TOOTH DEVELOPMENT:

LEARN FROM "THE NORMAL" AND "THE ABNORMAL"

by

Shih-Kai Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Oral Health Sciences)
in the University of Michigan
2014

Doctoral Committee:
Professor James P. Simmer, Chair
Professor Deborah L. Gumucio
Professor Paul H. Krebsbach
Associate Professor Yuji Mishina

Shih-Kai Wang
© ———————— 2014
All Rights Reserved

DEDICATION

To my parents

Su-Ching Wu and Wen-Da Wang

ACKNOWLEDGEMENTS

I would like to thank all individuals who encouraged and supported me throughout the doctoral training:

Dr. James Simmer and Dr. Jan Hu, as my research advisors, have been providing me with exceptional scientific resources and environments. Their personal commitment to excellence in science has greatly inspired me and is very much appreciated. Also, Dr. Jan Hu has given me personal guidance and support for life in the United States, which I am deeply grateful for.

I would like to express my appreciation to the members of my dissertation committee:

Dr. Deborah Gumucio, for her contributions as a thesis committee member and for being a wonderful teacher who inspired me in developmental biology.

Dr. Paul Krebsbach, for support of my research from the Biologic and Materials Sciences Department and for being a role model of leadership for me.

Dr. Yuji Mishina, for his continuous support and mentoring since my second lab rotation in 2009. His work ethics and attitude in science are always valued.

I have received tremendous and diverse support and training for experiments and scientific discussions and want to thank the following individuals for their generosity of sharing knowledge and experimental materials.

Present and past members of Simmer and Hu laboratory:

Dr. Charles Smith, for inspiring me with his personal commitment and enthusiasm to excellence in dental enamel research and for being a role model for me.

Dr. Yuanyuan Hu, for support in molecular biology, mouse genetics, and histology.

Dr. Yasuo Yamakoshi, for providing fundamental teaching in protein biochemistry.

Dr. Hui-Chen Chan, for providing fundamental teaching in human mutational analyses.

Amelia Richard, for support in molecular biology, mouse genetics, and histology and for her wonderful friendship and partnership.

Andrew Samann, for support in protein biochemistry and for his “fruddy” friendship.

Bryan Reid, for support in mutational analyses, molecular biology, and mouse genetics.

Soumya Pal, for support in mutational analyses and mouse genetics.

Dr. Stephanie Núñez, for the peer support and friendship in Oral Health Sciences Ph.D. program.

Dr. Enamul Kabir, Rachel Milkovich, Rangsiyakorn Lertlam, Dr. Yuhe Lu, Dr. Yonghee Chun, Fumiko Yamakoshi, Dr. Hsun-Liang Chan, Dr. Shuhei Tsuchiya, Dr. Jie Yang, for support in experiments and for creating a vibrant work environment.

Dr. Yoshihiro Komatsu, from Mishina laboratory, for inspiring me in science and for being a role model for me.

Dr. Nobuhiro Kamiya, from Mishina laboratory, for providing fundamental teaching in bone biology.

Dr. Hongxiang Liu, from Mistretta laboratory, for discussion in histology.

Dr. Peter Polverini, the former dean of School of Dentistry, University of Michigan, for creating a unique scientific environment in the school and for providing financial support in the first two years of my doctoral training.

Dr. Robert Lyons and Dr. Brendan Tarrier, at the DNA Sequencing Core, University of Michigan, for whole exome sequencing.

Dr. James Cavalcoli, Dr. Yongsheng Bai, and Dr. Manjusha Pande, at the Bioinformatics Core, University of Michigan, for analyses of exome sequencing data.

The Yale Center for Genome Analysis, Yale University, for whole exome sequencing.

Dr. Murim Choi, at Seoul National University, for analyses of exome sequencing data.

Dr. Kathrin Wilczak, at the Keck Biotechnology Resource Laboratory, Yale University, for conducting mass-spectrometry experiments and analyses.

I am very grateful to the staff of the Oral Health Sciences Research Office, Patricia Schultz, Landon Manette, Charlene Erickson, Kimberly Smith, and Misty Gravelin, for their administrative support and kindness.

All current and past Oral Health Sciences Ph.D. students are appreciated for their comradeship, particularly Christopher Donnelly, for his peer support and inspiring friendship.

I would like to thank faculty and staff of the Biologic and Materials Sciences Department for their support.

I am very grateful to the faculty of my alma mater, National Taiwan University School of Dentistry, for their commitment to dental education and for well preparing me with knowledge and skills for pursuing doctoral training.

Dr. Ming-Kuang Guo, for inspiring me to higher education and academic dentistry.

Dr. Bing-Yu Wang, for his inspiring friendship and for encouraging me to explore myself.

Finally, I want to express my deepest appreciation to my parents, family, and friends for their unconditional love and support.

PREFACE

This thesis includes significant work of Drs. Jan Hu, James Simmer, Hui-Chen Chan, Murim Choi, James Cavalcoli, Manjusha Pande, Yongsheng Bai, Rachel Milkovich, Yuanyuan Hu, Amelia Richardson, Bryan Reid, Soumya Pal, Yasuo Yamakoshi, M. Enamul Kabir, and Andrew Samann.

Description of my contribution to the work presented in this thesis:

I wrote all chapters of this thesis. Some work of Chapter 2 and most of Chapter 3 have been published as 6 manuscripts which work contribution can be referred to. I have updated the references and content to the discussion. Drs. James Simmer and Jan Hu contributed to the experimental design and thesis writing.

I wrote Chapter 1 "Introduction," summarizing the general hypotheses and specific aims, and providing background information about topics discussed in following chapters.

Chapter 2 and 3 describe studies in mutational analysis of human inherited dental defects, including familial tooth agenesis (FTA) and amelogenesis imperfecta (AI). The work presented in these two chapters is part of the ongoing human genetic projects in the lab. All IRB protocols and consents were accomplished by Dr. Jan Hu, who also established collaborations, recruited subjects, acquired patient information, oral photographs, radiographs, and DNA samples. Genomic DNA isolation from blood/saliva was performed by several lab members, including Dr.

Hui-Chen Chan, Rachel Milkovich, Bryan Reid, Soumya Pal, and myself. Target gene analyses of FTA described in Chapter 2 were mainly conducted by myself and partly by Rachel Milkovich, Soumya Pal, and Bryan Reid. Exome sequencing and analyses were conducted at 3 different institutes. Sequencing data generated at the University of Michigan DNA Sequencing Core was analyzed by Drs. James Cavalcoli, Manjusha Pande, and Yongsheng Bai at the University of Michigan Bioinformatics Core. Commercial services from Edge BioSystems (Gaithersburg, MD, U.S.A.) were also used. Dr. Murim Choi at Seoul National University analyzed all of the exome sequencing data generated from Yale Center for Genome Analysis. He called the sequence variants, filtered the data, and narrowed the list of reasonable candidate variants, occasionally even suggesting the specific variant that caused the disease. I routinely further analyzed the filtered data. The results of my analyses are provided as tables in the Chapter 2. Bryan Reid generally performed the PCR amplifications for segregation analyses of the final candidate genes.

In Chapter 4, the purified recombinant FAM83H protein used for kinase reaction was provided by Drs. Yasuo Yamakoshi and M. Enamul Kabir. The *in vitro* kinase assays of FAM83H were conducted mainly by Drs. Jan Hu and Yasuo Yamakoshi and partly by myself. All mass spectrometry experiments and data analyses were conducted by Keck Biotechnology Resource Laboratory at Yale University. Many expression constructs were generated by Custom DNA Constructs, LLC (Cleveland, OH, U.S.A.). All other work described in Chapter 4 was designed and performed by myself with advice and assistance from Drs. James Simmer, Jan Hu, Yasuo Yamakoshi, and Andrew Samann. The work from Chapter 4 is not yet published.

In Chapter 5 “Conclusion,” I summarized the work accomplishment of this thesis and discussed about prospects and potential future directions of the work.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

CHAPTER 2

CHAPTER 4

ABSTRACT

TOOTH DEVELOPMENT:
LEARN FROM "THE NORMAL" AND "THE ABNORMAL"

by

Shih-Kai Wang

Chair: James P. Simmer

For the past decades, tooth development has been extensively studied as a model for understanding organogenesis of ectoderm-derived structures. Much has been learned from the morphological patterning and molecular signaling of normal tooth development in model organisms, mainly rodents. On the other hand, human inherited dental anomalies also provide a valuable source for studying tooth development. Discerning the genetic etiology of these developmental defects not only enhances our understanding of normal tooth development but also provides a fundamental basis for developing potential therapeutic strategies for these disorders.

Familial tooth agenesis (FTA) and amelogenesis imperfecta (AI) are the two most prevalent inherited dental defects in humans, characterized by failed tooth development and

dental enamel malformations respectively. In this dissertation research, we described 7 FTA and 12 AI families and aimed to define the genetic etiology of the diseases through mutational analyses. By using target gene approaches and whole exome analyses, we successfully identified the disease-causing mutations in 2 FTA families and in all the AI cases. The results not only expanded the mutation spectrum of known disease-associated genes but also established novel candidate genes, revealing critical players in tooth and enamel development.

Nevertheless, although human genetic studies of inherited dental and enamel defects have revealed many genes associated with the diseases, the functions of many of these genes and their roles in normal development and pathological conditions remain to be elucidated. Recently, mutations in *FAM83H* (family with sequence similarity 83, member H) were identified to cause autosomal dominant hypocalcified amelogenesis imperfecta (ADHCAI). Although many disease-causing *FAM83H* mutations have been reported, the cellular functions of this gene and the pathogenesis of its associated enamel defects are completely unknown. In this study, we used a biochemical approach to study FAM83H protein-protein interactions and identified CK1 (casein kinase 1) and SEC16A as FAM83H interacting proteins, which indicated a potential cellular function of FAM83H in vesicle trafficking and protein transport. The results also provided an explanation for the high mutation homogeneity of *FAM83H* disease-causing mutations revealed by human genetic studies and suggested a potential pathological mechanism for *FAM83H*-associated enamel defects.

CHAPTER 1

INTRODUCTION

PROBLEM STATEMENT

During evolution, mammals, compared to ancestral vertebrates, have lost a significant capacity for tooth renewal (Jernvall and Thesleff, 2012). In humans, only two sets of dentitions develop during a lifetime. Therefore, for the past decades, significant efforts have been made to regenerate teeth for replacement of missing teeth (Young *et al.*, 2002; Sharpe and Young, 2005; Ikeda and Tsuji, 2008; Ikeda *et al.*, 2009). However, the way to achieve this ultimate goal of regenerative dentistry greatly depends upon our thorough comprehension of tooth development in which the genetic control of this developmental process has not yet been completely understood.

Most of our current knowledge regarding the cellular and genetic basis of tooth development has come from mouse studies. Extensive investigations in molecular regulation of tooth formation have been carried out with many genetically engineered mouse models (Fleischmannova *et al.*, 2008; Bei, 2009). In addition, mutational analysis for human developmental tooth defects has further provided more relevant insight into tooth development in humans (Cobourne and Sharpe, 2013). For example, the identification of *MSX1* mutations as a cause of human tooth agenesis (congenitally missing teeth) demonstrates a critical role of this

transcription factor in early tooth development (Vastardis *et al.*, 1996). Along the same line, discerning the genetic defects causing inherited dental anomalies, such as hereditary enamel and dentin defects, will significantly improve our understanding of the normal developmental processes during tooth formation.

Familial tooth agenesis, amelogenesis imperfecta (inherited enamel defects), and dentinogenesis imperfecta (inherited dentin defects) are so far the three most extensively-studied developmental tooth defects (Kim and Simmer, 2007; Nieminen, 2009; Chan *et al.*, 2011; Cobourne and Sharpe, 2013). Many genes have been shown to be associated with these conditions and to be critical for specific processes of tooth development. However, the genetic etiologies of many cases of these diseases and many other inherited dental defects, such as hyperdontia (formation of extra teeth) and microdontia (small-sized teeth), have not yet been defined, revealing our current incomplete knowledge about tooth development (Cobourne and Sharpe, 2013).

Furthermore, in spite of the great impact of human mutational analysis on discovering critical players in tooth development, the functions of many genes identified from human genetic studies have been largely unknown. For example, our group identified *FAM83H* mutations as the cause of autosomal dominant hypocalcified amelogenesis imperfecta (ADHCAI), demonstrating a significant role of this gene in dental enamel formation (Kim *et al.*, 2008). Although 20 different *FAM83H* mutations have been so far reported (Wang *et al.*, 2013), the molecular functions of *FAM83H* and its role in enamel formation are completely unknown. Also, how mutations in *FAM83H* lead to the enamel defects is still an open question. Therefore, the need for functional studies of specific genes discovered by human mutational analysis of inherited tooth defects is urgent to elucidate the genetic regulation of tooth development.

GENERAL HYPOTHESIS

Our group has been working, for more than a decade, on mutational analysis of various human hereditary dental anomalies, including tooth number abnormalities, inherited enamel and dentin defects. Among these disorders, familial tooth agenesis and amelogenesis imperfecta are the two that we have recruited more families with and studied more extensively. However, the genetic causes of many cases of these two diseases are still to be identified (Chan *et al.*, 2011). We hypothesized that by defining the genetic etiology in these cases through mutational analysis, we would be able to expand the mutational spectrum in known disease candidate genes and discover novel candidate genes.

Furthermore, as mentioned previously, our group identified the genetic defects of ADHCAI in a novel gene, *FAM83H*, of which the molecular functions are completely unknown. We hypothesized that by molecularly characterizing FAM83H protein, we would gain insight into its cellular function in enamel formation and the pathological mechanism of ADHCAI caused by *FAM83H* mutations.

SPECIFIC AIMS

Two specific aims are proposed to test the hypotheses:

SA 1 To define the genetic etiologies of familial tooth agenesis and amelogenesis imperfecta

Dr. Jan Hu has recruited families with familial tooth agenesis (congenital missing teeth) and amelogenesis imperfecta (inherited enamel defects). By conducting mutational analysis with target gene approaches and whole exome sequencing, we aim to identify disease-causing

mutations in these families. Many candidate genes have been reported to cause these two diseases. By accomplishing this specific aim, we expect to find novel mutations in known candidate genes and probably discover new disease-associated genes.

SA 2 To molecularly characterize FAM83H protein *in vitro*

Instead of being directly cloned from developing teeth, *FAM83H* was discovered with mutational analysis through a genome-wide search (Mendoza *et al.*, 2007; Kim *et al.*, 2008). The primary structure of FAM83H, a protein of 1179 amino acids in humans, gives little indication of its potential functions. However, it has no obvious signal peptide and is expected to be a non-secretory protein, which is distinct from the known enamel matrix proteins, such as amelogenin, ameloblastin, and enamelin. This uniqueness of intracellular localization of FAM83H implies its potential intracellular function rather than extracellularly-structural function of all the other enamel proteins. By conducting protein-protein interaction studies, we aim to define the protein interactome of FAM83H, identify FAM83H-binding partners which may provide insights into its intracellular functions, and characterize potential functional domains of FAM83H protein. By accomplishing this specific aim, we expect to know which physiological cellular processes FAM83H may be involved in and probably to unravel the pathological mechanism of *FAM83H*-associated enamel defects.

BACKGROUND AND SIGNIFICANCE

Early tooth development

Tooth development, like that of all the other ectodermal appendages, is a sequential process of reciprocal epithelial-mesenchymal interactions that involves intricate modulation of complex signaling pathways (Bei, 2009; Thesleff, 2013). Morphologically, tooth formation commences with a structure of dental epithelial thickening, the dental lamina, which subsequently proliferates and invaginates into underlying mesenchyme. At the same time, signals from thickened epithelium induce condensation of mesenchymal cells, which is recognized as a tooth germ. The condensed mesenchyme then guides further epithelial invagination and convolution to progress the enamel organ through the sequential bud, cap, and bell-shaped stages of tooth morphogenesis. During these stages, distinct anatomical and functional parts of the tooth form, and the basic shape of tooth crown is established. Eventually, the mesenchyme-derived odontoblasts and epithelium-derived ameloblasts differentiate at the epithelial-mesenchymal

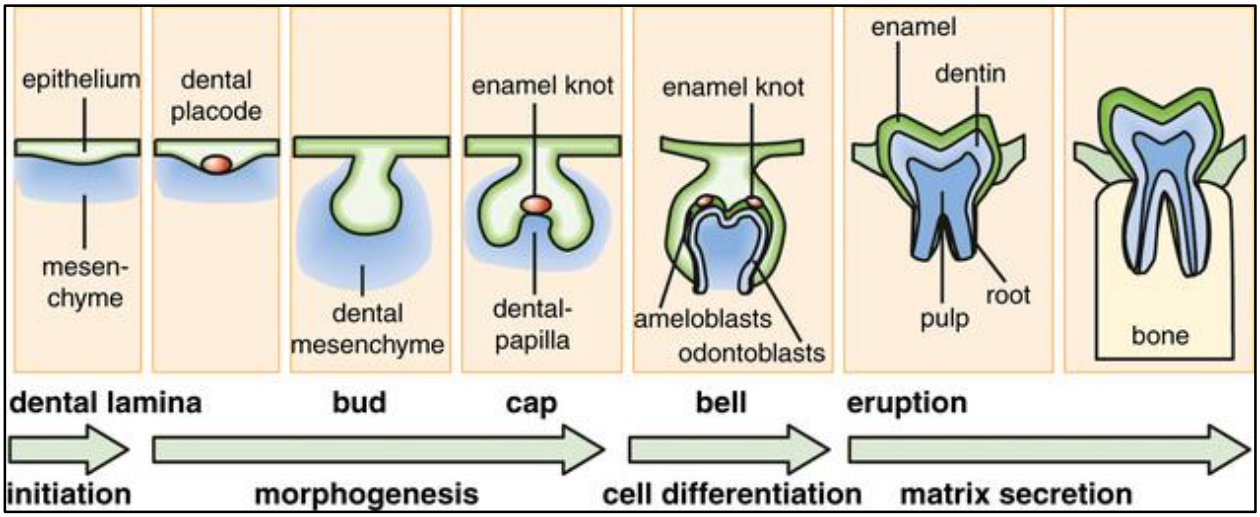


Figure 1.1: The process of tooth development.
The dental placode and enamel knots are signaling centers regulating tooth morphogenesis. This image is reproduced from Figure 1 of reference (Thesleff I. Current understanding of the process of tooth formation: transfer from the laboratory to the clinic. *Aust Dent J* 2013, Epub ahead of print) by permission of John Wiley & Sons Inc.

interface to form dentin and enamel respectively (Fig. 1.1) (Tucker and Sharpe, 2004; Nanci, 2008a; Thesleff, 2013).

Many genes and signaling pathways have been demonstrated to regulate morphological patterning and cell differentiation at specific stages of tooth development. Members of the transforming growth factor β (TGFβ), fibroblast growth factor (FGF), sonic hedgehog (Shh) and Wnt signaling pathways constitute the key pathways that mediate epithelial-mesenchymal interactions during tooth development (Fig. 1.2) (Jernvall and Thesleff, 2000; Bei, 2009; Tummers and Thesleff, 2009; Thesleff, 2013). Animal studies have shown that aberration of any of these pathways leads to developmental arrest of teeth at early stages of tooth formation (Fleischmannova *et al.*, 2008; Bei, 2009). For example, conditional inactivation of Shh in the dental epithelium results in the arrest of tooth development at the bud stage (Dassule *et al.*, 2000). Furthermore, transcription factors mediating these signaling networks are critical for early tooth development as well. Both *Msx1* and *Pax9* null mice exhibit failed tooth formation (Satokata and Maas, 1994; Peters *et al.*, 1998). In humans, environmental or genetic disturbances during the early process of tooth development also lead to the developmental arrest of tooth germs and cause tooth agenesis (De Coster *et al.*, 2009; Parkin *et al.*, 2009).

Familial tooth agenesis

Familial tooth agenesis refers to the inherited failure of teeth to develop. It is the most prevalent craniofacial anomaly in humans (Parkin *et al.*, 2009; Nieminen, 2009). Both the primary (deciduous) and secondary (permanent) dentitions may be affected, with a higher prevalence of the later. In addition, the number and class of teeth involved may vary

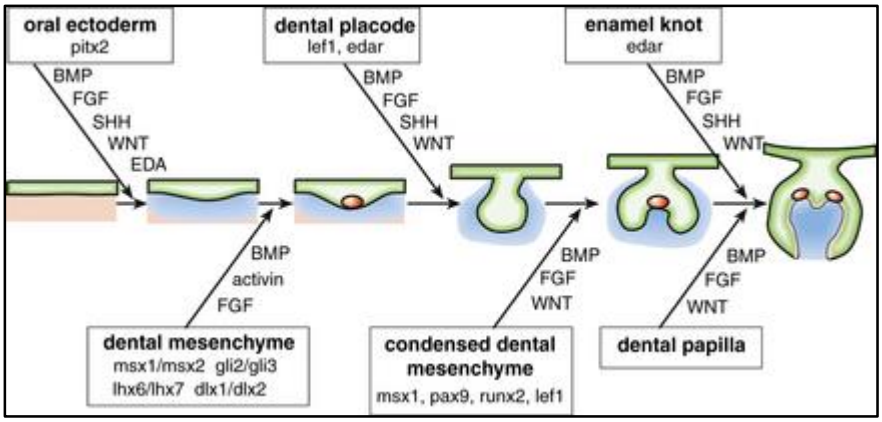


Figure 1.2: Signaling pathways and molecules critical for tooth development. Teeth form from oral epithelium (green) and underlying mesenchyme (blue) and interactions between these tissues regulate development. The most important signal molecules mediating this communication are BMP (bone morphogenetic protein), WNT, SHH (sonic hedgehog) and FGF (fibroblast growth factor). They regulate the expression of important transcription factors indicated in boxes. Loss of function of many of these genes arrests the process of tooth development in genetically modified mice, and their mutations cause tooth agenesis in humans. This image is reproduced from Figure 2 of reference (Thesleff I. Current understanding of the process of tooth formation: transfer from the laboratory to the clinic. *Aust Dent J* 2013, Epub ahead of print) by permission of John Wiley & Sons Inc.

significantly, with considerable demographic and geographic variations. Different patterns of disease inheritance and penetrance have also been described, with autosomal dominant pattern being the most common mode of inheritance. Several terms have been used to describe tooth agenesis with various disease severities. "Hypodontia" has been used as a general term for congenitally missing teeth but currently refers to a specific condition with less than six missing teeth. "Oligodontia" and "anodontia" are used to describe more severe forms of at least six teeth missing and complete lack of teeth respectively (Nieminen, 2009). These terms can refer to isolated disorders in the absence of non-dental phenotypes but can also be used to indicate manifestations in syndromes, such as ectodermal dysplasia. Although most of the cases of tooth agenesis are isolated (non-syndromic), many developmental syndromes include hypodontia as part of their phenotypic anomalies, more than 60 of which are listed in Online Mendelian Inheritance in Men (OMIM) (Cobourne, 2007; Cobourne and Sharpe, 2013).

Genetic defects in a number of genes have been reported to cause non-syndromic tooth agenesis, including *MSX1* (Msh homeobox 1) (STHAG1; MIM #106600) (Vastardis *et al.*, 1996), *PAX9* (paired box 9) (STHAG3; MIM #604625) (Stockton *et al.*, 2000), *EDA* (ectodysplasin A) (STHAGX1; MIM #313500) (Tao *et al.*, 2006), and *WNT10A* (Wingless-type MMTV integration site family, member 10A) (STHAG4; MIM #150400) (Bohring *et al.*, 2009). In addition, mutations in *AXIN2* (axis inhibition protein 2) and *LTBP3* (latent transforming growth factor beta binding protein 3) respectively lead to oligodontia-colorectal cancer syndrome (MIM #608615) (Lammi *et al.*, 2004) and selective tooth agenesis (MIM #613097) (Noor *et al.*, 2009), in which the affected individuals have tooth agenesis with variable expressivity of other non-dental manifestations. MSX1 and PAX9 are two transcription factors shown to be expressed at dental mesenchyme throughout all the stages of odontogenesis (Tummers and Thesleff, 2009). Both *Msx1* and *Pax9* null mice exhibit developmental arrest of teeth at bud stage with abolished mesenchymal expression of BMP4 (bone morphogenic protein 4), a morphogen critical for tooth development (Satokata and Maas, 1994; Peters *et al.*, 1998). EDA is a member of tumor necrosis factor (TNF) superfamily and involved in NF-κB signaling pathway, which is essential for ectodermal organogenesis (Mikkola and Thesleff, 2003). *EDA* mutations contribute to the human X-linked hypohidrotic ectodermal dysplasia, which is featured by EDA triad, hypotrichosis, hypodontia, and hypohidrosis (Kere *et al.*, 1996). However, affected hemizygous females sometimes only show congenital absence of teeth without other non-dental phenotypes, making *EDA* a candidate gene for isolated tooth agenesis (Tao *et al.*, 2006). Genetic defects of *WNT10A* were first identified to cause Odontoonychodermal dysplasia (OODD; MIM#257980), a rare autosomal recessive syndrome characterized by features of ectodermal dysplasia, including severe oligodontia (Adaimy *et al.*, 2007). However, it was recently reported that people carrying

heterozygous *WNT10A* mutations are prone to have tooth agenesis with no other OODD characteristics (Bohring *et al.*, 2009). These findings reaffirm the critical role of Wnt signaling in tooth development, which has been demonstrated in many mouse models (Fleischmannova *et al.*, 2008; Bei, 2009). Furthermore, another player of the Wnt signaling pathway, AXIN2, is also associated with tooth agenesis. However, it is worth noting that people with *AXIN2* mutations have not only severe oligodontia but also a predisposition to colorectal cancers (Lammi *et al.*, 2004). As TGFβ signaling has been known to be important for tooth development, homozygous *LTBP3* mutations were found in a family with oligodontia and short stature with increased bone density (Noor *et al.*, 2009). In addition to these genes associated with isolated tooth agenesis, many others are found in syndromic tooth agenesis, such as *PITX2* and *IKBKG* (NEMO), which reveals more critical players in early tooth development (Cobourne and Sharpe, 2013).

Enamel development (amelogenesis)

Dental enamel is the most highly mineralized tissue in human bodies, comprised of more than 96% calcium hydroxyapatite crystallites. Its hardness and strength come from not only the high mineral content, but also the highly-ordered organization of enamel crystallites. Therefore, dental enamel formation, amelogenesis, requires an intricately regulated orchestration of cellular and chemical events during the process to properly make enamel with the necessary mineral composition and organization (Smith, 1998, Simmer *et al.*, 2010). Environmental disturbances or genetic aberrations in this process will lead to developmental enamel defects (Winter and Brook, 1975; Witkop, 1988; Hu *et al.*, 2007).

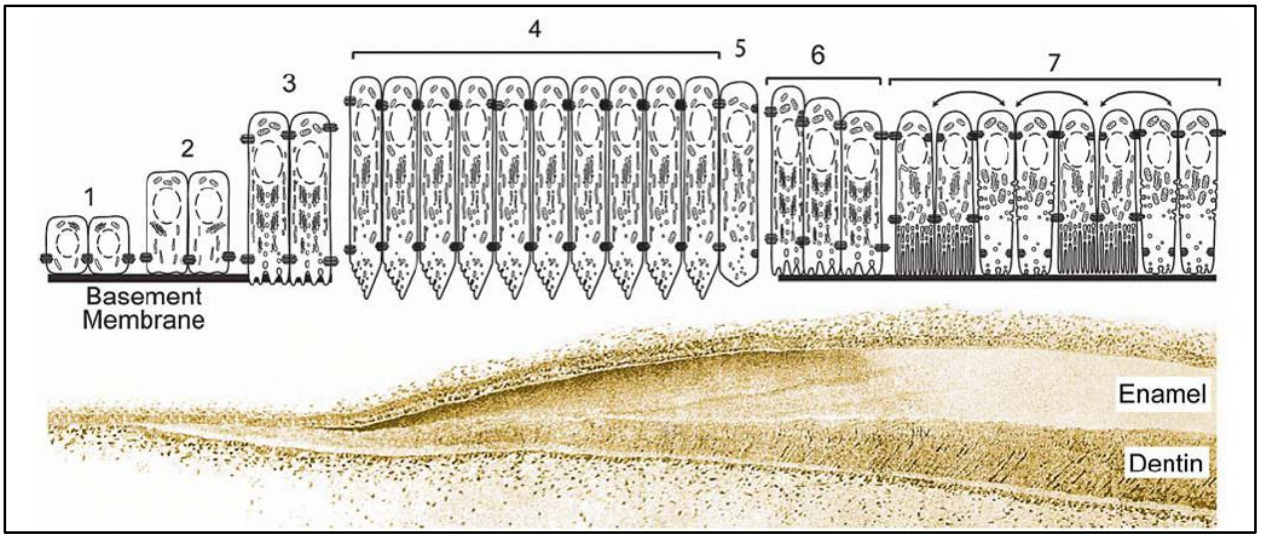


Figure 1.3: Ameloblast changes during enamel formation.
The epithelial cells of the inner enamel epithelium (1) rest on a basement membrane containing laminin. These cells increase in length as differentiating ameloblasts above the predentin matrix (2). Presecretory ameloblasts send processes through the degenerating basement membrane as they initiate the secretion of enamel proteins on the villous surface of mineralizing dentin (3). After establishing the dentinoenamel junction and mineralizing a thin layer of aprismatic enamel, secretory ameloblasts develop a secretory specialization, or Tomes' process. Along the secretory face of the Tomes' process, in place of the absent basement membrane, secretory ameloblasts secrete proteins at a mineralization front where the enamel crystals grow in length (4). Each enamel rod follows a retreating Tomes' process from a single ameloblast. At the end of the secretory stage, ameloblasts lose their Tomes' process and produce a thin layer of aprismatic enamel (5). At this point the enamel has achieved its final thickness. During the transition stage, the ameloblasts undergo a major restructuring that diminishes their secretory activity and changes the types of proteins secreted (6). KLK4 is secreted, which degrades the accumulated protein matrix and amelotin (AMTN) is secreted as part of the new basement membrane. During the maturation stage ameloblasts modulate between ruffled and smooth-ended phases (7). Their activities harden the enamel layer by promoting the deposition of mineral on the sides of enamel crystals laid down during the secretory stage. The histology of the developing tooth is adapted from Uchida et al. [1991]. This image is reproduced from Figure 1 of reference (Hu JC, Chun YH, Al Hazzazzi T, Simmer JP (2007). Enamel formation and amelogenesis imperfecta. *Cells Tissues Organs* 186:78-85.) by permission of S. Karger AG, Medical and Scientific Publishers.

Enamel development can be divided into four stages, which is defined by the morphology and function of ameloblasts, the enamel forming cells (Fig. 1.3) (Hu *et al.*, 2007). They are presecretory, secretory, transition, and maturation stages. After early morphogenesis of tooth development is complete, a layer of epithelial cells adjacent to underlying mesenchyme (inner enamel epithelium) starts to differentiate into ameloblasts. These cells gradually elongate in

morphology and reverse the cell polarity with nuclei facing away from the mesenchyme. At the same time, the differentiating cells, preameloblasts, extend cytoplasmic projections through the basement membrane, which is gradually degraded and removed. This presecretory stage prepares a proper environment for active matrix and ion deposition in subsequent stages of amelogenesis. During the secretory stage, the tall columnar ameloblasts secrete large amounts of enamel matrix proteins (amelogenin, ameloblastin, and enamelin), and the enamel crystallites form, in association with the secreted proteins. At the same time, ameloblasts retreat from the dentinoenamel junction (DEJ) to allow extension of crystal ribbons (appositional growth), which establishes the thickness of the enamel. Once the enamel reaches its full thickness, the secretory-stage ameloblasts retract their cytoplasmic processes and shorten into transition-stage ameloblasts, under which a basement membrane reforms. Subsequently, the transition-stage ameloblasts further transform into short cuboidal maturation-stage ameloblasts, which start modulating between ruffle and smooth-ended cells at the enamel surface. During the maturation stage, the enamel matrix proteins are further degraded by enamel proteases, mainly by kallikrein-related peptidase 4 (KLK4) and less by matrix metalloproteinase 20 (MMP20), and are removed to provide space for ion deposition onto the sides of enamel crystallites (enamel maturation), which establishes the hardness of the enamel (Fig. 1.3) (Simmer and Fincham, 1995; Smith, 1998; Fincham *et al.*, 1999; Hu *et al.*, 2007; Nanci, 2008b; Simmer *et al.*, 2010; Bartlett, 2013). As mentioned previously, these developmental processes are intricately regulated and, therefore, highly susceptible to environmental or genetic influences.

Amelogenesis imperfecta (AI)

Amelogenesis imperfecta (AI) is a group of inherited disorders characterized by enamel malformations typically in the absence of non-dental phenotypes (isolated AI). The term is also used to describe the enamel defects in syndromes (syndromic AI) (Witkop, 1988; Hu *et al.*, 2007). The prevalence of AI significantly varies geographically, ranging from 1:14,000 (U.S.A.) to 1:700 (North Sweden), depending on various populations (Witkop and Sauk, 1976; Witkop, 1988). In addition, the disorder exhibits a great variability in disease phenotypes, which is classified according to the thickness, hardness and smoothness of the affected enamel (Witkop, 1988; Aldred *et al.*, 2003; Hu *et al.*, 2007). Variations in these parameters are believed to result from differences in the timing of disturbances during amelogenesis. As the final thickness of dental enamel is established at secretory stage, disruption of this process will lead to insufficient enamel appositional growth and pathologically thin or *hypoplastic* enamel (hypoplastic AI). On the other hand, the maturation stage of amelogenesis is critical for development of enamel hardness. Aberration of this process will cause insufficient mineral deposits on the sides of enamel crystallites and leave enamel pathologically soft, a condition known as *hypomaturation* AI. In this case, the affected enamel is of normal thickness but easily abraded after tooth eruption due to the hardness defect. *Hypocalcified* AI is another form of AI in which the failure in mineralization is the most extreme. The enamel may have normal thickness but is rough, soft, and easily chipped off from tooth surfaces. Patients with AI often experience difficulty chewing and maintaining oral hygiene, have lower self-esteem due to poor dental appearance, and report compromised quality of life (Coffield *et al.*, 2005).

Defects in a number of genes have been reported to cause isolated AI, including *AMELX*, *ENAM*, *KLK4*, *MMP20*, *FAM83H*, *WDR72*, *C4orf26*, *SLC24A4*, and *LAMB3* (Aldred *et al.*, 1992; Rajpar *et al.*, 2001; Hart *et al.*, 2004; Kim *et al.*, 2005; Kim *et al.*, 2008; El-Sayed *et al.*,

2009; Parry *et al.*, 2012; Parry *et al.*, 2013; Poulter *et al.*, 2013). In addition, many other genes have been shown to be associated with developmental syndromes with enamel defects, such as *DLX3*, *CNNM4*, *ROGDI*, *FAM20A*, *STIM1*, and *ORAI1* (Cobourne and Sharpe, 2013). These findings reveal many critical players in dental enamel formation. *AMELX* (amelogenin) and *ENAM* (enamelin) are two genes encoding enamel matrix proteins that are actively secreted by ameloblasts at secretory stage. Mutations in these two genes lead to hypoplastic forms of AI (Hu *et al.*, 2007). MMP20 and KLK4 are two major enamel proteases expressed by mainly secretory-stage and maturation-stage ameloblasts respectively. While MMP20 processes enamel matrix proteins during the secretory stage, KLK4 is responsible for further degradation of residual proteins during the maturation stage (Hu *et al.*, 2007; Bartlett, 2013). Mutations in these two genes generally cause enamel hardness defects (hypomaturation AI) due to defective protein degradation and removal (Hu *et al.*, 2007). Except the four genes encoding enamel matrix proteins and proteases, all of the other AI associated genes were identified through human mutational analysis of AI kindreds (Kim *et al.*, 2006; Chan *et al.*, 2011). Therefore, the functions of most of these genes in enamel formation are largely unknown and under investigation, although some of them can be implied and predicted based upon limited current knowledge of the genes.

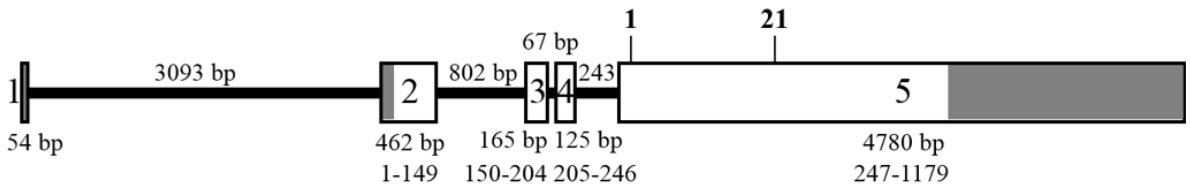
FAM83H **mutations and autosomal dominant hypocalcified AI (ADHCAI)**

Autosomal dominant hypocalcified AI (ADHCAI) is a specific form of AI in which the affected enamel is cheesy-soft and easily lost soon after tooth eruption (Witkop, 1988). Full thickness of enamel can only be seen on unerupted teeth from a radiograph or on erupting teeth without much attrition. The surface of enamel is rough and usually with brown-yellow

discoloration. Many efforts had been made to determine the genetic etiology of ADHCAI. However, target gene approaches failed to identify genetic defects in any of the known AI candidate genes encoding enamel matrix proteins and proteases (Kim *et al.*, 2006). In 2008, by linkage analysis and a gene-by-gene screening, our group identified *FAM83H* mutations as a cause of ADHCAI (Kim *et al.*, 2008).

FAM83H (family with sequence similarity 83, member H) is a gene that was first identified through bioinformatics prediction in Human Genome Project and was categorized into the *FAM83* gene family based upon sequence similarity. It is predicted to have 5 exons with a non-coding Exon 1 and a large Exon 5, and to encode a protein of 1179 amino acids with a calculated molecular mass of 127-kDa (Kim *et al.*, 2008). Domain analysis of FAM83H protein reveals neither a signal peptide nor any known functional domains except an N-terminal phospholipase D-like (PLD-like) domain (amino acids 17-281), which is the shared element among all the members in FAM83 family and gives the group its identity. However, the trace homology between this domain and PLD suggests the unlikeliness of FAM83H having PLD-like enzyme activity, and leaves the functions of this protein still unknown (Kim *et al.*, 2008; Ding *et al.*, 2009).

At least 20 different disease-causing *FAM83H* mutations have been so far identified (Fig. 1.4) (Wang *et al.*, 2013). All the mutations are either nonsense mutations (17/20) or frameshifts (3/20) located at 5' region of Exon 5. Introducing premature stop codons at the last exon, all reported mutations are predicted to produce mutant transcripts that can escape nonsense mediated decay and are expected to generate truncated proteins lacking the normal C-terminus (from p.287Ser* to p.Glu694*). There are no other types of loss-of-function mutations reported, such as missense mutations, splice-site mutations, or large deletions. Furthermore, several



#	Gene	cDNA	Protein	References
1	g.5411C>A	c.860C>A	p.S287*	(Wright *et al.*, 2009)
2	g. 5175T>A	c.891T>A	p.Y297*	(Lee *et al.*, 2008)
3	g. 5190T>A	c.906T>G	p.Y302*	(Haubek *et al.*, 2011; Song *et al.*, 2012)
4	g.5207_5208delTC	c.923_924delTC	p.L308fs*323	(Wright *et al.*, 2009)
5		c.924dupT	p.V309Rfs*324	(Song *et al.*, 2012)
6	g.5257C>T	c.973C>T	p.R325*	(Kim *et al.*, 2008; Song *et al.*, 2012)
7	g.5476C>T	c.1192C>T	p.Q398*	(Ding *et al.*, 2009; Hart *et al.*, 2009; Kim *et al.*, 2008; Wright *et al.*, 2011)
8	g.5528G>T	c.1243G>T	p.E415*	(Lee *et al.*, 2008)
9	g.5573C>A	c.1289C>A	p.S430*	(Wright *et al.*, 2011)
10	g.5614C>T	c.1330C>T	p.Q444*	(Ding *et al.*, 2009; Hart *et al.*, 2009)
11	g.5638C>T	c.1354C>T	pQ452*	(Chan *et al.*, 2011; Haubek *et al.*, 2011; Hyun *et al.*, 2009; Song *et al.*, 2012)
12	g.5650C>T	c.1366C>T	p.Q456*	(Hart *et al.*, 2009)
13	g.10653C>T	c.1369C>T	p.Gln457*	this thesis
14	g.5658C>A	c.1374C>A	p.Y458*	(El-Sayed *et al.*, 2010)
15	g.5663G>A	c.1379G>A	p.W460*	(Wright *et al.*, 2009)
16	g.5664G>A	c.1380G>A	p.W460*	(Lee *et al.*, 2008)
17	g.5692G>A	c.1408C>T	p.Q470*	(Wright *et al.*, 2009)
18	g.6156_6157delCC	c.1872_1873delCC	p.L625fs*703	(Wright *et al.*, 2009)
19	g.6277C>T	c.1993 C>T	p.Q665*	(Lee *et al.*, 2011)
20	g.6313C>T	c.2029 C>T	p.Q677*	(Chan *et al.*, 2011; Lee *et al.*, 2008; Lee *et al.*, 2011; Song *et al.*, 2012; Wright *et al.*, 2011)
21	g.6364G>T	c.2080G>T	p.E694*	(Wright *et al.*, 2009)

Figure 1.4: *FAM83H* disease-causing mutations.
FAM83H gene structure: numbered boxes indicate exons; introns are lines connecting the exons. The numbers above each intron indicate the length of the intron in basepairs (bp). The numbers below each exon show the length of the exon in bp and below that the range of amino acids encoded by it. Shaded exon regions are non-coding. The 21 *FAM83H* missense or frameshift mutations are located between the sites marked 1 and 21 in bold. The gene numbers start from the first nucleotide of the *FAM83H* reference sequence NG_016652.1. The cDNA numbers start from the translation initiation site of *FAM83H* reference sequence NM_198488.3. This image is reproduced and modified from Appendix Figure 5 of reference (Wang SK, Hu Y, Simmer JP, Seymen F, Estrella NM, Pal S, Reid BM, Yildirim M, Bayram M, Bartlett JD, Hu JC (2013). Novel KLK4 and MMP20 mutations discovered by whole-exome sequencing. *J Dent Res* 92:266-71.) by permission of SAGE Journals.

mutations have even been reported more than once from different ADHCAI kindreds, suggesting a high genetic homogeneity of *FAM83H* disease-causing mutations (Fig. 1.4) (Wang *et al.*, 2013).

The general goal of the studies incorporated into this thesis was to identify critical players of tooth formation and to understand their roles and functions during this developmental process. Our main approaches were to conduct mutational analyses in families with familial tooth agenesis and amelogenesis imperfecta and to characterize FAM83H protein at molecular level. These genetic and biochemical approaches were used to find new players in tooth and enamel formation and gather information to unravel the roles of FAM83H in normal development as well as in pathological conditions.

REFERENCES

Adaimy L, Chouery E, Megarbane H, Mroueh S, Delague V, Nicolas E, Belguith H, de Mazancourt P, Megarbane A (2007). Mutation in WNT10A is associated with an autosomal recessive ectodermal dysplasia: the odonto-onycho-dermal dysplasia. *Am J Hum Genet* 81:821-8.

Aldred MJ, Crawford PJ, Roberts E, Thomas NS (1992). Identification of a nonsense mutation in the amelogenin gene (AMELX) in a family with X-linked amelogenesis imperfecta (AIH1). *Hum Genet* 90:413-6.

Aldred MJ, Savarirayan R, Crawford PJ (2003). Amelogenesis imperfecta: a classification and catalogue for the 21st century. *Oral Dis* 9:19-23.

Bartlett JD (2013). Dental Enamel Development: Proteinases and Their Enamel Matrix Substrates. *ISRN Dent* 2013:684607.

Bei M (2009). Molecular genetics of tooth development. *Curr Opin Genet Dev* 19:504-10.

Bohring A, Stamm T, Spaich C, Haase C, Spree K, Hehr U, Hoffmann M, Ledig S, Sel S, Wieacker P, Röpke A (2009). WNT10A mutations are a frequent cause of a broad spectrum of ectodermal dysplasias with sex-biased manifestation pattern in heterozygotes. *Am J Hum Genet* 85:97-105.

Chan HC, Estrella NM, Milkovich RN, Kim JW, Simmer JP, Hu JC (2011). Target gene analyses of 39 amelogenesis imperfecta kindreds. *Eur J Oral Sci* 119 Suppl 1:311-23.

Cobourne MT (2007). Familial human hypodontia--is it all in the genes? *Br Dent J* 203:203-8.

Cobourne MT, Sharpe PT (2013). Diseases of the tooth: the genetic and molecular basis of inherited anomalies affecting the dentition. *Wiley Interdiscip Rev Dev Biol* 2:183-212.

Coffield KD, Phillips C, Brady M, Roberts MW, Strauss RP, Wright JT (2005). The psychosocial impact of developmental dental defects in people with hereditary amelogenesis imperfecta. *J Am Dent Assoc* 136:620-30.

Dassule HR, Lewis P, Bei M, Maas R, McMahon AP (2000). Sonic hedgehog regulates growth and morphogenesis of the tooth. *Development* 127:4775-85.

De Coster PJ, Marks LA, Martens LC, Huysseune A (2009). Dental agenesis: genetic and clinical perspectives. *J Oral Pathol Med* 38:1-17.

Ding Y, Estrella MR, Hu YY, Chan HL, Zhang HD, Kim JW, Simmer JP, Hu JC (2009). Fam83h is associated with intracellular vesicles and ADHCAI. *J Dent Res* 88:991-6.

El-Sayed W, Parry DA, Shore RC, Ahmed M, Jafri H, Rashid Y, Al-Bahlani S, Al Harasi S, Kirkham J, Inglehearn CF, Mighell AJ (2009). Mutations in the beta propeller WDR72 cause autosomal-recessive hypomaturation amelogenesis imperfecta. *Am J Hum Genet* 85:699-705.

El-Sayed W, Shore RC, Parry DA, Inglehearn CF, Mighell AJ (2010). Ultrastructural analyses of deciduous teeth affected by hypocalcified amelogenesis imperfecta from a family with a novel Y458X FAM83H nonsense mutation. *Cells Tissues Organs* 191:235-9.

Fincham AG, Moradian-Oldak J, Simmer JP (1999). The structural biology of the developing dental enamel matrix. *J Struct Biol* 126:270-99.

Fleischmannova J, Matalova E, Tucker AS, Sharpe PT (2008). Mouse models of tooth abnormalities. *Eur J Oral Sci* 116:1-10.

Hart PS, Hart TC, Michalec MD, Ryu OH, Simmons D, Hong S, Wright JT (2004). Mutation in kallikrein 4 causes autosomal recessive hypomaturation amelogenesis imperfecta. *J Med Genet* 41:545-9.

Hart PS, Becerik S, Cogulu D, Emingil G, Ozdemir-Ozenen D, Han ST, Sulima PP, Firatli E, Hart TC (2009). Novel FAM83H mutations in Turkish families with autosomal dominant hypocalcified amelogenesis imperfecta. *Clin Genet* 75:401-4.

Haubek D, Gjørup H, Jensen LG, Juncker I, Nyegaard M, Børglum AD, Poulsen S, Hertz JM (2011). Limited phenotypic variation of hypocalcified amelogenesis imperfecta in a Danish five-generation family with a novel FAM83H nonsense mutation. *Int J Paediatr Dent* 21:407-12.

Hu JC, Chun YH, Al Hazzazzi T, Simmer JP (2007). Enamel formation and amelogenesis imperfecta. *Cells Tissues Organs* 186:78-85.

Hyun HK, Lee SK, Lee KE, Kang HY, Kim EJ, Choung PH, Kim JW (2009). Identification of a novel FAM83H mutation and microhardness of an affected molar in autosomal dominant hypocalcified amelogenesis imperfecta. *Int Endod J* 42:1039-43.

Ikeda E, Tsuji T (2008). Growing bioengineered teeth from single cells: potential for dental regenerative medicine. *Expert Opin Biol Ther* 8:735-44.

Ikeda E, Morita R, Nakao K, Ishida K, Nakamura T, Takano-Yamamoto T, Ogawa M, Mizuno M, Kasugai S, Tsuji T (2009). Fully functional bioengineered tooth replacement as an organ replacement therapy. *Proc Natl Acad Sci USA* 106:13475-80.

Jernvall J, Thesleff I (2000). Reiterative signaling and patterning during mammalian tooth morphogenesis. *Mech Dev* 92:19-29.

Jernvall J, Thesleff I (2012). Tooth shape formation and tooth renewal: evolving with the same signals. *Development* 139:3487-97.

Kere J, Srivastava AK, Montonen O, Zonana J, Thomas N, Ferguson B, Munoz F, Morgan D, Clarke A, Baybayan P, Chen EY, Ezer S, Saarialho-Kere U, de la Chapelle A, Schlessinger D (1996). X-linked anhidrotic (hypohidrotic) ectodermal dysplasia is caused by mutation in a novel transmembrane protein. *Nat Genet* 13:409-16.

Kim JW, Simmer JP, Hart TC, Hart PS, Ramaswami MD, Bartlett JD, Hu JC (2005). MMP-20 mutation in autosomal recessive pigmented hypomaturation amelogenesis imperfecta. *J Med Genet* 42:271-5.

Kim JW, Simmer JP, Lin BP, Seymen F, Bartlett JD, Hu JC (2006). Mutational analysis of candidate genes in 24 amelogenesis imperfecta families. *Eur J Oral Sci* 114 Suppl 1:3-12; discussion 39-41, 379.

Kim JW, Simmer JP (2007). Hereditary dentin defects. *J Dent Res* 86:392-9.

Kim JW, Lee SK, Lee ZH, Park JC, Lee KE, Lee MH, Park JT, Seo BM, Hu JC, Simmer JP (2008). FAM83H mutations in families with autosomal-dominant hypocalcified amelogenesis imperfecta. *Am J Hum Genet* 82:489-94.

Lammi L, Arte S, Somer M, Jarvinen H, Lahermo P, Thesleff I, Pirinen S, Nieminen P (2004). Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am J Hum Genet* 74:1043-50.

Lee SK, Hu JC, Bartlett JD, Lee KE, Lin BP, Simmer JP, Kim JW (2008). Mutational spectrum of FAM83H: the C-terminal portion is required for tooth enamel calcification. *Hum Mutat* 29:E95-9.

Lee SK, Lee KE, Jeong TS, Hwang YH, Kim S, Hu JC, Simmer JP, Kim JW (2011). FAM83H mutations cause ADHCAI and alter intracellular protein localization. *J Dent Res* 90:377-81.

Mendoza G, Pemberton TJ, Lee K, Scarel-Caminaga R, Mehrian-Shai R, Gonzalez-Quevedo C, Ninis V, Hartiala J, Allayee H, Snead ML, Leal SM, Line SR, Patel PI (2007). A new locus for autosomal dominant amelogenesis imperfecta on chromosome 8q24.3. *Hum Genet* 120:653-62.

Mikkola ML, Thesleff I (2003). Ectodysplasin signaling in development. *Cytokine Growth Factor Rev* 14:211-24.

Nanci A (2008a). Development of the tooth and its supporting tissues. In: Ten Cate's Oral Histology Development, Structure, and Function. Nanci A editor. St. Louis: Mosby, pp.79-107.

Nanci A (2008b). Enamel: composition, formation, and structure. In: Ten Cate's Oral Histology Development, Structure, and Function. Nanci A editor. St. Louis: Mosby, pp.141-90.

Nieminen P (2009). Genetic basis of tooth agenesis. *J Exp Zool B Mol Dev Evol* 312B:320-42.

Noor A, Windpassinger C, Vitcu I, Orlic M, Rafiq MA, Khalid M, Malik MN, Ayub M, Alman B, Vincent JB (2009). Oligodontia is caused by mutation in LTBP3, the gene encoding latent TGF-beta binding protein 3. *Am J Hum Genet* 84:519-23.

Parkin N, Elcock C, Smith RN, Griffin RC, Brook AH (2009). The aetiology of hypodontia: the prevalence, severity and location of hypodontia within families. *Arch Oral Biol* 54 Suppl 1:S52-6.

Parry DA, Brookes SJ, Logan CV, Poulter JA, El-Sayed W, Al-Bahlani S, Al Harasi S, Sayed J, Raïf el M, Shore RC, Dashash M, Barron M, Morgan JE, Carr IM, Taylor GR, Johnson CA, Aldred MJ, Dixon MJ, Wright JT, Kirkham J, Inglehearn CF, Mighell AJ (2012). Mutations in C4orf26, encoding a peptide with in vitro hydroxyapatite crystal nucleation and growth activity, cause amelogenesis imperfecta. *Am J Hum Genet* 91:565-71.

Parry DA, Poulter JA, Logan CV, Brookes SJ, Jafri H, Ferguson CH, Anwari BM, Rashid Y, Zhao H, Johnson CA, Inglehearn CF, Mighell AJ (2013). Identification of mutations in SLC24A4, encoding a potassium-dependent sodium/calcium exchanger, as a cause of amelogenesis imperfecta. *Am J Hum Genet* 92:307-12.

Peters H, Neubüser A, Kratochwil K, Balling R (1998). Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes Dev* 12:2735-47.

Rajpar MH, Harley K, Laing C, Davies RM, Dixon MJ (2001). Mutation of the gene encoding the enamel-specific protein, enamelin, causes autosomal-dominant amelogenesis imperfecta. *Hum Mol Genet* 10:1673-7.

Satokata I, Maas R (1994). Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat Genet* 6:348-56.

Sharpe PT, Young CS (2005). Test-tube teeth. *Sci Am* 293:34-41.

Simmer JP, Fincham AG (1995). Molecular mechanisms of dental enamel formation. *Crit Rev Oral Biol Med* 6:84-108.

Simmer JP, Papagerakis P, Smith CE, Fisher DC, Rountrey AN, Zheng L, Hu JC (2010). Regulation of dental enamel shape and hardness. *J Dent Res* 89:1024-38.

Smith CE (1998). Cellular and chemical events during enamel maturation. *Crit Rev Oral Biol Med* 9:128-61.

Song YL, Wang CN, Zhang CZ, Yang K, Bian Z (2012). Molecular characterization of amelogenesis imperfecta in Chinese patients. *Cells Tissues Organs* 196:271-9.

Stockton DW, Das P, Goldenberg M, D'Souza RN, Patel PI (2000). Mutation of PAX9 is associated with oligodontia. *Nat Genet* 24:18-9.

Tao R, Jin B, Guo SZ, Qing W, Feng GY, Brooks DG, Liu L, Xu J, Li T, Yan Y, He L (2006). A novel missense mutation of the EDA gene in a Mongolian family with congenital hypodontia. *J Hum Genet* 51:498-502.

Thesleff I (2013). Current understanding of the process of tooth formation: transfer from the laboratory to the clinic. *Aust Dent J* [Epub ahead of print]

Tucker A, Sharpe P (2004). The cutting-edge of mammalian development; how the embryo makes teeth. *Nat Rev Genet* 5:499-508.

Tummers M, Thesleff I (2009). The importance of signal pathway modulation in all aspects of tooth development. *J Exp Zool B Mol Dev Evol* 312B:309-19.

Vastardis H, Karimbux N, Guthua SW, Seidman JG, Seidman CE (1996). A human MSX1 homeodomain missense mutation causes selective tooth agenesis. *Nat Genet* 13:417-21.

Wang SK, Hu Y, Simmer JP, Seymen F, Estrella NM, Pal S, Reid BM, Yildirim M, Bayram M, Bartlett JD, Hu JC (2013). Novel KLK4 and MMP20 mutations discovered by whole-exome sequencing. *J Dent Res* 92:266-71.

Winter GB, Brook AH (1975). Enamel hypoplasia and anomalies of the enamel. *Dent Clin North Am* 19:3-24.

Witkop CJ Jr, Sauk JJ Jr (1976). Hereditable defects of enamel. In: Oral Facial Genetics. Stewart RE and Prescott GH editors. St. Louis: Mosby, pp.151-226.

Witkop CJ Jr (1988). Amelogenesis imperfecta, dentinogenesis imperfecta and dentin dysplasia revisited: problems in classification. *J Oral Pathol* 17:547-53.

Wright JT, Frazier-Bowers S, Simmons D, Alexander K, Crawford P, Han ST, Hart PS, Hart TC (2009). Phenotypic variation in FAM83H-associated amelogenesis imperfecta. *J Dent Res* 88:356-60.

Wright JT, Torain M, Long K, Seow K, Crawford P, Aldred MJ, Hart PS, Hart TC (2011). Amelogenesis imperfecta: genotype-phenotype studies in 71 families. *Cells Tissues Organs* 194:279-83.

Young CS, Terada S, Vacanti JP, Honda M, Bartlett JD, Yelick PC (2002). Tissue engineering of complex tooth structures on biodegradable polymer scaffolds. *J Dent Res* 81:695-700.

CHAPTER 2

MUTATIONAL ANALYSIS – FAMILIAL TOOTH AGENESIS

ABSTRACT

Tooth agenesis is the most prevalent craniofacial congenital anomaly in humans. The term refers to an isolated disorder in the absence of non-dental phenotypes but is also used to describe the manifestation of missing teeth in syndromes. The affected individuals suffer from compromised masticatory functions and have decreased quality of life. Discerning the genetic etiology of tooth agenesis not only improves our understanding of normal tooth development but also provides a fundamental basis for developing potential therapeutic strategies for this anomaly. To date, *MSX1*, *PAX9*, *AXIN2*, *EDA*, and *WNT10A* have been established as candidate genes associated with non-syndromic tooth agenesis. However, there are still many cases for which the genetic mutations cannot be found in these genes, suggesting high genetic heterogeneity of this disorder. In this thesis, we described 7 families with non-syndromic tooth agenesis, in which there was considerable variation in the number and class of teeth that were involved, and aimed to define the causative mutations for the disease. We identified a novel *PAX9* mutation (c.43T>A, p.Phe15Ile) in a family of which the proband had 10 missing teeth and two reported *WNT10A* mutations (c.321C>A, p.Cys107*; c.682T>A, p.Phe228Ile) in another family of which two compound heterozygous individuals had 8 and 15 teeth missing. However,

for the other 5 families, we were still unable to define their genetic defects even with the exome sequencing analyses. Our results reaffirmed the genotypic and phenotypic heterogeneity of tooth agenesis. In the future, we need to recruit more families with similar phenotypes and collectively analyze their exome data at a "genetic pathway level," meaning that mutations in genes involved in similar pathways should be considered to probably cause similar disease phenotypes. For sporadic cases, the genetic cause should not be determined until the whole exome data has been scrutinized, especially for those who have mild phenotypes (missing only one or two teeth), since genotype-phenotype segregation cannot be discerned to confirm the causality in these cases.

INTRODUCTION

In vertebrates, tooth development starts when specific epithelial domains of the first branchial arch express inductive signals to underlying mesenchyme, which defines the "odontogenic fields" that determine the location of future teeth. Subsequently, the odontogenic potential shifts to the mesenchyme, and mesenchymal factors signal back to epithelium and direct tooth bud morphogenesis (Tucker and Sharpe, 2004; Nanci, 2008; Thesleff, 2013). Disturbances during this early process of tooth development will lead to developmental arrest of tooth buds and cause tooth agenesis (De Coster *et al.*, 2009; Parkin *et al.*, 2009).

Tooth agenesis is the most prevalent craniofacial congenital anomaly in humans. More than 20% of people fail to develop at least one of the third molars (wisdom teeth) and 3–10% one or more of the other permanent teeth (Parkin *et al.*, 2009; Nieminen, 2009). There is considerable variation in the number and class of teeth that can be involved. Except third molars, the lower second premolars are the most commonly affected, followed by upper lateral incisors and upper

second premolars in Caucasian populations (Nieminen, 2009; Cobourne and Sharpe, 2013). Different terms have been used to describe the severity of tooth agenesis. While "hypodontia" refers to the situation with less than six missing teeth, "oligodontia" is used to indicate a more severe condition with six or more missing teeth except third molars (Nieminen, 2009).

Tooth agenesis is usually an isolated abnormality, but it can also be one of the manifestations of syndromes, such as ectodermal dysplasia (Cobourne and Sharpe, 2013). Especially, a significant proportion of severe tooth agenesis happens in syndromes, in which the affected individuals suffer from compromised masticatory functions and decreased quality of life. Unraveling the genetic etiology of tooth agenesis not only gains our understanding of normal tooth development but also provides a fundamental basis for developing potential therapeutic strategies for this anomaly. To date, mutations in *MSX1* (Msh homeobox 1), *PAX9* (paired box 9), *AXIN2* (axis inhibition protein 2), *EDA* (ectodysplasin A), and *WNT10A* (Wingless-type MMTV integration site family, member 10A) have been identified to cause non-syndromic tooth agenesis (Vastardis *et al.*, 1996; Stockton *et al.*, 2000; Tao *et al.*, 2006; Bohring *et al.*, 2009), which suggested the significant functions of these genes in early tooth development. However, there is still a remarkable portion of tooth agenesis cases for which the disease-causing mutations cannot be identified in these candidate genes, highlighting the great genetic heterogeneity of this congenital disorder (Bergendal *et al.*, 2011; Arte *et al.*, 2013).

Our group has been working on mutational analysis of human inherited dental defects for decades, and Dr. Jan Hu has been actively recruiting families with these disorders. In this chapter of the thesis, we described families and sporadic individuals with tooth agenesis that we have conducted mutational analyses on to determine the genetic etiology of the disorder. We applied a target gene approach with known candidate genes for non-syndromic tooth agenesis as well as

whole exome sequencing when disease-causing mutations could not be identified in those genes. Through this research, we aimed for not only expanding the mutational spectrum of known candidate genes but also identifying novel genes, which have never been known to be critical for tooth development.

RESULTS

Family 1 and a *PAX9* mutation

The proband was an 8-year-old boy with a total of 10 missing teeth (tooth numbers 2, 5, 12, 13, 15, 18, 24, 25, 29, 31) excluding the third molars for which missing status could not be determined due to his young age (Fig. 2.1). The disease trait of tooth agenesis was inherited from his mother, who had 4 missing teeth (tooth numbers 4, 13, 24, 25) in addition to 4 missing third molars, suggesting a dominant pattern of inheritance. Using a target gene approach, we identified a heterozygous missense mutation (c.43T>A, p.Phe15Ile) in *PAX9* in the proband and his mother, but not the unaffected father. This mutation substituted a highly-conserved amino acid phenylalanine[15] with isoleucine in the paired box domain critical for DNA-binding ability of PAX9. Therefore, we concluded that this *PAX9* mutation causes the tooth agenesis phenotype in this family. The result of this work has been published (Wang *et al.*, 2012).

Family 2 and *WNT10A* mutations

Family 2 was a 4-generation Caucasian family with familial tooth agenesis (Fig. 2.2). The female proband (III:3), a mother of three children, had only one missing tooth (tooth number 4)

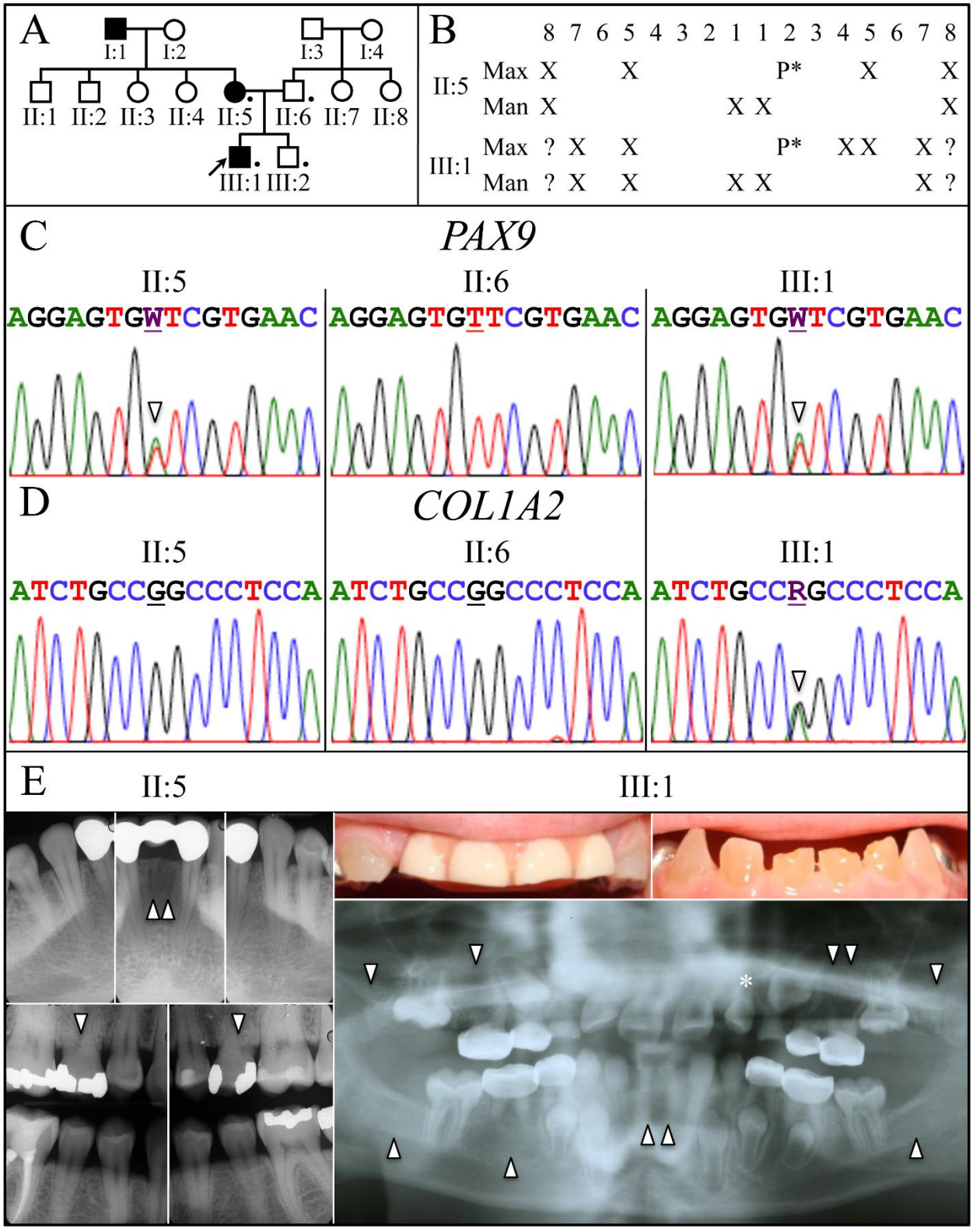


Figure 2.1: Family 1 pedigree, missing teeth, and disease causing mutations.
A: The family pedigree follows the tooth agenesis trait for 3 generations and is consistent with an autosomal-dominant pattern of inheritance. **Key**: A filled icon indicates tooth agenesis. A dot indicates

individual who donated samples. ***B:*** Chart of missing teeth in mother (II:5) and the proband (III:1). ***C:*** DNA sequencing chromatograms show that the affected mother (II:5) and proband (III:1) had a T or A (W) (arrowhead) at position g.5368 (NCBI Ref. Seq. NC_000014.8). This *PAX9* mutation (g.5368T>A; c.43T>A; p.Phe15Ile) caused the tooth agenesis. ***D:*** DNA sequencing chromatograms show both parents (II:5; II:6) had the wild-type G, while the proband had a G or A (R) (arrowhead) at position g.15941 (NCBI Ref. Seq. NG_007405.1). This spontaneous *COL1A2* mutation (g.15941G>A; c.1171G>A; p.Gly391Ser) caused the dentin defects in the proband. ***E:*** Radiographs of the mother (II:5) and proband (III:1) document the missing teeth (arrowheads) and the peg lateral (*) in the proband. Oral photos show the proband's primary anterior teeth show the brownish discoloration and attrition. His maxillary incisors were removed because of severe attrition and a pediatric partial denture was placed. The proband's radiographs show the bulbous crowns with cervical constrictions and thin, narrow roots. This image is reproduced from Figure 1 of reference (Wang SK, Chan HC, Makovey I, Simmer JP, Hu JC (2012). Novel PAX9 and COL1A2 missense mutations causing tooth agenesis and OI/DGI without skeletal abnormalities. *PLoS One* 7:e51533.).

excluding third molars. Her older sister (III:2) was also affected with 2 missing lower second bicuspids. While her husband (III:4) was unaffected and had full set of permanent teeth, two of her three children were missing multiple teeth. The first child (IV:3) had all permanent teeth except 2 missing upper third molars. The second child (IV:4) was a 10-year-old boy who had 8 missing teeth (tooth numbers 2, 4, 13, 15, 18, 24, 25, 31) and 2 peg lateral incisors (small teeth), and the third child, a 7-year-old girl (IV:5), had a total of 15 missing teeth (tooth numbers 2, 4, 7, 10, 12, 13, 15, 18, 20, 23, 24, 25, 26, 29, 31). The pedigree analysis suggested a dominant pattern of disease inheritance and a variation in disease expressivity between generations. In addition to dental phenotype, no other developmental anomalies were observed, except that the mother claimed that the hair growth of her two younger children was slow.

For mutational analysis, we first used a target gene approach but failed to find any potential disease-causing mutations in *MSX1*, *PAX9*, *AXIN2*, or *EDA*, the 4 candidate genes known to cause non-syndromic tooth agenesis at that time. Assuming that the genetic defect might be in a novel gene, we then submitted DNA samples from 4 of the family members (III:2, III:4, IV:4, IV:5) for exome sequencing and tried to find the disease-causing mutation through a

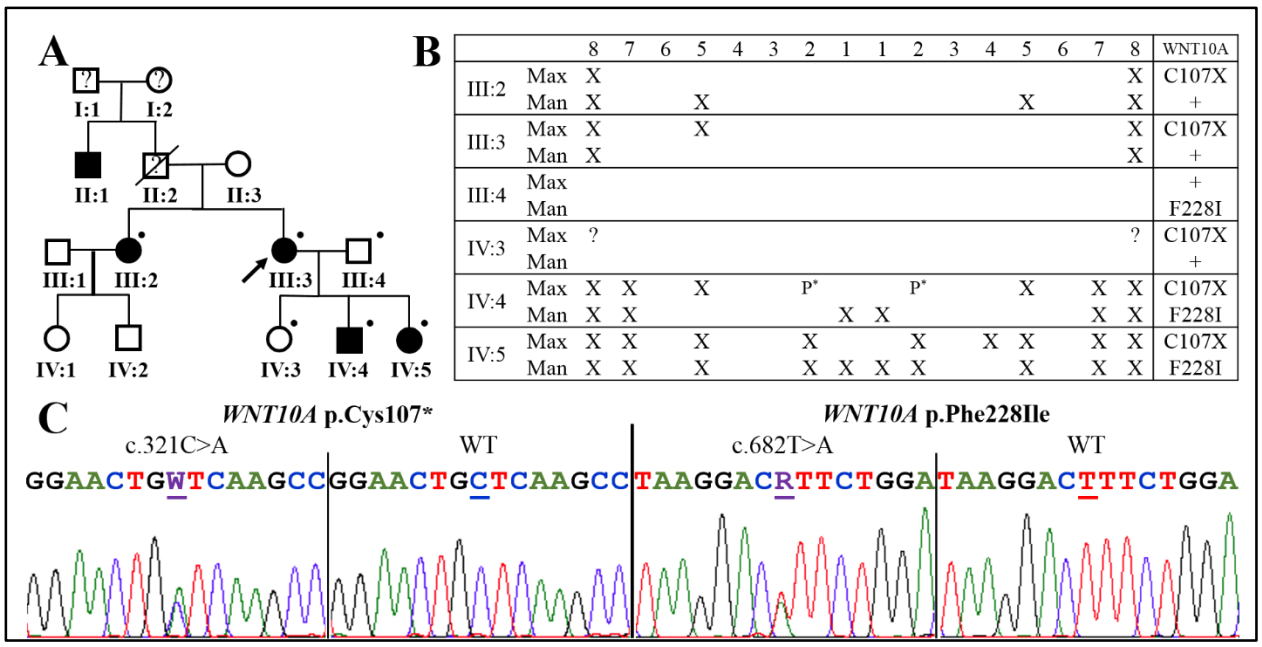


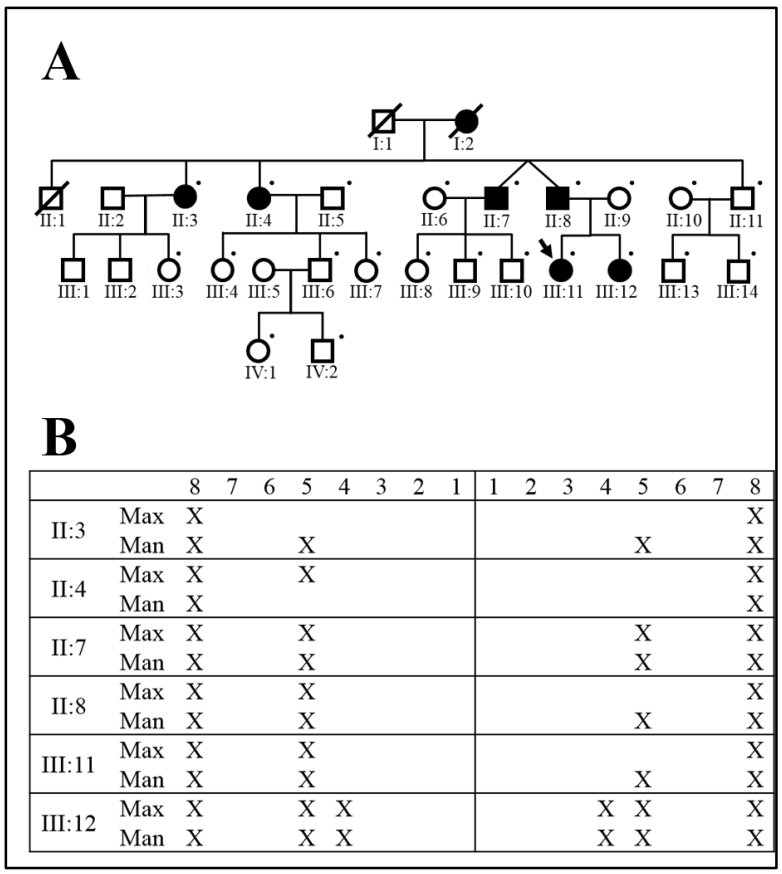
Figure 2.2: Family 2 pedigree, missing teeth, and disease causing mutations.
A: The family pedigree exhibits an autosomal-dominant pattern of inheritance. A filled icon indicates tooth agenesis. A question mark indicates undetermined disease phenotype. A slash indicates deceased individual. A dot indicates individual who donated samples. ***B:*** Chart of missing teeth in all participants. X: missing tooth; P*: peg lateral; ?: undetermined status. ***C:*** DNA sequencing chromatograms show the nonsense *WNT10A* mutation (g.6836C>A, c.321C>A, p.Cys107*) and the missense WNT10A mutation (g.14757T>A, c.682T>A, p.Phe228Ile). The mutation designations are with respect to the *WNT10A* genomic reference sequence NG_012179.1 and cDNA reference sequence NM_025216.2. W means C or A. R means T or A.

genome-wide search. However, at the same time, a report showed that people carrying a heterozygous loss-of-function mutation of *WNT10A*, a gene responsible for a specific type of ectodermal dysplasia when both alleles are defected, had missing teeth, which suggested *WNT10A* as a novel candidate gene for tooth agenesis (Bohring *et al.*, 2009). Therefore, we included *WNT10A* in our analysis and found two *WNT10A* mutations (g.6836C>A, c.321C>A, p.Cys107* and g.14757T>A, c.682T>A, p.Phe228Ile) running in this family (Fig. 2.2). While the proband (III:3) and her sister (III:2) were heterozygotes for the nonsense mutation (c.321C>A, p.Cys107*), both of the two affected children (IV:4, IV:5) were compound heterozygotes for these two mutations (c.321C>A, p.Cys107* and c.682T>A, p.Phe228Ile). More interestingly, the

unaffected father and the first child (III:4, IV:3) carried the missense mutation (c.682T>A, p.Phe228Ile) and the nonsense mutation (c.321C>A, p.Cys107*) respectively. These two *WNT10A* mutations have been reported to cause odonto-onycho-dermal dysplasia (OODD), a rare autosomal-recessive inherited form of ectodermal dysplasia, in a homozygous condition (Bohring *et al.*, 2009). Therefore, the diagnosis of the two affected children, who had more severe tooth agenesis phenotype, should be revised from non-syndromic to syndromic tooth agenesis, although they do not have clinically documented abnormalities of OODD outside of oligodontia.

Family 3

Family 3 is a 4-generation Caucasian family with 6 affected individuals of tooth agenesis (including a pair of identical twins) out of total 22 recruited family members (Fig.2.3). The pattern of missing teeth in the family was of missing only premolars. The proband (III:11) was a 17-year-old girl who had three missing premolars (tooth numbers 4, 20, 29), and her younger sister (III:12) missed all of her permanent premolars (tooth numbers 4, 5, 12, 13, 20, 21, 28, 29) but had a mesiodens (a supernumerary tooth between upper central incisors). The father (II:8) had three missing teeth (tooth numbers 4, 20, 29), while his twin brother (II:7) had four (tooth numbers 4, 13, 20, 29). The two affected aunts (II:3, II:4) had tooth numbers 20, 29 and tooth number 4 absent respectively. According to the mother (II:9), the hairs of her affected husband and two daughters were thin, but there was no problem of heat intolerance. The distribution of phenotypes in family members suggested the tooth agenesis was inherited in a dominant manner; however, there might be penetrance issue of disease phenotype since only 2 out of 11 potentially-affected offspring in the 3rd generation showed tooth agenesis (Fig. 2.3).



		8	7	6	5	4	3	2	1	1	2	3	4	5	6	7	8
II:3	Max	X															X
	Man	X			X									X			X
II:4	Max	X			X												X
	Man	X															X
II:7	Max	X			X									X			X
	Man	X			X									X			X
II:8	Max	X			X												X
	Man	X			X									X			X
III:11	Max	X			X												X
	Man	X			X									X			X
III:12	Max	X			X	X							X	X			X
	Man	X			X	X							X	X			X

Figure 2.3: Family 3 pedigree and missing teeth.
A: The family pedigree exhibits an autosomal-dominant pattern of inheritance with probable incomplete penetrance of disease phenotype. Individuals II:7 and II:8 are identical twins. A filled icon indicates tooth agenesis. A slash indicates deceased individual. A dot indicates individual who donated samples. ***B:*** Chart of missing teeth in all affected individuals. The involved teeth are premolars and third molars. An X indicates a missing tooth. Individual III:12 has a mesiodens between upper central incisors that has been surgically removed.

The pattern of missing teeth in the family (only premolars and third molars) made *MSX1* the most likely candidate gene responsible for the disorder, based upon previous reports of *MSX1* mutations (Vastardis *et al.*, 1996; Kim *et al.*, 2006). However, we failed to identify any potential disease-causing mutations in *MSX1* or *PAX9*. We then sequenced the whole exomes of 8 family members, 4 affected (II:3, II:4, III:11, III:12) and 4 unaffected (II:9, II:11, III:8, III:10), and selected potential disease-causing sequence variations by analyzing and comparing exome sequencing data from different individuals. The analyzed exome sequencing data and the gene lists of non-reference sequence variants after comparison were provided by Dr. Manjusha Pande at the University of Michigan Bioinformatics Core.

Combining the exome data from all the 8 individuals, there were total 3,148 sequence variants different from human reference sequence with a less than 5% minor allele frequency (MAF) in general populations. Among these, there were 487 variants shared by at least 3 affected individuals but no more than 2 unaffected individuals, and there were only 8 variants shared by all the affected but none of the unaffected individuals (Tab. 2.1). They were sequence variants in *MFSD9*, *MUC4*, *KIAA1109*, *HOXC6*, *SLC39A5*, *MYO1A*, *CDC27*, and *MUC16*. Unfortunately, at this stage, we were not able to confidently define the actual disease-causing mutations from these candidate sequence variants.

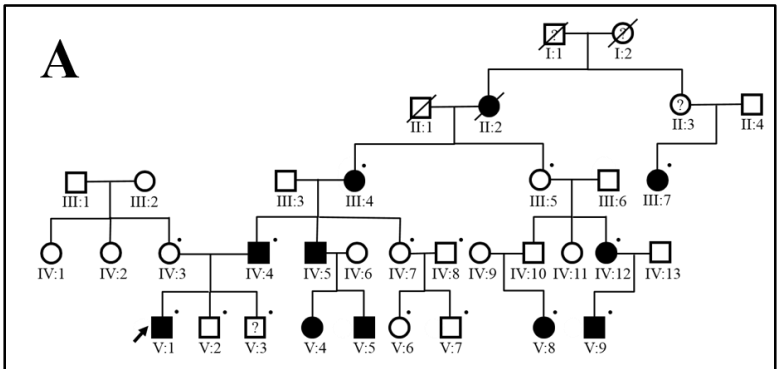
Table 2.1: Sequence variants from exome analysis of Family 3.
Whole exomes of 4 affected (II:3, II:4, III:11, III:12) and 4 unaffected (II:9, II:11, III:8, III:10) individuals were analyzed comparatively. A total of 3,148 sequence variants different from human reference sequence with a less than 5% minor allele frequency (MAF) in general populations was obtained. The listed 8 sequence variants are the ones shared by all the affected but none of the unaffected individuals. This table is derived from a gene list of exome analysis (Oligodontia_variants_3148.xlsx) provided by Dr. Manjusha Pande.

Chromosome	Position	Reference Allele	Sample Allele	Gene Symbol	Protein Variant	Translation Impact	SIFT Function Prediction	dbSNP ID
Chr2	103340233	C	A	***MFSD9***	p.G188V	missense	Damaging	55839709
Chr3	195506306	G	C	***MUC4***	p.L4049V	missense		
Chr4	123108639	T	C	***KIAA1109***	p.I200T	missense	Damaging	
Chr12	54422504	C	T	***HOXC6***	p.R67W	missense	Damaging	144307645
Chr12	56630985	T	C	***SLC39A5***	p.L447P	missense	Damaging	76216511
Chr12	57441459	G	A	***MYO1A***	p.R93*	stop gain		121909305
Chr17	45249316	T	C	***CDC27***	p.Y73C	missense	Tolerated	62077266
Chr19	8999560	T	C	***MUC16***	p.K13539E	missense		77501519

Family 4

Family 4 was a 5-generation Caucasian family of which the male proband (V:1) had only one missing tooth of lower right second premolar (Fig. 2.4). His younger brother (V:2) was unaffected, and the youngest brother (V:3) is too young to determine his dental status. However, the father (IV:4) was more severely affected with missing 4 premolars and one molar (tooth numbers 5, 13, 20, 29, 31). The uncle (IV:5) was reported to have 6 missing teeth, and his two children (V:4, V:5) were missing tooth number 4 and 29 respectively, but they were not successfully recruited to the study. Noticeably, although there were 4 affected individuals on the other side of the family, the phenotype is different from that of the proband's side. Instead of missing premolars, the individuals III:7, IV:12, V:8, and V:9 were reported to have anterior teeth missing (lateral incisors and canines) (Fig. 2.4).

We submitted 6 DNA samples, including 4 affected (III:4, V:1, III:7, V:8) and 2 unaffected (IV:7, III:5) individuals, for whole exome sequencing. The variation in the disease



B

		8	7	6	5	4	3	2	1	1	2	3	4	5	6	7	8
IV:4	Max	X				X								X			X
	Man	X	X		X									X			X
V:1	Max	?															?
	Man	?			X												?
V:8	Max	?						X			X						?
	Man	?															?

Figure 2.4: Family 4 pedigree and missing teeth.
A: The family pedigree exhibits an autosomal-dominant pattern of inheritance. A filled icon indicates tooth agenesis. A question mark indicates undetermined disease phenotype. A slash indicates deceased individual. A dot indicates individual who donated samples. ***B:*** Chart of missing teeth in affected individuals whose radiographs can be obtained and missing teeth confirmed. X: missing tooth; ?: undetermined status.

phenotypes made us suspect that there might be two phenotypic traits and causative mutations running in this family. Based upon this assumption, we analyzed the data from each side of the family separately and listed all the sequence variants sorted by different criteria. The analyzed exome sequencing data and the gene lists of non-reference sequence variants after comparison were provided by Dr. Murim Choi.

On the proband's side (trait of premolar missing), there were 19 sequence variants shared by the two affected individuals (III:4, V:1). Eight of these variants were not present in the unaffected member (IV:7) (Tab. 2.2), and are each potentially responsible for the trait of missing premolars. On the other side of the family (trait of anterior teeth missing), only 7 sequence variants were shared by two affected individuals (III:7, V:8), but all of them were also present in the unaffected member (III:5). Although several sequence variants appeared to be potentially disease-causing, none of them showed strong biological relevancy to tooth development. Therefore, we were unable to confidently define the actual disease-causing mutations from these candidate sequence variants.

Family 5, 6, and 7

The probands of Family 5, 6, and 7 all showed severe form of oligodontia in which multiple anterior and posterior teeth were involved. It is possible that the disease phenotypes of these families share the same genetic etiology. Also, considering the relatively small size of these families, we grouped them together for mutational analysis.

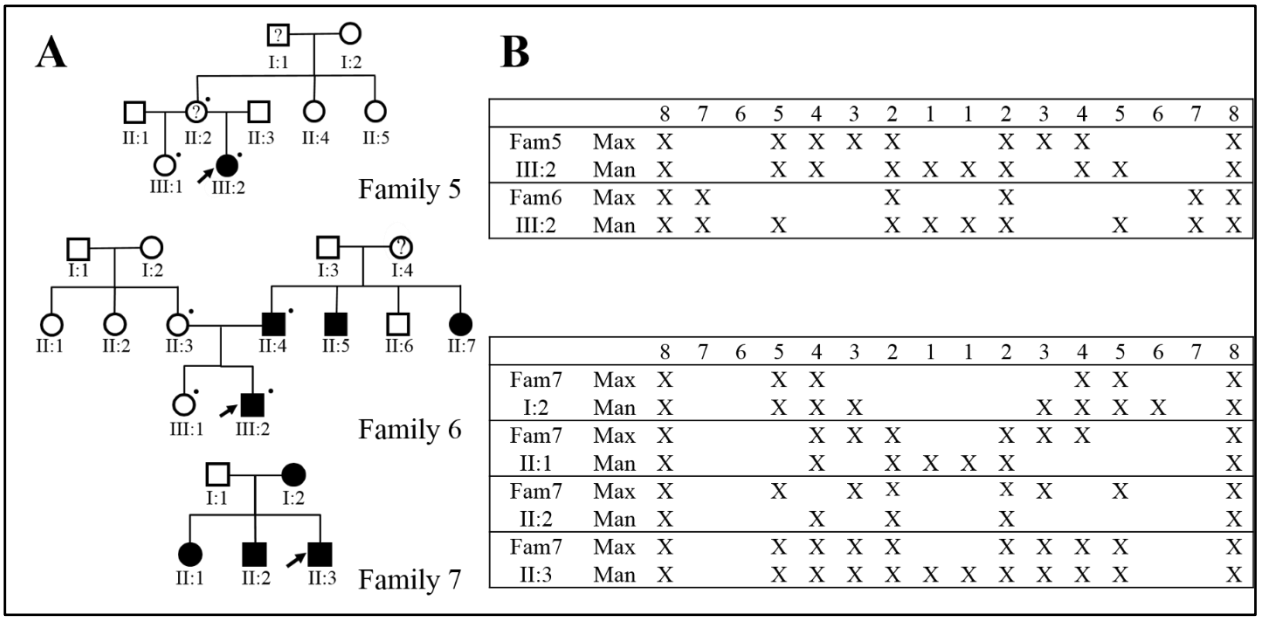
The proband of Family 5 (III:2) was the only family member whose dental record was available and phenotype confirmed. She had a total of 15 missing teeth (tooth numbers 4, 5, 6, 7,

Table 2.2: Sequence variants from exome analysis of Family 4.
Whole exomes of 2 affected (III:4, V:1) and 1 unaffected (IV:7) individuals from the left side of Family 4 pedigree (cases of premolar missing) were analyzed comparatively. The listed 8 sequence variants are the ones shared by the two affected but not the unaffected family members. This table is derived from gene lists of exome analyses (NOVELVAR_n29.xlsx) provided by Dr. Murim Choi.

Chromosome	Position	Base change	Het/Hom	Gene	Status	AA change	PhyloP	SIFT	Polyphen2
chr11	33612954	G>A	Het	***C11orf41***	Missense	E1283K	5.716	0.17	0.999
chr5	118511024	T>A	Het	***DMXL1***	Missense	H2250Q	0.613	0.18	0.051
chr3	52380742	+GG	Het	***DNAH1***	Indel	T637	-2.185		
chr14	102568306	A>G	Het	***HSP90AA1***	Missense	F91S	-1.204	0	0.023
chr16	75675599	C>G	Het	***KARS***	Missense	A29P	5.982	0.1	1.0
chr5	115831951	T>C	Het	***SEMA6A***	Missense	H113R	4.888	0.02	0.524
chr5	171520934	+AGA	Het	***STK10***	Indel	E346	1.583		
chr3	44685396	C>G	Het	***ZNF197***	missense	T925S	1.761	0.08	0.976

10, 11, 12, 20, 21, 23, 24, 25, 26, 28, 29) (Fig. 2.5). Both her parents were reported to have "good teeth," but the actual dental phenotypes could not be confirmed. Since the mode of disease inheritance could not be determined, we considered the proband as a sporadic case when conducting mutational analysis, meaning all the heterozygous and homozygous sequence variants were evaluated.

The proband of Family 6 (III:2) was a boy who had 12 missing teeth (tooth numbers 2, 7, 10, 15, 18, 20, 23, 24, 25, 26, 29, 31) (Fig. 2.5). The disease phenotype seemed to be inherited from the paternal side, since his father was reported to have 6 missing teeth, and two other siblings of the father were also reported to be affected. However, only the proband, his sister, and parents were successfully recruited to the study.



B

		8	7	6	5	4	3	2	1	1	2	3	4	5	6	7	8
Fam5	Max	X			X	X	X	X			X	X	X				X
III:2	Man	X			X	X		X	X	X	X		X	X			X
Fam6	Max	X	X					X			X					X	X
III:2	Man	X	X		X			X	X	X	X			X		X	X

		8	7	6	5	4	3	2	1	1	2	3	4	5	6	7	8
Fam7	Max	X			X	X							X	X			X
I:2	Man	X			X	X	X					X	X	X	X		X
Fam7	Max	X				X	X	X			X	X	X				X
II:1	Man	X				X		X	X	X	X						X
Fam7	Max	X			X		X	X			X	X		X			X
II:2	Man	X				X		X			X						X
Fam7	Max	X			X	X	X	X			X	X	X	X			X
II:3	Man	X			X	X	X	X	X	X	X	X	X	X			X

Figure 2.5: Pedigrees and missing teeth of Family 5, 6, and 7.
A: The Family 5 pedigree shows that the proband may be a sporadic case of tooth agenesis. The Family 6 and 7 pedigrees exhibit an autosomal-dominant pattern of inheritance. A filled icon indicates tooth agenesis. A question mark indicates undetermined disease phenotype. A dot indicates individual who donated samples. All members of Family 7 donated samples. ***B:*** Charts of missing teeth in the probands of Family 5 and 6, and affected individuals of Family 7. All patients are severe cases of oligodontia. While the proband of Family 5 has a total of 15 missing teeth, the proband of Family 6 is missing 12 teeth. In Family 7, the individuals I:2, II:1, II:2, and II:3 are missing 11, 11, 9, and 18 teeth respectively. Interestingly, although multiple teeth are involved, second molars are all present. An X indicates a missing tooth.

Family 7 was a small family of 5 members with 4 of them having severe tooth agenesis (Fig. 2.5). The proband (II:3) had a total of 18 missing permanent teeth (tooth numbers 4, 5, 6, 7, 10, 11, 12, 13, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29) with multiple over-retained primary teeth. His older sister (II:1) and brother (II:2) were also affected with missing 11 (tooth numbers 5, 6, 7, 10, 11, 12, 23, 24, 25, 26, 28) and 9 (tooth numbers 4, 6, 7, 10, 11, 13, 23, 26, 28) permanent teeth respectively. The disease trait seemed inherited from the mother (I:2), since she was missing 11 teeth (tooth numbers 4, 5, 12, 13, 19, 20, 21, 22, 27, 28, 29). Despite that the "unaffected" father (I:1) also had one missing tooth of upper left lateral incisor, the disease phenotype showed a dominant pattern of inheritance. Although the condition of tooth agenesis in

this family was significant, no abnormalities other than dental phenotype from any individual were reported.

For mutational analysis, we first searched for mutations in candidate genes, including *MSX1*, *PAX9*, *AXIN2*, *EDA*, *WNT10A*, *BMP4*, *OSR2*, and *HOXC6*, but no potential disease-causing sequence variations were observed in probands of all three families. We then sequenced and analyzed the exomes from the three probands, listed all the sequence variants for each individual, and inspected them individually. The analyzed exome sequencing data and the gene lists of non-reference sequence variants after comparison were provided by Dr. Murim Choi. Also, under the assumption that there might be two probands sharing the same genetic etiology, we then compared the sequence variants between any two out of the three exomes, and sorted out the candidate genes in which both exomes had sequence variants (Tab. 2.3). However, none of the approaches gave us strong candidates, so at this stage we were unable to confidently define the actual disease-causing mutation for any of the three families.

Table 2.3: Gene lists from exome comparison of Family 5, 6, and 7 probands.
Any 2 out of 3 exomes from different probands were compared. The listed genes are the ones in which the compared two exomes have non-reference sequence variants. Genes marked with red are genes in which all three probands have non-reference sequence variants. This table is derived from gene lists of exome analyses (NOVELVAR_n29.xlsx) provided by Dr. Murim Choi.

Compared probands	Genes in common
Family 5, Family 6	*CD24*, *CFTR*, *MUC4*, *PCNXL2*, *PRSS1*, *SGK1*, *SLC25A5*
Family 5, Family 7	*C8orf73*, *CD24*, *CFTR*, *MUC4*, *PABPC1*, *PRSS1*, *SLC25A5*, *UHRF1*, *VPS36*
Family 6, Family 7	*CD24*, *CFTR*, *MUC16*, *MUC4*, *PCLO*, *PRSS1*, *PRSS2*, *PRSS3*, *RECQL5*, *SLC25A5*

DISCUSSION

WNT10A mutations and non-syndromic tooth agenesis

Human *WNT10A* mutations, in a homozygous condition, were first identified to cause Odontoonychodermal dysplasia (OODD; MIM#257980), a rare autosomal recessive syndrome characterized by dry hair, severe hypodontia, smooth tongue with marked reduction of fungiform and filiform papillae, onychodysplasia, keratoderma, and hyperhidrosis (Adaimy *et al.*, 2007). With significant variation in disease severity, a brother and sister from one of the reported families had oligodontia and sparse body hair and eyebrows as their only manifestations. Interestingly, a subsequent study of *WNT10A* mutations reported that about half of the heterozygous carriers in OODD families showed a phenotype manifestation, including mainly tooth and nail anomalies (Bohring *et al.*, 2009), suggesting that tooth agenesis in patients without other abnormalities might result from a heterozygous *WNT10A* mutation.

With *WNT10A* being established as a candidate gene for non-syndromic tooth agenesis, we identified 2 reported *WNT10A* mutations (c.321C>A, p.Cys107*; c.682T>A, p.Phe228Ile) in a 3-generation Caucasian family (Family 2). Two affected children were compound heterozygotes for both mutations, although they didn't show apparent OODD abnormalities other than severe oligodontia, which made us revise the disease diagnosis. The mother and the aunt, who had 1 and 2 premolars missing respectively, were heterozygous carriers of the nonsense mutation (p.Cys107*), which explained their milder disease phenotypes, compared to those of the two affected children. However, interestingly, the unaffected father, who had full set of permanent teeth, and older sister, who had only two third molars missing, were carriers for the p.Phe228Ile and p.Cys107* mutations respectively, suggesting an incomplete penetrance of the

phenotype with heterozygous mutant *WNT10A* alleles. Genotypically, the disease in this family should be considered "autosomal recessive," although it appeared to be "autosomal dominant" phenotypically, which well explained the more severe disease phenotype in children's generation. The incomplete penetrance of disease phenotype from a mild mutant allele (*WNT10A* p.Phe228Ile in this case) sometimes can be misleading. Therefore, among the candidate genes of non-syndromic tooth agenesis, *WNT10A* might be the one in higher priority, when people are searching for genetic defects in families with variable disease severity between generations.

It has been well recognized that Wnt/β-catenin signaling is critical for tooth development. While ablation of Wnt/β-catenin signaling in mice causes tooth developmental arrest at early stage, constitutive activation of β-catenin leads to formation of supernumerary teeth (Sasaki *et al.*, 2005; Liu *et al.*, 2008). The discovery that *WNT10A* mutations are associated with human tooth agenesis reaffirmed the significance of Wnt signaling pathway in early tooth development.

Recently, many different *WNT10A* mutations were reported to cause non-syndromic tooth agenesis with remarkable variation in disease severity, and it has even been reported that mutations in *WNT10A* are present in more than half of isolated hypodontia cases based on study of 58 Caucasian subjects (van den Boogaard *et al.*, 2012; Song *et al.*, 2014). However, in many cases, the researchers performed target gene approaches with selected candidate genes and arbitrarily assigned the disease-causing mutations without confirming disease segregation, especially for sporadic cases, which makes the genetic causality suspicious. For example, some have reported that individuals with a heterozygous *WNT10A* p.Phe228Ile mutation have more than 10 missing teeth, with the maximum of 18 in the literature (van den Boogaard *et al.*, 2012), while others, including us, found that this specific mutation causes mild tooth agenesis or sometimes no missing teeth. Nevertheless, it is unlikely that a sequence variant can cause such a

wide spectrum of disease severity from no phenotype to severe phenotype, suggesting that the reported variant might not be the actual genetic defect responsible for the disorder. People may argue that there might be some genetic modifiers that affect the phenotypic expressivity of this variant. However, before these genetic modifiers are actually identified and confirmed, one needs to be cautious about determining disease-causing mutations in the cases whose phenotypes are much more extreme than those of reported cases. Therefore, additional investigations need to be conducted before one can conclude that *WNT10A* mutations are the major genetic cause of non-syndromic tooth agenesis (van den Boogaard *et al.*, 2012; Song *et al.*, 2014).

Variable genetic penetrance and expressivity in familial tooth agenesis

In our Family 3, the family pedigree indicated a dominant manner of disease inheritance. However, among total 16 individuals who had 50% chance to be affected, there were only 4 actually showed the disease phenotype (Fig. 2.3). This disobedience to Mendelian proportion for dominant inheritance highly suggested an incomplete phenotypic penetrance of the genetic defect in this family. This lack of penetrance is commonly seen in many of the human genetic diseases, including familial tooth agenesis, which makes it more difficult to identify the genetic etiology in some families. In the case of Family 3, the family size was relatively large, which theoretically increased the genetic power to identify the disease-causing mutation. However, the incomplete penetrance offset this advantage. Although we analyzed the exomes from 8 family members and selected certain candidate sequence variants, we were still not able to determine the genetic etiology, especially when little was known about the functional impacts of those variants. Therefore, in order to confidently define the disease-causing mutation in Family 3, we

might need more families with similar phenotypes and probably defects in the same causative gene to support the genetic causality.

Another interesting characteristic of familial tooth agenesis is the variable phenotypic expressivity (disease severity) within the family. Sometimes the variation could be very significant between individuals carrying the same genetic defect. For example, in Family 1, despite that the proband and his mother had the same *PAX9* p.Phe15Ile mutation, the mother had only 4 missing teeth, while the prohand had 10. In Genetics, this phenomenon has been explained by a hypothetical existence of "genetic modifiers," meaning that some unidentified genetic functional elements might affect or "modify" the observed phenotypes. Nevertheless, in spite of the underlying mechanism, the variation in expressivity sometimes could be misleading in mutational analysis. For instance, in Family 2, based upon disease phenotype, there seemed to be a significant expressivity variation between generation III and IV. However, the different disease severity actually resulted from carrying one or two defective alleles. Therefore, in such cases, carefully characterizing the disease phenotype and confirming the segregation between genotype and phenotype are particularly important to determine the genetic causality.

Phenotypic and genotypic heterogeneity in familial tooth agenesis

It has been observed that the pattern of missing teeth in familial tooth agenesis is, to some extent, associated with its genetic etiology. While second bicuspids and third molars are frequently involved in *MSX1*-associated tooth agenesis, *PAX9* mutations usually lead to tooth agenesis of second bicuspids, second molars, third molars, and some central incisors (Kim *et al.*, 2006). People with *EDA* mutations tend to have multiple anterior teeth missing but first molar

preserved (Han *et al.*, 2008). In *AXIN2*-associated tooth agenesis, the affected individuals usually have severe oligodontia, with multiple anterior and posterior teeth missing (Lammi *et al.*, 2004). Moreover, it was recently reported that *WNT10A* mutations cause tooth agenesis with lateral incisors and second premolars being frequently involved (Kantaputra and Sripathomsawat, 2011; Song *et al.*, 2014), although some reported that there is no specific pattern of missing teeth in *WNT10A*-associated tooth agenesis (Plaisancié et al., 2013; van den Boogaard *et al.*, 2012). This genotype-phenotype correlation is sometimes very useful for prioritizing the candidate genes in mutational analysis. For example, in our Family 1, the proband had a typical pattern of missing teeth associated with *PAX9* mutations, and we eventually identified the *PAX9* p.Phe15Ile mutation responsible for the disease. However, the phenotype is not always predictive for the genetic etiology, and given the potential for oligogenic effects, the mutational analyses may provide an incomplete picture of the genetic etiology. In Family 3, the pattern of missing teeth in affected individuals (missing of premolars and third molars) suggested *MSX1* might be the responsible gene, but we failed to find any potential disease-causing mutation in *MSX1*. However, this inconsistency suggested that the actual gene involved in the disease of this family might function in the related genetic pathway with *MSX1*, so when the gene is defected, the affected individual shows similar phenotype to that of *MSX1*-associated tooth agenesis. In this sense, families with similar phenotypes but different genetic causes may provide a valuable *in vivo* tool to study genetic pathways and gene-gene interactions. However, it is also possible that the mutation is at the regulatory region of the gene, which cannot be easily determined.

As mentioned previously, although mutations in *MSX1*, *PAX9*, *AXIN2*, *EDA*, and *WNT10A* have been shown to cause non-syndromic tooth agenesis, there are still many cases of which the genetic mutations cannot be found in these candidate genes, indicating high genetic

heterogeneity of this disorder (Bergendal *et al.*, 2011; Arte *et al.*, 2013). In this thesis, the genetic etiology of tooth agenesis in 5 out of 7 families still needs to be determined. Recently, whole exome sequencing has been demonstrated to be a powerful tool in human mutational analysis (Bamshad et al., 2011; Goh and Choi, 2012). In many cases, by combining the data from unrelated affected individuals with the same disease, the genetic power is significantly increased, and the responsible gene is readily identified (Rabbani *et al.*, 2012). However, for diseases with high genetic heterogeneity, this approach may not work, since similar phenotypes can result from distinct genetic causes. We combined and analyzed the exome data from the probands of Family 5, 6, and 7, but still failed to identify the causative mutation in any of the families. Sometimes, a larger family size can help to overcome the problem from genetic heterogeneity of the disease. With more affected and unaffected family members, the genetic linkage may provide significant power to locate the disease locus in the genome and help to identify the mutation (Bailey-Wilson and Wilson, 2011). For example, linkage analyses of two large families with tooth agenesis have mapped new disease locus to chromosome 16q12.1 and 10q11.2, although no candidate genes have yet been identified at these loci (Ahmad *et al.*, 1998; Liu *et al.*, 2001). Unfortunately, the sizes of most of the families we described here were relatively small, which makes linkage analysis impossible. Therefore, in order to actually identify the responsible genes in these families with unknown genetic defects, we may need supportive evidence from mouse models, linkage data of other large families, or more individual cases with similar disease phenotypes.

Prospects in mutational analysis of familial tooth agenesis

Discerning the genetic etiology of human tooth agenesis significantly gain our understanding of tooth development. All the candidate genes identified so far and their roles in

tooth formation have been extensively studied (Nieminen, 2009). However, as discussed above, due to the genotypic and phenotypic heterogeneity of the disease, the genetic etiology of a significant number of cases still waits to be determined. In the future, for large-sized families, we should take advantage of the genetic linkage power to define disease locus and further analyze with exome sequencing data to pinpoint the disease-causing mutation (Bailey-Wilson and Wilson, 2011; Wijsman, 2012). For relatively small-sized families, we need to recruit multiple families with similar phenotypes and collectively analyze their exome data at a "genetic pathway level," meaning that mutations in genes involved in similar pathways need to be considered to probably cause similar disease phenotypes (Gilissen *et al.*, 2012; Li *et al.*, 2012). For sporadic cases, the genetic causes should not be determined until the whole exome data has been scrutinized, especially for those who have mild phenotypes (missing only one or two teeth), since genotype-phenotype segregation cannot be discerned to confirm the causality in these cases. Admittedly, the current mutational analysis focuses on the protein-coding region of the genome. However, one can fully appreciate that the genetic defects may locate at non-coding regulatory regions of the genome in certain cases. Therefore, whole genome sequencing may be the alternative approach to address this issue (Cooper and Shendure, 2011; Lyon and Wang, 2012). However, at this stage, we have very limited ability to interpret the functional impact of sequence variants found in regulatory regions. In other words, many efforts still need to be made to prove the utility of whole genome sequencing in human mutational analysis (Lyon and Wang, 2012).

Furthermore, since familial tooth agenesis is sometimes considered as a multifactorial genetic disease, people have tried to identify disease-associated loci through case-control studies with multiple candidate genes or genome wide association studies (GWAS) (Vieira *et al.*, 2007;

Vieira *et al.*, 2008; Haga *et al.*, 2013; Liu *et al.*, 2013). For example, an SNP, rs1469622, at an intron of *THSD7B* (thrombospondin, type I, domain containing 7B) was shown to be associated with third molar agenesis in Japanese and Korean populations (Haga *et al.*, 2013). These association studies with large sample sizes can avoid some problems that traditional mutational analysis cannot conquer, such as genotypic and phenotypic heterogeneity of the disease, and may shed light on finding novel genes involved in tooth development. However, association studies only provide "genetic association," the validation of disease causality still depends upon traditional mutational analysis and functional studies of specific genes.

MATERIALS AND METHODS

Family recruitment and ethics statement

The human study protocol and subject consents were reviewed and approved by the Institutional Review Boards (IRB) at the University of Michigan and the University of Texas Health Science Center at San Antonio. Study participants signed appropriate written consents after an explanation of their contents and after their questions about the study were answered. Any minors age 8 or older signed a written assent form after their parent completed a written parental consent for participation of the minor. All the IRB processes were issued by Dr. Jan Hu.

Phenotypic data collection and family pedigree construction

During subject recruitment, we collected dental records, mainly radiographs, and took medical and dental history from each participant. When dental records were not available, determination of disease phenotype depended upon communication with the contact family

member, which compromised the validity of phenotyping. For each family, the genetic pedigree was then constructed based up the phenotypic data. All the family recruitments, including collecting subjects' samples and clinical data were conducted by Dr. Jan Hu.

DNA extraction from blood and saliva samples

Peripheral whole blood (5 cc) or saliva (2 cc) was obtained from recruited individuals of each family. Genomic DNA was isolated from blood or saliva with the QIAamp DNA Blood Maxi Kit (51194; Qiagen; Valencia, CA, U.S.A.) and the Saliva DNA Collection, Preservation and Isolation Kit (RU35700; Norgen Biotek Corporation; Thorold, Canada) respectively. The quality of the extracted DNA samples were determined by spectrophotometry at OD_{260} and OD_{280} and prepared for subsequent analyses. DNA extractions of blood/saliva samples from subjects described here were performed by Dr. Hui-Chen Chan, Rachel Milkovich, Soumya Pal, Bryan Reid, and myself.

Mutational analysis (target gene approach)

For each candidate gene of interest, the coding exons and intron junctions of the gene were amplified by polymerase chain reaction (PCR) with specific primer sets (Tab. 2.4). The amplification products were purified and characterized by direct DNA sequencing (Sanger sequencing) at the University of Michigan DNA Sequencing Core (Ann Arbor, MI, U.S.A.). The sequencing data was then examined by comparing it to the human reference sequence, and sequence variants called and evaluated. Target gene analyses for subjects described here were conducted by Rachel Milkovich, Soumya Pal, Bryan Reid, and myself.

Table 2.4: Primers and PCR conditions used for Sanger sequencing.
Primer sets used for amplification of exons and exon-intron junctions of target genes (*MSX1*, *PAX9*, *AXIN2*, *EDA*, *WNT10A*, *OSR2*, *BMP4*, and *HOXC6*) are listed. For large exons, more than one primer set may be applied. 5% or 10% Dimethyl sulfoxide (DMSO) is used in PCR reactions with high GC-content amplicons. PCR amplifications were done using the Platinum® PCR Supermix (11306-016; Invitrogen™ by Life Technologies; Grand Island, NY, U.S.A.). The reactions had a 5 min denaturation at 94 ˚C, followed by 35 cycles each with denaturation at 94 ˚C for 30 sec, primer annealing at 56-58 ˚C for 60 sec, and product extension at 72 ˚C for 90 sec. In the final cycle the 72 ˚C extension was for 7 min. PCR amplification products were purified by QIAquick PCR Purification Kit and protocol (28106; Qiagen; Valencia, CA, U.S.A.).

MSX1					
Primers	Oligonucleotide sequence	Size (bp)	Primers	Oligonucleotide sequence	Size (bp)
Ex1F	CTGGCCTCGCCTTATTAGC	766*	Ex2F	ACTTGGCGGCACTCAATATC	698
Ex1R	GCCTGGGTTCTGGCTACTC		Ex2R	CAGGGAGCAAAGAGGTGAAA	
Annealing temperature (Ta): 58°C; *: 10% DMSO is required.					
PAX9					
Primers	Oligonucleotide sequence	Size (bp)	Primers	Oligonucleotide sequence	Size (bp)
Ex1F	GGGTGGGGAACAATTACTGA	590	Ex2F	TTCTGAGATGTCCATCGTTCC	741
Ex1R	GCGGCTAAAAGGAGCAGTC		Ex2R	CTGACCCTTAGCGTGTTCCT	
Ex3F	GGGGACAGCCCCAGTAGTTA	857	Ex4F	TGGAAAGGCCTACTCTGAGG	499
Ex3R	GGGAAAGACAGTGTCCCTGA		Ex4R	GAAGGATCTGGCTCGTAGCA	
Ex5F	TCAGAGCATTGCTGGCTTAC	481			
Ex5R	CTTTCAAGGCAGAAGGGTTG				
Annealing temperature (Ta): 58°C					
AXIN2					
Primers	Oligonucleotide sequence	Size (bp)	Primers	Oligonucleotide sequence	Size (bp)
Ex1F	CAGGAGGAAGAGGCTGAATG	721	Ex2-1F	AGTGTGCAGGGAGCTCAGAT	667
Ex1R	CTGCGCTGTGGATTTAACTG		Ex2-1R	ATTGCAGGCAAACCAGAAGT	
Ex2-2F	CATCTCCGGATTCCCCTCT	684	Ex3F	GGCTGCCTCTGGAATACTCTC	549
Ex2-2R	TCCACCCATCCACCATACTT		Ex3R	ATACTCCCCTCCCACCAAAC	
Ex4F	AGCACCGATGGTATCTGGAG	457	Ex5F	AAGAACAAGCAGGGCCTCTC	600
Ex4R	TCACATCACTGTGCTCACCA		Ex5R	ATCCACACGCATATGCACAC	
Ex6F	CTTCTGCTTCCTGGGTCACT	673	Ex7F	GCCGCATTACAGGCATTTAG	567
Ex6R	CTGCCGCCCTCTTAGAAACT		Ex7R	ACATGAACAGGGGTCAGTGC	
Ex8F	GGAATGGCTGTTTTTGCAGT	603	Ex9F	GTCTTGGTTGGGTCTCCGTA	529
Ex8R	TTCTCATGGGAGGGTTTGAG		Ex9R	AGTTGAATGGGGGCAACTTT	
Ex10F	ATTGCAGCCCTAGTGTTTGG	544	Ex11F	AGGTGCTGGGTTGAAGATTG	533
Ex10R	GTTCACCTGGTGGAAAGAGC		Ex11R	CACTGGCCGATTCTTCCTTA	
Annealing temperature (Ta): 57°C					
EDA					
Primers	Oligonucleotide sequence	Size (bp)	Primers	Oligonucleotide sequence	Size (bp)
Ex1F	ATTCTTAGCCTCCCCCTCCT	1046*	Ex2F	CCCACCCATCATATCCTGTC	626
Ex1R	TGGTCCTGCCCTCTAAATTG		Ex2R	TGGTCCTCTACAGGCAAGGT	
Ex3F	TACAGTGGAGGGGAAGATGG	415	Ex4F	CGCATGACTCTTCAACCTCA	618
Ex3R	GGCTGGTTTTGAATTCCTCA		Ex4R	AAAAGAAGGGCAGGGAGAAG	
Ex5F	TCACCCGAAGTCAGGAGTTT	826	Ex6F	GGGTGCACTCTGACTCTTCC	333

Ex5R	TGGAGCTAGATGCTGGGAAT		Ex6R	GCTGTGAGTGAAAACCGTCA	
Ex7F	ACAGCTGCACAGTGCTTGAC	549	Ex8F	GGCCCCATAACAACAAAGAA	585
Ex7R	GGCATGATGGAGCAAAGAAT		Ex8R	GCAGGAAGTTAGCCATTGGA	
Ex9F	ATGCCTGTCACCTGTCCTTT	640			
Ex9R	GGCTCCATCAACTGTCCTGT				
Annealing temperature (Ta): 56°C; *: 5% DMSO is required.					
WNT10A					
Primers	Oligonucleotide sequence	Size (bp)	Primers	Oligonucleotide sequence	Size (bp)
Ex1F	CCTCGCGGTAACACATATCC	930*	Ex2F	GGCAGGATGATTGTGAGGAG	706
Ex1R	TCTACCCCAGCAAGAGCATC		Ex2R	TGACCCAGGAGTCCAGTTCT	
Ex3F	TTCCTTGTGCCAGACTCTCC	593	Ex4F	GCGTTTGCCTCTGTATAATGG	714*
Ex3R	CGTGGTCCTCAGAAGAGAGG		Ex4R	CCTCTTCCCAAGAGCCAAG	
Annealing temperature (Ta): 57°C; *: 5% DMSO is required.					
OSR2					
Primers	Oligonucleotide sequence	Size (bp)	Primers	Oligonucleotide sequence	Size (bp)
Ex1F	AAACCTCGAGCAAACCTGTG	633*	Ex2-1F	ACGGTCTCTCCCTCTCTTCC	560
Ex1R	CTTACACCCCCAAACCCTTC		Ex2-1R	TTGCTCAGGTCTCCCATCTT	
Ex2-2F	GTATCCCAATGTGCACGAGA	662	Ex3F	TCACCATGGGGCAAAGTTAT	284
Ex2-2R	CTGGGGGAAAGAAAGGGTAG		Ex3R	GTGTTGCTTTTCCCCACCTA	
Ex4F	GTGTAATCGGGCCTCTGTGT	556			
Ex4R	TTCTCAGTCCCCCAGAAGAA				
Annealing temperature (Ta): 57°C; *: 5% DMSO is required.					
BMP4					
Primers	Oligonucleotide sequence	Size (bp)	Primers	Oligonucleotide sequence	Size (bp)
Ex1F	TCCCATGGGTATTTTTGGAA	686	Ex2F	AGGGGCTGGAAGAAAAACAG	506
Ex1R	ACAGCCTGTGACCAGCTTCT		Ex2R	GCTCTCCCAGACAAGTTGGA	
Ex3F	CTTACTTTCAGGCCGTGCAT	687	Ex4-1F	TTTTCCCCCAGTAGGTTTCC	727
Ex3R	GGACTGGGGCTTTGATGTAA		Ex4-1R	CACATCGCTGAAGTCCACAT	
Ex4-2F	AGCAGCCAAACTATGGGCTA	640			
Ex4-2R	AGTGTGTGGGTGAGTGGATG				
Annealing temperature (Ta): 57°C					
HOXC6					
Primers	Oligonucleotide sequence	Size (bp)	Primers	Oligonucleotide sequence	Size (bp)
Ex1F	CACGCTACACAAATGCATCG	789	Ex2F	CAACAGAAGCAGAAGCGATTT	662
Ex1R	CCCCTCTTCTCCCTAACCAA		Ex2R	TTCTAGGGAAGCCGGTCATA	
Ex3F	CCTAAGGAGGCTGTGAGCTG	569			
Ex3R	GGTCCACGTTTGACTCCCTA				
Annealing temperature (Ta): 57°C					

Mutational analysis (whole exome sequencing)

Whole exome sequencing was conducted at the University of Michigan DNA Sequencing Core (Ann Arbor, MI, U.S.A.) for Family 2 and 3 and Yale Center for Genome Analysis (West

Haven, CT, U.S.A.) for Family 4, 5, 6, and 7. At the University of Michigan DNA Sequencing Core, genomic DNAs (3μg) from subjects were assessed of their quality and quantity and subjected to whole exome sequencing, which was conducted by using Illumina TruSeq Exome Enrichmemt system and HiSeq 2000 platform at 75 base paired-end sequencing. Sequence output was inspected and aligned against human reference genome hg19, and variants were filtered and annotated using Ingenuity Variant Analysis tool by the University of Michigan Bioinformatics Core. At the Yale Center for Genome Analysis, the exome sequencing and subsequent analysis were modified from a previous report (Choi *et al.*, 2009). Briefly, the genomic DNA was captured with NimbleGen v2.0 exome capture reagent (Roche/NimbleGen Incorporation; Madison, WI, U.S.A.) and sequenced with Illumina HiSeq 2000 for 75 base paired-end reads. Reads were aligned to human reference genome hg19 using ELAND v2. Single nucleotide variants and short insertions and deletions (indels) were called using SAMtools. The called variants were annotated using an in-house script. Next, the annotated results were first inspected to search for potential disease-causing sequence variations in the known candidate genes of syndromic and non-syndromic tooth agenesis. In families with multiple exomes being sequenced, the data from each were compared, and the sequence variations were further filtered based upon phenotype-genotype segregation. In some circumstances, exome data from unrelated affected individuals were compared and search for sequence variations of the same gene. Exome data analyses and generation of gene lists of non-reference sequence variants were conducted by Dr. Murim Choi. Eventually, potentially disease-causing sequence variations were confirmed in the probands and all participating family members by Sanger sequencing.

REFERENCES

Adaimy L, Chouery E, Megarbane H, Mroueh S, Delague V, Nicolas E, Belguith H, de Mazancourt P, Megarbane A (2007). Mutation in WNT10A is associated with an autosomal recessive ectodermal dysplasia: the odonto-onycho-dermal dysplasia. *Am J Hum Genet* 81:821-8.

Ahmad W, Brancolini V, ul Faiyaz MF, Lam H, ul Haque S, Haider M, Maimon A, Aita VM, Owen J, Brown D, Zegarelli DJ, Ahmad M, Ott J, Christiano AM (1998). A locus for autosomal recessive hypodontia with associated dental anomalies maps to chromosome 16q12.1. *Am J Hum Genet* 62:987-91.

Arte S, Parmanen S, Pirinen S, Alaluusua S, Nieminen P (2013). Candidate gene analysis of tooth agenesis identifies novel mutations in six genes and suggests significant role for WNT and EDA signaling and allele combinations. *PLoS One* 8:e73705.

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12:745-55.

Bergendal B, Klar J, Stecksén-Blicks C, Norderyd J, Dahl N (2011). Isolated oligodontia associated with mutations in EDARADD, AXIN2, MSX1, and PAX9 genes. *Am J Med Genet A* 155A:1616-22.

Bohring A, Stamm T, Spaich C, Haase C, Spree K, Hehr U, Hoffmann M, Ledig S, Sel S, Wieacker P, Röpke A (2009). WNT10A mutations are a frequent cause of a broad spectrum of ectodermal dysplasias with sex-biased manifestation pattern in heterozygotes. *Am J Hum Genet* 85:97-105.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 106:19096-101.

Cobourne MT, Sharpe PT (2013). Diseases of the tooth: the genetic and molecular basis of inherited anomalies affecting the dentition. *Wiley Interdiscip Rev Dev Biol* 2:183-212.

Cooper GM, Shendure J (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* 12:628-40.

De Coster PJ, Marks LA, Martens LC, Huysseune A (2009). Dental agenesis: genetic and clinical perspectives. *J Oral Pathol Med* 38:1-17.

Gilissen C, Hoischen A, Brunner HG, Veltman JA (2012). Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 20:490-7.

Goh G, Choi M (2012). Application of whole exome sequencing to identify disease-causing variants in inherited human diseases. *Genomics Inform* 10:214-9.

Haga S, Nakaoka H, Yamaguchi T, Yamamoto K, Kim YI, Samoto H, Ohno T, Katayama K, Ishida H, Park SB, Kimura R, Maki K, Inoue I (2013). A genome-wide association study of third molar agenesis in Japanese and Korean populations. *J Hum Genet* 58:799-803.

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. *Eur J Med Genet* 51:536-46.

Kantaputra P, Sripathomsawat W (2011). WNT10A and isolated hypodontia. *Am J Med Genet A* 155A:1119-22.

Kim JW, Simmer JP, Lin BP, Hu JC (2006). Novel MSX1 frameshift causes autosomal-dominant oligodontia. *J Dent Res* 85:267-71.

Lammi L, Arte S, Somer M, Jarvinen H, Lahermo P, Thesleff I, Pirinen S, Nieminen P (2004). Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer. *Am J Hum Genet* 74:1043-50.

Li MX, Gui HS, Kwan JS, Bao SY, Sham PC (2012). A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res* 40:e53.

Liu W, Wang H, Zhao S, Zhao W, Bai S, Zhao Y, Xu S, Wu C, Huang W, Chen Z, Feng G, He L (2001). The novel gene locus for agenesis of permanent teeth (He-Zhao deficiency) maps to chromosome 10q11.2. *J Dent Res* 80:1716-20.

Liu F, Chu EY, Watt B, Zhang Y, Gallant NM, Andl T, Yang SH, Lu MM, Piccolo S, Schmidt-Ullrich R, Taketo MM, Morrisey EE, Atit R, Dlugosz AA, Millar SE (2008). Wnt/beta-catenin signaling directs multiple stages of tooth morphogenesis. *Dev Biol* 313:210-24.

Liu H, Han D, Wong S, Nan X, Zhao H, Feng H (2013). rs929387 of GLI3 Is Involved in Tooth Agenesis in Chinese Han Population. *PLoS One* 8:e80860.

Lyon GJ, Wang K (2012). Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress. *Genome Med* 4:58.

Nanci A (2008). Development of the tooth and its supporting tissues. In: Ten Cate's Oral Histology Development, Structure, and Function. Nanci A editor. St. Louis: Mosby, pp.79-107.

Nieminen P (2009). Genetic basis of tooth agenesis. *J Exp Zool B Mol Dev Evol* 312B:320-42.

Parkin N, Elcock C, Smith RN, Griffin RC, Brook AH (2009). The aetiology of hypodontia: the prevalence, severity and location of hypodontia within families. *Arch Oral Biol* 54 Suppl 1:S52-6.

Plaisancié J, Bailleul-Forestier I, Gaston V, Vaysse F, Lacombe D, Holder-Espinasse M, Abramowicz M, Coubes C, Plessis G, Faivre L, Demeer B, Vincent-Delorme C, Dollfus H, Sigaudy S, Guillén-Navarro E, Verloes A, Jonveaux P, Martin-Coignard D, Colin E, Bieth E,

Calvas P, Chassaing N (2013). Mutations in WNT10A are frequently involved in oligodontia associated with minor signs of ectodermal dysplasia. *Am J Med Genet A* 161A:671-8.

Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I (2012). Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet* 57:621-32.

Sasaki T, Ito Y, Xu X, Han J, Bringas P Jr, Maeda T, Slavkin HC, Grosschedl R, Chai Y (2005). LEF1 is a critical epithelial survival factor during tooth morphogenesis. *Dev Biol* 278:130-43.

Song S, Zhao R, He H, Zhang J, Feng H, Lin L (2014). WNT10A variants are associated with non-syndromic tooth agenesis in the general population. *Hum Genet* 133:117-24.

Stockton DW, Das P, Goldenberg M, D'Souza RN, Patel PI (2000). Mutation of PAX9 is associated with oligodontia. *Nat Genet* 24:18-9.

Tao R, Jin B, Guo SZ, Qing W, Feng GY, Brooks DG, Liu L, Xu J, Li T, Yan Y, He L (2006). A novel missense mutation of the EDA gene in a Mongolian family with congenital hypodontia. *J Hum Genet* 51:498-502.

Thesleff I (2013). Current understanding of the process of tooth formation: transfer from the laboratory to the clinic. *Aust Dent J* [Epub ahead of print]

Tucker A, Sharpe P (2004). The cutting-edge of mammalian development; how the embryo makes teeth. *Nat Rev Genet* 5:499-508.

van den Boogaard MJ, Créton M, Bronkhorst Y, van der Hout A, Hennekam E, Lindhout D, Cune M, Ploos van Amstel HK (2012). Mutations in WNT10A are present in more than half of isolated hypodontia cases. *J Med Genet* 49:327-31.

Vastardis H, Karimbux N, Guthua SW, Seidman JG, Seidman CE (1996). A human MSX1 homeodomain missense mutation causes selective tooth agenesis. *Nat Genet* 13:417-21.

Vieira AR, Modesto A, Meira R, Barbosa AR, Lidral AC, Murray JC (2007). Interferon regulatory factor 6 (IRF6) and fibroblast growth factor receptor 1 (FGFR1) contribute to human tooth agenesis. *Am J Med Genet A* 143:538-45.

Vieira AR, Seymen F, Patir A, Menezes R (2008). Evidence of linkage disequilibrium between polymorphisms at the IRF6 locus and isolate tooth agenesis, in a Turkish population. *Arch Oral Biol* 53:780-4.

Wang SK, Chan HC, Makovey I, Simmer JP, Hu JC (2012). Novel PAX9 and COL1A2 missense mutations causing tooth agenesis and OI/DGI without skeletal abnormalities. *PLoS One* 7:e51533.

Wijsman EM (2012). The role of large pedigrees in an era of high-throughput sequencing. *Hum Genet* 131:1555-63.

CHAPTER 3

MUTATIONAL ANALYSIS – AMELOGENESIS IMPERFECTA

ABSTRACT

Amelogenesis imperfecta (AI) is a collection of genetic disorders featured by developmental enamel defects. While isolated AI describes the enamel malformations in the absence of non-dental phenotypes, syndromic AI refers to the enamel defects manifested in syndromes. Patients afflicted with AI suffer from esthetic and functional burdens and have compromised quality of life. Discerning the genetic etiology of AI not only will improve our understanding of dental enamel formation but may also provide fundamental information that will impact therapeutic strategies for this anomaly. With decades of efforts, many genes have been identified to be associated with different types of AI and shown to play significant roles in different stages of amelogenesis. However, there is still a significant number of AI cases of which the genetic causes cannot be identified in these candidate genes. In this thesis, we described 12 AI kindreds and performed mutational analyses to identify the genetic etiology for their enamel defects. We successfully identified novel mutations in *MMP20*, *KLK4*, *FAM83H*, *FAM20A*, *SLC24A4*, and *STIM1*. Particularly, we demonstrated that mutations in *FAM20A* can cause Enamel-Renal Syndrome (ERS), in addition to Amelogenesis Imperfecta Gingival Fibromatosis Syndrome (AIGFS), and suggested that these two disorders may be the same

disease entity with variable presentation of renal calcifications. We also reported, for the first time, that human Integrin beta 6 (*ITGB6*) mutations cause generalized hypoplastic AI, which demonstrates that cell-matrix interaction and integrin signaling are critical for enamel formation. However, the failure to identify genetic defects in many AI families our lab recruited but not described in this thesis reaffirmed the high genetic heterogeneity of AI and revealed our current inadequate knowledge about enamel formation. In the future, we need to recruit more AI families with relatively similar phenotypes and collectively analyze their exome data using a "genetic pathway" approach to help identify AI candidate genes, meaning that mutations in genes involved in similar pathways need to be considered as potential causes resulting in similar disease phenotypes. Furthermore, mouse models with enamel defects can also predictively and supportively facilitate human AI mutational analysis using a reverse genetics approach.

INTRODUCTION

In tooth development, after early morphogenesis is complete, the biomineralization process takes place to form three composing hard tissues of a tooth: enamel, dentin, and cementum (Tucker and Sharpe, 2004; Nanci, 2008a; Thesleff, 2013). Dental enamel is formed through a developmental process, amelogenesis, which can be broken down into roughly two main stages: secretory and maturation stages. While the thickness of enamel is accomplished at secretory stage, the hardness is established at maturation stage of amelogenesis (Hu *et al.*, 2007; Nanci, 2008b; Simmer *et al.*, 2010; Bartlett, 2013). This developmental process requires intricate genetic regulations and coordination to build dental enamel with precise shapes and accurate

compositions. Therefore, disturbances of this process result in malformation of dental enamel (Winter and Brook, 1975; Witkop, 1988; Hu *et al.*, 2007).

Amelogenesis imperfecta (AI) is a group of inherited disorders with enamel malformations in the absence of non-dental phenotypes (isolated AI), although the term also refers to the enamel phenotype in syndromes (syndromic AI) (Witkop, 1988; Hu *et al.*, 2007). The prevalence of AI ranges from 1:14,000 to 1:700, depending on various populations (Witkop and Sauk, 1976; Witkop, 1988). Patients afflicted with AI suffer from esthetic and functional burdens and have compromised quality of life (Coffield *et al.*, 2005). Phenotypically, there are mainly three types of AI: hypoplastic, hypomaturation, and hypocalcified. While hypoplastic AI refers to a reduced thickness of dental enamel (thin enamel), hypomaturation AI suggests a hardness defect in dental enamel (soft enamel). Hypocalcified AI is a special form of AI in which the malformed enamel is of normal thickness but cheesy-soft, and may be lost soon after tooth eruption (Witkop, 1988; Aldred *et al.*, 2003; Hu *et al.*, 2007). Unraveling the genetic etiology of different types of AI will not only advance our understanding of normal enamel development but also provide a foundation for developing potential treatments for inherited enamel defects.

The first several AI causative genes were identified with the appreciation that enamel matrix proteins and proteases people identified in developing teeth may play important roles in enamel formation. They were *AMELX* (amelogenin), *ENAM* (enamelin), *KLK4* (Kallikrein-Related Peptidase 4), and *MMP20* (matrix metalloproteinase 20) (Aldred *et al.*, 1992; Rajpar *et al.*, 2001; Hart *et al.*, 2004; Kim *et al.*, 2005). However, defects in these genes only accounted for 25% or less of all isolated AI cases (Kim *et al.*, 2006). Afterwards, with the realization that knowledge-based target gene approaches were not sufficient to comprehensively delineate the

genetic etiology of AI, people started to search for causative genes in AI kindreds through a genome-wide approach. *FAM83H* and *WDR72* were two AI candidate genes identified in large AI kindreds by traditional linkage analysis and chromosome walking (Kim *et al.*, 2008; El-Sayed *et al.*, 2009). More recently, with the emergence of next-generation sequencing, the rate of identifying AI causative genes has been significantly accelerated. With whole exome sequencing, *C4orf26* and *SLC24A4* were identified to be associated with different types of isolated AI (Parry *et al.*, 2012; Parry *et al.*, 2013), and *FAM20A* a type of syndromic AI (O'Sullivan *et al.*, 2011). Nevertheless, despite the fact that many candidate genes have been identified, there are about half of the AI cases of which the genetic causes are still unknown, indicating the genetic heterogeneity of AI and the complexity of enamel formation (Chan *et al.*, 2011).

Our group has been working on mutational analysis of human inherited dental defects for decades, and Dr. Jan Hu has been actively recruiting families with these disorders. In this chapter of the thesis, we described 12 families and sporadic cases of different types of AI, and conducted mutational analyses to determine the genetic etiology of the disorder. We used a target gene approach with known AI candidate genes as well as whole exome sequencing when disease-causing mutations could not be identified in these genes. Through this research, we aimed for not only expanding the mutational spectrum of known candidate genes but also identifying novel genes, which have never been known to be critical for enamel formation.

RESULTS

Family 1 – hypoplastic AI and a *MMP20* mutation

Family 1 was a 3-generation family of which the male proband (III:2) was the only affected individual. His enamel was generally thin but radiographically contrasted well with dentin, suggesting a typical phenotype of hypoplastic AI (Fig. 3.1). Whole exome sequencing identified an apparent disease-causing mutation (g.15390A>G; c.611A>G; p.His204Arg) in both alleles of *MMP20*. This missense mutation changed a highly-conserved amino acid, His^{204}, which was known to coordinate a structural zinc ion for *MMP20* protease activity. The result of this work has been published (Wang *et al.*, 2013b).

Family 2 and 3 – hypoplastic AI and *ITGB6* mutations

The proband of Family 2 (III:1) was an 8-year-old Hispanic girl who was the only person with enamel malformation in the family (Fig. 3.2). Clinically, her teeth were spaced with reduced thickness of enamel and surface roughness. Panoramic radiograph showed that the enamel layer of unerupted teeth, particularly the bicuspids, was thin but contrasted well with dentin. This family was recruited many years ago and has been screened for mutations in all AI candidate genes with no positive results.

The proband of Family 3, an 8-year-old Hispanic boy, (III:3) had a very similar enamel phenotype to that of Family 2 proband except that the maxillary incisors exhibited a "Hutchisonian" or screwdriver morphology (Fig. 3.3). In addition to dental phenotype, the proband was in good health but appeared to be undernourished with specific craniofacial features of anteverted pinnae and ptosis, suggesting a diagnosis of Nance–Horan syndrome (MIM #302350), a well-characterized X-linked syndrome that can explain all of the phenotypes in the proband except the severity of the enamel malformations (Burdon *et al.*, 2003). Therefore, we

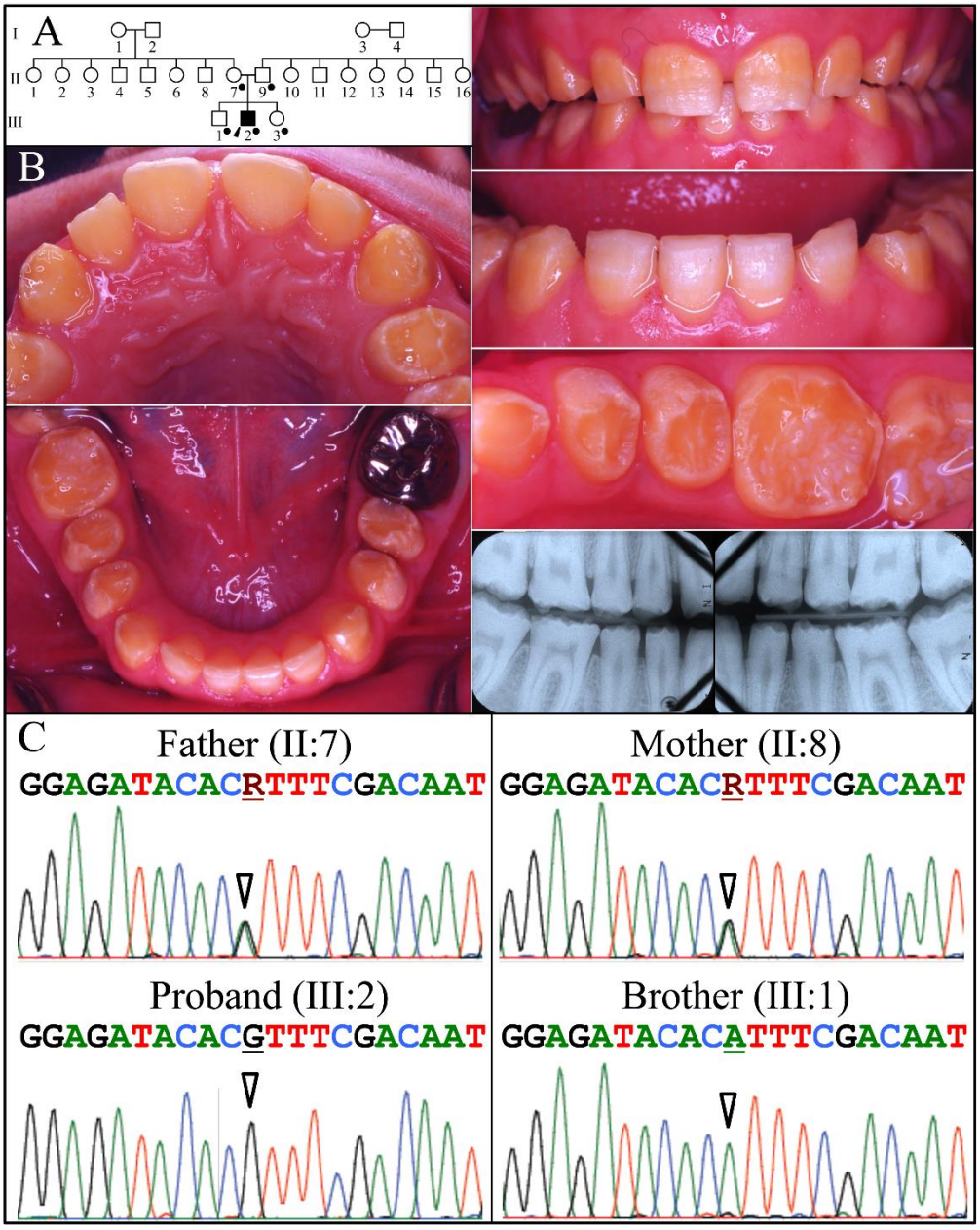


Figure 3.1: Family 1 pedigree, enamel defects, and disease causing mutations.
A: Pedigree of family with a simplex pattern of inheritance. DNA was obtained from the five persons in the nuclear family (black dots). ***B:*** Oral photographs and bitewing radiographs of the proband at age 14 yrs. Most of the occlusal enamel had abraded from the mandibular first molars. The radiographs show a thin layer of enamel that is only slightly more radio-opaque than dentin. ***C:*** DNA-sequencing chromatograms show that the parents were both heterozygous for the *MMP20* c.611A>G transition mutation, while the proband was homozygous for the mutation (p.His204Arg) and his older brother was homozygous-normal. The sequence analysis determined that the mutation was inherited in a recessive pattern. This image is reproduced from Figure 2 of reference (Wang SK, Hu Y, Simmer JP, Seymen F, Estrella NM, Pal S, Reid BM, Yildirim M, Bayram M, Bartlett JD, Hu JC (2013). Novel KLK4 and MMP20 mutations discovered by whole-exome sequencing. *J Dent Res* 92:266-71.) by permission of SAGE Journals.

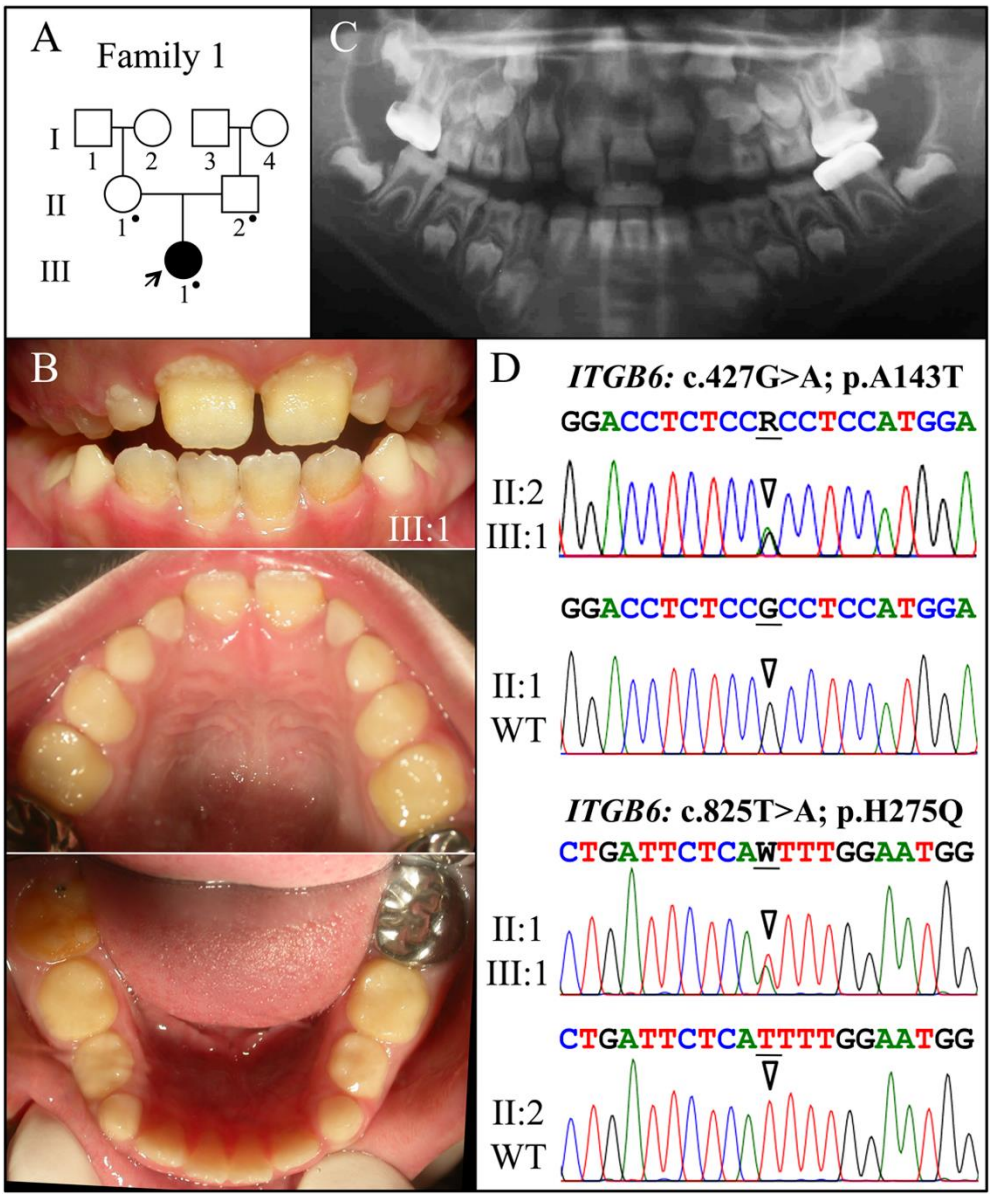


Figure 3.2: Family 2 pedigree, enamel defects, and disease causing mutations.
A: Pedigree. The proband (III:1) is the only person with enamel defects in the family. A dot marks the three study participants who donated samples for DNA sequencing. ***B:*** The proband at almost 8 years is in the mixed dentition stage with the permanent maxillary central incisors and all mandibular incisors and first molars erupted. Oral photographs of the proband show very little enamel covering dentin and signs of rapid attrition. ***C:*** Panoramic radiograph shows no contrasting enamel layer in erupted teeth and only a thin layer of enamel in unerupted teeth, which is characteristic of hypoplastic AI. ***D:*** Sequencing chromatograms of heterozygous ITGB6 mutations in Exon 4 (g.4545G>A; c.427G>A; p.Ala143Thr) and Exon 6 (g.27415T>A; c.825T>A; p.His275Gln). The proband was the only compound heterozygote. The father (II:2) had only the Exon 4 mutation; the mother (II:1) had only the Exon 6 mutation. The sequence variations are named relative to the ITGB6 genomic (NC_000002.11) and cDNA (NM_000888.3) reference sequences. R = G or A; W = A or T. This image is reproduced from Figure 1 of reference (Wang SK, Choi M, Richardson AS, Reid BM, Lin BP, Wang SJ, Kim JW, Simmer JP, Hu JC. ITGB6 loss-of-function mutations cause autosomal recessive amelogenesis imperfecta. *Hum Mol Genet* 2013, Epub ahead of print) by permission of Oxford University Press.

suspected that the enamel defects might be a distinct phenotypic trait other than that of Nance–Horan syndrome, which suggested a separate genetic defect contributing to the enamel phenotype.

With whole exome sequencing of the two probands, we identified three integrin beta 6 (*ITGB6*) mutations responsible for their enamel malformations. The family 2 proband was a compound heterozygote for two *ITGB6* missense mutations (g.4545G>A, c.427G>A, p.Ala143Thr; g.27415T>A, c.825T>A, p.His275Gln), and the Family 3 proband had a *ITGB6* nonsense mutation (g.73664C > T c.1846C > T p.Arg616*) in both alleles in addition to a missense mutation in *NHS* gene (g.355444T>C, c.1697T>C, p.Met566Thr) responsible for the Nance–Horan syndrome. These are the first disease-causing *ITGB6* mutations to be reported, demonstrating a significant role of integrin beta 6 in enamel formation. We also performed immunohistochemistry of ITGB6 on developing mouse teeth and showed its specific expression pattern at different stages of amelogenesis (Fig. 3.4). The result of this work has been published (Wang *et al.*, 2013d).

Family 4 – hypomaturation AI and a *KLK4* mutation

The proband of Family 4 (IV:1) was a 9-year-old girl from a first-cousin marriage. Her enamel was of normal thickness but appeared chipped-off on multiple teeth with brown discoloration and dental caries (Fig. 3.5). Radiographically, the enamel showed only slightly increased radiopacity than dentin, indicating a hypomaturation enamel defect. Whole exome sequencing identified a single-nucleotide deletion (g.6930delG; c.245delG; p.Gly82Alafs*87) in both *KLK4* alleles. The mutation shifted the reading frame and produced a premature stop codon

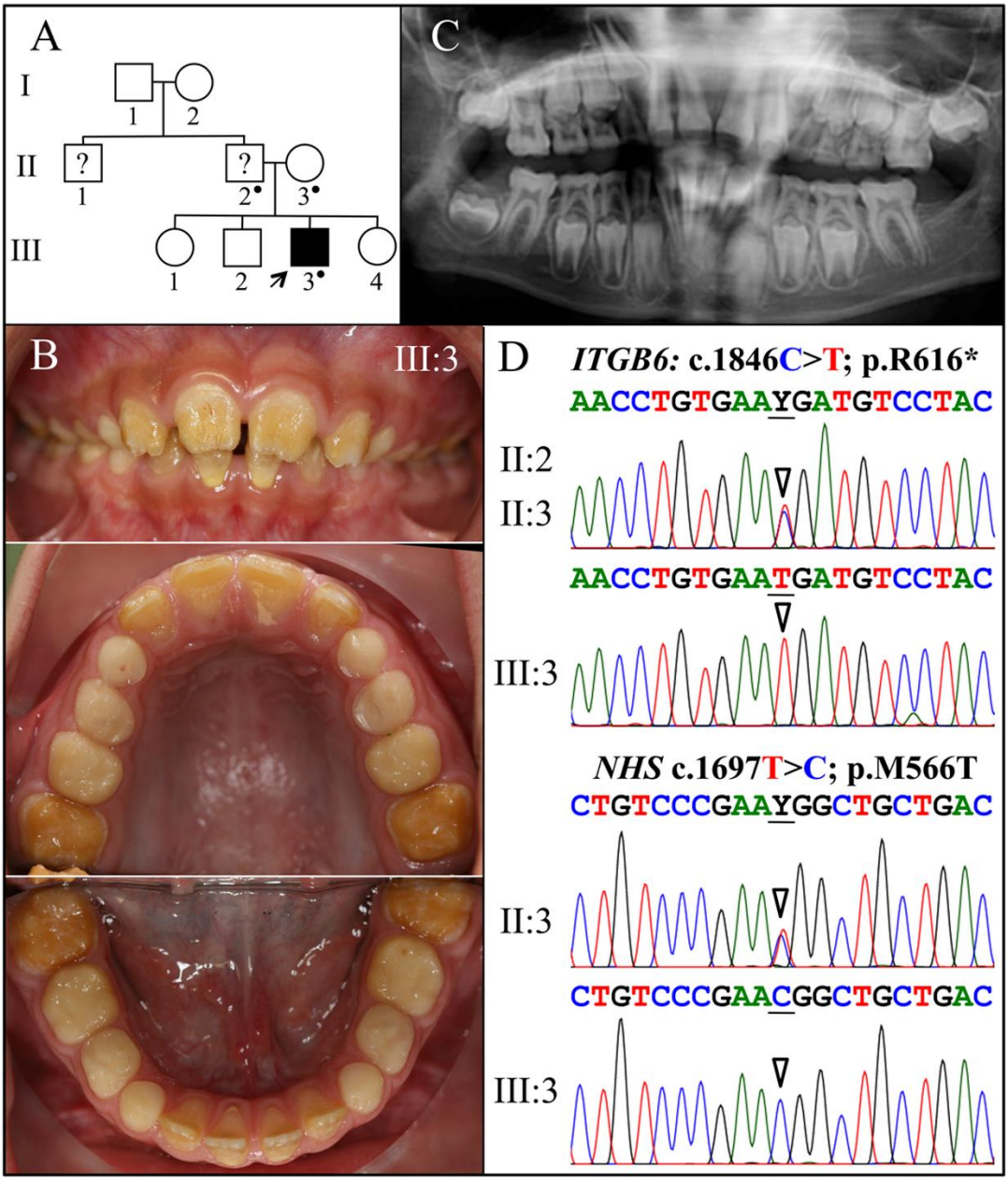


Figure 3.3: Family 3 pedigree, enamel defects, and disease causing mutations.
A: Pedigree. The proband (III:3) is the only person with known enamel defects in the family. The proband's father and uncle were reported by the mother to have 'yellow teeth' but their affection status is uncertain. A dot marks the three study participants who donated samples for DNA sequencing. ***B:*** The proband at age 8 is in the mixed dentition stage with all of the permanent maxillary and mandibular incisors and first molars erupted. Oral photographs of the proband show very little enamel covering dentin and signs of rapid attrition. ***C:*** Panoramic radiograph shows a thin, interrupted layer of contrasting enamel in some erupted teeth and a thin layer of continuous enamel in unerupted teeth. The mandibular left first molar is missing. ***D:*** Sequencing chromatograms of ITGB6 exon 11 (top) and NHS Exon 6 (bottom). The mother (II:3) was heterozygous for both of these sequence variations. Y = T or C. This image is reproduced from Figure 3 of reference (Wang SK, Choi M, Richardson AS, Reid BM, Lin BP, Wang SJ, Kim JW, Simmer JP, Hu JC. ITGB6 loss-of-function mutations cause autosomal recessive amelogenesis imperfecta. *Hum Mol Genet* 2013, Epub ahead of print) by permission of Oxford University Press.

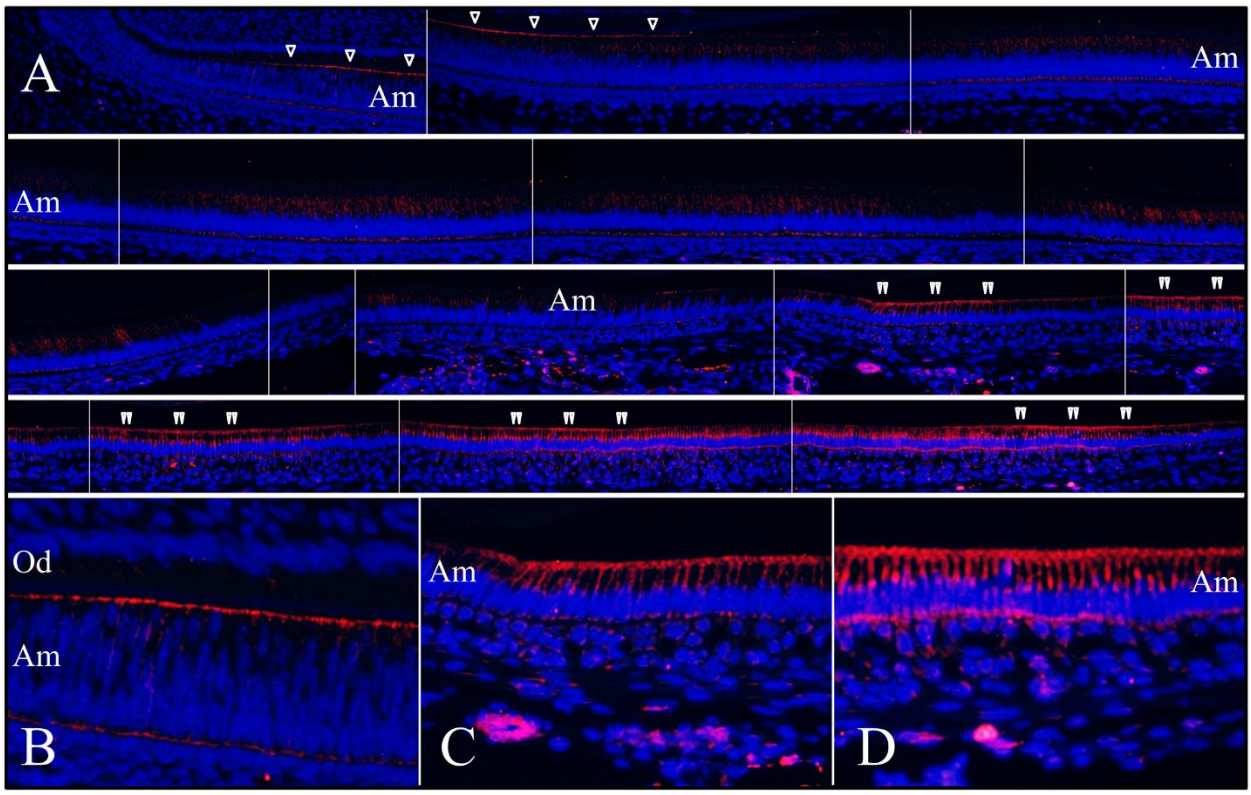


Figure 3.4: ITGB6 immunohistochemistry of Day 14 mouse mandibular incisors.
A: The top left panel shows the cervical loop. Subsequent images move incisally. Single arrowheads mark ITGB6 signal along distal membrane of polarizing ameloblasts (note the varying levels of the nuclei and close proximity to the opposing sheet of odontoblasts). The early signal continues until the ameloblasts are fully polarized and the distance between the ameloblasts and odontoblasts has increased. ITGB6 signal along the distal membrane diminishes and ends and ITGB6 appears to be internalized in secretory stage ameloblasts. Double arrowheads mark ITGB6 signal along the distal membrane in maturation stage ameloblasts. The signal along the distal membrane comes and goes, which presumably correlates with ameloblast modulations. ***B:*** Higher magnification view of ITGB6 along the distal membrane of differentiating ameloblasts. ***C:*** Higher magnification view of ITGB6 at the onset of enamel maturation. ***D:*** Higher magnification view of ITGB6 late in maturation stage. Note: blue is DAPI staining of nuclei; red is ITGB6 immunofluorescence; Od, odontoblasts; Am, ameloblasts. This image is reproduced from Figure 4 of reference (Wang SK, Choi M, Richardson AS, Reid BM, Lin BP, Wang SJ, Kim JW, Simmer JP, Hu JC. ITGB6 loss-of-function mutations cause autosomal recessive amelogenesis imperfecta. *Hum Mol Genet* 2013, Epub ahead of print) by permission of Oxford University Press.

on the mutant transcript, which was predicted to undergo nonsense mediated decay. This was the second human *KLK4* mutation reported so far. The result of this work has been published (Wang *et al.*, 2013b).

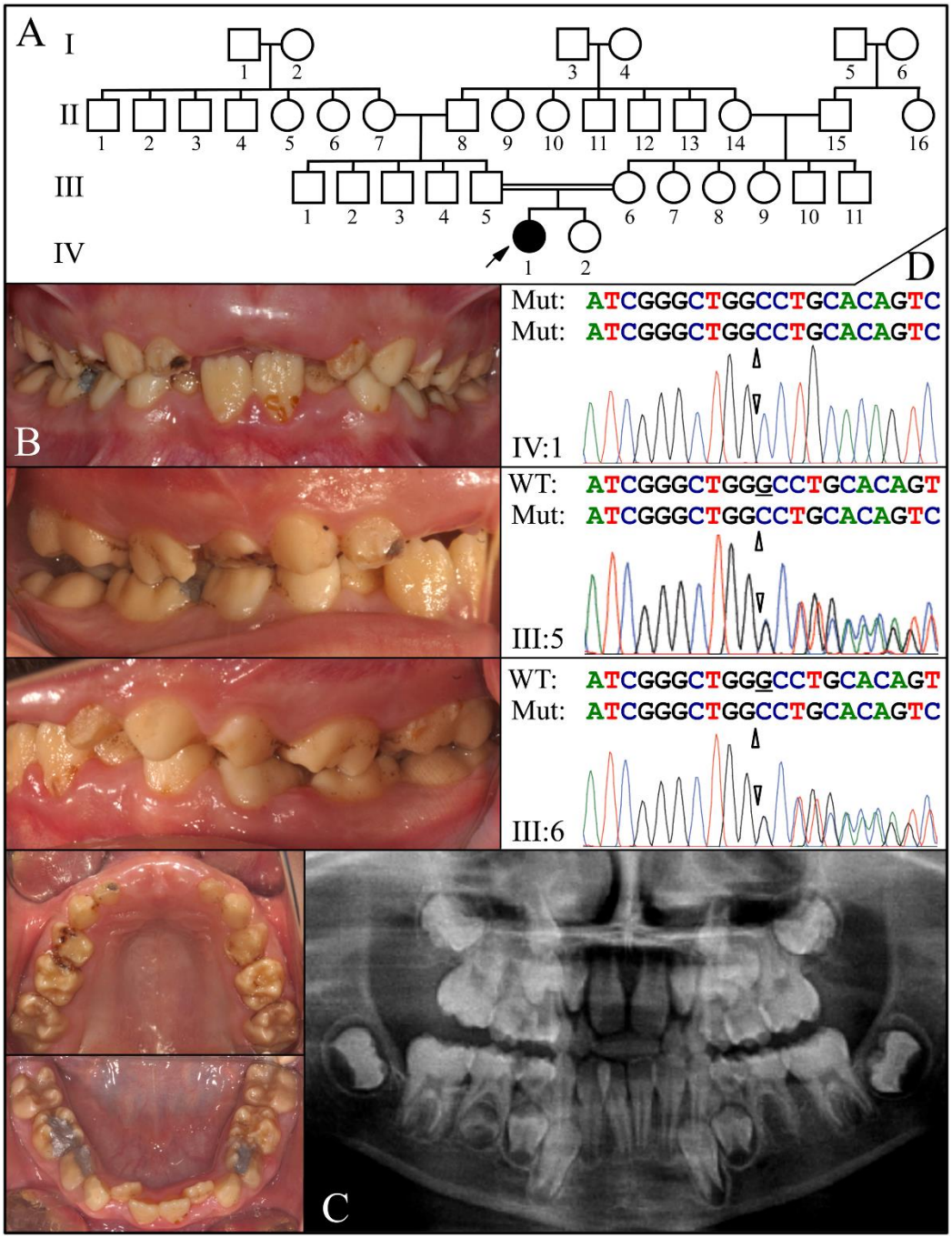


Figure 3.5: Family 4 pedigree, enamel defects, and disease causing mutations.
A: Pedigree of consanguineous family from Turkey with enamel malformations in the proband. ***B:*** Frontal, lateral, and occlusal photos of the proband at age 9 yrs. The enamel is chipped and shaded brown. ***C:*** Panorex radiograph of the mixed dentition showing enamel of normal thickness that contrasts only slightly with dentin. ***D:*** DNA sequencing chromatograms showing that the single-nucleotide deletion (c.245delG; position marked by arrowhead) is homozygous in the proband and heterozygous in his parents. This image is reproduced from Figure 1 of reference (Wang SK, Hu Y, Simmer JP, Seymen F, Estrella NM, Pal S, Reid BM, Yildirim M, Bayram M, Bartlett JD, Hu JC (2013). Novel KLK4 and MMP20 mutations discovered by whole-exome sequencing. *J Dent Res* 92:266-71.) by permission of SAGE Journals.

Family 5 – hypomaturation AI and a *SLC24A4* mutation

The proband of Family 5 (II:1) was a 5.5-year-old Turkish girl from a consanguineous marriage (Fig. 3.6). Her primary teeth appeared creamy-yellow in color with extensive carious defects. The enamel was soft and chipped off from tooth surface. The lower central incisors were of normal thickness of dental enamel, suggesting the loss of enamel thickness was due to post-eruption attrition. Radiographically, the enamel of unerupted permanent incisors and first molars showed full thickness but no contrast with underlying dentin, indicating a hypomaturation type of AI (Fig. 3.6).

Whole exome sequencing identified a homozygous missense mutation in *SLC24A4* (g.124552C>A; c.437C>T; p.Ala146Val), a newly-identified AI candidate gene, which encodes a potassium-dependent sodium-calcium exchanger. The mutated Ala^{146} is strictly conserved between all the SLC24A and SLC8A homologs of many species, suggesting that the amino acid is critical for the function of SLC24A4 and that the identified mutation is disease-causing. This is the third disease-causing *SLC24A4* mutation to be reported. We also performed immunohistochemistry of SLC24A4 on developing mouse teeth and showed its specific expression in maturation-stage ameloblasts, which demonstrated its critical role in enamel maturation (Fig. 3.7). The result of this work has been published (Wang *et al.*, 2014b).

Family 6 – syndromic hypomaturation AI and a *STIM1* mutation

The proband of Family 6 (IV:1) was a Turkish girl from a first-cousin marriage (Fig. 3.8). Soon after recruitment, the family moved to another country and was no longer contactable, so the information of disease phenotype was limited. The proband had a history of frequent throat

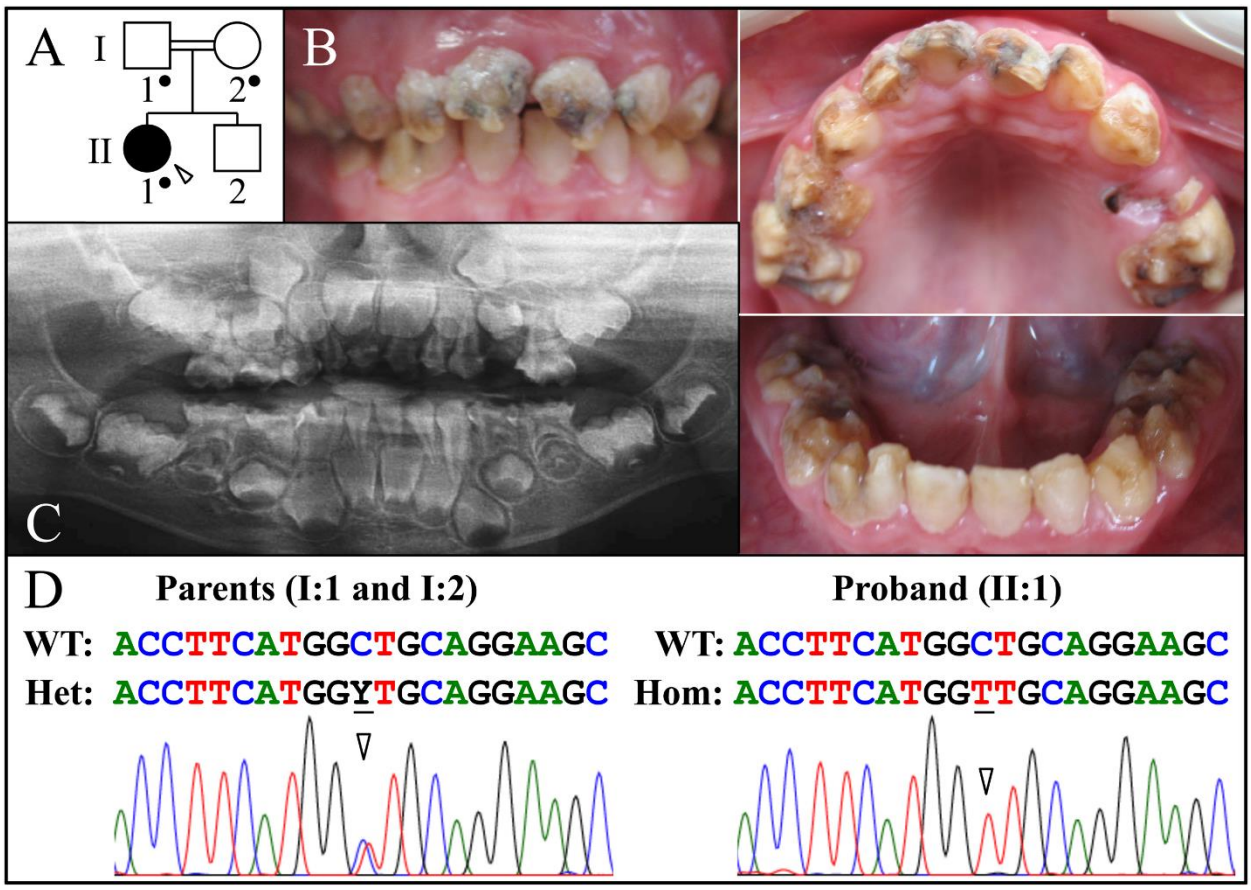


Figure 3.6: Family 5 pedigree, enamel defects, and disease causing mutations.
A: Pedigree. The arrowhead marks the proband, the only affected individual in the consanguineous family. Dots mark the three persons who donated samples for DNA sequencing. ***B:*** Oral photographs of the proband (II:1) at age 5.5. The teeth are yellow or cream-colored, show signs of attrition and dental caries. ***C:*** Panorex of the proband at age 5.5. The enamel of erupting first molars exhibits normal thickness but no contrast with underlying dentin, indicating maturation enamel defects. ***D:*** Sequence from *SLC24A4* Exon 5 showing the heterozygosity of the sequence variation g.124552C>A; c.437C>T; p.Ala146Val that occurred in the father (I:1) and mother (I:2) (left) and the homozygosity in the proband (II:1) (right). The mutation designations are with respect to the *SLC24A4* genomic reference sequence NG_023408.1 and cDNA reference sequence NM_153646.3 (for mRNA transcript variant 1). ***Key:*** arrowhead: mutation point; Y: T or C. This image is reproduced from Figure 3 of reference (Wang SK, Choi M, Richardson AS, Reid BM, Seymen F, Yildirim M, Tuna EB, Gencay K, Simmer JP, Hu JC. STIM1 and SLC24A4 are critical for enamel maturation. *J Dent Res* 2014, Epub ahead of print) by permission of SAGE Journals.

infections, but no immunological evaluation was performed. Her teeth looked creamy-brown, and the malformed enamel was of normal thickness but soft with extensive post-eruption

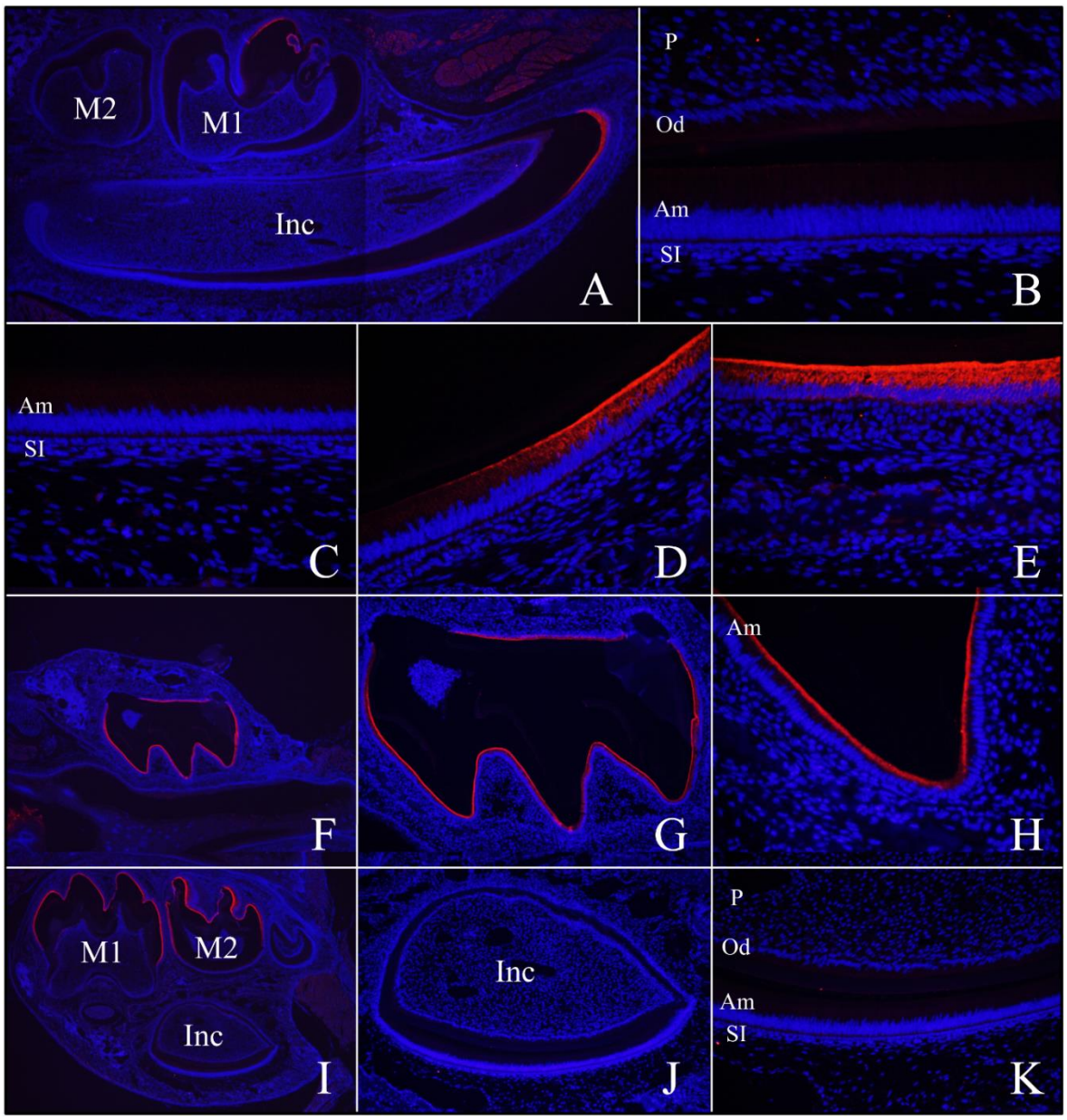


Figure 3.7: SLC24A4 immunohistochemistry of developing teeth of day 5 and 11 mice. ***A-E:*** At postnatal day 5 (PN5), SLC24A4 signal was observed in maturation stage ameloblasts at the incisal end of mandibular incisor and cusp tip of the first molar, and in skeletal muscles (A). High magnification views of differentiation and secretory stages toward the basal end of the mandibular incisor (B, C); transition to maturation stage (D); maturation stage (E). ***F-H:*** At postnatal day 11 (PN11), SLC24A4 signal was detected in the maxillary first molar where ameloblasts are in the maturation stage. ***I-K:*** At postnatal day 11 (PN11), only ameloblasts of the mandibular first and second molars (at maturation stage) were above the threshold for detection, but other tissues, including secretory stage ameloblasts (in the third molar and in the basal third of the incisor) were not (I). High magnification of secretory stage ameloblasts in a mandibular incisor (J, K). ***Key:*** Am: ameloblasts; Od: odontoblasts; P: pulp; SI: stratum intermedium. This image is reproduced from Figure 4 of reference (Wang SK, Choi M, Richardson AS, Reid BM, Seymen F, Yildirim M, Tuna EB, Gencay K, Simmer JP, Hu JC. STIM1 and SLC24A4 are critical for enamel maturation. *J Dent Res* 2014, Epub ahead of print) by permission of SAGE Journals.

attrition. Noticeably, nail dysplasia was evident on some fingers and toes, suggesting that the proband might be a case of syndromic AI.

Whole exome sequencing identified a homozygous missense mutation in *STIM1* (g.232598C>T; c.1276C>T; p.Arg426Cys), of which loss-of-function mutations have been shown to cause severe immunodeficiency, congenital myopathy, and ectodermal dysplasia with enamel defects (Picard *et al.*, 2009). The mutated amino acid (R^{426}) is strictly conserved among vertebral STIM1 orthologs and is located within a critical functional domain of STIM1 protein, suggesting that the identified mutation is disease-causing. We also performed immunohistochemistry of STIM1 on developing mouse teeth and showed its specific expression in maturation-stage ameloblasts, which demonstrated its critical role in enamel maturation (Fig. 3.9). The result of this work has been published (Wang *et al.*, 2014b).

Family 7 – hypocalcified AI and a *FAM83H* mutation

The proband of Family 7 was a 10-year-old Turkish boy from a consanguineous marriage (Fig. 3.10). He was the only individual with enamel malformations in the whole family, suggesting the disease was caused by a recessive or a *de novo* mutation. Clinically, the proband had a mixed dentition with enamel defects on both primary and permanent teeth. The malformed enamel was brown discolored, cheesy-soft, and chipped off from tooth surfaces. Most of the teeth showed extensive post-eruption attrition except the erupting lower right first premolar, which had a normal thickness of enamel. Radiographically, the enamel of unerupted teeth was of normal thickness but showed no contrast with underlying dentin. These findings suggested a diagnosis of hypocalcified AI.

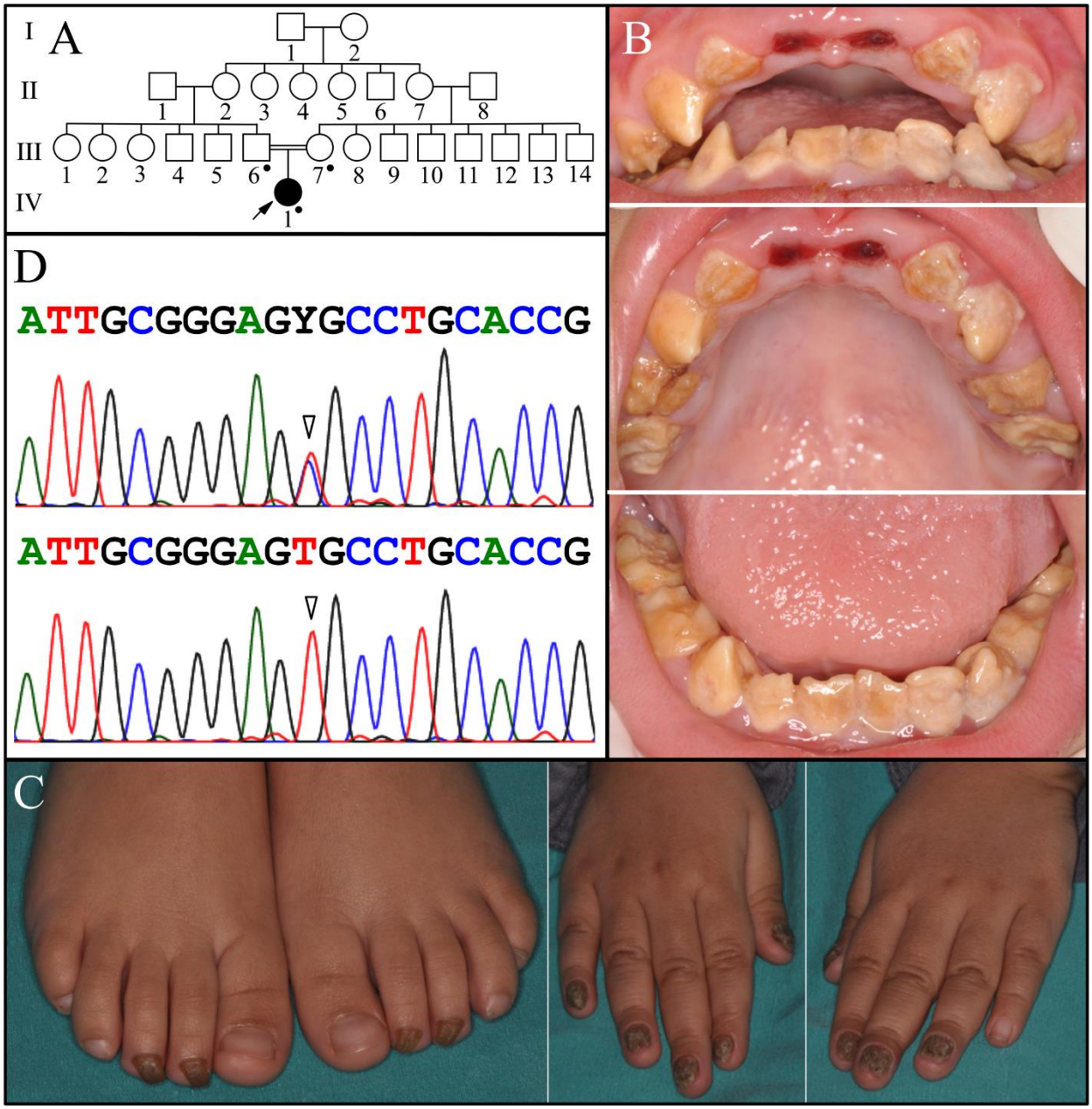


Figure 3.8: Family 6 pedigree, enamel defects, and disease causing mutations.
A: Pedigree. Dots mark the three persons who donated samples for DNA sequencing. ***B:*** Oral photographs of the proband (IV:1) at age 6. The teeth are normal in size and shape, but are brown or cream-colored, and have undergone attrition. ***C:*** Photographs of the hands and feet showing nail dysplasia. ***D:*** Sequence from *STIM1* Exon 10 revealing heterozygosity for the sequence variation g.232598C>T; c.1276C>T; p.Arg426Cys that occurred in the father (III:6) and mother (III:7) (top) and homozygosity in the proband (IV:1) (bottom). The mutation designations are with respect to the *STIM1* genomic reference sequence NG_016277.1 and cDNA reference sequence NM_001277961.1 (for mRNA transcript variant 1). ***Key:*** arrowhead: mutation point; Y: T or C. This image is reproduced from Figure 1 of reference (Wang SK, Choi M, Richardson AS, Reid BM, Seymen F, Yildirim M, Tuna EB, Gencay K, Simmer JP, Hu JC. STIM1 and SLC24A4 are critical for enamel maturation. *J Dent Res* 2014, Epub ahead of print) by permission of SAGE Journals.

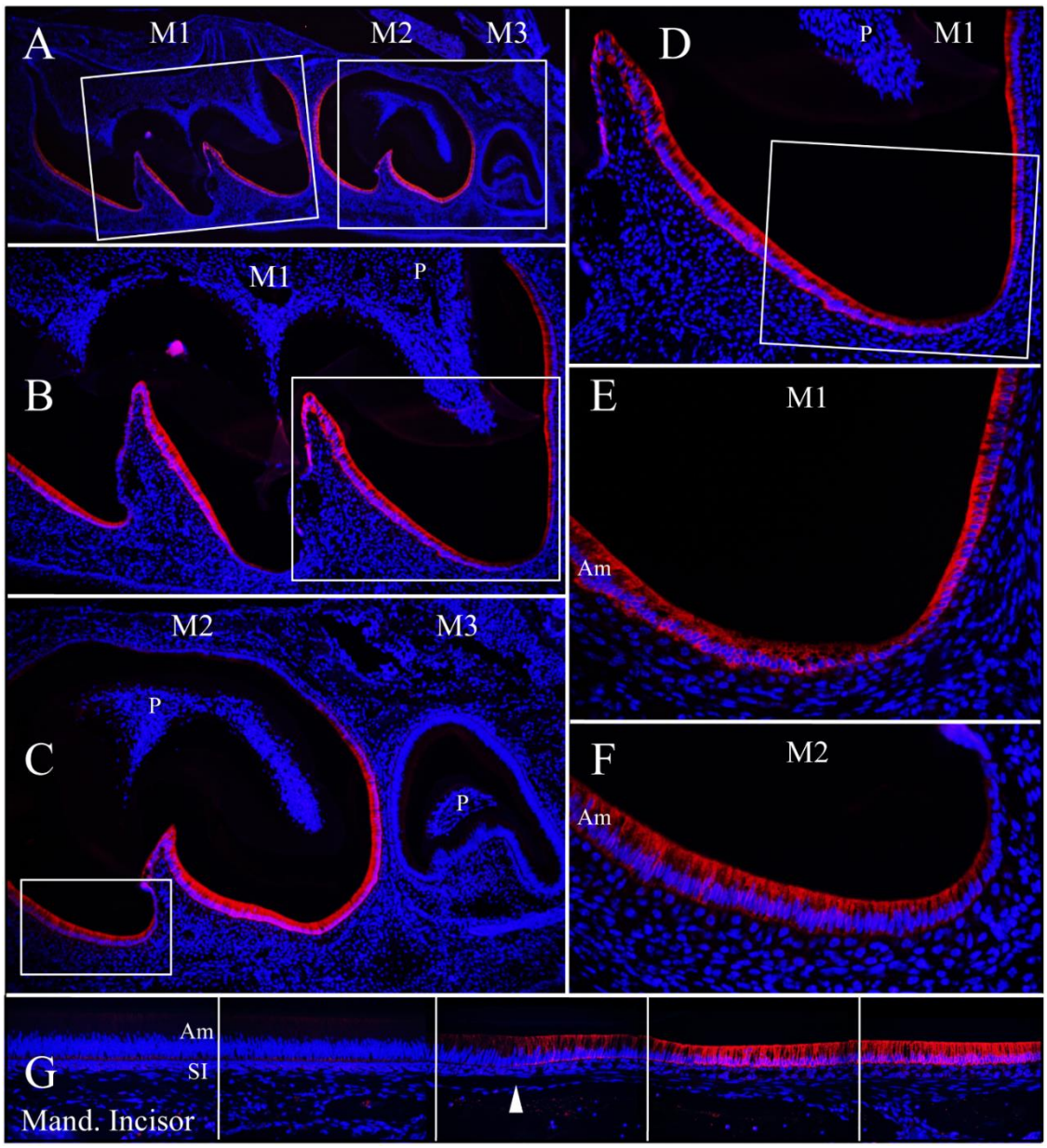


Figure 3.9: STIM1 immunohistochemistry of day 11 maxillary molars and mandibular incisor. ***A:*** Low magnification (40x) views of the maxillary first (M1), second (M2), and third (M3) molars. Boxes outline the higher magnification (100x) views in panels B and C. ***B:*** Maxillary first molar (M1) at 100x. The box outlines the higher magnification (200x) view in panel D. ***C:*** Maxillary second (M2) and third (M3) molars at 100x. The box outlines the highest magnification (400x) view shown in panel F. ***D:*** Distal cusp of M1 (200x). The box outlines the highest magnification (400x) view shown in panel E. ***E:*** Distal cusp tip of M1 (400x). ***F:*** Mesial cusp tip of M2 (400x). ***G:*** Longitudinal sections of a mandibular incisor (400x). Arrowhead marks the approximate onset of the maturation stage. Note that only maturation stage ameloblasts are positive for STIM1 in developing teeth. Secretory stage ameloblasts in M3 and the incisor are negative. ***Key:*** Am: ameloblasts; Od: odontoblasts; P: pulp; SI: stratum intermedium. This image is reproduced from Figure 2 of reference (Wang SK, Choi M, Richardson AS, Reid BM, Seymen F, Yildirim M, Tuna EB, Gencay K, Simmer JP, Hu JC. STIM1 and SLC24A4 are critical for enamel maturation. *J Dent Res* 2014, Epub ahead of print) by permission of SAGE Journals.

Whole exome sequencing of proband's DNA identified no potential disease-causing mutations in all AI candidate genes except a nonsense mutation in *FAM83H* (g.10653C>T, c.1369C>T, p.Gln457*), which explained the enamel phenotype. However, neither of his parents carried this sequence variation, suggesting that it is a spontaneous *de novo* mutation. In consistency with previously reported mutations, this one was a nonsense mutation in Exon5 of *FAM83H* and was predicted to produce a truncated protein without its C-terminus. The analyzed exome sequencing data and the gene list of non-reference sequence variants were provided by Dr. Murim Choi. The *FAM83H* mutation confirmation with Sanger sequencing was conducted by Bryan Reid.

Family 8, 9, 10, 11, and 12 – enamel-renal syndrome and *FAM20A* mutations

In 2011, O'Sullivan *et al.* identified *FAM20A* mutations, as a cause of Amelogenesis Imperfecta and Gingival Fibromatosis Syndrome (AIGFS; MIM #614253), an autosomal recessive disorder characterized by severe enamel hypoplasia, failed tooth eruption, intrapulpal calcifications, and gingival hyperplasia (O'Sullivan *et al.*, 2011). In our hands, we have recruited several AI families with these distinct phenotypes. However, interestingly, some of the probands from these families exhibited asymptomatic nephrocalcinosis in addition to the dental phenotypes of AIGFS. This finding suggested that *FAM20A* mutations may also cause Enamel-Renal Syndrome (ERS; MIM #204690), an autosomal recessive disorder closely resembling AIGFS except for the additional phenotype of renal calcifications.

Family 8, 9, 10, 11, and 12 are five AI families with distinct phenotypes of severely hypoplastic (thin) or aplastic enamel on both the primary and secondary dentitions, pulp stones,

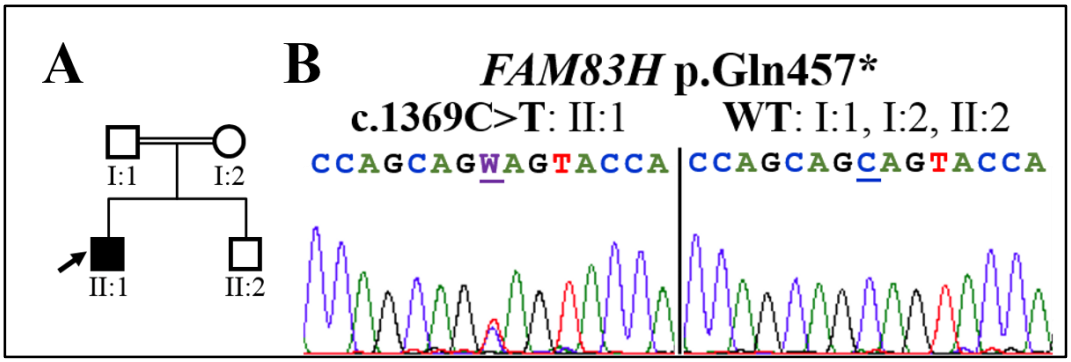


Figure 3.10: Family 7 pedigree, enamel defects, and disease causing mutations.
A: Pedigree: The proband (II:1) comes from a consanguineous marriage. ***B:*** *FAM83H* Exon 5 DNA sequencing chromatograms. The proband's parents (I:1 and I:2) and younger brother (II:1) were all wild-type (WT) at cDNA position 1369. The proband (II:I) had a heterozygous c.1369C>T mutation of *FAM83H*. This mutation is a *de novo* nonsense mutation (g.10653C>T, c.1369C>T, p.Gln457*). The gene numbers start from the first nucleotide of the *FAM83H* reference sequence NG_016652.1. The cDNA numbers start from the translation initiation site of *FAM83H* reference sequence NM_198488.3. (The *FAM83H* chromatograms are courtesy for Bryan Reid.)

and failed or delayed eruption of much of the permanent dentition, particularly the posterior teeth (Fig. 3.11; Fig. 3.12; Fig. 3.13; Fig. 3.14; Fig. 3.15). For kidney phenotypes, while probands from Family 9 and 10 exhibited nephrocalcinosis (Fig. 3.12), the proband of Family 11, at age of 12, did not. We have not been able to obtain any information concerning kidney calcifications from Family 8 and 12.

By directly screening for *FAM20A* mutations in these families, we identified, in each case, recessive *FAM20A* mutations: Family 8 (c.992G>A; g.63853G>A; p.Gly331Asp), Family 9 (c.720-2A>G; g.62232A>G; p.Gln241_Arg271del), Family 10 (c.406C>T; g.50213C>T; p.Arg136* and c.1432C>T; g.68284C>T; p.Arg478*), Family 11 (g.502011G>C; c.405-1G>C), and Family 12 (g.65094G>A; c.1207G>A; p.D403N), which demonstrated for the first time that *FAM20A* mutations cause ERS. All these 7 *FAM20A* mutations as well as other reported ones are expected to be loss-of-function mutations, although the function of FAM20A is still unknown. We also characterized teeth extracted from Family 3 proband and demonstrated that FAM20A$^{-/-}$

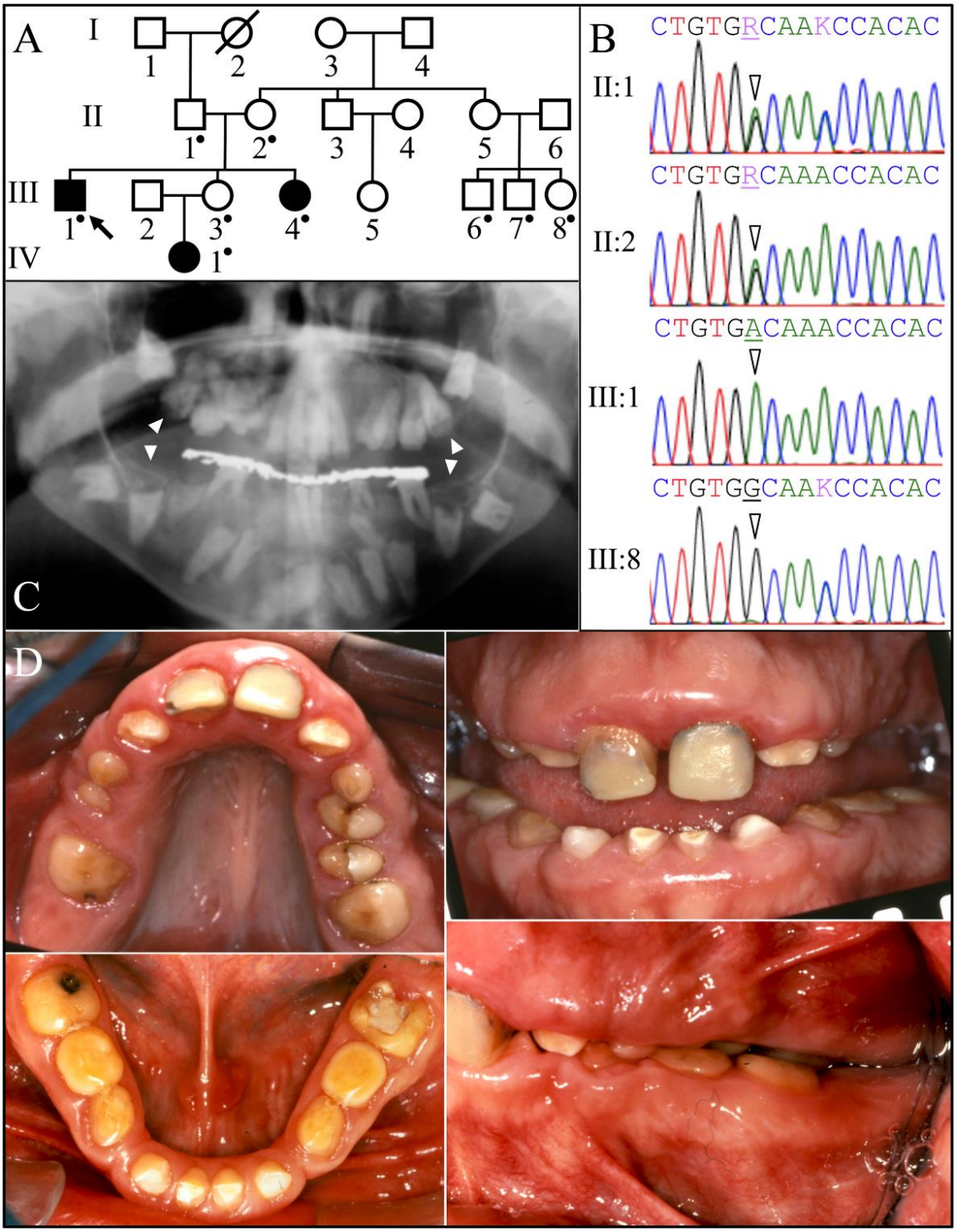


Figure 3.11: Family 8 pedigree, enamel defects, and disease causing mutations.
A: Pedigree. A dot marks person who donated samples for DNA sequencing. ***B:*** *FAM20A* exon 7 DNA sequencing chromatograms. The proband's parents (II:1 and II:2) were both heterozygous (R = A or G) at cDNA position 992 (arrowheads). The proband (III-1) had the c.992G>A transition mutation in both alleles of *FAM20A*. This mutation changed a conserved glycine with an aspartic acid (p.G331D). The proband's affected younger sister (III-4) and her infant niece (IV:1) were also homozygous for this

mutation (not shown). II:1 and III:8 were heterozygous for a recognized polymorphism (rs2302234) in exon 7 (K = A or C) unrelated to the phenotype. ***C:*** Proband's panoramic radiograph. Note the many unerupted teeth. The mandibular and maxillary unerupted second molars show concave occlusal surfaces without enamel (arrowheads) ***D:*** Proband's oral photos. The maxillary central incisors are restored. The clinical crowns were short with hypoplastic enamel. There was a deep anterior overbite, a posterior cross-bite, and retained mandibular primary molars (letters K, L, S, T). This image is reproduced from Figure 1 of reference (Wang SK, Aref P, Hu Y, Milkovich RN, Simmer JP, El-Khateeb M, Daggag H, Baqain ZH, Hu JC (2013). FAM20A mutations can cause enamel-renal syndrome (ERS). *PLoS Genet* 9:e1003302.).

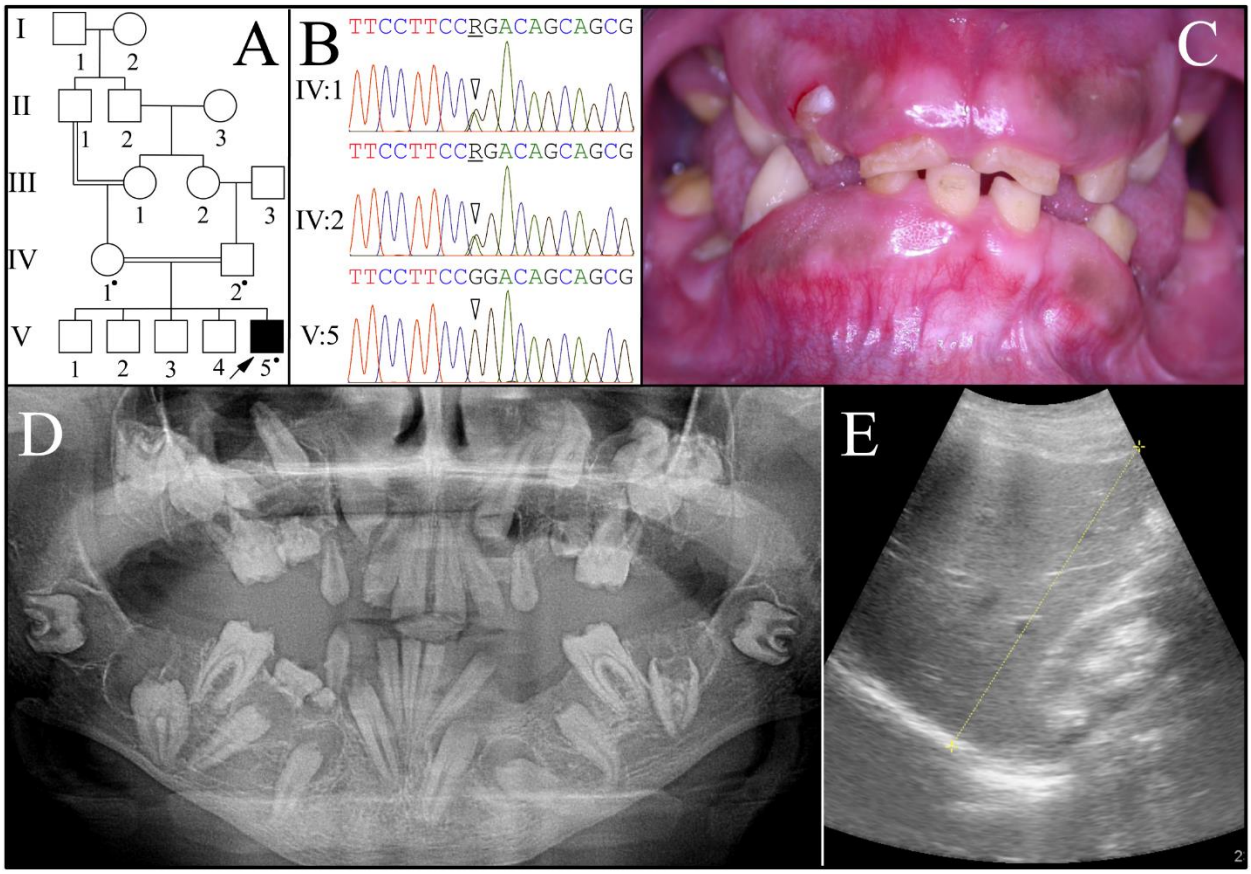


Figure 3.12: Family 9 pedigree, enamel defects, and disease causing mutations.
A: Pedigree: a dot marks person who donated samples for DNA sequencing. ***B:*** *FAM20A* intron 4 DNA sequencing chromatograms. The proband's parents (IV:1 and IV:2) were both heterozygous (R = A or G) at cDNA position 720 (2 arrowheads). The proband (V:5) had the c.720-2A>G transition mutation in both alleles of *FAM20A*. This mutation is predicted to cause the skipping of exon 5, which is predicted to delete 31 amino acids (Q241-R271) from the protein without shifting the reading frame. ***C:*** Proband's oral photo showing enamel hypoplasia, gingival enlargement and failed eruption. ***D:*** Proband's panoramic radiograph. Note the enamel hypoplasia, pulp calcifications, and unerupted teeth with pericoronal radiolucencies delimited by sclerotic borders. The left mandibular second molar (#18) shows apparent crown resorption. ***E:*** Ultrasound of proband's right kidney, located to the right of the yellow line. This image is reproduced from Figure 2 of reference (Wang SK, Aref P, Hu Y, Milkovich RN, Simmer JP, El-Khateeb M, Daggag H, Baqain ZH, Hu JC (2013). FAM20A mutations can cause enamel-renal syndrome (ERS). *PLoS Genet* 9:e1003302.).

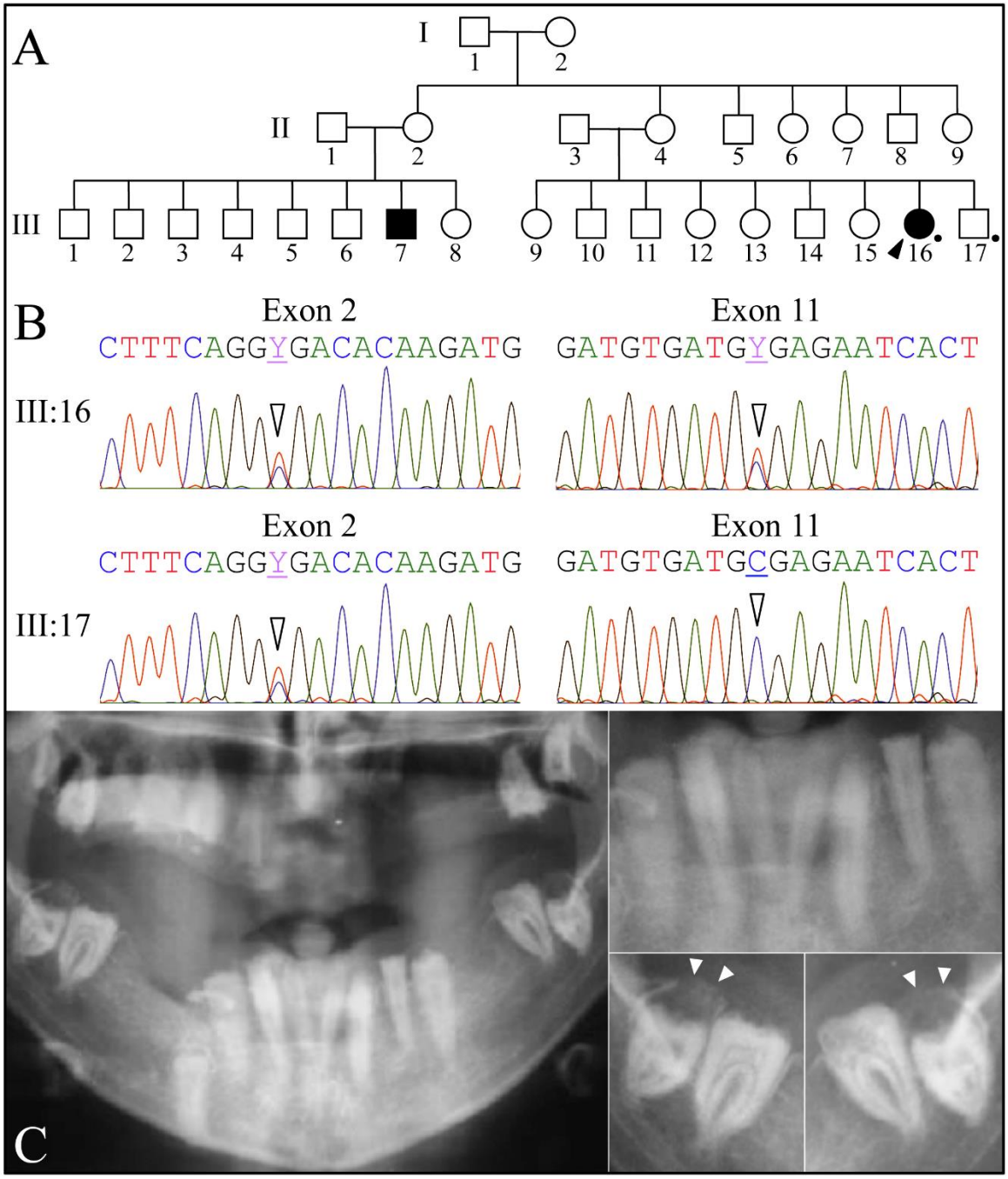


Figure 3.13: Family 10 pedigree, enamel defects, and disease causing mutations.
A: Pedigree consistent with a recessive pattern of inheritance. ***B:*** Exon 2 (left) and exon 11 DNA sequencing chromatograms. The proband (III:16) is heterozygous for nonsense mutations in exon 2 (c.406C>T) and exon 11 (c.1432C>T). The unaffected brother (III:17) is only heterozygous for the c.406C>T mutation in exon 2. ***C:*** Panoramic radiograph of proband. Note the lack of enamel, pericoronal radiolucencies over the unerupted mandibular third molars (arrowheads), and apparent crown resorption of the left mandibular second molar (#18). This image is reproduced from Figure 3 of reference (Wang SK, Aref P, Hu Y, Milkovich RN, Simmer JP, El-Khateeb M, Daggag H, Baqain ZH, Hu JC (2013). FAM20A mutations can cause enamel-renal syndrome (ERS). *PLoS Genet* 9:e1003302.).

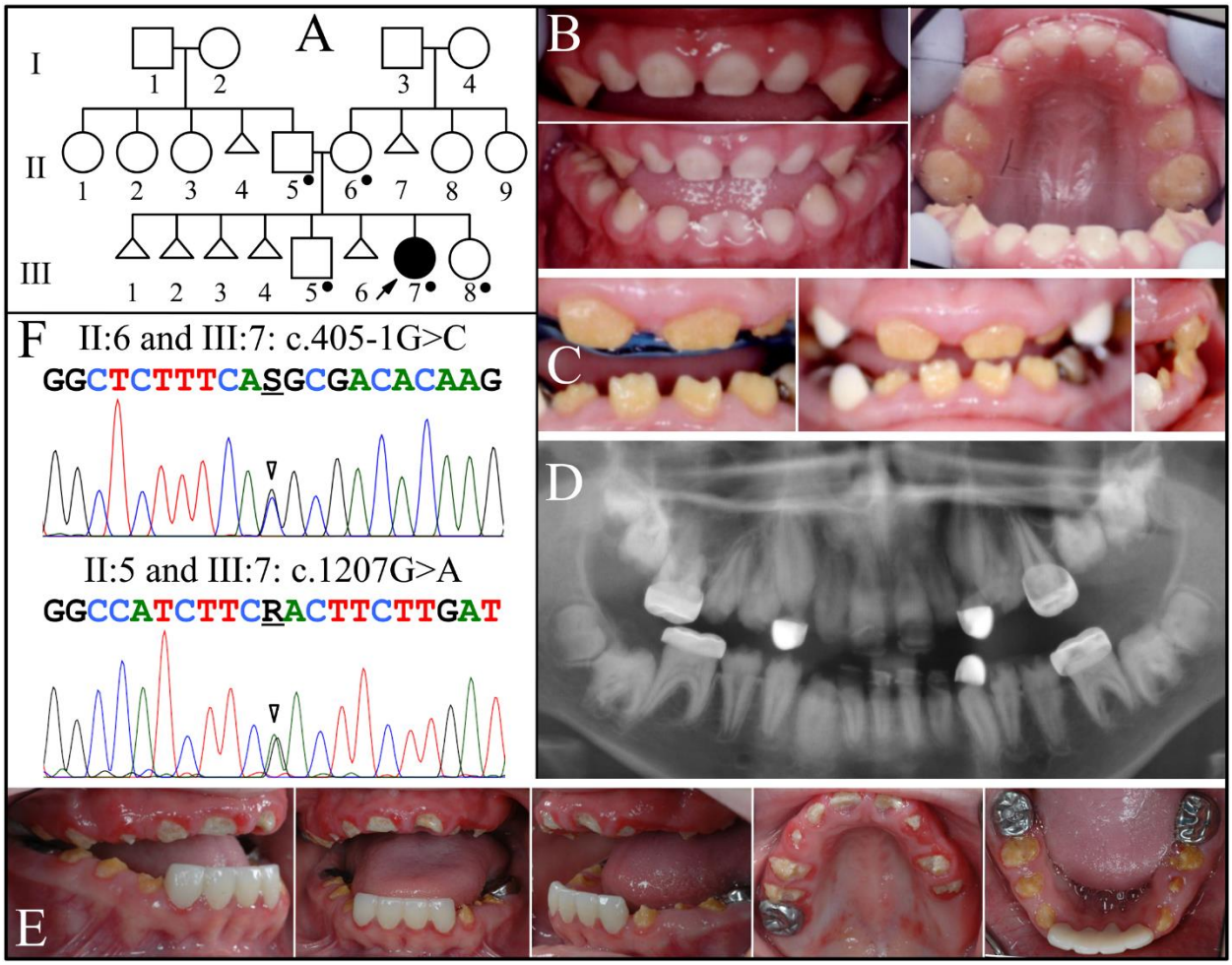


Figure 3.14: Family 11 pedigree, enamel defects, and disease causing mutations.
A: Pedigree. Dots mark the five persons who donated samples for DNA sequencing. Triangles represent stillborns. ***B:*** Oral photographs of the proband (III:7) at age 2.5 yrs. ***C:*** Oral photographs of the proband at age 8.5 yrs. The anterior incisors are present and exhibit severe enamel hypoplasia. The white cuspids are dental restorations. The attached gingiva is enlarged, but the impression is enhanced by the small sizes of the clinical crowns. ***D:*** Panorex of the proband at age 11.5 yrs. Enamel is missing or does not contrast with dentin throughout. Most tooth roots are short, and eruption is less than expected. Pulp stones are observed in many teeth, particularly in the first molars. ***E:*** Oral photographs of the proband at age 13 yrs. The gingival hyperplasia is minimal and could readily be missed in an oral examination. ***F:*** DNA sequencing chromatograms of Family 1. ***Top*****:** Sequence from the border of Exon 1 and Intron 1, revealing heterozygosity for a splice junction mutation (g.502011G>C; c.405-1G>C) that occurred in the mother and proband. ***Bottom*****:** Exon 8 sequence revealing heterozygosity for a missense mutation (g.65094G>A; c.1207G>A; p.D403N) that occurred in the father and proband. The unaffected brother (III:5) and sister (III:8) had neither of these mutations (data not shown). The mutation designations are with respect to the *FAM20A* genomic reference sequence NG_029809.1 and cDNA reference sequence NM_017565.3 (for mRNA transcript variant 1). ***Key*****:** arrowhead = mutation point; R = A or G; S = G or C. This image is reproduced from Figure 1 of reference (Wang SK, Reid BM, Dugan SL, Roggenbuck JA, Read L, Aref P, Taheri AP, Yeganeh MZ, Simmer JP, Hu JC (2014). FAM20A Mutations Associated with Enamel Renal Syndrome. *J Dent Res* 93:42-8.) by permission of SAGE Journals.

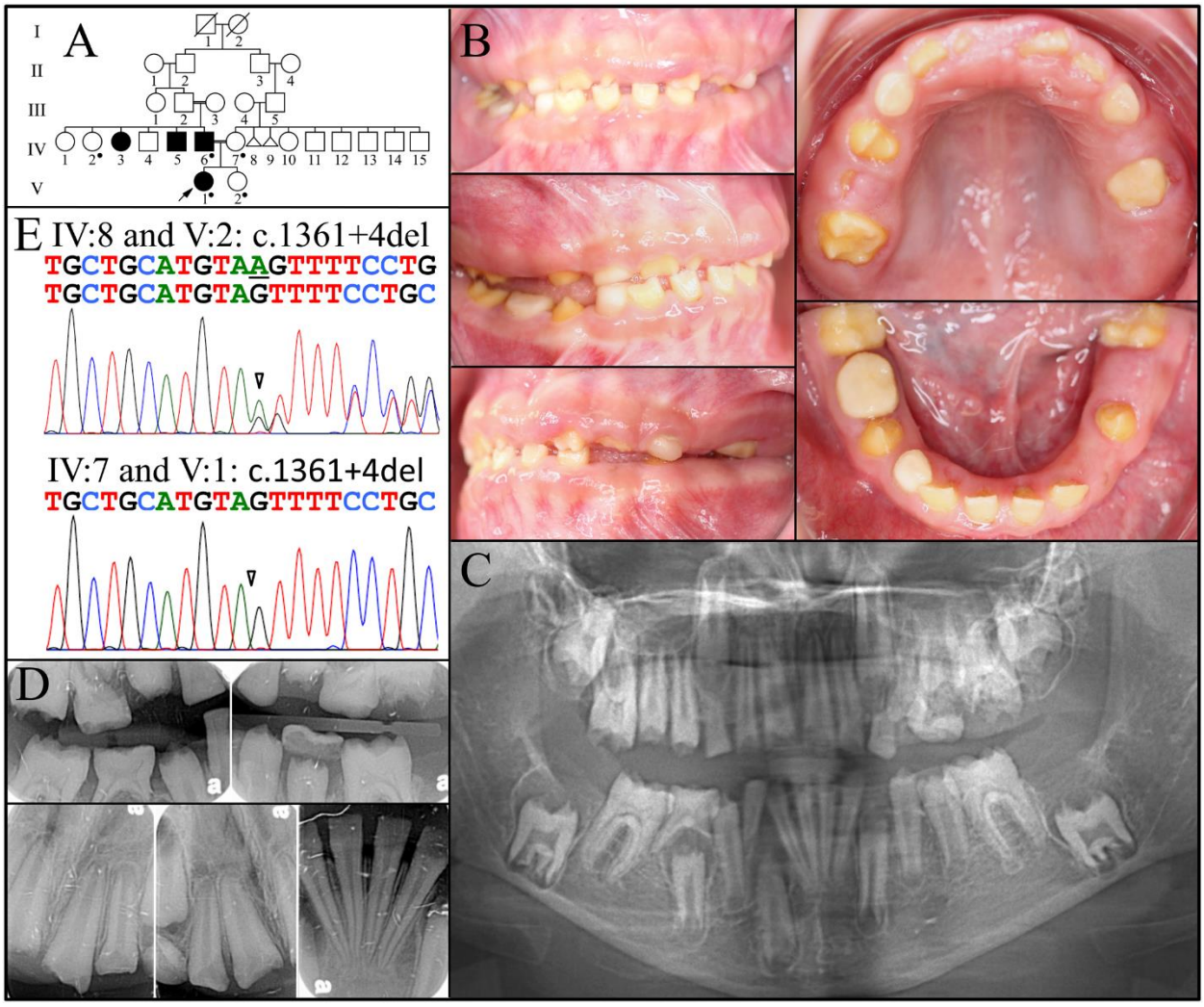


Figure 3.15: Family 12 pedigree, enamel defects, and disease causing mutations.
A: Pedigree. Dots mark the five persons who donated samples for DNA sequencing. Triangles represent stillborns. ***B:*** Oral photographs of the proband (V:I) at age 10 yrs. ***C:*** Panorex radiograph taken when the girl was 10 yrs old. ***D:*** Bitewings and periapical radiographs taken when the girl was 9 yrs old. ***E:*** DNA sequencing chromatograms for Family 2. ***Top*****:** Sequencing chromatogram from the border of Exon 10 and Intron 10 revealing heterozygosity for a splice junction deletion (g.66622del; c.1361+4del) in the proband's unaffected aunt (IV:2), mother (IV:7), and the sister (V:2). ***Bottom*****:** Sequencing chromatogram from the border of Exon 10 and Intron 10 revealing homozygosity for a splice junction deletion (g.66622del; c.1361+4del) in the proband (V:1) and her affected father (IV:6). The mutation designations are with respect to the*FAM20A* genomic reference sequence NG_029809.1 and cDNA reference sequence NM_017565.3 (for mRNA transcript variant 1). This image is reproduced from Figure 2 of reference (Wang SK, Reid BM, Dugan SL, Roggenbuck JA, Read L, Aref P, Taheri AP, Yeganeh MZ, Simmer JP, Hu JC (2014). FAM20A Mutations Associated with Enamel Renal Syndrome. *J Dent Res* 93:42-8.) by permission of SAGE Journals.

molars lacked true enamel, showed extensive crown and root resorption, hypercementosis, and partial replacement of resorbed mineral with bone or coalesced mineral spheres (Wang *et al.*, 2013a). Along with nephrocalcinosis in some cases, these findings suggested a critical role of FAM20A in the regulation of biomineralization processes. In addition, we performed immunohistochemistry of FAM20A in developing mouse heads and adult kidneys, and localized FAM20A in secretory-stage ameloblasts, odontoblasts, tooth eruption pathway, and in renal tubules of the kidney, which explained some of the disease phenotypes (Fig. 3.16). The result of this work has been published in 2 papers (Wang *et al.*, 2013a; Wang *et al.*, 2014a).

DISCUSSION

Hypoplastic AI and *MMP20* mutations

Hypoplastic AI refers to a thickness defect of dental enamel (thin enamel). Since the full thickness of enamel is established by appositional growth during secretory stage of amelogenesis, hypoplastic AI is usually caused by mutations in genes which play a significant role at the secretory stage, such as genes encoding enamel matrix proteins, *AMELX* (amelogenin) and *ENAM* (enamelin) (Hu *et al.*, 2007). Noticeably, an enamel-specific protease, MMP20, is also expressed by secretory-stage ameloblasts (Hu *et al.*, 2002). To date, there are a total of 7 disease-causing *MMP20* mutations reported, including the one we identified in our Family 1 (Gasse *et al.*, 2013). All of them appear to be loss-of-function mutations, suggesting that failed proteolytic processing of enamel matrix proteins at secretory stage is the cause of enamel defects. However, there is a high variation of enamel phenotypes among the reported cases.

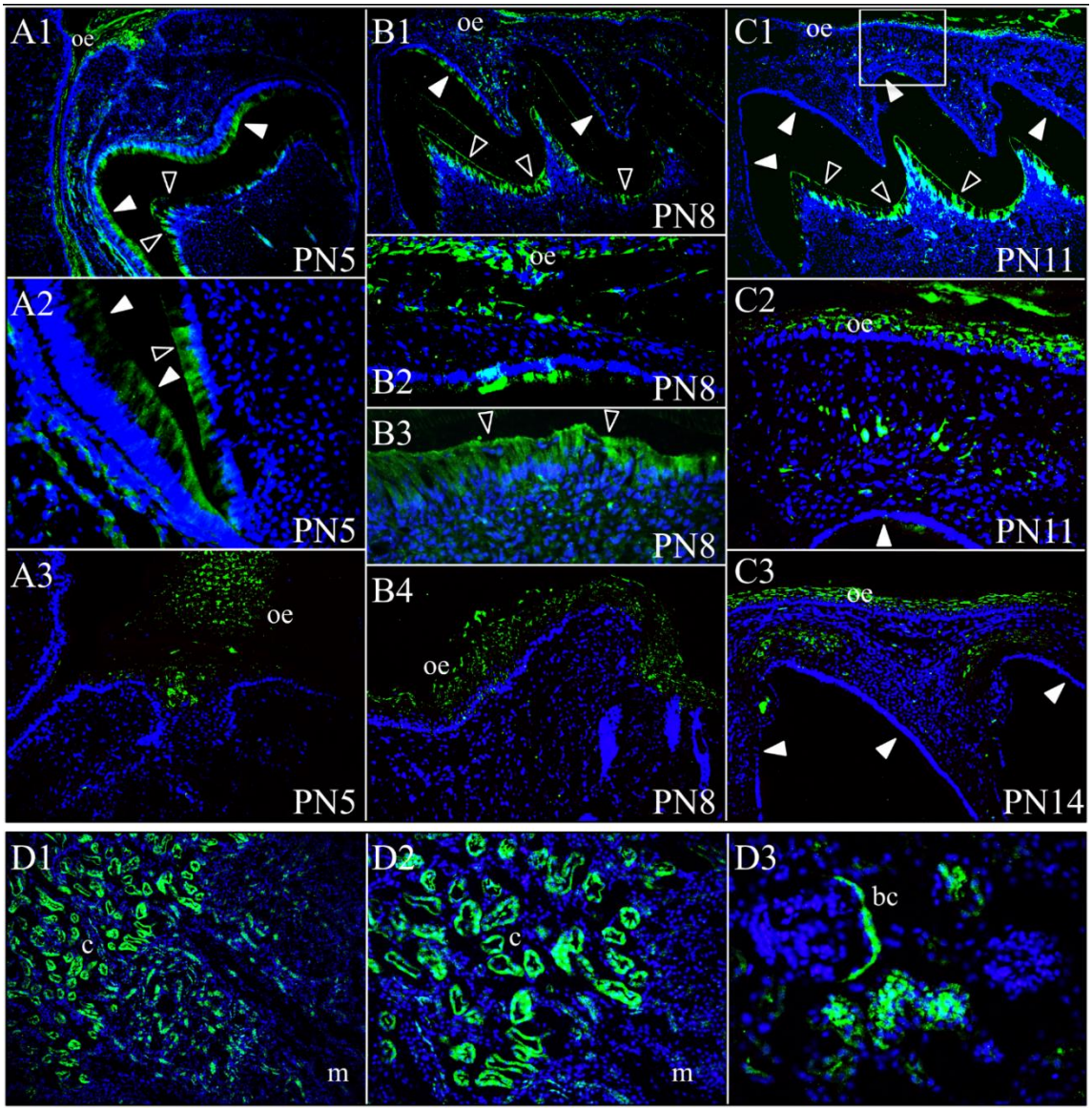


Figure 3.16: FAM20A immunohistochemistry.
Top*:** FAM20A localization during mouse first maxillary molar development. ***A1-A3: At post-natal day 5 (PN5), FAM20A signal was observed in secretory-stage ameloblasts, odontoblasts, and in oral epithelium (A3). ***B1-B4:*** At post-natal day 8 (PN8), FAM20A signal was minimal in transition or early maturation-stage ameloblasts, but strong in odontoblasts and the oral epithelium (B2 and B4). ***C1-C2:*** At post-natal day 11 (PN11), no FAM20A signal was observed in maturation-stage ameloblasts. FAM20A was detected in odontoblasts and in oral epithelium. A band of cells positive for FAM20A was observed in the connective tissue between the oral epithelium and the developing cusp tip. The box in C1 corresponds to the position of panel C2. ***C3:*** On post-natal day 14 (PN14), just prior to eruption, FAM20A signal was evident in the dental follicle above the molar cusp tips. ***D1-D3:*** FAM20A localization in four-week-old mouse kidney. ***Key*:** bc, Bowman's capsule; c, renal cortex; m, renal medulla; oe, oral epithelium. Blue is DAPI-stained nuclei, green is FAM20A signal, solid arrowheads indicate ameloblasts, and hollow arrowheads indicate odontoblasts. This image is reproduced from Figure 4 of reference (Wang SK, Reid BM, Dugan SL, Roggenbuck JA, Read L, Aref P, Taheri AP, Yeganeh MZ, Simmer JP, Hu JC (2014). FAM20A Mutations Associated with Enamel Renal Syndrome. *J Dent Res* 93:42-8.) by permission of SAGE Journals.

For instance, while the proband of Family 1, who had a p.His204Arg *MMP20* mutation, showed a typical hypoplastic AI phenotype with very thin enamel, the patient with p.Thr130Ile mutation reported by Gasse *et al.* had a typical hypomaturation AI phenotype (Gasse *et al.*, 2013). Some cases even showed combined hypoplastic-hypomaturation phenotypes (Papagerakis *et al.*, 2008).

Theoretically, one would expect that loss-of-function mutations in genes encoding enamel matrix proteases, *MMP20* and *KLK4*, would impede ion deposition at maturation stage due to defective protein degradation and lead to hypomaturation AI. However, the hypoplastic phenotype of Family 1 proband and the thin enamel of *Mmp20* null mice suggest that MMP20 may have functions other than degradation of matrix protein during secretory stage of enamel formation (Caterina *et al.*, 2002). It has been shown that *Mmp20* is expressed by preameloblasts, secretory-stage ameloblasts, and odontoblasts and the protease concentrates at the dentino-enamel junction (DEJ) in developing teeth (Hu *et al.*, 2002; Wang *et al.*, 2013c). Therefore, it is possible that MMP20 might be responsible for degrading the basement membrane (BM) prior to the secretory stage, and that failed BM degradation due to defective MMP20 function might impede matrix secretion and lead to a hypoplastic phenotype. However, further investigations need to be conducted to test this hypothesis.

Hypoplastic AI and *ITGB6* mutations

We reported the first human *ITGB6* (integrin beta 6) mutations causing autosomal recessive hypoplastic AI. Along with the enamel defects in *Itgb6* null mice, our finding demonstrated a significant role of integrin β6 in enamel formation (Wang *et al.*, 2013d). Moreover, our immunohistostaining of mouse developing incisors showed that ITGB6 was

present on the ameloblast distal membrane at the onset of enamel formation and during formation of the initial enamel, and appeared to be internalized about the time ameloblasts developed their Tomes' processes (Fig. 3.4). This window of expression, along with the previous finding that *Itgb6* null mice overexpressed the secreted enamel proteins amelogenin (21-fold) and enamelin (7.6-fold) but failed to make enamel rods, suggests that ITGB6 signaling regulates the onset of enamel formation by repressing the expression of enamel proteins and also regulates formation of the Tomes' process (an ameloblast distal membrane structure necessary for enamel rod/interrod organization) (Mohazab *et al.*, 2013; Wang *et al.*, 2013d).

Several integrins have been shown to be expressed at different stages of murine tooth development, including ITGA6 and ITGB4 (integrin α6β4), two protein components of hemidesmosome-anchoring filament complexes (Salmivirta *et al.*, 1996). Mutations in genes encoding these hemidesmosome complex proteins (*LAMA3*, *LAMB3*, *LAMC2*, *COL17A1*, *ITGA6* and *ITGB4*) cause junctional epidermolysis bullosa (JEB) in which enamel defects are one of the manifestations (Masunaga, 2006; Fine, 2010). Also, heterozygous carriers with these genetic defects were also reported to have enamel malformations without non-dental phenotypes. In this case, the affected enamel usually shows pitted defects instead of generalized hypoplasia (Kim *et al.*, 2013; Poulter *et al.*, 2014). However, in contrast, our patients with recessive *ITGB6* mutations exhibited generalized hypoplastic enamel, suggesting that integrin β6, unlike integrin α6β4, may have functions other than mechanically attaching epithelium to extracellular matrices by hemidesmosome protein complexes.

Soon after we reported the *ITGB6* mutations, Poulter *et al.* also reported an *ITGB6* mutation (p.Pro196Thr) causing enamel defects (Poulter *et al.*, 2013). However, the patient's enamel phenotype was quite different from those of our two cases. While our patients showed

generalized hypoplastic enamel, theirs had a phenotype of pitted hypomineralised AI. The basis of this phenotypic difference is unclear. More cases of *ITGB6* mutations and further investigations are needed to elucidate the pathological mechanism of *ITGB6*-associated AI.

Enamel maturation and hypomaturation AI

The hardness of dental enamel is established at the maturation stage of amelogensis, in which two principal activities were involved: degradation and re-absorption of the organic matrix and regulated movement of ions into and out of the matrix (Smith, 1998; Hu *et al.*, 2007; Simmer *et al.*, 2010). During the maturation stage, residual enamel proteins are degraded by KLK4 and, to a lesser extent, by MMP20. Therefore, mutations in *KLK4* and *MMP20* can cause hypomaturation AI (Hart *et al.*, 2004; Kim *et al.*, 2005). In this study, we reported the second human *KLK4* mutation (g.6930delG; c.245delG; p.Gly82Alafs*87), which reaffirmed the significant role of KLK4 in matrix protein degradation and enamel maturation (Wang *et al.*, 2013b).

In addition to degradation and re-absorption of the residual proteins, ion deposition, particularly calcium ion, onto the sides of enamel crystallites is the fundamental basis of enamel maturation. However, the underlying mechanism of calcium transport during amelogenesis is still largely unknown. STIM1 (stromal interaction molecule 1) and ORAI1 (ORAI calcium release-activated calcium modulator 1) mediate store-operated calcium entry (SOCE), which is a Ca^{2+} influx pathway critical for the normal functioning of many cell types, including T cells, muscle cells, and ameloblasts (Shaw *et al.*, 2013; Collins *et al.*, 2013). SLC24A4 is a potassium-dependent sodium-calcium exchanger reported to be expressed by ameloblasts (Hu *et al.*, 2012).

Genetic defects in human *STIM1*, *ORAI1*, and *SLC24A4* cause hypomaturation AI, demonstrating the critical roles of these three proteins in calcium transport during enamel maturation (Feske *et al.*, 2006; Picard *et al.*, 2009; Parry *et al.*, 2013). In this study, we identified novel *STIM1* and *SLC24A4* mutations in families with hypomaturation AI and showed, by immunohistostaining, that these two genes were specifically expressed by maturation-stage but not secretory-stage ameloblasts (Fig. 3.7; Fig. 3.9), suggesting that there are important differences in the calcium transcellular transport systems employed by secretory and maturation stage ameloblasts. We also proposed a working hypothesis that explains how calcium from the blood supply might be drawn through maturation stage ameloblasts to support mineralization (Wang *et al.*, 2014b). Perhaps SLC24A4 is responsible for the active transport of calcium ions out of the cell and into the enamel matrix by using the energy of a Na^+ gradient, which in turn lowers intracellular Ca^{2+}, depletes ER stores, and activates the SOCE system (STIM1/ORAI1) to replenish intracellular Ca^{2+} stores at the proximal side of maturation stage ameloblasts, nearest the blood supply (Wang *et al.*, 2014b).

Autosomal dominant hypocalcified AI and *FAM83H* mutations

Since the first *FAM83H* mutations were identified to cause autosomal dominant hypocalcified AI (ADHCAI) in 2008 (Kim *et al.*, 2008), many ADHCAI kindreds and disease-causing *FAM83H* mutations have been sequentially reported (Fig. 1.4). In Family 7, we identified a novel *FAM83H* mutation (p.Gln457*). Interestingly, all these disease-causing *FAM83H* mutations are either nonsense mutations or frameshifts leading to a premature stop codon. No other types of loss-of-function mutations have been reported. Also, all the reported mutations are located at a specific 5' region of Exon 5 (the last exon) and expected to produce a

truncated FAM83H protein with a specific range of lengths (from 287 to 694 amino acids) (Fig. 1.4). This genetic homogeneity of mutational spectrum highly suggests a dominant negative effect or a "gain-of-function" rather than haploinsufficiency as the pathological mechanism of ADHCAI.

This hypothesis is further supported by the reported nonsense sequence variants of *FAM83H* in general population. According to dbSNP, 1000 Genome Project, and NHLBI Exome Sequencing Project databases, several nonsense *FAM83H* mutations in earlier exons (Exons 2, 3, and 4) have been identified in general populations, with one of which (p.Gln201*; rs189033490) has a minor allele frequency of 0.1%. These early nonsense mutations, unlike those at the last exon, are expected to produce a mutant transcript which will undergo nonsense mediated decay and lead to a complete null allele. In other words, loss of half FAM83H doses seems not to cause a disease phenotype, meaning haploinsufficiency is not the pathological mechanism of ADHCAI. In contrast, only truncation mutations between p.Ser287* and p.Glu694* were reported to be disease-causing, which supports the hypothesis of dominant negative effects. For nonsense mutations at further 3' end of *FAM83H* Exon 5 (longer than p.Glu694*), there are two (p.Glu883*, p.Glu1156*) reported so far in COSMIC (Catalogue of Somatic Mutations in Cancer) database, suggesting that these mutations seem not to cause a disease phenotype as those at earlier exons.

In Chapter 4, we will discuss the molecular characterization of FAM83H protein and propose a working hypothesis of why only mutations between p.Ser287* and p.Glu694* are disease-causing and how mutant truncated FAM83Hs exert a dominant negative effect.

AIGFS, ERS, and FAM20A mutations

In this thesis, we reported 6 novel and 1 reported *FAM20A* mutations in 5 families and demonstrated that Enamel-Renal Syndrome (ERS) are caused by *FAM20A* mutations. Based upon the similar dental phenotypes between AIGFS and ERS, we also proposed that these two disorders may be of the same disease entity but with variable presentation of nephrocalcinosis (Wang *et al.*, 2013a; Wang *et al.*, 2014a).

FAM20A belongs to a small gene family that in human and mouse has three members: *FAM20A*, *FAM20B*, and *FAM20C*. Proteins encoded by this gene family are all predicted to have a potential kinase domain (Nalbant *et al.*, 2005). FAM20B was the first one in the family shown to have kinase activity: FAM20B was demonstrated to be a xylose kinase, required in the Golgi for the efficient addition of glycan attachments on secreted proteins (Koike *et al.*, 2009). More recently, FAM20C was identified as Golgi Casein Kinase (GCK) responsible for phosphorylation of many SCPP proteins (secretory calcium-binding phosphoproteins) critical for biomineralization (Tagliabracci *et al.*, 2012). Genetic defects in human *FAM20C* cause Raine syndrome (OMIM #259775), an autosomal recessive disorder characterized by major defects in biomineralization (Simpson *et al.*, 2007). In terms of FAM20A, although its defects have been shown to cause AIGFS and ERS (O'Sullivan *et al.*, 2011; Jaureguiberry *et al.*, 2012; Wang *et al.*, 2013a), its actual function is still unknown. To date, all the *FAM20A* mutations are loss-of-function mutations, with some of which are missense mutations of only one amino acid substitution (Wang *et al.*, 2014a). Particularly, some of these missense mutations happen at an amino acid position in which corresponding *FAM20C* mutations cause loss of kinase activity. For example, the *FAM20A* mutation (p.Asp403Asn) we identified in Family 11 has a corresponding *FAM20C* mutation (p.Asp451Asn) which has been shown to abolish FAM20C

kinase activity *in vitro* and to be disease-causing (Wang *et al.*, 2014a). These missense mutations highly suggest that FAM20A is also a kinase, although its substrates still need to be identified.

Assuming that FAM20A is indeed a kinase, our immunohistostaining further suggests a significant role of FAM20A-dependent phosphorylations in different stages of tooth and enamel development (Fig. 3.16). We demonstrate that FAM20A is expressed by secretory-stage but not maturation-stage ameloblasts, which explains the hypoplastic enamel phenotype, when both *FAM20A* alleles are defected. FAM20A is expressed in the dental follicle just above the molar cusp tips and absence of FAM20A causes failures in tooth eruption, suggesting that FAM20A-dependent phosphorylations may play a role in forming eruption pathways, although, tooth eruption seems to be relatively unaffected in the human primary dentition and in *Fam20a* null mice (Vogel *et al.*, 2012). Gingival fibromatosis is another feature often reported in *FAM20A* mutation patients. This manifestation might be associated with FAM20A expression in selected cells within the gingival tissues, rather than in the oral epithelium. We show that some *FAM20A* mutation patients have asymptomatic nephrocalcinosis and find FAM20A expression in renal tubules of adult mouse kidneys. *Fam20a* null mice exhibited widespread and severe ectopic calcification throughout the body, but most prominantly in the kidneys. Normal blood calcium and phosphorus in mutant mice and human patients suggested that the ectopic calcification might be caused by local rather than systemic effects (Vogel *et al.*, 2012).

Phenotypic and genotypic heterogeneity in AI

To date, many candidate genes, including *AMELX*, *ENAM*, *MMP20*, *KLK4*, *FAM83H*, *WDR72*, *C4orf26*, *SLC24A4*, and *LAMB3*, have been identified to be associated with isolated AI.

However, the genetic defects of about half of the AI cases are still unknown, indicating that there are more genes involved in amelogenesis waiting to be identified (Chan *et al.*, 2011). This high genetic heterogeneity of AI also suggests that amelogenesis is a delicate developmental process which is highly susceptible to many genetic and environmental alterations. For instance, *FAM83H* and *WDR72* are expressed by many tissues other than developing teeth. However, Patients with genetic defects in these two genes only show enamel malformations without clinically-detectable manifestations in non-dental tissues, suggesting the vulnerability of enamel formation (Kim *et al.*, 2008; El-Sayed *et al.*, 2009). Moreover, just because of this vulnerability, many syndromes have AI as one of the manifestations (Hu *et al.*, 2007; Cobourne and Sharpe, 2013). However, it is important for us to differentiate if the enamel defects in syndromes are due to primary or secondary effects. Many syndromes with kidney problems are also associated with enamel defects due to defective ion regulation (Subramaniam *et al.*, 2012). In other words, the enamel phenotype comes from the secondary effect of impaired renal functions rather than the genetic defect having direct impact on enamel formation.

In addition to genotypic heterogeneity, there is also high variability in AI phenotypes. As we mentioned previously, there are mainly three types of AI, featured by different characteristics of enamel defects. Even with the same genetic defects, the phenotypes can still be highly variable. *MMP20*-associated enamel defects are the best example of this high phenotypic variability (See the above discussion section of *MMP20* mutations). Sometimes, this variation is further exaggerated by the post-eruptional changes, such as discoloration and attrition. Therefore, for mutational analysis of AI, it is not practical to establish genotype-phenotype correlation due to high genotypic and phenotypic heterogeneities, which poses a problem in defining genetic etiology of many AI cases, especially the sporadic ones. Outside this thesis, our lab also analyzed

the exomes from the probands of different AI families separately and then collectively but still failed to identify the genetic defects in about half of the cases, although it is possible that the mutations are at the regulatory region of the gene, which is not easy to identify.

Prospects in mutational analysis of AI

As we discussed in Chapter 2 (mutational analysis of familial tooth agenesis), due to the genotypic and phenotypic heterogeneity of the disease, the genetic etiology of a significant number of cases still waits to be determined. In the future, for large-sized families, we should take advantage of the genetic linkage power and analyze with exome sequencing data to identify the disease-causing mutations. For small families or sporadic cases, we need to recruit multiple families with relatively similar phenotypes and collectively analyze their exome data at a "genetic pathway level."

Furthermore, enamel defects found in specific gene knockout mice can be highly suggestive of potential AI candidate genes in humans. For instance, we established *ITGB6* as a candidate gene for isolated AI, based upon the report that *Itgb6* null mice exhibited enamel malformations (Mohazab *et al.*, 2013). This approach demonstrates the power of reverse genetics (from genotype to phenotype) in unraveling the genetic etiology of human AI. Therefore, mouse models may provide supportive evidence for AI candidate genes we identify through human mutational analysis. Recently, the knockout mouse project (KOMP) and KOMP phenotyping plan were funded by National Human Genome Research Institute (NHGRI) and National Center for Research Resources (NCRR). These projects include a concerted, centralized, high-throughput phenotyping effort to extend the scientific value of the knockout ES cell resources,

which is also a precious resource for human genetics (Ayadi *et al.*, 2012; White *et al.*, 2013). Unfortunately, characterization of the dental phenotype is not part of these projects. If knockout mice could be routinely screened for apparent dental phenotypes, it could significantly advance our understanding of normal and pathological tooth development.

MATERIALS AND METHODS

For "Family recruitment and ethics statement," "Phenotypic data collection and family pedigree construction," "DNA extraction from blood and saliva samples," and "Mutational analysis (target gene approach)," please refer to the same sections in Chapter 2.

Mutational analysis (whole exome sequencing)

Whole exome sequencing was conducted at Yale Center for Genome Analysis (West Haven, CT, U.S.A.) and Edge BioSystems (Gaithersburg, MD, U.S.A.). For experimental processes at Yale Center for Genome Analysis, please refer to the same section in Chapter 2. For whole exome sequencing in Edge BioSystems, exome capture was conducted with Aligent SureSelect Target Enrichment System (Agilent Technologies; Santa Clara, CA, U.S.A.). For sequencing, 3μg of genomic DNA was nebulized into ~300 bp fragments, which were tailed and ligated to adapters, amplified, and analyzed using an Agilant 2100 Bioanalyzer prior to sequencing in paired-end reads, 75-100 bases per read (Life Technologies SOLiD™ System; Carlsbad, CA, U.S.A.).

Tissue preparation and sample sectioning for immunohistochemistry

Postnatal day 5, 8, 11, and 14 mouse heads were quickly dissected of skin, cut in half, and immersed in 4% paraformaldehyde fixative overnight at 4°C, washed in PBS 4-5 times at 4°C, and decalcified at 4°C by immersion in 1 L of 4.13% disodium ethylenediaminetetraacetic acid (EDTA, pH 7.3) with agitation. The EDTA solution was changed every other day for 8-9 d for day-5 mice, 19-21 d for day-8 mice, and 30 d for day-11 and day-14 mice. Kidneys from 4-week-old mice were dissected and fixed as described above. For cryosectioning, after decalcification the tissues were immersed in 30% sucrose overnight for cryoprotection and embedded in OCT/Tissue Tek (Sakura Finetek; Torrance, CA, U.S.A.). The blocks were cryosectioned at 10μm thickness at -20 °C. For paraffin-sectioning, samples underwent regular dehydration process and were embedded in paraffin. The blocks were sectioned at 5μm thickness.

Immunohistochemistry

For paraffin sections, prior to immunostaining, the sections were deparaffinized and heat-treated (95°C) with antigen retrieval solution (ab973; Abcam; Cambridge, MA, U.S.A.) for 30 min. For cryosections, the slides can be proceeded directly for immunostaining without prerequisite preparation.

The slides were rinsed with PBT buffer (0.1% Triton X-100 in PBS), blocked with 5% sheep serum (S22; Millipore Corporation; Billerica, MA, U.S.A.) in PBT for 30 min at room temperature, and serial sections were incubated overnight with primary antibodies in blocking solution. The sections were washed with PBT for 15 min and incubated for 30 min at room

temperature in solutions containing anti-rabbit IgG secondary antibody conjugated with Alexa Fluor 594 (1:500; A-11012; Molecular Probes® by Life Technologies). Sections were rinsed in PBT for 15 min, mounted with ProLong® Gold antifade reagent with DAPI (P-36941; Molecular Probes® by Life Technologies), and examined using an Olympus BX51 with fluorescence attachments and photographed using an Olympus DP71 camera with DP controller and manager software.

Primary antibodies used for immunohistochemistry include: anti-ITGB6 (HPA023626) (1:100), anti-FAM20A (SAB2100759) (1:100), anti-STIM1 (HPA012123) (1:100) from Sigma-Aldrich (St. Louis, MO, U.S.A.); anti-SLC24A4 (ab136968) (1:100) from abcam® (Cambridge, MA, U.S.A.)

Characterization of extracted teeth from Family 10 proband

The Micro-CT, SEM, and Backscatter SEM analyses were performed by Dr. Yuanyuan Hu. For detailed experimental processes, please refer to (Wang *et al.*, 2013a).

REFERENCES

Aldred MJ, Crawford PJ, Roberts E, Thomas NS (1992). Identification of a nonsense mutation in the amelogenin gene (AMELX) in a family with X-linked amelogenesis imperfecta (AIH1). *Hum Genet* 90:413-6.

Aldred MJ, Savarirayan R, Crawford PJ (2003). Amelogenesis imperfecta: a classification and catalogue for the 21st century. *Oral Dis* 9:19-23.

Ayadi A, Birling MC, Bottomley J, Bussell J, Fuchs H, Fray M, Gailus-Durner V, Greenaway S, Houghton R, Karp N, Leblanc S, Lengger C, Maier H, Mallon AM, Marschall S, Melvin D, Morgan H, Pavlovic G, Ryder E, Skarnes WC, Selloum M, Ramirez-Solis R, Sorg T, Teboul L, Vasseur L, Walling A, Weaver T, Wells S, White JK, Bradley A, Adams DJ, Steel KP, Hrabě de Angelis M, Brown SD, Herault Y (2012). Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mamm Genome* 23:600-10.

Bartlett JD (2013). Dental Enamel Development: Proteinases and Their Enamel Matrix Substrates. *ISRN Dent* 2013:684607.

Buchroithner B, Klausegger A, Ebschner U, Anton-Lamprecht I, Pohla-Gubo G, Lanschuetzer CM, Laimer M, Hintner H, Bauer JW (2004). Analysis of the LAMB3 gene in a junctional epidermolysis bullosa patient reveals exonic splicing and allele-specific nonsense-mediated mRNA decay. *Lab Invest* 84:1279-88.

Burdon KP, McKay JD, Sale MM, Russell-Eggitt IM, Mackey DA, Wirth MG, Elder JE, Nicoll A, Clarke MP, FitzGerald LM, Stankovich JM, Shaw MA, Sharma S, Gajovic S, Gruss P, Ross S, Thomas P, Voss AK, Thomas T, Gécz J, Craig JE (2003). Mutations in a novel gene, NHS, cause the pleiotropic effects of Nance-Horan syndrome, including severe congenital cataract, dental anomalies, and mental retardation. *Am J Hum Genet* 73:1120-30.

Caterina JJ, Skobe Z, Shi J, Ding Y, Simmer JP, Birkedal-Hansen H, Bartlett JD (2002). Enamelysin (matrix metalloproteinase 20)-deficient mice display an amelogenesis imperfecta phenotype. J Biol Chem 277:49598-604.

Chan HC, Mai L, Oikonomopoulou A, Chan HL, Richardson AS, Wang SK, Simmer JP, Hu JC (2010). Altered enamelin phosphorylation site causes amelogenesis imperfecta. *J Dent Res* 89:695-9.

Chan HC, Estrella NM, Milkovich RN, Kim JW, Simmer JP, Hu JC (2011). Target gene analyses of 39 amelogenesis imperfecta kindreds. *Eur J Oral Sci* 119 Suppl 1:311-23.

Cobourne MT, Sharpe PT (2013). Diseases of the tooth: the genetic and molecular basis of inherited anomalies affecting the dentition. *Wiley Interdiscip Rev Dev Biol* 2:183-212.

Coffield KD, Phillips C, Brady M, Roberts MW, Strauss RP, Wright JT (2005). The psychosocial impact of developmental dental defects in people with hereditary amelogenesis imperfecta. *J Am Dent Assoc* 136:620-30.

Collins HE, Zhu-Mauldin X, Marchase RB, Chatham JC (2013). STIM1/Orai1-mediated SOCE: current perspectives and potential roles in cardiac function and pathology. Am J Physiol Heart Circ Physiol 305:H446-58.

El-Sayed W, Parry DA, Shore RC, Ahmed M, Jafri H, Rashid Y, Al-Bahlani S, Al Harasi S, Kirkham J, Inglehearn CF, Mighell AJ (2009). Mutations in the beta propeller WDR72 cause autosomal-recessive hypomaturation amelogenesis imperfecta. *Am J Hum Genet* 85:699-705.

Feske S, Gwack Y, Prakriya M, Srikanth S, Puppel SH, Tanasa B, Hogan PG, Lewis RS, Daly M, Rao A (2006). A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. Nature 441:179-85.

Fine JD (2010). Inherited epidermolysis bullosa. *Orphanet J Rare Dis* 5:12.

Gasse B, Karayigit E, Mathieu E, Jung S, Garret A, Huckert M, Morkmued S, Schneider C, Vidal L, Hemmerlé J, Sire JY, Bloch-Zupan A (2013). Homozygous and compound heterozygous MMP20 mutations in amelogenesis imperfecta. *J Dent Res* 92:598-603.

Hart PS, Hart TC, Michalec MD, Ryu OH, Simmons D, Hong S, Wright JT (2004). Mutation in kallikrein 4 causes autosomal recessive hypomaturation amelogenesis imperfecta. *J Med Genet* 41:545-9.

Hu JC, Sun X, Zhang C, Liu S, Bartlett JD, Simmer JP (2002). Enamelysin and kallikrein-4 mRNA expression in developing mouse molars. *Eur J Oral Sci* 110:307-15.

Hu JC, Chun YH, Al Hazzazzi T, Simmer JP (2007). Enamel formation and amelogenesis imperfecta. *Cells Tissues Organs* 186:78-85.

Hu P, Lacruz RS, Smith CE, Smith SM, Kurtz I, Paine ML (2012). Expression of the sodium/calcium/potassium exchanger, NCKX4, in ameloblasts. *Cells Tissues Organs* 196:501-9.

Jaureguiberry G, De la Dure-Molla M, Parry D, Quentric M, Himmerkus N, Koike T, Poulter J, Klootwijk E, Robinette SL, Howie AJ, Patel V, Figueres ML, Stanescu HC, Issler N, Nicholson JK, Bockenhauer D, Laing C, Walsh SB, McCredie DA, Povey S, Asselin A, Picard A, Coulomb A, Medlar AJ, Bailleul-Forestier I, Verloes A, Le Caignec C, Roussey G, Guiol J, Isidor B, Logan C, Shore R, Johnson C, Inglehearn C, Al-Bahlani S, Schmittbuhl M, Clauss F, Huckert M, Laugel V, Ginglinger E, Pajarola S, Spartà G, Bartholdi D, Rauch A, Addor MC, Yamaguti PM, Safatle HP, Acevedo AC, Martelli-Júnior H, dos Santos Netos PE, Coletta RD, Gruessel S, Sandmann C, Ruehmann D, Langman CB, Scheinman SJ, Ozdemir-Ozenen D, Hart TC, Hart PS, Neugebauer U, Schlatter E, Houillier P, Gahl WA, Vikkula M, Bloch-Zupan A, Bleich M, Kitagawa H, Unwin RJ, Mighell A, Berdal A, Kleta R (2012). Nephrocalcinosis (enamel renal syndrome) caused by autosomal recessive FAM20A mutations. *Nephron Physiol* 122:1-6.

Kim JW, Simmer JP, Hart TC, Hart PS, Ramaswami MD, Bartlett JD, Hu JC (2005). MMP-20 mutation in autosomal recessive pigmented hypomaturation amelogenesis imperfecta. *J Med Genet* 42:271-5.

Kim JW, Simmer JP, Lin BP, Seymen F, Bartlett JD, Hu JC (2006). Mutational analysis of candidate genes in 24 amelogenesis imperfecta families. *Eur J Oral Sci* 114 Suppl 1:3-12; discussion 39-41, 379.

Kim JW, Lee SK, Lee ZH, Park JC, Lee KE, Lee MH, Park JT, Seo BM, Hu JC, Simmer JP (2008). FAM83H mutations in families with autosomal-dominant hypocalcified amelogenesis imperfecta. *Am J Hum Genet* 82:489-94.

Kim JW, Seymen F, Lee KE, Ko J, Yildirim M, Tuna EB, Gencay K, Shin TJ, Kyun HK, Simmer JP, Hu JC (2013). LAMB3 mutations causing autosomal-dominant amelogenesis imperfecta. J Dent Res 92:899-904.

Koike T, Izumikawa T, Tamura J, Kitagawa H (2009). FAM20B is a kinase that phosphorylates xylose in the glycosaminoglycan-protein linkage region. *Biochem J* 421:157-62.

Masunaga T (2006). Epidermal basement membrane: its molecular organization and blistering disorders. *Connect Tissue Res* 47:55-66.

Mohazab L, Koivisto L, Jiang G, Kytömäki L, Haapasalo M, Owen GR, Wiebe C, Xie Y, Heikinheimo K, Yoshida T, Smith CE, Heino J, Häkkinen L, McKee MD, Larjava H (2013). Critical role for αvβ6 integrin in enamel biomineralization. *J Cell Sci* 126:732-44.

Nalbant D, Youn H, Nalbant SI, Sharma S, Cobos E, Beale EG, Du Y, Williams SC (2005). FAM20: an evolutionarily conserved family of secreted proteins expressed in hematopoietic cells. *BMC Genomics* 6:11.

Nanci A (2008a). Development of the tooth and its supporting tissues. In: Ten Cate's Oral Histology Development, Structure, and Function. Nanci A editor. St. Louis: Mosby, pp.79-107.

Nanci A (2008b). Enamel: composition, formation, and structure. In: Ten Cate's Oral Histology Development, Structure, and Function. Nanci A editor. St. Louis: Mosby, pp.141-90.

O'Sullivan J, Bitu CC, Daly SB, Urquhart JE, Barron MJ, Bhaskar SS, Martelli-Júnior H, dos Santos Neto PE, Mansilla MA, Murray JC, Coletta RD, Black GC, Dixon MJ (2011). Whole-Exome sequencing identifies FAM20A mutations as a cause of amelogenesis imperfecta and gingival hyperplasia syndrome. *Am J Hum Genet* 88:616-20.

Papagerakis P, Lin HK, Lee KY, Hu Y, Simmer JP, Bartlett JD, Hu JC (2008). Premature stop codon in MMP20 causing amelogenesis imperfecta. *J Dent Res* 87:56-9.

Parry DA, Brookes SJ, Logan CV, Poulter JA, El-Sayed W, Al-Bahlani S, Al Harasi S, Sayed J, Raïf el M, Shore RC, Dashash M, Barron M, Morgan JE, Carr IM, Taylor GR, Johnson CA, Aldred MJ, Dixon MJ, Wright JT, Kirkham J, Inglehearn CF, Mighell AJ (2012). Mutations in

C4orf26, encoding a peptide with in vitro hydroxyapatite crystal nucleation and growth activity, cause amelogenesis imperfecta. *Am J Hum Genet* 91:565-71.

Parry DA, Poulter JA, Logan CV, Brookes SJ, Jafri H, Ferguson CH, Anwari BM, Rashid Y, Zhao H, Johnson CA, Inglehearn CF, Mighell AJ (2013). Identification of mutations in SLC24A4, encoding a potassium-dependent sodium/calcium exchanger, as a cause of amelogenesis imperfecta. *Am J Hum Genet* 92:307-12.

Picard C, McCarl CA, Papolos A, Khalil S, Lüthy K, Hivroz C, LeDeist F, Rieux-Laucat F, Rechavi G, Rao A, Fischer A, Feske S (2009). STIM1 mutation associated with a syndrome of immunodeficiency and autoimmunity. *N Engl J Med* 360:1971-80.

Poulter JA, Brookes SJ, Shore RC, Smith CE, Abi Farraj L, Kirkham J, Inglehearn CF, Mighell AJ (2013). A missense mutation in ITGB6 causes pitted hypomineralized amelogenesis imperfecta. *Hum Mol Genet* [Epub ahead of print]

Poulter JA, El-Sayed W, Shore RC, Kirkham J, Inglehearn CF, Mighell AJ (2014). Whole-exome sequencing, without prior linkage, identifies a mutation in LAMB3 as a cause of dominant hypoplastic amelogenesis imperfecta. *Eur J Hum Genet* 22:132-5.

Rajpar MH, Harley K, Laing C, Davies RM, Dixon MJ (2001). Mutation of the gene encoding the enamel-specific protein, enamelin, causes autosomal-dominant amelogenesis imperfecta. *Hum Mol Genet* 10:1673-7.

Salmivirta K, Gullberg D, Hirsch E, Altruda F, Ekblom P (1996). Integrin subunit expression associated with epithelial-mesenchymal interactions during murine tooth development. *Dev Dyn* 205:104-13.

Shaw PJ, Qu B, Hoth M, Feske S (2013). Molecular regulation of CRAC channels and their role in lymphocyte function. *Cell Mol Life Sci* 70:2637-56.

Simmer JP, Papagerakis P, Smith CE, Fisher DC, Rountrey AN, Zheng L, Hu JC (2010). Regulation of dental enamel shape and hardness. *J Dent Res* 89:1024-38.

Simpson MA, Hsu R, Keir LS, Hao J, Sivapalan G, Ernst LM, Zackai EH, Al-Gazali LI, Hulskamp G, Kingston HM, Prescott TE, Ion A, Patton MA, Murday V, George A, Crosby AH (2007). Mutations in FAM20C are associated with lethal osteosclerotic bone dysplasia (Raine syndrome), highlighting a crucial molecule in bone development. *Am J Hum Genet* 81:906-12.

Smith CE (1998). Cellular and chemical events during enamel maturation. *Crit Rev Oral Biol Med* 9:128-61.

Subramaniam P, Gupta M, Mehta A (2012). Oral health status in children with renal disorders. *J Clin Pediatr Dent* 37:89-93.

Tagliabracci VS, Engel JL, Wen J, Wiley SE, Worby CA, Kinch LN, Xiao J, Grishin NV, Dixon JE (2012). Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science* 336:1150-3.

Thesleff I (2013). Current understanding of the process of tooth formation: transfer from the laboratory to the clinic. *Aust Dent J* [Epub ahead of print]

Tucker A, Sharpe P (2004). The cutting-edge of mammalian development; how the embryo makes teeth. *Nat Rev Genet* 5:499-508.

Vogel P, Hansen GM, Read RW, Vance RB, Thiel M, Liu J, Wronski TJ, Smith DD, Jeter-Jones S, Brommage R (2012). Amelogenesis imperfecta and other biomineralization defects in Fam20a and Fam20c null mice. *Vet Pathol* 49:998-1017.

Wang SK, Aref P, Hu Y, Milkovich RN, Simmer JP, El-Khateeb M, Daggag H, Baqain ZH, Hu JC (2013a). FAM20A mutations can cause enamel-renal syndrome (ERS). *PLoS Genet* 9:e1003302.

Wang SK, Hu Y, Simmer JP, Seymen F, Estrella NM, Pal S, Reid BM, Yildirim M, Bayram M, Bartlett JD, Hu JC (2013b). Novel KLK4 and MMP20 mutations discovered by whole-exome sequencing. *J Dent Res* 92:266-71.

Wang SK, Samann AC, Hu JC, Simmer JP (2013c). FAM20C functions intracellularly within both ameloblasts and odontoblasts in vivo. *J Bone Miner Res* 28:2508-11.

Wang SK, Choi M, Richardson AS, Reid BM, Lin BP, Wang SJ, Kim JW, Simmer JP, Hu JC (2013d). ITGB6 loss-of-function mutations cause autosomal recessive amelogenesis imperfecta. *Hum Mol Genet* [Epub ahead of print]

Wang SK, Reid BM, Dugan SL, Roggenbuck JA, Read L, Aref P, Taheri AP, Yeganeh MZ, Simmer JP, Hu JC (2014a). FAM20A Mutations Associated with Enamel Renal Syndrome. *J Dent Res* 93:42-8.

Wang SK, Choi M, Richardson AS, Reid BM, Seymen F, Yildirim M, Tuna EB, Gencay K, Simmer JP, Hu JC (2014b). STIM1 and SLC24A4 are critical for enamel maturation. *J Dent Res* [Epub ahead of print]

White JK, Gerdin AK, Karp NA, Ryder E, Buljan M, Bussell JN, Salisbury J, Clare S, Ingham NJ, Podrini C, Houghton R, Estabel J, Bottomley JR, Melvin DG, Sunter D, Adams NC; Sanger Institute Mouse Genetics Project, Tannahill D, Logan DW, Macarthur DG, Flint J, Mahajan VB, Tsang SH, Smyth I, Watt FM, Skarnes WC, Dougan G, Adams DJ, Ramirez-Solis R, Bradley A, Steel KP (2013). Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* 154:452-64.

Winter GB, Brook AH (1975). Enamel hypoplasia and anomalies of the enamel. *Dent Clin North Am* 19:3-24.

Witkop CJ Jr, Sauk JJ Jr (1976). Hereditable defects of enamel. In: Oral Facial Genetics. Stewart RE and Prescott GH editors. St. Louis: Mosby, pp.151-226.

Witkop CJ Jr (1988). Amelogenesis imperfecta, dentinogenesis imperfecta and dentin dysplasia revisited: problems in classification. *J Oral Pathol* 17:547-53.

CHAPTER 4

MOLECULAR CHARACTERIZATION OF FAM83H

ABSTRACT

Recently, mutations in *FAM83H* (family with sequence similarity 83, member H) were identified to cause autosomal dominant hypocalcified amelogenesis imperfecta (ADHCAI). All of the 20 disease-causing mutations reported so far are expected to truncate FAM83H from 1179 to between 287 and 694 amino acids. This genetic homogeneity of the *FAM83H* mutational spectrum strongly suggested a dominant negative effect in the pathogenesis of ADHCAI. Although human mutational studies revealed the physiological function of FAM83H in enamel development, the functions of FAM83H in cellular and molecular levels are still largely unknown. Here, we used a biochemical approach to study FAM83H protein-protein interactions and the intracellular localization of wild-type and mutant FAM83Hs. We found that FAM83H interacts with itself, which suggested that the mutant protein may exert the dominant negative effect through interacting with the wild-type protein and prevent it from functioning. We also demonstrated that FAM83H interacts with casein kinase 1 (CK1) via an F-X-X-X-F-X-X-X-F sequence motif in its N-terminus and with SEC16A via a binding domain in the middle of the protein. With SEC16A being a protein component of COP II complex and involved in ER-to-Golgi membrane trafficking, we proposed that FAM83H may serve as a scaffold protein for

localizing CK1 to SEC16A so that CK1 can facilitate membrane trafficking at the ER exit site. We hypothesized that a truncated FAM83H that can bind to CK1 but not SEC16A will fail to correctly localize CK1, will disturb proper vesicle trafficking and protein transport from ER to Golgi, and lead to pathological enamel formation. Our study not only reveals a potential cellular function of FAM83H, but also provides a plausible pathological mechanism of ADHCAI through a dominant negative effect from mutant truncated FAM83H proteins.

INTRODUCTION

Mutational analysis of human hereditary enamel defects had been using target gene approaches of candidate genes encoding enamel matrix proteins and proteases that scientists had identified biochemically within developing enamel (Kim *et al.*, 2006). Few genome-wide approaches have been applied until recently. In 2007, by means of linkage analysis, Mendoza *et al.* mapped a new AI locus to chromosome 8q24.3 (Mendoza *et al.*, 2007). Based upon this linkage data, our group identified mutations in a novel gene, *FAM83H*, responsible for autosomal dominant hypocalcified amelogenesis imperfecta (ADHCAI) (Kim *et al.*, 2008), a specific form of AI in which the affected enamel is cheesy-soft and easily peeled off after tooth eruption. Subsequently, many disease-causing *FAM83H* mutations were reported by different groups, and *FAM83H*-associated AI seemed to be the most prevalent AI in North America (Lee *et al.*, 2008; Hart *et al.*, 2009; Wright *et al.*, 2009; Hyun *et al.*, 2009; El-Sayed *et al.*, 2010; Lee *et al.*, 2011; Wright *et al.*, 2011; Haubek *et al.*, 2011; Song *et al.*, 2012).

FAM83H (family with sequence similarity 83, member H) encodes a protein of 1179 amino acids with a calculated molecular mass of 127-kDa (Kim *et al.*, 2008). Unlike all of the

enamel matrix proteins and proteases known to be important for enamel formation, FAM83H does not have a signal peptide and is expected to be a non-secreted intracellular protein. However, the primary structure of FAM83H gives little indication of its potential function. Based upon bioinformatic structure and domain prediction, FAM83H has neither well-defined structural characteristics nor known functional domains except an N-terminal phospholipase D-like (PLD-like) domain, which is the shared element among all the members in FAM83 family and gives the group its identity (Fig. 4.1A). However, the homology between this domain and PLD is trace and probably only indicative of a similar three-dimensional fold. Therefore, it is very unlikely that FAM83H has PLD-like enzymatic activity (Ding *et al.*, 2009).

To date, 20 novel disease-causing mutations have been identified, and some have been reported more than once (Fig. 1.4). Noticeably, all the reported mutations are either nonsense mutations or frameshifts leading to a premature stop codon. No other types of loss-of-function mutations, such as missense mutations, have been reported to be disease-causing. More interestingly, all the *FAM83H* mutations are located within a specific 5' region of its last exon (Exon 5) and are predicted to generate a mutant transcript that can escape nonsense mediated decay and produce a truncated protein ending between amino acids 287 and 694 (Fig. 1.4; Fig. 4.1B, C). This mutational homogeneity highly suggests a dominant negative effect or a gain of function in the pathogenesis of ADHCAI. In contrast, haploinsufficiency is the less plausible pathological mechanism, since other types of loss-of-function mutations seem not to cause a disease phenotype.

So far, human mutational studies are the main evidence indicating the physiological significance of FAM83H. However, the actual functions of this protein inside the cell and the pathogenesis of its associated enamel defects are completely unknown. Therefore, in this study,

Figure 4.1 (A) (figure continued next page)

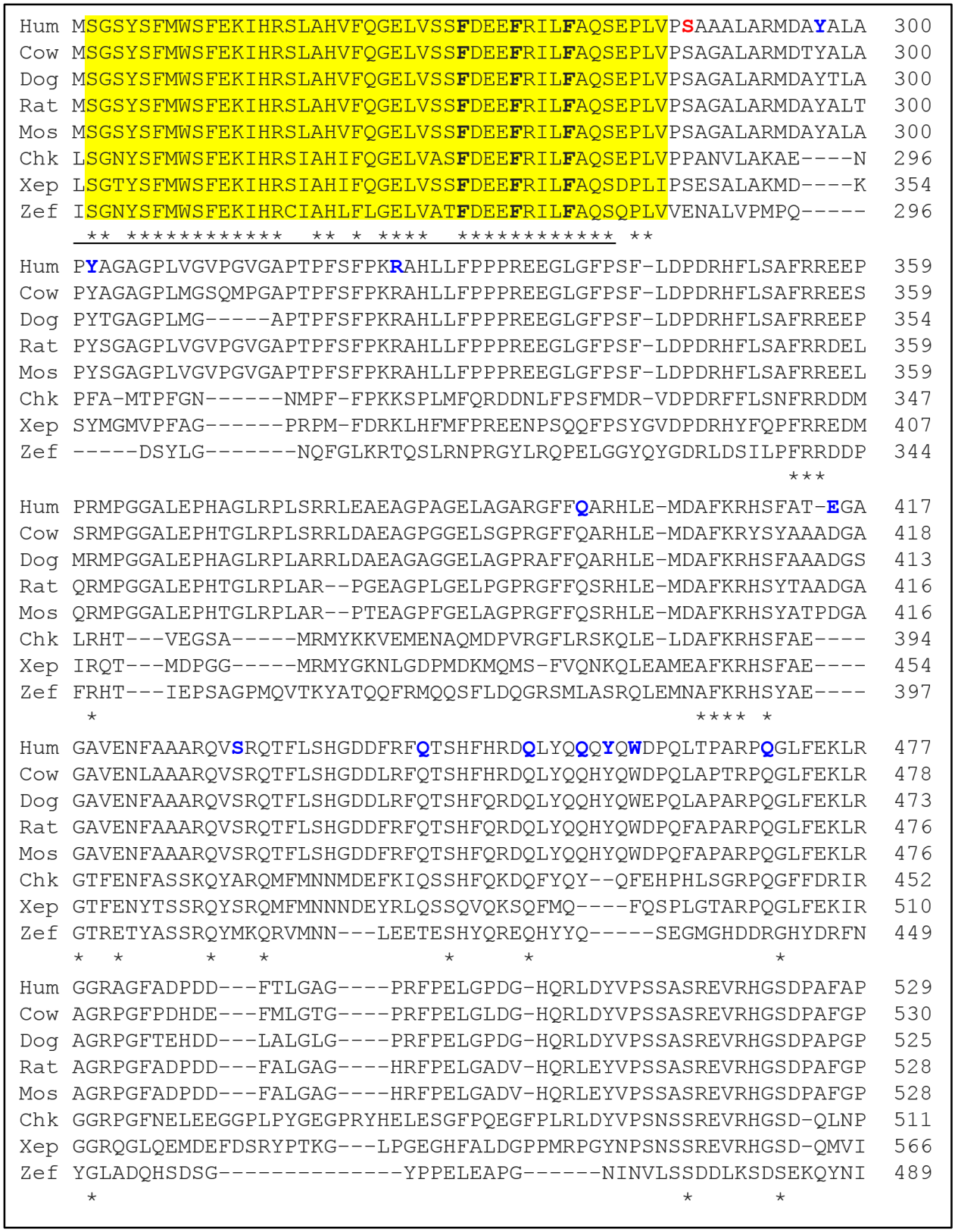


Figure 4.1 (B) (figure continued next page)

Figure 4.1 (C) (figure continued next page)

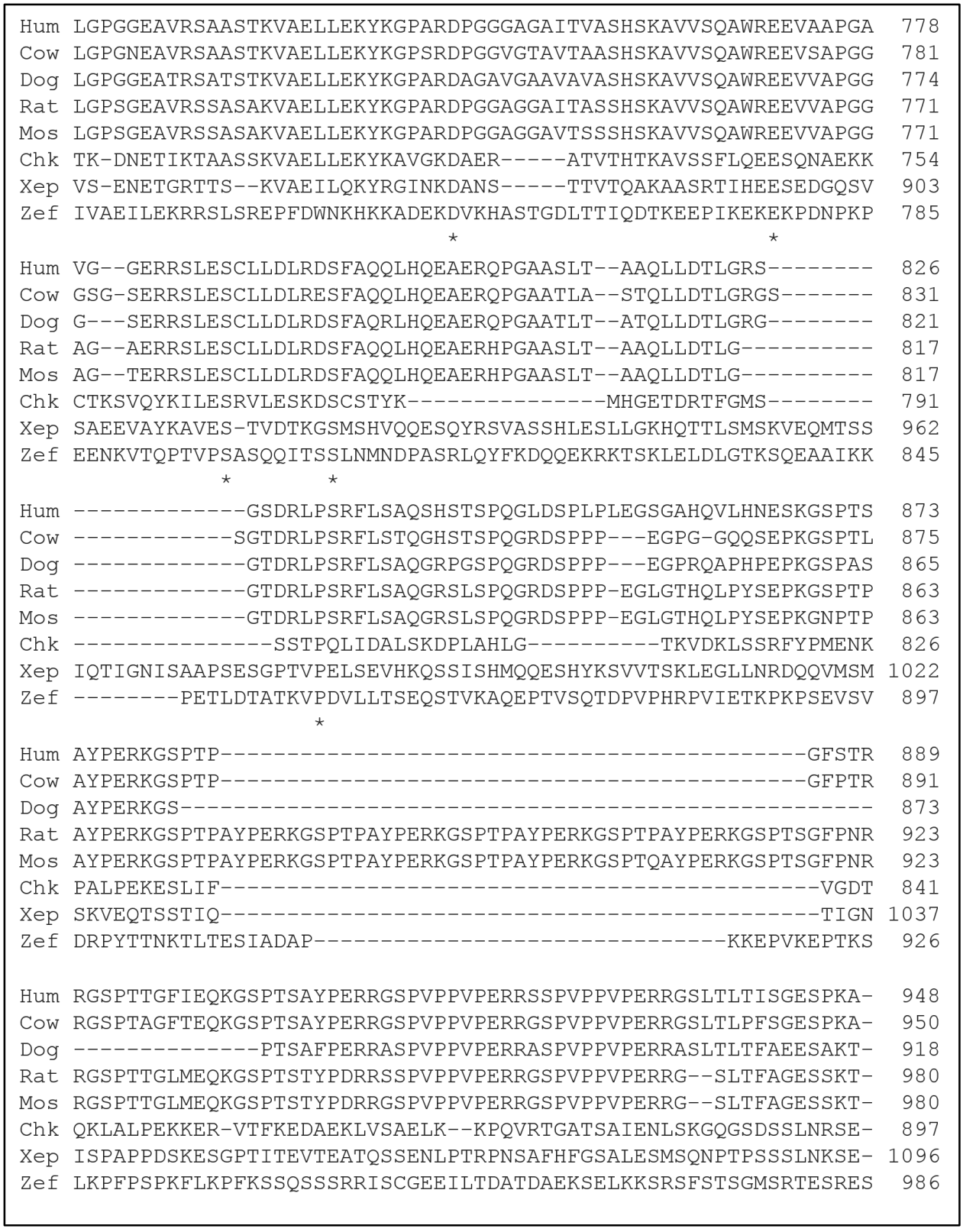


Figure 4.1 (D) (figure continued next page)


Hum ----GPAEEGPSGPMEVLRKGSLRLRQLLSPKGERRMEDEGGFPVPQENGQPESPRRLSL 1004
Cow ----GPTEEAVGGPMEVLRKGSARLRQLMSPKGERRADDEGSFPTPQENGQPESPQWPSL 1006
Dog ----GTAEESAGGPMEVLRKGSLRLRQLLSPKGERRAEEDGGFPAPQENGQPESPRRPSL 974
Rat ----GPTEEVSGGPMEVLRKGSLRLRQLLSPKSERRGEDEVSFPAPQENGQPESPRRPSL 1036
Mos ----GPTEEVSSGPMEVLRKGSLRLRQLLSPKNERRGEDEGSFPTPQENGQPESPRRPSL 1036
Chk ----EECSKQEQNTMEFLRKGSLRLKQFLNPKGEKKLEEEPNSEIGKSDKQPMGLKRSSM 953
Xep ----EDLAKTDQN---FFRKGSMRLKQFLQSKAEKKAEEDLASDNAKAEKQHSTLRRLSK 1149
Zef LSSLGNSESKDTKALDFLKKQTQRLKGILGPKGDKKHSGVSNSQEDKSMKTVPEVQEEIS 1046
Hum GQGDSTEAATE----------ERGPRARLSSATANALYSSNLRDDTKAILEQ 1046
Cow VRVDSTEAAAE----------ERGPRARVASATANALYSSNLRDDTKAILEQ 1048
Dog GRADSTEAAAAA----------DERGPRARTASATANALYSSNLRDDTKAILEQ 1018
Rat SRGDSTEAAAE----------ERGSRVRLASATANALYSSNLRDDTKAILEQ 1078
Mos SRGDSTEAAAE----------ERGSRVRLASATANALYSSNLRDDTKAILEQ 1078
Chk G--DCQEMLGE---EEKNHKFATLLPPKSSQPTQGRFPSSTANILYSSNLRDDTKVILEQ 1008
Xep S--DSQEVAASTDMEEKSAKSLSVSPPKTSSISQSRLSASTSNVIFSSNLRDDTKVILEQ 1207
Zef DKGKPSESISSS---------TAVENKPSAKPTTSRYQSSTSNIIFSSNLRDDTKVILEQ 1097
Hum ISAHGQKHRAVPAPSPGPTHNSPELGRPPAAGVLAPDMSDKDKCSAIFRSDSLGTQGRLS 1106
Cow ISAHGQKHRGVPAPGPTPAHSSPELGHSPVAGGLAPDMSDKDKCSAIFRSDSLGTQGRLS 1108
Dog ISAHGQKHRAVPAPAPGLAHSSPELGRSPTAGGLAPDMSDKDKCSAIFRSDSLGTQGRLS 1078
Rat ISAHGQKHRAVPAP--GPAHNSSDVGRPTTAGDLAPDMSDKDKCSAIFRSDSLGTQGRLS 1136
Mos ISAHGQKHRGVPAP--GPAHSSPDVGRPTTAGDLAPDMSDKDKCSAIFRSDSLGTQGRLS 1136
Chk ISANSQKNRAELAKQLPSTSNP-DLSKSTMSLERKTE---KEKSCNIHRSESFGSQKRNL 1064
Xep ISANSQKNRAEMVKQAQQIQATGDPDPATSKPESKTEGTASTDAAAITRTGSFLSRSRFS 1267
Zef ISANSQKTRQQNEESGKGDGGKDDVANSP----------FQSRNRFS 1134
Hum RTLPASAEERDRLLRRMESMRKEKRVYSRFEVFCKKEEASSPGAGEGPAEEGTRDSKVGK 1166
Cow RTLPAGAEERDRLLRRMESMRKEKRVYSRFEVFCKKEEPGALGAAEGPAEEDARDSKVGK 1168
Dog RTLPASAEDRDRLLRRMESMRKEKRVYSRFEVFCKKEEAGGPGAGEGPAEEDTRDSKVGK 1138
Rat RTLPASAEERDRLLRRMESMRKEKRVYSRFEVFCKKEDAGSSAAGDNLADEDTRDSKMGK 1196
Mos RTLPGSAEERDRLLRRMESMRKEKRVYSRFEVFCKKDEAGSSGAGDNLADEDTRDSKMGK 1196
Chk QRQPS--EDRDTLLKKMENMRKEKRVYSRFEVFCKKDE--HTSQSEEEYDTDAKDKKMGK 1120
Xep RPSPSSPEDRDILLKRMESIRKEKRVYSRFEVFCKKDE--QPSHAD-----DNDDKKAGK 1320
Zef -RAPVNPQERDNLLKRIESMRKEKKVYSRFEVLYRSREECCWERGS--------VEQANQ 1185
Hum FVPKILGTFKSKK 1179
Cow FMPKILGTFKSKK 1181
Dog FMPKILGTFKSKK 1151
Rat FVPKILGTFKSKK 1209
Mos FVPKILGTFKSKK 1209
Chk FMPKILGTFKTKK 1133
Xep IIPKLLGNLIKK- 1332
Zef FLIKSIE------ 1192

Figure 4.1 (E)

Figure 4.1: Protein sequence alignment of FAM83H orthologs from 7 vertebrates.
Asterisks (*) mark identical amino acids. The N-terminal PLD-like domain (amino acid 4-281 in human FAM83H) is underlined (A). Ser^{287} (the most N-terminal mutation) and Glu^{694} (the most C-terminal mutation) are marked red/bold (B, C). All the other nonsense mutations are marked blue/bold (B, C). Three phenylalanines (bold) of F^{270}-X-X-X-F^{274}-X-X-X-F^{278} motif are located at a highly conserved region right N-terminal to Ser^{287} (yellow-highlighted) (B). A conserved region N-terminal to Glu^{694} is highlighted with green (C), and two conserved areas at C-terminal end of FAM83H with blue (E). Protein sequences are from human (Hum) (*Homo sapiens*; gi|157311635, ref|NP_940890.3), cow (*Bos taurus*; gi|119906195, ref|XP_603315.3), dog (*Canis lupus familiaris*; gi|374091972, gb|AEY83660.1), rat (*Rattus norvegicus*; gi|194474062, ref|NP_001124037.1), mouse (Mos) (*Mus musculus*; gi|269914118, ref|NP_598848.2), chicken (Chk) (*Gallus gallus*; gi|118087467, ref|XP_423955.2), frog (Xep) (*Xenopus tropicalis*; gi|301629161, ref|XP_002943716.1), and zebrafish (Zef) (*Danio rerio*; gi|113682418, ref|NP_001038555.1).

we aim to define the functions of FAM83H in cellular and molecular levels, and to unravel the potential pathological mechanism of ADHCAI by means of biochemical approaches. Through characterizing protein-protein interactome of FAM83H and defining the intracellular localization of wild-type and mutant FAM83H proteins, we hope to know what cellular processes FAM83H is involved in and how mutant FAM83H proteins lead to a disease phenotype.

RESULTS

FAM83H self-interaction

In many pathological conditions, a mutant protein can exert a dominant negative effect by interacting with its wild-type protein and lead to a loss of function. With the suspicion that the reported *FAM83H* mutations lead to a dominant negative effect in the pathogenesis of its related enamel defects, we hypothesized that FAM83H might interact with itself to form dimers or multimers. Furthermore, the N-terminus of FAM83H contains a phospholipase D (PLD)-like domain (cd09188) with a predicted structure similar to that of PLD. Since PLD has been shown

to form dimers (Stuckey and Dixon, 1999), it is possible that FAM83H can dimerize through its N-terminal PLD-like domain.

In order to test this hypothesis, we first co-expressed Flag-tagged FAM83H and Myc-tagged FAM83H in HEK293 cells and performed anti-Flag pull-down assays using cell lysates. By immunoblotting with anti-Myc tag antibody, we demonstrated that Flag-tagged FAM83H could pull down Myc-tagged FAM83H, which suggested that FAM83H interacted with itself (Fig. 4.2A). We also tested this self-interaction with two truncated FAM83Hs ($FAM83H^{287X}$, $FAM83H^{697X}$). The results showed that both Flag-tagged $FAM83H^{287X}$ and $FAM83H^{697X}$ were able to pull down Myc-tagged FAM83H, indicating that the N-terminus of FAM83H (the first 287 amino acids) was responsible for the self-interaction (Fig. 4.2B).

We also performed protein-protein interaction modeling using SPRING ON-LINE, a template-based algorithm for protein-protein structure prediction (Guerler *et al.*, 2013). When the sequence of the first 287 amino acids of human FAM83H was used as an input, 6 interaction models were predicted. The one with the highest confidence score used phospholipase D from *S. typhimurium* as a template (1byrA) (Fig. 4.2C). This result suggested that FAM83H might form a dimer through its N-terminal PLD-like domain (the first 287 amino acids of human FAM83H).

FAM83H interactome

In addition to self (homomeric) interactions, a mutant protein might exert a dominant negative effect by competing with its wild-type protein for (heteromeric) interactions with another protein that leads to a loss of function. Also, identifying FAM83H binding partners may help to unravel its functions. Therefore, we aimed to identify the potential FAM83H interacting

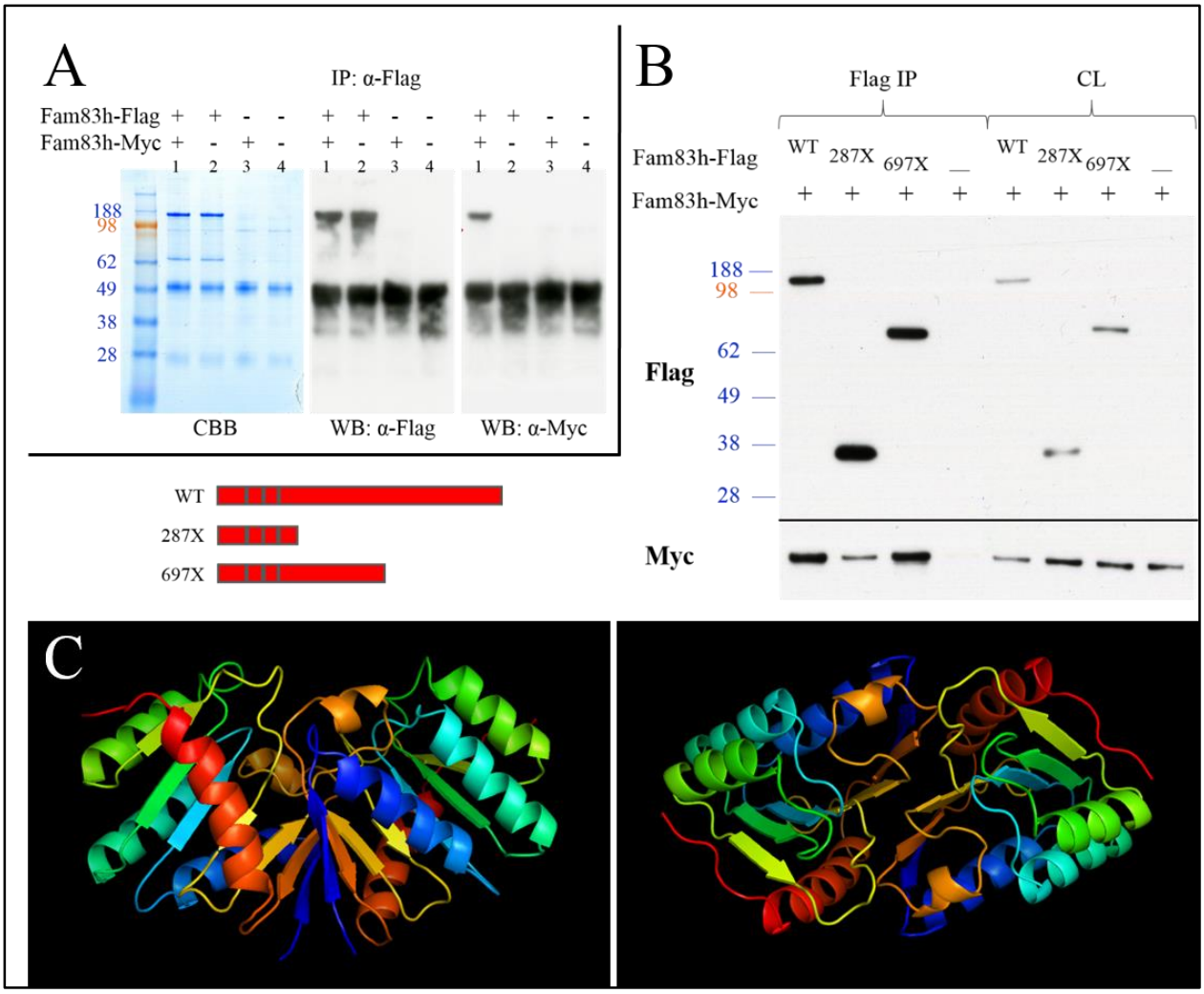


Figure 4.2: FAM83H self-interaction.
A: Pull-down assay of two different tagged full-length FAM83H proteins. Flag-tagged FAM83H and Myc-tagged FAM83H were co-expressed in HEK293 cells, the cell lysate then underwent immunoprecipitation (IP) with α-Flag antibody, and the precipitates were fractionated by SDS-PAGE and stained with Coomassie Brilliant Blue (CBB) (***Left***) or immunoblotted with α-Flag (***Middle***) and α-Myc (***Right***) antibodies (Lanes 1). 3 control experiments were included: Flag-tagged FAM83H expression only (Lanes 2), Myc-tagged FAM83H expression only (Lanes 3), and empty vector expression (Lanes 4). The α-Flag IP pulls down not only Flag-tagged but also Myc-tagged FAM83H (~130 kDa bands), suggesting that FAM83H interacts with itself. The ~50 kDa bands on western blots (WB) are signals from heavy chain of rabbit IgG used for α-Flag IP.
B: Pull-down assay of C-terminal truncated FAM83H proteins. Myc-tagged full-length FAM83H and one of the Flag-tagged FAM83H proteins with various length (WT: full-length; 287X: a.a.1-287; 697X: a.a.1-697; –: Flag tag only) were expressed and α-Flag immunoprecipitated. The IP products as well as initial cell lysates (CL) were blotted with α-Flag (***Top***) and α-Myc (***Bottom***) antibodies. All 3 different-length FAM83Hs (Flag-tagged) can pull down full-length FAM83H (Myc-tagged), suggesting that the first 287 amino acids of the protein are sufficient for FAM83H self-interaction. Bands of ~140 kDa, ~36 kDa, and ~75 kDa are full-length FAM83H, FAM83H^{287X}, and FAM83H^{697X} respectively.

C: Predicted human FAM83H^{287X} dimerization model. The interaction model was predicted by SPRING ON-LINE software (Guerler *et al.*, 2013). The first 287 amino acids of human FAM83H was used as an input, and the structure of phospholipase D from *S. typhimurium* (1byrA) was used as a modeling template. The F^{270}-X-X-X-F^{274}-X-X-X-F^{278} motif is located at the red α-helix. ***Left:*** Side view. ***Right:*** Top view.

proteins using affinity purification combined with mass-spectrometry (AP-MS) (Gingras *et al.*, 2007; Dunham *et al.*, 2012). We over-expressed Flag-tagged mouse FAM83H in HEK293 cells and performed immunoprecipitation with anti-Flag antibody. The immunoprecipitates were then resolved by SDS-PAGE, and 6 specific bands on the gel were sliced out and submitted for protein identification by mass-spectrometry. By this means, all of the proteins that co-immunoprecipitated with the affinity-purified FAM83H would be identified, and the potential binding partners of FAM83H could be determined by subsequent analyses. A total of 143 proteins were identified by mass spectrometry, including 89 matches with a high confidence score ($p<0.05$). The identified potential interacting proteins were involved in various cellular processes (Tab. 4.1).

Table 4.1: Potential FAM83H interacting proteins.
Over-expressed Flag-tagged mouse FAM83H protein (in HEK293 cells) underwent affinity purification followed by mass spectrometry (AP-MS) to identify potential binding partners of FAM83H. Only human proteins identified with a significant Mascot score (≥ 85 in this case) are listed. Note that endogenous human FAM83H protein can be co-immunoprecipitated by over-expressed mouse FAM83H protein, demonstrating that FAM83H interacts with itself as we show in Fig. 4.2.

Protein symbol	Protein name	Protein symbol	Protein name	Protein symbol	Protein name
ACTB	actin, beta	**HNRNPA1**	heterogeneous nuclear ribonucleoprotein A1	**RCN1**	reticulocalbin 1, EF-hand calcium binding domain
ACTG1	actin, gamma 1	**HNRNPC**	heterogeneous nuclear ribonucleoprotein C (C1/C2)	**RCN2**	reticulocalbin 2, EF-hand calcium binding domain
ACTR2	ARP2 actin-related protein 2 homolog (yeast)	**HSPA1A**	heat shock 70kDa protein 1A	**RFC3**	replication factor C (activator 1) 3, 38kDa
ARF4	ADP-ribosylation factor 4	**HSPA9**	heat shock 70kDa protein 9 (mortalin)	**RPL22**	ribosomal protein L22

ATAD3A	ATPase family, AAA domain containing 3A	**IGF2BP1**	insulin-like growth factor 2 mRNA binding protein 1	**RPL27**	ribosomal protein L27
ATAD3B	ATPase family, AAA domain containing 3B	**KRT1**	keratin 1	**RPL28**	ribosomal protein L28
CAD	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	**KRT10**	keratin 10	**RPLP0**	ribosomal protein, large, P0
CALU	calumenin	**KRT2**	keratin 2	**RPS13**	ribosomal protein S13
CAPZA1	capping protein (actin filament) muscle Z-line, alpha 1	**KRT9**	keratin 9	**RPS14**	ribosomal protein S14
CKAP5	cytoskeleton associated protein 5	**MYH10**	myosin, heavy chain 10, non-muscle	**RPS19**	ribosomal protein S19
CNOT1	CCR4-NOT transcription complex, subunit 1	**MYH14**	myosin, heavy chain 14	**RPS20**	ribosomal protein S20
CSNK1A1	casein kinase 1, alpha 1	**MYH9**	myosin, heavy chain 9, non-muscle	**RPS25**	ribosomal protein S25
CSNK1E	casein kinase 1, epsilon	**MYL6**	myosin, light chain 6, alkali, smooth muscle and non-muscle	**SEC16A**	SEC16 homolog A (S. cerevisiae)
DDX3X	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked	**NES**	nestin	**SLC25A13**	solute carrier family 25, member 13 (citrin)
DNAJB2	DnaJ (Hsp40) homolog, subfamily B, member 2	**NPM1**	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	**SNRNP200**	small nuclear ribonucleoprotein 200kDa (U5)
DNAJB6	DnaJ (Hsp40) homolog, subfamily B, member 6	**PCBP1**	poly(rC) binding protein 1	**SNRPD2**	small nuclear ribonucleoprotein D2 polypeptide 16.5kDa
FAM83H	family with sequence similarity 83, member H	**PPP1CA**	protein phosphatase 1, catalytic subunit, alpha isoform	**TIMM50**	translocase of inner mitochondrial membrane 50 homolog (S. cerevisiae)
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	**RBM14**	RNA binding motif protein 14	**TOMM22**	translocase of outer mitochondrial membrane 22 homolog (yeast)
HELZ	helicase with zinc finger	**RBM4**	RNA binding motif protein 4	**TRAFD1**	TRAF-type zinc finger domain containing 1

FAM83H and casein kinase 1 interaction

Based upon the AP-MS results, several members of casein kinase 1 (CK1) family could be identified in the immunoprecipitates of Flag-tagged FAM83H (FAM83H-Flag). In order to

validate this interaction, co-immunoprecipitation of endogenous CK1δ, and CK1ε (CSNK1D and CSNK1E) with over-expressed FAM83H-Flag in HEK293 cells was performed. The results showed that FAM83H-Flag could precipitate CK1δ, and CK1ε by Flag antibody immunoprecipitation, while the control-Flag could not, which demonstrated the specificity of the FAM83H-CK1 interaction (Fig. 4.3A).

Searching for potential CK1 binding sites, Okamura *et al.* identified a conserved docking motif for CK1 binding (F-X-X-X-F) in many of the CK1-interacting proteins (Okamura *et al.*, 2004). Interestingly, in human FAM83H, there are four of such a sequence motif in its N-terminus (F^{247}-X-X-X-F^{251}; F^{270}-X-X-X-F^{274}; F^{274}-X-X-X-F^{278}; F^{350}-X-X-X-F^{354}) (Fig. 4.1B), with the first three being highly conserved among the FAM83H orthologs during vertebrate evolution. The F^{270}-X-X-X-F^{274}-X-X-X-F^{278} motif is particularly conserved among the human FAM83 paralogs, suggesting that this motif might be a potential CK1 binding site in FAM83H. Noticeably, this motif is located at a highly-conserved sequence area right before Ser^{287} of FAM83H, which corresponds to the most N-terminal FAM83H truncation mutation reported so far (Fig. 4.1B). Therefore, we hypothesized that FAM83H interacts with CK1 through its F^{270}-X-X-X-F^{274}-X-X-X-F^{278} motif, and that all of the disease-causing truncation FAM83H, including the shortest and the longest one, can interact with CK1.

In order to test this hypothesis, two constructs expressing truncated mouse FAM83H with N-terminal Flag tag ($FAM83H^{287X}$, $FAM83H^{697X}$) were generated, and their ability to interact with CK1 was tested. $FAM83H^{287X}$ corresponded to the shortest disease-causing FAM83H truncation and $FAM83H^{697X}$ the longest. HEK293 cells were transfected with constructs expressing Myc-tagged CK1ε and one of the Flag-tagged FAM83Hs with various lengths (WT, $FAM83H^{287X}$, $FAM83H^{697X}$, and empty vector control). The cell lysate from each group then

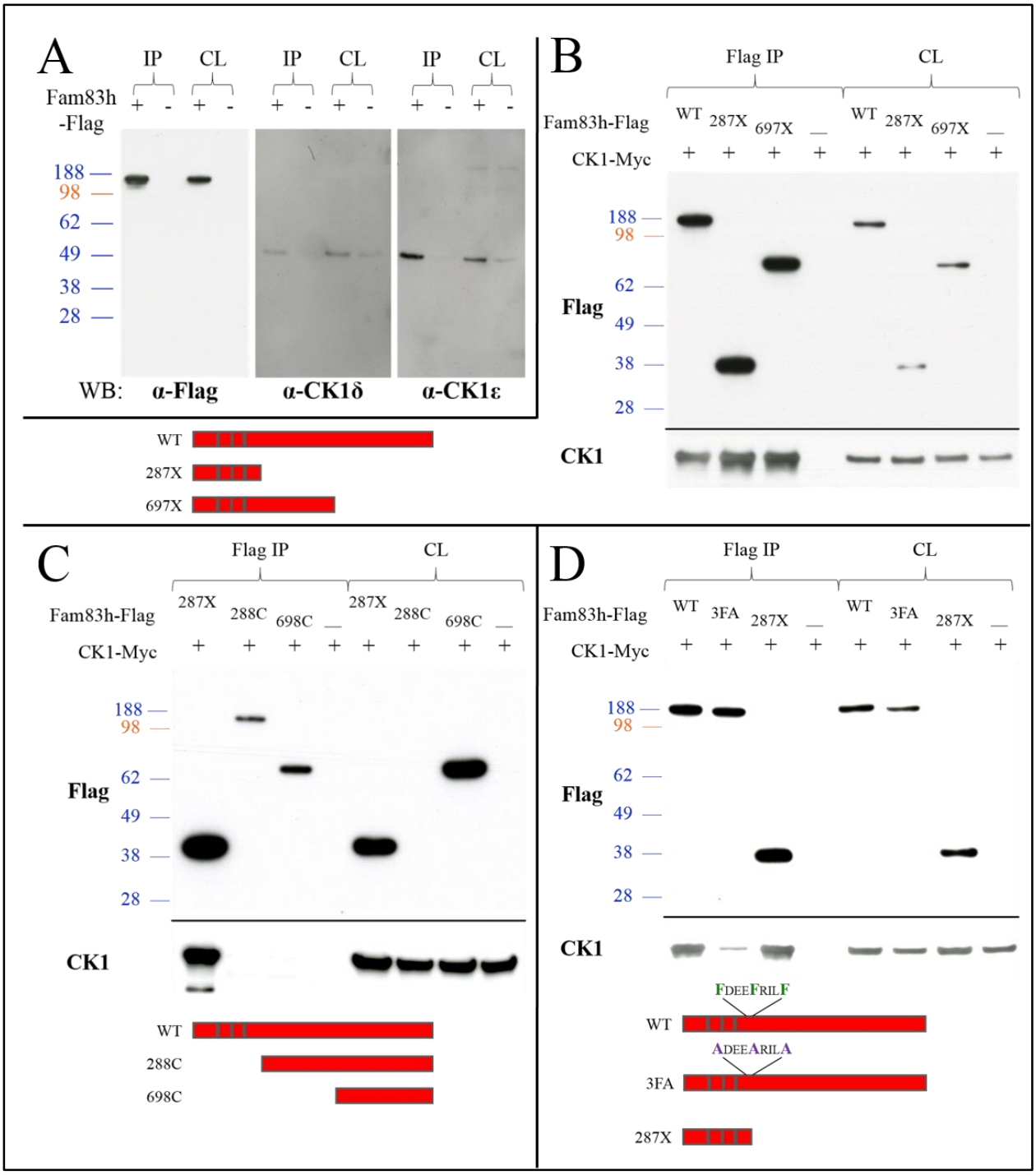


Figure 4.3: FAM83H-CK1 interaction.

A: Co-immunoprecipitation of FAM83H and CK1. Flag-tagged FAM83H was expressed in HEK293 cells, and cell lysates underwent α-Flag immunoprecipitation (IP). The IP products and initial cell lysates (CL) were fractionated and immunoblotted with α-Flag (***Left***), α-CK1δ (***Middle***), and α-CK1ε (***Right***) antibodies. Bands of ~130 kDa, ~49 kDa, and ~47 kDa are FAM83H, CK1δ, and CK1ε respectively. Flag-tagged FAM83H, but not control, IP can precipitate endogenous CK1δ and CK1ε, confirming the results from our AP-MS experiments.

B: Pull-down assay of C-terminal truncated FAM83H proteins. Myc-tagged CK1 and one of the Flag-tagged FAM83H proteins with various length (WT: full-length; 287X: a.a.1-287; 697X: a.a.1-697; –: Flag tag only) were expressed and α-Flag immunoprecipitated. The IP products as well as initial cell lysates (CL) were blotted with α-Flag (***Top***) and α-Myc (***Bottom***) antibodies. All 3 different-length FAM83Hs (Flag-tagged) can pull down CK1 (Myc-tagged), suggesting that the first 287 amino acids of the protein are sufficient for FAM83H-CK1 interaction. Bands of ~140 kDa, ~36 kDa, and ~75 kDa are full-length FAM83H, $FAM83H^{287X}$, and $FAM83H^{697X}$ respectively.

C: Pull-down assay of N-terminal truncated FAM83H proteins. Similar pull-down assays were performed with two different Flag-tagged FAM83H proteins without N-terminus (288C: a.a.288-1209; 698X: a.a.698-1209). While $FAM83H^{287X}$ (Flag-tagged) can pull down CK1 (Myc-tagged), neither $FAM83H^{288C}$ nor $FAM83H^{698C}$ can, confirming that the first 287 amino acids of the protein are necessary for FAM83H-CK1 interaction. ***Top:*** Bands of ~110 kDa and ~65 kDa are $FAM83H^{288C}$ and $FAM83H^{698C}$ respectively. ***Bottom:*** CK1 is ~50 kDa.

D: Pull-down assay of FAM83H proteins with site-directed mutagenesis. Similar pull-down assays were performed with a Flag-tagged FAM83H in which F^{270}, F^{274}, and ^{278} are substituted with Alanines (3FA: F^{270}-X-X-X-F^{274}-X-X-X-F^{278} to A^{270}-X-X-X-A^{274}-X-X-X-A^{278}). Compared to wild-type (WT) FAM83H, $FAM83H^{3FA}$ has significantly reduced ability to pull down CK1, suggesting the F^{270}-X-X-X-F^{274}-X-X-X-F^{278} motif is critical for FAM83H-CK1 interaction.

underwent immunoprecipitation with Flag antibody, and the immunoprecipitates were blotted with anti-CK1ε antibody (or anti-Myc tag antibody). As shown on the immunoblot, all three of the variable-length FAM83H truncations, but not the empty vector control, could pull down over-expressed CK1ε, meaning that the first 287 amino acids of FAM83H are sufficient to interact with CK1ε (Fig. 4.3B). We also tested the CK1-binding ability of two N-terminal truncated FAM83Hs ($FAM83H^{288\text{-}1209}$, $FAM83H^{698\text{-}1209}$) and showed that neither of these truncated proteins could pull down CK1ε, further demonstrating that the N-terminus (amino acids 1-287) of FAM83H is necessary and sufficient for FAM83H-CK1 interaction (Fig. 4.3C).

We mutagenized the three phenylalanines (F^{270}, F^{274}, F^{278}), potentially serving as a CK1 docking site in FAM83H, to alanines ($FAM83H^{3FA}$) by site-directed mutagenesis and evaluated this motif for its importance in CK1 binding. By using the same pull-down assays, we demonstrated that site-directed mutagenesis of F^{270}, F^{274}, F^{278} to A^{270}, A^{274}, A^{278} in FAM83H significantly attenuated its ability to interact with CK1ε (Fig. 4.3D). This result supported our

hypothesis that FAM83H interacts with CK1 through its F^{270}-X-X-X-F^{274}-X-X-X-F^{278} motif in N-terminus.

Mouse recombinant FAM83H phosphorylation by casein kinase 1 *in vitro*

The FAM83H-CK1 interaction raises the possibility that FAM83H might be phosphorylated by CK1, since there are many predicted CK1 phosphylation sites with a specific D/E/pS-X-X-Ser/Thr motif in FAM83H. In order to test if FAM83H is the substrate of CK1, we used bacterial-expressed mouse recombinant FAM83H and performed an *in vitro* kinase assay. Incubated with γ-^{33}P-ATP, the recombinant FAM83H showed much stronger radioactivity when CK1 was added, compared with a CK2 (casein kinase 2) and no kinase control, which demonstrated that FAM83H can be phosphorylated by CK1 *in vitro* (Fig. 4.4). The kinase assays were mainly conducted by Drs. Jan Hu and Yasuo Yamakoshi.

We submitted the CK1-phosphorylated mouse recombinant FAM83H (with cold ATP) for mass-spectrometry to identify the exact CK1 phosphorylation sites in FAM83H. The results showed that most of the CK1-phosphorylated Serines and Threonines were located at the C-terminus of FAM83H (Fig. 4.4).

FAM83H and SEC16A interaction

Previously, our group showed that over-expressed GFP-tagged mouse FAM83H partly localized in the trans-Golgi network of HEK293 cells, which suggested that FAM83H might be a peripheral membrane protein that plays a role in intracellular trafficking (Ding *et al.*, 2009).

Figure 4.4: FAM83H phosphorylation by CK1 *in vitro*.
Left: Autoradiograph of ^{33}P kinase assay. Purified mouse recombinant FAM83H proteins were incubated with CK1, CK2, or no enzyme (–) as well as radioactive ^{33}P-ATP at 30˚C for 60 min. The reaction samples were fractionated with SDS-PAGE, and the dried gel was exposed to a film. The CK1 reaction shows strong radioactivity at a specific band of ~130 kDa, demonstrating that CK1 can phosphorylate FAM83H *in vitro*. (The autoradiograph is courtesy for Drs. Jan Hu and Yasuo Yamakoshi.) ***Right:*** CK1 phosphorylation sites on FAM83H. The phosphorylation sites are determined by mass spectrometry of CK1-treated mouse FAM83H protein. The CK1-phosphorylated Serines and Threonines, which mainly locate at C-terminus of FAM83H, are marked red/bold.

Searching for protein candidates involved in membrane trafficking in our AP-MS data of FAM83H, we identified another potential FAM83H-binding protein, SEC16A, which might imply the function of FAM83H (Tab. 4.1). SEC16A is a protein that forms part of the COP II complex, which mediates vesicle formation and protein transport from endoplasmic reticulum (ER) to Golgi (Iinuma *et al.*, 2007; Miller and Barlowe, 2010; Zanetti *et al.*, 2012). SEC16A localized to the ER exit site (transitional ER), close to the Golgi, where GFP-tagged FAM83H had localized.

In order to validate this FAM83H-SEC16A interaction, we performed the same pull-down assays as we did for the FAM83H-CK1 interaction. Different-length Flag-tagged FAM83Hs (WT, FAM83H^{287X}, FAM83H^{697X}, and empty vector control) were tested for their ability to bind over-expressed Halo-tagged SEC16A. As shown on anti-SEC16A (as well as anti-

Halo tag) immunoblot, while full-length (WT) FAM83H and FAM83H^{697X} could pull down Halo-tagged SEC16A, FAM83H^{287X} could not, suggesting that FAM83H interacted with SEC16A and the middle part of the protein (amino acids 287-697 of mouse FAM83H) might be important for this interaction (Fig. 4.5A). This result was further validated by no interaction between Flag-tagged full-length FAM83H and control HaloTag® protein (~300 amino acids), which demonstrated the specificity of FAM83H-SEC16A interaction (Fig. 4.5B).

Previously, Lee *et al.* expressed different truncated GFP-tagged human FAM83Hs (WT, FAM83H^{325X}, FAM83H^{460X}, FAM83H^{677X}, FAM83H^{694X}) in HEK293 cells and showed altered intracellular localization of these truncated proteins (Lee *et al.*, 2011). In their results, while full-length (WT) FAM83H and FAM83H^{697X} localized to the cytoplasm, all of the other truncated proteins (FAM83H^{325X}, FAM83H^{460X}, FAM83H^{677X}) localized to the nucleus, which suggested that the area between amino acid 677 and 694 might be important for FAM83H to stay at its physiological location (cytoplasm). Correspondingly, there is a highly conserved sequence area in amino acids 664-688 of human FAM83H, which implies it has functional significance (Fig. 4.1C). Therefore, with our previous result showing that amino acids 287-697 of mouse FAM83H are important for SEC16A binding, we hypothesized that the conserved domain in the middle of FAM83H (amino acids 664-688 of human FAM83H) might serve as a SEC16A binding site and keep FAM83H at its physiological location.

To test this hypothesis, in addition to *FAM83H^{697X}*, we generated another construct, *FAM83H^{657X}*, to express a shorter form of truncated mouse FAM83H in which the conserved domain was deleted. With similar pull-down assays, we tested the SEC16A-binding ability of these two truncated FAM83Hs. However, inconsistent with what we expected, while

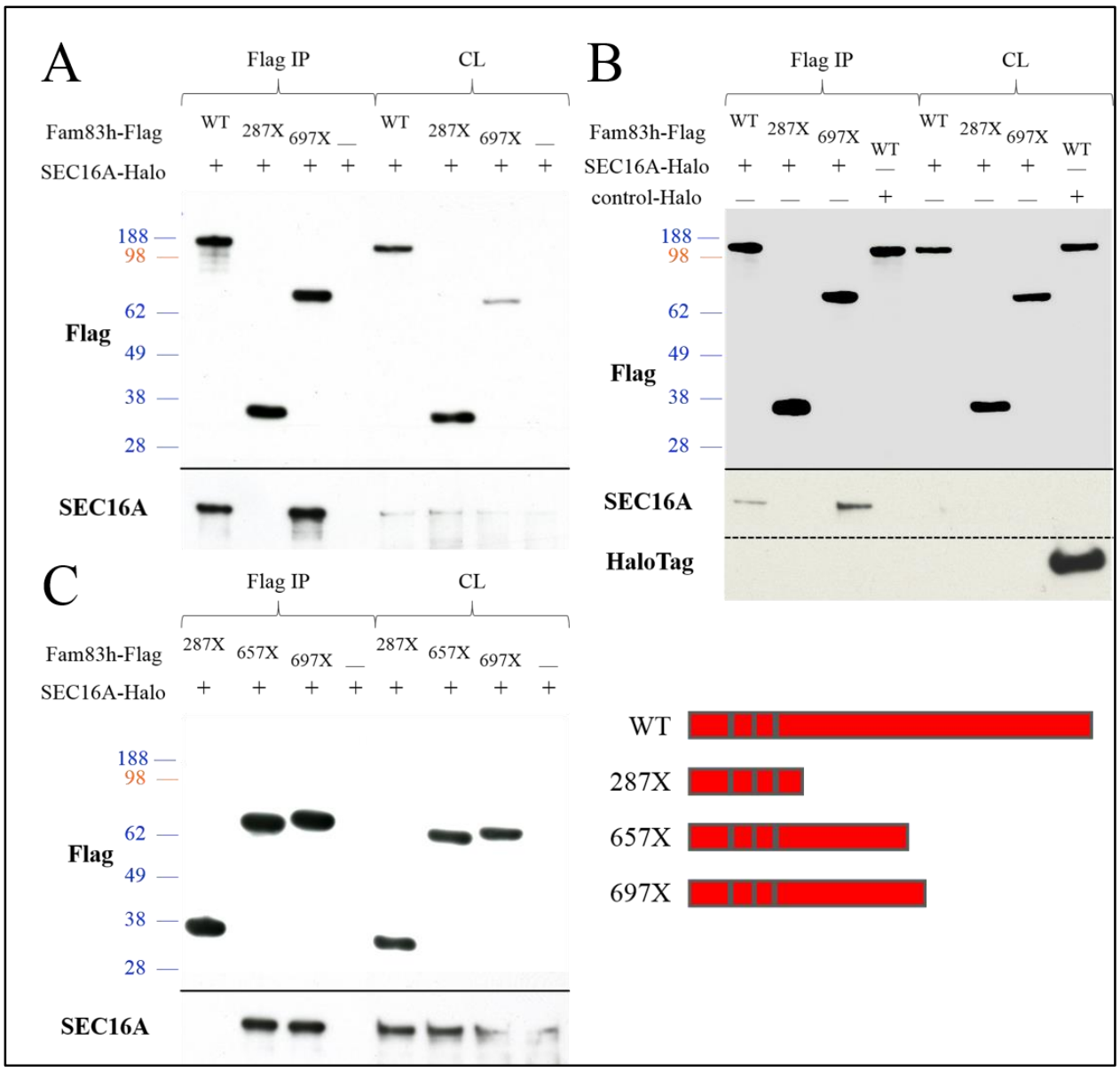


Figure 4.5: FAM83H-SEC16A interaction.

A: Pull-down assay of C-terminal truncated FAM83H proteins. Halo-tagged SEC16A and one of the Flag-tagged FAM83H proteins with various length (WT: full-length; 287X: a.a.1-287; 697X: a.a.1-697; –: Flag tag only) were expressed and α-Flag immunoprecipitated. The IP products as well as initial cell lysates (CL) were blotted with α-Flag (***Top***) and α-HaloTag® (***Bottom***) antibodies. While full-length FAM83H and FAM83H^{697X} (Flag-tagged) can pull down SEC16A (Halo-tagged), FAM83H^{287X} cannot, suggesting that the middle part of FAM83H (a.a.287-697) may be responsible for interacting with SEC16A. Bands of ~140 kDa, ~36 kDa, and ~75 kDa are full-length FAM83H, FAM83H^{287X}, and FAM83H^{697X} respectively.

B: Pull-down assay of FAM83H and control HaloTag® protein. Similar pull-down assays were performed with control HaloTag® protein (HaloTag® is ~300 amino acid long). Full-length FAM83H (Flag-tagged) can pull down SEC16A (Halo-tagged), but not HaloTag® protein, demonstrating the specificity of FAM83H-SEC16A interaction; the interaction is not through HaloTag®. ***Top:*** Bands of

~140 kDa, ~36 kDa, and ~75 kDa are full-length FAM83H, FAM83H^{287X}, and FAM83H^{697X} respectively. ***Bottom:*** SEC16A-Halo is ~270 kDa, and control HaloTag® protein is ~30 kDa.

C: Pull-down assay of FAM83H^{657X}. Similar pull-down assays were performed with another C-terminal truncated FAM83H (657X: a.a.1-657). Both FAM83H^{657X} and FAM83H^{697X} (Flag-tagged) can pull down SEC16A (Halo-tagged), suggesting that the region between amino acid 657 and 697 may not be necessary for FAM83H-SEC16A interaction. The ~70 kDa band is FAM83H^{657X}.

FAM83H^{287X} did not show interaction with SEC16A, both FAM83H^{657X} and FAM83H^{697X} were able to pull down Halo-tagged SEC16A, which suggested that amino acids 657-697 of mouse FAM83H (including the conserved domain) might not be critical for SEC16A binding (Fig. 4.5C).

We also performed immunostaining with anti-Flag antibody in HEK293 cells over-expressing Flag-tagged full-length FAM83H and FAM83H^{697X}. Unexpectedly, while full-length FAM83H showed localization in the cytoplasm, the FAM83H^{697X} localized in the nucleus, which is inconsistent with the results from previous report (Lee *et al.*, 2011). However, interestingly, we observed altered intracellular localization of endogenous CK1ε in HEK293 cells over-expressing FAM83H^{697X}. While CK1ε showed a diffuse cytoplasmic localization pattern when full-length FAM83H was over-expressed, in the presence of FAM83H^{697X}, CK1ε localized to the nucleus as FAM83H^{697X} itself did (Fig. 4.6).

DISCUSSION

FAM83H is the first AI candidate gene identified to encode a non-secretory, non-enamel-matrix protein. Human mutational studies are so far the only evidence indicating the physiological function of FAM83H, but the molecular function and the pathological mechanism of its associated enamel malformations are completely unknown. To date, all of the reported disease- causing mutations are nonsense mutations or frameshifts located between amino acids

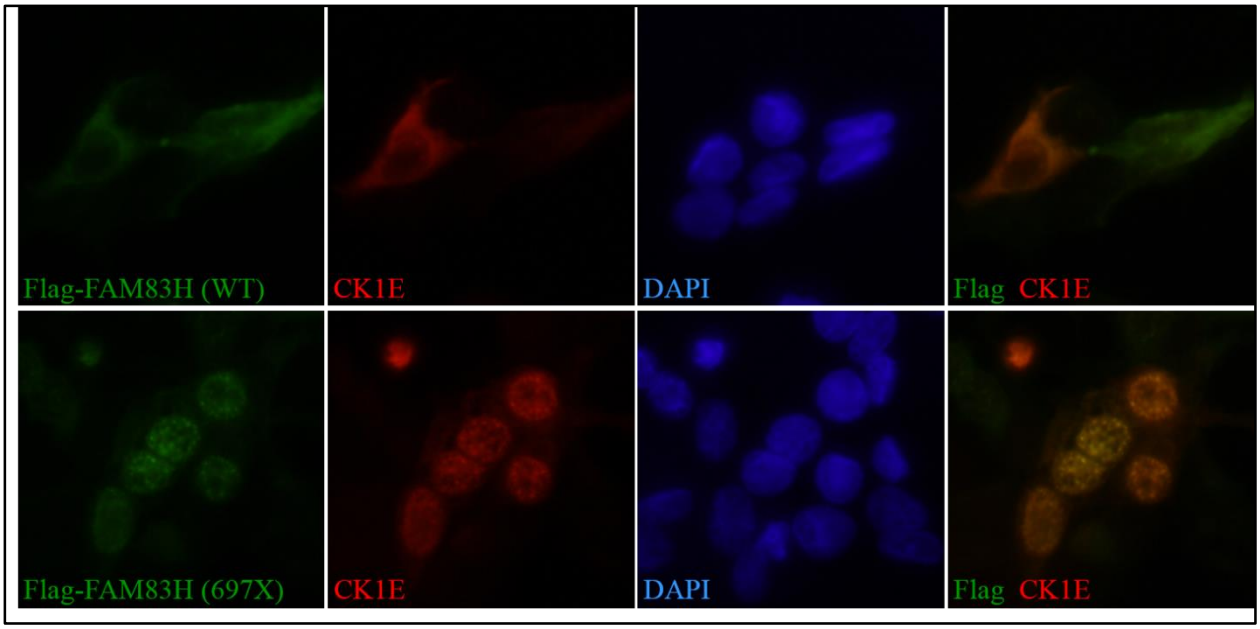


Figure 4.6: FAM83H and CK1E immunocytochemistry.
Flag-tagged full-length (WT) and truncated (697X) FAM83H proteins were expressed in HEK293 cells, and immunostained with α-Flag (green) and α-CK1E (red) antibodies. ***Top***: Both over-expressed WT FAM83H (green) and endogenous CK1E (red) localize at cytoplasm. ***Bottom***: Unlike WT FAM83H, over-expressed FAM83H^{697X} (green) localizes at nucleus, which endogenous CK1E (red) is also mis-localized to. The most right panels are the superimposition of Flag and CK1E images. Blue: DAPI nuclear staining.

287 and 694 in human FAM83H protein (Fig. 1.4). All of the mutations are expected to produce a mutant truncated protein. This genetic homogeneity of human *FAM83H* mutational spectrum not only suggests a dominant negative effect as a pathological mechanism but also provides valuable genetic information about the potential functional domains of this protein.

In this study, we demonstrate that FAM83H interacts with itself and probably dimerizes through interactions within its first 287 amino acids (the predicted PLD-like domain). This finding suggests that, under pathological conditions, the mutant truncated protein can interact with the wild-type protein and prevent it from functioning, which supports the dominant negative effect we proposed. Self-interactions in the N-terminal domain (1-287 amino acids) also provides a plausible explanation for why there are no reported disease-causing mutations N-terminal to

Ser^{287}. It is possible that the C-terminus of FAM83H may be the actual functional domain of this protein, while dimerization through the N-terminus may be critical for the protein to execute its function. Therefore, truncated proteins produced by disease-causing *FAM83H* mutations (from *FAM83H*^{287X} to *FAM83H*^{694X}) can still interact with the full-length FAM83H protein expressed from the wild-type allele, but the truncation-wild type protein complex fails to function. In contrast, any truncated FAM83H protein without a functional dimerization domain (shorter than FAM83H^{287X}) cannot bind to the wild-type protein. Half the normal number of wild-type dimers form so no disease phenotype is observed. However, further research needs to be conducted to demonstrate this hypothesis.

In addition to FAM83H self-interactions, we also demonstrated that FAM83H can interact with CK1 and the first 287 amino acids of FAM83H are necessary and sufficient for this interaction. A specific sequence motif, F^{270}-X-X-X-F^{274}-X-X-X-F^{278}, located at an evolutionarily conserved region N-terminal to Ser^{287}, seems to play a critical role in the FAM83H-CK1 interaction. This finding is further supported by a study of the CK1 interactome (Kategaya *et al.*, 2012) and a study about FAM83H in colorectal cancer (Kuga *et al.*, 2013). Furthermore, we showed that FAM83H interacts with SEC16A, a protein component of the COP II complex that is critical for ER-to-Golgi membrane trafficking, and that this interaction may be mediated by a domain in the middle of FAM83H (amino acids 287-657 in mouse). Along with a recent interesting finding that CK1 kinase activity is critical for vesicle trafficking between ER and Golgi (Lord *et al.*, 2011), our results provide an incomplete, but plausible pathological mechanism for *FAM83H*-associated AI.

It is known that the subcellular localization of CK1 is an important factor in its functional regulation (Knippschild *et al.*, 2005). Therefore, our finding that FAM83H interacts with CK1

and SEC16A via distinct domains suggest that FAM83H may function as a scaffold protein to bring CK1 to the place where SEC16A localizes, the ER exit site, so that CK1 can facilitate vesicle trafficking between ER and Golgi. If this hypothesis were true, all of the disease-causing *FAM83H* mutations (from *FAM83H*^{287X} to *FAM83H*^{694X}) would produce a truncated protein that is able to bind to CK1 through its N-terminus, but fails to localize to the ER exit site due to the absence of the SEC16A-interacting domain. In other words, the mutant truncated FAM83H would compete with wild-type FAM83H for binding to and correctly localizing CK1, causing the dominant negative effect. Our finding that over-expression of mouse FAM83H^{697X} leads to aberrant subcellular localization of CK1 supports this hypothesis.

This proposed mechanism also provides a plausible explanation for why there are no reported disease-causing mutations N-terminal to Ser^{287} or C-terminal to Glu^{694}. While a truncated protein shorter than FAM83H^{287X} lacks CK1-interacting domain and loses the ability to compete for CK1 binding so no dominant negative effect, a truncated protein longer than FAM83H^{694X} contains both CK1- and SEC16A-binding domains and is functionally sufficient to localize CK1 to the ER exit site. In both scenarios, the function of wild-type FAM83H would not be disturbed, so no pathology is observed. Instead, only truncated proteins with specific lengths (from FAM83H^{287X} to FAM83H^{694X}) that can bind to CK1 but not SEC16A would be disease-causing.

Based upon our proposed mechanism, we hypothesized that an evolutionarily-conserved sequence area immediately N-terminal to Glu^{694} of FAM83H might be responsible for interacting with SEC16A, so that a truncated protein shorter than human FAM83H^{694X} cannot bind to SEC16A, causing pathology. However, a truncated FAM83H without this conserved region (mouse FAM83H^{657X}) appeared to be sufficient for SEC16A binding, suggesting that this region

is not necessary for the FAM83H-SEC16A interaction. Moreover, if our proposed mechanism were true, any truncated FAM83H longer than human FAM83H^{694X} should be able to co-localize, at least partly, with SEC16A. Nevertheless, mouse FAM83H^{697X}, which contains the conserved region, showed aberrant localization to nucleus when over-expressed in HEK293 cells.

These unexpected results seemed to disprove our proposed pathological mechanism. However, there are several alternative explanations. First of all, we picked SEC16A for further investigation based upon our previous observation that over-expressed GFP-tagged FAM83H localized to the Golgi, which suggested that FAM83H might be involved in vesicle trafficking. However, the selection of potential FAM83H-interacting proteins for further study was arbitrary. It is possible that our hypothesis that FAM83H functions as a scaffold protein to bring CK1 to specific subcellular localizations might be still correct, but the localization is not through SEC16A. Recently, Kuga *et al.* reported that FAM83H interacted with cytokeratin and was associated with cytoskeletal organization (Kuga *et al.*, 2013), although we could not successfully reproduce the results with our experimental system (data not shown). Therefore, we may need to revisit FAM83H AP-MS data and test other plausible FAM83H-interacting proteins (including cytokeratins expressed by ameloblasts) to better assess the specific roles of FAM83H in cells.

Inconsistent with what was reported by Lee *et al.* (Lee *et al.*, 2011), our immunocytostaining showed nuclear localization of mouse FAM83H^{697X}. This discrepancy of protein localization might be due to artifacts from different protein-tag systems (GFP v.s. Flag) or because the FAM83H protein was from different species (human v.s. mouse). Also, in some cases, subcellular localization of proteins can be affected by cell-line differences and levels of protein over-expression. Therefore, our proposed mechanism that the truncated FAM83H longer

than human FAM83H^{694X} is able to correctly localize CK1 and is not disease-causing might still be valid. In order to demonstrate this molecular mechanism, in the future, we may consider using human FAM83H protein and ameloblast-like cell lines (or primary ameloblast culture) for *in vitro* protein localization experiments. Ideally, mouse models expressing different truncated FAM83H proteins would be the most physiologically relevant system for determining the localizations of mutant proteins in ameloblasts and for testing our proposed pathological mechanism, although such studies may not be cost-effective.

In this study, we also showed that recombinant mouse FAM83H can be phosphorylated by CK1 *in vitro*, and most of the phosphorylation sites are located in the C-terminus of FAM83H, although most of these phosphorylated Serines and Threonines are not evolutionarily conserved. Therefore, at this stage, we are not sure if FAM83H phosphorylation by CK1 is functionally or structurally significant. However, interestingly, there are two amino acid sequence areas (1025-1055, 1114-1139 in human FAM83H) at the C-terminus of FAM83H that show high sequence conservation during vertebrate evolution (Fig. 4.1E), suggesting that these regions may be important for FAM83H function. Given that FAM83H is a scaffold protein as we suspect, these two sequence areas may serve as protein docking sites. Using additional truncated FAM83H proteins to define the functions of these conserved domains seems to be the next logical pursuit.

In summary, by means of biochemical and molecular characterization, we showed here that FAM83H interacts with itself, CK1, and SEC16A through distinct domains, which suggests that FAM83H may serve as a scaffold protein for protein complex assembly. Based upon the protein-binding domains of FAM83H we defined as well as results from human mutational studies, we proposed a potential function for FAM83H that this protein is responsible for

correctly localizing CK1 to a specific subcellular domain and a pathological mechanism for *FAM83H*-associated AI that the disease phenotype comes from aberrant CK1 localization mediated by truncated FAM83H with specific lengths. Further investigations need to be conducted to test these hypotheses.

MATERIALS AND MATHODS

Expression constructs

Mouse *Fam83h* cDNA was cloned into phrGFP-C vector by Dr. Yumei Ding (Ding *et al.*, 2009). The *Fam83h* coding region from phrGFP-C-*Fam83h* was excised by double digestion with *Not*I and *Sal*I, and subcloned into pCMV-Tag4A and pCMV-Tag5A vectors (211174, 211175; Agilent Technologies; Santa Clara, CA, U.S.A.) to express C-terminal Flag-tagged and Myc-tagged mouse FAM83H. These two constructs were used for FAM83H self-interaction experiments. The pCMV-Tag4A-*Fam83h* was also used to over-express Flag-tagged FAM83H for affinity purification-mass spectrometry (AP-MS) experiment.

Seven constructs to express different domains, truncations, and mutations of FAM83H protein were generated by Custom DNA Constructs, LLC (Cleveland, OH, U.S.A.). They are pCMV-Tag2B-*Fam83h*, pCMV-Tag2B-$Fam83h^{287X}$, pCMV-Tag2B-$Fam83h^{657X}$, pCMV-Tag2B-$Fam83h^{697X}$, pCMV-Tag2B-$Fam83h^{288C}$, pCMV-Tag2B-$Fam83h^{698C}$, and pCMV-Tag2B-$Fam83h^{3FA}$. The pCMV-Tag2B-*Fam83h* construct encodes a full-length mouse FAM83H protein (1209 amino acids). The pCMV-Tag2B-$Fam83h^{287X}$, pCMV-Tag2B-$Fam83h^{657X}$, and pCMV-Tag2B-$Fam83h^{697X}$ constructs were made by introducing three respective premature stop codons into pCMV-Tag2B-*Fam83h* construct to express three truncated FAM83H proteins (1-287, 1-657, and 1-697 amino acids). The pCMV-Tag2B-$Fam83h^{288C}$ and pCMV-Tag2B-$Fam83h^{698C}$

constructs were designed to express two different-length C-terminal domain of FAM83H protein (288-1209 and 698-1209 amino acids). The pCMV-Tag2B-*Fam83h*^{3FA} construct was generated by introducing three site-directed mutations into pCMV-Tag2B-*Fam83h*, which encodes a full-length FAM83H protein with three Phenylalanine-to-Alanine substitutions (p.Phe270Ala, p.Phe274Ala, and p.Phe278Ala). All the encoded FAM83H proteins contain an N-terminal Flag tag. These six constructs were used for protein pull-down assays to study FAM83H protein-protein interaction.

Constructs expressing 6xMyc-tagged CK1δ and CK1ε (pcDNA3-CK1δ-6xMyc, pcDNA3-CK1ε-6xMyc) were gifts from Dr. Ying-Hui Fu (University of California San Francisco) (Kategaya *et al.*, 2012). The construct expressing Halo-tagged human SEC16A (FHC00048) and the HaloTag® control vector (G6591) were purchased from Promega Corporation (Madison, WI, U.S.A.).

Cell culture and plasmid transfection

Human HEK293 cells were cultured in Dulbeccos modified Eagle medium (DMED) (11995; Gibco® by Life Technologies; Grand Island, NY, U.S.A.) with 10% (v/v) fetal bovine serum (FBS) (16000; Gibco® by Life Technologies) in a 5% CO_2 humidified culture incubator. Cells were regularly passaged when reaching 95-100% confluency.

For transient plasmid transfection, cells in 2 mL DMEM was plated on 6-well plates the day before transfection so that the cell confluency could reach 75-80% on the day of transfection. 4 μg of plasmid in 10 μL of Lipofectamine® 2000 (11668; Invitrogen™ by Life Technologies) was diluted with 500 μL of Opti-MEM® I reduced serum medium (31985; Gibco® by Life Technologies), and incubated for 20 min at room temperature. The plasmid/Lipofectamine®

2000 complexes were then added to the culture media. After 6 h incubation, the culture media were changed to the regular media without transfection complexes, and the cells were further cultured for 42 h before harvested.

Immunoprecipitation (IP)

48 h following transfection, cells were washed twice with cold phosphate-buffered saline (PBS) and, for each well of 6-well plates, lysed with 500 μL NP40 cell lysis buffer (FNN0021; Novex® by Life Technologies) with 1 mM phenylmethylsulfonyl fluoride (PMSF) (P7626; Sigma-Aldrich; St. Louis, MO, U.S.A.) and 1X protease inhibitor cocktail (P2714; Sigma-Aldrich). After 30 min lysis, the lysates were collected and centrifuged at 15,000 rpm for 10 min. The supernatants were then used for subsequent analyses.

Dynabeads® protein A immunoprecipitation kit (10006D; Novex® by Life Technologies) was used for all the immunoprecipitation experiments. The experimental procedure followed the protocol provided by the manufacturer. In brief, anti-Flag antibody (1:20; F7425; Sigma-Aldrich) was incubated with protein A-attached Dynabeads for 30 min. The antibody/Dynabeads complexes were then mixed with the cell lysate and incubated for another 30 min at room temperature. After washed twice, the immunoprecipitates were eluted and assayed for SDS-PAGE and immunoblotting.

SDS-PAGE electrophoresis and Coomassie brilliant blue staining

Protein samples were prepared with 4X NuPAGE® LDS sample buffer (NP0007; Novex® by Life Technologies) in a 1:3 ratio and run on NuPAGE® 4-12% Bis-Tris protein gels (NP0322BOX; Novex® by Life Technologies) with NuPAGE® MES SDS running buffer

(NP0002; Novex® by Life Technologies). Electrophoresis was performed at a constant voltage of 200V for 30-35 min.

After electrophoresis, the protein gels were washed in distilled water for 3X 5 min, stained with SimplyBlue^{TM} SafeStain (LC6065; Novex® by Life Technologies) for 40-60 min, and de-stained in distilled water for several hours.

Western blot analysis

After SDS-PAGE, the proteins on the gel were transblotted onto a nitrocellulose membrane (LC2000; Novex® by Life Technologies) with NuPAGE® transfer buffer (NP0006; Novex® by Life Technologies) at a constant voltage of 30V for 70-100 min. The membrane was subsequently blocked with 5% non-fat milk (170-6404XTU; Bio-Rad Laboratories; Hercules, CA, U.S.A.) in TBS-Tween (20 mM Tris, 500mM NaCl, 0.1% Tween-20, pH 7.5) for 1 h, and incubated with primary antibody for 2 h at room temperature or overnight at 4˚C. After washed with TBS-Tween for 3X 10 min, the membrane was incubated with HRP-conjugated secondary antibody for 1 h and then washed for 3X 20 min. Blots were immersed with chemiluminescence ECL western blotting detection reagent (RPN2232; GE Healthcare; Little Chalfont, U.K.) for 5 min and exposed to a film.

Primary antibodies used for western blot analysis include: anti-FLAG® (F7425) (1:2000), anti-FLAG® M2 (F1804) (1:4000), anti-c-Myc (C3956) (1:2000) from Sigma-Aldrich (St. Louis, MO, U.S.A.); anti-Myc Tag, clone 4A6 (05-724) (1:4000) from Millipore Corporation (Billerica, MA, U.S.A.); anti- CK1α1 [EPR1961(2)] (ab108296) (1:2000), anti-CK1δ [AF12G4] (ab85320) (1:4000), anti-CK1ε [AF6C1] (ab82426) (1:2000), anti-SEC16A (ab70722) (1:2000) from abcam® (Cambridge, MA, U.S.A.); anti-HaloTag® (G9211) (1:1000)

from Promega Corporation (Madison, WI, U.S.A.). Two secondary antibodies were used: ECL anti-rabbit IgG (NA934V; GE Healthcare; Little Chalfont, U.K.) and anti-mouse IgG+IgM HRP (ab47827; abcam®; Cambridge, MA, U.S.A.).

Protein-protein interaction modeling

SPRING ON-LINE (http://zhanglab.ccmb.med.umich.edu/spring/) (Guerler *et al.*, 2013) from Zhang laboratory at University of Michigan was used for modeling potential FAM83H dimerization. Sequence of the first 287 amino acids (PLD-like domain) from human FAM83H protein was used for both query sequence A and query sequence B.

Affinity purification combined with mass spectrometry (AP-MS)

Flag-tagged mouse FAM83H protein was over-expressed by transiently-transfected HEK293 cells on a 10-cm Petri dish. The cell lysate from all the harvested cells underwent immunoprecipitation (with anti-FLAG® antibody), and the immunoprecipitate was further fractionated with SDS-PAGE, as described in the above sections. After Coomassie brilliant blue staining, 6 specific protein bands were sliced out and submitted to Keck Biotechnology Resource Laboratory at Yale University, where trypsinization, LC-MS/MS mass spectrometry for protein identification, and subsequent data analysis were performed.

***In vitro* kinase reaction**

For each kinase reaction, 0.2 mg purified mouse recombinant FAM83H was incubated with 5000 units of casein kinase 1 (P6030S), casein kinase 2 (P6010S; New England BioLabs®; Ipswich, MA, U.S.A.), or no enzyme and 0.05μCi ^{33}P-ATP in a total reaction volume of 20 μL.

The reactions were conducted at 30˚C for 60 min and subsequently fractionated with SDS-PAGE of Novex® 4-20% Tris-Glycine protein gels (EC60255BOX; Novex® by Life Technologies). After drying, the gel was exposed to a film for 20 min. The kinase assays were mainly conducted by Drs. Jan Hu and Yasuo Yamakoshi.

For determination of CK1 phosphorylation sites on FAM83H protein, mass spectrometry was used. The kinase reaction was conducted using above-mentioned method except that non-radioactive cold ATP was used. The reaction was fractionated with SDS-PAGE. After Coomassie brilliant blue staining, the band of FAM83H was sliced out and submitted to Keck Biotechnology Resource Laboratory at Yale University, where trypsinization, LC-MS/MS mass spectrometry for protein post-translational modification identification, and subsequent data analysis were performed.

Immunocytochemistry (ICC)

HEK293 cells were cultured in Lab-Tek chamber slides (1 chamber) with cover (70360-12; Electron Microscopy Sciences; Hatfield, PA, U.S.A.) and transfected with pCMV-Tag2B-*Fam83h*, pCMV-Tag2B-*Fam83h*^{697X}, or control empty vector. After 18 h, the cells were fixed with 100% methanol for 15 min at -20˚C, washed with PBS buffer for 3 times. Following blocking with 5% sheep serum (S22; Millipore Corporation; Billerica, MA, U.S.A.) in PBT buffer (0.1% Triton X-100 in PBS buffer) for 30 min at room temperature, anti-FLAG® antibody (1:200; F7425; Sigma-Aldrich; St. Louis, MO, U.S.A.) and anti-CK1ε antibody (1:200; ab82426; abcam®; Cambridge, MA, U.S.A.) were applied. After overnight incubation of primary antibody at 4˚C, the cells were washed with PBS buffer for 15 min and then incubated for 30 min at room temperature in solutions containing anti-rabbit IgG secondary antibody conjugated with Alexa

Fluor 488 (1:500; A-11008; Molecular Probes® by Life Technologies) and anti-mouse IgG1 secondary antibody conjugated with Alexa Fluor 594 (1:500; A-21125; Molecular Probes® by Life Technologies). The slides were then rinsed in PBS buffer for 15 min, mounted with ProLong® Gold antifade reagent with DAPI (P-36941; Molecular Probes® by Life Technologies), and examined under a Leica DM5000B fluorescence microscope.

REFERENCES

Ding Y, Estrella MR, Hu YY, Chan HL, Zhang HD, Kim JW, Simmer JP, Hu JC (2009). Fam83h is associated with intracellular vesicles and ADHCAI. *J Dent Res* 88:991-6.

Dunham WH, Mullin M, Gingras AC (2012). Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* 12:1576-90.

El-Sayed W, Shore RC, Parry DA, Inglehearn CF, Mighell AJ (2010). Ultrastructural analyses of deciduous teeth affected by hypocalcified amelogenesis imperfecta from a family with a novel Y458X FAM83H nonsense mutation. *Cells Tissues Organs* 191:235-9.

Gingras AC, Gstaiger M, Raught B, Aebersold R (2007). Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* 8:645-54.

Guerler A, Govindarajoo B, Zhang Y (2013). Mapping Monomeric Threading to Protein-Protein Structure Prediction. *J Chem Inf Model* 53:717-25.

Hart PS, Becerik S, Cogulu D, Emingil G, Ozdemir-Ozenen D, Han ST, Sulima PP, Firatli E, Hart TC (2009). Novel FAM83H mutations in Turkish families with autosomal dominant hypocalcified amelogenesis imperfecta. *Clin Genet* 75:401-4.

Haubek D, Gjørup H, Jensen LG, Juncker I, Nyegaard M, Børglum AD, Poulsen S, Hertz JM (2011). Limited phenotypic variation of hypocalcified amelogenesis imperfecta in a Danish five-generation family with a novel FAM83H nonsense mutation. *Int J Paediatr Dent* 21:407-12.

Hyun HK, Lee SK, Lee KE, Kang HY, Kim EJ, Choung PH, Kim JW (2009). Identification of a novel FAM83H mutation and microhardness of an affected molar in autosomal dominant hypocalcified amelogenesis imperfecta. *Int Endod J* 42:1039-43.

Iinuma T, Shiga A, Nakamoto K, O'Brien MB, Aridor M, Arimitsu N, Tagaya M, Tani K (2007). Mammalian Sec16/p250 plays a role in membrane traffic from the endoplasmic reticulum. *J Biol Chem* 282:17632-9.

Kategaya LS, Hilliard A, Zhang L, Asara JM, Ptáček LJ, Fu YH (2012). Casein kinase 1 proteomics reveal prohibitin 2 function in molecular clock. *PLoS One* 7:e31987.

Kim JW, Simmer JP, Lin BP, Seymen F, Bartlett JD, Hu JC (2006). Mutational analysis of candidate genes in 24 amelogenesis imperfecta families. *Eur J Oral Sci* 114 Suppl 1:3-12; discussion 39-41, 379.

Kim JW, Lee SK, Lee ZH, Park JC, Lee KE, Lee MH, Park JT, Seo BM, Hu JC, Simmer JP (2008). FAM83H mutations in families with autosomal-dominant hypocalcified amelogenesis imperfecta. *Am J Hum Genet* 82:489-94.

Knippschild U, Gocht A, Wolff S, Huber N, Löhler J, Stöter M (2005). The casein kinase 1 family: participation in multiple cellular processes in eukaryotes. *Cell Signal* 17:675-89.

Kuga T, Kume H, Kawasaki N, Sato M, Adachi J, Shiromizu T, Hoshino I, Nishimori T, Matsubara H, Tomonaga T (2013). A novel mechanism of keratin cytoskeleton organization through casein kinase Iα and FAM83H in colorectal cancer. *J Cell Sci* 126:4721-31.

Lee SK, Hu JC, Bartlett JD, Lee KE, Lin BP, Simmer JP, Kim JW (2008). Mutational spectrum of FAM83H: the C-terminal portion is required for tooth enamel calcification. *Hum Mutat* 29:E95-9.

Lee SK, Lee KE, Jeong TS, Hwang YH, Kim S, Hu JC, Simmer JP, Kim JW (2011). FAM83H mutations cause ADHCAI and alter intracellular protein localization. *J Dent Res* 90:377-81.

Lord C, Bhandari D, Menon S, Ghassemian M, Nycz D, Hay J, Ghosh P, Ferro-Novick S (2011). Sequential interactions with Sec23 control the direction of vesicle traffic. *Nature* 473:181-6.

Mendoza G, Pemberton TJ, Lee K, Scarel-Caminaga R, Mehrian-Shai R, Gonzalez-Quevedo C, Ninis V, Hartiala J, Allayee H, Snead ML, Leal SM, Line SR, Patel PI (2007). A new locus for autosomal dominant amelogenesis imperfecta on chromosome 8q24.3. *Hum Genet* 120:653-62.

Miller EA, Barlowe C (2010). Regulation of coat assembly--sorting things out at the ER. *Curr Opin Cell Biol* 22:447-53.

Okamura H, Garcia-Rodriguez C, Martinson H, Qin J, Virshup DM, Rao A (2004). A conserved docking motif for CK1 binding controls the nuclear localization of NFAT1. *Mol Cell Biol* 24:4184-95.

Song YL, Wang CN, Zhang CZ, Yang K, Bian Z (2012). Molecular characterization of amelogenesis imperfecta in Chinese patients. *Cells Tissues Organs* 196:271-9.

Stuckey JA, Dixon JE (1999). Crystal structure of a phospholipase D family member. *Nat Struct Biol* 6:278-84.

Wright JT, Frazier-Bowers S, Simmons D, Alexander K, Crawford P, Han ST, Hart PS, Hart TC (2009). Phenotypic variation in FAM83H-associated amelogenesis imperfecta. *J Dent Res* 88:356-60.

Wright JT, Torain M, Long K, Seow K, Crawford P, Aldred MJ, Hart PS, Hart TC (2011). Amelogenesis imperfecta: genotype-phenotype studies in 71 families. *Cells Tissues Organs* 194:279-83.

Zanetti G, Pahuja KB, Studer S, Shim S, Schekman R (2012). COPII and the regulation of protein sorting in mammals. *Nat Cell Biol* 14:20-8.

CHAPTER 5

CONCLUSION

SUMMARY

Dental caries and periodontal disease are two of the most prevalent infectious diseases in humans (Bagramian *et al.*, 2009; Dentino *et al.*, 2013). They cause destruction of tooth structures and supporting tissues and eventually lead to loss of teeth. To date, dental amalgam and composite resin are two major restoration materials for treatment of dental caries. However, the physical properties of these two materials are far from comparable to those of true enamel, resulting in various adverse consequences of restoration treatment (Sharif *et al.*, 2010a; Sharif *et al.*, 2010b). Therefore, producing and using true enamel as a restoration material have been aggressively pursued (Nör, 2006). Furthermore, many efforts have been made to regenerate whole functional teeth for patients who lose teeth due to extensive tooth decay or periodontal disease (Ikeda *et al.*, 2009; Galler and D'Souza, 2011). However, all of these goals can only be achieved, based on thorough comprehension of enamel formation and tooth development in which the genetic control of these developmental processes has not yet been completely understood.

Human inherited dental anomalies provide a valuable resource for studying tooth development (Cobourne and Sharpe, 2013). Discerning the genetic etiology of these

developmental defects not only improves our comprehension of normal tooth development but also provides a scientific foundation for developing potential preventive and therapeutic strategies for these disorders. While tooth agenesis is a tooth-number abnormality that is caused by disruptions at early tooth development, amelogenesis imperfecta is a collection of inherited enamel defects due to disturbances in enamel formation. Identification of genes associated with these two diseases has led to the discovery of many critical players in tooth and enamel development (Cobourne and Sharpe, 2013). However, there is still a significant number of cases for which genetic defects cannot be found in known disease candidate genes, indicating our current incomplete knowledge of the critical molecular participants in tooth and enamel formation (Chan *et al.*, 2011; Arte *et al.*, 2013).

In this thesis, we studied families with tooth agenesis and amelogenesis imperfecta (AI) and aimed to define the genetic etiology of the disorder in each family and to discover novel disease associated mutations and genes. We applied both target gene approaches and whole exome sequencing for mutational analysis. The results for tooth agenesis and AI were described in Chapter 2 and 3 respectively.

For mutational analysis of tooth agenesis, we described 7 families with non-syndromic tooth agenesis, in which there was considerable variation in the number and class of teeth that were involved. We identified a novel *PAX9* mutation (c.43T>A, p.Phe15Ile) in a proband with 10 missing teeth and two reported *WNT10A* mutations (c.321C>A, p.Cys107* and c.682T>A, p.Phe228Ile) in a family where the compound heterozygotes exhibited severe oligodontia without other non-dental OODD phenotypes. However, we were unable to determine the genetic defects in other families using target gene approaches of known candidate genes or by

comparative whole exome analyses. The results demonstrate a high genetic heterogeneity of tooth agenesis and the difficulty of establishing genotype-phenotype correlation in this disorder.

For mutational analysis of inherited enamel defects, we described 12 AI families with various enamel malformations and different patterns of disease inheritance. We successfully identified genetic defects in these families. Novel *MMP20* (c.611A>G, p.His204Arg) and *KLK4* (c.245delG, p.Gly82Alafs*87) mutations reaffirmed the critical roles of enamel matrix proteases in amelogenesis, especially enamel maturation (Hu *et al.*, 2007; Bartlett, 2013). The enamel phenotypic heterogeneity of *MMP20*-associated AI also suggests unrecognized functions of MMP20 other than processing and degradation of enamel matrix proteins (Wang *et al.*, 2013b). A novel *FAM83H* mutation (c.1369C>T, p.Gln457*) re-demonstrated the genetic homogeneity of *FAM83H* disease-causing mutations and supported the hypothesis of dominant negative effects as the pathological mechanism of ADHCAI. Novel *SLC24A4* (c.437C>T, p.Ala146Val) and *STIM1* (c.1276C>T, p.Arg426Cys) mutations, along with the specific expression of these two genes in maturation-stage ameloblasts demonstrated by immunohistochemistry, reveal the critical roles of these two Ca^{2+}-transport related proteins in enamel maturation and the important differences in the Ca^{2+} transcellular transport systems employed by secretory and maturation stage ameloblasts (Wang *et al.*, 2014b). Furthermore, we identified 7 *FAM20A* mutations in 5 families of AIGFS with/without nephrocalcinosis, established *FAM20A* mutations as a cause of Enamel-Renal Syndrome (ERS), and proposed that AIGFS and ERS are the same disease entity with variable presentation of renal calcifications (Wang *et al.*, 2013a; Wang *et al.*, 2014a). We also reported the first human Integrin beta 6 (*ITGB6*) mutations causing generalized hypoplastic AI, which demonstrates that cell-matrix interaction and integrin signaling are critical for enamel formation. The results strongly suggest that amelogenesis is a complicated biological process

depending upon extensive cell-cell and cell-matrix interactions, which poses significant challenges to the validity of growing enamel crystals "biomimetically" with only selected recombinant enamel matrix proteins (Wang *et al.*, 2013c). For many other AI families not described in this thesis, we are still not able to identify the genetic causes of their enamel defects, which reaffirms the high genetic heterogeneity of AI and revealed our current incomprehensive knowledge about enamel formation.

In addition to finding critical players in tooth and enamel development, researchers also aim for defining the functions of many genes identified by human mutational analyses of inherited dental anomalies, in order to understand the genetic regulation and molecular mechanisms of these developmental processes. Therefore, in this study, we attempted to determine the functions of *FAM83H*, a gene of which the mutations cause ADHCAI, at cellular and molecular levels and to unravel the pathological mechanism of *FAM83H*-associated AI (Kim *et al.*, 2008).

By studying FAM83H protein-protein interaction, we found that FAM83H interacts with itself, which suggests that the mutant protein may exert the dominant negative effect through interacting with the wild-type protein and prevent it from functioning. We also demonstrated that FAM83H interacts with casein kinase 1 (CK1) via an F-X-X-X-F-X-X-X-F sequence motif in its N-terminus and with SEC16A via a binding domain in the middle of the protein. Knowing SEC16A is involved in ER-to-Golgi membrane trafficking, we proposed that FAM83H may serve as a scaffold protein for localizing CK1 to SEC16A so that CK1 can facilitate membrane and protein trafficking at the ER exit site, which is an active cellular process during the secretory stage of amelogenesis. We hypothesized that truncated FAM83H that can bind to CK1 but not SEC16A will disturb localization of CK1, result in altered vesicle trafficking and protein

transport, and lead to a pathology in ameloblasts. However, further investigations are needed to confirm these potential cellular functions of FAM83H and to substantiate the hypothetical mechanism of dominant negative effect mediated pathogenesis of ADHCAI.

In summary, our group pursued a forward genetics-based approach in this study to investigate tooth and enamel development. We identified genes associated with developmental tooth defects and investigated the functions of genes in normal developmental processes as well as the pathological mechanisms of their associated disorders. By discerning the genetic etiology of tooth agenesis and AI through mutational analysis, we reaffirmed and discovered critical players in tooth and enamel formation, such as *MMP20*, *KLK4*, and *ITGB6*. For functional investigations, we focused on the recently-discovered AI-causing gene, *FAM83H*, defined a potential function of facilitating vesicle and protein trafficking, and tested a hypothetical model for the pathogenesis of FAM83H-associated ADHCAI. This phenotype-to-genotype approach, although bearing more difficulties for defining gene functions, provides the most relevant way to study tooth development.

PROSPECTS

Human mutational analysis for tooth agenesis has revealed many critical genes and signaling pathways in early tooth development (Cobourne and Sharpe, 2013). However, the actual roles of these players in tooth formation and the pathological mechanism of their associated tooth agenesis are not very clear. This is partly due to the limitations of widely-used mouse models for tooth development. Unlike humans with two dentitions, rodents have only one, which is analogous to humans' primary dentition (Jernvall and Thesleff, 2012). However, as

mentioned previously, human tooth agenesis affects permanent (secondary) dentition much more frequently than primary dentition (Nieminen, 2009). Also, the developmental molecular mechanisms for the formation of two dentitions appear to be different. While primary teeth initiate from *de novo* dental laminae, permanent teeth develop from the dental laminae of their predecessor primary teeth (Jernvall and Thesleff, 2012). Therefore, the mouse model is not optimal for studying secondary tooth formation and the pathological mechanisms of human tooth agenesis. For example, human loss-of-function mutations of *LTBP3* were reported to cause severe oligodontia and short stature (Noor *et al.*, 2009); however, the *Ltbp3* null mice exhibited no dental abnormalities (Dabovic *et al.*, 2002), which demonstrates the importance of using other animals as a disease model for human tooth agenesis. Pigs have long been used in dental research, although mainly for protein characterization of developing teeth (Yamakoshi *et al.*, 2006). Like humans, pigs have two dentitions, and their dental formulae (I_3-C_1-M_3; I_3-C_1-P_4-M_3) are similar to those of humans (I_2-C_1-M_2; I_2-C_1-P_2-M_3). Thus pigs offer important advantages over mice as an animal model for studying tooth number abnormalities. Recently, several genome engineering techniques, such as ZFN (zinc finger nucleases), TALENs (transcription activator-like effector nucleases), and CRISPR/Cas system, have made it possible to efficiently target and edit specific genomic areas in many organisms other than mice (Le Provost *et al.*, 2010; Joung and Sander, 2013; Mali *et al.*, 2013; Wei *et al.*, 2013), providing an avenue for using porcine models to study tooth renewal and human tooth number abnormalities.

For studying enamel development, scientists have taken advantage of the abundance of enamel matrix proteins in developing teeth to structurally and biochemically characterize these proteins to gain insights into their potential roles in dental enamel formation (Hu *et al.*, 2007). However, investigation of enamel matrix proteins from developing teeth seemed to have been

thoroughly explored. With traditional approaches, it is not easy to identify trace amount of proteins and low-expressed genes which may also play significant roles in enamel formation (Yamakoshi *et al.*, 2006). Therefore, recently, the direction and approaches to study enamel formation have been shifted from a biochemical aspect to a genetic aspect, meaning that by discerning genetic causes of inherited enamel defects, new critical players for enamel development can be identified (Wright *et al.*, 2011; Chan *et al.*, 2011). However, as demonstrated in this study, many efforts still need to be made to define the genetic defects of many AI cases with unknown etiology. As many mouse models have been shown to phenocopy human AI (Gibson *et al.*, 2001; Caterina *et al.*, 2002; Hu *et al.*, 2008; Simmer *et al.*, 2009), careful characterization of the dental phenotypes in high-throughput knockout mouse project, KOMP (Ayadi *et al.*, 2012), will identify more critical players in enamel formation and facilitate human genetic analyses into the causes of enamel malformations through the discovery of new candidate genes and the biological validation that defects in specific genes result in enamel defects. In other words, both forward-genetics (in humans) and reverse-genetics (in mice) based approaches need to be complementarily used to advance the enamel research. Furthermore, current knowledge about molecular and genetic regulation of the developmental process of enamel formation is fragmentary, mainly based upon the fact that individual gene defects cause enamel malformations. Therefore, following the list of AI candidate genes being expanded, more efforts need to be made to study gene-gene or protein-protein interactions in enamel formation from the perspective of molecular pathways or networks. For example, as mentioned previously, *SLC24A4*, *STIM1*, and *ORAI1* mutations cause hypomaturation AI, and all these genes are involved in calcium transport in the cell (Feske *et al.*, 2006; Picard *et al.*, 2009; Parry *et al.*, 2013). By elucidating how these molecules function cooperatively in ameloblasts, we will be

able to understand the underlying mechanism of calcium transport and homeostasis during enamel maturation (Wang *et al.*, 2014b).

Our *in vitro* molecular characterization of FAM83H suggests a potential cellular function of FAM83H and a hypothetical model for the pathogenesis of its associated enamel defects. However, these results and hypotheses cannot explain all observations and need to be further tested and confirmed by *in vivo* models. A conventional or conditional knockout (KO) mouse model is necessary to define the function of *Fam83h* in normal enamel development. However, in order to study the pathological mechanism of ADHCAI, a disease (truncation) mutation knockin (KI) model may be required, since a dominant negative effect from mutant FAM83H is suspected to be the pathogenesis of enamel defects. Therefore, our group has generated both mouse models, a conventional KO (*Fam83h*$^{-/-}$) and a disease mutation KI (*Fam83h*$^{Y297X/Y297X}$). With these two models, we will be able to test the *in vitro* results we presented here. For example, immunostaining of FAM83H and CK1 in KI mice will reveal if the mutant FAM83H alters CK1 localization in ameloblasts. Also, primary culture of ameloblasts from both KO and KI mice will allow us to conduct *ex vivo* experiments to define the normal and aberrant functions of FAM83H protein at molecular and cellular levels.

Recently, patient-specific induced pluripotent stem cells (iPSCs) have been used as a model to study human diseases, such as neurological disorders (Bellin *et al.*, 2012; Peitz *et al.*, 2013). Somatic cells harvested from patients are induced into iPS cells, which carry the patient-specific genetic defect, and re-differentiated into specific disease-affected cells and tissues, such as neurons for neurological disorders. These cells can be used not only to study the molecular mechanism of disease pathogenesis but also to develop potential therapeutics for the disease. This approach is especially valuable for studying diseases in which there is limited experimental

access to disease-affected human tissues (Bellin *et al.*, 2012). Ameloblasts are only present during enamel development and cannot be obtained following tooth eruption. Therefore, AI patient-specific iPS cells may serve as a potential disease model to study aberrant amelogenesis. However, although people have reprogrammed mouse iPS cells into "ameloblast-like" cells (Liu *et al.*, 2013), significant efforts still need to be made to establish specific protocols to differentiate human iPS cells into functional ameloblasts and recapitulate the processes of amelogenesis *ex vivo*.

REFERENCES

Arte S, Parmanen S, Pirinen S, Alaluusua S, Nieminen P (2013). Candidate gene analysis of tooth agenesis identifies novel mutations in six genes and suggests significant role for WNT and EDA signaling and allele combinations. *PLoS One* 8:e73705.

Ayadi A, Birling MC, Bottomley J, Bussell J, Fuchs H, Fray M, Gailus-Durner V, Greenaway S, Houghton R, Karp N, Leblanc S, Lengger C, Maier H, Mallon AM, Marschall S, Melvin D, Morgan H, Pavlovic G, Ryder E, Skarnes WC, Selloum M, Ramirez-Solis R, Sorg T, Teboul L, Vasseur L, Walling A, Weaver T, Wells S, White JK, Bradley A, Adams DJ, Steel KP, Hrabě de Angelis M, Brown SD, Herault Y (2012). Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. *Mamm Genome* 23:600-10.

Bagramian RA, Garcia-Godoy F, Volpe AR (2009). The global increase in dental caries. A pending public health crisis. *Am J Dent* 22:3-8.

Bartlett JD (2013). Dental Enamel Development: Proteinases and Their Enamel Matrix Substrates. *ISRN Dent* 2013:684607.

Bellin M, Marchetto MC, Gage FH, Mummery CL (2012). Induced pluripotent stem cells: the new patient? *Nat Rev Mol Cell Biol* 13:713-26.

Caterina JJ, Skobe Z, Shi J, Ding Y, Simmer JP, Birkedal-Hansen H, Bartlett JD (2002). Enamelysin (matrix metalloproteinase 20)-deficient mice display an amelogenesis imperfecta phenotype. J Biol Chem 277:49598-604.

Chan HC, Estrella NM, Milkovich RN, Kim JW, Simmer JP, Hu JC (2011). Target gene analyses of 39 amelogenesis imperfecta kindreds. *Eur J Oral Sci* 119 Suppl 1:311-23.

Cobourne MT, Sharpe PT (2013). Diseases of the tooth: the genetic and molecular basis of inherited anomalies affecting the dentition. *Wiley Interdiscip Rev Dev Biol* 2:183-212.

Dabovic B, Chen Y, Colarossi C, Obata H, Zambuto L, Perle MA, Rifkin DB (2002). Bone abnormalities in latent TGF-β binding protein (Ltbp)-3-null mice indicate a role for Ltbp-3 in modulating TGF-β bioavailability. *J Cell Biol* 156:227-32.

Dentino A, Lee S, Mailhot J, Hefti AF (2013). Principles of periodontology. *Periodontol 2000* 61:16-53.

Feske S, Gwack Y, Prakriya M, Srikanth S, Puppel SH, Tanasa B, Hogan PG, Lewis RS, Daly M, Rao A (2006). A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. Nature 441:179-85.

Galler KM, D'Souza RN (2011). Tissue engineering approaches for regenerative dentistry. *Regen Med* 6:111-24.

Gibson CW, Yuan ZA, Hall B, Longenecker G, Chen E, Thyagarajan T, Sreenath T, Wright JT, Decker S, Piddington R, Harrison G, Kulkarni AB (2001). Amelogenin-deficient mice display an amelogenesis imperfecta phenotype. *J Biol Chem* 276:31871-5.

Hu JC, Chun YH, Al Hazzazzi T, Simmer JP (2007). Enamel formation and amelogenesis imperfecta. *Cells Tissues Organs* 186:78-85.

Hu JC, Hu Y, Smith CE, McKee MD, Wright JT, Yamakoshi Y, Papagerakis P, Hunter GK, Feng JQ, Yamakoshi F, Simmer JP (2008). Enamel defects and ameloblast-specific expression in Enam knock-out/lacz knock-in mice. *J Biol Chem* 283:10858-71.

Ikeda E, Morita R, Nakao K, Ishida K, Nakamura T, Takano-Yamamoto T, Ogawa M, Mizuno M, Kasugai S, Tsuji T (2009). Fully functional bioengineered tooth replacement as an organ replacement therapy. *Proc Natl Acad Sci USA* 106:13475-80.

Jernvall J, Thesleff I (2012). Tooth shape formation and tooth renewal: evolving with the same signals. *Development* 139:3487-97.

Joung JK, Sander JD (2013). TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 14:49-55.

Kim JW, Lee SK, Lee ZH, Park JC, Lee KE, Lee MH, Park JT, Seo BM, Hu JC, Simmer JP (2008). FAM83H mutations in families with autosomal-dominant hypocalcified amelogenesis imperfecta. *Am J Hum Genet* 82:489-94.

Le Provost F, Lillico S, Passet B, Young R, Whitelaw B, Vilotte JL (2010). Zinc finger nuclease technology heralds a new era in mammalian transgenesis. *Trends Biotechnol* 28:134-41.

Liu L, Liu YF, Zhang J, Duan YZ, Jin Y (2013). Ameloblasts serum-free conditioned medium: bone morphogenic protein 4-induced odontogenic differentiation of mouse induced pluripotent stem cells. *J Tissue Eng Regen Med* [Epub ahead of print]

Mali P, Esvelt KM, Church GM (2013). Cas9 as a versatile tool for engineering biology. *Nat Methods* 10:957-63.

Nieminen P (2009). Genetic basis of tooth agenesis. *J Exp Zool B Mol Dev Evol* 312B:320-42.

Noor A, Windpassinger C, Vitcu I, Orlic M, Rafiq MA, Khalid M, Malik MN, Ayub M, Alman B, Vincent JB (2009). Oligodontia is caused by mutation in LTBP3, the gene encoding latent TGF-beta binding protein 3. *Am J Hum Genet* 84:519-23.

Nör JE (2006). Tooth regeneration in operative dentistry. *Oper Dent* 31:633-42.

Parry DA, Poulter JA, Logan CV, Brookes SJ, Jafri H, Ferguson CH, Anwari BM, Rashid Y, Zhao H, Johnson CA, Inglehearn CF, Mighell AJ (2013). Identification of mutations in SLC24A4, encoding a potassium-dependent sodium/calcium exchanger, as a cause of amelogenesis imperfecta. *Am J Hum Genet* 92:307-12.

Peitz M, Jungverdorben J, Brüstle O (2013). Disease-specific iPS cell models in neuroscience. *Curr Mol Med* 13:832-41.

Picard C, McCarl CA, Papolos A, Khalil S, Lüthy K, Hivroz C, LeDeist F, Rieux-Laucat F, Rechavi G, Rao A, Fischer A, Feske S (2009). STIM1 mutation associated with a syndrome of immunodeficiency and autoimmunity. *N Engl J Med* 360:1971-80.

Sharif MO, Merry A, Catleugh M, Tickle M, Brunton P, Dunne SM, Aggarwal VR (2010). Replacement versus repair of defective restorations in adults: amalgam. *Cochrane Database Syst Rev* (2):CD005970.

Sharif MO, Catleugh M, Merry A, Tickle M, Dunne SM, Brunton P, Aggarwal VR (2010). Replacement versus repair of defective restorations in adults: resin composite. *Cochrane Database Syst Rev* (2):CD005971.

Simmer JP, Hu Y, Lertlam R, Yamakoshi Y, Hu JC (2009). Hypomaturation enamel defects in Klk4 knockout/LacZ knockin mice. *J Biol Chem* 284:19110-21.

Wang SK, Aref P, Hu Y, Milkovich RN, Simmer JP, El-Khateeb M, Daggag H, Baqain ZH, Hu JC (2013a). FAM20A mutations can cause enamel-renal syndrome (ERS). *PLoS Genet* 9:e1003302.

Wang SK, Hu Y, Simmer JP, Seymen F, Estrella NM, Pal S, Reid BM, Yildirim M, Bayram M, Bartlett JD, Hu JC (2013b). Novel KLK4 and MMP20 mutations discovered by whole-exome sequencing. *J Dent Res* 92:266-71.

Wang SK, Choi M, Richardson AS, Reid BM, Lin BP, Wang SJ, Kim JW, Simmer JP, Hu JC (2013c). ITGB6 loss-of-function mutations cause autosomal recessive amelogenesis imperfecta. *Hum Mol Genet* [Epub ahead of print]

Wang SK, Reid BM, Dugan SL, Roggenbuck JA, Read L, Aref P, Taheri AP, Yeganeh MZ, Simmer JP, Hu JC (2014a). FAM20A Mutations Associated with Enamel Renal Syndrome. *J Dent Res* 93:42-8.

Wang SK, Choi M, Richardson AS, Reid BM, Seymen F, Yildirim M, Tuna EB, Gencay K, Simmer JP, Hu JC (2014b). *J Dent Res* [Epub ahead of print]

Wei C, Liu J, Yu Z, Zhang B, Gao G, Jiao R (2013). TALEN or Cas9 - rapid, efficient and specific choices for genome modifications. *J Genet Genomics* 40:281-9.

Wright JT, Torain M, Long K, Seow K, Crawford P, Aldred MJ, Hart PS, Hart TC (2011). Amelogenesis imperfecta: genotype-phenotype studies in 71 families. *Cells Tissues Organs* 194:279-83.

Yamakoshi Y, Hu JC, Zhang H, Iwata T, Yamakoshi F, Simmer JP (2006). Proteomic analysis of enamel matrix using a two-dimensional protein fractionation system. *Eur J Oral Sci* 114 Suppl 1:266-71.