Medical Sequencing of Candidate Genes for Nonsyndromic Cleft Lip and Palate


1 Department of Pediatrics, University of Iowa, Iowa City, Iowa, United States of America; 2 Interdisciplinary Genetics PhD Program, University of Iowa, Iowa City, Iowa, United States of America; 3 Latin American Collaborative Study of Congenital Malformations, Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil; 4 Latin American Collaborative Study of Congenital Malformations, Department of Genetics, Fiocruz, Rio de Janeiro, Brazil; 5 Center for Medical Education and Investigation, Buenos Aires, Argentina; 6 Cell and Molecular Genetics, Medical Research Council Human Genetics Unit, Edinburgh, United Kingdom; 7 Center for Oral Biology and Department of Biomedical Genetics, University of Rochester, Rochester, New York, United States of America; 8 Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America; 9 Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

Nonsyndromic or isolated cleft lip with or without cleft palate (CL/P) occurs in wide geographic distribution with an average birth prevalence of 1/700. We used direct sequencing as an approach to study candidate genes for CL/P. We report here the results of sequencing on 20 candidate genes for clefts in 184 cases with CL/P selected with an emphasis on severity and positive family history. Genes were selected based on expression patterns, animal models, and/or role in known human clefting syndromes. For seven genes with identified coding mutations that are potentially etiologic, we performed linkage disequilibrium studies as well as in 501 family triads (affected child/mother/father). The recently reported MSX1 P147Q mutation was also studied in an additional 1,098 cleft cases. Selected missense mutations were screened in 1,064 controls from unrelated individuals on the Centre d’Etude du Polymorphisme Humain (CEPH) diversity cell line panel. Our aggregate data suggest that point mutations in these candidate genes are likely to contribute to 6% of isolated clefts, particularly those with more severe phenotypes (bilateral cleft of the lip with cleft palate). Additional cases, possibly due to microdeletions or isodisomy, were also detected and may contribute to clefts as well. Sequence analysis alone suggests that point mutations in FOXE1, GLI2, JAG2, LHX8, MSX1, MSX2, SATB2, SKI, SPRY2, and TBX10 may be rare causes of isolated cleft lip with or without cleft palate, and the linkage disequilibrium data support a larger, as yet unspecified, role for variants in or near MSX2, JAG2, and SKI. This study also illustrates the need to test large numbers of controls to distinguish rare polymorphic variants and prioritize functional studies for rare point mutations.

Introduction

Nonsyndromic or isolated cleft lip with or without palate (CL/P) occurs in wide geographic distribution with an average birth prevalence of 1/700 [1]. CL/P is a complex trait determined by multiple, interacting loci, with additional environmental covariates. Recent work suggests that three to 14 interacting loci provide a good model for genetic effects in CL/P [2].

Studying candidate genes for CL/P selected from animal models and expression patterns is a common strategy [3]. To identify gene(s) involved in CL/P, investigators have used both association and linkage approaches to evaluate genetic contributions. To detect the very small gene effects on CL/P by linkage or linkage disequilibrium strategies, sample sizes need to be large and there needs to be either a common variant in association (for linkage disequilibrium) or a substantial degree of single locus contributions (for linkage). We used direct sequencing as an alternative approach to study candidate genes for CL/P hoping to identify genes with modest effects on the disease. The results of the direct sequencing of MSX1 [4,5] suggest that point mutations in this gene underlie approximately 2% of CL/P cases. We report here the results of sequencing 20 additional candidate genes for clefts. For seven genes with identified coding mutations that are potentially etiologic, we performed linkage disequilibrium studies as well. For the MSX1 P147Q mutation reported by Suzuki et al. [5], we investigated an additional 1,098 cleft cases.

Results

One hundred and forty-nine exons (representing 77,527 nucleotides of DNA sequencing), including exon–intron boundaries and untranslated regions, of 20 genes were screened for mutations in the Iowa and Philippine cleft populations. Table 2 summarizes the number of variants and
putative mutations observed. Of the 256 variants seen, 16
missense mutations in nine genes seemed to be of potential
etiologic importance. All 16 missense mutations were
observed in a single cleft lip and palate case, with the
exception of the SPRY2 D20A and TBX10 R354Q mutations
that were seen in two and three cases respectively. None were
seen in the 186 matched controls (Table 3). These mutation
sites are not highly conserved across species with the
exception of the SPRY2 and GLI2 mutations. Both SPRY2
mutation sites as well as three GLI2 mutation sites are
conserved from Xenopus to human (Figure 1; complete data
available at http://genes.uiowa.edu/publication.html). The
JAG2 and the TBX10 R354Q mutation sites are not conserved
in other species orthologs available for study. The sequence
surrounding the JAG2 A657H mutation site is likely a
calcium-binding EGF-like domain, which is present in a large
number of membrane-bound and extracellular proteins. Also,
the SPRY2 K68N mutation site is in the sprouty domain and
inhibits the Ras/mitogen-activated protein kinase (MAPK)
cascade, a pathway crucial for developmental processes
initiated by activation of various receptor tyrosine kinases.

All mutations were predicted to be benign by PolyPhen
(http://www.bork.embl-heidelberg.de/PolyPhen/) with the
exception of the JAG2 M597I and SPRY2 D20A that were
“possibly damaging” and GLI2 S1213Y that was “probably
damaging” (Table 3). However, with the exception of the
LHX8 E221A, GLI2 R426Q, and GLI2 S1213Y mutations, all
missense mutations appear to potentially disrupt splicing by
either creating or inactivating exonic splicing enhancer
sequences (complete information is available as supplemental
material at http://genes.uiowa.edu/publication.html). None of the
mutations identified in this study appear to disrupt possible
exonic splicing silencer sequences.

The SATB2 T190A mutation was not found in the panel of
1064 CEPH controls as well. We also tested the LHX8 E221A,
SKA388Y, SPRY2 D20A, and TBX10 R354Q mutations in the
panel of 1064 CEPH controls after not seeing it in 200
population matched controls. We found the LHX8 E221A
mutation in 17 samples, the SKA388Y mutation in nine
samples, the SPRY2 D20A mutation in 60 samples, and the
TBX10 R354Q mutation in six samples. (A complete list is

### Table 1. Candidate Genes Studied

<table>
<thead>
<tr>
<th>Gene</th>
<th>Map Location</th>
<th>KO*</th>
<th>L/LDb</th>
<th>CHc</th>
<th>Other Reason to Be Selected as Candidate</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOXE1</td>
<td>9q22</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>[9,28,47]</td>
</tr>
<tr>
<td>GLI2</td>
<td>2q14</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Mutations in GLI2 cause holoprosencephaly-like features with cleft lip and palate</td>
<td>[11]</td>
</tr>
<tr>
<td>ISGF3G</td>
<td>14q11.2</td>
<td>NA</td>
<td>—</td>
<td>—</td>
<td>Similar to IRF6</td>
<td>—</td>
</tr>
<tr>
<td>JAG2</td>
<td>14q32</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>[46]</td>
</tr>
<tr>
<td>Lhx8</td>
<td>1p31</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>[49]</td>
</tr>
<tr>
<td>MSX2</td>
<td>5q34-q35</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Similar to MSX1</td>
<td>—</td>
</tr>
<tr>
<td>NAT1</td>
<td>8p23.1–21.3</td>
<td>NA</td>
<td>+</td>
<td>—</td>
<td>Correlates with glucocorticoid-induced cleft palate in mice and is involved in smoking detoxification</td>
<td>[50,51]</td>
</tr>
<tr>
<td>OSR2</td>
<td>8q22.2</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>[28,52]</td>
</tr>
<tr>
<td>PAX9</td>
<td>19q12-q13</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>[54]</td>
</tr>
<tr>
<td>SATB2</td>
<td>2q32-q33</td>
<td>NA</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>[28,46]</td>
</tr>
<tr>
<td>SOX</td>
<td>1p36.3</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>[44,45]</td>
</tr>
<tr>
<td>SPRY1</td>
<td>4q26.1</td>
<td>NA</td>
<td>—</td>
<td>—</td>
<td>Sprouty expression is induced by the fibroblast growth factor receptor (FGFRI) pathway. Loss-of-function mutations in FGFRI cause a syndrome of clefting (Kallmann syndrome)</td>
<td>[35,55]</td>
</tr>
<tr>
<td>SPRY2</td>
<td>13q22.2</td>
<td>NA</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>[35,55]</td>
</tr>
<tr>
<td>TBX1</td>
<td>22q11.21</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>[56,57]</td>
</tr>
<tr>
<td>TBX10</td>
<td>11q13.1</td>
<td>NA</td>
<td>—</td>
<td>—</td>
<td>Ectopically expressed in Dancer cleft lip and palate mutant mice</td>
<td>[58]</td>
</tr>
<tr>
<td>TGFα</td>
<td>2p13</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>[28]</td>
</tr>
<tr>
<td>TGFβ1</td>
<td>19q13.2</td>
<td>—</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>[28]</td>
</tr>
<tr>
<td>TGFβ3</td>
<td>14q24.3</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>[43,59,60]</td>
</tr>
<tr>
<td>TGFβR</td>
<td>9q22</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>[28]</td>
</tr>
</tbody>
</table>

* A plus sign (+) indicates a knockout mouse with clefts; a negative sign (−) indicates a knockout mouse without clefts.
* A plus sign (+) indicates evidence of linkage/linkage disequilibrium; a negative sign (−) indicates no evidence of linkage/linkage disequilibrium.
* A plus sign (+) indicates evidence of chromosome rearrangements; a negative sign (−) indicates no evidence of chromosome rearrangements.

DOI: 10.1371/journal.pgen.0010064.s001
available at our Web site: http://genetics.uiowa.edu/publications.html).

The MSX1 P147Q mutation was not found in any of 1,671 controls but was found in two Filipino cleft families from a panel of 1,468 cleft cases from the Philippines, which indicates a frequency of 0.14%. The first family has no family history for clefting and the variant segregates from the unaffected mother. The second family has four affected with clefts. The variant was found in two cousins and segregates from the unaffected grandmother to her unaffected son and daughter of a sibship of nine. A first cousin of the proband is affected but does not carry the variant. The last affected, a third cousin once removed, also has a cleft but does not carry the variant (Figure 2).

The LHX8 and SATB2 mutations were originally seen in single cases from the Philippines. For SATB2 T190A, the variant was transmitted from the unaffected mother. The LHX8 E221A mutation was also transmitted from the unaffected mother and is present in one affected and one unaffected person but was absent in two unaffected siblings from a sibship of eight brothers and sisters (Figure 3). Both mutation sites are conserved between humans and mouse (see Appendix at http://genetics.uiowa.edu/publications.html). The SKI A398V mutation was seen in a case from Iowa. The mutation was transmitted from the mother and the mutation site is conserved between humans and species of fish and frog. Two cases from Iowa presented with the SPRY2 D20A mutation. Of these, one had parental DNA samples available and the mutation segregated from the unaffected mother. Three cleft individuals from Iowa presented the TBX10 R354Q mutation. In the two cases with parent samples available, one received the mutated allele from the mother and the other from the father (both unaffected).

Two additional interesting observations were made. An Iowa case presented with isolated cleft lip with cleft palate, and no family history of clefts or any features of DiGeorge syndrome was found to be homozygous for an intronic variant in the TBX1 gene (189 nucleotides into intron 8). This variant was not present in 186 matched controls. We tested this case for the presence of two copies of the ubiquitin fusion degradation gene (UPDIL), using an assay for DiGeorge or 22q syndrome [6], and the results were normal. Parental samples are not available to further study this case, nor is there enough DNA available to confirm a possible deletion by Southern blot analysis. The finding of a single rare homozygote variant suggests the possibility of a micro-deletion or segmental isodisomy of this region. We tested four microsatellite markers (D22S420, D22S1685, D22S683, and D22S445) on both proband and mother samples. The mother presents distinct genotypes from the proband for D22S420 and D22S683 markers, however this finding does not exclude a segmental maternal isodisomy because the interval between these two markers, which contains TBX1, is 32 cM (data not shown).

The second observation involved four Filipino cases that presented a missense mutation in the last SKI amino acid (P728L). This mutation was not found in 186 matched controls. One case appears to have a de novo SKI P728L mutation. This case presented with an isolated right cleft lip and cleft palate and positive family history for clefts. Neither parent had this variant and their biological relationship to the case was confirmed after testing twenty polymorphic markers. Of the other three cases, one had a positive family history for clefts. For this case, we tested the parents, the paternal grandparents and three siblings out of a 15 sibship progeny. One of the tested siblings is affected with a right cleft lip and cleft palate associated with microcephaly. The SKI P728L variant segregated from the unaffected grand-

### Table 2. Summary of Variants Found by Direct Sequence

<table>
<thead>
<tr>
<th>Genes</th>
<th>Location</th>
<th>Coding Variants</th>
<th>Non-Coding Variants</th>
<th>Frequency</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Syn Non-syn PE</td>
<td>Intron 5' UTR 3' UTR</td>
<td>&gt;5% &lt;5%</td>
<td></td>
</tr>
<tr>
<td>FOXE1</td>
<td>10 2</td>
<td>A207V, D285V</td>
<td>— — —</td>
<td>3 9 12</td>
<td></td>
</tr>
<tr>
<td>GLI2</td>
<td>3 10</td>
<td>R46H, R246G, G778D, S1213Y</td>
<td>— — —</td>
<td>5 8 13</td>
<td></td>
</tr>
<tr>
<td>ISGF3G</td>
<td>4 —</td>
<td>— — —</td>
<td>6 — 3</td>
<td>3 10 13</td>
<td></td>
</tr>
<tr>
<td>JAG2</td>
<td>7 10</td>
<td>M597L, A657H</td>
<td>11 1 10</td>
<td>7 32 39</td>
<td></td>
</tr>
<tr>
<td>LHXB1</td>
<td>1 1</td>
<td>E221A</td>
<td>2 — 1</td>
<td>2 3 5</td>
<td></td>
</tr>
<tr>
<td>MSX2</td>
<td>1 2</td>
<td>S63C</td>
<td>1 — —</td>
<td>2 2 4</td>
<td></td>
</tr>
<tr>
<td>NAT1</td>
<td>3 7</td>
<td>— — —</td>
<td>— — —</td>
<td>7 3 10</td>
<td></td>
</tr>
<tr>
<td>OSR2</td>
<td>— —</td>
<td>— — —</td>
<td>— — —</td>
<td>— — 0</td>
<td></td>
</tr>
<tr>
<td>P63</td>
<td>— —</td>
<td>— — —</td>
<td>17 2 —</td>
<td>9 10 19</td>
<td></td>
</tr>
<tr>
<td>PAI9</td>
<td>1 —</td>
<td>— — —</td>
<td>2 — —</td>
<td>2 2 2</td>
<td></td>
</tr>
<tr>
<td>SATB2</td>
<td>1</td>
<td>T190A</td>
<td>3 — —</td>
<td>1 2 4</td>
<td></td>
</tr>
<tr>
<td>SII</td>
<td>5 4</td>
<td>A388V</td>
<td>6 — 8</td>
<td>6 17 23</td>
<td></td>
</tr>
<tr>
<td>SPRY1</td>
<td>— —</td>
<td>— — —</td>
<td>— — 3</td>
<td>3 — 3</td>
<td></td>
</tr>
<tr>
<td>SPRY2</td>
<td>— 3</td>
<td>D20A, K66N</td>
<td>— 4 —</td>
<td>2 5 7</td>
<td></td>
</tr>
<tr>
<td>TBX1</td>
<td>5 1</td>
<td>— — —</td>
<td>7 8 11</td>
<td>14 18 32</td>
<td></td>
</tr>
<tr>
<td>TBX10</td>
<td>4 4</td>
<td>R354Q</td>
<td>15 — 3</td>
<td>13 9 22</td>
<td></td>
</tr>
<tr>
<td>TGF A</td>
<td>2 —</td>
<td>— — —</td>
<td>13 2 16</td>
<td>10 23 33</td>
<td></td>
</tr>
<tr>
<td>TGF B1</td>
<td>2 —</td>
<td>— — —</td>
<td>2 2 —</td>
<td>1 5 6</td>
<td></td>
</tr>
<tr>
<td>TGF B3</td>
<td>1 —</td>
<td>— — —</td>
<td>1 4 —</td>
<td>3 3 6</td>
<td></td>
</tr>
<tr>
<td>TGF B4</td>
<td>1 1</td>
<td>— — —</td>
<td>1 — 2</td>
<td>1 2 3</td>
<td></td>
</tr>
</tbody>
</table>

Non-syn, non-synonymous; PE, potentially etiologic; Syn, synonymous.

DOI: 10.1371/journal.pgen.0010064.t002
The mutations in this table were not found in 186 matched controls. None of the cases presented an affected parent.

* 1 indicates polydactyly; 2 indicates epicanthal folds; 3 indicates cafe au lait spots; 4 indicates hypertelorism.

*Mutations not found in 186 controls but disclosed when an extended control set of 1,204 samples were studied.

*Mutations not found in the extended control set of 1,204 samples.

**BL CL/P, bilateral cleft lip with cleft palate; FH, family history for clefts; L CL/P, unilateral left cleft lip with cleft palate; NA, information not available, sample not available, or sequencing failed; R CL/P, unilateral right cleft lip with cleft palate.**

DOI: 10.1371/journal.pgen.0010064.t003

Table 3. Potential Mutations Found in the Present Study

<table>
<thead>
<tr>
<th>Population (Number of Affected Individuals)</th>
<th>Gene</th>
<th>Amino Acid Change</th>
<th>PolyPhen Prediction</th>
<th>Phenotype</th>
<th>FH</th>
<th>Parents’ Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philippines (1/91)</td>
<td>FOXE1</td>
<td>A207V</td>
<td>Benign</td>
<td>BL CL/P</td>
<td>—</td>
<td>Yes 207AV 207AA</td>
</tr>
<tr>
<td>Iowa (1/93)</td>
<td>FOXE1</td>
<td>D285V</td>
<td>Benign</td>
<td>BL CL/P</td>
<td>—</td>
<td>Yes 285DD 285DV</td>
</tr>
<tr>
<td>Iowa (1/93)</td>
<td>GL2</td>
<td>R46H</td>
<td>Benign</td>
<td>R CL/P</td>
<td>—</td>
<td>No 46RR 46RH</td>
</tr>
<tr>
<td>Philippines (1/91)</td>
<td>GL2</td>
<td>R426Q</td>
<td>Benign</td>
<td>BL CL/P</td>
<td>1</td>
<td>Yes NA NA</td>
</tr>
<tr>
<td>Iowa (1/93)</td>
<td>GL2</td>
<td>G778D</td>
<td>Benign</td>
<td>R CL/P</td>
<td>—</td>
<td>No NA NA</td>
</tr>
<tr>
<td>Philippines (1/91)</td>
<td>GL2</td>
<td>S1213Y</td>
<td>Probably damaging</td>
<td>BL CL/P</td>
<td>—</td>
<td>Yes NA NA</td>
</tr>
<tr>
<td>Philippines (1/91)</td>
<td>JAG2</td>
<td>M597I</td>
<td>Benign</td>
<td>BL CL/P</td>
<td>—</td>
<td>No NA NA</td>
</tr>
<tr>
<td>Philippines (1/91)</td>
<td>JAG2</td>
<td>A657H</td>
<td>Benign</td>
<td>BL CL/P</td>
<td>—</td>
<td>Yes 657AA NA</td>
</tr>
<tr>
<td>Philippines (1/91)</td>
<td>MSX2</td>
<td>S63C</td>
<td>Benign</td>
<td>L CL/P</td>
<td>2</td>
<td>No NA NA</td>
</tr>
<tr>
<td>Philippines (1/91)</td>
<td>SATB2</td>
<td>T190A</td>
<td>Benign</td>
<td>BL CL/P</td>
<td>—</td>
<td>No 190TT 190TA</td>
</tr>
<tr>
<td>Iowa (1/93)</td>
<td>SKI</td>
<td>A388V</td>
<td>Benign</td>
<td>L CL/P</td>
<td>—</td>
<td>No 388AA 388AV</td>
</tr>
<tr>
<td>Iowa (2/93)</td>
<td>SPRY2</td>
<td>D20A</td>
<td>Possibly damaging</td>
<td>R CL/P</td>
<td>—</td>
<td>Yes NA NA</td>
</tr>
<tr>
<td>Iowa (1/93)</td>
<td>SPRY2</td>
<td>K68N</td>
<td>Benign</td>
<td>L CL/P</td>
<td>—</td>
<td>No NA NA</td>
</tr>
<tr>
<td>Iowa (3/93)</td>
<td>TBX10</td>
<td>R354Q2</td>
<td>Benign</td>
<td>NA</td>
<td>NA</td>
<td>NA NA NA</td>
</tr>
<tr>
<td>Philippines (2/1098)</td>
<td>MXS1</td>
<td>P147Q</td>
<td>Benign</td>
<td>R CL/P</td>
<td>—</td>
<td>Yes 147PQ 147PP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BL CL/P</td>
<td>4</td>
<td>No 147PP 147PQ</td>
</tr>
</tbody>
</table>

Discussion

Point mutations in the candidate genes FOXE1, GLI2, MSX2, SKI, SATB2, and SPRY2 appear in aggregate to contribute to as much as 6% of isolated cleft lip and palate cases, enriched for cases with bilateral cleft of the lip with cleft palate and a positive family history. The mutations found in this study are conserved in other mammals, may disrupt exonic splicing enhancer sequences, and were not found in between 400 to 2,000 control chromosomes. The JAG2 M597I and A657H mutations, although they appear to damage the JAG2 protein, according to the PolyPhen prediction, are not conserved in other species and may be rare polymorphic sites.

Testing a larger number of control samples proved to be a useful way to differentiate rare polymorphisms from etiologic mutations. The LHX8 E221A, SATB2 T190A, SKI A388V, SPRY2 D20A, and TBX10 R354Q variants initially were not observed in approximately 200 matched controls. This number of controls is commonly used to assume that if a variant is not present, it is likely causal despite models that have shown that a larger number of controls is useful in eliminating rare variants [8]. However, when we tested these variants in the extended set of 1,064 controls, we found the LHX8 E221A variant in 17 individuals, the SKI A388V mutation in nine, the SPRY2 D20A mutation in 60 individuals, and the TBX10 R354Q in six individuals. Although the presence of the amino acid changes in unaffected controls does not exclude them from playing a role in CL/P, it does place them in a lower priority group for additional functional
analysis and makes them difficult to use in any applied genetic counseling setting as they may be, at best, modifiers with low penetrance contributory alleles.

Some variants, such as the SKI A388V mutation, which is conserved back to X. laevis and T. tilapia but is found in controls, demonstrate that species conservation alone maybe not enough to argue for an etiologic role of a given variant.

Our study illustrates the difficulties of defining as causal a mutation rarely seen in the population. Many of the missense mutations found in the cases studied were seen only when we extended our screen to a larger control group comprised of samples from ethnically diverse groups from almost all parts of the world. In addition, none of the mutations segregated from an affected parent. Incomplete penetrance is likely the explanation for the mutations that may be causal as has been clearly shown for other genes that contribute to clefting such as MSX1 [9] and FGFR1 [10] mutations. It is likely that we found these mutations only because we tested cases more likely to present a stronger genetic contribution (cases with positive family history and bilateral cleft of the lip with cleft palate). Mutations like the MSX1 P147Q and others that appear to show incomplete penetrance are comparable to autosomal dominant disorders resulting from mutations in SCN5A [11], IRF6 [12], or NKX2.5 [13].

The MSX1 P147Q mutation was seen in two cleft cases from the Philippines, but in none of the over 1,600 controls. It appears that this specific mutation underlies approximately 0.15% cases of apparent isolated CL/P. As shown previously this variant results in variable expression and decreased penetrance that make prospective studies of its phenotypic outcome necessary before accurate genetic counseling risk can be measured [5].

Rigorous demonstration that a mutation disrupts a genuine exonic splicing enhancer requires that the sequence autonomously promote splicing and that enhancement be absent in the mutant. An advantage of the score matrix approach [14] is that it allows direct testing of predicted effects on individual putative enhancer sites, rather than having to characterize exonic splicing enhancers by testing multiple random
mutations and/or deletions along an exon. All 15 mutations on Table S2 appear to inactivate and/or create a predicted exon splicing enhancer of at least one of the four serine/arginine-rich (SR) proteins. However, the presence of a high-score motif in a sequence does not necessarily identify that sequence as an exonic splicing enhancer in its native context.

In TBX1, we found one rare intronic variant in homozygous form in an Iowa cleft case, which could indicate that clefts arise from recessive functional intronic mutations in TBX1, or microdeletions that cannot be visualized by direct sequencing. This case does not have a 22q deletion involving the UFD1L gene. Detecting this rare homozygote in the absence of this variant in any other of the 400 people tested suggests these individuals may be identical by descent at this locus and gene. This variant itself, or others in linkage disequilibrium in TBX1, might be a hypomorphic allele whose joint presence results in enough change in gene expression or function to trigger a phenotype. Therefore, other alleles in regulatory regions of TBX1 should be a priority for identification.

Previous work from our group has screened FGFR1, IRF6, MSX1, TGFA, and TGFβ3 for mutations on cleft cases. Point mutations in MSX1 appear to contribute approximately to 2% of all CL/P cases [4]. FGFR1 point mutations also appear to contribute to CL/P. In addition, FGFR1 loss-of-function mutations can cause forms of Kallmann syndrome that mimic isolated CL/P at birth and during childhood. Mutations in IRF6, which cause the syndromic forms of clefts, the Van der Woude and popliteal pterygium syndromes [12], were not found in the same collection of cleft cases as the present study, although rare non-coding variants in conserved regions were disclosed. However, IRF6 is strongly associated with CL/P, and it is likely a genetic modifier for clefts [15,16]. Previously for TGFA, five variants in conserved non-coding segments were found in individual cases but not seen in 278 controls [17]. In the present study we found another nine non-coding rare variants in single individuals, but we did not find the original five reported rare variants or any coding mutation that could be etiologic. If these variants are disease-causing mutations, they could explain, not only the conflicting results from association studies of isolated orofacial clefts and TGFα variants [18], but also the linkage studies that suggest a cleft susceptibility loci in 2p13, the TGFα locus [7]. For TGFβ3, we previously reported one missense mutation (K130R) in a cleft palate–only case not seen in 350 controls [19]. We did not find this or any other mutations in TGFβ3 in our current study population.

Loss-of-function mutations in GLI2 are associated with pituitary anomalies and holoprosencephaly-like features [20]. In this report, the three pedigrees segregating GLI2 loss-of-function mutations with complete clinical information presented orofacial clefts and polydactyly. We found four missense mutations in GLI2 in highly conserved amino acids. One of the cases also presented with polydactyly (Table 3).

We performed linkage disequilibrium studies in the genes that we found potentially disease-causing missense mutations. We found association between a marker (rs2843159) and a haplotype in SKI in the Filipino population. This association was also found in an independent population dataset from South America. The SKI locus, 1p36.3, was previously suggested as a cleft susceptibility loci in Caucasians [21,22]. Shi null mice present with clefts involving the lip [22], and the association we found appears to be stronger when cases with the involvement of the lip are included.

The possible trend for an association between clefts with a palate phenotype and SATB2 are in agreement with the cytogenetic evidence. Deletions and balanced translocations point to the existence of a locus on 2q32-q33, for which haploinsufficiency results in isolated cleft palate. A mutation analysis of SATB2 (located at 2q32-q33) in 70 unrelated patients with isolated cleft palate only did not reveal any coding region variants [23]. However, a meta-analysis of 13 genome scans for clefts indicated 2q32-q35 as a clefting susceptibility locus [7]. We studied 184 cleft lip and palate cases and found one missense mutation in SATB2 (T190A) that was not seen in approximately 1,200 controls. Based on the linkage disequilibrium and mutation analysis results of our study, we believe a regulatory element outside SATB2 coding regions may be implicated in clefting.
In summary, point mutations in six of the 20 candidate genes selected from expression, animal, and human data may to contribute to about 5% of isolated clefts, more likely those with more-severe phenotypes and/or a positive family history. Etiologic variants in regulatory elements of SKI, JAG2, and MSX2 may contribute to isolated clefts as well. Predictions by ESEfinder (http://rulai.cshl.edu/tools/ESE) and PolyPhen regarding the function of the missense mutations found in this study, as well as exonic splicing enhancement and protein damaging, are challenging to interpret.

A major challenge in these studies was the frequent absence of a cleft phenotype in near relatives of an affected proband with a cleft and a rare missense mutation. In some cases these variants may not be etiologic, but in others, reduced penetrance for the cleft may be an active force as has been seen commonly in clefts [9,10] and other birth defects such as congenital heart disease. Similarly these mutations may only be modifiers of the phenotype.

Cases due to microdeletions or isodisomy may contribute to clefts as well. This study illustrates the validity of testing greater numbers of controls to determine rare polymorphic variants and prioritize functional studies for rare point mutations. Given other recent data on the roles of FGFR1, IRF6, and MSX1 in isolated CLP, one can begin to consider sequencing of a panel of high-probability candidate genes for genetic counseling indication. Although issues of penetrance and even etiology for any given mutation are not yet resolved, progress in this direction is now measurable.

Materials and Methods

Two collections of CLP cases, 91 from the Philippines and 93 from Iowa, United States, were used to search for mutations in 20 candidate genes (Table 1). We selected the more-severe cases from those available to us, and the sequenced samples were enriched by bilateral cleft lip and palate cases with a positive family history for clefts (39/91 from the Philippines and 16/93 from Iowa). Two cases were later found to have associated features—one with Stickler syndrome in which family history was initially positive and a second with polydactyly. Cases from the Philippines were studied under the auspices of Operation Smile International [24]. Patients were seen and examined by a board certified clinical geneticist (JCM or colleagues; see Acknowledgments) at one of four sites within the Philippines (Cavite, Calo, Cebu, and Negros). Iowa cases were studied under the auspices of Operation Smile International [24]. Patients were seen and examined by a board certified clinical geneticist (JCM or colleagues; see Acknowledgments) at one of four sites within the Philippines (Cavite, Calo, Cebu, and Negros). Iowa cases were collected through the Iowa Birth Defects Registry [25]. Signed consents were obtained from all participants before a blood sample was obtained. DNA was extracted according to published protocols.

For each of the 20 candidate genes, all exons and 5' and 3' untranslated regions were sequenced in both directions. Primer sequences and PCR conditions are available on our Web site (http://genetics.uiowa.edu/publications.html). Primers for FOXE1, GLI2, MSX2, OR2, and TGFBR1 were obtained from the literature [20,26-29]. Cycle sequencing was performed in a 20-μl reaction using 4 μl of Applied Biosystems Big Dye Terminator sequencing reagent, 1 μl of 5 μM sequencing primer, 1 μl of DMSO, 4 μl of 2× Taq Buffer, and 2.5 ng/100 base pair of DNA template. Following a denaturation step at 96 C for 30 s, reactions were cycle sequenced at 96 C for 10 s, 50 C for 5 s, and 60 C for 4 min for 40 cycles. Cleanup was performed using standard protocols. Samples were resuspended in 40 to 100 μl of ddH2O, and 2.5 μl were then injected on an Applied Biosystems 3700 sequencer. The Applied Biosystems sequencing software (version 2.1.2) was used for lane tracking and first pass base calling. Chromatograms were transferred to a Unix workstation, base called with PHRED (version 0.960289), assembled with PHRAP (version 0.960731), scanned by POLYPHRED (version 0.970312), and the results viewed with CONSED (version 4.0) [30]. When the results indicated a possible new variant, the case sample was resequenced, as well as other available family members, and population controls. These data were analyzed using the same method.

For any coding variant, we performed direct sequencing in 186 population-matched controls. Control populations were collected as described above for the cases and consisted of individuals with no GLI1 or other recognized birth defect from adults at the same sites where cases were collected. If the variant was found in one or more controls, it was considered a polymorphism. To expand the number of controls tested, we developed allele-specific assays for the LFX1, E2F2, T190A, SKI, A388V, SPRY2, D20A, and TBRX10 mutations. We tested them in the CEPH Diversity Cell Panel [31], which is comprised of 1,604 DNA samples from cultured lymphoblastoid cell lines derived from individuals representing 51 different human populations.

We also developed an assay for the MSX1 P147Q missense mutation, recently reported in three Vietnamese cleft families [5]. Besides the 1,064 CEPH Diversity Cell Line Panel controls, we tested the MSX1 P147Q assay in an additional 607 Filipino controls and in a collection of 1,468 cleft cases from the Philippines as well.

For the nine genes with potentially etiologic missense mutations, we identified orthologs through BLAST search of the non-redundant database using Homo sapiens FOXE1, GLI2, LHX8, JAG2, MSX2, SATB2, SKI, SPRY2, and TBRX10, as reference sequences. We performed protein sequence comparisons with the available species. We also used the ESEfinder software available online to predict the presence of exonic splicing enhancers [14], which appear to be prevalent, and may be present in most, if not all exons [32,33]. We screened the 141 exonic splicing silencer decamers that were identified by Wang et al. [34] to check if any of those could be affected by the missense mutations we found. Finally, we used the PolyPhen software, also available online, to predict the impact of the amino acid substitutions identified on the structure and function of the human protein [35-37].

Two single nucleotide polymorphisms in weak linkage disequilibrium with each other were selected for each population to perform linkage disequilibrium studies in the genes with missense mutations in cases but not in controls. Four single nucleotide polymorphisms were chosen for GLI2 based on the International HapMap Project’s linkage disequilibrium pattern of the gene (data not shown). Frequency of the alleles can be found in the supplemental material (http://genetics.uiowa.edu/publications.html). TaqMan-based assays [38] were performed on Applied Biosystems 7900 HT Sequence Detection System (Applied Biosystems, Foster City, California, United States). For one marker in SKI (rs2843159), we used a kinetic polymerase chain reaction assay previously reported [39]. These linkage disequilibrium studies were composed of 296 complete triads (mother/father/affected child) from the Philippines and 205 from Iowa. These samples were obtained as described above for cases and controls investigated by sequencing. The Family Based Association Test (FBAT) [40-42] program was used for this analysis. Significance figures were accounted for using Bonferroni correction taking into account the number of tests carried out [43]. With the Bonferroni correction, alpha is 0.0003 (0.05/192 comparisons) for the individual marker analysis and 0.0001 (0.05/584 comparisons) for the haplotype analysis.

Linkage disequilibrium studies for FOXE1 were previously reported in Marazita et al. [7]. A third clefting population sample set of 434 case/mother pairs from South America was used to replicate any significant association. These population samples are derived from ECLAMC, which is a hospital-based birth defects registry study that includes sites in Argentina, Bolivia, Brazil, Chile, Ecuador, Paraguay, Uruguay, and Venezuela. This study population has previously been described in detail [44,45]. To analyze the ECLAMC samples, the likelihood ratio test (LRT) of Weinberg [46] was applied under the assumption that the distribution of paternal alleles is the same as maternal.

Supporting Information

Figure S1. Protein Comparisons of the Available Gene Orthologs for the Mutations Found in the Present Study

Green bars indicate degree of conservation in each site. Amino acids in red indicate the mutation sites. All mutation comparisons are available as supplemental material at http://genetics.uiowa.edu/publications.html.

Table S1. Primers and PCR Conditions

Table S2. Exonic Splicing Enhancer Prediction Analysis for the
Misseense Mutations Found in the Present Study. Nucleotides in red indicate the mutation sites.

Table S3. Mutations Screened in the CEPH Diversity Cell Line Panel

Table S4. Markers Selected for Linkage Disequilibrium Studies

Table S5. Haplotype Frequencies

Table S6. Linkage Disequilibrium Studies of Candidate Genes for Clefting

Table S7. Haplotype Analysis

Table S8. SKI LRT Results for the South American (ECLAMC) Clefting Samples

Accession Numbers

The National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) assigned accession numbers to the genes discussed in this paper as Human MUC1 (NM_004475), GLI2 (NM_0044655), JAG2 (NM_002226), LHX8 (AY449521), and MSX2 (NM_002249). SATB2 (NM_015265), SKI (AY334556), SPRY2 (NM_005842), and TBX10 (AY229977).

Acknowledgments

We are indebted to all the families that participated in this project. Buena Nepomucena and Paul Romitti supported the sample collection. Diana Caprau provided sample management. Carla Nimishura, Sally Santiago, John Allaman, Bonnie Ludwig, Kristin Aquilino, Kristin Orr, Amy Mach, and Darren Schipper provided technical support. Clinical exams in the Philippines were carried out on some cases by Howard Saal and Chintoo Fong. Nastar Malik and Heiner Westphal assisted with the LHX8 studies. Maurice Payne completed preliminary work on INGFG. Marianné Timm and Renata Lucia Ferreira de Lima did preliminary work on JAG2. Mark Kresowik did some preliminary work on MX2 and TBX11. Elliott Hill helped with TGFB3 sequencing. Satoshi Suzuki assisted with the MSX1 P147Q assay. Margaret Cooper helped with the LRT analysis. Financial support was provided by the National Institutes of Health grants DE08559, DE10213, ES18876, and 5 S31 TW00622.

Competing interests

The authors have declared that no competing interests exist.

Author contributions.

ARV, IMO, EEC, and JCM conceived and designed the experiments. ARV, JRA, ED, TMF, FR, JH, RRS, YW, MJ, JP, and SEO performed the experiments. ARV, JRA, SDH, ED, TMF, FR, JH, RRS, YW, MJ, JP, SEO, IMO, EEC, DRF, RJ, MLM, and JCM analyzed the data. SDH, DRF, RJ, MLM, and JCM contributed reagents/materials/analysis tools. ARV and JCM wrote the paper.

References


