American Journal of Medical Genetics Part A 143A:3228–3234 (2007)

Sequence Evaluation of FGF and FGFR Gene Conserved Non-Coding Elements in Non-Syndromic Cleft Lip and Palate Cases

Bridget M. Riley and Jeffrey C. Murray*
Department of Pediatrics, University of Iowa, Iowa City, Iowa

Received 21 May 2007; Accepted 6 June 2007

Non-syndromic cleft lip and palate (NS CLP) is a complex birth defect resulting from multiple genetic and environmental factors. We have previously reported the sequencing of the coding region of genes in the fibroblast growth factor (FGF) signaling pathway, in which missense and non-sense mutations contribute to approximately 5%–6% NS CLP cases. In this article we report the sequencing of conserved non-coding elements (CNEs) in and around 11 of the *FGF* and *FGFR* genes, which identified 55 novel variants. Seven of variants are highly conserved among ≥8 species and 31 variants alter transcription factor binding sites, 8 of which are important for craniofacial development. Additionally, 15 NS CLP patients had a combination of coding mutations and CNE variants, suggesting that an accumulation of variants in the FGF signaling pathway may contribute to clefting.
© 2007 Wiley-Liss, Inc.

Key words: cleft lip and palate; fibroblast growth factor; conserved non-coding elements; sequencing

How to cite this article: Riley BM, Murray JC. 2007. Sequence evaluation of FGF and FGFR gene conserved non-coding elements in non-syndromic cleft lip and palate cases. Am J Med Genet Part A 143A:3228–3234.

INTRODUCTION

Comparing the human genome with distantly related species is a productive method to identify evolutionarily conserved non-coding elements (CNEs) [Venkatesh et al., 2006]. CNEs contain regulatory regions important for transcription, translation, recombination, replication and repair; however, a large portion of these CNEs contain sequence whose function is not understood [Emison et al., 2005; Ponting and Lunter, 2006]. Although conservation over evolutionary distance does suggest functionality, not all functional sequence will be conserved [Ponting and Lunter, 2006]. It has been suggested that 3–15 mammalian genomes are sufficient to detect transcription factor binding sites, although comparisons among vertebrates such as chicken or zebrafish are regularly utilized [Eddy, 2005; Emison et al., 2005]. The identification of functional elements through CNEs does have obstacles; nevertheless, cross-species comparison is the best approach available.

Studies using comparative genomics have shown that 1.5% of the human genome codes for exons whereas 3% is devoted to CNEs, suggesting CNEs are highly important locations to explore for mutations [Emison et al., 2005]. Interestingly, approximately 20–30% of patients with rare Mendelian diseases have no identifiable mutations in the coding region of the gene associated with the disease, suggesting non-coding regions may harbor the etiologic mutation in a significant fraction of cases [Emison et al., 2005]. Using systematic comparisons of orthologous sequences, a common non-coding variant was identified in the receptor tyrosine kinase, *RET*, gene in families with Hirschsprung disease (HSCR), a multifactorial, non-Mendelian disorder [Emison et al., 2005]. HSCR is one of the first complex diseases to be genetically dissected; the complex, multigenic nature of HSCR makes it comparable to non-syndromic cleft lip and palate (NS CLP) and suggests we should utilize similar approaches to identify genes and mutations in NS CLP patients.

NS CLP is a common, complex disease affecting ~1/700 births worldwide, with substantial variation based on geographic origin and socioeconomic

This article contains supplementary material, which may be viewed at the American Journal of Medical Genetics website at http://www.interscience.wiley.com/jpages/1552-4825/suppmat/index.html.

Grant sponsor: NIH; Grant numbers: ES1087, R37-DE08559, P50-DE016215, 5T32GM008629; Grant sponsor: Center for Inherited Disease Research CIDR NIH; Grant number: N01-HG-65403.

*Correspondence to: Jeffrey C. Murray, 2182 Medical Laboratories, Department of Pediatrics, University of Iowa, Iowa City, IA 52242.
E-mail: jeff-murray@uiowa.edu

DOI 10.1002/ajmg.a.31965

status [Jugessur and Murray, 2005; Lidral and Moreno, 2005]. We previously reported on the sequencing of the coding regions of the fibroblast growth factors (*FGFs*) and receptor (*FGFR*) genes in NS CLP patients [Riley et al., 2007]. Thirty-seven point mutations were identified in the exons or at the intron–exon junctions of the *FGF* and *FGFR* genes, nine of the mutations were either missense or nonsense, accounting for about 5–6% of the cases examined. Comprehensive examination of the functional elements within each *FGF* or *FGFR* locus affords an opportunity to identify additional variation and/or genes playing a critical role in CLP and could have substantial clinical implications. This study explores the CNEs in *FGF* and *FGFR* genes for contributions to NS CLP.

METHODS

Samples

Ninety-one DNA samples from cases in the Philippines and another 93 from cases born in Iowa were utilized for the sequencing studies. The 184 cases are isolated, non-syndromic unilateral cleft lip (n = 1), bilateral cleft lip (n = 2), unilateral CLP (n = 60), bilateral CLP (n = 102), and CLP with unknown laterality (n = 19). Whole blood samples were collected by venipuncture. Subjects were reviewed by JCM to exclude any with syndromic features and have been described in more detail [Vieira et al., 2005]. Clinical aspects of sample collection have been previously described [Murray et al., 1997; Schultz et al., 2004]. The University of Iowa IRB gave approval for sample collection (approval numbers 9701068, 199804081, and 200003065) in conjunction with local approval in the Philippines.

Conserved Non-Coding Elements Selection

CNEs (2–10) were selected for the following genes: *FGF2*, *FGF3*, *FGF4*, *FGF7*, *FGF8*, *FGF9*, *FGF10*, *FGF18*, *FGFR1*, *FGFR2*, *FGFR3* (see the online Supplementary Table I at http://www.interscience.wiley.com/jpages/1552-4825/suppmat/index.html). Regions in the introns of the gene and up to 140 kb outside the gene were chosen using VISTA genome browser and the UCSC genome browser (see the online Supplementary Figs. 1–11 at http://www.interscience.wiley.com/jpages/1552-4825/suppmat/index.html for Vista plots). Regions were first selected based on greater than 50% conservation between human and frog. Next, regions were selected that had greater than 80% conservation between human and mouse and were longer than 100 bp in length. Finally, 3′- and 5′-UTRs were selected (FGFR2 and FGFR3) that were not fully sequenced in the original coding region sequencing project. A total of 52 regions were sequenced.

Sequencing

Cycle sequencing was performed in a 10 μl reaction using 0.25 μl of ABI Big Dye Terminator sequencing reagent (version 1.1), 0.5 μl of 5 μM sequencing primer, 0.5 μl DMSO, 1 μl of 5X buffer, and 6.75 μl of ddH_2O. Primers were designed from public sequence and are available on the Murray lab web site at http://genetics.uiowa.edu. Following a denaturation step at 96° for 30 sec, reactions were cycle sequenced at 96°C for 10 sec, 55°C for 5 sec, and 60°C for 4 min for 40 cycles. Cleanup with magnetic beads by standard protocols and injected on an Applied Biosystems 3730 capillary sequencer.

Sequence Analysis

Chromatograms were transferred to a Unix workstation (Sun Microsystems, Inc., Mountain View, CA), base called with PHRED (version 0.961028), assembled with PHRAP (version 0.960731), scanned by POLYPHRED (version 0.970312) and the results viewed with the CONSED program (version 4.0) [Nickerson et al., 1997].

RESULTS

Fifty-two CNEs were sequenced encompassing approximately 30,000 nucleotides of sequence in 184 NS CLP patients. Fifty-five novel variants that have not been previously reported in dbSNP were identified (listed in the online Supplementary Table II at http://www.interscience.wiley.com/jpages/1552-4825/suppmat/index.html) as well as 33 previously reported SNPs. Seven of the novel variants (found in five different CNEs) are conserved across eight species or more and are highlighted in bold in Table I. The five regions containing the seven highly conserved variants were chosen for additional sequencing in 184 controls from the Philippines (90) and a European ancestral population (94). Four of the seven variants were found in only one case and no controls, and the other three variants were found in multiple cases and controls.

Of the 55 novel variants, only one (FGF2 CNE 5 C > T) falls into a transcription factor binding site (FOXO1A) identified by the UCSC genome browser in the HMR Conserved Transcription Factor Binding Sites track. This track contains the location and score of transcription factor binding sites conserved in the human/mouse/rat alignment. The C > T nucleotide change in the FOXO1A binding site was found in 8 unrelated Iowa NS CLP cases and 4 of 159 controls (all 4 are Caucasian).

Additional analyses using the Genomatix MatInspector program (http://www.genomatix.de) were performed to search for transcription factor binding sites disrupted or created by the novel variants. These analyses predict 14 of the variants create new transcription factor binding sites, 16 of the variants

TABLE I. Novel Variants Conserved Across Six Species or More

Gene	CNE	Variant	Population	Case frequency	Control frequency	UCSC genome browser position[a]	Base pair conservation
FGF2	2	G > C	Iowa	6/167		Chr 4:123,972,190	M, R, Ra, D, A, E
	3	G > A	Iowa	7/171	13/176	Chr 4:123,981,273	M, R, Ra, D, A, E
	***3**	**G > T**	**Iowa**	**1/171**	**0/176**	**Chr 4:123,981,030**	**M, R, Ra, D, A, E, O, Xt, T**
	4	G > A	Iowa	1/177		Chr 4:123,995,377	M, R, D, A, E, O
	5	C > T	Iowa	8/158	4/159	Chr 4:124,025,315	M, R, Ra, D, E, A
FGF7	2	G > T	Phil	1/140		Chr 15:47,472,855	M, R, Ra, D, A, E, O
	2	A > T	Phil	2/140		Chr 15:47,473,045	M, R, Ra, D, A, E, O
	3	T > C	Phil	19/142		Chr 15:47,490,373	M, R, D, A, O, Xt
	4	G > T	Phil	2/149		Chr 15:47,502,707	Ra, D, A, E, O C, Xt
FGF8	2	A > G	Phil	1/164		Chr 10:103,523,734	M, R, Ra, D, O, C, Xt
	2	G > A	Iowa	1/164		Chr 10:103,523,419	M, R, Ra, O, C, Xt
FGF9	3	T > C	Phil	1/175		Chr 13:21,174,985	M, Ra, D, A, E, O, C
FGF10	***1**	**G > A**	**Phil**	**3/173**	**6/174**	**Chr 5:44,427,863**	**M, R, Ra, D, E, O, C, Xt**
	***1**	**A > G**	**Iowa**	**1/173**	**0/174**	**Chr 5:44,427,841**	**M, R, Ra, D, E, O, C, Xt**
	***1**	**C > T**	**Iowa**	**4/173**	**3/174**	**Chr 5:44,427,983**	**M, R, D, E, O, C, Xt, T**
	***3**	**G > C**	**Iowa**	**1/181**	**4/29**	**Chr 5:44,340,627**	**M, R, Ra, D, A, E, O, C, Xt**
FGFR2	1	T > C	Iowa	5/102		Chr 10:123.458,864	M, R, Ra, D, A, O
	1	G > A	Phil	8/102		Chr 10:123,458,555	M, R, A, E, O, C
	***5**	**C > A**	**Iowa**	**1/145**	**0/154**	**Chr 10:123,345,664**	**M, R, Ra, D, A, E, O, C**
	***8**	**G > A**	**Phil**	**1/152**	**0/155**	**Chr 10:123,274,030**	**M, R, Ra, D, A, E, O, C**
	10	A > G	Iowa	2/151		Chr 10:123,229,037	M, R, Ra, D, A, E
FGFR3	2	C > T	Iowa	1/129		Chr 4:1,772,043	M, R, D, A, E, C

Those variants highlighted in bold and indicated with an asterisk are conserved across eight species or more.
M, mouse; R, rat; Ra, rabbit; D, dog; A, armadillo; E, elephant; O, opossum; C, chicken; Xt, X. tropicalis; T, tetradon.
[a]UCSC genome browser position from the human March 2006 assembly.

remove transcription factor binding sites, and 3 of the variants change a binding site to a new binding site. Of the predicted transcription factor binding sites affected by the novel variants, eight are implicated in craniofacial development or the FGF signaling pathway: CBFA1/2, ARNT, PAX3, SOX5, STAT, MSX1/2, DLX, and MAZR. These transcription factors are highlighted in bold in Table II.

To further characterize possible functions of the novel variants, each variant was examined for presence in a human miRNA target site using the Memorial Sloan Kettering Human miRNA Targets—Search and View (http://cbio.mskcc.org/cgi-bin/mirnaviewer/mirnaviewer.pl). This online program creates a track on the UCSC genome browser. None of the identified variants were located in predicted miRNA target sites.

DISCUSSION

Mutation searches in human disease should include both coding regions of genes and neighboring non-coding elements [Emison et al., 2005]. We have thoroughly searched both the exons and non-coding conserved elements of 12 *FGF* and *FGFR* genes for mutations in NS CLP patients. Although many novel mutations in the exons have been identified in patients with NS CLP, these are likely rare mutations that explain, or partially explain, the clefting phenotype in those individuals. The results from CNE sequencing also identified many novel mutations, 31 of which are located in transcription factor binding sites (Table II). The remaining 25 novel variants should not be excluded from playing a role in etiology as these variants may be located in elements whose functional consequences that have not yet been characterized.

Of the 31 variants that alter transcription factor binding sites, 8 are located in binding sites important for craniofacial development. The change of CBFA1 to a CBFA2 site in the *FGF2* CNE2 may be important as mutations in *CBFA1* (OMIM 600211) cause cleidocranial dysplasia (defects in the development of the skull and collar bone) and can be associated with cleft palate. A change such as this may affect the regulation of the *FGF2* gene in the craniofacial region. ARNT (AhR nuclear translocator) has been significantly associated with non-syndromic clefting [Kayano et al., 2004]. Mutations in *PAX3* cause craniofacial-deafness-hand syndrome and it has been demonstrated that *Pax3* is a downstream target of *Fgf* signaling, suggesting a potential feedback loop [Asher et al., 1996; Firnberg and Neubuser, 2002; Monsoro-Burq et al., 2005]. Knockout mice for *Sox5* have cleft palate and it has been shown that SOX proteins regulate the expression of *Fgf* genes [Smits et al., 2001; Luster and Rizzino, 2003; Murakami et al., 2004]. The signal transducer and activator of transcription (STAT) proteins are downstream regulators of the FGF signaling pathway. Mutations in both *MSX1* and *MSX2* have been identified in NS CLP patients and are critical regulators of craniofacial development [van den Boogaard et al., 2000; Jezewski et al., 2003; Vieira et al., 2005]. The *DLX* genes are involved in craniofacial development and knockout mice for *Dlx1*, *Dlx2*, *Dlx1/Dlx2*, *Dlx5*, and *Dlx5/Dlx6* all have cleft palate in a proportion of the embryos [Qiu et al., 1997; Depew et al., 1999; Robledo et al., 2002]. The MAZR protein pairs with BACH2 to bind DNA and activate transcription; this pair is known to bind upstream of the *FGF4* gene and may regulate other *FGF* and *FGFR* genes [Kobayashi et al., 2000].

The seven variants that were highly conserved between eight species or more were also sequenced in control individuals and four of the seven variants which were found in only one case and were not found in the controls sequenced (Table I). The other three variants were found in both cases and controls. It is interesting to note that four of the highly conserved novel variants were identified in the *FGF10* gene and association studies for *FGF10* had borderline significant results with a *P*-value of 0.02 [Riley et al., 2007].

Kryukov et al. [2005] suggest that most individual variants in conserved non-coding regions are only slightly deleterious and do not have a large effect on fitness. However, they also propose that cumulatively, these CNE variants may have a significant impact on the fitness of the individual [Kryukov et al., 2005]. Of the highly conserved ($\geq$8 species) CNE variants and the CNE variants located in transcription factor binding sites implicated in craniofacial development, we find only two individuals with multiple CNE variants (Iowa-5 and Iowa-10, Table III). The Iowa-10 individual has two CNE variants in transcription factor binding sites (MAZR and FOXO1A) and the Iowa-5 individual has three CNE variants in transcription factor binding sites (PAX3, BARX2, and FOXO1A). Expanding the data to include all CNE variants identifies many individuals with multiple accumulated variants; however, it is difficult to determine if this accumulation has an affect on the fitness of the individual without functional testing. Further studies investigating the accumulation of CNE variants in individuals with NS CLP will help to determine the component of contributions of this class of variant to this common, complex disease.

Next, it has been suggested that coding and non-coding mutations work synergistically [Emison et al., 2005]. Table III lists 5 Filipino and 10 Iowa individuals that have both CNE variants and coding variants. While this is consistent with the hypothesis that coding and non-coding variants may work synergistically, functional testing is needed to accurately predict their effects.

Currently, it is not possible to discriminate between those mutations that contribute to disease and those

TABLE II. Predicted Transcription Factor Binding Sites

Gene	CNE	Variant	UCSC genome browser position[a]	UCSC HMR conserved transcription factor binding sites	Genomatix MatInspector prediction[b] created binding sites	Genomatix MatInspector prediction[b] removed binding sites
FGF2	2	G > C	Chr 4:123,971,990		**Changes CBFA1 (runt-related trx factor) to a CBFA2**	Removes INSM1 (zinc finger protein insulinoma-associated trx repressor)
	2	G > C	Ch r4:123,972,190		E4F (GLI-Krueppel regulator of adenovirus E4 promoter)	
	3	G > A	Chr 4:123,981,273		RREB (Ras responsive element binding protein)	
	4	G > A	Chr 4:123,995,377		MEL1 (MDS1/EVI1-like gene) DNA binding domain 2	
	4	C > A	Chr 4:123,995,164		ATBF1 (AT binding trx factor 1)	
	5	C > T	Chr 4:124,025,315	FOXO1A		
FGF3	1	A > G	Chr 11:69,346,124			TCF11/LCRF1/Nrf1
	1	T > C	Chr 11:69,345,933			IK2 (Ikaros 2)
FGF7	2	G > T	Chr 15:47,472,855			CREB (cAMP responsive element site)
	2	A > T	Chr 15:47,473,045		Changes MTBF (muscle specific MT binding site) to Lentivirus LTR TATA box	
FGF7	4	G > T	Chr 15:47,502,707		Changes CDP (trx repressor) and HNF1 (hepatic nuclear factor 1) to BCL6 (zinc finger trx repressor)	
FGF8	2	A > G	Chr 10:103,523,734			**ARNT (AhR nuclear translocator)**
	2	T > A	Chr 10:103,523,780		E2F (involved in cell cycle regulation)	
	2	G > A	Chr 10:103,523,419		**PAX3 (paired domain protein)**	
FGF9	3	T > C	Chr 13:21,174,985			OCT1, SOX5, and GCMF (glial chorionspecific trx factor)
FGF10	1	C > T	Chr 5:44,427,983			**STAT (signal transducer and activator of trx)**
FGFR1	1	C > A	Chr 8:38,448,960			CKROX (collagen krox protein)
	5	T > C	Chr 8:38,473,226			E2F
	6	A > G	Chr 8:38,449,599		OCT1	
FGFR2	1	G > A	Chr 10:123,458,555		**MSX1/2 binding sites**	GTF2I repeat 4 (TFII-IR4)
	3	G > T	Chr 10:123,365,553		SMARCA3 (SWI/SNF related)	
	3	C > T	Chr 10:123,365,264			NRSE (neural restrictive silencer element)
	5	C > A	Chr 10:123,345,664		GKLF (Gut enriched Krueppel-like factor) and NKX5-1	PSE/SNAP (proximal sequence element)
	8	G > A	Chr 10:123,274,030		DMTF1 (cyclin D-interacting myblike protein)	
	10	A > G	Chr 10:123,229,037			S8 homeodomain proteins (**DLX** and **MSX**), NKX2.5 and BARX2
FGFR3	3	C > T	Chr 4:1,766,606		NMP4 (nuclear matrix protein 4)	
	3	C > T	Chr 4:1,769,420			HIC1 (hypermethylated in cancer)
	3	C > G	Chr 4:1,769,576			WHN (winged helix protein)
	6	G > A	Chr 4:1,759,904			DMTF1 and ZAS domain trx factor (NFKB/HIVEP1)
	10	A > G	Chr 4:1,732,481			Meisb and Hoxa9
	10	C > T	Chr 4:1,732,679			WT1 (Wilms tumor suppressor) and **MAZR (myc associated zinc finger protein)**

Transcription factor binding sites highlighted in bold are implicated in craniofacial development. Only CNEs listed have identified transcription factor binding sites.
[a]UCSC genome browser position from the human March 2006 assembly.
[b]http://www.genomatix.de.

American Journal of Medical Genetics Part A: DOI 10.1002/ajmg.a

TABLE III. Individuals With Multiple FGF/FGFR CNE Variants and Coding Variants

Individual	Phenotype	Variant	Variant	Variant	Variant
Phil-1	BCLP	FGF8 CNE2 A > G (ARNT)	**FGF3 G30R**		
Phil-2	BCLP	FGF9 CNE3 T > C (SOX5)	**FGF18 P138P**		
Phil-3	BCLP	FGFR2 CNE1 G > A (MSX1/2)	**FGF18 P138P**		
Phil-4	UCLP	FGFR2 CNE1 G > A (MSX1/2)	**FGFR1 S107L**		
Phil-5	BCLP	FGFR2 CNE8 G > A (DMTF1)	**FGFR3 V329I**		
Iowa-1	UCLP	FGF2 CNE2 G > C (CBFA1/2)	**FGFR1 D200D**		
Iowa-2	BCLP	FGF2 CNE2 G > C (CBFA1/2)	**FGFR1 R609X**	**FGF4 A98A**	
Iowa-3	? CLP	FGF2 CNE2 G > C (CBFA1/2)	**NUDT6 K172N**		
Iowa-4	? CLP	FGF2 CNE2 G > C (CBFA1/2)	**FGF3 A23A**		
Iowa-5	BCLP	FGF8 CNE2 G > A (PAX3)	FGFR2 CNE10 A > G (BARX2)	FGF2 CNE5 C > T (FOXO1A)	**FGFR1 L754L**
Iowa-6	UCLP	FGF10 CNE1 C > T (STAT)	**FGF3 A23A**		
Iowa-7	BCLP	FGF10 CNE1 C > T (STAT)	**FGF18 P138P**	**FGF3 A23A**	
Iowa-8	? CLP	FGF10 CNE1 C > T (STAT)	**FGF3 A23A**		
Iowa-9	? CLP	FGFR2 CNE5 C > A (PSE, GKLF, NKX)	**FGF3 A23A**		
Iowa-10	BCLP	FGFR3 CNE10 C > T (MAZR)	FGF2 CNE5 C > T (FOXO1A)	**FGF3 A23A**	

Overlapping variants include: highly conserved (bold), located in transcription factor binding sites involved in craniofacial development (gray), coding variants (black). BCLP, bilateral cleft lip and palate; UCLP, unilateral cleft lip and palate; ? CLP, unknown laterality CLP.

that are normal polymorphisms without testing for function. We can make generalizations based on frequency in case versus control populations, however, this criterion may not be appropriate for a significant fraction of mutations because of penetrance and gene–gene interaction issues that must be accounted for in complex diseases such as CLP. In the future it will be necessary to design an assay to screen these CNE variants for functionality. Most recently, Fisher et al. [2006] developed an efficient method to test CNE function using a transposon-based transgenic assay in which the CNE can be screened for control of GFP expression in vivo in the zebrafish [Fisher et al., 2006]. An extension of this method would be to compare the wild-type CNE expression to the mutant CNE expression to determine if the identified variants alter gene expression.

ACKNOWLEDGMENTS

We would like to thank all of the families who participated and the many Operation Smile and HOPE volunteers who make this work possible. Thanks also to Brenna Trump and Lisa Raffensperger who helped perform some of the sequencing. And a special thanks to our lab administrators Susie McConnell, Melanie DeVore, and Dan Benton. This research was supported by NIH grants ES1087, R37-DE08559, P50-DE016215, and 5T32GM008629; and by support from the Center for Inherited Disease Research CIDR NIH contract N01-HG-65403.

REFERENCES

Asher JH Jr, Sommer A, Morell R, Friedman TB. 1996. Missense mutation in the paired domain of PAX3 causes craniofacial-deafness-hand syndrome. Hum Mutat 7:30–35.

Depew MJ, Liu JK, Long JE, Presley R, Meneses JJ, Pedersen RA, Rubenstein JL. 1999. Dlx5 regulates regional development of the branchial arches and sensory capsules. Development 126:3831–3846.

Eddy SR. 2005. A model of the statistical power of comparative genome sequence analysis. PLoS Biol 3:e10.

Emison ES, McCallion AS, Kashuk CS, Bush RT, Grice E, Lin S, Portnoy ME, Cutler DJ, Green ED, Chakravarti A. 2005. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. Nature 434:857–863.

Firnberg N, Neubuser A. 2002. FGF signaling regulates expression of Tbx2, Erm, Pea3, and Pax3 in the early nasal region. Dev Biol 247:237–250.

Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS. 2006. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312:276–279.

Jezewski PA, Vieira AR, Nishimura C, Ludwig B, Johnson M, O'Brien SE, Daack-Hirsch S, Schultz RE, Weber A, Nepomucena B, Romitti PA, Christensen K, Orioli IM, Castilla EE, Machida J, Natsume N, Murray JC. 2003. Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. J Med Genet 40:399–407.

Jugessur A, Murray JC. 2005. Orofacial clefting: Recent insights into a complex trait. Curr Opin Genet Dev 15:270–278.

Kayano S, Suzuki Y, Kanno K, Aoki Y, Kure S, Yamada A, Matsubara Y. 2004. Significant association between non-syndromic oral clefts and arylhydrocarbon receptor nuclear translocator (ARNT). Am J Med Genet Part A 130A:40–44.

Kobayashi A, Yamagiwa H, Hoshino H, Muto A, Sato K, Morita M, Hayashi N, Yamamoto M, Igarashi K. 2000. A combinatorial code for gene expression generated by transcription factor Bach2 and MAZR (MAZ-related factor) through the BTB/POZ domain. Mol Cell Biol 20:1733–1746.

Kryukov GV, Schmidt S, Sunyaev S. 2005. Small fitness effect of mutations in highly conserved non-coding regions. Hum Mol Genet 14:2221–2229.

Lidral AC, Moreno LM. 2005. Progress toward discerning the genetics of cleft lip. Curr Opin Pediatr 17:731–739.

Luster TA, Rizzino A. 2003. Regulation of the FGF-4 gene by a complex distal enhancer that functions in part as an enhanceosome. Gene 323:163–172.

Monsoro-Burq AH, Wang E, Harland R. 2005. Msx1 and Pax3 cooperate to mediate FGF8 and WNT signals during Xenopus neural crest induction. Dev Cell 8:167–178.

Murakami A, Shen H, Ishida S, Dickson C. 2004. SOX7 and GATA-4 are competitive activators of Fgf-3 transcription. J Biol Chem 279:28564–28573.

Murray JC, Daack-Hirsch S, Buetow KH, Munger R, Espina L, Paglinawan N, Villanueva E, Rary J, Magee K, Magee W. 1997. Clinical and epidemiologic studies of cleft lip and palate in the Philippines. Cleft Palate Craniofac J 34:7–10.

Nickerson DA, Tobe VO, Taylor SL. 1997. PolyPhred: Automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 25:2745–2751.

Ponting CP, Lunter G. 2006. Signatures of adaptive evolution within human non-coding sequence. Hum Mol Genet 15: R170–R175.

Qiu M, Bulfone A, Ghattas I, Meneses JJ, Christensen L, Sharpe PT, Presley R, Pedersen RA, Rubenstein JL. 1997. Role of the Dlx homeobox genes in proximodistal patterning of the branchial arches: Mutations of Dlx-1, Dlx-2, and Dlx-1 and -2 alter morphogenesis of proximal skeletal and soft tissue structures derived from the first and second arches. Dev Biol 185:165–184.

Riley BM, Mansilla MA, Ma J, Daack-Hirsch S, Maher BS, Raffensperger LM, Russo ET, Vieira AR, Dode C, Mohammadi M, Marazita ML, Murray JC. 2007. Impaired FGF signaling contributes to cleft lip and palate. Proc Natl Acad Sci USA 104:4512–4517.

Robledo RF, Rajan L, Li X, Lufkin T. 2002. The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. Genes Dev 16:1089–1101.

Schultz RE, Cooper ME, Daack-Hirsch S, Shi M, Nepomucena B, Graf KA, O'Brien EK, O'Brien SE, Marazita ML, Murray JC. 2004. Targeted scan of fifteen regions for nonsyndromic cleft lip and palate in Filipino families. Am J Med Genet Part A 125A:17–22.

Smits P, Li P, Mandel J, Zhang Z, Deng JM, Behringer RR, de Crombrugghe B, Lefebvre V. 2001. The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. Dev Cell 1:277–290.

van den Boogaard MJ, Dorland M, Beemer FA, van Amstel HK. 2000. MSX1 mutation is associated with orofacial clefting and tooth agenesis in humans. Nat Genet 24:342–343.

Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, Strausberg RL, Brenner S. 2006. Ancient noncoding elements conserved in the human genome. Science 314:1892.

Vieira AR, Avila JR, Daack-Hirsch S, Dragan E, Felix TM, Rahimov F, Harrington J, Schultz RR, Watanabe Y, Johnson M, Fang J, O'Brien SE, Orioli IM, Castilla EE, Fitzpatrick DR, Jiang R, Marazita ML, Murray JC. 2005. Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. PLoS Genet 1:e64.