

ORIGINAL ARTICLE

Variants on chromosome 4q21 near *PKD2* and *SIBLING*s are associated with dental caries

Scott Eckert^{1,19}, Eleanor Feingold^{2,3}, Margaret Cooper^{4,5}, Michael M Vanyukov^{2,6,7}, Brion S Maher⁸, Rebecca L Slayton⁹, Marcia C Willing¹⁰, Steven E Reis^{11,12}, Daniel W McNeil^{13,14}, Richard J Crout¹⁵, Robert J Weyant¹⁶, Steven M Levy^{17,18}, Alexandre R Vieira^{4,5}, Mary L Marazita^{2,4,5,7,12} and John R Shaffer²

A recent genome-wide association study (GWAS) for dental caries nominated the chromosomal region 4q21 near *ABCG2*, *PKD2* and the *SIBLING* (small integrin-binding ligand *N*-linked glycoprotein) gene family. In this investigation, we followed up and fine-mapped this region using a tag-SNP (single-nucleotide polymorphism) approach in 13 age- and race-stratified samples from 6 independent studies ($N = 4089$). Participants were assessed for dental caries via intraoral examination and 49 tag-SNPs were genotyped capturing much of the variation in the 4q21 locus. Linear models were used to test for genetic association, while adjusting for sex, age and components of ancestry. SNPs in and near *PKD2* showed significant evidence of association in individual samples of black adults (rs17013735, P -value = 0.0009) and white adults (rs11938025; P -value = 0.0005; rs2725270, P -value = 0.003). Meta-analyses across black adult samples recapitulated the association with rs17013735 (P -value = 0.003), which occurs at low frequency in non-African populations, possibly explaining the race specificity of the effect. In addition to race-specific associations, we also observed evidence of gene-by-fluoride exposure interaction effects in white adults for SNP rs2725233 upstream of *PKD2* ($P = 0.002$). Our results show evidence of regional replication, though no single variant clearly accounted for the original GWAS signal. Therefore, while we interpret our results as strengthening the hypothesis that chromosome 4q21 may impact dental caries, additional work is needed.

Journal of Human Genetics (2017) 62, 491–496; doi:10.1038/jhg.2016.161; published online 19 January 2017

INTRODUCTION

Dental caries (that is, tooth decay) is among the most common chronic disease affecting both children and adults across all populations. With appropriate treatment, decay may cause little or no negative consequences. In contrast, untreated decay may lead to many negative concomitants affecting quality of life, which disproportionately impacts vulnerable populations such as racial and ethnic minorities, those living in poverty or in rural areas, young children and the elderly. The central mechanism of decay is dissolution of mineral in the tooth caused by localized changes in pH owing to metabolic by-products of cariogenic bacteria. However, whether or not decay actually occurs depends on additional factors, such as diet and consumption habits, oral ecology, behavioral factors (for example, oral hygiene), endogenous factors (for example, enamel

quality, tooth morphology and saliva flow and buffering capacity), environmental exposures (for example, fluoride and medications) and socioeconomic and societal factors (for example, access to oral health care, cultural values, policy). Of these factors, fluoride exposures may be particularly important. Heritability estimates^{1–5} indicate that host genetics has a key role in susceptibility to dental caries, and genes are hypothesized to underlie many of the aforementioned factors influencing caries, although few specific caries-related genes have been discovered.

Previous genome-wide association studies (GWASs) of dental caries have nominated several loci,^{6–10} although few of these have been replicated in follow-up studies.¹¹ One such locus is the region on chromosome 4q21 (Supplementary Figure 1) near *PKD2* and *ABCG2* and immediately downstream of the dentin/bone extracellular matrix

¹Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; ²Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; ³Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; ⁴Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; ⁵Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; ⁶Department of Pharmaceutical Sciences, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA; ⁷Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; ⁸Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; ⁹Department of Pediatric Dentistry, School of Dentistry, University of Washington, Seattle, WA, USA; ¹⁰Division of Genetics and Genomics, Department of Pediatrics, School of Medicine, Washington University at St Louis, St Louis, MO, USA; ¹¹Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; ¹²Clinical and Translational Science Institute, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; ¹³Department of Dental Practice and Rural Health, West Virginia University, Morgantown, WV, USA; ¹⁴Department of Psychology, West Virginia University, Morgantown, WV, USA; ¹⁵Department of Periodontics, School of Dentistry, West Virginia University, Morgantown, WV, USA; ¹⁶Department of Dental Public Health and Information Management, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; ¹⁷Department of Preventive and Community Dentistry, University of Iowa College of Dentistry, Iowa City, IA, USA and ¹⁸Department of Epidemiology, University of Iowa College of Public Health, Iowa City, IA, USA

¹⁹Current address: Department of Biology, The College of New Jersey, Ewing Township, NJ 08618, USA.

Correspondence: DR JR Shaffer, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, 130 De Soto Street, Pittsburgh, PA 15261, USA. E-mail: john.r.shaffer@pitt.edu

Received 26 July 2016; revised 17 November 2016; accepted 7 December 2016; published online 19 January 2017

subfamily of secretory calcium-binding phosphoprotein (SCPP) gene cluster.⁸

The known biology of genes in this region suggests plausible roles in dental caries. Most relevant, the paralogous¹² bone/dentin extracellular matrix SCPP genes (that is, *SPP1*, *MEPE*, *IBSP*, *DMP1* and *DSPP*), also known as small integrin-binding ligand *N*-linked glycoproteins (SIBLINGs), are key genes involved in biomineralization.¹³ SIBLINGs are evolutionarily related to other subfamilies of the SCPP genes affecting tooth, including the enamel matrix genes, caseins and salivary and related genes.¹⁴ *SPP1* encodes osteopontin, is expressed in many tissues and is involved in mineralized tissue remodeling, among a variety of other functions. Less is known regarding the role of *MEPE*, which encodes osteoregulin, although it is expressed during odontogenesis¹⁵ as well as in skeletal tissues and tumors. *IBSP* encodes bone sialoprotein 2, a component of mineralized tissues, including dentin and cementin.¹⁶ *DMP1* and *DSPP* encode dentin matrix acidic phosphoprotein 1 and dentin sialophosphoprotein, respectively, and both are crucial for mineralization of dentin, with *DMP1* thought to regulate *DSPP*.¹⁷ Moreover, several SIBLINGs bind to and activate matrix metalloproteinase (MMP) partners, *MMP2*, *MMP3* and *MMP9*,¹⁸ which are in turn known to be involved in tooth development.^{19,20} In addition, SIBLINGs are expressed in the salivary gland.²¹ Overall, through their involvement in dentin mineralization and expression in oral environment, the SIBLINGs are sensible candidates for roles in susceptibility to tooth decay.

The other genes at the implicated locus are *PKD2* (polycystin-1) and *ABCG2* (a membrane transporter), both of which are likewise plausible candidates for influencing dental caries. For example, Khonsari *et al.*²² showed that loss of *PKD2* caused craniofacial and dental defects in mice, including irregular incisors, molar root fractures, alveolar bone loss and compressed temporomandibular joints. Moreover, mutations in *PKD2* cause autosomal-dominant polycystic kidney disease (ADPKD) in humans, and the same study showed differences in facial characteristics and asymmetry in ADPKD patients compared with controls.²² This finding suggests that *PKD2* may similarly impact craniofacial, and possibly dental, traits in humans. *ABCG2* is involved in trafficking molecules across membranes in multiple tissues. There is no direct evidence that *ABCG2* impacts characteristics of the teeth, although it is expressed in human dental pulp,²³ developing murine incisor²⁴ and human ameloblastic tumors,²⁵ which together suggest a possible role in growth of dental tissues.

Given that several genes in the region on chromosome 4q21 have functions potentially relevant to dental caries experience, we designed the present study to test and fine-map genetic association with variants in this region. Furthermore, because previous studies have identified statistical interaction effects on caries between other SCPP genes and fluoride exposures,^{26,27} we also explored whether fluoride may mediate the association of SIBLING variants with caries.

MATERIALS AND METHODS

Participant recruitment and data collection

Thirteen age- and race-stratified samples from six independent studies were recruited. Details of the study designs and data collection efforts for these samples have previously been reported.¹¹ In brief, the six studies are: (1) the Center for Oral Health Research in Appalachia, Cohort 1 (COHRA1; *N*=1910),²⁸ which recruited a multiracial (though primarily white) sample from rural West Virginia and Pennsylvania, (2) the Iowa Head Start Study (IHS; *N*=64),²⁹ which recruited lower income white children from Iowa, (3) the Iowa Fluoride Study (IFS; *N*=154),³⁰ which recruited almost exclusively white children from Iowa, (4) Dental Strategies Concentrating on Risk

Evaluation (Dental SCORE; *N*=530),³¹ which performed targeted recruitment of older blacks and whites from the Pittsburgh area, (5) the University of Pittsburgh Dental Registry and DNA Repository (DRDR; *N*=1169),^{10,11} which recruited a multiracial (though primarily white) sample of patients seeking treatment at the School of Dental Medicine, and (6) the Center for Education and Drug Abuse Research (CEDAR; *N*=262),³² which recruited a multiracial sample of children who have fathers with or without substance use disorder.

Participants were stratified into analysis groups comprising non-Hispanic whites and blacks based on self-reported and genetically confirmed race. This was to avoid spurious genetic associations owing to confounding by population structure. Owing to small sample sizes, participants reporting as other racial categories, including mixed, were excluded. Likewise, participants were stratified based on age, with children aged 3–12 years comprising the child samples, and adults aged ≥ 18 years comprising the adult samples. The one exception was the CEDAR study that included individuals aged ≥ 15 years who were grouped with the adults for the purposes of this study. This approach was to investigate caries in the primary and permanent dentitions separately under the hypothesis that risk factors may differ between dentitions.

All participants were recruited without regard to their oral health status and underwent dental caries assessment via intraoral examination. Each present tooth was scored for evidence of dental caries, including the occurrence of untreated decay and restorations indicative of past decay. Missing teeth were noted. From these assessments, two commonly used semiquantitative caries indices were generated: decayed, missing and filled teeth in the permanent dentition of adults, and decayed and filled teeth in the primary dentition of children. Note that the occurrence of missing teeth in children was not considered evidence of decay owing to the difficulty in determining the cause of missingness given primary tooth exfoliation. Third molars (that is, 'wisdom teeth') were excluded from assessments of caries in the permanent dentition. Our assessment approach and the decayed, missing and filled teeth/decayed and filled teeth phenotypes utilized are consistent with the recommendations of the PhenX toolkit (www.phenxtoolkit.org), designed to maximize interstudy comparability. Fluoride exposure data were collected for a subset of the COHRA1 samples and included fluoride concentration (p.p.m.) in a home water sample as measured by an ion-specific electrode and self- or parent-reported frequency of tooth brushing. Each of the two fluoride exposures was dichotomized into low- and high-risk classes. Home water source fluoride concentration was dichotomized as >0.7 p.p.m. vs <0.7 p.p.m., representing sufficient and insufficient fluoride concentrations, respectively. Tooth brushing frequency was dichotomized as daily or more frequent tooth brushing vs less frequent than daily brushing, which reflects the consensus in the literature that at least daily brushing is needed,³³ and is consistent with the dichotomization used in previous studies.^{9,27,34}

Genotype data collection

Forty-nine single-nucleotide polymorphisms (SNPs) in the chromosome 4q21 region were genotyped by the Center for Inherited Disease Research at Johns Hopkins University using the Illumina (San Diego, CA, USA) GoldenGate technology. These SNPs were part of a custom panel of variants chosen to follow up a variety of genetic associations with oral health-related outcomes as well as SNPs in *a priori* candidate genes of interest. The 49 SNPs in the chromosome 4q21 locus were selected to tag the common variation in this region. Criteria used to select these specific SNPs were (1) compatibility with the GoldenGate technology, (2) high 'designability' scores indicating likelihood of successful genotyping, (3) minor allele frequency >0.02 , (4) low redundancy in information with other genotyped SNPs in the region as determined by multiple correlation coefficient observed in the International HapMap Project,³⁵ and (5) physical proximity to other SNPs on the panel owing to technical limitations in the genotyping method. Details regarding genotype data quality control and the composition of the rest of the custom genotyping panel have been previously described.¹¹

Statistical analysis

Dental caries indices are semiquantitative measures that approximate continuous distributions owing to their broad range and high variance and, therefore, were analyzed using robust quantitative methods commonly used for

Table 1 Characteristics of the samples, mean (range) or percentage

Sample	N	Age, years	Female (%)	DFT/DMFT	Fluoridated water (%)	Tooth brushing per day
<i>Children</i>						
COHRA1 whites	667	7.3 (3.0–12.0)	46.70	2.3 (0–17)	60.21	1.59 (0–4)
COHRA1 blacks	92	7.6 (3.2–11.8)	46.90	1.8 (0–8)	86.79	1.60 (0–3)
IHS whites	41	4.1 (3.2–5.3)	58.50	6.3 (0–20)	—	—
IHS blacks	23	4.3 (3.4–5.6)	52.20	5.7 (0–17)	—	—
IFS whites	154	5.2 (4.4–6.8)	48.50	1.2 (0–16)	—	—
<i>Adults</i>						
COHRA1 whites	1061	34.3 (18.0–75.0)	62.80	10.5 (0–28)	58.82	1.47 (0–2)
COHRA1 blacks	90	36.2 (18.2–60.8)	70.90	9.3 (9–28)	88.00	1.53 (0–2)
Dental SCORE whites	293	64.0 (48.0–78.0)	63.20	16.4 (2–28)	—	—
Dental SCORE blacks	237	61.6 (47.0–79.0)	72.90	14.8 (1–28)	—	—
DRDR whites	928	43.0 (18.0–74.8)	50.00	16.6 (0–28)	—	—
DRDR blacks	241	44.5 (18.0–74.4)	57.80	16.5 (0–28)	—	—
CEDAR whites	186	20.4 (15.7–28.6)	31.20	5.4 (0–21)	—	—
CEDAR blacks	76	20.2 (15.6–27.8)	44.30	6.4 (0–16)	—	—

Abbreviations: CEDAR, Center for Education and Drug Abuse Research; COHRA1, Center for Oral Health Research in Appalachia, Cohort 1; Dental SCORE, Dental Strategies Concentrating of Risk Evaluation; DFT, decayed or filled primary teeth; DMFT, decayed, missing or filled permanent teeth; DRDR, Dental Registry and DNA Repository; IFS, Iowa Fluoride Study; IHS, Iowa Head Start Study.

these phenotypes. Genetic association was tested using linear regression under the additive genetic model while simultaneously adjusting for age, sex and two principal components (PCs) of ancestry (generated across all cohorts using PC analysis of 2663 SNPs in 71 genes of interest as well as 96 ancestry-informative SNPs specifically chosen for modeling ancestry). Analyses were performed separately in 13 age- (that is, child vs adult) and race-stratified samples; this decision was based on hypothesized differences in the genes underlying caries susceptibility in the primary and permanent dentitions^{3,36} as well as to guard against spurious results owing to population structure. Stouffer's inverse-variance weighted method was used to combine *P*-values across stratified analyses in order to determine statistical significance of associations across samples. This method was deemed appropriate given the differences in scale of caries indices across samples. Fixed- and random-effects inverse-variance weighted meta-analyses (based on effect sizes and s.e. estimates) were used to generate overall effect sizes across samples for associated SNPs. The method by Li and Ji³⁷ was used to define appropriate thresholds for declaring statistical significance based on adjustment for the number of independent SNP-wise tests performed.

In light of recent results documenting gene-by-fluoride exposure interaction effects on dental caries,^{26,27} we also tested for interaction effects with the two fluoride exposures in the samples of sufficient size for which fluoride exposure data were available (that is, COHRA1 white adults and COHRA1 white children). These fluoride exposures were dichotomized measures of home water source fluoride (<0.7 vs ≥0.7 p.p.m.) and tooth brushing frequency (once or more per daily vs less often than once per day). Interaction effects were tested using linear regression while simultaneously modeling SNP and fluoride main effects as well as sex, age and two PCs of ancestry. In order to have sufficient number of participants in each SNP-by-fluoride stratum necessarily to accurately model the interaction effect, for the interaction models only, we combined the heterozygote and minor allele homozygote (that is, assumed the dominant genetic model) for SNPs with minor allele frequencies <25%. In light of both the multiple comparisons issue and the linkage disequilibrium (LD) among the SNPs in this region of genome, the method by Li and Ji³⁷ was used to declare statistical significance at *P*-values <0.0032 based on the number of functionally independent SNPs. Interaction models showing *P*-values <0.05 were considered 'suggestive' trends.

RESULTS

Characteristics of the age- and race-stratified samples are presented in Table 1. The observed variation in dental caries experience across samples was expected given the differences in age and demography.

Tests for genetic association with 49 SNPs in the chromosome 4q21 region were performed separately in each sample (Supplementary Table 1), and results across samples were combined via meta-analysis (Figure 1). SNPs in and immediately downstream of *PKD2* showed significant evidence of association in COHRA1 adult blacks (rs17013735, *P*-value=0.0009) and Dental SCORE adult whites (rs11938025; *P*-value=0.0005; rs2725270, *P*-value=0.003). No significant associations were observed in children. Meta-analyses showed that the associated SNP observed in COHRA1 adult blacks was also significantly associated with dental caries across all adult black samples combined (rs17013735; *P*-value=0.003; Figure 1). Full results of associations for all SNPs across all samples are available in Supplementary Material. Genotype distributions for select SNPs are shown in Supplementary Table 2.

Figure 2 shows a forest plot of SNP rs17013735 in black adult samples; the caries-SNP association results are detailed in Supplementary Figure 2. The variant has large effects (that is, 2–5 carious teeth) in COHRA1, Dental SCORE and DRDR black adults, whereas in CEDAR black adults the point estimate near zero and wide confidence interval precludes any conclusion about the size or direction of the effect. The overall (meta-analysis) effect size is 1.99 and 2.07 carious teeth for fixed- and random-effects models, respectively. SNP rs17013735 has a higher minor allele frequency in blacks compared with whites (0.15 and 0.03 in African American and European ancestry groups, respectively, in dbSNP). It is located in an intron and is in high ($r^2 > 0.8$; $D' = 1.0$) LD (from the 1000 Genomes Project) with four SNPs (rs4484262, rs75400904, rs12500843, rs12500008; the latter two near *SPP1*) in regulatory elements (defined by DNase hypersensitivity and H3K27Ac signatures in ENCODE). However, it is not in LD with the leading SNP from the previously published GWAS ($r^2 < 0.2$ in all ancestry groups).

Though not meeting the threshold for statistical significance after adjustment for multiple comparisons, SNPs in *DSPP* showed evidence of association in the meta-analysis of all adults (for example, rs6532012, *P*-value=0.005), and a SNP in *MEPE* showed a trend in meta-analysis of all blacks (rs10018300, *P*-value=0.01).

Given the important protective role of fluoride exposure, we also tested for gene-by-fluoride exposure interaction effects in the

COHRA1 white samples (our largest samples for which fluoride data were available). One significant interaction with fluoride exposure as measured by frequency of tooth brushing was observed for the SNP

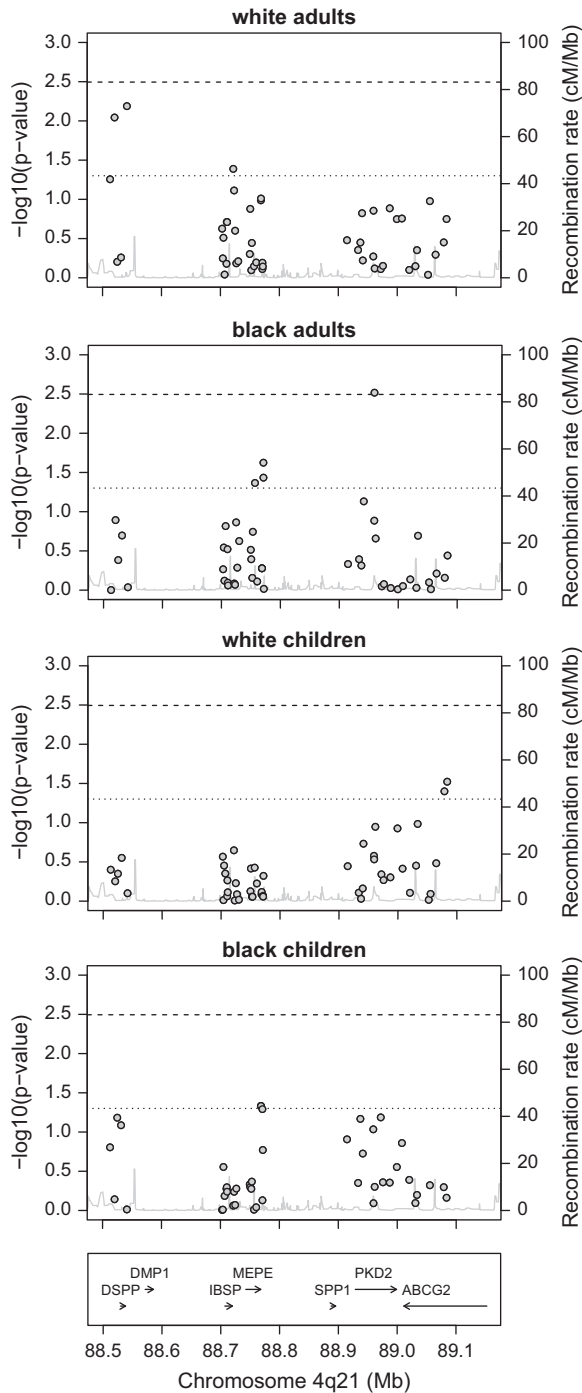


Figure 1 Evidence of association for 49 SNPs in the chromosome 4q21 region. Negative log₁₀-transformed *P*-values (left y axis) are shown for meta-analyses across white adults, black adults, white children and black children. The recombination rate overlay (right y axis) indicates the LD structure of the region. Note, plotted SNPs are in low LD ($r^2 < 0.2$) with each other. The horizontal dashed lines indicate (lower) *P*-value of 0.05 and (upper) adjusted threshold for significance given multiple comparisons as per the method by Li and Ji.³⁷ The arrows indicate the physical positions and directionality of genes of interest.

rs2725233 upstream of *PKD2* in COHRA1 white adults (SNP main effect *P*-value=0.005, interaction *P*-value=0.002). Additionally, though not significant after adjustment for multiple comparisons, we observed an interaction trend for rs4282132 (SNP main effect *P*-value=0.02, interaction *P*-value=0.01). Each of these interactions took the same form, whereby participants with two copies of the risk allele experienced greater dental caries only if in the low fluoride strata (Figure 3).

DISCUSSION

Here we report a follow-up study seeking to explore the region of chromosome 4q21 that was previously implicated in a GWAS of dental caries.⁸ The original report showed an association peak in *ABCG2* near *PKD2*. This locus was just downstream of the *SIBLINGs*, a cluster of autologous genes with key roles in biomineralization of dentin, which were considered strong candidates for the observed GWAS signal. Our results show significant association with multiple tag-SNPs in this region, notably in *PKD2*, which we interpret as evidence of regional replication. Moreover, one of the strongest associations (rs17013735) was race specific, with substantial differences in allele frequency between whites and blacks. This result suggests that risk variants that differ in frequency across ancestry groups may account for part of the disparity in caries experience across racial and ethnic strata that is not attributable to environmental factors. Given the low allele frequency, statistical power to detect genetic association of rs17013735 was low in whites (for example, 25% via meta-analysis of 2468 white adults assuming the same effect size and significance as in blacks).

The original GWAS signal was discovered in the COHRA1 white adult sample, which was included as 1 of the 13 samples in the present study. The evidence of genetic association (main effects) reported herein came from samples other than the COHRA1 whites, which did not show strong evidence of association. Though this observation may seem surprising, our study included different SNPs slightly centromeric of the original GWAS signal in order to follow up genes with plausible biological roles, and therefore the lack of significant associations in the COHRA1 white adult sample are consistent (by physical position) with the previous GWAS results in this group. Interestingly, we saw evidence of gene-by-fluoride exposure interaction effects for tooth brushing in the COHRA1 white adults, though we caution that these results are preliminary pending replication. The form of the interaction suggests that genetic risk may be important for individuals who lack adequate caries protection via fluoride or, equivalently, that fluoride is especially important in those with genetic predisposition for dental caries.

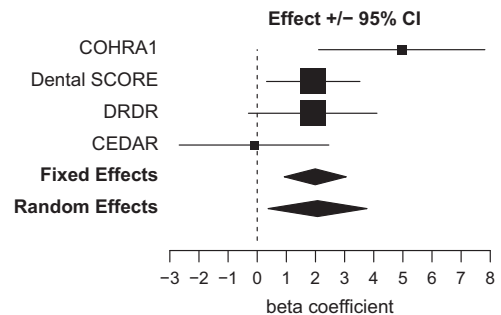


Figure 2 Forest plot showing the effects of the SNP rs17013735 in black adult samples. Beta-coefficients indicate the per allele increase in decayed, missing and filled teeth scores for the risk variant. Fixed- and random-effects meta-analyses show the overall effect size across all black adult samples.

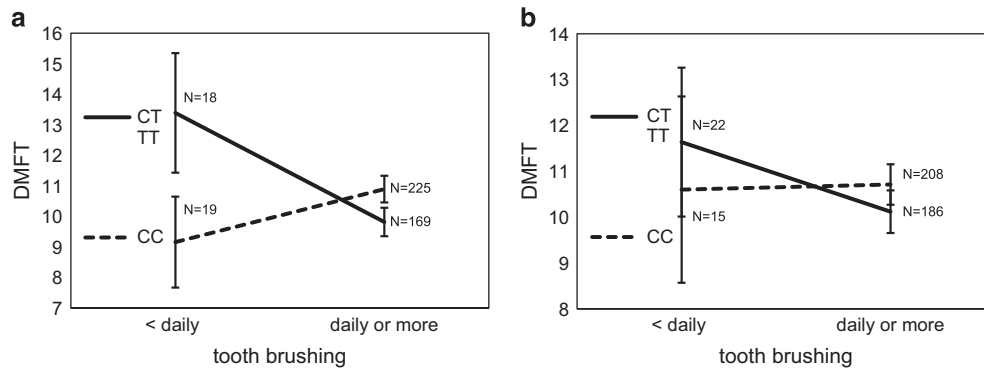


Figure 3 Gene-by-fluoride exposure interaction plots showing the mean (s.e.) of the genotype groups across the low and high fluoride strata. Interactions were observed between tooth brushing frequency and (a) rs2725233 in COHRA1 white adults ($N=431$, P -value=0.002) and (b) rs4282132 in COHRA1 white adults ($N=431$, P -value=0.01). The number of participants in each genotype-by-fluoride exposure stratum is annotated. For both SNPs, among participants reporting tooth brushing frequency less than once daily, decayed, missing and filled teeth (DMFT) scores differed between those with one or two copies of the rarer T allele compared with homozygotes for the common C allele. In contrast, among those reporting tooth brushing frequency of once or more per day, DMFT scores did not differ by genotype. This interaction was statistically significant (after considering multiple comparisons) for rs2725233 and was a suggestive trend for rs4282132.

The specific SNPs interrogated in this study were selected to capture the majority of variation in this genomic region while limiting redundancy of information and thus are not themselves expected to be causal but may be in LD with causal variants. For example, the SNP showing association in the meta-analysis of black adults was an intronic variant with no predicted functional role; however, it is in strong LD with other variants in regulatory elements (in some cell types) near *PKD2* and *SPPI*. Given the fact that gene targets of regulatory elements are not necessarily the genes nearest to the element, it is unclear which, if any, of the genes in this region is affected by these variants. Likewise, it is unknown if these regulatory elements are active in cell types relevant to dental caries (although regulatory elements are indeed frequently shared across cell types). The variants showing the strongest evidence of association were all in and near *PKD2*, which is a plausible candidate given the experimental evidence of craniofacial and dental defects in mice.²² Similarly, the nearby SIBLINGs are strong candidates based on their biological roles,¹³ with some variants showing modest statistical evidence. Overall, this study showed diffuse evidence of association across the region, with no single variant clearly accounting for the original GWAS signal. Indeed, associated SNPs observed here were not in high LD ($r^2 < 0.2$) with the leading SNP from the previous GWAS. Therefore, while we interpret our results as strengthening the hypothesis that chromosome 4q21 may impact dental caries, they do not point to a specific gene as the clear culprit.

This study benefited from a large sample size and by investigating caries in both primary and permanent dentitions, which, previous work has suggested, differ in their genetic risk.^{3,36} We, too, found different results between dentitions, although this could be explained by differences in power. This was also one of the few studies to consider genetic associations with dental caries in African Americans, which is an important population to study given the disproportionate rates of untreated decay and differences in frequencies of putative risk alleles, such as rs17013735 shown herein. Our analyses benefited from adjustment for two PCs of ancestry, which were generated across all samples combined using genetic data from candidate genes of interest and ancestry-informative SNPs. As a guard against spurious results owing to population structure, we view this as a strength, especially as many targeted/candidate gene association studies do not ascertain ancestry. However, we note that

these PCs were generated from far fewer data points than genome-wide studies; hence, we do not have sufficient resolution for capturing subtle population structure within non-admixed ancestry groups; for example, geographic clines within whites.

In addition to these strengths, there are some potential limitations affecting this research. Despite the large sample size overall, fluoride exposure data were only available for some cohorts, and interaction analyses were likely underpowered. Moreover, our two fluoride exposures, home water concentration and tooth brushing frequency, represent major sources of fluoride but are limited in that they fail to capture the duration of topical exposure to the tooth enamel. Therefore, SNP-by-fluoride interactions should be interpreted with caution, and negative results should be interpreted as lack of evidence (rather than evidence that interactions are absent). Additionally, there are some issues to consider related to the statistical model. For example, as a quantitative phenotype our caries measurement lacked precision and was modeled using linear regression, which, while robust in terms of validity, may yield suboptimal power if the error is non-normally distributed. Both of these issues may bias our association tests toward the null hypothesis of no effect but would not cause false positive associations.

In conclusion, we showed significant associations with multiple tag-SNPs in the chromosome 4q21 region, which we interpret as evidence of regional replication. No single causal variant (or proxy) was identified. Therefore, we interpret our results as strengthening the hypothesis that genetic variation in this region may impact dental caries and recognize that additional work is needed to determine the causal variant(s) and mechanisms through which they impact risk of decay.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We express our gratitude to the participants of the six studies, whose contributions have enabled this work. This effort was supported by the following National Institutes of Health grants: R03-DE024264, U01-DE018903, R01-DE014899, R01-DE009551, R01-DE012101, R01-DE018914, P50-DA005605, and R01-DA019157, as well as the National Science Foundation/Department of Defense grant DBI-1263020. The Dental Registry and DNA Repository is supported by the University of Pittsburgh School of

Dental Medicine. The Dental SCORE sample is partially supported by the Commonwealth of Pennsylvania Department of Health grant ME-02-384.

Author contributions: JRS and MLM conceived of the study; SE performed the statistical analyses; JRS, EF, MC and MLM designed the genotyping panel and performed the genetic data cleaning and quality control; EF, MC, MMV, BSM, RLS, MCW, SER, DWM, RJC, RJW, SML, ARV, MLM and JRS were involved in the study design, data collection and data cleaning of the parent studies; JRS wrote the manuscript; all the authors interpreted the results, revised the manuscript and approved the manuscript for publication.

- 1 Bretz, W. A., Corby, P. M., Melo, M. R., Coelho, M. Q., Costa, S. M., Robinson, M. *et al*. Heritability estimates for dental caries and sucrose sweetness preference. *Arch. Oral Biol.* **51**, 1156–1160 (2006).
- 2 Bretz, W. A., Corby, P. M., Schork, N. J., Robinson, M. T., Coelho, M., Costa, S. *et al*. Longitudinal analysis of heritability for dental caries traits. *J. Dent. Res.* **84**, 1047–1051 (2005).
- 3 Wang, X., Shaffer, J. R., Weyant, R. J., Cuenco, K. T., DeSensi, R. S., Crout, R. *et al*. Genes and their effects on dental caries may differ between primary and permanent dentitions. *Caries Res.* **44**, 277–284 (2010).
- 4 Shaffer, J. R., Wang, X., Desensi, R. S., Wendell, S., Weyant, R. J., Cuenco, K. T. *et al*. Genetic susceptibility to dental caries on pit and fissure and smooth surfaces. *Caries Res.* **46**, 38–46 (2012).
- 5 Shaffer, J. R., Feingold, E., Wang, X., Tcuenco, K. T., Weeks, D. E., DeSensi, R. S. *et al*. Heritable patterns of tooth decay in the permanent dentition: principal components and factor analyses. *BMC Oral Health* **12**, 7 (2012).
- 6 Zeng, Z., Feingold, E., Wang, X., Weeks, D. E., Lee, M., Cuenco, D. T. *et al*. Genome-wide association study of primary dentition pit-and-fissure and smooth surface caries. *Caries Res.* **48**, 330–338 (2014).
- 7 Zeng, Z., Shaffer, J. R., Wang, X., Feingold, E., Weeks, D. E., Lee, M. *et al*. Genome-wide association studies of pit-and-fissure- and smooth-surface caries in permanent dentition. *J. Dent. Res.* **92**, 432–437 (2013).
- 8 Shaffer, J. R., Feingold, E., Wang, X., Lee, M., Tcuenco, K., Weeks, D. E. *et al*. GWAS of dental caries patterns in the permanent dentition. *J. Dent. Res.* **92**, 38–44 (2013).
- 9 Shaffer, J. R., Wang, X., Feingold, E., Lee, M., Begum, F., Weeks, D. E. *et al*. Genome-wide association scan for childhood caries implicates novel genes. *J. Dent. Res.* **90**, 1457–1462 (2011).
- 10 Wang, X., Shaffer, J. R., Zeng, Z., Begum, F., Vieira, A. R., Noel, J. *et al*. Genome-wide association scan of dental caries in the permanent dentition. *BMC Oral Health* **12**, 57 (2012).
- 11 Stanley, B. O., Feingold, E., Cooper, M., Vanyukov, M. M., Maher, B. S., Slayton, R. L. *et al*. Genetic association of MPPED2 and ACTN2 with dental caries. *J. Dent. Res.* **93**, 626–632 (2014).
- 12 Fisher, L. W. DMP1 and DSPP: evidence for duplication and convergent evolution of two SIBLING proteins. *Cells Tissues Organs* **194**, 113–118 (2011).
- 13 Fisher, L. W. & Fedarko, N. S. Six genes expressed in bones and teeth encode the current members of the SIBLING family of proteins. *Connect Tissue Res.* **44**(Suppl 1): 33–40 (2003).
- 14 Kawasaki, K. & Weiss, K. M. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc. Natl Acad. Sci. USA* **100**, 4060–4065 (2003).
- 15 MacDougall, M., Simmons, D., Gu, T. T. & Dong, J. MEPE/OF45, a new dentin/bone matrix protein and candidate gene for dentin diseases mapping to chromosome 4q21. *Connect Tissue Res.* **43**, 320–330 (2002).
- 16 Qin, C., Brunn, J. C., Jones, J., George, A., Ramachandran, A., Gorski, J. P. *et al*. A comparative study of sialic acid-rich proteins in rat bone and dentin. *Eur. J. Oral Sci.* **109**, 133–141 (2001).
- 17 Gibson, M. P., Zhu, Q., Wang, S., Liu, Q., Liu, Y., Wang, X. *et al*. The rescue of dentin matrix protein 1 (DMP1)-deficient tooth defects by the transgenic expression of dentin sialophosphoprotein (DSPP) indicates that DSPP is a downstream effector molecule of DMP1 in dentinogenesis. *J. Biol. Chem.* **288**, 7204–7214 (2013).
- 18 Fedarko, N. S., Jain, A., Karadag, A. & Fisher, L. W. Three small integrin binding ligand N-linked glycoproteins (SIBLINGs) bind and activate specific matrix metalloproteinases. *FASEB J.* **18**, 734–736 (2004).
- 19 Caron, C., Xue, J., Sun, X., Simmer, J. P., Bartlett, J. D. & Gelatinase, A. MMP-2) in developing tooth tissues and amelogenin hydrolysis. *J. Dent. Res.* **80**, 1660–1664 (2001).
- 20 Feng, J., McDaniel, J. S., Chuang, H. H., Huang, O., Rakian, A., Xu, X. *et al*. Binding of amelogenin to MMP-9 and their co-expression in developing mouse teeth. *J. Mol. Histol.* **43**, 473–485 (2012).
- 21 Ogbureke, K. U. & Fisher, L. W. Expression of SIBLINGs and their partner MMPs in salivary glands. *J. Dent. Res.* **83**, 664–670 (2004).
- 22 Khonsari, R. H., Ohazama, A., Raouf, R., Kawasaki, M., Kawasaki, K., Pomtaveetus, T. *et al*. Multiple postnatal craniofacial anomalies are characterized by conditional loss of polycystic kidney disease 2 (Pkd2). *Hum. Mol. Genet.* **22**, 1873–1885 (2013).
- 23 Honda, M. J., Nakashima, F., Satomura, K., Shinohara, Y., Tsuchiya, S., Watanabe, N. *et al*. Side population cells expressing ABCG2 in human adult dental pulp tissue. *Int. Endod. J.* **40**, 949–958 (2007).
- 24 Li, L., Kwon, H. J., Harada, H., Ohshima, H., Cho, S. W. & Jung, H. S. Expression patterns of ABCG2, Bmi-1, Oct-3/4, and Yap in the developing mouse incisor. *Gene Expr. Patterns* **11**, 163–170 (2011).
- 25 Kumamoto, H. & Ohki, K. Detection of CD133, Bmi-1, and ABCG2 in ameloblastic tumors. *J. Oral Pathol. Med.* **39**, 87–93 (2010).
- 26 Anjomshoa, I., Briseno-Ruiz, J., Deeley, K., Poletta, F. A., Mereb, J. C., Leite, A. L. *et al*. Aquaporin 5 interacts with fluoride and possibly protects against caries. *PLoS ONE* **10**, e0143068 (2015).
- 27 Shaffer, J. R., Carlson, J. C., Stanley, B. O., Feingold, E., Cooper, M., Vanyukov, M. M. *et al*. Effects of enamel matrix genes on dental caries are moderated by fluoride exposures. *Hum. Genet.* **134**, 159–167 (2015).
- 28 Polk, D. E., Weyant, R. J., Crout, R. J., McNeil, D. W., Tarter, R. E., Thomas, J. G. *et al*. Study protocol of the Center for Oral Health Research in Appalachia (COHRA) etiology study. *BMC Oral Health* **8**, 18 (2008).
- 29 Slayton, R. L., Cooper, M. E. & Marazita, M. L. Tuffelin, mutans streptococci, and dental caries susceptibility. *J. Dent. Res.* **84**, 711–714 (2005).
- 30 Wang, X., Willing, M. C., Marazita, M. L., Wendell, S., Warren, J. J., Broffitt, B. *et al*. Genetic and environmental factors associated with dental caries in children: the Iowa Fluoride Study. *Caries Res.* **46**, 177–184 (2012).
- 31 Ayer, A. N., Kip, K. E., Marroquin, O. C., Mulukutla, S. R., Edmundowicz, D. & Reis, S. E. Racial differences in coronary artery calcification are not attributed to differences in lipoprotein particle sizes: the Heart Strategies Concentrating on Risk Evaluation (Heart SCORE) Study. *Am. Heart J.* **153**, 328–334 (2007).
- 32 Vanyukov, M. M., Maher, B. S., Devlin, B., Tarter, R. E., Kirillova, G. P., Yu, L. M. *et al*. Haplotypes of the monoamine oxidase genes and the risk for substance use disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **125B**, 120–125 (2004).
- 33 Attin, T. & Hornecker, E. Tooth brushing and oral health: how frequently and when should tooth brushing be performed? *Oral Health Prev. Dent.* **3**, 135–140 (2005).
- 34 Shaffer, J. R., Polk, D. E., Feingold, E., Wang, X., Cuenco, K. T., Weeks, D. E. *et al*. Demographic, socioeconomic, and behavioral factors affecting patterns of tooth decay in the permanent dentition: principal components and factor analyses. *Community Dent. Oral Epidemiol.* **41**, 364–373 (2013).
- 35 International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
- 36 Bayram, M., Deeley, K., Reis, M. F., Trombetta, V. M., Ruff, T. D., Sencak, R. C. *et al*. Genetic influences on dental enamel that impact caries differ between the primary and permanent dentitions. *Eur. J. Oral Sci.* **123**, 327–334 (2015).
- 37 Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb.)* **95**, 221–227 (2005).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)