GENETIC ANALYSES OF DENTAL ANOMALIES AND DENTAL CARIES IN

MULTIETHNIC POPULATIONS

by

Rasha N. Alotaibi

BS, King Saud University, 2007

MS, Old Dominion University, 2013

Submitted to the Graduate Faculty of

THE SCHOOL OF DENTAL MEDICINE in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

SCHOOL OF DENTAL MEDICINE

This dissertation was presented

by

Rasha N. Alotaibi

It was defended on

October 31, 2019

and approved by

John R. Shaffer, Ph.D., Assistant Professor,
Department of Human Genetics, Graduate School of Public Health
Department of Oral Biology, School of Dental Medicine
University of Pittsburgh

Seth Weinberg, Ph.D., Assistant Professor,
Department of Oral Biology, School of Dental Medicine
University of Pittsburgh

Katherine Neiswanger, Ph.D., Research Associate Professor,
Department of Oral Biology, School of Dental Medicine
University of Pittsburgh

Dissertation Director: Mary L. Marazita, Ph.D., Director, Center for Craniofacial and Dental
Genetics
Professor, Dept. of Oral Biology, School of Dental Medicine
Professor, Dept. of Human Genetics, Graduate School of Public Health
Professor, Dept. of Psychiatry and Dept. of Clinical and Translational Science,
School of Medicine
University of Pittsburgh

GENETIC ANALYSES OF DENTAL ANOMALIES AND DENTAL CARIES IN MULTIETHNIC POPULATIONS

Rasha Alotaibi, B.S, M.S, PhD

University of Pittsburgh, 2019

ABSTRACT

Tooth development is a long and complex process of several stages that starts between the sixth and eighth weeks in utero. Disruption during any of these stages can result in dental anomalies and/or increased risk of oral diseases such as dental caries. Previous studies indicate that genetic factors play an important role in the etiology of dental anomalies of size, shape, number, position, and of oral diseases including dental caries. Furthermore, previous studies have also shown that some dental anomalies are associated with each other in the same patient, suggesting that they may share common etiologic components.

This study evaluated the relationships between several structural dental anomalies and dental caries across the multi-ethnic cohort from The Pittsburgh Orofacial Cleft study (POFC; N= 3579), performed genome-wide association scans (GWAS) in POFC, including the multivariate approach of GWAS for the associated dental anomalies and dental caries, and performed GWAS for each of the included dental anomalies in POFC. Finally, we performed GWAS of dental caries in two cohorts; (POFC; N= 3579), and a cohort from The Center for Oral Health Research in Appalachia (COHRA1, N=1763), along with subsequent meta-analysis.

We found intercorrelations between four subsets of dental phenotypes: Agenesis, Impaction and Rotation (AIR); Hypoplasia, Displacement and Rotation (HDR); Displacement, Rotation and

Mamelons (DRM); and Dental caries, Agenesis and Hypoplasia (DAH). This was the first study to investigative genetic associations for multivariate patterns of correlated dental anomalies and dental caries, and we were able to identify suggestive association signals ($P < 1 \times 10^{-5}$) near genes with biological roles during tooth development; *ADAMTS9* and *PRICKLE2* were associated with AIR; *GLIS3*, *WDR72*, and *ROR2* were associated with HDR and DRM*; ROBO2* was associated with DRM*, BMP7* was associated with HDR*;* and *ROBO1, SMAD2* and *MSX2* were associated with DAH. We were also able to identify risk loci associated with each of the structural dental anomalies separately.

This is one of the largest multi-ethnic GWAS study of dental caries that helped in discovering several novel loci containing genes with plausible biological roles in tooth development and/or oral dental caries; such as *REL* gene ($P = 3.91E\text{-}09$) and were also able to confirm some of the previously identified genes *(MPPED2*, $P = 2.36E{-}6$; *NEED9*, $P = 7.60E\text{-}06$) that has been associated with dental caries.

Further studies are needed to replicate the results of this study in an independent cohort, which is strongly recommended. In addition, studies are needed to investigate the biological roles of these nominated genes, in human and animal models to fully understand their relevance to tooth development process and dental caries.

Table of Contents

List of Tables

List of Figures

Preface

I would like to express my thanks for Dr. Mary Marazita, my PhD dissertation Director. Dr. Marazita guided me throughout the years of my PhD training, which led to the conceptualization of this dissertation work. Your constructive comments have helped me a lot in shaping my research and have given me a great lesson. Indeed, this PhD dissertation would not be possible without her scientific contribution and her guidance.

I would like to express my thanks and gratitude for Dr. Katherine Neiswanger, she has been a great advisor and friend to me during my time at the Center for Craniofacial and Dental Genetics (CCDG). I admired her passion and caring about students.

I would also like to thank Dr. John Shaffer and Dr. Seth Weinberg for all of their ideas and advices during our meetings and for all of their support and encouragement throughout my study and research. Thanks to the CCDG team, including faculties, staff and students, for providing me with the resources that I needed to complete this dissertation, and for their help with my research.

I would also like to express my deepest thanks to my friend Ekaterina (Katya) Orlova. I am so grateful for having you in my life, thank you for your friendship, support, patience and encouragements.

A special thanks and gratitude goes to my dear sister Khlood, for her unending love and invaluable support from the start of my education. The PhD Journey would not have been bearable without her warm love, continued patience, and endless support.

Last, but most important, I would like to thank ALLAH the Almighty, for giving me the blessing, the strength, the chance and endurance to complete this dissertation and reach my goal. Without his blessings, this achievement would not have been possible.

1.0 OVERALL RESEARCH GOAL AND SPECIFIC AIMS

Different dental anomalies could negatively impact oral health quality of life because such anomalies might negatively affect the functionality of the dentition, and perhaps the aesthetic appearance of the person. Some dental anomalies seem to be associated in the same patient, perhaps because variants in some genes lead to different types of dental anomalies. One of the complications that can be caused by dental anomalies is increased susceptibility to oral diseases such as dental caries due to defects in tooth structure and/or crowding.   There is previous evidence that increased risk of dental caries might occur in teeth affected with enamel hypoplasia.

The central hypotheses of this study are: (i) that some <u>dental anomalies are highly correlated in the same subject, (ii) that enamel hypoplasia, one of the dental anomalies, could increase the risk of developing dental caries, and (iii) that there are genetic variants associated with dental anomalies, such as: tooth agenesis, enamel hypoplasia, correlated dental anomalies, and with dental caries.</u> To address these hypotheses, four specific aims were pursued:

Aim 1 Analyze the prevalence and the relationship among dental anomalies, and between dental caries and enamel hypoplasia across the multi-ethnic cohort from The Pittsburgh Orofacial Cleft study (POFC).

Aim 2 Conduct genome-wide association analyses (GWAS), a multivariate approach, in the POFC study cohort to identify common variants associated with the correlated dental anomalies and dental caries, and conduct genome-wide association analyses (GWAS) for each of the included dental anomalies separately.

Aim 3 Conduct genome-wide association analyses in two independent study samples POFC, and COHRA1 from the Center for Oral Health Research in Appalachia (COHRA) to identify common variants associated with dental caries.

Aim 4 Perform meta-analyses of dental caries GWAS, by combining the results from the two independent study cohorts (POFC, COHRA1).

This study will bring together two existing data sets: the COHRA1 population-based study population recruited from West Virginia and Pennsylvania, and unaffected controls and unaffected family members of orofacial cleft cases from the POFC study, recruited from the United States and from multiple international sites around the world.

2.0 INTRODUCTION

The process of tooth development for both primary and permanent teeth, is a complicated process that starts early in embryogenesis and involves signaling between epithelial and mesenchymal tissues. Both primary and permanent teeth develop from the oral ectoderm and the underlying neural mesenchymal cells, which migrate from the cranial neural crest to the facial processes. During tooth initiation starting in the sixth week of human development, the ectoderm thickens and forms the dental lamina that creates several buds which embed in the underlying neural-crest- derived mesenchyme. In each tooth bud, the epithelial compartment forms a cup-shaped structure, the enamel organ, with an inner part, the enamel reticulum, and an epithelial outer layer. During subsequent morphogenesis the enamel organ encompasses the dental papilla which is condensed mesenchymal tissue. During the differentiation stage, the mesenchymal cells of the dental papilla start to differentiate into odontoblasts and secrete dentin, the inner layer of teeth, while the epithelial cells adjacent to the differentiating odontoblasts start to differentiate into ameloblasts and secrete enamel, the outer layer of teeth. The central part of the dental papilla will form the dental pulp and the mesenchymal cells surrounding the enamel organ. The growing roots form the periodontal ligament, a fibrous structure that connects the tooth to the alveolar bone (Thesleff, 2003, Thesleff, et al., 1995).

Human and mice share a lot of similarities regarding tooth development, in fact, a mouse model was used for a long time to understand the molecular mechanisms underlying the tooth development process (Zhang, et al., 2005). During the stages of tooth development, in humans and mice, an interaction between the epithelium and mesenchymal cells occurs under the control of different families of signaling molecules and their receptors. These signaling families include transforming growth factor b (TGFb), bone morphogenetic proteins (BMP), fibroblast growth

factors (FGF), epidermal growth factor (EGF), and the hedgehog (Hh) and wingless (Wnt)

families (Hu, et al., 2013, Sarkar, et al.,2000). In addition, there are multiple genes that regulate

the communication and the interaction between epithelium and mesenchymal cells, such as

Msx1, Msx2, Pax9, Gli2, Gli3, Edar and *Runx2*. Disturbance during the signaling process or

mutations in any of the regulating genes may result in dental anomalies, including changes in

tooth development, structure, number, size, and morphology (Bailleul-Forestier, et al., 2008,

Thesleff, 2006).

2.1 STRUCTURAL DENTAL ANOMALIES

2.1.1 BACKGROUND

Structural dental anomalies such as tooth agenesis, enamel hypoplasia, impaction, rotation,

displacement, mamelons and supernumerary teeth (Table 1 provides definitions for those dental

anomalies) are presumed to be caused by interactions between genetic, epigenetic and

environmental factors during the process of tooth development (Brook, et al.,2009). Any

deficiencies in the mesenchymal tissue, blood supply, or in molecular signaling between the

dental lamina and the surrounding mesenchyme due to long-term disturbances in the oral

environment could cause structural dental anomalies. Dental anomalies can be associated with

other systemic skeletal or organic disorders (termed syndromic dental anomalies), i.e.

cleidocranial dysplasia, Gardener syndrome, or can be isolated (non-syndromic dental

anomalies) (Gallucci, et al., 2012). Dental anomalies affecting tooth structure can be classified

by the affected tissue (enamel versus dentin), and/or by their pattern of inheritance: autosomal

dominant (AD), autosomal recessive (AR), X-linked recessive (XLR), or polygenic modes of

inheritance (Bailleul-Forestier, et al., 2008).

Although dental anomalies might seem asymptomatic, these anomalies can lead to serious

clinical problems. Some of these problems include: delayed or non-eruption of the normal series

of teeth; breast feeding problems; temporomandibular joint pain and dysfunction; attrition, which

is the loss of tooth structure by mechanical forces from opposing teeth; malocclusion;

periodontal diseases because of excessive occlusal force; and increased susceptibility to dental

caries due to defects in tooth structure and/or crowding. All of these complications would affect

the aesthetic appearance of the person; speech ability and mastication; and the functionality of

the dentition all ultimately affecting the oral health-related quality of life of affected people,

especially children, in a negative way (Temilola, et al.,2014).

Table 1 Definitions of Structural Dental Anomalies to be Tested in the Study*

Dental Anomaly	Definition
Agenesis	The absence of one or more teeth due to the failure of any tooth to develop during embryonic growth and development due to the absence of primordial tissue.
Displaced	Any tilt of the long axis of the tooth either buccally or lingually from the normal anatomical position in the dental arch or in a way that it appears outside of the line of occlusion.
Enamel Hypoplasia	A quantitive defect that involves the enamel and associated with a reduced thickness of enamel. It can be localized or generalized and occur in the forms of pits, grooves, and striae.
Impacted	A tooth fails to erupt because of a positional deviation of its developing follicle or the presence of a physical barrier in its eruption path. In this study, the determination of impactions was used only for canines that were still not erupted even though permanent second molars were fully erupted.
Mamelons	Small tubercules that appear on newly erupted teeth due to the way they develop. They are often seen on the maxillary and mandibular incisors.
Rotated	Tooth turning along its long axis, either mesially or distally in relation to the line of occlusion.
Supernumerary	Presence of one or more teeth tooth additional to the normal number of teeth. Only erupted supernumeraries were counted in this study.

*Definitions of the dental anomalies are adapted from:

Howe BJ, Cooper ME, Vieira AR, et al. Spectrum of Dental Phenotypes in Nonsyndromic Orofacial Clefting. Journal of Dental Research. 2015;94(7):905-912. https://doi:10.1177/002203451558828

Dure-Molla, M., Fournier, B. P., Manzanares, M. C., Acevedo, A. C., Hennekam, R. C., … Bloch-Zupan, A. (2019). Elements of morphology: Standard terminology for the teeth and classifying genetic dental disorders. American Journal of Medical Genetics Part A. https://doi.org/10.1002/ajmg.a.61316

There are variations in the prevalence of structural dental anomalies in different populations, and there are several possible reasons behind these variations. These variations could arise because of the sample size, which in some studies is very small; age of the subjects in the study; design of the study; the study location and racial differences. In addition, different definitions of the

different dental anomalies could account for some of the variations in the results (Ezoddini, et al., 2009).

There have been several studies that investigated different dental anomalies in different ethnicities (Laganà, et al., 2017; Roslan, et al., 2018; Aren, et al.2015; Kuchler, et al., 2008; Cholitgul & Drummond, 2000; Temilola, et al., 2014). However, none of these studies succeeded in studying different dental anomalies in large multiethnic population like the population we have in our study. A summary of these studies is summarized in Table 2. The sample size in one of these studies (Laganà, et al., 2017) was the biggest (n=4706) compared to the other studies in table 1, and they found that tooth agenesis was the second most common dental anomaly (7.1%) in the whole sample and it was the most common one in the 9 years-old age group (8.4%). They also discovered associations among different dental anomalies, including an association between tooth agenesis and displacement of maxillary canines, and between tooth agenesis and tooth transposition. However, they focused only on young subjects, 8-12 years, so there is no information regarding the prevalence of these dental anomalies among older individual (Laganà, et al., 2017).

Most of the studies in Table 2 only focused on young subjects (Kuchler, et al., 2008; Cholitgul & Drummond, 2000, Temilola, et al., 2014, Laganà, et al., 2017). One study in Table 2 suffered from a small sample size (n=370), and they found out that impaction was the most common dental anomaly (14.32%), however, due to the small sample size, it is hard to make their findings representative of their studied population, moreover, they failed to report the age of the participants in the study (Roslan, et al., 2018). Most of these summarized studies in Table 2 had information regarding the medical and/or family history of the participants, except two studies (Aren, et al., 2015; Kuchler, et al., 2008) that failed to report if the participants have family

history of any medical conditions. It is really important to report the medical and family history of the subjects because several dental anomalies could be found in a subject as part of a syndrome.

Tooth agenesis was one of the most common dental anomalies (Aren, et al., 2015; Kuchler, et al., 2008; Laganà, et al., 2017), and its prevalence was 1.7%, 4.8%, and 7.1% respectively, and that actually falls in the range of the reported prevalence of tooth agenesis (1.6% to 9.6%) from the literature (Galluccio, et al., 2012; Zhang, et al., 2015). Enamel hypoplasia was the second most common dental anomaly, with prevalence of 16.1% (Temilola, et al., 2014), which is well within the range reported by the literature (Robles, et al., 2013).

It was mentioned earlier that there are several factors behind the variation in the prevalence of dental anomalies, including differences in ethnicities and differences in diagnostic methods. The variation in the prevalence of dental anomalies reported in these studies demonstrates the need to study structural dental anomalies in multiple diverse populations, including genetic studies, and to understand the reasons behind these variations across the different ethnicities.

Table 2 Summary of Dental Anomalies Studies

Study	Sample Size/Age	Ethnicity	Dental Anomalies investigated	Findings
Laganà, et al., 2017	4706 8-12 years	Caucasian	Supernumerary tooth, Impacted tooth, Tooth ankylosis, Odontomas, Taurodontism, Tooth transposition, Displacement of maxillary canine	Displacement of maxillary canine was the most frequent anomaly (7.5%), followed by tooth agenesis (7.1%). The most common dental anomalies associated with Supernumerary tooth was impaction and transposition.
Roslan, et al.,2018	370	Indian	Impaction, hypodontia, supernumerary, supraocclusion, infraocclusion, peg shaped tooth, dilacerations, enamel hypoplasia	Impaction was the most frequent anomaly (14.32%). There was significant association between dental anomalies and race.
Aren, et al.2015	2025 9-35 years	Turkey	Tooth agenesis, microdontia macrodontia, taurodontism and root abnormalities.	Tooth agenesis was the most common dental anomaly (1.77%)
Kuchler, et al.,2008	1198 6-12years	Brazil	Tooth agenesis, microdontia, peg-shaped teeth, supernumerary teeth, taurodontism, and transposition	4.8% of the subjects had tooth agenesis, which make it the most common dental anomaly
Cholitgul & Drummond, 2000	1607 10-15 years	New Zealand	Malpositioned teeth, Agenesis, misshaped teeth, and enamel hypoplasia.	The most frequent dental anomaly was teeth malplostioned
Temilola, et al., 2014).	1,036 4months-12years	Nigeria	Microdontia, Macrodontia, Fusion, Enamel hypoplasia, Dens evaginatus, Dens invaginatus, Supernumerary tooth, Hypodontia, Tooth transposition and other dental anomalies.	Enamel hypoplasia was the most frequent dental anomaly in this study (16.1%)

As we discussed earlier how the different stages of odontogenesis (tooth development process) are under the control of different signaling molecules and different genes and how any disturbance that occurs during the signaling process or mutations in any of the regulating genes may result in dental anomalies (Bailleul-Forestier, et al., 2008, Thesleff, 2006), we can see how genetics could influence the etiology of structural dental anomalies of number, size, position, and morphology. In addition, some dental anomalies seem to co-occur in the same patient, and this might happen because the same mutated gene or genetic polymorphism could lead to different manifestations of dental anomalies. Most of this evidence comes from family (Kurol & Bjerklin, 1982; Kurol, 1981; Vastardis, 2000), epidemiological (Baccetti, 1998; Becker, 1995; Garib, et al., 2009; Peck, et al., 1994; Peck, et al., 1998) and twin studies (Markovic, 1982 & Mossey, 1999).

2.1.2 TOOTH AGENESIS

Tooth agenesis (TA), the absence of one or more teeth due to developmental failure, is one of the most common dental anomalies (Dure-Molla, et al., 2019), see Table 2. TA can be classified as syndromic if there is the involvement of other organs, or as isolated/non-syndromic if only the dentition is affected. TA is also classified according to the number of missing teeth; hypodontia (less than 6 missing teeth other than third molars), oligodontia (6 and more missing teeth other than third molars) or the complete agenesis of all teeth i.e. anodontia (Matalova, et al., 2004). Previous epidemiological studies indicate that the prevalence of non-syndromic TA ranges from 1.6% to 9.6% in different geographic profiles and ethnicities (Galluccio, et al., 2012; Zhang, et al., 2015), with 1.37 times higher prevalence in females than in males (Polder, et al., 2004; Rakhshan, 2015). Further, previous studies indicated that TA is one of the structural dental

anomalies that appears to be frequently associated with other anomalies such as enamel hypoplasia, microdontia, ectopias, and delayed tooth development in the same individuals (Baccetti, 1998; Becker, 1995; Garib, et al., 2009; Peck, et al., 1994; Peck, et al., 1998; Vastardis, 2000).

2.1.3 ENAMEL HYPOPLASIA

Enamel hypoplasia can be defined as a quantitative defect associated with reduction in the thickness of the affected enamel. This defect usually occurs during the secretory stage of amelogenesis and can affect both dentitions (Suckling, 1989). Enamel hypoplasia will clinically appear as a shallow or deep fossa (i.e. depression on the tooth surface) with horizontal or vertical grooves and in some cases will cause a partial or complete absence of enamel. An increased risk of dental caries might occur in teeth affected with enamel hypoplasia (Li et al., 1996; Lai et al., 1997), because enamel hypoplasia may provide an enhanced environment for adhesion, colonization and retention of cariogenic bacteria. In combination with other caries risk factors, e.g. cariogenic diet and poor oral hygiene, dental caries may develop more rapidly (Li et al., 1996). In addition, hypoplastic enamel has higher acid solubility than normal enamel and is thus more susceptible to caries risk factors (Zheng et al., 1998; Hong, et al., 2009). Several studies have reported an association between enamel hypoplasia and dental caries (Milgrom et al., 2000; Montero et al., 2003; Daneshkazemi and Davari, 2005, Hong, et al., 2009). The prevalence of enamel hypoplasia varies depending on dentition, specific teeth, method of examination, and population studied (see Table 2). The prevalence of developmental defects of enamel, including

enamel hypoplasia, in developed countries and healthy children ranges between 24% to 49% in primary dentition, and between 9% to 63% in permanent teeth (Robles, et al., 2013).

2.1.4 OTHER DENTAL ANOMALIES

Impaction, is a condition where the tooth fails to erupt because of a positional deviation of its developing follicle or the presence of a physical barrier in its eruption path (Al-Abdallah, et al., 2018). A tooth will be classified as impacted when it remained in the jaw two years after the respective mean age of tooth eruption. The prevalence of impacted teeth, excluding third molars, has been reported to vary between 5.6 to 18.8% and the most frequently impacted teeth are the canines and second premolars (Al-Abdallah, et al., 2018 & Fardi, et al., 2011).

Supernumerary teeth can be defined as an extra tooth to the normal series of teeth and can be find in any region in both jaws. The most common supernumerary tooth is the one located in maxillary midline and is called a mesiodens. The prevalence of supernumerary teeth has been reported to range between 0.2–3% in the primary and permanent dentitions. Supernumerary teeth can cause some complications. One of these complications is tooth displacement and/or tooth rotation (Subasioglu, et al., 2015; Garvey, et al., 1999).

Tooth displacement, which is another dental anomaly that can exist as a complication of supernumerary teeth or alone, is a condition where the tooth is buccally or lingually deviated from the normal anatomical position in either jaw, or when the tooth appears outside of the line of occlusion (Howe, et al., 2015; Dure-Molla, et al., 2019).

Tooth rotation on the other hand is where the tooth deviated from the long axis of a tooth either mesially or distally in relation to the line of occlusion and it could be associated with tooth agenesis. Mamelons is a dental anomaly where there small tubercules on the newly erupted teeth.

This usually affects the incisors, either mandibular or maxillary, and they disappear or decrease in size as they with get worn away by mastication (Dure-Molla, et al., 2019).

2.1.5 PREVIOUS GENETIC STUDIES OF STRUCTURAL DENTAL ANOMALIES

To the best of our knowledge, no genome-wide association study (GWAS) for any of the previously descried dental anomalies in a diverse population has been reported. In addition, there are no studies using a multivariate approach for GWAS, which identified risk loci associated with multiple associated dental anomalies. There have been GWAS studies of tooth agenesis in one ethnicity, and they were focusing on agenesis of third molars only (Vukelic, et al., 2017; Haga et al., 2013). In a GWAS of tooth agenesis in European ancestry that used a sample from the POFC study for a replication, they were able to identify several risk variants near *ASCL5/CACNA1S, ARHGAP15, FOXI3, EDAR*, and *WNT10A* gene which were associated with tooth agenesis, excluding third molars (Jonsson, et al., 2018). Additionally, previous sequencing study of *WNT10A* indicated that nonsynonymous, nonsense, and missense mutations in *WNT10A* are associated with tooth agenesis of 1-3 teeth and also for 4 or more teeth (Arzoo, et al., 2014). A previous study also identified several genes that have been associated with non-syndromic forms of tooth agenesis, including *MSX1, PAX9, AXIN2*, and *EDA* (Matalova, et al.,2008); each of these genes play an important role during tooth development. It is notable that there have been GWASs of the timing of eruption of permanent and primary teeth (Geller, et al., 2011; Pillas, et al., 2010; Fatemifa, et al., 2013) that helped in identifying risk loci near *HMGA2* and *TNP1* and *ADK* genes. It is also mentionable that there was a recent GWAS in a Chinese sample that tested the associations between non-syndromic cleft lip with or without palate-susceptibility loci and

the occurrence of supernumerary teeth, more specifically the non-syndromic form of supernumerary teeth. They found out that there is an association between non-syndromic cleft lip with or without palate risk variants and supernumerary teeth (Kan, et al., 2019).

2.2 DENTAL CARIES

Dental caries is a common multifactorial disease, in which environment and genetics each play important roles. Dental caries is considered to be a worldwide public health problem and if left untreated can cause pain, infection, abscesses and loss of teeth (Edelstein BL, 2006). According to WHO, the prevalence of dental caries among schoolchildren is about 60-90% and approximately 100% among adults throughout the world (World Health Organization, 2012). The role of environmental and behavioral risk factors in causing dental caries is very well established, with factors such as lack of fluoride, poor oral hygiene habits, and unhealthy diets contributing to dental caries (Ahluwalia et al. 2004; Shaffer et al. 2012).

The risk for developing poor oral health, including dental caries, can also increase in children born with craniofacial defects (Mobley et al., 2009; Cheng et al., 2007), due to concomitant oral abnormalities, affecting the dental structure, shape and numbers of teeth. As a consequence, the child could develop teeth malposition and crowding, and that could eventually make it very hard to maintain good oral health.

2.2.1 DENTAL CARIES IN OROFACIAL CLEFTS (OFC)

Several studies have evaluated oral health in children with orofacial clefts (OFCs) versus controls, see Table 3 for detailed summaries of these studies. Some indicated that children with OFCs have significantly increased rates of caries experience when compared with controls (Mutarai et al., 2008; Al-Dajani, 2009; King et al., 2013), but others found no difference (Lucas et al., 2000; Kirchberg et al., 2012; Freitas et al., 2013; Howe et al., 2017). Some studies indicated that children with OFCs have significantly increased rates of oral health indices such as dental plaque, bacterial loads and gingival indices when compared with controls (Parapanisiou et al., 2009; Hazza'a et al., 2011; Chopra et al., 2014; Sundell, Ullbro, et al., 2015). Note that most of these studies were small, with most sample sizes less than 100 case/control pairs, and a few between 100 and 400. The largest of these studies (n=3326; with 639 OFC-affected individuals, 1549 unaffected OFC-family members, and 1138 controls, Howe et al., 2017) was done in the current study population, and indicated that individuals with non-syndromic OFCs and their family members do not have a higher dental caries risk when compared to the general population.

Table 3 Summary of studies comparing dental caries, dental plaque, and bacterial loads between children with OFCs and controls

Study	Study Design	Ethnicity	Number of cases/controls	Dental Phenotype measured *	Caries Differs between cases and controls
Al-Dajani, 2008	Case/Control	Syria	53 subjects with CL/P (12-29 years). Controls: 53 subjects without CL/P (12-29 years).	DMFT	Yes
Al-Wahadni et al., 2005	Case/Control	Middle Eastern (Jordan)	Cases: 32 with cleft and palate (ages: 10-28yr.). Controls: 32 non-cleft (ages: 10-28yr.).	DMFT	Yes
Britton & Welbury, 2010	Case/Control	Scotland	Cases: 209 (6months-6years old) with any type of Orofacial clefts. Controls: National data (6months-6years old) without clefts.	dmft	-Younger age groups did not differ -Older age groups did differ
Chopra et al., 2014	Case/Control	Asian (Indian)	Cases: 52 children with CLP (4-6-year-old). Controls: 48 unaffected subjects	-dmft -Plaque and gingival indices	Yes
Freitas et al., 2013	Case/Control	Brazil	Cases: 30 adolescent and young adults with CL/P (12 to 21 years). Controls: 30 unaffected adolescent and young adult without CL/P (12 to 21 years).	DMFT	No
Hazza'a et al., 2011	Case/control	Middle Eastern (Jordan)	Cases: 98 children with CLP (age: 4–23yr.). Controls: 98 unaffected subjects.	-dmft/DMFT -Plaque and gingival indices	Yes
Howe et al., 2017	Case/Control	Multiple sites	639 case probands, 1,549 unaffected relatives, and 1,138 controls	- Primary dft, dt - Permanent DFT, DT	No

King et al., 2013	Case/Control	Chinese	Cases: 132 children (2-7years) with CLP. Controls: 132 controls (2-7years) without any cleft.	DMFT	-No in 2-4 years -Yes in 5-7 years
Kirchberg et al., 2012	Case/Control	German	Cases: 295 children (1 to 6 years) with clefts. Controls: 548 (1- to 6-year) without clefts.	dmft	No
Lucas et al., 2000	Case/Control	UK	Cases: 60 children with CLP (3-15 years). Controls: 60 children without CLP (3-15 years).	dmfs	No
Mutarai et al., 2008	Case/Control	Thailand	Cases: 69 children with CL/P (1.5-3 years). Controls: 69 children without CL/P (1.5-3 years).	dmft	Yes
Parapanisiou et al., 2009	Case/Control	Greek	Cases: 41 (4-18 years-old) with CLP. Controls: 41 (4-18yr.) without CLP.	- Plaque index - Prevalence of initial/white spot and carious lesions.	Yes
Sundell, Ullbro, et al., 2015	Case/Control	European (Sweden)	Cases: 133 children (77 subjects aged 5 years and 56 aged 10 years). Controls: 297 non-clefts (133 aged 5 years and 164 aged 10 years).	Plaque Index, and saliva samples were analyzed for lactobacilli counts.	Yes

*DMFT/dmft: Total number of decayed, missing and filled teeth due to caries
*DFT/dft: Total number of decayed and filled teeth due to caries

2.2.2   GENETICS IN DENTAL CARIES

The successful completion of the Human Genome Project (HGP) in 2000 made a variety of research tools possible (Sham, et al., 2014). These research tools include databases such as the National Center for Biotechnology Information databases and the UCSC Genome Browser. Moreover, statistical approaches such as genome-wide association studies (GWAS), which generally utilize single-nucleotide polymorphisms (SNPs) along with imputation to infer genotypes at non-genotyped base pairs in a study population became possible after the completion of the HGP.

These new genomics tools and approaches are ideal for dissecting common, complex (non-single-gene) traits such as dental caries and structural dental anomalies where there is an interplay between genetics and environmental factors in causing these traits.

Applying genome-wide approaches to dental caries studies helped in identifying multiple regions associated at the genome-wide level. In the last decade, there have been several genome-wide association studies that helped to reveal potentially causal risk genes for dental caries. However, most of these previous GWASs have been done in a single ethnic group, specifically within populations of European descent. Results of some of the recent genome-wide association studies (GWASs) of dental caries are summarized in Table 4 (Shaffer, et al., 2011; Wang et al., 2012; Shaffer et al. 2013; Zeng, et al., 2013; Zeng, et al., 2014; Haworth et al., 2018; Shungin et al., 2019).

Previous genome-wide association studies (GWASs) of dental caries have sought to identify the genetic factors involved with dental caries in a single ethnic group and/or in a single dentition (Shaffer et al., 2013a; Shaffer, et al., 2013b; Shaffer et al., 2011). However, the GWAS in the current study leverages a large international OFC consortium with dental data for children and

adults with non-syndromic clefting, plus their unaffected siblings and parents, and controls with no personal nor family history of OFC.

Table 4 Genome-Wide Studies of Dental caries

Study	Sample Size	Dental Caries Phenotype	Genes	Genome-wide Significant Associations
Shaffer et al., 2011	1305 white children, age: 3-12	Binary affection status in primary dentition	*ACTN2, EDARADD, MPPED2, MTR,* and *LPO*	No
Wang et al., 2012	7443 whites Adult, age: 17-89	Permanent decayed, missing, filled surfaces index (DMFS)	*RPS6KA2, ISL1, TLR2 RHOU, FZD1, PTK2B,* and *ADMTS3,*	No
Shaffer et al., 2013	920 whites Adult, ages :18-75	Permanent cluster-based partial DMFS	*LYZL2, AJAP1*	Yes
			ABCG2, PKD2, the dentin/bone *SCPP* sub-family, *EDNRA, TJFBR1, NKX2-3, IFT88, TWSG1, IL17D,* and *SMAD7*	No
Zeng et al., 2013	1,017 whites Adult, age:14-56	Permanent decayed and filled teeth (dft) stratified to generate df-pitt and fissures (dfPF)and df-smooth surface (dfSM)	*BCOR, BCORL1, INHBA, CXCR1* and *CXCR2*	No
Zeng et al., 2014	1,006 white children, age :3-14	Primary decayed and filled teeth (dft) stratified to generate df-pitt and fissures (dfPF)and df-smooth surface (dfSM)	*KPNA4, ITGAL, PLUNC* family genes,	Yes
			MPPED2, AJAP, and PRS6KA2	No
Haworth et al., 2018	19,003 Primary analysis; 13,353 Permanent analysis European ancestry meta-analysis age: 2.5-18	Presence or absence of treated or untreated caries	*ALLC, NEDD9* for primary and permanent dentition respectively	Yes
Shungin et al., 2019	GLIDE and UKB (n = 26,792)	DMFS, DFSS, Nteeth	*C5orf66, CA12*	Yes
			KRTCAP2, WNT10A, FGF10, HLA, FOXL1, PBX3, MAMSTR	No

3.0 ANALYZE THE PREVALENCE AND ASSOCIATIONS AMONG STRUCTURAL DENTAL ANOMALIES, AND BETWEEN DENTAL CARIES AND ENAMEL HYPOPLASIA

3.1 METHODS AND MATERIAL

3.1.1 SAMPLE DESCRIPTION

POFC

The first cohort for this study (N= 3,579) comes from the Pittsburgh Orofacial Clefts Study (POFC). The POFC study populations were recruited from multiple cleft centers in the United States, including Colorado, Iowa, Pennsylvania, Texas, and Puerto Rico, and internationally, from Argentina, the Philippines, Colombia, Guatemala, and Hungary for the purpose of studying orofacial clefts. Institutional review board (IRB) approval was obtained at each site by the appropriate IRB process and committee, with a coordinating IRB at the University of Pittsburgh (IRB 0405013). In addition, Informed consent was obtained from all participants. The same data collection protocols were used for every site.

The total sample (N=3,579) included 1392 control individuals, including control probands, their parents, and siblings and 2187 unaffected relatives, including case parents, siblings, and spouses. The exclusion criteria were subjects younger than 7 years; any case affected with orofacial clefts, and edentulous subjects. Table 5 summarizes the basic descriptive statistics for the subjects involved in our study.

COHRA1

Our second cohort (N= 1763) comes from the Center for Oral Health Research in Appalachia (COHRA), which was designed to address oral health disparities in the Appalachian region in the US. Appalachia has been known to have the largest burden of oral health problems per capita in the United States. The participants for COHRA were recruited from rural Appalachia in West Virginia and Western Pennsylvania by using a household-based recruitment protocol, which required a minimum of one biological parent-child pair per household. The recruitment unit for COHRA was families, which were recruited from five sites across Pennsylvania and West Virginia in northern Appalachia. Institutional review board (IRB) approval was obtained at each site by the appropriate IRB process and committee, with a coordinating IRB at the University of Pittsburgh (IRB 0207073) In addition, Informed consent was obtained from all participants. All edentulous subjects and subjects younger than 7 years were excluded, to make this cohort comparable with the POFC cohort, since the majority of our subjects come through the POFC study. Table 6 summarize the basic descriptive statistics for the COHRA1 subjects.

3.1.2 DATA COLLECTION

POFC

Data regarding the subject's dental history, including dental extractions and orthodontic treatment, were collected from all the subjects by self-report. In addition, each participant had either an in-person dental exams or an intraoral photos taken. In-person dental exams were performed by trained and calibrated dentists or dental hygienists who inspected the oral cavity using dental mirrors and explorers. Cameras (Canon EF 100-mm f/2.8 macro USM lens, Canon

macro MR-14EX ring flash; Canon, Tokyo, Japan) and supplies for intraoral photo collection were provided to the recruitment sites. At least 5 photographs were taken per subject to appropriately display the entire oral cavity.

Table 5 POFC Cohort (1505 male, 2074 female)

	SITE	Subject Type				Total (%)
		Unaffected relatives Sex		Controls Sex		
		Male	Female	Male	Female	
USA	USA-COLORADO	15	20			35 (0.98)
	USA-IOWA	90	162	87	146	485 (13.55)
	USA-PITTSBURGH	31	49	45	65	190 (5.31)
	USA-TEXAS	119	161	0	1	281 (7.85)
	USA-PUERTO RICO	11	21	3	11	46 (1.28)
	Total	266	413	135	223	1037 (28.97)
INTERNATIONAL SITES	COLOMBIA	129	136	90	89	444 (12.4)
	GUATEMALA	55	94	90	185	424 (11.85)
	HUNGARY	142	163	205	231	741 (20.70)
	ARGENTINA	82	185	16	30	313 (8.75)
	PHILIPPINES	246	276	49	49	620 (17.32)
	Total	654	854	450	584	2542 (71.03)
ALL SITES		920	1267	585	807	
	TOTAL	2187 (61.1%)		1392(38.9%)		3579
	Age (mean± SD)	8-82yr 30.18± 14				

COHRA1

Self-reported questionnaires and interview data were collected from the subjects to cover oral health, medical health, family history, family pedigree, family relationship, and pregnancy history. Parents were asked to complete the questionnaires for children 10 years or younger. Children, age 11 and older completed the questionnaires themselves.

The clinical examination of the subjects was performed in an exam room equipped with a dental chair and dental examination light. Trained and calibrated dental examiners (either a dentist or dental hygienist plus an assistant) performed standardized periodontal and caries screenings. The assessment included documentation of oral health problems such as caries, restorations, plaque, and calculus.

Table 6 COHRA1 Cohort (752 male, 1011 female)

SITE	SEX		Total (%)
	Male	**Female**	
PENNSYLVANIA—BK	72	106	178 (10)
PENNSYLVANIA—BR	94	108	202 (11.5)
PENNSYLVANIA—BU	118	142	260 (14.8)
WV	468	655	1123(63.7)
Age	7-62yr	7-67yr	7-67yr
(mean + SD)	(23.63 ± 14)	(25.89 ± 12.76)	(24.93 ± 13.35)
Total	752	1011	1763

BK- Braddock, BR- Bradford, BU- Burgettstown, WV- West Virginia

3.1.3 STRUCTURAL DENTAL ANOMALIES

Structural dental anomalies in this study were obtained only in the POFC cohort. For the current study we investigated enamel hypoplasia, mamelons, rotated and displaced teeth, supernumerary teeth, and agenesis, all as "yes/no" binary traits. Table 1 provides definitions for the different dental anomalies.

Each dental anomaly was evaluated separately in each subject, then prevalence averaged over the total study and in unaffected relatives and controls separately. We tested to see if there is a significant difference in the prevalence values between unaffected relatives and controls (see section 3.3). Sex differences were tested for all the different dental anomalies. In addition, correlations between the different dental anomalies were tested (see section 3.3).

The calibration process of the intraoral dental exams and photos for dental anomalies was done in data from 15 participants, which were randomly chosen. Each photo of the participant teeth was rated 2 times by each rater (BJH, LMMU, and ARV). Results from ratings by LMMU and ARV were calibrated against the gold-standard rater, BJH. The Intra-rater reliability for BJH was 100% agreement, with kappa = 0.95, while the Inter-rater reliability between all 3 raters was 97.1% to 97.3% (kappa =0.91–0.93) (Howe, et al., 2015).

3.1.4 DENTAL CARIES

Dental caries phenotypes were assessed by trained dentists or dental hygienists plus an assistant in the COHRA1 cohort. In the POFC cohort, data were collected through in-person dental examinations and/or intraoral photos. As previously reported for POFC (Howe et al., 2017), at least 5 intraoral photographs were taken for participants (maxillary and mandibular occlusal, right and left lateral, anterior biting) to appropriately display the entire oral cavity. Each tooth surface was scored as either decayed, filled/restored, or missing. Each tooth was scored as Decayed if one or more surface had untreated decay, Filled/restored if at least one surface was filled/restored (but no surface with untreated decay), and Missing if at least one surface was missing (but no surface untreated decay nor filled/restored). In other words, for scoring a tooth, untreated Decay trumped Filled/restored, which trumped Missing.   The DFT indices per person were calculated as the total number of teeth with decayed, and/or filled/restored teeth, with both dentitions included. We used the total DFT+ dft index, in both POFC and COHRA1, instead of DMFT/dmft due to incomplete information regarding missing teeth for the majority of the participants. Dental caries was treated as a binary trait when we wanted to evaluate the relationship/correlation between it and different structural dental anomalies.

In POFC, calibration of the intra-oral photo caries rating was done in data from sixteen randomly chosen participants. Each photo was rated twice by each of three raters (BJH, NP, and TJC). Results from ratings by NP and TJC were calibrated against the gold-standard rater, BJH (note BJH's intra-rater reliability (kappa) was 0.96 (CI 0.92-0.99)). The Inter-rater reliability (kappa) between BJH and NP was 0.95 (0.91, 0.98), and between BJH and TJC was 0.96 (0.93, 0.99).

Thus, the reliability of the caries ratings from intra-oral photos was excellent (Howe, et al., 2017).

In COHRA1, calibration sessions for dental caries assessments occurred prior to the initiation of the study and periodically during the years of data collection. The Inter-rater reliability was determined by comparing assessments performed by each examiner/recorder team against the assessment from the gold standard examiner/recorder team on the same subject on two consecutive days. At each calibration session, screenings were performed on two children with caries, two adolescents with caries, and two adults with attachment loss that included periodontal pockets of at least 5 mm from the base of the pocket to the free gingival margin. Reliabilities were determined using Cohen's kappa, and the mean inter- and intra-rater reliabilities were 0.83 and 0.98 respectively, which indicate that the reliability of dental caries examination was very good (Polk, et al., 2008).

3.2 STATISTICAL METHODS

All descriptive and comparative statistical analyses were performed using the R statistical analysis environment (R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/). DFT mean scores were evaluated among different variables and comparisons using the t-test and chi-square (χ^2) test. The prevalence was assessed among different dental anomalies, and comparisons were performed by using χ^2 tests. P value $<1.3 \times 10^{-3}$ was selected as the threshold for significance after Bonferroni correction for 36 independent tests. The Pearson correlation coefficient was used to evaluate the relationship/correlation between the different structural dental anomalies; and between dental anomalies and dental caries, and P-values < 0.05 were considered as significant.

3.3 RESULTS

The POFC cohort comprised 1505 males (42.05%), and 2074 females (57.95) with an age range of 7–82 years and a mean age of 31 years (Table 5), Table A 1 and Table A 2 in the appendix show the age distribution across the different sites in POFC and the percentage of subjects according the different age groups, respectively. 991 subjects (27.69%) came from U.S recruitment sites, while 2,588 subjects (72.31%) came from international sites. Among the 3579 subjects, 3570 (99.75%) had data regarding dental caries phenotype (DFT score) and 3579 (100%) had dental anomaly data. In POFC, regression analysis showed that there were age, sites subjects' type; unaffected relatives vs. control; and sex effects on the DFT (dft) mean scores. General linear regression modeling of unaffected relatives-control status, allowing for adjustment of age, sex, and site, were completed by considering the DFT mean scores. Regression results for POFC are provided in the appendix Table A 3 . Figure 1 shows DFT scores distribution before and after the adjustment for covariates in POFC. Figure A1 and Figure A2 in the appendix shows the prediction of DFT across sites, age groups, subject types and sex.
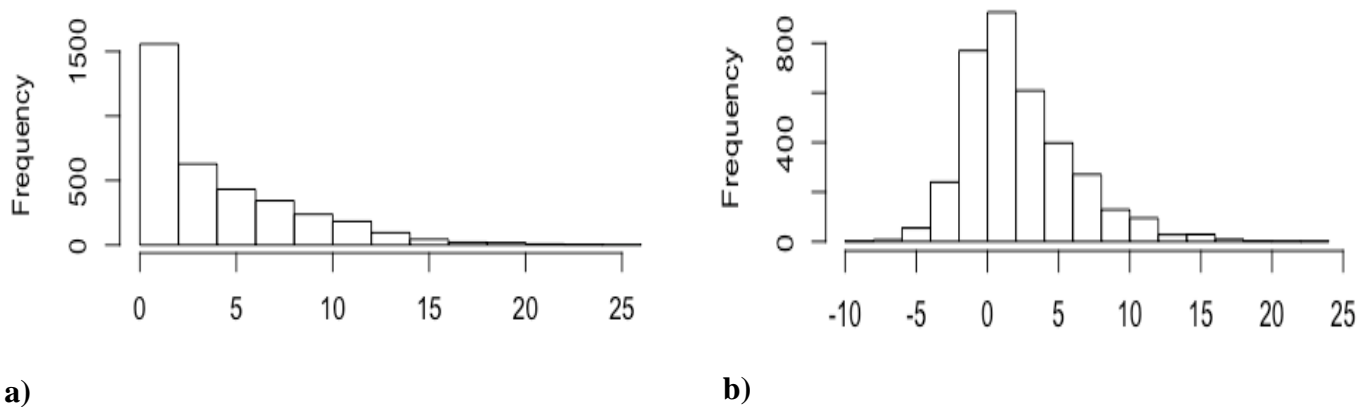

a)   b)

Figure 1 DFT scores Distribution a) Before and b) After Adjustment for Covariates in POFC

Table 7 DFT mean difference between unaffected relatives and controls) in POFC

SITE	Subject Type		Test of mean difference between unaffected relatives and controls*	
	Unaffected relatives	Controls	*P* value*a	95%CI
USA-COLORADO	4.12	--------------	-----------------	--------------------
USA-IOWA	0.17	0.61	0.09195	-0.95_0.07
USA-PITTSBURGH	3.22	3.12	0.8881	-1.34_1.55
USA-TEXAS	1.27	5.12 (1 observation)	--------------*b	-----------------*b
COLOMBIA	1.68	1.10	0.1334	-0.18_1.33
GUATEMALA	1.39	1.34	0.8918	-0.67_0.77
HUNGARY	1.78	1.15	0.04505	0.01_ 1.25
ARGENTINA	0.922	2.22	0.01115	-2.28_-0.30
PHILIPPINES	1.22	2.06	0.04845	-1.68_-0.01
PUERTO RICO	3.02	1.98	0.4069	-149_3.58
TOTAL	0.79	0.39	0.002163	0.15_0.66

Unaffected relatives (N=2187); controls (N=1392)

* the test of mean difference was done after the adjustment for the covariates

*a *P* value are based on t-test of mean differences

*b no enough observation to test for a difference in mean between unaffected relatives and controls

The DFT mean scores (Table 7) were slightly higher among the unaffected relatives (P = 0.002163) compared with controls, however this difference was not considered significant after Bonferroni correction for multiple testing. In addition, the difference between unaffected relatives and controls regarding the DFT mean scores in Argentina (DFT = 0.92 vs. 2.22, P = 0.01115) was not considered significant.

Table 8 shows the comparisons of the DFT mean scores between males and females across the different sites. There was no significant difference in the DFT mean scores between male and female (DFT = 1.13 vs. 1.54, P = 0.002177) across the different sites. There was a difference in

the DFT mean scores between male and female in Colombia (DFT = 1.56 vs. 2.56, P = 0.00463), and in Philippines (DFT = 1.03 vs. 1.65, P = 0.04114), however, both of these differences are not considered significant after Bonferroni correction for multiple testing.

Table 8 DFT mean difference between male and female in POFC

SITE	Sex		Test of mean difference between male and female	
	Male	Female	P value*	95%CI
USA-COLORADO	2.89	1.53	0.3048	-1.30_4.03
USA-IOWA	2.25	2.03	0.4249	-0.32_ 0.75
USA-PITTSBURGH	1.63	2.43	0.2918	-2.30_ 0.70
USA-TEXAS	1.14	1.39	0.5865	-1.16_0.66
COLOMBIA	1.56	2.65	0.00463	-1.83_-0.34
GUATEMALA	2.13	2.11	0.9504	-0.67_0.71
HUNGARY	1.83	2.36	0.07894	-1.12_ 0.06
ARGENTINA	2.02	2.16	0.7606	-1.0_0.75
PHILIPPINES	1.03	1.65	0.04114	-1.23_-0.02
PUERTO RICO	0.59	2.82	0.08914	-4.81_0.36
TOTAL	1.13	1.54	0.002177	-0.68_-0.15

*P value are based on t-test of mean differences

The total sample size from the COHRA1 cohort was 1763, 752 male (42.65%) and 1011 female (57.35%) (Table 6), the age range for the subjects was between 7-67 year with a mean age of 24.93, and the majority of subjects coming from West Virginia (63.70%). Table A 4 in the appendix shows the age distribution across the different sites in COHRA1. Regression analysis

showed that there were age and site effects on the DFT (dft) mean scores in COHRA1 cohort, and we adjust for these covariates in our analysis. Regression results for COHRA1 are provided in the appendix Table A 5.

Figure 2 shows the DFT scores distribution before and after the adjustment for covariates in COHRA1. The differences in the DFT mean scores between male and female (Table 9) across the different sites in COHRA1 (DFT = -0.62 vs. -0.6, P = 0.9011) was not significant.
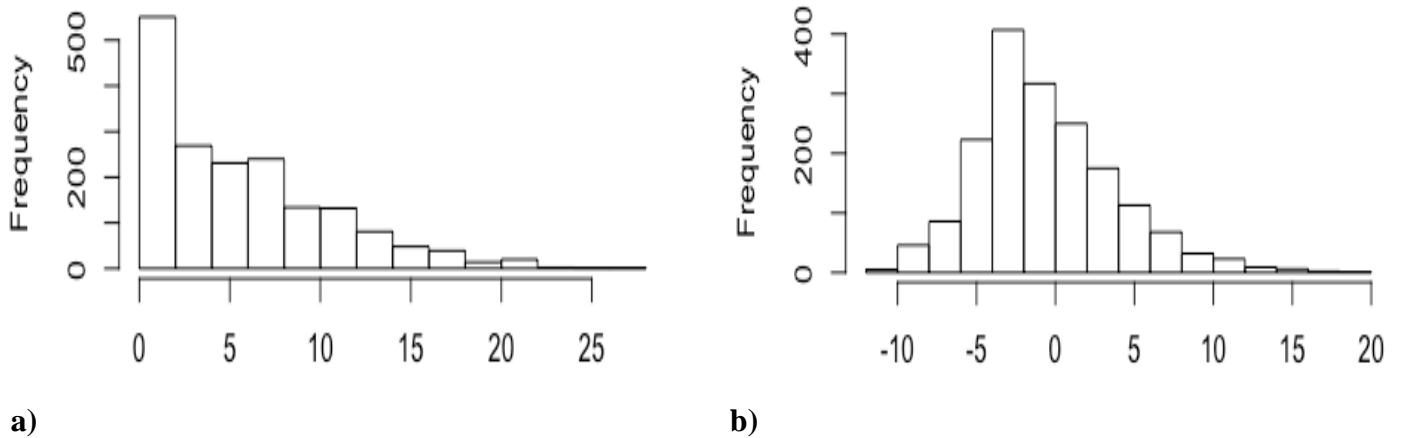

a)   b)

Figure 2 DFT scores Distribution a) Before and b) After Adjustment for Covariates in COHRA1

Table 9 DFT mean difference between male and female in COHRA1

SITE	Sex		Test of mean difference between male and female	
	Male	Female	*P* value*	95%CI
PENNSYLVANIA—BK	-2.02	-1.87	0.7564	-1.16_0.84
PENNSYLVANIA—BR	-0.34	0.23	0.3514	-1.79_0.64
PENNSYLVANIA—BU	-0.38	-0.36	0.9673	-1.20_ 1.15
WV	-0.52	-0.58	0.8386	-0.48_0.59
Total	-0.62	-0.60	0.9011	-0.44_ 0.39

BK- Braddock, BR- Bradford, BU- Burgettstown, WV- West Virginia

P value are based on t-test of mean differences

Age distribution was not significantly different between male and female in COHRA1 ($P =$ 0.01425), however, there was a significant age difference between male and female in BK site ($P = 3.74 \times 10^{-5}$) (appendix Table A 4).

Dental anomalies result in POFC indicate that Rotation (70.80%) and Displacement (40.77%) were the most prevalent dental anomalies. Significantly, females had more rotation ($P = 6.478^{-09}$) and displacement ($P = 1.003^{-06}$) when compared with males in the POFC cohort (Table 10) Unaffected relatives had more rotation ($P = 0.0016$) when compared to controls (Table 11), however, this difference was not considered significant.

Table 10 Distribution of dental anomalies by sex in POFC

Dental Anomaly	n (%)	Sex (%)		P-value
		M (%)	F (%)	
Agenesis	170 (4.7)	80 (2.2)	90 (2.5)	0.175
Hypoplasia	282 (7.8)	136 (3.8)	146 (4)	0.029
Impaction	41 (1.14)	15 (0.42)	26 (0.73)	0.476
Rotation	2534 (70.80)	1144 (32)	1390 (38.8)	6.478e-09*
Displacement	1459 (40.77)	685 (19.14)	774 (21.63)	1.003e-06*
Mamelons	354 (9.89)	155 (4.33)	199 (5.56)	0.486
Supernumerary	26 (0.73)	16 (0.45)	10 (0.28)	0.043
Dental Caries	2806 (78.40)	1,169 (41.66)	1,637 (58.34)	0.368

P values are based on χ2 tests

Table 11 Distribution of dental anomalies by subject type in POFC

Dental Anomaly	n (%)	Subject Type (%)		P-value
		Unaffected relatives	Controls	
Agenesis	170 (4.7)	116 (68.24)	54 (31.76)	0.051
Hypoplasia	282 (7.8)	177 (62.77)	105 (32.23)	0.551
Impaction	41 (1.14)	33 (80.49)	8 (19.51)	0.010
Rotation	2534 (70.80)	1,599 (63.10)	935 (36.90)	0.0016
Displacement	1459 (40.77)	893 (61.21)	566 (38.79)	0.919
Mamelons	354 (9.89)	233 (65.82)	121 (34.18)	0.055
Supernumerary	26(0.73)	22 (84.62)	4 (15.38)	0.014
Dental Caries	2,806 (78.40)	1,756 (62.58%)	1,050 (37.42)	0.0018

P values are based on χ2 tests

Based on the Pearson correlation coefficient, that was used to evaluate the relationships between the different dental anomalies, we found a correlation between agenesis, impaction and rotation (AIR); hypoplasia, displacement and rotation (HDR); and between displacement, rotation and mamelons (DRM). In addition, we found out that dental caries, as a binary trait, was also correlated with agenesis and hypoplasia (DAH). Although the correlation coefficient reflects a weak association between those anomalies, we used a multivariate GWAS method, which is recommended when genetic correlations between traits are weak. Correlation results are summarized in Table 12. We found that 84.4% of participants who had enamel hypoplasia had developed dental caries (Table 13). In addition, Table 14 shows the correlation results between dental caries (as a binary and quantitative trait) and enamel hypoplasia and indicate that enamel hypoplasia was correlated with dental caries.

Table 12 Correlations between the different dental anomalies and dental caries in POFC

	Dental Caries	Agenesis	Hypoplasia	Impaction	Displaced	Rotation	Mamelons	Supernumerary
Dental Caries		0.059***	0.043*	-0.007	0.019	0.029	-0.137	0.005
Agenesis			-0.002	0.038*	0.007	0.045**	-0.026	-0.004
Hypoplasia				0.017	0.057***	0.049**	-0.007	0.036
Impaction					0.028	0.029	-0.000	-0.009
Displaced						0.459***	0.112***	0.009
Rotation							0.192***	-0.003
Mamelons								-0.006
Supernumerary								

*P value <0.05, ** P value <0.01, ***P value <0.001

Table 13 Chi-square results for dental caries and enamel hypoplasia in POFC

Enamel hypoplasia	Subjects with caries	Subjects without caries	Total	
With hypoplasia	238	44	282	
	84.4	15.6	100	
Without hypoplasia	2568	720	3288	
	78.1	21.9	100	
Total	2806	764	3570	
	78.6	21.4	100	

chi^{2}X = 5.7501, *P*= 0.016

Table 14 Correlation results between dental caries and enamel hypoplasia in POFC

Tooth Outcome	Predicator variable	Cor. (95% CI)	*P*-value
DFT (Quantitive)	Enamel Hypoplasia	0.037 0.0049_0.0684	0.0284*
DFT (Binary)	Enamel Hypoplasia	0.043	0.0108*

*P value <0.05

3.4 DISCUSSION

Our study provided a large sample size to study dental caries phenotype in two cohorts, POFC and COHRA1, and study different structural dental anomalies in POFC. We also had the advantage of studying dental caries phenotype among Oral facial clefts relatives in POFC. Previous studies indicated that relatives of subjects with clefts do not have an increased risk of dental caries when compared to the general population (de Castilho et al. 2006; Al-Dajani 2009). Moreover, a previous study (Howe, et al., 2017) that was done in this study cohort (POFC) found

that unaffected relatives of the individuals with non-syndromic OFCs do not have a higher dental caries risk when compared to controls.

After the adjustment for covariates in POFC, the DFT scores were slightly higher among the unaffected relatives (DFT = 0.79 vs. 0.39, P = 0.002163) compared with controls, however, this difference was not significant after Bonferroni correction for multiple testing. In addition, the difference in the DFT percentage between unaffected relatives and controls (62.58% vs. 37.42%), was also not considered significant after Bonferroni correction for multiple testing (P = 0.0018). Our results are consistent with the previous study (Howe, et al., 2017) that was done in this study cohort (POFC) and found that unaffected relatives of the individuals with non-syndromic OFCs do not have a higher dental caries risk when compared to controls.

There was a trend for an increases risk to develop dental caries among females when compared to males (DFT = 1.13 vs. 1.54, P = 0.002177) across the different POFC sites. However, there was no difference in DFT scores between males and females in COHRA1 (DFT = 0.62 vs. 0.60, P = 0.9011). Higher dental caries prevalence among females when compared to males has been reported before. The possible explanations behind the increase risk for developing more dental caries among female are the fact that teeth erupt earlier in females, which allows for a longer exposure to a cariogenic oral environment; pregnancy and hormonal changes, female tend to eat snacks more; and different dentition characteristics, such as tooth enamel in females (Shaffer, et al., 2015; Martinez-Mier & Zandona, 2013; Lukacs & Largaespada, 2006). In addition, previous studies concluded that sex differences in dental caries can be explained by the differential effects of genes that influence dental caries (Ferraro & Vieira, 2010; Vieira, Marazita & Goldstein-McHenry, 2008).

Sex differences for all dental anomalies, including dental caries as a binary trait, were tested and we noticed that females had more rotation ($P = 6.478^{-09}$), and displacement ($P = 1.003^{-06}$) when compared to males. The tendency to develop more rotation and displacement in females could be explained by the observation that females had different characteristics of the dentition, i.e. differences in the dentin, enamel and mesiodistal teeth width, when compared to males (Martinez-Mier & Zandona, 2013; Lukacs & Largaespada, 2006). Although a previous study by one of our collaborators investigated the sex differences in dental anomalies among this cohort (POFC), it is important to point out that their comparisons excluded tooth rotation and displacement (Howe, et al., 2015), due to the large incidence of both tooth rotation and displacement in the study cohort.

Subject type, i.e. unaffected relatives vs. controls, was tested for all dental anomalies' variables, and we noticed that unaffected relatives had more rotation in their teeth ($P = 0.0016$) when compared to controls, however, this difference was not significant.

We also observed a trend for an increased rate in unaffected relatives for impaction (80.49% vs. 19.51%, $P = 0.01$), supernumerary teeth (84.62% vs. 15.38%, $P = 0.014$), and for dental caries (78.40% vs. 62.58%, $P = 0.0018$) compared with controls. However, we do not consider these differences significant in our study after correcting for multiple testing, which is also consistent with the findings from a previous study in this same cohort (Howe, et al., 2015). Since we did not find significant differences between unaffected relatives and controls regarding the different dental anomalies, we decided to combine these subsets of the cohort for our second aim in this study.

We observed a correlation between different structural dental anomalies, plus dental caries, although the correlation coefficient reflects a weak correlation between those anomalies. This is

because these different dental anomalies might share a very similar genetic background and that emphasizes the importance of investigating these anomalies on a genetic level. Interestingly, we also noticed that there is a significant relationship between dental caries and enamel hypoplasia, where subjects with enamel hypoplasia are more likely than subjects without hypoplasia to develop dental caries, chi^{2}X (1, N = 3570) = 5.7501, P = 0.016 (Table 13). This result is consistent with the previous studies that reported an association between enamel hypoplasia and dental caries (Milgrom et al., 2000; Montero et al., 2003; Daneshkazemi and Davari, 2005, Hong, et al., 2009).

3.5 CONCLUSION

Our results indicated that relatives of oral facial clefts do not have higher risk to develop dental caries when compared to the general population, and other environmental/behavioral factors could play a role in in increasing the risk of developing dental caries. In addition, dental caries become more prominent with age.

Correlation between different structural dental anomalies; including the correlation between dental caries and enamel hypoplasia, and between dental caries tooth agenesis; reflect the possibility of similar genetic background and indicate the importance of further investigation on a genetic level. Future studies are needed to investigate these correlations in bigger sample size and in different ethnicities.

4.0 CONDUCT A GENOME-WIDE ASSOCIATION ANALYSES IN THE POFC STUDY COHORT TO IDENTIFY COMMON VARIANTS ASSOCIATED WITH TOOTH AGENESIS, ENAMEL HYPOPLASIA AND WITH OTHER DENTAL ANOMALIES

4.1.1 SAMPLE DESCRIPTION

POFC

The sample description was previously described in section 3.1.1

4.1.2 DATA COLLECTION

DNA Collection, Genotyping and quality control

For both COHRA and POFC, DNA was collected from blood, mouthwash, buccal swab, or saliva samples. The genotyping for COHRA and POFC was performed by the Johns Hopkins University Center for Inherited Disease Research (CIDR). The chip that was used for the genotyping in the COHRA sample was the Illumina Human610-Quadv1_B BeadChip (Illumina, San Diego, CA, USA) (Wang, et al., 2012), while the chip that was used for the genotyping of the POFC study was the Illumina HumanCoreExomePlusCustom_Marazita_15050181 array (BPM annotation version A, genome build 37) (Marazita, 2015). Genotype data for COHRA and POFC went through an extensive process of cleaning, imputation, and quality assurance, performed by the CIDR Genetics Coordinating Center (GCC) at the University of Washington.

Standard quality-control criteria were applied to filter SNPs. SNPs were excluded if they had (i) a missing call rate of ≥5% in cases or controls; (ii) >1 discordant calls; (iii) SNPs with MAF < 1% in the population; (iv) significant deviation from Hardy–Weinberg equilibrium (*P*-values less than 10^{-4}); (v) >25 Mendelian errors; (vi) Imputed genotypes were filtered out if their probability was < 0.9; (vii) SNPs with INFO score < 0.5; and (viii) more than 1 HapMap replicate error. The total genotypes that were released by CIDR was 589,735 SNPs for COHRA, as part of the GENEVA consortium, and 539,473 SNPs for POFC (Marazita, 2015; Cornelis, et al., 2010; Bennett, et al., 2011).

4.1.3 DENTAL ANOMALIES

Structural Dental anomalies in this study were obtained only in the POFC cohort, and were restricted to hypoplasia; mamelons; rotated, and displaced teeth; supernumerary teeth; and agenesis. The binary traits for these dental anomalies were defined as "no" for not having the dental anomaly, and "yes" for having the specific dental anomaly (0/1). The third molars (wisdom teeth) were not included in our analysis.

Pearson correlation coefficient was used to evaluate the associations between the different dental anomalies (previously descried in section 3.4), and we found a correlation between agenesis, impaction and Rotation (AIR); hypoplasia, displacement and rotation (HDR); between displacement, rotation and mamelons (DRM); and between dental caries, agenesis, and hypoplasia (DAH). Although the correlation coefficient reflects a weak association between those anomalies, we used multivariate GWAS method, which is recommended when genetic correlations between traits are weak.

4.1.4 GWAS

GWAS is a hypothesis-free approach that has been used and become available after the completion of Human Genome Project (HGP). It is a very useful method for studying the contribution of common variants to complex traits and diseases. The design of most GWASs is based on the common disease-common variant hypothesis, which assumes that genetic susceptibility to a common disease is highly associated with common variants, and in most cases the common variants are single nucleotide polymorphisms (SNPs). Note that these common SNPs are unlikely to be the etiologic variants but instead are more likely to be in LD with the etiologic variant(s). There is also an alternative design of GWAS, which is based on the common disease/rare variant hypothesis, which state that complex traits are caused collectively by several rare variants with moderate to high penetrance. These rare variants might be etiologic but might also be in LD with the etiologic variant(s).

Under the GWAS approach, quantitative traits are usually analyzed using linear regression while binary (categorical) traits are analyzed using logistic regression. The methods of regressions allow us to control for covariates (such as age, sex), and can provide us with adjusted odds ratios as a measure of effect size for logistic regression and regression coefficient (beta) as a measure of effect size in linear regression (Bush, et al., 2012).

For the purpose of this study, structural dental anomalies were treated as a binary trait and the associations between SNP genotypes and dental anomalies were assessed using logistic regression with adjustment for age, age^2, sex, site, and Principal components (PCs) of genetic ancestry. Under the GWAS approach, millions of SNPs were tested and to avoid false positive results, multiple test correction was required and the conservative genome-wide significance

threshold of $P < 5 \times 10^{-8}$ (Bonferroni correction for a million tests) has become commonly used in GWASs.

The inclusion criteria for the genome- wide association analysis was subjects having assessment for each of the dental anomalies in POFC and for whom genetic data was available.

Since we wanted to analyze multiple traits simultaneously while accounting for correlations between those traits, we used multivariate GWAS. One major advantage of the multi-trait or multivariate GWAS is that phenotypic correlations can increase statistical power to detect association signals relative to univariate methods (Korte et al. 2012; O'Reilly et al., 2012; Zhou and Stephens 2014).

There are number of methods for simultaneous analysis of multiple correlated traits in population-based GWAS, some examples of these methods are: Multivariate test of association; Bayesian multiple phenotype test; Bayesian model comparison for multivariate regression, and the Trait-based Association Test, which uses Extended Simes procedure (TATES). We choose to use the multivariate test of association method (MQFAM) implemented in PLINK (Purcell, 2007).

MV-PLINK uses the canonical correlation analysis (CCA), which is a multivariate generalization of the Pearson product-moment correlation to measure the association between sets of variables. CCA extracts the linear combination of traits that explain the largest possible amount of the covariation between the genetic variants and all traits included in the analysis. The value of the F in the output of CCA is for testing the significance of the Wilks' lambda, which tests the significance of the canonical correlations, while the weights column reflects the correlation coefficients for each individual trait and the score correspond to the linear combination of traits

with maximum correlation with the genetic variant. The stronger the weight, the more an individual trait contributes to the association result (Ferreira & Purcell, 2009).

The CCA method is equivalent to multivariate analysis of variance (MANOVA) and it is most appropriate for the analysis of normally distributed traits. However, it also showed good performance when considering non-normal distributed traits (Ferreira & Purcell, 2009). In addition, MV- PLINK performed best for different scenarios, and outperformed other methods, such as TATES (Galesloot, et al., 2014). The only issue we had while using mv-PLINK was the long time it took to run the analysis. This is was because we had to run the analyses with permutation testing to correct for family structure.

In addition to running the Multivariate approach implemented in PLINK (mv-PLINK), we also ran a separate genome wide association analysis for each of the dental anomalies (univariate analysis). Since the POFC sample comes from a larger family studies with different ethnicities, we wanted to use exact association test statistics that could account for relatedness and population structure, i.e. the variance component approach implemented in Efficient Mixed-Model Association eXpedited (EMMAX) software (Kang, et al., 2010). EMMAX corrects for a broad range of sample structures by explicitly accounting for pair-wise relatedness between individuals by using population parameters previously determined to prevent repeatedly estimating variance components when performing each test by using the pre-estimated variance components from the null model. EMMAX keeps the heritability estimated from the null model fixed when testing individual SNPs (Wu, et al., 2011). Furthermore, it accounts for the structure on the scale of ethnic groups, samples from same ancestry group, within populations. EMMAX uses high- density genotypes to create a kinship matrix between every pair of individuals in the

study. Then, the kinship matrix is integrated into a linear mixed model to adjust for correlation in the phenotypic distribution during association mapping.

Previous studies showed that EMMAX has the highest power compared to principal components analysis (PCA) and genomic control methods that adjust for stratification. A previous study that evaluated the performance of different approaches to a genome-wide association in the context of consanguineous offspring found out that EMMAX had more power than within family-based association test only FBAT to rank the true SNP in the top genome-wide regions (Kang, et al., 2010). We opted to use EMMAX considering all the advantages of using EMMAX mentioned above as well as its fast-computational time for analyzing large GWAS datasets.

Since our dental anomalies are treated as binary traits, and because EMMAX is based on a linear mixed model rather than generalized mixed model, the effect size (beta) would not be meaningful. However, the *P*-values would be reliable.

4.1.5 STATISTICAL ANALYSES

R (http://www.r-project.org/) was used for the general statistical tests. The differences in the demographic variables, the genotype and allele frequencies of the SNPs between cases and controls was evaluated by using the χ^2 test (for categorical variables) or the t-test (for continuous variables). Hardy–Weinberg equilibrium of the genotype distributions of controls was estimated by a goodness-of-fit χ^2 test. Genomic inflation was assessed by calculating the genomic inflation factor lambda (λ), which is the ratio of the median of the observed distribution of the test statistic to the expected median, and visualized in a quantile-quantile plot in R. In addition, we used R to create Manhattan plots to visualize the association results. Variants with minor allele frequencies

lower than 5%, as well as variants with genotyping call rates less than 10% was filtered out. Regional plots were generated using LocusZoom to visualize association P-values for the regions of interest.

The threshold for genome-wide significance was set to P-value $\leq 5 \times 10^{-8}$, and for suggestive significance to P-value $< 10^{-5}$. The top associated loci were defined based on P-values of the suggestive and genome-wide significant associations and will be subsequently annotated.

4.1.6 GWAS RESULTS ANNOTATION AND INTERPRETATION

We investigated all the genes within ±500 kilobases (kb) of the top association signals (index SNPs) for putative connections to different dental anomalies and dental caries. To visualize regions showing genome-wide significance and regions showing suggestive significance, we used the regional plots generated by LocusZoom. For investigation of functionality, genes were investigated using the resources such as the Gene and PubMed databases at the National Center for Biotechnology Information (NCBI). To help prioritize potential causal genes in each associated region, genes were annotated for function in molecular, cellular, animal model and tissue/organ levels using multiple databases, including OMIM, UCSC, Genome Browser, EMAGE, Ensembl and ENCODE.

 To examine the genes near the association signals, we searched these different databases by using terms including the name of the gene or corresponding protein, plus terms relevant to dental development and dental anomalies. Moreover, these genes were examined within the Mouse Genome Informatics database to identify mouse phenotypes with presumed connections to different dental health and/or tooth development process.

4.2 RESULTS OF GWAS OF CORRELATED DENTAL ANOMALIES

Since the Pearson correlation coefficient that was used to evaluate the correlation between the different dental anomalies (previously descried in section 3.4), indicated a correlation between Agenesis, Impaction and Rotation (AIR); Hypoplasia, Displacement and Rotation (HDR); between Displacement, Rotation and Mamelons (DRM), and between Dental caries, Agenesis and Hypoplasia (DAH), we separated the results section to discuss the results for each of these four combinations of correlated dental anomalies separately.

4.2.1 AGENESIS, IMPACTION AND ROTATION (AIR)

The multivariate test of association was performed on cohort (N=3579) for whom genotypes data, phenotypes, and covariates data were available. A total of 5,802,671 SNPs was available for analysis after applying quality control criteria. Although no SNPs reached genome-wide significance ($P < 5 \times 10^{-8}$); there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Table 15. A visual representation of the association results of AIR is displayed in Figure 3. The genomic inflation factor (lambda) was =0.87, which indicate a possible deflation of the results.

Regions of the genome that showed suggestive significance with Agenesis, Impaction and Rotation (AIR), were further visualized using Regional Association (LocusZoom) plots in order to assess whether the region may have a possible role in tooth development and dental/oral health.

One of our suggestive variants was Chr3:64541255 ($P = 1.00E\text{-}06$), which is intronic to *ADAMTS9* (Figure 4). This gene is located in the extracellular matrix (ECM) and interacts tightly

with the ECM proteins, which are involved in neural crest formation and craniofacial morphogenesis. *ADAMTS9* may modify the extracellular environment, which allow proper proliferation and survival of neural crest-derived cells (Christian, Bahudhanapati & Wei, 2013). *Adamts9* expression was detected in developing craniofacial structures such as teeth and mandible in mice, and that might suggest a possible role of *ADMTS9* in tooth development (Jungers, et al., 2005) In addition, this signal is also approximately 300kb upstream of *PRICKLE2*, which is a planar cell polarity protein and has involvement in amelogenesis. A previous study that examined the expression of *Prickle2* in rats found that *Prickle2* is expressed in the differentiating inner enamel epithelial cells and early inner enamel-secretory ameloblasts (Nishikawa& Kawamoto, 2015). Interestingly, a previous study to identify genetic variants associated with missing teeth in the COHRA1 cohort found that a variant (rs7624909) in the same locus that we have here in our results of AIR pattern that could influence functional dentition (Everman, 2018). These observations indicate that *ADAMTS9* and *PRICKLE2* might have a role during odontogenesis and it might influence these correlated dental anomalies (AIR).

rs140220410 (P = 3.00E-06), is less than 100 kb downstream from *ARNT2* gene (Figure 5). *ARNT2* is a transcriptional regulator that is involved in several biological functions, including regulation of developmental genes and in expression study of Arnt2 in mouse tissues expressed in the molar and incisor teeth at a moderate level and in the odontoblasts, both inner and outer enamel epithelium, and in the stratum intermedium (Aitola & Pelto-Huikko, 2003). These findings might suggest a possible role of *ARNT2* gene during tooth development and could have a role also in influencing these correlated dental anomalies (AIR).

rs9913511 (P = 3.00E-06), which is intronic to *NTN1* (Figure 6), encodes a protein that belongs to a family of laminin-related secreted proteins. *NTN1* was found to affect the development of

craniofacial region based on animal models studies (Salminen, et al., 2000). *Ntn1* has been also shown to be expressed in the medial edges and oral sides of the palatal shelves in mice and was found to mediate a critical step in palatal fusion in mouse embryo (Serafini, et al., 1996; Ren, et al., 2004). Another study found that *ntn1* was expressed in the oral epithelium at the buccal side of the teeth germ where the vestibular sulcus formed in mice (Løes, et al., 2003). Previous studies have identified NTN1 as one of the susceptible genes of NSCL/P (Beaty et al., 2010; Beaty et al., 2011). Additionally, the expression pattern of *NTN1* has been found to be significantly upregulated in the dental pulp stem cell culture of NSCL/P patients (Sun, et al., 2015). There is supporting biological evidence that *NTN1* may have a role during odontogenesis and it may also influence these correlated dental anomalies (AIR).

chr22:31013419 (*P* = 3.00E-06) is upstream of *LIF* gene (Figure 7). Leukemia inhibitory factor (LIF), one of the major members of the interleukin-6 (IL-6) cytokine family that has an important role in the process of bone remodeling has been detected in vitro in many cell types, including fibroblasts, bone marrow stromal cells, osteoblasts and in the periodontal tissue, which suggests that *LIF* could play important role in periodontium remodeling (Liang, et al., 2011). However, we do not know if *LIF* has a role during tooth development and/or could have a role in increasing the risk of these associated dental anomalies (AIR).

rs2251904 (*P* = 4.00E-06) is downstream of the *TRPC4* gene (Figure 8) and the encoded protein of this gene plays a role during different biological processes, including neurotransmitter release and cell proliferation. In a sequencing analysis study of expression levels of canonical transient receptor potential (*Trpc4*) of rat during tooth germ development, *Trpc4* was highly expressed in the rat dental follicle and stellate reticulum cells during the early stage, and moderately expressed in odontoblasts (Yang, et al., 2017). These findings might suggest a possible role of

TRPC4 during tooth development and it might also influence these correlated dental anomalies (AIR).

rs1838002 (*P* = 6.00E-06) is upstream to *WNT2B* (Figure 9) which is one of the wingless-type MMTV integration site (WNT) family of signaling factors that play an important role in human development as well as tooth development. *WNT2B* plays an important role in the differentiation and proliferation of cementoblasts and odontoblasts (Yi Q, et al., 2017). Although there is no evidence of a direct relationship between *WNT2B* and AIR, *WNT2B* might influence those dental anomalies since it has a role during tooth development.

Another suggestive variant was rs6478094 (*P* = 1.10E-05), which is downstream to *COL27A1*(Figure 10) one of the fibrillary collagens family that encodes a proalpha chain of type XXVII collagen and is expressed in human and mouse epithelial cell layers in developing tissues, including tooth forming cells (ameloblast cells) (Boot-Handford, et al.,2003; Hjorten, et al.,2007). It is plausible that any alterations in the previously discussed genes may lead to defects in tooth morphology and structure.

We examined the candidate genes from previous sequencing and GWASs of tooth agenesis in our AIR GWAS module results to see if they had been replicated. GWAS of tooth agenesis was the only dental anomaly that was investigated before in a single ancestry and in a sample that has no other syndromes. Only limited evidence of suggestive associations was identified near tooth agenesis genes; *ASCL5/CACNA1S, ARHGAP15, FOXI3, EDAR*, *WNT10A, MSX1, PAX9,* and *AXIN2*. Association results for those genes are shown in Table 16. Regions of the genome that showed suggestive significance with tooth agenesis in the AIR model were further visualized using Regional Association (LocusZoom) plots (Figure 11).

Although we only wanted to focus on the multivariate approach of GWAS to investigate these associated dental anomalies and identify possible associated risk loci, we conducted a univariate

GWAS for each of these structural dental anomalies to identify possible risk loci associated with each one of them.

Table 15 Top GWAS hits for the Correlated Dental anomalies (AIR) in POFC

SNP	CHR	BP	F*	Weights*	*P*	Effect A.	MAF	Type
chr3:64541255	3	64541255	5.914	0.536,0.758,0.413	1.00E-06	G	0.0693	Imputed
rs412438	4	131979534	3.021	-0.478, -0.492,0.748	2.00E-06	A	0.09723	Imputed
rs140220410	15	81018892	2.508	-0.474, -0.736,0.524	3.00E-06	A	0.05183	Imputed
rs9913511	17	9019184	11.08	0.340, -0.322,0.883	3.00E-06	T	0.0898	Genotyped
chr22:31013419	22	31013419	7.703	0.311,0.506,0.805	3.00E-06	T	0.1309	Genotyped
rs2251904	13	38770991	7.978	0.422, -0.027,0.906	4.00E-06	A	0.1618	Imputed
rs1838002	1	112705328	8.079	0.274, -0.635,0.721	6.00E-06	G	0.3036	Imputed
rs8005462	14	97555659	8.112	0.124, -0.346,0.933	6.00E-06	G	0.4515	Genotyped
rs10749387	10	85381553	8.511	0.136, -0.344,0.932	7.00E-06	C	0.3272	Imputed
rs6483935	11	23422978	5.949	0.155,0.448,0.876	7.00E-06	G	0.3585	Imputed
rs78081702	16	21806883	8.663	0.150, -0.412,0.902	7.00E-06	G	0.3064	Imputed
rs2618611	20	17827285	6.189	-0.352, -0.133,0.930	7.00E-06	G	0.4027	Imputed
rs10929407	2	16154237	4.857	-0.099, -0.572,0.825	9.00E-06	A	0.09362	Imputed
rs4658356	1	90602946	7.725	0.378,0.129,0.925	1.10E-05	G	0.3485	Imputed
rs12563757	1	190642791	3.423	-0.489,0.260,0.818	1.10E-05	A	0.1013	Imputed
chr3:162308897	3	162308897	6.698	0.332, -0.569,0.739	1.10E-05	CTA	0.1988	Imputed
rs6478094	9	117387673	8.327	0.643, -0.006,0.766	1.10E-05	T	0.1312	Imputed

* F-Statistic tests the significance of the canonical correlations

* Weights reflect the correlation coefficients for each individual trait
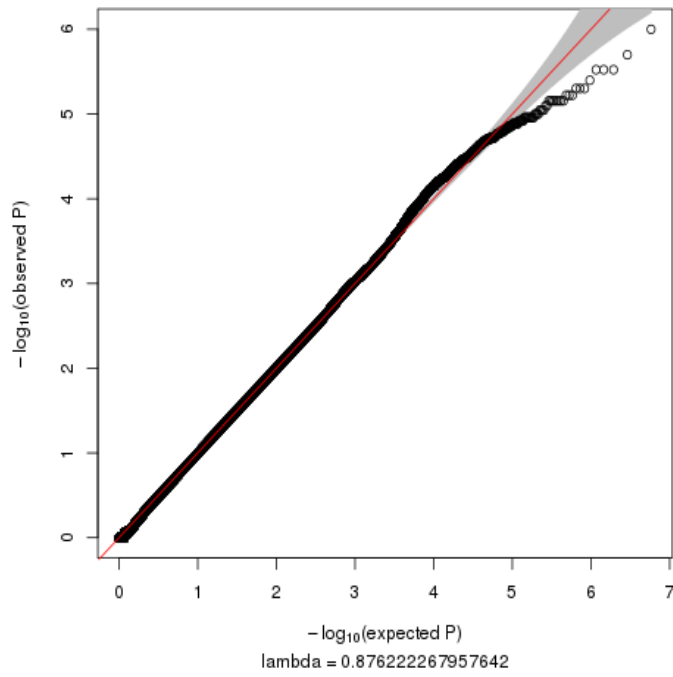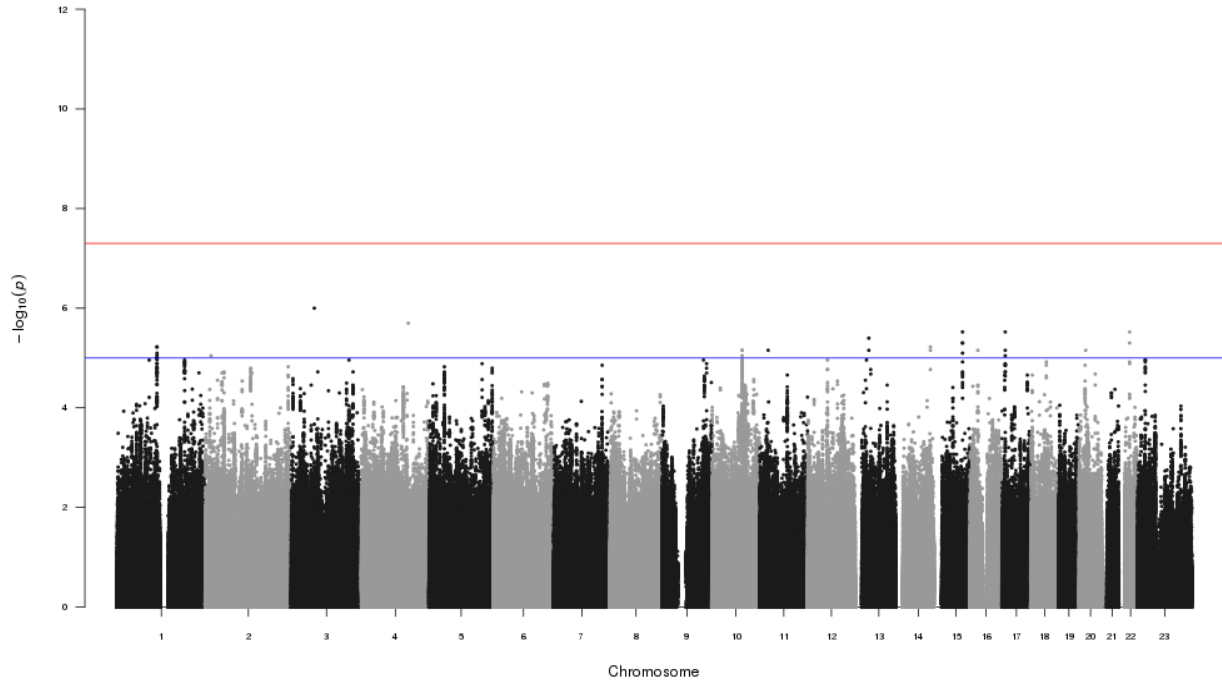
Figure 3 Manhattan and QQ plots for the Correlated Dental anomalies (AIR) in POFC
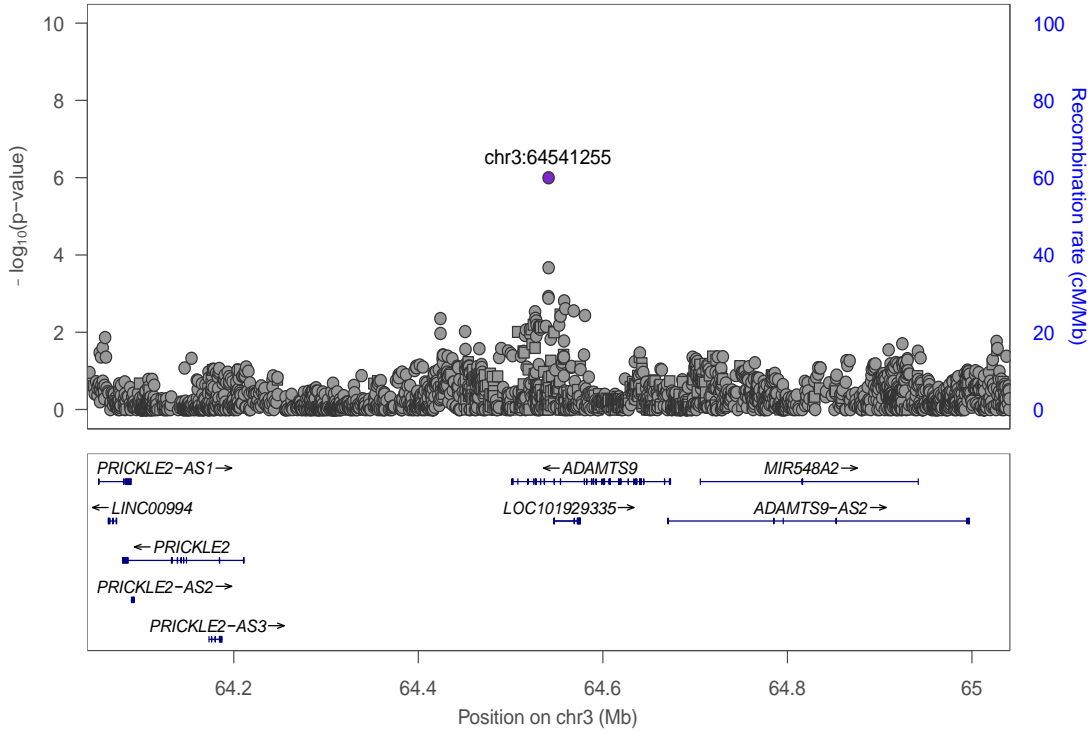
Figure 4 Regional Association Plot of chr3:6454125 from (AIR) results in POFC
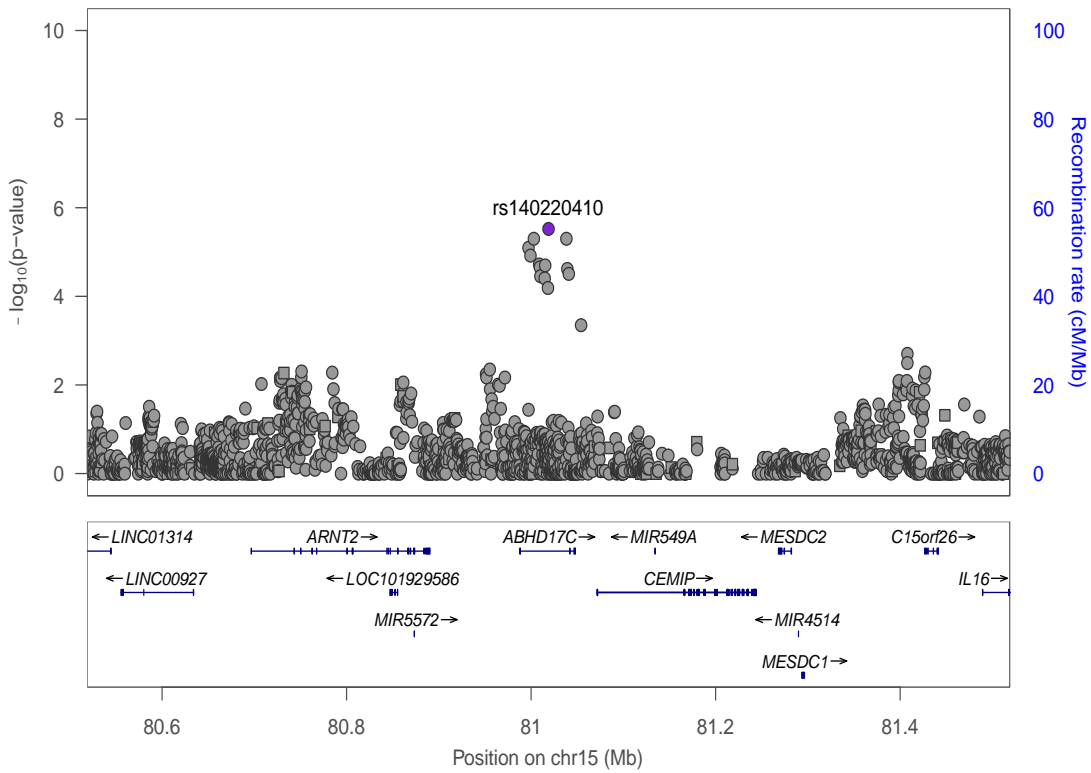
Figure 5 Regional Association Plot of rs140220410 from (AIR) results in POFC
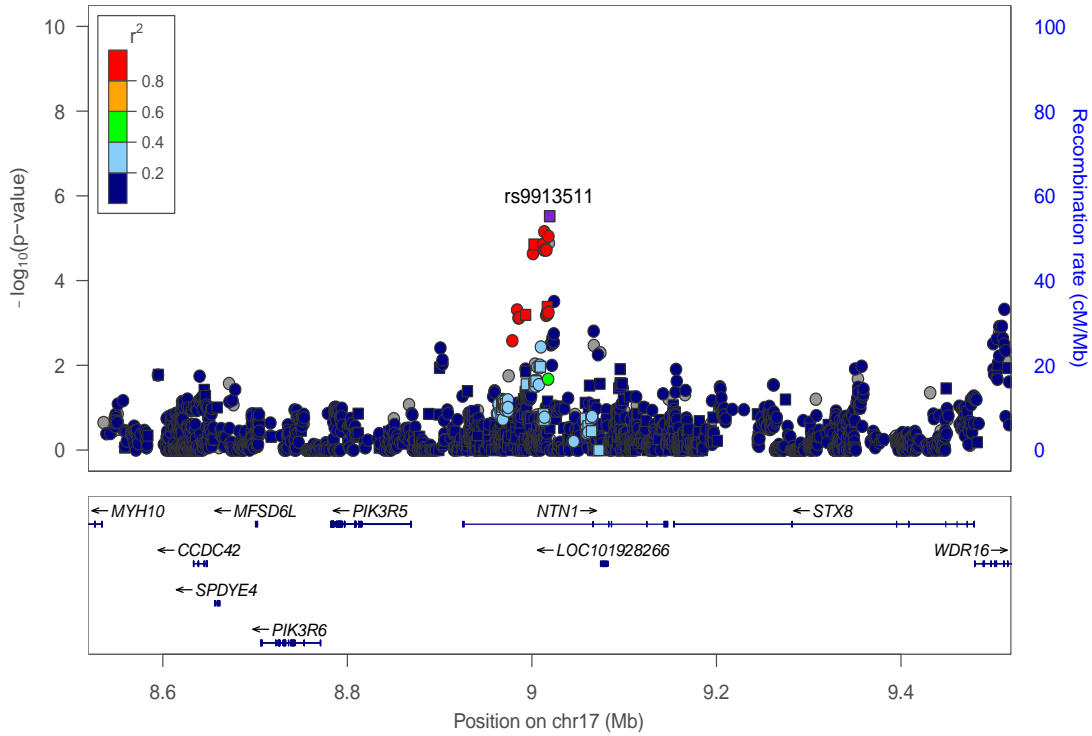
Figure 6 Regional Association Plot of rs9913511 from (AIR) results in POFC
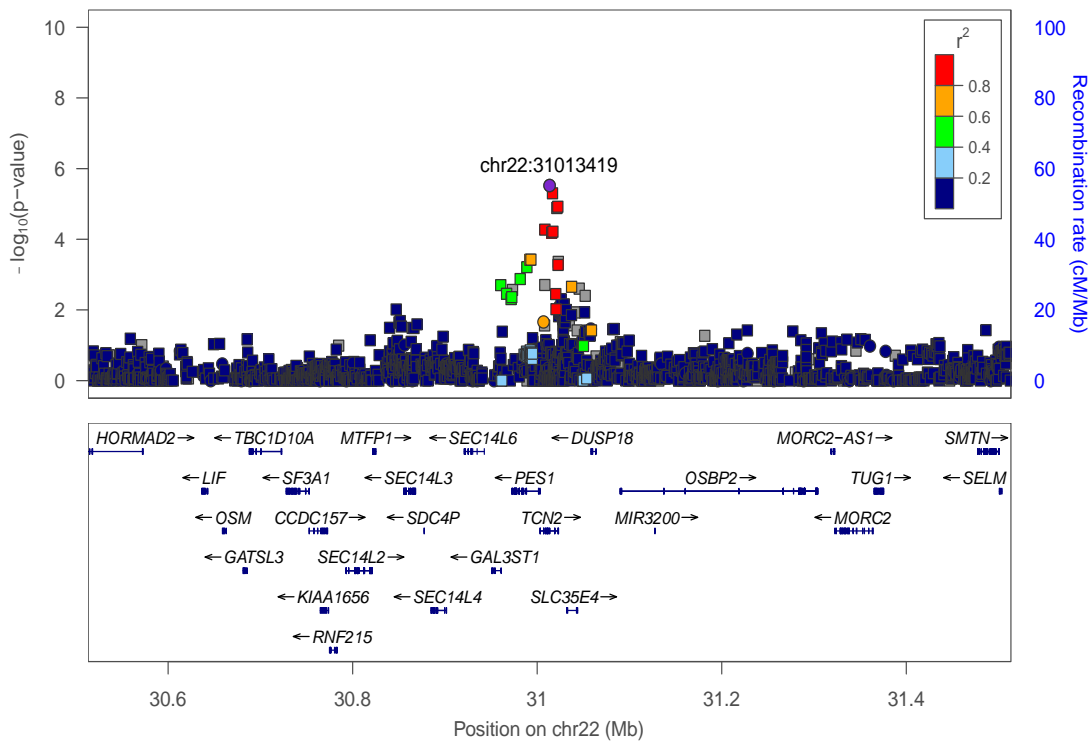
Figure 7 Regional Association Plot of chr22:31013419 from (AIR) results in POFC

64
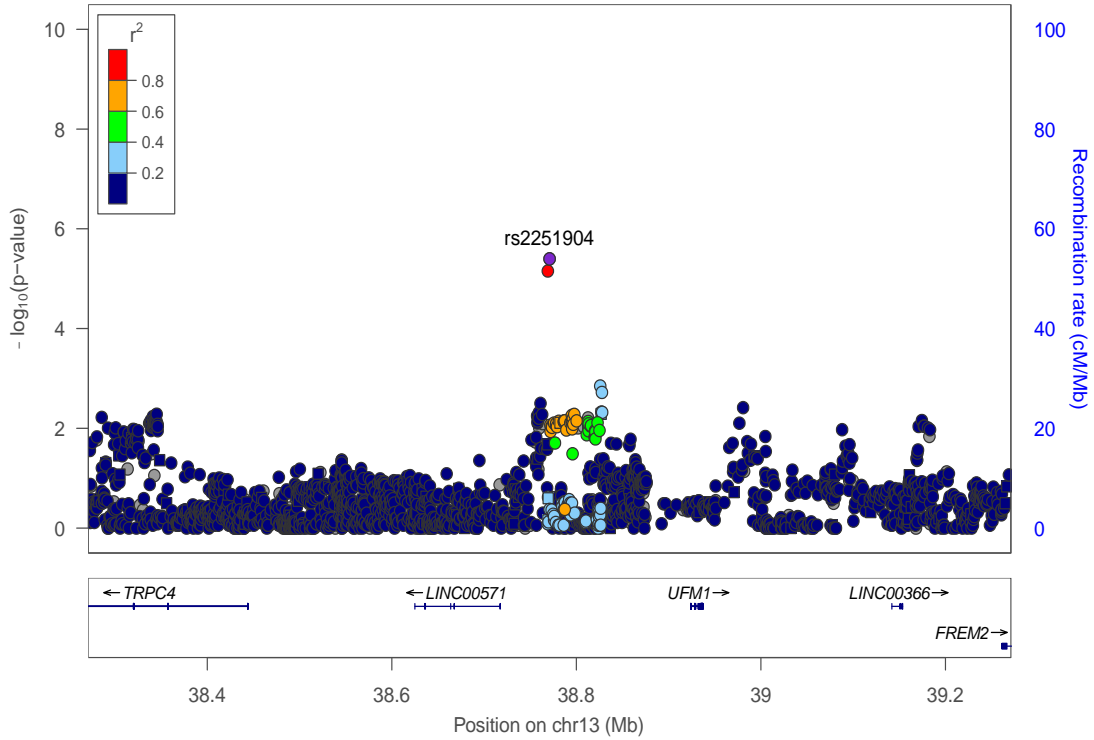
Figure 8 Regional Association Plot of rs2251904 from (AIR) results in POFC
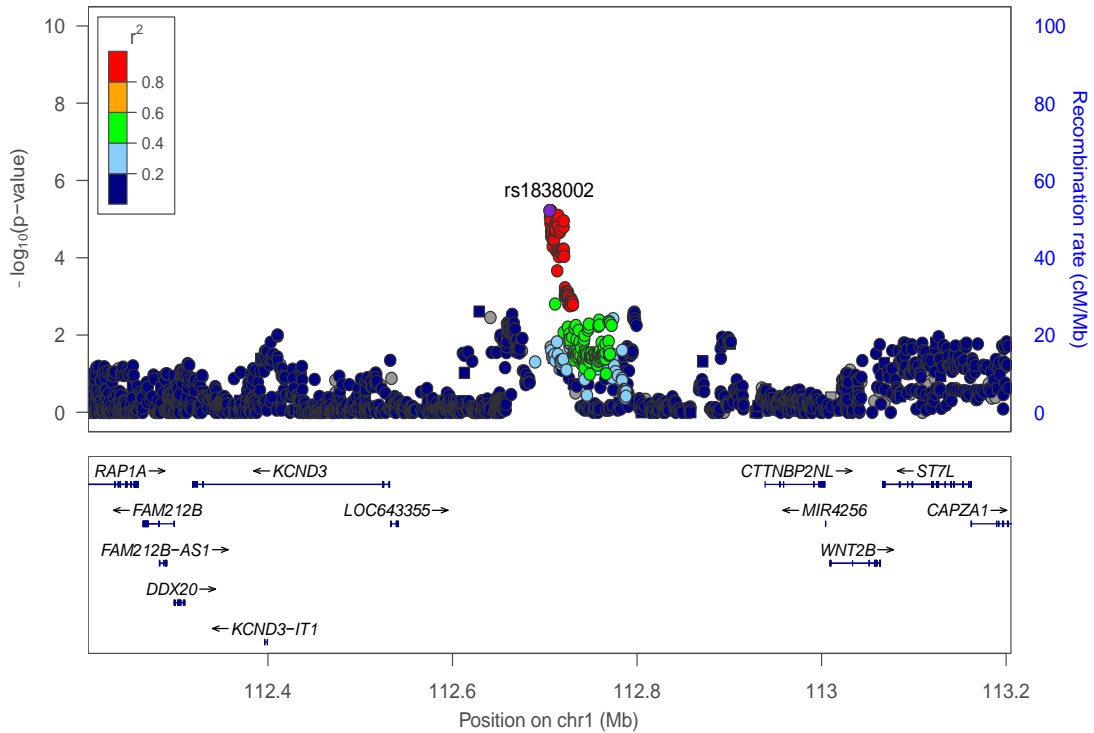
Figure 9 Regional Association Plot of rs1838002 from (AIR) results in POFC
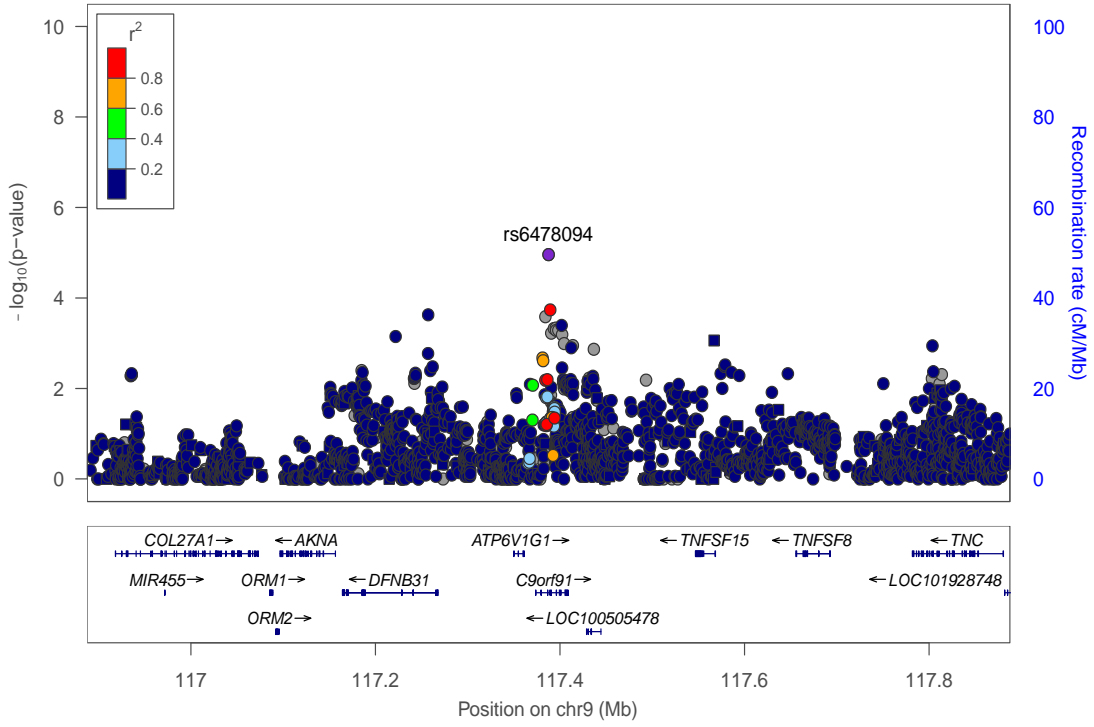
Figure 10 Regional Association Plot of rs6478094 from (AIR) results in POFC

Table 16 Association results of tooth Agenesis genes in (AIR) results in POFC

SNP	CHR	BP	*P*	Location to Gene	GENE	A1	Type
rs7572273	2	144174295	3.64E−4	Intronic	*ARHGAP15*	G	Imputed
rs55817509	1	201210545	3.99E−3	Downstream	*ASCL5/CACNA1S*	C	Imputed
rs28451574	17	63212699	4.6E−4	Upstream	*AXIN2*	T	Imputed
rs116280045	2	109563352	7.81E−3	Intronic	*EDAR*	T	Imputed
chr2:88344018	2	88344018	1.46E−2	Upstream	*FOXI3*	CAA	Genotyped
rs12106767	3	71745894	2.65E−4	Downstream	*FOXP1*	T	Imputed
rs57105923	4	109343076	2.3E−3	Downstream	*LEF1*	GAA	Imputed
rs6819062	4	4783737	2.57E−4	Upstream	*MSX1*	G	Imputed
rs4791130	17	65771213	7.13E−3	Downstream	*NOLL1*	T	Imputed
rs6571780	14	37547688	2.87E−3	Downstream	*PAX9*	A	Imputed
rs73993395	2	219879482	9.15E−3	Downstream	*WNT10A*	T	Imputed
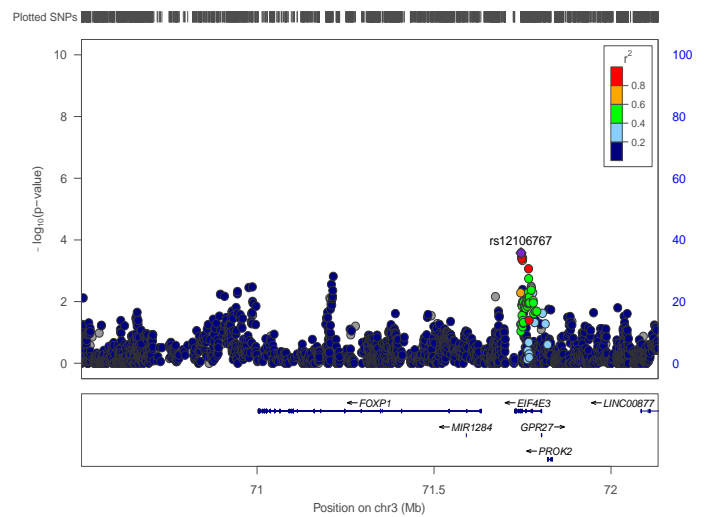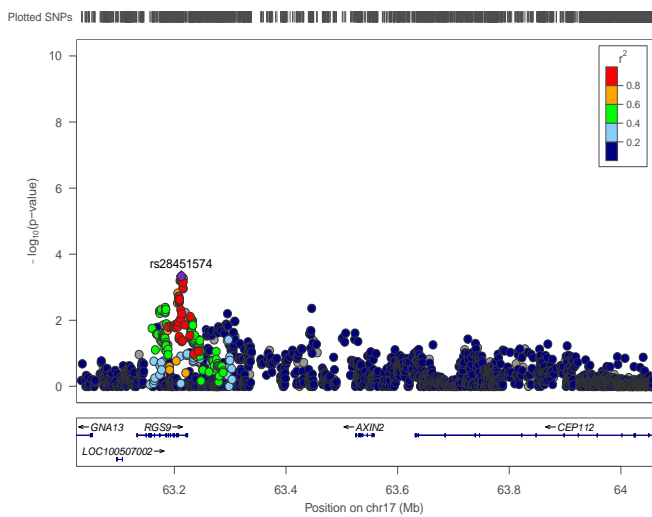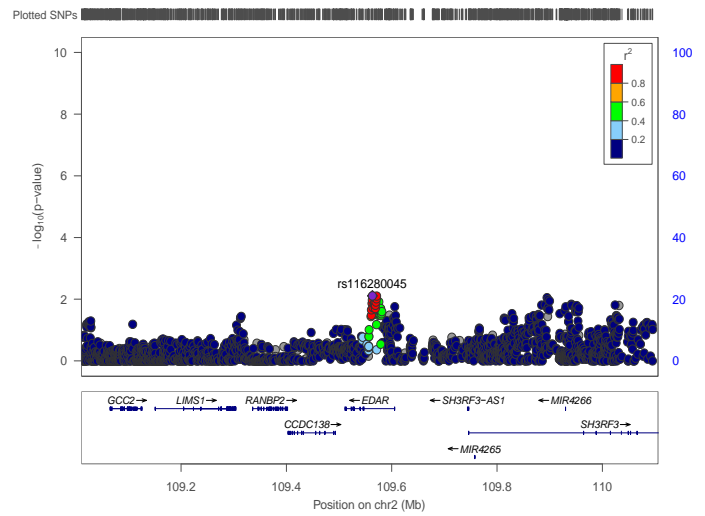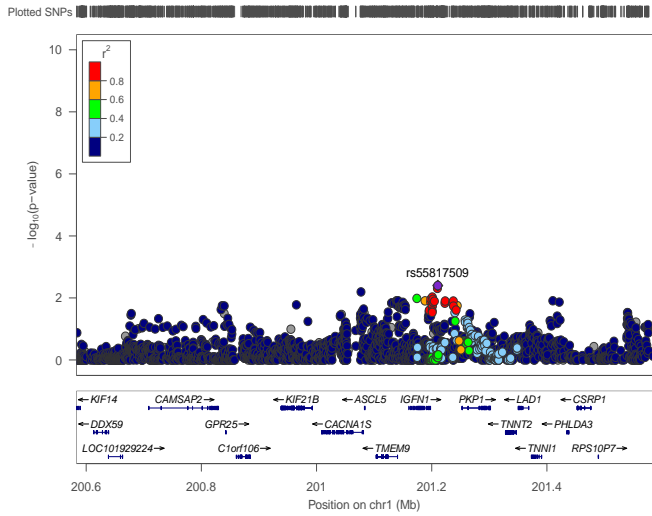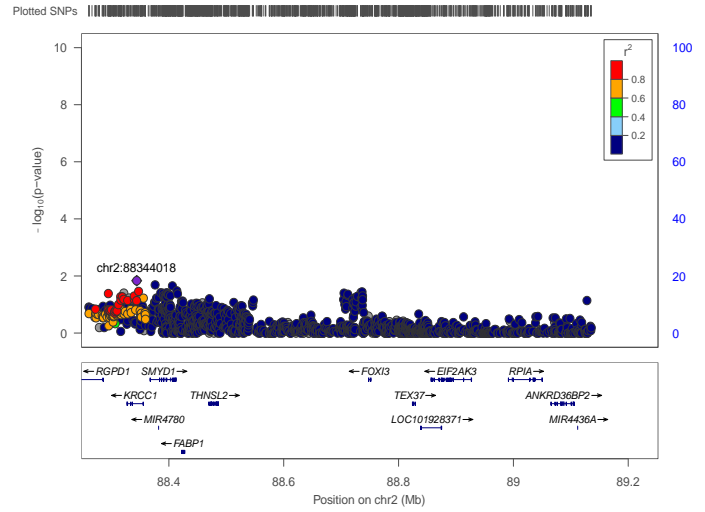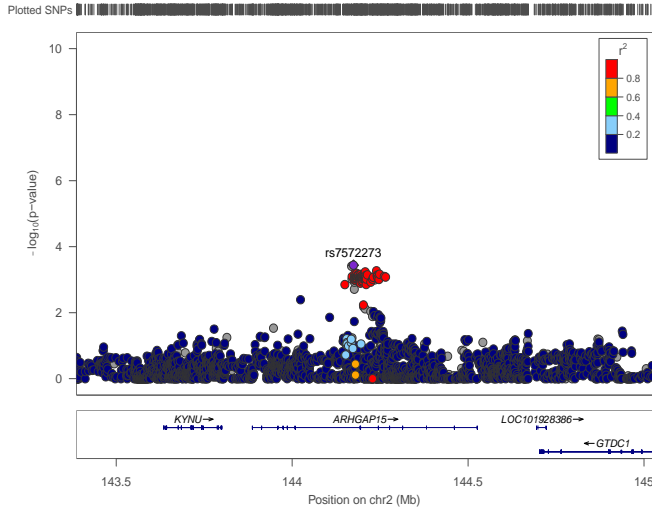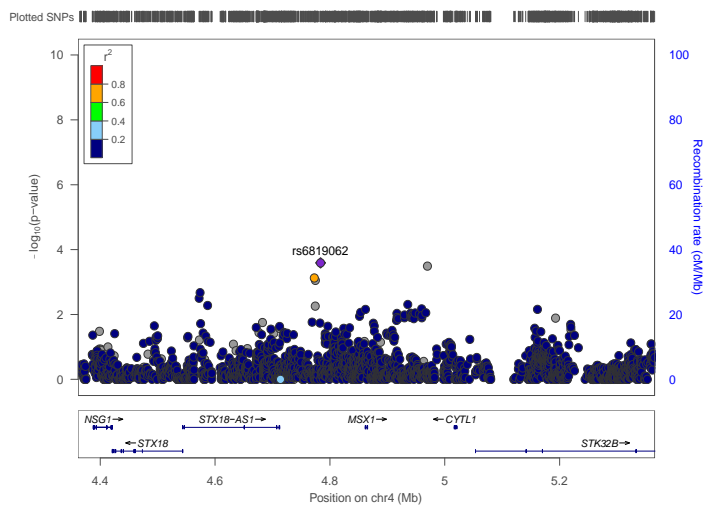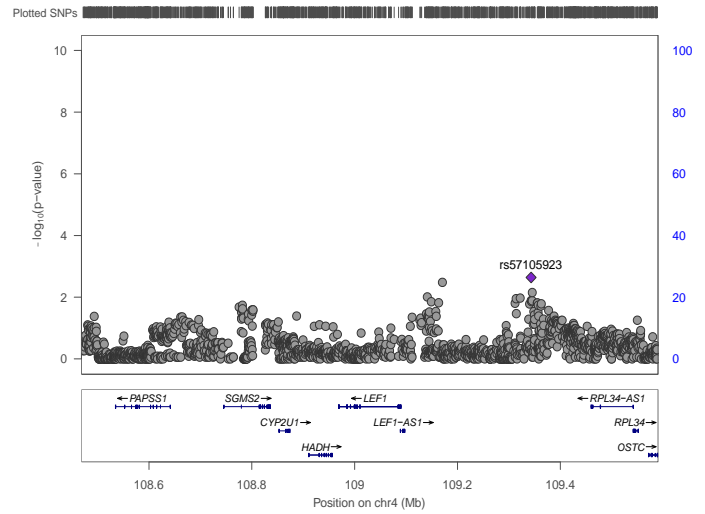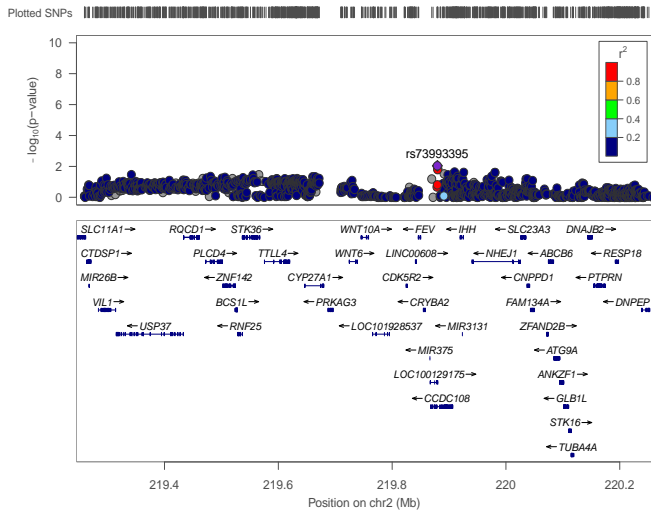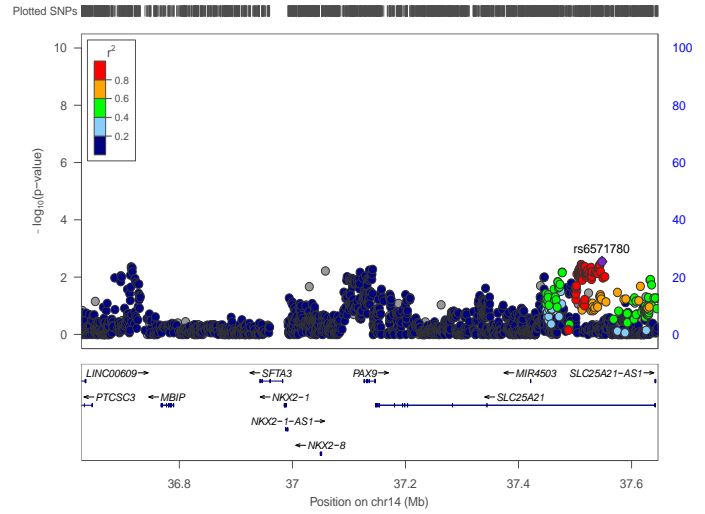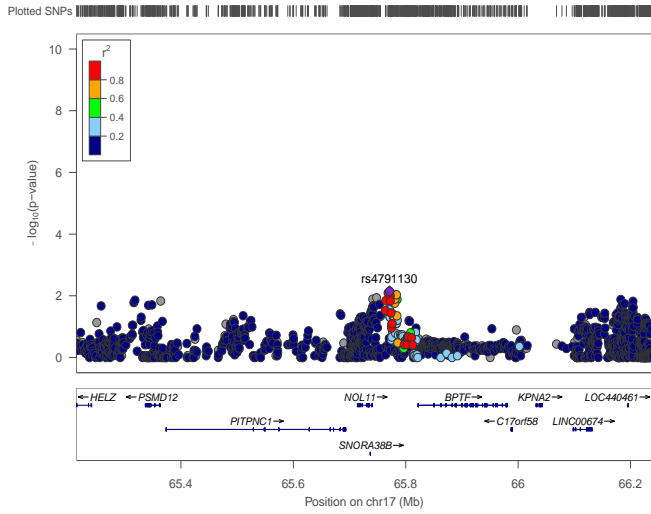
Figure 11 Regional Association Plots of Tooth Agenesis genes in (AIR) results

4.2.1.1 TOOTH AGENESIS GWAS RESULTS

The GWAS of tooth agenesis was performed on the POFC cohort (N=3579) for which genotypes data, phenotypes, and covariates data were available. A total of 7,168,909 SNPs was available for analysis after applying quality control criteria. In the GWAS of tooth agenesis, we were able to identify SNPs that had reached the genome-wide significance ($P < 5$ x 10^{-8}). In addition, there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Table 17. A visual representation of the association results of tooth agenesis is displayed in Figure 12. The genomic inflation factor (lambda) was =0.99, which indicative of no genomic inflation.

Table 17 Top GWAS hits for Tooth Agenesis in POFC

SNP	CHR	BETA	*P*	BP	A1	MAF	Type
rs201095681	8	-0.1808119	9.71E-09	7938916	T	0.2035	Imputed
rs36129019	7	-0.070572	2.74E-08	331086	A	0.2381	Imputed
rs200129154	13	-0.0897445	1.82E-07	19031198	TA	0.1021	Imputed
rs4868444	5	-0.0552329	2.62E-07	174160113	T	0.07077	Imputed
rs114372065	18	-0.0581726	2.78E-07	46542604	A	0.07131	Imputed
rs72890743	2	-0.0542171	7.23E-07	140776933	T	0.06084	Imputed
chr7:104473863	7	-0.0366013	8.84E-07	104473863	AAT	0.2198	Imputed
rs6758898	2	-0.0264772	8.90E-07	113382684	G	0.4478	Imputed
rs34284265	8	-0.0508968	9.71E-07	52775587	CT	0.1097	Imputed
rs56404075	2	-0.0482207	1.08E-06	60631365	C	0.07059	Imputed

lambda = 0.998369246115887
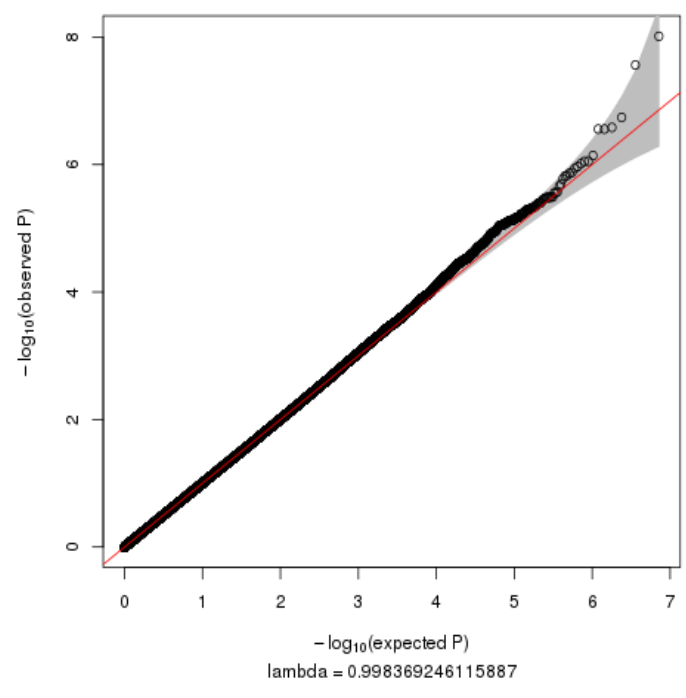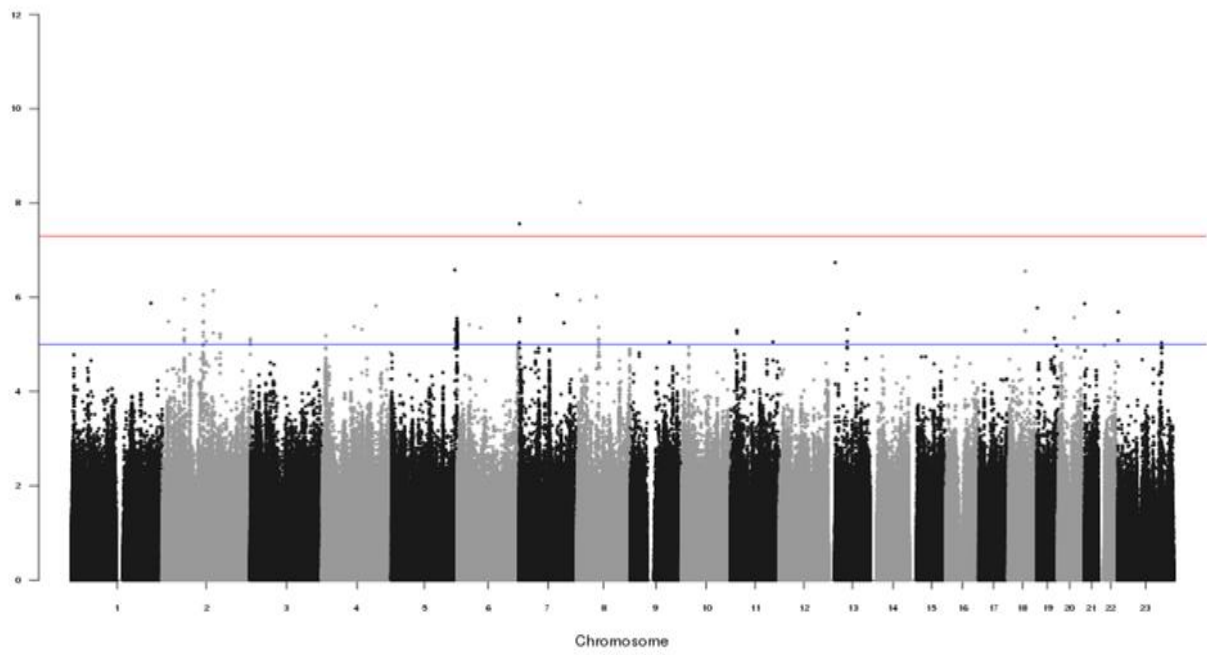
Figure 12 Manhattan and QQ plots for Tooth Agenesis in POFC

Regions of the genome that showed significant and suggestive significance with Agenesis were further visualized using Regional Association (LocusZoom) plots in order to assess whether the region may have a possible role in tooth development and dental/oral health.

One of our top significant variants was rs36129019 (beta = -0.070572, P = 2.74E-08), which is downstream to *FAM20C* (Figure 13), this gene is also known as dentin matrix protein 4 (*DMP4*), and the encoded protein binds calcium and phosphorylates proteins involved in bone mineralization. A previous expression study of *Fam20c* in mice found that *Fam20c* is highly expressed in bone and tooth and it plays a crucial role in the differentiation and mineralization processes of dental tissues by regulating molecules essential to the differentiation of tooth-formative cells (Wang, et al., 2012). A previous study indicated that mutations in the *FAM20C* gene could leads to bone and craniofacial/dental abnormalities, known as Raine syndrome (Simpson, et al., 2007). Another study that investigated the role that *FAM20C* plays in the odontoblastic differentiation of the human Dental Pulp Cells (hDPCs), found that *FAM20C* was elevated during odontoblastic differentiation. Knockdown *FAM20C* repressed the expression of DSPP and DMP1, those protein are essential for proper tooth development, which indicate that *FAM20C* was positively involved in regulating hDPC odontoblastic differentiation (Li, et al., 2018). It is plausible that *FAM20C* expression may influence tooth agenesis because of its role in tooth development, and alterations in its function may lead to increase the risk of developing tooth agenesis.

One of our suggestive signals was near rs4868444 (beta = -0.0552329, P = 2.62E-07), which is intronic to *MSX2* (Figure 14). The encoded protein of *MSX2* provides the balance between survival and apoptosis of neural crest-derived cells, which is necessary for proper craniofacial morphogenesis. In addition, *msx2* protein is part of the bone morphogenic protein (BMP)

signaling pathway, and this pathway regulate various processes, including odontogenesis (Davidson, 1995). A mutation in msx2 has been found in a family with amelogenesis imperfecta and impaired tooth eruption (Suda, et al., 2006), and *Msx2* knockout mice showed a tooth abnormality resembling amelogenesis imperfecta (Satokata, et al., 2000). All of these observations support the possible role that *MSX2* plays during tooth development and the possibility of *MSX2* influence on tooth agenesis.

Another suggestive signal was detected near rs114372065 (beta = -0.0581726, *P* = 2.78E-07), which is upstream to *SMAD7* gene (Figure 15). Previous investigation on a mouse module demonstrated that Smad7, a member of the smad family, plays an essential role in regulating TGF-β signaling during tooth development (Ito, et al., 2001). In addition, previous GWAS of dental caries in permanent dentition found that *SMAD7* is associated with dental caries (Shaffer et al. 2013). These results suggest that *SMAD7* plays a role during tooth development, although no evidence was found to link it to tooth agenesis specifically.

A suggestive signal was detected near rs6758898 (beta = -0.0264772, *P* = 8.90E-07) is downstream of *IL1A* and *IL1B* (Figure 16)*,* which belong to the interleukin 1 cytokine family. The encoded proteins of these genes have multiple and different functions within the immune system. Interleukin-1 alpha (*IL1A*) has a role in bone resorption, and positively affects the survival and differentiation of osteoclasts and odontoclasts (Song, et al., 2013), while Interleukin-1 beta (*IL1B*) has an important function as a mediator in the inflammatory response (Karimbux, et al., 2012). Previous study found that some variants in *IL1A* and *IL1B* have been associated with severe or progressive chronic periodontitis in Caucasians (Nikolopoulos, et al., 2008). While it is unclear how *IL1A* and *IL1B* may influence tooth agenesis health, there is supporting evidence that it could affect oral health.
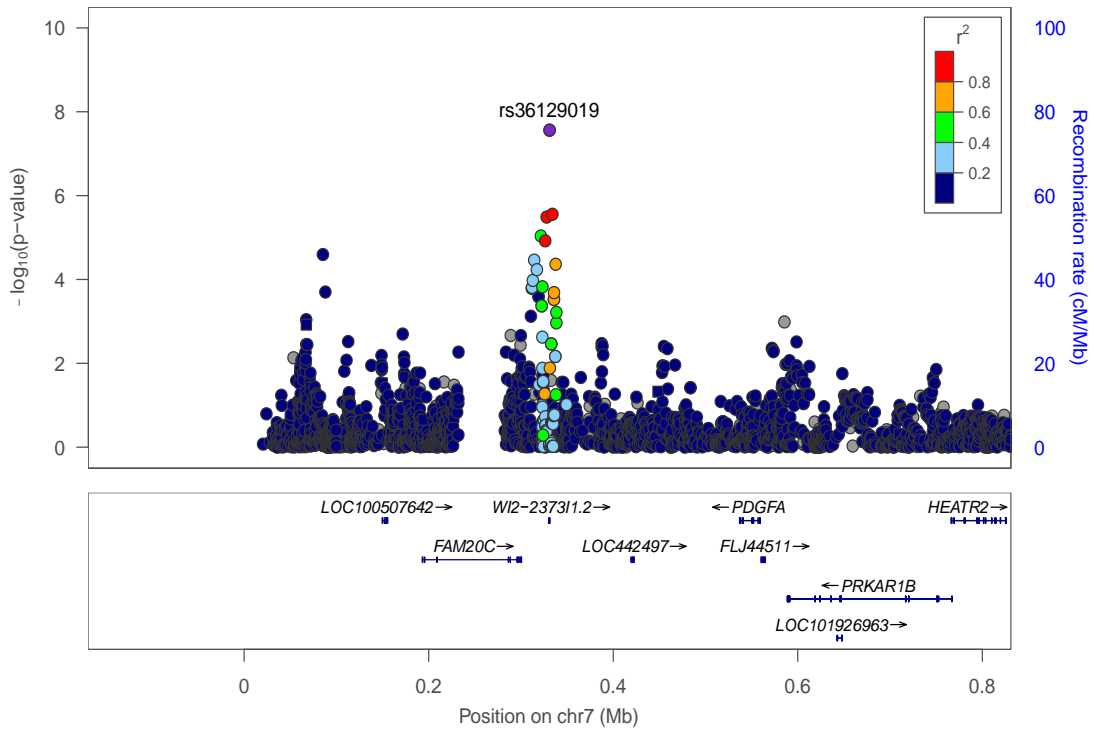
Figure 13 Regional Association Plot of rs36129019 from Tooth Agenesis results in POFC
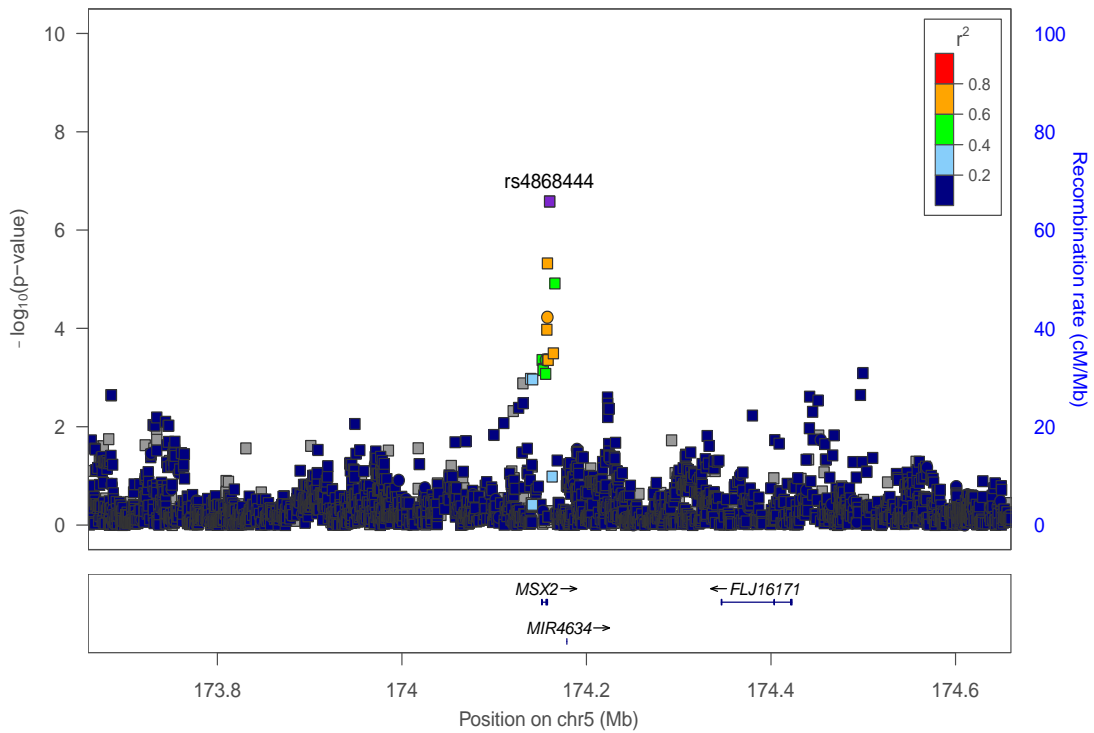
Figure 14 Regional Association Plot of rs4868444from Tooth Agenesis results in POFC
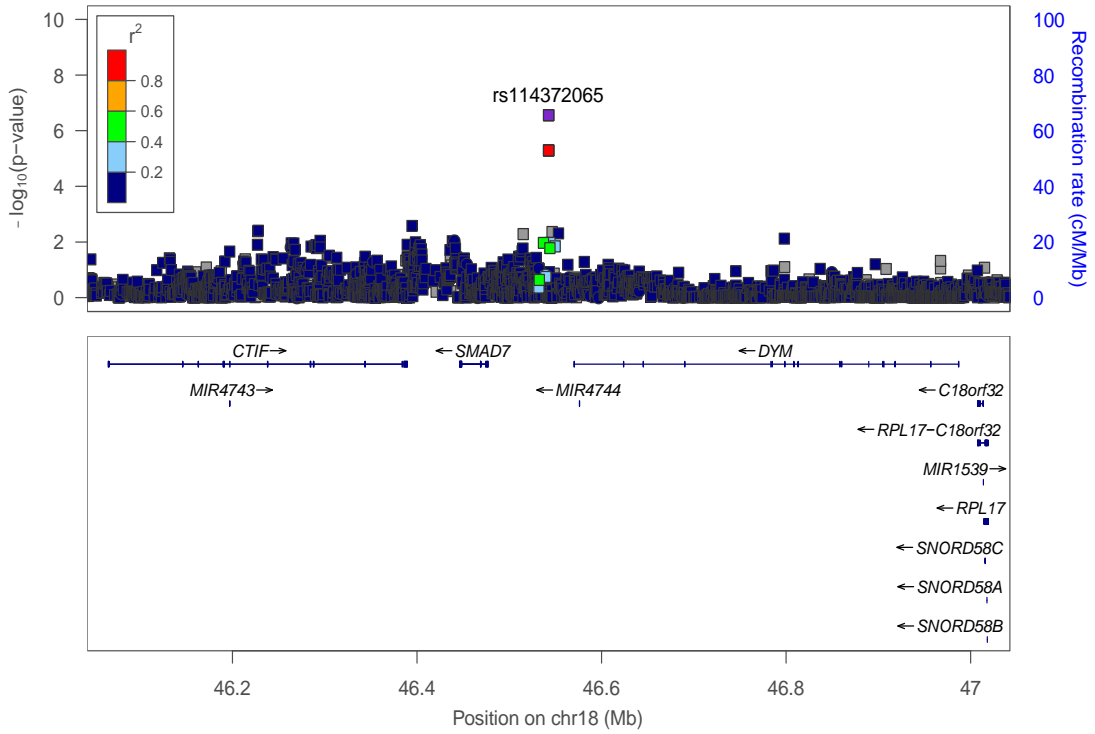
Figure 15 Regional Association Plot of rs114372065from Tooth Agenesis results in POFC
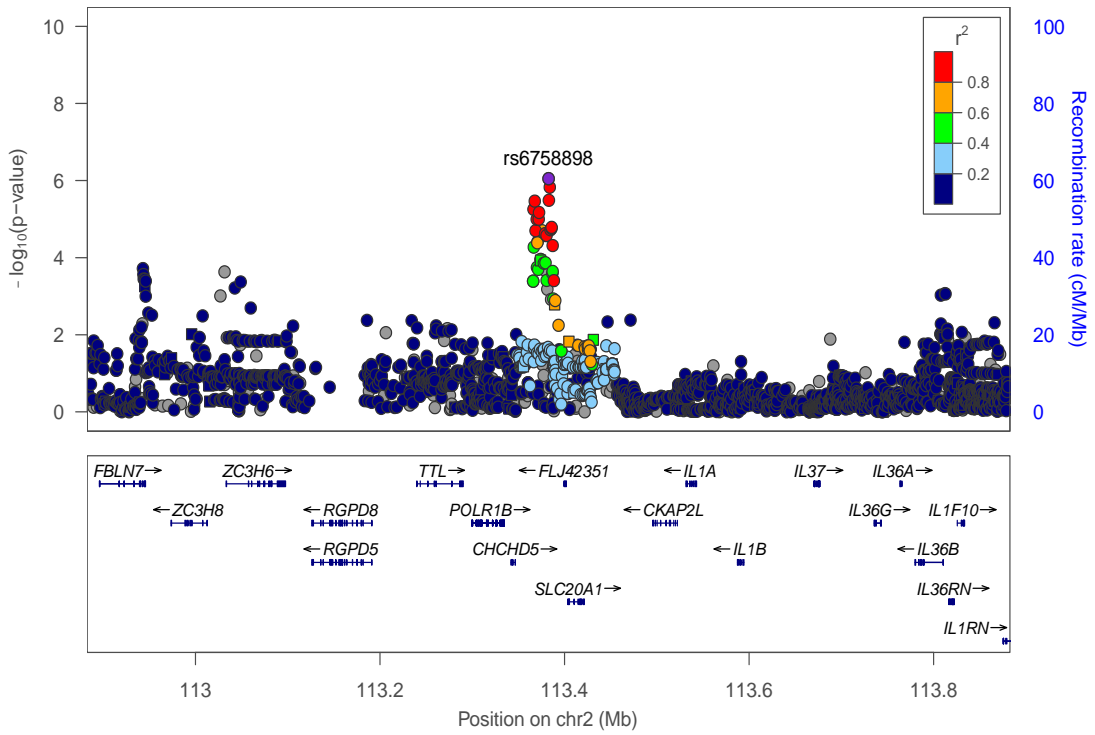
Figure 16 Regional Association Plot of rs6758898 from Tooth Agenesis results in POFC

We examined the candidate genes from previous sequencing and GWASs of tooth agenesis in our GWAS results of tooth agenesis to see if they could be replicated. We did not have association results for all of them. However, association results for the ones we have in our sample are shown in Table 18. Regions of the genome that showed suggestive significance with tooth agenesis genes in the GWAS results of tooth agenesis in our sample were further visualized using Regional Association (LocusZoom) plots (Figure 17).

Table 18 Association results of Tooth Agenesis genes in Tooth Agenesis GWAS in POFC

SNP	CHR	BP	*P*	Location to Gene	GENE	A1	Type
rs2034604	2	143958745	2.69E−4	Intronic	*ARHGAP15*	T	Imputed
rs58035895	2	109129763	6.59E−3	Upstream	*EDAR*	T	Imputed
rs371389899	2	88734499	2.1E−3	Intronic	*FOXI3*	T	Genotyped
rs9828296	3	70964533	1.24E−4	Upstream	*FOXP1*	T	Imputed
rs11723882	4	109126816	4.77E−6	Downstream	*LEF1*	C	Imputed
rs111378805	17	65428387	3.77E−4	Upstream	*NOLL1*	T	Imputed
rs148015836	2	219656787	8.68E−5	Upstream	*WNT10A*	TAC	Imputed
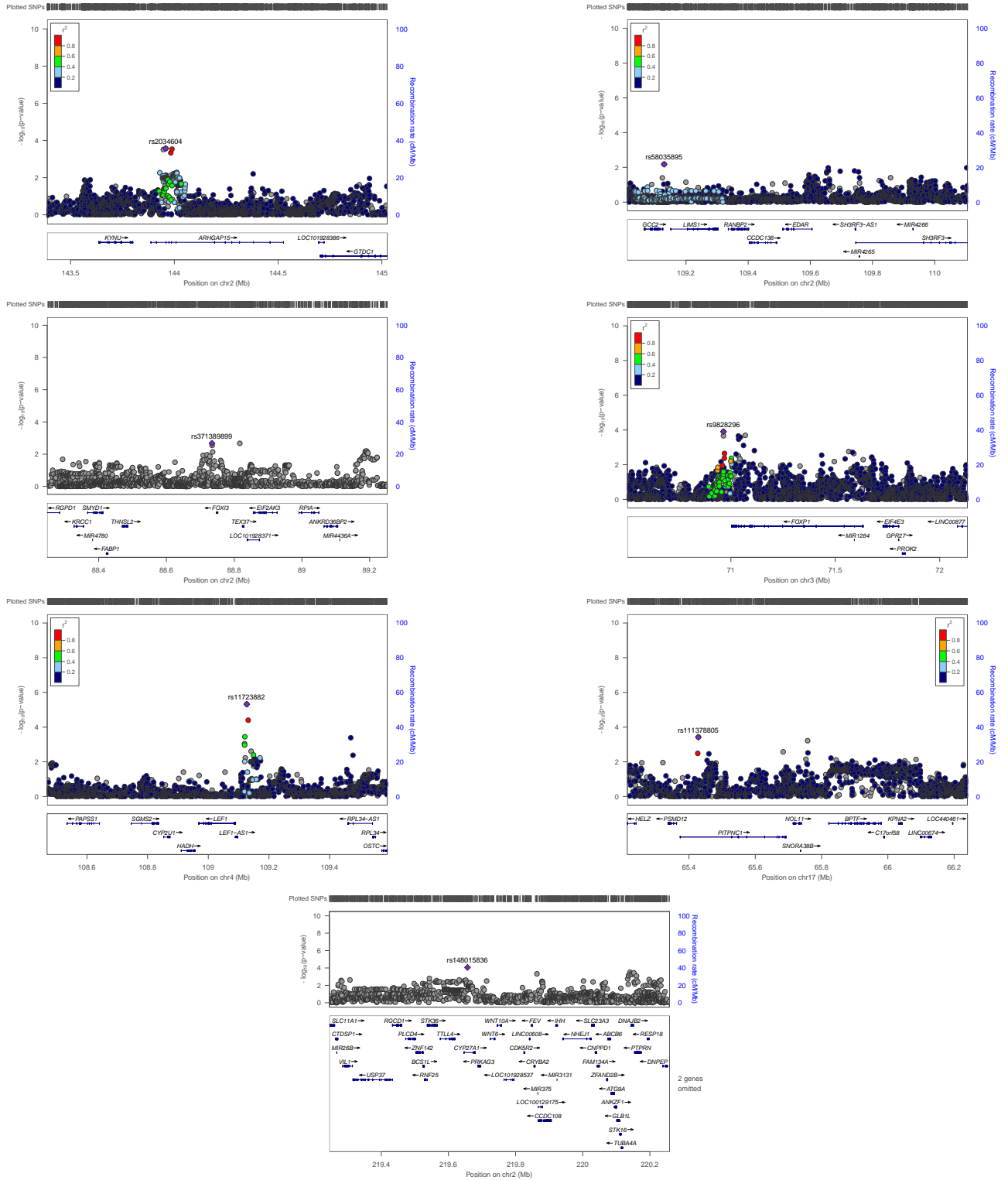
Figure 17 Regional Association Plots of Tooth Agenesis genes in Tooth Agenesis GWAS

4.2.1.2 IMPACTION GWAS RESULTS

The GWAS of tooth impaction was performed on cohort (N=3579) for whom genotypes data, phenotypes, and covariates data were available. A total of 7,168,909 SNPs was available for analysis after applying quality control criteria. In the GWAS of impaction, we were able to identify several SNPs that had reached the genome-wide significance ($P < 5$ x 10^{-8}). In addition, there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Table 19 Top GWAS hits for Tooth Impaction in POFC. A visual representation of the association results of tooth impaction is displayed in Figure 18. The genomic inflation factor (lambda) was =0.99, which indicate no genomic inflation.

Table 19 Top GWAS hits for Tooth Impaction in POFC

SNP	CHR	BP	*P*	Type	BETA	MAF	A1
rs2857355	14	106368614	2.24E-09	Imputed	-0.0710444	0.0731	T
rs61672846	7	110319796	3.53E-09	Imputed	-0.0383921	0.05384	C
rs6972307	7	25561172	3.65E-09	Imputed	-0.0359801	0.06329	T
chr16:68278089	16	68278089	1.94E-08	Imputed	-0.0316074	0.06847	C
rs34436159	11	78925040	2.36E-08	Imputed	-0.0414259	0.06919	A
rs4648737	1	1881019	2.38E-08	Imputed	-0.0371759	0.05925	G
rs9858484	3	67165389	2.68E-08	Imputed	-0.0301183	0.07014	T
chr7:6412423	7	6412423	2.92E-08	Imputed	-0.036966	0.05877	C
rs371522050	2	89336361	3.99E-08	Imputed	-0.0705058	0.06553	C
rs340040	5	145266742	6.73E-08	Imputed	-0.0400831	0.05096	C
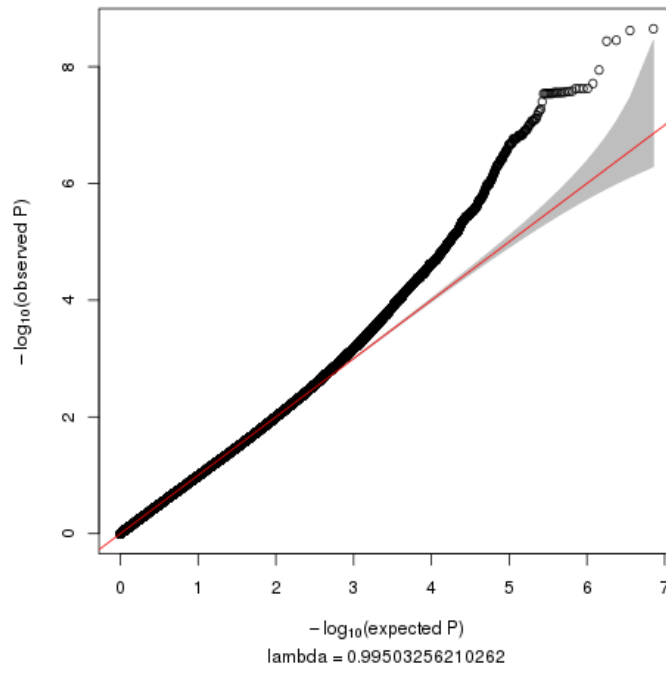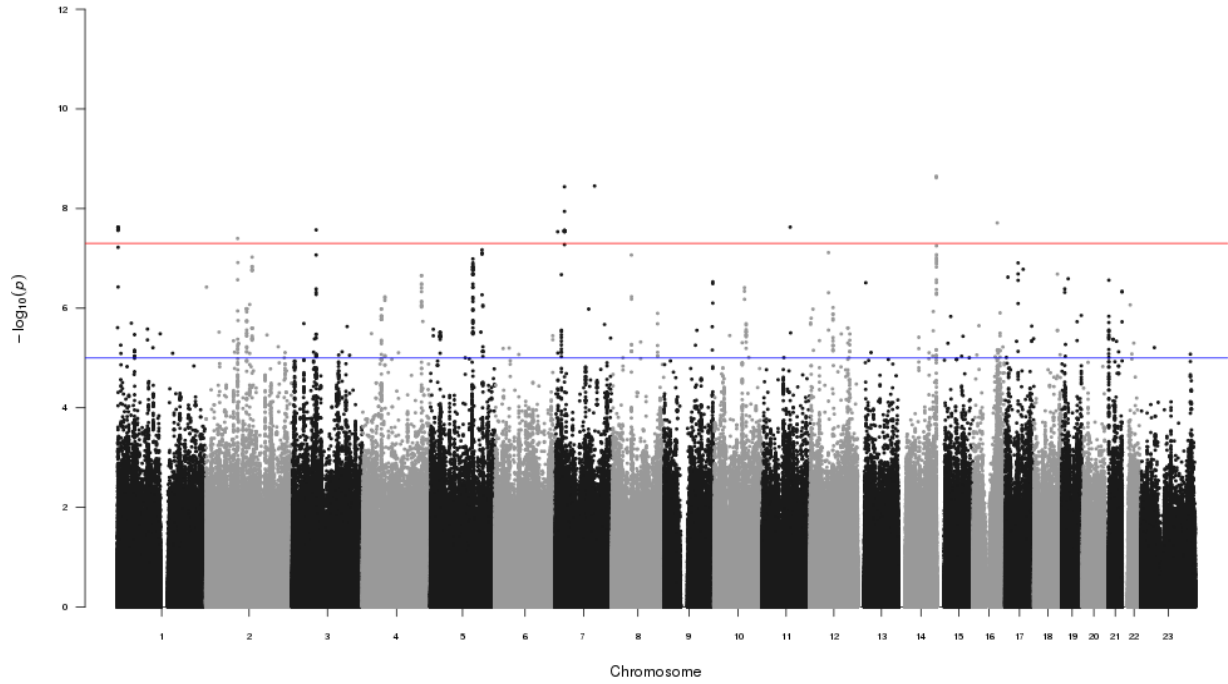
Figure 18 Manhattan and QQ plots for Impaction in POFC

One of the top significant association signals was near rs2857355 (beta = -0.0710444, *P* = 2.24E-09), which is intronic to *KIAA012* (Figure 19). The encoded protein of this gene is one of the noncoding RNAs that has been found to be expressed in Ameloblastoma and adenomatoid odontogenic tumor, which are tumor of the jaw that derived from the teeth forming apparatus (Dinize, et al., 2018). However, we are not certain if it has a role in tooth development and/or oral health.

Another significant association signal was detected near chr16:68278089 (beta = -0.0316074, *P* = 1.94E-08), which is upstream to *CDH1* gene (Figure 20). The encoded protein of *CDH1* is found within the membrane that surrounds epithelial cells, such as the cells inside of the mouth and it plays a role during craniofacial development, which includes the development of teeth (Figueiredo, et al., 2013). A study that investigated the association between *CDH1* and tooth agenesis and cleft lip/palate found that there is an association between *CDH1* and CL/P with tooth agenesis (Letra, et al., 2009). It is still however unclear how *CDH1* may affect dental health and/or influence impaction.
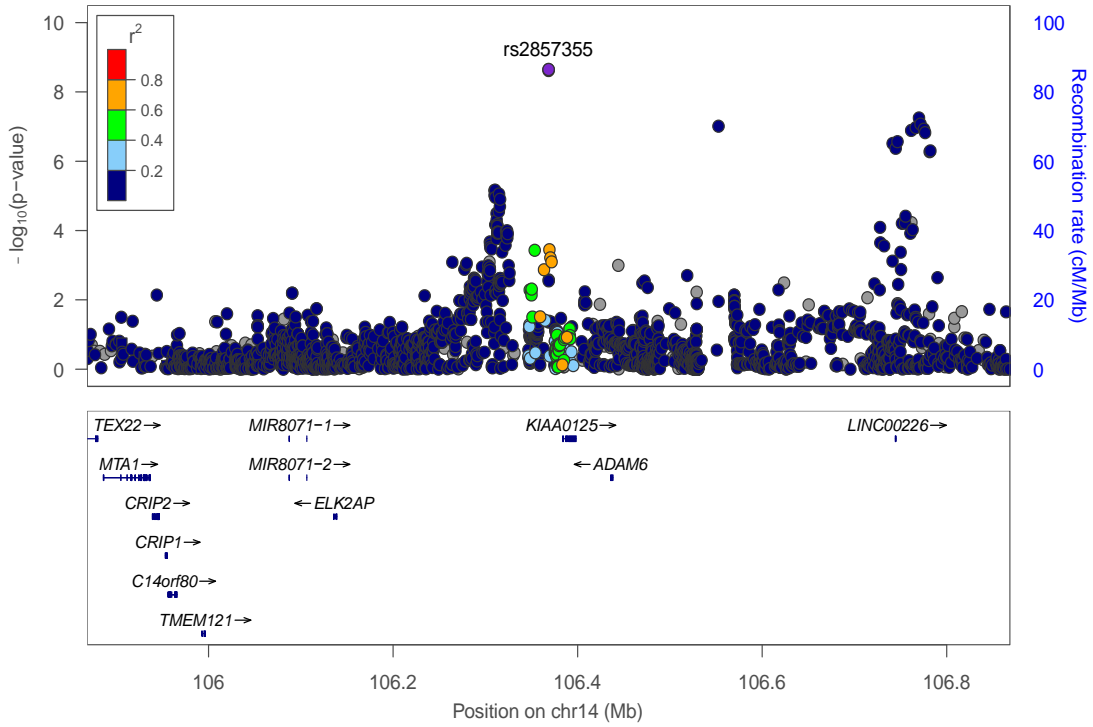
Figure 19 Regional Association Plot of rs2857355 from Impaction GWAS in POFC
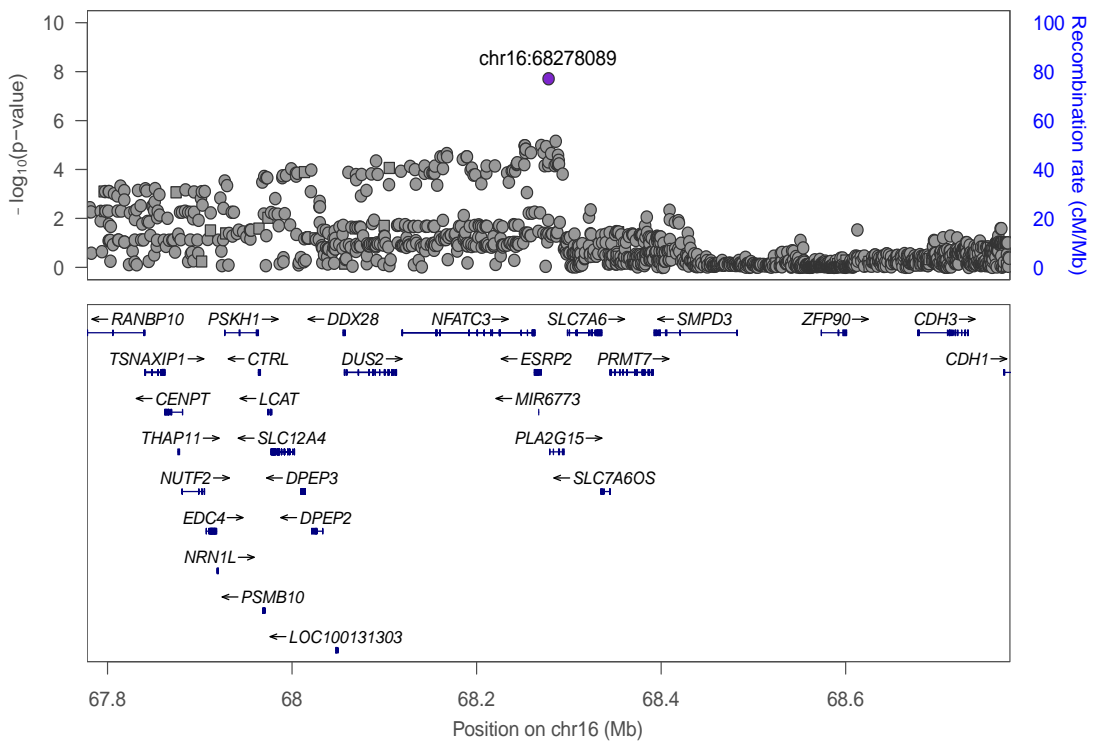
Figure 20 Regional Association Plot of chr16:68278089 from Impaction GWAS in POFC

4.2.1.3 ROTATION GWAS RESULTS

The GWAS of tooth rotation was performed on cohort (N=3579) for whom genotypes data, phenotypes, and covariates data were available. A total of 7,168,909 SNPs was available for analysis after applying quality control criteria. In the GWAS of rotation, we were able to identify SNPs that had reached the genome-wide significance ($P < 5 \times 10^{-8}$). In addition, there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Table 20. A visual representation of the association results of rotation is displayed in Figure 21. The genomic inflation factor (lambda) was=0.99, which indicate no genomic inflation.

Table 20 Top GWAS hits for Tooth Rotation in POFC

SNP	CHR	BETA	*P*	BP	A1	MAF	type
rs72697811	14	0.03419317	1.66E-08	86381697	T	0.2798	Imputed
rs12425757	12	0.04549168	2.32E-07	22007944	T	0.1057	Imputed
rs77781808	18	0.04985617	4.96E-07	5883833	G	0.07168	Imputed
rs1025623	12	0.030015	5.00E-07	31406872	A	0.242	Imputed
rs149049213	14	0.05357958	5.15E-07	97209699	ATTTT	0.06811	Imputed
rs4725189	7	-0.0476394	1.12E-06	9393563	T	0.06805	Imputed
rs60409848	4	-0.0288448	1.41E-06	22760979	C	0.2552	Imputed
rs28493495	4	-0.0475661	1.43E-06	116612312	T	0.07005	Imputed
rs2015683	10	0.04909655	1.64E-06	33943396	T	0.06986	Imputed
rs11767363	7	0.03630944	1.66E-06	116140275	T	0.1183	Imputed
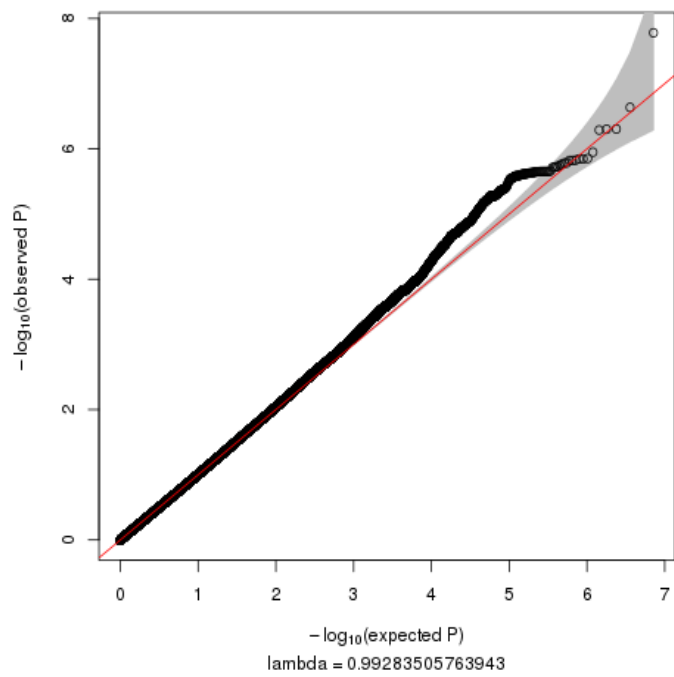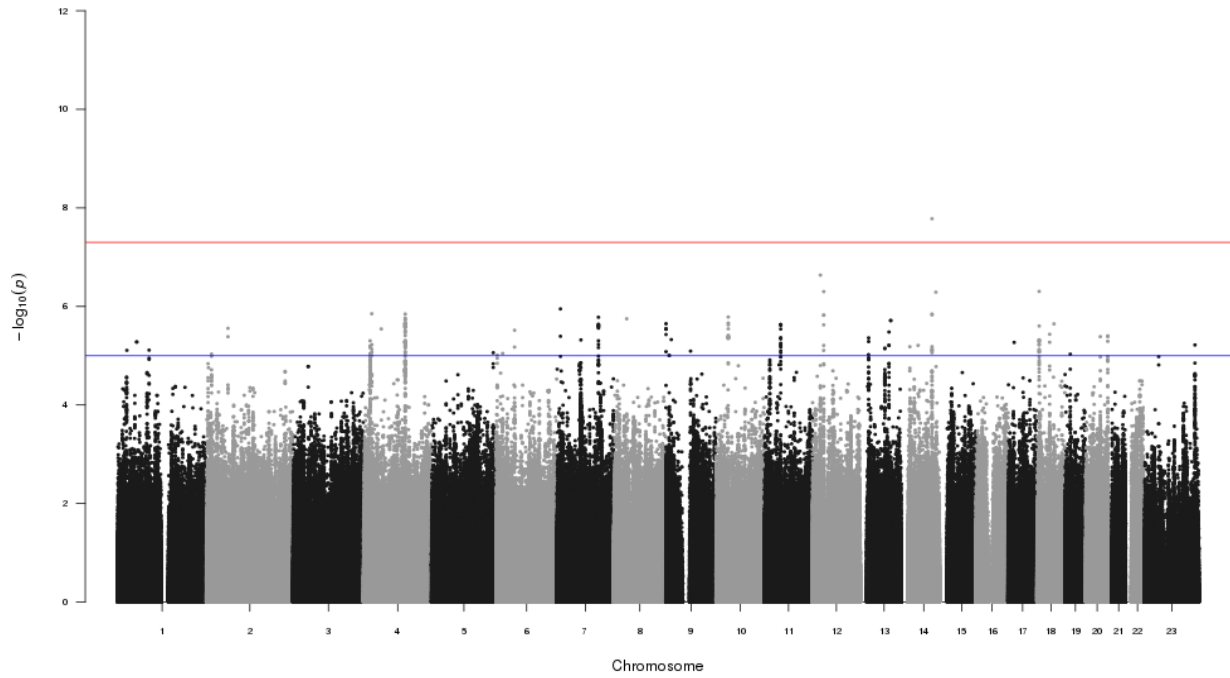
Figure 21 Manhattan and QQ plots for Rotation in POFC

One of the significant association signals was detected near rs72697811 (beta = 0.03419317, P = 1.66E-08), which is downstream of *FLRT2* gene (Figure 22), the encoded protein of this gene belongs to Fibronectin Leucine-Rich Transmembrane (Flrt) gene family, which has been previously reported to be expressed in the developing embryo. In a gene expression study of flrt2 in mice, it was found that flrt2 was expressed during early craniofacial development and it was highly expressed in the cranial neural crest cells. In addition, flrt2 was highly expressed in the neural crest-derived mesenchyme in the medial aspect of the developing frontonasal region in close relationships with the expression of Fgfr2, Shh, and Msx1, and those genes have shown previously to play critical roles during craniofacial development, and tooth development in particular (Gong, et al., 2009). Although no evidence was found to link *FLRT2* gene to tooth rotation, it is plausible that *FLRT2* could have a role in tooth development and could influence tooth rotation.
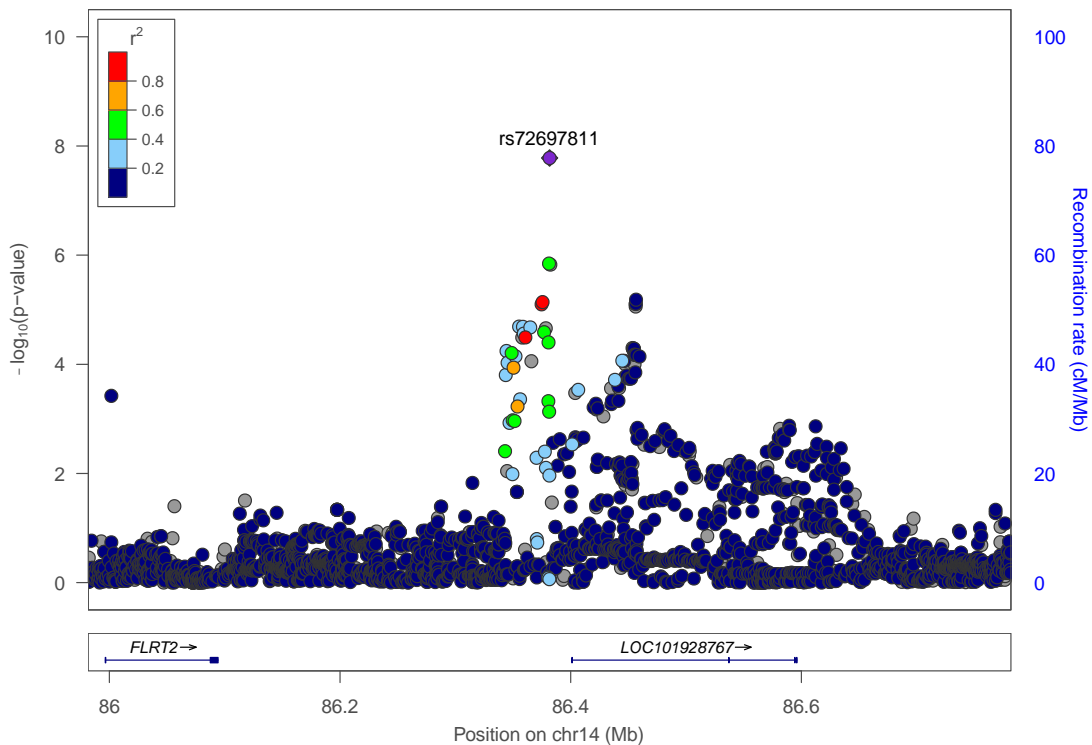
Figure 22 Regional Association Plot of rs72697811 from Tooth Rotation in POFC

One of the suggestive associations signals was detected near rs2015683 (beta = 0.04909655, P = 1.64E-06) which is upstream from *NRP1* gene (Figure 23). The encoded protein of *NRP1* plays a role in several signaling pathways that control cell migration and is essential for proper endothelial cells organization to form a vascular tube. It was found to be highly expressed in human teeth pulp with complete and incomplete root development (Gomez-Sosa, et al., 2019), which might indicate a possible role of *NPR1* in tooth development and could influence tooth rotation.
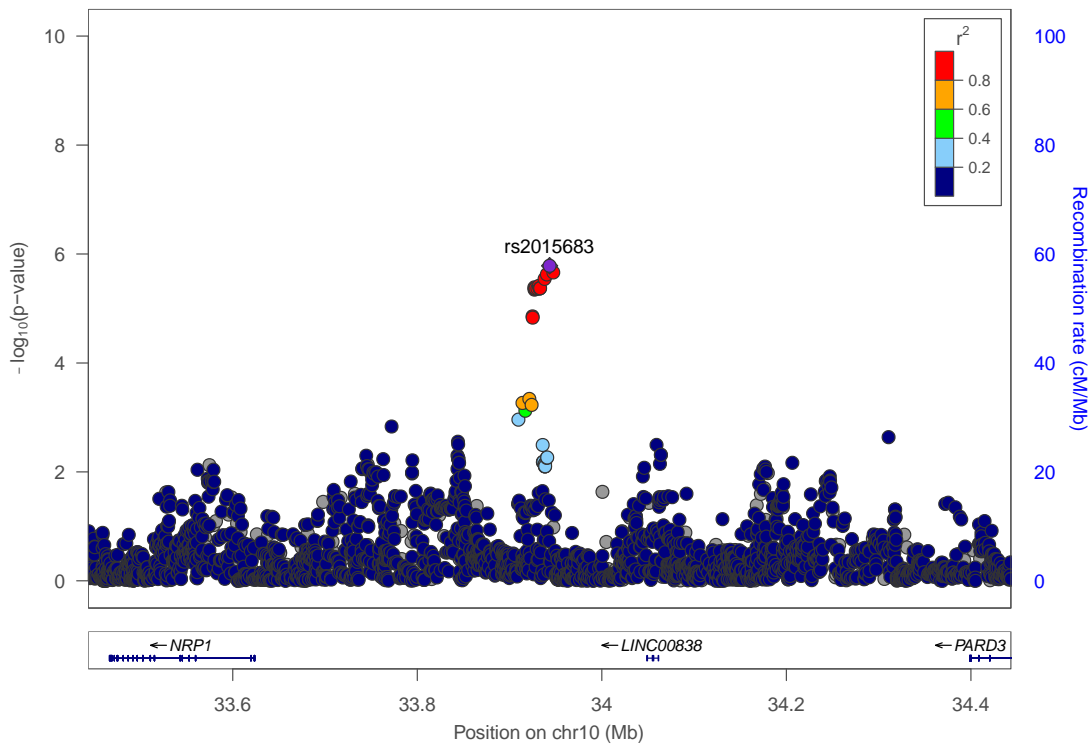
Figure 23 Regional Association Plot of rs2015683 from Tooth Rotation in POFC

4.2.2 HYPOPLASIA, DISPLACED, ROTATION (HDR)

Although no SNPs reached genome-wide significance ($P < 5 \times 10^{-8}$); there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Figure 24. A visual representation of the association results of HDR is displayed in Figure 24. The genomic inflation factor (lambda) was=0.87, which is indicative of deflation.

Variant rs12379966 ($P = 1.00$E-06) is downstream to the zinc finger protein *GLIS3* (Figure 25). The encoded protein of *GLIS3* regulates and improves osteoblast differentiation by acting interdependently with BMP2 and Shh. In addition, *GLIS3* promotes an increase in *FGF18* expression during osteoblast differentiation (Beak, et al., 2007). *GLIS3's* role in tooth development has not been studied yet, and it is unknown if this gene may affect or have a role in dental and oral health.

Variant rs6479408 is one of our suggestive variants ($P = 1.00$E-06). It is upstream from *ROR2* (Figure 26), an orphan tyrosine kinase that mediates Wnt5a-initiated noncanonical signaling and Wnt5a- inhibition of the Wnt canonical signaling, which is required during the growth, patterning, and differentiation of teeth. *Ror2* expression has been observed in mice developing teeth and a tooth development retardation was also observed in *Ror2* mutant mice (Lin, et al., 2011). This variant is also located downstream to *OMD* gene, Osteomodulin, which is a member of the small leucine-rich proteoglycan family distributed in the extracellular matrix (ECM). It has been found that *OMD* is expressed in polarized odontoblasts and alveolar bone during early crown formation and plays an essential role in modulating the osteo/odontoblastic differentiation of human dental pulp steam cells (Lin, et al., 2019). These studies suggest a plausible role of

ROR2 and *OMD* in tooth development and they might be associated with these correlated dental anomalies (HDR).

One of our suggestive variants was rs141429354 (*P* = 6.00E-06) is upstream of *WDR72* (Figure 27), a protein coding gene that plays an important role in the mineralization of tooth enamel. It is essential during the maturation phase of amelogenesis for normal formation of the enamel (El-Sayed, et al., 2009). Wdr72 expression was detected in the enamel organ of mouse incisors, and knockout mice (Wdr72 -/-) appeared to have defect in enamel maturation (Katsura, et al., 2014). It is been reported that mutations in this gene is associated with amelogenesis imperfecta hypomaturation (El-Sayed, et al., 2009). Although it is unclear the role *WDR72* that have in tooth development and/or oral health, there is supporting biological evidence that it may influence those correlated dental anomalies (HDR), specifically enamel hypoplasia.

Variant rs404727 (*P* = 8.00E-06), is located near to *BMP7* (Figure 28), which encodes a secreted ligand of the TGF-beta (transforming growth factor-beta) a superfamily of proteins that play a role in ectopic bone formation and odontogenesis. In a study of *BMP7* expression in mice they found out that *BMP7* is highly expressed during tooth development and knock out mice had morphological and functional changes in their teeth (Zurowski, et al., 2018). A previous study that investigated the role and the expression of *BMP7* in dental mesenchyme in human found out that *BMP7* was expressed in the late bell-stage human dental papilla and they might have a role in inducing the odontogenic differentiation of human dental pulp stem cells (Gao, et al., 2015). Although there is no evidence of a relationship between BMP7 and HDR, there are enough evidence that BMP7 might influence those dental anomalies since it plays a role during tooth development.

Variant rs1390193 (*P* = 1.50E-05) is located near to *WNT5A* (Figure 29), this gene belongs to WNT signaling family, and this particular gene signals through both the canonical and non-

canonical WNT pathways and plays an important role in regulating developmental pathways during embryogenesis. Previous study indicated that *WNT5A* expression was noticeable in early and later stages of tooth development in mice. In addition, they noticed that *WNT5A* mutant mice appeared to have smaller, and mis-patterned teeth, and their odontoblast differentiation process was delayed. Their results strongly suggest that Wnt5a regulates growth, patterning, and odontoblast differentiation during odontogenesis (Lin, et al., 2011) and that could possibly indicate that *WNT5A* could also influence any of those correlated dental anomalies in (HDR) pattern.

Variant rs7788761 ($P = 3.20E-05$) is located near to *DLX5* and *DLX6* (Figure 30), this gene belongs to Dlx family of homeobox genes, which has been shown to play critical roles in many embryonic developmental processes, including bone development. *Dlx5* has shown to be expressed, in mice and human cells, in developing and differentiating cartilage, bone, and teeth, which implicate the role for *Dlx5* in regulating the formation of these hard tissues (Ferrari et al., 1995, Ryoo et al., 1997, Newberry et al., 1998). Previous studies that investigated the role of *DLX5* in bone and tooth formation found that the lack of *DLX* protein caused dysmorphogenesis in teeth and almost all of the cranial bones (Acampora et al., 1999, Depew et al., 1999, Zhang 2003). These evidences suggest a plausible role of *DLXs* in tooth development and they might be also associated with these correlated dental anomalies (HDR).

Table 21 GWAS hits for the Correlated Dental anomalies (HDR) in POFC

SNP	CHR	BP	F*	Weights*	*P*	Effect A.	MAF	Type
rs6892150	5	162187646	8.876	-0.381, -0.452,0.7548	1.00E-06	A	0.1893	Imputed
rs12379966	9	3650939	9.534	0.135,0.825, -0.455	1.00E-06	G	0.1802	Genotyped
rs6479408	9	94903215	11.37	-0.01,0.792, -0.505	1.00E-06	G	0.4517	Imputed
rs34903076	8	69306322	10.05	-0.342, -0.796,0.418	3.00E-06	A	0.4181	Imputed
rs17088403	13	72264583	6.984	-0.689,0.696,0.296	3.00E-06	A	0.0759	Genotyped
rs141429354	15	54478011	8.714	-0.335,0.921, -0.0870	3.00E-06	T	0.05562	Imputed
rs12502364	4	86179966	9.654	0.451,0.488,0.785	4.00E-06	G	0.3469	Imputed
rs2518311	6	101993937	9.781	0.703,0.703,0.167	4.00E-06	A	0.4073	Imputed
rs17598358	4	109891009	8.567	-0.006,0.981, -0.074	5.00E-06	C	0.2128	Imputed
rs78669149	5	84401577	8.312	0.142, -0.786,0.510	5.00E-06	C	0.08562	Imputed
rs9502420	6	6142985	12.07	0.085, -0.438,0.840	7.00E-06	A	0.07251	Imputed
rs9651934	12	89355709	11.87	0.440, -0.313,0.788	7.00E-06	A	0.09472	Imputed
rs404727	20	55465434	10.12	0.407, -0.623,0.577	8.00E-06	A	0.251	Imputed
rs201590135	16	21748024	8.198	-0.435, -0.347,0.801	1.00E-05	C	0.1931	Imputed
rs76013619	1	233662732	8.901	-0.368, -0.469,0.752	1.10E-05	A	0.2855	Imputed
rs113402395	9	140135219	7.393	0.151,0.986,0.164	1.20E-05	A	0.09724	Imputed
rs1322468	9	7949546	8.862	-0.190, -0.621,0.683	1.30E-05	C	0.1661	Imputed
rs16998644	20	17050262	7.58	-0.022,0.894, -0.343	1.40E-05	A	0.2592	Genotyped
rs1390193	3	55741010	9.433	0.0470,0.442,0.938	1.50E-05	A	0.2433	Imputed
rs7788761	7	96372390	7.753	-0.799,0.166,0.610	3.20E-05	G	0.1045	Imputed

* F-Statistic tests the significance of the canonical correlations
* Weights reflect the correlation coefficients for each individual trait
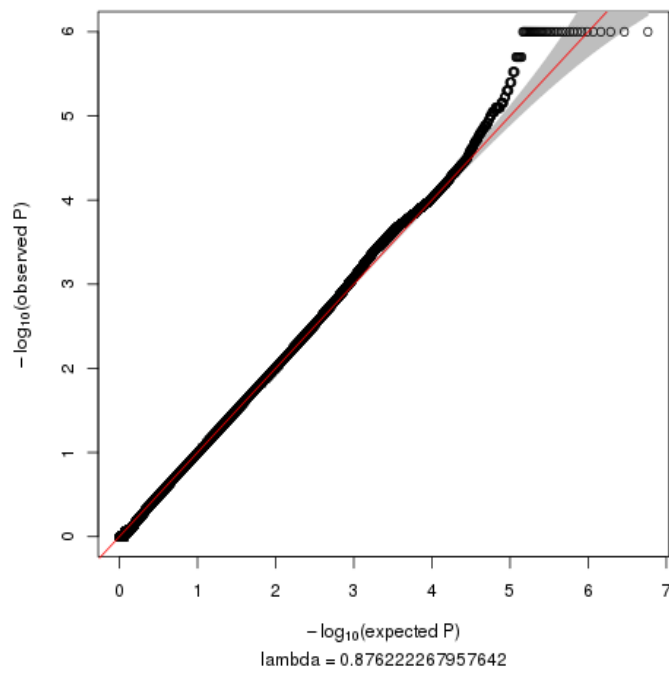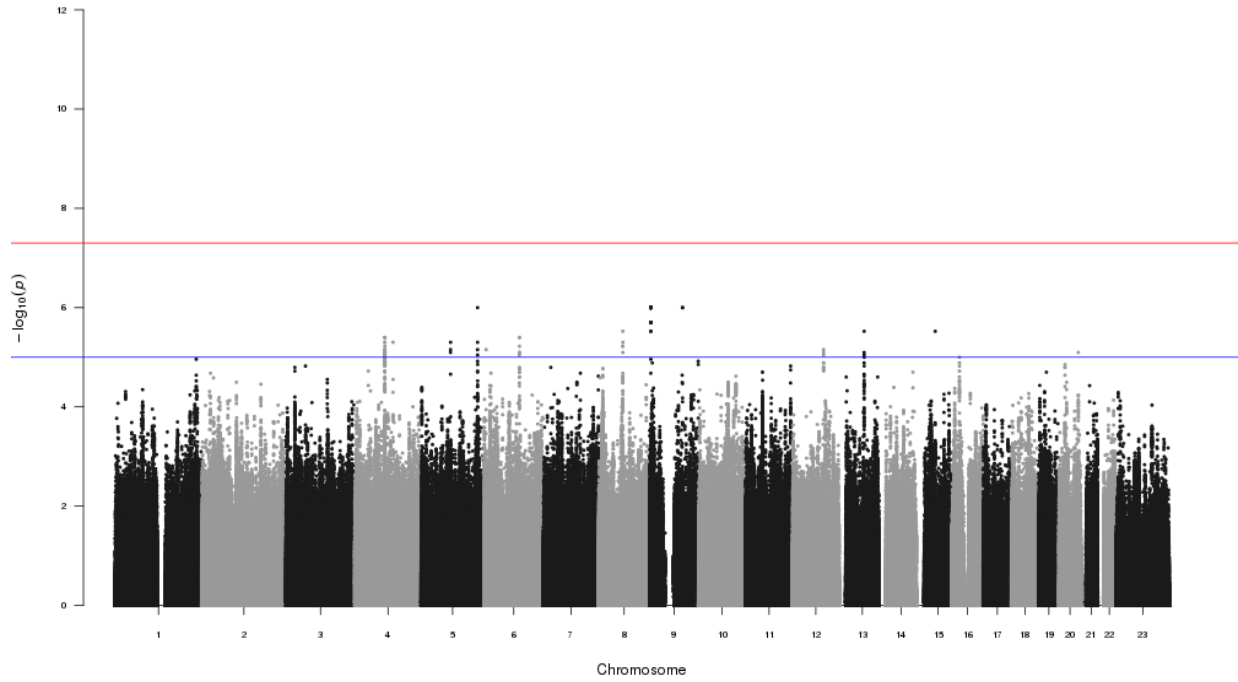
Figure 24 Manhattan and QQ plots for the Correlated Dental anomalies (HDR) in POFC
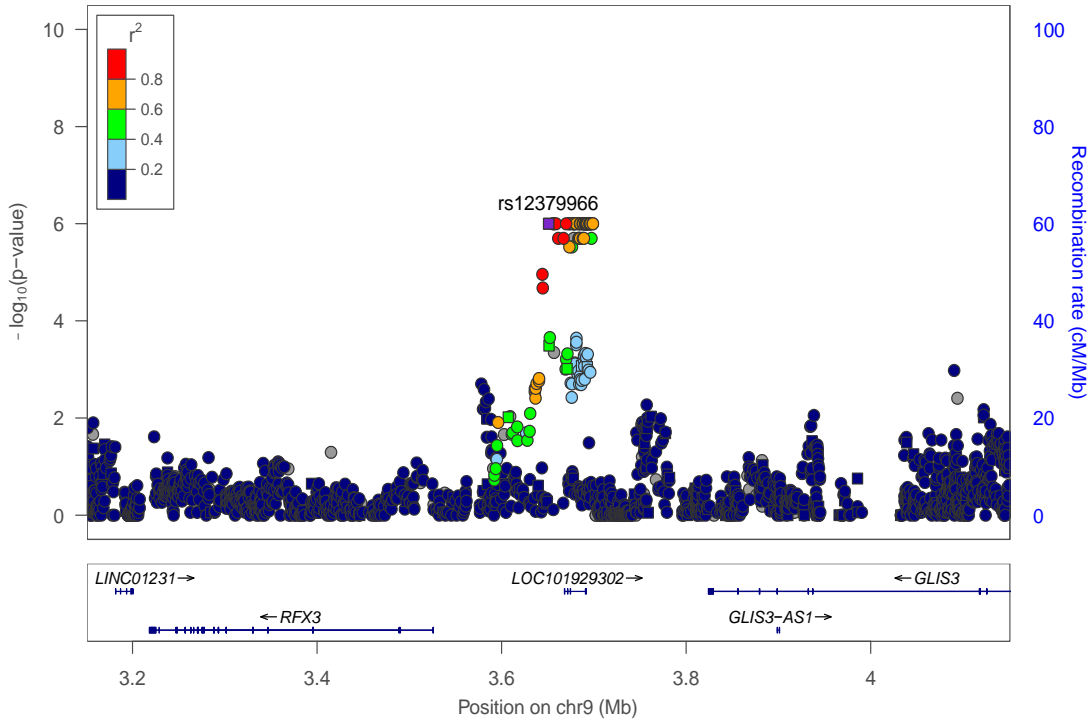
Figure 25 Regional Association Plot of rs12379966 from (HDR) results in POFC
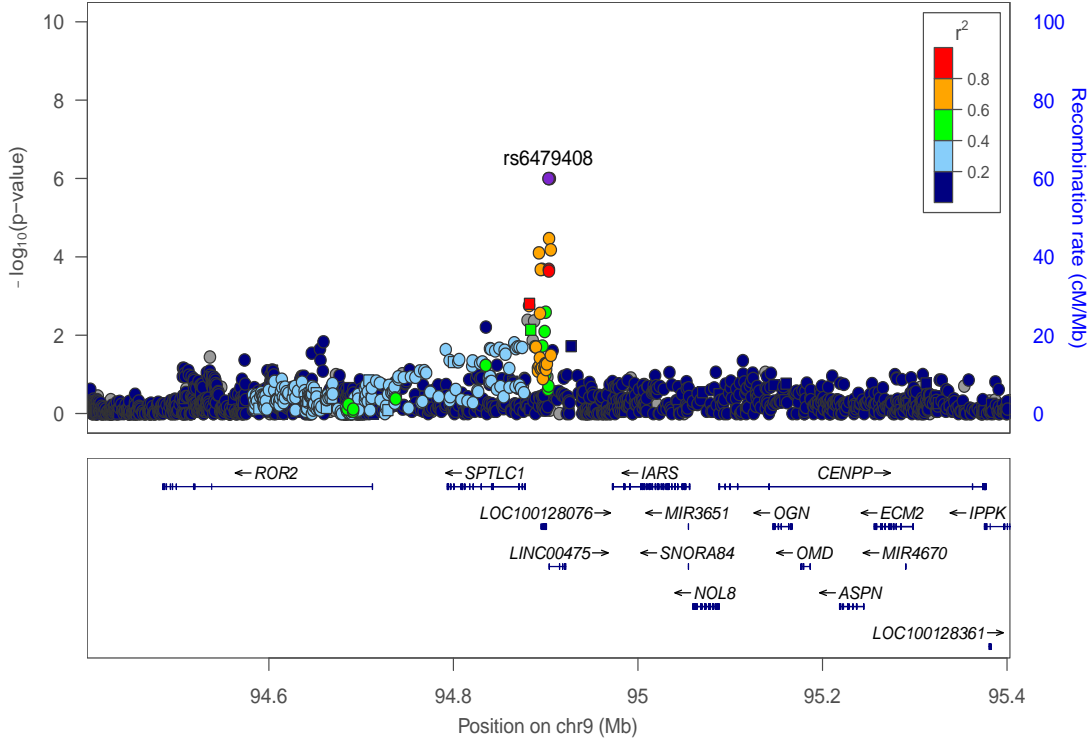
Figure 26 Regional Association Plot of rs6479408 from (HDR) results in POFC
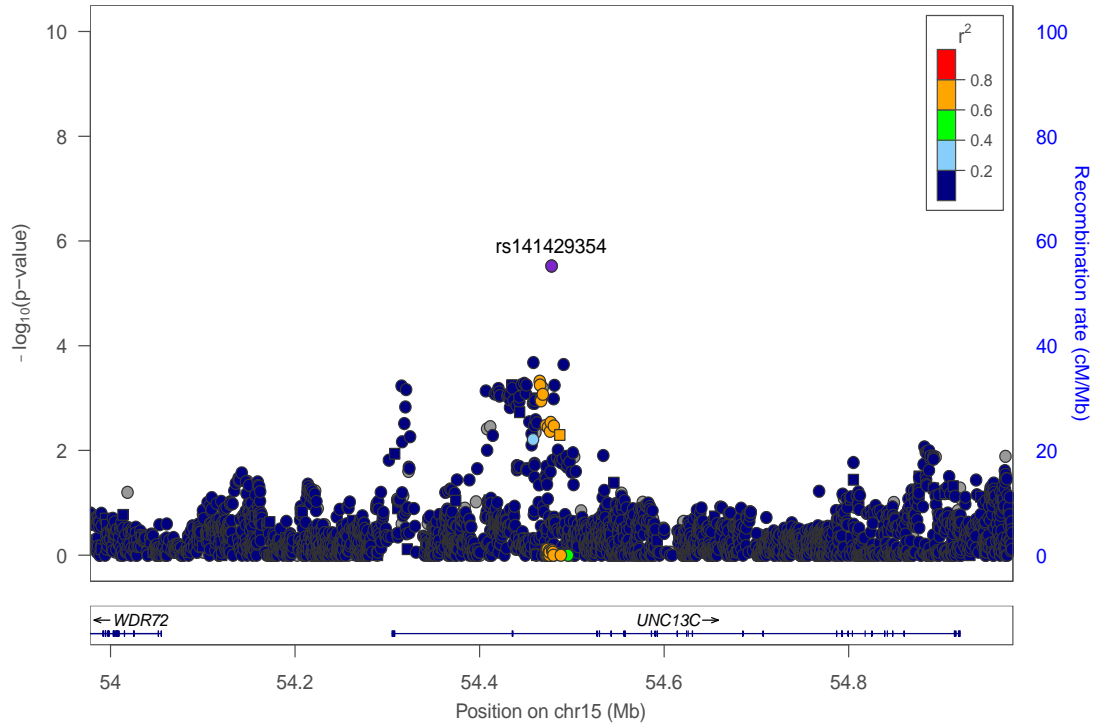
Figure 27 Regional Association Plot of rs141429354 from (HDR) results in POFC
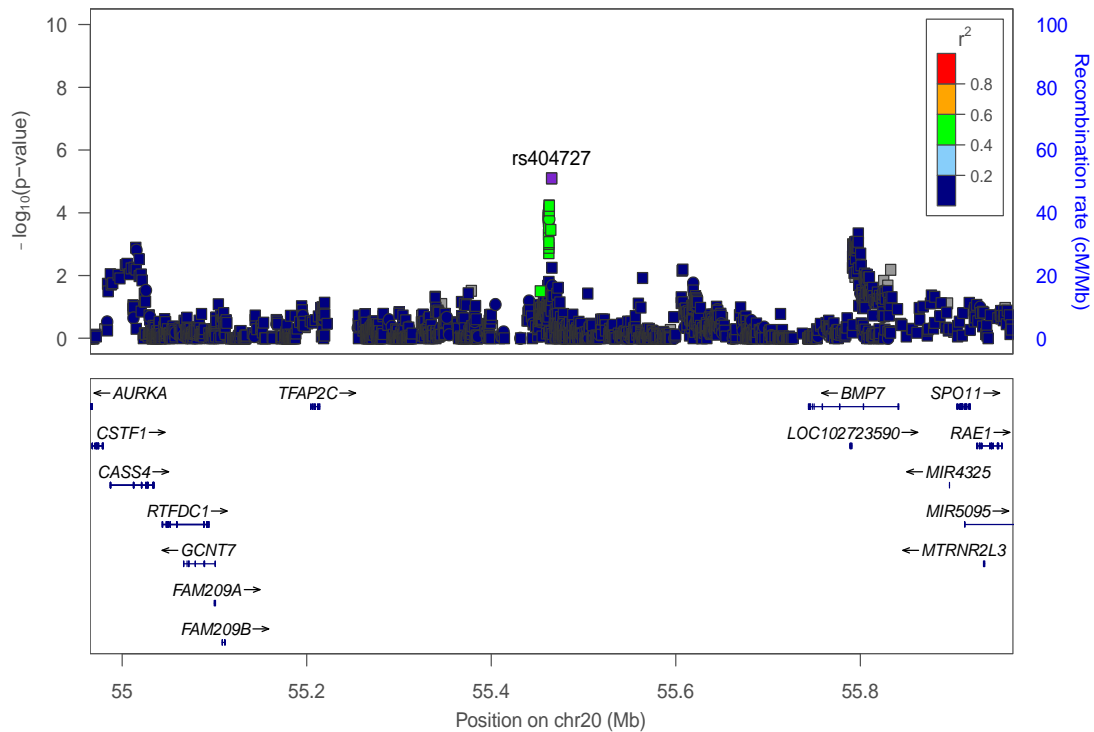
Figure 28 Regional Association Plot of rs404727 from (HDR) results in POFC
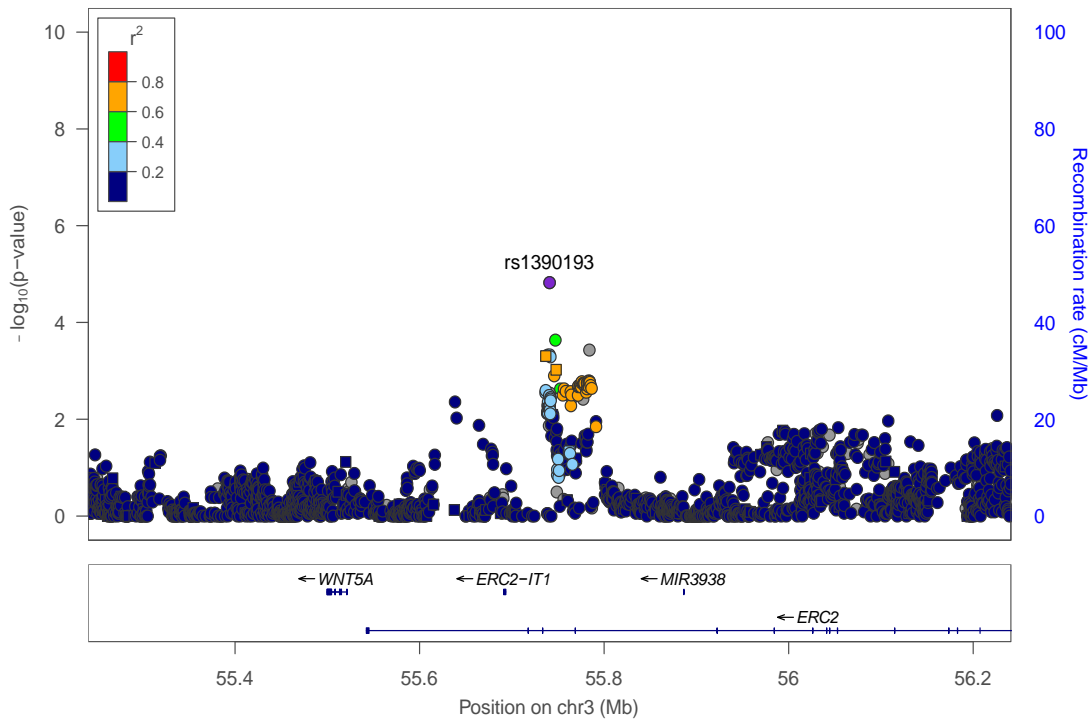
Figure 29 Regional Association Plot of rs1390193 from (HDR) results in POFC
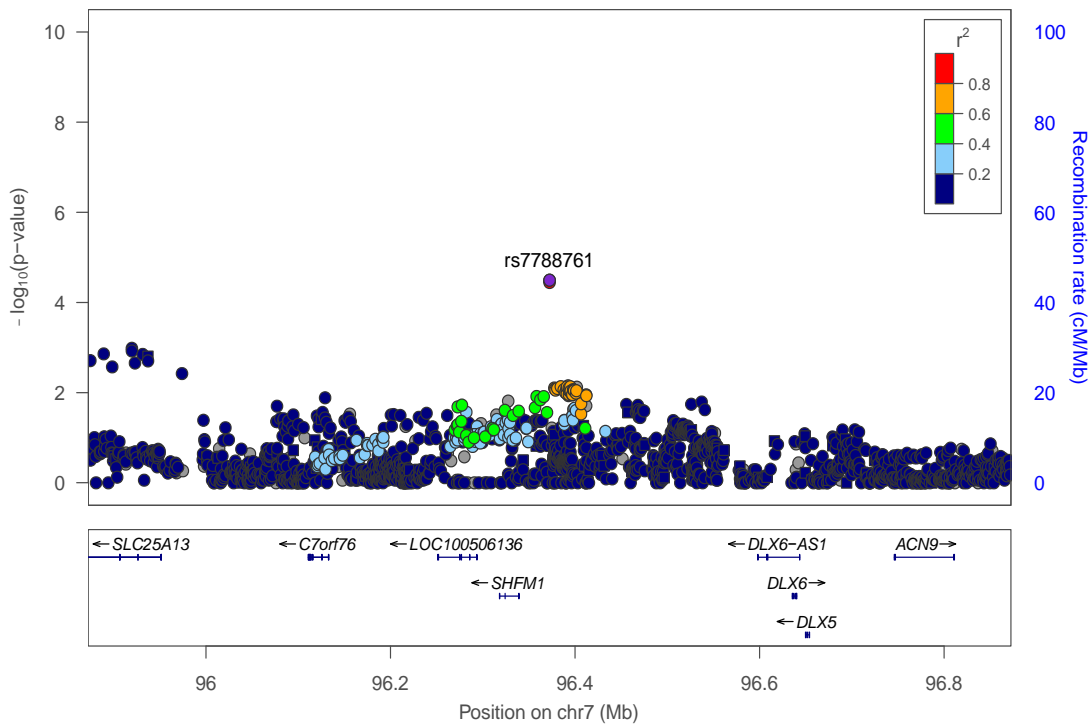
Figure 30 Regional Association Plot of rs7788761 from (HDR) results in POFC

4.2.2.1 HYPOPLASIA GWAS RESULTS

The GWAS of hypoplasia was performed on cohort (N=3579) for whom genotypes data, phenotypes, and covariates data were available. A total of 7,168,909 SNPs was available for analysis after applying quality control criteria. In the GWAS of hypoplasia, we were able to identify SNPs that had almost reached the genome-wide significance ($P < 5 \times 10^{-8}$). In addition, there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Table 22. A visual representation of the association results of hypoplasia is displayed in Figure 31. The genomic inflation factor (lambda) was =0.99, which indicate no inflation.

Table 22 Top GWAS hits for Hypoplasia in POFC

SNP	CHR	BP	*P*	BETA	MAF	A1	Type
rs60248638	3	41883692	5.87E-08	-0.0600436	0.1468	A	Imputed
rs12043922	1	154118497	7.45E-08	-0.0741821	0.06949	A	Imputed
rs2414459	15	56365357	9.58E-08	-0.0610045	0.1088	T	Imputed
rs9616163	22	47328584	3.22E-07	-0.0683888	0.06117	G	Imputed
rs55726235	13	34022292	5.12E-07	-0.0767129	0.06841	A	Imputed
rs34201461	9	34986416	5.14E-07	-0.0502732	0.1686	C	Imputed
rs73587996	6	125811869	5.86E-07	-0.0901956	0.0522	T	Imputed
rs71270085	8	84171075	8.09E-07	-0.0500024	0.1441	C	Imputed
rs9676005	18	2860891	9.89E-07	-0.0830724	0.07639	A	Imputed
rs146585982	10	85208452	1.30E-06	-0.0321393	0.3807	AC	Imputed

One of the top hits in the GWAS of hypoplasia was rs60248638 (beta = -0.0600436, P = 5.87E-08), which is located near to *CCK* gene (Figure 32). *CCK* encode member of the gastrin/cholecystokinin family of proteins and is best known as a hormone and neuropeptide associated with several functions including the release of digestive enzymes from the pancreas and serving as an autocrine growth factor (Nurbaeva, et al., 2018). In a previous genome-wide screening that compared enamel organ (EO) cells from secretory and maturation stage, *CCK* was identified as being highly up-regulated in enamel maturation (Lacruz, et al., 2012). Based on these studies, we could nominate *CCK* as a possible gene that could be associated with enamel hypoplasia.

Variant rs12043922 (beta = -0.0741821, P = 7.45E-08) is located downstream to *ADAR* gene (Figure 33), which makes RNA-specific adenosine deaminase 1 (*ADAR1*) protein. *Adar1* expression was very strong in mice dental papilla, ameloblasts, and odontoblasts during tooth development, and it was found that novel mutation in *ADAR1* caused dental anomalies, more specifically dens evaginatus and dens invaginatus (Kantaputra, et al., 2012). There is supporting biological evidence that *ADAR1* could be associated with enamel hypoplasia.

Another top association signal in the GWAS of hypoplasia was detected near rs2414459 (beta = -0.0610045, P = 9.58E-08), which is located near to *PYGO1* gene (Figure 34). It was found that mice lacking both Pygo1 and Pygo2 in epithelial cells developed teeth, however the tooth enamel was structurally disorganized, bright white and exhibited reduced iron content compared to control mice, and these phenotypes resemble to those of humans with amelogenesis imperfecta (AI) (Cantù, et al., 2017). We can see a possible connection between enamel hypoplasia and *PYOG1* gene, however additional studies are needed to confirm the role that *PYGO1* gene could play in the enamel formation during tooth development using human and mouse models.
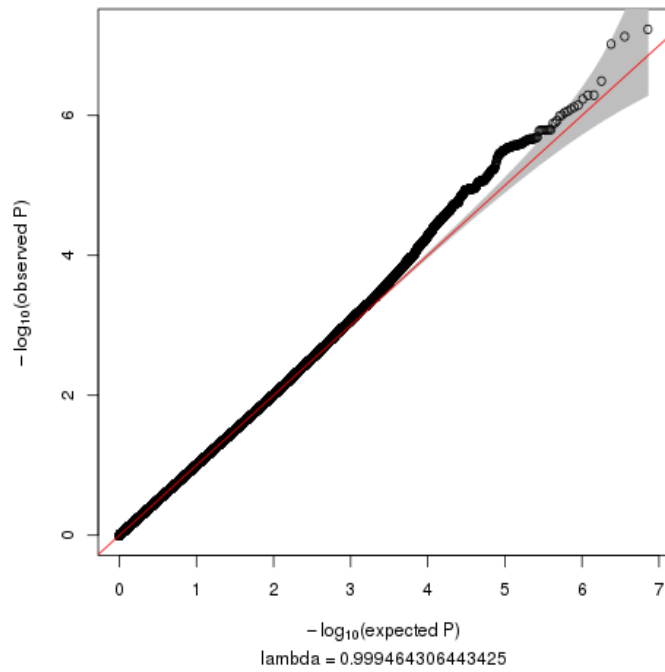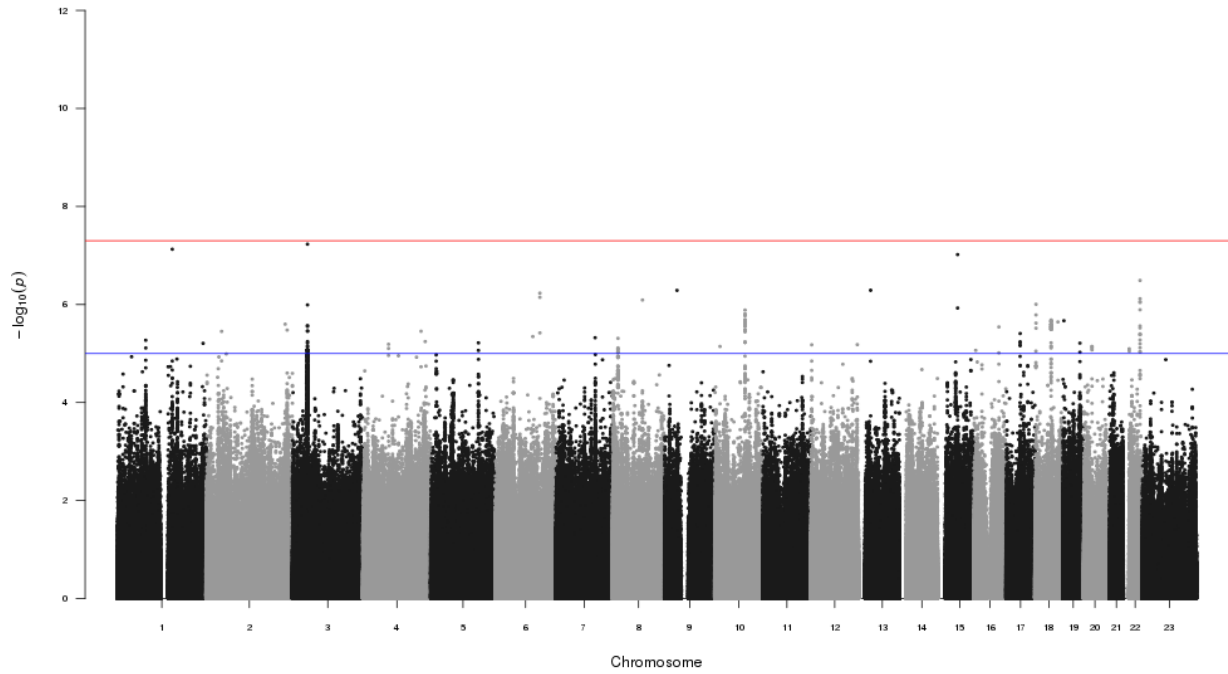
Figure 31 Manhattan and QQ plots for Hypoplasia GWAS in POFC
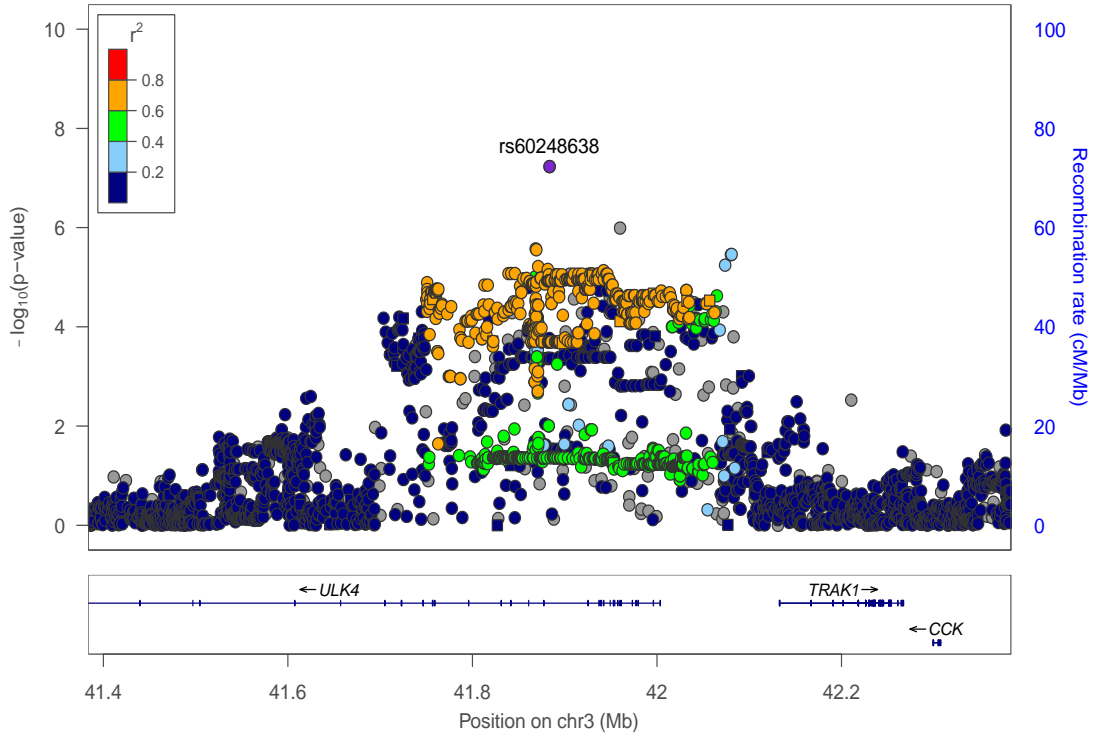
Figure 32 Regional Association Plot of rs60248638 from Hypoplasia GWAS in POFC
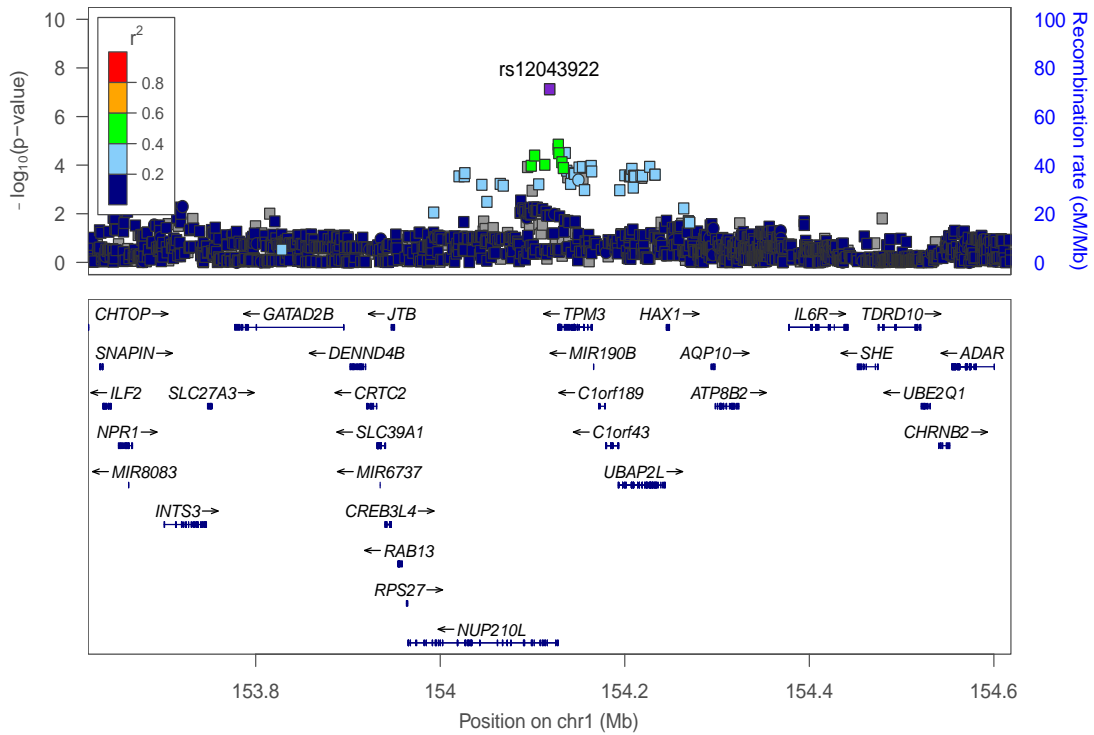
Figure 33 Regional Association Plot of rs12043922 from Hypoplasia GWAS in POFC

97
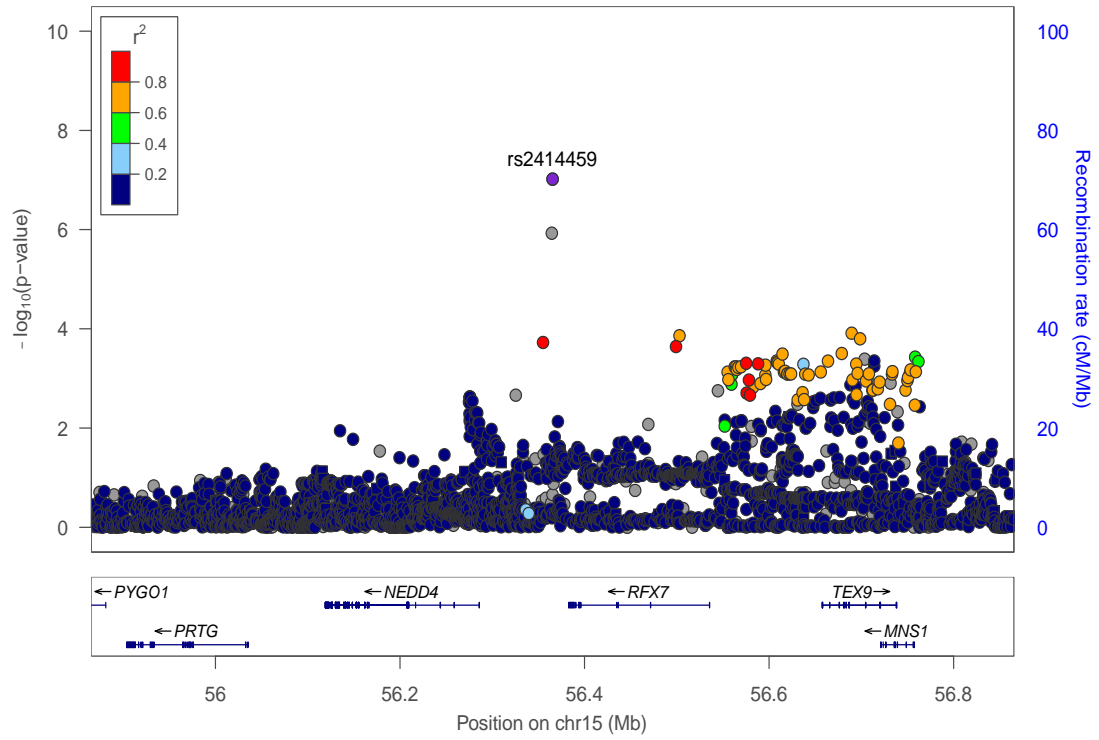
Figure 34 Regional Association Plot of rs2414459 from Hypoplasia GWAS in POFC

4.2.2.2 DISPLACEMENT GWAS RESULTS

The GWAS of tooth displacement was performed on cohort (N=3579) for whom genotypes data, phenotypes, and covariates data were available. A total of 7,168,909 SNPs was available for analysis after applying quality control criteria. In the GWAS of tooth displacement, we were able to identify SNPs that had almost reached the genome-wide significance ($P < 5 \times 10^{-8}$). In addition, there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Table 23. A visual representation of the association results of tooth displacement is displayed in Figure 35. The genomic inflation factor (lambda) was =0.99, which indicate no inflation.

Table 23 Top GWAS hits for of Tooth Displacement in POFC

SNP	CHR	BP	*P*	Type	BETA	MAF	A1
rs111809922	7	72309720	1.97E-07	Imputed	-0.1161277	0.06821	A
rs12631728	3	23085248	2.06E-07	Imputed	0.14731502	0.05079	T
rs9677256	2	135031863	1.11E-06	Imputed	-0.0763464	0.1776	G
rs17439917	9	131930777	1.65E-06	Imputed	0.09395418	0.07136	G
rs12102482	16	86231288	2.30E-06	Imputed	0.07390724	0.2048	A
rs34685308	19	21762945	2.96E-06	Imputed	0.09655418	0.06716	A
rs12700326	7	21955738	3.67E-06	Imputed	0.06742992	0.1495	A
rs143629334	21	28448615	4.03E-06	Imputed	-0.0800776	0.1262	C
rs6455824	6	162831336	4.11E-06	Imputed	0.08099162	0.1393	A
rs199970173	7	66518068	4.57E-06	Imputed	-0.0558996	0.4317	CTTT
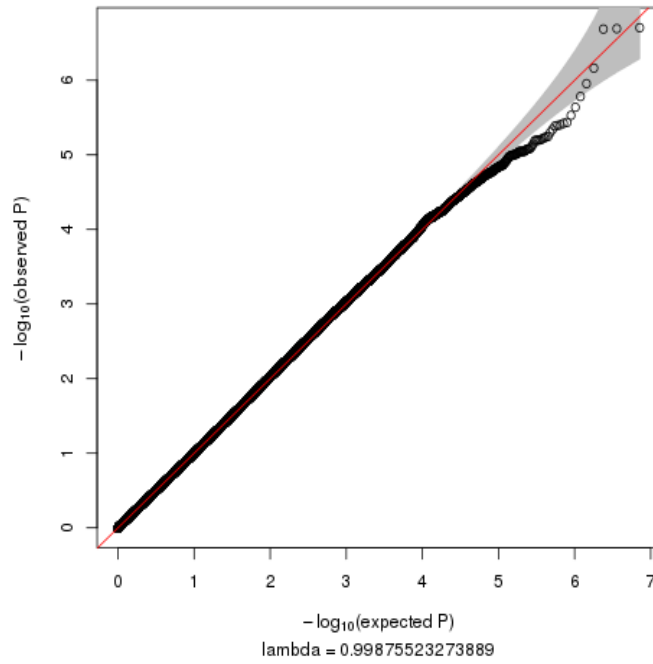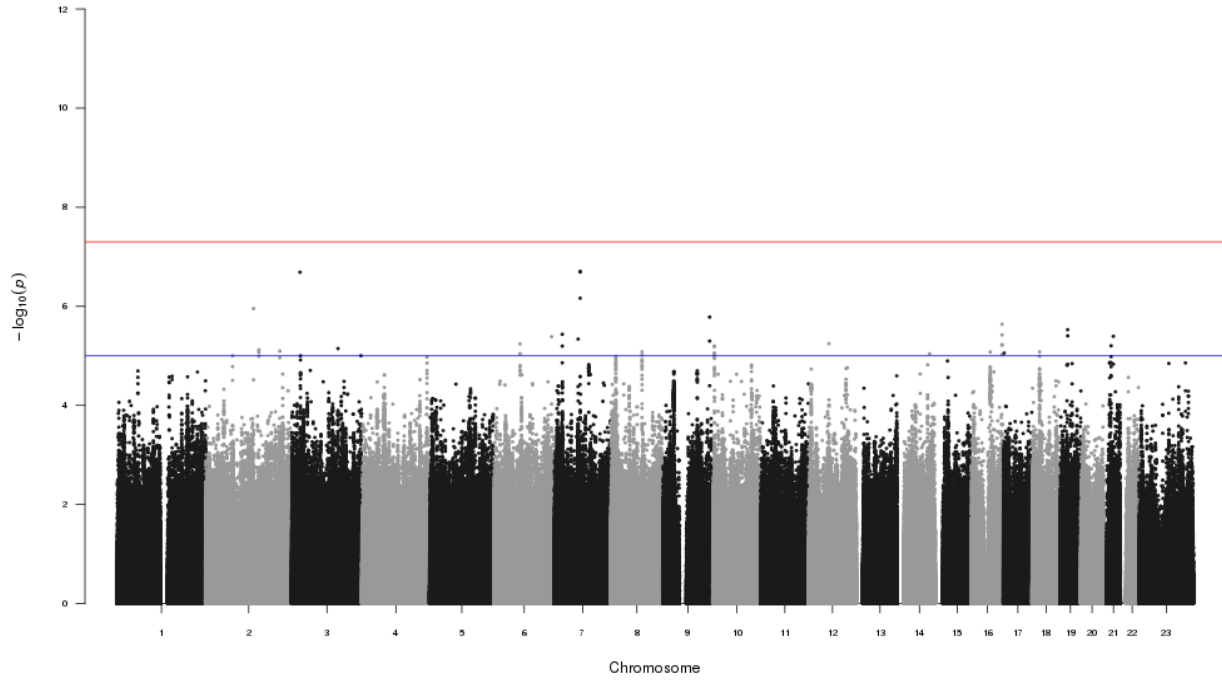
Figure 35 Manhattan and QQ plots for Displacment GWAS in POFC

One of the suggestive associations signal was detected at rs12102482 (beta = 0.07390724, P = 2.230E-06) is downstream from *IRF8*, which is a transcription factor of the interferon (IFN) regulatory factor (IRF) family (Figure 36). *IRF8* plays different roles, including a role as a negative regulator in cells of the immune system, and as a positive regulator of macroautophagy in dendritic cells (Agod, et al., 2018). *IRF8* has been found to be bound to regulatory regions of thousands of genes in osteoclast precursors, and it has also been found that a mutation in *IRF8* lead to increased osteoclast activity and could increase the risk of developing multiple idiopathic tooth root resorption, which is a form of periodontal disease (Thumbigere-Math, et al., 2019). Interestingly, rs12102482 also located upstream of *FOX* genes (*FOXF1, FOXC2, FOXL1*), and they are transcriptions factors with a highly conserved winged-helix/forkhead DNA-binding domain (Hong, et al., 1999). Cranial neural crest-derived cells (CNCs) expressions of foxc2 and foxf1 has been found to be positively regulated by Hedgehog signaling during murine craniofacial development, and this could indicate a possible role of Fox genes during facial skeletal development (Yamagishi, et al., 2003; Jeong et al., 2004). Previous studies investigated the expression of foxf1 in mice and found it to be highly expressed in developing tooth buds, and it has also been found that foxc2 and foxf1 mutant mice had developed cleft palate (Iida, et al., 1997, Wang, et al., 2003), while tooth loss was noticed in foxf1 mutant zebrafish (Xu, et al., 2018). Although it is still unclear the specific role of these genes during tooth development, it is plausible that FOX genes may have a role during tooth development, and they could also influence dental health and tooth displacement.

rs143629334 is another suggestive association variant (beta = -0.0800776, P = 4.030E-06) is located upstream of *ADAMTS* genes (Figure 37), which belong to a family of extracellular protease that function in cleaving proteoglycans, such as aggrecan and brevican. Adamst1-5 were

found to be expressed in mice dental pulp cells, odontoblasts, cementoblasts, cementocytes, periodontal ligament cells, osteoblasts and osteocytes (Sone, et al.,2005). From these findings, we can suggest that *ADAMTS* genes (*ADAMTS 1* and *ADMTS5* more specifically) could have a role in tooth development and could influence tooth displacement.
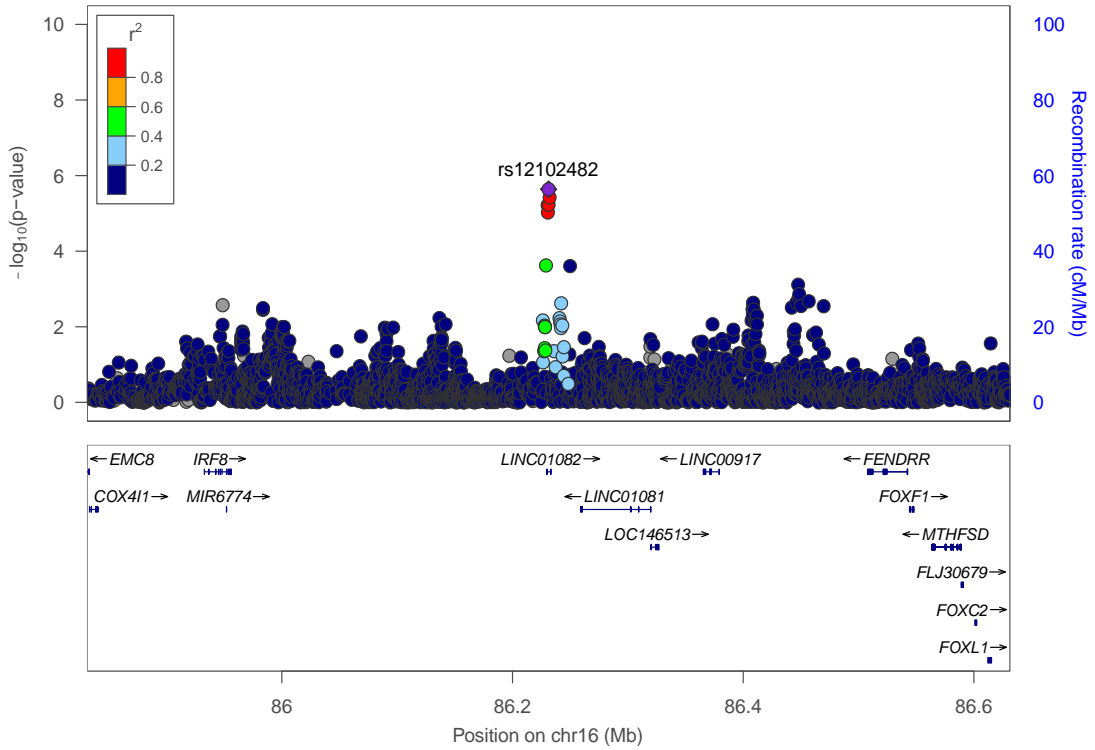
Figure 36 Regional Association Plot of rs12102482 from Displacment GWAS in POFC
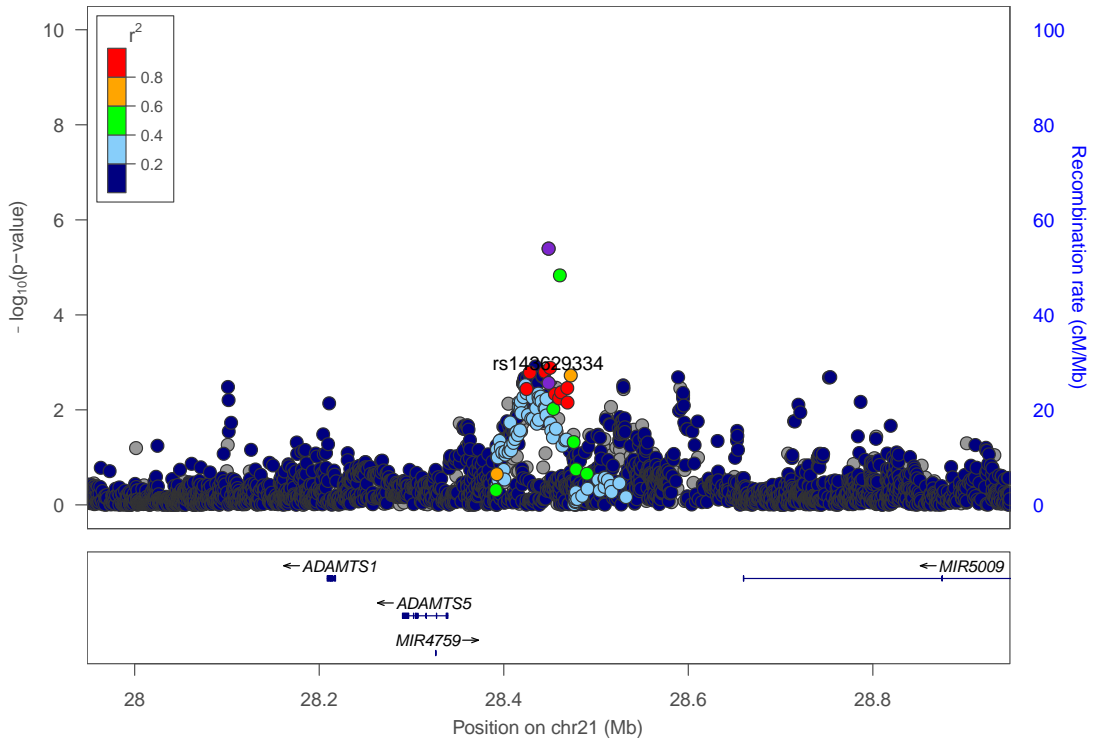
Figure 37 Regional Association Plot of rs143629334 from Displacment GWAS in POFC

4.2.3 DISPLACED, ROTATION, MAMELONS (DRM)

No SNPs reached genome-wide significance ($P < 5 \times 10^{-8}$) however, there were several variants that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Figure 38 visual representation of the association results of DRM is displayed in Figure 38. The genomic inflation factor (lambda) was= 0.82, which indicative of deflation.

One of our top suggestive variants was rs10511451 ($P = 1.00E-06$) is downstream of the zinc finger protein *GLIS3* (Figure 39). The protein product of *GLIS3* regulates and improves osteoblast differentiation by acting interdependently with BMP2 and Shh. In addition, *GLIS3* promote an increase in the FGF18 expression during osteoblast differentiation (Beak, et al., 2007). However, *GLIS3* has not been studied in tooth development content exclusively, and it is unknown if this protein may affect or have a role in dental and oral health. This association signal is overlapping with another association signal in the same loci (rs12379966; $P = 1.00E-06$), which was identified in the previously discussed pattern (HDR).

We also had a suggestive variant (rs141429354; $P = 3.00E-06$), upstream to *WDR72* (Figure 40), which is a protein coding gene that plays an essential role in the mineralization and maturation of tooth enamel (El-Sayed, et al., 2009). This association signal (rs141429354) is actually overlapping with another association signal (rs141429354; $P = 6.00E-06$) from the previously discussed correlated dental anomalies (HDR).

rs6479408 (Figure 41) is a suggestive variant ($P = 5.00E-06$), and this variant is also been identified in association with the previous associated dental anomalies model (HDR) ($P = 1.00E-06$).

Another suggestive variant is rs174814 (P = 9.00E-06), which is located upstream of *ROBO2*

(Figure 42). The encoded protein of *ROBO2* function as cell receptor for slit2, which have a role

olfactory bulb axon guidance during neuronal development, it also has a role in cell migration

(Nguyen-Ba-Charvet, et al., 2002). In a gene set enrichment analysis study, *ROBO2* was listed as

one of the genes that could "potentially" be associated with dental traits (Wang, et al., 2013). It

is unknown if *ROBO2* may influence those correlated dental anomalies (DRM) or have a role in

dental and oral health.

Table 24 Top GWAS hits for the Correlated Dental anomalies (DRM) in POFC

SNP	CHR	BP	F*	Weights*	*P*	Effect A.	MAF	Type
rs10511451	9	3672822	11.13	-0.787,0.522,0.058	1.00E-06	C	0.1631	Imputed
rs116341491	1	30337895	10.35	-0.573, -0.749, -0.502	2.00E-06	G	0.05458	Imputed
rs12022189	1	233651712	10.46	-0.371,0.643, -0.593	2.00E-06	A	0.2757	Imputed
rs12213712	6	148126753	10.5	0.401, -0.360,0.808	2.00E-06	C	0.1069	Imputed
rs6112681	20	19932932	8.84	-0.114, -0.512,0.820	2.00E-06	A	0.1129	Imputed
rs4851383	2	101560758	11.75	0.257, -0.265,0.909	3.00E-06	G	0.356	Imputed
rs17039589	4	109886367	9.146	-0.674,0.466,0.496	3.00E-06	A	0.1135	Imputed
rs141429354	15	54478011	8.516	0.941, -0.089,0.290	3.00E-06	T	0.05562	Imputed
rs676793	3	118542408	10.55	0.646,0.509, -0.567	5.00E-06	A	0.2978	Imputed
rs6479408	9	94903215	11.26	0.790, -0.508,0.029	5.00E-06	G	0.4517	Imputed
rs56073015	16	20679391	9.908	0.046, -0.19,0.966	5.00E-06	G	0.2071	Imputed
rs9502420	6	6142985	12.74	0.448, -0.813, -0.253	7.00E-06	A	0.07251	Imputed
rs71524668	7	85980761	9.201	-0.128, -0.123,0.973	8.00E-06	A	0.3509	Imputed
rs75487692	9	26651466	8.691	-0.125, -0.123,0.972	8.00E-06	A	0.1553	Imputed
rs4277017	10	133331750	10.88	-0.187,0.755, -0.554	8.00E-06	A	0.2488	Imputed
rs174814	3	76761788	9.488	0.196,0.769,0.680	9.00E-06	C	0.4874	Imputed

* F-Statistic tests the significance of the canonical correlations

* Weights reflect the correlation coefficients for each individual trait
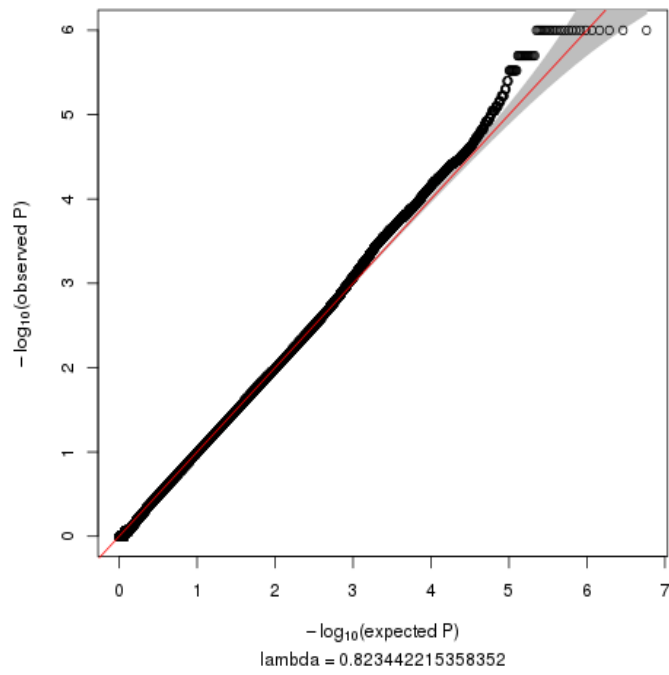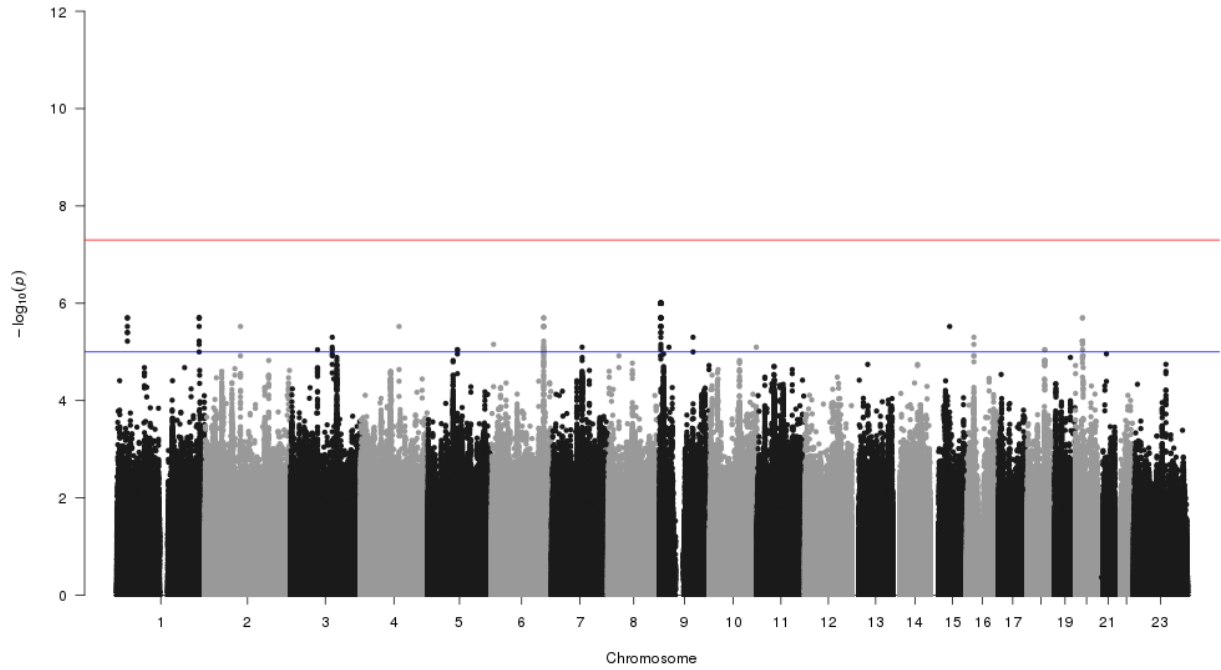
Figure 38 Manhattan and QQ plots for the Correlated Dental anomalies (DRM) in POFC

lambda = 0.823442215358352
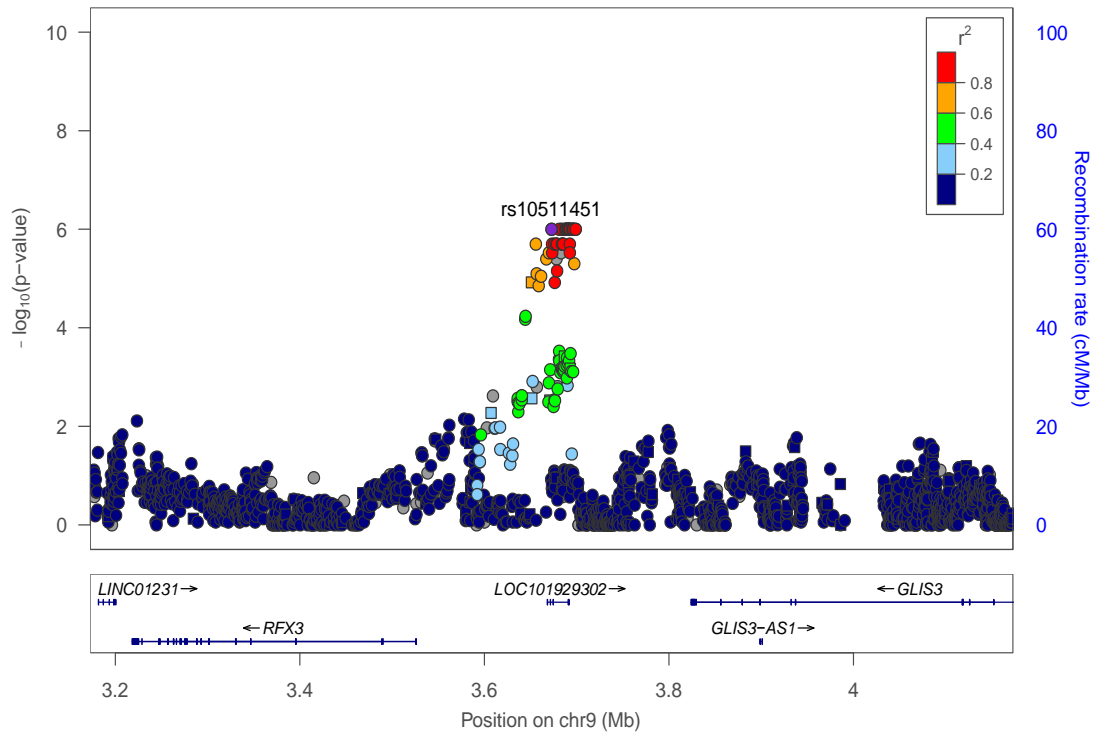
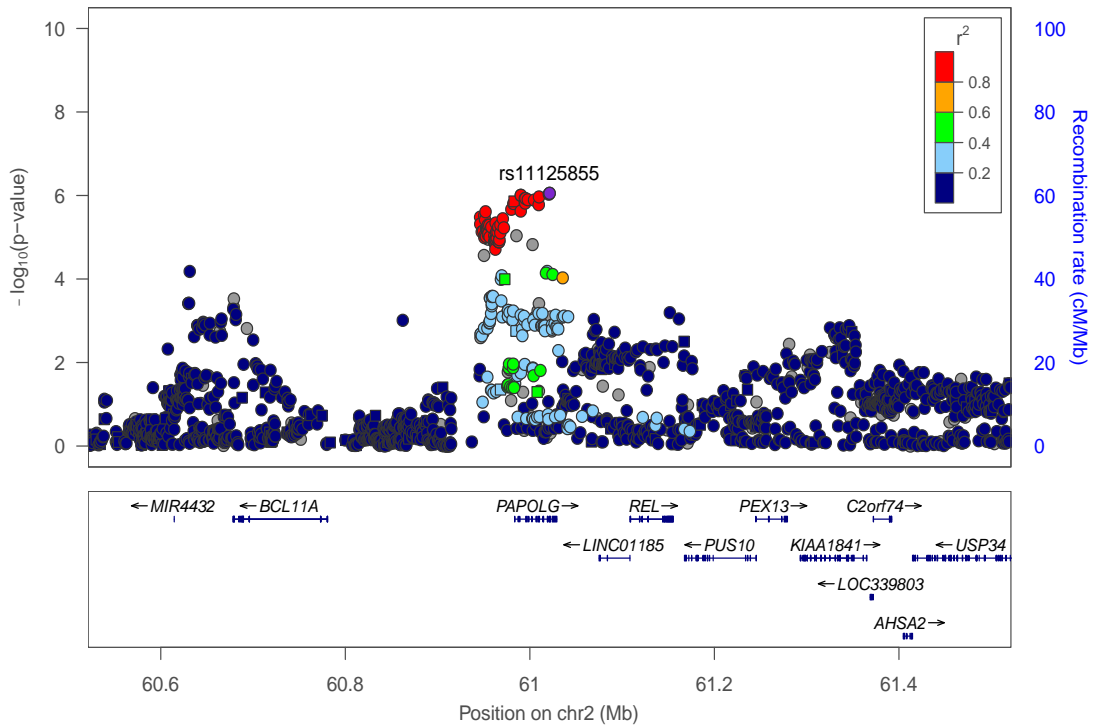**Figure 39 Regional Association Plot of rs10511451 from (DRM) results in POFC**
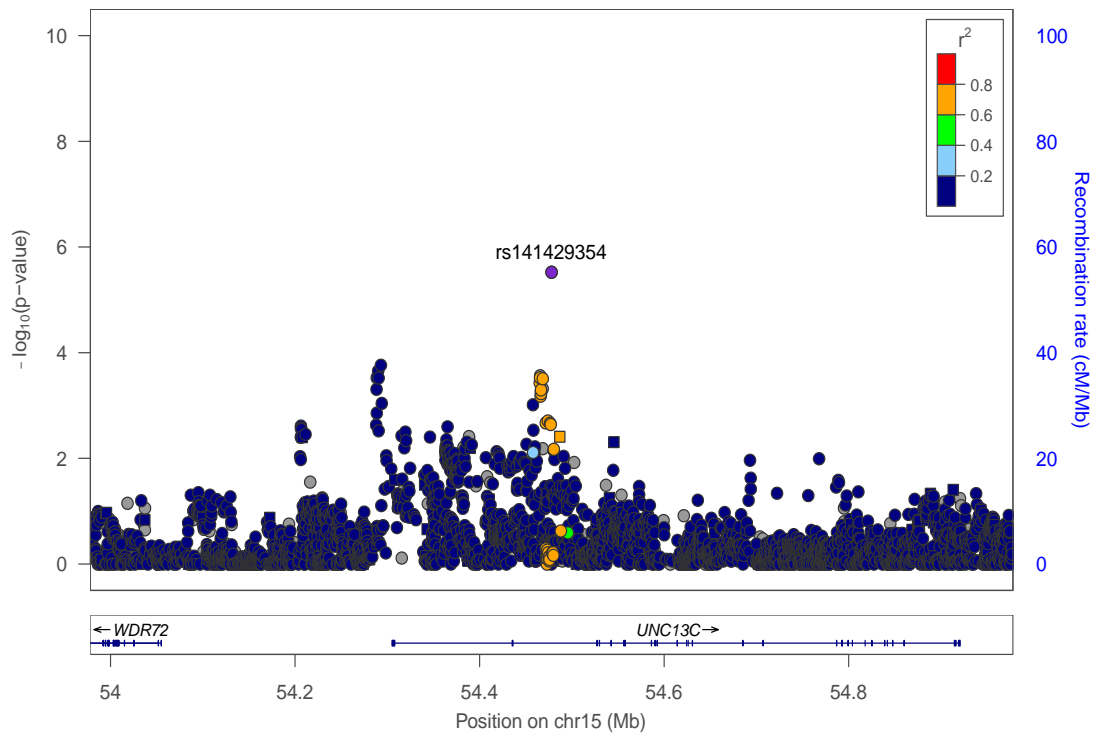
Figure 40 Regional Association Plot of rs1414293354 from (DRM) results in POFC
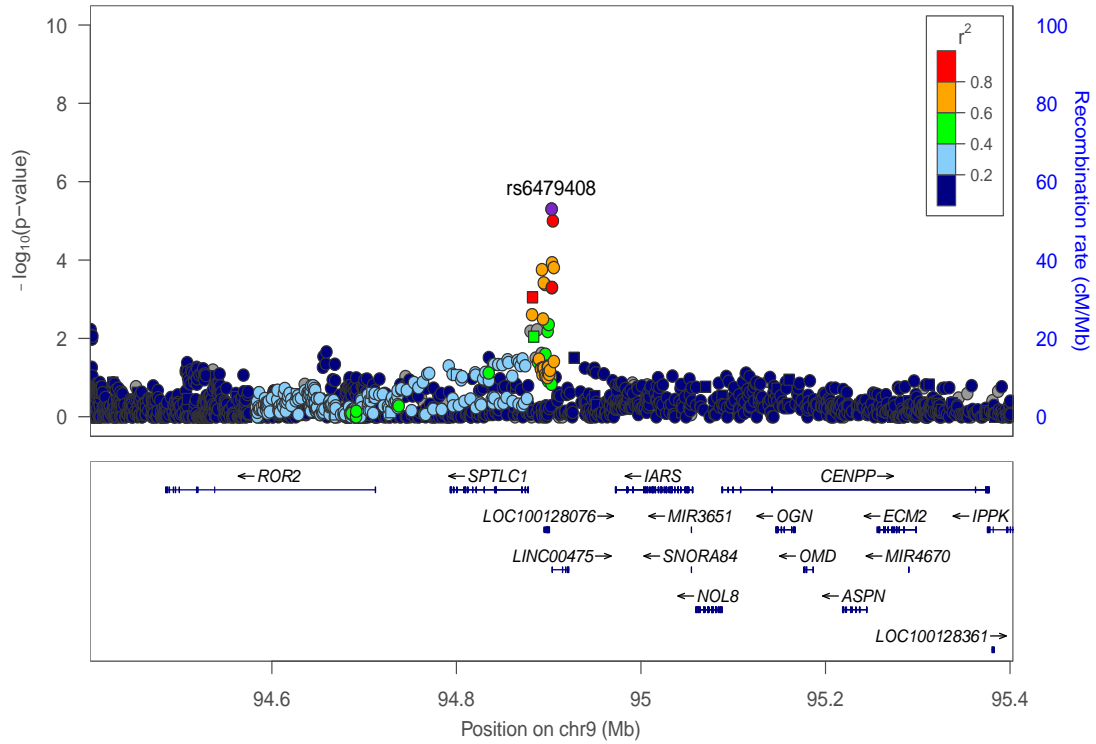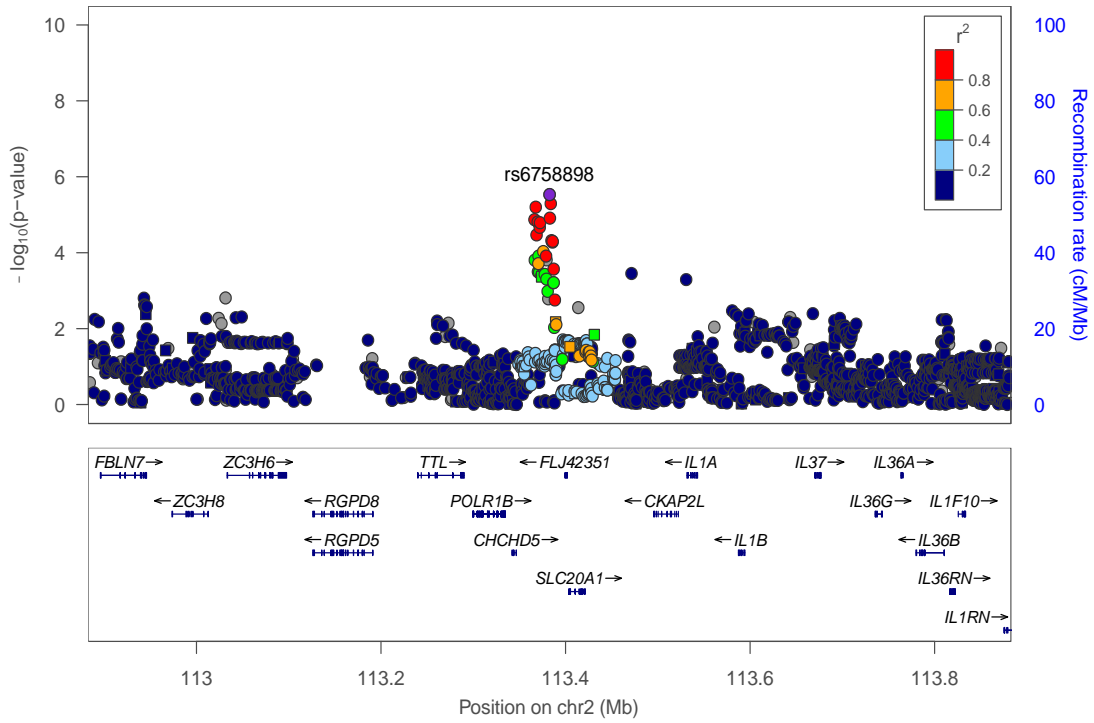
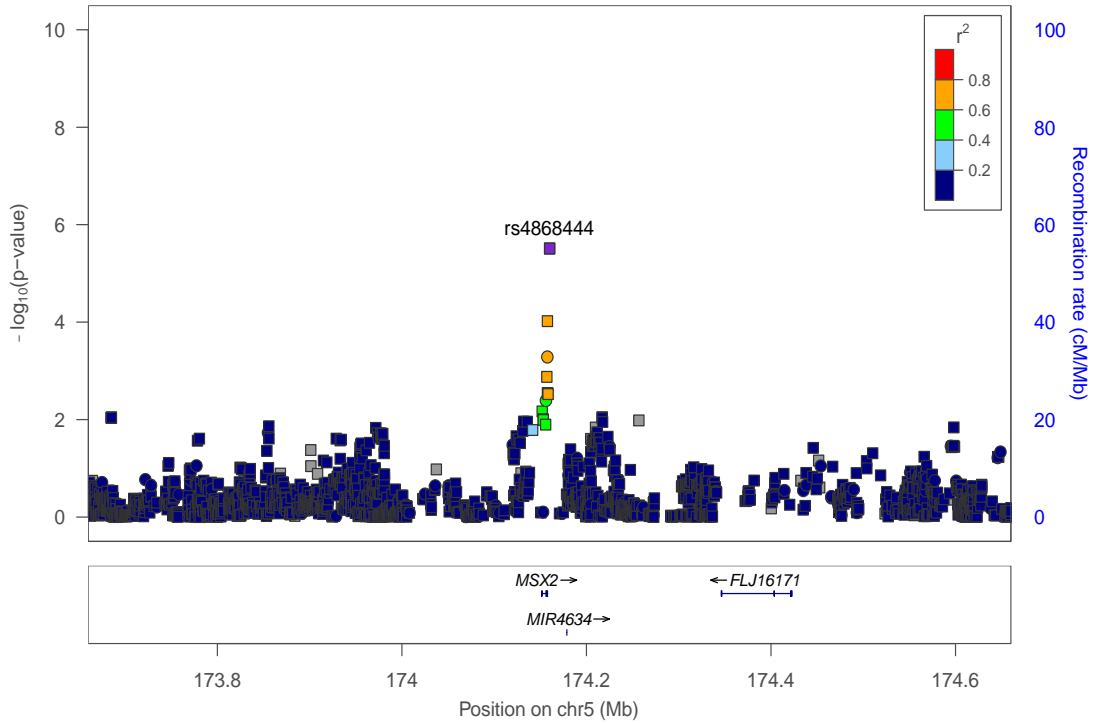**Figure 41 Regional Association Plot of rs6479408 from (DRM) results in POFC**
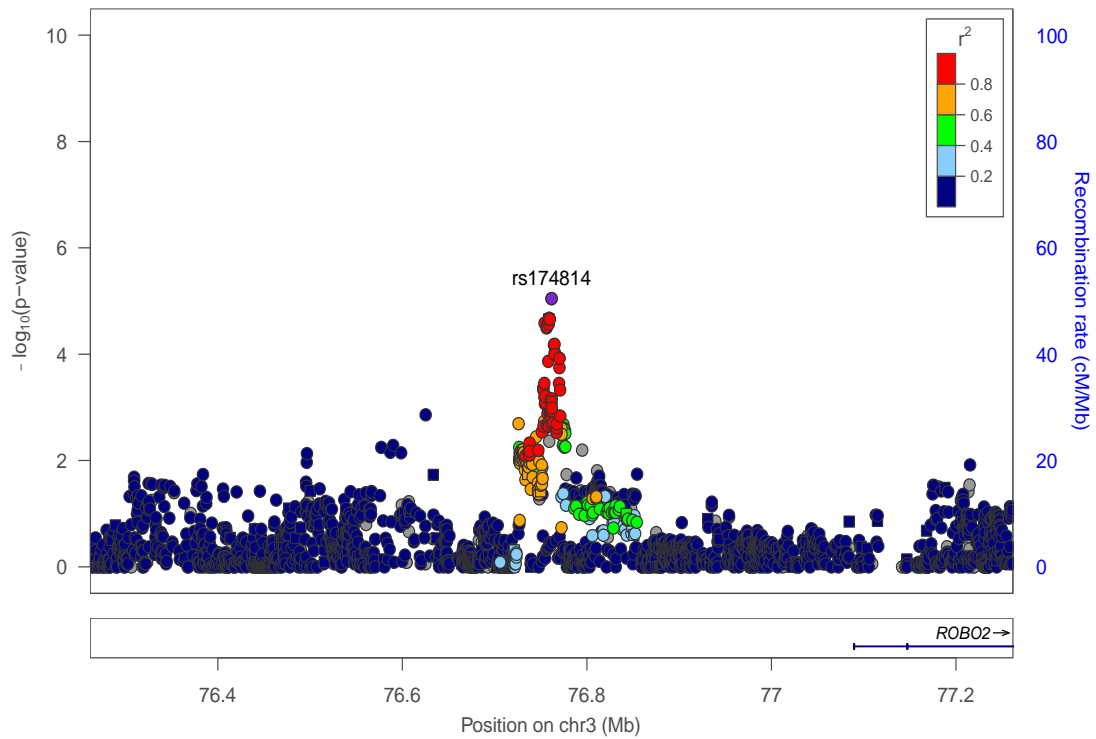
Figure 42 Regional Association Plot of rs174814 from (DRM) results in POFC

4.2.3.1 MAMELONS GWAS RESULTS

The GWAS of mamelons was performed on cohort (N=3579) for whom genotypes data, phenotypes, and covariates data were available. A total of 7,168,909 SNPs was available for analysis after applying quality control criteria. In the GWAS of mamelons, we were able to identify several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Table 25. A visual representation of the association results of mamelons is displayed in Figure 43. The genomic inflation factor (lambda) was=1.00, which indicate no inflation.

Table 25 Top GWAS hits for the Mamelons in POFC

SNP	CHR	BP	*P*	BETA	MAF	type
rs56090116	13	109513810	1.87E-07	-0.0686811	0.05865	imputed
rs9656931	8	120212453	2.92E-07	-0.032211	0.3817	imputed
rs75810764	7	28666301	6.96E-07	-0.0732547	0.05093	imputed
rs2428453	23	16248973	7.72E-07	-0.0470445	0.1004	imputed
rs112004983	7	5013163	8.12E-07	-0.0514651	0.09968	imputed
rs57157422	4	3686670	8.18E-07	0.04933682	0.1802	imputed
rs375366392	21	10434808	1.31E-06	-0.211756	0.05918	imputed
rs62045528	15	93666155	1.63E-06	0.06118702	0.06505	imputed
rs4747992	10	12630865	2.37E-06	-0.0676604	0.1782	imputed
rs61775597	1	11073323	2.85E-06	-0.0545779	0.08292	imputed
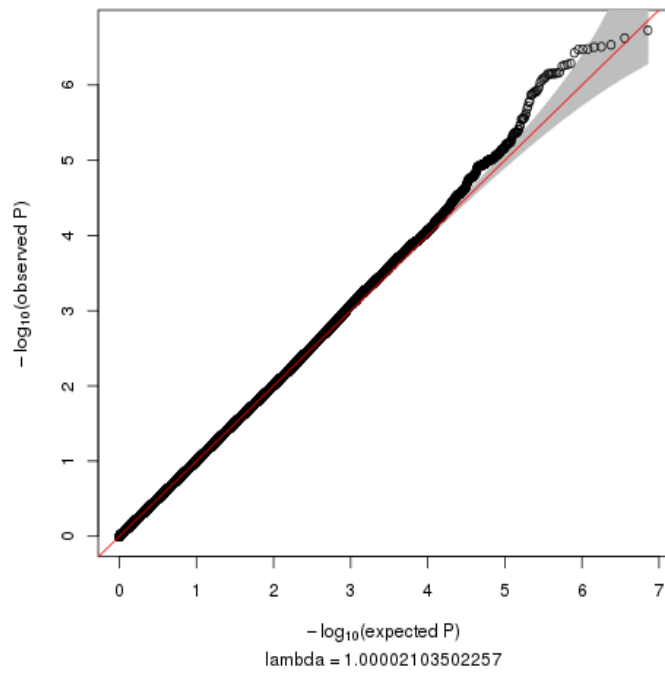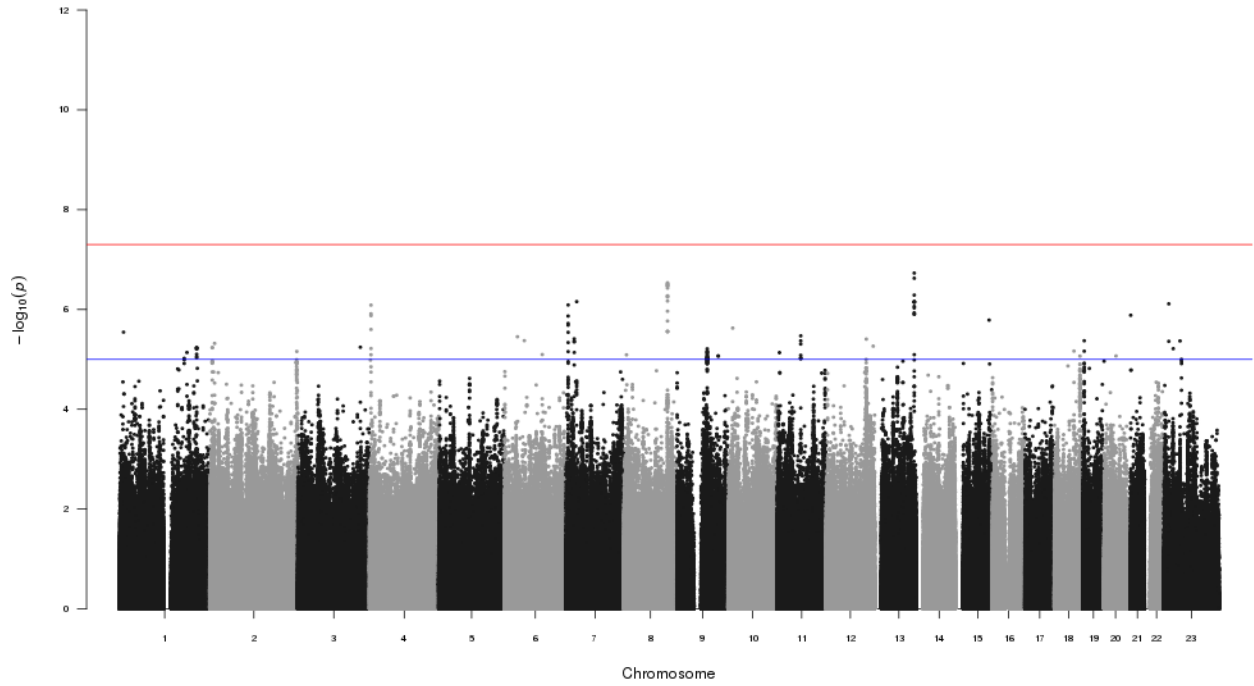
Figure 43 Manhattan and QQ plots for the Mamelons in POFC

One of the top suggestive signals was detected at rs9656931 (beta = -0.032211, P = 2.92E-07),

which is near *ENPP2* gene (Figure 44). This gene is also known as AUTOTAXIN (*ATX*), and the

protein encoded by this gene functions as a phosphodiesterase, and in producing signaling lipid

lysophosphatidic acid (LPA) in extracellular fluids (Perrakis and Moolenaar, 2014). In

development expression study of *ATX* in mice, it was found to be expressed during tooth

development, especially around dental papilla, some of tooth buds, and in ameloblasts (Dietmar,

et al., 1999). These findings could suggest a possible role of *ENPP2* in tooth development and it

could influence mamelons, however, more studies are needed to fully understand the role this

gene play in odontogenesis.
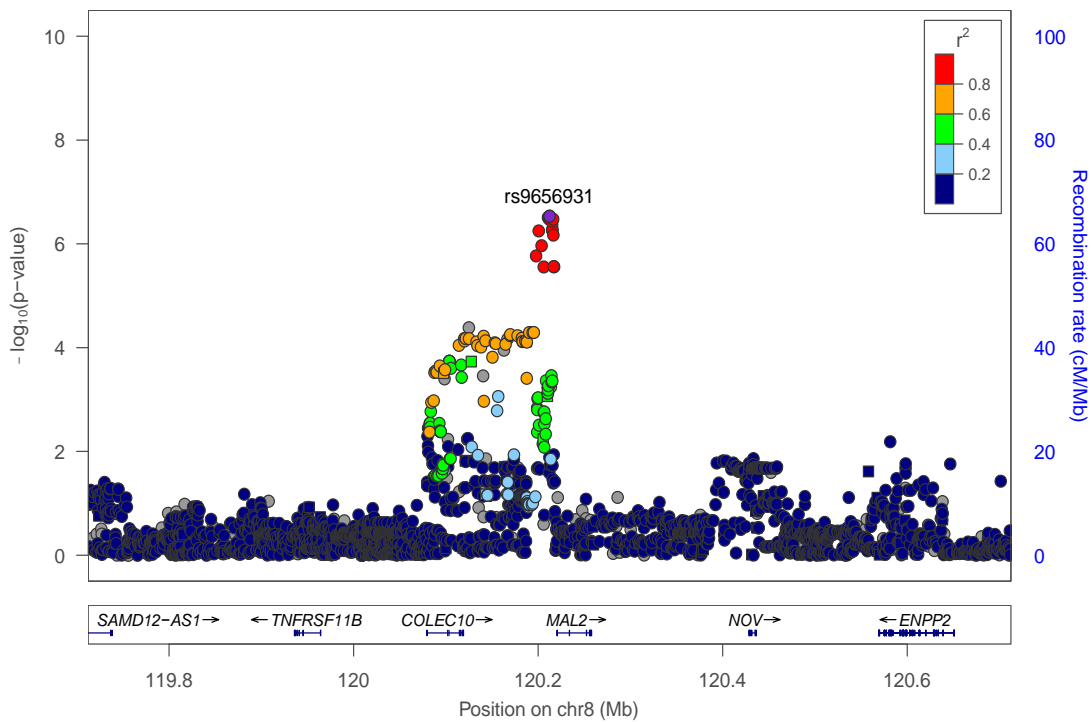
Figure 44 Regional Association Plot of rs9656931 from Mamelons results in POFC

4.2.4 DENTAL CARIES, AGENESIS, HYPOPLASIA (DAH)

Same as the previously discussed associated dental anomalies (AIR, HDR, DRM) we did not have SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$) however, we had several variants that showed suggestive significance ($P < 10^{-5}$). Association results for top hits SNPs are shown in Table 26. A visual representation of the association results of DAH is displayed in Figure 45. The genomic inflation factor (lambda) was=1.09, which indicate slight inflation.

rs79577009 ($P = 2.87E-07$) is located intronic to *SMAD2* gene (Figure 46). The encoded protein of *SMAD2* has an in important function in mediating the signal of the transforming growth factor (TGF)-beta. Therefore, *SMAD2* helps indirectly (TGFb) in regulating multiple cellular processes, such as cell proliferation, differentiation, and odontogenesis (Sarkar, et al., 2000). An experiment in mice provided evidence that smad2 had an essential involvement during early stages of tooth formation (Ito, et al., 2001). These evidences indicate that *SMAD2* may have a role during tooth development and may influence dental health and could also be associated with our correlated dental anomalies (DAH).

rs11125855 ($P = 8.82E-07$) is located upstream to *REL* gene (Figure 47) and the encoded protein belongs to the Rel homology domain/immunoglobulin-like fold plexin, transcription factor (RHD/IPT) family, and play roles in different biological processes such as apoptosis, inflammation, as well as the immune response. In addition, *REL* protein have an important role in the survival and proliferation of B lymphocytes. These B lymphocytes produce cytokine as a response to inflammation, such as in the case of dental caries, to regulate the intensity and the duration of the immune response (Jontell, et al., 1998). Functional analyses showed that *REL*, which is also known as c-rel, control the development of the epidermis and associated

appendages, such as teeth, during embryogenesis in mice (Gugasyan, et al., 2004). Based on these studies, it is plausible that *REL* might have a role in tooth development and might be associated with these correlated dental anomalies (DAH).

rs6758898 (*P* = 2.93E-06) is located near to *IL1A* and *IL1B* (Figure 48)*,* and they belong to the interleukin 1 cytokine family. The encoded protein of these genes has multiple and different functions within the immune system. Interleukin-1 alpha (*IL1A*) has a role in bone resorption, and positively affects the survival and differentiation of osteoclast and odontoclast (Song, et al., 2013). A systematic review of twenty-seven studies that investigated if there is an association between *IL-1* gene variants chronic periodontitis in white patients and they found that genetic variation in both of *IL-1a* and *IL-1b* could contribute to chronic periodontitis in whites (Karimbux, et al., 2012). Interleukin-1 beta (*IL1B*) has an important function as a mediator in the inflammatory response. In a case-control study in China that investigated the association between dental caries and *ILB1* variant and dental caries, they found that there is a significant association between dental caries and *ILB1* (Hu, et al.,2019). This association signal (rs6758898) is actually overlapping with another association signal (rs6758898; *P* = 8.90E-07) from the previously discussed GWAS of tooth agenesis. These evidences strongly suggest that *IL1A* and *IL1B* have a role in tooth development and could influence these correlated dental anomalies.

rs4868444 (*P* = 3.09E-06), is intronic to *MSX2* (Figure 49), the encoded protein provides the balance between survival and apoptosis of neural crest-derived cells, which is necessary for proper craniofacial morphogenesis. In addition, *msx2* protein is part of the bone morphogenic protein (BMP) signaling pathway, and this pathway regulate various processes, including odontogenesis (Davidson, 1995). A mutation in msx2 has been found in a family with amelogenesis imperfecta and impaired tooth eruption (Suda, et al., 2006), and *Msx2* knockout

mice showed a tooth abnormality resembling amelogenesis imperfecta (Satokata, et al., 2000). This association signal (rs4868444) is actually overlapping with another association signal (rs4868444; $P = 2.62E-07$) from the previously discussed GWAS of tooth agenesis. rs5850440 ($P = 3.37$ E-06) is located intronic to *ROBO1* (Figure 50), the encoded protein of *ROBO1* function as cell receptor for slit1, which have a role in axonal pathfinding. In mRNA expression of slit-1 and 2 in mice, it was found that both of slit1 and robo1 were evident in the primary enamel knot and during the cap stage, robo1 expression in the middle of tooth germ and dental papilla was noted. Moreover, it was found that even before birth both of robo1 and robo2 were localized in preodontoblast (Sigbjørn, et al., 2001).

Table 26 Top GWAS hits for the Correlated Dental anomalies (DAH) in POFC

SNP	CHR	BP	F*	Weights*	*P*	Effect A.	MAF	Type
rs79577009	18	45411221	11.13	-0.205, -0.003,0.981	2.87E-07	T	0.1441	Imputed
rs1267310	1	46987307	10.49	0.483, -0.006,0.867	7.20E-07	A	0.3107	Imputed
rs11125855	2	61021349	10.35	0.857,0.518, -0.197	8.82E-07	G	0.4123	Imputed
rs6944859	7	150726689	10.18	-0.525,0.533,0.648	1.13E-06	T	0.2477	Imputed
rs62405343	6	19219615	9.658	0.548,0.210,0.803	2.40E-06	T	0.1745	Imputed
rs6758898	2	113382684	9.52	0.038,0.967,0.236	2.93E-06	G	0.4478	Imputed
rs2017839	12	70497726	9.493	0.552, -0.434,0.690	3.04E-06	A	0.2603	Imputed
rs4868444	5	174160113	9.484	0.0487,0.956, -0.296	3.09E-06	T	0.07077	Imputed
rs5850440	3	79649799	9.422	-0.788, -0.545,0.361	3.37E-06	AT	0.3316	Imputed
rs35687319	2	227955084	9.257	0.437,0.041,0.890	4.27E-06	T	0.0954	Imputed
rs1518612	2	106093166	9.182	0.955,0.003,0.28	4.75E-06	A	0.08486	Imputed
rs60345613	6	163038671	9.161	-0.127, -0.149,0.984	4.90E-06	G	0.2413	Imputed
rs6585107	10	113486544	9.03	0.360,0.125,0.919	5.92E-06	A	0.4117	Imputed
rs4326107	5	179860038	8.922	0.379,0.863,0.359	6.92E-06	T	0.3829	Imputed
rs74663685	6	22000078	8.795	0.566, -0.231,0.781	8.29E-06	G	0.237	Imputed
rs141675959	12	78213869	8.792	-0.898, -0.347,0.332	8.36E-06	A	0.2167	Imputed
rs9365301	6	161997791	8.78	-0.258,0.700,0.652	8.47E-06	T	0.2099	Imputed
rs11293529	3	28827168	8.757	-0.479,0.430,0.757	8.78E-06	C	0.0725	Imputed
rs143358219	23	122701668	8.633	-0.147,0.934,0.308	1.05E-05	C	0.2559	Imputed

* F-Statistic tests the significance of the canonical correlations

* Weights reflect the correlation coefficients for each individual trait
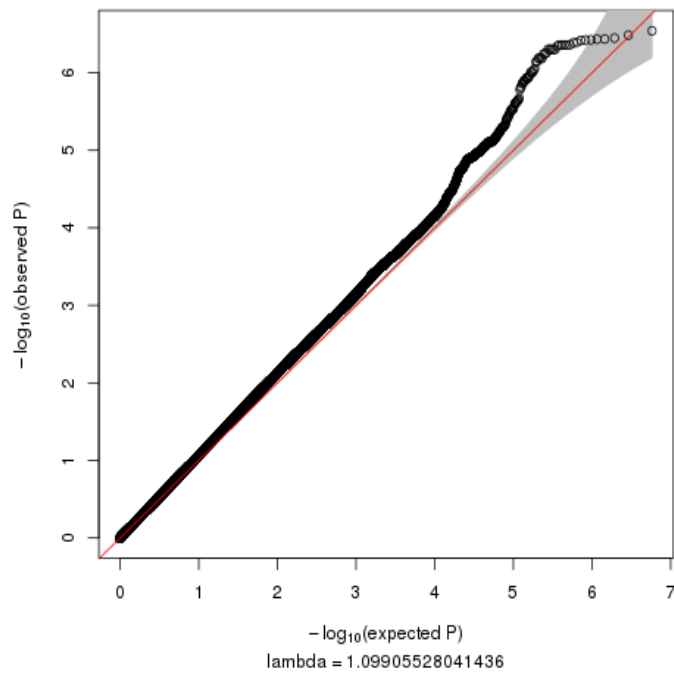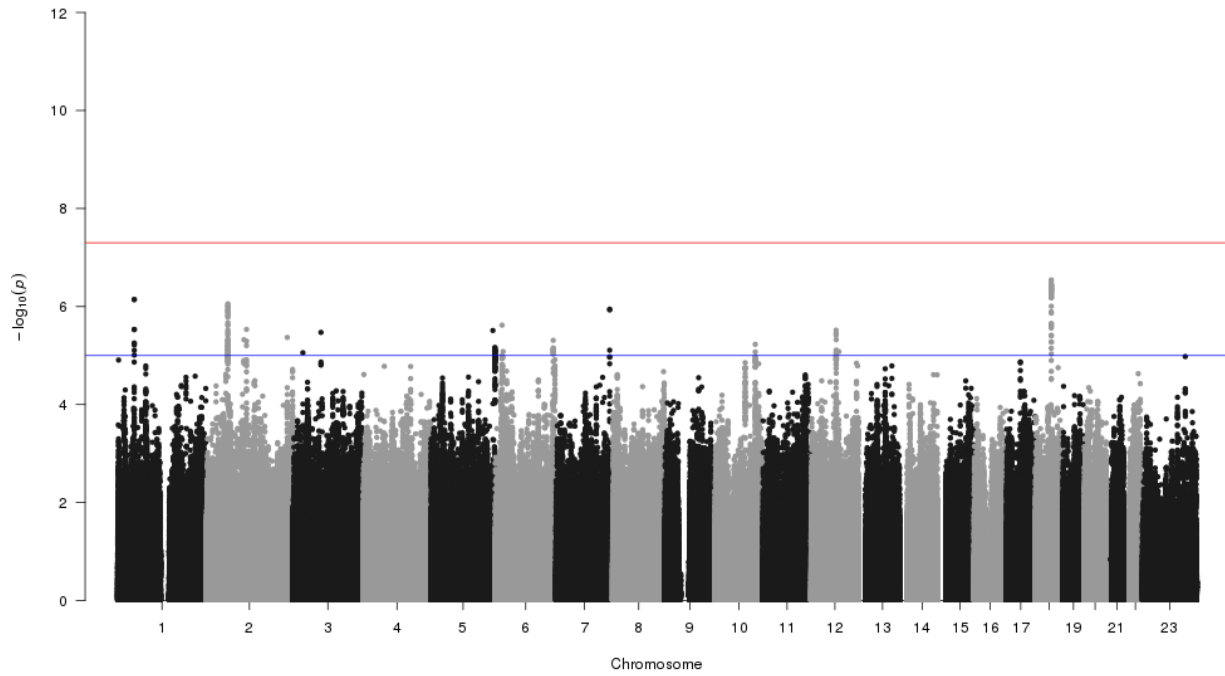
Figure 45 Manhattan and QQ plots for the Correlated Dental anomalies (DAH) in POFC
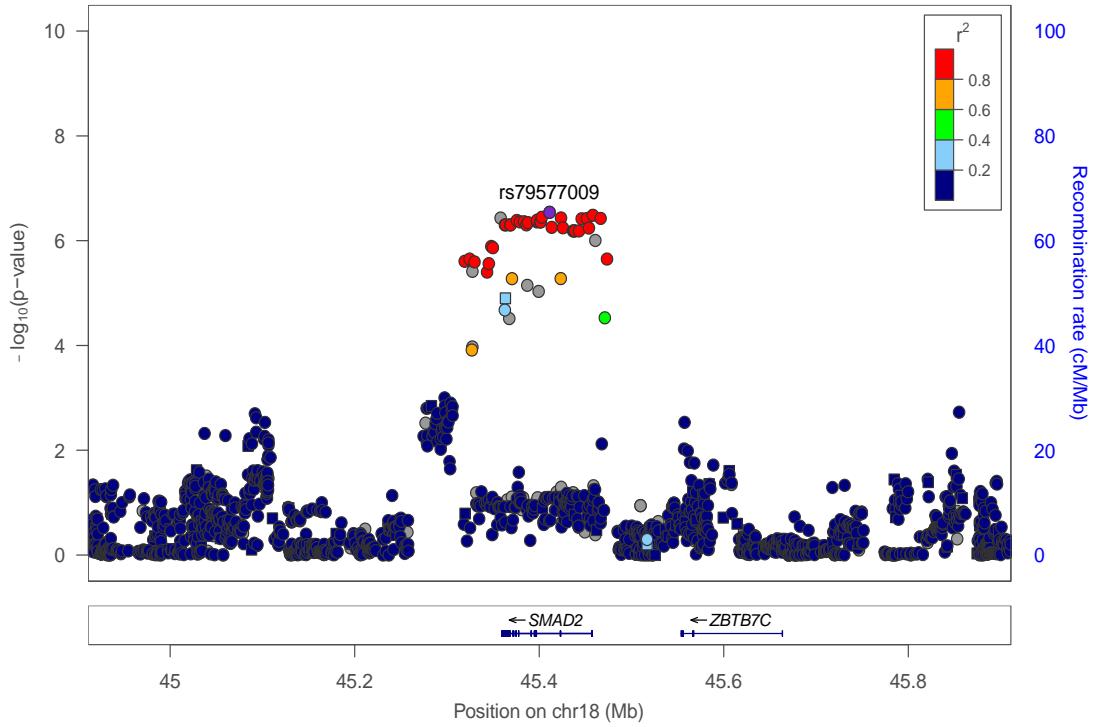
Figure 46 Regional Association Plot of rs79577009 from (DAH) results in POFC

Figure 47 Regional Association Plot of rs11125855 from (DAH) results in POFC

Figure 48 Regional Association Plot of rs6758898 from (DAH) results in POFC

Figure 49 Regional Association Plot of rs48868444 from (DAH) results in POFC
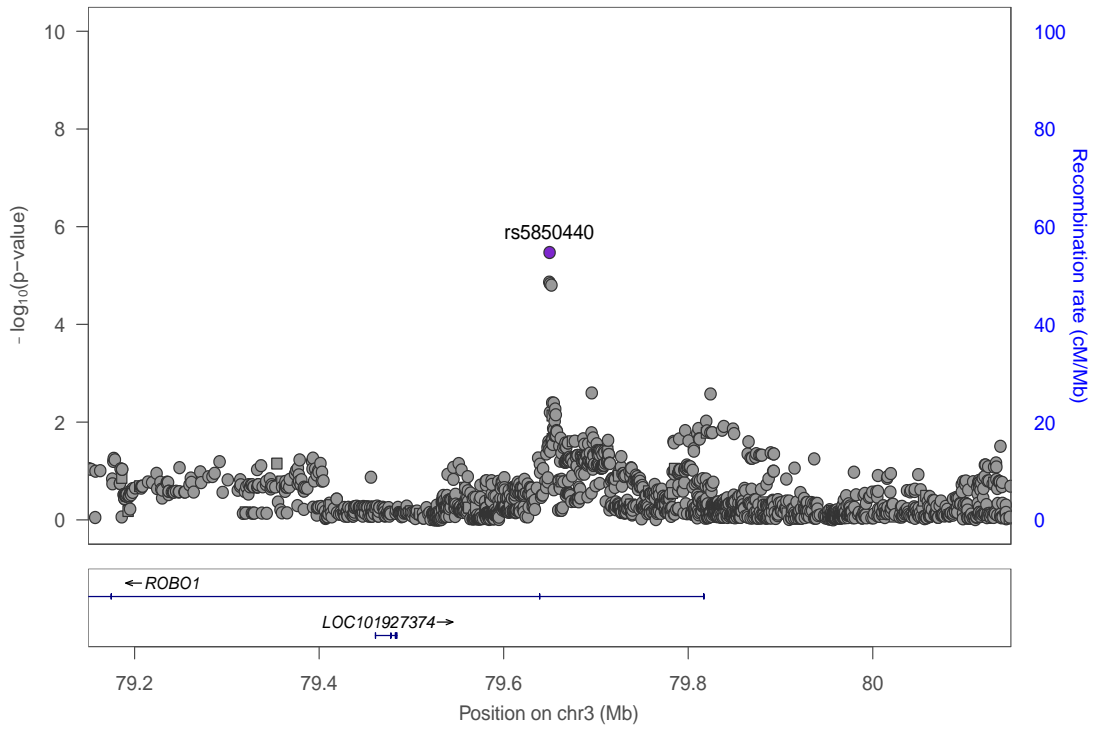
Figure 50 Regional Association Plot of rs5850440 from (DAH) results in POFC

We examined the candidate genes from previous sequencing and GWASs of tooth agenesis in our GWAS results of DAH to see if they could be replicated. We did not have association results for all of them. However, association results for the ones we have in our sample are shown in Table 27. Regions of the genome that showed suggestive significance with tooth agenesis genes in the GWAS results of DAH in our sample were further visualized using Regional Association (LocusZoom) plots (Figure 51).

Table 27 Association results of tooth Agenesis genes in DAH GWAS in POFC

SNP	CHR	BP	*P*	Location to Gene	GENE	A1	Type
rs1611002	2	144049021	1.39E−3	Intronic	*ARHGAP15*	G	Imputed
rs296556	1	200926196	2.22E−3	Upstream	*ASCL5/CACNA1S*	T	Imputed
rs260627	2	109528832	9.8E−3	Intronic	*EDAR*	G	Imputed
rs139528735	2	88440412	3.08E−4	Upstream	*FOXI3*	C	Imputed
rs9310221	3	71849696	9.78E−4	Downstream	*FOXP1*	A	Imputed
rs201012183	2	220211432	1.3E−3	Downstream	*WNT10A*	CTTA	Imputed
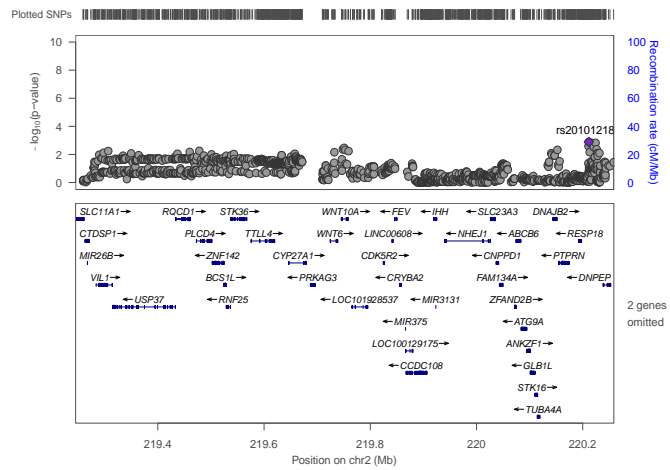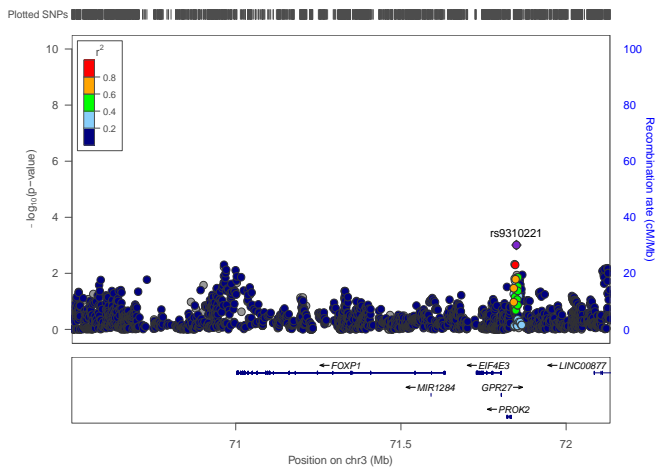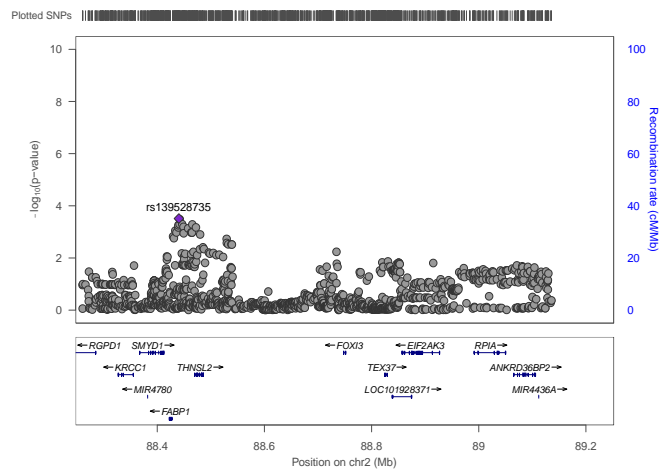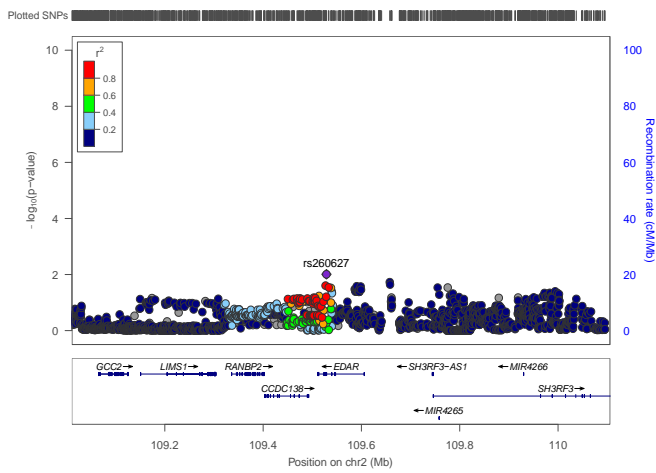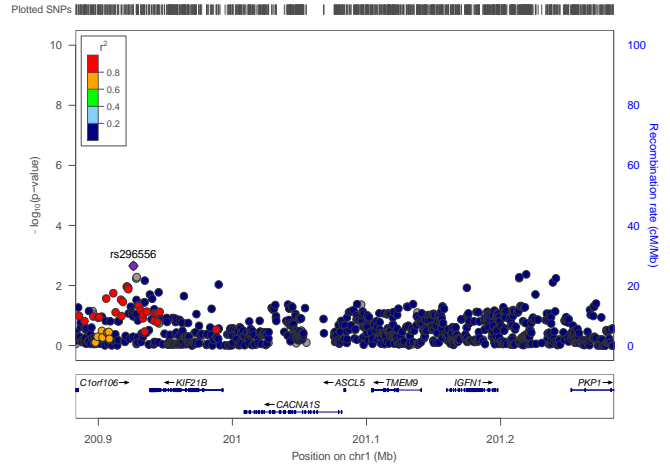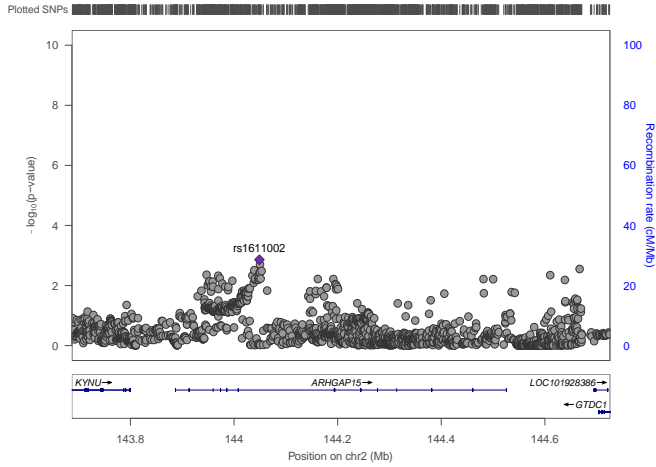
Figure 51 Regional Association Plots of Tooth Agenesis genes in DAH GWAS results

We also examined the previously identified genes from previous GWASs dental caries (Table 28) in our GWAS results of DAH to see if they could be replicated. We did not have association results for all of them. However, association results for the ones we have in our sample are shown in Table 27 Regions of the genome that showed suggestive significance with dental caries genes in the GWAS results of DAH in our sample were further visualized using Regional Association (LocusZoom) plots (Figure 52).

Table 28 Association results of identified dental caries genes in DAH

SNP	CHR	BP	*P*	GENES	Type
rs965399	1	236768404	7.23E−5	*ACTN2, EDARADD, MTR*	Imputed
rs911269	11	30773173	1.06E−4	*MPPED2*	Imputed
rs3744102	17	56383187	4.19E−4	*LPO*	Imputed
rs60565062	6	167691347	1.22E−5	*RPS6KA2*	Imputed
rs35795772	3	50800896	4.03E−3	*ISL1*	Imputed
rs9998672	4	154144630	5.26E−3	*TLR2*	Imputed
rs16849157	1	229035380	8.5E−3	*RHOU*	Imputed
rs500674	10	30561309	1.83E−3	*LYZL2*	Imputed
rs112745971	7	91162751	1.71E−4	*FZD1*	Imputed
rs368270943	3	26723006	1.06E−3	*PTK2B*	Imputed
rs600544	4	5336806	2.76E−4	*AJAP1*	Imputed
rs56361560	4	89118687	1E−2	*ABCG2, PKD2*	Imputed
rs116795308	4	148036545	3.96E−4	*EDNRA*	Imputed
rs61955513	13	21090605	5.09E−3	*IFT88, IL17D*	Imputed
rs663000	18	9638030	7.57E−3	*TWSG1*	Imputed
rs148954978	18	46908296	3.92E−3	*SMAD7*	Imputed
rs257113	7	41830901	8.77E−4	*INHBA*	Imputed
rs60355650	2	218617626	1.63E−3	*CXCR1, CXCR2*	Imputed
rs62272841	3	160314924	1.09E−4	*KPNA4*	Imputed
chr16:30518041	16	30518041	4.59E−3	*ITGAL*	Imputed
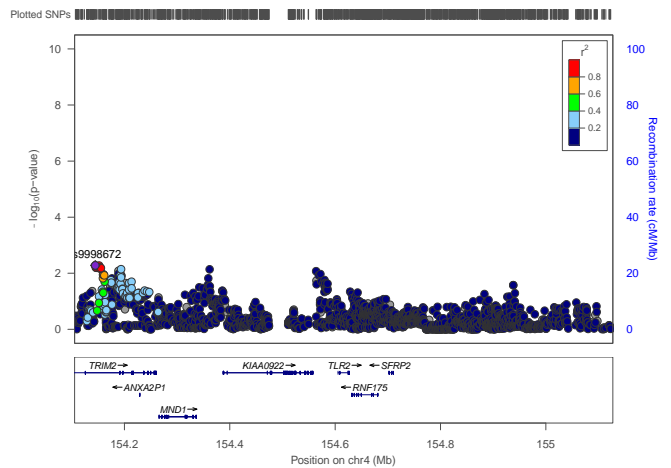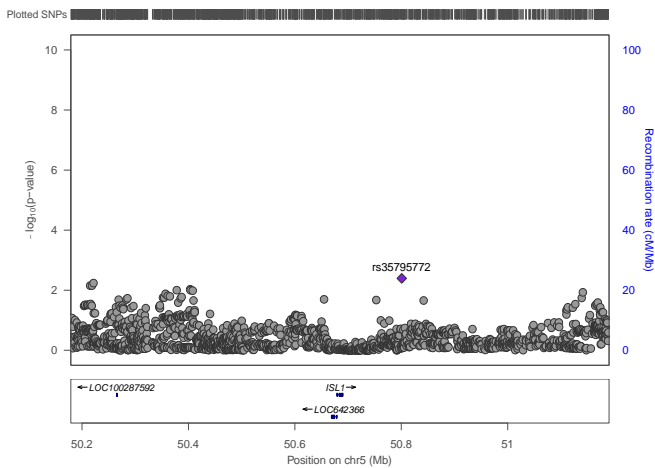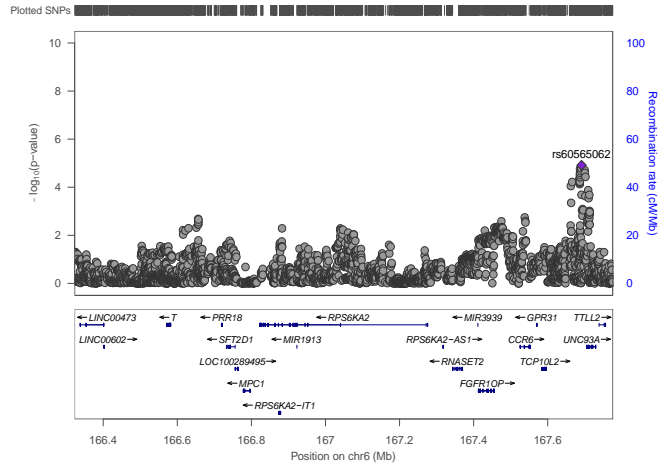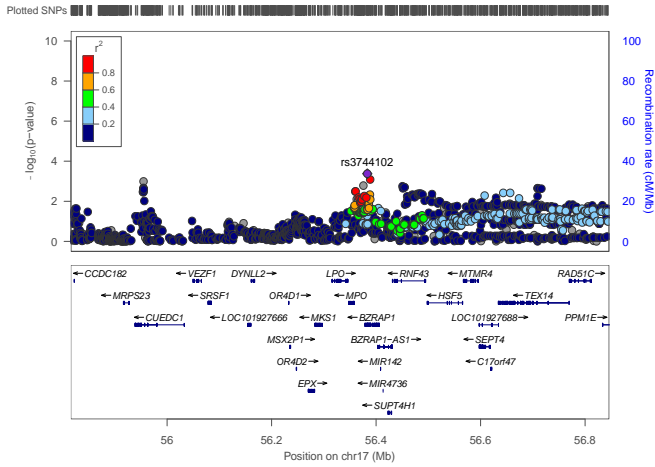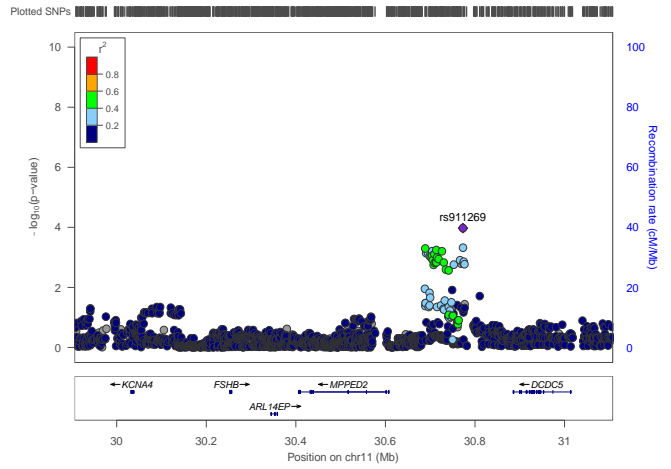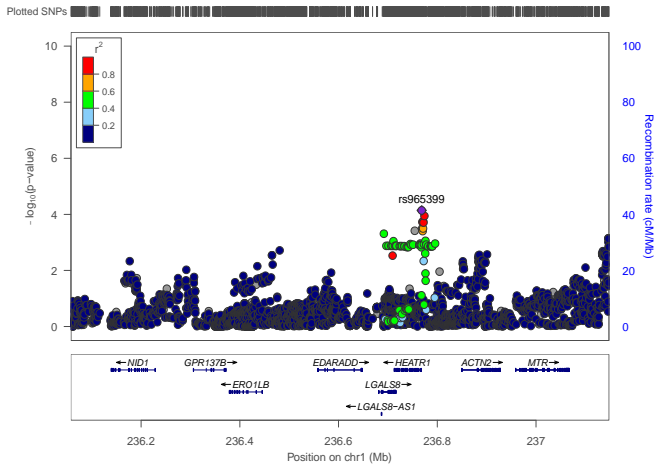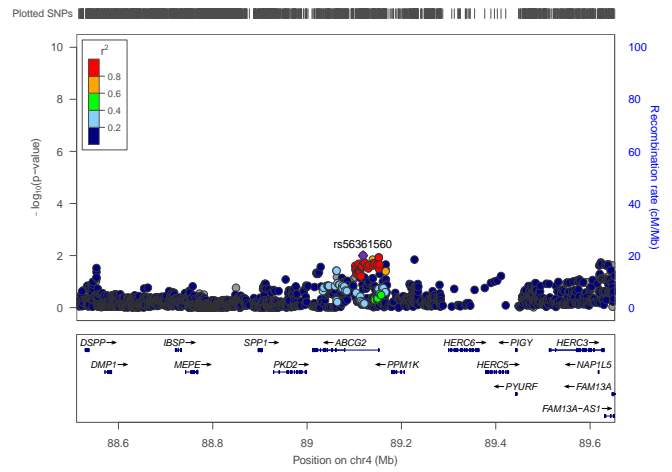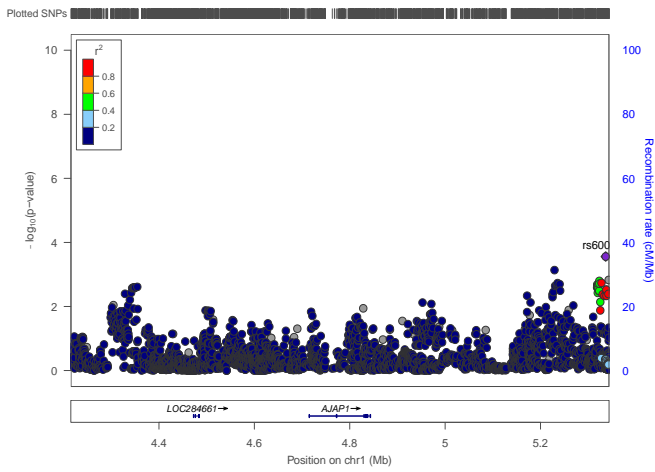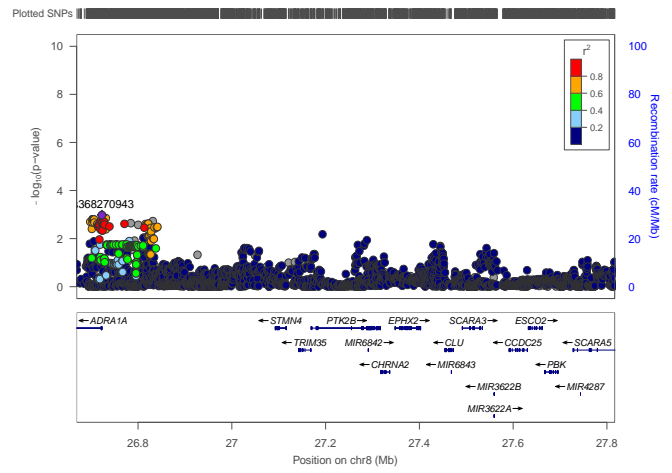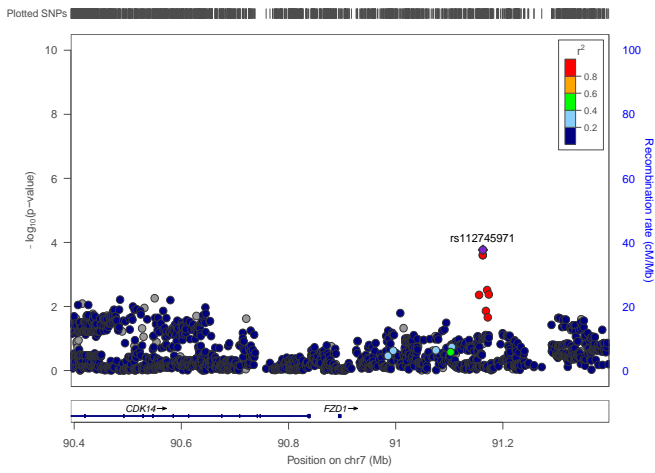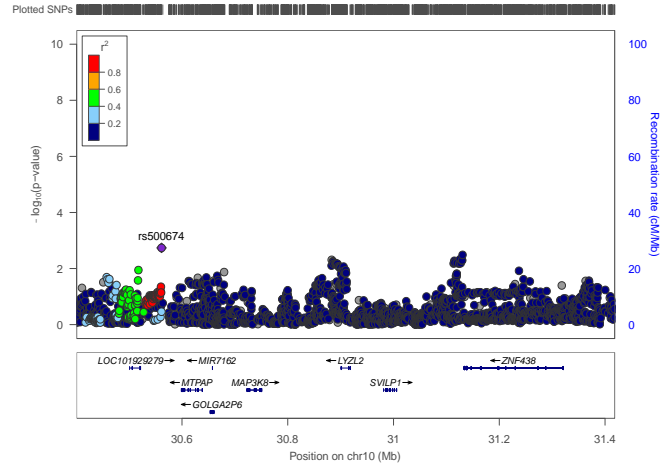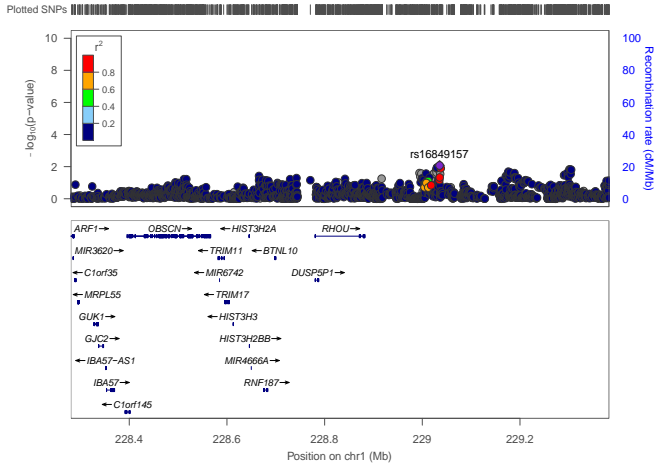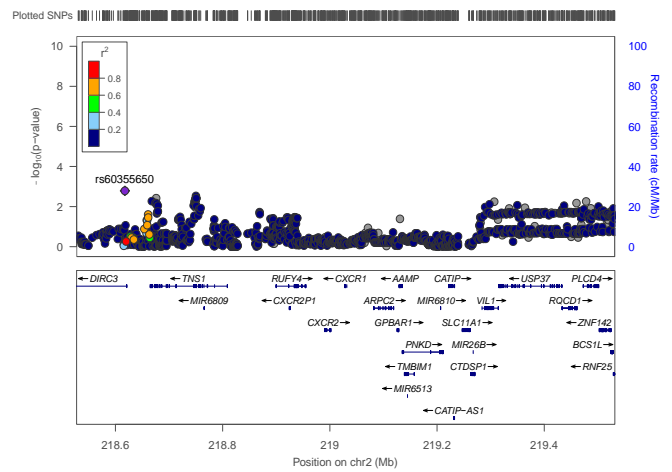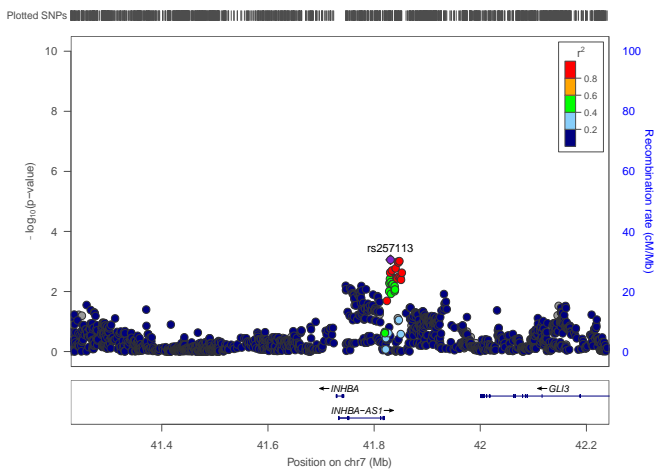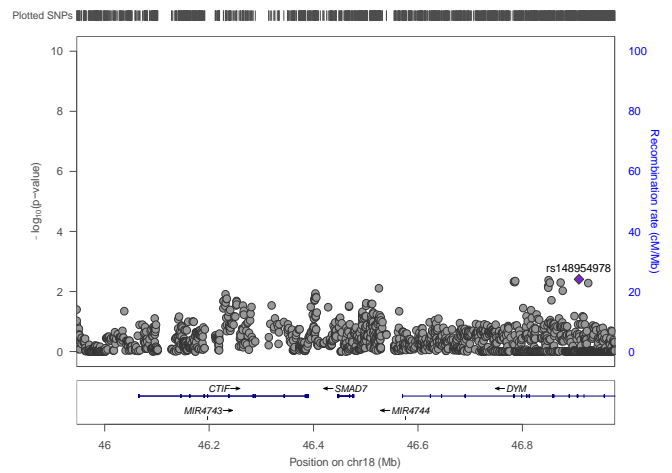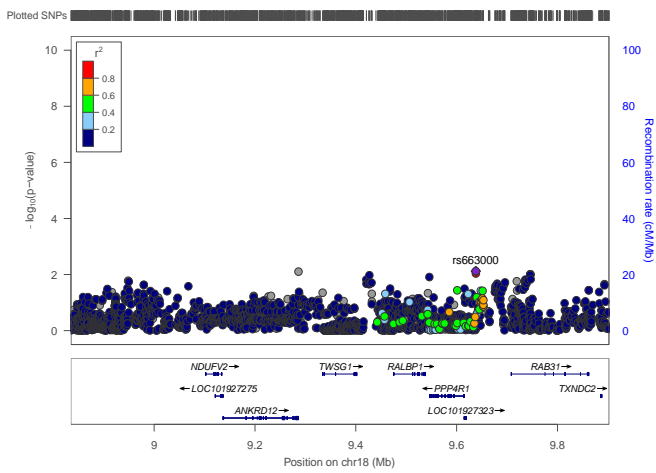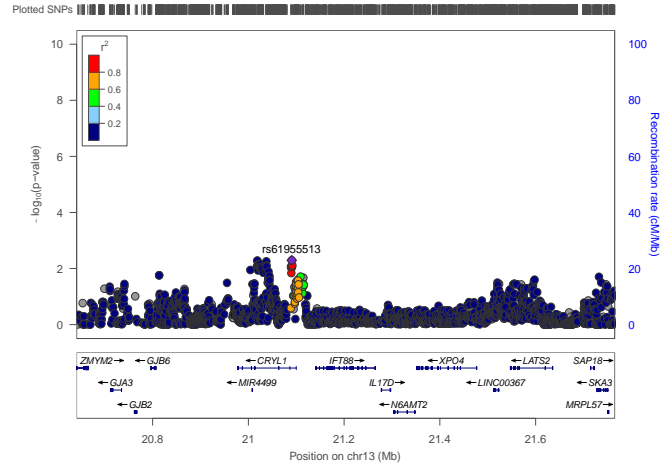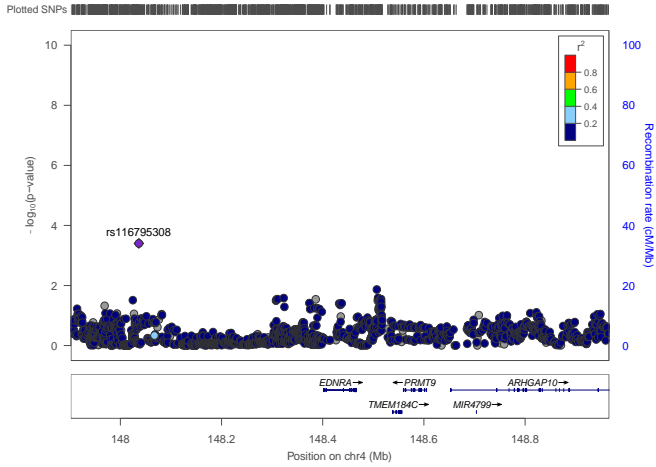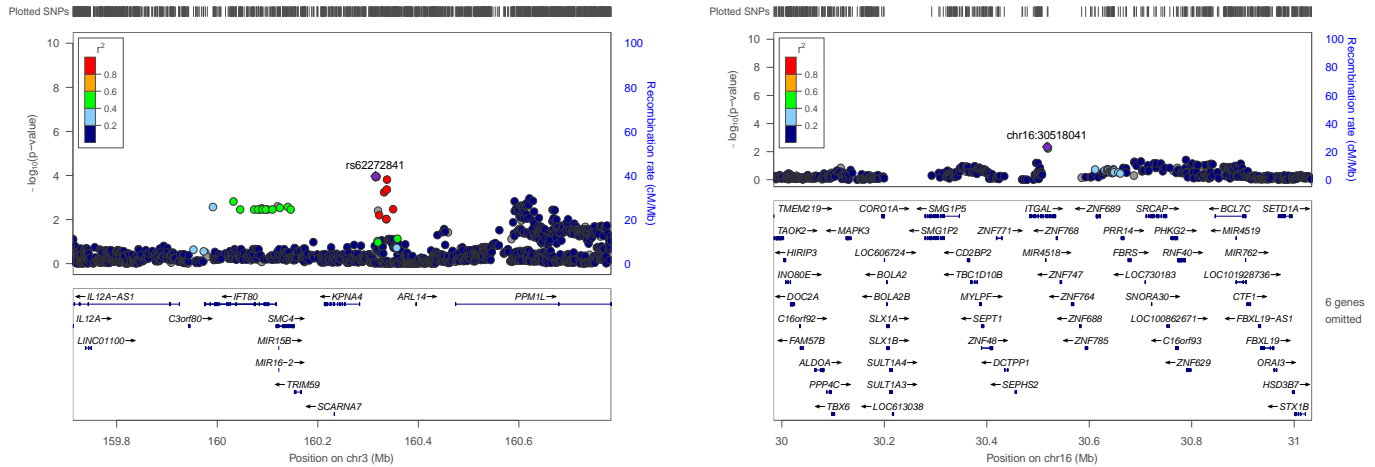
Figure 52 Regional Association Plots of dental caries genes in DAH results

4.3 DISCUSSION

This study is the first multivariate approach of GWAS that aimed to identify possible genetic loci that may be associated with the presence of patterns of correlated dental anomalies (AIR, DRM, HDR, or DAH). Although none of our multivariate GWAS models results reached the strict genome-wide significance level, there were a large number of suggestive variants relevant to odontogenesis, teeth and dental traits that we were able to identify in our study. The suggestive variants were annotated to help generate hypotheses and nominate these variants for further investigations. We did the annotation of these suggestive variants by using different bioinformatics databases (see section 4.1.6) We focused our search on genes with known biological functions related to tooth development process (odontogenesis) and oral/dental health, either in human or in rodents. However, there were other suggestive variants near genes with biological roles in the body that could be of interest for further investigation in future studies.

Regional Association plots of SNPs that reaching suggestive level near genes with unknown roles in dental and/or oral health are in the appendix Figure A3 through Figure A22.

The strongest association signal in the multivariate patterns was detected in the DAH pattern (dental caries, agenesis, and enamel hypoplasia). The lead variant, rs79577009 (P = 2.87E-07) is located intronic to *SMAD2* gene, which has an in important function in mediating the signal of the transforming growth factor (TGF)-beta and helps indirectly (TGFb) in regulating multiple cellular processes, including odontogenesis (Sarkar, et al.,2000). In addition, smad2 was found to be involved during the early stages of tooth development in mice (Ito, et al., 2001). Thus, we could hypothesize that *SMAD2* might increase or influence the chance of having any of these correlated dental anomalies; dental caries, agenesis, and hypoplasia, and we can nominate this gene to be associated with these correlated dental anomalies. Additionally, one of the suggestive association signals in DAH was detected near *REL* (rs11125855; P = 8.82E-07). One of the roles of *REL* gene is in regulation of the survival and proliferation of B lymphocytes and these B lymphocytes produce cytokine as a response to inflammation, such as in the case of dental caries, to regulate the intensity and the duration of the immune response (Jontell, et al., 1998). *REL* also seems to have a role in controlling the development of the epidermis and associated appendages, such as teeth, during embryogenesis in mice (Gugasyan, et al., 2004). Although it is still unclear the role that *REL* might have in tooth development and dental health, there is supporting biological evidence that it may influence tooth development and those correlated dental anomalies.

Many of the genes identified in the multivariate patterns (AIR, HDR, DRM, and DAH) in this study play some role in tooth development, either directly or indirectly. Genes like : *ADAMTS9* (Christian, Bahudhanapati & Wei, 2013; Jungers, et al., 2005), *PRICKLE2* (Nishikawa& Kawamoto, 2015; Everman, 2018), *ARNT2* (Aitola & Pelto-Huikko, 2003), *NTN1* (Serafini, et al., 1996; Ren, et al.,

129

2004; Løes, et al., 2003), *TRPC4* (Yang, et al., 2017), *WNT2B* (Yi Q, et al., 2017) in AIR pattern; *ROR2* (Lin, et al., 2011), *OMD* (Lin, et al., 2019), *WDR72* (Katsura, et al.,2014; El-Sayed, et al., 2009), *BMP7* (Zurowski, et al., 2018; Gao, et al., 2015), *WNT5A* (Lin, et al.,2011) in HDR pattern; *GLIS3* (Beak, et al., 2007), *ROR2* (Lin, et al., 2011), *OMD* (Lin, et al., 2019), *WDR72* (Katsura, et al.,2014; El-Sayed, et al., 2009) in DRM pattern; *SMAD2* (Sarkar, et al.,2000; Ito, et al.,2001), *MSX2* (Davidson, 1995; Suda, et al., 2006), and *IL1A & IL1B* (Song, et al.,2013; Karimbux, et al., 2012) in DAH pattern.

In addition to the multivariate GWAS that we did, we conducted a univariate GWAS for each of dental anomalies and had some interesting results near genes with possible roles in tooth development and it could be associated with or increase the risk of some of these structural dental anomalies. Among these results, we were able to identify genes that were associated only with specific dental anomalies. *FAM20C* gene that was positively involved in regulating hDPC odontoblastic differentiation (Li, et al., 2018) was associated with tooth agenesis in our study. In the GWAS of hypoplasia we were able to identify: *CCK* gene, which has been found to be highly regulated in enamel maturation (Lacruz, et al., 2012); *ADAR* gene, which was found to be in dental papilla, ameloblasts, and odontoblasts during tooth development in mice, and it was found that novel mutation in *ADAR1* caused dental anomalies, more specifically dens evaginatus and dens invaginatus (Kantaputra, et al., 2012); and *PYGO1* gene, which was found to be essential for developing normal enamel in mice (Cantù, et al., 2017). In the GWAS of impaction we were able to identify *CDH1* gene, which has been found before to have a vital role during craniofacial development and was found to be associated with CL/P with or without tooth agenesis (Letra, et al., 2009). In the GWAS of rotation we were able to identify *FLRT2* gene, which was found to be highly expressed in the neural crest-derived mesenchyme in the medial aspect of the developing frontonasal region in mice and it was in close relationships with the expression of Fgfr2, Shh,

and Msx1, and those genes have shown previously to play critical roles during craniofacial development, and tooth development in particular (Gong, et al., 2009). In the GWAS of displacement we were able to identify: *FOXF1*and *FOXC2,* which was found that they were expressed in in developing tooth buds in mice and needed for proper tooth development (Iida, et al., 1997, Wang, et al., 2003); and *ADAMTS* genes, which were found to be found to be expressed in mice dental pulp cells, odontoblasts, cementoblasts, cementocytes, periodontal ligament cells, osteoblasts and osteocytes (Sone, et al., 2005). Based on these biological evidences, it is plausible that these genes may influence tooth development, dental health, and dental anomalies.

An overlap in results of HDR GWAS and DRM GWAS model was detected. The results of these overlapped variants are presented in appendix Table A 6. One of the overlapped results that we found in both HDR and DRM was a suggestive association signals near *WDR72* gene, and this gene was found to have an essential role in the mineralization and maturation of tooth enamel (El-Sayed, et al., 2009). So, this gene can be influencing these correlated structural dental anomalies (HDR and DRM) and further investigation is needed to understand the exact role this gene has in tooth development and/or dental health. Another overlap in the results of both HDR and DRM was detected near *ROR2* and *OMD* genes, and both of these genes had a role in tooth development directly or indirectly (Lin, et al., 2011; Lin, et al., 2019) thus we can suggest that these genes might be influence tooth development and dental health and can also be associated with these dental anomalies. The overlap in results between HDR and DRM patterns can simply be explained by the fact that they share two of the correlated dental anomalies; displacement and rotation, which can majorly influence this overlap in results. The overlap between association signals has been also seen between the multivariate GWAS results and the univariate GWAS

results of dental anomalies. Signals that were suggestively significant in the DAH and agenesis seen near *IL1A* and *IL1B* and *MSX2* genes. Based on the biological studies that been previously discussed, it is plausible that these genes have a role in tooth development and could influence these dental anomalies.

We examined the candidate genes from previous sequencing and GWASs of tooth agenesis in our AIR GWAS module results, Agenesis GWAS results, and in DAH GWAS results to see if they had been replicated. We only had limited evidence of suggestive associations that was identified near tooth agenesis genes; *ASCL5/CACNA1S, ARHGAP15, FOXI3, EDAR*, *WNT10A, MSX1, PAX9,* and *AXIN2.* In addition, we also examined the previously identified genes from previous GWAS's dental caries (Table 28) in our GWAS results of DAH to see if they could be replicated. We only had limited evidence of suggestive associations that was identified near some of these genes, which are summarized in Table 4.

4.4 CONCLUSION

This is the first study to investigative genetic associations for multivariate patterns of correlated dental anomalies, and we were able to identify suggestive association signals near genes with biological roles during tooth development. Our multivariate GWAS identified suggestive association signals ($P < 1 \times 10^{-5}$) near genes with roles in tooth development and/or dental health, including *ADAMTS9* and *PRICKLE2* were associated with AIR; *GLIS3*, *WDR72*, and *ROR2* were associated with HDR and DRM*; ROBO2* was associated with DRM; *BMP7* was associated with HDR*;* and *ROBO1, SMAD2* and *MSX2* were associated with DAH. In addition,

there were genes with plausible role in tooth development that we were able to identify in this study, such as: *SMAD7, FLRT2, PYGO1, IRF8, ENPP2,* and *FOX* genes that were associated with specific dental anomalies. Further studies are needed to replicate the results in an independent cohort. Additional studies are needed to investigate the nominated genes in human and animal models to fully understand their relevance for tooth development, dental anomalies and dental caries.

5.0 CONDUCT GENOME-WIDE ASSOCIATION ANALYSES IN TWO INDEPENDENT STUDY SAMPLES, POFC AND COHRA1, TO IDENTIFY COMMON VARIANTS ASSOCIATED WITH DENTAL CARIES; FOLLOWED BY META-ANALYSIS

5.1.1 SAMPLE DESCRIPTION

POFC & COHRA1

The sample description was previously described in section 3.1.1

5.1.2 DATA COLLECTION

The data collection procedures were previously described in section 4.1.2

5.1.3 DENTAL CARIES PHENOTYPE

For this aim, we used dental caries phenotype (Total dft+DFT index) as a quantitative trait, i.e. the total number of teeth with decayed, and/or filled/restored surfaces. We used the total DFT+dft index, in both POFC and COHRA, instead of DMFT/dmft due to incomplete information regarding missing teeth for the majority of the participants.

5.1.4 GWAS

For the aim, dental caries was treated as a quantitative trait and the associations between SNP genotypes and dental caries was assessed using linear regression with adjustment for age, age^2, sex, and site. Under the GWAS approach, millions of SNPs were tested, and to avoid false positive results, multiple testing correction was required and the conservative genome-wide significance threshold of $P < 5 \times 10^{-8}$ (Bonferroni correction for a million tests) has become commonly used in GWAS's.

Since the POFC sample comes from a larger family study with different ethnicities, we used the exact association test statistics that could account for relatedness and population structure, i.e. the variance component approach implemented in Efficient Mixed-Model Association eXpedited (EMMAX) software (Kang, et al., 2010). The method was previously described in section 4.1.4

5.2 RESULTS:

Here we are going to report the results for POFC and COHRA1 separately

5.2.1 POFC

A total of 7,168,908 SNPs was available for analysis after applying quality control criteria. A genome-wide association analysis was implemented in EMMAX using linear regression. Manhattan and QQ-plots for the quantitative trait are shown in Figure 53. The genomic control

lambda was calculated to be $\lambda = 1.008$. Any loci that included a SNP with a P value $< 1 \times 10^{-5}$ for an association with dental caries was plotted using Locus Zoom to display the linkage disequilibrium. A genome-wide significance threshold ($P < 5 \times 10^{-8}$) and suggestive level of significance ($P < 1 \times 10^{-5}$) were assigned.

We had SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$); and there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for our top SNPs are shown in Table 29.

The strongest association signal that reached the whole genome significance level was detected near rs4672405 (beta = -0.6183, P = 3.91E-09) located on the chromosome 2, and lies near the *REL* gene (Figure 54). The encoded protein of *REL* belongs to the Rel homology domain/immunoglobulin-like fold plexin, transcription factor (RHD/IPT) family, which plays roles in different biological processes such as apoptosis, inflammation, as well as the immune response. *REL* protein more specifically has an important role in the survival and proliferation of B lymphocytes. These B lymphocytes produce cytokines as a response to inflammation, such as in the case of dental caries, to regulate the intensity and the duration of the immune response (Jontell, et al., 1998). Functional analyses showed that *REL*, which is also known as c-rel, controls the development of the epidermis and associated appendages, such as teeth, during embryogenesis in mice (Gugasyan, et al., 2004). Based on these studies, it is plausible that *REL* might play a role in the dental caries process.

Variant rs1518612 (beta = -0.9218575, P = 7.04E-08) on chromosome 2 is located near to *FHL2* gene (Figure 55), which encodes a member of the four-and-a-half-LIM-only protein family and depending on the protein partners involved it act as either an activator or repressor of tissue-

specific gene expression. It has been found that *FHL2* is expressed in odontoblasts and odontoblast-like cells in mature human teeth in both normal and pathological conditions (Du, et al., 2012). Another study found that *FHL2* might mediate mesenchyme cell differentiation by interacting with *Runx2* during early stages of tooth development and human dental pulp cell differentiation (Du, et al., 2016). These lines of evidence suggest that *FHL2* might have a role during tooth development and could influence dental caries and further studies are needed to fully investigate their role during tooth development.

rs60191353 (beta = -1.1457371, *P* = 4.15E-07) on chromosome 1 is located upstream to *PRDM16* (Figure 56), the encoded protein of this gene is a zinc finger transcription factor and it also has positive regulatory (PR), repressor and acidic domains (Mochizuki, et al., 2000). In mice, *Prdm16* is expressed in embryonic craniofacial tissue; including primary and secondary palates, Meckel's cartilage and teeth. It has been found that mutation in *Prdm16* gene caused a cleft in the secondary palate in mice and it been nominated as a candidate gene for mutation in human clefting, especially NSCP (Bjork, et al., 2010). It is still unclear the role that *PRDM16* may have in tooth development or in dental caries.

rs62290079 (beta = -0.6814, *P* = 1.82E-06) on chromosome 3 is located near *EPHB3* gene (Figure 57). The encoded protein of *EPHB3* belongs to a family of cell surface receptors with several roles in regulating different aspects of embryonic development, including palatogenesis (palate formation). Mouse models showed mutations in this locus result in craniofacial defects, including cleft palate (Risley, et al., 2009). *EPHB3* does not appear to have a known relevant role in dental caries, however.

Variant rs911269 (beta = 0.49855, *P* = 2.36E-06) on chromosome 11 is located near *MPPED2* gene (Figure 58). The encoded protein of *MPPED2* might have a function during brain

development. In a genome-wide association meta-analysis study that investigated several candidate genes, one of these genes was *MPPED2,* they found that *MPPED2* was significantly associated with dental caries in children (Sheffer et al., 2011). This evidence suggests a possible role of *MPPED2* in tooth development and dental caries.

Variant rs59286075 (beta = -0.94104, *P* = 6.02E-06) on chromosome 2, is located near *GLI2* (Figure 59)*,* and the encoded protein has a function as a mediator for the Sonic hedgehog (Shh) signaling pathway, which plays a regulatory role in several processes during embryogenesis, including odontogenesis (tooth development) (Mo, et al., 1997). *Gli2* were found to be expressed during early stages of murine tooth development and mutant mice were found to have abnormal development of maxillary incisors (Hardcastle, et al., 1998). These studies show that *GLI2* might have a role in tooth development.

Table 29 Top GWAS hits for the associated Dental Caries in POFC

SNP	CHR	BETA	*P*	BP	A1	MAF	Type
rs4672405	2	-0.6183138	3.91E-09	61031500	C	0.4692	Imputed
rs1518612	2	-0.9218575	7.04E-08	106093166	A	0.08486	Imputed
rs6449554	5	0.56356632	3.90E-07	61079024	T	0.398	Imputed
rs60191353	1	-1.1457371	4.15E-07	2967932	T	0.06529	Imputed
rs62290079	3	-0.6814766	1.82E-06	184361139	A	0.1405	Imputed
rs911269	11	0.49855408	2.36E-06	30773173	C	0.3144	Imputed
rs28653701	3	0.68137115	3.55E-06	178963302	G	0.1183	Imputed
rs190046398	11	-0.9078796	4.42E-06	56679814	C	0.08561	Imputed
rs35835898	15	-0.4372099	5.25E-06	89249994	T	0.4354	Imputed
rs9326285	11	0.4520562	5.35E-06	126665849	C	0.3702	Imputed
chr17:65180936	17	-0.9172645	5.64E-06	65180936	C	0.06116	Imputed
rs59286075	2	-0.9410422	6.02E-06	121122812	G	0.07256	Imputed
rs371022872	8	-0.4997932	7.88E-06	16528596	T	0.2517	Imputed
rs6531822	4	-0.7597502	8.05E-06	86368499	A	0.1241	Imputed
rs10441206	7	-1.1986287	8.24E-06	145193259	C	0.05145	Genotyped
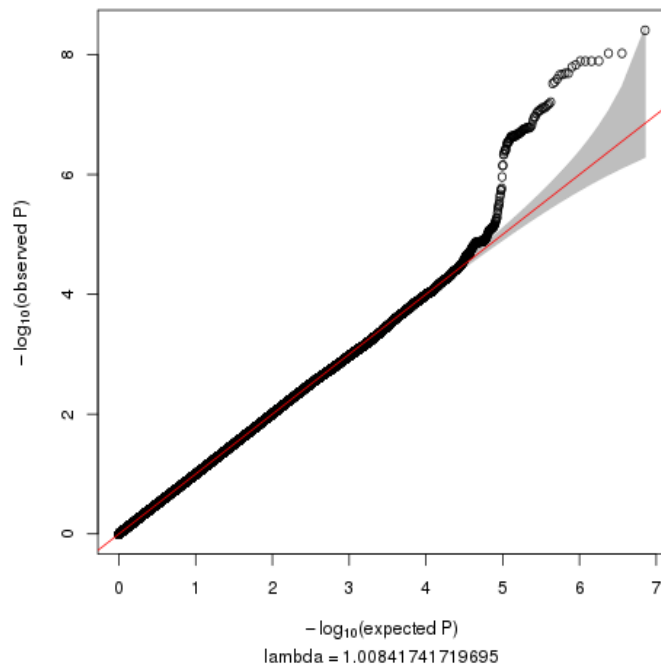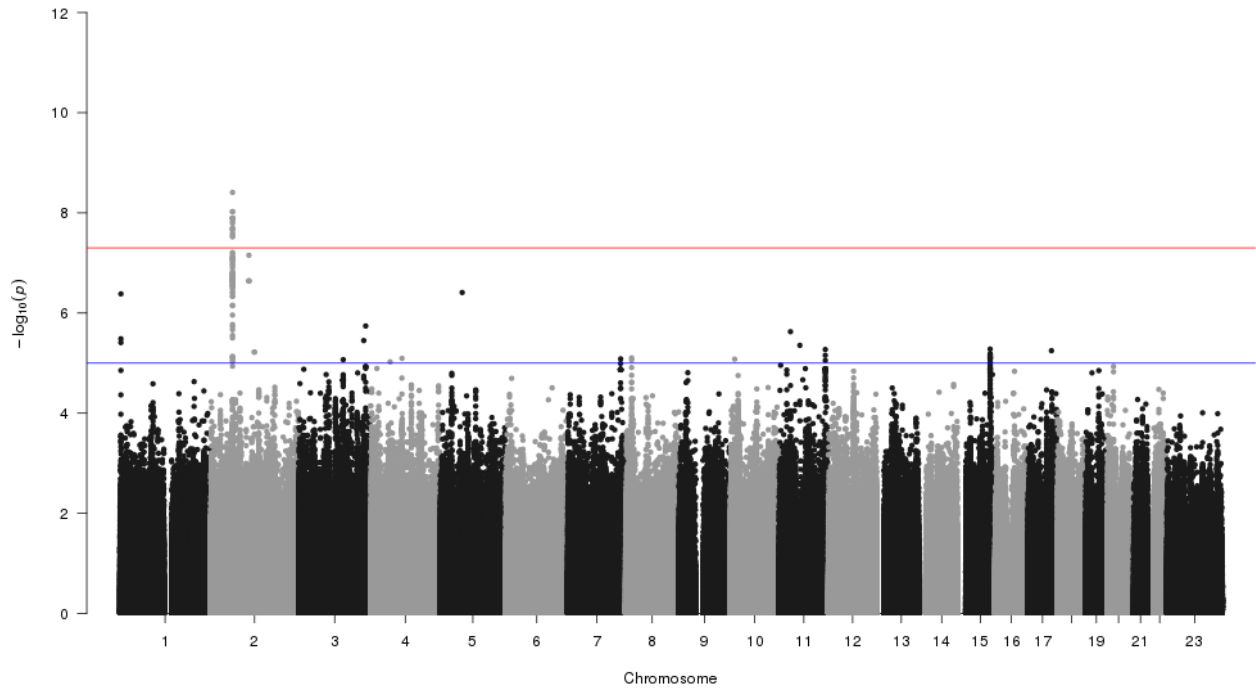
Figure 53 Manhattan and QQ plots for the associated Dental Caries in POFC
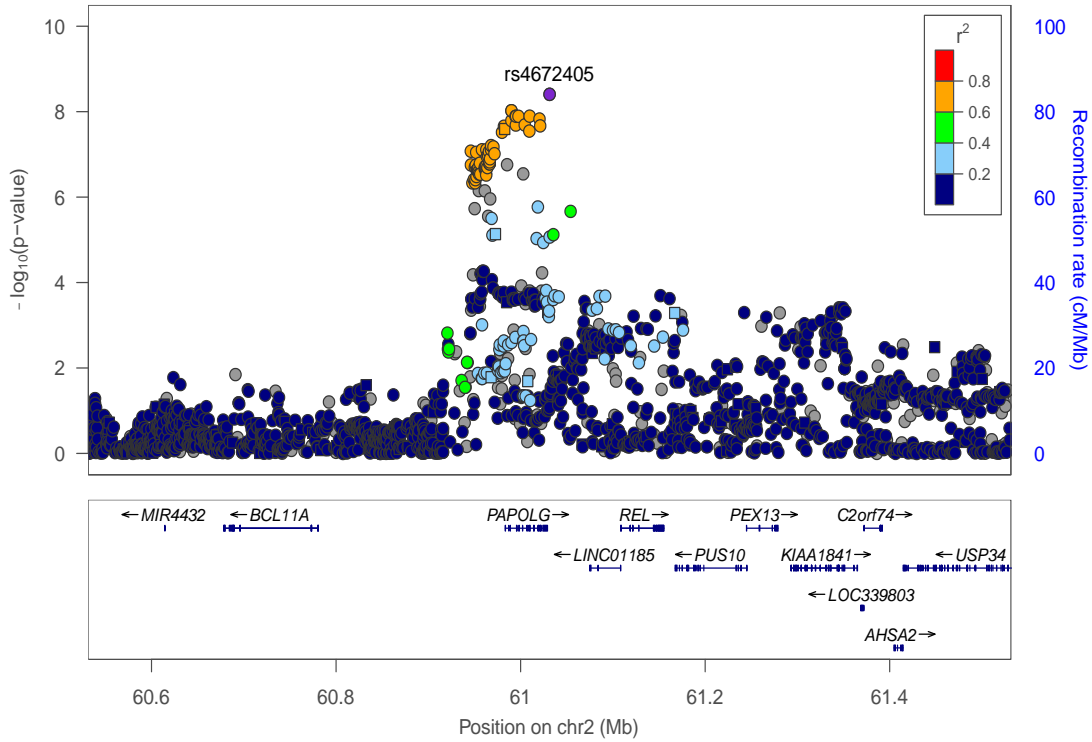
Figure 54 Regional Association Plot of rs4672405 from dental caries GWAS in POFC
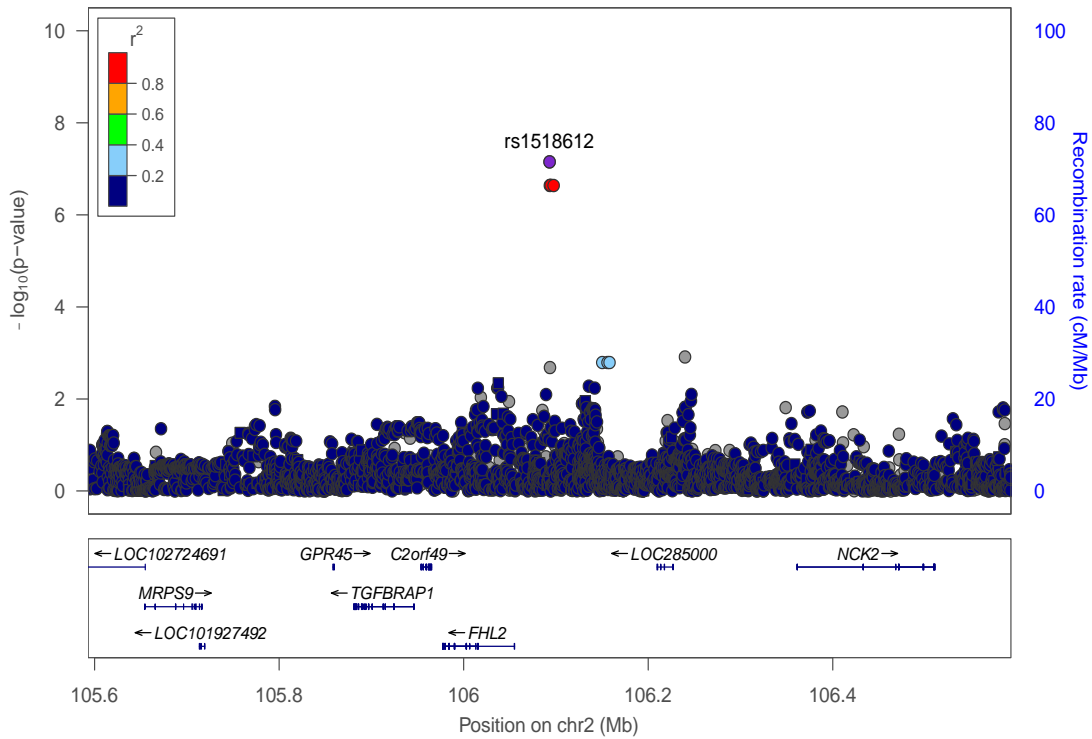
Figure 55 Regional Association Plot of rs1518612 from dental caries GWAS in POFC
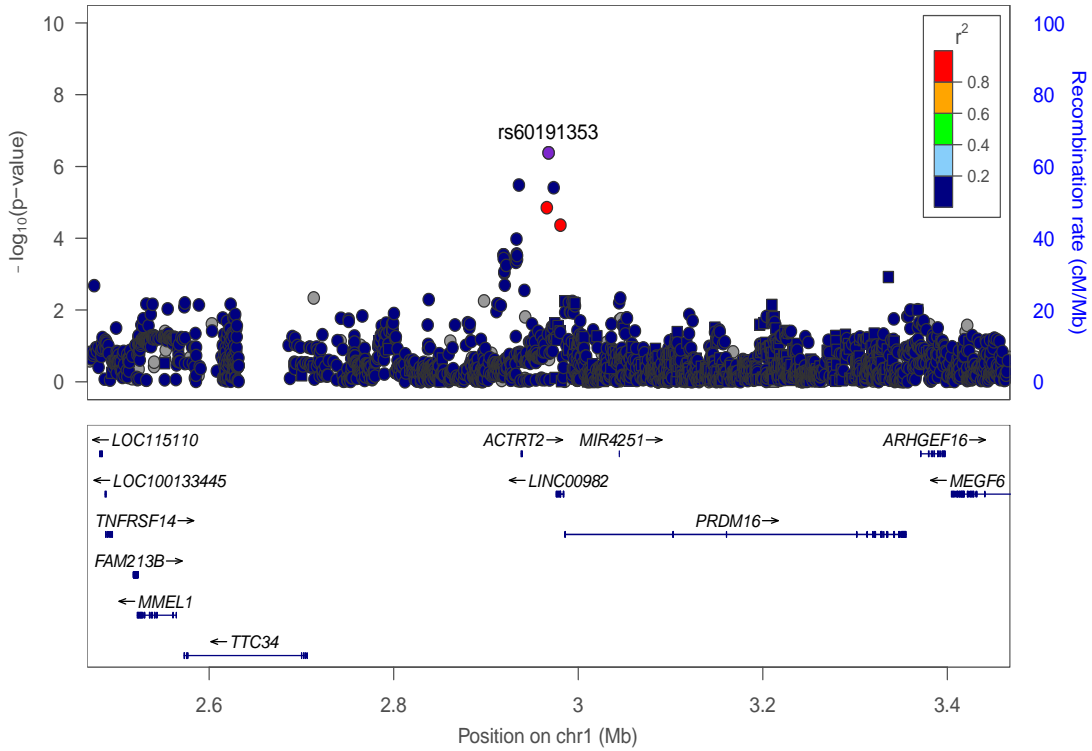
Figure 56 Regional Association Plot of rs60191353 from dental caries GWAS in POFC
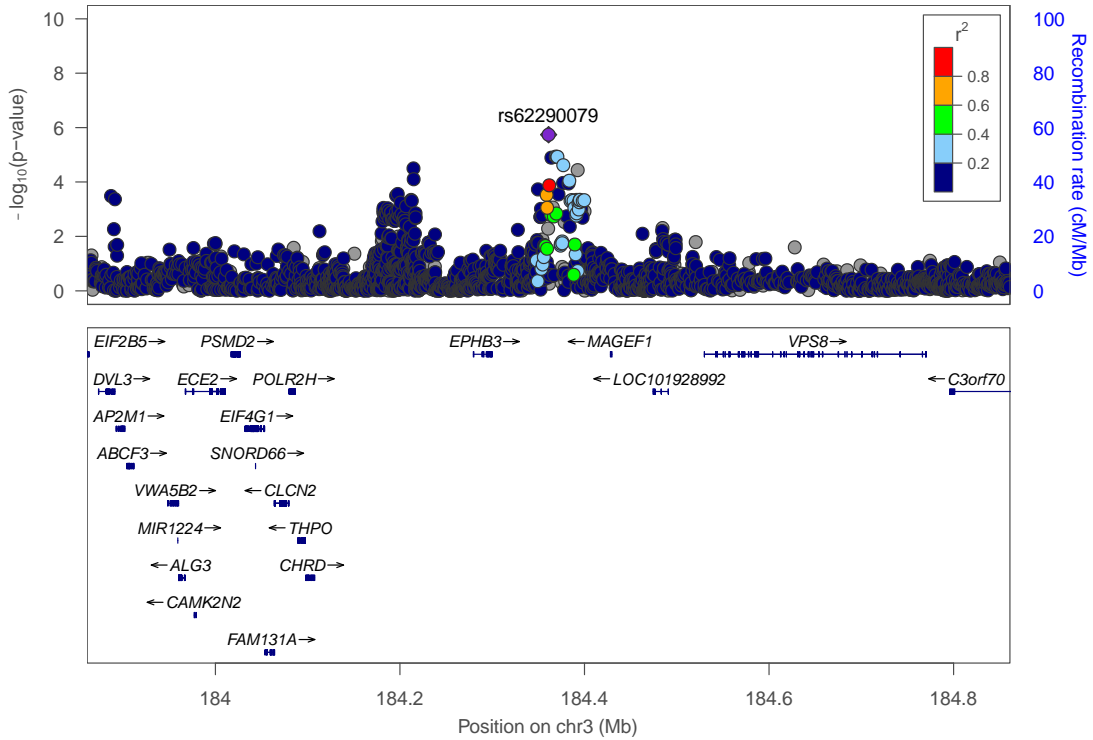
Figure 57 Regional Association Plot of rs62290079 from dental caries GWAS in POFC
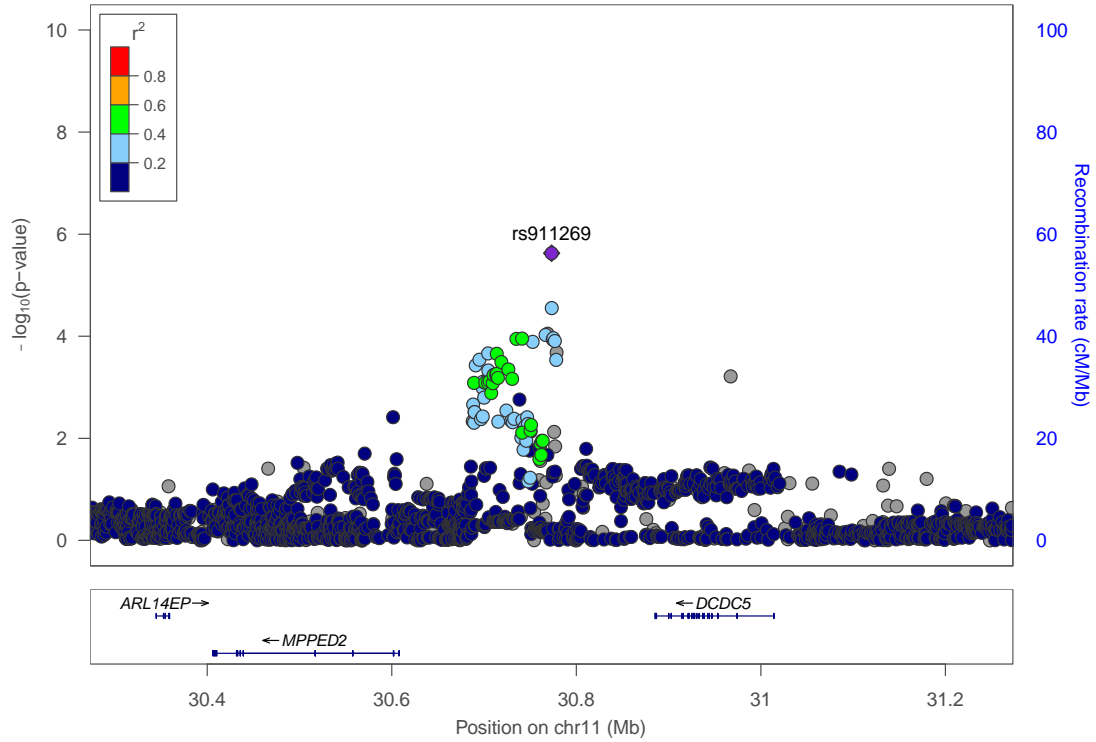
Figure 58 Regional Association Plot of rs911269 from dental caries GWAS in POFC
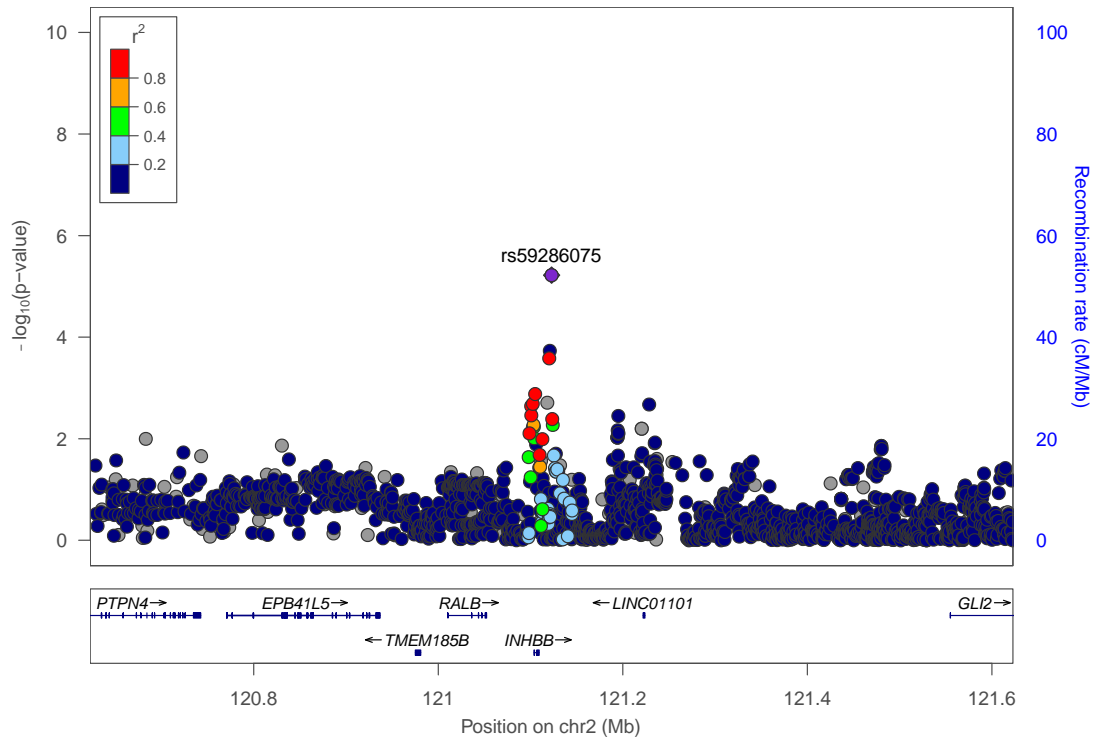
Figure 59 Regional Association Plot of rs59286075 from dental caries GWAS in POFC

5.2.2 COHRA1

A total of 6,300,597 SNPs was available for analysis after quality control. A genome-wide association analysis was implemented in EMMAX using linear regression (see section 5.1.4 for discussion of EMMAX). Manhattan and QQ-plots for the quantitative trait are shown in Figure 60. The genomic control lambda was calculated to be $\lambda = 1.045$. Any loci that included an SNP with a P value $< 1 \times 10^{-5}$ for the association with dental caries were plotted using Locus Zoom to display the linkage disequilibrium. A genome-wide significance threshold of ($P < 5 \times 10^{-8}$), and suggestive level of significance ($P < 1 \times 10^{-5}$) were assigned.

We did not have SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$); however, there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for our top SNPs are shown in Table 30.

Variant rs113242005 (beta = -1.5855, P = 2.71E-06) is one of the suggestive variants that is located near *CRISPLD2* (Figure 61) on chromosome 16, which encode a Protein Coding gene. In mice, expression study showed that *crispld2* was detected in mandible, palate and tooth germs (Chiquet, et al., 2007). Recent studies have suggested that genetic variation in *CRISPLD2* might have a role in the etiology of non-syndromic cleft with or without cleft palate (Letra, et al., 2011; Chiquet, et al., 2007). However, we do not have evidence that *CRISPLD2* have a role, directedly or indirectly, in tooth development or in dental caries. A suggestive signal was detected at rs72832935 (beta = 0.91147223, P = 7.60E-06) that is located near to *NEDD9* (Figure 62) on chromosome 6, which encodes a protein that plays an important role in regulating signaling complexes, neuronal differentiation, development and migration, as well as neural crest cell migration (Riccomagno, et al., 2014; Aquino, et al., 2009). In a large-scale consortium genome-

wide association study (GWAS) of dental caries for permanent teeth, an association between *NEDD9* and dental caries was detected (Haworth, et al., 2018); one of the cohorts that was used in that GWAS are from COHRA1. These findings strongly suggest that *NEDD9* could influence dental caries.

Another suggestive signal was detected on chromosome 4 at rs113059858 (beta =1.37167122, *P* = 8.03E-06) that located near *KDR* (Figure 63), which encodes one of the vascular endothelial growth factor (VEGF) receptors, and has an important role in mediating VEGF-induced endothelial proliferation, survival, migration, as well as tubular morphogenesis (Ferrara, et al., 2004). Both *VEGF* and *KDR* were found expressed in the dental pulp in the primary and permanent teeth (Mattuella, et al., 2007; Liu, et al., 2014) and these findings suggest that *KDR* could play a role during tooth development although it is still unclear and need to be investigated.

Table 30 Top GWAS hits for the associated Dental Caries in COHRA1

SNP3	CHR	BETA	*P*	BP	A2	MAF	Type
chr10:42739014	10	1.96547162	9.51E-07	42739014	G	0.06011	Imputed
rs2224554	9	-1.5487223	1.90E-06	13798290	G	0.09365	Imputed
rs72726727	9	-1.2906272	1.93E-06	30266548	T	0.09779	Imputed
rs113242005	16	-1.5855268	2.71E-06	85363165	G	0.05875	Imputed
rs77720033	7	-0.96007	2.85E-06	152770650	A	0.1821	Imputed
rs58977011	1	-2.3973391	3.19E-06	121387162	A	0.07107	Imputed
rs11045273	12	0.89062427	3.37E-06	20629112	G	0.2056	Imputed
rs10869748	9	-1.3038449	3.90E-06	71643486	C	0.09147	Imputed
rs67373295	2	-0.8741234	4.56E-06	239494101	C	0.2077	Imputed
rs72948765	18	1.09608349	5.87E-06	63198881	A	0.2173	Imputed
rs959167	15	1.06907753	6.44E-06	93978414	A	0.1221	Genotyped
rs72832935	6	0.91147223	7.60E-06	11397324	T	0.1792	Imputed
rs113059858	4	1.37167122	8.03E-06	55499439	C	0.06596	Imputed
rs8137223	22	-1.375967	8.56E-06	21319402	G	0.1349	Imputed
rs7074361	10	-0.6996466	8.80E-06	129024651	C	0.4341	Imputed
chr3:166737275	3	0.85368779	8.97E-06	166737275	G	0.2091	Imputed
rs8098843	18	-0.790188	9.32E-06	53089162	C	0.259	Imputed
rs3758709	11	-0.7798662	9.37E-06	76878915	C	0.2778	Imputed
rs4742727	9	-0.6901303	9.80E-06	101169375	G	0.4053	Genotyped
rs4094705	4	1.46372339	1.00E-05	49256887	T	0.2463	Imputed
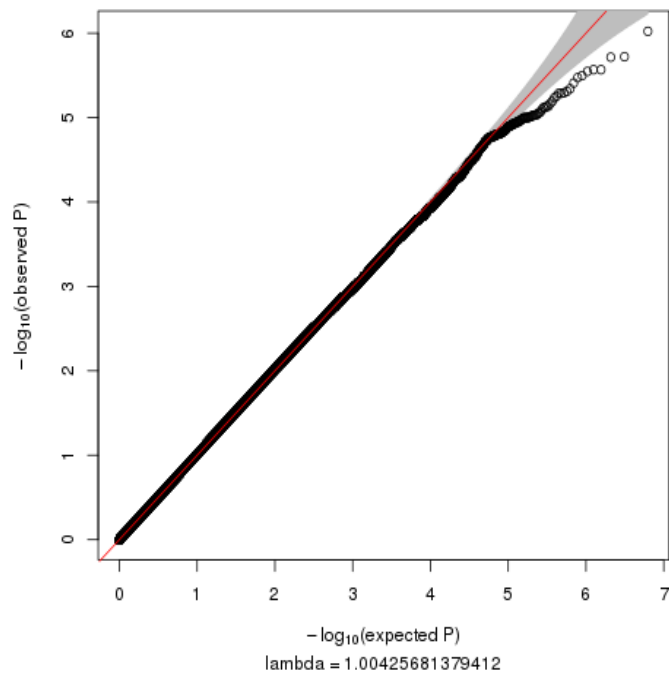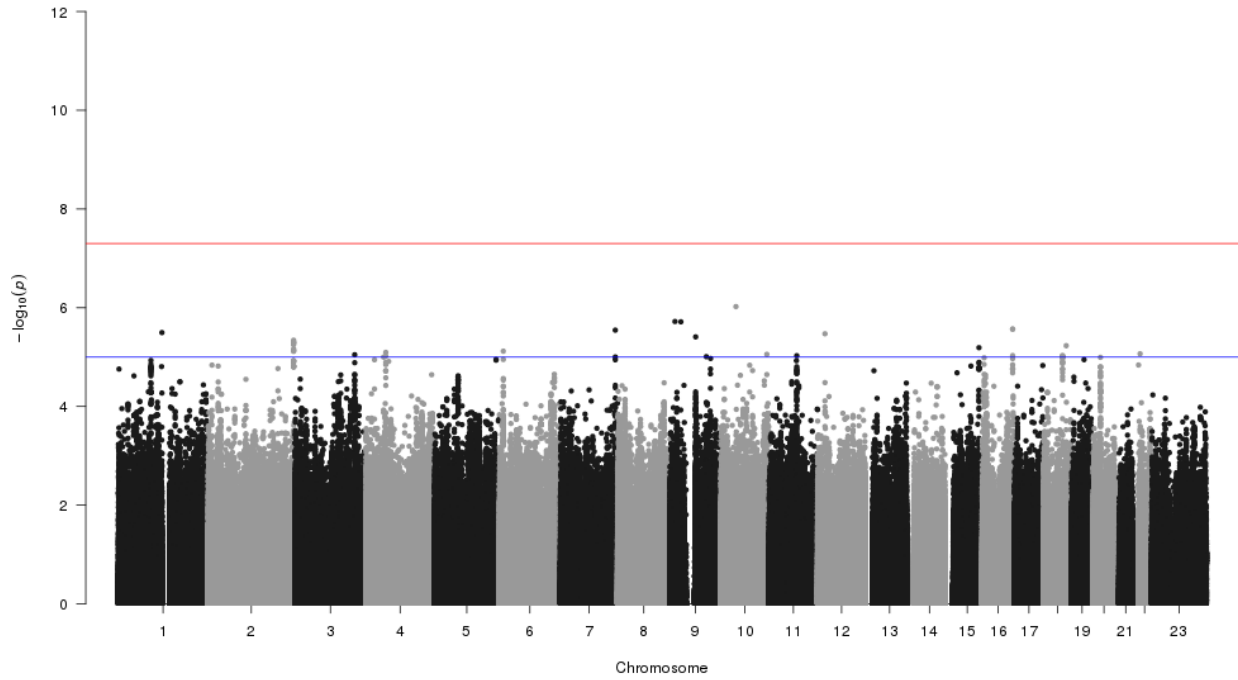
Figure 60 Manhattan and QQ plots for the associated Dental Caries in COHRA1
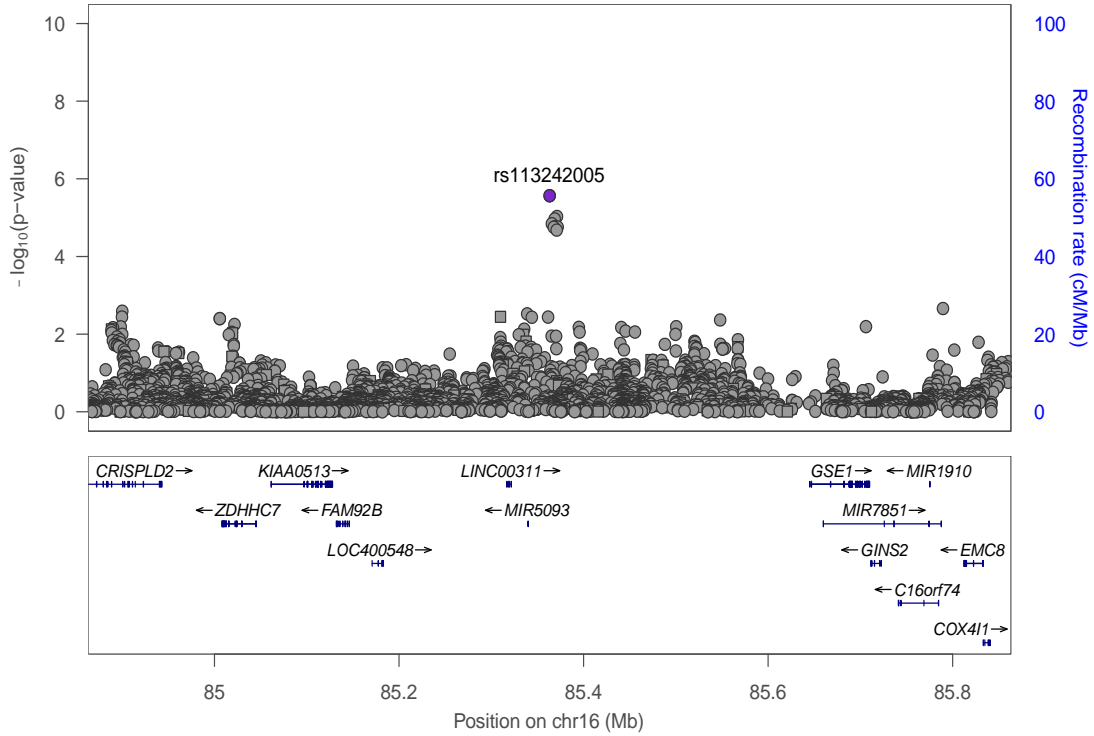
Figure 61 Regional Association Plot of rs113242005 from dental caries GWAS in COHRA1
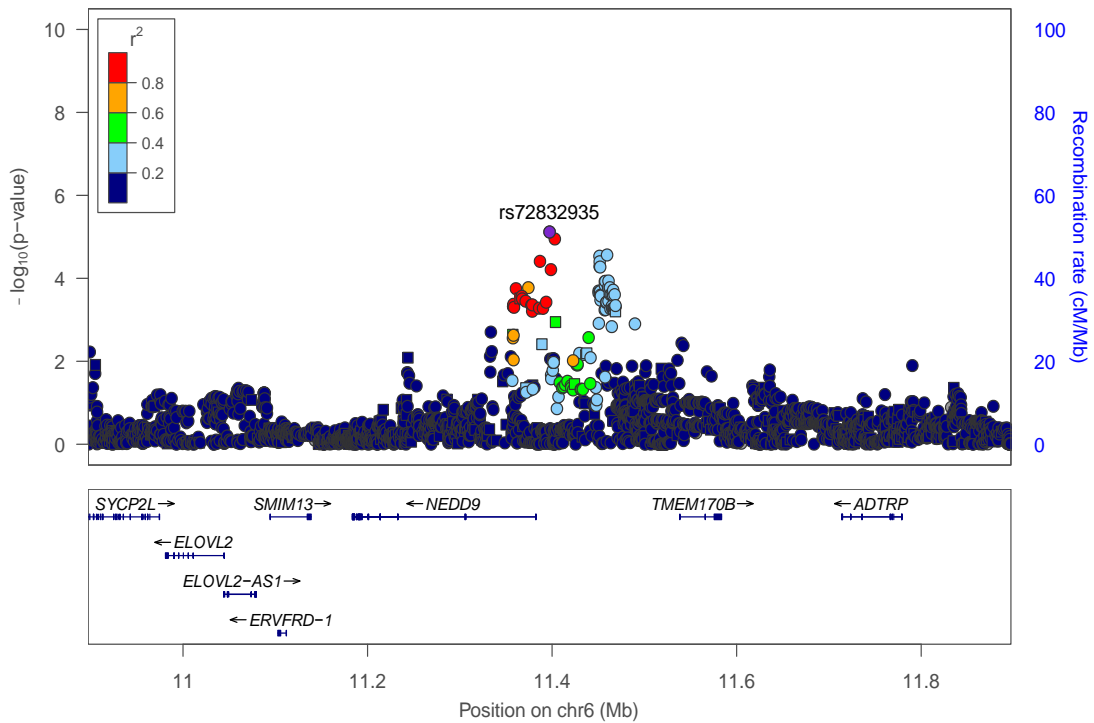
Figure 62 Regional Association Plot of rs72832935 from dental caries GWAS in COHRA1

148
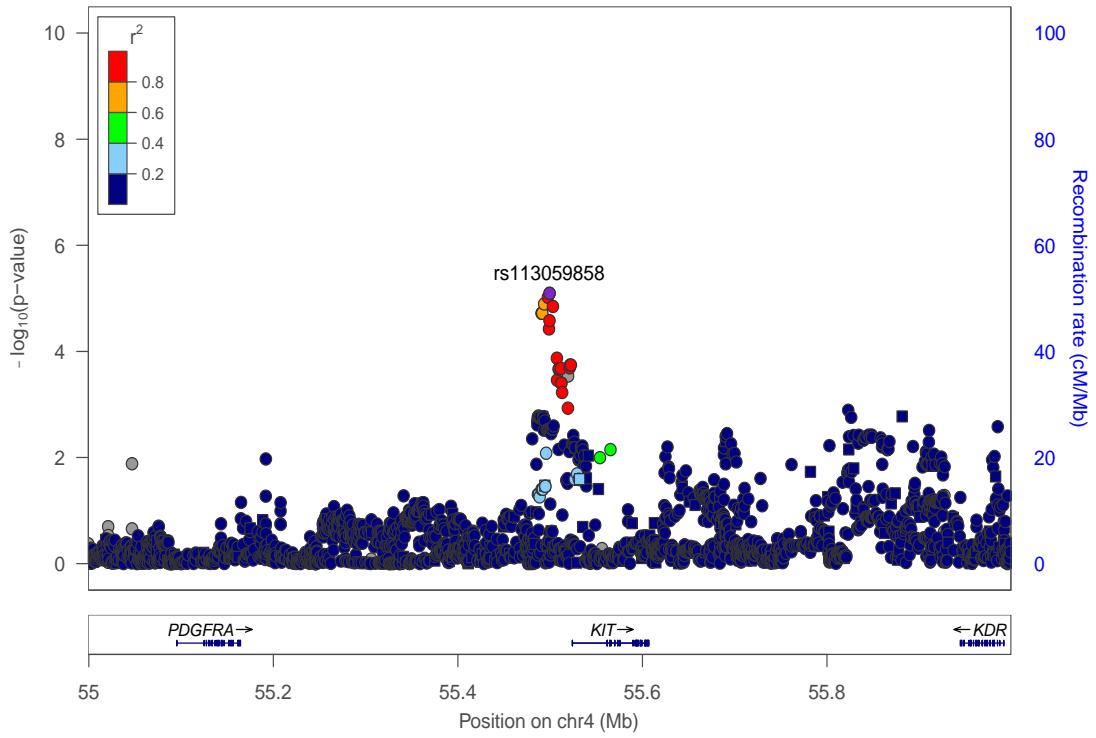
Figure 63 Regional Association Plot of rs113059858 from dental caries GWAS in COHRA1

5.3 META-ANALYSIS RESULTS

To assess and to further investigate the association between genetic variation and dental caries across multiple ethnic populations, a meta-analysis was conducted with results obtained from the two studies, POFC and COHRA1, which included subjects with different ethnic backgrounds. The Meta-analysis was performed using Stouffer's p-value-based meta-analysis as implemented in METAL software (Willer, et al., 2010) which combined the p-values across studies, while considering both of the sample size and the direction of effect into account. One of the major advantages of Meta-analysis is in the little or no loss of efficiency. In this study, we only focused on variants with known direction, either – or +, in both cohorts (POFC & COHRA1). If the variants were missing in one cohort, i.e. their direction were known in one cohort but were unknown in the other cohort, they were excluded.

Manhattan and QQ-plots for the total DFT+dft quantitative trait are shown in Figure 64. The genomic control lambda was calculated to be $\lambda = 1.014$. Any loci that included a SNP with a P value $< 1 \times 10^{-5}$ for the association with dental caries were plotted using Locus Zoom to display the linkage disequilibrium. A genome-wide significance threshold of ($P < 5 \times 10^{-8}$), and suggestive level of significance ($P < 1 \times 10^{-5}$) were assigned.

We did not have SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$); however, there were several regions of the genome that showed suggestive significance ($P < 10^{-5}$). Association results for our top SNPs are shown in Table 31.

Variant rs9386450 ($P = 1.68E-06$) is one of the suggestive variants that is located near *LIN28B* (Figure 65), and this gene is expressed in many developing tissues and involved in the biosynthesis of the microRNA let-7 family and have an important role in embryogenesis

processes. In mouse model, *LIN28B* found to be expressed in dental epithelium, tooth germ, ameloblast and odontoblasts during the different stages of tooth development (Dong, et al., 2017). These findings suggest that *LIN28B* may have a role in odontogenesis and dental caries, however further investigation, in both animal and human, is needed to confirm their function in tooth development.

Variant rs2087781 ($P = 2.07E-06$) is one of the suggestive variants that is located near *ARHGAP21* (Figure 66). Encoded protein of *arhgap21* gene act as a GTPase-activating protein (GAP) for CDC42, which is been suggested that it is required for proper odontogenesis and amelogenesis (Dubois, et al., 2005; Tian, et al., 2018). In a sequencing study in Italy, they found that a rare variant in *ARHGAP21* is associated with malocclusion (Perillo, et al., 2015). These findings suggest that *ARHGAP21* have a role in odontogenesis and further investigation, is needed to figure out their role in dental caries process.

Variant rs1518612 ($P = 2.45E-06$) is located on chromosome 2 near *FHL2* gene (Figure 67), which has been found to be expressed in odontoblasts and odontoblast-like cells in mature human teeth in both normal and pathological conditions (Du, et al., 2012) and it mediate mesenchyme cell differentiation by interacting with *Runx2* during early stages of tooth development and human dental pulp cell differentiation (Du, et al., 2016), which strongly suggest that *FHL2* might have a role during tooth development and might play a role in dental caries.

Variant rs62290079 ($P = 5.265E-06$) on chromosome 3 is located near *EPHB3* gene (Figure 68), and the encoded protein belongs to a family of cell surface receptors with several roles in regulating different aspects of embryonic development, including palatogenesis. Mouse models showed mutations in this locus result in craniofacial defects, including cleft palate (Risley, et al.,

2009). More studies are needed to investigate the potential role of *EPHB3* in tooth development

and dental caries.

Table 31 Top GWAS hits for meta-analysis of dental caries in POFC and COHRA1

SNP	*P*	Direction	CHR	BP	MAF_OFC	MAF_COHRA	Type
rs9386450	1.68E-06	++	6	105996852	0.2868	0.1747	Imputed
rs2087781	2.07E-06	--	10	25372340	0.4495	0.3435	Imputed
rs1518612	2.45E-06	--	2	106093166	0.08486	0.1396	Imputed
rs2224554	2.48E-06	--	9	13798290	0.07104	0.09365	Imputed
rs179145	4.55E-06	++	14	95983975	0.404	0.3722	Imputed
rs6645396	4.62E-06	--	23	116933059	0.2024	0.1692	Imputed
rs71599734	4.85E-06	++	5	163759402	0.05842	0.08153	Imputed
rs7590283	5.25E-06	--	2	216839840	0.3694	0.4868	Imputed
rs62290079	5.26E-06	--	3	184361139	0.1405	0.2412	Imputed
rs112770207	6.66E-06	--	17	79172119	0.07626	0.1064	Imputed
rs6973065	7.53E-06	--	7	93335949	0.3922	0.3638	Imputed
rs10064994	8.92E-06	++	5	21614817	0.1814	0.1765	Imputed
rs9592635	9.11E-06	++	13	69983244	0.06084	0.1027	Imputed
rs12575749	9.19E-06	++	11	36383440	0.1894	0.2913	Imputed
rs11619822	9.48E-06	++	13	111026454	0.09096	0.1183	Imputed
rs10810974	9.62E-06	++	9	18512570	0.4045	0.3883	Genotyped
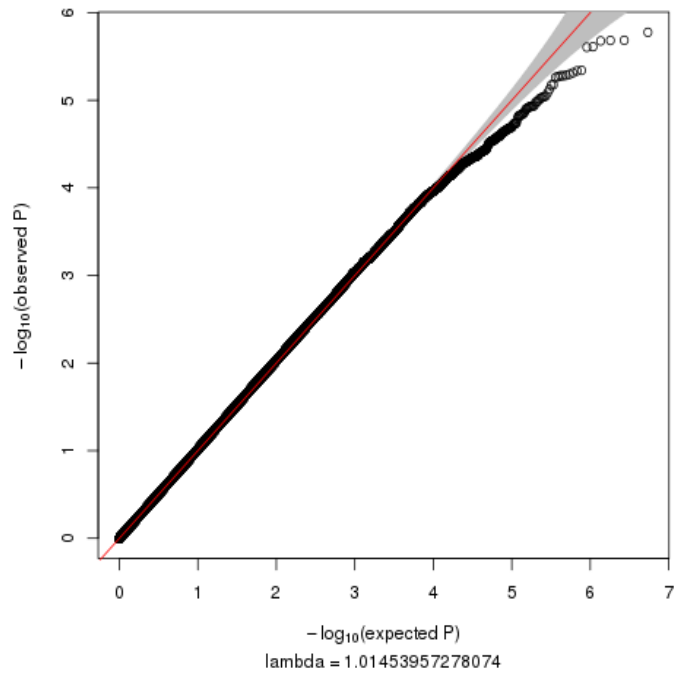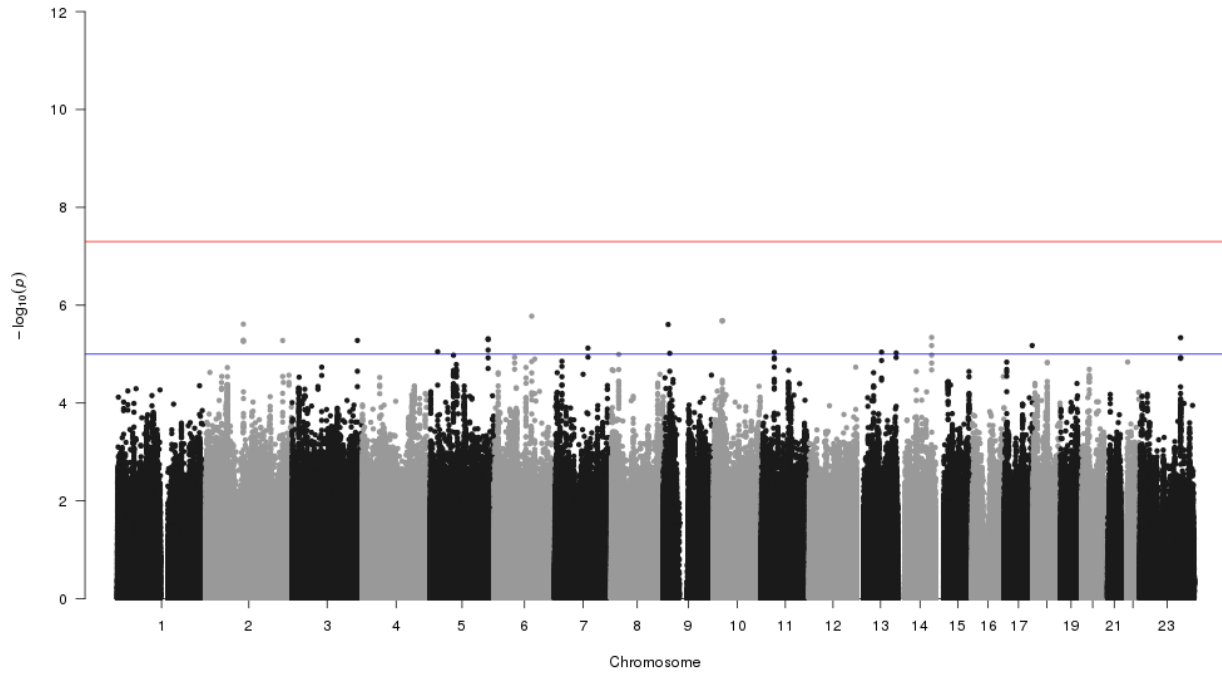
Figure 64 Manhattan and QQ plots for Dental Caries in Meta-analysis
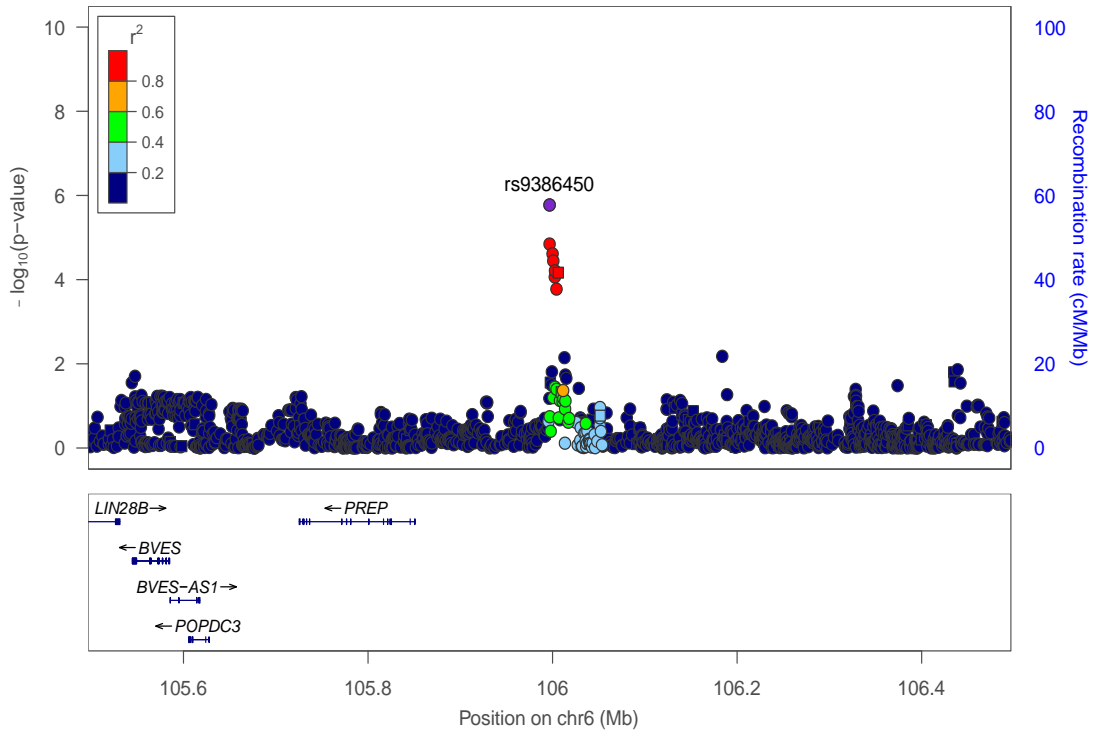
Figure 65 Regional Association Plot of rs9386450 from the Meta-analysis of dental caries
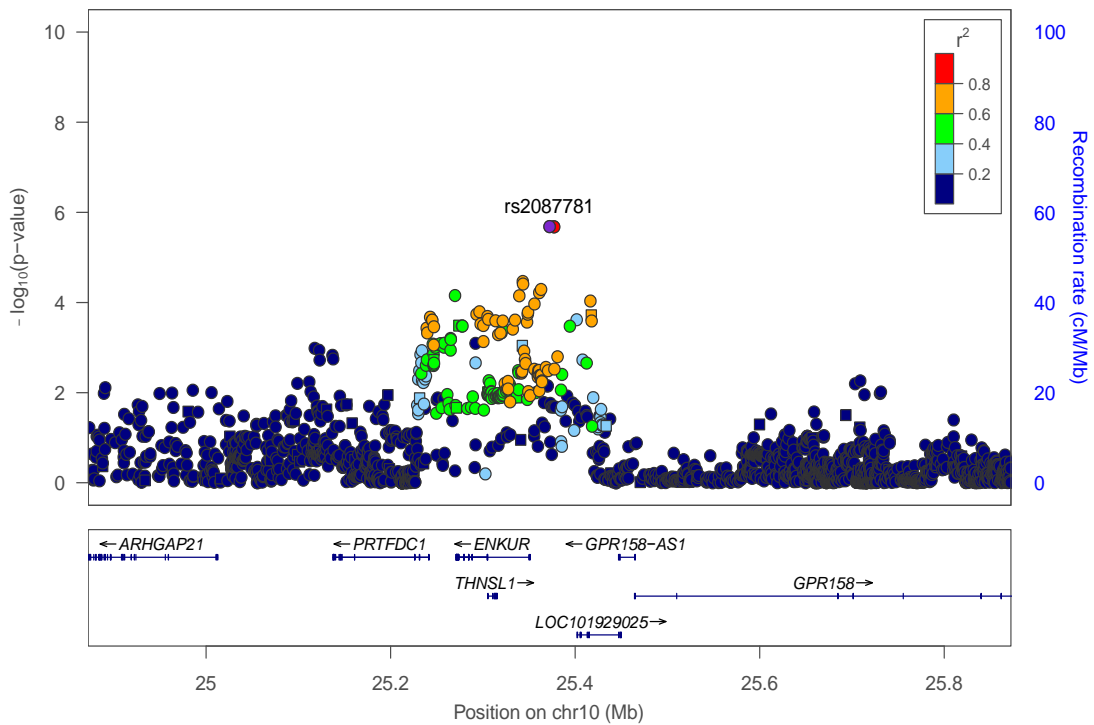
Figure 66 Regional Association Plot of rs2087781 from the Meta-analysis of dental caries

154
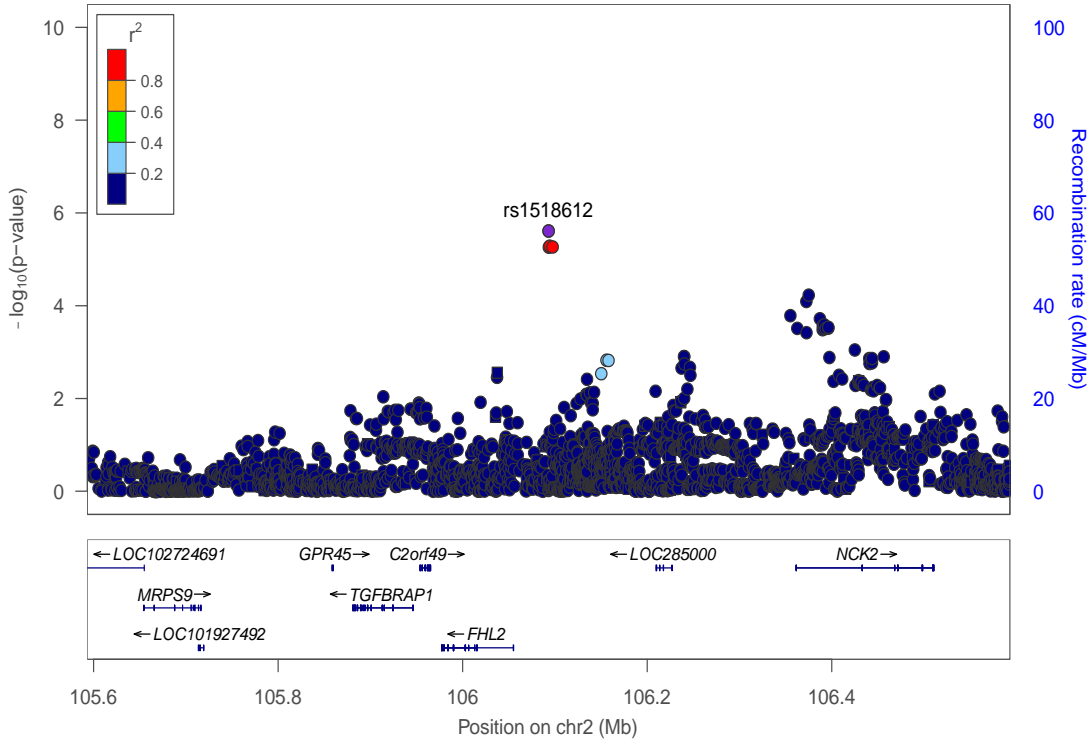
Figure 67 Regional Association Plot of rs1518612 from the Meta-analysis of dental caries
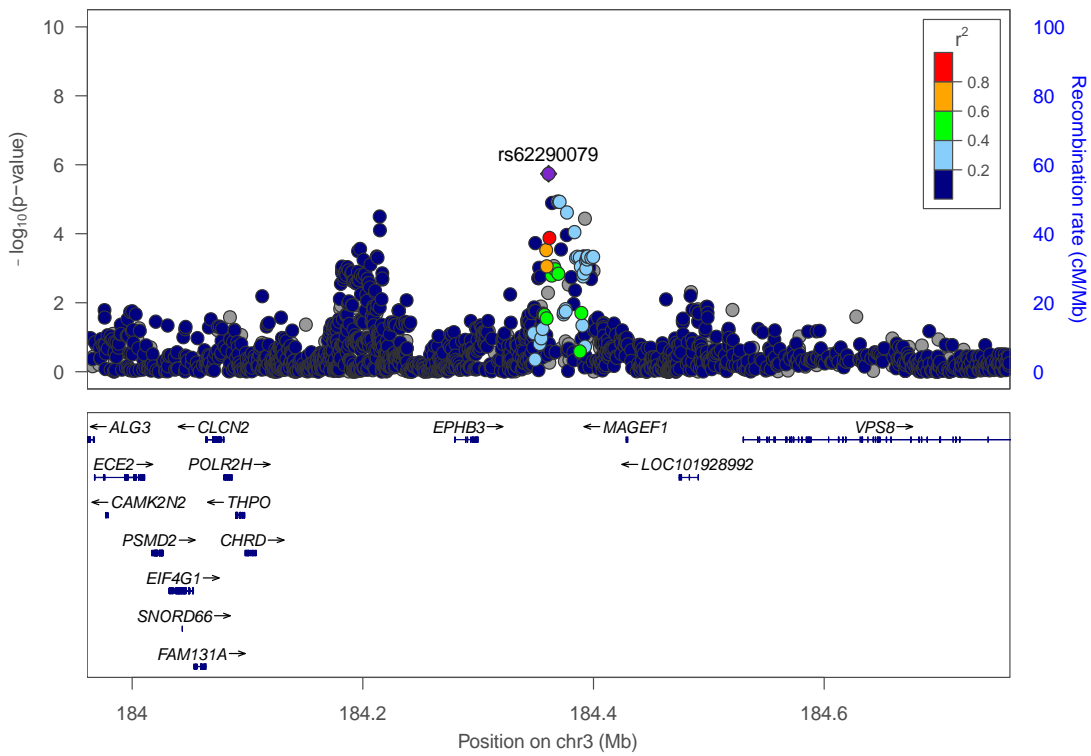
Figure 68 Regional Association Plot of rs62290079 from the Meta-analysis of dental caries

We searched for the identified genes from previous GWASs of dental caries in our dental caries meta GWAS results to see if they had been replicated (Table 32). Only limited evidence of suggestive associations was detected near those identified genes. Regions of the genome that showed suggestive significance with dental caries in our results were further visualized using Regional Association (LocusZoom) plots (Figure 69).

Table 32 Association results of identified dental caries genes in dental caries meta-analysis

SNP	CHR	BP	*P*	GENES	Type
rs35354317	1	236644901	4.94E−4	*ACTN2, EDARADD, MTR*	Imputed
rs10835721	11	30773173	6.85E−4	*MPPED2*	Imputed
rs60957147	17	56507204	1.31E−3	*LPO*	Imputed
rs3799639	6	166954433	1.14E−3	*RPS6KA2*	Imputed
rs10037522	5	50635643	8.72E−3	*ISL1*	Imputed
rs6853866	4	154581208	3.06E−3	*TLR2*	Imputed
rs517769	1	229176190	1.9E−3	*RHOU*	Imputed
rs140461609	10	30890978	3.25E−4	*LYZL2*	Imputed
rs10223932	7	90579518	5.52E−4	*FZD1*	Imputed
rs35838266	8	27520526	2.12E−3	*PTK2B*	Imputed
Chr1:4356392	1	4356392	5.09E−4	*AJAP1*	Imputed
rs56361560	4	89118687	9.69E−4	*ABCG2, PKD2*	Imputed
rs36005247	4	148586276	3.91E−4	*EDNRA*	Imputed
rs34974591	10	101523073	1.47E−3	*NKX2−3*	Imputed
rs12866627	13	21018715	9.64E−4	*IFT88, IL17D*	Imputed
rs9965664	18	9759074	3.53E−3	*TWSG1*	Imputed
rs9966529	18	46487501	9.2E−4	*SMAD7*	Imputed
rs66798958	7	41362241	2.44E−3	*INHBA*	Imputed
chr2:219320233	2	219320233	3.56E−4	*CXCR1, CXCR2*	Imputed
rs1447624	3	160621070	9.59E−5	*KPNA4*	Imputed
rs11574947	16	30524676	1.26E−3	*ITGAL*	Imputed
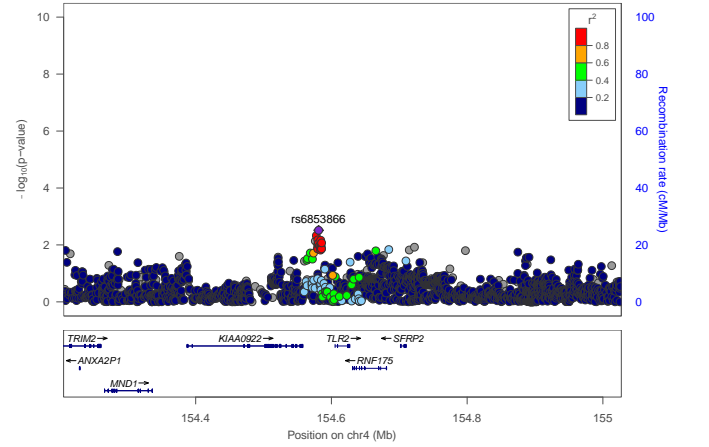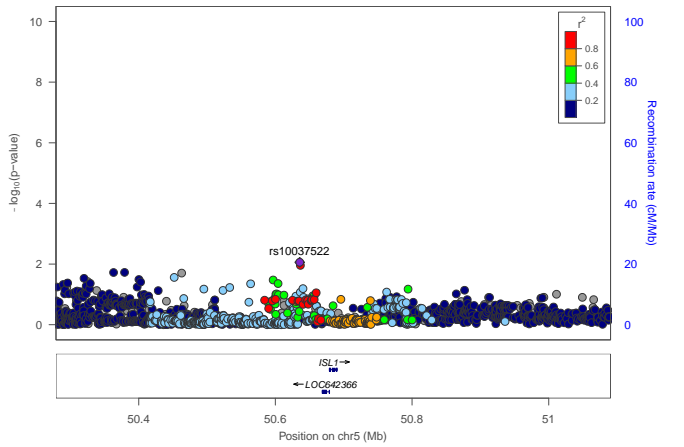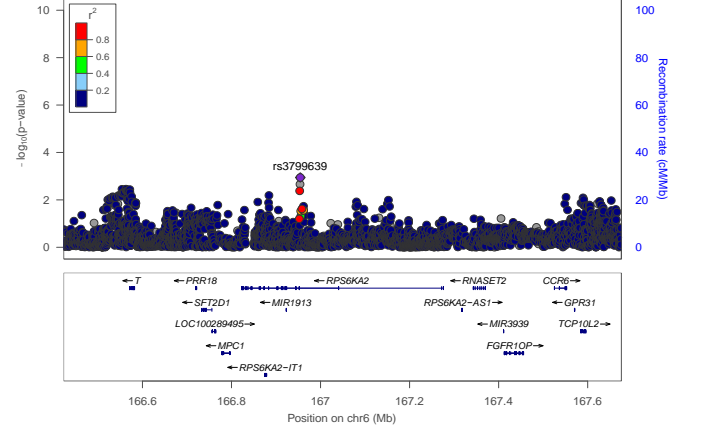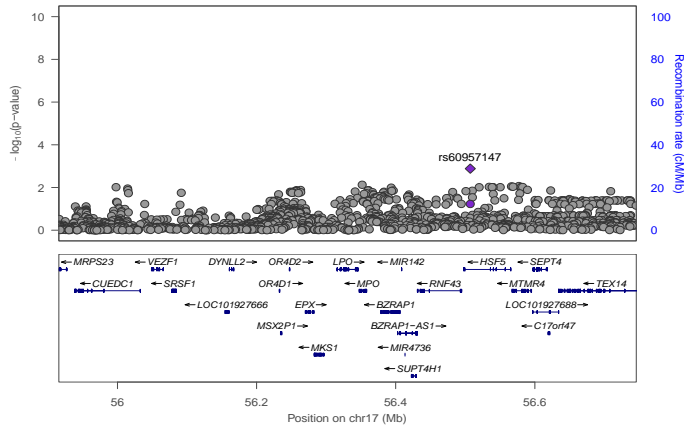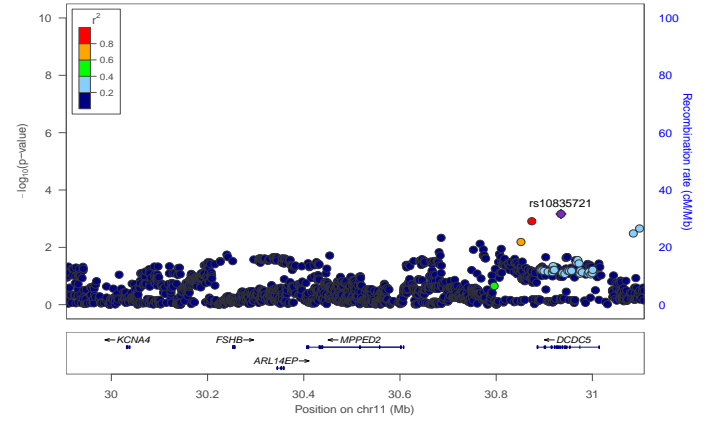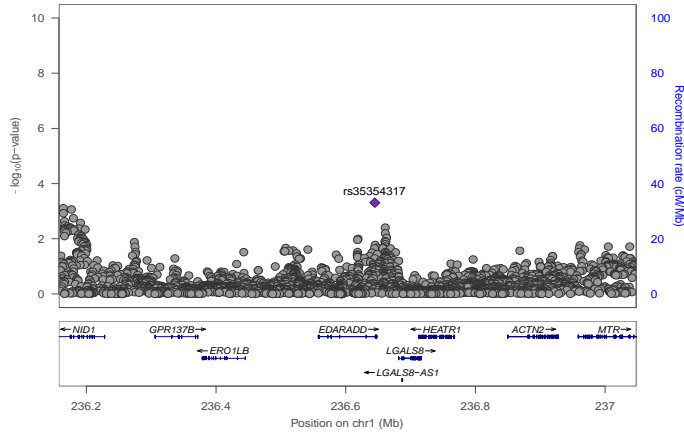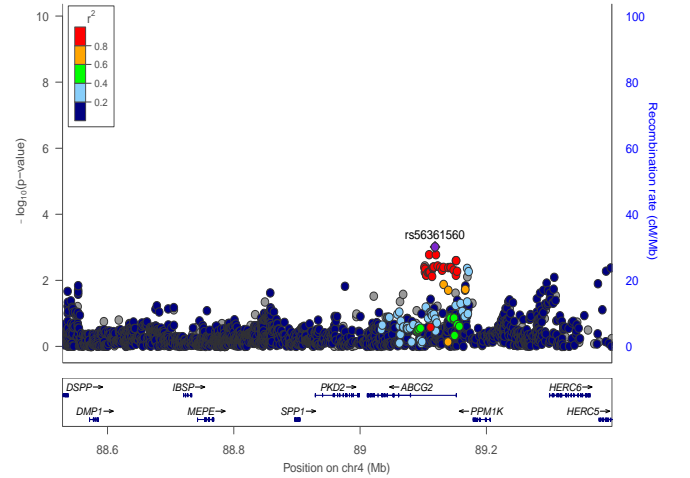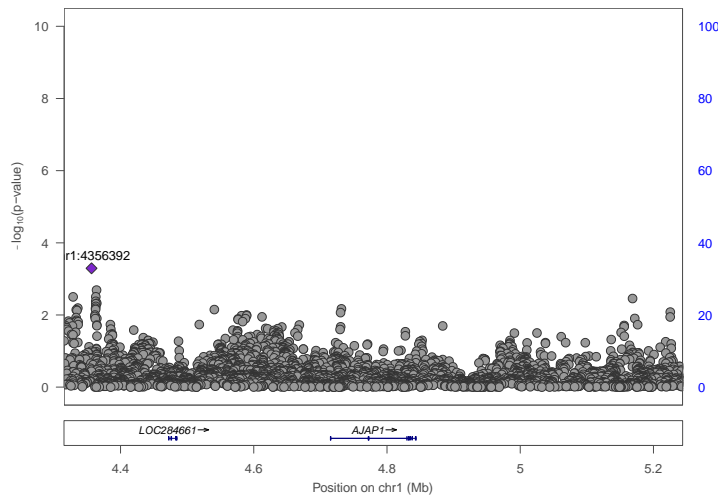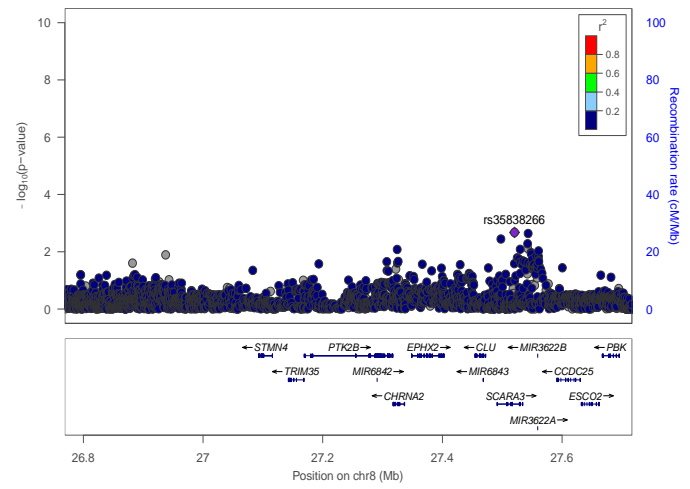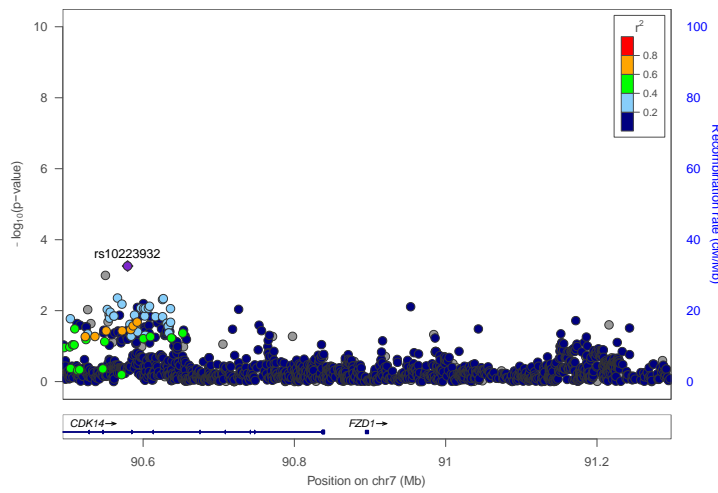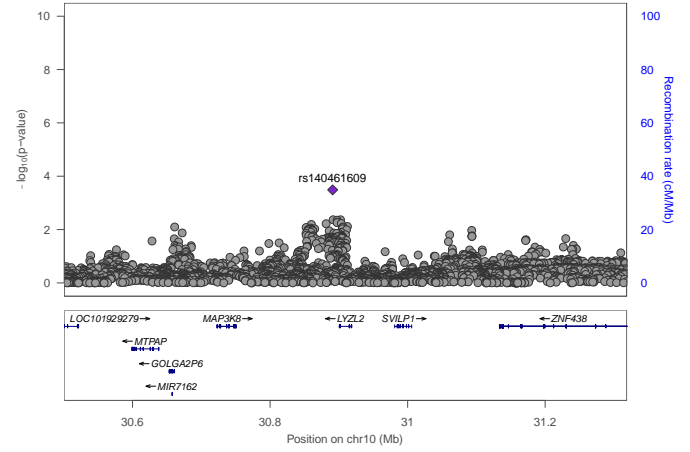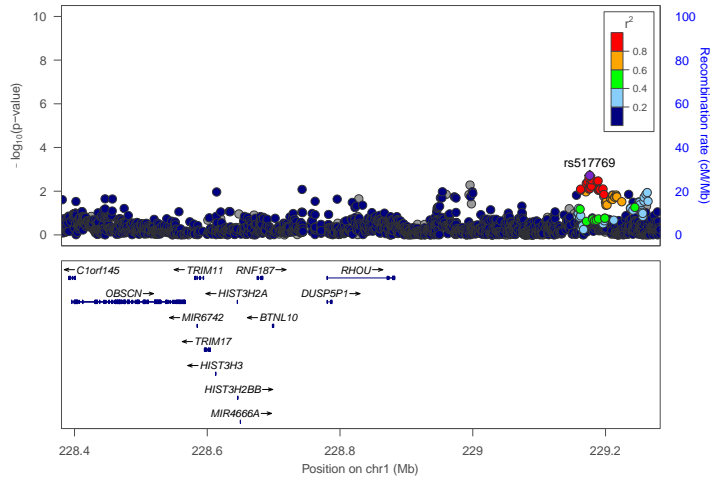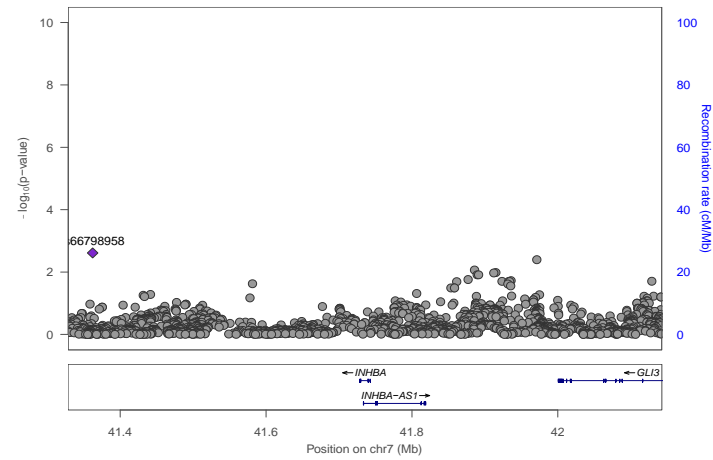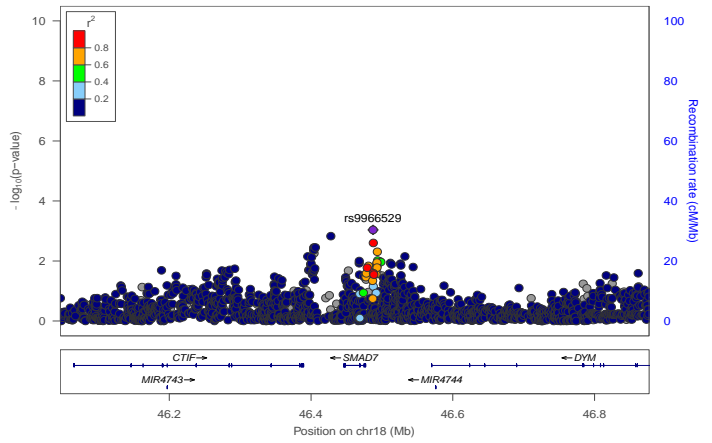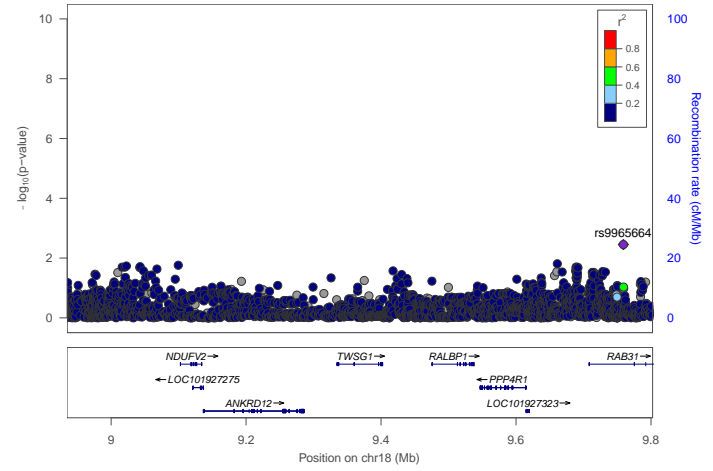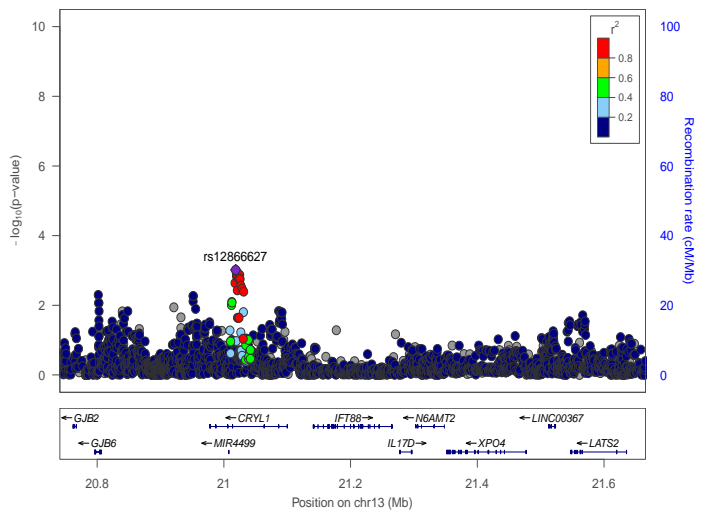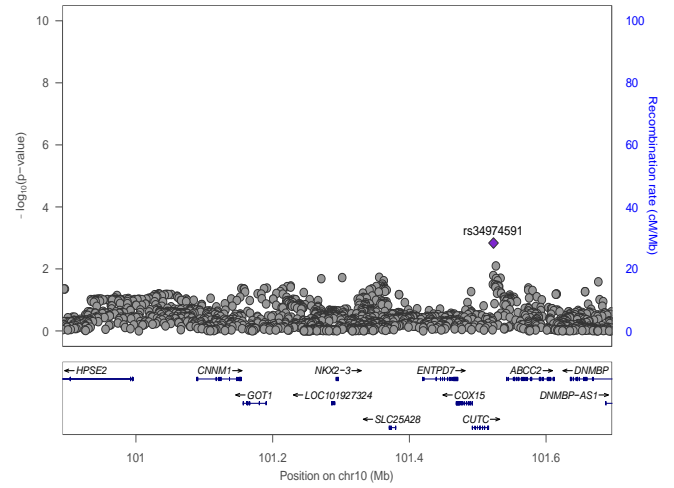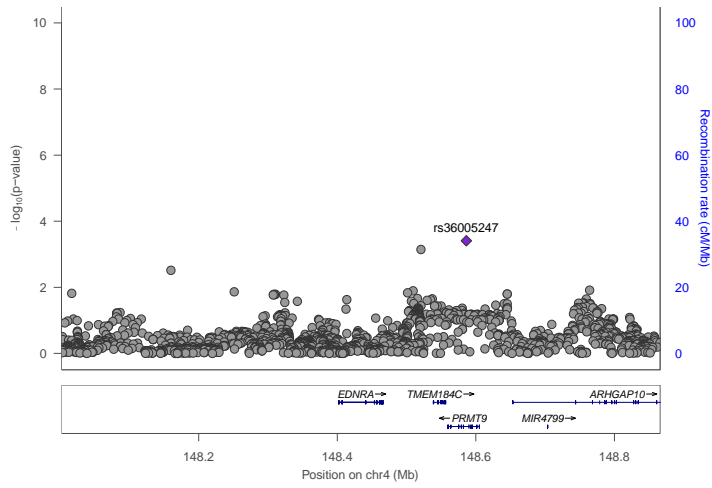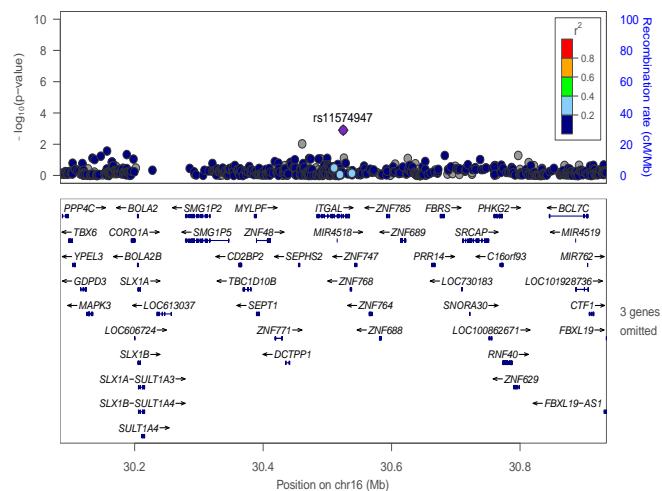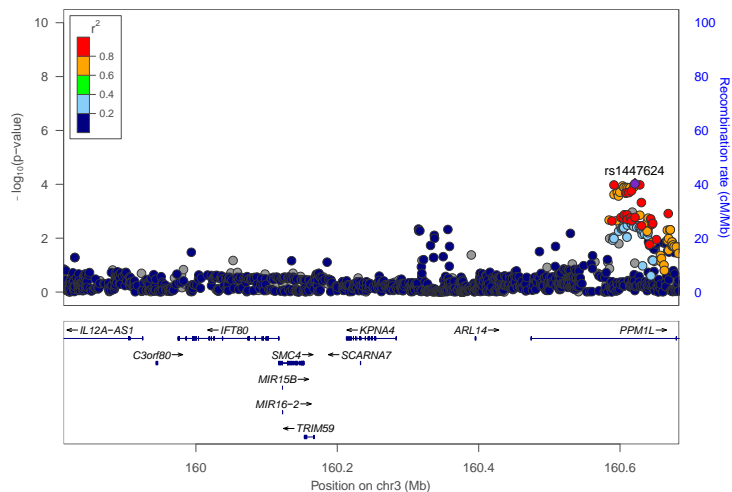
Figure 69 Regional Association Plots of dental caries genes in Meta-analysis GWAS results

5.4 DISCUSSION

This study is one of the largest GWAS that aimed to identify possible genetic loci that may be involved in increasing the risk of developing dental caries in multi-ethnic population. A few of association signals that reached the genome-wide significance level were near genes that have a role, directly or indirectly, in tooth development and possibly in cariology. Regional Association plots of SNPs that reaching suggestive level near genes with unknown roles in dental and/or oral health are in the appendix Figure A23 through Figure A41.

One of our strongest signals was detected in the POFC sample near rs4672405 ($P = 3.91E-09$) and lies near the *REL* gene. Although *REL* plays roles in multiple biological processes such as inflammation, as well as the immune response (Jontell, et al., 1998) and it has been shown that *REL* could control the development of the epidermis and associated appendages, including teeth during embryogenesis in mice (Gugasyan, et al., 2004), however its relevance to dental development and/or oral health is still unclear. *REL* was also one of the genes that also been identified near risk variants in the DAH pattern (see section 4.2.4 It is plausible from these evidences and our results to suggest that *REL* might have a role during tooth development, however, further investigations are needed to confirm its role.

Most of the suggestive variants that we had identified in this study were annotated to help generate hypotheses and nominate these variants for further investigations. We had number of suggestive variants that have been identified before in previous studies as relevant to the tooth development process and/or dental caries, such as *MPPED2, GLI2* in POFC, *and NEDD9* in COHRA. *MPPED2* has been one of the results of GWAS of dental caries in POFC and it has been identified in association with dental caries previously (Sheffer et al.,2011; Stanley, et al., 2014), and *NEDD9* has also been associated with dental caries in a large-scale consortium

162

genome-wide association study of dental caries for permanent teeth (Haworth, et al., 2018). *GLI2* in on the other hand has been investigated before and it was found that it has a function as a mediator for the Sonic hedgehog (Shh) signaling pathway, which plays a regulatory role in several processes during embryogenesis, including tooth development (Mo, et al., 1997). In addition, *Gli2* were found to be expressed during early stages of murine tooth development and mutant mice were found to have abnormal development of maxillary incisors (Hardcastle, et al., 1998), and these biological evidences strongly indicate that the *GLI2* have a role in tooth development and could possibly influence dental health. Another suggestive association signal that was detected in POFC and confirmed in the meta-analysis results was near *FHL2,* which has a role in mediating mesenchyme cell differentiation by interacting with *Runx2* during early stages of tooth development and human dental pulp cell differentiation (Du, et al., 2016).

In this study we also been able to identify suggestive variants near genes that has been suggested before to have roles in the palate formation, such as *PRDM16* (Bjork, et al., 2010), *EPHB3* (Risley, et al., 2009), and *CRISPLD2* (Chiquet, et al., 2007), although it is still unclear if there is a connection between these genes and the tooth development specifically and/or dental caries process.

The meta-analysis results did not show any significant association signals; however, we had several suggestive variants near genes with plausible roles during tooth development, such as *LIN28B,* which was found to be expressed in dental epithelium, tooth germ, ameloblast and odontoblasts during the different stages of tooth development in mice (Dong, et al., 2017), and *ARHGAP21*, which has a role as a GTPase-activating protein (GAP) for *CDC42*, which is been suggested that it is required for proper odontogenesis and amelogenesis (Dubois, et al., 2005; Tian, et al., 2018). These findings suggest that these genes could be associated with dental caries,

however further investigation, in both animal and human, is needed to confirm their function in tooth development and dental caries.

Our investigation for the identified genes from previous GWASs of dental caries (Table 32) in our dental caries meta-analysis results to see if they had been replicated only showed limited evidence of suggestive associations.

There were suggestive variants near genes with biological roles in other parts of the body, unrelated to dental health, which we did not include in our results. Regional Association plots of those variants and other variant that reached suggestive level of significance near genes with unknown roles in dental and/or oral health are in the appendix figure A30 through figure A48.

5.5 CONCLUSION

This is one of the largest multi-ethnic GWAS study of dental caries that helped in discovering several loci containing genes with plausible biological roles in tooth development and/or oral dental caries including *MPPED2, GLI2*, *LIN28B*, *FHL2* and *NEDD9*. Further studies are needed to replicate the results of this study in an independent cohort, which is strongly recommended. In addition, we nominate new genes, such as *REL* gene, to be further investigated, in human and animal model, to fully understand their biological roles in tooth development and dental caries. Finally, understanding the factors influencing dental caries susceptibility will lead to improvements in prediction/diagnosis, prevention and disease management.

6.0 SIGNIFICANCE, LIMITATIONS, AND FUTURE DIRECTIONS

6.1 STRUCTURAL DENTAL ANOMALIES

To the best of our knowledge, this is the first study that investigated the different types of dental anomalies by conducting a multivariate GWAS in a multiethnic population. We were also able to investigate the relationship/correlation between several different dental anomalies in a large and diverse sample. Although we only detected weak correlations between different dental anomalies, it gave us an opportunity to go further and investigate these weak correlations on a genetic level by applying the multivariate approach of GWAS. This study helped in identifying/nominating potential genetic variants associated with different patterns of correlated dental anomalies. We were able to highlight several genes that had involvement in biological processes, including tooth development and inflammatory pathways.

In addition to conducting the multivariate approach of GWAS to investigate and identify risk variants associated with the correlated dental anomalies, we also conducted GWAS (univariate) for each of the included dental anomalies in a multiethnic population, which, to the best of our knowledge, has never been done before. We were able to identify/nominate potential genetic variants associated with different dental anomalies that had involvement in biological processes, including tooth development and dental health.

Although we had some strength on being the first and also identifying/nominating variants with roles in dental development that were associated with multiple correlated dental anomalies, we unfortunately had some limitations in this study. The major limitations was that we did not have replication cohorts available for testing those potential variants associated with the correlated

dental anomalies to confirm the association signals. Additional studies in larger and more diverse cohorts are warranted to assess the effect of those potential variants identified in this study. Even though we were able identify potential genetic variants associated with different patterns of correlated dental anomalies that were near several genes that had involvement in biological processes, including tooth development, none of our potential variants reached the whole genome significance level.

Another limitation of this current study is the unavailability of panoramic radiographs of the subjects in the study to make sure of the diagnosis of the structural dental anomalies included in this study, specifically in the case of tooth agenesis. The in-person dental examination and intraoral photos forms are provided in Appendix 1.Panoramic radiographs can help in identifying if the subjects are missing those teeth congenitally or if they are just unerupted (impacted) teeth. Hopefully future studies will be able to confirm those genetic associations in larger/diverse population, and this will help in eventually in the screening process and treatment for people at risk of developing those dental anomalies. This work helps in understanding the genetics of those dental anomalies which may also help to decrease oral health disparities across diverse underprivileged communities and reduce the burden of oral diseases in the future.

6.2 DENTAL CARIES

This GWAS of dental caries in our study is considered to be one of the largest multi-ethnic GWAS that helped in discovering several novel loci containing genes with plausible biological roles in tooth development and/or oral dental caries. Our results summed risk loci associated with DFT+dft, i.e. all caries not just a particular dentition. Moreover, we were able to confirm

some of the previously identified genes *(MPPED2* and *NEDD9*) that has been investigated before and had an association with dental caries.

The biggest limitation of the GWAS/meta-analysis of dental caries was inability to replicate some of those association signals near genes with roles in dental caries and tooth development that we were able to identify separately in POFC and COHRA1. Maybe if we lowered our threshold of significance, we could've detected those associations signal in our meta-analysis. Further studies are warranted to replicate the results of this GWAS/meta-analysis of dental caries in an independent and more diverse cohort. In addition, functional studies are needed to further investigate the biological roles of those nominated genes, in human and animal model. These studies will eventually help in understanding the biological roles of those genes, especially the ones with relevant roles in tooth development process and dental caries.

ACKNOWLEDGEMENT

Table A 1 Age by the recruitment site and subject type in POFC

| SITE | Subject Type | | *P* value* | 95%CI | Total |
	Unaffected Relatives	Control			
COLOMBIA	34.96 8-81	32.15 8-68	0.05287	-0.04_ 5.67	33.83 8-81
COLORADO	29.06 8-63	0 0	----------------	--------------------	29.06 8-63
GUATEMALA	29.13 8-66	27.44 8-74	0.2009	-0.91_ 4.29	28.03 8-74
HUNGARY	33.82 8-74	31.98 8-79	0.09642	-0.33_3.99	32.74 8-79
IOWA	28.62 8-56	27.77 8-58	0.464	-1.43_ 3.12	28.81 8-58
ARGENTINA	31.52 8-78	25.46 8-56	0.004	2.02_10.12	30.63 8-78
PHILIPPINES	32.15 8-82	30.62 20-57	0.1536	-0.58_ 3.64	31.91 8-82
PITTSBURGH	32.03 8-73	33.15 8-73	0.6028	-5.36_ 3.12	32.67 8-73
PUERTORICO	32.15 20-48	32.49 8-50	0.9259	-7.94_7.26	32.25 8-50
TEXAS	28.62 8-87	19 19	----------------	--------------------	28.59 8-87
Total	31.53 8-82	30.18 8-79	0.005438	0.40_ 2.30	31.00 8-82

*P value are based on t-test of mean differences

Table A 2 Distribution of Subjects according to the age groups in POFC

Age Groups	Freq.	%
7-10	298	8.33
11-20	671	18.75
21-31	827	23.11
32-52	1534	42.86
53-73	241	6.73
74-82	8	0.22
Total	3579	100

Table A 3 Regression Results for the covariates in POFC

	Total DFT+ dft
SITE	-0.0943***
	(0.0280)
Age	0.107***
	(0.00469)
Sex	0.422**
	(0.135)
Subject Type	-0.372**
	(0.142)
N	3579

Standard errors in parentheses
* $P < 0.05$, **$P < 0.01$, ***$P < 0.001$

Table A 4 Age by the recruitment site and sex in COHRA1

| SITE | Sex | | *P* value* | 95%CI | Total |
	Male	Female			
BK	18.41 7-52	26.71 7-57	3.741e-05*	-12.17_4.43	23.35 7-57
BR	24.75 7-59	27.34 7-53	0.1858	-6.43_1.26	26.13 7-59
BU	25.08 7-57	27.44 7-58	0.2134	-6.09_ 1.37	26.37 7-58
WV	23.85 7-62	25.18 7-67	0.09056	-2.87_0.21	24.63 7-67
Total	23.63 7-62	28.89 7-67	0.01425	-1.08_-0.12	5.98 7-67

P value are based on t-test of mean differences

Table A 5 Regression Results for the covariates in COHRA1

	Total DFT+ dft
SITE	1.909*** (0.453)
Age	0.364*** (0.036)
Sex	0.0596446 (0.2146)
N	1763

Standard errors in parentheses
P < 0.05, **P* < 0.01, ***P* < 0.001

Table A 6 Overlaped results between HDR GWAS and DRM GWAS

SNP	CHR	BP	A1	MAF	F.HDR	*P* (HDR)	F.DRM	*P* (DRM)	Type
rs113402395	9	140135219	A	0.09724	7.393	1.20E-05	7.285	1.90E-05	Imputed
rs1322468	9	7949546	C	0.1661	8.862	1.30E-05	8.872	5.46E-05	Imputed
rs1390193	3	55741010	A	0.2433	9.433	1.50E-05	9.985	5.46E-05	Imputed
rs141429354	15	54478011	T	0.05562	8.714	3.00E-06	8.516	3.00E-06	Imputed
rs2828273	21	24946151	C	0.195	8.03	3.74E-05	9.861	1.10E-05	Imputed
rs34903076	8	69306322	A	0.4181	10.05	3.00E-06	9.007	1.70E-05	Imputed
rs35112680	15	42792340	CT	0.1754	6.485	7.61E-05	6.36	3.92E-05	Imputed
rs36159351	11	48536107	C	0.4343	9.25	2.00E-05	9.665	2.00E-05	Imputed
rs404727	20	55465434	A	0.251	10.12	8.00E-06	8.844	7.35E-05	Imputed
rs6479408	9	94903215	G	0.4517	11.37	1.00E-06	11.26	5.00E-06	Imputed
rs6957650	7	106450360	C	0.3206	6.27	2.10E-05	7.053	2.40E-05	Imputed
rs8016518	14	97548256	C	0.4565	9.336	2.00E-05	9.1	5.11E-05	Imputed
rs908329	1	235051501	A	0.3709	8.74	4.69E-05	9.371	2.40E-05	Genotyped
rs9502420	6	6142985	A	0.07251	12.07	7.00E-06	12.74	7.00E-06	Imputed
rs9509076	13	20747640	A	0.417	7.932	2.50E-05	8.608	3.80E-05	Genotyped

Appendix 1

OFC Dental Phenotype Form Part 2 Dental Professionals

20278

Recorded by

Dentist Initials

Study ID: Letters | Number

Individual ID: Number | – | Number

Date: mm / dd / yyyy

1. Dental Examination, Maxillary Teeth

☐ Part 2 Not Completed/Refused

Rate each tooth by marking the bubbles below. Fill-in the bubble for primary or permanent tooth. Teeth can either be missing, sound, decayed, or restored. If there is a supernumerary tooth, mark the box between the two adjacent teeth.

Quadrant 1	18 1	17 2	16 3	15 O 4 / 55 O A	14 O 5 / 54 O B	13 O 6 / 53 O C	12 O 7 / 52 O D	11 O 8 / 51 O E
Missing: Decay / Agenesis / Other / X-ray Needed	O O O O	O O O O	O O O O	O O O O	O O O O	O O O O	O O O O	O O O O
Sound / Decayed / Restored	O O O	O O O	O O O	O O O	O O O	O O O	O O O	O O O
Other: Fluorosis / Hypoplasia / Microdontia / Impacted / Rotation / Displaced / Other (specify)	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O
Extra Teeth	O	O	O	O	O	O	O	O

Quadrant 1 Notes

Quadrant 2	21 O 9 / 61 O F	22 O 10 / 62 O G	23 O 11 / 63 O H	24 O 12 / 64 O I	25 O 13 / 65 O J	26 14	27 15	28 16
Missing: Decay / Agenesis / Other / X-ray Needed	O O O O	O O O O	O O O O	O O O O	O O O O	O O O O	O O O O	O O O O
Sound / Decayed / Restored	O O O	O O O	O O O	O O O	O O O	O O O	O O O	O O O
Other: Fluorosis / Hypoplasia / Microdontia / Impacted / Rotation / Displaced / Other (specify)	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O	O O O O O O O
Extra Teeth	O	O	O	O	O	O	O	O

Quadrant 2 Notes

Appendix Figure I. In-person dental exam form

OFC Dental Phenotype Part 2 Intra-Oral Photos

Recorded by

Initials

Date	mm / dd / yyyy

Study ID — Letters — Numbers

Individual ID — Number — Number
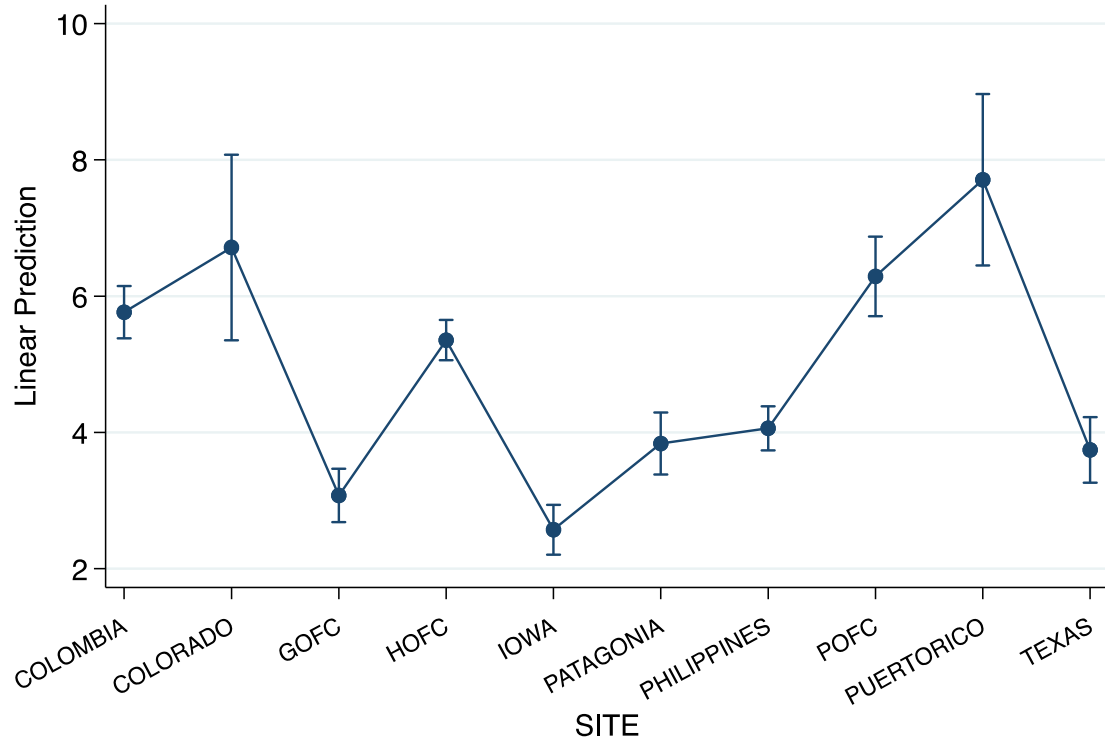
Overall Photo Quality
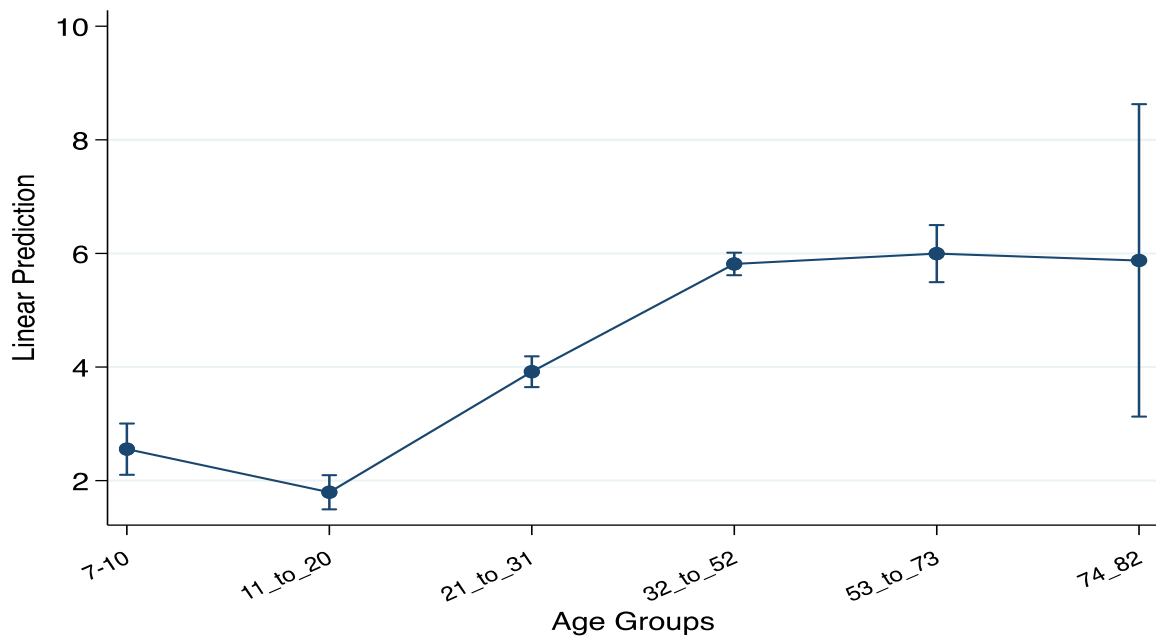
O Poor
O Good

1. Dental Examination, Maxillary Teeth

Rate each tooth by marking the bubbles below. Each tooth should have an entry. Fill-in the bubble for primary or permanent tooth. Teeth can either be missing or present. If there is space or a supernumerary tooth, mark the box between the two adjacent teeth.

Quadrant 1	18 1	17 2	16 3	15 O 4 / 55 O A	14 O 5 / 54 O B	13 O 6 / 53 O C	12 O 7 / 52 O D	11 O 8 / 51 O E
Missing: Agenesis / Other — **Present**	O O O	O O O	O O O	O O O	O O O	O O O	O O O	O O O
Present Status: Full coverage (crown) / Partial coverage (onlay, cusp replacement, veneers)	O O	O O	O O	O O	O O	O O	O O	O O
Filling (amalgam, composite)	O	O	O	O	O	O	O	O
Gross decay / Attrition more than 2/3 of the clinical crown	O O	O O	O O	O O	O O	O O	O O	O O
If present: Fluorosis / Hypoplasia / Hypocalcification / Microdontia / Impacted / Rotation / Displaced / Mammalons / Incisal Fissures / Other (specify below)	O O O O O O O O O O	O O O O O O O O O O	O O O O O O O O O O	O O O O O O O O O O	O O O O O O O O O O	O O O O O O O O O O	O O O O O O O O O O	O O O O O O O O O O
Confidence: Low / High	O O	O O	O O	O O	O O	O O	O O	O O

Extra Teeth	O	O	O	O	O	O	O	O	
Space Between Teeth	O	O	O	O	O	O	O	O	

Quadrant 1 Notes

Appendix Figure 2. Intraoral potograph evaluation form

a)

b)

Figure A1 Adjusted prediction of Total DFT+dft scores across a) **different sites and** b) **across different age groups in POFC**
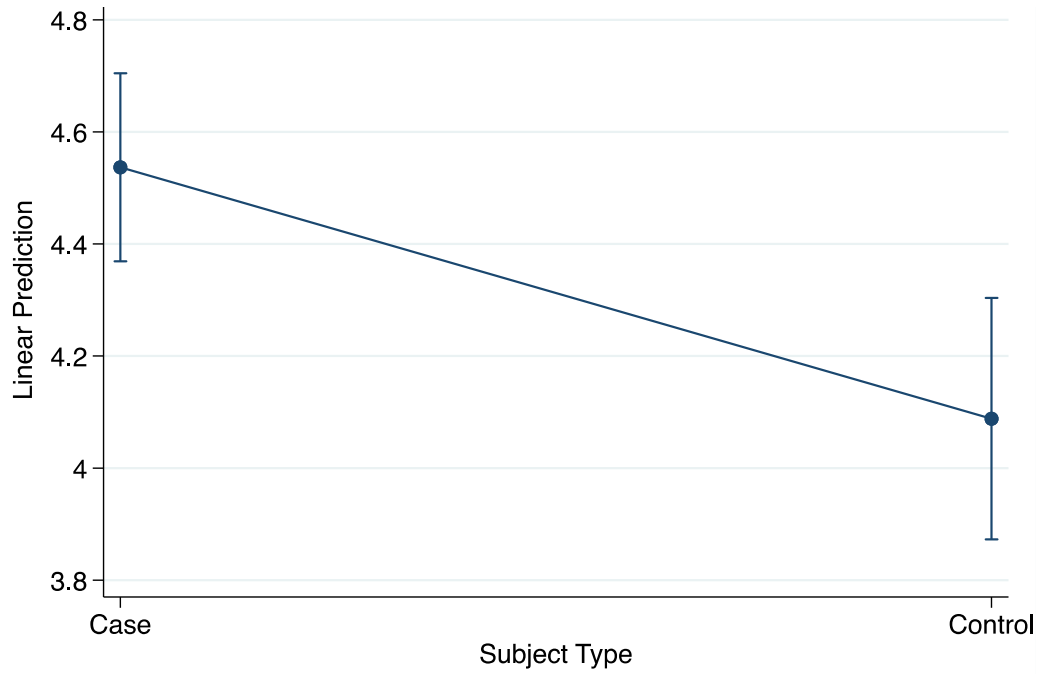
a)
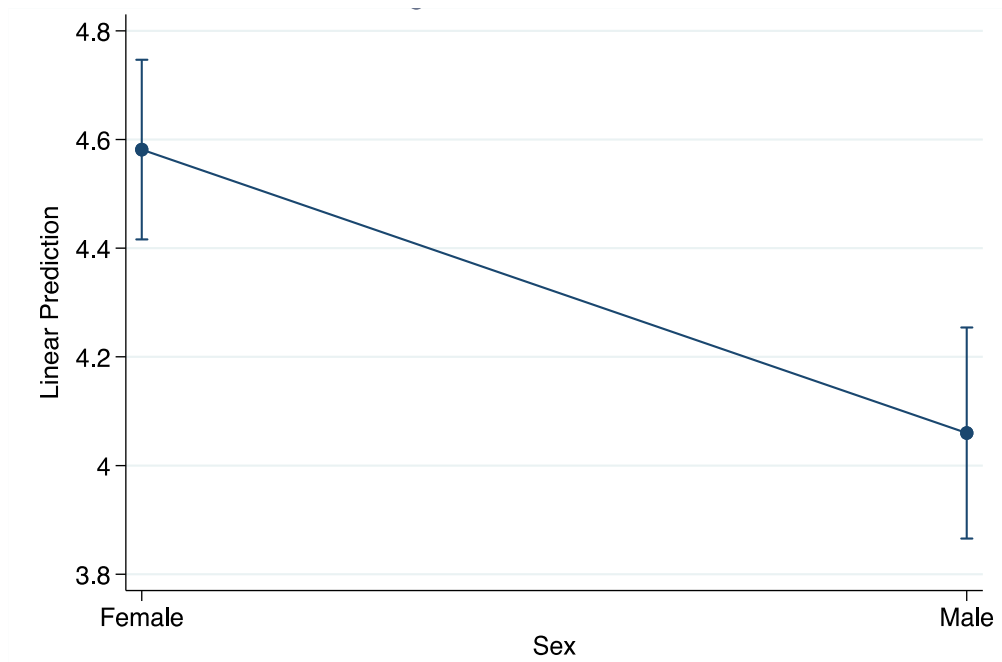
b)

Figure A2 Predictive margins of Total DFT+dft scores across a) **unaffected relatives and controls and** b) **between male and female in POFC**

Figure A3 Regional Association Plot rs8005462 from (AIR) results in POFC
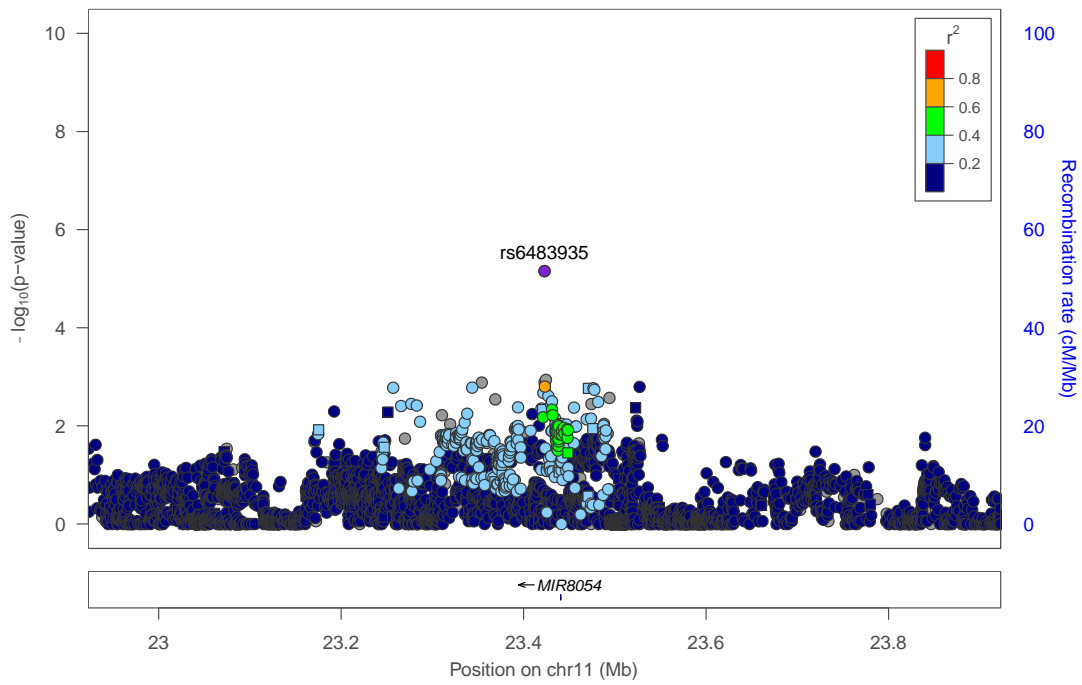
Figure A4 Regional Association Plot rs6483935 from (AIR) results in POFC

Figure A5 Regional Association Plot rs34903076 from (HDR) results in POFC
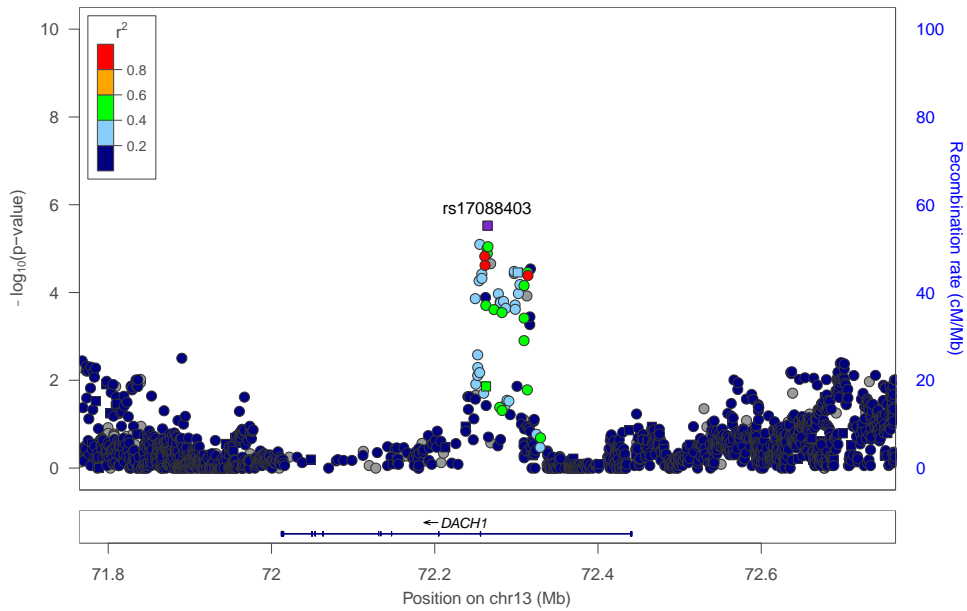
Figure A6 Regional Association Plot rs17088403 from (HDR) results in POFC
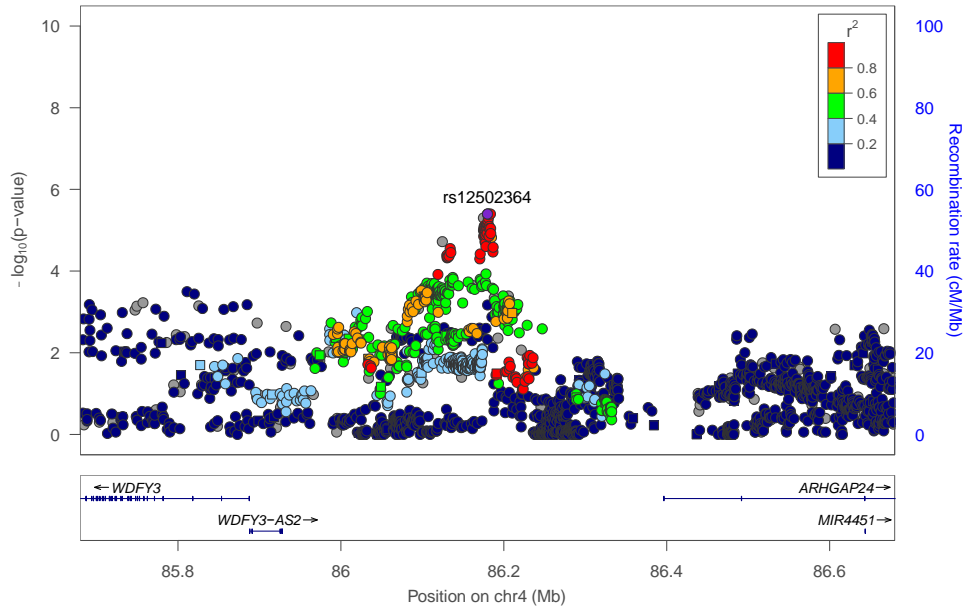
Figure A7 Regional Association Plot rs12502364 from (HDR) results in POFC
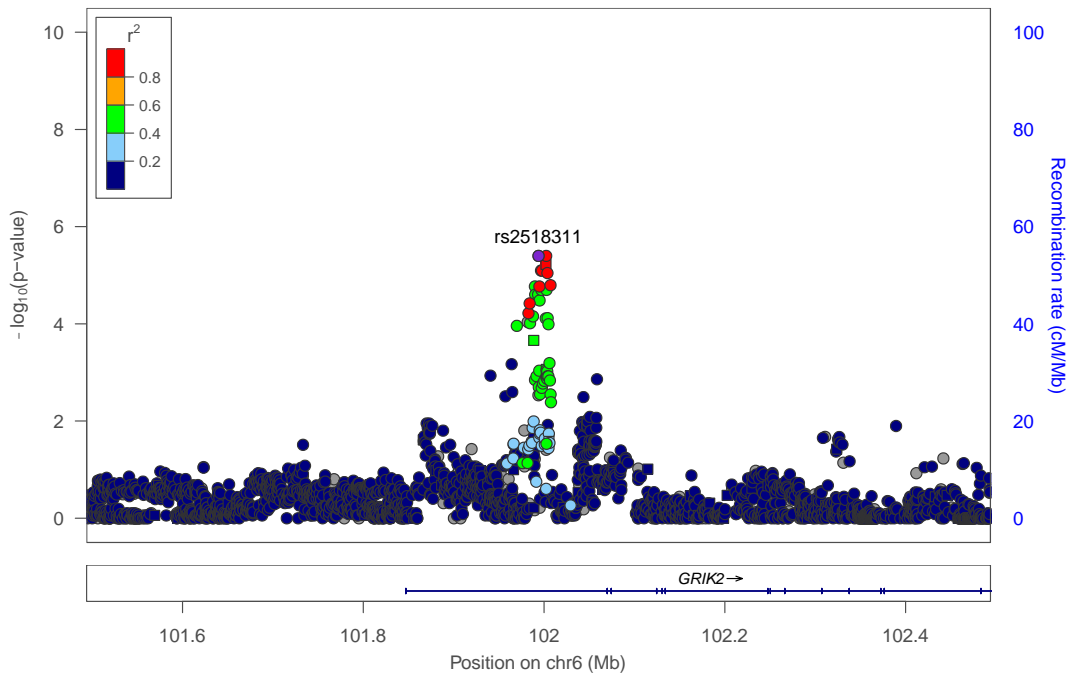
Figure A8 Regional Association Plot rs2518311 from (HDR) results in POFC
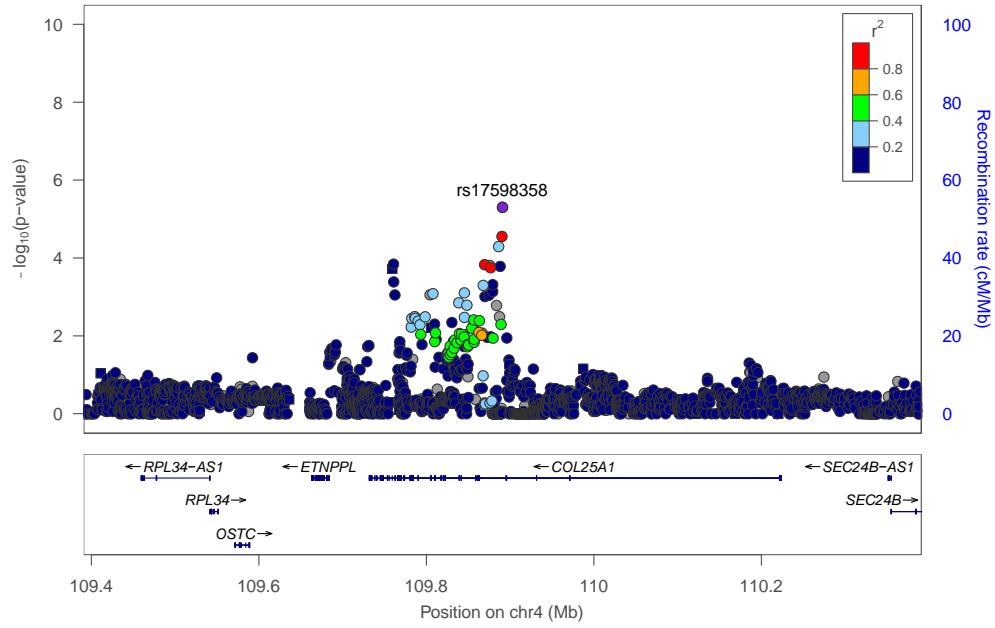
Figure A9 Regional Association Plot rs17598358 from (HDR) results in POFC
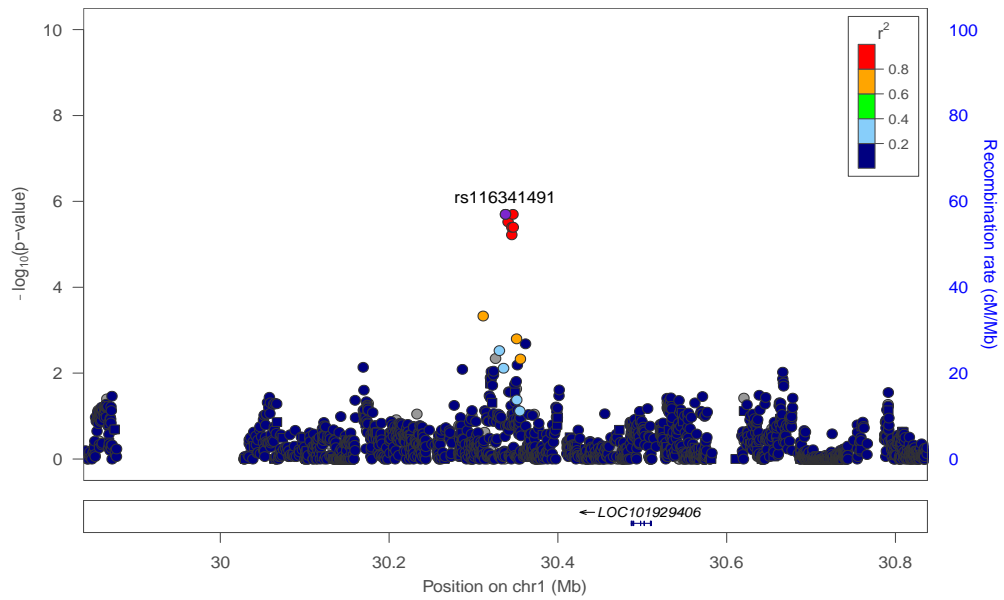
Figure A10 Regional Association Plot rs116341491 from (DRM) results in POFC
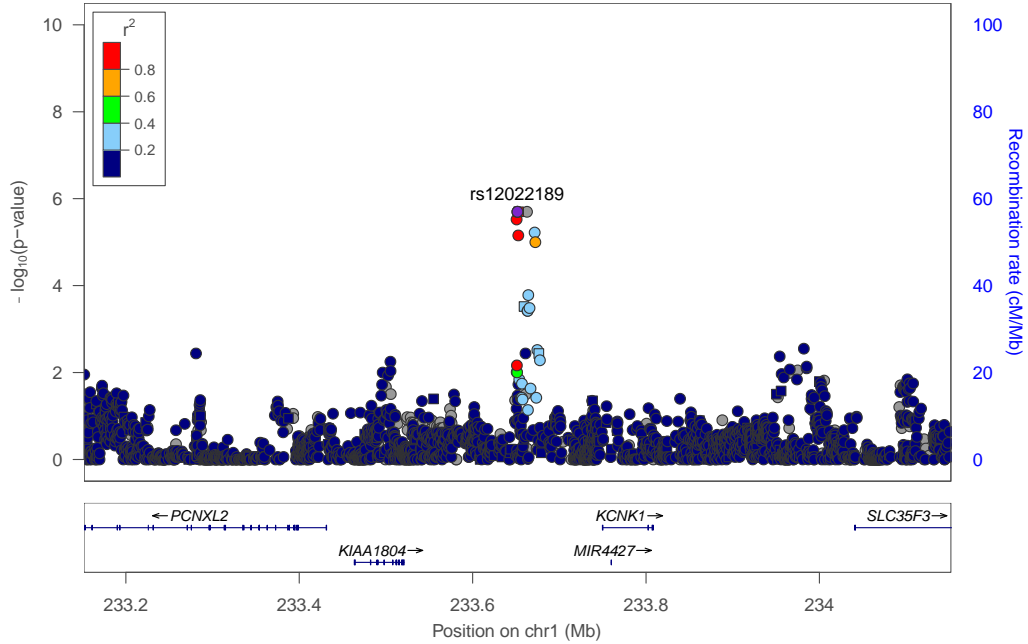
Figure A11 Regional Association Plot rs12022189 from (DRM) results in POFC
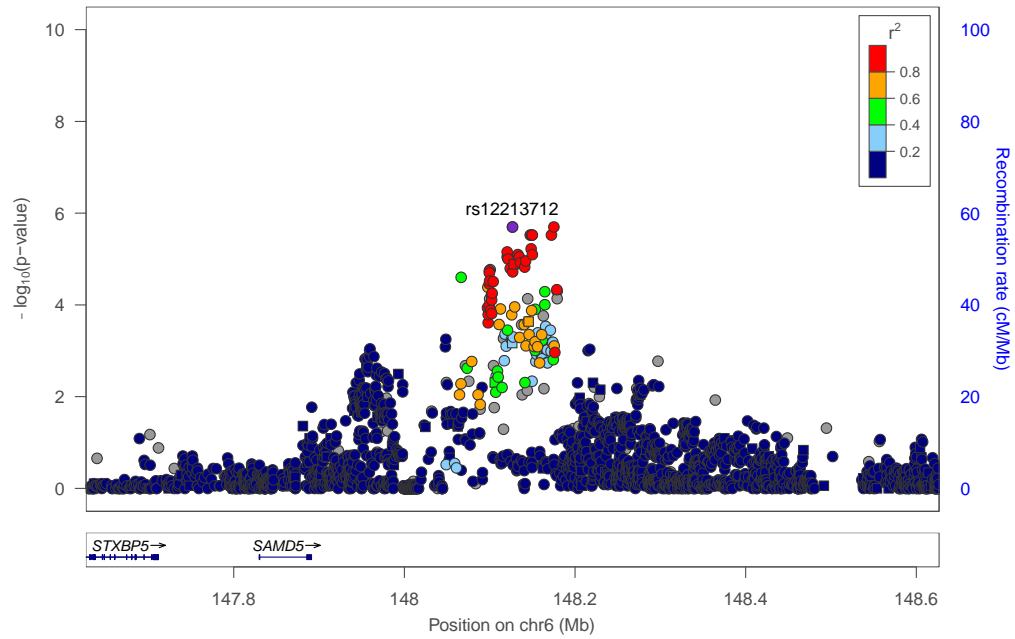
Figure A12 Regional Association Plot rs12213712 from (DRM) results in POFC
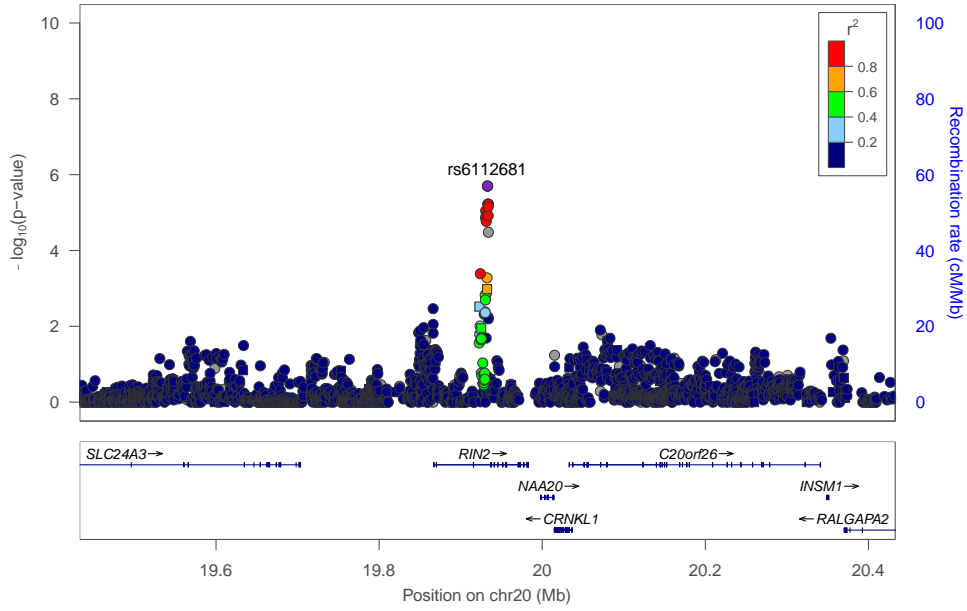
Figure A13 Regional Association Plot rs6112681 from (DRM) results in POFC
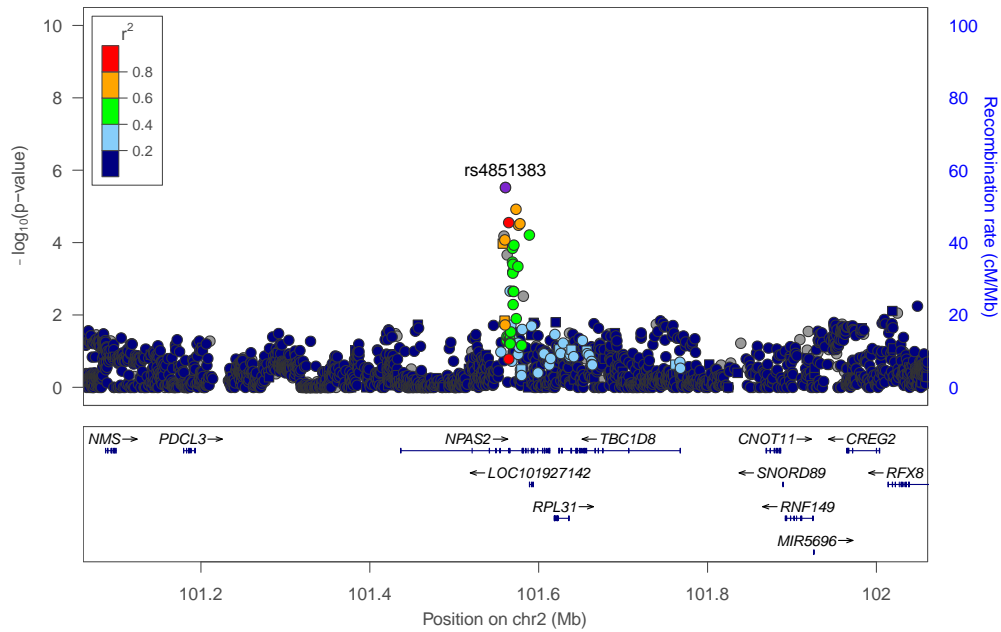
Figure A14 Regional Association Plot rs4851383 from (DRM) results in POFC
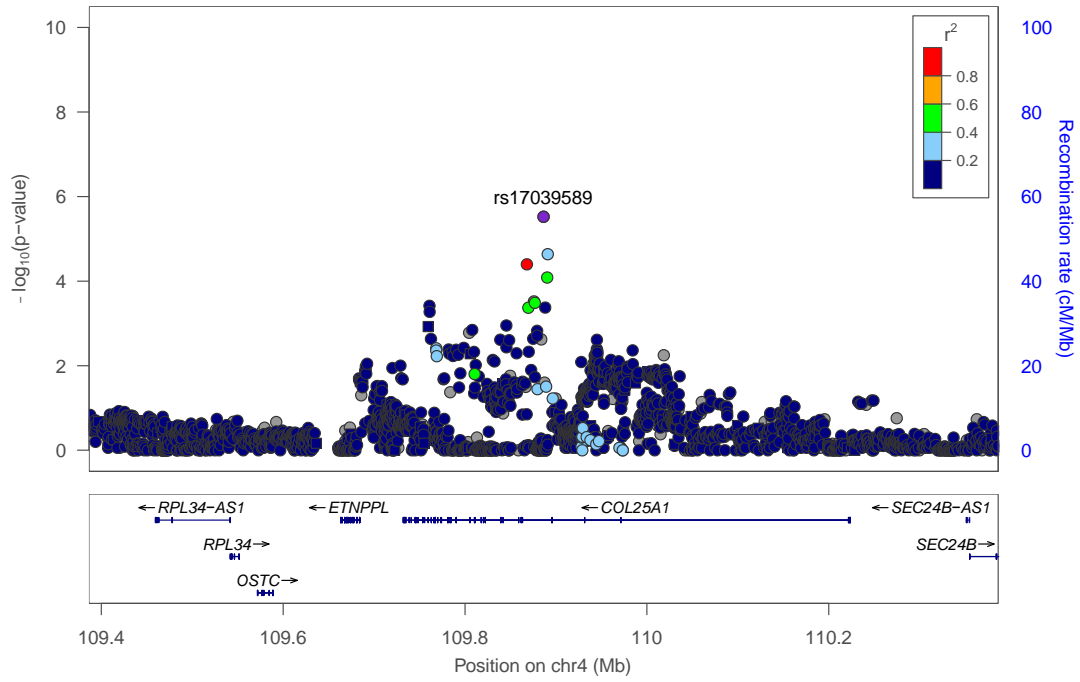
Figure A15 Regional Association Plot rs17039589 from (DRM) results in POFC
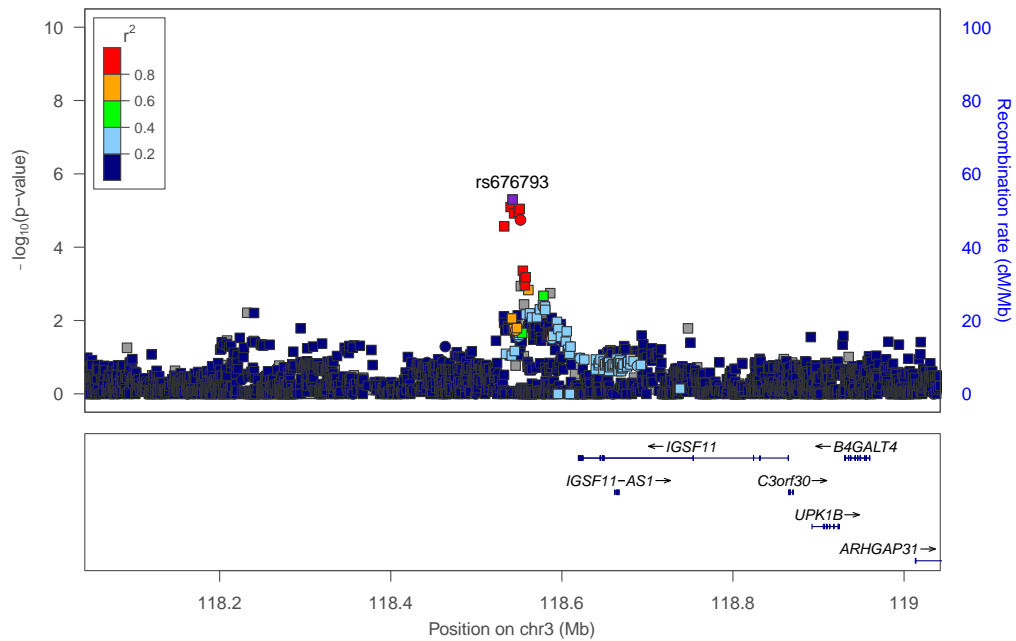
Figure A16 Regional Association Plot rs676793 from (DRM) results in POFC
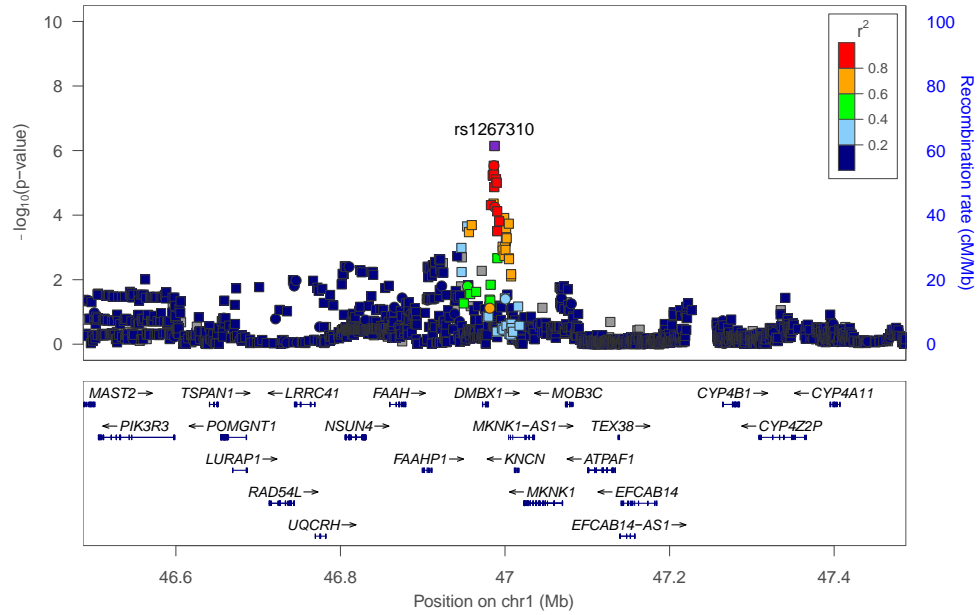
Figure A17 Regional Association Plot rs1267310 from (DAH) results in POFC
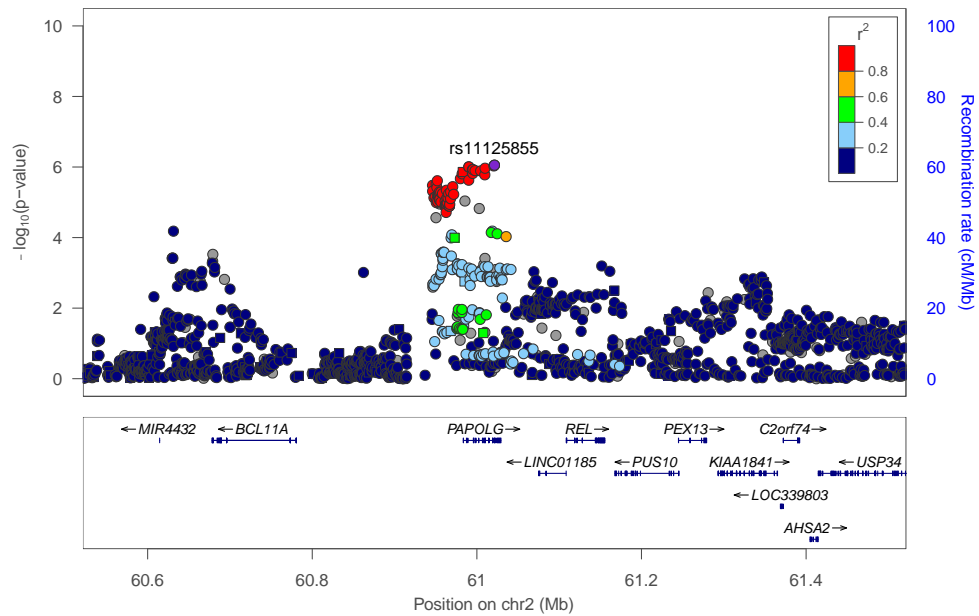
Figure A18 Regional Association Plot rs1112855 from (DAH) results in POFC
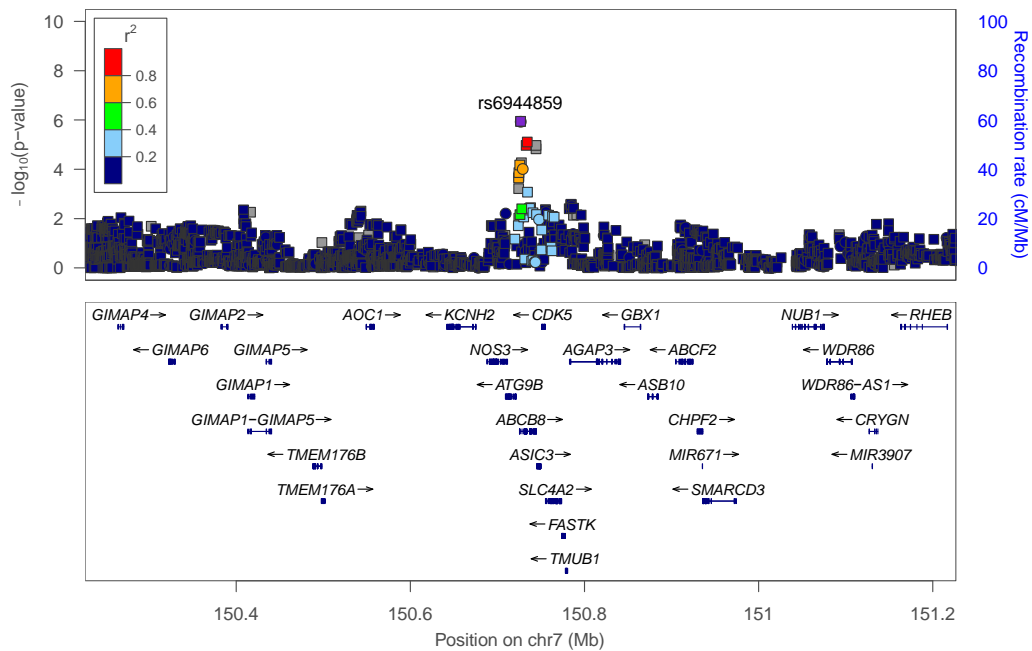
Figure A19 Regional Association Plot rs6944859 from (DAH) results in POFC
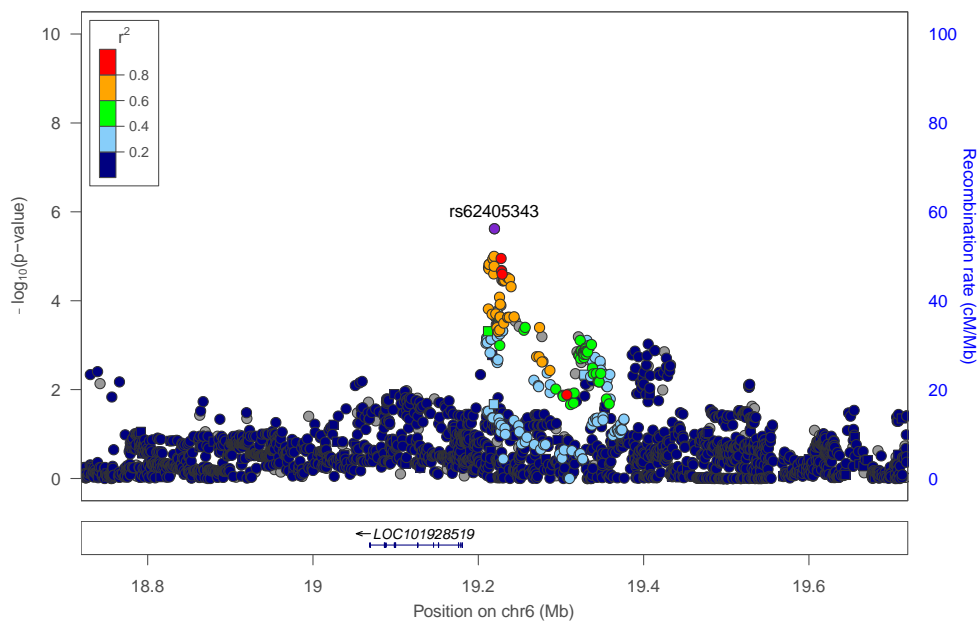
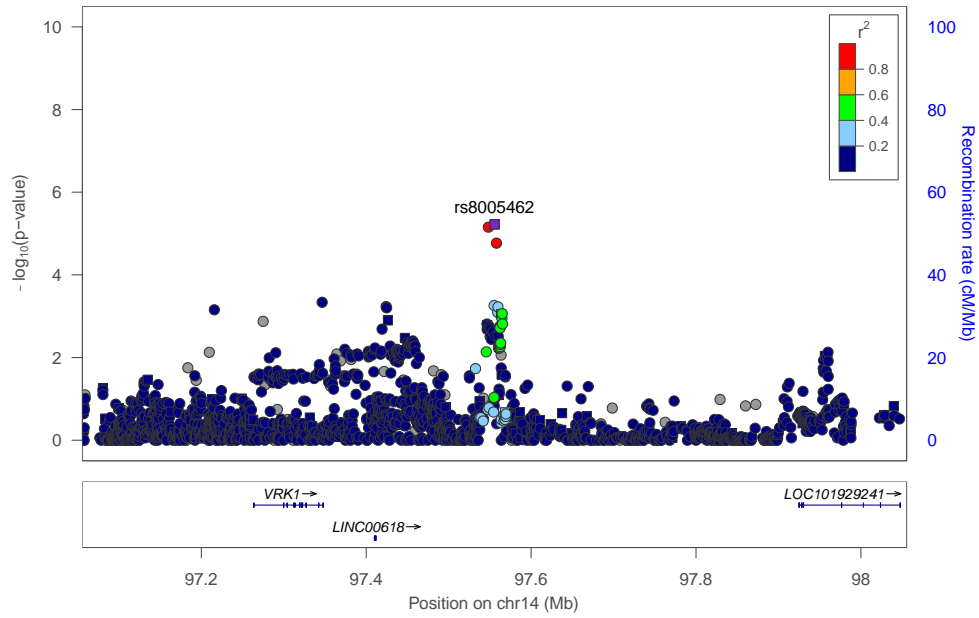**Figure A20** Regional Association Plot rs62405343 from (DAH) results in POFC

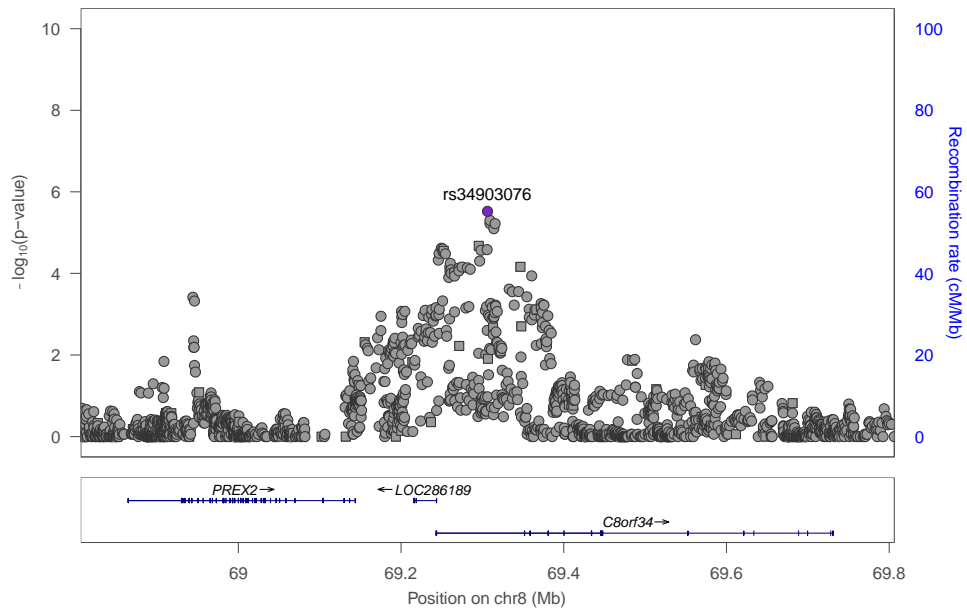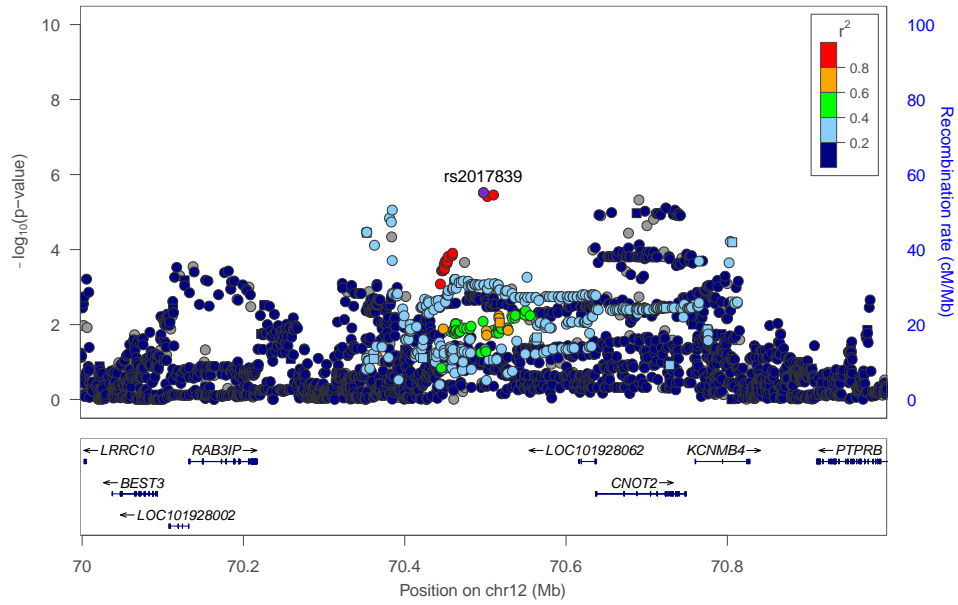**Figure A21** Regional Association Plot rs2017839 from (DAH) results in POFC
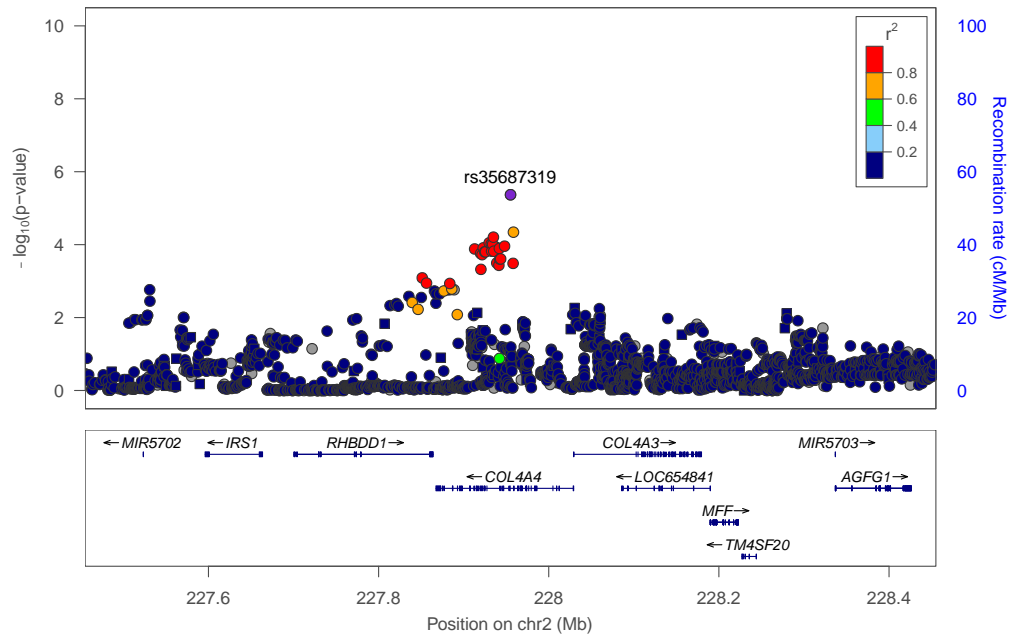
Figure A22 Regional Association Plot rs35687319 from (DAH) results in POFC
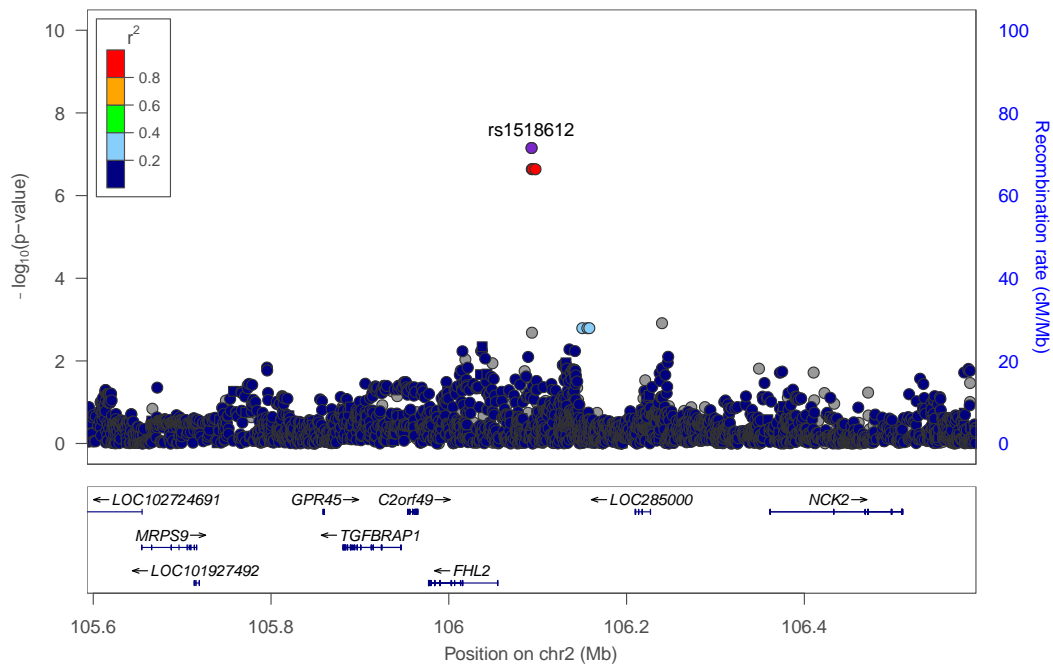
Figure A23 Regional Association Plot of rs1518612 from dental caries GWAS in POFC
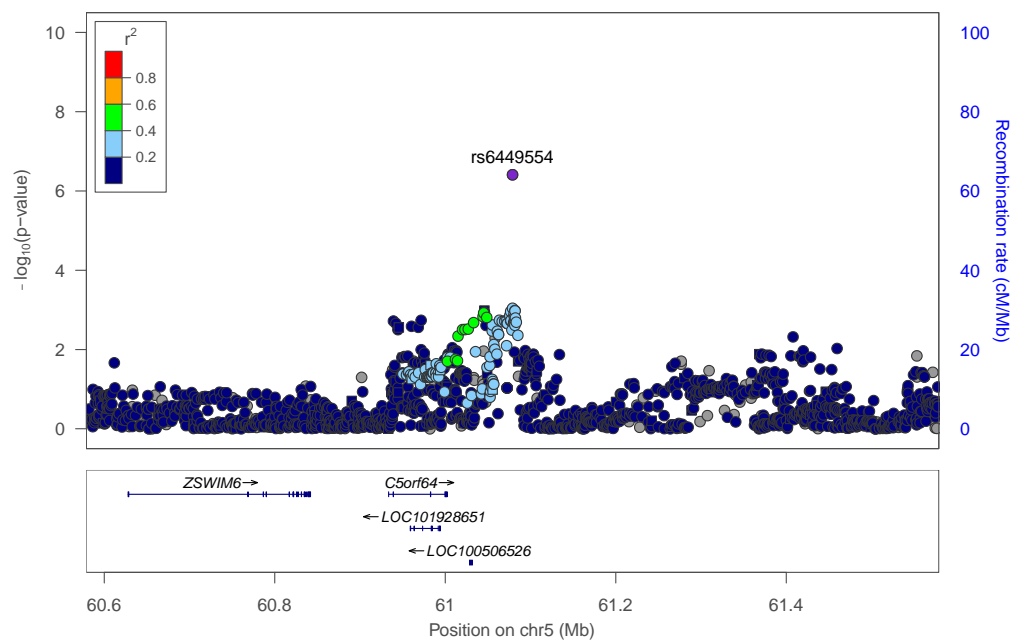
Figure A24 Regional Association Plot of rs6449554 from dental caries GWAS in POFC
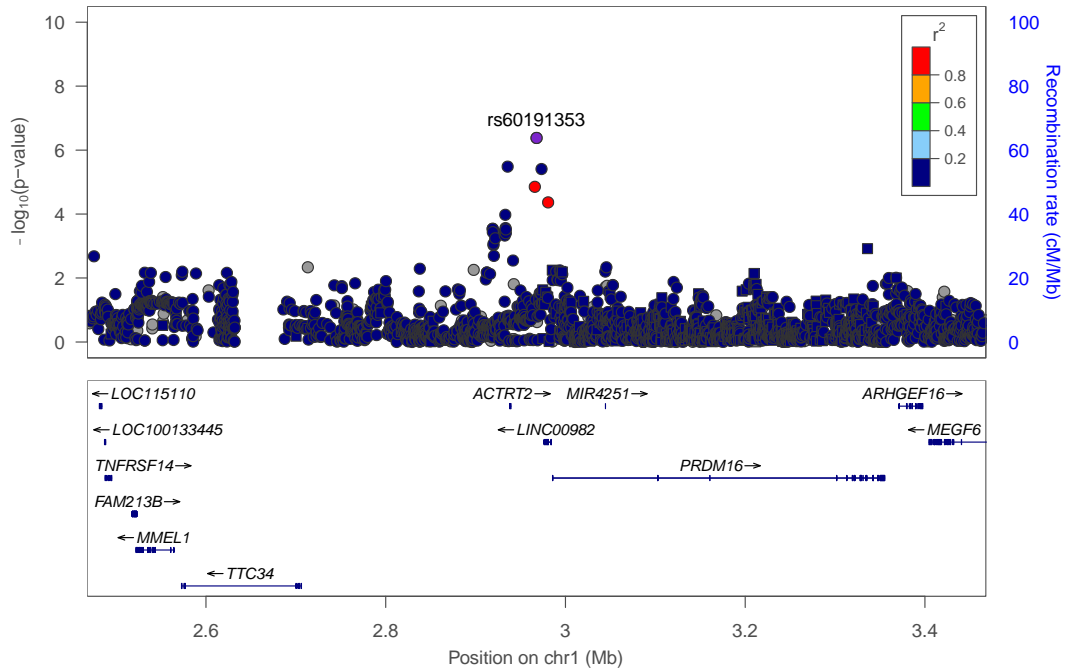
Figure A25 Regional Association Plot of rs6091353 from dental caries GWAS in POFC
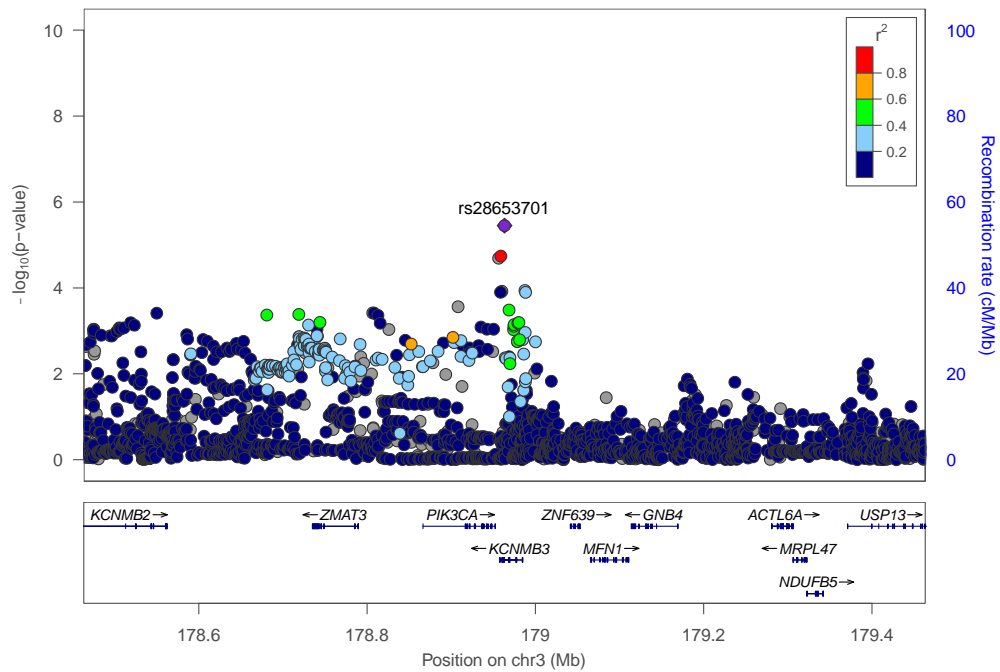
Figure A26 Regional Association Plot of rs28653701 from dental caries GWAS in POFC
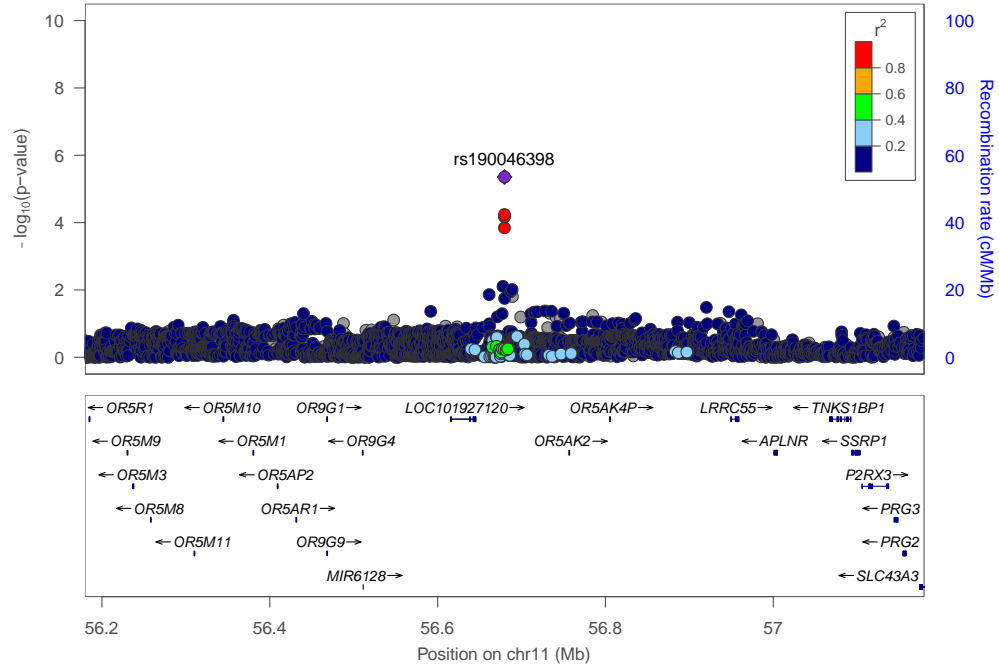
Figure A27 Regional Association Plot of rs190046398 from dental caries GWAS in POFC
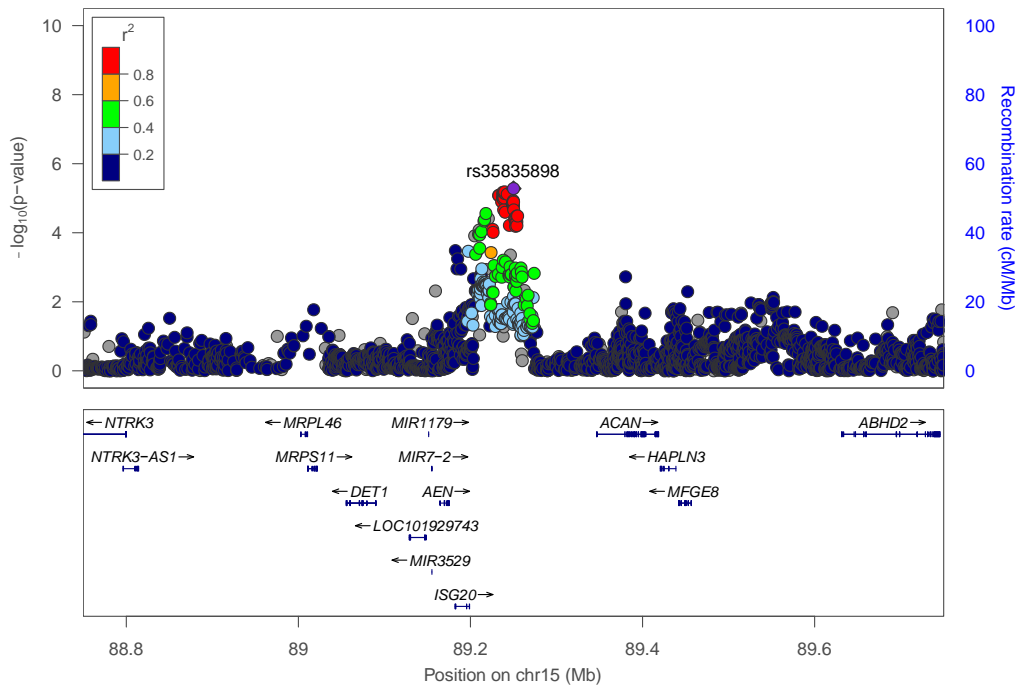
Figure A28 Regional Association Plot of rs35835898 from dental caries GWAS in POFC
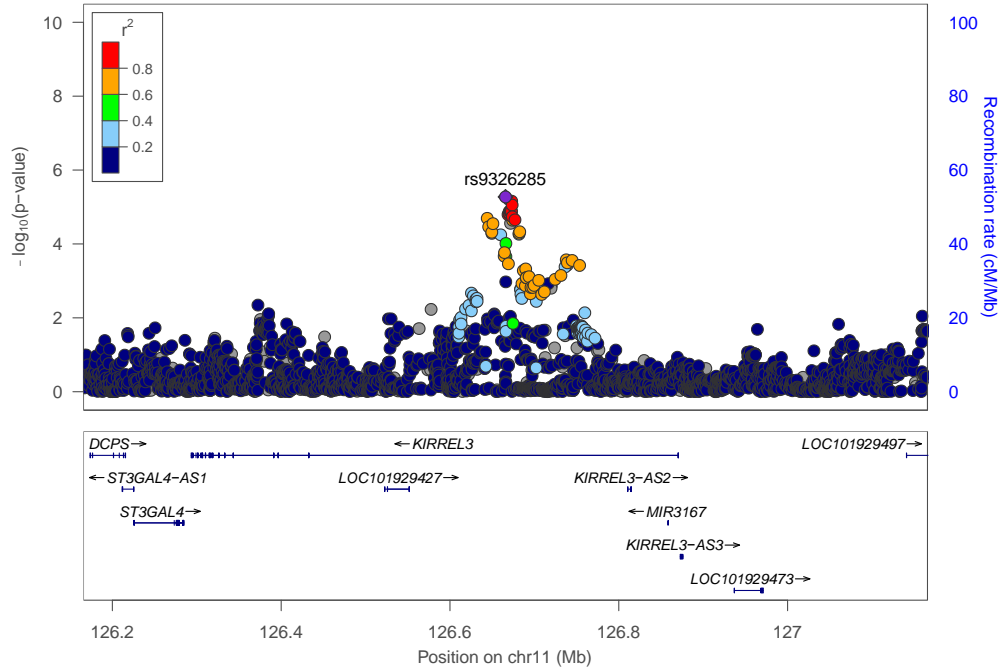
Figure A29 Regional Association Plot of rs9326285 from dental caries GWAS in POFC
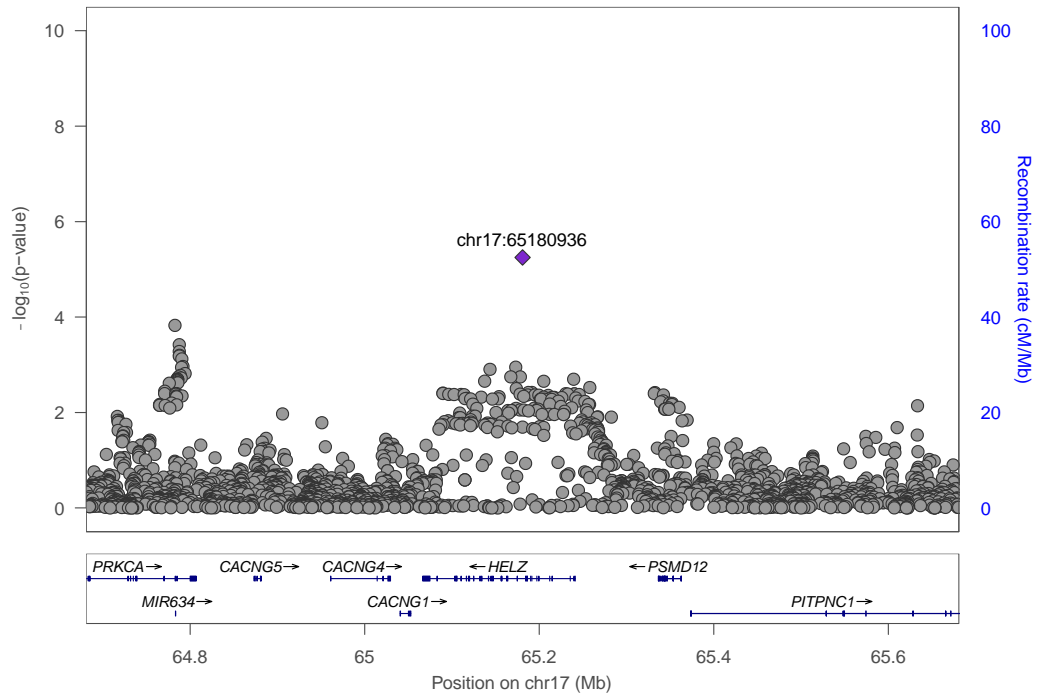
Figure A30 Regional Association Plot of chr17:65180936 from dental caries GWAS in POFC
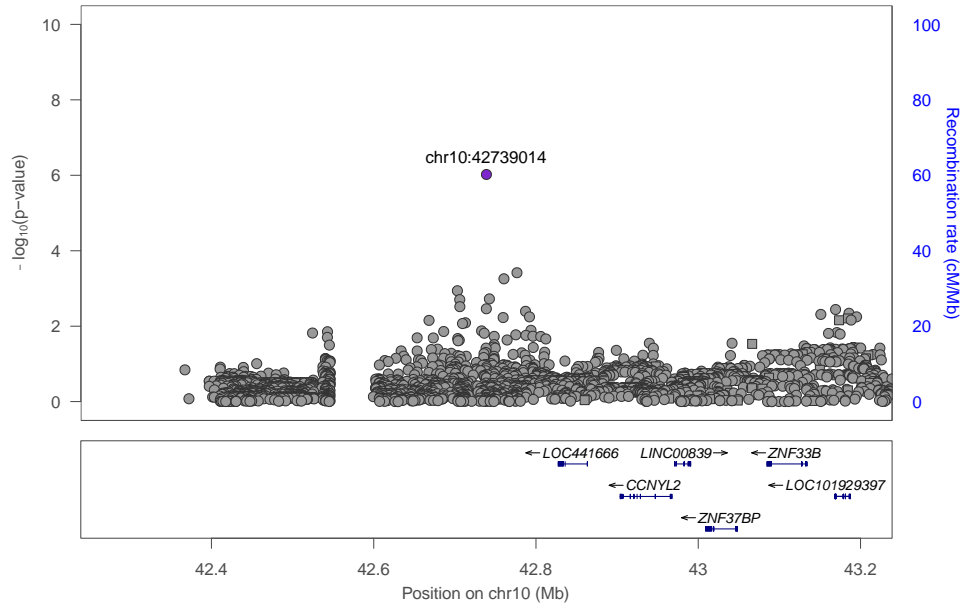
Figure A31 Regional Association Plot of chr10:42739014 from dental caries GWAS in COHRA1
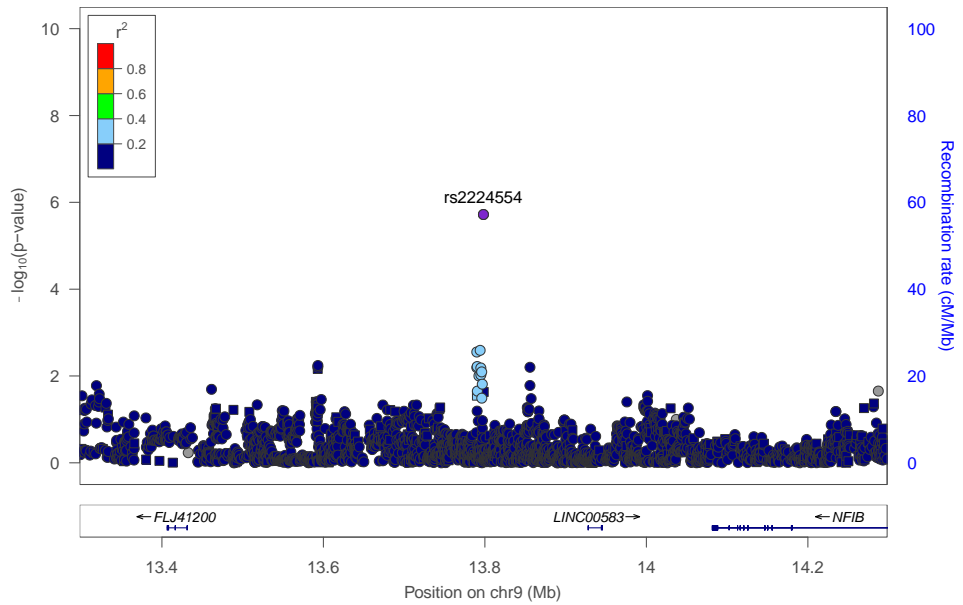
Figure A32 Regional Association Plot of rs2224554 from dental caries GWAS in COHRA1
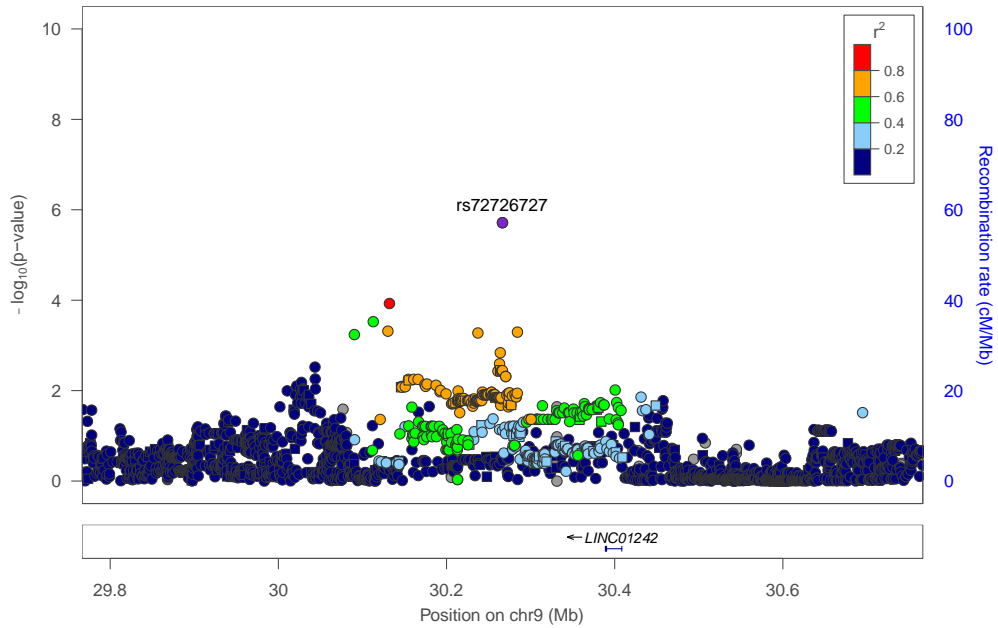
Figure A33 Regional Association Plot of rs72726727 from dental caries GWAS in COHRA1
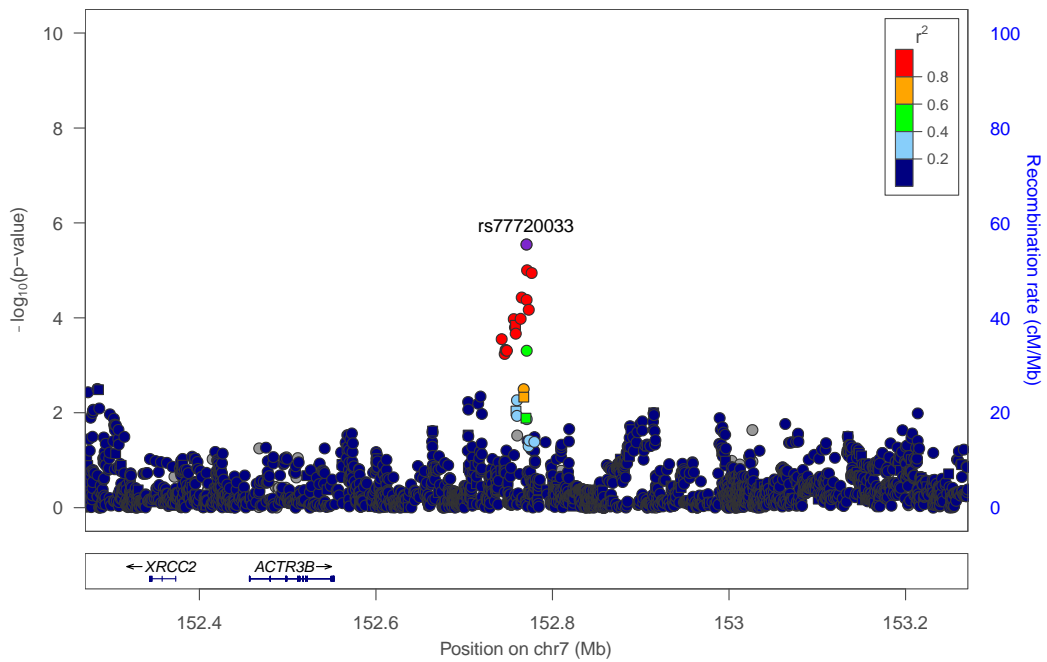
Figure A34 Regional Association Plot of rs777200033 from dental caries GWAS in COHRA1
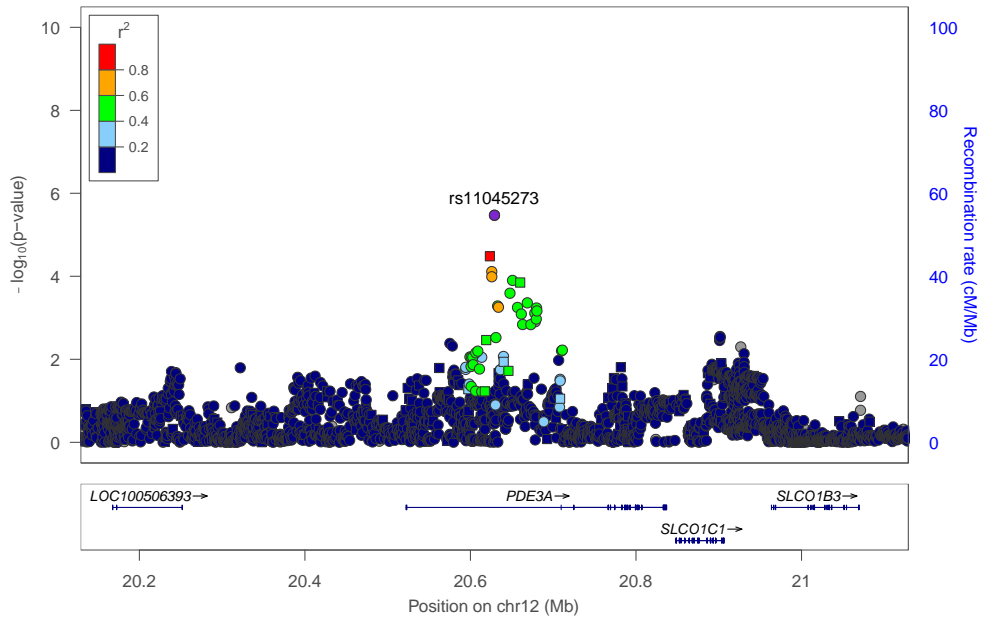
Figure A35 Regional Association Plot of rs11045273 from dental caries GWAS in COHRA1
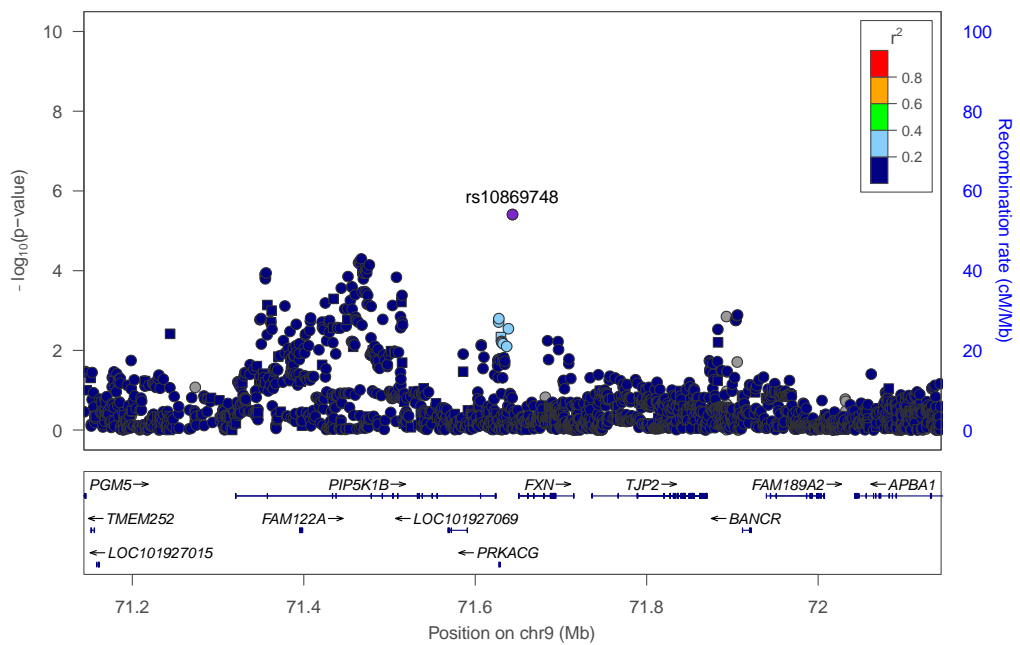
Figure A36 Regional Association Plot of rs10868748 from dental caries GWAS in COHRA1
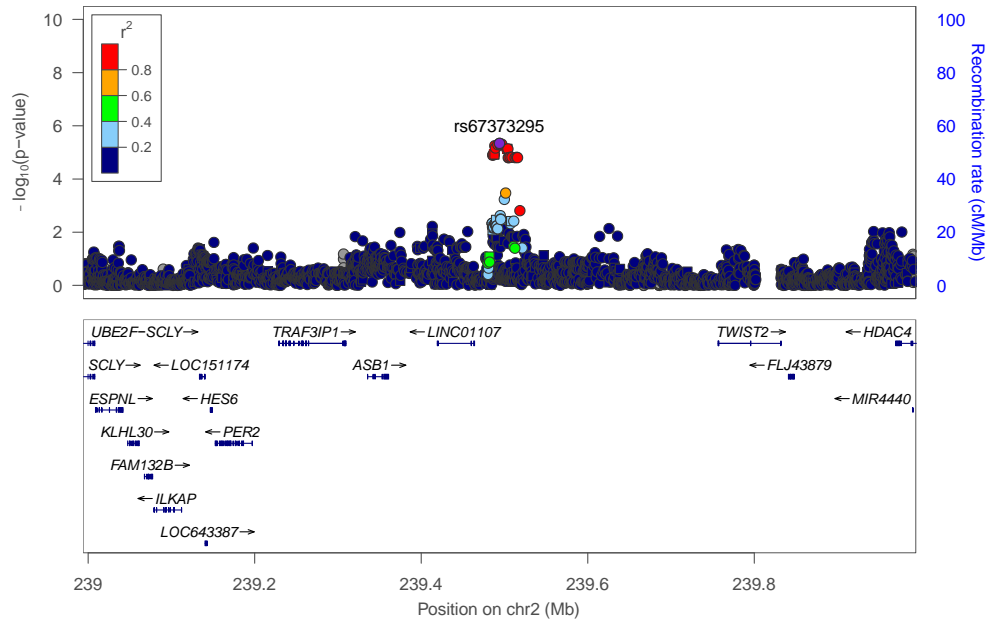
Figure A37 Regional Association Plot of rs67373295 from dental caries GWAS in COHRA1
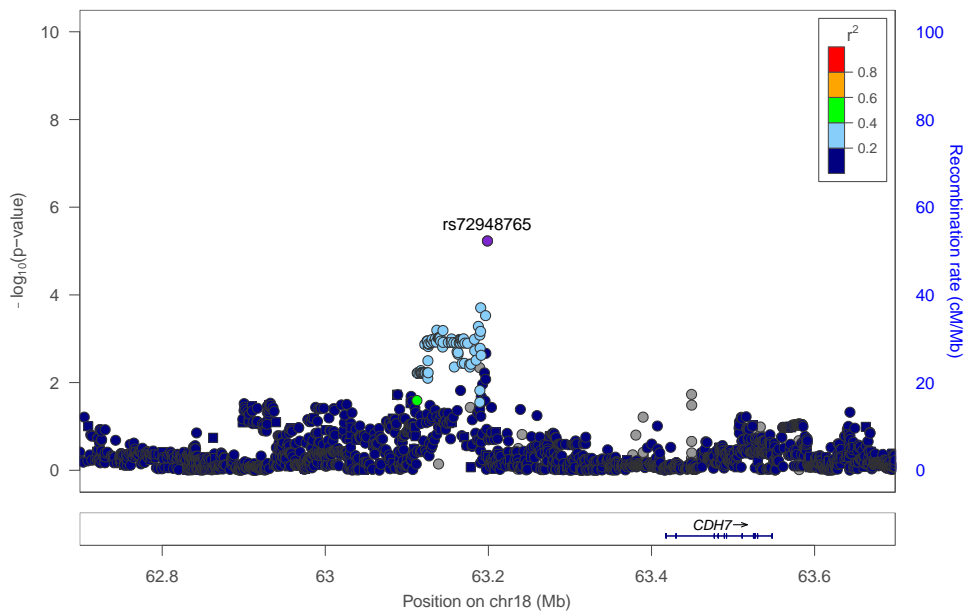
Figure A38 Regional Association Plot of rs72948765 from dental caries GWAS in COHRA1

194
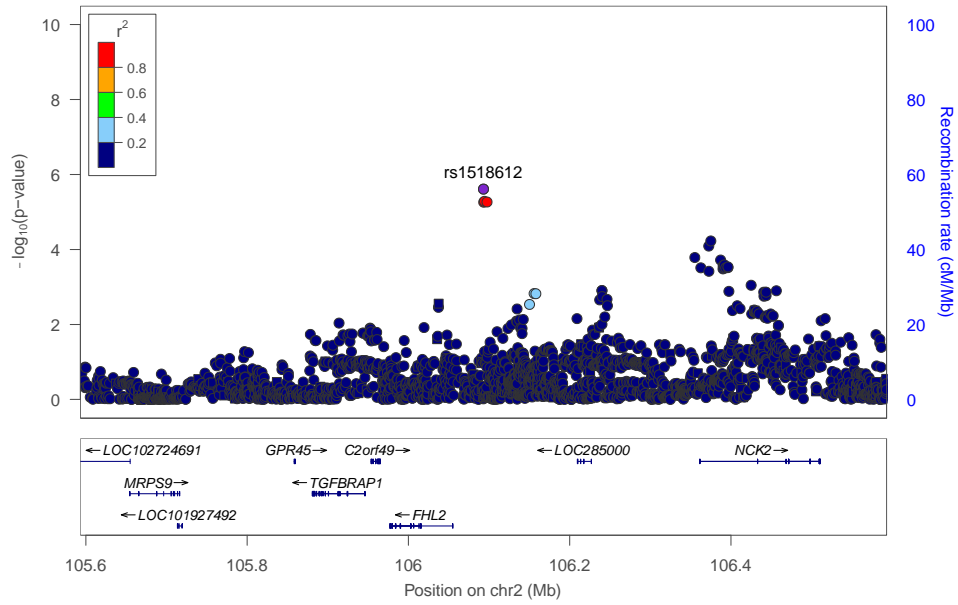
Figure A39 Regional Association Plot of rs1518612 from dental caries meta-analysis results
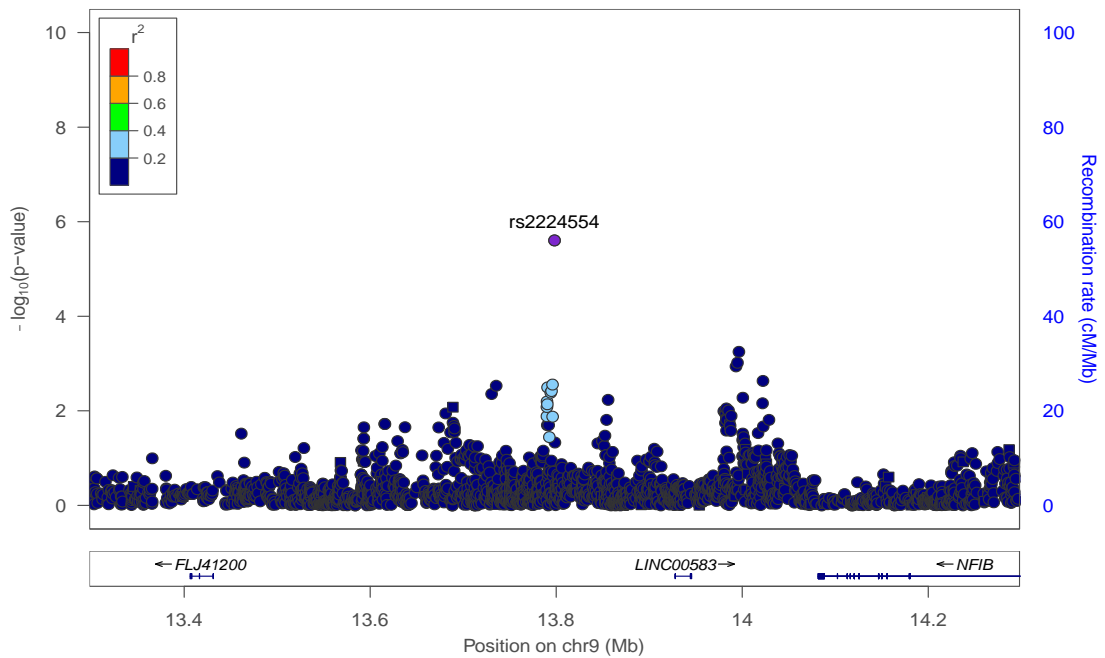
Figure A40 Regional Association Plot of rs2224554 from dental caries meta-analysis results
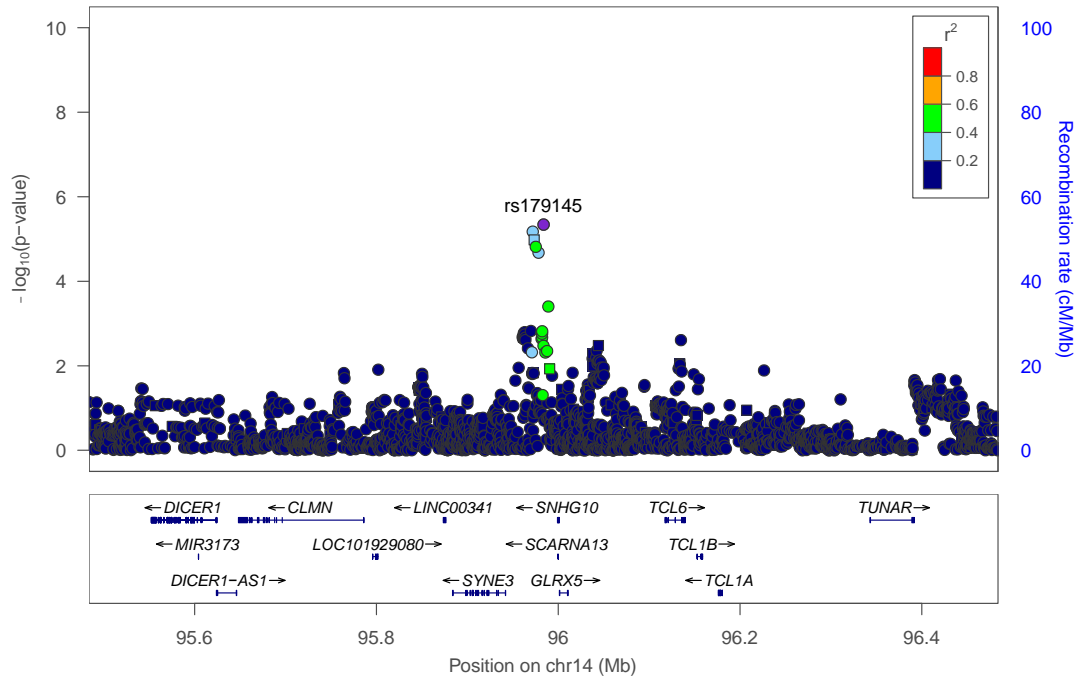
Figure A41 Regional Association Plot of rs179145 from dental caries meta-analysis results

Bibliography

Acampora, D., Merlo, G. R., Paleari, L., Zerega, B., Postiglione, M. P., Mantero, S., … Levi, G. (1999). Craniofacial, vestibular and bone defects in mice lacking the Distal-less-related gene Dlx5. *Development*.

Agod, Z., Pazmandi, K., Bencze, D., Vereb, G., Biro, T., Szabo, A., … Lanyi, A. (2018). Signaling lymphocyte activation molecule family 5 enhances autophagy and fine-tunes cytokine response in monocyte-derived dendritic cells via stabilization of interferon regulatory factor 8. Frontiers in Immunology. https://doi.org/10.3389/fimmu.2018.00062

Ahluwalia, M., Brailsford, S. R., Tarelli, E., Gilbert, S. C., Clark, D. T., Barnard, K., & Beighton, D. (2004). Dental caries, oral hygiene, and oral clearance in children with craniofacial disorders. Journal of Dental Research. https://doi.org/10.1177/154405910408300218

Aitola, M. H., & Pelto-Huikko, M. T. (2003). Expression of Arnt and Arnt2 mRNA in developing murine tissues. *Journal of Histochemistry and Cytochemistry*. https://doi.org/10.1177/002215540305100106

Al-Abdallah, M., AlHadidi, A., Hammad, M., & Dar-Odeh, N. (2018). What factors affect the severity of permanent tooth impaction? *BMC Oral Health*. https://doi.org/10.1186/s12903-018-0649-5

Al-Dajani, M. (2009). Comparison of dental caries prevalence in patients with cleft lip and/or palate and their sibling controls. *Cleft Palate-Craniofacial Journal*. https://doi.org/10.1597/08-003.1

Aren, G., Güven, Y., Tolgay, C. G., Ozcan, İ., Bayar, Ö. F., Kose, T. E., … Ak, G. (2015). The Prevalence Of Dental Anomalies In A Turkish Population. *Journal of Istanbul University Faculty of Dentistry*. https://doi.org/10.17096/jiufd.86392

Arzoo, P. S., Klar, J., Bergendal, B., Norderyd, J., & Dahl, N. (2014). WNT10A mutations account for 1/4 of population-based isolated oligodontia and show phenotypic correlations. *American Journal of Medical Genetics, Part A*. https://doi.org/10.1002/ajmg.a.36243

Baccetti, T. (1998). A controlled study of associated dental anomalies. *Angle Orthodontist*.

Bailleul-Forestier, I., Molla, M., Verloes, A., & Berdal, A. (2008). The genetic basis of inherited anomalies of the teeth. Part 1: Clinical and molecular aspects of non-syndromic dental disorders. *European Journal of Medical Genetics*. https://doi.org/10.1016/j.ejmg.2008.02.009

Beak, J. Y., Kang, H. S., Kim, Y. S., & Jetten, A. M. (2007). Krüppel-like zinc finger protein Glis3 promotes osteoblast differentiation by regulating FGF18 expression. *Journal of Bone and Mineral Research*. https://doi.org/10.1359/jbmr.070503

Beaty, T. H., Murray, J. C., Marazita, M. L., Munger, R. G., Ruczinski, I., Hetmanski, J. B., … Scott, A. F. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nature Genetics*. https://doi.org/10.1038/ng.580

Beaty, T. H., Taub, M. A., Scott, A. F., Murray, J. C., Marazita, M. L., Schwender, H., … Ruczinski, I. (2013). Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study. *Human Genetics*. https://doi.org/10.1007/s00439-013-1283-6

Becker, A. (1995). In defense of the guidance theory of palatal canine displacement. *Angle Orthodontist*.

Bennett, S. N., Caporaso, N., Fitzpatrick, A. L., Agrawal, A., Barnes, K., Boyd, H. A., … Williams, K. (2011). Phenotype harmonization and cross-study collaboration in GWAS consortia: The GENEVA experience. *Genetic Epidemiology*. https://doi.org/10.1002/gepi.20564

Boot-Handford, R. P., Tuckwell, D. S., Plumb, D. A., Farrington Rock, C., & Poulsom, R. (2003). A novel and highly conserved collagen (proα1(XXVII)) with a unique expression pattern and unusual molecular characteristics establishes a new clade within the vertebrate fibrillar collagen family. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M212889200

Brook, A. H., Griffin, R. C., Townsend, G., Levisianos, Y., Russell, J., & Smith, R. N. (2009). Variability and patterning in permanent tooth size of four human ethnic groups. *Archives of Oral Biology*. https://doi.org/10.1016/j.archoralbio.2008.12.003

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1002822

Cheng, L. L., Moor, S. L., & Ho, C. T. C. (2007). Predisposing factors to dental caries in children with cleft lip and palate: A review and strategies for early prevention. *Cleft Palate-Craniofacial Journal*. https://doi.org/10.1597/05-112

Chiquet, B. T., Lidral, A. C., Stal, S., Mulliken, J. B., Moreno, L. M., Arco-Burgos, M., … Hecht, J. T. (2007). CRISPLD2: A novel NSCLP candidate gene. *Human Molecular Genetics*. https://doi.org/10.1093/hmg/ddm176

Cholitgul, W., & Drummond, B. K. (2000). Jaw and tooth abnormalities detected on panoramic radiographs in New Zealand children aged 10-15 years. *The New Zealand Dental Journal*.

Chopra, A., Lakhanpal, M., Rao, N. C., Gupta, N., & Vashisth, S. (2014). Oral health in 4-6 years children with cleft lip/palate: A case control study. *North American Journal of Medical Sciences*. https://doi.org/10.4103/1947-2714.134371

Cornelis, M. C., Agrawal, A., Cole, J. W., Hansel, N. N., Barnes, K. C., Beaty, T. H., … Yu, K. (2010). The gene, environment association studies consortium (GENEVA): Maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genetic Epidemiology*. https://doi.org/10.1002/gepi.20492

Daneshkazemi, A. R., & Davari, A. (2005). Assessment of DMFT and enamel hypoplasia among junior high school children in Iran. *Journal of Contemporary Dental Practice*.

Davidson, D. (1995). The function and evolution of Msx genes: pointers and paradoxes. *Trends in Genetics*. https://doi.org/10.1016/S0168-9525(00)89124-6

Depew, M. J., Liu, J. K., Long, J. E., Presley, R., Meneses, J. J., Pedersen, R. A., & Rubenstein, J. L. R. (1999). Dlx5 regulates regional development of the branchial arches and sensory capsules. *Development*.

De Castilho, A. R. F., Das Neves, L. T., & De Carvalho Carrara, C. F. (2006). Evaluation of oral health knowledge and oral health status in mothers and their children with cleft lip and palate. *Cleft Palate-Craniofacial Journal*. https://doi.org/10.1597/04-205

Diekwisch, T. G. H. (2002). Pathways and fate of migratory cells during late tooth organogenesis. *Connective Tissue Research*.

Diniz, M. G., França, J. A., Vilas-Boas, F. A. S., de Souza, F. T. A., Calin, G. A., Gomez, R. S., … Gomes, C. C. (2019). The long noncoding RNA KIAA0125 is upregulated in ameloblastomas. *Pathology Research and Practice*. https://doi.org/10.1016/j.prp.2018.12.030

Dong, N., Liu, Y., Zhang, T., Zhao, L., Tian, J., & Ruan, J. (2017). Different expression patterns of Lin28 and Lin28b in mouse molar development. *Archives of Oral Biology*. https://doi.org/10.1016/j.archoralbio.2017.06.023

Du, J., Wang, Q., Yang, P., & Wang, X. (2016). FHL2 mediates tooth development and human dental pulp cell differentiation into odontoblasts, partially by interacting with Runx2. *Journal of Molecular Histology*. https://doi.org/10.1007/s10735-016-9655-6

Du, J., Wang, Q., Wang, L., Wang, X., & Yang, P. (2012). The expression pattern of FHL2 during mouse molar development. *Journal of Molecular Histology*. https://doi.org/10.1007/s10735-012-9409-z

Dubois, T., Paléotti, O., Mironov, A. A., Fraisier, V., Stradal, T. E. B., De Matteis, M. A., … Chavrier, P. (2005). Golgi-localized GAP for Cdc42 functions downstream of ARF1 to control Arp2/3 complex and F-actin dynamics. *Nature Cell Biology*. https://doi.org/10.1038/ncb1244

Edelstein, B. L. (2006). The dental caries pandemic and disparities problem. In *BMC Oral Health*. https://doi.org/10.1186/1472-6831-6-S1-S2

El-Sayed, W., Parry, D. A., Shore, R. C., Ahmed, M., Jafri, H., Rashid, Y., … Mighell, A. J. (2009). Mutations in the Beta Propeller WDR72 Cause Autosomal-Recessive Hypomaturation Amelogenesis Imperfecta. *American Journal of Human Genetics*. https://doi.org/10.1016/j.ajhg.2009.09.014

Everman, Amanda (2018) Genome-wide association study and meta-analysis of missing teeth and functional dentition. Master's Thesis, University of Pittsburgh. (Unpublished)

Ezoddini Ardakani, F., Sheikhha, M. H., & Ahmadi, H. (2007). Prevalence of dental developmental anomalies: A radiographic study. *Community Dental Health*.

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. *Human Molecular Genetics*. https://doi.org/10.1093/hmg/ddt231

Fardi, A., Kondylidou-Sidira, A., Bachour, Z., Parisis, N., & Tsirlis, A. (2011). Incidence of impacted and supernumerary teeth - A radiographic study in a North Greek population. *Medicina Oral, Patologia Oral y Cirugia Bucal*. https://doi.org/10.4317/medoral.16.e56

Figueiredo, J., Söderberg, O., Simões-Correia, J., Grannas, K., Suriano, G., & Seruca, R. (2013). The importance of E-cadherin binding partners to evaluate the pathogenicity of E-cadherin missense mutations associated to HDGC. *European Journal of Human Genetics*. https://doi.org/10.1038/ejhg.2012.159

Ferrara, N., Gerber, H. P., & LeCouter, J. (2003). The biology of VEGF and its receptors. *Nature Medicine*. https://doi.org/10.1038/nm0603-669

Ferrari, D., Sumoy, L., Gannon, J., Sun, H., Brown, A. M. C., Upholt, W. B., & Kosher, R. A. (1995). The expression pattern of the Distal-less homeobox-containing gene Dlx-5 in the developing chick limb bud suggests its involvement in apical ectodermal ridge activity, pattern formation, and cartilage differentiation. *Mechanisms of Development*. https://doi.org/10.1016/0925-4773(95)98113-O

Ferraro, M., & Vieira, A. R. (2010). Explaining Gender Differences in Caries: A Multifactorial Approach to a Multifactorial Disease. *International Journal of Dentistry*. https://doi.org/10.1155/2010/649643

Ferreira, M. A. R., & Purcell, S. M. (2009). A multivariate test of association. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btn563

Freitas, A. B. D. A., De Barros, L. M., Fiorini, J. E., Boriollo, M. F. G., Moreira, A. N., & Magalhães, C. S. (2013). Caries experience in a sample of adolescents and young adults with cleft lip and palate in Brazil. *Cleft Palate-Craniofacial Journal*. https://doi.org/10.1597/11-143

Galesloot, T. E., Van Steen, K., Kiemeney, L. A. L. M., Janss, L. L., & Vermeulen, S. H. (2014). A comparison of multivariate genome-wide association methods. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0095923

Galluccio, G., Castellano, M., & La Monaca, C. (2012). Genetic basis of non-syndromic anomalies of human tooth number. *Archives of Oral Biology*. https://doi.org/10.1016/j.archoralbio.2012.01.005

Gao, B., Zhou, X., Zhou, X., Pi, C., Xu, R., Wan, M., … Zheng, L. (2015). BMP7 and EREG contribute to the inductive potential of dental mesenchyme. *Scientific Reports*. https://doi.org/10.1038/srep09903

Garib, D. G., Peck, S., & Gomes, S. C. (2009). Increased occurrence of dental anomalies associated with second-premolar agenesis. *Angle Orthodontist*. https://doi.org/10.2319/021308-87.1

Garvey, M. T., Barry, H. J., & Blake, M. (1999). Supernumerary teeth--an overview of classification, diagnosis and management. *Journal (Canadian Dental Association)*.

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 Genetics*. https://doi.org/10.1371/journal.pgen.1002275

Gomez-Sosa, J. F., Caviedes-Bucheli, J., Diaz-Barrera, L. E., & Munoz, H. R. (2019). Gene expression of growth factors with angiogenic potential in human dental pulp tissue from teeth with complete and incomplete root development. *International Endodontic Journal*, *0*(0). https://doi.org/10.1111/iej.13188

Gong, S. G., Mai, S., Chung, K., & Wei, K. (2009). Flrt2 and Flrt3 have overlapping and non-overlapping expression during craniofacial development. *Gene Expression Patterns*. https://doi.org/10.1016/j.gep.2009.07.009

Grando Mattuella, L., de Figueiredo, J. A. P., Nör, J. E., de Araujo, F. B., & Fossati, A. C. M. (2007). Vascular Endothelial Growth Factor Receptor-2 Expression in the Pulp of Human Primary and Young Permanent Teeth. *Journal of Endodontics*. https://doi.org/10.1016/j.joen.2007.08.019

Gugasyan, R., Voss, A., Varigos, G., Thomas, T., Grumont, R. J., Kaur, P., … Gerondakis, S. (2004). The Transcription Factors c-rel and RelA Control Epidermal Development and Homeostasis in Embryonic and Adult Skin via Distinct Mechanisms. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.24.13.5733-5745.2004

Haga, S., Nakaoka, H., Yamaguchi, T., Yamamoto, K., Kim, Y. Il, Samoto, H., … Inoue, I. (2013). A genome-wide association study of third molar agenesis in Japanese and Korean populations. *Journal of Human Genetics*. https://doi.org/10.1038/jhg.2013.106

Hardcastle, Z., Mo, R., Hui, C. C., & Sharpe, P. T. (1998). The Shh signalling pathway in tooth development: Defects in Gli2 and Gli3 mutants. *Development*.

Hazza'a, A. M., Rawashdeh, M. A., Al-Nimri, K., & Al Habashneh, R. (2011). Dental and oral hygiene status in Jordanian children with cleft lip and palate: A comparison between unilateral and bilateral clefts. International Journal of Dental Hygiene, 9(1), 30–36. https://doi.org/10.1111/j.1601-5037.2009.00426.x

Hjorten, R., Hansen, U., Underwood, R. A., Telfer, H. E., Fernandes, R. J., Krakow, D., … Pace, J. M. (2007). Type XXVII collagen at the transition of cartilage to bone during skeletogenesis. *Bone*. https://doi.org/10.1016/j.bone.2007.06.024

Hong, H. K., Lass, J. H., & Chakravarti, A. (1999). Pleiotropic skeletal and ocular phenotypes of the mouse mutation congenital hydrocephalus (ch/Mf1) arise from a winged helix/forkhead transcription factor gene. *Human Molecular Genetics*. https://doi.org/10.1093/hmg/8.4.625

Hong, L., Levy, S. M., Warren, J. J., & Broffitt, B. (2009). Association between enamel hypoplasia and dental caries in primary second molars: A cohort study. *Caries Research*. https://doi.org/10.1159/000231571

Howe, B. J., Cooper, M. E., Vieira, A. R., Weinberg, S. M., Resick, J. M., Nidey, N. L., … Moreno Uribe, L. M. (2015). Spectrum of dental phenotypes in nonsyndromic orofacial clefting. *Journal of Dental Research*. https://doi.org/10.1177/0022034515588281

Howe, B. J., Cooper, M. E., Wehby, G. L., Resick, J. M., Nidey, N. L., Valencia-Ramirez, L. C., … Moreno Uribe, L. M. (2017). Dental Decay Phenotype in Nonsyndromic Orofacial Clefting. *Journal of Dental Research*. https://doi.org/10.1177/0022034517709961

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. *Journal of International Medical Research*. https://doi.org/10.1177/0300060519828450

Hu, X., Zhang, S., Chen, G., Lin, C., Huang, Z., Chen, Y., & Zhang, Y. (2013). Expression of SHH signaling molecules in the developing human primary dentition. *BMC Developmental Biology*. https://doi.org/10.1186/1471-213X-13-11

Iida, K., Koseki, H., Kakinuma, H., Kato, N., Mizutani-Koseki, Y., Ohuchi, H., … Miura, N. (1997). Essential roles of the winged helix transcription factor MFH-1 in aortic arch patterning and skeletogenesis. *Development*.

Ito, Y., Zhao, J., Mogharei, A., Shuler, C. F., Weinstein, M., Deng, C., & Chai, Y. (2001). Antagonistic Effects of Smad2 Versus Smad7 are Sensitive to Their Expression Level during Tooth Development. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M011424200

Liang, Y., Zhou, Y., Jiang, T., Zhang, Z., Wang, S., & Wang, Y. (2011). Expression of LIF and LIFR in periodontal tissue during orthodontic tooth movement. *Angle Orthodontist*. https://doi.org/10.2319/102510-622.1

Løes, S., Luukko, K., Kvinnsland, I. H., Salminen, M., & Kettunen, P. (2003). Developmentally regulated expression of Netrin-1 and -3 in the embryonic mouse molar tooth germ. *Developmental Dynamics*. https://doi.org/10.1002/dvdy.10317

Lukacs, J. R., & Largaespada, L. L. (2006). Explaining sex differences in dental caries prevalence: Saliva, hormones, and "life history" etiologies. *American Journal of Human Biology*. https://doi.org/10.1002/ajhb.20530

Jonsson, L., Magnusson, T. E., Thordarson, A., Jonsson, T., Geller, F., Feenstra, B., … Stefansson, K. (2018). Rare and Common Variants Conferring Risk of Tooth Agenesis. *Journal of Dental Research*. https://doi.org/10.1177/0022034517750109

Jontell, M., Okiji, T., Dahlgren, U., & Bergenholtz, G. (1998). Immune defense mechanisms of the dental pulp. *Critical Reviews in Oral Biology and Medicine*. https://doi.org/10.1177/10454411980090020301

Jungers, K. A., Le Goff, C., Somerville, R. P. T., & Apte, S. S. (2005). Adamts9 is widely expressed during mouse embryo development. *Gene Expression Patterns*. https://doi.org/10.1016/j.modgep.2005.03.004

Kan, S, Zhu, G, Du, Y, et al. (2019). Non-syndromic cleft lip with or without palate-susceptible SNPs is associated with hyperdontia. *Oral Dis*, 00: 1– 8. https://doi-org.pitt.idm.oclc.org/10.1111/odi.13149

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., … Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*. https://doi.org/10.1038/ng.548

Kantaputra, P. N., Chinadet, W., Ohazama, A., & Kono, M. (2012). Dyschromatosis symmetrica hereditaria with long hair on the forearms, hypo/hyperpigmented hair, and dental anomalies: Report of a novel ADAR1 mutation. *American Journal of Medical Genetics Part A*, *158A*(9), 2258–2265. https://doi.org/10.1002/ajmg.a.35488

Karimbux, N. Y., Saraiya, V. M., Elangovan, S., Allareddy, V., Kinnunen, T., Kornman, K. S., & Duff, G. W. (2012). Interleukin-1 Gene Polymorphisms and Chronic Periodontitis in Adult Whites: A Systematic Review and Meta-Analysis. *Journal of Periodontology*. https://doi.org/10.1902/jop.2012.110655

Katsura, K. A., Horst, J. A., Chandra, D., Le, T. Q., Nakano, Y., Zhang, Y., … DenBesten, P. K. (2014). WDR72 models of structure and function: A stage-specific regulator of enamel mineralization. *Matrix Biology*. https://doi.org/10.1016/j.matbio.2014.06.005

King, N. M., Wong, W. L., & Wong, H. M. (2013). Caries experience of chinese children with cleft lip and palate. The Cleft Palate-Craniofacial Journal: Official Publication of the American Cleft Palate-Craniofacial Association, 50(4), 448–55. https://doi.org/10.1597/11-133

Kirchberg, A., Makuch, A., Hemprich, A., & Hirsch, C. (2014). Dental caries in the primary dentition of German children with cleft lip, alveolus, palate. *Cleft Palate-Craniofacial Journal*. https://doi.org/10.1597/12-106

Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q., & Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics*. https://doi.org/10.1038/ng.2376

Küchler, E. C., Risso, P. A., Costa, M. de C., Modesto, A., & Vieira, A. R. (2008). Studies of dental anomalies in a large group of school children. *Archives of Oral Biology*. https://doi.org/10.1016/j.archoralbio.2008.04.003

Kurol, J., & Bjerklin, K. (1982). Ectopic eruption of maxillary first permanent molars: familial tendencies. *ASDC Journal of Dentistry for Children*.

Kurol, J. (1981). Infraocclusion of primary molars: an epidemiologic and familial study. *Community Dentistry and Oral Epidemiology*. https://doi.org/10.1111/j.1600-0528.1981.tb01037.x

Lacruz, R. S., Smith, C. E., Bringas, P., Chen, Y. B., Smith, S. M., Snead, M. L., … Paine, M. L. (2012). Identification of novel candidate genes involved in mineralization of dental enamel by genome-wide transcript profiling. *Journal of Cellular Physiology*. https://doi.org/10.1002/jcp.22965

Laganà, G., Venza, N., Borzabadi-Farahani, A., Fabi, F., Danesi, C., & Cozza, P. (2017). Dental anomalies: Prevalence and associations between them in a large sample of non-orthodontic subjects, a cross-sectional study. *BMC Oral Health*. https://doi.org/10.1186/s12903-017-0352-y

Lai, P. Y. (1997). Enamel hypoplasia and dental caries in very-low birthweight children: A case-controlled, longitudinal study. *Pediatric Dentistry*.

Christian, L., Bahudhanapati, H., & Wei, S. (2013). Extracellular metalloproteinases in neural crest development and craniofacial morphogenesis. *Critical Reviews in Biochemistry and Molecular Biology*. https://doi.org/10.3109/10409238.2013.838203

Leslie, E. J., Taub, M. A., Liu, H., Steinberg, K. M., Koboldt, D. C., Zhang, Q., … Murray, J. C. (2015). Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. *American Journal of Human Genetics*. https://doi.org/10.1016/j.ajhg.2015.01.004

Letra, A., Menezes, R., Granjeiro, J. M., & Vieira, A. R. (2009). AXIN2 and CDH1 polymorphisms, tooth agenesis, and oral clefts. *Birth Defects Research Part A - Clinical and Molecular Teratology*. https://doi.org/10.1002/bdra.20489

Letra, A., Menezes, R., Cooper, M. E., Fonseca, R. F., Tropp, S., Govil, M., … Marazita, M. L. (2011). CRISPLD2 variants including a C471T silent mutation may contribute to nonsyndromic

cleft lip with or without cleft palate. *Cleft Palate-Craniofacial Journal*. https://doi.org/10.1597/09-227

Li, Q., Yi, B., Feng, Z., Meng, R., Tian, C., & Xu, Q. (2018). FAM20C could be targeted by TET1 to promote odontoblastic differentiation potential of human dental pulp cells. *Cell Proliferation*. https://doi.org/10.1111/cpr.12426

Li, Y., Navia, J. M., & Bian, J. Y. (1996). Caries Experience in Deciduous Dentition of Rural Chinese Children 3–5 Years Old in Relation to the Presence or Absence of Enamel Hypoplasia. *Caries Research*. https://doi.org/10.1159/000262130

Lin, M., Li, L., Liu, C., Liu, H., He, F., Yan, F., … Chen, Y. P. (2011). Wnt5a regulates growth, patterning, and odontoblast differentiation of developing mouse tooth. *Developmental Dynamics*. https://doi.org/10.1002/dvdy.22550

Lin, W., Gao, L., Jiang, W., Niu, C., Yuan, K., Hu, X., … Huang, Z. (2019). The role of osteomodulin on osteo/odontogenic differentiation in human dental pulp stem cells. *BMC Oral Health*, *19*(1), 22. https://doi.org/10.1186/s12903-018-0680-6

Liu, W., Gong, Q., Ling, J., Zhang, W., Liu, Z., & Quan, J. (2014). Role of MiR-424 on angiogenic potential in human dental pulp cells. *Journal of Endodontics*. https://doi.org/10.1016/j.joen.2013.09.035

Lucas, V. S., Gupta, R., Ololade, O., Gelbier, M., & Roberts, G. J. (2000). Dental health indices and caries associated microflora in children with unilateral cleft lip and palate. *Cleft Palate-Craniofacial Journal*. https://doi.org/10.1597/1545-1569(2000)037<0447:DHIACA>2.0.CO;2

Marazita, M.L. (2015). Genetics of Orofacial Clefts and Related Phenotypes Quality Control Report, issued by the University of Washington GCC. http://www.ccdg.pitt.edu/docs/Marazita_ofc_QC_report_feb2015.pdf.

Markovic, M. (1982). Hypodontia in twins. *Swedish Dental Journal. Supplement*.
Martinez-Mier, E. A., & Zandona, A. F. (2013). The impact of gender on caries prevalence and risk assessment. *Dental Clinics of North America*. https://doi.org/10.1016/j.cden.2013.01.00

Matalova, E., Fleischmannova, J., Sharpe, P. T., & Tucker, A. S. (2008). Tooth agenesis: From molecular genetics to molecular dentistry. *Journal of Dental Research*. https://doi.org/10.1177/154405910808700715

Milgrom, P., Riedy, C. A., Weinstein, P., Tanner, A. C. R., Manibusan, L., & Brass, J. (2000). Dental caries and its relationship to bacterial infection, hypoplasia, diet, and oral hygiene in 6- to 36-month-old children. *Community Dentistry and Oral Epidemiology*. https://doi.org/10.1034/j.1600-0528.2000.280408.x

Mo, R., Freer, A. M., Zinyk, D. L., Crackower, M. A., Michaud, J., Heng, H. H. Q., … Hui, C. C. (1997). Specific and redundant functions of Gli2 and Gli3 zinc finger genes in skeletal patterning and development. *Development*.

Mobley, C., Marshall, T. A., Milgrom, P., & Coldwell, S. E. (2009). The Contribution of Dietary Factors to Dental Caries and Disparities in Caries. *Academic Pediatrics*. https://doi.org/10.1016/j.acap.2009.09.008

Mochizuki, N., Shimizu, S., Nagasawa, T., Tanaka, H., Taniwaki, M., Yokota, J., & Morishita, K. (2000). A novel gene, MEL1, mapped to 1p36.3 is highly homologous to the MDS1/EVI1 gene and is transcriptionally activated in t(1;3)(p36;q21)-positive leukemia cells. *Blood*.

Montero, M. J., Douglass, J. M., & Mathieu, G. M. (2003). Prevalence of dental caries and enamel defects in Connecticut head start children. *Pediatric Dentistry*.

Mossey, P. A. (1999). The heritability of malocclusion: part 2. The influence of genetics in malocclusion. *British Journal of Orthodontics*. https://doi.org/10.1093/ortho/26.3.195

Mutarai, T., Ritthagol, W., & Hunsrisakhun, J. (2008). Factors influencing early childhood caries of cleft lip and/or palate children aged 18 to 36 months in southern Thailand. *Cleft Palate-Craniofacial Journal*. https://doi.org/10.1597/07-017.1

Newberry, E. P., Latifi, T., & Towler, D. A. (1998). Reciprocal regulation of osteocalcin transcription by the homeodomain proteins Msx2 and Dlx5. *Biochemistry*. https://doi.org/10.1021/bi981878u

Nguyen-Ba-Charvet, K. T., Plump, A. S., Tessier-Lavigne, M., & Chédotal, A. (2002). Slit1 and slit2 proteins control the development of the lateral olfactory tract. *Journal of Neuroscience*.

Nikolopoulos, G. K., Dimou, N. L., Hamodrakas, S. J., & Bagos, P. G. (2008). Cytokine gene polymorphisms in periodontal disease: A meta-analysis of 53 studies including 4178 cases and 4590 controls. *Journal of Clinical Periodontology*. https://doi.org/10.1111/j.1600-051X.2008.01298.x

Nishikawa, S., & Kawamoto, T. (2015). Localization of core planar cell polarity proteins, PRICKLEs, in ameloblasts of rat incisors: Possible regulation of enamel rod decussation. *Acta Histochemica et Cytochemica*. https://doi.org/10.1267/ahc.14046

Nurbaeva, M. K., Eckstein, M., Devotta, A., Saint-Jeannet, J. P., Yule, D. I., Hubbard, M. J., & Lacruz, R. S. (2018). Evidence that calcium entry into calcium-transporting dental enamel cells is regulated by cholecystokinin, acetylcholine and ATP. *Frontiers in Physiology*. https://doi.org/10.3389/fphys.2018.00801

O'Reilly, P. F., Hoggart, C. J., Pomyen, Y., Calboli, F. C. F., Elliott, P., Jarvelin, M. R., & Coin, L. J. M. (2012). MultiPhen: Joint model of multiple phenotypes can increase discovery in GWAS. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0034861

Parapanisiou, V., Gizani, S., Makou, M., & Papagiannoulis, L. (2009). Oral health status and behaviour of Greek patients with cleft lip and palate. *European Archives of Paediatric Dentistry : Official Journal of the European Academy of Paediatric Dentistry*. https://doi.org/10.1007/BF03321606

Peck, S., Peck, L., & Kataja, M. (1998). Mandibular lateral incisor-canine transposition, concomitant dental anomalies, and genetic control. *Angle Orthodontist*.

Peck, S., Peck, L., & Kataja, M. (1994). The palatally displaced canine as a dental anomaly of genetic origin. *Angle Orthodontist*.

Perillo, L., Monsurrò, A., Bonci, E., Torella, A., Mutarelli, M., & Nigro, V. (2015). Genetic association of ARHGAP21 gene variant with mandibular prognathism. *Journal of Dental Research*. https://doi.org/10.1177/0022034515572190

Perrakis, A., & Moolenaar, W. H. (2014). Autotaxin: structure-function and signaling. *Journal of Lipid Research*. https://doi.org/10.1194/jlr.r046391

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 Genetics. https://doi.org/10.1371/journal.pgen.1000856

Polder, B. J., Van't Hof, M. A., Van Der Linden, F. P. G. M., & Kuijpers-Jagtman, A. M. (2004). A meta-analysis of the prevalence of dental agenesis of permanent teeth. *Community Dentistry and Oral Epidemiology*. https://doi.org/10.1111/j.1600-0528.2004.00158.x

Polk, D. E., Weyant, R. J., Crout, R. J., McNeil, D. W., Tarter, R. E., Thomas, J. G., & Marazita, M. L. (2008). Study protocol of the Center for Oral Health Research in Appalachia (COHRA) etiology study. *BMC Oral Health*. https://doi.org/10.1186/1472-6831-8-18

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. https://doi.org/10.1086/519795

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rakhshan, V. (2015). Congenitally missing teeth (hypodontia): A review of the literature concerning the etiology, prevalence, risk factors, patterns and treatment. *Dental Research Journal*. https://doi.org/10.4103/1735-3327.150286

Risley, M., Garrod, D., Henkemeyer, M., & McLean, W. (2009). EphB2 and EphB3 forward signalling are required for palate development. *Mechanisms of Development*. https://doi.org/10.1016/j.mod.2008.10.009

Robles, M. J., Ruiz, M., Bravo-Perez, M., González, E., & Peñalver, M. A. (2013). Prevalence of enamel defects in primary and permanent teeth in a group of schoolchildren from Granada (Spain). *Medicina Oral, Patologia Oral y Cirugia Bucal*. https://doi.org/10.4317/medoral.18580

Roslan, A. A., Rahman, N. A., & Alam, M. K. (2018). Dental anomalies and their treatment modalities/planning in orthodontic patients. *Journal of orthodontic science*, 7, 16. https://doi:10.4103/jos.JOS_37_18

Ryoo, H. M., Hoffmann, H. M., Beumer, T., Frenkel, B., Towler, D. A., Stein, G. S., … Lian, J. B. (1997). Stage-specific expression of Dlx-5 during osteoblast differentiation: Involvement in regulation of osteocalcin gene expression. *Molecular Endocrinology*. https://doi.org/10.1210/mend.11.11.0011

Salminen, M., Meyer, B. I., Bober, E., & Gruss, P. (2000). netrin 1 is required for semicircular canal formation in the mouse inner ear. *Development*.

Sarkar, L., Cobourne, M., Naylor, S., Smalley, M., Dale, T., & Sharpe, P. T. (2000). Wnt/Shh interactions regulate ectodermal boundary formation during mammalian tooth development. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.97.9.4520

Satoh, K., Hamada, S., & Shimosegawa, T. (2012). MSX2 in pancreatic tumor development and its clinical application for the diagnosis of pancreatic ductal adenocarcinoma. *Frontiers in Physiology*. https://doi.org/10.3389/fphys.2012.00430

Satokata, I., Ma, L., Ohshima, H., Bei, M., Ian, W., Nishizawa, K., … Maas, R. (2000). Msx2 deficiency in mice causes pleiotropic defects in bone growth and ectodermal organ formation. *Nature Genetics*. https://doi.org/10.1038/74231

Serafini, T., Colamarino, S. A., Leonardo, E. D., Wang, H., Beddington, R., Skarnes, W. C., & Tessier-Lavigne, M. (1996). Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell*. https://doi.org/10.1016/S0092-8674(00)81795-X

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. *Journal of Dental Research*. https://doi.org/10.1177/0022034512463579

Shaffer, J. R., Leslie, E. J., Feingold, E., Govil, M., McNeil, D. W., Crout, R. J., … Marazita, M. L. (2015). Caries Experience Differs between Females and Males across Age Groups in Northern Appalachia. *International Journal of Dentistry*. https://doi.org/10.1155/2015/938213

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 Research*. https://doi.org/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. *Journal of Dental Research*. https://doi.org/10.1177/0022034511422910

Sham, P. C., & Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3706

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. *Nature communications*, *10*(1), 2773. https://doi:10.1038/s41467-019-10630-1

Sigbjørn, S., Luukko, K., Kvinnsland, I. H., & Kettunen, P. (2001). Slit1 is specifically expressed in the primary and secondary enamel knots during molar tooth cusp formation. *Mechanisms of Development*. https://doi.org/10.1016/S0925-4773(01)00454-3

Simpson, M. A., Hsu, R., Keir, L. S., Hao, J., Sivapalan, G., Ernst, L. M., … Crosby, A. H. (2007). Mutations in FAM20C are associated with lethal osteosclerotic bone dysplasia (Raine syndrome), highlighting a crucial molecule in bone development. *American Journal of Human Genetics*. https://doi.org/10.1086/522240

Sone, S., Nakamura, M., Maruya, Y., Takahashi, I., Mizoguchi, I., Mayanagi, H., & Sasano, Y. (2005). Expression of Versican and ADAMTS During Rat Tooth Eruption. *Journal of Molecular Histology*, *36*(4), 281–288. https://doi.org/10.1007/s10735-005-5534-2

Song, J. S., Hwang, D. H., Kim, S. O., Jeon, M., Choi, B. J., Jung, H. S., … Choi, H. J. (2013). Comparative Gene Expression Analysis of the Human Periodontal Ligament in Deciduous and Permanent Teeth. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0061231

Stanley, B. O. C., Feingold, E., Cooper, M., Vanyukov, M. M., Maher, B. S., Slayton, R. L., … Shaffer, J. R. (2014). Genetic association of MPPED2 and ACTN2 with dental caries. *Journal of Dental Research*. https://doi.org/10.1177/0022034514534688

Subasioglu, A., Savas, S., Kucukyilmaz, E., Kesim, S., Yagci, A., & Dundar, M. (2015). Genetic background of supernumerary teeth. *European Journal of Dentistry*. https://doi.org/10.4103/1305-7456.149670

Suckling, G. W. (1989). Developmental defects of enamel--historical and present-day perspectives of their pathogenesis. *Advances in Dental Research*. https://doi.org/10.1177/08959374890030022901

Suda, N., Kitahara, Y., & Ohyama, K. (2006). A case of amelogenesis imperfecta, cleft lip and palate and polycystic kidney disease. *Orthodontics and Craniofacial Research*. https://doi.org/10.1111/j.1601-6343.2006.00337.x

Sun, Y., Huang, Y., Yin, A., Pan, Y., Wang, Y., Wang, C., … Yang, Y. (2015). Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. *Nature Communications*. https://doi.org/10.1038/ncomms7414

Sundell, A. L., Ullbro, C., Marcusson, A., & Twetman, S. (2015). Comparing caries risk profiles between 5- and 10- year-old children with cleft lip and/or palate and non-cleft controls. *BMC Oral Health*. https://doi.org/10.1186/s12903-015-0067-x

Temilola, D. O., Folayan, M. O., Fatusi, O., Chukwumah, N. M., Onyejaka, N., Oziegbe, E., … Agbaje, H. (2014). The prevalence, pattern and clinical presentation of developmental dental hard-tissue anomalies in children with primary and mix dentition from Ile-Ife, Nigeria. *BMC Oral Health*. https://doi.org/10.1186/1472-6831-14-125

Thesleff, I. (2003). Epithelial-mesenchymal signalling regulating tooth morphogenesis. *Journal of Cell Science*. https://doi.org/10.1242/jcs.00410

Thesleff, I., & Nieminen, P. (1996). Tooth morphogenesis and cell differentiation. *Current Opinion in Cell Biology*. https://doi.org/10.1016/S0955-0674(96)80086-X

Thesleff, I., Vaahtokari, A., & Partanen, A. M. (1995). Regulation of organogenesis. Common molecular mechanisms regulating the development of teeth and other organs. *International Journal of Developmental Biology*.

Thesleff, I. (2006). The genetic of tooth development and dental defects. In *American Journal of Medical Genetics, Part A*. https://doi.org/10.1002/ajmg.a.31360

Tian, Z., Lv, X., Zhang, M., Wang, X., Chen, Y., Tang, P., … Zhang, L. (2018). Deletion of epithelial cell-specific Cdc42 leads to enamel hypermaturation in a conditional knockout mouse model. *Biochimica et Biophysica Acta - Molecular Basis of Disease*. https://doi.org/10.1016/j.bbadis.2018.04.015

Thumbigere-Math, V., Foster, B. L., Bachu, M., Yoshii, H., Brooks, S. R., Coulter, A., … Somerman, M. J. (2019). Inactivating Mutation in IRF8 Promotes Osteoclast Transcriptional Programs and Increases Susceptibility to Tooth Root Resorption. *Journal of Bone and Mineral Research*. https://doi.org/10.1002/jbmr.3690

Vieira, A. R., Marazita, M. L., & Goldstein-McHenry, T. (2008). Genome-wide scan finds suggestive caries loci. Journal of Dental Research. https://doi.org/10.1177/154405910808700506

Vastardis, H. (2000). The genetics of human tooth agenesis: new discoveries for understanding dental anomalies. *American Journal of Orthodontics and Dentofacial Orthopedics : Official Publication of the American Association of Orthodontists, Its Constituent Societies, and the American Board of Orthodontics*. https://doi.org/10.1016/S0889-5406(00)70173-9

Vukelic, A., Cohen, J. A., Sullivan, A. P., & Perry, G. H. (2017). Extending genome-wide association study results to test classic anthropological hypotheses: Human third molar agenesis

and the "probable mutation effect." *Human Biology*. https://doi.org/10.13110/humanbiology.89.2.03

Wang, Q., Jia, P., Cuenco, K. T., Zeng, Z., Feingold, E., Marazita, M. L., … Zhao, Z. (2013). Association Signals Unveiled by a Comprehensive Gene Set Enrichment Analysis of Dental Caries Genome-Wide Association Studies. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0072653

Wang, T., Tamakoshi, T., Uezato, T., Shu, F., Kanzaki-Kato, N., Fu, Y., … Miura, N. (2003). Forkhead transcription factor Foxf2 (LUN)-deficient mice exhibit abnormal development of secondary palate. *Developmental Biology*. https://doi.org/10.1016/S0012-1606(03)00176-3

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*. https://doi.org/10.1186/1472-6831-12-57

Wang, X., Wang, S., Lu, Y., Gibson, M. P., Liu, Y., Yuan, B., … Qin, C. (2012). FAM20C plays an essential role in the formation of murine teeth. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M112.386862

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

World Health Organization. Oral health report. April 2012. Available from: https://www.mah.se/CAPP/Oral-Health-Promotion/WHO-Oral-Health-Fact-Sheet1/[Last accessed on 2018 October 31].

Wu, C., Dewan, A., Hoh, J., & Wang, Z. (2011). A Comparison of Association Methods Correcting for Population Stratification in Case-Control Studies. *Annals of Human Genetics*. https://doi.org/10.1111/j.1469-1809.2010.00639.x

Xu, P., Balczerski, B., Ciozda, A., Louie, K., Oralova, V., Huysseune, A., & Gage Crump, J. (2018). Fox proteins are modular competency factors for facial cartilage and tooth specification. *Development (Cambridge)*. https://doi.org/10.1242/dev.165498

Yang, J., Cai, W., Lu, X., Liu, S., & Zhao, S. (2017). RNA-sequencing analyses demonstrate the involvement of canonical transient receptor potential channels in rat tooth germ development. *Frontiers in Physiology*. https://doi.org/10.3389/fphys.2017.00455

Yi, Q., Liu, O., Yan, F., Lin, X., Diao, S., Wang, L., … Fan, Z. (2017). Analysis of senescence-related differentiation potentials and gene expression profiles in human dental pulp stem cells. *Cells Tissues Organs*. https://doi.org/10.1159/000448026

Zhang, Z., Song, Y., Zhang, X., Tang, J., Chen, J., & Chen, Y. P. (2003). Msx1/Bmp4 genetic pathway regulates mammalian alveolar bone formation via induction of Dlx5 and Cbfa1. *Mechanisms of Development*. https://doi.org/10.1016/j.mod.2003.09.002

Zeng, Z., Feingold, E., Wang, X., Weeks, D. E., Lee, M., Cuenco, K. T., … Shaffer, J. R. (2014). Genome-wide association study of primary dentition pit-and-fissure and smooth surface caries. *Caries Research*. https://doi.org/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. *Journal of Dental Research*. https://doi.org/10.1177/0022034513481976

Zhang, J., Liu, H. C. he., Lyu, X., Shen, G. H. u., Deng, X. X. i., Li, W. R. a., … Feng, H. L. a. (2015). Prevalence of tooth agenesis in adolescent Chinese populations with or without orthodontics. *The Chinese Journal of Dental Research: The Official Journal of the Scientific Section of the Chinese Stomatological Association (CSA)*.

Zhang, Y. D., Chen, Z., Song, Y. Q., Liu, C., & Chen, Y. P. (2005). Making a tooth: Growth factors, transcription factors, and stem cells. *Cell Research*. https://doi.org/10.1038/sj.cr.7290299

Zhou, X., & Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods*. https://doi.org/10.1038/nmeth.2848

Zurowski, C., Jamniczky, H., Graf, D., & Theodor, J. (2018). Deletion/loss of bone morphogenetic protein 7 changes tooth morphology and function in Mus musculus: Implications for dental evolution in mammals. *Royal Society Open Science*. https://doi.org/10.1098/rsos.170761