Investigating the genetics of dental caries incidence, development over time, and variability

by

Tianyu Zou

BM, Sun Yat-Sen University, 2015

MS, Peking Union Medical College, 2018

Submitted to the Graduate Faculty of the

School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Tianyu Zou

It was defended on

May 27, 2022

and approved by

Ryan L. Minster, PhD, MSIS, Assistant Professor,
Department of Human Genetics, Graduate School of Public Health,
University of Pittsburgh

Seth M. Weinberg, PhD, Associate Professor,
Department of Oral and Craniofacial Sciences, School of Dental Medicine,
Department of Human Genetics, Graduate School of Public Health,
University of Pittsburgh

Mary L. Marazita, PhD, Professor,
Department of Human Genetics, Graduate School of Public Health,
Department of Oral and Craniofacial Sciences, School of Dental Medicine,
Clinical and Translational Sciences, School of Medicine,
University of Pittsburgh

Dissertation Director: John R. Shaffer, PhD, Assistant Professor,
Department of Human Genetics, Graduate School of Public Health,
Department of Oral and Craniofacial Sciences, School of Dental Medicine,
University of Pittsburg

Investigating the genetics of dental caries incidence, development over time, and variability

Tianyu Zou, PhD

University of Pittsburgh, 2022

Dental caries is the most common chronic disease of childhood, worldwide. A better understanding of the etiology of dental caries is needed to improve caries prevention and oral health at both individual and population levels. This study performed GWASs using phenotypes including time-to-first caries incidence and repeated caries measurements at multiple time points, and conducted genome-wide scans of heteroscedasticity in dental caries experience followed by targeted gene-by-environment interaction (GEI) modeling.

Genome-wide survival analysis in Aim 1 showed the genetic underpinnings of caries incidence, with heritability (i.e., the proportion of variance explained by genetics) of time-to-first caries estimated to be 54.4% and 14 association signals identified at the suggestive level (P<1E-5), some of which were located near genes with potential roles in dental caries, including *COL5A1, ASIC2,* and *ESR1*. GWAS of repeated dental caries measurements in Aim 2 was the first longitudinal GWAS of caries that demonstrated the genetic influence on caries development over time. The heritability estimate of longitudinal caries trajectories was 54.7%, and there were 1 genome-wide (P<5E-8) and 12 suggestive (P<1E-6) associations, some of which were located near genes with potential functions in caries including *SOD2, WNK1, CTSD, WWP2,* and *BTF3*. Genome-wide scans for variance quantitative trait loci (vQTL) – i.e., genetic variants associated with the variance rather than mean of the trait – in Aim 3 identified a total of 39 independent vQTLs, some of which were located in or near genes with potential functions in dental caries (e.g., *IGFBP7, SLC5A8,* and *SHH* involved in tooth development and enamel mineralization). In the

GEI analysis using the prioritized vQTLs and self-reported environmental factors, we found that children with certain genotypes of prioritized variants exhibited higher caries experience if they had lower parental educational attainment, had lower household/parental income, brushed their teeth less frequently, consumed sugar-sweetened beverages more frequently, were not breastfed, and were female.

This study has broadened our knowledge of the genetic architecture of caries onset and development over time, as well as GEI, and may provide the foundation for better early detection, risk assessment, dental care, and effective public health interventions.

Table of Contents

List of Tables

List of Figures

Preface

I would like to express my thanks to my supervisor, Dr. John Shaffer. Dr. Shaffer gave me the precious opportunity to work in his lab and has guided me throughout the years of my training with a great amount of patience and brilliance. His scientific insights and experience have led me the way to scientific thinking and helped me in shaping all my research projects. He is available whenever I have questions about my projects, and he always provides the clearest explanations and the best suggestions. I would certainly not be possible to complete my PhD degree without his support. Thanks for being the best mentor that I have ever had.

I would like to express my gratitude to Dr. Mary Marazita, Dr. Seth Weinberg and Dr. Ryan Minster, my dissertation committee. Dr. Marazita has a profound experience in the genetic studies of oral health, which has deepened my projects with additional aspects of analyses. Dr. Weinberg has an extensive knowledge in genetics of human craniofacial anatomy and has provided plenty of fresh perspectives to improve my research. Dr. Minster is an excellent expert in statistical genetics and has given me practical advice to make sure that my work is solid. Their patient and insightful guidance have helped me enormously throughout my dissertation projects and given me the confidence to pursue my career goals in the future.

I would also like to thank everyone at CCDG, an amazing research center that gives me everything I need to complete my research. I am grateful to my colleagues in the Shaffer lab and my fellow schoolmates for their warm support and friendship all these years.

I am incredibly lucky to work with so many lovely and brilliant people and finish my doctoral degree in HUGEN at Pitt. It is an awesome journey for me.

1.0 Overall Research Goal and Specific Aims

Dental caries is a multifactorial disease presenting a global health challenge. Genetic studies have demonstrated a strong genetic component to dental caries, with genetic factors explaining 20%-70% of variation in the dental caries traits (Bretz et al., 2006; Bretz, Corby, Schork, et al., 2005; Haworth et al., 2020; Shaffer, Feingold, et al., 2012; Shaffer, Wang, et al., 2012; X. Wang et al., 2010). Genome-wide association studies (GWASs) of caries have also identified several association signals located near genes with plausible biological roles in dental caries. However, most GWASs have been performed cross-sectionally with phenotypes including binary traits (e.g., presence or absence of caries) and/or indices of **d**ecayed, **m**issing, and **f**illed **t**eeth or **s**urfaces (DMFT/DMFS representing the permanent dentition and dmft/dmfs representing the primary dentition) (Haworth et al., 2018; Shaffer et al., 2011; X. Wang et al., 2012; Zeng et al., 2014; Zeng et al., 2013); these phenotypes may not have fully captured the genetic effects on all aspects of cariogenesis; indeed, study designs investigating the genetics of dental caries incidence, development of caries over time, and dental caries variability (variance) may capture previously under-studied aspects of the underlying genetic architecture of caries. Furthermore, genetic effects on dental caries may differ between individuals depending on environmental factors due to gene-environment interactions (GEI). Few studies have explored GEI for dental caries. Filling these gaps in the current knowledge may provide the foundation for better early detection, risk assessment, dental care, and effective public health interventions.

The overall goal of this study is to improve the genetic understanding of dental caries. Specifically, this study explored the genetic basis of caries using three approaches, including time-to-event analyses of dental caries incidence, longitudinal analyses of dental caries development,

and analyses dental caries variability (variance), by leveraging multiple advances in statistical modeling in order to pursue the following specific aims:

Aim 1. Identify genetic variants associated with caries incidence in the primary dentition using genome-wide survival analysis.

Aim 2. Identify genetic variants associated with longitudinal caries development using repeated caries measurements at multiple time points as phenotype.

Aim 3. Identify genetic variants associated with caries variability in the primary dentition by performing genome-wide scan of vQTLs and examine GEI between prioritized variants, putative environmental risk factors, and dental caries.

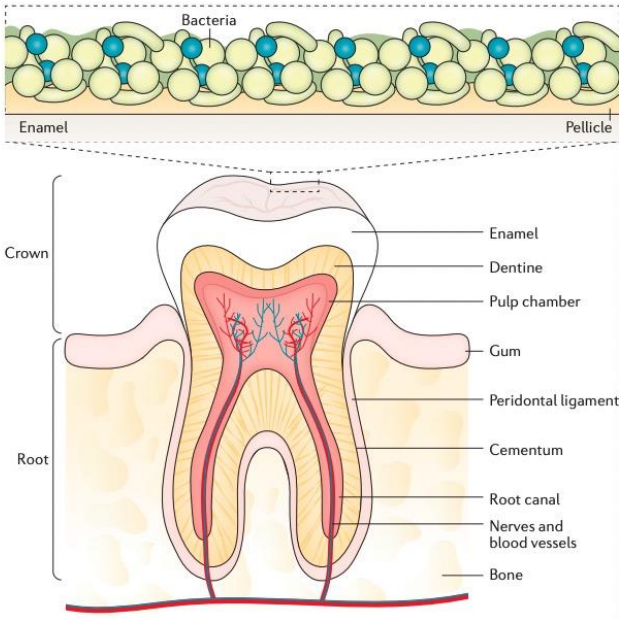
This study used caries incidence and repeated caries measurements that were generated from longitudinal cohorts in genome-wide association analyses, which broadened the knowledge of the genetic architecture of dental caries onset and development over time. Moreover, this study explored the genetic effects on caries variability and the underlying GEI effects of the identified vQTLs, which provided an alternative method to assess GEI in dental caries.

2.0 Introduction

2.1 Dental caries

Dental caries is the damage to a tooth that can be caused by acids that are generated by cariogenic oral bacteria. It is a chronic disease that can affect the hard tissue including enamel, cementum, and dentine (Figure 1). Dental caries can be detected in both the crown and root of primary and permanent teeth, as well as on smooth (adjacent to cheeks, lips and tongue) and pit-fissure (small hollows and grooves on the biting side) surfaces. In general, the etiology of dental caries is complicated by uneven risk across dentitions, therefore, studies of dental caries often investigate primary and permanent dentitions separately for a better understanding of cariogenesis.

Figure 1 Normal tooth structure



Adapted from Pitts et al., 2017 (Pitts et al., 2017). Enamel is composed of mineral covering the crown, cementum composed of mineral and collagen is the outermost layer of the root, dentine is beneath both enamel and cementum. The dental pellicle is the salivary protein-rich film that covers the tooth crown. Over time a bacterial biofilm develops over the pellicle. In humans, primary teeth erupt around age 6-33 months and gradually exfoliate and are replaced by permanent teeth around age 6-12 years.

2.1.1 Epidemiology

Dental caries is one of the most common chronic diseases in both children and adults, worldwide. In 2010, the global prevalence of untreated dentine carious lesions in the primary dentition was around 9% and the incidence was 15,205 cases per 100,000 person-years, while in the permanent dentition, the prevalence was about 35% and incidence was 27,257 cases per 100,000 person-years. Worldwide, untreated caries in the primary dentition affected 621 million children, which ranked 10th most prevalent health condition and untreated caries in the permanent

dentition was the most prevalent health condition affecting 2.4 billion people (Kassebaum et al., 2015).

The prevalence of dental caries can be influenced by multiple factors, such as age, ethnicity, socioeconomic status and development degree of countries. In the US, untreated caries affected 8.8% of children aged 2-5 years, 15.3% aged 6-11 years and 13.4% aged 12-19 years in 2015-2016 (Fleming & Afful, 2018). In adults, the prevalence of untreated caries was approximately 26% from age 20-64 years and 18% above age 65 years. Dental caries is more prevalent in populations with Hispanic ethnicity, non-Hispanic blacks and lower socioeconomic status. According to WHO, many developing countries experience an increasing prevalence of caries, mostly due to increasing sugar consumption, unhealthy diet pattern and lack of access to dental care.

A traditional measurement of caries experience is DMF/dmf (Decayed, Missing and Filling) index, which is the total count of affected teeth (DMFT/dmft) or surfaces (DMFS/dmfs). The upper-case acronym is usually used to represent the index for the permanent dentition and the lower-case acronym is usually used to represent the index for the primary dentition. This convention is used in this dissertation. DMF index represents cumulative caries experience including both current and past caries. According to the National Health and Nutrition Examination Survey (NHANES) in 1999-2004, the average dmft/DMFT is 1.17 among children aged 2-5 years, 2.29 in youth aged 6-11 years (dmft in the primary dentition + DMFT in the permanent dentition), 0.54 in youth aged 12-19 years, 3.28 in adults aged 20-64 years and 9.24 in seniors 65 and older. The lower average DMFT from age 12-19 is due to the exfoliation of affected primary teeth and limited exposure duration of newly erupted permanent teeth. In epidemiological studies, different thresholds of DMF index can be used to determine the proportion of affected individuals. For example, the proportion of affected children at age 15 years varies from 11% to

52% based on different DMF thresholds (Pitts et al., 2017). Therefore, comparisons of prevalence and distribution of dental caries can be complicated by diagnostic criteria.

Dental caries remains a huge burden to society. In the US, the costs of dental services have increased to $143.2 billion in 2019 (*National Health Expenditures 2019 Highlights*, [accessed 2020]). Untreated caries can impact quality of life significantly in both children and adults. In 1984 National Health Interview Survey report, children and adolescents under 18 years had 5 million restricted activity days, over 1.6 million days in bed, and over 1.7 million missed school days due to acute dental conditions (Waldman, 1987). Moreover, untreated caries can lead to dental pain, chronic pain, loss of teeth, poor school performance, poor social relationships, and decreased success later in life. The consequence of untreated dental caries is not limited to children. Employed adults lose 92.4 million work hours per year in seeking dental care (Kelekar & Naavaal, 2018) in the US. One third of adults aged 50 years and older had problems with eating and social communication and were dissatisfied with their oral health (Locker, 1992). The impact of dental caries on children and adults highlights the urges of the management and prevention of dental caries.

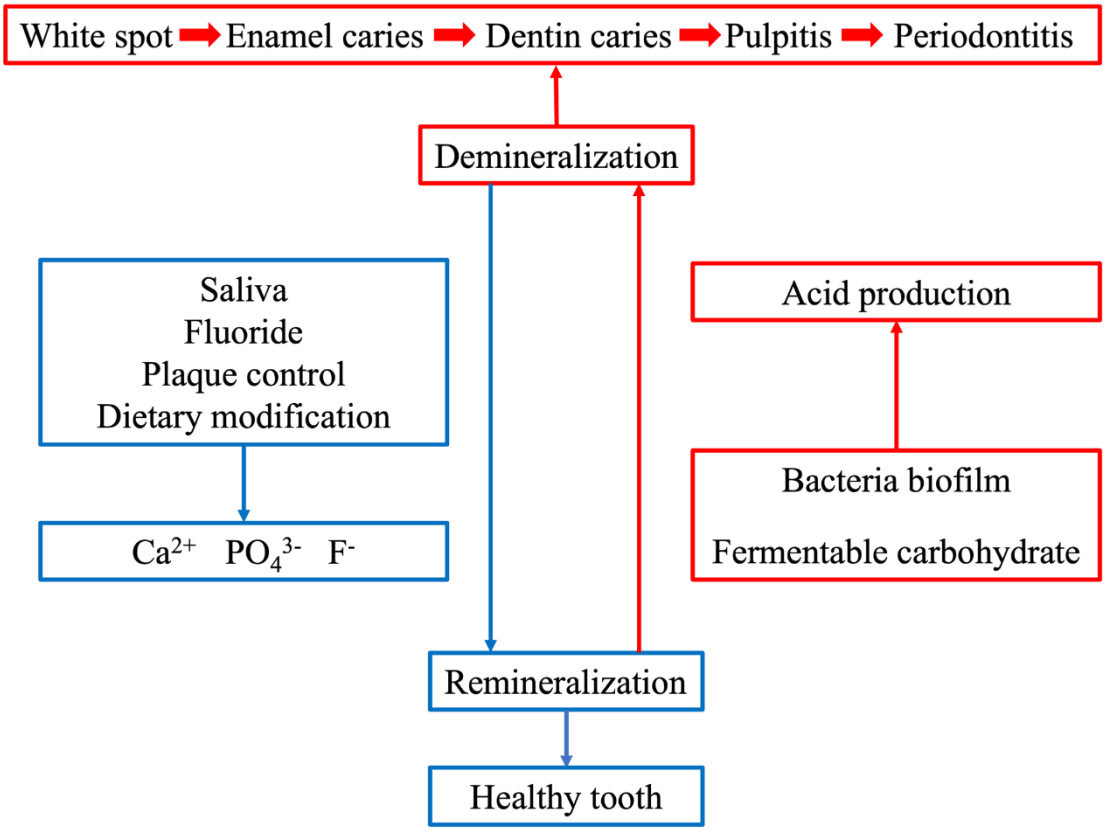
2.1.2 Pathogenesis

Dental caries is a complex biofilm-mediated disease that results from interactions over time among multiple factors including cariogenic biofilm, fermentable carbohydrate, poor oral hygiene, and insufficient fluoride exposure (Pitts et al., 2017). Typically, the process of dental caries starts at the enamel surface, where bacteria live in microcolonies, and their by-products accumulate in a biofilm. Endogenous bacteria (e.g., *Streptococcus mutans*) generate organic acids while metabolizing fermentable carbohydrates (e.g., glucose). The build-up acids lead to local pH drop

to a certain value resulting in the demineralization of crystalline mineral structure on tooth surface. At the early stage, demineralization can be arrested or reversed. First, the products from the dissolution (i.e., calcium, phosphate and carbonate) increase saturation degree and moderately prevent the surface from further demineralization. Second, fluoride helps remineralization by acting as a catalyst for the calcium and phosphate to rebuild the crystalline mineral structure. Third, saliva serves as a buffer to neutralize the biofilm pH, which stops demineralization. Figure 2 demonstrates the process of caries involving repeated cycles of demineralization and remineralization.

The process of demineralization and remineralization is dynamic and influences the development of dental caries. If causative factors (e.g., sugar, cariogenic bacteria) persist for too long, the progression of demineralization will continue resulting in increased porosity and softening surface, which allows acids to go deeper into the subsurface, then a white spot (incipient carious lesion) appears due to mineral loss. A white spot is an important stage for clinical practice when the lesion can be repaired by correcting cariogenic factors and applying preventive methods. If the caries develops further, micro-cavitations or even physical cavitation will be formed in the enamel or in root. At this extensive stage, a lesion may be still arrested by treatments like drilling out and filling the lesion. However, if caries reaches to an irreversible stage, dental pulp will be damaged eventually and operative interventions are needed, such as root canal treatment and tooth extraction. Therefore, it is critical to monitor the status of teeth and prevent caries development in the early stage.

Figure 2 Diagram of repeated cycle of demineralization and remineralization



The persistence of causative factors breaks the balance between demineralization and remineralization, leading to dental caries progression. Carious lesions can be corrected and/or arrested in the early stage.

2.1.3 Risk factors

Dental caries is a complex disease influenced over time by multiple factors and interactions among them. Environmental risk factors, such as socioeconomic status (SES) and fluoride exposure, have been reported to be associated caries risk by a range of robust evidence (Schwendicke et al., 2015; ten Cate & Featherstone, 1991). Occupation, education, and income are commonly used as SES indicators. Previous studies showed that lower parental education was associated with a higher risk of dental caries and lesser use of dental services (Hooley, Skouteris,

Boganin, Satur, & Kilpatrick, 2012). People with higher educational level are more receptive to oral health knowledge, which leads to better dental care. Similarly, families with lower income level have a higher risk of caries, which may be due to more cariogenic diets and/or eating habits and/or more limited access to education and dental services. The use of fluoride in caries prevention and control is well-established worldwide. As described above, fluoride helps with the remineralization process to arrest or reverse the dissolution. Fluoride delivery methods include community-based (e.g., water fluoridation) and individual-level (e.g., fluoride toothpaste) approaches, which are proven to be beneficial towards overall caries reduction in many countries (Rugg-Gunn et al., 2016; Slade, Grider, Maas, & Sanders, 2018).

Behavioral risk factors related to dental caries include poor oral hygiene (Pitts et al., 2017), cariogenic diet (Blostein, Jansen, Jones, Marshall, & Foxman, 2020) and methods of feeding infants. Tooth brushing helps with the removal of dental biofilm where the metabolic activities leading to mineral loss occur. In addition, brushing with fluoridated toothpaste is a convenient and cost-effective approach to deliver fluoride to the surface of the teeth. The combination of regular tooth brushing and fluoridated toothpaste is associated with a lower risk of dental caries (Cury & Tenuta, 2014; Kumar, Tadakamadla, & Johnson, 2016). Cariogenic diet, especially sugar consumption, is a critical risk factor for caries onset and development. The two main factors that strengthen the role of cariogenic diet are the age at first sugar introduced and the frequency of sugar intake. Longitudinal studies have shown that sugar consumption in the first year of life is associated with the occurrence of dental caries in subsequent years (Feldens, Rodrigues, de Anastácio, Vítolo, & Chaffee, 2018), and the higher frequency of sugar-sweetened beverage (SSB) intake is associated with increment (i.e., increase across time) of DMFT (Bernabé, Vehkalahti, Sheiham, Aromaa, & Suominen, 2014) and amount of root caries (Kaye et al., 2015). Early

introduction of sugar boosts the colonization of cariogenic microbiota on newly erupted teeth and promotes children's desire for sugar in their later life. Higher frequency of sugar consumption increases the level of produced acids leading to a lower the pH value in the dental biofilm. A systematic review meta-analyzed 5 cross-sectional, 1 case-control and 1 cohort study to investigate the associations of breast and bottle feeding with dental caries reported that breastfed children had less caries than bottle fed children (Avila, Pordeus, Paiva, & Martins, 2015). However, there are some studies demonstrating that prolonged breastfeeding increases the risk of dental caries in infants (Jain et al., 2015; Kato et al., 2015), while others have reported no association between breastfeeding practices and caries (Nirunsittirat et al., 2016; Perera, Fernando, Warnakulasooriya, & Ranathunga, 2014). Further longitudinal studies are needed to confirm the association.

Biological risk factors that are linked to dental caries experience include previous caries experience (Mejàre et al., 2014), oral microbiome (Pitts et al., 2017; Ramadugu et al., 2020), insufficient salivary flow (Buzalaf et al., 2020) and genetic factors. Previous caries experience is a well-established predictor for caries development in the future, which has been shown by several epidemiological studies and reviews (Fontana, Young, & Wolff, 2009). On sound tooth surfaces, the component of microbiota is mainly non-*mutans streptococci* and *Actinomyces*, thus the level of acids is low to mild. When sugar is introduced, the increased acidification changes the composition of the oral microbiota with increasing acidogenic and aciduric bacteria (e.g., *mutans streptococci*, *lactobacilli*, and yeast). Saliva is one of the most essential biological factors for protection against caries. As described above, saliva as a buffer restores the dental biofilm pH and promotes remineralization with diffusion of calcium and phosphate. Furthermore, salivary flow removes the acids on tooth surfaces, and salivary proteins form the enamel pellicle that serves as

a barrier between acids and tooth surfaces (Pyati, Naveen Kumar, Kumar, Praveen Kumar, & Parveen Reddy, 2018). Therefore, hyposalivation increases caries risk.

2.2 Genome-wide association studies of dental caries

A strong genetic component to dental caries has been demonstrated by twin studies (Boraas, Messer, & Till, 1988; Bretz, Corby, Hart, et al., 2005; Bretz et al., 2006; Bretz, Corby, Schork, et al., 2005; Conry, Messer, Boraas, Aeppli, & Bouchard, 1993; Haworth et al., 2020), family-based studies (Shaffer, Feingold, et al., 2012; Shaffer, Wang, et al., 2012; X. Wang et al., 2010), and animal models (Nariyama, Shimizu, Uematsu, & Maeda, 2004). These studies have reported a wide range of heritability estimates indicating that the genetic factors explain 20%-70% of variation in the dental caries traits. This wide range of heritability estimates may be due to different study designs, study populations, sample sizes, or caries phenotyping methods.

To explore the genetic basis of dental caries, several "hypothesis-free" GWASs have been performed over the past decade. Appendix Table 1 summarizes the GWASs of dental caries, to date, and the identified risk genes. For the most part, these prior GWASs have investigated the genetic effects on dental caries in the primary and permanent dentitions separately based on the assumption that the genetic determinants of caries may differ between primary and permanent dentition (X. Wang et al., 2010). The major findings for each dentition are summarized, below.

GWASs of caries in the primary dentition have investigated clinical phenotypes, including binary traits (e.g., presence or absence of caries) and dft/dfs indices, in samples ranging from approximately 100 to 19,000 individuals (Zeng et al., 2014) (Ballantine et al., 2018; Haworth et al., 2018). These studies have identified several genetic variants associated with caries phenotypes

in the primary dentition, but few of the identified genetic variants were genome-wide significant and none have been replicated in independent studies. Moreover, the SNP-based heritability (i.e., the proportion of the phenotypic variance explained by a given set of 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% (95% CI: 0%-7%) (Haworth et al., 2018).

Across the GWASs of caries in the permanent dentition, the commonly used phenotypes were decayed, missing, and filled teeth/surfaces (DMFT/DMFS). The first study in 7443 white adults did not find any genome-wide significant associations, but 8 suggestive loci were observed including *RPS6KA2, PTK2B, CNIH, ISL1, RHOU, FZD1, ADAMTS2, and* TLR2 (X. Wang et al., 2012). In addition to global DMFT/DMFS, surface-specific caries phenotypes were also used in GWASs, such as partial DMFS scores for pit-and-fissure and smooth surfaces (Zeng et al., 2013), and partial DMFS scores for subsets of tooth surfaces grouped together based on co-occurrence of caries lesions (Shaffer, Feingold, Wang, Lee, et al., 2013; Shaffer, Feingold, Wang, Weeks, et al., 2013). Besides the white population, a GWAS in 11,754 Hispanic/Latino adults identified several genome-wide significant loci with DMFT and DMFS as phenotypes, including *SYPL1*, *NAMPT*, *BMP7*, *SPO11*, *IGSF10*, and *AADACL2* (Morrison et al., 2016). Though GWASs of caries in the permanent dentition have identified more genome-wide associations than in the primary dentition, the SNP-based heritability estimates of caries in the permanent dentition were still low. A GWAS in 13,353 individuals using a binary phenotype (affected vs. unaffected) reported the consortium-wide SNP-based heritability estimate was only 6% (95% CI: 0%-13%) (Haworth et al., 2018). The

latest and largest GWAS in 26,792 adults reported the SNP-based heritability estimate of DMFS was less than 10% (Shungin et al., 2019).

Altogether, GWASs for dental caries have identified genetic variants influencing caries risk and nominated new genes with biologically plausible roles in dental caries. However, the identified genetic loci, to date, cumulatively explain only a small portion of the heritability, and very few nominated genes are proven to play roles in dental caries. In addition, prior GWASs were cross-sectional analyses using phenotypes collected at a single time point. Therefore, these studies were not able to consider some aspects of dental caries experience, such as the time to first caries incidence or caries development over time. Likewise, the GWASs of quantitative caries phenotypes performed to date have tested the genetic effects on the phenotypic mean, whereas testing the effects on phenotypic variability has not been performed. These understudied aspects of caries experience represent gaps in our knowledge of the genetic contributors to dental caries.

As described above, dental caries is a multifactorial disease. However, people may be more susceptible or resistant to dental caries when exposed to the same levels of environmental risk factors. Similarly, genetic effects on dental caries can differ between individuals depending on environmental risk factors due to gene-environment interactions (GEI). To date, few studies have investigated GEI for dental caries. Shaffer et al. (Shaffer, Carlson, et al., 2015) tested the interaction effects between 4 enamel matrix genes with fluoride exposure on caries experience. They found participants with risk alleles of two genetic variants (upstream of *TUFT1* and missense in *AMBN*) showed higher levels of dental caries only if they lacked exposures to fluoride. In addition to candidate gene approaches, a genome wide GEI study was conducted evaluating the interaction between presence of oral *Streptococcus mutans*, a putative caries-promoting bacteria, and host caries susceptibility (Meng, Wu, Billings, Kopycka-Kedzierawski, & Xiao, 2019). Three

suggestive loci in genes *IL32, GALK2, CELF4* were identified that interacted with *S. mutans*. Prior studies had limited success in explaining GEI effects on dental caries either with narrowly tested genes or environmental factors. Furthermore, it is difficult to identify the GEI because an accurate record of environmental exposures is challenging to collect, and modeling GEI requires larger sample sizes (H. Wang et al., 2019) than modeling main effects. An alternative approach is needed to address the problem.

2.3 Rationale for the specific aims

2.3.1 Aim 1. Identify the genetic variants associated with caries incidence

Time has been considered as one of the most important components in caries development. However, most genetic studies have used a cross-sectional design and coded caries experience as binary (Yes/No affected status) or a semi-quantitative (DMFT/DMFS/dft/dfs) outcomes, which do not contain complete information about the timing of decay, even with age adjustments or age-restricted samples, which makes drawing causal inferences difficult. Survival analysis is a method for analyzing the duration of time until an event happens, which retains "time at risk" or "survival time" information. Moreover, survival analysis methods can handle censored data, so participants without complete follow-up can also provide information.

Survival analysis has been used to identify risk factors associated with dental caries via Cox proportional hazards regression models (Cortellazzi et al., 2013; T. S. Ghazal et al., 2019; H. J. Lee, Kim, Jin, Paik, & Bae, 2015; Ollila & Larmas, 2007). A 7-year survival analysis study in 183 children, for instance, found consumption of candies and lack of daily toothbrushing affected

the time to caries incidence in both primary and permanent molars (Ollila & Larmas, 2007). Survival analysis has also been used in GWASs (survival GWAS) to identify genetic variants associated with survival time of specific diseases such as several cancers, including pancreatic, lung, and breast, as well as age-related macular degeneration (Song et al., 2015; Tang et al., 2015; Wu et al., 2014; Yan et al., 2018).

Survival GWAS allows us to investigate the genetic influence on disease development and progression yet has not been applied to dental caries studies previously. Efforts in survival GWAS of time to first caries incidence are needed to answer questions such as to what degree genetics contributes to caries incidence and what specific genes are associated with time to first caries incidence. Answering these questions will broaden our understanding of the genetic underpinnings of dental caries onset.

2.3.2 Aim 2. Identify genetic variants associated with caries development

Longitudinal studies with repeated measurements at multiple time points are crucial in clinical and epidemiological research since they have more power to detect causal relationships than cross-sectional studies. With the interest in the genetics of longitudinal traits, longitudinal methods have been used in GWAS (longitudinal GWAS) in order to explore the continuous effects of genetic variants on the development and changes of human traits over time (Fan et al., 2012; Mei et al., 2012; Sikorska et al., 2013; Smith et al., 2010).

One of the most critical advantages of longitudinal GWAS is the power gain over cross-sectional GWAS. Longitudinal GWAS not only evaluates the phenotypic mean across genotype groups at the baseline, but also evaluates the rates of the phenotypic changes among the genotypic groups. Therefore, longitudinal GWAS may provide more precise estimates of the associations

between genetic variants and phenotypes, and more power to detect associations than cross-sectional GWAS. For example, Xu et al. (Xu, Shen, & Pan, 2014) compared longitudinal GWASs and cross-sectional GWASs of 56 brain-wide imaging phenotypes. At the genome-wide significant level, cross-sectional GWASs of baseline phenotypes identified two SNPs associated with 5 brain phenotypes, while longitudinal GWASs from baseline to month 48 identified an additional associated locus, and those three loci were cumulatively associated with 46 phenotypes. Thus, compared to cross-sectional GWAS, longitudinal GWAS both expanded the list of associated loci and deepened the understanding of their phenotypic effects.

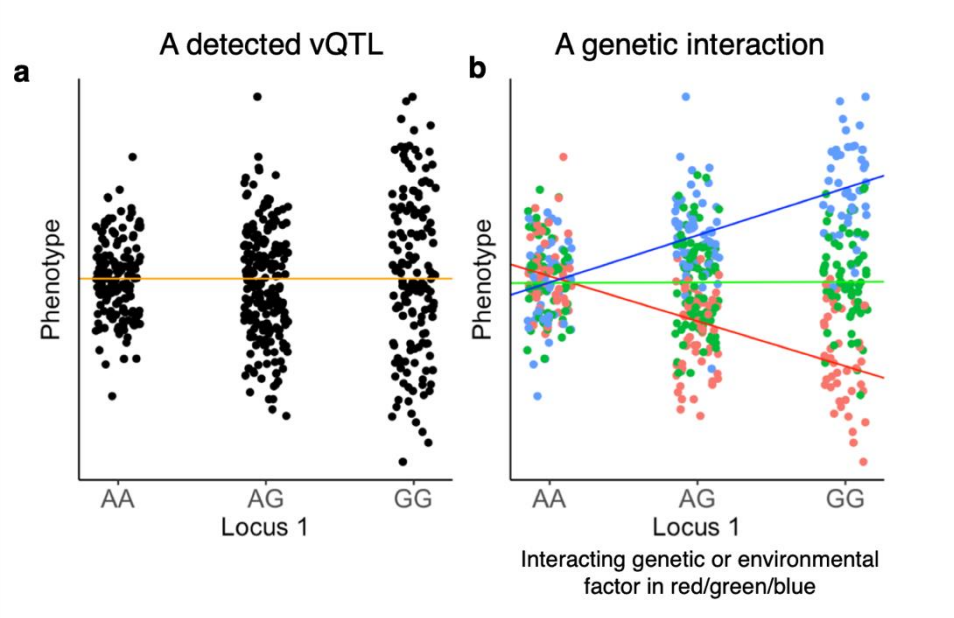
Though mixed models are commonly used to handle the correlated nature of repeated measurements in longitudinal studies, there remain challenges in computational efficiency when fitting mixed models to millions of SNPs in the context of a GWAS. Generalized linear mixed model association tests (GMMAT) (Chen et al., 2016) is an R package using generalized linear mixed models (GLMMs) that scale to the genome by first fitting a GLMM with covariate adjustment and random effects to account for population structure and relatedness, and then performing score tests for each genetic variant. It is an efficient tool which takes only several hours to test millions of genetic variants in thousands of people. Longitudinal methods have not previously been used to model the genetic underpinnings of dental caries development and changes over time; therefore, the first longitudinal GWAS on dental caries across multiple time points is needed.

2.3.3 Aim 3. Identify genetic variants associated with caries variability

Traditional GWAS seeks to identify genetic variants influencing the phenotypic mean with the assumption that all genotype groups have the same residual variance. However, the residual

variance can be different across groups of individuals due to extrinsic (e.g., environmental factors) and intrinsic (e.g., sex or genetics) factors. Some specific genetic variants, known as variance heterogeneity quantitative trait loci (vQTL) have been identified with effects on the variability of a trait (Rönnegård & Valdar, 2012), i.e., showing genetic variance heterogeneity (Figure 1a). Furthermore, that genetic variance heterogeneity may be partially explained by interaction effects, therefore, vQTL can be used to prioritize SNPs for interaction testing (A. A. Brown et al., 2014; Forsberg & Carlborg, 2017; J. Yang et al., 2012) (Figure 1b). For example, a genome-wide vQTL scan of body mass index (BMI) reported the difference in variance for BMI was approximately 7% between opposite homozygous groups at the *FTO* locus. This result was consistent with reported *FTO* by environment interactions for BMI (J. Yang et al., 2012). Another recent vQTL scan of 13 quantitative traits in the UK Biobank identified 75 significant vQTLs for 9 traits and found vQTLs were strongly enriched with GEI effects. Their findings suggested that GEI can be inferred from vQTL without knowledge of the environmental measurements (H. Wang et al., 2019). Therefore, genome-wide vQTL analysis may not only help find more associated genetic variants but may also aid the detection of GEI. Variance heterogeneity studies have not been considered for dental caries genetics, and the application of genome-wide vQTL scans for dental caries is needed to help overcome the challenges in detecting GEI described above.

Figure 3 Schematic of the differences in variance among genotype groups in the presence of vQTL and interaction effects



Adapted from Marderstein et al., 2021 (Marderstein et al., 2021).

2.4 Public Health Relevance

Dental caries is a complex disease representing a huge burden to the society. Caries prevention is of public health importance as it has both short- and long-term impact on people physically and financially. On the bright side, dental caries is one of the most preventable diseases. Understanding the genetic basis of dental caries is one of the key steps in enabling the design of personalized interventions, for example, those leading to the changes in behaviors and dietary patterns, or even exposures to environmental risk factors. However, there are major gaps in our knowledge of the genetics of dental caries, especially the onset and development of dental caries. This study focused on the genetic effects on dental caries incidence (time to first caries incidence),

development over time (repeated caries measurements at multiple times) and variability (caries variance). To our knowledge, no previous research has investigated caries genetics in these areas. Filling these gaps that may provide the foundation for better early detection, risk assessment, risk predication, dental care, and effective public health interventions, and thereby lower the disease burden, alleviate health inequities, and improve the overall health.

3.0 Identifying the genetic variants associated with caries incidence in the primary dentition

3.1 Background

This chapter discusses the potential risk factors and genetics variants associated with caries incidence. As described above, 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. Behavioral risk factors related to dental caries include poor oral hygiene and cariogenic diet, especially sugar intake, which is a key component influencing caries development. Moreover, biological factors including the oral microbiome and genetic factors are associated with caries experience (Selwitz, Ismail, & Pitts, 2007).

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 (Meurman & Pienihäkkinen, 2010; Peltzer, Mongkolchati, Satchaiyan, Rajchagool, & Pimpak, 2014), previous caries experience, consumption of sweetened foods/drinks (Corrêa-Faria, Paixão-Gonçalves, Paiva, & Pordeus, 2016; Lim, Tellez, & Ismail, 2015), *Streptococcus mutans* level (Meurman & Pienihäkkinen, 2010) and fluoridation in drinking water (Peltzer et al., 2014). 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 (Ollila & Larmas, 2007), the time to DMFT increment (H. J. Lee et al., 2015) and the time to first caries in permanent teeth (T. S. Ghazal et al., 2019). 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. Moreover, prior GWASs used phenotypes collected at a single time point, thus were not able to evaluate the genetic basis of time-to-caries incidence. Survival analysis with 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.

3.2 Method

3.2.1 Study sample

The study design and recruitment descriptions of the Center for Oral Health Research in Appalachia, cohort 2 (COHRA2) have been reported before (Neiswanger et al., 2015). Briefly, COHRA2 is a longitudinal study and has been recruiting and prospectively following pregnant women 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.

3.2.2 Phenotype and genotype

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 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.

Participants were genotyped using the Infinuim Multi-Ethnic Global-8 v1.0 Array. Ungenotyped 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 (HRC r1.1) 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.

3.2.3 Risk factor identification

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]), annual household income (<$50,000 or ≥$50,000), home water source (water company or well), home fluoride level (ppm, measured in a water sample), mother's educational

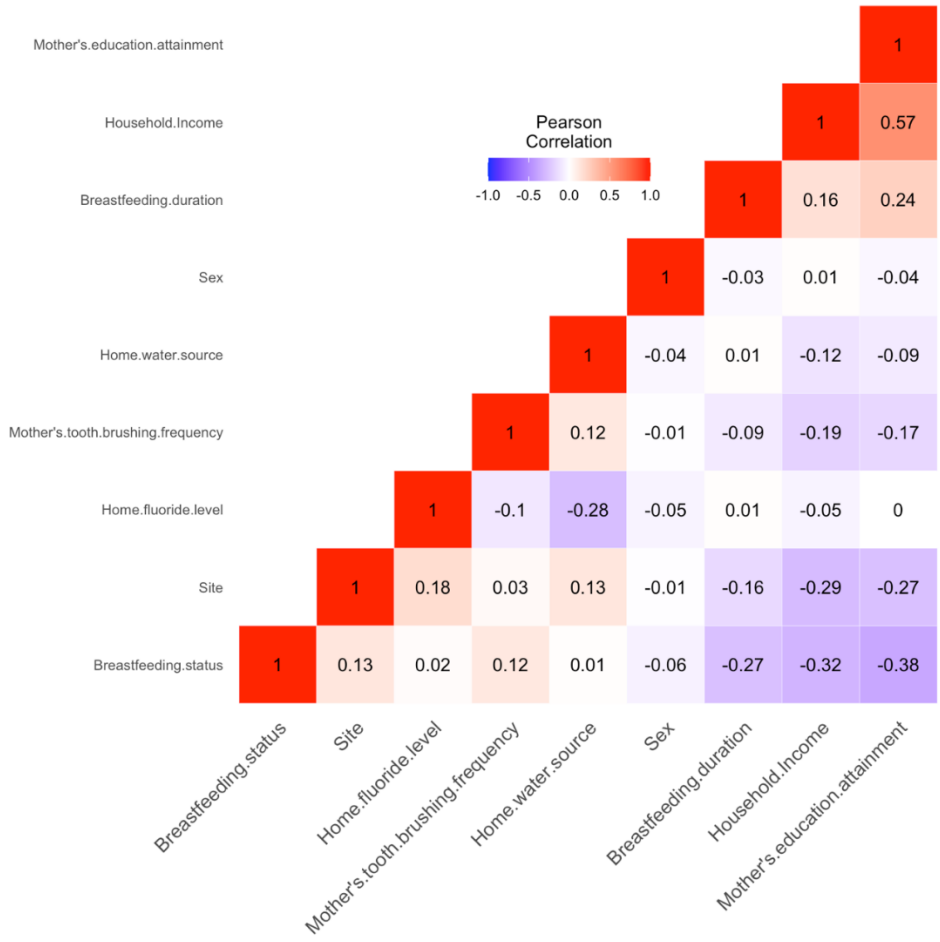
attainment (<high school, high school or some college, or college degree or more), mother's tooth brushing frequency (>1/day, 1/day, or <1/day), breastfeeding status at baseline (yes or no), and breastfeeding duration (months). The correlations among factors were calculated using Pearson's correlation method. There were no strong correlations (Figure 2).

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 Cox modeling was performed via the R package "My.stepwise". In the bidirectional stepwise (with iterations between the "forward" and "backward" steps) Cox modeling, first, all potential risk factors were put on the "variable list" to be selected, the significance levels for entry into (SLE) and for retention in (SLS) 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 stepwise regression can result in poor model fit out-of-sample, we also performed regularized Cox modeling as a second variable selection procedure as confirmation of the variables included in the final model using the R package glmnet (Friedman, Hastie, & Tibshirani, 2010). 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.

Figure 4 Pearson's correlations among factors in COHRA2 children



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).

3.2.4 Survival GWAS

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" (Rizvi et al., 2019), while adjusting for sex, 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 (PCA) 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 R. The regional association plots showing nearby genes were visualized using LocusZoom (Pruim et al., 2010). We defined a SNP is located near a gene if this gene is located within a 500kb window around this SNP.

3.2.5 Gene-based and gene set enrichment analyses

We performed a gene-based analysis using MAGMA (de Leeuw, Mooij, Heskes, & Posthuma, 2015) and a gene set enrichment analysis using Genomic Regions Enrichment of Annotations (GREAT) (McLean et al., 2010) with 14 independent SNPs in suggestive signals 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.

3.2.6 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 (Speed & Balding, 2019), 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.

3.2.7 Transcriptome-wide association study (TWAS)

The purpose of TWAS is to test the association between predicted gene expression and caries incidence in order to assess the mediating effects of gene expression levels on caries incidence in different tissues. The design of TWAS not only benefits the gene prioritization, but also the power gain from the cumulative genetic effects. However, TWAS does not provide evidence for causal inference and there are some concerns when applying TWAS to dental caries studies. First, tissue bias is one of the vulnerabilities of TWAS. It is ideal to use the most mechanistically relevant tissue to build gene expression prediction model for better performance. However, dental caries may act through multiple tissues, which has not been fully understood, thus, the most relevant tissue for dental caries remains unclear. Under this circumstance, we tested 49 available tissues in GTEx using S-PrediXcan as recommended. Second, the sample size of eQTL data influences largely on the power of TWAS (Gusev et al., 2016). In GTEx, the sample sizes for

49 tissues vary from approximately 70 to 360. There is a chance that the sample sizes for the relevant tissues that are involved in dental caries are too small to have enough power to detect weaker associations.

S-PrediXcan software was used to conduct TWAS, which provides 49 GTEx tissue types. Both 49 single-tissue TWASs and multiple-tissue TWAS (meta-analysis of 49 single-tissue TWAS) were performed using the Multivariate Adaptive Shrinkage in R (MASHR) model, which uses only putative causal (e.g., fine-mapped and biological relevant) eQTLs as predictors. All SNPs from the survival GWAS were included in the TWAS. The genome-wide significance threshold was set at P<2.4E-6 (i.e., 0.05/20807, the Bonferroni correction for 20807 tested genes) for both single-tissue and multiple-tissue TWAS.

3.2.8 Candidate SNP replication

In addition to the GWAS approach, a candidate SNP replication study was conducted to test the associations between time to first caries incidence and genetic variants that were reported to be significantly associated with dental caries traits in previous association studies using the same Cox models and covariates as described above. Candidate SNPs with minor allele frequencies of 0% in our sample were excluded. For variants without genotyping information in our data, proxy SNPs in high LD ($r^2 \geq 0.75$) with candidate SNPs were tested instead. In total, 118 variants were selected for the candidate SNP replication analysis; among those, 23 variants were from 5 GWASs (Haworth et al., 2018; Morrison et al., 2016; Shaffer, Feingold, Wang, Lee, et al., 2013; Shungin et al., 2019; J. Yang, Zhu, Chen, Zhang, & Wu, 2014), 37 variants were from 5 linkage studies and follow-up fine-mapping studies (Briseño-Ruiz et al., 2013; Küchler et al., 2013; Küchler et al., 2014; Shimizu et al., 2013; Weber et al., 2014), and 58 variants were from 22 candidate gene

studies (Anjomshoaa et al., 2015; Antunes et al., 2014; Chaussain et al., 2014; Deeley et al., 2008; Doetzer et al., 2015; Ergöz et al., 2014; Fine et al., 2013; Hu et al., 2019; Jeremias et al., 2013; Kang, Yoon, Lee, & Cho, 2011; Li, Hu, Zhou, Xie, & Zhang, 2015; Navarra et al., 2016; Ohta, Nishimura, & Asada, 2015; Patir et al., 2008; Romanos et al., 2015; Shaffer, Carlson, et al., 2015; Shimizu et al., 2012; Tannure, Küchler, Falagan-Lotsch, et al., 2012; Tannure, Küchler, Lips, et al., 2012; X. Wang et al., 2012; Wendell et al., 2010; Yildiz, Ermis, Calapoglu, Celik, & Türel, 2016). The category-wide significance thresholds for replicating associations nominated in prior GWAS, linkage studies, and candidate gene studies were set at $P<0.0022$, $P<0.0013$, and $P<0.00086$, respectively (i.e., 0.05 divided by the number of tested SNPs, the Bonferroni correction for tested variants). The candidate SNP study-wide significance threshold was $P<0.00042$ (i.e., 0.05/118). The reason we set category-wide significance levels is that the three categories use different approaches to test the associations between genetic variants and dental caries, providing unique evidence in each category. Therefore, we assessed the significance of the associations within each category first, then in the combined set of tested candidate SNPs.

3.3 Results

3.3.1 Sample summary

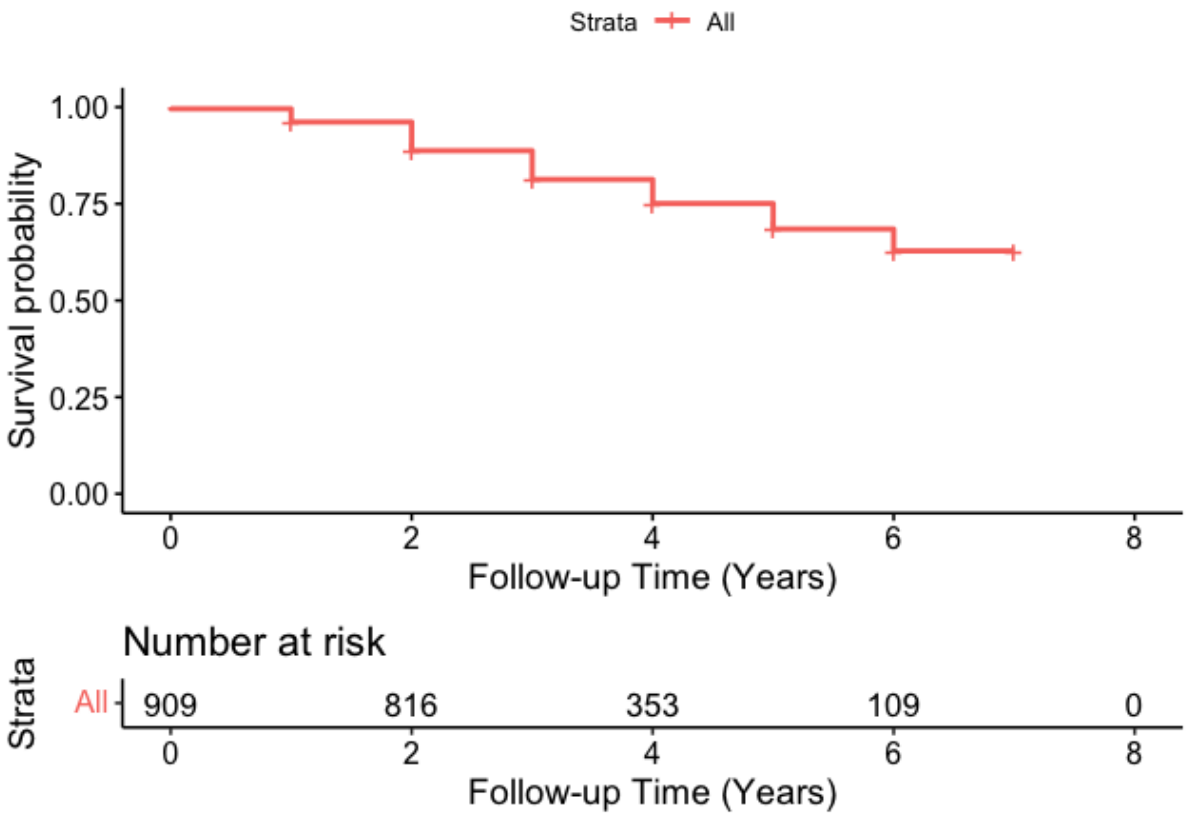
The basic characteristics of the study sample are summarized in Table 1. A total of 196 of 909 children (21.56%) had their first primary tooth caries event during the follow-up period. The caries-free survival rates were 99.6% at 1-year follow-up, 96.1% at 2-year follow-up, 88.7% at 3-

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 3).

Table 1 Basic characteristics of study sample in COHRA2 children

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

Figure 5 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.

3.3.2 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 (Table 3). Higher household income 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 4) compared to municipal water. The results of the regularized Cox model were the same as the stepwise Cox model (Figure 5). 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.

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)

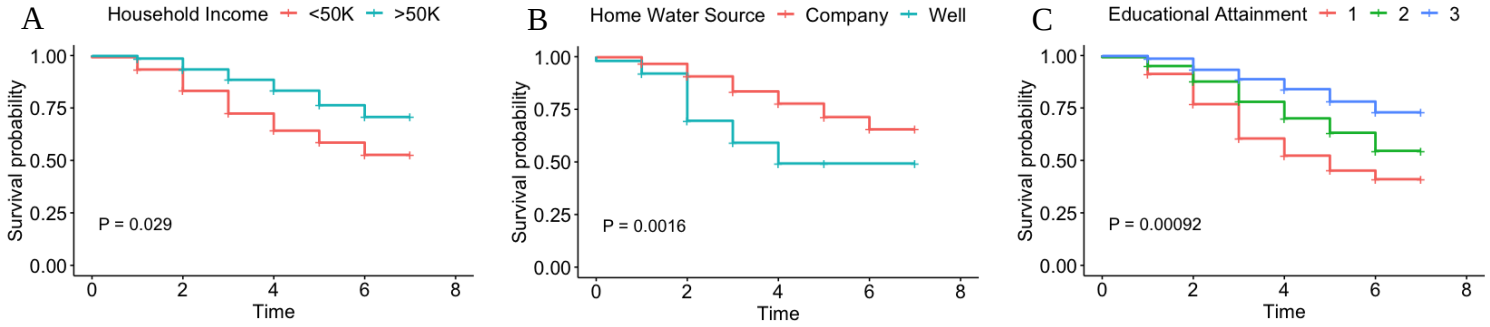
Note: PA, Pennsylvania; WV, West Virginia; SE, standard error; HR, hazard ratio; CI, confidence interval.

Table 3 The final stepwise Cox model

Variables	Reference group	Alternative group	Estimate	SE	P	HR (95% CI)
Household income	<50,000/year	≥50,000/year	-0.415	0.199	0.029	0.65 (0.44-0.96)
Home water source	Water company	Well	0.826	0.261	0.0016	2.28 (1.37-3.81)
Mother's educational attainment	<high school	≥high school	-0.434	0.125	0.00092	0.66 (0.52-0.84)

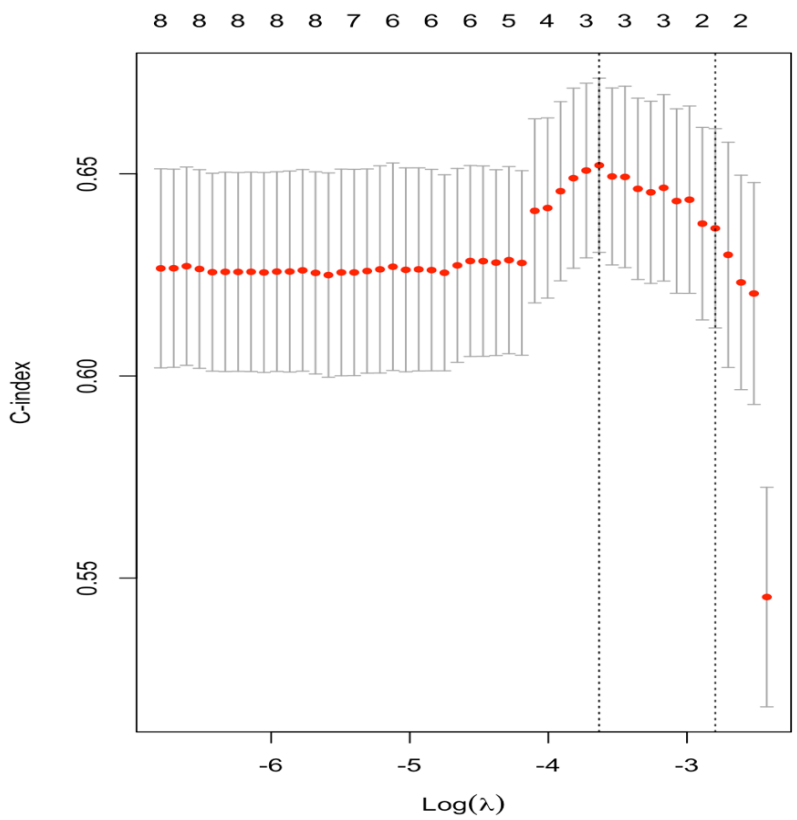
Note: SE, standard error; HR, hazard ratio; CI, confidence interval.

Figure 6 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).

Figure 7 The C-index generated from cross-validation under different values of λ



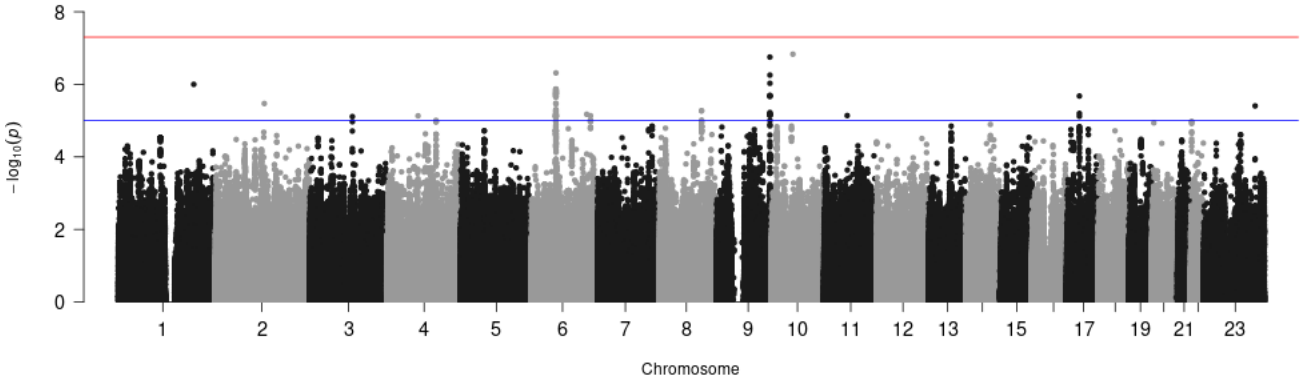
The y-axis is the C-index, the x-axis is the natural logarithm of λ, which is the tuning parameter controlling the overall strength of the penalty. The integers above the plot indicate the number of variables included in the Cox model at the current Log(λ). The left vertical line shows the highest C-index 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.

3.3.3 Heritability estimation, survival GWAS, and candidate SNP replication

The SNP-based heritability estimate of the time to first caries incidence was 0.544 (standard deviation [SD]: 0.41) indicating that genetics may have a large role in the timing or pace of the

cariogenic processes. Figure 6 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; Appendix Figure 1). 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 4 shows the 14 suggestive loci and their nearby genes. The regional association plots for these loci are provided in Appendix Figure 2. 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 level (X. Wang et al., 2012).

Figure 8 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 -log10-transformed P value, the x-axis is the physical position of each variant organized by chromosome.

Table 4 Gene nearby and biological relevance of 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 55520375	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 (Vaché et al., 2020)
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 (Giunta et al., 2002) and canine agenesis (Barbato et al., 2018) *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 (Levanti et al., 2016) and is associated with gingival inflammation (S. Zhang et al., 2016)
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 (Lamon, Wallace, Léger, & Russell, 2009) and arteriogenesis (Troidl et al., 2009) *OXR1:*is an important regulator of neuronal survival in response to oxidative stress (Oliver et al., 2011)

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 (Ward et al., 2005) *ADGRG6*: is associated with earlobe attachment (Shaffer et al., 2017) *GJE1*: is associated with hearing loss (Locher et al., 2015) *NMBR*: a member of the mammalian bombesin receptor family involved in several physiological effects (Park, Kim, Kim, Bae, & Bae, 2013)
rs11230097 11 59646339	T/C 0.182	7.26E-6	0.60 (0.48-0.75)	*TCN1*: is associated with colorectal cancer (Zhu et al., 2020) *OOSP* family: is associated with postfertilization development (Abbasi et al., 2020) *CBLIF*: encodes a glycoprotein *MRPL16*: is associated with colorectal tumors (Kim et al., 2008) *STX3*: participates in granule-granule fusion and is associated with Sjögren's syndrome (Barrera et al., 2012) *OR10V1*: encodes olfactory receptor involved in smell perception *MS4A3*: is associated with myeloid differentiation in human hematopoiesis (Ishibashi et al., 2018)
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 (Shaffer et al., 2014), and dental fluorosis (Dalledone et al., 2019) *SYNE1*: encodes spectrin repeat containing protein associated with spinocerebellar ataxia (Indelicato et al., 2019)
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).

In the candidate SNP replication, 118 variants were tested for the association with caries incidence, however, none of them showed association with time to first caries incidence (Table 5).

Table 5 Candidate SNP replication results for caries incidence

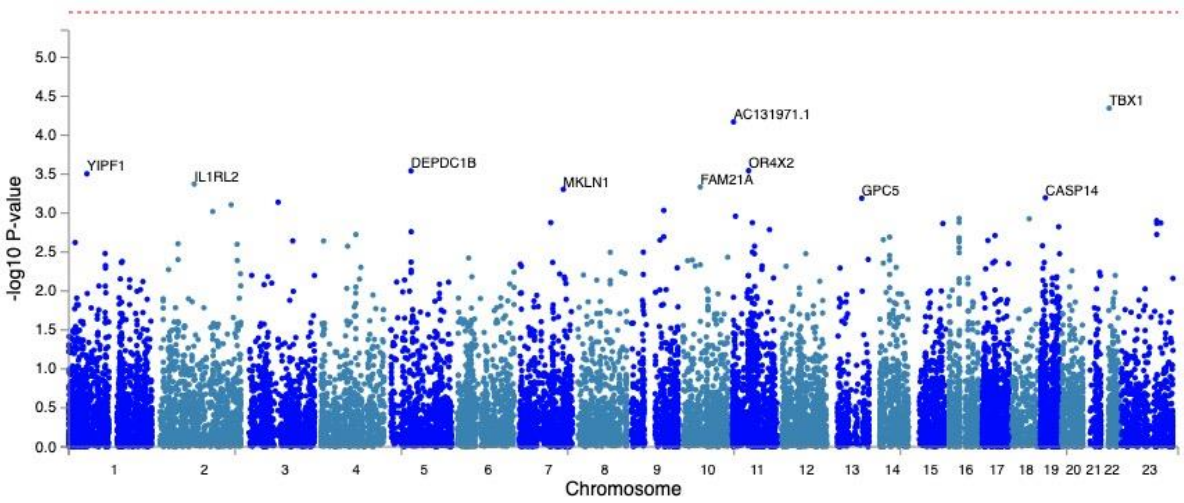
Study	Number of studies	Number of tested SNPs	Number of significant SNPs	SNP with smallest P value (P value)	Gene (Trait)
GWASs	5	23	0	rs138769355 (0.009)	NA (DMFT)
Linkage studies & follow-up fine-mapping	5	37	0	rs1077430 (0.0028)	ESRRB (high caries vs. low caries)
Candidate gene studies	22	58	0	rs1126478 (0.01)	LTF (Yes vs. No caries)

3.3.4 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<2.74E-6 (0.05/18942). There were no genes reaching genome-wide significance threshold (Figure 7). The gene with the smallest P-value was *TBX1* (P=4.49E-5). *TBX1* has been identified as the candidate gene for 22q11.2 DiGeorge syndrome (22q11.2 DS) with features including craniofacial dysmorphisms (e.g., cleft palate, tooth defects) (Jerome & Papaioannou, 2001). *TBX1* encodes a member of the T-box gene family involved in tooth morphogenesis by

regulating the proliferation, differentiation, and maturation of ameloblasts (Gao, Li, & Amendt, 2013). There was no significant enrichment in the gene set enrichment analyses.

Figure 9 Manhattan plot for gene-based analysis using MAGMA with survival GWAS summary statistics (labeling top ten genes)

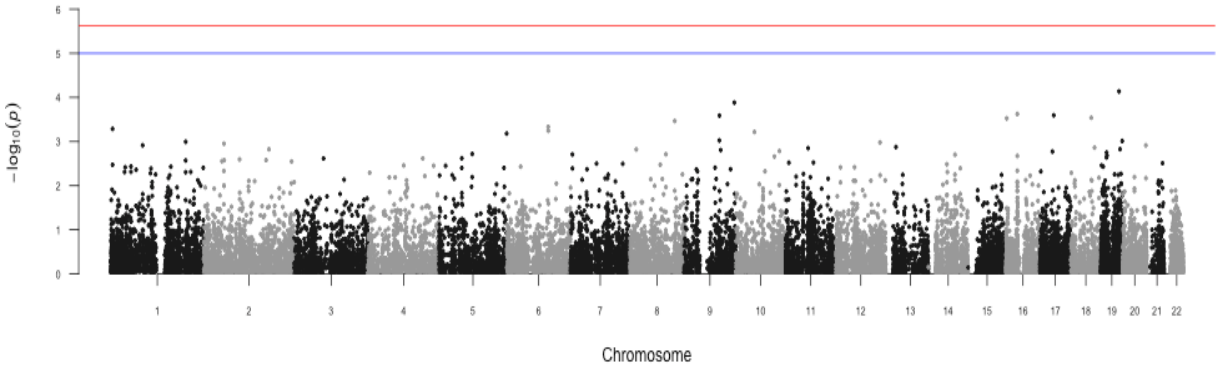


The dotted red line indicates the genome-wide significance threshold (P=2.74E-6). The y-axis is the -log10(P value), the x-axis is genes organized by physical position on each chromosome.

3.3.5 TWAS

In single-tissue TWAS, there was no genome-wide genes associated with caries incidence. Similarly, there was no genome-wide significant gene expression associated with caries incidence in multiple-tissue TWAS (Figure 10). Given the small sample size in this study and the survival GWAS results, these results are not unexpected. Further investigations with larger sample sizes and more available tissues related to caries are needed in the future.

Figure 10 Manhattan plot for multiple-tissue TWAS



The red line indicates the genome-wide significance threshold (P = 2.4E-6). The y-axis is the -log10-transformed P value, the x-axis is the physical position of each gene organized by chromosome.

3.4 Discussion

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 (Schwendicke et al., 2015), 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 (Peltzer et al., 2014) and/or that participants living in homes serviced by well water reside in more rural areas and may experience decreased access to dental care (Heaton et al., 2017). Further studies are needed to investigate the role of home water source in caries incidence and disentangle its potentially causal effects from confounding indicators of rurality. Moreover, despite the caries-preventive benefit of water fluoridation (Slade et al., 2018), 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 (Qadri, Nourallah, & Splieth, 2012) and prolonged breastfeeding (Jain et al., 2015; Kato et al., 2015) increased caries risk, while others showing no associations for breastfeeding practices (Nirunsittirat et al., 2016; Perera et al., 2014; Valaitis, Hesch, Passarelli, Sheehan, & Sinton, 2000). 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 survival GWAS for caries incidence in the primary dentition. The SNP-based heritability estimate of time to first caries incidence was 0.544 (SD: 0.41), which is similar to the heritability estimates of caries scores measured at a single point in time in the primary dentition (54%-70%) from a family study (X. Wang et al., 2010), slightly higher than the heritability estimates of the incremental increase in caries scores (30%) from a twin study (Bretz,

Corby, Schork, et al., 2005), and lightly higher than the SNP-based heritability estimate of caries status (yes or no) in the primary dentition (28% [95%CI: 9%-48%]) from a GWAS (Haworth et al., 2018). 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 more precisely estimate the heritability 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 (Mak, Png, & Lee, 2016). Mutations in *COL5A1* have been reported to be associated with Ehlers-Danlos syndrome (EDS) (Lin, Zeng, & Wang, 2019; Symoens et al., 2012). EDS patients experience a higher prevalence of oral problems including caries (De Coster, Martens, & De Paepe, 2005), tooth fractures, and periodontal disease (Hagberg, Berglund, Korpe, & Andersson-Norinder, 2004), as well as pulp calcification and abnormal tooth roots (Kapferer-Seebacher, Schnabl, Zschocke, & Pope, 2020). *ASIC2* encodes acid-sensing ion channel 2 detected in the nerves supplying the taste buds of zebrafish (Levanti et al., 2016) and in human odontoblasts implying it has a role in tooth mechanosensitivity (Solé-Magdalena et al., 2011). Its association with severe gingival inflammation has been identified in a GWAS (S. Zhang et al., 2016) and its involvement in the inflammation process has been shown in knockout mice (Gannon, McKey, Stec, & Drummond, 2015). *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 (Jedeon, Loiodice, et al., 2016). Associations with *ESR1* have been identified for several dental traits including chronic periodontitis (Shaffer et al., 2014), dental fluorosis (Dalledone et al., 2019; Saha, Goswami, Majumdar, Sikdar, & Pramanik, 2021), and tooth size (Cunha et al., 2021). 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 (X. Wang et al., 2012). 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 (Jerome & Papaioannou, 2001). *TBX1* encodes a member of the T-box gene family involved in tooth morphogenesis by regulating the proliferation, differentiation, and maturation of ameloblasts (Gao et al., 2013). 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* over-expressed mice (Gao et al., 2015). 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 (Hughes et al., 2007), 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 (Appendix Table 4). 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.

4.0 Identifying the genetic variants associated with caries development

4.1 Background

This chapter focuses on exploring the risk factors and genetic variants that are associated with caries development over time using longitudinal data analysis approaches. Longitudinal studies, in which repeated measurements are collected from the same person over a certain follow-up period, have a critical role in clinical and epidemiological research, including assessing the change in phenotypes of interest over time and the temporal association between exposure and outcome. Application of longitudinal data in GWAS can help us with identifying the genetic variants that influence the change in traits of participants over time.

The nature of longitudinal data is the dependency of repeated measurements, which are generated from the same participant at multiple time points. Therefore, cross-sectional linear regression modeling is not applicable for longitudinal data, given its assumption that the observations are independent of each other. Mixed-effects modeling is one of the methods to handle longitudinal data. Mixed-effects modeling simultaneously models fixed and random effects. Fixed effects stand for average (population-level) trends that persist across time, while random effects represent different trends among the participants. The general idea of mixed-effects modeling is that the unexplained variance in the outcome can be divided into random intercepts and random slopes. The first step in a mixed-effects model is to build a normal distribution around the intercepts and estimate the variance of the normal distribution. The variance is called the random intercept. The random intercept allows different intercepts among participants, yet the regression coefficients can also be different among participants, thus an interaction occurs between

the coefficient and the participant. Therefore, the next step is to build a normal distribution around the different coefficients and estimate the variance of the normal distribution. This variance is called the random slope. Inviting random effects in a regression model solves the non-independence problem. In longitudinal data, the response over time within participants and the variation in the time trends between participants are expected to be estimated.

Though mixed models can handle the nature of longitudinal data, fitting mixed models for millions of SNPs is time-consuming. There are several existing programs to perform longitudinal GWAS efficiently including GMMAT, HiGWAS, fGWAS and EB_{EAPML0}. Comparison of these different methods is outside the scope of this aim. GMMAT was used in this study due to the simple data structure of the input and the short computational time for running the GWAS.

Several epidemiological studies have used data from longitudinal cohorts to identify potential risk factors that are associated with the incidence of dental caries and/or the increment of DMFT/DMFS. Those identified risk factors include SES, previous caries experience, consumption of sweetened foods/drinks, *Streptococcus mutans* level, fluoridation, and tooth brushing frequency (Bernabé et al., 2014; Corrêa-Faria et al., 2016; T. Ghazal et al., 2015; Kumar et al., 2016; Leroy, Bogaerts, Martens, & Declerck, 2012; John J. Warren et al., 2009). However, to date, no longitudinal GWAS of dental caries has been carried out to investigate the genetic underpinnings of caries development over time. Therefore, this is the first longitudinal GWAS on dental caries across multiple time points.

4.2 Methods

4.2.1 Study sample

The study design and recruitment descriptions of the Iowa Fluoride Study (IFS) have been reported before (Broffitt, Levy, Warren, & Cavanaugh, 2013; Chankanka et al., 2011; J. J. Warren, Levy, & Kanellis, 2002). Briefly, IFS is a longitudinal birth cohort that recruited mothers and newborns from 8 Iowa hospital postpartum units between 1992-1995 and has been collecting fluoride exposure, diet, and other information related to dental fluorosis and caries by questionnaire surveys every 6 months since the children were 1.5 months old. Dental examinations were conducted at approximate ages of 5, 9, 13, 17 and 23 to assess dental caries experience (J. J. Warren et al., 2002). A subset of 414 non-Hispanic white individuals who had at least two dental examinations at ages 5, 9, 13 and 17 was included in this analysis. All participants provided assent, and Institutional Review Boards approved all aspects of the study.

4.2.2 Phenotype and genotype

The participants were phenotyped based on intra-oral examinations according to caries criteria that distinguished cavitated from non-cavitated lesions. The total count of decayed and filled surfaces (dfs+DFS) were computed for each participant including all tooth surfaces present across the primary, mixed, and permanent dentitions. The phenotype in this study was the repeated measurements of total dfs+DFS at the 4 time points. The phenotype will be referred as longitudinal DFS, hereafter. We included participants with dfs+DFS data at two or more time points, and set

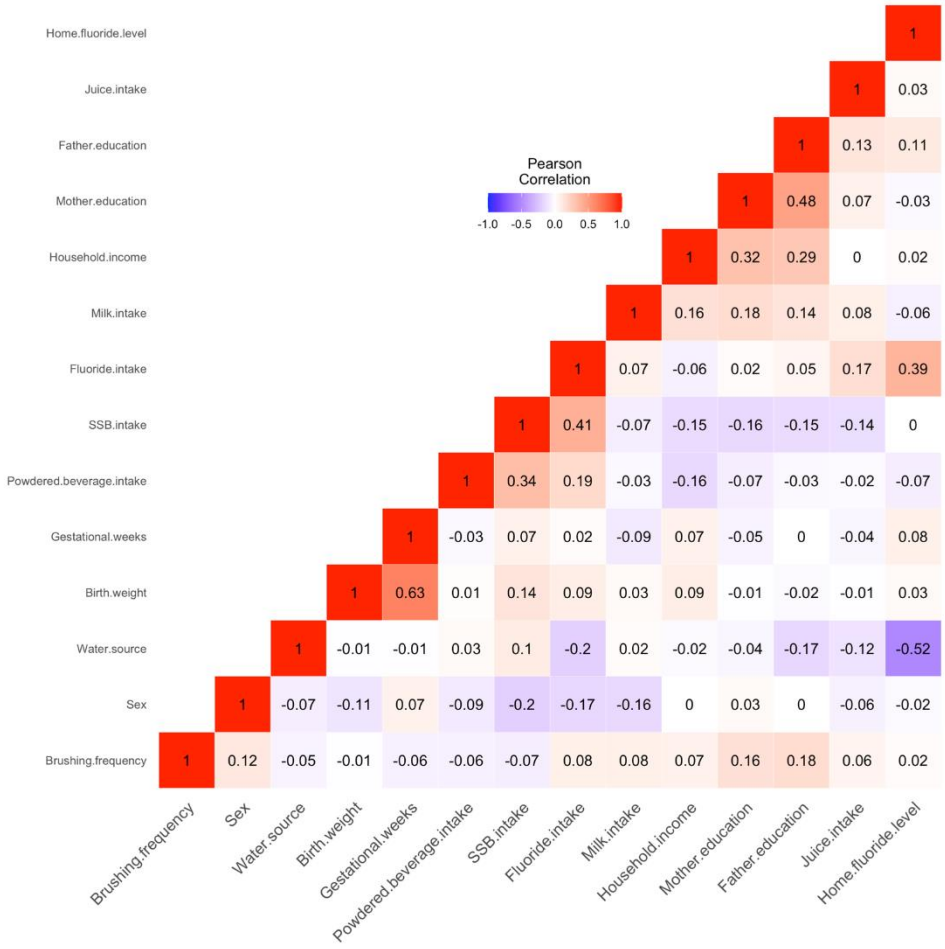
dfs+DFS at the missing time points to “NA”. The mixed-effect models that were used for risk factor identification and longitudinal GWAS could be fitted with dfs+DFS from at least two visits.

IFS was genotyped for about 580,000 SNPs by the Center for Inherited Disease Research at Johns Hopkins University using the Illumina Human610-Quadv1_B BeadChip (Illumina Inc., San Diego, Calif., USA). Ungenotyped SNPs and sporadic missing data of genotyped autosomal SNPs were imputed for a total of 16.2 million genetic variants using Michigan Imputation Server (Das et al., 2016) and Haplotype Reference Consortium r1.1 as the reference panel (McCarthy et al., 2016). Quality control procedures were the same as described in Chapter 3.2.2. A total of 5.4 million genetic variants were included in this study after filtering.

4.2.3 Risk factor identification

Demographic, behavioral and environmental factors that putatively relate to oral health were collected including sex (male or female), annual household income (<$20,000 or $20,000-39,999 or ≥$40,000), mother’s educational attainment (up to high school or some college or above four-year degree), father’s educational attainment (up to high school or some college or above four-year degree), water source (City/Public or well), home fluoride level (ppm), daily brushing frequency, birth weight (kg), gestational weeks at birth, daily milk intake (oz), daily 100% juice intake (oz), daily sugar-sweetened beverage (SSB) intake (oz), daily fluoride intake (mg), daily powdered beverage intake (oz). The Person’s correlations between 14 factors were shown in Figure 11. There were no strong correlations.

Figure 11 Person's correlations among factors in IFS



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).

Univariate linear mixed-effects models were used to identify the relationships between risk factors and longitudinal DFS with putting one of the 14 factors in the model at a time. Multiple variable and bidirectional stepwise linear mixed-effects modeling was used to identify the risk factors with independent effects on longitudinal DFS. Multiple variable modeling put all 14 factors simultaneously in the model to test the overall relationship between longitudinal DFS and risk factors, and the contribution of each independent factor to the relationship, while bidirectional

stepwise modeling was used to investigate the best combination of independent factors to predict the longitudinal DFS from iterations of forward and reverse variable selection. Both univariate and multible variable linear mixed-effects models were performed using “lme” function in R package “nlme” and stepwise mixed-effects model was performed using “stepcAIC” function in R package “cAIC4”.

4.2.4 Longitudinal GWAS and candidate SNP replication

Linear mixed-effect models were used to assess the associations between longitudinal DFS and 5.4 million genetic variants in the R package “GMMAT”, while adjusting for age, sex, the first 6 PCs of ancestry derived from a PCA of genome-wide genetic data, household income, daily 100% juice intake and daily SSB intake. The last three adjusted factors were significantly associated with longitudinal DFS in the final mixed-effect model. The number of PCs that were 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-6. The results were summarized and visualized using the same steps as described in Chapter 3.2.4.

For candidate SNP replication analyses, the same list of candidate SNPs presented in Chapter 3.2.4 was used with the exception that some SNPs were excluded from the candidate SNP analyses because they were originally reported in studies that used the same IFS sample as in this study. The post-GWAS procedures including heritability estimation, gene-based and gene set analyses, and TWAS were performed using the same procedures as described in Chapter 3.2.5, 3.2.6, and 3.2.7 using summary statistics from the GWAS of longitudinal DFS.

4.3 Results

4.3.1 Sample summary

Table 6 and Table 7 show the descriptive characteristics of IFS sample at baseline and time-varying variables at the four visits, respectively. Participation rates have remained high with the availability of 333 individuals' DFS data (80%) at age 17. The total DFS, which includes both primary and permanent teeth present at a given time, increased from age 5 to age 9, decreased slightly from age 9 to age 13, and increased rapidly from age 13 to age 17. The decline of total DFS from age 9 to age 13 was expected due to the exfoliation of affected primary teeth and the limited duration of exposure time for newly erupted permanent teeth. The large SD of total DFS at each point indicated the high variability in caries experience among participants.

Table 6 Descriptive characteristics of the IFS sample at baseline

Variables	n (%) or mean ± SD
Sex	
Male	199 (48.07)
Female	215 (51.93)
Household income	
<$20,000/year	47 (11.35)
$20,000-$39,999/year	149 (35.99)
≥$40,000/year	202 (48.79)
Mother's educational attainment	
≤high school	77 (18.60)
high school-some college	153 (36.96)
≥ Four-year college	184 (44.44)
Father's educational attainment	

≤high school	105 (26.38)
high school-some college	120 (30.15)
≥ Four-year college	173 (43.47)
Home water source	
City/public	311 (77.75)
Well	89 (22.25)
Birth weight	3.50 ± 0.58
Gestational weeks	39.41 ± 2.19

Table 7 Summary of time-varying variables in IFS sample

Variables	Age 5 (n (%) or mean ± SD)	Age 9 n (%) or mean ± SD	Age 13 n (%) or mean ± SD	Age 17 n (%) or mean ± SD
Sample size	414	398	380	333
Age	5.14 ± 0.41	9.23 ± 0.73	13.48 ± 0.59	17.72 ± 0.69
Total DFS	2.02 ± 4.86	2.69 ± 3.75	2.16 ± 3.25	6.54 ± 8.33
Home fluoride level	0.78 ± 0.37	0.80 ± 0.37	0.81 ± 0.36	0.71 ± 0.27
Toothbrushing frequency	1.50 ± 0.49	1.55 ± 0.51	1.65 ± 0.51	1.69 ± 0.51
Daily milk intake	12.79 ± 6.10	11.87 ± 6.48	12.46 ± 8.76	10.43 ± 7.87
Daily 100% juice intake	5.69 ± 4.15	2.32 ± 2.44	2.24 ± 2.57	1.36 ± 2.35
Daily SSB intake	4.85 ± 3.96	9.94 ± 6.98	12.40 ± 8.49	14.50 ± 12.46
Daily fluoride intake	0.71 ± 0.35	0.72 ± 0.40	0.85 ± 0.46	0.96 ± 0.50
Daily powdered beverage intake	1.40 ± 3.74	1.59 ± 3.38	1.87 ± 4.45	1.01 ± 3.03

Note: SD, standard deviation; SSB, sugar-sweetened beverages

4.3.2 Risk factor identification

In the univariate linear mixed-effect model, household income, mother's educational attainment, father's educational attainment, home fluoride level, toothbrushing frequency, daily

100% juice intake and daily SSB intake were individually significantly associated with longitudinal DFS ($P<0.05$, Table 7). However, in the multiple variable model, only household income, daily 100% juice intake and daily SSB intake were significantly associated with longitudinal DFS ($P<0.05$, Table 8), and these three risk factors were the only factors that were retained in the final stepwise mixed-effect model (Table 9). In the final stepwise model, higher household income group and higher daily 100% juice intake were significantly associated with lower total DFS ($P=0.0001$ and $P<0.0001$, respectively), while higher daily SSB intake was significantly associated with higher total DFS ($P=0.0052$). These results indicate that household income and daily 100% juice intake may have independent (though not necessarily causal) protective effects on caries development, while daily SSB intake may be an independent risk factor for caries development over time.

Table 8 Univariate linear mixed-effect model of risk factors and longitudinal DFS

Factors	Reference group	Alternative group	Estimate	SE	P
Sex	Male	Female	0.203	0.351	0.563
Household income	<$20,000	$20,000-$39,999	-0.757	0.378	0.045
	<$20,000	≥$40,000	-1.343	0.377	0.0004
Mother's educational attainment	≤high school	Some college	-0.173	0.489	0.724
	≤high school	≥ Four-year college	-1.578	0.473	0.0009
Father's educational attainment	≤high school	Some college	-0.912	0.472	0.054
	≤high school	≥ Four-year college	-1.453	0.436	0.0009
Home water source	Well	City/Public	0.154	0.245	0.531
Home fluoride level	per ppm		-1.202	0.429	0.0052
Toothbrushing frequency	per time		-0.966	0.296	0.0011
Birth weight	per kg		0.467	0.303	0.124
Gestational weeks	per week		0.073	0.081	0.370
Daily milk intake	per oz		-0.034	0.0207	0.098
Daily 100% juice intake	per oz		-0.137	0.037	0.0023
Daily SSB intake	per oz		0.072	0.028	0.0043
Daily fluoride intake	per oz		0.0007	0.338	0.998

Daily powdered beverage intake	per oz		0.053	0.036	0.143

Note: SSB: sugar-sweetened beverages

Table 9 Multiple variable linear mixed-effects model of risk factors and longitudinal DFS

Factors	Reference group	Alternative group	Estimate	SE	P
Sex	Male	Female	0.500	0.323	0.1224
Household income	<$20,000	$20,000-$39,999	-1.107	0.431	0.0105
	<$20,000	≥$40,000	-1.337	0.443	0.0027
Mother's educational attainment	≤high school	Some college	-0.276	0.472	0.5594
	≤high school	≥ Four-year college	-0.929	0.504	0.0661
Father's educational attainment	≤high school	Some college	-0.632	0.440	0.1517
	≤high school	≥ Four-year college	-0.434	0.450	0.3347
Home water source	Well	City/Public	0.004	0.296	0.9887
Home fluoride level	per ppm		0.035	0.448	0.9381
Toothbrushing frequency	per time		-0.450	0.270	0.0961
Birth weight	per kg		0.116	0.348	0.7384
Gestational weeks	per week		0.030	0.093	0.7482

Daily milk intake	per oz	0.009	0.018	0.6241
Daily 100% juice intake	per oz	-0.117	0.039	0.0033
Daily SSB intake	per oz	0.081	0.030	0.0084
Daily fluoride intake	per oz	-0.618	0.388	0.1116
Daily powdered beverage intake	per oz	0.035	0.021	0.0944

Note: SSB: sugar-sweetened beverages

Table 10 Final stepwise model of risk factors and longitudinal DFS

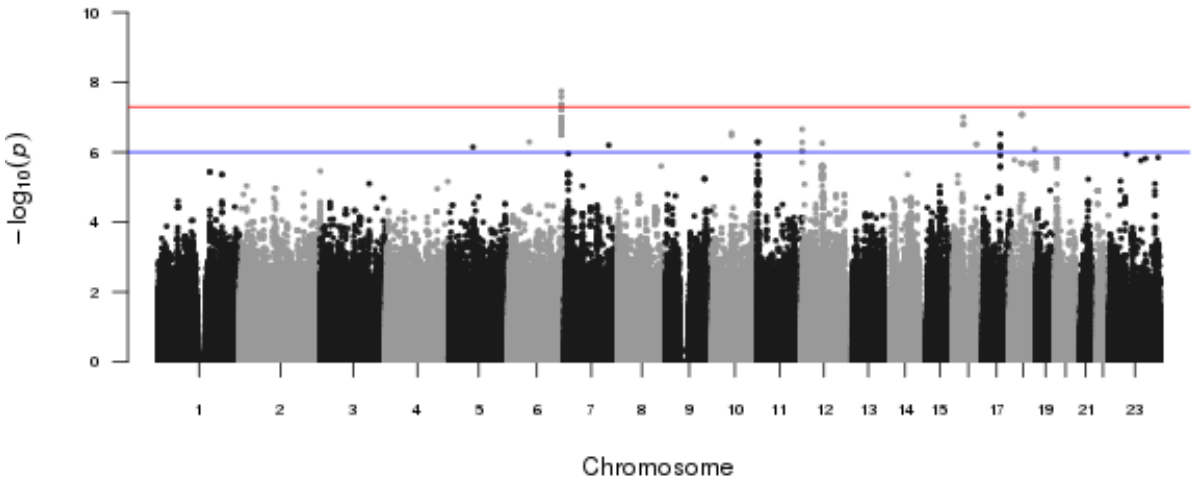
Factors	Reference group	Alternative group	Estimate	SE	P
Household income	<$20,000	$20,000-$39,999	-0.932	0.385	0.0157
	<$20,000	≥$40,000	-1.493	0.386	0.0001
Daily 100% juice intake	per oz		-0.152	0.036	<0.0001
Daily SSB intake	per oz		0.076	0.027	0.0052

Note: SSB: sugar-sweetened beverages

4.3.3 Heritability estimation, longitudinal GWAS and candidate SNP replication

The SNP-based heritability estimate of longitudinal DFS was 0.547 (SD: 0.88) indicating that genetic factors may play an important role in the caries progression over time. Figure 12 shows the Manhattan plot of the longitudinal GWAS results for longitudinal DFS. The genomic inflation factor was 1.01 indicating no evidence of genomic inflation (Appendix Figure 3 shows the Q-Q plot for the longitudinal GWAS). There were 1 signal reaching the genome-wide significance threshold (P<5E-8) and 12 signals at suggestive level (P<1E-6). Table 10 lists the 13 loci and any nearby genes with biological relevance to caries. The regional association plots for these loci are provided in Appendix Figure 4. In particular, some lead SNPs of the association signals were within or near genes with potential functions in dental caries including rs55849333 (P=1.79E-8, located near the gene *SOD2*), rs12311979 (P=2.2E-7, located near the gene *WNK1*), rs7112439 (P=5.09E-7, located near the gene *CTSD*), rs2047597 (P=5.90E-7, located in the gene *WWP2*) and rs819562 (P=7.14E-7, located near the gene *BTF3*). In addition, one lead SNP (rs80225586, P=9.82E-8) was previously reported to be associated with dental caries in prior GWAS (Zeng et al., 2014). Moreover, a locus at the suggestive level (lead SNP: rs184589040 located at chr10:57361698) was about 2Mb from the top locus in the survival GWAS (lead SNP: rs35324031 located at chr10: 55529375) from Aim 1, and the only gene located near this region is *PCDH15* (chr10:55580860-56561051).

Figure 12 Manhattan plot for the longitudinal GWAS of DFS at 4 time points in IFS



The red line indicates the genome-wide significance threshold (P = 5E-8), and the blue line indicates the suggestive significance threshold (P=1E-6). The y-axis is the -log10-transformed P value, the x-axis is the physical position of each variant organized by chromosome.

Table 11 Longitudinal GWAS results for loci showing genome-wide and suggestive evidence of association with longitudinal caries

SNP CHR BP	EA/RA MAF	P	SCORE (VAR)	Gene nearby and biological relevance
rs55849333 6 159942091	T/C 0.077	1.79E-08	-12.75 (5.13)	***SOD2***: Upregulated in periodontal infection as protective mechanism against exaggerated apoptosis; associated with dental fluorosis; higher expressed in periodontitis patients
rs4800033 18 36193503	G/A 0.058	8.38E-08	-10.37 (3.74)	NA
rs80225586 16 30406410	T/C 0.052	9.82E-08	-9.43 (3.13)	*SEPT1, ZNF48, MYLPF, ZNF771, TBC1D10B, CD2BP2, DCTPP1, SEPHS2, ITGAL, ANF768, ANF747, ANF764, ANF688, ANF785, NPIPB13, SLX1A, SULT1A3* **This locus is the top signal in GWAS of SM.dfs in the primary dentition**
rs12311979	T/C	2.20E-07	-14.23 (7.54)	***WNK1***: regulates HCO3- flux in saliva which affects caries

12 839290	0.144			*NINJ2, B4GALNT3, RAD52*
rs184589040	T/C	2.82E-07	-9.69 (3.56)	*MTRNR2L5, PCDH15*
10 57361698	0.057			**This locus is near the top signal in survival GWAS**
rs73390786	A/G	3.00E-07	-8.98 (3.07)	*KIF2B*
17 52087047	0.052			
rs80077770 6 62656352	T/G 0.057	5.08E-07	-9.24 (3.38)	*KHDRBS2*
rs7112439 11 1633277	T/C 0.087	5.09E-07	-11.16 (4.94)	***CTSD***: one of the proteases in saliva, anti-inflammatory activity involved in dentine modification and destruction of peri-implant connective tissue KRTAP5-1~6, IFITM10, CTSD, DUSP8, MOB2, BRSK2
rs1377613 12 61922878	G/T 0.260	5.50E-07	-16.88 (11.35)	*TAFA2*
rs2047597 16 69907848	G/A 0.071	5.90E-07	-10.32 (4.26)	***WWP2***: enriched in ligase activity forming carbon nitrogen bonds (GO term) associated with dental caries *CLEC18A, NOB1, NQO1, NFAT5*
rs275836 7 132177444	C/G 0.056	6.28E-07	-9.02 (3.27)	*PLXNA4*
rs819562 5 72818408	T/C 0.054	7.14E-07	-8.97 (3.27)	***BTF3***: associated with caries experience and harder enamel, enamel more resistant to demineralization, and enamel that more efficiently regain mineral and remineralize; detected expression in whole saliva *FOXD1, ANKRA2, UTP15, ARHGEF28*
rs9966477 18 75101231	G/A 0.072	8.25E-07	-9.98 (4.10)	*GALR1*

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

In the candidate SNP replication analysis, 109 SNPs from previous association studies were tested. There were no SNPs showing significant association with longitudinal DFS (Table 11).

Table 12 Candidate SNP replication results for longitudinal DFS

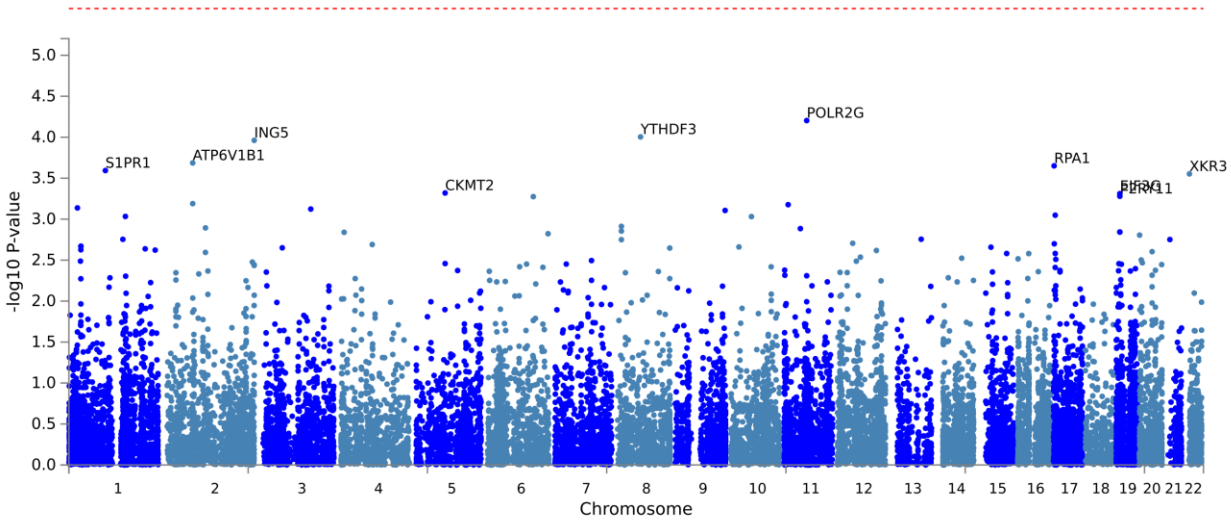
Study	Number of studies	Number of tested SNPs	Number of significant SNPs	SNP with smallest P value (P value)	Gene (Trait)
GWASs	4	22	0	rs1122171 (0.0046)	PITX-AS1 (DMFS)
Linkage studies & follow-up fine-mapping	5	37	0	rs17074565 (0.0056)	13q31.1 (High vs. Low)
Candidate gene studies	18	50	0	rs3759129 (0.009)	AQP5 (Binary)

Note: DMFS: decayed, missing, and filling surfaces in the permanent dentition; High vs. Low means high caries experience vs. low caries experience; Binary means presence or absence of caries. Some of the candidate SNPs tested in survival GWAS were excluded in this replication due to the same study sample.

4.3.4 Gene-based and gene set enrichment analyses

In the gene-based genetic association analysis, the SNPs from the longitudinal GWAS were mapped to 18,592 protein coding genes. The genome-wide significance threshold for the gene-based analysis was P<2.69E-6 (0.05/18,592). There were no genes reaching the genome-wide significance threshold (Figure 13). There was no significant enrichment in the gene set enrichment analyses.

Figure 13 Manhattan plot for the gene-based analysis of longitudinal DFS using MAGMA (labeling top ten genes)

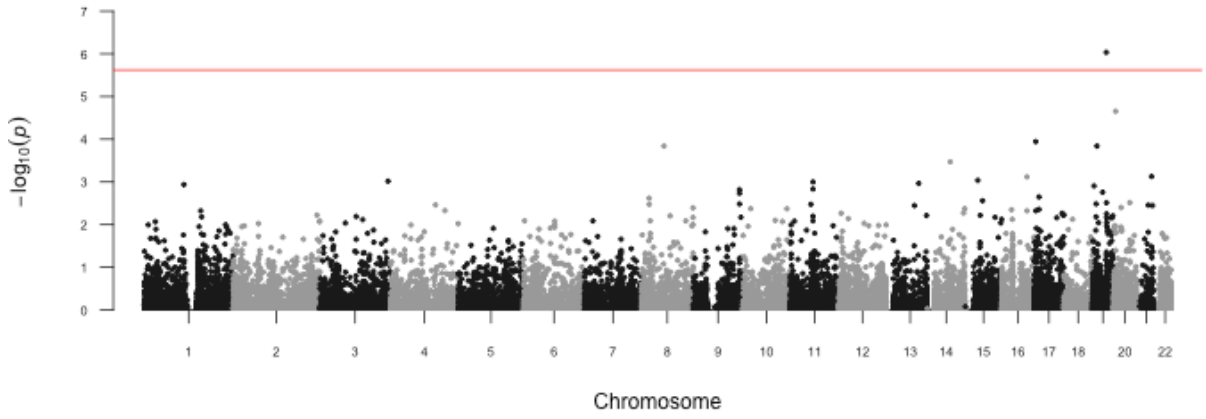


The dotted red line indicates the genome-wide significance threshold (P=2.74E-6). The y-axis is the -log10(P value), the x-axis is genes organized by physical position on each chromosome.

4.3.5 TWAS

In single-tissue TWAS, there was no genome-wide significant gene expression and no study-wide significant genes associated with longitudinal DFS. However, there was one genome-wide significant gene *EIF3K* (P=9.20E-7) associated with longitudinal DFS in the multiple-tissue TWAS (Figure 14). *EIF3K* encodes subunit K of eukaryotic initiation factor 3 (eIF3), which contains at least 12 subunits and plays an important role in protein synthesis by binding directly to the 40S ribosomal subunit and promoting formation of the 43 S preinitiation complex. Previous studies have shown that eIF3k may be involved in cancer development. The expression of eIF3K increased at late stage of nasopharyngeal carcinoma comparing to the early stage, which provides evidence for therapeutic target. The function of EIF3K in dental caries is not clear.

Figure 14 Manhattan plot for multiple tissue TWAS of longitudinal DFS



The red line indicates the genome-wide significance threshold (P = 2.4E-6). The y-axis is the -log10-transformed P value, the x-axis is the physical position of each gene organized by chromosome.

4.4 Discussion

In this study, we identified household income, daily 100% juice intake and daily SSB intake were associated with longitudinal caries and conduced the first longitudinal GWAS on dental caries to investigate genetic effects on caries development over time using mixed-effect models in a longitudinal cohort.

In Chapter 3, household income was found to be associated with caries incidence in young children from the COHRA2 cohort. Here, we provided further evidence that household income in childhood was also significantly associated with caries development over time, this time in the IFS cohort. As discussed in Chapter 3.4, income is commonly used as an indicator for socioeconomic status (SES). A few longitudinal studies have reported the significant association between SES/income and dental caries in different cohorts (Östberg & Petzold, 2020; Peres et al., 2018; Peres, Peres, de Barros, & Victora, 2007; Peres et al., 2011; Thomson et al., 2004; John J. Warren

et al., 2009). For example, Peres et al. used data from a birth cohort that followed participants from age 6 to age 18 years and modeled longitudinal family income into different trajectories including stable high, stable low, downward and upward. They found that stable high income trajectory was associated with lower caries increase (Peres et al., 2018). Another life-course study showed that children who grew up in low SES families had a threefold increase in caries level in adult life (Poulton et al., 2002). A 3-year period longitudinal study with 259,448 participants in western Sweden demonstrated that persistently low SES was associated with higher risk of new and accumulated caries (Östberg & Petzold, 2020). Thus, SES is an important risk factor that influences not only caries onset, but also caries development over time.

Beverage intake, especially sugar-sweetened beverages intake, is another important risk factor for dental caries. In this study, we identified daily SSB intake was a risk factor for caries development over time, while daily 100% juice intake had a protective effect on longitudinal caries. The mechanism of sugar affecting dental caries onset and development, and SSB as a risk factor for dental caries are discussed in Chapter 2. Here we focus on the effect of 100% fruit juice on dental caries. To date, the association between 100% juice and dental caries is not conclusive. A systematic review of 100% fruit juice and dental health found that prospective cohort studies on caries incidence showed either no association or an inverse association, and the results of case-control studies on caries and fruit juice were also mixed (Liska, Kelley, & Mah, 2019). 100% juice that contains natural sugars has been often combined with other SSB in caries risk assessment. However, 100% juice provides other nutrients from the fruit, which may benefit dental health. Also, 100% juice intake increases acute saliva production, which is a protective factor for caries. Further, the American Academy of Pediatrics recommended that 100% fruit juice as a part of

healthy diet (Heyman & Abrams, 2017). Our findings support that 100% fruit juice may function differently than SSB in dental caries.

Here we also present the first longitudinal GWAS to date for caries development over time. The SNP-based heritability estimate of longitudinal DFS was 0.547 (SD: 0.88), which is similar to the heritability estimates of caries incidence in Chapter 3 and slightly higher than the heritability estimates of the incremental increase in caries scores (30%) from a longitudinal twin study (Bretz, Corby, Schork, et al., 2005). The heritability estimate of longitudinal DFS reflects the large genetic component to caries development over time. However, the SD of the heritability estimate here (0.88) is even larger than the SD of heritability estimate of caries incidence (0.41), which may not represent the accurate heritability due to the small sample size. Similarly, future studies with larger sample size are needed to explore genetic contributions for caries development.

There were 1 locus reaching genome-wide significance threshold (P<5E-8) and 12 loci at suggestive significance level (5E-8<P<1E-6). Among the 13 identified loci, 5 loci are located within or near genes with potential functions in dental caries (*SOD2*, *WNK1*, *CTSD*, *WWP2*, and *BTF3*). *SOD2* encodes mitochondrial superoxide dismutase that is the major antioxidant enzyme involved in homeostasis of reactive oxygen species (ROS) under inflammatory conditions. *Streptoccocus mutans* (*S. mutans*) is one of the major bacteria for caries development and *SOD2* is enriched during the aerobic growth of *S. mutans* upon oxygen exposure (De Vendittis et al., 2010). In addition, previous studies showed that *SOD2* expression was upregulated in periodontitis and absence of *SOD2* increased inflammasome components (Yoon, Kim, Lee, & Kim, 2018), and *SOD2* had a protective effect on apoptosis (Rath-Deschner et al., 2021). The SNPs in *SOD2* have also been reported interact with fluoride exposure in dental fluorosis (Du et al., 2021). Given the involvement of *SOD2* in *S. mutans* growth, periodontitis and dental fluorosis, further studies are

needed to verify the function of *SOD2* in dental caries development. *WNK1* encodes a member of the WNK family of protein kinases. Under low Cl^- level, *WNK1* is activated leading to an increase of HCO_3^- concentration in salivary gland fluid (M. G. Lee, Ohana, Park, Yang, & Muallem, 2012). A high HCO_3^- concentration results in high salivary pH that prevents dental caries and facilitates the lubrication of chewed food (Shin et al., 2020). *CTSD* encodes one of the proteases in saliva, which is a collagen-degrading protein involving in dentine modification with aging and destruction of peri-implant connective tissue (Jágr, Eckhardt, Pataridis, & Mikšík, 2012; Roediger et al., 2009). *CTSD* expresses functions in tooth structure and soft tissue, but the direct effects of *CTSD* on caries is not clear yet. *WWP2* is a member of ligase activity pathways and influences the degradation of PTEN that is expressed in mouse oral development (Cho et al., 2008). Similarly, the biological functions of *WWP2* in dental caries have not been investigated yet, but a gene set enrichment analysis of caries GWAS identified that *WWP2* was enriched in gene ontology (GO) terms that were significantly associated with dental caries. Those GO terms included ligase activity forming carbon nitrogen bonds, ubiquitin protein ligase activity and small conjugating protein ligase activity (Zeng et al., 2013). There are several studies providing evidence of *BTF3* in caries pathogenesis. A genome-wide linkage scan identified that the interval 5q12.1-5q13.3 were linked to low caries susceptibility and a following fine-mapping study of this region found that *BTF3* was significantly associated with DMFT in the Filipino cohort. This association was also successfully replicated in other independent populations. In addition, the expression of *BTF3* was detected in the whole saliva and was associated with caries experience (Shimizu et al., 2013). A study of further exploration of *BTF3* function in dental caries showed that SNPs in *BTF3* were associated with enamel hardness, resistance of enamel to demineralization, efficiency of enamel to remineralization (Vieira et al., 2017). Notably, the locus at chromosome 10 (the 5^{th} signal in

longitudinal GWAS) is located near the top signal in the survival GWAS in Chapter 3. The only gene located near this region is *PCDH15*. Investigations of the biological function of *PCDH15* in dental caries are needed because this region shows significant associations with both caries incidence and caries development over time in different populations. Another interesting finding was that one identified loci (the 3rd signal) in this longitudinal GWAS was previously reported to be associated with dental caries in GWASs that used the same participants (IFS cohort data) as part of their study sample (Zeng et al., 2014). Previous GWASs of caries used phenotypes including pit-and-fissure caries (dfs_{PF}) and smooth surface caries (dfs_{SM}), which were collected at a single time point. Our results provided further support that this region at chromosome 16 was not only associated with dfs_{SM}, but also caries development over time.

The major limitation in this study is the small sample size. This study could serve as a pilot study that demonstrates the successful application of longitudinal data in the genetic study of dental caries. Second, this study also lacks external data sets with the necessary data (i.e., longitudinal caries assessments and genotyping data) to replicate the identified variants in this study. Thus, independent studies with larger sample size are needed in the future.

In conclusion, we reported household income, daily 100% juice intake and daily SSB intake as risk factors for caries development and nominated several genes with biologically plausible roles in caries development over time for future investigations. Understanding the risk factors and genetic contributions to caries development may provide further foundation for caries prevention, accurate risk assessment and personalized interventions, thereby improving overall oral health.

5.0 Identifying the genetic variants associated with caries variability in the primary dentition

5.1 Background

This chapter presents the results of genome-wide scan of genetic variants associated with caries variability in the primary dentition and the gene-by-environment interactions (GEI) effects on caries using the identified vQTLs. The main aim of this study is to prioritize variants using genome-wide scans of vQTLs to detect significant GEI associated with dental caries in the primary dentition.

Dental caries is a multifactorial disease. People may be more susceptible or resistant to dental caries when exposed to the same level of extrinsic risk factors due to their genetics. Equivalently, the genetic effects on caries may differ among individuals across environmental strata due to GEI. Though currently little is known about specific GEI effects on dental caries, its complex etiology with factors potentially operating across many mechanisms including tooth structure/anatomy, salivary composition and flow rate, diet, oral hygiene/behaviors, and microbiome/host defense suggests multiple ways through which GEI may impact susceptibility. For example, several lines of evidence suggest that GEI may play critical roles in the genetics of dental caries, including: (1) differences in the heritability of dental caries (e.g., across populations, sexes, and dentitions) (Haworth et al., 2018; Morrison et al., 2016; Shaffer, Wang, et al., 2015; X. Wang et al., 2010); (2) differences in the extrinsic risk factors for dental caries across ancestry groups (i.e., different genetic predisposition to dental caries) (Martignon et al., 2021; Tan, Teo, Tan, & Gao, 2021); and (3) the observation from candidate gene studies and GWAS that significant

associations between genetic variants and dental caries are replicated in some, but not all independent cohorts (Haworth et al., 2018; Morrison et al., 2016; X. Wang et al., 2012).

To date, few studies have investigated GEI for dental caries. Using candidate gene approaches, Shaffer et al. (Shaffer, Carlson, et al., 2015) tested the interaction effects between 4 enamel matrix genes with fluoride exposure on caries experience. They found participants with risk alleles of two genetic variants (upstream of *TUFT1* and missense in *AMBN*) showed higher levels of dental caries only if they lacked exposures to fluoride. Moreover, Yildiz et al. found that the interactions between 3 genetic variants in gene *CA6, DEFB1*, and *TAS2R38* with dental plaque, lactobacilli count, age, and saliva buffer capacity might be associated with caries experience in 154 adults (Yildiz et al., 2016). In addition to candidate gene approaches, a genome-wide GEI study of 709 US children was conducted evaluating the interaction between presence of oral *Streptococcus mutans*, a putative caries-promoting bacteria, and host caries susceptibility (Meng et al., 2019). Three suggestive loci in genes *IL32, GALK2, CELF4* were identified that interacted with *S. mutans*, but no genome-wide significant signals were observed. Prior studies had limited success in explaining GEI effects on dental caries either with narrowly tested genes or environmental factors or limited study sample size.

Efforts to identify GEI in dental caries may aid in discovering additional susceptibility loci for dental caries and identifying populations with high risk of dental caries, thus developing more accurate personalized genetic prediction. However, in general, it is difficult to detect GEI because (1) the accurate record of environmental factors is challenging to collect and (2) doing so requires larger sample sizes compared to detecting main effects due to the potentially small effect sizes of GEI and increased dimensionality (i.e., joint genetic-environment strata) across which the effects are modeled. Moreover, these challenges are exacerbated in the context of GWAS, where millions

of SNPs, and potentially a large number of environments, are tested for GEI effects. An efficient strategy is needed to address these problems.

vQTL analysis provides a method to reduce the scope of GEI testing by prioritizing SNPs mostly likely to be involved in interactions, and can be performed even in the absence of data on the interaction risk factor. The idea is that if a genetic variant is associated with the variance of a quantitative phenotype (i.e., a significant vQTL), then this variant may influence the phenotype by interacting with environmental factors (i.e., GEI). Thus, this approach has two steps: (1) performing a genome-wide scan of variants with vQTL effects on the phenotype; (2) using identified vQTLs to detect GEI associated with the phenotype (Marderstein et al., 2021). This strategy has been successfully applied in previous studies as described in Chapter 2.3.3.

For vQTL analysis, commonly used methods include Bartlett's test (Bartlett & Fowler, 1937), Levene's test (M. B. Brown & Forsythe, 1974), and Fligner-Killeen test (Fligner & Killeen, 1976). Recently, double generalized linear model (DGLM) (Rönnegård, Felleki, Fikse, Mulder, & Strandberg, 2010) and a likelihood-based test (Cao, Wei, Bailey, Kauwe, & Maxwell, 2014) have become available as flexible parametric methods. Previous studies suggested that Levene's test is robust to skewness and kurtosis of the phenotype distribution (Corty & Valdar, 2018). A simulation study reported that no inflation in false-positive rate was observed in Levene's test under the null (i.e., no vQTL effect) regardless of the phenotype distribution or the SNP effect on the phenotypic mean (H. Wang et al., 2019). DMF index (e.g., dfs) in the primary dentition is usually right skewed and zero-inflated. We chose Levene's test to do the vQTL analysis because it is more robust to phenotypic distribution compared to other methods, and even more robust for a version of the test statistic using the group median. The null hypothesis of Levene's test is that variances are equal across groups. The test statistic, W, is defined as:

$$W = \frac{(N-k)}{(k-1)} * \frac{\sum_{i=1}^{k} N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2}$$

N is the total sample size, k is the number of different groups, N_i is the sample size of the i-th subgroup. $Z_{ij} = \left| Y_{ij} - \bar{Y}_{i.} \right|$ Where Y_{ij} is the value of the phenotype for the j-th case from the i-th group, and $\bar{Y}_{i.}$ can be either a mean of the i-th group or a median of the i-th group (i.e., Levene's test with mean or with median). $Z_{i.}$ is the mean of Z_{ij} for group i and $Z_{..}$ is the mean of all Z_{ij}. W is approximately F-distributed with k-1 and N-k degrees of freedom.

5.2 Methods

5.2.1 Study samples, genotyping, and phenotyping

The genome-wide vQTL scans of dfs in the primary dentition were conducted in three cohorts including IFS (n=396), COHRA1 (n=328) and COHRA2 (n=773). Details on the study design, data collection, and genotyping information for COHRA2 and IFS were described above in Chapter 3 and Chapter 4, respectively. COHRA1 is a population-based cohort that recruited 862 northern Appalachian families. Children in COHRA1 were genotyped using Illumina Human610-Quadv1_B BeadChip (Illumina Inc., San Diego, Calif., USA) platform by the Center for Inherited Disease Research at Johns Hopkins University. COHRA1 was imputed using the Michigan Imputation Server (Minimac 4), and HRC r1.1 as the reference panel. Across the three cohorts, 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 < 10%.

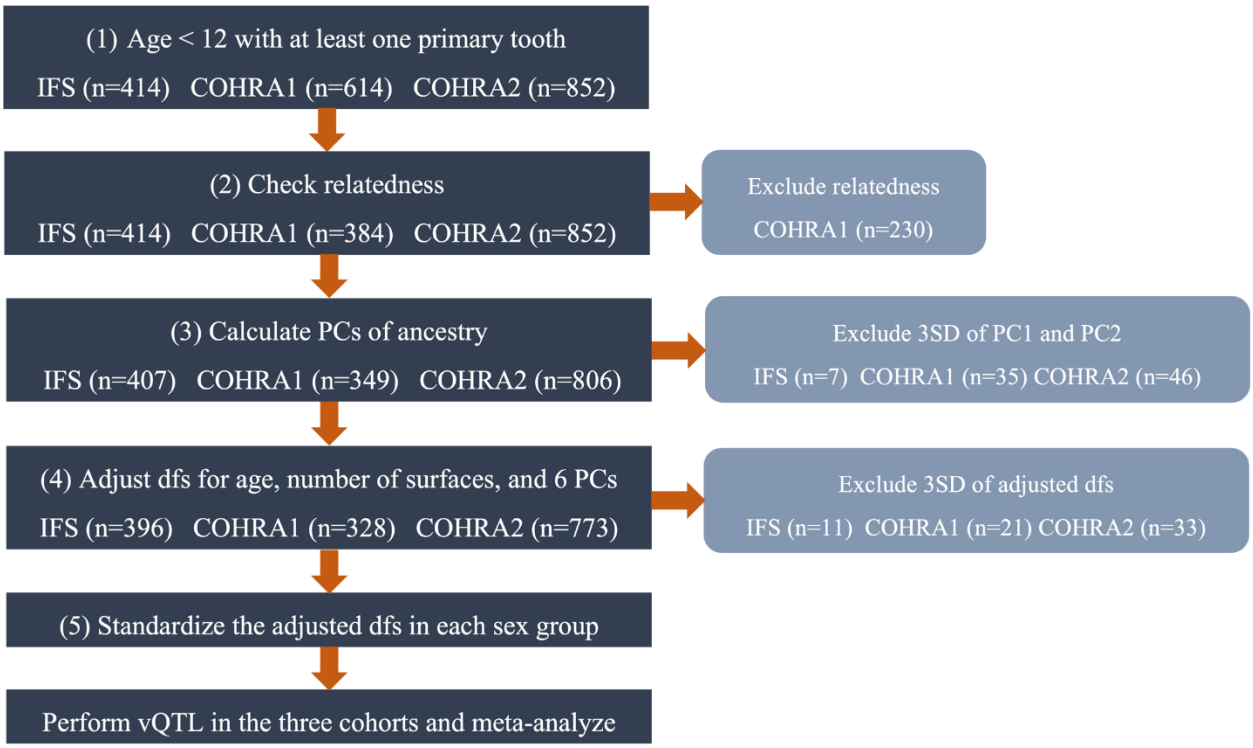
The phenotype across all three cohorts was dfs in the primary dentition, which was obtained through intra-oral examinations. The same participant inclusion criteria were applied across three cohorts including the following: (1) included children had at least one primary tooth and were age < 12 years old; (2) biological relatives were excluded (with one member per kinship retained at random) in the cohort because relatedness will cause inflation in the vQTL results; (3) population structure outliers were excluded based on calculating PCs of ancestry using PCA of genome-wide genetic data and excluding participants with PC1 and PC2 more than 3 SDs from the mean; (4) phenotype outliers were excluded by adjusting the raw phenotype (i.e., dfs) for age, number of surfaces and the first 6 PCs of ancestry (determined by the same procedure in Chapter 3.2.4) and excluding participants with adjusted phenotypes more than 3 SDs from the mean. The adjusted phenotype was standardized to z scores with mean 0 and variance 1 in each sex group. Note, the phenotype processing steps not only removed the effects of age, number of surfaces and the first 6 PCs on the phenotype, but also the differences in mean and variance between the two sex groups. The number of excluded participants in each step was summarized in Figure 15.

5.2.2 Genome-wide vQTL scan and meta-analysis

Levene's test with median as implemented in OSCA software (F. Zhang et al., 2019) was used to test the associations between genetic variants with the variance of dfs in the three cohorts separately. After completing the individual genome-wide vQTL analyses in the three cohorts, we conducted a Stouffer's P-value based meta-analysis using the summary statistics from three cohorts via METAL (Willer, Li, & Abecasis, 2010) in order to increase the statistical power. The threshold for genome-wide significance was P-value<5E-8, and for suggestive significance was P-value<1E-6. The results were summarized and visualized using the same steps as in Chapter 3.2.4.

We prioritized vQTLs with P<1E-6, which were generated from three cohorts and meta-analysis of three cohorts, then used these prioritized vQTL at suggestive level in the following GEI analysis. The workflow for vQTL analysis was shown in Figure 15.

Figure 15 Workflow of genome-wide vQTL scans in three cohorts



5.2.3 Environmental factors in three cohorts

In the following GEI analysis, we tested the interaction effects on dental caries in the primary dentition separately in the three cohorts using prioritized SNPs and environmental factors. For the environmental factors, we only included demographic, behavioral and environmental factors that putatively relate to oral health across the three cohorts. The included environmental factors were different across three cohorts.

In IFS, as described in Chapter 4.2.3, there were 14 factors including sex (male or female), annual household income (<$20,000 or $20,000-39,999 or ≥$40,000), mother's educational attainment (up to high school or some college or above four-year degree), father's educational attainment (up to high school or some college or above four-year degree), water source (City/Public or well), home fluoride level (ppm), daily brushing frequency, birth weight (kg), gestational weeks at birth, daily milk intake (oz), daily 100% juice intake (oz), daily sugar-sweetened beverage (SSB) intake (oz), daily fluoride intake (mg), daily powdered beverage intake (oz).

In COHRA2, besides the 9 factors described in Chapter 3.2.3 including sex (male or female), recruitment site (Pennsylvania [PA] or West Virginia [WV]), annual household income (<$50,000 or ≥$50,000), 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), two versions of composite SSB variables were generated for the included participants, which were added to the list of tested factors (i.e., 11 factors in total) in this study. For COHRA2 children, phone interviews were used to collect intake information on ten beverage types from their parents, which included the frequency of consumption ([1] Never or once; [2] Every few days; [3] Once a day; and [4] Several times a day) and how the beverage was sweetened ([1] Sugar sweetened; [2] Artificial; [3] Unsweetened). The ten beverage types were plant milk, flavored water, sports drink, juice drink, powdered mix beverage, soda, coffee, tea, meal drink, and energy drink. The first composite SSB variable (SSB1) was created in the following three steps: (1) We rescored each beverage type incorporating information on both the consumption frequency and whether it was sweetened. If

sugar sweetened, we retained the consumption frequency (i.e., 1 to 4, as defined above); if not sweetened, we rescored as 1. (2) We defined the composite SSB1 variable as the highest sweetened beverage consumption value reported across the ten beverage types, ranging 1 to 4. (3) If the highest SSB consumption across the ten beverage types was 3 (i.e., "once a day"), and the participant had more than 3 sugar-sweetened beverage types scored as 3, we rescored the SSB1 variable to 4. This step was to calibrate SSB1 because having multiple SSB types once a day could be considered equivalent as having one SSB type several times a day. For analysis, the final SSB1 variable was dichotomized with SSB1=1 as No, and SSB1=2/3/4 as Yes.

The SSB1 variable accounted for both frequency and sweetened type. However, there was a chance that the parents did not report or did not know how the beverage type was sweetened. Therefore, another SSB variable (SSB2) was created using only the consumption frequency information. The process was similar as in SSB1 except without performing step 1.

In COHRA1, there were 10 factors putatively relate to oral health including sex (male or female), recruitment site (Pennsylvania [PA] or West Virginia [WV]), annual mother's income (<$15,000 or ≥$15,000), annual father's income (<$25,000 or ≥$25,000), mother's educational attainment (<high school or ≥high school), father's educational attainment (<high school or ≥high school), water source (water company or well), home fluoride level (ppm), daily brushing frequency, and fluoride supplement (Yes or No).

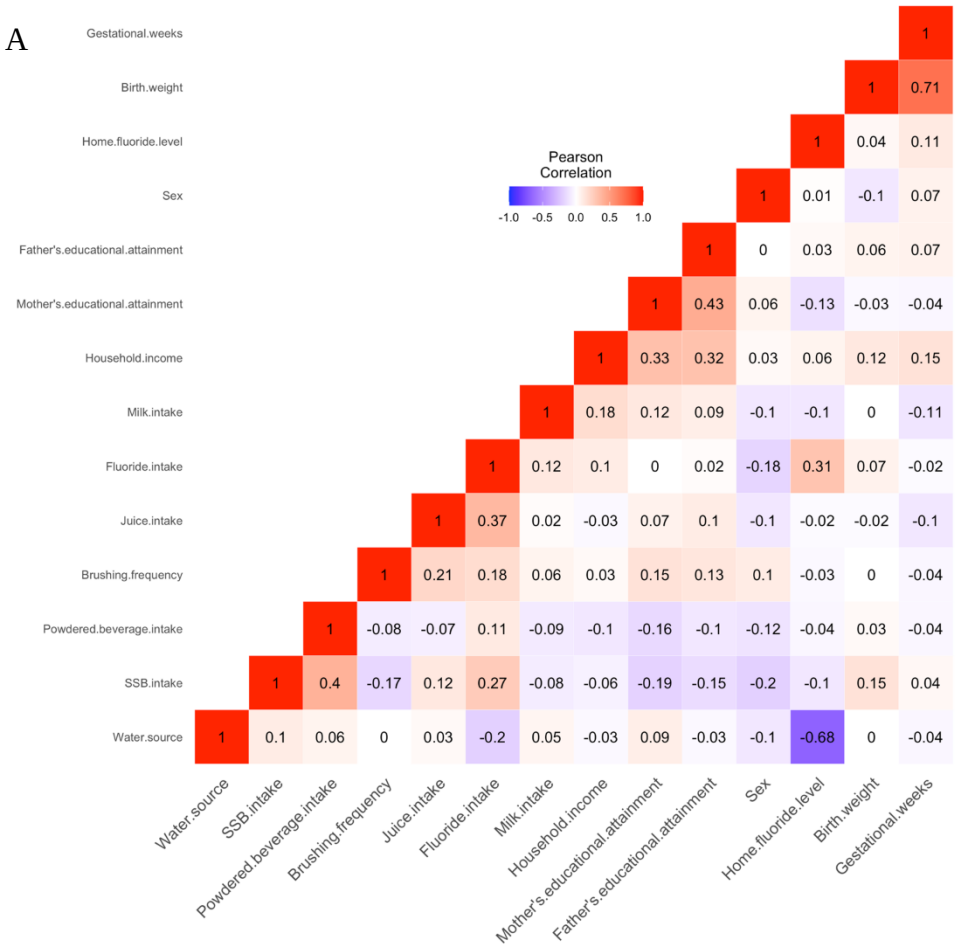
As described in Chapter 3 and Chapter 4, we first checked the correlations among the included factors in the three cohorts separately to make sure that there were no strong correlations. Figure 16 showed the correlations among factors in the three cohorts. There were no strong correlations (defined as correlation coefficient greater than 0.8) in IFS and COHRA1. In COHRA2, SSB1 and SSB2 were strongly correlated (Coefficient=0.84, Figure 15C), which was expected

since the two variables were created from the same underlying beverage consumption responses using similar rubrics. We included both SSB variables in the following GEI analysis in COHRA2.

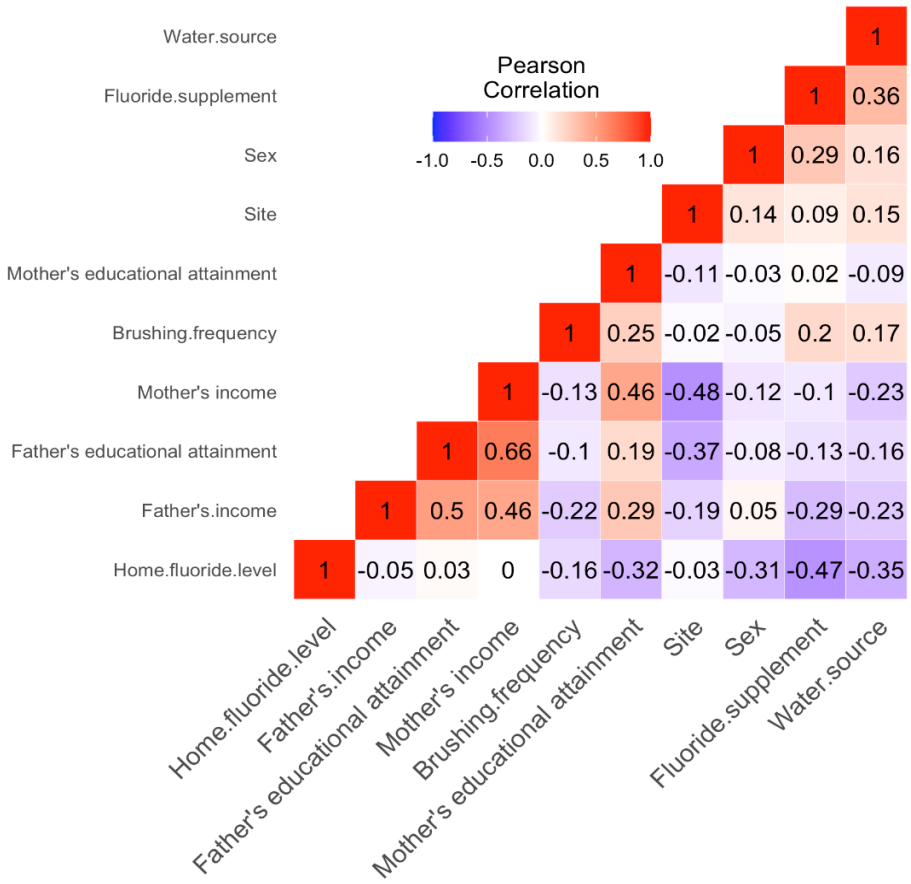
Figure 16 Pearson's correlations among the risk factors in the three cohorts.

A. IFS; B. COHRA1; C. COHRA2

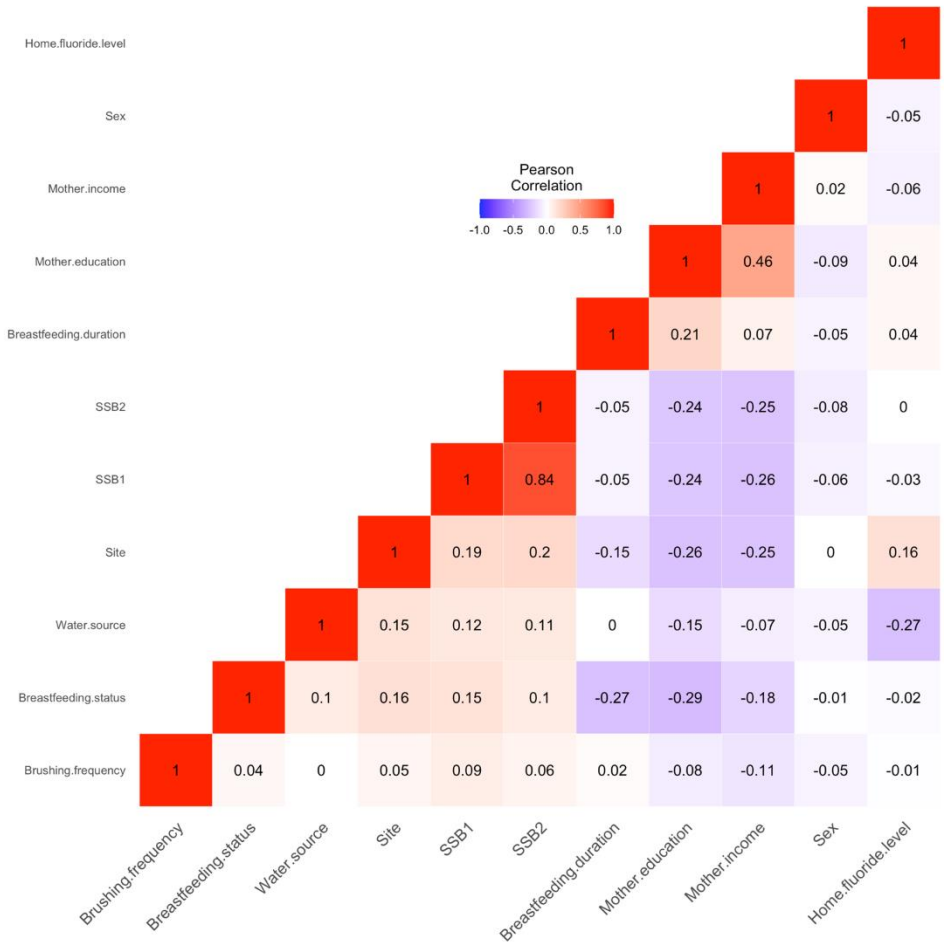
A



B


Pearson Correlation
-1.0 -0.5 0.0 0.5 1.0
Water.source
Fluoride.supplement
Sex
Site
Mother's educational attainment
Brushing.frequency
Mother's income
Father's educational attainment
Father's.income
Home.fluoride.level
Home.fluoride.level
Father's.income
Father's educational attainment
Mother's income
Brushing.frequency
Mother's educational attainment
Site
Sex
Fluoride.supplement
Water.source

C



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).

5.2.4 Gene-by-environment interaction analysis

We used linear regression to individually test the interactions between the prioritized vQTLs and each of the environmental factors in the three cohorts separately. Environmental factors in each cohort were described above. The model for GEI analysis was:

$$y = \mu + \beta_G x_G + \beta_E x_E + \beta_{GE} x_G x_E + e$$

where y is the standardized phenotype, μ is the mean term, x_G is the mean-centered SNP genotype, x_E is the mean-centered environmental factor. βs are the coefficients for the genotype, the environmental factor and the GEI. A stringent Bonferroni correction (i.e., 0.05 divided by the number of SNPs and number of factors tested for GEI) was used to determine the significance thresholds for GEI effects, which were 9.16E-5 (i.e., 0.05/39 SNPs/14 factors) for IFS, 1.28E-4 (i.e., 0.05/39 SNPs/10 factors) for COHRA1 and 1.17E-4 (i.e., 0.05/39 SNPs/11 factors) for COHRA2.

5.3 Results

5.3.1 Sample summary

The summary of environmental factors in the three cohorts is listed in Table 12. After quality control for the participants, a total of 1497 children from three cohorts were included in this study. There were 6 overlapping factors across three cohorts including sex, parental education information, income information, home fluoride level, toothbrushing frequency and water source.

Table 13 Basic characteristics of study sample in the three cohorts

Parameters	IFS (n=396)	COHRA1 (n=328)	COHRA2 (n=773)
Number of tested factors	14	10	11
dfs	1.36 ± 2.81	1.47 ± 3.22	0.68 ± 2.04

Age at examination	5.14 ± 0.41	3.53 ± 1.52	3.69 ± 1.40
Sex (Male vs. Female)	189/207	166/162	405/368
Site (PA vs. WV)	NA	94/234	437/336
Mother's educational attainment	<4-year college: 215 ≥4-year college: 181	≤high school: 133 >high school: 113	<4-year college: 336 ≥4-year college: 434
Father's educational attainment	<4-year college: 214 ≥4-year college: 169	≤high school: 90 >high school: 59	NA
Mother's income	NA	<$15,000: 108 ≥$15,000: 68	<$25,000: 422 ≥$25,000: 339
Father's income	NA	<$25,000: 62 ≥$25,000: 72	NA
Household Income	<$40,000: 185 ≥$40,000: 196	NA	NA
Home fluoride level (ppm)	0.80 ± 0.40	0.66 ± 0.43	0.78 ± 0.29
Brushing frequency (≥2/day vs. <2/day)	205/175	176/124	594/76
Water source (water company vs. well)	297/85	249/62	661/40
Birth weight (kg)	3.50 ± 0.58	NA	NA
Gestational weeks	39.40 ± 2.10	NA	NA
Daily milk intake (oz)	13.00 ± 6.10	NA	NA
Daily 100% juice intake (oz)	5.72 ± 4.17	NA	NA

Daily SSB intake (oz)	4.83 ± 3.92	NA	NA
Daily fluoride intake (oz)	0.71 ± 0.35	NA	NA
Daily powdered beverage intake (oz)	1.37 ± 3.70	NA	NA
Breastfeeding duration (months)	NA	NA	3.92 ± 5.61
Breastfeeding status (Yes vs. No)	NA	NA	643/106
SSB 1	NA	NA	No: 460 Yes: 313
SSB 2	NA	NA	No: 390 Yes: 383

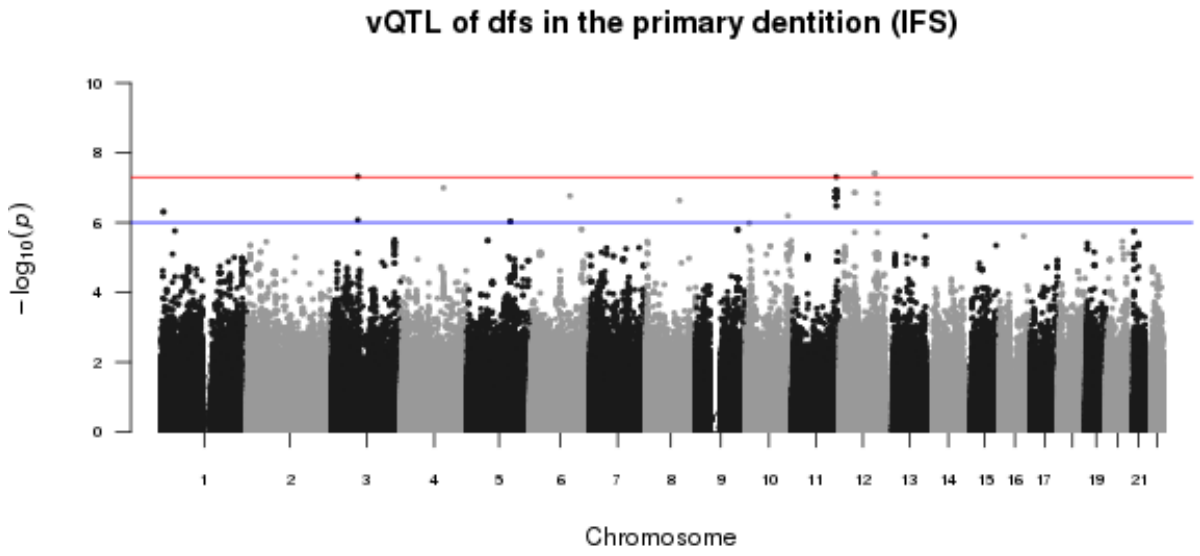
Note: SSB1 and SSB2 are two composite variables for sugar-sweetened beverages. For categorized variables, showing the number of participants in each group; for numeric variables, showing the mean ± SD. NA: not available in the cohort.

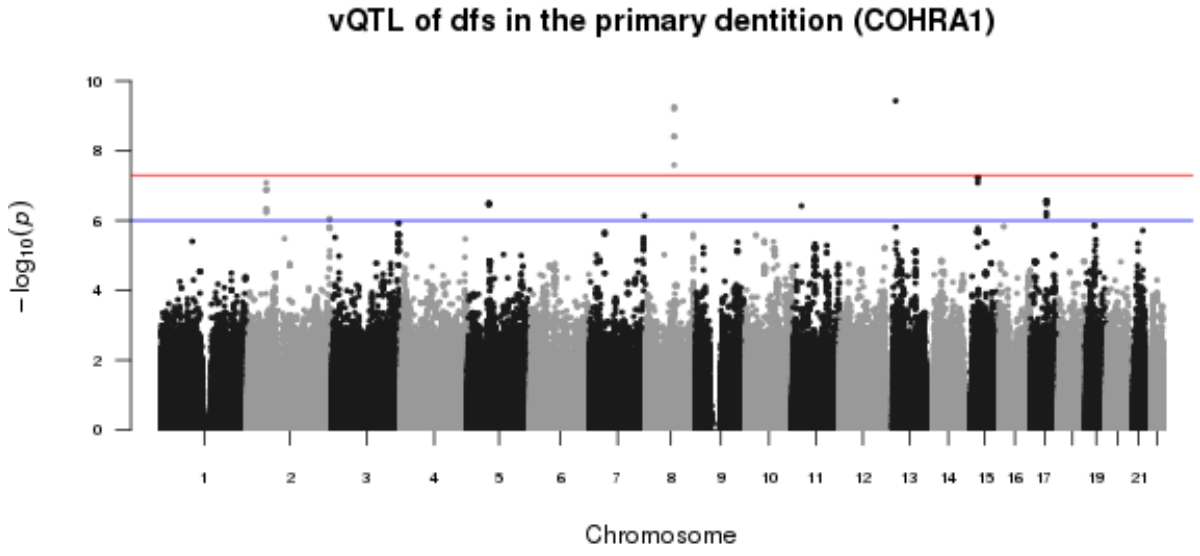
5.3.2 Genome-wide vQTL scan and meta-analysis

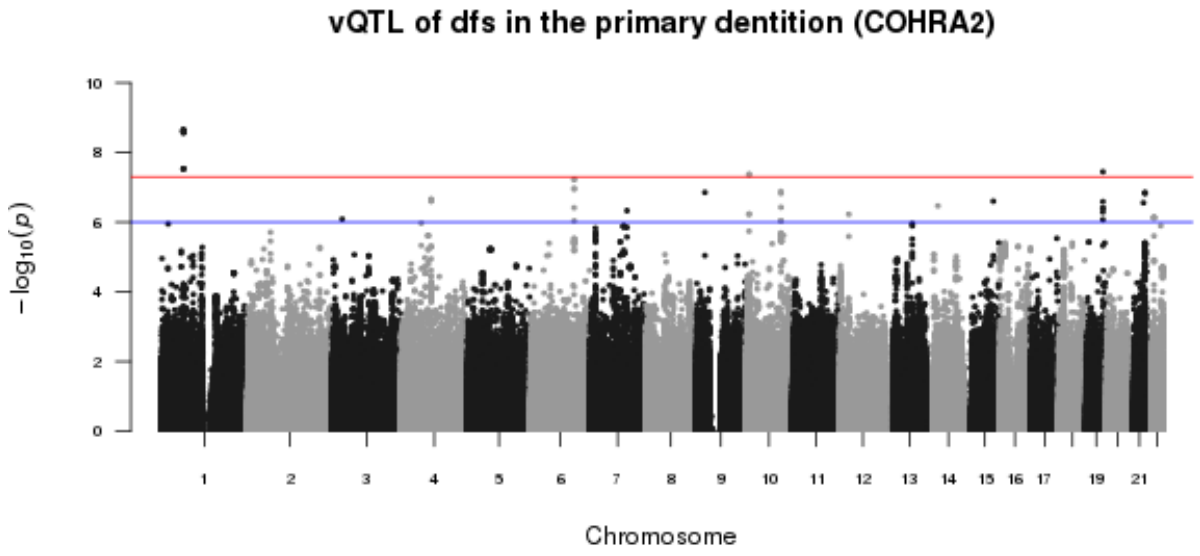
We conducted three genome-wide vQTL scans of dfs in IFS, COHRA1 and COHRA2 using Levene's test and a meta-analysis to combine results of the three cohorts. The Manhattan plots for the vQTL analysis in three cohorts and meta-analysis are shown in Figure 17. The genomic inflation factors were 1.00, 1.02, and 1.03 in IFS, COHRA1 and COHRA2, respectively, indicating there was no genomic inflation in the three cohorts (Appendix Figure 5 shows the Q-Q plot for vQTL scans). There were 3 loci reaching the genome-wide significance threshold (P<5E-

8) in IFS, 2 loci in COHRA1, 3 loci in COHRA2 and 2 loci in the meta-analysis. In addition, there were 8 signals at suggestive level (5E-8<P<1E-6) in IFS, 7 signals in COHRA1, 12 signals in COHRA2 and 3 signals in meta-analysis. In total there were 40 signals, 10 at the genome-wide significance threshold and 30 at the suggestive significance threshold. Notably, rs2090166 (top 1 SNP from COHRA2) and rs3862191 (from the meta-analysis) were in high LD (r^2=0.98). Therefore, a total of 39 independent SNPs were used in the following GEI analysis. Table 13 lists results for the 40 top SNPs from the three cohorts and meta-analysis. Appendix Table 3 shows the top 5 loci in meta-analysis and their results in three cohorts. The regional association plots showing the nearby genes for these loci are provided in Appendix Figure 6.

Figure 17 Manhattan plot for genome-wide vQTL scans of dfs in 3 cohorts and meta-analysis






vQTL of dfs in the primary dentition (COHRA2)
$-\log_{10}(p)$
0
2
4
6
8
10
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
21
Chromosome

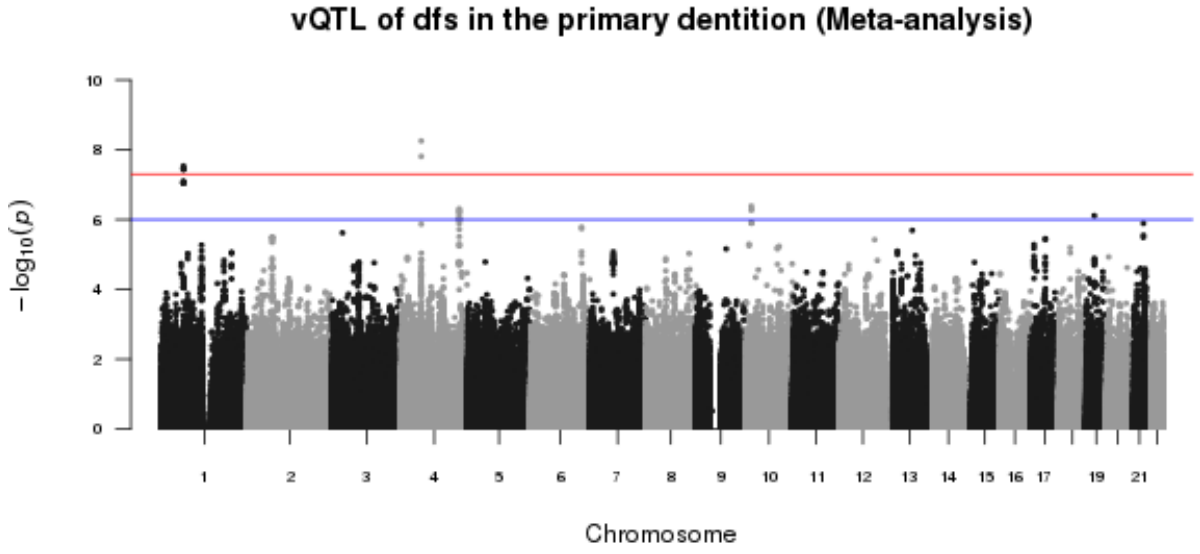

vQTL of dfs in the primary dentition (Meta-analysis)
$-\log_{10}(p)$
0
2
4
6
8
10
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
21
Chromosome

Table 14 40 Top SNPs (P<1E-6) from vQTL of dfs in three cohorts and meta-analysis

No	CHR	SNP	A1	A2	freq	BP	Beta	SE	P	NMISS	Cohort
1	12	rs59190052	A	G	0.23	101640969	0.447	0.08	**3.92E-08**	396	IFS
2	3	rs9830884	C	A	0.29	73612432	0.411	0.08	**4.82E-08**	396	IFS
3	11	rs77322490	T	C	0.19	125701344	0.48	0.09	**4.92E-08**	396	IFS
4	4	rs6844159	C	T	0.3	121794535	0.4	0.08	1.01E-07	396	IFS
5	12	rs3947271	A	C	0.19	43735470	0.458	0.09	1.39E-07	396	IFS
6	12	rs1089941	T	G	0.16	108759443	0.491	0.09	1.47E-07	396	IFS
7	6	rs1491071	C	T	0.22	113566803	0.436	0.08	1.73E-07	396	IFS
8	8	rs2018981	T	A	0.13	98266984	0.535	0.1	2.32E-07	396	IFS
9	1	rs11587481	G	T	0.19	7611159	0.443	0.09	4.93E-07	396	IFS
10	10	rs11199332	A	G	0.11	122186330	0.557	0.11	6.38E-07	396	IFS
11	5	rs11241707	A	G	0.45	123005420	-0.34	0.07	9.34E-07	396	IFS
12	13	rs12429729	A	C	0.1	27568586	0.769	0.12	**3.68E-10**	325	COHRA1
13	8	rs7463853	G	A	0.23	82662694	0.547	0.09	**5.48E-10**	326	COHRA1
14	15	rs690435	C	T	0.31	41132310	0.438	0.08	5.78E-08	328	COHRA1
15	2	rs12994450	C	T	0.29	53649620	0.44	0.08	8.26E-08	328	COHRA1
16	17	rs11654217	G	T	0.23	45531884	0.462	0.09	2.73E-07	328	COHRA1
17	5	rs264532	C	T	0.14	61624327	-0.55	0.11	3.33E-07	328	COHRA1
18	11	rs12797571	G	C	0.51	25880475	0.382	0.08	3.80E-07	328	COHRA1
19	7	rs11970843	A	G	0.21	155917252	0.497	0.1	7.33E-07	278	COHRA1
20	2	rs4663531	G	A	0.25	235893092	0.429	0.09	8.73E-07	328	COHRA1
21	1	**rs2090166**	C	T	0.14	64750226	0.434	0.07	**2.33E-09**	773	COHRA2
22	19	rs3786738	T	C	0.1	48716187	0.458	0.08	**3.59E-08**	772	COHRA2
23	10	rs11817228	C	T	0.11	10340951	0.44	0.08	**4.22E-08**	771	COHRA2
24	6	rs512158	G	A	0.12	125415660	0.424	0.08	5.82E-08	769	COHRA2
25	10	rs622516	C	T	0.15	102082135	0.374	0.07	1.31E-07	760	COHRA2
26	9	rs71508615	G	A	0.11	24787071	0.423	0.08	1.40E-07	772	COHRA2
27	21	rs9982623	T	C	0.13	47691216	0.392	0.07	1.44E-07	773	COHRA2
28	4	rs2869342	T	C	0.31	86320413	0.281	0.05	2.20E-07	768	COHRA2

29	15	rs17536922	T	G	0.11	85461106	0.418	0.08	2.50E-07	728	COHRA2
30	21	rs10651815	G	C	0.27	43045398	-0.29	0.06	2.77E-07	759	COHRA2
31	14	rs1958016	C	A	0.19	33605479	0.338	0.07	3.42E-07	713	COHRA2
32	7	rs73723358	G	A	0.11	106399401	0.403	0.08	4.68E-07	752	COHRA2
33	12	rs7972868	C	T	0.18	26695988	0.326	0.07	5.96E-07	769	COHRA2
34	22	rs73157913	G	T	0.12	25284280	0.384	0.08	7.39E-07	740	COHRA2
35	3	rs11923408	C	G	0.28	28647903	0.282	0.06	8.14E-07	745	COHRA2
36	4	rs9685188	T	C	0.31	58066166	5.829	NA	**5.57E-09**	1446	Meta
37	1	**rs3862191**	T	C	0.17	64753505	5.547	NA	**2.90E-08**	1486	Meta
38	10	rs11592458	C	G	0.12	17053179	5.064	NA	4.10E-07	1487	Meta
39	4	rs1497945	A	T	0.35	167043469	5.03	NA	4.91E-07	1496	Meta
40	19	rs1978471	T	C	0.18	24024926	4.942	NA	7.74E-07	1489	Meta

Note: freq: frequency of effect allele. rs2090166 and rs3862191 (bolded in SNP column) are in high LD with r^2=0.98; Bolded numbers in P column indicating genome-wide significance.

5.3.3 Gene-by-environment interaction analysis

A total of 39 independent SNPs were prioritized for interaction testing with 14, 10, and 11 available environmental factors in IFS, COHRA1, and COHRA2, respectively. GEI results for each cohort are shown separately in the following section.

5.3.3.1 GEI in IFS

The significance threshold for GEI effects was set as 9.16E-5. Table 14 shows the P values for the interaction terms (SNP * factor) that are associated with standardized dfs. There were 3 significant GEI effects including rs9830884 with household income (P = 4.24E-5), rs1491071 with father's educational attainment (P = 4.25E-6) and rs1978471 with toothbrushing frequency (P = 3.15E-5). Figure 18 shows the form of the interactions between the SNPs and environmental

factors. Children with the CC genotype of rs1491071 had higher standardized dfs than the CT and TT genotypes when their father's educational attainment was less then four-year college. Similarly, children with the TT genotype of rs1967471 exhibited higher dfs than the CT and CC genotypes when their toothbrushing frequency was less than 2 times a day and children with the CC genotype of rs9830883 had higher caries compared to the AC and CC genotypes when their household income was less than $40,000/year.

Table 15 P values of interactions between SNPs and factors associated with dfs in IFS

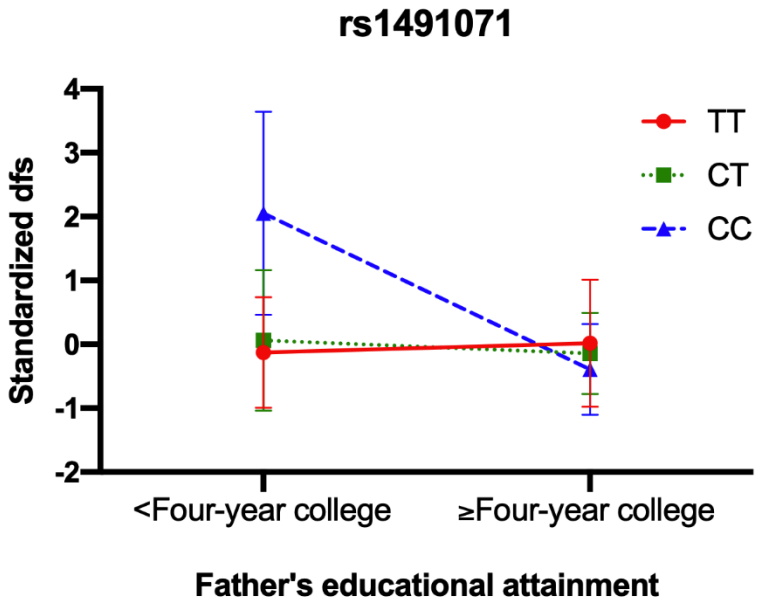
No.	SNP	Sex	Mother edu.	Father edu.	Income	Fluoride level	Brushing	Water source	Birth weight	Gest. weeks	Milk intake	100% juice intake	SSB intake	Fluoride intake	Powder drink intake
1	rs59190052	0.443	0.987	0.113	0.580	0.153	0.355	0.421	0.998	0.370	0.068	0.015	0.040	0.784	0.294
2	rs9830884	0.646	0.360	0.250	**4.24E-05**	0.133	0.277	0.767	0.794	0.709	0.311	0.039	0.148	0.282	0.677
3	rs77322490	0.746	0.051	0.114	0.645	0.936	0.021	0.206	0.708	0.379	0.962	0.497	0.037	0.586	0.854
4	rs6844159	0.851	0.014	0.057	0.051	0.370	0.034	0.050	0.706	0.603	0.275	0.255	0.403	0.401	0.041
5	rs3947271	0.349	0.028	0.028	0.032	0.798	0.052	0.260	0.022	0.448	0.403	0.656	0.475	0.409	0.002
6	rs1089941	0.896	0.058	0.381	0.206	0.654	0.865	0.991	0.580	0.887	0.050	0.179	0.332	0.010	0.199
7	rs1491071	0.326	0.001	**4.25E-06**	0.202	0.991	0.719	0.253	0.416	0.954	0.136	0.813	0.060	0.138	0.195
8	rs2018981	0.092	0.291	0.655	0.605	0.106	0.456	0.143	0.865	0.537	0.081	0.186	0.412	0.160	0.011
9	rs11587481	0.595	0.474	0.495	0.467	0.513	0.493	0.012	0.838	0.402	0.001	0.100	0.638	0.354	0.672
10	rs11199332	0.355	0.081	0.634	0.226	0.202	0.150	0.077	0.874	0.907	0.065	0.605	0.308	0.075	0.773
11	rs11241707	0.263	0.111	0.182	0.022	0.380	0.006	0.366	0.965	0.515	0.377	0.175	0.024	0.100	0.025
12	rs12429729	0.046	0.919	0.913	0.963	0.766	0.070	0.294	0.558	0.525	0.765	0.183	0.741	0.832	0.469
13	rs7463853	0.651	0.074	0.921	0.385	0.292	0.517	0.981	0.207	0.491	0.644	0.816	0.515	0.353	0.330
14	rs690435	0.960	0.134	0.265	0.121	0.989	0.117	0.400	0.083	0.507	0.699	0.486	0.472	0.818	0.036
15	rs12994450	0.108	0.868	0.675	0.453	0.043	0.154	0.122	0.730	0.161	0.519	0.469	0.782	0.936	0.623
16	rs11654217	0.800	0.517	0.790	0.739	0.918	0.750	0.423	0.929	0.147	0.485	0.522	0.873	0.295	0.372
17	rs264532	0.561	0.297	0.589	0.383	0.710	0.326	0.786	0.241	0.552	0.876	0.292	0.600	0.135	0.544
18	rs12797571	0.661	0.619	0.986	0.358	0.145	0.236	0.725	0.443	0.715	0.696	0.445	0.786	0.609	0.273
19	rs11970843	0.859	0.234	0.056	0.179	0.421	0.877	0.346	0.541	0.566	0.861	0.296	0.092	0.627	0.002
20	rs4663531	0.107	0.728	0.054	0.379	0.063	0.372	0.275	0.718	0.257	0.056	0.891	0.640	0.855	0.043
21	rs2090166	0.109	0.270	0.181	0.975	0.303	0.476	0.394	0.695	0.536	0.848	0.731	0.092	0.332	0.034
22	rs3786738	0.629	0.624	0.036	0.137	0.042	0.696	0.940	0.947	0.086	0.337	0.834	0.340	0.900	0.892

23	rs11817228	0.608	0.926	0.139	0.095	0.559	0.652	0.512	0.852	0.814	0.643	0.440	0.353	0.997	0.041
24	rs512158	0.248	0.931	0.961	0.386	0.146	0.473	0.435	0.837	0.660	0.661	0.133	0.872	0.301	0.027
25	rs622516	0.252	0.369	0.601	0.533	0.612	0.845	0.249	0.728	0.943	0.347	0.304	0.342	0.531	0.032
26	rs71508615	0.004	0.984	0.549	0.966	0.788	0.774	0.961	0.039	0.249	0.011	0.904	0.977	0.680	0.594
27	rs9982623	0.083	0.610	0.242	0.414	0.690	0.860	0.204	0.512	0.886	0.902	0.815	0.310	0.847	0.893
28	rs2869342	0.454	0.387	0.384	0.329	0.459	0.306	0.528	0.638	0.186	0.349	0.625	0.542	0.227	0.684
29	rs17536922	0.674	0.260	0.124	0.294	0.083	0.470	0.101	0.520	0.686	0.329	0.112	0.128	0.814	0.554
30	rs10651815	0.177	0.593	0.829	0.514	0.806	0.399	0.567	0.654	0.415	0.003	0.752	0.954	0.687	0.333
31	rs1958016	0.478	0.482	0.517	0.779	0.822	0.739	0.692	0.281	0.037	0.298	0.557	0.343	0.458	0.160
32	rs73723358	0.542	0.735	0.343	0.813	0.156	0.301	0.263	0.915	0.582	0.978	0.301	0.289	0.743	0.501
33	rs7972868	0.771	0.562	0.972	0.287	0.406	0.983	0.980	0.634	0.540	0.890	0.669	0.064	0.977	0.039
34	rs73157913	0.792	0.221	0.611	0.143	0.491	0.898	0.488	0.926	0.743	0.405	0.343	0.217	0.566	0.602
35	rs11923408	0.507	0.657	0.897	0.283	0.220	0.591	0.623	0.730	0.951	0.915	0.685	0.863	0.054	0.225
36	rs9685188	0.672	0.853	0.552	0.408	0.469	0.089	0.010	0.722	0.233	0.736	0.535	0.725	0.498	0.413
37	rs3862191	0.109	0.270	0.181	0.975	0.303	0.476	0.394	0.695	0.536	0.848	0.731	0.092	0.332	0.034
38	rs11592458	0.551	0.792	0.218	0.060	0.662	0.151	0.153	0.633	0.487	0.679	0.532	0.005	0.096	0.842
39	rs1497945	0.183	0.390	0.529	0.203	0.420	0.612	0.541	0.412	0.386	0.138	0.083	0.507	0.190	0.007
40	rs1978471	0.045	0.002	0.002	0.076	0.628	**3.15E-05**	0.325	0.374	0.584	0.057	0.815	0.311	0.502	0.001

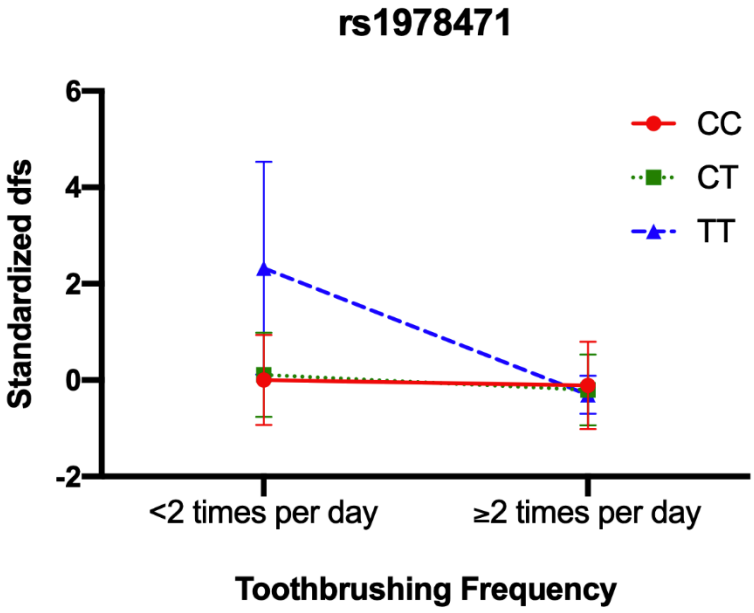
Note: Bolded numbers are the significant GEI. Income: household income; Mother edu.: mother's educational attainment; Father edu.: father's educational attainment; Brushing: Toothbrushing frequency; Gest.weeks: Gestational weeks; Powder drink intake: Powdered beverage intake. The order of the 40 SNPs was the same as in Table 14.

Figure 18 Interaction plots for SNPs and factors in IFS. A. rs1491071 with father's educational attainment. B. rs1978471 with toothbrushing frequency. C. rs9830884 with household income

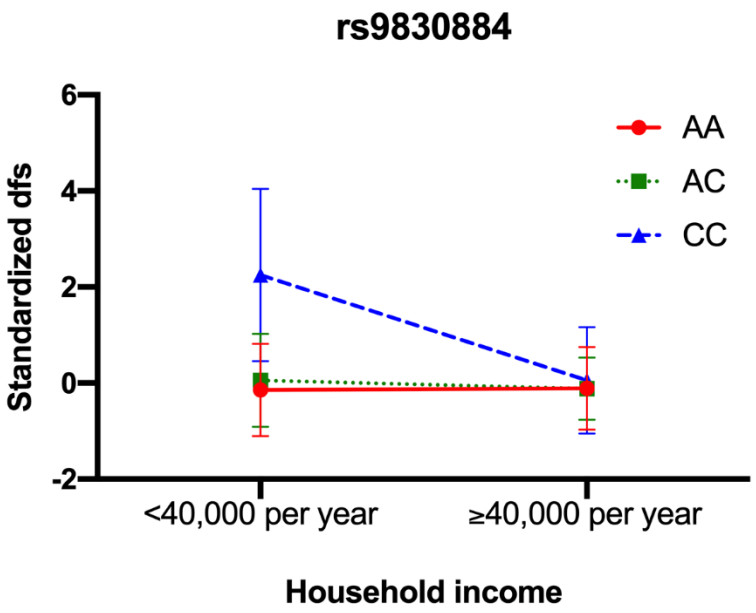
A



B



C



5.3.3.2 GEI in COHRA1

There was 1 significant GEI in COHRA1, which was rs7463853 with father's income (P = 3.50E-5, Table 15). Children with the GG genotype of rs7463853 had higher caries experience in the primary dentition compared to those with the AA and AG genotypes when their father's income was less than $25,000/year (Figure 19).

Table 16 P values of interactions between SNPs and factors associated with dfs in COHRA1

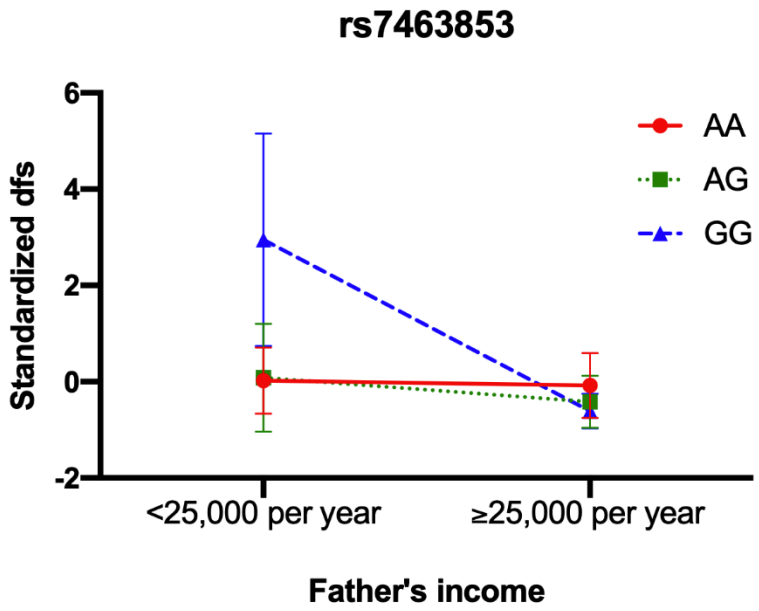
No	SNP	Sex	Site	Mother edu.	Father edu.	Mother income	Father income	Fluoride.level	Brushing	Water Source	Fluoride supplement
1	rs59190052	0.616	0.956	0.182	0.063	0.884	0.956	0.216	0.563	0.851	0.077
2	rs9830884	0.757	0.104	0.560	0.264	0.305	0.354	0.566	0.515	0.126	0.757
3	rs77322490	0.258	0.874	0.092	0.305	0.800	0.776	0.013	0.348	0.731	0.751
4	rs6844159	0.651	0.609	0.542	0.574	0.163	0.876	0.101	0.225	0.697	0.595
5	rs3947271	0.678	0.289	0.108	0.344	0.636	0.886	0.698	0.472	0.565	0.667
6	rs1089941	0.674	0.074	0.410	0.235	0.223	0.773	0.951	0.109	0.412	0.694
7	rs1491071	0.619	0.925	0.641	0.432	0.599	0.483	0.828	0.655	0.537	0.053
8	rs2018981	0.253	0.807	0.781	0.344	0.468	0.104	0.688	0.467	0.711	0.315
9	rs11587481	0.386	0.127	0.222	0.707	0.771	0.866	0.820	0.911	0.568	0.011

10	rs11199332	0.081	0.698	0.476	0.933	0.535	0.936	0.431	0.025	0.649	0.858*
11	rs11241707	0.648	0.724	0.222	0.698	0.593	0.651	0.970	0.234	0.854	0.474
12	rs12429729	0.353	0.004	0.023	0.003	0.008	0.003	0.954	0.030	0.001	0.887
13	rs7463853	0.418	0.021	0.167	3.1E-04	0.033	**3.50E-05**	0.688	0.047	0.135	0.06
14	rs690435	0.936	0.165	0.841	0.062	0.209	0.192	0.434	0.060	0.005	0.156
15	rs12994450	0.312	0.869	0.711	0.177	0.137	0.001	0.093	0.057	0.124	0.091
16	rs11654217	0.634	0.023	0.004	0.083	0.055	0.067	0.897	0.320	0.128	0.217
17	rs264532	0.861	0.384	0.616	0.286	0.901	0.090	0.003	0.005	4.5E-04	0.113*
18	rs12797571	0.296	0.263	0.084	0.016	0.150	0.003	0.943	0.258	0.050	0.737
19	rs11970843	0.514	0.418	0.004	0.482	0.537	0.129	0.925	0.711	0.721	0.042
20	rs4663531	0.021	0.294	0.457	0.262	0.295	0.874	0.102	0.462	0.012	0.802
21	rs2090166	0.330	0.405	0.108	0.811	0.402	0.229	0.054	0.354	0.403	0.716
22	rs3786738	0.877	0.817	0.836	0.806	0.064	0.191	0.744	0.311	0.207	0.824
23	rs11817228	0.528	0.339	0.913	0.952	0.170	0.646	0.704	0.743	0.674	0.589
24	rs512158	0.828	0.921	0.697	0.922	0.597	0.603	0.647	0.965	0.760	0.259
25	rs622516	0.649	0.483	0.670	0.977	0.594	0.424	0.107	0.272	0.236	0.093
26	rs71508615	0.378	0.861	0.866	0.335	0.556	0.555	0.651	0.084	0.295	0.77
27	rs9982623	0.406	0.155	0.494	0.320	0.220	0.677	0.622	0.848	0.307	0.125
28	rs2869342	0.755	0.865	0.517	0.642	0.540	0.928	0.117	0.330	0.293	0.164
29	rs17536922	0.449	0.644	0.788	0.802	0.506	0.243	0.865	0.208	0.346	0.134*
30	rs10651815	0.856	0.681	0.174	0.860	0.821	0.227	0.101	0.181	0.158	0.582
31	rs1958016	0.109	0.924	0.914	0.894	0.704	0.977	0.121	0.909	0.432	0.222
32	rs73723358	0.334	0.343	0.103	0.121	0.768	0.195	0.394	0.951	0.240	0.791*
33	rs7972868	0.714	0.774	0.224	0.468	0.869	0.961	0.666	0.099	0.291	0.348
34	rs73157913	0.472	0.945	0.176	0.687	0.532	0.679	0.157	0.440	0.903	0.946*
35	rs11923408	0.314	0.545	0.505	0.902	0.381	0.570	0.256	0.298	0.545	0.614
36	rs9685188	0.602	0.679	0.444	0.885	0.899	0.791	0.497	0.239	0.328	0.535
37	rs3862191	0.366	0.399	0.103	0.795	0.414	0.219	0.068	0.344	0.413	0.7
38	rs11592458	0.244	0.085	0.685	0.428	0.630	0.396	0.370	0.602	0.306	0.631
39	rs1497945	0.458	0.342	0.590	0.410	0.172	0.161	0.295	0.915	0.267	0.875
40	rs1978471	0.804	0.256	0.236	0.218	0.019	0.078	0.578	0.326	0.021	0.109

Note: Bolded numbers are the significant GEI. Mother edu.: mother's educational attainment; Father edu.: father's educational attainment; Brushing: Toothbrushing frequency. The order of the 40 SNPs was the same as in Table 14.

* There were no participants having two effect alleles of these SNPs and taking fluoride supplement. For these SNPs, we combined genotype groups of two effect alleles and one affect allele, then compared to zero affect allele.

Figure 19 Interaction plot for rs7463853 with father's income



5.3.3.3 GEI in COHRA2

There were 4 significant GEI effects in COHRA2 including rs71508615 with sex (P = 6.78E-6) and mother's income (P = 1.25E-4), rs9685188 with breastfeeding status (P = 8.03E-5), and rs73723358 with SSB1 (P = 7.39E-5). Table 16 lists all the P values of tge interactions that are associated with standardized dfs in COHRA2. Children with the GG genotype of rs71508615 exhibited higher standardized dfs than the AA and AG genotypes when their sex was female, and their mother's income was less than $25,000/year. Similarly, children with the TT genotype of rs9685188 had higher standardized dfs than the CT and CC genotypes when they were not breastfed and children with the GG genotype of rs73723358 experienced higher caries experience

then the AG and AA genotypes when had greater consumption of sugar-sweetened beverages. Table 18 summarizes all the 8 significant GEI from the three cohorts.

Table 17 P values of interactions between SNPs and factors associated with dfs in COHRA2

No	SNP	Sex	Site	Mother edu.	Mother income	Water source	Fluoride level	Brushing	BF.duration	BF.status	SSB1	SSB2
1	rs59190052	0.151	0.327	0.637	0.565	0.211	0.693	0.865	0.884	0.050	0.492	0.438
2	rs9830884	0.104	0.170	0.049	0.281	0.025	0.683	0.853	0.334	0.607	0.997	0.742
3	rs77322490	0.390	0.469	0.862	0.911	0.013	0.684	0.776	0.368	0.858	0.137	0.117
4	rs6844159	0.001	0.275	0.966	0.232	0.395	0.697	0.027	0.285	0.211	0.812	0.815
5	rs3947271	0.042	0.522	0.889	0.907	0.287	0.695	0.444	0.695	0.578	0.091	0.077
6	rs1089941	0.089	0.555	0.061	0.045	0.006*	0.291	0.189	0.793	0.102	0.079	0.338
7	rs1491071	0.978	0.895	0.205	0.144	0.499*	0.685	0.423	0.422	0.172	0.026	0.214
8	rs2018981	0.776	0.729	0.807	0.856	0.713	0.699	0.658	0.832	0.295	0.097	0.157
10	rs11199332	0.627	0.747	0.818	0.287	0.021*	0.692	0.550	0.067	0.942	0.607	0.693
11	rs11241707	0.703	0.928	0.156	0.611	0.097	0.690	0.742	0.262	0.218	0.581	0.631
12	rs12429729	0.499	0.750	0.899	0.807	0.266*	0.694	0.302	0.403	0.541	0.315	0.257
13	rs7463853	0.201	0.263	0.049	0.510	0.107*	0.896	0.337	0.726	0.298	0.729	0.542
14	rs690435	0.102	0.272	0.056	0.650	0.028	0.274	0.004	0.039	0.902	0.043	0.186
15	rs12994450	0.901	0.286	0.599	0.662	0.083	0.689	0.287	0.930	0.330	0.958	0.893
16	rs11654217	0.670	0.239	0.627	0.252	0.761*	0.601	0.335	0.811	0.281	0.661	0.925
17	rs264532	0.350	0.321	0.309	0.734	0.020	0.694	0.359	0.268	0.021	0.128	0.241
18	rs12797571	0.782	0.917	0.063	0.577	0.843	0.400	0.541	0.224	0.146	0.176	0.262
20	rs4663531	0.490	0.749	0.334	0.993	0.191	0.277	0.630	0.689	0.564	0.008	0.062
21	rs2090166	0.418	0.388	0.326	0.916	0.343	0.289	0.569	0.725	0.503	0.050	0.441
22	rs3786738	0.195	0.452	0.111	0.190	0.506	0.689	0.512	0.170	0.004	0.001	0.021
23	rs11817228	0.043	0.735	0.604	0.090	0.850	0.690	0.660	0.040	0.262	0.360	0.345
24	rs512158	0.402	0.185	0.007	0.259	0.010	0.267	0.479	0.628	0.522	0.004	0.002
25	rs622516	0.00037	0.063	0.060	0.013	0.290*	0.683	0.407	0.523	0.058	0.458	0.957
26	rs71508615	**6.78E-06**	0.735	0.00037	**1.25E-04**	0.150	0.678	0.029	0.003	0.041	0.089	0.194

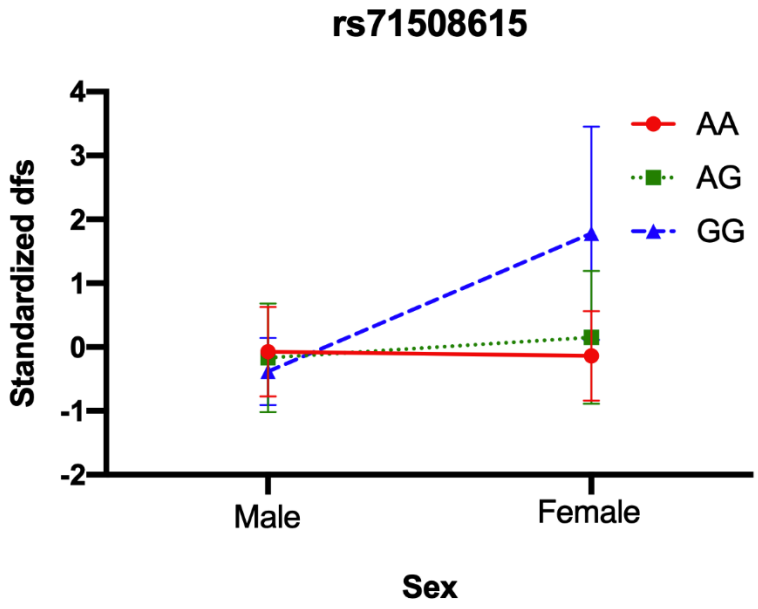
27	rs9982623	0.186	0.077	0.969	0.094	0.173	0.381	0.812	0.289	0.833	0.241	0.180
28	rs2869342	0.751	0.902	0.022	0.600	0.314	0.684	0.801	0.118	2.4E-04	0.817	0.692
29	rs17536922	0.116	0.957	0.390	0.260	0.576	0.676	0.886	0.296	0.511	0.861	0.384
30	rs10651815	0.006	0.371	0.161	0.185	0.193	0.240	0.680	0.686	0.305	0.241	0.292
31	rs1958016	0.648	0.174	0.055	0.060	0.187*	0.682	0.601	0.801	0.608	0.753	0.781
32	rs73723358	0.775	0.003	0.004	0.00014	0.022	0.676	0.461	0.052	0.814	**7.39E-05**	0.047
33	rs7972868	0.025	0.731	0.037	0.004	0.358	0.689	0.835	0.957	0.003	0.070	0.021
34	rs73157913	0.337	0.237	0.223	0.433	0.047*	0.687	0.060	0.787	0.465	0.054	0.019
35	rs11923408	0.354	0.381	0.919	0.925	0.443	0.686	0.657	0.932	0.328	0.098	0.251
36	rs9685188	0.175	0.783	0.001	0.007	0.858	0.686	0.424	0.556	**8.03E-05**	0.00033	0.004
37	rs3862191	0.472	0.394	0.343	0.938	0.343	0.289	0.575	0.801	0.433	0.059	0.510
38	rs11592458	0.025	0.162	0.006	0.123	0.064	0.689	0.684	0.115	0.207	0.614	0.914
39	rs1497945	0.201	0.449	0.027	0.780	0.022	0.405	0.612	0.174	0.302	0.255	0.399
40	rs1978471	0.216	0.513	0.631	0.149	0.751	0.607	0.769	0.153	0.695	0.934	0.832

Note: Bolded numbers are the significant GEI. Mother edu.: mother's educational attainment; Brushing: Toothbrushing frequency; BF.duration: Breastfeeding duration; BF.status: Breastfeeding status; SSB1 and SSB2: sugar-sweetened beverages 1 and 2. The order of the 40 SNPs was the same as in Table 14.

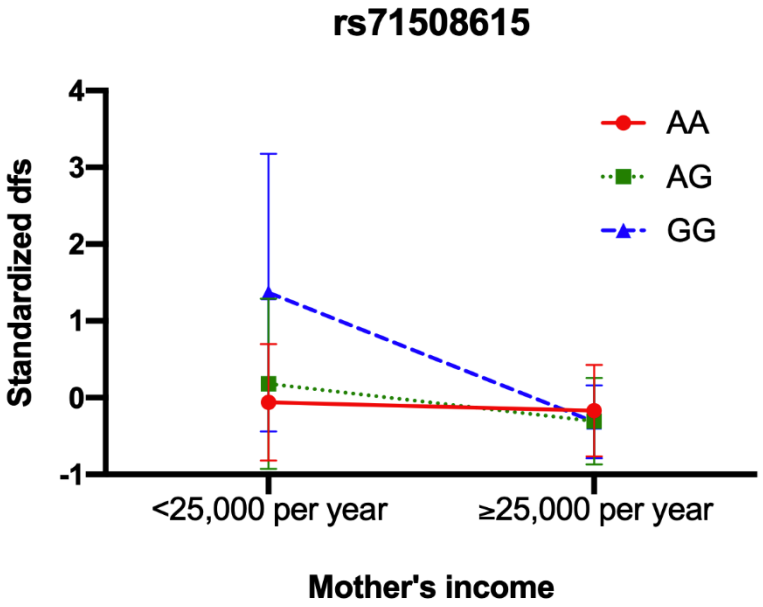
* There were no participants having two effect alleles of these SNPs and with water source from well. For these SNPs, we combined genotype groups of two effect alleles and one affect allele, then compared to zero affect allele.

Figure 20 Interaction plots for SNPs and factors in COHRA2. A. rs71508615 with sex. B. rs71508615 with mother's income. C. rs9685188 with breastfeeding status. D. rs73723358 with SSB1

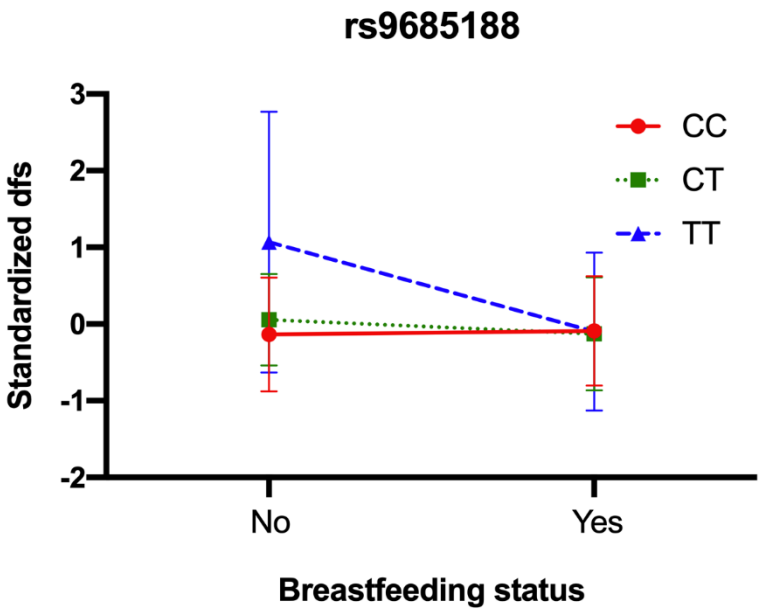
A



B



C


rs9685188
Standardized dfs
3
2
1
0
-1
-2
CC
CT
TT
No
Yes
Breastfeeding status

D

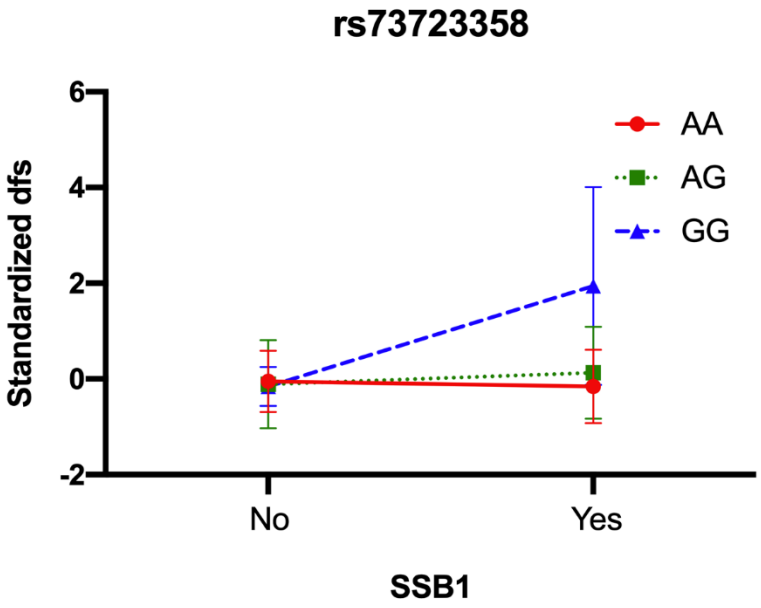

rs73723358
Standardized dfs
6
4
2
0
-2
AA
AG
GG
No
Yes
SSB1

Table 18 Summary of significant GEI from three cohorts

No.	Cohort	SNP	Factor	SNP significance level and cohort	SNP nearby genes
1	IFS	rs1491071	Father's education	Suggestive from IFS	NA
2	IFS	rs1978471	Tooth brushing frequency	Suggestive from Meta	*RPSAP58, ZNF675*
3	IFS	rs9830884	Household income	Genome-wide from IFS	*PDZRN3*
4	COHRA1	rs7463853	Father's income	Genome-wide from COHRA1	*SNX16, IMPA1*
5	COHRA2	rs71508615	Sex	Suggestive from in COHRA2	*IZUMO3*
6	COHRA2	rs71508615	Mother's income	Suggestive from COHRA2	*IZUMO3*
7	COHRA2	rs9685188	Breastfeeding status	Genome-wide from Meta	*IGFBP7*
8	COHRA2	rs73723358	SSB1	Suggestive from COHRA2	*NAMPT, PIK3CG*

Note: NA, intergenic (gene desert).

5.4 Discussion

This study, for the first time to date, explored the genetic effects on the variance of dental caries experience and presented the application of genome-wide vQTL scanning for prioritizing SNPs for GEI detection. Most of the identified loci in this study were only detected through the vQTL scans, which indicates the unique value of studies focusing on phenotypic variance rather than on phenotypic means, whereas one locus was reported to be associated with caries in the permanent dentition previously, which provides greater validity to this identified locus. In addition, using genome-wide vQTL scan to prioritize SNPs, we found several significant GEI between prioritized SNPs and environmental factors.

The genome-wide vQTL scans of dental caries in the three separate cohorts and the meta-analysis of the three cohorts cumulatively identified 39 independent loci that were significantly

associated with caries variance. The top SNPs from the 39 identified vQTL signals were used to test the GEI effects on dental caries in the primary dentition, yielding 8 significant GEI in total across the three cohorts (Table 18). Two of the SNPs from the 8 significant GEI are located in or near the genes that were reported to be involved in dental caries with relatively strong evidence in previous studies. First, the SNP rs73723358 identified as a possible vQTL at the suggestive level (chr7: 106399401, No. 32 in Table 14) in COHRA2 is located near the gene *NAMPT* that was previously reported to be associated with dental caries in the permanent dentition (Morrison et al., 2016) in an Hispanic/Latino population. *NAMPT* is a pro-inflammatory adipokine that is expressed in periodontal tissue (Damanaki et al., 2014) and is involved in periodontal healing and regulating matrix-degrading enzymes (S. Yang et al., 2015). Though the direct role of *NAMPT* in dental caries is not clear and rs73723358 does not have a known eQTL or regulatory effects on *NAMPT*, this SNP interacted with SSB intake influencing caries in the primary dentition in our study. As discussed in Chapter 2, sugar in food and drinks is a well-known risk factor for dental caries and periodontal disease (Lula, Ribeiro, Hugo, Alves, & Silva, 2014), which is used by bacteria as energy and increases the level of produced acids leading to a lower the pH in the dental biofilm and demineralization. Children with the GG genotype of rs73723358 had higher caries experience compared to AG and AA only if they had higher SSB consumption, suggesting that the effect of the genetic variant rs73723358 on caries experience in the primary dentition was mediated by the consumption of SSB (and vice versa).

Second, the SNP rs9685188 was identified as a vQTL from the meta-analysis of the three cohorts at genome-wide significance level (chr4: 58066166, No. 36 in Table 14), and is located near the gene *IGFBP7*, but rs9685188 and its proxies have no known eQTL or regulatory effects on *IGFBP7*. *IGFBP7* encodes a member of the insulin-like growth factor-binding protein family.

IGFBP7 was found to be expressed in whole cytoplasm of odontoblasts and the intercellular space of maturation-stage ameloblasts during tooth germ mineralization. Knock-down of *IGFBP7* expression in rats promoted dentin matrix mineralization suggesting the negative regulation of *IGFBP7* in enamel and dentin formation (Moon et al., 2021). Given the role of *IGFBP7* in tooth development and mineralization, it is plausible that *IGFBP7* impacts caries development through abnormal mineralization. In addition, this SNP influenced caries through significantly interacting with breastfeeding status in our study. Children with the TT genotype of rs9685188 had higher caries scores compared to CT and CC only if they were not breastfed. Breastfeeding is considered as a protective factor for dental caries via various pathways. For example, the antibodies in breastmilk may reduce the growth of cariogenic bacteria (Rugg-Gunn, Roberts, & Wright, 1985). Also, the milk is released into the throat via breastfeeding, as opposed to bottle feeding, during which the milk delivered into the front of the mouth and around the teeth. Our results suggest that the effects of *IGFBP7* on dental caries, potentially through its role in tooth development and mineralization, may be different between breastfed and non-breastfed children. However, the direct relationship between caries and this GEI is not clear yet.

Our findings of the effects of GEI on dental caries provided additional insights into its etiology. First, we identified vQTL signals that might represent novel caries susceptibility loci with effects that were not able to be detected via GWAS. Second, the nominated genes might play roles in the biology of dental caries, expanding the list of putatively caries-associated genes for further investigation. Third, the environmental factors interacting with genetic variants in dental caries expands our knowledge of cariogenesis and could potentially help with identifying genetic subgroups with higher exposure-specific caries risk for prevention efforts (i.e., increased environmental-specific risk for individuals with a certain genotype). For instance, SES might

influence caries risk through living conditions, behavior and dietary patterns. Here, we also detected that SES indicators (e.g., income, education) significantly interacted with some of the genetic variants across all the three cohorts. Our results indicated that people with lower SES might be more susceptible to dental caries when they carry a certain genotype. And same for two other risk factors, toothbrushing frequency and sex; people having lower toothbrushing frequency or being female experience higher caries risk when they have a certain genotype. However, the detailed mechanisms of these identified GEI in dental caries need to be further explored.

Besides the findings on GEI effects on dental caries, the vQTL scans, themselves, identified several signals located in or near genes with plausible biological roles in dental caries. For example, in the IFS cohort, the top 1 SNP at genome-wide significance level (rs59190052, chr12: 101640969; No. 1 in Table 14) has eQTL effects on nearby genes *ANO4* and *UTP20*, and is located near the gene *SLC5A8*, which was reported to be expressed in ameloblasts at the maturation stage and directly involved in enamel mineralization (Jedeon, Loiodice, et al., 2016). Abnormal *SLC5A8* expression disrupted ion and pH homeostasis and inhibited enamel crystal growth, which led to enamel hypomineralization, thus increasing susceptibility to caries development. In addition, *SLC5A8* has been shown to be expressed in epithelial cells and stimulate apoptosis, thus it may also promote apoptosis and autophagy of ameloblast during amelogenesis (Jedeon, Houari, et al., 2016). Likewise, in COHRA1, the vQTL SNP rs11970843, which was identified at the suggestive level (chr7: 155917252; No. 19 in Table 14), is located near the gene *SHH* and has putative regulatory effects on the enhancer of *SHH*, which encodes the Sonic Hedgehog signaling molecule. The Shh signaling pathway plays an essential role in human tooth development from initiation through to root development, which includes setting the boundaries between odontogenic and non-odontogenic epithelium at the initiation stage, and regulates cell

cycle, differentiation, and morphogenesis at the stage of tooth germ establishment. Moreover, recent studies have shown that Shh signaling pathway promotes morphogenetic movement via cell polarization and impacts both cuspal and root development at later stages, and defines tooth number at the maturation stage (Hosoya, Shalehin, Takebe, Shimo, & Irie, 2020). Given the role of *SHH* in tooth development, it is possible that *SHH* influences dental caries through abnormal tooth development. Another possible vQTL, observed at the suggestive level in COHRA2, rs622516 (chr10: 102082135, No. 25 in Table 14), is located in the gene *PKD2L1* and has eQTL effects on *PKD2L1* in testis, frontal cortex and cortex tissues. *PKD2L1* encodes a member of the transient receptor potential family of ion channels and is a potential candidate gene for sour taste that detects acids in foods and drinks, which may influence dietary patterns and potentially influence dental caries. Knockout *PKD2L1* in a mouse model reduces the responses of the gustatory nerve to sour taste stimuli (Horio et al., 2011). Lastly, in the meta-analysis, an identified vQTL at the suggestive level, SNP rs1497945 (chr4: 167043469; No. 39 in Table 14), is located near the gene *TLL1* and has a regulatory feature on open chromatin. *TLL1* encodes a member of the tolloid-like proteinase family. Though the role of *TLL1* in dental caries in not known, *TLL1* was reported to be required for maintaining periodontal homeostasis (J. Wang et al., 2017), and inactivation of *TLL1* led to dental and periodontal defects in mice (H. Zhang, Jani, Liang, Lu, & Qin, 2017).

There are several limitations of this study. The sample sizes are relatively small in the three cohorts, and even in the meta-analysis. SNPs with small effect sizes on dental caries variance might not have been detected due to low power. There were no overlapping vQTLs among the three cohorts (i.e., the identified loci in one cohort were not replicated in other two cohorts even at nominal p-values), even with the same phenotype and similar aged participants, and the same

with the significant GEI. This may also be a consequence of the low power problem. Another explanation is that the vQTL effects and GEI are cohort specific. Participants were recruited from different regions in different decades and were exposed to different levels of environmental factors. Thus, the identified vQTL or GEI in one population may not be generalizable to other populations. Another limitation is that variance heterogeneity can be explained by multiple mechanisms including genuine interaction effects, confounding by phenotypic distribution, and induction by nearby causal variants. We are not able to disentangle the mechanisms and make direct interpretations of the identified vQTLs in each cohort, thus the different mechanisms in vQTLs may also partially explain the fact that there were no overlapping loci where the causal mechanisms were different. After prioritizing the SNPs via the vQTL scans, we only explored the GEI effects on dental caries. Gene-by-gene interactions can also lead to vQTLs, which is out of the scope of this study but is worth to examining in the future. The phenotype we used was dfs, the caries score in the primary dentition in children. We did not test the caries experience in the permanent dentition, which is important and meaningful for the genetics of dental caries and needs to be studied in the future. Another potential limitation is that for some of the vQTLs, we may not have collected data on the truly interacting environmental factor, hence, we were not able to detect the corresponding GEI for those vQTLs. Similarly, there were missing values for some of the environmental factors in COHRA1 (e.g., mother's and father's educational attainment) and unbalanced number of participants in the strata of factors and genotype in COHRA1 and COHRA2, which might cause problems in detecting the accurate effects of GEI on dental caries and replicating the GEI effects in other cohorts.

In conclusion, we conducted the first genome-wide vQTL scan of dental caries and identified 39 independent loci that were associated with caries variance. Some of the identified

loci are located in or near genes with possible functions in dental caries. Using these prioritized SNPs, we detected several environmental factors that significantly interacted with the vQTL SNPs to influence dental caries experience, including income, education, toothbrushing frequency, sex, breastfeeding, and SSB consumption. Our results expanded the understanding of the genetic basis of dental caries.

6.0 Conclusion

6.1 Summary

This dissertation investigated the genetic basis of dental caries incidence, development over time and variance by leveraging longitudinal data and advanced statistical methods. In Chapter 3, we used survival analysis to investigate the effects of risk factors and genetic variants on dental caries incidence. Income, education and water source were found to be associated with dental caries incidence. We highlighted the large contribution of genetic factors to dental caries incidence and nominated several novel genes with potential roles in dental caries for further investigation. Similar approaches were applied in Chapter 4, except we used repeated dental caries measurements at multiple time points as the phenotype and mixed-effects models to test the associations of risk factors and genetic variants with caries development. Income was also one of the significant risk factors for caries development, along with 100% juice and SSB intake. We also found a high heritability estimate for caries development and several genes with promising evidence of functions related to dental caries development or susceptibility. Notably, one locus at the suggestive level was near the top locus from the survival GWAS presented in Chapter 3. In Chapter 5, we performed the first genome-wide vQTL scan of dental caries. We not only identified novel loci through the vQTL scan, but also detected significant GEI for dental caries using the prioritized variants. Remarkably, income and education were also significantly interacting with genetic variants to impact caries experience. Our findings provided further support for the value of testing variance effects and the usefulness of our prioritization strategy in detecting GEI for dental caries.

6.2 Significance

This dissertation filled current gaps in the knowledge of the genetic basis of dental caries in three aspects: (1) by showing the large contribution of genetic factors to dental caries incidence and nominating several novel genes for future studies; (2) by demonstrating the genetic contribution to dental caries development over time and nominating several novel genes with promising evidence of biological functions on dental caries; (3) for the first time exanimating the genetic effects on caries variability and identifying GEI that were associated with dental caries. Our results expanded the current understanding of the genetic architecture of dental caries, which will provide a foundation for further investigations into the mechanisms of dental caries onset and progression. In addition, we also explored the potential environmental risk factors for caries incidence and development over time. The combined understanding of the genetic and environmental factors for dental caries, which was expanded in this work, may ultimately help with better early detection, risk assessment, dental care, and effective public health interventions.

Appendix A Supplementary materials for Chapter 2

Appendix Table 1 GWASs of dental caries and identified genes

Study	Sample	Phenotype	Gene
(Shaffer et al., 2011)	1305 white children	Binary (Yes/No affected)	*ACTN2, EDARADD, EPHA7, LPO, MPPED2, MTR, TFIP11, ZMPSTE24*
(X. Wang et al., 2012)	1483 white adults	DMFS	*CNIH1, ISL1, PTK2B, RPS6KA2*
	5960 older adults	Proportion of affected surfaces, self-report	*FZD1, RHOU*
	7443 white adults	DMFS, Proportion of affected surfaces, self-reported	*ADAMTS3*
(Zeng et al., 2013)	1004 white adults	Smooth DMFS	*8q21.3, BCORL1, CXCR1, CXCR2*
		Pit-and-fissure DMFS	*BCOR, INHBA*
(Shaffer, Feingold, Wang, Lee, et al., 2013)	920 white adults	Cluster-based partial DMFS	*LYZL2*, AJAP1*, ABCG2, EDNRA, IFT88, IL17D, NKX2-3, NR4A3, PKD2, SCPP family, SMAD7, TGFBR1, TWSG1*
(J. Yang et al., 2014)	1483 white adults	DMFS	*CNTN5, COL4A2*
(Zeng et al., 2014)	1006 white children	Smooth dfs	*AJAP1, ITGAL, RPS6KA2, PLUNC family*
	979 white children	Pit-and-fissure dfs	*KPNA4*, MPPED2*
(Morrison et al., 2016)	11754 Hispanic/Latino adults	DMFT	*SYPL1*, NAMPT*, BMP7*, SPO11**
		DMFS	*IGSF10*, AADACL2*, SYPL1*, NAMPT**

(Ballantine et al., 2018)	212 children	ECC	*CLDN14*
(Govil et al., 2018)	464 families, 2616 individuals	PRIM (binary)	*KAT2B*
		QTOT1	*CACNA1E, LAMC2, ALMS1, STAMBP, GXYLT2, SLC12A2, MEGF10, IL family, TMEM181, ARID1B, NLRP2, NLRP7, NLRP, KIR, LILR families*
		QTOT2	*BCL11A*
(Haworth et al., 2018)	19003 children	Binary (Yes/No affected)	*ALLC**
	13353 adults	Binary (Yes/No affected)	*NEDD9**
(Orlova et al., 2019)	109 adults	DMFT	*PTGER3, ZRANB2, MIR186, TAF1B, GRHL1, KLF11, PDIA6, PGAP1, HES1, LSG1, MIR1305, IQGAP2, S100Z, CRHBP, F2R, NDUFA4, THSD7A, MIR148, SNX10, IFRD1, PTCH1, SLC5A12, CCDC34, LGR4, APOBR, SH2B1, PLCG2, CDH13, CTSB, AIRE, TRPM2, TSPEAR*
		DMFS	*ARHGEF16, TAF1B, GRHL1, KLF11, PDIA6, HES1, LSG1, IQGAP2, S100Z, SNORA47, CRHBP, F2R, MIR148, FKBP14, NOD1, HS3ST4, POLD1, ACPT, KLK1, KLK4, SIGLEC9, CD33*
	96 children	dft	*CD1D, CD1A, CD1C, PYHIN1, RARB, LAPR7, MIR302A, RUFY1, MAML1, GRM6, ADAMTS2, MIR6874, HPVC1, VSTM2A, EGFR, ANGPT2, DEF family, MIR346, FAT3, MTNR1B, IGF1, CLDN10, HS6ST3, SMAD6, SMAD3, MAP2K5,*

		dfs	*MEF2A, IGF1R, SLC9A3R1, CHST8, KCTD15, NLRP5, ZNF582* *RARB, RBMS3, IRX4, IRX2, TRAF3IP2, MIR346, FAT3, MTNR1B, CD300E, SLC9A3R1, CHST8, KCTD15*
(Shungin et al., 2019)	26792 adults	DMFS	*C5orf66*, CA12*, KRTCAP2, WNT10A, FGF10, HLA, FOXL1, PBX3, MAMSTR*

Note:

DMFT/DMFS: Decayed, missing, and filled teeth/surfaces

dft/s: decayed, filled deciduous teeth/surfaces

ECC: early childhood caries defined as the presence of one or more decayed (non-cavitated or cavitated lesions), missing (due to caries), or filled tooth surfaces in any primary tooth in a child under the age of six.

PRIM: dichotomized as zero versus one or more affected primary teeth

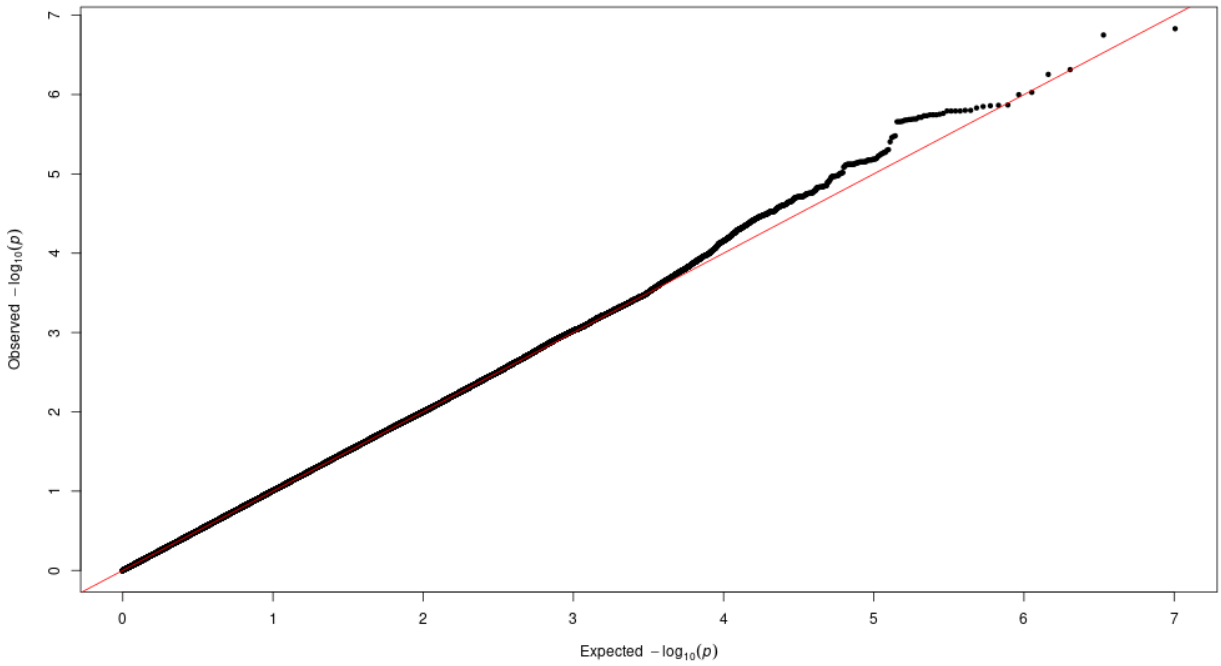
QTOT1: age-adjusted quantitative caries measure for both primary and permanent dentitions including pre-cavitated lesions

QTOT2: age-adjusted quantitative caries excluding pre-cavitated lesions

*: genome-wide significant associations (P-value ≤ 5E-8)

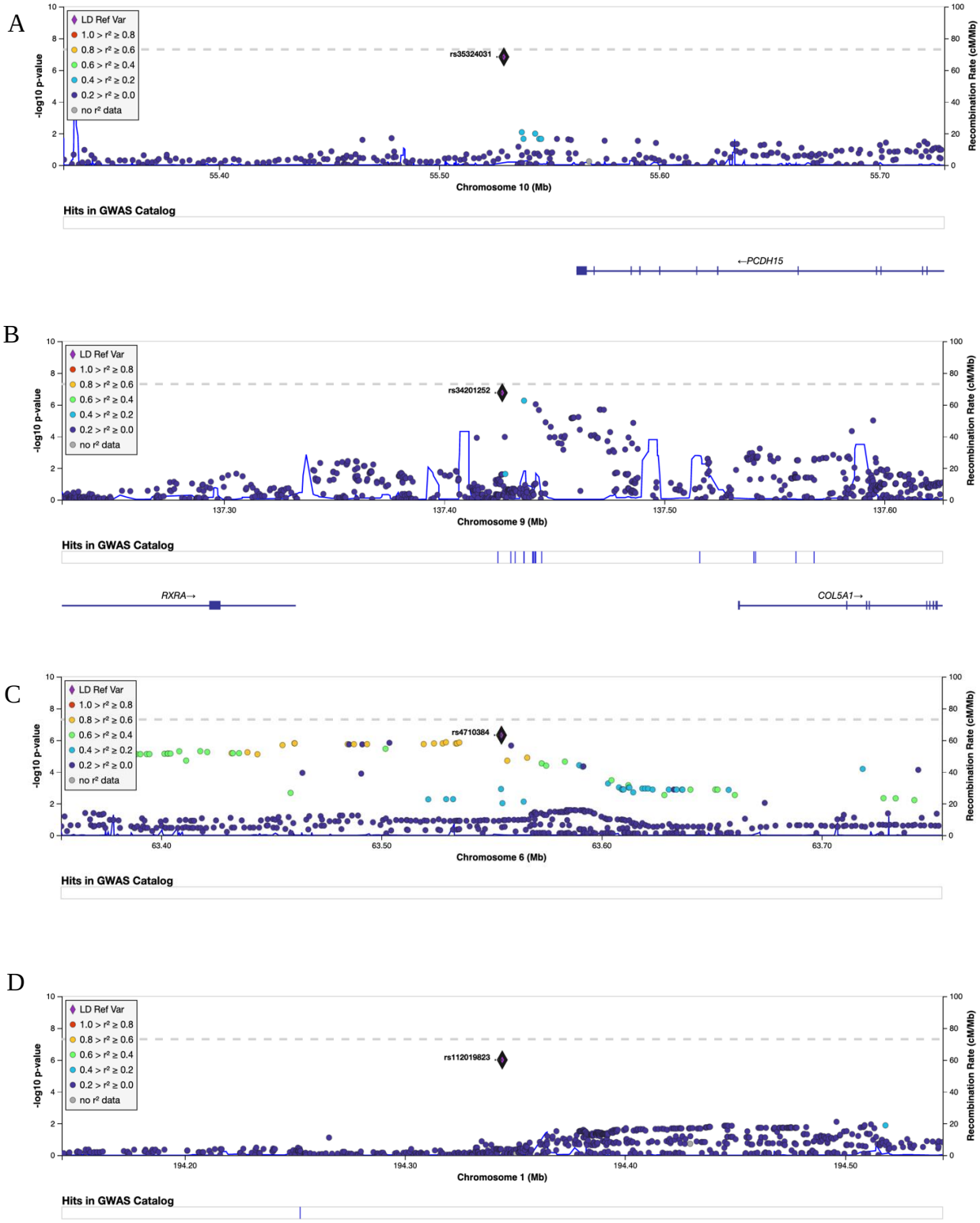
Appendix B Supplementary materials for Chapter 3

Appendix Figure 1 Quantile-quantile plot for the survival GWAS of time to first caries incidence

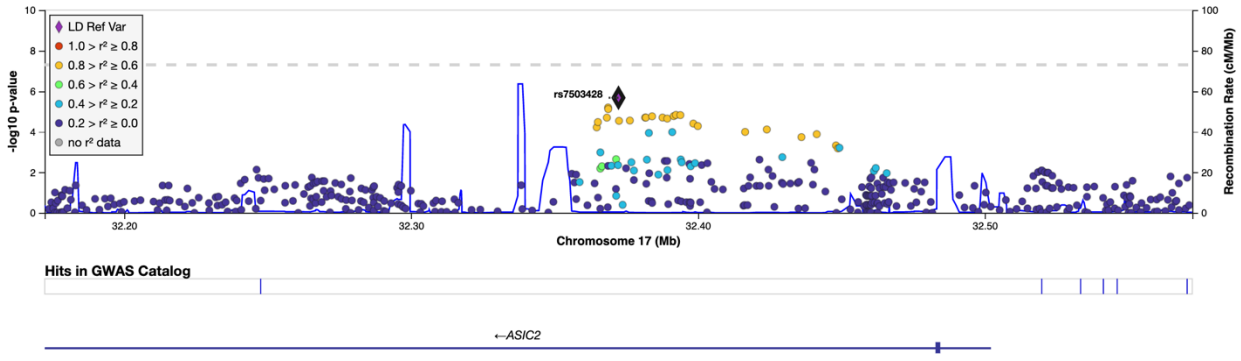


The genomic inflation factor was 1.00.

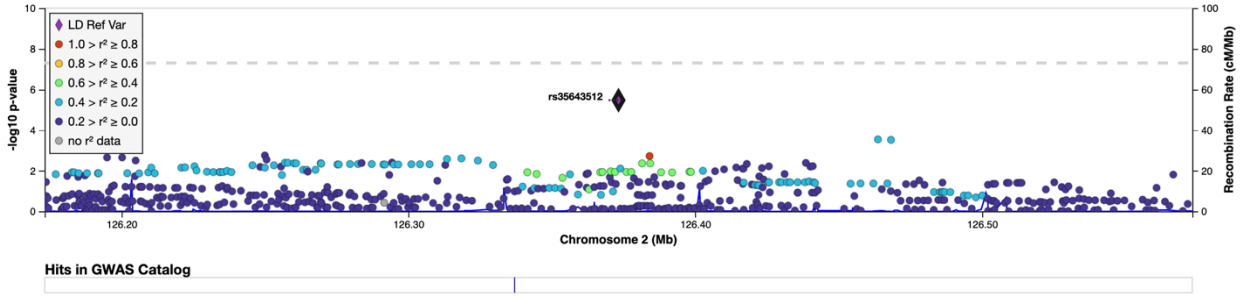
Appendix Figure 2 Regional association plots for the 14 lead SNPs at suggestive loci



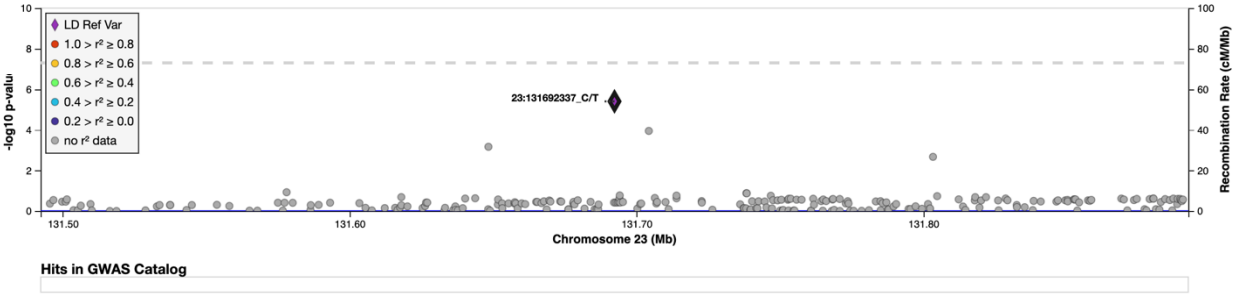
E


LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
rs7503428
-log10 p-value
Recombination Rate (cM/Mb)
32.20
32.30
32.40
32.50
Chromosome 17 (Mb)
Hits in GWAS Catalog
←ASIC2

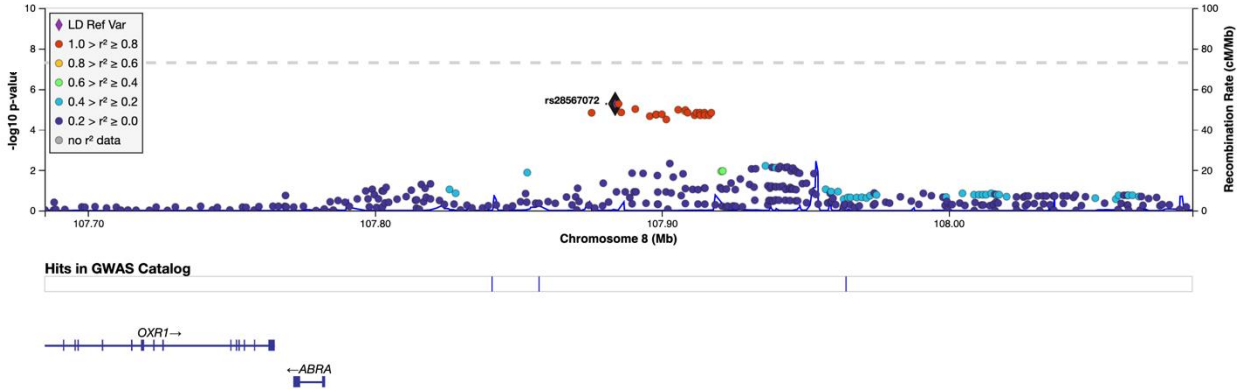
F


LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
rs35643512
-log10 p-value
Recombination Rate (cM/Mb)
126.20
126.30
126.40
126.50
Chromosome 2 (Mb)
Hits in GWAS Catalog

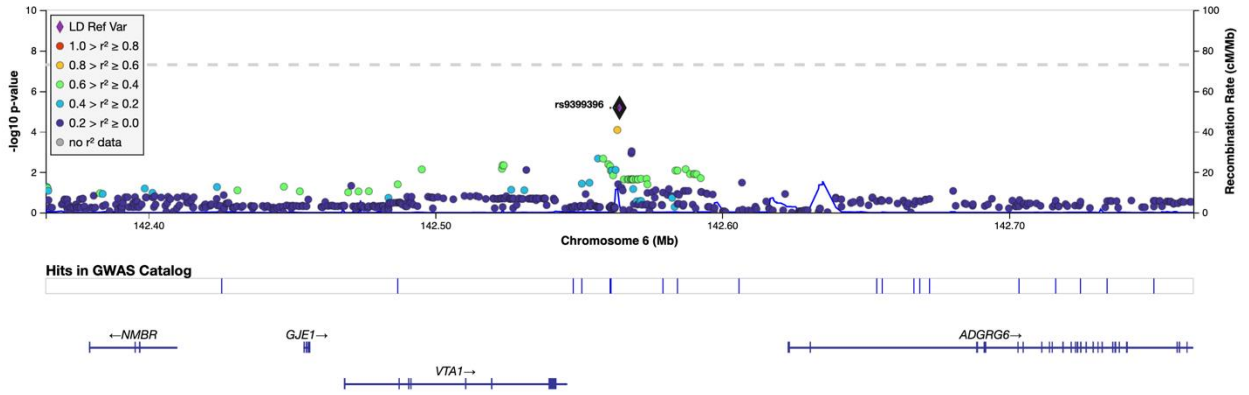
G


LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
23:131692337_C/T
Recombination Rate (cM/Mb)
131.50
131.60
131.70
131.80
Chromosome 23 (Mb)
Hits in GWAS Catalog

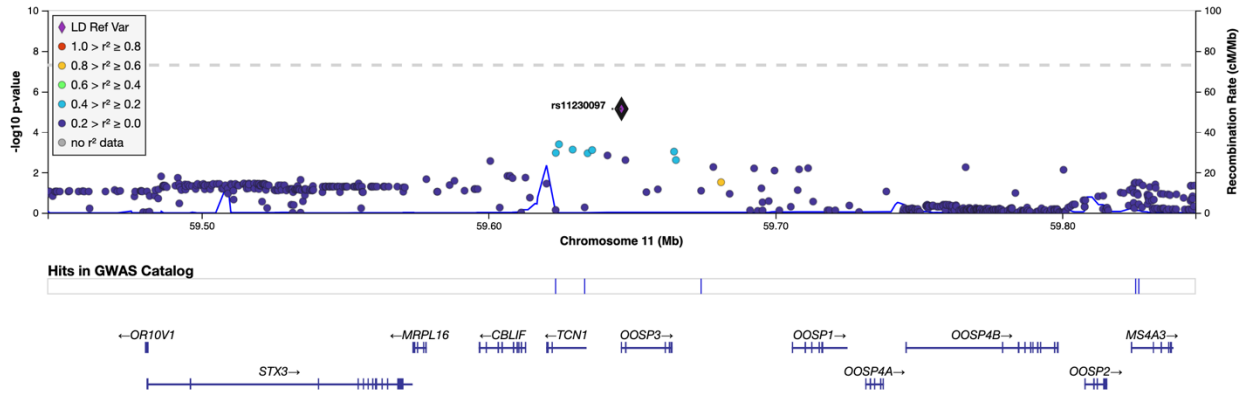
H


LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
rs28567072
Recombination Rate (cM/Mb)
107.70
107.80
107.90
108.00
Chromosome 8 (Mb)
Hits in GWAS Catalog
OXR1→
←ABRA

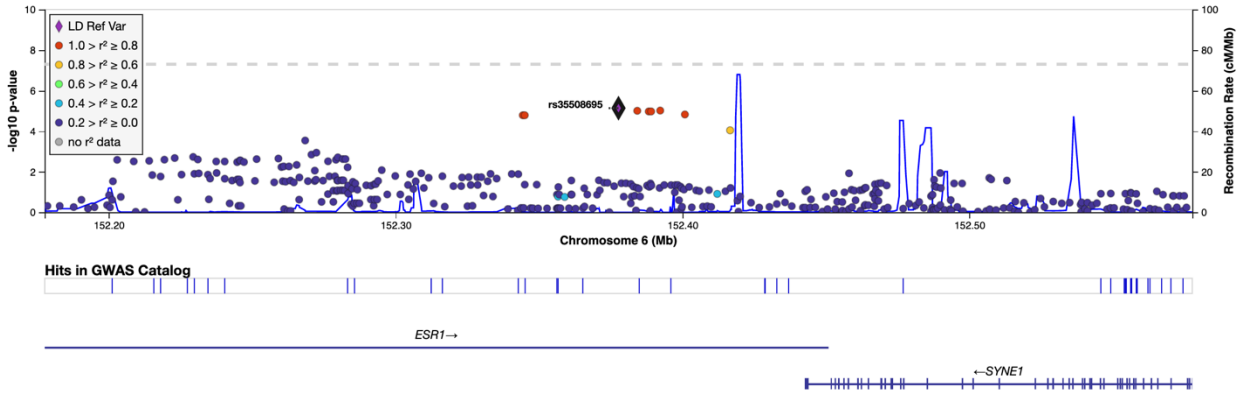
I


LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
rs9399396
-log10 p-value
Recombination Rate (cM/Mb)
142.40
142.50
142.60
142.70
Chromosome 6 (Mb)
Hits in GWAS Catalog
←NMBR
GJE1→
VTA1→
ADGRG6→

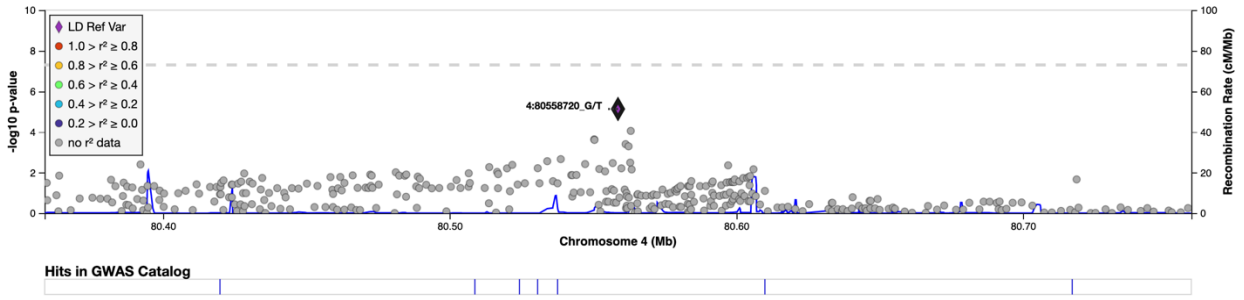
J


LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
rs11230097
-log10 p-value
Recombination Rate (cM/Mb)
59.50
59.60
59.70
59.80
Chromosome 11 (Mb)
Hits in GWAS Catalog
←OR10V1
STX3→
←MRPL16
←CBLIF
←TCN1
OOSP3→
OOSP1→
OOSP4A→
OOSP4B→
OOSP2→
MS4A3→

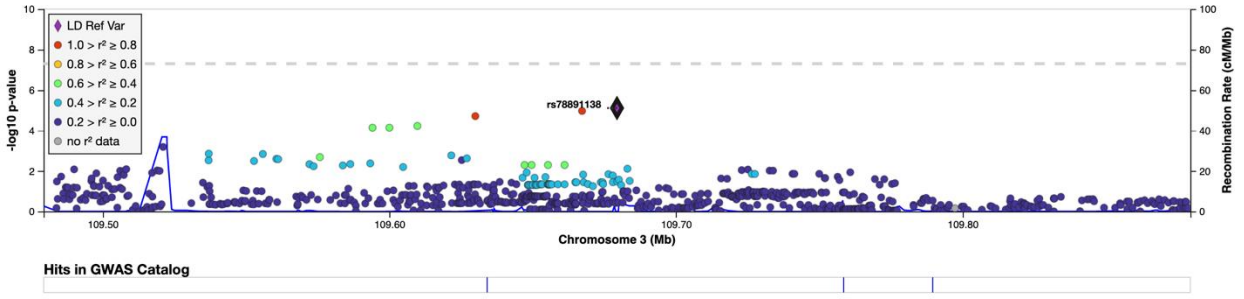
K


LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
rs35508695
-log10 p-value
Recombination Rate (cM/Mb)
152.20
152.30
152.40
152.50
Chromosome 6 (Mb)
Hits in GWAS Catalog
ESR1→
←SYNE1

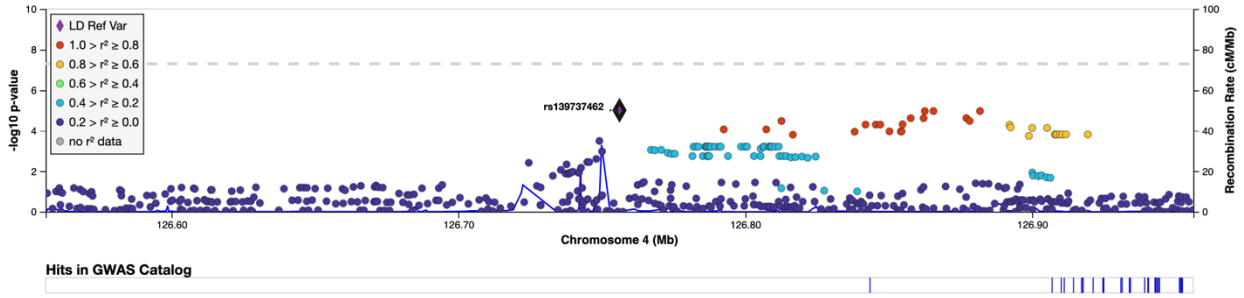
L



M



N



A. rs35324031 (chr10:55529375); B. rs34201252 (chr9:137426383); C. rs4710384 (chr6:63554552); D. rs112019823 (chr1:194344257); E. rs7503428 (chr17:32372318); F. rs35643512 (chr2:126373265); G. rs12834574 (chr23:131692337); H. rs28567072 (chr8:107883740); I. rs9399396 (chr6:142564217); J. rs11230097 (chr11:59646339); K. rs35508695 (chr6:152377709); L. rs3111790 (chr4:80558720); M. rs78891138 (chr3:109679425); N. rs139737462 (chr4:126756165). Each point represents the significance (–log10-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 [r2]) 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.

Appendix 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 (Pillas et al., 2010)
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 (Geller et al., 2011)
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)	
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)	Age at first tooth eruption and/or number of teeth at 12 months (Fatemifar et al., 2013)
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)	

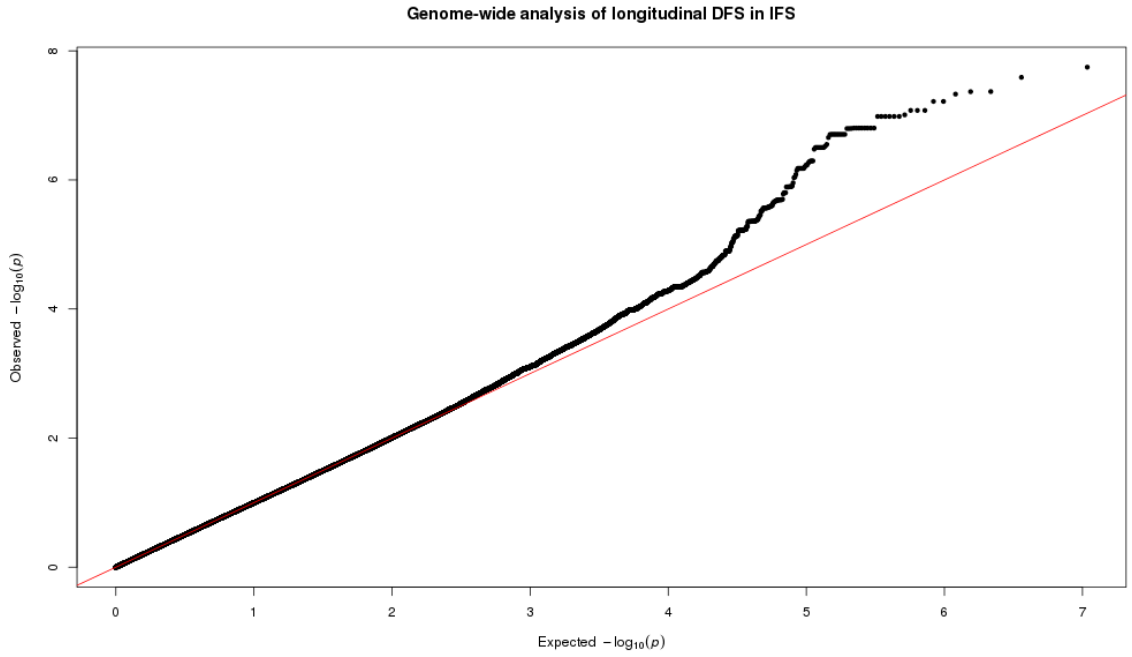
rs17563	14	54417522	G/A	0.464	0.996	1.00 (0.81-1.23)
rs1994969	17	47080431	T/G	0.474	0.819	0.98 (0.8-1.19)
rs412000	17	56709058	C/G	0.458	0.675	0.96 (0.78-1.17)
rs8080944	17	68185586	A/G	0.421	0.437	0.92 (0.75-1.13)
rs11796357	23	68798703	A/G	0.210	0.708	0.96 (0.8-1.16)

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

Significance threshold: 0.05/25=0.002. No tooth eruption-related SNPs were significantly associated with caries incidence.

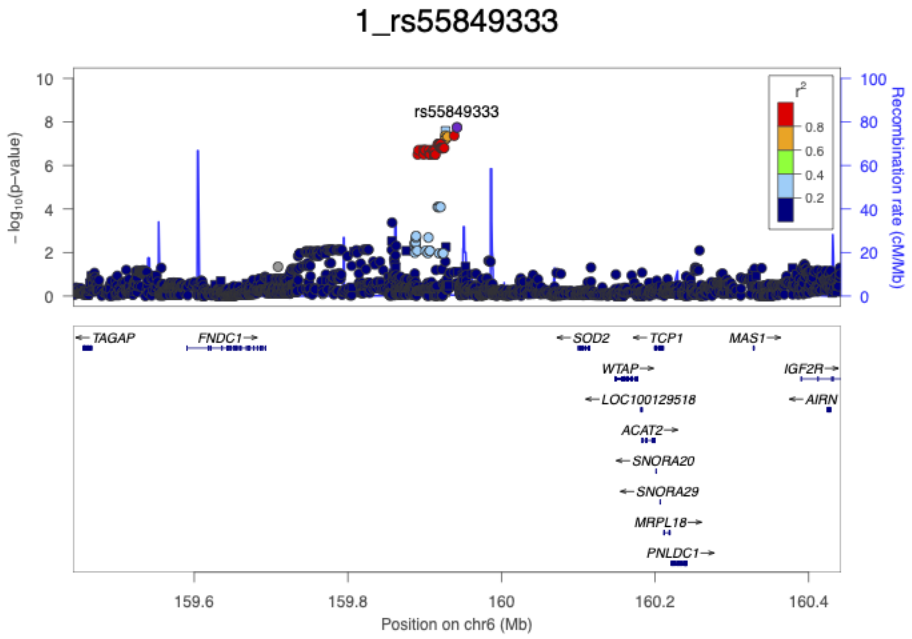
Appendix C Supplementary materials for Chapter 4

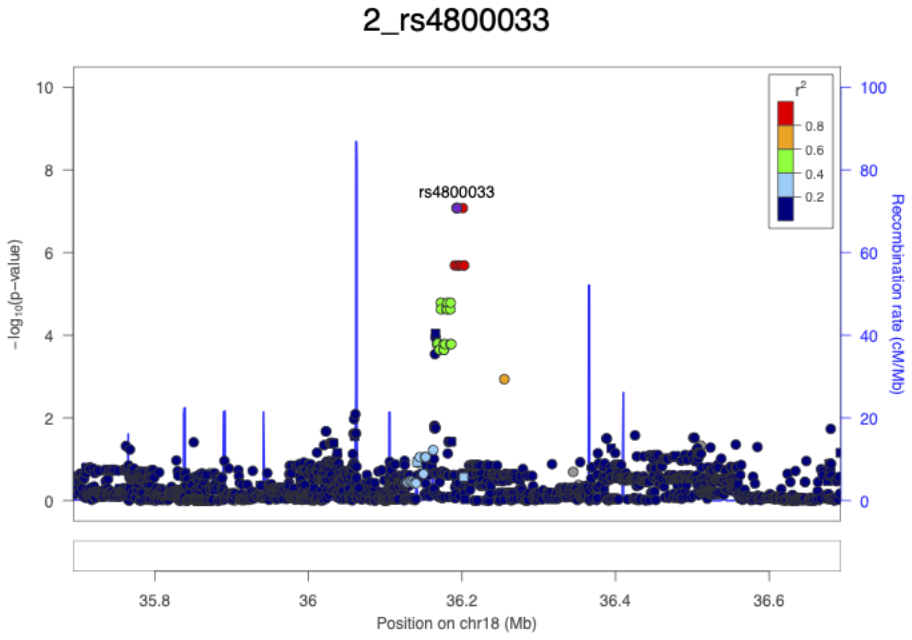
Appendix Figure 3 Quantile-quantile plot for the longitudinal GWAS



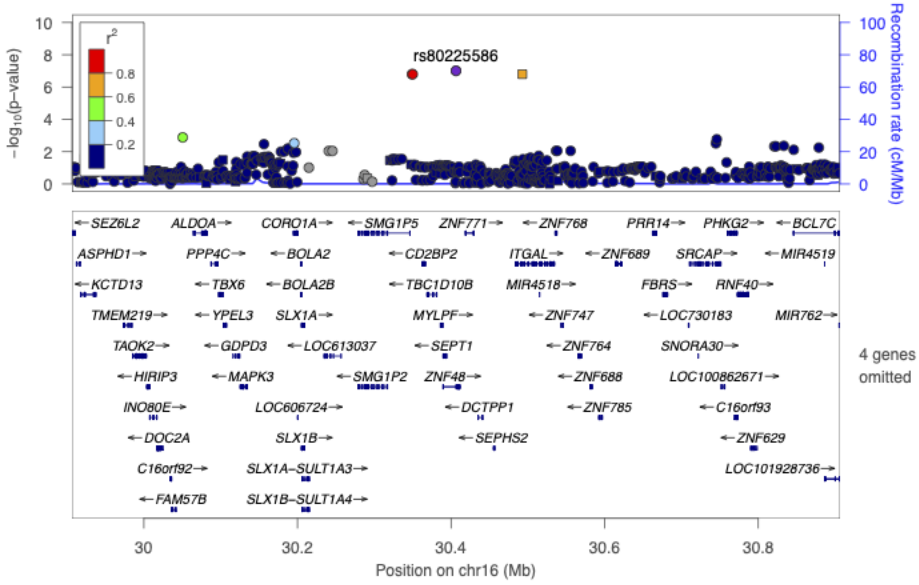
The genomic inflation factor was 1.01.

Appendix Figure 4 Reginal association plots for 13 SNPs at 1 genome-wide and 12 suggestive loci from longitudinal GWAS of DFS at 4 time points

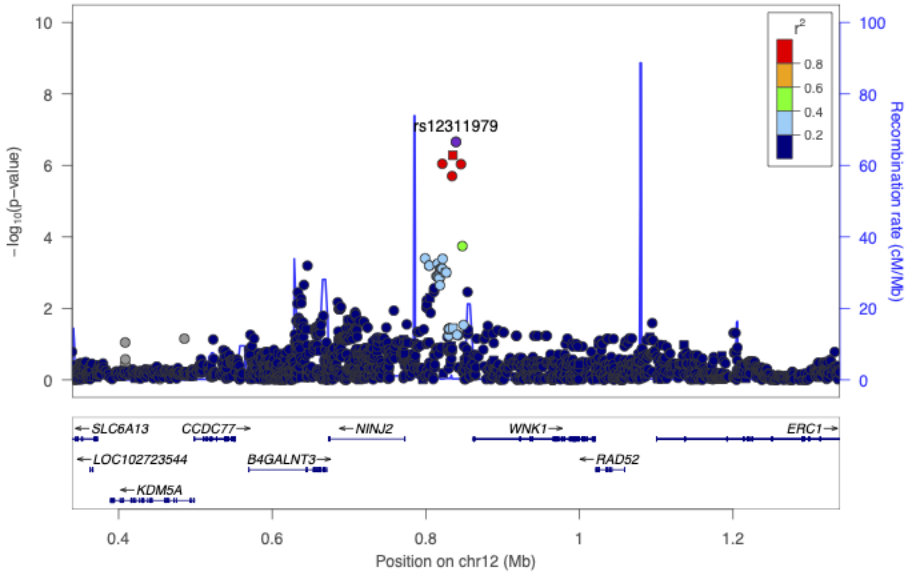




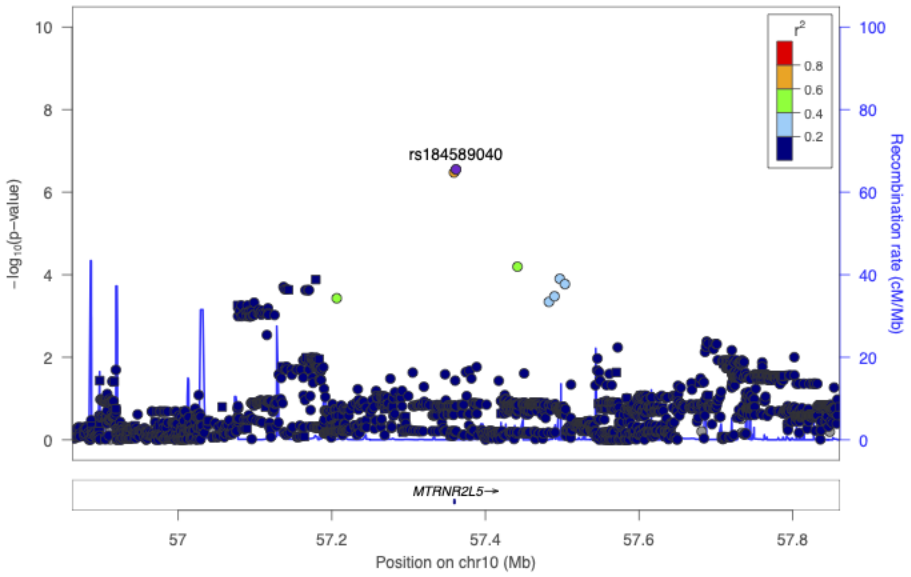
3_rs80225586



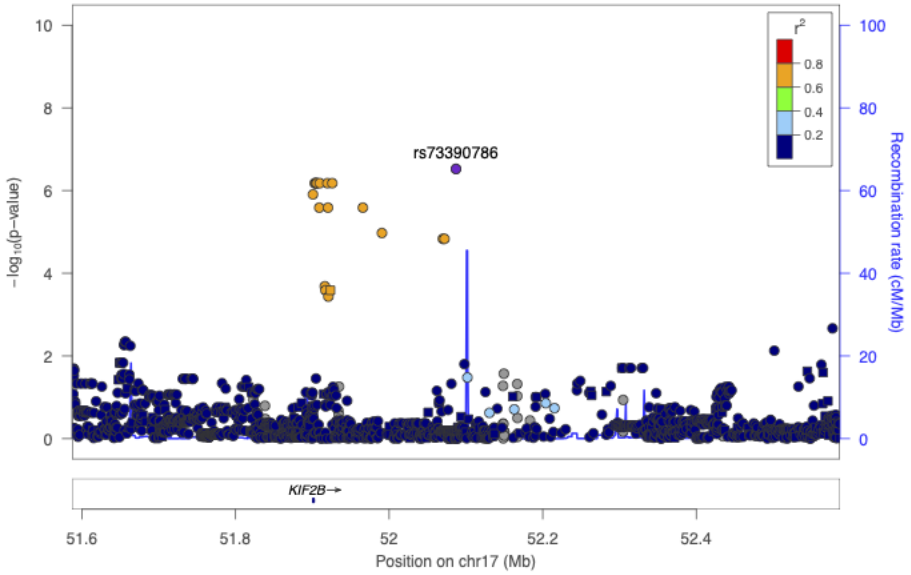
4_rs12311979



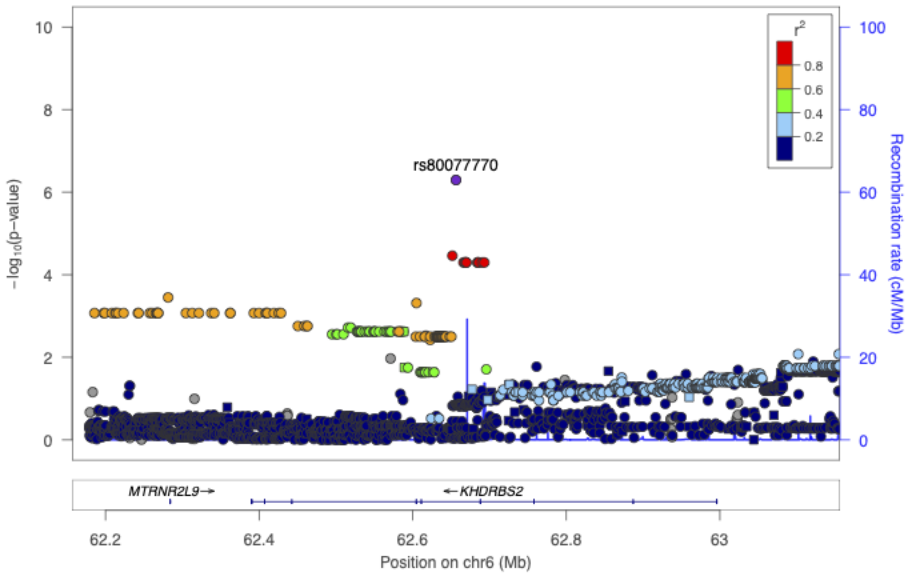
5_rs184589040



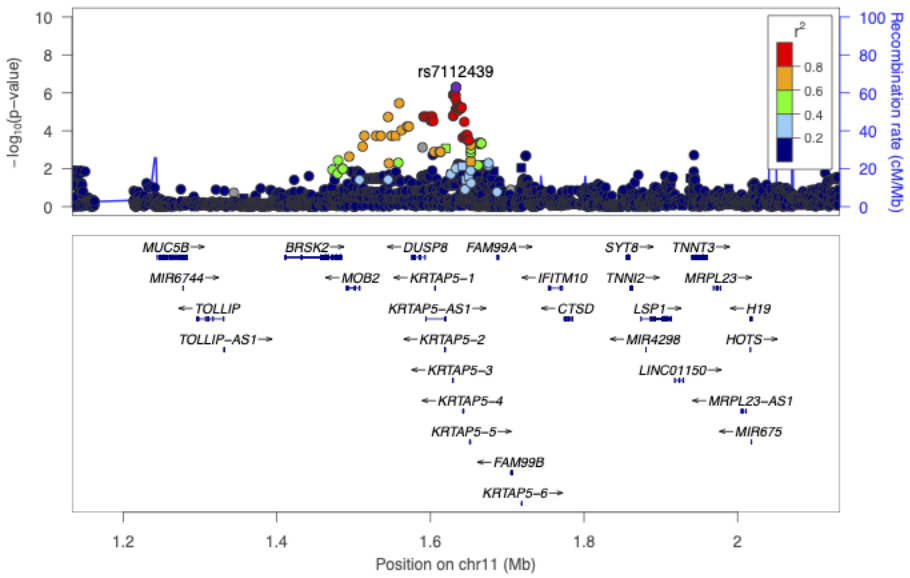
6_rs73390786



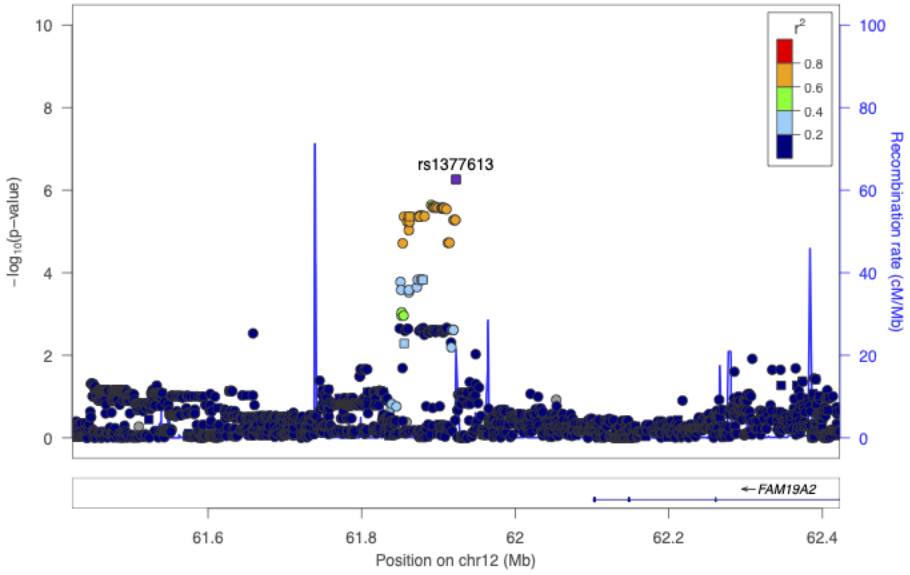
7_rs80077770



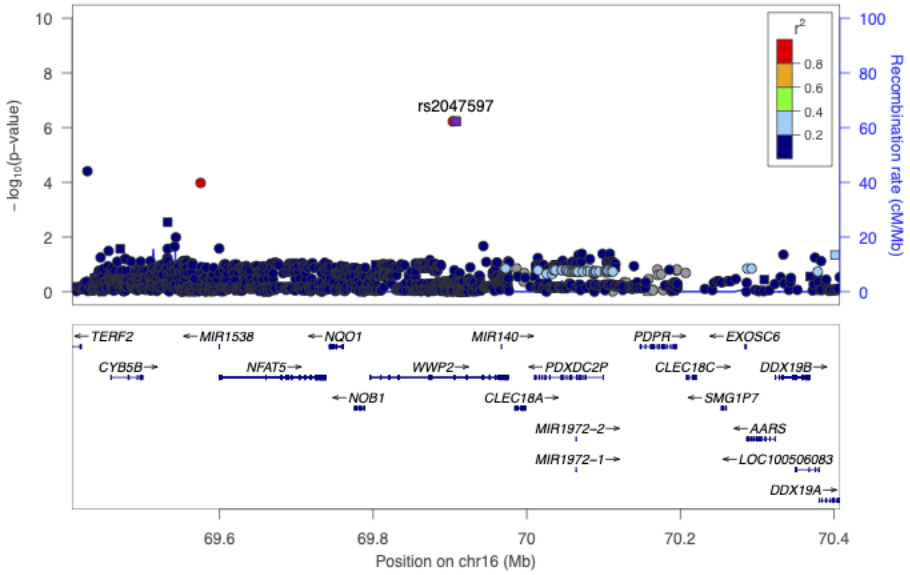
8_rs7112439



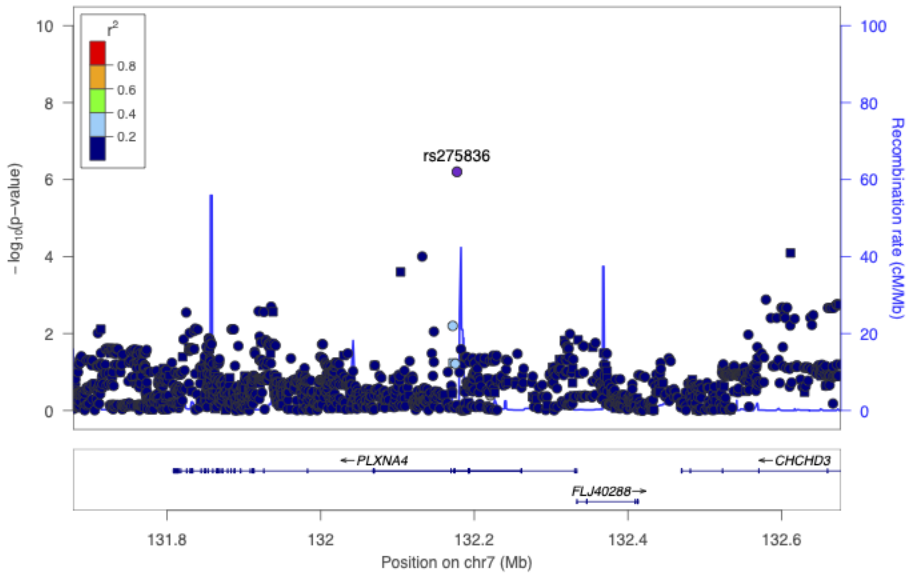
9_rs1377613



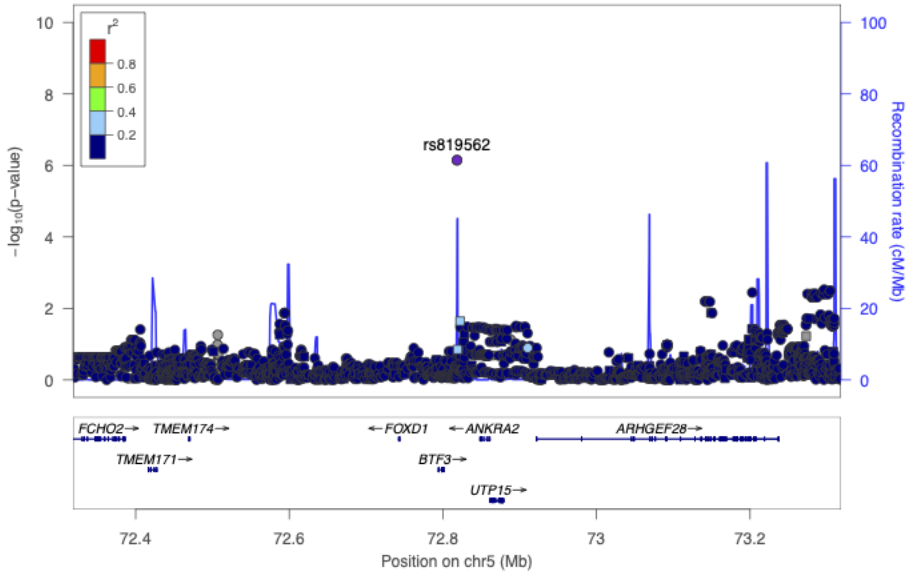
10_rs2047597

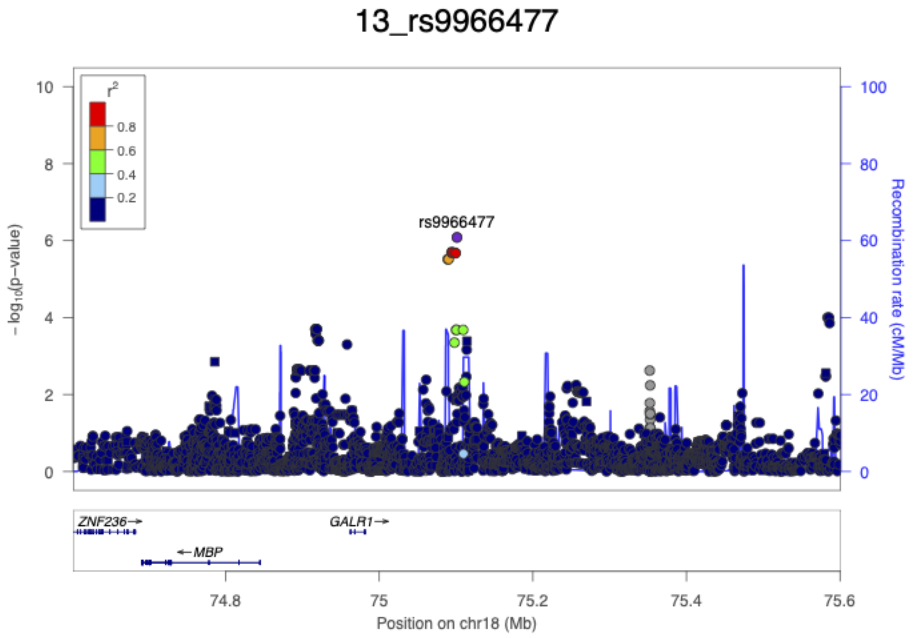


11_rs275836



12_rs819562

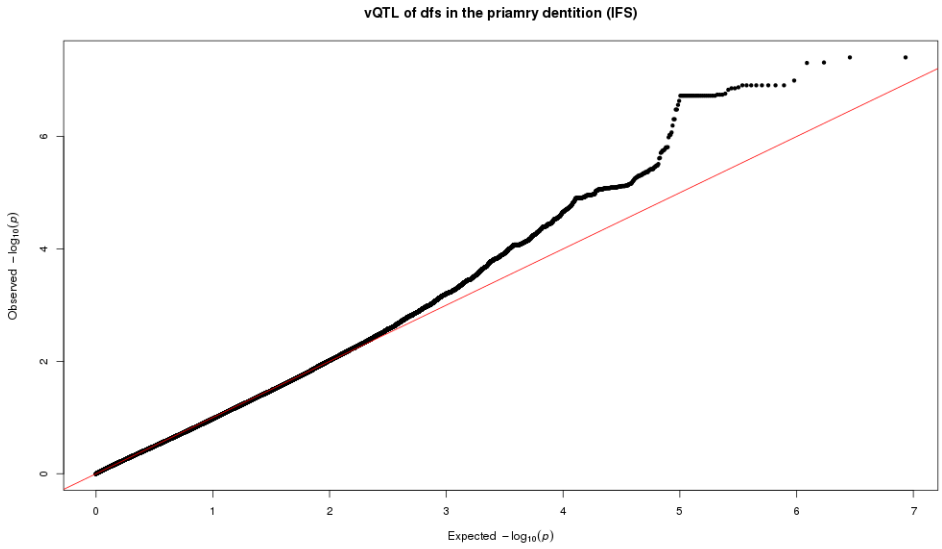


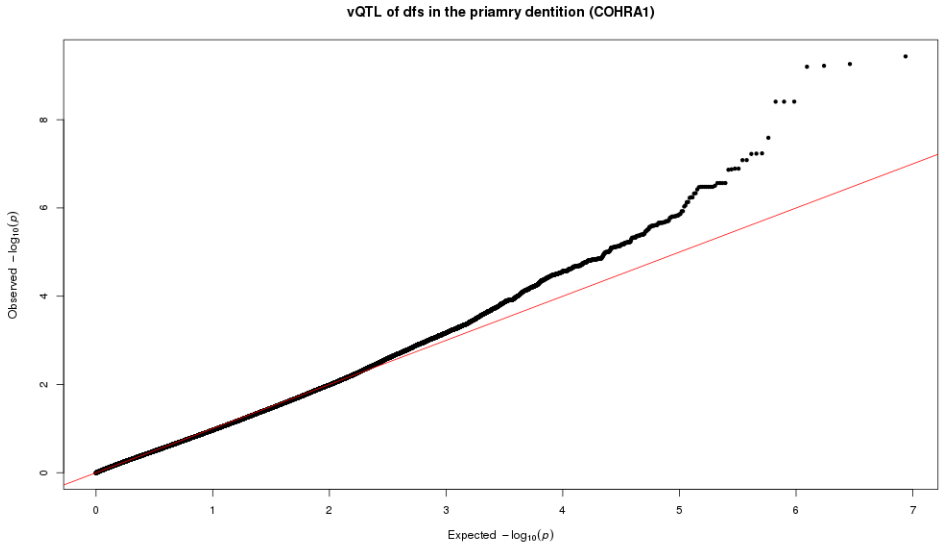


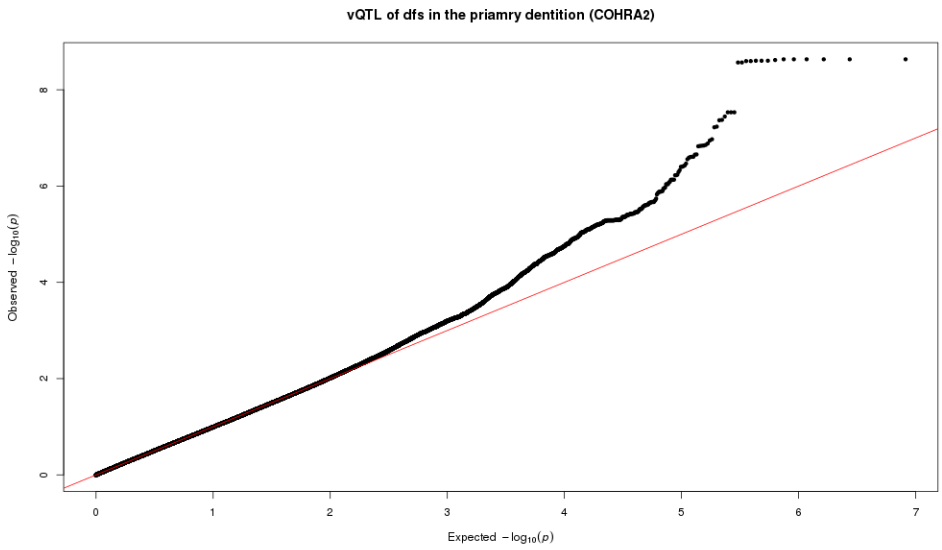
Each point represents the significance (–log10-transformed P value; left Y-axis) of a SNP. The purple diamond point with rsID label is the lead SNP in the association region. The color of the points represents the linkage disequilibrium (i.e., correlation [r2]) of the SNP with the lead SNP. Right Y-axis (recombination rate) provides information about the LD structure of region. Genes near the association signal are shown on the bottom.

Appendix D Supplementary materials for Chapter 5

Appendix Figure 5 Quantile-quantile plot for genome-wide vQTL analysis of DFS in IFS, COHRA1, and COHRA2

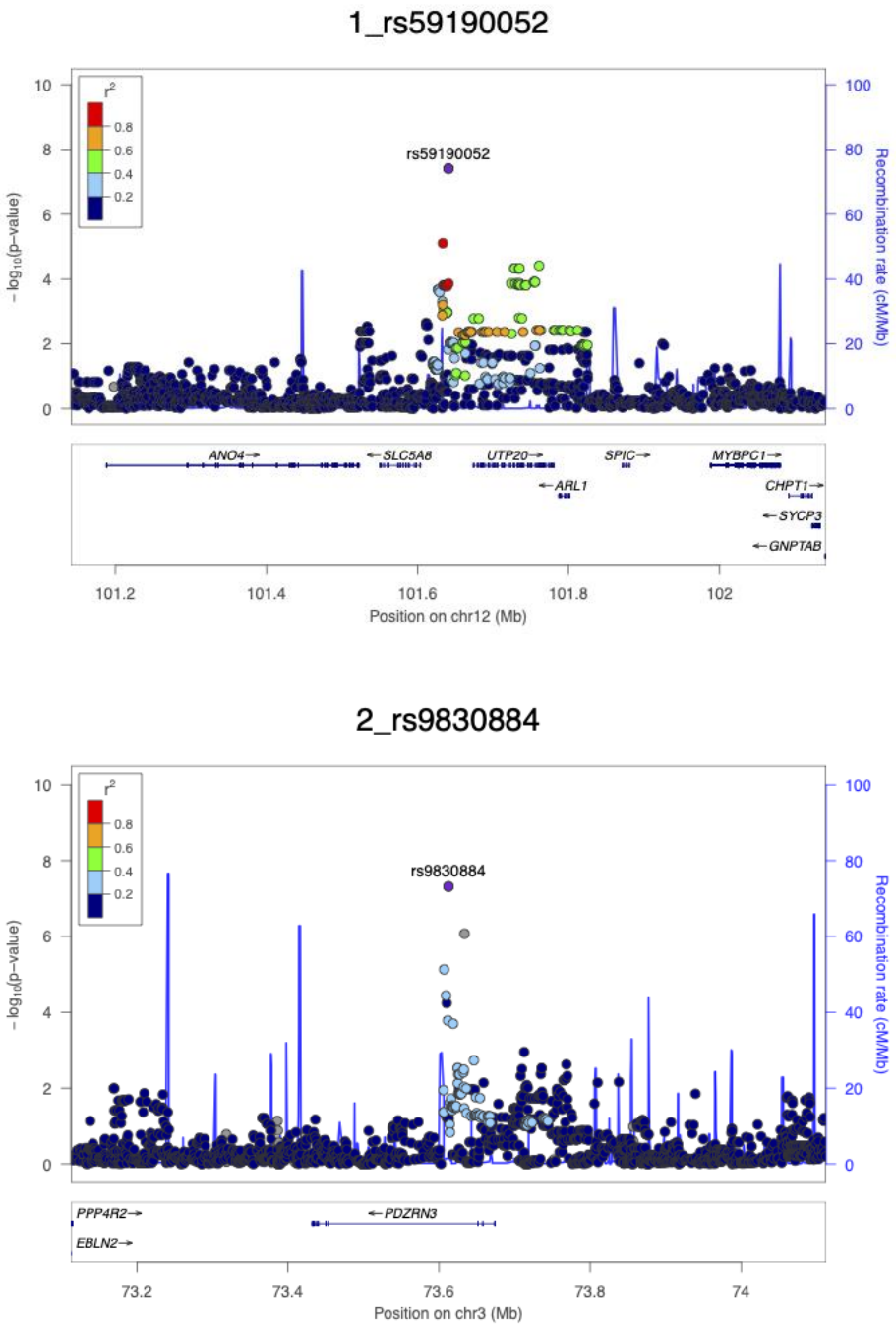




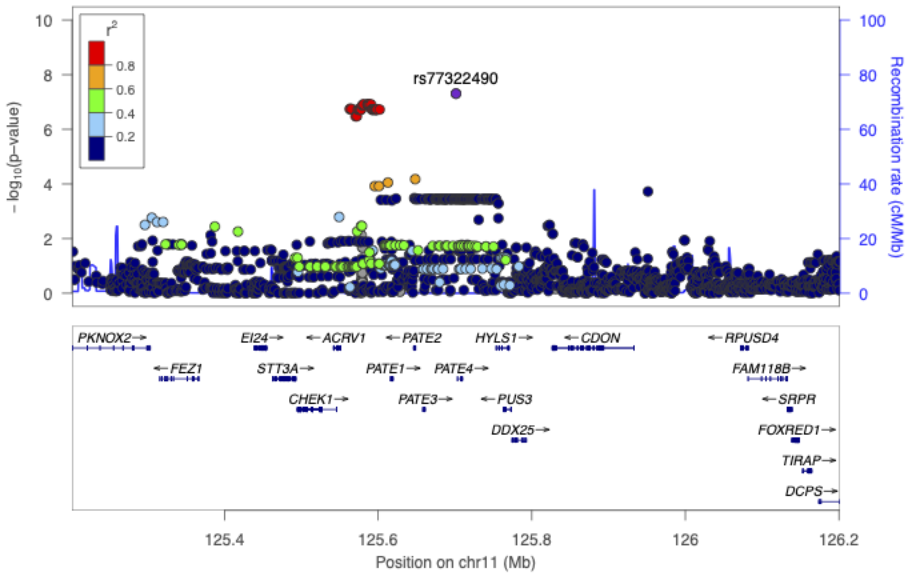


The genomic inflation factors were 1.00, 1.02, and 1.03 in IFS, COHRA1, and COHRA2, repectively.

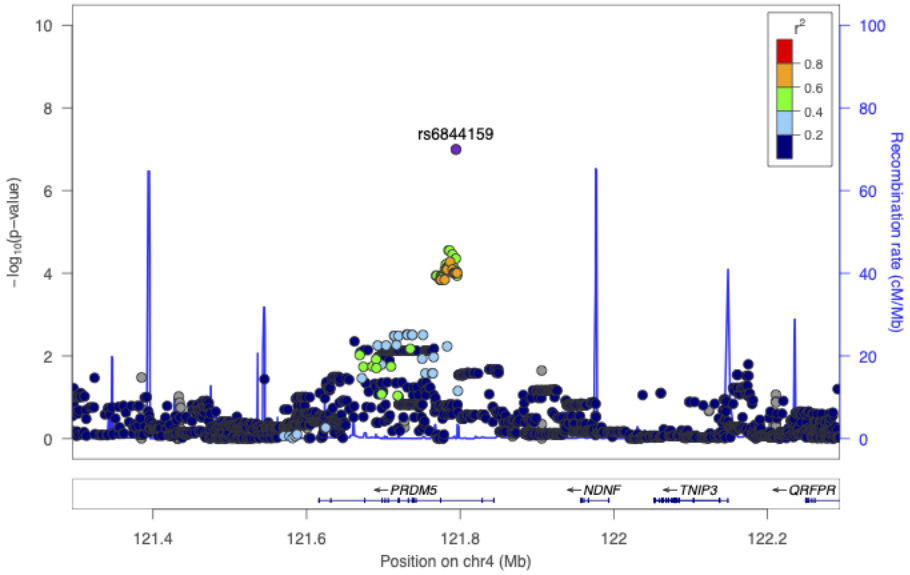
Appendix Figure 6 Reginal association plots for 40 top SNPs extracted from vQTL of DFS in IFS, COHRA1, COHRA2 and meta-analysis



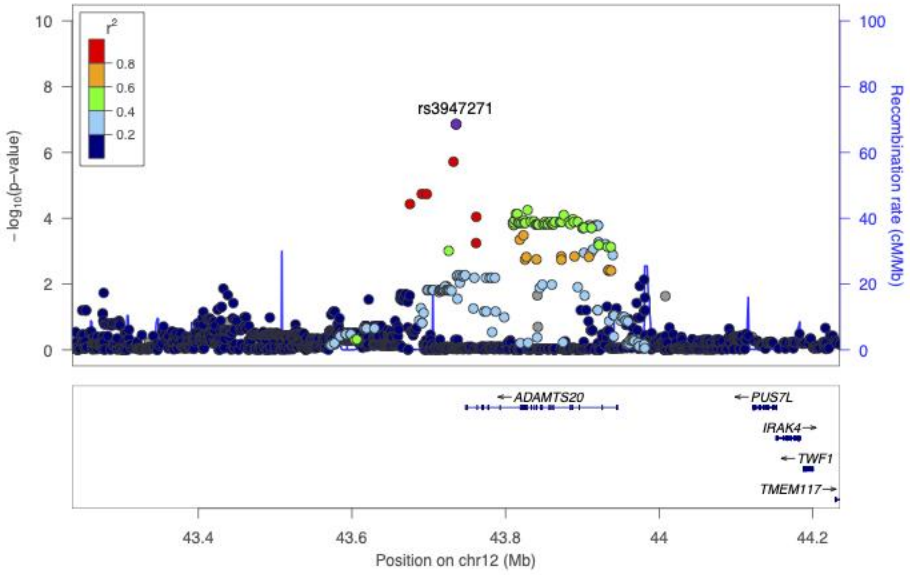
3_rs77322490



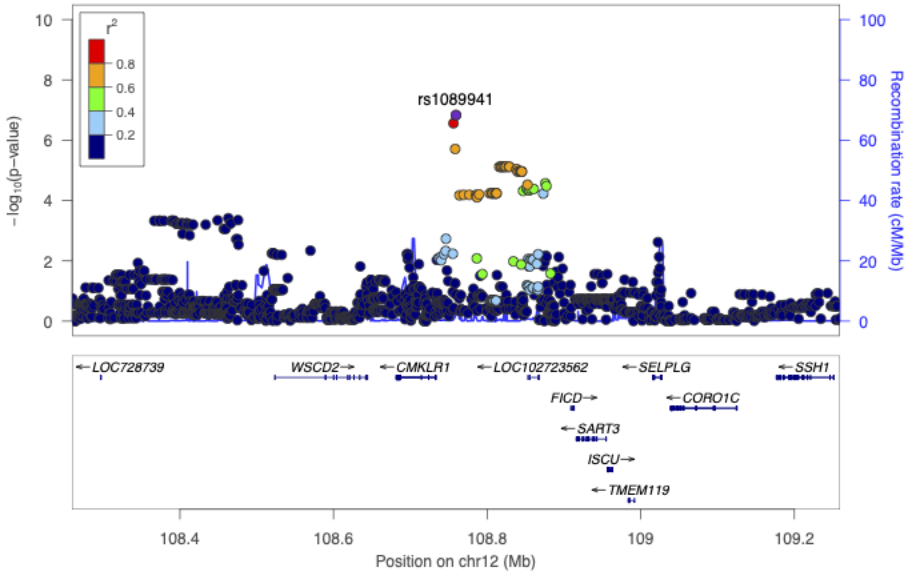
4_rs6844159



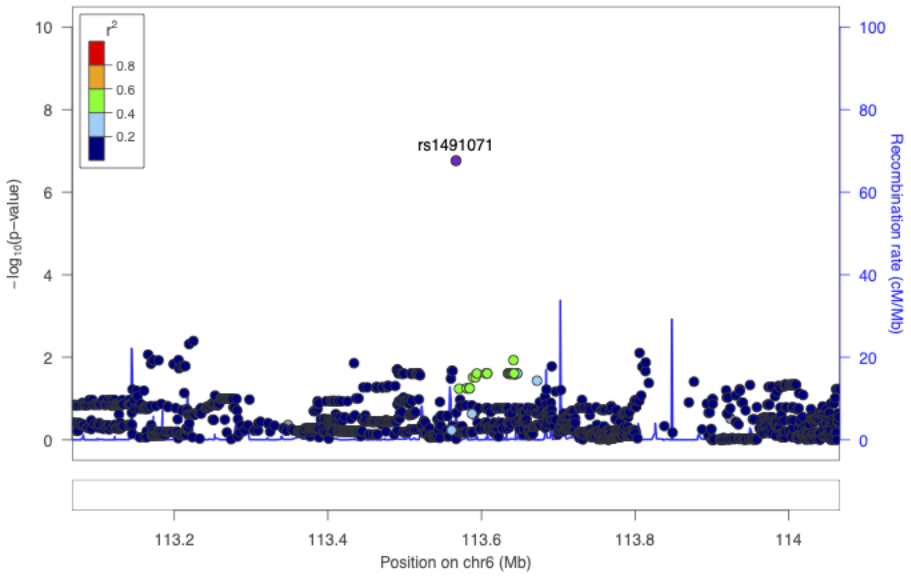
5_rs3947271



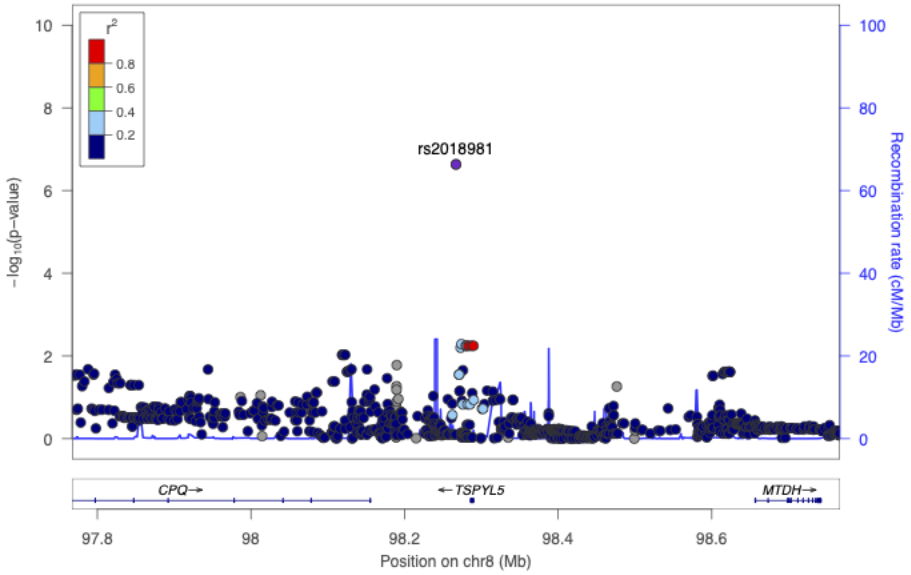
6_rs1089941



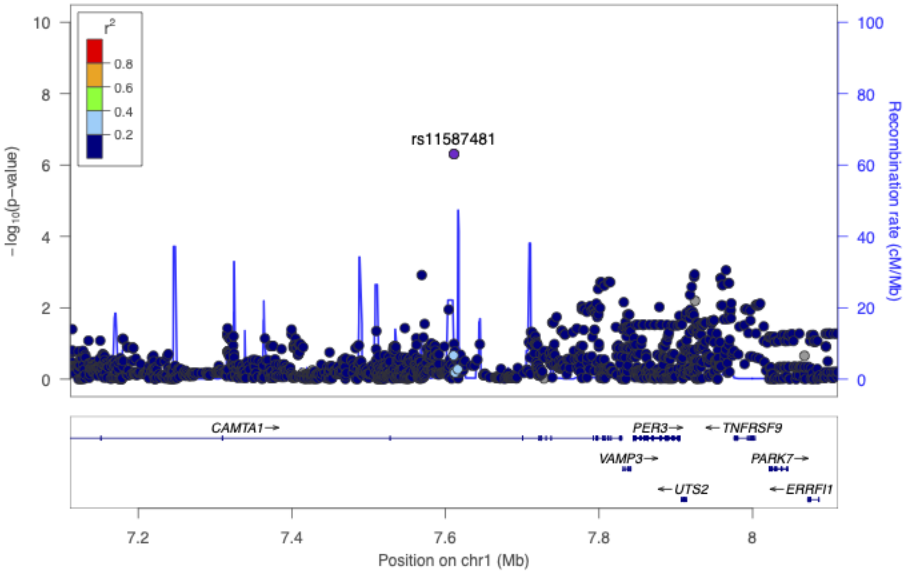
7_rs1491071



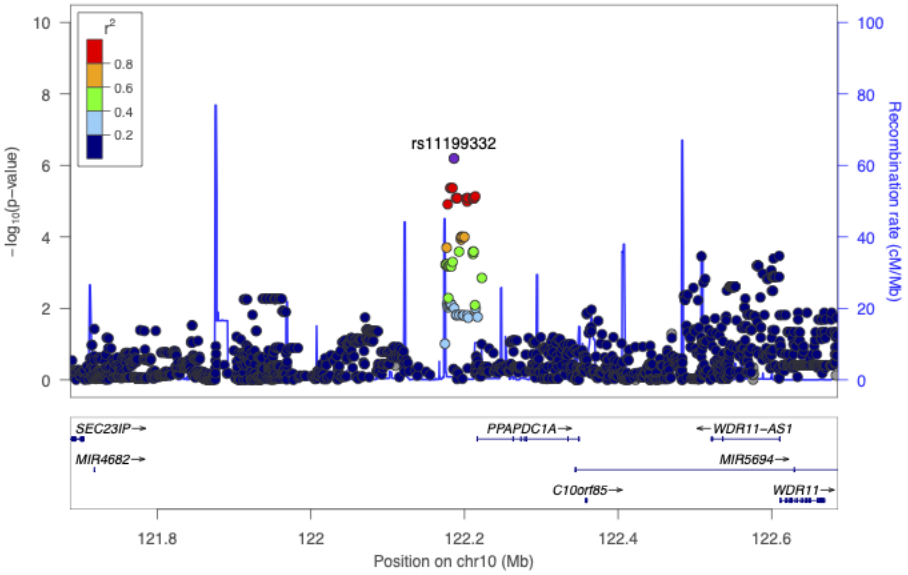
8_rs2018981



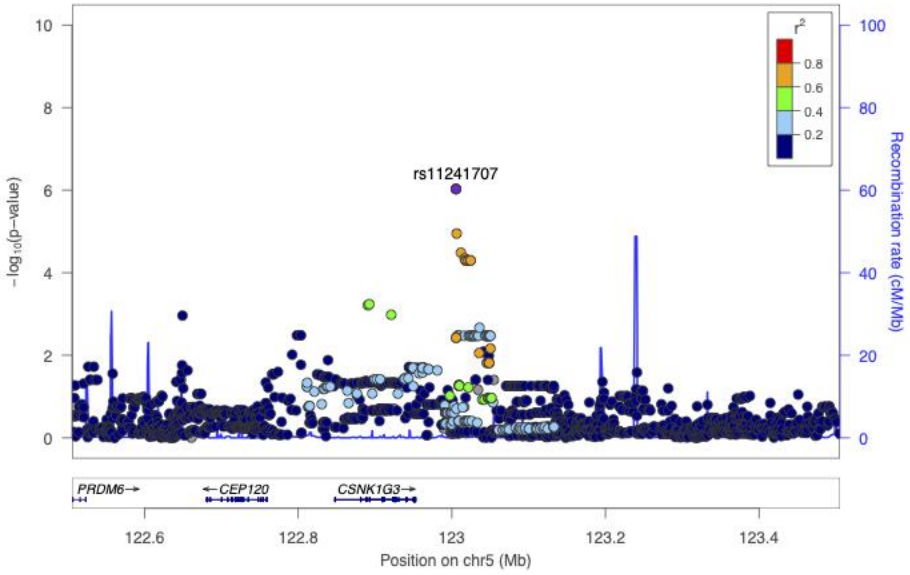
9_rs11587481



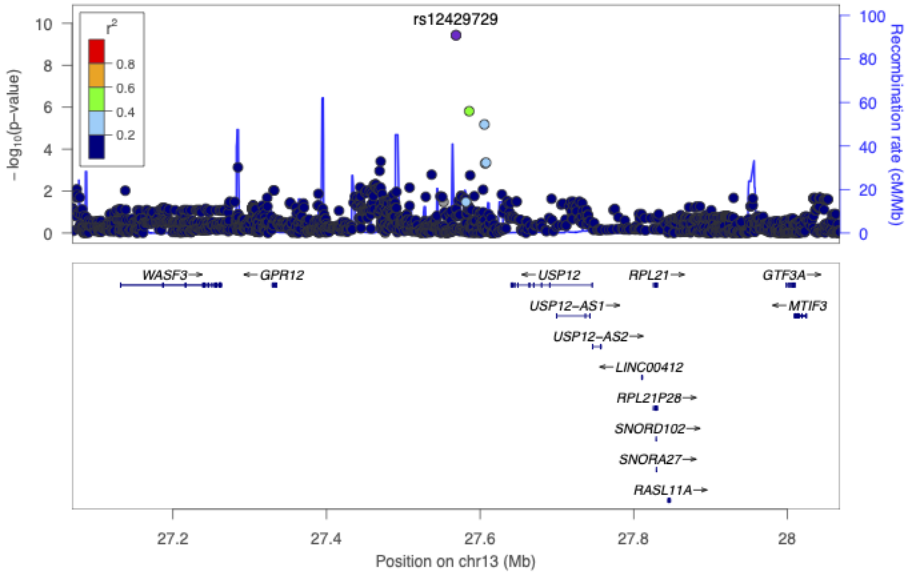
10_rs11199332



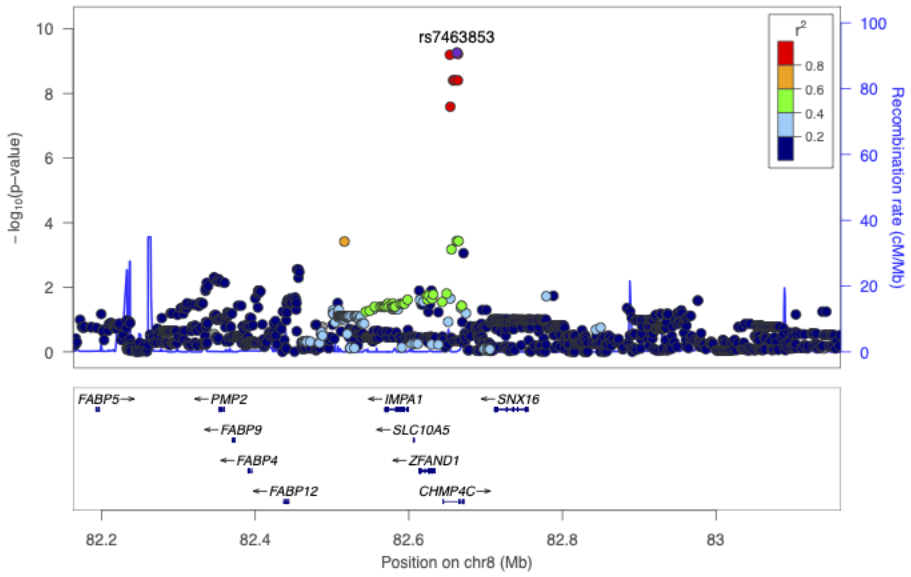
11_rs11241707



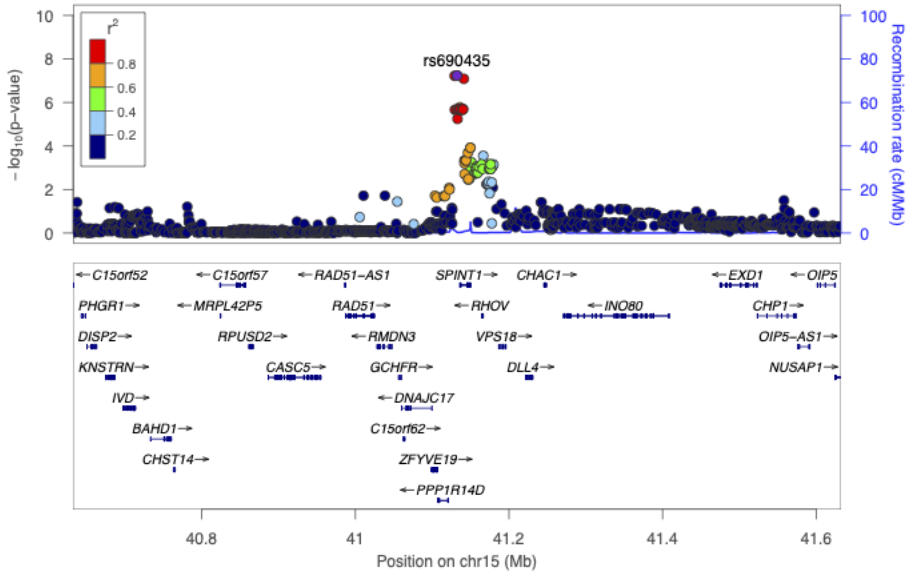
12_rs12429729



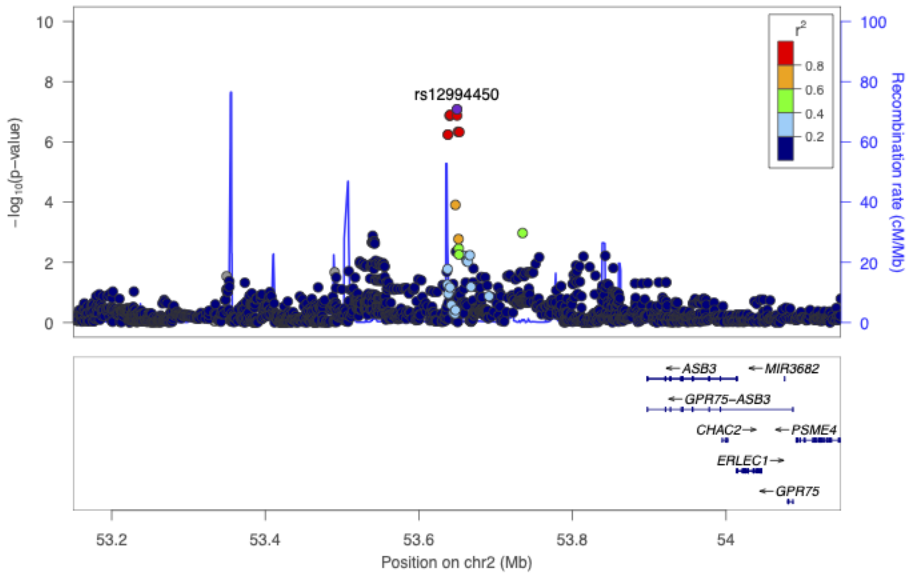
13_rs7463853



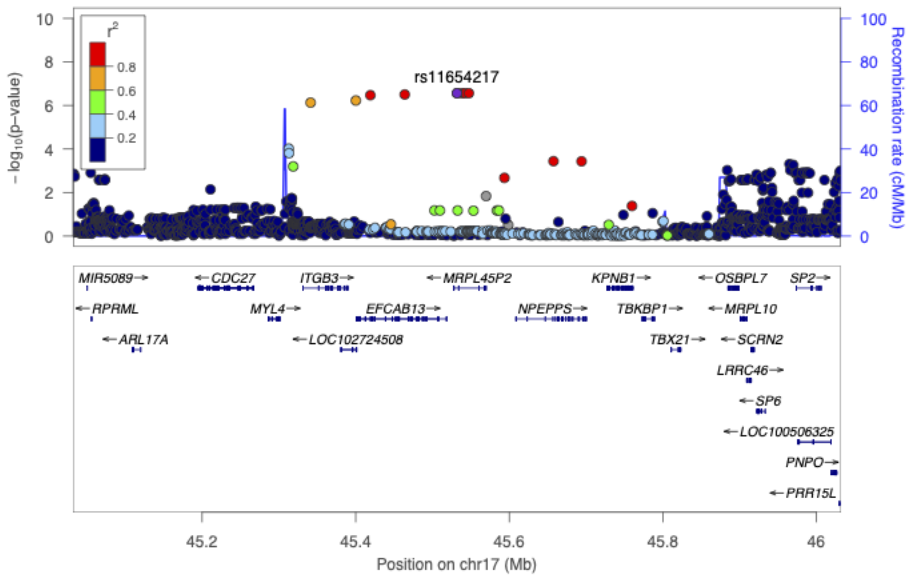
14_rs690435



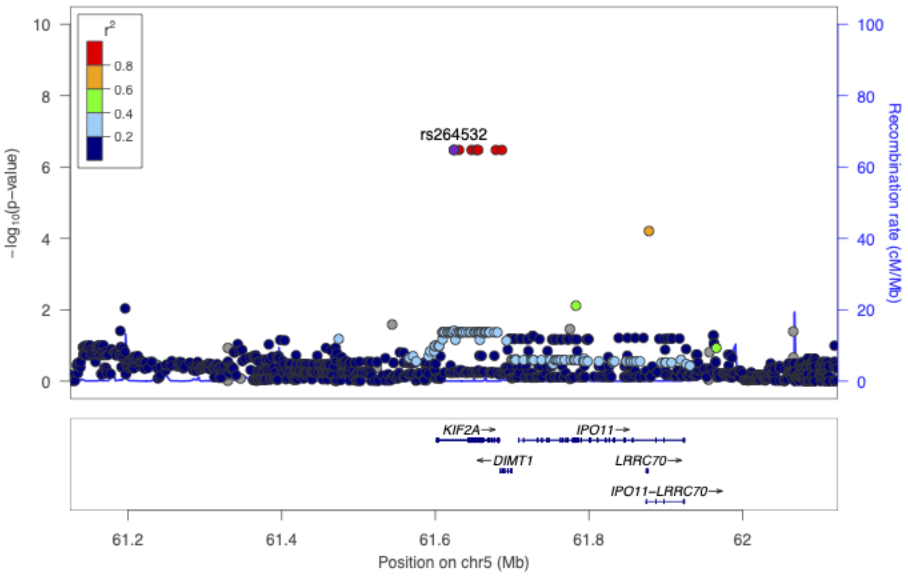
15_rs12994450



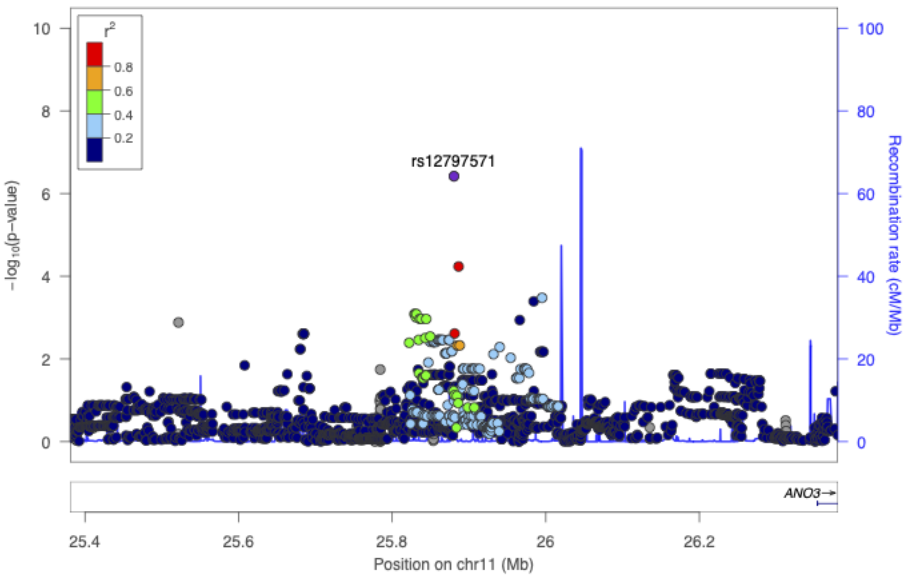
16_rs11654217



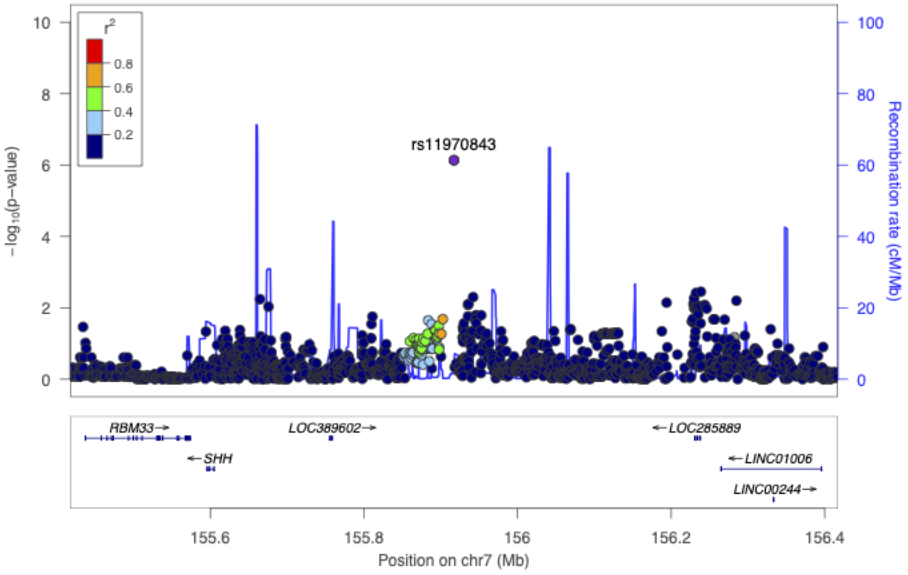
17_rs264532



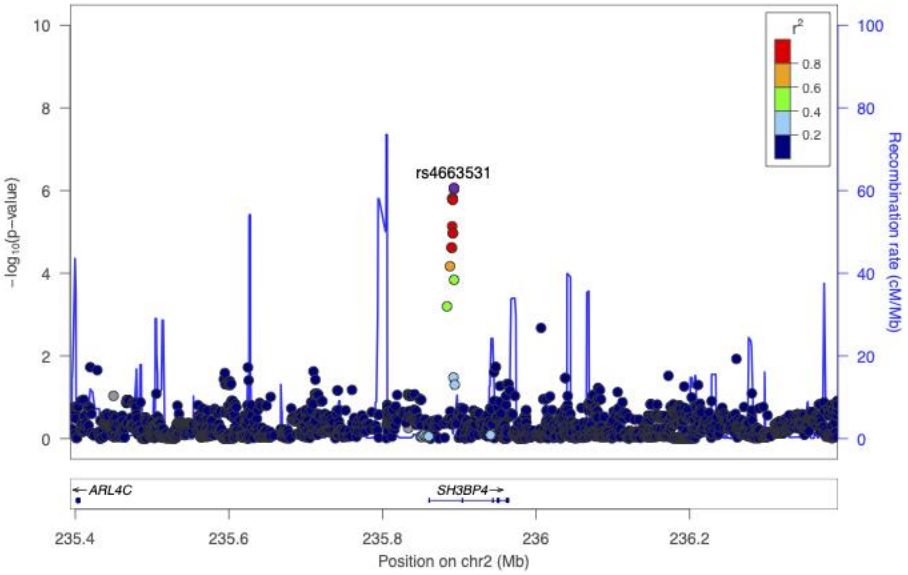
18_rs12797571



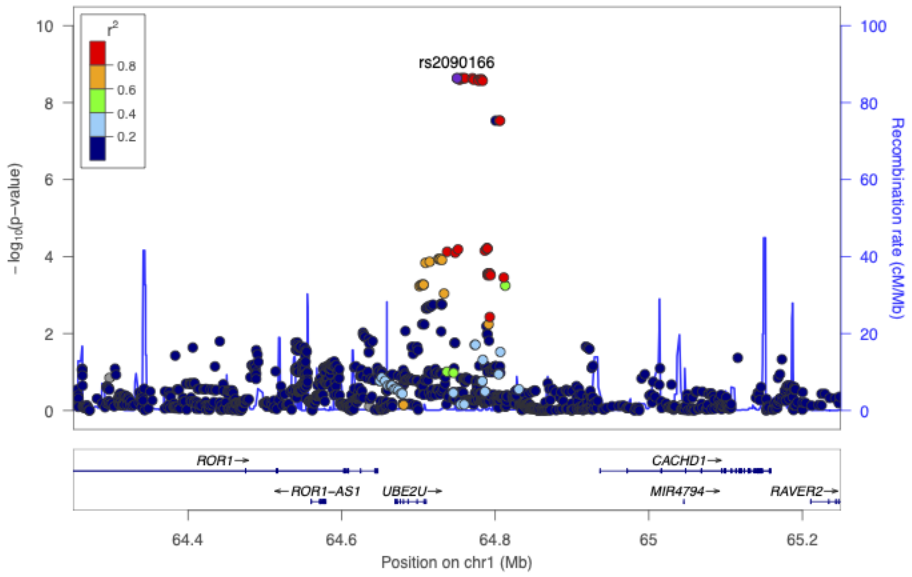
19_rs11970843



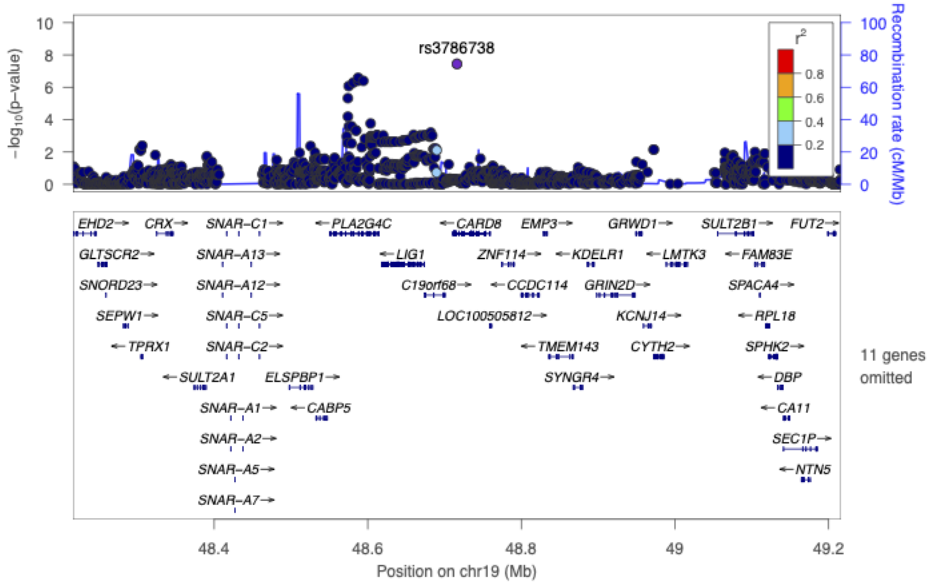
20_rs4663531



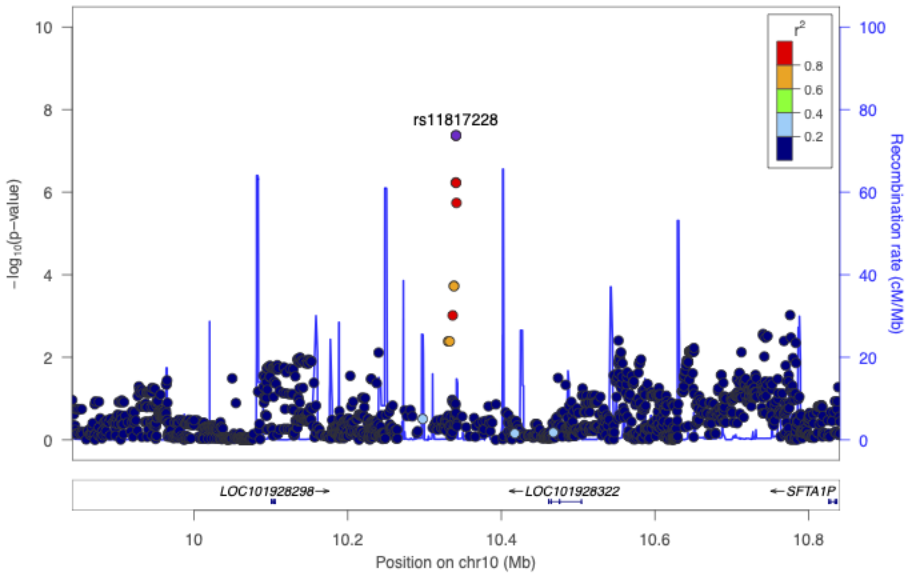
21_rs2090166



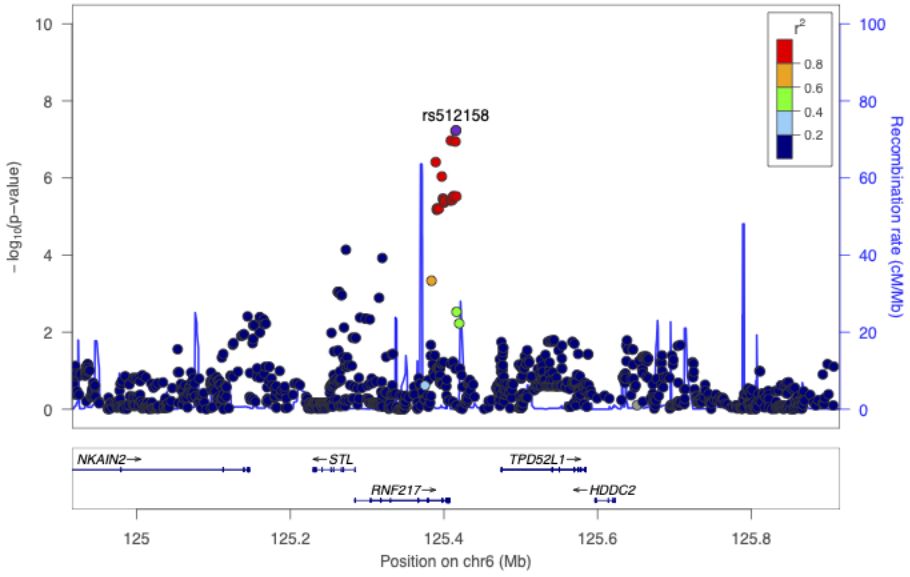
22_rs3786738



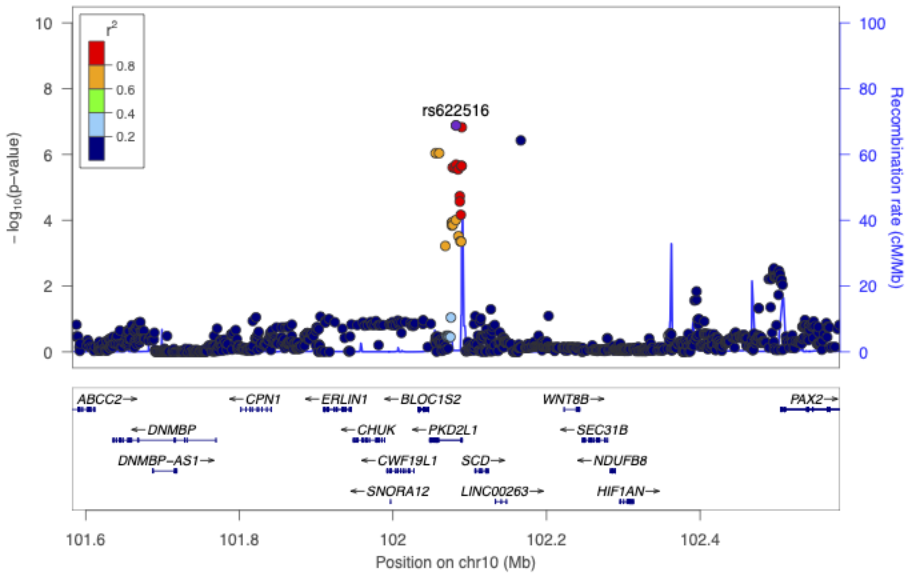
23_rs11817228



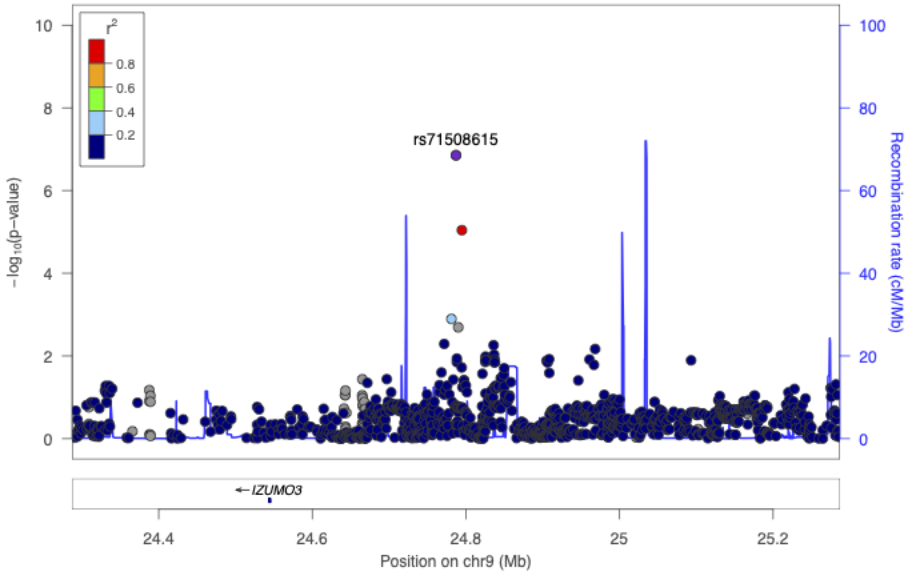
24_rs512158



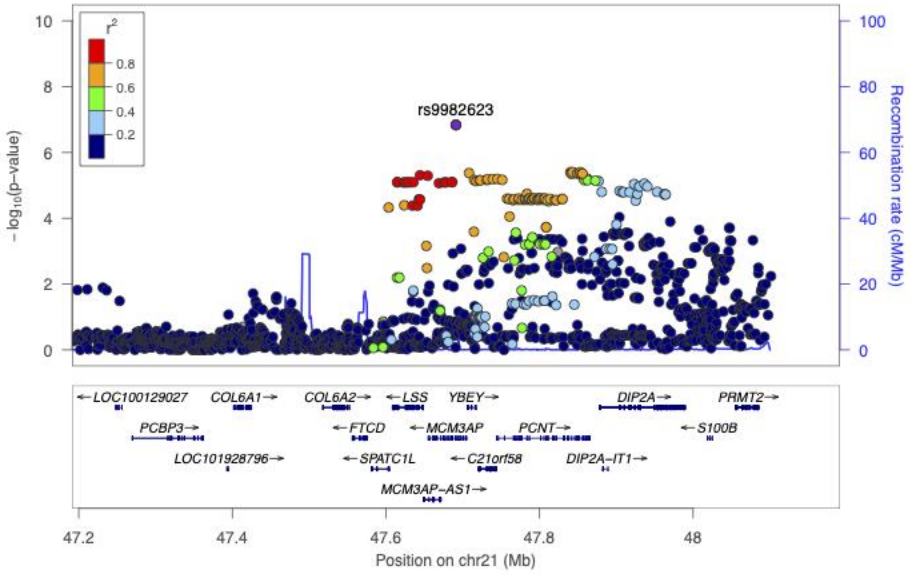
25_rs622516



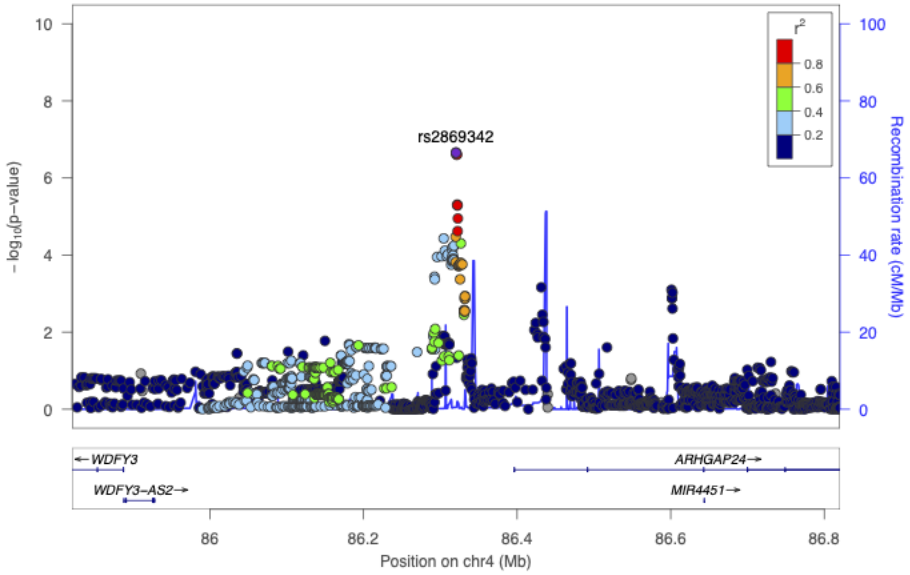
26_rs71508615



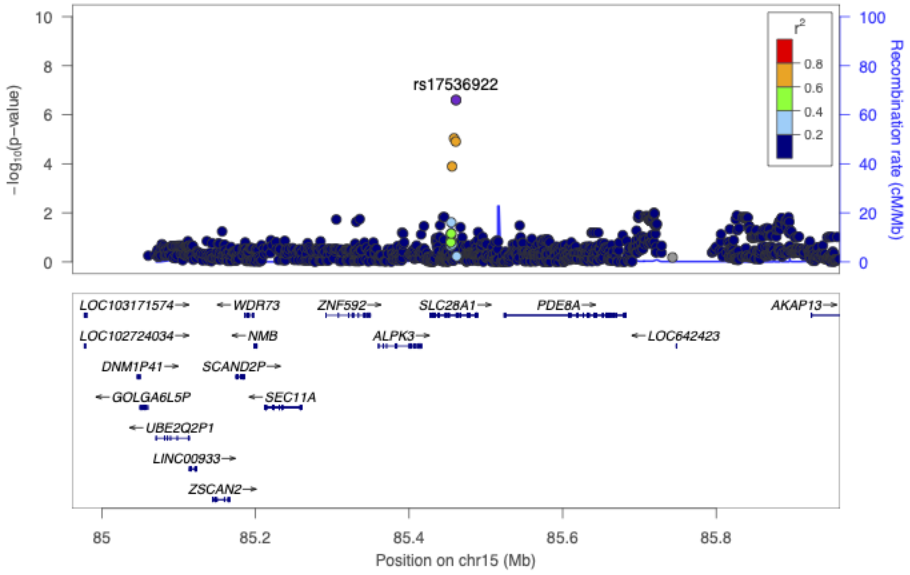
27_rs9982623



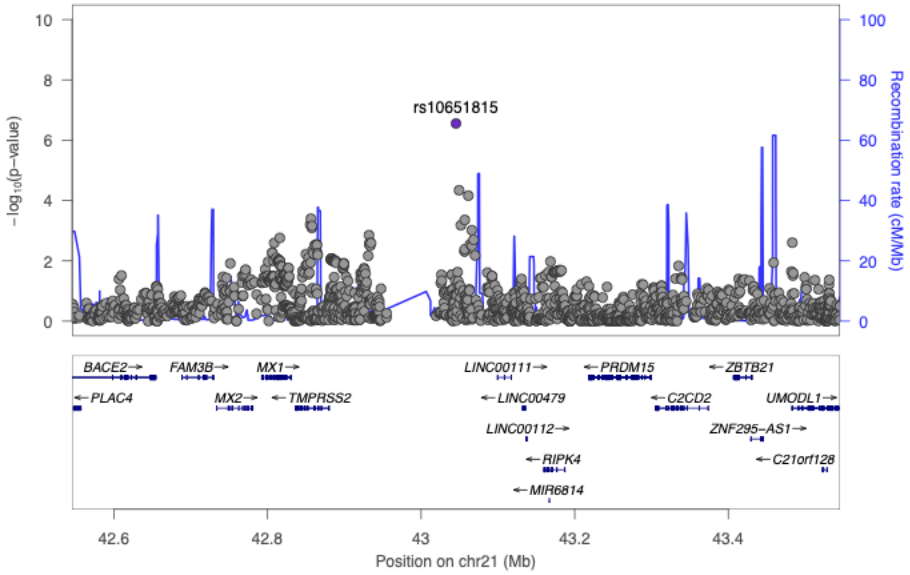
28_rs2869342



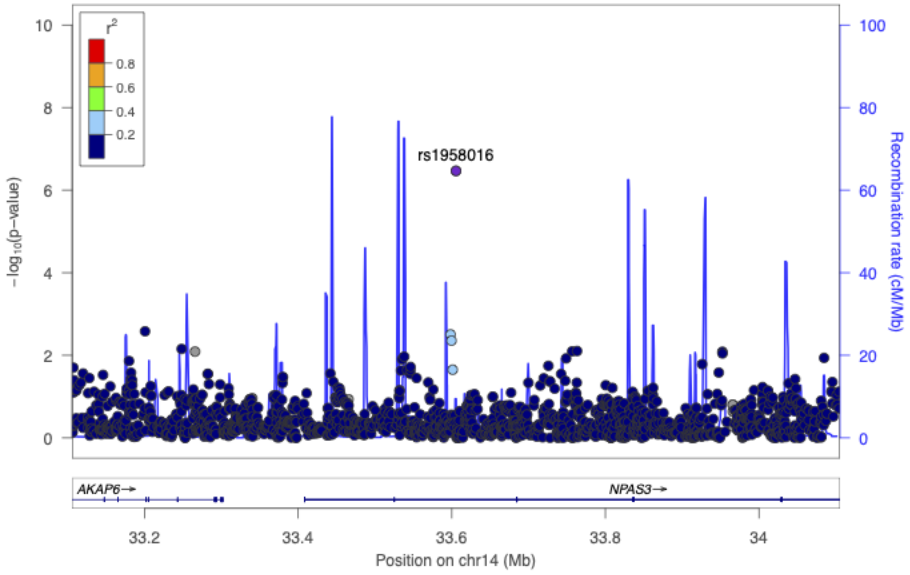
29_rs17536922



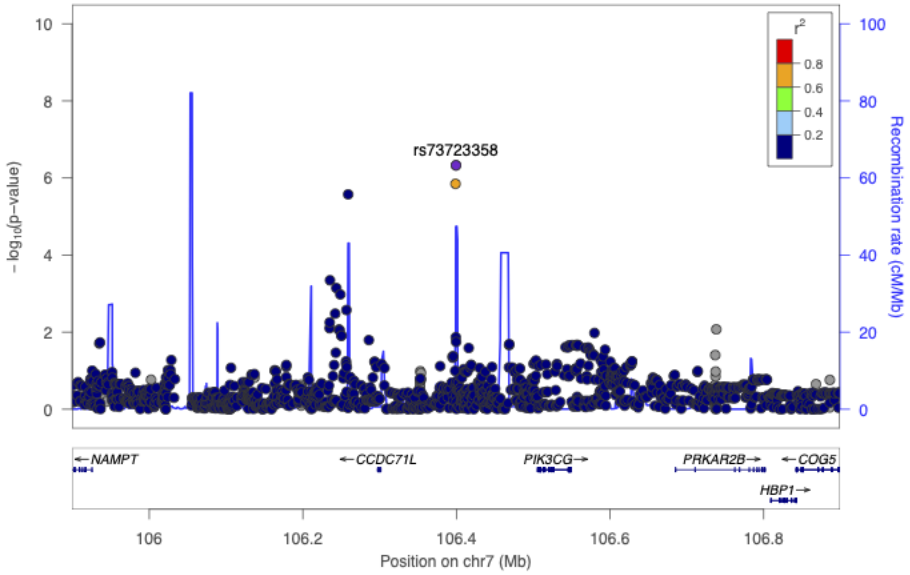
30_rs10651815



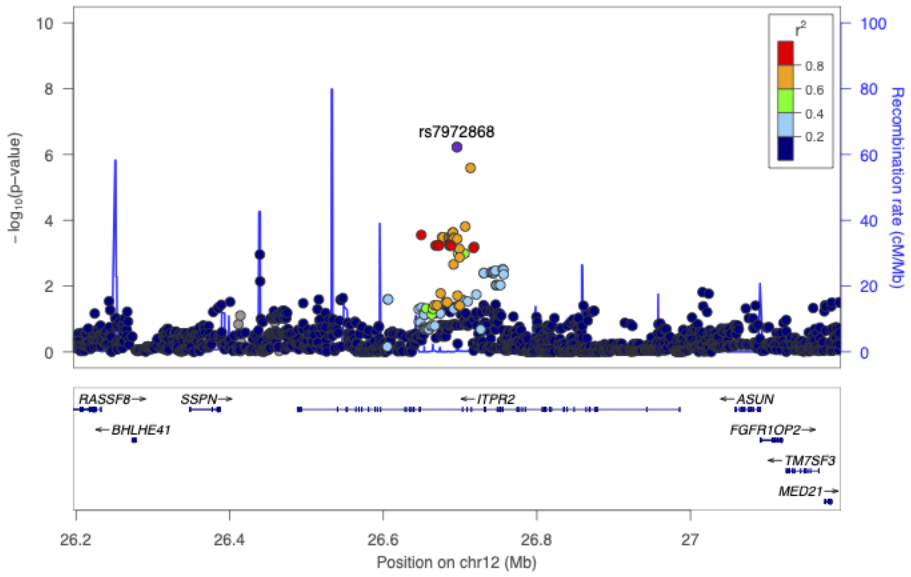
31_rs1958016



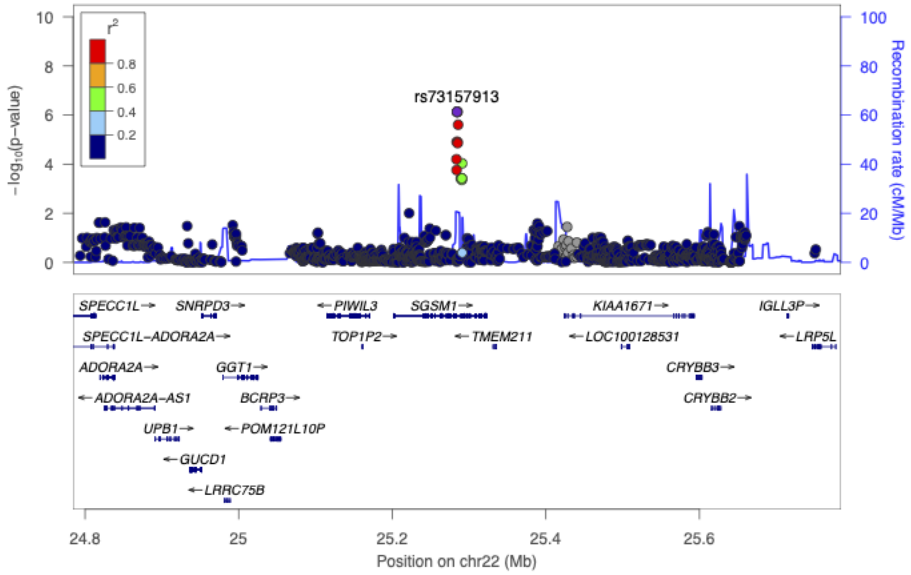
32_rs73723358



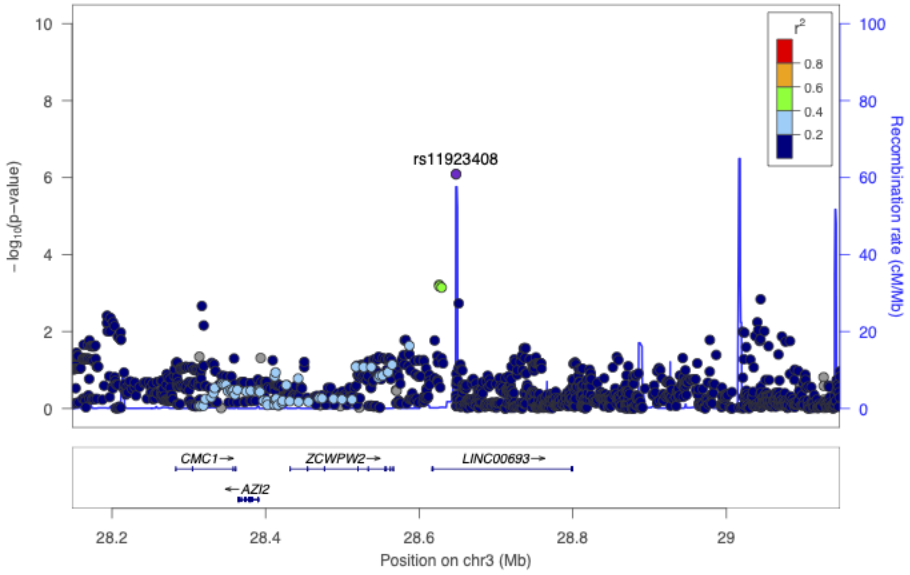
33_rs7972868



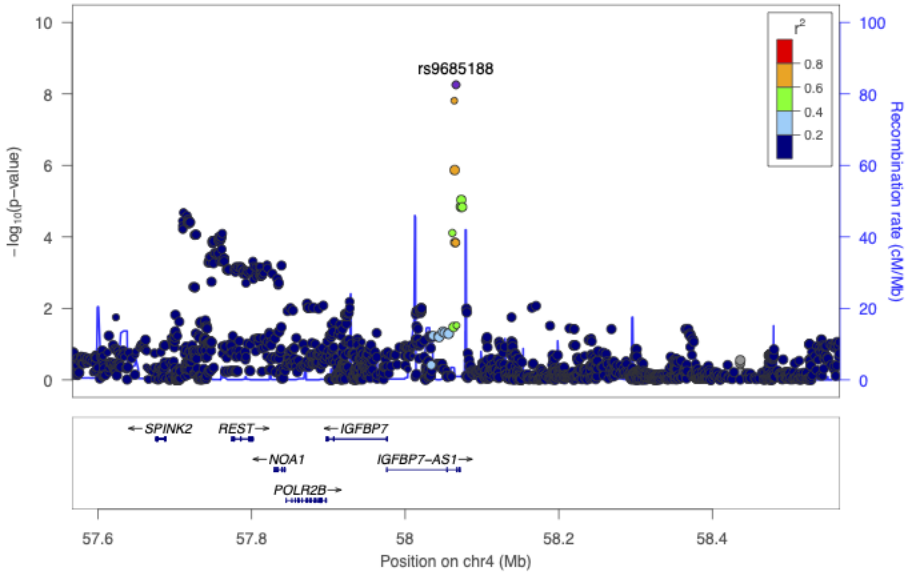
34_rs73157913



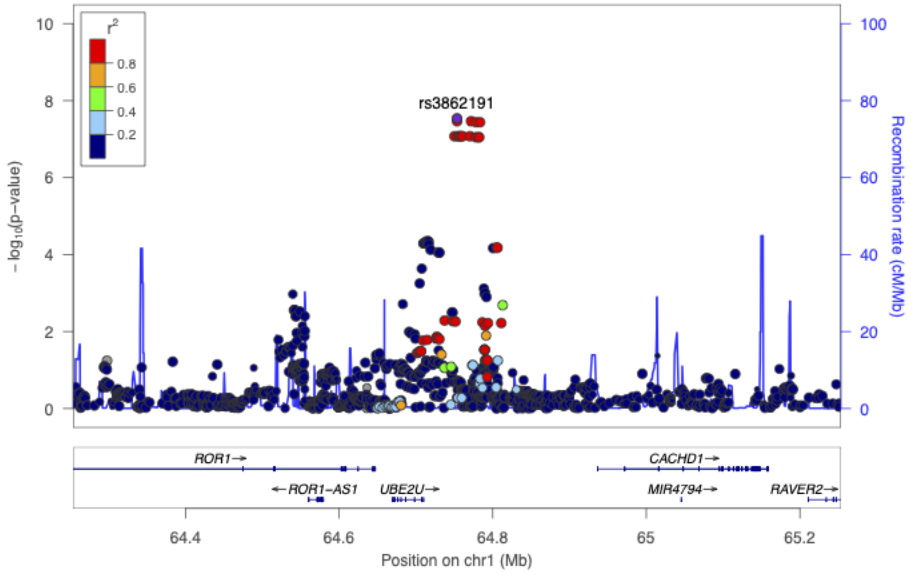
35_rs11923408



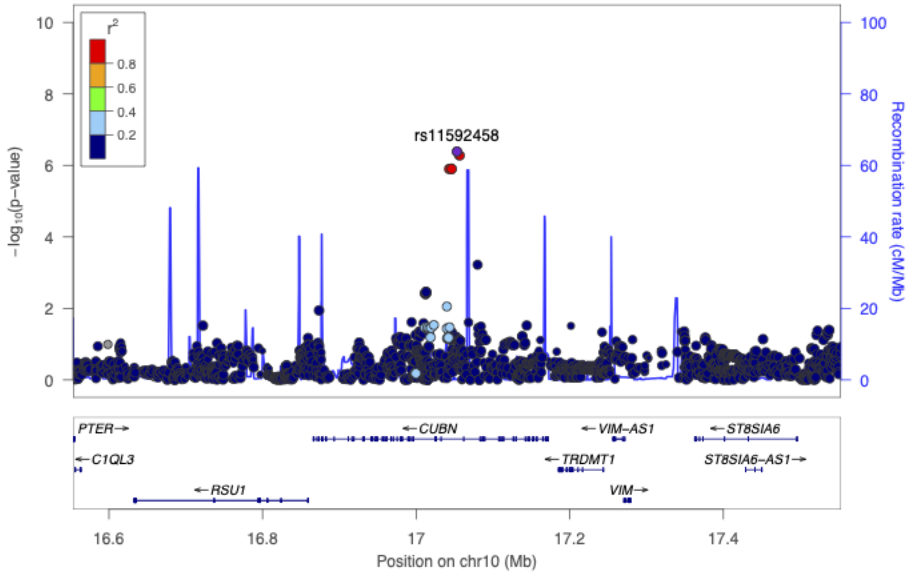
36_rs9685188

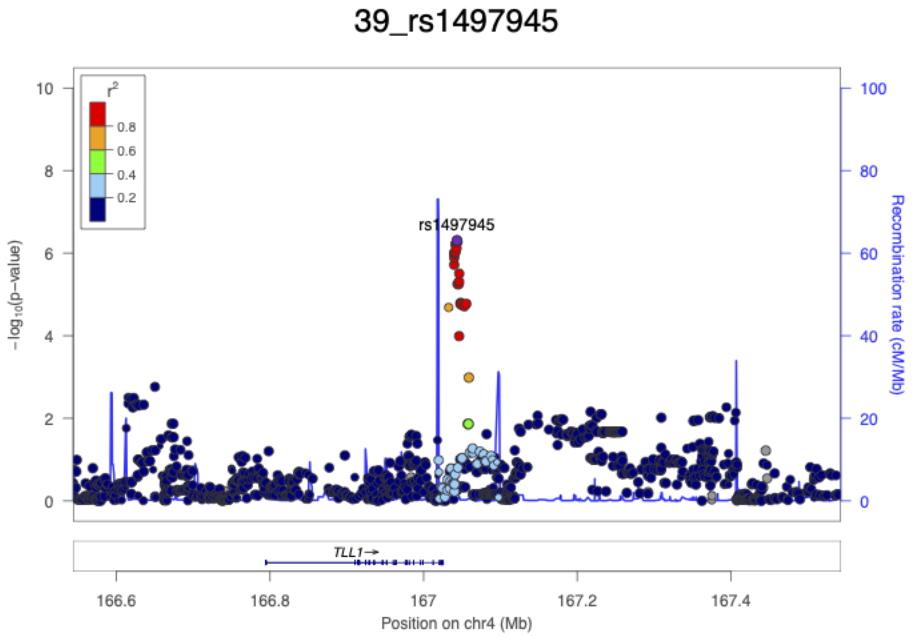


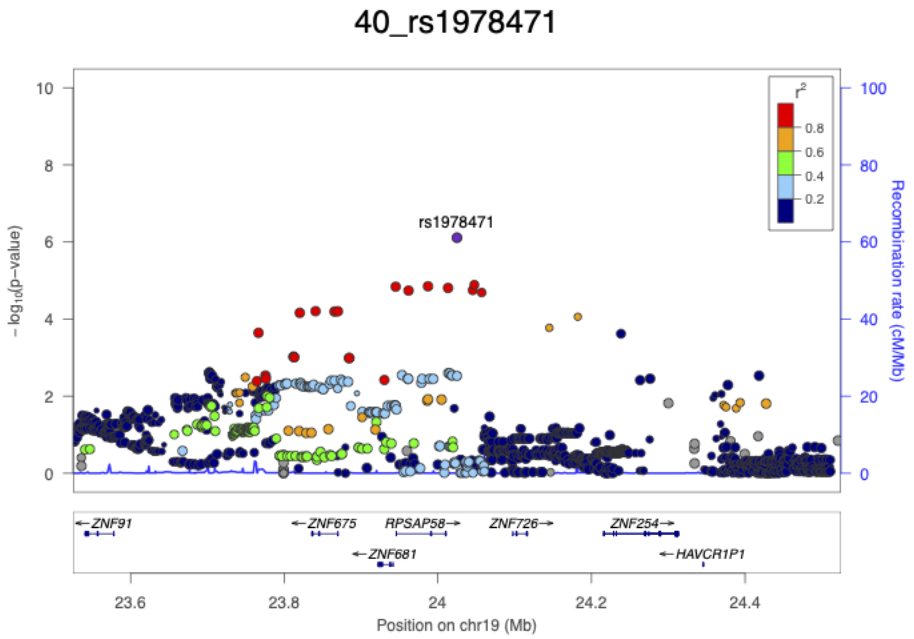
37_rs3862191



38_rs11592458







Among the 40 top SNPs, 1-11 were from IFS, 12-20 were from COHAR1, 21-35 were from COHRA2 and 36-40 were from meta-analysis

Each point represents the significance (–log10-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 [r2]) of the SNP with the lead SNP. Right Y-axis (recombination rate)

provides information about the LD structure of region. Genes near the association signal are shown on the bottom.

Appendix Table 3 Top SNPs from meta-analysis and their results in three cohorts

						Meta-analysis				IFS				COHRA1				COHRA2			
SNP	CHR	BP	freq	A1	A2	N	Zscore	P	Direction	Beta	SE	P	N	Beta	SE	P	N	Beta	SE	P	N
rs9685188	4	58066166	0.308	t	c	1446	5.829	5.57E-09	+++	0.275	0.076	0.00028	396	0.084	0.084	0.314	325	0.268	0.055	1.07E-06	725
rs3862191	1	64753505	0.173	t	c	1486	5.547	2.90E-08	+++	0.125	0.094	0.183	396	0.126	0.101	0.213	327	0.435	0.073	2.41E-09	763
rs11592458	10	17053179	0.116	c	g	1487	5.064	4.10E-07	+++	0.328	0.110	0.0027	396	0.286	0.110	0.009	327	0.244	0.076	0.0013	764
rs1497945	4	167043469	0.354	a	t	1496	5.03	4.91E-07	+++	0.221	0.073	0.0027	396	0.246	0.082	0.0027	328	0.155	0.053	0.0037	772
rs1978471	19	24024926	0.176	t	c	1489	4.942	7.74E-07	+++	0.362	0.092	7.78E-05	396	0.420	0.099	2.21E-05	328	0.084	0.066	0.202	765

Bibliography

Abbasi, F., Kodani, M., Emori, C., Kiyozumi, D., Mori, M., Fujihara, Y., & Ikawa, M. (2020). CRISPR/Cas9-Mediated Genome Editing Reveals Oosp Family Genes are Dispensable for Female Fertility in Mice. *Cells, 9*(4). doi:10.3390/cells9040821

Anjomshoaa, I., Briseño-Ruiz, J., Deeley, K., Poletta, F. A., Mereb, J. C., Leite, A. L., . . . Vieira, A. R. (2015). Aquaporin 5 Interacts with Fluoride and Possibly Protects against Caries. *PLoS One, 10*(12), e0143068. doi:10.1371/journal.pone.0143068

Antunes, L. S., Tannure, P. N., Antunes, L. A., Reis, M. F., Costa, M. C., Gouvêa, C. V., . . . Küchler, E. C. (2014). Genetic association for caries susceptibility among cleft lip and/or palate individuals. *J Contemp Dent Pract, 15*(3), 288-293. doi:10.5005/jp-journals-10024-1530

Avila, W. M., Pordeus, I. A., Paiva, S. M., & Martins, C. C. (2015). Breast and Bottle Feeding as Risk Factors for Dental Caries: A Systematic Review and Meta-Analysis. *PLoS One, 10*(11), e0142922. doi:10.1371/journal.pone.0142922

Ballantine, J. L., Carlson, J. C., Ferreira Zandoná, A. G., Agler, C., Zeldin, L. P., Rozier, R. G., . . . Divaris, K. (2018). Exploring the genomic basis of early childhood caries: a pilot study. *Int J Paediatr Dent, 28*(2), 217-225. doi:10.1111/ipd.12344

Barbato, E., Traversa, A., Guarnieri, R., Giovannetti, A., Genovesi, M. L., Magliozzi, M. R., . . . Caputo, V. (2018). Whole exome sequencing in an Italian family with isolated maxillary canine agenesis and canine eruption anomalies. *Arch Oral Biol, 91*, 96-102. doi:10.1016/j.archoralbio.2018.04.011

Barrera, M. J., Sánchez, M., Aguilera, S., Alliende, C., Bahamondes, V., Molina, C., . . . González, M. J. (2012). Aberrant localization of fusion receptors involved in regulated exocytosis in salivary glands of Sjögren's syndrome patients is linked to ectopic mucin secretion. *J Autoimmun, 39*(1-2), 83-92. doi:10.1016/j.jaut.2012.01.011

Bartlett, M. S., & Fowler, R. H. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences, 160*(901), 268-282. doi:10.1098/rspa.1937.0109

Bernabé, E., Vehkalahti, M. M., Sheiham, A., Aromaa, A., & Suominen, A. L. (2014). Sugar-sweetened beverages and dental caries in adults: a 4-year prospective study. *J Dent, 42*(8), 952-958. doi:10.1016/j.jdent.2014.04.011

Blostein, F. A., Jansen, E. C., Jones, A. D., Marshall, T. A., & Foxman, B. (2020). Dietary patterns associated with dental caries in adults in the United States. *Community Dent Oral Epidemiol, 48*(2), 119-129. doi:10.1111/cdoe.12509

Boraas, J. C., Messer, L. B., & Till, M. J. (1988). A genetic contribution to dental caries, occlusion, and morphology as demonstrated by twins reared apart. *J Dent Res, 67*(9), 1150-1155. doi:10.1177/00220345880670090201

Bretz, W. A., Corby, P. M., Hart, T. C., Costa, S., Coelho, M. Q., Weyant, R. J., . . . Schork, N. J. (2005). Dental caries and microbial acid production in twins. *Caries Res, 39*(3), 168-172. doi:10.1159/000084793

Bretz, W. A., Corby, P. M., Melo, M. R., Coelho, M. Q., Costa, S. M., Robinson, M., . . . Hart, T. C. (2006). Heritability estimates for dental caries and sucrose sweetness preference. *Arch Oral Biol, 51*(12), 1156-1160. doi:10.1016/j.archoralbio.2006.06.003

Bretz, W. A., Corby, P. M., Schork, N. J., Robinson, M. T., Coelho, M., Costa, S., . . . Hart, T. C. (2005). Longitudinal analysis of heritability for dental caries traits. *J Dent Res, 84*(11), 1047-1051. doi:10.1177/154405910508401115

Briseño-Ruiz, J., Shimizu, T., Deeley, K., Dizak, P. M., Ruff, T. D., Faraco, I. M., Jr., . . . Vieira, A. R. (2013). Role of TRAV locus in low caries experience. *Hum Genet, 132*(9), 1015-1025. doi:10.1007/s00439-013-1313-4

Broffitt, B., Levy, S. M., Warren, J., & Cavanaugh, J. E. (2013). Factors associated with surface-level caries incidence in children aged 9 to 13: the Iowa Fluoride Study. *J Public Health Dent, 73*(4), 304-310. doi:10.1111/jphd.12028

Brown, A. A., Buil, A., Viñuela, A., Lappalainen, T., Zheng, H. F., Richards, J. B., . . . Durbin, R. (2014). Genetic interactions affecting human gene expression identified by variance association mapping. *Elife, 3*, e01381. doi:10.7554/eLife.01381

Brown, M. B., & Forsythe, A. B. (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association, 69*(346), 364-367. doi:10.1080/01621459.1974.10482955

Buzalaf, M. A. R., Ortiz, A. C., Carvalho, T. S., Fideles, S. O. M., Araújo, T. T., Moraes, S. M., . . . Reis, F. N. (2020). Saliva as a diagnostic tool for dental caries, periodontal disease and cancer: is there a need for more biomarkers? *Expert Rev Mol Diagn, 20*(5), 543-555. doi:10.1080/14737159.2020.1743686

Cao, Y., Wei, P., Bailey, M., Kauwe, J. S. K., & Maxwell, T. J. (2014). A versatile omnibus test for detecting mean and variance heterogeneity. *Genet Epidemiol, 38*(1), 51-59. doi:10.1002/gepi.21778

Chankanka, O., Cavanaugh, J. E., Levy, S. M., Marshall, T. A., Warren, J. J., Broffitt, B., & Kolker, J. L. (2011). Longitudinal associations between children's dental caries and risk factors. *J Public Health Dent, 71*(4), 289-300. doi:10.1111/j.1752-7325.2011.00271.x

Chaussain, C., Bouazza, N., Gasse, B., Laffont, A. G., Opsahl Vital, S., Davit-Béal, T., . . . Sire, J. Y. (2014). Dental caries and enamelin haplotype. *J Dent Res, 93*(4), 360-365. doi:10.1177/0022034514522060

Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., . . . Lin, X. (2016). Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum Genet, 98*(4), 653-666. doi:10.1016/j.ajhg.2016.02.012

Cho, K. W., Cho, S. W., Lee, J. M., Lee, M. J., Gang, H. S., & Jung, H. S. (2008). Expression of phosphorylated forms of ERK, MEK, PTEN and PI3K in mouse oral development. *Gene Expr Patterns, 8*(4), 284-290. doi:10.1016/j.gep.2007.12.001

Conry, J. P., Messer, L. B., Boraas, J. C., Aeppli, D. P., & Bouchard, T. J., Jr. (1993). Dental caries and treatment characteristics in human twins reared apart. *Arch Oral Biol, 38*(11), 937-943. doi:10.1016/0003-9969(93)90106-v

Corrêa-Faria, P., Paixão-Gonçalves, S., Paiva, S. M., & Pordeus, I. A. (2016). Incidence of dental caries in primary dentition and risk factors: a longitudinal study. *Braz Oral Res, 30*(1). doi:10.1590/1807-3107BOR-2016.vol30.0059

Cortellazzi, K. L., Tagliaferro, E. P., Pereira, S. M., Ambrosano, G. M., Guerra, L. M., de Vazquez, F., . . . Pereira, A. C. (2013). A cohort study of caries incidence and baseline socioeconomic, clinical and demographic variables: a Kaplan-Meier survival analysis. *Oral Health Prev Dent, 11*(4), 349-358. doi:10.3290/j.ohpd.a30480

Corty, R. W., & Valdar, W. (2018). QTL Mapping on a Background of Variance Heterogeneity. *G3 (Bethesda), 8*(12), 3767-3782. doi:10.1534/g3.118.200790

Cunha, A. S., Vertuan Dos Santos, L., Schaffer Pugsley Baratto, S., Abbasoglu, Z., Gerber, J. T., Paza, A., . . . Küchler, E. C. (2021). Human permanent tooth sizes are associated with genes encoding oestrogen receptors. *J Orthod, 48*(1), 24-32. doi:10.1177/1465312520958710

Cury, J. A., & Tenuta, L. M. (2014). Evidence-based recommendation on toothpaste use. *Braz Oral Res, 28 Spec No*, 1-7. doi:10.1590/s1806-83242014.50000001

Dalledone, M., Cunha, A. S., Ramazzotto, L. A., Pecharki, G. D., Nelson-Filho, P., Scariot, R., . . . Brancher, J. A. (2019). Estrogen receptor gene is associated with dental fluorosis in Brazilian children. *Clin Oral Investig, 23*(9), 3565-3570. doi:10.1007/s00784-018-2778-2

Damanaki, A., Nokhbehsaim, M., Eick, S., Götz, W., Winter, J., Wahl, G., . . . Deschner, J. (2014). Regulation of NAMPT in human gingival fibroblasts and biopsies. *Mediators Inflamm, 2014*, 912821. doi:10.1155/2014/912821

Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A. E., Kwong, A., . . . Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nat Genet, 48*(10), 1284-1287. doi:10.1038/ng.3656

De Coster, P. J., Martens, L. C., & De Paepe, A. (2005). Oral health in prevalent types of Ehlers-Danlos syndromes. *J Oral Pathol Med, 34*(5), 298-307. doi:10.1111/j.1600-0714.2004.00300.x

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol, 11*(4), e1004219. doi:10.1371/journal.pcbi.1004219

De Vendittis, A., Amato, M., Mickniewicz, A., Parlato, G., De Angelis, A., Castellano, I., . . . De Vendittis, E. (2010). Regulation of the properties of superoxide dismutase from the dental pathogenic microorganism Streptococcus mutans by iron- and manganese-bound co-factor. *Mol Biosyst, 6*(10), 1973-1982. doi:10.1039/c003557b

Deeley, K., Letra, A., Rose, E. K., Brandon, C. A., Resick, J. M., Marazita, M. L., & Vieira, A. R. (2008). Possible association of amelogenin to high caries experience in a Guatemalan-Mayan population. *Caries Res, 42*(1), 8-13. doi:10.1159/000111744

Doetzer, A. D., Brancher, J. A., Pecharki, G. D., Schlipf, N., Werneck, R., Mira, M. T., . . . Trevilatto, P. C. (2015). Lactotransferrin Gene Polymorphism Associated with Caries Experience. *Caries Res, 49*(4), 370-377. doi:10.1159/000366211

Du, Y., Fu, X., Jin, J., Li, Z., Xu, K., Guo, M., . . . Ba, Y. (2021). Effects of SNPs in SOD2 and SOD3 interacted with fluoride exposure on the susceptibility of dental fluorosis. *Int J Hyg Environ Health, 239*, 113879. doi:10.1016/j.ijheh.2021.113879

Ergöz, N., Seymen, F., Gencay, K., Tamay, Z., Deeley, K., Vinski, S., & Vieira, A. R. (2014). Genetic variation in Ameloblastin is associated with caries in asthmatic children. *Eur Arch Paediatr Dent, 15*(3), 211-216. doi:10.1007/s40368-013-0096-6

Fan, R., Zhang, Y., Albert, P. S., Liu, A., Wang, Y., & Xiong, M. (2012). Longitudinal association analysis of quantitative traits. *Genet Epidemiol, 36*(8), 856-869. doi:10.1002/gepi.21673

Fatemifar, G., Hoggart, C. J., Paternoster, L., Kemp, J. P., Prokopenko, I., Horikoshi, M., . . . Evans, D. M. (2013). Genome-wide association study of primary tooth eruption identifies pleiotropic loci associated with height and craniofacial distances. *Hum Mol Genet, 22*(18), 3807-3817. doi:10.1093/hmg/ddt231

Feldens, C. A., Rodrigues, P. H., de Anastácio, G., Vítolo, M. R., & Chaffee, B. W. (2018). Feeding frequency in infancy and dental caries in childhood: a prospective cohort study. *Int Dent J, 68*(2), 113-121. doi:10.1111/idj.12333

Fine, D. H., Toruner, G. A., Velliyagounder, K., Sampathkumar, V., Godboley, D., & Furgang, D. (2013). A lactotransferrin single nucleotide polymorphism demonstrates biological activity that can reduce susceptibility to caries. *Infect Immun, 81*(5), 1596-1605. doi:10.1128/iai.01063-12

Fleming, E., & Afful, J. (2018). Prevalence of Total and Untreated Dental Caries Among Youth: United States, 2015-2016. *NCHS Data Brief*(307), 1-8.

Fligner, M. A., & Killeen, T. J. (1976). Distribution-Free Two-Sample Tests for Scale. *Journal of the American Statistical Association, 71*(353), 210-213. doi:10.1080/01621459.1976.10481517

Fontana, M., Young, D. A., & Wolff, M. S. (2009). Evidence-based caries, risk assessment, and treatment. *Dent Clin North Am, 53*(1), 149-161, x. doi:10.1016/j.cden.2008.10.003

Forsberg, S. K. G., & Carlborg, Ö. (2017). On the relationship between epistasis and genetic variance heterogeneity. *J Exp Bot, 68*(20), 5431-5438. doi:10.1093/jxb/erx283

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw, 33*(1), 1-22.

Gannon, K. P., McKey, S. E., Stec, D. E., & Drummond, H. A. (2015). Altered myogenic vasoconstriction and regulation of whole kidney blood flow in the ASIC2 knockout mouse. *Am J Physiol Renal Physiol, 308*(4), F339-348. doi:10.1152/ajprenal.00572.2014

Gao, S., Li, X., & Amendt, B. A. (2013). Understanding the role of Tbx1 as a candidate gene for 22q11.2 deletion syndrome. *Curr Allergy Asthma Rep, 13*(6), 613-621. doi:10.1007/s11882-013-0384-6

Gao, S., Moreno, M., Eliason, S., Cao, H., Li, X., Yu, W., . . . Amendt, B. A. (2015). 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, 24*(8), 2330-2348. doi:10.1093/hmg/ddu750

Geller, F., Feenstra, B., Zhang, H., Shaffer, J. R., Hansen, T., Esserlind, A. L., . . . Melbye, M. (2011). Genome-wide association study identifies four loci associated with eruption of permanent teeth. *PLoS Genet, 7*(9), e1002275. doi:10.1371/journal.pgen.1002275

Ghazal, T., Levy, S. M., Childers, N. K., Broffitt, B., Cutter, G. R., Wiener, H. W., . . . Cavanaugh, J. E. (2015). Factors associated with early childhood caries incidence among high caries-risk children. *Community dentistry and oral epidemiology, 43*(4), 366-374. doi:10.1111/cdoe.12161

Ghazal, T. S., Levy, S. M., Childers, N. K., Carter, K. D., Caplan, D. J., Warren, J. J., & Kolker, J. L. (2019). Survival analysis of caries incidence in African-American school-aged children. *J Public Health Dent, 79*(1), 10-17. doi:10.1111/jphd.12289

Giunta, C., Nuytinck, L., Raghunath, M., Hausser, I., De Paepe, A., & Steinmann, B. (2002). Homozygous Gly530Ser substitution in COL5A1 causes mild classical Ehlers-Danlos syndrome. *Am J Med Genet, 109*(4), 284-290. doi:10.1002/ajmg.10373

Govil, M., Mukhopadhyay, N., Weeks, D. E., Feingold, E., Shaffer, J. R., Levy, S. M., . . . Marazita, M. L. (2018). Novel caries loci in children and adults implicated by genome-wide analysis of families. *BMC Oral Health, 18*(1), 98. doi:10.1186/s12903-018-0559-6

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W., . . . Pasaniuc, B. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet, 48*(3), 245-252. doi:10.1038/ng.3506

Hagberg, C., Berglund, B., Korpe, L., & Andersson-Norinder, J. (2004). Ehlers-Danlos Syndrome (EDS) focusing on oral symptoms: a questionnaire study. *Orthod Craniofac Res, 7*(3), 178-185. doi:10.1111/j.1601-6343.2004.00288.x

Haworth, S., Esberg, A., Lif Holgerson, P., Kuja-Halkola, R., Timpson, N. J., Magnusson, P. K. E., . . . Johansson, I. (2020). Heritability of Caries Scores, Trajectories, and Disease Subtypes. *J Dent Res, 99*(3), 264-270. doi:10.1177/0022034519897910

Haworth, S., Shungin, D., van der Tas, J. T., Vucic, S., Medina-Gomez, C., Yakimov, V., . . . Timpson, N. J. (2018). Consortium-based genome-wide meta-analysis for childhood dental caries traits. *Hum Mol Genet, 27*(17), 3113-3127. doi:10.1093/hmg/ddy237

Heaton, B., Crawford, A., Garcia, R. I., Henshaw, M., Riedy, C. A., Barker, J. C., & Wimsatt, M. A. (2017). Oral health beliefs, knowledge, and behaviors in Northern California American Indian and Alaska Native mothers regarding early childhood caries. *J Public Health Dent, 77*(4), 350-359. doi:10.1111/jphd.12217

Heyman, M. B., & Abrams, S. A. (2017). Fruit Juice in Infants, Children, and Adolescents: Current Recommendations. *Pediatrics, 139*(6). doi:10.1542/peds.2017-0967

Hooley, M., Skouteris, H., Boganin, C., Satur, J., & Kilpatrick, N. (2012). Parental influence and the development of dental caries in children aged 0-6 years: a systematic review of the literature. *J Dent, 40*(11), 873-885. doi:10.1016/j.jdent.2012.07.013

Horio, N., Yoshida, R., Yasumatsu, K., Yanagawa, Y., Ishimaru, Y., Matsunami, H., & Ninomiya, Y. (2011). Sour taste responses in mice lacking PKD channels. *PLoS One, 6*(5), e20007. doi:10.1371/journal.pone.0020007

Hosoya, A., Shalehin, N., Takebe, H., Shimo, T., & Irie, K. (2020). Sonic Hedgehog Signaling and Tooth Development. *Int J Mol Sci, 21*(5). doi:10.3390/ijms21051587

Hu, X. P., Song, T. Z., Zhu, Y. Y., Wu, L. L., Zhang, X., Zhou, J. Y., & Li, Z. Q. (2019). Association of ENAM, TUFT1, MMP13, IL1B, IL10 and IL1RN gene polymorphism and dental caries susceptibility in Chinese children. *J Int Med Res, 47*(4), 1696-1704. doi:10.1177/0300060519828450

Hughes, T. E., Bockmann, M. R., Seow, K., Gotjamanos, T., Gully, N., Richards, L. C., & Townsend, G. C. (2007). Strong genetic control of emergence of human primary incisors. *J Dent Res, 86*(12), 1160-1165. doi:10.1177/154405910708601204

Indelicato, E., Nachbauer, W., Fauth, C., Krabichler, B., Schossig, A., Eigentler, A., . . . Boesch, S. (2019). SYNE1-ataxia: Novel genotypic and phenotypic findings. *Parkinsonism Relat Disord, 62*, 210-214. doi:10.1016/j.parkreldis.2018.12.007

Ishibashi, T., Yokota, T., Satoh, Y., Ichii, M., Sudo, T., Doi, Y., . . . Kanakura, Y. (2018). Identification of MS4A3 as a reliable marker for early myeloid differentiation in human hematopoiesis. *Biochem Biophys Res Commun, 495*(3), 2338-2343. doi:10.1016/j.bbrc.2017.12.117

Jágr, M., Eckhardt, A., Pataridis, S., & Mikšík, I. (2012). Comprehensive proteomic analysis of human dentin. *Eur J Oral Sci, 120*(4), 259-268. doi:10.1111/j.1600-0722.2012.00977.x

Jain, M., Namdev, R., Bodh, M., Dutta, S., Singhal, P., & Kumar, A. (2015). Social and Behavioral Determinants for Early Childhood Caries among Preschool Children in India. *J Dent Res Dent Clin Dent Prospects, 9*(2), 115-120. doi:10.15171/joddd.2014.023

Jedeon, K., Houari, S., Loiodice, S., Thuy, T. T., Le Normand, M., Berdal, A., & Babajko, S. (2016). Chronic Exposure to Bisphenol A Exacerbates Dental Fluorosis in Growing Rats. *J Bone Miner Res, 31*(11), 1955-1966. doi:10.1002/jbmr.2879

Jedeon, K., Loiodice, S., Salhi, K., Le Normand, M., Houari, S., Chaloyard, J., . . . Babajko, S. (2016). Androgen Receptor Involvement in Rat Amelogenesis: An Additional Way for Endocrine-Disrupting Chemicals to Affect Enamel Synthesis. *Endocrinology, 157*(11), 4287-4296. doi:10.1210/en.2016-1342

Jeremias, F., Koruyucu, M., Küchler, E. C., Bayram, M., Tuna, E. B., Deeley, K., . . . Vieira, A. R. (2013). Genes expressed in dental enamel development are associated with molar-incisor hypomineralization. *Arch Oral Biol, 58*(10), 1434-1442. doi:10.1016/j.archoralbio.2013.05.005

Jerome, L. A., & Papaioannou, V. E. (2001). DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nat Genet, 27*(3), 286-291. doi:10.1038/85845

Kang, S. W., Yoon, I., Lee, H. W., & Cho, J. (2011). Association between AMELX polymorphisms and dental caries in Koreans. *Oral Dis, 17*(4), 399-406. doi:10.1111/j.1601-0825.2010.01766.x

Kapferer-Seebacher, I., Schnabl, D., Zschocke, J., & Pope, F. M. (2020). Dental Manifestations of Ehlers-Danlos Syndromes: A Systematic Review. *Acta Derm Venereol, 100*(7), adv00092. doi:10.2340/00015555-3428

Kassebaum, N. J., Bernabé, E., Dahiya, M., Bhandari, B., Murray, C. J., & Marcenes, W. (2015). Global burden of untreated caries: a systematic review and metaregression. *J Dent Res, 94*(5), 650-658. doi:10.1177/0022034515573272

Kato, T., Yorifuji, T., Yamakawa, M., Inoue, S., Saito, K., Doi, H., & Kawachi, I. (2015). Association of breast feeding with early childhood dental caries: Japanese population-based study. *BMJ Open, 5*(3), e006982. doi:10.1136/bmjopen-2014-006982

Kaye, E. K., Heaton, B., Sohn, W., Rich, S. E., Spiro, A., 3rd, & Garcia, R. I. (2015). The Dietary Approaches to Stop Hypertension Diet and New and Recurrent Root Caries Events in Men. *J Am Geriatr Soc, 63*(9), 1812-1819. doi:10.1111/jgs.13614

Kelekar, U., & Naavaal, S. (2018). Hours Lost to Planned and Unplanned Dental Visits Among US Adults. *Prev Chronic Dis, 15*, E04. doi:10.5888/pcd15.170225

Kim, J. C., Kim, S. Y., Roh, S. A., Cho, D. H., Kim, D. D., Kim, J. H., & Kim, Y. S. (2008). Gene expression profiling: canonical molecular changes and clinicopathological features in sporadic colorectal cancers. *World J Gastroenterol, 14*(43), 6662-6672. doi:10.3748/wjg.14.6662

Küchler, E. C., Deeley, K., Ho, B., Linkowski, S., Meyer, C., Noel, J., . . . Vieira, A. R. (2013). Genetic mapping of high caries experience on human chromosome 13. *BMC Med Genet, 14*, 116. doi:10.1186/1471-2350-14-116

Küchler, E. C., Feng, P., Deeley, K., Fitzgerald, C. A., Meyer, C., Gorbunov, A., . . . Vieira, A. R. (2014). Fine mapping of locus Xq25.1-27-2 for a low caries experience phenotype. *Arch Oral Biol, 59*(5), 479-486. doi:10.1016/j.archoralbio.2014.02.009

Kumar, S., Tadakamadla, J., & Johnson, N. W. (2016). Effect of Toothbrushing Frequency on Incidence and Increment of Dental Caries: A Systematic Review and Meta-Analysis. *J Dent Res, 95*(11), 1230-1236. doi:10.1177/0022034516655315

Lamon, S., Wallace, M. A., Léger, B., & Russell, A. P. (2009). Regulation of STARS and its downstream targets suggest a novel pathway involved in human skeletal muscle hypertrophy and atrophy. *J Physiol, 587*(Pt 8), 1795-1803. doi:10.1113/jphysiol.2009.168674

Lee, H. J., Kim, J. B., Jin, B. H., Paik, D. I., & Bae, K. H. (2015). Risk factors for dental caries in childhood: a five-year survival analysis. *Community Dent Oral Epidemiol, 43*(2), 163-171. doi:10.1111/cdoe.12136

Lee, M. G., Ohana, E., Park, H. W., Yang, D., & Muallem, S. (2012). Molecular mechanism of pancreatic and salivary gland fluid and HCO3 secretion. *Physiol Rev, 92*(1), 39-74. doi:10.1152/physrev.00011.2011

Leroy, R., Bogaerts, K., Martens, L., & Declerck, D. (2012). Risk factors for caries incidence in a cohort of Flemish preschool children. *Clin Oral Investig, 16*(3), 805-812. doi:10.1007/s00784-011-0579-y

Levanti, M., Randazzo, B., Viña, E., Montalbano, G., Garcia-Suarez, O., Germanà, A., . . . Abbate, F. (2016). Acid-sensing ion channels and transient-receptor potential ion channels in zebrafish taste buds. *Ann Anat, 207*, 32-37. doi:10.1016/j.aanat.2016.06.006

Li, Z. Q., Hu, X. P., Zhou, J. Y., Xie, X. D., & Zhang, J. M. (2015). Genetic polymorphisms in the carbonic anhydrase VI gene and dental caries susceptibility. *Genet Mol Res, 14*(2), 5986-5993. doi:10.4238/2015.June.1.16

Lim, S., Tellez, M., & Ismail, A. I. (2015). Dental caries development among African American children: results from a 4-year longitudinal study. *Community dentistry and oral epidemiology, 43*(3), 200-207. doi:10.1111/cdoe.12140

Lin, Z., Zeng, J., & Wang, X. (2019). Compound phenotype of osteogenesis imperfecta and Ehlers-Danlos syndrome caused by combined mutations in COL1A1 and COL5A1. *Biosci Rep, 39*(7). doi:10.1042/bsr20181409

Liska, D., Kelley, M., & Mah, E. (2019). 100% Fruit Juice and Dental Health: A Systematic Review of the Literature. *Front Public Health, 7*, 190. doi:10.3389/fpubh.2019.00190

Locher, H., de Groot, J. C., van Iperen, L., Huisman, M. A., Frijns, J. H., & Chuva de Sousa Lopes, S. M. (2015). Development of the stria vascularis and potassium regulation in the human fetal cochlea: Insights into hereditary sensorineural hearing loss. *Dev Neurobiol, 75*(11), 1219-1240. doi:10.1002/dneu.22279

Locker, D. (1992). The burden of oral disorders in a population of older adults. *Community Dent Health, 9*(2), 109-124.

Lula, E. C., Ribeiro, C. C., Hugo, F. N., Alves, C. M., & Silva, A. A. (2014). Added sugars and periodontal disease in young adults: an analysis of NHANES III data. *Am J Clin Nutr, 100*(4), 1182-1187. doi:10.3945/ajcn.114.089656

Mak, K. M., Png, C. Y., & Lee, D. J. (2016). Type V Collagen in Health, Disease, and Fibrosis. *Anat Rec (Hoboken), 299*(5), 613-629. doi:10.1002/ar.23330

Marderstein, A. R., Davenport, E. R., Kulm, S., Van Hout, C. V., Elemento, O., & Clark, A. G. (2021). Leveraging phenotypic variability to identify genetic interactions in human phenotypes. *Am J Hum Genet, 108*(1), 49-67. doi:10.1016/j.ajhg.2020.11.016

Martignon, S., Roncalli, A. G., Alvarez, E., Aránguiz, V., Feldens, C. A., & Buzalaf, M. A. R. (2021). Risk factors for dental caries in Latin American and Caribbean countries. *Braz Oral Res, 35*(suppl 01), e053. doi:10.1590/1807-3107bor-2021.vol35.0053

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., . . . Durbin, R. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet, 48*(10), 1279-1283. doi:10.1038/ng.3643

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., . . . Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol, 28*(5), 495-501. doi:10.1038/nbt.1630

Mei, H., Chen, W., Jiang, F., He, J., Srinivasan, S., Smith, E. N., . . . Berenson, G. S. (2012). Longitudinal replication studies of GWAS risk SNPs influencing body mass index over the course of childhood and adulthood. *PLoS One, 7*(2), e31470. doi:10.1371/journal.pone.0031470

Mejàre, I., Axelsson, S., Dahlén, G., Espelid, I., Norlund, A., Tranæus, S., & Twetman, S. (2014). Caries risk assessment. A systematic review. *Acta Odontol Scand, 72*(2), 81-91. doi:10.3109/00016357.2013.822548

Meng, Y., Wu, T., Billings, R., Kopycka-Kedzierawski, D. T., & Xiao, J. (2019). Human genes influence the interaction between Streptococcus mutans and host caries susceptibility: a genome-wide association study in children with primary dentition. *Int J Oral Sci, 11*(2), 19. doi:10.1038/s41368-019-0051-4

Meurman, P. K., & Pienihäkkinen, K. (2010). Factors associated with caries increment: a longitudinal study from 18 months to 5 years of age. *Caries Res, 44*(6), 519-524. doi:10.1159/000320717

Moon, J. S., Nam, Y. S., Kang, J. H., Yang, D. W., Kim, D. Y., Lee, S. Y., . . . Kim, S. H. (2021). Regulatory role of insulin-like growth factor-binding proteins in odontogenic mineralization in rats. *J Mol Histol, 52*(1), 63-75. doi:10.1007/s10735-020-09923-3

Morrison, J., Laurie, C. C., Marazita, M. L., Sanders, A. E., Offenbacher, S., Salazar, C. R., . . . Shaffer, J. R. (2016). Genome-wide association study of dental caries in the Hispanic Communities Health Study/Study of Latinos (HCHS/SOL). *Hum Mol Genet, 25*(4), 807-816. doi:10.1093/hmg/ddv506

Nariyama, M., Shimizu, K., Uematsu, T., & Maeda, T. (2004). Identification of chromosomes associated with dental caries susceptibility using quantitative trait locus analysis in mice. *Caries Res, 38*(2), 79-84. doi:10.1159/000075929

National Health Expenditures 2019 Highlights. ([accessed 2020]). Retrieved from https://www.cms.gov/files/document/highlights.pdf

Navarra, C. O., Robino, A., Pirastu, N., Bevilacqua, L., Gasparini, P., Di Lenarda, R., & Crovella, S. (2016). Caries and Innate Immunity: DEFB1 Gene Polymorphisms and Caries Susceptibility in Genetic Isolates from North-Eastern Italy. *Caries Res, 50*(6), 589-594. doi:10.1159/000450965

Neiswanger, K., McNeil, D. W., Foxman, B., Govil, M., Cooper, M. E., Weyant, R. J., . . . Marazita, M. L. (2015). Oral Health in a Sample of Pregnant Women from Northern Appalachia (2011-2015). *Int J Dent, 2015*, 469376. doi:10.1155/2015/469376

Nirunsittirat, A., Pitiphat, W., McKinney, C. M., DeRouen, T. A., Chansamak, N., Angwaravong, O., . . . Pimpak, T. (2016). Breastfeeding Duration and Childhood Caries: A Cohort Study. *Caries Res, 50*(5), 498-507. doi:10.1159/000448145

Ohta, M., Nishimura, H., & Asada, Y. (2015). Association of DLX3 gene polymorphism and dental caries susceptibility in Japanese children. *Arch Oral Biol, 60*(1), 55-61. doi:10.1016/j.archoralbio.2014.08.020

Oliver, P. L., Finelli, M. J., Edwards, B., Bitoun, E., Butts, D. L., Becker, E. B., . . . Davies, K. E. (2011). Oxr1 is essential for protection against oxidative stress-induced neurodegeneration. *PLoS Genet, 7*(10), e1002338. doi:10.1371/journal.pgen.1002338

Ollila, P., & Larmas, M. (2007). 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, 65*(1), 29-35. doi:10.1080/00016350600963590

Orlova, E., Carlson, J. C., Lee, M. K., Feingold, E., McNeil, D. W., Crout, R. J., . . . Shaffer, J. R. (2019). Pilot GWAS of caries in African-Americans shows genetic heterogeneity. *BMC Oral Health, 19*(1), 215. doi:10.1186/s12903-019-0904-4

Östberg, A. L., & Petzold, M. (2020). A longitudinal study of the impact of change in socioeconomic status on dental caries in the permanent dentition of Swedish children and adolescents. *Community dentistry and oral epidemiology, 48*(4), 271-279. doi:10.1111/cdoe.12529

Park, H. J., Kim, M. K., Kim, S. R., Bae, S. K., & Bae, M. K. (2013). Hypoxia regulates the expression of the neuromedin B receptor through a mechanism dependent on hypoxia-inducible factor-1α. *PLoS One, 8*(12), e82868. doi:10.1371/journal.pone.0082868

Patir, A., Seymen, F., Yildirim, M., Deeley, K., Cooper, M. E., Marazita, M. L., & Vieira, A. R. (2008). Enamel formation genes are associated with high caries experience in Turkish children. *Caries Res, 42*(5), 394-400. doi:10.1159/000154785

Peltzer, K., Mongkolchati, A., Satchaiyan, G., Rajchagool, S., & Pimpak, T. (2014). Sociobehavioral factors associated with caries increment: a longitudinal study from 24 to 36 months old children in Thailand. *Int J Environ Res Public Health, 11*(10), 10838-10850. doi:10.3390/ijerph111010838

Perera, P. J., Fernando, M. P., Warnakulasooriya, T. D., & Ranathunga, N. (2014). Effect of feeding practices on dental caries among preschool children: a hospital based analytical cross sectional study. *Asia Pac J Clin Nutr, 23*(2), 272-277. doi:10.6133/apjcn.2014.23.2.13

Peres, M. A., Liu, P., Demarco, F. F., Silva, A. E. R., Wehrmeister, F. C., Menezes, A. M., & Peres, K. G. (2018). Income trajectories affect treatment of dental caries from childhood to young adulthood: a birth cohort study. *Braz Oral Res, 32*, e36. doi:10.1590/1807-3107bor-2018.vol32.0036

Peres, M. A., Peres, K. G., de Barros, A. J., & Victora, C. G. (2007). The relation between family socioeconomic trajectories from childhood to adolescence and dental caries and associated oral behaviours. *J Epidemiol Community Health, 61*(2), 141-145. doi:10.1136/jech.2005.044818

Peres, M. A., Peres, K. G., Thomson, W. M., Broadbent, J. M., Gigante, D. P., & Horta, B. L. (2011). The influence of family income trajectories from birth to adulthood on adult oral health: findings from the 1982 Pelotas birth cohort. *Am J Public Health, 101*(4), 730-736. doi:10.2105/ajph.2009.184044

Pillas, D., Hoggart, C. J., Evans, D. M., O'Reilly, P. F., Sipilä, K., Lähdesmäki, R., . . . Jarvelin, M. R. (2010). Genome-wide association study reveals multiple loci associated with primary tooth development during infancy. *PLoS Genet, 6*(2), e1000856. doi:10.1371/journal.pgen.1000856

Pitts, N. B., Zero, D. T., Marsh, P. D., Ekstrand, K., Weintraub, J. A., Ramos-Gomez, F., . . . Ismail, A. (2017). Dental caries. *Nat Rev Dis Primers, 3*, 17030. doi:10.1038/nrdp.2017.30

Poulton, R., Caspi, A., Milne, B. J., Thomson, W. M., Taylor, A., Sears, M. R., & Moffitt, T. E. (2002). Association between children's experience of socioeconomic disadvantage and adult health: a life-course study. *Lancet, 360*(9346), 1640-1645. doi:10.1016/s0140-6736(02)11602-3

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., . . . Willer, C. J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics, 26*(18), 2336-2337. doi:10.1093/bioinformatics/btq419

Pyati, S. A., Naveen Kumar, R., Kumar, V., Praveen Kumar, N. H., & Parveen Reddy, K. M. (2018). Salivary Flow Rate, pH, Buffering Capacity, Total Protein, Oxidative Stress and Antioxidant Capacity in Children with and without Dental Caries. *J Clin Pediatr Dent, 42*(6), 445-449. doi:10.17796/1053-4625-42.6.7

Qadri, G., Nourallah, A., & Splieth, C. H. (2012). Early childhood caries and feeding practices in kindergarten children. *Quintessence Int, 43*(6), 503-510.

Ramadugu, K., Blostein, F., Bhaumik, D., Jiang, W., Davis, E., Salzman, E., . . . Foxman, B. (2020). Co-occurrence of yeast, streptococci, dental decay, and gingivitis in the post-partum period: results of a longitudinal study. *J Oral Microbiol, 12*(1), 1746494. doi:10.1080/20002297.2020.1746494

Rath-Deschner, B., Nogueira, A. V. B., Memmert, S., Nokhbehsaim, M., Augusto Cirelli, J., Eick, S., . . . Damanaki, A. (2021). Regulation of Anti-Apoptotic SOD2 and BIRC3 in Periodontal Cells and Tissues. *Int J Mol Sci, 22*(2). doi:10.3390/ijms22020591

Rizvi, A. A., Karaesmen, E., Morgan, M., Preus, L., Wang, J., Sovic, M., . . . Sucheston-Campbell, L. E. (2019). gwasurvivr: an R package for genome-wide survival analysis. *Bioinformatics, 35*(11), 1968-1970. doi:10.1093/bioinformatics/bty920

Roediger, M., Miró, X., Geffers, R., Irmer, M., Huels, A., Miosge, N., & Gersdorff, N. (2009). Profiling of Differentially Expressed Genes in Peri-implantitis and Periodontitis in vivo by Microarray Analysis. *Journal of Oral Biosciences, 51*(1), 31-45. doi:https://doi.org/10.1016/S1349-0079(09)80018-2

Romanos, H. F., Antunes, L. S., Lopes, L. B., Sabóia Tde, M., Tannure, P. N., Lips, A., . . . Küchler, E. C. (2015). BMP2 Is Associated with Caries Experience in Primary Teeth. *Caries Res, 49*(4), 425-433. doi:10.1159/000371715

Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2010). Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *Genet Sel Evol, 42*(1), 8. doi:10.1186/1297-9686-42-8

Rönnegård, L., & Valdar, W. (2012). Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genet, 13*, 63. doi:10.1186/1471-2156-13-63

Rugg-Gunn, A. J., Roberts, G. J., & Wright, W. G. (1985). Effect of human milk on plaque pH in situ and enamel dissolution in vitro compared with bovine milk, lactose, and sucrose. *Caries Res, 19*(4), 327-334. doi:10.1159/000260863

Rugg-Gunn, A. J., Spencer, A. J., Whelton, H. P., Jones, C., Beal, J. F., Castle, P., . . . Zusman, S. P. (2016). Critique of the review of 'Water fluoridation for the prevention of dental caries' published by the Cochrane Collaboration in 2015. *Br Dent J, 220*(7), 335-340. doi:10.1038/sj.bdj.2016.257

Saha, D., Goswami, R., Majumdar, K. K., Sikdar, N., & Pramanik, S. (2021). 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, 199*(1), 1-8. doi:10.1007/s12011-020-02116-9

Schwendicke, F., Dörfer, C. E., Schlattmann, P., Foster Page, L., Thomson, W. M., & Paris, S. (2015). Socioeconomic inequality and caries: a systematic review and meta-analysis. *J Dent Res, 94*(1), 10-18. doi:10.1177/0022034514557546

Selwitz, R. H., Ismail, A. I., & Pitts, N. B. (2007). Dental caries. *Lancet, 369*(9555), 51-59. doi:10.1016/s0140-6736(07)60031-2

Shaffer, J. R., Carlson, J. C., Stanley, B. O., Feingold, E., Cooper, M., Vanyukov, M. M., . . . Marazita, M. L. (2015). Effects of enamel matrix genes on dental caries are moderated by fluoride exposures. *Hum Genet, 134*(2), 159-167. doi:10.1007/s00439-014-1504-7

Shaffer, J. R., Feingold, E., Wang, X., Lee, M., Tcuenco, K., Weeks, D. E., . . . Marazita, M. L. (2013). GWAS of dental caries patterns in the permanent dentition. *J Dent Res, 92*(1), 38-44. doi:10.1177/0022034512463579

Shaffer, J. R., Feingold, E., Wang, X., Tcuenco, K. T., Weeks, D. E., DeSensi, R. S., . . . Marazita, M. L. (2012). Heritable patterns of tooth decay in the permanent dentition: principal components and factor analyses. *BMC Oral Health, 12*, 7. doi:10.1186/1472-6831-12-7

Shaffer, J. R., Feingold, E., Wang, X., Weeks, D. E., Weyant, R. J., Crout, R., . . . Marazita, M. L. (2013). Clustering tooth surfaces into biologically informative caries outcomes. *J Dent Res, 92*(1), 32-37. doi:10.1177/0022034512463241

Shaffer, J. R., Li, J., Lee, M. K., Roosenboom, J., Orlova, E., Adhikari, K., . . . Weinberg, S. M. (2017). Multiethnic GWAS Reveals Polygenic Architecture of Earlobe Attachment. *Am J Hum Genet, 101*(6), 913-924. doi:10.1016/j.ajhg.2017.10.001

Shaffer, J. R., Polk, D. E., Wang, X., Feingold, E., Weeks, D. E., Lee, M. K., . . . Marazita, M. L. (2014). Genome-wide association study of periodontal health measured by probing depth in adults ages 18-49 years. *G3 (Bethesda), 4*(2), 307-314. doi:10.1534/g3.113.008755

Shaffer, J. R., Wang, X., Desensi, R. S., Wendell, S., Weyant, R. J., Cuenco, K. T., . . . Marazita, M. L. (2012). Genetic susceptibility to dental caries on pit and fissure and smooth surfaces. *Caries Res, 46*(1), 38-46. doi:10.1159/000335099

Shaffer, J. R., Wang, X., Feingold, E., Lee, M., Begum, F., Weeks, D. E., . . . Marazita, M. L. (2011). Genome-wide association scan for childhood caries implicates novel genes. *J Dent Res, 90*(12), 1457-1462. doi:10.1177/0022034511422910

Shaffer, J. R., Wang, X., McNeil, D. W., Weyant, R. J., Crout, R., & Marazita, M. L. (2015). Genetic susceptibility to dental caries differs between the sexes: a family-based study. *Caries Res, 49*(2), 133-140. doi:10.1159/000369103

Shimizu, T., Deeley, K., Briseño-Ruiz, J., Faraco, I. M., Jr., Poletta, F. A., Brancher, J. A., . . . Vieira, A. R. (2013). Fine-mapping of 5q12.1-13.3 unveils new genetic contributors to caries. *Caries Res, 47*(4), 273-283. doi:10.1159/000346278

Shimizu, T., Ho, B., Deeley, K., Briseño-Ruiz, J., Faraco, I. M., Jr., Schupack, B. I., . . . Vieira, A. R. (2012). Enamel formation genes influence enamel microhardness before and after cariogenic challenge. *PLoS One, 7*(9), e45022. doi:10.1371/journal.pone.0045022

Shin, D. H., Kim, M., Kim, Y., Jun, I., Jung, J., Nam, J. H., . . . Lee, M. G. (2020). Bicarbonate permeation through anion channels: its role in health and disease. *Pflugers Arch, 472*(8), 1003-1018. doi:10.1007/s00424-020-02425-x

Shungin, D., Haworth, S., Divaris, K., Agler, C. S., Kamatani, Y., Keun Lee, M., . . . Johansson, I. (2019). Genome-wide analysis of dental caries and periodontitis combining clinical and self-reported data. *Nat Commun, 10*(1), 2773. doi:10.1038/s41467-019-10630-1

Sikorska, K., Rivadeneira, F., Groenen, P. J., Hofman, A., Uitterlinden, A. G., Eilers, P. H., & Lesaffre, E. (2013). Fast linear mixed model computations for genome-wide association studies with longitudinal data. *Stat Med, 32*(1), 165-180. doi:10.1002/sim.5517

Slade, G. D., Grider, W. B., Maas, W. R., & Sanders, A. E. (2018). Water Fluoridation and Dental Caries in U.S. Children and Adolescents. *J Dent Res, 97*(10), 1122-1128. doi:10.1177/0022034518774331

Smith, E. N., Chen, W., Kähönen, M., Kettunen, J., Lehtimäki, T., Peltonen, L., . . . Murray, S. S. (2010). Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet, 6*(9), e1001094. doi:10.1371/journal.pgen.1001094

Solé-Magdalena, A., Revuelta, E. G., Menénez-Díaz, I., Calavia, M. G., Cobo, T., García-Suárez, O., . . . Vega, J. A. (2011). Human odontoblasts express transient receptor protein and acid-sensing ion channel mechanosensor proteins. *Microsc Res Tech, 74*(5), 457-463. doi:10.1002/jemt.20931

Song, N., Choi, J. Y., Sung, H., Jeon, S., Chung, S., Park, S. K., . . . Kang, D. (2015). Prediction of breast cancer survival using clinical and genetic markers by tumor subtypes. *PLoS One, 10*(4), e0122413. doi:10.1371/journal.pone.0122413

Speed, D., & Balding, D. J. (2019). SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat Genet, 51*(2), 277-284. doi:10.1038/s41588-018-0279-5

Symoens, S., Syx, D., Malfait, F., Callewaert, B., De Backer, J., Vanakker, O., . . . De Paepe, A. (2012). Comprehensive molecular analysis demonstrates type V collagen mutations in over 90% of patients with classic EDS and allows to refine diagnostic criteria. *Hum Mutat, 33*(10), 1485-1493. doi:10.1002/humu.22137

Tan, S. H. X., Teo, Y. Y., Tan, M. H. X., & Gao, X. (2021). Childhood Factors and Dental Caries in the Permanent Dentition: Findings of an 8-Year Study Under a Nationwide School Dental Service. *Int Dent J, 71*(6), 508-515. doi:10.1016/j.identj.2021.01.008

Tang, S., Pan, Y., Wang, Y., Hu, L., Cao, S., Chu, M., . . . Shen, H. (2015). Genome-wide association study of survival in early-stage non-small cell lung cancer. *Ann Surg Oncol, 22*(2), 630-635. doi:10.1245/s10434-014-3983-0

Tannure, P. N., Küchler, E. C., Falagan-Lotsch, P., Amorim, L. M., Raggio Luiz, R., Costa, M. C., . . . Granjeiro, J. M. (2012). MMP13 polymorphism decreases risk for dental caries. *Caries Res, 46*(4), 401-407. doi:10.1159/000339379

Tannure, P. N., Küchler, E. C., Lips, A., Costa Mde, C., Luiz, R. R., Granjeiro, J. M., & Vieira, A. R. (2012). Genetic variation in MMP20 contributes to higher caries experience. *J Dent, 40*(5), 381-386. doi:10.1016/j.jdent.2012.01.015

ten Cate, J. M., & Featherstone, J. D. (1991). Mechanistic aspects of the interactions between fluoride and dental enamel. *Crit Rev Oral Biol Med, 2*(3), 283-296. doi:10.1177/10454411910020030101

Thomson, W. M., Poulton, R., Milne, B. J., Caspi, A., Broughton, J. R., & Ayers, K. M. (2004). Socioeconomic inequalities in oral health in childhood and adulthood in a birth cohort. *Community dentistry and oral epidemiology, 32*(5), 345-353. doi:10.1111/j.1600-0528.2004.00173.x

Troidl, K., Rüding, I., Cai, W. J., Mücke, Y., Grossekettler, L., Piotrowska, I., . . . Troidl, C. (2009). Actin-binding rho activating protein (Abra) is essential for fluid shear stress-induced arteriogenesis. *Arterioscler Thromb Vasc Biol, 29*(12), 2093-2101. doi:10.1161/atvbaha.109.195305

Vaché, C., Puechberty, J., Faugère, V., Darmaisin, F., Liquori, A., Baux, D., . . . Roux, A. F. (2020). A 4.6 Mb Inversion Leading to PCDH15-LINC00844 and BICC1-PCDH15 Fusion Transcripts as a New Pathogenic Mechanism Implicated in Usher Syndrome Type 1. *Front Genet, 11*, 623. doi:10.3389/fgene.2020.00623

Valaitis, R., Hesch, R., Passarelli, C., Sheehan, D., & Sinton, J. (2000). A systematic review of the relationship between breastfeeding and early childhood caries. *Can J Public Health, 91*(6), 411-417. doi:10.1007/bf03404819

Vieira, A. R., Bayram, M., Seymen, F., Sencak, R. C., Lippert, F., & Modesto, A. (2017). In Vitro Acid-Mediated Initial Dental Enamel Loss Is Associated with Genetic Variants Previously Linked to Caries Experience. *Front Physiol, 8*, 104. doi:10.3389/fphys.2017.00104

Waldman, H. B. (1987). Another perspective on children's dental needs and demand for services during the 1980s. *ASDC J Dent Child, 54*(5), 344-348.

Wang, H., Zhang, F., Zeng, J., Wu, Y., Kemper, K. E., Xue, A., . . . Yang, J. (2019). Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci Adv, 5*(8), eaaw3538. doi:10.1126/sciadv.aaw3538

Wang, J., Massoudi, D., Ren, Y., Muir, A. M., Harris, S. E., Greenspan, D. S., & Feng, J. Q. (2017). BMP1 and TLL1 Are Required for Maintaining Periodontal Homeostasis. *J Dent Res, 96*(5), 578-585. doi:10.1177/0022034516686558

Wang, X., Shaffer, J. R., Weyant, R. J., Cuenco, K. T., DeSensi, R. S., Crout, R., . . . Marazita, M. L. (2010). Genes and their effects on dental caries may differ between primary and permanent dentitions. *Caries Res, 44*(3), 277-284. doi:10.1159/000314676

Wang, X., Shaffer, J. R., Zeng, Z., Begum, F., Vieira, A. R., Noel, J., . . . Marazita, M. L. (2012). Genome-wide association scan of dental caries in the permanent dentition. *BMC Oral Health, 12*, 57. doi:10.1186/1472-6831-12-57

Ward, D. M., Vaughn, M. B., Shiflett, S. L., White, P. L., Pollock, A. L., Hill, J., . . . Kaplan, J. (2005). The role of LIP5 and CHMP5 in multivesicular body formation and HIV-1 budding in mammalian cells. *J Biol Chem, 280*(11), 10548-10555. doi:10.1074/jbc.M413734200

Warren, J. J., Levy, S. M., & Kanellis, M. J. (2002). Dental caries in the primary dentition: assessing prevalence of cavitated and noncavitated lesions. *J Public Health Dent, 62*(2), 109-114. doi:10.1111/j.1752-7325.2002.tb03430.x

Warren, J. J., Weber-Gasparoni, K., Marshall, T. A., Drake, D. R., Dehkordi-Vakil, F., Dawson, D. V., & Tharp, K. M. (2009). A longitudinal study of dental caries risk among very young low SES children. *Community dentistry and oral epidemiology, 37*(2), 116-122. doi:10.1111/j.1600-0528.2008.00447.x

Weber, M. L., Hsin, H. Y., Kalay, E., BroŽková, D. S., Shimizu, T., Bayram, M., . . . Vieira, A. R. (2014). Role of estrogen related receptor beta (ESRRB) in DFN35B hearing impairment and dental decay. *BMC Med Genet, 15*, 81. doi:10.1186/1471-2350-15-81

Wendell, S., Wang, X., Brown, M., Cooper, M. E., DeSensi, R. S., Weyant, R. J., . . . Marazita, M. L. (2010). Taste genes associated with dental caries. *J Dent Res, 89*(11), 1198-1202. doi:10.1177/0022034510381502

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics, 26*(17), 2190-2191. doi:10.1093/bioinformatics/btq340

Wu, C., Kraft, P., Stolzenberg-Solomon, R., Steplowski, E., Brotzman, M., Xu, M., . . . Wolpin, B. M. (2014). Genome-wide association study of survival in patients with pancreatic adenocarcinoma. *Gut, 63*(1), 152-160. doi:10.1136/gutjnl-2012-303477

Xu, Z., Shen, X., & Pan, W. (2014). Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes. *PLoS One, 9*(8), e102312. doi:10.1371/journal.pone.0102312

Yan, Q., Ding, Y., Liu, Y., Sun, T., Fritsche, L. G., Clemons, T., . . . Chen, W. (2018). Genome-wide analysis of disease progression in age-related macular degeneration. *Hum Mol Genet, 27*(5), 929-940. doi:10.1093/hmg/ddy002

Yang, J., Loos, R. J., Powell, J. E., Medland, S. E., Speliotes, E. K., Chasman, D. I., . . . Visscher, P. M. (2012). FTO genotype is associated with phenotypic variability of body mass index. *Nature, 490*(7419), 267-272. doi:10.1038/nature11401

Yang, J., Zhu, W., Chen, J., Zhang, Q., & Wu, S. (2014). Genome-wide two-marker linkage disequilibrium mapping of quantitative trait loci. *BMC Genet, 15*, 20. doi:10.1186/1471-2156-15-20

Yang, S., Ryu, J. H., Oh, H., Jeon, J., Kwak, J. S., Kim, J. H., . . . Chun, J. S. (2015). NAMPT (visfatin), a direct target of hypoxia-inducible factor-2α, is an essential catabolic regulator of osteoarthritis. *Ann Rheum Dis, 74*(3), 595-602. doi:10.1136/annrheumdis-2013-204355

Yildiz, G., Ermis, R. B., Calapoglu, N. S., Celik, E. U., & Türel, G. Y. (2016). Gene-environment Interactions in the Etiology of Dental Caries. *J Dent Res, 95*(1), 74-79. doi:10.1177/0022034515605281

Yoon, Y., Kim, T. J., Lee, J. M., & Kim, D. Y. (2018). SOD2 is upregulated in periodontitis to reduce further inflammation progression. *Oral Dis, 24*(8), 1572-1580. doi:10.1111/odi.12933

Zeng, Z., Feingold, E., Wang, X., Weeks, D. E., Lee, M., Cuenco, D. T., . . . Shaffer, J. R. (2014). Genome-wide association study of primary dentition pit-and-fissure and smooth surface caries. *Caries Res, 48*(4), 330-338. doi:10.1159/000356299

Zeng, Z., Shaffer, J. R., Wang, X., Feingold, E., Weeks, D. E., Lee, M., . . . Marazita, M. L. (2013). Genome-wide association studies of pit-and-fissure- and smooth-surface caries in permanent dentition. *J Dent Res, 92*(5), 432-437. doi:10.1177/0022034513481976

Zhang, F., Chen, W., Zhu, Z., Zhang, Q., Nabais, M. F., Qi, T., . . . Yang, J. (2019). OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol, 20*(1), 107. doi:10.1186/s13059-019-1718-z

Zhang, H., Jani, P., Liang, T., Lu, Y., & Qin, C. (2017). Inactivation of bone morphogenetic protein 1 (Bmp1) and tolloid-like 1 (Tll1) in cells expressing type I collagen leads to dental and periodontal defects in mice. *J Mol Histol, 48*(2), 83-98. doi:10.1007/s10735-016-9708-x

Zhang, S., Divaris, K., Moss, K., Yu, N., Barros, S., Marchesan, J., . . . Offenbacher, S. (2016). The Novel ASIC2 Locus is Associated with Severe Gingival Inflammation. *JDR Clin Trans Res, 1*(2), 163-170. doi:10.1177/2380084416645290

Zhu, X., Yi, K., Hou, D., Huang, H., Jiang, X., Shi, X., & Xing, C. (2020). Clinicopathological Analysis and Prognostic Assessment of Transcobalamin I (TCN1) in Patients with Colorectal Tumors. *Med Sci Monit, 26*, e923828. doi:10.12659/msm.923828