19390095) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) and by Solution-oriented Research of Science and Technology (SORST) from Japan Science and Technology Agency (JST). K.Y. was supported in part by a Grant-in-Aid for Scientific Research (on Priority Areas – Transportsome, No. 18059027) from MEXT. The authors are particularly grateful to the following clinicians: Drs. Katsumi Tsuneda, Minoru Arakaki, Keiko Onizuka, Yoshimi Noda, Syogo Kaji, Junji Tsuruta, Hiroshi Tsuchida, Hiroshi Mori, Hiroshi Yoshimoto, Masaki Fujioka, Kayoko Kondo, Daiki Nonaka, and Kenji Hayashida, for providing samples and clinical information.

References

1. Emter R, Natsch A: **The sequential action of a dipeptidase and a beta-lyase is required for the release of the human body odorant 3-methyl-3-sulfanylhexan-1-ol from a secreted Cys-Gly-(S) conjugate by Corynebacteria.** *J Biol Chem* 2008, **283:**20645-20652.
2. Matsunaga E: **The dimorphism in human normal cerumen.** *Ann Hum Genet* 1962, **25:**273-286.
3. Petrakis NL, Molohon KT, Tepper DJ: **Cerumen in American Indians: Genetic implications of sticky and dry types.** *Science* 1967, **158:**1992-1993.
4. Petrakis NL, Pingke U, Petrakis SJ, Petrakis SL: **Evidence for a genetic cline in cerumen type in the Middle East and Southeast Asia.** *Am J Phys Anthropol* 1971, **35:**141-144.
5. Omoto K: **Polymorphic traits in peoples of eastern Asia and the Pacific.** *Isr J Med Sci* 1973, **9:**1195-1215.
6. Yoshiura K, Kinoshita A, Ishida T, Ninokata A, Ishikawa T, Kaname T, Bannai M, Tokunaga K, Sonoda S, Komaki R, Ihara M, Saenko VA, Alipov GK, Sekine I, Komatsu K, Takahashi H, Nakashima M, Sosonkina N, Mapendano CK, Ghadami M, Nomura M, Liang DS, Miwa N, Kim DK, Garidkhuu A, Natsume N, Ohta T, Tomita H, Kaneko A, Kikuchi M, Russomando G, Hirayama K, Ishibashi M, Takahashi A, Saitou N, Murray JC, Saito S, Nakamura Y, Niikawa N: **A SNP in the *ABCC11* gene is the determinant of human cerumen type.** *Nature Genet* 2006, **38:**324-330.
7. Meyer zum.Gottesberge A, Meyer zum Gottesberge AM: **Cerumen from the anthropologic viewpoint.** *Laryngorhinootologie* 1995, **74:**50-53.
8. Yabuuchi H, Shimizu H, Takayanagi S, Ishikawa T: **Multiple splice variants of two human ATP-binding cassette transporters, *ABCC11* and *ABCC12*.** *Biochem Biophys Res Commun* 2001, **288:**933-939.
9. Tammur J, Prades C, Arnould I, Rzhetsky A, Hutchinson A, Adachi M, Schuetz JD, Swoboda KJ, Ptacek LJ, Rosier M, Dean M, Allikmets R: **Two new genes from the human ATP-binding cassette transporter superfamily, *ABCC11* and ABCC12 tandemly duplicated on chromosome 16q12.** *Gene* 2001, **273:**89-96.
10. Bera TK, Lee S, Salvatore G, Lee B, Pastan I: **MRP8, a new member of ABC transporter superfamily, identified by EST database mining and gene prediction program, is highly expressed in breast cancer.** *Mol Med* 2001, **7:**509-516.
11. Kruh GD, Belinsky MG: **The MRP family of drug efflux pumps.** *Oncogene* 2003, **22:**7537-7552.
12. Chen ZS, Guo Y, Belinsky MG, Kotova E, Kruh GD: **Transport of bile acids, sulfated steroids, estradiol 17-β-D-glucuronide, and leukotriene C4 by human multidrug resistance protein 8 (ABCC11).** *Mol Pharmacol* 2005, **67:**545-557.
13. Bortfeld M, Rius M, König M, Herold-Mende C, Nies AT, Keppler D: **Human multidrug resistance protein 8 (MRP8/ABCC11), an apical efflux pump for steroid sulfates, is an axonal protein of the CNS and peripheral nervous system.** *Neuroscience* 2006, **137:**1247-1257.
14. Zeng XN, Leyden JJ, Lawley HJ, Sawano K, Nohara I, Preti G: **Analysis of characteristic odors from human male axillae.** *J Chem Ecol* 1991, **17:**1469-1492.
15. Natsch A, Gfeller H, Gygax P, Schmid J, Acuna G: **A specific bacterial aminoacylase cleaves odorant precursors secreted in the human axilla.** *J Biol Chem* 2003, **278:**5718-5727.
16. Natsch A, Derrer S, Flachsmann F, Schmidt J: **A broad diversity of volatile carboxylic acids, released by a bacterial amynoacylase from axilla secretions, as candidate molecules for the determination of human-body odor type.** *Chem Biodiv* 2006, **3:**1-20.
17. Toyoda Y, Sakurai A, Mitani Y, Nakashima M, Yoshiura K, Nakagawa H, Sakai Y, Ota I, Lezhava A, Hayashizaki Y, Niikawa N, Ishikawa T: **Earwax, osmidrosis, and breast cancer: Why does one SNP (538G>A) in the human ABC transporter *ABCC11* gene determine earwax type?** *FASEB J* 2009, **23:**2001-2013.
18. Wigginton JE, Cutler DJ, Abecasis GR: **A note on exact tests of Hardy-Weinberg equilibrium.** *Am J Hum Genet* 2005, **76:**887-893.
19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC: **PLINK: a toolset for whole-genome association and population-based linkage analysis.** *Am J Hum Genet* 2007, **81:**559-575.

Searching for Genes for Cleft Lip and/or Palate Based on Breakpoint Analysis of a Balanced Translocation t(9;17)(q32;q12)

Junichiro Machida, D.D.S., Ph.D., Têmis M. Félix, M.D., Ph.D., Jeffrey C. Murray, M.D., Koh-ichiro Yoshiura, M.D., Ph.D., Mitsuyo Tanemura, M.D., Ph.D., Munefumi Kamamoto, D.D.S., Kazuo Shimozato, D.D.S., Ph.D., Shin-ichi Sonta, Ph.D., Takao Ono, Ph.D.

Objective: Identification of the breakpoints of disease-associated chromosome rearrangements can provide informative clues to a positional cloning approach for genes responsible for inherited diseases. Recently, we found a three-generation Japanese family segregating balanced chromosome translocation t(9;17)(q32;q12). One of the subjects had cleft lip and palate. We examined whether regions near the breakpoint could be associated with cleft lip and/or palate.

Methods: We determined the breakpoints involved in the translocation by fluorescence *in situ* hybridization analysis and subsequent long-range polymerase chain reaction. In order to study the role of these disrupted regions in nonsyndromic cleft lip and/or palate, we performed mutation analysis and a haplotype-based transmission disequilibrium test using tagging single-nucleotide polymorphisms in the flanking regions of the breakpoints in white and Filipino nonsyndromic cleft lip and/or palate populations.

Results: Sequence analysis demonstrated that two genes, *SLC31A1* (solute carrier family 31 member 1) on chromosome 9 and *CCL2* (chemokine ligand 2) on chromosome 17, were rearranged with the breaks occurring within their introns. It is interesting that *SLC31A1* lies closed to *BSPRY* (B-box and SPRY domain), which is a candidate for involvement with cleft lip and/or palate. Some of the variants in *BSPRY* and *CCL2* showed significant *p* values in the cleft lip and/or palate population compared with the control population. There was also statistically significant evidence of transmission distortion for haplotypes on both chromosomes 9 and 17.

Conclusions: The data support previous reports that genes on chromosomal regions of 9q and 17q play an important role in facial development.

KEY WORDS: *association analysis, balanced chromosomal translocation,* BSPRY, CCL2, *cleft lip and palate, haplotype, HapMap,* SLC31A1

Nonsyndromic cleft lip and/or palate (CL/P) is a common congenital anomaly of complex etiology, with the birth prevalence being approximately 1 per 700 live births (Mossey and Little, 2002; Murray, 2002). Previous family and population studies indicate that both genetic and environmental factors are involved in the occurrence of this malformation. Prior sequencing analysis of CL/P has indicated roles for mutations in *MSX1*, *FOXE1*, *GLI2*, *MSX2*, *SKI*, *SPRY2*, *RYK*, *FGF1*, *FGF2*, and *FGF8* in the etiology of the birth defect (Ichikawa et al., 2006; Jezewski et al., 2003; Riley et al., 2007; Vieira et al., 2005). In addition, polymorphisms in *IRF6* have been found to be associated with CL/P (Blanton et al., 2005; Ghassibe et al., 2005; Park et al., 2007; Scapoli et al., 2005; Zucchero et al., 2004).

Dr. Machida is Associate Professor, Department of Maxillofacial Surgery, School of Dentistry, Aichi-Gakuin University, Nagoya, Japan. Dr. Félix is Clinical Geneticist, Medical Genetics Service of Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil. Dr. Murray is Professor, Department of Pediatrics, University of Iowa, Iowa City, Iowa. Dr. Yoshiura is Professor, Department of Human Genetics, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan. Dr. Tanemura is Associate Professor, Department of Obstetrics and Gynecology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan. Dr. Kamamoto is postgraduate student and Dr. Shimozato is Professor, Department of Maxillofacial Surgery, School of Dentistry, Aichi-Gakuin University, Nagoya, Japan. Dr. Sonta is Chief Scientist, Division of Cytogenetics, Department of Genetics, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan. Dr. Ono is Senior Research Scientist, Chromosome Dynamics Laboratory, RIKEN Discovery Research Institute, Wako, Saitama, and Senior Researcher, Division of Clinical Genetics, Department of Genetics, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan.

Funding was provided by the Grant-in-Aid for Young Scientists of Japan Society for the Promotion of Science 16791271 and National Institutes of Health grants DE08559 (J.C.M.) and DE016215 (J.C.M.).

Submitted August 2008; Accepted January 2009.

Address correspondence to: Dr. Junichiro Machida, Department of Maxillofacial Surgery, School of Dentistry, Aichi-Gakuin University, 2-11 Suemori-dori, Chikusa-ku, Nagoya 464-8651, Japan. E-mail jmachida@dpc.aichi-gakuin.ac.jp.

DOI: 10.1597/08-047.1

Because clefts are complex and caused by multiple interacting genes, we do not expect that single gene disruptions in all individuals with a chromosomal break will manifest as CL/P. However, in some cases, identification of the breakpoints of disease-associated chromosome translocations can be an effective way to identify genes and/or chromosomal regions contributing to the occurrence of CL/P. The positional cloning approach for disease-associated genes using chromosome rearrangements has successfully identified four relevant genes (*CLPTM1* [Yoshiura et al., 1998], *SATB2* [FitzPatrick et al., 2003], *SUMO1* [Alkuraya et al., 2006], and *FGFR1* [Kim et al., 2005]) as candidates for CL/P. These findings suggest that identification of the genes disrupted in a cleft/translocation case can give us insight into some of the pathways involved in this birth defect and provide candidates for analysis in cases independent of the index family.

Recently, we identified a Japanese family segregating a balanced chromosome translocation t(9;17)(q32;q12). Of three generations we examined, one individual is affected with cleft lip and palate. Genome-wide linkage studies have suggested that the region 9q21 has significant linkage with CL/P (Marazita et al., 2004). Therefore, genes in this region of chromosome 9 may have an important role in facial development. In this study, we identified the breakpoints of t(9;17)(q32;q12) in the patient by fluorescence *in situ* hybridization (FISH) analysis and identified candidate genes by polymerase chain reaction (PCR) and DNA sequencing. We then carried out a mutation search and a case-control association study of candidate genes. In addition, extended single-nucleotide polymorphism (SNP) analysis of the chromosomal regions adjacent to the translocation breakpoint was performed.

Materials and Methods

Clinical Report

The proband was a male infant born prematurely by cesarean section at 24 weeks gestation after an uneventful pregnancy. The maternal age was 25 years and paternal age 26 years at the time of his birth. The parents were nonconsanguineous, and there was no family history of malformations. His birth weight and height were 706 g and 31.8 cm, respectively (10th centile = 560 g, 90th centile = 774 g). The patient was in normal development at 24 weeks gestation. The circumferences of his head and chest were 23 cm and 21 cm at birth, respectively. He had apparent ocular hypertelorism, a bilateral cleft lip and palate, and a mild nasal flattening. Echocardiography showed a small atrial septal defect, which had closed by 2 years of age. At 6 years after birth, his weight and height were normal for age at 20.1 kg and 114.6 cm, respectively. He shows no evidence of developmental delay. Routine chromosome-banding analysis of the patient revealed an apparently balanced translocation between the long arm of chromosome 9 and chromosome 17: 46,XY, t(9;17)(q32;q12). His father and paternal grandmother had the same translocation seen in the patient; whereas, his mother had a normal karyotype (Fig. 1). Neither of the other translocation carriers had a cleft or malformations, and there was no family history of CL/P.

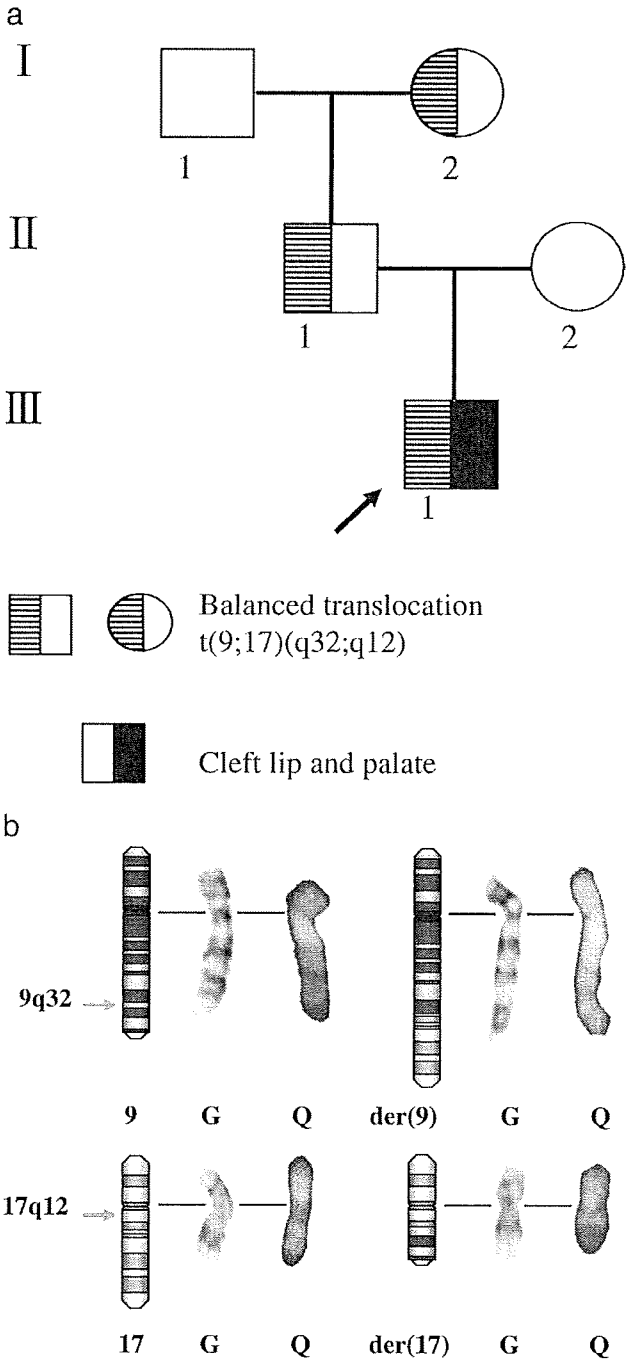

FIGURE 1 Case presentation. a: Pedigree of the family examined. b: Chromosomal breakpoints in the patient (III-1). Ideogram and a partial karyotype are shown. Chromosomal breakpoints are indicated by arrows.

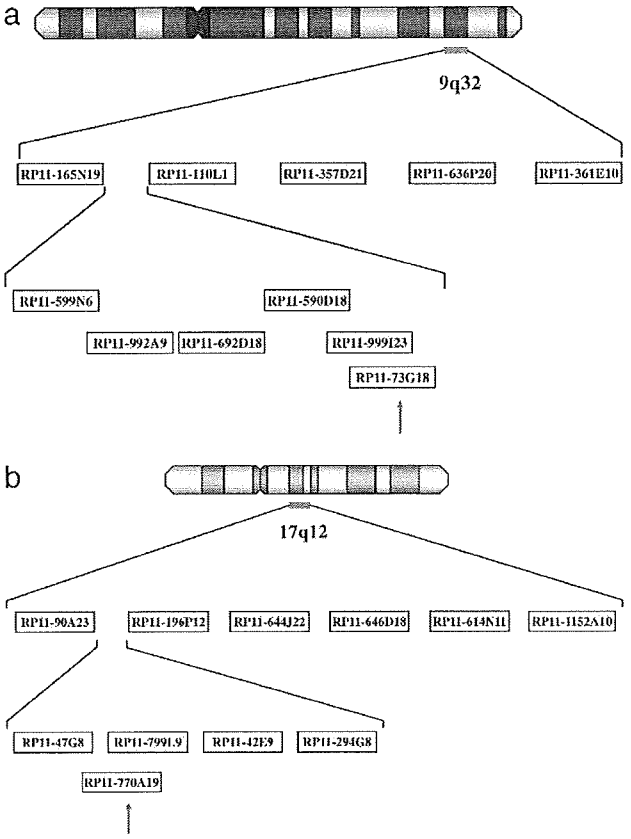

FIGURE 2 BAC clones used for FISH analysis. a: A physical map of BAC clones covering the 9q32 subchromosomal band. b: A physical map of BAC clones covering the 17q12 subchromosomal band.

Identification of Translocation Breakpoint

Fluorescence In Situ *Hybridization Analysis*

Breakpoint mapping of the translocation in the patient was initiated by FISH analysis, according to published procedures (Ono et al., 1997). Bacterial artificial chromosome (BAC) clones were identified using the genome maps provided by the National Center for Biotechnology Information (http://www.ncbi.nlm.org) and University of California, Santa Cruz (UCSC) (http://genome.cse.ucsc.edu) genome browsers. Figure 2 shows the BAC clones on chromosomes 9 and 17 used in the present study.

Cloning of the Translocation Breakpoint

We carried out a series of long-range PCR amplifications of DNA fragments to detect the recombinant sequence caused by the translocation. Primer sequences were designed based on the FISH results, with one forward primer on chromosome 9 and one reverse primer on chromosome 17, and were evaluated by RepeatMasker (http://www.repeatmasker.org/) to avoid nonspecific amplification of repetitive sequence. The PCR products then were subcloned into the pCR2.1-TOPO vector (Invitrogen, Carlsbad, CA). Additional nested primers were designed to obtain a DNA fragment including the breakpoint on der(9) using the subcloned plasmid DNA from the patient and genomic DNA of members of his family.

DNA Sequencing

The PCR products were labeled with the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit according to the manufacturer's instruction (Applied Biosystems, Carlsbad, CA). The products were purified, and then analyzed on an ABI 3700 automated sequencer (Applied Biosystems). The Applied Biosystems sequence software (version 2.1.2) was used for lane tracking. Chromatograms were transferred to a UNIX workstation, base-called with PHRED 4.0 (http://droog.mbt.washington.edu/poly_doc40.html) and results viewed with the CONSED program (version 4.0) (Nickerson et al., 1997). The UCSC and the Ensembl databases (http://www.ensembl.org) were used to detect potential genes around the breakpoint.

Mutation Search of Candidate Genes in Case/Control CL/P Population

To examine the association of the disrupted genes with the occurrence of CL/P in individuals unrelated to this family, we screened for mutations of the untranslated regions (UTR), exons and exon-intron boundaries of the two rearranged genes on chromosome 9 and chromosome 17. In addition, any plausible candidate gene adjacent to the breakpoint was screened. We sequenced 90 individuals affected with isolated CL/P from the Philippines and 90 from Iowa. The control group comprised 90 samples from unrelated Filipinos and whites provided by Centre d'Etudes du Polymorphisme Humaine (Dausset et al., 1990). Standard chi-square tests of association were used to compare the frequencies of each variant found by sequencing between groups of patients and controls.

Haplotype Analysis Adjacent to the Translocation Breakpoint

To determine whether the chromosomal regions adjacent to the breakpoint could play a role in the development of CL/P, we carried out the haplotype-based transmission disequilibrium test. Four tagging SNP markers (rs974230, rs1330691, rs4596714, rs3750534) adjacent to the breakpoint on chromosome 9 and three markers (rs16561, rs725276, rs1029719) on chromosome 17 were selected, based on the HapMap (http://www.hapmap.org/index.html.en) database. The TaqMan genotyping for these SNPs was performed with the ABI PRISM 7900 Sequence Detection System (Applied Biosystems) using 371 case/parent families from the Philippines and 206 from Iowa. Haplotype-based transmission disequilibrium statistics

were calculated using the software FBAT (Horvath et al., 2001; Rabinowitz and Laird, 2000).

Bioethics Approval

Committee approval was obtained from both the ethical committee at Toyohashi Municipal Hospital and Aichi-Gakuin University to obtain whole blood from the patient, his parents, and his paternal grandparents after signed informed consent. For the association analysis, patients were identified at the University of Iowa Hospital and Clinics and in the Philippines through Operation Smile (Murray et al., 1997). Blood samples were obtained with informed consent following institutional review board (IRB) approval in the United States (University of Iowa, IRB Committee) and in the Philippines (Hope Foundation, Bacolod City, Negros, Philippines).

Results

Identification of the Translocation Breakpoint

We found that two clones, RP11-73G18 (164,488 base pairs [bp]) and RP11-999I23 (160,162 bp), included the breakpoint region, judging from a hybridization pattern in which FISH signals were detected simultaneously on the normal chromosome 9 as well as both der(9) and der(17) chromosomes (Fig. 3a). The breakpoint on chromosome 17 was covered by two clones, RP11-770A19 (194,408 bp) and RP11-799L9 (178,732 bp), according to the hybridization pattern in which FISH signals were detected simultaneously on the normal chromosome 17 as well as both der(9) and der(17) chromosomes (Fig. 3b).

To examine sequences of the translocation breakpoints, we tried PCR amplification involving the breakpoint region (Fig. 4a). A primer set of F2 (5′-TCTTAACTTACCTGTTGCTGCCGTTCCCTG-3′) derived from chromosome 9 and R1 (5′-TGGGCTGAGATTGTGGTAACCGTAGGAAAG-3′) derived from chromosome 17 successfully amplified a DNA fragment of approximately 3 kbp in length. The PCR product amplified by 9f1 (5′-GTGTATGATACCATAACAATTAC-3′) and 17r1 (5′-GCACTGAGCAAACAGGAAGTG-3′) was sequenced after subcloning into the pCR2.1-TOPO vector. The nucleotide sequence of this PCR product demonstrated that the breakpoint of chromosome 9 lies in intron 1 of *SLC31A1* (solute carrier family 31 member 1; GenBank accession number NM001859, between nucleotides 113,072,186 and 113,072,187) according to the UCSC database (hg17). The breakpoint on chromosome 17 is assigned to the 5′UTR of chemokine ligand 2 (*CCL2*; GenBank accession number MN002982, between nucleotides 29,568,964 and 29,568,965). These three exons also form exons 4, 5, and 6 of a predicted six-exon gene of which *CCL2* may be one isoform, according to the Ensembl

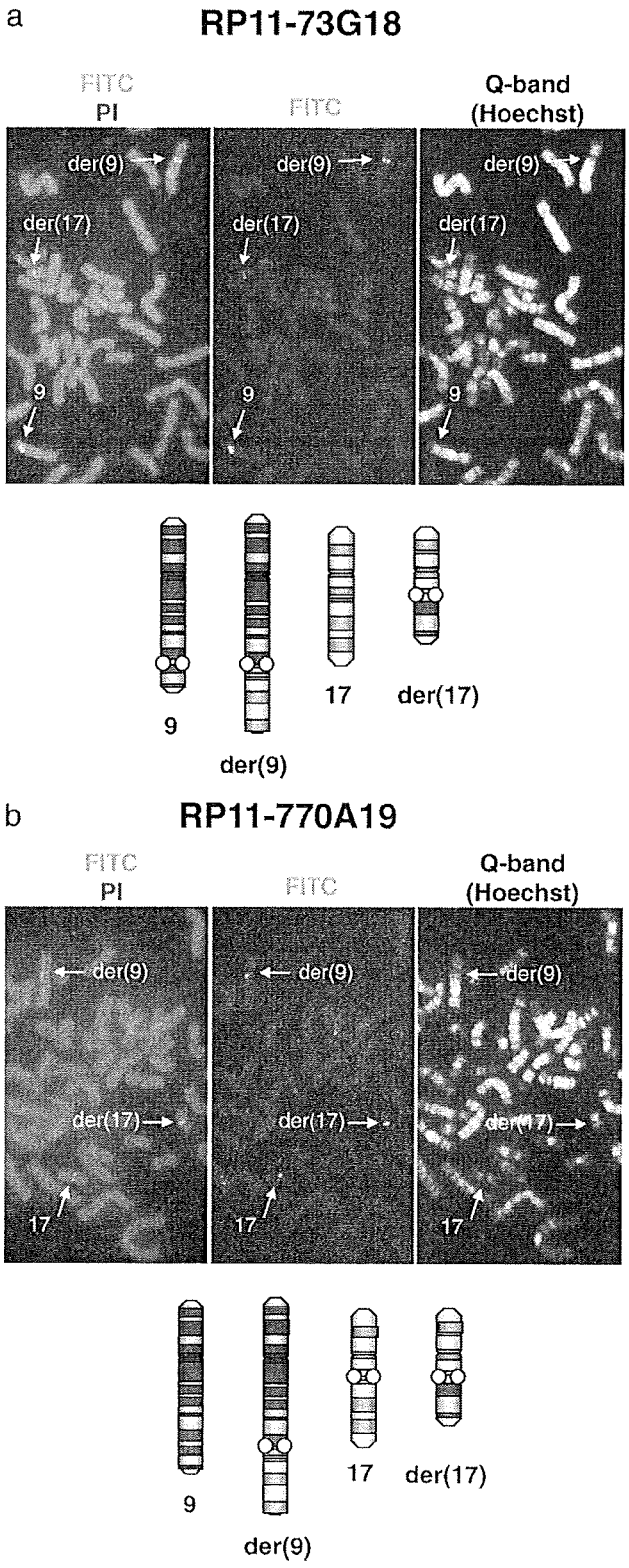

FIGURE 3 Molecular cytogenetic characterization of the t(9;17)(q32;q12). Each left panel shows FISH signals (FITC) on propidium iodide-stained chromosomes (PI). Each right panel exhibits Q-banded chromosomes (Hoechst). The RP11-73G18 clone simultaneously hybridized to normal 9, der(9), and der(17) chromosomes. The RP11-770A19 clone simultaneously hybridized to normal 17, der(9), and der(17) chromosomes. Bottom panels show schematic diagrams of FISH results.

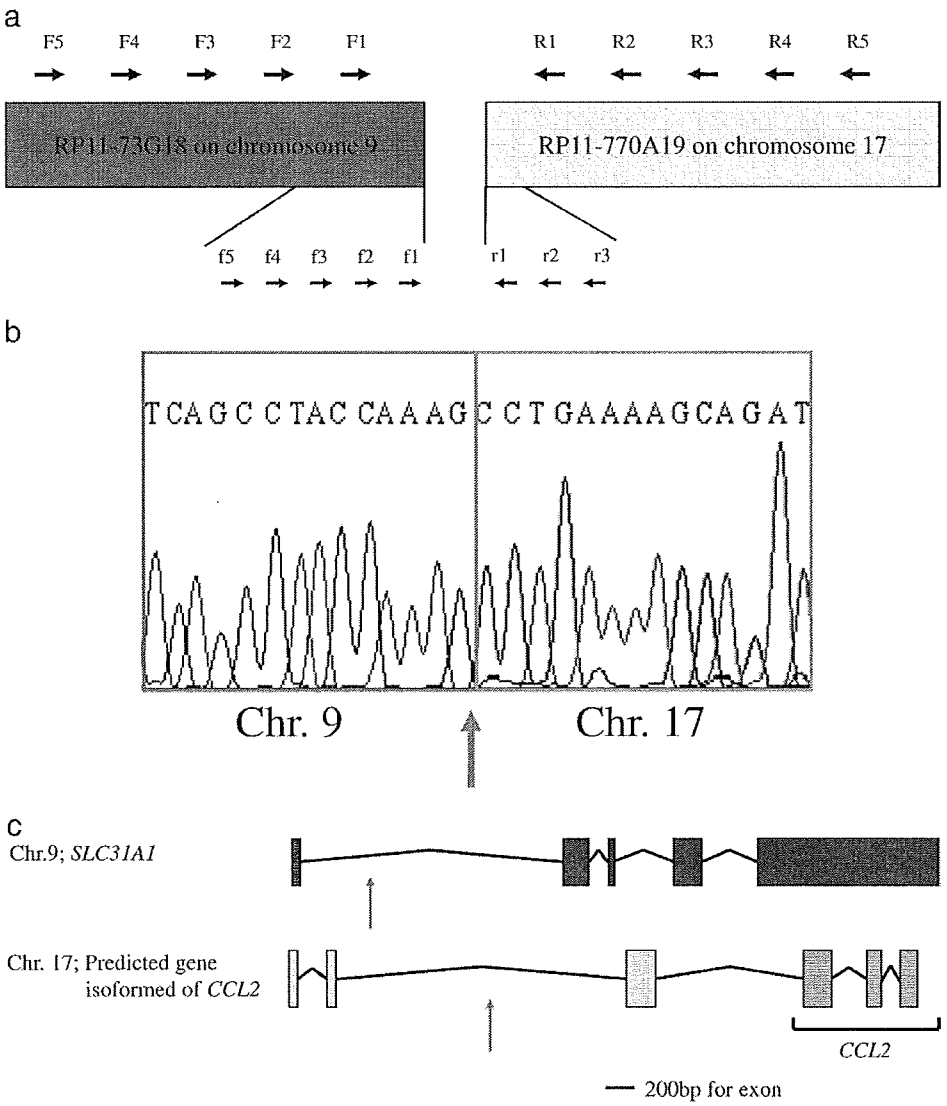

FIGURE 4 Fine mapping of the breakpoints. a: Primers used for long-range PCR. Forward and reverse primers were designed every 5 kbp of RP11-73G18 on chromosome 9 and RP11-770A19 on chromosome 17, respectively (arrows in upper). Additional primers flanked regions of the breakpoint (arrows in lower). b: Identification of the breakpoint on der(9), using a PCR product with a primer set of 9f1 and 17r1. c: The breakpoint on chromosome 9 is located between exons 1 and 2 of *SLC3A1*. The breakpoint on chromosome 17 is located between exons 2 and 3 of a predicted gene, according to the Ensembl database. Arrows indicate the positions of the breakpoints.

database (Transcript ID; GENSCAN00000016010). The function of this predicted gene is unknown. The translocation examined in this study shows no gain or loss of nucleotides at the breakpoint, confirming a perfectly balanced reciprocal translocation in translocation in the proband, his father, and his paternal grandmother (Fig. 4b and 4c).

Mutation Analysis of Candidate Genes in Case/Control CL/P Population

In order to determine whether genes at or near the breakpoint play a role in CL/P, we screened for mutations in three candidate genes: *SCL31A1* and *BSPRY* on chromosome 9 and the predicted gene that includes exons of *CCL2* on chromosome 17. Although *BSPRY* is located approximately 120 kbp downstream from the breakpoint, we considered it a plausible candidate because it is induced by the fibroblast growth factor receptor (*FGFR*) pathway (Welsh et al., 2007) and impaired FGF signaling has been found to be associated with 3% to 5% of CL/P (Riley et al., 2007). In addition, variants in a related gene, *SPRY2*, that also has a sprouty domain, already have been identified in CL/P (Vieira et al., 2005). We hypothesized that disruption of regulatory elements of *BSPRY* might lead to a cleft phenotype.

Table 1 shows nucleotide variants found in *BSPRY* and the predicted gene that include *CCL2*. In *BSPRY*, two

TABLE 1 Polymorphisms Found (Entries in the Table are Numbers of Individuals With Homozygous Common Allele/Heterozygous Rare Allele/Homozygous Rare Allele)

	Filipino Genotyping Distributions			*White Genotyping Distributions*			
Variant (Position)*	*Case (AF†)*	*Control (AF)*	*p Value*	*Case (AF)*	*Control (AF)*	*p Value*	*Comment*
BSPRY							
new (113202320)	62/0/0 (1.000)	68/5/0 (0.966)	.012	62/0/0 (1.000)	55/0/0 (1.000)	~.0	
rs752757 (113202508)	41/19/2 (0.814)	33/32/6 (0.690)	.058	55/22/2 (0.835)	25/7/1 (0.864)	0.756	
new (113202593)	49/2/0 (0.980)	65/3/0 (0.978)	.895	75/1/0 (0.993)	25/0/0 (1.000)	.449	
rs2296074 (113210279)	45/23/1 (0.819)	33/18/0 (0.825)	.565	42/4/0 (0.957)	33/6/0 (0.923)	.341	
Predicted gene (GENESCAN00000016010) containing *CCL2*							
rs4795884 (29527481)	52/32/6 (0.756)	42/36/1 (0.747)	.103	17/47/17 (0.500)	24/40/19 (0.530)	.396	
new (29527634)	82/0/0 (1.000)	79/0/0 (1.000)	~.0	74/6/0 (0.963)	57/4/0 (0.967)	.828	R15C‡
rs4795885 (29527655)	83/2/0 (0.988)	79/0/0 (1.000)	.103	53/23/3 (0.816)	38/20/3 (0.787)	.830	
new (29540781)	48/30/4 (0.768)	16/12/4 (0.688)	.364	60/19/3 (0.848)	53/25/2 (0.819)	.489	M33T‡
new (29540817)	79/6/0 (0.965)	69/0/0 (1.000)	.007	57/21/4 (0.823)	53/29/1 (0.813)	.198	N45S‡
new (29540841)	86/0/0 (1.000)	69/0/0 (1.000)	~.0	81/1/0 (0.994)	84/0/0 (1.000)	.234	T53M‡
new (29541048)	86/0/0 (1.000)	69/0/0 (1.000)	~.0	69/11/0 (0.931)	75/7/1 (0.981)	.290	
rs1860188 (29602669)	31/21/3 (0.755)	35/21/2 (0.784)	.833	49/12/2 (0.873)	22/13/0 (0.814)	.073	
new (29602875)	54/4/1 (0.949)	47/5/0 (0.952)	.463	63/0/0 (1.000)	73/0/0 (1.000)	~.0	
rs28730833 (29606607)	64/1/0 (0.992)	80/0/0 (1.000)	.204	56/0/0 (1.000)	80/0/0 (1.000)	~.0	
rs4586 (29607382)	10/19/8 (0.527)	11/27/14 (0.471)	.755	20/30/5 (0.636)	5/9/4 (0.528)	.363	

* Position according to UCSC database (hg17).
† AF = allele frequency.
‡ Amino acid change in predicted gene containing *CCL2* according to the Ensembl database (Transcript ID; GENSCAN00000016010).

common polymorphisms (rs752757 and rs2296074) and two new variants (positions 113202320 and 113202593) were detected. One variant (position 113202320) shows a significant p value (.012) in the Filipino CL/P population compared with matched controls. In the predicted gene containing *CCL2*, six new variants were detected. Four variants (positions 29527634, 29540781, 29540817, and 29540841) possibly result in a change of amino acid, R15C, M33T, N45S, and T53M, respectively, according to the Ensembl database (Transcript ID; GENSCAN-00000016010). R15C, N45S, and T53M were not found in Filipino controls, and T53M was not found in either Filipino or white controls. Five known variants were also identified (rs4795884, rs4795885, rs1860188, rs28730833, and rs4586). The variant at position 29540817 was present with a significant p value (.007) in an affected Filipino population compared with control population. No variants were identified in *SLC31A1* (data not shown).

Haplotype Analysis in Markers Adjacent to the Translocation Breakpoint

Although no statistically significant evidence of transmission distortion was seen for either SNP individually, there was statistically significant evidence of transmission distortion for haplotype h2-F9 (C-G-T-T; bi-allelic p value = .017) and haplotype h2-C9 (G-A-C-T; biallelic p value = .028) on chromosome 9 region (Table 2). We also confirmed statistically significant evidence for haplotype h5-F17 (G-G-T; biallelic p value = .042) and haplotype h8-F17 (G-A-A; biallelic p value = .017) on chromosome 17 region (Table 3). These p values, however, did not reach the level of significance, if we account for the number of tests done and use the conservative Bonferroni correction (0.0071).

Discussion

It has been demonstrated that CL/P is associated with trisomy 9q (Metzke-Heidemann et al., 2004), tetrasomy 9q (Wyandt et al., 2000), and chromosomal translocations involving the 9q region: t(6;9)(p23;q22.3) (Donnai et al., 1992); 46,XX,−22,+der(9)t(9;22)(q22;q11.2) (Pivnick et al., 1990); and 46,X,t(X;9)(p22.1;q32) (Zori et al., 1993). Naritomi et al. (1989) divided trisomy 9q syndrome into four groups (group 1: trisomy for 9q11>q32, group 2: 9q32>qter, group 3: 9q34, and group 4: 9q13/q21>qter) according to the length of the trisomic segment. Of these, the patients in groups 2 and 4 showed cleft palate, suggesting that the 9q32 segment may have important roles for facial development. Genome-wide linkage studies have revealed that the region 9q21 has significant linkage with facial clefting (Marazita et al., 2004). A number of candidate genes for CL/P in this region, 9q21–23, have been reported previously (*ROR2*, *PTCH*, *FOXE1*, and *TGFBR1*) (Ichikawa et al., 2006; Loeys et al., 2006; Mansilla et al., 2006; Mizuguchi et al., 2004; van Bokhoven et al., 2000; Vieira et al., 2005). Using a positional cloning approach, we demonstrated that *SLC31A1* on chromosome 9 was disrupted within its intron. The *SLC31A1* encodes a copper transporter and is itself not a strong candidate for involvement in facial development; although, unbalanced copper regulation can disturb cell metabolism (Lee et al., 2001). However, it is increasingly recognized that point mutations in regulatory elements located far from gene structural elements can be damaging. For example, mutations located up to ~1 Mbp from the gene *SHH* are capable of causing congenital abnormalities (Gurnett et al., 2007; Lettice et al., 2003) and campomelic dysplasia has resulted from a balanced translocation breakpoint 1.3 Mbp

TABLE 2 Results of Haplotype-based Transmission Disequilibrium Test for Single-Nucleotide Polymorphisms (SNPs) Around the Breakpoint on Chromosome 9

	SNP					
Filipino Haplotype	*a**	*b*	*c*	*d*	*Estimated Frequency*	*p Values (Biallelic)*
h1-F9	C	A	T	T	0.282	.456 (−)
h2-F9	C	G	T	T	0.246	.017 (+)
h3-F9	C	G	C	T	0.206	.095 (−)
h4-F9	C	A	T	C	0.053	.809 (+)
h5-F9	C	G	T	C	0.052	.892 (+)
h6-F9	G	A	T	T	0.047	.638 (−)
h7-F9	C	G	C	C	0.031	.093 (−)
h8-F9	G	G	C	T	0.027	.367 (+)
h9-F9	C	A	C	T	0.015	.594 (+)
h10-F9	G	G	T	T	0.013	.815 (+)
h11-F9	G	A	T	C	0.010	.956 (−)
h12-F9	G	G	C	C	0.010	.476 (+)
	SNP					
White Haplotype	*a*	*b*	*c*	*d*	*Estimated Frequency*	*p Values (Biallelic)*
h1-C9	G	A	T	T	0.137	.111 (−)
h2-C9	G	A	C	T	0.121	.028 (−)
h3-C9	G	A	T	C	0.110	.441 (+)
h4-C9	C	A	C	T	0.102	.342 (−)
h5-C9	G	A	C	C	0.087	.224 (+)
h6-C9	C	A	T	T	0.085	.122 (+)
h7-C9	G	G	T	C	0.065	.103 (+)
h8-C9	G	G	T	T	0.062	.624 (+)
h9-C9	G	G	C	T	0.059	.246 (+)
h10-C9	C	G	T	T	0.056	.198 (−)
h11-C9	C	G	C	T	0.039	.819 (+)

* a = rs1330692 C/G; b = rs3750534 A/G; c = rs4596714 C/T; d = rs974230 C/T.

downstream of the *SOX9* gene (Velagaleti et al., 2005). In the present report, although we did not detect any mutations in the *SLC31A1* gene itself, we did detect a significant *p* value for CL/P in the case-control study of a strong candidate for craniofacial development, the *BSPRY* gene, which is located 120 kbp downstream from the breakpoint (Table 1). Thus the translocation break may disrupt not the gene itself but regulatory regions. Moreover, there was significant evidence of transmission distortion for haplotypes on chromosome 9 (Table 2). In summary, it is possible that this region of chromosome 9q contains one or more genes playing a role in facial development.

TABLE 3 Results of Haplotype-Based Transmission Disequilibrium Test for Single-Nucleotide Polymorphisms (SNPs) Around the Breakpoint on Chromosome 17

	SNP				
Filipino Haplotype	*e**	*f*	*g*	*Estimated Frequency*	*p Values (Biallelic)*
h1-F17	A	G	T	0.322	.451 (+)
h2-F17	A	A	T	0.316	.991 (+)
h3-F17	A	A	A	0.103	.782 (+)
h4-F17	A	G	A	0.102	.973 (+)
h5-F17	G	G	T	0.065	.042 (−)
h6-F17	G	A	T	0.054	.450 (−)
h7-F17	G	G	A	0.025	.765 (−)
h8-F17	G	A	A	0.013	.017 (+)
	SNP				
White Haplotype	*e*	*f*	*g*	*Estimated Frequency*	*p Values (Biallelic)*
h1-C17	A	A	A	0.180	.275 (+)
h2-C17	A	A	T	0.164	.592 (+)
h3-C17	G	A	T	0.136	.597 (+)
h4-C17	G	A	A	0.120	.684 (−)
h5-C17	A	G	A	0.115	.899 (+)
h6-C17	G	G	T	0.112	.547 (−)
h7-C17	A	G	T	0.099	.534 (−)
h8-C17	G	G	A	0.075	.161 (−)

* e = rs16561 A/G; f = rs1029719 A/G; g = rs725276 A/T.

It is also possible that facial development may be influenced by genes on chromosome 17q. Case reports of chromosome translocations involving 17q24 (Czako et al., 2004; Luke et al., 1992), 17q25 (Bridge et al., 1985), 17q21.1 (Martinet et al., 2006), and 17q23.3 (Stalker et al., 2001) are associated with CL/P. Additively, genome-wide linkage analysis also shows significant association of the 17q21 region to the occurrence of CL/P (Marazita et al., 2004). Three candidate genes for CL/P in this region, 17q23–25, have been described (*SEPT9*, *MKS1*, and *GAA*) (Huie et al., 1999; Jeannet et al., 2001; Paavola et al., 1995). In addition, *RARA*, which is located approximately 6 Mbp downstream from the breakpoint studied here, is associated with CL/P (Chenevix-Trench et al., 1992). In this study we identified a predicted gene containing *CCL2* in the 17q breakpoint region. *CCL2* (monocyte chemotactic protein-1) is a member of the small inducible gene family and plays a role in the recruitment of monocytes to sites of injury and infection and so is not an obvious candidate for CL/P. We did, however, observe multiple amino acid variants in the predicted gene. One of them, a heterozygous variant in position 29540817 (possibly N45S), showed a significant *p* value in the Filipino CL/P population compared with the control population (Table 1). Moreover, there was statistically significant evidence of transmission distortion for two haplotypes on chromosome 17 in Filipinos but not in the white population (Table 3). These data suggest that the participation of this region in facial development may vary among different populations, and these results warrant an extension of these studies to larger numbers for confirmation.

Although three generations were examined here segregating the same chromosome translocation, only one individual was clearly affected with cleft lip and palate. CL/P is commonly nonpenetrant even in disorders caused by single gene mutations. Mutations in *IRF6* cause the autosomal dominant van der Woude syndrome, but there is only 70% penetrance for the CL/P phenotype (Burdick, 1986; Zucchero et al., 2004). Mutations in *MSX1* and several of the FGF and FGFR genes may also cause CL/P, but again only a subset of individuals with the mutation have a cleft phenotype (van den Boogaard et al., 2000). Isolated clefts are complex traits, not single-gene Mendelian disorders. They are likely caused by several interacting genes, each with small effects. In our case, the presumed loss of 50% of gene product may be the tipping point in an individual already having several other of the predisposing alleles/genes that may be absent in other family members with the translocation, hence their unaffected status. Clearly, future work is required to develop a better understanding of penetrance in CL/P.

In summary, we report an extensive study of candidate genes for CL/P based on a family with CL/P and segregating a balanced chromosome translocation t(9;17)(q32;q12). Analysis of the translocation shows disruption of *SLC31A1* on chromosome 9 and a predicted gene that includes *CCL2* on chromosome 17. It is intriguing that *SLC31A1* lies close to *BSPRY*, an excellent candidate gene for CL/P, and the results of a case-control study support a role for *BSPRY* in CL/P. In the aggregate, these data may provide additional support for an intensive search for genes/mutations and also for microdeletions in the 9q/17q region that would play a role in CL/P.

Acknowledgments. We are especially grateful to this family, many affected subjects, and their family members for their willingness to participate in this study. We thank Marla Johnson, Susie McConnell, and Adela Mansilla (University of Iowa) for technical assistance; and Qing Ying Li, Chiaki Nakagawa. Kyoko Takeshima, and Reiko Kimura (Aichi Human Service Center) for assistance in chromosomal analysis. We also thank Tadashi Yamamoto and Atsuo Kaetsu (Toyohashi Municipal Hospital) for clinical support.

References

Alkuraya FS, Saadi I, Lund JJ, Turbe-Doan A, Morton CC, Maas RL. SUMO1 haploinsufficiency leads to cleft lip and palate. *Science.* 2006;313:1751.

Blanton SH, Cortez A, Stal S, Mulliken JB, Finnell RH, Hecht JT. Variation in IRF6 contributes to nonsyndromic cleft lip and palate. *Am J Med Genet A.* 2005;137:259–262.

Bridge J, Sanger W, Mosher G, Buehler B, Hearty C, Olney A, Fordyce R. Partial duplication of distal 17q. *Am J Med Genet.* 1985;22:229–235.

Burdick AB. Genetic epidemiology and control of genetic expression in van der Woude syndrome. *J Craniofac Genet Dev Biol.* 1986;Suppl 2:99–105.

Chenevix-Trench G, Jones K, Green AC, Duffy DL, Martin NG. Cleft lip with or without cleft palate: associations with transforming growth factor alpha and retinoic acid receptor loci. *Am J Hum Genet.* 1992;51:1377–1385.

Czako M, Riegel M, Morava E, Bajnoczky K, Kosztolanyi G. Opitz "C" trigonocephaly-like syndrome in a patient with terminal deletion of 2p and partial duplication of 17q. *Am J Med Genet A.* 2004;131:310–312.

Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R. Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics.* 1990;6:575–577.

Donnai D, Heather LJ, Sinclair P, Thakker Y, Scambler PJ, Dixon MJ. Association of autosomal dominant cleft lip and palate and translocation 6p23;9q22.3. *Clin Dysmorphol.* 1992;1:89–97.

FitzPatrick DR, Carr IM, McLaren L, Leek JP, Wightman P, Williamson K, Gautier P, McGill N, Hayward C, Firth H, et al. Identification of SATB2 as the cleft palate gene on 2q32–q33. *Hum Mol Genet.* 2003;12:2491–2501.

Ghassibe M, Bayet B, Revencu N, Verellen-Dumoulin C, Gillerot Y, Vanwijck R, Vikkula M. Interferon regulatory factor-6: a gene predisposing to isolated cleft lip with or without cleft palate in the Belgian population. *Eur J Hum Genet.* 2005;13:1239–1242.

Gurnett CA, Bowcock AM, Dietz FR, Morcuende JA, Murray JC, Dobbs MB. Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. *Am J Med Genet A.* 2007;143:27–32.

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

Huie ML, Kasper JS, Arn PH, Greenberg CR, Hirschhorn R. Increased occurrence of cleft lip in glycogen storage disease type II (GSDII): exclusion of a contiguous gene syndrome in two patients by presence of intragenic mutations including a novel nonsense mutation Gln58Stop. *Am J Med Genet.* 1999;85:5–8.

Ichikawa E, Watanabe A, Nakano Y, Akita S, Hirano A, Kinoshita A, Kondo S, Kishino T, Uchiyama T, Niikawa N, et al. PAX9 and

TGFB3 are linked to susceptibility to nonsyndromic cleft lip with or without cleft palate in the Japanese: population-based and family-based candidate gene analyses. *J Hum Genet.* 2006;51:38–46.

Jeannet PY, Watts GD, Bird TD, Chance PF. Craniofacial and cutaneous findings expand the phenotype of hereditary neuralgic amyotrophy. *Neurology.* 2001;57:1963–1968.

Jezewski PA, Vieira AR, Nishimura C, Ludwig B, Johnson M, O'Brien SE, Daack-Hirsch S, Schultz RE, Weber A, Nepomucena B, et al. Complete sequencing shows a role for MSX1 in non-syndromic cleft lip and palate. *J Med Genet.* 2003;40:399–407.

Kim HG, Herrick SR, Lemyre E, Kishikawa S, Salisz JA, Seminara S, MacDonald ME, Bruns GA, Morton CC, Quade BJ, et al. Hypogonadotropic hypogonadism and cleft lip and palate caused by a balanced translocation producing haploinsufficiency for FGFR1. *J Med Genet.* 2005;42:666–672.

Lee J, Prohaska JR, Thiele DJ. Essential role for mammalian copper transporter Ctr1 in copper homeostasis and embryonic development. *Proc Natl Acad Sci U S A.* 2001;98:6842–6847.

Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet.* 2003;12:1725–1735.

Loeys BL, Schwarze U, Holm T, Callewaert BL, Thomas GH, Pannu H, De Backer JF, Oswald GL, Symoens S, Manouvrier S, et al. Aneurysm syndromes caused by mutations in the TGF-beta receptor. *N Engl J Med.* 2006;355:788–798.

Luke S, Bennett HS, Pitter JH, Verma RS. A new case of monosomy for 17q25—qter due to a maternal translocation [t(3;17)(p12;q24)]. *Ann Genet.* 1992;35:48–50..

Mansilla MA, Cooper ME, Goldstein T, Castilla EE, Lopez Camelo JS, Marazita ML, Murray JC. Contributions of PTCH gene variants to isolated cleft lip and palate. *Cleft Palate Craniofac J.* 2006;43:21–29.

Marazita ML, Murray JC, Lidral AC, Arcos-Burgos M, Cooper ME, Goldstein T, Maher BS, Daack-Hirsch S, Schultz R, Mansilla MA, et al. Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32–35. *Am J Hum Genet.* 2004;75:161–173.

Martinet D, Vial Y, Thonney F, Beckmann JS, Meagher-Villemure K, Unger S. Fetus with two identical reciprocal translocations: description of a rare complication of consanguinity. *Am J Med Genet A.* 2006;140:769–774.

Metzke-Heidemann S, Kuhling-von Kaisenberg H, Caliebe A, Janssen D, Jonat W, Grote W, von Kaisenberg CS. Phenotypical variation in cousins with the identical partial trisomy 9 (pter-q22.2) and 7 (q35-qter) at 16 and 23 weeks gestation. *Am J Med Genet A.* 2004;126: 197–203.

Mizuguchi T, Collod-Beroud G, Akiyama T, Abifadel M, Harada N, Morisaki T, Allard D, Varret M, Claustres M, Morisaki H, et al. Heterozygous TGFBR2 mutations in Marfan syndrome. *Nat Genet.* 2004;36:855–860.

Mossey P, Little J. Epidemiology of oral clefts: an international perspective. In: Wyszynski DF, ed. *Cleft Lip and Palate: From Origin to Treatment.* New York: Oxford University Press; 2002:127–157.

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

Murray JC, Daack-Hirsch S, Buetow KH, Munger R, Espina L, Paglinawan N, Villanueva E, Rary J, Magee K, Magee W. Clinical and epidemiologic studies of cleft lip and palate in the Philippines. *Cleft Palate Craniofac J.* 1997;34:7–10.

Naritomi K, Izumikawa Y, Goya Y, Gushiken M, Shiroma N, Hirayama K. Trisomy 9q3 syndrome: a case report and review of the literature. *Clin Genet.* 1989;35:293–298.

Nickerson DA, Tobe VO, Taylor SL. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.* 1997;25: 2745–2751.

Ono T, Kawabe T, Sonta S, Okamoto T. Assignment of MARK3 alias KP78 to human chromosome band 14q32.3 by in situ hybridization. *Cytogenet Cell Genet.* 1997;79:101–102.

Paavola P, Salonen R, Weissenbach J, Peltonen L. The locus for Meckel syndrome with multiple congenital anomalies maps to chromosome 17q21–q24. *Nat Genet.* 1995;11:213–215.

Park JW, McIntosh I, Hetmanski JB, Jabs EW, Vander Kolk CA, Wu-Chou YH, Chen PK, Chong SS, Yeow V, Jee SH, et al. Association between IRF6 and nonsyndromic cleft lip with or without cleft palate in four populations. *Genet Med.* 2007;9:219–227.

Pivnick EK, Wilroy RS, Summitt JB, Tucker B, Herrod HG, Tharapel AT. Adjacent-2 disjunction of a maternal t(9;22) leading to duplication 9pter—22 and deficiency of 22pter—q11.2. *Am J Med Genet.* 1990; 37:92–96.

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

Riley BM, Mansilla MA, Ma J, Daack-Hirsch S, Maher BS, Raffensperger LM, Russo ET, Vieira AR, Dode C, Mohammadi M, et al. Impaired FGF signaling contributes to cleft lip and palate. *Proc Natl Acad Sci U S A.* 2007;104:4512–4517.

Scapoli L, Palmieri A, Martinelli M, Pezzetti F, Carinci P, Tognon M, Carinci F. Strong evidence of linkage disequilibrium between polymorphisms at the IRF6 locus and nonsyndromic cleft lip with or without cleft palate, in an Italian population. *Am J Hum Genet.* 2005;76:180–183.

Stalker HJ, Gray BA, Zori RT. Dominant transmission of a previously unidentified 13/17 translocation in a five-generation family with Robin cleft and other skeletal defects. *Am J Med Genet.* 2001;103:339–341.

van Bokhoven H, Celli J, Kayserili H, van Beusekom E, Balci S, Brussel W, Skovby F, Kerr B, Percin EF, Akarsu N, et al. Mutation of the gene encoding the ROR2 tyrosine kinase causes autosomal recessive Robinow syndrome. *Nat Genet.* 2000;25:423–426.

van den Boogaard MJ, Dorland M, Beemer FA, van Amstel HK. MSX1 mutation is associated with orofacial clefting and tooth agenesis in humans. *Nat Genet.* 2000;24:342–343.

Velagaleti GV, Bien-Willner GA, Northup JK, Lockhart LH, Hawkins JC, Jalal SM, Withers M, Lupski JR, Stankiewicz P. Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. *Am J Hum Genet.* 2005;76: 652–662.

Vieira AR, Avila JR, Daack-Hirsch S, Dragan E, Felix TM, Rahimov F, Harrington J, Schultz RR, Watanabe Y, Johnson M, et al. Medical sequencing of candidate genes for nonsyndromic cleft lip and palate. *PLoS Genet.* 2005;1:e64.

Welsh IC, Hagge-Greenberg A, O'Brien TP. A dosage-dependent role for Spry2 in growth and patterning during palate development. *Mech Dev.* 2007;124:746–761.

Wyandt HE, Lebo RV, Fenerci EY, Sadhu DN, Milunsky JM. Tandem duplication/deletion in a maternally derived chromosome 9 supernumerary derivative resulting in 9p trisomy and partial 9q tetrasomy. *Am J Med Genet.* 2000;93:305–312.

Yoshiura K, Machida J, Daack-Hirsch S, Patil SR, Ashworth LK, Hecht JT, Murray JC. 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.* 1998;54:231–240.

Zori RT, Gray BA, Bent-Williams A, Driscoll DJ, Williams CA, Zackowski JL. Preaxial acrofacial dysostosis (Nager syndrome) associated with an inherited and apparently balanced X;9 translocation: prenatal and postnatal late replication studies. *Am J Med Genet.* 1993;46:379–383.

Zucchero TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, Caprau D, Christensen K, Suzuki Y, Machida J, et al. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. *N Eng J Med.* 2004;351:769–780.

Journal of Human Genetics (2009) 54, 499–503

www.nature.com/jhg

ORIGINAL ARTICLE

Japanese map of the earwax gene frequency: a nationwide collaborative study by Super Science High School Consortium

The Super Science High School Consortium[1]

Wet/dry types of earwax are determined by the c.538G>A single-nucleotide polymorphism (SNP) in the *ABCC11* gene; GA and GG genotypes give the wet type and AA the dry type. The Japanese population may have a dual structure comprising descendants of mixtures between the ancient 'Jomon' and 'Yayoi' populations. We hypothesized that the dry type was introduced by the Yayoi people to the Jomon population where the wet type was predominant, and as the mixture of the two populations has not yet been complete, the allele-A frequency (f^A) would even now be higher along a putative Yayoi man's peopling route within Japanese islands. To know the frequency, a nationwide Super Science High School (SSH) Consortium collected 1963 fingernail samples of pupils/students from at least one high school/university in every prefecture. All further procedures, DNA extraction, SNP genotyping and gene frequency estimation, were carried out by trained SSH pupils. Although the allele-A frequency varied among the 47 prefectures, the Gifu/Kyoto and Okinawa prefectures showed the highest and lowest values, respectively. Areas with high frequencies included Northeastern Kyushu, Northern Shikoku and Kinki districts, showing a belt-like zone, whereas those with low frequencies other than Okinawa were the Southwestern Kyushu, Hiroshima prefecture and Tohoku districts. The f^A value in Kinki district was statistically higher than those in prefectures westward and east-northward from it. The result may provide another line of evidence supporting a possible route of the Yayoi-man's peopling in Japan.
Journal of Human Genetics (2009) **54**, 499–503; doi:10.1038/jhg.2009.62; published online 31 July 2009

Keywords: *ABCC11* gene; earwax type; gene frequency; Japanese history; Japanese population

INTRODUCTION

Human earwax is classified into two distinct phenotypes: wet and dry types. The wet earwax is a real secretory product from the ceruminous apocrine gland, whereas the dry type is the phenotype of lacking or reduced ceruminous secretion.[1,2] In 1907, Kishi[3] first described the different nature of earwax between the Japanese and Europeans. Adachi[4] studied earwax and axillary odor phenotypes in the Japanese and nearby ethnic populations, and not only found that the dry type is more frequently seen than the wet type among the Japanese but also that the wet type is linked to axillary odor. Matsunaga[5] provided evidence that the human earwax is a bimorphic Mendelian trait, and that the wet type is completely dominant to the dry type. Subsequent studies of various ethnic populations revealed that the dry type is highly specific to East Asians with a frequency of 80–95%, whereas it is rare (0–3%) in populations of European or African origin, and intermediate in values (30–50%) in Southeast Asian, Oceanian, Central Asian and Middle-East populations as well as in native Americans and Inuit people.[6–12]

By a genetic linkage analysis of eight Japanese families, Tomita *et al.*[13] successfully mapped the human earwax gene locus to 16p11.2-q12.1. Yoshiura *et al.*[2] at the same laboratory then identified the earwax determining gene, *ABCC11* (for ATP-binding cassette, sub-family C, member 11). A functional single-nucleotide polymorphism (SNP), c.538G>A (rs17822931), of the gene determines the earwax types; that is, AA homozygotes for allele A at rs17822931 have dry earwax, and GA heterozygotes and GG homozygotes have wet earwax. The allele-A frequencies calculated for 33 different ethnic populations around the world varied but showed a downward cline from the highest areas (the frequency of 1.00) in Shanxi Province (Northern Han Chinese) and Taegu City (Koreans) toward Japan and Southern Asia and toward Central Asia and Europe.[2] This map corresponded to the phenotypical data of earlier studies mentioned above.[4–12] The map may also indicate a possible route of migration and peopling of ancient Northeast Mongoloid with dry earwax.

There are two main models proposed for the origin of the variation in the current Japanese population: the dual-structure model versus the single-origin hypothesis.[14,15] According to the former model, the modern Japanese are descendants of mixtures between at least two populations, the 'Jomon' and 'Yayoi' people (both are named after the ware types they used). The Jomon people formed the native Japanese

Correspondence: T Nagashima, Nagasaki Nishi High School, Takenokubo 12-9, Nagasaki 852-8014, Japan.
E-mail: ssh01@nagasaki-nishi.ed.jp
[1]Members of the SSH Consortium (asterisk) and other high schools/universities that participated in the SSH Consortium study are listed in the Appendix.
Received 6 April 2009; revised 16 June 2009; accepted 17 June 2009; published online 31 July 2009

population inhabited widely in Japan islands since 16 000 years ago, and the Yayoi people were those who came over 3000–1800 years ago. Other immigrants, who came over largely from the Korean Peninsula, further joined in the society in the third to eighth centuries. On the other hand, the single-origin hypothesis proposes that the divergence in various districts may have led to the current genetic variation in Japan. On the basis of data from earlier studies,[2,5,9] we favored the dual structure theory and hypothesized that the dry earwax of the Japanese was introduced by the Yayoi people to the wet-type predominant Jomon population background. As the mixture of these populations has not yet been complete, the dry-type allele would be more frequently observed even now along a migration route of the newcomer population within Japan's islands.

As there has been no gene-based Japanese map of earwax types, a nationwide consortium of Super Science High School (SSH) was recently constructed to study the earwax gene frequency in every prefecture of Japan. Here, we report the result of a large-scale collaborative study by the SSH Consortium.

MATERIALS AND METHODS

The consortium composed of 48 SSHs that are located in various prefectures of Japan was established in 2006. The SSHs are natural science oriented high schools that are selected, approved and supported for 5 years by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan and the Japan Science and Technology Agency (JST). At its first assembly in 2006 in Nagasaki City, the SSH Consortium started a 2-year collaborative study with a common protocol that had been approved by the Committee for the Ethical Issues on Human Genome and Gene Analysis in Nagasaki University and also in the Health Sciences University of Hokkaido.

The consortium collected several (up to 10) fingernail clippings from each of 19–113 pupils and/or teachers from at least one SSH in every prefecture after obtaining written informed consent from them. In some prefectures where an SSH was not available or found it difficult to cooperate for the study, nail samples were obtained from the pupils of other high schools or from adult volunteers in respective prefectures under assist by some universities. The fingernail clippings obtained were put into a plastic tube, and the sample tubes were then numbered anonymously at each school/university and sent by mail to Nagasaki Nishi High School, one of the SSH Consortium members. Pupils of this school and some other SSHs were provided with minimum essential knowledge and methods of molecular genetic analysis of the human earwax gene during their summer/winter vacations by the staff members of the Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences.

A total of 1963 nail samples were collected from 55 high schools (48 SSHs and seven other high schools) and five universities (Table 1). The nail clippings were once frozen in liquid nitrogen and crushed into fine pieces manually with scissors, or into fine powder using Multi-beads Shocker (Yasui Kikai, Osaka, Japan). The piece/powder was dissolved in a urea-lysis solution (2 M urea; 0.5% SDS; 10 mM Tris-HCl, pH 7.5; 50 mM EDTA) containing 1 mg ml^{-1} proteinase K and 40 mM DTT at 55 °C overnight. Nail DNA in lysis solution was extracted with the QIAamp DNA Mini Kit (Qiagen, Tokyo, Japan). To genotype at the rs17822931 locus in a large number of samples, PCR was performed by the use of hydrolyzing TaqMan probes and a set of amplification primers. Their sequences (5′–3′) were as follows: CAGTGTACTCGGGCCAG and CAGT GTACTCAGGCCAG for hydrolyzing VIC-labeled TaqMan MGB wet probe and FAM-labeled dry probe, respectively; and CTTCTGGGCATCTGCTTCTG and CAAACCTCACCAAGTCTGCCA for EW-ampF and EW-ampR primers, respectively. Reactions were carried out using TaqMan Universal PCR Master Mix (AppliedBiosystems, Foster City, CA, USA). Thermal cycling was performed initially at 95 °C for 10 min followed by 45 cycles of 95 °C for 15 s and 62 °C for 1 min on LightCycler480 (Roche Diagnostics, Basel, Switzerland). Genotypes were determined according to the FAM/VIC fluorescence intensity ratio (Figure 1). The allele-A frequency was calculated with the formula of $(2A+B)/2C$, where A is the number of 'AA' homozygotes, B is the number of 'GA' heterozygotes and C is the total number of individuals examined.

Table 1 Allele-A frequency at each prefecture

	No. of genotypes at rs17822931				
Prefectures	*AA*	*GA*	*GG*	*Total*	*Allele-A frequency ± 95% confidence interval*
Hokkaido	47	8	0	55	0.927±0.069
Aomori	80	17	1	98	0.903±0.059
Iwate	21	20	2	43	0.721±0.134
Miyagi	17	6	2	25	0.800±0.157
Akita	23	6	0	29	0.897±0.111
Yamagata	38	14	0	52	0.865±0.093
Fukushima	45	18	1	64	0.844±0.089
Ibaragi	73	19	2	94	0.878±0.066
Tochigi	36	10	1	47	0.872±0.096
Gunma	55	19	0	74	0.872±0.076
Saitama	21	9	0	30	0.850±0.128
Chiba	22	7	0	29	0.879±0.119
Tokyo	21	4	0	25	0.920±0.106
Kanagawa	20	7	0	27	0.870±0.127
Niigata	25	9	0	34	0.868±0.114
Toyama	13	6	0	19	0.842±0.164
Ishikawa	21	7	0	28	0.875±0.123
Fukui	42	6	0	48	0.938±0.068
Yamanashi	27	11	2	40	0.813±0.121
Nagano	21	4	1	26	0.885±0.123
Gifu	22	1	0	23	0.978±0.060
Shizuoka	60	18	2	80	0.863±0.075
Aichi	18	6	0	24	0.875±0.132
Mie	11	2	0	13	0.923±0.145
Shiga	21	4	0	25	0.920±0.106
Kyoto	26	1	0	27	0.981±0.051
Osaka	19	4	0	23	0.913±0.115
Hyogo	24	4	0	28	0.929±0.095
Nara	19	4	0	23	0.913±0.115
Wakayama	99	13	1	113	0.934±0.046
Tottori	37	15	2	54	0.824±0.102
Shimane	15	7	0	22	0.841±0.153
Okayama	17	8	0	25	0.840±0.144
Hiroshima	28	22	1	51	0.765±0.116
Yamaguchi	62	18	2	82	0.866±0.074
Tokushima	31	9	0	40	0.888±0.098
Kagawa	41	7	1	49	0.908±0.081
Ehime	40	5	0	45	0.944±0.067
Kohchi	44	16	1	61	0.852±0.089
Fukuoka	21	3	1	25	0.900±0.118
Saga	18	6	1	25	0.840±0.144
Nagasaki	36	6	1	43	0.907±0.087
Kumamoto	37	11	2	50	0.850±0.099
Oita	22	2	0	24	0.958±0.080
Miyazaki	45	5	1	51	0.931±0.070
Kagoshima	19	6	0	25	0.880±0.127
Okinawa	13	11	1	25	0.740±0.172
Total/average	1,513	421	29	1,963	0.878±0.014

High schools and universities that participated in this study are listed in APPENDIX.

RESULTS AND DISCUSSION

This study confirmed that the Japanese population has two distinct earwax types. The genotype detection rate among 1963 nail samples was 99.6%, and the average allele-A frequency (f^A) among 47 prefectures was 0.878. Thus, the dry-type frequency among the Japanese is 77.1%, which is given as the square value of f^A by assuming

the Hardy–Weinberg equilibrium, being comparable with those estimated from earlier phenotypical studies of earwax.[4,5] The f^A value with 95% confidence interval in each prefecture is shown in Table 1. A Japanese map by prefecture on which these values are depicted in color, that is, the lighter color, the higher frequency, is shown in Figure 2. The map shows the highest and the lowest allele-A frequencies in the Gifu/Kyoto prefectures and in the Okinawa prefecture, respectively. Areas with relatively high frequencies included Northeastern Kyushu, Northern Shikoku along the Seto Inland Sea, and Kinki district between the Hyogo and Gifu prefectures, showing a belt-like lighter-color zone, whereas those with relatively low frequencies other than the Okinawa prefecture were the Southwestern Kyushu, Chugoku (around Hiroshima prefecture) and Tohoku districts. The f^A value in Kinki district (Hyogo, Osaka, Nara, Wakayama, Kyoto and Shiga prefectures) is statistically higher than those in prefectures westward from Kinki (P=0.0030 when including Tokyo and Hokkaido, and P=0.0016 when excluding them, using Mann–Whitney's U-test) and east-northward (P=0.0053 when including Okinawa, and P=0.067 when excluding it). The f^A values look higher in the westward prefectures than that in the east-northward prefectures, but they were not significantly different (P=0.7308 and P=0.8360, when including and excluding Tokyo/Hokkaido and Okinawa prefectures, respectively, using Mann–Whitney's U-test).

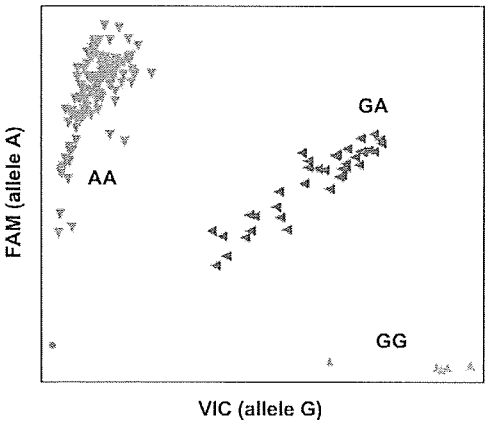

Figure 1 Three different genotypes at the rs17822931 locus by the TaqMan PCR method. 'AA', 'GA' and 'GG' are AA homozygotes (phenotypically dry type) GA heterozygotes (wet type) and GG homozygotes (wet type), respectively.

As far as the earwax variation among the Japanese is concerned, the results in this study cannot absolutely rule out either the dual structure model or the single-origin hypothesis. Our earlier analysis on three SNP sites (rs17822931–rs6500380–ss49784070) around the *ABCC11* locus showed a common haplotype in each of dry-type individuals and wet-type individuals.[2] Therefore, considering these earlier haplotyping data suggesting a founder effect of the allele-A,[2] the present data may favor the dual-structure model.

According to our initial hypothesis of this study, prefectures with lighter color would indicate the areas with more descendants of the Yayoi population, whereas those with darker color indicate the areas with more descendants of the Jomon population. The earwax gene frequency map tended to correspond to the distribution for

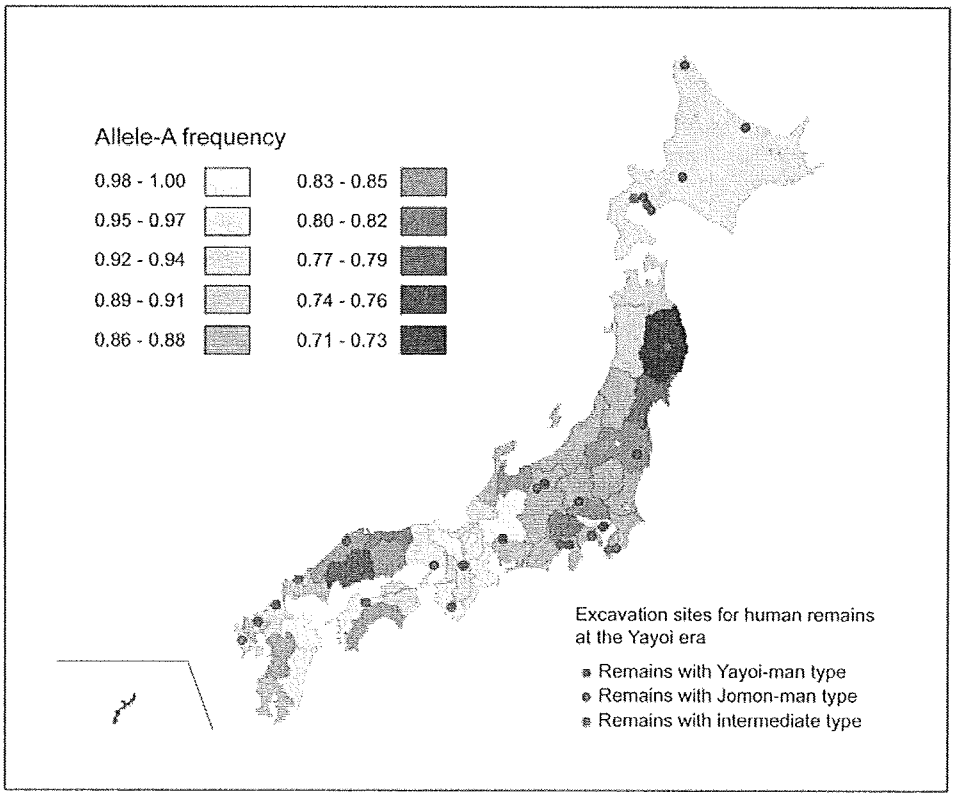

Figure 2 Japanese map of the allele-A (dry-type allele) frequency, merged with excavation sites for human remains at the Yayoi era. The lighter color, the higher frequency of allele A.

archeological sites (data from the National Science Museum, Tokyo, http://shinkan.kahaku.go.jp/index_jp.jsp), where Jomon (blue circle) and Yayoi (red circle) remains have been discovered. Excavation sites for human remains in the Yayoi era are mostly located in Western Japan, whereas those for Jomon-man remains tend to exist in Eastern Japan. It is of great interest that the distribution of Yayoi dig sites overlaps with the areas of higher allele-A frequencies even in the present time, and *vice versa* for Jomon dig sites with lower frequencies, although no statistical difference (P=8980, Mann–Whitney's U-test) was observed in allele frequency between the two integrated areas of prefectures where the two groups of dig-sites were discovered, respectively. Considering the dual-structure model, the overlapping distribution might reflect incomplete mixtures between the two populations with or without wet earwax during the past 3000 years. This argument is supported by the most recent finding on the genetic and anthropological structure of the modern Japanese population.[16] By genotyping at about 140 000 SNP loci among 7001 Japanese, Yamaguchi-Kabata *et al.*[16] found that the Japanese population consists mainly of two clusters, the Ryukyu (Okinawa) and the Hondo (main islands) clusters, which can be characterized by SNP genotypes at the hair thickness gene (*EDAR*) and the earwax gene (*ABCC11*) loci. Earlier studies on carrier rates of adult T-cell leukemia virus (HTLV-1) in the Japanese population showed a geographical distribution similar to that of the wet earwax in this study, and suggested that the characteristics of the native Japanese (the Jomon man) tended to still remain in habitants in Southwestern Kyushu and Southern Shikoku districts.[17,18]

A belt showing high allele-A frequency (lighter color) in Western Japan (Figure 2) may reflect the hypothetical route of the Yayoi people's migration within Japan's islands. They may have been continually coming over through the Korean Peninsula or Southern China first to Northern Kyushu since thousands of years, and migrated along the Seto Inland Sea toward the Kinki area, moving to other areas along the Japan Sea and Pacific Ocean, and finally reaching Northeast Japan. The reasons why the Hiroshima and Aomori prefectures show considerably low and high allele-A frequencies, respectively, remain unknown. Although small sample size and/or high-school selection bias cannot be ruled out, it might suggest an alternative route through Shikoku or a sea route by the Black Current of Pacific Ocean from Eastern Kyushu to Kinki district and far to the north.

In conclusion, a Japanese map of the earwax gene frequency was made by this SSH study. It may provide another line of evidence that suggests a possible route of the Yayoi-man's peopling in Japan. The SSH Consortium strongly hopes that high-school pupils will learn in the near future the achievement of this interdisciplinary study between humanity (Japanese history) and science (genetics) accomplished by the pupils themselves.

ACKNOWLEDGEMENTS

The SSH Consortium expresses its great gratitude to high-school pupils and teachers, university students and to adult volunteers for their donations of fingernail clippings to the study. This study was supported in part for the SSH Consortium by the SSH Program from Japan Science and Technology Agency (JST), and for N Niikawa by SORST from JST and by a Grant-in-Aid for Scientific Research (Category B, no. 19390095) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

1 Shugyo, Y., Sudo, N., Kanai, K., Yamashita, T., Kumazawa, T. & Kanamura, S. Morphological differences between secretory cells of wet and dry type of human ceruminous glands. *Am. J. Anat.* **181**, 377–384 (1988).
2 Yoshiura, K., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T. *et al.* SNP in the *ABCC11* gene is the determinant of human earwax type. *Nat. Genet.* **38**, 324–330 (2006).
3 Kishi, K. Anatomie des Ohres der Japaner. I. Die Drüsen des äusseren Gehörgans. *Arch. f. Ohrenheilk.* **70**, 205–210 (1907).
4 Adachi, B. Das Ohrenschmalz als Rassenmerkmal und der Rassengeruch ('Achselgeruch') nebst dem Rassenunterschied der Schweissdrüsen. *Z. Rassenk.* **6**, 273–307 (1937).
5 Matsunaga, E. The dimorphism in human normal cerumen. *Ann. Hum. Genet.* **25**, 273–286 (1962).
6 Petrakis, N. L., Molohon, K. T. & Tepper, D. J. Cerumen in American Indians: genetic implications of sticky and dry types. *Science* **158**, 1992–1993 (1967).
7 Petrakis, N. L. Dry cerumen: a prevalent genetic trait among American Indians. *Nature* **222**, 1080–1081 (1969).
8 Petrakis, N. L., Pingke, U., Petrakis, S. J. & Petrakis, S. L. Evidence for a genetic cline in earwax type in the Middle East and Southeast Asia. *Am. J. Phys. Anthropol.* **35**, 141–144 (1971).
9 Omoto, K. Polymorphic traits in peoples of eastern Asia and the Pacific. *Israel J. Med. Sci.* **9**, 1195–1215 (1973).
10 Patel, S. Study on some genetic traits of Tibetans at a Chandragiri, district Ganjam, Orissa. *Am. J. Phys. Anthopol.* **38**, 755–756 (1973).
11 Norakmal, I. & Tan, S. G. Cerumen polymorphism in the three major ethnic groups of Malaysia. *Jpn. J. Hum. Genet.* **24**, 119–121 (1979).
12 Spitsyn, V. A. & Afanasèva, I. Genetic geography of inherited dimorphism of ear wax by its consistency. *Genetika* **25**, 1854–1860 (1989).
13 Tomita, H-A., Yamada, K., Ghadami, M., Ogura, T., Yanai, Y., Nakatomi, K. *et al.* Mapping of the wet/dry earwax locus to the pericentromeric region of chromosome 16. *Lancet* **359**, 2000–2002 (2002).
14 Hanihara, K. The population history of the Japanese (in Japanese). *Nippon Ronen Igakkai Zasshi* **30**, 923–931 (1993).
15 Omoto, K. & Saitou, N. Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *Am. J. Phys. Anthropol.* **102**, 437–446 (1997).
16 Yamaguchi-Kabata, Y., Nakazono, K., Takahashi, A., Saito, S., Hosono, N., Kubo, M. *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.* **83**, 445–456 (2008).
17 Hinuma, Y. Seroepidemiology of adult T-cell leukemia virus (HTLV-I/ATLV): Origin of virus carriers in Japan. *AIDS Res.* **2**(Suppl 1), S17–S22 (1986).
18 Takatsuki, K., Matsuoka, M. & Yamaguchi, K. Adult T-cell leukemia in Japan. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **13**(Suppl 1), S15–S19 (1996).

APPENDIX

Satoshi Sakai, Sapporo Kaisei High School, Hokkaido; Kazumi Imai, *Obihiro Hakuyo High School, Hokkaido; Takashi Ogawa, *Hachinohe Kita High School, Aomori; Hiroshi Iwaoka, Sanbongi Agricultural High School, Aomori; Mikiko Ishii, *Mizusawa High School, Iwate; Shinichiro Komori, *Aizu High School, Fukushima; Toyohiko Yoshida, *Fukushima High School, Fukushima; Hideyuki Jumonji, *Seishingakuen High School, Ibaragi; Keiji Iizumi, Sakai High School, Ibaragi; Hiroshi Ohshima, *Sano Nihondaigaku High School, Tochigi; Toshiaki Maeda, *Takasaki High School, Gunma; Akira Kanno, *Urawa Daiichi Joshi High School, Saitama; Kenji Takahashi, *Kashiwa High School, Chiba; Hiroto Kubota, *Tokyo Metropolitan Koishikawa High School, Tokyo; Yoko Inoue, Hiroyoshi Ikeda, *Seisho High School, Kanagawa; Yoshiyuki Takahashi, *Niigata Minami High School, Niigata; Hiroshi Onoda, *Takaoka High School, Toyama; Rie Uchiyama, *Nanao High School, Ishikawa; Michihiko Matsuda, *Koshi High School, Fukui; Takashi Akazawa, *Fujishima High School, Fukui; Naohiro Kawamura, *Tsuru High School, Yamanashi; Toru Odagiri, *Yashiro High School, Nagano; Yasuharu Watanabe, *Ena High School, Gifu; Yukihiro Matsumoto, *Iwata Minami High School, Shizuoka; Seiji Shinoda, *Shimizu Higashi High School, Shizuoka; Masato Terada, *Ichinomiya High School, Aichi; Manabu Matsuoka, *Yokkaichi High School, Mie; Chikara Ueno, Takada Senior High School, Mie; Etsuo Ozaki, *Zeze High School, Shiga; Sadafusa Takaya, *Senior High School Attached to Kyoto University of Education, Kyoto; Tetsuharu Takeyama, *Rakuhoku High School, Kyoto; Toshiaki Hujita, *Ten-noji High School, Osaka; Kazuya Kawakatsu, *Kakogawa Higashi High School, Hyogo; Junichi Takemura, *Nara High School, Nara; Hitonori Maekawa, *Koyo High School, Wakayama; Tomoko Doei, *Hidaka High School, Wakayama; Shigeru

Ihara, *Toin High School, Wakayama; Yuji Sakaguchi, *Tottori Higashi High School, Tottori; Yasuyuki Hirota, *Masuda High School, Shimane; Akihiko Shindo and Hiroko Araki, *Okayama Ichinomiya High School, Okayama; Junko Miura and Tatsumi Morita, *Kokutaiji High School, Hiroshima; Takayuki Fujiwara, *Ube High School, Yamaguchi; Haruhiko Akiyama, *Jonan High School, Tokushima; Shinya Itome, *Sanbonmatsu High School, Kagawa; Yoshihisa Tanaka and Kazunori Nakagawa, *Matsuyama Minami High School, Ehime; Sumito Okamoto and Shushi Yamamoto, *Ozu High School, Kohchi; Takaaki Aoyagi, Munakata High School, Fukuoka; Toru Noda, *Chienkan High School, Saga; Shinichi Inoue, Hokuyodai High School, Nagasaki; Isao Hirota, Kiyoshi Tanaka, Tetsuya Nagashima, Iwao Koga and Kayo Watanabe, *Nagasaki Nishi High School, Nagasaki; Hideto Kusadome, *Kumamoto Daini High School, Kumamoto; Hirosi Otsuka, *Maizuru High School, Oita; Tosifumi Takayama, *Miyazaki Kita High School, Miyazaki; Hiroshi Miwa, Kinkowan High School, Kagoshima; Atsushi Hamakawa and Katsunori China, *Kaiho High School, Okinawa; Norio Niikawa, Tohru Ohta, Dmytro Starenki, Nadiya Sosonkina, and Ken Umehara, Health Sciences University of Hokkaido, Hokkaido; Kensuke Yamada, Tohoku University, Miyagi; Yoichi Shimada, Hiroyuki Nagasawa and Takashi Minato, Akita University, Akita; Toshihiko Ogino, Yamagata University, Yamagata; and Koh-ichiro Yoshiura, Nobutomo Miwa, Masayo Nomura, Hideo Kuniba, Yasuko Noguchi, Shinji Ono, Masayoshi Tsuda, Mitsuko Nakashima, Taeko Kikuchi, Daisuke Satoh, Tatsuya Kishino, Shinji Kondo and Akira Kinoshita, Nagasaki University, Nagasaki.

Search for Genomic Alterations in Monozygotic Twins Discordant for Cleft Lip and/or Palate

Jane W. Kimani,[1*] Koh-ichiro Yoshiura,[2] Min Shi,[3] Astanand Jugessur,[4] Danilo Moretti-Ferreira,[5] Kaare Christensen,[6] and Jeffrey C. Murray[1]

[1] *Department of Pediatrics, University of Iowa, United States of America*
*(*currently at Department of Pathology and Lab Medicine, University of North Carolina, United States of America)*
[2] *Department of Human Genetics, Nagasaki University Graduate School of Biomedical sciences, Nagasaki, Japan*
[3] *Biostatistics Branch, National Institute of Environmental Health Sciences (NIEHS), United States of America*
[4] *Craniofacial Development, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Australia*
[5] *Servico de Aconselhamento, Genetico da Universidade Estadual Paulista, Botucatu, Brazil*
[6] *Department of Epidemiology, Institute of Public Health, University of Southern Denmark, Denmark*

Phenotypically discordant monozygotic twins offer the possibility of gene discovery through delineation of molecular abnormalities in one member of the twin pair. One proposed mechanism of discordance is postzygotically occurring genomic alterations resulting from mitotic recombination and other somatic changes. Detection of altered genomic fragments can reveal candidate gene loci that can be verified through additional analyses. We investigated this hypothesis using array comparative genomic hybridization; the 50K and 250K Affymetrix GeneChip® SNP arrays and an Illumina custom array consisting of 1,536 SNPs, to scan for genomic alterations in a sample of monozygotic twin pairs with discordant cleft lip and/or palate phenotypes. Paired analysis for deletions, amplifications and loss of heterozygosity, along with sequence verification of SNPs with discordant genotype calls did not reveal any genomic discordance between twin pairs in lymphocyte DNA samples. Our results demonstrate that postzygotic genomic alterations are not a common cause of monozygotic twin discordance for isolated cleft lip and/or palate. However, rare or balanced genomic alterations, tissue-specific events and small aberrations beyond the detection level of our experimental approach cannot be ruled out. The stability of genomes we observed in our study samples also suggests that detection of discordant events in other monozygotic twin pairs would be remarkable and of potential disease significance.

Keywords: monozygotic twins, discordant, cleft lip and palate, genome-wide

Classical twin research studies comparing disease concordance rates between monozygotic (MZ) and dizygotic (DZ) twins have been extensively applied to estimate the contribution of genetic and environmental factors to many complex traits (Boomsma et al., 2002). MZ twin concordance for common diseases and traits rarely reaches 100%, an observation that is often attributed to differential environmental exposures (Wong et al., 2005). However, there are several molecular mechanisms that could underlie phenotypic discordance between MZ twins, such as *de novo* somatic mutations (Kondo et al., 2002), chromosomal anomalies (Gilbert et al., 2002), skewed X chromosome inactivation (De Gregorio et al., 2005), imprinting defects (Weksberg et al., 2002) and differential gene expression (Mak et al., 2004). The identification of molecular genetic differences between discordant MZ twins suggests that the utility of twin studies could be extended beyond heritability studies to gene discovery.

Postzygotic mitotic recombination and other somatic events such as deletions, nondisjunction, gene conversion, mobile genetic elements, fragile sites and repeat expansions have been proposed as a cause of MZ twin discordance (Cote & Gyftodimou, 1991; Kastern & Kryspin-Sorensen, 1988). Occurrence of somatic genomic changes in the early embryo can have significant implications for the phenotypes of the twins and the twinning process. If a genomic alteration occurs prior to twinning, unequal allocation of cells with the lesion to the two embryos could account for discordant phenotypes, with evidence of mosaicism, while the segregation of two genetically different populations of cells may trigger the twinning process, resulting in co-twins with discordant phenotypes (Machin, 1996). Since detection of chromosomal abnormalities relies on the comparison of a test genome with a reference genome sequence, the identical genomes of MZ twins ensure that any genomic differences can be established with confidence and

Received 11 November, 2008; accepted 17 June, 2009.

Address for correspondence: Jeffrey C. Murray, MD, University of Iowa, Department of Pediatrics, S. Grand Avenue, 2182 ML, Iowa City IA 52242 USA. E-mail: jeff-murray@uiowa.edu

accuracy. This approach is particularly advantageous for delineating disease causing copy number variants that could arise in one member of the twin pair. Any genomic mismatches between the twins' genomes could be revealed as amplifications, deletions or loss of heterozygosity. Recent identification of novel cleft lip and/or palate (CLP) candidate genes by genome-wide array-CGH highlights the impact of genomic alterations in the etiology of this complex disorder (Osoegawa et al., 2008).

CLP occurs in approximately 1 out of 700 births worldwide and has a significant clinical and economic impact. Approximately 40% of MZ twins are discordant for the disease phenotype (Christensen & Fogh-Andersen, 1993a). Genetic factors contributing to disease etiology include *IRF6* (Zucchero et al., 2004); *MSX1* (Jezewski et al., 2003); *RYK* (Watanabe et al., 2006) and genes in the *FGF* pathway (Riley et al., 2007). Additionally, nongenetic factors such as maternal cigarette smoking and nutrition also play a role in CLP etiology (Jugessur & Murray, 2005).

The complex nature of CLP complicates traditional mapping approaches such as linkage and association, which rely on population-matched or family-based controls to determine if a sequence variant is disease-causing. Based on identical genetic backgrounds, the unaffected twin in a discordant pair provides a well-matched control for studying a complex disease such as CLP thereby eliminating the need for external controls. We have previously reported our model using MZ twins discordant for CLP in gene discovery (Mansilla et al., 2005). In this report, we describe our search for genomic alterations in MZ twins discordant for CLP using array CGH and high-density SNP genotyping arrays.

Materials and Methods

Study Subjects

A summary of the discordant MZ twin samples and the respective analyses employed for each twin pair is presented in Table 1. The variability in the methods used to analyze the different samples resulted from multiple stages of the study being conducted over a period of time. All affected individuals were nonsyndromic cases of CLP. We used DNA previously

Table 1

Twin Samples Analyzed in This Study

Twin sample data				Analysis method			
Twin pair no.	Country	Sex	Phenotype	Array CGH	Affymetrix 50K	Affymetrix 250K	Custom Illumina SNPs
1	P	F	CL+P	✓			✓
2	P	M	CLO	✓	✓		✓
3	P	M	CL+P	✓			✓
4	P	F	CLO	✓		✓	✓
5	P	F	CL+P	✓	✓		✓
6	C	M	CL+P	✓		✓	
7	C	M	CLO			✓	
8	P	F	CL+P			✓	✓
9	B	M	CL+P			✓	
10	B	F	CLO			✓	✓
11	B	M	CLO			✓	✓
12	B	F	CLO			✓	
13*	B	F	CL+P			✓	
14	P	M	CL+P			✓	✓
15	A	F	U/T				✓
16	A	M	U/T				✓
17	A	M	U/T				✓
18	D	M	CLO				✓
19	D	M	CLO				✓
20	D	F	CLO				✓
21	D	F	CLO				✓
22	D	F	CLO				✓
23	D	M	CLO				✓
24	D	F	CL+P				✓
25**	D	F	CPO				✓

Note: **Father's DNA sample unavailable; **Triplets; A-Australia; B-Brazil; C-Colombia; D-Denmark; P-Philippines; U/T-Untyped cleft phenotype.*

extracted from peripheral blood lymphocytes. Genotyping a set of DNA markers had previously established that the twins were MZ. Informed consent was obtained for all study participants and the institutional review board approved the study.

Array CGH

CGH is a molecular cytogenetic technique for analyzing DNA copy number variations. DNA from a test and reference sample are differentially labeled and hybridized to an array spotted with a genomic representation that allows detection of copy number differences between the two samples at specific genomic locations (Pinkel & Albertson, 2005). For our analysis, Koh-ichiro Yoshiura et al. at Nagasaki University developed an array spotted with 2,173 genomic BAC clones. Samples from six pairs of twins were analyzed: for each pair, the DNA samples were differentially labeled, hybridized, and then scanned. Reverse sample labeling was also performed. The average normalized inter-locus fluorescence ratio (ANILFR) between the affected and unaffected twin samples was calculated. Prior preliminary experiments using five sets of normal/normal control samples had established the thresholds for copy number gain and loss at 0.86 and 1.18 with a standard deviation (*SD*) of 0.06, so the normal ANILFR range was defined as within ± 2 *SD*.

SNP Genotyping Arrays

Large scale SNP genotyping allows detection of allelic imbalances such as loss of heterozygosity (LOH) and copy number changes through hybridization signal intensities. In our analysis, we utilized a custom BeadArray™ platform from Illumina, Inc., (San Diego, CA, USA) and Affymetrix GeneChip® Human Mapping 50K and 250K sets (Santa Clara, CA, USA).

Illumina custom SNP genotyping. The Illumina BeadChip is a platform for performing multiplex gene analyses using oligonucleotides attached to silica beads. Defined SNPs can be chosen for a custom genotyping array based on the interests of the researcher (Steemers & Gunderson, 2007). For our analysis, we selected 1,536 SNPs representing 388 CLP candidate genes for a single genotyping array. The experiments were carried out at the Center for Inherited Disease Research (CIDR) according to the manufacturer's instructions and included 20 pairs of discordant MZ twins. Data were analyzed using Illumina's BeadStudio v2 genotyping and LOH plus module software that allow detection of chromosomal aberrations and allelic imbalance in paired samples (www.illumina.com).

Affymetrix SNP genotyping. Affymetrix GeneChips employ a whole genome sampling analysis method to genotype thousands of SNPs on synthetic oligonucleotide arrays by allele-specific hybridization. DNA samples were prepared for analysis and hybridization according to the manufacturer's instructions (www.affymetrix.com). We used the GeneChip® Human Mapping 50K *Xba* I (n = 2 twin pairs) and the 250K *Nsp* I (n = 10 twin pairs) arrays. SNP allele calls were assigned using a Dynamic Model mapping algorithm, a highly accurate genotype calling method that uses a one-sided Wilcoxon signed rank test to provide a confidence score (CS) for each genotype. The CS measures the reliability of a genotype call thus filtering out SNPs with a high error rate as 'no-calls' (Matsuzaki et al., 2004). Gender status for each sample is inferred based on X chromosome heterozygosity. We performed a paired analysis for LOH and copy number changes using the Affymetrix Chromosome Copy Number Analysis Tool 4.0 (CNAT 4.0).

DNA Sequencing

Genotypes generated using the Affymetrix 50K and 250K GeneChips that had discordant allele calls and a confidence score of ≤ 0.05 were identified as candidates for DNA sequencing to verify the SNP genotypes. After ranking by the significant confidence scores, at least 10 SNPs that did not fall within repeat elements were selected for sequencing in each twin pair. DNA samples from the parents were included in the sequencing analysis to verify Mendelian segregation of alleles. PCR reactions were performed on Applied Biosystems Gene Amp PCR System 9700 with 20ng

Table 2

BAC Clones With Significantly High Signal Differences From aCGH

BAC Clone	Locus	Twin pair no.					
		1[a]	2[b]	3[a]	4[a,c]	5[a,b]	6[c]
GS-98C4	XpYpter			✓	✓	✓	
RP11-89P7	2q32.1 - 2q32.2			✓		✓	
RP11-1145M16	7q11.2					✓	✓
RP11-117N14	8q21.11	✓			✓		
RP11-613G2	11p15.5		✓				✓
RP11-75H6	19p13.3			✓			✓
RP11-17K15	19p13.3		✓			✓	
RP11-197O4	19p13.2			✓	✓		

Note: Samples also genotyped with the Illumina genotyping panel[a], the 50K Affymetrix GeneChip[b], or the 250K Affymetrix GeneChip[c].

Table 3

Summary of Affymetrix GeneChips Genotyping Results

50K Affymetrix GeneChip				
Twin pair no.	SNP Call rate (affected/unaffected) average: 91.4%	Genotype concordance average -99.7%	No. of discordant genotypes	Discordant genotypes with ≤ 0.05 confidence score
2	90.7 / 92.7	99.7	180	4
5	91.3 / 91.5	99.8	137	5
250K Affymetrix GeneChip				
Twin pair no.	SNP Call rate Average –91.5%	Genotype concordance Average -98.8%	No. of discordant genotypes	Discordant genotypes with ≤ 0.05 confidence score
4	95.7 / 93.9	99.3	1818	13
6	87.7 / 88.3	98.6	3668	16
7	86.0 / 85.7	98.0	5349	64
8	93.4 / 94.0	99.2	2126	7
9	90.8 / 90.7	98.6	3607	32
10	91.9 / 92.1	98.9	2945	27
11	95.3 / 93.6	99.2	2027	15
12	94.5 / 92.4	99.0	2528	24
13	93.2 / 93.0	99.0	2711	35
14	92.6 / 90.5	98.8	3052	35

DNA and 0.25 units of Biolase (Bioline, Randolph, MA). Sequencing was carried out with the Big Dye™ Terminator cycle sequencing and run on ABI Prism 3730 DNA Analalyzer (Applied Biosystems). Sequences were assembled using the Phred-Phrap package and visualized using the Consed program.

Results

Array CGH Results

With 2,173 BAC clones spread across the genome, the aCGH provided a resolution of approximately 1-Mb. Table 2 shows 8 BAC clones with fluorescence ratios that deviated significantly (> 3 SD) in at least two twin pairs. The regions covered by these BACs do not overlap with any known CLP candidate loci, nor contain genes showing high craniofacial expression according to COGENE — the Craniofacial and Oral Gene Expression Network. Genomic region 19p13 looked especially interesting with 3 BACs showing high signal differences, so we genotyped two SNPs of high heterozygosity within each BAC in the twins and the parents, but did not find evidence for allelic imbalance.

Illumina Genotyping Results

The average SNP call rate on the Illumina genotyping panel was approximately 89%. We did not observe any discordant genotypes between pairs of MZ twins. A paired analysis with the Beadstudio software was performed for LOH and copy number changes at a 1 Mb window size. Although the sparse SNP coverage of ~1,500 SNPs greatly limited the power of this analysis, there was no indication of LOH in the twin pairs. A few genomic regions revealed possible copy number changes for twin pairs 3–6, but the more comprehensive genotype data generated from Affymetrix arrays disproved that observation.

Affymetrix Genotyping Results

Genotyping data and concordance rates for the samples scanned with the Affymetrix GeneChips are summarized in Table 3. The call rate from 58,960 and 262,264 SNPs for the 50K and 250K GeneChips respectively was > 90% for all samples. The average proportion of genotypes that were concordant between twin pairs was ~99% for both GeneChips. This was comparable to the 98% degree of genotype concordance observed in comparing two independent 250K array scans of the unaffected individual in twin pair no. 6, which revealed high reproducibility of the genotype calls.

Sequencing Results

DNA sequencing was carried out for a total of 107 regions surrounding SNPs that had received discordant genotype calls from the Affymetrix GeneChip analyses within twin pairs. Results revealed 181 SNP genotypes that were concordant between twin pairs (Table 4). Additionally, sequencing of DNA samples from the parents showed consistency with Mendelian inheritance.

Discussion

The advent of genome scanning tools allows comprehensive analysis of chromosomal rearrangements, dependent on the resolution of the experimental approach. The aCGH experiment using genomic

Table 4

Summary of SNPs Genotyped by Sequencing to Confirm Concordance Between Twin Pairs

Twin pair	Region (no. SNPs)												
2	2q23.3 (1)	5q12 (1)	10q21 (1)	16p12.1 (1)									
5	8q24.12 (3)	10q21.1 (1)	12q13.13 (1)	15q26.2 (1)									
4	2q14.3 (1)	3q.27.2 (1)	10p15.2 (1)	10q.26.2 (1)	13q31.1 (1)	16q.23.3 (1)	21q21.1 (2)						
6	1q32.1 (4)	4p15.32 (3)	4q32.3 (1)	5q21.3 (1)	7p12.1 (3)	7q31.32 (2)	8q22.1 (2)	9p24.2 (2)	10q11.23 (1)	10q21.1 (4)	11q24.2 (1)	12q15 (1)	14q31.3 (1)
7	2q21.3 (1)	5p15.33 (2)	6q14.1 (1)	7p21.1 (1)	12q21.2 (1)	12q21.32 (2)	12q23.2 (2)	16q22.3 (1)	18q12.3 (1)	19q13.32 (4)			
8	2p24.1 (1)	2q21.3 (1)	4p13 (1)	5q21.3 (1)	8q24.22 (3)	11p13 (2)	12q21.33 (1)	21q21.1 (2)	Xq25 (2)				
9	2q13 (2)	2q31.2 (2)	4q31.23 (1)	5p13.3 (1)	5q14.1 (3)	7q21.11 (1)	10q23.31 (3)	11p15.4 (2)	11p14.2 (2)	17q11.2 (1)			
10	1q32.1 (4)	1q32.1 (5)	2p16.1 (3)	3p14.2 (1)	5p13.2 (1)	5q14.3 (3)	5q35.2 (3)	7p21.2 (1)	8q23.2 (1)	14q21.3 (2)			
11	1q25.3 (1)	2q34 (1)	4q22.3 (3)	5q12.1 (3)	7p21.2 (1)	7q21.3 (2)	11p14.1 (1)	12p12.3 (1)	12q12 (1)	Xq28 (2)			
12	1p36.12 (1)	1q44 (3)	2p24.3 (5)	4p16.3 (1)	6q13 (3)	6q16.1 (1)	7p15.1 (1)	10q21.3 (1)	10q26.11 (1)	18q12.3 (1)			
13	1p12 (1)	1q24.2 (4)	3q23 (2)	3q25.33 (1)	6p25.3 (1)	6q16.3 (2)	9p21.3 (1)	10p15.1 (1)	16q12.2 (2)	17q22 (1)			
14	3p12.3 (1)	3q13.33 (1)	4p12 (2)	6q16.1 (1)	8p23.2 (2)	8q23.1 (2)	8q23.2 (1)	10p12.32 (1)	10q21.3 (2)	11p11.2 (1)			

BACs provided extensive coverage of the genome at a resolution of ~1 Mb, so any smaller chromosomal aberrations would not be detected. SNP arrays not only provide genotypes for thousands of SNPs, but can also be used to detect copy number changes based on hybridization signal intensities. Genotype data is useful for detecting loss of heterozygosity in chromosomal regions with deletions or uniparental disomy. At an average inter-marker distance of ~60 kb and ~12 kb for the 50K and 250K GeneChips respectively, and additional genotyping through the Illumina SNP panel, our analysis provide sufficient resolution to detect submicroscopic structural variants that are defined in the range of ~10 kb to 3 Mb (Feuk et al., 2006). Several samples were interrogated using more than one method (table 1), thus allowing data comparison and verification.

Our study using aCGH and genotyping arrays did not reveal any genomic alterations within MZ twins discordant for nonsyndromic CLP. Genomic alterations could be confined to specific tissues depending on the timing of the mutational event, resulting in mosaicism. Since acquisition of DNA samples from lip and palatal tissues is difficult, our analysis was limited to analysis of DNA samples collected from peripheral blood lymphocytes, which may not accurately represent the target tissues affected in CLP. Additionally, our experimental approaches were incapable of detecting balanced variants such as those resulting from inversions and translocations, and chromosomal aberrations that involve regions with only homozygous alleles. Routine karyotyping which can generally detect such chromosomal abnormalities could not be performed due to unavailability of living cells from a blood sample.

We are unaware of definitive reports of postzygotic genomic rearrangements underlying MZ twin discordance. A recent report of copy number variants arising between MZ twins during somatic development is consistent with our rationale for using discordant MZ twins in disease gene identification (Bruder et al., 2008). Changes in the somatic genome are well recognized as a source of diversity within the immunoglobulin and T-receptor genes (Kastern & Kryspin-Sorensen, 1988). In disease states, rearrangements can induce a phenotype by directly interrupting a gene sequence, altering gene dosage, or gene expression through position effects (Lupski & Stankiewicz, 2005). Mitotic recombination is especially relevant in tumor development, since it can lead to the expression of recessive tumor suppressor genes and/or amplification of protooncogenes (Gupta et al., 1997). Analysis of MZ twins concordant for cancer can reveal DNA rearrangements that are common to both twins as potential candidates for susceptibility loci (el-Rifai et al., 1999). In addition to oncogenesis, somatic mutations mediated through LOH can potentially contribute to other biological processes such as aging (Grist et al., 1992).