ORIGINAL ARTICLE

Variants on chromosome 4q21 near PKD2 and SIBLINGs are associated with dental caries

Scott Eckert1,19, Eleanor Feingold2,3, Margaret Cooper4,5, Michael M Vanyukov2,6,7, Brion S Maher8, Rebecca L Slayton9, Marcia C Willing10, Steven E Reis11,12, Daniel W McNeil13,14, Richard J Crout15, Robert J Weyant16, Steven M Levy17,18, Alexandre R Vieira4,5, Mary L Marazita2,4,5,7,12 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 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, although few of these have been replicated in follow-up studies. 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 Pharmaceutical Sciences, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA; 6Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 7Department of Biostatistics, Graduate School of Public Health, 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; 10School of Medicine, University of Washington, Seattle, WA, USA; 11Division of Genetics and Genomics, Department of Pediatrics, School of Medicine, University of Washington, Seattle, WA, USA; 12Department of Dental Practice and Rural Health, West Virginia University, Morgantown, WV, USA; 13Department of Psychology, West Virginia University, Morgantown, WV, USA; 14Department of Periodontics, School of Dentistry, West Virginia University, Morgantown, WV, USA; 15Department of Preventive and Community Dentistry, University of Iowa College of Dentistry, Iowa City, IA, 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; 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.

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. The known biology of genes in this region suggests plausible roles in dental caries. Most relevant, the paralogous 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. SIBLINGs are evolutionarily related to other subfamilies of the SCPP genes affecting tooth, including the enamel matrix genes, caseins and salivary and related genes. 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 odontogenesis as well as in skeletal tissues and tumors. IBSP encodes bone sialoprotein 2, a component of mineralized tissues, including ameloblastic tumors, which together suggest a possible role in patients compared with controls. This showed differences in facial characteristics and asymmetry in ADPKD showed that loss of ABCG2 and related genes. The other genes at the implicated locus are ABCG2 (a membrane transporter), both of which are likewise plausible candidates for influencing dental caries. For example, Khonsari et al. 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. 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, developing murine incisor and human ameloblastic tumors, 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, 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. In brief, the six studies are: (1) the Center for Oral Health Research in Appalachia, Cohort 1 (COHRA1; N=1910), 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), 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 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.

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, and is consistent with the dichotomization used in previous studies.

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, 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.

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
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 = 0.0003; 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 = 0.0003; 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 DRD 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, rs12500068; 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
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. 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 allelic 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 allelic 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 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. The arrows indicate the physical positions and directionality of genes of interest.

**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.
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 intrinsic 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 SPP1. 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.3,6 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
Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)