1 *CRISPLD2* **VARIANTS INCLUDING A C471T SILENT MUTATION MAY**

2 **CONTRIBUTE TO NONSYNDROMIC CLEFT LIP WITH OR WITHOUT CLEFT**

3 **PALATE.**

4

5 Ariadne Letra, D.D.S., Ph.D., Research Assistant Professor[1,2], Renato Menezes, D.D.S., Ph.D.,

6 Research Assistant Professor[11,2], Margaret E. Cooper, M.S., M.S.I.S., Research Instructor[1,2],

7 Renata F. Fonseca, M.S., Graduate Student[3], Stephen Tropp, B.S., Graduate Student[1], Manika

8 Govil, Ph.D., Research Assistant Professor[1,2], Jose M. Granjeiro, D.D.S., Ph.D., Associate

9 Professor[4], Sandra R. Imoehl, B.S., Dental Student[5], M. Adela Mansilla, Ph.D., Research

10 Assistant[6], Jeffrey C. Murray, M.D., Professor[6], Eduardo E. Castilla, M.D., Ph.D., Directorr[7],

11 Iêda M. Orioli, M.D., Ph.D., Professor[3], Andrew E. Czeizel, M.D., Ph.D., Director[8], Lian Ma,

12 D.D.S., Professor[9], Brett T. Chiquet, B.A., Graduate Student and Dental Student[10], Jacqueline T.

13 Hecht, M.S., Ph.D., Professor[10] , Alexandre R. Vieira, D.D.S., Ph.D., Assistant Professor[1,2,11,12],

14 Mary L. Marazita, Ph.D., Professor [1,2,12,13]

15 [1] Department of Oral Biology, and [2]Center for Craniofacial and Dental Genetics, School of

16 Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA

17 [3] Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

18 [4] Department of Cell and Molecular Biology, Universidade Federal Fluminense, Niterói, RJ,

19 Brazil

20 [5] College of Dentistry, University of Iowa, Iowa City, IA, USA

21 [6] Department of Pediatrics, University of Iowa, Iowa City, IA, USA

22 [7] Department of Genetics, FIOCRUZ, Rio de Janeiro, RJ, Brazil

23 [8] Foundation for the Community Control of Hereditary Diseases, Budapest, Hungary

24

25 [9] School of Stomatology, Beijing University, Beijing, China

26 [10] Department of Pediatrics and Pediatric Research Center, University of Texas Medical School,,

27 Houston, TX, USA

28 [11] Department of Pediatric Dentistry School of Dental Medicine, University of Pittsburgh,

29 Pittsburgh, PA, USA

30 [12] Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh,

31 Pittsburgh, PA, USA

32 [13] Department of Psychiatry, and Clinical and Translational Sciences Institute, School of

33 Medicine, University of Pittsburgh, Pittsburgh, PA, USA

34

35 Corresponding author: Mary L. Marazita, Ph.D.

36 Center for Craniofacial and Dental Genetics

37 University of Pittsburgh

38 Suite 500, Bridgeside Point Building

39 100 Technology Dr.

40 Pittsburgh, PA 15219

41 Phone: 412-648-8380, FAX: 412-648-8779

42 e-mail: marazita@pitt.edu

43

44 **Running Title:** *CRISPLD2* and CL(P)

45 **Grant Support:** This work was supported by grants from the National Institutes of Health

46 (NIH): K99-DE018954 (A.L.), K99-DE018913 (R.M.), R21-DE016718 (A.R.V.), R01-

47 DE016148 (M.L.M, A.R.V), P50-DE016215 (M.L.M., J.C.M.), R21-DE016930 (M.L.M.), R01-

48 DE09886 (M.L.M., L.M.), R01-DE012472 (M.L.M.), R01-DE016148-04A1 (J.T.H.), K99-

49 DE018085 (M.G.). Additional support provided by FAPERJ E-26/152.831/2006, CNPq

50 308885/2006-0, 401467/2004-0 (I.M.O.); and CAPES, Brazil (R.F.F.). The content is solely the

51 responsibility of the authors and does not necessarily represent the official views of the National

52 Institute of Dental and Craniofacial Research or the National Institutes of Health.

53

54 **ABSTRACT**

55 **Objective:** To assess association between nonsyndromic (NS) cleft lip with or without cleft
56 palate (CL(P)) and SNPs within the *CRISPLD2* gene (cysteine-rich secretory protein LCCL
57 domain containing 2).

58 **Design:** Four SNPs within the *CRISPLD2* gene domain (rs1546124, rs8061351, rs2326398,
59 rs4783099) were genotyped to test for association via family-based association methods.

60 **Participants:** 5,826 individuals from 1,331 families in which one or more family member is
61 affected with CL(P).

62 **Results:** Evidence of association was seen for SNP rs1546124 in USA (p=0.02) and Brazilian
63 (p=0.04) Caucasian cohorts. We also found association of SNP rs1546124 with cleft palate alone
64 (CP) in South Americans (Guatemala and ECLAMC) and combined Hispanics (Guatemala,
65 ECLAMC and Texas Hispanics) (p=0.03 for both comparisons), and with both cleft lip with cleft
66 palate (CLP; p=0.04) and CL(P) (p=0.02) in North Americans. Strong evidence of association
67 was found for SNP rs2326398 with CP in Asian populations (p=0.003) and with CL(P) in
68 Hispanics (p=0.03), and also with bilateral CL(P) in the Brazilians (p=0.004). In the Brazilians,
69 SNP rs8061351 showed association with cleft subgroups incomplete CL(P) (p=0.004), and
70 unilateral incomplete CL(P) (p=0.003). Prediction of SNP functionality revealed that the C allele
71 in the C471T silent mutation (overrepresented in cases with CL(P) presents two putative exonic
72 splicing enhancer motifs and creates a binding site AP-2 alpha, a transcription factor involved in
73 craniofacial development.

74 **Conclusions:** Our results support the hypothesis that variants in the *CRISPLD2* gene may be
75 involved in the etiology of NS CL(P).

76 **Key words**: *CRISPLD2* gene, cleft lip, cleft palate, subphenotypes

77

78 **INTRODUCTION**

79 Craniofacial anomalies, and in particular oral-facial clefts including cleft lip (CL) and
80 cleft palate (CP), are major human structural birth defects with a worldwide frequency of 1 in
81 700 live births and substantial clinical impact. The possible etiologies are many, including
82 single-gene disorders, chromosome aberrations, exposure to teratogens, and sporadic conditions
83 of unknown cause (Murray, 2002). Oral-facial clefts can be further classified as nonsyndromic
84 (NS, i.e. isolated) or syndromic based on the presence of other structural anomalies.
85 Approximately 30% of all clefts are associated with one of more than 400 described syndromes
86 (Gorlin et al., 2001) while the remaining 70% are isolated defects. It is generally accepted that
87 CL with or without CP (CL(P)) and cleft palate alone (CP) are developmentally distinct
88 phenotypes. CL(P) is more common, affecting 1–2/1000 births and presenting considerable
89 differences in prevalence, with Native Americans and Asians showing the highest rate and
90 Africans the lowest. CP is less common, with a prevalence of approximately 1/1500–2000 births
91 in Caucasians, less variable among different ethnic backgrounds (Forrester and Merz, 2004).

92 The nature of the genetic contribution to the etiology of NS CL(P) and CP is still a
93 subject for discussion and investigation. Analyses of familial recurrence risk patterns in CL(P)
94 have estimated that 3–14 genes interacting multiplicatively may be involved, indicating that
95 CL(P) is a heterogeneous disorder (Schliekelman and Slatkin, 2002). Nevertheless, despite the
96 evidence for a genetic role in the etiology of CL(P) and CP, environmental factors such as
97 smoking and maternal nutrition are also thought to influence this structural birth defect, possibly
98 in an interactive manner (Maestri et al., 1997; Wyszynski et al., 1997). Therefore, etiological
99 heterogeneity has probably been a major confounding factor for identifying clefting
100 susceptibility loci.

101 Recent advances in high-throughput genotyping technologies and powerful statistical
102 approaches have accelerated the discovery of loci conferring susceptibility for complex diseases
103 through the use of genome scans (Altmuller et al., 2001).The first CL(P) scan was conducted
104 with 92 British sib pairs and identified nine regions with suggestive results, including a region on
105 chromosome 16q (Prescott et al., 2000; Prescott et al., 2001). Other genomic scans followed,
106 and the region of 16q21-24 reached genome-wide statistical significance for linkage with CL(P)
107 in multiple studies (Field et al., 2004; Marazita et al., 2004a; Marazita et al., 2004b; Marazita
108 et al., 2009).

109 *CRISPLD2* (cysteine-rich secretory protein containing LCCL domain 2) gene is located
110 on chromosome 16q24.1 and has been recently associated with nonsyndromic CL(P) in U.S.
111 Caucasian and Hispanic populations (Chiquet et al., 2007). Moreover, the authors detected
112 *CRISPLD2* expression in the mandible, palate and nasopharynx regions during craniofacial
113 development at E13.5-E17.5, and have suggested *CRISPLD2* as a novel candidate gene for the
114 etiology of NS CL(P). Although the function of *CRISPLD2* remains to be elucidated, its structure
115 featuring the presence of a LCCL (Limulus factor C, Coch-5b2 and Lgl1) domain has been
116 suggested to play a structural or immunologic role, or even be involved in cell motility
117 (Liepinsh et al., 2001; Nagai et al., 2007). Interestingly, cell motility is required for effective
118 cell migration which, together with apoptosis, accounts for the cellular mechanism responsible
119 for the disappearance of medial edge epithelia cells prior to palatal fusion (Chai and Maxson,
120 2006).

121 Given the observed population differences in CL(P) and CP birth prevalences and other
122 characteristics, it is of interest to expand studies of orofacial cleft etiology to diverse populations.
123 In this study, we performed association studies with *CRISPLD2* and NS CL(P) and NS CP in a

124 large cohort sample consisting of distinct populations from North America, South America, Asia,
125 Northern and Eastern Europe.
126
127
128 **SUBJECTS AND METHODS**
129 **Subjects** The study population consisted of a total of 5,826 genotyped individuals from distinct
130 population sets. Included were 1,023 multiplex families from the Philippines, Guatemala,
131 Europe (Spain, Hungary, Turkey) and the USA (Iowa, Pennsylvania, Missouri, and Texas). In
132 addition, there were 308 nuclear trios, 84 from China plus 224 from Argentina, Brazil and Chile
133 ascertained through ECLAMC (Latin American Collaborative Study of Congenital
134 Malformations). There were also 610 unrelated individuals (328 NS CL(P) cases and 282
135 controls with no known family history of clefting) of Caucasian ethnicity from Brazil. Table 1
136 summarizes all of the populations, families and individuals. With the exception of ECLAMC
137 which is a hospital-based birth defects registry study, all families were ascertained by
138 recruitment, either from community-based ascertainment (the Philippines) or through clinical-
139 facility-based ascertainment (remainder of sites).
140 All cases had NS CL, CP or CL with CP (CLP). CL and CLP combined comprise CL(P).
141 Informed consent was obtained from each study subject after approval by the appropriate
142 institutional review boards in the USA (University of Pittsburgh, Pittsburgh, PA; University of
143 Iowa, IA; Washington University, St. Louis, MO; University of Texas, Houston, TX) and each
144 participating international site (details provided on request).
145 Blood or saliva samples were collected to obtain genomic DNA. Procedures for DNA
146 extraction were performed according to standard protocols.

147 **Genotyping**

148 Four single nucleotide polymorphisms (SNPs) within the *CRISPLD2* gene domain

149 previously shown to be associated with CL(P) (Chiquet et al., 2007) were genotyped using

150 Taqman chemistry on an automatic sequence-detection instrument (ABI Prism 7900HT, Applied

151 Biosystems, Foster City, CA). Assays and reagents were supplied by Applied Biosystems

152 (Applied Biosystems, Foster City, CA). Details of the studied polymorphisms are presented in

153 Table 2. Further, Table 3 presents the allelic and genotypic frequencies by population sub-

154 groups, plus the results of tests of Hardy Weinberg Equilibrium (HWE).

155 **Statistical Analysis**

156 Families were analyzed separately according to population and then in subgroups. We

157 combined all the North American Caucasian samples from Pittsburgh, Saint Louis, Iowa and

158 Texas to form a North American group. The European Caucasians (Turkey, Spain and Hungary)

159 were added to the North American group to form an overall Caucasian group. The data from

160 Guatemala and ECLAMC formed the South American group. Note that the Guatemalans and the

161 registry participants from ECLAMC are of European (Spain and Portugal) extractions as well as

162 a mixture of Amer-Indians and Africans. A Hispanic group was formed with the self-reported

163 Hispanics from Texas and the South American group, to address any inquiries as to ethnic

164 environmental considerations.

165 For some analyses the families were divided into the following non-overlapping subsets:

166 CL=families in which all affecteds had CL only; CLP=all affecteds had CL plus CP; CLCLP=at

167 least one affected with CL plus at least one with CLP; CP=at least one affected with CP alone.

168 Another grouping was designated CL(P) and consisted of the combination of the CL, CLP and

169 CLCLP subsets.

170 The properties of the SNPs were assessed in the proband trios using Haploview (Barrett
171 et al., 2005), ie., to assess LD between the SNPs, to test for Hardy-Weinberg Equilibrium
172 (HWE), and to estimate the minor allele frequencies (MAF).
173 The association between clefting and SNPs in the *CRISPLD2* gene was deternined by
174 Transmission Disequilibrium Tests (TDT) in the family data, as implemented in FBAT (Family
175 Based Association Test, version 1.7.3) (Laird et al., 2000; Rabinowitz and Laird, 2000;
176 Horvath et al., 2001). Allelic, genotypic, and haplotype TDT analyses were performed with the
177 empirical analysis option in order to adjust for families with multiple parent-child trios. Parent
178 of origin effects were assessed in the proband trios using PLINK (Purcell et al., 2007). For the
179 Brazilian case-control sample, allelic/genotypic associations were assessed with chi-square tests
180 as implemented in SAS (version 9.1.3).
181 Nominal p-values are reported. With a Bonferroni correction for multiple testing
182 (considering the number of variables and tests performed), p-values below 0.0006 (0.05/80)
183 would be considered statistically significant.

184 **RESULTS**

185 **Preliminary analyses** Pairwise linkage disequilibrium (D' values) between the four
186 SNPs were calculated for each population using the GOLD (Graphical Overview of Linkage
187 Disequilibrium) program (Abecasis and Cookson, 2000) and ranged from 0.02 to 0.23; thus it is
188 likely that very little redundant information was obtained from the data (data not shown). Table
189 3 presents the allelic and genotypic frequencies, and the p-values from the tests of HWE by each
190 subpopulation. There was no evidence of deviation from Hardy-Weinberg equilibrium for any of
191 the SNP/population group combinations, except for SNP rs4783099 in the North American
192 Caucasian and ALL Caucasian groups.

Association analyses The results of case-control and family-based allelic association analyses stratified by population are summarized in Table 4. None of these results reached formal Bonferroni-adjusted significance (i.e. p-value < 0.0006), however results in 2 SNPs (rs1546124 and rs8061351) were nominally significant ($p<0.05$) for the Brazilian case-control sample (Caucasian), the USA Caucasians, and suggestive ($0.5 \leq p<0.10$) in Turkey; in addition to the original results from Texas (Chiquet et al., 2007). Interestingly, the results from Asia and the Latin American admixed populations (Guatemala and ECLAMC) were not significant for any SNP.

Table 5 shows the results from TDT allelic association analyses by cleft family subgroups. There are some suggestions in the results that the CLP subgroup (i.e. families in which all affecteds have CL plus CP) have the greatest statistical significance in the Caucasian populations for SNP rs1546124. Of interest, in the Asian subgroup there was a suggestive result in the CP family subgroup ($p =0.003$) however the sample size was very small (15 informative families out of 37 total CP families). The haplotype TDT association results are not presented in detail because in no case did the significance levels improve over the individual SNP results. Similarly, there was no significant evidence of parent-of-origin effects for any SNP, therefore those results are not presented in detail.

Table 6 presents the Brazilian case-control results for both allelic and genotypic association for all cases and for subgroups of the cases based on cleft laterality and completeness (Letra et al., 2007). Again, no formal significance was seen, but results for three of the SNPs (rs1546124, rs8061351 and rs2326398) were nominally significant or suggestive. The most significant results were with SNP rs8061351 for Incomplete clefts (allelic p-value=0.009,

215 genotypic 0.004), and for Unilateral Incomplete clefts (genotypic p-value=0.003); and with SNP
216 rs2326398 for Bilateral (allelic p-value=0.004) and Bilateral Complete (allelic p-value=0.007).
217 **Prediction of SNP Functionality** In order to verify if the synonymous mutation C471T
218 (rs8061351) in exon 4 of the *CRISPLD2* gene associated with incomplete cleft lip/palate in the
219 Brazilian individuals and in the Hispanic families from Texas (Table 4) could disrupt DNA-
220 binding sites and further affect *CRISPLD2* protein expression, two transcription binding site
221 prediction methods, FASTSNP and AliBaba 2.1, were used (Matys et al., 2006; Yuan et al.,
222 2006). FASTSNP identified that the C allele, overrepresented in cases with cleft lip/palate,
223 affects splicing regulation by altering exonic splicing enhancer motifs; FASTSNP attributes a
224 low-medium risk for this allele. Further, AliBaba 2.1 identified a Sp1 binding site with either C
225 or T alleles, however the C allele also harbored a binding site for transcription factor AP-2 alpha.
226
227 **DISCUSSION**
228 The current study sought to replicate the recent study (Chiquet et al., 2007) that
229 identified the *CRISPLD2* (cysteine-rich secretory protein containing LCCL domain 2) gene as a
230 novel candidate gene for NS CL(P). *CRISPLD2* is located on chromosome 16q24.1, spans
231 approximately 8.95kb, and contains 14 exons coding for a 497 amino acid polypeptide. The
232 exact function of this gene is not yet known, nevertheless i*n situ* hybridization of mouse tissues
233 showed *CRISPLD2* expression in the naso- and oropharynx at E13.5, the mandible at E14.5, and
234 the palate and cartilage primordia of the nasal septum at E17.5. (Chiquet et al., 2007).
235 We investigated the SNPs associated in the original study (Chiquet et al., 2007) in a
236 large NS CL(P) and CP cohort consisting of twelve distinct populations (see Table 1). Although
237 not reaching formal Bonferroni-adjusted significance, several results were suggestive for CL(P)

238 in the Caucasian populations but interestingly not in the Asian nor admixed populations, while
239 the Asian populations had suggestive results for CP. Of note, statistical significance was
240 increased when cases were stratified based on cleft phenotypes.

241 One polymorphism (rs1546124), located at the 5' UTR of the *CRISPLD2* gene, showed
242 significant altered transmission in the Caucasian cohort in the original study (Chiquet et al.,
243 2007) and was the only SNP significantly associated with CL(P) in that study. In the current
244 study, a synonymous mutation in exon 4 (C471T, rs8061351) showed association with the case-
245 control cohort from Brazil, and was also shown to be associated with the Hispanic cohort in the
246 original study (Chiquet et al., 2007). Furthermore, our results also suggest that SNP rs1546124
247 may have a stronger effect on individuals of Caucasian ethnicity whereas SNP rs8061351 seems
248 to have a stronger effect on individuals of Hispanic or South American origin. A possible
249 explanation for this discrepancy is that distinct populations may have different risk alleles.

250 Recently a study of 31 SNPs in 12 candidate genes including *CRISPLD2* was conducted
251 in an Irish study, comprising 509 CL(P) and 383 CP case-trios, and 926 population-based-
252 controls, and including three of the four SNPs analyzed in the current study (Carter et al., 2010).
253 This study population is most comparable to the ALL Caucasian subgroup in the current study.
254 Notably, for the three SNPs in common between the studies (rs1546124, rs8061351, rs4783099)
255 the minor alleles (G, C, T respectively) and MAF values in the current study (see Table 3) were
256 comparable to those presented in Table 2 of the Irish study (Carter et al., 2010). However, the
257 Irish study found no statistically significant association with any of the three SNPs in common,
258 nor with three additional *CRISPLD2* SNPs assessed only in the Irish study. One of the SNPs in
259 common, rs1546124 showed suggestive association in the current study Caucasians. These
260 differing results could be in part because the study designs differed. Our study population was

261 enriched in multiplex families since most were ascertained as part of a linkage study (Marazita
262 et al., 2009), whereas 7.5% of the Irish study cases had a positive family history of orofacial
263 clefts. Notably, the region of 16q21-24 reached genome-wide statistical significance for linkage
264 with CL(P) in multiple studies (Field et al., 2004; Marazita et al., 2004a; Marazita et al.,
265 2004b; Marazita et al., 2009), consistent with an hypothesis that the *CRISPLD2* relationship
266 with CL(P) may be most important in familial cases.

267 Of note, several chromosomal abnormalities involving chromosome 16 have been
268 described that include clefting as part of the clinical phenotype. Duplications of 16p12-13
269 (Brewer et al., 1999), as well as trisomy and translocations of chromosome 16 (Ducos et al.,
270 2004) have been associated with cleft palate. In addition, one marker at approximately 7 kb from
271 the *MMP25* gene on chromosome 16p13.3 has shown association with NS CL(P) (Blanton et al.,
272 2004). More specifically, the region between 16q21-24 has been identified in genomic scans as
273 harboring susceptibility genes for CL(P) (Prescott et al., 2000; Field et al., 2004; Marazita et
274 al., 2004a; Marazita et al., 2004b; Marazita et al., 2009). Further, other investigators have also
275 demonstrated the presence of CP with mutations in genes located in or around 16q21-24. Hecht
276 et al (Hecht et al., 1991) observed a correlation between cleft palate and variations at 16q22.1,
277 while mutations in the *FOXC2* gene located at 16q24.3 have been shown to cause distichiasis,
278 lymphedema, and CP (Bahuau et al., 2002).

279 Although the function of *CRISPLD2* is unknown, effects of mutations in its structure may
280 be predicted. It has been proposed that SNP rs1546124, a C/G polymorphism located at the 5'
281 UTR, may have a functional effect on *CRISPLD2* transcription by the presence of a Sp1 binding
282 site whenever a G allele is present (Nagai et al., 2007). We further verified that SNP rs8061351,
283 which denotes a C471T silent mutation with both nucleotides coding for a proline, also implies

284 differences in functionality depending on the allele present. We observed that the C allele, which
285 is overrepresented in the cases with clefts of the lip and of the primary palate, alters splicing
286 regulation by diminishing exonic splicing enhancer motifs. Moreover, the C allele also creates a
287 binding site for transcription factor AP-2 alpha, involved in craniofacial development and
288 recently suggested as an important regulatory element for *IRF6* expression and in turn, for the
289 occurrence of CL(P). Nevertheless, although synonymous changes should not affect gene
290 expression nor the final protein product, they may lead to the synthesis of a protein product with
291 the same amino acid sequence but different structural and functional properties and thus should
292 not be neglected in determining the risk of development of various diseases (Komar, 2007b;
293 Komar, 2007a).

294 Whether variants in the *CRISPLD2* gene are causal agents for CL(P) is yet to be
295 confirmed. So far, the only biological evidence for a possible role in palate development is the
296 expression of *CRISPLD2* at E17.5 in mice, a time point that does not seem to be critical since the
297 process of palatogenesis is known to be complete by E15 (Dudas et al., 2007). Nonetheless, the
298 associations observed in this and other genetic studies (Blanton et al., 2004; Chiquet et al.,
299 2007; Marazita et al., 2009), and the observations on SNP functionality warrant further
300 investigations to clarify the role for *CRISPLD2* in NS CL(P) and CP suggested by our results..

301 **CONCLUSIONS**

302 This study further demonstrates that variants in the *CRISPLD2* gene may be involved in
303 the pathogenic mechanism of NS CL(P), utilizing a much larger sample size and additional
304 ethnicities beyond the original publication (Chiquet et al., 2007). Nevertheless, the exact
305 biological functions and the contribution of *CRISPLD2* to the clefting phenotype are still to be
306 clarified.

307 **ACKNOWLEDGMENTS**

308 We gratefully acknowledge individuals and families for their valuable collaboration.
309 Thanks to research coordinators and staff at each collection site. Our research in Guatemala was
310 made possible by the support of Children of the Americas. Part of this paper is based on a thesis
311 submitted to the graduate faculty, Federal University of Rio de Janeiro, in partial fulfillment of
312 the requirements for the PhD degree (R.F.F.).

313

314 .

References

Abecasis GR, Cookson WO. GOLD--graphical overview of linkage disequilibrium. 2000. Bioinformatics 16(2):182-183.

Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. 2001. Am J Hum Genet 69(5):936-950.

Bahuau M, Houdayer C, Tredano M, Soupre V, Couderc R, Vazquez MP. FOXC2 truncating mutation in distichiasis, lymphedema, and cleft palate. 2002. Clin Genet 62(6):470-473.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. 2005. Bioinformatics 21(2):263-265.

Blanton SH, Bertin T, Serna ME, Stal S, Mulliken JB, Hecht JT. Association of chromosomal regions 3p21.2, 10p13, and 16p13.3 with nonsyndromic cleft lip and palate. 2004. Am J Med Genet A 125(1):23-27.

Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D. A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality--and tolerance of segmental aneuploidy--in humans. 1999. Am J Hum Genet 64(6):1702-1708.

338

339 Carter TC, Molloy AM, Pangilinan F, Troendle JF, Kirke PN, Conley MR, Orr DJ, Earley M,
340 McKiernan E, Lynn EC, Doyle A, Scott JM, Brody LC, Mills JL. Testing reported
341 associations of genetic risk factors for oral clefts in a large Irish study population. 2010.
342 Birth Defects Res A Clin Mol Teratol 88(2):84-93.

343

344 Chai Y, Maxson RE, Jr. Recent advances in craniofacial morphogenesis. 2006. Dev Dyn
345 235(9):2353-2375.

346

347 Chiquet BT, Lidral AC, Stal S, Mulliken JB, Moreno LM, Arcos-Burgos M, Valencia-Ramirez
348 C, Blanton SH, Hecht JT. CRISPLD2: a novel NSCLP candidate gene. 2007. Hum Mol
349 Genet 16(18):2241-2248.

350

351 Ducos A, Pinton A, Berland HM, Seguela A, Brun-Baronnat C, Bonnet N, Darre R, Milan D.
352 Cleft palate associated with an unbalanced karyotype in piglets sired by a heterozygous
353 carrier boar with a balanced constitutional reciprocal translocation. 2004. Vet Rec
354 154(21):659-661.

355

356 Dudas M, Li WY, Kim J, Yang A, Kaartinen V. Palatal fusion - where do the midline cells go? A
357 review on cleft palate, a major human birth defect. 2007. Acta Histochem 109(1):1-14.

358

359 Field LL, Ray AK, Cooper ME, Goldstein T, Shaw DF, Marazita ML. Genome scan for loci
360 involved in nonsyndromic cleft lip with or without cleft palate in families from West
361 Bengal, India. 2004. Am J Med Genet A 130(3):265-271.
362
363 Forrester MB, Merz RD. Descriptive epidemiology of oral clefts in a multiethnic population,
364 Hawaii, 1986-2000. 2004. Cleft Palate Craniofac J 41(6):622-628.
365
366 Gorlin RJ, Cohen MM, Hennekam RCM. *Syndromes of the Head and Neck*. New York, NY:
367 Oxford University Press; 2001.
368
369 Hecht JT, Wang YP, Blanton SH, Michels VV, Daiger SP. Cleft lip and palate: no evidence of
370 linkage to transforming growth factor alpha. 1991. Am J Hum Genet 49(3):682-686.
371
372 Horvath S, Xu X, Laird NM. The family based association test method: strategies for studying
373 general genotype--phenotype associations. 2001. Eur J Hum Genet 9(4):301-306.
374
375 Komar AA. Genetics. SNPs, silent but not invisible. 2007a. Science 315(5811):466-467.
376
377 Komar AA. Silent SNPs: impact on gene function and phenotype. 2007b. Pharmacogenomics
378 8(8):1075-1080.
379
380 Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of
381 association. 2000. Genet Epidemiol 19 Suppl 1:S36-42.

382

383 Letra A, Menezes R, Granjeiro JM, Vieira AR. Defining subphenotypes for oral clefts based on
384 dental development. 2007. J Dent Res 86(10):986-991.
385
386 Liepinsh E, Trexler M, Kaikkonen A, Weigelt J, Banyai L, Patthy L, Otting G. NMR structure of
387 the LCCL domain and implications for DFNA9 deafness disorder. 2001. EMBO J
388 20(19):5347-5353.
389
390 Maestri NE, Beaty TH, Hetmanski J, Smith EA, McIntosh I, Wyszynski DF, Liang KY, Duffy
391 DL, VanderKolk C. Application of transmission disequilibrium tests to nonsyndromic
392 oral clefts: including candidate genes and environmental exposures in the models. 1997.
393 Am J Med Genet 73(3):337-344.
394
395 Marazita ML, Field LL, Tuncbilek G, Cooper ME, Goldstein T, Gursu KG. Genome-scan for
396 loci involved in cleft lip with or without cleft palate in consanguineous families from
397 Turkey. 2004a. Am J Med Genet A 126(2):111-122.
398
399 Marazita ML, Lidral AC, Murray JC, Field LL, Maher BS, Goldstein McHenry T, Cooper ME,
400 Govil M, Daack-Hirsch S, Riley B, Jugessur A, Felix T, Moreno L, Mansilla MA, Vieira
401 AR, Doheny K, Pugh E, Valencia-Ramirez C, Arcos-Burgos M. Genome Scan, Fine-
402 Mapping, and Candidate Gene Analysis of Non-Syndromic Cleft Lip with or without
403 Cleft Palate Reveals Phenotype-Specific Differences in Linkage and Association Results.
404 2009. Hum Hered 68(3):151-170.

405

406 Marazita ML, Murray JC, Lidral AC, Arcos-Burgos M, Cooper ME, Goldstein T, Maher BS,
407 Daack-Hirsch S, Schultz R, Mansilla MA, Field LL, Liu YE, Prescott N, Malcolm S,
408 Winter R, Ray A, Moreno L, Valencia C, Neiswanger K, Wyszynski DF, Bailey-Wilson
409 JE, Albacha-Hejazi H, Beaty TH, McIntosh I, Hetmanski JB, Tuncbilek G, Edwards M,
410 Harkin L, Scott R, Roddick LG. Meta-analysis of 13 genome scans reveals multiple cleft
411 lip/palate genes with novel loci on 9q21 and 2q32-35. 2004b. Am J Hum Genet
412 75(2):161-173.
413
414 Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev
415 D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE,
416 Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene
417 regulation in eukaryotes. 2006. Nucleic Acids Res 34(Database issue):D108-110.
418
419 Murray JC. Gene/environment causes of cleft lip and/or palate. 2002. Clin Genet 61(4):248-256.
420
421 Nagai H, Sugito N, Matsubara H, Tatematsu Y, Hida T, Sekido Y, Nagino M, Nimura Y,
422 Takahashi T, Osada H. CLCP1 interacts with semaphorin 4B and regulates motility of
423 lung cancer cells. 2007. Oncogene 26(27):4025-4031.
424
425 Prescott NJ, Lees MM, Winter RM, Malcolm S. Identification of susceptibility loci for
426 nonsyndromic cleft lip with or without cleft palate in a two stage genome scan of affected
427 sib-pairs. 2000. Hum Genet 106(3):345-350.

428

429 Prescott NJ, Winter RM, Malcolm S. Nonsyndromic cleft lip and palate: complex genetics and
430 environmental effects. 2001. Ann Hum Genet 65(Pt 6):505-515.
431
432 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de
433 Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and
434 population-based linkage analyses. 2007. Am J Hum Genet 81(3):559-575.
435
436 Rabinowitz D, Laird N. A unified approach to adjusting association tests for population
437 admixture with arbitrary pedigree structure and arbitrary missing marker information.
438 2000. Hum Hered 50(4):211-223.
439
440 Schliekelman P, Slatkin M. Multiplex relative risk and estimation of the number of loci
441 underlying an inherited disease. 2002. Am J Hum Genet 71(6):1369-1385.
442
443 Wyszynski DF, Duffy DL, Beaty TH. Maternal cigarette smoking and oral clefts: a meta-
444 analysis. 1997. Cleft Palate Craniofac J 34(3):206-210.
445
446 Yuan HY, Chiou JJ, Tseng WH, Liu CH, Liu CK, Lin YJ, Wang HH, Yao A, Chen YT, Hsu CN.
447 FASTSNP: an always up-to-date and extendable service for SNP function analysis and
448 prioritization. 2006. Nucleic Acids Res 34(Web Server issue):W635-641.
449

450

Table 1: Summary of participants by population and phenotypic subgroups.

		FAMILIES						**NUMBER OF INDIVIDUALS GENOTYPED**		
				Number of families in phenotype subgroups [a]						
Population Group	**Country or State (USA)**	**Number**	**Mean family size ±s.d.**	**CL**	**CLP**	**CLCLP**	**CP**	**Affected**	**Unaffected**	**Unknown**
Asian	China	84	3.0 ± 0.0	13	47	n/a	24	84	168	0
	Philippines	204	13.4 ± 8.3	15	72	104	13	564	1127	13
Central Am.	Guatemala	95	6.11 ± 3.9	15	70	7	3	80	304	0
South Am.	ECLAMC	224	3.0 ± 0.0	43	132	n/a	49	209	423	0
	Brazil [b]	n/a	n/a	n/a	n/a	n/a	n/a	328	282	0
Europe	Spain	39	3.8 ± 0.94	5	24	4	6	49	97	0
	Hungary	31	5.0 ± 0.98	6	8	10	7	52	71	0
	Turkey	40	9.7 ± 2.7	7	15	7	11	39	66	0
USA	IA	140	3.0 ± 0.0	39	56	3	42	124	222	33
	PA + MO	130	5.3 ± 2.6	14	45	42	29	137	304	0
	TX- Cauc [c]	262	4.8 ± 3.8	45	191	24	2	335	506	1
	TX- Hispanic [c]	82	5.12 ± 3.7	8	71	2	1	87	131	0
TOTAL		**1,331**	**5.8 ± 5.3**	**211**	**731**	**203**	**187**	**2,088**	**3,701**	**47**

a Phenotype subgroups: CL=those pedigrees in which all affected members have CL only; CLP = affected members all have CL with CP; CLCLP= at least one affected member with CL and one with CLP; CP=at least one affected member with CP. NOTE: the analysis group CL(P)= CL + CLP + CLCLP.
b Brazilian study subjects were unrelated cases (CL=4; CLP=324), and controls (i.e. not families).
c Families originally reported (Chiquet et al., 2007)

Table 2. Summary of the SNPs studied in the *CRISPLD2* gene.

SNP Marker	**Base Position**[a]	**Region**	**Base Change**[b]	**Average Heterozygozity**[a]	**Type of Assay**
rs1546124	83,429,542	5' UTR	C/G	0.473 +/- 0.113	Taqman OD[c]
rs8061351	83,440,593	Exon 4	C/T[d]	0.493 +/- 0.059	Taqman OD
rs2326398	83,460,208	Intron 8	A/G	0.499 +/- 0.022	Taqman OD
rs4783099	83,473,468	3' UTR	C/T	0.471 +/- 0.117	Taqman OD

[a] According to the USCS Genome Browser Human 2004 May Assembly.
[b] Ancestral allele listed first.
[c] Assay-on-demand.
[d] Silent mutation.

Table 3: *CRISPLD2* SNPs: HardyWeinberg Equilibrium (HWE), allele and genotype frequencies as calculated from the proband trios.

	SNP MINOR ALLELE		SNP GENOTYPE FREQUENCY			
Population[a] / SNP	**Minor Allele**	**MinorAllele Frequency (MAF)**	**Minor Allele Homozygote**	**Heterozygote**	**Common Allele Homozygote**	**HWE p-value**
South/Central American						
rs1546124	G	29%	10%	37%	53%	0.15
rs8061351	ng	ng	ng	ng	ng	ng
rs2326398	G	22%	5%	36%	59%	0.29
rs4783099	ng	ng	ng	ng	ng	ng
Hispanic						
rs1546124	G	30%	11%	38%	51%	0.10
rs8061351	C	33%	11%	43%	46%	0.84
rs2326398	G	24%	6%	36%	58%	1.0
rs4783099	T	43%	16%	54%	30%	0.38
ALL Caucasian						
rs1546124	G	32%	11%	41%	48%	0.12
rs8061351	C	28%	8%	39%	53%	0.62
rs2326398	G	33%	10%	45%	45%	0.63
rs4783099	T	36%	11%	50%	39%	0.0006
North American Caucasian						
rs1546124	G	32%	11%	41%	48%	0.14
rs8061351	C	28%	8%	39%	53%	0.49
rs2326398	G	32%	11%	43%	46%	0.72
rs4783099	T	37%	11%	50%	39%	0.006
Asian						
rs1546124	C	47%	21%	51%	28%	0.70
rs8061351	C	38%	16%	44%	40%	0.25
rs2326398	A	26%	7%	38%	55%	0.81
rs4783099	C	42%	18%	50%	33%	0.75

[a] Population subgroups (see Table 1): **South/Central American**=Guatemala and ECLAMC; **Hispanic**=Guatamala, ECLAMC, Texas-Hispanic; **ALL Caucasian**=Iowa, Texas-Caucasian, Pittsburgh and St Louis, Madrid, Hungary, Turkey; **North American Caucasian**=Iowa, Texas-Caucasian, Pittsburgh and St Louis;**Asian**= Philippines, China
[b]ng = not genotyped

Table 4: Association of *CRISPLD2* and NS CL(P): Results of chi-square analysis on total Brazilian case-control sample and TDT analysis in families from multiple populations

	Case-Control (χ^2 p-value)	**p-values from family-based TDT for CL(P)[a]**									
		UNITED STATES									
	BRAZIL	**Texas [b]**		**Total USA[c]**	**ASIA**		**EUROPE**			**CENTRAL AMERICA**	**SOUTH AMERICA**
Ethnicity	**Cauc.**	**Cauc**	**Hisp**	**Cauc**	**Phil**	**China**	**Spain**	**Hungary**	**Turkey**	**Guatemala**	**ECLAMC[d]**
Number of Families	**n/a**	**262**	**82**	**532**	**204**	**84**	**39**	**31**	**40**	**95**	**224**
SNP:											
rs1546124	***0.04***	***0.01***	0.53	***0.02***	0.37	0.27	0.87	n/a	**0.09**	0.98	0.30
rs8061351	***0.04***	0.76	***0.02***	0.85	0.40	0.59	0.27	0.85	0.76	0.43	---
rs2326398	0.11	0.19	**0.06**	0.20	0.24	0.16	0.83	n/a	0.10	0.19	0.43
rs4783099	1.00	0.17	0.22	0.54	0.80	0.90	0.41	0.68	0.13	0.89	---

n/a = not applicable because not enough informative families; Cauc=Caucasian ethinicity, Hisp=Hispanic, Phil=the Philippines
a p-values from FBAT analysis of families, values in ***bold italics*** reach nominal significance (ie p-value<0.05), values in **bold** are suggestive (0.05<p-value<0.10).
b Results from original *CRISPLD2* report (Chiquet et al., 2007)
c "Total USA Cauc" includes Pittsburgh, St. Louis, Iowa and Texas-Caucasian
[d] ECLAMC families were not genotyped for rs8061351 and rs4783099

Table 5. Results from TDT allelic association analyses within cleft family subgroups and population subgroups for SNPs rs1546124 and rs2326398

		CLEFT FAMILY SUBGROUP [a]									
		CL		**CLP**		**CLCLP**		**CL(P)**		**CP**	
POPULATION [b]	**SNP**	**p-value (allele)** [c]	**N** [d]	**p-value (allele)**	**N**	**p-value (allele)**	**N**	**p-value (allele)**	**N**	**p-value (allele)**	**N**
South/Central America	rs1546124	0.56 (G)	32	0.23 (C)	71	NA	4	0.43 (C)	107	***0.03*** (C)	23
	rs2326398	**0.07 (A)**	17	0.66 (A)	61	NA	1	0.17 (A)	79	0.67 (A)	16
Hispanic	rs1546124	0.56 (G)	32	0.13 (C)	85	NA	4	0.26 (C)	121	***0.03*** (C)	23
	rs2326398	**0.09** (A)	19	0.21 (A)	79	NA	2	***0.03*** (A)	100	0.67 (A)	16
North American Caucasian	rs1546124	0.88 (C)	33	***0.04*** (C)	70	0.13 (C)	32	***0.02*** (C)	135	0.59 (C)	24
	rs2326398	0.89 (G)	43	0.12 (A)	96	0.75 (A)	31	0.20 (A)	170	0.81 (G)	23
ALL Caucasian	rs1546124	0.67 (G)	39	0.11 (C)	95	0.36 (C)	42	0.13 (C)	176	0.38 (C)	26
	rs2326398	0.54 (G)	54	**0.10** (A)	122	0.65 (A)	40	0.22 (A)	216	0.74 (A)	34
Asian	rs1546124	0.16 (G)	18	0.35 (C)	61	0.78 (C)	60	0.72 (C)	139	0.81 (G)	20
	rs2326398	0.83 (G)	16	1.00 (G)	64	0.49 (G)	60	0.57 (G)	140	***0.003*** (G)	15

[a] Cleft Family subgroups : CL=those pedigrees in which all affected members have CL only; CLP = affected members all have CL with CP; CLCLP= at least one affected member with CL and one with CLP; CL(P)= cleft with or without cleft palate familes; i.e. a combination of the CL, CLP, CLCLP subgroups; CP=at least one affected member with CP.

[b] Population subgroups: **South/Central America**=Guatemala and ECLAMC; **Hispanic**=Guatamala, ECLAMC, Texas-Hispanic; **North American Caucasian**=Iowa, Texas-Caucasian, Pittsburgh and St Louis; **ALL Caucasian**=Iowa, Texas-Caucasian, Pittsburgh and St Louis, Madrid, Hungary, Turkey; **Asian**= Philippines, China

[c] p-value (allele) = TDT p-value and associated allele. P-values in ***bold italics*** reach nominal significance (ie p-value<0.05), values in **bold** are suggestive (0.05<p-value<0.10).

[d] n: the number of informative families in the TDT (see Table 1 for the number of families in each cleft family subgroup)

Table 6: Brazilian case-control results for allelic and genotypic association with CL(P) subphenotypes (Letra et al., 2007).

			P-VALUES FROM CHI-SQUARE ANALYSES OF EACH *CRISPLD2* SNP [a]							
			rs1546124		**rs8061351**		**rs2326398**		**rs4783099**	
Phenotype	**cases**	**controls**	**Allelic**	**Genotypic**	**Allelic**	**Genotypic**	**Allelic**	**Genotypic**	**Allelic**	**Genotypic**
Subgroup [a]	**N [b]**	**N**	(C, 70%)[c]	(C/C, 49%)	(C, 40%)	(C/C, 18%)	(A, 73%)	(A/A, 55%)	n/a	n/a
TOTAL	328	282	***0.04***	**0.07**	***0.04***	**0.03**	0.11	0.27	1.00	1.00
Unilateral	157	n/a	***0.02***	**0.05**	0.14	**0.07**	0.96	0.88	0.80	0.63
Bilateral	125	n/a	0.54	0.83	***0.04***	**0.05**	***0.004***	***0.02***	0.29	0.47
Complete	86	n/a	0.12	0.33	0.20	0.19	0.14	0.24	1.00	0.97
Incomplete	196	n/a	**0.09**	**0.07**	***0.009***	***0.004***	0.34	0.30	0.35	0.38
Uni+Comp	64	n/a	**0.05**	0.17	0.64	0.82	0.69	0.22	0.46	0.73
Uni+InComp	93	n/a	**0.08**	0.12	***0.04***	***0.003***	0.67	0.40	0.62	0.39
Bi+Comp	22	n/a	0.62	0.86	0.12	**0.08**	***0.007***	***0.03***	0.47	0.60
Bi+InComp	103	n/a	0.64	0.43	**0.05**	0.13	0.20	0.43	0.25	0.53

[a] p-values from case-control analyses, values in ***bold italics*** reach nominal significance (ie p-value<0.05), values in **bold** are suggestive (0.05<p-value<0.10).

[b] Phenotype Subgroups: groupings of the cleft cases, TOTAL = all cases, Unilateral and Bilateral refer to Laterality, Complete and Incomplete refer to the completeness of the cleft; Uni+Comp=unilateral and complete; Uni+InComp=Unilateral and Incomplete, Bi+Comp=bilateral and complete; Bi+InComp=bilateral and incomplete

[c] N=number or cases or controls, note that some of the phenotype subgroup numbers do not add up to the TOTAL because some of the cases did not have laterality or completeness information

[d] in parentheses is the associated allele or genotype for those SNPs in which one or more p-value < 0.