GWAS of Dental Caries Patterns in the Permanent Dentition

J.R. Shaffer, E. Feingold, X. Wang, M. Lee, K. TCuenco, D.E. Weeks, R.J. Weyant, R. Crout, D.W. McNeil and M.L. Marazita

J DENT RES 2013 92: 38 originally published online 11 October 2012
DOI: 10.1177/0022034512463579

RESEARCH REPORTS
Clinical

J.R. Shaffer[1,2]*, E. Feingold[1,2,3], X. Wang[2,4,5], M. Lee[2,4,5], K. TCuenco[2,4,5], D.E. Weeks[1,3], R.J. Weyant[2,6], R. Crout[2,7], D.W. McNeil[8], and M.L. Marazita[1,2,4,5,9]

[1]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, 130 DeSoto St., Pittsburgh, PA 15261, USA; [2]Center for Oral Health Research in Appalachia, University of Pittsburgh, PA, and West Virginia University, Morgantown, WV, USA; [3]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, PA, USA; [4]Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, PA, USA; [5]Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, PA, USA; [6]Department of Dental Public Health and Information Management, University of Pittsburgh, School of Dental Medicine, Pittsburgh, PA, USA; [7]Department of Periodontics, West Virginia University School of Dentistry, Morgantown, WV, USA; [8]Dental Practice and Rural Health, West Virginia University School of Dentistry, Morgantown, WV, USA; and [9]Clinical and Translational Science Institute, and Department of Psychiatry, School of Medicine, University of Pittsburgh, PA, USA; *corresponding author, jrs51@pitt.edu

J Dent Res 92(1):38-44, 2013

GWAS of Dental Caries Patterns in the Permanent Dentition

ABSTRACT

The importance of susceptibility genes in the risk for dental caries has been clearly established. While many candidate caries genes have been proposed, to date, few of them have been rigorously validated through observational and experimental studies. Moreover, most genetic epidemiological studies have analyzed global caries phenotypes that ignore the possibility that genes may exert differential effects across tooth surfaces of the dentition. Therefore, we performed genome-wide association studies (GWAS) of 5 novel dental caries phenotypes (developed by clustering the permanent dentition into categories of tooth surfaces based on co-occurrence of caries) to nominate new candidate caries genes. GWAS was performed in 920 self-reported white participants, aged 18 to 75 years, with genotype data on 518,997 genetic variants. We identified a significant genetic association between dental caries of the anterior mandibular teeth and *LYZL2* (p value = 9e-9), which codes a bacteriolytic agent thought to be involved in host defense. We also identified a significant genetic association between caries of the mid-dentition tooth surfaces and *AJAP1* (p value = 2e-8), a gene possibly involved in tooth development. Suggestive genetic associations were also observed for *ABCG2, PKD2*, the dentin/bone *SCPP* sub-family, *EDNRA, TJFBR1, NKX2-3, IFT88, TWSG1, IL17D*, and *SMAD7* (p values < 7e-6). We nominate these novel genes for future study.

KEY WORDS: genome-wide association study, genetic association, genomics, hierarchical clustering, tooth decay, cluster analysis.

DOI: 10.1177/0022034512463579

Received June 29, 2012; Last revision September 11, 2012; Accepted September 12, 2012

INTRODUCTION

Dental caries is the most common chronic disease worldwide and affects approximately 90% of adults in the US (Beltran-Aguilar *et al.*, 2005). Many factors contribute to disease, including environmental, endogenous, and behavioral factors such as diet, bacterial flora, fluoride exposures, oral hygiene, salivary flow and composition, morphological and positional characteristics of the dentition, host immune response, and preventive interventions. Genes and gene-by-environmental interactions are widely acknowledged to contribute to caries susceptibility through mechanisms affecting these and other unknown risk factors. The heritability of dental caries is high (30-55%) (Boraas *et al*., 1988; Shaffer *et al*., 2012a), and previous studies, including a genome-wide association study (GWAS) (Shaffer *et al*., 2011) have sought to identify the specific genetic factors involved. However, to date, the causal roles of caries genes have not been rigorously established. The vast majority of genetic variants affecting disease remain unknown.

To date, a potential limitation of most genetic epidemiological studies of dental caries is the global phenotype definition. Most genetics studies have used a single measure of decay, such as a binary (yes/no) affection status or the DMFT/S index (calculated as the sum of decayed, missing due to decay, or filled/restored teeth/surfaces). As global measures of decay, phenotypes such as these ignore the fact that tooth surfaces across the dentition exhibit differences in susceptibility to dental caries and are differentially affected by risk factors. By ignoring the patterns of decay, global caries measures may be poor phenotypes for identifying genetic variants affecting dental caries, which are likely to exert individually weak and differential effects across the various surfaces of the permanent dentition. In a related study presented in this issue of the *Journal*, we describe the use of hierarchical clustering to generate novel caries phenotypes that capture the non-uniform risk of caries across tooth surfaces of the dentition (Shaffer *et al*., 2013). In this study, we performed GWAS for these novel dental caries phenotypes to nominate candidate caries genes for future observational and experimental studies.

METHODS

Sample Recruitment

The sample was recruited as part of an initiative by the Center for Oral Health Research in Appalachia (COHRA) to study the factors affecting oral health in a

Table 1. Novel Dental Caries Phenotypes Generated *via* Cluster Analysis on Tooth-surface-level Caries Data

		Partial DMFS		Heritability			GWAS Results (# loci)	
Phenotype	Tooth Surfaces Included	Mean	SD	h^2	p value	λ	Significant[a]	Suggestive[b]
DMFS1	pit and fissure molar	6.37	3.75	0.27	0.057	1.022	-	-
DMFS2	mandibular anterior (incisors, canines, first pre-molar)	0.66	2.07	0.54	**0.003**	1.030	1	8
DMFS3	posterior (molars and maxillary pre-molars excluding molar pit and fissure)	7.78	8.84	0.43	**0.004**	1.053	0	5
DMFS4	maxillary anterior (incisors)	2.42	3.98	0.00	0.500	0.997	-	-
DMFS5	mid-dentition (pre-molars and canines)	3.70	6.01	0.40	**0.008**	1.047	0	10
$DMFS5_{max}$	maxillary mid-dentition	1.86	3.73	0.34	**0.041**	1.041	1	6
$DMFS5_{mand}$	mandibular mid-dentition	1.84	2.97	0.31	**0.016**	1.038	0	10

[a]p value threshold for genome-wide significant associations = $10^{-7.3^{\lambda}}$.
[b]p value threshold for suggestive associations = $10^{-5^{\lambda}}$.
Mean = mean number of carious surfaces within a cluster.
SD = standard deviation.
h^2 = heritability estimate.
p value = significance of test that heritability estimate differs from zero; bold indicates p value < 0.05.
λ = genomic inflation factor.

rural, underserved population. Details regarding recruitment and data collection for this sample have been described previously (Polk *et al*., 2008) and are summarized in a separate article (Shaffer *et al*., 2013). Nine hundred twenty self-reported whites, aged 18 to 75 yrs, with dental caries assessments and genome-wide genotype data were included in this study.

Novel Caries Phenotypes

To study the genetic factors influencing dental caries, which we hypothesized may exert differential effects across tooth surfaces of the permanent dentition, we first developed novel caries phenotypes that capture the patterns of decay. Development of these novel phenotypes is fully described in a separate article (Shaffer *et al*., 2013). In summary, we used hierarchical clustering to group tooth surfaces based on co-occurrence of dental caries. We hypothesized that tooth surfaces that co-vary with respect to dental caries may be similarly influenced by genetic risk factors. Cluster analysis yielded 5 stable clusters (Table 1), one of which, the mid-dentition surfaces, could be further partitioned into sub-clusters comprised of maxillary and mandibular tooth surfaces. For each cluster (or sub-cluster), we generated the "partial DMFS index", which is the count of carious surfaces (*i.e.*, non-cavitated or cavitated decay, missing due to decay, or restored) for a given cluster. We denote these novel phenotypes as DMFS1, DMFS2, DMFS3, DMFS4, and DMFS5 (which is sub-divided into $DMFS5_{max}$ and $DMFS5_{mand}$).

Genotyping

Genotyping was performed at the Center of Inherited Disease Research (CIDR) of Johns Hopkins University. Data cleaning and quality assurance procedures were conducted in conjunction with the CIDR data cleaning center at the University of Washington. The Illumina Human610-Quadv1_B BeadChip (Illumina, Inc., San Diego, CA, USA) and Illumina Infinium II assay protocol were used. In general, genotyping quality was excellent; 518,997 genetic markers (single-nucleotide polymorphisms; SNPs) passing quality control and analysis filters (participant call rates > 90%; SNP call rates > 99%; Hardy-Weinberg p values > 0.0001; minor allele frequency > 0.02) were analyzed in this study. Details regarding genotyping and data cleaning are publicly available (www.genevastudy.org).

Statistical Analysis

GWAS were performed for each heritable dental caries phenotype (*i.e.*, DMFS2, DMFS3, DMFS5, $DMFS5_{max}$, $DMFS5_{mand}$). Genetic association for each of the 518,997 SNPs was tested by linear regression in PLINK (Purcell *et al*., 2007), adjusted for the effects of sex, age, and age^2. The genomic inflation factor, λ, and Manhattan plots were generated in R (R Foundation for Statistical Computing, Vienna, AU). Given the issue of multiple comparisons, conservative p value thresholds for statistical significance were set equal to 5e-8 after adjustment for genomic inflation (*i.e.*, $p \leq 10^{-7.3^{\lambda}}$). The p value thresholds for suggestive significance were set equal to 1e-5 after adjustment for genomic inflation (*i.e.*, $p \leq 10^{-5^{\lambda}}$). Under the GWAS approach, associated markers are not expected to be causal, but are assumed to be physically proximal and in linkage disequilibrium (*i.e.*, correlated due to ancestry) with unobserved causal variants. Moreover, where a causal variant may be situated with respect to the gene it influences is currently unknown, but, typically, is thought to be also physically proximal. Therefore, we explored and report the known biological functions of genes near our significant and suggestive hits for plausible roles in dental caries.

RESULTS

The study sample included 920 self-reported whites, aged 18 to 75 yrs, with available data for caries phenotypes, covariates (*i.e.*, age, age^2, and sex), and genetic markers. Novel dental

caries phenotypes reflecting decay in different categories of tooth surfaces were generated by cluster analysis. A separate article describes the development and heritability estimation of the phenotypes used herein (Shaffer *et al*., 2013). Descriptive statistics for novel caries phenotypes are shown in Table 1. The genomic inflation factor (λ) ranged from 1.00 to 1.05, indicating minimal to minor inflation of p values compared with that expected by chance alone. Thresholds for genome-wide and suggestive statistical significance were adjusted for the genomic inflation factor observed for each GWAS. Manhattan plots illustrating the GWAS results for the 5 heritable dental caries phenotypes (*i.e*., DMFS2, DMFS3, DMFS5, $DMFS5_{max}$, and $DMFS5_{mand}$) are shown in the Fig.

Two loci exceeded the threshold for genome-wide significance (Table 2). The most significant association that we observed was for DMFS2 (mandibular anterior tooth surfaces) and *LYZL2* on chromosome 10p11.23 (rs399593; p value = 9.4e-9). Though very little is known about *LYZL2* specifically, it belongs to the family of c-type lysosomes, which are well-recognized bacteriolytic factors of host defense (Zhang *et al*., 2005). It is currently unknown whether *LYZL2* affects dental caries, although its putative antibacterial function suggests a biologically plausible mechanism through which *LYZL2* may influence cariogenesis. The second most significant association that we observed was for $DMFS5_{max}$ and the upstream un-translated region of *AJAP1* on chromosome 1p (rs3896439; p value = 2.4e-8). The protein product of *AJAP1*, SHREW1, interacts with basigin, a mediator of matrix metalloprotease (MMP) activity (Schreiner *et al.*, 2007) involved in tooth development in rat (Schwab *et al.*, 2007) and mouse models (Kumamoto and Ooya, 2006). This association was also observed for DMFS5 at the suggestive level of significance (rs3896439; p value = 1.1e-6). Likewise, associated genetic variants upstream of *AJAP1* were among the top hits in an independent GWAS of dental caries in COHRA children (unpublished observations). These lines of evidence suggest that *AJAP1* may influence dental caries, potentially through the regulation tooth development.

In addition to the 2 genome-wide significant associations described above, we also observed several suggestive association signals. The strongest of these was observed for $DMFS5_{mand}$ and a locus on chromosome 4q22.1 harboring several tooth-related genes. The top SNP in this region (rs3114018; p value = 7.478e-8) was in *ABCG2*, a transporter protein and stem-cell marker expressed in human dental pulp (Honda *et al.*, 2007) and in developing mouse incisors (Li *et al*., 2011). Increased expression in ameloblastic tumors suggests that *ABCG2* may regulate the maintenance of odontogenic tissues (Kumamoto and Ohki, 2010). The adjacent gene, *PKD2*, has also been implicated in craniofacial development: *PKD2* mutations in mice cause dental loss, root fractures, and other craniofacial anomalies, and are associated with facial asymmetry in humans (Khonsari; unpublished observations presented at IADR 2012, http://iadr.confex.com/iadr/2012rio/webprogram/Paper160464). The associated locus also includes the 5 paralogous genes of the dentin/bone *SCPP* sub-family (*i.e., SPP1, MEPE, IBSP, DMP1*, and *DSPP*), which are extracellular matrix proteins involved in tooth (and bone) formation (Kawasaki and Weiss, 2008). Genetic variation at this locus may affect one or more of these genes related to tooth development, which, in turn, may affect susceptibility to dental caries.

Many additional suggestive associations were observed, including several in or near genes with biologically plausible roles in caries susceptibility, as summarized in Table 2. Although none of these genes has previously been implicated in dental caries, their putative roles in tooth, salivary gland, and craniofacial development, and in response to bacterial infection, suggest plausible mechanisms through which they may influence cariogenesis. Other suggestive associations were observed for genetic loci that have no obvious biological functions related to dental caries.

DISCUSSION

Dental caries is an extremely multi-factorial disease, whereby genes operating through numerous biological avenues may influence susceptibility. In this study, we performed genome-wide association scans for novel dental caries phenotypes developed under the hypothesis that genetic variants (as is the case with environmental risk factors [Shaffer *et al*., 2012b]) may differentially affect caries risk across tooth surfaces of the permanent dentition. Two genome-wide significant loci were observed: *LYZL2* for mandibular anterior surfaces (*i.e*., surfaces of the incisors, canines, and premolars), and *AJAP1* for mid-dentition surfaces (*i.e*., surfaces of the premolars and maxillary canines). Numerous suggestive loci were also observed.

A theme among the associated loci was genes thought to be involved in craniofacial, salivary gland, and tooth development, including *AJAP1, ABCG2, PKD2*, the dentin/bone *SCPP* subfamily, *EDNRA, TJFBR1, NKX2-3, IFT88, TWSG1*, and *SMAD7*. Developmental genes such as these add to the growing list of host genes, including taste preference, enamel-related, and bactericidal genes, that have been chosen for investigation. In light of the observational study design, we cautiously interpret our results to nominate these loci as putative caries genes, pending confirmation in additional epidemiological and experimental studies. More work, including functional studies, will be required to determine whether associated loci may be useful for disease prediction, diagnosis, or treatment. Either way, the novel associations identified here may lead to better understanding of the biological mechanisms influencing cariogenesis.

For the most part, the significant and suggestive associations identified in this study did not overlap with loci identified *via* GWAS of global caries phenotypes of the permanent dentition in this sample (unpublished observations) or global phenotypes of the primary dentition in a related sample of children from the same population (Shaffer *et al*., 2011). This observation suggests that data-driven caries phenotypes that attempt to capture biologically informative patterns of tooth decay, such as those generated for this study, may be useful for identifying and understanding the genetic contributions to dental caries. Indeed, the X-linked association between DMFS5 and *BCORL1* was also observed in a GWAS of smooth-surface caries, a similar caries phenotype defined based on *a priori* tooth-surface classifications (unpublished observations).

In the interpretation of our results, a few design aspects of the GWAS approach warrant discussion. First, we view GWAS as a

discovery-driven approach, with the goal of generating new hypotheses (as opposed to the paradigm of testing hypotheses). Therefore, in addition to scrutinizing our genome-wide significant hits (which met a very strict significance threshold, even considering the multiple testing), we also examined and report suggestive associations. While the genome-wide significant associations are extremely unlikely to occur by chance alone, associations meeting suggestive significance may occur by chance (though likely not as many as we observed). Therefore, we expect that some of the suggestive associations observed in this study may be spurious, whereas others are likely true reflections of genetic risk. Ideally, these associations should be replicated in an independent sample; however, we analyzed novel caries phenotypes developed specifically for this GWAS from tooth-surface-level caries data not currently available in any other adult sample. Therefore, replication in an independent sample is not currently possible. This limitation is partly countered by the corroborating evidence of gene functions connecting many of our associated loci to tooth-related biological processes. In short, many of our associations make sense, and therefore, we propose prioritizing them among the growing list of caries candidate genes. Additional work is needed to explore the independent or interacting effects of these candidate genes with known environmental risk factors, such as fluoride exposures, dietary and oral hygiene behaviors, and demographics.

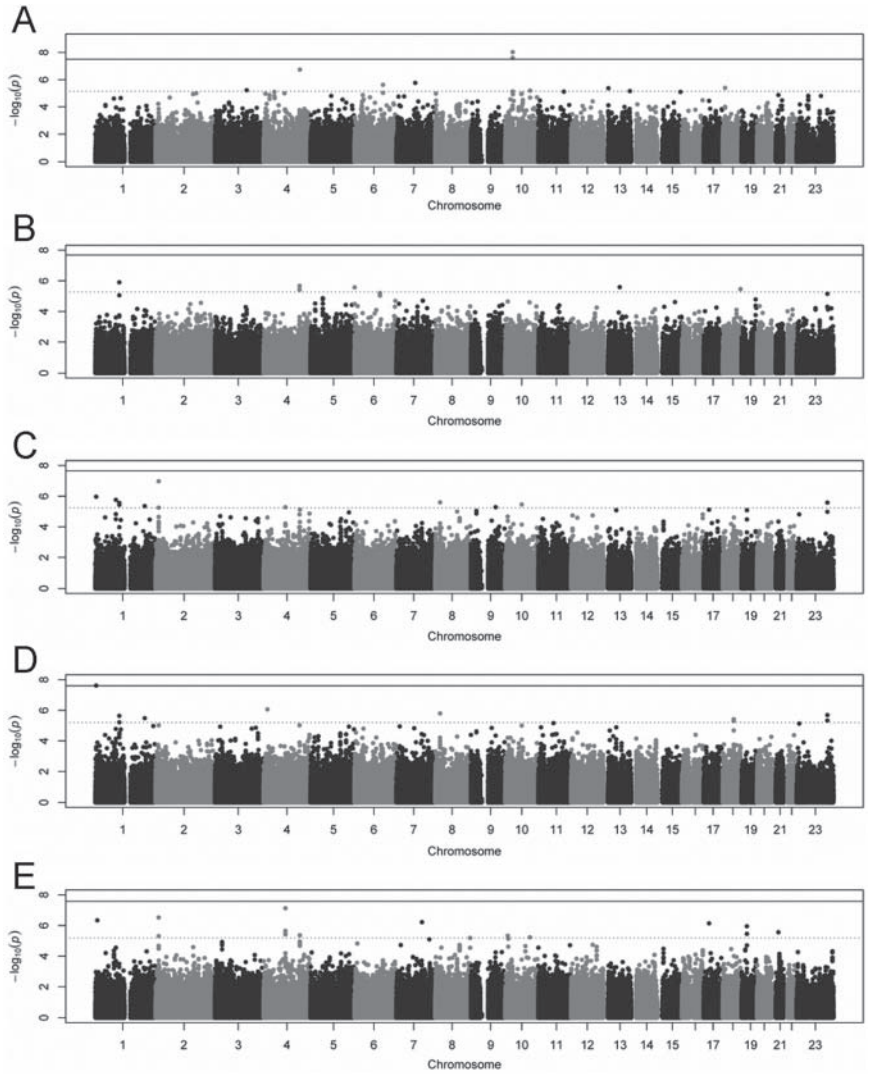

Figure. Manhattan plots showing GWAS results for **(A)** DMFS2, **(B)** DMFS3, **(C)** DMFS5, **(D)** $DMFS5_{max}$, and **(E)** $DMFS5_{mand}$. Solid lines represent thresholds for genome-wide significance (p value < $10^{-7.3}$). Dotted lines represent thresholds for suggestive significance (p value < 10^{-5}).

A major strength and innovation of this study was the use of novel caries phenotypes, which were developed by hierarchical clustering analysis on tooth-surface-level data. We have previously shown that grouping surfaces based on co-occurrence of caries resulted in sensible categories, which were reproducible in NHANES 1999-2000, an independent national cohort (Shaffer *et al*., 2013). Some of the novel caries phenotypes were heritable and/or exhibited significant correlations with putative environmental predictors such as age, sex, educational attainment, and toothbrushing behaviors (Shaffer *et al*., 2013). Here, we showed that these novel caries outcomes were also useful for identifying genetic loci, which may exert differential effects across categories of tooth surfaces. For example, bacteriolytic *LYZL2* and pro-inflammatory *IL17D*, both genes involved in host response to bacterial infection, were implicated in DMFS2, which reflects caries of the mandibular anterior tooth surfaces (*i.e.*, incisors, canines, and first premolars). These surfaces are similar in their positions in the mouth and their comparatively low prevalence of caries. In contrast, several tooth development genes were implicated for DMFS5 (and sub-clusters $DMFS5_{max}$ and $DMFS5_{mand}$), reflecting caries of the mid-dentition surfaces (*i.e.*, pre-molars and canines, which exhibit moderate caries prevalence). DMFS3, which reflects posterior surfaces excluding the molar pit and fissure surfaces, showed associations with a number of genetic loci, though no obvious candidate genes were identified. In contrast, DMFS1 (reflecting molar pit and fissure surfaces, which exhibit the greatest caries prevalence) and DMFS4 (reflecting maxillary incisors, which exhibit moderate caries prevalence) were not significantly heritable, and therefore, GWAS for these outcomes were not considered. These two categories of tooth surfaces are likely strongly influenced by important environmental factors (*e.g.*, sealants and

Table 2. Significant (p value and gene symbol in bold) and Suggestive Associated Loci

Chr.	Phenotype(s)	Top SNP	BP	p value	Gene(s)	Corroborative Evidence/Plausible Role in Dental Caries
1p36	DMFS5$_{max}$, DMFS5	**rs3896439**	**4568530**	**2E-8**	*AJAP1*	**Interacts with basagin (Schreiner *et al.*, 2007), a mediator of *MMP* activity during tooth development (Kumamoto and Ooya, 2006; Schwab *et al.*, 2007)**
1p36	DMFS5$_{mand}$	rs9308447	9354977	5E-7	*SPSB1*	Unknown
1p22	DMFS5	rs1750491	84998887	2E-6	*LPAR3*	Unknown
1p21	DMFS3, DMFS5, DMFS5$_{max}$	rs11166135	99121424	1E-6	*LPPR5*	Unknown
1q32	DMFS5, DMFS5$_{max}$	rs7552806	203050641	3E-6	*NFASC*	Unknown
2p24	DMFS5, DMFS5$_{mand}$	rs10180496	12886351	1E-7	*TRIB2*	Unknown
3q21	DMFS2	rs9810890	130135243	6E-6	*ACAD9*	Unknown
4p15	DMFS5$_{max}$	rs2531154	15627422	9E-7	*PROM1*	Unknown
4q22	DMFS5$_{mand}$, DMFS5	rs3114018	89283605	7E-8	*ABCG2*	Transporter protein and stem-cell marker expressed in human dental pulp (Honda *et al.*, 2007) and in developing mouse incisor (Li *et al.*, 2011); increased expression in ameloblastic tumors suggests that *ABCG2* may regulate the maintenance of odontogenic tissues (Kumamoto and Ohki, 2010)
					PKD2	*Pkd2* mutations in mice caused dental loss, root fractures (unpublished observations presented at IADR 2012 by Khonsari, http://iadr.confex.com/iadr/2012rio/webprogram/Paper160464.html)
					dentin/bone sub-family SCPP (*SPP1, MEPE, IBSP, DMP1*, and *DSPP*)	*SCPP* sub-family consisting of paralogous genes coding extracellular matrix proteins of dentin/bone; shares homology with enamel *SCPP* sub-family (Kawasaki and Weiss, 2008)
4q31	DMFS3	rs11100904	147123754	2E-6	*ZNF827*	Unknown
4q31	DMFS2, DMFS5$_{mand}$	rs1429138	148501792	2E-7	*EDNRA*	Signaling gene expressed during early craniofacial development; mice knock-outs exhibit severe craniofacial defects (Ruest *et al.*, 2004)
6p25	DMFS3	rs2476842	511741	3E-6	*EXOC2*	Unknown
6q22	DMFS2	rs1204798	116650540	2E-6	*NT5DC1* (6q22.1)	Unknown; modest genetic association (p value = 0.02) in a candidate gene study following up linkage signal (Vieira *et al.*, 2008)
7q11	DMFS2	rs848452	77434748	2E-6	*PHTF2*	Unknown
7q22	DMFS5$_{mand}$	rs10242311	105060956	6E-7	*ATXN7L1*	Unknown
8p22	DMFS5$_{max}$, DMFS5	rs10111661	20380853	2E-6	*LZTS1*	Unknown
9q22	DMFS5	rs649057	101274144	5E-6	300 Kb from *TGFBR1*	Strongly expressed in ameloblasts; promotes *MMP20* expression during amelogenesis (Gao *et al.*, 2009)
					300 Kb from *NR4A3*	Transcription factor up-regulated in dental follicle cells during osteogenic differentiation (Morsczeck *et al.*, 2009)
10p14	DMFS5$_{mand}$	rs11256676	10641866	5E-6	*SFTA1P*	Unknown
10p11	**DMFS2**	**rs399593**	**30952036**	**9E-9**	*LYZL2*	**Bacteriolytic factor thought to be involved in host defense (Zhang *et al.*, 2005)**
10q22	DMFS5	rs2441755	67835273	4E-6	*CTNNA3*	Unknown
10q24	DMFS5$_{mand}$, DMFS2	rs7078219	101264355	6E-6	*NKX2-3*	Involved in salivary gland and tooth morphogenesis (Biben *et al.*, 2002)

(continued)

Table 2. (Continued)

Chr.	Phenotype(s)	Top SNP	BP	p value	Gene(s)	Corroborative Evidence/Plausible Role in Dental Caries
13q12	DMFS2	rs735539	20178034	4E-6	*IFT88*	Mutation in *IFT88* leads to increased *SHH* signaling during development, resulting in ectopic extra molars (Ohazama *et al.*, 2009)
					IL17D	Cytokine that enhances pro-inflammatory response during bacterial infection *via* up-regulation of *TLR4* (Guzzo *et al.*, 2012)
13q21	DMFS3	rs2875517	66594146	3E-6	*PCDH9*	Unknown
13q33	DMFS2	rs17485138	108300850	7E-6	*MYO16*	Unknown
17q11	$DMFS5_{mand}$	rs12602978	22591207	7E-7	*WSB1*	Unknown
18p11	DMFS2	rs2864527	9278456	4E-6	*TWSG1*	Regulates *BMP* signaling in the mandibular arch; mouse knock-outs exhibit craniofacial defects and salivary gland dysmorphogenesis (Melnick *et al.*, 2006; MacKenzie *et al.*, 2009)
18q21	$DMFS5_{max}$	rs357894	44833968	4E-6	*SMAD7*	Regulator of *TGF-beta* mediated tooth development (Xu *et al.*, 2003); expressed in human tooth bud (Bao *et al.*, 2003); regulator of enamel deposition (Klopcic *et al.*, 2007)
18q23	DMFS3	rs13381277	72447598	4E-6	*ZNF516*	Unknown
19p12	$DMFS5_{mand}$	rs931608	22405962	1E-6	*ZNF98*	Unknown
21q21	$DMFS5_{mand}$	rs2829459	25201222	3E-6	(gene desert)	Unknown
Xq26	DMFS5, $DMFS5_{max}$	rs3788848	129027493	3E-6	*BCORL1*	Shares homology with *BCOR*, a gene affecting tooth development (Fan *et al.*, 2009); association observed in GWAS of related dental caries phenotype in this cohort

dietary choices) that may render them less informative for discovering genetic risk factors. Overall, we take these results to support our premise that susceptibility genes exert their effects differentially across the permanent dentition and that modeling tooth-surface-level caries data can benefit gene-mapping efforts.

In conclusion, we identified many novel genetic associations for dental caries and nominated several new caries genes. Moreover, we observed that different genetic variants were associated with caries across the different categories of tooth surfaces, which supports our central hypothesis. This study demonstrated the utility of novel dental caries phenotypes for identifying genetic risk factors, which we believe may prove useful for understanding the biological mechanisms predisposing to dental caries, and, ultimately, may lead to improvements in disease prevention, detection, and treatment.

ACKNOWLEDGMENTS

This work was funded through NIH grants U01-DE018903, R01-DE014899, and R03-DE021425. Genotyping was performed as part of the GENEVA consortium (www.genevastudy.org) by the Center for Inherited Disease Research (www.cidr.jhmi.edu) through an NIH contract (HHSN268200782096C). Phenotype harmonization and genotype data cleaning and quality assurance were conducted in conjunction with the GENEVA coordinating center (funded through U01-HG004446). The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

REFERENCES

Bao L, Niu Z, Shi J (2003). [The expression and function of Smad7 during human tooth germ development]. *Hua Xi Kou Qiang Yi Xue Za Zhi* 21:438-440.

Beltran-Aguilar ED, Barker LK, Canto MT, Dye BA, Gooch BF, Griffin SO, *et al*. (2005). Surveillance for dental caries, dental sealants, tooth retention, edentulism, and enamel fluorosis—United States, 1988-1994 and 1999-2002. *MMWR Surveill Summ* 54:1-43.

Biben C, Wang CC, Harvey RP (2002). NK-2 class homeobox genes and pharyngeal/oral patterning: Nkx2-3 is required for salivary gland and tooth morphogenesis. *Int J Dev Biol* 46:415-422.

Boraas JC, Messer LB, Till MJ (1988). A genetic contribution to dental caries, occlusion, and morphology as demonstrated by twins reared apart. *J Dent Res* 67:1150-1155.

Fan Z, Yamaza T, Lee JS, Yu J, Wang S, Fan G, *et al*. (2009). BCOR regulates mesenchymal stem cell function by epigenetic mechanisms. *Nat Cell Biol* 11:1002-1009.

Gao Y, Li D, Han T, Sun Y, Zhang J (2009). TGF-beta1 and TGFBR1 are expressed in ameloblasts and promote MMP20 expression. *Anat Rec (Hoboken)* 292:885-890.

Guzzo C, Ayer A, Basta S, Banfield BW, Gee K (2012). IL-27 enhances LPS-induced proinflammatory cytokine production via upregulation of TLR4 expression and signaling in human monocytes. *J Immunol* 188:864-873.

Honda MJ, Nakashima F, Satomura K, Shinohara Y, Tsuchiya S, Watanabe N, *et al*. (2007). Side population cells expressing ABCG2 in human adult dental pulp tissue. *Int Endod J* 40:949-958.

Kawasaki K, Weiss KM (2008). SCPP gene evolution and the dental mineralization continuum. *J Dent Res* 87:520-531.

Klopcic B, Maass T, Meyer E, Lehr HA, Metzger D, Chambon P, *et al*. (2007). TGF-beta superfamily signaling is essential for tooth and hair morphogenesis and differentiation. *Eur J Cell Biol* 86:781-799.

Kumamoto H, Ohki K (2010). Detection of CD133, Bmi-1, and ABCG2 in ameloblastic tumors. *J Oral Pathol Med* 39:87-93.

Kumamoto H, Ooya K (2006). Immunohistochemical detection of MT1-MMP, RECK, and EMMPRIN in ameloblastic tumors. *J Oral Pathol Med* 35:345-351.

Li L, Kwon HJ, Harada H, Ohshima H, Cho SW, Jung HS (2011). Expression patterns of ABCG2, Bmi-1, Oct-3/4, and Yap in the developing mouse incisor. *Gene Expr Patterns* 11:163-170.

MacKenzie B, Wolff R, Lowe N, Billington CJ Jr, Peterson A, Schmidt B, *et al*. (2009). Twisted gastrulation limits apoptosis in the distal region of the mandibular arch in mice. *Dev Biol* 328:13-23.

Melnick M, Petryk A, Abichaker G, Witcher D, Person AD, Jaskoll T (2006). Embryonic salivary gland dysmorphogenesis in Twisted gastrulation deficient mice. *Arch Oral Biol* 51:433-438.

Morsczeck C, Schmalz G, Reichert TE, Vollner F, Saugspier M, Viale-Bouroncle S, *et al*. (2009). Gene expression profiles of dental follicle cells before and after osteogenic differentiation in vitro. *Clin Oral Investig* 13:383-391.

Ohazama A, Haycraft CJ, Seppala M, Blackburn J, Ghafoor S, Cobourne M, *et al*. (2009). Primary cilia regulate Shh activity in the control of molar tooth number. *Development* 136:897-903.

Polk DE, Weyant RJ, Crout RJ, McNeil DW, Tarter RE, Thomas JG, *et al*. (2008). Study protocol of the Center for Oral Health Research in Appalachia (COHRA) etiology study. *BMC Oral Health* 8:18.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al*. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559-575.

Ruest LB, Xiang X, Lim KC, Levi G, Clouthier DE (2004). Endothelin-A receptor-dependent and -independent signaling pathways in establishing mandibular identity. *Development* 131:4413-4423.

Schreiner A, Ruonala M, Jakob V, Suthaus J, Boles E, Wouters F, *et al*. (2007). Junction protein shrew-1 influences cell invasion and interacts with invasion-promoting protein CD147. *Mol Biol Cell* 18:1272-1281.

Schwab W, Harada H, Goetz W, Nowicki M, Witt M, Kasper M, *et al*. (2007). 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* 128:195-203.

Shaffer JR, Wang X, Feingold E, Lee M, Begum F, Weeks DE, *et al*. (2011). Genome-wide association scan for childhood caries implicates novel genes. *J Dent Res* 90:1457-1462.

Shaffer JR, Feingold E, Wang X, Tcuenco K, Weeks DE, Desensi RS, *et al*. (2012a). Heritable patterns of tooth decay in the permanent dentition: principal components and factor analyses. *BMC Oral Health* 12:7.

Shaffer JR, Polk DE, Feingold E, Wang X, T.Cuenco K, Weeks DE, *et al*. (2012b). Demographic, socioeconomic, and behavioral factors affecting patterns of tooth decay in the permanent dentition: principal components and factor analyses. *Community Dent Oral Epidemiol* (in press).

Shaffer JR, Feingold E, Wang X, Weeks DE, Weyant RJ, Crout R, *et al*. (2013). Clustering tooth surfaces into biologically informative caries outcomes. *J Dent Res* 92:32-37.

Vieira AR, McHenry TG, Daack-Hirsch S, Murray JC, Marazita ML (2008). Candidate gene/loci studies in cleft lip/palate and dental anomalies finds novel susceptibility genes for clefts. *Genet Med* 10:668-674.

Xu X, Jeong L, Han J, Ito Y, Bringas P Jr, Chai Y (2003). Developmental expression of Smad1-7 suggests critical function of TGF-beta/BMP signaling in regulating epithelial-mesenchymal interaction during tooth morphogenesis. *Int J Dev Biol* 47:31-39.

Zhang K, Gao R, Zhang H, Cai X, Shen C, Wu C, *et al*. (2005). Molecular cloning and characterization of three novel lysozyme-like genes, predominantly expressed in the male reproductive system of humans, belonging to the c-type lysozyme/alpha-lactalbumin family. *Biol Reprod* 73:1064-1071.