ORIGINAL ARTICLE

Contribution of *MSX1* variants to the risk of non-syndromic cleft lip and palate in a Malay population

Iman Salahshourifar[1,2], Ahmad Sukari Halim[2], Wan Azman Wan Sulaiman[2] and Bin Alwi Zilfalil[1]

Oral clefts are clinically and genetically heterogeneous disorders that are influenced by both genetic and environmental factors. The present family-based association study investigated the role of the *MSX1* and *TGFB3* genes in the etiology of non-syndromic oral cleft in a Malay population. No transmission distortion was found in the transmission disequilibrium analysis for either *MSX1*-CA or *TGFB3*-CA intragenic markers, whereas *TGFB3*-CA exhibited a trend to excess maternal transmission. In sequencing the *MSX1* coding regions in 124 patients with oral cleft, five variants were found, including three known variants (A34G, G110G and P147Q) and two novel variants (M37L and G267A). The P147Q and M37L variants were not observed in 200 control chromosomes, whereas G267A was found in one control sample, indicating a very rare polymorphic variant. Furthermore, the G110G variant displayed a significant association between patients with non-syndromic cleft lip, with or without cleft palate, and normal controls ($P=0.001$, odds ratio$=2.241$, 95% confidence interval, 1.357–3.700). Therefore, these genetic variants may contribute, along with other genetic and environmental factors, to this condition.

INTRODUCTION

Oral clefts are clinically and genetically heterogeneous disorders, involving both genetic and environmental factors. *MSX1* and *TGFB3* have emerged as candidate genes for oral cleft, based on expression studies[1,2] and animal models.[3,4] Their involvement is further supported by linkage and association studies in populations of different ethnic origin.[5–8] Although no deleterious mutation has been reported in *TGFB3*,[7,9] *MSX1* has been suggested to function upstream of *TGFB3* in the development of oral cleft.[10]

The aim of the current research was to investigate the genetic association of *MSX1* and *TGFB3* in a Malay population with non-syndromic orofacial cleft. The association study was performed using intragenic CA markers for *MSX1* and *TGFB3*. The coding regions of *MSX1* were also directly sequenced.

MATERIALS AND METHODS

Subjects

A total of 124 Malay families were enrolled in the study by the Reconstructive Sciences Unit of Hospital Universiti Sains Malaysia, Kelantan. Of these families, 90 complete case–parent trios had non-syndromic cleft of the lip, with or without cleft palate (CL/P), 22 incomplete case–parent trios also had non-syndromic CL/P but were missing a sample from at least one parent and 12 families had CP only (CPO). Among all cleft cases, 41 families had a positive family history of oral cleft. Only completed families were included in genotyping of the CA markers, whereas all 124 affected cases were included in direct sequencing. All subjects were prospectively recruited in a consecutive manner since 2008. Our center is a tertiary referral center, and cases with variable types of orofacial clefts were referred to us throughout the Kelantan state. Kelantan is bordered to the north by Narathiwat Province of Thailand, thus, increasing the likelihood of some population mixture. However, 95% of Kelantanese, and the majority of the Narathiwat Province population, have a Malay ethnic background. All patients with a minor anomaly, such as low-set ears, mild hypertelorism, clinodactyly and single palmar crease, were included in the study. Patients identified during clinical assessment by a team of specialists in the fields of craniofacial surgery, pediatrics and genetics to have heart disease, a known syndrome or other major or more than two minor defects were excluded from this study. In total 100 healthy volunteers, with neither cleft nor a family history of cleft, were used as controls. The study was approved by Human Ethics Committee of Universiti Sains Malaysia, and informed consent from subjects was obtained before taking blood samples.

Genotyping of *MSX1*-CA and *TGFB3*-CA repeats

In total, 90 patient–parent trios with non-syndromic CL/P, excluding CP only and uncompleted CL/P families, were included for genotyping of *MSX1*-CA and *TGFB3*-CA repeats. The amplification of both intragenic CA repeats was performed using previously described primers for both *MSX1*-CA[9] and *TGFB3*-CA repeats,[11] with minor modifications. Multiplex-PCR amplification

using fluorescent primers was performed using the Type-It Microsatellite Amplification Kit (QIAGEN, GmbH, Hilden, Germany). The PCR products were run on an ABI 3100 sequencer (Applied Biosystems, Foster City, CA, USA) and analyzed using GeneScan v3.0 (Applied Biosystems). Direct DNA sequencing was conducted to determine the size of the MSX1-CA repeat allele.

Mutation screening of MSX1

All 124 affected individuals with oral cleft (12 cases with CP only and 112 cases with CL/P) were included for mutation screening of MSX1. Specific primers were designed, via the Primer3 program (v. 0.4.0),[12] to cover the coding regions of MSX1. Both exons were amplified by standard PCR using the following primers: exon-1 (forward: 5′-GCCAGTGCTGCGGCAGAA-3′, reverse: 5′-CGC CTGGGTTCTGGCTACTCACT-3′) and exon-2 (forward: 5′-TGATCATGCTCC AATGCTTCT-3′, reverse: 5′-ACCAGGGCTGGAGGAATC-3′). The amplification was carried out in a total volume of $25\,\mu l$, containing $40\,ng$ of genomic DNA, 2% Q-solution (QIAGEN), $0.5\,\mu M$ of each primer, $0.2\,mM$ dNTPs, $2\,mM$ $MgCl_2$, $1.25\,U$ of AmpliTaq Gold polymerase (Applied Biosystems) and $1\times$ PCR buffer supplied by the manufacturer. The Q-solution was applied only for the amplification of exon 1. The PCR conditions were as follows: initial denaturation for $10\,min$ at $95\,°C$; followed by 35 cycles of denaturation at $95\,°C$ for $30\,s$, annealing at $61\,°C$ for $2\,min$ (exon 2) or $64\,°C$ for $2\,min$ (exon 1) and elongation at $72\,°C$ for $1\,min$; and final elongation at $72\,°C$ for $10\,min$.

The PCR products were purified using the QiaQuik PCR purification kit (QIAGEN), sequenced directly using the BigDye terminator V3.1 sequencing standard kit (Applied Biosystems), as recommended by the manufacturer, and run on an automated ABI Prism 3100 Genetic Analyzer (Applied Biosystems). DNA sequences were analyzed using BLAST with reference sequences for MSX1 cDNA (NM_002448.3) and genomic DNA (NT_006051.18). The samples with variations were resequenced with both forward and reverse primers.

Bioinformatics analysis

The potential consequence of each variant on the MSX1 protein structure was predicted by PolyPhen (http://www.bork.embl-heidelberg.de/PolyPhen). The multiple alignments were performed using the ClustalX program version 2.0.12 (University College Dublin, Dublin, Ireland). The Homo sapiens MSX1 gene protein (NP_002439) sequence was aligned with that of the chimpanzee (Pan troglodytes; XP_517087.2), pig (Sus scrofa; NP_001156359.1), mouse (Mus musculus; NP_034965.2), rat (Rattus norvegicus; NP_112321.1), cow (Bos taurus; NP_777223.1) and chicken (Gallus gallus; NP_990819.1). The ESE finder program was used to predict exonic splicing enhancers (http://rulai.cshl.edu/tools/ESE2/).

Statistical analysis

The χ^2-test, odds ratio (OR) and 95% confident interval (CI) were calculated using an online program, available at http://statpages.org/, to compare results between affected patients and controls. The standard transmission disequilibrium test and linkage disequilibrium between markers were computed using the FAMHAP program version 19 beta (University of Bonn, Bonn, Germany). Subsequently, we used a log-linear model to estimate the single- and double-dose effects of alleles using the HAPLIN program, version 3.5.[13] This software uses a previously introduced model[14] to compute both the relative risk (RR) and the likelihood ratio. We used a reciprocal reference with the threshold set to 0.01 (1%) for haplotype frequency.

RESULTS

A total of 90 trios with non-syndromic CL/P were analyzed for CA repeats in the MSX1 and TGFB3 genes. Both markers were found to be in Hardy–Weinberg equilibrium ($P > 0.05$). Five alleles were identified for MSX1-CA, including: allele 1, 12 CA repeats (190 bp); allele 2, 11 CA repeats (188 bp); allele 3, 10 CA repeats (186 bp); allele 4, 9 CA repeats (184 bp); and allele 5, 8 CA repeats (182 bp). Allele 4 of the MSX1-CA repeat was the most common. TGFB3-CA was less polymorphic: three alleles containing 163, 165 or 167 base pairs were identified. The results of transmission disequilibrium test did not indicate preferential transmission for either the MSX1-CA or the TGFB3-CA markers. When the analysis

was restricted to the maternal or paternal transmission, the TGFB3-CA exhibited a trend toward maternal transmission ($P = 0.058$). The 163-bp allele was overtransmitted ($P_{value} = 0.04$, $OR_{transmission} = 2.1$), whereas the 165-bp allele was undertransmitted ($P_{value} = 0.06$, $OR_{transmission} = 0.52$) from heterozygous mothers to the affected patients. A maternal double dose of the 163-bp allele was shown to have a slightly higher RR (RR = 1.18, 95% CI, 0.534–2.61) compared with the 165-bp allele (RR = 0.843, 95% CI, 0.39–1.89), but this difference did not reach statistical significance. Furthermore, a fetal double dose of the 163-bp allele (RR = 1.81, 95% CI, 0.814–4.04) demonstrated a higher RR compared with the 165-bp allele (RR = 0.553, 95% CI, 0.249–1.26), consistent with a maternal genotype effect.

In a total screening of 124 affected individuals with oral cleft (112 CL/P and 12 CP only cases), five variants were found in the coding regions of MSX1. Of these variants, two common variants of 101C > G (A34G) and 330C > T (G110G) in exon 1 have been previously reported and were observed in both affected patients and controls (Table 1). The difference in allele frequency was not significant for 101C > G ($P = 0.813$), whereas the 330C > T variant showed a significant association with cleft ($P = 0.001$, OR = 2.241, 95% CI, 1.357–3.700). The A34G was also found to be heterozygous in eight control samples (Table 1).

Of the remaining MSX1 variants, three heterozygous, non-synonymous variants, 440C > A (P147Q), 109A > C (M37L) and 800G > C (G267A), were found. With the exception of the G267A variant being found in one control sample, these variants were not observed in 200 control chromosomes.

The patient who harbored the P147Q variant had right unilateral cleft lip with CP and mild hypertelorism. Interestingly, similar to the proband, his otherwise healthy, non-cleft sister had mild hypertelorism. The P147Q variant was not found in this sister; however, his other non-cleft sister, without hypertelorism, was heterozygous for this variant. This observation is not surprising, given that disruption of normal facial development is common among cases with orofacial clefts, and hypertelorism is commonly present to some degree in both affected cases and their non-cleft relatives.[15,16] This variant was inherited heterozygously from the non-cleft father who had neither dental anomaly nor other subphenotypes of cleft. Dental anomalies and other subphenotypes of cleft were not found in either the proband or family members.

The M37L variant was found in a patient with bilateral CLP without additional anomalies. This variant was segregated from a non-cleft father who was homozygous for this variant. In this family,

Table 1 Comparison of specifies alleles of 101C > G and 330C > T among 112 cases with isolated CL/P and 100 normal controls

Variant	Allele 1	Allele 2	χ^2-test (d.f. 1) (P-value)	Odds ratio (OR)
101C > G (A34G)	C	G		
Cases	214	10	0.056 (0.813)	0.892 (95% CI, 0.355–2.239)
Controls	192	8		
330C > T (G110G)	C	T		
Cases	195	29	**10.13 (0.001)**	2.241 (95% CI, 1.357–3.700)
Controls	150	50		

Abbreviation: CI, confidence interval; CL/P, with or without cleft palate; d.f., degree of freedom. This table shows the frequency of the variants among 200 control chromosomes and 224 case chromosomes. The 330C > T indicated a significant association (in bold).

the grandfather's siblings had oral cleft from the paternal side, but samples were not available for tracing the variant. Upon medical examination of the father, neither dental abnormality nor other subphenotypes of cleft were found.

The patient with G267A had incomplete cleft lip, without other associated dysmorphic features. The variant was inherited from a heterozygous non-cleft father who had a high-arched palate and torus palatinus. There was no family history of cleft.

DISCUSSION

Several studies using *MSX1*-CA and *TGFB3*-CA markers have found significant associations with non-syndromic orofacial cleft,[7,17,18] whereas others have found no significant association.[8,11,19,20] In the present research, no transmission distortion was found in the transmission disequilibrium analysis for either *MSX1*-CA or *TGFB3*-CA intragenic markers, but *TGFB3*-CA exhibited a trend to excess maternal transmission. Our findings suggest that both the maternal and fetal RR results are in agreement for both *TGFB3*-CA alleles, as the 163-bp allele had a deleterious recessive effect, whereas the 165-bp allele had a protective recessive effect. The present results are consistent with a previously report that found a parent of origin effect for the *TGFB3* gene among central European populations.[21]

Upon sequencing the *MSX1* coding regions in 124 patients with oral cleft, five variants were found, including three known variants (A34G, G110G and P147Q) and two novel variants (M37L and G267A) as summarized in Table 2. Furthermore, in agreement with a previous report on Asian populations,[22] the G110G variant displayed a significant association between patients with non-syndromic cleft lip with or without CP and normal controls.

In addition to its well-known homeodomain, six other highly conserved domains have recently been identified within the human *MSX1* protein.[23] It appears that amino-acid changes within the homeodomain lead to the development of tooth agenesis with high penetrance and a dominant mode of inheritance. Almost all reported mutations within the homeodomain to date have been detected in cases with tooth agenesis.[23,24] In addition, orofacial clefts have been seen as associated anomalies among cases with tooth agenesis bearing an *MSX1* mutation.[25] Therefore, it appears that *MSX1* has an overlapping role in tooth, lip and palate development. In spite of a significant association between non-syndromic orofacial clefts and *MSX1*, previous studies have not identified clear Mendelian inheritance or a genotype–phenotype correlation.[17,22,26] In contrast to the tooth agenesis, most of the *MSX1* mutations that have been identified among non-syndromic clefts cases are found outside the most highly conserved homeodomain.[17,22,26]

Similar to other studies, we have found three variants without clear Mendelian inheritance in single-cleft cases. Of these variants, only the P147Q variant, which has been detected in a number of Southeast Asian populations, is a relatively common variation in the *MSX1* gene. In addition to the present study, this variant has been detected in three Vietnamese families,[17] two Filipino families[9] and three Thai families.[26] The prevalence of this variant may reflect a Founder effect in these populations. The P147Q variant was previously reported as a damaging variant;[9,17] however, evidence from a Thai population suggested that it may have no disease-causing effect because it was detected in controls.[26] Although the population admixture of our subjects with the Thai population was expected, we did not detect the P147Q variant in 200 control chromosomes. There is no clear correlation between genotype and phenotype of individuals who are carrying the P147Q variant. Consistent with the present study, it has been detected across affected and non-affected members of the same family and in unrelated healthy controls.[9,17,26] Although this variant is located within a conserved phosphorylation motif within the MH3 domain,[23] it is revealed that it may not be pathogenic.[26] Therefore, it, at best, may contribute to clefting with other genetic and environmental risk factors in a complex inheritance pattern.

In the present study, the M37 amino acid was substituted with leucine (another hydrophobic amino acid). Methionine is a very hydrophobic amino acid and is conserved at position 37, across species as far back as amphibians.[23] Although this variant was predicted to be benign using the Polyphen program, it is located exactly one amino acid after the polyalanine tract of *MSX1* at the N-terminal, from nucleotides 29 to 36 (29-AAAATAAA-36). Polyalanine tracts are found in the homeodomain of many transcription factors,[27] and have been shown to be associated with several human disorders.[28] Current evidence suggests that polyalanine tracts function by changing the normal folding, degradation and cytoplasmic aggregation of the mutant proteins.[28] Both alanine repeats and the flanking amino acids are conserved, suggesting that both are important to the function of the polyalanine tract.[28] Furthermore, the alanine at position 34 is highly conserved and located inside the polyalanine tract (29-AAAATAAA-36) (Figure 1). Similar to M37, its conservation also extends back to the amphibians.[23] The A34G variant was observed in both cases and controls as a non-pathogenic variant in populations of different ethnic background.[17,22,26] However, the M37L variant was not found in 200 control chromosomes. Furthermore, the non-cleft father who was homozygous for this variant did not exhibit subphenotypes of cleft or dental anomaly. Therefore, the M37L variant appears to be very rare and may be detectable across a large number of subjects. Most *MSX1* variants have been found in only single CL/P cases, and therefore, a large control group would be needed to obtain significant results.[24]

As in the Thai study,[26] we also found an amino-acid substitution at the position 267 of the MSX1 protein. In the previous study, the glycine at this position was substituted with cysteine (G267C).[26] This 267 variant was predicted to be possibly damaging in the Thai study but to be benign in the present research. This difference may be due to the fact that cysteine is a more hydrophobic amino acid than glycine and alanine. This amino acid is not highly conserved but is located in a block that is conserved among species (Figure 1). The patient

Table 2 Summarized clinical and genetic data of patients with *MSX1* variant

Exon	Patient's genotype		Patient's phenotype	Inheritance	Father's genotype	Father's phenotype
1	109A>C Het	M37L	Bilateral CLP	Paternal	Homo	No cleft, dental anomaly and other subphenotypes
1	440C>A Het	P147Q	Unilateral CLP, mild hypertelorism	Paternal	Het	No cleft, dental anomaly and other subphenotypes
2	800G>C Het	G267A	Incomplete CL	Paternal	Het	No cleft and dental anomaly, having high arched palate and torus palatinus

Abbreviations: CL, cleft lip; CLP, cleft lip with cleft palate; Het, heterozygous; Homo, homozygous.

Species	Protein accession no.	A34G M37L		P147Q	G267A
Homo sapiens	NP_002439.2	SAAAAT**A**AA**M**GADEEGAKPKVS		RTPWMQSPRFS**P**PPARRL	VAAAAGASLY**G**ASGPFQRAAL
Pan troglodytes	XP_517087.2	SAAAAT**A**AA**M**GADEEGAKPKVS		RTPWMQSPRFS**P**PPARRL	VAAAAGASLY**G**ASGPFQRAAL
Sus scrofa	NP_001156359	STTAAT**A**AA**M**GADEEGAKPKVS		RTPWMQNPRFS**P**PPARRL	VAAAAGASLY**G**ASGPFQRAAL
Mus musculus	NP_034965.2	GAAAAT**A**TA**M**GTDEEGAKPKVP		RTPWMQSPRFS**P**PPARRL	VAAAAGASLY**S**ASGPFQRAAL
Rattus norvegicus	NP_112321.1	GAAAAT**A**TA**M**GTDEEGAKPKVP		RTPWMQSPRFS**P**PPARRL	VAAAAGASLY**S**ASGPFQRAAL
Bos taurus	NP_777223.1	STTAAT**A**AA**M**GADAEGAKPKVS		RTPWMQNPRFS**P**PPSRRL	VAAPAGASLY**G**ASGPFQRAAL
Gallus gallus	NP_990819.1	---LPV**A**AA**M**GGEEESDKPKVS		RSPWMQSPRFS**P**PPPRRL	----AGASLY**G**ASSPFQRAGL

Figure 1 The multi-alignment of the human *MSX1* protein with its orthologs among various species, generated by using the ClustalX program version 2.0.12. The methionine at position 37 is not located in a highly conserved block, although it is conserved among different species. The glycine at position 267 is less conserved but is located in a conserved block of amino acids. The proline at position of 147 is located in a highly conserved block and is highly conserved among different species.

with this variant had a mild form of cleft and no dental abnormality. The father of the patient had a slightly high-arched palate and torus palatinus, but not cleft lip or dental anomalies. Furthermore, this variant was found in one control sample, suggesting a very rare polymorphic site.

In conclusion, with the exception of the P147Q variant, several single, rare variants have been reported in either families or specific populations with oral clefts.[17,22,26,29] Oral clefts are more severe than tooth agenesis, and recent data suggest that *MSX1* causative mutations for tooth agenesis may not be sufficient to cause oral clefts.[29] In addition, the change in amino acid itself may not be a sufficient criterion to predict disease in complex traits such as oral clefts.[24] Therefore, these rare variants could, at best, contribute to clefting as part of a complex inheritance pattern, with both additional genes and environmental factors having a role.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

1 Nugent, P. & Greene, R. M. MSX-1 gene expression and regulation in embryonic palatal tissue. *In. Vitro. Cell. Dev. Biol. Anim.* **34,** 831–835 (1998).
2 Bodo, M., Baroni, T., Carinci, F., Becchetti, E., Bellucci, C., Pezzetti, F. *et al.* TGFbeta isoforms and decorin gene expression are modified in fibroblasts obtained from non-syndromic cleft lip and palate subjects. *J. Dent. Res.* **78,** 1783–1790 (1999).
3 Satokata, I. & Maas, R. Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat. Genet.* **6,** 348–356 (1994).
4 Kaartinen, V., Voncken, J. W., Shuler, C., Warburton, D., Bu, D., Heisterkamp, N. *et al.* Abnormal lung development and cleft palate in mice lacking TGF-beta 3 indicates defects of epithelial-mesenchymal interaction. *Nat. Genet.* **11,** 415–421 (1995).
5 Moreno, L. M., Arcos-Burgos, M., : Marazita, M. L., Krahn, K., Maher, B. S., Cooper, M. E. *et al.* Genetic analysis of candidate loci in non-syndromic cleft lip families from Antioquia-Colombia and Ohio. *Am. J. Med. Genet. A.* **125A,** 135–144 (2004).
6 Schultz, R. E., Cooper, M. E., Daack-Hirsch, S., Shi, M., Nepomucena, B., Graf, K. A. *et al.* Targeted scan of fifteen regions for nonsyndromic cleft lip and palate in Filipino families. *Am. J. Med. Genet. A.* **125A,** 17–22 (2004).
7 Lidral, A. C., Romitti, P. A., Basart, A. M., Doetschman, T., Leysens, N. J., Daack-Hirsch, S. *et al.* Association of MSX1 and TGFB3 with nonsyndromic clefting in humans. *Am. J. Hum. Genet.* **63,** 557–568 (1998).
8 Jugessur, A., Lie, R. T., Wilcox, A. J., Murray, J. C., Taylor, J. A., Saugstad, O. D. *et al.* Variants of developmental genes (TGFA, TGFB3, and MSX1) and their associations with orofacial clefts: a case-parent triad analysis. *Genet. Epidemiol.* **24,** 230–239 (2003).
9 Vieira, A. R., Avila, J. R., Daack-Hirsch, S., Dragan, E., Felix, T. M., Rahimov, F. *et al.* Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. *PLoS. Genet.* **1,** e64 (2005).
10 Vieira, A. R. Unraveling human cleft lip and palate research. *J. Dent. Res.* **87,** 119–125 (2008).
11 Lidral, A. C., Murray, J. C., Buetow, K. H., Basart, A. M., Schearer, H., Shiang, R. *et al.* Studies of the candidate genes TGFB2, MSX1, TGFA, and TGFB3 in the etiology of cleft lip and palate in the Philippines. *Cleft. Palate. Craniofac. J.* **34,** 1–6 (1997).
12 Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods. Mol. Biol.* **132,** 365–386 (2000).
13 Gjessing, H. K. & Lie, R. T. Case-parent triads: estimating single- and double-dose effects of fetal and maternal disease gene haplotypes. *Ann. Hum. Genet.* **70,** 382–396 (2006).
14 Weinberg, C. R., Wilcox, A. J. & Lie, R. T. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am. J. Hum. Genet.* **62,** 969–978 (1998).
15 Yoon, Y. J., Perkiomaki, M. R., Tallents, R. H., Barillas, I., Herrera-Guido, R., Fong, C. T. *et al.* Transverse craniofacial features and their genetic predisposition in families with nonsyndromic unilateral cleft lip and palate. *Cleft. Palate. Craniofac. J.* **41,** 256–261 (2004).
16 Weinberg, S. M., Naidoo, S. D., Bardi, K. M., Brandon, C. A., Neiswanger, K., Resick, J. M. *et al.* Face shape of unaffected parents with cleft affected offspring: combining three-dimensional surface imaging and geometric morphometrics. *Orthod. Craniofac. Res.* **12,** 271–281 (2009).
17 Suzuki, Y., Jezewski, P. A., Machida, J., Watanabe, Y., Shi, M., Cooper, M. E. *et al.* In a Vietnamese population, MSX1 variants contribute to cleft lip and palate. *Genet. Med.* **6,** 117–125 (2004).
18 Vieira, A. R., Orioli, I. M., Castilla, E. E., Cooper, M. E., Marazita, M. L. & Murray, J. C. MSX1 and TGFB3 contribute to clefting in South America. *J. Dent. Res.* **82,** 289–292 (2003).
19 Blanco, R., Suazo, J., Santos, J. L., Paredes, M., Sung, H., Carreno, H. *et al.* Association between 10 microsatellite markers and nonsyndromic cleft lip palate in the Chilean population. *Cleft. Palate. Craniofac. J.* **41,** 163–167 (2004).
20 Romitti, P. A., Lidral, A. C., Munger, R. G., Daack-Hirsch, S., Burns, T. L. & Murray, J. C. Candidate genes for nonsyndromic cleft lip and palate and maternal cigarette smoking and alcohol consumption: evaluation of genotype-environment interactions from a population-based case-control study of orofacial clefts. *Teratology.* **59,** 39–50 (1999).
21 Reutter, H., Birnbaum, S., Mende, M., Lauster, C., Schmidt, G., Henschke, H. *et al.* TGFB3 displays parent-of-origin effects among central Europeans with nonsyndromic cleft lip and palate. *J. Hum. Genet.* **53,** 656–561 (2008).
22 Jezewski, P. A., Vieira, A. R., Nishimura, C., Ludwig, B., Johnson, M., O'Brien, S. E. *et al.* Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. *J. Med. Genet.* **40,** 399–407 (2003).
23 Finnerty, J. R., Mazza, M. E. & Jezewski, P. A. Domain duplication, divergence, and loss events in vertebrate Msx paralogs reveal phylogenomically informed disease markers. *BMC. Evol. Biol.* **9,** 18 (2009).
24 Wilkie, A. O. Pitfalls in the phylogenomic evaluation of human disease-causing mutations. *J. Biol.* **8,** 26 (2009).
25 van den Boogaard, M. J., Dorland, M., Beemer, F. A. & van Amstel, H. K. MSX1 mutation is associated with orofacial clefting and tooth agenesis in humans. *Nat. Genet.* **24,** 342–343 (2000).
26 Tongkobpetch, S., Siriwan, P. & Shotelersuk, V. MSX1 mutations contribute to nonsyndromic cleft lip in a Thai population. *J. Hum. Genet.* **51,** 671–676 (2006).
27 Zhang, H., Catron, K. M. & Abate-Shen, C. A role for the Msx-1 homeodomain in transcriptional regulation: residues in the N-terminal arm mediate TATA binding protein interaction and transcriptional repression. *Proc. Natl Acad. Sci. USA* **93,** 1764–1769 (1996).
28 Amiel, J., Trochet, D., Clement-Ziza, M., Munnich, A. & Lyonnet, S. Polyalanine expansions in human. *Hum. Mol. Genet.* **13,** R235–243 (2004).
29 Kamamoto, M., Machida, J., Yamaguchi, H., Kimura, M., Ono, T. & Jezewski, P. A. Clinical and functional data implicate the Arg(151)Ser variant of MSX1 in familial hypodontia. *Eur. J. Hum. Genet.* **30** 2011.