

ARTICLE

# Maternal transmission effects of the *PAX* genes among cleft case–parent trios from four populations

Jae Woong Sull<sup>1,2</sup>, Kung-Yee Liang<sup>2</sup>, Jacqueline B Hetmanski<sup>2</sup>, Margaret Daniele Fallin<sup>2</sup>, Roxanne G Ingersoll<sup>2,3</sup>, Jiwan Park<sup>4</sup>, Yah-Huei Wu-Chou<sup>5</sup>, Philip K Chen<sup>5</sup>, Samuel S Chong<sup>6</sup>, Felicia Cheah<sup>6</sup>, Vincent Yeow<sup>7</sup>, Beyoung Yun Park<sup>8</sup>, Sun Ha Jee<sup>1,2</sup>, Ethylin W Jabs<sup>3</sup>, Richard Redett<sup>3</sup>, Alan F Scott<sup>3</sup> and Terri H Beaty<sup>\*,2</sup>

<sup>1</sup>Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea; <sup>2</sup>Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; <sup>3</sup>Johns Hopkins School of Medicine, Baltimore, MD, USA; <sup>4</sup>Department of Molecular and Cellular Biology, Sungkyunkwan University School of Medicine, Suwon, Korea; <sup>5</sup>Chang Gung Memorial Hospital, Taoyuan, Taiwan; <sup>6</sup>Department of Pediatrics, National University of Singapore, Singapore, Singapore; <sup>7</sup>KK Women's & Children's Hospital, Singapore, Singapore and <sup>8</sup>Yonsei University College of Medicine, Seoul, Korea

Isolated cleft lip with or without cleft palate (CL/P) is among the most common human birth defects, with a prevalence of 1 in 700 live births. The paired box (*PAX*) genes have been suggested as candidate genes for CL/P based largely on mouse models; however, few human studies have focused on this gene family. This study tests for association between markers in four *PAX* genes and CL/P using a case-parent trio design considering parent-of-origin effects. Trios from four populations (76 from Maryland, 146 from Taiwan, 35 from Singapore, and 40 from Korea) were genotyped for 34 single nucleotide polymorphisms (SNPs) in the *PAX3*, *PAX6*, *PAX7*, and *PAX9* genes. We performed the transmission disequilibrium test (TDT) on individual SNPs. Parent-of-origin effects were assessed using the transmission asymmetry test (TAT) and the parent-of-origin likelihood ratio test (PO-LRT). TDT analysis showed one SNP (rs766325) in *PAX7* yielding evidence of linkage and association when parent-of-origin was not considered, with an OR(transmission) = 1.62 ( $P = 0.003$ ), and five SNPs in *PAX6* (including two pairs in near perfect linkage disequilibrium). TAT analysis of all trios revealed two SNPs in *PAX7* and four SNPs in *PAX3* showing significant excess maternal transmission. For these six SNPs, the maternal OR(transmission) ranged between 1.74 and 2.40, and PO-LRT was also significant ( $P$ -values = 0.035–0.012). When this analysis was limited to trios with male cases, SNPs in *PAX7* showed higher maternal OR(transmission) and greater significance. *PAX* genes may influence the risk of CL/P through maternal effects, possibly imprinting, which seems to be stronger among male cases.

*European Journal of Human Genetics* (2009) 17, 831–839; doi:10.1038/ejhg.2008.250; published online 14 January 2009

**Keywords:** *PAX* genes; oral cleft; maternal transmission effects; parent-of-origin

\*Correspondence: Terri H Beaty, Department of Epidemiology, School of Public Health, Johns Hopkins University, 615 N Wolfe Street, Baltimore, MD 21205, USA. Tel: +1 410 955 6960; Fax: +1 410 955 0863; E-mail: tbeaty@jhsph.edu

Received 20 February 2008; revised 13 November 2008; accepted 20 November 2008; published online 14 January 2009

## Introduction

Oral clefts are one of the most common birth defects in humans and represent a significant public health burden in terms of both medical and economic burden for affected individuals and their families. Non-syndromic cleft lip with or without cleft palate (CL/P) is complex in its

etiology, and both genes and environmental risk factors influence the risk.<sup>1</sup> Although several candidate genes have been studied extensively in different populations (*TGFA*, *IRF6*, *BCL*, *RARA*, etc), relatively few genes have been shown to contain truly causal mutations (*MSX1*, *PVRL1*, etc), and these are individually rare and often show incomplete penetrance.<sup>2,3</sup> Recently, several studies have reported that genes responsible for Mendelian malformation syndromes that include CL/P (eg *IRF6*, which accounts for the majority of Van der Woude syndrome) may also be associated with non-syndromic clefts.<sup>2,4</sup>

Paired box (*PAX*) genes, termed as the *PAX* gene family, encode for specific DNA-binding transcription factors, which typically contain a *PAX* domain (an octapeptide) and a paired-type homeodomain.<sup>5</sup> The mammalian *PAX* gene family includes nine genes encoding DNA-binding transcriptional regulatory proteins.<sup>6</sup> These nine individual *PAX* genes are assigned to four subgroups based on conservation of their primary structure: (1) *PAX1/PAX9*, (2) *PAX2/PAX5/PAX8*, (3) *PAX3/PAX7*, and (4) *PAX4/PAX6*.<sup>6</sup>

*PAX* genes play critical roles during fetal development and in the growth of cancer cells. Mutations in the *PAX3* (MIM 606597) gene have been associated with Waardenburg syndrome, which can include CL/P.<sup>7</sup> *PAX3* has also been associated with craniofacial-deafness-hand syndrome.<sup>8</sup> *PAX7* (MIM 167410) plays a crucial function during neural crest development.<sup>9</sup> *PAX3* and *PAX7* have also been associated with alveolar rhabdomyosarcoma.<sup>10</sup> Mutations in *PAX6* (MIM 607108) have been associated with aniridia and development of the central nervous system.<sup>11,12</sup> Several studies have reported that *PAX* genes are associated with CL/P in animals.<sup>9,13</sup> However, to date few studies have focused on *PAX* genes as risk factors for CL/P in humans.<sup>14</sup>

It is important to consider parent-of-origin effects when studying birth defects because maternal genotype controls the *in utero* environment of the developing fetus, and separating maternal genotypic effects from imprinting effects remains an important scientific question.<sup>15,16</sup> Maternal parent-of-origin effects have been suggested for several genes associated with non-syndromic CL/P.<sup>17,18</sup> Males are more often affected with CL/P than females<sup>19,20</sup> however, the underlying cause of this aberrant sex ratio remains unclear. In this paper, we tested for association between single nucleotide polymorphisms (SNPs) in *PAX3*, *PAX6*, *PAX7*, and *PAX9* genes and risk of CL/P in 297 case–parent trios, specifically considering parent-of-origin effects in the total sample and stratified by the proband's gender.

## Materials and methods

### Sample description

As part of an international study of oral clefts, we collected data on case–parent trios recruited through treatment

**Table 1** Gender among 297 non-syndromic CL/P cases from four populations

Population	Total	CL/P cases	
		Male	Female
Taiwan	146	95	51
Singapore	35	24	11
Korea	40	22	18
Maryland	76	44	32
Total	297	185	112

centers in Maryland (MD): Johns Hopkins and University of Maryland; Taiwan (TW): Chang Gung Memorial Hospital; Singapore (SP): KK Women's and Children's Hospital, and Korea (KR): Yonsei Medical Center. Research protocols were reviewed and approved by institutional review boards at each institution. Table 1 lists the gender of all CL/P probands. All parents of probands in the Singaporean, Taiwan, and Korean trios were unaffected, but 4 parents among the 76 MD trios also had an oral cleft. The racial background of case families from MD was 80% European American, 16% African American, and 4% 'other'. All probands underwent a clinical genetics evaluation (including checking for other congenital anomalies or major developmental delays), and were classified as having an isolated, non-syndromic CL/P. Among the total collection of 297 cases (5% of whom did not specify laterality), 17% of CL cases and 23% of CLP cases were bilateral.

### SNP selection, DNA, and genotyping

Single nucleotide polymorphic markers were obtained from literature review and the NCBI dbSNP database (<http://www.ncbi.nlm.nih.gov/SNP/>), using a NorthStar Searchlet from Genetic Software Innovations (Cicero, NY, USA), which identified SNPs within each gene based on definitions used in LocusLink and EntrezGene. SNPs were selected with primary consideration given to the spacing between known SNPs and the amount of sequence data available at that time in public databases. SNPs with multiple submitters and higher heterozygosity levels were given priority. SNPs with high 'design scores' (a predictor of useable genotypes provided by Illumina, Inc., San Diego, CA, USA), heterozygosity above 0.1 in both Caucasian and Asian populations, and HapMap validation were included. SNPs were selected in and around four *PAX* genes with the goal of identifying one SNP per 5 kb: 7 SNPs were genotyped for *PAX7* on chromosome 1p36.2–p36.12, 13 for *PAX3* on chromosome 2q35–q37, 7 for *PAX6* on chromosome 11p13, and 7 for *PAX9* on chromosome 14q12–q13. A total of 45 SNPs were identified, and 35 were polymorphic in all populations. The call rate we considered acceptable was  $\geq 80\%$ . One SNP had unacceptably high

rates of missing data (71%), leaving only 34 SNPs with reasonable heterozygosity for analysis (Table 2).

Genomic DNA samples were prepared from peripheral blood by the protein precipitation method described earlier.<sup>21</sup> DNA concentration was determined using the PicoGreen<sup>®</sup> dsDNA Quantitation Kit (Molecular Probes, Inc., Eugene, OR, USA) and all DNA samples were stored at  $-20^{\circ}\text{C}$ . A 4- $\mu\text{g}$  aliquot of each genomic DNA sample (concentration of 100 ng/ $\mu\text{l}$ ) was dispensed into bar-coded 96-well plates and genotyped for SNP markers using the Illumina Golden-Gate chemistry with Sentrix<sup>®</sup> Array Matrices<sup>22</sup> at the SNP Center of the Genetic Resources Core Facility, part of the McKusick–Nathans Institute of Genetic Medicine at the Johns Hopkins School of Medicine. Two duplicates and four CEPH controls were included on each plate to evaluate genotyping consistency within and between plates. Genotypes were generated on a BeadLab 1000 system.<sup>23</sup> All SNPs were inspected, and

poorly performing SNPs were dropped. No Mendelian inconsistencies were found for these 34 SNPs when checked with the SIB-PAIR program.<sup>24</sup>

### Statistical analysis

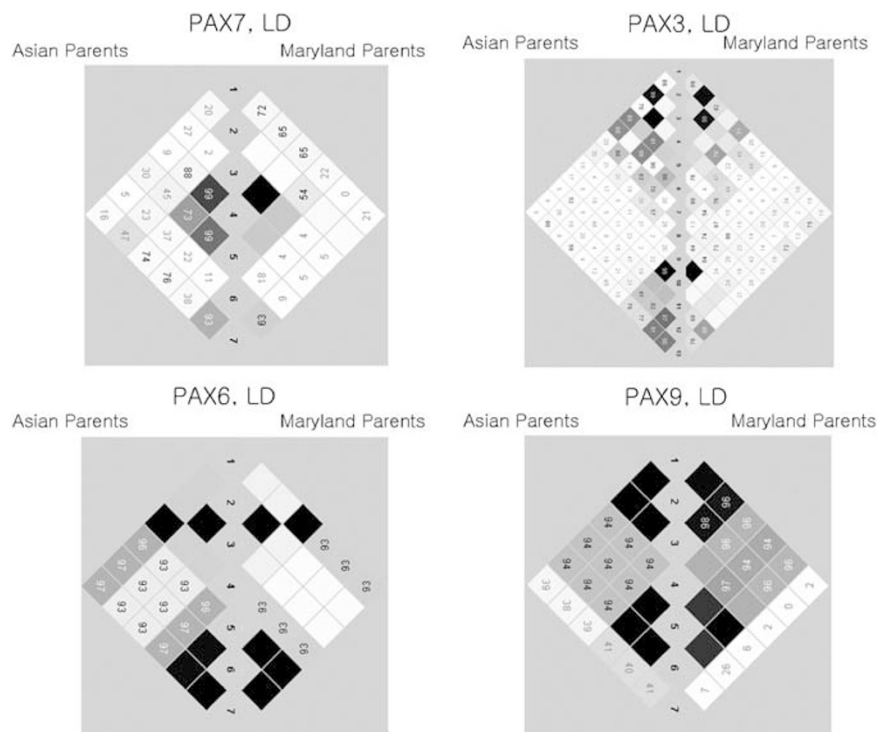
Within each population, minor allele frequencies (MAFs) were computed among parents. Pairwise linkage disequilibrium (LD) was computed as  $r^2$  for all SNPs using the Haploview program,<sup>25</sup> and blocks were identified in Asian and MD population separately (Figure 1). Clayton's extension of the transmission disequilibrium test (TDT) as incorporated into STATA 8.2<sup>26,27</sup> was used on individual SNPs to test for evidence of linkage and LD in the total sample of 297 CL/P trios. From this TDT analysis, we calculated the odds ratio of transmission, OR (transmission), and defined a 'high-risk' allele as that allele over-transmitted to cases (regardless of its statistical significance).

**Table 2** SNP minor allele frequencies among parents of 297 CL/P cases from four populations

Gene	No.	SNP name	Physical location <sup>a</sup>	Minor allele	High-risk allele <sup>b</sup>	Minor allele frequency			
						Taiwan	Singapore	Korea	Maryland
PAX7	1	rs766325	18701764	1	1	0.172	0.172	0.138	0.574
	2	rs880810	18739860	1	2	0.139	0.095	0.107	0.077
	3	rs618941	18751458	2	2	0.306	0.319	0.307	0.056
	4	rs553934	18758087	2	2	0.257	0.276	0.250	0.056
	5	rs545793	18761828	2	2	0.372	0.362	0.336	0.204
	6	rs624761	18790183	2	2	0.195	0.207	0.118	0.169
PAX3	7	rs609959	18791125	2	2	0.089	0.069	0.053	0.095
	1	rs3731858	222891375	1	1	0.254	0.271	0.289	0.194
	2	rs7565552	222900773	1	2	0.089	0.143	0.132	0.391
	3	rs1370924	222904097	1	1	0.256	0.284	0.263	0.218
	4	rs1013262	222916284	2	1	0.089	0.144	0.132	0.401
	5	rs1965791	222922214	1	2	0.346	0.415	0.380	0.709
	6	rs6717432	222927785	1	1	0.167	0.237	0.178	0.126
	7	rs1367411	222931347	2	1	0.162	0.161	0.243	0.489
	8	rs1367414	222934296	1	2	0.382	0.398	0.368	0.517
	9	rs7609007	222940422	2	2	0.348	0.297	0.395	0.094
	10	rs1430662	222946520	1	1	0.349	0.297	0.392	0.094
	11	rs4674639	222949598	2	1	0.356	0.398	0.270	0.213
	12	rs930140	222973797	1	2	0.404	0.458	0.327	0.510
13	rs7600206	222982092	2	1	0.323	0.368	0.257	0.194	
PAX6	1	rs3026393	31768791	2	1	0.491	0.422	0.428	0.528
	2	rs644242	31769377	1	2	0.194	0.181	0.132	0.052
	3	rs667773	31771938	1	2	0.194	0.181	0.132	0.052
	4	rs2239789	31772472	2	1	0.491	0.422	0.427	0.524
	5	rs592859	31775911	1	1	0.259	0.333	0.382	0.164
	6	rs3026354	31787233	2	2	0.252	0.333	0.380	0.164
	7	rs2071164	31790118	2	2	0.253	0.333	0.367	0.165
PAX9	1	rs2073241	36198687	1	1	0.491	0.466	0.428	0.353
	2	rs2073247	36200496	1	1	0.489	0.466	0.412	0.348
	3	rs2295218	36204242	2	2	0.491	0.466	0.428	0.351
	4	rs12892031	36207131	1	1	0.218	0.276	0.322	0.391
	5	rs1955734	36208379	1	1	0.217	0.293	0.322	0.436
	6	rs8004187	36210255	2	2	0.217	0.276	0.322	0.389
	7	rs7144276	36214358	2	2	0.221	0.138	0.280	0.067

<sup>a</sup>Based on NCBI Human Genome build 35.1.

<sup>b</sup>High-risk allele was defined as that allele showing over-transmission to cases in TDT (ie OR(transmission) > 1.0 ignoring parent-of-origin) regardless of statistical significance.



**Figure 1** Linkage disequilibrium as measured by  $r^2$  in *PAX7*, *PAX3*, *PAX6*, and *PAX9* among parents of CL/P children from Asian and Maryland populations. White:  $r^2=0$ . Shades of gray:  $0 < r^2 < 1$ . Black:  $r^2=1$ .

Parent-of-origin analyses were conducted in the total sample in several ways. As an initial screening, parent-of-origin effects were examined using the transmission asymmetry test (TAT), suggested by Weinberg *et al*,<sup>28</sup> which is similar to the TDT but excludes mating between two heterozygotes (where transmission can be ambiguous). TAT was stratified into separate paternal and maternal allelic tests. Next, we used the likelihood-based approach proposed by Weinberg<sup>29</sup> to confirm these parent-of-origin effects. This log-linear model considers the three mating types where the mother and father carry different numbers of variant alleles, stratified by the number of alleles in the child. This log-linear model is used to compute a parent-of-origin likelihood ratio test (PO-LRT), which tests maternal genotypic effects on the phenotype of the fetus (which could otherwise confound assessment of parent-of-origin effects) along with a separate term for imprinting.<sup>29</sup> Here imprinting reflects a differential transmission of alleles to the affected child from mothers versus fathers. PO-LRT was executed using the LEM software.<sup>30</sup> We also tested for parent-of-origin effect in the sample stratified by proband's gender, with separate analyses for trios with male and female cases.

The FAMHAP package was used to estimate haplotype frequencies, while testing for excess transmission of multi-SNP haplotypes.<sup>31</sup> In this haplotype analysis, 2–5 SNP haplotypes using sliding window were analyzed using

FAMHAP. For this FAMHAP analysis, MD and Asian trios were analyzed separately. The haplotype analysis was carried out ignoring parent-of-origin first. Then the haplotype analysis was conducted for maternal and paternal transmissions separately. FAMHAP calculates maximum likelihood estimates of haplotype frequencies from nuclear families through the expectation–maximization algorithm and is robust in handling missing SNPs.<sup>32</sup> This tool provides a haplotype-based test, where the test statistic is based on simulations that randomly permute transmitted and non-transmitted genotypes/haplotypes in each replicate.<sup>33</sup> In this analysis, we used max-TDT, which analyzes each haplotype separately and relies on the maximum TDT as the test statistic.<sup>33,34</sup> The program yields *P*-values corrected for multiple haplotypes.

## Results

Among these 34 SNPs, there was considerable variation in allele frequencies among parents from MD and the three Asian populations (Table 2). From the allele frequencies shown in Table 2, it is clear that some markers showed sharp distinctions between MD and Asian samples, whereas others did not. Taiwan, Korean, and Singaporean parents had very similar haplotype frequencies; therefore all Asian trios were combined for haplotype analysis. Patterns of LD

across each gene were similar in all populations, with some adjacent SNPs in each gene in perfect LD, rendering them redundant (see Figures 1 and Table 2).

When individual markers were screened with the TDT in the total sample, one SNP in *PAX7* and five SNPs in *PAX6* were nominally significant when parent-of-origin was ignored (Table 3). The OR(transmission) was 1.62 ( $P=0.003$ ) for rs766325 in *PAX7*. The five SNPs in *PAX6* showing significant evidence of linkage and LD included two separate pairs of SNPs in perfect LD. The most significant SNP (rs3026354) gave an OR(transmission) = 1.47 ( $P=0.008$ ) ignoring parent-of-origin (Table 3). When analyzed separately in each of the four populations, the association was less strong because of the smaller sample sizes, but the patterns of OR(transmission) were similar (data not shown).

**Table 3** Number of transmitted or non-transmitted minor alleles in 297 CL/P cases (all populations combined) from TDT and estimated odds ratios of transmission, OR(transmission) ignoring parent-of-origin

Gene	No.	SNP name			TDT		OR <sup>a</sup>
			T	NT	$\chi^2$	P-value	
PAX7	1	rs766325	97	60	8.72	<b>0.003</b>	1.62
	2	rs880810	55	46	0.80	0.370	1.20
	3	rs618941	101	79	2.69	0.101	1.28
	4	rs553934	99	75	3.31	0.069	1.32
	5	rs545793	127	107	1.71	0.191	1.19
	6	rs624761	70	68	0.29	0.865	1.03
	7	rs609959	44	34	1.28	0.258	1.29
PAX3	1	rs3731858	99	87	0.77	0.379	1.14
	2	rs7565552	63	51	1.26	0.261	1.24
	3	rs1370924	86	82	0.09	0.757	1.05
	4	rs1013262	67	60	0.39	0.534	1.12
	5	rs1965791	115	111	0.07	0.790	1.04
	6	rs6717432	84	70	1.27	0.259	1.20
	7	rs1367411	91	75	1.54	0.214	1.21
	8	rs1367414	135	118	1.14	0.285	1.14
	9	rs7609007	101	101	0.00	1.000	1.00
	10	rs1430662	101	96	1.27	0.722	1.05
	11	rs4674639	121	101	1.80	0.179	1.20
	12	rs930140	126	104	2.10	0.147	1.21
	13	rs7600206	88	71	1.82	0.177	1.24
PAX6	1	rs3026393	118	110	0.28	0.596	1.07
	2	rs644242	71	47	4.88	<b>0.027</b>	1.51
	3	rs667773	71	47	4.88	<b>0.027</b>	1.51
	4	rs2239789	116	110	0.16	0.690	1.05
	5	rs592859	115	80	6.28	<b>0.012</b>	1.44
	6	rs3026354	115	78	7.09	<b>0.008</b>	1.47
	7	rs2071164	114	78	6.75	<b>0.009</b>	1.46
PAX9	1	rs2073241	120	115	0.11	0.744	1.04
	2	rs2073247	117	114	0.04	0.843	1.03
	3	rs2295218	118	116	0.02	0.895	1.02
	4	rs12892031	100	80	2.22	0.136	1.25
	5	rs1955734	105	85	2.11	0.146	1.24
	6	rs8004187	100	84	1.39	0.238	1.19
	7	rs7144276	74	70	1.11	0.738	1.06

T, transmitted; NT, not transmitted; OR, odds ratio. Shading in the table indicates inferred LD blocks. <sup>a</sup>O (transmission): odds ratio of transmission for the high-risk allele. Bold values represent results significant at the  $P<0.05$  level.

Parent-of-origin effects were first investigated by stratifying informative transmissions and non-transmissions by parental source for all SNPs in the total dataset (Table 4). TAT (ie where heterozygous × heterozygous matings were dropped) revealed three SNPs showing excess maternal transmission significant at the  $P<0.01$  level (rs618941, rs553934 in *PAX7*, and rs1367414 in *PAX3*; see Figure 2 and Table 4), and three others (rs4674639, rs930140, and rs7600206) in *PAX3* showing slightly less significant transmission from mothers. For these six SNPs, estimated maternal OR(transmission) was statistically significant (ranging from 1.74 to 2.40) in TAT analysis. The PO-LRTs were also significant for these six SNPs ( $P$ -values ranging from 0.035 to 0.012) and gave estimated risk ratios for imprinting ranging between 2.08 and 2.78 for these six SNPs, suggesting excess maternal transmission of this region in *PAX7* and *PAX3* (Table 4). Parent-of-origin effects for markers in *PAX6* and *PAX9* were not significant (data not shown).

Separate analyses were conducted for trios with male and female cases. For two SNPs in *PAX7*, the estimated OR(transmission) from mothers to male case was statistically significant (OR = 4.50,  $P=0.0003$  for rs553934; and OR = 4.20,  $P=0.0017$  for rs618941). Both of these SNPs gave significant PO-LRTs ( $P=0.028$  for rs618941 and  $P=0.027$  for rs553934). Among trios with a female case, however, OR(transmission) and PO-LRT were non-significant for these two SNPs (data not shown).

In the haplotype analysis using sliding window (ignoring parent-of-origin), haplotypes of two SNPs (rs766325 and rs880810) in *PAX7* showed evidence of excess transmission of the 1–2 haplotype to the case among Asian trios ( $P=0.036$ ). MD trios showed similar transmission patterns, but were not statistically significant (data not shown). In *PAX6*, haplotypes of three SNPs (rs592859, rs3026354, and rs2071164) showed strong evidence of excess transmission of the 1–2–2 haplotype to the case among Asian trios ( $P=0.011$ ). The MD trios again showed similar transmission patterns, but were not statistically significant (data not shown). Haplotypes in *PAX3* and *PAX9* genes were not significant at  $P$ -value = 0.05 (data not shown).

Next, we conducted the haplotype analyses stratified by parent-of-origin using 2–5 sliding windows (Table 5). In *PAX7*, haplotypes of rs880810 and rs618941 were most significant. The 2–2 haplotype showed evidence of excess maternal transmission to the CL/P child among Asian trios ( $P=0.049$ ) and among MD trios ( $P=0.041$ ), whereas no haplotypes showed deviation from expected when inherited from fathers. Analysis of two SNPs in *PAX3* (rs4674639 and rs930140) showed evidence of excess maternal transmission of the 1–2 haplotype, with stronger evidence coming from Asian trios (here again MD trios showed similar but non-significant patterns of over-transmission).

**Table 4** Number of transmitted or non-transmitted associated alleles to 297 CL/P cases (all populations combined)<sup>a</sup> from TAT and estimated odds ratios, and parent-of-origin likelihood ratio test to test for inequality of maternal *versus* paternal transmission

Gene	No.	SNP name	Paternal TAT				Maternal TAT				PO-LRT <sup>c</sup>	
			T	NT	P-value	OR <sup>b</sup>	T	NT	P-value	OR <sup>b</sup>	OR <sup>d</sup>	P-value
PAX7	1	rs766325	28	16	0.070	1.75	33	20	0.074	1.65	0.90	0.814
	2	rs880810	15	17	0.724	0.88	22	19	0.639	1.16	1.28	0.602
	3	rs618941	27	29	0.789	0.93	33	15	<b>0.009</b>	2.20	2.31	<b>0.033</b>
	4	rs553934	31	30	0.898	1.03	36	15	<b>0.003</b>	2.40	2.38	<b>0.034</b>
	5	rs545793	32	26	0.431	1.23	39	29	0.225	1.34	1.10	0.791
	6	rs624761	32	23	0.225	1.39	22	29	0.327	0.76	0.56	0.134
	7	rs609959	19	10	0.095	1.90	19	18	0.869	1.06	0.56	0.250
PAX3	1	rs3731858	32	21	0.131	1.52	34	27	0.370	1.26	0.82	0.613
	2	rs7565552	20	14	0.303	1.43	18	24	0.355	0.75	0.53	0.179
	3	rs1370924	27	20	0.307	1.35	29	24	0.492	1.21	0.89	0.782
	4	rs1013262	21	18	0.631	1.17	21	27	0.386	0.78	0.67	0.349
	5	rs1965791	31	30	0.898	1.03	33	38	0.553	0.87	0.85	0.651
	6	rs6717432	35	25	0.197	1.40	34	26	0.302	1.31	0.93	0.843
	7	rs1367411	24	19	0.446	1.26	33	30	0.705	1.10	0.88	0.739
	8	rs1367414	35	37	0.814	0.95	42	21	<b>0.008</b>	2.00	2.08	<b>0.035</b>
	9	rs7609007	31	33	0.803	0.94	30	24	0.414	1.25	1.33	0.441
	10	rs1430662	31	30	0.898	1.03	30	24	0.414	1.25	1.21	0.613
	11	rs4674639	25	33	0.294	0.76	41	23	<b>0.024</b>	1.78	2.22	<b>0.032</b>
	12	rs930140	24	35	0.152	0.69	40	23	<b>0.032</b>	1.74	2.54	<b>0.012</b>
	13	rs7600206	17	26	0.170	0.65	33	17	<b>0.024</b>	1.94	2.78	<b>0.018</b>

T, transmitted; NT, not transmitted; OR, odds ratio.

<sup>a</sup>TAT analysis was used on trios with matings between two heterozygous parents deleted.

<sup>b</sup>OR(transmission): odds ratio of transmission of the high-risk allele.

<sup>c</sup>Parent-of-origin likelihood ratio tests (PO-LRT) include separate terms for imprinting.

<sup>d</sup>OR: odds ratio for imprinting effect (ie differential transmission from mothers and from fathers).

Shading in the table indicates inferred LD blocks.

Bold values represent results significant at the  $P < 0.05$  level.

## Discussion

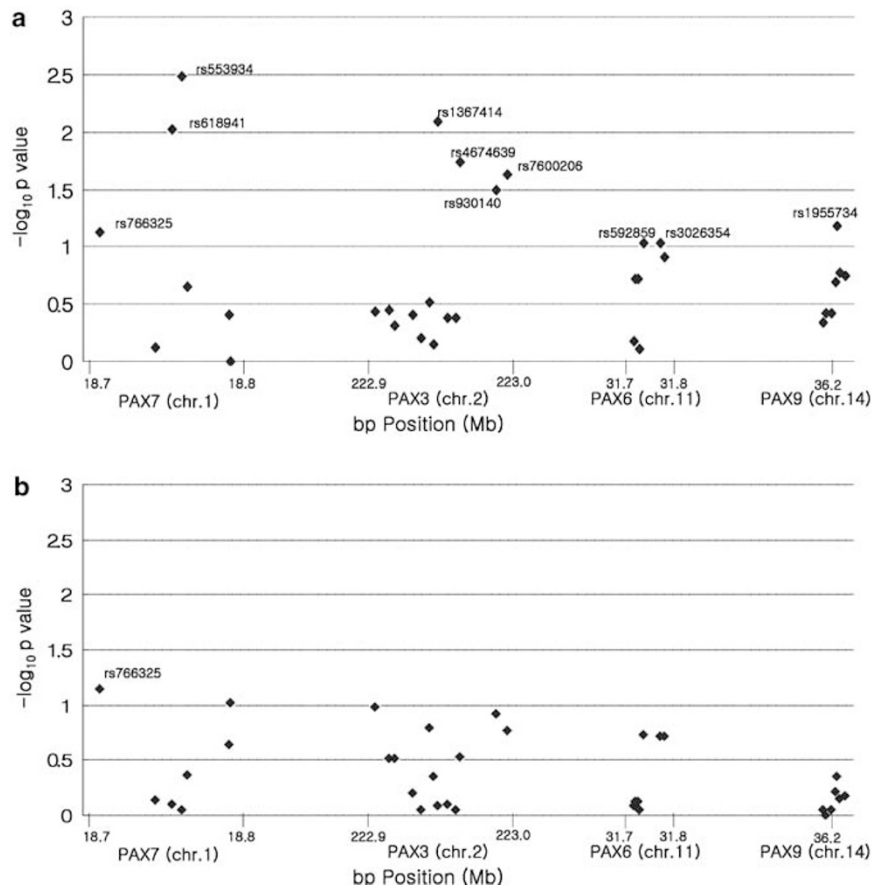
Our study of case–parent trios from different populations (comprising a total of 297 CL/P trios) showed evidence of linkage in the presence of LD for multiple SNPs in the *PAX7* and *PAX3* genes only when parent-of-origin effects were considered. In this study, ignoring parent-of-origin made the *PAX7* and *PAX3* genes look relatively uninteresting. Only a single SNP in *PAX7* showed any evidence of linkage and LD. However, considering parent-of-origin revealed two SNPs in *PAX7* and three SNPs in *PAX3* yielding strong evidence of linkage and LD when transmitted from mothers but not from fathers. This evidence was more dramatic among male cases. Other studies also report that ignoring parent-of-origin could lead to overlooking important genes. In a case–parent trio study for bipolar I disorder, TDT analysis revealed no statistically significant association with SNPs on chromosome 18p11. However, when parent-of-origin was considered, evidence of association was seen involving two potentially causal loci.<sup>35</sup>

In screening for parent-of-origin effects, we found suggestive evidence of excess maternal transmission for several SNPs in *PAX3* and *PAX7*, which are closely related and are important in mammalian embryogenesis.<sup>36</sup> Relaix *et al*<sup>37</sup> identified a new cell population expressing transcription factors *PAX3* and *PAX7*, but

no skeletal muscle-specific markers. These cells are maintained as a proliferating cell line throughout development in embryonic and fetal muscles of the trunk and limbs.

Excess maternal transmission could reflect genomic imprinting or direct maternal genotype effects on the developing fetus. Maternal genotypic effects for non-syndromic CL/P have also been reported for several other candidate genes (MTHFR, CBS, and TGFB3), but these are yet to be confirmed.<sup>17,18,38</sup>

In this study, log-linear models discriminating between maternal genotype and child genotype effects revealed a possible maternal imprinting effect for multiple SNPs in *PAX7* and *PAX3*. Estimates of maternal genotype effects were generally non-significant for the 19 SNPs in *PAX7* and *PAX3*, except for a single SNP (rs1367414). Genomic imprinting is defined as the differential expression of alleles depending on parent-of-origin.<sup>39</sup> A common feature of imprinted genes is DNA sequence carrying a gametic methylation imprint, known as gametic DMR (Differentially DNA-Methylated Region).<sup>40</sup> Parental allele-specific DNA methylation has been found at most imprinted clusters examined thus far. For example, the *IGF2* cluster has a gametic DMR located 2 kb upstream from the H19 nc RNA promoter, which is methylated only in the paternal



**Figure 2** Empirical *P*-values for individual SNPs from *PAX7*, *PAX3*, *PAX6*, and *PAX9* genes in 297 CL/P case–parent trios from four populations (Maryland, Singapore, Taiwan, and Korea) combined. (a) Only maternal transmission was considered; (b) only paternal transmission was considered.

**Table 5** Testing for excess transmission of haplotypes of SNPs rs880810 and rs618941 in *PAX7* and SNPs rs4674639 and rs930140 in *PAX3* in 297 CL/P case–parent trios using the program FAMHAP with maternal and paternal transmission considered separately

Gene	Population	Haplotype	Frequency	Maternal			Paternal		
				T	NT	Maximum TDT (P-value)	T	NT	Maximum TDT (P-value)
PAX7	Asian_combined	2 1	0.591	44.9	53.0	4.803 (0.049)	42.9	44.5	0.029 (0.996)
		2 2	0.280	48.2	29.0		40.0	38.6	
		1 1	0.093	14.6	22.6		14.7	15.0	
	Maryland	1 2	0.036	5.4	8.5		5.5	5.0	
		2 1	0.863	2.0	10.0	5.333 (0.041)	8.0	6.0	0.667 (0.782)
		2 2	0.069	4.0	0.0		4.0	4.0	
PAX3	Asian_combined	1 1	0.069	6.0	2.0		2.0	4.0	
		1 2	0.584	58.4	36.0	5.331 (0.046)	41.5	49.9	1.393 (0.557)
		2 1	0.305	36.0	53.4		41.9	41.5	
	Maryland	2 2	0.030	4.6	5.0		9.6	5.1	
		1 1	0.081	14.1	18.5		16.1	12.6	
		1 2	0.468	15.3	9.0	2.954 (0.138)	9.4	10.0	0.760 (0.780)
		2 1	0.182	3.9	10.4		9.8	7.6	
		2 2	0.027	0.7	2.0		0.6	2.0	
		1 1	0.323	10.1	8.6		11.2	11.4	

Bold values represent results significant at the  $P < 0.05$  level.

gamete and is maintained thereafter in all somatic tissues.<sup>41</sup> Kurmasheva *et al*<sup>42</sup> suggested *PAX3* gene methylation may be correlated with gene inactivation.

In a variety of animal species, maternal transcripts and proteins control early embryonic development in the developing oocyte.<sup>43</sup> In the leech *Helobdella*, Woodruff

*et al*<sup>44</sup> found that *Hau-Pax3/7A* is present as a maternal transcript in both ectodermal and mesodermal progenitor cells. They suggested *Hau-Pax3/7A* plays an important role in mesoderm development. Helobdella embryos receive a large contribution of maternal *Hau-Pax3/7A* RNA, but its function remains unknown.<sup>44</sup>

Many congenital anomalies occur more often in one gender. Males are more often affected with CL/P than females.<sup>19,20</sup> Rittler *et al*<sup>20</sup> reported that infants with CL/P were more frequently female when the father was older, and among CL cases, this shift in sex ratio was highly significant. In our results from markers in *PAX7* and *PAX3* genes, boys showed stronger evidence of possible imprinting than female cases.

Even though this candidate gene study involved a modest number of SNPs in each gene, addressing the issue of multiple comparisons is necessary before an overall statement about the significance of our findings can be made. Here we relied on a hypothesis-driven approach for single SNP analysis and haplotype-based test statistics. SNPs in strong LD typically have highly correlated *P*-values, adjusting significance levels through Bonferroni correction is overly conservative. Therefore, following the strategy in Sull *et al*,<sup>45</sup> we adjusted empirical *P*-values for the number of LD blocks rather than the number of SNPs. In this study, we have 10 LD blocks in these four genes (three for *PAX7*, three for *PAX3*, two for *PAX6*, and two for *PAX9*). In the second block with two SNPs in *PAX7* gene (as shown in Table 4), we found evidence against the null hypothesis only for maternal transmission (the empirical *P*-value of 0.006 would still be marginally significant after correcting for the number of LD blocks). We also used haplotype-based test statistics based on permutation analysis of case–parent trio data. Salyakina *et al*<sup>46</sup> argue that permutation tests are generally preferred over adjustments of asymptotic *P*-values based on the estimated correlation structure among multiple markers or on conventional Bonferroni adjustment (which can be too conservative).<sup>47</sup> The case–parent trio design offers the advantage of testing directly for maternal versus paternal effects, and allows separating these from effects of the fetal genotype versus parental origin in a robust manner.<sup>26,48</sup> Another advantage of this design is that it minimizes confounding that plagues traditional case–control designs. This permits pooling trios from four diverse populations into a combined test of allelic effects on OR(transmission), while testing for parent-of-origin effects. The present study showed excess maternal transmission of markers in *PAX7* and *PAX3*, suggesting that these genes may influence the risk of CL/P, possibly through imprinting. Independent confirmation is still needed to determine the ultimate impact of these *PAX* genes on risk to CL/P.

#### Acknowledgements

This research was supported by R21-DE-013707 and R01-DE-014581 from the National Institute of Dental & Craniofacial Research and

Korean Research Foundation (2004-041-E00104), and the Seoul City R&BD program (10526) in Korea. We thank all participants who donated samples for this multi-center study of oral clefts, as well as the staff at each participating site and institution. We also thank Gerald Raymond for his assistance in screening CL/P patients at Hopkins.

#### Web Resources

Online Mendelian Inheritance in Man (OMIM): <http://www.ncbi.nlm.nih.gov/Omim/>

#### References

- Cobourne MT: The complex genetics of cleft lip and palate. *Eur J Orthod* 2004; **26**: 7–16.
- Zuccherro TM, Cooper ME, Maher BS *et al*: Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. *N Engl J Med* 2004; **351**: 769–780.
- Carinci F, Scapoli L, Palmieri A, Zollino I, Pezzetti F: Human genetic factors in nonsyndromic cleft lip and palate: an update. *Int J Pediatr Otorhinolaryngol* 2007; **71**: 1509–1519.
- Stanier P, Moore GE: Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Hum Mol Genet* 2004; **13**: R73–R81.
- Volcik KA, Blanton SH, Kruzel MC *et al*: Testing for genetic associations with the *PAX* gene family in a spina bifida population. *Am J Med Genet* 2002; **110**: 195–202.
- Underhill DA: Genetic and biochemical diversity in the *Pax* gene family. *Biochem Cell Biol* 2000; **78**: 629–638.
- Tassabehji M, Read AP, Newton VE *et al*: Mutations in the *PAX3* gene causing Waardenburg syndrome type 1 and type 2. *Nat Genet* 1993; **3**: 26–30.
- Asher Jr JH, Sommer A, Morell R, Friedman TB: Missense mutation in the paired domain of *PAX3* causes craniofacial-deafness-hand syndrome. *Hum Mutat* 1996; **7**: 30–35.
- Basch ML, Bronner-Fraser M, Garcia-Castro MI: Specification of the neural crest occurs during gastrulation and requires *Pax7*. *Nature* 2006; **441**: 218–222.
- Barr FG, Nauta LE, Davis RJ, Schafer BW, Nycum LM, Biegel JA: *In vivo* amplification of the *PAX3-FKHR* and *PAX7-FKHR* fusion genes in alveolar rhabdomyosarcoma. *Hum Mol Genet* 1996; **5**: 15–21.
- Chao LY, Mishra R, Strong LC, Saunders GF: Missense mutations in the DNA-binding region and termination codon in *PAX6*. *Hum Mutat* 2003; **21**: 138–145.
- Heins N, Malatesta P, Ceconi F *et al*: Glial cells generate neurons: the role of the transcription factor *Pax6*. *Nat Neurosci* 2002; **5**: 308–315.
- Peters H, Neubuser A, Kratochwil K, Balling R: *Pax9*-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes Dev* 1998; **12**: 2735–2747.
- Ichikawa E, Watanabe A, Nakano Y *et al*: *PAX9* and *TGFB3* are linked to susceptibility to nonsyndromic cleft lip with or without cleft palate in the Japanese: population-based and family-based candidate gene analyses. *J Hum Genet* 2006; **51**: 38–46.
- Weinberg CR, Umbach DM: A hybrid design for studying genetic influences on risk of diseases with onset early in life. *Am J Hum Genet* 2005; **77**: 627–636.
- Wilkins JF, Haig D: What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet* 2003; **4**: 359–368.
- van Rooij IA, Vermeij-Keers C, Kluijtmans LA *et al*: Does the interaction between maternal folate intake and the methylenetetrahydrofolate reductase polymorphisms affect the risk of cleft lip with or without cleft palate? *Am J Epidemiol* 2003; **157**: 583–591.
- Rubini M, Brusati R, Garattini G *et al*: Cystathionine beta-synthase c.844ins68 gene variant and non-syndromic cleft lip and palate. *Am J Med Genet* 2005; **136**: 368–372.



- 19 Niswander JD, MacLean CJ, Chung CS, Dronamraju K: Sex ratio and cleft lip with or without cleft palate. *Lancet* 1972; **2**: 858–860.
- 20 Rittler M, Lopez-Camelo J, Castilla EE: Sex ratio and associated risk factors for 50 congenital anomaly types: clues for causal heterogeneity. *Birth Defects Res A Clin Mol Teratol* 2004; **70**: 13–19.
- 21 Bellus GA, Hefferon TW, Ortiz de Luna RI *et al*: Achondroplasia is defined by recurrent G380R mutations of FGFR3. *Am J Hum Genet* 1995; **56**: 368–373.
- 22 Oliphant A, Barker DL, Stuenkel JR, Chee MS: BeadArray™ Technology: enabling an accurate, cost-efficient approach to high-throughput genotyping. *Biotechniques* 2002; **32**: S56–S61.
- 23 Fan JB, Oliphant A, Shen R *et al*: Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* 2003; **68**: 69–78.
- 24 Duffy DL: *Sib-pair, Version 1.00a*. Brisbane: Queensland Institute of Medical Research, 2007.
- 25 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 26 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.
- 27 Cordell HJ, Barratt BJ, Clayton DG: Case/pseudocontrol analysis in genetic association studies: a unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet Epidemiol* 2004; **26**: 67–185.
- 28 Weinberg CR, Wilcox AJ, Lie RT: 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* 1998; **62**: 969–978.
- 29 Weinberg CR: Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. *Am J Hum Genet* 1999; **65**: 229–235.
- 30 van Den Oord EJ, Vermunt JK: Testing for linkage disequilibrium, maternal effects, and imprinting with (in)complete case-parent triads, by use of the computer program LEM. *Am J Hum Genet* 2000; **66**: 335–338.
- 31 Becker T, Baur MP, Knapp M: Detection of parent-of-origin effects in nuclear families using haplotype analysis. *Hum Hered* 2006; **62**: 64–76.
- 32 Becker T, Knapp M: Maximum-likelihood estimation of haplotype frequencies in nuclear families. *Genet Epidemiol* 2004; **27**: 21–32.
- 33 Zhao H, Zhang S, Merikangas KR *et al*: Transmission/disequilibrium tests using multiple tightly linked markers. *Am J Hum Genet* 2000; **67**: 936–946.
- 34 Schaid DJ: General score tests for associations of genetic markers with disease using cases and their parents. *Genet Epidemiol* 1996; **13**: 423–449.
- 35 Mulle JG, Fallin MD, Lasseter VK, McGrath JA, Wolyniec PS, Pulver AE: Dense SNP association study for bipolar I disorder on chromosome 18p11 suggests two loci with excess paternal transmission. *Mol Psychiatry* 2007; **12**: 367–375.
- 36 Mansouri A, Hallonet M, Gruss P: Pax genes and their roles in cell differentiation and development. *Curr Opin Cell Biol* 1996; **8**: 851–857.
- 37 Relaix F, Rocancourt D, Mansouri A, Buckingham M: A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature* 2005; **435**: 948–953.
- 38 Reutter H, Birnbaum S, Mende M *et al*: TGFB3 displays parent-of-origin effects among central Europeans with nonsyndromic cleft lip and palate. *J Hum Genet* 2008; **53**: 656–661.
- 39 Iwasa Y: The conflict theory of genomic imprinting: how much can be explained? *Curr Top Dev Biol* 1998; **40**: 255–293.
- 40 Parham DM, Ellison DA: Rhabdomyosarcomas in adults and children: an update. *Arch Pathol Lab Med* 2006; **130**: 1454–1465.
- 41 Ferguson-Smith AC, Sasaki H, Cattanach BM, Surani MA: Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* 1993; **362**: 751–755.
- 42 Kurmasheva RT, Peterson CA, Parham DM, Chen B, McDonald RE, Cooney CA: Upstream CpG island methylation of the PAX3 gene in human rhabdomyosarcomas. *Pediatr Blood Cancer* 2005; **44**: 328–337.
- 43 Davidson EH: *Gene Activity in Early Development*. Orlando, FL: Academic Press, 1986.
- 44 Woodruff JB, Mitchell BJ, Shankland M: Hau-Pax3/7A is an early marker of leech mesoderm involved in segmental morphogenesis, nephridial development, and body cavity formation. *Dev Biol* 2007; **306**: 824–837.
- 45 Sull JW, Liang KY, Hetmanski JB *et al*: Differential parental transmission of markers in RUNX2 among cleft case-parent trios from four populations. *Genet Epidemiol* 2008; **32**: 505–512.
- 46 Salyakina D, Seaman SR, Browning BL, Dudbridge F, Muller-Myhsok B: Evaluation of Nyholt's procedure for multiple testing correction. *Hum Hered* 2005; **60**: 19–25.
- 47 Nyholt DR: A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004; **74**: 765–769.
- 48 Starr JR, Hsu L, Schwartz SM: Assessing maternal genetic associations: a comparison of the log-linear approach to case-parent triad data and a case-control approach. *Epidemiology* 2005; **16**: 294–303.