nature genetics

Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate

Elisabeth Mangold[1,26], Kerstin U Ludwig[1,2,26], Stefanie Birnbaum[1,26], Carlotta Baluardo[3], Melissa Ferrian[3], Stefan Herms[1,2], Heiko Reutter[1], Nilma Almeida de Assis[1], Taofik Al Chawa[1], Manuel Mattheisen[1,2,4], Michael Steffens[4], Sandra Barth[1,2], Nadine Kluck[1,2], Anna Paul[5], Jessica Becker[1,2], Carola Lauster[6], Gül Schmidt[6], Bert Braumann[7], Martin Scheer[8], Rudolf H Reich[9], Alexander Hemprich[10], Simone Pötzsch[11], Bettina Blaumeiser[12], Susanne Moebus[13], Michael Krawczak[14], Stefan Schreiber[14], Thomas Meitinger[15], Hans-Erich Wichmann[16–18], Regine P Steegers-Theunissen[19–23], Franz-Josef Kramer[24], Sven Cichon[1,2], Peter Propping[1], Thomas F Wienker[4], Michael Knapp[4], Michele Rubini[3], Peter A Mossey[25], Per Hoffmann[1,2] & Markus M Nöthen[1,2]

© 2010 Nature America, Inc. All rights reserved.

We conducted a genome-wide association study for nonsyndromic cleft lip with or without cleft palate (NSCL/P) in 401 affected individuals and 1,323 controls, with replication in an independent sample of 793 NSCL/P triads. We report two new loci associated with NSCL/P at 17q22 (rs227731, combined $P = 1.07 \times 10^{-8}$, relative risk in homozygotes = 1.84, 95% CI 1.34–2.53) and 10q25.3 (rs7078160, combined $P = 1.92 \times 10^{-8}$, relative risk in homozygotes = 2.17, 95% CI 1.32–3.56).

NSCL/P is one of the most common human birth defects. In European populations, NSCL/P has a prevalence ranging from 1 in 700 to 1 in 1,000. We recently reported a susceptibility locus for NSCL/P at chromosome 8q24.21 from a genome-wide association study in 224 individuals with NSCL/P (cases) and 383 population-based controls[1]. This locus is the second susceptibility locus to have been unequivocally identified for NSCL/P to date, the first being the *IRF6* locus[2].

To identify additional cleft susceptibility loci, we enlarged our sample by genotyping an additional set of 177 NSCL/P cases and adding the genotypes of 940 population-based controls of central European origin. Genotyping was performed using Illumina BeadChips (Human610-Quad and HumanHap 550k).

Following quality control (**Supplementary Methods** and **Supplementary Fig. 1**), association analysis of 521,288 SNPs having a minor allele frequency (MAF) of ≥1% in controls was performed in 399 cases and 1,318 controls.

After excluding markers from the previously described 8q24.21 locus, 20 SNPs with $P < 10^{-5}$ remained. Five chromosomal loci (8q12.3, 10q25.3, 13q31.1, 15q13.3 and 17q22) were located within these 20 top SNPs, and the associations at these loci were further supported by at least three more SNPs with $P < 10^{-4}$ (**Supplementary Fig. 2** and **Supplementary Table 1**). Two additional regions were considered to be promising NSCL/P susceptibility loci (6p22.1, 11q14.2), as they contained at least four markers with $P < 10^{-4}$.

To replicate the genome-wide association study (GWAS) findings, we selected the 20 top SNPs ($P < 10^{-5}$) as well as additional backup markers for each of the seven previously mentioned loci, resulting in two replication assays. We included additional SNPs with $P < 10^{-4}$ in the two replication assays, giving highest priority to SNPs with the lowest *P* values. Thus, a total of 56 markers were genotyped in a replication sample of 793 NSCL/P triads of European origin. Genotyping using matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Sequenom Inc.) was successful for 45 markers (representing 32 different loci), which were then analyzed by the transmission-disequilibrium test in 665 triads (128 triads were excluded after quality control, **Supplementary Methods**).

Of the 45 SNPs successfully genotyped, 11 (representing six different loci) showed $P < 0.05$ in the replication sample (**Supplementary Table 2**). Two of these SNPs remained significant after correction for multiple testing by a conservative Bonferroni procedure (17q22: rs227731, $P_{corr} = 0.01039$ and 10q25.3: rs7078160, $P_{corr} = 0.04999$). The probability that 6 or more out of 32 loci would generate *P* values <0.05 by chance alone is 0.0046. It is therefore likely that true association was detected. After combining the GWAS and replication samples, genome-wide significant evidence of association was found

[1]Institute of Human Genetics and [2]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [3]Department of Experimental and Diagnostic Medicine, Medical Genetics Unit, University of Ferrara, Ferrara, Italy. [4]Institute of Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany. [5]Department of Otolaryngology, Asklepios Klinik Altona, Hamburg, Germany. [6]Department of Cleft Lip and Cleft Palate Surgery, Humboldt University of Berlin, Berlin, Germany. [7]Department of Orthodontics and [8]Department of Oral and Maxillofacial Surgery, University of Cologne, Cologne, Germany. [9]Department of Oral and Maxillofacial Plastic Surgery, University of Bonn, Bonn, Germany. [10]Department of Oral and Maxillofacial Surgery, University of Leipzig, Leipzig, Germany. [11]Monitoring of Congenital Malformations Saxony Anhalt, University of Magdeburg, Magdeburg, Germany. [12]Department of Medical Genetics, University Hospital and University of Antwerp, Antwerp, Belgium. [13]Institute of Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, University Duisburg-Essen, Essen, Germany. [14]PopGen Biobank, Christian-Albrechts University Kiel, Kiel, Germany. [15]Department of Human Genetics, Klinikum rechts der Isar der Technischen Universität München, Munich, Germany. [16]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [17]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [18]Klinikum Grosshadern, Munich, Germany. [19]Department of Obstetrics and Gynecology, [20]Department of Epidemiology, [21]Department of Pediatrics and [22]Department of Clinical Genetics, Erasmus Medical Center, University Medical Center, Rotterdam, The Netherlands. [23]Department of Epidemiology, Radboud University Medical Center, Nijmegen, The Netherlands. [24]Department of Oral and Maxillofacial Surgery, University of Göttingen, Göttingen, Germany. [25]Dental Hospital and School, University of Dundee, Dundee, UK. [26]These authors contributed equally to this work. Correspondence should be addressed to E.M. (e.mangold@uni-bonn.de).

Received 29 July; accepted 16 November; published online 20 December 2009; doi:10.1038/ng.506

Table 1 GWAS and replication for the most significantly associated markers at five NSCL/P susceptibility loci

SNP Position in build 36	Chr.	Alleles[a]	Sample	MAF[b] cases/ MAF controls	T/NT[c]	*P* value[d]	RR_{het}[e]	95% CI[f]	RR_{hom}[g]	95% CI[h]
rs227731 52128237	17	A/**C**	GWAS	0.496 (A)/0.416 (C)		1.58×10^{-5}	1.55	1.18–2.03	1.98	1.44–2.72
			Replication		349/258	2.21×10^{-4}	1.26	0.99–1.60	1.84	1.34–2.53
			Combined			$\mathbf{1.07 \times 10^{-8}}$	1.38	1.21–1.56	1.91	1.63–2.24
rs7078160 118817550	10	**A**/G	GWAS	0.233 (A)/0.163 (A)		9.50×10^{-6}	1.40	1.10–1.79	2.91	1.74–4.84
			Replication		243/176	1.06×10^{-3}	1.32	1.05–1.65	2.17	1.32–3.56
			Combined			$\mathbf{1.92 \times 10^{-8}}$	1.36	1.21–1.53	2.50	1.95–3.21
rs7590268 43393629	2	**G**/T	GWAS	0.297 (G)/0.221 (G)		1.11×10^{-5}	1.59	1.26–2.02	1.97	1.26–3.10
			Replication		235/179	5.92×10^{-3}	1.26	1.00–1.60	1.85	1.18–2.90
			Combined			8.62×10^{-8}	1.42	1.26–1.59	1.95	1.56–2.44
rs9574565 79566875	13	**C**/T	GWAS	0.185 (T)/0.263 (T)		9.22×10^{-6}	1.34	0.75–2.40	2.19	1.24–3.85
			Replication		227/177	1.29×10^{-2}	1.25	0.77–2.03	1.62	0.97–2.68
			Combined			3.44×10^{-7}	1.31	1.01–1.70	1.89	1.45–2.47
rs1258763 30837715	15	**A**/G	GWAS	0.247 (G)/0.337 (G)		1.30×10^{-6}	2.42	1.42–4.12	3.32	1.96–5.62
			Replication		269/223	3.81×10^{-2}	1.15	0.80–1.64	1.41	0.95–2.10
			Combined			1.14×10^{-6}	1.52	1.23–1.86	2.00	1.62–2.49

Chr., chromosome.

[a]Risk allele in bold. [b]Minor allele frequency. [c]Transmission/non-transmission of risk allele. [d]Genome-wide significant *P* values in bold. [e]Relative risk for the heterozygous genotype. [f]95% confidence interval, RR_{het}. [g]Relative risk for the homozygous genotype. [h]95% confidence interval, RR_{hom}.

© 2010 Nature America, Inc. All rights reserved.

using the combined haplotype relative risk method[3] for three SNPs at two loci (17q22: rs227731, $P_{comb} = 1.07 \times 10^{-8}$; and 10q25.3: rs7078160, $P_{comb} = 1.92 \times 10^{-8}$ and rs4752028, $P_{comb} = 2.48 \times 10^{-8}$) (**Table 1**, **Supplementary Table 3** and **Supplementary Methods**). Two further loci (13q31.1, 15q13.3) were replicated, although they fell short of achieving genome-wide significance (13q31.1: rs9574565, $P_{comb} = 3.44 \times 10^{-7}$ and 15q13.3: rs1258763, $P_{comb} = 1.14 \times 10^{-6}$). Similarly, the single marker rs7590268 on 2p21 failed to reach genome-wide significance in the combined sample ($P_{comb} = 8.62 \times 10^{-8}$). For the two significant markers on 6q27 (rs2197100 and rs7740603, **Supplementary Table 2**), the putative risk alleles in the replication study differed from those in the GWAS, and these markers were therefore not considered to have been replicated. There was no evidence for imprinting or a maternal genotype effect for any of the 45 SNPs included in the family-based replication step (**Supplementary Methods**).

The relative risk (RR) in the replication sample for rs227731 (17q22) was 1.26 (95% CI 0.99–1.6) for the heterozygous genotype and 1.84 (95% CI 1.34–2.53) for the homozygous genotype. The RR in the replication sample for rs7078160 (10q25.3) was 1.32 (95% CI 1.05–1.65) for the heterozygous genotype and 2.17 (95% CI 1.32–3.56) for the homozygous genotype (**Table 1**).

The population attributable risk (PAR) estimated from the combined sample was 23.9% for rs227731 (17q22) and 12.3% for rs7078160 (10q25.3). The joint PAR for these two new loci, the key susceptibility locus at 8q24.21 and the *IRF6* locus estimated from the combined sample was 54.6% (**Supplementary Methods**). Although there may be biases in estimating the PAR from these discovery cohort samples, these results suggest that together these four loci may explain a substantial proportion of the risk for NSCL/P. However, because the summary PAR is usually less than the sum of the individual PARs, it is also likely that additional genetic variants contribute to NSCL/P risk and remain to be identified.

The chromosome 17q22 region contains the gene encoding the noggin protein, *NOG*, which is located 100 kb centromeric of rs227731 (**Fig. 1a**). NOG is an antagonist of members of the transforming growth factor beta (TGF-β) superfamily, which includes proteins such as the bone morphogenetic protein 4 (BMP4). *BMP4* has been

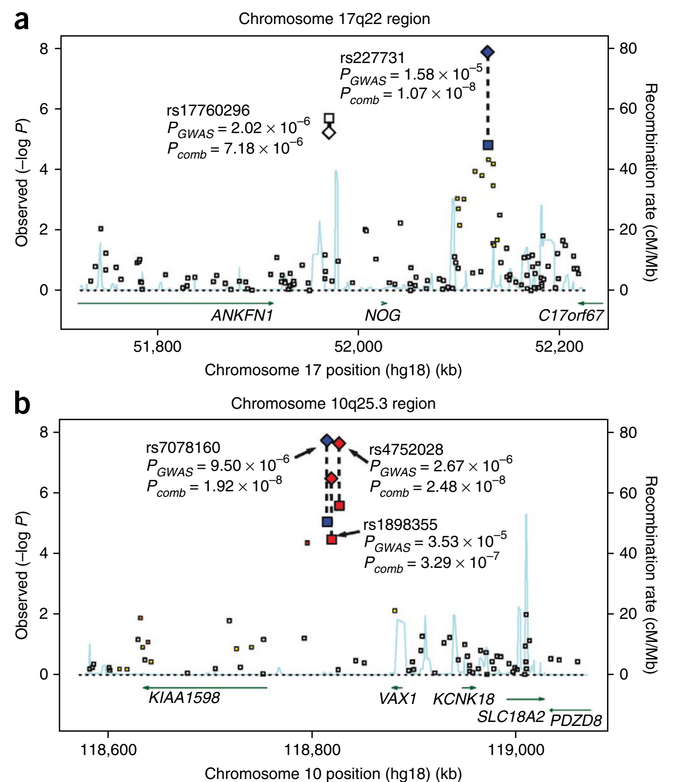

Figure 1 Details of the loci showing genome-wide significant association with NSCL/P in the combined sample. (**a**,**b**) Each panel shows single-marker association statistics (as $-\log_{10} P$; left *y* axis) from the GWAS (squares) and from the combined analysis (diamonds). Linkage disequilibrium (r^2) to the most significantly associated regional SNP (blue diamond), as estimated from the GWAS control genotypes, is color-coded (red: $r^2 > 0.8$; orange: $r^2 = 0.5$–0.8; yellow: $r^2 = 0.2$–0.5; white: $r^2 < 0.2$). Recombination rates across each region in HapMap CEU are shown in light blue (right *y* axis). The chromosomal locations and relative positions of genes according to hg18 are shown (*x* axis).

shown to regulate mammalian palatogenesis[4] and has been reported to be associated with clefting in humans[5].

The four markers at 10q25.3 with a $P_{GWAS} < 10^{-4}$ are located within a 30-kb region and show high intermarker linkage disequilibrium (**Fig. 1b**). Two genes are located in close vicinity to this region: *KIAA1598* (40 kb centromeric) and *VAX1* (ventral anterior homeobox 1; 53 kb telomeric). Mice with homozygous *Vax1* mutations display craniofacial malformations including cleft palate[6]. Two individuals with a 10q terminal deletion syndrome with breakpoints in 10q25 have been reported, one with a submucous cleft palate[7] and the other with a cleft lip[8].

Three further loci were successfully replicated but did not achieve genome-wide significance in the combined analysis. *FMN1* (formin-1), a gene with an unknown function, and *GREM1* (gremlin-1), the gene coding for another known antagonist of BMP4, are located on chromosome 15q13.3 (**Supplementary Fig. 3a**). The 13q31.1 locus lies in a gene desert (**Supplementary Fig. 3b**). *SPRY2* (sprouty homolog 2) is located 241 kb telomeric of rs9574565. In mice, palate development is sensitive to *Spry2* dosage[9], and *Spry2* overexpression results in craniofacial defects[10]. Resequencing of *SPRY2* in NSCL/P cases has suggested the presence of rare, and possibly detrimental, variants in *SPRY2* (ref. 11). The SNP on 2p21 (rs7590268) is located within intron 31 of *THADA* (thyroid adenoma associated), which may be involved in the cell death receptor pathway and apoptosis[12] (**Supplementary Fig. 3c**). Notably, rs7590268 maps to a region that was duplicated in two individuals presenting with cleft palate and other anomalies[13,14].

The most significant SNP at the 8q24.21 locus[1] (rs987525) and the functional *IRF6* variant[2] (rs642961) were also genotyped in the replication sample in order to test for a possible interaction. Suggestive evidence ($P = 0.005$) was found for an interaction between the *IRF6* variant rs642961 and the SNP near *GREM1* at 15q13.3 (rs1258763). No evidence was found for any other interactions.

Recent genetic data[2] support the hypothesis that NSCL/P may be separable into two sub-phenotypes: cleft lip only (NSCLO) and cleft lip with cleft palate (NSCLP). The genotype distribution in the NSCLP subsample (318 cases) did not differ significantly from that in the NSCLO subsample (81 cases) (data not shown) for any of the nine reported SNPs at the five newly identified loci (**Supplementary Table 3**).

Because NSCL/P and nonsyndromic cleft palate only (NSCPO) may have an etiological overlap, we genotyped the replication marker panel in 295 NSCPO triads. None of the SNPs showed evidence of association (**Supplementary Table 4**).

In summary, we have identified two new NSCL/P susceptibility loci with genome-wide significance on 17q22 and 10q25.3, and three further loci (13q31.1, 15q13.3 and 2p21) for which there is suggestive evidence. Promising candidate genes at these loci include *NOG* (noggin), *VAX1* (ventral anterior homeobox 1), *GREM1* (gremlin 1), *SPRY2* (sprouty homolog 2) and *THADA* (thyroid adenoma associated). Given the intergenic location of the associated SNPs in the present study, further studies should test for allele-specific expression of these candidate genes and resequence their coding regions in order to identify possible functional variants.

Accession numbers. GenBank: *BMP4*, NM_001202 (NM_130850, NM_130851); *FMN1*, NM_001103184; *GREM1*, NM_013372; *IRF6*, NM_006147; *KIAA1598*, NM_018330 (NM_001127211); *NOG*, NM_005450; *SPRY2*, NM_005842; *THADA*, NM_022065; *VAX1*, NM_199131 (NM_001112704).

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS
We thank all affected individuals and their families for their participation in this study as well as the German support group for people with cleft lip and/or palate (Deutsche Selbsthilfevereinigung für Lippen-Gaumen-Fehlbildungen e.V.). We acknowledge our collaborating clinical partners, all colleagues who contributed to the European Collaboration on Craniofacial Anomalies (EUROCRAN) project, our laboratory technicians and colleagues responsible for database management (for a complete list of all individuals see **Supplementary Note**). The study was supported by the Deutsche Forschungsgemeinschaft (FOR 423 and individual grants MA 2546/3-1, KR 1912/7-1, NO 246/6-1 and WI 1555/5-1). The Heinz Nixdorf Recall cohort was established with the generous support of the Heinz Nixdorf Foundation, Germany (Chairman: G. Schmidt). POPGEN biobank is financed by the local Ministry of Science, Economy and Transport of Schleswig-Holstein, Germany. The Kooperative Gesundheitforschung in der Region Augsburg (KORA) research platform was initiated and financed by the Helmholtz Center Munich, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the state of Bavaria. The work of KORA is supported by the German Federal Ministry of Education and Research (BMBF) in the context of the German National Genome Research Network (NGFN-2 and NGFN-plus). The EUROCRAN study was supported by the European Commission FP5 (contract no. QLG1-CT-2000-01019), and the ITALCLEFT study was supported by a FAR-2008 grant from the University of Ferrara, Italy. N.A.d.A. is supported by the Conselho Nacional de Pesquisa (CNPq), Brasil. T.A.C. is supported by a grant from the Ministry of Higher Education, Syrian Arab Republic. Additional acknowledgments are found in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS
E.M., F.-J.K., T.F.W., P.P. and M.M.N. initiated the study. E.M., S. Birnbaum, K.U.L., P.H., M. Knapp, M.R., P.A.M. and M.M.N. contributed to the study design. M.M.N., E.M., S.C., P.H., K.U.L. and S. Birnbaum coordinated the work and prepared the manuscript, with feedback from the other authors. S. Birnbaum., H.R., A.P., C.L., G.S., M. Scheer., B. Braumann, R.H.R., A.H., S.P., B. Blaumeiser, R.P.S.-T., F.-J.K., M.R. and P.A.M. clinically characterized the cleft families and collected blood samples. S.M., M. Krawczak, S.S., T.M. and E.M. characterized and recruited the controls. K.U.L., C.B., M.F., N.A.d.A., T.A.C., S. Barth, N.K. and J.B. prepared the DNA and performed the molecular genetic experiments. M. Knapp., S.H., M. Steffens and M.M. conducted the statistical analysis. M.M.N., E.M., M. Knapp., T.F.W., S. Birnbaum, K.U.L., M.R. and P.P. analyzed and interpreted the data.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Birnbaum, S. *et al. Nat. Genet.* **41**, 473–477 (2009).
2. Rahimov, F. *et al. Nat. Genet.* **40**, 1341–1347 (2008).
3. Guo, C.Y. *et al. Genet. Epidemiol.* **33**, 54–62 (2009).
4. Zhang, Z. *et al. Development* **129**, 4135–4146 (2002).
5. Suzuki, S. *et al. Am. J. Hum. Genet.* **84**, 406–411 (2009).
6. Hallonet, M. *et al. Genes Dev.* **13**, 3106–3114 (1999).
7. Petersen, B. *et al. Am. J. Med. Genet.* **77**, 60–62 (1998).
8. Mulcahy, M.T. *et al. Clin. Genet.* **21**, 33–35 (1982).
9. Welsh, I.C. *et al. Mech. Dev.* **124**, 746–761 (2007).
10. Goodnough, L.H. *et al. Dev. Dyn.* **236**, 1918–1928 (2007).
11. Vieira, A.R. *et al. PLoS Genet.* **1**, e64 (2005).
12. Drieschner, N. *et al. Gene* **403**, 110–117 (2007).
13. Yunis, E. *et al. Hum. Genet.* **48**, 241–244 (1979).
14. Magee, A.C. *et al. Clin. Genet.* **54**, 65–69 (1998).

© 2010 Nature America, Inc. All rights reserved. npg