RESEARCH REPORTS
Biological

Z. Zeng[1,3], J.R. Shaffer[1]*, X. Wang[2], E. Feingold[1,3], D.E. Weeks[1,3], M. Lee[2], K.T. Cuenco[1,2], S.K. Wendell[2], R.J. Weyant[4], R. Crout[5], D.W. McNeil[6], and M.L. Marazita[1,2,7]*

[1]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; [2]Center for Craniofacial & Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; [3]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; [4]Department of Dental Public Health and Information Management, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; [5]Department of Periodontics, School of Dentistry, West Virginia University, Morgantown, WV, USA; [6]Dental Practice and Rural Health, West Virginia University, Morgantown, WV, USA; and [7]Clinical and Translational Science, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; *corresponding authors, marazita@pitt.edu, jrs51@pitt.edu

Genome-wide Association Studies of Pit-and-Fissure- and Smooth-surface Caries in Permanent Dentition

J Dent Res DOI: 10.1177/0022034513481976

APPENDIX

MATERIALS & METHODS

Samples

Study participants were identified 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 *et al.,* 2008). 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 (all participants provided assent with written parental informed consent, and all study procedures were approved by Institutional Review Boards at the pertinent universities). To maintain statistical power, we did not exclude related samples. In total, 1,063 self-reported whites had dental caries assessments and data on all covariates for the PF-surface analysis, while 1,049 had caries and covariate data for the SM-surface analysis. Among them, 1,017 individuals had genotype data and were included in the PF-surface scan, while 1,004 were included in the SM-surface scan. Imputation of autosomal SNPs was done only for individuals who were of genetically determined European ancestry as defined by principal component analysis (996 individuals in the PF-surface scan, and 982 individuals in the SM-surface scan).

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 *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. Inter- and intra-examiner concordance was high (Wendell *et al*., 2010).

Covariate Selection and Description

Eight variables in our dataset were considered as potential covariates in construction of the models for the genetic association tests: age at examination, sex, education group, saliva flow, source of water (*i.e.,* city/public, well, and other), home water fluoride level, toothbrushing frequency, and the presence of *Streptococcus mutans*. We applied a forward step-wise selection strategy to determine the covariate(s) in the linear regression models for PF- and SM-surface analyses, respectively. First, age was forced to be in the initial models, since its effect had been established previously. Each of the other covariates as well as age^2 was added to the model and tested individually. The most significant covariate (with the smallest partial F-test p value) was retained in the model. Then, the procedure was repeated until no incoming covariate was significant (partial F-test p value < .05). In the final models, age, sex, and the presence of *Streptococcus mutans* were included for PF D1MFS, and age, education group, and the presence of *Streptococcus mutans* were

Appendix Table 1. 66 SNPs for PF Surface Scans Meeting Suggestive Significance (*i.e.*, *p* value < 5E-5)

CHR	BP	SNP	A1	A2	A1_freq	Beta	*p*	Imputed
1	75533665	rs10493567	C	T	0.9505	–2.4614	3.61E-05	Yes
2	205437485	rs12328369	A	T	0.0203	4.3552	1.75E-05	Yes
2	205437540	rs12327977	C	T	0.9833	–4.3196	2.36E-05	Yes
3	59976454	rs17061812	A	G	0.0463	2.9681	7.18E-06	Yes
3	59976865	rs9311745	C	T	0.0466	2.9914	5.02E-06	Yes
3	62666659	rs13068742	A	T	0.1053	–1.7671	4.00E-05	Yes
3	72389275	rs10212587	A	T	0.7369	–1.2793	4.17E-05	Yes
3	72393455	rs6549449	C	T	0.739	–1.2861	3.42E-05	Yes
3	129687641	rs2335052	A	G	0.1474	1.501	4.14E-05	No
3	129723644	rs7433900	A	G	0.1994	1.443	1.07E-05	No
3	129728399	rs9819402	A	C	0.1704	1.4483	3.01E-05	Yes
3	129734368	rs4431128	C	T	0.8307	–1.4839	1.73E-05	Yes
3	129743240	rs4857855	T	C	0.1698	1.431	3.04E-05	No
3	129759438	rs4857907	A	G	0.1832	1.5327	6.14E-06	Yes
3	129761493	rs2335050	T	C	0.1813	1.511	7.87E-06	No
3	129765387	rs6806253	A	G	0.8295	–1.5595	5.67E-06	Yes
3	129768184	rs17344939	T	C	0.1522	1.58	1.22E-05	No
3	129772898	rs2713589	A	G	0.1465	1.595	1.58E-05	No
3	129774903	rs2734046	A	G	0.1507	1.5674	1.59E-05	Yes
5	5288028	rs2913630	T	C	0.2295	1.289	2.91E-05	No
6	511741	rs2476842	T	C	0.4871	1.077	1.57E-05	No
6	522820	rs9504361	G	A	0.4504	1.156	5.77E-06	No
6	39338549	rs4711589	C	A	0.323	–1.157	1.82E-05	No
6	39338991	rs11961538	C	T	0.3511	–1.1486	1.30E-05	Yes
6	39354834	rs2758873	A	G	0.3294	–1.1184	3.24E-05	Yes
6	39368775	rs2815060	A	T	0.6714	1.1066	3.86E-05	Yes
7	11479349	rs2189310	C	T	0.2555	1.259	1.75E-05	No
7	41776860	rs10951660	A	G	0.3896	1.1296	1.46E-05	Yes
7	41778433	rs10486722	C	T	0.3349	1.215	6.50E-06	No
7	137238909	rs273969	A	G	0.2992	1.3448	3.02E-05	Yes
8	22295813	rs12545568	G	A	0.1827	1.438	1.09E-05	No
8	22317574	rs11987482	G	A	0.181	1.328	4.29E-05	No
8	22319611	rs7819400	G	A	0.1908	1.374	1.56E-05	No
8	27576103	rs492786	T	C	0.3807	1.053	4.83E-05	No
8	27576664	rs542876	A	G	0.3807	1.053	4.83E-05	No
8	56645014	rs13251620	C	T	0.7297	1.2421	3.85E-05	Yes
8	56897479	rs7829716	A	G	0.0262	3.4449	2.62E-05	Yes
8	56928693	rs10087354	A	G	0.9742	–3.3874	4.05E-05	Yes
8	90280216	rs2046315	T	C	0.1322	1.871	6.50E-07	No
8	127035340	rs10956273	A	G	0.9348	2.1817	4.00E-05	Yes
8	127037469	rs10505464	C	A	0.06345	–2.212	3.27E-05	No
9	2170694	rs10965113	C	T	0.1385	1.5509	4.38E-05	Yes
9	17407517	rs2593395	A	G	0.06903	2.137	1.83E-05	No
9	27792678	rs10114706	C	T	0.9687	–3.0511	4.78E-05	Yes
9	27800414	rs4878586	C	T	0.9637	–3.0614	4.95E-05	Yes
9	70577584	rs10511964	A	C	0.4717	–1.0711	2.22E-05	Yes
9	70580991	rs12350270	A	G	0.4891	1.1532	1.03E-05	Yes
9	70658077	rs10869420	A	C	0.467	1.083	2.28E-05	No
9	70661732	rs6560397	T	C	0.4671	–1.289	2.40E-07	No
9	70703782	rs4745375	T	G	0.2657	–1.211	1.99E-05	No
9	70704367	rs883951	A	G	0.7277	1.2949	4.84E-06	Yes
10	1502856	rs6560737	T	C	0.2075	–1.339	2.95E-05	No
10	29330964	rs11007352	A	C	0.9264	2.0409	2.10E-05	Yes
12	55732112	rs324021	C	A	0.4313	1.081	3.02E-05	No
12	55736077	rs1044931	C	T	0.5674	–1.1019	2.34E-05	Yes
13	58708493	rs4884323	G	T	0.4012	1.14	1.41E-05	No
13	58708732	rs9570071	A	G	0.374	1.11	2.18E-05	No
14	82206638	rs1777696	C	T	0.489	–1.041	4.68E-05	No

(continued)

Appendix Table 1. (Continued)

CHR	BP	SNP	A1	A2	A1_freq	Beta	*p*	Imputed
15	91461811	rs6416579	G	A	0.3313	1.081	4.67E-05	No
15	91788851	rs11074186	A	G	0.1657	–1.568	1.13E-05	No
18	11283890	rs987890	A	C	0.1135	–1.617	4.94E-05	No
22	35528968	rs5750309	T	C	0.3187	1.206	1.34E-05	No
22	35529752	rs2022068	A	G	0.3926	1.1831	6.61E-06	Yes
22	35536287	rs4820254	G	T	0.4928	–1.3933	1.23E-05	Yes
22	35538606	rs5750310	A	C	0.567	–1.3572	3.96E-05	Yes
23	39770574	rs17145638	C	T	0.1169	–2.329	1.79E-07	No

included for SM D1MFS. During the covariate selection step, we did not consider the effect of any SNP.

The presence of *Streptococcus mutans* was either tested by means of a Dentocult®SM Strip mutans kit (Orion Diagnostica, Espoo, Finland) with saliva samples or determined genetically by a real-time PCR assay with DNA samples extracted from saliva (Vieira *et al.,* 2011). The measure of *Streptococcus mutans* was qualitative (*i.e*., the covariate was coded as 1 if present, and 0 if not). Education group was coded in discrete numbers ranging from 1 to 3, and was treated as ordinal values in the model; "1" indicated "up to high school", "2" indicated "some college", "3" indicated "four-year degree or beyond".

In addition to the strategy described above, we performed sensitivity analyses using several other covariate models, including models that used all covariates and models that used only age. None of these modeling variations produced any significant difference in the qualitative genetic results – the list of top genes was unchanged, and the *p* values differed only very slightly.

Genotyping, Statistical Analysis, and Result Annotation

The genotype data are publicly available from dbGap (http://www.ncbi.nlm.nih.gov/gap, study accession designation phs 000095.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* values under .001; 32,417 SNPs were filtered out due to missing rates higher than 10%; 69,818 SNPs were filtered out due to minor allele frequency less than 2%, yielding a total of 548,012 SNPs passing all the filters for the PF-surface analysis, and 548,051 for the SM-surface analysis.

The GENEVA Coordinating Center performed imputation with BEAGLE (Browning and Browning, 2009) using the HapMap Phase III reference panel for autosomal SNPs. Certain sample-chromosome combinations were excluded where a gross chromosomal anomaly was detected or when the chromosome-specific missing call rate was ≥ 5%. All chromosomes for a given sample were excluded when the missing call rate across all chromosomes was ≥ 5%. Only study participants of European ancestry as defined *via* principal components analysis were selected for imputation analysis. European ancestry individuals were defined as those falling within 5 standard deviations of the mean of eigenvectors 1 and 2 for self-identified white participants. Sporadic missingness of SNPs was also imputed. To determine the imputation quality of BEAGLE, we performed masked SNP analysis (*i.e*., comparison of imputed and experimentally determined genotypes) and analysis of Mendelian inconsistencies for imputed SNPs among relatives. 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).

Association between PF- or SM-surface caries scores and each SNP was tested with PLINK (Purcell *et al.,* 2007). To see if different statistical tests dramatically affected X chromosome SNPs rs17145638 and rs3788848 *p* values, we compared results from the PLINK test, which puts men and women on different scales (men coded as 0/1 for genotypes A/B, women coded as 0/1/2 for genotypes AA/AB/BB), linear regressions putting men and women on the same scale (men coded as 0/2 for A/B, women coded as 0/1/2 for AA/AB/BB), and meta-analysis on stratified analysis by sex (*i.e.,* applied the PLINK test to men and women separately). In any of these tests, *p* values of the 2 SNPs stayed within the suggestive range (<5 E-5). We chose to report the *p* values from PLINK test. Association tests were conducted for genotyped data and imputed data separately, and the results were merged, with the genotyped results retained whenever there was an overlap.

Manhattan plots were generated by Haploview (Barrett *et al.,* 2005). Quantile-quantile plots and genomic inflation factors (λ) were generated by the R statistical package (R Foundation for Statistical Computing, Vienna, Austria). We chose to maintain statistical power and included all samples while carefully monitoring type I error, which may be subject to inflation if there is a substantial number of related samples, by setting a relatively stringent suggestive significance threshold (α = 5E-5). Also, no severe skewness of quantile-quantile plots was observed either

Appendix Table 2. 117 SNPs for SM Surface Scans Meeting Suggestive Significance (*i.e., p* value < 5E-5)

CHR	BP	SNP	A1	A2	A1_freq	Beta	*p*	Imputed
1	9432408	rs10779726	C	G	0.8237	–3.6881	2.02E-05	Yes
1	9433113	rs11121423	C	T	0.8236	–3.6925	1.98E-05	Yes
1	84998887	rs1750491	A	G	0.2815	3.186	1.85E-05	No
1	104688236	rs11811323	G	A	0.03387	7.438	3.74E-05	No
1	208435759	rs12094311	G	A	0.1508	3.912	3.41E-05	No
1	208437213	rs12060567	A	G	0.1508	3.912	3.41E-05	No
1	208956814	rs11119577	A	T	0.1501	4.0053	1.29E-05	Yes
1	208960258	rs6681860	A	C	0.8537	–3.9032	3.16E-05	Yes
1	208961270	rs12022982	A	G	0.1326	4.193	1.23E-05	No
1	208963316	rs7555360	T	G	0.1337	4.022	2.57E-05	No
1	208964372	rs10863852	A	C	0.1344	4.2433	7.79E-06	Yes
1	208964886	rs1934620	A	G	0.8656	–4.2462	7.70E-06	Yes
1	208966802	rs10863853	A	G	0.8646	–4.2301	8.38E-06	Yes
1	244336191	rs6681900	T	C	0.08158	5.058	2.57E-05	No
1	244336569	rs6689428	G	A	0.08015	5.186	1.81E-05	No
1	244348150	rs6686745	C	T	0.9144	–4.8267	4.72E-05	Yes
1	244348886	rs12404212	C	T	0.9143	–4.8269	4.72E-05	Yes
2	15302935	rs9287655	T	C	0.406	2.852	2.10E-05	No
2	15344624	rs4668892	C	T	0.5568	–2.8106	3.29E-05	Yes
2	15345075	rs4668893	C	T	0.4397	2.7452	4.09E-05	Yes
2	48956648	rs995146	A	C	0.9306	–7.3272	1.69E-05	Yes
2	85192530	rs3893079	C	A	0.05492	6.312	1.41E-05	No
2	108245143	rs1470874	A	G	0.3625	2.793	4.25E-05	No
2	185640560	rs263767	G	A	0.2419	–3.181	2.82E-05	No
2	218073793	rs4372880	T	C	0.1322	–4.156	2.44E-05	No
2	218079446	rs9989823	T	C	0.136	–4.088	2.86E-05	No
2	218093085	rs2373077	A	C	0.1279	–4.129	3.43E-05	No
2	218827379	rs7600989	G	T	0.9612	–8.0739	2.26E-06	Yes
2	218834980	rs1567869	C	T	0.0416	8.0814	1.99E-06	Yes
2	218836750	rs2292549	C	T	0.0397	8.0833	1.93E-06	Yes
2	218838758	rs1079204	A	G	0.0378	8.4526	1.90E-06	Yes
2	218878706	rs1017697	A	G	0.0404	7.8129	3.34E-06	Yes
2	218879706	rs2014597	A	G	0.0404	7.8309	3.21E-06	Yes
2	218896913	rs6708662	C	T	0.0424	7.6107	5.24E-06	Yes
2	218898957	rs10192690	A	G	0.959	–7.9885	4.27E-06	Yes
2	229584711	rs16825564	A	G	0.9024	–4.8062	9.08E-06	Yes
2	229591741	rs7563172	A	C	0.0967	4.7758	1.05E-05	Yes
2	229593322	rs10490035	C	A	0.09685	4.762	1.34E-05	No
3	70017021	rs17006578	A	G	0.9899	–14.3399	4.62E-05	Yes
3	126170457	rs1909586	G	T	0.3824	2.9042	3.92E-05	Yes
3	168209459	rs6781033	A	G	0.785	3.3438	3.53E-05	Yes
3	168252662	rs1519943	C	T	0.7467	3.1428	3.61E-05	Yes
3	176307233	rs4894477	A	G	0.4077	2.809	1.79E-05	No
3	176310488	rs6782155	A	G	0.4084	2.858	1.29E-05	No
4	15626996	rs2286458	G	T	0.0677	5.9704	3.83E-06	Yes
4	15627380	rs2677789	A	C	0.06651	5.808	8.54E-06	No
4	15627422	rs2531154	G	A	0.06638	5.808	8.45E-06	No
4	15630636	rs1829271	A	G	0.9357	–6.3378	1.78E-06	Yes
4	15631638	rs2677780	A	G	0.9356	–6.3471	1.74E-06	Yes
4	15645094	rs6816182	C	T	0.1085	4.4903	4.78E-05	Yes
4	40668284	rs4466078	A	T	0.9856	–17.8273	1.54E-05	Yes
4	66795224	rs4289486	A	G	0.05344	6.291	2.06E-05	No
4	78101455	rs17002297	C	T	0.7844	–3.3733	3.20E-05	Yes
4	78121149	rs4241597	A	G	0.7832	–3.3672	3.17E-05	Yes
4	141929596	rs11930453	A	C	0.3961	2.7239	4.77E-05	Yes
4	147095001	rs723794	G	T	0.2686	3.645	5.94E-07	No
5	131900972	rs739718	C	T	0.0754	5.1874	3.47E-05	Yes
5	156123792	rs1845479	G	A	0.03721	7.77	1.36E-05	No
5	156124842	rs11953631	T	C	0.03677	7.8	1.42E-05	No

(continued)

Appendix Table 2. (Continued)

CHR	BP	SNP	A1	A2	A1_freq	Beta	*p*	Imputed
5	158485488	rs1582508	A	G	0.4885	2.796	2.33E-05	No
5	158500688	rs2420355	A	C	0.4372	–2.7032	4.19E-05	Yes
5	170173001	rs11134654	A	C	0.1735	4.6139	2.12E-07	Yes
5	170176011	rs1422160	A	G	0.1718	4.322	1.32E-06	No
5	170195941	rs11745293	C	T	0.8527	–4.0611	1.18E-05	Yes
5	170202257	rs888811	C	T	0.1452	4.0888	1.30E-05	Yes
6	31298517	rs9263985	A	G	0.3321	–2.86	4.67E-05	No
6	88894335	rs9353524	C	G	0.7982	–3.6251	1.44E-05	Yes
6	167085776	rs10946186	T	C	0.4948	–2.814	1.69E-05	No
6	167086923	rs9295368	A	G	0.5766	–2.7747	3.20E-05	Yes
6	167096924	rs388372	T	C	0.3569	2.995	1.72E-05	No
6	167097412	rs635808	T	C	0.2266	–3.394	1.44E-05	No
8	26137422	rs4275231	T	C	0.1656	3.791	1.78E-05	No
8	27375822	rs11778371	T	C	0.05153	–6.018	4.57E-05	No
8	67695752	rs2467750	A	G	0.04298	6.955	2.40E-05	No
8	90062604	rs10429371	C	T	0.2142	3.8	3.00E-06	No
8	90084328	rs1487791	A	G	0.735	–3.33	1.68E-05	Yes
8	90087157	rs12676566	C	G	0.735	–3.3252	1.73E-05	Yes
8	90280216	rs2046315	T	C	0.1322	5.187	7.85E-08	No
8	134554009	rs7835464	A	G	0.04723	6.371	4.01E-05	No
8	140045364	rs7834262	G	T	0.2366	3.387	2.03E-05	No
9	22761006	rs10965448	T	G	0.04437	6.533	4.71E-05	No
9	22783225	rs10811775	G	A	0.05072	6.405	2.65E-05	No
9	101274144	rs649057	T	G	0.05057	6.806	7.72E-06	No
10	5391308	rs7081234	C	T	0.5211	2.7347	2.42E-05	Yes
10	5394347	rs10904478	C	T	0.4051	–2.8759	1.21E-05	Yes
10	5414727	rs11253143	A	G	0.6054	3.036	1.39E-05	Yes
10	10425015	rs7087371	C	T	0.3587	–2.9197	4.78E-05	Yes
10	13893858	rs7088455	A	G	0.06155	5.933	1.58E-05	No
10	17294838	rs7080366	T	C	0.437	–2.71	3.76E-05	No
10	21607442	rs12358291	T	G	0.06584	5.565	2.49E-05	No
10	129997306	rs1255136	A	G	0.5118	2.7693	2.48E-05	Yes
10	131569383	rs4142058	A	G	0.3559	–2.83	4.64E-05	No
11	80889400	rs11232701	C	G	0.0955	5.3955	1.03E-05	Yes
11	80904206	rs10897833	G	A	0.09351	4.945	2.65E-05	No
11	80909584	rs11232711	G	A	0.09303	4.851	3.93E-05	No
11	80910473	rs4944304	C	A	0.09447	5.059	1.58E-05	No
11	80912798	rs2032381	G	T	0.9212	–5.6947	3.68E-06	Yes
11	132607863	rs7933745	A	G	0.1613	–3.821	1.25E-05	No
11	132626147	rs2078454	A	C	0.2094	–3.603	6.05E-06	No
13	19971094	rs7335998	C	G	0.0112	12.5701	2.95E-05	Yes
13	19972980	rs9506503	A	C	0.0133	12.4141	4.27E-05	Yes
13	19973813	rs6490590	C	T	0.9889	–12.5924	2.85E-05	Yes
13	20178034	rs735539	T	C	0.2476	3.241	2.31E-05	No
13	20183009	rs6490605	A	G	0.2471	3.2794	1.71E-05	Yes
13	58925818	rs17056606	A	G	0.9893	–13.836	1.56E-05	Yes
18	45458441	rs8082881	G	A	0.05391	6.29	2.02E-05	No
18	61341576	rs12962841	A	G	0.4269	–2.8258	3.58E-05	Yes
18	67590053	rs17085106	G	T	0.9843	–11.6027	1.34E-05	Yes
18	67592624	rs8099373	A	G	0.9846	–12.071	9.54E-06	Yes
18	72447295	rs13381274	C	T	0.0665	5.8385	2.43E-05	Yes
18	72447598	rs13381277	G	A	0.06107	5.889	1.57E-05	No
20	41425015	rs2010809	C	T	0.9883	–14.0668	1.83E-06	Yes
20	41435864	rs6103260	G	T	0.9883	–14.0537	1.87E-06	Yes
20	41444367	rs6017025	A	G	0.9867	–11.6564	3.00E-05	Yes
20	41447413	rs6103268	C	G	0.0131	12.5835	1.13E-05	Yes
23	129022933	rs6637684	T	C	0.2493	3.652	2.98E-05	No
23	129027493	rs3788848	G	A	0.2605	3.807	1.01E-05	No

Appendix Table 3. Caries Scores of the Most Significant SNPs for X Chromosome Genes *BCOR* and *BCORL1* by Sex and Genotype

SNP (phenotype)	Major Allele	Minor Allele	Sex	Genotype	Genotype Count	Phenotype Mean	95% CI from *t* distribution
rs17145638 (PF D1MFS)	T	C	Females	CC	9	6	(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	254	13.7	(11.6, 15.8)
				AA	325	10.6	(9.1, 12.1)
			Males	G	98	15.1	(11.3, 18.9)
				A	293	11.1	(9.5, 12.6)

Note: not adjusted for covariates.

from genomic inflation factors λ or by visual inspection. In addition, we focused more on the rank of association signals than on their nominal *p* values, and the ranks were not affected by relatedness.

LocusZoom (http://csg.sph.umich.edu/locuszoom/) (Pruim *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 strict hypothesis-testing approach), we set our suggestive significance threshold at α = 5E-5, to diminish Type 1 error and generate a reasonable number of SNPs to be annotated. All genes within the 400-kb 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 nominate as possible susceptibility genes.

APPENDIX REFERENCES

Barrett JC, Fry B, Maller J, Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-265.

Browning BL, Browning SR (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 84:210-223.

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.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, *et al.* (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26:2336-2337.

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.

Vieira AR, Deeley KB, Callahan NF, Noel JB, Anjomshoaa I, Carricato WM, *et al.* (2011). Detection of *Streptococcus mutans* genomic DNA in human DNA samples extracted from saliva and blood. *ISRN Dent* 2011:543561.

Wang X, Shaffer JR, Weyant RJ, Cuenco KT, DeSensi RS, Crout R, *et al.* (2010). Genes and their effects on dental caries may differ between primary and permanent dentitions. *Caries Res* 44:277-284.

Wendell S, Wang X, Brown M, Cooper ME, DeSensi RS, Weyant RJ, *et al.* (2010). Taste genes associated with dental caries. *J Dent Res* 89:1198-1202.