Int J Mol Epidemiol Genet 2023;14(2):19-33
www.ijmeg.org /ISSN:1948-1756/IJMEG0150137

Original Article

Potential risk factors and genetic variants associated with dental caries incidence in Appalachia using genome-wide survival analysis

Tianyu Zou[1], Katherine Neiswanger[2], Eleanor Feingold[1,3], Betsy Foxman[4], Daniel W McNeil[5], Mary L Marazita[1,2,6], John R Shaffer[1,2]

[1]Department of Human Genetics, School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; [2]Center for Craniofacial and Dental Genetics, Department of Oral and Craniofacial Sciences, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; [3]Department of Biostatistics, School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; [4]Center for Molecular and Clinical Epidemiology of Infectious Diseases, Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA; [5]Department of Community Dentistry and Behavioral Science, College of Dentistry, University of Florida, Gainesville, FL, USA; [6]Clinical and Translational Sciences, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

Received March 14, 2023; Accepted July 6, 2023; Epub August 15, 2023; Published August 30, 2023

Abstract: Objective: The aim of this study was to identify the potential risk factors and genetic variants associated with dental caries incidence using survival analysis. Methods: The Center for Oral Health Research in Appalachia recruited and prospectively followed pregnant women and their children. A total of 909 children followed from birth for up to 7 years were included in this study. Annual intra-oral examinations were performed to assess dental caries experience including the approximate time to first caries incidence in the primary dentition. Cox proportional hazards models were used to assess the associations of time to first caries incidence with self-reported risk factors and 4.9 million genetic variants ascertained using a genome-wide genotyping array. Results: A total of 196 of 909 children (21.56%) had their first primary tooth caries event during follow-up. Household income, home water source, and mother's educational attainment were significantly associated with time to first caries incidence in the stepwise Cox model. The heritability (i.e., proportion of variance explained by genetics) of time to first caries was 0.54. Though no specific genetic variants were associated at the genome-wide significance level (P < 5E-8), we identified 14 loci at the suggestive significance level (5E-8 < P < 1E-5), some of which were located within or near genes with plausible biological functions in dental caries. Conclusion: Our findings indicate that household income, home water source, and mother's educational attainment are independent risk factors for dental caries incidence. We nominate several suggestive loci for further investigation.

Keywords: Risk factors, genetics, bioinformatics, genomics, epidemiology, child dentistry

Introduction

Dental caries is the most common chronic disease of childhood affecting 21.4% of US children by age 5 years and 50.5% by age 11 years [1]. Dental caries can lead to chronic pain, loss of teeth, poor school performance and decreased success later in life [2, 3]. A better understanding the etiology of dental caries incidence in the primary dentition is needed to improve caries prevention at both individual and population levels.

Dental caries is a complex disease influenced by multiple factors. Environmental risk factors, such as low socioeconomic indicators and inadequate fluoride exposure, have been identified to be associated with caries prevalence [4, 5]. Behavioral risk factors related to dental caries include poor oral hygiene [6] and cariogenic diet [7], especially sugar intake, which is a key component influencing caries development [8, 9]. Moreover, biological factors including the oral microbiome [6, 10] and genetic factors are associated with caries experience.

Risk factors specifically for caries incidence - that is the proportion of new cases of dental caries across a defined period of time - in the primary dentition include parental income, education, and occupation [11, 12], previous caries experience, consumption of sweetened foods/drinks [13-15], *Streptococcus mutans* level [12, 16], fluoridation in drinking water [11], and tooth brushing frequency [17]. However, most longitudinal risk factor studies have coded caries incidence in the primary dentition as binary traits (e.g., presence or absence of caries) and/or incremental increase of decayed, missing, and filled teeth or surfaces (dmft/dmfs) between two time points, which does not fully capture timing information. By contrast, survival analysis retains "time at risk" information and has been used to identify risk factors for time-to-event of caries incidence including the time to caries incidence in both primary and permanent molars [18], the time to caries increment [19] and the time to first caries in permanent teeth [20]. However, the small sample sizes and/or short follow-up periods of these studies limit their power to detect and model the effects of risk factors.

Genetic studies have reported a strong genetic component to dental caries in the primary dentition, with genetic factors explaining 30%-76% of variation in the dental caries traits [21-25]; the wide range is estimates may be due to differences in study design, populations, sample sizes, or dental caries phenotyping methods. Several "hypothesis-free" genome-wide association studies (GWASs) of dental caries in the primary dentition have been performed over the past decade to investigate clinical phenotypes including caries affection status (affected vs. unaffected) and quantitative caries indices (e.g., counts of the number of affected teeth or tooth surfaces), in samples ranging from approximately 100 to 19,000 individuals [26-30]. These studies have identified several association signals located near genes with plausible biological roles in dental caries, however, to date, few of these associations reached the genome-wide significance threshold or have been rigorously replicated. Moreover, the SNP-based heritability (i.e., the proportion of the phenotypic variance explained by the set of common genetic variants) of caries in the primary dentition was much lower than previously reported family-based heritability estimates. For instance, a consortium GWAS with the largest combined sample size (n = 19,003) identified only one genome-wide significant association, in *ALLC*, using a binary phenotype. The consortium-wide estimated heritability was only 1% [28], in stark contrast to the aforementioned family-based estimates, and lower than the individual heritability estimates observed in the constituent studies that formed the consortium. In addition, prior GWASs used phenotypes collected at a single time point, thus were not able to evaluate the genetic basis of time-to-caries incidence. As mentioned above, survival analysis takes time information into account and its application in GWAS is an approach that may broaden our understanding of the genetic underpinnings of dental caries onset. Therefore, we used the time to first caries incidence in the primary dentition as the outcome in order to identify its potential risk factors and associated genetic variants using survival analysis.

Methods

Study sample

The study design and recruitment descriptions of the Center for Oral Health Research in Appalachia, cohort 2 (COHRA2) have been reported before [31]. Briefly, COHRA2 is a longitudinal study and has been recruiting and prospectively following pregnant white women from Pennsylvania and West Virginia and their children starting in 2011. This cohort has information on genetic, microbial, behavioral, and environmental factors involved in oral health. A subset of 909 children with available data on potential risk factors and genetics was included in this analysis. Each participant was followed annually from birth with 7 years being the longest follow-up, to date, in this ongoing study. Annual intra-oral examinations were performed to assess dental caries experience. Parental consent was provided for all participants, and Institutional Review Boards and Ethics Committees approved all aspects of the study.

Phenotype and genotyping

The participants were phenotyped based on intra-oral examinations according to caries criteria that distinguished cavitated from non-cavitated lesions (i.e., white spots/incipient

lesions). The time to first caries incidence was defined as the time (in years) from birth to the first visit when one or more carious lesions were observed. If no carious lesions were present by the end of follow-up period, time to first caries incidence was coded as censored at the last available visit. For example, if the child had carious lesions at the third visit (3 years-old), then his/her survival time is 3 years and the outcome was 1 (i.e., with carious lesions); if no carious lesions were observed during the 7 years, then survival time is 7 years and the outcome was 0 (i.e., without carious lesions).

Participants were genotyped using the Infinuim Multi-Ethnic Global-8 v1.0 Array. Ungenotyped single nucleotide polymorphisms (SNPs) and sporadic missing data of genotyped SNPs were imputed using the Michigan Imputation Server (Minimac 3 for autosomal chromosomes and Minimac 4 for chromosome X), and Haplotype Reference Consortium Phase 1 as the reference panel. Variants were excluded if they had low imputation quality (INFO score < 0.3), departed from Hardy-Weinberg equilibrium (P < 1E-6), or had minor allele frequencies < 5%. A total of 4.9 million genetic variants were included in this study after filtering.

Risk factor modeling

Demographic, behavioral and environmental factors that putatively relate to oral health were collected including sex (male or female), recruitment site (Pennsylvania [PA] or West Virginia [WV]), household income (< $50,000 per year or ≥ $50,000 per year), home water source (water company or well), home fluoride level (ppm, measured in a water sample), mother's educational attainment (< high school, high school or some college, or college degree or more), mother's tooth brushing frequency (> 1/day, 1/day, or < 1/day), breastfeeding status at baseline (yes or no), and breastfeeding duration (months). The correlations among factors were calculated using Pearson's correlation method. There were no strong correlations (**Figure 1**). Univariate Cox proportional hazards regression models with one risk factor included at a time were used to evaluate the relationships between risk factors and the time to first caries incidence. To identify the independent risk factors, bidirectional stepwise (with iterations between the "forward" and "backward" steps) Cox modeling was performed via the R package "My.stepwise". First, all potential risk factors were put on the "variable list" to be selected, the significance levels for entry into and for retention in the model were set at 0.15. Then, the final Cox model was obtained by dropping the variables with *P* value > 0.05 one at a time until all included variables were significant at alpha level of 0.05. Potential multicollinearity among the variables was evaluated using variance inflation factor (VIF). VIF > 10 in continuous variables or VIF > 2.5 in categorical variables was considered indicative of a multicollinearity issue among variables. In the final model, VIF values of selected variables (household income, home water source, and mother's educational attainment) were 1.45, 1.01, 1.43, respectively, indicating no multicollinearity problem. Because previous studies have shown that stepwise regression can result in poor model fit out-of-sample [32], we also performed regularized Cox modeling as a second variable selection procedure (see Supplementary Materials for methods details of regularized Cox modeling) as confirmation of the variables included in the final model. The results of the regularized Cox model were the same as the stepwise Cox model (Supplementary Figure 1).

Survival GWAS and candidate SNP replication

Cox proportional hazards regression models were used to assess the associations between time to first caries incidence and genetic variants in the R package "gwasurvivr" [33], while adjusting for household income, home water source, and mother's educational attainment (i.e., the three risk factors that were significantly associated with time to first caries incidence in the final Cox model), and the first 6 principal components (PCs) of ancestry derived from a principal component analysis of the genome-wide genetic data. The number of PCs sufficient to capture the population structure was determined based on the scatterplots and scree plots of PCs. The threshold for genome-wide significance was *P*-value < 5E-8, and for suggestive significance was *P*-value < 1E-5. The results were summarized and visualized using Manhattan plots and quantile-quantile (Q-Q) plots created in the R statistical environment (R Foundation for Statistical Computing, Vienna, AU). The regional association plots showing

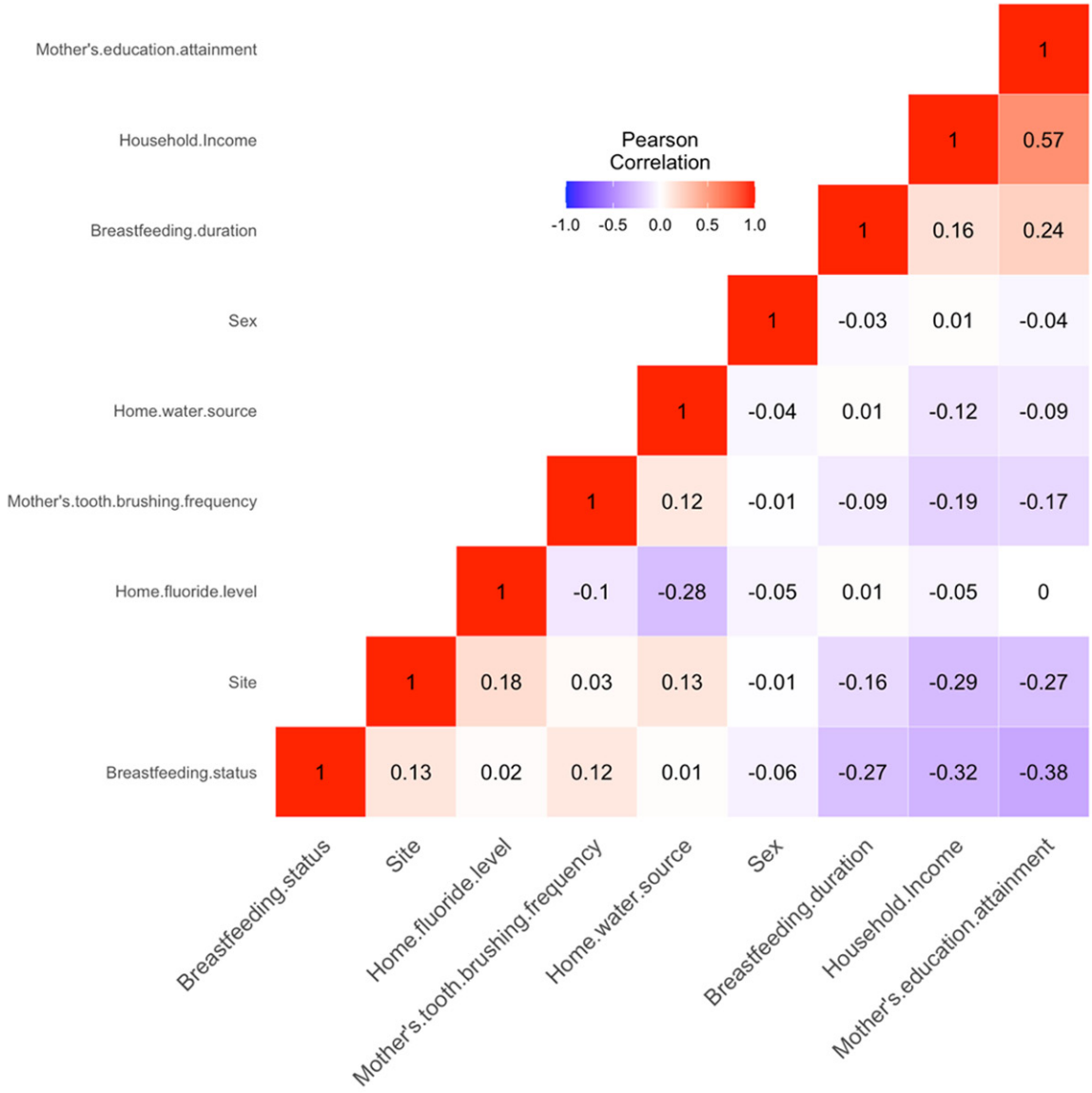


Figure 1. Pearson's correlations among environmental risk factors. The number in each block indicates the correlation coefficient between two factors. The color shows the degree of the correlation (red represents positive correlation and blue represents negative correlation).

mapped genes were visualized using LocusZoom [34].

In addition to the GWAS approach, we conducted a candidate SNP replication study to test the associations between time to first caries incidence and 23 genetic variants that were reported to be associated with dental caries traits in previous GWASs [28, 35-38] at the genome-wide significant level using the same Cox models and covariates as described above (methods details described in the Supplementary Materials). The significance level for candidate SNP replication was set at *P*-value < 0.0022 (i.e., 0.05/23, the Bonferroni correction for 23 tested variants).

Gene-based and gene set enrichment analyses

We performed a gene-based analysis using MAGMA [39] that uses multiple linear principal components regression model to compute the gene-level *p*-values and a gene set enrichment analysis using Genomic Regions Enrichment of Annotations [40] with 14 independent SNPs in

Table 1. Basic characteristics of study sample

Variables	n (%) or mean ± SD
Sex	
Male	504 (55.45)
Female	405 (44.55)
Recruitment site	
PA	483 (53.14)
WV	426 (46.86)
Annual household income	
< $50,000	375 (43.60)
≥ $50,000	485 (56.40)
Home water source	
Water company	766 (93.87)
Well	50 (6.13)
Home fluoride level (ppm)	0.79 ± 0.29
Mother's educational attainment	
< High school	171 (18.98)
High school-Some college	260 (28.86)
College degree or more	470 (52.16)
Mother's tooth brushing frequency	
> 2/day	661 (73.36)
1/day	216 (23.97)
< 1/day	24 (2.66)
Breastfeeding status at baseline	
Yes	733 (83.58)
No	144 (16.42)
Breastfeeding duration (months)	3.66 ± 5.48

Note: PA: Pennsylvania; WV: West Virginia; SD: standard deviation.

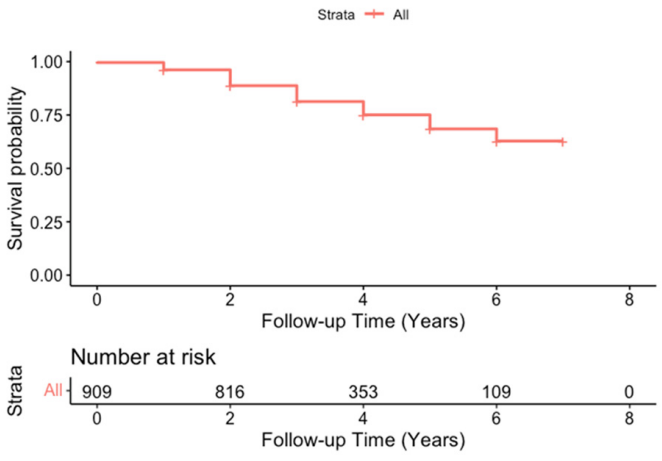


Figure 2. Kaplan-Meier survival curve, using first decayed or filled primary tooth surface as the event. Number at risk on the bottom means the number of individuals that are caries-free (in the risk of having the first caries) at the follow-up time. Our follow-up period was 7 years, therefore there were no children at risk at 8 years.

suggestive signals (P < 1E-5) as input. The gene-based analysis aggregates genetic variants at the gene level and tests the joint association of aggregated variants with the phenotype. In the gene set enrichment analysis, genes are aggregated to gene groups based on biological functions, which can provide evidence of specific pathways involved in the phenotype. The advantages of aggregation include reduced number of tests and more power to detect weak associations.

Heritability estimation

We calculated the SNP-based heritability estimate (i.e., the proportion of the phenotypic variance explained by a given set of genetic variants) of time to first caries incidence trait using SumHer [41], which used the summary statistics from the survival GWAS as input and reference linkage disequilibrium scores from the HapMap3 reference panel using the linkage disequilibrium-adjusted kinship method.

Results

Sample summary

The basic characteristics of the study sample are summarized in **Table 1**. A total of 196 of 909 children (21.56%) had their first primary tooth caries event during the follow-up period. The caries-free survival rates were 99.6% at 1-year follow-up, 96.1% at 2-year follow-up, 88.7% at 3-year follow-up, 81.3% at 4-year follow-up, 75.1% at 5-year follow-up, 68.5% at 6-year follow-up, and 62.8% at 7-year follow-up (**Figure 2**).

Risk factor modeling

In the univariate Cox models, household income, home water source, mother's educational attainment, mother's tooth brushing frequency, breastfeeding status at baseline, and breastfeeding duration were individually significantly associated with time to first caries incidence ($P < 0.05$, **Table 2**). However, in the final stepwise Cox model, only household income, home water source, and mother's educational attainment were retained. Higher household income

Table 2. Univariate Cox models of risk factors and time to first caries incidence

Factors	Reference group	Alternative group	Estimate	SE	P	HR (95% CI)
Sex	Male	Female	-0.041	0.144	0.773	0.959 (0.724-1.271)
Recruitment site	PA	WV	0.207	0.143	0.147	1.230 (0.929-1.629)
Household income	< 50,000/year	≥ 50,000/year	-0.813	0.149	< 0.0001	0.443 (0.331-0.594)
Home water source	Water company	Well	0.989	0.244	< 0.0001	2.689 (1.666-4.339)
Home fluoride level		per ppm	0.069	0.249	0.782	1.072 (0.657-1.747)
Mother's educational attainment	< High school	High school or some college	-0.596	0.182	0.001	0.551 (0.385-0.788)
Mother's educational attainment	< High school	College degree or more	-1.259	0.177	< 0.0001	0.284 (0.200-0.402)
Mother's tooth brushing frequency	< 1/day	1/day	0.066	0.169	0.698	1.068 (0.767-1.488)
Mother's tooth brushing frequency	< 1/day	> 1/day	1.216	0.316	< 0.0001	3.375 (1.818-6.265)
Breastfeeding status	Yes	No	0.723	0.173	< 0.0001	2.061 (1.468-2.893)
Breastfeeding duration		per month	-0.036	0.015	0.017	0.964 (0.936-0.994)

Note: PA: Pennsylvania; WV: West Virginia; SE: standard error; HR: hazard ratio; CI: confidence interval.

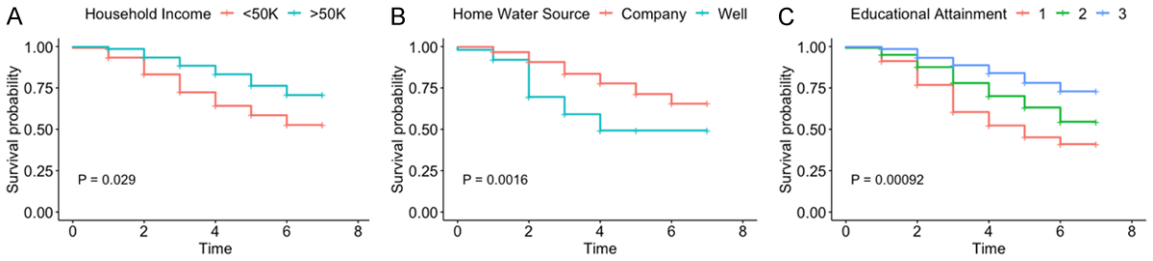


Figure 3. Survival curves for time to first caries incidence stratified by significant risk factors in the final stepwise Cox model. The x-axis is the follow-up time in years. The y-axis is the probability of remaining caries free. The lines represent survival curves of stratified groups. A. Stratified by household income. B. Stratified by home water source. C. Stratified by mother's educational attainment (1 = < high school, 2 = high school or some college, 3 = college degree or more).

group and higher levels of mother's educational attainment were significantly associated with lower hazards for time to first caries incidence (Hazard ratio [HR] = 0.65 and 0.66, P = 0.029 and 0.00092, respectively), while well water as the home water source was significantly associated with greater hazards for time to first caries incidence (HR = 2.28, P = 0.0016, **Figure 3**) compared to municipal water. These results indicate that household income, home water source, and mother's educational attainment may have independent (though not necessarily causal) effects on time to first caries incidence. These three final risk factors were included in the following survival GWAS.

Heritability estimation, survival GWAS, and candidate SNP replication

The SNP-based heritability estimate of the time to first caries incidence was 0.54 (standard deviation [SD]: 0.41) indicating that genetics may have a large role in the timing or pace of the cariogenic processes. **Figure 4** shows the Manhattan plot of the survival GWAS results for time to first caries incidence. There was no evidence of genomic inflation (genomic inflation factor was 1.00; Supplementary Figure 2). Though there were no specific genetic variants associated with time to first caries incidence at the genome-wide significance level (P < 5E-8), 14 loci at the suggestive level were identified (P < 1E-5). **Table 3** shows the 14 suggestive loci and their nearby genes and annotation. Notably, some lead SNPs of the suggestive association signals were within or near genes with biologically plausible roles in dental caries including rs34201252 (P = 1.77E-7, located near the gene *COL5A1*), rs7503428 (P = 2.10E-6, located in the gene *ASIC2*), and rs35508695 (P = 7.30E-6, located in the gene *ESR1*). In addition, the lead SNP rs28567072 on chromosome 8 (P = 5.36E-6) is located near the gene *ABRA*, which was previously reported to be associated with dental caries in the permanent dentition at the suggestive significance

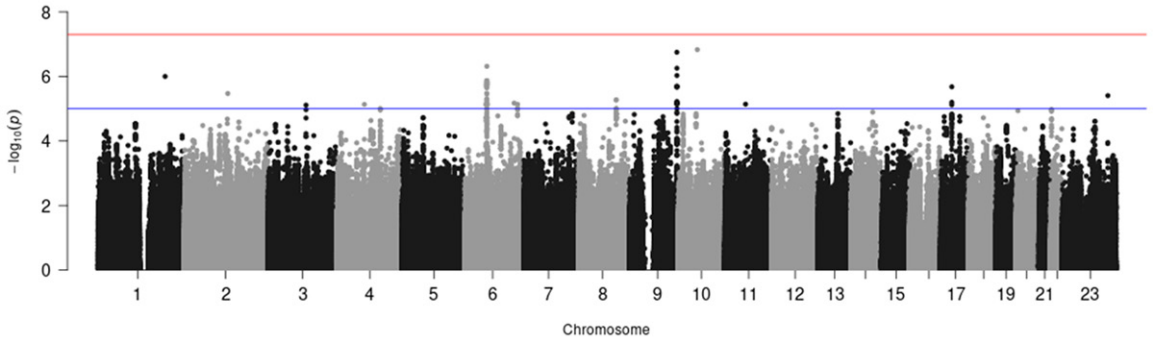


Figure 4. Manhattan plot for the survival GWAS of time to first caries incidence. The red line indicates the genome-wide significance threshold (P = 5E-8), and the blue line indicates the suggestive significance threshold (P = 1E-5). The y-axis is the -$\log_{10}$-transformed *P* value, the x-axis is the physical position of each variant organized by chromosome.

level [42]. The regional association plots for these 4 loci near or in the genes with potential roles in dental caries are provided in **Figure 5**.

We also tested genetic variants that have been previously reported to be associated with dental caries traits in the primary or permanent dentitions at the genome-wide significance level (P < 5E-8). There were 23 tested variants, however, none of them showed association with time to first caries incidence in this analysis (Supplementary Table 1).

Gene-based genetic association, and gene set enrichment analyses

In the gene-based genetic association analysis, the SNPs from the survival GWAS were mapped to 18,942 protein coding genes. The genome-wide significance threshold for gene-based analysis was *P*-value < 2.74E-6 (0.05/18942). There were no genes reaching genome-wide significance threshold. The gene with the smallest *P*-value was *TBX1* (P = 4.49E-5). There was no significant enrichment in the gene set enrichment analyses.

Discussion

Dental caries is a complex disease that can be impacted by environmental, behavioral and genetic factors. All the factors are important in the etiology of dental caries. We have a cohort with relatively larger sample size and longer follow-up period, which provides statistical power to detect risk factors and genetic factors for dental caries incidence. It is beneficial to determine behavioral/environmental factors first because these significant risk factors could explain some non-genetic variation. Accounting for such non-genetic sources of variance would in turn improve our ability to detect genetic effects.

In this study, we identified that household income, home water source, and mother's educational attainment were significantly associated with caries incidence in the primary dentition. Household income and education are commonly used as indicators of the socioeconomic status (SES) of research participants. It is well-established that populations with lower SES have higher caries prevalence [4], both within specific communities and across communities, worldwide. Our results provided further support that SES level was associated with time to first caries incidence. Higher family SES may reduce children's caries incidence through promoting healthy living conditions, such as better access to dental care, promotion of positive oral health behaviors including regular tooth brushing, and reduced exposure to environmental stressors and risk factors including lower or less frequent consumption of sugary foods and beverages.

Regarding the observed association with home water source, we speculate that this association could be explained by well water differing in fluoridation level from municipal water [11] and/or that participants living in homes serviced by well water reside in more rural areas and may experience decreased access to dental care [43]. Further studies are needed to investigate the role of home water source in caries incidence and disentangle its potentially

Table 3. Survival GWAS results for loci showing suggestive evidence of association with time to first caries incidence

SNP	CHR	BP	EA/RA	MAF	P	HR (95% CI)	Gene nearby and biological relevance
rs35324031	10	55529375	G/A	0.060	1.48E-7	0.41 (0.30-0.58)	*PCDH15:* is associated with Usher type 1 syndrome involved in congenital hearing loss [69]
rs34201252	9	137426383	G/A	0.054	1.77E-7	0.40 (0.28-0.56)	*COL5A1*: is associated with Ehlers-Danlos syndrome presenting with dental pathology [70] and canine agenesis [71] *RXRA:* encodes retinoid X receptor α involved in retinoic acid-mediated gene activation
rs4710384	6	63554552	G/A	0.155	4.84E-7	0.56 (0.44-0.70)	NA
rs112019823	1	194344257	C/T	0.114	1.00E-6	0.53 (0.41-0.68)	NA
rs7503428	17	32372318	G/C	0.362	2.10E-6	0.61 (0.50-0.75)	*ASIC2:* encodes a member of acid-sensing ion channels involved in the perception of sour taste [56] and is associated with gingival inflammation [58]
rs35643512	2	126373265	G/T	0.162	3.38E-6	0.56 (0.44-0.72)	NA
rs12834574	23	131692337	C/T	0.073	3.94E-6	0.57 (0.44-0.72)	NA
rs28567072	8	107883740	A/G	0.094	5.36E-6	0.52 (0.39-0.69)	*ABRA:* encodes a muscle specific actin-binding protein involved in skeletal muscle hypertrophy [72] and arteriogenesis [73] *OXR1:* is an important regulator of neuronal survival in response to oxidative stress [74]
rs9399396	6	142564217	C/T	0.201	6.69E-6	0.59 (0.47-0.75)	*VTA1*: encodes a protein involved in trafficking of the multivesicular body and is associated with arthrogryposis [75] *ADGRG6*: is associated with earlobe attachment [76] *GJE1*: is associated with hearing loss [77] *NMBR*: a member of the mammalian bombesin receptor family involved in several physiological effects [78]
rs11230097	11	59646339	T/C	0.182	7.26E-6	0.60 (0.48-0.75)	*TCN1*: is associated with colorectal cancer [79] *OOSP* family: is associated with postfertilization development *CBLIF*: encodes a glycoprotein *MRPL16*: is associated with colorectal tumors [80] *STX3*: participates in granule-granule fusion and is associated with Sjögren's syndrome [81] *OR10V1*: encodes olfactory receptor involved in smell perception *MS4A3*: is associated with myeloid differentiation in human hematopoiesis [82]
rs35508695	6	152377709	G/A	0.071	7.30E-6	0.51 (0.38-0.69)	*ESR1*: encodes estrogen receptor alpha and is associated with periodontitis [61], and dental fluorosis [62] *SYNE1*: encodes spectrin repeat containing protein associated with spinocerebellar ataxia [83]
rs3111790	4	80558720	G/T	0.387	7.39E-6	0.63 (0.52-0.77)	NA
rs78891138	3	109679425	T/C	0.055	7.75E-6	0.46 (0.33-0.65)	NA
rs139737462	4	126756165	G/A	0.056	9.85E-6	0.45 (0.32-0.64)	NA

Note: EA: effect allele; RA: reference allele; MAF: minor allele frequency; HR: hazard ratios; NA: intergenic (gene desert).

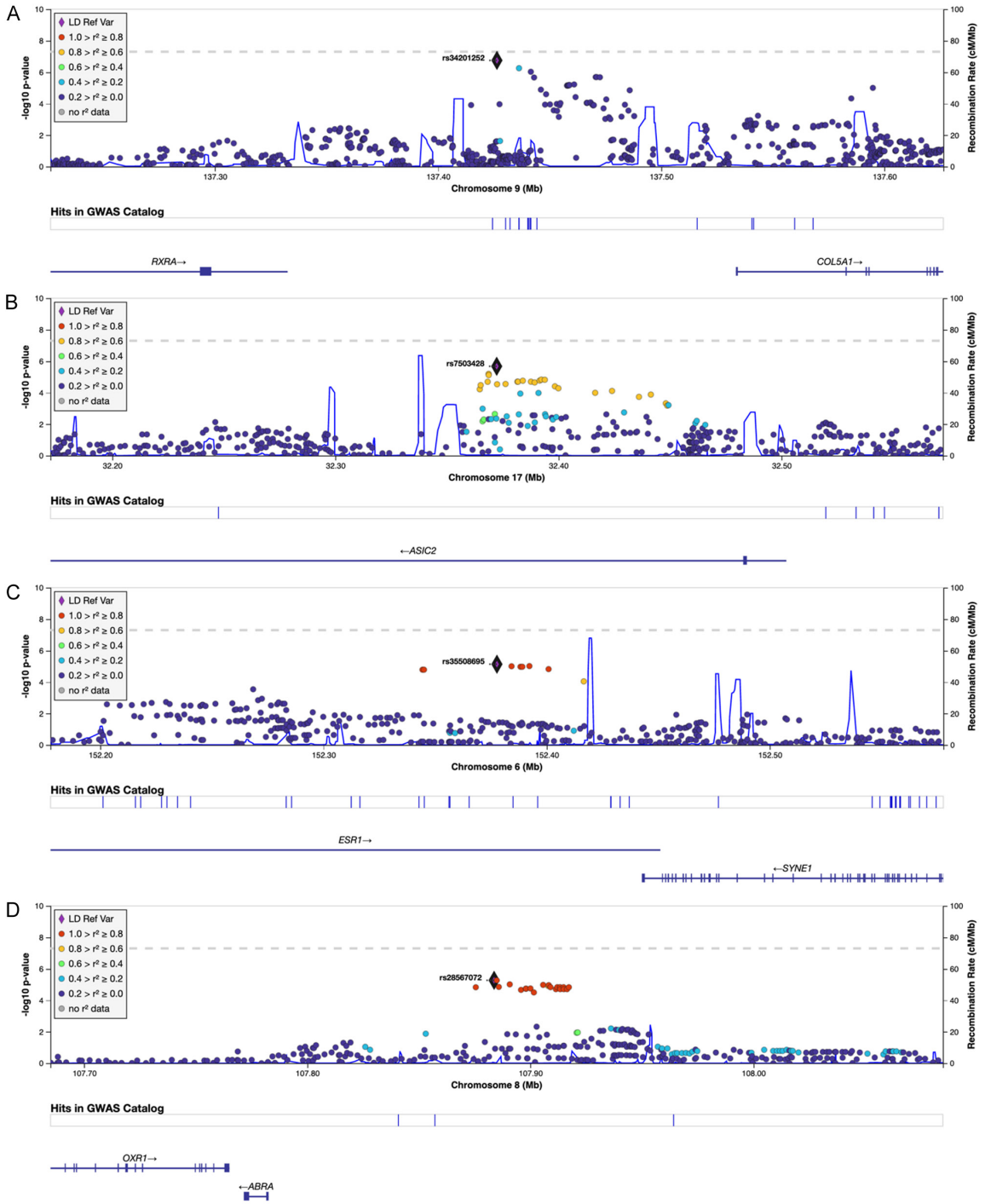


Figure 5. Regional association plots for the 4 loci located in or near genes with plausible function in dental caries. A. rs34201252 (chr9:137426383) near the gene *COL5A1*; B. rs7503428 (chr17:32372318) in the gene *ASIC2*; C. rs35508695 (chr6:152377709) in the gene *ESR1*; D. rs28567072 (chr8:107883740) near the gene *ABRA*. Each point represents the significance (-$\log_{10}$-transformed *P* value; left Y-axis) of a SNP. The purple diamond point with rsID label is the lead SNP in the association region. The color of the points represents the linkage disequilibrium (i.e., correlation [r^2]) of the SNP with the lead SNP. Right Y-axis (recombination rate) provides information about the LD structure of region. The rug plot labeled "Hits in GWAS Catalog" below the plot marks the positions of SNPs that have been previously reported in the GWAS Catalog to be associated with a trait. Genes near the association signal are shown on the bottom.

causal effects from confounding indicators of rurality. Moreover, despite the caries-preventive benefit of water fluoridation [44], we did not find a significant association between home fluoride level and caries incidence, which is likely due to the small variance of home fluoride level in our sample.

There are mixed results regarding the association between breastfeeding practices and dental caries in previous studies. Some studies showed that bottle feeding [45] and prolonged breastfeeding [46, 47] increased caries risk, while others showing no associations for breastfeeding practices [48-50]. In our study, breastfeeding and longer breastfeeding duration decreased the hazards for caries incidence in univariate models, but they were not significant in the final stepwise model. The correlations between breastfeeding practices and SES variables that were retained in our model make it difficult to independently assess the effect of breastfeeding practices.

Here we also present the first survival GWAS to date for caries incidence in the primary dentition. The SNP-based heritability estimate of time to first caries incidence was 0.54 (SD: 0.41), which is similar to the heritability estimates of caries scores measured at a single point in time in the primary dentition (0.54-0.70) from a family study [21], slightly higher than the heritability estimates of the incremental increase in caries scores (0.30) from a twin study [22], and slightly higher than the SNP-based heritability estimate of caries status (yes or no) in the primary dentition (0.28 [95% CI: 0.09-0.48]) from a GWAS [28]. The SNP-based heritability estimate in our study only reflects the genetic contributions of common variants; other genetic components (e.g., low frequency and rare variants, copy number variants, structural variants, or genetic interactions) were not considered. In addition, the relatively large SD of the heritability estimate indicates the uncertainty of heritability modeling. One explanation for the uncertainty is that our sample size (under 5000) was too small to calculate the precise estimate, hence the large SD. Future studies with larger sample size are needed to estimate the heritability more precisely and further explore the genetic contributions of caries incidence.

Of the 14 identified loci at the suggestive significance level, 3 loci are located within or near genes with plausibly biological functions in dental caries (*COL5A1*, *ASIC2*, and *ESR1*). *COL5A1* encodes type V collagen which is a dominant regulator of collagen fibrillogenesis [51]. Mutations in *COL5A1* have been reported to be associated with Ehlers-Danlos syndrome (EDS) [52]. EDS patients experience a higher prevalence of oral problems including caries [53], tooth fractures, and periodontal disease [54], as well as pulp calcification and abnormal tooth roots [55]. *ASIC2* encodes acid-sensing ion channel 2 detected in the nerves supplying the taste buds of zebrafish [56] and in human odontoblasts implying it has a role in tooth mechanosensitivity [57]. Its association with severe gingival inflammation has been identified in a GWAS [58] and its involvement in the inflammation process has been shown in knockout mice [59]. *ESR1* encodes estrogen receptor alpha that is expressed in dental tissues and cells, such as ameloblasts, odontoblasts, and dental pulp and is involved in enamel formation [60]. Associations with *ESR1* have been identified for several dental traits including chronic periodontitis [61], dental fluorosis [62, 63], and tooth size [64]. Though these genes have been shown to be related to dental traits, it is currently unknown whether they are involved in genetic susceptibility to dental caries, specifically. Notably, the locus at chromosome 8 is located near gene *ABRA* which has been reported to be associated with dental caries in the permanent dentition at the suggestive significance level ($P \leq 10E\text{-}5$) in a prior GWAS [42]. However, there were no significant associations between time to first caries incidence and previously reported SNPs associated with caries traits at the more liberal significance level in the candidate SNP replication analysis. Failure to replicate these previous genetic associations may be due to the different phenotypes between this time-to-event study and prior studies, population differences, low statistical power to detect weak effects, or prior false positive results.

Though there were no genome-wide significant genes associated with caries incidence in the gene-based analysis, we found the top gene *TBX1* is a major candidate gene for 22q11.2 Deletion Syndrome, which causes craniofacial malformations including dental defects and

cleft palate [65]. *TBX1* encodes a member of the T-box gene family involved in tooth morphogenesis by regulating the proliferation, differentiation, and maturation of ameloblasts [66]. In vivo, microdontia with decreased stem cell proliferation was observed in *Tbx1* conditional knockout mice, while increased dental epithelial progenitor cells was shown in *Tbx1* overexpressed mice [67]. Given the role of *TBX1* in tooth development, it is plausible that this gene has effects on dental caries onset by influencing the timing of tooth eruption.

There are some limitations in this study. First, the sample size was very small for a GWAS. Any variants with small effect sizes or low frequency may not have been detectable in this survival GWAS. Second, external study samples with the necessary data (i.e., longitudinal caries assessments and genotyping data) are not currently available to replicate the identified variants in this study. Thus, independent studies with larger sample size are needed in the future. Third, timing of tooth eruption, which is variable and has been shown to be influenced by genetic factors [68], impacts the time that primary teeth are at-risk, and thus may affect caries incidence. Although we do not have data on the timing of tooth eruption across the dentition, and therefore we cannot incorporate this into our models of caries incidence, we tested the associations between caries incidence and tooth eruption-related SNPs that were reported in previous studies. No tooth eruption-associated SNPs were significantly associated with caries incidence after Bonferroni correction (Supplementary Table 2). In addition, we calculated the association and genetic correlation between caries incidence and timing of the eruption of the first tooth using the subset of our sample with such data (n = 469). Note, the first tooth to erupt is the only tooth for which we collected timing data; this is most often a mandibular central incisor, a very low-risk tooth and likely not to exhibit incident dental caries, but the timing of its eruption may serve as an indicator of the initiation of dental eruption, in general. Strictly speaking, the timing of the first tooth eruption was not associated with caries incidence (HR: 0.22 [95% CI: 0.05-1.04], P = 0.06), however, the trend was suggestive of a protective effect (i.e., later tooth eruption is protective). A weak genetic correlation (Rg: 0.22, SD: 1.09) was observed, indicating that the timing of initiation of dental eruption, while possibly contributing, does not fully explain the heritability of caries incidence, suggesting other mechanisms are at play.

In conclusion, we reported household income, home water source, and mother's educational attainment as risk factors for caries incidence and nominated several genes with biologically plausible roles in caries incidence for future investigations. Understanding the risk factors and genetic contributions to caries incidence may provide the foundation for better early detection, risk assessment and personalized interventions, and thereby improving the overall oral health.

Acknowledgements

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: R01-DE014899 and X01-HG009878.

Disclosure of conflict of interest

None.

Address correspondence to: Mary L Marazita and John R Shaffer, Department of Human Genetics, School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. E-mail: marazita@pitt.edu (MLM); john.r.shaffer@pitt.edu (JRS)

References

[1] Fleming E and Afful J. Prevalence of total and untreated dental caries among youth: United States, 2015-2016. NCHS Data Brief 2018; 1-8.

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

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

[4] Schwendicke F, Dorfer CE, Schlattmann P, Foster Page L, Thomson WM and Paris S. Socioeconomic inequality and caries: a systematic review and meta-analysis. J Dent Res 2015; 94: 10-18.

[5] Whelton HP, Spencer AJ, Do LG and Rugg-Gunn AJ. Fluoride revolution and dental caries: evo-

lution of policies for global use. J Dent Res 2019; 98: 837-846.
[6] Pitts NB, Zero DT, Marsh PD, Ekstrand K, Weintraub JA, Ramos-Gomez F, Tagami J, Twetman S, Tsakos G and Ismail A. Dental caries. Nat Rev Dis Primers 2017; 3: 17030.
[7] Blostein FA, Jansen EC, Jones AD, Marshall TA and Foxman B. Dietary patterns associated with dental caries in adults in the United States. Community Dent Oral Epidemiol 2020; 48: 119-129.
[8] Sheiham A and James WP. A reappraisal of the quantitative relationship between sugar intake and dental caries: the need for new criteria for developing goals for sugar intake. BMC Public Health 2014; 14: 863.
[9] Bernabe E, Vehkalahti MM, Sheiham A, Aromaa A and Suominen AL. Sugar-sweetened beverages and dental caries in adults: a 4-year prospective study. J Dent 2014; 42: 952-958.
[10] Ramadugu K, Blostein F, Bhaumik D, Jiang W, Davis E, Salzman E, Srinivasan U, Marrs CF, Neiswanger K, McNeil DW, Marazita ML and Foxman B. Co-occurrence of yeast, streptococci, dental decay, and gingivitis in the post-partum period: results of a longitudinal study. J Oral Microbiol 2020; 12: 1746494.
[11] Peltzer K, Mongkolchati A, Satchaiyan G, Rajchagool S and Pimpak T. Sociobehavioral factors associated with caries increment: a longitudinal study from 24 to 36 months old children in Thailand. Int J Environ Res Public Health 2014; 11: 10838-10850.
[12] Meurman PK and Pienihakkinen K. Factors associated with caries increment: a longitudinal study from 18 months to 5 years of age. Caries Res 2010; 44: 519-524.
[13] Ismail AI, Sohn W, Lim S and Willem JM. Predictors of dental caries progression in primary teeth. J Dent Res 2009; 88: 270-275.
[14] Leroy R, Bogaerts K, Martens L and Declerck D. Risk factors for caries incidence in a cohort of Flemish preschool children. Clin Oral Investig 2012; 16: 805-812.
[15] Lim S, Tellez M and Ismail AI. Dental caries development among African American children: results from a 4-year longitudinal study. Community Dent Oral Epidemiol 2015; 43: 200-207.
[16] Cogulu D, Ersin NK, Uzel A, Eronat N and Aksit S. A long-term effect of caries-related factors in initially caries-free children. Int J Paediatr Dent 2008; 18: 361-367.
[17] Ghazal T, Levy SM, Childers NK, Broffitt B, Cutter GR, Wiener HW, Kempf MC, Warren J and Cavanaugh JE. Factors associated with early childhood caries incidence among high caries-risk children. Community Dent Oral Epidemiol 2015; 43: 366-374.
[18] Ollila P and Larmas M. A seven-year survival analysis of caries onset in primary second molars and permanent first molars in different caries risk groups determined at age two years. Acta Odontol Scand 2007; 65: 29-35.
[19] Lee HJ, Kim JB, Jin BH, Paik DI and Bae KH. Risk factors for dental caries in childhood: a five-year survival analysis. Community Dent Oral Epidemiol 2015; 43: 163-171.
[20] Ghazal TS, Levy SM, Childers NK, Carter KD, Caplan DJ, Warren JJ and Kolker JL. Survival analysis of caries incidence in African-American school-aged children. J Public Health Dent 2019; 79: 10-17.
[21] Wang X, Shaffer JR, Weyant RJ, Cuenco KT, DeSensi RS, Crout R, McNeil DW and Marazita ML. Genes and their effects on dental caries may differ between primary and permanent dentitions. Caries Res 2010; 44: 277-284.
[22] Bretz WA, Corby PM, Schork NJ, Robinson MT, Coelho M, Costa S, Melo Filho MR, Weyant RJ and Hart TC. Longitudinal analysis of heritability for dental caries traits. J Dent Res 2005; 84: 1047-1051.
[23] Bretz WA, Corby PM, Melo MR, Coelho MQ, Costa SM, Robinson M, Schork NJ, Drewnowski A and Hart TC. Heritability estimates for dental caries and sucrose sweetness preference. Arch Oral Biol 2006; 51: 1156-1160.
[24] Bretz WA, Corby PM, Hart TC, Costa S, Coelho MQ, Weyant RJ, Robinson M and Schork NJ. Dental caries and microbial acid production in twins. Caries Res 2005; 39: 168-172.
[25] Shaffer JR, Wang X, Desensi RS, Wendell S, Weyant RJ, Cuenco KT, Crout R, McNeil DW and Marazita ML. Genetic susceptibility to dental caries on pit and fissure and smooth surfaces. Caries Res 2012; 46: 38-46.
[26] Shaffer JR, Wang X, Feingold E, Lee M, Begum F, Weeks DE, Cuenco KT, Barmada MM, Wendell SK, Crosslin DR, Laurie CC, Doheny KF, Pugh EW, Zhang Q, Feenstra B, Geller F, Boyd HA, Zhang H, Melbye M, Murray JC, Weyant RJ, Crout R, McNeil DW, Levy SM, Slayton RL, Willing MC, Broffitt B, Vieira AR and Marazita ML. Genome-wide association scan for childhood caries implicates novel genes. J Dent Res 2011; 90: 1457-1462.
[27] Zeng Z, Shaffer JR, Wang X, Feingold E, Weeks DE, Lee M, Cuenco KT, Wendell SK, Weyant RJ, Crout R, McNeil DW and Marazita ML. Genome-wide association studies of pit-and-fissure- and smooth-surface caries in permanent dentition. J Dent Res 2013; 92: 432-437.
[28] Haworth S, Shungin D, van der Tas JT, Vucic S, Medina-Gomez C, Yakimov V, Feenstra B, Shaffer JR, Lee MK, Standl M, Thiering E, Wang C, Bonnelykke K, Waage J, Jessen LE, Norrisgaard PE, Joro R, Seppala I, Raitakari O, Dud-

ding T, Grgic O, Ongkosuwito E, Vierola A, Eloranta AM, West NX, Thomas SJ, McNeil DW, Levy SM, Slayton R, Nohr EA, Lehtimaki T, Lakka T, Bisgaard H, Pennell C, Kuhnisch J, Marazita ML, Melbye M, Geller F, Rivadeneira F, Wolvius EB, Franks PW, Johansson I and Timpson NJ. Consortium-based genome-wide meta-analysis for childhood dental caries traits. Hum Mol Genet 2018; 27: 3113-3127.

[29] Ballantine JL, Carlson JC, Ferreira Zandona AG, Agler C, Zeldin LP, Rozier RG, Roberts MW, Basta PV, Luo J, Antonio-Obese ME, McNeil DW, Weyant RJ, Crout RJ, Slayton RL, Levy SM, Shaffer JR, Marazita ML, North KE and Divaris K. Exploring the genomic basis of early childhood caries: a pilot study. Int J Paediatr Dent 2018; 28: 217-225.

[30] Orlova E, Carlson JC, Lee MK, Feingold E, McNeil DW, Crout RJ, Weyant RJ, Marazita ML and Shaffer JR. Pilot GWAS of caries in African-Americans shows genetic heterogeneity. BMC Oral Health 2019; 19: 215.

[31] Neiswanger K, McNeil DW, Foxman B, Govil M, Cooper ME, Weyant RJ, Shaffer JR, Crout RJ, Simhan HN, Beach SR, Chapman S, Zovko JG, Brown LJ, Strotmeyer SJ, Maurer JL and Marazita ML. Oral health in a sample of pregnant women from Northern Appalachia (2011-2015). Int J Dent 2015; 2015: 469376.

[32] Smith G. Step away from stepwise. J Big Data 2018; 5: 32.

[33] Rizvi AA, Karaesmen E, Morgan M, Preus L, Wang J, Sovic M, Hahn T and Sucheston-Campbell LE. Gwasurvivr: an R package for genome-wide survival analysis. Bioinformatics 2019; 35: 1968-1970.

[34] Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR and Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; 26: 2336-2337.

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

[36] Yang J, Zhu W, Chen J, Zhang Q and Wu S. Genome-wide two-marker linkage disequilibrium mapping of quantitative trait loci. BMC Genet 2014; 15: 20.

[37] Morrison J, Laurie CC, Marazita ML, Sanders AE, Offenbacher S, Salazar CR, Conomos MP, Thornton T, Jain D, Laurie CA, Kerr KF, Papanicolaou G, Taylor K, Kaste LM, Beck JD and Shaffer JR. Genome-wide association study of dental caries in the Hispanic Communities Health Study/Study of Latinos (HCHS/SOL). Hum Mol Genet 2016; 25: 807-816.

[38] Shungin D, Haworth S, Divaris K, Agler CS, Kamatani Y, Keun Lee M, Grinde K, Hindy G, Alaraudanjoki V, Pesonen P, Teumer A, Holtfreter B, Sakaue S, Hirata J, Yu YH, Ridker PM, Giulianini F, Chasman DI, Magnusson PKE, Sudo T, Okada Y, Volker U, Kocher T, Anttonen V, Laitala ML, Orho-Melander M, Sofer T, Shaffer JR, Vieira A, Marazita ML, Kubo M, Furuichi Y, North KE, Offenbacher S, Ingelsson E, Franks PW, Timpson NJ and Johansson I. Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data. Nat Commun 2019; 10: 2773.

[39] de Leeuw CA, Mooij JM, Heskes T and Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015; 11: e1004219.

[40] McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM and Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010; 28: 495-501.

[41] Speed D and Balding DJ. SumHer better estimates the SNP heritability of complex traits from summary statistics. Nat Genet 2019; 51: 277-284.

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

[43] Heaton B, Crawford A, Garcia RI, Henshaw M, Riedy CA, Barker JC and Wimsatt MA; Native Oral Health Project. Oral health beliefs, knowledge, and behaviors in Northern California American Indian and Alaska Native mothers regarding early childhood caries. J Public Health Dent 2017; 77: 350-359.

[44] Slade GD, Grider WB, Maas WR and Sanders AE. Water fluoridation and dental caries in U.S. children and adolescents. J Dent Res 2018; 97: 1122-1128.

[45] Qadri G, Nourallah A and Splieth CH. Early childhood caries and feeding practices in kindergarten children. Quintessence Int 2012; 43: 503-510.

[46] Kato T, Yorifuji T, Yamakawa M, Inoue S, Saito K, Doi H and Kawachi I. Association of breast feeding with early childhood dental caries: Japanese population-based study. BMJ Open 2015; 5: e006982.

[47] Jain M, Namdev R, Bodh M, Dutta S, Singhal P and Kumar A. Social and behavioral determinants for early childhood caries among pre-

school children in India. J Dent Res Dent Clin Dent Prospects 2015; 9: 115-120.
[48] Nirunsittirat A, Pitiphat W, McKinney CM, DeRouen TA, Chansamak N, Angwaravong O, Patcharanuchat P and Pimpak T. Breastfeeding duration and childhood caries: a cohort study. Caries Res 2016; 50: 498-507.
[49] Valaitis R, Hesch R, Passarelli C, Sheehan D and Sinton J. A systematic review of the relationship between breastfeeding and early childhood caries. Can J Public Health 2000; 91: 411-417.
[50] Perera PJ, Fernando MP, Warnakulasooriya TD and Ranathunga N. Effect of feeding practices on dental caries among preschool children: a hospital based analytical cross sectional study. Asia Pac J Clin Nutr 2014; 23: 272-277.
[51] Mak KM, Png CY and Lee DJ. Type V collagen in health, disease, and fibrosis. Anat Rec (Hoboken) 2016; 299: 613-629.
[52] Symoens S, Syx D, Malfait F, Callewaert B, De Backer J, Vanakker O, Coucke P and De Paepe A. Comprehensive molecular analysis demonstrates type V collagen mutations in over 90% of patients with classic EDS and allows to refine diagnostic criteria. Hum Mutat 2012; 33: 1485-1493.
[53] De Coster PJ, Martens LC and De Paepe A. Oral health in prevalent types of Ehlers-Danlos syndromes. J Oral Pathol Med 2005; 34: 298-307.
[54] Hagberg C, Berglund B, Korpe L and Andersson-Norinder J. Ehlers-danlos syndrome (EDS) focusing on oral symptoms: a questionnaire study. Orthod Craniofac Res 2004; 7: 178-185.
[55] Kapferer-Seebacher I, Schnabl D, Zschocke J and Pope FM. Dental manifestations of Ehlers-Danlos syndromes: a systematic review. Acta Derm Venereol 2020; 100: adv00092.
[56] Levanti M, Randazzo B, Viña E, Montalbano G, Garcia-Suarez O, Germanà A, Vega JA and Abbate F. Acid-sensing ion channels and transient-receptor potential ion channels in zebrafish taste buds. Ann Anat 2016; 207: 32-37.
[57] Solé-Magdalena A, Revuelta EG, Menénez-Díaz I, Calavia MG, Cobo T, García-Suárez O, Pérez-Piñera P, De Carlos F, Cobo J and Vega JA. Human odontoblasts express transient receptor protein and acid-sensing ion channel mechanosensor proteins. Microsc Res Tech 2011; 74: 457-463.
[58] Zhang S, Divaris K, Moss K, Yu N, Barros S, Marchesan J, Morelli T, Agler C, Kim SJ, Wu D, North KE, Beck J and Offenbacher S. The novel ASIC2 locus is associated with severe gingival inflammation. JDR Clin Trans Res 2016; 1: 163-170.
[59] Gannon KP, McKey SE, Stec DE and Drummond HA. Altered myogenic vasoconstriction and regulation of whole kidney blood flow in the ASIC2 knockout mouse. Am J Physiol Renal Physiol 2015; 308: F339-348.
[60] Jedeon K, Loiodice S, Salhi K, Le Normand M, Houari S, Chaloyard J, Berdal A and Babajko S. Androgen receptor involvement in rat amelogenesis: an additional way for endocrine-disrupting chemicals to affect enamel synthesis. Endocrinology 2016; 157: 4287-4296.
[61] Shaffer JR, Polk DE, Wang X, Feingold E, Weeks DE, Lee MK, Cuenco KT, Weyant RJ, Crout RJ, McNeil DW and Marazita ML. Genome-wide association study of periodontal health measured by probing depth in adults ages 18-49 years. G3 (Bethesda) 2014; 4: 307-314.
[62] Dalledone M, Cunha AS, Ramazzotto LA, Pecharki GD, Nelson-Filho P, Scariot R, Trevilatto PC, Vieira AR, Küchler EC and Brancher JA. Estrogen receptor gene is associated with dental fluorosis in Brazilian children. Clin Oral Investig 2019; 23: 3565-3570.
[63] Saha D, Goswami R, Majumdar KK, Sikdar N and Pramanik S. Evaluating the association between dental fluorosis and polymorphisms in bone development and mineralization genes among population from a fluoride endemic region of Eastern India. Biol Trace Elem Res 2021; 199: 1-8.
[64] Cunha AS, Vertuan Dos Santos L, Schaffer Pugsley Baratto S, Abbasoglu Z, Gerber JT, Paza A, Matsumoto MAN, Scariot R, Stuani MB and Küchler EC. Human permanent tooth sizes are associated with genes encoding oestrogen receptors. J Orthod 2021; 48: 24-32.
[65] Jerome LA and Papaioannou VE. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. Nat Genet 2001; 27: 286-291.
[66] Gao S, Li X and Amendt BA. Understanding the role of Tbx1 as a candidate gene for 22q11.2 deletion syndrome. Curr Allergy Asthma Rep 2013; 13: 613-621.
[67] Gao S, Moreno M, Eliason S, Cao H, Li X, Yu W, Bidlack FB, Margolis HC, Baldini A and Amendt BA. TBX1 protein interactions and microRNA-96-5p regulation controls cell proliferation during craniofacial and dental development: implications for 22q11.2 deletion syndrome. Hum Mol Genet 2015; 24: 2330-2348.
[68] Hughes TE, Bockmann MR, Seow K, Gotjamanos T, Gully N, Richards LC and Townsend GC. Strong genetic control of emergence of human primary incisors. J Dent Res 2007; 86: 1160-1165.
[69] Vache C, Puechberty J, Faugere V, Darmaisin F, Liquori A, Baux D, Blanchet C, Garcia-Garcia G, Meunier I, Pellestor F, Koenig M and Roux AF. A 4.6 Mb inversion leading to PCDH15-LINC00844 and BICC1-PCDH15 fusion transcripts as a new pathogenic mechanism impli-

cated in usher syndrome type 1. Front Genet 2020; 11: 623.
[70] Giunta C, Nuytinck L, Raghunath M, Hausser I, De Paepe A and Steinmann B. Homozygous Gly530Ser substitution in COL5A1 causes mild classical ehlers-danlos syndrome. Am J Med Genet 2002; 109: 284-290.
[71] Barbato E, Traversa A, Guarnieri R, Giovannetti A, Genovesi ML, Magliozzi MR, Paolacci S, Ciolfi A, Pizzi S, Di Giorgio R, Tartaglia M, Pizzuti A and Caputo V. Whole exome sequencing in an Italian family with isolated maxillary canine agenesis and canine eruption anomalies. Arch Oral Biol 2018; 91: 96-102.
[72] Lamon S, Wallace MA, Leger B and Russell AP. Regulation of STARS and its downstream targets suggest a novel pathway involved in human skeletal muscle hypertrophy and atrophy. J Physiol 2009; 587: 1795-1803.
[73] Troidl K, Ruding I, Cai WJ, Mucke Y, Grossekettler L, Piotrowska I, Apfelbeck H, Schierling W, Volger OL, Horrevoets AJ, Grote K, Schmitz-Rixen T, Schaper W and Troidl C. Actin-binding rho activating protein (Abra) is essential for fluid shear stress-induced arteriogenesis. Arterioscler Thromb Vasc Biol 2009; 29: 2093-2101.
[74] Oliver PL, Finelli MJ, Edwards B, Bitoun E, Butts DL, Becker EB, Cheeseman MT, Davies B and Davies KE. Oxr1 is essential for protection against oxidative stress-induced neurodegeneration. PLoS Genet 2011; 7: e1002338.
[75] Ward DM, Vaughn MB, Shiflett SL, White PL, Pollock AL, Hill J, Schnegelberger R, Sundquist WI and Kaplan J. The role of LIP5 and CHMP5 in multivesicular body formation and HIV-1 budding in mammalian cells. J Biol Chem 2005; 280: 10548-10555.
[76] Shaffer JR, Li J, Lee MK, Roosenboom J, Orlova E, Adhikari K; 23andMe Research Team; Gallo C, Poletti G, Schuler-Faccini L, Bortolini MC, Canizales-Quinteros S, Rothhammer F, Bedoya G, Gonzalez-Jose R, Pfeffer PE, Wollenschlaeger CA, Hecht JT, Wehby GL, Moreno LM, Ding A, Jin L, Yang Y, Carlson JC, Leslie EJ, Feingold E, Marazita ML, Hinds DA, Cox TC, Wang S, Ruiz-Linares A and Weinberg SM. Multiethnic GWAS reveals polygenic architecture of earlobe attachment. Am J Hum Genet 2017; 101: 913-924.
[77] Locher H, de Groot JC, van Iperen L, Huisman MA, Frijns JH and Chuva de Sousa Lopes SM. Development of the stria vascularis and potassium regulation in the human fetal cochlea: insights into hereditary sensorineural hearing loss. Dev Neurobiol 2015; 75: 1219-1240.
[78] Park HJ, Kim MK, Kim SR, Bae SK and Bae MK. Hypoxia regulates the expression of the neuromedin B receptor through a mechanism dependent on hypoxia-inducible factor-1alpha. PLoS One 2013; 8: e82868.
[79] Zhu X, Yi K, Hou D, Huang H, Jiang X, Shi X and Xing C. Clinicopathological analysis and prognostic assessment of transcobalamin I (TCN1) in patients with colorectal tumors. Med Sci Monit 2020; 26: e923828.
[80] Kim JC, Kim SY, Roh SA, Cho DH, Kim DD, Kim JH and Kim YS. Gene expression profiling: canonical molecular changes and clinicopathological features in sporadic colorectal cancers. World J Gastroenterol 2008; 14: 6662-6672.
[81] Barrera MJ, Sanchez M, Aguilera S, Alliende C, Bahamondes V, Molina C, Quest AF, Urzua U, Castro I, Gonzalez S, Sung HH, Albornoz A, Hermoso M, Leyton C and Gonzalez MJ. Aberrant localization of fusion receptors involved in regulated exocytosis in salivary glands of Sjogren's syndrome patients is linked to ectopic mucin secretion. J Autoimmun 2012; 39: 83-92.
[82] Ishibashi T, Yokota T, Satoh Y, Ichii M, Sudo T, Doi Y, Ueda T, Nagate Y, Hamanaka Y, Tanimura A, Ezoe S, Shibayama H, Oritani K and Kanakura Y. Identification of MS4A3 as a reliable marker for early myeloid differentiation in human hematopoiesis. Biochem Biophys Res Commun 2018; 495: 2338-2343.
[83] Indelicato E, Nachbauer W, Fauth C, Krabichler B, Schossig A, Eigentler A, Dichtl W, Wenning G, Wagner M, Fanciulli A, Janecke A and Boesch S. SYNE1-ataxia: novel genotypic and phenotypic findings. Parkinsonism Relat Disord 2019; 62: 210-214.

Supplementary Materials

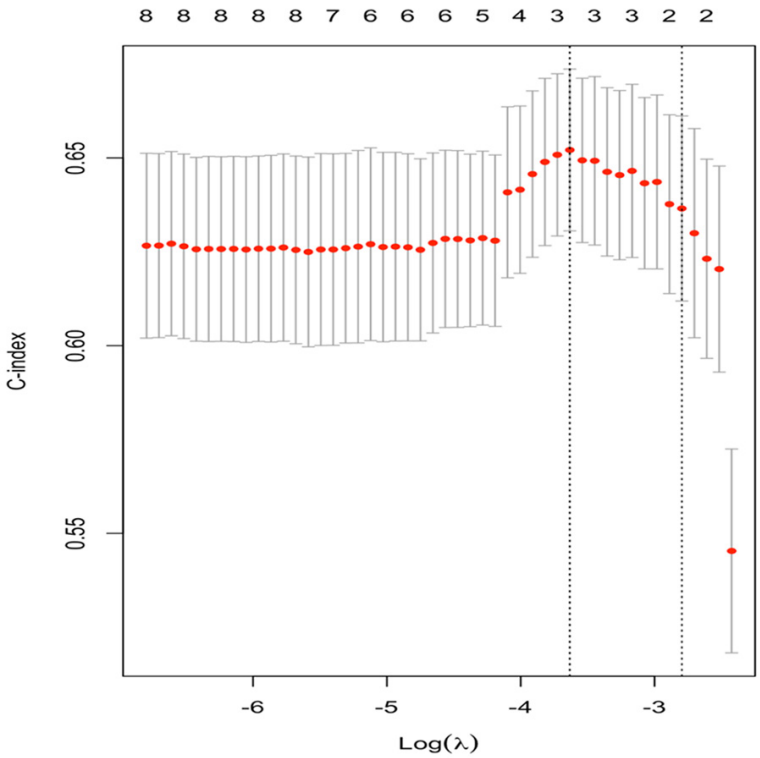
Methods

Risk factor modeling

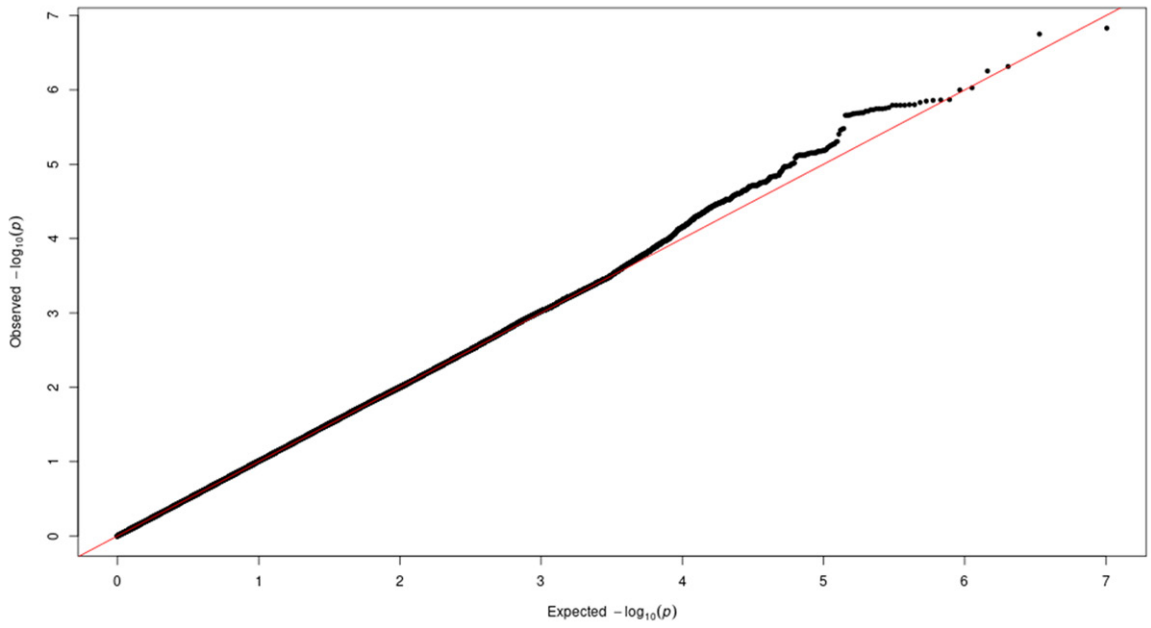
To verify the variable selection results, we performed a second procedure, regularized Cox modeling, using the R package glmnet [1]. Glmnet penalizes the negative log of the partial likelihood with an elastic net penalty and uses cyclical coordinate descent (i.e., optimizing function over each parameter with others fixed successively and cycling repeatedly until convergence). We first used the function "glmnet" to compute the solution path under default settings including all covariates, then performed cross-validation using the Harell C index to evaluate model performance. Higher C index indicates better performance. We chose the model with highest C index as the best regularized Cox model. The three variables selected in the best regularized Cox model were household income, home water source, and mother's educational attainment, which was the same set of variables as in the stepwise Cox model.

Candidate SNP replication

We tested the associations between time to first caries incidence and variants reported to be associated with dental caries traits at the genome-wide significance level ($P < 5E\text{-}8$) in previous GWASs. In total, 25 variants were reported in 5 prior GWASs [2-6]. Among these 25 variants, 2 variants had minor allele frequencies of 0% and 3 variants had no genotyping information in our data. We tested proxy SNPs based on high LD ($r^2 \geq 0.75$) for the 3 variants without genotyping information. Thus, 23 variants were tested in this study.



Supplementary Figure 1. The C-index generated from cross-validation under different values of λ. The y-axis is the C-index, the x-axis is the natural logarithm of λ, which is the tuning parameter controlling the overall strength of the penalty. The integers above the plot indicate the number of variables included in the Cox model at the current Log (λ). The left vertical line shows the highest C-index indicating the best Cox model. The 3 variables included in the best Cox model were household income, home water source, and mother's educational attainment. The results of regularized Cox model were the same as the stepwise Cox model.



Supplementary Figure 2. Quantile-quantile plot for the survival GWAS of time to first caries incidence. The genomic inflation factor was 1.00.

Supplementary Table 1. Associations between time to first caries incidence with previously reported variants associated with dental caries in both primary and permanent dentitions

SNP	Chr	BP	EA/RA	MAF	P	HR (95% CI)	Associated traits in previous GWASs
rs3896439	1	4668670	G/A	0.062	0.58	1.13 (0.74-1.72)	Partial DMFS[a] [2]
rs399593	10	30912030	T/G	0.134	0.4	1.14 (0.84-1.55)	
rs7607421	2	220792320	C/T	0.381	0.742	1.04 (0.84-1.28)	DMFS [3]
rs10790497	11	99033861	A/G	0.349	0.144	1.18 (0.95-1.46)	
rs17236529	3	160226307	T/C	0.018	0.451	1.46 (0.54-3.95)	PF.dfs[b] [7]
rs190395159[f]	7	105964857	A/G	0.001	0.586	1.73 (0.24-12.42)	DMFT [4]
rs138769355	3	151367491	A/G	0.002[e]	0.009	2.97 (1.31-6.70)	DMFS [4]
rs138642966	1	235695611	T/C	0.02	0.588	1.25 (0.56-2.82)	
rs116717469	10	62237360	C/T	0.001[e]	0.485	0.50 (0.07-3.55)	
rs71381322	17	48900824	G/A	0.024	0.82	1.08 (0.56-2.09)	
rs16946661	18	26946870	T/C	0.001[e]	0.343	0.39 (0.05-2.77)	
rs1594318	2	3733944	C/G	0.4	0.764	1.03 (0.84-1.26)	Binary in pimary[c] [5]
rs7738851	6	11241788	A/T	0.147	0.566	1.09 (0.81-1.47)	Binary in permanent[d] [5]
rs4971099[f]	1	155155731	A/G	0.426	0.358	1.10 (0.90-1.35)	DMFS [6]
rs121908120	2	219755011	T/A	0.028[e]	0.675	0.89 (0.53-1.51)	
rs1122171	5	134509987	T/C	0.439	0.738	0.97 (0.79-1.18)	
rs1482698	5	44539453	G/C	0.385	0.754	1.03 (0.84-1.27)	
rs9366651	6	26336696	T/G	0.47	0.974	1.00 (0.82-1.22)	
rs10987008[f]	9	128661600	A/G	0.36	0.464	1.08 (0.88-1.33)	
rs72748935	15	63639416	T/C	0.472	0.172	0.87 (0.71-1.06)	
rs10048146	16	86710660	A/G	0.168	0.607	1.07 (0.82-1.41)	
rs28822480	18	57924823	G/A	0.287	0.496	0.93 (0.75-1.15)	
rs11672900	19	49220323	A/G	0.496	0.398	1.09 (0.89-1.33)	

Note: Chr: chromosome; BP: base pair position; EA: effect allele; RA: reference allele; MAF: minor allele frequency; HR: hazard ratios; DMFS: decayed, missing, filled surfaces in the permanent dentition; DMFT: decayed, missing, filled teeth in the permanent dentition. [a]DMFS in mid-dentition (pre-molars and canines). [b]dfs in Pit-and Fissure surfaces. [c]binary trait (Yes/No affected) in the primary dentition. [d]binary trait (Yes/No affected) in the permanent dentition. [e]MAF < 0.05, excluded in the survival GWAS. [f]No genotyping information in our data. We listed the original rsID from previous GWASs in the table and tested their closest proxies ($r^2 \geq 0.75$). The proxy of rs190395159 is rs375025137 ($r^2 = 0.75$, distance = 854 bp). The proxy of rs4971099 is rs4971100 ($r^2 = 0.91$, distance = 123 bp). The proxy of rs10987008 is rs10987017 ($r^2 = 1$, distance = 7115 bp).

Supplementary Table 2. Associations between time to first caries incidence with previously reported variants associated with tooth eruption

SNP	Chr	BP	EA/RA	MAF	P	HR (95% CI)	Associated traits in previous GWASs
rs8079702	17	68190826	A/G	0.445	0.682	0.96 (0.78-1.17)	Age at first tooth eruption and/or number of teeth at 12 months [8]
rs4844096	23	68805318	G/A	0.463	0.311	1.09 (0.93-1.28)	
rs5936487	23	68892916	A/G	0.452	0.380	1.07 (0.92-1.26)	
rs10506525	12	65783378	C/T	0.393	0.618	1.05 (0.86-1.29)	
rs9674544	17	47084711	A/G	0.497	0.956	0.99 (0.82-1.21)	
rs1956529	14	68788924	T/C	0.387	0.031	0.80 (0.65-0.98)	
rs12424086	12	66364509	T/C	0.214	0.390	0.90 (0.71-1.14)	Number of permanent teeth [9]
rs4491709	2	217894756	T/C	0.271	0.702	0.96 (0.76-1.2)	
rs2281845	1	201081943	C/T	0.401	0.596	0.95 (0.78-1.15)	
rs7924176	10	76295789	A/G	0.429	0.714	1.04 (0.85-1.26)	
rs10932688	2	217863481	C/G	0.238	0.772	1.04 (0.81-1.32)	Age at first tooth eruption and/or number of teeth at 12 months [10]
rs6568401	6	106188818	T/C	0.269	0.778	1.03 (0.83-1.29)	
rs1799922	7	128415195	T/G	0.397	0.244	0.89 (0.72-1.09)	
rs10740993	10	18442482	C/T	0.456	0.188	0.87 (0.72-1.07)	
rs7924176	10	76295789	A/G	0.429	0.714	1.04 (0.85-1.26)	
rs4937076	11	125826702	A/G	0.472	0.990	1.00 (0.82-1.23)	
rs12229918	12	65762058	G/C	0.365	0.801	1.03 (0.84-1.26)	
rs17101923	12	66338202	G/T	0.246	0.240	0.88 (0.7-1.09)	
rs9316505	13	51390598	A/G	0.437	0.619	0.95 (0.78-1.16)	
rs997154	14	23464482	G/A	0.222	0.511	0.92 (0.73-1.17)	
rs17563	14	54417522	G/A	0.464	0.996	1.00 (0.81-1.23)	
rs1994969	17	47080431	T/G	0.474	0.819	0.98 (0.8-1.19)	
rs412000	17	56709058	C/G	0.458	0.675	0.96 (0.78-1.17)	
rs8080944	17	68185586	A/G	0.421	0.437	0.92 (0.75-1.13)	
rs11796357	23	68798703	A/G	0.210	0.708	0.96 (0.8-1.16)	

Note: Chr: chromosome; BP: base pair position; EA: effect allele; RA: reference allele; MAF: minor allele frequency; HR: hazard ratios.

References

[1] Friedman J, Hastie T and Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010; 33: 1-22.

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

[3] Yang J, Zhu W, Chen J, Zhang Q and Wu S. Genome-wide two-marker linkage disequilibrium mapping of quantitative trait loci. BMC Genet 2014; 15: 20.

[4] Morrison J, Laurie CC, Marazita ML, Sanders AE, Offenbacher S, Salazar CR, Conomos MP, Thornton T, Jain D, Laurie CA, Kerr KF, Papanicolaou G, Taylor K, Kaste LM, Beck JD and Shaffer JR. Genome-wide association study of dental caries in the Hispanic Communities Health Study/Study of Latinos (HCHS/SOL). Hum Mol Genet 2016; 25: 807-816.

[5] Haworth S, Shungin D, van der Tas JT, Vucic S, Medina-Gomez C, Yakimov V, Feenstra B, Shaffer JR, Lee MK, Standl M, Thiering E, Wang C, Bønnelykke K, Waage J, Jessen LE, Nørrisgaard PE, Joro R, Seppälä I, Raitakari O, Dudding T, Grgic O, Ongkosuwito E, Vierola A, Eloranta AM, West NX, Thomas SJ, McNeil DW, Levy SM, Slayton R, Nohr EA, Lehtimäki T, Lakka T, Bisgaard H, Pennell C, Kühnisch J, Marazita ML, Melbye M, Geller F, Rivadeneira F, Wolvius EB, Franks PW, Johansson I and Timpson NJ. Consortium-based genome-wide meta-analysis for childhood dental caries traits. Hum Mol Genet 2018; 27: 3113-3127.

[6] Shungin D, Haworth S, Divaris K, Agler CS, Kamatani Y, Keun Lee M, Grinde K, Hindy G, Alaraudanjoki V, Pesonen P, Teumer A, Holtfreter B, Sakaue S, Hirata J, Yu YH, Ridker PM, Giulianini F, Chasman DI, Magnusson PKE, Sudo T, Okada Y, Völker U, Kocher T, Anttonen V, Laitala ML, Orho-Melander M, Sofer T, Shaffer JR, Vieira A, Marazita ML, Kubo M, Furuichi Y, North KE, Offenbacher S, Ingelsson E, Franks PW, Timpson NJ and Johansson I. Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data. Nat Commun 2019; 10: 2773.

[7] Zeng Z, Feingold E, Wang X, Weeks DE, Lee M, Cuenco DT, Broffitt B, Weyant RJ, Crout R, McNeil DW, Levy SM, Marazita ML and Shaffer JR. Genome-wide association study of primary dentition pit-and-fissure and smooth surface caries. Caries Res 2014; 48: 330-338.

[8] Pillas D, Hoggart CJ, Evans DM, O'Reilly PF, Sipilä K, Lähdesmäki R, Millwood IY, Kaakinen M, Netuveli G, Blane D, Charoen P, Sovio U, Pouta A, Freimer N, Hartikainen AL, Laitinen J, Vaara S, Glaser B, Crawford P, Timpson NJ, Ring SM, Deng G, Zhang W, McCarthy MI, Deloukas P, Peltonen L, Elliott P, Coin LJ, Smith GD and Jarvelin MR. Genome-wide association study reveals multiple loci associated with primary tooth development during infancy. PLoS Genet 2010; 6: e1000856.

[9] Geller F, Feenstra B, Zhang H, Shaffer JR, Hansen T, Esserlind AL, Boyd HA, Nohr EA, Timpson NJ, Fatemifar G, Paternoster L, Evans DM, Weyant RJ, Levy SM, Lathrop M, Smith GD, Murray JC, Olesen J, Werge T, Marazita ML, Sørensen TI and Melbye M. Genome-wide association study identifies four loci associated with eruption of permanent teeth. PLoS Genet 2011; 7: e1002275.

[10] Fatemifar G, Hoggart CJ, Paternoster L, Kemp JP, Prokopenko I, Horikoshi M, Wright VJ, Tobias JH, Richmond S, Zhurov AI, Toma AM, Pouta A, Taanila A, Sipila K, Lähdesmäki R, Pillas D, Geller F, Feenstra B, Melbye M, Nohr EA, Ring SM, St Pourcain B, Timpson NJ, Davey Smith G, Jarvelin MR and Evans DM. Genome-wide association study of primary tooth eruption identifies pleiotropic loci associated with height and craniofacial distances. Hum Mol Genet 2013; 22: 3807-3817.