Genome-wide association study of primary dentition pit-and-fissure and smooth surface caries

Zhen Zeng[1,2], **Eleanor Feingold**[1,2], **Xiaojing Wang**[3], **Daniel E. Weeks**[1,2], **Myoungkeun Lee**[3], **Karen T. Cuenco**[1,3], **Barbara Broffitt**[4], **Robert J. Weyant**[5], **Richard Crout**[6], **Daniel W. McNeil**[7,8], **Steven M. Levy**[4,9], **Mary L. Marazita**[1,3], and **John. R. Shaffer**[1]

[1]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

[2]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

[3]Center for Craniofacial & Dental Genetics, Department of Oral Biology, School of Dental Medicine, Pittsburgh, PA, USA

[4]Department of Preventive and Community Dentistry, University of Iowa College of Dentistry, Iowa City, IA, USA

[5]Department of Dental Public Health and Information Management, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA

[6]Department of Periodontics, School of Dentistry, West Virginia University, Morgantown, W V, USA

[7]Department of Dental Practice and Rural Health, West Virginia University, Morgantown, WV, USA

[8]Department of Psychology, West Virginia University, Morgantown, W V, USA

[9]Department of Epidemiology, University of Iowa College of Public Health, Iowa City, IA, USA

Abstract

Dental caries continues to be the most common chronic disease in children today. Despite the substantial involvement of genetics in the process of caries development, the specific genes contributing to dental caries remain largely unknown.

We performed separate genome-wide association studies of smooth and pit-and-fissure tooth surface caries experience in the primary dentitions of self-reported white children in two samples from Iowa and rural Appalachia. In total, 1006 children (ages 3-12 years) were included for smooth surface analysis, and 979 children (ages 4-14 years) for pit-and-fissure surface analysis. Associations were tested for more than 1.2 million single nucleotide polymorphisms, either genotyped or imputed.

Corresponding authors: John R. Shaffer, Ph.D., Department of Human Genetics, University of Pittsburgh, 130 DeSoto Street, A300 Crabtree Hall, Pittsburgh, PA 15261, jrs51@pitt.edu.

We detected genome-wide significant signals in *KPNA4* (p-value = 2.0E-9), and suggestive signals in *ITGAL* (p-value = 2.1E-7) and *PLUNC* family genes (p-value = 2.0E-6), thus nominating these novel loci as putative caries susceptibility genes. We also replicated associations observed in previous studies for *MPPED2* (p-value = 6.9E-6), *AJAP1* (p-value = 1.6E-6) and *RPS6KA2* (p-value = 7.3E-6). Replication of these associations in additional samples, as well as experimental studies to determine the biological functions of associated genetic variants, are warranted. Ultimately, efforts such as this may lead to a better understand of caries etiology, and could eventually facilitate the development of new interventions and preventive measures.

Keywords

Introduction

Dental caries (i.e., tooth decay) is the most common chronic disease in children globally, and despite advances in oral healthcare, the incidence of dental caries in young children in the US has increased in recent decades [Beltran-Aguilar et al., 2005; Dye et al., 2007]. The effects of childhood dental caries on quality of life can be profound, including chronic pain, tooth loss, difficulty hearing, eating, and sleeping, and failure to thrive [Acs et al., 1992; Acs et al., 1999], as well as substandard school performance, poor social relationships, and decreased success later in life [Low et al., 1999]. In addition, disease burden and associated co-morbidities vary considerably across U.S. socioeconomic and ethnic strata, making dental caries in children a focal issue in efforts to reduce public health disparities.

Cariogenesis is multi-factorial [Anderson, 2002; Hunter, 1988], influenced by sex [Ferraro and Vieira, 2010] and environmental factors, such as bacterial flora, dietary behaviors, fluoride intake and exposures, oral hygiene, salivary composition and flow rate, and tooth positional and morphological features, as well as genetic factors, and gene-by-environment interactions. In addition, the etiology of dental caries is further complicated by the non-uniform risk across tooth surfaces of the primary dentition. Tooth surfaces can be grouped according to the sequence and pattern of decay, which demonstrates a surface hierarchy in susceptibility to dental caries [Batchelor and Sheiham, 2004]. In general, pit-and-fissure surfaces exhibit much greater risk of developing carious lesions than do smooth surfaces [Batchelor and Sheiham, 2004], and progression of decay differs between these surface types.

Just as environmental risk factors exert differential effects between pit-and-fissure and smooth surfaces, (e.g., exposure to fluoride better protects smooth surfaces [Jiang et al., 2005], tooth brushing frequency and sugary drinks have greater effects on pit-and-fissure surfaces [Warren et al., 2006]), we hypothesize that genetic factors may also differentially affect pit-and-fissure and smooth surfaces. For example, genetic factors that modulate the differential effects of environmental exposures (e.g., fluoride-sensitivity genes and taste-preference genes), or that are involved in the patterning of tooth morphology during tooth development, may affect pit-and-fissure and smooth surfaces differently. This hypothesis is

supported by family studies showing that the heritability of dental caries between pit-and-fissure and smooth surfaces is only partly shared [Shaffer et al., 2012]. Investigating pit-and-fissure and smooth surfaces separately may be beneficial for advancing our understanding of cariogenesis.

Although the importance of genetic factors on dental caries is well established [Townsend et al., 1998], and heritability has been estimated to be 30 to 50% [Bretz et al., 2005; Conry et al., 1993; Shaffer et al., 2012; Wang et al., 2010], few specific genes influencing disease risk have been discovered. Previous candidate gene studies and a recent genome-wide association scan [Shaffer et al., 2011] have implicated several loci, although few of these findings have been replicated in follow-up samples or corroborated via functional analyses. The complexity of the phenotype and the possible heterogeneity in risk factors presents obstacles for investigating the genetics of caries. Because specific genes could exert stronger effects on one surface type than the other, and because risk and progression of decay differs by surface type, a reasonable approach to studying the genetic causes of dental caries is to investigate pit-and-fissure and smooth surfaces separately. As part of the Gene Environment Association Studies consortium (GENEVA), we carried out genome-wide association studies (GWAS) to explore the genetic factors affecting dental caries of primary tooth pit-and-fissure and smooth surfaces.

Subjects and Methods

Participants were drawn from two studies: the Center for Oral Health Research in Appalachia (COHRA) [Polk et al., 2008] and the Iowa Fluoride Study (IFS) [Levy et al., 2003]. All participants provided assent with written parental informed consent, and all study procedures were approved by Institutional Review Boards at the pertinent Universities.

All surfaces of all primary teeth were individually scored based on visual inspection for each participant by dental experts who were calibrated across study sites annually. Two classes of tooth surfaces were defined by similarities in morphology and risk of developing carious lesions: pit-and-fissure surfaces and smooth surfaces. The traditional primary dentition caries index, decayed and filled surfaces (dfs), was partitioned to generate two phenotypes: dfs_{PF} and dfs_{SM}, calculated as the summation of surfaces scored as white spot lesion, decayed, or filled, respectively, for pit-and-fissure and smooth surface types.

The COHRA sample was genotyped by the Center for Inherited Disease Research (CIDR) at Johns Hopkins University using the Illumina Human610-Quadv1_B BeadChip (Illumina, Inc., San Diego, CA, USA). Two sub-samples of IFS (denoted as IFS1 and IFS2) were genotyped at different labs. IFS1 was genotyped together with the COHRA sample at CIDR. IFS2 was genotyped separately by the Genetic Analysis Platform at the Broad Institute of MIT and Harvard (Broad) using the same Illumina SNPChip. Additional un-genotyped autosomal SNPs and sporadic missing data of genotyped autosomal SNPs were imputed for COHRA and IFS1 samples using BEAGLE software [Browning and Browning, 2009] and the HapMap Phase III reference panel. To ensure the appropriateness of the reference panel, only study subjects of European ancestry defined via principal components analysis were included for imputation. Imputation quality scores, defined as the average posterior

probability of the most likely genotype across all samples, were high. 531,025 genotyped SNPs passed all the filters and were included for dfs_{SM} analysis and 531,230 for dfs_{PF} analysis. Including imputed SNPs, the numbers were increased to 1,216,189 and 1,216,074, respectively. Because the IFS1 and IFS2 samples were from the same cohort and genotyped using the same platform, they were merged together and mega-analyzed using the participant-level data. Since the COHRA and IFS samples differ in many aspects, including age, social-economic status, and living environment, we combined results using meta-analysis, which eliminates the risk of possible confounding effects introduced by these differences. Specifically, we combined the COHRA and IFS association results using Stouffer's p-value based method of meta-analysis as implemented in the METAL software [Willer et al., 2010].

All the analyses were limited to self-reported white children with genetically verified European ancestry. Teeth having only smooth surfaces, such as central and lateral incisors, erupt and are exfoliated before teeth having pit-and-fissure surfaces, such as molars. Due to this eruption and exfoliation pattern of the primary dentition, different age (at examination) ranges were considered for analyses of smooth and pit-and-fissure surfaces. Children aged 3 to 12 years and having at least one smooth surface present were considered for smooth surface analysis. Children aged 4 to 14 years and having at least one pit-and-fissure surface present were considered for pit-and-fissure surface analysis. A total of 1006 children were included for smooth surface analysis (among them, 919 also having imputed data), and 979 children were included for pit-and-fissure surface analysis (among them, 895 also having imputed data). Due to the household- and community-based study design, the COHRA sample contains some nuclear and extended biological relatives. To maintain statistical power, related samples were not excluded, though theoretically this source of population substructure among participants could lead to inflated p-values. However, by monitoring p-values for evidence of genomic inflation, we found that the inclusion of relatives did not adversely impact the results. Moreover, by re-analyzing SNPs of interest using family-based methods that condition on the relatedness among participants in the sample [Almasy and Blangero, 1998], we observed that unadjusted p-values were very similar to family-adjusted p-values (see supplemental tables). Sample sizes for each subset are shown in Table 1.

Linear regression was used to test the genetic association between each SNP and dfs_{SM} or dfs_{PF} under additive genetic model. Due to the comparatively wider age range in COHRA, the varying number of exposed surfaces, and the varying durations of exposure to cariogenic processes, we adjusted for age and age^2 in the analyses of COHRA sample. In contrast, we only adjusted for age in the analyses of IFS sample, due to the comparatively narrower age range of the IFS sample (from 2.9 to 7.6). A conservative threshold for genome-wide significance was set at $\alpha = 5.0E-8$. A threshold for suggestive significance was set at $\alpha = 1.0E-5$. Genome-wide significant associations, as well as suggestive associations near plausible caries genes, were reported. The effect of non-synonymous SNPs on protein function (tolerated or damaging) was predicted using SIFT [Kumar et al., 2009].

Additional details of the phenotype assessment, genotyping and quality assurance, and statistical analysis are described in the supplemental methods.

Results

Previously, we have reported results of GWAS for a dichotomized (yes/no) childhood caries phenotype in the COHRA and IFS samples [Shaffer et al., 2011]. In the present study, we extend this work by analyzing dfs scores in smooth and pit-and-fissure surfaces, separately, according to our hypothesis that genes may differentially affect risk of caries by surface morphology.

On average, COHRA children had 3.0 carious smooth surfaces (dfs$_{SM}$; SD = 5.5) and 2.5 carious pit-and-fissure surfaces (dfs$_{PF}$; SD = 3.4). In contrast, IFS children had 0.8 carious smooth surfaces (dfs$_{SM}$; SD = 2.6) and 1.0 carious pit-and-fissure surfaces (dfs$_{PF}$; SD = 2.1), reflecting the differences in age, demography and environmental factors between the samples. Additional descriptive statistics for the two samples are summarized in Table 1.

Manhattan and quantile-quantile plots of dfs$_{SM}$ and dfs$_{PF}$ GWAS for combined genotyped and imputed SNPs are shown in Figure 1. The genomic inflation factors (λ) for dfs$_{SM}$ and dfs$_{PF}$ GWAS are 1.031 and 1.028, respectively, indicating moderately inflated p-values. These GWAS scans implicated several novel loci in addition to the putative caries loci identified in previous GWAS of other caries phenotypes (Table 2).

One genome-wide significant association was observed for dfs$_{PF}$ in the intronic region of *KPNA4*, encoding a nuclear protein importer (3q26.1, rs17236529, p-value = 2.0E-9; Fig. 2a). No function of this gene or adjacent genes is currently known to be related to cariogenesis.

In contrast, several suggestive association signals (p-values   1.0E-5) near plausible caries genes were observed. Notably, a missense SNP rs1064524 in *ITGAL* (16p11.2, p-value = 2.1E-7) was suggestively associated with dfs$_{SM}$ (Fig. 2b). The T allele causes an amino acid change from arginine to tryptophan, which was predicted to be damaging with high confidence by SIFT. Although associations are of the same direction for the two samples, the effect sizes differ dramatically. On average, each T allele predicted an increase of 0.54 carious surfaces in the COHRA sample and 2.62 carious surfaces in the Iowa sample. *ITGAL* encodes the integrin alpha L chain of the integrin lymphocyte function-associated antigen-1 (LFA-1). LFA-1 is expressed on all leukocytes and plays a central role in leukocyte intercellular adhesion and also functions in lymphocyte co-stimulatory signaling. A study that investigated the expression of ITGAL in peripheral blood mononuclear cells found a significantly higher level within the CD4(+) and CD8(+) T cells in chronic periodontitis and aggressive periodontitis patients than in healthy controls, suggesting the participation of ITGAL in the pathogenesis of periodontal lesions [Lima et al., 2011]. It is unknown that if ITGAL may affect dental caries in a similar fashion.

A suggestive association with dfs$_{SM}$ was observed for rs17124372 (p-value = 2.0E-6) and several linked SNPs on 20q11.21 in a region harboring 9 genes of the *PLUNC* gene family, such as *PLUNC* and *BASE* (Fig. 2c). BASE is normally expressed only in the salivary gland [Egland et al., 2003], and PLUNC is predominantly expressed in upper airways, nose and mouth. These genes were suggested to be involved in innate immunity and host defense against pathogens in oral and nasal cavities [Bingle and Craven, 2002; Fabian et al., 2012].

While genes in the *PLUNC* family have not previously been implicated in dental caries, their potential roles in oral pathogen defense pose a plausible mechanism by which they could affect risk of caries.

A suggestive association was also observed for *MPPED2* on 11p14.1 in the pit-and-fissure surface scan (rs7121800, p-value = 6.9E-6) (Fig. 2d). This association was the strongest signal in our previous GWAS of primary tooth decay using the binary tooth-level caries outcome (p-value = 1.6E-7) [Shaffer et al., 2011]. *MPPED2* encodes a metallophosphoesterase. Decreased expression of this gene has been reported in oral epithelial cells exposed to periodontopathogens [Milward et al., 2007].

We observed a suggestive association with a region near *AJAP1* on 1p36 for dfs$_{SM}$ (rs4654438, p-value = 1.6E-6) (Fig. 2e). This association represents replication of a result from a recent GWAS involving a non-overlapping subset of adult individuals from the COHRA sample [Shaffer et al., 2013a]. GWAS of novel caries patterns identified by clustering analysis [Shaffer et al., 2013b] in the permanent dentition implicated the region near *AJAP1* for caries of the maxillary pre-molars and canines (p-value = 2.4E-8). Although no study directly links the function of *AJAP1* to dental caries, its product interacts with basigin [Schreiner et al., 2007], a plasma membrane protein expressed in both human and rat tooth germs that appears to serve as an inducer of matrix metalloproteinase activity during tooth development [Kumamoto and Ooya, 2006; Schwab et al., 2007].

We also observed a SNP 88 kb from the 5′ end of *RPS6KA2* suggestively associated with dfs$_{SM}$ (rs3798305, p-value = 7.3E-6) (Fig. 2f). The product of *RPS6KA2* is a kinase in p38-dependent MAPK signaling important for oral-related diseases including dental caries. SNPs in an intron of this gene have been found suggestively associated with both tooth-level caries and smooth surface caries scores in independent studies of the permanent dentition [Wang et al., 2012; Zeng et al., 2013].

Associations with several loci were suggestive for both pit-and-fissure and smooth caries, including the genome-wide significant locus on 3q26.1 (rs17236529, p-value = 2.0E-9 for dfs$_{PF}$, p-value = 3.2E-6 for dfs$_{SM}$), as well as 18q12.2 (rs11082098, p-value = 2.6E-7 for dfs$_{PF}$, p-value = 1.7E-6 for dfs$_{SM}$), and Xq21.2 (rs5967638, p-value = 1.3E-6 for dfs$_{PF}$, p-value = 8.3E-7 for dfs$_{SM}$). No genes at or near these loci have known biological functions obviously relevant to cariogenesis. A full list of SNPs attaining the suggestive significance threshold can be found in the supplemental tables, including family-adjusted p-values for the COHRA sample.

Discussion

To explore our hypothesis that genes may exert differential effects across tooth surfaces of differing morphology, we performed separate GWAS for smooth and pit-and-fissure surface caries in the primary dentition in children sampled from two populations, a high-risk Appalachian population (COHRA), and comparatively lower risk Iowan population (IFS). We performed meta-analyses to combine results of COHRA and IFS. Novel genes and loci

for childhood caries were nominated and putative caries genes previously identified in other studies were replicated.

The association of SNP rs1064524 in *ITGAL* with dfs_{SM} is of particular interest. Not only was the associated variant a missense SNP, but its associations in the two samples were in the same direction, with the predicted damaging allele (T) associated with more severe caries experience. This suggests that rs1064524 might be a causal mutation. Further functional studies are warranted to investigate its role in cariogenesis.

Suggestive associations within or near *AJAP1* and *RPS6KA2* have been observed for permanent dentition caries phenotypes in other studies involving adult participants from the COHRA sample [Shaffer et al., 2013a; Wang et al., 2012]. We detected suggestive association signals near the two genes (but different associated SNPs) for dfs_{SM} in the current study. These findings made the associations of the two genes more convincing because the samples for primary and permanent caries studies comprise different participants from the same underlying population. The two genes could reflect common genetic risk factors for caries in both the primary and permanent dentition.

Although we nominated some different genes for pit-and-fissure and smooth surface caries, we also observed a number of suggestive loci for caries of both types of surfaces. These findings are consistent with previously reported estimates of genetic correlation between dfs_{PF} and dfs_{SM} based on the COHRA sample [Shaffer et al., 2012], suggesting that approximately 58% of the heritability of the two caries scores was explained by common genetic factors and 42% by surface-specific genetic factors.

Moderate genomic inflation was observed ($\lambda = 1.031$ for smooth surface scan and $\lambda = 1.028$ for pit-and-fissure surface scan), which is common for GWAS of complex diseases, and may be due to a variety of potential causes. These include population stratification, known or cryptic relatedness, and model misspecification (such as non-normal phenotype distributions). Additionally, under polygenic inheritance some degree of genomic inflation is expected, depending on sample size, heritability, linkage disequilibrium structure, and the number of true causal variants [Yang et al., 2011]. We speculate that all or some of these causes contribute to the observed genomic inflation in our study. Rather than attempting to identify and adjust for the causes of this inflation, we instead chose to report original p-values. This was because the main aim of the current study was to identify new caries genes and generate hypotheses to be tested in additional studies. Therefore, we focused on the rank of SNPs rather than their nominal p-values. In light of this moderate inflation, caution should be exerted when interpreting p-values from the current GWAS with regard to strict thresholds for statistical significance. Note that, after adjusting for λ, the top SNP for pit-and-fissure surface scan (rs17236529) stays genome-wide significant.

Though the number of participants was modest (for a GWAS), gene-mapping of dental caries is still in its infancy, so pioneering efforts in cohorts such as ours are valuable. Indeed, our sample size of approximately 1,000 was comparable to other GWAS efforts for oral health phenotypes [Divaris et al., 2012; Schaefer et al., 2010; Shaffer et al., 2013a; Shaffer et al., 2011; Wang et al., 2012; Zeng et al., 2013]. We retained related participants in order

to maintain statistical power. While not feasible at the genome-level, family-based likelihood methods [Almasy and Blangero, 1998], which condition of the biological relationships among the sample, were used to confirm the top hits. P-values for unadjusted and family-adjusted associations were not meaningfully different and did not alter interpretation of the results (see supplemental tables).

In summary, we performed GWAS of surface-level caries scores in the primary dentition and nominated *KPNA4*, *ITGAL*, and *PLUNC* family genes as caries susceptibility genes, as well as replicated the associations for *MPPED2*, *AJAP1* and *RPS6KA2*. Replications in additional samples are warranted to confirm the associations of newly nominated genes with dental caries. More studies on the biological functions of these genes are needed to disentangle their roles in cariogenesis. Understanding the genetic nature of caries etiology could ultimately lead to personalized interventions and preventive measures, and could advance the shift in paradigm of dental care from restoration to caries prevention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Acs G, Lodolini G, Kaminsky S, Cisneros GJ. Effect of nursing caries on body weight in a pediatric population. Pediatr Dent. 1992; 14:302–305. [PubMed: 1303533]

Acs G, Shulman R, Ng MW, Chussid S. The effect of dental rehabilitation on the body weight of children with early childhood caries. Pediatr Dent. 1999; 21:109–113. [PubMed: 10197335]

Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. American journal of human genetics. 1998; 62:1198–1211. [PubMed: 9545414]

Anderson M. Risk assessment and epidemiology of dental caries: Review of the literature. Pediatric Dentistry. 2002; 24:377–385. [PubMed: 12412952]

Batchelor PA, Sheiham A. Grouping of tooth surfaces by susceptibility to caries: A study in 5-16 year-old children. BMC Oral Health. 2004; 4:2. [PubMed: 15511295]

Beltran-Aguilar ED, Barker LK, Canto MT, Dye BA, Gooch BF, Griffin SO, Hyman J, Jaramillo F, Kingman A, Nowjack-Raymer R, Selwitz RH, Wu T. Surveillance for dental caries, dental sealants, tooth retention, edentulism, and enamel fluorosis--united states, 1988-1994 and 1999-2002. MMWR Surveill Summ. 2005; 54:1–43. [PubMed: 16121123]

Bingle CD, Craven CJ. Plunc: A novel family of candidate host defence proteins expressed in the upper airways and nasopharynx. Hum Mol Genet. 2002; 11:937–943. [PubMed: 11971875]

Bretz WA, Corby PM, Schork NJ, Robinson MT, Coelho M, Costa S, Melo Filho MR, Weyant RJ, Hart TC. Longitudinal analysis of heritability for dental caries traits. J Dent Res. 2005; 84:1047–1051. [PubMed: 16246939]

Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American journal of human genetics. 2009; 84:210–223. [PubMed: 19200528]

Conry JP, Messer LB, Boraas JC, Aeppli DP, Bouchard TJ Jr. Dental caries and treatment characteristics in human twins reared apart. Arch Oral Biol. 1993; 38:937–943. [PubMed: 8297257]

Divaris K, Monda KL, North KE, Olshan AF, Lange EM, Moss K, Barros SP, Beck JD, Offenbacher S. Genome-wide association study of periodontal pathogen colonization. J Dent Res. 2012; 91:21S–28S. [PubMed: 22699663]

Dye BA, Tan S, Smith V, Lewis BG, Barker LK, Thornton-Evans G, Eke PI, Beltran-Aguilar ED, Horowitz AM, Li CH. Trends in oral health status: United states, 1988-1994 and 1999-2004. Vital Health Stat. 2007; 11:1–92.

Egland KA, Vincent JJ, Strausberg R, Lee B, Pastan I. Discovery of the breast cancer gene base using a molecular approach to enrich for genes encoding membrane and secreted proteins. Proc Natl Acad Sci U S A. 2003; 100:1099–1104. [PubMed: 12538848]

Fabian TK, Hermann P, Beck A, Fejerdy P, Fabian G. Salivary defense proteins: Their network and role in innate and acquired oral immunity. Int J Mol Sci. 2012; 13:4295–4320. [PubMed: 22605979]

Ferraro M, Vieira AR. Explaining gender differences in caries: A multifactorial approach to a multifactorial disease. Int J Dent. 2010; 2010:649643. [PubMed: 20339488]

Hunter PB. Risk factors in dental caries. Int Dent J. 1988; 38:211–217. [PubMed: 3063664]

Jiang H, Tai B, Du M, Peng B. Effect of professional application of apf foam on caries reduction in permanent first molars in 6-7-year-old children: 24-month clinical trial. J Dent. 2005; 33:469–473. [PubMed: 15935266]

Kumamoto H, Ooya K. Immunohistochemical detection of mt1-mmp, reck, and emmprin in ameloblastic tumors. J Oral Pathol Med. 2006; 35:345–351. [PubMed: 16762015]

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the sift algorithm. Nat Protoc. 2009; 4:1073–1081. [PubMed: 19561590]

Levy SM, Warren JJ, Broffitt B, Hillis SL, Kanellis MJ. Fluoride, beverages and dental caries in the primary dentition. Caries Res. 2003; 37:157–165. [PubMed: 12740537]

Lima PM, Souza PE, Costa JE, Gomez RS, Gollob KJ, Dutra WO. Aggressive and chronic periodontitis correlate with distinct cellular sources of key immunoregulatory cytokines. J Periodontol. 2011; 82:86–95. [PubMed: 20809862]

Low W, Tan S, Schwartz S. The effect of severe caries on the quality of life in young children. Pediatr Dent. 1999; 21:325–326. [PubMed: 10509332]

Milward MR, Chapple IL, Wright HJ, Millard JL, Matthews JB, Cooper PR. Differential activation of nf-kappab and gene expression in oral epithelial cells by periodontal pathogens. Clin Exp Immunol. 2007; 148:307–324. [PubMed: 17355248]

Polk DE, Weyant RJ, Crout RJ, McNeil DW, Tarter RE, Thomas JG, Marazita ML. Study protocol of the center for oral health research in appalachia (cohra) etiology study. BMC Oral Health. 2008; 8:18. [PubMed: 18522740]

Schaefer AS, Richter GM, Nothnagel M, Manke T, Dommisch H, Jacobs G, Arlt A, Rosenstiel P, Noack B, Groessner-Schreiber B, Jepsen S, Loos BG, Schreiber S. A genome-wide association study identifies glt6d1 as a susceptibility locus for periodontitis. Hum Mol Genet. 2010; 19:553–562. [PubMed: 19897590]

Schreiner A, Ruonala M, Jakob V, Suthaus J, Boles E, Wouters F, Starzinski-Powitz A. Junction protein shrew-1 influences cell invasion and interacts with invasion-promoting protein cd147. Mol Biol Cell. 2007; 18:1272–1281. [PubMed: 17267690]

Schwab W, Harada H, Goetz W, Nowicki M, Witt M, Kasper M, Barth K. Immunocytochemical and biochemical detection of emmprin in the rat tooth germ: Differentiation-dependent co-expression

with mmps and co-localization with caveolin-1 in membrane rafts of dental epithelial cells. Histochem Cell Biol. 2007; 128:195–203. [PubMed: 17684756]

Shaffer JR, Feingold E, Wang X, Lee M, Tcuenco K, Weeks DE, Weyant RJ, Crout R, McNeil DW, Marazita ML. Gwas of dental caries patterns in the permanent dentition. J Dent Res. 2013a; 92:38–44. [PubMed: 23064961]

Shaffer JR, Feingold E, Wang X, Weeks DE, Weyant RJ, Crout R, McNeil DW, Marazita ML. Clustering tooth surfaces into biologically informative caries outcomes. J Dent Res. 2013b; 92:32–37. [PubMed: 23064960]

Shaffer JR, Wang X, Desensi RS, Wendell S, Weyant RJ, Cuenco KT, Crout R, McNeil DW, Marazita ML. Genetic susceptibility to dental caries on pit and fissure and smooth surfaces. Caries Res. 2012; 46:38–46. [PubMed: 22286298]

Shaffer JR, Wang X, Feingold E, Lee M, Begum F, Weeks DE, Cuenco KT, Barmada MM, Wendell SK, Crosslin DR, Laurie CC, Doheny KF, Pugh EW, Zhang Q, Feenstra B, Geller F, Boyd HA, Zhang H, Melbye M, Murray JC, Weyant RJ, Crout R, McNeil DW, Levy SM, Slayton RL, Willing MC, Broffitt B, Vieira AR, Marazita ML. Genome-wide association scan for childhood caries implicates novel genes. J Dent Res. 2011; 90:1457–1462. [PubMed: 21940522]

Townsend GC, Aldred MJ, Bartold PM. Genetic aspects of dental disorders. Aust Dent J. 1998; 43:269–286. [PubMed: 9775475]

Wang X, Shaffer JR, Weyant RJ, Cuenco KT, DeSensi RS, Crout R, McNeil DW, Marazita ML. Genes and their effects on dental caries may differ between primary and permanent dentitions. Caries Res. 2010; 44:277–284. [PubMed: 20516689]

Wang X, Shaffer JR, Zeng Z, Begum F, Vieira AR, Noel J, Anjomshoaa I, Cuenco KT, Lee MK, Beck J, Boerwinkle E, Cornelis MC, Hu FB, Crosslin DR, Laurie CC, Nelson SC, Doheny KF, Pugh EW, Polk DE, Weyant RJ, Crout R, McNeil DW, Weeks DE, Feingold E, Marazita ML. Genome-wide association scan of dental caries in the permanent dentition. BMC Oral Health. 2012; 12:57. [PubMed: 23259602]

Warren JJ, Levy SM, Broffitt B, Kanellis MJ. Longitudinal study of non-cavitated carious lesion progression in the primary dentition. J Public Health Dent. 2006; 66:83–87. [PubMed: 16711625]

Willer CJ, Li Y, Abecasis GR. Metal: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]

Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, Smith AV, Ingelsson E, O'Connell JR, Mangino M, Magi R, Madden PA, Heath AC, Nyholt DR, Martin NG, Montgomery GW, Frayling TM, Hirschhorn JN, McCarthy MI, Goddard ME, Visscher PM. Consortium G. Genomic inflation factors under polygenic inheritance. European journal of human genetics : EJHG. 2011; 19:807–812. [PubMed: 21407268]

Zeng Z, Shaffer JR, Wang X, Feingold E, Weeks DE, Lee M, Cuenco KT, Wendell SK, Weyant RJ, Crout R, McNeil DW, Marazita ML. Genome-wide association studies of pit-and-fissure- and smooth-surface caries in permanent dentition. J Dent Res. 2013; 92:432–437. [PubMed: 23470693]
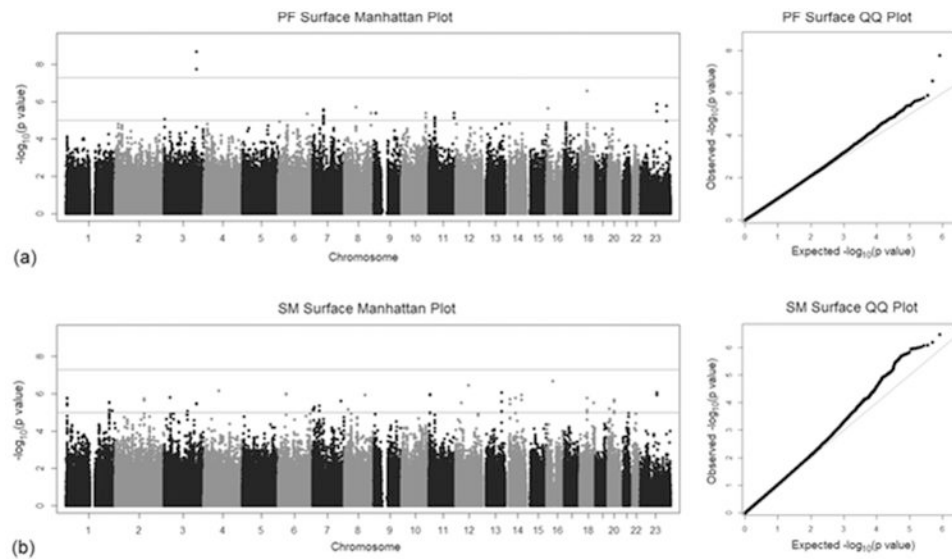
Figure 1.
Manhattan plots (left panel) and quantile-quantile (QQ) plots (right panel) for GWAS of dfs$_{PF}$ and dfs$_{SM}$. **(a)** pit-and-fissure surface scan and **(b)** smooth surface scan. Manhattan plots show -log$_{10}$-transformed p-values (y-axis) across the whole genome (x-axis). The genomic inflation factors (λ) are 1.028 for pit-and-fissure surface scan and 1.031 for smooth surface scan. Percentage of p-values used to calculate λ was set at 0.95. The two horizontal lines in Manhattan plots represent suggestive significance threshold (p-value < 1.0E-5) and genome-wide significance threshold (p-value < 5.0E-8) respectively. Diagonal lines in QQ plots represent the expected distribution of p-values under the null hypothesis of no association. Genotyped and imputed SNPs are plotted together.
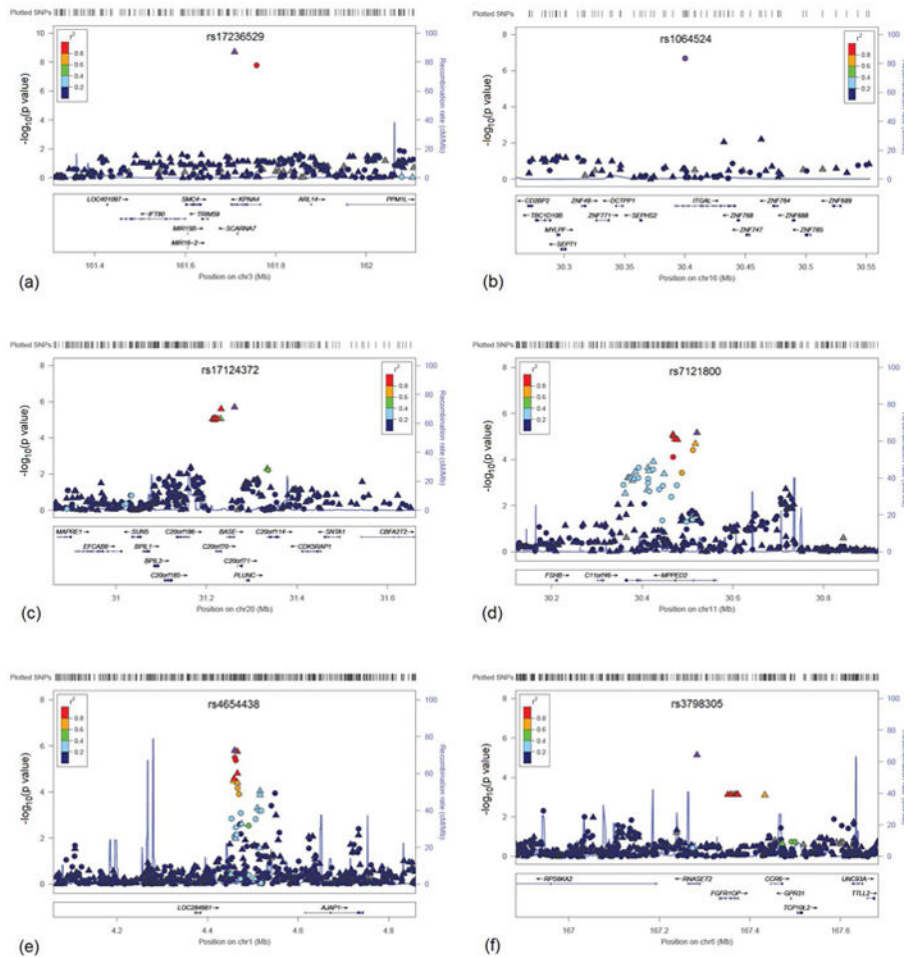
Figure 2.
LocusZoom plots for regions of interest. **(a)** Genome-wide significant locus in *KPNA4* for dfs$_{PF}$; **(b)** Suggestive locus in *ITGAL* for dfs$_{SM}$; **(c)** Suggestive locus near *PLUNC* family genes for dfs$_{SM}$; **(d)** Suggestive locus in *MPPED2* for dfs$_{PF}$; **(e)** Suggestive locus near *AJAP1* for dfs$_{SM}$; **(f)** Suggestive locus near *RPS6KA2* for dfs$_{SM}$. Negative log$_{10}$-transformed p values and physical positions for SNPs in the regions are shown. The recombination rate overlay as well as colors indicating linkage disequilibrium between the top SNP (purple) and other SNPs are based on 1000 Genomes CEU data (hg18). The rug plot indicates regional SNP density. Gene positions and directions of transcription are annotated. Circles represent genotyped SNPs, and triangles represent imputed SNPs.

Table 1

Characteristics of COHRA and IFS samples by caries phenotype.

Phenotype	Sample Characteristic	COHRA	IFS1	IFS2
	Age at exam (yrs)[*]	7.31 ± 2.43	5.13 ± 0.40	
	Age range	3.0-12.0	2.9-7.6	
	Sex ratio (males to females)	1.13	1.00	
	Smooth caries prevalence (%)	48.3%	18.7%	
dfs_{SM}	dfs_{SM}[*]	3.00 ± 5.48	0.81 ± 2.60	
	Genotyping center	CIDR	CIDR	Broad
	Total sample size	596	397	84
	Samples genotyped	547	397	62
	Samples with imputed data	525	394	NA
	Age at exam (yrs)[*]	7.97 ± 2.44	5.13 ± 0.40	
	Age range	4.0-14.0	2.9-7.6	
	Sex ratio (males to females)	1.10	1.00	
	Pit-and-fissure caries prevalence (%)	57.3%	31.2%	
dfs_{PF}	dfs_{PF}[*]	2.51 ± 3.38	1.02 ± 2.05	
	Genotyping center	CIDR	CIDR	Broad
	Total sample size	572	397	84
	Samples genotyped	520	397	62
	Samples with imputed data	501	394	NA

[*] mean ± SD

Table 2

Genome-wide significance and suggestive loci: association results for the top ranking SNP at each locus are shown.

Phenotype	Chr	SNP	BP[1]	Beta (IFS)	Beta (COHRA)	p-value (Meta)	Gene of interest
dfs$_{SM}$	1	rs4654438 [2]	4,459,761	-0.30	-1.63	1.60E-6	*AJAP1*
	6	rs3798305 [2]	167,283,463	2.75	4.83	7.26E-6	*RPS6KA2*
	16	rs1064524	30,400,324	2.62	0.54	2.07E-7	*ITGAL*
	20	rs17124372 [2]	31,262,098	-0.89	-5.20	1.99E-6	*PLUNC family genes*
dfs$_{PF}$	3	rs17236529 [2]	161,709,001	2.81	1.78	2.03E-9	*KPNA4*
	11	rs7121800 [2]	30,521,153	0.49	0.71	6.85E-6	*MPPED2*

[1] base pair position for reference genome build 36.3 hg18

[2] imputed SNP