Am. J. Hum. Genet. 75:161–173, 2004

Meta-Analysis of 13 Genome Scans Reveals Multiple Cleft Lip/Palate Genes with Novel Loci on 9q21 and 2q32-35

Mary L. Marazita,[1,2] Jeffrey C. Murray,[3] Andrew C. Lidral,[4,5] Mauricio Arcos-Burgos,[6] Margaret E. Cooper,[1] Toby Goldstein,[1] Brion S. Maher,[1] Sandra Daack-Hirsch,[3] Rebecca Schultz,[3] M. Adela Mansilla,[3] L. Leigh Field,[7,8] You-e Liu,[9] Natalie Prescott,[10] Sue Malcolm,[10] Robin Winter,[10] Ajit Ray,[11] Lina Moreno,[4,5] Consuelo Valencia,[6] Katherine Neiswanger,[1] Diego F. Wyszynski,[12,13] Joan E. Bailey-Wilson,[14] Hasan Albacha-Hejazi,[17] Terri H. Beaty,[15] Iain McIntosh,[16] Jacqueline B. Hetmanski,[15] Gökhan Tunçbilek,[18] Matthew Edwards,[19] Louise Harkin,[22] Rodney Scott,[20] and Laurence G. Roddick[21]

[1]Center for Craniofacial and Dental Genetics, Division of Oral Biology, School of Dental Medicine, and [2]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh; Departments of [3]Pediatrics and [4]Orthodontics and [5]Dows Institute for Dental Research, University of Iowa, Iowa City; [6]University of Antioquia, Medellín, Colombia; [7]Department of Medical Genetics, University of British Columbia, and [8]British Columbia Research Institute for Children's and Women's Health, Vancouver; [9]Zhabei Genetics Institute, Shanghai; [10]Clinical and Molecular Genetics Unit, Institute of Child Health, London; [11]Department of Anthropology (Emeritus), University of Toronto, Toronto; [12]Genetics Program, Department of Medicine, and [13]Department of Epidemiology, Boston University, Boston; [14]Statistical Genetics Section, National Human Genome Research Institute, National Institutes of Health, [15]Department of Epidemiology, School of Public Health, and [16]Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore; [17]Ibn Al-Nafees Hospital, Damascus; [18]Department of Plastic Surgery, Hacettepe University, Ankara, Turkey; [19]Hunter Genetics, Hunter Area Health Service, and University of Newcastle, and [20]Discipline of Medical Genetics, School of Biomedical Sciences, University of Newcastle, and the Hunter Medical Research Institute, [21]John Hunter Children's Hospital and University of Newcastle, Newcastle, Australia; and [22]Department of Genetic Medicine, Women's and Children's Hospital, and Department of Paediatrics, University of Adelaide, Adelaide, Australia

Isolated or nonsyndromic cleft lip with or without cleft palate (CL/P) is a common birth defect with a complex etiology. A 10-cM genome scan of 388 extended multiplex families with CL/P from seven diverse populations (2,551 genotyped individuals) revealed CL/P genes in six chromosomal regions, including a novel region at 9q21 (heterogeneity LOD score [HLOD] = 6.6). In addition, meta-analyses with the addition of results from 186 more families (six populations; 1,033 genotyped individuals) showed genomewide significance for 10 more regions, including another novel region at 2q32-35 ($P = .0004$). These are the first genomewide significant linkage results ever reported for CL/P, and they represent an unprecedented demonstration of the power of linkage analysis to detect multiple genes simultaneously for a complex disorder.

Introduction

Orofacial clefts, particularly cleft lip (CL), cleft palate (CP), and cleft lip with or without cleft palate (CL/P), are very common structural birth defects with a complex etiology (Murray 2002). Birth prevalence ranges from 1/500 to 1/2,000, depending on the population, with Native American and Asian populations having the highest prevalence and African populations having the lowest. Individuals with these disorders often require multidisciplinary treatment into adulthood. In addition to the financial costs of such treatment, individuals with orofacial clefts and their families experience health burdens, such as increased morbidity and mortality, and a variety of psychosocial implications (Berk and Marazita 2002). Despite >200 years of research, the complex etiology of orofacial clefts in humans remains unclear, although the pace of discovery has quickened in recent years.

In 1757, the first description of a family with several affected members was published (Trew 1757). Many other studies followed that evaluated the family patterns of orofacial clefts (Marazita 2002*a*). The consensus from these studies was that there is a familial component to orofacial clefts but not a straightforward single gene-dominant or -recessive pattern. Indeed, statistical modeling of recurrence risk data in families with orofacial clefts suggests that 2–14 loci are likely to be involved (Schliekelman and Slatkin 2002).

Recently, investigators have been attempting to locate the genes predisposing to orofacial clefting, using linkage and/or association methods. The groundwork for these

Received February 26, 2004; accepted for publication May 7, 2004; electronically published June 4, 2004.

Address for correspondence and reprints: Dr. Mary L. Marazita, Center for Craniofacial and Dental Genetics, University of Pittsburgh, 100 Technology Drive, Suite 500 Cellomics Building, Pittsburgh, PA 15219. E-mail: marazita@sdmgenetics.pitt.edu

0002-9297/2004/7502-0003$15.00

Table 1

Summary of the 13 Populations Included in the GSMA

	Population Studied																	
	East Asians				Whites of European Descent					Other								
Population Description	Philippines[a]	China-A[b]	China-B[c]	Total	England[d]	United States (Pittsburgh)[e]	United States (Ohio)[f]	Australia[g]	Total	Colombia[f]	India[h]	Turkey-A[i]	Turkey-B[c]	Argentina, Mexico, and United States (Iowa)[j]	Syria[k]	Total	CIDR-7 Totals ($N = 7$ Studies)	Grand Totals ($N = 13$ Studies)
Genome scan performed by	CIDR-7	Marshfield	CIDR-7		In-house[l]	CIDR-7	CIDR-7	In-house[l]	...	CIDR-7	Marshfield/CIDR-7	Marshfield	CIDR-7	CIDR	CIDR	...	...	...
No. of families	220	36	75	331	92	24	9	1	126	48	38/11	18	1	10	2	117	388	574
Type of families	Extended multiplex	Extended multiplex	Extended multiplex	...	Affected sib pairs	Extended multiplex	Extended multiplex	Extended multiplex	...	Extended multiplex	Extended multiplex, some consanguineous	Consanguineous	Consanguineous	Extended multiplex	Extended multiplex, some consanguineous	...	...	...
No. of individuals	2,770	466	504	3,740	372	163	145	29	709	497	551/162	185	9	102	197	1,541	4,250	5,990
No. of individuals genotyped	1,676	310	347	2,333	338	122	61	19	540	261	272/81	55	3	59	61	711	2,551	3,584
No. of affected individuals genotyped	567	80	98	745	207	36	26	3	272	98	82/18	21	2	29	18	250	845	1,267
No. of unaffected individuals genotyped	1,109	230	249	1,588	131	86	35	16	268	163	190/63	34	1	30	43	461	1,706	2,317
Overall weight[m]	3.023	1.300	1.375	...	1.358	.816	.577	.322	...	1.193	1.218	.548	.128	.567	.577	...	...	...
Subgroup weight[m]	1.591	.684	.724	...	1.768	1.062	.751	.419	...	1.692	1.727	.777	.181	.804	.818	...	...	...

[a] Schultz et al. 2003.
[b] Marazita et al. 2002.
[c] M.L.M. and L.L.F., unpublished data.
[d] Prescott et al. 2000.
[e] Neiswanger et al. 2003; Weinberg et al. 2003.
[f] Moreno et al. 2003.
[g] Edwards et al. 2003.
[h] Field et al., in press.
[i] Marazita et al. 2004.
[j] Zeiger et al. 2003.
[k] Wyszynski et al. 2003.
[l] In-house = genotyping performed in the principal investigator's laboratory or by in-house genotyping service.
[m] Weighting factor for the GSMA, calculated as the square root of the number of individuals genotyped in a particular study, divided by the mean across all studies (Levinson et al. 2003). Weighting factors were calculated separately for each of the three subgroups as well as for the total study.

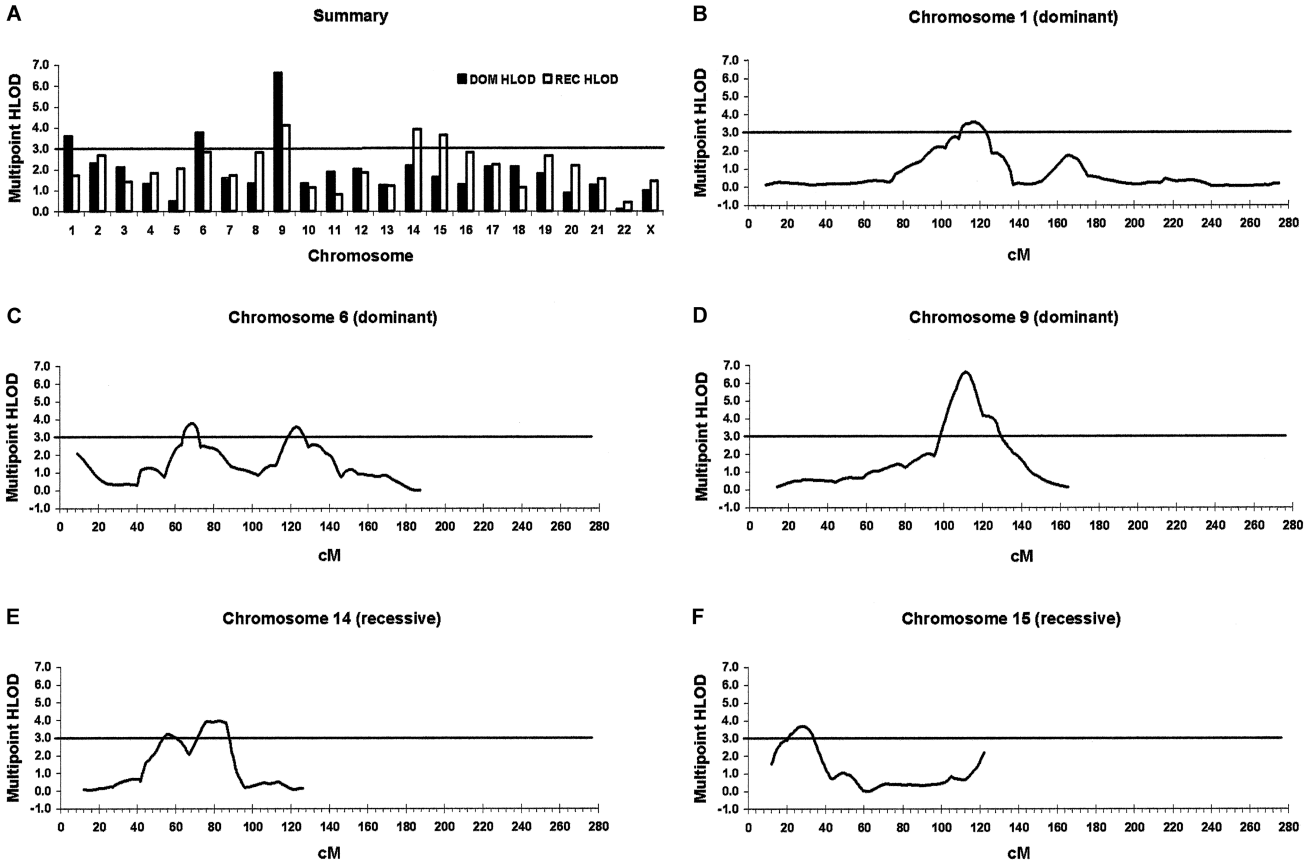

Figure 1 Summary of the 10-cM genome scan of CL/P in the CIDR-7 studies. *A,* Summary of the maximum summed multipoint HLOD for each chromosome, under both dominant (DOM) and recessive (REC) genetic models for CL/P. *B–F,* Summed multipoint HLOD plots for each chromosome that had a maximum summed HLOD ≥3.2 (under the best genetic model for each chromosome).

studies was laid by application of complex statistical analyses that allowed direct comparison of multifactorial and single-gene models in CL/P (reviewed by Marazita [2002*b*]), which demonstrated that single-gene models provided an equal or better fit to the family data than did the "large number of equal and additive loci" specified by standard multifactorial models (Marazita 2002*b*). The possibility that one or a few loci might explain orofacial clefts made linkage and association studies feasible.

After several negative linkage and association studies, the first positive linkage finding with orofacial clefting was at the F13A locus on chromosome 6p (Eiberg et al. 1987), and the first positive allelic association was between CL/P and a *Taq*I RFLP in the transforming growth factor α (TGFA) locus (Ardinger et al. 1989). Since those early linkage and association studies, there have been many others (Marazita 2002*a*, 2002*b*). Positive linkage and/or association results in CL/P, CP, or both have been found, primarily on seven chromosomes (1, 2, 4, 6, 14, 17, and 19) (Marazita 2002*a*).

The first CL/P genome scan reported results in English affected sib pairs (Prescott et al. 2000); since then, 12 additional genome scans have been performed (see table 1), each of which identified a number of positive signals, although no single study produced results that reached the standard levels of genomewide significance. Clearly, for complex traits such as orofacial clefts, large numbers of families and/or combined results across studies are essential. Therefore, a consortium of four research groups (M.L.M./L.L.F., J.C.M., A.C.L., and M.A.-B.) was formed to undertake genome scans of a large number of extended multiplex families with nonsyndromic CL/P (NS CL/P) from seven populations (see table 1*)*.

In addition, we performed a meta-analysis in which results from the seven populations were combined with six other published genome scans of CL/P (see table 1*)*. Various methods have been proposed for meta-analysis in genetic studies, dating back to Fisher (1925), that require either tests of the same hypotheses or the same statistical tests across studies. The 13 available genome scans of NS CL/P did not test the same hypotheses (i.e., different marker sets were used) and did not all calculate the same statistics. Therefore, we applied the

genome scan meta-analysis (GSMA) method (Wise et al. 1999; Levinson et al. 2003). Simulation studies that used this method detected linkage with power comparable to or greater than that obtained by performance of a combined linkage analysis of all data (Levinson et al. 2003). The major goal of the current study was to identify those genomic regions that contain genes predisposing to CL/P across populations.

Material and Methods

Study Populations

There were 13 study populations with genome scans available for families with NS CL/P (summarized in table 1). There were 574 families in the 13 populations, with 5,990 individual family members. Of the total family members, 3,584 were genotyped (1,267 affected; 2,317 unaffected). Most of the studies involved extended multiplex kindreds; that is, multigenerational families with at least two affected individuals. One study involved affected sib pairs (England), and all or part of the families in four studies were consanguineous (India, Turkey-A, Turkey-B, and Syria). Two countries each had two genome scans (China-A and China-B; Turkey-A and Turkey-B). Note that each genome scan consisted of unique families—that is, there is no overlap between the families of China-A and China-B or between those of Turkey-A and Turkey-B. The India study had 38 families genotyped by Marshfield (Center for Medical Genetics), of which 11 families were also genotyped as part of CIDR-7 (see below).

The phenotype analyzed in each study was NS CL/P; in all studies, families were ascertained through probands, and additional relatives were recruited. For families to be included, it was necessary that the proband have isolated CL with or without CP (i.e., no other anomalies) and that no other family member have an indication of a clefting syndrome (e.g., lip pits). Each study included evaluations of family members by clinical geneticists to rule out syndromic forms of CL/P. Each study had approval by the appropriate institutional review boards; to participate, all study subjects provided informed consent. Refer to each study reference for more specific details on ascertainment, genotyping, and family structures (see table 1).

Genome-Scan Genotyping

Seven of the studies (Philippines, China-B, India, Colombia, United States-Pittsburgh, United States-Ohio, and Turkey-B; 388 families) (see table 1) were part of a collaborative genome-scan project (between M.L.M./L.L.F., J.C.M., A.C.L., and M.A.-B.) and thus were all genotyped by the Center for Inherited Disease Research (CIDR) at the same time. We will refer to these studies as the "CIDR-7 studies." The other six studies (186 additional families) were genotyped either by CIDR, by Marshfield Laboratories Mammalian Genotyping Service (Center for Medical Genetics), or in-house by study investigators (see table 1 for details). In most cases, Weber screening-set markers were used (Yuan et al. 1997); that is, high-quality STRPs with an average spacing of ~9 cM (with a range of 1–19 cM). For details of the markers and genotyping, refer to the Web sites for CIDR and Marshfield and to the references for each genome scan (see table 1).

Statistical Genetic Analyses of Genome-Scan Markers

Each of the 13 studies analyzed the genome-scan data in a variety of ways, by use of both parametric and model-free methods of linkage analysis and, in most cases, also by application of tests of linkage in the presence of association, such as the transmission/disequilibrium test (TDT) (see individual references for details). When different studies have exactly the same markers and exactly the same analysis methods, it is straightforward to combine results. The CIDR-7 families were genotyped for exactly the same markers and were analyzed using the same methods as in M.L.M.'s research group. For the other six studies, different markers were used, and, in some cases, different statistical analyses were performed. Therefore, results across all 13 studies were combined into a meta-analysis (see the "GSMA" section).

For purposes of combining the results from the CIDR-7 studies, we utilized multipoint parametric linkage statistics, in particular the heterogeneity LOD score (HLOD). HLODs are based on the admixture heterogeneity test (Smith 1963), in which the recombination fraction (θ) and the proportion of linked families (α) are estimated simultaneously. Recent 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 (Greenberg and Abreu 2001; Vieland et al. 2001; Hodge et al. 2002). The descent-graph method (Sobel et al. 1996; Sobel and Lange 1996; Lange 2002) implemented in the computer program SIMWALK2 was used for the multipoint HLOD calculations of the CIDR-7 data.

The inheritance of each marker in each of the CIDR-7 study populations was analyzed with PedCheck (O'Connell and Weeks 1998) to test for inconsistencies due to nonpaternity or other errors. Parametric linkage approaches such as the HLOD approach used here require estimates of marker-allele frequencies and parameters of the genetic model. Allele frequencies were estimated in the founders of the families, separately by study population because the seven populations were of diverse ethnicities. HLOD calculations were done under

the best-fitting dominant and recessive models for each study population (estimated from segregation analysis; see individual references). Maximizing LOD scores over a range of genetic 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 (Hodge and Elston 1994). 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 in a recessive one.

To combine the multipoint results across the CIDR-7 studies, we summed the multipoint HLODs. Because each study population had different allele-frequency estimates and different genetic-model parameters, it was not possible simply to perform a combined linkage analysis over all families. Further, simulation studies by Vieland et al. (2001) showed that combined HLOD analyses that pool multiple data sets resulted in a loss of power for detecting linkage, compared with summing the HLODs from individual studies.

The meta-analysis method that was used to combine the results from all 13 studies (see the "GSMA" section) can combine any statistics; for maximum consistency across studies, we utilized linkage statistics, including affected–sib pair tests (identical-by-descent–based statistics) from the England study, model-free two-point NPL statistics from the Argentina/Mexico/Iowa study, and two-point parametric LOD scores from all other studies. For the CIDR-7 families, we calculated two-point LOD scores using the Elston-Stewart algorithm (Elston and Stewart 1971), employing the LINKAGE program with recent updates to speed calculations (VITESSE and FASTLINK) (Cottingham et al. 1993; Terwilliger and Ott 1994; O'Connell and Weeks 1995).

GSMA

To combine the genome-scan results across all 13 studies, a meta-analysis procedure was necessary. There are many proposed methods for combining results across diverse studies. Fisher (1925) proposed a simple and elegant method for combining probabilities from tests of significance across multiple studies. Province (2001) supplied an update to incorporate certain modern linkage and association statistics that are truncated at 0.0. In Fisher's method, it is not necessary that the same statistic be used for calculating *P* values across studies; however, it is necessary that the *P* values arise from tests of the same hypothesis—that is, tests of linkage or association to the same marker. To combine *P* values across "*i*" studies, sum the quantity: $(-2 \ln P_i)$, where P_i is the *P* value in the *i*th study (Fisher 1925); for parameters that are truncated at 0.0, use $P = 1/(2 \ln 2) = .72$ (Province 2001). Various other methods have also been proposed (reviewed by Wise et al. [1999], Province [2001], and Levinson et al. [2003]) that require either tests of the same hypotheses or that the same statistical tests be applied across studies.

The 13 available genome scans of NS CL/P did not test the same hypotheses (i.e., different marker sets were used) and did not all calculate the same statistics. Therefore, we applied the GSMA method (Wise et al. 1999; Levinson et al. 2003) to combine data across studies. 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 (Levinson et al. 2003).

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 that 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 (130 bins across the genome). For each of the 13 studies, bins were assigned a rank (*R*, with values of 1–130) according to the maximum-linkage statistic of markers in each bin. Any tied bins were assigned equal *R*s on the basis of the mean of the sequential ranks for those bins.

Because the 13 study populations covered a wide range of sample sizes (see table 1), we weighted the *R* statistics on the basis of sample size. Optimal weighting strategies are not determined for the GSMA (Wise et al. 1999; Levinson et al. 2003), but simulation studies showed that weighting increased the power of the GSMA to detect linked loci (Levinson et al. 2003). We used a weighting factor based on the total number of genotyped individuals—the ranks within each study were multiplied by

$$\frac{\sqrt{N \text{ genotyped}}}{\text{mean}\sqrt{N \text{ genotyped}}} .$$

Weighting factors for each study are listed in table 1.

For combining the rank results across studies, the average rank (R_{avg}) for each bin was calculated across all 13 studies. The probability distribution for unweighted *R* is derived (Wise et al. 1999); however, for weighted *R*s, the probability distribution must be determined empirically. To determine statistical significance, the resulting R_{avg} distribution was then compared with its empiric probability distribution, derived from 1,000 simulations under the assumption that ranks were assigned randomly (GSMA *P* values [denoted "P_{avg}" by Levinson et al. (2003)]). The P_{ord} (Levinson et al. 2003) probabilities also were calculated for each bin to provide additional in-

formation about interpreting the GSMA results with respect to linkage to each bin (see the "Statistical Significance" section below). The P_{ord} probability for a rank is the $P_{avg\ Rank/order}$—in other words, that bin's place in the order of average ranks—and quantifies the consistency of the R_{avg} across simulations.

The GSMA procedure was done across the entire set of 13 studies, then repeated for three subgroups: "East Asian," "white," and "other." Table 1 lists the composition of the subgroups, including the weighting factors used within each subgroup. Note that, unlike the East Asian and white subgroups, the "other" subgroup is a catch-all category (with no particular biological meaning). Within the "other" subgroup, there are studies of South Americans, Indians, Turks, and Syrians, but there are not enough families in any of those groups to have sufficient power to analyze them separately.

The GSMA identified 30-cM bins that are best supported across the studies. To narrow the regions of positive findings, we propose 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). We repeated the GSMA twice, shifting the starting point for the binning procedure first to 10 cM and then to 20 cM. This determined the 10-cM MRMS for each positive finding. Given that all genome scans averaged 10 cM between markers, 10 cM is the limit of resolution for this meta-analysis.

Candidate-Gene–Analysis Methods (Markers on 9q21)

To pursue the highly significant results on 9q21, five candidate genes that lie within or immediately adjacent to the one- and two-LOD intervals of the peak HLOD score were selected for additional studies. SNPs for four of these genes (*PTCH, ROR2, TGFBR1,* and *ZNF189*) were genotyped in 219 multiplex Filipino families; microsatellite markers for two of the genes (*FOXE1* and *TGFBR1*) were genotyped in 99 Chinese, 50 Indian, and 18 Turkish families.

The TDT was used to assess association in the presence of linkage disequilibrium between the markers and CL/P. The family-based association-test (FBAT) extension of the TDT (Laird et al. 2000; Rabinowitz and Laird 2000; Horvath et al. 2001) was used to assess association between alleles at each marker and CL/P, plus association between CL/P and haplotypes of SNPs within candidate genes. Linkage was assessed using the methods described in sections above. Since *TGFBR1* was assessed in all the data sets, the Fisher (1925) method was used to combine the *TGFBR1* results across studies.

Statistical Significance

For determining genomewide significance for the multipoint-linkage calculations of the CIDR-7 families, standard guidelines (Lander and Kruglyak 1995) were followed. The desired α-level of .05 was divided by 400 (the approximate number of genome-scan markers) to yield .000125; therefore, the corresponding LOD-score threshold for genomewide significance was 3.2.

For the GSMA, result-interpretation guidelines presented by Levinson et al. (2003) were followed. By use of a Bonferoni correction, the criterion for genomewide significance for any particular bin would be .05/130 = .0004. In the situation where more than one bin is likely to contain a linked locus, Levinson et al. (2003) determined (from simulation studies) that, if the GSMA *P* value ("P_{avg}" in the notation of Levinson et al. [2003]) and the P_{ord} were both $\leqslant$.05 for a particular bin, then that bin was highly likely to contain a linked gene. Therefore, we considered the GSMA results to have genomewide significance either if the GSMA/MRMS *P* value was $\leqslant$.0004 or if both the *P* value and P_{ord} were $\leqslant$.05.

Results

Figure 1 summarizes the results of the combined multipoint linkage analyses of NS CL/P in the CIDR-7 populations. Six regions on five chromosomes had summed HLODs $\geqslant$3.2 (chromosomes 1p12-13, 6p23, 6q23-25, 9q21, 14q21-24, and 15q15). There were nine additional regions with summed HLODs between 2.0 and 3.2 (see figs. 1*A* and A1 [online only]).

Figure 2 summarizes the results of the initial GSMA for all 13 studies. Depicted are the R_{avg} for each of the 130 bins, across all chromosomes, with the thresholds for *P* values of .05 and .01 indicated. Figure 3 provides a graphical representation of the bin-shifting procedure and the MRMS results for the six chromosomes with the most significant GSMA/MRMS results (see fig. A2 [online only] for all chromosomes). Under the MRMS procedure, the significance increased for most of the regions that were positive in the initial GSMA, whereas some regions that were not significant in the initial GSMA were significant after rebinning. Supplemental figure A3 (online only) shows the GSMA results, by population subgroup, for each chromosome. Table 2 summarizes the genomewide significant results from the combined analyses of the CIDR-7 studies as well as the GSMA/MRMS of the 13 studies and of the population subgroups.

Of the six regions of genomewide significance in the CIDR-7 analyses, two also had genomewide significance in the meta-analysis (6q23-25 and 14q21-24). One region, 2q32-35, had a genomewide-significant GSMA *P* value ($P = .0004$), whereas nine additional regions met the combined genomewide significance criteria of both *P* and $P_{ord} \leqslant .05$ (Levinson et al. 2003) (table 2).

For the markers in 9q21 genes *PTCH, ROR2, TGFBR1,* and *ZNF189* (see fig. 4), linkage and associa-

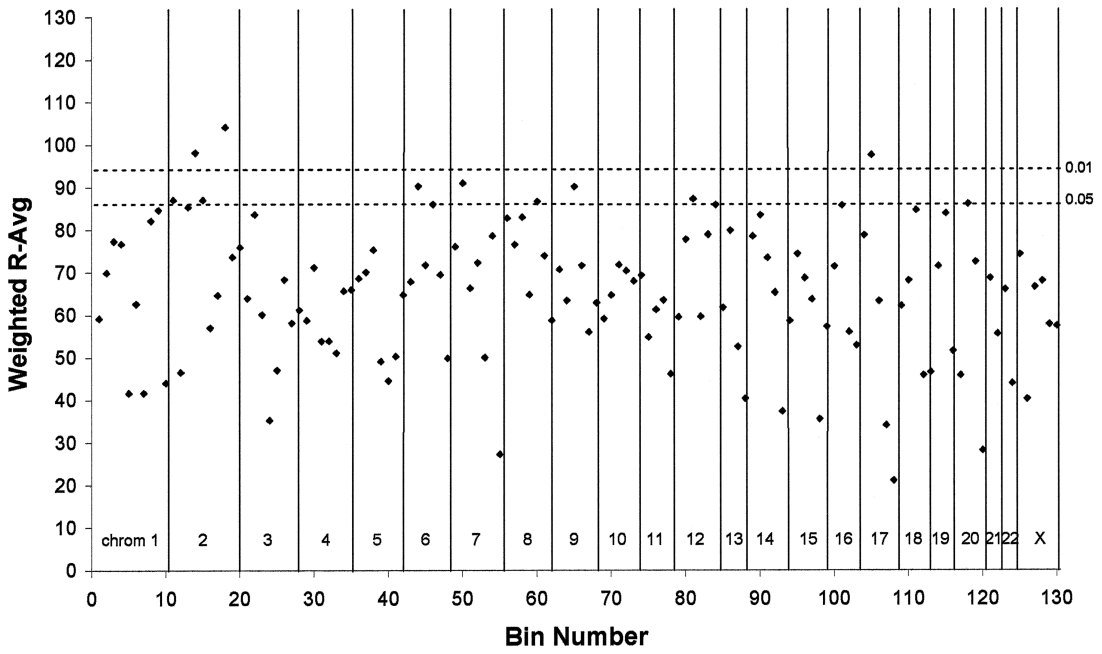

Figure 2 GSMA of 13 CL/P-study populations. Depicted are the R_{avg} for each of the 130 bins, across all chromosomes. The horizontal lines indicate the R_{avg} values corresponding to empirical *P* values of .05 and .01, as determined by simulation.

tion were assessed and results combined across studies by use of the Fisher (1925) method. Positive associations were found with four of the candidate genes (*ROR2* $P = .02$; *PTCH* SNP haplotype $P = .007$; *TGFBR1* $P = .02$; and *FOXE1* $P = .002$), and suggestive linkage was found with *FOXE1* (summed maximum LOD = 1.54; $\theta = 0.05$).

Discussion

Combined results from 13 genome scans were remarkable, given the genetic complexity of CL/P; one region (9q21) had an HLOD of 6.6, and another (2q32-35) had a meta-analysis *P* value of .0004. This demonstrates the great utility of the GSMA in combining disparate data sets. Further, we summarize here a new extension of the GSMA to determine the MRMS. The combined GSMA/MRMS allowed us to narrow the regions of significance from 30 cM to 10 cM (fig. 3 and table 2). It is interesting that some statistically significant regions were discovered only during one of the repeat bin–shifted GSMAs used to determine the MRMS (table 2), highlighting the importance of utilizing bin shifting when applying the GSMA.

The chromosomal regions that had statistical genomewide significance in this study include many of the regions reported elsewhere for NS CL/P but also identify two highly significant novel regions, 9q21 and 2q32-35. Positive association results (indicating close proximity to a CL/P locus) were also found with four of the five selected candidate genes on 9q21. Overall, 16 regions on 12 chromosomes had genomewide-significant results (table 2). Six other regions on four chromosomes had suggestive results (HLOD between 2 and 3 and/or GSMA/MRMS *P* value <.05 but $P_{ord} > .05$). The finding of multiple chromosomal regions with statistically significant results is consistent with an oligogenic model for CL/P, as proposed elsewhere (Farrall and Holder 1992; Mitchell and Risch 1992; Schliekelman and Slatkin 2002).

Table 2 summarizes all results of genomewide significance and indicates potential cleft candidate genes in each positive region. A brief discussion of the most notable results by chromosome follows. Although not summarized in detail, several regions also have chromosomal rearrangements reported in CL/P (Brewer et al. 1998, 1999).

Chromosome 1

The 1q32 region is the location for interferon regulatory factor–6 (*IRF6*) that was identified recently as the locus involved in van der Woude syndrome (VDWS [MIM 119300]) (Kondo et al. 2002). Our group recently has found highly significant evidence of association between *IRF6* and CL/P (Zucchero et al., in press), with TDT $P < 10^{-9}$ in a sample of 3,350 nuclear families (some families from CIDR-7). The current results that demonstrate linkage to the *IRF6* region provide pow-

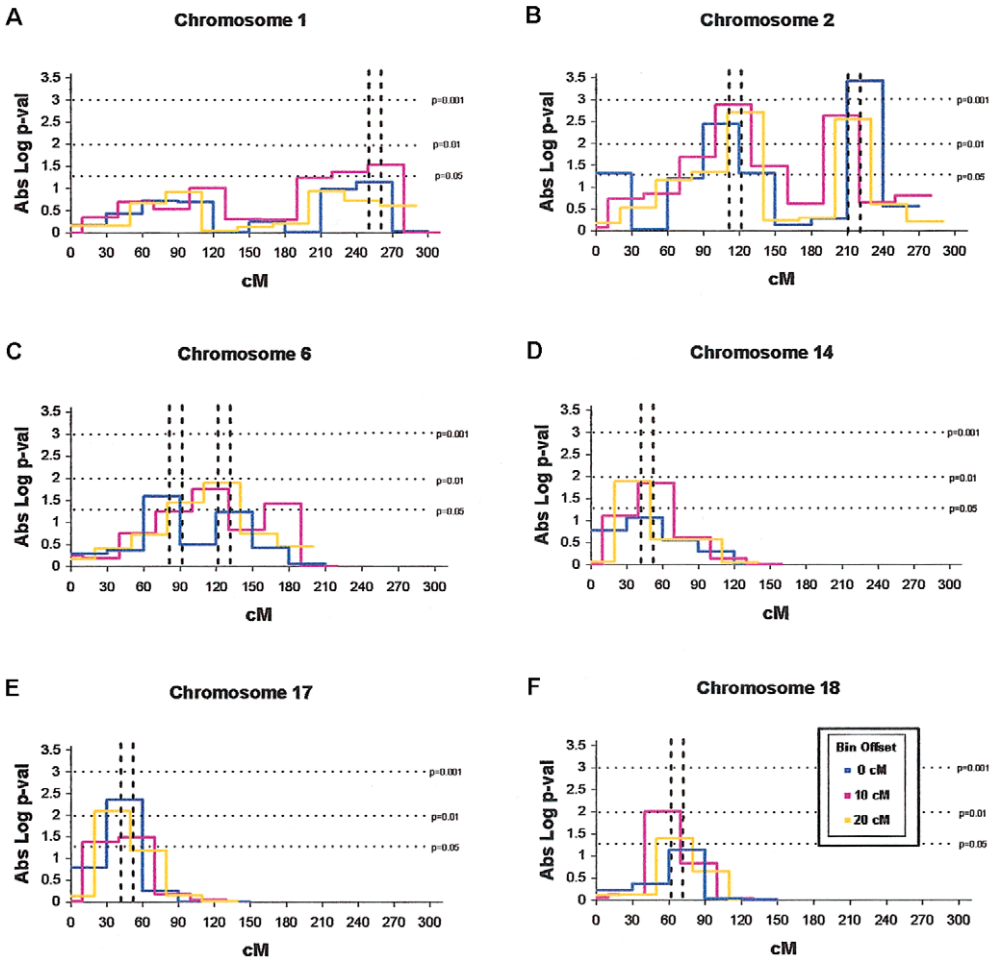

Figure 3 Graphs of the MRMSs for the six chromosomes with the most significant GSMA/MRMS results for the 13 CL/P-study populations. These graphs summarize the MRMS process, repeating the GSMA with bin shifting to narrow the statistically significant region. The dashed vertical lines indicate the 10-cM MRMS for each chromosome. *A–F,* Graph of chromosomes 1, 2, 6, 14, 17, and 18, respectively.

erful support for the *IRF6*-association findings. It is interesting that there was only a weak linkage signal to the *IRF6* locus itself (LOD < 1.0) (Zucchero et al., in press); thus, the current study demonstrates the utility of combining genome scans across studies to detect subtle effects.

Chromosome 2

The 2q32-35 region had the most highly significant GSMA results in the current study ($P = .0004$) and contains the gene for the DNA-binding protein SATB2 (a.k.a. "KIAA1034") that has been identified elsewhere, through translocation-breakpoint analysis, as a gene involved in cleft palate (FitzPatrick et al. 2003) and that also shows site- and stage-specific expression in murine palate development.

The 2p13 region contains *TGFA,* the gene with the first reported association with CL/P (Ardinger et al. 1989) and numerous confirmatory reports (Mitchell 1997; Marazita 2002*a*). *TGFA* has ~40% sequence homology with epidermal growth factor (EGF) and competes with EGF for binding to the EGF receptor. The results for the 2p13 region did not reach genomewide significance but were suggestive (HLOD = 2.67; GSMA $P = .001$; $P_{ord} = .18$).

Chromosome 4

Chromosome 4 is a notable omission from the list of chromosomes with positive results in this study. Two regions on chromosome 4 have had positive linkage or association results in the literature: homeobox 7 (*MSX1* 4p16.1 [MIM 106600] and 4q21-31) (Marazita 2002*a*). Regarding 4p16.1, positive associations between CL/P and *MSX1* have been reported (Marazita 2002*a*), hypodontia with oral clefts is due to *MSX1,* mouse *Msx1* knockouts have cleft palate, and complete sequencing of

Table 2

Genomewide Significant Linkage and Meta-Analysis Results

	CIDR-7 Results[a]				GSMA/MRMS[b]		
Chromosomal Region	Maximum Summed HLOD[d]	Peak[e] (cM)	Model[f]	Average α[g] (%)	10-cM Interval[h] (P, P_{ord})	Subgroup with $P \leq .05$[i,j]	Candidate Genes in Region (OMIM Designation)[c]
1p12–13	3.58	116	Dominant	42.9	…	…	*COL11A1*
1q32	…	…	…	…	260–270 (.030, .001)	EA, O	*IRF6*
2q32–35	…	…	…	…	210–220 (.0004, .049)	EA, O	*SATB2, WNT6, DLX2*
3p25	…	…	…	…	60–70 (.017, .001)	O	*WNT7A, RARB, THRB, TGFBR2*
6p23	3.74	68	Dominant	52.1	…	…	*F13A1, TFAP2A, BMP6, EDN1, COL11A2* (OFC1 [MIM 119530])
6q23–25	3.55	123	Dominant	19.3	120–130 (.013, .033)	EA	
7p12	…	…	…	…	70–80 (.027, .001)		*EGFR*
8p21	…	…	…	…	60–70 (.027, .001)	EA	*FGFR1, EGR3, PPP3CC, FZD3, NRG1*
8q23	…	…	…	…	140–150 (.014, .001)	W	*FZD6*
9q21	6.62	111	Dominant	42.1	…	…	*PTCH, ROR2, TGFBR1 ZNF189, TTF2(FOXE1), clf2* homology
12p11	…	…	…	…	80–90 (.018, .008)	O	
14q21–24	3.92	83	Recessive	15.0	40–50 (.013, .005)	EA, W	*TGFB3, BMP4*
15q15	3.65	28	Recessive	27.1	…	…	*ZNF291, NRG4*
17q21	…	…	…	…	40–50 (.004, .017)	W, O	*RARA, clf1* homology, *WNT9B*
18q21	…	…	…	…	60–70 (.010, .015)		*MADH(SMAD2)*
20q13	…	…	…	…	50–60 (.021, .001)	EA	*EDN3, BMP7*

[a] $N = 7$ studies; 388 families.

[b] $N = 13$ studies; 574 families. Results of genomewide significance from GSMA of 13 total studies (i.e., either $P \leq .0004$ or both P and $P_{ord} \leq .05$ [Levinson et al. 2003]).

[c] List of potential cleft candidate genes in regions. One of the regions has the designation "orofacial cleft 1" (OFC1) on Online Mendelian Inheritance in Man (OMIM).

[d] Maximum multipoint HLOD summed for the seven CIDR-7 studies (see table 1 for list; see figs. 1 and A1 [online only] for full HLOD plots); presented are those regions with HLODs ≥3.2 (threshold for genomewide significance [Lander and Kruglyak 1995]).

[e] Location (cM) of the maximum summed multipoint HLOD.

[f] Genetic model for the maximum summed HLOD.

[g] Estimated proportion of linked families at the peak, averaged over the CIDR-7 studies.

[h] The 10-cM interval best supported by the GSMA/MRMS, with the corresponding P value and P_{ord} (see figs. 3 and A2 [online only] for MRMS graphs).

[i] Population subgroups: EA = East Asian; W = white; O = "other." (See table 1 for list of populations in each subgroup; see fig. A3 for population-specific GSMA/MRMS graphs.)

[j] For chromosomes 7p12 and 18q21, none of the population-specific GSMA results were significant (none ≤.05), although the overall results were significant.

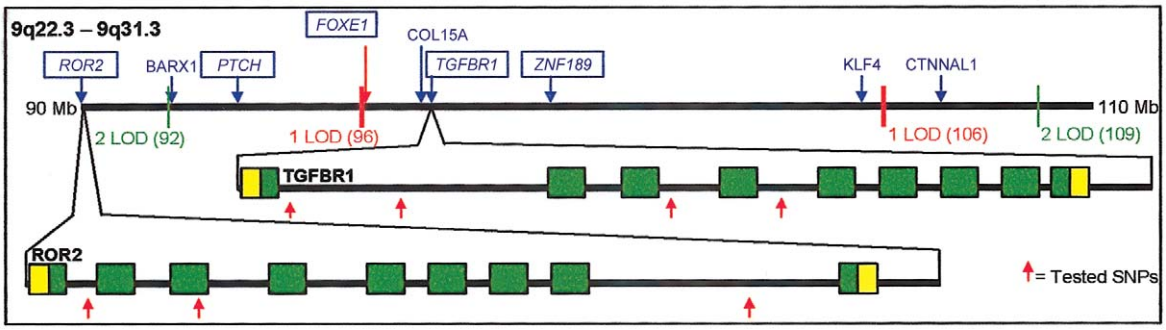

Figure 4 Diagram of chromosome 9q21, with one- and two-LOD support intervals and candidate genes depicted

the *MSX1* gene showed mutations in ~2% of NS CL/P cases (Jezewski et al. 2003). If an *MSX1* variant is a low-frequency gene predisposing to CL/P or if *MSX1* is a modifier of expression, it would be difficult to detect by use of linkage methods. The fact that relationships between CL/P and both *MSX1* and *IRF6* were detected using nonlinkage approaches highlights the necessity of applying diverse, complementary approaches for adequately delineating complex traits.

MSX1-related CL/P may be limited to certain population groups; in the subgroup-specific GSMA/MRMS analyses, there was a borderline significant result (P = .05) near the *MSX1* region in the "other" subgroup but not in the white or East Asian subgroups.

Chromosome 9

Region 9q21 had the most significant linkage result and also was suggestive in the GSMA/MRMS (P = .01; P_{ord} = .058). Mutations in the gene for the human homolog of *patched* (*PTCH* [9q22.3]) cause basal cell nevus syndrome (Gorlin syndrome [MIM 109400]), in which ~5% of cases have CL/P. Furthermore, *PTCH* maps to the region homologous to the *clf2* locus in the mouse (Juriloff 2002). Receptor tyrosine kinase–like orphan receptor 2 (*ROR2* [9q22]) is selectively expressed in—and particularly important for—the chondrocyte lineage. Transforming growth factor β receptor type 1 (*TGFBR1* [9q33-q34]) mediates the induction of several genes involved in cell-matrix interactions, including the *TGFB* superfamily (thought to be developmental regulators of cell proliferation and differentiation). The gene for zinc finger protein 189 (*ZNF189* [9q22-q31]) is one of the estimated 300–700 ZNF genes in the human genome. ZNFs bind nucleic acids and regulate transcription. Mutations in the forkhead domain–containing transcription factor *FOXE1* (a.k.a. *"TTF2"* or *"TITF2"* [9q22]) are associated with congenital hypothyroidism, thyroid agenesis, and cleft palate in humans and mice (Castanet et al. 2002; Dathan et al. 2002). 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" (Lehman et al. 2003) and regulates many diverse developmental processes in eukaryotes. We assessed linkage and association with each of these five potential candidates, and we found positive associations with *PTCH, ROR2, TGFBR1,* and *FOXE1* and suggestive linkage with *FOXE1.* A major current focus of our research group is fine mapping of this region.

Chromosome 14

The region containing transforming growth factor β 3 (*TGFB3*) (14q24) was significant in both CIDR-7 and the GSMA/MRMS. The mouse *Tgfb3* knockout has cleft palate, and there are some positive results from association studies of CL/P and *TGFB3* (Marazita 2002*a*). We also recently found suggestive evidence that the CL/P phenotype within ectrodactyly-ectodermal dysplasia-clefting syndrome (EEDC1) maps to 14q, indicating that genes involved in NS CL/P may modify the phenotypic expression of the *EEC1* gene (Ray et al., in press).

Chromosome 17

The 17q21 region had significant results in the GSMA/MRMS analyses and suggestive results in the CIDR-7 linkage analysis (HLOD = 2.23). It is notable that this region is syntenic with the region harboring the mouse *clf1* mutation (Juriloff 2002). Further, one of the early associations observed with NS CL/P was with the gene for retinoic acid receptor α that maps near this region (*RARA* [17q12]) (Marazita 2002*a*).

Chromosome 19

Linkage and association have been seen in a few studies between CL/P and *APOC2, BCL3,* or nearby anonymous markers (19q13) (Marazita 2002*a*). The current results for 19q13 are also suggestive (HLOD = 2.66; GSMA

$P = .012$; $P_{ord} = .088$). Genes in this region include *TGFB1* (19q13.1) and polio-virus receptor (*PVR*)/poliovirus receptor–like 2 (*PVRL2*), both at 19q13.2-q13.3. It is interesting that mutations of *PVRL1* (11q23-q24) cause the autosomal recessive Margarita Island clefting syndrome, in which it has been suggested that heterozygotes have an increased risk of clefting (Sozen et al. 2001). A family segregating a balanced translocation and clefting has its breakpoint in the *CLPTM1* gene in this region (Yoshiura et al. 1998).

* * *

The studies summarized here combine results from diverse populations of families with CL/P. Therefore, the positive results necessarily will be those that are in common across populations. Note that a number of chromosomes had population subgroup–specific differences in the patterns of GSMA results (summarized in table 2). There is no formal test of heterogeneity within the GSMA, but the results summarized in table 2 (also see fig. A3 [online only]) could guide prioritization for fine-mapping and gene identification in particular population subgroups.

In summary, the remarkably significant evidence for linkage from this large study demonstrates the advantage of combining data sets to further our understanding of the genetics of a complex trait such as CL/P. This is an unprecedented demonstration of the power of linkage analysis to detect simultaneously multiple genes for a truly non-Mendelian disease in which no single major genes are evident. To identify and characterize those genes, fine mapping and candidate-gene sequencing has begun in our research groups, as have investigations of gene-gene interactions and genotype-phenotype correlations. After almost 250 years of research regarding the familial nature of CL/P, evidence is accumulating that a relatively small number of chromosomal regions contain causal genes for NS CL/P. With the molecular and statistical tools available today, and with our rich resource of extended multiplex families with CL/P, this becomes a manageable number of chromosomal regions to characterize in order to identify the causal genes for CL/P.

Acknowledgments

This study is dedicated to the memory of coauthor Robin Winter, who died, at age 53 years, during the preparation of this manuscript. Robin was a tireless clinician and scientist who made enormous contributions to phenotyping of birth defects and was someone who made possible wide-ranging collaborations, such as this one, through his talents, wit, and fellowship. He will be greatly missed. Many thanks are due to the families all over the world who participated in these studies. This study benefited from many fruitful discussions during the joint University of Iowa/University of Pittsburgh Craniofacial Genetics Journal Club (via weekly video conferences, supported by National Institutes of Health [NIH] grant P60-DE13076). This study was supported by NIH grants R01-DE-08559 (J.C.M. and M.L.M.), R01-DE-09886 (M.L.M., L.L.F., and Y.L.), R01-DE-12472 (M.L.M., L.L.F., and G.T.), R01-DE-14667 (A.C.L., M.A.-B., and M.L.M.), P60-DE-13076 (J.C.M., M.L.M., and A.C.L.), P60-DE-13078 (T.H.B., D.F.W., and J.B.-W.), and RR-00084 (MLM); by Medical Research Council of Canada grant MT-11263 (L.L.F.); by March of Dimes grant 6F601-616 (A.C.L., M.A.-B.); by a grant from United Kingdom Action Research, Birth Defects Foundation (N.P. and R.W.); and by a grant from the Area Clinical Services Sector, Hunter Area Health Service, Newcastle, Australia (M.E., L.H., R.S., and L.R.). Genotyping was provided by grants from the Mammalian Genotyping Service (see table 1) and CIDR (see table 1).

Electronic-Database Information

The URLs for data presented herein are as follows:

Center for Inherited Disease Research (CIDR), http://www.cidr.jhmi.edu/

Center for Medical Genetics, http://research.marshfieldclinic.org/genetics/ (for the Marshfield Mammalian Genotyping Service)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for VDWS, *MSX1*, Gorlin syndrome, and OFC1)

References

Ardinger HH, Buetow KH, Bell GI, Bardach J, VanDemark DR, Murray JC (1989) Association of genetic variation of the transforming growth factor alpha gene with cleft lip and palate. Am J Hum Genet 45:348–353

Berk NW, Marazita ML (2002) Costs of cleft lip and palate: personal and societal implications. In: Wyszynski DF (ed) Cleft lip and palate: from origin to treatment. Oxford University Press, Oxford, pp 458–467

Brewer C, Holloway S, Zawalnyski P, Schinzel A, FitzPatrick D (1998) A chromosomal deletion map of human malformations. Am J Hum Genet 63:1153–1159

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

Castanet M, Park SM, Smith A, Bost M, Leger J, Lyonnet S, Pelet A, Czernichow P, Chaterjee K, Polak M (2002) A novel loss-of-function mutation in TTF-2 is associated with congenital hypothyroidism, thyroid agenesis and cleft palate. Hum Mol Genet 11:2051–2059

Cottingham RW Jr, Idury RM, Schaffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53: 252–263

Dathan N, Parlato R, Rosica A, De Felice M, Di Lauro R (2002) Distribution of *titf2/foxe1* gene product is consistent with an important role in the development of foregut endoderm, palate, and hair. Dev Dyn 224:450–456

Edwards M, Roddick L, Scott R (2003) Autosomal dominant nonsyndromic cleft lip and palate linked to chromosome 4. Am J Hum Genet Suppl 73:672

Eiberg H, Bixler D, Nielsen LS, Conneally PM, Mohr J (1987) Suggestion of linkage of a major locus for nonsyndromic orofacial cleft with F13A and tentative assignment to chromosome 6. Clin Genet 32:129–132

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Farrall M, Holder S (1992) Familial recurrence-pattern analysis of cleft lip with or without cleft palate. Am J Hum Genet 50:270–277

Field LL, Ray A, Cooper ME, Goldstein T, Neiswanger K, Shaw DF, Marazita ML. Genome-scan for loci involved in nonsyndromic cleft lip with or without cleft palate in families from West Bengal, India. Am J Med Genet (in press)

Fisher RA (1925) Statistical methods for research workers, 13th ed. Oliver & Loyd, London, p 99

FitzPatrick DR, Carr IM, McLaren L, Leek JP, Wightman P, Williamson K, Gautier P, McGill N, Hayward C, Firth H, Markham AF, Fantes JA, Bonthron DT (2003) Identification of *SATB2* as the cleft palate gene on 2q32–q33. Hum Molec Genet 12:2491–2501

Greenberg DA, Abreu P (2001) Determining trait locus position from multipoint analysis: accuracy and power of three different statistics. Genet Epidemiol 21:299–314

Hodge SE, Elston RC (1994) Lods, wrods, and mods: the interpretation of lod scores calculated under different models. Genet Epidemiol 11:329–342

Hodge SE, Vieland VJ, Greenberg DA (2002) HLODs remain powerful tools for detection of linkage in the presence of genetic heterogeneity. Am J Hum Genet 70:556–558

Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype-phenotype associations. Eur J Hum Genet 9:301–306

Jezewski P, Vieira AR, Nishimura C, Ludwig B, Johnson M, O'Brien SE, Daack-Hirsch S, Schultz RE, Weber A, Nepomucena B, Romitti PA, Christensen K, Orioli IM, Castilla EE, Machida J, Natsume N, Murray JC (2003) Complete sequencing demonstrates a role for MSX1 in nonsyndromic cleft lip and palate. J Med Genet 40:399–407

Juriloff DM (2002) Mapping studies in animal models. In: Wyszynski DF (ed) Cleft lip and palate: from origin to treatment. Oxford University Press, Oxford, pp 265–282

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Valencia C, Houdayer C, Bahuau M, Moretti-Ferreira D, Richieri-Costa A, Dixon MJ, Murray JC (2002) Mutations in *IRF6* cause Van der Woude and popliteal pterygium syndromes. Nat Genet 32:285–289

Laird NM, Horvath S, Xu X (2000) Implementing a unified approach to family-based tests of association. Genet Epidemiol Suppl 19:36–42

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lange K (2002) Descent graph methods. In: Mathematical and statistical methods for genetic analysis, 2nd ed. Springer-Verlag, New York, pp 167–198

Lehman OJ, Sowden JC, Carlsson P, Jordan T, Bhattacharya SS (2003) Fox's in development and disease. Trends Genet 19:339–344

Levinson DF, Levinson MD, Segurado R, Lewis CM (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder. I. Methods and power analysis. Am J Hum Genet 73: 17–33

Marazita ML (2002*a*) Genetic etiologies of facial clefting. In: Mooney MP, Siegel MI (eds) Understanding craniofacial anomalies: the etiopathogenesis of craniosynostoses and facial clefting. John Wiley and Sons, New York, pp 147–162

Marazita ML (2002*b*) Segregation analyses. In: Wyszynski DF (ed) Cleft lip and palate: from origin to treatment. Oxford University Press, Oxford, pp 222–233

Marazita ML, Field LL, Cooper ME, Tobias R, Maher BS, Peanchitlertkajorn S, Liu Y-e (2002) Genome scan for loci involved in cleft lip with or without cleft palate, in Chinese multiplex families. Am J Hum Genet 71:349–364

Marazita ML, Field LL, Tunçbilek G, Cooper ME, Goldstein T, Gürsu KG (2004) Genome-scan for loci involved in cleft lip with or without cleft palate in consanguineous families from Turkey. Am J Med Genet 126A:111–122

Mitchell LE (1997) Transforming growth factor α locus and nonsyndromic cleft lip with or without cleft palate: a reappraisal. Genet Epidemiol 14:231–240

Mitchell LE, Risch N (1992) Mode of inheritance of nonsyndromic cleft lip with or without cleft palate: a reanalysis. Am J Hum Genet 51:323–332

Moreno L, Arcos-Burgos M, Marazita ML, Cooper ME, Goldstein T, Krahn K, Valencia C, Lidral AC (2003) Genome scan for loci in multiplex nonsyndromic cleft lip families from Colombia and Ohio. Am J Hum Genet Suppl 73:478

Murray JC (2002) Gene/environment causes of cleft lip and/or palate. Clin Genet 61:248–256

Neiswanger K, Ford MD, Cooper ME, Goldstein TH, Petiprin SS, Bardi KM, Brandon CA, Giles RF, Marazita ML (2003) Genome scan of cleft lip with or without cleft palate (CL/P). II. Broadening the phenotype to include velopharyngeal incompetence (VPI). Am J Hum Genet Suppl 73:493

O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266

——— (1995) The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. Nat Genet 11:402–408

Prescott NJ, Lees MM, Winter RM, Malcolm S (2000) Identification of susceptibility loci for nonsyndromic cleft lip with or without cleft palate in a two stage genome scan of affected sib-pairs. Hum Genet 106:345–350

Province MA (2001) The significance of not finding a gene. Am J Hum Genet 69:660–663

Rabinowitz D, Laird N (2000) A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered 50:211–223

Ray AK, Marazita ML, Pathak R, Beever CL, Cooper ME, Goldstein T, Shaw DF, Field LL. TP63 mutation and clefting

modifier genes in an EEC syndrome family. Clin Genet (in press)

Schliekelman P, Slatkin M (2002) Multiplex relative risk and estimation of the number of loci underlying an inherited disease. Am J Hum Genet 71:1369–1385

Schultz RE, Cooper ME, Riley BM, Goldstein T, Daack-Hirsch S, Marazita ML, Murray JC (2003) A two stage genome-wide linkage scan for nonsyndromic cleft lip and palate in 215 Filipino families shows evidence of linkage to 2p21, 6q23, and 8p21. Am J Hum Genet Suppl 73:536

Smith CAB (1963) Testing for heterogeneity of recombination fraction values in human genetics. Ann Hum Genet 27:175–182

Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323–1337

Sobel E, Lange K, O'Connell JR, Weeks DE (1996) Haplotyping algorithms. In: Speed TP, Waterman MS (eds) Genetic mapping and DNA sequencing. IMA volumes in mathematics and its applications. Springer-Verlag, New York, pp 89–110

Sozen MA, Suzuki K, Tolarova MM, Bustos T, Fernandez Iglesias JE, Spritz RA (2001) Mutation of *PVRL1* is associated with sporadic, non-syndromic cleft lip/palate in northern Venezuela. Nat Genet 29:141–142

Terwilliger JD, Ott J (1994) Handbook of human genetic linkage. Johns Hopkins University Press, Baltimore

Trew CJ (1757) Sistens plura exempla palati deficientis. Nova Acta Physico-Medica Academiae Caesarae Leopoldion-Carolinae 1:445–447

Vieland VJ, Wang K, Huang J (2001) Power to detect linkage based on multiple sets of data in the presence of locus heterogeneity: comparative evaluation of model-based linkage methods for affected sib pair data. Hum Hered 51:199–208

Weinberg SM, Neiswanger K, Mooney MP, Cooper ME, Goldstein TH, Bowen A, Martin RA, Marazita ML (2003) Genome scan of cleft lip with or without cleft palate (CL/P). I. Broadening the phenotype to include lip muscle defects. Am J Hum Genet Suppl 73:291

Wise LH, Lanchbury JS, Lewis CM (1999) Meta-analysis of genome searches. Ann Hum Genet 63:263–272

Wyszynski DF, Albacha-Hejazi H, Aldirani M, Hammod M Shkair H, Karm A, Alaskar J, Holms TN, Pugh EW, Doheny KF, McIntosh I, Beaty TH, Bailey-Wilson JE (2003) A genome-wide scan for loci predisposing to non-syndromic cleft lip with or without cleft palate in two large Syrian families. Am J Med Genet 123A:140–147

Yoshiura K, Machida J, Daack-Hirsch S, Patil SR, Ashworth LK, Hecht JT, Murray JC (1998) Characterization of a novel gene disrupted by a balanced chromosomal translocation t(2;19)(q11.2;q13.3) in a family with cleft lip and palate. Genomics 54:231–240

Yuan B, Vaske D, Weber JL, Beck J, Sheffield VC (1997) Improved set of short-tandem repeat polymorphisms for screening the human genome. Am J Hum Genet 60:459–460

Zeiger JS, Hetmanski JB, Beaty TH, VanderKolk CA, Wyszynski DF, Bailey-Wilson JE, Ortiz de Luna R, Perandones C, Tolarova MM, Mosby T, Bennun R, Segovia M, Calda P, Pugh EW, Doheny K, McIntosh I (2003) Evidence for linkage of nonsyndromic cleft lip with or without cleft palate to a region on chromosome 2. Eur J Hum Genet 11:835–839

Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, et al Interferon regulatory factor 6 (*IRF6*) gene variants confer risk for isolated cleft lip and palate. New Engl J Med (in press)