Published in final edited form as:

Birth Defects Res A Clin Mol Teratol. 2009 January ; 85(1): 42–51. doi:10.1002/bdra.20571.

Identification of Microdeletions in Candidate Genes for Cleft Lip and/or Palate

Min Shi^{†,1}, Adrianna Mostowska^{†,2,4}, Astanand Jugessur³, Marla K. Johnson⁴, Maria Adela Mansilla⁴, Kaare Christensen⁵, Rolv T. Lie⁶, Allen J. Wilcox⁷, and Jeffrey C. Murray⁴

¹Biostatistics Branch, National Institute of Environmental Health Sciences, National Institutes of Health, MD A3-03, P.O. Box 12233, Research Triangle Park, NC 27709

²Department of Biochemistry and Molecular Biology University of Medical Sciences in 60-781 Poznan, Poland

³Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Australia

⁴University of Iowa, Carver College of Medicine, Department of Pediatrics, S. Grand Avenue, 2182 ML, Iowa City, Iowa 52242

⁵Institute of Public Health, University of Southern Denmark, J.B. Winslows Vej 9, DK-5000 Odense C. Denmark

⁶Section for Epidemiology and Medical Statistics, Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway

⁷Epidemiology Branch, National Institute of Environmental Health Sciences MD A3-05, PO Box 12233, Research Triangle Park, NC 27709

Abstract

BACKGROUND—Genome-wide association studies are now used routinely to identify genes implicated in complex traits. The panels used for such analyses can detect single nucleotide polymorphisms and copy number variants, both of which may help to identify small deleted regions of the genome that may contribute to a particular disease

METHODS—We performed a candidate gene analysis involving 1221 SNPs in 333 candidate genes for orofacial clefting using 2823 samples from 725 two- and three-generation families with a proband with clefts of the lip and/or palate. We used SNP genotyping, DNA sequencing, high-resolution DNA microarray analysis and long-range PCR to confirm and characterize the deletion events

RESULTS—This dataset had a high duplicate reproducibility rate (99.98%), high Mendelian consistency rate (99.93%), and low missing data rate (0.55%), which provided a powerful opportunity for deletion detection. Apparent Mendelian inconsistencies between parents and child suggested deletion events in 15 individuals in 11 genomic regions. We confirmed deletions involving *CYP11B1*, *FGF10*, *SP8*, *SUMO1*, *TBX1*, *TFAP2A*, and *UGT7A1*, including both *de novo* and familial cases. Deletions of *SUMO1*, *TBX1*, and *TFAP2A* are likely to be etiologic

CONCLUSIONS—These deletions suggest the potential roles of genes or regulatory elements contained within deleted regions in the etiology of clefting. Our analysis took advantage of genotypes from a candidate-gene-based SNP survey and proved to be an efficient analytical approach to interrogate genes potentially involved in clefting. This can serve as a model to find genes playing a role in complex traits in general.

Corresponding Author: Jeffrey C. Murray, MD, University of Iowa Carver College of Medicine, Department of Pediatrics, S Grand Avenue, 2182 ML, Iowa City, Iowa 52242, Phone: 319-335-6897, FAX: 319-335-6970 jeff-murray@uiowa.edu.

[†]These authors contributed equally to this work

Presented at the 57th annual meeting of the American Society of Human Genetics, October 23-27, 2007, San Diego, CA.

Keywords

cleft lip; cleft palate; microdeletion; SNPs; CNV; candidate genes

Introduction

Clefts of the upper lip and/or palate (CL/P) are common birth defects with a complex etiology (Murray, 2002). Although a number of genes have been suggested to play a role in clefting, they account for only a small proportion of the recognized etiologies (Jugessur and Murray, 2005). Among many possible mechanisms, microdeletions have been hypothesized to play a significant role. Microdeletions may be identified by an association of clefts with other anomalies when karyotyping or comparative genomic hybridization is performed for clinical indications (Stanier and Moore, 2004). Copy number variants (CNVs) are being increasingly well defined for the human genome (Kidd et al., 2008), lending further power to their use in microdeletion detection. Certain genes may be involved in both syndromic and nonsyndromic forms of clefting. Cases where the deletion is confined to one gene only may present with a cleft only phenotype, while cases where the deletion encompasses multiple genes may show physical or developmental anomalies in addition to the cleft. With the availability of high-throughput assays that can interrogate many SNPs simultaneously, multiple candidate genes (or even genome-wide evaluations) can be investigated to search for deletions. When family-based samples are used, these deletions can be found either by dosage differences in the cases compared to controls or by evidence of apparent non-Mendelian transmissions from parents to a hemizygous child. In this study, we carried out a candidate gene analysis to search for potential microdeletions as part of a larger study looking for SNP associations with clefting. Population-based familial samples from two Scandinavian populations (Norwegian and Danish) were used in this study, and genotype data for 1221 SNPs in 333 candidate genes for orofacial clefts were available for an evaluation of microdeletions.

Samples and Methods

Case material

DNA samples from two relatively homogeneous populations from Scandinavia (Denmark and Norway) were used in this study and both syndromic and nonsyndromic cases were included. The Norwegian samples, collected from a population-based case-control study of facial clefts in Norway (1996-2001), provided a total of 478 case-parent triads and 516 population-derived healthy control infant-parent triads for analysis. Detailed descriptions of the overall study design, including patient and control characteristics, have been published previously (Jugessur et al., 2008; Nguyen et al., 2007; Wilcox et al., 2007). Of the 478 case-parent triads, 100 were cleft palate only (CPO) triads (plus 67 triads of syndromic CPO) and 264 were cleft lip with or without cleft palate (CL/P) triads (plus 47 triads of syndromic CL/P). Danish samples came from families in which a child was born with a cleft in the period 1991-2001. These cases were identified through the Danish Facial Cleft Database (Bille et al., 2005). The samples comprise three-generation families made up of the affected offspring, the parents, and the parents of the mother. A total of 68 CPO triads (plus 15 triads of syndromic CPO) and 153 CL/P triads (plus 11 triads of syndromic CL/P) were used in the analysis.

Genotypes

Genotyping was performed by the Center for Inherited Disease Research (CIDR) using Illumina's GoldenGate genotyping technology (Illumina; San Diego, CA). Genotypes were

checked for duplicate reproducibility rates, missing data rates, Hardy-Weinberg disequilibrium, and Mendelian inconsistencies. The genotype data were first cleaned by removing families showing non-paternity, evidenced by >10 Mendelian inconsistencies in the same family. Further, we removed SNPs that showed low duplicate reproducibility rates (<98%), low call rates (<95%), and ambiguous genotype clustering in the PCR amplification plots.

Detecting deletions using a candidate-gene-based genotype survey

Mendelian inconsistency patterns have been successfully applied to deletion detection in high-resolution surveys of dense SNP genotype data from case-parent triads (Conrad et al., 2006; McCarroll et al., 2006). Adopting Conrad et al.'s useful classification of transmission patterns, we categorized each transmission event into the following groups: **A** and **B**: Mendelian inconsistency compatible with potential deletion of the maternal or paternal alleles respectively, **C**: Mendelian inconsistency indicating no deletion transmission, **D**: Consistent with Mendelian inheritance with no information on potential deletions, **E** and **F**: Consistent with Mendelian inheritance and compatible with potential deletion of the maternal or paternal alleles respectively, and **G**: Consistent with Mendelian inheritance and indicating no deletion transmission. Patterns **A** and **B** are informative for potential deletion transmission. Patterns **A** and **E** are supportive of maternal transmission of a deleted allele, while patterns **B** and **F** are supportive of paternal transmission. We adapted these deletion detection rules to accommodate our relatively sparse but gene-oriented data as follows: families were labeled as harboring potential deletions when either of the following two scenarios occurred: 1) events of more than two consecutive **A**'s or **B**'s or 2) events of **A**'s or **B**'s occurring across two generations in the same pedigree (*i.e.*, from grandparents to mother and from mother to child). The above set of rules is designed to detect the transmission of a deleted allele. Therefore, one expects to see evidence of maternal/paternal transmission of a deleted allele across several consecutive SNPs, but not across intermingled SNPs that are informative for both maternal and paternal transmissions. For example, SNPs showing transmission pattern **E** (maternal evidence) may be seen in the middle of pattern **A** (maternal evidence) but not in the middle of pattern **B** (paternal evidence). For *de novo* deletion identification, patterns **E** and **F** are not informative. Instead, one looks for clusters of **A**'s or **B**'s, possibly separated by **E**'s and **F**'s. This approach, however, can only nominate potential deletions. One still needs to perform confirmatory studies, such as direct sequencing, complementary microarray analysis, or additional individual SNP genotyping to verify their authenticity.

DNA sequencing

To confirm the presence of potential deletions and replicate the GoldenGate data, SNPs showing Mendelian inconsistencies were sequenced in the affected offspring and their parents (SNPs: rs163078, rs10916, rs162556, rs4663888, rs7572563, rs7597496, rs876688, rs1448035, rs1621700, rs10807242, rs2529753, rs1978060, rs11082639 and rs4803763). Approximately 20 ng of template genomic DNA was amplified by PCR and the products were used directly in sequencing reactions. Cycle sequencing was carried out according to the manufacturer's instructions using ABI PRISM Big Dye™ Terminator Sequencing kit (Applied Biosystems; Foster City, CA) on an ABI Prism 3730 capillary sequencer. Data were collected and analyzed by the software Polyphred (version 0.970312; <http://droog.gs.washington.edu/PolyPhred.html>) and Consed (version 4.0; <http://www.genome.washington.edu/UWGC/analysistools/consed.cfm>). Information on sequencing primers is available upon request.

SNP genotyping

When available, TaqMan® SNP genotyping assays (Applied Biosystems; Foster City, CA) were also used for technical verification and deletion boundary localization. PCR reactions were set up as specified by the manufacturer and performed on an ABI Prism 7900HT instrument (Applied Biosystems; Foster City, CA). Genotypes from affected offspring and their parents were called using the accompanying Sequence Detection Systems (SDS) 2.2 software.

Long-range PCR

To determine the exact size of the deletion containing the cytochrome P450 1B1 gene (*CYP1B1*), long-range PCR was performed using different combinations of primers that were selected from regions suspected to flank the boundaries of the putative deletion. PCR using primers CTTTCTTACCACTTAGGCCTTCTGTGA and CCATGCCAACTAACTAAGACCTTCAAC was done using TaKaRa *LA Taq*TM DNA Polymerase according to the manufacturer's instructions (TAKARA BIO; Madison, WI). A 1485-bp PCR product (the deleted allele) detected in one affected individual was subsequently used in direct sequencing.

DNA microarray analysis

To determine the boundaries of putative deletions containing UDP-glycosyltransferase 1A7 (*UGT1A7*), T-box 1 (*TBX1*) and transcription factor AP2 alpha (*TFAP2A*), genome-wide copy-number scan was performed on DNA samples of the four affected individuals using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, CA, USA). The results were analyzed using the Partek Genomics Suite software (<http://www.partek.com/software>). Nucleotide numbering in this paper is according to UCSC Build 36.1 human reference sequence (of March 2006).

Results

Candidate genes and SNPs selection

Candidate genes for orofacial clefts were selected 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. A detailed description of the selection criteria and a list of studied genes/SNPs in the overall study are provided in Supplement table 1.

Initial deletion scan based on SNP survey of candidate genes

The high quality of this dataset, as reflected by the high duplicate reproducibility rate (99.98%) and low missing data rate (0.55%), is essential to the success of deletion detection using Mendelian inconsistency patterns. After data-cleaning, we had genotypes for 1221 SNPs on autosomal chromosomes in 1241 families (725 case families and 516 control families) (Table 1).

We identified 11 potential deletions (Supplement table 2) by scanning genotypes for signatures of segregating deletion events based on Mendelian inconsistency patterns. Five of these, labeled by consecutive **A** or **B** inheritance patterns, are in the regions around the following genes: *CYP1B1* on chromosome 2, the small ubiquitin-like modifier 1 gene (*SUMO1*) on chromosome 2, *UGT1A7* on chromosome 2, *TFAP2A* on chromosome 6, and *TBX1* on chromosome 22. The remaining six deletion events, labeled by potential trans-generational transmission of the deleted allele (where transmission patterns for both

grandmother-to-mother and mother-to-child are **A**'s or **B**'s), are the transforming growth factor beta receptor type 2 (*TGFBR2*) gene on chromosome 3, fibroblast growth factor 10 (*FGF10*) on chromosome 5, transcription factor Sp8 (*SP8*) on chromosome 7, mothers against decapentaplegic, drosophila, homolog 2 (*SMAD2*) on chromosome 8, poliovirus receptor-like 2 (*PVRL2*) on chromosome 19, and *TBX1* on chromosome 22 (Supplement table 2).

Confirmation of the deletion candidates

Danish families—Confirmatory analyses in the Danish families validated four of the eight identified potential deletions. The first is a 140 kb-wide heterozygous deletion on chromosome 2 that was identified in a three-generation family with CLP in the child. This pedigree shows trans-generational transmission of the deleted allele. To confirm and further characterize this deletion, we genotyped 40 SNPs within a 0.53 Mb region encompassing the deletion using DNA from the offspring and the parents to narrow down the break points. The proximal breakpoint of the deletion was narrowed down to a 914-bp region between a newly identified marker located at nucleotide position 38127498 bp and rs11124631, and the distal breakpoint was localized to a 15-kb region between rs13385748 and rs232587 (Table 2). With this knowledge in hand, we performed long-range PCR to further demarcate the breakpoints of this deletion. Sequencing of the PCR product revealed the exact breakpoints of the deletion to be chr2: 38,128,460 - 38,268,588 bp (Figure 1A). We designed a set of PCR primers located at the boundaries of the deletion region to amplify a 518-bp PCR product only in those samples carrying the deleted allele (the normal allele is too large to be amplified). We confirmed deletions in the affected offspring, his mother and grandmother using this assay (Figure 1B). This deletion encompasses the entire length of *CYP11B1*, the hypothetical gene *MGC34824*, and the last exon of the gene encoding the hypothetical protein FAM82A.

The second deletion confirmed on chromosome 2 was present in a child affected with CPO. We genotyped 27 SNPs within a 478-kb region in this child and his parents in order to narrow down the proximal breakpoint of the deletion to a 16.7-kb region between markers rs6717044 and rs4675272, and the distal breakpoint to a 79.5-kb region between markers rs35662035 and rs6435143 (Table 3). The maximum estimated size of the deletion, based on informative markers (markers identified in the proband to be heterozygous), is 124.2 kb (chr 2:202,778,312-202,902,501) and contains the genes *SUMO1* and *NAP5/NOP58*. The minimum estimated deletion size based on SNPs showing Mendelian inconsistencies is 28 kb (chr 2: 202,794,974 - 202,823,039) and contains a portion of the intron and last exon of *SUMO1* (Figure 2). This Mendelian inconsistency pattern of consecutive B's confirms the presence of a deletion in the child and suggests that this deletion may have been transmitted from the father. However, the deletion is also compatible with the occurrence of a *de novo* event in the affected child.

The third deletion was identified 12.4 kb upstream of the *SP8* gene (chr7: 20,805,488-20,805,494). This novel 7-bp nucleotide change (delAGGTGGG) disrupts the GoldenGate assay, with the deleted allele failing to be amplified upon PCR (Figure 2A). This deletion is present in an affected child with CL/P, her mother and grandmother, as confirmed by sequencing (Figure 2B).

Lastly, we confirmed the deletion in the second intron of *FGF10* (chr 5:44,347,532-44,347,538). This 7-bp deletion (del AAAACTA) disrupts the GoldenGate assay and the corresponding sequences are shown in Figure 3A. It is present in a child affected with CPO, her mother and grandmother (Figure 3B).

Norwegian families—Analyses performed on the Norwegian families confirmed three of the four potential deletions. We first confirmed the presence of Mendelian inconsistencies by DNA sequencing and then performed a genome-wide deletion scan analysis on the four probands of the investigated families (Table 4) to further characterize the deletions. Approximately 1.8 million genetic markers (900,000 SNPs and 900,000 CNVs) were genotyped in each person using Affymetrix Genome-Wide Human SNP Array 6.0. The average SNP call rate was 99.1%. Genome-wide copy number scans revealed the exact size of the novel deletions. One deletion, identified in a child affected with cleft lip (CL) and congenital heart defects, is located on chr6: 9,085,657 – 12,517,402. This large 3.4 Mb deletion encompasses several genes, one of which is *TFAP2A*—a gene previously implicated in clefting. The second deletion (chr2: 234,213,750 – 234,304,972) was identified in two families with CLP and CL, respectively. The 91-kb deletion found in the affected children contains the first shared exon of *UGT1A9*, *UGT1A7*, *UGT1A6*, *UGT1A5*, *UGT1A4* and *UGT1A3*, as well as a fragment of the first intron of *UGT1A8* and *UGT1A10*. Due to the unavailability of DNA samples, we could not perform microarray scans on the parents to verify these two deletions. Although the inheritance patterns (pattern A) are consistent with maternal transmission of the deleted allele, they could also represent *de novo* deletions.

We also confirmed the occurrence of a deletion in chromosome 22 (Chr 22:17,022,145-19,872,615). This deletion, identified in a patient with the 22q11-syndrome form of cleft palate, spans *TBX1* among several other genes. Analysis of the inheritance pattern for this family showed that SNPs consistent with maternal transmission of the deleted allele (E) resided among SNPs compatible with paternal transmission of the deleted allele (B) (Supplement table 2). This suggests that the deletion is a *de novo* event in the affected child. DNA was not available to confirm the deletion in the fourth Norwegian family.

Conclusion and Discussion

This study investigated microdeletions in cases of orofacial clefts using data from a large-scale genetic association analysis, in which 1221 SNPs in 333 cleft candidate genes were genotyped in 725 families from two Scandinavian populations of shared ancestry. The initial scan identified 11 regions containing putative genomic deletions, based on patterns of Mendelian inconsistency among the families. Upon further investigation using dense SNP genotyping, long-range PCR, direct sequencing and DNA microarray analysis, we confirmed and characterized seven of the deletions. Five of these were found in a total of 585 nonsyndromic cleft cases and an additional two were found in 140 syndromic cases. Mendelian inconsistencies will identify, at best, only 25% of deletions that include any specific SNP. Our study did not include comprehensive coverage of entire genes with tagging SNPs for most of the 333 genes studied, and we required evidence for two consecutive SNPs being deleted before further investigation. Thus, these results showing a deletion rate of 7/725 (1%) should be considered a conservative estimate. Higher resolution panels with the ability to detect smaller deletions using CNV approaches may identify a higher rate of deletion, especially in genome-wide association panels where deletion detection does not depend on *a priori* selection of specific candidate genes.

Deletions spanning 7 bp to 3.4 Mb were confirmed in seven regions. Five of the deletions include at least some portion of one or more adjacent genes, increasing the likelihood of a wider range of phenotypes. Indeed, two of the five cases were found to be syndromic forms of clefting. Of the seven confirmed deletion regions, six are compatible with transmission of the deleted allele and one appeared to be *de novo*. Two of the deletions are small (7 bp) and likely to represent rare normal variants. The presence of the deletion in one parent was

confirmed for three deleted regions (regions containing *CYP1B1*, *FGF10* and *SP8*) and unknown for three other deleted regions (regions containing *SUMO1*, *TFAP2A*, and *UGT1A7*). Even though some deletions may turn out to be normal variants, such as the 7-bp deletions, there is evidence to support the potential roles played by the larger deletions. The three deletions lacking informative data on the parents could be *de novo*. Additionally, phenotype data on the other family members may not be as detailed as those on the cases, and consequently, minor phenotypes might have been missed. Etiological involvement of a microdeletion may also require interactions between the deleted alleles and other variants in the genome or exposure variables that are only possessed or experienced by the affected child. Two of these deletions are small and may represent rare normal variants (*SP8* and *FGF10*). Two other provide the first evidence to date of a potential association between *CYP1B1* and *UGT7A1* with clefting, whereas *SUMO1*, *TBX1*, and *TFAP2A*, provide confirmatory evidence.

CYP1B1 and *UGT1A7* are involved in phase I and II detoxification reactions. *CYP1B1* is a member of the cytochrome P450-related enzymes which are important in metabolizing a variety of endogenous and xenobiotic compounds. *UGT1A7*, a member of the UDP-glycosyltransferase family, catalyzes conjugation reactions in which hydrophobic chemicals are converted into water-soluble compounds. Among the growing number of environmental exposures reported to increase the risk of orofacial clefting, cigarette smoking has been the most consistent (Lie et al., 2008; Shi et al., 2007). A reduced capacity to biotransform toxins in tobacco smoke, as a result of microdeletions in these detoxification genes, may represent a plausible mechanism for the adverse effects of smoking on pregnancy outcomes. On the other hand, the deletions observed in *CYP1B1* and *UGT1A7* might also represent rare and innocuous variants with no connection to clefting.

TFAP2A is highly expressed in craniofacial structures and knockout mice show multiple facial anomalies (Schorle et al., 1996). In addition, mutations in *TFAP2A* have recently been reported to underlie the Branchio-Oculo-Facial Syndrome, which includes clefts as part of the phenotype (Milunsky et al., 2008).

FGF10 is a member of the fibroblast growth factor signaling family. This pathway plays an important role in craniofacial development and mutations in several members of this pathway, including *FGF10*, have been associated with orofacial clefts (Riley et al., 2007). Similarly, several members of the T-box gene family have been linked to human clefting disorders, including *TBX22* in cleft palate with ankyloglossia (Packham and Brook, 2003) and *TBX1* in DiGeorge syndrome (Chieffo et al., 1997). The patient with 22q deletion in our sample was previously reported to be a 22q11 deletion patient (Sivertsen et al., 2007). This was independently confirmed in our current analysis, demonstrating the reliability of using multi-SNP genotype data in the fine-mapping of deletion events.

An important role for *SUMO1* in CL/P has recently been reported (Alkuraya et al., 2006). This gene was interrupted by a balanced reciprocal translocation in a patient born with unilateral cleft lip and palate and who was otherwise phenotypically normal. Alkuraya and colleagues used whole-mount *in situ* hybridization experiments in mice to support a causative role for *SUMO1* haploinsufficiency in the pathogenesis of CL/P (Alkuraya et al., 2006). The identification of a second case with clefts harboring a disruption of the *SUMO1* gene provides additional evidence for its role in human clefting. Interestingly, a recent review hypothesized possible synergistic interactions between the FGF signaling pathway, SUMO modification, and environmental risk factors in the causation of CL/P (Pauws and Stanier, 2007). In this study, we independently identified deletions in genes involved in all of these interacting genes. The supportive data for genes such as *TFAP2A* and *SUMO1* promote further searches for a role for regulatory or sequence variants in these genes in the

etiology of common, isolated forms of clefting. For all of the deletions observed, it is also possible that genes or regulatory elements flanking the deleted region may be the etiologic sites, and identification of these will require additional genetic and functional studies.

In conclusion, this study explored the possible involvement of deletions in the etiology of orofacial clefts and identified several deletions by scanning for patterns of Mendelian inconsistencies across multiple SNPs in a large candidate-gene study of clefting. This study serves as a proof-of-principle that genotype data from whole-genome surveys can be used for deletion detection, supplementing the association data generated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to especially thank the many families and children who participated in this study. This research was supported by NIH grants DE08559, P60 DE13076, NIH P30 ES05605, the Norwegian Research Council and the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. This work was also made possible by a donation from the Iowa Order of the Eastern Star. We also thank Susie McConnell for secretarial support, Dorthe Grosen for help with the Danish samples, Dr. Clarice Weinberg for helpful discussions, Drs. Andrew Lidral and Bridget Riley for SNP selection, and Sandra Daack-Hirsch, Cathy Dragan and Kathy Frees for preparing and shipping samples. Genotyping services were provided by the Center for Inherited Disease Research (CIDR) and we greatly appreciate the work of David Valle, Corinne Boehm, Kim Doheny, Ivy McMullen and the other staff in this project. CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, Contract Number N01-HG-65403.

Grant Information and Grant Numbers: NIH grants DE08559, P60 DE13076, NIH P30 ES05605, Norwegian Research Council (NFR) and Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (NIEHS), and a donation from the Iowa Order of the Eastern Star. Genotyping services were provided by the Center for Inherited Disease Research (CIDR), funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, Contract Number N01-HG-65403. The US National Institute of Dental and Craniofacial Research (NIDCR) underwrote a significant proportion of the genotyping costs by CIDR.

References

- Alkuraya FS, Saadi I, Lund JJ, et al. SUMO1 haploinsufficiency leads to cleft lip and palate. *Science*. 2006; 313:1751. [PubMed: 16990542]
- Bell SM, Schreiner CM, Waclaw RR, et al. Sp8 is crucial for limb outgrowth and neuropore closure. *Proc Natl Acad Sci U S A*. 2003; 100:12195–12200. [PubMed: 14526104]
- Bille C, Knudsen LB, Christensen K. Changing lifestyles and oral clefts occurrence in Denmark. *Cleft Palate Craniofac J*. 2005; 42:255–259. [PubMed: 15865458]
- Bush JO, Lan Y, Jiang R. The cleft lip and palate defects in Dancer mutant mice result from gain of function of the Tbx10 gene. *Proc Natl Acad Sci U S A*. 2004; 101:7022–7027. [PubMed: 15118109]
- Chieffo C, Garvey N, Gong W, et al. Isolation and characterization of a gene from the DiGeorge chromosomal region homologous to the mouse Tbx1 gene. *Genomics*. 1997; 43:267–277. [PubMed: 9268629]
- Conrad DF, Andrews TD, Carter NP, et al. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet*. 2006; 38:75–81. [PubMed: 16327808]
- Jugessur A, Murray JC. Orofacial clefting: recent insights into a complex trait. *Curr Opin Genet Dev*. 2005; 15:270–278. [PubMed: 15917202]
- Jugessur A, Rahimov F, Lie RT, et al. Genetic variants in IRF6 and the risk of facial clefts: single-marker and haplotype-based analyses in a population-based case-control study of facial clefts in Norway. *Genet Epidemiol*. 2008
- Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature*. 2008; 453:56–64. [PubMed: 18451855]

- Lie RT, Wilcox AJ, Taylor J, et al. Maternal Smoking and Oral Clefts: The Role of Detoxification Pathway Genes. *Epidemiology*. 2008
- McCarroll SA, Hadnott TN, Perry GH, et al. Common deletion polymorphisms in the human genome. *Nat Genet*. 2006; 38:86–92. [PubMed: 16468122]
- Milunsky JM, Maher TA, Zhao G, et al. TFAP2A mutations result in branchiooculo-facial syndrome. *Am J Hum Genet*. 2008; 82:1171–1177. [PubMed: 18423521]
- Murray JC. Gene/environment causes of cleft lip and/or palate. *Clin Genet*. 2002; 61:248–256. [PubMed: 12030886]
- Nguyen RH, Wilcox AJ, Moen BE, et al. Parent's occupation and isolated orofacial clefts in Norway: a population-based case-control study. *Ann Epidemiol*. 2007; 17:763–771. [PubMed: 17664071]
- Packham EA, Brook JD. T-box genes in human disorders. *Hum Mol Genet*. 2003; 12(Spec No 1):R37–44. [PubMed: 12668595]
- Pauws E, Stanier P. FGF signalling and SUMO modification: new players in the aetiology of cleft lip and/or palate. *Trends Genet*. 2007; 23:631–640. [PubMed: 17981355]
- Riley BM, Mansilla MA, Ma J, et al. Impaired FGF signaling contributes to cleft lip and palate. *Proc Natl Acad Sci U S A*. 2007; 104:4512–4517. [PubMed: 17360555]
- Schorle H, Meier P, Buchert M, et al. Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature*. 1996; 381:235–238. [PubMed: 8622765]
- Schultz RE, Cooper ME, Daack-Hirsch S, et al. Targeted scan of fifteen regions for nonsyndromic cleft lip and palate in Filipino families. *Am J Med Genet*. 2004; 125A:17–22. [PubMed: 14755461]
- Shi M, Christensen K, Weinberg CR, et al. Orofacial cleft risk is increased with maternal smoking and specific detoxification-gene variants. *Am J Hum Genet*. 2007; 80:76–90. [PubMed: 17160896]
- Sivertsen A, Lie RT, Wilcox AJ, et al. Prevalence of duplications and deletions of the 22q11 DiGeorge syndrome region in a population-based sample of infants with cleft palate. *Am J Med Genet A*. 2007; 143:129–134. [PubMed: 17163526]
- Stanier P, Moore GE. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Hum Mol Genet*. 2004; 13(Spec No 1):R73–81. [PubMed: 14722155]
- Treichel D, Schock F, Jackle H, et al. mBtd is required to maintain signaling during murine limb development. *Genes Dev*. 2003; 17:2630–2635. [PubMed: 14597661]
- Wilcox AJ, Lie RT, Solvoll K, et al. Folic acid supplements and risk of facial clefts: national population based case-control study. *Bmj*. 2007; 334:464. [PubMed: 17259187]

Figure 1.

Confirmation of *CYP11B1* gene deletion

1A. Long-range PCR and sequencing analysis. Long-range PCR was done using primers CYP11B1F: CTTTCTTACCACTTAGGCCTTCTGTGA and CYP11B1R: CCATGCCAACTAACTAAGACCTTCAAC. The amplified product of the mutated allele is 1485 bp long. The normal allele is not amplifiable due to its length (140,645 bp). Sequencing of the 1485 bp product identified the exact breakpoints of the deleted allele as chr2:del38,128,460 - 38,268,588. M: molecular size marker; 1: proband with CLP.

1B. Segregation of the mutation. To confirm the presence of deletion in the proband's family, PCR was done on all available family members using primer pair CYP11B1F1: AAAGAAGCTGGATGCAAAACA and CYP11B1R2: GACACTCTGTAGTAGGTAGCCATGA. The amplified product of the deleted allele was 518 bp. The analysis revealed that the 140,127-bp deletion was present in the affected proband (individual 1), his mother (individual 3), and grandmother (individual 5).

A. chr7: del20,805,488-20,805,494

Figure 2.

Detection of deletion chr7:del20,805,488 - 20,805,494

2A. **Sequencing analysis.** Chromatograms showing the heterozygous 7-bp deletion (AGGTGGG) in the affected child and her mother. Arrows indicate the location of rs2529753. The deleted nucleotides are underlined.

2B. **Segregation of the mutation.** The deletion is present in the proband, her mother and grandmother.

Figure 3.

Detection of deletion chr5: del44,347,532 - 44,347,538

3A. **Sequencing analysis.** Chromatograms showing the heterozygous 7-bp deletion (AAAAACT) present in the affected child and her mother. Arrows indicate the location of rs1448035. The deleted nucleotides are underlined.

3B. **Segregation of the mutation.** The deletion is present in the proband, her mother and grandmother.

Table 1

Study participants by cleft type and population

Population	Phenotype	Number of triads		Comments
		nonsyndromic	syndromic	
Danish ^e	CPO ^a	68	15	223 three-generation families (mother, father, child, and mother's parents)
	CL ^b	64	5	
	CLP ^c	89	6	
Norwegian	CPO	100	67	426 complete triads and 52 triads with a missing parent
	CL	112	15	
	CLP	152	32	
	Control ^d	516		326 dyads and 190 triads

^aCPO= Cleft Palate Only^bCL= Cleft Lip only^cCLP= Cleft Lip and cleft Palate^dControl fathers were first asked to provide a swab sample only after November 1998, two years after the onset of the study. Hence, a substantial proportion of the control families initially recruited consist of only mother-child dyads.^eSubphenotype information for one Danish case was not available.

Table 2

Genotype analysis of SNPs around potential *CYP1B1* deletion

rs number ^a	SNP location ^b	SNP location	Genotype ^c			
			CH ^d	F ^d	M ^d	TP ^e
rs7579777	38,092,831	intron 10 of <i>FAM82A</i>	TT	CT	CT	D
rs6544134	38,092,860	intron 10 of <i>FAM82A</i>	AA	AG	AG	D
rs100461	38,119,482	intron 10 of <i>FAM82A</i>	CT		CC	G
rs2116232	38,122,222	intron 10 of <i>FAM82A</i>	GT	TT	GG	G
rs2121260	38,122,267	intron 10 of <i>FAM82A</i>	AC	CC	AA	G
NEW	38,125,384	intron 10 of <i>FAM82A</i>	AC	AA	AC	G
rs4670232	38,125,537	intron 10 of <i>FAM82A</i>	CG	GG	CC	G
rs4670234	38,125,981	intron 10 of <i>FAM82A</i>	CT	TT	CC	G
rs4670235	38,126,032	intron 10 of <i>FAM82A</i>	CT	TT	CC	G
rs10192788	38,126,176	intron 10 of <i>FAM82A</i>	CT	TT	CC	G
NEW	38,126,296	intron 10 of <i>FAM82A</i>	HET	N	N	G
rs2082255	38,127,262	intron 10 of <i>FAM82A</i>	GT	TT	GG	G
rs2082254	38,127,398	intron 10 of <i>FAM82A</i>	AC	AA	CC	G
NEW	38,127,498	intron 10 of <i>FAM82A</i>	CT	CC	CT	G
rs11124631	38,128,412	intron 10 of <i>FAM82A</i>	TT	TT	CC	A
rs985257	38,136,732	intron 10 of <i>FAM82A</i>	AA	AA	AA	D
rs163078	38,140,449	intron 10 of <i>FAM82A</i>	GG	GG	GG	D
rs10916	38,150,674	3'UTR of <i>CYP1B1</i>	TT	TT	GG	A
rs4646430	38,159,919	upstream of <i>CYP1B1</i>	GG	GG	CC	A
rs162556	38,159,958	upstream of <i>CYP1B1</i>	CC	CC	TT	A
rs232557	38,180,411	upstream of <i>CYP1B1</i>	AA	AA	GG	A
rs232545	38,198,965	upstream of <i>CYP1B1</i>	CC		TT	A
rs2432661	38,206,143	upstream of <i>CYP1B1</i>	AA		GG	A
rs232600	38,230,073	intron 2 of <i>MGC34824</i>	CC	CT	CC	D
rs232572	38,243,815	intron 2 of <i>MGC34824</i>	AA	AA	AA	D

rs number ^a	SNP location ^b	SNP location	Genotype ^c			
			CH ^d	F ^d	M ^d	TP ^e
rs432252	38,258,273	intron 2 of <i>MGC34824</i>	CC	CC	CC	D
rs2194381	38,258,299	intron 2 of <i>MGC34824</i>	GG	GG	AA	A
rs13385748	38,259,373	intron 2 of <i>MGC34824</i>	AA	AA	GG	A
rs68352	38,261,470	exon3 of <i>MGC34824</i>	CC	CC	CC	D
rs2447739	38,263,774	downstream of <i>MGC34824</i>	GG	GG	GG	D
rs170813	38,269,593	downstream of <i>MGC34824</i>	GG	GG	GG	D
rs232588	38,274,177	downstream of <i>MGC34824</i>	GG	GG	GG	D
rs232587	38,274,357	downstream of <i>MGC34824</i>	TT	TT	AT	D
rs4670244	38,303,513	downstream of <i>MGC34824</i>	GG	GG	GG	D
rs7573277	38,395,628	intron 5 of <i>ARL6IP2</i>	AG	AA	AG	G
rs12712583	38,396,077	intron 4 of <i>ARL6IP2</i>	CT	CC	CT	G
rs11892771	38,524,944	upstream of <i>ARL6IP2</i>	CC	CT	CC	D
rs6761215	38,591,755	upstream of <i>ARL6IP2</i>	CT	CT	CC	G
rs11124640	38,591,845	upstream of <i>ARL6IP2</i>	CG	CC	CG	G
rs6761542	38,591,857	upstream of <i>ARL6IP2</i>	AG	AG	AA	G

^aShaded SNPs indicate the regions of breakpoints. The minimum size of the deletion (between rs11124631 and rs13385748) is 130,961 bp. The maximum size of the deletion (between NEW at bp 38,128,412 and rs7573277) is 268,130 bp.

^bSNP location according to UCSC Genome Browser, human sequence assembly of March 2006.

^cAccording to dbSNP build 128 (forward strand). Blank cells represent missing genotypes.

^dCH: child, F: father, M: mother.

^eTransmission pattern: A, Mendelian inconsistency in agreement with deletion of the maternal allele; D, Mendelian inheritance with no information on potential deletions; G, Mendelian inheritance indicating no deletion involvement. The deletion is present in affected child and mother.

Table 3

Genotype analysis of SNPs around potential *SUMO1* deletion

rs number ^a	SNP location ^b	SNP location	Genotype ^c				TP ^e
			CH ^d	F ^d	M ^d		
rs6747814	202,688,963	upstream of <i>LOC339809</i>	AG	GG	AA		G
rs12988414	202,770,323	3-UTR of <i>LOC339809</i>	CC		CC		D
rs2350358	202,772,739	downstream of <i>SUMO1</i>	GG	GG	GG		D
rs6717044	202,778,312	downstream of <i>SUMO1</i>	TT	TC			G
rs3769817	202,787,552	intron 2 of <i>SUMO1</i> ^f	GG	GG	GG		D
rs7580433	202,791,910	intron 2 of <i>SUMO1</i> ^f	GG	GG	GG		D
rs4675272	202,794,974	intron1 of <i>SUMO1</i>	GG	CC	GG		B
rs6742786	202,795,116	intron1 of <i>SUMO1</i>	GG	AA	GG		B
rs2350360	202,795,948	intron1 of <i>SUMO1</i>	TT	TT	TT		D
rs7599810	202,797,513	intron1 of <i>SUMO1</i>	AA	GG	AA		B
rs6755690	202,798,838	intron1 of <i>SUMO1</i>	GG	CC	GG		B
rs3754931	202,811,646	upstream of <i>SUMO1</i>	CC	TT	CC		B
rs4675274	202,816,113	upstream of <i>SUMO1</i>	TT	TT	TT		D
rs35662035	202,823,039	upstream of <i>SUMO1</i>	CC	TT	CC		B
rs10166594	202,823,230	upstream of <i>SUMO1</i>	TT	TT	TT		D
rs6435139	202,828,113	upstream of <i>SUMO1</i>	GG	GG	GG		D
rs6758247	202,834,698	upstream of <i>SUMO1</i>	CC	CC	CC		D
rs7581542	202,840,182	intron 14 of <i>NOP5/NOP58</i>	AA	AA	AA		D
rs13392645	202,843,084	intron 14 of <i>NOP5/NOP58</i>	AA	AA	AA		D
rs7572505	202,860,600	intron 10 of <i>NOP5/NOP58</i>	GG	GG	GG		D
rs3731700	202,864,181	exon 8 of <i>NOP5/NOP58</i>	AA	AA	AA		D
rs6753156	202,873,779	intron 2 of <i>NOP5/NOP58</i>	CC	CC	CC		D
rs10084427	202,879,917	downstream of <i>NOP5/NOP58</i>	GG	GG	GG		D
rs6435143	202,902,501	downstream of <i>NOP5/NOP58</i>	AC		CC		G

			Genotype ^c			
rs number ^a	SNP location ^b	SNP location	CH ^d	F ^d	M ^d	TP ^e
rs6435146	202,920,474	downstream of <i>NOP5/NOP58</i>	AA	AA	AA	D
rs6747299	202,993,402	intron 12 of <i>BMPR2</i>	TT	GT	TT	G

^aShaded SNPs indicate the regions of breakpoints. The minimum size of the deletion (between rs4675272 and rs35662035) is 28,065 bp. The maximum size of the deletion (between rs6717044 and rs6435143) is 124,189 bp.

^bSNP location according to UCSC Genome Browser, human sequence assembly of March 2006

^cAccording to dbSNP build 128 (forward strand). Blank cells represent missing genotypes.

^dCH: child, F: father, M: mother.

^eTransmission pattern: **B**, Mendelian inconsistency in agreement with deletion of the paternal allele; **D**, Mendelian inheritance with no information on potential deletions; **G**, Mendelian inheritance indicating no deletion involvement. The deletion is present in the affected child and probably in his father.

^f*SUMO1* isoform 1 = NM_001005782

Summary of identified deletions

Table 4

Family	Individuals with deletion ^a	Phenotype of child ^b	Chromosomal region ^c	Known genes involved	Regions	CNV database
A Danish three-generation family	GM-M-CH	CLP	Chr 2: 38,128,460 - 38,268,588	<i>CYP11B1</i> , <i>MGC34824</i> , and <i>FAM82A</i>	whole gene deleted	not reported
A Danish three-generation family	CH ^d	CPO	Chr 2:202,778,312/202,794,974 -202,823,039/202,902,501	<i>SUMO1</i>	whole gene or the last exon deleted	a CNV was reported upstream of this region. CNV database Variation_6020
A Danish three-generation family	GM-M-CH	CLP	Chr 7:20,805,488-20,805,494 (delAGGTGGG)	<i>SP8</i>	1244bp upstream of the gene	not reported
A Danish three-generation family	GM-M-CH	CPO	Chr 5:44,347,532-44,347,538 (delAAACTA)	<i>FGF10</i>	intron 2	not reported
A Norwegian triad	CH ^d	CL syndromic	Chr 6: 9,085,657-12,517,402	<i>TFAP2A</i> and several other genes	multiple genes deleted	several CNV reported in the region but none covering <i>AP2</i> gene
Two Norwegian triads	CH ^d	CPO CL	Chr 2:234,213,750-234,304,972	<i>UGT1A10A9</i> , 8,7,6,5,4,3)	one exon deleted	not reported
A Norwegian triad	CH <i>de novo</i>	CPO syndromic	Chr 22:17,022,145-19,872,615	<i>TBX1</i> and many other genes	multiple genes deleted	a CNV reported. CNV database Variation_22959

^aM= mother; CH= child; GM= mother's mother

^bCPO= Cleft Palate Only; CL= Cleft Lip only; CLP= Cleft Lip and cleft Palate

^cNucleotide positions are based on UCSC Genome Browser (March 2006 assembly)

^dDeletion status of the parents is unknown