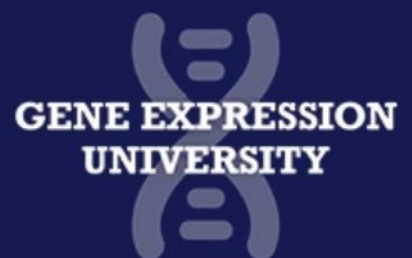




December 1st, 12pm EST

**THE ULTIMATE
VIRTUAL SEMINAR
SERIES IN GENE
EXPRESSION STUDIES**



[Register Now>>](#)

**applied
biosystems**
by Thermo Fisher Scientific

ThermoFisher
SCIENTIFIC

WILEY

Spectrum of pathogenic variants and founder effects in amelogenesis imperfecta associated with *MMP20*

Georgios Nikolopoulos^{1,2}  | Claire E. L. Smith¹  | James A. Poulter¹  |
Gina Murillo³  | Sandra Silva⁴ | Teresa Lamb⁵ | Ian R. Berry⁶ |
Catriona J. Brown⁷ | Peter F. Day^{8,9}  | Francesca Soldani⁹ |
Suhaila Al-Bahlani¹⁰ | Sarah A. Harris^{11,12}  | Mary J. O'Connell^{13,14}  |
Chris F. Inglehearn¹  | Alan J. Mighell² 

¹Division of Molecular Medicine, Leeds Institute of Medical Research, University of Leeds, Leeds, UK

²Department of Oral Biology, School of Dentistry, St James's University Hospital, University of Leeds, Leeds, UK

³School of Dentistry, Universidad de Costa Rica, Ciudad Universitaria Rodrigo Facio, San Pedro Montes De Oca, Costa Rica

⁴Cellular and Molecular Biology Centre (CBCM), Universidad de Costa Rica, Ciudad Universitaria Rodrigo Facio, San Pedro Montes De Oca, Costa Rica

⁵Oxford University Hospitals NHS Foundation Trust, Oxford, UK

⁶Leeds Genetics Laboratory, St James's University Hospital, Leeds, UK

⁷Birmingham Dental Hospital, Birmingham, UK

⁸Department of Paediatric Dentistry, Leeds Dental Institute, University of Leeds, Leeds, UK

⁹Community Dental Service, Horton Park Health Centre, Bradford District Care NHS Foundation Trust, Bradford, UK

¹⁰Dental & O.M.F.S Clinic, Al Nahdha Hospital, Ministry of Health, Muscat, Oman

¹¹School of Physics, University of Leeds, Leeds, UK

¹²Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK

¹³School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK

¹⁴School of Life Sciences, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, UK

Correspondence

Alan J. Mighell, Department of Oral Biology,
School of Dentistry, James's University
Hospital, University of Leeds, Worsley Bldg,
Leeds LS2 9LU, UK.

Email: A.J.Mighell@leeds.ac.uk

Abstract

Amelogenesis imperfecta (AI) describes a heterogeneous group of developmental enamel defects that typically have Mendelian inheritance. Exome sequencing of 10 families with recessive hypomaturation AI revealed four novel and one known variants in the matrix metalloproteinase 20 (*MMP20*) gene that were predicted to be pathogenic. *MMP20* encodes a protease that cleaves the developing extracellular enamel matrix and is necessary for normal enamel crystal growth during amelogenesis. New homozygous missense changes were shared between four families of Pakistani heritage (c.625G>C; p.(Glu209Gln)) and two of Omani origin (c.710C>A; p.(Ser237Tyr)). In two families of UK

Georgios Nikolopoulos and Claire E. L. Smith contributed equally to this study.

Chris F. Inglehearn and Alan J. Mighell are joint senior authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Human Mutation* published by Wiley Periodicals LLC.

origin and one from Costa Rica, affected individuals were homozygous for the previously reported c.954-2A>T; p.(Ile319Phefs*19) variant. For each of these variants, microsatellite haplotypes appeared to exclude a recent founder effect, but elements of haplotype were conserved, suggesting more distant founding ancestors. New compound heterozygous changes were identified in one family of the European heritage: c.809_811+12delinsCCAG; p.(?) and c.1122A>C; p.(Gln374His). This report further elucidates the mutation spectrum of *MMP20* and the probable impact on protein function, confirms a consistent hypomaturation phenotype and shows that mutations in *MMP20* are a common cause of autosomal recessive AI in some communities.

KEYWORDS

amelogenesis imperfecta, enamel, founder effect, hypomaturation AI, *MMP20*

1 | INTRODUCTION

The hardness of dental enamel is remarkable and is a result of its high mineral and low protein content (C. E. Smith, 1998). The process of enamel formation, amelogenesis, begins with the production of a secreted soft enamel protein matrix, which is then transformed into a mature, organized structure of rod and interrod enamel crystallites almost devoid of protein. Crucial to this transformation process is the secretory stage enamel protease, matrix metalloproteinase 20 (*MMP20*) [NM_004771.4, ENST00000260228.2] (MIM #604629).

MMP20 is a zinc-dependent endopeptidase that is secreted in trace amounts during the secretory and transition stages of amelogenesis by ameloblasts (Llano et al., 1997; Seymen et al., 2015). Once activated, it selectively cleaves the secreted enamel proteins: amelogenin, enamelin, and ameloblastin into several products with distinct functional roles (Simmer & Hu, 2002). *MMP20* is also thought to facilitate ameloblast movement during secretion, through cell–cell communication, and may influence ameloblast gene expression (Guan & Bartlett, 2013). These activities create a newly voided space within which enamel crystallites are able to grow in width and thickness, and finally, to interlock (Bartlett, 2013). If they are not removed, the enamel proteins occupy the enamel volume and restrict the growth of the enamel crystallites. This produces immature enamel that fails prematurely due to its inability to resist the mechanical stress resulting from biting and chewing.

Mutations in *MMP20* and 19 other genes have been shown to cause nonsyndromic amelogenesis imperfecta (AI) with autosomal dominant, autosomal recessive, or X-linked inheritance (J. W. Kim et al., 2019; C. E. L. Smith et al., 2017, 2020; and <http://dna2.leeds.ac.uk/LOVD/genes>), while perhaps as many again have been implicated in syndromic AI (Dubail et al., 2018; Wright et al., 2015). Some of these genes encode proteins processed by *MMP20*, such as the enamel matrix proteins AMELX (MIM #300391), AMBN (MIM #610259), and ENAM (MIM #606585). Other genes implicated in AI include the second enamel matrix protease, KLK4 (MIM #603767), and cell adhesion proteins, such as LAMA3 (MIM #600805), LAMB3 (MIM #150310), and COL17A1 (MIM #113811). AI, therefore, describes a heterogeneous group of conditions

characterized by inherited enamel defects in both dentitions. It can present as a hypoplastic phenotype, where a deficit in secretion results in the formation of a reduced volume of mineralized enamel, or as hypomineralized AI, whereby a failure in maturation results in the enamel of full thickness but which is soft or brittle and fails prematurely (Gadhia et al., 2012). Hypomineralized AI can be further subclassified as hypomaturation, caused by incomplete removal of protein from the developing enamel, or hypocalcification, caused by insufficient transport of calcium ions into the forming enamel (Smith et al., 2017), though these phenotypes often overlap. Mutations in *MMP20* cause autosomal recessive hypomaturation AI. Fourteen pathogenic variants have been reported to date (Table S1), most of which are located in the catalytic domain of *MMP20* and are thought to affect the stability and functionality of the protein structure.

This study describes the identification of pathogenic *MMP20* variants in 10 families segregating autosomal recessive hypomaturation AI, providing additional insights into the spectrum of *MMP20* mutations and associated phenotype. Four novel variants are reported, two of which are relatively common founder mutations in specific populations. Additionally, we perform molecular dynamics (MD) simulations of variants in the catalytic domain of *MMP20* to examine the predicted effect that these changes have on the protein. This study increases the total reported pathogenic *MMP20* variants to 17, suggests that defects in *MMP20* are a more common cause of AI than was previously reported and enriches our understanding of the effects that mutations in the catalytic domain of *MMP20* can have on its functionality.

2 | MATERIALS AND METHODS

2.1 | Patients

Individuals from each of the 10 families were recruited following informed consent in accordance with the principles outlined in the Declaration of Helsinki, with local ethical approval (REC 13/YH/0028). These 10 families are part of a larger AI cohort presenting with a variety of AI phenotypes. A diagnosis of AI was made by a dentist after clinical

examination, based on the physical appearance of the dentition, and was confirmed by dental X-ray. Genomic DNA was obtained via venous blood samples, using conventional extraction techniques, or from saliva using Oragene® DNA Sample Collection kits (DNA Genotek) and extracted according to the manufacturer's instructions.

2.2 | Whole-exome sequencing (WES) and analysis

Three micrograms of genomic DNA from a single individual from each family (marked with an arrow on the pedigrees shown in

Figure 1) were subjected to WES and analyzed as described previously (C. E. L. Smith et al., 2019). Variants present in the dbSNP150 database of NCBI or the Genome Aggregation Database (gnomAD; <http://gnomad.broadinstitute.org>) with a minor allele frequency $\geq 1\%$ were excluded. The potential pathogenic effect of the variants was predicted using Combined Annotation-Dependent Depletion (CADD v1.3; <https://cadd.gs.washington.edu>) (Rentzsch et al., 2019), the Sorting Intolerant From Tolerant algorithm (SIFT; Sim et al., 2012), Protein Variation Effect Analyzer (PROVEAN; Choi et al., 2012), and MutPred2 (<http://mutpred.mutdb.org/index.html>). The potential effect on splicing for each

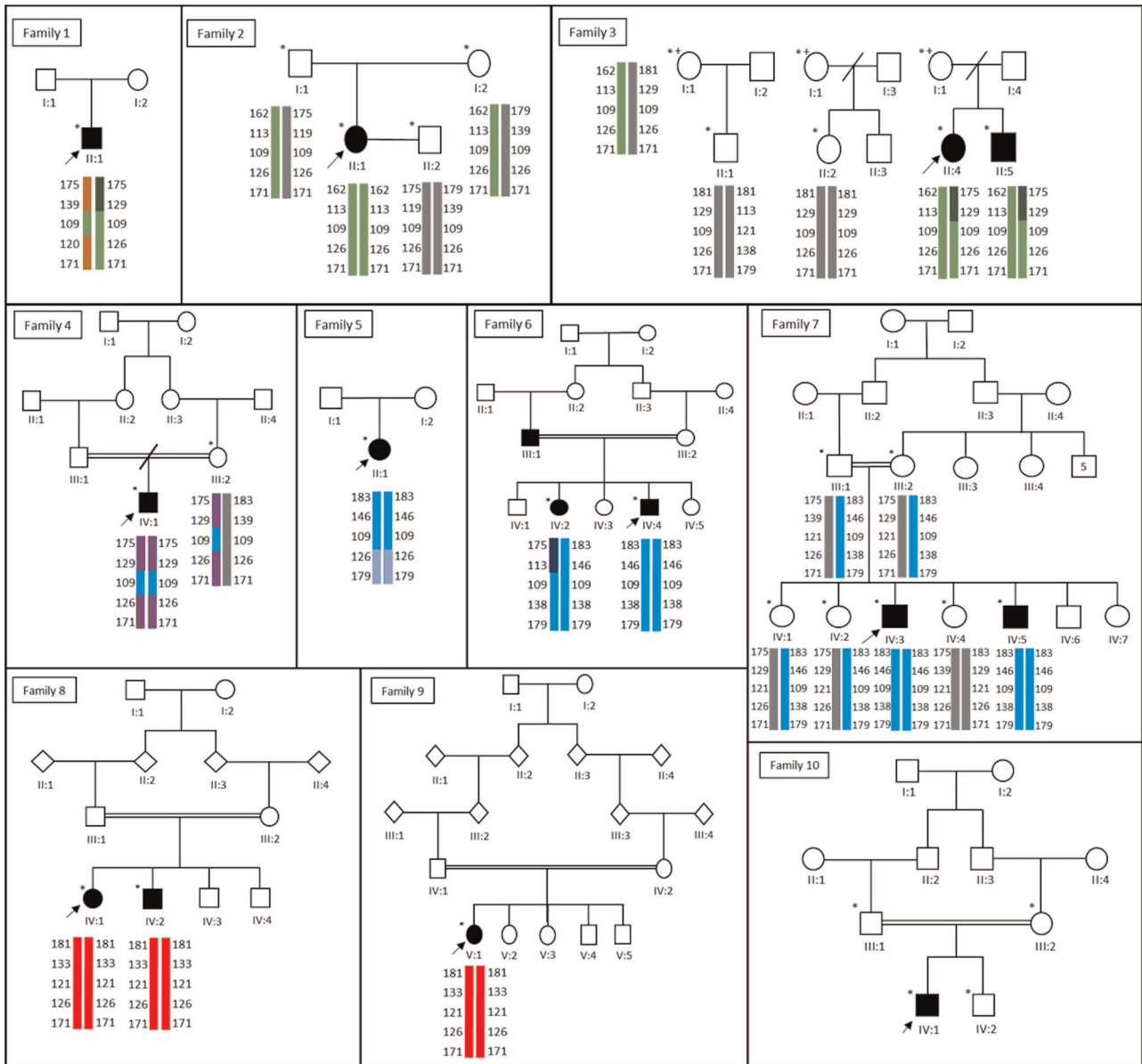


FIGURE 1 Pedigrees of the 10 families recruited for this study. The genotypes determined from the microsatellite markers are presented beneath each individual examined. Families 1, 2, and 3 carry the c.954-2A>T variant; Families 4, 5, 6, and 7 carry the c.625G>C variant; Families 8 and 9 carry the c.710C>A variant; and Family 10 has the c.809_811+12delinsCCAG and c.1122A>C variants. The common haplotypes are presented in the same color and the haplotypes without a pathogenic variant are colored gray. The markers are presented in the order: 11cen, D11S940, D11S1339, MMP20, D11S4108, D11S4159, D11S4161, 11qter. The recruited family members are marked with an asterisk and the proband in each family is indicated with a black arrow

intronic variant was predicted by the Human Splicing Finder v3.1 (<http://umd.be/HSF3/HSF.shtml>).

2.3 | Polymerase chain reaction (PCR) and Sanger sequencing

Mutations were confirmed and segregation was performed for all available family members, marked with (*) on each pedigree of Figure 1. Primer sequences can be found in Table S2. Sanger sequencing was performed using the BigDye Terminator v3.1 kit (Life Technologies) according to the manufacturer's instructions and resolved on an ABI3130xl sequencer (Life Technologies). Results were analyzed using SeqScape v2.5 (Life Technologies).

2.4 | Microsatellite analysis

Primer sequences for markers were obtained from the UCSC genome browser and standard HEX-tagged primers (Sigma-Aldrich; Table S2) were used to assess the flanking haplotypes of *MMP20*. The analysis was performed as described previously (Nikolopoulos et al., 2020).

2.5 | Protein structure analysis

The tertiary structure of *MMP20* has been determined in atomistic detail by nuclear magnetic resonance (NMR) (PDB: 2JSD) (Arendt et al., 2007), which provides the starting structure for MD simulations. Amino acid substitutions were made in the wild type (WT) structure using the Chimera visualization tool (Pettersen et al., 2004). MD simulations were performed using AMBER18 (Case et al., 2018). Protocols used to perform the MD simulations are in the Supporting Information Methods. Rhapsody was used for in silico saturation mutagenesis analysis (Ponzoni et al., 2020) and the prediction of the changes in solvent accessibility, residue occlusion, and free energy ($\Delta\Delta G$) of the mutated proteins was performed with Site-Directed Mutator (SDM) (Pandurangan et al., 2017).

3 | RESULTS

Ten unrelated families presenting with features consistent with autosomal recessive hypomaturation AI, in the absence of any cosegregating disease, were recruited for the study (Figures 1 and 2). Genomic DNA from affected members of each family was subjected to exome sequencing. Detailed coverage statistics can be found in



FIGURE 2 Clinical images for nine probands at different chronological ages, illustrating the spectrum of variations that can be observed within the overall, typical features of hypomaturation amelogenesis imperfecta that have been subjected to different post-eruptive environments. (a–c) Families 1–3, who all have European heritage and share the same homozygous *MMP20* c.954-2A>T; p.(Ile319Phefs*19) variant. (d–g) Families 4–7, who all have Pakistani heritage and share the same homozygous *MMP20* c.625G>C; p.(Glu209Gln) variant. (h) Family 9 who, along with Family 8 (no clinical images available), has Omani heritage, and both have the same homozygous *MMP20* c.710C>A; p.(Ser237Tyr) variant. (i) Family 10, of European heritage, which was found to segregate the compound heterozygous *MMP20* c.809_811+12delinsCCAG; p.(?) and c.1122A>C; p.(Gln374His) variants

Table S3. Samples were sequenced in different batches over a period of some years, with considerable variation in coverage between batches. Variant files were filtered to select rare variants with high predicted pathogenicity, then variants in known AI-causing genes were highlighted. This revealed biallelic mutations in *MMP20* (Table S4) in each family. The position of these variants in the gene is shown in Figure 3, along with representative electropherograms for each novel variant. PCR and Sanger sequencing confirmed the variants segregated with AI in all available family members.

In Families 1, 2, and 3, a homozygous frameshift variant (NM_004771: c.954-2A>T, NP_004762: p.(Ile319Phefs*19)) was identified in intron 6. This has been published previously as a cause of autosomal recessive hypomaturation AI and is expected to lead to retention of the sixth intron (J. W. Kim et al., 2005). To confirm this hypothesis, we attempted to perform reverse-transcriptase PCR of the *MMP20* transcript on control blood complementary DNA (cDNA). However, no amplification was achieved, suggesting that the level of *MMP20* expression in blood is below the threshold for detection.

A novel homozygous missense mutation, c.625G>C, p.(Glu209Gln), was identified in exon 4 as the cause of disease in Families 4, 5, 6, and 7. This is known variant rs199788797, which has not previously been associated with a disease phenotype. In the gnomAD database, this variant has a frequency of 0.000457 in the South Asian population but is absent from all other reported populations. E209 is fully conserved in the mammalian clade, as shown in Figure S1, using the sequences listed in Table S5. Additionally, the mutation is predicted to be damaging (Table S6).

In Families 8 and 9, a novel homozygous missense mutation, c.710C>A, p.(Ser237Tyr) was identified in exon 5. This variant is absent from databases, is evolutionarily conserved in all the mammalian species analyzed (Figure S1) and is predicted to be damaging by all algorithms used (Table S6).

The affected individual in Family 10 was found to be a compound heterozygote for a novel missense mutation c.1122A>C, p.(Gln374His) in exon 8 and a novel deletion–insertion (delins) variant: c.809_811+12delinsCCAG, p(?), spanning the splice donor site of intron 5. Both are absent from variant databases. Variant p.(Gln374His) is predicted to be damaging (Table S6), and Q374 is conserved in the mammalian clade (Figure S1). Interestingly, mouse *Mmp20* has a histidine in the equivalent position. This presumably does not have a detrimental effect on the function of the mouse protein, perhaps due to other changes of the mouse gene that balance the protein or reflecting differences in enamel development and function between mouse teeth, which grow throughout life, and human teeth which do not. Nevertheless, mutation prediction software does classify this variant as pathogenic; it is absent from the gnomAD database, and when paired with a second *MMP20* variant it has given rise to a hypomaturation phenotype consistent with biallelic *MMP20* disease. The delins variant is predicted by Human Splicing Finder to disrupt the intron 5 splice donor site, possibly leading to retention of the fifth intron (Figure S2).

The families that share variants (Families 1, 2, and 3 with c.954-2A>T, Families 4–7 with c.625G>C, and Families 8 and 9 with c.710C>A) also originate from the same ethnic backgrounds. To determine whether these families share common founder haplotypes at the *MMP20* locus, we genotyped five microsatellite markers across a 1.5 cM/3 Mb region of chromosome 11q22 in each family, in the order: 11cen, D11S940, D11S1339, *MMP20*, D11S4108, D11S4159, D11S4161, 11qter. The haplotype analysis is shown in Figure 1.

In Omani families 8 and 9, carrying the c.710C>A variant, haplotyping suggests they are closely related. However, the picture is more complex in the remaining two family groups. Families 1–3 (the first from Costa Rica, the remaining two from the United Kingdom) carry the previously published c.954-2A>T variant. The proband in

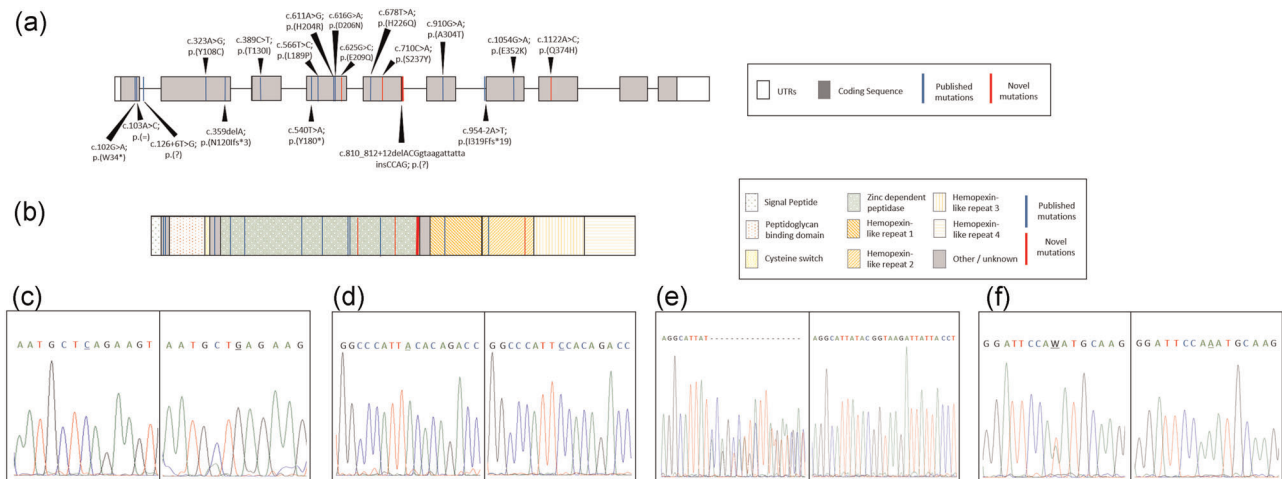


FIGURE 3 (a) *MMP20* gene diagram. The positions of all known mutations are marked on the gene with a vertical bar, while novel variants are highlighted in red. (b) Representation of the protein domains with all known mutations marked with vertical bars, as in (a). The zinc-dependent peptidase domain is the catalytic domain of the protein, where most pathogenic variants have been found. (c–f) A representative electropherogram of each of the novel variants found in this study, (c) c.625G>C variant and wild type (WT), (d) c.710C>A variant and WT, (e) c.809_811+12delinsCCAG variant and WT, (f) c.1122A>C variant and WT

Family 2 is homozygous for a haplotype also seen in Family 3, but the second haplotype in Family 3 has proximal recombination. Family 1 could share the recombinant haplotype observed in Family 3 together with an unrelated haplotype, but without phase information this cannot be confirmed. Interestingly, all three families are homozygous for D11S4108 100 kb from *MMP20*. Families 4–7, of UK–Pakistani origin, carry the c.625G>C variant. The proband in Family 6 is homozygous for the same haplotype segregating in Family 7. However, a second haplotype with distal recombination is seen in the second affected sibling in Family 6, suggesting the affected (unsampled) father carries both haplotypes. Family 5 is homozygous for a third haplotype, again identical at the proximal end to that in Family 7 but recombinant at the distal end. In contrast, Family 4 is homozygous for a fourth haplotype identical to the Family 7 haplotype at the distal end but proximally recombinant. Again, all four families are homozygous for the immediately adjacent marker D11S4108.

The catalytic center of *MMP20* is a 160-residue domain containing the zinc-dependent peptidase active site. It contains one catalytic and one structural zinc ion; the structures modeled here additionally contain two calcium ions, which are also structural. We assessed how missense variants affect the protein function by performing MD simulations of the relevant mutated proteins. In addition, to demonstrate the key role of metal ions in maintaining *MMP20* structure, we performed MD simulations in which the metal ions were removed. We compared the protein structures and dynamics observed for the WT with metal ions present, with those obtained for p.(Glu209Gln) and p.(Ser237Tyr) from this study (see Figure 4a) and p.(Thr130Ile) and p.(Leu189Pro) from the literature (Figure 4b); and also with results of the WT in the absence of metal ions. Figure 4c shows the root mean squared deviation (RMSD) of each of the simulations from their starting structures. The averages calculated for the last 300 ns of each repeat are in Table S7, and the repeats for the entire 900-ns simulation are shown separately in Figure S3. The calculated RMSD values are continuously adjusted and recalculated during the 900 ns of the simulation. As such, the final 300 ns, during which the plots have stabilized, are selected as the representative RMSD values for each simulation, and so the average values presented in Table S7 are used for the comparison among the variants. An increase in RMSD relative to the WT, implying decreased protein stability, was observed for all variants, apart from p.(Thr130Ile). We also analyzed the changes in three key interatomic interactions between the WT and the variants, atomic fluctuations (Figure S4a), hydrogen bonding interactions (Table S8), and salt bridges (Table S9), to provide insight into why these particular variants cause functionally deleterious changes in protein structure. The most significant structural distortions were observed in the simulations performed in the absence of structural zinc and calcium ions (see Figure S4b,c).

In silico saturation mutagenesis of the *MMP20* active site performed by Rhapsody shows that there are regions of the protein that are significantly more likely to cause a pathogenic effect when the residues located there are altered (Figure S5). These regions largely

correlate with the sites of the known and novel variants and have an increased PolyPhen-2 score (Figure S5). The Rhapsody analysis was limited to the catalytic domain of *MMP20* because it relies on the availability of a tertiary structure. The results of the SDM analysis are presented in Table S10, showing the changes of free energy, residue occlusion, and solvent accessibility for the WT and each mutant, respectively.

4 | DISCUSSION

Exome sequencing in a cohort of nonsyndromic AI families revealed biallelic *MMP20* variants in 10 autosomal recessive families. *MMP20* is a protease that plays an essential role during the secretory and early transition stages of amelogenesis. Defects in *MMP20* cause AI in both humans and mice, due to a failure to process, degrade, and remove proteins from the extracellular matrix scaffold upon which the developing dental enamel is formed. The hypomaturation phenotype observed in these families fits with this hypothesis, with the dental enamel being of normal volume but characterized by a loss of translucency, discoloration, and premature enamel loss. Post-eruptive changes probably determine the pattern of premature enamel failure. The phenotypes observed in the families presented are consistent with those described in previous reports of AI due to *MMP20* variants.

The variant identified in Families 1–3, (c.954-2A>T), has been reported previously (Gasse et al., 2017; J. W. Kim et al., 2005; Prasad et al., 2015; Wright et al., 2011), but it has not been possible to confirm the effect on splicing experimentally since *MMP20* is not expressed in blood. The prediction that the mutation most likely results in the loss of exon 7 from the *MMP20* transcript (J. W. Kim et al., 2005), therefore, remains speculative. If confirmed, this would break the reading frame and lead to an abnormal transcript, which would be expected to be subject to nonsense-mediated decay. *MMP20* cDNA cannot be obtained from the blood due to undetectable expression levels; however, it could possibly be obtained from tissues, such as tonsil or appendix, as expression has been detected there and these are routinely removed through routine surgery. However, altered splicing in these tissues would only suggest that splicing is altered in ameloblasts, instead CRISPR–Cas9 editing of this gene within ameloblast cell lines might be a better model to determine whether splicing is affected.

The missense variant identified in Families 4–7, p.(Glu209Gln), is novel and replaces a negatively charged glutamic acid residue with neutral glutamine. E209 is a coordinating ligand for calcium ion binding and is, therefore, critical to the function of *MMP20* (Andreini et al., 2004; Arendt et al., 2007; Yamakoshi et al., 2013). Replacement with glutamine decreases protein stability relative to WT in the MD simulations (see Table S7). Inspection of the WT atomistic structure shows that E209 has a pair of carboxylate oxygen atoms oriented toward the nearby Ca²⁺ ion. In E209, the average distance is <4 Å for both MD repeats (Table S9). However, the mutant Q209 only has one oxygen; in one of the two duplicate MD trajectories, the

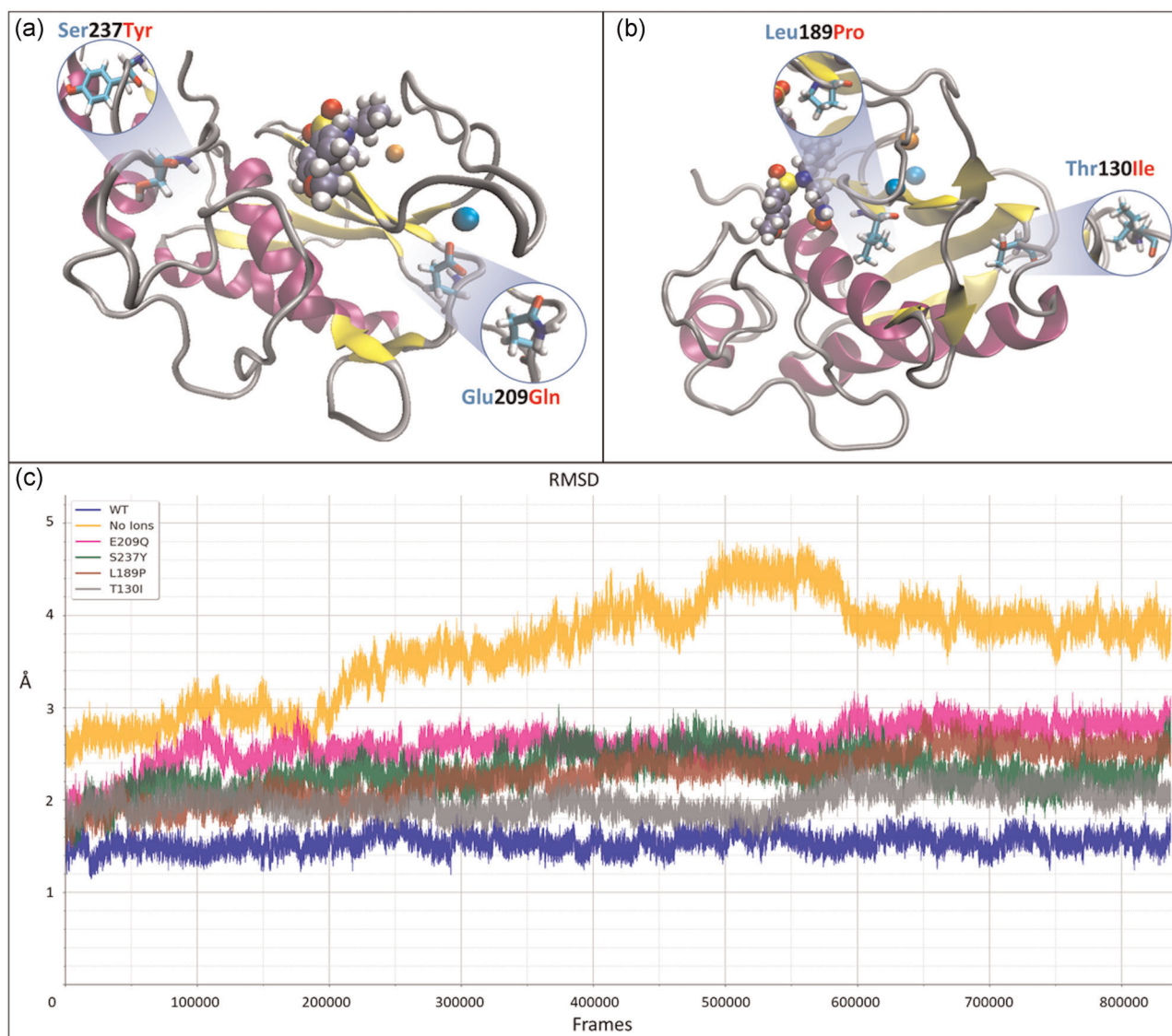


FIGURE 4 The tertiary structure of the catalytic domain of MMP20, based on the PDB: 2JSD nuclear magnetic resonance model. (a) The wild-type (WT) protein structure with the two novel variants, c.625G>C; p.(Glu209Gln) of Families 4–7 and c.710C>A; p.(Ser237Tyr) of Families 8 and 9, of the active site are presented in the inlays. (b) The WT protein structure with the inlays showing two of the known variants of the active site of MMP20, c.389C>T; p.(Thr130Ile), and c.566T>C; p.(Leu189Pro). (c) The root mean square deviation (RMSD) of each modeled variant during molecular dynamics simulations of 900 ns. An increase of RMSD value corresponds to a loss of stability, with the WT MMP20 structure being the most stable and the MMP20 structure in the absence of structural zinc and calcium ions being the least stable

interaction between this amide oxygen and Ca^{2+} remains strong (an interatomic distance of 2.5 Å) but in the other it is completely disrupted (interatomic distance of 6.9 Å), allowing it to form non-native hydrogen bonds with other residues (see Table S8).

The novel missense variant identified in Families 8 and 9 replaces a small serine residue with a larger aromatic tyrosine, p.(Ser237Tyr) in a highly conserved part of the catalytic domain of MMP20. Correspondingly, the SDM analysis identifies an increase in the folding free energy of the protein relative to the WT (Table S10), implying that the variant is less stable. Moreover, in the MD simulations, the larger and more hydrophobic Y237 results in an increased RMSD (2.35 Å) relative to the WT (1.55 Å). In these trajectories, the main backbone

hydrogen bond formed by S237 (with M244) is also formed by Y237. However, the number of side-chain hydrogen bonds is reduced in Y237 compared with S237 (see Table S8) because the corresponding tyrosine oxygen protrudes too far into the solvent to participate in interatomic interactions within the protein.

In Family 10, the delins variant disrupts the intron 5' splice donor site and is therefore again likely to lead to a transcript that is subject to nonsense-mediated decay. The novel c.1122A>C, p.(Gln374His) variant changes a hydrophilic glutamine residue for a hydrophobic histidine in the C-terminal hemopexin-like domain, as shown in Figure 3a, disrupting the folding of the protein and potentially leading to a loss of function.

Microsatellites were used to determine whether the variants shared in each of the three family groups were founder alleles, implying these families are related rather than that these sites are mutation hotspots. Family 1 is from Costa Rica, the population of which is largely of mixed European and indigenous American descent. Families 2 and 3 are Caucasian European families from the United Kingdom. The c.954-2A>T variant has been reported previously in at least six families from France (Gasse et al., 2017; Prasad et al., 2015) and one from the United States (J. W. Kim et al., 2005). Ethnicity is given as Caucasian by Prasad and co-workers but is not given in the remaining reports. Our haplotype analysis shows that this variant is present on three different chromosomal backgrounds, but elements of the haplotype, in particular the genotype of D11S4108 100 kb from *MMP20*, appear conserved between families. This suggests they may be related but through a distant common ancestor. The picture is similar for Families 4, 5, 6, and 7, UK families of Pakistani origin. In contrast, Families 8 and 9 from Oman share an identical haplotype across the region tested. These findings, therefore, suggest that *MMP20* c.710C>A represents a relatively recent founder mutation in the Omani population, while c.954-2A>T and c.625G>C result from older founder mutations in the Caucasian European and Pakistani populations.

In addition to the novel p.(Glu209Gln) and p.(Ser237Tyr) variants, we also performed MD for known variants p.(Leu189Pro) and p.(Thr130Ile) to see if the simulations could provide equivalent structural insight into the loss of function. Other variants within the active site of *MMP20* that change the ability of zinc ions to form coordination bonds were not modeled because this will either lead to substantial structural distortion (as seen with MD in the absence of calcium or zinc ions), or destabilize the key zinc metal center essential for catalysis.

For p.(Leu189Pro), while there is only a modest increase in RMSD relative to the WT, simulation of the behavior of the bound ligand in the WT and variant shows that replacing bulky lysine with more compact proline changes the shape of the active site. This leads to substantial structural rearrangement of the ligand, which is likely to affect substrate recognition and catalytic activity of the protease against natural substrates.

The only pathogenic missense variant modeled that did not show a significantly larger RMSD than the WT was p.(Thr130Ile). While minor changes in hydrogen bonding interactions were detected (Table S8), no substantial changes in overall protein structure or perturbations to interatomic interactions with the ligand or metal ions were observed. Consequently, our MD simulations do not provide a structural mechanism for why this variant should be pathogenic. However, p.(Thr130Ile) is located at the edge of the catalytic domain, close to regions of the protein that are absent from the available structure and which, therefore, do not feature in these simulations.

The 14 *MMP20* variants previously reported to cause AI are shown in Figure 3, together with those newly described herein, indicating their location in the gene and protein. Previously

reported pathogenic variants consisted of seven missense variants, two premature stop codons, two frameshifts, and two putative splice variants. Here, we report a further three missense and one splice variants, consistent with and extending the spectrum of mutations in *MMP20* causing autosomal recessive AI. The three splice variants have not been verified beyond in silico prediction and therefore remain likely but unproven. These 17 variants, and in particular, the missense variants, cluster primarily in or near to the zinc-dependent peptidase domain (Figure 3). Our analyses suggest that pathogenic mutations in this domain alter protein stability, while others (Y. J. Kim et al., 2017) have shown that variants in the catalytic domain can lead to a reduction or complete loss of enzymatic function. The combination of recessive inheritance, missense variants that reduce or abolish function, and the lack of significant difference between phenotypes associated with missense and nonsense variants, all point to a loss-of-function phenotype, where lack of functional *MMP20* gives rise to a consistent hypomaturational AI phenotype.

In summary, we have identified 10 families segregating AI due to four novel and one known mutations in *MMP20*, and reviewed previously reported mutations, raising the total number of AI-causing *MMP20* variants to 18. This expands the spectrum of *MMP20* variants implicated in AI and confirms the association with a hypomaturational phenotype. Haplotype analysis suggests that the c.710C>A, p.(Ser237Tyr); c.625G>C, p.(Glu209Gln); and c.954-2A>T, p.(Ile319Phefs*19) variants may be relatively common founder mutations in the Omani, Pakistani, and Caucasian populations, respectively. Biochemical modeling and MD analyses show that missense mutations in the catalytic domain alter protein stability, suggesting that this form of AI is the result of the loss-of-function of *MMP20*.

ACKNOWLEDGMENTS

The authors thank the families involved in this study. This study was supported by the Wellcome Trust (Grant: 093113). Georgios Nikolopoulos was funded by a PhD scholarship funded by Leeds Dental School, Faculty of Medicine and Health, University of Leeds. CELS was funded by a Wellcome Trust Institutional Strategic Support Fund award. Funding for Open Access publication charges for this article was provided by the Wellcome Trust. This study was undertaken on ARC3, part of the High-Performance Computing Facilities at the University of Leeds, UK.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.










DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in ClinVar at <https://www.ncbi.nlm.nih.gov/clinvar/>, accession numbers: SCV001338799-SCV001338802 and in the AI Leiden Open Variation Database (LOVD) at <http://dna2.leeds.ac.uk/LOVD/> with reference numbers: 0000000313-0000000317.

WEB RESOURCES

- The AI Leiden Open Variation Database (LOVD; <http://dna2.leeds.ac.uk/LOVD/>).
- ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>).
- The Genome Aggregation Database (gnomAD; <http://gnomad.broadinstitute.org>).
- Combined Annotation-Dependent Depletion (CADD v1.3; <https://cadd.gs.washington.edu>).
- MutPred2 (<http://mutpred.mutdb.org/index.html>).
- Human Splicing Finder v3.1 (<http://umd.be/HSF3/HSF.shtml>).

ORCID

- Georgios Nikolopoulos  <http://orcid.org/0000-0003-3166-8372>
- Claire E. L. Smith  <http://orcid.org/0000-0001-8320-5105>
- James A. Poulter  <http://orcid.org/0000-0003-2048-5693>
- Gina Murillo  <http://orcid.org/0000-0003-2976-6301>
- Peter F. Day  <http://orcid.org/0000-0001-9711-9638>
- Sarah A. Harris  <http://orcid.org/0000-0002-2812-1651>
- Mary J. O'Connell  <http://orcid.org/0000-0002-1877-1001>
- Chris F. Inglehearn  <https://orcid.org/0000-0002-5143-2562>
- Alan J. Mighell  <https://orcid.org/0000-0002-9624-6923>

REFERENCES

- Andreini, C., Banci, L., Bertini, I., Luchinat, C., & Rosato, A. (2004). Bioinformatic comparison of structures and homology-models of matrix metalloproteinases. *Journal of Proteome Research*, 3(1), 21–31. <https://doi.org/10.1021/pr0340476>
- Arendt, Y., Banci, L., Bertini, I., Cantini, F., Cozzi, R., Del Conte, R., & Gonnelli, L. (2007). Catalytic domain of MMP20 (Enamelysin)—The NMR structure of a new matrix metalloproteinase. *FEBS Letters*, 581(24), 4723–4726. <https://doi.org/10.1016/j.febslet.2007.08.069>
- Bartlett, J. D. (2013). Dental enamel development: Proteinases and their enamel matrix substrates. *ISRN Dentistry*, 2013, 1–24. <https://doi.org/10.1155/2013/684607>
- Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghoreishi, D., & Gilson, M. K. (2018). *AMBER 2018*. University of California, San Francisco: <http://ambermd.org/CiteAmber.php>
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLOS One*, 7(10), e46688. <https://doi.org/10.1371/journal.pone.0046688>
- Dubail, J., Huber, C., Chantepie, S., Sonntag, S., Tüysüz, B., Mihci, E., Gordon, C. T., Steichen-Gersdorf, E., Amiel, J., Nur, B., Stolte-Dijkstra, I., van Eerde, A. M., van Gassen, K. L., Breugem, C. C., Stegmann, A., Lekszas, C., Maroofian, R., Karimiani, E. G., Bruneel, A., ... Cormier-Daire, V. (2018). SLC10A7 mutations cause a skeletal dysplasia with amelogenesis imperfecta mediated by GAG biosynthesis defects. *Nature Communications*, 9(1), 3087. <https://doi.org/10.1038/s41467-018-05191-8>
- Gadhia, K., McDonald, S., Arkutu, N., & Malik, K. (2012). Amelogenesis imperfecta: An introduction. *British Dental Journal*, 212(8), 377–379. <https://doi.org/10.1038/sj.bdj.2012.314>
- Gasse, B., Prasad, M., Delgado, S., Huckert, M., Kawczynski, M., Garret-Bernardin, A., Lopez-Cazaux, S., Bailleul-Forestier, I., Manière, M. C., Stoetzel, C., Bloch-Zupan, A., & Sire, J. Y. (2017). Evolutionary analysis predicts sensitive positions of MMP20 and validates newly- and previously-identified MMP20 mutations causing amelogenesis imperfecta. *Frontiers in Physiology*, 8, 398. <https://doi.org/10.3389/fphys.2017.00398>
- Guan, X., & Bartlett, J. D. (2013). MMP20 modulates cadherin expression in ameloblasts as enamel develops. *Journal of Dental Research*, 92(12), 1123–1128. <https://doi.org/10.1177/0022034513506581>
- Kim, J. W., Simmer, J. P., Hart, T. C., Hart, P. S., Ramaswami, M. D., Bartlett, J. D., & Hu, J. C. C. (2005). MMP-20 mutation in autosomal recessive pigmented hypomaturation amelogenesis imperfecta. *Journal of Medical Genetics*, 42(3), 271–275. <https://doi.org/10.1136/jmg.2004.024505>
- Kim, J. W., Zhang, H., Seymen, F., Koruyucu, M., Hu, Y., Kang, J., Kim, Y. J., Ikeda, A., Kasimoglu, Y., Bayram, M., Zhang, C., Kawasaki, K., Bartlett, J. D., Saunders, T. L., Simmer, J. P., & Hu, J. C. (2019). Mutations in RELT cause autosomal recessive amelogenesis imperfecta. *Clinical Genetics*, 95(3), 375–383. <https://doi.org/10.1111/cge.13487>
- Kim, Y. J., Kang, J., Seymen, F., Koruyucu, M., Gencay, K., Shin, T. J., Hyun, H. K., Lee, Z. H., Hu, J. C. C., Simmer, J. P., & Kim, J. W. (2017). Analyses of MMP20 missense mutations in two families with hypomaturation amelogenesis imperfecta. *Frontiers in Physiology*, 8, 229. <https://doi.org/10.3389/fphys.2017.00229>
- Llano, E., Pendás, A. M., Knäuper, V., Sorsa, T., Salo, T., Salido, E., Murphy, G., Simmer, J. P., Bartlett, J. D., & López-Otín, C. (1997). Identification and structural and functional characterization of human enamelysin (MMP-20). *Biochemistry*, 36(49), 15101–15108. <https://doi.org/10.1021/bi972120y>
- Nikolopoulos, G., Smith, C. E. L., Brookes, S. J., El-Asrag, M. E., Brown, C. J., Patel, A., Murillo, G., O'Connell, M. J., Inglehearn, C. F., & Mighell, A. J. (2020). New missense variants in RELT causing hypomineralised amelogenesis imperfecta. *Clinical Genetics*, 97(5), 688–695. <https://doi.org/10.1111/cge.13721>
- Pandurangan, A. P., Ochoa-Montaño, B., Ascher, D. B., & Blundell, T. L. (2017). SDM: A server for predicting effects of mutations on protein stability. *Nucleic Acids Research*, 45(W1), W229–W235. <https://doi.org/10.1093/nar/gkx439>
- Petterson, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Ponzoni, L., Peñaherrera, D. A., Oltvai, Z. N., & Bahar, I. (2020). Rhapsody: Predicting the pathogenicity of human missense variants. *Bioinformatics*, 36, 3084–3092. <https://doi.org/10.1093/bioinformatics/btaa127>
- Prasad, M. K., Geoffroy, V., Vicaire, S., Jost, B., Dumas, M., Le Gras, S., Switala, M., Gasse, B., Laugel-Haushalter, V., Paschaki, M., Leheup, B., Droz, D., Dalstein, A., Loing, A., Grollemund, B., Muller-Bolla, M., Lopez-Cazaux, S., Minoux, M., Jung, S., ... Bloch-Zupan, A. (2015). A targeted next-generation sequencing assay for the molecular diagnosis of genetic disorders with orofacial involvement. *Journal of Medical Genetics*, 53(2), 98–110. <https://doi.org/10.1136/jmedgenet-2015-103302>
- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1), D886–D894. <https://doi.org/10.1093/nar/gky1016>
- Seymen, F., Park, J. C., Lee, K. E., Lee, H. K., Lee, D. S., Koruyucu, M., Gencay, K., Bayram, M., Tuna, E. B., Lee, Z. H., Kim, Y. J., & Kim, J. W. (2015). Novel MMP20 and KLK4 mutations in amelogenesis imperfecta. *Journal of Dental Research*, 94(8), 1063–1069. <https://doi.org/10.1177/0022034515590569>
- Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40(W1), W452–W457. <https://doi.org/10.1093/nar/gks539>

- Simmer, J. P., & Hu, J. C. C. (2002). Expression, structure, and function of enamel proteinases. *Connective Tissue Research*, 43(2–3), 441–449. <https://doi.org/10.1080/713713530>
- Smith, C. E. (1998). Cellular and chemical events during enamel maturation. *Critical Reviews in Oral Biology and Medicine*, 9(2), 128–161. <https://doi.org/10.1177/10454411980090020101>
- Smith, C. E. L., Poulter, J. A., Antanaviciute, A., Kirkham, J., Brookes, S. J., Inglehearn, C. F., & Mighell, A. J. (2017). Amelogenesis imperfecta: genes, proteins, and pathways. *Frontiers in Physiology*, 8, 435. <https://doi.org/10.3389/fphys.2017.00435>
- Smith, C. E. L., Poulter, J. A., Brookes, S. J., Murillo, G., Silva, S., Brown, C. J., Patel, A., Hussain, H., Kirkham, J., Inglehearn, C. F., & Mighell, A. J. (2019). Phenotype and variant spectrum in the LAMB3 form of amelogenesis imperfecta. *Journal of Dental Research*, 98(6), 698–704. <https://doi.org/10.1177/0022034519835205>
- Smith, C. E. L., Whitehouse, L. L. E., Poulter, J. A., Hewitt, L. W., Nadat, F., Jackson, B. R., Manfield, I. W., Edwards, T. A., Rodd, H. D., Inglehearn, C. F., & Mighell, A. J. (2020). A missense variant in specificity protein 6 (SP6) is associated with amelogenesis imperfecta. *Human Molecular Genetics*, 29, 1417–1425. <https://doi.org/10.1093/hmg/ddaa041>
- Wright, J. T., Carrion, I. A., & Morris, C. (2015). The molecular basis of hereditary enamel defects in humans. *Journal of Dental Research*, 94(1), 52–61. <https://doi.org/10.1177/0022034514556708>
- Wright, J. T., Torain, M., Long, K., Seow, K., Crawford, P., Aldred, M. J., Hart, P. S., & Hart, T. C. (2011). Amelogenesis imperfecta: Genotype-phenotype studies in 71 families. *Cells Tissues Organs*, 194(2–4), 279–283. <https://doi.org/10.1159/000324339>
- Yamakoshi, Y., Simmer, J. P., Bartlett, J. D., Karakida, T., & Oida, S. (2013). MMP20 and KLK4 activation and inactivation interactions in vitro. *Archives of Oral Biology*, 58(11), 1569–1577. <https://doi.org/10.1016/j.archoralbio.2013.08.005>

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Nikolopoulos, G., Smith, C. E. L., Poulter, J. A., Murillo, G., Silva, S., Lamb, T., Berry, I. R., Brown, C. J., Day, P. F., Soldani, F., Al-Bahlani, S., Harris, S. A., O'Connell, M. J., Inglehearn, C. F., & Mighell, A. J. (2021). Spectrum of pathogenic variants and founder effects in amelogenesis imperfecta associated with MMP20. *Human Mutation*, 42, 567–576. <https://doi.org/10.1002/humu.24187>