GENOME-WIDE ASSOCIATION STUDIES OF PIT AND FISSURE SURFACE CARIES AND SMOOTH SURFACE CARIES IN THE PERMANENT DENTITION

by

Zhen Zeng

B.E., Harbin Institute of Technology, China, 2007

M.S., Shanghai Jiao Tong University, China, 2010

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Zhen Zeng

It was defended on

October 1, 2012

and approved by

Thesis Advisor: Eleanor Feingold, Ph.D., Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Committee Member: John R. Shaffer, Ph.D., Research Assitant Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Committee Member: Xiaojing Wang, Ph.D., Research Assitant Professor, Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh

GENOME-WIDE ASSOCIATION STUDIES OF PIT AND FISSURE SURFACE CARIES AND SMOOTH SURFACE CARIES IN THE PERMANENT DENTITION

Zhen Zeng, M.S.

University of Pittsburgh, 2013

ABSTRACT

Dental caries (i.e., tooth decay) is one of the most prevalent public health problems. According to the 1999-2004 National Health and Nutrition Examination Survey, 92% adults 20 to 64 in the United States have had dental caries in their permanent teeth. The important role of genetics in the multi-factorial etiology of dental caries is well-established. However, few specific caries genes have been discovered and validated. Recent family-based analyses have suggested differential genetic factors for primary dentition caries and permanent dentition caries, as well as for pit and fissure surface caries and smooth surface caries. Therefore, to identify genetic variants implicated in dental caries, we performed separate genome-wide association studies for caries indices of the two types of surfaces in the permanent dentition in 1,017 self-reported whites (ages 14 to 56 years), adjusted for the effects of age, sex, presence of *Streptococcus mutans*, and home water fluoride level. Caries indices were derived based on visual assessment of each surface of each tooth. More than 1.2 million SNPs were either successfully genotyped or imputed and were tested for association. Two homologous genes located on separate arms of the X chromosomes were suggestively associated with dental caries: *BCOR* (Xp11.4) was implicated for pit and fissure surfaces (P-value = 3.9E-7), and *BCORL1* (Xq26.1) was implicated for smooth surfaces (P-value = 5.5E-6). Mutations in *BCOR* cause oculofaciocardiodental syndrome, a

Mendelian disease involving multiple dental anomalies, along with other facial and developmental defects. Associations of other genes with plausible biological functions in cariogenesis were also observed for pit and fissure surfaces (e.g., *INHBA*, P-value = 5.1E-6) and for smooth surfaces (e.g., *CXCR1* and *CXCR2*, P-value = 1.5E-6). In summary, this study lends additional support for the notion that genes differentially affect cariogenesis across the surfaces of the permanent dentition, and nominates several novel caries genes for further investigation. Understanding the role of genes in dental caries may ultimately lead to improved detection of high-risk individuals for better and earlier preventative interventions and treatments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I give the sincerest thank to my advisor, Dr. Eleanor Feingold, and my committee members, Dr. John Shaffer and Dr. Xiaojing Wang. I cherish the time working with the awesome team, GENEVA dental caries research group, lead by Dr. Mary Marazita. That was an invaluable experience in my life.

I also thank my family: my parents, my wife and my boy. Their selfless support made my dream possible. This thesis is dedicated to them.

Support of this work was provided by the NIDCR as part of the GENEVA consortium (U01-DE018903; https://www.genevastudy.org/). Genotyping was done by the Center for Inherited Disease Research (http://www.cidr.jhmi.edu/), funded by NIH contract (HHSN268200782096C). Assistance with phenotype harmonization and genotype cleaning was provided by the GENEVA Coordinating Center (U01-HG004446) and by NCBI. Additional support was provided by NIDCR grants R01-DE 014899 and R03-DE021425.

At last, I warmly thank all the participants in the study and thank all the researchers who have contributed to this work.

1.0 INTRODUCTION

Dental caries, also known as tooth decay or cavities, is a demineralization of enamel, dentin and cementum, accompanied with lesions of soft tissues, usually induced by bacterial infection. Cariogenic bacteria process food residues on the tooth surfaces and produce acid to consume healthy teeth. *Streptococcus mutans* is one of the most common cariogenic bacteria. In the early phase of caries development, the affected area of a tooth surface appears to be a white spot, indicating a reversible demineralization. A white spot could be restored to normal by re-mineralization or further develop to a cavity. Once a tooth cavity forms, it is irreversible. If left untreated, caries lesions can lead to pain, tooth loss, and oral infection, or other co-morbidities. Dental caries remains to be one of the most common diseases worldwide. In the US, the prevalence in adults is approximately 90% (Beltran-Aguilar, Barker et al. 2005). According to the 1999-2004 National Health and Nutrition Examination Survey, 92% US adults 20 to 64 have had dental caries in their permanent teeth. In some high-risk populations, the prevalence is even higher. Treatment of dental caries costs significant resources each year.

The development of dental caries is a multi-factorial process, affected by a combination of environmental and behavioral factors, including bacterial flora, fluoride intake and exposures, salivary composition and flow rate, tooth positional and morphological features, dietary behaviors, and oral hygiene. In addition, genetic predisposition and gene-by-environment interactions (Anderson 2002; Mobley, Marshall et al. 2009) also contribute to cariogenesis. The

importance of genetic factors to the risk of dental caries has been well-established through a series of twin studies (Shuler 2001). The estimates of heritability of different caries indices vary from 30% to 50% (Boraas, Messer et al. 1988; Conry, Messer et al. 1993; Bretz, Corby et al. 2005; Wang, Shaffer et al. 2010; Shaffer, Wang et al. 2012), indicating a significant contribution of genes. Therefore, genetic studies of dental caries genetics are of great significance.

Genes hypothetically modify caries risk by affecting host factors in a variety of ways. According to the aspects they may affect, caries susceptibility genes could be classified but not limited to the following categories: immunologic response genes, tooth development and enamel formation genes, taste perception genes, genes influencing saliva secretion and components, and genes influencing oral-health-related behaviors. Candidate gene studies have reported a number of potential genes associated with dental caries in the past decades. However, most of the studies were of small sample size and the results were either inconclusive or lack of replication. Major histocompatibility complex (MHC) class II alleles and *Amelogenin* (*AMELX*) were the two most extensively studies candidates. MHC class II alleles were are immune system components, and were thought to be linked to caries through modifying host defensive ability to cariogenic bacteria. The majority of studies, however, failed to detect significant associations (Acton, Dasanayake et al. 1999; Altun, Guven et al. 2008; Bagherian, Nematollahi et al. 2008; Valarini, Maciel et al. 2012). *AMELX* is a strong biological candidate gene because of its role in tooth enamel development, and mutations in the gene can cause a Mendelian enamel disorder. Despite lots of trying, the genetic association studies of *AMELX* cumulatively show inconclusive results: association with high caries experience has been reported in three populations of Guatemalans, Turkish children, and Koreans living in fluoridated area (Deeley, Letra et al. 2008; Patir, Seymen et al. 2008; Kang, Yoon et al. 2011), but not in Polish and US children (Slayton, Cooper et al.

2005; Olszowski, Adler et al. 2012). Reported candidates of significant association but lack of replication include anti-infection genes *DEFB1* (Ozturk, Famili et al. 2010) and *LTF* (Azevedo, Pecharki et al. 2010), enamel and tooth development genes *MMP13*, *MMP20*, *DSPP*, *KLK4* and *AQP5* (Tannure, Kuchler et al. 2012; Tannure, Kuchler et al. 2012; Wang, Willing et al. 2012), and taste genes *TAS2R38* and *TAS1R2* (Wendell, Wang et al. 2010). Other candidates of biological plausibility but failed in association studies are not listed.

The severity of caries can be assessed quantitatively using a dfs/DMFS system which was initially introduced in 1938 and modified later by World Health Organization. The dfs index counts the total number of tooth surfaces that are decayed or filled due to caries in the primary dentition. The DMFS index counts the total number of tooth surfaces that are decayed, filled due to caries or missing due to caries in the permanent dentition. If pre-cavitated surfaces with white spots are also counted, the indices are denoted as d1fs and D1MFS.

Although dfs and DMFS indices measure overall caries severity very well, they may not always be optimal for genetic studies, in that both the environmental and genetic factors for caries are not the same across tooth surfaces. Grouping tooth surfaces with similar properties together to form novel caries phenotypes for genetic association studies may increase power to detect effects of both genetic and environmental factors. Since it is easy to divide tooth surfaces into two types: pit and fissure (PF) surfaces and smooth (SM) surfaces, the caries indices can be naturally divided into SM dfs/DMFS and PF dfs/DMFS. More complicated caries phenotypes can also be generated by machine learning tools, such as principle component analysis (PCA), factor analysis (FA) or clustering algorithms. Shaffer et al. (Shaffer, Feingold et al. 2012) applied PCA and FA to discover underlying caries patterns in the permanent dentition, and identified a number of interpretable patterns. Interestingly, some of these identified patterns are heritable

while the others are not, which may correspond to environmentally determined caries etiologies and genetically determined ones. These methods are probably useful to "concentrate" genetic effect and thereby increase the power of detecting caries-related genes.

A study assessed genetic correlation between primary and permanent caries phenotypes in 2,600 individuals from 740 families and showed that both shared and unique genetic risk factors may affect dental caries of the primary dentition and permanent dentition (Wang, Shaffer et al. 2010). Two genome-wide association studies (GWAS) were performed separately and nominated different novel biologically plausible caries genes for the two dentitions (Shaffer, Wang et al. 2011; Wang, Shaffer et al. 2012). GWAS in the primary dentition nominated *ACTN2* (possibly involved in enamel formation), *MTR* and *EDARADD* on 1q43, *MPPED2* on 11p14.1 and *LPO* (against oral bacterial) on 17q22. GWAS in the permanent dentition highlighted genes related to two pathways: p38-depenedent MAPK signaling (*RPS6KA2* and *PTK2B*) and Wnt signaling cascade (*RHOU* and *FZD1*), which have been implicated in cariogenesis, and genes involved in tooth development (*ADMTS3* and *ISL1*) and immune response to oral pathogens (*TLR2*).

However, even if all the findings are true positive ones, these genes cumulatively explained only a fraction of the genetic variance of dental caries. Recent work has suggested that the effects of genetic factors on dental caries may also differ between PF and SM tooth surfaces (Shaffer, Wang et al. 2012). The estimated genetic correlation between PF and SM surface caries was significantly less than 100% for the primary dentition (point estimate: 0.76), though was not significant for the permanent dentition (point estimate: 1.00). It is well-established that PF surfaces exhibit much greater risk of developing carious lesions than SM surfaces (Batchelor and Sheiham 2004; Shaffer, Feingold et al. 2012; Shaffer, Wang et al. 2012), and progression of

decay varies between these surface types. In order to further study potential differences in genetic contributions to PF and SM surface caries risk, we performed separate GWAS in the permanent dentition for the two types of surfaces.

2.0 METHODS

2.1 SAMPLES

Subjects were collected as part of a program by the Center for Oral Health Research in Appalachia (COHRA), a partnership between the University of Pittsburgh and West Virginia University, which has been described previously (Polk, Weyant et al. 2008). Poor oral health has been seen in many parts of Appalachia, so the population is of high caries risk. All residents of an eligible household having at least one biological parent-child pair were invited to participate without regard to oral health status or other biological and/or legal relationships. Written informed consent was provided by all adult participants, and assent with parent or guardian written consent was provided by all child participants. All forms and protocols were approved by the COHRA research committee and Institutional Review Boards of the University of Pittsburgh and West Virginia University. To maintain statistical power, we did not exclude related samples. In total, 1,063 self-reported whites (ages 14 to 56 years) had dental caries assessments and data on all covariates (i.e., age, sex, presence of *Streptococcus mutans*, and home water fluoride level).

2.2 PHENOTYPES

Dentists, or research dental hygienists who were annually calibrated to a reference dentist, assessed dental caries in participants by intra-oral examination as previously described in detail (Wang, Shaffer et al. 2010). All surfaces of all permanent teeth were individually scored based on visual inspection as sound (i.e., no evidence of decay), pre-cavitated, decayed (according to a four-level classification of lesion progression), filled, missing due to decay, hypoplastic, or missing due to causes other than decay. Intra-class correlation coefficient (ICC) was applied to quantify the consistency of caries assessments among and within the examiners. Very high concordance rates were observed for both inter-examiner reliability ($0.86 < ICC < 0.99$) and intra-examiner reliability ($ICC > 0.99$) (Wendell, Wang et al. 2010).

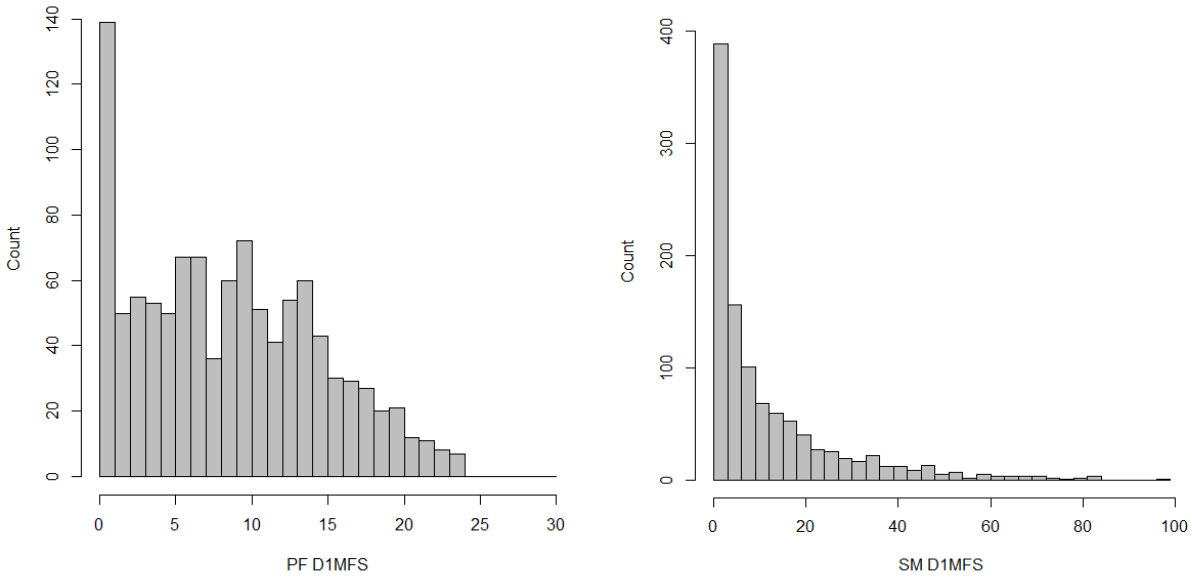

Figure 1. Histograms of caries index

Two classes of tooth surfaces were defined by similarities in morphology and risk of developing carious lesions: PF and SM surfaces. PF surfaces included buccal and occlusal surfaces of mandibular molars and lingual and occlusal surfaces of maxillary molars. SM surfaces included all other tooth surfaces. Two caries indices were generated, PF D1MFS and SM D1MFS, calculated as the total number of surfaces scored as pre-cavitated, decayed, missing due to decay, or filled for PF and SM surface types. Distributions of caries indices are shown in Figure 1.

2.3 COVARIATES

The participant ages used as covariates are the ages at examination. We observed a strong non-linear relationship between age and caries scores, so quadratic terms were included to better model the effect of age. The relationships of caries indices versus ages are shown in Figure 2. Other second-order terms were not considered because there was not any strong indication to include them. The presence of *Streptococcus mutans* was either tested using Dentocult®SM Strip mutans kit with saliva samples or determined genetically using a real-time PCR assay with DNA samples extracted from saliva (Vieira, Deeley et al. 2011). The measure of *Streptococcus mutans* was qualitative (i.e., the covariate was coded as 1 if present, and 0 if not). Fluoride was measured from household water samples provided by each COHRA household with a fluoride-specific electrode. The level of home water fluoride was considered as low if less than 0.7 mg/L and as sufficient if greater than or equal to 0.7 mg/L. The low home water fluoride and sufficient home water fluoride covariates were coded as indicator variables (i.e., low home water fluoride was coded as “1” for those fluoride level less than 0.7 mg/L and as “0” otherwise, and similarly,

sufficient home water fluoride was coded as “1” if fluoride level was greater than or equal to 0.7 mg/L and as “0” otherwise.). A portion of the sample was lacking home water fluoride data, and was coded as “0” for both fluoride covariates.

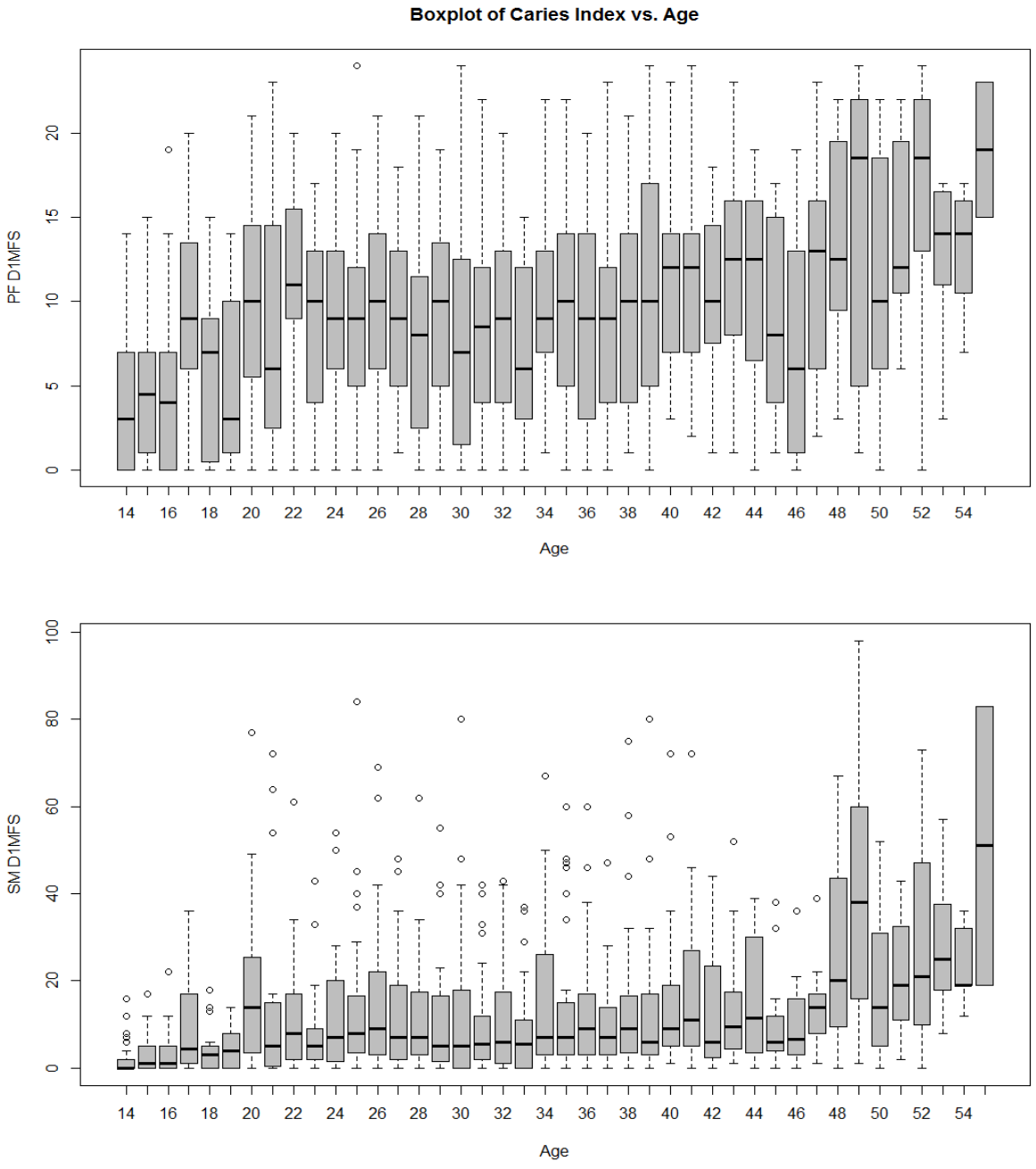

Figure 2. Box plots of caries index versus age

Other potential covariates were excluded because they exhibited non-significant effects in the regression analysis and their missing values would diminish our sample size of GWAS. Table 2 lists the P-values and coefficients for all included covariates.

Table 1. P-values and betas for covariates in linear regression

Covariates		Age	Age^2	Sex	Presence/absence of *Streptococcus mutans*	Low home water fluoride	Sufficient home water fluoride
P-value	SM D1MFS	0.661	0.204	0.414	0.0003	0.962	0.091
	PF D1MFS	0.038	0.840	0.019	0.0025	0.278	0.074
Beta	SM D1MFS	0.109	0.005	0.765	4.055	-0.058	1.770
	PF D1MFS	0.202	-0.0003	0.863	1.312	0.511	0.732

2.4 GENOTYPING AND IMPUTATION

Details regarding allele calling, data cleaning, and quality assurance metrics have been previously described in detail (Shaffer, Wang et al. 2011) and are also available on the Gene Environment Association Studies (GENEVA) consortium website (http://www.genevastudy.org/). We also provide a brief description here. The dataset is publicly available from dbGap (http://www.ncbi.nlm.nih.gov/gap, study accession designation phs000095.v1.p1).

The Illumina Human610-Quadv1_B BeadChip (Illumina, Inc., San Diego, CA, USA) and Illumina Infinium II assay protocol were used for this study. All genotyping was carried out on behalf of the NIH Gene Environment Association Studies (GENEVA) consortium by the Johns

Hopkins University Center for Inherited Disease Research (CIDR). Data cleaning, quality assessment, and imputation were conducted jointly with the GENEVA consortium Coordinating Center. Among 620,901 SNPs released by CIDR, 2,671 SNPs were filtered out due to Hardy-Weinberg equilibrium P-value less than 0.001; 32,417 SNPs were filtered out due to missing rate higher than 10%; 69,818 SNPs were filtered out due to minor allele frequency less than 2%, yielding a total of 548,051 SNPs passing all the quality control filters (548,012 non-mitochondrial SNPs). Of 1,063 individuals, 1,017 were successfully genotyped.

The GENEVA Coordinating Center performed imputation using BEAGLE (Browning and Browning 2009) with the HapMap Phase III reference panel for autosomal SNPs of 996 individuals who were genetically determined as of European ancestry by PCA. Since suggestive SNPs near *BCOR* and *BCORL1* were located on the X chromosome, we performed imputation for these two regions with MaCH (Li, Willer et al. 2010) using the same reference panel. To determine the imputation quality of BEAGLE, masked SNP analysis (i.e., comparison of imputed and experimentally determined genotypes) and analysis of Mendelian inconsistencies for imputed SNPs among relatives were performed. The imputation quality was high. Details are available online in the "GENEVA Dental Caries project imputation report" (http://www.genevastudy.org/docs/DentalCaries_imputation_report_final.pdf). For SNPs imputed by MaCH, only those with Rsq>0.3 were included in association test. A total of more than 1.2 million SNPs were either genotyped or imputed.

2.5 STATISTICAL ANALYSIS AND RESULT ANNOTATION

Linear regression was used to model additive genetic effects while adjusting for the following covariates: age, age^2, sex, presence or absence of *Streptococcus mutans*, low home water fluoride and sufficient home water fluoride. Association between PF or SM surface caries scores and each SNP was tested with PLINK (Purcell, Neale et al. 2007).

Manhattan plots were generated by Haploview (Barrett, Fry et al. 2005). Quantile-quantile plots and genomic inflation factors (λ) were generated by the R statistical package (R Foundation for Statistical Computing, Vienna, AU). We chose to maintain statistical power by including all samples while carefully monitoring type I error, which may be subject to inflation if there is a substantial number of related samples. In addition, we focused on the rank of association signals more rather than their nominal P-values, and the ranks are not affected by relatedness.

We generated scatter plot for -$\log_{10}$ (P-value) from SM and PF surface GWAS scans and calculated squared Pearson correlation coefficient (r^2) to show: 1) an overlap of overall association signal between two traits. Although most of the SNPs are truly without association, we still observed a considerable correlation of log P-values between the two traits (r^2=0.191) and 2) different "detectable" association signals for the two traits (top-ranked SNPs for each trait are off the 45-degree line in Figure 3).

LocusZoom (http://csg.sph.umich.edu/locuszoom/) (Pruim, Welch et al. 2010) was used to plot association signals for loci of interest. Following the adopted standard in the field, we chose a genome-wide significance threshold of α=5E-8, which is an extremely conservative choice based on a Bonferroni correction for one million SNPs. Following the perspective that GWAS is useful as a hypothesis-generating approach (in contrast to a strictly hypothesis-testing

approach) we set our suggestive significance threshold at α=5E-5 in order to avoid Type 1 error and generate a reasonable number of SNPs to be annotated. All genes within the 400kb flanking region of suggestive SNPs were annotated. Based on gene function (such as existing biological data that supported the role of certain genes in tooth development or other caries-related processes, including previous association with dental malformation, previous expression studies in relevant tissues, and previous mouse studies) and proximity, we selected promising genes to report as possible susceptibility genes.

3.0 RESULTS

Sample characteristics are summarized in Table 2. The mean D1MFS scores were 12.1 for SM surfaces and 9.0 for PF surfaces. The caries prevalence was 82% for SM surfaces and 92% for PF surfaces. Females exhibited both more PF and SM surface caries than males (9.3 versus 8.5 and 12.3 versus 11.9, respectively). The squared Pearson correlation coefficient (r^2) for $-\log_{10}$ (P-value) between SM and PF surface GWAS scans was 0.191, indicating that 80.9% of the variation of $\log_{10}$ (P-value) was surface-specific (Figure 3).

Table 2. Sample characteristics

Characteristics	All	Male	Female
Sample size	1,063	408	655
Age at exam (mean ± SD, yrs)	31.46±10.29	31.66±11.24	31.34±9.67
SM D1MFS (mean ± SD)	12.14±15.37	11.88±15.07*	12.29±15.57*
SM caries prevalence (%)	81.9	82.1	81.8
PF D1MFS (mean ± SD)	8.98±6.07	8.50±5.97*	9.28±6.12*
PF caries prevalence (%)	92.0	91.4	92.4
Presenting *Streptococcus mutans* (%)	79.0	79.2	78.9
High fluoride (%)	29.4	31.1	28.2
Low fluoride (%)	19.5	21.1	18.5

*P-values of Wilcoxon rank-sum test for males and females equal median are 0.514 for SM D1MFS and 0.047 for PF D1MFS.

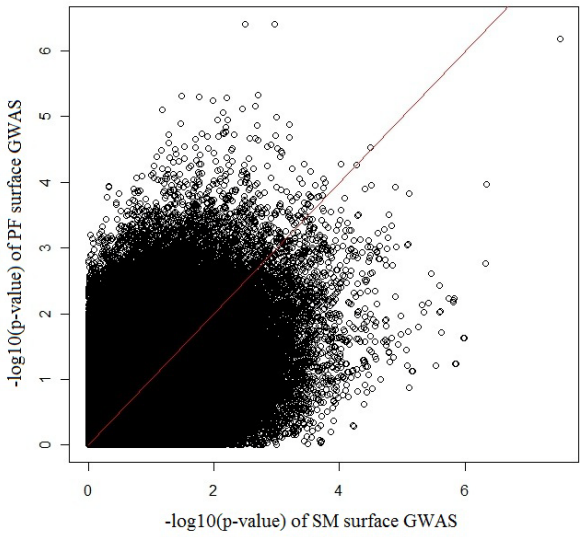

Figure 3. Scatter-plot of -$\log_{10}$ (P-value) from PF and SM surface scans

Red line indicates slope equal to 1.

3.1 PF SURFACE GWAS

No genome-wide significant signal was observed for the PF surface scan. The most significant SNP was rs6560397 on chromosome 9q21.11 (P-value=3.85E-7). This SNP is located in an intron of *PIP5K1B* (phosphatidylinositol-4-phosphate 5-kinase, type I, beta), which is involved in the biosynthesis of a cell membrane signaling protein, but has not previously been implicated in cariogenesis. The genomic inflation factor (λ) was 1.020 for PF surface scan (Figure 4a). This very slight inflation of P-values may be due to related individuals in the sample.

The second ranked SNP for PF D1MFS was rs17145638 on chromosome Xp11.4 (P-value=3.89E-7) located in the 3' downstream region of the *BCOR* gene (BCL6 corepressor) (Figure 5a). BCOR is a key transcription regulator during early embryonic development (Ng, Thakker et al. 2004). Null mutations in *BCOR* are the sole cause of the X-linked dominant

Mendelian disorder oculofaciocardiodental (OFCD) syndrome (Ng, Thakker et al. 2004; Hilton, Johnston et al. 2009), which results in dental abnormalities including radiculomegaly, hypodontia, fusion and duplication of teeth, persistent primary teeth, delayed tooth development and eruption, and defective tooth enamel, as well as other developmental defects (e.g., microphthalmia, congenital cataracts, cardiac and digital defects) (Schulze, Horn et al. 1999; Oberoi, Winder et al. 2005). Bcor expression was observed during mouse development in multiple tissues that are affected in OFCD patients (Wamstad and Bardwell 2007), and normal Bcor expression is required for differentiation of multiple tissue lineages (Wamstad, Corcoran et al. 2008). Further, the osteo-dentinogenic potential was found to be increased in mesenchymal stem cells isolated from an OFCD patient with a *BCOR* mutation (Fan, Yamaza et al. 2009). A study exploring the role of Bcor in tooth development of mice found that silenced *Bcor* expression by RNA interference in dental tissues caused dentinogenesis defects and retardation of tooth root development (Cai, Kwak et al. 2010). No direct role of BCOR in dental caries has been established. However, since null mutations in *BCOR* cause severe dental abnormalities, it is very plausible that genetic variation within or near this gene may impact tooth development so as to modify susceptibility to dental caries.

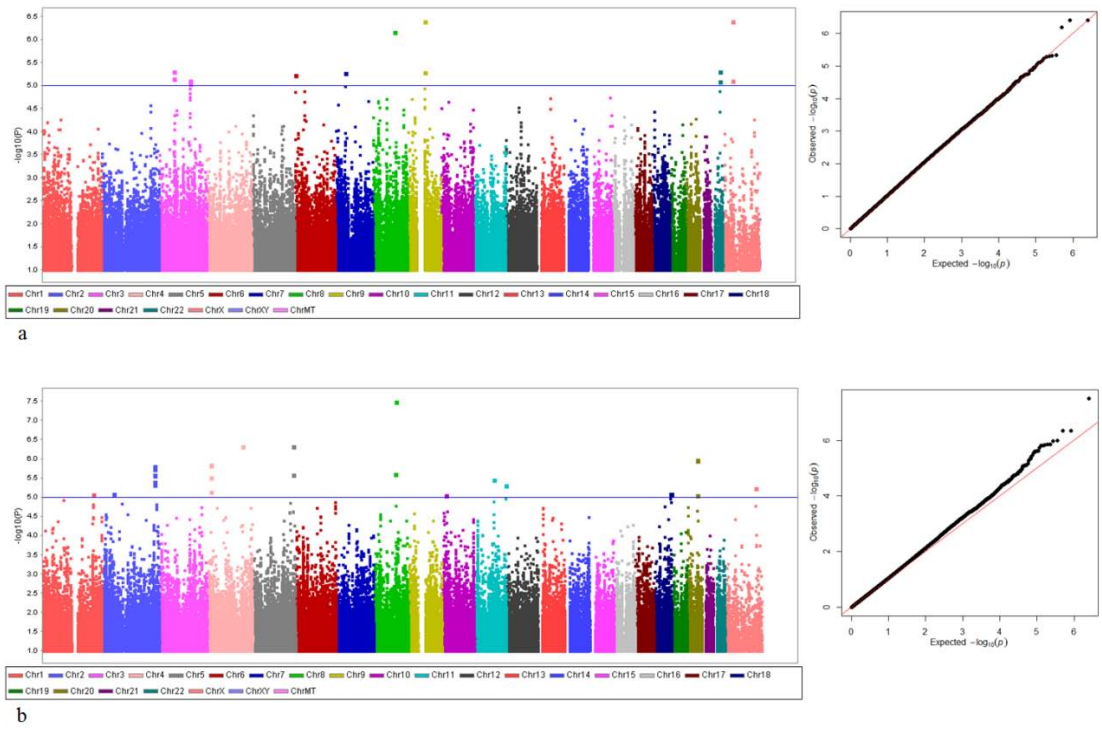

Figure 4. Manhattan plots and quantile-quantile plots for GWAS of PF and SM surface caries

a. PF surface scan; b. SM surface scan. The genomic inflation factors (λ) are 1.020 for PF surface scan and 1.044 for SM surface scan. Percentage of P-values used to calculate λ was set at 0.95. Genotyped and imputed SNPs are plotted together.

Another suggestive gene associated with PF D1MFS was *INHBA* (inhibin, beta A) on chromosome 7p14.1. The most significant SNP in this region was rs10486722, located in the 5' upstream of *INHBA* and also in a non-coding RNA *INHBA-AS1*, with a P-value of 5.14E-6 (Figure 5b). *INHBA* encodes a subunit of activin and inhibin, members of the TGFβ superfamily, that play roles in reproduction and development (Mather, Moore et al. 1997). Expression of Inhba in the mesenchymal cells responsible for early tooth development is essential for tooth bud formation. Mouse knockouts for *Inhba* demonstrate disruption of tooth eruption of incisors and mandibular molars (Ferguson, Tucker et al. 1998; Brown, Houston-Hawkins et al. 2000). Moreover, inhibition of activin signaling in tooth development by exogenous soluble receptors as

well as mutation of activin receptors IIA and IIB, and Smad2, which are effectors in the activin signaling pathway, produced similar dental aberrations (Ferguson, Tucker et al. 2001). Possibly, INHBA participates in the pathogenesis of caries by affecting tooth morphology.

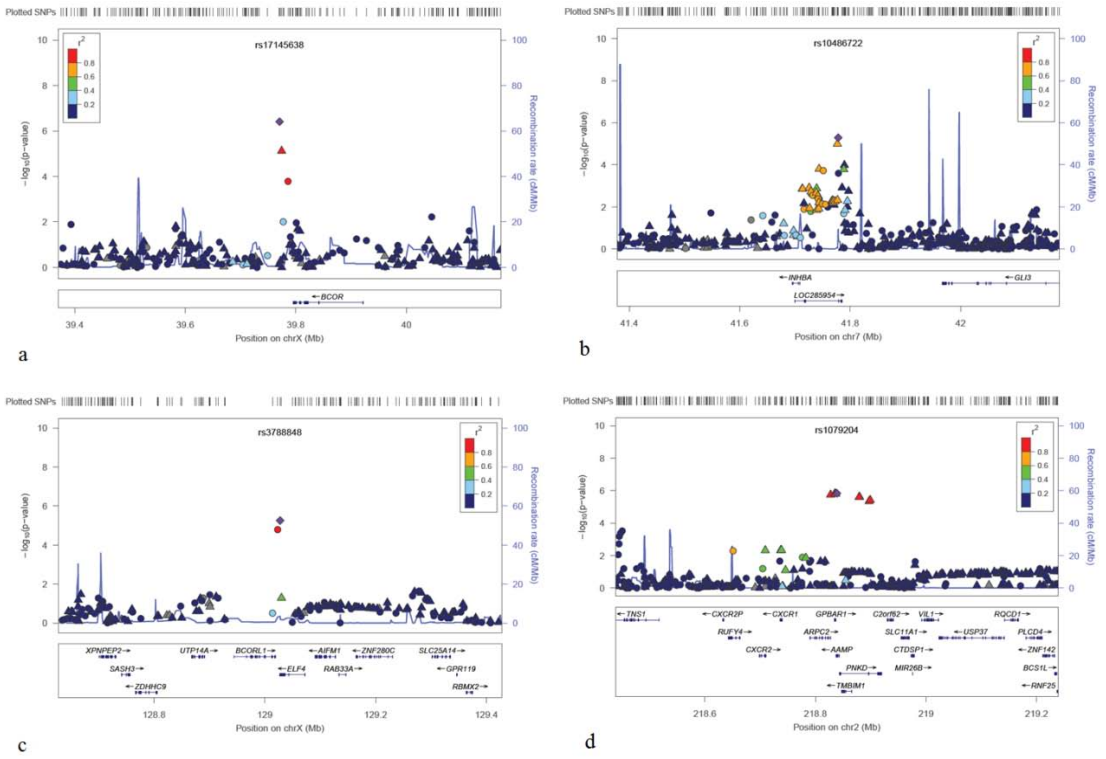

Figure 5. LocusZoom plots for regions of interest

a. Suggestive locus near *BCOR* for PF surface GWAS; b. Suggestive locus near *INHBA* for PF surface GWAS; c. Suggestive locus near *BCORL1* for SM surface GWAS. d. Suggestive locus near *CXCR1* and *CXCR2* for SM surface GWAS. 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 (diamond) and other SNPs are based on HapMap Phase II CEU data (hg18) for chromosome X, or 1000 Genomes CEU data (hg18) for other chromosomes. The rug plot indicates regional SNP density. Gene positions and directions of transcription are annotated. Round points represent genotyped SNPs, and triangles represent imputed SNPs.

See Supplemental Table 1 for a full list of SNPs with P-values less than 5E-5 for the PF surface GWAS.

3.2 SM SURFACE GWAS

The genomic inflation factor (λ) was 1.044 for the SM surface scan (Figure 4b). The most significant association for SM D1MFS was with rs2046315 on chromosome 8q21.3 (P-value=3.08E-8), which was the only SNP meeting genome-wide significance (5E-8). This SNP was also suggestively associated with PF D1MFS (P-value=6.51E-7). However, no known gene is close to this SNP. The nearest gene, *RIPK2* (receptor-interacting serine-threonine kinase 2), is 560 kb away. To further annotate SNP rs2046315, we searched the eQTL database (SNP and CNV Annotation Database, http://www.scandb.org/newinterface/about.html). No gene's expression is predicted by rs2046315 with P-value less than 0.0001 in CEPH (samples with northern and western European ancestry). The Mammal Conservation Score for rs2046315 is -0.128, corresponding to a P-value of 0.745 under a null hypothesis of neutral evolution. In other words, this SNP is not conserved across species. In addition, this SNP does not reside on any genomic element identified by the ENCODE project.

A suggestive association was observed on chromosome Xq26.1 near *BCORL1* (BCL6 corepressor-like 1; rs3788848; P-value=5.54E-6) (Figure 5c). SNP rs3788848 is located in the 3' downstream of *BCORL1* and also in the 3'UTR of *ELF4*. Of note, chromosome Xq has been reported to be linked to low caries susceptibility. In a genome-wide linkage scan, a marker in Xq27.1 yielded a non-parametric P-value of 5E-5 (Vieira, Marazita et al. 2008). *BCORL1* is strong transcriptional repressor so named because of its sequence similarity with *BCOR* (Pagan,

Arnold et al. 2007). Three homologous regions between BCOR and BCORL1 have been identified: a small amino-terminal region and two substantial regions spanning 600 amino acids within the central region and 336 amino acids at the carboxyl terminus, respectively (Pagan, Arnold et al. 2007). No known function of BCORL1 directly relates to cariogenesis. However, it is the only gene similar to *BCOR* in the entire human genome. Furthermore, both the association signals of *BCOR* and *BCORL1* are located on the 3' end of the genes. It is interesting to observe such suggestive associations near two closely related but physically separated genes.

Several imputed SNPs on chromosome 2q35 were associated with SM D1MFS at the suggestive significance level (e.g., rs1079204, located in an intron of *AAMP*; P-value=1.48E-6) near two chemokine receptor genes, *CXCR1* and *CXCR2* (Figure 5d). These genes are major receptors of interleukin 8, a chemokine that is an essential mediator of the inflammation response. A number of studies have shown potential roles of this chemokine and its receptors in the process of oral infection. Increased expression of interleukin 8 has been reported in inflamed human dental pulps (Huang, Potente et al. 1999). *CXCR1* and *CXCR2* mRNA expressions were found to be related to the presence of certain periodontopathic bacteria in inflamed gingival tissues (Noda, Hamachi et al. 2007). Also, *CXCR2* was reported to be associated with periodontitis in a Brazilian population (Viana, Kim et al. 2010). These lines of evidence support the view that *CXCR1* and *CXCR2* contribute to cariogenesis by influencing host susceptibility to oral bacteria.

Supplemental Table 2 provides a full list of SNPs with P-values less than 5E-5 for the SM surface GWAS.

4.0 DISCUSSION

In this study we performed separate GWAS scans to detect genetic variants associated with dental caries of PF and SM surfaces in the permanent dentition. We successfully identified several potential caries genes specific for PF or SM surfaces, including *BCOR* and *BCORL1* (two X-linked genes with sequence similarity), *INHBA*, *CXCR1* and *CXCR2*. Nominated genes were involved in a variety of possible cariogenic processes, such as tooth morphology, tooth development, and immune defense, which is consistent with the prevailing view that the genetic etiology of dental caries includes many genes acting through multiple mechanisms.

Table 3. SNPs of interest showing genetic associations

SNP	CHR	BP	Description	PF P-value	PF beta	SM P-value	SM beta	genotyped or imputed
rs6560397	9	70661732	top hit for PF, *PIP5K1B*	3.85E-07	-1.27	0.003	-1.89	genotyped
rs2046315	8	90280216	top hit for SM, no gene	6.51E-07	1.87	3.08E-08	5.32	genotyped
rs17145638	X	39770574	suggestive hit for PF, *BCOR*	3.89E-07	-2.28	0.001	-3.76	genotyped
rs10486722	7	41778433	suggestive hit for PF, *INHBA*	5.14E-06	1.23	0.006	1.89	genotyped
rs3788848	X	129027493	suggestive hit for SM, *BCORL1*	0.016	0.81	5.54E-06	3.90	genotyped
rs1079204	2	218838758	suggestive hit for SM, *CXCR1* and *CXCR2*	0.006	1.92	1.48E-06	8.44	imputed

Associations shown in this table are all in the same direction for PF and SM surfaces. Top hit: SNP with the smallest P-value in a scan. Suggestive hit: SNP with P-value less than 5E-5 in a scan.

A previous family-based study reported that the genetic correlation between PF and SM surface caries scores did not significantly differ from 100%. Although this estimate of correlation might be inflated, the two surface types are likely to be influenced by the same genetic components (Shaffer, Wang et al. 2012). However, it is still possible to observe different association patterns and to identify different genes for PF and SM surface caries. For instance, under single gene inheritance pattern, where each phenotype is mostly determined by its main-effect gene, the two main-effect genes may be in linkage to each other and transmitted simultaneously within families, resulting in a high genetic correlation. The case of dental caries probably follows polygenic inheritance pattern, where phenotypes are influenced by numerous genes with small to moderate effects. A few genes may be detectable under the study power, but in total their effects are small compared to the total genetic influence. If these detectable genes exert preferential impact on different types of surfaces, and the majority of genetic components are the same for both phenotypes, we can identify distinct genes for different tooth surfaces in spite of high genetic correlation. This scenario fits the findings of our study very well. The P-value scatter plot (Figure 3) shows that most top SNPs for the two surface types are inconsistent, while the correlation of P-values between the two surface types is substantial (r=0.437). Table 3 tabulates the top SNPs within the regions of interest mentioned in the Results section. By comparing the P-values from PF and SM surface scans, we can see that all of the SNPs actually have P-values less than 0.05 for both types of surfaces, but some are more significantly associated (P-values<5E-5) with one type of surface. Therefore, our results do not conflict with the reported high genetic correlation and show that subdivision of complex phenotypes may be more powerful for detecting genes with sub-phenotype-specific effects, even though there is a high genetic correlation between the subdivided phenotypes.

Table 4. Caries indices of the most significant SNPs for X chromosome genes *BCOR* and *BCORL1* by gender and genotype

SNP (phenotype)	Major allele	Minor allele	Gender	Genotype	Genotype count	Phenotype mean	95% CI from t-distribution
rs17145638 (PF D1MFS)	T	C	females	CC	9	6.0	(1.9, 10.1)
				TC	128	7.1	(6.1, 8.2)
				TT	486	9.9	(9.4, 10.5)
			males	C	42	6.9	(5.3, 8.5)
				T	352	8.7	(8.1, 9.3)
rs3788848 (SM D1MFS)	A	G	females	GG	34	22.3	(14.8, 29.7)
				GA	257	13.6	(11.5, 15.6)
				AA	332	10.5	(9.0, 12.0)
			males	G	98	15.1	(11.3, 18.9)
				A	296	11.0	(9.4, 12.6)

Note: not adjusted for covariates.

There is consistent evidence across populations for higher caries incidence in females than males. A similar sex difference was also observed in our current study. While this difference is generally attributed to environmental factors and genetic factors that work through changing individual environmental factors (Lukacs and Largaespada 2006; Lukacs 2010), Vieira et al. (Vieira, Marazita et al. 2008) hypothesized that X-linked genetic variants may partly explain the gender differences in dental caries susceptibility. Our observation of suggestive associations between the X-linked genes *BCOR* and *BCORL1* and caries phenotypes supports their hypothesis. We summarized caries scores by genotypes of the most significant SNPs for the two genes (Table 4). The risk alleles appeared to affect homozygous females more severely than hemizygous males. The effect of X-linked genes on sex differences in dental caries is an open question and deserves further investigation.

Table 5. Overlapping SNPs between PF or SM scan and Wang et al.'s Meta 1 analysis (P-value < 5E-5)

SNP	P-value for different surfaces			Gene
	SM	PF	All	
rs2046315	3.08E-08	6.51E-07	3.28E-06	-
rs1079204	1.48E-06		3.59E-05	*AAMP*
rs2292549	1.53E-06		3.85E-05	*AAMP*
rs2010809	1.03E-06		3.88E-05	*PPIAL*
rs1567869	1.53E-06		4.16E-05	*GPBAR1*
rs7600989	1.80E-06		3.90E-05	*ARPC2*
rs635808	1.62E-05		1.06E-07	*RPS6KA2*
rs4466078	1.85E-05		7.61E-06	*APBB2*

We compared our results with a contemporary GWAS of total dental caries (caries index not divided by tooth surface types) in the permanent dentition conducted by Wang et al. (2012), which was based on the same sample as well as additional cohorts. Table 5 lists overlapping SNPs with P-value < 5E-5 for PF or SM scan and Wang et al.'s Meta 1 analysis, which used compatible DMFS index. There are 112 SNPs with P-value < 5E-5 for total surface caries, 121 SNP for SM, and 64 for PF. Among them, 8 SNPs overlap between SM and total surface caries, and 1 SNP overlaps between PF and total surface caries. It is noteworthy that rs2046315, the only SNP reached genome-wide significance, was also among suggestive for total caries. Two highlighted regions in Wang et al.'s study also showed suggestive associations in ours. One is in an intron of *RPS6KA2* for SM surface caries, and the other is near *PTK2B* for both SM and PF surface caries (Table 6). Both of these two genes are involved in the p38-dependent MAPK pathway important for oral-related diseases including dental caries.

Table 6. Suggestive SNPs in the region for *PTK2B* on chromosome 8

Surface	SNP	Position	P-value
All	rs17057381	27472718	4.02E-07
SM	rs11778371	27375822	3.63E-05
	rs492786	27576103	3.23E-05
	rs542876	27576664	3.23E-05
PF	rs492786	27576103	2.93E-05
	rs542876	27576664	2.93E-05

Although this study has many strengths, including well-defined caries phenotypes, high-quality GWAS data, and important environmental covariates, such as the presence of *Streptococcus mutans* and home water fluoride level, a major limitation is the lack of replication. However, this issue is partly alleviated by the compelling biological evidence for the involvement of the nominated genes in mechanisms related to cariogenesis. In particular, we observed intriguing suggestive associations with *BCOR*, the gene for OFCD syndrome, and *BCORL1*, its only known homologue, for separate caries phenotypes.

5.0 ADDITIONAL AND FUTURE WORK

As an expansion of current project, we also performed GWAS of SM and PF caries in the primary dentition on children in the COHRA sample and an additional sample from Iowa. 547 self-reported whites for SM caries (age from 3 to 12 years) and 520 for PF caries (age from 4 to 14 years) were included from COHRA sample. 459 children were included for both SM and PF caries (age from 2.9 to 7.6 years) from Iowa sample. The two samples were analyzed separately and then meta-analyzed. Preliminary results implicated novel genes and loci as well as re-emphasized some genes and loci identified in previous GWAS.

A missense SNP rs1064524 (on 16p11.2, P-value = 2.07E-07, the allele T causing a change of Arginine to Tryptophan) in *ITGAL* was suggestively associated with SM caries. T allele of rs1064524 was predicted as damaging with high confidence by SIFT. As the number of T alleles increase by 1, 0.54 more caries were predicted among the COHRA sample and 2.62 more caries were predicted among 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 costimulatory signaling. A study investigated the expression of ITGAL in peripheral blood mononuclear cells, and 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, Souza et al. 2011).

A group of SNPs on 20q11.21 (e.g., rs17124372, P-value = 1.99E-06) highlighted several physically adjacent *PLUNC* family genes for SM caries: *BPIFA1*, *BPIFB2*, *BPIFB6*, *BPIFA4P*. BPIFA4P is normally expressed only in salivary gland (Egland, Vincent et al. 2003), and BPIFA1 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, Hermann et al. 2012).

In the previous GWAS of childhood tooth decay conducted by Shaffer et al. with binary dft index, the strongest signal was located in the gene *MPPED2*. *MPPED2* hypothetically encodes a metallophosphoesterase. A decreased expression of this gene has been reported in oral epithelial cells exposed to periodontopathogens (Milward, Chapple et al. 2007). Suggestive association was also observed for this gene in our PF caries scan (rs7121800, on 11p14.1, P-value = 6.85E-06).

As discussed above, SNPs in an intron of *RPS6KA2*, whose product is a kinase in the p38-depenedent MAPK signaling, were suggestively associated with both DMFS index and SM DMFS index in the permanent dentition. In the primary SM caries GWAS, we observed another suggestive signal near the 5' end of this gene (rs3798305, P-value = 7.26E-06). This finding made the association of this region with dental caries much more convincing because the samples for primary dentition caries are totally independent from the samples for primary dentition (they were of non-overlapping age ranges).

Several other loci were nominated for both PF and SM caries, although there are no genes supported by independent biological evidence within these associated regions, including the only

genome-wide significant locus in the primary dentition SM and PF caries scans (on 3q26.1, rs17236529, P-value = 2.03E-09 for PF caries, P-value = 3.23E-06 for SM caries, located in the intronic region of *KPNA4*), 18q12.2 (rs11082098, P-value = 2.63E-07 for PF caries, P-value = 1.68E-06 for SM caries), and Xq21.2 (rs5967638, P-value = 1.31E-06 for PF caries, P-value = 8.33E-07 for SM caries).

The GWAS of surface level caries phenotypes in both the permanent dentition and the primary dentition nominated a number of genes and loci possibly associated with dental caries. Replication of selected SNPs in these interesting genes and loci in additional samples to confirm our findings is in progress. As mentioned in the introduction section, dental caries is a complex disease. Genetic factors may influence cariogenesis through a way that various factors synergize and/or interact with each other. Although simple single-variant association test like ordinary GWAS is good as a start point to explore the role of genetics, more sophisticated methods should be considered in future studies. Systematic biology approaches, such as pathway analysis and gene-gene interaction analysis, are promising to uncover genes that do not have prominent individual effects. It is also of primary interest to collect data on a more thorough list of important environmental variables. Not only would it allow controlling for confounding factors to increase genetic power, but enable investigation of gene-environment interactions. Identification of gene-environment interactions could potentially help explain the observed oral health disparities among different races, genders, and living conditions. Moreover, other than qualitative genetic variations, quantitative change of genes (e.g., copy number variations) may also have impact on caries susceptibility. However, this is still an untouched area so far. Finally, if any of the suggestively associated genes/loci are validated in the follow-up studies, it might be appropriate to apply sequencing analysis to search for causal variants in the targeted regions. In

summary, given that the nature of cariogenesis is extremely complicated, future studies should use a combination of various approaches and take advantage of recently emerged techniques to investigate genetics of caries in a systematical manner.

APPENDIX A

SUPPLEMENTAL TABLE 1

Table S1: 64 SNPs for PF surface scan meeting suggestive significance (i.e., P-value<5E-5)

CHR	BP	SNP	A1	A2	A1_freq	BETA	P-value	imputed
2	205437485	rs12328369	A	T	0.0203	4.2669	2.64E-05	Yes
2	205437540	rs12327977	C	T	0.9833	-4.2294	3.55E-05	Yes
3	59976454	rs17061812	A	G	0.0463	2.9733	6.83E-06	Yes
3	59976865	rs9311745	C	T	0.0466	2.998	4.72E-06	Yes
3	62666659	rs13068742	A	T	0.1053	-1.7624	4.23E-05	Yes
3	72389275	rs10212587	A	T	0.7369	-1.2786	4.12E-05	Yes
3	72393455	rs6549449	C	T	0.739	-1.2854	3.39E-05	Yes
3	129723644	rs7433900	A	G	0.1996	1.411	1.72E-05	No
3	129728399	rs9819402	A	C	0.1704	1.4133	4.74E-05	Yes
3	129734368	rs4431128	C	T	0.8307	-1.4716	2.01E-05	Yes
3	129743240	rs4857855	T	C	0.1701	1.413	3.83E-05	No
3	129759438	rs4857907	A	G	0.1832	1.5072	8.69E-06	Yes
3	129761493	rs2335050	T	C	0.1814	1.486	1.12E-05	No
3	129765387	rs6806253	A	G	0.8295	-1.5382	7.60E-06	Yes
3	129768184	rs17344939	T	C	0.1519	1.567	1.44E-05	No
3	129772898	rs2713589	A	G	0.146	1.576	1.98E-05	No
3	129774903	rs2734046	A	G	0.1507	1.5583	1.76E-05	Yes
5	5288028	rs2913630	T	C	0.2281	1.267	4.28E-05	No
6	511741	rs2476842	T	C	0.4887	1.084	1.36E-05	No
6	522820	rs9504361	G	A	0.4508	1.155	5.75E-06	No
6	39338549	rs4711589	C	A	0.323	-1.146	2.17E-05	No
6	39338991	rs11961538	C	T	0.3511	-1.147	1.31E-05	Yes

Supplemental Table S1 continued

6	39354834	rs2758873	A	G	0.3294	-1.0998	4.34E-05	Yes
7	11479349	rs2189310	C	T	0.2579	1.235	2.54E-05	No
7	41776860	rs10951660	A	G	0.3896	1.1499	1.02E-05	Yes
7	41778433	rs10486722	C	T	0.3353	1.229	5.14E-06	No
7	137238909	rs273969	A	G	0.2992	1.3709	2.13E-05	Yes
8	503226	rs6984685	C	T	0.4336	1.1421	4.72E-05	Yes
8	22295813	rs12545568	G	A	0.1839	1.397	2.09E-05	No
8	22319611	rs7819400	G	A	0.1927	1.348	2.27E-05	No
8	27576103	rs492786	T	C	0.381	1.085	2.93E-05	No
8	27576664	rs542876	A	G	0.381	1.085	2.93E-05	No
8	56645014	rs13251620	C	T	0.7297	1.2634	2.88E-05	Yes
8	56897479	rs7829716	A	G	0.0262	3.5011	1.93E-05	Yes
8	56928693	rs10087354	A	G	0.9742	-3.4428	3.02E-05	Yes
8	90280216	rs2046315	T	C	0.1323	1.871	6.51E-07	No
8	127035340	rs10956273	A	G	0.9348	2.1794	4.09E-05	Yes
8	127037469	rs10505464	C	A	0.06441	-2.215	3.24E-05	No
9	17407517	rs2593395	A	G	0.07115	2.139	1.88E-05	No
9	27792678	rs10114706	C	T	0.9687	-3.0564	4.73E-05	Yes
9	27800414	rs4878586	C	T	0.9637	-3.0649	4.94E-05	Yes
9	70577584	rs10511964	A	C	0.4717	-1.0538	3.01E-05	Yes
9	70580991	rs12350270	A	G	0.4891	1.1475	1.14E-05	Yes
9	70658077	rs10869420	A	C	0.467	1.072	2.81E-05	No
9	70661732	rs6560397	T	C	0.4671	-1.268	3.85E-07	No
9	70703782	rs4745375	T	G	0.266	-1.214	1.89E-05	No
9	70704367	rs883951	A	G	0.7277	1.2917	5.03E-06	Yes
10	1502856	rs6560737	T	C	0.208	-1.338	3.07E-05	No
10	29330964	rs11007352	A	C	0.9264	2.034	2.20E-05	Yes
10	132821242	rs10829913	C	A	0.3552	1.126	3.22E-05	No
12	55732112	rs324021	C	A	0.4322	1.071	3.57E-05	No
12	55736077	rs1044931	C	T	0.5674	-1.0882	2.94E-05	Yes
13	58708493	rs4884323	G	T	0.4002	1.126	1.82E-05	No
13	58708732	rs9570071	A	G	0.3736	1.091	3.11E-05	No
15	91788851	rs11074186	A	G	0.1629	-1.533	1.81E-05	No
15	96681892	rs1982259	A	G	0.5974	1.8146	4.30E-05	Yes
16	50806428	rs12932608	C	T	0.3622	1.0813	4.67E-05	Yes

Supplemental Table S1 continued

18	11283890	rs987890	A	C	0.1141	-1.647	3.60E-05	No
22	35528968	rs5750309	T	C	0.3181	1.207	1.29E-05	No
22	35529752	rs2022068	A	G	0.3926	1.1999	4.85E-06	Yes
22	35536287	rs4820254	G	T	0.4928	-1.4255	7.83E-06	Yes
22	35538606	rs5750310	A	C	0.567	-1.3629	3.65E-05	Yes
23	39770574	rs17145638	C	T	0.1146	-2.275	3.89E-07	No
23	39774306	rs5917340	C	G	0.879	1.5646	7.57E-06	Yes

APPENDIX B

SUPPLEMENTAL TABLE 2

Table S2:121 SNPs for SM surface scan meeting suggestive significance (i.e., P-value<5E-5)

CHR	BP	SNP	A1	A2	A1_freq	BETA	P-value	imputed
1	62770189	rs10458569	A	C	0.0449	6.3993	4.10E-05	Yes
1	84998887	rs1750491	A	G	0.2832	3.238	1.15E-05	No
1	208435759	rs12094311	G	A	0.1495	3.832	4.70E-05	No
1	208437213	rs12060567	A	G	0.1495	3.832	4.70E-05	No
1	208956814	rs11119577	A	T	0.1501	3.9314	1.75E-05	Yes
1	208960258	rs6681860	A	C	0.8535	-3.8715	3.43E-05	Yes
1	208961270	rs12022982	A	G	0.1347	4.14	1.46E-05	No
1	208963316	rs7555360	T	G	0.1363	3.956	3.23E-05	No
1	208964372	rs10863852	A	C	0.1347	4.2168	8.15E-06	Yes
1	208964886	rs1934620	A	G	0.8654	-4.2193	8.07E-06	Yes
1	208966802	rs10863853	A	G	0.8644	-4.2148	8.24E-06	Yes
1	244336191	rs6681900	T	C	0.08161	4.912	4.11E-05	No
1	244336569	rs6689428	G	A	0.08014	5.011	3.24E-05	No
2	15302935	rs9287655	T	C	0.4071	2.802	2.71E-05	No
2	15344624	rs4668892	C	T	0.5575	-2.8142	2.98E-05	Yes
2	15345075	rs4668893	C	T	0.439	2.7489	3.71E-05	Yes
2	20657560	rs7595132	A	G	0.1013	4.535	3.17E-05	No
2	48956648	rs995146	A	C	0.9306	-7.5904	7.82E-06	Yes
2	85192530	rs3893079	C	A	0.05315	6.289	1.43E-05	No
2	108245143	rs1470874	A	G	0.3612	2.851	2.65E-05	No
2	185640560	rs263767	G	A	0.2448	-3.093	3.85E-05	No
2	218827379	rs7600989	G	T	0.9607	-8.0519	1.80E-06	Yes
2	218834980	rs1567869	C	T	0.0421	8.0696	1.53E-06	Yes

Supplemental Table S2 continued

2	218836750	rs2292549	C	T	0.0402	8.0618	1.53E-06	Yes
2	218838758	rs1079204	A	G	0.0383	8.4379	1.48E-06	Yes
2	218878706	rs1017697	A	G	0.0408	7.8069	2.56E-06	Yes
2	218879706	rs2014597	A	G	0.0409	7.8261	2.45E-06	Yes
2	218896913	rs6708662	C	T	0.0428	7.5719	4.52E-06	Yes
2	218898957	rs10192690	A	G	0.9586	-7.9391	3.79E-06	Yes
2	229584711	rs16825564	A	G	0.9031	-4.6921	1.47E-05	Yes
2	229591741	rs7563172	A	C	0.0959	4.6516	1.77E-05	Yes
2	229593322	rs10490035	C	A	0.09587	4.664	2.01E-05	No
3	70017021	rs17006578	A	G	0.99	-14.6429	3.28E-05	Yes
3	126170457	rs1909586	G	T	0.3829	2.9058	3.59E-05	Yes
3	176307233	rs4894477	A	G	0.4094	2.741	2.69E-05	No
3	176310488	rs6782155	A	G	0.41	2.803	1.80E-05	No
4	15626996	rs2286458	G	T	0.0678	5.9978	2.94E-06	Yes
4	15627380	rs2677789	A	C	0.06749	5.826	6.98E-06	No
4	15627422	rs2531154	G	A	0.06742	5.828	6.82E-06	No
4	15630636	rs1829271	A	G	0.9355	-6.3516	1.42E-06	Yes
4	15631638	rs2677780	A	G	0.9355	-6.3597	1.40E-06	Yes
4	40668284	rs4466078	A	T	0.9857	-17.7148	1.85E-05	Yes
4	66795224	rs4289486	A	G	0.0531	6.057	3.86E-05	No
4	78101455	rs17002297	C	T	0.7842	-3.3618	3.14E-05	Yes
4	78121149	rs4241597	A	G	0.783	-3.3451	3.30E-05	Yes
4	147095001	rs723794	G	T	0.2679	3.671	4.50E-07	No
4	182431572	rs7678192	C	T	0.3105	3.1141	1.83E-05	Yes
4	189458201	rs6830572	G	A	0.2653	-3.054	4.94E-05	No
5	131900972	rs739718	C	T	0.0745	5.1647	4.04E-05	Yes
5	156123792	rs1845479	G	A	0.03638	7.542	2.43E-05	No
5	156124842	rs11953631	T	C	0.03593	7.572	2.53E-05	No
5	158485488	rs1582508	A	G	0.4857	2.869	1.37E-05	No
5	158500688	rs2420355	A	C	0.437	-2.6798	4.61E-05	Yes
5	170173001	rs11134654	A	C	0.1753	4.4237	4.63E-07	Yes
5	170176011	rs1422160	A	G	0.1691	4.15	2.54E-06	No
5	170195941	rs11745293	C	T	0.8506	-3.8709	2.26E-05	Yes
5	170202257	rs888811	C	T	0.1472	3.9066	2.39E-05	Yes
6	88894335	rs9353524	C	G	0.7984	-3.5507	1.87E-05	Yes
6	104569248	rs2205748	C	T	0.3967	2.855	3.04E-05	No
6	104599177	rs9377596	C	T	0.3787	2.7822	4.42E-05	Yes
6	104604086	rs9373743	C	G	0.3793	2.9285	3.22E-05	Yes
6	167085776	rs10946186	T	C	0.4966	-2.781	1.92E-05	No

Supplemental Table S2 continued

6	167086923	rs9295368	A	G	0.5766	-2.7905	2.50E-05	Yes
6	167096924	rs388372	T	C	0.354	3.013	1.29E-05	No
6	167097412	rs635808	T	C	0.2262	-3.345	1.62E-05	No
8	26137422	rs4275231	T	C	0.1647	3.698	2.67E-05	No
8	27375822	rs11778371	T	C	0.05211	-6.027	3.63E-05	No
8	27576103	rs492786	T	C	0.381	2.761	3.23E-05	No
8	27576664	rs542876	A	G	0.381	2.761	3.23E-05	No
8	67695752	rs2467750	A	G	0.04331	6.697	4.03E-05	No
8	90062604	rs10429371	C	T	0.2144	3.812	2.40E-06	No
8	90084328	rs1487791	A	G	0.735	-3.3248	1.61E-05	Yes
8	90087157	rs12676566	C	G	0.735	-3.3197	1.65E-05	Yes
8	90280216	rs2046315	T	C	0.1323	5.319	3.08E-08	No
8	134554009	rs7835464	A	G	0.04769	6.357	3.91E-05	No
9	22783225	rs10811775	G	A	0.05129	6.386	2.64E-05	No
9	101274144	rs649057	T	G	0.04916	6.221	3.99E-05	No
10	5391308	rs7081234	C	T	0.5232	2.6468	3.88E-05	Yes
10	5394347	rs10904478	C	T	0.404	-2.7551	2.52E-05	Yes
10	5414727	rs11253143	A	G	0.6058	2.8528	3.93E-05	Yes
10	10404535	rs4747852	G	T	0.4523	-2.698	4.96E-05	No
10	10425015	rs7087371	C	T	0.3585	-2.9829	3.11E-05	Yes
10	13893858	rs7088455	A	G	0.06146	6.1	8.61E-06	No
10	17294838	rs7080366	T	C	0.44	-2.665	4.51E-05	No
10	21607442	rs12358291	T	G	0.06637	5.575	2.27E-05	No
10	129997306	rs1255136	A	G	0.5116	2.7105	3.60E-05	Yes
10	130034129	rs10829385	A	G	0.5134	2.7292	4.10E-05	Yes
10	130035334	rs7079081	G	T	0.5133	2.7352	3.97E-05	Yes
10	130047859	rs1936418	A	G	0.8165	3.5652	4.37E-05	Yes
11	13226368	rs10832006	A	G	0.7494	3.0821	3.99E-05	Yes
11	13276470	rs7949336	G	A	0.2493	-3.062	4.17E-05	No
11	80889400	rs11232701	C	G	0.0948	5.323	1.26E-05	Yes
11	80904206	rs10897833	G	A	0.09046	4.789	4.33E-05	No
11	80910473	rs4944304	C	A	0.09145	4.895	2.70E-05	No
11	80912798	rs2032381	G	T	0.9212	-5.6853	3.43E-06	Yes
11	132607863	rs7933745	A	G	0.1608	-3.838	1.07E-05	No
11	132626147	rs2078454	A	C	0.2085	-3.636	4.77E-06	No

Supplemental Table S2 continued

13	19971094	rs7335998	C	G	0.0116	12.6267	1.89E-05	Yes
13	19972980	rs9506503	A	C	0.0136	12.4914	2.69E-05	Yes
13	19973813	rs6490590	C	T	0.9885	-12.6436	1.83E-05	Yes
13	20183009	rs6490605	A	G	0.2474	3.1401	3.45E-05	Yes
13	58925818	rs17056606	A	G	0.9895	-13.2634	3.77E-05	Yes
13	62834221	rs9570849	A	G	0.3265	2.86	3.41E-05	No
13	98117502	rs7996576	G	A	0.1211	4.171	4.68E-05	No
14	100827448	rs4906092	C	T	0.12	4.134	3.19E-05	No
18	45458441	rs8082881	G	A	0.0531	6.343	1.71E-05	No
18	61341576	rs12962841	A	G	0.4288	-2.8026	3.70E-05	Yes
18	67590053	rs17085106	G	T	0.984	-11.6204	1.07E-05	Yes
18	67592624	rs8099373	A	G	0.9843	-12.0501	7.95E-06	Yes
18	72447295	rs13381274	C	T	0.0667	6.0139	1.29E-05	Yes
18	72447598	rs13381277	G	A	0.06244	6.082	7.86E-06	No
20	4320446	rs6084726	A	G	0.5252	2.8423	1.80E-05	Yes
20	4327086	rs3904864	A	G	0.4705	-2.741	2.93E-05	No
20	41425015	rs2010809	C	T	0.9884	-14.4122	1.03E-06	Yes
20	41435864	rs6103260	G	T	0.9885	-14.396	1.06E-06	Yes
20	41444367	rs6017025	A	G	0.9868	-11.8564	2.21E-05	Yes
20	41447413	rs6103268	C	G	0.0129	12.773	8.42E-06	Yes
20	44347669	rs6017764	C	T	0.9758	-8.9563	3.97E-05	Yes
23	41123863	rs787084	C	T	0.4146	3.008	3.57E-05	No
23	129022933	rs6637684	T	C	0.2463	3.754	1.60E-05	No
23	129027493	rs3788848	G	A	0.2579	3.899	5.54E-06	No

BIBLIOGRAPHY

Acton, R. T., A. P. Dasanayake, et al. (1999). "Associations of MHC genes with levels of caries-inducing organisms and caries severity in African-American women." Hum Immunol 60(10): 984-989.

Altun, C., G. Guven, et al. (2008). "Human leukocyte antigen class II alleles and dental caries in a child population." Pediatr Dent 30(2): 154-159.

Anderson, M. (2002). "Risk assessment and epidemiology of dental caries: review of the literature." Pediatr Dent 24(5): 377-385.

Azevedo, L. F., G. D. Pecharki, et al. (2010). "Analysis of the association between lactotransferrin (LTF) gene polymorphism and dental caries." J Appl Oral Sci 18(2): 166-170.

Bagherian, A., H. Nematollahi, et al. (2008). "Comparison of allele frequency for HLA-DR and HLA-DQ between patients with ECC and caries-free children." J Indian Soc Pedod Prev Dent 26(1): 18-21.

Barrett, J. C., B. Fry, et al. (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics 21(2): 263-265.

Batchelor, P. A. and A. Sheiham (2004). "Grouping of tooth surfaces by susceptibility to caries: a study in 5-16 year-old children." BMC Oral Health 4(1): 2.

Beltran-Aguilar, E. D., L. K. Barker, 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(3): 1-43.

Bingle, C. D. and C. J. Craven (2002). "PLUNC: a novel family of candidate host defence proteins expressed in the upper airways and nasopharynx." Hum Mol Genet 11(8): 937-943.

Boraas, J. C., L. B. Messer, et al. (1988). "A genetic contribution to dental caries, occlusion, and morphology as demonstrated by twins reared apart." J Dent Res 67(9): 1150-1155.

Bretz, W. A., P. M. Corby, et al. (2005). "Longitudinal analysis of heritability for dental caries traits." J Dent Res 84(11): 1047-1051.

Brown, C. W., D. E. Houston-Hawkins, et al. (2000). "Insertion of Inhbb into the Inhba locus rescues the Inhba-null phenotype and reveals new activin functions." Nat Genet 25(4): 453-457.

Browning, B. L. and S. R. Browning (2009). "A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals." Am J Hum Genet 84(2): 210-223.

Cai, J., S. Kwak, et al. (2010). "Function analysis of mesenchymal Bcor in tooth development by using RNA interference." Cell Tissue Res 341(2): 251-258.

Conry, J. P., L. B. Messer, et al. (1993). "Dental caries and treatment characteristics in human twins reared apart." Arch Oral Biol 38(11): 937-943.

Deeley, K., A. Letra, et al. (2008). "Possible association of amelogenin to high caries experience in a Guatemalan-Mayan population." Caries Res 42(1): 8-13.

Egland, K. A., J. J. Vincent, et al. (2003). "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 100(3): 1099-1104.

Fabian, T. K., P. Hermann, et al. (2012). "Salivary defense proteins: their network and role in innate and acquired oral immunity." Int J Mol Sci 13(4): 4295-4320.

Fan, Z., T. Yamaza, et al. (2009). "BCOR regulates mesenchymal stem cell function by epigenetic mechanisms." Nat Cell Biol 11(8): 1002-1009.

Ferguson, C. A., A. S. Tucker, et al. (1998). "Activin is an essential early mesenchymal signal in tooth development that is required for patterning of the murine dentition." Genes Dev 12(16): 2636-2649.

Ferguson, C. A., A. S. Tucker, et al. (2001). "The role of effectors of the activin signalling pathway, activin receptors IIA and IIB, and Smad2, in patterning of tooth development." Development 128(22): 4605-4613.

Hilton, E., J. Johnston, et al. (2009). "BCOR analysis in patients with OFCD and Lenz microphthalmia syndromes, mental retardation with ocular anomalies, and cardiac laterality defects." Eur J Hum Genet 17(10): 1325-1335.

Huang, G. T., A. P. Potente, et al. (1999). "Increased interleukin-8 expression in inflamed human dental pulps." Oral Surg Oral Med Oral Pathol Oral Radiol Endod 88(2): 214-220.

Kang, S. W., I. Yoon, et al. (2011). "Association between AMELX polymorphisms and dental caries in Koreans." Oral Dis 17(4): 399-406.

Li, Y., C. J. Willer, et al. (2010). "MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes." Genet Epidemiol 34(8): 816-834.

Lima, P. M., P. E. Souza, et al. (2011). "Aggressive and chronic periodontitis correlate with distinct cellular sources of key immunoregulatory cytokines." J Periodontol 82(1): 86-95.

Lukacs, J. R. (2010). "Sex differences in dental caries experience: clinical evidence, complex etiology." Clin Oral Investig.

Lukacs, J. R. and L. L. Largaespada (2006). "Explaining sex differences in dental caries prevalence: saliva, hormones, and "life-history" etiologies." Am J Hum Biol 18(4): 540-555.

Mather, J. P., A. Moore, et al. (1997). "Activins, inhibins, and follistatins: further thoughts on a growing family of regulators." Proc Soc Exp Biol Med 215(3): 209-222.

Milward, M. R., I. L. Chapple, et al. (2007). "Differential activation of NF-kappaB and gene expression in oral epithelial cells by periodontal pathogens." Clin Exp Immunol 148(2): 307-324.

Mobley, C., T. A. Marshall, et al. (2009). "The contribution of dietary factors to dental caries and disparities in caries." Acad Pediatr 9(6): 410-414.

Ng, D., N. Thakker, et al. (2004). "Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in BCOR." Nat Genet 36(4): 411-416.

Noda, D., T. Hamachi, et al. (2007). "Relationship between the presence of periodontopathic bacteria and the expression of chemokine receptor mRNA in inflamed gingival tissues." J Periodontal Res 42(6): 566-571.

Oberoi, S., A. E. Winder, et al. (2005). "Case reports of oculofaciocardiodental syndrome with unusual dental findings." Am J Med Genet A 136(3): 275-277.

Olszowski, T., G. Adler, et al. (2012). "MBL2, MASP2, AMELX, and ENAM gene polymorphisms and dental caries in Polish children." Oral Dis 18(4): 389-395.

Ozturk, A., P. Famili, et al. (2010). "The antimicrobial peptide DEFB1 is associated with caries." J Dent Res 89(6): 631-636.

Pagan, J. K., J. Arnold, et al. (2007). "A novel corepressor, BCoR-L1, represses transcription through an interaction with CtBP." J Biol Chem 282(20): 15248-15257.

Patir, A., F. Seymen, et al. (2008). "Enamel formation genes are associated with high caries experience in Turkish children." Caries Res 42(5): 394-400.

Polk, D. E., R. J. Weyant, et al. (2008). "Study protocol of the Center for Oral Health Research in Appalachia (COHRA) etiology study." BMC Oral Health 8: 18.

Pruim, R. J., R. P. Welch, et al. (2010). "LocusZoom: regional visualization of genome-wide association scan results." Bioinformatics 26(18): 2336-2337.

Purcell, S., B. Neale, et al. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet 81(3): 559-575.

Schulze, B. R., D. Horn, et al. (1999). "Rare dental abnormalities seen in oculo-facio-cardio-dental (OFCD) syndrome: three new cases and review of nine patients." Am J Med Genet 82(5): 429-435.

Shaffer, J. R., E. Feingold, et al. (2012). "Heritable patterns of tooth decay in the permanent dentition: principal components and factor analyses." BMC Oral Health 12(1): 7.

Shaffer, J. R., X. Wang, et al. (2012). "Genetic susceptibility to dental caries on pit and fissure and smooth surfaces." Caries Res 46(1): 38-46.

Shaffer, J. R., X. Wang, et al. (2011). "Genome-wide association scan for childhood caries implicates novel genes." J Dent Res 90(12): 1457-1462.

Shuler, C. F. (2001). "Inherited risks for susceptibility to dental caries." J Dent Educ 65(10): 1038-1045.

Slayton, R. L., M. E. Cooper, et al. (2005). "Tuftelin, mutans streptococci, and dental caries susceptibility." J Dent Res 84(8): 711-714.

Tannure, P. N., E. C. Kuchler, et al. (2012). "MMP13 polymorphism decreases risk for dental caries." Caries Res 46(4): 401-407.

Tannure, P. N., E. C. Kuchler, et al. (2012). "Genetic variation in MMP20 contributes to higher caries experience." J Dent 40(5): 381-386.

Valarini, N., S. M. Maciel, et al. (2012). "Association of Dental Caries with HLA Class II Allele in Brazilian Adolescents." Caries Res 46(6): 530-535.

Viana, A. C., Y. J. Kim, et al. (2010). "Association of haplotypes in the CXCR2 gene with periodontitis in a Brazilian population." DNA Cell Biol 29(4): 191-200.

Vieira, A. R., K. B. Deeley, et al. (2011). "Detection of Streptococcus mutans Genomic DNA in Human DNA Samples Extracted from Saliva and Blood." ISRN Dent 2011: 543561.

Vieira, A. R., M. L. Marazita, et al. (2008). "Genome-wide scan finds suggestive caries loci." J Dent Res 87(5): 435-439.

Wamstad, J. A. and V. J. Bardwell (2007). "Characterization of Bcor expression in mouse development." Gene Expr Patterns 7(5): 550-557.

Wamstad, J. A., C. M. Corcoran, et al. (2008). "Role of the transcriptional corepressor Bcor in embryonic stem cell differentiation and early embryonic development." PLoS One 3(7): e2814.

Wang, X., J. R. Shaffer, et al. (2010). "Genes and their effects on dental caries may differ between primary and permanent dentitions." Caries Res 44(3): 277-284.

Wang, X., J. R. Shaffer, et al. (2012). "Genome-wide association Scan of dental caries in the permanent dentition." BMC Oral Health 12(1): 57.

Wang, X., M. C. Willing, et al. (2012). "Genetic and environmental factors associated with dental caries in children: the Iowa Fluoride Study." Caries Res 46(3): 177-184.

Wendell, S., X. Wang, et al. (2010). "Taste genes associated with dental caries." J Dent Res 89(11): 1198-1202.