Please cite this article in press as: Parry et al., Mutations in *C4orf26*, Encoding a Peptide with In Vitro Hydroxyapatite Crystal Nucleation and Growth Activity, Cause Am..., The American Journal of Human Genetics (2012), http://dx.doi.org/10.1016/j.ajhg.2012.07.020

REPORT

Mutations in *C4orf26*, Encoding a Peptide with In Vitro Hydroxyapatite Crystal Nucleation and Growth Activity, Cause Amelogenesis Imperfecta

David A. Parry,[1] Steven J. Brookes,[2] Clare V. Logan,[1] James A. Poulter,[1] Walid El-Sayed,[1,2,3] Suhaila Al-Bahlani,[4] Sharifa Al Harasi,[5] Jihad Sayed,[2] El Mostafa Raïf,[2] Roger C. Shore,[2] Mayssoon Dashash,[6] Martin Barron,[7] Joanne E. Morgan,[1] Ian M. Carr,[1] Graham R. Taylor,[1] Colin A. Johnson,[1] Michael J. Aldred,[8] Michael J. Dixon,[7] J. Tim Wright,[9] Jennifer Kirkham,[2] Chris F. Inglehearn,[1] and Alan J. Mighell[1,2,*]

Autozygosity mapping and clonal sequencing of an Omani family identified mutations in the uncharacterized gene, *C4orf26,* as a cause of recessive hypomineralized amelogenesis imperfecta (AI), a disease in which the formation of tooth enamel fails. Screening of a panel of 57 autosomal-recessive AI-affected families identified eight further families with loss-of-function mutations in *C4orf26*. *C4orf26* encodes a putative extracellular matrix acidic phosphoprotein expressed in the enamel organ. A mineral nucleation assay showed that the protein's phosphorylated C terminus has the capacity to promote nucleation of hydroxyapatite, suggesting a possible function in enamel mineralization during amelogenesis.

Mature enamel consists almost exclusively of highly organized calcium hydroxyapatite ($Ca_{10}[PO_4]_6[OH]_2$) (HA) crystals that give teeth their key functional properties. Enamel formation (amelogenesis) involves biomineralization of an ephemeral protein extracellular matrix secreted by ameloblasts. Failure of normal amelogenesis presents clinically as amelogenesis imperfecta (AI [MIM 104530]), a genetically and phenotypically heterogeneous group of inherited conditions that have a prevalence ranging from 1/700 to 1/14,000[1,2] and that can have a profoundly negative impact on affected individuals and their families.[3] Classification into hypoplastic or hypomineralized AI distinguishes forms with either diminished enamel volume or near-normal enamel volumes and poor mineralization leading to abnormally soft enamel, respectively. Mixed hypomineralized and hypoplastic phenotypes can occur, although posteruption changes can confound phenotype interpretation.

Understanding of the molecular pathways involved in amelogenesis has informed the identification of genes mutated in nonsyndromic AI; these genes include *AMELX*[4] (MIM 300391) and *ENAM*[5] (MIM 606585), which encode enamel matrix proteins, and *KLK4*[6] (MIM 603767) and *MMP20*[7] (MIM 604629), encoding proteases that digest enamel-matrix proteins. However, the etiology of a large proportion of AI cases still remains to be discovered.[8,9] Furthermore, recent genetic studies have identified mutations in uncharacterized genes with no clear functional roles,[10–16] reflecting our limited understanding of key events in amelogenesis. Identification of additional components of amelogenesis will provide insights into biomineralization and will further facilitate the development of biomimetic materials with applications for enamel repair and skeletal-tissue engineering.[17]

We identified a consanguineous Omani family (AI-46) affected by autosomal-recessive hypomineralized AI (Figure 1). Affected enamel was yellowish brown and hypomineralized with partial, mild developmental enamel hypoplasia that was prone to rapid functional failure with posteruptive volume loss (Figure 2). DNA samples were obtained from Oragene DNA sample collection kits (DNA Genotek, Ontario, Canada). With informed consent, the study was performed according to the principles of the declaration of Helsinki after ethics approval was obtained in the countries where the families were living.

DNA from affected individuals IV:9 and IV:10 and an equimolar mixture of DNA from affected individuals IV:12, IV:13, and IV:15 (Figure 1A) were subjected to Affymetrix 6.0 SNP microarray analysis. Analysis with AutoSNPa[18] identified a shared region of homozygosity (Figure 1B) on chromosome 4 between SNPs rs7679122 and rs17030727 (chr4: 69,820,283–101,963,249). Microsatellite genotyping confirmed that all affected individuals shared a homozygous haplotype that was not shared by unaffected members of the family. This region encompassed several genes, including *ENAM*, implicated in enamel biomineralization; *ENAM* variants are associated with hypoplastic forms of AI. Sanger sequencing of exonic

[1]Leeds Institute of Molecular Medicine, St. James's University Hospital, University of Leeds, LS9 7TF Leeds, UK; [2]Leeds Dental Institute, University of Leeds, LS2 9LU Leeds, UK; [3]Oral Biology, Dental School, Suez Canal University, Ismailia, Egypt; [4]Al-Nahda Hospital, P.O. Box 937, Muscat, P.C. 111, Sultanate of Oman; [5]Military Dental Centre, P.O. Box 454, Seeb, P.C. 121, Sultanate of Oman; [6]Department of Paediatric Dentistry, Damascus University, Syria; [7]Faculty of Medical & Human Sciences and Faculty of Life Sciences, Manchester Academic Health Sciences Centre Michael Smith Building, University of Manchester, Oxford Road, Manchester, M13 9PT, UK; [8]Dorevitch Pathology, 18 Banksia Street, Heidelberg, Victoria 3084, Australia; [9]Department of Pediatric Dentistry, School of Dentistry, University of North Carolina, Chapel Hill, NC 27599, USA
*Correspondence: a.j.mighell@leeds.ac.uk
http://dx.doi.org/10.1016/j.ajhg.2012.07.020.

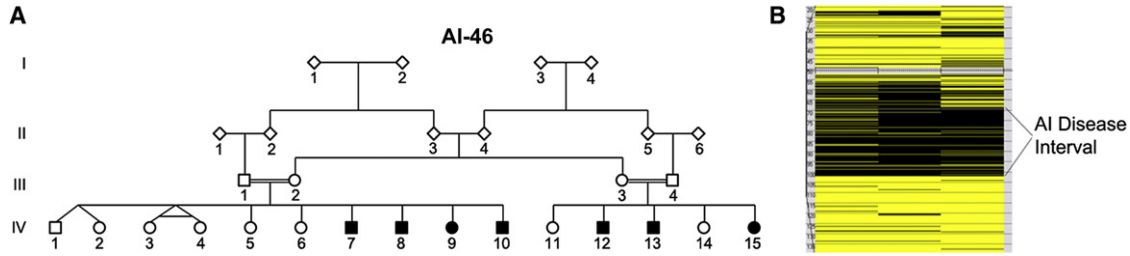

Figure 1. Identification of a Recessive AI Locus on Chromosome 4q
(A) Pedigree of family AI-46.
(B) AutoSNPa[18] output for chromosome 4 for individuals IV:9 and IV:10 and an equimolar combination of DNA from IV:12, IV:13, and IV:15. Heterozygous SNPs are highlighted in yellow, homozygous SNPs are shown in black, and chromosome position (Mb) is shown on the left. The disease interval is marked between SNPs rs7679122 and rs17030727 (chr4: 69,820,283–101,963,249).

regions and intron-exon boundaries of *ENAM* (MIM 606585), *AMBN* (MIM 601259), *AMTN* (MIM 610912), and *ODAM* did not reveal any novel or potentially pathogenic variation.

We designed a custom SureSelect Target Enrichment reagent (Agilent Technologies, Edinburgh UK) targeting coding exons within the disease interval in parallel with the capture of disease intervals for seven other unrelated disorders. Upon the merging of overlapping exons, a total of 1,693 coding intervals were identified in the region comprising 27.4 kb. We were able to target a total of 1,684 intervals while avoiding designing baits with homology to repeat masked sequence, leaving only 847 bp of coding sequence not targeted by the reagent. Three micrograms of genomic DNA was processed according to the Agilent SureSelect Library Prep protocol version 1.0.1 (October 2009). In brief, DNA was sheared and ligated to Illumina paired-end adapters before size selection (200–300 bp) and PCR amplification. The amplified library was hybridized for 24 hr at 65°C with the SureSelect reagent, and captured DNA was selected with streptavidin-coated magnetic beads and purified. A posthybridization amplification step was performed before clean up with Ampure SPRI beads (Beckman). The library was hybridized to a single-read flow cell (Illumina, California, USA) and sequenced for 80 cycles on an Illumina GAIIx with an adapted single-read protocol.

Raw data files were processed by the Illumina pipeline (version 1.3.4), and sequencing reads were aligned to the human reference sequence (hg19) with Novoalign software (Novocraft Technologies, Selangor, Malaysia). The resulting alignment was processed in the SAM/BAM format[19] with Picard and the Genome Analysis Toolkit (GATK)[20,21] java programs for correcting alignments around indel sites and marking potential PCR duplicates. After postprocessing and duplicate removal, a mean depth of 112 reads

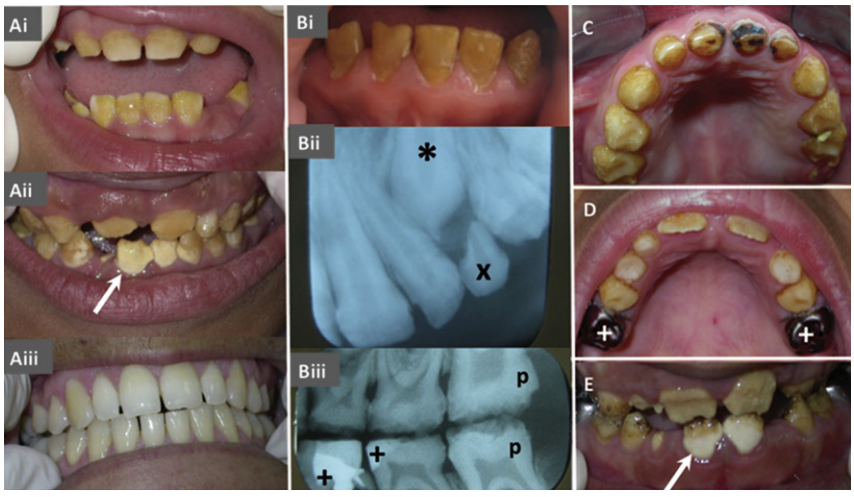

Figure 2. Clinical Dental Phenotypes Due to *C4orf26* Mutations
(A) Family AI-46 from Oman. The appearance is within normal limits for the heterozygous (carrier) mother (A.iii) (see III:2 in Figure S1) of the affected individuals (A.i and A.ii) (see IV:13 and IV:12, respectively, in Figure S1). By contrast, all the other clinical images are from individuals with homozygous mutations. The affected enamel is yellowish brown and hypomineralized with partial, mild developmental enamel hypoplasia. The enamel is prone to rapid functional failure with posteruptive volume loss. Dental plaque deposits (→) can be extensive, reflecting difficulties in maintaining good oral hygiene as the teeth fail and become sensitive.
(B) Individual II:4 (see Figure S1) from Iraqi family AI-1. A dental radiograph (B.ii) illustrates a near-normal crown morphology for the unerupted permanent upper canine tooth (*), which is in the process of replacing the almost exfoliated deciduous canine (x). In the permanent canine there is absence of the expected contrast between enamel and dentine radiodensity. These appearances are consistent with enamel hypomineralization. Bitewing radiographs (B.iii) illustrate poor-quality enamel and premature enamel failure involving the occlusal surfaces of permanent (p) and deciduous teeth. Dental restorations (+) are marked.
(C) Individual IV:2 (see Figure S1) from Syrian family AI-42.
(D) Individual IV:1 (see Figure S1) from Omani family AI-47.
(E) Individual II:2 (see Figure S1) from Omani family AI-51.

with a minimum base-phred quality score of 17 and minimum read-mapping quality of 20 was achieved for target regions; 97% of target bases were covered by greater than 5 such reads.

Variants within the candidate regions were called in the variant-call format with the Unified Genotyper and DINDEL[22] functions of GATK. A total of 191 single-nucleotide and indel variants in the target regions were identified in or within 20 bp of coding exons in the target region. Variants present in dbSNP129 or with a minor allele frequency of 1% or higher were removed, as were variants present in two or more of the seven unrelated samples sequenced alongside the proband.

Upon filtering of common variation, only two potentially pathogenic variants remained. The first, a homozygous missense change (c.95C>A [p.Pro32His]; RefSeq accession number NM_000670.3) in *ADH4* (MIM 103740), was subsequently shown to be heterozygous in affected individual IV:8 by Sanger sequencing, suggesting a modest refinement of the candidate region. The remaining variant was a homozygous nonsense change (c.229C>T [p.Arg77*]; RefSeq NM_178497.3) in the uncharacterized *C4orf26* and segregated with AI in the AI-46 Omani family and was absent from 160 ethnically-matched and 78 Pakistani control DNA samples. To determine the population frequency of this variant, we extracted alignment regions for *C4orf26* from BAM files at the 1000 Genomes Project[23] and called genotypes. Genotypes were confidently called (genotype phred quality score $\geq$ 30) for 1,135 individuals who were all homozygous for the reference nucleotide.

We sequenced *C4orf26* coding exons and flanking intronic sequence in DNA from 57 apparently unrelated probands who had recessive or sporadic AI and for whom mutations in other recessive AI-associated genes had been excluded (Figure 3). The same homozygous nonsense variant (c.229C>T [p.Arg77*]) as identified in family AI46 was also identified in three further Omani families (AI-47, AI-51, and AI-67 [Figure S1, available online]) and was shown to segregate with hypomineralized AI. We found an Iraqi family (AI-1 [Figure S1]) in which a different homozygous nonsense change (c.129C>A [p.Cys43*]) segregated with AI. Mutation of the splice acceptor site of exon 2 (c.68-2A>T [p.?]) was identified and shown to segregate with AI in a Syrian family (AI-42 [Figure S1]). Neither of these changes could be found in 110 Turkish, 50 Jordanian, or 78 Pakistani control DNA samples. The former change (c.129C>A [p.Cys43*]) was found to be homozygous for the reference base in 1,131 samples (from the 1000 Genomes Project) for which the genotype could be confidently called. Similarly, the splice-acceptor-site mutation was excluded in all 1,135 individuals (from the 1000 Genomes Project) for whom the genotype could be confidently called.

Affected individuals from two families (AI-102 and AI-103 [Figure S1]) from the United States were found to be compound heterozygotes for a nonsense change (c.318G>A [p.Trp106*]) and an indel creating a frameshift and premature termination (c.51_56delGGTAACins ATGCTGGTTACTGGTA [p.Val18Cysfs*23]). Affected individuals from another family from the United States (AI-108 [Figure S1]) were found to be homozygous for the same indel (c.51_56delGGTAACinsATGCTGGTTACTG GTA [p.Val18Cysfs*23]). Both the c.318G>A variant and the c.51_56delGGTAACinsATGCTGGTTACTGGTA variant were absent from 91 control individuals of European descent. Upon the calling of 1000 Genomes sample genotypes for the nonsense mutation (c.318G>A [p.Trp106*]), all 1,133 sample genotypes with a minimum genotype phred quality score of 30 were homozygous for the reference base. This variant has been identified in the heterozygous state in 2 of 10,758 chromosomes sequenced by the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (Exome Variant Server, NHLBI Exome Sequencing Project, Seattle, WA) and has been given the dbSNP identifier rs146645381. The indel variant was not identified in 1,197 individuals for whom a genotype could be called (these individuals include 1,139 samples with a minimum genotype phred quality score of 30), although it should be noted that a complex indel such as this might be considered unlikely to align and be detected with clonal sequencing methods.

In total, we identified nine families affected by five unique mutations in *C4orf26*. In each case, the mutations segregated as would be expected for recessively inherited AI and appear to represent rare variation. The identification of nine families affected by *C4orf26* mutations indicates discovery of a potential major cause of recessive AI. We note that individuals with 4q syndrome sometimes present with enamel defects. However, because individuals with heterozygous *C4orf26* mutations in our study do not have a defect in enamel mineralization, we consider it unlikely that deletion of one copy of the gene would result in an enamel deficiency without mutation of the remaining allele. With the possible exception of a reported case with del(4)(q21.1q21.3),[24] haploinsufficiency of either *ENAM* or *DSPP* (MIM 125485), which are mutated in dominantly inherited conditions, appears more likely to be responsible for enamel defects observed in individuals with deletions of the long arm of chromosome 4.

After the identification of mutations in *C4orf26* (RefSeq NM_178497.3), a further RefSeq transcript (NM_001206981.1) that includes an additional exon resulting in a frameshift was reported. The two encoded proteins share only the first 22 residues and bear no similarity throughout the length of the remaining sequences. Putative protein orthologs for this second transcript could only be identified in *Pan troglodytes* and *Nomascus leucogenys*. Although the splice-site mutation (c.68-2A>T) and indel mutation (c.51_58delGGTAACTGinsATGCTGGTTA CTGGTATG) identified in AI-affected families are likely to result in loss of function or truncation of this additional gene product, the remaining truncating mutations

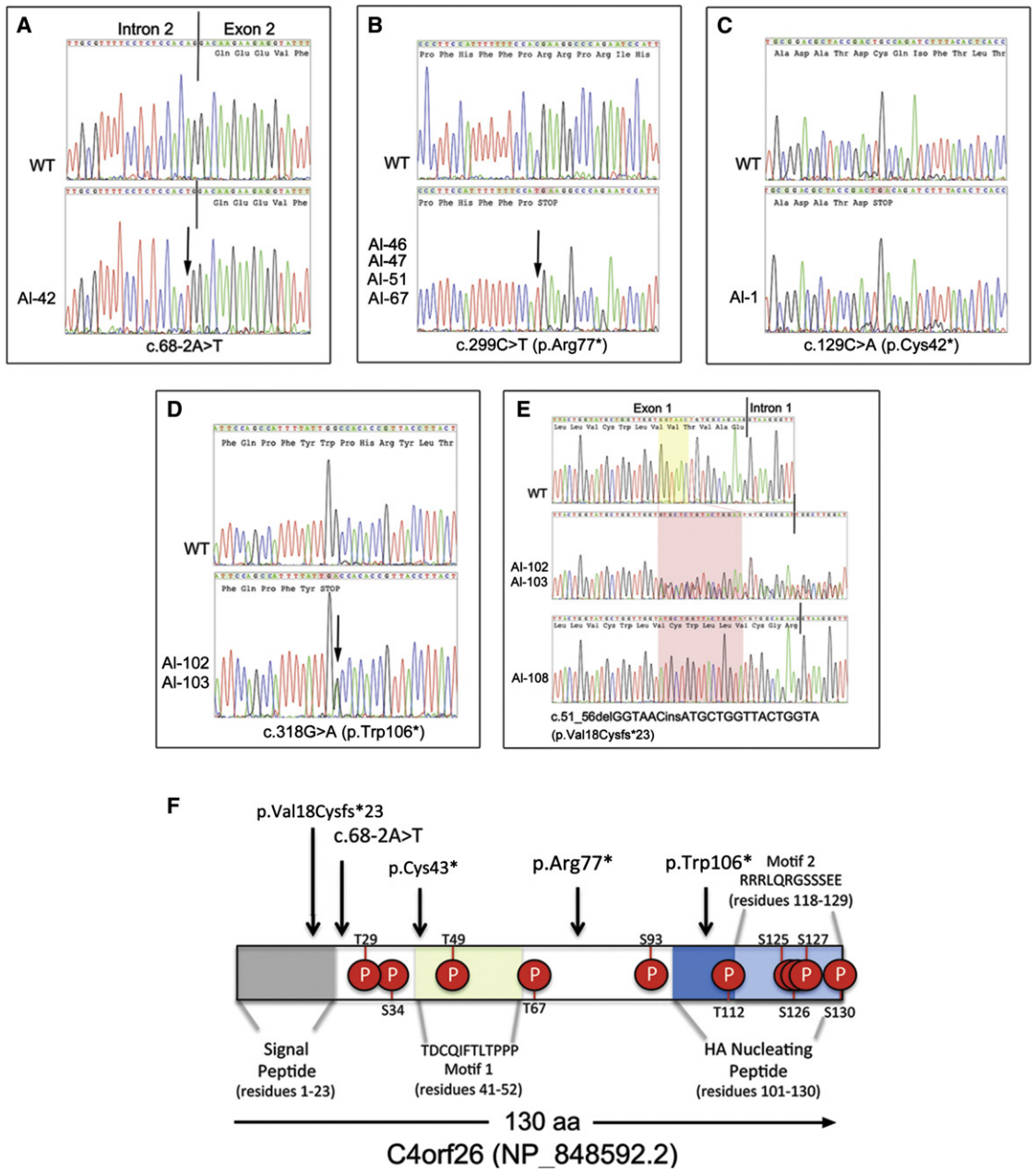

Figure 3. *C4orf26* mutations
(A–E) Representative electropherograms for *C4orf26* mutations uncovered in AI-affected families.
(F) Schematic representation of human C4orf26 protein. Sequence changes found in AI-affected families are marked by vertical arrows. Predicted phosphorylated sites are shown as red circles. Shading indicates the predicted signal peptide (gray), two conserved motifs (green and light blue), and the C-terminal peptide (dark and light blue) used in HA-nucleation experiments.

identified in NM_178497.3 result in synonymous or missense variation in NM_001206981.1. It is thus unclear whether this latter transcript plays any role in amelogenesis or has any functional similarity to NM_178497.3.

To confirm expression of *C4orf26* during amelogenesis, we screened cDNA obtained from a continually erupting rat incisor enamel organ by using RT-PCR. The rat ortholog was identified by comparison of the alignment of the mouse ortholog (Gm1045) with the rat genome (Baylor 3.4) and splice-site prediction with NetGene2.[25] The predicted gene was subsequently found to correspond with CONTRAST[26] prediction CONTRAST.chr14.175. The rat ortholog was clearly expressed in both secretory- and maturation-stage enamel organs (Figure 4A and data not shown) and was not expressed in the rat heart, kidney, or spleen. In a panel of 20 human tissues, not including the enamel organ, expression was most obvious in the placenta (Figure S2). Family AI-46 is large with no history of premature placental failure.

C4orf26 encodes a 130 amino acid protein with a predicted signal peptide (residues 1–23) but no other recognizable domains or motifs. Protein-protein BLAST results

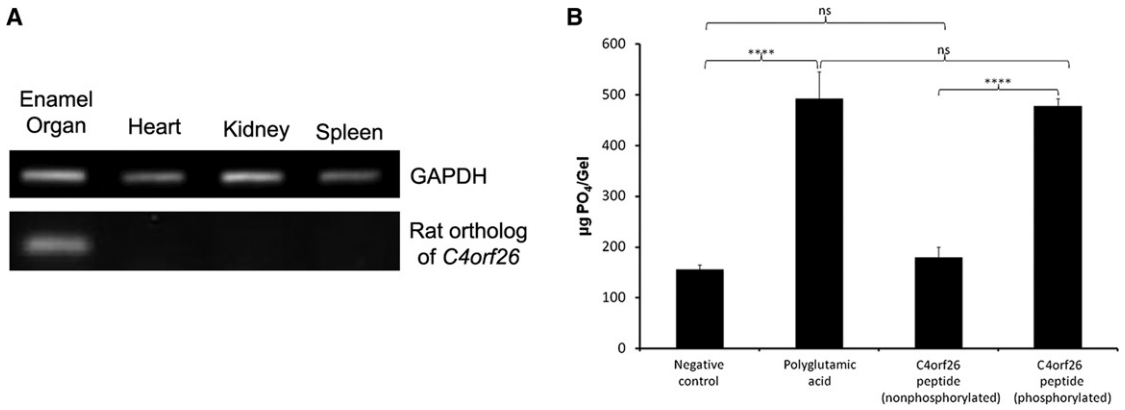

Figure 4. Expression and Functional Characterization of C4orf26
(A) RT-PCR demonstrating expression of the rat ortholog of *C4orf26* in the enamel organ; band identity is confirmed by sequencing. (B) Representative in vitro mineralization data showing the mass of phosphate precipitated with no peptide (negative control), the polyglutamic-acid positive control, and nonphosphorylated and phosphorylated peptides corresponding to the C-terminal 30 residues of C4orf26. Data (n = 6) were analyzed with an unpaired one-tailed Student's t test. ****$p < 0.0001$. The following abbreviation is used: ns, not significant.

indicate that C4orf26 is conserved in mammals. Two motifs, TDCQ[IV][FL]TLTPPP (residue 41–52 of human NP_848592.2) and R[RG][RQNK]xx[RSG][GD]SSSEE (residue 118–130 of NP_848592.2) are strongly conserved (Figure S3).

Sequence analysis with NetPhos 2.0[27] predicts C4orf26 to be an acidic phosphoprotein containing ten possible phosphorylation sites resulting in a predicted isoelectric point of 6.6. Prediction with NetPhos 2.0 suggests that phosphorylation of the cluster of three C-terminal serine residues is conserved in mammalian orthologs. Acidic phosphoproteins promote HA nucleation[28] across a diverse range of biomineralized tissues, yet at high concentrations, they inhibit crystal growth.[29] It is also notable that C4orf26 has a high proline content of 14.6%, which is reminiscent of other known enamel-matrix proteins: ENAM, AMBN, ODAM, AMTN, and AMELX.

We investigated the mineral-nucleation potential of synthetic peptides corresponding to the predicted phosphorylated C terminus of C4orf26 (residues 101–130) by using a steady-state in vitro nucleation assay carried out in modified equilibrium dialysis cells as described previously.[30] Phosphorylated C4orf26 (Cys-Phe-Gln-Pro-Phe-Tyr-Trp-Pro-His-Arg-Tyr-Leu-(p)Thr-Tyr-Arg-Tyr-Phe-Pro-Arg-Arg-Arg-Leu-Gln-Arg-Gly-(p)Ser-(p)Ser-(p)Ser-Glu-Glu-Ser) and the corresponding nonphosphorylated peptide (Eurogentec, Southampton, UK) were dissolved at 7 μM (~30 μg/ml) in 1% agarose solution containing 20 mM HEPES and 150 mM NaCl (pH 7.4) at 37°C. Polyglutamic acid (Sigma, Poole, UK) dissolved in the same buffered agarose at a concentration of 10 μg/ml was used as a positive control, and buffered agarose containing no peptide was used as negative control. The peptide/agarose solutions were poured into the central cavities of the modified dialysis cells and were allowed to set at room temperature. Dialysis cells were connected in parallel, and a peristaltic pump was used for pumping calcium-containing buffer (20 mM HEPES, 150 mM NaCl, and 6.5 mM $Ca(NO_3).4H_2O$ at pH 7.4) through the dialysis cell reservoirs in contact with one side of the agarose plugs, whereas phosphate-containing buffer (20 mM HEPES, 150 mM NaCl, and 3.9 mM $Na_2HPO_4.2H_2O$ at pH 7.4) was pumped in contra flow direction through the adjacent reservoirs in contact with the other side of the agarose plugs. The buffers were pumped continuously at a flow rate of 1 ml/hr/cell at 37°C for 7 days. During this period, calcium and phosphate ions diffuse from opposite ends of the agarose plugs and mineral precipitates in the gels to a degree dependent on the nucleating potential of the peptides present. We quantified mineral precipitation in the plugs by determining phosphate concentration in each plug by using the phosphomolybdate method.[30] These experiments showed that phosphorylated peptide was able to promote HA nucleation and support crystal growth, whereas the nonphosphorylated peptide showed no such activity (Figure 4B).

Posteruptive enamel loss is rapid in AI as a result of *C4orf26* mutations, and there are obvious abnormalities on scanning electron microscopy.[31] Transmission electron microscopy (Figure S5) reveals thin crystallites reminiscent of those normally found in the secretory and early-maturation stages of enamel formation prior to final maturation, where crystallites grow in width and thickness until the tissue volume is occluded. The presence of crystallites in *C4orf26*-linked AI suggests that HA nucleation is not the key role of C4orf26 in amelogenesis but that it does influence crystallite growth. C4orf26's predicted export peptide signal and similarities with known enamel-matrix extracellular proteins might reflect roles in the control of crystallite growth within the developing enamel matrix.

In conclusion, although the precise role of C4orf26 will require further experimental exploration, the discovery of mutations as a major cause of AI and the ability of a phosphorylated peptide to promote HA nucleation and crystal

growth in vitro offer new insights into biomineralization. This is relevant not only to this poorly understood group of clinical conditions that have a profoundly negative impact on affected individuals but also for the development of biomimetic materials with wide-ranging applications.

Supplemental Data

Supplemental Data include five figures and three tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

The authors thank the families for participating in this study. This work was supported by grants from The Wellcome Trust (to A.J.M., J.K., M.D., M.B., C.I., and S.B.) and the Sir Jules Thorn Award for Biomedical Research (to C.A.J, G.R.T., and C.I.).

Received: April 20, 2012
Revised: June 19, 2012
Accepted: July 18, 2012
Published online: August 16, 2012

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes Project, http://www.1000genomes.org/
Agilent eArray, https://earray.chem.agilent.com/earray/
AutoSNPa, http://dna.leeds.ac.uk/autosnpa/
CONTRAST, http://contra.stanford.edu/contrast/
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
Exome Variant Server, NHLBI Exome Sequencing Project, http://evs.gs.washington.edu/EVS/
NetGene2, http://www.cbs.dtu.dk/services/NetGene2/
UCSC Genome Browser, http://genome.ucsc.edu/

References

1. Witkop, C.J., and Sauk, J.J. (1976). Heritable defects of enamel. In Oral Facial Genetics, G. Prescott and R. Stewart, eds. (St. Louis: CV Mosby Company), pp. 151–226.
2. Bäckman, B., and Holm, A.K. (1986). Amelogenesis imperfecta: Prevalence and incidence in a northern Swedish county. Community Dent. Oral Epidemiol. *14*, 43–47.
3. Coffield, K.D., Phillips, C., Brady, M., Roberts, M.W., Strauss, R.P., and Wright, J.T. (2005). The psychosocial impact of developmental dental defects in people with hereditary amelogenesis imperfecta. J. Am. Dent. Assoc. *136*, 620–630.
4. Lagerström, M., Dahl, N., Nakahori, Y., Nakagome, Y., Bäckman, B., Landegren, U., and Pettersson, U. (1991). A deletion in the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1). Genomics *10*, 971–975.
5. Rajpar, M.H., Harley, K., Laing, C., Davies, R.M., and Dixon, M.J. (2001). Mutation of the gene encoding the enamel-specific protein, enamelin, causes autosomal-dominant amelogenesis imperfecta. Hum. Mol. Genet. *10*, 1673–1677.
6. Hart, P.S., Hart, T.C., Michalec, M.D., Ryu, O.H., Simmons, D., Hong, S., and Wright, J.T. (2004). Mutation in kallikrein 4 causes autosomal recessive hypomaturation amelogenesis imperfecta. J. Med. Genet. *41*, 545–549.
7. Kim, J.W., Simmer, J.P., Hart, T.C., Hart, P.S., Ramaswami, M.D., Bartlett, J.D., and Hu, J.C. (2005). MMP-20 mutation in autosomal recessive pigmented hypomaturation amelogenesis imperfecta. J. Med. Genet. *42*, 271–275.
8. Kim, J.W., Simmer, J.P., Lin, B.P., Seymen, F., Bartlett, J.D., and Hu, J.C. (2006). Mutational analysis of candidate genes in 24 amelogenesis imperfecta families. Eur. J. Oral Sci. *114* (*Suppl 1*), 3–12, discussion 39–41, 379.
9. Wright, J.T., Torain, M., Long, K., Seow, K., Crawford, P., Aldred, M.J., Hart, P.S., and Hart, T.C. (2011). Amelogenesis imperfecta: Genotype-phenotype studies in 71 families. Cells Tissues Organs (Print) *194*, 279–283.
10. El-Sayed, W., Parry, D.A., Shore, R.C., Ahmed, M., Jafri, H., Rashid, Y., Al-Bahlani, S., Al Harasi, S., Kirkham, J., Inglehearn, C.F., and Mighell, A.J. (2009). Mutations in the beta propeller WDR72 cause autosomal-recessive hypomaturation amelogenesis imperfecta. Am. J. Hum. Genet. *85*, 699–705.
11. O'Sullivan, J., Bitu, C.C., Daly, S.B., Urquhart, J.E., Barron, M.J., Bhaskar, S.S., Martelli-Júnior, H., dos Santos Neto, P.E., Mansilla, M.A., Murray, J.C., et al. (2011). Whole-Exome sequencing identifies FAM20A mutations as a cause of amelogenesis imperfecta and gingival hyperplasia syndrome. Am. J. Hum. Genet. *88*, 616–620.
12. Kim, J.W., Lee, S.K., Lee, Z.H., Park, J.C., Lee, K.E., Lee, M.H., Park, J.T., Seo, B.M., Hu, J.C., and Simmer, J.P. (2008). FAM83H mutations in families with autosomal-dominant hypocalcified amelogenesis imperfecta. Am. J. Hum. Genet. *82*, 489–494.
13. Parry, D.A., Mighell, A.J., El-Sayed, W., Shore, R.C., Jalili, I.K., Dollfus, H., Bloch-Zupan, A., Carlos, R., Carr, I.M., Downey, L.M., et al. (2009). Mutations in CNNM4 cause Jalili syndrome, consisting of autosomal-recessive cone-rod dystrophy and amelogenesis imperfecta. Am. J. Hum. Genet. *84*, 266–273.
14. Polok, B., Escher, P., Ambresin, A., Chouery, E., Bolay, S., Meunier, I., Nan, F., Hamel, C., Munier, F.L., Thilo, B., et al. (2009). Mutations in CNNM4 cause recessive cone-rod dystrophy with amelogenesis imperfecta. Am. J. Hum. Genet. *84*, 259–265.
15. Mory, A., Dagan, E., Illi, B., Duquesnoy, P., Mordechai, S., Shahor, I., Romani, S., Hawash-Moustafa, N., Mandel, H., Valente, E.M., et al. (2012). A nonsense mutation in the human homolog of Drosophila rogdi causes Kohlschutter-Tonz syndrome. Am. J. Hum. Genet. *90*, 708–714.
16. Schossig, A., Wolf, N.I., Fischer, C., Fischer, M., Stocker, G., Pabinger, S., Dander, A., Steiner, B., Tönz, O., Kotzot, D., et al. (2012). Mutations in ROGDI Cause Kohlschütter-Tönz Syndrome. Am. J. Hum. Genet. *90*, 701–707.
17. Kirkham, J., Firth, A., Vernals, D., Boden, N., Robinson, C., Shore, R.C., Brookes, S.J., and Aggeli, A. (2007). Self-assembling peptide scaffolds promote enamel remineralization. J. Dent. Res. *86*, 426–430.
18. Carr, I.M., Flintoff, K.J., Taylor, G.R., Markham, A.F., and Bonthron, D.T. (2006). Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. Hum. Mutat. *27*, 1041–1046.
19. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
20. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly,

M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

21. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.
22. Albers, C.A., Lunter, G., MacArthur, D.G., McVean, G., Ouwehand, W.H., and Durbin, R. (2011). Dindel: accurate indel calls from short-read data. Genome Res. *21*, 961–973.
23. 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.
24. Strehle, E.M., and Middlemiss, P.M. (2007). Children with 4q-syndrome: The parents' perspective. Genet. Couns. *18*, 189–199.
25. Brunak, S., Engelbrecht, J., and Knudsen, S. (1991). Prediction of human mRNA donor and acceptor sites from the DNA sequence. J. Mol. Biol. *220*, 49–65.
26. Gross, S.S., Do, C.B., Sirota, M., and Batzoglou, S. (2007). CONTRAST: A discriminative, phylogeny-free approach to multiple informant de novo gene prediction. Genome Biol. *8*, R269.
27. Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. *294*, 1351–1362.
28. Gajjeraman, S., Narayanan, K., Hao, J., Qin, C., and George, A. (2007). Matrix macromolecules in hard tissues control the nucleation and hierarchical assembly of hydroxyapatite. J. Biol. Chem. *282*, 1193–1204.
29. Boskey, A.L., Maresca, M., Doty, S., Sabsay, B., and Veis, A. (1990). Concentration-dependent effects of dentin phosphophoryn in the regulation of in vitro hydroxyapatite formation and growth. Bone Miner. *11*, 55–65.
30. Hunter, G.K., and Goldberg, H.A. (1993). Nucleation of hydroxyapatite by bone sialoprotein. Proc. Natl. Acad. Sci. USA *90*, 8562–8565.
31. Wright, J.T. (1985). Analysis of a kindred with amelogenesis imperfecta. J. Oral Pathol. *14*, 366–374.