Evaluating SKI as a candidate gene for non-syndromic cleft lip with or without cleft palate


Non-syndromic cleft lip with or without cleft palate (NSCL/P) is one of the most common of all congenital malformations and has a multifactorial etiology. Findings in mice suggest that the v-ski sarcoma viral oncogene homolog (SKI) gene is a candidate gene for orofacial clefing. In humans, a significant association between rs2843159 within SKI and NSCL/P has been reported in patients from the Philippines and South America. In the South American patients, the association was driven by the subgroup of patients with non-syndromic cleft lip only (NSCLO). Here we investigated the association with rs2843159 in a Mayan Mesoamerican population (172 NSCL/P patients and 366 controls). In addition, we analyzed the phenotypic subgroups NSCLO and non-syndromic cleft of lip and palate (NSCLP). A trend towards association between rs2843159 and NSCL/P was observed in the Mayan cohort (P = 0.097), and we found a stronger association in the NSCLP subgroup (P = 0.072) despite a limited sample size. To investigate whether other common variants within the SKI gene contribute to NSCL/P susceptibility in European and Asian populations, we also analyzed genotypic data from two recent genome-wide association studies using set-based statistical approaches. These analyses detected a trend toward association in the European population. Our data provide limited support for the hypothesis that common SKI variants are susceptibility factors for NSCL/P.

Orofacial clefts are a common type of congenital malformation and represent a major public health challenge in terms of patient care. The estimated worldwide prevalence of orofacial clefts among newborns is 1 in 600 (1). Cleft lip with or without cleft palate (CL/P) is the most frequent type of orofacial clefting. Although CL/P may occur as part of a complex malformation syndrome, most cases are non-syndromic and have a multifactorial etiology that includes both genetic and environmental factors (2). In terms of genetic factors, conclusive evidence for involvement in non-syndromic CL/P (NSCL/P) has been obtained for the interferon regulatory factor 6 (IRF6) gene and for five chromosomal loci identified through genome-wide association studies (GWAS) (2). There is some indication that the strength of the associations might vary between populations, and not all loci identified by GWAS appear to be involved in both European and Asian populations (3). However, these susceptibility loci explain a considerable proportion of all NSCL/P cases. Nevertheless, additional risk variants still await identification.

The v-ski sarcoma viral oncogene homolog (SKI) gene was first implicated as a candidate gene for orofacial clefting by Berk et al. (4). Their study showed that Ski−/− murine mutants display a complex phenotype that includes severe defects in the patterning of
vertebral and craniofacial skeletal structures. The occurrence of encephaly or midline facial clefting, but never both, in these mice suggests that the genetic background influences the penetrance of the mutation (4). This hypothesis was confirmed in a subsequent analysis of Ski−/− mice, which showed that the incidence of facial clefting increased from generation to generation when mutated mice were backcrossed (5). Notably, the human SKI gene maps to the genomic region that is commonly deleted in the 1p36 deletion syndrome. The phenotypic spectrum of this syndrome includes orofacial clefting (6).

On the basis of this evidence, the SKI gene has been included in a comprehensive resequencing and linkage study of multiple clefting candidate genes (7). A highly significant association between a common single-nucleotide polymorphism (SNP) within SKI (rs2843159), and NSCL/P was reported. This association was found predominantly in patients from the Philippines \( (P = 0.000004) \) and was replicated in a South American NSCL/P sample \( (P = 0.02) \).

Epidemiologic and embryologic data support the hypothesis that NSCL/P can be subdivided into non-syndromic cleft lip only (NSCLO) and non-syndromic cleft of lip and palate (NSCLP), the latter affecting both lip and palate, and that distinct genetic factors may be involved in their respective development (2). Interestingly, in the above-mentioned South American NSCL/P sample, the association with the SKI gene was driven by the subgroup of patients with NSCLO \( (P = 0.004) \).

A strong Native American maternal contribution has been reported in the South American clefting population (8). This finding, together with the observation of variable penetrance depending on the genetic background, suggests that a replication attempt in a Native American cleft sample is warranted.

The aim of the present study was to analyze the association with rs2843159 in a case–control cohort of Mayan descent. To investigate whether other common variants within the SKI gene might contribute to NSCL/P susceptibility in European and Asian populations, we also extracted genotypic data from two GWAS \( (9, 10) \) and analyzed them using set-based statistical approaches.

Material and methods

Mayan sample

The study included 172 NSCL/P patients (152 with NSCLP and 20 with NSCLO; 111 male patients and 61 female patients) and 366 unaffected controls (127 male subjects and 239 female subjects). Patients and controls were recruited from confined areas of San Cristóbal de las Casas and Tuxtla Gutiérrez in the State of Chiapas (Mexico). Most patients were ascertained within the context of surgical outreach programs. A detailed patient questionnaire was completed to identify possible prenatal contributory factors, such as maternal ingestion of known teratogenic medications or toxins, and no affected families were identified. Unaffected controls were recruited at two outpatient clinics, which they were attending for various medical indications. The ethnic background of patients and controls was assessed on the basis of the grandparents’ descent. Only those individuals whose four grandparents were all of Mayan descent (Mestizo; Mestizo admixed with Chol, Tzotzil, Tzeltal, Mam; or Chol, Tzotzil, Tzeltal, Mam) were included. We were not aware of any specific environmental risk factors or protective factors in the controls.

Ethical approval for the study was obtained from the respective Medical Faculty Ethics Committees, and all individuals provided written informed consent. For children <18 yr of age, consent was obtained from their parents. All participating individuals were clinically assessed by one of four medical geneticists (A.R.M., H.R., O.C.C., or S.N.) to determine cleft status and, if affected, to exclude any underlying syndrome. Patients with cleft palate only (CPO) were excluded from the present study. None of the controls displayed orofacial clefting or any minor form of clefting, such as small defects of the lip or alveolar arch, or scarlike ridges above the lip or submucosal cleft palate.

Genomic DNA preparation

A whole-blood sample of 10–30 ml was collected into ethylenediaminetetraacetic acid (EDTA)-coated vacutainer tubes according to standard venipuncture methods. Blood was drawn from all affected individuals, from their parents (if possible), and from all controls. Genomic DNA was extracted from peripheral leukocytes using the salting-out method and a standardized genomic DNA isolation protocol.

Genotyping

Genotyping was performed using restriction fragment length polymorphism (RFLP) analysis. The SNP rs2843159 (NM_003036.3:c.1475-60T>C) was genotyped after initial amplification of the SKI intron 4 by PCR. The PCR was performed using the FastStart Taq DNA Polymerase kit (Roche Applied Science, Indianapolis, IN, USA) under standard conditions and with the primers SKI_intron4_F: 5’-TCCTACAAGACGTTTGAGACG-3’ and SKI_intron4_R: 5’-GAGCTGGATGAGGTAAAG-GAC-3’, yielding a product length of 498 bp. The PCR product was then subjected to restriction digestion with BsrNI (restriction site: CCWGG; New England Biolabs, Ipswich, MA, USA). Based on human reference sequence data, two BsrNI restriction sites are present in the amplicon, yielding restriction fragments of 112, 165, and 221 bp. In the presence of the T allele, one restriction site is lost, yielding fragments that are 165 and 333 bp in size. The digested PCR products were separated by electrophoresis through 3% agarose gels, and genotypes were assigned by two independent investigators. Detailed information of PCR amplification and the genotyping procedure are obtainable on request.

In-silico set-based tests on GWAS data

The GWAS data were available from (i) our own case–control GWAS of 399 patients and 1,318 controls of European descent (10) and (ii) the publically available database.
Analyses of the subgroups NSCLO and NSCLP revealed evidence for a stronger effect of rs2843159 in patients with NSCLP (Table 1). Compared with the overall analysis of NSCL/P, the P-value in the NSCLP subgroup decreased to 0.072, despite the smaller sample size. No significant association, or trend toward association, was observed in the NSCLO group (Table 1).

**Set-based analysis of GWAS data**

The results of the gene-based analysis for SKI are shown in Table 2. A trend toward association was observed in the European population in both the case–control data (P = 0.0612) and the trio (P = 0.108) data. The lowest P-value was observed for rs9039160 – an intronic variant within SKI – in the European case-control data (P = 0.00633). In the family-based data set the lowest P-value was observed for rs4648831 (P = 0.00592), which is located downstream of SKI. However, these single SNP results did not remain significant after Bonferroni correction for the number of included SNPs (n = 41).

**Discussion**

We followed up on previous findings reporting strong evidence for an association between rs2843159, a SNP within the SKI gene, and NSCL/P (7). The association was identified in two independent samples of Filipino and South American subjects. As research suggests that there is a strong maternal Native American contribution to clefting in samples from South America (8), we investigated the role of this common variant in a Native American case-control cohort of Mayan descent.

Although we observed a trend towards association for rs2843159 in the NSCL/P group, the P-value fell short of nominal significance. To address whether this may be attributed to power issues, we recalculated the power for the present study assuming an OR of 1.2, as observed in the original report (7). This analysis yielded a power of 28.5%, which supports the hypothesis that the failure to reach statistical significance was a result of the type 2 error.
of the limited sample size. Notably, the association observed in the present study is in the same allelic direction as that reported by Vieira et al., with the T allele being more frequently observed in cases. This further supports the possible contribution of rs2843159 to NSCL/P in samples from South America and Middle America.

When analyzing the NSCL/P subgroups NSCLO and NSCLP, the P-value in the NSCL/P group was smaller compared with that for the entire NSCL/P sample and also smaller compared with that for the NSCLO group. This is in contrast to the initial findings in the South American population, where the effect of rs2843159 was mainly driven by the NSCLO subgroup. This might be explained by the limited sample size (especially for the NSCLO subgroup) in the present study. Alternatively, it might reflect the variable phenotypic expressivity that has already been demonstrated in Sk1−/− mice (5). Of note, a recent meta-analysis of GWAS data suggests that distinct genetic factors contribute to the NSCL/P subgroups NSCLO and NSCLP, further suggesting that subphenotyping is important in future genetic analysis in the field of orofacial clefting (3).

We also investigated whether common variants contribute to NSCL/P in other populations. A previous analysis of rs2843159 in a case–control sample of European ethnicity from Iowa yielded no positive findings (7). However, this study might have failed as a result of genetic heterogeneity at the SKI locus, if our hypothesis that other common variants within SKI are of etiological relevance to NSCL/P is correct. To address this question we performed a gene-based analysis of recent genome-wide data from European and Asian NSCL/P patients. The results from the European case–control data provided scant support for our hypothesis, as although the gene-based P-value approached nominal significance, it failed to reach the P = 0.05 criterion. The most strongly associated SNP in this data set, rs903916, which is located in intron 1 of SKI, showed a P-value of 0.00633, which did not withstand correction for multiple testing for the number of included SNPs.

Table 2
Gene-based test of the v-ski sarcoma viral oncogene homolog (SKI) gene in data sets from genome-wide association studies (10)

<table>
<thead>
<tr>
<th>Sample (study)</th>
<th>Number of SNPs</th>
<th>Number of simulations*</th>
<th>Gene-based test statistics</th>
<th>Gene-based P-value</th>
<th>Best SNP</th>
<th>Best SNP P-value</th>
<th>Location relative to SKI</th>
</tr>
</thead>
<tbody>
<tr>
<td>European case–control</td>
<td>41</td>
<td>100,000</td>
<td>71.48</td>
<td>0.0612</td>
<td>rs903916</td>
<td>0.00633</td>
<td>Intron 1</td>
</tr>
<tr>
<td>European case–parent</td>
<td>42</td>
<td>1,000</td>
<td>63.06</td>
<td>0.108</td>
<td>rs4648831</td>
<td>0.00592</td>
<td>Downstream of SKI intronic of MORN1</td>
</tr>
<tr>
<td>Asian case–parent</td>
<td>42</td>
<td>1,000</td>
<td>46.48</td>
<td>0.327</td>
<td>rs263526</td>
<td>0.0234</td>
<td>Upstream</td>
</tr>
</tbody>
</table>

Membrane Occupation and Recognition Nexus 1 gene (MORN1); SNP, single nucleotide polymorphism.

*VEGAS increases the number of simulations, depending on the result of the previous round of simulations.

No additional support was provided from either the European or the Asian family-based data sets. In both of these analyses, the gene-based P-values were statistically non-significant and the most significant SNPs were located outside the genomic region of SKI. Although our results do not demonstrate a strong contribution of common variants within SKI in European and Asian samples, subtle effects of SKI SNPs cannot be completely ruled out. The lack of significance might be a result of the limited power of the samples. Alternatively, it might be that none of the SNPs present in the arrays is in strong linkage disequilibrium (LD) with a putative causal variant. In this case, the presence of some SNPs with nominal significance, either within or outside the SKI gene, might be explained by a different sample structure/population background in the analyzed GWAS sets.

Apart from the initial investigation by Vieira et al. (7), only one independent study, analyzing variants within SKI and NSCL/P, has been published to date (12). In that study, the authors first sequenced the SKI gene in 100 controls. One non-synonymous variant (rs28384811) was identified and this was subsequently genotyped in 148 NSCL/P cases and 147 controls. This analysis yielded a minor allele frequency of 10.5% in NSCL/P cases and of 17.7% in controls. Unfortunately, the only information provided concerning the genetic background of these individuals was that they were born in California (12), which renders interpretation of these findings difficult.

Accumulating evidence suggests that rare variants with high penetrance contribute to complex phenotypes such as orofacial clefting. Accordingly, Vieira et al. (7) attempted to identify such variants in subjects with NSCL/P by sequencing the SKI coding region, including exon–intron boundaries and untranslated regions. Within SKI, four non-synonymous variants were observed, one of which, A388V, was claimed by the authors to be of potential etiological relevance. However, this mutation was subsequently identified in nine of 1,064 European controls, and analysis of family members of the A288V index patient revealed that it had
been inherited from the unaffected mother. Although these data do not support the involvement of rare SKI variants in NSCL/P, SKI remains an interesting candidate gene in pedigrees with multiple affected family members, particularly those showing linkage to the SKI chromosomal region on 1p36.3. The fact that linkage of NSCL/P to a region between 1p36.1 and 1p36.3 has indeed already been reported (13) further supports the idea that private mutations within SKI might be responsible for the linkage findings in these families. Also, sequencing of SKI in patients with the 1p36 deletion syndrome might be envisaged, in particular in those 17% of patients with clefting anomalies (6). Rare, recessive alleles of SKI might be unmasked in these patients and could explain some of the facial anomalies (5).

In summary, the present data provide some limited support for the hypothesis that common SKI variants are susceptibility factors for NSCL/P, particularly in Native American populations. However, the phenotypic consequences of predisposing variants in any given individual are likely to be dependent on the respective genetic background. Thus, further studies in different ethnicities are warranted to elucidate the contribution of SKI variants to orofacial clefting.

Acknowledgements – We sincerely thank all of the patients and families at each recruitment site for their participation in this study, and we gratefully acknowledge the invaluable assistance of all clinical, field, and laboratory staff who contributed to this effort. We also thank S. Raeder and A. Reinscheid for their laboratory work. The study was supported by the Deutsche Forschungsgemeinschaft (FOR 423 and individual grants MA 2546/3-1, KR 1912/7-1, NO 246/6-1, WI 1555/5-1). H.R. was supported by a research grant from the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF). No Mexican funds were granted for this work. Data sets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000094.v1.p1. Additional details on dbGaP are provided as Supporting Information (Fund- through dbGaP accession number phs000094.v1.p1. Additional details on dbGaP are provided as Supporting Information (Fund-

References


Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Funding support: Additional details on dbGaP.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.