BRA 2016 Final Report

Type of Award: Biomedical Research Award
Title: Sequencing SNAI3 and TWIST1 to Identify Functional Variants Underlying Malocclusion Phenotypes
Name of Principal Investigator: Lina M. Moreno Uribe.
Institution: University of Iowa
Period of AAOF support: July 1st 2016- December 31, 2018.
Amount of AAOF Funding: $30,000.

Signature: **Date:** December 31, 2018

Summary/Abstract

Previously we found significant associations between a quantitative phenotype ranging from severe convexity or concavity of the skeletal profile as well as dento-alveolar relations ranging between positive and negative overjets (PC2) and various SNPs located within a ~12Kb interval surrounding *SNAI3*. Also, association between a quantitate phenotype depicting variation from a large to a small mandibular body (PC3) was also detected with a ~15Kb interval surrounding *TWIST1.* Within this project we have completed Sanger sequencing of the SNAI3 and TWIST1 associated intervals and respective association analyses with both common and rare variants. For SNAI3 sequencing was performed in 272/277 attempted (F=191, M=81). Of these 55 are skeletal Class I, 129 are skeletal Class II and 88 are skeletal Class III. The age range is 12-68 with a mean age of 30. For TWIST1 we sequenced 277 individuals (F=195, M=82). Of these 56 are Skeletal Class I, 131 are skeletal Class II and 90 are skeletal Class III. The age range is from 12-68 with a mean age of 30. Association results for SNAI3 common variants showed that rs4287555 and rs9931509 are the most highly associated and bioinformatics analyses indicated that these two variants are eQTLs for SNAI3-AS1, SNAI3 and RNF166. New data from SysFACE shows that RNF166 is highly express in maxilla, mandible and palate of both mice and human. Association results for SNAI3 rare variants returned non-significant results for changes in PC scores at the mean level, for any of the rare variant count categories regardless of PC evaluated. For the odds of falling into a specific PC threshold, significant results (P<0.05) were obtained with the odds of having a PC2 score in the 90th percentile or above and the presence of any rare variant vs. none (P=0.03). Amongst the 105 rare variants within the SNAI3 interval we identified 4 coding mutations. One of these Chr16:88747702; G166E is a new mutation not previously reported in EXACT or GNOMAD. The phenotypic description of individuals carrying the same rare coding variant does not seem to correlate suggesting that these rare variants may not be related to a specific skeletal phenotype. Association results for TWIST1 common variants retrieved the best association between PC3 and rs2189000 (P=0.0003) located 3.9 kb upstream of TWIST1. *In sílico* analysis predicted that SNP rs2189000 disrupts a transcription factor bindings site (TFBS) for PITX2. Association results for TWIST1 rare variants showed that for changes in PC scores at the mean level, linear regression models returned significant results (P<0.05) for PC3 with the variables total rare variant count (P=0.048), and the presence of 3 or more rare variants vs. cero (P=0.029). For the odds of falling into a specific PC threshold, significant results (P<0.05) were obtained with the odds of having a PC1 score in the 25th percentile or below and the presence of 1 or more rare variant vs. cero (P=0.015). Also for the odds of having a PC1 score in the 75th percentile or above and the variable total rare variant count (P=0.017) and 1 or more rare variants vs. cero (P=0.011). Moreover, significant results were also obtained for the odds of having a PC3 score in the 25th percentile or below with the variable total rare variant count (P=0.03). Quantile regression returned significant results (P<0.05) also for PC1 and PC3. For PC1 significant results were seen for PC1Q75 with total rare variant count (P=0.047), for PC1Q90 with the variable any rare variant (P=0.041), for PC1Q10 with 1 or more (P=0.005) and 2 or more (P=0.001) rare variants. Finally for PC3 significant quantile regression results were seen at PC3Q10 (P=0.003) and PC3Q25 (P=0.005) with 3 or more rare variants. For skeletal Class risk, significant results were seen for Class II vs all others (Class I and III combined) and the presence of 2 or more rare variants (P=0.05). Also a significant result was seen for Class II vs. all others and the presence of 3 or more rare variants (P=0.038). Of the 72 rare variants identified none of them were found in the coding region of *TWIST1*. Within the scope of this project, functional analyses efforts were focused on the best associated common variants rs9931509 (*SNAI3*) and rs2189000 (*TWIST1*). The presence of RNA transcripts for SNAI3, RNF166 and TWIST1 was verified by RT-qPCR, in 2 human craniofacial cell lines, one oral epithelial cell line (GMSM-K,) and one embryonic palatal mesenchymal

cell line (HEPM). For experiments with TWIST1 rs2189000, a Chinese hamster ovary (CHO) cells as well as 293T and LS8 were also utilized. Functional experiments with rs9931509 (*SNAI3*) did not worked as planned unfortunately (see below). We had difficulties with transfection efficiency for luciferase assays. For TWIST1 on the other hand, ChIP and luciferase reporter assay results performed in multiple cell lines indicated that SNP rs2189000 located 5' of TWIST1 is a functional SNP that disrupts a transcription factor binding site for PITX2. Thus PITX2 directly regulates expression of TWIST1 by binding preferentially the consensus sequence with the common allele A. PITX2 binding is disrupted and TWIST1 expression is largely decreased in the presence of the rare allele G. Dysregulation of TWST1 expression might play an important role in determining the size of the body of the mandible. Further confirmation on these results in vivo will continue to clarify the role of TWIST1 in the growth of the maxilla-madibular complex.

Specific Aims:

Aim 1. Identify human variants in SNAI3 and TWIST1 underlying quantitative malocclusion phenotypes.
Previously we found significant associations between a quantitative phenotype ranging from severe convexity or concavity of the skeletal profile as well as dento-alveolar relations ranging between positive and negative overjets and various SNPs located within a ~12Kb interval surrounding *SNAI3*. Also association between a quantitate phenotype depicting variation from a large to a small mandibular body was also detected with a ~15Kb interval surrounding *TWIST1*. Preliminary fine mapping with additional SNPs within these regions identified one more SNPs near SNAI3 with more significant association results than in our initial study. Thus additional variants not yet identified in the region can better account for the association signals. Thus we propose in depth sequencing of the association peaks around *SNAI3* and *TWIST1* to identify and functionally annotate all the genetic variation within these regions.

Aim 2. Functional characterization of genetic variants identified within *SNAI3* and *TWIST1*.
Bioinformatics analyses of associated variants will be performed for annotation and prioritizing of functional studies in cell lines that are relevant for craniofacial development.
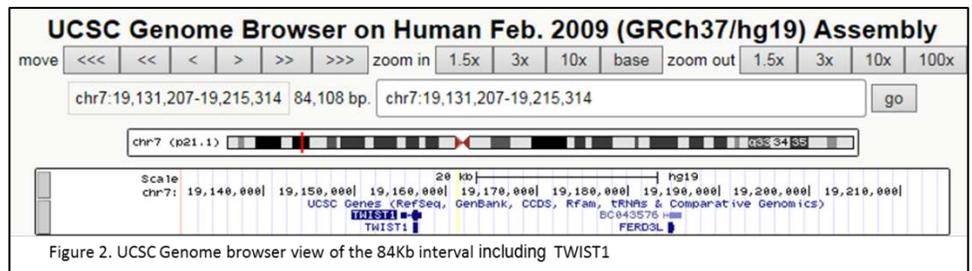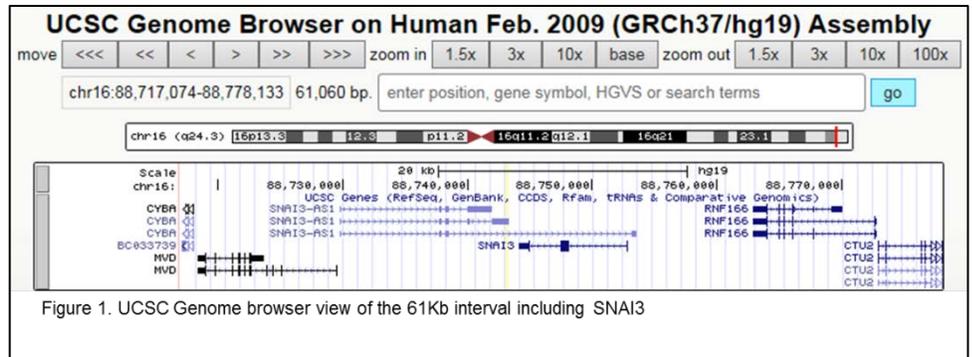
Studies and Results:

Specific Aim 1: Our previous study (da Fontoura et al., 2015) aimed at identifying genotype-phenotype correlations between 198 SNPs within 71 candidate genes for craniofacial development and dento-skeletal phenotypes derived from pretreatment lateral cephs on 269 Caucasian individuals of European descent. Phenotype-genotype associations were tested via multivariate linear regression for continuous phenotypes via principal component analysis (PCA) and multinomial logistic regression for skeletal malocclusion class. PCA resulted in 4 principal components (PCs) explaining 69% of the total skeletal facial variation. PC1 explained 32.7% of the variation and depicted vertical discrepancies ranging from skeletal deep to open bites. PC1 was associated with a SNP near PAX5 (P = 0.01). PC2 explained 21.7% and captured horizontal maxilla-mandibular discrepancies. PC2 was associated with SNPs upstream of SNAI3 (rs4287555 P = 0.0002) and MYO1H (P = 0.006). Particularly the minor allele of rs4287555 upstream of *SNAI3* was associated with a severe class II phenotype and accentuated convex profile. PC3 explained 8.2% and captured variation in ramus height, body length, and anterior cranial base orientation. PC3 was associated with TWIST1 (rs2189000 P = 0.000076), particularly the minor allele of *TWIST1* rs2189000 was associated with a shorter ramus, a larger body length, and a steep anterior cranial base orientation. Finally, PC4 explained 6.6% and detected variation in condylar inclination as well as symphysis projection. PC4 was associated with PAX7 (P = 0.007). Furthermore, skeletal class II risk increased relative to class I with the minor alleles of SNPs in FGFR2 (odds ratio [OR] = 2.1, P = 0.004) and declined with SNPs in EDN1 (OR = 0.5, P = 0.007). Conversely, skeletal class III risk increased versus class I with SNPs in FGFR2 (OR 2.2, P = 0.005) and COL1A1 (OR = 2.1, P = 0.008) and declined with SNPs in TBX5 (OR = 0.5, P = 0.014). In summary, PAX5, SNAI3, MYO1H, TWIST1, and PAX7 were associated with craniofacial skeletal variation among patients with malocclusion, while FGFR2, EDN1, TBX5, and COL1A1 were associated with type of skeletal malocclusion.

To follow up our SNAI3 and TWIST1 results subsequent genotyping in additional individuals (N=277) was performed in Dr. Moreno's lab to better define association intervals within SNAI3 and TWIST1. This new round of genotyping found that a new SNP rs9931509 located in the 3' UTR of *SNAI3* yielded better association results (P=0.00002) with PC2 than the previously found SNP rs4287555 and confirmed the TWIST1 association with SNP rs2189000 and PC3. Thus, the main goal of the current proposal was to follow up on the previous associations by sequencing the entire intervals for SNAI3 and TWIST1. Also, in order to not miss any phenotype-genotype correlations not only the previously associated phenotypes PC2 and PC3 but also all
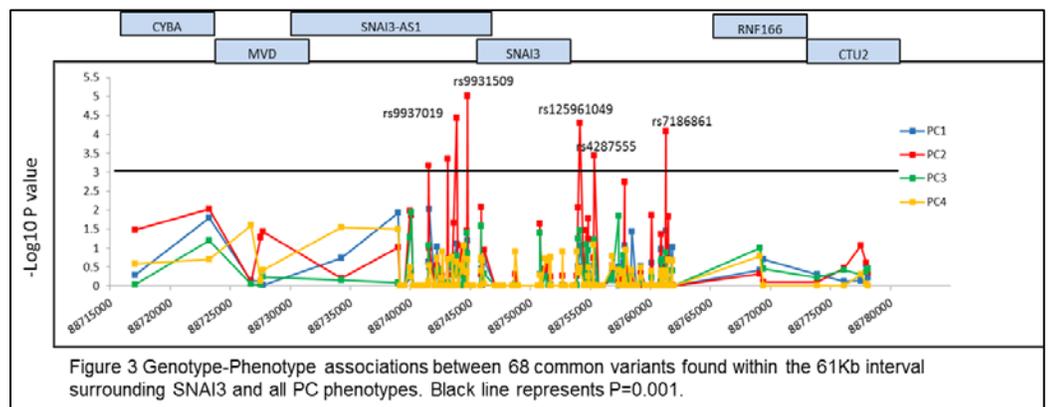
previous continuous phenotypes (PC1-PC4) as well as Skeletal Class I, II and III were analyzed against the SNAI3 and TWIST1 sequencing data. With the previous genotypic data (da Fontoura et al., 2015 plus added genotyping) in addition to the sequencing performed within this award for both SNAI3 (~28Kb) and TWIST1 (~32Kb) we have secured coverage for 61Kb around SNAI3 (16q24.3: 88717074-88778133 bps) at an average resolution of a called variant per every 350 bps and 84Kb including TWIST1 (7p21.1: 19131207-19215314 bps) with an average resolution of a called variant per every 667 bps in 272/277 individuals attempted (Figures 1 and 2).
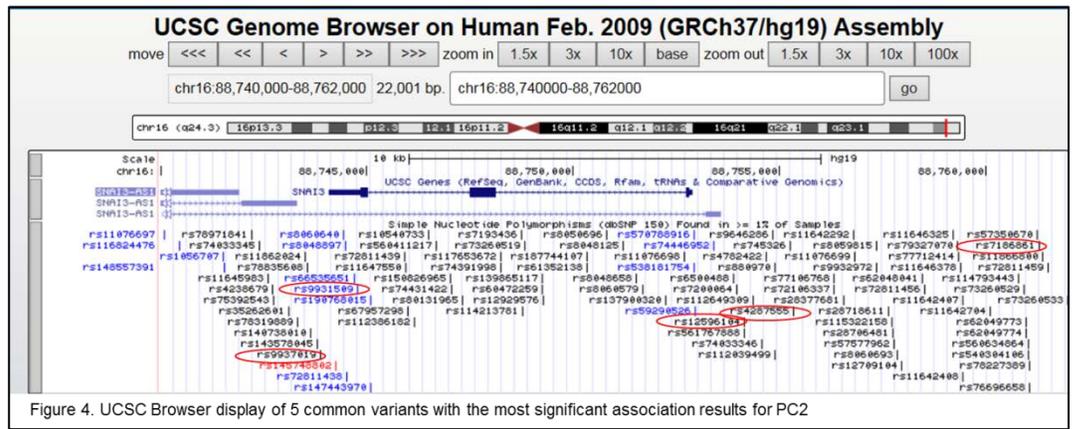

Figure 1. UCSC Genome browser view of the 61Kb interval including SNAI3

Figure 2. UCSC Genome browser view of the 84Kb interval including TWIST1

SNAI3 Sequencing Results: The 61 Kb interval around SNAI3 (chr16q24.3) includes the following genes in order from centromere to telomere CYBA, MVD, SNAI3-AS1 (non-coding RNA), SNAI3, RNF166 and CTV2 (Figure 1). Within the 61 Kb interval we have genotypic calls for 173 variants including SNPs and insertion deletion variations in 272 individuals (Angle Class I=55, Class II=129, Class III=88). Similarly to our previous studies, genotypes were recoded to 0, 1 or 2 depending on the copies of the minor allele present on an individual's genotype. Of 173 variants called, 105 occurred at a frequency of less than 5%. Analyses of the 68/173 remaining common variants for all phenotypes derived previously, confirmed significant associations with PC2 (Figure 3). Associations with other phenotypes (PC1, PC3, PC4 and Skeletal Class) only showed marginal results (P<0.05) for the SNAI3 interval, with the best associations noted for a new SNP rs_chr16_88741562 with

Figure 3 Genotype-Phenotype associations between 68 common variants found within the 61Kb interval surrounding SNAI3 and all PC phenotypes. Black line represents P=0.001.

PC1 (P=0.009), rs11076697 with PC3 and rs4782394 (P=0.02) with PC4. Skeletal Class showed that the risk of Class III vs. Class I decreased with more copies of the minor allele for rs11642704 (OR=0.56, P=0.03).

As discussed above our most significant associations were found with PC2, a phenotypic component that explained 21.7% of the variation and captured horizontal maxilla-mandibular discrepancies ranging from severe convexity to concavity of the skeletal profile as well

Figure 4. UCSC Browser display of 5 common variants with the most significant association results for PC2

as dento-alveolar relations ranging between positive and negative overjets. Five common SNPs rs9937019, rs9931509, rs125961049, rs4287555 and rs7186861 showed the strongest associations with PC2 with the best signal observed for rs9931509 (P=0.00001) (Figure 3) which we had found in our prior association studies. Thus the sequencing effort in this proposal did not reveal any new common variants with better associations but did confirm and solidify the prior findings and helped redefine the associated interval more clearly. SNP rs9931509 is located between SNAI3-AS1 and SNAI3 which are transcribed in opposite directions. Thus rs9931509 is approximately 2.6 kb downstream of SNAI3-AS and 345 bp downstream of SNAI3 (Figure 4).

Table 1 shows detailed information on the 5 common variants with the best association signals including their linkage disequilibrium

Table 1. Information on the 5 most significantly associated variants around SNAI3								
Variant	Variant type/function	Location	MA=MAF 1000 Genomes	MA=MAF Study sample	Assoc. P value	LD with rs9931509 (D'/R²)	Regulome DB score
rs9937019	Intronic for SNAI3-AS1	88743872	T=36%	T=28%	0.00003	1/0.97	4
rs9931509	Intronic for SNAI3-AS1 & 3' UTR for SNAI3	88744756	C=44%	C=30%	0.00001	1/1	2b
rs125961049	1.2 Kb upstream of SNAI3	88754144	C=37%	C=30%	0.00005	0.92/0.8	3a
rs4287555	2.4Kb upstream of SNAI3	88755325	G=35%	G=46%	0.0003	1/0.39	1f
rs7186861	8.4 Kb upstream of SNAI3	88761282	A=36%	A=29%	0.00008	0.89/0.75	5

values against the SNP with the strongest association rs9931509 and Regulome DB scores which range from 1-6 representing higher to lower regulatory potential. All 5 SNPs seemed to be in large LD (D'=1) indicating high correlation with each other and thus likely representing the same association signal. Based on the Regulome scores, 2/5 SNPs rs4287555 (Score=1f) and the strongest associated SNP rs9931509 (Score=2b) seem to have reasonable regulatory potential. A score of 1f indicates that the variant rs4287555 is likely to affect binding of a transcription factor, regulate the expression of a transcript via eQTL function, or chromatin structure. The variant rs4287555 is a known cis eQTL for monocytes (Zeller et al., 2010) and also for various tissues including muscle-skeletal tissue. It appears to regulate transcripts for SNAI3-AS1, SNAI3 and RNF166.

Figure 5 shows examples of eQTL box plots for expression of the SNAI3-AS1 in the heart-left ventricle and SNAI3 in lung and muscle-skeletal tissue (data from https://www.gtexportal.org/home/). Both hets and homozygous for the rare allele of rs4287555 showed a significant increase in expression level (effect size
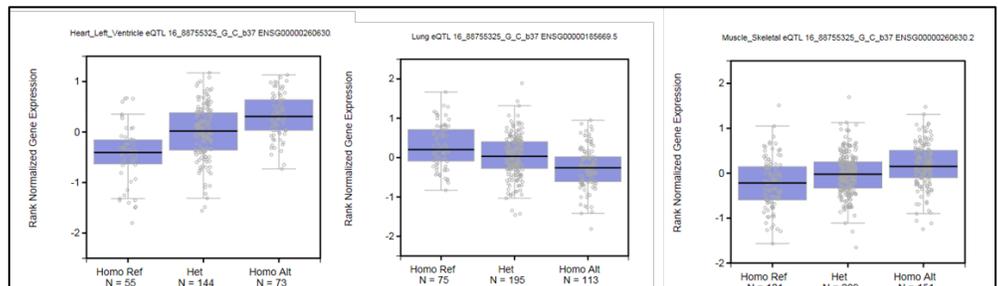

Figure 5. eQTL box plots from https://www.gtexportal.org/home/ showing the changes in expression associated with SNP rs4287555 genotypes in Heart-left ventricle (SNAI3-AS1), Lung and Muscle skeletal tissues (SNAI3).

0.48, P=1.1e-15) for SNAI3-AS1 in the heart-left ventricle. Conversely, both hets and homozygous for the rare allele of rs4287555 showed a significant decrease in SNAI3 expression for lung tissue (effect size -0.34, P=1.7e-12) and a mild increase for muscle-skeletal tissue (effect size 0.23 P= 1.4 e-10). Also rs4287555 is

located in an area of an H3K27Ac mark often found near active regulatory elements for different cells. Finally rs4287555 is within a region of protein factor mediated chromatin interaction as seen in 5 human cancer cells and also seems to function in histone modifications by repressing or weakly repressing PolyComb group proteins in various tissue groups including epithelial, brain, digestive, muscle and blood (ENCODE, REMC). A score of 2b indicates that rs9931509 is likely to affect binding of another protein, in this case USF1 (ENCODE). USF1 has been associated with hyperlipidemia and Alzheimer's disease and although it is mildly expressed in palatal tissue, its role in craniofacial development has not been described. SNP rs9931509 is also an eQTL for the CTU2 gene in various tissues and for the SNAI3-AS1 and the RNF166 gene in heart left ventricle and transformed fibroblasts respectively (box plots not shown). CTU2 expression in the facial structures has not been shown, however, RNF166 is highly expressed in maxilla, mandible frontonasal processes and palatal tissues in mice and human per SysFACE data (https://bioinformatics.udel.edu/research/sysface/). SNP rs9931509 also participates in modification of chromatin structure in various tissue groups. Furthermore, the Decipher track of UCSC browser showed that individuals with deletions or duplications in the genomic region encompassing these 5 SNPs present with craniofacial abnormalities including depressed nasal bridge, hypertelorism, microgntahia and midface retrusion, everted lower lips, thin upper lip vermillion, small, widely space or malformed teeth, frontal bossing and high palate.

In summary the most promising common variants for functional analyses are rs4287555 and rs9931509. They both are eQTLs for SNAI3-AS1, SNAI3 and RNF166. New data from SysFACE shows that RNF166 is highly express in maxilla, mandible and palate of both mice and human, and although the SNPs within RNF166 genotyped so far did not yield significant results for our phenotypes, it could well be that the phenotypic associations are driven by regulatory SNPs near SNAI3 yet the target gene is RNF166. Future functional research involving both SNAI3 and RNF166 will help clarify these results.

One hundred and five rare variants within the SNAI3 interval were analyzed in different categories based on the counts of rare variants in any given individual. Categories analyzed included: Total number of rare variants per individual, at least one rare variant vs. none, the presence of 0 vs. 1 vs. 2 vs. 3 or more rare variants and finally the presence of 0 vs. 1 vs. 2 or more rare variants. Of the 272 individuals, 128/272 had cero rare variants, 103/272 had at least one rare variant, 28/272 had 2 rare variants and 13/272 individuals had 3 or more rare variants. The dependent variable was changes in PC scores per each principal component analyzed both at the mean level using linear regression and also at the extremes of the variation. That is whether individuals fall on the 10th percentile or below or 90th percentile or above and also on the 25th percentile or below or 75th percentile or above. Further we performed quantile regression to test changes in quantiles of each PC (0.1, 0.25, 0.5, 0.75 and 0.9) for the different rare variant count categories. Finally, risk for Class III or Class II vs. Class I was also analyzed for the same categories of rare variants. For changes in PC scores at the mean level, linear regression models returned none significant results (P<0.05) for any of the rare variant count categories regardless of PC evaluated. Interestingly however, PC2 our highly associated component above, had the smallest P values (P~0.1) when compared to all other PCs (P values >0.5), indicating a possible association trend between PC2 and counts of rare variants for SNAI3. For the odds of falling into a specific PC threshold, significant results (P<0.05) were obtained with the odds of having a PC2 score in the 90th percentile or above and the presence of any rare variant vs. none (P=0.03), 1 rare variant vs. cero (P=0.017) and 1 rare variant vs. 0 vs. 2 or more (P=0.018). Quantile regression returned significant results (P<0.05) for PC4 and changes in quantiles. Recall that PC4 explained 6.6% of the variation and depicts changes in condylar inclination as well as symphysis projection. Significant results were seen for the variables total rare variant count with q10 (P=0.003), 2 or more rare variants vs. cero with q10 (P=0.006) and 3 or more rare variants (P<0.001). For skeletal Class risk, no significant results were obtained for any of the rare variant count categories analyzed. Lastly, within the 105 rare variants we identified 4 coding mutations in 6 individuals described in table below. One of these Chr16:88747702; G166E is a new mutation not previously reported in EXACT or GNOMAD. At first glance the phenotypic distribution of individuals carrying the same rare coding variant does not seem to follow a specific pattern, suggesting that these rare variants may not be related to a specific phenotypic pattern. Efforts are ongoing for haplotype tests and bioinformatics analyses of these

variants in an attempt to select the best *SNAI3* candidates for the functional analyses proposed in specific aim 2.

Table 2 Coding variants in SNAI3				
Coding Variant	**Position**	**Pathogenic**	**Ind**	**Phenotype**
rs145970254; R229C	88747514	Deleterious missense	1het	Class II, ANB 2, Witts1, Overjet 5, Neg. PC2.
*Chr16:88747702; G166E	88747702	Deleterious missense	1het	Class II, ANB 6, Witts1.3, Overjet 4, Neg. PC2
			1het	Class III, ANB -2, Witts -4.7, Overjet 0, Neg. PC2
rs142561823; G151S	88747748	Tolerated missense	1het	Class II, ANB 4.7, Witts 2, Overjet 4, Neg. PC2
			1het	Class III, ANB -1, Witts -7.0, Overjet 0.5, Pos.PC2
rs150728595; Q24	88752743	Synonymous	1het	Class III Negative ANB 0, -2.9, Overjet 3, Pos. PC2
*New mutation not found in GNOMAD.				

TWST1 Sequencing Results: The 84 Kb interval including *TWIST1* (chr7p21.1: 19131207-19215314 bps) also includes the gene FERD3L (Figure 2). Within this 84 Kb interval we have genotypic calls for 126 variants including SNPs and insertion deletion variations. Similarly to our previous studies, genotypes were recoded to 0, 1 or 2 depending on the copies of the minor allele present on an individual's genotype. Of 126 variants called, 72 occurred at a frequency of less than 5%. Analyses of the 54/126 remaining common variants for all phenotypes derived previously confirmed significant associations with PC3 (Figure 6). Associations with other phenotypes only showed marginal results (P<0.05) for
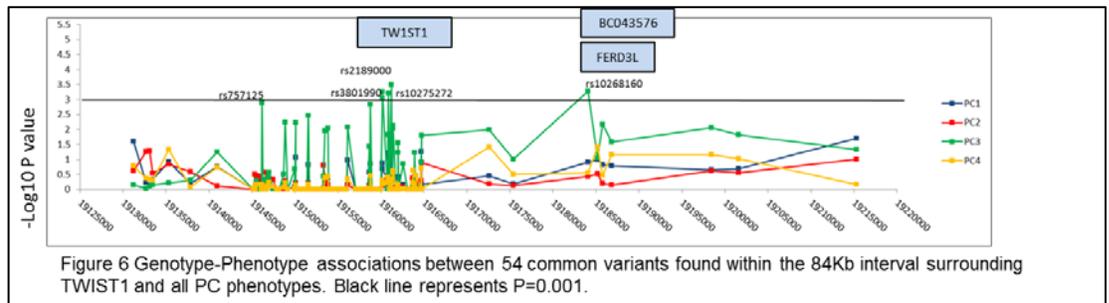

Figure 6 Genotype-Phenotype associations between 54 common variants found within the 84Kb interval surrounding TWIST1 and all PC phenotypes. Black line represents P=0.001.

the TWIST1 interval, with the best associations noted for rs7788015 with PC1 (P=0.02), rs10225279 with PC2 (P=0.05) and rs1157350 and PC4 (P=0.01). Skeletal Class showed that the risk of Class III vs Class I decreased with the minor allele of rs117390026 (OR=0.17, P=0.04).

As discussed above our most significant associations were found with PC3, a phenotypic component that explained 8.2% of the variation and captured differences in ramus height, body length, and anterior cranial base orientation. Five common SNPS rs3801991, rs3801990, rs10275272, rs2189000, rs10268160 yielded the best associations (P<0.001), with the best signal observed for rs2189000 (P=0.0003) (Figure 6) which we had found in our prior association studies. Thus similarly to SNAI3, sequencing efforts in this proposal did not reveal any new common variants with better associations but did confirm and
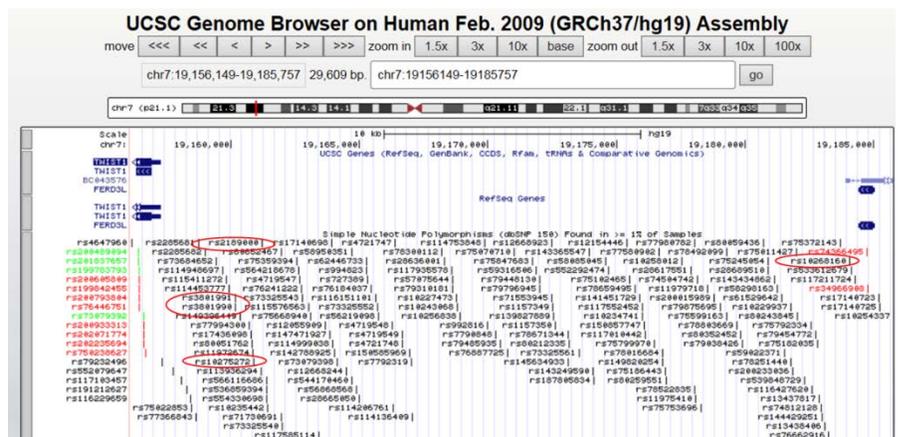

Figure 7. UCSC Browser display of 5 common variants with the most significant association results for PC3

solidify the prior findings and helped redefine the associated interval more clearly. TWIST1 SNP rs2189000 (P=0.0003) is located 3.9 kb upstream of TWSIT1 (Figure 7).

Table 3 shows the 5 SNPs with the best association signals including their linkage disequilibrium values against the SNP with the strongest association rs2189000 and Regulome DB scores which range from 1-6 representing higher to lower regulatory potential. All 5 SNPs seemed to be in large LD (D'=1) indicating high correlation with each other and thus likely representing the same association signal. Based on the Regulome scores all 5 SNPS

Table 3. Information on the 5 most significantly associated variants around TWIST1							
Variant	Variant type/function	Location	MA=MAF in 1000 Genomes	MA=MAF in study sample	Assoc. P value	LD with rs2189000 (D'/R^2)	Regulome DB score
rs3801991	Variant 2.8 Kb upstream of TWIST1	19160177	A=16.5%	A=12.4%	0.0009	1/0.95	5
rs3801990	Variant 2.8 Kb upstream of TWIST1	19160183	G=16.5%	G=12.6%	0.0005	1/0.95	5
rs10275272	Variant 3.6Kb upstream of TWIST1	19160897	T=28%	T=14.4%	0.0006	1/1	4
rs2189000	Variant 3.9Kb upstream of TWIST1	19161218	G=27%	G=16.6%	0.0003	1/1	5
rs10268160	Non-coding transcript variant	19184059	G=32%	G=20%	0.0005	1/0.814	4

seem to have minimal regulatory potential and none of them seem to be an eQTL. Two other variants rs10254337 (P=0.006) and rs4647960 (P=0.008) also in LD with rs2189000 (not shown) showed Regulome scores of 3a and 2b respectively indicating moderate regulatory potential. Similar to the SNAI3 region above, the Decipher and Copy Number Variant tracks of the UCSC browser shows that individuals with deletions, duplications or copy number variations in the genomic region encompassing these highly associated SNPs present with craniofacial abnormalities including: micrognathia, narrow mouth, craniosynostosis, skull asymmetry, abnormal facial shape, cleft palate, abnormal dental enamel, flat occiput, low set ears, microcephaly and microtia. According to OMIM the area including rs2189000 is also a locus for ectodermal dysplasia-syndactyly syndrome 2.

Seventy two rare variants (frequency <5%) within the TWIST1 interval were analyzed in different categories based on the counts of rare variants in any given individual. These categories were: Total rare variant count (0-5) variants per individual, at least one rare variant vs. none, the presence of 0 vs. 1 vs. 2 vs. 3 or more rare variants and finally the presence of 0 vs. 1 vs. 2 or more rare variants. Successful sequencing was obtained for 271/272 individuals attempted. Of these, 144/271 had cero rare variants, 96/271 had one rare variant, 21/271 had 2 rare variants and 10/271 individuals had 3 or more rare variants. The dependent variables were changes in PC scores both at the mean level using linear regression and also at the extremes of the variation. That is whether individuals fall on the 10th percentile or below, 25th percentile or below, 75th percentile or above and 90th percentile or above. Furthermore, we performed quantile regression to test changes in quantiles at each PC (0.1, 0.25, 0.5, 0.75 and 0.9) for the different rare variant count categories. Finally, risk for Class III or Class II vs. Class I was also analyzed for the same categories of rare variants. For changes in PC scores at the mean level, linear regression models returned significant results (P<0.05) for PC3 with the variables total rare variant count (P=0.048), and 3 or more rare variants vs. cero (P=0.029). For the odds of falling into a specific PC threshold, significant results (P<0.05) were obtained with the odds of having a PC1 score in the 25th percentile or below and the presence of 1 or more rare variant vs. cero (P=0.015). Also for the odds of having a PC1 score in the 75th percentile or above and the variable total rare variant count (P=0.017) and 1 or more rare variants vs. cero (P=0.011). Moreover, significant results were also obtained for the odds of having a PC3 score in the 25th percentile or below with the variable total rare variant count (P=0.03). Quantile regression returned significant results (P<0.05) also for PC1 and PC3. For PC1 significant results were seen for PC1Q75 with total rare variant count (P=0.047), for PC1Q90 with the variable any rare variant (P=0.041), for PC1Q10 with 1 or more (P=0.005) and 2 or more (P=0.001) rare variants. Finally for PC3 significant quantile regression results were seen at PC3Q10 (P=0.003) and PC3Q25 (P=0.005) with 3 or more rare variants. For skeletal Class risk, significant results was seen for Class II vs all others and the presence of 2 or more rare variants (P=0.05). Also a significant result was seen for Class II vs. all others (Class I and III combined) and the presence of 3 or more rare variants (P=0.038). Of the 72 rare variants identified none of them were found in the coding region of TWIST1.

In summary both common and rare variant counts around the TWST1 interval are associated with quantitative malocclusion phenotypes. TWIST1 common SNP rs2189000 was selected for functional analyses as described below in specific aim 2.

Specific Aim 2: Within the scope of this project, functional analyses efforts were focused on the best associated common variants rs9931509 (*SNAI3*) and rs2189000 (*TWIST1*).

SNAI3 Functional Analyses:

Confirming expression of *SNAI3* and *RNF166* in craniofacial cell lines GMSM-K and HEPMs:

To test for the expression of the genes of interest SNAI3 and RNF166 in human fetal oral epithelial cells (GMSM-K) and human embryonic palatal mesenchymal cells HEPMs total RNA was isolated using RNeasy mini kit (Qiagen, Hilden,Germany) and treated with RNease-free DNaseI (Promega, Madison, WI, USA).Reverse transcription was performed by High Capacity cDNA Reverse Transcription Kits (Applied BIosystems, Foster city, CA, USA). Real-time PCR was performed in CFX96 TouchReal-Time PCR Detection System using SYBR-Green qPCR (Bio-rad). Reactions were done in duplicates. For SYBR-Green qPCR, primers were synthesized by IDT (Coralville, IA, USA). The Beta-actin gene was used as internal control. cDNA amplicons were confirmed with Sanger sequencing.
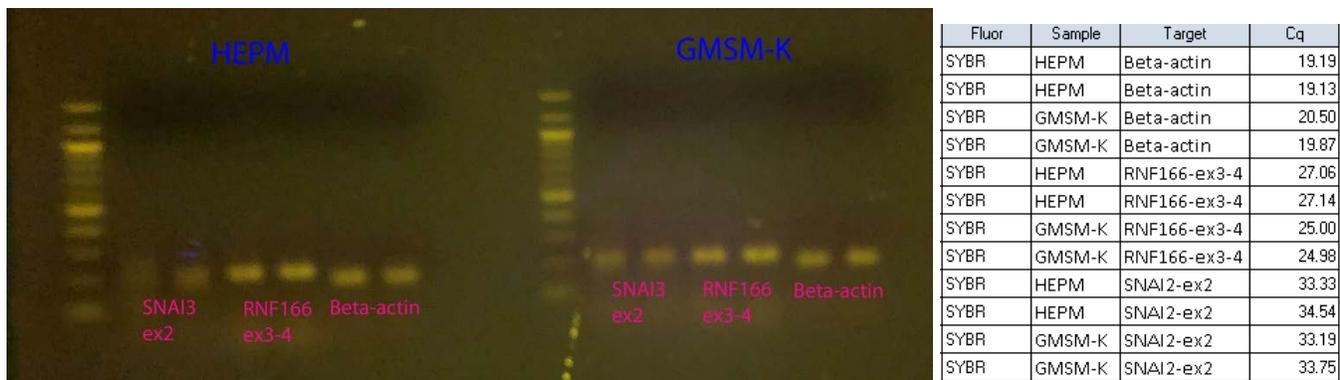

Fluor	Sample	Target	Cq
SYBR	HEPM	Beta-actin	19.19
SYBR	HEPM	Beta-actin	19.13
SYBR	GMSM-K	Beta-actin	20.50
SYBR	GMSM-K	Beta-actin	19.87
SYBR	HEPM	RNF166-ex3-4	27.06
SYBR	HEPM	RNF166-ex3-4	27.14
SYBR	GMSM-K	RNF166-ex3-4	25.00
SYBR	GMSM-K	RNF166-ex3-4	24.98
SYBR	HEPM	SNAI2-ex2	33.33
SYBR	HEPM	SNAI2-ex2	34.54
SYBR	GMSM-K	SNAI2-ex2	33.19
SYBR	GMSM-K	SNAI2-ex2	33.75

Figure 8. RT-qPCR showing expression of *SNAI3* and *RNF166* in GMSM-K and HEPM cell lines

Testing for SNPrs993159 enhancer activity *in vitro*:

First we amplified an 856 bp (chr16:88,744,167-88,745,010) DNA element which harbors expressing chromatin marks indicative of enhancer activity as well as containing the risk-associated SNP rs9931509. We tested the enhancer activity of this element in vitro utilizing human fetal oral epithelial cells GMSM-K. Both the elements containing the common rs9931509_A allele and the rare rs9931509_C allele were synthetized and subcloned into a pGL3-Promoter firefly luciferase vector (Biobasic) and electroporated for transfection together with a plasmid that drives constitutive expression of renilla luciferase (transfection control) for three biological replicates. Three different protocols for electroporation were tested. Plasmids were electroporated into GMSM-K cells (500,000 cells) with Amaxa Cell Line Nucleofector Kit V (Lonza, Cologne, Germany) using Nucleofector II (Lonza) programs: C-017 (500,000 cells, 1ug fire fly plasmids A and C, 1ug renilla plasmid), program: T-020 (500,000 cells, 3ug fire fly plasmids A and C, 0.7ug renilla plasmid) and program: X-005 (500,000 cells, 1ug fire fly plasmids A and C, 1ug renilla plasmid). The Dual-Luciferase Reporter Assay System (Promega, Madison, WI, USA) and 20/20n Luminometer (Turner Biosystems, Sunnyvale, CA, USA) were employed to evaluate the luciferase activity when cells were approximately 60% confluent at 72 h post-transfection following the manufacturer's instructions. Relative luciferase activities were calculated by the ratio between the value for Firefly and Renilla luciferase activities. Three measurements were made for the lysate from each transfection group. Unfortunately results indicated low transfection efficiency with Firefly/Renilla values below 300 which indicates background levels for programs C-017 and X-005. The T-020 electroporation protocol resulted in complete cell death and therefore it was not tested for luciferase activity. Efforts are ongoing to trouble shoot transfection efficiency at this time.

TWIST1 Functional Analyses:

We previously identified significant associations between SNPs rs2189000, 5' of *TWIST1* with mandibular body length in patients with moderate or severe malocclusion (p<0.0001). In this study we evaluated the associated SNP rs2189000 for its regulatory potential on *TWIST1* function.

First, *in sílico* analysis was used for identification of target-specific transcription factors predicted to bind to the rs2189000 site. Next, a Western Blot technique was used to identify the presence of TWIST1 in human embryonic palatal mesenchyme (HEPM) and Chinese hamster ovary (CHO) cells. Specificity tests were also carried out with endogenous chromatin immunoprecipitation (ChIP) assay using primers upstream of the *TWIST1* promoter. ChIP assays were performed using the ChIP Assay Kit (Upstate) and DNA from the precipitants were subject to PCR to evaluate relative enrichment. Finally, a luciferase reporter assays were used in multiple cell lines (HEPM, GMSM-K, CHO, 293T and LS8) to measure promoter activity of *TWIST1* and determine if SNP rs2189000 is functional in regulating *TWIST1* promoter activity.

Results:

In silico analysis predicted that the *TWIST1* SNP rs2189000 disrupts a transcription factor bindings site (TFBS) for *PITX2*. *PITX2* is a transcription factor that controls cell proliferation in a tissue-specific manner and is involved in the development of the eye, tooth and abdominal organs. Mutations in PITX2 are associated with Axenfeld-Rieger syndrome. Patients with this condition can present with craniofacial anomalies such as maxillary hypoplasia. A Western blot analysis identified the presence of TWIST1 protein in human embryonic palatal mesenchyme (HEPM) and Chinese hamster ovary (CHO) cells (Figure 9). ChIP assay results showed that endogenous *PITX2* indeed binds to this consensus binding site 5' of *TWIST1* sequence containing the rs2189000 common allele (A) (Figure 10). Therefore, the data suggests that the SNP might disrupt the ability of the *PITX2* to associate with *TWIST1* promoter and stimulate its expression. Quantitative PCR - performed in duplicates -revealed an increase of 6-fold enrichment of the *TWIST1* ChIP product over that of the IgG control (Figure 11). The luciferase reporter assay using a construct bearing SNP rs2189000 rare allele G in the *TWIST1* enhancer region showed that co-transfection of cells with PITX2 led to decrease in reporter activity when compared to the transfection performed with the reporter construct containing the common allele A (P= 0.00573912 - fig 12.) Co-transfection of the PITX2 gene with TWIST1 TK Luc construction resulted in 8.45-fold activation (Fig. 11; comparison is to transfection with empty vector). Mutation of the PITX2 binding site 5' of TWIST1 decreased TWIST1 expression (Fig. 12). These results indicate that PITX2 directly regulates TWIST1 promoter activity.
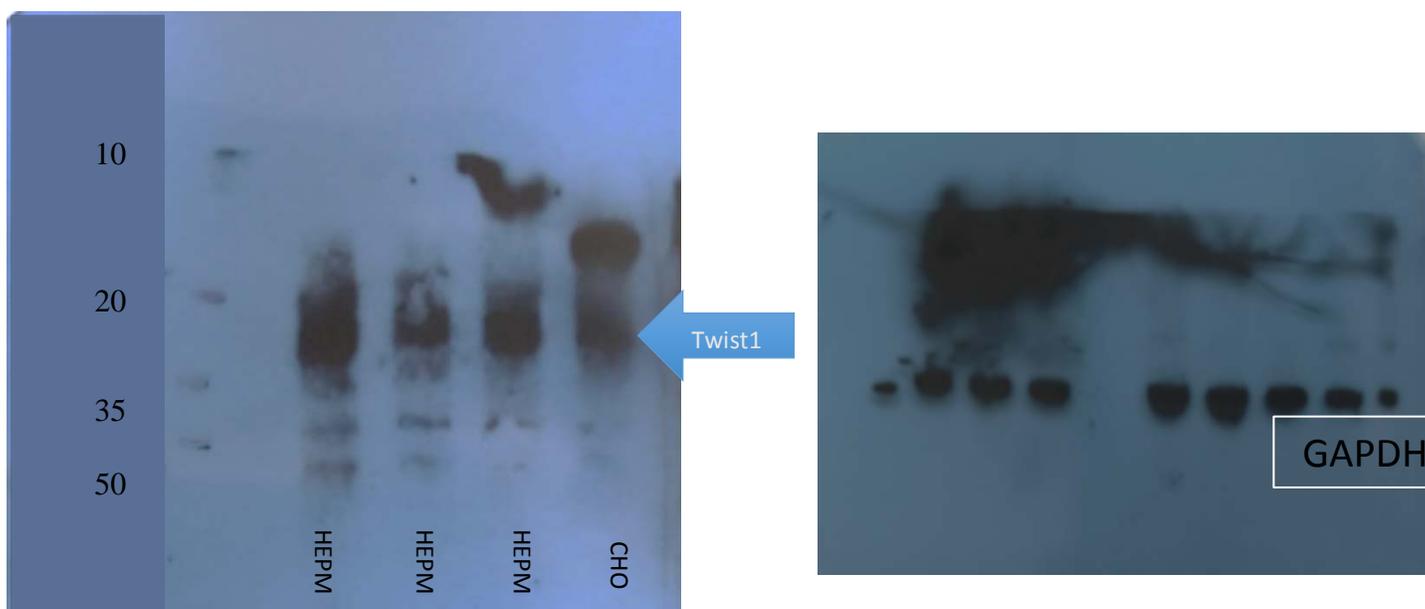

Fig 9. Western blot analyses identified the presence of *TWIST1* protein in HEPM and CHO cells (left). The GAPDH gene is used as control.
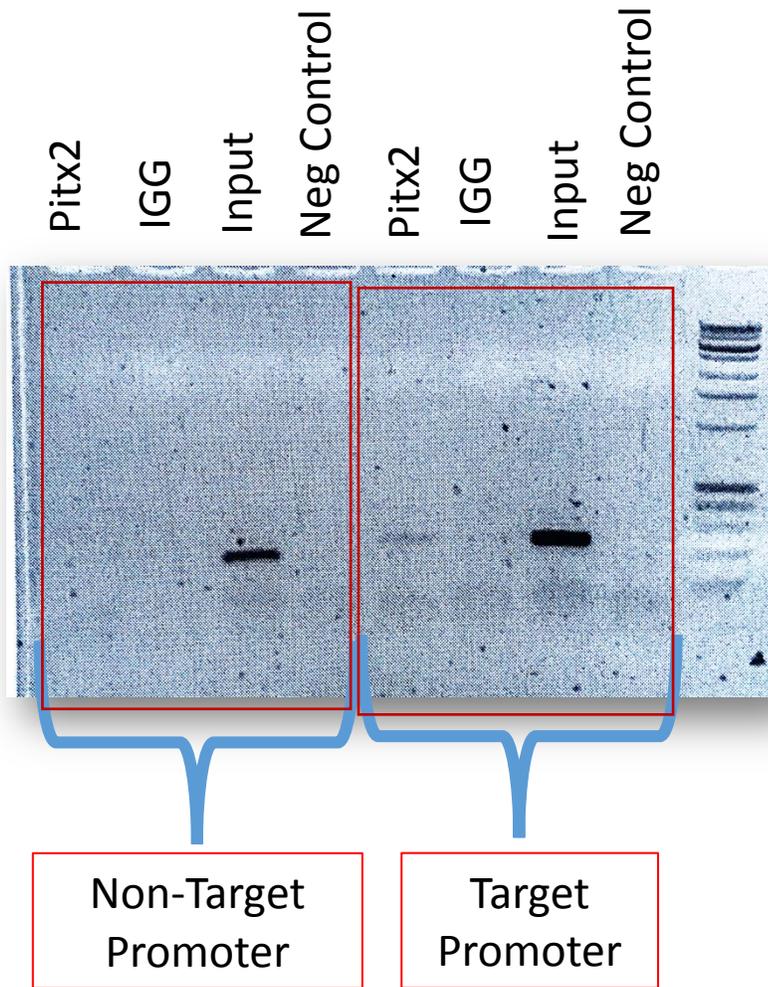
Fig 10. ChIP assay showed that the TWIST1 sequence containing the common allele (A)- right box- immunoprecipitated chromatin while the non- target TWIST1 -left box- failed to pull down the PITX2 chromatin, and the control IgG.

	Input normalized	Yield	Control adjusted Ct	Fold
Promoter	**DCt**	**% In**	**DDCt**	**Enrichment**
31.8	13.47	0.01	0.00	1.00
33.18667	14.86	0.00	1.40	0.38

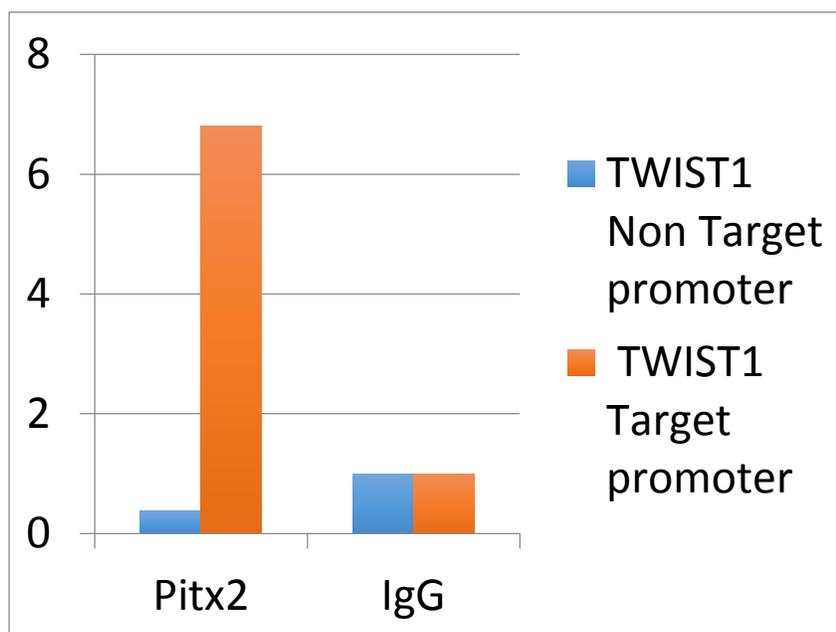|		**Target***	Input normalized	Yield	Control adjusted Ct	**Fold**
Antibody		**Promoter**	**DCt**	**% In**	**DDCt**	**Enrichment**
NS* IgG		40.0	12.63	0.02	0.00	1.00
Your* Ab		37.2333	9.86	0.11	-2.77	6.81
1 % Input	34.0	6.64	1.00	-5.98		

Fig 11. Quantitative PCR demonstrated a 6-fold enrichment of the *TWIST1* ChIP product over that of the IgG control.
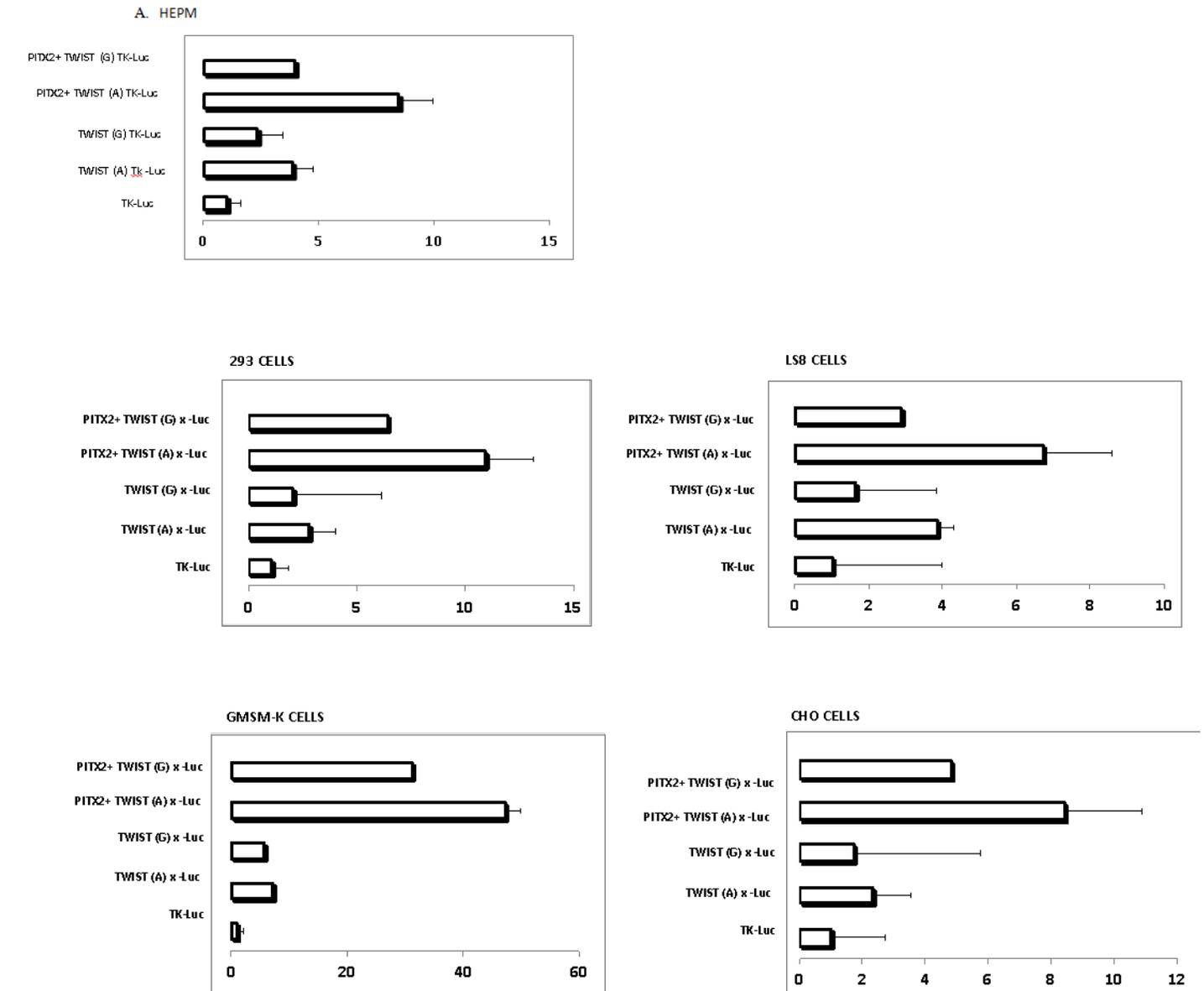
Fig. 12. A. Luciferase assay results indicate that PITX2 directly regulates the TWIST1 promoter activity (0.00075856) allowing an increase of 4.62-fold change (PITX2+ TWIST (A) x -Luc vs TWIST (A) x -**Luc** p= 0.00075856). Co-transfection of the PITX2 gene with TWIST1 -TK Luc construction in HEPM cells resulted in 8.45-fold activation compared to empty vector. Co-transfection of HEPM cells using a construct bearing SNP rs2189000_G in the *TWIST1* enhancer region showed that with *PITX2* there was a decrease in reporter activity when compared to the transfection performed with the reporter construct containing the common allele A (p= 0.00573912). B. Results are consistent throughout different cells lines. These results indicate that PITX2 directly regulates the TWIST1 promoter activity.

HEPM	Average	Average %	stdev	p-value	T-test
TK-Luc empyt vector	1.00	100.0%	0		
TK-Luc WT Twist1	3.83	383.2%	0.611486		
TK-Luc MT Twist1	2.33	232.7%	0.918753	0.000759	TWIST1 WT vs TWIST 1 WT + PITX2
TK-Luc WT Twist1 + PITX2	8.45	844.8%	1.116594	0.160869	TWIST1 MT vs TWIST1 MT + PITX2
TK-Luc MT Twist1 + PITX2	3.94	394.3%	1.489262	0.005739	TWIST1 WT + PITX2 vs TWIST1 MT + PITX2
GMSM-K					
TK-Luc empyt vector	1.00	100.0%	0		
TK-Luc WT Twist1	7.14	713.7%	1.232247		
TK-Luc MT Twist1	5.80	580.3%	0.240794	0.001052	TWIST1 WT vs TWIST 1 WT + PITX2
TK-Luc WT Twist1 + PITX2	47.22	4721.9%	0.416626	0.011522	TWIST1 MT vs TWIST1 MT + PITX2
TK-Luc MT Twist1 + PITX2	31.05	3105.4%	2.723609	0.027847	TWIST1 WT + PITX2 vs TWIST1 MT + PITX2
CHO					
TK-Luc empyt vector	1.00	100.0%	0		
TK-Luc WT Twist1	2.31	230.8%	1.688671		
TK-Luc MT Twist1	1.71	170.8%	1.205624	0.051754	TWIST1 WT vs TWIST 1 WT + PITX2
TK-Luc WT Twist1 + PITX2	8.44	844.0%	4.048048	0.092643	TWIST1 MT vs TWIST1 MT + PITX2
TK-Luc MT Twist1 + PITX2	4.83	483.0%	2.422491	0.233218	TWIST1 WT + PITX2 vs TWIST1 MT + PITX2
293T					
TK-Luc empyt vector	1.00	100.0%	0		
TK-Luc WT Twist1	2.76	276.2%	0.796368		
TK-Luc MT Twist1	2.05	204.7%	1.195372	0.049869	TWIST1 WT vs TWIST 1 WT + PITX2
TK-Luc WT Twist1 + PITX2	10.97	1096.9%	4.099406	0.070123	TWIST1 MT vs TWIST1 MT + PITX2
TK-Luc MT Twist1 + PITX2	6.43	643.0%	2.224443	0.240769	TWIST1 WT + PITX2 vs TWIST1 MT + PITX2
LS8					
TK-Luc empyt vector	1.00	100.0%	0		
TK-Luc WT Twist1	3.87	386.5%	2.995253		
TK-Luc MT Twist1	1.64	164.3%	0.419337	0.228922	TWIST1 WT vs TWIST 1 WT + PITX2
TK-Luc WT Twist1 + PITX2	6.73	673.2%	2.183438	0.29429	TWIST1 MT vs TWIST1 MT + PITX2
TK-Luc MT Twist1 + PITX2	2.89	289.1%	1.833965	0.058411	TWIST1 WT + PITX2 vs TWIST1 MT + PITX2

Methods

DNA cloning, shRNA, cell culture, transient transfection, luciferase, beta-galactosidase assay and Western blotting

A 4.72kb *TWIST1* DNA fragment- in which the Pitx2 binding site is present as well as the *TWIST1 rs2189000* mutated SNP- were cloned upstream the pTK-Luc vector (Promega) using the following primers: 5'-TCAGTGGATCCCTTTCATGACCCTGGTAGCC -3' and 5'- ACAGTAAGCTTGCCTGCTCATGTTGTTCAGA -3'.

HEPM, GMSM-K, CHO, LS-8 cells (oral epithelial-like cell) and 293T cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin and were transfected by lipofectamine and PEI. Transfected cells were incubated for 48 h in a.6 well culture dishes: 0.5×10^6 and fed with 10% FBS and DMEM and then lysed and assayed for reporter activity and protein content by Bradford assay (Bio-Rad). Luciferase was measured using reagents from Promega. β-galactosidase was measured using the Galacto-Light Plus reagents (Tropix Inc.). All luciferase activities were normalized to β-galactosidase activity. For Western blot assay, cell lysates (10 μg) were separated on a 10% SDS–polyacrylamide gel and the proteins were transferred to PVDF filters (Millipore), and immunoblotted using the following antibodies TWIST1, (GXT, 1:1000), PITX2 (Capra Sciences, 1:1000), GAPDH (Santa Cruz, 1:1000). ECL reagents from GE HealthCare were used for detection.

Chromatin Immunoprecipitation assay (ChIP)

The ChIP assays were performed as previously described using the ChIP Assay Kit (Upstate) with the following modification. HEPM cells were plated in 60 mm dishes and fed 24 h prior to the experiment, harvested and plated in 60 mm dishes. Cells were cross-linked with 1% formaldehyde for 10 min at 37°C. Cross-linked cells were sonicated three times to shear the genomic DNA to DNA fragments average ranged between 200 and 1000 bp. DNA/protein complex were immunoprecipitated with specific antibody (Pitx2 antibody, Capra Sciences, PA-1023; TWIST1 antibody, GTX, GTX127310). DNAs from the precipitants were subject to PCR to evaluate the relative enrichment.

Response to the following questions:

1. **Were the original, specific aims of the proposal realized?**
Yes. However some planned experiments did not work. I plan to continue to trouble shoot our techniques and send an updated report to the AAOF as soon as I get the results.

2. **Were the results published?**
The results are not published yet, but we have started the manuscript organization. Additionally, some of the data will be used for a planned R01 application

3. **Have the results of this proposal been presented?**
I am planning to present these results in the upcoming AAO meeting of 2019.

4. **To what extent have you used, or how do you intend to use, AAOF funding to further your career?**
I am currently preparing a manuscript with the data generated in this proposal in addition to the generation of preliminary data that I plan to use in future applications. Funding from the AAOF has been instrumental in advancing my career and to the understanding of the genetic variation underlying malocclusion conditions.