Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype-specific differences in linkage and association results

M. L. Marazita
A. C. Lidral
J. C. Murray

Please see article for additional authors.

DOI: https://doi.org/10.1159/000224636
Genome Scan, Fine-Mapping, and Candidate Gene Analysis of Non-Syndromic Cleft Lip with or without Cleft Palate Reveals Phenotype-Specific Differences in Linkage and Association Results

Mary L. Marazita\textsuperscript{a–c} Andrew C. Lidral\textsuperscript{e} Jeffrey C. Murray\textsuperscript{f} L. Leigh Field\textsuperscript{i} Brion S. Maher\textsuperscript{h}
Toby Goldstein McHenry\textsuperscript{a} Margaret E. Cooper\textsuperscript{a} Manika Govil\textsuperscript{a} Sandra Daack-Hirsch\textsuperscript{g}
Bridget Riley\textsuperscript{j} Astanand Jugessur\textsuperscript{k} Temis Felix\textsuperscript{l} Lina Morene\textsuperscript{e} M. Adela Mansilla\textsuperscript{f}
Alexandre R. Vieira\textsuperscript{a,b,d} Kim Doheny\textsuperscript{m} Elizabeth Pugh\textsuperscript{m} Consuelo Valencia-Ramirez\textsuperscript{n}
Mauricio Arcos-Burgos\textsuperscript{o}

\textsuperscript{a}Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, and \textsuperscript{b}Department of Human Genetics, Graduate School of Public Health, \textsuperscript{c}Department of Psychiatry, School of Medicine, and \textsuperscript{d}Department of Pediatric Dentistry, School of Dental Medicine, University of Pittsburgh, Pittsburgh, Pa.; Departments of \textsuperscript{e}Orthodontics, College of Dentistry, \textsuperscript{f}Pediatrics College of Medicine, and \textsuperscript{g}College of Nursing, University of Iowa, Iowa City, Iowa; \textsuperscript{h}Department of Psychiatry, Virginia Commonwealth University, Richmond, Va., USA; \textsuperscript{i}Department of Medical Genetics, University of British Columbia, and B.C. Research Institute for Children’s and Women’s Health, Vancouver, BC, Canada; \textsuperscript{j}New York University School of Medicine, New York, N.Y., USA; \textsuperscript{k}Craniofacial Development, Musculoskeletal Disorders, Murdoch Childrens Research Institute, Royal Children’s Hospital, Melbourne, Australia; \textsuperscript{l}Hospital de Clínicas Porto Allegre-RS, Universidade Federal do Rio Grande do Sul, Brazil; \textsuperscript{m}Center for Inherited Disease Research, Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Md., USA; \textsuperscript{n}College of Dentistry, Universidad de Antioquia, Medellin, Colombia, and \textsuperscript{o}Division of Child Psychiatry, Department of Psychiatry and Behavioral Sciences, Leonard M. Miller School of Medicine, University of Miami, Miami, Fla., USA

Abstract

Objectives: Non-syndromic orofacial clefts, i.e. cleft lip (CL) and cleft palate (CP), are among the most common birth defects. The goal of this study was to identify genomic regions and genes for CL with or without CP (CL/P). Methods: We performed linkage analyses of a 10 cM genome scan in 820 multiplex CL/P families (6,565 individuals). Significant linkage results were followed by association analyses of 1,476 SNPs in candidate genes and regions, utilizing a weighted false discovery rate (wFDR) approach to control for multiple testing and incorporate the genome scan results. Results: Significant (multipoint HLOD $\geq 3.2$) or genome-wide-significant (HLOD $\geq 4.02$) linkage results were found for regions 1q32, 2p13, 3q27-28, 9q21, 12p11, 14q21-24 and 16q24. SNPs in $\text{IRF6}$ (1q32) and in or near $\text{FOXE1}$ (9q21) reached formal genome-wide wFDR-adjusted significance. Further, results were phenotype dependent in that the $\text{IRF6}$ region results were most significant for families in which affected individuals have CL alone, and the $\text{FOXE1}$ region results were most significant in families in which some or all of the affected individuals have CL with CP. Conclusions: These results highlight the importance of careful phenotypic delineation in large samples of families for genetic analyses of complex, heterogeneous traits such as CL/P.

Mary L. Marazita, PhD
Center for Craniofacial and Dental Genetics
University of Pittsburgh, Suite 500 Bridgeside Point
100 Technology Dr., Pittsburgh, PA 15219 (USA)
Tel. +1 412 648 8380, Fax +1 412 648 8779, E-Mail marazita@pitt.edu
Introduction

Orofacial clefts (OFC) are a major public health problem, affecting one in every 500–1000 births worldwide. It has become increasingly apparent that the genetic contribution to OFC is complex, probably heterogeneous, and likely due to interacting effects of multiple loci coupled to environmental covariates. Over 400 syndromes have been described in which a cleft of the lip and/or palate is a feature [1]. Among these syndromes are numerous genetic examples such as cytogenetic abnormalities (e.g., trisomies 13 and 18, 4p–), and single gene Mendelian disorders (e.g. van der Woude syndrome, Stickler syndrome). Non-syndromic (sometimes termed ‘isolated’) cleft lip with or without cleft palate (NS CL/P) is the most prevalent type of OFC, and is the focus of the current study.

Familial aggregation of NS CL/P has been reported in the scientific literature for 250 years [2–7]. Since then, segregation analyses of CL/P have supported models that include genes of major effect [8–10], and analyses of recurrence risk patterns have been consistent with estimates ranging from 3 to 14 genes (possibly interacting) for CL/P [11–15].

Mutation screens of more than 20 NS CL/P candidate genes find that 2–6% of the total number of individuals with NS CL/P have mutations in genes such as MSX1, FOXE1, GLI2, JAG2, LHX8, SATB2, RYK1 and others [16–18]. That is, the large majority of individuals with NS CL/P (94–98%) do not have mutations in any of a wide range of plausible candidate genes. In parallel, many candidate gene association studies have also been carried out seeking specific polymorphic variants that increase the risk of NS CL/P [1, 19–22].

Most notably, the gene identified in van der Woude syndrome (IRF6 [23]) has been shown by our group [24] and confirmed in multiple other populations (Italy [25]; Belgium [26]; US [27]; Thailand [28]; US/Taiwan/Singapore/Korea [29]; South America [30]; Norway [31]) to show highly significant association with NS CL/P and may explain about 12–18% of NS CL/P [24]. Recently we have identified a specific SNP (rs642961) in IRF6 that disrupts the binding site for the transcription factor AP-2α, and that represents the etiologic locus within IRF6, at least in some populations [32].

In addition to analyses of candidate genes/loci, several genome-wide linkage screens of NS CL/P have been published [33–38]. About 8–10 regions have positive, although not highly significant, results in the individual studies. We performed a genome scan meta-analysis of these 5 published genome scans plus data from our additional studies in six countries and found several regions with genome-wide significant results [38], notably a novel region on 9q21. The purpose of this study is to describe a larger genome-wide scan for NS CL/P with twice as many families and individuals as were reported previously [38], to summarize the results of our follow-up fine mapping and candidate gene efforts, and to perform phenotype-specific analyses. Thus the aim was to identify those genetic regions that appear to contribute to OFC overall, and also those regions that appear to have an effect on specific OFC phenotypes.

Subjects and Methods

Study Populations

This study was conducted in two phases: (1) linkage genome-scan, and (2) fine mapping and candidate gene studies. For the genome-scan, there were 820 families ascertained in six countries (Philippines, Colombia, China, India, Turkey, USA), with 6,565 total individuals (summarized in table 1). Of the total family members, 4,373 were genotyped (1,514 affected and 2,859 unaffected). For fine-mapping and candidate gene studies, there were 861 families, with 7,047 total individuals (summarized in table 2), representing a subset of the 820 genome-scan families plus additional families that were ascertained after the genome-scan was completed. Most of the families were extended multiplex kindreds, i.e. multigenerational families with ≥2 affected individuals. The largest family had 51 individuals; there were 5 pedigrees with 40–50 individuals, 10 with 30–40, 46 with 20–30, many with 10–20 individuals. Each study had approval by the appropriate institutional review boards, and all study subjects provided informed consent to participate. Note that the original 388 genome-scan families [38] were also included in the current study (denoted ‘CIDR-7 families’ in the previous report [38]).

The phenotype was NS CL/P, i.e. for families to be included, it was necessary that the proband have NS CL/P (i.e. no other anomalies), and that no other family member have an indication of an OFC syndrome (e.g. lip pits). Each study included evaluations of family members by clinical geneticists to rule out syndromic forms of CL/P.

For some analyses, the total families were divided into the following non-overlapping subsets based on family members reported to be affected (whether or not all affected family members participated in these studies): CL, those families in which all affected family members had CL only; CLP (all affected family members had CL plus CP); CL+CLP (at least one affected family member had CL only and at least one had CL plus CP). The numbers of families and individuals in each of these subsets are summarized in table 1 (genome-scan dataset) and table 2 (fine-mapping dataset). There was one additional subset that had very few families and was therefore not included in the subset studies reported here, i.e. families in which at least one affected family member had CP only.

Genome Scan Genotyping

The genome scans were conducted in multiple batches (the previous report [38] summarized results from the first two batch-
of genotyping. Microsatellite genotyping for all batches was performed at the Center for Inherited Disease Research (CIDR) using STRP markers with average spacing of about 9 cM (1–19 cM). Marker set details and methods are available at www.cidr.jhmi.edu.

Trace data was reviewed and genotypes were called using the software package Genotyper from Applied Biosystems. The raw genotype sizes for each genotype were exported from Genotyper and binned using SAS. Study samples were plated with 2 CEPH Utah control samples 133 1 - 1 and 133 1 - 2 repeated on each plate. For each microsatellite marker the CEPH control sample sizes for each plate were compared to the project average and used to adjust the raw sizes for all data for that plate. Data with similar raw sizes were grouped into proposed ‘binned’ alleles. Plots were made for each marker showing the overall as well as by plate allele sizes and proposed bins for both the raw and adjusted size data. Each plot was manually reviewed by at least two data analysts. Trace data for unexpected sizes for CEPH controls and outliers that didn’t fall within bins were reviewed in Genotyper. Trace data was also reviewed for markers with high numbers of mendelian errors according to Pedcheck [39]. Binning parameters and plate adjustments were done manually if needed.

After the entire dataset was genotyped, all alleles were re-binned across batches before performing the analyses reported here. The final combined dataset consisted of 375 STRP markers which were genotyped in all three sample submissions. Quality

### Table 1. Summary of families and individuals ascertained for genome-scan linkage analyses

<table>
<thead>
<tr>
<th>Country</th>
<th>Families</th>
<th>Individuals</th>
<th>Subsets&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>affected</td>
<td>CL</td>
<td>CLP</td>
</tr>
<tr>
<td></td>
<td>genotyped</td>
<td>n fams</td>
<td>n fams</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>total</td>
</tr>
<tr>
<td>China</td>
<td>72</td>
<td>98</td>
<td>17</td>
</tr>
<tr>
<td>Colombia</td>
<td>163</td>
<td>266</td>
<td>15</td>
</tr>
<tr>
<td>India</td>
<td>11</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Philippines</td>
<td>462</td>
<td>928</td>
<td>44</td>
</tr>
<tr>
<td>Turkey</td>
<td>6</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>USA</td>
<td>106</td>
<td>196</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>820</td>
<td>1,514</td>
<td>104</td>
</tr>
</tbody>
</table>

<sup>a</sup> Subsets: CL = those families in which all affected members had CL alone; CLP = those families in which all affected members had CL and CP; CL+CLP = those families in which at least one affected member had CL alone and at least one had CL+CP; n fams = number of families.

### Table 2. Summary of families and individuals ascertained for fine-mapping studies

<table>
<thead>
<tr>
<th>Country</th>
<th>Families</th>
<th>Individuals</th>
<th>Subsets&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>affected</td>
<td>CL</td>
<td>CLP</td>
</tr>
<tr>
<td></td>
<td>genotyped</td>
<td>n fams</td>
<td>n fams</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
<td>total</td>
</tr>
<tr>
<td>China</td>
<td>104</td>
<td>168</td>
<td>33</td>
</tr>
<tr>
<td>Colombia</td>
<td>232</td>
<td>349</td>
<td>12</td>
</tr>
<tr>
<td>India</td>
<td>53</td>
<td>94</td>
<td>42</td>
</tr>
<tr>
<td>Philippines</td>
<td>221</td>
<td>600</td>
<td>47</td>
</tr>
<tr>
<td>Spain</td>
<td>36</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>Turkey</td>
<td>28</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>USA</td>
<td>187</td>
<td>231</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>861</td>
<td>1,515</td>
<td>165</td>
</tr>
</tbody>
</table>

<sup>a</sup> Subsets: CL = those families in which all affected members had CL alone; CLP = those families in which all affected members had CL and CP; CL+CLP = those families in which at least one affected member had CL alone and at least one had CL+CP; n fams = number of families.
control was monitored by investigator-provided blind duplicate samples. 220 blind duplicate pairs across all three sample submissions resulted in a reproducibility rate of 99.95%.

**Candidate Gene SNP Genotyping**

After the results of the linkage genome scan, the families were genotyped by CIDR for a custom panel of 1,536 single nucleotide polymorphisms (SNPs), of which 1,476 were analyzed (list available on request, see below for reasons some SNPs were not analyzed).

SNPs were chosen to saturate the 1-LOD regions for each genome-wide significant region found in the linkage scan, targeting all known genes in each region, with 2–6 SNPs per gene and filling intergenic gaps >100 kb with 1 SNP per 20 kb. In addition, SNPs in cleft candidate genes were included in the custom panel, and were chosen from a variety of resources, including published linkage and association studies on clefts, genome-wide scans, gene knockout experiments in mice, studies of chromosomal rearrangements in humans, and gene-expression analyses in human and mouse embryonic tissues [20, 40–43], including the Craniofacial and Oral Gene Expression Network [44] (COGENE). We also searched the Serial Analysis of Gene Expression (SAGE) libraries to see whether a particular gene of interest is expressed in the relevant embryonic tissues at the pertinent developmental stage (weeks 5–6 for fusion of the embryonic lip and weeks 7–10 for fusion of the palatal shelves [45]). Data from two gene expression approaches were also considered, the ENU project [46], and the Developmental Genome Anatomy Project (DGAP [47]). Finally genes involved in Mendelian forms of clefting, particularly those that may manifest as phenocopies of NS CL/P (identified through OMIM) were considered.

To guide the selection of SNPs within the candidate genes and regions, the genome browser of the International HapMap Consortium was used. SNPs were prioritized according to prior evidence of an association with clefting, a preference for coding SNPs and those located in putative regulatory regions, haplotype-tagging properties (htSNPs), and minor allele frequency (MAF) of at least 10%, ideally across the multiple ethnicities represented in this study.

A combination of software, including HAPLOVIEW version 2.05, BEST (Best Enumeration of SNP Tags [48]), and SNP Browser (Applied Biosystems; Foster City, Calif., USA), was used to evaluate MAF, deviations from Hardy-Weinberg equilibrium (HWE), inter-marker distance, as well as LD patterns and haplotype block structures (for the selection of htSNPs). SNP assays were designed by Illumina (San Diego, Calif., USA) and a ‘design score’ was computed for each SNP using an algorithm that rigorously tests the performance of that SNP on an Illumina GoldenGate platform.

SNP genotyping was performed at CIDR on a BeadLab system using the Illumina GoldenGate technology [49] (Illumina; San Diego, Calif., USA). Genotypic data for 1,489 of 1,536 attempted SNPs were released by CIDR. The overall rate of missing genotypic data was 0.35%. As part of quality control, CIDR genotyped a parent-child trio plus one duplicate child, which yielded a duplicate error rate of 0.009% and an overall parent-child discordance rate of zero. Of the 1,489 SNPs released after CIDR quality control checks, 13 more were not included in our analysis due to poor performance in one or more of the individual study populations (e.g. low MAF or high rates of mendelian inconsistencies). Therefore, results are reported here for the remaining 1,476 SNPs.

### Statistical Genetic Analyses

**Preliminary Analyses**

The inheritance of each marker (genome-wide STRPs and custom SNPs) was assessed with PedCheck [39] to test for inconsistencies due to non-paternity or other errors. Marker allele frequencies are required by linkage analysis approaches and were estimated in the founders of the families, separately by country due to the diverse ethnicities. Other genetic model parameters are summarized in table 3, and were taken from the results of segregation analysis (Philippines – unpublished results; China [10]; India [50]; Colombia [51]; Caucasians [9, 52, 53]).

**Parametric Linkage Analyses**

Standard multipoint parametric linkage statistics were calculated at each of the 375 genome-scan STRPs, in particular the heterogeneity LOD score or HLOD. HLODs are based on the admixture heterogeneity test [54] where the recombination fraction (θ) and the proportion of linked families (α) are simultaneously estimated. Simulation studies have shown that although the estimate of the proportion of linked families may not be precise, HLODs are a powerful method for detecting linkage in the presence of heterogeneity [55–57]. The descent graph method [58] implemented in computer program SIMWALK2 was used for the multipoint HLOD calculations.

### Table 3. Summary of genetic model parameters used for parametric linkage analysis in each country, plus weights for meta-analysis

<table>
<thead>
<tr>
<th>Country</th>
<th>Recessive</th>
<th></th>
<th></th>
<th>Dominant</th>
<th></th>
<th></th>
<th></th>
<th>Weights for meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FreqA</td>
<td>penAA</td>
<td>penAB</td>
<td>penBB</td>
<td>FreqA</td>
<td>penAA</td>
<td>penAB</td>
<td>penBB</td>
</tr>
<tr>
<td>China</td>
<td>0.05</td>
<td>0.002</td>
<td>0.002</td>
<td>0.9</td>
<td>0.001</td>
<td>0.002</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Colombia</td>
<td>0.05</td>
<td>0.002</td>
<td>0.002</td>
<td>0.9</td>
<td>0.001</td>
<td>0.002</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>India</td>
<td>0.05</td>
<td>0.002</td>
<td>0.002</td>
<td>0.9</td>
<td>0.001</td>
<td>0.002</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Philippines</td>
<td>0.04</td>
<td>0.002</td>
<td>0.002</td>
<td>0.9</td>
<td>0.002</td>
<td>0.002</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Turkey</td>
<td>0.05</td>
<td>0.002</td>
<td>0.002</td>
<td>0.9</td>
<td>0.001</td>
<td>0.002</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>USA; Pittsburgh</td>
<td>0.05</td>
<td>0.002</td>
<td>0.002</td>
<td>0.9</td>
<td>0.001</td>
<td>0.002</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>USA; other than Pittsburgh</td>
<td>0.05</td>
<td>0.002</td>
<td>0.002</td>
<td>0.9</td>
<td>0.001</td>
<td>0.002</td>
<td>0.6</td>
<td>0.6</td>
</tr>
</tbody>
</table>

154 Hum Hered 2009;68:151–170

Marazita et al.
HLOD calculations were done under the best-fitting dominant and recessive models for each study population (estimated from segregation analysis, see preliminary analyses section above and table 3). Maximizing LOD scores over a range of models is valid for simultaneously evaluating linkage and determining the most likely genetic model (without adjustment to significance levels and without need to correct for ascertainment) provided that there is indeed linkage [59]. Furthermore, if an oligogenic model is suspected (as seems likely for CL/P), or if significant heterogeneity exists, some of the causal genes may act in a dominant fashion and others recessive.

To combine the multipoint results across the study populations we summed the multipoint HLODs. Because each study population had different allele frequency estimates and different genetic model parameters it was not optimal to perform a combined linkage analysis pooling all families across all populations. Simulation studies [55] have shown that HLOD analyses that pooled multiple datasets resulted in a loss of power for detecting linkage compared to summing the HLODs from individual studies.

For determining genome-wide significance for the multipoint linkage calculations of the families, standard guidelines [60] were followed. A Bonferroni correction was applied to account for the multiple markers, data subsets, and genetic models tested. To do so, the desired \( \alpha \) level of 0.05 was divided by \( \frac{375}{(\text{number of markers}) \times 4 \times (\text{number of data subsets}) \times 2 \times (\text{number of genetic models tested})} \) to yield a Bonferroni-adjusted alpha of 0.000017. The \( \chi^2 \) corresponding to the p value (i.e. 18.5) was divided by \( 2 \log(10) = 4.605 \) to yield 4.02 as the LOD score threshold for genome-wide significance for this study. Values between 3.2 and 4.02 were considered as significant but not genome-wide significant.

**Genome Scan Meta-Analysis**

As an additional method used primarily to estimate the weighting factors for the weighted FDR approach summarized below, we used the genome scan meta-analysis method (GSMA [38, 61, 62]). The GSMA is a nonparametric rank ordering method that can combine genome scan methods across studies with different markers and different statistical tests. In simulation studies the GSMA detected linkage with power comparable to or greater than that obtained by performing a combined linkage analysis of all data [62, 63].

For the GSMA procedure, the genome was divided into bins with bin-width selected such that there were at least two bins on the smallest chromosome and at least one marker was genotyped within each bin. Therefore, for combining the current 10 cm genome scans a bin width of 30 cm was selected (i.e. 130 bins across the genome). For each of the individual populations, bins were assigned a rank (R, with values from 1 to 130) according to the maximum linkage statistic of markers in each bin (multipoint HLOD scores). Any tied bins were assigned equal R’s based on the mean of the sequential ranks for those bins.

Because the study populations covered a wide range of sample sizes (see table 1), we weighted the R statistics based on sample size. Various weighting strategies have been proposed for the GSMA [61, 62], and simulation studies showed that weighting increased the power of the GSMA to detect linked loci [62]. We used a weighting factor based on the total number of genotyped individuals (N-genotyped) – the ranks within each population were multiplied by the square root of N-genotyped divided by the mean of N-genotyped across all studies. The weighting factors calculated for each population are listed in table 3.

The GSMA identified 30 cm bins that are best supported across the studies. In order to narrow the regions of positive findings, we used an extension of the GSMA that involves repeating the GSMA with different bin starting points and then determining the Minimum Region of Maximum Significance (MRMS [64]). In brief, we repeated the GSMA twice more shifting the starting point for the binning procedure to first 10 cm and then to 20 cm. This determined the 10 cm MRMS for each positive finding. Given that the genome scan STRP panel averaged 10 cm between markers, 10 cm was the limit of resolution for the meta-analyses.

**Association Analysis Methods (Fine-Mapping and Candidate Genes)**

The transmission disequilibrium test (TDT) was used to assess association in the presence of linkage disequilibrium between the 1,476 fine-mapping markers and CL/P. The Family Based Association Test extension of the TDT (FBAT [65–67]) was used to assess association between each SNP and CL/P. To control for multiple testing and to include information from the prior linkage genome scan we performed a weighted approach to false discovery rate (wFDR [68]). In this approach, per marker FBAT p values were adjusted using weights derived from the GSMA/MRMS results, and allowing for a maximum 10% FDR.

**Results**

**Multipoint Linkage Results (HLOD and GSMA)**

Figure 1 shows a summary of the multipoint HLOD results for the total dataset, and for the CL, CLP and CL+CLP subsets. Depicted are the largest multipoint HLODs on each chromosome, under dominant and recessive models. The HLOD genome scan revealed genome-wide significant linkage results (i.e. multipoint HLOD \( \geq 4.02 \)) in the regions 3q27–28 (under a dominant model for CL/P), 9q21 (dominant model), and 14q21–24 (recessive model). Three additional regions reached nominal significance (i.e. multipoint HLOD \( \geq 3.2 \)): 1q32 (under a dominant model for CL/P), 2p13 (dominant model), and 16q24 (recessive model). Of those regions, 1q32, 9q21, 12p11, 14q21–24 and 16q24 were also statistically significant in the GSMA analysis. Figure 2A–G shows the detailed HLOD and GSMA results for each of the chromosomal regions with significant results.

As can be seen in figure 1, the phenotypic subsets have differing results. In the CL subset, the region on chromosome 1q32 was significant under a dominant model (see fig. 2A). In the CL+CLP subset, regions on chromosome 9q21 (dominant) and 16q24 (recessive) were significant (see fig. 2D and G for details). In the CLP subset a region
on chromosome 12p11 was significant under a dominant model (see fig. 2E). Table 4 summarizes the chromosomal regions with significant linkage results in the TOTAL dataset and the phenotypic subsets, plus potential CL/P candidate genes in those significant regions.

**Fine-Mapping and Candidate Gene Analyses**

Figure 3 shows the weighted p values and weighted FDR alpha levels for the total dataset, and the CLP and CL+CLP datasets. One SNP in IRF6 and 3 SNPs in or near FOXE1 were the only ones reaching formal weighted-FDR-adjusted significance (p < 10^-7, and p < 10^-6, respectively) in the total dataset. Although not reaching formal genome-wide significance, additional SNPs on 1q, 6q, and 9q were near significant (p < 0.001, results not shown in detail).

Of the phenotypic subsets, only CLP had SNPs reaching genome-wide significance: i.e., 5 SNPs in or near FOXE1 on 9q. Although not reaching genome-wide significance, the most significant SNP in both the CL and CL+CLP phenotypic subsets was in IRF6 (p < 0.001 and p < 0.002, respectively), and was the same SNP significant in the TOTAL dataset. Table 5 summarizes the genome-wide significant SNPs in the total dataset and in the CLP subset.

**Discussion**

The genome scan revealed multiple significant linkage results (i.e. multipoint HLOD ≥3.2) in the regions 1q32, 2p13, 3q27–28, 9q21, 14q21–24 and 16q24 for the TOTAL dataset, with the 3q, 9q and 14q regions also genome-wide significant (HLOD ≥4.02). The 1q32 region result was also significant in the CL subset but not the others, implying that the significant linkage was due to the CL families. Similarly, the 9q21 and 16q24 results were also genome-wide significant in the CL+CLP subset. In the CLP subset, an additional region of significance was found for the
The remaining two regions (2p11, 3q27–28) were not significant in any individual subset, implying that these regions may be involved in OFC overall, rather than any specific phenotype. Also, note that in each case where there were significant findings in one of the phenotypic subgroups, the estimated proportion of linked families (α) was larger in the subgroup than in the total dataset (see table 4), further supporting the notion that phenotypic sub-grouping may be a useful approach to reduce heterogeneity across cleft families.

Table 4. Significant linkage and meta-analysis results

<table>
<thead>
<tr>
<th>Chromosomal region</th>
<th>Phenotypic subset</th>
<th>Maximum summed multipoint HLOD</th>
<th>Linkage peak: cM location</th>
<th>Model</th>
<th>Average α</th>
<th>GSMA/MRMS peak (p value)</th>
<th>Candidate genes in region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q32</td>
<td>TOTAL</td>
<td>3.35</td>
<td>170.3</td>
<td>dom</td>
<td>0.29</td>
<td>150–160 (0.02)</td>
<td>IRF6</td>
</tr>
<tr>
<td></td>
<td>CL</td>
<td>3.24</td>
<td>171.0</td>
<td>dom</td>
<td>0.47</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>2p13</td>
<td>TOTAL</td>
<td>3.25</td>
<td>69.0</td>
<td>dom</td>
<td>0.31</td>
<td>–</td>
<td>TGFα</td>
</tr>
<tr>
<td>3q27–28</td>
<td>TOTAL</td>
<td>4.13</td>
<td>204.2</td>
<td>dom</td>
<td>0.34</td>
<td>–</td>
<td>TP63</td>
</tr>
<tr>
<td>9q21</td>
<td>TOTAL</td>
<td>5.37</td>
<td>118.2</td>
<td>dom</td>
<td>0.41</td>
<td>100–110 (0.0007)</td>
<td>FOXE1, PTCH, ROR2, TGFBR1, ZNF189</td>
</tr>
<tr>
<td></td>
<td>CL+CLP</td>
<td>5.44</td>
<td>116.4</td>
<td>dom</td>
<td>0.46</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>12p11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14q21–24</td>
<td>TOTAL</td>
<td>4.18</td>
<td>75.1</td>
<td>rec</td>
<td>0.31</td>
<td>60–70 (0.01)</td>
<td>PAX9, TGFβ3, BMP4</td>
</tr>
<tr>
<td>16q24</td>
<td>TOTAL</td>
<td>3.56</td>
<td>84.5</td>
<td>rec</td>
<td>0.16</td>
<td>100–110 (0.01)</td>
<td>CRSPDL2, FOXC2</td>
</tr>
<tr>
<td></td>
<td>CL+CLP</td>
<td>3.42</td>
<td>81.0</td>
<td>rec</td>
<td>0.24</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

12p11 region. The remaining two regions (2p11, 3q27–28) were not significant in any individual subset, implying that these regions may be involved in OFC overall, rather than any specific phenotype. Also, note that in each case where there were significant findings in one of the phenotypic subgroups, the estimated proportion of linked families (α) was larger in the subgroup than in the total dataset (see table 4), further supporting the notion that phenotypic sub-grouping may be a useful approach to reduce heterogeneity across cleft families.

Follow-up fine-mapping association studies found SNPs in IRF6 (chromosome 1q) and in or near FOXE1 (chromosome 9q) that reached formal FDR-adjusted significance (see table 5), and SNPs in 6q were near signifi-

Table 5. Genome-wide significant SNP results (from weighted FDR analyses of FBAT results) in the TOTAL dataset and CLP phenotypic subset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Chromosome</th>
<th>SNP name</th>
<th>SNP</th>
<th>Associated allele (freq)</th>
<th>Gene</th>
<th>Chromosome location</th>
<th>Weighted p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL</td>
<td>1q32.3-q41</td>
<td>rs2013162</td>
<td>A/C</td>
<td>C (62%)</td>
<td>IRF6</td>
<td>208035307</td>
<td>1.0 × 10⁻⁷</td>
</tr>
<tr>
<td></td>
<td>9q22</td>
<td>rs1443434</td>
<td>G/T</td>
<td>T (75%)</td>
<td>FOXE1</td>
<td>99657300</td>
<td>7.0 × 10⁻⁷</td>
</tr>
<tr>
<td></td>
<td>9q22</td>
<td>rs993501</td>
<td>C/T</td>
<td>T (81%)</td>
<td>FOXE1</td>
<td>99663198</td>
<td>2.1 × 10⁻⁶</td>
</tr>
<tr>
<td></td>
<td>9q22.33</td>
<td>rs6586</td>
<td>C/T</td>
<td>T (63%)</td>
<td>C9orf156</td>
<td>99706752</td>
<td>5.8 × 10⁻⁵</td>
</tr>
<tr>
<td>CLP</td>
<td>9q22</td>
<td>rs1443434</td>
<td>G/T</td>
<td>T (72%)</td>
<td>FOXE1</td>
<td>99657300</td>
<td>4.0 × 10⁻⁷</td>
</tr>
<tr>
<td></td>
<td>9q22.33</td>
<td>rs6586</td>
<td>C/T</td>
<td>T (59%)</td>
<td>C9orf156</td>
<td>99706752</td>
<td>7.5 × 10⁻⁶</td>
</tr>
<tr>
<td></td>
<td>9q22</td>
<td>rs993501</td>
<td>C/T</td>
<td>T (77%)</td>
<td>C9orf156</td>
<td>99663198</td>
<td>1.4 × 10⁻⁵</td>
</tr>
<tr>
<td></td>
<td>9q22</td>
<td>rs1877431</td>
<td>A/G</td>
<td>T (66%)</td>
<td>C9orf156</td>
<td>99573968</td>
<td>3.3 × 10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>9q22.3</td>
<td>rs4743106</td>
<td>T/MOD</td>
<td>T (78%)</td>
<td>TMOD</td>
<td>99353614</td>
<td>3.3 × 10⁻⁴</td>
</tr>
</tbody>
</table>
significant. Consistent with the linkage results, the fine-mapping results were also phenotype dependent. The IRF6 SNP rs2013162 (significant in the TOTAL dataset) was only positive in the CL and CL+CLP subsets (although not reaching formal genome-wide significance in those subsets). Similarly, FOXE1 SNPs (significant in the TOTAL dataset) were genome-wide significant only in the CLP subset.

Following is a brief discussion of the most notable results by chromosome. Although not summarized in detail, several of these regions also have chromosomal rearrangements reported in CL/P [47, 69, 70].

**Chromosome 1q**

The 1q32 region with significant linkage and association results is the location for the IRF6 gene that was identified as the etiologic locus for van der Woude syndrome (VDWS, MIM# 119300 [23]) and also significantly associated with non-syndromic CL/P (SNP rs2235371 [22, 24–26, 28–30, 71]).

Recently we have identified a specific IRF6 SNP (rs642961) where the ancestral allele is highly evolutionarily conserved and where functional assays show disruption of an AP-2α binding site that likely represents the major NS CL/P etiologic locus within IRF6 [32]. Notably the results with SNP rs642961 were significant only in families with one or more CL-alone affected members.

---

**Fig. 2.** Summed multipoint HLOD plots for each chromosome that had a maximum summed HLOD ≥3.2 (under the best genetic model for each chromosome), in the entire dataset (TOTAL), and the subsets – see definitions in figure 1. For each of those chromosomes, also shown are the graphs of the Minimum Regions of Maximal Significance (MRMS). These graphs summarize the process of repeating the Genome Scan Meta Analyses (GSMA), shifting the bins in order to narrow the region of potential involvement with CL/P. For those chromosomes with statistically significant GSMA/MRMS results (i.e. p values ≤0.05), dashed vertical lines indicate the 10 cM MRMS. a Chromosome 1 under a dominant model for TOTAL and for the CL subset.
consistent with the linkage and association results reported here. In the current study, \textit{IRF6} SNP rs2013162 reached genome-wide significant association with CL/P in the TOTAL dataset; further the association results were positive in only the CL and CL+CLP subsets (although not reaching formal genome-wide significance in those subsets). Based on Haploview version 4.1 [72], the current study SNP (rs2013162) lies between the other two SNPs, but is not in LD with them.

The current results demonstrating significant linkage to the \textit{IRF6} region provide powerful support for the \textit{IRF6} association findings. Interestingly, until the current study we only found weak linkage signals to the 1q32 region [38] and/or the \textit{IRF6} locus itself [24] (LODs <1.0); the current study thus demonstrates the utility of large sample sizes with careful sub-phenotyping in detecting subtle effects that otherwise are only detectable with association methods.

\textbf{Chromosome 2p and 2q}

Note that in our previous linkage report [38], a 2q32–35 region had highly significant GSMA results. In the current study with double the sample size there are no longer genome-wide significant results with this region. The 2q32–35 region contains the gene for DNA-binding protein SATB2 (a.k.a. \textit{KIAA1034}) that has been identified through translocation breakpoint analysis as a gene involved in cleft palate [73], and that also shows site- and
stage-specific expression in murine palate development. It may be that some population or phenotypic subsets led to the previous positive results and additional analyses are on-going to continue to investigate SATB2.

The 2p13 region with significant linkage results contains TGFA, the gene with the first reported association with CL/P [74] and numerous confirmatory reports [19, 75]. There were no genome-wide significant association results with TGFA fine-mapping SNPs (rs3732253, rs1807968, rs374640, rs377122), but the positive linkage results warrant continued study of this region.

**Chromosome 3q**

The 3q27–28 region had genome-wide significant linkage results, consistent with our previously reported result near this region in Chinese families [76], but there have been no other published reports of either positive linkage or positive association of non-syndromic CL/P to this region. A potential candidate gene in this region is TP63.

TP63 encodes a member of the p53 family of transcription factors. An animal model, p63−/− mice, has been useful in defining the role this protein plays in the development and maintenance of stratified epithelial tissues. p63−/− mice have several developmental defects which include the lack of limbs and other tissues, such as teeth and...
mammary glands, which develop as a result of interactions between mesenchyme and epithelium. Furthermore, recently completed detailed expression studies [77] during murine craniofacial development to dissect the molecular pathogenesis of the bilateral cleft lip and cleft palate seen in Tp63-deficient mice. Analysis of key signaling molecules revealed that the craniofacial defects resulted from increased Bmp4 signaling acting antagonistically on Fgf8 and Shh [77].

In humans, mutations in the TP63 gene are associated with ectodermal dysplasia and cleft lip/palate syndrome 3 (EEC3 [78, 79]); split-hand/foot malformation 4 (SHFM4); Hay Wells syndrome [80] (ankyloblepharon-ectodermal defects-cleft lip/palate); ADULT syndrome (acro-dermato-ungual-lacrimal-tooth); limb-mammary syndrome; Rap-Hodgkin syndrome (RHS [81, 82]) and NS OFC [83]. Both alternative splicing and the use of alternative promoters results in multiple transcript variants encoding different proteins, which may underlie the wide phenotypic spectrum associated with mutations in TP63 [84, 85].

Therefore, the custom SNP panel included 10 SNPs within TP63: rs4396880, rs1920266, rs4575879, rs7616178, rs4607088, rs7619526, rs4686529, rs7619549, rs9810322, and rs1515490. None of these SNPs reached genome-wide significance in the wFDR analyses, but given the strong linkage signal and the biological plausibility of this gene, our group is continuing analyses in this region. Further-
more, given the interacting nature of the involvement of Tp63 in the murine model [77] we are simultaneously exploring TP63, BMP4, FGF8 and SHH in our multiplex families. Notably, as described below, we also saw genome-wide significant linkage to region 14q which is near the location for BMP4.

**Chromosome 9q**

Region 9q21 had the most significant linkage result in our previous report [35]. This region also reached genome-wide significance in the current study, and showed the most positive HLOD and GSMA results in the TOTAL and CL+CLP datasets. There are a number of possible candidate genes in this region including the human homolog of patched (PTCH, 9q22.3), receptor tyrosine kinase-like orphan receptor 2 (ROR2, 9q22), transforming growth factor beta receptor type 1 (TGFBR1, 9q33-q34), zinc finger protein 189 (ZNF189, 9q22-q31) and FOXE1 (a.k.a. TTF2, TITF2; 9q22). Therefore multiple SNPs from the 1-LOD and 2-LOD intervals around the 9q linkage peak were genotyped, as well as SNPs in each of these candidate genes.

The only SNPs genotyped in the 9q region to reach genome-wide significant association were in or near FOXE1 (see table 5), and were significantly associated with CL/P in the TOTAL and CLP datasets. FOXE1 is a forkhead domain-containing transcription factor (a.k.a. TTF2, TITF2; 9q22); mutations in FOXE1 are associated with congenital hypothyroidism, thyroid agenesis and
Cleft palate in humans (Bamford-Lazarus syndrome, MIM 241850) and mice [86–88]. The forkhead gene family (Fox), originally identified in Drosophila, encodes transcription factors with a conserved 100-amino acid DNA-binding motif called the ‘forkhead domain’ [89] and regulates diverse developmental processes in eukaryotes. Rare missense mutations in FOXE1 have been associated with isolated clefting [17, 90].

**FOXE1** is a major current focus of our research group, including studies of multi-species conservation, structure, function, and etiology. Genotyping the FOXE1 locus at a greater SNP density has excluded adjacent genes and narrowed the mutation to a 43 kb region including and upstream of FOXE1 (Moreno et al., unpublished results). Figure 4 shows the relative positions of the 9q21 SNPs that were genome-wide significant in the current study (see table 5 for list), as well as three additional fine-mapping SNPs recently found to be significantly associated with CL/P (rs894673; rs3758249; rs1867278; Moreno et al. unpublished results), along with the LD patterns estimated in the HapMap CEU population (Data Release 23a/phaseII Mar08, on NCBI B36 assembly, dbSNP b126) utilizing Haploview version 4.1 [72]. Of the recently studied SNPs, rs3758249 is the one most strongly associated with CL/P in multiple populations (Moreno et al., unpublished results) and the associated SNPs from the current study are in a haplotype block with rs3758249 (see fig. 4). Notably, in both the current study and our recent FOXE1

---

**Fig. 2.** Summed multipoint HLOD plots for each chromosome that had a maximum summed HLOD ≥3.2 (under the best genetic model for each chromosome), in the entire dataset (TOTAL), and the subsets – see definitions in figure 1. For each of those chromosomes, also shown are the graphs of the Minimum Regions of Maximal Significance (MRMS). These graphs summarize the process of repeating the Genome Scan Meta Analyses (GSMA), shifting the bins in order to narrow the region of potential involvement with CL/P. For those chromosomes with statistically significant GSMA/MRMS results (i.e. p values ≤0.05), dashed vertical lines indicate the 10 cM MRMS. f Chromosome 14 (recessive).
studies, the strongest associations are seen with families in which one or more affected family members have CL plus CP, and little or no evidence of association in families with CL alone or CP alone. Further, the involvement of FOXE1 during primary palatogenesis is supported by the previously uncharacterized epithelial expression in the medial nasal and maxillary processes that will undergo fusion (Moreno et al., unpublished results).

**Chromosome 12p11**

This region did not achieve significant linkage results in the TOTAL dataset, however, it did reach significance in the CLP subset. Chromosomal rearrangements involving the 12p region result in phenotypes including CL/P or CP [69, 70], including Pallister-Killian syndrome (mosaic isochromosome 12p) [91] and others (del 12p [92]; Fryns syndrome [93]). However this region has not been previously investigated in non-syndromic CL/P and warrants further investigation.

**Chromosome 14q**

The 14q21–24 region had genome-wide significant results in our previous report, and also reached genome-wide significant linkage results in the current study. This region contains PAX9, BMP4 and transforming growth factor beta 3 (TGFβ3). Both Pax9 and Tgfβ3 when inactivated in mice results in clefts of the secondary palate [94]. Positive results have been seen in association studies.
of CL/P and TGFB3, and one missense mutation has been reported [19, 95]. However, none of the one PAX9 or five TGFB3 SNPs tested in the current study reached genome-wide significant association in any of the datasets.

**BMP4** (14q22–23) is another plausible candidate gene in this region. Multiple murine model studies have demonstrated a role for the Bmp4 pathway in lip and palate fusion [77]. In a Bmp4 conditional knockout mouse model [96], all embryos had bilateral cleft lip at 12 days post-conception but by 14.5 days only 22% still exhibited cleft lip, suggesting in utero healing. Our group has found BMP4 nonsense and missense mutations in CL/P, microform, and subclinical cleft cases that are absent in controls [97] and a recent report showed association of a common variant missense change with isolated cleft lip and palate in a Chinese population [98]. Only one BMP4 SNP was included in the custom SNP panel (rs2147105), and was not significantly associated with CL/P in this study.

Given the strong linkage finding, and the biological plausibility of the possible candidates in this region, we are continuing to investigate BMP4 and other candidates in this region, and will incorporate GxG analyses as well (given the interacting patterns in the murine model recently reported [77]).

**Chromosome 16q**

Chromosome 16q24.1 was first identified in a genome scan of Caucasian NSCLP sib pairs [33]. Two candidate genes in this region were subsequently investigated: IRF8 and CRISPLD2 [99]. No association was seen with IRF8, but CRISPLD2 showed association in Caucasian and Hispanic families. Although the function of CRISPLD2 is

**Fig. 3.** Summaries of the weighted False Discovery Rate (wFDR) results for 1,476 SNPs selected within candidate genes or to fine-map the linkage peaks. Shown are graphs for the TOTAL dataset, and the CLP and CL+CLP subsets, i.e. those subsets in which there were genome-significant wFDR results. Shown are the $-\log_{10}$ (wFDR p values), with the SNPs arrayed across the genome from chromosome 1 (left side of X axis) to chromosome 22.
unknown, its structure featuring the presence of a LCCL (Limulus factor C, Coch-5b2 and Lgl1) domain has been suggested to play a structural or immunologic role, or even be involved in cell motility which is required for effective cell migration. In situ hybridization studies [99] showed that CRISPLD2 is expressed in the mandible, palate, and nasopharynx regions during craniofacial development [99].

**Fig. 4.** Depicted are the LD patterns estimated in the HapMap CEU population for multiple SNPs in FOXE1. The SNPs marked with asterisks (*) are those that were found to be genome-wide significantly associated with CL/P in the current study (TOTAL and CL+CLP subset). The other SNPs have recently been found by our group to be strongly associated with CL/P (Moreno et al., unpublished results).

FOXC2 is another plausible candidate in this region since FOXC2 mutations cause lymphedema-distichiasis which has incomplete penetrance of cleft palate [100–102], the knockout has cleft palate, and decreased Foxc2 expression in the lateral nasal processes occurs in mice with SHH signaling inactivated in the cranial neural crest cells [103]. We are currently testing FOXC2 and CRISPLD2 SNPs in our study populations, as these were not included in the fine-mapping SNPs genotyped in the current study.

**Summary and Future Plans**

The current study utilized linkage approaches in a large sample of multiplex CL/P families to identify six genomic regions with genome-wide significant findings, plus one additional region significant in one phenotypic subset. A follow-up fine-mapping SNP panel identified two genome-wide significant associations with SNPs in or near candidate genes (IRF6 and FOXE1), which have led to identification of the likely etiologic variants (IRF6 [71]; FOXE1 – Moreno et al., unpublished results). Note that the fine-mapping approach utilized here would only detect relatively common variants associated with CL/P. Therefore, sequencing of candidate genes is now required to find those variants that may have resulted in linkage signals but with family-specific etiologic variants that could not be detected by LD approaches such as the wFDR utilized here. We are also examining the genotyped SNPs for evidence of micродelletions that are causal and that may be flagged from apparent non-mendelian inheritance. As described above, we are also continuing to pursue those regions with significant linkage results with further fine-mapping efforts.

A striking result from these studies is that some of the significant findings could be attributed to specific phenotypic subsets. Indeed, some findings were significant only in specific phenotypic subgroups. Non-syndromic CL/P is a complex trait, with etiologic heterogeneity, so these results highlight the importance of careful phenotypic delineation in large samples of families as one method to begin to understand the observed etiologic heterogeneity in CL/P families, and as a method to implement in other human disorders with similar levels of etiologic heterogeneity.

There are a number of additional sub-clinical phenotypes now under analysis in our group that suggest increased risk for CL/P: dental anomalies [104, 105], deficiencies of the upper lip orbicularis oris muscle [106, 107], asymmetry [108, 109], and specific craniofacial measurements [110] in the non-overtly-affected relatives of individuals affected with CL/P. Such sub-clinical phenotyping holds great promise for addressing etiologic heterogeneity in CL/P families by allowing a subdivision of families into potentially more homogeneous subsets based on both the overt and sub-clinical phenotypes present, thus markedly increasing the power and specificity of our genetic studies.

Finally, identification of specific genes such as IRF6 and FOXE1 as well as the association with specific cleft phenotypes leads to a realistic expectation that improved
recurrence risk and OFC prognoses may soon be possible using a combination of molecular testing and improved phenotyping in families.

Acknowledgements

Many thanks to the families in our world-wide study sites who graciously agreed to participate, and without whom these studies would not be possible. The staff and collaborators at each study site were critical for successful completion of these studies: Dr. You-e Liu and Dr. Dan-ning Hu (China); Dr. Ajit Ray (India); Dr. Tuncbilek (Turkey); Carla Brandon, Kathy Bardi, Judith Resick, Dr. Katherine Neiswanger, Dr. Joseph Losee (USA, Pittsburgh); Katy Krahn, Amy Mach and Tamara Busch (USA, Iowa), Dora Rivera and Dr. Mauricio Camargo (Colombia). In the Philippines, Operation Smile International, Operation Smile Philippines, the HOPE Foundation, Bill and Kathy Magee, Edith Villanueva, Bue-na Nepomuceno, Henrietta Gamboa, Salie Onggada, Rachel Lim, and Gloria Melocoton were all critical to sample collection. Many thanks to Dr. Peter Chines of NHGRI for SNP mapping and annotation.

References

3 Darwin C: The Variation of Plants and Animals under Domestication. New York, Appleton, 1875.
6 Fogh-Andersen P: The Inheritance of Hare Lip and Cleft Palate; Contributions to the Elucidation of the Congenital Clefts of the Face. Copenhagen, Nyt Nordisk Forlag Arnold Busck, 1942.


55 Hodge SE, Vieland VJ, Greenberg DA: Hlods
56 Levinson DF, Levinson MD, Segurado R, Rabinowitz D, Laird N: A unified approach
57 Linkage and Fine-Mapping of CL/P Hum Hered 2009;68:151–170
64 Cooper ME, Goldstein TH, Maher BS, Marazita ML: Identifying genomic regions for fine-mapping using genome scan meta-analysis (GMSA) to identify the minimum regions of maximum significance (MRMS) across populations. BMC Genet 2005;6(suppl 1):S42.


