A Variant in the *IRF6* Promoter Associated with the Risk for Orofacial Clefting

Journal of Dental Research
2023, Vol. 102(7) 806–813

DOI: 10.1177/00220345231165210
journals.sagepub.com/home/jdr

M.-J. Li[1*], P. Kumari[2*], Y.-S. Lin[1*], M.-L. Yao[1], B.-H. Zhang[1], B. Yin[1], S.-J. Duan[1], R.A. Cornell[2], M.L. Marazita[3], B. Shi[1], and Z.-l. Jia[1]

Abstract

The single-nucleotide polymorphism (SNP) rs2235371 (*IRF6* V274I) is associated with nonsyndromic cleft lip with or without cleft palate (NSCL/P) in Han Chinese and other populations but appears to be without a functional effect. To find the common etiologic variant or variants within the haplotype tagged by rs2235371, we carried out targeted sequencing of an interval containing *IRF6* in 159 Han Chinese with NSCL/P. This study revealed that the SNP rs12403599, within the *IRF6* promoter, is associated with all phenotypes of NSCL/P, especially nonsyndromic cleft lip (NSCLO) and a subphenotype of it, microform cleft lip (MCL). This association was replicated in 2 additional much larger cohorts of cases and controls from the Han Chinese. Conditional logistic analysis indicated that association of rs2235371 with NSCL/P was lost if rs12403599 was excluded. rs12403599 contributes the most risk to MCL: its G allele is responsible for 38.47% of the genetic contribution to MCL, and the odds ratios of G/C and G/G genotypes were 2.91 and 6.58, respectively, for MCL. To test if rs12403599 is functional, we carried out reporter assays in a fetal oral epithelium cells (GMSM-K). Unexpectedly, the risk allele G yielded higher promoter activity in GMSM-K. Consistent with the reporter studies, expression of *IRF6* in lip tissues from NSCLO and MCL patients with the G/G phenotype was higher than in those from patients with the C/C phenotype. These results indicate that rs12403599 is tagging the risk haplotype for NSCL/P better than rs2235371 in Han Chinese and supports investigation of the mechanisms by which the allele of rs12403599 affects *IRF6* expression and tests of this association in different populations.

Keywords: orofacial, deep sequencing, genetic association studies, logistic model, promoter region, genetic susceptibility

Introduction

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is one of the most common orofacial clefts, with an average incidence of 1/1,000 live births worldwide (Dixon et al. 2011) and about 1.44/1,000 in China (Dai et al. 2010). It extensively damages oral and maxillofacial functions, including mastication, swallowing, speech, and hearing. In addition to these impacts in early life, NSCL/P is also associated with an increased risk for mental health disorders, cancer, and death rate from all causes throughout life (Christensen et al. 2004). Treatment for NSCL/P requires long-term serial interventions from a team of specialists in surgery, speech, and orthodontics. Thus, NSCL/P seriously influences life quality of those affected and imposes substantial economic burdens (Wehby and Cassell 2010).

While environmental factors contribute to the occurrence of NSCL/P, genetic predisposition plays a major role in its etiology (Grosen et al. 2011; Beaty et al. 2016). In 2002, it was discovered that a single-dose deficiency of *IRF6* interfered with oral and facial development, leading to the occurrence of Van der Woude syndrome (OMIM: 119300) and popliteal pterygium syndrome (OMIM: 119500) (Kondo et al. 2002). In 2004, via association analysis among 1,968 NSCL/P families from 10 populations, Zucchero et al. (2004) showed that the C allele of rs2235371 (p.V274I) of *IRF6* was significantly

[1]State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases & Department of Cleft Lip and Palate, West China Hospital of Stomatology, Sichuan University, Chengdu, China
[2]Department of Oral Health Sciences, University of Washington, Seattle, WA, USA
[3]Centre for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA, USA
*Authors contributing equally to this article.

A supplemental appendix to this article is available online.

Corresponding Author:
Z.-l. Jia, State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases & Dept. of cleft lip and palate, West China Hospital of Stomatology, Sichuan University, No. 14, Section 3, Ren Min Nan Road, Chengdu 610041, People's Republic of China.
Email: zhonglinjia@sina.com

overtransmitted in NSCL/P patients in the entire population data set and explains about 12% of heritability for NSCL/P. The association between rs2235371 (*IRF6*) and NSCL/P has been well established by both genome-wide association studies (GWASs) and over 20 candidate gene studies (Park et al. 2007; Zhou et al. 2013; Sun et al. 2015; de Souza et al. 2016; Wu-Chou et al. 2019). However, functional studies of rs2235371 in zebrafish showed that it is benign and unlikely to directly cause orofacial clefts (Li et al. 2017). Importantly, the lead single-nucleotide polymorphism (SNP) identified by GWAS is not necessarily the pathogenic variant within the haplotype block; other SNPs in strong linkage disequilibrium (LD) with it are as likely to be the pathogenic variant (Dickson et al. 2010; Sun et al. 2011; Chang and Keinan 2012). For instance, rs642961, located in an enhancer upstream of *IRF6*, is in strong LD with rs2235371 and may be functional because it alters the binding of transcription factor AP2-α in vitro (Rahimov et al. 2008). However, it remains unknown whether there are other functional SNPs in the haplotype tagged by rs2235371. To address this question, further screening the region around rs2235371 for potential functional variants is necessary.

In this study, to find etiologic variants near *IRF6*, we conducted targeted sequencing of an interval containing *IRF6* in a cohort of 159 NSCL/P cases. We carried out a replication studies in larger cohorts. We also performed bioinformatic and functional analyses. Together, the studies support rs12403599, which is in strong LD with rs2235371, as being functional (Appendix Fig. 1).

Materials and Methods

Sample Collection

In the initial phase of targeted sequencing, we recruited 159 NSCL/P patients (79 cases of nonsyndromic cleft lip with cleft palate [NSCLP] and 80 cases of nonsyndromic cleft lip [NSCLO]) who underwent surgical treatment in the West China Hospital of Stomatology from 2016 to 2018, as well as 542 unaffected controls from Novogene WGS data with 39.89× coverage. Then, the first-round replication for 5 significant SNPs was conducted in 2,004 NSCL/P patients and 1,823 controls; later, the second-round replication for the most significant SNP was conducted in an independent sample of 1,729 patients with NSCL/P, 995 patients with nonsyndromic cleft palate (NSCPO), and 996 healthy controls; the affected individuals were from the West China Hospital of Stomatology from 2008 to 2016, while the healthy controls were from the West China Women's and Children's Hospital. All those recruited, both patients and healthy controls, were Han Chinese, as determined by self-report. More details are shown in Appendix Table 1 and Appendix Figure 2.

A signed informed consent was obtained from all patients or their guardians involved in this study. This study was approved by the Hospital Ethics Committee (HEC) of the West China Hospital of Stomatology, Sichuan University (WCHSIRB-D-2016-012R1) and conformed to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

In addition, genotyping data of Asian populations (Singapore, Taiwan, Philippines, Korea, and China) from GENEVA OFC data (Leslie et al. 2017), including 889 NSCL/P and 234 NSCPO trios, were also recruited in this study.

Next-Generation Sequencing Library Preparation and Sequencing

Based on the LD structure in CHB/JPT HapMap around *IRF6*, we selected chr1:209,923,704-210,382,027 (Hg19) (458.323 kb) as the targeted region, which comprised both the coding and noncoding regions (Appendix Fig. 3).

By using the Agilent liquid capture system (Agilent SureSelectXT Custom Kit), the next-generation sequencing (NGS) library for targeted sequencing was efficiently captured and enriched from 1.0 μg genomic DNA. To be specific, the genomic DNA sample was sheared into 180- to 280-bp fragments by Covaris crusher, and after end repair and A-tailing, indexing-specific adapters were ligated to both ends of the fragments to prepare a DNA library, which was then hybridized with biotin-labeled probes, and the targeted region was captured by streptomycin-containing magnetic beads. Later, the captured region was amplified by polymerase chain reaction (PCR) and subjected to library quality inspection, then sequenced on an Illumina Hiseq 4000 platform for paired-end 150-bp reads.

Genotyping in the Replication Studies

Genotyping of the SNPs selected for replication was performed by the SNPscan method (Genesky Biotechnologies), based on double ligation and multiplex fluorescence PCR. Duplicate analyses were performed to assess for genotyping quality by randomly choosing 4% of samples with high DNA quality.

Statistical Analysis

Hardy–Weinberg equilibrium (HWE) test, association analysis, sliding window analysis for haplotype association, and conditional logistic analysis were performed for SNP (with minor allele frequency [MAF] ≥0.05) by PLINK (Purcell et al. 2007). For the association analysis, we selected 5 significant SNPs from the discovery phase for further replication based on the following criteria: 1) minimum minor allele frequency (GnomAD_EAS_AF and 1000 genome_EAS)=0.3, and 2) based on the LD block from CHS&CHB in the 1000 Genomes database, tagging SNPs were chosen to make maximum coverage of all the significant SNPs. After Bonferroni multiple correction, the significance threshold of the *P* value for association analysis in the targeted sequencing phase (P_{trs}) was 4.59E-05 (1 phenotypic group * 1,090 SNPs), whereas P_{rep1} was 0.0033 (0.05 / [3 phenotypic groups * 5 SNPs]), P_{rep2} and P_{comb} was 0.0042 (0.05 / 12 phenotypic groups * 1 SNP), respectively.

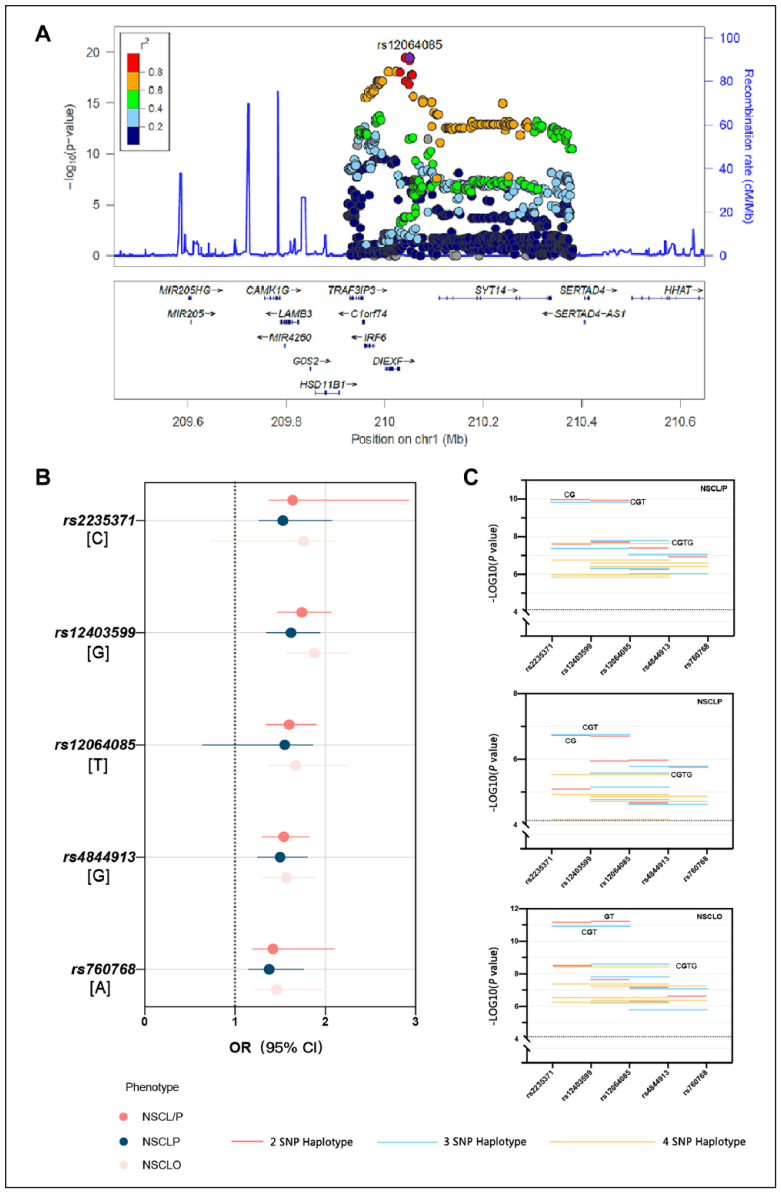

Figure 1. Targeted sequencing for the haplotype around *IRF6*. (**A**) Regional plot of association single-nucleotide polymorphisms (SNPs) and linkage disequilibrium patterns at the targeted region in discovery phase. (**B**) Association analysis of the selected SNPs in the first-round replication. The plot depicts the corresponding odds ratio and 95% confidence interval for the risk allele. (**C**) Results of sliding window analysis for haplotype association. The significance threshold is set as 2.59E-04 (0.05/193), and the dashed line shows the logarithm to base 10 of the significance threshold. NSCLO, nonsyndromic cleft lip only; NSCLP, nonsyndromic cleft lip and palate; NSCL/P, nonsyndromic cleft lip with or without palate.

Pairwise LD was calculated by Haploview, while the population attributable risk percentage (PAR%) was calculated based on the following formula: PAR%=[*P*e (RR − 1) / RR]×100 (*P*e is the prevalence of the allele in cases, and RR is the relative risk; odds ratio [OR] was used as a proxy for the RR).

Quantitative Real-Time PCR Analysis

Lip tissues from 54 NSCLO and 32 microform cleft lip (MCL) patients were collected when they were undergoing cleft lip repair surgery around 1 y of age. Total RNA of each sample was extracted using the RNA-easy Isolation Reagent (Vazyme). Reverse transcription to complementary DNA (cDNA) and further quantitative real-time PCR (RT-qPCR) were performed using the PrimeScript RT reagent kit and TB Green Fast qPCR mix (Takara), respectively (primers are shown in Appendix Table 2). Finally, relative expression of *IRF6* was analyzed by the $2^{-\Delta\Delta Ct}$ method with *GAPDH* used as reference gene for normalization.

Results

Targeted Sequencing for the Haplotype Around IRF6

In the initial phase of targeted sequencing, we found 1,263 single-nucleotide variants (SNVs) among 159 NSCL/P patients (Appendix Fig. 4).

First, we focused on deleterious low-frequency variants and enrolled them into gene-based burden analysis, but none of genes was associated with NSCL/P ($P>0.05$) (Appendix Table 3). Then, from those deleterious low-frequency variants, we tried to identify *de novo* variants. We did not discover any *de novo* deleterious variants but did find 2 novel deleterious variants in *C1orf74* (NM_152485: c.T323C:p.L108P) and *DIEXF* (NM_014388: c.T139G:p.C47G), both of which were inherited from one of their unaffected parents (Appendix Fig. 5). Neither of these 2 genes has been previously implicated in craniofacial development, so it is unclear if the variants are etiologic.

Switching to the common SNPs, we evaluated association with NSCL/P of 1,090 SNPs conforming to HWE ($P>0.000001$). Among these, 426 SNPs demonstrated a potential association ($P_{trs}<4.59E-05$) (Appendix Table 4). The relevant regional LD plot of *P* values is displayed in Figure 1A.

Replication of the Significant SNPs Identified in Association Analysis

Based on the *P* values and LD, we selected 5 SNPs (rs2235371, rs12403599, rs12064085, rs4844913, and rs760768) for replication in an independent cohort of 2,004 NSCL/P patients and 1,823 normal controls. All the SNPs passed the HWE test ($P>0.05$) (Appendix Table 5) and were included in the subsequent analysis. All 5 SNPs were significantly associated with NSCLO, NSCLP, and NSCL/P under both the allelic model and genotypic model (Appendix Tables 6 and 7) ($P_{rep1}<0.0042$). In each phenotype, the SNP with the lowest *P* value and highest OR was rs12403599 (Fig. 1B, Appendix Table 6).

To determine if the risk alleles of all 5 SNPs were inherited within the same haplotype block, we performed a pairwise linkage disequilibrium test. The results showed that rs12403599

Table. Association Analysis of rs12403599 among NSCL/P, NSCPO, and Controls.

	Replication 1			Replication 2[a]			Combined		
Phenotype	A1	OR (95% CI)	*P* Value	A1	OR (95% CI)	*P* Value	A1	OR (95% CI)	*P* Value
NSCL/P	G	1.74 (1.47–2.06)	1.22E-10	C	0.51 (0.46–0.57)	3.06E-36	G[b]	1.73 (1.57–1.91)	1.12E-53
NSCLP	G	1.62 (1.35–1.94)	1.94E-07	G	1.79 (1.55–2.07)	1.24E-15	G	1.58 (1.45–1.73)	3.37E-24
NSCLO	G	1.88 (1.57–2.25)	7.14E-12	G	2.02 (1.8–2.26)	6.35E-34	G	1.85 (1.70–2.00)	1.62E-51
BCL	G	2.19 (1.56–3.08)	4.86E-06	G	2.19 (1.66–2.9)	1.89E-08	G	2.07 (1.68–2.54)	2.58E-12
UCL	G	1.71 (1.42–2.07)	2.34E-08	G	2.02 (1.79–2.28)	3.54E-31	G	1.78 (1.63–1.94)	4.47E-40
RCL	G	1.80 (1.42–2.28)	1.09E-06	G	2.13 (1.8–2.51)	4.08E-19	G	1.88 (1.66–2.13)	1.55E-23
LCL	G	1.67 (1.36–2.04)	8.37E-07	G	1.96 (1.71–2.25)	3.65E-22	G	1.73 (1.56–1.91)	1.87E-27
CCL	G	1.14 (0.83–1.57)	0.4105	G	1.73 (1.25–2.38)	0.00081	G	1.29 (1.04–1.60)	0.01841
ICL	G	2.24 (1.78–2.81)	3.29E-12	G	2.36 (2.00–2.79)	2.13E-24	G	2.17 (1.93–2.45)	4.26E-37
MCL	G	2.82 (2.11–3.77)	9.28E-13	G	2.46 (1.95–3.11)	7.30E-15	G	2.48 (2.09–2.93)	5.06E-27
NSCPO	/	/	/	G	0.64 (0.57–0.72)	1.59E-12	/	/	/
HSCP	/	/	/	G	0.66 (0.57–0.76)	7.31E-09	/	/	/

A1, minor allele; BCL, bilateral cleft lip; CCL, complete cleft lip; CI, confidence interval; HSCP, hard/soft cleft palate; ICL, incomplete cleft lip; LCL, left-side cleft lip; MCL, microform cleft lip; NSCLP, nonsyndromic cleft lip and palate; NSCL/P, nonsyndromic cleft lip with or without palate; NSCPO, nonsyndromic cleft palate only; OR, odds ratio; RCL, right-side cleft lip; UCL, unilateral cleft lip; /, data is unavailable.
[a]This replication was specifically for rs12403599. The G allele frequency of rs12403599 in NSCPO is 30.70% and 40.89% in unaffected controls.
[b]Minor allele is C.

was linked with rs2235371 among NSCL/P ($D'=0.99$, $r^2=0.49$), NSCLP ($D'=0.99$, $r^2=0.49$), and NSCLO ($D'=0.99$, $r^2=0.49$) individuals (Appendix Fig. 6).

Regardless of whether haplotypes are composed of 2, 3, or 4 alleles, the haplotypes most strongly associated with each phenotype always contained the "G" allele of rs12403599 and have a higher frequency in cases than that in controls. For example, the *P* value of the CG haplotype of rs2235371 and rs12403599 in NSCL/P is 1.09E-10 (F_A=54.55%, F_U=40.76%); the *P* value of the CGT haplotype of rs2235371, rs12403599, and rs12064085 in NSCLP is 1.75E-07 (F_A=4.17%, F_U=4.06%); and the *P* value of the GT haplotype of rs12403599 and rs12064085 in NSCLO is 5.46E-12 (F_A=52.59%, F_U=36.87%) (Fig. 1C, Appendix Table 8).

Next, as the risk alleles of rs2235371 and rs12403599 are frequently but not always inherited together, we conducted a conditional analysis using a logistic model. Without adding rs2235371 as a covariate, rs12403599 was associated with NSCLO (*P*=4.34E-05; OR=1.71; 95% confidence interval [CI], 1.32–2.20) and NSCL/P (*P*=0.0002; OR=1.56; 95% CI, 1.23–1.97). On the other hand, without adding rs12403599 as a covariate, rs2235371 did not show significance, whereas by adding rs12403599, rs2235371 showed significance in NSCLO and NSCL/P ($P<0.0033$), suggesting that rs12403599 is tagging the risk haplotype for nonsyndromic orofacial cleft much better than rs2235371 in Han Chinese population (Appendix Table 9).

Further Replication of rs12403599 among NSCL/P and NSCPO

We next tested the association between rs12403599 and NSCL/P, as well as NSCPO, in a third cohort, composed of 1,729 NSCL/P patients, 995 NSCPO patients, and 996 healthy controls. As predicted, rs12403599 was associated with all the phenotypes of NSCL/P ($P_{rep2}<0.0042$), NSCPO ($P_{rep2}<0.0042$), and NSCPO subphenotype: hard/soft cleft palate (HSCP) ($P_{rep2}<0.0042$) (Table). Interestingly, while the G allele of rs12403599 was associated with risk for NSCL/P and its subphenotypes ($OR>1$), this allele was the protective (i.e., non-risk associated) allele in NSCPO and its subphenotype (Table).

rs12403599 Is a Novel Susceptibility SNP and Plays a Substantial Role in NSCL/P

To increase the statistical power, we combined the sample sets of replication 1 and replication 2 and reanalyzed whether rs12403599 was associated with NSCL/P, NSCLP, and NSCLO. Since rs12403599 had the highest OR in NSCLO ($OR_{comb\text{-}NSCLO}=1.85$; $OR_{comb\text{-}NSCL/P}=1.73$; $OR_{comb\text{-}NSCLP}=1.58$) (Fig. 2A, Table), we next evaluated the association of rs12403599 with multiple subphenotypes of NSCLO, including bilateral cleft lip (BCL), unilateral cleft lip (UCL), right-side cleft lip (RCL), left-side cleft lip (LCL), complete cleft lip (CCL), incomplete cleft lip (ICL), and MCL. Among all these subphenotypes, association between rs12403599 and CCL had the lowest OR ($P_{comb}=0.01841$; $OR_{comb}=1.29$; 95% CI_{comb}, 1.04–1.60), whereas the highest OR was from MCL ($P_{comb}=5.06E\text{-}27$; $OR_{comb}=2.48$; 95% CI_{comb}, 2.09–2.93) (Fig. 2A, Table). According to the effect size, the G allele of rs12403599 (or SNP in strong LD with it) appears to confer over 13% of genetic contribution (i.e., heritable risk) for each phenotypic group: NSCL/P (23.59%), NSCLP (19.69%), NSCLO (26.41%), BCL (31.14%), UCL (24.80%), RCL (27.10%), LCL (23.58%), CCL (26.21%), ICL (13.81%), and MCL (38.47%) (Appendix Table 10). Furthermore, we detected a possible dosage effect for the G allele of rs12403599 in the genotypic model. The OR of the G/C genotype was 1.77 versus

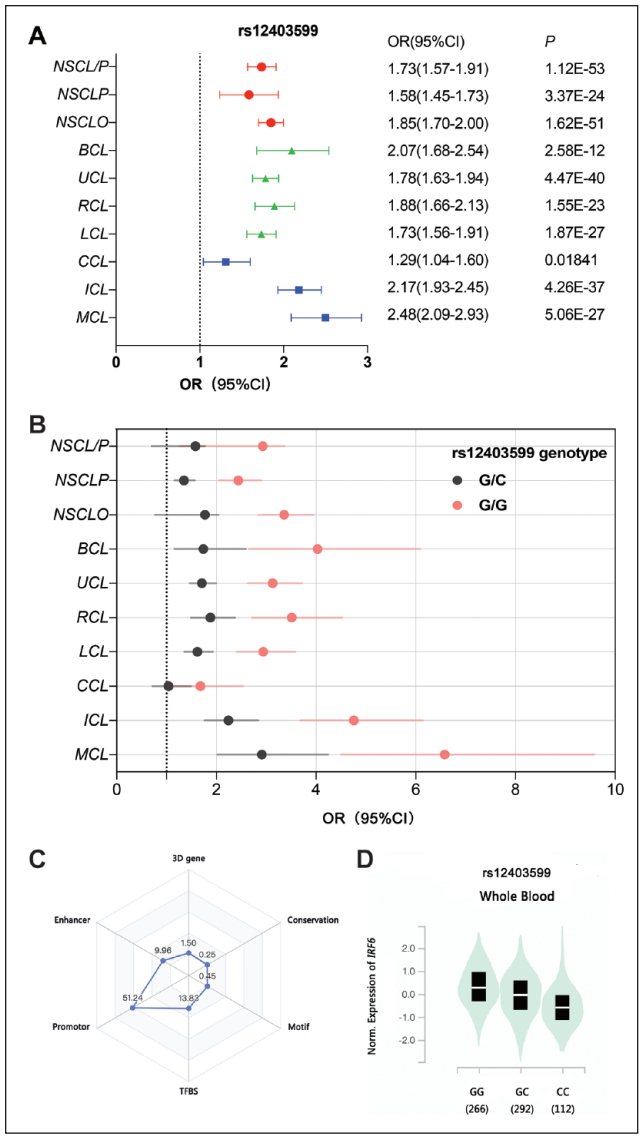

Figure 2. Statistical analysis and Functional prediction of rs12403599. (**A**) The forest plot displays the combination of 2 rounds of association analyses for rs12403599. (**B**) The genotypic odds ratio (OR) for cleft phenotypic subsets and rs12403599. (**C**) Annotation for rs12403599 by 3-dimensional SNP. (**D**) Quantitative trait locus result of rs12403599 in whole blood. BCL, bilateral cleft lip; CCL, complete cleft lip; ICL, incomplete cleft lip; LCL, left-side cleft lip; MCL, microform cleft lip; NSCLO, nonsyndromic cleft lip only; NSCLP, nonsyndromic cleft lip and palate; NSCL/P, nonsyndromic cleft lip with or without palate; RCL, right-side cleft lip; UCL, unilateral cleft lip.

3.36 for the G/G genotype in NSCLO and 2.91 versus 6.58 in MCL (Fig. 2B, Appendix Table 11).

Functional Analysis of rs12403599

We next sought to determine if rs12403599 is functional. Three-dimensional (3D) SNP annotation revealed that rs12403599 lies in 21 transcription factor binding sites. In 26 cell types, rs12403599 is within chromatin region that appears to have enhancer function. In 53 cell types, including normal human embryonic keratinocytes (NHEKs), it lies in a promoter state (ENCODE data). In addition, rs12403599 is linked to 7 genes via 3D chromatin interactions. Overall, the corresponding score is highest in the "promoter" category (Fig. 2C). According to the GTEx database, rs12403599 is an expression quantitative trait locus (eQTL); in whole blood, the G allele is associated with higher expression of *IRF6* (Fig. 2D).

We used reporter assays to determine if the allele of rs12403599 alters activity of the *IRF6* promoter. We engineered a 997-bp element centered on rs12403599 and harboring either the G or C allele of rs12403599 into a promoterless vector containing cDNA encoding firefly luciferase. We transfected these plasmids (separately) with a transfection control plasmid into fetal oral epithelium cells (GMSM-K), as *IRF6* is expressed in embryonic oral epithelium (Biggs et al. 2014). The reporter construct of the risk-allele G exhibited greater promoter activity than the one harboring the non-risk allele C (Fig. 3A).

Finally, we used RT-qPCR to evaluate *IRF6* expression levels in lip tissues isolated from patients with NSCLO or MCL and correlated this with the rs12403599 genotype. Tissues with the G/G genotype had higher average *IRF6* expression level than those with the C/C genotype. *IRF6* expression level in MCL patients was greater than that in NSCLO patients, potentially reflecting a slightly different anatomic location of the excised tissue (Fig. 3B). Together, the results support rs12403599 as an etiologic variant whose G allele, which is a risk for NSCL/P and protective for NSCPO, enhances *IRF6* promoter activity, leading to relatively higher *IRF6* expression level.

Discussion

To deepen the understanding of the etiology of NSCL/P, this study was designed with 2 objectives. The first was to identify additional variants in the Han Chinese population associated with NSCL/P near the *IRF6* gene. The second was to identify the functional SNP or SNP in the same haplotype block as rs2235371, a lead SNP near *IRF6* detected in a GWAS for NSCL/P in the Han Chinese population (Sun et al. 2015) and shown to be associated with NSCL/P in a mixed-ethnicity cohort (Zucchero et al. 2004).

We discovered 2 novel deleterious rare coding variants in genes neighboring *IRF6*, *C1orf74*, and *UTP25/DIEXF*. These genes have not been previously implicated in craniofacial development, although homozygous deletion of *UTP25/DIEXF* is embryonic lethal in mice (Lundborg 1989). Further research is necessary to determine if these coding and noncoding variants around *IRF6* (e.g., in *IRF6* enhancer regions) contribute to NSCL/P etiology.

As for the common variants, using stringent thresholds of 4.59E-05 derived from multiple corrections, a total of 426 SNPs were found to be significantly associated with NSCL/P in the discovery phase. Among these, rs2235371 continued to show a significant association with NSCL/P in the replication phase among large samples.

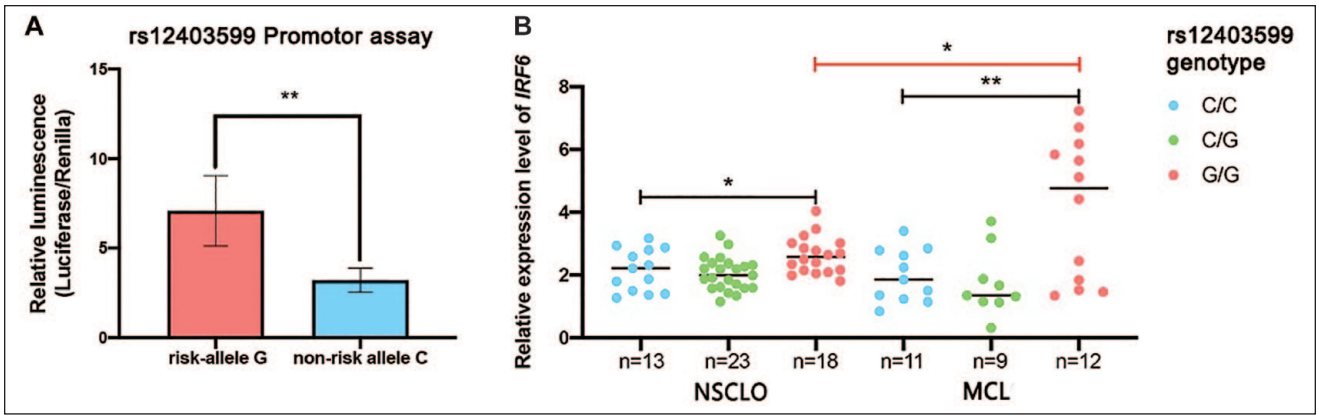

Figure 3. Functional analysis of rs12403599. (**A**) Result of the dual Luciferase reporter assay. (**B**) Genotype-specific quantitative real-time polymerase chain reaction analysis of lip tissue from patients with nonsyndromic cleft lip only (NSCLO) and microform cleft lip (MCL). **P* < 0.05, ***P* < 0.01.

Intriguingly, in the replication phase, the minor allele of rs12403599 was the risk allele (OR > 1), whereas that of rs2235371 was protective (OR < 1) for all phenotypes, and rs12403599 had the lowest *P* value and highest OR in NSCLO compared with NSCL/P and NSCLP. Since evidence of genetic heterogeneity is observed among the subphenotypes of NSCLO (Carlson et al. 2017; Yu et al. 2017; Zhang and Arneja 2017; Huang et al. 2019; Yin et al. 2021), we examined the associations and calculated the PAR% of rs12403599 within NSCLO subphenotypes, and the results further confirmed the genetic heterogeneity.

Because rs12403599 was associated with NSCLO and its subphenotypes (OR > 1), we were not surprised that it was also associated with NSCPO and its subphenotype HSCP. It was unexpected, however, that the G allele of rs12403599 that confers risk for NSCLO is protective for NSCPO. This finding was further confirmed among Asian populations (889 NSCL/P and 234 NSCPO trios), with the G allele of rs12403599 being a risk for NSCL/P (*P* = 2.09E-14; OR = 0.58; 95% CI, 0.51–0.68) but protective for NSCPO (*P* = 0.002755; OR = 0.66; 95% CI, 0.5–0.87). As for the opposite effects of direction, similar findings were also reported on the other genes, with rs1339063 (*PAX7*), rs12632559 (*DLG1*), rs9291207 (*LIMCH1*), rs4422437 (*SHROOM3*), and rs4794658 (*NOG*) appearing to be associated with both CL/P and CPO, but opposite directions were observed according to the relative risk; also, rs11083400 (*MIR302F*), rs11191818 (*SH3PXD2A*), rs2802532 (*MIR137HG*), and rs72741048 (*IRF6*) were associated with both CLO and CPO but in opposite directions (Huang et al. 2019; Ray et al. 2021). This might be attributed to the temporally and spatially separate morphogenesis of the lip and palate, as well as distinct epidemiologic patterns of CLO and CPO. Nevertheless, whether the *IRF6* expression is relevant risk for CPO and CLO in embryonic tissues is not certain. In the present study, the G allele of rs12403599 is associated with higher expression of *IRF6* in the postnatal tissues we examined and in whole blood (GTEx database). Our previous GWAS and other studies have detected similar trend before (Ingraham et al. 2006; Iwata et al. 2013; Huang et al. 2019), but more investigation of this topic is clearly warranted.

rs12403599 lies in noncoding DNA and is therefore likely to affect the function of the *cis* regulatory elements (CREs) (e.g., enhancers and promoters), which strictly regulate the spatiotemporal expression of genes (Chatterjee and Ahituv 2017; Nowosad et al. 2020). SNPs that potentially affect the function of CREs have been reported to be involved in the occurrence of multiple diseases, including cancer and leukemia (Patel et al. 2014; Liu et al. 2017; Hua et al. 2018; Liu et al. 2020). Moreover, an orofacial cleft–associated SNP that has allele-specific effects on an oral epithelium enhancer that contacts the promoter of *ARHGAP29* may directly affect the risk for NSCL/P (Liu et al. 2017). rs12403599 is within 1 kb of the *IRF6* transcription start site, and in NHEKs, the region it lies in harbors chromatin marks indicative of promoter function (ENCODE data). Our reporter studies suggest rs12403599 is a functional SNP, with the G allele driving stronger promoter activity, and the G allele is associated with higher *IRF6* expression in patient-derived oral mucosal tissue. A linkage disequilibrium test indicated that rs2235371 and rs12403599 are within the same haplotype block, and a conditional analysis indicated that the association of rs2235371 with NSCL/P falls away if rs12403599 is excluded. rs642961, which is also in LD with rs2235371, was reported to account for an 18% attributable risk for NSCLO (Rahimov et al. 2008). Here we have found an even higher attributable risk of rs12403599 for NSCLO at 26.41% and higher still for MCL (38.47%). We conclude that rs12403599 is a strong candidate as a pathogenic genetic variant for NSCL/P, and rs12403599 tags the risk haplotype for NSCL/P much better than rs2235371 in Han Chinese.

Conclusion

Via targeted sequencing on LD structure around *IRF6* and further association analysis, we here identify rs12403599, an SNP in the *IRF6* promoter, as being strongly associated with all phenotypes of NSCL/P, particularly NSCLO and its subphenotype MCL. Dual luciferase report assay and RT-qPCR analysis supported the SNP being functional, and the conditional logistic analysis indicated that rs12403599 is the main variant tagging the risk haplotype in Han Chinese, but the joint functional role

of rs12403599 and rs2235371 remains unknown. In addition, rs12403599 conferred a greater attributable risk for NSCLO than rs642961. Thus, we suggest that rs12403599 is a directly functional SNP in an NSCL/P-associated haplotype near *IRF6* (Appendix Fig. 7). The current study was conducted in the Han Chinese population. It will be interesting to determine if the association between rs12403599 and NSCL/P is present in different populations. In addition, the specific mechanism behind its regulation of *IRF6* expression should be further explored.

Author Contributions

M.-J. Li, contributed to acquisition and data interpretation, drafted and critically revised the manuscript; P. Kumari, contributed to data interpretation, critically revised the manuscript; Y.-S. Lin, contributed to acquisition and data interpretation, critically revised the manuscript; M.-L. Yao, B.-H. Zhang, B. Yin, S.-J. Duan, M.L. Marazita, contributed to data acquisition, critically revised the manuscript; R.A. Cornell, B. Shi, contributed to data conception, critically revised the manuscript; Z.-l. Jia, contributed to conception and design, data acquisition, analysis, and interpretation, drafted and critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.

Acknowledgments

The authors thank all the participants who donated samples in this study; Professor Jeffrey C. Murray at the University of Iowa and Professor Mary L. Marazita at the University of Pittsburg for sharing the GENEVA OFC data; and Jeffrey C. Murray for help with English editing.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by the National Science Funds of China (No. 81600849 and No. 81271118) and National Institutes of Health R01 DE027362 (R.A. Cornell).

ORCID iDs

P. Kumari https://orcid.org/0000-0002-7210-172X
M.-L. Yao https://orcid.org/0000-0002-2429-1709
Z.-l. Jia https://orcid.org/0000-0002-0129-0367

References

Beaty TH, Marazita ML, Leslie EJ. 2016. Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. F1000Research. 5:2800.

Biggs LC, Naridze RL, DeMali KA, Lusche DF, Kuhl S, Soll DR, Schutte BC, Dunnwald M. 2014. Interferon regulatory factor 6 regulates keratinocyte migration. J Cell Sci.127(Pt 13):2840–2848.

Carlson JC, Taub MA, Feingold E, Beaty TH, Murray JC, Marazita ML, Leslie EJ. 2017. Identifying genetic sources of phenotypic heterogeneity in orofacial clefts by targeted sequencing. Birth Defects Res. 109(13):1030–1038.

Chang D, Keinan A. 2012. Predicting signatures of "synthetic associations" and "natural associations" from empirical patterns of human genetic variation. PLoS Comput Bio. 8(7):e1002600.

Chatterjee S, Ahituv N. 2017. Gene regulatory elements, major drivers of human disease. Ann Rev Genomics Hum Genet. 18:45–63.

Christensen K, Juel K, Herskind AM, Murray JC. 2004. Long term follow up study of survival associated with cleft lip and palate at birth. BMJ. 328(7453):1405.

Dai L, Zhu J, Mao M, Li Y, Deng Y, Wang Y, Liang J, Tang L, Wang H, Kilfoy BA, et al. 2010. Time trends in oral clefts in Chinese newborns: data from the Chinese National Birth Defects monitoring network. Birth Defects Res A Clin Mol Teratol. 88(1):41–47.

de Souza LT, Kowalski TW, Ferrari J, Monlléo IL, Ribeiro EM, de Souza J, Fett-Conte AC, de Araujo TK, Gil-da-Silva-Lopes VL, Ribeiro-Dos-Santos ÂK, et al. 2016. Study of IRF6 and 8q24 region in non-syndromic oral clefts in the Brazilian population. Oral Dis. 22(3):241–245.

Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. 2010. Rare variants create synthetic genome-wide associations. PLoS Biol. 8(1):e1000294.

Dixon MJ, Marazita ML, Beaty TH, Murray JC. 2011. Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet. 12(3):167–178.

Grosen D, Bille C, Petersen I, Skytthe A, von Bornemann Hjelmborg J, Pedersen JK, Murray JC, Christensen K. 2011. Risk of oral clefts in twins. Epidemiology. 22(3):313–319.

Hua JT, Ahmed M, Guo H, Zhang Y, Chen S, Soares F, Lu J, Zhou S, Wang M, Li H, et al. 2018. Risk SNP-mediated promoter-enhancer switching drives prostate cancer through lncRNA PCAT19. Cell. 174(3): 564–575.

Huang L, Jia Z, Shi Y, Du Q, Shi J, Wang Z, Mou Y, Wang Q, Zhang B, Wang Q, et al. 2019. Genetic factors define CPO and CLO subtypes of nonsyndromicorofacial cleft. PLoS Genet. 15(10):e1008357.

Ingraham CR, Kinoshita A, Kondo S, Yang B, Sajan S, Trout KJ, Malik MI, Dunnwald M, Goudy SL, Lovett M, et al. 2006. Abnormal skin, limb and craniofacial morphogenesis in mice deficient for interferon regulatory factor 6 (Irf6). Nat Genet. 38(11):1335–1340.

Iwata J, Suzuki A, Pelikan RC, Ho TV, Sanchez-Lara PA, Urata M, Dixon MJ, Chai Y. 2013. Smad4-Irf6 genetic interaction and TGFβ-mediated IRF6 signaling cascade are crucial for palatal fusion in mice. Development. 140(6):1220–1230.

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, et al. 2002. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet. 32(2):285–289.

Leslie EJ, Carlson JC, Shaffer JR, Butali A, Buxó CJ, Castilla EE, Christensen K, Deleyiannis FW, Leigh Field L, Hecht JT, et al. 2017. Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. Hum Genet. 136(3):275–286.

Li EB, Truong D, Hallett SA, Mukherjee K, Schutte BC, Liao EC. 2017. Rapid functional analysis of computationally complex rare human IRF6 gene variants using a novel zebrafish model. PLoS Genet. 13(9):e1007009.

Liu H, Duncan K, Helverson A, Kumari P, Mumm C, Xiao Y, Carlson JC, Darbellay F, Visel A, Leslie E, et al. 2020. Analysis of zebrafish periderm enhancers facilitates identification of a regulatory variant near human KRT8/18. Elife. 9:e51325.

Liu H, Leslie EJ, Carlson JC, Beaty TH, Marazita ML, Lidral AC, Cornell RA. 2017. Identification of common non-coding variants at 1p22 that are functional for non-syndromic orofacial clefting. Nat Commun. 8:14759.

Lundborg G. 1989. Experimental research in hand surgery. J Hand Surg Br. 14(1):3–5.

Nowosad K, Hordyjewska-Kowalczyk E, Tylzanowski P. 2020. Mutations in gene regulatory elements linked to human limb malformations. J Med Genet. 57(6):361–370.

Park JW, McIntosh I, Hetmanski JB, Jabs EW, Vander Kolk CA, Wu-Chou YH, Chen PK, Chong SS, Yeow V, Jee SH, et al. 2007. Association between IRF6 and nonsyndromic cleft lip with or without cleft palate in four populations. Genet Med. 9(4):219–227.

Patel B, Kang Y, Cui K, Litt M, Riberio MS, Deng C, Salz T, Casada S, Fu X, Qiu Y, et al. 2014. Aberrant TAL1 activation is mediated by an interchromosomal interaction in human T-cell acute lymphoblastic leukemia. Leukemia. 28(2):349–361.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81(3):559–575.

Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C, et al. 2008. Disruption of an AP-2α binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet. 40(11):1341–1347.

Ray D, Venkataraghavan S, Zhang W, Leslie EJ, Hetmanski JB, Weinberg SM, Murray JC, Marazita ML, Ruczinski I, Taub MA, et al. 2021. Pleiotropy method reveals genetic overlap between orofacial clefts at multiple novel loci from GWAS of multi-ethnic trios. PLoS Genet. 17(7):e1009584.

Sun Y, Huang Y, Yin A, Pan Y, Wang Y, Wang C, Du Y, Wang M, Lan F, Hu Z, et al. 2015. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun. 6:6414.

Sun X, Namkung J, Zhu X, Elston RC. 2011. Capability of common SNPs to tag rare variants. BMC Proc. 5(Suppl 9):S88.

Wehby G, Cassell CH. 2010. The impact of orofacial clefts on quality of life and healthcare use and costs. Oral Dis. 16(1):3–10.

Wu-Chou YH, Lu YC, Chen KP, Chang HF, Lin YT, Lo LJ. 2019. Association studies between regulatory regions of IRF6/TP63 genes and nonsyndromic oral clefts. Cleft Palate Craniofac J. 56(6):778–785.

Yin B, Shi JY, Lin YS, Shi B, Jia ZL. 2021. SNPs at TP63 gene was specifically associated with right-side cleft lip in Han Chinese population. Oral Dis. 27(3):559–566.

Yu Y, Zuo X, He M, Gao J, Fu Y, Qin C, Meng L, Wang W, Song Y, Cheng Y, et al. 2017. Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nat Commun. 8:14364.

Zhang JX, Arneja JS. 2017. Evidence-based medicine: the bilateral cleft lip repair. Plast Reconstr Surg. 140(1):152e–165e.

Zhou Q, Li M, Zhu W, Guo J, Wang Y, Li Y, Li S. 2013. Association between interferon regulatory factor 6 gene polymorphisms and nonsyndromic cleft lip with or without cleft palate in a Chinese population. Cleft Palate Craniofac J. 50(5):570–576.

Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J, et al. 2004. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. N Engl J Med. 351(8):769–780.