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 worldwide. The highest incidence has been found in Finland, 1.01 per 1000 livebirths. 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. Differences are even more striking when analysing the birth places of the grandparents of probands.

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. Although some pedigrees have shown autosomal recessive inheritance 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, msx1 and tgfβ3 among them. In humans, defects in genes encoding collagens, a fibroblast growth factor receptor, a sulphate transporter, a nuclear protein, and a thyroid transcription factor 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. In the Danish population, no association between CPO and MSX1 was found but an association between the risk for CP and variation at the TGFβ3 locus was detected. 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. An MSX1 missense mutation has been shown to cause selective tooth agenesis. No mutations were found in coding regions of MSX1 and TGFβ3 in patients with CPO in Iowa. Two genome scans for cleft lip with or without cleft palate (CL/P) have been published. A genome wide scan has not been performed in families with CPO. Nine percent of patients with deletion in 22q11 manifest cleft palate. Patients also have other signs like velopharyngeal insufficiency, hypocalcaemia, thymic hypoplasia, cardiac problems, renal anomalies, and abnormal facies (CATCH 22 syndrome). The size of the commonly deleted region is 3 Mb, but so far the smallest deletion found has been 20 kb. The deletion of 20 kb removed exons 1 to 3 of the UFDB1 gene and the patient had typical features of 22q11 deletion. In a study of this chromosomal region, no 22q11 hemizygosity was detected in patients with isolated cleft palate.

Cleft palate has also been described in patients having other chromosomal deletions, but always with other symptoms or signs. Recently, Brewer et al 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.

Subjects and Methods

In this study, we have focused on four candidate regions. TGFβ3 and MSX1 were chosen because knock-out mice models exhibit cleft palate. 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 shared haplotypes within and between subsets of the families.

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

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 TGFβ3 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.
- We conclude that none of the four loci is a major genetic determinant of CPO in Finnish patients.
by HK to recognise previously misdiagnosed syndromes. Cleft palates of all subtypes were included. Ancestors of the patients were traced three to six generations back. The parents of the proband were related to each other in only one pedigree. A common ancestor for probands previously not known to be related to each other could be found in two pedigrees. Altogether, the families consisted of 63 affected and 112 unaffected subjects (a total of 175 subjects).

DNA was extracted from blood samples by a non-enzymatic method. Fifty ng of genomic DNA was amplified either by the Weissenbach or Morrow protocol (http://www.gdb.org). The PCR products were fractionated on 6% polyacrylamide gels. The gels were silver stained and alleles were numbered. The PCR products of markers D22S425, D22S306, and D22S308 were labelled with fluorescent dyes and the gels were run in an ABI 377 sequencing machine (PE Biosystems, CA, USA).

The 2q11 region was studied by nine microsatellite polymorphisms: D22S1638, D22S1623, D22S941, D22S944, D22S264, D22S3511, D22S306, D22S308, and D22S425. These markers are inside the 3 Mb region which is commonly deleted in patients with the velocardiofacial syndrome.24,25

TGFB3 was studied by using polymorphic markers D14S273 and D14S561. TGFB3 is located between these markers in the YAC 746B4 within 1800 kb.25 The MSXI region was studied using a polymorphic marker D4S394, which is located approximately 7 cM proximal to MSXI, and with an intragenic dinucleotide repeat polymorphism. In addition, the entire chromosome 4 was analysed by using 20 polymorphic markers from the modified Weber set 6 (http://www.pebio.com/ga/). The mean distance between the markers was 11.3 cM. Chromosome 2 was also analysed by using markers from the ABI marker set. The mean distance between the 21 markers was 13.8 cM. The critical 2q32 region was further analysed by using the polymorphic markers D25S111, D25S348, D25S392, and D25S115.26,27

Unambiguous parameters for a model of inheritance cannot be set. Therefore, we used non-parametric linkage analysis and examined allele sharing among the affected by GENEHUNTER, which calculates the deviation from the random distribution over the possible inheritance vectors (NPL score, Z).27 One pedigree was too large to be computerized and was used in linkage disequilibrium analysis only. The transmission disequilibrium test (TDT) was also done with GENEHUNTER.

We tested the data sets for allele and haplotype association using standard χ² tests for 2 × 2 contingency tables and estimated our localisation power by means of a computer simulation. Since population level association could be assumed, we also used transmission/disequilibrium tests to test for linkage with genetic markers.28 Finally, we used computer simulations to assess our power to localise the disease gene in a similar set up, if the localisation is based solely on linkage disequilibrium between subjects of different families.

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 symptoms.28 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.

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...

Figure 1 Linkage results for chromosome 4 showing the NPL (Z) scores for all the 24 families. The markers used from left to right are D4S2366-HOX7-D4S394-D4S2639-D4S2397-D4S408-D4S1627-GATA28F03-D4S2367-D4S1647-D4S2623-D4S2394-D4S1644-D4S1625-D4S1629-D4S2368-D4S2431-D4S2417-D4S408-D4S1652.

Figure 2 Linkage results for chromosome 2 and the NPL (Z) scores for all the 24 families. The markers used from left to right are D2S1780-D2S423-D2S1400-D2S405-D2S1788-D2S1356-D2S441-D2S1394-D2S1777-D2S1790-D2S410-D2S442-D2S1326-D2S1776-D2S1391-D2S311-D2S15-D2S1348-D2S2392-D2S1384-D2S1649-D2S434-D2S4279-D2S338-D2S125.
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 two 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 permuted 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 associated and control do not differ. From each data set, ran-

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Transmission disequilibrium test for the four regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome</td>
<td>Marker</td>
</tr>
<tr>
<td>2q32</td>
<td>D2S442</td>
</tr>
<tr>
<td>MSX1</td>
<td>D4S1647</td>
</tr>
<tr>
<td>TGFβ3</td>
<td>D14S61</td>
</tr>
<tr>
<td>22q11</td>
<td>D22S264</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Association analysis results. The allele size is given in bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>Marker</td>
</tr>
<tr>
<td>2q32</td>
<td>D2S1356</td>
</tr>
<tr>
<td>MSX1</td>
<td>D4S394</td>
</tr>
<tr>
<td>TGFβ3</td>
<td>D14S273</td>
</tr>
<tr>
<td>22q11</td>
<td>D22S311</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Power to detect allele or haplotype association as determined empirically by simulation. The length (number of markers) in a haplotype is given in the left column and the power is given as a function of sharing (% of chromosomes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>Power</td>
</tr>
<tr>
<td>2q32</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.35</td>
</tr>
<tr>
<td>3</td>
<td>0.36</td>
</tr>
<tr>
<td>MSX1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
</tr>
<tr>
<td>3</td>
<td>0.23</td>
</tr>
<tr>
<td>TGFβ3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.46</td>
</tr>
<tr>
<td>22q11</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>0.18</td>
</tr>
</tbody>
</table>

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...
Figure 3. Cumulative distribution functions of 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.

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

Authors’ affiliations
H Kaatla, 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 Hakki, Cleft Centre, University Hospital of Helsinki, Helsinki, Finland
H Kaatla, 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; juha.kere@biosci.ki.se

REFERENCES


www.jmedgenet.com


