SUPPLEMENTAL MATERIAL

Materials and methods

Animal models. DLX3$^{K14\text{-}cKO}$ mice were generated as described previously (17). *KRT75*^{tm1Der}KI mice were provided by Dr. Jiang Chen and Dr. Dennis Roop (30). All animal work was approved by the NIAMS Animal Care and Use Committee.

Microarray and RNA-Seq analyses. Mandibles were dissected from P10 mice, transferred to RNAlater® solution (Life Technologies) and stored at 4C. Enamel organs were dissected from mandibles and homogenized in Trizol® reagent (Invitrogen) using the Precellys® 24 (Bertin Technologies). Total RNA was extracted and further purified using the RNAeasy mini kit (Qiagen). Microarray analysis was performed as described previously (33) using three WT and three cKO animals. RNA-seq analysis was performed using the Mondrian SP kit (Illumina) and the Illumina HiSeq 2000 sequencing system. Fold-changes were calculated using Partek Genomics Suite (http://www.partek.com).

Immunostaining. Mouse mandibles were dissected at P2, embedded in OCT and frozen. 10 µm-thick sections were prepared using a Leica CM 1850 cryostat (Leica). Human teeth were ground, polished and etched for 10 seconds in 37% Phosphoric acid to expose cross-sectioned enamel rods. Enamel tufts were isolated after complete demineralization of human molars in 0.1M EDTA. Immunohistochemical analysis was performed on tissue sections, polished teeth and isolated tufts using a blocking solution containing 5% goat serum and 7.5% BlockHen II (Aves Labs) in 1X PBS. Primary antibodies used: Guinea-pig anti-KRT75 antibody (LifeSpan BioSciences Inc.), rabbit anti-K14 antibody

(Covance). Secondary antibodies used: Alexa 555 conjugated goat anti-guinea-pig antibody and Alexa 488 conjugated goat anti-rabbit antibody (Life technologies). Image acquisition was performed on a Leica SP5 NLO confocal microscope (Leica).

Scanning electron microscopy analysis of enamel rod sections. Ground, polished and etched human teeth were prepared for SEM as described previously (33) and analyzed under a Field Emission Scanning Electron Microscope S4800 (Hitashi) at 10 kV.

Transmission electron microscopy analysis of isolated enamel tufts and rod sheaths. Enamel tufts and rod sheaths were isolated from decalcified human teeth, fixed for 48 hours at 4^0C in 2% glutaraldehyde in 0.1M cacodylate buffer (pH 7.4) and washed with cacodylate buffer three times. The tissues were fixed with 2% OsO_4 for two hours, washed again with 0.1M cacodylate buffer three times, washed with water and placed in 1% uranyl acetate for one hour. The specimens were subsequently serially dehydrated in ethanol and propylene oxide and embedded in EMBed 812 resin (Electron Microscopy Sciences). Thin sections, approx. 80 nm, were obtained by utilizing the Leica ultracut-UCT ultramicrotome (Leica) and placed onto 300 mesh copper grids and stained with saturated uranyl acetate in 50% methanol and then with lead citrate. The grids were viewed in the JEM-1200EXII electron microscope (JEOL Ltd) at 80kV and images were recorded on the XR611M, mid mounted, 10.5Mpixel, CCD camera (Advanced Microscopy Techniques Corp).

Human subject recruitment and data collection. The Center for Oral Health Research in Appalachia (COHRA) study was initiated to investigate the community-, family-, and individual-level contributors to oral health (34). Participants from rural counties of

Pennsylvania and West Virginia were enrolled via a household-based recruitment strategy, with eligible households required to include at least one biological parent-child pair. All other members of eligible households were invited to participate without regard to biological or legal relationships, or oral health status. Intra-oral examinations of all participants were performed by licensed dentists and/or research dental hygienists. Each surface of each tooth (excluding third molars) was examined for evidence of decay, from which dental caries indices were generated. Three measurements of caries experience were considered: (1) the number of surfaces with untreated decay (DS/ds); (2) the traditional DMFS/dfs indices which represent the number of decayed (D/d), missing due to decay (M), and filled (F/f) tooth surfaces (S/s) in the permanent (DMFS) and primary (dfs) dentitions; and (3) the partial DMFS and dfs indices limited to the molars and premolar pit and fissure surfaces which are at high risk of decay. DNA samples were collected via blood, saliva or buccal swab. Genotyping for approximately 600,000 polymorphisms was performed by the Center for Inherited Disease Research at Johns Hopkins University using the Illumina Human610-Quadv1_B BeadChip (Illumina). Extensive data cleaning and quality assurance analyses were performed as previously reported (35). Imputation of unobserved polymorphisms, including rs2232387 and rs2232398 in *KRT75*, was performed using the 1000 Genomes Project reference sample. Written informed consent was provided by all adult participants, and verbal assent with parental written consent was provided by all child participants. All procedures, forms and protocols were approved by the Institutional Review Boards of the University of Pittsburgh and West Virginia University.

Statistical analysis for human data. Linear regression was used to test the association of dental caries experience with genetic polymorphisms under the additive genetic model while adjusting for age and sex. Analyses were preformed separately for dental caries in the permanent dentition of adults (ages 25-50 years) and the primary dentition of children (ages 6-12 years). All analyses were limited to self-reported non-Hispanic whites (mixed European descent); self-reported race was consistent with genetically-determined ancestry. Analyses were performed in the R statistical environment (R Foundation for Statistical Computing).

Human teeth collection and sequencing of the KRT75 gene. Teeth were collected from over 200 patients undergoing teeth extraction in the Oral and Maxillofacial Surgery unit at the Walter Reed National Military Medical Center (WRNMMC) in Bethesda, MD. Basic information collected from the patients included age, sex, race and hair type. Teeth were received on the day of surgery. Soft tissue was collected from the periodontal ligament area and used for extraction of genomic DNA before fixing specimens in 4% paraformaldehyde in 1X PBS for three days at 4°C. PCR and automated DNA sequencing were used to amplify and sequence the region of the *KRT75* gene surrounding the position A161, as described previously (23): (Forward) 5'-TCAGTGGCCCCAGCTTCCCCGTGTGTC-3' and (Reverse) 5'-TGTTCCTCAGTTCAGCTTCAAGCCTG-3'. All procedures, forms, and protocols were approved by Institutional Review Boards of the NIAMS and WRNMMC.

Microhardness testing and Scanning Electron Microscopy (SEM) analysis. Molars were sectioned in the mesial-distal plane using an Isomet slow speed saw (Buehler). 2 mm sections from the central portion of a crown containing mesial and distal groves were

initially polished by hand using 1200 grit Carbmet abrasive paper (Buehler) mounted on a model 900 grinder/polisher (Electron Microscopy Sciences). The fine polishing for the SEM analysis was carried out using Metadi diamond suspensions of 6, 1 and 0.25 μm and Texmet and Microcloth polishing cloth on an Isomet 1000 polisher (Buehler). Samples were stored in Hanks solution at 4°C prior to the microindentation tests. The microindentations were carried out using IdentaMet 1105 microhardness tester with a CCD camera and OmniMet software (Buehler). The tests were conducted using Vickers diamond tip at 0.1 kgf (0.98N) load and 5 sec dwell time. Tests were conducted in the occlusal inner, middle and outer enamel. Five replicate indentations per location were carried out in each sample, with the distance between individual indentations no less than 5 indent diagonals from each other. The average microhardness values for each location was determined and the data were statistically analyzed using linear regression to model simultaneously the effects of race and *KRT75* allele dosage under the additive genetic model. Prior to SEM analysis, teeth sections were air-dried mounted on aluminum stubs and carbon coated. The sections were studied using Jeol 6335F scanning electron microscope (Jeol) in a backscattered electron mode (BS SEM) at 10KV and a working distance of 7 mm.

Micro-CT analysis and 3-D reconstructions. Micro-CT analysis of fixed molars was performed as described previously (33) using the Skyscan 1172 desktop X-ray microfocus CT scanner and the following parameters: 0.5mm AI + 0.1mm Cu filters, 100 kV, 100 uA, 8.00 micron resolution, 0.4 degrees rotation step over 180 degrees). CTan and CTvox software (Bruker microCT) were used for data analysis, measurements and

3D reconstruction. Molars from twenty patients (9 with $KRT75^{GG}$, 6 with $KRT75^{GA}$ and 5 with $KRT75^{AA}$) were analyzed.

Gram staining on human tooth sections. Sections of fixed (4% PFA) molars were stained for bacteria using a stabilized gram stain set (Fisher Scientific).

33. Duverger O, et al. Neural Crest Deletion of Dlx3 Leads to Major Dentin Defects through Down-regulation of Dspp. *J Biol Chem*. 2012; 287(15):12230-12240.
34. Polk DE, et al. Study protocol of the Center for Oral Health Research in Appalachia (COHRA) etiology study. *BMC Oral Health*. 2008; 8:18.
35. Shaffer JR, et al. Genome-wide association scan for childhood caries implicates novel genes. *J Dent Res*. 2011; 90(12):1457-1462.

Supplemental figures

Supplemental Figure 1: Molecular consequences of Dlx3 deletion in the enamel organ as determined by RNA-seq analysis
A) Effects of Dlx3 deletion on the expression of keratins in mouse enamel organ at P10, as determined by microarray analysis. **B)** Effects of Dlx3 deletion on the expression of keratins in mouse enamel organ at P10, as determined by RNA-seq analysis. Consistent with microarray analysis performed on independently isolated tissues at the same stage, no difference in *Krt14* expression was observed while *Krt25*, *Krt27* and *Krt75* were significantly reduced (see Figure 1A). In addition, RNA-seq analysis revealed an increase in *Krt17* and *Krt80* expression and a decrease in *Krt71* and *Krt76* expression, even though these changes did not pass the correction for multiple hypothesis test (q-value >0.05). **C)** Genome browser alignment of the RNA-seq data presented in A.

Supplemental Figure 2: Detection of KRT75 in mouse enamel organ and effect of *Krt75* mutation on enamel prism assembly in *Krt75^{tm1Der}* mice
A) Detection of KRT14 and KRT75 in mouse ameloblasts at the secretory stage. DM, dental mesenchyme; Am, ameloblasts; SI, stratum intermedium. Scale bar: 50 μm. **B)** Alignment of the human and mouse sequences of the KRT75 protein showing the proximity between the mutation in *Krt75^{tm1Der}* mice and the human polymorphism that was shown to be strongly associated with pseudofolliculitis barbae. *Krt75^{tm1Der}* mice were generated by deletion of an asparagine at position 159 (N159del) in the *Krt75* gene. The human and mouse mutations are two amino acids apart and located in a highly conserved region of the 1A α-helical segment of KRT75 protein. **C)** Genotyping of WT and *Krt75^{tm1Der}* mice by PCR. **D)** Appearance of hair shaft from WT and *Krt75^{tm1Der}* mice. *Krt75^{tm1Der}* mice exhibit a high proportion of puffy hair when compared to WT littermates. **E)** Scanning electron microscopy analysis of broken incisors showing defects in enamel prism structure in Krt75^{tm1Der} mice when compared to WT littermates. Scale bar: 5 μm.

Supplemental Figure 3: *KRT75* genotype-effect and race-effect on enamel hardness
Vickers enamel hardness measured on molar sections from individuals with *KRT75^{GG}*, *KRT75^{GA}* or *KRT75^{AA}* genotype. For each genotype, three teeth from patients of mixed European descent and three teeth from African Americans were analyzed. As presented in Figure 3D, inner enamel hardness was significantly affected by the *KRT75* genotype (GG>GA>AA). Also, significant effect of the race was observed for inner enamel, African Americans exhibiting significantly lower hardness independently of the genotype. Genotype and race had no significant effects when comparing enamel hardness in middle enamel and outer enamel.

Supplemental Figure 4: Gram staining of bacteria on human tooth sections
Sections of fixed molars from participants carrying the *KRT75* missense A allele (rs2232387) stained with a Stabilized Gram Stain Set. Dark purple staining marks Gram-positive bacteria located in the holes present in the enamel. Staining could also be seen in lesions that reached the dentin (central panel). E, enamel; D, dentin.

Supplemental tables:

Supplemental Table 1: Genetic association of SNP rs2232387 with dental caries in permanent dentition of adults (25-50 years) and primary dentition of children (6-12 years).

Supplemental Table 2: Genetic association of SNP rs2232398 with dental caries in permanent dentition of adults (25-50 years) and primary dentition of children (6-12 years).

Supplemental Table 3: Genetic association plus interaction of SNPs rs2232387 and rs2232398 with dental caries in permanent dentition of adults (25-50 years) and primary dentition of children (6-12 years).

Supplemental Table 4: Multiple regression model of Vicker's hardness in 18 participants (9 of mixed European descent and 9 African Americans)

Supplemental videos:

Video 1: 3D reconstruction of a molar from a patient with $KRT75^{GG}$ genotype. Enamel is colored in blue while dentin is colored in red based on density. A thin layer at the surface of the enamel appears red due to the transition between air density and enamel density. The video shows a 360° rotation of the crown followed by a progressive planar clipping of the crown through its entire thickness.

Video 2: Same as Video 1 performed on a molar from a patient with $KRT75^{GA}$ genotype. Invagination of peripheral red staining inside the enamel reflects the presence of tubular projections from the pit and fissure area of the tooth deep into the enamel and all the way to the dentin.

Video 3: Same as Video 2 performed on a molar from a patient with $KRT75^{AA}$ genotype.

Video 4: 3D reconstruction of a molar from a patient with $KRT75^{GG}$ genotype. Enamel is transparent while dentin is colored in red based on density. A thin layer at the surface of the enamel appears red due to the transition between air density and enamel density. The video shows a 360° rotation of the crown followed by a cylinder clipping of the center of the crown and a 360° rotation of the clipped tooth.

Video 5: Same as Video 4 performed on a molar from a patient with $KRT75^{GA}$ genotype. The clipped area reveals the 3D distribution of the air tubules projecting into the enamel.

Video 6: Same as Video 5 performed on a molar from a patient with $KRT75^{AA}$ genotype.

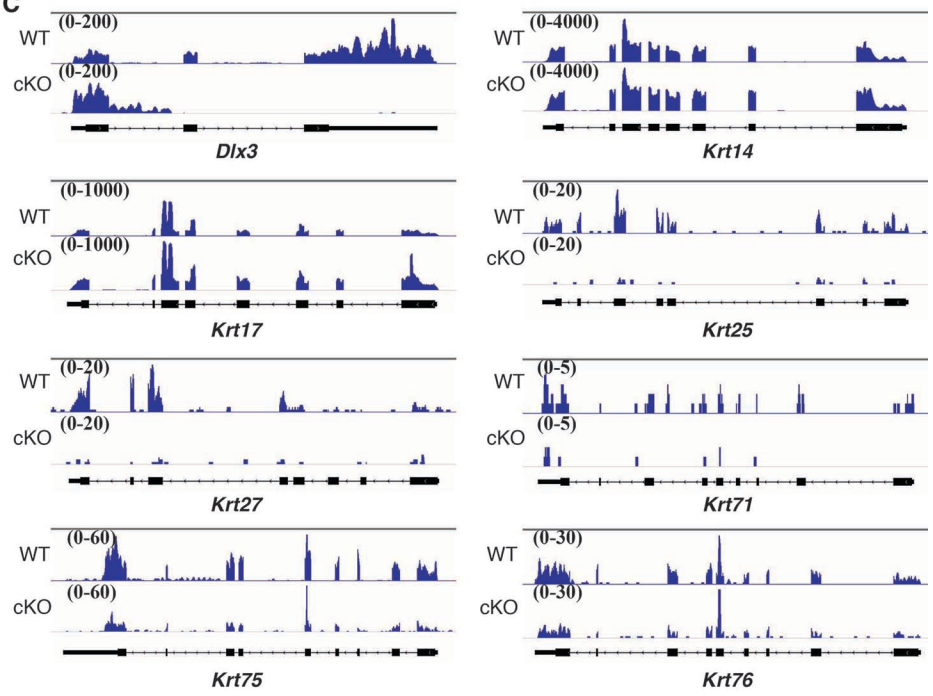
A **Microarray analysis**

Gene Symbol	RefSeq Transcript ID	p-value (cKO vs. WT)	Fold-Change (cKO vs. WT)
Dlx3	NM_010055	2.6E-05	-47.81
Krt25	NM_133730	0.0011	-10.66
Krt27	NM_010666	0.0005	-9.71
Krt75	NM_133357	0.0087	-5.13
Krt5	NM_027011	0.2377	-1.33
Krt14	NM_016958	0.5050	-1.11

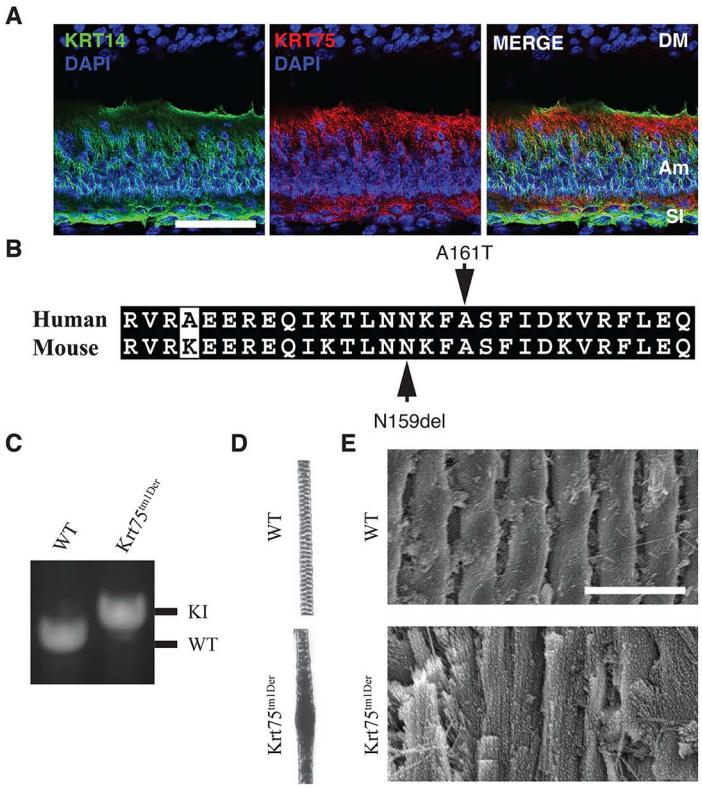
B **RNA-seq analysis**

Gene Symbol	Transcript	p-value (cKO vs. WT)	qvalue (p-value(cKO vs. WT))	Fold-Change (cKO vs. WT)
Dlx3	NM_010055	2.7E-06	0.0019	-3.88
Krt14	NM_016958	0.5799	0.5731	1.07
Krt17	NM_010663	0.0020	0.0575	2.76
Krt25	NM_133730	1.3E-05	0.0052	-11.36
Krt27	NM_010666	0.0006	0.0361	-6.68
Krt71	NM_019956	0.0015	0.0508	-6.01
Krt75	NM_133357	0.0005	0.0360	-2.90
Krt76	NM_001033177	0.0376	0.2005	-2.34
Krt80	NM_028770	0.0109	0.1217	2.28

C

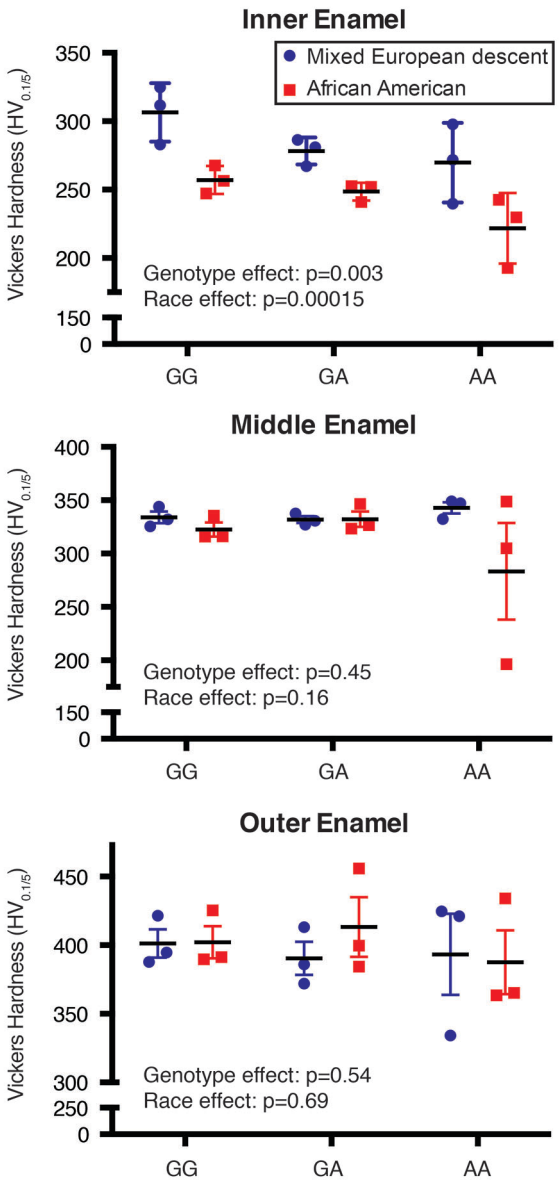


Supplemental Figure 1:
Molecular consequences of Dlx3 deletion in the enamel organ as determined by RNA-seq analysis
A) Effects of Dlx3 deletion on the expression of keratins in mouse enamel organ at P10, as determined by microarray analysis. **B)** Effects of Dlx3 deletion on the expression of keratins in mouse enamel organ at P10, as determined by RNA-seq analysis. Consistent with microarray analysis performed on independently isolated tissues at the same stage, no difference in *Krt14* expression was observed while *Krt25*, *Krt27* and *Krt75* were significantly reduced (see Figure 1A). In addition, RNA-seq analysis revealed an increase in *Krt17* and *Krt80* expression and a decrease in *Krt71* and *Krt76* expression, even though these changes did not pass the correction for multiple hypothesis test (q-value >0.05). **C)** Genome browser alignment of the RNA-seq data presented in A.

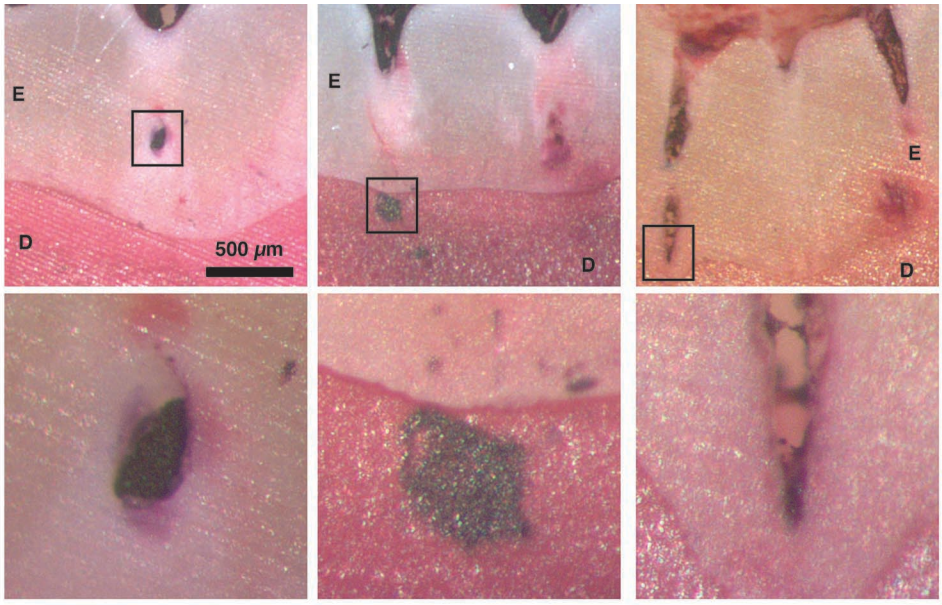


Supplemental Figure 2: Detection of KRT75 in mouse enamel organ and effect of *Krt75* mutation on enamel prism assembly in *Krt75*^{tm1Der}KI mice

A) Detection of KRT14 and KRT75 in mouse ameloblasts at the secretory stage. DM, dental mesenchyme; Am, ameloblasts; SI, stratum intermedium. Scale bar: 50 µm. **B)** Alignment of the human and mouse sequences of the KRT75 protein showing the proximity between the mutation in *KRT75*^{tm1Der}KI mice and the human polymorphism that was shown to be strongly associated with pseudofolliculitis barbae. Krt75tm1DerKI mice were generated by deletion of an asparagine at position 159 (N159del) in the *Krt75* gene. The human and mouse mutations are two amino acids apart and located in a highly conserved region of the 1A α-helical segment of KRT75 protein. **C)** Genotyping of WT and KRT75tm1DerKI mice by PCR. **D)** Appearance of hair shaft from WT and KRT75tm1DerKI mice. KRT75tm1DerKI mice exhibit a high proportion of puffy hair when compared to WT littermates. **E)** Scanning electron microscopy analysis of broken incisors showing defects in enamel prism structure in KRT75tm1DerKI mice when compared to WT littermates. Scale bar: 5 µm.



Supplemental Figure 3: *KRT75* genotype-effect and race-effect on enamel hardness
Vickers enamel hardness measured on molar sections from individuals with *KRT75*^{GG}, *KRT75*^{GA} or *KRT75*^{AA} genotype. For each genotype, three teeth from patients of mixed European descent and three teeth from African Americans were analyzed. As presented in Figure 3D, inner enamel hardness was significantly affected by the *KRT75* genotype (GG>GA>AA). Also, significant effect of the race was observed for inner enamel, African Americans exhibiting significantly lower hardness independently of the genotype. Genotype and race had no significant effects when comparing enamel hardness in middle enamel and outer enamel.



Supplemental Figure 4: Gram staining of bacteria on human tooth sections
Sections of fixed molars from participants carrying the *KRT75* missense A allele (rs2232387) stained with a Stabilized Gram Stain Set. Dark purple staining marks Gram-positive bacteria located in the holes present in the enamel. Staining could also be seen in lesions that reached the dentin (central panel). E, enamel; D, dentin.

Supplementary Table 1: Genetic association of SNP rs2232387 with dental caries in permanent dentition of adults (25-50 years) and primary dentition of children (6-12 years).

dentition	phenotype	N	β estimate	SE	p-value
permanent	DS	703	2.17	0.77	0.005
	DMFS	703	3.91	1.64	0.017
	pit and fissure DMFS	706	1.34	0.48	0.006
primary	ds	386	0.05	0.28	0.848
	dfs	386	-0.52	0.70	0.463
	pit and fissure dfs	386	0.02	0.36	0.960

Note: All models are adjusted for age and sex.

Supplementary Table 2: Genetic association of SNP rs2232398 with dental caries in permanent dentition of adults (25-50 years) and primary dentition of children (6-12 years).

dentition	phenotype	N	β estimate	SE	p-value
permanent	DS	703	2.24	4.15	0.589
	DMFS	703	1.01	8.89	0.910
	pit and fissure DMFS	706	1.24	2.62	0.636
primary	ds	386	6.70	1.82	0.0003
	dfs	386	4.18	5.40	0.439
	pit and fissure dfs	386	4.73	2.17	0.030

Note: All models are adjusted for age and sex.

Supplementary Table 3: Genetic association plus interaction of SNPs rs2232387 and rs2232398 with dental caries in permanent dentition of adults (25-50 years) and primary dentition of children (6-12 years).

dentition	phenotype	N	predictor	β estimate	SE	p-value
permanent	DS	703	rs2232387	43.26	16.42	0.009
			rs2232398	0.03	4.26	0.994
			interaction	41.26	16.48	0.013
	DMFS	703	rs2232387	31.21	35.37	0.378
			rs2232398	0.14	9.18	0.988
			interaction	27.41	35.50	0.440
	pit and fissure DMFS	706	rs2232387	4.26	10.42	0.683
			rs2232398	1.38	2.70	0.611
			interaction	2.92	10.46	0.780
primary	ds	386	rs2232387	-6.99	4.92	0.156
			rs2232398	7.89	1.99	0.00009
			interaction	-7.12	4.94	0.150
	dfs	386	rs2232387	-6.18	14.59	0.672
			rs2232398	5.09	5.92	0.390
			interaction	-5.92	14.66	0.687
	pit and fissure dfs	386	rs2232387	-5.87	6.22	0.346
			rs2232398	5.59	2.35	0.018
			interaction	-5.96	6.24	0.341

Note: All models are adjusted for age and sex.

Supplementary Table 4: Multiple regression model of Vicker's hardness in 18 participants (9 of mixed European descent and 9 African Americans)

	rs2232387			race			model		
phenotype	β estimate	SE	p-value	β estimate	SE	p-value	R^2	F	p-value
inner enamel	-18.01	5.16	0.003	-42.37	8.43	0.00015	0.71	18.7	0.00008
middle enamel	-7.57	9.67	0.45	-23.52	15.79	0.16	0.16	1.42	0.27
outer enamel	-5.64	8.97	0.54	6	14.64	0.69	0.04	0.28	0.76

R^2 = proportion of phenotype variance explained by rs2232387 genotype and race combined