Genome-Wide Association Study (GWAS) of dental caries in diverse populations

Rasha N. Alotaibi1,2,3*, Brian J. Howe4,5, Jonathan M. Chernus1,6, Nandita Mukhopadhyay1,2, Carla Sanchez1,2, Frederic W. B. Deleyiannis7, Katherine Neiswanger1,2, Carmencita Padilla8, Fernando A. Poletta9, Ieda M. Orioli10, Carmen J. Buxó11, Jacqueline T. Hecht1,2, George L. Wehby13, Ross E. Long14, Alexandre R. Vieira1, Seth M. Weinberg1,2, John R. Shaffer1,2,6, Lina M. Moreno Uribe4,15 and Mary L. Marazita1,2,6

Abstract
Background: Dental caries is one of the most common chronic diseases and is influenced by a complex interplay of genetic and environmental factors. Most previous genetic studies of caries have focused on identifying genes that contribute to dental caries in specific ethnic groups, usually of European descent.

Methods: The aim of this study is to conduct a genome-wide association study (GWAS) to identify associations affecting susceptibility to caries in a large multiethnic population from Argentina, the Philippines, Guatemala, Hungary, and the USA, originally recruited for studies of orofacial clefts (POFC, N = 3686). Ages of the participants ranged from 2 to 12 years for analysis of the primary dentition, and 18–60 years for analysis of the permanent dentition. For each participant, dental caries was assessed by counts of decayed and filled teeth (dft/DFT) and genetic variants (single nucleotide polymorphisms, SNPs) were genotyped or imputed across the entire genome. Caries was analyzed separately for the primary and permanent dentitions, with age, gender, and presence/absence of any type of OFC treated as covariates. Efficient Mixed-Model Association eXpedited (EMMAX) was used to test genetic association, while simultaneously accounting for relatedness and stratification.

Results: We identified several suggestive loci (5 × 10^{-8} < P < 5 × 10^{-6}) within or near genes with plausible biological roles for dental caries, including a cluster of taste receptor genes (TAS2R38, TAS2R3, TAS2R4, TASR25) on chromosome 7 for the permanent dentition analysis, and DLX3 and DLX4 on chromosome 17 for the primary dentition analysis. Genome-wide significant results were seen with SNPs in the primary dentition only; however, none of the identified genes near these variants have known roles in cariogenesis.

Conclusion: The results of this study warrant further investigation and may lead to a better understanding of cariogenesis in diverse populations, and help to improve dental caries prediction, prevention, and/or treatment in future.

Keywords: Genomics, Genetics, Dental caries, Ethnicity

Background
Dental caries is a common multifactorial disease, in which environment and genetics each play important roles. Dental caries is a global public health problem and if left untreated, can lead to serious complications including pain, infection, abscess and loss of teeth [1]. There are at least 2.3 billion people globally affected with caries of the permanent dentition, and more than...
530 million children with caries of the primary dentition [2]. The 2010 Global Burden of Disease, Injury and Risk Factors (GBD) study estimated that oral diseases, such as untreated dental caries and severe periodontitis, accounted for 15 million disability-adjusted life-years (DALYs) globally among people older than 60 years old; and untreated dental caries in permanent dentition was the most common condition in the entire 2010 GBD study (global prevalence for all ages is 35%) [3].

The role of environmental risk factors in causing dental caries is very well established, and includes factors such as lack of fluoride, poor oral hygiene habits, and consuming a diet high in sucrose [4, 5]. In addition, disparities in terms of income, educational level and access to dental services in areas such as Appalachia in the United States could be associated with an increased risk of oral diseases, especially dental caries [6]. Moreover, a systematic review study indicated that socioeconomic factors, such as educational level, occupation and income, are highly associated with dental caries among adults [7].

Over the last decade, human genetic studies have made significant progress identifying genetic risk factors associated with dental caries. Previous genome-wide association studies (GWAS) of dental caries, summarized in Table 1, have reported numerous associations. However, these prior studies were limited to a single ethnic group and/or in a single dentition [8–10]. Therefore, the aim of the current GWAS study is to investigate both primary and permanent dentitions in a multiethnic cohort of children and adults.

**Methods**

**Sample**

The study sample (N = 3686) comprised individuals recruited as part of a large international study of orofacial clefting. Individuals recruited for this study included affected cases, their unaffected family members, and unrelated healthy controls from the United States (Colorado, Iowa, Pennsylvania, Texas, and Puerto Rico) and internationally (Argentina, the Philippines, and Hungary). All participant provided written informed consent for themselves and for children younger than 18 years, informed consent was obtained from their parents or their legally authorized representative. Local ethical approval was obtained at each site and all methods in this study were performed in accordance with the Institutional Review Board policies and guidelines of the University of Pittsburgh and all of the other sites.

| Study                    | Sample size                      | Dental caries phenotype                                      | Genes                                         | Genome-wide significant associations |
|--------------------------|----------------------------------|-------------------------------------------------------------|-----------------------------------------------|-------------------------------------|
| Shaffer et al. 2011      | 1305 white children, age 3–12    | Binary affection status in primary dentition               | ACTN2, EDARADD, MPPED2, MTR, and LPO         | No                                  |
| Wang et al. 2012         | 7443 whites Adult, age 17–89     | Permanent decayed, missing, filled surfaces index (DMFS)    | RPS6KA2, ISL1, TLR2 RHOU, FZD1, PTK28, and ADMTS3, | No                                  |
| Shaffer et al. 2013      | 920 whites Adult, ages 18–75     | Permanent cluster-based partial DMFS                       | LYZL2, AJAP1                                  | Yes                                 |
| Zeng et al. 2013         | 1017 whites Adult, age 14–56     | Permanent decayed and filled teeth (dft) stratified to generate df-pit and fissures (dfPF) and df-smooth surface (dfSM) | BCOR, BCORL1, INHBA, CXCRI1 and CXCRI2       | No                                  |
| Zeng et al. 2014         | 1006 white children, age 3–14    | Primary decayed and filled teeth (dft) stratified to generate df-pit and fissures (dfPF) and df-smooth surface (dfSM) | KRTCAP2, ITGAL, PLUNC family genes           | Yes                                 |
| Haworth et al. 2018      | 19,003 Primary analysis; 13,353 Permanent analysis European ancestry meta-analysis age 2.5–18 | Presence or absence of treated or untreated caries | MPPED2, AJAP, and PRS6KA2 | No                                  |
| Shungin et al. 2019      | GLIDE and UKB (n = 26,792)       | DMFS, DFSS, Nteeth                                          | C5orf66, CA12, KRTCAP2, WNT10A, FGF10, HLA, FOXI1, PBX3, MAMSTR | Yes                                 |
The total sample for the primary dentition analysis included 1116 children (age 2–12 years) and 2570 adults for the permanent dentition analysis (ages 18–60 years). Table 2 summarizes the sex and cleft status breakdown for each analysis. Since our participants came from multiple sites, Table 3 summarizes the distribution of participants for the different recruitment sites.

**Dental caries assessment**

The same data collection protocols were used for every site. Dental caries scores \( (dft/DFT) \) were assessed by trained dentists or dental hygienists, with data collected either through in-person dental examinations and/or intraoral photos (at least 5 per participant, maxillary and mandibular occlusal, right and left lateral, anterior biting) to appropriately cover the entire oral cavity, as previously reported [11]. The \( dft/DFT \) index was calculated as the total number of teeth with decayed, and/or filled/restored surfaces. In addition, questionnaires recording dental history were collected for all participants in the study. We used the \( dft/DFT \) index instead of \( dmft/DMFT \) due to incomplete information regarding missing teeth for the majority of the participants. Note that for the children, primary dentition caries scores \( (dft) \) were obtained only from the primary teeth, any permanent teeth in these children were not included. Table 3 summarizes the dental caries mean scores for the different recruitment sites in analyses of both the primary and permanent dentitions.

All of the training and calibration for the intra-oral photos and the in-person dental exam was completed prior to the start of data collection. Each photo rater (BJH, LMU, and ARV) rated randomly selected participants \( (n=15) \) for calibration twice. Results from ratings by LMU and ARV were calibrated against the gold-standard rater, BJH (note BJH’s intra-rater reliability \( \kappa = 0.95 \)). The Inter-rater reliability \( (\kappa) \) between all the 3 raters \( = 0.91–0.93 \). Data from 158 participants who had both intra-oral photos and in-person dental exams were used to measure the consistency between in-person dental exams and the ratings from intra-oral photos; the results showed excellent agreement \( (\kappa > 90\%) \) between the two methods [11].

**Genotyping**

Participants were genotyped on the Illumina Human610-Quadv1_B BeadChip (Illumina, Inc., San Diego, CA, USA) and Illumina Infinium II assay protocol by the Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Data cleaning and quality assurance procedures were conducted in conjunction with the CIDR data cleaning center at the University of Washington. 529,285 SNPs (single nucleotide polymorphism) were genotyped, and additional SNPs were imputed based on the 1000 Genomes Project reference panel resulting in 35,153,445 SNPs [12]. Of those, 8,859,951 SNPs passed quality control and analysis filters (participant call rates > 90%; SNP call rates > 99%; Hardy–Weinberg \( p \) values > 0.0001; MAF > 2%) and were therefore analyzed in this study.

### Table 2  Basic characteristics of the study cohorts

| Male, n (%) | Female, n (%) | Age, range Mean ± SD | No cleft | With cleft | dft/DFT, range Mean ± SD | Total |
|-------------|---------------|----------------------|----------|------------|--------------------------|-------|
| Primary     |               |                      |          |            |                          |       |
| 619 (55.5%) | 497 (44.5%)   | 2–12 6.9 ± 2.6       | 687      | 429        | 0–20 2.6 ± 3.6           | 1116  |
| Permanent   |               |                      |          |            |                          |       |
| 1001 (38.9%)| 1569 (61.1%)  | 18–60 34.4 ± 9.9     | 2408     | 162        | 0–25 5 ± 4.3             | 2570  |

### Table 3  Distribution of participants across different sites

| Site                  | Primary | dft/DFT, range Mean ± SD | Permanent | DFT, range Mean ± SD |
|-----------------------|---------|--------------------------|-----------|----------------------|
| US sites              |         |                          |           |                      |
| Colorado              | 26      | 1.42 ± 1.55              | 32        | 8.12 ± 3.87          |
| Iowa                  | 225     | 0.65 ± 1.48              | 358       | 3.51 ± 3.56          |
| Pittsburgh            | 78      | 1.16 ± 2.54              | 147       | 7.48 ± 5.10          |
| Texas                 | 130     | 2.55 ± 3.21              | 196       | 4.74 ± 4.69          |
| Puerto Rico           | 55      | 2.09 ± 2.80              | 121       | 8.53 ± 4.06          |
| Total                 | 514 (46.1%) | 1.40 ± 2.47              | 854 (33.2%) | 5.36 ± 4.66          |
| Non-US sites          |         |                          |           |                      |
| Hungary               | 213     | 2.31 ± 2.94              | 565       | 6.45 ± 4.48          |
| Argentina             | 106     | 2.46 ± 3.09              | 246       | 4.32 ± 3.95          |
| Philippines           | 159     | 6.35 ± 4.86              | 563       | 4.20 ± 3.71          |
| Guatemala             | 124     | 3.76 ± 3.34              | 342       | 3.26 ± 3.67          |
| Total                 | 478 (53.9%) | 3.70 ± 4.00              | 1374 (66.8%) | 4.78 ± 4.19          |
| Total across all sites| 1116    | 2.6 ± 3.6                | 2570      | 5 ± 4.3              |
Statistical analysis
Genetic analyses
Caries scores in the primary dentition (dft) were analyzed separately from caries scores in the permanent dentition (DFT), both as quantitative traits. The analysis of dft in the primary dentition included participants of aged 2–12 years, and the analysis of DFT in the permanent dentition included participants aged 18–60 years. The statistical software used, namely, the Efficient Mixed-Model Association eXpedited (EMMAX) [13] uses a variance component-based method for testing the SNP main effects along with a genetic relationship matrix (GRM) and covariates (in this case, sex, recruitment site, age, age², and cleft status). Results of modeling of those covariates to determine if they are needed to be included to adjust our primary dft and permanent DFT analyses are shown in Table 4. Adjustment for those covariates were included in the final analyses. The genomic inflation factor (λ), and Manhattan plots were generated using the R v3.4.1 statistical analysis environment (R Core Team, 2019). Given the issue of multiple comparisons, conservative p-value thresholds for genome-wide statistical significance were set to $P < 5 	imes 10^{-8}$. The p-value thresholds for suggestive significance were set equal to $5 	imes 10^{-8} < P < 1 	imes 10^{-6}$.

Bioinformatic approaches for the genetic results
Different bioinformatic and visualization tools were used in this study to interpret the GWAS results at both the SNP and gene level. Loci/genes of interest were identified and reported based on physical proximity of ±500 kb windows from the lead SNP at each of the loci associated with dental caries. We used LocusZoom [14], to visualize the associations results in the regions around the lead SNPs at each locus. To gain more information regarding the biological consequences of the lead SNPs in the risk loci, variant annotation tools, such as HaploReg, were used to annotate regulatory information, such as enhancer/promoter regions, expression quantitative trait loci (eQTLs), and transcription factor binding sites [15]. Other regulatory databases, such as the Genotype-Tissue Expression (GTEx) [16], and Encyclopedia of DNA Elements [17] were used, too, to help in interpreting the associations results.

Results
Cohort characteristics and covariate modeling
The total study sample comprised 3686 genotyped and phenotyped individuals. The primary dentition cohort (n = 1116) included 616 males (55.5%), and 497 females (44.5%) with an age range of 2–12 years and a mean age of 6.9 years (Table 2), 46.1% of the participants were from U.S recruitment sites, and 53.1% from international sites. The permanent dentition cohort (n = 2570) included 1001 males (38.9%), and 1569 females (61.1%) with an age range of 18–60 years and a mean age of 34.4 years (Table 2), with 33.2% of the participants from U.S recruitment sites, and 66.8% of participants from international sites. General linear regression analysis (details in Table 4) showed that age, recruitment site, cleft status (affected vs. unaffected with cleft) and sex had an impact on the DFT and dft mean scores; therefore, these factors were included as covariates in the genetic analyses.

Genetic results
Manhattan plots illustrating the GWAS results for the primary dft and permanent DFT are shown in Fig. 1a, c. The genomic inflation factor (λ) for the primary dentition analysis was 0.98, and for the permanent dentition analysis was 0.99, indicating no inflation of p values (Fig. 1b, d). Note that associations identified under the GWAS approach do not specify which SNP at that locus is the causal SNP (cause the association) or which gene is affected by the causal SNP. Here we report any known biological functions of genes near our significant and suggestive results that have plausible roles in dental caries or have some biological relevance to tooth development and/or oral health.

Primary dentition
For the primary dft, SNPs at multiple loci exceeded the threshold for genome-wide significance ($P < 5 	imes 10^{-8}$) and others exhibited suggestive significance ($5 	imes 10^{-8} < P < 1 	imes 10^{-6}$). The biological roles of potential candidate genes at these loci are described in Table 5. Genes at two of the genome-wide significant loci have no known role in cariogenesis and warrant further investigation. One of the lead SNPs (rs113021760, $P = 3.36 	imes 10^{-8}$), is located in the intron of the SUSD1 gene (Sushi Domain Containing 1), so it might play a

---

Table 4 Regression results for the covariates

| Phenotype      | Covariate | Beta    | SE     | P         |
|---------------|-----------|---------|--------|-----------|
| Primary dft   | Age       | 1.69641 | 0.19039| <2e−16 ***|
|               | Age²      | -0.12954| 0.01357| <2e−16 ***|
|               | Sex       | -0.10053| 0.19735| 0.611     |
|               | Site      | 2.27537 | 0.19744| <2e−16 ***|
|               | Cleft status | 0.80861| 0.20120| 6.24e−05 ***|
| Permanent DFT | Age       | 0.3832854| 0.0531329| 7.14e−13 ***|
|               | Age²      | -0.0039532| 0.0007238| 5.17e−08 ***|
|               | Sex       | 0.6350039| 0.1723444| 0.000234 ***|
|               | Site      | 0.3680415| 0.1865084| 0.048566*  |
|               | Cleft status | 1.4312329| 0.3496957| 4.39e−05 ***|

Significant P values codes: 0.000*** 0.001** 0.01* 0.05
regulatory role. In addition, this SNP shows enhancer chromatin marks in different tissues, including osteoblasts, which are found in the periodontal ligament and alveolar bone in the oral cavity.

One of the strongest suggestive association signals for the primary dft was at 17q21.33 (rs16948495, \(P = 5.16 \times 10^{-8}\)) which is less than 100 kb downstream from DLX3 and upstream from DLX4 (see Fig. 2a).

These genes belong to the Distal-less homeobox family, known to be expressed in cranial neural crest cells and in the craniofacial mesenchyme; and important in tooth development [18–20]. DLX3, which plays an essential role in skeletal formation and development, is also required for tooth morphogenesis [21, 22]. Furthermore, a recent study also showed that DLX3 plays an important role in dentinogenesis. Mutations in this gene cause...
tricho-dento-osseous (TDO) syndrome, a rare syndrome that affects the teeth and bones, including enamel hypoplasia and dentin hypoplasia [23]. These severe outcomes of DLX3 mutations emphasize the important role DLX3 plays in the development of bones and teeth.

DLX4 is highly expressed in human dental pulp cells (DPCs) and is associated with abnormal human tooth
formation [24]. DLX4 also plays an important role in forebrain and craniofacial development. A mutation in DLX4 was reported to cause a non-syndromic form of cleft lip and palate [25]. Rs16948495 shows enhancer chromatin marks in osteoblasts. These lines of evidence suggest that both DLX3 and DLX4 may play important roles in the dental caries process by the regulation tooth development [21–23, 25].

Fig. 2. LocusZoom of regions of interest. Panels a, b and c are association results from the primary dentition GWAS, panels d and e are from the permanent dentition GWAS. a suggestive locus near DLX3 and DLX4 genes on chromosome 17. b suggestive locus near TLN1 and CA9 on chromosome 9. c suggestive locus near NFX1 gene on chromosome 9. d suggestive locus near REL gene on chromosome 2. e suggestive locus near Taste receptor genes (TAS2R38, TAS2R3, TAS2R4, TAS2R5) on chromosome 7. The genome build used for the recombination rate was based on 1000 Genomes November 2014 EUR data. All of the gene positions and directions of transcription are annotated on the plots.
There was also a strong suggestive association signal for primary dft and a SNP at 9p13.3 (chr9:35,753,170, \( P = 1.89 \times 10^{-7} \), Fig. 2b). The gene CA9 (CA IX) is located less than 200 kb upstream from the lead SNP at this locus: CA9 encodes a family of enzymes with at least 14 different isoforms that participate in several biological processes including the secretion of saliva and pH homeostasis [26]. In candidate gene studies, a different member of the CA family (CA VI) was investigated due to its role in saliva secretion and regulating salivary pH and was found to be significantly associated with dental caries in children [27–29]. When we looked-up CA6 gene in our association results, we did find a SNP (rs58579969), 300 kb downstream from it. However, this SNP only showed nominal evidence of an association with primary dft (\( P = 1.2 \times 10^{-5} \)).

Another gene at the 9p13.3 locus is TLN1; a cytoskeletal protein that has been implicated in several biological functions, including migration of fibroblasts and osteoclasts [30, 31]. The expression of tln1 in zebrafish was observed in the craniofacial cartilage structures, including the palate. Tln1 mutant zebrafish had malformation in their palate and craniofacial muscles, which indicate the possible role of tln1 in craniofacial morphogenesis [32]. Neither TLN1 nor CA9 genes have been implicated in dental caries before, however, their plausible roles in craniofacial morphogenesis and pH regulation could suggest that they may influence cariogenesis.

**Permanent dentition**

In the permanent dentition analysis, there were no SNPs meeting the genome-wide significance level, but there were results with multiple loci that were within the range for the genome-wide suggestive level (\( 5 \times 10^{-8} < P < 1 \times 10^{-6} \)). Summaries of the lead SNPs and nearby genes from the permanent analysis are presented in Table 5.

One of the strongest suggestive association signals for the permanent DFT was at the 9p13.3 locus (rs11197981; \( P = 1.11 \times 10^{-6} \)), approximately 200 kb downstream of the lead SNP at 7q34 (lead SNP rs11686767, \( P = 3.11 \times 10^{-8} \)) which is located within APTAX (Fig. 2c), and the encoded protein of this gene is involved in the repair of DNA damage in cells [33]; however, no role is currently known for this gene in odontogenesis or dental caries. The lead SNP is a strong eQTL (expression quantitative trait locus) for APTAX (per an eQTL database [34]). It is also a strong eQTL for other genes (BAGALT1, SMUI, RP11-54K16.2) [16]. The lead SNP is also approximately 300 kb upstream from the NFX1 gene, which plays an important role in regulating the duration of inflammatory responses [35].

Another suggestive association signal at 2p16.1 (rs11686767, \( P = 3.11 \times 10^{-8} \)), is intronic to PAPOLG (Fig. 2d) which plays a role in catalyzing template-independent extension of a DNA/RNA strand [36]. The SNP is a strong eQTL for PAPOLG [16], however, no known role yet of this gene in odontogenesis or dental caries. Finally, rs11686767 was also 100 kb upstream to REL gene. The protein encoded by this gene, proto-oncogene c-Rel, is involved in many important cellular processes such as apoptosis, inflammation, and the immune response [37]. Those genes; NFX1 and REL could contribute to dental caries by influencing host susceptibility to oral microorganism.

An interesting suggestive association for the permanent dentition DFT was at 7q34 (lead SNP rs11197981; \( P = 1.11 \times 10^{-6} \)), approximately 200 kb downstream of taste receptor genes (TAS2R3, TAS2R3, TAS2R4, TAS2R25) (Fig. 2e), which have plausible roles in dental caries. Specifically, TAS2R38 has been associated with dental caries in previous studies [38–41]. Further, taste receptor genes, specifically TAS2R38, influence bitter perception and dietary habits that are hypothesized to lead to consuming more sweets in the diet and thus leading to dental caries. The lead SNP at this locus is also within the OR9A4 gene, which is responsible for activating the neural response that initiates the perception of smell [42].

To further investigate these significant or suggestive SNPs, we used the FUMA platform (Functional Mapping and Annotation of Genome-Wide Association Studies) [43] to prioritize, annotate, and interpret the genomic variants and genes from the GWAS results of both primary and permanent analyses. Gene-based testing using GWAS results was computed by MAGMA (Multi-marker Analysis of GenoMic Annotation) using the default settings implemented in FUMA. After mapping the top SNPs from the GWAS analyses to 19,182 protein coding genes, the genome-wide significance level was calculated based on the number of tested genes and it was set at \( 0.05/19182 = 2.61 \times 10^{-6} \). However, none of the genes in both analyses reached the genome wide significance level. FUMA results of the risk genomic loci for both analyses are presented in the Additional file 1: Tables S1 and S2.

**Discussion**

Dental caries is one of the most common chronic diseases that affects both children and adults, and is influenced by a complex interplay of diet, bacteria, salivary flow, genetic factors and other environmental factors. Untreated dental caries could lead to tooth loss which can negatively impact the oral health quality of life (OHRQoL) because it affects function, aesthetics and satisfaction. OHRQoL is a multidimensional concept that includes the impact of dental health on the individual’s functional well-being, emotional health, satisfaction, and self-esteem. OHRQoL would...
is very important because of its usefulness in addressing the impact of oral health disparities and access to care on overall health and quality of life to communicate it with policymakers to improve oral health by increasing access to care [44, 45].

There are now a number of effective preventive methods for dental caries (good oral hygiene, balanced diet, regular dental care) [46, 47]. Most notably, stannous fluoride, such as in toothpaste and water, has been an important public health advance by reducing the prevalence of dental caries [48]. However, dental caries is still wide spread worldwide [3], requiring additional approaches to understanding risk factors for dental caries.

The current study took a genome-wide approach to investigate genetic factors contributing to dental caries in the primary and permanent dentitions. Several previous genetic studies have investigated and identified the role of different genes in cariogenesis. Most of these past studies involved samples that were ethnically homogeneous, mainly of European descent [49–53], unlike the current multiethnic study (see also Table 1).

Notably, results in the primary dentition nominated a few SNPs that showed association at genome-wide significance, and multiple loci that showed suggestive evidence for association. In the permanent dentition, no variants reached genome-wide significance, but several reached suggestive significances. Importantly, none of the top associated SNPs in either of the analyses (dft and DFT) were replicated in the other one, full results are presented in the Additional file 1: Tables S3 and S4. This confirms that genetic loci that influence the susceptibility to dental caries differ between primary and permanent dentitions [54]. Although none of the genome-wide significant association signals were near genes with known roles in dental caries or tooth development, some of the suggestive signals were near genes with potential roles in dental caries.

Some of the new associations identified in this study were in or near genes that have important roles in the inflammatory process, such as NFX1 and REL genes (permanent dentition). Those genes could contribute to dental caries by influencing host susceptibility to oral microorganism. This study also identified association signals near genes that have a confirmed role in tooth and craniofacial development, such as TLN1 and DLX genes. Heterozygous variants in DLX4 have been implicated before in Orofacial clefting [25]. This could indicate that both dental caries and OFC share common genetic risk loci. This is plausible because some of the genes that had been previously implicated in dental caries and OFCs are craniofacial development genes. PKD2 gene is one of the genes that has been associated with dental caries and had an implication in craniofacial development [8] and mutation in PKD2 has been linked to craniofacial anomalies and tooth loose in mice [55]. Thus, our future investigations will focus on exploring the relationship between dental caries and OFCs on both phenotypic and genetic levels.

This study also supported the role of taste receptor genes in dental caries (TAS2R38, TAS2R3, TAS2R4, TASR25, see Table 5), which have been associated with dental caries in previous studies [38–41]. On the other hand, several loci/genes associated with dental caries from previous large meta-GWAS studies of dental caries [50, 53] were reviewed in our results (the lead SNPs ± 500 kb, a total of 5935 SNPs), but none of these variants reached statistical significance (see Additional file 1: Tables S5, S6 for details).

In the permanent dentition subset of the current study population, females had a significantly higher rate of dental caries than males (P = 0.000234) consistent with earlier studies [56, 57]. However, there was no gender difference in the primary dentition. Possible explanations for the increased rates of dental caries among adult females include (i) the fact that teeth tend to erupt earlier in females, which allows for a longer exposure to a cariogenic oral environment, and (ii) hormonal changes during pregnancy affect saliva pH leading to a decrease in the concentration of calcium, phosphorus, magnesium, and chloride, which, in turn might lead to increased caries. (iii) other behavioral/environmental factors during pregnancy such as consuming a diet high in sucrose and poor oral hygiene could also increase the risk of developing more dental caries.

**Conclusion**

In summary, this study found multiple associations near genes that may play important roles in dental caries, such as TAS2R38, TAS2R3, TAS2R4, TASR25, DLX3 and DLX4, several of which have plausible biological functions relevant to tooth development and cariogenesis. These findings contribute to our understanding of the genetic mechanisms of cariogenesis, providing targets for follow-up translational studies that will eventually improve prevention and treatment of this highly prevalent chronic disease worldwide.

**Abbreviations**

dft: Number of decayed or filled primary tooth; DFT: Number of decayed or filled permanent teeth; GWAS: Genome-wide association study; MAF: Minor allele frequency; SNP: Single nucleotide polymorphism.
Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12903-021-01670-5.

Additional file 1. Supplementary material.

Acknowledgements

Many thanks to the participating families world-wide who made this project possible. The tireless efforts by the dedicated field staff and collaborators was similarly essential to the success of this study. Special thanks to Margaret Cooper, Judith Resick, Dr. Eduardo Castilla and Dr. Andrew Czeese.

Authors’ contributions

RNA generated the data set for the analyses. RNA, JMC analyzed data. RNA, MLM, JMC, JRS, SMW helped define outcomes to be studied. RNA wrote the first draft of the manuscript. MLM, JRS, SMW, B.J.H, L.M.U, N.M, and I.N critically reviewed the manuscript. RNA & MLM designed the study. RNA interpreted data. MLM, CS, FWBD, KN, CP, FAP, IMO, CJR, JTH, GLW, SMW, ARV, L.M.U, REL collected and interpreted data. RNA & MLM generated the final draft of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Institutes of Health (NIH) including: R01-DE016148 [MLM, SMW], X01-HG007485 [MLM, EF], U01-DE024425 [MLM], R37-DE008559 [JCM, MLM], R21-DE016930 [MLM], R01-DE014727 [MLM], R01-DE011931 [JTH], U01-DD000295 [GLW], R00-DE024571 [CJR], S21-MD001830 [CJR], US4-MD007587 [CJR], R56-DE027055 [JRS]. Genotyping and data cleaning were provided via an NIH/NIDCR contract to the Johns Hopkins Center for Inherited Disease Research: HHSN268201000081. Additional support provided by an intramural grant from the Research Institute of the Children’s Hospital of Colorado [FWD], additional operating costs support in the Philippines was provided by the Institute of Human Genetics, National Institutes of Health, University of the Philippines, Manila [CP].

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the dbGaP repository, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000774.v2.p1

Declarations

Ethics approval and consent to participate

Study approval was granted by the University of Pittsburgh Institutional Review Board (coordinating center approval # CR19030367-003, Pittsburgh site approval # CR19080127-003). All participant provided written informed consent for themselves and for children younger than 18 years, informed consent was obtained from their parents or their legally authorized representative. Local ethical approval was obtained at each site and all methods in this study were performed in accordance with the Institutional Review Board policies and guidelines of the University of Pittsburgh and all of the other sites.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

1Department of Oral and Craniofacial Sciences, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA. 2Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA. 3Dental Health Department, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia. 4Department of Family Dentistry, College of Dentistry, University of Iowa, Iowa City, IA, USA. 5The Iowa Center for Oral Health Research, College of Dentistry, University of Iowa, Iowa City, IA, USA. 6Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. 7UC Health Medical Group, Colorado Springs, Denver, CO, USA. 8Department of Pediatrics, College of Medicine, University of the Philippines, Manila, Philippines. 9ECLAMC/INAGEMP At Center for Medical Education and Clinical Research (CEMIC-CINCEG), Buenos Aires, Argentina. 10Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. 11Dental and Craniofacial Genomics Core, School of Dental Medicine, University of Puerto Rico, San Juan, PR, USA. 12Department of Pediatrics, University of Texas Health Science Center At Houston, Houston, TX, USA. 13Department of Health Management and Policy, College of Public Health, University of Iowa, Iowa City, IA, USA. 14Lancaster Cleft Palate Clinic, Lancaster, PA, USA. 15Department of Orthodontics, School of Dentistry, University of Iowa, Iowa City, IA, USA.

Received: 30 March 2021 Accepted: 30 May 2021

Published online: 26 July 2021

References

1. Leal SC, Bronkhorst EM, Fan M, Frencken JE. Untreated cavitated dentine lesions: impact on children’s quality of life. Caries Res. 2012;46(2):102–6.

2. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1789–838.

3. Maricene W, Kassebaum NJ, Bernabe E, Flaxman A, Naghavi M, Lopez A, Murray CJ. Global burden of oral conditions in 1990–2010: a systematic analysis. J Dent Res. 2013;92(7):592–7.

4. Ahluwalia M, Braillford SP, Tarelli E, Gilbert SC, Clark DT, Barnard K, Beighton D. Dental caries, oral hygiene, and oral clearance in children with craniofacial disorders. J Dent Res. 2004;83:175–9.

5. Shaffer JR, Wang X, DeSensi RS, Wendell S, Weyant RJ, Cuenco KT, Crouch RM, Marazita ML. Genetic susceptibility to dental caries on pit and fissure and smooth surfaces. Caries Res. 2012;46:38–46.

6. Martin B, Willkerston AH, Patterson G, Nahar VK, Sharma M. Need for addressing oral health disparities in rural Appalachia. Health Promot Perspect. 2017;7(4):178–80.

7. Costa SM, Martins CC, Bonfim Mde L, Zina LG, Paiva SM, Pordeus IA, Abreu MH. A systematic review of socioeconomic indicators and dental caries in adults. Int J Environ Res Public Health. 2012;9(10):3540–7.

8. Shaffer JR, Feingold E, Wang X, Lee M, Tuencko K, Weeks DE, Weyant RJ, Crouch RM, McNeil DW, Marazita ML. GWAS of dental caries patterns in the permanent dentition. J Dent Res. 2013;92(1):38–44.

9. Shaffer JR, Polk DE, Feingold E, Wang X, Cuenco KT, Weeks DE, DeSensi RS, Weyant RJ, Crouch RM, McNeil DW, et al. Demographic, socioeconomic, and behavioral factors affecting patterns of tooth decay in the permanent dentition: principal components and factor analyses. Community Dent Oral Epidemiol. 2013;41(4):664–73.

10. Shaffer JR, Wang X, Feingold E, Lee M, Begum F, Weeks DE, Cuenco KT, Barnada MM, Wendell SK, Crosslin DR, et al. Genome-wide association scan for childhood caries implicates novel genes. J Dent Res. 2011;90:1457–62.

11. Howe BJ, Cooper ME, Wehby GL, Resick JM, Nidey NL, Valencia-Ramirez LC, Lopez-Palacio AM, Rivera D, Viera AR, Weinberg SM, et al. Dental decay phenotype in nonsyndromic orofacial clefting. J Dent Res. 2017;96:1106–14.

12. Leslie EJ, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, Jain D, Laurie CC, Doherty KF, McHenry T, et al. A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p22.2, 17q23 and 19q13. Hum Mol Genet. 2016;25(13):2862–72.

13. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods. 2014;11:407–9.

14. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336–7.

15. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):D950–934.
16. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):646–60.

17. Consortium EP . An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.

18. Chai Y, Jiang X, Ito Y, Bringas P Jr, Han J, Rowitch DH, Soriano P, McMahon AP, Suvoc HM. Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. Development. 2000;127(8):1671–9.

19. Duverger O, Zai A, Isaac J, Sun HW, Bartels AK, Jian J, Bendal A, Hwang J, Morasso MI. Neural crest deletion of Dlx3 leads to major dentin defects through down-regulation of Dspp. J Biol Chem. 2012;287(15):12230–40.

20. Zhao Z, Stock D, Buchanan A, Weiss K. Expression of Dlx genes during the development of the murine dentition. Develop Genes Evol. 2000;210(5):270–5.

21. Isaac J, Erthal J, Gordon J, Duverger O, Sun HW, Lichtler AC, Stein GS, Gian J, Morasso MI. Dlx3 regulates bone mass by targeting genes supporting osteoblast differentiation and mineral homeostasis in vivo. Cell Death Differ. 2014;21(9):1365–76.

22. Li Y, Han D, Zhang H, Liu H, Wongs N, Zhao N, Qiu L, Feng H. Morphological analyses and a novel de novo Dlx3 mutation associated with tricho-dento-osseous syndrome in a Chinese family. Eur J Oral Sci. 2015;123(4):228–34.

23. Tamaoki N, Takakahi K, Aoki H, Iida K, Kawauchi T, Hatakeyama D, Inden M, Chosa N, Ishiaki A, Kunisada T, et al. The homeobox gene Dlx4 promotes generation of human induced pluripotent stem cells. Sci Rep. 2014;4:7283–7283.

24. Wu D, Mandal S, Choi A, Anderson A, Prochazkova M, Perry H, Gil-Da-Silva-Lopes VL, Lao R, Wang E, Tsal T, et al. DLX4 is associated with orofacial clefting and abnormal jaw development. Hum Mol Genet. 2015;24(15):4340–52.

25. Pastorekova S, Parkiloka S, Pastorek J, Supuran CT. Review article. J Enzyme Inhib Med Chem. 2004;19(3):199–229.

26. Frasseto F, Parisotto TM, Peres RC, Marques MR, Line SRP, Parisotto TM. Nobre Dos Santos M: Children with a higher activity of carbonic anhydrase 6 gene with salivary buffer capacity, dental plaque pH, and flow rate, biofilm pH and caries in primary dentition. Caries Res. 2012;46(3):194–200.

27. Peres RC, Camargo G, Mofatto LS, Cortellazzi KL, Santos MC, Nobre-dos Santos M, Bergamaschi CC, Line SR. Association of polymorphisms in the carbonic anhydrase 6 gene with saliva buffer capacity, dental plaque pH, and caries index in children aged 7–9 years. Pharmacogenomics. J. 2010;11(2):114–9.

28. Picco DCR, Lopes LM, Rocha Marques M, Line SR. Nobre Dos Santos M: Craniofacial anomalies are characterized by conditional loss of polycystic kidney disease 1 gene. J Biochem. 2001;213(16):5614–23.

29. Yamashita S, Fujii K, Zhao C, Takagi H, Katakura Y. Involvement of the protein. Ann Neurol. 2004;55(2):241–9.

30. Wong K, Shaffer JR, Fong Z, Begum F, Vieira AR, Noel J, Anjomshoaa I, Cuenco KT, Lee MK, Beck J, et al. Genome-wide association Scan of dental caries management. Stomatologija. 2015;17(4):107–12.

31. Wang X, Shaffer JR, Zeng Z, Begum F, Vieira AR, Noel J, Anjomshoaa I. Biological approach of dental caries management. Stomatologija. 2015;17(4):107–12.

32. Pastorekova S, Parkiloka S, Pastorek J, Supuran CT. Review article. J Enzyme Inhib Med Chem. 2004;19(3):199–229.

33. Wu D, Mandal S, Choi A, Anderson A, Prochazkova M, Perry H, Gil-Da-Silva-Lopes VL, Lao R, Wang E, Tsal T, et al. DLX4 is associated with orofacial clefting and abnormal jaw development. Hum Mol Genet. 2015;24(15):4340–52.

34. Pastorekova S, Parkiloka S, Pastorek J, Supuran CT. Review article. J Enzyme Inhib Med Chem. 2004;19(3):199–229.

35. Rhedoro H, Oral health-related quality of life—what, why, how, and future implications. J Dent Res. 2011;90(11):1264–70.

36. Grigalaukienė R, Slabinskiene E, Vasiliauskienė I. Biological approach of dental caries management. Stomatologija. 2015;17(4):107–12.

37. Wendell S, Wang X, Brown M, Cooper ME, DeSenss RI, Cruot P, McNeil DW, Marazita ML. Taste genes associated with dental caries. J Dent Res. 2010;89(11):1198–202.

38. Yildiz G, Ermis RB, Calapoglu NS, Celik EU, Turel GY. Gene-environment interactions in the etiology of dental caries. J Dent Res. 2016;95(1):74–9.

39. Watanabe K, Taskesen E, van Bochoven A, Posthumus D. Functional mapping and annotation of genetic associations with FLNA. Nat Commun. 2017;8(1):1826.

40. Fiorillo L. Oral health: the first step to well-being. Medicina (Kaunas). 2019;55(10):679.

41. Sischo L, Broder HL. Oral health-related quality of life: what, why, how, and future implications. J Dent Res. 2011;90(11):1264–70.

42. Grigalaukienė R, Slabinskiene E, Vasiliauskienė I. Biological approach of dental caries management. Stomatologija. 2015;17(4):107–12.

43. Wranne B, Shingun L, van der Tas JT, Lucic V, Medina-Gomez C, Yakimov V, Feenstra B, Shaffer JR, Lee MK, Standall M, et al. Consortium-based genome-wide meta-analysis for childhood dental caries traits. Human Mol Genet. 2018;27(17):3113–27.

44. Gassé B, Grabar S, Lafont AG, Quinquins I, Opsahl Vital S, Davit-Béal T, Mousili E, Chabadel O, Hennequin M, Courson F, et al. Common SNPs of AmelogeninX (AMELX) and dental caries susceptibility. J Dent Res. 2013;92(5):418–24.

45. Haword H, Shungun L, van der Tas JT, Lucic V, Medina-Gomez C, Yakimov V, Feenstra B, Shaffer JR, Lee MK, Standall M, et al. Consortium-based genome-wide meta-analysis for childhood dental caries traits. Human Mol Genet. 2018;27(17):3113–27.

46. Grigalaukienė R, Slabinskiene E, Vasiliauskienė I. Biological approach of dental caries management. Stomatologija. 2015;17(4):107–12.

47. Kastovsky J, Bonilova Linhartova P, Musilova K, Lukova L, Kukla L, Izakovicova Holla L. Lack of association between BMP2/DLX3 gene polymorphisms and dental caries in primary and permanent dentitions. Caries Res. 2015;49(4):417–24.

48. Kastovsky J, Bonilova Linhartova P, Musilova K, Zackova L, Lukova K, Kukla L, Izakovicova Holla L. Lack of association between BMP2/DLX3 gene polymorphisms and dental caries in primary and permanent dentitions. Caries Res. 2017;51(6):590–5.

49. Shungun L, Haword H, Divaris K, Agler CS, Kamatani Y, Keun Lee M, Grinde K, Hindy G, Aaladunjakoi V, Pesonen P, et al. Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data. Nat Commun. 2019;10(1):2773–2773.

50. Wang X, Shaffer JR, Zeng Z, Begum F, Vieira AR, Noel J, Anjomshoaa I, Cuenco KT, Lee MK, Beck J, et al. Genome-wide association Scan of dental caries in the permanent dentition. BMC Oral Health. 2012. https://doi.org/10.1186/1472-6831-12-57.

51. Pkontes M, Dhont C. Natural language processing to navigate dental caries research literature: a systematic review. PLoS One. 2015;10(2):118661.

52. Pkontes M, Dhont C. Natural language processing to navigate dental caries research literature: a systematic review. PLoS One. 2015;10(2):118661.

53. Pkontes M, Dhont C. Natural language processing to navigate dental caries research literature: a systematic review. PLoS One. 2015;10(2):118661.

54. Pkontes M, Dhont C. Natural language processing to navigate dental caries research literature: a systematic review. PLoS One. 2015;10(2):118661.

55. Pkontes M, Dhont C. Natural language processing to navigate dental caries research literature: a systematic review. PLoS One. 2015;10(2):118661.

56. Pkontes M, Dhont C. Natural language processing to navigate dental caries research literature: a systematic review. PLoS One. 2015;10(2):118661.

57. Pkontes M, Dhont C. Natural language processing to navigate dental caries research literature: a systematic review. PLoS One. 2015;10(2):118661.