European Journal of Human Genetics (2009) 17, 774–784
© 2009 Macmillan Publishers Limited All rights reserved 1018-4813/09 $32.00
www.nature.com/ejhg

ARTICLE

The *PDGF-C* regulatory region SNP rs28999109 decreases promoter transcriptional activity and is associated with CL/P

Sun J Choi[1], Mary L Marazita[2], P Suzanne Hart[3], Pawel P Sulima[1], L Leigh Field[4], Toby Goldstein McHenry[2], Manika Govil[2], Margaret E Cooper[2], Ariadne Letra[2], Renato Menezes[2], Somnya Narayanan[2], Maria Adela Mansilla[5], José M Granjeiro[6], Alexandre R Vieira[2], Andrew C Lidral[7], Jeffrey C Murray[5] and Thomas C Hart*[,1]

[1]*Human Craniofacial Genetic Section, Craniofacial and Skeletal Diseases Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA;* [2]*Department of Oral Biology, Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA;* [3]*Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA;* [4]*Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada;* [5]*Department of Pediatrics, University of Iowa, Iowa City, IA, USA;* [6]*Department of Cell and Molecular Biology, Fluminense Federal University, Niterói, Rio de Janeiro, Brazil;* [7]*Department of Orthodontics, University of Iowa, Iowa City, IA, USA*

Human linkage and association studies suggest a gene(s) for nonsyndromic cleft lip with or without cleft palate (CL/P) on chromosome 4q31–q32 at or near the platelet-derived growth factor-C (*PDGF-C*) locus. The mouse *Pdgfc*$^{-/-}$ knockout shows that *PDGF-C* is essential for palatogenesis. To evaluate the role of *PDGF-C* in human clefting, we performed sequence analysis and SNP genotyping using 1048 multiplex CL/P families and 1000 case–control samples from multiple geographic origins. No coding region mutations were identified, but a novel −986 C>T SNP (rs28999109) was significantly associated with CL/P ($P=0.01$) in cases from Chinese families yielding evidence of linkage to 4q31–q32. Significant or near-significant association was also seen for this and several other *PDGF-C* SNPs in families from the United States, Spain, India, Turkey, China, and Colombia, whereas no association was seen in families from the Philippines, and Guatemala, and case–controls from Brazil. The −986T allele abolished six overlapping potential transcription regulatory motifs. Transfection assays of *PDGF-C* promoter reporter constructs show that the −986T allele is associated with a significant decrease (up to 80%) of *PDGF-C* gene promoter activity. This functional polymorphism acting on a susceptible genetic background may represent a component of human CL/P etiology.
European Journal of Human Genetics (2009) **17,** 774–784; doi:10.1038/ejhg.2008.245; published online 17 December 2008

Keywords: *PDGF-C*; SNP; CL/P; promoter activity

*Correspondence: Dr TC Hart, National Institute of Dental and Craniofacial Research, National Institutes of Health, Building 10, Room 5N-102, 10 Center Drive, Bethesda, MD 20892-1423, USA.
Tel: +1 301 402 0262; Fax: +1 301 480 2055;
E-mail: thart@mail.nih.gov
Received 26 May 2008; revised 21 October 2008; accepted 20 November 2008; published online 17 December 2008

Introduction

Isolated or nonsyndromic cleft lip with or without cleft palate (CL/P) is a common birth defect that affects ∼1/700 newborns worldwide.[1,2] Although the identification of genes for nonsyndromic CL/P is far from complete, genetic linkage and association studies in humans and animal

models have identified at least 16 candidate loci for CL/P.[1,3–5]

Rare point mutations or significant association has been found between human nonsyndromic CL/P and missense mutations or polymorphic variants in several genes including *PVRL1*, *IRF6*, *MSX1*, *RUNX2* and *FGF* signaling genes.[1,3,4,6] Only the *IRF6* variant finding has been consistently replicated, and recent evidence ascribes its affect to a point mutation in a *TFAP2A* binding site in an enhancer 10 kb upstream of the *IRF6* promoter.[6]

Multiple lines of evidence support the existence of a human CL/P gene on distal 4q, including significant associations with deletions of 4q31-qter,[7,8] linkage to D4S192[9] and significant allelic association for CL/P with D4S192 in Caucasian[10] and Chilean[11] case–control studies. The most significant result from multipoint linkage analysis of genome scan markers in Chinese multiplex CL/P families was found in 4q32, with the multipoint linkage peak at 158 Mb (Figure 1).[12] The earlier positive linkage and association findings[9–11] with D4S192 are not inconsistent with the Chinese results because earlier studies did not assess any markers distal to D4S192.

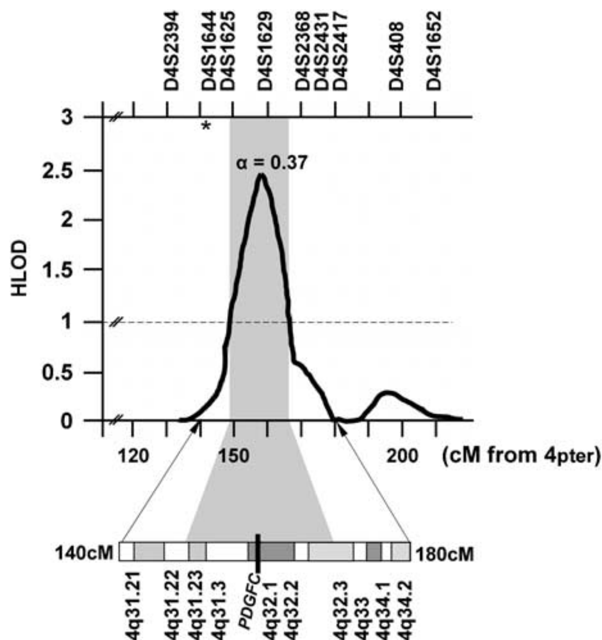

Figure 1 Multipoint linkage analysis results for chromosome 4q in multiplex Chinese families.[12] Depicted are the multipoint heterogeneity LOD scores. α = the estimated proportion of linked families at the peak. The black bar across cytogenetic band 4q32.1 denotes the map position of the *PDGF-C* locus. The vertical gray bar denotes the 1-LOD interval. Chromosomal marker locations and cytogenetic correlations are according to the Ensemble Database. The asterisk denotes the location of D4S192 (142 Mb from 4pter) reported to be significant in previous linkage[12] and association[10,11] reports for CL/P in Caucasian and Chilean populations.

The human *PDGF-C* locus (157.89–158.12 Mb) maps to the linkage peak found in the Chinese families. Mouse knockout studies show that *PDGF-C* is required for palatogenesis.[13] Although human studies support an etiologic role for several genes in CL/P etiology (*PVRL1*, *IRF6*, and *MSX1*),[14–16] expression levels of the mouse homologs of these genes were unaltered in *Pdgfc−/−* mutant embryos that develop clefts, suggesting that their activity is not related to *PDGF-C* signaling in palatogenesis, so *PDGF-C* signaling is a new pathway in palatogenesis, independent of those identified earlier.[13]

These findings led us to study the possible role of *PDGF-C* gene variants in human CL/P. Sequence analysis of the gene in Chinese CL/P cases and controls did not identify any coding region mutations in the *PDGF-C* gene. We did identify a novel SNP in the human *PDGF-C* gene promoter region that was associated with CL/P in Chinese families showing linkage to chromosome 4q32.1. Evaluation of the role of this SNP on the transcriptional regulation of human *PDGF-C* gene expression indicates that it may have a contributory role in some human CL/P cases. To evaluate the generality of these findings, multiple SNPs within *PDGF-C* were assessed for association with CL/P in 1048 multiplex CL/P families from Europe, United States, Asia, Central and South America, and 1000 case–control samples from Brazil.

Materials and methods

Participants

Study samples were drawn from 1048 multiplex families, that is, those with two or more individuals affected with nonsyndromic CL/P, from the following populations that have been described previously:[17,18] US and European Caucasians (Iowa; Texas; Pittsburgh; St Louis; Madrid, Spain; and Turkey), Asia (Shanghai and Beijing, China; West Bengal, India; the Philippines), Central America (Guatemala), and South America (Colombia). An additional set of 500 cases and 500 controls from Brazil was also studied.[19] The families are summarized in Table 1. All study subjects provided informed consent as approved by institutional review boards in both the United States (University of Pittsburgh, University of Iowa) and each of the other countries involved.

A total of 108 affected individuals from the Shanghai, China study population were chosen for complete sequencing of the *PDGF-C* gene (one from each of the 104 families summarized in Table 1 plus an additional 4 cases who were the only participating individuals from their families). A total of 26 cases came from 'linked' families, that is, families that individually had positive LOD scores with anonymous STRP markers in 4q31–q32, and 82 came from 'unlinked' families.[12] A total of 113 Chinese controls were also sequenced, 19 from the same Shanghai study population and 94 from the Coriell Han Chinese Human

Table 1 Summary of families and individuals used in these analyses

		Individuals					
Population[a]	*Families*	*Genotyped affected*	*Genotyped unaffected*	*Untyped affected*	*Untyped unaffected*	*Untyped unknown*	*Total*
Caucasian							
Lowa	117	126	220	4	9	0	359
Pittsburgh	98	132	292	13	89	1	527
Madrid, Spain	36	43	90	0	3	0	136
Turkey	29	32	55	6	195	0	288
ASIA							
India	53	100	220	37	386	0	743
Philippines	242	664	1337	65	1366	0	3432
China	164	240	552	59	394	38	1283
Central America							
Guatemala	77	82	310	11	111	0	514
South America							
Colombia	232	349	577	15	368	0	1309
Total	1048	1768	3653	210	2920	40	8591
Case–controls (Brazil)	0	500	500	0	0	1000	

[a]'Iowa' includes families from Iowa and Texas; 'Pittsburgh' includes families from Pittsburgh, PA, and St Louis, MO; 'China' includes families from Shanghai and Beijing, China.

Variation panel (Coriell Cell Repository, Camden, NJ, URL http://ccr.coriell.org/nigms).

Families from Pittsburgh, St Louis, Spain, India, Turkey, Guatemala, and the Philippines, and the Brazilian case-control samples were used for SNP genotyping of the novel SNP identified from sequencing these Chinese cases. All populations summarized in Table 1 except those from St Louis, Guatemala, and Brazil also had SNP genotyping available for 12 additional SNPs within *PDGF-C* to further investigate the association between CL/P and *PDGF-C* (see Table 3).

DNA sequence analysis of *PDGF-C*

Oligonucleotide primers were designed to amplify 1150 bases of the 5′-regulatory region and all seven exons of the *PDGF-C* gene, including intron–exon boundaries (primer sequences available on request). Primers were designed using Oligo (Molecular Biology Insights, Cascade, CO, USA) and the *PDGF-C* genomic sequence contained in the chromosome 4 contig (NT_016354). All nucleotide numbering assumes the A of the ATG start codon as nucleotide 1. PCR amplification products were subjected to DNA sequencing using ABI dye-terminator chemistry and analyzed using an ABI 3100 DNA analyzer (Applied Biosystems, Foster City, CA, USA). Sequence results were compared with the reference sequences using the Sequencer Program (Genecodes, Ann Arbor, MI, USA). Identification of transcription factor binding sites was determined computationally using AliBaba2.[20]

Generation of human *PDGF-C* promoter reporter constructs

The human *PDGF-C* cDNA sequence (3 kb) was used as a probe for bioinformatics screening of the *PDGF-C* promoter DNA sequence using the NCBI database NT_016354.18. A 20 kb *PDGF-C* promoter sequence was identified and two *PDGF-C* containing bacterial artificial chromosome (BAC) clones were selected as described by Choi *et al.*[21] BAC clones (CTD-3161F17, CTD-2510M6, Open Biosystems, Huntsville, AL, USA) containing the *PDGF-C* promoter were subjected to further analysis. Amplified genomic DNA (130–170 kb insert) from the BAC clones was digested with restriction enzymes *Hin*dIII and *Sst*I (New England Biolabs, Ipswich, MA, USA), respectively. An 11.4 kb band digested with *Sst*I and an 8.3 kb band digested with *Hin*dIII were eluted, subcloned (pBS KSII vector, Stratagene, La Jolla, CA, USA), and sequence-verified. A 8 kb human *PDGF-C* promoter DNA double-digested with *Eco*RV (partial digestion) and *Sac*I and a 4 kb human *PDGF-C* promoter DNA double-digested with *Eco*RV (from multiple cloning site of pBSKSII vector) and *Sac*I (New England Biolabs) were cloned into *Sma*I and *Sac*I sites of pGL2 Enhnacer vector (Promega, Madison, WI, USA). Point mutagenesis substituting T for C at −986 in the human *PDGF-C* promoter was generated using the QuikChange® XL Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's protocol, followed by sequence verification. Constructs −986C and −986T were generated for both −1116C and −1116G backgrounds, so that four haplotypes were evaluated:

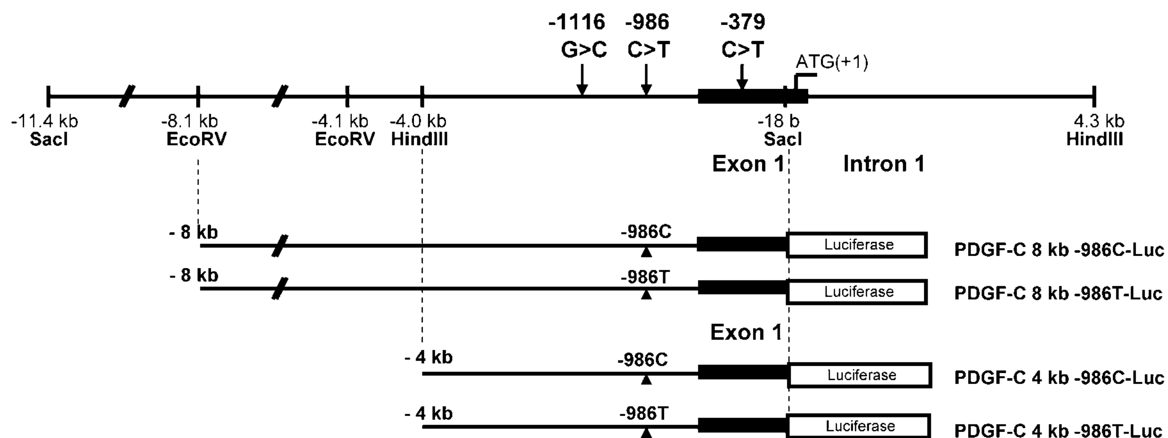

Figure 2 Genomic structure of human *PDGF-C* promoter and schematic diagrams of −986C and −986T allele human *PDGF-C* promoter Luciferase reporter constructs. Eight-kb human *PDGF-C* promoter DNA double-digested with *Eco*RV (partial digestion) and *Sac*I and 4 kb human *PDGF-C* promoter DNA double-digested with *Eco*RV (from the multiple cloning site of pBSKSII vector) and *Sac*I were cloned into *Sma*I and *Sac*I site of pGL2 enhancer reporter vector.

−1116G-986C, −1116G-986T, −1116C-986C, and −1116C-986T (see Figure 2). We generated two independent *PDGF-C* promoter reporter constructs (4 and 8 kb) to examine the −986C > T SNP effects on its transcription activity.

Measurement of −986C and −986T allele human *PDGF-C* promoter activity

Using a Lipofectamine Plus (Invitrogen, Carlsbad, CA, USA) kit, −986C and −986T allele human *PDGF-C* promoter Luciferase reporter constructs were transfected into MC3T3, C2C12, Hela, and 293 cells (ATCC, Manassas, VA, USA) as reported previously.[22] Briefly, MC3T3, C2C12, Hela, and HEK293 cells (5×10^5/well) in six-well plates were transfected with human *PDGF-C* promoter Luciferase reporter constructs (2 μg) and TK-*Renilla* Luciferase construct (0.2 μg). After 72 h of incubation in 5% CO_2 at 37°C, clear cell lysates were harvested by microcentrifugation, and Firefly and *Renilla* Luciferase activity were measured using a dual-Luciferase reporter assay system (Promega) and the relative activity (ratio between Firefly and *Renilla* Luciferase activity) was calculated. Results are reported as the mean ± SE for three replicate samples using five independent −986C and −986T reporter constructs and were analyzed by Student's *t*-test. Results were considered significantly different for $P < 0.05$.

PDGF-C SNP genotyping

To assess the generality of *PDGF-C* SNP associations in our study populations, the novel −986 SNP (rs28999109) identified in the Chinese case–control study was genotyped in families from India, Turkey, Columbia, the Philippines, Pittsburgh, St Louis, Guatemala, Spain, and China, and case–controls from Brazil, using TaqMan® chemistry with the Assay-On-Demand™ C_61773848_10 (Applied Biosystems). Additional *PDGF-C* SNPs were analyzed to further investigate *PDGF-C* in CL/P (Table 3). As part of a larger CL/P fine-mapping project (ACL, MLM, JCM, LLF), DNA samples from the kindreds listed in Table 1 were genotyped for multiple SNPs within *PDGF-C* (see Table 3) by the Center for Inherited Disease Research using the Illumina system (http://www.cidr.jhmi.edu/). Three additional SNPs were genotyped in some families using Assay-On-Demand probes, Applied Biosystems (rs3733486: C_25803666_10; rs1002091: C_7427164_10; and rs13133399: C_9306502_10).

Statistical analysis

Preliminary analyses and case–control comparisons Each SNP was assessed with PedCheck[23] to test for inconsistencies due to nonpaternity or other errors. Standard χ^2 tests and Fisher exact tests were used to compare the SNP allele frequencies (from the sequencing studies) between the Chinese cases and controls, and between the Brazilian cases and controls. *P*-values of 0.05 or less were considered significant.

Allelic association Alleles at each *PDGF-C* SNP were tested for association with CL/P using the Family-Based Association Test (FBAT).[24] Association was assessed for each SNP in each population, as well as the pooled data from all the populations, plus Caucasian, Asian, and Central/South American subsets. In addition to the individual SNPs, association was also tested using the haplotype version of FBAT (HBAT)[24] in the individual populations and pooled subsets for sliding windows across the *PDGF-C* SNPs.

Results

Sequencing of the human *PDGF-C* gene

Direct DNA sequence analysis of the *PDGF-C* coding regions did not identify any coding region mutations in Chinese cases or controls, but three SNPs were identified.

Table 2 Genotyping results for rs28999109 in the human *PDGF-C* gene promoter in 108 affected individuals from Chinese CL/P families and 113 Chinese controls

		Genotype				
	N	*C/C*	*C/T*	*T/T*	χ^2 *(1 d.f.)*	*P-value*
Cases – from linked families[a]	26	21	2	3	6.10	0.014
Cases – from unlinked families[a]	82	71	9	2	0.91	0.341
Cases – total	108	92	11	5	2.84	0.090
Controls – total	113	102	9	2	—	—

[a]Cases are from multiplex CL/P families from Shanghai, China. 'Linked families' are families with evidence of linkage to anonymous STRP marker D4S1629 in 4q31–q32, $HLOD_{MAX}$ is 158 Mb from 4pter.[12]

A known C>T SNP (rs3733486) was detected in the noncoding region of exon 1 (−379 from the ATG start codon), in one unaffected and two affected individuals. Two additional SNPs were identified in the 5′-regulatory domain of the *PDGF-C* gene: a known G>C substitution at −1116 (rs1002091), and a novel C>T substitution at −986 (submitted to dbSNP; assigned rs28999109; see Figure 2). There was no significant difference in the frequency of the −1116 G>C variant between cases and controls ($P=0.21$). Table 2 summarizes the frequency of the C>T SNP variant at −986 in CL/P cases from 4q31–q32 linked and 4q31–q32 unlinked Chinese families, and among Chinese controls. The frequency of the T allele in cases from linked Chinese families was significantly greater than in controls (case allele frequency = 0.15; control = 0.06; $P=0.01$), but not for cases from unlinked families (frequency = 0.08; $P=0.34$), whereas the frequency in all cases was borderline significant (frequency = 0.10; $P=0.09$).

The presence of the wild-type cytosine nucleotide allele at −986 from the ATG start codon is associated with six overlapping transcription factor-binding consensus motifs, including EGR-1, Sp1, WT1, USF, HEB, and ETF (Figure 3). Substitution of thymine at the nucleotide −986 position (−986T) abolishes these cytosine-rich regulatory motifs.

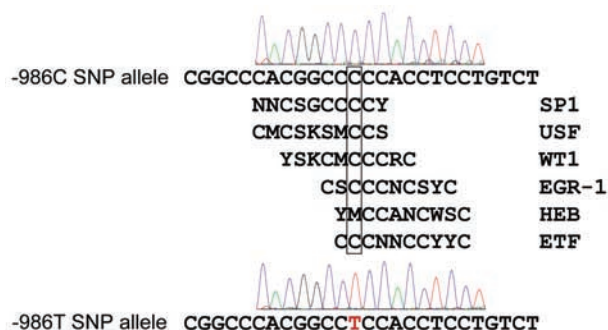

Figure 3 A C>T SNP (rs28999109) at −986 in the human *PDGF-C* promoter regulatory region abolishes consensus binding motifs for Sp1 (Specific protein 1); USF (upstream stimulatory factor), WT1 (Wilms tumor zinc-finger protein-1), EGR-1 (early growth response factor-1), HEB (human B-HLH factor), and ETF (epidermal growth factor receptor (EGFR)-specific transcription factor) (M = A,C; S = C,G; K = G,T; Y = C,T; R = A,G; W = A,T; N = A,C,G,T).

Effects of −986C and −986T SNPs in the human *PDGF-C* promoter on the transcriptional regulation of *PDGF-C* gene

To determine the effects of the −986C/T alleles on *PDGF-C* transcriptional regulation, −986C *PDGF-C* and −986T *PDGF-C* allele Luciferase reporter constructs were cotransfected with the TK-*Renilla* Luciferase construct into MC3T3, C2C12, and HEK-293 cell lines. The relative activity of 4 kb −986T allele *PDGF-C* promoter was significantly decreased in MC3T3 cell lines (about 80%) (Figure 4a) and in C2C12 (Figure 4c) and HEK293 (Figure 4e) cell lines (about 50%) compared with those of 4 kb −986C *PDGF-C* constructs.

Moreover, relative activities of the 8 kb −986T *PDGF-C* promoter constructs are also significantly decreased compared with those of the 8 kb −986C allele *PDGF-C* promoter (Figure 4b, d and f) in all three cell lines. Although transcriptional activities of the 4 kb *PDGF-C* 5′-promoter constructs are more than fivefold greater than all activities associated with the 8 kb promoter constructs, transcriptional activities of all constructs containing the −986T allele were consistently decreased by 40% or more compared with −986C constructs (Figure 4), suggesting that this region (from −8 to −4 kb) of the human *PDGF-C* promoter may contain tissue-specific negative regulatory domain(s). Although there was no statistically significant association with affection status for either allele at −1116, given its close proximity to the −986 SNP within the regulatory region, we measured *PDGF-C* gene promoter activity containing all four possible haplotypes of 4 kb reporter constructs (−1116G −986C; −1116G −986T; −1116C −986C; and −1116C −986T) to evaluate the effect of the −1116 G>C SNP on the transcriptional regulation of *PDGF-C* gene expression. As shown in Figure 4, −986T allele *PDGF-C* reporter constructs significantly decreased promoter activities compared with those of −986C constructs regardless of −1116G or −1116C in MC3T3

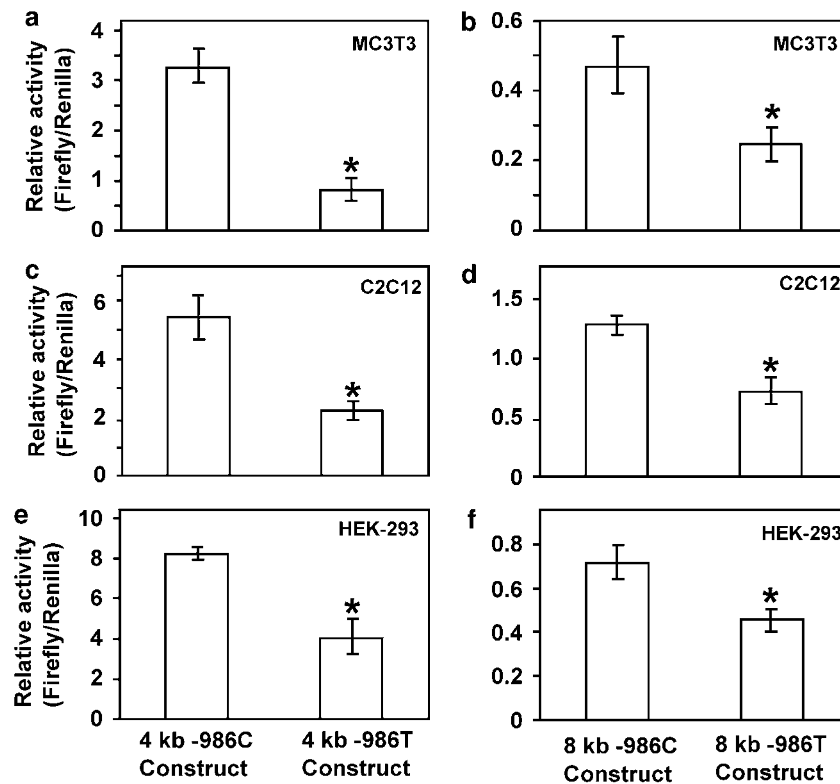

Figure 4 Presence of the −986T SNP (rs28999109) in the human *PDGF-C* promoter reduces promoter activity. Human *PDGF-C* Luciferase reporter constructs (2 μg) containing either the −986C or the −986T allele were cotransfected with the TK-*Renilla* Luciferase construct (0.2 μg) into mouse osteoblastic precursor cells (MC3T3) (**a**, **b**), mouse myoblastic cells (C2C12) (**c**, **d**), and human embryonic kidney epithelial cells (HEK293) (**e**, **f**) using Lipofectamine Plus kits. Relative activities of the 4 kb −986T allele *PDGF-C* promoter construct are significantly decreased (>50%) compared with that of the 4 kb −986C allele *PDGF-C* promoter construct in all three cell lines (**a**, **c**, **e**). Relative activities of 8 kb −986T allele *PDGF-C* promoter construct are also significantly decreased (>40%) compared with those of 8 kb −986C *PDGF-C* promoter construct in these cell lines (**b**, **d**, **f**) (*$P<0.05$).

(Figure 5a), HEK-293 (Figure 5b), and C2C12 (Figure 5c) cell lines. In contrast, −986T allele *PDGF-C* reporter constructs did not decrease the promoter activity of human *PDGF-C* gene in HeLa cell line, suggesting that Hela cells may not express transcriptional regulatory machinery for the human *PDGF-C* gene (Figure 5d).

Allelic association analyses of *PDGF-C* SNPs and CL/P

The results of the FBAT analyses of multiple *PDGF-C* SNPs are summarized in Table 3. The novel −986 *PDGF-C* SNP (rs28999109) was near-significantly associated with CL/P in the West Bengal, India families (P-value = 0.06) and Spanish families (P-value = 0.07) but not in the Turkish, Pittsburgh, St Louis, Guatemala, or Filipino families (P-values > 0.21). Other *PDGF-C* SNPs were also significantly associated with CL/P in some of the populations, notably within the Colombian, St Louis, Spanish, and Chinese samples (see Table 3). Furthermore, results showed a trend toward significance in additional populations (Iowa, Pittsburgh, Turkey). The results from the Brazilian case–control sample were not significant and are not shown in detail. Interestingly, case–control analyses from the Chinese families showed significant association with the −986 SNP, and family-based analyses also showed association with an adjacent SNP (rs10020901) in samples from China (see Table 3, P-value = 0.02), St Louis (P-value = 0.04) and near-significant association in Pittsburgh and Turkey. The pooled FBAT analyses showed a trend toward significance for some of the SNPs in the TOTAL and Caucasian subsets. Haplotype analyses also had population-specific patterns in significance. Table 4 shows the most notable results: a three-SNP haplotype in the Caucasians (comprising SNPs rs983473, rs10517653, and rs10517653) was significantly associated with clefting (P-value 0.04), as was a two-SNP haplotype in the Chinese (the −986 SNP rs28999109 and rs1002091; P-value = 0.04).

Discussion

Animal studies show that PDGF signaling is important in palatal development, with a specific role for *PDGF-C*.[13,25] Evidence from genetic linkage, association and cytogenetic deletions provide support for a human CL/P locus in the

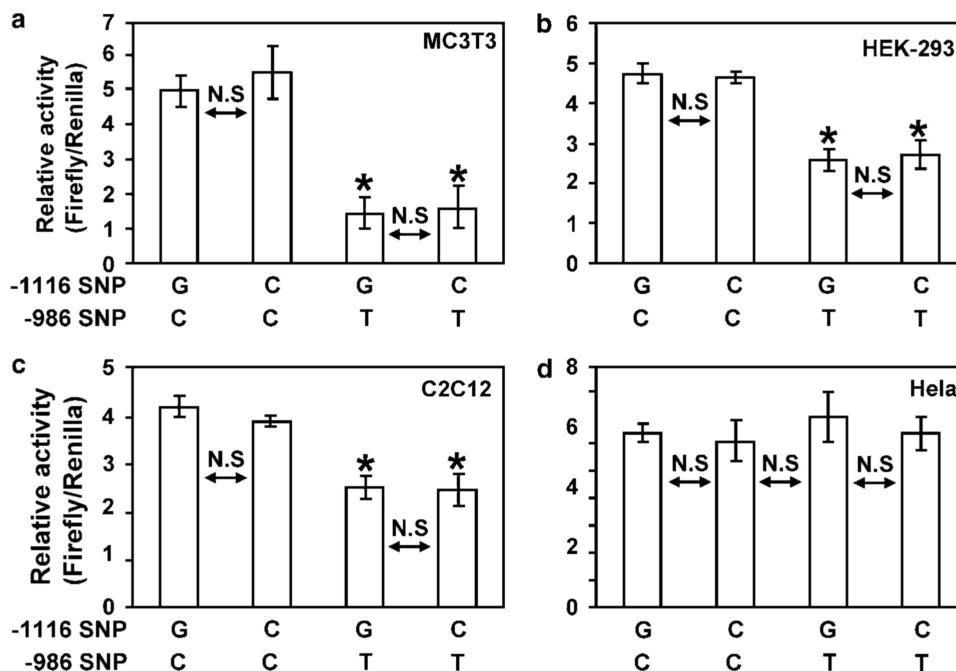

Figure 5 Effects of −1116 G>C SNP on *PDGF-C* promoter activity. Presence of the −986T SNP in the human *PDGF-C* promoter reduces promoter activity regardless of genotype at the −1116 position. Four haplotypes (−1116G −986C; −1116C −986C; −1116G −986T; −1116C −986T) were generated using the 4 kb human *PDGF-C* promoter reporter construct. The presence of the −986T SNP resulted in a significant decrease the *PDGF-C* promoter activity in MC3T3 cell line (**a**), 293 cell line (**b**), and C2C12 cell line (**c**) regardless of the −1116 genotype. However, the −986 T allele did not affect *PDGF-C* promoter activity in HeLa cell line (**d**) (*$P<0.05$).

Table 3 Summary of FBAT analyses of PDGFC SNPs in multiplex CL/P families, by population

						Individual populations						*Combined results*[a]		
Name	*SNP*	*MB location*	*Iowa*	*Pgh*	*St L*	*Spain*	*Turk*	*India*	*Phil*	*China*	*Colom*	*Guat*	*Total*	*Cauc*	*Asian*
rs3815861	T/C	157.903698	—	0.14	—	**0.007**	—	—	—	—	—	0.79	**0.011**	**0.0047**	—
rs1425486	T/C	157.903135	0.40	0.27	—	0.22	0.20	0.68	0.48	0.26	**0.01**	—	0.32	0.94	0.78
rs983473	T/C	157.914842	0.81	0.13	—	**0.016**	0.27	0.54	0.50	0.37	**0.03**	—	0.18	0.16	0.74
rs10517653	T/G	157.953440	0.38	0.16	—	**0.04**	0.32	0.39	0.71	0.39	**0.05**	—	0.33	0.57	0.98
rs2113992	T/C	157.977205	0.20	0.44	—	0.43	0.14	0.82	0.87	0.50	**0.02**	—	0.25	0.43	0.68
rs342318	A/G	158.008405	0.52	0.16	—	0.76	0.32	0.70	0.92	0.33	0.08	—	0.17	0.49	0.78
rs894588	A/G	158.036593	0.60	0.71	—	0.61	0.18	0.84	0.43	0.63	0.83	—	0.52	0.71	0.36
rs11728198	A/G	158.044847	0.75	0.71	—	0.62	0.11	0.71	0.47	0.42	0.94	—	0.54	0.76	0.33
rs6845322	T/C	158.103555	0.42	0.85	—	0.61	0.35	0.43	0.57	0.77	0.78	—	0.32	0.60	0.53
rs3733486	G/A	158.111884	—	—	—	—	—	—	0.42	—	—	—	0.42	—	0.42
rs28999109	C/T	158.124911	—	0.32	0.21	0.07	0.16	0.06	0.39	b	—	0.41	0.72	0.23	0.41
rs1002091	C/G	158.112621	—	0.18	**0.04**	0.35	0.06	—	0.79	**0.02**	—	0.59	0.27	0.39	0.30
rs13133399	T/G	158.117234	—	0.30	—	0.47	—	—	—	—	—	0.53	0.40	0.21	—

'Pgh' = Pittsburgh, 'St L' = St Louis, MO, 'Turk' = Turkey, 'Phil' = Philippines, 'Colom' = Colombia, 'Guat' = Guatemala.
Brazil case–control results not shown as there was no statistical significant association seen.
Results highlighted in dark gray/bold are statistically significant ($P \leq 0.05$), results highlighted in light gray exhibit a trend near significance ($0.05 < P \leq 0.20$).
— Denotes SNP either not genotyped or not polymorphic, or number of informative families too small for valid inference.
[a]Total = results from all families combined; 'Cauc' = Caucasian = results from Iowa+Pittsburgh+St Louis+Spain+Turkey+India; Asian = results from China+Philippines.
[b]See Table 2 for association results with this SNP in the Chinese families.

chromosome 4q31-ter region containing the *PDGF-C* locus.[7–12] These observations led us to evaluate the *PDGF-C* gene for genetic variants that may be etiologic for human CL/P. Sequence analysis identified a novel SNP polymorphism (rs28999109) in the proximal domain of the *PDGF-C* gene in Chinese CL/P families showing maximum linkage to the *PDGF-C* locus.[12] The presence of the less frequent rs288999109 T allele disrupts a highly

Table 4 Sliding window haplotype association results for Caucasian and Chinese subsets

Caucasians

rs3815861	rs1425486	rs983473	rs10517653	rs2113992	rs342318	rs894588	rs11728198	rs6845322	rs28999109	rs1002091	rs13133399
(A) Three-SNP windows in Caucasians											
	0.099775**										
		0.183495									
			0.036706*								
				0.294555							
					0.345681						
						0.734077					
							0.297771				
								0.693876			
									0.524789		
										0.632203	

Chinese

rs1425486	rs983473	rs10517653	rs2113992	rs342318	rs894588	rs11728198	rs6845322	rs28999109	rs1002091
(B) Two-SNP windows in the Chinese									
0.522842									
	0.629394								
		0.683704							
			0.577984						
				0.585417					
					0.571157				
						0.840722			
							0.697151		
								0.036375*	

*Statistically significant ($P<0.05$)
**Borderline significant ($P<0.10$).

conserved cytosine-rich regulatory motif in the 5′-proximal regulatory region of *PDGF-C* that has consensus sequence for several DNA-binding transcription factors that modulate *PDGF* expression.[26,27] Reporter constructs containing the T allele show significantly decreased promoter transcription compared with constructs containing the more common C allele ($P<0.05$). Decreased *PDGF-C* expression has been correlated with orofacial clefting in several mouse models.[13,28–30] Our finding of a positive association between the rarer rs28999109 T allele in the *PDGF-C* 5′-regulatory region and CL/P in a subset of Chinese families is consistent with a plausible etiologic mechanism for a complex trait such as CL/P. Further evidence for a role of *PDGF-C* in CL/P is provided by our results showing statistically significant association between additional *PDGF-C* SNPs and CL/P in study samples from St Louis, Spain, and Colombia, and a trend near significance in other study populations (Iowa, Pittsburgh, and Turkey). Larger study samples will be necessary to confirm the trends toward significance. The results of this study imply population-specific associations with *PDGF-C* in that one of the largest study samples (the Philippines) showed no evidence of association. Interestingly, two of the populations showing associations with the novel −986 SNP (China and India) were both of Asian origin, but the remaining Asian population (the Philippines) did not show significant association with any *PDGF-C* SNPs. The populations that showed significant or near-significant association with other *PDGF-C* SNPs were all Caucasian (Spain, Iowa, Pittsburgh, and Turkey) or mixed Caucasian/native American (Colombia). Although results in data pooled across populations was either borderline or nonsignificant, this is not unexpected given the population-specific patterns in the results, which may also contribute to the well-known differences in the epidemiological characteristics of nonsyndromic CL/P.[1,2] We estimated the population-attributable risk for the −986 allele in the Chinese to range from 4.2 to 6.4%; therefore, one possible explanation for the population-specific differences in the association results may be differences in proportion of risk.

Platelet-derived growth factors play distinct roles at successive stages of mammalian organogenesis.[31] PDGF ligands regulate biological processes by binding to and activating PDGF receptors (PDGFRs). This activates the tyrosine kinase domain contained within the intracellular portion of PDGFRs and initiates intracellular signaling events (Erk/MAPK and Akt/PKB) that trigger cellular responses, such as proliferation, migration, contraction, and cell survival, essential for numerous biological processes.[32] This PDGF signaling specificity is mediated through the activity of multiple immediate early genes (IEGs).[33] PDGF-dependent tissues include the vasculature, kidney, neural crest-derived skeleton, and thoracic skeleton as well as the branchial arches and craniofacial mesenchyme. Identification and validation of PDGF transcriptional

targets have been determined, and the specificity of downstream function is dependent upon PDGF receptor/ligand activation of specific IEGs.[33,34]

PDGF-C can bind and activate PDGFR-αα and -αβ. PDGFR-α is required for neural crest cell development and normal craniofacial development.[35] PDGF-C has been characterized as a key component of PDGFR-α signaling by biochemical analyses[36] and *in vivo* gene-targeting.[13] Nonsyndromic cleft palate derives from an embryopathy with consequent failure of the palatal shelf fusion.[29] PDGFR-α and PDGF-C are key regulators for embryonic and postnatal development, and are required for normal palatogenesis.[13,25,29,30,33,35,37] Animal models show that disruption of Pdgfr-α signaling through a variety of etiologies, including genetic mutation of *Pdgfr-α*,[35] micro-RNA suppression of *Pdgfr-α* translation,[38] genetic mutation of *PDGF-C* ligand,[13,28] and suppression of *PDGF-C* transcription[29] and PDGF-C protein expression by retinoic acid,[29,30] are all associated with disruption of palatogenesis and the presence of orofacial clefting. Although mutations and knockout studies of *Pdgfr-α* and *PDGF-C* are etiologic for defective palatogenesis and orofacial clefting, knockout studies in mice indicate that other genes etiologic for CL/P (*PVRL1*, *IRF6*, *MSX1*) are not altered, indicating that Pdgf-c signaling is a new and independent mechanism that regulates palatogenesis.[1,13]

The proximal domain of *PDGF* gene promoters can directly modulate gene expression.[39] *PDGF-A* and *PDGF-C* share common gene regulatory mechanisms, and their expression is controlled by the zinc-finger transcription factors Sp1 and EGR-1, which have affinity for overlapping nucleotide recognition elements.[26,27,40] Additionally, through its ability to repress expression of EGR-1, WT1 may function as a component in a transcription factor complex that regulates *PDGF-C* expression.[41] The presence of the more common cytosine nucleotide in rs28999109 at nucleotide position −986 from the ATG start codon in the *PDGF-C* gene preserves six overlapping transcription factor-binding consensus motifs (including Sp1, EGR-1, and WT1) within a highly conserved 500 bp interval. Substitution of a thymidine nucleotide at this position abolishes the consensus sequence for multiple DNA-binding transcription factors, including Sp1, EGR-1, and WT1. The transcriptional activities of *PDGF-C* promoter reporter constructs containing the −986 T allele were consistently decreased more than 40% compared with constructs containing the C allele in mesenchymal C2C12 cells and in osteoprogenic MC3T3 cells differentiated from mesenchyme, suggesting that disruption of the consensus sequence may reduce *PDGF-C* expression in these cells. Decreased *PDGF-C* transcription secondary to the less frequent rs28999109 T allele may act on a susceptible genetic background to impair PDGFR-α-PDGF-C signaling, increasing the susceptibility to CL/P. Genetic variants in downstream IEGs critical to craniofacial and palatal development could contribute to susceptibility as shown in the mouse model.[33]

Mutations in *Pdgfr-α*,[35] as well as mutations of their *Pdgf* ligands,[13,28] including *Pdgf-c*, disrupt craniofacial development and are associated with craniofacial clefting and defective palatogenesis. PDGF signaling specificity is mediated through immediate early genes.[33] Mutations of specific immediate response genes (*Arid5b*, *Tiparp*, *Sgpl1*, *BCo55757*, *Axud1*, *Mzf6d*, and *Schip1*), which are the downstream gene targets controlled by the PDGF pathway,[33,34] are also associated with anomalies of craniofacial development. The associated phenotypes are consistent with the facial clefting seen in *Pdgfrα−/−* mice, but are less severe and less penetrant. Additionally, although *Pdgfrα−/−* mice show severe skeletal defects, loss of one copy of *Pdgfrα* (*Pdgfrα−/+*) increased the severity of skeletal defects (including palatal clefting) in many lines with IEG mutations (*Arid5b*, *Tiparp*, *BCo55757*, *BC058969*, and *Schip1*) and created skeletal malformations in one IEG knockout line that had no previous skeletal defects (*Plekha1*).[33] These studies show that mice with mutations in the primary gene targets of PDGF signaling, the immediate early genes (IEGs), show phenotypes in the same structures and cell types as seen in *PDGF* receptor mutants, supporting the notion that target genes control specific processes downstream of individual receptor tyrosine kinases (RTKs). IEGs are specific for individual RTKs and responsible for particular downstream functions. Additionally, defective craniofacial development was seen as a result of dosage-sensitive genetic interactions with PDGF signaling genes and their downstream targets (IEGs), suggesting that these genes work collectively to implement PDGF function in development.[33]

Nonsyndromic cleft palate derives from an embryopathy with consequent failure of palatal shelf fusion.[29] The etiology of this condition is complex, and multiple genes and environmental factors are most likely involved.[3,4] As a key component of the PDGFR-α signaling pathway, *PDGF-C* is an important regulator of cell proliferation, survival, and migration as well as deposition and maintenance of extracellular matrix.[32] Inactivation of *PDGF-C* or reduced *PDGF-C* expression in genetic (Pdgf-c−/−) or teratogen (retinoic acid) models perturbs regulation of MMPs and TIFs in palatal mesenchyme during branchial arch development.[30,42,43] *PDGF-C* is a potent mitogen and it promotes proliferation of mouse embryonic palatal mesenchymal cells and is required for branchial arch morphogenesis.[29] *PDGF-C* expression is also influenced by FGF signaling and small ubiquitin-like modifier modification (SUMO).[44] As alterations of FGF signaling are etiologic for some forms of orofacial clefting,[45,46] SUMO and *PDGF-C* may also interact with environmental risk factors to influence CL/P susceptibility.

The findings of this study show the importance of applying both animal and human model approaches in the

search for developmental and etiological mechanisms underlying complex traits. In the same way as mutations in different genes in the PDGF signaling pathway work collectively to implement PDGF function in development,[33] mutations and functionally significant genetic polymorphisms in different genes along the pathway (receptor, ligand, and IEGs) can collectively result in defective craniofacial development, including craniofacial clefting. Taken together, these findings provide evidence for a role of the rs28999109 *PDGF-C* promoter SNP variant in the etiology of CL/P and highlight the potential importance of regulatory regions in complex traits.[6,47] In addition, the associations seen between other *PDGF-C* SNPs and CL/P provide further evidence of a role for *PDGF-C* in some forms of human CL/P.

Acknowledgements

We acknowledge and express their sincere appreciation to all individuals who participated in these studies. The staff and collaborators at each study site were critical for successful completion of these studies: Dr You-e Liu and Dr Dan-ning Hu (China), Dr Ajit Ray (India), Dr Tuncbilek (Turkey), Carla Brandon, Kathy Bardi, Judith Resick, Dr Katherine Neiswanger, Dr Joseph Losee (Pittsburgh, USA), Dr Consuelo Valencia-Ramirez, Dr Mauricio Camargo, Dr Mauricio Arcos-Burgos, Dora Rivera (Colombia), Sybill Naidoo, Dr Rick Martin, Dr Alex Kane (St Louis, USA). Operation Smile International, Operation Smile Philippines, the HOPE Foundation, Bill and Kathy Magee, Edith Villanueva, Buena Nepomuceno, Henrietta Gamboa, Salie Onggada, Rachel Lim, and Gloria Melocoton were all critical to sample collection in the Philippines. These studies were supported by National Institutes if Health Grants R01-DE09886, R01-DE012472, R37-DE08559, R01-DE016148, R01-DE014667, P50-DE016215; and by the Intramural Program of the National Institute of Dental and Craniofacial Research Z01-DE000711. Some of the genotyping was provided by the Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University; contract number N01-HG-65403. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Dental and Craniofacial Research or the National Institutes of Health.

References

1 Jugessur A, Murray JC: Orofacial clefting: recent insights into a complex trait. *Curr Opin Genet Dev* 2005; **15**: 270–278.
2 Mossey P: Epidemiology underpinning research in the aetiology of orofacial clefts. *Orthod Craniofac Res* 2007; **10**: 114–120.
3 Juriloff DM, Harris MJ, Brown CJ: Unraveling the complex genetics of cleft lip in the mouse model. *Mamm Genome* 2001; **12**: 426–435.
4 Murray JC, Schutte BC: Cleft palate: players, pathways, and pursuits. *J Clin Invest* 2004; **113**: 1676–1678.
5 Vieira AR, Avila JR, Daack-Hirsch S *et al*: Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. *PLoS Genet* 2005; **1**: 651–659.
6 Rahimov F, Marazita ML, Visel A *et al*: Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet* 2008; **40**: 1341–1347.
7 Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick DA: Chromosomal deletion map of human malformations. *Am J Hum Genet* 1998; **63**: 1153–1159.
8 Lin AE, Garver KL, Diggans G *et al*: Interstitial and terminal deletions of the long arm of chromosome 4: further delineation of phenotypes. *Am J Med Genet* 1988; **31**: 533–548.
9 Beiraghi S, Foroud T, Diouhy S *et al*: Possible localization of a major gene for cleft lip and palate to 4q. *Clin Genet* 1994; **46**: 255–256.
10 Mitchell LE, Healey SC, Chenevix-Trench G: Evidence for an association between nonsyndromic cleft lip with or without cleft palate and a gene located on the long arm of chromosome 4. *Am J Hum Genet* 1995; **57**: 1130–1136.
11 Paredes M, Carreno H, Sola JA, Segu J, Palomino H, Blanco R: Association between nonsyndromic cleft lip/palate with microsatellite markers located in 4q. *Rev Med Chil* 1999; **127**: 1431–1438.
12 Marazita ML, Field LL, Cooper ME *et al*: Genome-scan for loci involved in cleft lip with or without cleft palate in Chinese multiplex families. *Am J Hum Genet* 2002; **71**: 349–364.
13 Ding H, Wu X, Bostrom H *et al*: A specific requirement for PDGF-C in palate formation and PDGFR-alpha signaling. *Nat Genet* 2004; **36**: 1111–1116.
14 Avila JR, Jezewski PA, Vieira AR *et al*: PVRL1 variants contribute to non-syndromic cleft lip and palate in multiple populations. *Am J Med Genet* 2006; **140**: 2562–2570.
15 Zucchero TM, Cooper ME, Maher BS *et al*: Interferon regulatory factor 6 (*IRF6*) gene variants confer risk for isolated cleft lip and palate. *New Engl J Med* 2004; **351**: 769–780.
16 Jezewski PA, Vieira AR, Nishimura C *et al*: Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. *J Med Genet* 2003; **40**: 399–407.
17 Marazita ML, Murray JC, Lidral AC *et al*: Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32–35. *Am J Hum Genet* 2004; **75**: 161–173.
18 Neiswanger K, Deleyiannis FW, Avila JR *et al*: Candidate genes for oral-facial clefts in Guatemalan families. *Ann Plast Surg* 2006; **56**: 518–521.
19 Letra A, Menezes R, Granjeiro JM *et al*: Defining subphenotypes for oral clefts based on dental development. *J Dent Res* 2007; **86**: 986–991.
20 Grabe N: AliBaba2: context specific identification of transcription factor binding sites. *In Silico Biol* 2002; **2**: S1–S15.
21 Choi SJ, Oba T, Callander NS, Jelinek DF, Roodman GD: AML-1A and AML-1B regulation of MIP-1alpha expression in multiple myeloma. *Blood* 2003; **101**: 3778–3783.
22 Choi SJ, Song IS, Ryu OH *et al*: A 4 bp deletion mutation in DLX3 enhances osteoblastic differentiation and bone formation *in vitro*. *Bone* 2008; **42**: 162–171.
23 O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
24 Horvath S, Xu X, Laird NM: The family based association test method: strategies for studying general genotype – phenotype associations. *Eur J Hum Genet* 2001; **9**: 301–306.
25 Betsholtz C, Karlsson L, Lindahl P: Developmental roles of platelet-derived growth factors. *Bioessays* 2001; **23**: 494–507.
26 Khachigian LM, Williams AJ, Collins T: Interplay of Sp1 and Egr-1 in the proximal platelet-derived growth factor A-chain promoter in cultured vascular endothelial cells. *J Biol Chem* 1995; **270**: 27679–27686.
27 Midgley VC, Khachigian LM: Fibroblast growth factor-2 induction of platelet-derived growth factor-C chain transcription in vascular smooth muscle cells is ERK-dependent but not JNK-dependent and mediated by Egr-1. *Biol Chem* 2004; **279**: 40289–40295.
28 Wu X, Ding H: Generation of conditional knockout alleles for PDGF-C. *Genesis* 2007; **45**: 653–657.

29 Han J, Xiao Y, Lin J, Li Y: PDGF-C controls proliferation and is down-regulated by retinoic acid in mouse embryonic palatal mesenchymal cells. *Birth Defects Res* 2006; **77**: 438–444.
30 Han J, Li L, Zhang Z *et al*: Platelet-derived growth factor C plays a role in the branchial arch malformations induced by retinoic acid. *Birth Defects Res A Clin Mol Teratol* 2007; **79**: 221–230.
31 Hoch RV, Soriano P: Roles of PDGF in animal development. *Development* 2003; **130**: 4769–4784.
32 Reigstad LJ, Varhaug JE, Lillehaug JR: Structural and functional specificities of PDGF-C and PDGF-D, the novel members of the platelet-derived growth factors family. *FEBS J* 2005; **272**: 5723–5741.
33 Schmahl J, Raymond CS, Soriano P: PDGF signaling specificity is mediated through multiple immediate early genes. *Nat Genet* 2007; **39**: 52–60.
34 Chen WV, Delrow J, Corrin PD, Frazier JP, Soriano P: Identification and validation of PDGF transcriptional targets by microarray-coupled gene-trap mutagenesis. *Nat Genet* 2004; **36**: 304–312.
35 Soriano P: The PDGF alpha receptor is required for neural crest cell development and for normal patterning of the somites. *Development* 1997; **124**: 2691–2700.
36 Li X, Pontén A, Aase K *et al*: PDGF-C is a new protease-activated ligand for the PDGF alpha-receptor. *Nat Cell Biol* 2000; **2**: 302–309.
37 Liu L, Korzh V, Balasubramaniyan NV, Ekker M, Ge R: Platelet-derived growth factor A (pdgf-a) expression during zebrafish embryonic development. *Dev Genes Evol* 2002; **212**: 298–301.
38 Eberhart JK, He X, Swartz ME *et al*: MicroRNA Mirn140 modulates Pdgf signaling during palatogenesis. *Nat Genet* 2008; **40**: 290–298.
39 Maul RS, Zhang H, Reid IV JD, Pedigo NG, Kaetzel DM: Identification of a cell type-specific enhancer in the distal 5′-region of the platelet-derived growth factor A-chain gene. *J Biol Chem* 1998; **273**: 33239–33246.
40 Sanchez-Guerrero E, Midgley VC, Khachigian LM: Angiotensin II induction of PDGF-C expression is mediated by AT1 receptor-dependent Egr-1 transactivation. *Nucleic Acids Res* 2008; **36**: 1941–1951.
41 Scharnhorst V, Menke AL, Attema J *et al*: EGR-1 enhances tumor growth and modulates the effect of the Wilms' tumor 1 gene products on tumorigenicity. *Oncogene* 2000; **19**: 791–800.
42 Han J, Li L, Zhang Z, Xiao Y, Lin J, Li Y: PDGF-C participates in branchial arch morphogenesis and is down-regulated by retinoic acid. *Toxicol Lett* 2006; **166**: 248–254.
43 Jinnin M, Ihn H, Mimura Y, Asano Y, Yamane K, Tamaki K: Regulation of fibrogenic/fibrolytic genes by platelet-derived growth factor C, a novel growth factor, in human dermal fibroblasts. *J Cell Physiol* 2005; **202**: 510–517.
44 Reigstad LJ, Martinez A, Varhaug JE, Lillehaug JR: Nuclear localisation of endogenous SUMO-1-modified PDGF-C in human thyroid tissue and cell lines. *Exp Cell Rse* 2006; **312**: 782–795.
45 Alkuraya FS, Saadi I, Lund JJ, Turbe-Doan A, Morton CC, Maas RL: SUMO1 haploinsufficiency leads to cleft lip and palate. *Science* 2006; **313**: 1751.
46 Pauws E, Stanier P: FGF signaling and SUMO modification: new players in the aetiology of cleft lip and/or palate. *Trends Genet* 2007; **23**: 631–640.
47 Tapia-Páez I, Tammimies K, Massinen S, Roy AL, Kere J: The complex of TFII-I, PARP1, and SFPQ proteins regulates the DYX1C1 gene implicated in neuronal migration and dyslexia. *FASEB J* 2008; **22**: 3001–3009.