287**(13): 10590-10601.

Auton, A., L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean and G. R. Abecasis (2015). "A global reference for human genetic variation." Nature **526**(7571): 68-74.

Avery, K., P. F. Steele and N. Avery (2002). Oral Development and Histology. New York, NY, USA, Thieme.

Backman, B. and A. K. Holm (1986). "Amelogenesis imperfecta: prevalence and incidence in a northern Swedish county." Community Dent Oral Epidemiol **14**(1): 43-47.

Bader, P. I., S. Dougherty, N. Cangany, G. Raymond and C. E. Jackson (2000). "Infantile refsum disease in four Amish sibs." Am J Med Genet **90**(2): 110-114.

Bagattoni, S., G. D'Alessandro, A. Prete, G. Piana and A. Pession (2014). "Oral health and dental late adverse effects in children in remission from malignant disease. A pilot case-control study in Italian children." Eur J Paediatr Dent **15**(1): 45-50.

Balmer, R. and S. A. Fayle (2007). "Enamel defects and ectopic eruption in a child with Usher syndrome and a cochlear implant." Int J Paediatr Dent **17**(1): 57-61.

Balmer, R., J. Toumba, J. Godson and M. Duggal (2012). "The prevalence of molar incisor hypomineralisation in Northern England and its relationship to socioeconomic status and water fluoridation." Int J Paediatr Dent **22**(4): 250-257.

Barron, M. J., S. J. Brookes, J. Kirkham, R. C. Shore, C. Hunt, A. Mironov, N. J. Kingswell, J. Maycock, C. A. Shuttleworth and M. J. Dixon (2010). "A mutation in the mouse Amelx tri-tyrosyl domain results in impaired

secretion of amelogenin and phenocopies human X-linked amelogenesis imperfecta." Hum Mol Genet **19**(7): 1230-1247.

Bartlett, J. D. (2013). "Dental Enamel Development: Proteinases and Their Enamel Matrix Substrates." ISRN Dent **2013**: 684607.

Bartlett, J. D. and J. P. Simmer (1999). "Proteinases in developing dental enamel." Crit Rev Oral Biol Med **10**(4): 425-441.

Bartlett, J. D. and J. P. Simmer (2015). "New perspectives on amelotin and amelogenesis." J Dent Res **94**(5): 642-644.

Bartlett, J. D. and C. E. Smith (2013). "Modulation of cell-cell junctional complexes by matrix metalloproteinases." J Dent Res **92**(1): 10-17.

Bartlett, J. D., Y. Yamakoshi, J. P. Simmer, A. Nanci and C. E. Smith (2011). "MMP20 cleaves E-cadherin and influences ameloblast development." Cells Tissues Organs **194**(2-4): 222-226.

Bateman, J. B., E. D. Riedner, L. S. Levin and I. H. Maumenee (1980). "Heterogeneity of retinal degeneration and hearing impairment syndromes." Am J Ophthalmol **90**(6): 755-767.

Baux, D., C. Blanchet, C. Hamel, I. Meunier, L. Larrieu, V. Faugere, C. Vache, P. Castorina, B. Puech, D. Bonneau, S. Malcolm, M. Claustres and A. F. Roux (2014). "Enrichment of LOVD-USHbases with 152 USH2A genotypes defines an extensive mutational spectrum and highlights missense hotspots." Hum Mutat **35**(10): 1179-1186.

Baux, D., L. Larrieu, C. Blanchet, C. Hamel, S. Ben Salah, A. Vielle, B. Gilbert-Dussardier, M. Holder, P. Calvas, N. Philip, P. Edery, D. Bonneau, M. Claustres, S. Malcolm and A. F. Roux (2007). "Molecular and in silico analyses of the full-length isoform of usherin identify new pathogenic alleles in Usher type II patients." Hum Mutat **28**(8): 781-789.

Beall, A. E. (2007). "Can a new smile make you look more intelligent and successful?" Dent Clin North Am **51**(2): 289-297, vii.

Bei, M. and R. Maas (1998). "FGFs and BMP4 induce both Msx1-independent and Msx1-dependent signaling pathways in early tooth development." Development **125**(21): 4325-4333.

Belkadi, A., A. Bolze, Y. Itan, A. Cobat, Q. B. Vincent, A. Antipenko, L. Shang, B. Boisson, J. L. Casanova and L. Abel (2015). "Whole-genome

sequencing is more powerful than whole-exome sequencing for detecting exome variants." Proc Natl Acad Sci U S A **112**(17): 5473-5478.

Bennett, S. R., J. C. Folk, A. E. Kimura, S. R. Russell, E. M. Stone and E. M. Raphtis (1990). "Autosomal dominant neovascular inflammatory vitreoretinopathy." Ophthalmology **97**(9): 1125-1135; discussion 1135-1126.

Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, E. C. M. Chiara, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. Huw Jones, G. D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. Chris Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy, V. H. Rae, S. R. Rawlings, A. Chiva Rodriguez, P. M. Roe, J. Rogers, M. C. Rogert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S.

W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. Ernest Sohna Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. Vandevondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin and A. J. Smith (2008). "Accurate whole human genome sequencing using reversible terminator chemistry." Nature **456**(7218): 53-59.

Berendse, K., M. S. Ebberink, L. Ijlst, B. T. Poll-The, R. J. Wanders and H. R. Waterham (2013). "Arginine improves peroxisome functioning in cells from patients with a mild peroxisome biogenesis disorder." Orphanet J Rare Dis **8**: 138.

Berendse, K., M. Engelen, S. Ferdinandusse, C. B. Majoie, H. R. Waterham, F. M. Vaz, J. H. Koelman, P. G. Barth, R. J. Wanders and B. T. Poll-The (2016). "Zellweger spectrum disorders: clinical manifestations in patients surviving into adulthood." J Inherit Metab Dis **39**(1): 93-106.

Berkman, M. D. and A. Singer (1971). "Demonstration of the lyon hypothesis in X-linked dominant hypoplastic amelogenesis imperfecta." Birth Defects Orig Artic Ser **7**(7): 204-209.

Beyeler, M., C. Schild, R. Lutz, M. Chiquet and B. Trueb (2010). "Identification of a fibronectin interaction site in the extracellular matrix protein ameloblastin." Exp Cell Res **316**(7): 1202-1212.

Bhattacharya, G., C. Miller, W. J. Kimberling, M. M. Jablonski and D. Cosgrove (2002). "Localization and expression of usherin: a novel basement membrane protein defective in people with Usher's syndrome type IIa." Hear Res **163**(1-2): 1-11.

Biggs, L. C. and M. L. Mikkola (2014). "Early inductive events in ectodermal appendage morphogenesis." Semin Cell Dev Biol **25-26**: 11-21.

Birschmann, I., A. K. Stroobants, M. van den Berg, A. Schafer, K. Rosenkranz, W. H. Kunau and H. F. Tabak (2003). "Pex15p of Saccharomyces cerevisiae provides a molecular basis for recruitment of the AAA peroxin Pex6p to peroxisomal membranes." Mol Biol Cell **14**(6): 2226-2236.

Bonnet, C., M. Grati, S. Marlin, J. Levilliers, J. P. Hardelin, M. Parodi, M. Niasme-Grare, D. Zelenika, M. Delepine, D. Feldmann, L. Jonard, A. El-Amraoui, D. Weil, B. Delobel, C. Vincent, H. Dollfus, M. M. Eliot, A. David, C.

Calais, J. Vigneron, B. Montaut-Verient, D. Bonneau, J. Dubin, C. Thauvin, A. Duvillard, C. Francannet, T. Mom, D. Lacombe, F. Duriez, V. Drouin-Garraud, M. F. Thuillier-Obstoy, S. Sigaudy, A. M. Frances, P. Collignon, G. Challe, R. Couderc, M. Lathrop, J. A. Sahel, J. Weissenbach, C. Petit and F. Denoyelle (2011). "Complete exon sequencing of all known Usher syndrome genes greatly improves molecular diagnosis." Orphanet J Rare Dis **6**: 21.

Borovjagin, A. V., J. Dong, M. J. Passineau, C. Ren, E. Lamani, O. A. Mamaeva, H. Wu, E. Keyser, M. Murakami, S. Chen and M. MacDougall (2011). "Adenovirus gene transfer to amelogenesis imperfecta ameloblast-like cells." PLoS One **6**(10): e24281.

Boskey, A. L., M. Maresca, S. Doty, B. Sabsay and A. Veis (1990). "Concentration-dependent effects of dentin phosphophoryn in the regulation of in vitro hydroxyapatite formation and growth." Bone Miner **11**(1): 55-65.

Bosshardt, D. D. and N. P. Lang (2005). "The junctional epithelium: from health to disease." J Dent Res **84**(1): 9-20.

Boughman, J. A., M. Vernon and K. A. Shaver (1983). "Usher syndrome: definition and estimate of prevalence from two high-risk populations." J Chronic Dis **36**(8): 595-603.

Braun, T. A., R. F. Mullins, A. H. Wagner, J. L. Andorf, R. M. Johnston, B. B. Bakall, A. P. Deluca, G. A. Fishman, B. L. Lam, R. G. Weleber, A. V. Cideciyan, S. G. Jacobson, V. C. Sheffield, B. A. Tucker and E. M. Stone (2013). "Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease." Hum Mol Genet **22**(25): 5136-5145.

Braverman, N., G. Steel, C. Obie, A. Moser, H. Moser, S. J. Gould and D. Valle (1997). "Human PEX7 encodes the peroxisomal PTS2 receptor and is responsible for rhizomelic chondrodysplasia punctata." Nat Genet **15**(4): 369-376.

Bronckers, A. L., D. Lyaruu, R. Jalali, J. F. Medina, B. Zandieh-Doulabi and P. K. DenBesten (2015). "Ameloblast Modulation and Transport of Cl-, Na+, and K+ during Amelogenesis." J Dent Res **94**(12): 1740-1747.

Brookes, S. J., M. J. Barron, R. Boot-Handford, J. Kirkham and M. J. Dixon (2014). "Endoplasmic reticulum stress in amelogenesis imperfecta and phenotypic rescue using 4-phenylbutyrate." Hum Mol Genet **23**(9): 2468-2480.

Brookes, S. J., C. Robinson, J. Kirkham and W. A. Bonass (1995). "Biochemistry and molecular biology of amelogenin proteins of developing dental enamel." Arch Oral Biol **40**(1): 1-14.

Brooks, S. A., T. M. Carter, L. Royle, D. J. Harvey, S. A. Fry, C. Kinch, R. A. Dwek and P. M. Rudd (2008). "Altered glycosylation of proteins in cancer: what is the potential for new anti-tumour strategies." Anticancer Agents Med Chem **8**(1): 2-21.

Brunton, P. A., R. P. Davies, J. L. Burke, A. Smith, A. Aggeli, S. J. Brookes and J. Kirkham (2013). "Treatment of early caries lesions using biomimetic self-assembling peptides--a clinical safety trial." Br Dent J **215**(4): E6.

Callaway, E. (2016). "UK scientists gain licence to edit genes in human embryos." Nature **530**(7588): 18.

Caton, J. and A. S. Tucker (2009). "Current knowledge of tooth development: patterning and mineralization of the murine dentition." J Anat **214**(4): 502-515.

Chan, H. C., N. M. Estrella, R. N. Milkovich, J. W. Kim, J. P. Simmer and J. C. Hu (2011). "Target gene analyses of 39 amelogenesis imperfecta kindreds." Eur J Oral Sci **119 Suppl 1**: 311-323.

Chan, H. C., L. Mai, A. Oikonomopoulou, H. L. Chan, A. S. Richardson, S. K. Wang, J. P. Simmer and J. C. Hu (2010). "Altered enamelin phosphorylation site causes amelogenesis imperfecta." J Dent Res **89**(7): 695-699.

Chang, C. C., W. H. Lee, H. Moser, D. Valle and S. J. Gould (1997). "Isolation of the human PEX12 gene, mutated in group 3 of the peroxisome biogenesis disorders." Nat Genet **15**(4): 385-388.

Chaves, A. M., A. Rosenblatt and O. F. Oliveira (2007). "Enamel defects and its relation to life course events in primary dentition of Brazilian children: a longitudinal study." Community Dent Health **24**(1): 31-36.

Chen, C. L., K. M. Bromley, J. Moradian-Oldak and J. J. DeYoreo (2011). "In situ AFM study of amelogenin assembly and disassembly dynamics on charged surfaces provides insights on matrix protein self-assembly." J Am Chem Soc **133**(43): 17406-17413.

Chen, L. S., R. I. Couwenhoven, D. Hsu, W. Luo and M. L. Snead (1992). "Maintenance of amelogenin gene expression by transformed epithelial cells of mouse enamel organ." Arch Oral Biol **37**(10): 771-778.

Chen, Y., M. Bei, I. Woo, I. Satokata and R. Maas (1996). "Msx1 controls inductive signaling in mammalian tooth morphogenesis." Development **122**(10): 3035-3044.

Cheon, J. Y., J. Mozersky and R. Cook-Deegan (2014). "Variants of uncertain significance in BRCA: a harbinger of ethical and policy issues to come?" Genome Med **6**(12): 121.

Chilamakuri, C. S., S. Lorenz, M. A. Madoui, D. Vodak, J. Sun, E. Hovig, O. Myklebost and L. A. Meza-Zepeda (2014). "Performance comparison of four exome capture systems for deep sequencing." BMC Genomics **15**: 449.

Cho, S. Y., Y. P. Chang, J. Y. Park, H. D. Park, Y. B. Sohn, S. W. Park, S. H. Kim, S. Ji, S. J. Kim, E. W. Choi, C. H. Kim, A. R. Ko, K. H. Paik and D. K. Jin (2011). "Two novel PEX1 mutations in a patient with Zellweger syndrome: the first Korean case confirmed by biochemical, and molecular evidence." Ann Clin Lab Sci **41**(2): 182-187.

Choi, Y., G. E. Sims, S. Murphy, J. R. Miller and A. P. Chan (2012). "Predicting the functional effect of amino acid substitutions and indels." PLoS One **7**(10): e46688.

Ciniawsky, S., I. Grimm, D. Saffian, W. Girzalsky, R. Erdmann and P. Wendler (2015). "Molecular snapshots of the Pex1/6 AAA+ complex in action." Nat Commun **6**: 7331.

Coffield, K. D., C. Phillips, M. Brady, M. W. Roberts, R. P. Strauss and J. T. Wright (2005). "The psychosocial impact of developmental dental defects in people with hereditary amelogenesis imperfecta." J Am Dent Assoc **136**(5): 620-630.

Collins, C. S. and S. J. Gould (1999). "Identification of a common PEX1 mutation in Zellweger syndrome." Hum Mutat **14**(1): 45-53.

Conrad, D. F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T. D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C. H. Ihm, K. Kristiansson, D. G. Macarthur, J. R. Macdonald, I. Onyiah, A. W. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, C. Tyler-Smith, N. P. Carter,

C. Lee, S. W. Scherer and M. E. Hurles (2010). "Origins and functional impact of copy number variation in the human genome." Nature **464**(7289): 704-712.

Cooper, G. M., B. P. Coe, S. Girirajan, J. A. Rosenfeld, T. H. Vu, C. Baker, C. Williams, H. Stalker, R. Hamid, V. Hannig, H. Abdel-Hamid, P. Bader, E. McCracken, D. Niyazov, K. Leppig, H. Thiese, M. Hummel, N. Alexander, J. Gorski, J. Kussmann, V. Shashi, K. Johnson, C. Rehder, B. C. Ballif, L. G. Shaffer and E. E. Eichler (2011). "A copy number variation morbidity map of developmental delay." Nat Genet **43**(9): 838-846.

Crawford, P. J., M. Aldred and A. Bloch-Zupan (2007). "Amelogenesis imperfecta." Orphanet J Rare Dis **2**: 17.

Crombie, F., D. Manton and N. Kilpatrick (2009). "Aetiology of molar-incisor hypomineralization: a critical review." Int J Paediatr Dent **19**(2): 73-83.

Cruvinel, V. R., D. B. Gravina, T. D. Azevedo, C. S. Rezende, A. C. Bezerra and O. A. Toledo (2012). "Prevalence of enamel defects and associated risk factors in both dentitions in preterm and full term born children." J Appl Oral Sci **20**(3): 310-317.

Cui, J., J. Xiao, V. S. Tagliabracci, J. Wen, M. Rahdar and J. E. Dixon (2015). "A secretory kinase complex regulates extracellular protein phosphorylation." Elife **4**: e06120.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean and R. Durbin (2011). "The variant call format and VCFtools." Bioinformatics **27**(15): 2156-2158.

Daubert, D. M., J. L. Kelley, Y. G. Udod, C. Habor, C. G. Kleist, I. K. Furman, I. N. Tikonov, W. J. Swanson and F. A. Roberts (2016). "Human enamel thickness and ENAM polymorphism." Int J Oral Sci **8**(2): 93-97.

Davit-Beal, T., A. S. Tucker and J. Y. Sire (2009). "Loss of teeth and enamel in tetrapods: fossil record, genetic data and morphological adaptations." J Anat **214**(4): 477-501.

Dawes, C., A. M. Pedersen, A. Villa, J. Ekstrom, G. B. Proctor, A. Vissink, D. Aframian, R. McGowan, A. Aliko, N. Narayana, Y. W. Sia, R. K. Joshi, S. B. Jensen, A. R. Kerr and A. Wolff (2015). "The functions of human saliva: A review sponsored by the World Workshop on Oral Medicine VI." Arch Oral Biol **60**(6): 863-874.

de la Dure-Molla, M., M. Quentric, P. M. Yamaguti, A. C. Acevedo, A. J. Mighell, M. Vikkula, M. Huckert, A. Berdal and A. Bloch-Zupan (2014). "Pathognomonic oral profile of Enamel Renal Syndrome (ERS) caused by recessive FAM20A mutations." Orphanet J Rare Dis **9**: 84.

de la Pena, V. A. and M. C. Valea (2011). "Treatment of enamel hypoplasia in a patient with Usher syndrome." J Am Dent Assoc **142**(8): 938-941.

de Ligt, J., P. M. Boone, R. Pfundt, L. E. Vissers, T. Richmond, J. Geoghegan, K. O'Moore, N. de Leeuw, C. Shaw, H. G. Brunner, J. R. Lupski, J. A. Veltman and J. Y. Hehir-Kwa (2013). "Detection of clinically relevant copy number variants with whole-exome sequencing." Hum Mutat **34**(10): 1439-1448.

Dean, M. C. (1998). "Comparative observations on the spacing of short-period (von Ebner's) lines in dentine." Arch Oral Biol **43**(12): 1009-1021.

Debela, M., N. Beaufort, V. Magdolen, N. M. Schechter, C. S. Craik, M. Schmitt, W. Bode and P. Goettig (2008). "Structures and specificity of the human kallikrein-related peptidases KLK 4, 5, 6, and 7." Biol Chem **389**(6): 623-632.

Debela, M., V. Magdolen, V. Grimminger, C. Sommerhoff, A. Messerschmidt, R. Huber, R. Friedrich, W. Bode and P. Goettig (2006). "Crystal structures of human tissue kallikrein 4: activity modulation by a specific zinc binding site." J Mol Biol **362**(5): 1094-1107.

Den Besten, P. K., C. H. Mathews, C. Gao and W. Li (1998). "Primary culture and characterization of enamel organ epithelial cells." Connect Tissue Res **38**(1-4): 3-8; discussion 35-41.

DenBesten, P. K., C. Gao, W. Li, C. H. Mathews and D. C. Gruenert (1999). "Development and characterization of an SV40 immortalized porcine ameloblast-like cell line." Eur J Oral Sci **107**(4): 276-281.

DenBesten, P. K., D. Machule, Y. Zhang, Q. Yan and W. Li (2005). "Characterization of human primary enamel organ epithelial cells in vitro." Arch Oral Biol **50**(8): 689-694.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D.

Altshuler and M. J. Daly (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." Nat Genet **43**(5): 491-498.

Diekwisch, T. G., B. J. Berman, S. Gentner and H. C. Slavkin (1995). "Initial enamel crystals are not spatially associated with mineralized dentine." Cell Tissue Res **279**(1): 149-167.

Dodt, G., N. Braverman, C. Wong, A. Moser, H. W. Moser, P. Watkins, D. Valle and S. J. Gould (1995). "Mutations in the PTS1 receptor gene, PXR1, define complementation group 2 of the peroxisome biogenesis disorders." Nat Genet **9**(2): 115-125.

Dodt, G. and S. J. Gould (1996). "Multiple PEX genes are required for proper subcellular distribution and stability of Pex5p, the PTS1 receptor: evidence that PTS1 protein import is mediated by a cycling receptor." J Cell Biol **135**(6 Pt 2): 1763-1774.

Dos Santos Neves, J., R. M. Wazen, S. Kuroda, S. Francis Zalzal, P. Moffatt and A. Nanci (2012). "Odontogenic ameloblast-associated and amelotin are novel basal lamina components." Histochem Cell Biol **137**(3): 329-338.

Dreyer, B., V. Brox, L. Tranebjaerg, T. Rosenberg, A. M. Sadeghi, C. Moller and O. Nilssen (2008). "Spectrum of USH2A mutations in Scandinavian patients with Usher syndrome type II." Hum Mutat **29**(3): 451.

Dreyer, B., L. Tranebjaerg, V. Brox, T. Rosenberg, C. Moller, M. Beneyto, M. D. Weston, W. J. Kimberling, C. W. Cremers, X. Z. Liu and O. Nilssen (2001). "A common ancestral origin of the frequent and widespread 2299delG USH2A mutation." Am J Hum Genet **69**(1): 228-234.

Du, C., G. Falini, S. Fermani, C. Abbott and J. Moradian-Oldak (2005). "Supramolecular assembly of amelogenin nanospheres into birefringent microribbons." Science **307**(5714): 1450-1454.

Ebberink, M. S., J. Kofster, R. J. Wanders and H. R. Waterham (2010). "Spectrum of PEX6 mutations in Zellweger syndrome spectrum patients." Hum Mutat **31**(1): E1058-1070.

Ebberink, M. S., J. Koster, G. Visser, F. Spronsen, I. Stolte-Dijkstra, G. P. Smit, J. M. Fock, S. Kemp, R. J. Wanders and H. R. Waterham (2012). "A novel defect of peroxisome division due to a homozygous non-sense mutation in the PEX11beta gene." J Med Genet **49**(5): 307-313.

Ebberink, M. S., P. A. Mooijer, J. Gootjes, J. Koster, R. J. Wanders and H. R. Waterham (2011). "Genetic classification and mutational spectrum of more than 600 patients with a Zellweger syndrome spectrum disorder." Hum Mutat **32**(1): 59-69.

Ebermann, I., J. B. Phillips, M. C. Liebau, R. K. Koenekoop, B. Schermer, I. Lopez, E. Schafer, A. F. Roux, C. Dafinger, A. Bernd, E. Zrenner, M. Claustres, B. Blanco, G. Nurnberg, P. Nurnberg, R. Ruland, M. Westerfield, T. Benzing and H. J. Bolz (2010). "PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome." J Clin Invest **120**(6): 1812-1823.

Ek, J., B. F. Kase, A. Reith, I. Bjorkhem and J. I. Pedersen (1986). "Peroxisomal dysfunction in a boy with neurologic symptoms and amaurosis (Leber disease): clinical and biochemical findings similar to those observed in Zellweger syndrome." J Pediatr **108**(1): 19-24.

El-Sayed, W., D. A. Parry, R. C. Shore, M. Ahmed, H. Jafri, Y. Rashid, S. Al-Bahlani, S. Al Harasi, J. Kirkham, C. F. Inglehearn and A. J. Mighell (2009). "Mutations in the beta propeller WDR72 cause autosomal-recessive hypomaturation amelogenesis imperfecta." Am J Hum Genet **85**(5): 699-705.

Ellingford, J. M., S. Barton, S. Bhaskar, J. O'Sullivan, S. G. Williams, J. A. Lamb, B. Panda, P. I. Sergouniotis, R. L. Gillespie, S. P. Daiger, G. Hall, T. Gale, I. C. Lloyd, P. N. Bishop, S. C. Ramsden and G. C. Black (2016a). "Molecular findings from 537 individuals with inherited retinal disease." J Med Genet.

Ellingford, J. M., S. Barton, S. Bhaskar, S. G. Williams, P. I. Sergouniotis, J. O'Sullivan, J. A. Lamb, R. Perveen, G. Hall, W. G. Newman, P. N. Bishop, S. A. Roberts, R. Leach, R. Tearle, S. Bayliss, S. C. Ramsden, A. H. Nemeth and G. C. Black (2016b). "Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease." Ophthalmology **123**(5): 1143-1150.

Eudy, J. D., M. D. Weston, S. Yao, D. M. Hoover, H. L. Rehm, M. Ma-Edmonds, D. Yan, I. Ahmad, J. J. Cheng, C. Ayuso, C. Cremers, S. Davenport, C. Moller, C. B. Talmadge, K. W. Beisel, M. Tamayo, C. C. Morton, A. Swaroop, W. J. Kimberling and J. Sumegi (1998). "Mutation of a gene encoding a protein with extracellular matrix motifs in Usher syndrome type IIa." Science **280**(5370): 1753-1757.

Exome Aggregation Consortium, L. Lek, K. Karczewski, E. Minikel, K. Samocha, E. Banks, T. Fennell, A. O'Donnell-Luria, J. Ware, A. Hill, B. Cummings, T. Tukiainen, D. Birnbaum, J. Kosmicki, L. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. Kurki, A. Levy Moonshine, P. Natarajan, L. Orozco, G. Peloso, R. Poplin, M. Rivas, V. Ruano-Rubio, S. Rose, D. Ruderfer, K. Shakir, P. Stenson, C. Stevens, B. Thomas, G. Tiao, M. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, E. Roberto, J. Florez, S. Gabriel, G. Getz, S. Glatt, C. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. McCarthy, D. McGovern, R. McPherson, B. Neale, S. Palotie, D. Saleheen, J. Scharf, P. Sklar, P. Sullivan, J. Tuomilehto, M. Tsuang, H. Watkins, J. Wilson, M. Daly and D. MacArthur (2016) "Analysis of protein-coding genetic variation in 60,706 humans." bioRxiv DOI: http://dx.doi.org/10.1101/030338.

Fang, P. A., R. S. Lam and E. Beniash (2011). "Relationships between dentin and enamel mineral at the dentino-enamel boundary: electron tomography and high-resolution transmission electron microscopy study." Eur J Oral Sci **119 Suppl 1**: 120-124.

Faust, P. L., D. Banka, R. Siriratsivawong, V. G. Ng and T. M. Wikander (2005). "Peroxisome biogenesis disorders: the role of peroxisomes and metabolic dysfunction in developing brain." J Inherit Metab Dis **28**(3): 369-383.

Fawcett, R. S., S. Linford and D. L. Stulberg (2004). "Nail abnormalities: clues to systemic disease." Am Fam Physician **69**(6): 1417-1424.

Ferguson, C. A., A. S. Tucker and P. T. Sharpe (2000). "Temporospatial cell interactions regulating mandibular and maxillary arch patterning." Development **127**(2): 403-412.

Finn, S. B. (1938). "Hereditary Opalescent Dentin. I. An Analysis of the Literature on Hereditary Anomalies of Tooth Color." JADA **25**(8): 1240-1249.

Fong, C. D., R. Cerny, L. Hammarstrom and I. Slaby (1998). "Sequential expression of an amelin gene in mesenchymal and epithelial cells during odontogenesis in rats." Eur J Oral Sci **106 Suppl 1**: 324-330.

Fong, C. D. and L. Hammarstrom (2000). "Expression of amelin and amelogenin in epithelial root sheath remnants of fully formed rat molars." Oral Surg Oral Med Oral Pathol Oral Radiol Endod **90**(2): 218-223.

Fong, C. D., I. Slaby and L. Hammarstrom (1996). "Amelin: an enamel-related protein, transcribed in the cells of epithelial root sheath." J Bone Miner Res **11**(7): 892-898.

Foster, B. L., F. H. Nociti, Jr. and M. J. Somerman (2014). "The rachitic tooth." Endocr Rev **35**(1): 1-34.

Fujiki, Y., N. Miyata, N. Matsumoto and S. Tamura (2008). "Dynamic and functional assembly of the AAA peroxins, Pex1p and Pex6p, and their membrane receptor Pex26p involved in shuttling of the PTS1 receptor Pex5p in peroxisome biogenesis." Biochem Soc Trans **36**(Pt 1): 109-113.

Fukae, M., T. Tanabe, T. Uchida, Y. Yamakoshi and M. Shimizu (1993). "Enamelins in the newly formed bovine enamel." Calcif Tissue Int **53**(4): 257-261.

Fukumoto, S., T. Kiba, B. Hall, N. Iehara, T. Nakamura, G. Longenecker, P. H. Krebsbach, A. Nanci, A. B. Kulkarni and Y. Yamada (2004). "Ameloblastin is a cell adhesion molecule required for maintaining the differentiation state of ameloblasts." J Cell Biol **167**(5): 973-983.

Gaildrat, P., A. Killian, A. Martins, I. Tournier, T. Frebourg and M. Tosi (2010). "Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants." Methods Mol Biol **653**: 249-257.

Ganss, B. and N. Abbarin (2014). "Maturation and beyond: proteins in the developmental continuum from enamel epithelium to junctional epithelium." Front Physiol **5**: 371.

Gardner, B. M., S. Chowdhury, G. C. Lander and A. Martin (2015). "The Pex1/Pex6 complex is a heterohexameric AAA+ motor with alternating and highly coordinated subunits." J Mol Biol **427**(6 Pt B): 1375-1388.

Gasse, B., Y. Chiari, J. Silvent, T. Davit-Beal and J. Y. Sire (2015a). "Amelotin: an enamel matrix protein that experienced distinct evolutionary histories in amphibians, sauropsids and mammals." BMC Evol Biol **15**: 47.

Gasse, B., E. Karayigit, E. Mathieu, S. Jung, A. Garret, M. Huckert, S. Morkmued, C. Schneider, L. Vidal, J. Hemmerle, J. Y. Sire and A. Bloch-Zupan

(2013). "Homozygous and compound heterozygous MMP20 mutations in amelogenesis imperfecta." J Dent Res **92**(7): 598-603.

Gasse, B., X. Liu, E. Corre and J. Y. Sire (2015b). "Amelotin Gene Structure and Expression during Enamel Formation in the Opossum Monodelphis domestica." PLoS One **10**(7): e0133314.

Gasse, B., J. Silvent and J. Y. Sire (2012). "Evolutionary analysis suggests that AMTN is enamel-specific and a candidate for AI." J Dent Res **91**(11): 1085-1089.

Genome of the Netherlands Consortium (2014). "Whole-genome sequence variation, population structure and demographic history of the Dutch population." Nat Genet **46**(8): 818-825.

Ghenea, S. T., M.; Motoyama, J.; Sasamoto, K.; Kunau, W.-H.; Kamiryo, T.; Bun-ya, M. (2001). "The cDNA Sequence and Expression of the AAA-family Peroxin Genes pex-1 and pex-6 from the Nematode Caenorhabditis elegans." Zoological Science **18**(5): 675-681.

Gillings, B. and M. Buonocore (1961). "An investigation of enamel thickness in human lower incisor teeth." J Dent Res **40**: 105-118.

Giros, M., F. Roels, J. Prats, M. Ruiz, A. Ribes, M. Espeel, R. J. Wanders, R. B. Schutgens and T. Pampols (1996). "Long survival in a case of peroxisomal biogenesis disorder with peroxisome mosaicism in the liver." Ann N Y Acad Sci **804**: 747-749.

Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, C. Russ, S. Gabriel, D. B. Jaffe, E. S. Lander and C. Nusbaum (2009). "Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing." Nat Biotechnol **27**(2): 182-189.

Goetz, S. C. and K. V. Anderson (2010). "The primary cilium: a signalling centre during vertebrate development." Nat Rev Genet **11**(5): 331-344.

Goldberg, M. and D. Septier (2002). "Phospholipids in amelogenesis and dentinogenesis." Crit Rev Oral Biol Med **13**(3): 276-290.

Goller, T., U. K. Seibold, E. Kremmer, W. Voos and W. Kolanus (2013). "Atad3 function is essential for early post-implantation development in the mouse." PLoS One **8**(1): e54799.

Gould, S., G. Raymond and D. Valle (2001). The Peroxisome Biogenesis Disorders in The Metabolic and Molecular Bases of Inherited Disease. New York, NY, USA, McGraw-Hill.

Grant, P., B. Ahlemeyer, S. Karnati, T. Berg, I. Stelzig, A. Nenicu, K. Kuchelmeister, D. I. Crane and E. Baumgart-Vogt (2013). "The biogenesis protein PEX14 is an optimal marker for the identification and localization of peroxisomes in different cell types, tissues, and species in morphological studies." Histochem Cell Biol **140**(4): 423-442.

Green, R. C., J. S. Berg, W. W. Grody, S. S. Kalia, B. R. Korf, C. L. Martin, A. L. McGuire, R. L. Nussbaum, J. M. O'Daniel, K. E. Ormond, H. L. Rehm, M. S. Watson, M. S. Williams and L. G. Biesecker (2013). "ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing." Genet Med **15**(7): 565-574.

Gregory-Evans, C. Y., M. Moosajee, M. D. Hodges, D. S. Mackay, L. Game, N. Vargesson, A. Bloch-Zupan, F. Ruschendorf, L. Santos-Pinto, G. Wackens and K. Gregory-Evans (2007). "SNP genome scanning localizes oto-dental syndrome to chromosome 11q13 and microdeletions at this locus implicate FGF3 in dental and inner-ear disease and FADD in ocular coloboma." Hum Mol Genet **16**(20): 2482-2493.

Grillet, N., W. Xiong, A. Reynolds, P. Kazmierczak, T. Sato, C. Lillo, R. A. Dumont, E. Hintermann, A. Sczaniecka, M. Schwander, D. Williams, B. Kachar, P. G. Gillespie and U. Muller (2009). "Harmonin mutations cause mechanotransduction defects in cochlear hair cells." Neuron **62**(3): 375-387.

Guan, X. and J. D. Bartlett (2013). "MMP20 modulates cadherin expression in ameloblasts as enamel develops." J Dent Res **92**(12): 1123-1128.

Guan, X., M. Xu, S. E. Millar and J. D. Bartlett (2016). "Beta-catenin is essential for ameloblast movement during enamel development." Eur J Oral Sci **124**(3): 221-227.

Gudbjartsson, D. F., P. Sulem, H. Helgason, A. Gylfason, S. A. Gudjonsson, F. Zink, A. Oddson, G. Magnusson, B. V. Halldorsson, E. Hjartarson, G. T. Sigurdsson, A. Kong, A. Helgason, G. Masson, O. T. Magnusson, U. Thorsteinsdottir and K. Stefansson (2015). "Sequence variants from whole genome sequencing a large group of Icelanders." Sci Data **2**: 150011.

Guergolette, R. P., C. C. Dezan, W. T. Frossard, F. B. Ferreira, A. Cerci Neto and K. B. Fernandes (2009). "Prevalence of developmental defects of enamel in children and adolescents with asthma." J Bras Pneumol **35**(4): 295-300.

Guo, Y., Q. Sheng, D. C. Samuels, B. Lehmann, J. A. Bauer, J. Pietenpol and Y. Shyr (2013). "Comparative study of exome copy number variation estimation tools using array comparative genomic hybridization as control." Biomed Res Int **2013**: 915636.

Handsaker, R. E., V. Van Doren, J. R. Berman, G. Genovese, S. Kashin, L. M. Boettger and S. A. McCarroll (2015). "Large multiallelic copy number variations in humans." Nat Genet **47**(3): 296-303.

Hanson, P. I. and S. W. Whiteheart (2005). "AAA+ proteins: have engine, will work." Nat Rev Mol Cell Biol **6**(7): 519-529.

Harada, H., Y. Ichimori, T. Yokohama-Tamaki, H. Ohshima, S. Kawano, K. Katsube and S. Wakisaka (2006). "Stratum intermedium lineage diverges from ameloblast lineage via Notch signaling." Biochem Biophys Res Commun **340**(2): 611-616.

Harbuz, R., F. Bilan, D. Couet, V. Charraud, A. Kitzis and B. Gilbert-Dussardier (2013). "Osteogenesis imperfecta, tricho-dento-osseous syndrome and intellectual disability: a familial case with 17q21.33-q22 (COL1A1 and DLX3) deletion and 7q32.3-q33 duplication resulting from a reciprocal interchromosomal insertion." Am J Med Genet A **161A**(10): 2504-2511.

Hart, P. S., M. J. Aldred, P. J. Crawford, N. J. Wright, T. C. Hart and J. T. Wright (2002). "Amelogenesis imperfecta phenotype-genotype correlations with two amelogenin gene mutations." Arch Oral Biol **47**(4): 261-265.

Hart, P. S., T. C. Hart, M. D. Michalec, O. H. Ryu, D. Simmons, S. Hong and J. T. Wright (2004). "Mutation in kallikrein 4 causes autosomal recessive hypomaturation amelogenesis imperfecta." J Med Genet **41**(7): 545-549.

Hart, S., T. Hart, C. Gibson and J. T. Wright (2000). "Mutational analysis of X-linked amelogenesis imperfecta in multiple families." Arch Oral Biol **45**(1): 79-86.

Hart, T. C., P. S. Hart, M. C. Gorry, M. D. Michalec, O. H. Ryu, C. Uygur, D. Ozdemir, S. Firatli, G. Aren and E. Firatli (2003). "Novel ENAM mutation

responsible for autosomal recessive amelogenesis imperfecta and localised enamel defects." J Med Genet **40**(12): 900-906.

Harutunian, K., R. Figueiredo and C. Gay-Escoda (2011). "Tuberous sclerosis complex with oral manifestations: a case report and literature review." Med Oral Patol Oral Cir Bucal **16**(4): e478-481.

Hashem, A., A. Kelly, B. O'Connell and M. O'Sullivan (2013). "Impact of moderate and severe hypodontia and amelogenesis imperfecta on quality of life and self-esteem of adult patients." J Dent **41**(8): 689-694.

Hayes, J. L., A. Tzika, H. Thygesen, S. Berri, H. M. Wood, S. Hewitt, M. Pendlebury, A. Coates, L. Willoughby, C. M. Watson, P. Rabbitts, P. Roberts and G. R. Taylor (2013). "Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation." Genomics **102**(3): 174-181.

Haze, A., A. L. Taylor, A. Blumenfeld, E. Rosenfeld, Y. Leiser, L. Dafni, B. Shay, Y. Gruenbaum-Cohen, E. Fermon, S. Haegewald, J. P. Bernimoulin and D. Deutsch (2007). "Amelogenin expression in long bone and cartilage cells and in bone marrow progenitor cells." Anat Rec (Hoboken) **290**(5): 455-460.

He, J., C. C. Mao, A. Reyes, H. Sembongi, M. Di Re, C. Granycome, A. B. Clippingdale, I. M. Fearnley, M. Harbour, A. J. Robinson, S. Reichelt, J. N. Spelbrink, J. E. Walker and I. J. Holt (2007). "The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization." J Cell Biol **176**(2): 141-146.

Heckel, J. and P. Jugelt (1984). "Quantitative analysis of bulk samples without standards by using peak-to-background ratios." X-Ray Spectrometry **13**: 159-165.

Heimler, A., J. E. Fox, J. E. Hershey and P. Crespi (1991). "Sensorineural hearing loss, enamel hypoplasia, and nail abnormalities in sibs." Am J Med Genet **39**(2): 192-195.

Hentschel, J., D. Tatun, D. Parkhomchuk, I. Kurth, B. Schimmel, R. Heinrich-Weltzien, S. Bertzbach, H. Peters and C. Beetz (2016). "Identification of the first multi-exonic WDR72 deletion in isolated amelogenesis imperfecta, and generation of a WDR72-specific copy number screening tool." Gene.

Hetz, C., F. Martinon, D. Rodriguez and L. H. Glimcher (2011). "The unfolded protein response: integrating stress signals through the stress sensor IRE1alpha." Physiol Rev **91**(4): 1219-1243.

Hiller, C. R., C. Robinson and J. A. Weatherell (1975). "Variations in the composition of developing rat incisor enamel." Calcif Tissue Res **18**(1): 1-12.

Holcroft, J. and B. Ganss (2011). "Identification of amelotin- and ODAM-interacting enamel matrix proteins using the yeast two-hybrid system." Eur J Oral Sci **119 Suppl 1**: 301-306.

Holm, S. (1979). "A simple sequentially rejective multiple test procedure." Scandinavian Journal of Statistics **6**(2): 65-70.

Holme, R. H. and K. P. Steel (2002). "Stereocilia defects in waltzer (Cdh23), shaker1 (Myo7a) and double waltzer/shaker1 mutant mice." Hear Res **169**(1-2): 13-23.

Holtzman, N. A. (2013). "ACMG recommendations on incidental findings are flawed scientifically and ethically." Genet Med **15**(9): 750-751.

Hong, L., S. M. Levy, J. J. Warren and B. Broffitt (2011). "Amoxicillin use during early childhood and fluorosis of later developing tooth zones." J Public Health Dent **71**(3): 229-235.

Honsho, M., S. Tamura, N. Shimozawa, Y. Suzuki, N. Kondo and Y. Fujiki (1998). "Mutation in PEX16 is causal in the peroxisome-deficient Zellweger syndrome of complementation group D." Am J Hum Genet **63**(6): 1622-1630.

Hu, C. C., M. Fukae, T. Uchida, Q. Qian, C. H. Zhang, O. H. Ryu, T. Tanabe, Y. Yamakoshi, C. Murakami, N. Dohi, M. Shimizu and J. P. Simmer (1997). "Sheathlin: cloning, cDNA/polypeptide sequences, and immunolocalization of porcine enamel sheath proteins." J Dent Res **76**(2): 648-657.

Hu, J. C., Y. Hu, C. E. Smith, M. D. McKee, J. T. Wright, Y. Yamakoshi, P. Papagerakis, G. K. Hunter, J. Q. Feng, F. Yamakoshi and J. P. Simmer (2008). "Enamel defects and ameloblast-specific expression in Enam knock-out/lacz knock-in mice." J Biol Chem **283**(16): 10858-10871.

Hu, J. C., X. Sun, C. Zhang, S. Liu, J. D. Bartlett and J. P. Simmer (2002). "Enamelysin and kallikrein-4 mRNA expression in developing mouse molars." Eur J Oral Sci **110**(4): 307-315.

Hu, J. C. and Y. Yamakoshi (2003). "Enamelin and autosomal-dominant amelogenesis imperfecta." Crit Rev Oral Biol Med **14**(6): 387-398.

Hu, J. C., C. Zhang, X. Sun, Y. Yang, X. Cao, O. Ryu and J. P. Simmer (2000). "Characterization of the mouse and human PRSS17 genes, their relationship to other serine proteases, and the expression of PRSS17 in developing mouse incisors." Gene **251**(1): 1-8.

Huckert, M., C. Stoetzel, S. Morkmued, V. Laugel-Haushalter, V. Geoffroy, J. Muller, F. Clauss, M. K. Prasad, F. Obry, J. L. Raymond, M. Switala, Y. Alembik, S. Soskin, E. Mathieu, J. Hemmerle, J. L. Weickert, B. B. Dabovic, D. B. Rifkin, A. Dheedene, E. Boudin, O. Caluseriu, M. C. Cholette, R. McLeod, R. Antequera, M. P. Gelle, J. L. Coeuriot, L. F. Jacquelin, I. Bailleul-Forestier, M. C. Maniere, W. Van Hul, D. Bertola, P. Dolle, A. Verloes, G. Mortier, H. Dollfus and A. Bloch-Zupan (2015). "Mutations in the latent TGF-beta binding protein 3 (LTBP3) gene cause brachyolmia with amelogenesis imperfecta." Hum Mol Genet **24**(11): 3038-3049.

Hutchins, M. U., M. Veenhuis and D. J. Klionsky (1999). "Peroxisome degradation in Saccharomyces cerevisiae is dependent on machinery of macroautophagy and the Cvt pathway." J Cell Sci **112 (Pt 22)**: 4079-4087.

Huysseune, A. and J. Y. Sire (1998). "Evolution of patterns and processes in teeth and tooth-related tissues in non-mammalian vertebrates." Eur J Oral Sci **106 Suppl 1**: 437-481.

Iizuka, S., Y. Kudo, M. Yoshida, T. Tsunematsu, Y. Yoshiko, T. Uchida, I. Ogawa, M. Miyauchi and T. Takata (2011). "Ameloblastin regulates osteogenic differentiation by inhibiting Src kinase via cross talk between integrin beta1 and CD63." Mol Cell Biol **31**(4): 783-792.

Ikeda, E., R. Morita, K. Nakao, K. Ishida, T. Nakamura, T. Takano-Yamamoto, M. Ogawa, M. Mizuno, S. Kasugai and T. Tsuji (2009). "Fully functional bioengineered tooth replacement as an organ replacement therapy." Proc Natl Acad Sci U S A **106**(32): 13475-13480.

Imamura, A., N. Shimozawa, Y. Suzuki, Z. Zhang, T. Tsukamoto, Y. Fujiki, T. Orii, T. Osumi, R. J. Wanders and N. Kondo (2000). "Temperature-sensitive mutation of PEX6 in peroxisome biogenesis disorders in complementation group C (CG-C): comparative study of PEX6 and PEX1." Pediatr Res **48**(4): 541-545.

Imamura, A., S. Tamura, N. Shimozawa, Y. Suzuki, Z. Zhang, T. Tsukamoto, T. Orii, N. Kondo, T. Osumi and Y. Fujiki (1998a). "Temperature-sensitive mutation in PEX1 moderates the phenotypes of peroxisome deficiency disorders." Hum Mol Genet **7**(13): 2089-2094.

Imamura, A., T. Tsukamoto, N. Shimozawa, Y. Suzuki, Z. Zhang, T. Imanaka, Y. Fujiki, T. Orii, N. Kondo and T. Osumi (1998b). "Temperature-sensitive phenotypes of peroxisome-assembly processes represent the milder forms of human peroxisome-biogenesis disorders." Am J Hum Genet **62**(6): 1539-1543.

Innis, J. W., P. A. Sieving, P. McMillan and R. A. Weatherly (1998). "Apparently new syndrome of sensorineural hearing loss, retinal pigment epithelium lesions, and discolored teeth." Am J Med Genet **75**(1): 13-17.

Ionita-Laza, I., K. McCallum, B. Xu and J. D. Buxbaum (2016). "A spectral approach integrating functional genomic annotations for coding and noncoding variants." Nat Genet **48**(2): 214-220.

Ishikawa, H. O., A. Xu, E. Ogura, G. Manning and K. D. Irvine (2012). "The Raine syndrome protein FAM20C is a Golgi kinase that phosphorylates bio-mineralization proteins." PLoS One **7**(8): e42988.

Isken, O. and L. E. Maquat (2007). "Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function." Genes Dev **21**(15): 1833-1856.

Isken, O. and L. E. Maquat (2008). "The multiple lives of NMD factors: balancing roles in gene and genome regulation." Nat Rev Genet **9**(9): 699-712.

Iwasaki, K., E. Bajenova, E. Somogyi-Ganss, M. Miller, V. Nguyen, H. Nourkeyhani, Y. Gao, M. Wendel and B. Ganss (2005). "Amelotin--a Novel Secreted, Ameloblast-specific Protein." J Dent Res **84**(12): 1127-1132.

Iwase, M., S. Kaneko, H. Kim, Y. Satta and N. Takahata (2007). "Evolutionary history of sex-linked mammalian amelogenin genes." Cells Tissues Organs **186**(1): 49-59.

Iwata, T., Y. Yamakoshi, J. C. Hu, I. Ishikawa, J. D. Bartlett, P. H. Krebsbach and J. P. Simmer (2007). "Processing of ameloblastin by MMP-20." J Dent Res **86**(2): 153-157.

Jacobsen, P. E., T. B. Henriksen, D. Haubek and J. R. Ostergaard (2013). "Developmental enamel defects in children prenatally exposed to anti-epileptic drugs." PLoS One **8**(3): e58213.

Jaijo, T., E. Aller, S. Oltra, M. Beneyto, C. Najera, C. Ayuso, M. Baiget, M. Carballo, G. Antinolo, D. Valverde, F. Moreno, C. Vilela, H. Perez-Garrigues, A. Navea and J. M. Millan (2006). "Mutation profile of the MYO7A gene in Spanish patients with Usher syndrome type I." Hum Mutat **27**(3): 290-291.

Jalili, I. K. (2010). "Cone-rod dystrophy and amelogenesis imperfecta (Jalili syndrome): phenotypes and environs." Eye (Lond) **24**(11): 1659-1668.

Janones, D. S., L. F. Massa and V. E. Arana-Chavez (2005). "Immunocytochemical examination of the presence of amelogenin during the root development of rat molars." Arch Oral Biol **50**(5): 527-532.

Jarvinen, E., M. Tummers and I. Thesleff (2009). "The role of the dental lamina in mammalian tooth replacement." J Exp Zool B Mol Dev Evol **312B**(4): 281-291.

Jaureguiberry, G., M. De la Dure-Molla, D. Parry, M. Quentric, N. Himmerkus, T. Koike, J. Poulter, E. Klootwijk, S. L. Robinette, A. J. Howie, V. Patel, M. L. Figueres, H. C. Stanescu, N. Issler, J. K. Nicholson, D. Bockenhauer, C. Laing, S. B. Walsh, D. A. McCredie, S. Povey, A. Asselin, A. Picard, A. Coulomb, A. J. Medlar, I. Bailleul-Forestier, A. Verloes, C. Le Caignec, G. Roussey, J. Guiol, B. Isidor, C. Logan, R. Shore, C. Johnson, C. Inglehearn, S. Al-Bahlani, M. Schmittbuhl, F. Clauss, M. Huckert, V. Laugel, E. Ginglinger, S. Pajarola, G. Sparta, D. Bartholdi, A. Rauch, M. C. Addor, P. M. Yamaguti, H. P. Safatle, A. C. Acevedo, H. Martelli-Junior, P. E. dos Santos Netos, R. D. Coletta, S. Gruessel, C. Sandmann, D. Ruehmann, C. B. Langman, S. J. Scheinman, D. Ozdemir-Ozenen, T. C. Hart, P. S. Hart, U. Neugebauer, E. Schlatter, P. Houillier, W. A. Gahl, M. Vikkula, A. Bloch-Zupan, M. Bleich, H. Kitagawa, R. J. Unwin, A. Mighell, A. Berdal and R. Kleta (2012). "Nephrocalcinosis (enamel renal syndrome) caused by autosomal recessive FAM20A mutations." Nephron Physiol **122**(1-2): 1-6.

Jedeon, K., M. De la Dure-Molla, S. J. Brookes, S. Loiodice, C. Marciano, J. Kirkham, M. C. Canivenc-Lavier, S. Boudalia, R. Berges, H. Harada, A. Berdal and S. Babajko (2013). "Enamel defects reflect perinatal exposure to bisphenol A." Am J Pathol **183**(1): 108-118.

Jeremias, F., R. A. Pierri, J. F. Souza, C. M. Fragelli, M. Restrepo, L. S. Finoti, D. G. Bussaneli, R. C. Cordeiro, R. Secolin, C. V. Maurer-Morelli, R. M. Scarel-Caminaga and L. Santos-Pinto (2016). "Family-Based Genetic Association for Molar-Incisor Hypomineralization." Caries Res **50**(3): 310-318.

Jernvall, J., P. Kettunen, I. Karavanova, L. B. Martin and I. Thesleff (1994). "Evidence for the role of the enamel knot as a control center in mammalian tooth cusp formation: non-dividing cells express growth stimulating Fgf-4 gene." Int J Dev Biol **38**(3): 463-469.

Josephsen, K. and O. Fejerskov (1977). "Ameloblast modulation in the maturation zone of the rat incisor enamel organ. A light and electron microscopic study." J Anat **124**(Pt 1): 45-70.

Josephsen, K., Y. Takano, S. Frische, J. Praetorius, S. Nielsen, T. Aoba and O. Fejerskov (2010). "Ion transporters in secretory and cyclically modulating ameloblasts: a new hypothesis for cellular control of preeruptive enamel maturation." Am J Physiol Cell Physiol **299**(6): C1299-1307.

Kachi, S., Y. Oshima, N. Esumi, M. Kachi, B. Rogers, D. J. Zack and P. A. Campochiaro (2005). "Nonviral ocular gene transfer." Gene Ther **12**(10): 843-851.

Kallenbach, E. (1977). "Fine structure of secretory ameloblasts in the kitten." Am J Anat **148**(4): 479-511.

Katsura, K., J. Horst, D. Chandra, T. Le, Y. Nakano, Y. Zhang, O. Horst, L. Zhu, M. Le and P. K. DenBesten (2014). "WDR72 models of structure and function: A stage-specific regulator of enamel mineralization." Matrix Biol.

Kawano, S., T. Morotomi, T. Toyono, N. Nakamura, T. Uchida, M. Ohishi, K. Toyoshima and H. Harada (2002). "Establishment of dental epithelial cell line (HAT-7) and the cell differentiation dependent on Notch signaling pathway." Connect Tissue Res **43**(2-3): 409-412.

Kawasaki, K. and T. Suzuki (2011). "Molecular evolution of matrix metalloproteinase 20." Eur J Oral Sci **119 Suppl 1**: 247-253.

Kawasaki, K., T. Suzuki and K. M. Weiss (2004). "Genetic basis for the evolution of vertebrate mineralized tissue." Proc Natl Acad Sci U S A **101**(31): 11356-11361.

Kawasaki, K. and K. M. Weiss (2003). "Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster." Proc Natl Acad Sci U S A **100**(7): 4060-4065.

Kawasaki, K. and K. M. Weiss (2006). "Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family." J Exp Zool B Mol Dev Evol **306**(3): 295-316.

Keats, B. J. and D. P. Corey (1999). "The usher syndromes." Am J Med Genet **89**(3): 158-166.

Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler and D. Haussler (2002). "The human genome browser at UCSC." Genome Res **12**(6): 996-1006.

Kim, J. W., S. K. Lee, Z. H. Lee, J. C. Park, K. E. Lee, M. H. Lee, J. T. Park, B. M. Seo, J. C. Hu and J. P. Simmer (2008). "FAM83H mutations in families with autosomal-dominant hypocalcified amelogenesis imperfecta." Am J Hum Genet **82**(2): 489-494.

Kim, J. W., S. H. Nam, K. T. Jang, S. H. Lee, C. C. Kim, S. H. Hahn, J. C. Hu and J. P. Simmer (2004). "A novel splice acceptor mutation in the DSPP gene causing dentinogenesis imperfecta type II." Hum Genet **115**(3): 248-254.

Kim, J. W., F. Seymen, K. E. Lee, J. Ko, M. Yildirim, E. B. Tuna, K. Gencay, T. J. Shin, H. K. Kyun, J. P. Simmer and J. C. Hu (2013). "LAMB3 mutations causing autosomal-dominant amelogenesis imperfecta." J Dent Res **92**(10): 899-904.

Kim, J. W., J. P. Simmer, T. C. Hart, P. S. Hart, M. D. Ramaswami, J. D. Bartlett and J. C. Hu (2005). "MMP-20 mutation in autosomal recessive pigmented hypomaturation amelogenesis imperfecta." J Med Genet **42**(3): 271-275.

Kimberling, W. J., M. S. Hildebrand, A. E. Shearer, M. L. Jensen, J. A. Halder, K. Trzupek, E. S. Cohn, R. G. Weleber, E. M. Stone and R. J. Smith (2010). "Frequency of Usher syndrome in two pediatric populations: Implications for genetic screening of deaf and hard of hearing children." Genet Med **12**(8): 512-516.

Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper and J. Shendure (2014). "A general framework for estimating the relative pathogenicity of human genetic variants." Nat Genet **46**(3): 310-315.

Kirkham, J., J. Zhang, S. J. Brookes, R. C. Shore, S. R. Wood, D. A. Smith, M. L. Wallwork, O. H. Ryu and C. Robinson (2000). "Evidence for charge domains on developing enamel crystal surfaces." J Dent Res **79**(12): 1943-1947.

Kitagawa, M., T. Ando, A. Subarnbhesaj, T. Uchida, M. Miyauchi and T. Takata (2016). "N-terminal region of human ameloblastin synthetic peptide promotes bone formation." Odontology.

Kitagawa, M., S. Kitagawa, A. Nagasaki, M. Miyauchi, T. Uchida and T. Takata (2011). "Synthetic ameloblastin peptide stimulates differentiation of human periodontal ligament cells." Arch Oral Biol **56**(4): 374-379.

Kivirikko, S., J. A. McGrath, C. Baudoin, D. Aberdam, S. Ciatti, M. G. Dunnill, J. R. McMillan, R. A. Eady, J. P. Ortonne, G. Meneguzzi and et al. (1995). "A homozygous nonsense mutation in the alpha 3 chain gene of laminin 5 (LAMA3) in lethal (Herlitz) junctional epidermolysis bullosa." Hum Mol Genet **4**(5): 959-962.

Kjoelby, M., I. Thesleff, C. Sahlberg, O. Fejerskov and K. Josephsen (1994). "Degradation of the dental basement membrane during mouse tooth development in vitro." Int J Dev Biol **38**(3): 455-462.

Kobayashi, K., Y. Yamakoshi, J. C. Hu, K. Gomi, T. Arai, M. Fukae, P. H. Krebsbach and J. P. Simmer (2007). "Splicing determines the glycosylation state of ameloblastin." J Dent Res **86**(10): 962-967.

Kolkman, J. A. and W. P. Stemmer (2001). "Directed evolution of proteins by exon shuffling." Nat Biotechnol **19**(5): 423-428.

Kollar, E. J. and G. R. Baird (1970). "Tissue interactions in embryonic mouse tooth germs. I. Reorganization of the dental epithelium during tooth-germ reconstruction." J Embryol Exp Morphol **24**(1): 159-171.

Komatsu, N., M. Takata, N. Otsuki, T. Toyama, R. Ohka, K. Takehara and K. Saijoh (2003). "Expression and localization of tissue kallikrein mRNAs in human epidermis and appendages." J Invest Dermatol **121**(3): 542-549.

Korbel, J. O., A. E. Urban, J. P. Affourtit, B. Godwin, F. Grubert, J. F. Simons, P. M. Kim, D. Palejev, N. J. Carriero, L. Du, B. E. Taillon, Z. Chen, A.

Tanzer, A. C. Saunders, J. Chi, F. Yang, N. P. Carter, M. E. Hurles, S. M. Weissman, T. T. Harkins, M. B. Gerstein, M. Egholm and M. Snyder (2007). "Paired-end mapping reveals extensive structural variation in the human genome." Science **318**(5849): 420-426.

Krause, C., H. Rosewich and J. Gartner (2009). "Rational diagnostic strategy for Zellweger syndrome spectrum patients." Eur J Hum Genet **17**(6): 741-748.

Krebsbach, P. H., S. K. Lee, Y. Matsuki, C. A. Kozak, K. M. Yamada and Y. Yamada (1996). "Full-length sequence, localization, and chromosomal mapping of ameloblastin. A novel tooth-specific gene." J Biol Chem **271**(8): 4431-4435.

Krumm, N., P. H. Sudmant, A. Ko, B. J. O'Roak, M. Malig, B. P. Coe, A. R. Quinlan, D. A. Nickerson and E. E. Eichler (2012). "Copy number variation detection and genotyping from exome sequence data." Genome Res **22**(8): 1525-1532.

Krysko, O., L. Hulshagen, A. Janssen, G. Schutz, R. Klein, M. De Bruycker, M. Espeel, P. Gressens and M. Baes (2007). "Neocortical and cerebellar developmental abnormalities in conditions of selective elimination of peroxisomes from brain or from liver." J Neurosci Res **85**(1): 58-72.

Kuga, T., M. Sasaki, T. Mikami, Y. Miake, J. Adachi, M. Shimizu, Y. Saito, M. Koura, Y. Takeda, J. Matsuda, T. Tomonaga and Y. Nakayama (2016). "FAM83H and casein kinase I regulate the organization of the keratin cytoskeleton and formation of desmosomes." Sci Rep **6**: 26557.

Kuilman, T., A. Velds, K. Kemper, M. Ranzani, L. Bombardelli, M. Hoogstraat, E. Nevedomskaya, G. Xu, J. de Ruiter, M. P. Lolkema, B. Ylstra, J. Jonkers, S. Rottenberg, L. F. Wessels, D. J. Adams, D. S. Peeper and O. Krijgsman (2015). "CopywriteR: DNA copy number detection from off-target sequence data." Genome Biol **16**: 49.

Kumar, P., S. Henikoff and P. C. Ng (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-1081.

Lacruz, R. S., Y. Nakayama, J. Holcroft, V. Nguyen, E. Somogyi-Ganss, M. L. Snead, S. N. White, M. L. Paine and B. Ganss (2012a). "Targeted

overexpression of amelotin disrupts the microstructure of dental enamel." PLoS One **7**(4): e35200.

Lacruz, R. S., C. E. Smith, P. Bringas, Jr., Y. B. Chen, S. M. Smith, M. L. Snead, I. Kurtz, J. G. Hacia, M. J. Hubbard and M. L. Paine (2012b). "Identification of novel candidate genes involved in mineralization of dental enamel by genome-wide transcript profiling." J Cell Physiol **227**(5): 2264-2275.

Lacruz, R. S., C. E. Smith, Y. B. Chen, M. J. Hubbard, J. G. Hacia and M. L. Paine (2011). "Gene-expression analysis of early- and late-maturation-stage rat enamel organ." Eur J Oral Sci **119 Suppl 1**: 149-157.

Lagerlof, F. (1983). "Effects of flow rate and pH on calcium phosphate saturation in human parotid saliva." Caries Res **17**(5): 403-411.

Lagerstrom, M., N. Dahl, Y. Nakahori, Y. Nakagome, B. Backman, U. Landegren and U. Pettersson (1991). "A deletion in the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1)." Genomics **10**(4): 971-975.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock,

Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen and J. Szustakowki (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Landrum, M. J., J. M. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin-Salomon, W. Rubinstein and D. R. Maglott (2016). "ClinVar: public archive of interpretations of clinically relevant variants." Nucleic Acids Res **44**(D1): D862-868.

Le, M. H., R. Warotayanont, J. Stahl, P. K. Den Besten and Y. Nakano (2016). "Amelogenin Exon4 Forms a Novel miRNA That Directs Ameloblast and Osteoblast Differentiation." J Dent Res **95**(4): 423-429.

Le Quesne Stabej, P., Z. Saihan, N. Rangesh, H. B. Steele-Stallard, J. Ambrose, A. Coffey, J. Emmerson, E. Haralambous, Y. Hughes, K. P. Steel, L. M. Luxon, A. R. Webster and M. Bitner-Glindzicz (2012). "Comprehensive sequence analysis of nine Usher syndrome genes in the UK National Collaborative Usher Study." J Med Genet **49**(1): 27-36.

Lee, S. K., P. H. Krebsbach, Y. Matsuki, A. Nanci, K. M. Yamada and Y. Yamada (1996). "Ameloblastin expression in rat incisors and human tooth germs." Int J Dev Biol **40**(6): 1141-1150.

Lee, S. K., F. Seymen, H. Y. Kang, K. E. Lee, K. Gencay, B. Tuna and J. W. Kim (2010). "MMP20 hemopexin domain mutation in amelogenesis imperfecta." J Dent Res **89**(1): 46-50.

Lenassi, E., Z. Saihan, M. Bitner-Glindzicz and A. R. Webster (2014). "The effect of the common c.2299delG mutation in USH2A on RNA splicing." Exp Eye Res **122**: 9-12.

Lenassi, E., A. Vincent, Z. Li, Z. Saihan, A. J. Coffey, H. B. Steele-Stallard, A. T. Moore, K. P. Steel, L. M. Luxon, E. Heon, M. Bitner-Glindzicz and A. R. Webster (2015). "A detailed clinical and molecular survey of subjects with nonsyndromic USH2A retinopathy reveals an allelic hierarchy of disease-causing variants." Eur J Hum Genet **23**(10): 1318-1327.

Lertsirivorakul, J., M. Wongswadiwat and P. Treesuwan (2014). "Oral manifestations and dental management of a child with Zellweger syndrome." Spec Care Dentist **34**(1): 46-50.

Letunic, I., T. Doerks and P. Bork (2015). "SMART: recent updates, new developments and status in 2015." Nucleic Acids Res **43**(Database issue): D257-260.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, Z., M. Yu and W. Tian (2013). "An inductive signalling network regulates mammalian tooth morphogenesis with implications for tooth regeneration." Cell Prolif **46**(5): 501-508.

Liang, T., T. Meng, S. Wang, C. Qin and Y. Lu (2016). "The LPV Motif Is Essential for the Efficient Export of Secretory DMP1 From the Endoplasmic Reticulum." J Cell Physiol **231**(7): 1468-1475.

Lima, L. H., I. A. Barbazetto, R. Chen, L. A. Yannuzzi, S. H. Tsang and R. F. Spaide (2011). "Macular dystrophy in Heimler syndrome." Ophthalmic Genet **32**(2): 97-100.

Liu, B., S. Chen, D. Cheng, W. Jing and J. A. Helms (2014). "Primary cilia integrate hedgehog and Wnt signaling during tooth development." J Dent Res **93**(5): 475-482.

Liu, X., O. V. Bulgakov, K. N. Darrow, B. Pawlyk, M. Adamian, M. C. Liberman and T. Li (2007). "Usherin is required for maintenance of retinal photoreceptors and normal development of cochlear hair cells." Proc Natl Acad Sci U S A **104**(11): 4413-4418.

Liu, X. Z., J. Walsh, P. Mburu, J. Kendrick-Jones, M. J. Cope, K. P. Steel and S. D. Brown (1997a). "Mutations in the myosin VIIA gene cause non-syndromic recessive deafness." Nat Genet **16**(2): 188-190.

Liu, X. Z., J. Walsh, Y. Tamagawa, K. Kitamura, M. Nishizawa, K. P. Steel and S. D. Brown (1997b). "Autosomal dominant non-syndromic deafness caused by a mutation in the myosin VIIA gene." Nat Genet **17**(3): 268-269.

Liu, Y., M. Jiang, W. Hao, W. Liu, L. Tang, H. Liu and Y. Jin (2013). "Skin epithelial cells as possible substitutes for ameloblasts during tooth regeneration." J Tissue Eng Regen Med **7**(12): 934-943.

Llano, E., A. M. Pendas, V. Knauper, T. Sorsa, T. Salo, E. Salido, G. Murphy, J. P. Simmer, J. D. Bartlett and C. Lopez-Otin (1997). "Identification and structural and functional characterization of human enamelysin (MMP-20)." Biochemistry **36**(49): 15101-15108.

Logan, W. H. G. and R. Kronfeld (1933). "Development of the Human Jaws and Surrounding Structures from Birth to the Age of Fifteen Years." The Journal of the American Dental Association **20**(3): 379-428.

Lu, Y., P. Papagerakis, Y. Yamakoshi, J. C. Hu, J. D. Bartlett and J. P. Simmer (2008). "Functions of KLK4 and MMP-20 in dental enamel formation." Biol Chem **389**(6): 695-700.

Lundwall, A. (2013). "Old genes and new genes: the evolution of the kallikrein locus." Thromb Haemost **110**(3): 469-475.

Lupas, A. N. and J. Martin (2002). "AAA proteins." Curr Opin Struct Biol **12**(6): 746-753.

Lygidakis, N. A., G. Dimou and D. Marinou (2008). "Molar-incisor-hypomineralisation (MIH). A retrospective clinical study in Greek children. II. Possible medical aetiological factors." Eur Arch Paediatr Dent **9**(4): 207-217.

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

MacDougall, M., D. Simmons, T. T. Gu, K. Forsman-Semb, C. K. Mardh, M. Mesbah, N. Forest, P. H. Krebsbach, Y. Yamada and A. Berdal (2000). "Cloning, characterization and immunolocalization of human ameloblastin." Eur J Oral Sci **108**(4): 303-310.

MacDougall, M., D. Simmons, X. Luan, J. Nydegger, J. Feng and T. T. Gu (1997). "Dentin phosphoprotein and dentin sialoprotein are cleavage products expressed from a single transcript coded by a gene on human chromosome 4. Dentin phosphoprotein DNA sequence determination." J Biol Chem **272**(2): 835-842.

Maerker, T., E. van Wijk, N. Overlack, F. F. Kersten, J. McGee, T. Goldmann, E. Sehn, R. Roepman, E. J. Walsh, H. Kremer and U. Wolfrum (2008). "A novel Usher protein network at the periciliary reloading point between molecular transport machineries in vertebrate photoreceptor cells." Hum Mol Genet **17**(1): 71-86.

Mahajan, V. B., J. M. Skeie, A. G. Bassuk, J. H. Fingert, T. A. Braun, H. T. Daggett, J. C. Folk, V. C. Sheffield and E. M. Stone (2012). "Calpain-5 mutations cause autoimmune uveitis, retinal neovascularization, and photoreceptor degeneration." PLoS Genet **8**(10): e1003001.

Majewski, J. and J. Ott (2002). "Distribution and characterization of regulatory elements in the human genome." Genome Res **12**(12): 1827-1836.

Majewski, J., Z. Wang, I. Lopez, S. Al Humaid, H. Ren, J. Racine, A. Bazinet, G. Mitchel, N. Braverman and R. K. Koenekoop (2011). "A new ocular phenotype associated with an unexpected but known systemic disorder and mutation: novel use of genomic diagnostics and exome sequencing." J Med Genet **48**(9): 593-596.

Mardh, C. K., B. Backman, G. Holmgren, J. C. Hu, J. P. Simmer and K. Forsman-Semb (2002). "A nonsense mutation in the enamelin gene causes local hypoplastic autosomal dominant amelogenesis imperfecta (AIH2)." Hum Mol Genet **11**(9): 1069-1074.

Mathur, P. and J. Yang (2015). "Usher syndrome: Hearing loss, retinal degeneration and associated abnormalities." Biochim Biophys Acta **1852**(3): 406-420.

Matsumoto, N., S. Tamura and Y. Fujiki (2003). "The pathogenic peroxin Pex26p recruits the Pex1p-Pex6p AAA ATPase complexes to peroxisomes." Nat Cell Biol **5**(5): 454-460.

Matsuo, S., H. Ichikawa, S. Wakisaka and M. Akai (1992). "Changes of cytochemical properties in the Golgi apparatus during in vivo differentiation of the ameloblast in developing rat molar tooth germs." Anat Rec **234**(4): 469-478.

Matsuzono, Y., N. Kinoshita, S. Tamura, N. Shimozawa, M. Hamasaki, K. Ghaedi, R. J. Wanders, Y. Suzuki, N. Kondo and Y. Fujiki (1999). "Human PEX19: cDNA cloning by functional complementation, mutation analysis in a patient with Zellweger syndrome, and potential role in peroxisomal membrane assembly." Proc Natl Acad Sci U S A **96**(5): 2116-2121.

Maurano, M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul and J. A. Stamatoyannopoulos (2012). "Systematic localization of common disease-associated variation in regulatory DNA." Science **337**(6099): 1190-1195.

Maxwell, M. A., T. Allen, P. B. Solly, T. Svingen, B. C. Paton and D. I. Crane (2002). "Novel PEX1 mutations and genotype-phenotype correlations in Australasian peroxisome biogenesis disorder patients." Hum Mutat **20**(5): 342-351.

Maxwell, M. A., P. B. Leane, B. C. Paton and D. I. Crane (2005). "Novel PEX1 coding mutations and 5' UTR regulatory polymorphisms." Hum Mutat **26**(3): 279.

Maxwell, M. A., P. V. Nelson, S. J. Chin, B. C. Paton, W. F. Carey and D. I. Crane (1999). "A common PEX1 frameshift mutation in patients with disorders of peroxisome biogenesis correlates with the severe Zellweger syndrome phenotype." Hum Genet **105**(1-2): 38-44.

McCarroll, S. A., F. G. Kuruvilla, J. M. Korn, S. Cawley, J. Nemesh, A. Wysoker, M. H. Shapero, P. I. de Bakker, J. B. Maller, A. Kirby, A. L. Elliott, M. Parkin, E. Hubbell, T. Webster, R. Mei, J. Veitch, P. J. Collins, R. Handsaker, S. Lincoln, M. Nizzari, J. Blume, K. W. Jones, R. Rava, M. J. Daly, S. B. Gabriel and D. Altshuler (2008). "Integrated detection and population-genetic analysis of SNPs and copy number variation." Nat Genet **40**(10): 1166-1174.

McGee, T. L., B. J. Seyedahmadi, M. O. Sweeney, T. P. Dryja and E. L. Berson (2010). "Novel mutations in the long isoform of the USH2A gene in patients with Usher syndrome type II or non-syndromic retinitis pigmentosa." J Med Genet **47**(7): 499-506.

McGrath, J. A., B. Gatalica, A. M. Christiano, K. Li, K. Owaribe, J. R. McMillan, R. A. Eady and J. Uitto (1995). "Mutations in the 180-kD bullous pemphigoid antigen (BPAG2), a hemidesmosomal transmembrane collagen (COL17A1), in generalized atrophic benign epidermolysis bullosa." Nat Genet **11**(1): 83-86.

McGrath, J. A., B. Gatalica, K. Li, M. G. Dunnill, J. R. McMillan, A. M. Christiano, R. A. Eady and J. Uitto (1996). "Compound heterozygosity for a dominant glycine substitution and a recessive internal duplication mutation in the type XVII collagen gene results in junctional epidermolysis bullosa and abnormal dentition." Am J Pathol **148**(6): 1787-1796.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M. A. DePristo (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

McPherson, J. D., M. Marra, L. Hillier, R. H. Waterston, A. Chinwalla, J. Wallis, M. Sekhon, K. Wylie, E. R. Mardis, R. K. Wilson, R. Fulton, T. A. Kucaba, C. Wagner-McPherson, W. B. Barbazuk, S. G. Gregory, S. J. Humphray, L. French, R. S. Evans, G. Bethel, A. Whittaker, J. L. Holden, O. T. McCann, A. Dunham, C. Soderlund, C. E. Scott, D. R. Bentley, G. Schuler, H. C. Chen, W. Jang, E. D. Green, J. R. Idol, V. V. Maduro, K. T. Montgomery, E.

Lee, A. Miller, S. Emerling, Kucherlapati, R. Gibbs, S. Scherer, J. H. Gorrell, E. Sodergren, K. Clerc-Blankenburg, P. Tabor, S. Naylor, D. Garcia, P. J. de Jong, J. J. Catanese, N. Nowak, K. Osoegawa, S. Qin, L. Rowen, A. Madan, M. Dors, L. Hood, B. Trask, C. Friedman, H. Massa, V. G. Cheung, I. R. Kirsch, T. Reid, R. Yonescu, J. Weissenbach, T. Bruls, R. Heilig, E. Branscomb, A. Olsen, N. Doggett, J. F. Cheng, T. Hawkins, R. M. Myers, J. Shang, L. Ramirez, J. Schmutz, O. Velasquez, K. Dixon, N. E. Stone, D. R. Cox, D. Haussler, W. J. Kent, T. Furey, S. Rogic, S. Kennedy, S. Jones, A. Rosenthal, G. Wen, M. Schilhabel, G. Gloeckner, G. Nyakatura, R. Siebert, B. Schlegelberger, J. Korenberg, X. N. Chen, A. Fujiyama, M. Hattori, A. Toyoda, T. Yada, H. S. Park, Y. Sakaki, N. Shimizu, S. Asakawa, K. Kawasaki, T. Sasaki, A. Shintani, A. Shimizu, K. Shibuya, J. Kudoh, S. Minoshima, J. Ramser, P. Seranski, C. Hoff, A. Poustka, R. Reinhardt and H. Lehrach (2001). "A physical map of the human genome." Nature **409**(6822): 934-941.

Meynert, A. M., M. Ansari, D. R. FitzPatrick and M. S. Taylor (2014). "Variant detection sensitivity and biases in whole genome and exome sequencing." BMC Bioinformatics **15**: 247.

Michalski, N., V. Michel, A. Bahloul, G. Lefevre, J. Barral, H. Yagi, S. Chardenoux, D. Weil, P. Martin, J. P. Hardelin, M. Sato and C. Petit (2007). "Molecular characterization of the ankle-link complex in cochlear hair cells and its role in the hair bundle functioning." J Neurosci **27**(24): 6478-6488.

Michelakakis, H. M., D. I. Zafeiriou, M. S. Moraitou, J. Gootjes and R. J. Wanders (2004). "PEX1 deficiency presenting as Leber congenital amaurosis." Pediatr Neurol **31**(2): 146-149.

Millan, J. M., E. Aller, T. Jaijo, F. Blanco-Kelly, A. Gimenez-Pardo and C. Ayuso (2011). "An update on the genetics of usher syndrome." J Ophthalmol **2011**: 417217.

Mina, M. and E. J. Kollar (1987). "The induction of odontogenesis in non-dental mesenchyme combined with early murine mandibular arch epithelium." Arch Oral Biol **32**(2): 123-127.

Mitsiadis, T. A. and H. U. Luder (2011). "Genetic basis for tooth malformations: from mice to men and back again." Clin Genet **80**(4): 319-329.

Mitsiadis, T. A., A. S. Tucker, C. De Bari, M. T. Cobourne and D. P. Rice (2008). "A regulatory relationship between Tbx1 and FGF signaling during tooth morphogenesis and ameloblast lineage determination." Dev Biol **320**(1): 39-48.

Moffatt, P., C. E. Smith, R. St-Arnaud, D. Simmons, J. T. Wright and A. Nanci (2006). "Cloning of rat amelotin and localization of the protein to the basal lamina of maturation stage ameloblasts and junctional epithelium." Biochem J **399**(1): 37-46.

Mokry, M., H. Feitsma, I. J. Nijman, E. de Bruijn, P. J. van der Zaag, V. Guryev and E. Cuppen (2010). "Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries." Nucleic Acids Res **38**(10): e116.

Moradian-Oldak, J. (2012). "Protein-mediated enamel mineralization." Front Biosci (Landmark Ed) **17**: 1996-2023.

Moradian-Oldak, J., I. Jimenez, D. Maltby and A. G. Fincham (2001). "Controlled proteolysis of amelogenins reveals exposure of both carboxy- and amino-terminal regions." Biopolymers **58**(7): 606-616.

Morotomi, T., S. Kawano, T. Toyono, C. Kitamura, M. Terashita, T. Uchida, K. Toyoshima and H. Harada (2005). "In vitro differentiation of dental epithelial progenitor cells through epithelial-mesenchymal interactions." Arch Oral Biol **50**(8): 695-705.

Moser, H. W. (1999). "Genotype-phenotype correlations in disorders of peroxisome biogenesis." Mol Genet Metab **68**(2): 316-327.

Motley, A. M. and E. H. Hettema (2007). "Yeast peroxisomes multiply by growth and division." J Cell Biol **178**(3): 399-410.

Motley, A. M., E. H. Hettema, E. M. Hogenhout, P. Brites, A. L. ten Asbroek, F. A. Wijburg, F. Baas, H. S. Heijmans, H. F. Tabak, R. J. Wanders and B. Distel (1997). "Rhizomelic chondrodysplasia punctata is a peroxisomal protein targeting disease caused by a non-functional PTS2 receptor." Nat Genet **15**(4): 377-380.

Murakami, C., N. Dohi, M. Fukae, T. Tanabe, Y. Yamakoshi, K. Wakida, T. Satoda, O. Takahashi, M. Shimizu, O. H. Ryu, J. P. Simmer and T. Uchida (1997). "Immunochemical and immunohistochemical study of the 27- and 29-kDa calcium-binding proteins and related proteins in the porcine tooth germ." Histochem Cell Biol **107**(6): 485-494.

Murrell, D. F., A. M. Pasmooij, H. H. Pas, P. Marr, S. Klingberg, E. Pfendner, J. Uitto, S. Sadowski, F. Collins, R. Widmer and M. F. Jonkman (2007). "Retrospective diagnosis of fatal BP180-deficient non-Herlitz junctional epidermolysis bullosa suggested by immunofluorescence (IF) antigen-mapping of parental carriers bearing enamel defects." J Invest Dermatol **127**(7): 1772-1775.

Nakata, A., T. Kameda, H. Nagai, K. Ikegami, Y. Duan, K. Terada and T. Sugiyama (2003). "Establishment and characterization of a spontaneously immortalized mouse ameloblast-lineage cell line." Biochem Biophys Res Commun **308**(4): 834-839.

Nakayama, Y., J. Holcroft and B. Ganss (2015). "Enamel Hypomineralization and Structural Defects in Amelotin-deficient Mice." J Dent Res **94**(5): 697-705.

Nam, A. S., Y. Yin, Z. von Marschall and L. W. Fisher (2014). "Efficient trafficking of acidic proteins out of the endoplasmic reticulum involves a conserved amino terminal IleProVal (IPV)-like tripeptide motif." Connect Tissue Res **55 Suppl 1**: 138-141.

Nanci, A. (2012). Ten Cate's Oral Histology. St Louis, Missouri, USA, Mosby Elsevier.

Nashiro, C., A. Kashiwagi, T. Matsuzaki, S. Tamura and Y. Fujiki (2011). "Recruiting mechanism of the AAA peroxins, Pex1p and Pex6p, to Pex26p on the peroxisomal membrane." Traffic **12**(6): 774-788.

National Centre Biotechnology Information. (2009, 05/05/2009). "[dbsnp-announce] dbSNP Human Build 130." Retrieved 03/03/2016, 2016, from http://www.ncbi.nlm.nih.gov/mailman/pipermail/dbsnp-announce/2009q2/000088.html.

National Centre Biotechnology Information. (2016). "Variation Glossary." Retrieved 03/03/2016, 2016, from http://www.ncbi.nlm.nih.gov/variation/docs/glossary/.

Nelson, P. S., L. Gan, C. Ferguson, P. Moss, R. Gelinas, L. Hood and K. Wang (1999). "Molecular cloning and characterization of prostase, an androgen-regulated serine protease with prostate-restricted expression." Proc Natl Acad Sci U S A **96**(6): 3114-3119.

Neubuser, A., H. Peters, R. Balling and G. R. Martin (1997). "Antagonistic interactions between FGF and BMP signaling pathways: a mechanism for positioning the sites of tooth formation." Cell **90**(2): 247-255.

Ng, P. C., S. Levy, J. Huang, T. B. Stockwell, B. P. Walenz, K. Li, N. Axelrod, D. A. Busam, R. L. Strausberg and J. C. Venter (2008). "Genetic variation in an individual human exome." PLoS Genet **4**(8): e1000160.

Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. A. Nickerson, J. Shendure and M. J. Bamshad (2010). "Exome sequencing identifies the cause of a mendelian disorder." Nat Genet **42**(1): 30-35.

Nievers, M. G., R. Q. Schaapveld and A. Sonnenberg (1999). "Biology and function of hemidesmosomes." Matrix Biol **18**(1): 5-17.

Nunes, R. (2006). "Deafness, genetics and dysgenics." Med Health Care Philos **9**(1): 25-31.

Nunez, S. M., Y. P. Chun, B. Ganss, Y. Hu, A. S. Richardson, J. E. Schmitz, R. Fajardo, J. Yang, J. C. Hu and J. P. Simmer (2015). "Maturation stage enamel malformations in Amtn and Klk4 null mice." Matrix Biol.

Nuttall, J. M., A. M. Motley and E. H. Hettema (2014). "Deficiency of the exportomer components Pex1, Pex6, and Pex15 causes enhanced pexophagy in Saccharomyces cerevisiae." Autophagy **10**(5): 835-845.

O'Sullivan, J., C. C. Bitu, S. B. Daly, J. E. Urquhart, M. J. Barron, S. S. Bhaskar, H. Martelli-Junior, P. E. dos Santos Neto, M. A. Mansilla, J. C. Murray, R. D. Coletta, G. C. Black and M. J. Dixon (2011). "Whole-Exome sequencing identifies FAM20A mutations as a cause of amelogenesis imperfecta and gingival hyperplasia syndrome." Am J Hum Genet **88**(5): 616-620.

Ogura, T., S. W. Whiteheart and A. J. Wilkinson (2004). "Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases." J Struct Biol **146**(1-2): 106-112.

Ohshima, H., T. Maeda and Y. Takano (1998). "Cytochrome oxidase activity in the enamel organ during amelogenesis in rat incisors." Anat Rec **252**(4): 519-531.

Ohyama, Y., J. H. Lin, N. Govitvattana, I. P. Lin, S. Venkitapathi, A. Alamoudi, D. Husein, C. An, H. Hotta, M. Kaku and Y. Mochida (2016). "FAM20A binds to and regulates FAM20C localization." Sci Rep **6**: 27784.

Olivares-Navarette, R. H. H., S. L.; Almaguer-Flores, A.; Mauth, C.; Gemperli, A. C.; Boyan, B. D.; Schwartz, Z. (2013). "Amelogenin peptide extract increases differentiation and angiogenic and local factor production and inhibits apoptosis in human osteoblasts." ISRN Biomaterials **2013**(2013): 11.

Ong, K. R., S. Visram, S. McKaig and L. A. Brueton (2006). "Sensorineural deafness, enamel abnormalities and nail abnormalities: a case report of Heimler syndrome in identical twin girls." Eur J Med Genet **49**(2): 187-193.

Oshima, M., M. Mizuno, A. Imamura, M. Ogawa, M. Yasukawa, H. Yamazaki, R. Morita, E. Ikeda, K. Nakao, T. Takano-Yamamoto, S. Kasugai, M. Saito and T. Tsuji (2011). "Functional tooth regeneration using a bioengineered tooth unit as a mature organ replacement regenerative therapy." PLoS One **6**(7): e21531.

Ozdemir, D., P. S. Hart, E. Firatli, G. Aren, O. H. Ryu and T. C. Hart (2005a). "Phenotype of ENAM mutations is dosage-dependent." J Dent Res **84**(11): 1036-1041.

Ozdemir, D., P. S. Hart, O. H. Ryu, S. J. Choi, M. Ozdemir-Karatas, E. Firatli, N. Piesco and T. C. Hart (2005b). "MMP20 active-site mutation in hypomaturation amelogenesis imperfecta." J Dent Res **84**(11): 1031-1035.

Pang, A. W., J. R. MacDonald, D. Pinto, J. Wei, M. A. Rafiq, D. F. Conrad, H. Park, M. E. Hurles, C. Lee, J. C. Venter, E. F. Kirkness, S. Levy, L. Feuk and S. W. Scherer (2010). "Towards a comprehensive structural variation map of an individual human genome." Genome Biol **11**(5): R52.

Papagerakis, P., H. K. Lin, K. Y. Lee, Y. Hu, J. P. Simmer, J. D. Bartlett and J. C. Hu (2008). "Premature stop codon in MMP20 causing amelogenesis imperfecta." J Dent Res **87**(1): 56-59.

Parry, D. A., S. J. Brookes, C. V. Logan, J. A. Poulter, W. El-Sayed, S. Al-Bahlani, S. Al Harasi, J. Sayed, M. Raif el, R. C. Shore, M. Dashash, M. Barron, J. E. Morgan, I. M. Carr, G. R. Taylor, C. A. Johnson, M. J. Aldred, M. J. Dixon, J. T. Wright, J. Kirkham, C. F. Inglehearn and A. J. Mighell (2012). "Mutations in C4orf26, encoding a peptide with in vitro hydroxyapatite crystal

nucleation and growth activity, cause amelogenesis imperfecta." Am J Hum Genet **91**(3): 565-571.

Parry, D. A., J. A. Poulter, C. V. Logan, S. J. Brookes, H. Jafri, C. H. Ferguson, B. M. Anwari, Y. Rashid, H. Zhao, C. A. Johnson, C. F. Inglehearn and A. J. Mighell (2013). "Identification of mutations in SLC24A4, encoding a potassium-dependent sodium/calcium exchanger, as a cause of amelogenesis imperfecta." Am J Hum Genet **92**(2): 307-312.

Parry, D. A., C. E. L. Smith, W. El-Sayed, J. A. Poulter, R. C. Shore, C. V. Logan, C. Mogi, K. Sato, F. Okajima, A. Harada, H. Zhang, M. Koruyucu, F. Seymen, J. C.-C. Hu, J. P. Simmer, M. Ahmed, H. Jafri, C. A. Johnson, C. F. Inglehearn and A. J. Mighell (2016). "Mutations in the pH Sensing G-protein Coupled Receptor GPR68 cause Amelogenesis Imperfecta." American Journal of Human Genetics **99**(4): 984-990.

Pasmooij, A. M., H. H. Pas, G. H. Jansen, H. H. Lemmink and M. F. Jonkman (2007). "Localized and generalized forms of blistering in junctional epidermolysis bullosa due to COL17A1 mutations in the Netherlands." Br J Dermatol **156**(5): 861-870.

Patel, M., S. T. McDonnell, S. Iram and M. F. Chan (2013). "Amelogenesis imperfecta - lifelong management. Restorative management of the adult patient." Br Dent J **215**(9): 449-457.

Pavlic, A., T. Battelino, K. Trebusak Podkrajsek and M. Ovsenik (2011). "Craniofacial characteristics and genotypes of amelogenesis imperfecta patients." Eur J Orthod **33**(3): 325-331.

Peltz, S. W., M. Morsy, E. M. Welch and A. Jacobson (2013). "Ataluren as an agent for therapeutic nonsense suppression." Annu Rev Med **64**: 407-425.

Peters, H., A. Neubuser, K. Kratochwil and R. Balling (1998). "Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities." Genes Dev **12**(17): 2735-2747.

Pieke-Dahl, S., C. G. Moller, P. M. Kelley, L. M. Astuto, C. W. Cremers, M. B. Gorin and W. J. Kimberling (2000). "Genetic heterogeneity of Usher syndrome type II: localisation to chromosome 5q." J Med Genet **37**(4): 256-262.

Pispa, J., H. S. Jung, J. Jernvall, P. Kettunen, T. Mustonen, M. J. Tabata, J. Kere and I. Thesleff (1999). "Cusp patterning defect in Tabby mouse teeth and its partial rescue by FGF." Dev Biol **216**(2): 521-534.

Plagnol, V., J. Curtis, M. Epstein, K. Y. Mok, E. Stebbings, S. Grigoriadou, N. W. Wood, S. Hambleton, S. O. Burns, A. J. Thrasher, D. Kumararatne, R. Doffinger and S. Nejentsev (2012). "A robust model for read count data in exome sequencing experiments and implications for copy number variant calling." Bioinformatics **28**(21): 2747-2754.

Platta, H. W., S. Grunau, K. Rosenkranz, W. Girzalsky and R. Erdmann (2005). "Functional role of the AAA peroxins in dislocation of the cycling PTS1 receptor back to the cytosol." Nat Cell Biol **7**(8): 817-822.

Poll-The, B. T., J. Gootjes, M. Duran, J. B. De Klerk, L. J. Wenniger-Prick, R. J. Admiraal, H. R. Waterham, R. J. Wanders and P. G. Barth (2004). "Peroxisome biogenesis disorders with prolonged survival: phenotypic expression in a cohort of 31 patients." Am J Med Genet A **126A**(4): 333-338.

Pollak, C., M. Floy and B. Say (2003). "Sensorineural hearing loss and enamel hypoplasia with subtle nail findings: another family with Heimler's syndrome." Clin Dysmorphol **12**(1): 55-58.

Pope, F. M., A. C. Nicholls, J. McPheat, P. Talmud and R. Owen (1985). "Collagen genes and proteins in osteogenesis imperfecta." J Med Genet **22**(6): 466-478.

Poulter, J. A., S. J. Brookes, R. C. Shore, C. E. L. Smith, L. Abi Farraj, J. Kirkham, C. F. Inglehearn and A. J. Mighell (2014a). "A missense mutation in ITGB6 causes pitted hypomineralized amelogenesis imperfecta." Hum Mol Genet **23**(8): 2189-2197.

Poulter, J. A., W. El-Sayed, R. C. Shore, J. Kirkham, C. F. Inglehearn and A. J. Mighell (2014b). "Whole-exome sequencing, without prior linkage, identifies a mutation in LAMB3 as a cause of dominant hypoplastic amelogenesis imperfecta." Eur J Hum Genet **22**(1): 132-135.

Poulter, J. A., G. Murillo, S. J. Brookes, C. E. Smith, D. A. Parry, S. Silva, J. Kirkham, C. F. Inglehearn and A. J. Mighell (2014c). "Deletion of ameloblastin exon 6 is associated with amelogenesis imperfecta." Hum Mol Genet.

Poulter, J. A., G. Murillo, S. J. Brookes, C. E. L. Smith, D. A. Parry, S. Silva, J. Kirkham, C. F. Inglehearn and A. J. Mighell (2014d). "Deletion of

ameloblastin exon 6 is associated with amelogenesis imperfecta." Hum Mol Genet **23**(20): 5317-5324.

Poulter, J. A., C. E. L. Smith, G. Murrillo, S. Silva, S. Feather, M. Howell, L. Crinnion, D. T. Bonthron, I. M. Carr, C. M. Watson, C. F. Inglehearn and A. J. Mighell (2015). "A distinctive oral phenotype points to FAM20A mutations not identified by Sanger sequencing." Mol Genet Genomic Med **3**(6): 543-549.

Powers, J. M. and H. W. Moser (1998). "Peroxisomal disorders: genotype, phenotype, major neuropathologic lesions, and pathogenesis." Brain Pathol **8**(1): 101-120.

Prajapati, S., J. Tao, Q. Ruan, J. J. De Yoreo and J. Moradian-Oldak (2016). "Matrix metalloproteinase-20 mediates dental enamel biomineralization by preventing protein occlusion inside apatite crystals." Biomaterials **75**: 260-270.

Prasad, M. K., V. Geoffroy, S. Vicaire, B. Jost, M. Dumas, S. Le Gras, M. Switala, B. Gasse, V. Laugel-Haushalter, M. Paschaki, B. Leheup, D. Droz, A. Dalstein, A. Loing, B. Grollemund, M. Muller-Bolla, S. Lopez-Cazaux, M. Minoux, S. Jung, F. Obry, V. Vogt, J. L. Davideau, T. Davit-Beal, A. S. Kaiser, U. Moog, B. Richard, J. J. Morrier, J. P. Duprez, S. Odent, I. Bailleul-Forestier, M. M. Rousset, L. Merametdijan, A. Toutain, C. Joseph, F. Giuliano, J. C. Dahlet, A. Courval, M. El Alloussi, S. Laouina, S. Soskin, N. Guffon, A. Dieux, B. Doray, S. Feierabend, E. Ginglinger, B. Fournier, M. de la Dure Molla, Y. Alembik, C. Tardieu, F. Clauss, A. Berdal, C. Stoetzel, M. C. Maniere, H. Dollfus and A. Bloch-Zupan (2016). "A targeted next-generation sequencing assay for the molecular diagnosis of genetic disorders with orodental involvement." J Med Genet **53**(2): 98-110.

Prescott, R. S., R. Alsanea, M. I. Fayad, B. R. Johnson, C. S. Wenckus, J. Hao, A. S. John and A. George (2008). "In vivo generation of dental pulp-like tissue by using dental pulp stem cells, a collagen scaffold, and dentin matrix protein 1 after subcutaneous transplantation in mice." J Endod **34**(4): 421-426.

Preuss, N., U. Brosius, M. Biermanns, A. C. Muntau, E. Conzelmann and J. Gartner (2002). "PEX1 mutations in complementation group 1 of Zellweger spectrum patients correlate with severity of disease." Pediatr Res **51**(6): 706-714.

Price, J. A., D. W. Bowden, J. T. Wright, M. J. Pettenati and T. C. Hart (1998). "Identification of a mutation in DLX3 associated with tricho-dento-osseous (TDO) syndrome." Hum Mol Genet **7**(3): 563-569.

Price, J. A., J. T. Wright, S. J. Walker, P. J. Crawford, M. J. Aldred and T. C. Hart (1999). "Tricho-dento-osseous syndrome and amelogenesis imperfecta with taurodontism are genetically distinct conditions." Clin Genet **56**(1): 35-40.

Pulkkinen, L., A. M. Christiano, T. Airenne, H. Haakana, K. Tryggvason and J. Uitto (1994a). "Mutations in the gamma 2 chain gene (LAMC2) of kalinin/laminin 5 in the junctional forms of epidermolysis bullosa." Nat Genet **6**(3): 293-297.

Pulkkinen, L., A. M. Christiano, D. Gerecke, D. W. Wagman, R. E. Burgeson, M. R. Pittelkow and J. Uitto (1994b). "A homozygous nonsense mutation in the beta 3 chain gene of laminin 5 (LAMB3) in Herlitz junctional epidermolysis bullosa." Genomics **24**(2): 357-360.

Purdue, P. E., J. W. Zhang, M. Skoneczny and P. B. Lazarow (1997). "Rhizomelic chondrodysplasia punctata is caused by deficiency of human PEX7, a homologue of the yeast PTS2 receptor." Nat Genet **15**(4): 381-384.

Raas-Rothschild, A., R. J. Wanders, P. A. Mooijer, J. Gootjes, H. R. Waterham, A. Gutman, Y. Suzuki, N. Shimozawa, N. Kondo, G. Eshel, M. Espeel, F. Roels and S. H. Korman (2002). "A PEX6-defective peroxisomal biogenesis disorder with severe phenotype in an infant, versus mild phenotype resembling Usher syndrome in the affected parents." Am J Hum Genet **70**(4): 1062-1068.

Rajpar, M. H., K. Harley, C. Laing, R. M. Davies and M. J. Dixon (2001). "Mutation of the gene encoding the enamel-specific protein, enamelin, causes autosomal-dominant amelogenesis imperfecta." Hum Mol Genet **10**(16): 1673-1677.

Rao, S. and C. J. Witkop, Jr. (1971). "Inherited defects in tooth structure." Birth Defects Orig Artic Ser **7**(7): 153-184.

Ratbi, I., K. D. Falkenberg, M. Sommen, N. Al-Sheqaih, S. Guaoua, G. Vandeweyer, J. E. Urquhart, K. E. Chandler, S. G. Williams, N. A. Roberts, M. El Alloussi, G. C. Black, S. Ferdinandusse, H. Ramdi, A. Heimler, A. Fryer, S. A. Lynch, N. Cooper, K. R. Ong, C. E. L. Smith, C. F. Inglehearn, A. J. Mighell, C. Elcock, J. A. Poulter, M. Tischkowitz, S. J. Davies, A. Sefiani, A. A. Mironov, W.

G. Newman, H. R. Waterham and G. Van Camp (2015). "Heimler Syndrome Is Caused by Hypomorphic Mutations in the Peroxisome-Biogenesis Genes PEX1 and PEX6." Am J Hum Genet **97**(4): 535-545.

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, J. Zhang, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer and M. E. Hurles (2006). "Global variation in copy number in the human genome." Nature **444**(7118): 444-454.

Redza-Dutordoir, M. and D. A. Averill-Bates (2016). "Activation of apoptosis signalling pathways by reactive oxygen species." Biochim Biophys Acta **1863**(12): 2977-2992.

Reiners, J., E. van Wijk, T. Marker, U. Zimmermann, K. Jurgens, H. te Brinke, N. Overlack, R. Roepman, M. Knipper, H. Kremer and U. Wolfrum (2005). "Scaffold protein harmonin (USH1C) provides molecular links between Usher syndrome type 1 and type 2." Hum Mol Genet **14**(24): 3933-3943.

Reith, E. J. (1967). "The early stage of amelogenesis as observed in molar teeth of young rats." J Ultrastruct Res **17**(5): 503-526.

Reith, E. J. (1970). "The stages of amelogenesis as observed in molar teeth of young rats." J Ultrastruct Res **30**(1): 111-151.

Reuber, B. E., E. Germain-Lee, C. S. Collins, J. C. Morrell, R. Ameritunga, H. W. Moser, D. Valle and S. J. Gould (1997). "Mutations in PEX1 are the most common cause of peroxisome biogenesis disorders." Nat Genet **17**(4): 445-448.

Richards, S., N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding and H. L. Rehm (2015). "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." Genet Med **17**(5): 405-424.

Risnes, S. (1998). "Growth tracks in dental enamel." J Hum Evol **35**(4-5): 331-350.

Ritchie, G. R., I. Dunham, E. Zeggini and P. Flicek (2014). "Functional annotation of noncoding sequence variants." Nat Methods **11**(3): 294-296.

Ritchie, H. H., J. E. Berry, M. J. Somerman, C. T. Hanks, A. L. Bronckers, D. Hotton, P. Papagerakis, A. Berdal and W. T. Butler (1997). "Dentin sialoprotein (DSP) transcripts: developmentally-sustained expression in odontoblasts and transient expression in pre-ameloblasts." Eur J Oral Sci **105**(5 Pt 1): 405-413.

Rivolta, C., E. A. Sweklo, E. L. Berson and T. P. Dryja (2000). "Missense mutation in the USH2A gene: association with recessive retinitis pigmentosa without hearing loss." Am J Hum Genet **66**(6): 1975-1978.

Robinson, C., H. D. Briggs, P. J. Atkinson and J. A. Weatherell (1979). "Matrix and mineral changes in developing enamel." J Dent Res **58**(Spec Issue B): 871-882.

Robinson, C., S. Connell, J. Kirkham, S. J. Brookes, R. C. Shore and A. M. Smith (2004). "The effect of fluoride on the developing tooth." Caries Res **38**(3): 268-276.

Robinson, C., J. Kirkham, S. J. Brookes, W. A. Bonass and R. C. Shore (1995). "The chemistry of enamel development." Int J Dev Biol **39**(1): 145-152.

Robinson, C., R. C. Shore, S. J. Brookes, S. Strafford, S. R. Wood and J. Kirkham (2000). "The chemistry of enamel caries." Crit Rev Oral Biol Med **11**(4): 481-495.

Robinson, C., J. A. Weatherell and A. S. Hallsworth (1971). "Variation in composition of dental enamel within thin ground tooth sections." Caries Res **5**(1): 44-57.

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

Rong, W., X. Chen, K. Zhao, Y. Liu, X. Liu, S. Ha, W. Liu, X. Kang, X. Sheng and C. Zhao (2014). "Novel and recurrent MYO7A mutations in Usher syndrome type 1 and type 2." PLoS One **9**(5): e97808.

Roux, A. F., V. Faugere, C. Vache, D. Baux, T. Besnard, S. Leonard, C. Blanchet, C. Hamel, M. Mondain, B. Gilbert-Dussardier, P. Edery, D. Lacombe,

D. Bonneau, M. Holder-Espinasse, U. Ambrosetti, H. Journel, A. David, G. Lina-Granade, S. Malcolm and M. Claustres (2011). "Four-year follow-up of diagnostic service in USH1 patients." Invest Ophthalmol Vis Sci **52**(7): 4063-4071.

Rowley, R., F. J. Hill and G. B. Winter (1982). "An investigation of the association between anterior open-bite and amelogenesis imperfecta." Am J Orthod **81**(3): 229-235.

Ruch, J. V., H. Lesot and C. Begue-Kirn (1995). "Odontoblast differentiation." Int J Dev Biol **39**(1): 51-68.

Ruiz-Martinez, M. C., J. Berka, A. Belenkii, F. Foret, A. W. Miller and B. L. Karger (1993). "DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection." Anal Chem **65**(20): 2851-2858.

Ryu, O. H., A. G. Fincham, C. C. Hu, C. Zhang, Q. Qian, J. D. Bartlett and J. P. Simmer (1999). "Characterization of recombinant pig enamelysin activity and cleavage of recombinant pig and mouse amelogenins." J Dent Res **78**(3): 743-750.

Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander and D. Altshuler (2001). "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." Nature **409**(6822): 928-933.

Saffian, D., I. Grimm, W. Girzalsky and R. Erdmann (2012). "ATP-dependent assembly of the heteromeric Pex1p-Pex6p-complex of the peroxisomal matrix protein import machinery." J Struct Biol **179**(2): 126-132.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich (1988). "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase." Science **239**(4839): 487-491.

Salido, E. C., P. H. Yen, K. Koprivnikar, L. C. Yu and L. J. Shapiro (1992). "The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes." Am J Hum Genet **50**(2): 303-316.

Samarakoon, P. S., H. S. Sorte, B. E. Kristiansen, T. Skodje, Y. Sheng, G. E. Tjonnfjord, B. Stadheim, A. Stray-Pedersen, O. K. Rodningen and R. Lyle (2014). "Identification of copy number variants from exome sequence data." BMC Genomics **15**: 661.

Sambrook, J. and D. W. Russell (2000). Molecular Cloning: A Laboratory Manual. New York, USA, Cold Spring Harbour Laboratory Press.

Samiy, N. (2014). "Gene therapy for retinal diseases." J Ophthalmic Vis Res **9**(4): 506-509.

Sanchez, A. R., R. S. Rogers, 3rd and P. J. Sheridan (2004). "Tetracycline and other tetracycline-derivative staining of the teeth and oral cavity." Int J Dermatol **43**(10): 709-715.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.

Sanghavi, D. (2006). "Wanting babies like themselves, some parents choose genetic defects." N Y Times Web: F5, F8.

Santos, M. C., P. S. Hart, M. Ramaswami, C. M. Kanno, T. C. Hart and S. R. Line (2007). "Exclusion of known gene for enamel development in two Brazilian families with amelogenesis imperfecta." Head Face Med **3**: 8.

Santos, M. J., T. Imanaka, H. Shio, G. M. Small and P. B. Lazarow (1988). "Peroxisomal membrane ghosts in Zellweger syndrome--aberrant organelle assembly." Science **239**(4847): 1536-1538.

Sarkar, J., E. J. Simanian, S. Y. Tuggy, J. D. Bartlett, M. L. Snead, T. Sugiyama and M. L. Paine (2014). "Comparison of two mouse ameloblast-like cell lines for enamel-specific gene expression." Front Physiol **5**: 277.

Sasaki, S., T. Takagi and M. Suzuki (1991). "Cyclical changes in pH in bovine developing enamel as sequential bands." Arch Oral Biol **36**(3): 227-231.

Sasaki, T. (1984). "Synthesis and secretion of the enamel matrix precursor by the kitten secretory ameloblast." Acta Anat (Basel) **120**(3): 98-102.

Sasano, Y. (1986). "Dynamic behavior of ciliated centrioles in rat incisor ameloblasts during cell differentiation." Arch Histol Jpn **49**(4): 437-448.

Schell-Steven, A., K. Stein, M. Amoros, C. Landgraf, R. Volkmer-Engert, H. Rottensteiner and R. Erdmann (2005). "Identification of a novel, intraperoxisomal pex14-binding site in pex13: association of pex13 with the docking complex is essential for peroxisomal matrix protein import." Mol Cell Biol **25**(8): 3007-3018.

Schultz, J., F. Milpetz, P. Bork and C. P. Ponting (1998). "SMART, a simple modular architecture research tool: identification of signaling domains." Proc Natl Acad Sci U S A **95**(11): 5857-5864.

Schwartz, R. A., G. Fernandez, K. Kotulska and S. Jozwiak (2007). "Tuberous sclerosis complex: advances in diagnosis, genetics, and management." J Am Acad Dermatol **57**(2): 189-202.

Schwartz, S., E. Hall and G. Ast (2009). "SROOGLE: webserver for integrative, user-friendly visualization of splicing signals." Nucleic Acids Res **37**(Web Server issue): W189-192.

Schwarz, J. M., C. Rodelsperger, M. Schuelke and D. Seelow (2010). "MutationTaster evaluates disease-causing potential of sequence alterations." Nat Methods **7**(8): 575-576.

Sculean, A., F. Schwarz, J. Becker and M. Brecx (2007). "The application of an enamel matrix protein derivative (Emdogain) in regenerative periodontal therapy: a review." Med Princ Pract **16**(3): 167-180.

Seymen, F., K. E. Lee, C. G. Tran Le, M. Yildirim, K. Gencay, Z. H. Lee and J. W. Kim (2014). "Exonal deletion of SLC24A4 causes hypomaturation amelogenesis imperfecta." J Dent Res **93**(4): 366-370.

Seymen, F., J. C. Park, K. E. Lee, H. K. Lee, D. S. Lee, M. Koruyucu, K. Gencay, M. Bayram, E. B. Tuna, Z. H. Lee, Y. J. Kim and J. W. Kim (2015). "Novel MMP20 and KLK4 Mutations in Amelogenesis Imperfecta." J Dent Res **94**(8): 1063-1069.

Shapiro, M. B. and P. Senapathy (1987). "RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression." Nucleic Acids Res **15**(17): 7155-7174.

Sharp, P. A. (1981). "Speculations on RNA splicing." Cell **23**(3): 643-646.

Sharpe, P. T. (1995). "Homeobox genes and orofacial development." Connect Tissue Res **32**(1-4): 17-25.

Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski and K. Sirotkin (2001). "dbSNP: the NCBI database of genetic variation." Nucleic Acids Res **29**(1): 308-311.

Shi, Y. and J. Majewski (2013). "FishingCNV: a graphical software package for detecting rare copy number variations in exome-sequencing data." Bioinformatics **29**(11): 1461-1462.

Shields, E. D., D. Bixler and A. M. el-Kafrawy (1973). "A proposed classification for heritable human dentine defects with a description of a new entity." Arch Oral Biol **18**(4): 543-553.

Shimozawa, N., Y. Suzuki, Z. Zhang, A. Imamura, K. Ghaedi, Y. Fujiki and N. Kondo (2000). "Identification of PEX3 as the gene mutated in a Zellweger syndrome patient lacking peroxisomal remnant structures." Hum Mol Genet **9**(13): 1995-1999.

Shimozawa, N., Y. Suzuki, Z. Zhang, A. Imamura, R. Toyama, S. Mukai, Y. Fujiki, T. Tsukamoto, T. Osumi, T. Orii, R. J. Wanders and N. Kondo (1999). "Nonsense and temperature-sensitive mutations in PEX13 are the cause of complementation group H of peroxisome biogenesis disorders." Hum Mol Genet **8**(6): 1077-1083.

Shimozawa, N., T. Tsukamoto, T. Nagase, Y. Takemoto, N. Koyama, Y. Suzuki, M. Komori, T. Osumi, G. Jeannette, R. J. Wanders and N. Kondo (2004). "Identification of a new complementation group of the peroxisome biogenesis disorders and PEX14 as the mutated gene." Hum Mutat **23**(6): 552-558.

Shimozawa, N., T. Tsukamoto, Y. Suzuki, T. Orii, Y. Shirayoshi, T. Mori and Y. Fujiki (1992). "A human gene responsible for Zellweger syndrome that affects peroxisome assembly." Science **255**(5048): 1132-1134.

Shin, J., G. L. Ming and H. Song (2014). "Decoding neural transcriptomes and epigenomes via high-throughput sequencing." Nat Neurosci **17**(11): 1463-1475.

Shiozawa, K., N. Goda, T. Shimizu, K. Mizuguchi, N. Kondo, N. Shimozawa, M. Shirakawa and H. Hiroaki (2006). "The common phospholipid-binding activity of the N-terminal domains of PEX1 and VCP/p97." FEBS J **273**(21): 4959-4971.

Shiozawa, K., N. Maita, K. Tomii, A. Seto, N. Goda, Y. Akiyama, T. Shimizu, M. Shirakawa and H. Hiroaki (2004). "Structure of the N-terminal domain of PEX1 AAA-ATPase. Characterization of a putative adaptor-binding domain." J Biol Chem **279**(48): 50060-50068.

Shore, R. C., J. Kirkham, S. J. Brookes, S. R. Wood and C. Robinson (2000). "Distribution of exogenous proteins in caries lesions in relation to the pattern of demineralisation." Caries Res **34**(2): 188-193.

Simmer, J. P. and A. G. Fincham (1995). "Molecular mechanisms of dental enamel formation." Crit Rev Oral Biol Med **6**(2): 84-108.

Simmer, J. P. and J. C. Hu (2002). "Expression, structure, and function of enamel proteinases." Connect Tissue Res **43**(2-3): 441-449.

Simmer, J. P., Y. Hu, R. Lertlam, Y. Yamakoshi and J. C. Hu (2009). "Hypomaturation enamel defects in Klk4 knockout/LacZ knockin mice." J Biol Chem **284**(28): 19110-19121.

Simmer, J. P., A. S. Richardson, S. K. Wang, B. M. Reid, Y. Bai, Y. Hu and J. C. Hu (2014). "Ameloblast transcriptome changes from secretory to maturation stages." Connect Tissue Res **55 Suppl 1**: 29-32.

Sire, J. Y., T. Davit-Beal, S. Delgado and X. Gu (2007). "The origin and evolution of enamel mineralization genes." Cells Tissues Organs **186**(1): 25-48.

Skobe, Z. (1976). "The secretory stage of amelogenesis in rat mandibular incisor teeth observed by scanning electron microscopy." Calcif Tissue Res **21**(2): 83-103.

Small, J., K. Heinrich, C. Fiori, R. Myklebust, D. Newbury and M. Dilmore (1978). The production and characterization of glass fibers and spheres for microanalysis. Scanning Electron Microscopy. O. Johari. AMF O'Hare, IL, USA, SEM Inc. **1:** 445-454.

Small, K. W., A. P. DeLuca, S. S. Whitmore, T. Rosenberg, R. Silva-Garcia, N. Udar, B. Puech, C. A. Garcia, T. A. Rice, G. A. Fishman, E. Heon, J. C. Folk, L. M. Streb, C. M. Haas, L. A. Wiley, T. E. Scheetz, J. H. Fingert, R. F. Mullins, B. A. Tucker and E. M. Stone (2016). "North Carolina Macular Dystrophy Is Caused by Dysregulation of the Retinal Transcription Factor PRDM13." Ophthalmology **123**(1): 9-18.

Smith, A. J., N. Cassidy, H. Perry, C. Begue-Kirn, J. V. Ruch and H. Lesot (1995). "Reactionary dentinogenesis." Int J Dev Biol **39**(1): 273-280.

Smith, C. E. (1998). "Cellular and chemical events during enamel maturation." Crit Rev Oral Biol Med **9**(2): 128-161.

Smith, C. E., M. Issid, H. C. Margolis and E. C. Moreno (1996). "Developmental changes in the pH of enamel fluid and its effects on matrix-resident proteinases." Adv Dent Res **10**(2): 159-169.

Smith, C. E. and A. Nanci (1989). "A method for sampling the stages of amelogenesis on mandibular rat incisors using the molars as a reference for dissection." Anat Rec **225**(3): 257-266.

Smith, C. E. and A. Nanci (1995). "Overview of morphological changes in enamel organ cells associated with major events in amelogenesis." Int J Dev Biol **39**(1): 153-161.

Smith, C. E. and H. Warshawsky (1977). "Quantitative analysis of cell turnover in the enamel organ of the rat incisor. Evidence for ameloblast death immediately after enamel matrix secretion." Anat Rec **187**(1): 63-98.

Smith, C. E. L., G. Murillo, S. J. Brookes, J. A. Poulter, S. Silva, J. Kirkham, C. F. Inglehearn and A. J. Mighell (2016a). "Deletion of amelotin exons 3-6 is associated with amelogenesis imperfecta." Hum Mol Genet.

Smith, C. E. L., J. A. Poulter, A. V. Levin, J. E. Capasso, S. Price, T. Ben-Yosef, R. Sharony, W. G. Newman, R. C. Shore, S. J. Brookes, A. J. Mighell and C. F. Inglehearn (2016b). "Spectrum of PEX1 and PEX6 variants in Heimler syndrome." Eur J Hum Genet **24**(11): 1565-1571.

Smith, J. J. and J. D. Aitchison (2013). "Peroxisomes take shape." Nat Rev Mol Cell Biol **14**(12): 803-817.

Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. Kent and L. E. Hood (1986). "Fluorescence detection in automated DNA sequence analysis." Nature **321**(6071): 674-679.

Somogyi-Ganss, E., Y. Nakayama, K. Iwasaki, Y. Nakano, D. Stolf, M. D. McKee and B. Ganss (2012). "Comparative temporospatial expression profiling of murine amelotin protein during amelogenesis." Cells Tissues Organs **195**(6): 535-549.

Sorensen, E. S., P. Hojrup and T. E. Petersen (1995). "Posttranslational modifications of bovine osteopontin: identification of twenty-eight phosphorylation and three O-glycosylation sites." Protein Sci **4**(10): 2040-2049.

Soviero, V., D. Haubek, C. Trindade, T. Da Matta and S. Poulsen (2009). "Prevalence and distribution of demarcated opacities and their sequelae in permanent 1st molars and incisors in 7 to 13-year-old Brazilian children." Acta Odontol Scand **67**(3): 170-175.

Spahr, A., S. P. Lyngstadaas, I. Slaby and G. Pezeshki (2006). "Ameloblastin expression during craniofacial bone formation in rats." Eur J Oral Sci **114**(6): 504-511.

Statham, P. and J. Pawley (1978). A new method for particle x-ray micro-analysis based on peak-to-background measurement. Scanning Electron Microscopy. O. Johari. AMF O'Hare, IL, USA, SEM Inc. **1:** 469-478.

Steinberg, S., L. Chen, L. Wei, A. Moser, H. Moser, G. Cutting and N. Braverman (2004). "The PEX Gene Screen: molecular diagnosis of peroxisome biogenesis disorders in the Zellweger syndrome spectrum." Mol Genet Metab **83**(3): 252-263.

Steinberg, S. J., G. V. Raymond, N. E. Braverman and A. B. Moser (2012). Peroxisome Biogenesis Disorders, Zellweger Syndrome Spectrum. Seattle, WA, USA, University of Washington, Seattle.

Stelzig, I., S. Karnati, K. P. Valerius and E. Baumgart-Vogt (2013). "Peroxisomes in dental tissues of the mouse." Histochem Cell Biol **140**(4): 443-462.

Stenson, P. D., M. Mort, E. V. Ball, K. Shaw, A. Phillips and D. N. Cooper (2014). "The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine." Hum Genet **133**(1): 1-9.

Stranger, B. E., M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazley, N. Thorne, R. Redon, C. P. Bird, A. de Grassi, C. Lee, C. Tyler-Smith, N. Carter, S. W. Scherer, S. Tavare, P. Deloukas, M. E. Hurles and E. T. Dermitzakis (2007). "Relative impact of nucleotide and copy number variation on gene expression phenotypes." Science **315**(5813): 848-853.

Sundell, S. and G. Koch (1985). "Hereditary amelogenesis imperfecta. I. Epidemiology and clinical classification in a Swedish child population." Swed Dent J **9**(4): 157-169.

Sundell, S. and J. Valentin (1986). "Hereditary aspects and classification of hereditary amelogenesis imperfecta." Community Dent Oral Epidemiol **14**(4): 211-216.

Tagliabracci, V. S., J. L. Engel, J. Wen, S. E. Wiley, C. A. Worby, L. N. Kinch, J. Xiao, N. V. Grishin and J. E. Dixon (2012). "Secreted kinase phosphorylates extracellular proteins that regulate biomineralization." Science **336**(6085): 1150-1153.

Takano, Y., M. A. Crenshaw, J. W. Bawden, L. Hammarstrom and S. Lindskog (1982). "The visualization of the patterns of ameloblast modulation by the glyoxal bis(2-hydroxyanil) staining method." J Dent Res **Spec No**: 1580-1587.

Tamura, S., S. Yasutake, N. Matsumoto and Y. Fujiki (2006). "Dynamic and functional assembly of the AAA peroxins, Pex1p and Pex6p, and their membrane receptor Pex26p." J Biol Chem **281**(38): 27693-27704.

Tan, R., Y. Wang, S. E. Kleinstein, Y. Liu, X. Zhu, H. Guo, Q. Jiang, A. S. Allen and M. Zhu (2014). "An evaluation of copy number variation detection tools from whole-exome sequencing data." Hum Mutat **35**(7): 899-907.

Tekin, M., B. O. Hismi, S. Fitoz, H. Ozdag, F. B. Cengiz, A. Sirmaci, I. Aslan, B. Inceoglu, E. B. Yuksel-Konuk, S. T. Yilmaz, O. Yasun and N. Akar (2007). "Homozygous mutations in fibroblast growth factor 3 are associated with a new form of syndromic deafness characterized by inner ear agenesis, microtia, and microdontia." Am J Hum Genet **80**(2): 338-344.

Thesleff, I. (2003). "Epithelial-mesenchymal signalling regulating tooth morphogenesis." J Cell Sci **116**(Pt 9): 1647-1648.

Thesleff, I. and M. Mikkola (2002). "The role of growth factors in tooth development." Int Rev Cytol **217**: 93-135.

Thompson, J. D., D. G. Higgins and T. J. Gibson (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-4680.

Tischkowitz, M., C. Clenaghan, S. Davies, L. Hunter, J. Potts and S. Verhoef (1999). "Amelogenesis imperfecta, sensorineural hearing loss, and Beau's lines, a second case report of Heimler's syndrome." J Med Genet **36**(12): 941-943.

Tompkins, K. (2006). "Molecular mechanisms of cytodifferentiation in mammalian tooth development." Connect Tissue Res **47**(3): 111-118.

Townsend, A., S. Adam, P. H. Birch and J. M. Friedman (2013). "Paternalism and the ACMG recommendations on genomic incidental findings: patients seen but not heard." Genet Med **15**(9): 751-752.

Tran, D., W. Greenhill and S. Wilson (2011). "Infantile refsum disease with enamel defects: a case report." Pediatr Dent **33**(3): 266-270.

Trincavelli, J. and R. Van Grieken (1994). "Peak-to-Background method for standardless electron microprobe analysis of particles." X-Ray Spectrometry **23**: 254-260.

Trumpp, A., M. J. Depew, J. L. Rubenstein, J. M. Bishop and G. R. Martin (1999). "Cre-mediated gene inactivation demonstrates that FGF8 is required for cell survival and patterning of the first branchial arch." Genes Dev **13**(23): 3136-3148.

Tucker, A. S., K. L. Matthews and P. T. Sharpe (1998). "Transformation of tooth type induced by inhibition of BMP signaling." Science **282**(5391): 1136-1138.

Tucker, A. S., G. Yamada, M. Grigoriou, V. Pachnis and P. T. Sharpe (1999). "Fgf-8 determines rostral-caudal polarity in the first branchial arch." Development **126**(1): 51-61.

Turk, B. E., D. H. Lee, Y. Yamakoshi, A. Klingenhoff, E. Reichenberger, J. T. Wright, J. P. Simmer, J. A. Komisarof, L. C. Cantley and J. D. Bartlett (2006). "MMP-20 is predominately a tooth-specific enzyme with a deep catalytic pocket that hydrolyzes type V collagen." Biochemistry **45**(12): 3863-3874.

Uchida, T., T. Tanabe, M. Fukae, M. Shimizu, M. Yamada, K. Miake and S. Kobayashi (1991). "Immunochemical and immunohistochemical studies, using antisera against porcine 25 kDa amelogenin, 89 kDa enamelin and the 13-17 kDa nonamelogenins, on immature enamel of the pig and rat." Histochemistry **96**(2): 129-138.

Vache, C., T. Besnard, C. Blanchet, D. Baux, L. Larrieu, V. Faugere, M. Mondain, C. Hamel, S. Malcolm, M. Claustres and A. F. Roux (2010). "Nasal epithelial cells are a reliable source to study splicing variants in Usher syndrome." Hum Mutat **31**(6): 734-741.

van Binsbergen, E. (2011). "Origins and breakpoint analyses of copy number variations: up close and personal." Cytogenet Genome Res **135**(3-4): 271-276.

van Wijk, E., R. J. Pennings, H. te Brinke, A. Claassen, H. G. Yntema, L. H. Hoefsloot, F. P. Cremers, C. W. Cremers and H. Kremer (2004). "Identification of 51 novel exons of the Usher syndrome type 2A (USH2A) gene that encode multiple conserved functional domains and that are mutated in patients with Usher syndrome type II." Am J Hum Genet **74**(4): 738-744.

van Wijk, E., B. van der Zwaag, T. Peters, U. Zimmermann, H. Te Brinke, F. F. Kersten, T. Marker, E. Aller, L. H. Hoefsloot, C. W. Cremers, F. P. Cremers, U. Wolfrum, M. Knipper, R. Roepman and H. Kremer (2006). "The DFNB31 gene product whirlin connects to the Usher protein network in the cochlea and retina by direct association with USH2A and VLGR1." Hum Mol Genet **15**(5): 751-765.

VanSaun, M. N. and L. M. Matrisian (2006). "Matrix metalloproteinases and cellular motility in development and disease." Birth Defects Res C Embryo Today **78**(1): 69-79.

Veis, A. (2003). "Amelogenin gene splice products: potential signaling molecules." Cell Mol Life Sci **60**(1): 38-55.

Veis, A., K. Tompkins, K. Alvares, K. Wei, L. Wang, X. S. Wang, A. G. Brownell, S. M. Jengh and K. E. Healy (2000). "Specific amelogenin gene splice products have signaling effects on cells in culture and in implants in vivo." J Biol Chem **275**(52): 41263-41272.

Venkatraman, E. S. and A. B. Olshen (2007). "A faster circular binary segmentation algorithm for the analysis of array CGH data." Bioinformatics **23**(6): 657-663.

von Marschall, Z., S. Mok, M. D. Phillips, D. A. McKnight and L. W. Fisher (2012). "Rough endoplasmic reticulum trafficking errors by different classes of mutant dentin sialophosphoprotein (DSPP) cause dominant negative effects in both dentinogenesis imperfecta and dentin dysplasia by entrapping normal DSPP." J Bone Miner Res **27**(6): 1309-1321.

Wakeling, E. L., S. N. Abu-Amero, P. Stanier, M. A. Preece and G. E. Moore (1998). "Human EGFR, a candidate gene for the Silver-Russell

syndrome, is biallelically expressed in a wide range of fetal tissues." Eur J Hum Genet **6**(2): 158-164.

Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." EMBO J **1**(8): 945-951.

Walter, C., J. Gootjes, P. A. Mooijer, H. Portsteffen, C. Klein, H. R. Waterham, P. G. Barth, J. T. Epplen, W. H. Kunau, R. J. Wanders and G. Dodt (2001). "Disorders of peroxisome biogenesis due to mutations in PEX1: phenotypes and PEX1 protein levels." Am J Hum Genet **69**(1): 35-48.

Wanders, R. J. and H. R. Waterham (2006). "Biochemistry of mammalian peroxisomes revisited." Annu Rev Biochem **75**: 295-332.

Wang, M. and R. J. Kaufman (2016). "Protein misfolding in the endoplasmic reticulum as a conduit to human disease." Nature **529**(7586): 326-335.

Wang, S., M. Choi, A. S. Richardson, B. M. Reid, F. Seymen, M. Yildirim, E. Tuna, K. Gencay, J. P. Simmer and J. C. Hu (2014a). "STIM1 and SLC24A4 Are Critical for Enamel Maturation." J Dent Res.

Wang, S. K., P. Aref, Y. Hu, R. N. Milkovich, J. P. Simmer, M. El-Khateeb, H. Daggag, Z. H. Baqain and J. C. Hu (2013a). "FAM20A mutations can cause enamel-renal syndrome (ERS)." PLoS Genet **9**(2): e1003302.

Wang, S. K., M. Choi, A. S. Richardson, B. M. Reid, B. P. Lin, S. J. Wang, J. W. Kim, J. P. Simmer and J. C. Hu (2014b). "ITGB6 loss-of-function mutations cause autosomal recessive amelogenesis imperfecta." Hum Mol Genet **23**(8): 2157-2163.

Wang, S. K., Y. Hu, J. P. Simmer, F. Seymen, N. M. Estrella, S. Pal, B. M. Reid, M. Yildirim, M. Bayram, J. D. Bartlett and J. C. Hu (2013b). "Novel KLK4 and MMP20 mutations discovered by whole-exome sequencing." J Dent Res **92**(3): 266-271.

Wang, S. K., Y. Hu, J. Yang, C. E. Smith, S. M. Nunez, A. S. Richardson, S. Pal, A. C. Samann, J. C. Hu and J. P. Simmer (2015). "Critical roles for WDR72 in calcium transport and matrix protein removal during enamel maturation." Mol Genet Genomic Med **3**(4): 302-319.

Wang, S. K., Y. Hu, J. Yang, C. E. Smith, A. S. Richardson, Y. Yamakoshi, Y. L. Lee, F. Seymen, M. Koruyucu, K. Gencay, M. Lee, M. Choi, J. W. Kim, J. C. Hu and J. P. Simmer (2016). "Fam83h null mice support a neomorphic mechanism for human ADHCAI." Mol Genet Genomic Med **4**(1): 46-67.

Warren, D. S., J. C. Morrell, H. W. Moser, D. Valle and S. J. Gould (1998). "Identification of PEX10, the gene defective in complementation group 7 of the peroxisome-biogenesis disorders." Am J Hum Genet **63**(2): 347-359.

Warshawsky, H. and C. E. Smith (1974). "Morphological classification of rat incisor ameloblasts." Anat Rec **179**(4): 423-446.

Waterham, H. R. and M. S. Ebberink (2012). "Genetics and molecular basis of human peroxisome biogenesis disorders." Biochim Biophys Acta **1822**(9): 1430-1441.

Waterham, H. R., J. Koster, C. W. van Roermund, P. A. Mooyer, R. J. Wanders and J. V. Leonard (2007). "A lethal defect of mitochondrial and peroxisomal fission." N Engl J Med **356**(17): 1736-1741.

Waters, P. J. (2001). "Degradation of mutant proteins, underlying "loss of function" phenotypes, plays a major role in genetic disease." Curr Issues Mol Biol **3**(3): 57-65.

Watson, C. M., L. A. Crinnion, A. Tzika, A. Mills, A. Coates, M. Pendlebury, S. Hewitt, S. M. Harrison, C. Daly, P. Roberts, I. M. Carr, E. G. Sheridan and D. T. Bonthron (2014). "Diagnostic whole genome sequencing and split-read mapping for nucleotide resolution breakpoint identification in CNTNAP2 deficiency syndrome." Am J Med Genet A **164A**(10): 2649-2655.

Weerheijm, K. L., B. Jalevik and S. Alaluusua (2001). "Molar-incisor hypomineralisation." Caries Res **35**(5): 390-391.

Wei, H., S. Kemp, M. C. McGuinness, A. B. Moser and K. D. Smith (2000). "Pharmacological induction of peroxisomes in peroxisome biogenesis disorders." Ann Neurol **47**(3): 286-296.

Weil, D., S. Blanchard, J. Kaplan, P. Guilford, F. Gibson, J. Walsh, P. Mburu, A. Varela, J. Levilliers, M. D. Weston and et al. (1995). "Defective myosin VIIA gene responsible for Usher syndrome type 1B." Nature **374**(6517): 60-61.

Weinmann, J. P., J. F. Svobada and R. W. Woods (1945). "Hereditary disturbances of enamel formation and calcification." J Am Dent Assoc **32**(7): 397-418.

Welch, E. M., E. R. Barton, J. Zhuo, Y. Tomizawa, W. J. Friesen, P. Trifillis, S. Paushkin, M. Patel, C. R. Trotta, S. Hwang, R. G. Wilde, G. Karp, J. Takasugi, G. Chen, S. Jones, H. Ren, Y. C. Moon, D. Corson, A. A. Turpoff, J. A. Campbell, M. M. Conn, A. Khan, N. G. Almstead, J. Hedrick, A. Mollin, N. Risher, M. Weetall, S. Yeh, A. A. Branstrom, J. M. Colacino, J. Babiak, W. D. Ju, S. Hirawat, V. J. Northcutt, L. L. Miller, P. Spatrick, F. He, M. Kawana, H. Feng, A. Jacobson, S. W. Peltz and H. L. Sweeney (2007). "PTC124 targets genetic disorders caused by nonsense mutations." Nature **447**(7140): 87-91.

Weleber, R. G., M. E. Pennesi, D. J. Wilson, S. Kaushal, L. R. Erker, L. Jensen, M. T. McBride, T. R. Flotte, M. Humphries, R. Calcedo, W. W. Hauswirth, J. D. Chulay and J. T. Stout (2016). "Results at 2 Years after Gene Therapy for RPE65-Deficient Leber Congenital Amaurosis and Severe Early-Childhood-Onset Retinal Dystrophy." Ophthalmology **123**(7): 1606-1620.

Wendland, M. and S. Subramani (1993). "Presence of cytoplasmic factors functional in peroxisomal protein import implicates organelle-associated defects in several human peroxisomal disorders." J Clin Invest **92**(5): 2462-2468.

Weston, M. D., J. D. Eudy, S. Fujita, S. Yao, S. Usami, C. Cremers, J. Greenberg, R. Ramesar, A. Martini, C. Moller, R. J. Smith, J. Sumegi and W. J. Kimberling (2000). "Genomic structure and identification of novel mutations in usherin, the gene responsible for Usher syndrome type IIa." Am J Hum Genet **66**(4): 1199-1210.

Whatling, R. and J. M. Fearne (2008). "Molar incisor hypomineralization: a study of aetiological factors in a group of UK children." Int J Paediatr Dent **18**(3): 155-162.

Wheway, G., D. A. Parry and C. A. Johnson (2014). "The role of primary cilia in the development and disease of the retina." Organogenesis **10**(1): 69-85.

Wiese, S., T. Gronemeyer, R. Ofman, M. Kunze, C. P. Grou, J. A. Almeida, M. Eisenacher, C. Stephan, H. Hayen, L. Schollenberger, T. Korosec, H. R. Waterham, W. Schliebs, R. Erdmann, J. Berger, H. E. Meyer, W. Just, J. E. Azevedo, R. J. Wanders and B. Warscheid (2007). "Proteomics

characterization of mouse kidney peroxisomes by tandem mass spectrometry and protein correlation profiling." Mol Cell Proteomics **6**(12): 2045-2057.

Winter, G. B. and A. H. Brook (1975). "Enamel hypoplasia and anomalies of the enamel." Dent Clin North Am **19**(1): 3-24.

Wise, G. E., F. Lin and W. Fan (1992). "Localization of epidermal growth factor and its receptor in mandibular molars of the rat prior to and during prefunctional tooth eruption." Dev Dyn **195**(2): 121-126.

Witkop, C. J., Jr. (1988). "Amelogenesis imperfecta, dentinogenesis imperfecta and dentin dysplasia revisited: problems in classification." J Oral Pathol **17**(9-10): 547-553.

Witkop, C. J. and J. J. Sauk (1976). Heritable defects of enamel. St. Louis, Missouri, USA, CV Mosby Company.

Wong, F. S., P. Anderson, H. Fan and G. R. Davis (2004). "X-ray microtomographic study of mineral concentration distribution in deciduous enamel." Arch Oral Biol **49**(11): 937-944.

Wright, J. T., K. I. Hall and M. Yamauche (1997). "The enamel proteins in human amelogenesis imperfecta." Arch Oral Biol **42**(2): 149-159.

Wright, J. T., M. Torain, K. Long, K. Seow, P. Crawford, M. J. Aldred, P. S. Hart and T. C. Hart (2011). "Amelogenesis imperfecta: genotype-phenotype studies in 71 families." Cells Tissues Organs **194**(2-4): 279-283.

Wu, Y., H. Zhou, X. Fan, Y. Zhang, M. Zhang, Y. Wang, Z. Xie, M. Bai, Q. Yin, D. Liang, W. Tang, J. Liao, C. Zhou, W. Liu, P. Zhu, H. Guo, H. Pan, C. Wu, H. Shi, L. Wu, F. Tang and J. Li (2015). "Correction of a genetic disease by CRISPR-Cas9-mediated gene editing in mouse spermatogonial stem cells." Cell Res **25**(1): 67-79.

Xiao, S., C. Yu, X. Chou, W. Yuan, Y. Wang, L. Bu, G. Fu, M. Qian, J. Yang, Y. Shi, L. Hu, B. Han, Z. Wang, W. Huang, J. Liu, Z. Chen, G. Zhao and X. Kong (2001). "Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in DSPP." Nat Genet **27**(2): 201-204.

Yahraus, T., N. Braverman, G. Dodt, J. E. Kalish, J. C. Morrell, H. W. Moser, D. Valle and S. J. Gould (1996). "The peroxisome biogenesis disorder group 4 gene, PXAAA1, encodes a cytoplasmic ATPase required for stability of the PTS1 receptor." EMBO J **15**(12): 2914-2923.

Yamakoshi, Y. (2011). "Porcine Amelogenin : Alternative Splicing, Proteolytic Processing, Protein - Protein Interactions, and Possible Functions." J Oral Biosci **53**(3): 275-283.

Yamakoshi, Y., J. C. Hu, M. Fukae, F. Yamakoshi and J. P. Simmer (2006). "How do enamelysin and kallikrein 4 process the 32-kDa enamelin?" Eur J Oral Sci **114 Suppl 1**: 45-51; discussion 93-45, 379-380.

Yamakoshi, Y., J. P. Simmer, J. D. Bartlett, T. Karakida and S. Oida (2013). "MMP20 and KLK4 activation and inactivation interactions in vitro." Arch Oral Biol **58**(11): 1569-1577.

Yamakoshi, Y., T. Tanabe, S. Oida, C. C. Hu, J. P. Simmer and M. Fukae (2001). "Calcium binding of enamel proteins and their derivatives with emphasis on the calcium-binding domain of porcine sheathlin." Arch Oral Biol **46**(11): 1005-1014.

Yamakoshi, Y., F. Yamakoshi, J. C. Hu and J. P. Simmer (2011). "Characterization of kallikrein-related peptidase 4 glycosylations." Eur J Oral Sci **119 Suppl 1**: 234-240.

Yang, X., W. Yan, Y. Tian, P. Ma, L. A. Opperman and X. Wang (2016). "Family with sequence similarity member 20C is the primary but not the only kinase for the small-integrin-binding ligand N-linked glycoproteins in bone." FASEB J **30**(1): 121-128.

Ye, K., M. H. Schulz, Q. Long, R. Apweiler and Z. Ning (2009). "Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads." Bioinformatics **25**(21): 2865-2871.

Yeo, G. and C. B. Burge (2004). "Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals." J Comput Biol **11**(2-3): 377-394.

Yik, W. Y., S. J. Steinberg, A. B. Moser, H. W. Moser and J. G. Hacia (2009). "Identification of novel mutations and sequence variation in the Zellweger syndrome spectrum of peroxisome biogenesis disorders." Hum Mutat **30**(3): E467-480.

Yin, H., W. Xue, S. Chen, R. L. Bogorad, E. Benedetti, M. Grompe, V. Koteliansky, P. A. Sharp, T. Jacks and D. G. Anderson (2014a). "Genome

editing with Cas9 in adult mice corrects a disease mutation and phenotype." Nat Biotechnol **32**(6): 551-553.

Yin, K., J. G. Hacia, Z. Zhong and M. L. Paine (2014b). "Genome-wide analysis of miRNA and mRNA transcriptomes during amelogenesis." BMC Genomics **15**: 998.

Yoshida, K., J. Sato, R. Takai, O. Uehara, Y. Kurashige, M. Nishimura, I. Chiba, M. Saitoh and Y. Abiko (2015). "Differentiation of mouse iPS cells into ameloblast-like cells in cultures using medium conditioned by epithelial cell rests of Malassez and gelatin-coated dishes." Med Mol Morphol **48**(3): 138-145.

Yoshizaki, K. and Y. Yamada (2013). "Gene evolution and functions of extracellular matrix proteins in teeth." Orthod Waves **72**(1): 1-10.

Yousef, G. M., C. V. Obiezu, L. Y. Luo, M. H. Black and E. P. Diamandis (1999). "Prostase/KLK-L1 is a new member of the human kallikrein gene family, is expressed in prostate and breast tissues, and is hormonally regulated." Cancer Res **59**(17): 4252-4256.

Yuen, W. Y., A. M. Pasmooij, C. Stellingsma and M. F. Jonkman (2012). "Enamel defects in carriers of a novel LAMA3 mutation underlying epidermolysis bullosa." Acta Derm Venereol **92**(6): 695-696.

Zaki, M. S., R. Heller, M. Thoenes, G. Nurnberg, G. Stern-Schneider, P. Nurnberg, S. Karnati, D. Swan, E. Fateen, K. Nagel-Wolfrum, M. I. Mostafa, H. Thiele, U. Wolfrum, E. Baumgart-Vogt and H. J. Bolz (2016). "PEX6 is Expressed in Photoreceptor Cilia and Mutated in Deafblindness with Enamel Dysplasia and Microcephaly." Hum Mutat **37**(2): 170-174.

Zhang, R., L. Chen, S. Jiralerspong, A. Snowden, S. Steinberg and N. Braverman (2010). "Recovery of PEX1-Gly843Asp peroxisome dysfunction by small-molecule compounds." Proc Natl Acad Sci U S A **107**(12): 5569-5574.

Zhang, X., T. G. Diekwisch and X. Luan (2011). "Structure and function of ameloblastin as an extracellular matrix protein: adhesion, calcium binding, and CD63 interaction in human and mouse." Eur J Oral Sci **119 Suppl 1**: 270-279.

Zhao, C., E. A. Matveeva, Q. Ren and S. W. Whiteheart (2010). "Dissecting the N-ethylmaleimide-sensitive factor: required elements of the N and D1 domains." J Biol Chem **285**(1): 761-772.

Zheng, Q. Y., J. D. Scarborough, Y. Zheng, H. Yu, D. Choi and P. G. Gillespie (2012). "Digenic inheritance of deafness caused by 8J allele of myosin-VIIA and mutations in other Usher I genes." Hum Mol Genet **21**(11): 2588-2598.

Zou, J., L. Luo, Z. Shen, V. A. Chiodo, B. K. Ambati, W. W. Hauswirth and J. Yang (2011). "Whirlin replacement restores the formation of the USH2 protein complex in whirlin knockout photoreceptors." Invest Ophthalmol Vis Sci **52**(5): 2343-2351.

Appendices

Appendix 1: Non-syndromic AI genes

Gene	Inheritance pattern	Human enamel phenotype	Proposed function in amelogenesis	Proposed disease mechanism	Number of mutations reported[a]	References
Amelogenin, X linked (*AMELX*)	X linked dominant	Females: -stripes of normal and AI affected enamel -hypoplastic (some hypomaturation). Males: -Large deletions and N terminal mutations cause hypomaturation AI with variable hypoplasia -C terminal mutations cause hypoplastic AI	-EMP -Structural component of developing enamel matrix. -Alternative splicing and cleavage by proteinases leads to formation of peptide products with distinct roles.	**Loss of function:** Variety of mutations identified including large deletions, frameshift, stop and missenses. **? Dominant negative:** One mutation known to cause a change in splicing bias towards more frequent inclusion of exon 4 in transcripts, may result in decrease of miRNA originating from exon 4 that affects *AMTN* and *ODAM* expression.	21	Brookes et al., 1995; Hart et al., 2000; Veis et al., 2000; Lagerstrom et al., 2001; Hart et al., 2002; Veis et al., 2003; Cho et al., 2014; Le et al., 2016.

Gene	Inheritance pattern	Human enamel phenotype	Proposed function in amelogenesis	Proposed disease mechanism	Number of mutations reported[a]	References
Ameloblastin (*AMBN*)	Autosomal recessive	-hypoplastic AI	-EMP -Influences ameloblast proliferation and differentiation. -Undergoes PTM and proteinase cleavage to form products with distinct roles, including one able to bind Ca^{2+}. -Possible roles in cell adhesion and extracellular signalling.	**Loss of function:** Mutations include an in-frame deletion of exon 6 and a mutation affecting splicing believed to either cause retention of intron 6 or to result in the skipping of exon 7.	2	Smith et al., 1998; Bartlett and Simmer, 1999; Fukumoto et al., 2004; Beyeler et al., 2010; Poulter et al., 2014; Prasad et al., 2016;
Enamelin (*ENAM*)	Autosomal dominant and recessive	Dominant: -severe smooth hypoplastic AI or milder local hypoplastic AI Recessive: -severe hypoplastic AI.	-EMP -Undergoes cleavage by proteinases to form peptides with distinct roles, including binding HA crystals.	**Dominant negative:** Dominant and recessive severe form. **Loss of function:** Local milder form caused by missense mutations.	18	Rajpar et al., 2001; Mardh et al., 2002; Hart et al., 2003; Hu and Yamakoshi, 2003; Ozdemir et al., 2005.
Amelotin (*AMTN*)	Autosomal dominant	-hypomineralised AI	-Formation of compact aprismatic enamel	**? Dominant negative gain of function:** In-frame deletion reported that is hypothesised to lead to accumulation of protein aggregates within the ER.	1	Abbarin et al., 2015; Nakayama et al., 2015; Núñez et al., 2015; Smith et al., 2016.

Gene	Inheritance pattern	Human enamel phenotype	Proposed function in amelogenesis	Proposed disease mechanism	Number of mutations reported[a]	References
Matrix metalloproteinase 20 (*MMP20*)	Autosomal recessive	-hypomaturation AI	-Secretory stage cleavage of EMPs. -Activation of KLK4 by cleavage. -Mediation of ameloblast cell movement by cleavage of extracellular domains of cadherins. -May affects ameloblast development.	**? Loss of function:** Includes stop and frameshift and missense variants.	8	Simmer and Hu, 2002; Kim et al., 2005; Bartlett et al., 2011; Guan and Bartlett, 2013; Yamakoshi et al., 2013; Guan et al., 2016.
Kallikrein-related peptidase 4 (*KLK4*)	Autosomal recessive	-hypomaturation AI	-Maturation stage enamel matrix proteinase. -Further degrades MMP20 peptide cleavage products. -Inactivation of secretory stage proteinase, MMP20.	**? Loss of function:** Frameshift or stop variants including one within the final exon.	3	Hu et al., 2000; Hu et al., 2002; Hart et al., 2004; Simmer et al., 2009; Yamakoshi et al., 2013.
WD repeat-containing protein 72 (*WDR72*)	Autosomal recessive	-hypomaturation AI	-Endocytosis by maturation stage ameloblasts. -Removes EMP degradation products to allow enamel to mineralise.	**Loss of function:** All null or frameshift variants that lead to NMD apart from one missense found in a compound heterozygous individual in *trans* with a null variant.	9	El-Sayed et al., 2009; Katsura et al. 2014; Wang et al., 2015; Prasad et al. 2016.

Gene	Inheritance pattern	Human enamel phenotype	Proposed function in amelogenesis	Proposed disease mechanism	Number of mutations reported[a]	References
Family with sequence similarity 83, member H (*FAM83H*)	Autosomal dominant	-hypocalcified AI	-Associates with perinuclear vesicles. -Cytoskeletal reorganisation, desmosome formation.	**Dominant negative:** All variants within final exon. All except one are a stop or a frameshift.	26	Foster and Xu, 2003; Mendoza et al., 2007; Kim et al., 2008; Ding et al., 2009; Kuga et al., 2016; Wang et al., 2016.
Solute carrier family 24 (Sodium/potassium/calcium exchanger), member 4 (*SLC24A4*)	Autosomal recessive	-hypomaturation / hypomoineralised AI	-Transport of Ca^{2+} by RE.	**Loss of function:** Both missense and stop variants identified.	5	Parry et al., 2013; Wang et al., 2014; Bronckers et al., 2015.
Chromosome 4 open reading frame 26 (*C4orf26*)	Autosomal recessive	-hypomineralised AI	-Unknown. -Promotes nucleation and crystallisation *in vitro.*	**Loss of function:** All frameshift or stop variants predicted to escape NMD.	5	Parry et al., 2012.
G-protein coupled receptor 68 (*GPR68*)	Autosomal recessive	-hypomineralised AI	-Functions as a pH sensor. -Expressed throughout amelogenesis.	**Loss of function:** Includes a large deletion, frameshift and missense variants.	3	Parry et al., 2016

Gene	Inheritance pattern	Human enamel phenotype	Proposed function in amelogenesis	Proposed disease mechanism	Number of mutations reported[a]	References
Integrin, beta 6 (*ITGB6*)	Autosomal recessive	-hypoplastic / hypmineralised AI	-Cell surface adhesion receptor. -Mediates cell-cell and cell-extracellular matrix interactions. -Able to activate TGF-β1 and other signalling molecules. -Localises to maturation stage ameloblasts.	**Loss of function:** Includes a stop and missense variants.	5	Munger et al., 1999; Poulter et al., 2014; Wang et al., 2014.
Laminin, alpha 3 (*LAMA3*)	Autosomal dominant	-hypoplastic AI -grooved and pitted	-Control of ameloblast differentiation and adhesion to the enamel surface.	**Haploinsufficiency:** Frameshift variants predicted to be subject to NMD.	1 (+1 with *COL17A1* variant)	Carter et al., 1991; Marchisio et al., 1993; Yoshiba et al., 1998; Yuen et al., 2012.
Laminin, beta 3 (*LAMB3*)	Autosomal dominant	-hypoplastic AI -grooved and pitted	-Control of ameloblast differentiation and adhesion to the enamel surface.	**Dominant negative / Haploinsufficiency:** All but one are frameshift and stop mutations predicted to escape NMD.	7[b]	Carter et al., 1991; Marchisio et al., 1993; McGrath et al., 1995; Yoshiba et al., 1998; Kim et al., 2013; Poulter et al., 2014.
Collagen, type XVII, alpha-1 (*COL17A1*)	Autosomal dominant	-Hypoplastic AI -grooved and pitted	-Ligand of LM-332 (heterotrimeric protein consisting of LAMA3, LAMB3 and LAMC2) see *LAMA3* / *LAMB3*	**Haploinsufiiciency:** Frameshift, stop and missense variants identified.	10 (+1 with *LAMA3* variant)	McGrath et al., 1996; Floeth and Bruckner-Tuderman, 1999; Murrell et al., 2007.

Gene	Inheritance pattern	Human enamel phenotype	Proposed function in amelogenesis	Proposed disease mechanism	Number of mutations reported[a]	References
Family with sequence similarity 20, member A (*FAM20A*)	Autosomal recessive	-hypoplastic AI[c]	-Predicted to be a pseudokinase that might enhance the action of FAM20C, a Golgi kinase with targets associated with mineralisation.	**Loss of function:** Majority are frameshift or stop variants predicted to lead to NMD. Variants predicted to affect splicing also identified.	38	Jaureguiberry et al., 2012; Wang et al., 2013; Cui et al., 2015.

Non-syndromic AI genes.

The enamel phenotype, the proposed function of the encoded protein in amelogenesis, the proposed disease mechanism and the number of mutations reported, are listed. The list includes amelotin for which mutations are reported in chapter 4 of this work. Note that not all of the references for the identification of human mutations are cited. See Leeds NS AI LOVD for further details of the mutations identified.

[a] Number of mutations reported is as of 25th May 2016, apart from for *GPR68*, for which the report identifying the first three mutations is in review and *FAM20A*, for which the LOVD entries are not yet complete (*FAM20A* is curated by Dr. Poulter (University of Leeds).

[b] Due to the severity of the JEB phenotype, the tooth phenotype of carriers is not often reported / reported in sufficient detail.

[c] There are also often a variety of associated oral defects present, including delayed tooth eruption, impaction, microdontia and gingival overgrowth, suggesting that not only amelogenesis but also tooth morphogenesis is affected. Dental pulp calcification and ectopic gingival and dental follicle calcification have also been associated with ERS (de la Dure-Molla et al., 2014).

Appendix 2: Mouse models of non-syndromic AI

Gene	Human Inheritance pattern	Proposed function in amelogenesis	Mouse enamel phenotype	Mouse model references
Amelogenin, X linked (*AMELX*)	X linked dominant	-EMP -Structural component of developing enamel matrix. -Alternative splicing and cleavage by proteinases leads to formation of peptide products with distinct roles.	***Amelx$^{Y/-}$*** -less than 10% of WT enamel volume. -enamel lacked normal prismatic structure ("Flat, plate like"). ***Amelx$^{Y/Y64H}$*** -roughened, opaque, chalky white enamel.	Gibson et al., 2001; Barron et al., 2010.
Ameloblastin (*AMBN*)	Autosomal recessive	-EMP -Influences ameloblast proliferation and differentiation. -Undergoes PTM and proteinase cleavage to form products with distinct roles, including one able to bind Ca^{2+}. -Possible roles in cell adhesion and extracellular signalling.	***Ambn$^{-5,6/-5,6}$*** -severely hypoplastic enamel. -ameloblasts lost contact with the enamel matrix. -Tomes' processes failed to develop. ***p.Krt14:Ambn$^{+/+}$*** -altered rod structure. -loss of surface aprismatic enamel. -altered craniofacial development.	Fukumoto et al., 2004; Smith et al., 2009; Wazen et al., 2009; Paine et al., 2003; Atsawasuwan et al., 2013.
Enamelin (*ENAM*)	Autosomal dominant and recessive	-EMP -Undergoes cleavage by proteinases to form peptides with distinct roles, including binding HA crystals.	***Enam* LoF** -hypomaturation AI in heterozygotes. -total lack of enamel in homozygotes. ***Enam$^{-1-7/-1-7}$*** -homozygotes lacked true enamel. -heterozygotes had discoloured incisors that wore rapidly.	Masuya et al., 2005; Seedorf et al., 2007; Hu et al., 2008.
Amelotin (*AMTN*)	Autosomal dominant	-Formation of compact aprismatic enamel.	***Amtn$^{-/-}$*** -chalky maxillary incisors. -rough irregular surface enamel that is easily chipped away. ***pAmelx:Amtn$^{+/+}$*** brittle, thin enamel. defective enamel surface layer.	Lacruz et al., 2012; Nakayama et al., 2015; Núñez et al., 2016.

Gene	Human Inheritance pattern	Proposed function in amelogenesis	Mouse enamel phenotype	Mouse model references
Matrix metalloproteinase 20 (*MMP20*)	Autosomal recessive	-Secretory stage cleavage of EMPs. -Activation of KLK4 by cleavage. -Mediation of ameloblast cell movement by cleavage of extracellular domains of cadherins. -May affect ameloblast development.	***Mmp20*$^{-/-}$** -defective AMELX processing. -loss of the typical rod structure. -reduced mineral content. -enamel softer than WT. ***pAmelx:Mmp20*$^{+/+}$** -hypomineralised enamel. -large amounts of smaller EMP cleavage products.	Caterina et al., 2002; Shin et al., 2014; Hu et al., 2015.
Kallikrein-related peptidase 4 (*KLK4*)	Autosomal recessive	-Maturation stage enamel matrix proteinase. -Further degrades MMP20 peptide cleavage products. -Inactivation of secretory stage proteinase, MMP20.	***Klk4*$^{-/-}$** -hypomaturation AI. -enamel crystallites did not function as a unit. -enamel easily abraded. -enamel breaks just above DEJ.	Simmer et al., 2009; Yamakoshi et al., 2013; Hu et al., 2015; Núñez et al., 2016.
WD repeat-containing protein 72 (*WDR72*)	Autosomal recessive	-Endocytosis by maturation stage ameloblasts. -Removes EMP degradation products to allow enamel to mineralise.	***Wdr72*$^{-/-}$** -hypomaturation AI. -EMP degradation products retained within the enamel during maturation. -RE may not form. -enamel remains unmineralised. -disrupted cell matrix attachment to the enamel matrix during maturation.	Katsura et al., 2014; Wang et al., 2015.
Family with sequence similarity 83, member H (*FAM83H*)	Autosomal dominant	-Associates with perinuclear vesicles. -Cytoskeletal reorganisation, desmosome formation.	***Fam83h*$^{-/-}$** -majority die within 2 weeks. -no defect in enamel or dentine. -hair present around base of teeth. -scruffy coat. -smaller pulp volume. ***pActb:Fam83h*$^{+/+}$** -no defects in enamel or dentine.	Kweon et al., 2013; Wang et al., 2015.

Gene	Human Inheritance pattern	Proposed function in amelogenesis	Mouse enamel phenotype	Mouse model references
Solute carrier family 24 (sodium/ potassium/ calcium exchanger), member 4 (*SLC24A4*)	Autosomal recessive	-Transport of Ca^{2+} by RE.	***Nckx4*$^{-/-}$** -enamel breaks off incisor after eruption to expose underlying dentine.	Stephan et al., 2012; Parry et al., 2013.
Chromosome 4 open reading frame 26 (*C4orf26*)	Autosomal recessive	-Unknown. -Promotes nucleation and crystallisation *in vitro.*	No model yet characterised.	N/A
G-protein coupled receptor 68 (*GPR68*)	Autosomal recessive	-Functions as a pH sensor. -Expressed throughout amelogenesis.	***Ogr1*$^{-/-}$** -no major defects. -incisor enamel structure mildly affected. -delayed yellowing of maxillary incisor teeth.	Mogi et al., 2009; Parry et al., 2016.
Integrin, beta 6 (*ITGB6*)	Autosomal recessive	-Cell surface adhesion receptor. -Mediates cell-cell and cell-extracellular matrix interactions. -Able to activate TGF-β1 and other signalling molecules. -Localises to maturation stage ameloblasts.	***Itgb6*$^{-/-}$** -hypomineralised AI with disorganised enamel prisms and pits. -retained AMELX within the enamel.	Mohazab et al., 2013.
Laminin, alpha 3 (*LAMA3*)	Autosomal dominant	-Control of ameloblast differentiation and adhesion to the enamel surface.	***Lama3*$^{-/-}$** -smaller ameloblasts. -abnormal enamel deposition. -disorganised reduced enamel epithelium.	Ryan et al., 1999.
Laminin, beta 3 (*LAMB3*)	Autosomal dominant	-Control of ameloblast differentiation and adhesion to the enamel surface.	***Lamb3*$^{-/-}$** -model dies within 24hr of birth. -no assessment of the teeth carried out.	Kuster et al., 1997.

Gene	Human Inheritance pattern	Proposed function in amelogenesis	Mouse enamel phenotype	Mouse model references
Collagen, type XVII, alpha-1 (*COL17A1*)	Autosomal dominant	-Ligand of LM-332 (heterotrimeric protein consisting of LAMA3, LAMB3 and LAMC2). -See *LAMA3* / *LAMB3.*	***Col17a1*$^{-/-}$** -no tooth phenotype described.	Hurskainen et al., 2012.
Family with sequence similarity 20, member A (*FAM20A*)	Autosomal recessive	-Predicted to be a Golgi kinase with targets specific to mineralisation, calcium transport and proteoglycan synthesis or a pseudokinase that might enhance the action of FAM20C.	***Fam20a*$^{-/-}$** -ameloblast layer disorganised and detached from the DEJ. -pitted, thin enamel.	Vogel et al., 2012.

Mouse models of non-syndromic AI gene function.

The proposed roles of the proteins encoded by genes in which mutations are reported to cause non-syndromic AI (note that amelotin is also included) and their respective mouse models and phenotypes are described. Mouse models are defined using standard nomenclature where possible. Overexpression models where the promoter of another gene is driving expression are denoted: p(gene promoter driving expression):(gene under control). Where specific exons have been removed this is reflected in the nomenclature used. For *Amelx*, which is located on the X chromosome, the nomenclature used refers to hemizygous male mutants. Abbreviations used: DEJ dentine enamel junction; EMP enamel matrix protein; HA hydroxy(l)apatite; hr hour; LOF loss of function; PTM post translational modification; RE ruffle ended ameloblasts; WT wild-type.

Appendix 3: Usher syndrome classification

USH type	Representation relative to total cases	Phenotype			Genes associated	MIM reference
		Hearing loss	Vestibular function	Ocular		
I	Common	Profound, congenital	Absent	Onset before puberty	*MYO7A* *USH1C* *CDH23* *PCDH15* *USH1G* *CIB2*	*276903 *605242 *605516 *605514 *607696 *605564
II	Most common	Prelingual, moderate to severe, especially high frequencies	Normal	Onset in 2nd decade	*USH2A* *ADGRV1* *DFNB31*	*608400 *602851 *607928
III	Rare except in certain populations, e.g. Finnish, Ashkenazi.	Variable onset, progressive to profound	Variable	Variable onset	*CLRN1* *HARS*	*606397 *142810
Atypical	Rare	Variable	Variable	Variable	Various	Various

Usher syndrome phenotype classifications and their associated genes.

Appendix 4: Whole exome sequencing commands

Alignment

Align fastq files to GRCh37 reference genome and produce SAM file:

```
$ novoalign -c 12 -d human_g1k_v37.nix -f sample_R1_001_fastq.gz
sample_R2_001_fastq.gz -o SAM
$'@RG\tID:sample_novo\tSM:sample\tPL:Illumina\tLB:sample_exome' -k -K
mismatches_sample_novo.txt 2>novostats_sample_novo.txt >
sample_novo.sam
```

Sort alignment and convert SAM to bam using Picard:

```
$ java -Xmx2g -jar /picard/picard-tools/SortSam.jar I=input_unsorted.sam
O=input.bam SO=coordinate CREATE_INDEX=TRUE
```

Mark duplicates using Picard:

```
$ java -Xmx4g -jar /picard/picard-tools/MarkDuplicates.jar I=input.bam
O=input.rmdups.bam M=input.rmdups.metrics CREATE_INDEX=TRUE
```

Create indel realigner targets using GATK:

```
$ java -Xmx4g -jar GenomeAnalysisTK.jar -T RealignerTargetCreator -R
human_g1K_v37.fasta -known 1000G_phase1.indels.b37.vcf -known
Mills_and_1000G_gold_standard.indels.b37.sites.vcf -I input.rmdups.bam -o
input.rmdups.indelrealign.intervals
```

Perform Indel realignment using GATK:

```
$ java -Xmx4g -jar GenomeAnalysisTK.jar -T IndelRealigner -R
human_g1k_v37.fasta -known 1000G_phase1.indels.b37.vcf -known
Mills_and_1000G_gold_standard.indels.b37.sites.vcf -I input.rmdups.bam
-targetIntervals input.rmdups.indelrealign.intervals -o
input.rmdups.indelrealign.bam
```

Recalibrate Base Quality Scores using GATK.
Obtain the recalibration model:
$ java -Xmx4g -jar GenomeAnalysisTK.jar -T BaseRecalibrator -I input.rmdups.indelrealign.bam -R human_g1k_v37.fasta -o input.rmdups.indelrealign.recal.grp -knownSites dbSnp142.b37.vcf.gz -knownSites 1000G_phase1.indels.b37.vcf -known Mills_and_1000G_gold_standard.indels.b37.sites.vcf

Check the recalibration model:
$ java -Xmx4g -jar GenomeAnalysisTK.jar -T BaseRecalibrator -I input.rmdups.indelrealign.bam -R human_g1k_v37.fasta -BQSR input.rmdups.indelrealign.recal.grp -o input.rmdups.indelrealign.postrecal.grp -knownSites dbSnp142.b37.vcf.gz -knownSites 1000G_phase1.indels.b37.vcf -known Mills_and_1000G_gold_standard.indels.b37.sites.vcf

$ java -Xmx4g -jar GenomeAnalysisTK.jar -T AnalyzeCovariates -R human_g1k_v37.fasta -before input.rmdups.indelrealign.recal.grp -after input.rmdups.indelrealign.postrecal.grp -plots input.rmdups.indelrealign.postrecal.plots.pdf

Apply the recalibration:
$ java -Xmx4g -jar GenomeAnalysisTK.jar -T PrintReads -R human_g1k_v37.fasta -o input.rmdups.indelrealign.recal.bam -I input.rmdups.indelrealign.bam -BQSR input.rmdups.indelrealign.recal.grp

Variant calling

Call variants using Haplotype caller:
$ java -Xmx4g - jar GenomeAnalysisTK.jar -T HaplotypeCaller -R human_g1k_v37.fasta -D dbSnp142.b37.vcf.gz -L all_Ref_Seq_genes.b37.bed -stand_call_conf 30 -stand_emit_conf 10 -I input.rmdups.indelrealign.recal.bam -o var.input.rmdups.indelrealign.recal.raw.vcf

Select indels and SNP variants using GATK and split multiallelic variants using VCFhacks.

Select SNPs:

```
$ java -Xmx4g - jar GenomeAnalysisTK.jar -T SelectVariants -R
human_g1k_v37.fasta --variant var.input.rmdups.indelrealign.recal.raw.vcf
-selectType SNP -o var.input.rmdups.indelrealign.recal.raw.SNPs.vcf
```

Split SNPs:

```
$ perl /home/vcfhacks/splitMultiallelicVariants.pl -i
var.input.rmdups.indelrealign.recal.raw.SNPs.vcf -o
var.input.rmdups.indelrealign.recal.raw.SNPs.split.vcf
```

Select indels:

```
$ java -Xmx4g - jar GenomeAnalysisTK.jar -T SelectVariants -R
human_g1k_v37.fasta --variant var.input.rmdups.indelrealign.recal.raw.vcf
-selectType INDEL -selectType MNP -o
var.input.rmdups.indelrealign.recal.raw.indels.vcf
```

Split indels:

```
$ perl /home/vcfhacks/splitMultiallelicVariants.pl -i
var.input.rmdups.indelrealign.recal.raw.indels.vcf -o
var.input.rmdups.indelrealign.recal.raw.indels.split.vcf
```

Perform hard filtering separately on SNPs and indels using GATK.

SNPs:

```
$ java -Xmx4g - jar GenomeAnalysisTK.jar -T VariantFiltration -R
human_g1k_v37.fasta -V var.input.rmdups.indelrealign.recal.raw.SNPs.split.vcf
--filterExpression “QD < 2.0 || FS > 60.0 || MQ < 40.0 || HaplotypeScore > 13.0
|| MappingQualityRankSum < -12.5 || ReadPosRankSum < -8.0” --filterName
“snp_hard_filter” -o var.input.rmdups.indelrealign.recal.filtered.SNPs.split.vcf
```

Indels:

```
$ java -Xmx4g - jar GenomeAnalysisTK.jar -T VariantFiltration -R
human_g1k_v37.fasta -V var.input.rmdups.indelrealign.recal.raw.indels.split.vcf
```

--filterExpression “QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0”
--filterName “indel_hard_filter” -o
var.input.rmdups.indelrealign.recal.filtered.indels.split.vcf

Combine SNPs and indels using GATK:
$ java -Xmx4g - jar GenomeAnalysisTK.jar -T CombineVariants -R
human_g1k_v37.fasta --variant
var.input.rmdups.indelrealign.recal.filtered.SNPs.split.vcf --variant
var.input.rmdups.indelrealign.recal.filtered.indels.split.vcf -o
var.input.rmdups.indelrealign.recal.filtered.combined.vcf --genotypemergeoption
UNSORTED

Variant filtering and annotation

For recessive disease, variants were filtered if present at ≥1% in publically available databases. For dominant disease, variants were filtered if present at ≥0.1% in publically available databases. The examples below show the filtering strategy for recessive disease.

Filter on dbSNP142 at 1%:
$ perl /home/vcfhacks/annotateSnps.pl -d human_g1k_v37.fasta
clinvar_20150330.vcf.gz -f 1 -pathogenic -i
var.input.rmdups.indelrealign.recal.filtered.combined.vcf -o
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_1pc
(Optional: -b 129)

Filter EVS at 1%:
$ perl /home/vcfhacks/filterOnEvsMaf.pl -d /home/ref/evs/ -f 1 --progress -i
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_1pc -o
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_1pc

Filter on ExAC at 1%:
$ perl /home/vcfhacks/filterVcfOnVcf.pl -f
ExAC_release/release0.3/ExAC.r0.3/sites.vep.vcf.gz -w -y 0.01 -i
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_1pc

-o
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc

Filter on control samples:
$ perl /home/vcfhacks/filterOnSample.pl -i
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc -r samples_to_filter -o
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_samplefilter

Annotate variants with VEP:
$ perl /home/variant_effect_predictor/variants_effect_predictor.pl --offline --vcf --everything --dir /home/variant_effect_predictor/vep_cache --plugin Condel,/home/variant_effect_predictor/vep_cache/Plugins/config/Condel/config/ --plugin SpliceConsensus --assembly GRCh37 -input
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc -o
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_VEP

Get functional variants:
$ /home/vcfhacks/getFunctionalVariantsVep.pl -i
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_VEP --consensus_splice_site -o
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_VEP_func

Find biallelic variants in the same gene in more than one family:
$ /home/vcfhacks/findBiallelicVep.pl -i
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_VEP_func -s sample_name -f name.ped -n 2 -o
var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_VEP_func_biall_fam

Annotate with Ensembl Gene Annotator:

$ perl /home/vcfhacks/ensemblGeneAnnotator -d /home/ref/ensAnnotator/ -i var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_VEP_func_biall_fam -o var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_VEP_func_biall_fam.geneanno

Convert to Excel file format (.xlsx):

$ perl /home/vcfhacks/annovcfToSimpleVep.pl -i var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_VEP_func_biall_fam.geneanno --vep --gene_anno -o var.input.rmdups.indelrealign.recal.filtered.combined.vcf_dbSNP142_EVS_ExAC_1pc_VEP_func_biall_fam.geneanno.simple

Depth of Coverage

$ java -Xmx10g -jar GATK.jar -T DepthOfCoverage -R human_g1k_v37.fasta -I input.rmdups.indelrealign.bam -o output.coveragedepth.txt -L SSV4/5_regions_b37.bed -ct 4 -ct 9 -ct 14 -ct 19 -ct 24 -ct 29

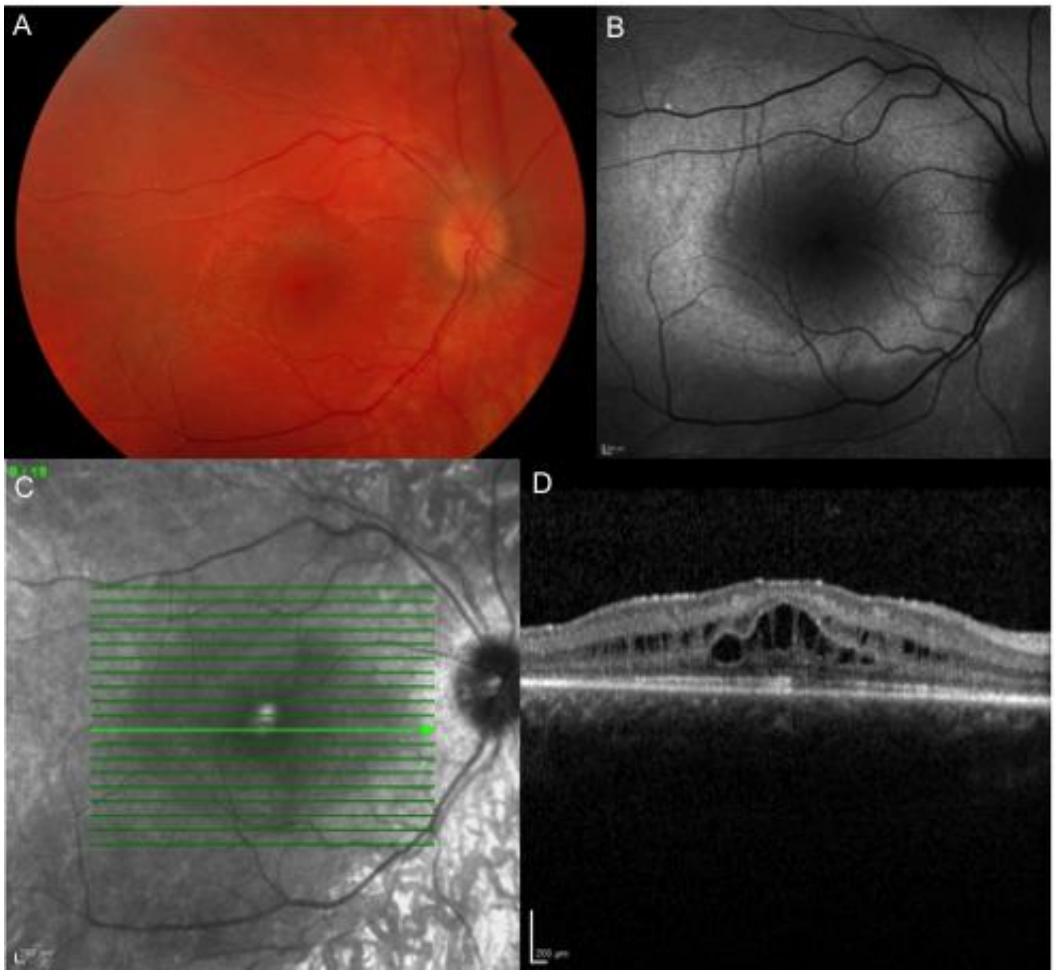
Appendix 5: Primer sequences for Chapter 3

Gene	Exon	Forward Sequence (5′-3′)	Reverse Sequence (5′-3′)	Product (bp)	$[Mg^{2+}]/A_T$/Add.
USH2A	13	TCATTTTCCCATCCTCACCT	ACCCCTAATTTGCAAGGACA	496	1.5/59/-
USH2A	22	AAGAATCAACAAAACCCCAGC	CCTCAGTACCAGGCACCTAC	244	1.5/59/-
USH2A	60	TGCAAAAGGACAGGTTAAAATAA	TCCTCACTGGTAGGTGCTACA	451	1.5/59/-
USH2A	61	GTGTGCAGCTGTCACTGGTT	TTCATATATTTAAGCCCTAAGTGAAGA	507	1.5/60/-
MYO7A	2	AGCATGACATGGTCTCTCTCC	ACAGCAGAGGGAACAGGTTT	264	1.5/58/-
MYO7A	13	TGATGGGGATAGCTTGCTAA	GGGCTCAAAGCAGGGAAG	389	1.5/58/-
MYO7A	16	CTGTCCCTCAAACCCTGACC	CATCCTTCACTCCCACCCTA	344	1.5/58/-
MYO7A	32	GAATCAGTGAGAAGCCTGGG	AGGTTCAGATGAGGCCTGG	397	1.5/58/B
MYO7A	35	GGCCAGCTCTGACTTAGCCT	CATAAATCTCCCAGCCTCCA	490	1.5/59/B
MYO7A	38	CTCACAACCTGGGGGTTTCT	AAGGCAGTGTGCAGACGAA	300	1.5/58/-
MYO7A	42	AGGGGGCTCAGTATAGGAGG	CTCTCTTCCGTCTGCTGTCC	288	1.5/58/-
ENAM	6	TGGCTGAGTTTTAGAGGCTGA	AGTGTGTATATGGGGGTGGC	433	1.5/59/-
PEX1	10	ACTGGGAAGGCAAAATTCAG	TGGTCAAAACCCAAAGAAAGAAAGAT	296	1.5/59/-
PEX1	13	GCGATACCATACATACTGCACC	GAAGCATAAATTTAAAGCCACGA	375	3/59/H D
PEX1	19	ACCTGGCAGAAGTAAAGCTCA	GCTGCTACTAGTCGCCCTGA	166	3/58/H
PEX6	1(a)	CCTTTTCCTCTGGCCTCC	GAGCTCAGTCACAGCCAGC	597	3/55/H D
PEX6	1(b)	GTCTTGCGGGTCCTGGAG	GGCTGTGAAGTGTTCGATGA	698	3/58/H D
PEX6	1(c)	ACTCGGCTGGCTGTGACT	CAAAGTCCGGGATGATATGG	536	3/58/H
PEX6	4-5	TCCATCTGCATTCCTTTTCC	AGGAAGGTCCTCCCAATCC	585	3/58/H
PEX6	8	CCTAGCACCCACCTCACTTC	TCACAAGGCAACAGGACTGA	387	3/58/H
PEX6	13	CATGTATGTGGGCCAAAGTG	TGGACTCTGAAGACTGCTGTG	483	3/58/H
PEX6	16	CATGCAACATGCAGGATGAG	GGTCTCTCTGTGTTGCCCA	371	3/58/H

Primers sequences used for research presented in Chapter 3.

Abbreviations are as follows: $[Mg^{2+}]$ Mg^{2+} concentration present in the PCR reaction (mM); AT Annealing Temperature used in ^{o}C; Add. Additives used to enhance the PCR; H Hot Shot Mastermix (Section 2.5.3); B 1M betaine; D 10% DMSO. Exons listed as 1(a), 1(b), 1(c) show that one large exon was amplified using multiple primers that produced overlapping PCR products to cover the entire exon.

Appendix 6: Clinical details of patients in Chapter 3



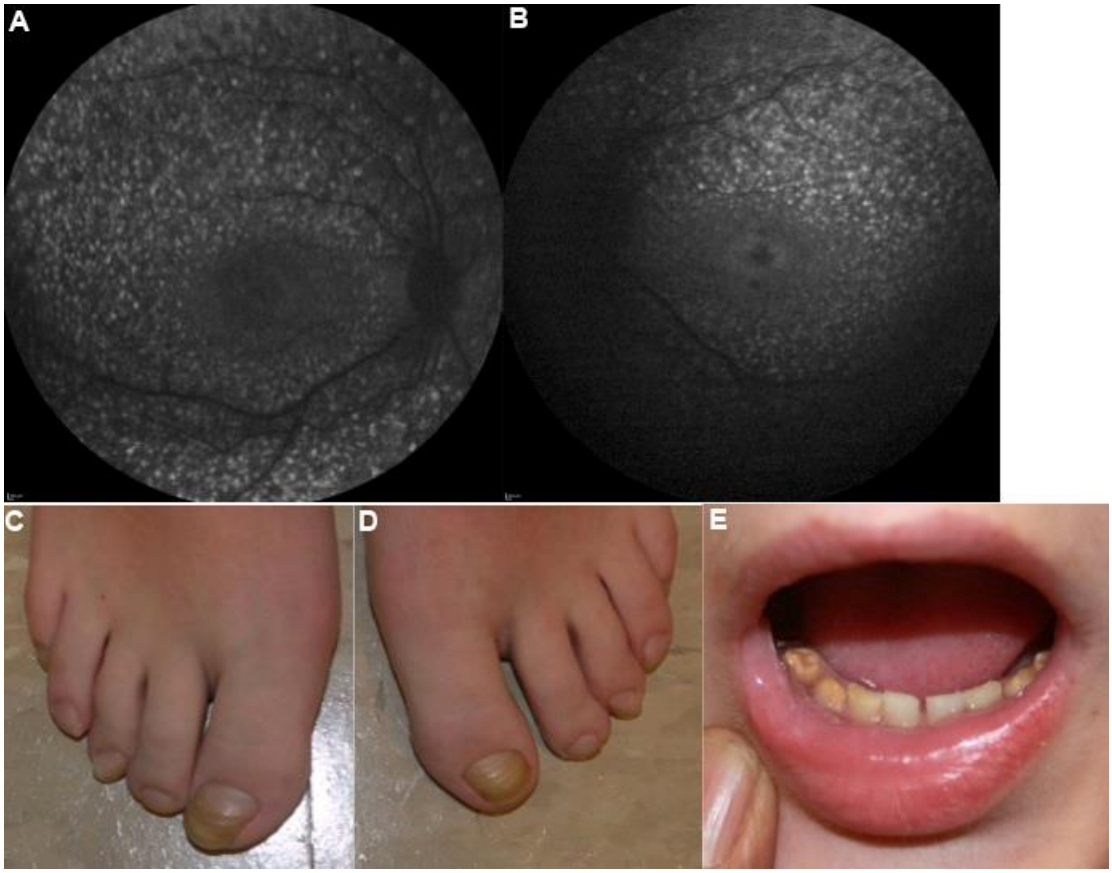
Clinical detail of the retinal phenotype of individual 2975 (family AI-138).

Only the right eye is shown.

A: Fundus image showing peripheral retinal changes without marked pigmentation and a normal optic nerve.

B: Fundus autofluorescence image showing central hypofluorescence and peripheral hyperfluorescence.

C and D: Optical coherence tomography images, the bold green line shown in C indicates the plane (through the macula) at which the image in D was obtained. D shows macular oedema leading to disruption of the retinal layers.



Clinical images of 3000 (family AI-146).

A and B: Fundus autofluorescence of right (A) and left eye (B). Note the spots of abnormal hyperfluorescence.

C and D: Imagess of right (C) and left (D) feet show toenails with Beau's lines that are most apparent on the halluces.

E: Image of part of the lower dentition. AI is most apparent in the canine and molar teeth.

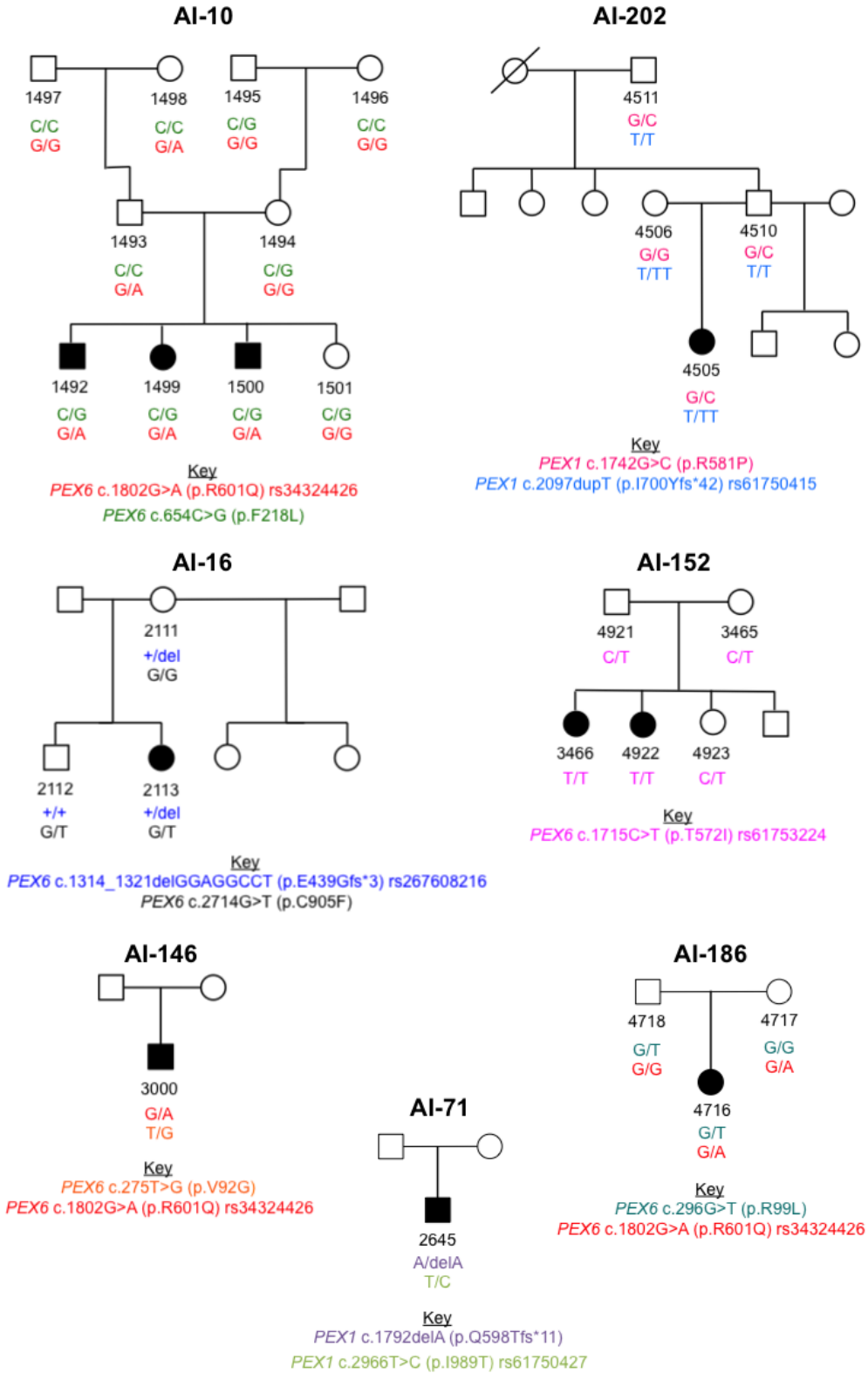
Clinical Images of 2645 (family AI-71).

A and B: Images of hands that display dry, split skin.

C to E: Images of lower limbs; right leg (C), legs (D) and feet (E) that display dry, split skin.

F to N: Images of dentition: thirteen teeth have been extracted and those that remain display AI.

Appendix 7: Segregation of *PEX1* and *PEX6* variants



Segregation analysis of *PEX1* and *PEX6* variants in families AI-10, AI-16, AI-71, AI-146, AI-152, AI-186 and AI-202.

Affected individuals are denoted by shading. Genotypes for each *PEX1* or *PEX6* variant identified are also presented.

Appendix 8: Whole exome sequencing alignment statistics for Heimler syndrome samples

Family	Sample	Input DNA /μg	Capture reagent	Mean depth of coverage	% bases covered by >4 reads	% bases covered by >9 reads	% bases covered by >14 reads	% bases covered by >19 reads	% bases covered by >24 reads	% bases covered by >29 reads	Total read count
AI-10	1499	2.7	SSV4	89.76	99.4	98.0	95.9	92.9	89.4	85.4	4,594,835,166
AI-10	1500	1.6	SSV5	37.84	98.0	93.9	86.8	77.9	68.3	58.8	1,905,827,228
AI-11	1503	3.0	SSV4	84.89	99.3	97.9	95.5	92.3	88.5	84.3	4,345,645,192
AI-11	1504	3.0	SSV4	80.24	99.2	97.4	94.7	91.1	86.9	82.3	4,107,490,638
AI-12	1509	2.0	SSV4	91.75	99.5	98.1	95.9	92.9	89.3	85.3	4,696,371,551
AI-12	1514	2.7	SSV4	103.79	99.5	98.4	96.6	94.2	91.2	87.9	5,313,092,258
AI-16	2113	3.0	SSV5	65.44	98.8	97.5	95.5	92.3	88.2	83.2	3,295,512,156
AI-71	2645	3.0	SSV5	67.92	98.5	96.9	94.6	91.3	87.2	82.3	3,420,312,345
AI-138	2975	3.0	SSV5	76.63	98.9	97.9	96.3	93.9	90.7	86.9	3,859,017,557
AI-146	3000	1.9	SSV5	58.23	98.9	97.5	94.8	90.5	84.9	78.5	2,932,584,366
AI-152	3466	3.0	SSV5	64.68	98.6	97.3	95.1	91.9	87.5	82.4	3,257,378,994
AI-152	4922	3.0	SSV5	76.82	99.0	98.3	97.1	95.3	92.6	89.0	3,868,923,901
AI-186	4716	3.0	SSV5	83.23	98.9	98.0	96.7	94.8	92.2	89.0	4,191,686,607
AI-202	4505	3.0	SSV5	61.56	98.4	97.0	94.7	91.1	86.4	80.8	3,099,974,781

Details of DNA input, capture reagent and alignment statistics for WES of HS DNA samples.

The standard input for the WES protocol used is 3μg. Alignment statistics were generated using the regions targetted by the relevant capture reagent as the interval. Abbreviations used: SSV4 SureSelectXT Human All Exon V4; SSV5 SureSelectXT Human All Exon V5.

Appendix 9: Family AI-12 SNVs and indels

Genomic variant (GRCh37)	dbSNP142 reference	Amino acid change	Gene	CADD v1.3	Provean	SIFT	Polyphen2 (HumVar)	Mutation Taster	Ensembl transcript	RefSeq protein
1:215901562delTG	N/A	p.Q3959*	*USH2A*	35	N/A	N/A	N/A	Disease causing (1.000)	ENST 00000 307340	NP_996816
1:216420437delC	rs80338903	p.E767X	*USH2A*	33	N/A	N/A	N/A	Disease causing (1.000)	ENST 00000 307340	NP_996816
1:216270469G>A	rs111033333	p.L1572F	*USH2A*	26.8	Neutral (-2.39)	Damaging (0.011)	Probably damaging (0.885)	Disease causing (0.979)	ENST 00000 307340	NP_996816
15:22862978G>A	rs148108126	p.G667S	*TUBGCP5*[a]	15.22	Neutral (-0.44)	Tolerated (0.187)	Benign (0.276)	Disease causing (0.999)	ENST 00000 283645	NP_443135
11:1016934G>A	N/A	p.S1956L	*MUC6*[a]	15.21	Neutral (-1.63)	Tolerated (0.092)	Benign (0.001)	Polymor -phism (0.999)	ENST 00000 421673	NP_005952

Details of the five genomic variants identified by WES as shared in individuals 1509 and 1514 (family AI-12) after filtering.

Biallelic variants present in 1509 and 1514 (family AI-12) were selected. These were filtered by removing those with a MAF of 1% or more in dbSNP142, EVS or ExAC and by selecting only biallelic variants not present in two unrelated control samples.

The list is restricted to variants scoring 15 or more when scored with CADD v1.3. All variants listed are heterozygous.

Provean, SIFT and Mutation Taster annotations were based on Ensembl transcript references, Polyphen-2 annotations were based on RefSeq transcript references.

[a]The corresponding biallelic variants for these genes scored <15 and so are not included here.

All scores were obtained 11th February 2016.

Appendix 10: Family AI-138 SNVs and indels

Genomic variant (GRCh37)	Zygo-sity	dbSNP142 reference	Amino acid change	Gene	CADD v1.3	Provean	SIFT	Polyphen 2 (HumVar)	Ensembl transcript	RefSeq transcript / protein
12:42499692 A>T	Het.	rs77044712	p.Y264*	*GXYLT1*	36	N/A	N/A	N/A	ENST 00000398675	NP_775872
14:64066363 G>A	Het.	N/A	p.R100*	*WDR89*	35	N/A	N/A	N/A	ENST 00000394942	NP_542397
12:42499835 C>T	Het.	N/A	p.D217N	*GXYLT1*	34	Deleterious (-4.52)	Damaging (0)	Probably damaging (1)	ENST 00000398675	NP_775872
1:216420437 delC	Het.	rs80338903	p.E767X	*USH2A*	33	N/A	N/A	N/A	ENST 00000307340	NP_996816
12:42499802 C>A	Het.	rs78540738	p.D228Y	*GXYLT1*	29.8	Deleterious (-5.55)	Damaging (0)	Possibly damaging (0.885)	ENST 00000398675	NP_775872
18:5890864 C>A	Het.	rs577682765	p.S400I	*TMEM200C*	27.2	Neutral (0.09)	Damaging (0.032)	Neutral (0.154)	ENST 00000383490	NP_001073678
1:216270469 G>A	Het.	rs111033333	p.L1572F	*USH2A*	26.8	Neutral (-2.39)	Damaging (0.011)	Probably damaging (0.885)	ENST 00000307340	NP_996816
14:103999171 C>A	Het.	N/A	p.A195D	*TRMT61A*[a]	26.8	Deleterious (-4.85)	Damaging (0.007)	Probably damaging (0.978)	ENST 00000389749	NP_689520
12:42499825 A>C	Het.	rs76034661	p.I220S	*GXYLT1*	26.5	Deleterious (-3.87)	Damaging (0.002)	Possibly damaging (0.579)	ENST 00000398675	NP_775872
16:2120487 G>A	Het.	rs1800729	p.A583T	*TSC2*	26.2	Deleterious (-3.67)	Damaging (0)	Probably damaging (0.997)	ENST 00000219476	NP_000539
1:215901511 G>A	Het.	rs142381713	p.T3976M	*USH2A*	25.5	Neutral (-2.43)	Tolerated (0.056)	Probably damaging (0.982)	ENST 00000307340	NP_996816

Genomic variant (GRCh37)	Zygo-sity	dbSNP142 reference	Amino acid change	Gene	CADD v1.3	Provean	SIFT	Polyphen 2 (HumVar)	Ensembl transcript	RefSeq transcript / protein
12:42499690 T>C	Het.	rs79044728	p.Y265C	*GXYLT1*	25.5	Deleterious (-8.29)	Damaging (0)	Probably damaging (0.986)	ENST 00000398675	NP_775872
12:42499701 C>A	Het.	rs74583427	p.R261S	*GXYLT1*	25.1	Deleterious (-5.1)	Damaging (0.045)	Probably damaging (0.823)	ENST 00000398675	NP_775872
12:42499694 A>T	Het.	N/A	p.Y264N	*GXYLT1*	25.0	Deleterious (-7.75)	Damaging (0.079)	Probably damaging (0.978)	ENST 00000398675	NP_775872
14:106780706 C>T	Het.	rs199610585	p.W53*	*IGHV4-28*[a]	25.0	N/A	N/A	N/A	ENST 00000390612	N/A
16:2124321 C>A	Het.	rs45517238	p.L826M	*TSC2*	24.4	Neutral (-1.15)	Damaging (0.01)	Probably damaging (0.996)	ENST 00000219476	NP_000539
1:215914880 C>T	Het.	N/A	p.splice	*USH2A*	24.1	N/A	N/A	N/A	ENST 00000307340	NP_996816
11:1016892 G>A	Het.	rs78784632	p.S1970F	*MUC6*	24.1	Neutral (-2.11)	Tolerated (0.06)	Probably damaging (0.982)	ENST 00000421673	NP_005952
19:8999468 C>G	Het.	N/A	p.W13569C	*MUC16*	23.8	Deleterious (-3.07)	Damaging (0.017)	Possibly damaging (0.88)	ENST 00000397910	NP_078966
19:52888076 A>ATG	Het.	N/A	p.K415Mfs*15	*ZNF880*	23.3	N/A	N/A	N/A	ENST 00000422689	NP_001138906
18:5890865 T>G	Het.	rs543455233	p.S400R	*TMEM200C*	23.1	Neutral (-1.08)	Damaging (0.0333)	Benign (0.054)	ENST 00000383490	NP_001073678
19:52888077 A>AGGTCAGGAGATCG	Het.	N/A	p.A416Vfs*7	*ZNF880*	23.0	N/A	N/A	N/A	ENST 00000422689	NM_001145434 NP_001138906
17:38911309 C>T	Het.	N/A	p.R72Q	*KRT25*	22.9	Neutral (0.06)	Tolerated (0.167)	Benign (0.002)	ENST 00000312150	NP_853512

Genomic variant (GRCh37)	Zygo-sity	dbSNP142 reference	Amino acid change	Gene	CADD v1.3	Provean	SIFT	Polyphen 2 (HumVar)	Ensembl transcript	RefSeq transcript / protein
11:1016871 G>T	Het.	N/A	p.P1977H	*MUC6*	22.7	Neutral (-1.49)	Tolerated (0.05)	Benign (0.091)	ENST 00000421673	NP_005952
19:8999488 C>T	Het.	rs200071553	p.D13563N	*MUC16*	22.3	Neutral (-0.69)	Tolerated (0.171)	Probably damaging (0.968)	ENST 00000397910	NP_078966
4:3076609 G>GCAGCA GCAGCAGC AGCAGCAG	Hom.	N/A	p.?	*HTT*	18.12	Deleterious (-2.95)	N/A	N/A	ENST 00000355072	NP_002102
19:8999514 C>T	Het.	rs548751371	p.R13554H	*MUC16*	17.77	Neutral (-1.52)	Tolerated (0.277)	Possibly damaging (0.715)	ENST 00000397910	NP_078966
17:38907503 T>C	Het.	N/A	p.T249A	*KRT25*	17.49	Deleterious -3.31	Damaging (0.05)	Benign (0.58)	ENST 00000312150	NP_853512
14:64066367 T>A	Het.	N/A	p.D98E	*WDR89*	15.72	Unscored	Damaging (0)	Possibly damaging (0.695)	ENST 00000349942	NP_542397
11:1016934 G>A	Het.	N/A	p.S1956L	*MUC6*	15.21	Deleterious (-1.63)	Tolerated (0.092)	Probably damaging (0.945)	ENST 00000421673	NP_005952

Details of the thirty genomic variants identified by WES in individual 2975 (family AI-138) after filtering.

Biallelic variants present in 2975 were selected and were filtered by removing those with a MAF of 1% or more in dbSNP142, EVS or ExAC and by selecting only biallelic variants not present in two unrelated control samples. The list is restricted to variants scoring 15 or more when scored with CADD v1.3.

Provean and SIFT annotations were based on Ensembl transcript references, Polyphen-2 annotations were based on RefSeq transcript references.

[a]The corresponding biallelic variants for these genes scored <15 and so are not included here.

Abbreviations: Het. heterozygous; Hom. homozygous.

All scores were obtained 11th February 2016.

Appendix 11: Whole genome sequencing alignment statistics

Family	Sample	Depth of coverage	% bases covered by >1 reads	% bases covered by >2 reads	% bases covered by >3 reads	% bases covered by >4 reads	Total read count
AI-10	1492	0.52	39.5	9.5	1.7	0.3	1481295764
AI-11	1502	0.72	49.9	16.2	4.0	0.9	2070178231
AI-16	2113	0.65	46.6	13.7	3.0	0.6	1863736476
AI-71	2645	0.76	51.5	17.4	4.5	1.0	2174646482
AI-152	3466	0.42	33.5	6.6	1.0	0.2	1195708362
AI-186	4716	0.58	43.0	11.3	2.2	0.4	1653366880

Alignment statistics for low coverage whole genome sequencing of HS DNA samples.

Alignment statistics were generated using the GRCh37 assembly as the interval.

Appendix 12: Family AI-11 SNVs and indels

Genomic variant (GRCh37)	dbSNP142 reference	Amino acid change	Gene	CADD v1.3	Provean	SIFT	Polyphen 2 (HumVar)	Mutation Taster	Ensembl transcript	RefSeq protein
14:64066363 G>A	N/A	p.R100*	*WDR89*	35	N/A	N/A	N/A	Disease causing (1)	ENST 00000 394942	NP_542397
16:48221232 G>C	rs55713504	p.S938*	*ABCC11*	35	N/A	N/A	N/A	Disease causing (1)	ENST 00000 394747	NP_149163
7:20823926 G>A	rs368503334	p.R486C	*SP8*[a]	25.4	Neutral (-2.25)	Damaging (0.001)	Probably damaging (0.973)	Disease causing (0.999)	ENST 00000 361443	NP_874359
1:12907518 T>A	rs201750607	p.S209C	*HNRNPCL1*	23	Neutral (-1.27)	Damaging (0.037)	Probably damaging (0.994)	Disease causing (0.969)	ENST 00000 317869	NP_001139653
8:110503307 T>C	N/A	p.I3364T	*PKHD1L1*	22.9	Deleterious (-3.24)	Damaging (0.006)	Benign (0.15)	Polymor -phism (0.836)	ENST 00000 378402	NP_803875
16:48226474 G>A	rs200992288	p.T888M	*ABCC11*	22.7	Deleterious (-4.44)	Damaging (0.029)	Benign (0.148)	Polymor -phism (0.999)	ENST 00000 394747	NP_115972
1:12907519 C>A	rs6702447	p.Q208H	*HNRNPCL1*	22.1	Neutral (0)	Tolerated (0.096)	Probably damaging (0.989)	Disease causing (0.830)	ENST 00000 317869	NP_001139653
1:12907446 C>T	rs116578678	p.E233K	*HNRNPCL1*	20.4	Neutral (-1.02)	Tolerated (0.117)	Benign (0.031)	Polymor -phism (0.561)	ENST 00000 317869	NP_001139653
8:110498978 G>A	rs200880307	p.G3270S	*PKHD1L1*	19.90	Neutral (-0.68)	Tolerated (0.646)	Benign (0.017)	Polymor -phism (0.973)	ENST 00000 378402	NP_803875
11:1016881 C>T	N/A	p.G1974R	*MUC6*[a]	19.33	Neutral (-1.05)	Damaging (0.019)	Benign (0.114)	Polymor -phism (0.999)	ENST 00000 421673	NP_005952

Genomic variant (GRCh37)	dbSNP142 reference	Amino acid change	Gene	CADD v1.3	Provean	SIFT	Polyphen 2 (HumVar)	Mutation Taster	Ensembl transcript	RefSeq protein
14:64066367 A>T	N/A	p.D98E	*WDR89*	15.72	Deleterious (-3.53)	Damaging (0)	Possibly damaging (0.695)	Disease causing (0.999)	ENST 00000 394942	NP_542397

Details of the eleven genomic variants identified by WES as shared in individuals 1503 and 1504 (family AI-11) after filtering.

Biallelic variants present in 1503 and 1504 (family AI-11) were selected. These were filtered by removing those with a MAF of 1% or more in dbSNP142, EVS or ExAC and by selecting only biallelic variants not present in two unrelated control samples. The list is restricted to variants scoring 15 or more when scored with CADD v1.3. All variants listed are heterozygous.

Provean, SIFT and Mutation Taster annotations were based on Ensembl transcript references, Polyphen-2 annotations were based on RefSeq transcript references.

[a]The corresponding biallelic variants for these genes scored <15 and so are not included here.

All scoring accessed 12th February 2016.

Appendix 13: Family AI-11 *ENAM* variant details

Genomic variant (GRCh37)	Transcript variant	Amino acid change	Exon	dbSNP142 reference	CADD v1.3	Provean	SIFT	Polyphen2 (HumVar)	Mutation Taster
4:71500146A>C	c.332A>C	p.K111T	6	rs532584416	16.01	Neutral (-0.29)	Tolerated (0.087)	Benign (0.368)	Polymorphism (0.999)

Details of the heterozygous genomic *ENAM* variant identified by WES in family AI-11.

The list is restricted to variants scoring 15 or more when scored with CADD v1.3.

Provean, SIFT and Mutation Taster annotations were based on Ensembl transcript ENST00000396073, Polyphen-2 annotations were based on RefSeq protein NP_114095.

All scores were obtained 12th February 2016.

Genomic variant (GRCh37)	dbSNP142 reference; number of variant alleles (total alleles at this position); allele frequency (all reported individuals heterozygous)	ClinVar (release 07/01/2016)	ExAC (v0.3) number of variant alleles (total alleles at this position); allele frequency (all reported individuals heterozygous)	EVS number of variant alleles (total alleles at this position); allele frequency (all reported individuals heterozygous)
4:71500146A>C	rs532584416; 1 (5008); 2 x 10^{-4}	N/A	31 (121310); 2.56 x10^{-4}	N/P

Details of the known variation in public databases for the *ENAM* variant identified in family AI-11.

Abbreviation is as follows: N/P Not present.

All databases accessed 12th February 2016.

Appenndix 14: Family AI-202 SNVs and indels

Genomic variant (GRCh37)	Zygo-sity	dbSNP142 reference	Predicted amino acid change	Gene	CADD v1.3	Provean	SIFT	Polyphen2 (HumVar)	Mutation Taster	Ensembl transcript	RefSeq transcript / protein
11:1017041 G>T	Het.	rs75482640	p.Y1920*	*MUC6*[a]	42	N/A	N/A	N/A	Disease causing (1)	ENST 00000 421673	NP_ 005952
7:92132483 T>TA	Het.	N/A	p.I700Yfs* 42	*PEX1*	27.6	N/A	N/A	N/A	Disease causing (1)	ENST 00000 248633	NP_ 000457
2:70524515 T>A	Het.	rs201866753	p.D108V	*FAM136A*[a]	25.9	Deleterious (-6.13)	Damaging (0.001)	Possibly damaging (0.643)	Disease causing (0.999)	ENST 00000 037869	NP_ 116211
6:117128308 C>T	Het.	rs145460545	p.R187Q	*GPRC6A*	25.3	Deleterious (-3.9)	Damaging (0)	Probably damaging (1)	Disease causing (0.996)	ENST 00000 310357	NP_ 683766
20:50803572 C>T	Het.	N/A	p.D29N	*ZFP64*	23.4	Neutral (-0.95)	Damaging (0.035)	Probably damaging (0.990)	Disease causing (0.999)	ENST 00000 361387	NP_ 060667
18:21483967 C>T	Het.	rs199684626	p.S2130F	*LAMA3*[a]	22.6	Deleterious (-5.11)	Damaging (0.001)	Probably damaging (0.999)	Disease causing (0.999)	ENST 00000 313654	NP_ 937762
2:141116420 C>T	Het.	rs150879175	p.G3743S	*LRP1B*	22.1	Deleterious (-3.74)	Tolerated (0.1)	Probably damaging (0.998)	Disease causing (0.999)	ENST 00000 389484	NP_ 061027
7:92136369 C>G	Het.	N/A	p.R581P	*PEX1*	21.8	Deleterious (-3.16)	Damaging (0.006)	Probably damaging (0.923)	Disease causing (0.926)	ENST 00000 248633	NP_ 000457
16:70164435 C>A	Het.	N/A	p.N239K	*PDPR*	21.1	Deleterious (-5.17)	Damaging (0.004)	Probably damaging (0.958)	Disease causing (0.999)	ENST 00000 288050	NP_ 060460

Genomic variant (GRCh37)	Zygo-sity	dbSNP142 reference	Predicted amino acid change	Gene	CADD v1.3	Provean	SIFT	Polyphen2 (HumVar)	Mutation Taster	Ensembl transcript	RefSeq transcript / protein
6:117113765 T>TTTCC	Het.	N/A	p.FS	*GPRC6A*	21.0	N/A	N/A	N/A	Disease causing (1)	ENST 00000 310357	NP_ 683766
2:97851077 C>A	Het.	N/A	p.S655*	*ANKRD36*[a]	20.8	N/A	N/A	N/A	Disease causing (1)	ENST 00000 420699	NP_ 001157787
2:141032021 T>A	Het.	rs149644677	p.N4372Y	*LRP1B*	20.2	Neutral (-1.44)	Damaging (0.008)	Benign (0.123)	Disease causing (0.896)	ENST 00000 389484	NP_ 061027
16:70162749 C>A	Het.	rs199978223	p.R142C	*PDPR*	18.82	Deleterious (-5.01)	Damaging (0.001)	Probably damaging (1)	Disease causing (0.999)	ENST 00000 288050	NP_ 060460
20:50781298 T>TAA	Hom.	rs72411807	p.splice	*ZFP64*	18.75	N/A	N/A	N/A	Disease causing (1)	ENST 00000 361387	NP_ 060667
3:126722305 G>A	Het.	rs138997671	p.E504K	*PLXNA1*[a]	17.72	Deleterious (-2.8)	Tolerated (0.084)	Benign (0.132)	Disease causing (0.999)	ENST 00000 393409	NP_ 115618
1:228509579 C>T	Het.	rs199935165	p.R5970W	*OBSCN*	17.50	Deleterious (-5.54)	Damaging (0)	Probably damaging (0.997)	Disease causing (0.994)	ENST 00000 570156	NP_ 001092093
11:48387284 G>A	Het.	N/A	p.S245L	*OR4C5*[a]	16.37	Deleterious (-5.45)	Damaging (0)	Not scored	Polymorphi sm (0.999)	ENST 00000 319813	N/A
6:32489745 CGGCGCGC CTGTCT>C	Het.	N/A	p.D99Rfs* 26	*HLA-DRB5*[a]	15.81	N/A	N/A	N/A	Disease causing (1)	ENST 00000 374975	NP_ 002116
13:25671163 T>A	Het.	N/A	p.L276H	*PABPC3*[a]	15.65	Deleterious (-6.67)	Damaging (0)	Probably damaging (0.999)	Disease causing (0.999)	ENST 00000 281589	NP_ 112241

Genomic variant (GRCh37)	Zygo-sity	dbSNP142 reference	Predicted amino acid change	Gene	CADD v1.3	Provean	SIFT	Polyphen2 (HumVar)	Mutation Taster	Ensembl transcript	RefSeq transcript / protein
1:228467095 T>C	Het.	rs200720682	p.V2878A	*OBSCN*	15.50	Deleterious (-2.52)	Damaging (0.041)	Possibly damaging (0.622)	Disease causing (1)	ENST 00000 570156	NP_ 001092093

Details of the twenty genomic variants identified by WES in individual 4505 (family AI-202) after filtering.

Biallelic variants present in 4505 (family AI-202) were selected. These were filtered by removing those with a MAF of 1% or more in dbSNP142, EVS or ExAC and by selecting only biallelic variants not present in two unrelated control samples. The list is restricted to variants scoring 15 or more when scored with CADD v1.3.

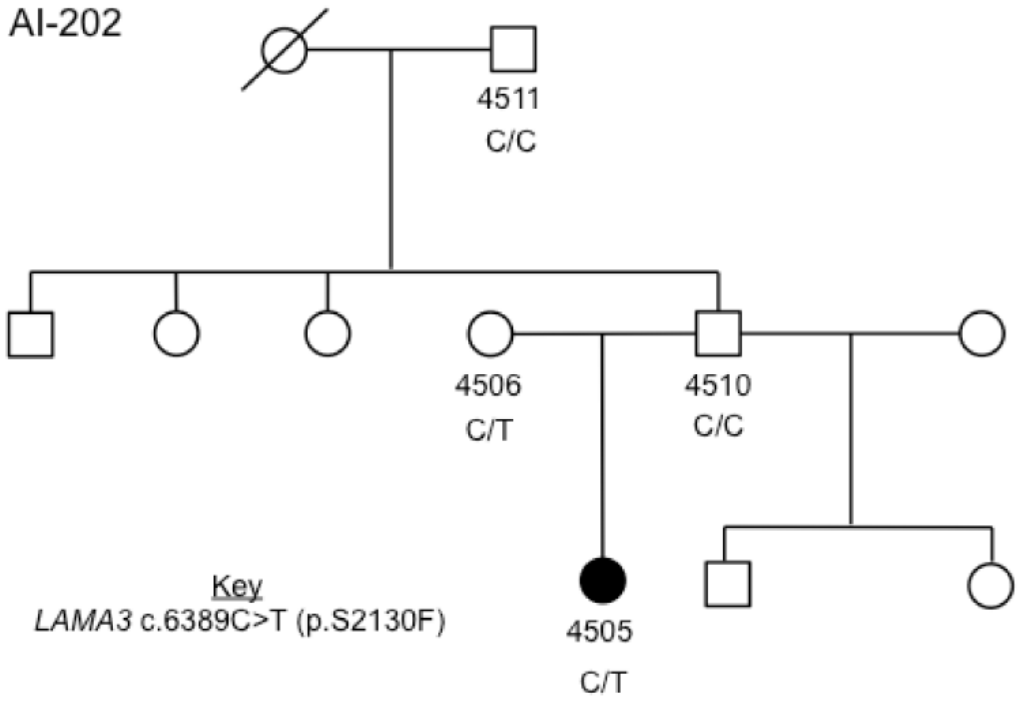
Provean, SIFT and Mutation Taster annotations were based on Ensembl transcript references, Polyphen-2 annotations were based on RefSeq transcript references.

[a]The corresponding biallelic variants for these genes scored <15 and so are not included here.

Abbreviations: Het. heterozygous; Hom. homozygous.

All scoring accessed 15th February 2016.

Appendix 15: Family AI-202 *LAMA3* variant segregation



Segregation of heterozygous *LAMA3* c.6389C>T variant in family AI-202.

Refseq transcript: NM_198129; Refseq protein: NP_937762.

Appendix 16: Markers and SNPs used to analyse the haplotype surrounding the *PEX6* c.1802G>A variant

Marker	Physical location (hg19 / GRCh37)	Genetic location			Forward primer (5′-3′)	Reverse primer (5′-3′)	Size (bp)
		Genethon	Marshfield	Decode			
D6S1552 AFM157XE11	41955110-41955373	63.50	63.28	N/A	AGCCTGAACGACAGAACAAG	CTGCTTAACTTNAGATCTTTGGTAT	156-184
D6S400 UT2081	42156381-42156758	N/A	63.28	N/A	AGCCTGGGTAACTTAGTGAG	TCTTTGCTCAAGTCTTTCTCC	182
D6S1582 AFMA272ZB5	43098617-43098961	65.20	65.14	N/A	CTAGGTAGTCAGGTGGTCATAGTC	AGTAGGGCTGGAACCTTCT	133-151
D6S282 AFM184XA11	43234575-43234945	66.40	66.37	66.10	ATGGCCCAGACAGTGGGTAT	ATGGTTTGTGCAGGTTCAGA	108-126
D6S271 AFM136YF8	43500818-43501193	66.40	66.37	66.44	AACAATTGGGAAATGGCTTA	TAGGTTGTGGTGGGTGTTAC	166-208
D6S1604 AFMB007XA9	43679700-43680083	66.40	66.37	N/A	CTGGGACTACAGGCATGAGC	CTAGGACTGGGCAGGATTTG	189-205
D6S1650 AFMB334YG9	44130031-44130426	68.70	68.65	N/A	GGGCTCCACTGTTTAACC	CCCTCGGGCTATGATTAC	113-126

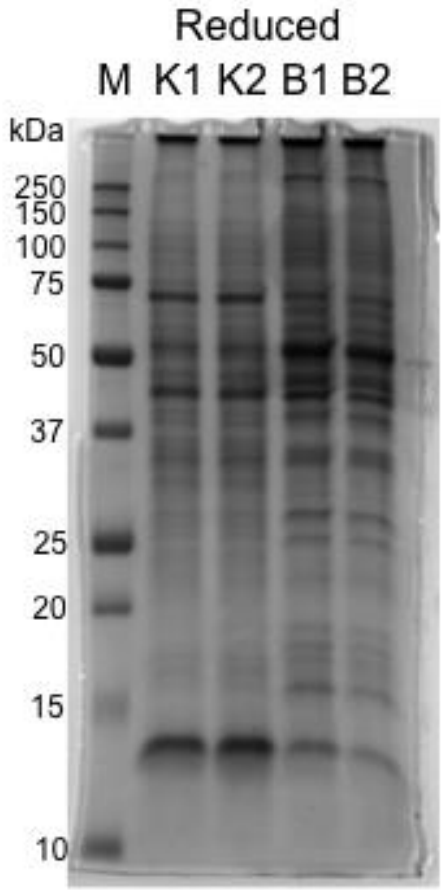
Summary of chromosome 6 Sequence Tagged Site (STS) markers genotyped to analyse the haplotype surrounding the shared *PEX6* c.1802G>A variant identified in families AI-10, AI-146 and AI-186.

dbSNP build 142	Physical location (hg19 / GRCh37)	Reference (allele frequency)	Alternative (allele frequency)	Forward Primer (5′-3′)	Reverse Primer (5′-3′)
rs7770760	42915403	C (52.935%)	T (47.064%)	TCCTGGGTTCACGCCATT	TGCTAGCAGATCCACCCAC
rs9471978	42929533	C (50.319%)	T (49.681%)	GATGAAGTCGTGCTCTGTCG	CCATTTGAGAGACAGACAAGCT
rs34324426	42935188	C (99.900%)	T (0.100%)	CCTAGCACCCACCTCACTTC	TCACAAGGCAACAGGACTGA

Summary of the Single Nucleotide Polymorphisms (SNPs) on chromosome 6 sequenced to analyse the haplotype surrounding the shared *PEX6* c.1802G>A variant (rs34324426) identified in families AI-10, AI-146 and AI-186.

Allele frequencies relate to dbSNP build 142.

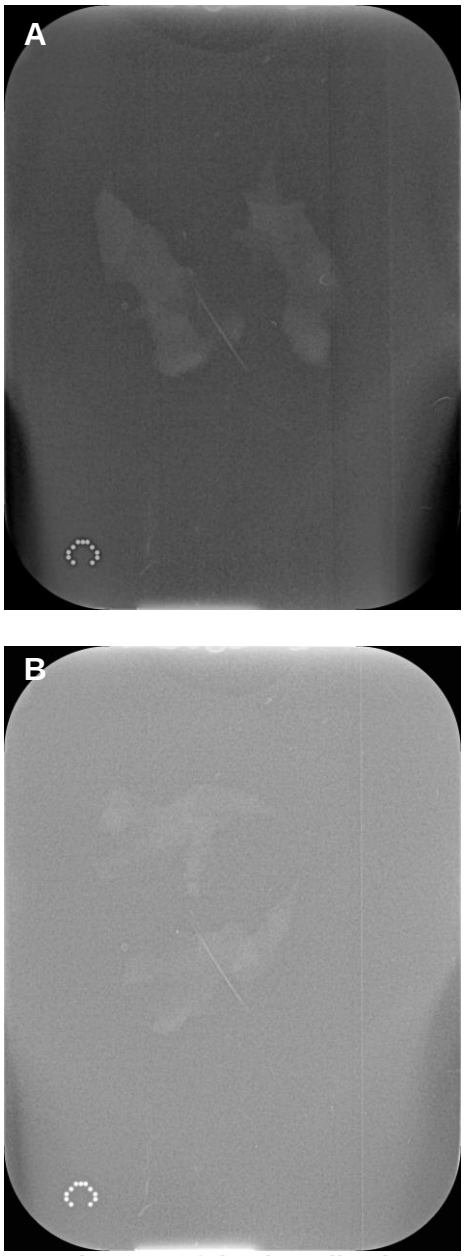
Appendix 17: Coomassie blue stained SDS PAGE gel



Coomassie stained SDS PAGE gel.

50µg total protein samples from both kidney (K) and brain (B) from two mice were electrophoresed by SDS PAGE under reducing conditions.

Appendix 18: X-ray photographs of decalcified mouse jaws



X-ray images of demineralised mouse jaws.

A and B: Radiographs of a demineralised mouse jaw. An overexposed image is represented in B to show that demineralisation is complete. Note the absence of any distinct region of incisor or molar tooth that is more radio-dense than the remainder of the jaw.

Appendix 19: Previously reported *ENAM* variants

Genomic variant (GRCh37)	Transcript variant	Predicted amino acid change	Reference(s) including dbSNP142 (frequency)	CADD v1.3	Provean	SIFT	Polyphen-2	Mutation Taster
4:71497439delA	c.107delA	p.(N36Ifs*22)	Simmer et al., 2013	29.2	N/A	N/A	N/A	Disease causing (1)
4:71497456G>A	c.123+1G>A	p.?	Prasad et al., 2016	26.9	N/A	N/A	N/A	Disease causing (1)
4:71497581delA	c.139delA	p.(M47Cfs*11)	Wang et al., 2015	35	N/A	N/A	N/A	Disease causing (1)
4:71497599A>T	c.157A>T	p.(K53*)	Mårdh et al., 2002 Kim et al., 2006 rs121908109 (no frequency data)	43	N/A	N/A	N/A	Disease causing (1)
4:71500023A>C	c.211-2A>C	p.(M71_Q157del)	Kim et al., 2005	23.9	N/A	N/A	N/A	Disease causing (1)
4:71500146A>C	c.332A>C	p.(K111T)	This study: rs532584416 (2×10^{-4})	16.02	Neutral (-0.29)	Tolerated (0.087)	Benign (0.368)	Polymorphism (1)
4:71500172C>T	c.358C>T	p.(Q120*)	Seymen et al., 2014	37	N/A	N/A	N/A	Disease causing (1)
4:71500221_71500221insTCAAAAAAGCCGACCACAA	c.407_408insTCAAAAAAGCCGACCACAA	p.(K136Nfs*16)	Wang et al., 2015	25.8	N/A	N/A	N/A	Disease causing (1)
4:71500268G>T	c.454G>T	p.(Q152*)	Seymen et al., 2014	39	N/A	N/A	N/A	Disease causing (1)
4:71501612G>A	c.534+1G>A	p.(A158_Q178del)	Rajpar et al., 2001 Song et al., 2012	26.7	N/A	N/A	N/A	Disease causing (1)
4:71503505A>G	c.535-2A>G	p.(R179-N196del)	Wright et al., 2011	23.3	N/A	N/A	N/A	Disease causing (1)
4:71503508G>T	c.536G>T	p.R179M	Guitierrez et al., 2007	21.5	Neutral (-0.87)	Tolerated (0.084)	Probably damaging (0.911)	Polymorphism (0.996)

Genomic variant (GRCh37)	Transcript variant	Predicted amino acid change	Reference(s) including dbSNP142 (frequency)	CADD v1.3	Provean	SIFT	Polyphen-2	Mutation Taster
4:71503561delG	c.588+1delG	p.(N197Ifs*81) p.(?)	Kida et al., 2002 Hart PS et al., 2003 Kim et al., 2005 Pavlic et al., 2007 Wright et al., 2011	24.3	N/A	N/A	N/A	Disease causing (1)
4:71507790C>T	c.647G>T	p.(S216L)	Chan et al., 2010	26.4	Deleterious (-5.93)	Damaging (0.002)	Probably damaging (1)	Disease causing (0.996)
4:71507880C>A	c.737C>A	p.(S246*)	Ozdemir et al., 2005	36	N/A	N/A	N/A	Disease causing (1)
4:71508164_71508165insGTCAGTACCAGTACTGTGTCA	c.1021_1022insGTCAGTACCAGTACTGTGTC	p.(V340_M341insSQYQYCV)	Ozdemir et al., 2005	0.001	Deleterious (-13.38)	N/A	N/A	Polymorphism (1)
4:71508402_71508403insAG	c.1259_1260insAG	p.(V422Pfs*27)	Hart, TC et al., 2003 Ozdemir et al., 2005 Pavlic et al., 2007 Kang et al., 2009 Chan et al., 2010 Lindemeyer et al., 2010	23.3	N/A	N/A	N/A	Disease causing (1)
4:71510134delT	c.2991delT	p.(L998Wfs*65)	Kang et al., 2009	26.1	N/A	N/A	N/A	Disease causing (1)

***ENAM* variants reported in individuals with AI and in AI-11.**

The variant identified in AI-11 is shaded grey.

Provean, SIFT and Mutation Taster annotations were based on the Ensembl transcript ENST00000396073, Polyphen-2 annotations were based on the RefSeq protein NP_114095.

All scores were obtained 2nd March 2016.

Appendix 20: *USH2A* c.2299delG haplotypes analysis for families AI-12 and AI-138

Variant	dbSNP142	Family AI-12	Aller et al. (2010) haplotype		Family AI-138	Aller et al. (2010) haplotype
			M6	M7		M3
c.373G>A p.A125T	rs10779261	A/A	A	A	A/A	A
c.504A>G p.T168=	rs4253963	G/G	G	G	G/G	G
c.1419C>T p.T473=	rs1805050	C/C	C	C	C/C	C
c.2299delG p.E767Sfs*21	rs80338903	G/-	-	-	G/-	-
c.3157+35G>A p.?	rs1324330	A/A	A	A	A/A	A
c.4457G>A p.R1486K	rs1805049	A/A	A	A	A/A	A
c.4714C>T p.L1572F	rs111033333	C/T	T	T	C/T	T
c.6506T>C p.I2169T	rs10864219	T/C	T	C	T/C	T
c.6857G>A p.R2292H	rs41277210	G/G	G	G	G/G	G
c.10232A>C p.E3411A	rs10864198	C/C	C	C	A/A	A
c.11602A>G p.M3868V	rs35309576	A/A	A	A	A/A	A
c.11677C>A p.P3893T	rs41303285	C/C	C	C	C/C	C
c.12612A>G p.T4204=	rs2797235	A/G	G	G	A/A	A
c.12666A>G p.T4222=	rs2797234	A/A	A	A	A/A	A
c.13191G>A p.E4397=	rs2009923	G/A	G	G	A/A	A

Comparison of the sequences identified at specific SNPs within the *USH2A* gene for families AI-12 and AI-138 with the genotypes previously identified in European patients.

Aller et al., (2010) described ten haplotypes associated with the *USH2A* c.2299delG mutation (labelled M1 to M10). The genotypes of the SNPs used to identify the ten haplotypes are listed for AI-12 and AI-138 from available WES data. For simplicity, the transcript variant (Ensembl ENST00000307340) not the genomic variant is presented. The predicted protein change is based on RefSeq protein NP_996816. Underlining indicates the allele of each heterozygous call that the c.2299delG mutation is present on (confirmed by Sanger sequencing). For the genotypes highlighted grey, it is uncertain which allele carries the c.2299delG mutation. A selection of the haplotypes identified by Aller et al., (2010) are presented for comparison (M6, M7 and M3) and represent the closest matches to the SNP calls obtained for AI-12 and AI-138 WES data.

Appendix 21: Primer sequences for Chapter 4

Amplicon	Forward primer (5′-3′)	Reverse primer (5′-3′)	Size (bp)	[$MgCl_2$]/A_T/Add
exon 2	TTGACCATGTTTCAAGTAGAACTTTT	AGGGCAGGAGAGAGGTAGGT	246	3/59/H
exon 3	AACCTGGATATAAATGGACACAAA	AACGTGTTTTCTTTACCGGTTT	239	3/59/H
exon 4	GAGGATATCCACCACTGACAGA	CACCCATATACATGCTCCCC	228	3/59/H
exon 5	TGTATTTACATAGCAACTCCTTCCTTT	GGGCAAAATTTTCAAGAAGC	311	3/59/H
exon 6-7	CAGTGAGCCAAGACCATGC	TTAGGATTTATGGTCAAAAGAGACTAT	713	3/59/H
exon 8	CTACCCAAACTTGCTCCCCT	CAAGGCAAGACAAAGAACTCCT	428	3/59/H
exon 9	GGATGCTCCTAATCCTGCAA	CCAATGTGTCTCAATCTATTCCA	211	3/59/H
exon 2F-exon 6-7R	TTGACCATGTTTCAAGTAGAACTTTT	TTAGGATTTATGGTCAAAAGAGACTAT	10,638 (Wild-Type) Approx 2,000 (Mutant)	3/56/H
intron 2-intron 6	TGTCCTCCTTGAAAGAAGTTGTC	CAGCATATGATGTGAGACTGCT	9,073 (Wild-Type) 399 (Mutant)	3/59/H
exon 3-exon 6 (deletion 5′ breakpoint)	CAAATGGTTCTTTCACTTCGTCA	CCTGCCCCACTAGCTCTCTA	225 (Wild-Type)	3/59/H

Primers sequences used for research presented in Chapter 4.

All amplicons refer to *AMTN*. Abbreviations are as follows: [Mg^{2+}] represents the Mg^{2+} concentration present in the PCR reaction (mM); A_T represents the annealing temperature used in ^{o}C. Add. details the use of additives to enhance the PCR. H Hot Shot Mastermix (Section 2.5.3).

Appendix 22: Clinical details of AI-154 unaffected individual

Deciduous dentition of family AI-154 unaffected individual IV:2.

Appendix 23: Whole exome sequencing alignment statistics for family AI-154 samples

Sample	Input DNA /μg	Capture reagent	Depth of coverage	% bases covered by >4 reads	% bases covered by >9 reads	% bases covered by >14 reads	% bases covered by >19 reads	% bases covered by >24 reads	% bases covered by >29 reads
II:2	3.0	SSV5	60.06	98.9	97.6	95.4	91.9	87.1	81.3
II:3	0.2	SSV5	89.77	99.6	99.0	98.1	96.6	94.5	91.8
III:3	3.0	SSV5	66.44	98.7	97.3	94.9	91.6	87.3	82.2
IV:1	0.2	SSV5	110.54	99.8	99.4	98.8	97.8	96.5	94.8

Details of DNA input, capture reagent and alignment statistics for WES of AI-154 DNA samples.

The standard input for the WES protocols used is either 3μg or 200ng. Alignment statistics were generated using the regions targetted by the capture reagent as the interval. Abbreviation used: SSV5 SureSelectXT Human All Exon V5.

Appendix 24: Family AI-154 SNVs and indels

Genomic variant (GRCh37)	dbSNP reference	Amino acid change	Gene	CADD v1.3	SIFT	Polyphen	RefSeq transcript
1:45471767G>C	N/A	p.R583G	*HECTD3*	32	Deleterious (0)	Probably damaging (0.982)	NM_024602
1:155028673C>T	N/A	p..R288C	*ADAM15*	32	Deleterious (0)	Probably damaging (1)	NM_207197
20:2796081T>G	rs117426258	p.L84R	*C20orf141*	28.5	Deleterious (0)	Possibly damaging (0.883)	NM_080739
10:129681992G>A	rs141771225	p.P126L	*CLRN3*	27.1	Deleterious (0)	Probably damaging (1)	NM_152311
3:49845339T>C	rs561051731	p.I849V	*UBA7*	26.4	Deleterious (0)	Probably damaging (0.999)	NM_003335
19:12429806C>A	N/A	p.G345C	*ZNF563*	26.0	Deleterious (0)	Probably damaging (0.999)	NM_145276
7:55266431T>C	N/A	p.M908T	*EGFR*	25.5	Deleterious (0.02)	Probably damaging (1)	NM_005228
1:159899551C>T	N/A	p.V735M	*IGSF9*	25.1	Deleterious (0)	Possibly damaging (0.632)	NM_001135050
18:59221687T>C	rs147507320	p.V722A	*CDH20*	24.9	Deleterious (0.03)	Probably damaging (0.996)	NM_031891
7:73038619C>G	N/A	p.Q68H	*MLXIPL*	23.9	Tolerated (0.37)	Possibly damaging (0.620)	NM_032951
1:167343355C>T	rs375705900	p.T138M	*POU2F1*	23.8	Tolerated (0.14)	Unknown (0)	NM_002697
1:35650194G>A	N/A	p.R663C	*SFPQ*	23.4	Tolerated (0.15)	Benign (0.217)	NM_005066
18:33785103CCT>C	N/A	p.?	*MOCOS*	23.2	N/A	N/A	NM_017947
20:3145149T>G	N/A	p.K612T	*LZTS3*	23.3	Tolerated (0.14)	Possibly damaging (0.859)	NM_001282533
20:3145150T>C	N/A	p.K612E	*LZTS3*	23.1	Tolerated (0.18)	Possibly damaging (0.801)	NM_001282533
1:190234086G>A	rs200575097	p.T176M	*BRINP3*	22.9	Deleterious 0.03	Benign (0.235)	NM_199051
6:37605158C>T	rs201115181	p.A952T	*MDGA1*	22.9	Tolerated low confidence (0.19)	Benign (0.040)	NM_153487
2:171641303C>T	rs551957588	p.P59L	*ERICH2*	22.8	Deleterious low confidence (0.01)	Benign (0.269)	NM_001289947
3:75788040G>A	rs111880168	p.A195V	*ZNF717*	22.6	Tolerated (0.06)	Possibly damaging (0.739)	NM_0012890209
17:48195540 ATCCGAGCCTTCACCATCC>A	N/A	p.E59_S64del	*SAMD14*	22.1	N/A	N/A	NM_174920
17:48687281G>A	N/A	p.A1582T	*CACNA1G*	22.0	Tolerated (0.41)	Benign (0.027)	NM_018896
11:85435159C>T	N/A	p.E1305K	*SYTL2*	21.6	N/A	Benign (0.005)	NM_001162953

Genomic variant (GRCh37)	dbSNP reference	Amino acid change	Gene	CADD v1.3	SIFT	Polyphen	RefSeq transcript
20:23472425T>C	N/A	p.S41P	*CST8*	21.5	Deleterious (0)	Probably damaging (0.991)	NM_005492
1:115226899C>A	rs139582106	p.Q189H	*AMPD1*	18.62	Tolerated (0.29)	Probably damaging (0.928)	NM_000036
6:80341219T>A	N/A	p.V14E	*SH3BGRL2*	17.56	Tolerated (0.38)	Benign (0.112)	NM_031469
11:70172724A>C	rs534856862	p.S224R	*PPFIA1*	15.91	Tolerated (0.34)	Benign (0.035)	NM_003626

Details of the twenty-six genomic variants identified by WES after filtering.

Variants shared by II:2, III:3 and IV:1, but not present in II:3 were selected and these remaining variants were filtered by removing those with a MAF of 0.1% or more in dbSNP142, EVS or ExAC, that are homozygous and those located on the X chromosome. The list is restricted to variants scoring 15 or more when scored with CADD v1.3.

Appendix 25: AI-154 copy number variants

Gene	CNV type	Span (GRCh37)	No. of affected; total; protein coding exons[a]	Bayes Factor		Reads ratio	
				III:3	IV:1	III:3	IV:1
AMTN[b]	Deletion	4: 71388473-71394475	4; 9; 8	13.6	24.4	0.538	0.519
NAT8[c]	Deletion	2: 73868073-73868755	1; 2; 1	4.97	0.59	16	0.538

Details of the CNVs identified by ExomeDepth from AI-154 WES data that remained after filtering.

CNVs were retained only if present in III:3 and IV:1 and not present in II:3. Any CNV with a boundary that overlapped in III:3 and IV:1 was retained. For a CNV to be excluded, the boundary specified in II:3 had to include the entire region present in both III:3 and IV:1. CNV data from individual II:2 was not included since the number of control samples available for ExomeDepth analysis was insufficient.

[a] Column details the number of exons deleted by each CNV followed by the total number of exons in the transcript specified and the number of protein coding exons.

[b] *AMTN* Refseq transcript NM_212557.

[c] *NAT8* Refseq transcript NM_003960.

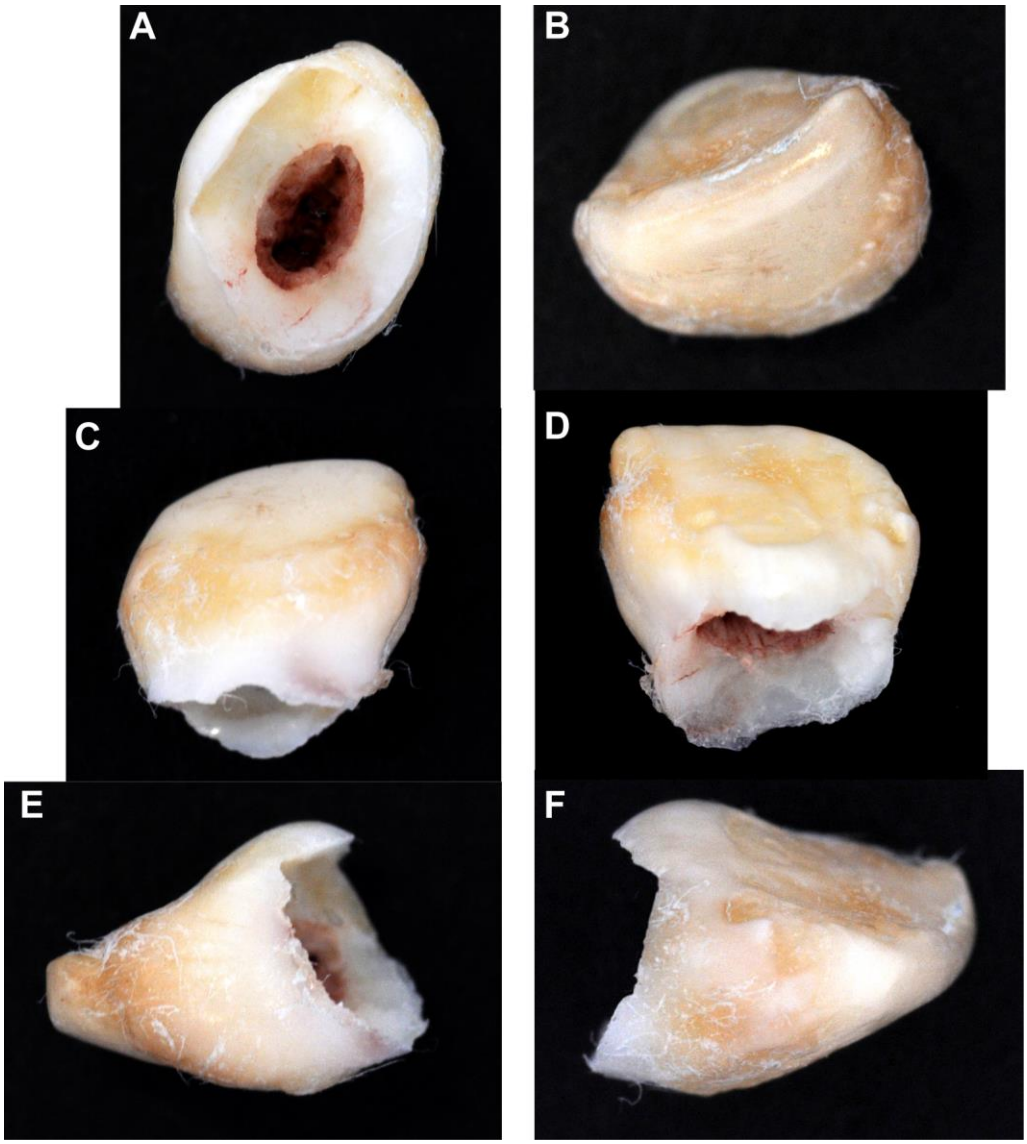
Appendix 26: Additional families screened for *AMTN* variants

Family	Country recruited	Inheritance	Phenotype	Ethnic origin	Subsequent findings
AI-26	UK	AD	Pitted with hypoplasia	Caucasian	
AI-29	UK	AD	Hypomaturation with some hypoplasia	Caucasian	
AI-32	UK	AD	Hypomaturation with some hypoplasia	Caucasian	
AI-34	UK	AD	Crowned (hypomaturation?)	Caucasian	*FAM83H* het. c.1354C>T, p.(Q452*)
AI-57	UK	Probable AD	Hypomineralised	Pakistani	
AI-73	UK	AD	Pitted with hypoplasia	Caucasian	
AI-75	Pakistan	AD	Hypomaturation / hypomineralisation	Pakistani	
AI-83	Pakistan	N/K	Hypomaturation	Pakistani	
AI-115	UK	AD	N/K	Caucasian	
AI-153	UK	AD	Hypomaturation	Caucasian	WES - no clear candidate - ongoing
AI-157	UK	AD	Hypomaturation	Caucasian	WES - no clear candidate - ongoing
AI-165	UK	AD	Hypomineralised (hypocalcified)	Caucasian	*FAM83H* het. c.2028C>T, p.(Q677*)
AI-168	UK	AD	N/K	N/K	
AI-192	UK	Probable AD	Hypocalcified	Pakistani	
AI-197	UK	AD	N/K	Caucasian	
AI-210	UK	Probable AD	Hypomineralised	Caucasian	
AI-211	UK	AD	Hypomineralised with some hypoplasia	N/K	
AI-212	UK	AD	Hypoplastic and hypomineralised	N/K	*ENAM* het. c.92T>G, p.(L31R)
AI-213	UK	AD	Hypoplastic	Caucasian	
AI-214	UK	AD	Hypoplastic	Caucasian	*ENAM* het. c.92T>G, p.(L31R)
AI-215	UK	Probable AD	Hypoplastic, pitted	Caucasian	Actually AR consanguineous: *FAM20C* hom. c.1376G>A, p.R459H
AI-236	UK	AD	Hypoplastic and hypomineralised	Caucasian	
AI-241	UK	AD	Hypoplastic and hypomineralised	Caucasian	
AI-244	UK	AD	Hypoplastic	Caucasian	
AI-246	Costa Rica	AD	N/K	N/K	*FAM83H* het. c.1363C>T, p.(Q455*)
AI-251	UK	AD	N/K	N/K	*FAM83H* het. c.1282C>T, p.(Q428*)
AI-253	UK	AD	N/K	Asian	
AI-254	UK	AD	N/K	N/K	
AI-255	UK	AD	N/K	N/K	
AI-271	Slovenia	AD	Hypomaturation	Caucasian	WES - no clear candidate - ongoing
AI-272	UK	AD	N/K	N/K	Dentinogenesis imperfecta - excluded from study
AI-273	UK	AD	Hypoplastic	Caucasian	
AI-274	UK	AD	Hypoplastic	Caucasian	*AMELX* het./hemi. c.100G>A, p.(E34K)
AI-275	UK	AD	Hypoplastic	Caucasian	*ENAM* het. c.92T>G, p.(L31R)
AI-276	UK	AD	Hypomaturation	Caucasian	

Details of the families screened for *AMTN* mutations.

Abbreviations used: AD autosomal dominant; AR autosomal recessive; hemi. hemizygous; het. heterozygous; hom. homozygous; N/K not known; WES whole exome sequencing.

Appendix 27: Photographs of tooth D from individual IV:1



Photographs of exfoliated deciduous lateral incisor tooth D from individual IV:1 (family AI-154).

A Cervical view.

B Coronal view.

C Labial view.

D Lingual view.

E and F Mesial and distal views (exact tooth position within the mouth unknown).

Appendix 28: Relative positions of the signal peptide and the IPV-like motif in selected secreted proteins

Protein	Group	Amino acid sequence	Reference sequence	Location
AMELX	Enamel	MGTWILFACLLGAAFAMP<>LPPHP	NP_001133 / NM_001142.2	Xp22.2
AMBN	Enamel	MSASKIPLFKMKDLILILCLLEMSFAVP<>FFPQQ	NP_057603 / NM_016515.5	4q13.3
AMTN	Enamel	MRSTILLFCLLGSTRSLP<>QLKPA	NP_997722 / NM_212557.3	4q13.3
ENAM	Enamel	MLVLRCRLGTSFPKLDNLVPKGKMKILLVFLGLLGNSVAMP<>MHMPR	NP_114095 / NM_031889.2	4q13.3
ODAM	Enamel	MKIIILLGFLGATLSAP<>LIPQR	NP_060325 / NM_017855.3	4q13.3
DMP1	Dentine / Bone	MKISILLMFLWGLSCALP<>VTRYQ	NP_004398 / NM_004407.3	4q13.3
DSPP	Dentine / Bone	MKIITYFCIWAVAWAIP<>VPQSK	NP_055023 / NM_014208.3	4q22.1
SPP1	Dentine / Bone	MRIAVICFCLLGITCAIP<>VKQAD	NP_000573 / NM_000582.2	4q22.1
MEPE	Dentine / Bone	MRVFCVGLLLFSVTWAAP<>TFQPQ	NP_064588 / NM_020203.3	4q22.1
BGN	Dentine / Bone	MWPLWRLVSLLALSQALPFEQRG	NP_001702 / NM_001711.5	Xq28
DCN	Dentine / Bone	MKATIILLLLAQVSWAGPFQQRG	NP_001911 / NM_001920.4	12q21.33
SPARCL1	Other*	MKTGLFFLCLLGTAAAIP<>TNARL	NP_004675 / NM_004684.5	4q22.1
SPARC	Dentine / Bone*	MRAWIFFLLCLAGRALAAP<>QQEAL	NP_003109 / NM_003118.3	5q33.1
C3	Blood protein	MGPTSGPSLLLLLLTHLPLALGSP<>MYSII	NP_000055 / NM_000064.3	19p13.3
AHSG	Blood protein	MKSLVLLLCLAQLWGCHSAPHGPGL	NP_001613 / NM_001622.2	3q27.3

Protein	Group	Amino acid sequence	Reference sequence	Location
SERPINC1	Blood protein	MYSNVIGTVTSGKRKVYLLSLLLIGFWDCVTCHGSPVDICT	NP_000479 / NM_000488.3	1q25.1
PRL	Pro-hormone	MNIKGSPWK<>GSLLLLLVSNLLLCQSVAPLPICPGG	NP_000939 / NM_000948.5	6p22.3
LEP	Pro-hormone	MHWGTLCGFLWLWPYLFYVQAVPIQKVQ	NP_000221 / NM_000230.2	7q32.1
GH1	Pro-hormone	MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLS	NP_000506 / NM_000515.4	17q23.3
CHGA	Neuroendocrine secretion-associated protein	MRSAAVLALLLCAGQ<>VTALPVNSPM	NP_001266 / NM_001275.3	14q32.12
CHGB	Neuroendocrine secretion-associated protein	MQPTLLLSLLGAVGLA<>AVNSMPVDNRN	NP_001810 / NM_001819.2	20p12.3
FGF23	Neuroendocrine secretion-associated protein	MLGARLRLWVCALCSVCSMSVLRAYPNASPL	NP_065689 / NM_020638.2	12p13.32

Positions of the signal peptide relative to the IPV-like motif for selected proteins of the enamel matrix, dentine matrix and other secreted proteins.

The sequence shown is limited to that spanning from the N-terminus to 4 amino acids beyond the IPV-like motif.

The signal peptide is highlighted yellow, the IPV-like motif is underlined and the positions of exon boundaries are denoted by chevrons.

The signal peptide was identified using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/) and IPV-like motifs were identified from previous reports by von Marschall et al., 2012, Nam et al., 2014 and Liang et al., 2015.

Note that for ENAM, a signal peptide was not detected, but a peak, indicating a predicted cleavage site, was present at 39-40, although this peak did not reach significance. The inferred position of the signal peptide agreed with the considerable literature on the murine and porcine ENAM proteins.

The reference sequences listed include both the amino acid sequence used to identify the signal peptide and the transcript used to identify the positions of exon boundaries relative to the amino acid sequence.

SPARCL1 and SPARC are included here since both the enamel and dentine / bone matrix genes arose from these genes (Kawasaki & Weiss, 2003).

Appendix 29: Clustal Omega multiple sequence alignment of homologous protein sequences for AMTN

```
CLUSTAL O(1.2.1) multiple sequence alignment

Guinea pig              1      MKTVILLLGLLGWAQSLPKQLNSALEVSPTKPAPDQVTPLSQQQPTQVFPSLSLISLTQL 60
Mouse                   1      MKTMILLLCLLGSAQSLPKQLNPASGVPATKPTPGQVTPLPQQQPNQVFPSISLIPLTQL 60
Rat                     1      MKTVVLLLCLLGSAQSLPRQLSPALGAPATKPTPGQVTPLTQQQPNQVFPSISLIPLTQL 60
Elephant                1      MKTTILLFCLLGSTQSLP-QLNPALGLSAAKLVPDQATLLNQQQPNQVFPSLSLIPLTQL 59
Rhesus macaque          1      MKTMILLFCLLGSTQSLP--LKPALGLPPTKLAPDQGTLPNQQQSNQVFPSLSLIPLTQM 58
Gorilla                 1      MRSTILLFCLLGSTRSLP-QLKPALGLPPTKLAPDQGTLPNQQQSNQVFPSLSLIPLTQM 59
Human                   1      MRSTILLFCLLGSTRSLP-QLKPALGLPPTKLAPDQGTLPNQQQSNQVFPSLSLIPLTQM 59
Chimpanzee              1      MRSTILLFCLLGSTRSLP-QLKPALGLPPTKPAPDQETLPNQQQSNQVFPSLSLIPLTQM 59
Horse                   1      MKTVILLFCLLGSTLSLPMQF-PALGLPPTKPVPDQATQLNQQQPNQVFPSLNLIPLTQM 59
Cat                     1      MKTMILLLYLLGSTQSLPTQLNPALGLP-TKLALDQATLLNQQQLNQVFPSLSLIPLTQM 59
Dog                     1      MKTMILLLYLLGSTQSLPAQLNPALGLPPTKLTPHQATLLNQQQPNQVFPSLSLIPLTQM 60
Wild boar               1      MKTTILLFCLLGSTLSLPMQFNPVLGLPPTKLVPDQATLRNQQQPNQVFPSLSLIPLTQM 60
Sheep                   1      MKAAILLFCLLGSTLSLPMQLNPALVLPPTKLVPDQATLLNPQQPSQVFSSLSLIPLTHM 60
Cow                     1      MKAAILLFCLLGSTLSLPMQLNPALVLPPTKLVPDQATLLNPQQPNQVFSSLSLIPLTHM 60
                               *:: :**: *** : ***  :  .     :* .  * *    ** .*** *:.** **::

Guinea pig             61      LTLASDLQLLTPASGTKPGAQTLQLSLGALQGQNQLQPQLLPIFVAQLGAPAAILSSEEL 120
Mouse                  61      LTLGSDLPLFNPAAG-PHGAHTLPFTLGPLNGQQQLQPQMLPIIVAQLGAQGALLSSEEL 119
Rat                    61      LTLGSDLPLFNPATM-PHGTQTLPFTLGPLNGQQQLQPQMLPIIVAQLGAQGALLSSEEL 119
Elephant               60      LTLGSDLQQLNPATGMAPGTQTFPLNLGRLTIKQQLQSQLLPIIVAQLGAQGAILSSEEL 119
Rhesus macaque         59      LTLGPDLHLLNPAAGMTPGTQAHPLTLGVLNLPQQLQPHMLPIFVTQLGAQGTILSSEEL 118
Gorilla                60      LTLGPDLHLLNPAAGMTPGTQTHPLTLGGLNVQQQLHPHMLPIFVTQLGAQGTILSSEEL 119
Human                  60      LTLGPDLHLLNPAAGMTPGTQTHPLTLGGLNVQQQLHPHVLPIFVTQLGAQGTILSSEEL 119
Chimpanzee             60      LTLGPDLHLLNPAAGMTPGTQTHPLTLGGLNVQQQLHPHVLPIFVTQLGAQGTILSSEEL 119
Horse                  60      LKLGSDLQLLNPAAGVAPGAQTLPLTLGGLKAQQQLQSQMLPIFVAQFGAQGTILSSEEL 119
Cat                    60      ITLGADLQLLNPPAGLAPGTQTLPLTLGGLNTQQPLQAQMLPVIVAHLGAQGTILSSEEL 119
Dog                    61      FTLASDLQLLNPAAGMASGTQTLPLSLGVLNTQQQLQPQMLPVIVAHLGAHGAILSSEEL 120
Wild boar              61      LTLGSDLQLLNPAIGMVPSSQTLPLTLGALNVQQQLQPQMIPVIVAHLGAQGAILSSEEL 120
Sheep                  61      LTLGSDLQLLNPAAGMPSGTQTLPLALGGLKVQQQLQPQMLPVIVAHFGAQGTILSSEEL 120
Cow                    61      LTLGSNLQLLNPAAGMPSGTQTLPLALGGLKVQQQLQPQMLPVIVAHFGAQGTILSSEEL 120
                               :.*. :*  :.*      .:::  : ** *   : *: :::*::*:::** .::******
```

```
Guinea pig            121    PLAPQIFTGLLLNPLFPEGTLPTSQAGASPNIQEEAIPAGQTGVNTNMQVTTEGQLSTPG 180
Mouse                 120    PLASQIFTGLLIHPLFPGAIPPSGQAGTKPDVQNGVLPTRQAGAKAVNQGTTPGHVTTPG 179
Rat                   120    PLASQIFTGLLIHPLFPGAIQPSGQTGAKPDVQNGALPTRQAGASPANQATTPGHT-TPA 178
Elephant              120    PMAPQIFAGLLIQPLFPGAILPTSLAGATPEVQEGILPAGQAGLNPAIQRTPEKHPSTSS 179
Rhesus macaque        119    P---QIFTSLIIHSLFPGGILPTSQAGANPDVQDGSLPAGQAGVNPAIQGTPAGRLPTPS 175
Gorilla               120    P---QIFTSLIIHSLFPGGILPTSQAGANPDVQDGSLPAGGAGVNPAIQGTPAGRLPTPS 176
Human                 120    P---QIFTSLIIHSLFPGGILPTSQAGANPDVQDGSLPAGGAGVNPATQGTPAGRLPTPS 176
Chimpanzee            120    P---QIFTSLIIHSLFPGGILPTSQAGANPDVQDGSLPAGGAGVNPATQGTPAGRLPTPS 176
Horse                 120    PAAPQIFAGLIFQPLFPGGILPTSQ--ATPDVQNGILPAGQGGVKPAIQGTSESPLPTTS 177
Cat                   120    PMAPQIFTGLIFQPLFPGSTLPNSQ--ANPDAQNGILPAGQAGMNPAIQGTSEGFSPTPS 177
Dog                   121    PGSPQIFTGLIFQPLFPGAILPTSP--ANPDAQNGILPAGQAGGNPAIQGTPESFSTTPS 178
Wild boar             121    PATRQILTGLIFHTLFPGAILPPSP--AKPDAQNGIHPAGQAGANPAVQGTPRGPFPTSS 178
Sheep                 121    QGTSQILTGLIFHPLFPGAILPTSQ--ANPDAQNGILPAGQAGANPATQGTPEDPFSTPS 178
Cow                   121    QGTSQILTGLIFHPLFPGAILPTSQ--ANPDAQNGILPAGQAGANPAAQGTPEDPFSTPS 178
                                 **::.*::. *** .  * .   :.*: *:   *:   * .   * *      * .

Guinea pig            181    VT-DDDFEVTTTLGIQKVTHTPEKTTTGSPNCEFL------------------------ 214
Mouse                 180    VTDDDDYEMSTPAGLRRATHTTEGTTIDPPNRTQ------------------------- 213
Rat                   179    VTDDDDYEMSTPAGLQRATHTTEGTTMDPPNRTK------------------------- 212
Elephant              180    DT-DSVFGVTTPAGLQRGMRTTGETTTESPNDEPQNLLYVVLQPDTSNR----LELAQ-- 232
Rhesus macaque        176    GT-DDDFAVTTPAGIQRSTHATEETTTE------------------------------- 202
Gorilla               177    GT-DDDSAVTTPAGIQRSTRAIEETTIESANGIQ------------------------- 209
Human                 177    GT-DDDFAVTTPAGIQRSTHAIEEATTESANGIQ------------------------- 209
Chimpanzee            177    GT-DDDFAVTTPADIQRSTRAIEEATTESANGIQ------------------------- 209
Horse                 178    DT-DDDFGGTTPAGIQRGMRTTEETITKSPNEFSTDQLLALNFRDLFSRPLPRTSATKHS 236
Cat                   178    DT-DDDFEVTAPTGIRRGMHTTQETTTGPPNGNQ------------------------ 210
Dog                   179    DT-DDDFGVTAPAGIQRGTHTTQETTSGPPNGKFSEPGKCC-------RSHRRVGTG--- 227
Wild boar             179    GT-DDDFDVTTPAGLQRGTHATEETTTGSPNGMQ------------------------ 211
Sheep                 179    GT-DDDFAATTPAGIQRGRQTTEEAPTGSPKGRFSKPGRHP-------RAMRVICGVREK 230
Cow                   179    GT-DDDFASTTPAGIQRGRPTTEETPTGSPKGIQ------------------------ 211
                              * *.    ::   :::   :   :
```

```
Guinea pig          214    ----------------------------------   214
Mouse               213    ----------------------------------   213
Rat                 212    ----------------------------------   212
Elephant            232    ----------------------------------   232
Rhesus macaque      202    ----------------------------------   202
Gorilla             209    ----------------------------------   209
Human               209    ----------------------------------   209
Chimpanzee          209    ----------------------------------   209
Horse               237    GLVHIAMPPLTFHFGVPTLDALRQCRGRSIPESHF   271
Cat                 210    ----------------------------------   210
Dog                 228    -F-LV-------------LP---------------   232
Wild boar           211    ----------------------------------   211
Sheep               231    GSLSC-------------LDKYRS-----------   241
Cow                 211    ----------------------------------   211
```

Clustal Omega multiple sequence alignment of homologous protein sequences for AMTN.

The red boxed region indicates the position of the signal peptide. Yellow shaded residues indicate those that span the IPV-like motif. The green boxed region indicates the region of the human protein deleted (if translated) by deletion exons 3-6. Green shaded residues indicate the IPLT motif that is predicted to be an O-glycosylation site. Magenta shaded residues indicate the SXE motif. An asterisk (*) indicates positions which have a single, fully conserved residue. A colon (:) indicates conservation between groups of strongly similar properties. A full stop (.) indicates conservation between groups of weakly similar properties.

AMTN sequences used: Guinea pig, *Cavia porcellus* XP_003467586.2; Mouse, *Mus musculus* NP_082069.1; Rat, *Rattus norvegicus* NP_001037761.1; Elephant, *Loxodonta africana* XP_010592490.1; Rhesus macaque, *Macaca mulatta* XP_014994036.1; Gorilla, *Gorilla gorilla* XP_004038827.1; Human, *Homo sapiens* NP_997722.1; Chimpanzee, *Pan troglodytes* XP_001160426.1; Horse, *Equus caballus* XP_014594023.1; Cat, *Felis Catus* XP_00395385.2; Dog, *Canis lupus familiaris* XP_854599.1; Wild boar, *Sus scrofa* XP_01384473.1; Sheep, *Ovis aries* XP_012035497.1; Cow, *Bos taurus* NP_001289882.1.