

# ACP4 Variants In Hypoplastic Amelogenesis Imperfecta

**Lu Liu**

University of Leeds <https://orcid.org/0009-0008-3593-8409>

**Cheuk Wang Au**

University of Leeds <https://orcid.org/0009-0007-0781-820X>

**Ummey Hany**

University of Leeds <https://orcid.org/0000-0002-4486-1625>

**Alice L. Rigby**

University of Leeds <https://orcid.org/0000-0001-6888-2255>

**Anesha Chauhan**

Birmingham Dental Hospital

**Catriona Brown**

Birmingham Dental Hospital <https://orcid.org/0009-0009-9963-4321>

**Jessie Sims**

St James's University Hospital <https://orcid.org/0009-0008-9335-1382>

**Gina Murillo**

University of Costa Rica <https://orcid.org/0000-0003-2976-6301>

**María Gabriela Acosta de Carmargo**

University of Carabobo <https://orcid.org/0000-0001-7615-918X>

**Chris F. Inglehearn**

University of Leeds <https://orcid.org/0000-0002-5143-2562>

**Christopher M. Watson**

University of Leeds <https://orcid.org/0000-0003-2371-1844>

**Alan J. Mighell**

University of Leeds <https://orcid.org/0000-0002-9624-6923>

**Claire E. L. Smith**

[c.e.l.smith@leeds.ac.uk](mailto:c.e.l.smith@leeds.ac.uk)

University of Leeds <https://orcid.org/0000-0001-8320-5105>

---

## Research Article

**Keywords:** ACP4, amelogenesis imperfecta, enamel, acid phosphatase 4

**Posted Date:** February 6th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-8797936/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** The authors declare no competing interests.

---

# Abstract

Amelogenesis imperfecta (AI) is a group of rare inherited conditions causing tooth enamel defects. Human acid phosphatase 4 (ACP4) is a transmembrane protein involved in maintaining appositional enamel growth. Variants in *ACP4* cause recessive hypoplastic AI. Here we identify further families and review published *ACP4* variants causing AI.

In three Pakistani families we identified a new *ACP4* variant, c.254T > C, p.(Pro85Leu), which long-read sequencing revealed to be a founder variant. Two further families were homozygous for previously reported pathogenic *ACP4* variants. Further details are also reported for two families previously listed in a technical/cohort study by this group. In total seventeen *ACP4* variants had been reported in the literature causing AI in seventeen families prior to this study. This report adds an eighteenth variant and brings the total to 22 families, suggesting that *ACP4* variants can be expected to be encountered as a cause of recessive AI, with specific variants more relevant to AI pathogenesis in particular populations.

*ACP4* variants implicated in AI include fifteen missense, one splice and two frame-breaking deletions. Missense variants lie mainly in the acid phosphatase domain, with one in the transmembrane domain. The consistent hypoplastic phenotype suggests a single mutational mechanism, and the report of a family with a homozygous frameshift variant likely to be subject to nonsense mediated decay points to loss of function. Missense variants alter amino acids at the catalytic core or affect protein stability, homodimerisation or membrane localisation, all likely to result in functional insufficiency.

## Background

Human acid phosphatase 4 (ACP4, OMIM\*606362) is a member of a family of transmembrane proteins responsible for the hydrolysis of orthophosphoric acid esters [1]. ACP4 serves essential functions only during amelogenesis (enamel formation), although it is also expressed in non-dental tissues [2]. During the secretory stage of amelogenesis, oral epithelium-derived ameloblasts secrete a highly proteinaceous matrix composed of multiple specialised phosphoproteins, including amelogenin, ameloblastin, and enamelin [3]. ACP4 is essential for both initiation and elongation of enamel mineralisation [4]. Previous research has detected acid phosphatase activity to be highest within the Tomes' process of secretory stage ameloblasts [5], with ACP4 specifically found to localise to the Tomes' process and lateral membrane of secretory stage ameloblasts [4]. As ameloblasts transition to maturation stage, secretion of enamel matrix proteins (EMPs) ceases and proteolytic enzymes degrade and remove the EMPs, which are transported to lysosomes for further degradation [6]. Concomitantly, hydroxyapatite crystals grow in width and thickness, replacing tissue fluid during maturation [7].

Amelogenesis imperfecta (AI) is a heterogeneous group of rare Mendelian disorders which affect the enamel of both dentitions [8]. Non-syndromic AI represents a disease subgroup in the absence of other morphological or biochemical changes in the body [9]. AI results in abnormalities in enamel structure, appearance and function [10]. Phenotyping AI teeth is challenging due to their wide spectrum of clinical presentation resulting from post-eruptive breakdown [11, 12]. The hypoplastic form is distinguished by

quantitative defects in mineralised enamel. Clinical presentations can vary from a pitted appearance to a thinner enamel layer or the complete absence of enamel [12]. In the hypomineralised form, dental enamel is soft or mottled due to insufficient calcification during mineralisation [13]. By contrast, hypomaturational AI displays normal enamel thickness but defects in the hardness of enamel with disturbed mineralisation at the maturation stage [14, 15].

Multiple studies have identified autosomal recessive mutations in *ACP4* causing non-syndromic hypoplastic AI (type IJ). To date, 17 families with 17 homozygous or compound heterozygous pathogenic *ACP4* variants have been reported by various groups (Table 1) [4, 16–20]. The majority of variants (14/17) result in missense changes. Kim et al. showed that three out of four pathogenic missense *ACP4* variants that they tested resulted in significantly decreased protein expression and decreased acid phosphatase activity levels, suggesting that *ACP4* variants may cause disease through functional insufficiency [17]. From phenotyping studies of the *Acp4*<sup>R110C/R110C</sup> mouse model, it has been demonstrated that ACP4 functions during appositional growth of enamel during amelogenesis, although the exact function of ACP4 remains unclear [4].

Here we report a novel homozygous missense founder variant and two previously reported variants, causing AI in five families with recessive hypoplastic AI, as well as additional information on two families previously only listed in a technical/cohort study by this group. We also review the mutation spectrum of *ACP4* variants identified in individuals with AI published to date.

## Materials and Methods

### Recruitment and sample preparation

Clinical phenotypes were assessed by the recruiting dentist. Participants' (affected individuals and relatives if available) personal information and blood/saliva for DNA extraction were collected with informed consent obtained, adhering to the principles of the Declaration of Helsinki. Ethical approval was obtained from the Yorkshire and Humber - Leeds East Research Ethics Committee (reference 13/YH/0028; IRAS project ID82448). Genomic DNA was obtained from saliva using Oragene® DNA Sample Collection kits (DNA Genotek, Ottawa, ON, Canada) following the manufacturer's protocol. One family described herein (Family 4) was screened as part of the UK National Health Service (NHS) R340 (Supplement Table S1) genetic test for AI which uses DNA obtained from blood (<https://nhsgms-panelapp.genomicsengland.co.uk/panels/269/v3.0>).

### Variant nomenclature

All variant nomenclature for *ACP4*/*ACP4* refers to *ACP4* transcript NM\_033068.3 and ACP4 protein sequence NP\_149059.1.

## Massively parallel sequencing

Proband genomic DNA was processed using the Twist exome comprehensive capture kit (Twist Bioscience, San Francisco, CA, USA), SureSelect All Exons version 5 kit (Agilent, Santa Clara, CA, USA) or using our custom single-molecule molecular inversion probes (smMIPs) protocol as described by Hany et al [18]. Data were analysed by short-read next generation sequencing of either whole-exome sequencing (WES), or smMIPs data generated on HiSeq 3000, NextSeq 500, or NextSeq 2000 sequencers (Illumina).

WES reads were first processed by Trim Galore (v.0.6.10) and Cutadapt (v.4.4) [21]. The sequences were aligned to the indexed human reference genome hg38 using the BWA software (v.0.7.17) [22]. PCR duplicates were removed using Picard (v.3.1.0) (Broad Institute, 2019). Non-reference bases from 96 samples were combined into cohort data before being genotyped by Genome Analysis Tool Kit (GATK) HaplotypeCaller (v.4.4.0.0) according to recommended best practice workflows [23]. Identified sequence variants were then annotated using Ensembl Variant Effect Predictor (VEP) (v.110.1), along with the population allele frequencies, CADD (v.1.4), and 5' UTR according to the Genome Aggregation Database (gnomAD; v.3.1.2) and the dbSNP build 138 [24–26]. Annotated data were filtered using VASE (v.0.5.1) (<https://github.com/david-a-parry/vase>) by chromosomal regions of known and candidate AI genes, CADD score and allele frequency in gnomAD, 1000 Genomes, and dbSNP databases [27, 28]. Variants were filtered for a list of known AI genes and those causative for AI in animal models (Supplement Table S1) reflecting that participants had non-syndromic AI [29]. A minor allele frequency (MAF) of less than 0.01 was used to filter out any variants that were too common in the population to be a plausible cause of recessively inherited AI. Variants with a CADD score (prediction of variant deleteriousness) of less than 15.0 were excluded [25]. Variant pathogenicity status was classified by the updated ClinVar database and the web-based platform Franklin using American College of Medical Genetics and Genomics (ACMG) guidelines [30, 31]. The Linux command line for WES analysis is provided in Supplement.

## Single-molecule molecular inversion probes (smMIPs) analysis

The smMIPs data was processed by an in-house bioinformatics pipeline following the manufacturer's instructions and as previously described by Hany et al [18].

### Polymerase chain reaction (PCR) amplification of full length ACP4 fragments

To determine the ancestral haplotype of the c.245T > C p.(Leu85Pro) *ACP4* variant, long-read sequencing of a 9,647 bp long-range PCR product (spanning chr19:50,788,531 – 50,798,177) was performed using a GridION sequencer (Oxford Nanopore Technologies (ONT), Oxford, UK).

Each PCR consisted of 1 µL genomic DNA (~ 1-100 ng/µL), 2 µL of 10× SequelPrep reaction buffer (Invitrogen, Paisley, UK), 0.36 µL of 5 U/µL SequelPrep long polymerase (Invitrogen), 0.4 µL of dimethyl

sulfoxide (Invitrogen), 1 µL of 10× SequelPrep Enhancer A (Invitrogen), 13.24 µL of nuclease free water and 1 µL each of 10µM forward (dCATCTCCAGGGGCTAGAACG) and reverse (dGGCTCTCAAGGACAGACACC) primers. Thermocycling conditions comprised a denaturation step at 94°C for 2 minutes followed by 10 cycles of 94°C for 10 seconds, 57°C for 30 seconds and 68°C for 10 minutes then 20 cycles of 94°C for 10 seconds, 57°C for 30 seconds and 68°C for 16 minutes with an additional 20 seconds added per cycle, before a final extension step at 72°C for 5 minutes. Amplification products were excised from a 0.8% agarose TAE gel and purified using a QIAquick column following manufacturer's protocols (Qiagen).

To create a nanopore sequencing library using the SQK-LSK114 kit, an end-prep reaction was first performed. This comprised 25 µL of purified amplification product, 3.5 µL Ultra II End-prep Reaction Buffer (New England Biolabs (NEB), Ipswich, Massachusetts, USA) and 1.5 µL of Ultra II End-prep Enzyme Mix (NEB) with the reaction being incubated at 20°C for 5 minutes then 65°C for 5 minutes. The reaction was cleaned-up with 30 µL of AMPure XP beads (Beckman Coulter, Indianapolis, Indiana, USA) using two 200 µL 80% ethanol washes and eluted in 31 µL of nuclease-free water. Sequencing adapters were ligated to double stranded DNA in a reaction that comprised 30 µL of treated PCR products, 12.5 µL of Ligation Buffer (LNB; ONT), 2.5 µL of Ligation Adapter (LA; ONT) and 5.0 µL of Salt-T4 DNA Ligase (NEB). The reaction was incubated at room temperature for 10 minutes. An AMPure XP bead (Beckman Coulter) cleanup was next carried out with two 125 µL Long Fragment Buffer washes before the library was eluted using 7 µL of buffer EB (ONT). Flongle flowcells were primed using a solution that comprised 117 µL of Flow Cell Flush (FCF) reagent combined with 3 µL of Flow Cell Tether (FCT) reagent. A sequencing mix comprising 5 µL of the final library, 15 µL of Sequencing Buffer (SB; ONT) and 10 µL of Library Beads (LIB; ONT) was then loaded to the flowcell before a 24-hour sequencing run was initiated using MinKNOW software v.24.11.8.

Offline basecalling was performed using Dorado v.1.1.1 (<https://github.com/nanoporetech/dorado>), with the high accuracy calling model (v.5.2.0), to convert raw pod5 files to FASTQ format. The resulting sequences were selected based on their length (9,147 – 10,147 bp) and quality (Q12) using chopper v.0.11.0 (<https://github.com/wdecoster/chopper>) [32]. Reads were next aligned to an indexed human reference genome (build GRCh38) using minimap2 v.2.30 (<https://github.com/lh3/minimap2>), then sorted by alignment coordinate and indexed using samtools v.1.22.1 (<http://www.htslib.org/>) [33]. Sequencing metrics were evaluated using NanoStat v.1.6.0 (<https://github.com/wdecoster/nanostat>) and mapped sequenced reads were visualised using the Integrative Genomics Viewer (IGV) v.2.16.2 (<https://igv.org/doc/desktop/>) [34, 35].

## Sanger sequencing and segregation analyses

Primer pairs were designed using AutoPrimer3 (<https://github.com/david-a-parry/autoprimers3>) and synthesised by IDT (Leuven, Belgium) to bracket each variant as shown in Supplemental Table S2. Targeted regions were amplified by PCR using Q5® High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA, USA). PCR amplicons were evaluated by agarose gel electrophoresis and purified by

ExoSAP-IT (ThermoFisher Scientific, Waltham, MA, USA). Sanger sequencing was performed using BigDye Terminator v.3.1 and resolved on the ABI3130xl Genetic Analyser (Applied Biosystems, Paisley, UK). The electropherograms were analysed by 4Peaks (v.1.8) (Nucleobytes).

## Computational structural analysis

Computational modelling of ACP4 variants were performed using Rosetta 3.10, as previously described by Au et al [36]. All calculations were conducted on the University of Leeds AIRE high-performance computing cluster. As there is no crystal structure available for ACP4, an initial model was generated using AlphaFold3 based on the amino acid sequence, and this model served as the starting point for all subsequent modelling [37].

To obtain a low-energy configuration of the structure prior to mutagenesis, the wild-type structure was subjected to the Rosetta Relax protocol. A total of 1,000 relaxations were generated, and the lowest-energy model was selected for further modelling. Single point mutations were introduced using the RosettaScripts framework with the *MutateResidue* function. Each mutant structure was subjected to the same scoring function to ensure local backbone and side-chain optimisation. For each variant, 1,000 independent models were generated. The scoring function provides Rosetta Energy Unit (REU) scores, and the wildtype REU was subtracted from the mutant REU to obtain the  $\Delta$ REU for each variant. The locations of the phosphatase active site, signal peptide, and membrane-bound regions were identified using Interpro [38].

## Results

### Dental phenotyping

All probands for which DNA was sequenced from Families 1–7 were diagnosed with hypoplastic AI based on clinical examination and images. Images of teeth from affected individuals from Families 4, 6 and 7 show hypoplastic AI with reduced quantity of enamel in both primary and permanent dentitions. The posterior teeth had yellow discolouration from thin or absent enamel. Anterior teeth were also hypoplastic but to a more limited extent with reduced discolouration and evidence of pitting in some cases (Fig. 1).

#### Identification of ACP4 variants

Three *ACP4* variants were identified among five families, as a result of on-going screening using various approaches in a cohort of over 400 AI families (Fig. 2). The cohort includes probands with syndromic and non-syndromic AI and is composed both of locally ascertained patients and cases recruited by national and international collaborators.

A novel homozygous variant, c.254T > C, p.(Leu85Pro) in exon 3 of *ACP4*, was identified in three UK families (Families 1–3) of Pakistani origin (Fig. 2a). This alteration is classified as a Variant Of Unknown

Significance (VOUS) by Franklin. However, the variant clearly co-segregates with AI in multiple affected family members in each family and given the shared ethnic origin of these families, it is likely they are related, implying further co-segregation with disease between families. When, in light of these findings, criterion PP1 (Pathogenic supporting) is manually changed to strongly supporting, this variant is reclassified likely pathogenic. Furthermore, it is predicted to affect protein function (Sorting Intolerant From Tolerant score = 0.00) [39] and sequence alignments of orthologues indicate that Leu 85 in *ACP4* is conserved across 14 mammalian species (Supplement Fig. S1).

The proband in Family 4 was born in the UK from consanguineous parents of Pakistani origin. NHS R340 genetic testing revealed a homozygous *ACP4* variant c.746C > T, p.(Pro249Leu) (rs1085307111), which has been reported previously by Smith et al in another UK Pakistani consanguineous family (Fig. 2b) [20]. The Family 5 proband was also born from consanguineous parents of UK Pakistani origin. A homozygous pathogenic variant, c.331C > T, p.(Arg111Cys) (rs202073531), previously identified by Seymen et al. in a Turkish family, was identified in proband II:4 following whole exome sequencing analysis (Fig. 2c) [16].

*ACP4* variants identified by smMIPs analysis in Families 6 and 7 were reported in brief as part of a larger cohort analysis by Hany et al [18], but no family, phenotype or ethnicity details were included. In Family 6, originating from Costa Rica, compound heterozygous *ACP4* variants were identified, including a variant identical to that identified in Family 5, c.331C > T, p.(Arg111Cys) (rs202073531) and a previously unreported frameshift variant, c.433delC, p.(Val146Trpfs\*7) (absent in gnomAD) (Fig. 2d). Finally, a previously unreported homozygous missense variant, c.845T > C, p.(Met282Thr) was identified in the proband of Family 7, who are Venezuelan (Fig. 2e).

A list of all published *ACP4* variants and *ACP4* associated AI cases reported to date are reported in Tables 1 and 2 respectively, including those in this report. Their locations on the transcript are shown in Fig. 2f. Positions of the altered amino acid residues are shown in Fig. 2g. All *ACP4* variants identified in this study were submitted to ClinVar: SCV007338208-SCV007338211.

## Haplotype analysis using long read nanopore sequencing

The presence of the c.254T > C, p.(Leu85Pro) variant in three families from the same geographical area suggests that they may share a common ancestor. This was further investigated by determining the haplotype of flanking variants using long-read PCR. A 9,647 bp DNA segment spanning exons 1–11 of *ACP4* was PCR amplified and analysed by long-read sequencing in probands from Families 1, 2, and 3, all of whom were homozygous for c.254T > C, p.(Leu85Pro). An identical haplotype consisting of 21 nonreference nucleotides arranged in cis with c.254T > C and spanning at least 9,647 bp was shared by the three families analysed, suggesting a shared common ancestor (<https://www.ncbi.nlm.nih.gov/bioproject/1403257>) (Fig. 3, Table 3).

Upon analysis of the flanking SNPs identified within *ACP4*, another downstream missense variant, c.973\_974delinsAA, p.(Ala325Asn) rs386810263, with a CADD score of 13.98, was identified in all three

samples. This variant was not present in gnomAD, but two common missense variants, rs55716643 and rs55735528, were reported causing two different substitutions of the same amino acid, changing it from Ala to Asp or to Thr. These variants have near identical allele frequencies (0.29) and one substitutes the first nucleotide of the triplet, G, for A, while the other substitutes the second nucleotide, a C, for A. Together they would therefore replicate the c.973\_974delinsAA variant, suggesting they are in fact one variant, as observed in our sequence data, but have been misclassified as two. If so, this variant has a true frequency of 0.29, excluding it as a cause of disease.

## Structural modelling

Computational modelling was carried out using Rosetta to predict the structural implications of the fifteen likely pathogenic missense *ACP4* variants published here or in previous reports, via their impact on predicted stability. For comparison, we also assessed three *ACP4* missense changes listed in gnomAD but considered too common to be disease causing (allele frequency > 0.05 in gnomAD and the likely misclassified p.(Ala325Asn) variant) and predicted to be benign by Franklin (Supplement Table S3). The Rosetta Energy Units (REU) score resembles kcal/mol, and predicted  $\Delta$ REU values obtained *in silico* have been found to be comparable to those obtained experimentally [40]. The substituted amino acids have a range of effects on stability, which are reflected in the  $\Delta$ REU calculations (Fig. 4). We also assessed the change to the local environment of each substituted amino acid, including predicted interactions with surrounding residues and secondary structures (Fig. 5) and produced a video to show the residue positions substituted as a result of all the pathogenic variants reported to date (<https://doi.org/10.5518/1814>).

Of the variants described in this study, Leu85Pro displayed the strongest destabilisation, due to the disruption of an  $\alpha$ -helical region that is important for maintaining fold integrity. Leu85Pro introduces backbone constraints and removes an amide hydrogen required for hydrogen bonding to ensure helix stabilisation [41]. This, in addition to the introduction of a pyrrolidine ring, could lead to a substantial helix distortion. In contrast, Arg111Cys had only slight energetic perturbation, which suggests that the overall structure stability is maintained, despite the loss of a charged side chain. Arg111 is involved in polar contacts but is not buried deeply within the loop, suggesting a structural tolerance for residue substitutions. The change to a cysteine residue likely leads to the loss of some electrostatic interactions.

The effects of Pro249Leu and Met282Thr were observed to be intermediate, or only causing local effects, rather than a total structural collapse. Pro249 enforces loop geometry, and the change to a leucine introduces flexibility, causing loss of backbone rigidity [42]. Pro249Leu also introduces a bulky, hydrophobic side chain into the loop, which could lead to a slight repacking of the protein, inferring a conformational change. Met282Thr introduces a smaller, polar hydroxyl group into a hydrophobic region, which can lead to the exposure of the core to water [43].

Other published variants had varying predicted effects on stability, with the substitutions involving proline, Ala128Pro and Leu209Pro, having the largest impacts due to the introduction of a rigid amino acid with structural constraint.

Modelling of variations predicted to be benign due to their reported allele frequency and/or CADD score was included for comparison with variations predicted to be pathogenic. The median  $\Delta$ REU value obtained for likely pathogenic variants identified in individuals with AI was 5.06 kcal/mol, whereas for the predicted benign variants, the median value was 4.57 kcal/mol.

## Discussion

Here we describe the identification of a novel homozygous missense founder variant and four additional previously reported variants in *ACP4* (MIM \*606362) in seven families with recessive hypoplastic AI. The homozygous c.254T > C, p.(Leu85Pro) variant, identified in three different UK-Pakistani families (Family 1–3), segregated with AI in all available additional family members. Long-read sequencing confirmed an identical haplotype among three probands from Family 1–3, suggesting they are distantly related. Notably, another mutation, c.713C > T, p.(Ser238Leu), has been detected in four Turkish families to date, suggesting that this *ACP4* variant may also be a founder variant within that population [16, 17]. The other four variants identified include previously published variants, c.746C > T, p.(Pro249Leu) (rs1085307111) identified as biallelic in Family 4 and c.331C > T, p.(Arg111Cys) (rs202073531) identified in as biallelic in Family 5 and as heterozygous in Family 6 [16, 20]. The p.(Pro249Leu) variant has now been identified in two British families of Pakistani heritage to date and the p.(Arg111Cys) variant has been identified in 3 families, including those of Turkish and Costa Rican heritage. These findings add to the current mutation spectrum for *ACP4* and further underline its critical role in pathological amelogenesis, although the function of this phosphatase during enamel formation remains unclear [17].

To date, including families reported here, 22 families with AI due to segregating *ACP4* pathogenic variants have been reported in the literature, summarised in Table 2. Among the nine families with *ACP4* variants identified in Leeds, seven are of Pakistani origin, one Costa Rican and one Venezuelan. This reflects a bias in the Leeds AI cohort towards local families of Pakistani heritage, which results primarily from the increased incidence of consanguinity and the resulting recessive disease in the Bradford Asian community [44]. Previously published reports of *ACP4* AI reflect a wide heritage. Accordingly, *ACP4* variants can be expected to be encountered as a cause of recessive AI. Current understanding is that *ACP4* variants are not causative of disease in other tissues and organs, unlike many other genes causative for recessive hypoplastic AI. As additional individuals with AI due to *ACP4* variants are identified, including via state-driven genomic testing, it will become clearer if there is ever an impact beyond enamel formation. Improved understanding of the optimum patient pathway for affected individuals will follow.

All variants appear, from the reported hypoplastic enamel phenotypes, to perturb the secretory phase of amelogenesis. However, the individual who carried compound heterozygous *ACP4* variants: c.626T > C, p.(Leu209Pro);c.1199C > A, p.(Ala400Asp), also had agenesis of teeth 18 and 28, consistent with a *WNT10A* phenotype. The individual also carries a heterozygous *WNT10A* variant: NM\_025216.3:c.682T > A, p.(Phe228Ile). In addition to selective tooth agenesis, *WNT10A* variants have also been tentatively associated with hypoplastic enamel phenotypes in mice, although the effect might result from reduced

dentine volume [45]. Of the *ACP4* variants reported to date, c.626T > C, p.(Leu209Pro), c.1199C > A, p.(Ala400Asp) and the intron 6 splice variant c.645 + 1G > A have only been reported as part of a large cohort study which did not detail family pedigrees, nor confirm segregation [19]. Therefore, the causal association of the *ACP4* c.626T > C, p.(Leu209Pro) and c.1199C > A, p.(Ala400Asp) variants with AI require further validation.

In all, fifteen missense, one splice and two frame-breaking deletion *ACP4* pathogenic variants have been identified. Most variants are clustered in exons 3 (3 missense variants), 4 (6 missense, 1 frameshift) and 7 (3 missense, 1 frameshift) in the histidine phosphatase domain, with all but one variant (c.1199C > A, p.(Ala400Asp)) located within the extracellular portion of the protein. It has been suggested that biallelic *ACP4* variants causing AI do so through functional insufficiency [17]. One possible mechanism by which amino-acid substitutions can cause loss of function is through alterations in protein stability and/or binding interactions. To test for alterations in protein stability, we performed computational modelling of the fifteen likely pathogenic missense *ACP4* variants and, for comparison, three likely *ACP4* missense polymorphisms, using Rosetta 3.10 [36]. Rosetta's *MutateResidue* protocol and associated scoring function, which evaluate the impact of amino acid substitutions on protein stability, were used in this study. This analysis suggests that protein stability does play a role in AI disease causation for some, but not all, pathogenic variants, with a range of effects varying from no effect on stability to very substantial changes. The median score for disease causing variants was 5.06 kcal/mol, while benign variants were predicted to have less impact on stability, with a median score of 4.57 kcal/mol. This indicates that these residues may be more tolerant to substitutions but the relatively similar value may suggest that protein stability is only significantly affected by particular variants which are also more likely to exert a pathogenic effect. Consequently, it should be noted that alterations in binding affinity are not captured in the values obtained from this analysis. Binding interactions are also likely to be important to its proper function. *ACP4* is also known to function as a homodimer, with residues at the interface between the monomers important to its structure and residues at its catalytic core fundamental to its function [16, 17].

The extreme destabilisation predicted for c.254T > C, p.(Leu85Pro) and c.626T > C, p.(Leu209Pro) is likely to cause loss of function due to the inclusion of a proline residue, which would be expected to disrupt wild-type secondary structure by affecting folding or stability of the protein. Conversely, Arg111Cys is predicted to maintain stability similar to wild-type. However, the InterPro Protein family database predicts that residues 41, 44, 111, 288 and 289 make up the catalytic core, meaning that even minor changes in local electrostatic interactions or changes of binding interactions at residue 111 are likely to be sufficient to cause loss of function (European Bioinformatics Institute (EMBL-EBI), <http://www.ebi.ac.uk/>) [46]. The c.1199C > A, p.(Ala400Asp) also shows only a minor effect on stability, but this variant, which is in the transmembrane domain (residues 394–426), replaces a hydrophobic alanine with a hydrophilic aspartic acid, and may therefore disrupt insertion of the protein into the membrane. Overall, these findings are therefore consistent with a functional insufficiency mechanism and suggest that the missense variants documented have a range of different effects on the *ACP4* protein.

Phenotypically, ACP4-related AI teeth revealed a generalised hypoplastic appearance. Clinical information suggests a characteristic yellow appearance of the posterior molars, with evidence of striated or pitted enamel present on the incisors. To categorise phenotypes of ACP4-related AI accurately, radiographs before eruption are required to measure the enamel volume and exclude any post-eruptive breakdown of the enamel. A previous study examining an extracted tooth from an individual carrying a homozygous *ACP4* c.746C > T, p.(P249L) variant estimated the enamel layer of a molar to be approximately 10% of the thickness of a healthy enamel layer of a matched tooth, although this was calculated from teeth that had experienced post-eruptive wear [20]. Further phenotypic studies of unerupted teeth are required to reveal the phenotypic effects of *ACP4* variants in humans.

In summary we have identified a new founder variant in *ACP4* in the South Asian population and the genetic cause of AI for 7 families. Our screening methods include use of the UK National Health Service screen for AI and a cost-effective, custom made smMIPs capture reagent, as well as exome sequencing [18]. These data suggest that biallelic *ACP4* variants account for a significant proportion of recessive AI families. We have modelled the effects of variants on the stability and structure of ACP4, showing that some variants are likely to alter protein stability while others act through different mechanisms, all of which are likely to lead to functional insufficiency. These findings provide benefits in aiding the genetic diagnosis and management of AI in the future.

## Declarations

### Author Contributions

Conceptualization: Lu Liu, Cheuk W. Au, Ummey Hany, Chris F. Inglehearn, Christopher M. Watson, Alan J. Mighell, Claire E.L. Smith; Methodology: Lu Liu, Cheuk W. Au, Ummey Hany, Chris F. Inglehearn, Christopher M. Watson, Alan J. Mighell, Claire E.L. Smith; Formal Analysis: Lu Liu, Cheuk W. Au, Ummey Hany, Christopher M. Watson; Investigation: Lu Liu, Cheuk W. Au, Ummey Hany, Alice L Rigby, Anesha Chauhan, Catriona Brown, Jessie Sims, Gina Murillo, Maria Gabriela Acosta de Carmargo, Chris F. Inglehearn, Christopher M. Watson, Alan J. Mighell, Claire E.L. Smith; Resources: Anesha Chauhan, Catriona Brown, Gina Murillo, Maria Gabriela Acosta de Carmargo, Alan J. Mighell; Data Curation: Lu Liu, Cheuk W. Au, Ummey Hany, Christopher M. Watson, Claire E.L. Smith; Writing – Original Draft Preparation: Lu Liu, Cheuk W. Au, Claire E.L. Smith; Writing – Review & Editing: Lu Liu, Cheuk W. Au, Ummey Hany, Alice L Rigby, Anesha Chauhan, Catriona Brown, Jessie Sims, Gina Murillo, Maria Gabriela Acosta de Carmargo, Chris F. Inglehearn, Christopher M. Watson, Alan J. Mighell, Claire E.L. Smith; Visualization: Lu Liu, Cheuk W. Au, Ummey Hany, Alice L Rigby, Christopher M. Watson; Supervision: Catriona Brown, Gina Murillo, Maria Gabriela Acosta de Carmargo, Chris F. Inglehearn, Christopher M. Watson, Alan J. Mighell, Claire E.L. Smith; Project Administration: Chris F. Inglehearn, Alan J. Mighell, Claire E.L. Smith; Funding Acquisition: Ummey Hany, Alice L Rigby, Chris F. Inglehearn, Christopher M. Watson, Alan J. Mighell.

Claire E.L. Smith is the guarantor of the paper.

All authors revised the paper critically for intellectual content and approved the final version. All authors agree to be accountable for the work and to ensure that any questions relating to the accuracy and integrity of the paper are investigated and properly resolved.

### **Competing Interests**

The authors have no relevant financial or non-financial interests to disclose.

### **Acknowledgement**

This research was funded by the Wellcome Trust (WT093113MA), the Rosetrees Trust (PGS19-2/10111 and PGL23 100152) and a Leeds Doctoral Scholarship.

The authors thank the families involved in this study.

The authors would like to dedicate this paper to the memory of Sandra Silva, University of Costa Rica, who greatly contributed to this study, but who sadly passed away before this manuscript was written.

### **Ethics Approval**

Informed consent was obtained from participating families adhering to the principles of the Declaration of Helsinki with ethical approval granted (REC reference 13/YH/0028; IRAS project ID82448).

### **Data Accessibility Statement**

All non-identifiable sequencing data is available. Exome sequencing, smMIPs sequencing and NHS R340 sequencing data is not available as it is deemed to be identifying.

Nanopore sequencing data was uploaded to the Sequence Read Archive, available at:  
<https://www.ncbi.nlm.nih.gov/bioproject/1403257>

The data associated with this paper are openly available from the University of Leeds Data Repository:  
<https://doi.org/10.5518/1814>

## **References**

1. Yousef GM, Diamandis M, Jung K, Diamandis EP (2001) Molecular Cloning of a Novel Human Acid Phosphatase Gene (ACPT) That Is Highly Expressed in the Testis. *Genomics* 74:385–395.  
<https://doi.org/https://doi.org/10.1006/geno.2001.6556>
2. Mu Y, Huang X, Liu R, Gai Y, Liang N, Yin D, Shan L, Xu S, Yang G (2021) ACPT gene is inactivated in mammalian lineages that lack enamel or teeth. *PeerJ* 9:e10219.  
<https://doi.org/10.7717/peerj.10219>
3. Fincham AG, Moradian-Oldak J, Simmer JP (1999) The structural biology of the developing dental enamel matrix. *J Struct Biol* 126:270–299. <https://doi.org/10.1006/jsbi.1999.4130>

4. Liang T, Wang S-K, Smith C, Zhang H, Hu Y, Seymen F, Koruyucu M, Kasimoglu Y, Kim J-W, Zhang C, Saunders TL, Simmer JP, Hu JCC (2022) Enamel defects in Acp4R110C/R110C mice and human ACP4 mutations. *Sci Rep* 12:16477. <https://doi.org/10.1038/s41598-022-20684-9>
5. Hammarström LE, Wade CWR, Hanker JS, Toverud SU (1971) Cellular Differences in Acid Phosphatase Isoenzymes in Bone and Teeth. *Clin Orthop and Relat Res*® 78
6. Nanci A, Slavkin HC, Smith CE (1987) Application of High-Resolution Immunocytochemistry To the Study of the Secretory, Resorptive, and Degradative Functions of Ameloblasts. *Adv Dent Res* 1:148–161. <https://doi.org/10.1177/08959374870010020301>
7. Robinson C (2014) Enamel maturation: a brief background with implications for some enamel dysplasias. *Front Physiol* 5:388. <https://doi.org/10.3389/fphys.2014.00388>
8. Gadhia K, McDonald S, Arkutu N, Malik K (2012) Amelogenesis imperfecta: an introduction. *Br Dent J* 212:377–379. <https://doi.org/10.1038/sj.bdj.2012.314>
9. Bailleul-Forestier I, Molla M, Verloes A, Berdal A (2008) The genetic basis of inherited anomalies of the teeth: Part 1: Clinical and molecular aspects of non-syndromic dental disorders. *Eur J Med Genet* 51:273–291. <https://doi.org/10.1016/j.ejmg.2008.02.009>
10. Hu JC-C, Chun Y-HP, Al Hazzazzi T, Simmer JP (2007) Enamel Formation and Amelogenesis Imperfecta. *Cells Tissues Organs* 186:78–85. <https://doi.org/10.1159/000102683>
11. Witkop CJ Jr. (1988) Amelogenesis imperfecta, dentinogenesis imperfecta and dentin dysplasia revisited: problems in classification. *J Oral Pathol Med* 17:547–553. <https://doi.org/https://doi.org/10.1111/j.1600-0714.1988.tb01332.x>
12. Crawford PJM, Aldred M, Bloch-Zupan A (2007) Amelogenesis imperfecta. *Orphanet J Rare Dis* 2:17. <https://doi.org/10.1186/1750-1172-2-17>
13. Brook A (2009) Multilevel complex interactions between genetic, epigenetic and environmental factors in the aetiology of anomalies of dental development. *Archives Oral Biol* 54:S3–S17
14. Seow W (2014) Developmental defects of enamel and dentine: challenges for basic science research and clinical management. *Aust Dent J* 59:143–154. <https://doi.org/https://doi.org/10.1111/adj.12104>
15. Wang YL, Lin HC, Liang T, Lin JC, Simmer JP, Hu JC, Wang SK (2024) ENAM Mutations Can Cause Hypomaturation Amelogenesis Imperfecta. *J Dent Res* 103:662–671. <https://doi.org/10.1177/00220345241236695>
16. Seymen F, Kim YJ, Lee YJ, Kang J, Kim T-H, Choi H, Koruyucu M, Kasimoglu Y, Tuna EB, Gencay K, Shin TJ, Hyun H-K, Kim Y-J, Lee S-H, Lee ZH, Zhang H, Hu JCC, Simmer JP, Cho E-S, Kim J-W (2016) Recessive Mutations in ACPT, Encoding Testicular Acid Phosphatase, Cause Hypoplastic Amelogenesis Imperfecta. *Am J Hum Genet* 99:1199–1205. <http://doi.org/10.1016/j.ajhg.2016.09.018>
17. Kim YJ, Lee Y, Kasimoglu Y, Seymen F, Simmer JP, Hu JC, Cho ES, Kim JW (2022) Recessive Mutations in ACP4 Cause Amelogenesis Imperfecta. *J Dent Res* 101:37–45. <https://doi.org/10.1177/00220345211015119>

18. Hany U, Watson CM, Liu L, Nikolopoulos G, Smith CEL, Poulter JA, Antanaviciute A, Rigby A, Balmer R, Brown CJ, Patel A, de Camargo MGA, Rodd HD, Moffat M, Murillo G, Mudawi A, Jafri H, Mighell AJ, Inglehearn CF (2025) Genetic Screening of a Nonsyndromic Amelogenesis Imperfecta Patient Cohort Using a Custom smMIP Reagent for Selective Enrichment of Target Loci. *Hum Mutat* 2025:8942542. <https://doi.org/10.1155/humu/8942542>
19. Bloch-Zupan A, Rey T, Jimenez-Armijo A, Kawczynski M, Kharouf N, O-Rc D-M, Noirrit ML, Hernandez E, Joseph-Beaudin M, Lopez C, Tardieu S, Thivichon-Prince C, Dostalova BECC, Macek T, Alloussi MIC, Qebibo ME, Morkmued L, Pungchanchaikul S, Orellana P, Manière BU, Gérard M-C, Bugueno B, Laugel-Haushalter IM, Alembik V, Ahossi Y, Bailleul-Forestier V, Blanchet I, Berdal I, Boileau A, Chassaing MJ, Clauss N, Delfosse F, De-Saint-Martin C, Dahlet A, Doray J-C, Davideau B, Davit-Béal J-L, Dollfus T, Duprez H, de La Dure Molla J-P, Dieterich M, Droz K, El Chehadeh D, Etienne S, Euvrard O, Faivre E, Fournier L, Garot B, Grollemund E, Guffon-Fouilhoux B, Hernandez N, Huckert M, Isidor M, Joseph-Beaudin B, Jung C, Lacombe S, Lavillaurex D, Lebrun A, Leheup M, Loing B, Lopez A, Marlin S, Morrier S, Muller-Bolla J-J, Noirrit M, Odent E, Paule Gelle S, Piard M, Pons J, Richard L, Rossi B, Sadones M, Schaefer P, Sixou E, Soskin J-L, Strub S, Tardieu M, Thivichon-Prince C, Toutain B, Verloes A, Vaysse A, Wagner F, Amar D, Dostalova JL, El Alloussi T, Macek M, Morkmued M, Noura S, Pungchanchaikul Z, Qebibo P, Revencu L, Tunisie N, Urzúa S, Orellana B (2023) Amelogenesis imperfecta: Next-generation sequencing sheds light on Witkop's classification. *Front Physiol* 14. <https://doi.org/10.3389/fphys.2023.1130175>
20. Smith CE, Whitehouse LL, Poulter JA, Brookes SJ, Day PF, Soldani F, Kirkham J, Inglehearn CF, Mighell AJ (2017) Defects in the acid phosphatase ACPT cause recessive hypoplastic amelogenesis imperfecta. *Eur J Hum Genet* 25:1015–1019. <https://doi.org/10.1038/ejhg.2017.79>
21. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011 17:3. <https://doi.org/10.14806/ej.17.1.200>
22. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinform* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
23. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498. <https://doi.org/10.1038/ng.806>
24. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F (2016) The Ensembl Variant Effect Predictor. *Genome Biol* 17:122. <https://doi.org/10.1186/s13059-016-0974-4>
25. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M (2018) CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 47:D886–D894. <https://doi.org/10.1093/nar/gky1016>
26. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nat* 581:434–443. <https://doi.org/10.1038/s41586-020-2308-7>

27. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nat* 536:285–291. <https://doi.org/10.1038/nature19057>
28. Consortium GP (2015) A global reference for human genetic variation. *Nature* 526:68
29. Smith CEL, Poulter JA, Antanaviciute A, Kirkham J, Brookes SJ, Inglehearn CF, Mighell AJ (2017) Amelogenesis Imperfecta; Genes, Proteins, and Pathways. *Front Physiol* 8. <https://doi.org/10.3389/fphys.2017.00435>
30. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 44:D862–D868. <https://doi.org/10.1093/nar/gkv1222>
31. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24–26. <https://doi.org/10.1038/nbt.1754>
32. De Coster W, Rademakers R (2023) NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinform* 39. <https://doi.org/10.1093/bioinformatics/btad311>
33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinform* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
34. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinform* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>
35. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192. <https://doi.org/10.1093/bib/bbs017>
36. Au CW, Manfield I, Webb ME, Paci E, Turnbull WB, Ross JF (2024) The Mutagenic Plasticity of the Cholera Toxin B-Subunit Surface Residues: Stability and Affinity. *Toxins (Basel)* 16. <https://doi.org/10.3390/toxins16030133>
37. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C-C, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Židek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nat* 630:493–500. <https://doi.org/10.1038/s41586-024-07487-w>
38. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A,

- Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong SY, Finn RD (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 47:D351–d360. <https://doi.org/10.1093/nar/gky1100>
39. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31:3812–3814. <https://doi.org/10.1093/nar/gkg509>
40. Kellogg EH, Leaver-Fay A, Baker D (2011) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* 79:830–838. <https://doi.org/10.1002/prot.22921>
41. Gray TM, Arnoys EJ, Blankespoor S, Born T, Jagar R, Everman R, Plowman D, Stair A, Zhang D (1996) Destabilizing effect of proline substitutions in two helical regions of T4 lysozyme: leucine 66 to proline and leucine 91 to proline. *Protein Sci* 5:742–751. <https://doi.org/10.1002/pro.5560050419>
42. Ruvinsky AM, Vakser IA (2010) Sequence composition and environment effects on residue fluctuations in protein structures. *J Chem Phys* 133. <https://doi.org/10.1063/1.3498743>
43. Shroff NP, Bera S, Cherian-Shaw M, Abraham EC (2001) Substituted hydrophobic and hydrophilic residues at methionine-68 influence the chaperone-like function of alphaB-crystallin. *Mol Cell Biochem* 220:127–133. <https://doi.org/10.1023/a:1010834107809>
44. Sheridan E, Wright J, Small N, Corry PC, Oddie S, Whibley C, Petherick ES, Malik T, Pawson N, McKinney PA, Parslow RC (2013) Risk factors for congenital anomaly in a multiethnic birth cohort: an analysis of the Born in Bradford study. *Lancet* 382:1350–1359. [https://doi.org/10.1016/s0140-6736\(13\)61132-0](https://doi.org/10.1016/s0140-6736(13)61132-0)
45. Yoshinaga K, Yasue A, Mitsui SN, Minegishi Y, Oyadomari S, Imoto I, Tanaka E (2023) Effects of Wnt10a and Wnt10b Double Mutations on Tooth Development. *Genes (Basel)* 14. <https://doi.org/10.3390/genes14020340>
46. Pearce M, Basutkar P, Neto Renato Caminha J, Subramoniam Vijay V, Neis K, Tutis I, Hermjakob H (2025) EBI Search: providing discovery tools for biological metadata in 2025. *Nucleic Acids Res* 53:W273–W276. <https://doi.org/10.1093/nar/gkaf359>

## Tables

Tables are available in the Supplementary Files section.

## Figures

Family 4 - VI:1

Family 6 - II:2

Family 7 - II:1

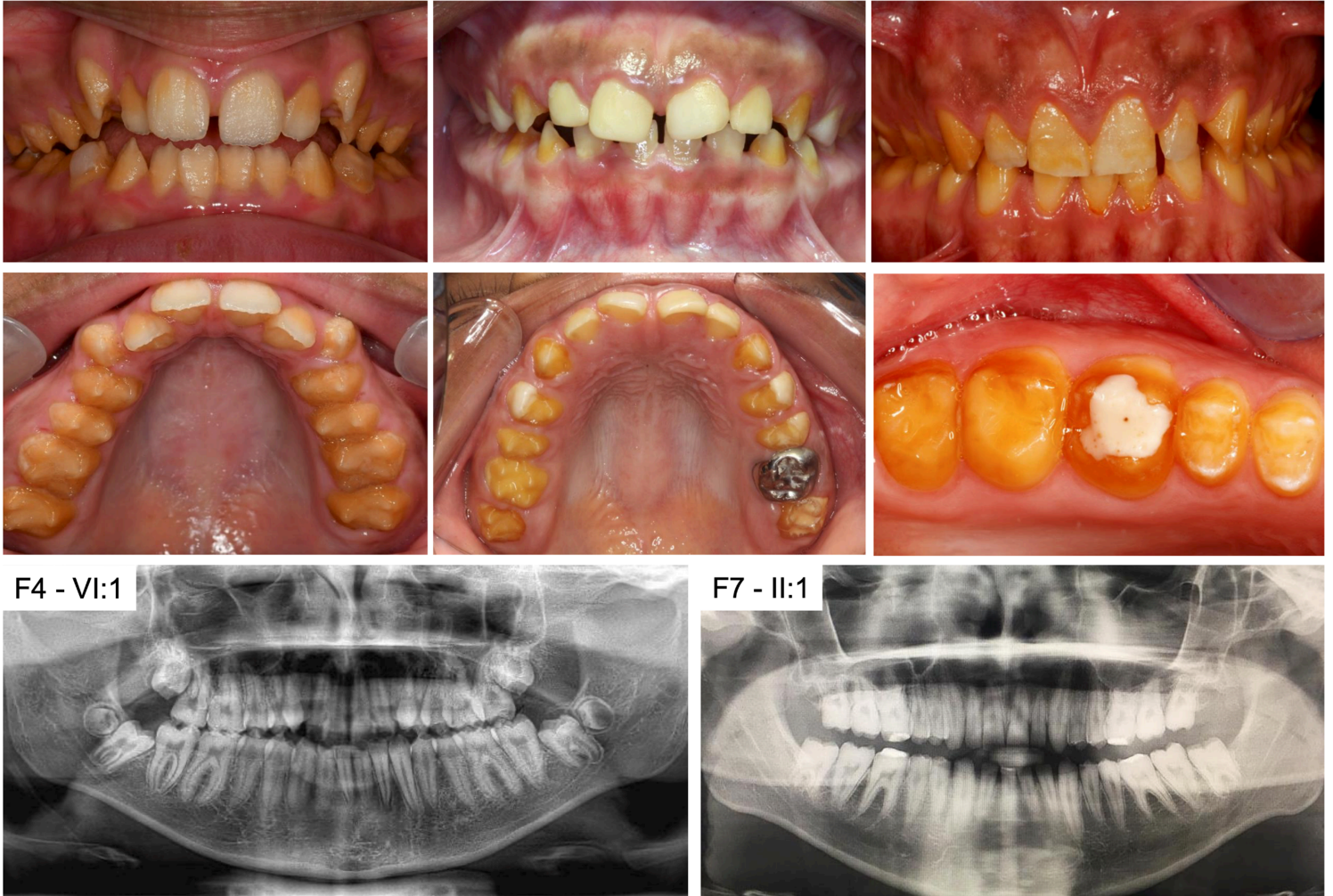
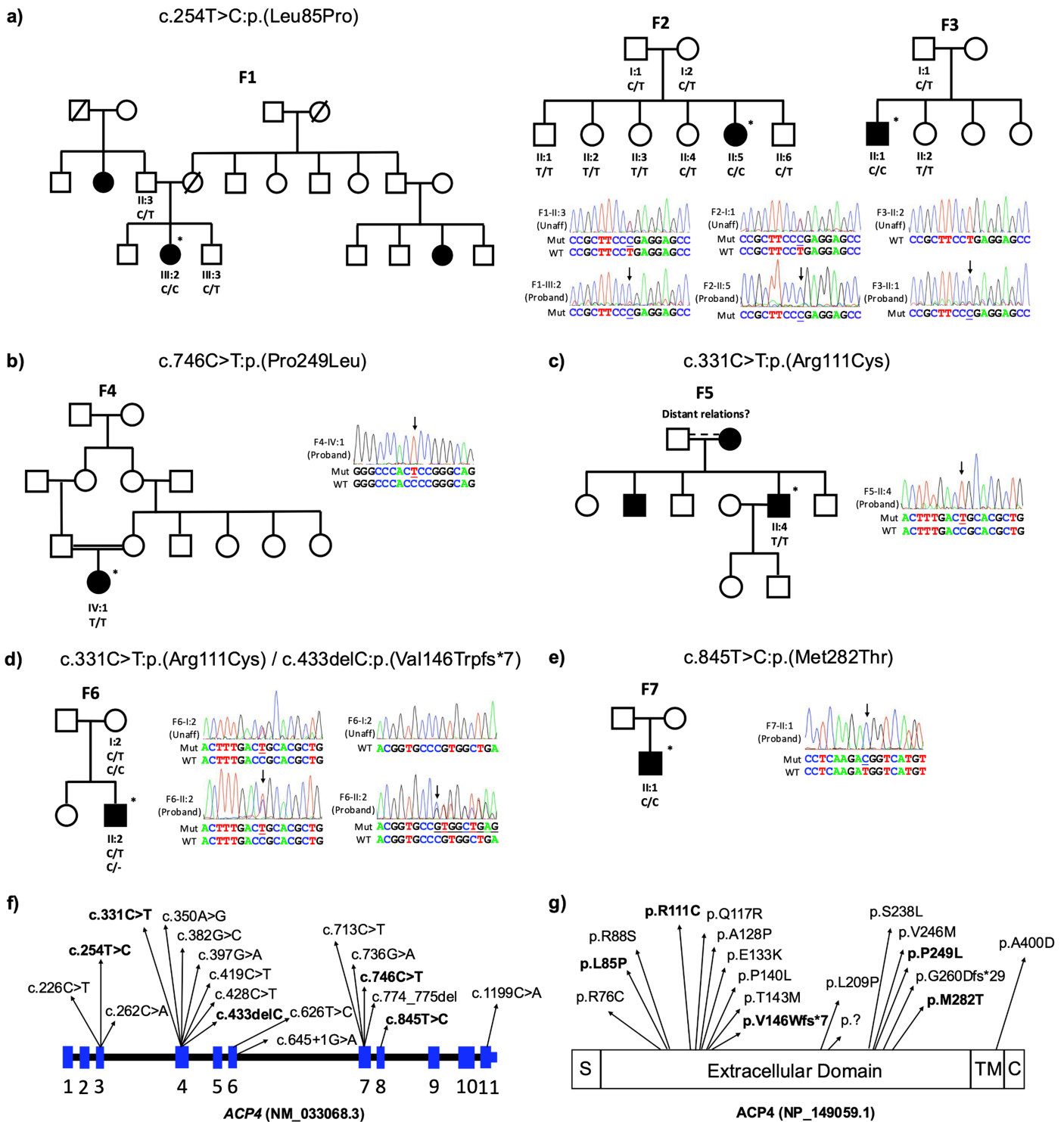


Figure 1

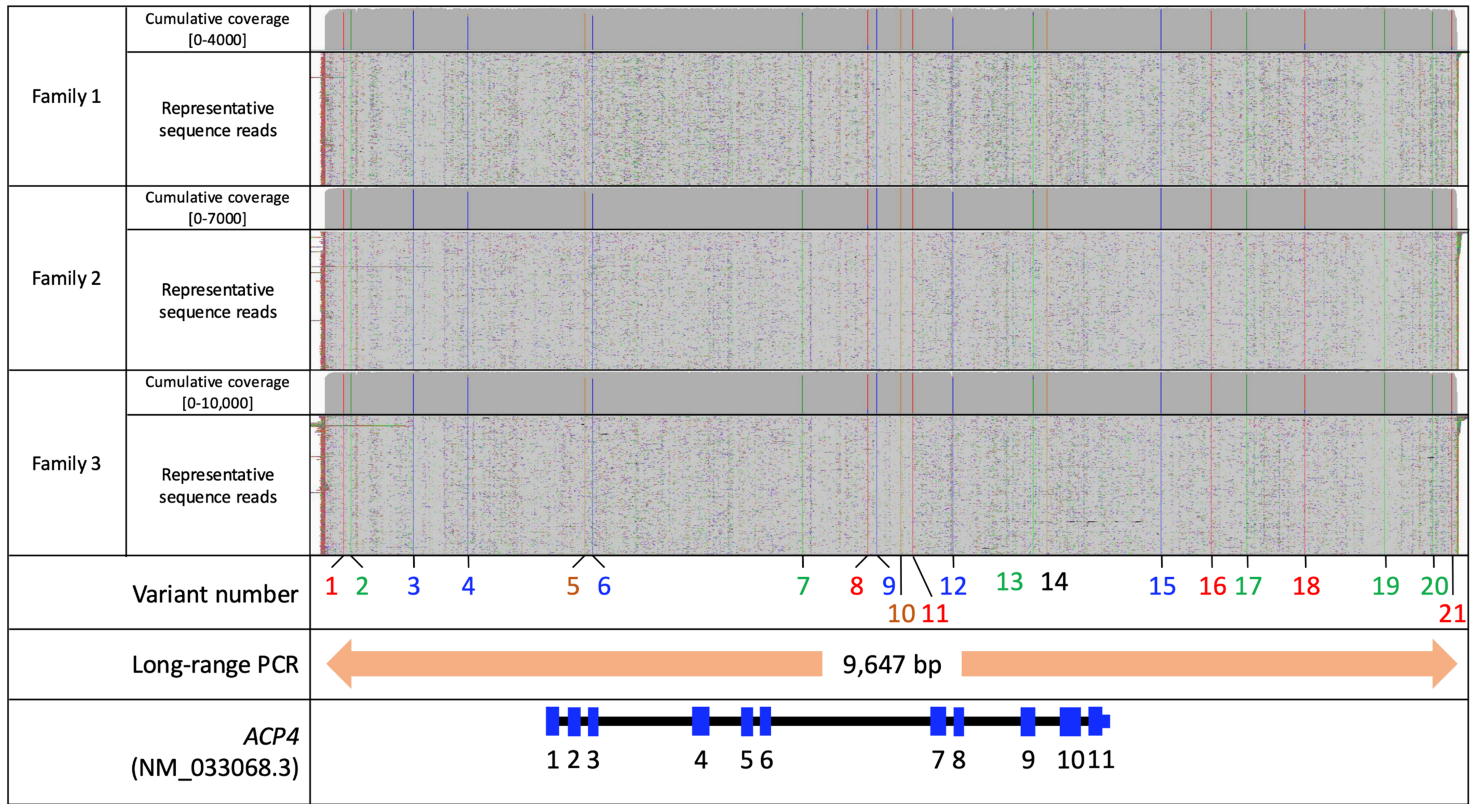
Clinical images and radiographs of teeth from affected individuals. Clinical images from probands F4-V:1, F6-II:2, and F7-II:1 show reduced quantity of enamel on molars with yellow appearance and thin enamel on anterior teeth. Radiographs from probands F4-VI:1 and F7-II:1 illustrate a thin enamel layer on permanent dentition.



**Figure 2**

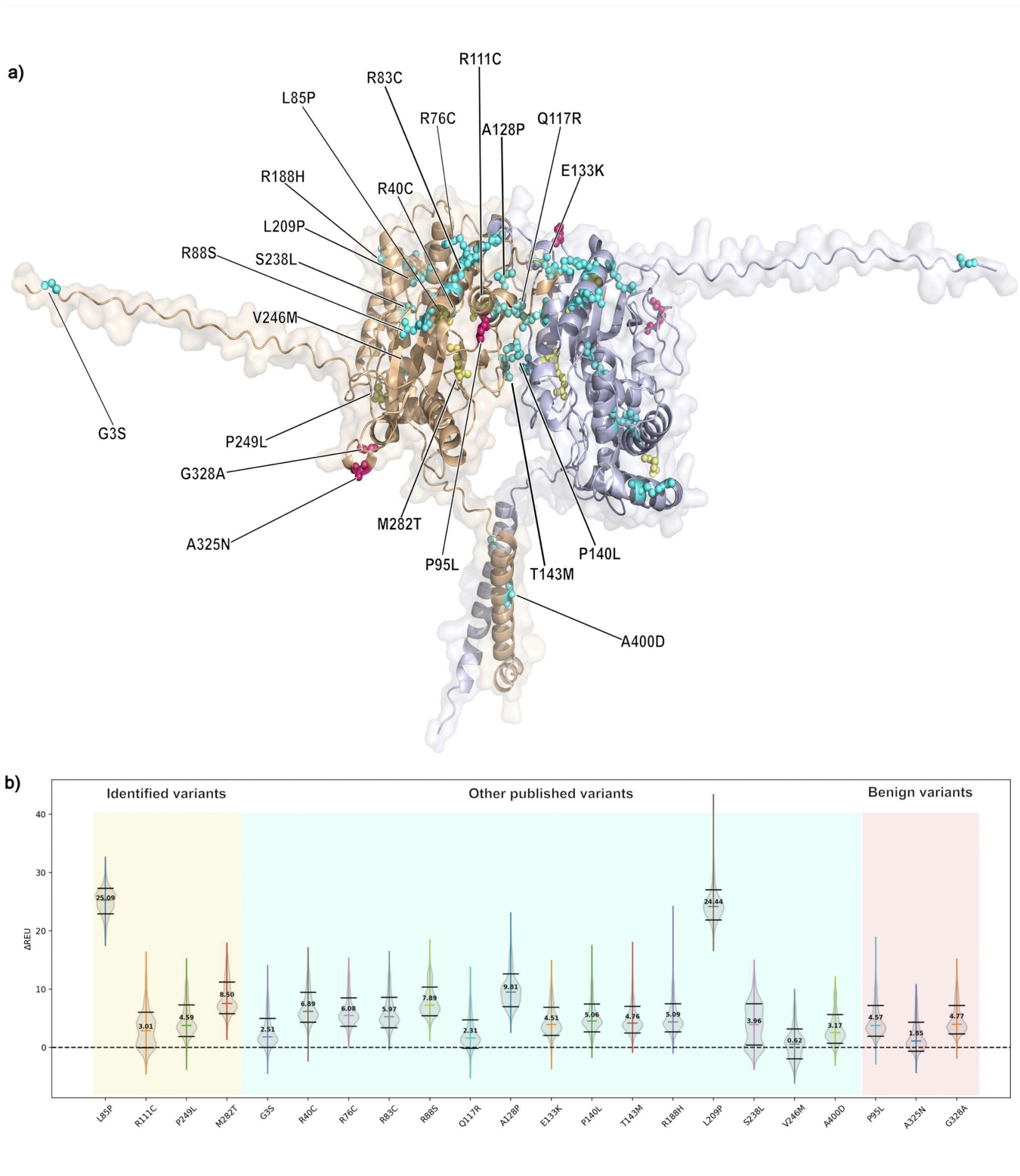
Family pedigrees, variant Sanger sequencing traces and schematic of ACP4 transcript. Family 1-3 (a), c.254T>C, p.(Leu85Pro) was homozygous in the probands and heterozygous in recruited parental samples. Family 4 (b), homozygous c.746C>T, p.(Pro249Leu) identified from consanguineous parents. Family 5 (c), homozygous c.331C>T, p.(Arg111Cys) confirmed in proband II:4. Family 6 (d), heterozygous c.331C>T, p.(Arg111Cys) confirmed in II:2 and I:2. Heterozygous c.433delC, p.(Val146Trpfs\*7) only

present in proband II:2. Family 7 (e), homozygous c.845T>C, p.(Met282Thr) was confirmed in proband II:1. (f) schematic of all published ACP4 variants and their locations (NM\_033068.3). (g) ACP4 domain structure labelled with the position of altered residues from all published variants (NP\_149059.2). S = signal peptide; TM = transmembrane domain; C = cytoplasmic domain.



**Figure 3**

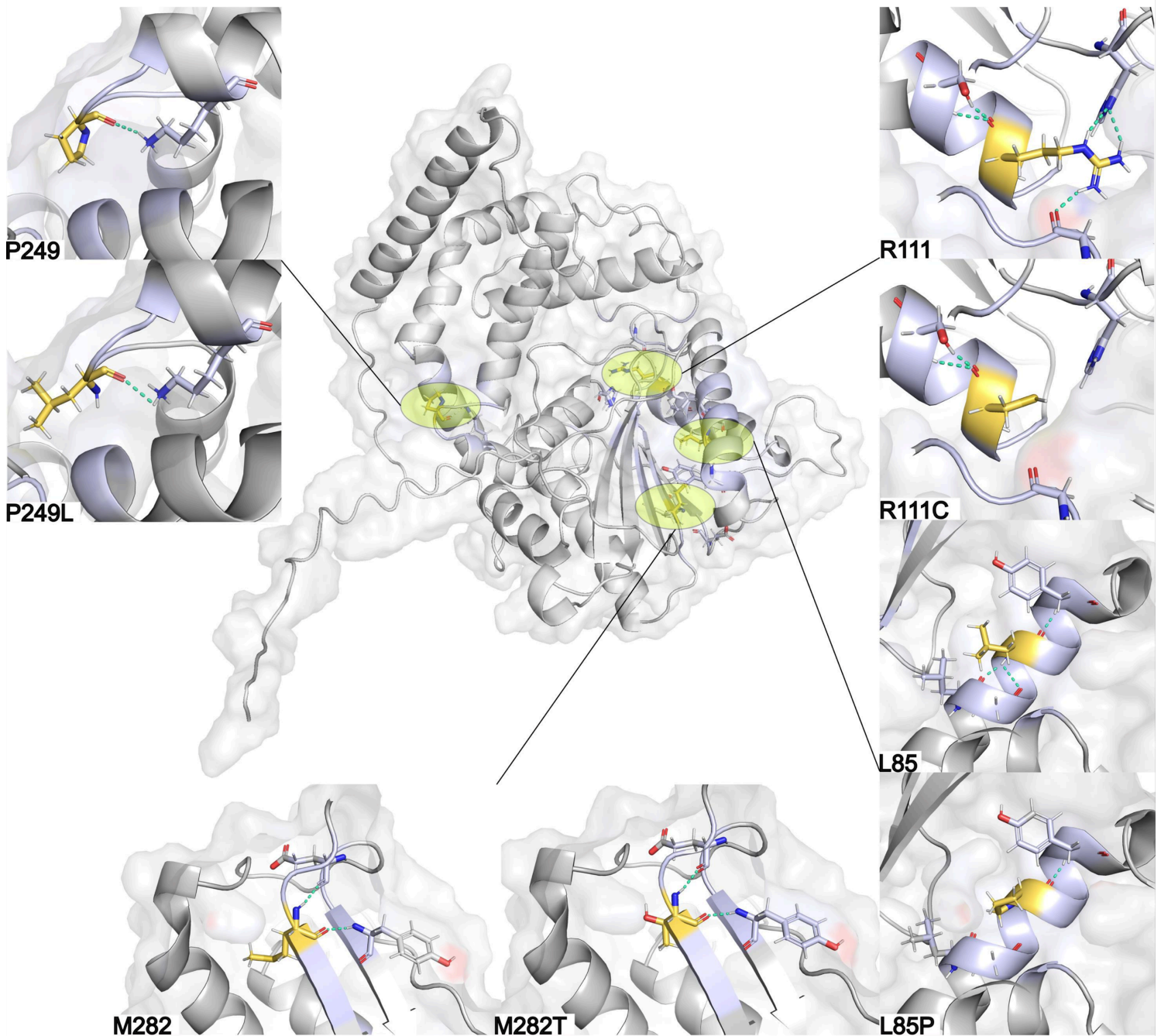
Long-read sequencing analysis of ACP4. Integrated Genomics Viewer traces of a 9.6-kb amplification product from the ACP4 locus spanning exons 1-11 for families 1-3. Nucleotide positions are reported according to human genome build hg38.



**Figure 4**

Simulated ACP4 homodimer and mapping of variants. Structural model of the ACP4 homodimer (a) highlighting ACP4 variants identified in this study (yellow), previously reported pathogenic ACP4 variants (blue) and benign ACP4 variants (red). Variants are color-coded to match the violin plot below. (b) Violin graphs of  $\Delta$ REU are shown for each variant relative to the wild type. Results are shown for all variants identified in this study, all other published ACP4 variants identified in individuals with AI and for three

benign variants. Horizontal bars are used to show the standard deviation (black), and the median  $\Delta$ REU of each variant are displayed. The dashed line at  $\Delta$ REU = 0 represents the wild type REU.



**Figure 5**

Simulated ACP4 structure and predicted effects of variants. The locations of the four missense variants identified in this study (labelled L85P, R111C, P249L, M282T) are highlighted (yellow) within the AlphaFold3-generated prediction of ACP4 monomer structure. The local environments of wildtype and mutants are displayed using insets (top and bottom, respectively) and side chains are depicted using sticks. Dashed lines indicate hydrogen bonds and polar contacts.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ACP4Supplement.docx](#)
- [Tables.docx](#)