STUDY OF CHROMOSOME 11q22.2q22.3 FOR LINKAGE TO CLASS III MALOCCLUSION IN SOUTH AMERICANS

by

Loulwah Khalid AlOthman

D.D.S, King Saud University, 2009

Submitted to the Graduate Faculty of

The School of Dental Medicine in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

SCHOOL OF DENTAL MEDICINE

This thesis was presented

by

Loulwah Khalid AlOthman

It was defended on

March 27, 2015

and approved by

James K. Hartsfield, Jr., DMD, PhD., Professor, E. Preston Hicks Endowed Chair in Orthodontics and Oral Health Research, University of Kentucky, College of Dentistry

Mary Marazita, PhD., Professor, Vice Chair of the Department of Oral Biology, School of Dental Medicine, University of Pittsburgh

Mark Mooney, PhD., Professor, Chair of the Department of Oral Biology, School of Dental Medicine, University of Pittsburgh

Thesis Director: Manika Govil, PhD., Assistance Professor, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh

STUDY OF CHROMOSOME 11q22.2q22.3 FOR LINKAGE TO CLASS III MALOCCLUSION IN SOUTH AMERICANS

Loulwah Khalid AlOthman, D.D.S, M.S.

University of Pittsburgh, 2015

Class III malocclusion is one of the dentofacial deformities that represents a challenge for orthodontists in terms of treatment and prognosis. Due to its complexity and aesthetic involvement, a lot of research have been undertaken to understand the mechanisms underlying the development of this growth deformity. Several studies have suggested a strong genetic contribution in the formation of class III malocclusion. Previous studies have implicated a region on chromosome 11 (11q22.2-q22.3) that is linked with class III phenotype in a Hispanic cohort (Frazier-Bowers et al., 2009). To further investigate the region and find genes that might affect the incidence of class III malocclusion, Dr. Hartsfield and Dr. Lorri Ann Morford at the University of Kentucky have selected and genotyped 4 single nucleotide polymorphisms (SNPs; rs666723, rs578169, rs1386719 and rs12416856) within the 11q22.2-22.3 region on two multi-generational family-based cohorts from Brazil and Colombia for multipoint linkage analysis. The families in each cohort had a high prevalence of class III malocclusion; and varied greatly in the size, structure, and number of affected individuals. Class III affected and unaffected individuals were diagnosed based on cephalometric measurements, models, photographs and/or oral

examination. Maximum maximized LOD score **(***MMLS*) and multipoint heterogeneity LOD scores (*HLODs*) maximized over different levels of heterogeneity, and two genetic models (reduced penetrance dominant and recessive), were generated using SimWalk2. To estimate the empirical significance of these multipoint *HLODs*, 1000 replicates of unlinked genotype data based on real data pedigree structures, affection status and pattern of missing genotypes were simulated for the Brazilian and Colombian cohort using SLINK and SIMULATE respectively. These replicates were then analyzed using SimWalk2 with the original maximizing mode of inheritance. Power was estimated similarly for each cohort by generating 1000 replicates of pedigree data linked to the SNP with the highest *HLOD*. The corresponding cohort-specific mode of inheritance was used for the power simulation genetic parameters. For the Brazilian cohort, the *MMLS* was observed for rs12416856 at 191.6 cM (*HLOD*=1.84), under a recessive mode of inheritance. The empirical significance for this *HLOD* was a *p*-value <0.001 and the empirical type 1 error threshold for α=0.05, was an *HLOD* equal to 1.6. The power for suggestive linkage (*HLOD*≥2) was 80%.

For the Colombian cohort, the maximum MMLS was observed for rs578169 at 188.4 cM (*HLOD*=0.51), under a recessive mode of inheritance. The empirical significance for this *HLOD* was a *p*-value of 0.023 and the empirical **type 1 error** threshold for α=0.05, was an *HLOD* equal to 1.5. These results support potential linkage on chromosome 11.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude and appreciation to everybody who have made completing my thesis possible. First, my overwhelming thanks goes to Dr. Manika Govil. Dr. Govil has been more than a mentor to me. She has been with me in every step of the way in completing my Master's thesis. She gave me the courage and the support all the way during this journey. Thanks for giving me the chance to learn a totally new area for me and answering all my questions every time I encountered a problem. In my heart, your encouraging words "what does not kill you makes you stronger" will always remain. Although I have finally triumphed in my thesis, I will apply your principles in other areas of life as well.

I would like to thank Ms. Nandita Mukhopadhyay for her invaluable support in the lab. Learning programing language was like learning a new alphabet for me, yet she made it fun. She took me like a sister and guided me step by step during the last two years of my studies. Her support made learning and completing my thesis not only a dream come true but in the best environment possible.

I take this opportunity to pass my warmest regards to our collaborators from University of Kentucky Dr. James Hartsfield and Dr. Lorri Morford. Thanks for being there whenever I needed the help from your side.

I would like to acknowledge the support of Dr. Mark Mooney and Dr. Mary Marazita for giving me a thoughtful academic advice as committee members for the thesis.

The academic journey of studying abroad has not been simple. Completing a Master's program in a country away from home was overwhelmingly difficult but worthwhile. Taking care of my daughter and academics was not an easy ride. However, the entire university fraternity gave me the help I needed to do it all. Perhaps the lowest point in my life abroad came when I experienced a health problem in second year of my Masters. A health problem, being away from home, a daughter to take care of, and a thesis ahead of me almost overwhelmed me to a point of giving up. The challenges of studying abroad have taught me to fight for my goals. I have finally succeeded not because I did not make mistakes but because I failed a thousand times and a thousand times, I rose up stronger and better.

Finally, I would like to pass special thanks to my dear husband Omar, the source of my strength and energy. Without his support, this thesis would never have started, much less finished. Thanks to my family for their unending love and support during my studies. Ultimately, I give all the glory and honor to God for giving me this golden opportunity and most importantly, giving me the strength and courage to face all the challenges.

1.0 INTRODUCTION

1.1 CLASS III MALOCCLUSION

The malocclusion of the teeth is defined as the misalignment between the teeth of the two dental arches, and it is considered to be one of the most common problems encountered in the field of dentistry. Severe malocclusion has a huge impact on the quality of life if left untreated (Samsonyanova and Broukal, 2014) that may vary from social impacts to serious health impacts such as sleep apnea, muscle trismus, and digestive difficulties (Joshi et al., 2014).

In 1890, Edward H. Angle published the first classification of malocclusion based on the position of the permanent mandibular first molar buccal groove relative to the mesiobuccal cusp of the permanent maxillary first molar (Figure 1). In class III malocclusion, the permanent mandibular first molar buccal groove is located anterior to the mesiobuccal cusp of the permanent maxillary first molar. Thus, class III malocclusion as defined by Angle may also reflect the forward positioning of the mandible relative to the maxilla, although Angle only defined it in terms of the sagittal relationship of the permanent first molars to each other. The purist may take umbrage at the term “skeletal” Angle Class III malocclusion, since Angle only based it on the dentition, and not any possible underlying skeletal relationship. The purists would prefer references to the skeletal relationship of the mandible anterior to the maxilla in the sagittal

plane or relative to the cranial base such as mandibular prognathism, or maxillary retrusion or hypoplasia. Still, it has generally been recognized that this Class III dental relationship is often observed with a corresponding skeletal relationship as well. This skeletal deformity varies in its degree of severity and can be affected by both environmental and genetic factors (Otero et al., 2014).

Angle's classification of malocclusion is the most common method used, although it has limitations because it does not define the skeletal bases of malocclusion Skeletally, class III malocclusion can be classified into 3 types according to the position of the maxilla in relation to the mandible. True mandibular prognathism, which characterized by overgrown mandible and normal maxilla. Class III malocclusion can also results from excess growth of both mandible and maxilla or an undergrowth maxilla "maxillary hypoplasia" . Treatment modalities are vary among each category (Park and Baik, 2001).

The prevalence rate of the class III malocclusion varies among different ethnic groups. For example, the prevalence in the United States Caucasians is between 3.4-9.5%, 22.9-40% in Asians, and 16% in the Africans (Hartsfield et al., 2013). Although class III cases are found less frequently in orthodontic clinics as compared with the other types of malocclusions, they are most often cited as inherited traits (Frazier-Bowers et al., 2009). The genetic contribution to class III malocclusion has been reported in both animal and human studies (Dohmoto et al., 2002). Many studies have shown a genetic influence on class III traits using different modes of inheritance that vary from monogenic to multiple gene influence. The reported modes of inheritance include autosomal recessive, autosomal dominant, and autosomal dominant showing

incomplete penetrance (Cruz et al., 2008; El-Gheriani et al., 2003; Wolff et al., 1993). In different ethnic groups, these inheritance patterns may vary in both their penetrance and expressivity.

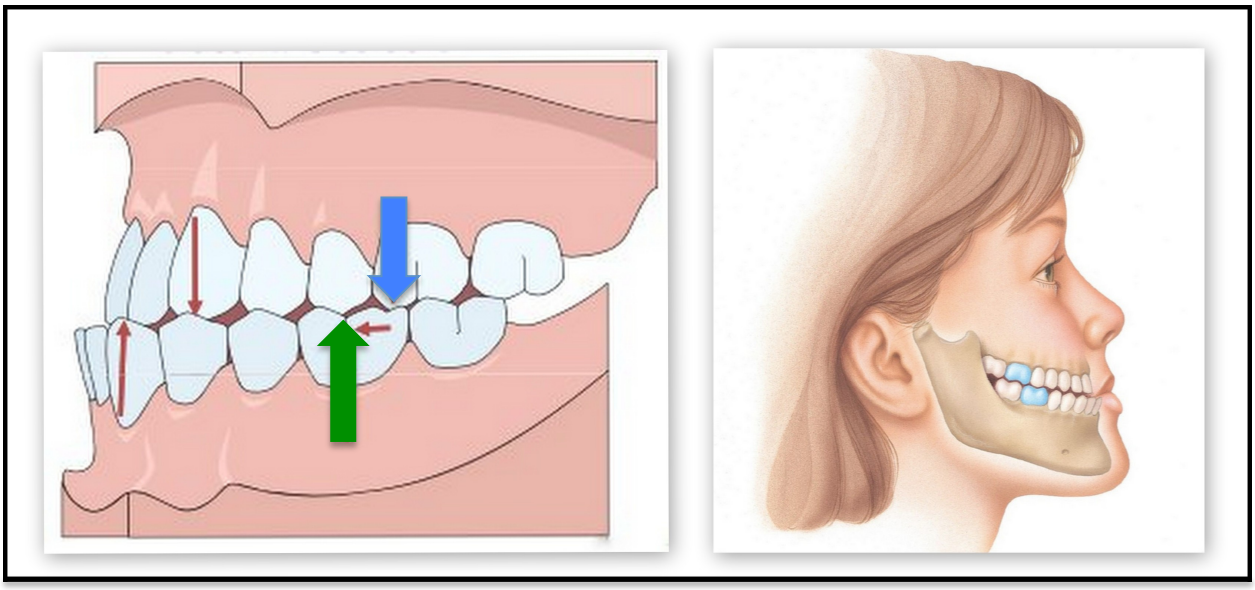

Figure 1. -Class III Malocclusion- Maxillary permanent first molar mesial cusp "blue arrow" is set posteriorly in relation to the mandibular permanent first molar buccal groove "green arrow". The lateral view shows a forward positioning of the mandible in relation to the maxilla in sagittal plane. *Reproduced with Permission from (Kasia, 2015).*

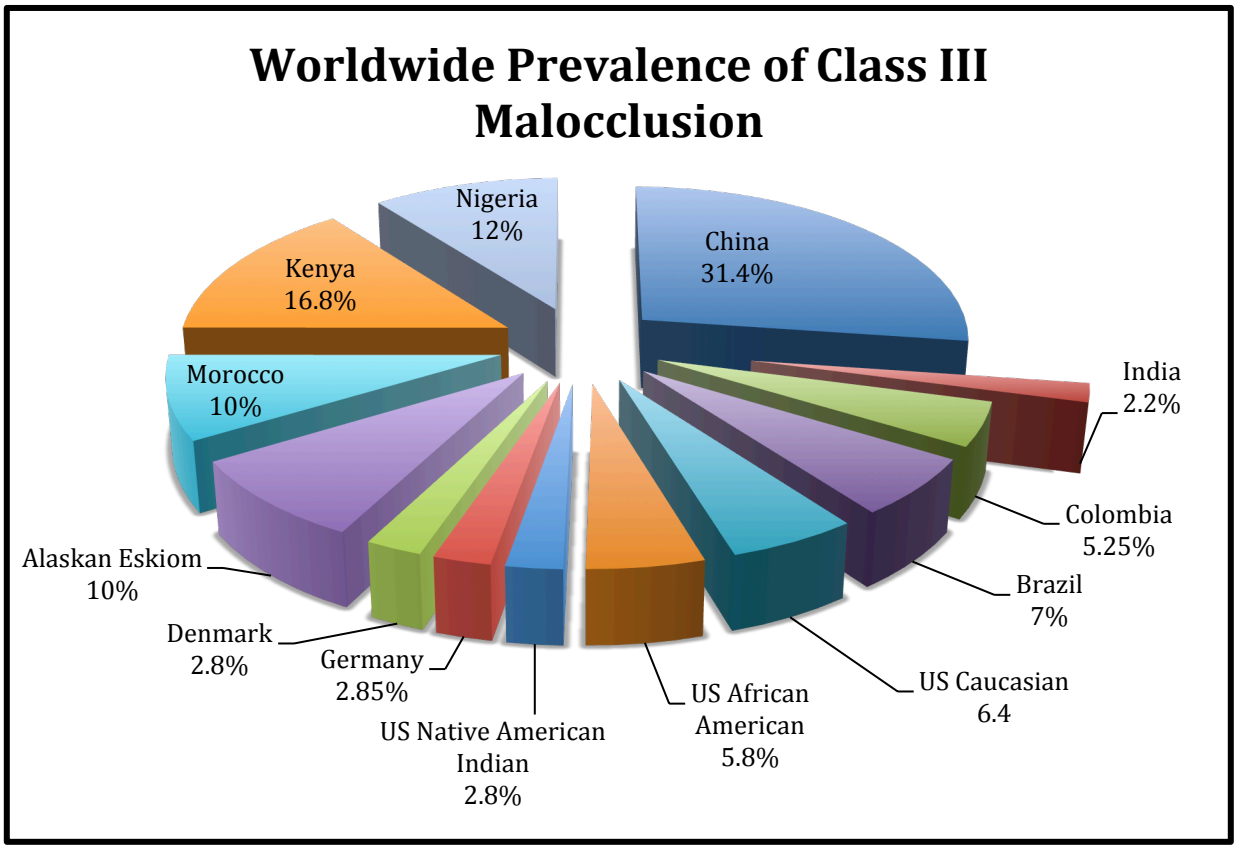

Figure 2. -Worldwide Prevalence of Class III Malocclusion based on geographic distribution and ethnic background- Eskimos, African and Asians populations have a high prevalence of class III phenotype. Low prevalence observed among Native American, Indians and Caucasians. *Data From (Hartsfield et al., 2013).*

Different loci of interest with respect to class III malocclusion have been identified in several chromosomes including, but not limited to: chromosome 1,3, 12 and 19 (Hartsfield et al., 2013). Genetic studies of the Hispanic population have implicated a unique locus at human chromosome 11q22 (Frazier-Bowers et al., 2009). In mice, the quantitative trait loci studies have shown that chromosome 11 is correlated with the mandibular prognathism (Dohmoto et al., 2002).

Although class III phenotype shows a strong genetic contribution, the majority of the previous studies have focused on treatment modalities rather than genetic etiology. Identifying the genes that contribute to the development of class III malocclusion will provide an alternate beneficial diagnostic tool in the orthodontic clinic and support orthodontists when making decisions pertaining to early intervention in the treatment of class III malocclusion.

Based on Frazier-Bowers et al findings, Dr. Hartsfield and Dr. Morford at the University of Kentucky have selected a region on chromosome 11q22.2-q22.3 to determine the gene on chromosome 11 that might affect the incidence of class III malocclusion in 17 Brazilian families and 15 Colombian families with a high prevalence of the class III phenotype. We conducted a linkage analysis and a simulation study to evaluate the performance of the linkage results.

1.2 BACKGROUND ON LINKAGE

Genetic linkage analysis is a powerful tool that is used to determine causal genetic loci on a certain chromosome with the help of genetic markers whose locations are known *a priori* (Rao and Province, 2001). The basic idea of genetic linkage is that when genetic loci are located proximal to each other on a chromosome, they have a greater tendency to be inherited together during meiosis i.e. the cell division process that produces haploid gametes (Rao and Province, 2001). This result takes advantage of an important violation of Gregor Mendel's second law. In 1856, Mendel started series of experiments to study heredity. Based on his observation on pea plants, Mendel had developed the theory of what is called *particles of inheritance*. Mendel had not used the word "genes" yet because genes dose not get invented for much longer. The theory stated that for every trait, parent had two such particles where the offspring get one from each parent. The choice of which of these two alleles is transmit to the offspring is a random draw (Mendel, 1901).

Thomas Hunt Morgan first described linkage in 1910. At the Columbia University, he preformed breeding experiments with the fruit fly, Drosophila Melanogaster. (Morgan, 1910). Morgan's experiments showed that genes are physically located on chromosomes and account for specific phenotypes. Further, his findings have violated Mendel's second law of independent segregation by proving that genes on the same chromosome do not always independently assort. He suggested that genes located proximal to each other have a greater chance of being inherited together in the offspring. Likewise, when genes are far apart on the chromosome, the probability

of being inherited together will decrease. Thus, linkage depends on the physical distance between the genes: the closer the genes, the stronger the linkage.

Morgan's conclusions inspired his undergraduate research student, Alfred Henry Sturtevant, to construct the first known genetic map in 1911 (Sturtevant, 1913). Sturtevant constructed the genetic map by relating crossover frequency to distance between two genes. The farther apart two genes lie on a chromosome; the more likely they are to be separated by recombination. In a genetic map, the distance between two genes on a chromosome can be measured by the probability of a crossover occurring between these genes. Sturtevant laid the groundwork for genetic map construction in other species. Genetic map construction depends on the recombination fraction theta (θ), which is a measure of the distance between two genes on the same chromosome. Theta is defined " as the probability that the alleles at two loci will appear in new combinations not observed in the parental generation" (Rao and Province, 2001). When the two loci lie apart from each other, it is more likely they are going to recombine during meiosis. In this case we say the two loci are unlinked and the expected recombination fraction is 0.5. Furthermore, when the two loci are close to one another on the same chromosome, the original combination of he parental alleles are more likely to be inherited together. In this case we say these two loci are linked ($\theta<1/2$). LOD score analysis is used to tests this recombination fraction.

In 1955, Newton E. Morton developed a statistical method for testing linkage. The *LOD* [logarithm (base 10) of odds of linkage] score compares the likelihood of obtaining the observed data, if the two given loci are linked, to the likelihood of observing the same data purely by chance (Rao and Province, 2001). The test procedure in his *LOD* score method was sequential;

families were added successively to the sample until it is possible to decide whether the two loci are linked or unlinked (Morton, 1955). Morton specified a threshold of 3 to conclude linkage and threshold of -2 to exclude or conclude no linkage. His justification of theses specific threshold was based on "the theory of sequential test" developed by (Wald 1947). Morton showed that when " $A = 1-\beta / \alpha$ " and " $B=\beta /1-\alpha$ ", where A and B are the critical values, and α and β represent type 1 and type 2 error, respectively; one should continue adding families as long as the *LOD* score value is falling between A and B. When the LOD score is larger than A, linkage is declared or concluded and if it its fall below B linkage is excluded. Morton has set a low level of type 1 error $\alpha = 0.001$ to keep the incorrect rejection of no linkage at a low level. When $\alpha = 0.001$ and $1-\beta$ (power) = 0.99; thus, log (A) = 3 and log (B) = -2 which are the thresholds specified by Morton. *LOD* score of 3 threshold is applied only for simple Mandelian inheritance. Nowadays, researchers are looking at multi-factorial traits where the traditional *LOD* score method no longer applies in linkage analysis.

As mentioned above, LOD score method is a useful tool for detecting linkage when the mode of inheritance of a disease or trait is known and it has been successfully used in the past for mapping single gene disorders. Clerget-Darpoux et al. (1986) showed that to use LOD score analysis to detect linkage in complex traits, LOD score function must be maximized to include other trait model parameters beside the recombination fraction θ. These parameters include trait gene frequency, trait penetrance (the probability of the trait or phenotype being expressed by individuals having the genotype corresponding to the trait), and trait heterogeneity (Clerget-Darpoux et al., 1986).

1.3 NEED FOR EMPIRICAL THRESHOLDS IN CURRENT DAY LINKAGE ANALYSIS

The following assumptions were proposed in the threshold specified by Morton. First, one family at a time is sequentially added to the test until *LOD* score below or above the threshold is achieved. Second, all the *LOD* score are calculated for a single fixed theta, thus, the analysis in this case is a two- point not a multipoint analysis. Third, only one single genetic model under the condition of homogeneity can be tested. Thus, the threshold proposed by Morton cannot be applied on current day linkage.

Linkage analysis nowadays can utilize a complete data set and maximize the likelihood over the space of theta θ. It is now possible to preform multipoint linkage analysis where multiple markers can be taken into account simultaneously. Moreover, current linkage analysis accounts for heterogeneity by including the heterogeneity parameter α in the likelihood i.e. *HLOD* method developed by (Smith, 1963). Finally, it takes care of unknown or uncertain prevalence and mode of inheritance by maximizing over several genetic model parameters e.g. the *MMLS* or maximized maximum LOD score approach (Greenberg, 1989) and the *MOD* or maximized *LOD* score (Clerget-Darpoux et al., 1986).

1.4 LINKAGE ANALYSIS APPROACHES

Statistical methods have been developed to test linkage in genetically heterogeneous complex traits such as class III malocclusion. The goal of these methods is to gather all available information from the pedigrees to provide linkage evidence between a trait and marker in a known location on a chromosome. Testing linkage one marker at a time to the trait is called two-point linkage mapping, whereas testing linkage with two or more markers to the trait is called multipoint linkage mapping. Further, methods for linkage analysis can also be divided into two categories: parametric and nonparametric.

Parametric linkage analysis, by definition, takes into consideration the trait parameters and requires the specification of a genetic model including disease allele frequency and penetrance. Because the true trait parameters are not known in complex traits, several parametric methods have been developed (Figure 3). One of these approaches is the maximum-likelihood or *LOD*-score (Z) where the likelihood is maximized over θ while fixing the other trait parameters (Ott, 1999). Specification of the mode of inheritance is required and considered a limitation in this method. Another limitation is that genetic heterogeneity is not taken into account. To accommodate genetic heterogeneity, the heterogeneity *LOD* or the *HLOD* (Smith, 1963) takes into consideration the admixture parameter α and it is maximized to estimate the other genetic parameters. When *LOD* score method is based on a wrong genetic model the power to detect linkage tend to reduce (Clerget-Darpoux et al., 1986). To overcome this limitation investigators have developed what it is called the maximum maximized *LOD* score or the *MMLS* (Greenberg, 1989). *MMLS* approach maximizes the *HLOD* over two genetic models (dominant and recessive)

and use the highest statistical result as a test for linkage. It has been shown that when the maximization considers an average mode of inheritance i.e. dominant and recessive models with reduced penetrance, the power to detect linkage is not affected (Greenberg, 1989). The power to detect linkage in this case is almost the same as what one would expect when the true mode of inheritance is known. Moreover, the effect of multiple testing preformed in the *MMLS* on type 1 error is minimum and not of a concern (Greenberg et al., 1998).

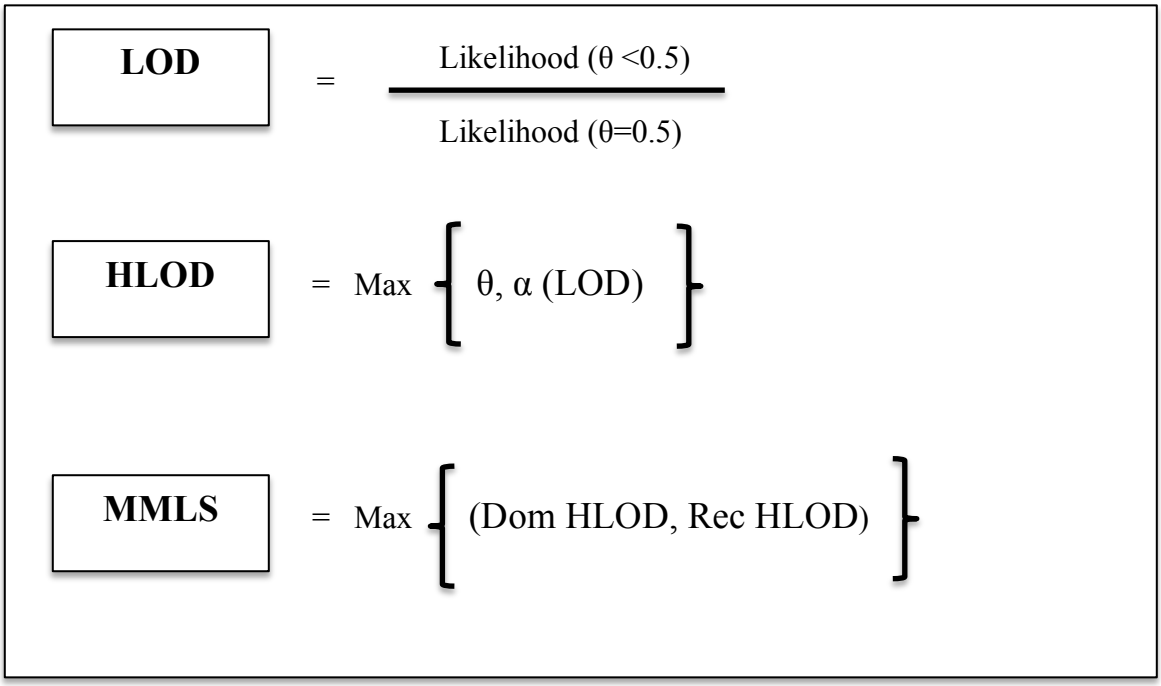

Figure 3. Different Statistical Approach in Testing Linkage

LOD score method takes the recombination fracture θ into consideration. Heterogeneity *LOD* score approach includes the recombination fracture θ and the admixture parameter α in the likelihood. *MMLS* approach maximized the likelihood over incomplete dominant and recessive genetic models, respectively

Nonparametric linkage methods however, do not need the specification of genetic model parameters. One class of non-parametric linkage methods, namely the affected-relative pair class of methods detects linkage by analyzing the degree of excessive allele sharing by descent in affected relatives (Goheen et al., 1992; Weeks and Harby, 1995). The basic idea of the affected-relative pair approach is that when a marker is linked to a disease locus, the two affected siblings will share the same marker allele more often than expected by chance (Rao and Province, 2001).

2.0 MATERIALS AND METHODS

2.1 DATA

Subjects consist of two multi-generational family-based cohorts located in Brazil and Colombia with a high prevalence of class III malocclusion. Based on previous data implicating a region of chromosome 11 in the Class III phenotype, 4 single nucleotide polymorphisms (SNPs; rs666723, rs578169, rs1386719 and rs12416856) were genotyped within 17 Brazilian families (178 individuals) and 15 Colombian families (248 individuals). The families in each cohort varied greatly in the size, structure and the number of affected individuals.

2.2 ASCERTAINMENT OF FAMILIES

The following universities were involved in this collaborative study: University of Kentucky, Lexington, KY, U.S.A. where the subjects were genotyped, Pontifica Universidad Javeriana, Bogota, Columbia, South America and Universidade de Brasilia, Brasilia, Brazil, South America where the subjects were recruited. The statistical analysis preformed at the University of Pittsburgh, Pittsburgh, Pennsylvania, U.S.A.

Study participation required that the subjects allowed access to their personal and family dental/facial information, which was used to diagnose the members of the family that exhibited the class III phenotype, and provide a sample of genomic DNA. Patients and relatives aged 6–70 years were evaluated to determine their dental/facial phenotypes. Subjects who exhibited syndromes and pathologies such as cleft/lip palate or general physical diseases were excluded from the study. The affected status of subjects who had undergone surgery to correct their Class III malocclusion was confirmed using dental charts. A complete pedigree was constructed for each family that included all available subjects.

2.3 GENOTYPING

After obtaining informed consent, the DNA samples were collected either by peripheral blood via venous puncture, buccal swab collection from the inner cheek, or saliva collection in association with either Pontificia Universidad Javeriana or Universidade de Brasília. Subjects were excluded if they were unable or unwilling to supply personal and familial data related to the study. Genomic DNA was extracted from the buffy coat obtained after the centrifugation of peripheral blood using a commercial GFX™ Genomic Blood DNA Purification Kit (GE Healthcare Life Sciences). Genomic DNA was isolated from buccal swabs using the Puregene method (Gentra Systems, Minneapolis, MN). Saliva (2–4 mL) was collected using Oragene®-DNA self-collection vials (DNA Genotek, Ottawa, Ontario, Canada) and genomic DNA was isolated by ethanol precipitation according to the manufacturer's instructions. The DNA was resuspended in 10 mM Tris–HCl with 1 mM EDTA (pH 8.0; Fisher Scientific, Fair Lawn, NJ,

USA). The DNA concentrations were determined using a NanoDrop-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

The purified DNA was subjected to single nucleotide polymorphism (SNP) analysis using Taqman genotyping assay kits and reagents (Applied Biosystems) with a Roche LightCycler 480® system (Roche Applied Science, Indianapolis, IN, USA).

2.4 TRAIT CLINICAL EVALUATION

The Colombian population was diagnosed with class III malocclusion based on lateral cephalometric radiographs, dental casts, and/ or facial and intraoral photographs. A negative ANB angle on the lateral cephalometric radiograph was required for the diagnosis of class III malocclusion. For Brazilian population, examining dental casts, and/or facial and intraoral photographs from orthodontic records determined class III malocclusion. When no orthodontic records were available, Class III malocclusion in the relatives was determined by the presence of an edge-to edge incisor relationship, anterior crossbite, concave facial profile and visual inspection. The control group included unaffected members of the same family and unrelated individuals without Class III malocclusion.

2.5 STATISTICAL ANALYSIS OF REAL DATA

2.5.1 Data Cleaning

We performed thorough cleaning of genotype to resolve Mendelian errors. Mendelian error checking within the genotyping of parent-offspring trios was conducted with PLINK (version 1.07) and Pedcheck (version 1.2) software (Purcell et al., 2007). Mendel (version 13.2) software (Lange et al., 2013) was used to estimate maximum likelihood marker allele frequencies separately in the 2 cohorts (Table.2). Since several studies have showed that an accurate genetic map has a strong impact on the power of linkage analysis (He et al., 2011), a population specific map was constructed based on reference genetic maps obtained from Rutgers using Kelvin software (Vieland et al., 2011), to account for population differences with respect to observed recombination frequencies (map construction and acquisition protocol was obtained from private communication with Dr. Veronica Vieland).

2.5.2 Linkage Analysis

Nonparametric linkage analysis and parametric multipoint linkage analysis were both preformed for the families. MERLIN (version 1.1.2) (Abecasis et al., 2002) was used for the nonparametric , using an allele-sharing and affected-only model to test for linkage. Both NPL-all and pairs statistics were used to test for excessive allele sharing among affected individuals. *LOD* scores were calculated using Kong and Cox linear model (Kong and Cox, 1997).

SIMWALK2 (version 2.91) software (Sobel and Lange, 1996) was used for multipoint linkage analysis based on either dominant or recessive mode of inheritance with grid of heterogeneity parameter α ranging from 0-1 in a steps of 0.5. In dominant model, the disease allele frequency was set on 0.03 and penetrance of 0.002, 0.5, and 0.5. For the recessive model the disease allele frequency was set on 0.27 and penetrance of 0.002, 0.002, and 0.5. Mega2 (version 4.6, (Mukhopadhyay et al., 2005) was used to prepare the formatted input files needed for Pedcheck, Merlin, Mendel and SimWalk2.

2.6 SIMULATION: SIGNIFICANCE, TYPE 1 ERROR AND POWER CALCULATIONS OF PARAMETRIC LINKAGE

In order to estimate the **significance** of the parametric linkage results observed from the real data, we conducted a simulation study. **Type 1 error** thresholds were estimated for each cohort by simulation of 1000 replicates of the genotype data based on real data pedigree size, structure, affection status and pattern of missing genotypes. Replicates were simulated unconditioned on disease status (assuming no linkage between the trait and the marker) using SLINK and SIMULATE software (Ott, 1989; Weeks D.E., 1990). *HLOD* thresholds were calculated for a type 1 error of 0.05, 0.01, and 0.001 for each population. This was done to see the probability of obtaining a linkage signal similar or better than the signal observed from the real data without simulation.

Power was calculated conditionally on the disease status by generating 1000 replicates. Recessive model was specified based on the maximum *HLOD* obtained from the multipoint

parametric linkage analysis. As mentioned above, recessive model the disease allele frequency was set on 0.27 and penetrance of 0.002, 0.002, and 0.5. *HLOD* thresholds were calculated for power of 90%, 80%, and 70% for Brazilian cohort.

2.7 FOLLOW UP ON CANDIDATE GENES IN SPECIFIED REGION

UCSC genome browser was used in order to explore the candidate genes (Kent et al., 2002). Region 11q22.2-22.3 was specified in the "search term" box. The UCSC Genes track was chosen as a display mode. This displays genes prediction based on GenBank and RefSeq. I used OMIM, GENECARDS and UCSC sites to check the functionality of each gene and to search for any potential effect on the growth of the mandible.

3.0 RESULTS

3.1 DATA DESCRIPTION

From the Brazilian cohort, 17 out of 18 families were informative for linkage with a total of 178 individuals documented in multi-generational pedigrees for these families and 176 of these individuals genotyped, 85 individuals were diagnosed with the Class III phenotype. From the Colombian cohort, 18 families out of 23 were informative for linkage with a total of 276 individuals documented in multi-generational pedigrees, 153 of these individuals genotyped and 112 individuals diagnosed with the Class III phenotype (Figure.4).

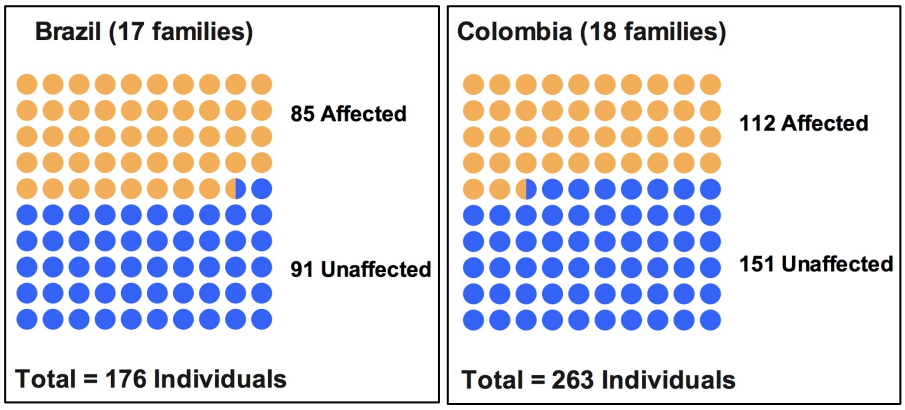

Figure - 4 - Pedigree and Genotype Data for Colombian and Brazilian Cohorts

3.2 DATA CLEANING

The minor allele frequencies for both Brazilian and Colombian cohorts examined in this study for each SNP are shown in (Table.1). For population specific map construction, we ended up using the reference map because of an insufficient number of crossovers detected due to marker proximity.

Table.1 SNP Position and Minor Allele Frequency for Colombian and Brazilian Cohorts

SNP	Position (kb)	Brazil Minor Allele Frequency		Colombia Minor Allele Frequency	
rs666723	104,827	A	0.29	A	0.43
rs578169	104,862	A	0.35	A	0.40
rs1386719	109,559	G	0.28	G	0.27
rs12416856	109,593	A	0.41	A	0.30

3.3 NONPARAMETRIC LINKAGE RESULTS

Linkage was not observed using the nonparametric analysis in both Colombian and Brazilian cohorts.

Table.2 Nonparametric results for SNPs; rs666723, rs578169, rs1386719 and rs12416856 for Brazilian and Colombian population

	Brazil			Colombia		
Marker	***p*-value**	**Position**	**Position (kb)**	***p*-value**	**Position**	**Position (kb)**
rs666723	0.125	104,826	104,827	0.21	104,826,604	104,827
rs578169	0.124	104,862	104,862	0.24	104,861,643	104,862
rs1386719	0.116	109,559	109,559	0.13	109,559,457	109,559
rs12416856	0.117	109,593	109,593	0.14	109,593,305	109,593

3.4 PARAMETRIC LINKAGE RESULTS

The Brazilian cohort was analyzed separate from the Colombian cohort due to potential differences in ethnic composition within each population.

Observed *MMLS* results: For Brazilian cohort, a **peak *MMLS* of 1.84 (empirical P-value < 0.001)** was observed for rs12416856 at 191.6 cM (Figure.5). , The maximizing model is a recessive mode of inheritance.

Simulation results: The empirical **type 1 error** threshold for α=0.05 is *HLOD* equal to 1.6 (Table.3). Other critical empirical threshold values are presented in table 3. Figure 6 shows the HLOD score distribution for simulated linked data for the Brazil pedigrees. The expected **power of this set of pedigrees for an** *HLOD*=2 is ≥80% (Figure. 6)

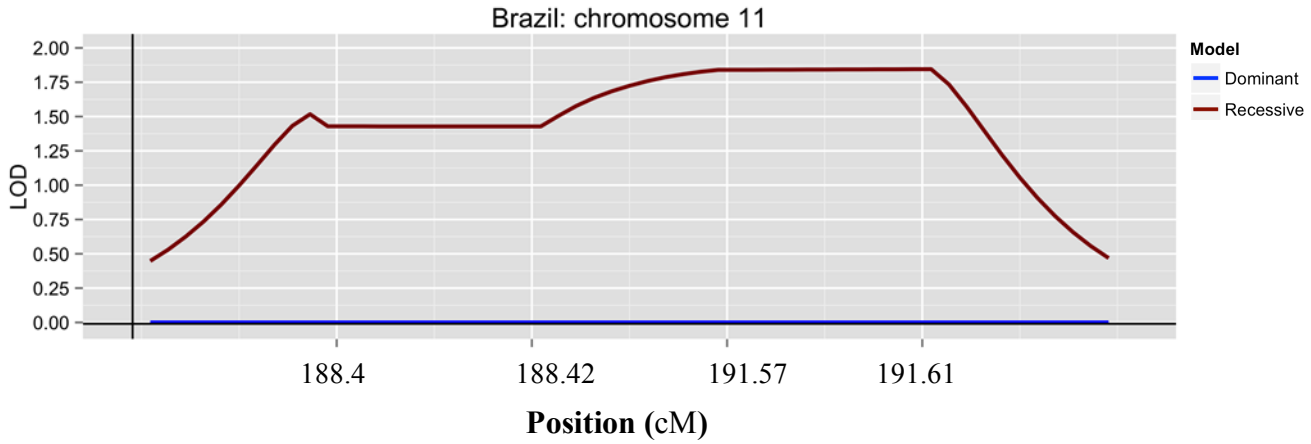

Figure 5. -Brazilian Population Parametric Linkage Analysis Results For Dominant and Recessive Models- *LOD* scores shown at SNPs; max *HLOD*=1.84 (SNP rs12416856) at 191.6 cM under a recessive model.

Table.3 -Brazilian Cohort Type 1 Error Results- empirically derived *HLOD* threshold values over 1000 replicates

Type 1 Error	*HLOD*
0.05	0.5
0.01	1.1
0.001	1.6

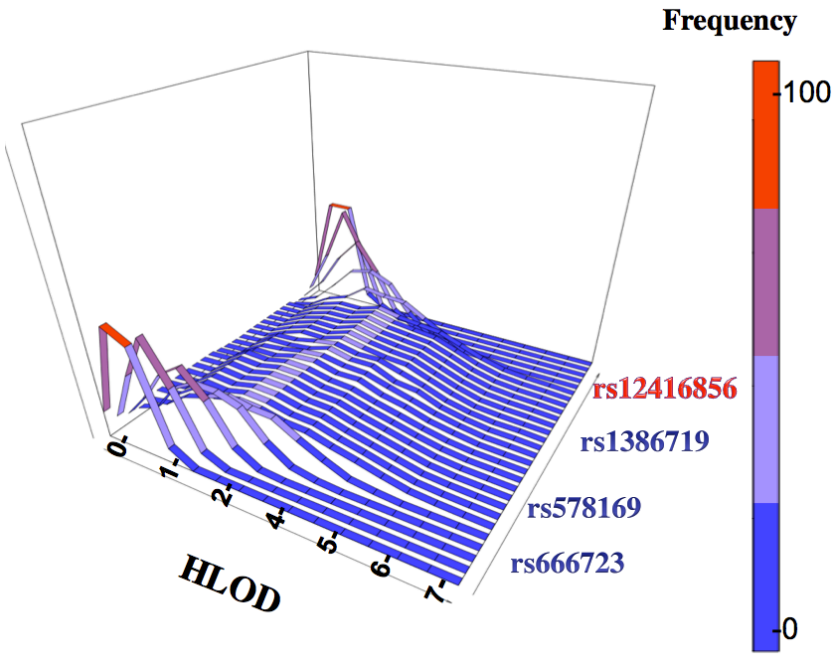

Figure 6. 3D Surface Plots of Brazilian Cohort Power Results- At rs12416856 for *HLOD* ≥ 1.5, 2 and 3 the power was 91.6%, 83% and 68%, respectively.

Observed *MMLS* results: For Colombian cohort, a **peak *MMLS* of 0.51 (empirical P-value = 0.023)** was observed for rs578169 at 188.4 cM (Figure.7). The maximizing model is a recessive mode of inheritance.

Simulation results: The empirical **type 1 error** threshold for α=0.05 is *HLOD* equal to 1.5 (Table.4). Other critical empirical threshold values are presented in table 4. Due to computational limitation, Colombian cohort **power** calculations have been delayed.

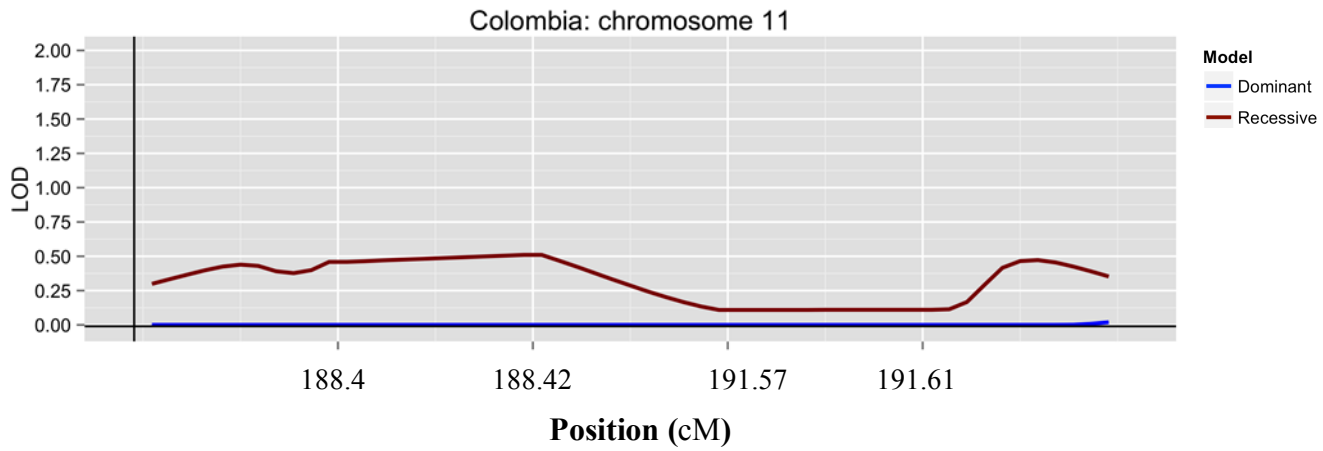

Figure 7. Colombian Population Parametric Linkage Analysis Results For Dominant and Recessive Models *LOD* scores are shown for (SNPs; rs666723, rs578169, rs1386719 and rs12416856). The max *HLOD*=0.51 was found on rs578169 at 188.4 cM under a recessive model.

Table.4 – Colombian Cohort Type 1 Error Results- *HLOD* were empirically derived over 1000 replicates.

Type 1 Error	*HLOD*
0.05	0.3
0.01	0.84
0.001	1.5

3.5 CANDIDATE GENES IN LINKAGE REGION

We found potential candidate genes on chromosome 11q22.2-22.3 using UCSC genome browser (Figure .8). Matrix metalloproteinase MMP-1 and MMP-8 are believed to have a role in bone growth and metabolism. They are also involved in extracellular matrix breakdown, which is a very important mechanism in embryonic development and tissue remodeling.

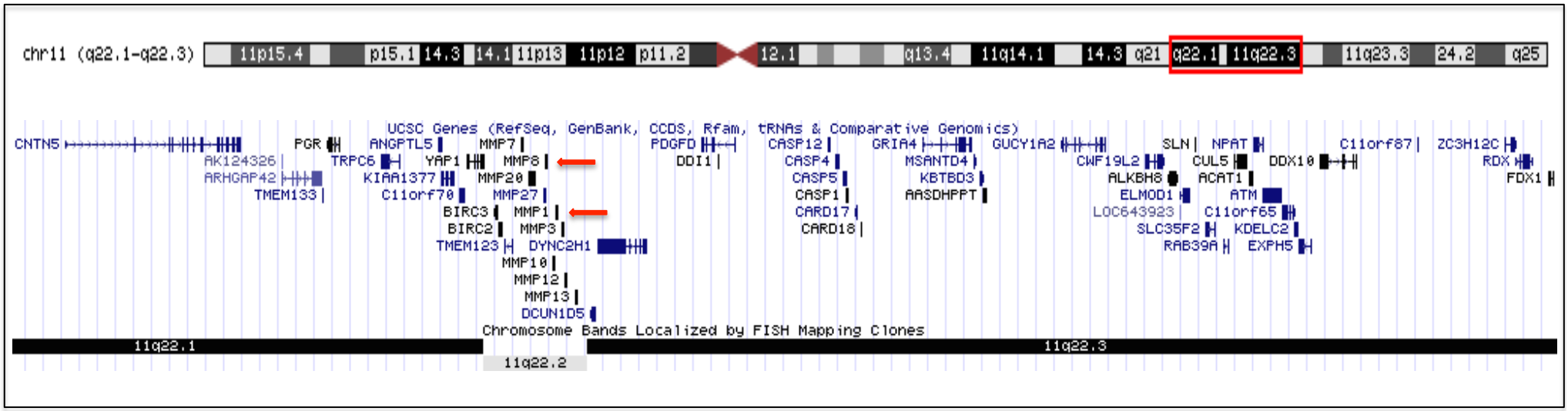

Figure 8. -The Potential Candidate Genes on Chromosome 11q22.2-22.3 as shown from UCSC genome browser – MMP1 and MMP8 were found within the region q22.2-22.3.

4.0 DISCUSSION

Despite the fact that class III malocclusion have been reported as the most inheritable type of malocclusion, most of the investigations have concentrated on class III phenotype treatment modalities and outcomes while little is being done to understand the genetic aspect of the trait. Many human and animal studies have shown the evidence of genetic contribution of developing class III phenotype.

Previous genetic studies of class III in Hispanic population reported area of interest on chromosome11. Microsatellites, D11s1886 and D11s4204, have been implicated in linkage to Class III (Frazier-Bowers et al., 2009). Therefore, the focus here was to study 11q22.2-q22.3 chromosomal region and genotyped single nucleotide polymorphisms (SNPs) close to D11s4204 and D11s1886 (SNPs: rs1386719, rs578169, rs666723 and rs12416856).

The results from our parametric linkage analysis for both Brazilian and Colombian cohorts support Frazier-Bowers et al. 2009 findings. Potential genetic heterogeneity was taken into consideration through every step in the analysis. The heterogeneity of class III phenotype came from different factors. First, The skeletal aspect of class III varies from only mandibular prognathism, maxillary hypoplasia or a combination of both. Second, in this study, the Brazilian

cohort was diagnosed based on dental history evaluation using dental charts, casts, and intraoral photographs. In contrast the Colombian cohort was diagnosed based on skeletal evaluation using cephalometric radiographs. Third, it has been shown that class III malocclusion prevalence and the severity of phenotype may vary according to sex and ethnicity (Hartsfield et al., 2013). Class III malocclusion also exhibits genetic heterogeneity with evidence of multiple modes of inheritance. These include autosomal recessive, autosomal dominant, and autosomal dominant with incomplete penetrance (Cruz et al., 2008; El-Gheriani et al., 2003; Wolff et al., 1993). In addition, South America is one of the most diverse regions in the world since its inhabitants came from different population ancestries, ethnic groups and races. All these factors account for difficulties during the investigation and the analysis of the trait. In order to minimize the effect of theses factors several steps have been done before and during the analysis. Population specific marker allele frequencies were estimated. Also, population specific genetic maps were constructed based on the data available from the pedigrees. To account for genetic heterogeneity of the trait, the *MMLS* was calculated by analyzing the data under two genetic models, dominant and recessive, with incomplete penetrance (Greenberg et al., 1998). Greenberg demonstrated that this approach has strong power to detect linkage for complex traits..

The MMLS results show an empirically significant linkage peaks max *HLOD* of 1.84 and 0.51 for both Brazilian and Colombian cohorts respectively. The results from the two populations support each other since both maximized under a recessive model. Further, the two linkage peaks are 3 cM apart (~3 X 10^6 bp) and the Brazilian region of interest includes the Colombian peak region. Given a type 1 error *HLOD* threshold of 1.6 (corresponding to α=0.05),

the simulation showed ≥ *90% power* for the Brazilian cohort, which indicates enough sample size to detect linkage.

Due to the presence of inbreeding loops in the Colombian cohort pedigrees, linked data could not be simulated to calculate power. This issue could have been addressed either by dropping or ignoring the multiple inbreeding or marriage loops so the likelihood estimation can be computed, however, this will result in loosing a lot of genetic information and therefore reducing the power to detect linkage.

We found potential candidate genes on chromosome 11 using UCSC genome browser. *MMP-1* and *MMP-8* have been shown to influence mandibular condyle growth in animal models (Patil et al., 2012). Proteins of *MMP* family are involved in the break down of the extra cellular matrix. This is an essential step during many physiological processes in human body like tissue remodeling and embryonic development. The protein encoded by *MMP* family cleaves type II collagen and plays a role in articular cartilage turnover (OMIM, 2015). A gene expression study has showed a 2.33-fold up regulation of MMP-1 in condylar cartilage of New Zealand rabbits when a mandibular anterior device was applied (Patil et al., 2012). Mandibular anterior device is used as an early intervention for undergrowth mandible in children to enhance the mandibular condyle growth, which result in forward positing of the mandible. Another study showed a high expression of *MMP-8* in condylar cartilage of juvenile pigs after the application of anterior mandibular displacement device (Gredes et al., 2012).

5.0 CONCLUSIONS

Our results suggest evidence for linkage of class III malocclusion to **11q22.2q22.3**. Current linkage results are based on *MMLS* maximized over two genetic *mod*els. We suggest taking the entire genetic parameter space into consideration by calculating *MODs* for each population and each pedigree. This would help in further understanding the relation between each pedigree *MOD* distribution and phenotypic reality for these families.

Class III phenotype expresses itself with different skeletal patterns and severity, which highly affect the trait homogeneity. A strict definition of class III malocclusion is needed to enhance the trait homogeneity thus aid in understanding the genetic bases of class III malocclusion. After a clear definition of the trait, one might consider categorizing the cases based on severity and/or dental and skeletal features.

Theses results were specific for Hispanic cohorts. It would be interesting to explore the genes that influence the incidence of class III trait on the same region on chromosome 11 in other populations and ethnicities. Further, since many genes on different chromosomes have been associated with class III malocclusion, future research should consider studying the relation between different skeletal and dental pattern of class III trait and the genes associated with them.

BIBLIOGRAPHY

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics* 30(1)*:*97-101.

Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J (1986). Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* 42(2)*:*393-399.

Cruz RM, Krieger H, Ferreira R, Mah J, Hartsfield J, Jr., Oliveira S (2008). Major gene and multifactorial inheritance of mandibular prognathism. *American journal of medical genetics Part A* 146A(1)*:*71-77.

Dohmoto A, Shimizu K, Asada Y, Maeda T (2002). Quantitative trait loci on chromosomes 10 and 11 influencing mandible size of SMXA RI mouse strains. *Journal of dental research* 81(7)*:*501-504.

El-Gheriani AA, Maher BS, El-Gheriani AS, Sciote JJ, Abu-Shahba FA, Al-Azemi R *et al.* (2003). Segregation analysis of mandibular prognathism in Libya. *Journal of dental research* 82(7)*:*523-527.

Frazier-Bowers S, Rincon-Rodriguez R, Zhou J, Alexander K, Lange E (2009). Evidence of linkage in a Hispanic cohort with a Class III dentofacial phenotype. *Journal of dental research* 88(1)*:*56-60.

Goheen MP, Blumershine R, Bartlett MS, Hull MT, Smith JW (1992). Improved intracellular morphology of Pneumocystis carinii from rat lung by postfixation with a mixture of potassium ferrocyanide and osmium tetroxide. *Biotechnic & histochemistry : official publication of the Biological Stain Commission* 67(3)*:*140-148.

Gredes T, Mack H, Spassov A, Kunert-Keil C, Steele M, Proff P *et al.* (2012). Changes in condylar cartilage after anterior mandibular displacement in juvenile pigs. *Archives of oral biology* 57(6)*:*594-598.

Greenberg DA (1989). Inferring mode of inheritance by comparison of lod scores. *American journal of medical genetics* 34(4)*:*480-486.

Greenberg DA, Abreu P, Hodge SE (1998). The power to detect linkage in complex disease by means of simple LOD-score analyses. *American journal of human genetics* 63(3)*:*870-879.

Hartsfield JK, Morforda LA, Oterob LM, Fardoc DW (2013). Genetics and non-syndromic facial growth. *Journal of Pediatric Genetics* 2013(2)*:*1–12.

He C, Weeks DE, Buyske S, Abecasis GR, Stewart WC, Matise TC *et al.* (2011). Enhanced genetic maps from family-based disease studies: population-specific comparisons. *BMC medical genetics* 12(15.

Joshi N, Hamdan AM, Fakhouri WD (2014). Skeletal malocclusion: a developmental disorder with a life-long morbidity. *Journal of clinical medicine research* 6(6)*:*399-408.

Kasia (2015). Dentodontics Chronicles of a Dental Student (http://dentodontics.com).

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM *et al.* (2002). The human genome browser at UCSC. *Genome research* 12(6)*:*996-1006.

Kong A, Cox NJ (1997). Allele-sharing models: LOD scores and accurate linkage tests. *American journal of human genetics* 61(5)*:*1179-1188.

Lange K, Papp JC, Sinsheimer JS, Sripracha R, Zhou H, Sobel EM (2013). Mendel: the Swiss army knife of genetic analysis programs. *Bioinformatics* 29(12)*:*1568-1570.

Mendel G (1901). Versuche über Pflanzen-Hybriden. *J R Hortic Soc* 26(1–32.

Morgan TH (1910). Sex Limited Inheritance in Drosophila. *Science* 32(812)*:*120-122.

Morton NE (1955). Sequential tests for the detection of linkage. *American journal of human genetics* 7(3)*:*277-318.

Mukhopadhyay N, Almasy L, Schroeder M, Mulvihill WP, Weeks DE (2005). Mega2: data-handling for facilitating genetic linkage and association analyses. *Bioinformatics* 21(10)*:*2556-2557.

OMIM (2015). Online Mendelian Inheritance in Man. McKusick-Nathans Institute of Genetic Medicine: Johns Hopkins University (Baltimore, MD); http://omim.org/.

Otero LM, Morford LA, Falcão-Alencar G, Hartsfield Jr JK (2014). Chapter 1. Family History and Genetics of Mandibular Prognathism. In: Orthodontic Treatment of Class III Malocclusion eBook. PW Ngan, T Deguchi and EW Roberts editors: Bentham eBooks, pp. 3-24.

Ott J (1989). Computer-simulation methods in human linkage analysis. *Proceedings of the National Academy of Sciences of the United States of America* 86(11)*:*4175-4178.

Ott J (1999). Analysis of human genetic linkage. 3rd ed. Baltimore: Johns Hopkins University Press.

Park JU, Baik SH (2001). Classification of Angle Class III malocclusion and its treatment modalities. *The International journal of adult orthodontics and orthognathic surgery* 16(1)*:*19-29.

Patil A, Sable R, Kothari R (2012). Genetic expression of MMP-Matrix-mettalo-proteinases (MMP-1 and MMP-13) as a function of anterior mandibular repositioning appliance on the growth of mandibular condylar cartilage with and without administration of Insulin like growth factor (IGF-1) and Transforming growth factor-B (TGF-beta). *The Angle orthodontist* 82(6)*:*1053-1059.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 81(3)*:*559-575.

Rao DC, Province MA (2001). Genetic dissection of complex traits San Diego: Academic Press.

Samsonyanova L, Broukal Z (2014). A systematic review of individual motivational factors in orthodontic treatment: facial attractiveness as the main motivational factor in orthodontic treatment. *International journal of dentistry* 2014(938274.

Smith CA (1963). Testing for Heterogeneity of Recombination Fraction Values in Human Genetics. *Annals of human genetics* 27(175-182.

Sobel E, Lange K (1996). Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *American journal of human genetics* 58(6)*:*1323-1337.

Sturtevant AH (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *Journal of Experimental Zoology* 14(1)*:*43-59.

Vieland VJ, Huang Y, Seok SC, Burian J, Catalyurek U, O'Connell J *et al.* (2011). KELVIN: a software package for rigorous measurement of statistical evidence in human genetics. *Human heredity* 72(4)*:*276-288.

Weeks D.E. OJaLGM (1990). SLINK: a general simulation program for linkage analysis. *American journal of human genetics* 47:A204(204.

Weeks DE, Harby LD (1995). The affected-pedigree-member method: power to detect linkage. *Human heredity* 45(1)*:*13-24.

Wolff G, Wienker TF, Sander H (1993). On the genetics of mandibular prognathism: analysis of large European noble families. *Journal of medical genetics* 30(2)*:*112-116.