

GENETIC DETERMINANTS OF SHOVEL-SHAPED INCISORS AND
CARABELLI'S CUSP

BY

FATMA NUR ERBIL

BA, Ankara University, 2012
MA, Colorado State University, 2017

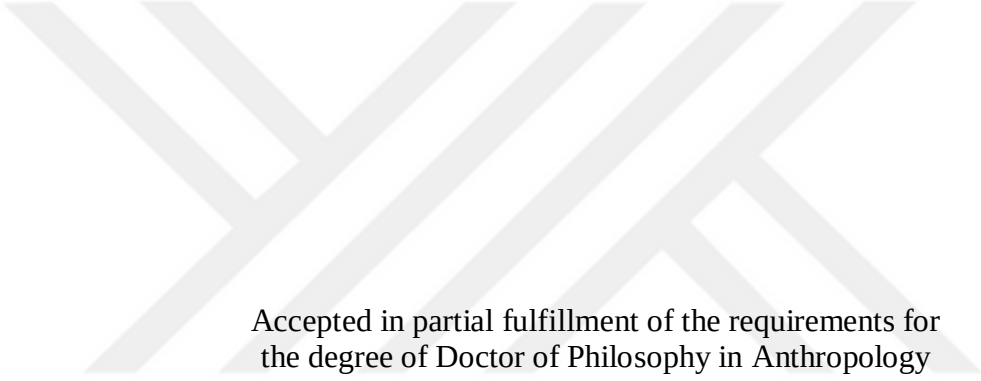
DISSERTATION

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Anthropology
in the Graduate School of
Binghamton University
State University of New York
2022



© Copyright by Fatma Nur Erbil 2022

All Rights Reserved



Accepted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Anthropology
in the Graduate School of
Binghamton University
State University of New York
2022

May 9, 2022

Dr. D. Andrew Merriwether, Chair
Department of Anthropology, Binghamton University

Dr. Elizabeth A. DiGangi, Member
Department of Anthropology, Binghamton University

Dr. Rolf Quam, Member
Department of Anthropology, Binghamton University

Dr. Thomas Powell, Outside Examiner
Department of Biology, Binghamton University

Abstract

Shovel-shaped incisors and Carabelli's cusp are two of the most commonly used dental non-metric traits in anthropological studies. Their frequencies and geographic patterns of variation have been well characterized. Although these two dental nonmetric traits have been assumed to be under strong genetic control and have long been used to differentiate between geographic human populations, genetic mechanisms that cause these two traits are not completely understood. In this study, six candidate genes *NKX2-3*, *SOSTDC1*, *BMP4*, *FGF3*, *FGF4*, and *WNT10A* which were previously identified as being involved in dental development, are investigated to identify if variation in their basal DNA sequences has any impact on whether or not an individual expresses shovel-shaped incisors and/or Carabelli's cusp. Study methods involved collection of saliva samples and dental data from 36 participants. The Arizona State University Dental Anthropology System (ASUDAS) was used to score shoveling and Carabelli's cusp expression. Next-generation sequencing (NGS) methods were utilized to sequence the entire gene region of the six candidate genes. The Spearman's correlation test is utilized to determine whether the degree of trait expression demonstrates any relationship with the presence or absence of reference and alternate alleles in the genotype of an individual. Results shows that fifteen SNPs/INDELs belonging to *SOSTDC1*, *FGF3*, *FGF4* and *WNT10A* genes are significantly associated with shoveling and/or Carabelli's cusp expression. No SNPs/INDELs were detected in the genes *BMP4* and *NKX2-3*. Due to the small sample size in this study, the conclusion may be preliminarily drawn that variation in the genes *FGF3*, *FGF4*, *SOSTDC1* and *WNT10A* might be involved in the formation of shoveling and Carabelli's cusp. However, more studies are needed to confirm the role of *NKX2-3* and *BMP4* in the production of these two dental traits.



To Sophie.

Acknowledgements

I would like to thank to my advisor Dr. Andrew Merriwether for his mentorship, support, and patience during the completion of this dissertation. I cannot express how thankful I am for Dr. Elizabeth DiGangi for her comments that made this dissertation better. I am hugely indebted to Dr. Rolf Quam for his great knowledge and encouragement he provided during my studies at Binghamton University. I would like to thank to my outside examiner Dr. Thomas Powell for taking the time to read and comment on this dissertation.

A special thanks to Joel Mora for keeping me sane during COVID-19 and always being there for me. I will be grateful forever for all your help, support, and comments you provided for this dissertation. Thank you for believing in me and making this work possible. A special thanks to Dr. Matthew Emery for his help with the bioinformatics analysis of the sequencing data. Thanks to your introduction to such a complex topic, data analysis has been smooth. I would like to thank my mom, sister and brother for the love, support, and patience overseas.

I would like to thank the Turkish Ministry of National Education for providing the financial support for my studies in the United States. Finally, I would like to thank Binghamton University's Transdisciplinary Areas of Excellence Program for funding this project.

Table of Contents

List of Figures.....	ix
List of Tables.....	xi
Chapter 1 Introduction	1
1.1 Background and Statement of the Problem	1
1.2 Significance	3
1.3 Scope and limitations	3
Chapter 2 Literature Review	5
2.1 Introduction.....	5
2.2 Terminology and Tooth Anatomy	7
2.3 Tissues of a tooth	8
2.4 Orientational Terms.....	10
2.5 Global variation of Carabelli's cusp and shovel-shaped incisors.....	11
2.6 Issues related to the use of dental traits in ancestry estimation.....	12
2.7 Embryonic tooth development	14
2.8 Patterning Cascade Model	17
2.9 Identification of genes involved in dental development	20
2.10 Pleiotropic effects of genes associated with tooth development	24
2.11 Linkage Disequilibrium	25
2.12 Candidate genes	26
2.12.1 NKX2-3.....	26
2.12.2 FGF3 and FGF4.....	28
2.12.3 WNT10A.....	30
2.12.4 BMP4.....	32
2.12.5 SOSTDC1	33
2.13 Research questions and hypotheses to be addressed	35
Chapter 3 Materials and Methods.....	37
3.1 IRB approval and Participant Recruitment	37
3.2 Dental data collection and trait scoring	39
3.3 Saliva sample collection and DNA extraction	41
3.4 Candidate gene approach.....	41
3.5 Primer Design for Long-range PCR.....	42
3.6 Polymerase Chain Reaction (PCR) Optimization and Agarose Gel Electrophoresis	44
3.7 PCR Product Purification, DNA Quantification and Amplicon Pooling.....	45

3.8 Library Preparation	46
3.9 Next-generation DNA Sequencing.....	46
3.10 Data analysis	47
3.10.1 Bioinformatics pipeline for NGS data	47
3.10.1.1 Data pre-processing (Trimming and Merging)	47
3.10.1.2 Mapping and Alignment.....	48
3.10.1.3 Removing soft-clipped reads	48
3.10.1.4 BAM cleaning and preparation (Samtools)	48
3.10.1.5 Visualization in Integrative Genomics Viewer (IGV)	49
3.10.1.6 BAM cleaning and preparation (Picard and GATK).....	49
3.10.1.7 Identifying genotype calls and creating GVCFs	50
3.10.1.8 Consolidating GVCFs and joint genotyping.....	51
3.10.1.9 Callset refinement	51
3.10.2 Spearman's Rank Analysis.....	52
3.10.3 Linkage disequilibrium analysis	54
Chapter 4 Results	55
4.1 Dental traits.....	55
4.2 Statistical analysis	57
4.3 Linkage Disequilibrium Analysis.....	63
Chapter 5 Discussion and Conclusion.....	69
5.1 Issues with INDEL Calling.....	70
5.2 Significant Single Nucleotide Polymorphisms (SNPs) and Insertion or Deletion (INDELs)	72
5.3 Linkage Disequilibrium of Variants	74
5.4 Location and consequence of variants.....	76
5.5 Conclusion	77
5.6 Future directions	79
Appendix A	80
Appendix B	86
Appendix C	96
References	100

List of Figures

Figure 2.1 Basic components of a tooth	8
Figure 2.2 Basic terms of orientation for upper and lower teeth	11
Figure 2.3 Some of the signaling molecules, receptors and transcription factors involved in tooth development	16
Figure 2.4 Stages of tooth development	17
Figure 3.1 ASUDAS plaque for shoveling in upper first (central) incisors	40
Figure 3.2 ASUDAS plaque for shoveling in upper second (lateral) incisors.....	41
Figure 3.3 ASUDAS plaque for Carabelli's cusp in upper molars	41
Figure 3.4 Visualization of positive and negative Spearman's correlations in a scatterplot	54
Figure 4.1 Distribution of Carabelli's cusp grades in upper left first molars	55
Figure 4.2 Distribution of shoveling grades in upper left central incisors	56
Figure 4.3 Distribution of shoveling grades in upper left lateral incisors	57
Figure 4.4 Locations of the variants detected in this study.....	59
Figure 4.5 Linkage disequilibrium heatmap showing correlations between SOSTDC1 variants.....	63
Figure 4.6 Five most frequent haplotypes for SOSTDC1 variants.	65
Figure 4.7 Linkage disequilibrium heatmap showing correlations between FGF3 and FGF4 variants.	66
Figure 4.8 Linkage disequilibrium results for the WNT10A variants.....	68
Figure B.1 Approval letter from the Binghamton University IRB for this dissertation project	87
Figure B.2 Informed consent form used in this dissertation. (Total of seven pages).....	94
Figure B.3 Subject information form used in this dissertation.	95
Figure C.1 Copyright permission document for using the visual in Figure 2.2.....	97
Figure C.2 Copyright permission document for using the visual in Figure 2.3.....	98

Figure C.3 Copyright permission document for using the visual in Figure 2.4.....99



List of Tables

Table 3.1 Self-reported ancestries of study participants.....	39
Table 3.2 List of the candidate genes that were sequenced in this dissertation project. ...	42
Table 3.3 Primer sequences, genomic locations, and length of amplicons.	43
Table 3.4 PCR reaction mix for a single 50 μ L reaction.....	44
Table 3.5 PCR cycling conditions.	45
Table 4.1 Spearman's rank correlation coefficients.....	59
Table 4.2 Ensemble Variant Effect Predictor results with possible consequences of the variants detected in this study.	60
Table 4.3 Genotype and allele frequencies for detected variants.....	63
Table 4.4 Linkage disequilibrium D' values for SOSTDC1 variants.	64
Table 4.5 Linkage disequilibrium r ² values for SOSTDC1 variants.	64
Table 4.6 Linkage disequilibrium D' values for FGF3 and FGF4 variants.....	67
Table 4.7 Linkage disequilibrium r ² values for FGF3 and FGF4 variants.....	67
Table A.1 Distribution of Carabelli's cusp (UM1) grades in each genotype at Chr:218886215 in WNT10A.....	80
Table A.2 Distribution of shoveling (UI1L and UI2L) grades in each genotype at rs10270805.....	80
Table A.3 Distribution of shoveling (UI1L and UI2L) grades in each genotype at rs9791629.....	81
Table A.4 Distribution of shoveling (UI1L) grades in each genotype at rs146663171. ...	81
Table A.5 Distribution of shoveling (UI1L and UI2L) grades in each genotype at rs201295694.....	81
Table A.6 Distribution of Carabelli's cusp (UM1) grades in each genotype at rs12699800.	82
Table A.7 Distribution of Carabelli's cusp (UM1) grades in each genotype at rs11761180.	82

Table A.8 Distribution of shoveling (UI1L and UI2L) grades in each genotype at rs6945425.....	82
Table A.9 Distribution of shoveling (UI2L) grades in each genotype at rs9666584.	83
Table A.10 Distribution of Carabelli's cusp (UM1) grades in each genotype at rs868979042.....	83
Table A.11 Distribution of shoveling (UI1L and UI2L) and Carabelli's cusp (UM1) grades in each genotype at rs6971172.....	84
Table A.12 Distribution of Carabelli's cusp (UM1) grades in each genotype at rs12699802.....	84
Table A.13 Distribution of shoveling (UI2L) grades in each genotype at rs7349332.	84
Table A.14 Distribution of Carabelli's cusp (UM1) grades in each genotype at Chr11:69811473.	84
Table A.15 Distribution of Carabelli's cusp (UM1) grades in each genotype at Chr11:69813652.	85

Chapter 1 Introduction

1.1 Background and Statement of the Problem

Dental development or odontogenesis is a highly regulated, multidimensional process that involves many molecular factors interacting with each other throughout embryonic and postnatal development. These include genetic, epigenetic and environmental factors, which can all significantly influence the final dental phenotype (Brook et al., 2014). Features of the tooth crown emerge as a result of this complicated process involving many genes, which are strictly controlled through a complex interplay of chemical signals between adjacent and distant cells as an organism continues through its development (Hughes & Townsend, 2013). Normal dental variation is therefore the result of polygenic and environmental influences working simultaneously. This makes parsing apart the causal biological mechanisms behind the presence or absence of a characteristic and those underlying the degree of expression of those characteristics difficult (Scott et al., 2018b).

Dental nonmetric traits are the morphological characters of the teeth that are assumed to be under strong genetic control and therefore show patterned geographic variation (Scott et al 2018b). They are usually assumed to be selectively neutral meaning that they are not adaptive and not affected by environmental stressors (Irish et al., 2020). It is due to their neutrality and strong genetic component that they have been utilized as

population markers in anthropological studies (Hanihara, 2008; Scott et al 2018b).

Shovel-shaped incisors and Carabelli's cusp that are investigated in this dissertation are the two dental traits that have received the most attention from anthropologists due to their marked geographic variation (Carayon et al., 2019). Shovel-shaped incisors or shoveling are named in this manner because the marginal ridges of the incisors are pronounced, which creates a fossa on the lingual surface of the tooth reminiscent of a shovel (Hasegawa et al., 2009). This trait is exceptionally common, almost reaching 100% in some North and East Asian and Native American populations (Hlusko et al., 2018; Scott et al., 2018b). Carabelli's cusp, on the other hand, is a trait that appears on the mesiolingual side of the upper molar and commonly observed in populations with primarily European descent (Scott & Irish, 2017). The geographic distribution of shovel-shaped incisors and Carabelli's cusp has been well documented (Hrdlička, 1920; Scott, 1980; Scott et al., 2018b; Tsai et al., 1996).

Technological advances have enabled the detection of hundreds of tooth development specific genes (Pillas et al., 2010). Studies such as genome-wide association studies (GWAS)¹, gene expression², gene knockout³, mutant and transgenic animal experiments⁴ has helped elucidate the molecular and cellular basis of not only tooth development but also development of other ectodermal organs such as hair, feathers, sweat and mammary glands (Kimura et al., 2015; Rostampour et al., 2019; Vohr & Green, 2013;

¹ Genome-wide association study (GWAS) is a genetic approach that scans the entire genome of individuals in order to identify the genetic markers that cause a disease/trait (Uffelmann et al., 2021).

² Gene expression analysis is a quantitative method for measuring the levels and types of RNA molecules to discover their transcriptional behavior in cells (Lovén et al., 2012).

³ Gene knockout refers to the method used to permanently inactivate the function of genes in an organism to observe the consequences of the removal of the genes (Shaffer, 2021).

⁴ Mutant and transgenic animal experiments refer to the techniques that either remove or introduce foreign DNA sequences to the genome of the animals (Boverhof et al., 2011).

Wang & Fan, 2011). Previous studies have demonstrated that dental nonmetric traits arise through a complex interplay of gene induction and silencing processes (i.e., activation and deactivation of a gene, respectively) that take place during development (Jernvall & Thesleff, 2000; Jussila & Thesleff, 2012; Scott et al., 2018b). The stages of embryonic dental development involve evolutionarily conserved signaling molecules and pathways between the dental epithelium and the underlying mesenchyme which are the two tissue layers involved in tooth development. Today, we know that more than three hundred genes are expressed during odontogenesis (Thesleff, 2006). However, we still do not know which genes are involved in the formation of the specific nonmetric dental characters. Due to the need for further research on this matter in the literature, this dissertation seeks to identify the genes that play a role in this complex process. Most specifically, this project seeks to find the genes that cause shovel- shaped incisors and Carabelli's cusp.

1.2 Significance

The study at hand seeks to provide an initial investigative framework for understanding the genetic and genomic regulatory mechanisms which induce the production of two anthropologically and forensically relevant discrete dental traits: shovel-shaped incisors and Carabelli's cusp. Data generated in this project will inform future research that assumes dental nonmetric traits are under strong genetic control. This data generated by this project will help us understand which genes influence tooth crown morphology during development and the genetic basis of geographic differences of tooth morphology.

1.3 Scope and limitations

This study was conducted during the peak of COVID-19 pandemic between May

2020 and 2022. The study methods had to be modified multiple times to conform with the pandemic restrictions. Therefore, the number of study volunteers fell under the anticipated number of participants. Additionally, although more than three hundred genes play roles during odontogenesis; this study investigates only six of them.



Chapter 2 Literature Review

This chapter provides a review of the literature regarding the global variation, embryonic and postnatal development, and various genetic aspects of teeth with a focus on shovel-shaped incisors and Carabelli's cusp. The chapter starts with a brief introduction that provides definitions of these two dental traits and that describes the general contexts in which these two traits have been utilized by previous studies. Anatomical terminology of teeth, directional terms and process of embryonic tooth development as it relates to the formation of dental non-metric traits are also summarized. The patterning cascade model (PCM) as an explanatory mechanism for the dental morphological variation and cusp formation in mammals is described with a focus on Carabelli's cusp variation. Studies that identified the genetic factors that contribute to the dental morphological variation in humans, and the candidate genes and their role in tooth development is reviewed. Finally, the research questions and hypotheses in this dissertation are introduced at the end of the chapter.

2.1 Introduction

Analysis of tooth morphology in living people and dental remains collected from archaeological and forensic contexts provides the evidentiary basis for answering a wide range of anthropological research questions. This includes questions pertaining to forensic

investigations that aim to determine a person's identity, research into biomedical-anthropological genetics of living populations, the identification and classification of living and extinct primate species, and in reconstructing phylogenetic relationships of humans among and between different populations (Delgado et al., 2019; Gross & Edgar, 2019; Hanihara, 2008; Irish, 1997, 1998; Scott, 1980; Scott et al., 2018a).

Morphological characteristics under investigation fall into one of two categories: non-metric and metric traits. Dental non-metric traits are heritable small details in the shape of the tooth crown that demonstrate a significant degree of variation within and between populations (Scott et al., 2018b). Non-metric dental variation has long been used as a population marker to identify and differentiate between major human populations (Haneji et al., 2007; Hanihara, 2008; Rathmann et al., 2017; Rathmann & Reyes-Centeno, 2020).

Carabelli's cusp and shovel-shaped incisors, dental traits under investigation in this dissertation, are two of the most commonly used dental non-metric traits in biodistance studies that measure the relatedness of different human populations (Guatelli-Steinberg et al., 2013; Hanihara, 2008). Carabelli's cusp was first defined and named by an Austrian dentist, Georg Carabelli, in 1842 (Bhavyaa et al., 2020). This trait occurs on the mesiodistal aspect of the cusp protocone in the upper molars and shows a wide range of expression from minor grooves to a large free-standing cusp (Figure 3.3) (Scott et al., 2018b; Scott & Irish, 2017). Carabelli's cusp is more commonly present in the upper first molars, but it can be present on the second and third molars with a decreased frequency and degree of expression.

The other dental trait of interest, shovel-shaped incisors, is characterized by the

development of distal and mesial marginal ridges on the lingual surface of both upper and lower central and lateral incisors and canines (Figure 3.1 and 3.2) (Scott & Irish, 2017). Scoring shoveling expression on the lower incisors and both upper and lower canines has been shown to be problematic (Scott and Irish, 2017). Therefore, upper central and lateral incisors are usually considered as the key teeth for shoveling expression.

2.2 Terminology and Tooth Anatomy

This section reviews the terms and concepts that are specific to adult human permanent teeth. In the field of dental anthropology, permanent tooth names (i.e., incisor, canine, premolar, and molar) are commonly abbreviated in capital letters. The letter “I” for incisors, “C” for canines, “P” for premolars, and “M” for molars are used to identify each tooth class (White & Folkens, 2005). Primary teeth, also known as deciduous, baby or milk teeth, are usually depicted as lowercase to distinguish them from the permanent teeth. It is also a common practice to add a lowercase letter “d” in front of the letter abbreviation that identifies each tooth such as “dm” for deciduous molar and “di” for deciduous incisor (White & Folkens, 2005).

Figure 2.1 illustrates the basic components of a human tooth. The **anatomical crown** of a tooth is the part that is covered by enamel and visible to the naked eye inside of a person’s oral cavity. The **gingiva**, also called **gums**, are the pink-colored mucosal tissue that cover the alveolar bone of the maxilla and mandible at the base of the teeth in the oral cavity. The **root** is the part that is embedded in the jawbone and is covered with a thin layer of cementum. The point where the crown and the root meet is called the **neck** or **cervix** (Scheid & Weiss, 2011; White & Folkens, 2005).

2.3 Tissues of a tooth

The tooth is composed of four major tissues: enamel, dentin, cementum and pulp. The **enamel** is the external layer of the crown and is primarily composed of an inorganic mineral matrix called calcium hydroxy-apatite. Dental enamel is the hardest tissue in the body with its 96% mineralized and calcified content (Guatelli-Steinberg, 2016; Scheid & Weiss, 2011). The remaining 4% is composed of water and organic material. Enamel is formed during a process called amelogenesis at the late bell stage of embryonic tooth formation (Figure 2.4). Specialized cells called ameloblasts differentiate from dental epithelial cells and then secrete the enamel matrix (Antoine & Hillson, 2015). Enamel is already partly mineralized (around 30%) when it is initially secreted by the ameloblasts (Antoine & Hillson, 2015). As the maturation continues, the amount of calcium hydroxy-apatite crystals reaches around 96%. For this reason, enamel is highly mineralized when it reaches maturity.

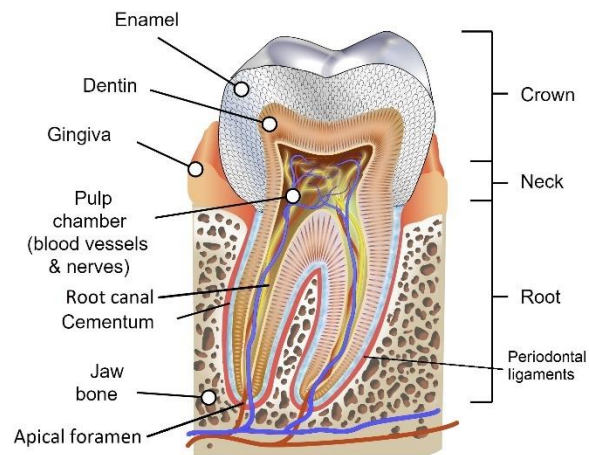


Figure 2.1 Basic components of a tooth (Figure by KDS4444 - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=52624520>. Minor changes were made on the figure).

Dentin is the hard tissue that underlies the enamel of the tooth crown and the cementum of the root. It surrounds the pulp cavity/chamber (Tang et al., 2016) and makes up the major bulk of a tooth. Dentin is composed of 70% inorganic material (calcium hydroxyapatite), 18% organic material (collagen fibers) and the remaining 12% is water (Tang et al., 2016). It is less mineralized and brittle and more elastic than enamel which makes it more resistant to fracture by absorbing the stress and force during mastication (Tang et al., 2016). Dentin originates from dental papilla cells (i.e., condensed mesenchymal cells that eventually form the dentin and pulp of the tooth), which originate in the mesoderm. These cells give rise to odontoblast cells that form dentin during a process called dentinogenesis.

Cementum or **cement** is the thin layer that encapsulates the root. It is formed by a type of specialized cells called cementoblasts and composed of 65% inorganic material (calcium hydroxyapatite), 35% organic matter (collagen fibers) and the remaining 12% is water (Scheid & Weiss, 2011; Tang et al., 2016). It functions as an anchorage for the periodontal ligaments, and it attaches the tooth to the alveolar bone of the jaw.

Pulp is the soft connective tissue that is not calcified or mineralized unlike enamel, dentin and cementum (Scheid & Weiss, 2011). Pulp is formed by embryonic dental papilla cells, which originate from mesoderm. Blood vessels within the pulp tissue provide dentin with nutrition through the odontoblast cells within the pulp, and nerves that extend into the dentin have sensory functions, which, for example, generate sensations of pain caused by decay, infection, temperature changes etcetera (Legge & Hardin, 2015). In addition to these nutritional and sensory functions, the pulp produces secondary dentin throughout a person's life via the layer of odontoblast cells within the pulp tissue (Legge & Hardin,

2015). The pulp tissue is contained within the tooth in a space called the **pulp cavity**. The pulp cavity has two main portions: **pulp chamber** of the crown and **pulp canal** or **root canal** of the root. The **apical foramen** is the opening at the root tip that allows blood vessels and the nerves to enter the pulp cavity.

2.4 Orientational Terms

Below is a list of orientational terms that are commonly used in the field of dental anthropology and this dissertation. Descriptions of the terms were adapted from Hillson (1996) and Scheid & Weiss (2011). Please see Figure 2.2 for a visual representation of the terms.

- **Mesial** refers to the surface that is closest to the midline (i.e., line that divides the skull into right and left halves) where the central incisors contact each other.
- **Distal** is used as the exact opposite of mesial: the distal surface of a tooth is away from the midline. The mesial surface of a tooth contacts or faces the distal surface of the neighboring tooth in the same jaw.
- **The lingual** surface is known as the surface that faces the tongue, and **labial** is the opposite of lingual which means toward the lips. The term labial is exclusively used for incisors and canines.
- The term **buccal** means toward the cheeks and is also used as the opposite of lingual. However, buccal is used for premolars and molars only.
- **The occlusal** surface of a tooth refers to the surface that contacts the teeth in the opposing jaw during mastication.
- The **incisal** is used to define the cutting surface of the incisor teeth (definition was taken from merriam-webster.com).

- **The apical surface** is toward the tip of the root which is the exact opposite of the occlusal aspect of a tooth.

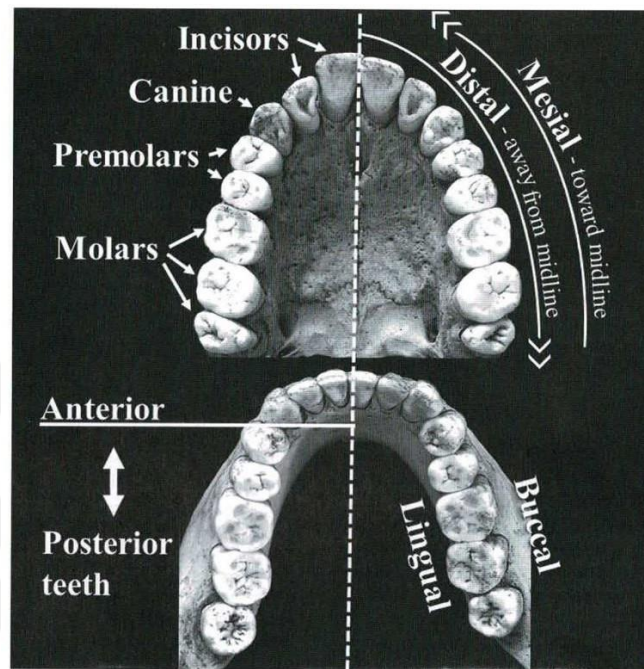


Figure 2.2 Basic terms of orientation for upper and lower teeth. Figure from Scott and Irish (2017) was used with permission from Cambridge University Press. See Appendix C for the permission document.

2.5 Global variation of Carabelli's cusp and shovel-shaped incisors

The global variation of Carabelli's trait displays a more complicated picture than previously thought. Although this trait has been assumed to be a diagnostic feature of European populations in the early 1900s, later studies have shown that it is also present in other human populations (Scott & Irish, 2017). The frequency of this trait has been shown to be 40-50% in Eastern Europeans and more than 50% in Western Europeans (Scott & Irish, 2017). Although Carabelli's trait is also observed in other major human populations, Europeans show a unique pattern of correlation between total trait frequency and mean degree of expression (Scott & Irish, 2017). For instance, Carabelli's trait is frequently present among Native Americans, but full cusp forms of this trait are not common among

these populations. Similarly, Pacific populations exhibit the cusp form of this trait although the total trait frequency is relatively low compared to European populations (Scott & Irish, 2017).

Unlike Carabelli's cusp, the world variation of shovel-shaped incisors is quite patterned: East and North Asians and Native Americans exhibit exceptionally high frequencies of this trait. In some Native American populations, this trait is almost invariant where every individual exhibits some degree of shoveling expression (Scott et al., 2018b). Western Eurasians, on the other hand, fall at the low end of the spectrum showing very low frequencies of shoveling (10-20%). The remaining groups that were examined fall in between these two extremes of the global range of shoveling expression (Scott et al., 2018b).

2.6 Issues related to the use of dental traits in ancestry estimation

Historically, anthropologists have routinely used dental metric and nonmetric traits in estimating biological race, or with a more euphemistic term "ancestry" (Hanihara, 2008; Hrdlička, 1920). Based on the assumption that dental traits reflect global population structure, contemporary forensic anthropologists rely on these traits in their calculations of ancestry estimation. However, DiGangi and Bethard (2021) state that the majority of anthropologists tend to ignore how much structural racism is embedded in some of the practices/methods that they have been traditionally employing without even questioning their validity. DiGangi and Bethard (2021) further argue that forensic anthropologists contribute to perpetuation of biological race concept and the ideology of white supremacy by practicing ancestry estimation using dental and skeletal traits, although they may not be purposefully doing so. It is not to say that forensic anthropologists explicitly and

intentionally promote the concept of biological race when they do ancestry estimation. Instead, when forensic anthropologists need to write reports for law enforcement about an unidentified person's ancestry, intentionally or not, they normalize and validate the use of biological race concept and ignore the structural racism embedded in their practice of ancestry estimation.

Anthropological studies that deal with patterns of human dental variation seek to characterize geographic regions based on trait frequencies (Hanihara, 2008). For example, the term Sinodont is used to describe the dental pattern of East and North Asian populations (Scott et al., 2016). Another major Asian dental complex, called Sundadonty, has been commonly observed among Southeast Asians and Pacific populations and usually used as the opposite of Sinodonty. In such studies, populations are usually imagined as being morphologically very different from each other as if they are mutually exclusive with no contact, similarity, or admixture with each other and this is inherently problematic.

It should be noted here that patterns of global dental variation exist, but this pattern does not equate to the socially constructed racial categories. Race is a social construct and should not have anything to do with the concept of human biological variation. When anthropologists categorize populations based on their tooth morphology without being critical of the typological approach they employ, their work explicitly or implicitly may support the misconception that biological race is real (DiGangi and Bethard, 2021; Bethard and DiGangi, 2020). Therefore, anthropologists should be aware of the issues related to the typological approaches to human variation.

Similar to the patterns of global dental variation, patterns of global genetic variation exist as well. In other words, frequency of genetic markers show population-specific

distributions. However, this pattern cannot be used to promote the biological race concept. Again, race is a socially constructed term and there is not a one-to-one relationship between biological race and patterns of global genetic variation.

Although the reports of dental trait frequencies in major human populations might be useful for description purposes, it is important to investigate how the global variational patterns occur rather than just to describe the differences and similarities between populations. As long as we do not try to explain the how and why questions (i.e. how these traits form and why), only reporting the frequencies of trait expression and placing human populations in arbitrary categories will lead to this problematic typological approach where researchers describe human populations for the sake of describing them (DiGangi and Bethard, 2021). In other words, investigating the causal microevolutionary mechanisms (i.e., mutation, selection, genetic drift, and gene flow) that lead one population to have higher frequencies of a certain trait than the other populations are what future studies such as this one should focus on in order to explore how the observed global variation come about and what kind of evolutionary forces shape this global variation we observe today.

2.7 Embryonic tooth development

Most of our knowledge regarding the process of tooth development comes from the genetic studies conducted on mice (Peterkova et al., 2014; Thesleff, 2014). Although we have gained an immense amount of knowledge from such studies, the mouse dentition is different from that of humans in many aspects. For instance, mice teeth are reduced to one incisor, no canines, no premolars and three molars in each quadrant with the dental formula of 1.0.0.3/1.0.0.3. Numbers in the dental formula indicate the total number of teeth in each tooth class: incisors, canines, premolars, and molars respectively. Left and right sides of

the forward slash refer to the teeth in upper and lower quadrants of the mouth (Plikus et al., 2005). In contrast, humans possess two incisors, one canine, two premolars and three molars with the adult dental formula of 2.1.2.3/2.1.2.3 (Catón & Tucker, 2009; Plikus et al., 2005). In addition, the mouse incisors and molars are separated by a diastema (i.e., gap that form between teeth), and mice grow only one generation (i.e., set) of teeth as opposed to two generations in humans. Despite these differences, many studies have demonstrated that mostly the same sets of genes and molecules are involved in embryonic tooth development across many mammal species including voles (Keränen et al., 1998) and hamsters (Karg et al., 1997) owing to the evolutionary conservative nature of homologous organ development (Thesleff, 2014).

Odontogenesis or tooth development starts at around embryonic day 11.5 (E11.5) in mice (Yu & Klein, 2020) and 6th week of intrauterine life in humans (Koussoulakou et al., 2009; Tucker & Sharpe, 2004). The whole developmental process involves oral epithelium, which is derived from the embryonic tissue known as ectoderm, and ectomesenchyme. The cells of these two tissues stay in communication with each other through the expression of members of many signaling families including bone morphogenetic protein (BMP), fibroblast growth factor (FGF), sonic hedgehog (Shh), and wingless-related (Wnt) signaling families (Figure 2.3) (Brook et al., 2014; Jernvall & Jung, 2000).

Thickening of the oral epithelium to form the dental lamina is the earliest evidence of tooth development. The dental lamina is a horseshoe-shaped condensed epithelial cell layer which reflects the shape of the future dental arches. During this thickening stage, the number and location of each tooth on the dental lamina is determined through a series of

complex molecular and cellular interactions (Koussoulakou et al., 2009). As odontogenesis advances, teeth go through a series of developmental stages which are named according to the shape of the tooth germ: sequentially, the bud, cap and bell stages (Figure 2.4).

During the bud stage at around embryonic day 13 (E13) in mice, epithelial cells proliferate and invaginate into the underlying mesenchyme forming a bud shape (Catón & Tucker, 2009). At this stage, mesenchymal cells underneath the epithelial cells start to condense. At the onset of bud to cap stage, a signaling center at the tip of the bud forms (Tucker & Sharpe, 2004). This signaling center is called the primary enamel knot (pEK) (Ahtiainen et al., 2016). The cells consisting of the enamel knots are non-dividing epithelial cells that do not proliferate but induce the proliferation of the surrounding cells (Jernvall et al., 1994). As the morphogenesis proceeds, secondary enamel knots (sEK) form in multicuspid teeth (i.e., premolars and molars). During the cap (E14) and bell stages (E16), the enamel knots control the whole process of cusp development and determine the final size and shape of the tooth crown (Jernvall et al., 1994).

	INITIATION	MORPHOGENESIS	
SIGNALLING MOLECULES	Activin beta A; Bmp 2, 4, 7; Eda; Fgf 1, 2, 4, 8, 9; Shh; Wnt 3, 4, 5, 6, 7b, 10(a,b)	Activin beta A; Bmp 2, 4, 6, 7; Eda; Fgf 3, 4, 8, 9, 10; Shh; Wnt 4, 5a, 6, 7b, 10(a,b)	Activin beta A; Bmp 2, 3, 6, 7; Eda; Fgf 1, 2, 3, 4, 8, 9, 10; Shh; Wnt 3, 4, 5a, 6, 7b, 10(a,b)
RECEPTORS	Edar; Fgfr 1, 2; Notch 1, 2, 3; Patched 1	Edar; Fgfr 1, 2; Notch 1, 2, 3; Patched 1, 2	Edar; Fgfr 1, 2; Notch 1, 2; Patched 1, 2
TRANSCRIPTION FACTORS	Lef 1; Msx 1, 2; Pax-9; Runx 2; Epiprotein/Sp 6	Catenin beta; Lef 1; Msx 1, 2; Pax-9; Runx 1, 2, 3; Epiprotein/Sp 6	Catenin beta; Runx 1, 2, 3; Epiprotein/Sp 4, 6

Figure 2.3 Some of the signaling molecules, receptors and transcription factors involved in tooth development. Figure adapted from Brook et al. (2014) with permission from John Wiley and Sons. See Appendix C for the permission document.

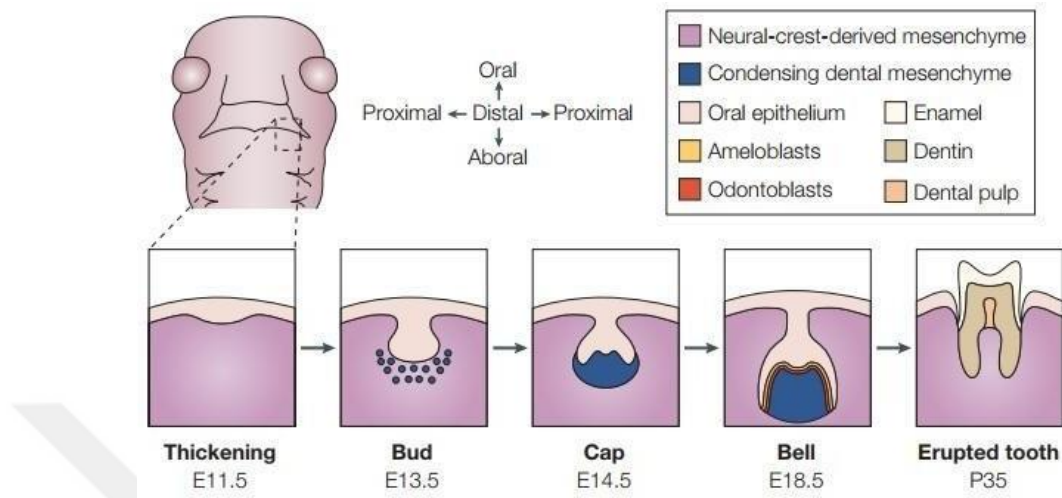


Figure 2.4 Stages of tooth development. E, embryonic day; P, postnatal day in mice. Figure adapted with permission from Springer Nature. See Appendix C for the permission document.

Studies on monozygotic (MZ) and dizygotic (DZ) twins have demonstrated that tooth morphology is highly heritable and under strong genetic control (Hughes & Townsend, 2013; Scott et al., 2018b). However, there have been many instances where discordances in dental trait expressions between both MZ and DZ twin pairs have been observed (Dempsey et al., 1995; Townsend et al., 2005). Since MZ twin pairs share 100% of their DNA, these discordances usually have been attributed to the environmental and epigenetic events that occur during the embryonic development of the dentition (Townsend et al., 2009).

2.8 Patterning Cascade Model

A morphodynamic model for tooth formation, namely the Patterning Cascade Model (PCM), was proposed as an evolutionary developmental framework to explain the dental morphological variation in mammals (Jernvall & Jung, 2000; Salazar-Ciudad & Jernvall, 2002). The model suggests that the enamel knots, which are the sites of future

cuspid tips, control the formation of other enamel knots in a tooth through secretion of activator and inhibitor molecules. As the activators induce the formation of new enamel knots in a developing multicuspid tooth, the inhibitors around the cells expressing activator genes will block the formation of new enamel knots by forming an inhibition zone. Inhibitors will prevent new enamel knots from forming adjacent to the previously formed enamel knot but will promote proliferation of mesenchymal cells. As the tooth development proceeds, the differential rate of epithelial and mesenchymal cell proliferation will promote epithelial folding. In order for a new enamel knot to form, it must escape from the inhibition zone (Salazar-Ciudad & Jernvall, 2002).

According to the PCM, relative intercuspid distances as measured by the distance between the cuspid tips, overall crown size and shape, the number of cusps, and the time duration of tooth development are all interdependent variables and influence each other during tooth development (Hunter et al., 2010). The PCM predicts that if the size of the tooth germ is relatively large, and the previously forming cusps are closely located to each other, then there will be enough space for more cusps to form. In addition, if the tooth development is prolonged, the tooth will have sufficient time to develop more cusps. The PCM is based on the idea that the interaction between gene networks, timing and spatial context all contribute to the final phenotype of the tooth crown rather than genes alone (Hunter et al., 2010).

The predictive power of the PCM has been tested on the Carabelli's cuspid expression in humans by Hunter et al., 2010. Hunter and colleagues (2010) hypothesized that if a Carabelli's trait enamel knot is involved during the development of this trait, this trait should be observable at the dentine-enamel junction (DEJ) (i.e., the dentin surface of the

tooth when the enamel is removed). Assuming that the cuspal expression of this trait forms via the enamel knot signaling activity, then the PCM would be expected to successfully predict manifestations of the Carabelli's cusp trait (Guatelli-Steinberg et al., 2013).

Hunter et al. (2010) found that if the distances between the earlier forming four main cusps of the upper molars are small, a Carabelli's cusp will likely form after the four main cusps are formed. On the contrary, if the distances between the four main cusps are too large, there will not be enough space for an extra cusp to form. These results support the predictive power of the PCM. Moormann et al. (2013)'s study also tested the ability of the PCM to predict the presence of Carabelli's cusp in the first (M1) and second (M2) permanent molars in adults. In agreement with the PCM's prediction that teeth with larger total crown sizes are expected to develop more cusps than teeth with smaller total crown sizes, Moorman and colleagues (2013) found that M1 is more likely to express Carabelli's trait than M2 due to the larger crown size of M1s compared to M2s. This study is significant in that it confirms the PCM's predictive power by predicting a well-known fact that M1 exhibits Carabelli's cusp more frequently than M2.

On the other hand, Paul et al. (2016)'s results showed mixed support for the PCM in the mixed dentition with permanent first molars (M1) and deciduous second molars (dm2). They found that dm2s exhibit bigger Carabelli's cusp than M1s, a result both contradicts and supports the PCM's predictions. It contradicts because since the overall crown size of M1s is larger than dm2s, it would be expected that the M1s would exhibit bigger Carabelli's cusp. On the other hand, the results are not surprising because if we consider dm2s as the anteriormost member of the molar district, it is commonly observed that the grade of Carabelli's trait expression decreases toward the posterior end of the molar

district (i.e. M3). Carabelli's cusp is often called a "cusp", but it takes many forms and shapes from a slight pit or a groove to a full-sized cusp as large as the other four main cusps of a tooth (Guatelli-Steinberg et al., 2013; Scott et al., 2018b). Due to the variable nature of this trait, developmental mechanisms that form a Carabelli's cusp might be slightly different than the mechanisms that produce the main cusps of the tooth. In other words, the PCM might not be the only underlying mechanism that forms this trait. Another possibility is that different independent cascades of gene expression are at play for different parts of the crown (Jernvall & Jung, 2000). Paul et al. (2016) mention that the conflicting results in their study might be due to the fact that Carabelli's trait is not always in a cuspal form. Earlier studies observed that M1 exhibits a thicker enamel than dm2 (Grine, 2005). Therefore, features that are visible at the DEJ might not be visible at the outer enamel surface (OES) due to the thick enamel (Bailey et al., 2011; Skinner et al., 2008).

2.9 Identification of genes involved in dental development

Technological advances have enabled the detection of hundreds of tooth development specific genes (Pillas et al., 2010). Studies such as genome-wide association studies (GWAS), gene expression, gene knockout, mutant and transgenic animal experiments has helped elucidate the molecular and cellular basis of not only tooth development but also development of other ectodermal organs such as hair, feathers, sweat and mammary glands (Kimura et al., 2015; Rostampour et al., 2019; Vohr & Green, 2013; Wang & Fan, 2011). Previous studies have demonstrated that dental nonmetric traits arise through a complex interplay of gene induction and silencing processes that take place during development (Jernvall & Thesleff, 2000; Jussila & Thesleff, 2012; Scott et al., 2018b).

Previous studies have shown that the genes that direct the development of teeth are highly evolutionarily conserved and therefore they should be understood as controlling many downstream developmental pathways, not just in odontogenesis but in the development of other tissue types as well. Any disruptions during embryonic development, such as those caused by genetic mutations, can result in a range of possible outcomes from mild phenotypic variations to severe dental defects (Salazar-Ciudad & Jernvall, 2004; Tucker & Sharpe, 2004). Disorders or diseases are usually polygenic but are ultimately caused by one or a small number of mutations in one or several genes with simple modes of inheritance that “tip the scales” toward the aberrant phenotype. This is not the case for most non-metric dental traits (Hughes & Townsend, 2013). Normal dental variation is a result of polygenic and environmental influences working simultaneously and the genetic underpinnings cannot be parsed within a simple present or non- present framework, as can be done with many diseases. In addition to variation within the basal genetic sequence of multiple genes, epigenetic factors that influence the regulation of gene expression during tooth crown development have also been previously noted to influence dental phenotypes (Cao et al., 2010; Lin et al., 2018; Oommen et al., 2012).

With the increasing application of genome-wide association studies (GWAS) on many dental phenotypes, several loci affecting primary tooth development have been identified. In one such study, Pillas et al. (2010) analyzed two birth cohorts and looked for genetic determinants of age at first tooth eruption and number of teeth at the age of one year. They found several loci that regulate the primary tooth development including KCNJ2, EDA, RAD51L1, IGF2BP1, HOXB2 and MSRB3. Mutations in many of the identified genes have been previously associated with altered dental development (Mikkola

& Thesleff, 2003). Interestingly, the genes, RAD51L1 and IGF2BP1, have been previously associated with breast cancer (Thomas et al., 2009) and ovarian cancer (Köbel et al., 2007), respectively. A shared genetic link between abnormal tooth development and colorectal cancer has been previously pointed out as well (Lammi et al., 2004).

The ectodysplasin A receptor (EDAR) gene is known for its role in the formation of ectodermal organs including teeth, hair, sweat and mammary glands (Hlusko et al., 2018; Kamberov et al., 2013; Tucker et al., 2000, 2004). Mutations in this gene have been shown to cause hypohidrotic ectodermal dysplasia which results in abnormal development of teeth, hair and sweat glands (Mikkola, 2009). A nonsynonymous variant, T1540C (rs3827760), in the EDAR gene has been detected in high frequencies in East Asian populations and has been associated with Asian specific hair thickness (Park et al., 2012; Tan et al., 2014). Kimura et al. (2009) aimed to understand whether this variant is also associated with shovel-shaped incisors in Japanese individuals. They found that the number of copies of the T1540C allele was significantly associated with shoveling and double shoveling in incisors. This allele explained 18.9% of the total variance in the expression of shoveling in this population. The heritability of this trait has been previously calculated as 0.75. Therefore, this allele is responsible for almost 25% of the previously calculated heritability (i.e., 0.75) for this trait.

PAX9 (paired box 9) is a member of the PAX gene family which encodes for transcription factors that are involved in early development (Chhabra et al., 2014; Klein et al., 2013; Suryadeva & Khan, 2015). This gene is expressed in dental mesenchyme during tooth development. Mutations in this gene has been associated with congenital absence of posterior teeth including third molars (Miletich & Sharpe, 2003). Lee et al. (2012)

examined five SNPs in the PAX9 gene in Japanese and Korean individuals to determine if any of these SNPs were associated with third molar agenesis, crown size and nonmetric dental traits. They found that four of the SNPs they examined were associated with crown size and they were in strong linkage disequilibrium indicating that the expression of these variants is not independent during tooth development. They also found that three of the five SNPs were associated with shoveling of upper central incisors (Lee et al., 2012).

The genetics of tooth agenesis or congenitally missing teeth have been well documented (Chhabra et al., 2014; Klein et al., 2013; Miletich & Sharpe, 2003). Non-syndromic hypodontia, i.e., absence of one to six teeth excluding third molars, and oligodontia, i.e., absence of six or more teeth excluding third molars, are the most common types of congenital tooth agenesis (Chhabra et al., 2014). Several genes including MSX1 (Hu et al., 1998), PAX9 (Mostowska et al., 2006; Nieminen et al., 2001), AXIN2 (Lammi et al., 2004) and EDA (Callea et al., 2014; Han et al., 2008) causing hypodontia and oligodontia have been previously identified. MSX1 is a highly conserved gene that is expressed in tissues of multiple organs in vertebrates. Five point mutations have been identified in this gene that might be causing hypodontia but the severity of the effects of them is unknown (Chhabra et al., 2014). PAX9 is a paired-box domain containing transcription factor that regulates the expression of downstream genes by binding to their enhancer sequences. Several mutations that cause a change in the amino acid sequence and alter the protein product have been discovered to cause hypodontia in both permanent and primary teeth have been identified in the PAX9 gene. Interactions between MSX1 and PAX9 at both gene and protein level play a key role in morphogenesis of the developing tooth (Chhabra et al., 2014). Mutations in either of these genes can result in disruption of

the normal tooth development. It was shown that dysfunction of MSX1 can influence all teeth whereas PAX9 mostly influences molar teeth (Chhabra et al., 2014).

2.10 Pleiotropic effects of genes associated with tooth development

It would be possible to suggest that most dental traits, regardless of whether or not they have a function mastication, come with a “package” of multiple unrelated traits that are influenced by the same DNA variant due to the effects of pleiotropy or linkage disequilibrium⁵ of many variants that are expressed during odontogenesis (Kimura et al., 2009, 2015). A good example of the pleiotropic effect is the V370A variant found in the EDAR gene. This variant was determined to have originated 30,000 years ago in China and has been subject to strong positive selection (i.e., selection of advantageous traits/alleles) in the eastern Asian and Native American populations (Kamberov et al., 2013). Previous investigations associated this variant with a wide variety of features including Asian specific hair thickness, incisor shoveling and double-shoveling (i.e., pronounced mesial and distal marginal ridges on the labial surface of the incisors), presence of hypoconulid (i.e., 5th cusp that occasionally occur on the lower molars), increased mammary gland ductal branching, and sweat gland density (Hlusko et al., 2018; Kamberov et al., 2013; Kimura et al., 2009; Park et al., 2012; Scott and Irish, 2017; Vohr & Green, 2013). We still do not know the true selective pressure for the EDAR V370A variant among east Asian and Native American populations. It is difficult to determine which one(s) of the mentioned traits is/are the actual target of the demonstrated positive selection, but it has been suggested that those not the actual target of natural selection might be “hitchhiking phenotypes” (Park et al., 2012:508) that make up a package of many features

⁵ Linkage disequilibrium is defined as the nonrandom inheritance of alleles in a population (Bush and Moore, 2012).

that are likely to be inherited together. Although the true selective pressure for this variant is still unknown, Park and colleagues (2012) speculate that the effect of this variant on hair morphology might be under strong selection, and the other features that come with the variant, such as the dental features, are only by-products. Another possibility they mention is that an unknown effect of this variant might be under strong genetic control, and its effects on hair morphology and dental phenotypes are hitchhiking phenotypes. Before coming to any conclusion, more studies regarding the true selective pressure of the EDAR V370A variant are needed.

2.11 Linkage Disequilibrium

Due to many factors such as non-random mating, selection, lack of meiotic recombination, population structure in a population, alleles might be non-randomly correlated with each other. When two SNPs are non-randomly correlated or linked with each other, this indicates that they are non-randomly inherited together (Bush and Moore, 2012). In genotype-phenotype association studies, it is important to consider the possibility of linkage disequilibrium occurring between alleles from different loci. It's important because LD between alleles determines how accurate the results of an association study are and influences how much we can trust the detected genetic markers' true role in producing a trait or disease (Qanbari, 2020).

Genetic markers that are in LD with each other will stay that way until they are broken apart by recombination events during meiosis. Over time, as the number of generations increase in a population, recombination events will break more genetic markers apart until the population moves away from linkage disequilibrium (i.e., nonrandom

association of alleles) and get closer to the linkage equilibrium state (i.e., no associations between alleles) (Bush and Moore, 2012).

If two SNPs are in LD with each other, that might have two possible outcomes in a genotype-phenotype association study (Bush and Moore, 2012). The first outcome is that the SNP detected in a study might be the real functional variant that cause the disease/trait. This type of association is called direct association (Bush and Moore, 2012). On the other hand, a SNP might not necessarily cause a trait but still might be found to be significantly associated to a phenotype only because it is associated with the real causal SNP somewhere else on the same chromosome with the detected SNP (Bush and Moore, 2012). Therefore, it is important to distinguish the trait-causing variants from the variants that are in LD in genotype-phenotype association studies in order to accurately identify the true causal variants.

2.12 Candidate genes

This section overviews the literature related to the candidate genes that might be involved in the formation of shovel-shaped incisors and Carabelli's cusp. Although these genes are involved in development of many organs, this review is focused on their role in tooth development.

2.12.1 NKX2-3

The NK class of homeobox genes are a family of transcription factors that contain a common homeodomain for DNA binding (Han et al., 2018). Members of the NK-2 class are essential during embryonic development of the spleen, pharynx, salivary glands, heart as well as teeth (Harvey, 1996). The *NK-1* and *NK-2* genes are two distinct classes that are shared among a wide variety of organisms including fruit flies, chickens, mice, nematodes,

and humans (Harvey, 1996). Based on the comparisons of homeodomain proteins of *NK1* and *NK2* genes of many organisms, it has been suggested that these two distinct classes are ancient in origin (Harvey, 1996).

Within the NK-2 class, two distinct clades have been defined: the “neural” clade, which is expressed in the central nervous system, and the “cardiac” clade expressed in the cells of the heart and visceral mesoderm. Genes within these two clades are also expressed in the pharyngeal endoderm in an overlapping pattern. Several members of the NK-2 class, including *NKX2-1*, *NKX2-2*, *NKX2-3*, *NKX2-5* and *NKX2-6*, have been shown to play a key role in the pharyngeal development in mice (Biben et al., 2002).

The *NKX2-3* gene is a protein coding member of the cardiac clade that is highly expressed during pharyngeal development (Biben et al., 2002; Han et al., 2018). In mice, the expression of *Nkx2-3* is first evident in the ectoderm of the first branchial arches, and then the expression becomes more restricted to the odontogenic epithelium (Biben et al., 2002). Biben and colleagues (2002) examined the genetic role of *Nkx2-3* during pharyngeal development in mice. Their study found that the expression of *Nkx2-3* was the highest in the molar region of the lower jaw. In *Nkx2-3* mutant mice, the sublingual gland morphogenesis was abnormal and delayed. In addition, the cusps of the lower molars were absent, and the third molar was occasionally missing. The authors argue that normal development of the body and the root of the molars, but defected development of cusps suggest that *Nkx2-3* expression is necessary during the later stages of tooth morphogenesis (Biben et al., 2002). They further argue that *Nkx2-3* might be essential for enamel knot formation in molars (Biben et al., 2002).

In a more recent study, Han and colleagues (2018) used microarray analysis to

identify expression of tooth specific genes in mice at embryonic day 14 (E14). They found that the expression levels of *Nkx2-3* were higher particularly in teeth compared to other organs examined in the study including the heart, skin, kidney, lung, hair, brain and stomach. In addition, using qRT-PCR, they confirmed that the expression levels of *Nkx2-3* at E14 are elevated in dental epithelium compared with mesenchyme (Han et al., 2018). Unlike the other members of the NK2 family, *Nkx2-3* showed a tooth-specific expression indicating its importance in tooth development. Furthermore, its expression was localized in the cells of the primary and secondary enamel knots through the bud (E13), cap (E14) and bell stages (E16). Han and colleagues (2018)'s study also showed that down-regulation of *Nkx2-3* expression interrupts with cusp formation in developing teeth in mice.

2.12.2 FGF3 and FGF4

FGFs (Fibroblast growth factors) are a family of polypeptides that play a pivotal role during development of many organs including teeth (Charles et al., 2009; Itoh & Ornitz, 2008). Twenty-two *FGF* genes are known to be expressed in both mice and humans (Itoh & Ornitz, 2008). Most of the members of the FGF family are extracellular proteins that bind to the tyrosine kinase fibroblast growth factor receptors (FGFRs) on the cell surface. However, a few FGFs are intracellular proteins (iFGFs) that work independently from the FGFRs (Itoh & Ornitz, 2008).

FGFs are expressed in both dental epithelium and mesenchyme in a developing tooth (Du et al., 2018). The enamel knots express many members of FGFs including *Fgf3*, 4, and 9. FGFs promote the proliferation of the cells around the enamel knots and control the folding of the epithelium. Although the enamel knots induce cell proliferation in the cells around them, the cells of the enamel knots lack *Fgf* receptors on their cell surface.

Lacking Fgf receptors makes the enamel knots insensitive to the Fgf signals and therefore they remain non-proliferative (Jussila & Thesleff, 2012). Due to their pivotal role during epithelial cell proliferation, precise control of the FGF expression is crucial for determination of the final phenotype of teeth (Jernvall & Thesleff, 2000).

The expression of Fgf4 is exclusively restricted in the cells of the enamel knots while the Fgf3 expression is also observed in the mesenchyme of developing mice tooth (Jernvall & Thesleff, 2000). Since the secondary enamel knots in a developing tooth mark the location of the future cusp tips, a highly localized expression pattern of FGF4 in the enamel knots was interpreted as FGF4 being a major promoter of cusp formation (Jernvall & Thesleff, 2000). In addition, it has been shown that the expression levels of FGF4 is significantly reduced in the epithelial cells of Lef1-deficient mice (Kratochwil et al., 2002). Kratochwil and colleagues (2002) observed that lack of FGF4 in the epithelial cells results in an arrest of mesenchymal condensation at E13 (Du et al., 2018). Based on this result, Kratochwil and colleagues (2002) wanted to further examine the effects of FGF4 in the mesenchyme. They treated the mesenchymal cells with exogenous FGF4 proteins. They observed that the developmental arrest in the mesenchymal cells is rescued due to the increased levels of FGF4 protein by inducing the expression of *Fgf3* gene. This indicates that Fgf3 may be a direct target of FGF4 (Kratochwil et al., 2002).

Mutations in the Fgf3 gene have been reported to cause major morphological anomalies in molar teeth of both mice and humans (Charles et al., 2009; Tekin et al., 2007). In mice, a decrease in Fgf3 dosage results in altered molar size and shape (Charles et al., 2009). Homozygous *Fgf3*^{-/-} (i.e., loss of function of the Fgf3 gene) mice exhibit smaller molars and loss of cusps compared to the heterozygous *Fgf3*^{+/-} and homozygous wild type

Fgf3^{+/+} molars. Similarly, Fgf3 null (i.e, Fgf3^{-/-} or Fgf3-deficient) human patients display smaller upper first and second molars, and a missing hypocone cusp. Because of the missing hypocone, molar teeth display a triangular shape with only three cusps which resembles the molar morphology of primitive primates, such as the Asian *Bahinia* genus from 40 million years ago. Based on this observation, Charles and colleagues (2009) suggest that a decrease in Fgf3 gene dosage results in reappearance of ancestral anthropoid molar morphology with a missing hypocone. They argue that the evolutionary appearance of hypocone in nonhuman primates and humans might be linked to the increase of Fgf3 gene dosage during embryonic tooth development.

2.12.3 WNT10A

WNTs belong to a large signaling molecule family consisting of nineteen different proteins identified in mammals. The abbreviation Wnt is a combination of the “Wingless” gene in fruit flies and the “Int-1” (integration site 1) gene in mice (Tamura & Nemoto, 2016). Initially, these two genes were discovered separately (Nusse & Varmus, 1982; Nüsslein-Volhard & Wieschaus, 1980). Later studies have shown that the two genes are homologues of each other, and therefore, the two names were combined (Rijsewijk et al., 1987). WNTs are crucial in early development of many organs, especially teeth. Many publications have pointed out the oncogenic potential of WNTs (Rijsewijk et al., 1987) and the significance of WNTs in developing and postnatal skin, hair, and teeth morphogenesis (Andl et al., 2002; Cai et al., 2011; Tamura & Nemoto, 2016; Yamashiro et al., 2007).

WNT10A is a member of the WNT signaling family and was first discovered in the 1990s in mice (Wang & Shackleford, 1996). WNT10A has been associated with tooth development (Tamura & Nemoto, 2016). Mutations in this gene have been shown to cause

a rare type of ectodermal dysplasia (ED) known as odonto-onycho-dermal dysplasia (OODD) (Adaimy et al., 2007). OODD is an autosomal recessive syndrome that causes dry hair, congenitally missing teeth, smooth tongue (i.e., reduced lingual structures such as papillae on the tongue) and hyperhidrosis (Adaimy et al., 2007). It has been reported that the severity of tooth agenesis is elevated if an individual is homozygous for the disease causing Wnt10a loci or if the disease causing Wnt10a alleles are combined with other ED causing alleles in several other genes including Eda (ectodysplasin A), EDAR (ectodysplasin A receptor), and EDARADD (EDAR- associated death domain) (Yang et al., 2015).

WNT10A is a ligand (i.e., signaling molecule that recognizes and binds to a receptor) that is expressed in both primary and secondary enamel knots (Tamura & Nemoto, 2016) and Wnt10a expression was not found in dental mesenchyme (Yamashiro et al., 2007). Localized expression of Wnt10a has been observed in secretory odontoblasts indicating that Wnt10a regulates the differentiation of odontoblasts and the expression of dentin sialophosphoprotein (Dspp) (Yamashiro et al., 2007). This indicates that Wnt10a is an important regulator of dentinogenesis and tooth development in general.

Yang et al. (2015) used WNT10A knockout mice and human patients to investigate the dental phenotypes that are associated with this gene. Since previous studies associated this gene with tooth agenesis, (Yang et al., 2015) expected that the WNT10A null mice in their study might also be born with missing teeth. However, contrary to expectations, one of the WNT10A null mice developed forth molars distal to the third molar in the lower jaw. In addition, the crown and root morphology, and cusp patterning in the WNT10A null mice were altered.

Kimura et al. (2015) investigated whether the SNPs in the THADA, FRAS1, WNT10A, NAF1 and FGFR2 genes, which were previously associated with hair morphology, are also associated with tooth morphology. The authors used genetic and dental metric and nonmetric data from Japanese and Korean individuals. Their results showed that one SNP in the WNT10A gene, rs7349332, was significantly associated with crown size. They further investigated the WNT10A gene and found another SNP, rs10177996, that was significantly associated with crown traits including maxillary length, UI1 shoveling, LP2 distolingual cusp, UM1 5th cusp, and hypoconulid (i.e., cusp 5 or the distal cusp of the lower molars) in LM2 (Kimura et al., 2015). The worldwide distribution of the rs10177996 showed that the derived allele which was associated with crown size is widespread in Europe whereas it is not common in African populations (Kimura et al., 2015). This contradicts previous studies that found that the crown sizes of European and Asian teeth are smaller than those of the Africans (Hanihara and Ishida, 2005). The authors interpret this as an indication that other genetic factors are likely involved in producing the different morphologies of Eurasian and African teeth (Kimura et al., 2015).

2.12.4 BMP4

Bone Morphogenic Proteins (BMPs) are a family of growth factors that belong to the Transforming Growth Factor- β (TGF- β) signaling protein superfamily (Wang et al., 2014). BMPs were discovered in the 1960s after recognizing their potential to produce bone and cartilage tissue from mesenchymal cells. Later, it was discovered that BMPs are pivotal in induction of not only bones and cartilage formation but formation of a wide variety of tissue types throughout the body in all animals (Murashima-Suginami et al., 2008; Wang et al., 2014). BMPs are essential during embryonic development for cell

growth, cell death and differentiation. In addition, they are important regulators of adult tissue homeostasis including tissue repair and vascular remodeling (Wang et al., 2014).

BMP signaling is one of the four major signaling families that regulate tooth development (Malik et al., 2018). Murashima-Suginami et al. (2008) reported that the rat gene USAG-1 (aka ectodin or SOSTDC1 as the human ortholog) acts as an inhibitor of BMP and Wnt signaling during odontogenesis. The authors observed that BMP and Wnt signaling is enhanced in USAG-1-deficient mice. In the absence of this important inhibitor, BMPs and Wnts are over- expressed which leads to supernumerary tooth formation in mice (Murashima-Suginami et al., 2008).

It has been previously reported that tooth development does not proceed beyond the bud stage in *Msx1*-deficient mice (Bei & Maas, 1998). Expression of many transcription and growth factors and ligands including BMP4 and FGF3 requires *Msx1* expression. Bei and colleagues (2000) cultured wild-type and *Msx1*-deficient tooth germs with exogenous *Bmp4* and *Fgf3* in vitro to investigate whether the expression of BMP4 and FGFs is dependent on *Msx1* expression. They observed that exposure to exogenous BMP4, but not FGFs, rescued tooth development in *Msx1*-deficient mice (Bei et al., 2000). Treating *Msx1*-deficient tooth germs with BMP4 allowed the tooth germs to establish their subsequent developmental stages (Bei et al., 2000). This observation indicates that *Bmp4* expression is required for primary enamel knot formation so that the tooth germ can proceed beyond the bud stage (Jia et al., 2013).

2.12.5 SOSTDC1

The *SOSTDC1* (sclerostin domain containing 1) gene, also known as *Usag1*, ectodin and wise, has been reported to be involved in tooth development in mice (Kassai

et al., 2005; Laurikkala et al., 2003). This gene acts as an inhibitor of BMPs and Wnts by binding to their receptors and neutralizing them (Arikan et al., 2018). *Sostdc1* expression has been reported in the cells of epithelium and mesenchyme of many developing ectodermal organs including teeth, hair, and mammary glands (Kassai et al., 2005; Laurikkala et al., 2003).

Kassai and colleagues (2005) have shown that loss of ectodin (*SOSTDC1*), which is a well-known BMP inhibitor, results in excess BMP and p21 expression, which causes unbalanced activators and inhibitors in and around the enamel knots. This in turn interferes with the reaction-diffusion mechanism of tooth development, which determines the final cusp pattern. Their results showed that *SOSTDC1* deficient mice developed larger enamel knots, molar teeth with smaller intercuspal regions, fused and supernumerary cusps and extra teeth (Kassai et al., 2005).

In addition to Kassai and colleagues' (2005) results, multiple other studies (Munne et al., 2009; Murashima-Suginami et al., 2021) have also confirmed that *Sostdc1*/ectodin knockout mice develop extra incisors and molars adjacent to the existing incisors and molars, respectively. Mice develop only one set of teeth during their lifetime (Catón & Tucker, 2009; Peterkova et al., 2014). They lost their ability to replace teeth during their evolutionary history. Mice have specialized dentition with only molars and continuously growing incisors. A diastema exists between incisors and molars in place of the missing canines and premolars. Studies have shown that rudimentary dental placodes form in the diastema region during the very early stages of normal tooth development (Munne et al., 2009). However, these tooth rudiments do not grow into whole functional teeth because they are evolutionarily programmed to undergo apoptosis during the later stages of

embryonic development. Previous work has shown that downregulation of the inhibitor *Sostdc1* can result in “rescue” of these tooth rudiments before they undergo programmed cell death (Kassai et al., 2005; Munne et al., 2009).

Munne and colleagues (2009) observed that duration of Wnt activity is longer and Wnt expressing cells are more wide-spread around the primary incisor enamel knot in *SOSTDC1*-deficient mice compared to the wild-type mice (wild-type refers to the *Sostdc1*^{+/+} heterozygous mice). As a result, *SOSTDC1* deficiency allows mice to develop extra teeth where the vestigial incisor and premolar primordia are located (Munne et al., 2009; Murashima-Suginami et al., 2021). In the normal wild-type mice, these high-WNT and -BMP and low-*SOSTDC1*-activity areas undergo intense apoptosis, and therefore, extra teeth never have an opportunity to grow in these regions.

2.13 Research questions and hypotheses to be addressed

This dissertation project will address the following questions: 1) Are there any short variants (i.e., SNPs and INDELs) within the candidate genes that show strong correlations with the dental traits of interest in this study? 2) Does the reference allele or alternate allele(s) within the detected variants show stronger correlation with the degree of trait expression? 3) Are the detected variants in linkage disequilibrium (LD) with each other? In other words, are the detected variants nonrandomly coinherited?

This study will test the following hypotheses:

1) Since the literature has shown that several types of variants, such as the Single Nucleotide Polymorphisms (SNPs) and Insertion-Deletions (INDELs), in many genes that are involved in tooth development can cause variation in tooth morphology, it would be expected that the candidate genes in this study might have roles in the formation of

shoveling and Carabelli's cusp in humans. Therefore, I hypothesize that some short variants (i.e., SNPs or INDELs) within the candidate genes are strongly associated with the presence of shovel-shaped incisors and Carabelli's cusp,

2) Alleles detected in a variant in the six candidate genes are expected to have differential effects on the final morphology. One allele might have a more substantial effect on the degree of expression of shoveling or Carabelli's cusp than the other allele. I hypothesize that the reference allele (i.e., the allele observed in the reference genome GRCh38) is more strongly associated with trait expression than the alternate allele (i.e., the allele observed in the study participants' genotypes) in a detected locus, or vice versa,

3) If multiple variants are detected in a chromosome, I hypothesize that several of them are expected to be in strong linkage disequilibrium (LD) with each other (i.e., they are nonrandomly inherited together).

Chapter 3 Materials and Methods

3.1 IRB approval and Participant Recruitment

This dissertation project was approved by the Binghamton University Institutional Review Board (IRB) (STUDY00002536) with special Covid-19 restrictions in effect (See Figures B.1 and B.2 in Appendix B). The original study methods had to be modified to meet the requirements of the pandemic restrictions.

Due to the COVID-19 restrictions, only the individuals who were affiliated members of the Binghamton University Campus Community (faculty, staff, students) were allowed to participate in the study between November 2020 and July 2021. After the restrictions were lifted in July 2021, participants outside of the campus were also recruited.

Equal numbers of people from any sex were targeted. Regardless of gender, adults 18 years old and older who possess shovel-shaped incisors and/or Carabelli's cups in molars were included in the positive control group in the study. Adult individuals who showed interest in participating but did not possess the dental traits of interest were included in the negative control group. The goal was a sample size of 50 individuals per dental trait with a total of 100 individuals for both shovel-shaped incisors and Carabelli's trait. However, due to the Covid-19 pandemic only 36 individuals volunteered to participate in the study.

Sample collection which involved both saliva and dental data collection from the study participants took place in the state of New York, United States. between January 2021 and August 2021. Table 3.1 shows the self-reported demographics of the study participants. Participants comprised of people who-self reported as European (N=20), Asian (N=9), European Hispanic (N=4), Hispanic (N=2), and African American (N=1) individuals. Ancestries were self-reported by the participants (Table 3.1). They received an open-ended question in the subject information form that asked about their maternal and paternal ethnicities (See Figure B.3 in Appendix B). Therefore, the categories in Table 3.1 are based on the responses of the participants. If the participants had different maternal and paternal ethnicity, they were assigned to both of their parents' ancestry such as "European Hispanic". Participants who put specific country names as their ancestry were put in broader categories such as "European" and "Asian".

The reason why the participants' reported responses were put into broader categories is that there were so few individuals in the study that reporting the exact responses provided by study participants might reveal personally identifiable information and violate the terms laid out in the IRB protocol. The study population included a number of individuals from groups that have been historically marginalized and exploited and therefore this was taken as a measure to protect their privacy, confidentiality, and bio-data.

Ancestry	Number of participants (N)	Percentage (%)
European	20	55.56%
Asian	9	25%
European Hispanic	4	11.11%
Hispanic	2	5.56%
African American	1	2.78%

Total:	36	100%
--------	----	------

Table 3.1 Self-reported ancestries of study participants.

3.2 Dental data collection and trait scoring

The initial aim in this project was to collect dental impressions from the participants. However, due to the COVID-19 restrictions, I had to switch to a remote dental photo collection rather than collecting actual dental impressions from the participants. Participants were asked to submit photos of their teeth to a private Google Drive to determine their eligibility to participate in the study. The photos were also used to score the dental traits of interest in this study.

After the COVID-19 restrictions were lifted in July 2021, the Binghamton University IRB approved collection of dental impressions from the participants. In addition to collecting dental photos, dental impressions were taken from 17 participants to make dental casts. Negative impressions of both the subject's upper and lower dentitions were taken by following standard dental procedures (Mandikos, 1998). An alginate alternative dental impression material, vinyl polysiloxane (VPS) was used, which comes in ready-to-use 50 ml cartridges. This alginate alternative material compared to the traditional alginate impression material eliminated the need for mixing the alginate powder with water. The traditional alginate mixing procedure is time consuming and requires a lot of practice. Therefore, to shorten the participant's time commitment in the study, VPS impression material was chosen in this study.

Scoring of the dental traits was completed utilizing the Arizona State University Dental Anthropology System (ASUDAS) plaques (Turner et al., 1991). Only the left incisors and molars were scored. The photographs of the ASUDAS plaques of shoveling in upper central (UI1) and lateral incisors (UI2) and Carabelli's cusp on upper molars (UM)

are shown in Figures 3.1, 3.2, and 3.3, respectively. ASUDAS plaques have been commonly used as a standard for scoring dental traits in the field of dental anthropology. ASUDAS is a standardized system that involves ordinal scales for dental nonmetric trait expression. The scale for each trait starts with the lowest level of expression (i.e., absence or Grade 0) and ends with the highest level of expression. The plaques provide straightforward observations on teeth and decreased the intra- and inter-observer errors.



Figure 3.1 ASUDAS plaque for shoveling in upper first (central) permanent incisors.



Figure 3.2 ASUDAS plaque for shoveling in upper second (lateral) permanent incisors.



Figure 3.3 ASUDAS plaque for Carabelli's cusp in upper permanent molars

3.3 Saliva sample collection and DNA extraction

A mouthwash saliva collection protocol was used to collect saliva. The participants rinsed their mouth with 10 ml mouthwash for 30 seconds and spit into a 50 ml collection tube. DNA was extracted from the collected saliva using the QIAamp DNA Mini Kit (Qiagen, Germany) following the manufacturer's protocol.

3.4 Candidate gene approach

In this project, the candidate gene approach was employed. This approach required selection of the pre-specified genes of interest that were previously reported to be expressed during odontogenesis by the earlier studies (Tabor et al., 2002). The six strongest candidate genes were picked according to some stringent criteria (Table 3.2). The first and most important criterion for picking a candidate gene is the gene's involvement during dental development. Another criterion is the length of the gene. Due to the difficulties of efficiently amplifying very long DNA templates, genes that are no longer than 10,000 bp ($\pm 2,000$ bp) were selected as the candidate genes in this project except for the *WNT10A*

gene which is almost 20,000 bp. Due to the length of this gene, it was divided into three smaller sections and separate primer sets were designed for each section. However, due to the unreliable amplification issues encountered for the first two sections during PCR, only the last section which gave more reliable and consistent amplification results were included in this study.

Gene	Gene length (bp)	Genomic position (GRCh38.p13)	Type of molecule	References
NKX2-3	3583	Chr10:99532942..99536524	Transcription factor	Han et al., 2018
FGF3	9449	Chr11:complement(69809968..69819416)	Ligand	Jernvall & Thesleff, 2000
FGF4	4320	Chr11:complement(69771022..69775341)	Ligand	Du et al., 2018; Jernvall & Thesleff, 2000
BMP4	7156	Chr14:complement(53949736..53956891)	Ligand	Jernvall & Jung, 2000; Ramanathan et al., 2018
SOSTDC1	4258	Chr7:complement(16461481..16465738)	Ligand	Kassai et al., 2005
WNT10A	19816	Chr2:218874116..218893931	Ligand	Kimura et al., 2015

Table 3.2 List of the candidate genes that were sequenced in this dissertation project.

3.5 Primer Design for Long-range PCR

Primers for long-range PCRs were designed with high specificity using the National Center for Biotechnology Information's (NCBI) Primer-BLAST tool which is freely available (Ye et al., 2012). Primers were designed to amplify the entire gene sequence including the intronic (i.e., segments of DNA that does not code for amino acids) and exonic (i.e., coding) regions of the genes of interest. To increase specificity, primers

were designed according to the criteria that were recommended by the TaKaRa LA (Takara Bio Inc.) Taq manual. TaKaRa LA (Long and Accurate) Taq polymerase enzyme (Takara Bio Inc.) and reagents were used to perform long-range PCRs. This enzyme allows for amplification of very long DNA templates with high fidelity. Long-range polymerase chain reaction (PCR) was performed to amplify the entire gene region of *NKX2-3*, *FGF3*, *FGF4*, *BMP4*, and *SOSTDC1*. Due to its length, only the last third of *WNT10A* gene was included in the study. Table 3.3 shows the primer sequences, genomic locations, and length of amplicons.

Gene name	Genomic location (GRCh38/hg38)	Sequence (5'→3')	Length of amplicon (bp)
Fgf3	Chr11:69809200-69820343	Forward: CCAAGTGCCAGGAGAGGTTAGTACACTGC Reverse: GGGACAGAGGACCAGGAAGCAAGAGAAA	11144
Fgf4	Chr11:69770778-69776854	Forward: TACAGTGC GGGAATGGCGTGAATTAGC Reverse: AGACAACACAGCAAGTGAGGGATGGGT	6077
Bmp4	Chr14:53949462-53959648	Forward: CATCCCAGTGT TTTCTCCAAGGCATGTGT Reverse: GGGCAGGACCAGGAAGTCTGCATTTTCATC	10187
Nkx2-3	Chr10:99532030-99537060	Forward: TTTGCCTCATTCAACCCTAGCAACAACCA Reverse: CTCCGCAAGTGACAAGGAGCCGCATA	5031
Sostdc1	Chr7:16458386-16465969	Forward: TCTCACACCGAGCATCCTAAGTCACCTC Reverse: GCGTCGGCTCACAGACAAGTGATGAAGT	7584
Wnt10a	Chr2:218886144-218894785	Forward: TGTACCCAGAGAGGTGAGCTGGTGCAA Reverse: CACAAGAGGCCCCAGGAAGAATGTGCCC	8642

Table 3.3 Primer sequences, genomic locations, and length of amplicons.

3.6 Polymerase Chain Reaction (PCR) Optimization and Agarose Gel Electrophoresis

PCRs were performed using a GeneAmp[®] PCR System 9700 (Applied Biosystems, Foster city, CA). Table 3.4 shows the manufacturer's recommended general PCR reaction mix for TaKaRa LA Taq. Each gene region was amplified in 50 µl volumes containing the PCR mix components listed in Table 3.4. A 2-step PCR technique with combined annealing and extension stages was used to amplify all the target sequences except for *WNT10A*. Due to the specificity issues encountered during amplification of *WNT10A*, amplification of this target was performed using the touch-down (TD) PCR technique. TD is a special PCR technique that starts with a temperature that is several degrees higher than the calculated T_m (melting temperature) of the primers, and the temperature is gradually decreased until the optimum T_m is reached (Korbie and Mattick, 2008). This technique allows for higher specificity and prevention of heteroduplex formation (i.e., amplification of homolog sequences rather than the target sequence) during PCR. Table 3.5 shows the PCR cycling conditions used for each target region. Success of the PCR amplifications was confirmed via 0.4% agarose gel electrophoresis.

Reagent	Volume	Final Concentration
TaKaRa LA Taq [™] DNA Polymerase (5 units/µL)	0.5 µL	2.5 units/50 µL
10X LA PCR Buffer II (Mg ²⁺ plus)	5.0 µL	1X
dNTP Mix (2.5 mM each)	8.0 µL	400 µM
Template DNA	1-5 µL	2.5-500 ng/50 µL
Primers	1-5 µL	0.2 µM each
Sterile ddH ₂ O	Up to 50 µL	

Table 3.4 PCR reaction mix for a single 50 µL reaction (adapted from TaKaRa Bio Inc.'s website).

Target name	Cycle temperatures and times	Cycle Number	Cycle name
FGF3	94 °C 1 min	1X	Initial denaturation
	98 °C 10 sec + 68 °C 10 min	33X	Amplification
	72 °C 10 min	1X	Final elongation
WNT10A	94 °C 1 min	1X	Initial denaturation
	98 °C 20 sec + 73 °C 5 min 40 sec	5X	Amplification
	98 °C 20 sec + 71 °C 5 min 40 sec	5X	Amplification
	98 °C 20 sec + 69 °C 5 min 40 sec	25X	Amplification
	72 °C 10 min	1X	Final elongation
NKX2-3 and FGF4	94 °C 1 min	1X	Initial denaturation
	98 °C 10 sec + 69 °C 4 min 33 sec	30X	Amplification
	72 °C 10 min	1X	Final elongation
BMP4 and FGF3	94 °C 1 min	1X	Initial denaturation
	98 °C 10 sec + 68 °C 9 min	30X	Amplification
	72 °C 10 min	1X	Final elongation

Table 3.5 PCR cycling conditions.

3.7 PCR Product Purification, DNA Quantification and Amplicon Pooling

Amplicons were purified using Agencourt AMPure XP PCR Purification systems following the manufacturers protocol (Beckman Coulter, Pasadena, CA). A 1.8:1 ratio of AMPure beads to sample was used. Briefly, 10 µl of sample was added to 18 µl of beads, and the bead-sample mixture was incubated at room temperature. The mixture was placed on a magnetic stand and washed with 200 µl 70% ethanol twice. Lastly, the sample was eluted with 40 µl nuclease-free water, and the elution which contains the purified PCR product was collected.

The purified PCR product was quantified on a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) using Qubit® dsDNA HS Assay Kit. The manufacturer's recommended protocol for quantifying double stranded DNA was followed. Briefly, 1:200 Qubit® dsDNA HS reagent to Qubit® dsDNA HS buffer was mixed to prepare a working solution. 198 µl of working solution was mixed with 2 µl of

sample to perform the quantification. Each amplicon (i.e., purified PCR product) from 36 participants were diluted to 1 ng/μl before pooling. 5 μl of each amplicon belonging to a single participant was pooled and used as the starting material for library preparation described below.

3.8 Library Preparation

Libraries were prepared using the Nextera XT DNA library preparation kit (Illumina, Inc.). Each participant was treated as a single library, and therefore, indexed separately using IDT for Illumina DNA/RNA UD indexes (Illumina). 1 ng of the pooled amplicons was used for library preparation. The main steps of the Nextera XT DNA library preparation workflow includes tagmentation, amplification, clean up, library quality check, normalization, and library pooling. It should be noted here that the library quality check step of the library prep workflow had to be skipped due to the technical issues encountered with the Bioanalyzer equipment. It is recommended in Illumina's Nextera XT DNA library prep reference guide (Document # 15031942 v05) that 600 bp can be used as the average library size to calculate the molarity of the libraries if the libraries were not qualified on a Bioanalyzer. Therefore, the average library size of 600 bp was assumed for each library to calculate the molarity of the final pooled libraries. Finally, the libraries were pooled with a final concentration of 2nM. The pool also included PhiX at 12.5 pM concentration as a sequencing control.

3.9 Next-generation DNA Sequencing

Pooled libraries with a concentration of 2 nM were further denatured and diluted to 10 pM of loading concentration for sequencing. Sequencing was performed using MiSeq Reagent Kit v2 with a 2x149-cycle paired-end run configuration (Illumina, Inc.) by

utilizing an Illumina MiSeq next generation sequencer in Dr. D. Andrew Merriwether's Molecular Anthropology Lab at Binghamton University. The reason for using a 2x149 read length configuration rather than a more common 2x151 read length is because IDT for Illumina UD indexes were used during library prep. IDT for Illumina UD indexes are 10 bp long compared to Nextera XT indexes which are 8 bp long which required adjustments to the read length parameters.

3.10 Data analysis

3.10.1 Bioinformatics pipeline for NGS data

This section describes the analytical pipeline that utilizes various programs/tools to process the raw NGS output data used in this dissertation (Van der Auwera & O'Connor, 2020; DePristo et al., 2011; McKenna et al., 2010; Poplin et al., 2018; Van der Auwera et al., 2013). The process starts with a FASTQ file which is a text-based file that contains raw sequencing reads along with the quality scores associated with each base call (Cock et al., 2010). The final output of the whole pipeline is a Variant Call Format (VCF) file. A VCF file is also a text file, and it contains all the known variants that were detected in the sample. The VCF file only contains the variable regions of the genome rather than the non-variable regions which make up the majority of the human genome. This file format is commonly used in bioinformatic analyses, especially for the ones that deal with large-scale sequencing data such as whole-genome sequencing data.

3.10.1.1 Data pre-processing (Trimming and Merging)

The adapter sequences that were added during library preparation need to be removed from the raw reads for further processing. The AdapterRemoval tool (Lindgreen, 2012) was used for this purpose. This tool can be used for both single-end and pair-end reads. The following command was used to trim the adapter sequences from the ends and

merge overlapping paired- end reads to obtain longer consensus sequences:

```
AdapterRemoval --file1 [read1.fastq] --file2 [read2.fastq] --basename  
[output] --trimns -- trimqualities --collapse --adapter1 [adapter sequence 1] --  
adapter2 [adapter sequence 2]
```

3.10.1.2 Mapping and Alignment

The trimmed and merged reads were mapped and aligned to Grch38/hg38 human reference genome. The reference sequence, dbSNP and INDEL data was obtained freely from the Broad Institute's GATK resource bundle (<https://gatk.broadinstitute.org/hc/en-us/articles/360035890811-Resource-bundle>). The alignment software Burrows-Wheeler Aligner MEM (BWA) tool was utilized for mapping and alignment (Li & Durbin, 2009). This step generates a Sequence Alignment Map (SAM) file as the output. The output file from the previous step (i.e., trimming and merging) is used as the input for the alignment:

```
bwa mem reference.fasta input > output.sam
```

3.10.1.3 Removing soft-clipped reads

The *awk* command below was used to remove soft-clipped reads which are the mismatches that occur during the alignment. This step converts SAM file into a BAM file format (Binary Alignment Map). A BAM file is the compressed binary version of SAM file. Unlike a SAM file, a BAM file is only machine readable.

```
awk 'BEGIN {{OFS="\t"}} {{split($6,C,/ [0-9]*/); split($6,L,/ [SMDIN]/); if  
(C[2]=="S"){{ $10=substr($10,L[1]+1); $11=substr($11,L[1]+1)}}; if  
(C[length(C)]=="S"){{ L1=length($10)-L[length(L)-1]; $10=substr($10,1,L1);  
$11=substr($11,1,L1); }}; gsub(/ [0-9]*S/, "", $6); print }}' input.sam >  
output.bam
```

3.10.1.4 BAM cleaning and preparation (Samtools)

The output BAM file was further processed using the Samtools software (Li et al., 2009). The command lines below were used for;

Filtering:

```
samtools view -bh -F4 input.bam > output.filtered.bam samtools view -bh -q 30  
input.bam > output.q30.bam
```

Sort alignment:

```
samtools sort -o input.bam output.sorted.bam
```

Duplicate removal:

```
samtools rmdup -s input.bam output.removeduplicate.bam
```

3.10.1.5 Visualization in Integrative Genomics Viewer (IGV)

Integrative Genomics Viewer (IGV) was utilized for the visualization and verification of the BAM files (Robinson et al., 2011). IGV requires a .bam file and an associated .bai file to visualize the reads. Following command was used to index the .bam file to generate a .bai file:

```
samtools index -b input.bam
```

3.10.1.6 BAM cleaning and preparation (Picard and GATK)

Steps in this section are taken to further manipulate the BAM files to prepare them for the downstream analyses described below. Picard (<https://sourceforge.net/projects/picard/>) and Genome Analysis Toolkit (GATK) programs were utilized during these steps (Van der Auwera et al., 2013). Picard's MarkDuplicates function is used to identify and remove reads that have identical sequences and coordinates. AddOrReplaceReadGroups function assigns separate libraries to separate read groups. The commands are as follows:

Mark duplicates:

```
java -jar picard.jar MarkDuplicates -I input.bam -O output.bam -M  
output_metrics.txt
```

Add (or change) read groups:

```
java -jar picard.jar AddOrReplaceReadGroups -I input.bam -O output.bam --
```

```
RGLB LIB[output] --RGPL ILLUMINA --RGPU Unit1 --RGSM output
```

where RGLB= Read-Group library; RGPL= Read-Group platform such as ILLUMINA; RGPU= Read-Group Platform unit (eg. run barcode); RGSM= Read-Group sample name

Base Quality Score Recalibration (BQSR):

The sequencing machine (i.e., Illumina MiSeq) assigns each base call to a base quality score (i.e., Phred quality score). The assigned base quality score is an indication of how likely the machine called a correct base each time. Although this process is quite accurate most of the time, systematic errors can occur during when each base call is assigned to a quality score. Therefore, Base Quality Score Recalibration (BQSR) is a process to detect and remove those errors made by the sequencing machine. BQSR works based on a machine learning model that uses a set of known variants and associated quality scores to adjust and improve the accuracy of base qualities. This is a two-step procedure which involves GATK's 1) BaseRecalibrator tool to build a model and 2) ApplyBQSR tool to apply the model to the sequencing data and produce a BAM file. Known sites for SNPs and INDELs were obtained from the Broad Institute's GATK resource bundle. Used commands are as follows:

```
java -jar gatk-4.2.5.0/gatk-package-4.2.5.0-local.jar BaseRecalibrator -R  
reference.fasta -I input.bam --known-sites dbSNP.vcf --known-sites  
INDELs.vcf.gz -O recal_data_output.table
```

Apply the model generated by BQSR:

```
java -jar gatk-4.2.5.0/gatk-package-4.2.5.0-local.jar ApplyBQSR -R  
reference.fasta -I input.bam --bqsr-recal-file recal_data_output.table -O  
output.bam
```

3.10.1.7 Identifying genotype calls and creating GVCFs

This step involves GATK's HaplotypeCaller tool which detects the variable (or active) regions in the sequencing reads and calls them. It calls both the SNPs and INDELs

simultaneously. The resulting file is a GVCF format which is an intermediate output that contains only the active regions of the sequence reads.

```
java -jar gatk-4.2.5.0/gatk-package-4.2.5.0-local.jar HaplotypeCaller -R
reference.fasta -I input.bam -O output.g.vcf.gz -bamout output.bam -ERC
GVCF
```

3.10.1.8 Consolidating GVCFs and joint genotyping

Consolidate GVCFs across samples:

The command line below combines GVCFs from multiple samples:

```
java -jar gatk-4.2.5.0/gatk-package-4.2.5.0-local.jar GenomicsDBImport -R
reference.fasta -V input.g.vcf.gz --genomicsdb-workspace-path my-database-
chr# --intervals chr#
```

Joint genotyping on a cohort:

This step uses the output database file created by the previous step for joint genotyping and generates a VCF file.

```
java -jar gatk-4.2.5.0/gatk-package-4.2.5.0-local.jar GenotypeGVCFs -R
reference.fasta -V gendb://my-database-chr# -O chr#.vcf
```

3.10.1.9 Callset refinement

Separate SNPs from Indels:

```
java -jar gatk-4.2.5.0/gatk-package-4.2.5.0-local.jar SelectVariants -V
input.vcf -select-type SNP -O output-SNPs.vcf
```

```
java -jar gatk-4.2.5.0/gatk-package-4.2.5.0-local.jar SelectVariants -V
input.vcf -select-type INDEL -O output-INDELs.vcf
```

VariantRecalibrator (VQSR):

Similar to BQSR, VariantRecalibrator (VQSR) utilizes a series of high-quality known sites for an organism as “true sites” and feeds the relationship between the true sites and their annotations (QD, MQ etc.) into a machine learning model as training data. VQSR then builds a recalibration model to evaluate the probability of each variant being real and assigns a log-odds ratio (VQSLOD) score in the INFO field. This step was repeated for

SNPs and INDELs separately:

```
java -jar gatk-4.2.5.0/gatk-package-4.2.5.0-local.jar VariantRecalibrator -R
reference.fasta -V input.vcf --trust-all-polymorphic -tranche 100.0 -tranche
99.95 -tranche 99.9 -tranche 99.8 -tranche 99.6 -tranche 99.5 -tranche 99.4 -
tranche 99.3 -tranche 99.0 -tranche 98.0 -tranche 97.0 -tranche 90.0 -an QD -
an MQ -an ReadPosRankSum -an FS -an SOR -an DP -mode SNP --max-
gaussians 5 --
resource:hapmap,known=false,training=true,truth=true,prior=15.0
hapmap_3.3.hg38.vcf.gz --
resource:omni,known=false,training=true,truth=false,prior=12.0
1000G_omni2.5.hg38.vcf.gz --
resource:1000G,known=false,training=true,truth=false,prior=10.0
1000G_phase1.snps.high_confidence.hg38.vcf.gz --
resource:db SNP,known=true,training=false,truth=false,prior=2.0 db SNP138.vcf
-O output.recal --tranches-file VQSR-all-samples-chr#.tranches
```

ApplyVQSR:

Following command line applies the machine learning model generated by the previous step to the sequencing reads. This step was repeated for SNPs and INDELs separately.

```
java -jar gatk-4.2.5.0/gatk-package-4.2.5.0-local.jar ApplyVQSR -V input.vcf -
O output.vcf --truth-sensitivity-filter-level 99.7 --tranches-file VQSR-all-
samples-chr#.tranches --recal-file VQSR-all-samples-chr2-SNPs.recal -mode
SNP --create-output-variant-index true
```

3.10.2 Spearman's Rank Analysis

The variant calling pipeline scanned a total of 48,665 bp DNA sequence from the six candidate genes. This number is the sum of all the amplicon lengths of the six candidate genes that were amplified and sequenced in this project (see Table 3.3 for the amplicon lengths). The variant calling pipeline identified 277 variants in the 48,665 bp DNA sequence. The remaining 48,388 bp did not show any variable regions and therefore was not included in the further statistical analysis. 277 variants that were detected by the variant calling pipeline were analyzed using the Spearman's rank correlation test.

The effect and significance of a detected genetic variant on a dental phenotype was

estimated using Spearman's rank correlation coefficient (ρ) analysis. The Spearman's rank correlation test is used to measure the strength and direction of relationship between two ranked or ordinal variables. The Spearman "correlation coefficient is represented by the Greek letter rho, ρ " (Corder & Foreman, 2011:140). Correlation coefficient values range between 1 to -1. 1 describes a perfect positive relationship where one variable increases, the other variable increases as well. On the contrary, -1 indicates a perfect negative relationship where one variable increases the other variable decreases. A 0 value indicates no relationship between variables.

In this project, the genotypes are ranked according to the number of copies of the alternate and reference allele: 0: homozygotes for the reference allele (RR); 1: heterozygotes (RA) and 2: homozygotes for the alternate allele (AA) (Kimura et al., 2015). Dental trait expression, the other ranked variable, is already by design in the ranked format in the ASUDAS system (i.e., starting with a rank of "Grade 0" and ending with the highest expression of the trait). Figure 3.4 shows a visual representation of positive and negative genotype-trait expression grade correlations. If an individual's genotype at a locus is positively correlated with the dental trait grade, it means that as the trait expression grade increases, the number of alternate alleles increases. If the correlation coefficient is a negative value, it indicates that as the number of alternate alleles increases, trait expression grade decreases.

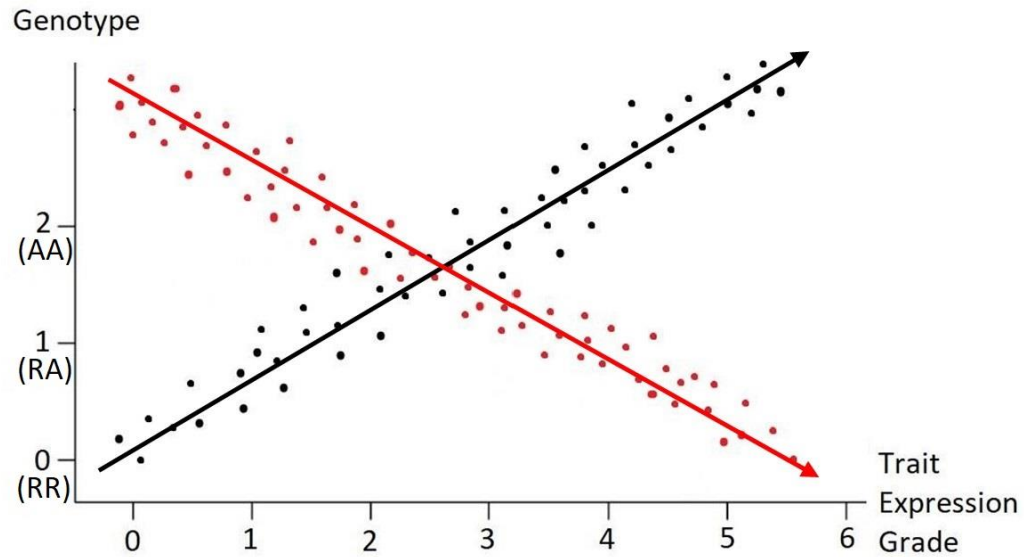


Figure 3.4 Visualization of positive (black dots) and negative (red dots) Spearman's correlations in a scatterplot. RR: homozygote for the reference allele, RA: heterozygote, AA: homozygote for the alternate allele. This chart does not contain actual data generated during this study; it is included purely for illustrative purposes.

3.10.3 Linkage disequilibrium analysis

In order to differentiate the causative variants from the variants that are in linkage disequilibrium (LD), LDlink software was used in this project (Machiela & Chanock, 2015). LDlink is a freely available and easy-to-use web-based tool that analyzes LD between SNPs (<https://ldlink.nci.nih.gov/>). LDlink consists of different modules for different purposes. LDmatrix, LDhap and LDpair modules of the LDlink software are utilized for different purposes in this study. The LDmatrix module is used to generate a heatmap which is a graphical display of the LD measures between two or more SNPs. The LDhap module calculates the haplotypes that were most frequently observed for a list of SNPs in a specific population. Lastly, LDpair is used to calculate the haplotype and allele frequencies and LD measures for a pair of SNPs (Machiela & Chanock, 2015). LDlink requires SNP reference (rs) numbers or genomic positions as input. The population and the version of the reference genome (e.g., GRCh38 and GRCh37) can be specified.

Chapter 4 Results

4.1 Dental traits

Figures 4.1, 4.2, and 4.3 below show the distribution of shoveling and Carabelli's cusp expression grades observed in the study participants. Individuals who did not exhibit any degree of shoveling and Carabelli's cusp (i.e., grade 0) were not included in the figures.

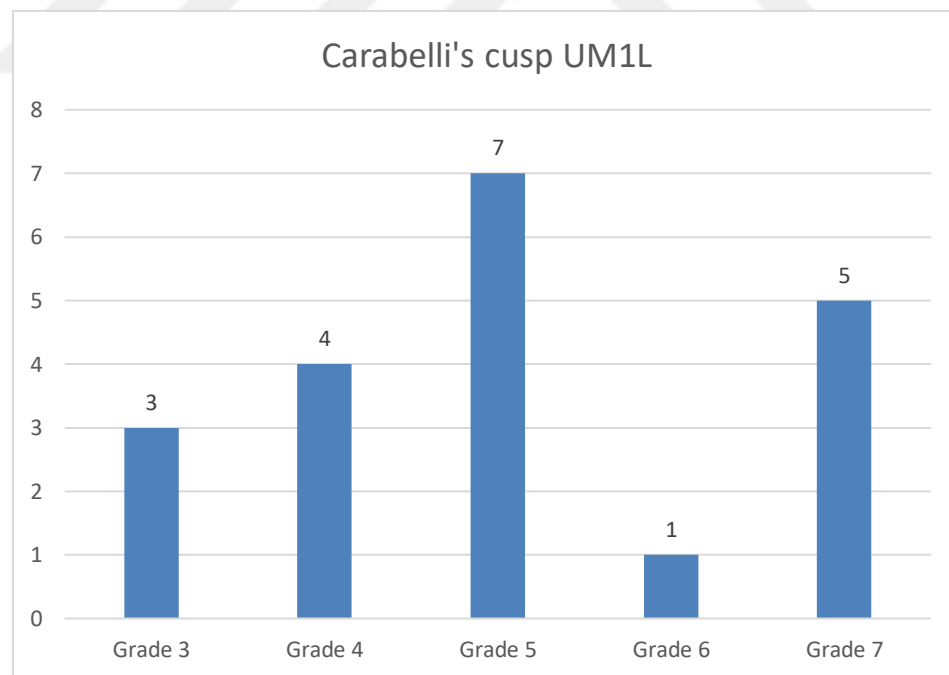


Figure 4.1 Distribution of Carabelli's cusp grades in upper left first molars (UM1L)

Figure 4.1 shows that three individuals exhibited grade 3, and four individuals had grade 4 Carabelli's cusp expression. A total of thirteen participants exhibited grades 5 (seven individuals), 6 (one individual), and 7 (five individuals) which are considered as the cusp form of Carabelli's trait (Marado and Campanacho, 2013).

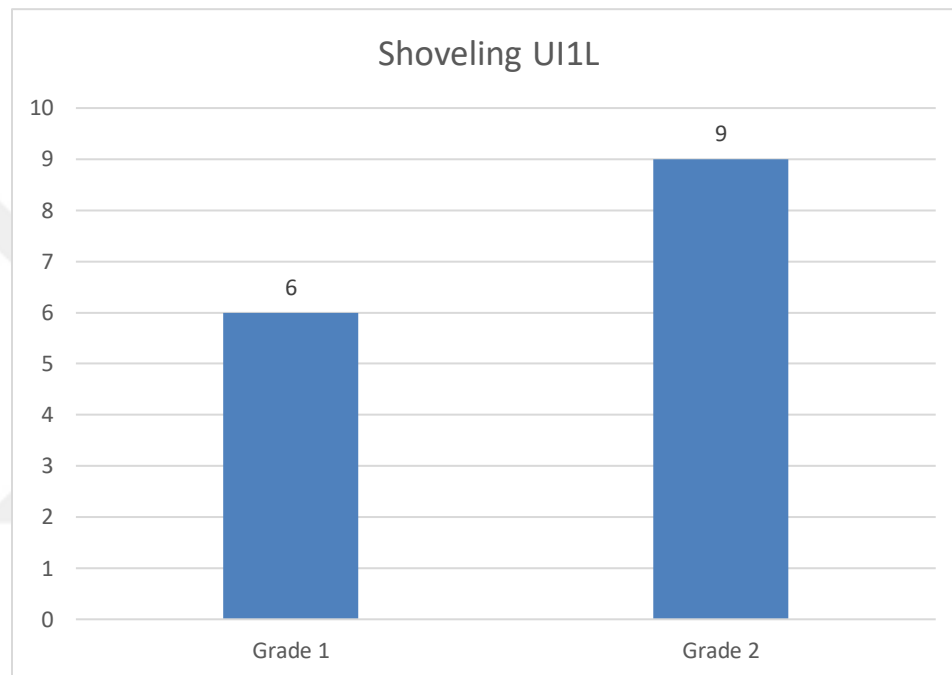


Figure 4.2 Distribution of shoveling grades in upper left central incisors (UI1L)

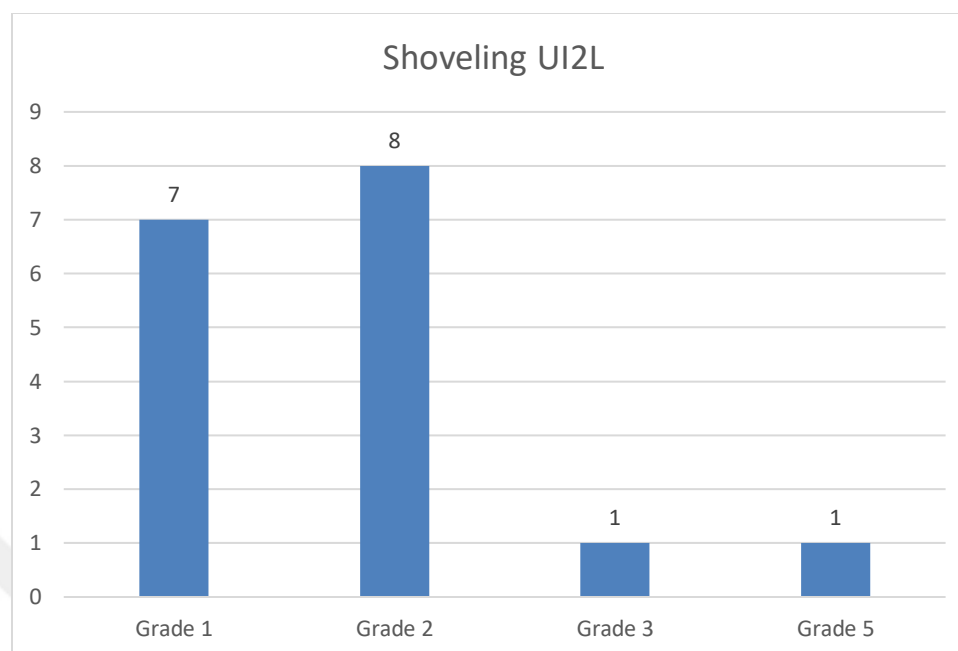


Figure 4.3 Distribution of shoveling grades in upper left lateral incisors (UI2L)

Majority of the participants had weak expression of shoveling (i.e., grades 1 and 2) in both upper central and lateral incisors. Six individuals had grade 1, and nine individuals had grade 2 shoveling in their upper left central incisors (UI1L) (Figure 4.2). Figure 4.3 shows that a total of fifteen individuals exhibit grade 1 (seven individuals) and grade 2 (eight individuals) shoveling expression in their lateral incisors (UI2L). One individual displayed grade 3, and one individual expressed grade 5 shoveling in the upper left lateral incisors (UI2L).

4.2 Statistical analysis

The Spearman's rank correlation test yielded moderate significant associations between fifteen loci from four different genes and the two dental traits (i.e., shoveling in UI1L and UI2L and Carabelli's cusp in UM1L). Two loci in WNT10A, seven loci in SOSTDC1, one locus in FGF4, and five loci in FGF3 were significantly correlated with the dental traits. Both negative and positive correlations were found for several loci (Table

4.1). Figure 4.4 shows the locations of the variants that were significantly correlated with the dental traits. The variant calling pipeline did not detect any variation in BMP4 and NKX2-3 and therefore these genes were not included in the statistical analysis.

Among the detected variants, three of them were INDELs. Their rs ID numbers gave inconsistent results in the NCBI's website (<https://www.ncbi.nlm.nih.gov/>). Therefore, instead of the rs numbers, genomic positions were used throughout the text. Issues related to INDEL calling were discussed in section 5.1 of the Discussion chapter in this dissertation.

Chr:Position	SNP/INDEL ID	Gene	Rho	Sig. (2-tailed)	N	Associated trait(s)
2:218891661	rs7349332	WNT10A	-.437	.026	26	Shoveling UI2L
2: 218886215	N/A	WNT10A	-.458	.019	26	Carabelli's cusp UM1L
7:16460517	rs10270805	SOSTDC1	.378	.023	36	Shoveling UI1L
			.430*	.009	36	Shoveling UI2L
7:16460784	rs9791629	SOSTDC1	.378	.023	36	Shoveling UI1L
			.430*	.009	36	Shoveling UI2L
7:16463435	rs12699800	SOSTDC1	-.329	.050	36	Carabelli's cusp UM1L
7:16464081	rs11761180	SOSTDC1	-.329	.050	36	Carabelli's cusp UM1L
7:16464648	rs6971172	SOSTDC1	.340	.043	36	Shoveling UI1L
			.388	.019	36	Shoveling UI2L
			-.366	.028	36	Carabelli's cusp UM1L
7:16464881	rs12699802	SOSTDC1	-.329	.050	36	Carabelli's cusp UM1L
7:16465716	rs6945425	SOSTDC1	.416	.012	36	Shoveling UI1L
			.341	.042	36	Shoveling UI2L
11:69775191	rs9666584	FGF4	.450*	.006	36	Shoveling UI2L
11:69813640	rs868979042	FGF3	.354	.034	36	Carabelli's cusp UM1L
11:69813824	rs146663171	FGF3	.342	.041	36	Shoveling UI1L
11:69813844	rs201295694	FGF3	.437*	.008	36	Shoveling UI1L

			.412	.013	36	Shoveling UI2L
11:69811473	N/A	FGF3	-.349	.037	36	Carabelli's cusp UM1L
11:69813652	N/A	FGF3	.405	.014	36	Carabelli's cusp UM1L

Table 4.1 Spearman's rank correlation coefficients (ρ), p values and dental traits that are associated with each locus. Asterisks (*) indicate correlations significant at 0.01 level. The rest of the correlations are significant at the 0.05 level. UI1L= Upper left first (central) incisor, UI2L= Upper left second (lateral) incisor, UM1L= Upper left first molar.

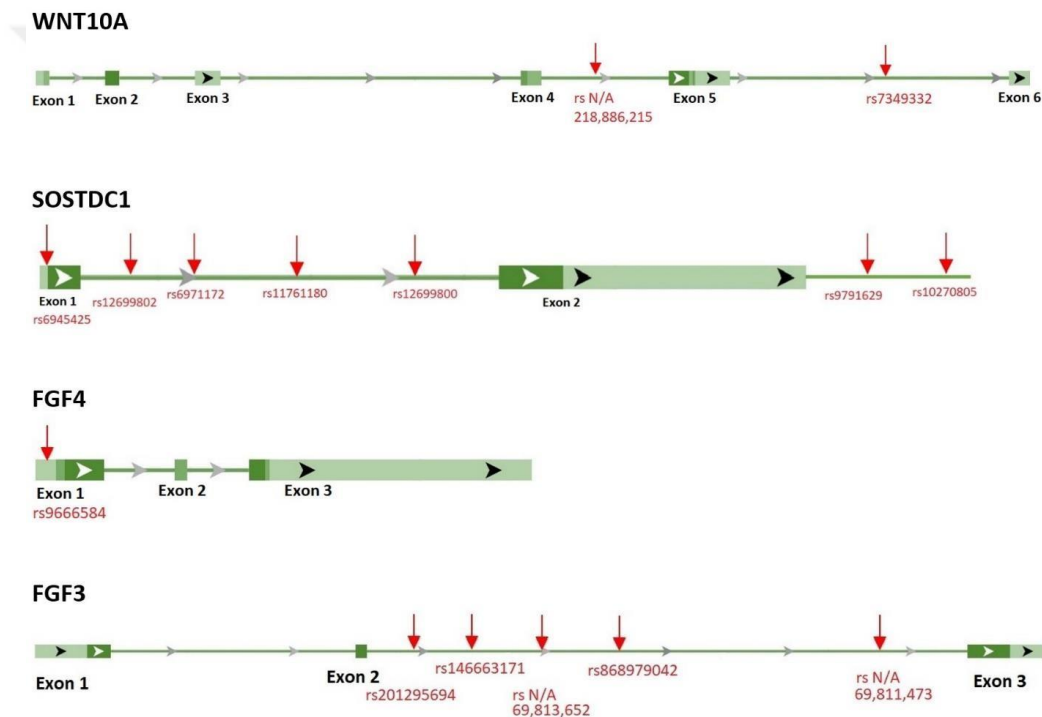


Figure 4.4 Locations of the detected variants (Red arrows). Dark green bars indicate coding exons, and the light green bars indicate non-coding untranslated regions (UTRs). Horizontal lines that connect exons show the introns. White and black arrows indicate the direction of transcription from 5' to 3'. Grey arrows in the introns indicate the direction of splicing. Green bars and connecting lines showing exon and intron regions were taken from the National Center for Biotechnology Information's website (<https://www.ncbi.nlm.nih.gov/>).

As Table 4.2 shows, most of the variants detected in this study (i.e., 11) are intron variants. Two of the detected variants are downstream gene variants, and two of them are 5 prime UTR variants. Three of the variants were INDELs and the rest of the variants were

SNPs. No reference SNP information (rs number) was available in the National Center for Biotechnology (NCBI)'s dbSNP database for the variants located at Chr2: 218886215, Chr11: 69811473 and Chr11:69813652 with the reference and alternate alleles shown in Table 4.2. Therefore, the Variant ID columns for these variants were denoted with N/A.

Genomic Position	Variant ID	Variant type	Reference allele	Alternate allele	Gene	Consequence
Chr2:218891661	rs7349332	SNP	C	T	WNT10A	intron variant
Chr2:218886215	N/A	INDEL	GCA	G	WNT10A	intron variant
Chr7:16460517	rs10270805	SNP	A	G	SOSTDC1	downstream gene variant
Chr7:16460784	rs9791629	SNP	C	T	SOSTDC1	downstream gene variant
Chr7:16463435	rs12699800	SNP	C	T	SOSTDC1	intron variant
Chr7:16464081	rs11761180	SNP	A	G	SOSTDC1	intron variant
Chr7:16464648	rs6971172	SNP	G	C	SOSTDC1	intron variant
Chr7:16464881	rs12699802	SNP	T	C	SOSTDC1	intron variant
Chr7:16465716	rs6945425	SNP	T	C	SOSTDC1	5 prime UTR variant
Chr11:69775191	rs9666584	SNP	A	G	FGF4	5 prime UTR variant
Chr11:69813640	rs868979042	SNP	G	*,A	FGF3	intron variant
Chr11:69813824	rs146663171	SNP	G	*,A	FGF3	intron variant
Chr11:69813844	rs201295694	SNP	T	A,*	FGF3	intron variant
69811473	N/A	INDEL	GA	G	FGF3	intron variant
69813652	N/A	INDEL	ATGGATGGATGGGTGGATGGC	*,A	FGF3	intron variant

Table 4.1 Ensemble Variant Effect Predictor (McLaren et al., 2010) results with possible consequences of the variants detected in this study. * indicates an allele deletion.

Table 4.3 shows the frequencies of genotypes and alleles for the fifteen variants detected in this study. Distributions of shoveling and Carabelli's cusp grades in each genotype were provided through Tables A.1 and A.15 in the Appendix A as well.

Genomic position	Variant ID	Ref Allele	Alt Allele	Genotype Frequency	Allele Frequency
Chr2: 218891661	rs7349332	C	T	C/C=9 (34.6%) C/T=15 (57.69%) T/T= 2 (7.69%)	C=63.46% T=36.54%
Chr2: 218886215	N/A	GCA	G	GCA/GCA=2 (7.69%) GCA/G=24 (92.31%)	GCA=53.85% G=46.15%
Chr7: 16460517	rs10270805	A	G	A/A=1 (2.78%) A/G=8 (22.22%) G/G=27 (75%)	A=13.89% G=86.11%
Chr7: 16460784	rs9791629	C	T	C/C=1 (2.78%) C/T=8 (22.22%) T/T=27 (75%)	C=13.89% T=86.11%
Chr7: 16463435	rs12699800	C	T	C/C=5 (13.89%) C/T=18 (50%) T/T=13 (36.11%)	C=38.89% T=61.11%
Chr7: 16464081	rs11761180	A	G	A/A=5 (13.89%) A/G=18 (50%) G/G=13 (36.11%)	A=38.89% G=61.11%
Chr7: 16464648	rs6971172	G	C	G/G=1 (2.78%) G/C=7 (19.44%) C/C=28 (77.78%)	G=12.5% C=87.5%
Chr7: 16464881	rs12699802	T	C	T/T=5 (13.89%) T/C=18 (50%) C/C=13 (36.11%)	T=38.89% C=61.11%
Chr7: 16465716	rs6945425	T	C	T/T=1 (2.78%) T/C=9 (25%) C/C=26 (72.22%)	T=15.28% C=84.72%
Chr11: 69775191	rs9666584	A	G	A/A=4 (11.11%) A/G=19 (52.78%) G/G=13 (36.11%)	A=37.5% G=62.5%
Chr11: 69813640	rs868979042	G	*,A	G/G=26 (72.22%) G/A=1 (2.78%) G/=7 (19.44%) */*=2 (5.56%)	G=83.33% A=1.39% *=15.28%
Chr11:69813824	rs146663171	G	*,A	G/G=28 (77.78%) G/=7 (19.44%) G/A=1 (2.78%)	G=88.89% *=9.72% A=1.39%

11:6981384 4	rs201295694	T	A,*	T/T=30 (83.33%) T/A=4 (11.11%) T/=2 (5.56%)	T=91.67% A=5.56% *=2.78%
11:6981147 3	N/A	GA	G	GA/GA=26 (72.22%) GA/G=10 (27.78%)	GA=86.11% G=13.89%
11:6981365 2	N/A	ATG GAT GGA TGG GTG GAT GGC	*,A	Ref/Ref=28 (77.78%) Ref/=5 (13.89%) Ref/A=2 (5.56%) */*=1 (2.78%)	Ref=87.5% *=9.72% A=2.78%

Table 4.3 Genotype and allele frequencies for detected variants. * indicates an allele deletion. The word “Ref” was used instead of the actual sequence in the last variant.

4.3 Linkage Disequilibrium Analysis

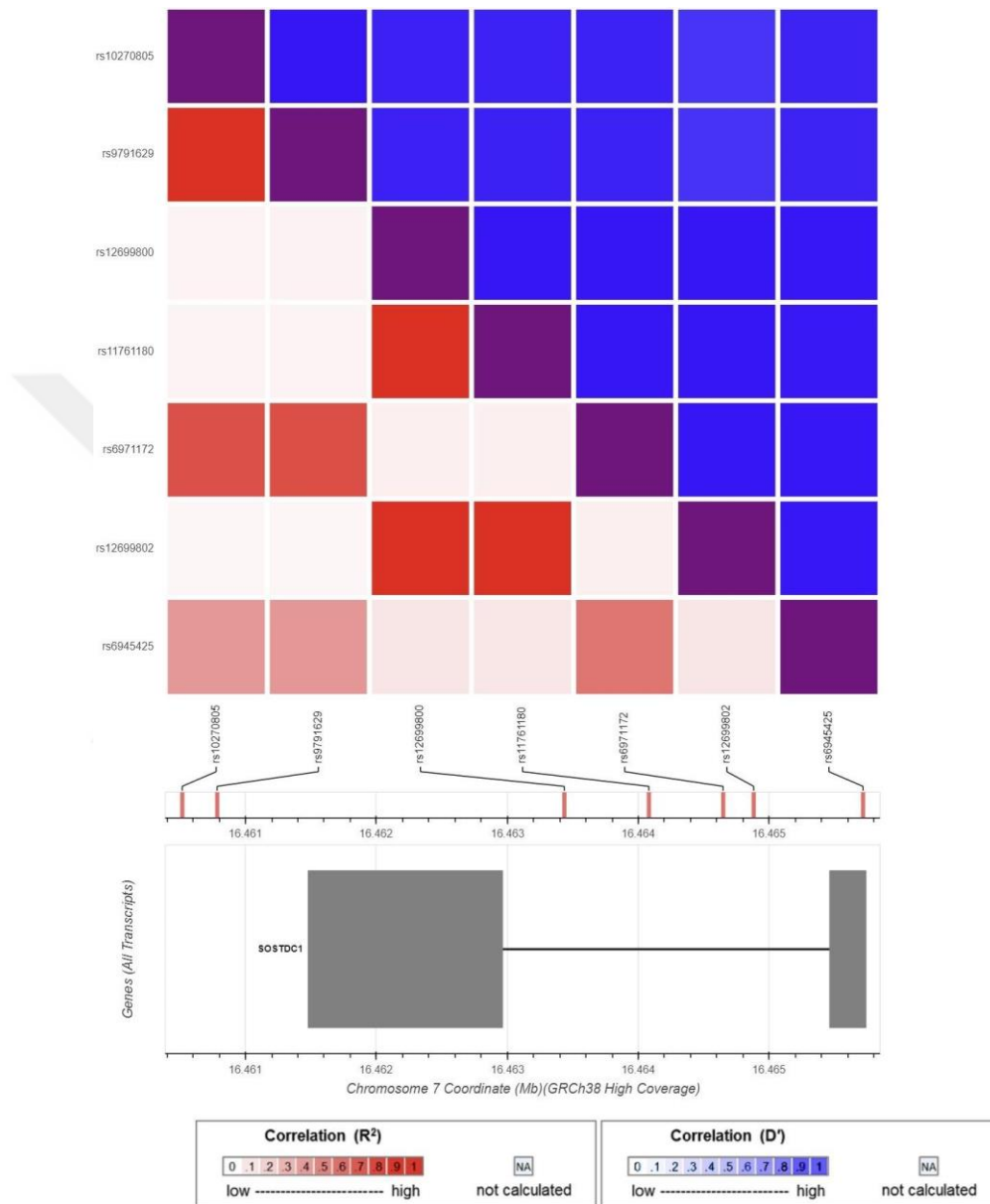


Figure 4.5 Linkage disequilibrium heatmap showing correlations between SOSTDC1 variants.

RS_number	rs10270805	rs9791629	rs12699800	rs11761180	rs6971172	rs12699802	rs6945425
rs10270805	1.0	0.99	0.909	0.902	0.906	0.813	0.89
rs9791629	0.99	1.0	0.923	0.915	0.918	0.826	0.904
rs12699800	0.909	0.923	1.0	0.994	1.0	1.0	0.988
rs11761180	0.902	0.915	0.994	1.0	1.0	0.998	0.984
rs6971172	0.906	0.918	1.0	1.0	1.0	1.0	0.988
rs12699802	0.813	0.826	1.0	0.998	1.0	1.0	0.988
rs6945425	0.89	0.904	0.988	0.984	0.988	0.988	1.0

Table 4.4 Linkage disequilibrium D' values for SOSTDC1 variants.

RS_number	rs10270805	rs9791629	rs12699800	rs11761180	rs6971172	rs12699802	rs6945425
rs10270805	1.0	0.977	0.044	0.043	0.74	0.036	0.416
rs9791629	0.977	1.0	0.046	0.045	0.759	0.037	0.427
rs12699800	0.044	0.046	1.0	0.983	0.059	0.985	0.1
rs11761180	0.043	0.045	0.983	1.0	0.059	0.977	0.099
rs6971172	0.74	0.759	0.059	0.059	1.0	0.06	0.567
rs12699802	0.036	0.037	0.985	0.977	0.06	1.0	0.101
rs6945425	0.416	0.427	0.1	0.099	0.567	0.101	1.0

Table 4.5 Linkage disequilibrium r² values for SOSTDC1 variants.

RS Number	Position (GRCh38 High Coverage)	Allele Frequencies	Haplotypes				
rs10270805	chr7:16460517	G=0.935, A=0.065	G	G	A	G	G
rs9791629	chr7:16460784	T=0.934, C=0.066	T	T	C	T	T
rs12699800	chr7:16463435	C=0.569, T=0.431	C	T	C	C	C
rs11761180	chr7:16464081	A=0.57, G=0.43	A	G	A	A	A
rs6971172	chr7:16464648	C=0.928, G=0.072	C	C	G	C	G
rs12699802	chr7:16464881	T=0.565, C=0.435	T	C	T	T	T
rs6945425	chr7:16465716	C=0.881, T=0.119	C	C	T	T	T
Haplotype Count			2232	2132	297	231	58
Haplotype Frequency			0.4457	0.4257	0.0593	0.0461	0.0116

Figure 4.6 Five most frequent haplotypes for SOSTDC1 variants.

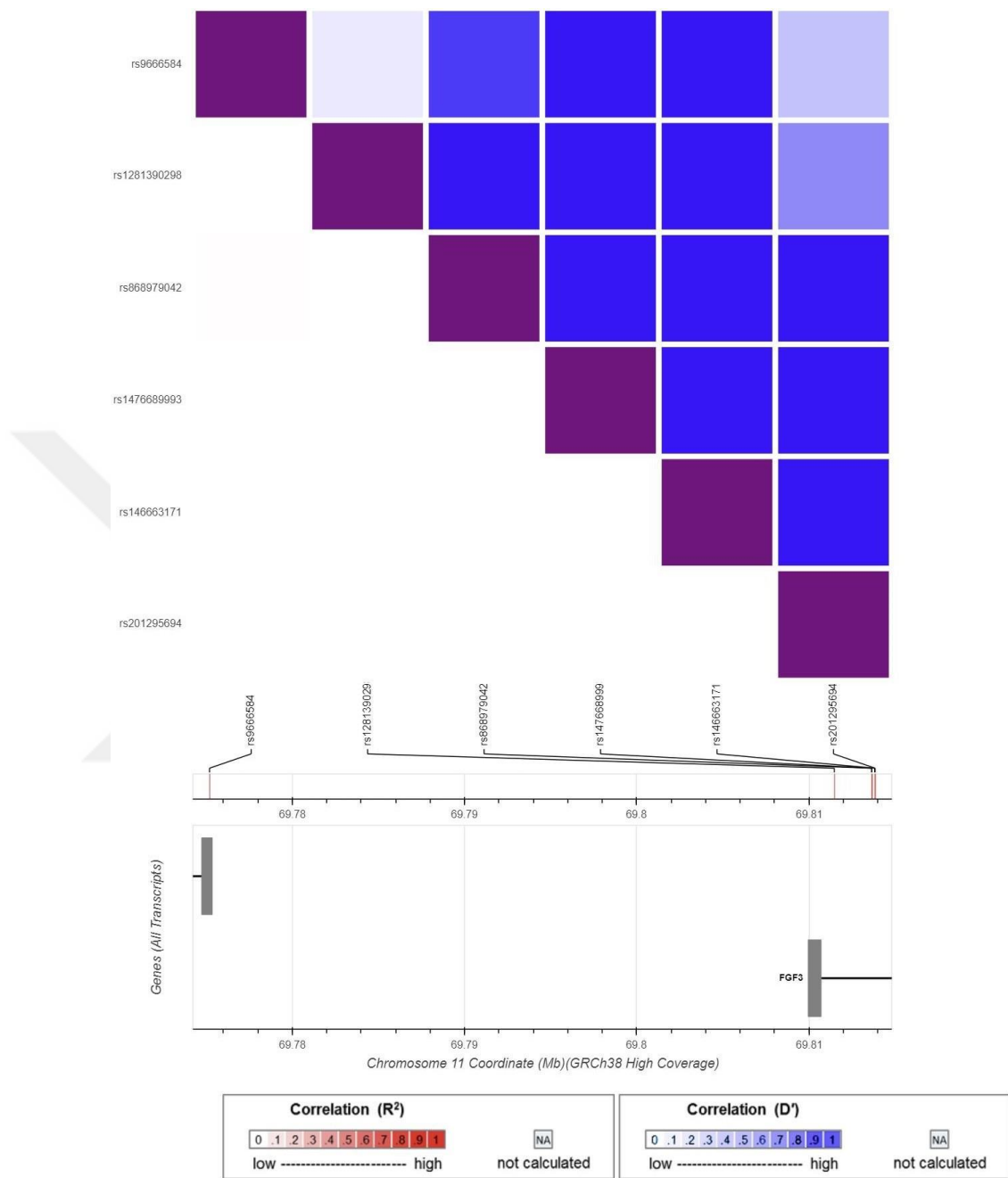


Figure 4.7 Linkage disequilibrium heatmap showing correlations between FGF3 and FGF4 variants.

RS_number	rs9666584	rs1281390298	rs868979042	rs1476689993	rs146663171	rs201295694
rs9666584	1.0	0.086	0.794	1.0	1.0	0.243
rs1281390298	0.086	1.0	1.0	1.0	1.0	0.467
rs868979042	0.794	1.0	1.0	1.0	1.0	1.0
rs1476689993	1.0	1.0	1.0	1.0	1.0	1.0
rs146663171	1.0	1.0	1.0	1.0	1.0	1.0
rs201295694	0.243	0.467	1.0	1.0	1.0	1.0

Table 4.6 Linkage disequilibrium D' values for FGF3 and FGF4 variants.

RS_number	rs9666584	rs1281390298	rs868979042	rs1476689993	rs146663171	rs201295694
rs9666584	1.0	0.001	0.002	0.0	0.0	0.001
rs1281390298	0.001	1.0	0.0	0.0	0.0	0.001
rs868979042	0.002	0.0	1.0	0.0	0.0	0.0
rs1476689993	0.0	0.0	0.0	1.0	0.0	0.0
rs146663171	0.0	0.0	0.0	0.0	1.0	0.0
rs201295694	0.001	0.001	0.0	0.0	0.0	1.0

Table 4.7 Linkage disequilibrium r2 values for FGF3 and FGF4 variants.

As was mentioned in the section 4.2 Statistical analysis, genomic positions for the INDELs at chr2:218886215, chr11:69811473 and chr11:69813652 rather than the associated rs number were used for the LD analysis. When the genomic positions were entered in the LDlink's LDmatrix tool, the software automatically associated the genomic positions with a rs number. The associated rs numbers were rs962630813 for chr2:218886215, rs1281390298 for chr11:69811473 and rs1476689993 for chr11:69813652.

Figures 4.5 and 4.7 depict the linkage disequilibrium (LD) values in heatmaps for the genes Sostdc1, Fgf3 and Fgf4. Since Fgf3 and Fgf4 are located in the same chromosome

(i.e., chromosome 11), they were included in the same LD analysis. The color intensity in the figures indicates the strength of the LD between the variants. Red color shows the r^2 measurements for each variant pair, and the blue color displays the D' measurements for each variant pair. Tables 4.4, 4.5, 4.6, and 4.7 represent the numerical values of the heatmaps in Figures 4.5 and 4.7. Figure 4.6 shows the most frequently observed haplotypes for the five SOSTDC1 variants. Due to the issues related to the rs numbers of three indels in the WNT10A and FGF3 genes, haplotypes for these genes were not included in the analysis. Figure 4.8 shows the LD measures for the Wnt10a variant pairs.

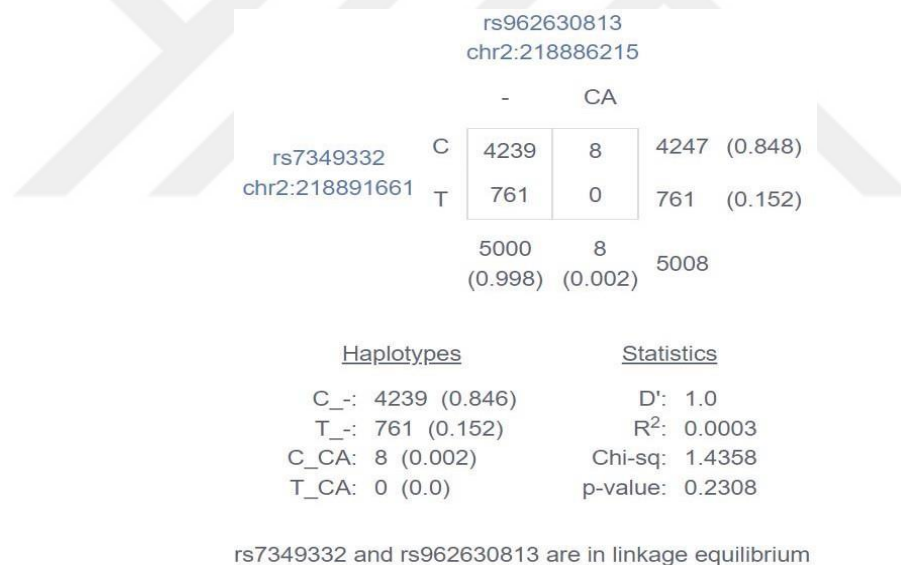


Figure 4.8 Linkage disequilibrium results for the WNT10A variants.

Chapter 5 Discussion and Conclusion

The genetic pathways that are involved in the embryonic development of teeth have been described in detail over the past few decades. The hundreds of genes that ultimately produce teeth have been identified and the gene network interactions that cascade during embryonic development to produce teeth are understood (Pillas et al., 2010). However, only few genes that play a significant role in the production of common dental nonmetric variation have been identified to this day (Kimura et al., 2009, 2015; Lee et al., 2012). The etiology behind many nonmetric dental traits still remains to be elucidated.

Shovel-shaped incisors and Carabelli's cusp in molars, the two dental traits that have been investigated in this dissertation, are the most commonly used dental traits in biodistance, kinship, and phylogenetic studies and have been traits of interest since the early 1900s (Hrdlička, 1920; Marado & Campanacho, 2013). Observations of marked differences in their expression between populations as well as twin and family studies that demonstrate morphological concordances between twin pairs and family members led researchers to assume that shoveling and Carabelli's cusp are highly genetically controlled without direct genetic sequence data. However, despite their very frequent use in anthropological studies, very few attempts have been made to discover the genetic underpinnings of shovel-shaped incisors and Carabelli's cusp.

In this study, six candidate genes previously identified as being involved in dental development, *NKX2-3*, *SOSTDC1*, *BMP4*, *FGF3*, *FGF4*, and *WNT10A*, were investigated to identify if variation in their basal DNA sequences has any impact on whether or not an individual expresses shovel-shaped incisors and/or Carabelli's cusp and whether or not this genetic variation can explain the level of expression of these traits.

For each candidate gene present in each of the 36 participants in the study, the entire gene region including the 5' and 3' untranslated regions (UTRs), all intronic and exonic regions, and significant sections of the upstream and downstream regions adjacent to the candidate genes were sequenced. The genes were targeted and amplified using long-range PCR and sequenced using next-generation sequencing (NGS) techniques. The spearman's correlation test was utilized to determine whether the degree of trait expression demonstrates any relationship with the presence or absence of reference and alternate alleles in the genotype of an individual. Results showed that fifteen SNPs/INDELs belonging to *SOSTDC1*, *FGF3*, *FGF4* and *WNT10A* genes are significantly associated with shoveling and Carabelli's cusp expression.

5.1 Issues with INDEL Calling

One challenge that was encountered during data analysis was that the reference/alternate alleles and rs ID numbers for the INDELs at Chr2:218886215, Chr11:69811473, and Chr11:69813652 that were detected by the GATK variant calling pipeline did not match the reference and alternate alleles in the NCBI's dbSNP database (build 155). For example, the GATK pipeline in this study detected a variant at Chr2:218886215 with the rs ID number of rs962630813. When this rs ID number is entered in the NCBI's dbSNP database to further investigate this variant, it showed that the

reference allele is G and the alternate allele is A at this locus. However, the variant that was detected at this exact locus in this study had GCA as the reference allele and G for the alternate allele. Similarly, the same issue was encountered with the variant at Chr11:69811473 with the rs ID number of rs1281390298. NCBI's dbSNP website shows that this rs number is associated with G reference allele and A alternate allele whereas the detected alleles in this study were GA for reference and G for alternate alleles. Lastly, the GATK detected a variant at Chr11:69813652 in this study but there was not a rs ID number associated with this locus. When this genomic position (i.e., Chr11:69813652) is entered in the NCBI, further investigation of this locus reveals that this region of the human genome is highly variable containing many duplicate SNPs and INDELs with different rs IDs at the same locus. Due to these errors occurring during INDEL calling, rs numbers for these variants were excluded and only the genomic positions of the variants were used throughout this dissertation to prevent any possible misunderstandings.

There can be many reasons for the INDEL errors which make INDEL calling inaccurate. Fang et al. (2014) mention that errors that are likely to occur during PCR amplification in NGS library preparation might be one of the sources for inaccurate INDEL calling. In addition, there are still so many unknown regions and gaps in the human reference genome that complicates the detection of rare INDEL variants. Realignment errors and incomplete reference genome issues are some of the challenges that have been encountered by previous studies (Li, 2014) and need to be addressed by future work. Since the GATK variant calling pipeline is heavily based on aligning the novel reads to the reference genome, it is important to use the same reference genome version, which are also called assemblies or builds, during every step of the research methodology. In this study,

the latest build of the human reference genome, GRCh38/hg38, was used from the PCR primer design step to the variant calling analysis. Therefore, the possibility that the INDEL errors that occurred in this study is due to the inconsistent reference genome versions can be ruled out.

5.2 Significant Single Nucleotide Polymorphisms (SNPs) and Insertion or Deletion (INDELs)

Seven variants, rs7349332 and Chr2:218886215 in WNT10A, rs12699800, rs11761180, rs6971172, rs12699802 in SOSTDC1, and Chr11:69811473 in FGF3 are negatively associated with shoveling and Carabelli's cusp (see Table 4.1 in Results). The rho values range between -.329 and -.458 which indicates a moderate negative correlation. A negative Spearman's rho value means that an individual having an increased number of the alternate allele(s) is correlated with decreased level of trait expression. Of the seven negatively correlated variants, six of them were associated with the general expression of Carabelli's cusp in any form and likewise further the degree of expression of Carabelli's cusp. One of the seven aforementioned variants was associated with shoveling (UI2L). This result is interesting because Carabelli's cusp expression and degree of expression seems to be consistently showing negative associations with the alternate alleles in these six loci. In other words, higher grades of Carabelli's cusp expression are more correlated with the reference allele than the alternate allele at these loci. These negative correlations need to be verified in the future in a study with a larger sample size and a more geographically diverse group of study participants.

Variants within the genes *NKX2-3* and *BMP4* showed no correlations with any of the traits in this study. However, this result should not mislead the readers to believe that these two genes have no function specific to the production of shoveling and Carabelli's

cuspid. No correlations with these genes and the dental traits in this study might be due to the small sample size. Future work with larger sample size is needed to confirm their role in dental trait formation.

Nine variants rs10270805, rs9791629, rs6971172, rs6945425 in SOSTDC1, rs9666584 in FGF4, rs868979042, rs146663171, rs201295694 and Chr11:69813652 in FGF3 are positively correlated with shoveling and Carabelli's cusp (see Table 4.1 in Results). Correlations are moderate ranging between .340 and .450. Positive correlations indicate that the number of alternate alleles is associated with the higher grades of shoveling and Carabelli's cusp.

Five SNPs showed correlations with more than one trait. SNPs rs10270805, rs9791629, rs6971172, rs6945425 and rs201295694 were positively correlated with shoveling in both central and lateral incisors. These results are not surprising given the previous observations that shoveling expression in central and lateral incisors are highly correlated with each other (Hasegawa et al., 2009). In other words, if an individual expresses shoveling in the upper central incisors, there is a high likelihood that their lateral incisors will exhibit some degree of shoveling as well. This observation has been previously interpreted as the same genetic factors influence both central and lateral incisor morphology (Hasegawa et al., 2009). The SNP rs6971172 was the only variant that was correlated with both shoveling in both central and lateral incisors and Carabelli's trait in the upper first molar (Table 4.1 in Results). This SNP was positively correlated with shoveling and negatively correlated with Carabelli's cusp. This result also is not surprising because shoveling and Carabelli's cusp are usually considered mutually exclusive in terms of their manifestations across Asian and European populations. Individuals from these

geographies typically display either shovel shaping or Carabelli's cusp but not both in most cases.

Most of the variants detected in this study have no publications about them and no reports in ClinVar database at the NCBI's website (<https://www.ncbi.nlm.nih.gov/clinvar/>). ClinVar is a public database of genetic variation and identifies any disease that has been previously associated with genetic variants. The fact that the variants detected in this study have no publications about them and yet are included in the ClinVar database is most likely due to the possibility that these SNPs/INDELs might have been identified by previous GWAS (Genome- Wide Association Studies) but the functions of these SNPs/INDELs are yet to be discovered by future studies. SNP rs7349332 is the exception to this; it is the only variant that has been mentioned in previous publications. Previous publications associated this SNP with hair loss, hair shape, and increased risk of colorectal cancer (Li et al., 2017). In addition, as already mentioned in Chapter 2, Kimura et al. (2015) found that rs7349332 in the *WNT10A* gene was significantly correlated with crown size. In addition to crown size, the alternate allele of this SNP was associated with hair curl (Eriksson et al., 2010).

5.3 Linkage Disequilibrium of Variants

Several variants identified in this study show high degree of linkage disequilibrium (LD) with each other (See Figures 4.5, 4.7, and 4.8 and also Tables 4.4 through 4.7). LD is defined as non-random coinheritance of different loci across the genome (McEvoy et al., 2011) and can be caused by many mechanisms including natural selection, genetic drift, and population structure (Loh et al., 2013). LD decays over time due to segregation and recombination events during meiosis (Baird, 2015; McEvoy et al., 2011). D' (D prime) and r^2 (R squared) are the most common measures to estimate LD but they are not equivalent

to each other. Researchers might favor one measure over the other depending on the purpose of their research. Mueller (2004) argues that r^2 is a better predictor than D' for gene mapping and association studies when small sample size is the issue. D' , on the other hand, tends to give higher values because it does not account for allele frequencies in a population. Therefore, caution should be exercised when interpreting unmatching r^2 and D' values.

Low r^2 and high D' values for the same variants were the case for *FGF3* and *FGF4* variants in this study (Figure 4.7, and Tables 4.6 and 4.7). For the *SOSTDC1* variants, r^2 and D' values are more consistent with each other showing strong LD. High r^2 and D' values indicate that these variants are in linkage disequilibrium however it is important to stress that they might not be the true causal variants that ultimately result in shoveling or Carabelli's cusp. The fact that they are all linked to one another could be immaterial in terms of the fundamental biological mechanism behind the production of shovel shaping and Carabelli's cusp. They could also all be potentially linked to an as-of-yet unelucidated variant that functions as the actual biological mechanism underlying the production of shoveling or Carabelli's cusp. A larger sample size and a broader range of genes screened will help resolve these ambiguities about whether or not these variants are the cause of Carabelli's cusp and shoveling.

Identifying any population-specific genotype and phenotype pattern is tricky due to the small sample size in this study. Although this preliminary study is based on a small sample size and lacks power, fifteen variants belonging to four different genes were significantly associated with the two dental nonmetric traits. It is useful to be cautious when interpreting these results since the significant results might only have occurred by chance

and not reflect the reality due to the small sample size. However, some of the significant variants identified here (e.g., rs7349332) were also found to be significantly associated with dental traits by previous studies such as rs7349332 in the *WNT10A* gene (Kimura et al., 2015).

5.4 Location and consequence of variants

Most of the variants detected in this study are intron variants (Table 4.2). See Figure 4.4 in the Results chapter for a visualization of the location of the variants detected. Introns are segments of DNA within a gene that do not code for amino acids. Introns were previously considered “junk” DNA because they are excised by endonucleases during transcription (Parenteau & Abou Elela, 2019). However, the important functions of introns during gene expression have been recognized in recent years (Jo & Choi, 2015). During transcription, DNA is transcribed into a pre-mRNA molecule which contains both intron and exon sequences. Subsequently, because introns are not needed for translation of RNA into proteins, they are excised from the pre-mRNA transcript during a process called splicing. After the introns are removed, the exons are joined together to form a mature mRNA that will produce the final protein product. However, splicing does not always occur in the same way. “Alternative splicing” is a term that is used to describe a highly regulated molecular mechanism during which the introns and exons are spliced and joined in different locations. Alternative splicing is very common in humans occurring in 95% of the multi-exon genes (Jo & Choi, 2015).

It then begs the question, if the introns are not translated into proteins, then why do the cells spend all that energy to transcribe, splice and degrade them? Why do we even keep the introns in our genome? Although there is no single consensus among researchers

on this issue, there have been attempts to explain why introns exist. One explanation is that introns are evolutionarily advantageous because they are an important source for new genes and gene products. The mechanism of alternative splicing can lead to the production of different isoforms of proteins that are produced from the same gene. In other words, many forms of proteins can be generated from a single gene. In addition, studies showed that gene expression is enhanced in genes with introns compared to the genes that do not contain any introns (Shabalina et al., 2010).

Therefore, although the intron variants found in this study might not have any consequences in terms of the proteins' amino acid sequences and therefore cannot alter protein geometry and potentially function, it might be plausible to expect that those variants somehow alter to gene expression via the process of alternative splicing and gene expression enhancement. This points to the importance of deep sequencing (or sequencing with deep coverage) of the entire gene including both introns and exons rather than only focusing on the exonic regions. Variation in introns, as well as the 3' and 5' untranslated regions can certainly alter gene expression.

5.5 Conclusion

This preliminary study was an attempt to elucidate the genetic underpinnings of shovel- shaped incisors and Carabelli's trait. Significant associations were found between fifteen short variants in four genes and the two dental traits of interest despite the small sample size. The hypothesis that some short variants (SNPs and INDELs) are strongly associated with the presence of shovel-shaped incisors and Carabelli's cusp is supported by the results of this study. This hypothesis can be preliminarily accepted for *FGF3*, *FGF4*, *SOSTDC1* and *WNT10A* but the results do not support this hypothesis for *NKX2-3* and

BMP4. Therefore, the conclusion may be preliminarily drawn that variation in the genes *FGF3*, *FGF4*, *SOSTDC1* and *WNT10A* might be involved in the formation of shoveling and Carabelli's cusp. However, more studies are needed to confirm if the genes *NKX2-3* and *BMP4* are involved in the production of shoveling and Carabelli's cusp.

Further, the hypothesis that the reference alleles are more strongly associated with trait expression than the alternate allele in a detected locus, or that alternate alleles confer a lower degree of trait expression, is supported by the results in this study. The spearman's rank correlation test, which measures the strength and direction of association, indicates that either the reference or alternate alleles in the fifteen loci confer a greater degree of Carabelli's cusp development and/or shovel shaping in individuals bearing the reference allele or a lower degree of Carabelli's cusp and/or shovel shaping in individuals with the alternate allele.

Finally, LD analysis provided strong support for the hypothesis which proposed that the variants detected in this study are non-randomly associated with each other, or in other words, they are in strong linkage disequilibrium (LD). This indicates that the variants detected in this study may not actually be causing shovel-shaping and Carabelli's cusp. Instead, they are "hitchhiking" meaning that they non-randomly travel together. The fact that the variants are in strong LD complicates the identification of the causal genes. Therefore, it is not clear which SNPs/INDELs are the real cause of the observed phenotypes. How these genes influence expression of shovel-shaping and Carabelli's cusp, and how they interact with other genes need to be elucidated in more detail by future studies.

5.6 Future directions

The majority of genotype-phenotype association studies focus only on the exonic regions of the candidate genes. However, significant roles that the introns, and the 3' and 5' untranslated regions (UTRs) play during gene expression are overlooked. More studies on the functional genomic assays to determine which regions of the introns and the 3' and 5' UTRs might be playing in gene expression are needed. The methodology might include chromatin immunoprecipitation sequencing (ChIP Seq) for identifying all protein-DNA interactions in order to determine the regions of protein and DNA interactions during transcription and splicing and potentially alternative splicing of these genes.

In addition, targeted sequencing and exome sequencing miss many structural variants (SVs) such as deletions, duplications, and copy number variations that potentially contribute to the dental phenotypes (Fang et al., 2014). Whole genome analysis with more participants for statistical significance can reveal SVs as well as short variants such as SNPs and INDELs.

A line of research that can benefit from this project is the field of tooth tissue engineering. Most of the tooth regeneration studies are done on mice. The molecular signals that control tooth development and morphology in humans are not completely understood. The genetic variants I found in this project may contribute to the existing knowledge of the genetics of tooth crown formation and help future studies that seek to regenerate whole teeth in humans with a predetermined shape and size.

Appendix A

	Genotype	
Carabelli's cusp UM1	GCA GCA	GCA G
0	-	12
3	-	3
4	-	3
5	-	5
6	1	-
7	1	1

Table A.1 Distribution of Carabelli's cusp (UM1) grades in each genotype at Chr:218886215 in WNT10A.

	Genotype		
Shoveling UI1L	AA	AG	GG
0	1	7	13
1	-	1	5
2	-	-	9
Shoveling UI2L			
0	1	7	11
1	-	1	6
2	-	-	8
3	-	-	1
5	-	-	1

Table A.2 Distribution of shoveling (UI1L and UI2L) grades in each genotype at rs10270805.

	Genotype		
Shoveling UI1L	CC	CT	TT
0	1	7	13
1	-	1	5
2	-	-	9
Shoveling UI2L			
0	1	7	11
1	-	1	6
2	-	-	8
3	-	-	1
5	-	-	1

Table A.3 Distribution of shoveling (UI1L and UI2L) grades in each genotype at rs9791629.

	Genotype		
Shoveling UI1L	GG	G*	GA
0	18	3	-
1	5	-	1
2	5	4	-

Table A.4 Distribution of shoveling (UI1L) grades in each genotype at rs146663171.

	Genotype		
Shoveling UI1L	TT	TA	T*
0	20	1	-
1	5	1	-
2	5	2	2
Shoveling UI2L			
0	18	1	-
1	7	-	-
2	3	3	2
3	1	-	-
5	1	-	-

Table A.5 Distribution of shoveling (UI1L and UI2L) grades in each genotype at rs201295694.

	Genotype		
Carabelli's cusp UM1	CC	CT	TT
0	2	4	10
3	-	3	-
4	1	2	1
5	1	6	-
6	-	1	-
7	1	2	2

Table A.6 Distribution of Carabelli's cusp (UM1) grades in each genotype at rs12699800.

	Genotype		
Carabelli's cusp UM1	AA	AG	GG
0	2	4	10
3	-	3	-
4	1	2	1
5	1	6	-
6	-	1	-
7	1	2	2

Table A.7 Distribution of Carabelli's cusp (UM1) grades in each genotype at rs11761180.

	Genotype		
Shoveling UI1L	TT	TC	CC
0	1	8	12
1	-	1	5
2	-	-	9
Shoveling UI2L			
0	1	7	11
1	-	1	6
2	-	1	7
3	-	0	1
5	-	0	1

Table A.8 Distribution of shoveling (UI1L and UI2L) grades in each genotype at rs6945425.

	Genotype		
Shoveling UI2L	AA	AG	GG
0	3	13	3
1	1	2	4
2	-	4	4
3	-	-	1
5	-	-	1

Table A.9 Distribution of shoveling (UI2L) grades in each genotype at rs9666584.

	Genotype			
Carabelli's cusp UM1	GG	GA	G*	**
0	14	-	2	-
3	2	-	1	-
4	3	-	-	1
5	4	-	2	1
6	1	-	-	-
7	2	1	2	-

Table A.10 Distribution of Carabelli's cusp (UM1) grades in each genotype at rs868979042.

	Genotype		
Shoveling UI1L	GG	GC	CC
0	1	6	14
1	-	1	5
2	-	-	9
Shoveling UI2L			
0	1	6	12
1	-	1	6
2	-	-	8
3	-	-	1
5	-	-	1
Carabelli's cusp UM1			
0	1	-	15
3	-	-	3
4	-	1	3
5	-	4	3

6	-	-	1
7	-	2	3

Table A.11 Distribution of shoveling (UI1L and UI2L) and Carabelli's cusp (UM1) grades in each genotype at rs6971172.

	Genotype		
Carabelli's cusp UM1	TT	TC	CC
0	2	4	10
3	-	3	-
4	1	2	1
5	1	6	-
6	-	1	-
7	1	2	2

Table A.12 Distribution of Carabelli's cusp (UM1) grades in each genotype at rs12699802.

	Genotype		
Shoveling UI2L	CC	CT	TT
0	1	10	1
1	3	2	-
2	3	3	1
3	1	-	-
5	1	-	-

Table A.13 Distribution of shoveling (UI2L) grades in each genotype at rs7349332.

	Genotype	
Carabelli's cusp UM1	GA GA	GA G
0	9	7
3	1	2
4	4	-
5	7	-
6	1	-
7	4	1

Table A.14 Distribution of Carabelli's cusp (UM1) grades in each genotype at Chr11:69811473.

Carabelli's cusp UM1	Genotype			
	Ref Ref	Ref *	Ref A	* *
0	15	1	-	-
3	3	-	-	-
4	3	1	-	-
5	3	2	1	1
6	1	-	-	-
7	2	2	1	-

Table A.15 Distribution of Carabelli's cusp (UM1) grades in each genotype at Chr11:69813652.

Appendix B



BINGHAMTON UNIVERSITY INSTITUTIONAL REVIEW BOARD
PO BOX 6000
BINGHAMTON, NEW YORK 13902-6000
HSRRC@BINGHAMTON.EDU
CALL: (607) 777-3818 OR (607) 777-3918

IRB EXPEDITED CONDITIONAL APPROVAL

Dear Fatma Nur Erbil:

On 11/11/2020, the IRB reviewed the following submission:

Type of Review:	Expedited
Title of Study:	Identification of the genetic determinants of shovel-shaped incisors and Carabelli's cusp
Investigator:	Fatma Nur Erbil
IRB ID:	STUDY00002536
Funding:	Name: Binghamton University
Grant ID:	
Documents Reviewed:	• Genetics of tooth morphology informed consent, Category: Consent Form; • Covid-19 Screening Survey.docx, Category: Other; • covidrisksheet.docx, Category: Other; • Email to recruit participants, Category: Recruitment Materials; • Faculty advisor statement, Category: Other; • Flyer, Category: Recruitment Materials; • Genetics of tooth morphology protocol narrative, Category: IRB Protocol; • Instructions for mouthwash saliva collection.docx, Category: Other; • Instructions for taking dental photos.docx, Category: Other; • Qualtrics Links.docx, Category: Other; • Social Media advertisement, Category: Recruitment Materials; • SUBJECT INFORMATION FORM.docx, Category: Other;

The IRB has considered your submission for the project referenced above and has issued a conditional Expedited approval. Please note, you may only conduct this research in accordance with the IRB approved study procedures. The in-person research activity within this protocol is conditionally approved, with the strict understanding that in-person study procedures cannot commence until 1) the University suspension on in-person human subjects research (on or off campus) in which social distancing cannot be maintained has been lifted and/or 2) approval has been received via the Return to Research application process. This process was enacted on May 15, 2020 to enable research and creative activities to resume in line with standards set forth by New York state and the Centers for Disease Control and Prevention.

Expedited studies do not require continuing review. However, the IRB may still require continuing review on a study-specific basis if it is determined that continuing review would enhance the protection of human subjects involved in research. Justifications will be provided to the investigators at the time of approval if the IRB determines that continuing review is required.

Please note, you may only conduct this research in accordance with the IRB approved study procedures. Should you need to make a modification to your research, this must be submitted through PACS for IRB review prior to any changes being executed. Once approval has been granted your study can move forward with those changes. You can submit a Modification by navigating to the “Active” tab and clicking Create Modification/CR.

When your project has concluded, please complete the Protocol closure process found within the Create Modification/CR selection in PACS.

Additional guidance for researchers is provided in the IRB [Policy and Procedures](#) document, which can be found at the Division of Research home page under Research Compliance and Human Subjects. If you have any questions, please contact the Human Subjects Research Review Office at the phone number or e-mail address in the header.

Figure B.1 Approval letter from the Binghamton University IRB for this dissertation project (Total of 2 pages in pages 87 and 88).

Consent for Research Participation

Study Title: Identification of the genetic determinants of shovel-shaped incisors and Carabelli's cusp.
Sponsor: This research project is funded by Binghamton University's Transdisciplinary Area of Excellence Seed Grant (TAE) program.
Researcher(s): Fatma Nur Erbil (student principal investigator) and Dr Andrew Merriwether (faculty advisor), Department of Anthropology, Binghamton University.
Researcher Contact Info: Fatma Nur Erbil, ferbil1@binghamton.edu, [REDACTED]
Dr David Andrew Merriwether, andym@binghamton.edu, [REDACTED]

You are being invited to take part in a research study conducted by the researchers named above. Below is detailed information for you to consider when determining whether or not to participate. Carefully consider all of this information and ask any questions you may have about it before deciding whether to participate or not.

Key Information for You to Consider

- **Voluntary Consent:** You are being asked to volunteer for a research study. It is your choice whether to participate or not. There will be no penalty or loss of benefits to which you are otherwise entitled if you choose not to participate or to discontinue participation.
- **Purpose:** The purpose of this research is to identify the genes that cause shovel-shaped incisors and Carabelli's cusp, which have been extensively used in forensic and ethnic studies. You are being invited to participate because you are 18 and older and may or may not exhibit shovel-shaped incisors and/or Carabelli's cusp. Regardless of gender, adults 18 years old and older who possess shovel-shaped incisors and/or Carabelli's cups in their molars will be included in the positive control group in the study. Adult individuals who show interest in participating but do not possess the dental traits of interest in this study will be included in the negative control group in this research study. It is expected that approximately 60 individuals will be participating in this research.
- **Procedures and Activities:** If you agree to participate, you will be asked to:
 1. Fill out the subject information form and COVID-19 Screening Survey,
 2. Provide a saliva sample in-person,
 3. Provide dental photos in-person,
 4. Provide dental impressions in-person.

Please see the section "What happens if I agree to participate in this research?" below for more detailed information about the procedures and activities.
- **Duration:** Not including travel time, your time commitment will be approximately 5 minutes for completing the surveys, 5 minutes for saliva collection, 5 minutes for providing dental photos and 10 minutes for providing dental impressions. In total, the anticipated time commitment will be around 25 minutes.
- **Risks:** Some of the foreseeable risks or discomforts of your participation include: 1) risk of COVID-19 infection during oral-rinse saliva collection, dental photo and dental impression procedures, 2) discomfort during saliva collection, 3) discomfort and pressure during taking dental impressions, 4) risk of swallowing dental impression material, 5) minimal risk of loss of

confidentiality. Risks for COVID infection have been minimized in a number of ways. Please reference the Participant COVID Risk section for further details. The subjects may feel burning sensation in their mouth due to the normal alcohol content in the mouthwash. If the subjects feel uncomfortable during oral-rinse saliva collection, they can spit the mouthwash in the tube any time. In order to minimize discomfort during taking dental impressions, we will pick a dental tray size that matches the subject's size best. To mitigate the risk of swallowing the dental impression material, we will use a manual oral suction pump which will help protect your airway in case of an excessive drooling and/or gag reflex occurs. It is very important that you breathe through your nose rather than your mouth when the dental tray is placed in your mouth. We will use sterilized single-use suctioning tubes for each subject to minimize the risk of any infections. Confidentiality and privacy protections will be explained below.

- **Benefits:** There are no direct benefits to participants, however the findings of this research may help in understanding the genetic and developmental underpinnings of tooth crown morphology.
- **Alternatives:** Participation is voluntary and the only alternative is to not participate.

What happens if I agree to participate in this research?

If you agree to participate, you will be responsible for the following:

1. Fill out the subject information form and COVID-19 Screening Survey:

Once the informed consent is provided, the research team will assign you a randomly generated Subject ID# which will be provided to you via email. An appointment will be scheduled for you to provide a saliva sample, dental photos and dental impressions (see the steps #2, #3, and #4 below). Within 24 hours of your appointment an email will be sent to you that includes an anonymous, online subject information form and COVID-19 screening survey, and instructions for saliva collection. The Subject Information Form asks for basic information about you such as your age, ethnicity and any known allergies and dental pathologies. You will use your Subject ID# when completing the Qualtrics form/survey. If you report in the COVID screening survey that you have at least one of the symptoms specified, we will cancel your appointment and ask you to reschedule at a later date (14 days after showing no symptoms or testing negative for COVID-19).

2. Saliva sample collection:

To provide the in-person saliva sample you can either visit Dr Merriwether's Molecular Anthropology Lab (Science 3 Room 247C) on the Binghamton University campus, or the study team can visit you at your house or a location you identify. The study team is fully vaccinated and will need to enter your house or the location you specify to collect a saliva sample and to take dental photos and dental impressions. The study team will be wearing face masks, face shields and gloves at all times during the visit.

All COVID-19 safety precautions will be in place during your individual appointment. If the sample collection takes place in Dr Merriwether's Molecular Anthropology Lab, you must arrive wearing a face mask and will be provided gloves and a lab coat upon your arrival. If the sample collection takes place in your house or a location you identify, you must wear a face mask except when you are asked to take it off to give a saliva sample, dental photos and impressions. You are asked to abstain from eating and drinking at least 30 minutes prior to saliva sampling. The researcher will sterilize a saliva collection tube and disposable funnel and have an unopened bottle of travel-size mouthwash for you to use. You will be asked to remove your mask briefly to rinse your mouth vigorously with the mouthwash for 30 seconds and then spit into the funnel placed inside the collection tube. Once you have spit, you will put your mask back on, discard the funnel, and wipe the outside of the tube with 70% isopropyl alcohol to sterilize it. The anticipated time commitment for giving saliva sample is about 5 minutes.

3. Provide dental photos:

~~In order to~~ take photos of your teeth, the student PI will need to breach 6-foot social distancing. We will take at least 5 photos of the upper molars, 5 photos of the lower and 5 photos of upper incisors. We will not include any facial features in the photos. The student PI will examine the photos and score the teeth for dental nonmetric traits. The photos are for scoring purposes only. As soon as the scoring procedure is complete, we will delete the pictures permanently. The anticipated time commitment for taking pictures of your teeth is approximately 5 minutes.

4. Provide dental impressions:

The student PI will take impressions of your upper and lower teeth following common procedures used by dentists. All the tools including various sizes of dental trays will be cleaned and sterilized by UV irradiation. We will only use single-use disposable dental trays.

If dental impression procedure is taking place in our lab, we will have a dedicated space for it. The space will be cleaned and sterilized with 10% bleach solution, and the room will be UV irradiated between participants. If the procedure is taking place in the participant's house, we will ask you to show us a flat surface in your house (e.g., kitchen table, countertop, dining table, desk etc.) where we can prepare for taking dental impressions. We will clean and sanitize the surfaces with 70% isopropyl alcohol and water before starting.

Before loading the dental tray with the impression material, we will try a dental tray in your mouth ~~in order to~~ determine the tray size that fits inside of your mouth best. Dental bibs and clips will be provided to you to protect your clothes.

Once the tray size is determined, the student PI will load a dental tray and place it in your mouth. The impression material needs exactly one minute to set. During this step, while the tray is still in your mouth, we will use a manual hand operated oral suction pump to protect your airway. The tubing and collection cups are single-use and disposable. After one minute, the tray will be removed from your mouth. The same steps will be repeated for the opposite dentition. You will be provided with tissues, wet ~~wipes~~ and water to clean up. Taking dental

impressions for both upper and lower dentitions is expected to take 10 minutes in total. At that point the study is complete.

We will tell you about any new information that may affect your willingness to continue participation in this research.

What happens to the information collected for this research?

We will extract DNA from your saliva sample to look for any associations between your dental morphology and genetic variants. The findings will be submitted for publication and presented at conferences.

How will my privacy and data confidentiality be protected?

The student PI will be the only member of the research team authorized to review photos of participant dental traits. You will be assigned a Subject ID# to be used on all study surveys or forms. The research team will exclusively use the subject ID# on all research related materials. No names or other identifying information will be used in publications or for storing data.

Please note that email communication is neither private nor secure. Though we are taking precautions to protect your privacy, you should be aware that information sent through e-mail could be read by a third party.

We will take measures to protect the security of all your personal information. The access to the research materials will be limited to only the student PI, study team member Joel Mora and the PI's faculty advisor, Dr. Andrew Merriwether. Informed consent documents and the list of Subject ID#s and participant names will be stored separate from any raw data. Electronic records will be stored on encrypted, password protected hard drives.

Despite these measures to protect your privacy and confidentiality, we can never fully guarantee either. Individuals and organizations that conduct or monitor this research may be permitted access to inspect the research records, including accessing your private information. These individuals and organizations include approved study team members and the Binghamton University Institutional Review Board (IRB). This research will remain confidential unless we are required by New York State Law to report harm to yourself or your children.

What if I want to stop participating in this research?

Taking part in this research study is your decision. Your participation is voluntary. You do not have to take part in this study, but if you do, you can choose not to participate in any study activity or to completely withdraw at any point without penalty or loss of benefits to which you are otherwise entitled. Your decision whether or not to participate will not affect your relationship with the researcher(s) or Binghamton University.

What if I am injured because of participating in this research?

If you are injured as a result of participating in this study you and your insurance carrier assume full financial responsibility for any injury that you may suffer as a result of participation. No other form of compensation is being offered, however that does not mean you are giving up any of your legal rights.

Will I be paid for participating in this research?

All individuals who provide a saliva sample, dental photos and dental impressions for the study will receive a 10\$ Amazon gift card via their email. The full compensation of \$10 gift card will be given to the participant after all the study procedures are completed even if the participant decides to withdraw from the study later.

By accepting payment(s) for participating in this study, certain identifying information about you may be made available to professional auditors to satisfy federal and state reporting requirements, but confidentially will be preserved. Please note that if you earn \$600 or over in a calendar year as a research subject, these earnings will be reported to the Internal Revenue Service.

Consent to be contacted for participation in future research:

Researchers may wish to keep your contact information to invite you to be in future research projects that may be similar to or completely different from this research project. It is expected that in the future we will conduct a second part of this study involving collecting dental impressions.

- May we contact you again about potential participation in future studies? ☐ YES ☐ NO

Important Information about COVID-19 and Research Participation

At Binghamton University, the protection of research participants is of utmost importance, which is why additional information regarding COVID-19 is being provided to you. COVID-19 refers to A Coronavirus that is spread between people in our communities. We need to tell you about ways your study participation might change because of COVID-19 related risk.

If you are considering joining a study at this time or are currently enrolled in a study, it is important that you consider the following information to determine if study participation is right for you.

How is COVID-19 spread? COVID-19 is a respiratory virus spread by respiratory droplets, primarily from person-to-person. This can happen between people who are in close contact with one another (less than 6 feet). It is also possible that a person can get COVID-19 by touching a surface or object (such as a doorknob or counter surface) that has the virus on it, then touching their mouth, nose or eyes.

Can COVID-19 be prevented? Current ways to minimize the risk of exposure to COVID-19 include social distancing, wearing a face covering, hand washing, and use of hand sanitizers to decrease the potential for direct exposure to others who may have been exposed to COVID-19. It is important to understand that study participation may increase your exposure to others and your exposure to COVID-19. At this time there is no vaccination to prevent COVID-19 infection.

What are the risks of COVID-19? COVID-19 can affect anyone, and the disease can cause symptoms

that are varying and can range from none to leading to death. Some people may be more likely to have severe illness than others because they have characteristics or medical conditions that increase their risk. Examples include older aged individuals or those having certain underlying medical conditions. You should familiarize yourself with these conditions, and if you have any such condition, you should not participate in in-person research.

Mitigations to address COVID-19? There are several actions that Binghamton University researchers will take to try to minimize your risk related to COVID including the following:

- Eliminating/limiting in-person contact to the extent possible. If possible, research teams will limit the number of times you have to come to a research site.
- Selecting study sites that allow for better social distancing.
- Online screening of study subjects to evaluate health and potential risk for COVID-exposure before their study visit. If you have any symptoms of illness, you should discuss them with a healthcare provider and the study visit will be postponed until you are symptom free for 14 days.
- Daily screening of study team members to evaluate health and potential risk for COVID-exposure before their engagement. A daily log including contact information of entry and exit to the research site by study team members will be maintained. These logs will only be used in the event a study team member is found to test positive for COVID to allow for contact tracing.
- Study appointments will be scheduled to avoid overlap/contact with other study subjects and will be conducted one-on-one.
- You will be encouraged to text or call the research team when you arrive for an appointment so that a researcher may greet you and minimize time spent in waiting areas; a phone number to do this will be provided.
- Social distancing of 6 feet apart will be maintained during in-person research activities (i.e. interviews, computer activities, observations) except for taking dental photos and impressions.
- Facial coverings are required to be worn by both study team members and you during in-person interactions. You should put on a facial covering before arriving for your visit. If you arrive and are not already wearing a facial covering, one will be provided to you by the research team.
- Research procedure areas, data collection areas, and equipment will be carefully cleaned and disinfected prior to and following use by study team members and/or study subjects. This process includes a regular wipe down of shared research equipment and spaces (e.g. desk and table tops, chairs, keyboard/mouse, tablets, clipboards, pens, door handles) after each visit plus a wipe down of shared research equipment and spaces at the end of the day.
- If you wish to further wipe down an area you come in contact with, you will be provided with disinfectant wipes.
- If you use a shared piece of equipment (e.g., keyboard, mouse, touch screen), you will be offered hand sanitizer containing at least 60% alcohol or have access to a sink for hand-washing. Gloves will also be available and are an option to use during a study visit. Alternatively, the shared equipment may be shielded with a disposable covering.

The information related to risks of COVID-19 changes every day. This is being monitored so that the most current information is being provided to you through the research teams. If you have specific questions regarding your participation in research, please talk to the research team you are working with or contact IRB Associate Director Linda Reynolds at lreynold@binghamton.edu or 607-777-3918.

Who can answer my questions about this research?

If you have questions, concerns, or have experienced a research related injury, contact the research team at:

- Fatma Nur Erbil
- 607 759 2921 / ferbil1@binghamton.edu

Binghamton University's IRB is overseeing this research. The IRB is a group of people who perform independent reviews of research studies to ensure the rights and welfare of participants are protected. If you have questions about your rights or wish to speak with someone other than the research team, you may contact:

- The Binghamton University IRB
- hsrrc@binghamton.edu / (607) 777 - 3818

STATEMENT OF CONSENT

I have had the opportunity to read and consider the information outlined in this form and understand that the research team will collect a saliva sample and take photos and impressions of my teeth. I understand that the research team will be breaching 6-foot social distancing ~~in order to~~ take photos of my teeth and dental impressions. I have asked any questions necessary to ~~make a decision~~ about my participation. I understand that I can ask additional questions throughout my participation. I understand that by signing below, I volunteer to participate in this research. I understand that I am not waiving any legal rights. I have been provided with a copy of this consent form.

☐ I am at least 18 years ~~old~~ and I consent to participate in this study.

Signature of the Research Participant:

Participant Email*:

Participant Phone Number*:

*To be used by the research team, only, for communication of study procedures and compensation.

Figure B.2 Informed consent form used in this dissertation. (Total of 7 pages in pages 88 through 94).

SUBJECT INFORMATION FORM

1. SUBJECT ID NUMBER:
2. AGE:
3. GENDER:
4. ETHNICITY:
MATERNAL ETHNICITY:
PATERNAL ETNICITY:
5. ANY ALLERGIES:
6. ANY DISEASES/GENETIC DISORDERS (PLEASE CHECK ALL THAT MAY APPLY):
ENAMEL HYPOPLASIA:
CONGENITAL TOOTH LOSS:
SUPERNUMERARY TEETH:
OTHER:

Figure B.3 Subject information form used in this dissertation.

Appendix C



PARTIES:

1. **Cambridge University Press** [CompanyNumber] (Licensor); and
2. **Fatma Nur Erbil** (Licensee).

Thank you for your recent permission request. Some permission requests for use of material published by the Licensor, such as this one, are now being facilitated by PLSclear.

Set out in this licence cover sheet (the **Licence Cover Sheet**) are the principal terms under which Licensor has agreed to license certain Licensed Material (as defined below) to Licensee. The terms in this Licence Cover Sheet are subject to the attached General Terms and Conditions, which together with this Licence Cover Sheet constitute the licence agreement (the **Licence**) between Licensor and Licensee as regards the Licensed Material. The terms set out in this Licence Cover Sheet take precedence over any conflicting provision in the General Terms and Conditions.

Free Of Charge Licence Terms

Licence Date:
PLSclear Ref No: 66119

The Licensor

Company name: Cambridge University Press
Address: University Printing House
Shaftesbury Road
Cambridge
CB2 8BS
GB

The Licensee

Licensee Contact Name: Fatma Nur Erbil
Licensee Address: [REDACTED]
Vestal
13850
United States

Licensed Material

title: Human Tooth Crown and Root Morphology The Arizona
State University Dental Anthropology System
ISBN: 9781107480735
publisher: Cambridge University Press

Are you requesting permission to reuse the cover of the publication?	No
Figure number & title	Figure i Basic terms of orientation for upper and lower teeth
Page numbers	4
Are you requesting permission to reuse your own work?	No

For Use In Licensee's Publication(s)

usage type	Book, Journal, Magazine or Academic Paper-Thesis / Dissertation
Will your dissertation be placed in an online repository?	No
Author	Fatma Nur Erbil
Language	English
Title of dissertation/thesis	Genetic determinants of shovel-shaped incisors and Carabelli's cusp
University or institution	Binghamton University
Unlimited circulation?	No

Rights Granted

Exclusivity:	Non-Exclusive
Format:	Thesis / Dissertation
Language:	English
Territory:	USA
Duration:	Lifetime of Licensee's Edition
Maximum Circulation:	10

Figure C.1 Copyright permission document obtained through PLS Clear (<https://www.plsclear.com/>) for using the visual in Figure 2.2. (Total of two pages in pages 96 and 97)

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS
May 08, 2022

This Agreement between Ms. Nur Erbil ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5165610564053
License date	Oct 10, 2021
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Australian Dental Journal
Licensed Content Title	The dentition: the outcomes of morphogenesis leading to variations of tooth number, size and shape
Licensed Content Author	GC Townsend, TE Hughes, RN Smith, et al
Licensed Content Date	Mar 20, 2014
Licensed Content Volume	59
Licensed Content Issue	s1
Licensed Content Pages	12
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	Ph.D. student
Institution name	Binghamton University
Expected presentation date	May 2022
Portions	Fig. 1 Ms. Nur Erbil [REDACTED]
Requestor Location	VESTAL, NY 13850 United States Attn: Ms. Nur Erbil
Publisher Tax ID	EU826007151
Total	0.00 USD

Figure C.2 Copyright permission document obtained through RightsLink (<https://www.copyright.com/publishers/rightslink-scientific/>) for using the visual in Figure 2.3.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
May 08, 2022

This Agreement between Ms. Nur Erbil ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5165601475358
License date	Oct 10, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Genetics
Licensed Content Title	The cutting-edge of mammalian development; how the embryo makes teeth
Licensed Content Author	Abigail Tucker et al
Licensed Content Date	Jul 1, 2004
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
High-res required	no
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Ph.D. student
Institution name	Binghamton University
Expected presentation date	May 2022
Portions	Figure 2 Ms. Nur Erbil [REDACTED]
Requestor Location	VESTAL, NY 13850 United States Attn: Ms. Nur Erbil
Total	0.00 USD

Figure C.3 Copyright permission document obtained through RightsLink (<https://www.copyright.com/publishers/rightslink-scientific/>) for using the visual in Figure 2.4.

References

- Adaimy, L., Chouery, E., Mégarbané, H., Mroueh, S., Delague, V., Nicolas, E., Belguith, H., de Mazancourt, P., & Mégarbané, A. (2007). Mutation in WNT10A Is Associated with an Autosomal Recessive Ectodermal Dysplasia: The Odonto-onycho-dermal Dysplasia. *The American Journal of Human Genetics*, 81(4), 821–828.
- Ahtiainen, L., Uski, I., Thesleff, I., & Mikkola, M. L. (2016). Early epithelial signaling center governs tooth budding morphogenesis. *Journal of Cell Biology*, 214(6), 753–767.
- Andl, T., Reddy, S. T., Gaddapara, T., & Millar, S. E. (2002). WNT signals are required for the initiation of hair follicle development. *Developmental Cell*, 2(5), 643–653.
- Antoine, D., & Hillson, S. (2015). Enamel structure and properties. *A companion to dental anthropology*, 223–243.
- Arikan, V., Cumaogullari, O., Ozgul, B.M., & Oz, F.T. (2018). Investigation of SOSTDC1 gene in non- syndromic patients with supernumerary teeth. *Medicina Oral, Patología Oral y Cirugía Bucal*, 23(5), e531.
- Bailey, S. E., Skinner, M. M., & Hublin, J.J. (2011). What lies beneath? An evaluation of lower molar trigonid crest patterns based on both dentine and enamel expression. *American Journal of Physical Anthropology*, 145(4), 505–518.
- Baird, S. J. (2015). Exploring linkage disequilibrium. *Molecular Ecology Resources*, 15(5), 1017-1019.
- Bei, M., Kratochwil, K., & Maas, R. L. (2000). BMP4 rescues a non-cell-autonomous function of Msx1 in tooth development. *Development*, 127(21), 4711–4718.
- Bei, M., & Maas, R. (1998). FGFs and BMP4 induce both Msx1-independent and Msx1-dependent signaling pathways in early tooth development. *Development*, 125(21), 4325–4333.
- Bethard, J. D., & DiGangi, E. A. (2020). Letter to the editor—Moving beyond a Lost Cause: Forensic anthropology and ancestry estimates in the United States. *Journal of Forensic Sciences*, 65(5), 1791-1792.
- Bhavyaa, R., Sujitha, P., Muthu, M. S., & Kirthiga, M. (2020). Prevalence of Cusp of Carabelli and its caries susceptibility – an ambidirectional cohort study. *Australian Dental Journal*, 65(4), 294–301.
- Biben, C., Wang, C.C., & Harvey, R. P. (2002). NK-2 class homeobox genes and pharyngeal/oral patterning: Nkx2-3 is required for salivary gland and tooth morphogenesis. *International Journal of Developmental Biology*, 46(4), 415–422.

- Boverhof, D. R., Chamberlain, M. P., Elcombe, C. R., Gonzalez, F. J., Heflich, R. H., Hernandez, L. G., Jacobs, A.C., Jacobson-Kram, D., Luijten, M., Maggi, A., Manjanatha, M.G., Benthem, J., and Gollapudi, B. B. (2011). Transgenic animal models in toxicology: historical perspectives and future outlook. *Toxicological sciences*, 121(2), 207-233.
- Brook, A. H., Jernvall, J., Smith, R. N., Hughes, T. E., & Townsend, G. C. (2014). The dentition: the outcomes of morphogenesis leading to variations of tooth number, size and shape. *Australian Dental Journal*, 59, 131–142.
- Brook, A. H., O'Donnell, M. B., Hone, A., Hart, E., Hughes, T. E., Smith, R. N., & Townsend, G. C. (2014). General and craniofacial development are complex adaptive processes influenced by diversity. *Australian Dental Journal*, 59, 13–22.
- Bush WS, Moore JH (2012). Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*, 8(12), e1002822.
- Cai, J., Mutoh, N., Shin, J.O., Tani-Ishii, N., Ohshima, H., Cho, S.W., & Jung, H.S. (2011). Wnt5a plays a crucial role in determining tooth size during murine tooth development. *Cell and Tissue Research*, 345(3), 367–377.
- Callea, M., Nieminen, P., Willoughby, C., Clarich, G., Yavuz, I., Vinciguerra, A., Di Stazio, M., Giglio, S., Sani, I., Maglione, M., & Pensiero, S. (2014). A novel INDEL mutation in the EDA gene resulting in a distinct X-linked hypohidrotic ectodermal dysplasia phenotype in an Italian family. *Journal of the European Academy of Dermatology and Venereology*, 30(2), 341–343.
- Cao, H., Wang, J., Li, X., Florez, S., Huang, Z., Venugopalan, S. R., Elangovan, S., Skobe, Z., Margolis, H. C., Martin, J. F., & Amendt, B. A. (2010). MicroRNAs play a critical role in tooth development. *Journal of dental research*, 89(8), 779–784.
- Carayon, D., Adhikari, K., Monsarrat, P., Dumoncel, J., Braga, J., Duployer, B., Delgado, M., Fuentes-Guajardo, M., de Beer, F., Hoffman, J. W., Oettlé, A. C., Donat, R., Pan, L., Ruiz-Linares, A., Tenailleau, C., Vaysse, F., Esclassan, R., & Zanolli, C. (2019). A geometric morphometric approach to the study of variation of shovel-shaped incisors. *American Journal of Physical Anthropology*, 168(1), 229–241.
- Catón, J., & Tucker, A. S. (2009). Current knowledge of tooth development: patterning and mineralization of the murine dentition. *Journal of Anatomy*, 214(4), 502–515.
- Charles, C., Lazzari, V., Tafforeau, P., Schimmang, T., Tekin, M., Klein, O., & Viriot, L. (2009). Modulation of Fgf3 dosage in mouse and men mirrors evolution of mammalian dentition. *Proceedings of the National Academy of Sciences*, 106(52), 22364–22368.
- Chhabra, N., Goswami, M., & Chhabra, A. (2014). Genetic basis of dental agenesis-molecular genetics patterning clinical dentistry. *Medicina oral, patologia oral y*

cirugia bucal, 19(2), e112.

- Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6), 1767–1771.
- Corder, G. W., & Foreman, D. I. (2011). Nonparametric statistics for non-Statisticians: A Step-by-Step Approach (Second edition). *John Wiley & Sons*.
- Delgado, M., Ramírez, L. M., Adhikari, K., Fuentes-Guajardo, M., Zanolli, C., Gonzalez-José, R., Canizales, S., Bortolini, M.C., Poletti, G., Gallo, C., Rothhammer, F., Bedoya, G., & Ruiz-Linares, A. (2019). Variation in dental morphology and inference of continental ancestry in admixed Latin Americans. *American Journal of Physical Anthropology*, 168(3), 438–447.
- Dempsey, P. J., Townsend, G. C., Martin, N. G., & Neale, M. C. (1995). Genetic covariance structure of incisor crown size in twins. *Journal of Dental Research*, 74(7), 1389–1398.
- DePristo, M. A., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43, 491–498.
- DiGangi, E. A., & Bethard, J. D. (2021). Uncloaking a lost cause: Decolonizing ancestry estimation in the United States. *American Journal of Physical Anthropology*, 175, 422–436.
- Du, W., Du, W., & Yu, H. (2018). The role of fibroblast growth factors in tooth development and incisor renewal. *Stem Cells International*, 2018.
- Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., ... & Mountain, J. (2010). Web-based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits. *PLoS genetics*, 6(6), e1000993.
- Fang, H., Wu, Y., Narzisi, G., O’Rawe, J. A., Barrón, L. T. J., Rosenbaum, J., Ronemus, M., Iossifov, I., Schatz, M.C. and Lyon, G. J. (2014). Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome medicine*, 6(10), 1–17.
- Grine, F. E. (2005). Enamel thickness of deciduous and permanent molars in modern *Homo sapiens*. *American Journal of Physical Anthropology*, 126, 14–31.
- Gross, J. M., & Edgar, H. J. (2019). Informativeness of dental morphology in ancestry estimation in African Americans. *American Journal of Physical Anthropology*, 168(3), 521–529.

- Guatelli-Steinberg, D. (2016). Dental Stress Indicators from Micro- to Macroscopic In J. D. Irish & G. R. Scott (Eds.), *A companion to Dental Anthropology* (pp. 450-464). Wiley Blackwell.
- Guatelli-Steinberg, D., Hunter, J. P., Durner, R. M., Moormann, S., Weston, T. C., & Betsinger, T. K. (2013). Teeth, morphogenesis, and levels of variation in the human Carabelli trait. In G. R. Scott & J. D. Irish (Eds.), *Anthropological Perspectives on Tooth Morphology: Genetics, Evolution, Variation* (pp. 69–91). Cambridge University Press.
- Han, D., Gong, Y., Wu, H., Zhang, X., Yan, M., Wang, X., Qu, H., Feng, H., & Song, S. (2008). Novel EDA mutation resulting in X-linked non-syndromic hypodontia and the pattern of EDA-associated isolated tooth agenesis. *European Journal of Medical Genetics*, 51(6), 536-546.
- Han, X., Yoshizaki, K., Miyazaki, K., Arai, C., Funada, K., Yuta, T., Tian, T., Chiba, Y., Saito, K., Iwamoto, T., & Yamada, A. (2018). The transcription factor NKX2-3 mediates p21 expression and ectodysplasin-A signaling in the enamel knot for cusp formation in tooth development. *Journal of Biological Chemistry*, 293(38), 14572–14584.
- Haneji, K., Hanihara, T., Sunakawa, H., Toma, T., & Ishida, H. (2007). Non-metric dental variation of Sakishima Islanders, Okinawa, Japan: A comparative study among Sakishima and neighboring populations. *Anthropological Science*, 115(1), 35–45.
- Hanihara, T. (2008). Morphological variation of major human populations based on nonmetric dental traits. *American Journal of Physical Anthropology*, 136(2), 169-182.
- Hanihara, T. and Ishida, H. (2005). Metric dental variation of major human populations. *American Journal of Physical Anthropology*, 128, 287–298.
- Harvey, R. P. (1996). NK-2 homeobox genes and heart development. *Developmental Biology*, 178(2), 203–216.
- Hasegawa, Y., Terada, K., Kageyama, I., Tsukada, S., Uzuka, S., Nakahara, R., & Nakahara, S. (2009). Influence of shovel-shaped incisors on the dental arch crowding in Mongolian females. *Okajimas Folia Anatomica Japonica*, 86(2), 67–72.
- Hillson, S. (1996). *Dental anthropology*. Cambridge University Press.
- Hlusko, L. J., Carlson, J. P., Chaplin, G., Elias, S. A., Hoffeecker, J. F., Huffman, M., Jablonski, N. G., Monson, T. A., O'Rourke, D. H., Pilloud, M. A., & Scott, G. R. (2018). Environmental selection during the last ice age on the mother-to-infant transmission of vitamin D and fatty acids through breast milk. *Proceedings of the National Academy of Sciences*, 115(19), 4426–4432.

- Hrdlička, A. (1920). Shovel-shaped teeth. *American Journal of Physical Anthropology*, 3(4), 429–465.
- Hu, G., Vastardis, H., Bendall, A. J., Wang, Z., Logan, M., Zhang, H., Nelson, C., Stein, S., Greenfield, N., Seidman, C. E., Seidman, J. G., & Abate-Shen, C. (1998). Haploinsufficiency of MSX1: a Mechanism for Selective Tooth Agenesis. *Molecular and Cellular Biology*, 18(10), 6044–6051.
- Hughes, T. E., & Townsend, G. C. (2013). Twin and family studies of human dental crown morphology: genetic, epigenetic, and environmental determinants of the modern human dentition. In G. R. Scott & J. D. Irish (Eds.), *Anthropological Perspectives on Tooth Morphology: Genetics, Evolution, Variation* (pp. 31–68). Cambridge University Press.
- Hunter, J. P., Guatelli-Steinberg, D., Weston, T. C., Durner, R., & Betsinger, T. K. (2010). Model of tooth morphogenesis predicts Carabelli cusp expression, size, and symmetry in humans. *PloS One*, 5(7), e11844.
- “Incisal.” *Merriam-Webster.com Dictionary*, Merriam-Webster, <https://www.merriam-webster.com/dictionary/incisal>. Accessed 8 May. 2022.
- Irish, J. D. (1997). Characteristic high- and low-frequency dental traits in sub-Saharan African populations. *American Journal of Physical Anthropology*, 102(4), 455–467.
- Irish, J. D. (1998). Ancestral dental traits in recent Sub-Saharan Africans and the origins of modern humans. *Journal of Human Evolution*, 34(1), 81–98.
- Irish, J. D., Morez, A., Flink, L. G., Phillips, E. L. W., & Scott, G. R. (2020). Do dental nonmetric traits actually work as proxies for neutral genomic data? Some answers from continental- and global-level analyses. *American Journal of Physical Anthropology*, 172(3), 347–375.
- Itoh, N., & Ornitz, D. M. (2008). Functional evolutionary history of the mouse Fgf gene family. *Developmental Dynamics*, 237(1), 18–27.
- Jernvall, J., & Jung, H. S. (2000). Genotype, phenotype, and developmental biology of molar tooth characters. *American Journal of Physical Anthropology*, 113(S31), 171–190.
- Jernvall, J., Kettunen, P., Karavanova, I., Martin, L. B., & Thesleff, I. (1994). Evidence for the role of the enamel knot as a control center in mammalian tooth cusp formation: non-dividing cells express growth stimulating Fgf-4 gene. *International Journal of Developmental Biology*, 38(3), 463–469.
- Jernvall, J., & Thesleff, I. (2000). Reiterative signaling and patterning during mammalian tooth morphogenesis. *Mechanisms of Development*, 92(1), 19–29.

- Jia, S., Zhou, J., Gao, Y., Baek, J.A., Martin, J. F., Lan, Y., & Jiang, R. (2013). Roles of Bmp4 during tooth morphogenesis and sequential tooth formation. *Development*, 140(2), 423–432.
- Jo, B. S., & Choi, S. S. (2015). Introns: the functional benefits of introns in genomes. *Genomics & Informatics*, 13(4), 112–118.
- Jussila, M., & Thesleff, I. (2012). Signaling networks regulating tooth organogenesis and regeneration, and the specification of dental mesenchymal and epithelial cell lineages. *Cold Spring Harbor Perspectives in Biology*, 4(4), a008425.
- Kamberov, Y. G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., Yang, Y., Li, S., Tang, K., Chen, H., & Powell, A. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell*, 152(4), 691–702.
- Karg, H. A., Burger, E. H., Lyaruu, D. M., Wöltgens, J. H. M., & Bronckers, A. L. J. J. (1997). Gene expression and immunolocalisation of amelogenins in developing embryonic and neonatal hamster teeth. *Cell and Tissue Research*, 288(3), 545–555.
- Kassai, Y., Munne, P., Hotta, Y., Penttilä, E., Kavanagh, K., Ohbayashi, N., Takada, S., Thesleff, I., Jernvall, J., & Itoh, N. (2005). Regulation of mammalian tooth cusp patterning by ectodin. *Science*, 309(5743), 2067–2070.
- Keränen, S. V. E., Åberg, T., Kettunen, P., Thesleff, I., & Jernvall, J. (1998). Association of developmental regulatory genes with the development of different molar tooth shapes in two species of rodents. *Development Genes and Evolution*, 208(9), 477–486.
- Kimura, R., Watanabe, C., Kawaguchi, A., Kim, Y. I., Park, S. B., Maki, K., Ishida, H., & Yamaguchi, T. (2015). Common polymorphisms in WNT10A affect tooth morphology as well as hair shape. *Human Molecular Genetics*, 24(9), 2673–2680.
- Kimura, R., Yamaguchi, T., Takeda, M., Kondo, O., Toma, T., Haneji, K., Hanihara, T., Matsukusa, H., Kawamura, S., Maki, K., Osawa, M., Ishida, H., & Oota, H. (2009). A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *American Journal of Human Genetics*, 85(4), 528–535.
- Klein, O. D., Oberoi, S., Huysseune, A., Hovorakova, M., Peterka, M., & Peterkova, R. (2013). Developmental disorders of the dentition: an update. *American Journal of Medical Genetics*, 163(4), 318–332.
- Köbel, M., Weidensdorfer, D., Reinke, C., Lederer, M., Schmitt, W. D., Zeng, K., Thomssen, C., Hauptmann, S., & Hüttelmaier, S. (2007). Expression of the RNA-binding protein IMP1 correlates with poor prognosis in ovarian carcinoma. *Oncogene*, 26(54), 7584–7589.
- Korbie, D. J., & Mattick, J. S. (2008). Touchdown PCR for increased specificity and

- sensitivity in PCR amplification. *Nature protocols*, 3(9), 1452-1456.
- Koussoulakou, D. S., Margaritis, L. H., & Koussoulakos, S. L. (2009). A curriculum vitae of teeth: evolution, generation, regeneration. *International Journal of Biological Sciences*, 5(3), 226-243.
- Kratochwil, K., Galceran, J., Tontsch, S., Roth, W., & Grosschedl, R. (2002). FGF4, a direct target of LEF1 and Wnt signaling, can rescue the arrest of tooth organogenesis in *Lef1*^{-/-} mice. *Genes & Development*, 16(24), 3173–3185.
- Lammi, L., Arte, S., Somer, M., Järvinen, H., Lahermo, P., Thesleff, I., Pirinen, S., & Nieminen, P. (2004). Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *The American Journal of Human Genetics*, 74(5), 1043–1050.
- Laurikkala, J., Kassai, Y., Pakkasjärvi, L., Thesleff, I., & Itoh, N. (2003). Identification of a secreted BMP antagonist, ectodin, integrating BMP, FGF, and SHH signals from the tooth enamel knot. *Developmental Biology*, 264(1), 91–105.
- Lee, W.C., Yamaguchi, T., Watanabe, C., Kawaguchi, A., Takeda, M., Kim, Y.I., Haga, S., Tomoyasu, Y., Ishida, H., Maki, K., Park, S.B., & Kimura, R. (2012). Association of common PAX9 variants with permanent tooth size variation in non-syndromic East Asian populations. *Journal of Human Genetics*, 57(10), 654–659.
- Legge, S. S., & Hardin, A. M. (2015). The Pulp Cavity and Its Contents. In J. D. Irish & G. R. Scott (Eds.), *A companion to Dental Anthropology* (pp. 191-203). Wiley Blackwell.
- Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, 30(20), 2843-2851.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- Li, P., Liu, W., Xu, Q., & Wang, C. (2017). Clinical significance and biological role of Wnt10a in ovarian cancer. *Oncology Letters*, 14(6), 6611-6617.
- Lin, Y., Zheng, L., Fan, L., Kuang, W., Guo, R., Lin, J., Wu, J., & Tan, J. (2018). The epigenetic regulation in tooth development and regeneration. *Current Stem Cell Research & Therapy*, 13(1), 4–15.
- Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing

- reads. *BMC Research Notes*, 5(1), 337.
- Loh, P. R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., & Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics*, 193(4), 1233-1254.
- Lovén, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., Levens, D.L., Lee, T.I. and Young, R. A. (2012). Revisiting global gene expression analysis. *Cell*, 151(3), 476-482.
- McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., & Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, 26(16), 2069-2070.
- Machiela, M. J., & Chanock, S. J. (2015). LDlink: a web-based application for exploring population- specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, 31(21), 3555–3557.
- Malik, Z., Alexiou, M., Hallgrimsson, B., Economides, A. N., Luder, H. U., & Graf, D. (2018). Bone morphogenetic protein 2 coordinates early tooth mineralization. *Journal of Dental Research*, 97(7), 835–843.
- Mandikos, M. N. (1998). Polyvinyl siloxane impression materials: an update on clinical use. *Australian Dental Journal*, 43(6), 428–434.
- Marado, L. M., & Campanacho, V. (2013). Carabelli's trait: Definition and review of a commonly used dental non-metric variable. *Cadernos do GEEvH*, 2(1), 24-39.
- McEvoy, B. P., Powell, J. E., Goddard, M. E., & Visscher, P. M. (2011). Human population dispersal “Out of Africa” estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Research*, 21(6), 821–829.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297– 1303.
- Mikkola, M. L. (2009). Molecular aspects of hypohidrotic ectodermal dysplasia. *American Journal of Medical Genetics*, 149(9), 2031-2036.
- Mikkola, M. L., & Thesleff, I. (2003). Ectodysplasin signaling in development. *Cytokine & Growth Factor Reviews*, 14(3), 211–224.
- Miletich, I., & Sharpe, P. T. (2003). Normal and abnormal dental development. *Human molecular genetics*, R69–R73.
- Moormann, S., Guatelli-Steinberg, D., & Hunter, J. (2013). Metamerism, morphogenesis, and the expression of Carabelli and other dental traits in humans. *American*

Journal of Physical Anthropology, 150(3), 400–408.

- Mostowska, A., Biedziak, B., & Trzeciak, W. H. (2006). A novel mutation in PAX9 causes familial form of molar oligodontia. *European Journal of Human Genetics*, 14(2), 173–179.
- Mueller, J. C. (2004). Linkage disequilibrium for different scales and applications. *Briefings in bioinformatics*, 5(4), 355–364.
- Munne, P. M., Tummers, M., Järvinen, E., Thesleff, I., & Jernvall, J. (2009). Tinkering with the inductive mesenchyme: Sostdc1 uncovers the role of dental mesenchyme in limiting tooth induction. *Development*, 136(3), 393–402.
- Murashima-Suginami, A., Kiso, H., Tokita, Y., Mihara, E., Nambu, Y., Uozumi, R., Tabata, Y., Bessho, K., Takagi, J., Sugai, M., & Takahashi, K. (2021). Anti-USAG-1 therapy for tooth regeneration through enhanced BMP signaling. *Science Advances*, 7(7), eabf1798.
- Murashima-Suginami, A., Takahashi, K., Sakata, T., Tsukamoto, H., Sugai, M., Yanagita, M., Shimizu, A., Sakurai, T., Slavkin, H. C., & Bessho, K. (2008). Enhanced BMP signaling results in supernumerary tooth formation in USAG-1 deficient mouse. *Biochemical and Biophysical Research Communications*, 369(4), 1012–1016.
- Nieminen, P., Arte, S., Tanner, D., Paulin, L., Alaluusua, S., Thesleff, I., & Pirinen, S. (2001). Identification of a nonsense mutation in the PAX9 gene in molar oligodontia. *European Journal of Human Genetics*, 9(10), 743–746.
- Nusse, R., & Varmus, H. E. (1982). Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell*, 31(1), 99–109.
- Nüsslein-Volhard, C., & Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287(5785), 795–801.
- Oommen, S., Otsuka-Tanaka, Y., Imam, N., Kawasaki, M., Kawasaki, K., Jalani-Ghazani, F., Anderegg, A., Awatramani, R., Hindges, R., Sharpe, P. T., & Ohazama, A. (2012). Distinct roles of microRNAs in epithelium and mesenchyme during tooth development. *Developmental Dynamics*, 241(9), 1465–1472.
- Parenteau, J., & Abou Elela, S. (2019). Introns: good day junk is bad day treasure. *Trends in Genetics*, 35(12), 923–934.
- Park, J. H., Yamaguchi, T., Watanabe, C., Kawaguchi, A., Haneji, K., Takeda, M., Kim, Y. I., Tomoyasu, Y., Watanabe, M., Oota, H., Hanihara, T., Ishida, H., Maki, K., Park, S.-B., & Kimura, R. (2012). Effects of an Asian-specific nonsynonymous EDAR variant on multiple dental traits. *Journal of Human Genetics*, 57(8), 508–514.

- Paul, K. S., Astorino, C. M., & Bailey, S. E. (2016). The Patterning Cascade Model and Carabelli's trait expression in metameres of the mixed human dentition: Exploring a morphogenetic model. *American Journal of Physical Anthropology*, 162(1), 3–18.
- Peterkova, R., Hovorakova, M., Peterka, M., & Lesot, H. (2014). Three-dimensional analysis of the early development of the dentition. *Australian Dental Journal*, 59, 55–80.
- Pillas, D., Hoggart, C. J., Evans, D. M., O'Reilly, P. F., Sipilä, K., Lähdesmäki, R., Millwood, I. Y., Kaakinen, M., Netuveli, G., Blane, D., Charoen, P., Sovio, U., Pouta, A., Freimer, N., Hartikainen, A.-L., Laitinen, J., Vaara, S., Glaser, B., Crawford, P., ... Jarvelin, M.-R. (2010). Genome-wide Association Study Reveals Multiple Loci Associated with Primary Tooth Development During Infancy. *PLoS Genetics*, 6(2), e1000856.
- Plikus, M. V., Zeichner-David, M., Mayer, J.-A., Reyna, J., Bringas, P., Thewissen, J. G. M., Snead, M. L., Chai, Y., & Chuong, C.-M. (2005). Morphoregulation of teeth: modulating the number, size, shape and differentiation by tuning Bmp activity. *Evolution & Development*, 7(5), 440–457.
- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 1-22.
- Qanbari, S. (2020). On the extent of linkage disequilibrium in the genome of farm animals. *Frontiers in Genetics*, 10, 1304.
- Rathmann, H., & Reyes-Centeno, H. (2020). Testing the utility of dental morphological trait combinations for inferring human neutral genetic variation. *Proceedings of the National Academy of Sciences*, 117(20), 10769–10777.
- Rathmann, H., Reyes-Centeno, H., Ghirotto, S., Creanza, N., Hanihara, T., & Harvati, K. (2017). Reconstructing human population history from dental phenotypes. *Scientific Reports*, 7(1), 1–9.
- Rijsewijk, F., Schuermann, M., Wagenaar, E., Parren, P., Weigel, D., & Nusse, R. (1987). The *Drosophila* homology of the mouse mammary oncogene *int-1* is identical to the segment polarity gene *wingless*. *Cell*, 50(4), 649–657.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26.
- Rostampour, N., Appelt, C. M., Abid, A., & Boughner, J. C. (2019). Expression of new genes in vertebrate tooth development and p63 signaling. *Developmental*

Dynamics, 248(8), 744–755.

- Salazar-Ciudad, I., & Jernvall, J. (2002). A gene network model accounting for development and evolution of mammalian teeth. *Proceedings of the National Academy of Sciences*, 99(12), 8116–8120.
- Salazar-Ciudad, I., & Jernvall, J. (2004). How different types of pattern formation mechanisms affect the evolution of form and development. *Evolution & Development*, 6(1), 6–16.
- Scheid, R. C., & Weiss, G. (2011). *Woelfel's Dental Anatomy* (8th edition). Lippincott Williams & Wilkins.
- Scott, G. R. (1980). Population variation of Carabelli's trait. *Human biology*, 63–78.
- Scott, G. R., & Irish, J. D. (2017). *Human tooth crown and root morphology: The Arizona State University Dental Anthropology System*. Cambridge University Press.
- Scott, G. R., Schmitz, K., Heim, K. N., Paul, K. S., Schomberg, R., & Pilloud, M. A. (2018). Sinodonty, Sundadonty, and the Beringian Standstill model: Issues of timing and migrations into the New World. *Quaternary International*, 466, 233–246.
- Scott, G. R., Schmitz, K., Heim, K. N., Paul, K. S., Schomberg, R., & Pilloud, M. A. (2018a). Sinodonty, Sundadonty, and the Beringian Standstill model: Issues of timing and migrations into the New World. *Quaternary International*, 466, 233–246.
- Scott, G. R., Turner II, C. G., Townsend, G. C., & Martín-Torres, M. (2018b). *The Anthropology of Modern Human Teeth: Dental Morphology and its Variation in Recent and Fossil Homo sapiens* (2nd ed.). Cambridge University Press.
- Shabalina, S. A., Ogurtsov, A. Y., Spiridonov, A. N., Novichkov, P. S., Spiridonov, N. A., & Koonin, E. V. (2010). Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. *Molecular biology and evolution*, 27(8), 1745–1749.
- Shaffer, C. (2021). Gene Knockout Method. AZoLifeSciences. Retrieved on May 08, 2022 from <https://www.azolifesciences.com/article/Gene-Knockout-Method.aspx>.
- Skinner, M. M., Wood, B. A., Boesch, C., Olejniczak, A. J., Rosas, A., Smith, T. M., & Hublin, J.-J. (2008). Dental trait expression at the enamel-dentine junction of lower molars in extant and fossil hominoids. *Journal of Human Evolution*, 54(2), 173–186.
- Suryadeva, S., & Khan, M. B. (2015). Role of homeobox genes in tooth morphogenesis: a review. *Journal of Clinical and Diagnostic Research*: 9(2), ZE09–ZE12.

- Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Review Genetics* 3(5): 391-397.
- Tamura, M., & Nemoto, E. (2016). Role of the Wnt signaling molecules in the tooth. *Japanese Dental Science Review*, 52(4), 75–83.
- Tan, J., Peng, Q., Li, J., Guan, Y., Zhang, L., Jiao, Y., Yang, Y., Wang, S., & Jin, L. (2014). Characteristics of dental morphology in the Xinjiang Uyghurs and correlation with the EDARV370A variant. *Science China Life Sciences*, 57(5), 510–518.
- Tang, N., Cabec, A., & Antoine, D. (2016). Dentine and Cementum Structure and Properties. In J. D. Irish & G. R. Scott (Eds.), *A companion to Dental Anthropology* (pp. 204– 222). Wiley Blackwell.
- Tekin, M., Hişmi, B. Ö., Fitoz, S., Özdağ, H., Cengiz, F. B., Sırmacı, A., Aslan, İ., İnceoğlu, B., Yüksel-Konuk, E. B., Yılmaz, S. T., Yasun, Ö., & Akar, N. (2007). Homozygous Mutations in Fibroblast Growth Factor 3 Are Associated with a New Form of Syndromic Deafness Characterized by Inner Ear Agenesis, Microtia, and Microdontia. *The American Journal of Human Genetics*, 80(2), 338–344.
- Thesleff, I. (2006). The genetic basis of tooth development and dental defects. *American Journal of Medical Genetics*, 140(23), 2530–2535.
- Thesleff, I. (2014). Current understanding of the process of tooth formation: transfer from the laboratory to the clinic. *Australian Dental Journal*, 59, 48–54.
- Thomas, G., Jacobs, K. B., Kraft, P., Yeager, M., Wacholder, S., Cox, D. G., Hankinson, S. E., Hutchinson, A., Wang, Z., Yu, K., Chatterjee, N., Garcia-Closas, M., Gonzalez-Bosquet, J., Prokunina-Olsson, L., Orr, N., Willett, W. C., Colditz, G. A., Ziegler, R. G., Berg, C. D., Buys S.S., McCarty C.A., Feigelson H.S., Calle E.E., Thun M.J., Diver R., Prentice R., Jackson R., Kooperberg C., Chlebowski R., Lissowska J., Peplonska B., Brinton L.A., Sigurdson A., Doody M., Bhatti P., Alexander B.H., Buring J., Lee I.M., Vatten L.J., Hveem K., Kumle M., Hayes R.B., Tucker M., Gerhard D.S., Fraumeni J.F. Jr, Hoover R.N., Chanock S.J., Hunter, D. J. (2009). A multi- stage genome-wide association in breast cancer identifies two novel risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nature Genetics*, 41(5), 579–584.
- Townsend, G. C., Richards, L., Hughes, T., Pinkerton, S., & Schwerdt, W. (2005). Epigenetic influences may explain dental differences in monozygotic twin pairs. *Australian Dental Journal*, 50(2), 95– 100.
- Townsend, G., Hughes, T., Luciano, M., Bockmann, M., & Brook, A. (2009). Genetic and environmental influences on human dental variation: a critical evaluation of studies involving twins. *Archives of Oral Biology*, 54, S45–S51.

- Tsai, P. L., Hsu, J. W., Lin, L. M., & Liu, K. M. (1996). Logistic analysis of the effects of shovel trait on Carabelli's trait in a Mongoloid population. *American Journal of Physical Anthropology*, 100(4), 523–530.
- Tucker, A. S., Headon, D. J., Courtney, J. M., Overbeek, P., & Sharpe, P. T. (2004). The activation level of the TNF family receptor, Edar, determines cusp number and tooth number during tooth development. *Developmental Biology*, 268(1), 185–194.
- Tucker, A. S., Headon, D. J., Schneider, P., Ferguson, B. M., Overbeek, P., Tschopp, J., & Sharpe, P. T. (2000). Edar/Eda interactions regulate enamel knot formation in tooth morphogenesis. *Development*, 127(21), 4691–4700.
- Tucker, A., & Sharpe, P. (2004). The cutting-edge of mammalian development; how the embryo makes teeth. *Nature Reviews Genetics*, 5(7), 499–508.
- Turner, C. G., Nichol, C. R., & Scott, G. R. (1991). Scoring Procedures for Key Morphological Traits of the Permanent Dentition: The Arizona State University dental anthropology system. *Advances in Dental Anthropology*, 13-31.
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 1-21.
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics*, 43(1), 11-10.
- Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the Cloud: using Docker, GATK, and WDL in Terra (1st edition)*. O'Reilly Media.
- Vohr, S. H., & Green, R. E. (2013). A mouse following in the footsteps of human prehistory. *Cell*, 152(4), 667–668.
- Wang, J., & Shackleford, G. M. (1996). Murine Wnt10a and Wnt10b: Cloning and expression in developing limbs, face and skin of embryos and in adults. *Oncogene*, 13(7), 1537–1544.
- Wang, R. N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S., Idowu, O., Li, M., Shen, C., Hu, A., Haydon, R. C., Kang, R., Mok, J., Lee, M. J., Luu, H. L., & Shi, L.L. (2014). Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes & Diseases*, 1(1), 87–105.
- Wang, X.P., & Fan, J. (2011). Molecular genetics of supernumerary tooth formation. *Genesis*, 49(4), 261–277.

- White, T. D., & Folkens, P. A. (2005). *The human bone manual*. Elsevier.
- Yamashiro, T., Zheng, L., Shitaku, Y., Saito, M., Tsubakimoto, T., Takada, K., Takano-Yamamoto, T., & Thesleff, I. (2007). Wnt10a regulates dentin sialophosphoprotein mRNA expression and possibly links odontoblast differentiation and tooth morphogenesis. *Differentiation*, 75(5), 452–462.
- Yang, J., Wang, S.-K., Choi, M., Reid, B. M., Hu, Y., Lee, Y.-L., Herzog, C. R., Kim-Berman, H., Lee, M., Benke, P. J., Lloyd, K. C. K., Simmer, J. P., & Hu, J. C.-C. (2015). Taurodontism, variations in tooth number, and misshapened crowns in Wnt10a null mice and human kindreds. *Molecular Genetics & Genomic Medicine*, 3(1), 40–58.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13:134, 1-11.
- Yu, T., & Klein, O. D. (2020). Molecular and cellular mechanisms of tooth development, homeostasis and repair. *Development*, 147(2), dev184754, 1-15.