LETTER TO JMG

Linkage and linkage disequilibrium searched for between non-syndromic cleft palate and four candidate loci

H Koillinen, V Ollikainen, J Rautio, J Hukki, J Kere

Cleft palate (CP) is one of the most common congenital malformations. It can occur as part of a recognisable syndrome, associated with other malformations or, most commonly, be non-syndromic (CPO) (MIM 119540). The birth prevalence of CPO varies between and among populations but it is seen world wide. The highest incidence has been found in Finland, 1.01 per 1000 livebirths.1 Within Finland, there are regional differences in birth prevalence; the Oulu region (in northern central Finland) and central western Finland are over-represented, with up to twice the incidence compared to the average.2 Differences are even more striking when analysing the birth places of the grandparents of probands.3

The gene defects or susceptibility genes for clefts remain largely unknown and the basic mechanism causing the failure of the secondary palate to close is still poorly understood. Extrinsic factors like advanced paternal age, maternal smoking, and overall intake of medicines during the first trimester have been suggested to increase the risk of CPO.4 Although some pedigrees have shown autosomal dominant segregation and X linked recessive inheritance, the risk of recurrence for relatives in large series of cases differs greatly from expected values calculated on the basis of a simple Mendelian mode of inheritance.

Targeted mutations in mice have shown that malfunction of very different types of genes can lead to cleft palate, msxl and tgfb3 among them.5–7 In humans, defects in genes encoding collagens,8 a fibroblast growth factor receptor,9 a sulphate transporter,10 a nucleolar protein,11 and a thyroid transcription factor12 have appeared to be associated with syndromes where cleft palate can be involved. For non-syndromic cleft palate, neither mutations nor linkage to a specific chromosomal region has yet been established. However, linkage disequilibrium was suggested in two studies between MSX1 and non-syndromic CP.13–15 In the Danish population, no association between CPO and MSX1 was found but an association between the risk for CP and variation at the TGFB3 locus was detected.16 A mutation in MSX1 leading to a preterm stop codon was found to cosegregate with cleft lip with or without cleft palate (CL/P) in a large pedigree.17 An MSX1 missense mutation has been shown to cause selective tooth agenesis.18 No mutations were found in coding regions of MSX1 and TGFB3 in patients with CPO in Iowa.19 Two genome scans for cleft lip with or without cleft palate (CL/P) have been published.20,21 A genome wide scan has not been performed in families with CPO.

Nine percent of patients with deletion in 22q11 manifest cleft palate.21 Patients also have other signs like velocardiofacial insufficiency, hypocalcaemia, thymic hypoplasia, cardiac problems, renal anomalies, and abnormal facies (CATCH 22 syndrome).22 The size of the commonly deleted region is 3 Mb,23 but so far the smallest deletion found has been 20 kb.24 The deletion of 20 kb removed exons 1 to 3 of the UFDL1 gene and the patient had typical features of 22q11 deletion.25 In a study of this chromosomal region, no 22q11 hemizygosity was detected in patients with isolated cleft palate.26

Cleft palate has also been described in patients having other chromosomal deletions, but always with other symptoms or signs. Recently, Breuer et al20 reported two patients with cleft palate and balanced translocations in 2q32 between the loci D2S311 and D2S116, within a region of approximately 2.5 Mb. They also analysed data on cleft palate patients with chromosomal deletions from the Human Cytogenetics Database (HCDB). Regions 2q32, 4p13-p16, and 4q31-q35 were significantly associated with cleft palate.21

Key points

• The gene defects or susceptibility genes for cleft palate (CPO) remain largely unknown. We performed linkage and linkage disequilibrium analyses with Finnish multiplex families affected by non-syndromic cleft palate, focusing on four candidate loci: the genes TGFB3 and MSX1 and the chromosomal regions 22q11 and 2q32. We also analysed the whole of chromosomes 2 and 4 by linkage. Finally, we estimated whether the power of our tools, pedigrees and markers, was sufficient to show shared haplotypes within and between subsets of the families.

• None of the four candidate loci showed significant association between the phenotype and marker alleles or haplotypes, even though the power of the set up turned out to be adequate if haplotypes of more than one marker were moderately over-represented among the disease associated haplotypes.

• We conclude that none of the four loci is a major genetic determinant of CPO in Finnish patients.

SUBJECTS AND METHODS

In this study, we have focused on four candidate regions. TGFB3 and MSX1 were chosen because knock-out mice models exhibit cleft palate.13–15 The 22q11 and 2q32 regions were chosen on the basis of reported human phenotypes associated with chromosomal deletions and translocations, respectively. We performed TDT as well as linkage and linkage disequilibrium analyses with Finnish multiplex families affected by non-syndromic cleft palate. We also analysed the whole of chromosomes 2 and 4 by linkage. Finally, we estimated whether the power of our tools, pedigrees and markers, was sufficient to show a shared genotype or haplotype within and between a subset of the families.

The treatment of cleft patients in Finland has been centralised since 1948 to a national Cleft Centre. Data concerning multiplex cleft palate families were collected from the years 1967-1996 from the Cleft Register at the Centre and 250 patients were contacted by sending them a letter. The 24 largest pedigrees originating from different regions in Finland were chosen for the analysis. Nuclear families were examined

www.jmedgenet.com
RESULTS AND DISCUSSION

In 10 out of 24 families we observed null alleles for the marker D22S944. Fourteen affected and 17 unaffected subjects carried null alleles. The results were verified by synthesising a new set of primers over the same polymorphic repeat sequence, yielding identical results. Subjects carrying null alleles were recognised because they seemed to be homozygous for that locus, but Mendelian inheritance errors could be detected, when one of the parents and the offspring looked homozygous for a different allele. In all cases, non-paternity could be excluded by inspection of other markers. No other Mendelian errors occurred in chromosome 22. Interestingly, in a previous study, the marker D22S944 was the most common marker in the 22q11 region to be deleted among velocardiofacial syndrome patients with psychiatric syndromes. To understand the role of the null alleles in cleft palate, we genotyped the same marker in controls (28 Finnish patients with systemic lupus erythematosus and 63 of their healthy relatives) and again found similar null alleles in six patients and five unaffected subjects. Thus, we conclude that the null alleles of marker D22S944 are not specifically associated with cleft palate but occur commonly in the population.

We then continued to analyse genetic linkage using the non-parametric mode of GENEHUNTER, but found only weak evidence for linkage between cleft palate and candidate loci. In the 2q11 region, the highest Z value was 1.36 (p=0.09) and the information content was 0.77 at that point. The Z score was 0.80 (p=0.20, information content 0.56) in the TGFB3 region on chromosome 14. The MSX1 region on chromosome 4 showed negative Z scores (fig. 1), and the highest Z value was 1.64 (p=0.06) in chromosome 4 (fig 1). In chromosome 2, the Z score was at its maximum (1.34, p=0.09, information content 0.61) for locus D2S423 (fig 2). Markers in the candidate region 2q32 showed no strong evidence of linkage; Z max was 0.54 (p=0.29) and the information content was 0.78. None of these scores exceed the threshold for expected levels by random fluctuation.

We then used TDT which did not show any significant difference from the expected values. For each candidate region, the allele whose transmission count deviated most from the expected value is presented in table 1. The corrected p values are based on a permutation test of 1000 iterations. TDT analysis of haplotypes of two or three adjacent markers also failed to produce significant results (table 1).

Because weak genetic effects may sometimes be detected as associations with nearby marker alleles, especially in isolated founder populations, we searched next for association between the affected status and both marker alleles and haplotypes of all lengths. The possible bias caused by unequal numbers of affected subjects in the pedigrees was reduced by randomly selecting one affected person from each sibship. The most likely haplotypes of these subjects, as obtained from the Genehunter package, were considered as disease associated haplotypes, while the control haplotypes were constructed...
from the corresponding untransmitted parental alleles. Because this construction of 86 disease associated and a similar number of control haplotypes contained some randomness, we performed it 20 times, and show the medians of the highest $\chi^2$ values in table 2, all of which turned out to be related to individual alleles. The differences between the 20 repetitions were, however, small.

To obtain corrected $p$ values, we generated 100 data sets, where association between haplotype status and marker alleles was purely coincidental. The subjects in these data sets were selected in a similar way as the real data, but, in addition, the status column of the haplotypes was permutated before computing the highest $\chi^2$ value. The observed $\chi^2$ values were then compared to the resulting empirical distribution. After this correction for multiple testing, all of the observed weak associations can be considered non-significant.

Since we failed to show allelic association, we analysed the power of our set up by means of a computer simulation. For this purpose, 100 data sets of 172 chromosomes were sampled from the pedigrees in the way described before. Here, in presence of a clearly negative result, we could assume that the haplotype frequencies in the chromosomes labelled as disease associated and control do not differ. From each data set, random haplotypes $H$ of length 1, 2, and 3 were picked. Each of these haplotypes was then enriched one at a time in the disease associated chromosomes by replacing the corresponding alleles in each chromosome with haplotype $H$ with probability $P$ of 10%, 20%, and 30%. With probability $1-P$, each disease associated haplotype remained unchanged. Thus, probability $P$ represents the extent at which an artificially introduced disease associated haplotype is over-represented in the affected sample, and is analogous to $P_{\text{excess}} = (P_{\text{affected}} - P_{\text{normal}})/(1 - P_{\text{normal}})$, where $P_{\text{affected}}$ and $P_{\text{normal}}$ denote allele frequency in patient and control chromosomes, respectively. For each enriched data set, the highest $\chi^2$ value was computed, and, finally, it was counted how often these highest values exceeded the corresponding critical thresholds for $p=0.05$ obtained from a permutation test (based on 100 iterations). This ratio corresponds to the power to detect linkage disequilibrium at a type 1 error rate of 0.05. The empirical power levels are summarised in table 3.

The possibility of detecting allelic association between families depends on the relationship between the families, the density of the marker map, and the genetic heterogeneity of the disease. We performed an experiment on using simple tests of association as a means of localisation, provided that significant association has been shown by other means, such as permutation tests. We considered a set up where one affected subject is sampled from each family and repeatedly generated pedigrees where the distance of each subject taken into the analysis was separated by exactly 12 meiotic steps from the common ancestor. As a result, the kinship between two arbitrary affected subjects was 24 meiotic steps, which, knowing the population history of the Finnish isolates in focus, is a rather conservative assumption. To analyse the effect of phenocopy rate, we repeated the experiments using two different numbers of affected mutation carriers, namely six and 12, the remaining 18 and 12 affected subjects being phenocopies. In each simulation, the region around the disease locus sharing IBD was computed for each founder by repeatedly simulating crossovers using Haldane’s model. For each simulated pedigree, a marker map of 50 cM was made by setting intermarker distances to 1 cM, and the polymorphism information content (PIC) of each marker to 0.6. Next, a single affected founder haplotype was created, and the alleles within the IBD region of affected subjects were copied from

### Table 1 Transmission disequilibrium test for the four regions

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Marker</th>
<th>Allele</th>
<th>No trans</th>
<th>No not</th>
<th>$\chi^2$</th>
<th>$p_{\text{corr}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2q32</td>
<td>D2S442</td>
<td>196</td>
<td>8</td>
<td>1</td>
<td>5.44</td>
<td>0.36</td>
</tr>
<tr>
<td>MSX1</td>
<td>D4S1647</td>
<td>144</td>
<td>21</td>
<td>7</td>
<td>7.00</td>
<td>0.46</td>
</tr>
<tr>
<td>TGFβ3</td>
<td>D14S31</td>
<td>112</td>
<td>2</td>
<td>9</td>
<td>4.45</td>
<td>0.66</td>
</tr>
<tr>
<td>22q11</td>
<td>D22S264</td>
<td>102</td>
<td>0</td>
<td>4</td>
<td>4.00</td>
<td>0.68</td>
</tr>
</tbody>
</table>

### Table 2 Association analysis results. The allele size is given in bp

<table>
<thead>
<tr>
<th>Region</th>
<th>Marker</th>
<th>Allele</th>
<th>No of affected</th>
<th>No of controls</th>
<th>$\chi^2$</th>
<th>$p_{\text{corr}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2q32</td>
<td>D2S1356</td>
<td>236</td>
<td>37</td>
<td>16</td>
<td>6.34</td>
<td>0.50</td>
</tr>
<tr>
<td>MSX1</td>
<td>D4S394</td>
<td>237</td>
<td>17</td>
<td>5</td>
<td>6.43</td>
<td>0.68</td>
</tr>
<tr>
<td>TGFβ3</td>
<td>D14S273</td>
<td>104</td>
<td>4</td>
<td>0</td>
<td>3.71</td>
<td>0.66</td>
</tr>
<tr>
<td>22q11</td>
<td>D22S311</td>
<td>112</td>
<td>13</td>
<td>3</td>
<td>5.62</td>
<td>0.65</td>
</tr>
</tbody>
</table>

### Table 3 Power to detect allele or haplotype association as determined empirically by simulation.

<table>
<thead>
<tr>
<th>Length</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>2q32</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>0.36</td>
</tr>
<tr>
<td>MSX1</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td>TGFβ3</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>0.46</td>
</tr>
<tr>
<td>22q11</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.18</td>
</tr>
</tbody>
</table>
of carriers 535/1000, indicating that with the higher carrier frequency,
the true associated allele/haplotype and all other alleles/haplotypes that cover no
more than 10 cM altogether in the genetic map.

this founder haplotype. Alleles in the remaining marker loci as well as all alleles of the controls were drawn randomly using the allele frequencies specified in the map.

The capability to localise the disease locus within the region was first evaluated by calculating the proportion of replicates where the allele or haplotype that showed the most significant association was actually part of the affected founder haplotype. The level of association was measured by \( \chi^2 \) test statistic computed from a 2×2 contingency table. For six mutation carriers, this proportion was 104/1000 and for 12 haplotypes that produce higher carrier frequency the most significant association shows the true locus in most cases.

Next, we computed the total size of the regions identified by the true associated allele/haplotype and all other alleles/haplotypes that produce higher \( \chi^2 \) test statistic values. The respective cumulative distribution functions are shown in fig. 3, indicating that, for example, if there are 12 carriers of the common mutation, in 90% of cases the true disease locus could be found by focusing on the most significantly associated alleles/haplotypes that cover no more than 10 cM altogether in the genetic map. If the number of carriers decreases to six, the corresponding localisation power is, however, poor (33%).

Mapping a gene responsible for a genetically complex disease is a demanding task. During embryogenesis, a cascade of genes is probably needed for complete palatal closure and, therefore, a mutation in any of these genes could lead to cleft palate. Traditional linkage analysis is difficult to apply in the absence of unambiguous parameters. Linkage disequilibrium analysis might be more powerful, especially in isolated populations like Finland. But, as our simulations above showed, strict criteria in phenotyping and a dense marker map are necessary. None of the four candidate regions tested showed significant association between the phenotype and marker alleles or haplotypes, even though the power of the set up turned out to be adequate if haplotypes of more than one marker were moderately over-represented (Pexcess at least 0.2) in the disease associated haplotypes. Unfortunately, the chance of finding an existing disease haplotype in a complex disease is far too small, if samples are collected from different population subgroups with different histories (and possibly different disease alleles) and if a marker distance of 10 cM is used. Even if the map is much denser (1 cM), a low phenoype rate (50% or less) is required for reliable localisation using samples of 24 affected and 24 control subjects. It is worth remembering that differences between phenotypes owing to distinctive underlying gene defects can be clinically indistinguishable. In this respect, the shared segment method might be more suitable if there is a possibility of constructing large pedigrees and of finding distant, affected relatives.

ACKNOWLEDGEMENTS

We thank the families for their participation in the study. We thank Dr Elisabeth Widen and Ms Riitta Lehtimäki for discussions and laboratory assistance. This work was supported by the Academy of Finland and Sigrid Juselius Foundation.

Authors’ affiliations

H Kaatainen, J Kere, Department of Medical Genetics, University of Helsinki, Helsinki, Finland

V Ollikainen, J Kere, Finnish Genome Centre, University of Helsinki, Helsinki, Finland

V Ollikainen, Department of Computer Science, University of Helsinki, Helsinki, Finland

J Routio, J Hukki, Cleft Centre, University Hospital of Helsinki, Helsinki, Finland

H Kaatainen, Department of Child Neurology, University of Turku, Turku, Finland

J Kere, Department of Biosciences at Novum, Karolinska Institute, Stockholm, Sweden

Correspondence to: Professor J Kere, Department of Biosciences at Novum, Karolinska Institute, 14157 Huddinge, Sweden; jku.kere@biocsi.ki.se

REFERENCES


www.jmedgenet.com


Linkage and linkage disequilibrium searched for between non-syndromic cleft palate and four candidate loci
H Koillinen, V Ollikainen, J Rautio, J Hukki and J Kere

doi: 10.1136/jmg.40.6.464

Updated information and services can be found at:
http://jmg.bmj.com/content/40/6/464

These include:

**References**
This article cites 25 articles, 3 of which you can access for free at:
http://jmg.bmj.com/content/40/6/464#BIBL

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**
Articles on similar topics can be found in the following collections
- Genetic screening / counselling (887)
- Calcium and bone (307)
- Drugs: endocrine system (107)
- Immunology (including allergy) (604)
- Metabolic disorders (329)
- Molecular genetics (1254)

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/