Sequencing the *GRHL3* coding region reveals rare truncating mutations a common susceptibility variant for nonsyndromic cleft palate

Elisabeth Mangold,[1,17] Anne C. Böhmer,[1,2] Nina Ishorst,[1,2] Ann-Kathrin Hoebel,[1,2] Pinar Gültepe,[1,2] Hannah Schuenke,[1,2] Johanna Klamt,[1,2] Andrea Hofmann,[1,2] Lina Gölz,[3] Ruth Raff,[1] Peter Tessmann,[1,2] Stefanie Nowak,[1] Heiko Reutter,[1,2,4] Alexander Hemprich,[5] Thomas Kreusch,[6] Franz-Josef Kramer,[7] Bert Braumann,[8] Rudolf Reich,[9] Gül Schmidt,[10] Andreas Jäger,[3] Sibylle Brosch,[11] Janis Stavusis,[12] Miho Ishida,[13] Rimante Seselgyte,[13] Gudrun E. Moore,[13] Markus M. Nöthen,[1,2] Guntram Borck,[14] Khalid A. Aldhorae,[15] Baiba Lace,[12] Philip Stanier,[13] Michael Knapp,[16] Kerstin U. Ludwig[1,2]

[1]Institute of Human Genetics, University of Bonn, Bonn 53127, Germany; [2]Department of Genomics, Life & Brain Center, University of Bonn, Bonn 53127, Germany; [3]Department of Orthodontics, University of Bonn, Bonn 53127, Germany; [4]Department of Neonatology and Pediatric Intensive Care, Children's Hospital, University of Bonn, Bonn 53113, Germany; [5]Clinic for Maxillofacial Surgery, University Hospital Leipzig, 04103 Leipzig, Germany; [6]Department of Oral and Maxillofacial Surgery, Head and Neck Centre, Asklepios Klinik Nord, Heidberg, 22417 Hamburg, Germany; [7]Department of Oral and Maxillofacial Surgery, University of Göttingen, 37075 Göttingen, Germany; [8]Department of Orthodontics, University of Cologne, 50931 Cologne, Germany; [9]Department of Oral and Maxillo-Facial-Plastic Surgery, University of Bonn, 53111 Bonn, Germany; [10]Department of Cleft Lip and Cleft Palate Surgery, Humboldt University of Berlin, 13353 Berlin, Germany; [11]Section of Phoniatrics and Pedaudiology, Department of Otolaryngology - Head and Neck Surgery, University of Ulm, 89070 Ulm, Germany; [12]Medical Genetics and Mitochondrial Research Group; Latvian Biomedical Research and study Centre, Riga LV-1067, Latvia ; [13] Genetics and Genomic Medicine Programme, Institute of Child Health; University College London, London WC1N 1EH, UK ; [14]Institute of Human Genetics, Ulm University, Ulm 89073, Germany; [15]Orthodontic Department, College of Dentistry, Thamar University, Thamar, Yemen; [16]Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn 53127, Germany.

[17]Present address: Institute of Human Genetics, Biomedical Center, University of Bonn, Sigmund-Freud-Strasse 25, D-53127 Bonn, Germany.

Correspondence: e.mangold@uni-bonn.de, kludwig1@uni-bonn.de

Abstract

Nonsyndromic cleft lip with/without cleft palate (nsCL/P) and nonsyndromic cleft palate only (nsCPO) are the most frequent subphenotypes of orofacial clefts. A common syndromic form of orofacial clefting is Van der Woude syndrome (VWS) where individuals have CL/P or CPO, often but not always associated with lower lip pits. Recently, ~5% of VWS-affected individuals were identified with mutations in the grainy head-like 3 gene (*GRHL3*). To investigate *GRHL3* in nonsyndromic clefting, we sequenced its coding region in 576 Europeans with nsCL/P and 96 with nsCPO. Most strikingly, nsCPO-affected individuals had a higher minor allele frequency for rs41268753 (0.099) than control subjects (0.049; P=1.24×10^{-2}). This association was replicated in nsCPO/control cohorts from Latvia, Yemen and the UK ($P_{combined}$ 2.63×10^{-5}; $OR_{allelic}$ 2.46 (95% CI 1.6-3.7)), and reached genome-wide significance in combination with imputed data from a GWAS in nsCPO triads (P=2.73×10^{-9}). Notably, rs41268753 is not associated with nsCL/P (P=0.45). rs41268753 encodes the highly conserved p.Thr454Met (c.1361C> T) (GERP=5.3) which prediction programs denote as deleterious, has a CADD score of 29.6 and increases protein binding capacity in silico. Sequencing also revealed four novel truncating *GRHL3* mutations including two that were de novo in four families, where all nine individuals harboring mutations had nsCPO. This is important for genetic counseling: given that VWS is rare compared to nsCPO, our data suggests that dominant *GRHL3* mutations are more likely to cause nonsyndromic than syndromic CPO. Thus, with rare dominant mutations and a common risk variant in the coding region, we have identified an important contribution for *GRHL3* in nsCPO.

Main text

Orofacial clefts are among the most common congenital malformations in humans.[1] The two common distinct cleft phenotypes are cleft lip with or without cleft palate (CL/P) and cleft palate only (CPO).[2] These two phenotypes most often occur without additional clinical features and are described as nonsyndromic clefts (i.e. nsCL/P and nsCPO) that present with a multifactorial etiology from both diverse genetic and environmental factors.[2] Based on previous epidemiological observations and also genetic findings, nsCL/P and nsCPO are considered to have only very limited overlap in terms of their genetic etiology. This is further supported by findings from genome-wide association studies (GWASs), which have identified 15 risk loci for nsCL/P[3-9] although no genome-wide significant findings have been described for nsCPO.[10] The only risk locus for which there is evidence of an overlapping effect for both phenotypes is a region on 9q21, which had been initially identified in a linkage study of nonsyndromic clefting families.[11; 12] Subsequent fine mapping identified two SNPs around *FOXE1* (MIM: 602617) that potentially accounted for linkage to both nsCL/P and nsCPO, with larger effect sizes in nsCL/P.[13] This finding was independently confirmed,[14] although, association with nsCPO did not reach genome-wide significance.

CL/P and CPO can also occur in the context of complex malformation syndromes. One of the most common of these is Van der Woude Syndrome (VWS [MIM: 119300]). Mutations in the interferon regulatory factor 6 (*IRF6*) gene (MIM: 607199) account for ~70% of individuals with VWS, and recently grainy head-like 3 (*GRHL3* [MIM: 608317]) mutations were identified in a further ~5%.[15] Individuals with VWS have either a CL/P or a CPO, with pits of the lower lip being the sole additional symptom distinguishing individuals from those with nonsyndromic clefts. Lip pits seen in VWS-affected individuals have an incomplete penetrance or can be subtle and sometimes difficult to recognize, especially after surgical treatment. Thus, individuals with an *IRF6* or *GRHL3* mutation might sometimes be considered to have a nonsyndromic cleft. Consequently several studies have screened apparently nonsyndromic clefting for *IRF6* mutations to investigate their prevalence.[16; 17] These studies showed that the majority of individuals carrying *IRF6* mutations but classified as nonsyndromic turned out to have lip pits when a posteriori reviewed, i.e. they were misdiagnosed VWS.[17] Only a minority (0.24–0.44% of the families) remained truly nonsyndromic from a clinical perspective. No such studies have yet been presented for *GRHL3*. Of note, among individuals with *GRHL3* mutations, CPO is more frequent, while both CL/P and lip pits are less frequent compared to individuals with *IRF6* mutations.[15] Therefore, we speculated that individuals harboring *GRHL3* mutations would be even more likely to "mimic" nonsyndromic clefts.

Based on this hypothesis, we evaluated a large sample of individuals with nonsyndromic clefts for mutations in *GRHL3*. Of note, recent calculations on genic intolerance toward mutations in data from the Exome Aggregation Consortium (EXAC)[18] reveals that *GRHL3* is extremely intolerant towards loss-of-function mutations (pLI=1) and also shows some degree of intolerance toward missense mutations. The *GRHL3* coding region was Sanger-sequenced in a total of 672 unrelated individuals with clefting of Central European origin with either nsCL/P (576 individuals) or nsCPO (96 individuals). This study was approved by the Ethics Committee of the Medical Faculty of the University of Bonn. Informed consent was obtained from all participants or their legal representatives, which H also applies to all further samples used throughout this study. For these, study approval was obtained from the respective institutional review boards.

Sequencing was performed at Macrogen using 13 amplicons that targeted all 18 exons and UTRs of the four known transcripts of *GRHL3* (**Table S1**). Data analysis was performed using the SeqMan pro software provided by DNASTAR. Rare variants with minor allele frequencies (MAF) < 1% were subsequently confirmed in-house using BigDye v3.1 kit and an ABI 3130XL Capillary Sequencer.

Sequencing revealed 30 variants in total, 18 of which were rare in our sample (**Table S2**). Twelve variants were more frequent (MAF ≥ 1%, **Table S3**) and already assigned rs numbers. For all of the more common variants, MAFs were comparable between individuals with clefts and European control subjects in public databases (ExAC, 1000 genomes,[19] EVS and GoNL,[20] Web Resources, **Table S3**). However, when dividing the cleft cohort according to their cleft phenotype (nsCL/P or nsCPO), one of the variants, rs41268753, was observed to have a significantly higher MAF of 0.99 in individuals with ns CPO compared to control subjects. The highest MAF for this variant among the four control cohorts was reported in the GoNL dataset (genotype distribution 0/41/457, MAF = 0.041, Armitage trend test for association with rs41268753: $P = 5.79\times10^{-04}$). Given the population-specific occurrence of rare and low-frequency variants and to rule out population stratification[21] we genotyped rs41268753 in an additional 267 controls of Central European origin, using a KASP assay on demand (LGC, **Table 1**). In this dataset, the MAF for rs41268753 was 0.049 and therefore slightly higher than indicated in the public databases, although the association remained significant (p = 0.0124, **Table 1**).

In order to replicate our finding, we evaluated three independent nsCPO case-control cohorts for association of rs41268753 (**Table 1**). These included: (1) 47 individuals with nsCPO and 187 controls from Yemen, (2) 94 individuals with nsCPO and 177 controls of European ethnicity recruited in the London area and (3) 51 individuals with nsCPO and 94 controls from Latvia. In each of the samples, an association was found ($p < 0.1$), and the combination of the initial sample with the replication samples by fixed effects meta-analysis[22] resulted in a $P_{combined}$ of 2.63×10^{-5} with an $OR_{allelic}$ of 2.5 (95% CI 1.6-3.7, **Table 1**). To further evaluate this finding, we imputed GWAS data from nsCPO triads of European origin[10] that we had downloaded from dbGaP upon approved data access. Imputation was performed using IMPUTE2,[23] and analysis of imputed data ($p = 1.18\times10^{-4}$) was performed according to the FBATdosage method.[24] The imputed data ($p = 1.18\times10^{-4}$) confirmed our previous observation gained in the genotyped samples for rs41268753. Combining the genotyped datasets with the imputation by the Z-score method[24] led to a genome-wide significant association signal for rs41268753 in nsCPO individuals ($p = 2.73\times10^{-9}$).

Next, we evaluated rs41268753 in subjects with submucous cleft palate (SMCP), which is considered a mild subphenotype of cleft palate, and has also been described in families with VWS and dominant *GRHL3* mutations.[15] We genotyped 116 Central European patients with NSSMCP using PCR followed by direct genotyping, but did not find a significant association when compared to the 267 Central European controls (p = 0.94, **Table 1**). Given the statistical power to detect an association of the effect size found in the nsCPO sample (0.87, calculated according to Jackson et al. 2002)[21], we conclude that rs41268753 does not significantly contribute to SMCP.

The newly identified susceptibility variant, rs41268753, is present in all four known isoforms of *GRHL3* and encodes the p.Thr454Met (c.1361C>T) missense alteration (GenBank: NP_937817.3), located between the DNA binding and the dimerization domain of GRHL3.[15] Analysis of rs41268753 using the Variant Effect Predictor generated strong evidence for deleteriousness ("probably damaging" by PolyPhen-2, "deleterious" by SIFT, MutationTaster, Provean, Condel). Moreover, the variant occurs at a highly conserved nucleotide position (GERP score 5.27) and the CADD score is reported to be 29.6. In the absence of any experimental structures of the GRHL3 protein, we attempted to use Swiss-Model[26] for protein structure homology modeling. However, available templates included only the first hundreds of amino acids, making modeling at position 454 impossible. We next aimed at assessing the functional impact of the p.Thr454Met variant for (1) secondary structure, and (2) protein stability via PSIPRED Protein Sequence Analysis Workbench[27] and I-Mutant 2.0,[28] respectively. Although these results did not show a measurable effect of the p.Thr454Met variant on protein stability, we did observe a more "disorganized" state and an

increased protein binding capacity[27,29] for the corresponding part of the GRHL3 protein (**Figure 1 and S1, Table S4**). This strongly suggests a functional effect implicating rs41268753 as a causative variant at the *GRHL3* locus.

Given the evidence implicating a functional role for rs41268753 in nsCPO, we then evaluated its possible role in nsCL/P, using both the sequencing data generated in the present study and imputed data from our large in-house study that was comprised of 399 individuals with nsCL/P and 1,318 control subjects.[3] In our sequencing study, the rs41268753 major (non-risk) allele was overrepresented in the nsCL/P group when compared to the 267 Central European control subjects ($p = 6.31\times10^{-2}$), and similar results were obtained in the genome-wide imputed nsCL/P data showing no association (*P*=0.45). These findings indicate that rs41268753 does not confer an effect in individuals with nsCL/P and is therefore the first risk locus that uniquely contributes to nsCPO.

We hypothesized that common variant(s) located in the regulatory region up- or downstream of *GRHL3* might also contribute to either nsCL/P or nsCPO. This would parallel the situation described for *IRF6*, where the promoter variant rs642961 upstream of *IRF6* affects binding of the transcription factor AP2alpha and increases risk for nsCL/P.[9] We therefore analyzed the imputed data from both genome-wide datasets on nsCL/P[3] and nsCPO[10] for the *GRHL3* coding region (GenBank: NM_198174.2; chr1: 24,645,812-24,690,970) up to +/- 200 kb. All variants with SNP info scores > 0.4 and a MAF > 1% were extracted and plotted with LocusZoom,[30] which resulted in 1,401 variants for nsCL/P and 1,379 variants for nsCPO (**Figure S2**). In nsCL/P, the lowest p value observed was p = 0.0024 for rs4648988, located in the adjacent *NIPAL3* gene. This association did not withstand correction for multiple testing, suggesting that common variants at the *GRHL3* locus do not contribute to nsCL/P etiology. For nsCPO, the coding variant rs41268753 described above provided the strongest association, followed by a second variant, rs113965554 that is in high linkage disequilibrium ($r^2 > 0.8$, **Figure S2**). The latter is located within an intron of the adjacent *STPG1* gene (MIM: 615826), which could indicate a regulatory effect. Notably, neither rs41268753 nor rs113965554 were present on any of the genome-wide arrays used for nsCPO genotyping so far, potentially explaining why the locus had not been detected up to now. However, the comprehensive annotation software RegulomeDB v1.1[31] annotated rs113965554 with a score of 6, thus not indicating a regulatory effect. Collectively, our data of do not support an association of regulatory common variants at the *GRHL3* locus with nsCPO.

Apart from the common variants, our sequencing approach in 672 Central European individuals with nonsyndromic clefting identified 18 rare variants with allele frequencies below 1% (**Table S2**), six of them not listed in dbSNP142. Among these six variants were four novel (unlisted in EVS, 1kGP, ExAc or GoNL) truncating *GRHL3* mutations in four index individuals from unrelated families, of which two occurred de novo (**Table 2**, **Figure S3**). Three of these novel variants were splice site mutations occurring in three unrelated individuals with nsCPO (**Table S2**). For variant c.840+1G>T (p.?) (family BN-042), analysis of parental DNA revealed de novo occurrence in the index individual. Variant c.738C>T (family BN-248) is a synonymous variant (p.Gly246=) creating a cryptic splice site as suggested in silico by the splice site prediction NNSPLICE v0.9.[32] It occurred de novo in the mother and was transmitted to her daughter, and both were affected with nsCPO (**Figure S3**). Transcript analysis in blood confirmed the biological effect of the predicted splice site mutation (data not shown). In both families with de novo mutations, paternity was confirmed using a set of 16 microsatellites (PowerPlex 16 HS, Promega). In family BN-813, we identified c.1285+2delT (p.?), which is predicted to affect a splice donor site and was present in three affected individuals, i.e., two half-sisters and their mother (**Figure 2**). In a fourth family (BN-912), a frameshift mutation (c.916dupC [p.Arg306Profs*11) was identified in two affected brothers and their affected mother, whereas the unaffected brother did not carry this variant. In addition, a missense mutation (c.937C>T

[p.Arg313Trp], **Table S2**) classified as "probably damaging" (Polyphen-2) and "deleterious" (SIFT) was identified in an individual with no affected relatives. This variant was not listed in any of the queried databases, but segregation analysis detected the variant also in the unaffected mother and sibling. Therefore, although it is not clear if this variant can be causal but with reduced penetrance, this is similar to findings reported for VWS in three of the eight families described by Peyrard and colleagues.[15] In index individual BN-136, a variant (c.17+93G>A [p.?]) was identified, which is an intronic variant in two *GRHL3* isoforms and located in the UTR of one further isoform (GenBank: NM_001195010.1). This and the fact that this variant is also present in the unaffected mother suggests that it is not causative.

In contrast to the mutations described in VWS-affected individuals, which are distributed over *GRHL3*, all probably pathogenic (i.e., truncating) mutations and the missense mutation of yet unknown significance identified in our individuals with nonsyndromic clefting have located within the *GRHL3* DNA binding domain. Further investigation will be required to determine whether or not this reflects a genuine genotype-phenotype correlation.

Because a posteriori reviews of individuals harboring *IRF6* mutations in previous studies of individuals with apparently nonsyndromic clefting have revealed unreported lip pits (and thus misdiagnosed VWS-affected families),[17] we personally re-visited all affected members of the four families with truncating mutations. According to the initial clinical data available at time of recruitment, eight of the nine individuals harboring mutations had an isolated cleft palate, and one (BN-813-II-5) was reported to be unaffected (**Figure 2**). Initial reports on affected structures were confirmed where four individuals had a cleft soft palate and four had a cleft hard and soft palate. None of these individuals showed the typical lip pits seen with VWS or even minor forms, which we also confirmed in available pre-operative photographs, and lip pits in relatives were denied. Notably, after revisiting, family member BN-813-II-5 (initially reported to be unaffected) she was found to have a broad uvula and nasal speech strongly suggestive of a submucous palatal cleft. Interestingly, her daughters (who were half-sisters BN-813-III-7 and BN-813-III-8), showed slightly asymmetrical lower lips with elevations on the left side when opening their mouths (**Figure 2**). Without a molecular diagnosis, this very subtle symptom would not have indicated VWS. However, these elevations might represent very subtle forms of lower lip pits or transverse sulci, which also have been described in VWS-affected subjects,[33] so a posteriori this might be called a "transitional" phenotype. Therefore the combination of thorough clinical and genetic assessment revealed potentially important findings with implications for counseling and diagnostics. Obviously *GRHL3* mutations are more implicated in 'nonsyndromic' forms of clefting than *IRF6*.

Another aspect concerns the family structure, where three of the four families with truncating mutations had two or three affected individuals in consecutive generations. In general, recurrence risks for nsCPO in first-degree relatives are quite low (above 0.03)[34], so multiply-affected families remain quite rare for nsCPO. In this study, 27 of the 96 sequenced nsCPO index individuals have one of more affected relative(s) (defined as either an affected first-, second- or third degree relative), which is higher than in the average population and therefore represents an ascertainment bias. Our data allow for a preliminary assumption of the frequency of truncating *GRHL3* mutations among nsCPO-affected individuals where the a priori (i.e. without knowledge of family history) chance to identify a *GRHL3* mutation in individuals with nsCPO can be estimated as 0.017 (calculated by adding the estimated chance that an individual with an nsCPO is sporadic and has a *GRHL3* mutation of 0.97 x (1: 69), to the estimated chance that an individual with an nsCPO has a positive family history and a *GRHL3* mutation of 0.03 x (3: 37)). This is lower than the reported frequency of causative *GRHL3* mutations in VWS of around 0.05.[15] However, nsCPO has a frequency of around one in 2,400,[2] whereas VWS has been estimated to account for roughly one in 30,000 to one in 100,000 in the European and Asian populations (see GeneReview in Web Resources). Based on these estimates, in

the general population one in 141,000 people has a *GRHL3* mutation with an apparent nonsyndromic CPO should outnumber *GRHL3*-related VWS-CPO-affected subjects by around 4 to 14. Thus, *GRHL3* mutations, initially identified from study of a syndrome, might in fact more often lead to an isolated, nonsyndromic cleft. Similar observations have been made for other genes, for example for *CEP290* (MIM: 610142), initially identified as a gene underlying a form of Joubert syndrome, which is associated in all patients with congenital amaurosis, and later identified as a frequent cause of isolated Leber congenital amaurosis.[36]

As a result of our sequencing study, we report here on an autosomal-dominantly inherited form of clefting in individuals with nsCPO, which has important implications for genetic counseling. For individuals with an nsCPO, and especially in those with a positive family history, the option of *GRHL3* mutation screening should be discussed when affected individuals ask for cleft risks in their offspring. Of note, the four families identified with truncating mutations in this study have up to three affected family members. It is quite conceivable that in these families some descendants will have lip pits and/or a CL/P, changing the clinical diagnosis for the family from "truly nonsyndromic" to "VWS". Unless proven otherwise, individuals with *GRHL3* mutations, regardless of whether apparently "nonsyndromic" or not, should be counseled for a considerably higher cleft recurrence risk with the potential for other VWS symptoms such as lip pits in their offspring. The risk figures based on the VWS-affected families reported by Peyrard et al.[15] are from a very limited number of observations, and our observations suggest that the true risks might be lower. It is also not known whether these risk estimates even apply to the "nonsyndromic" cleft individual with a *GRHL3* mutation. It will therefore be very valuable to perform further sequencing studies on individuals with nonsyndromic clefting, especially multiplex families, taking great care to carefully phenotype individuals with a *GRHL3* mutation.

In summary, we identified both rare dominant mutations and a common risk variant in the coding region of *GRHL3* as causative in individuals with nsCPO. At a more general level, our results demonstrate that a proportion of nsCPOs will represent a variety of monogenic syndromic forms, but with reduced penetrance of accompanying symptoms. This hypothesis is supported by the observation that GWAS and other genetic studies on nsCPO have so far lacked any convincing association findings, despite reasonable sample sizes, in particular when compared to nsCL/P. Systematic sequencing approaches such as exome or genome sequencing are likely to play an important role in the delineation of further monogenic causes of nsCPO.

Accession numbers

Four novel truncating pathogenic mutations have been deposited with the accession numbers ClinVar: SCV000258550, SCV000258551, SCV000258552, and SCV000258553. Two variants with no confirmed pathogenic effect have been deposited in dbSNP: ss1961068732 and ss1961068733. Accession numbers are listed in Table S2.

Supplemental Data

Supplemental Data include three figures, four tables, and Supple- mental Acknowledgments and can be found with this article on- line at http://dx.doi.org/10.1016/j.ajhg.2016.02.013.

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 fur Lippen-Gaumen-Fehlbildungen e.V.). We thank Prof. Bernd Potzsch for help with collecting DNA samples from anonymous blood donors. We acknowledge our collaborating clinical partners, our laboratory technicians, and colleagues responsible for data- base management. In particular, we would like to thank Susanne Raeder and Dietlinde Stienen for their technical assistance. The study was supported by the Deutsche Forschungsgemeinschaft (FOR 423, individual grants MA 2546/3-1, MA 2546/5-1, LU 1944/2-1, KR 1912/7-1, NO 246/6-1, and WI 1555/5-1). K.U.L. was supported by the BONFOR program of the Medical Faculty Bonn (O-149.0097). P.S. and G.E.M. are supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. For the UK study, we are indebted to members of the North Thames Regional Cleft Centre for both recruitment and sample collection, which was approved by the Great Ormond Street Hospital and the Institute of Child Health Research Ethics Committee (REC No. 08H0713/46) and the Imperial College Hospital Ethics Committee (REC ref: 12/WA/0196), respectively. Datasets used for the genome-wide nsCPO analysis described in this manuscript were obtained from dbGaP: phs000094.v1.p1. Supplemental Acknowledgments are available online.

Web Resources

The URLs for data presented herein are as follows:
1000 Genomes (accessed 9/12/15), http://browser.1000genomes.org
Berkeley Drosophila Genome Project NNSplice 0.9, http://www.fruitfly.org/seq_tools/splice.html
dbGaP, http://www.ncbi.nlm.nih.gov/gap
ExAC Browser (accessed 9/12/15), http://exac.broadinstitute.org/
GeneReviews, Schutte et al. (1993). IRF6-Related Disorders,
http://www.ncbi.nlm.nih.gov/books/NBK1407/
GoNL (Genomes of the Netherlands) (accessed 9/12/15), http://www.nlgenome.nl/search/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server (accessed 9/12/15),
http://evs.gs.washington.edu/EVS/
OMIM, http://www.omim.org/

PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
PSIPRED, http://bioinf.cs.ucl.ac.uk/psipred/
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq
RegulomeDB, http://RegulomeDB.org/ SIFT, http://sift.bii.a-star.edu.sg/
SWISS-MODEL, http://swissmodel.expasy.org/
UCSC Human Genome Browser, http://genome.ucsc.edu/cgi-bin/hgGateway
Variant Effect Predictor, http://useast.ensembl.org/Homo_sapiens/Tools/VEP

References

1. Mossey, P.A., and Little, J. (2002). Epidemiology of oral clefts: an international perspective. In Cleft Lip and Palate: From Origin to Treatment, D.F. Wyszynski, ed. (Oxford, Oxford University Press), pp 127–144.
2. Mangold, E., Ludwig, K.U., and Nöthen, M.M. (2011). Breakthroughs in the genetics of orofacial clefting. Trends Mol Med 17, 725-733.
3. Mangold, E., Ludwig, K.U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., Reutter, nsCL/P., de Assis, N.A., Chawa, T.A., Mattheisen, M., et al. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat Genet 42, 24-26.
4. Birnbaum, S., Ludwig, K.U., Reutter, nsCL/P., Herms, S., Steffens, M., Rubini, M., Baluardo, C., Ferrian, M., Almeida de Assis, N., Alblas, M.A., et al. (2009). Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. Nat Genet 41, 473-477.
5. Beaty, T.nsCL/P., Murray, J.C., Marazita, M.L., Munger, R.G., Ruczinski, I., Hetmanski, J.B., Liang, K.nsCL/P., Wu, T., Murray, T., Fallin, M.D., et al. (2010). A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. Nat Genet 42, 525-529.
6. Ludwig, K.U., Mangold, E., Herms, S., Nowak, S., Reutter, nsCL/P., Paul, A., Becker, J., Herberz, R., AlChawa, T., Nasser, E., et al. (2012). Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. Nat Genet 44, 968-971.
7. Beaty, T.nsCL/P., Taub, M.A., Scott, A.F., Murray, J.C., Marazita, M.L., Schwender, nsCL/P., Parker, M.M., Hetmanski, J.B., Balakrishnan, P., Mansilla, M.A., et al. (2013). Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study. Hum Genet 132, 771-781.
8. Sun, nsCL/P., Huang, nsCL/P., Yin, A., Pan, nsCL/P., Wang, nsCL/P., Wang, C., Du, nsCL/P., Wang, M., Lan, F., Hu, Z., et al. (2015). Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun 6, 6414.
9. Rahimov, F., Marazita, M.L., Visel, A., Cooper, M.E., Hitchler, M.J., Rubini, M., Domann, F.E., Govil, M., Christensen, K., Bille, C., et al. (2008). Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. Nat Genet 40, 1341-1347.
10. Beaty, T.nsCL/P., Ruczinski, I., Murray, J.C., Marazita, M.L., Munger, R.G., Hetmanski, J.B., Murray, T., Redett, R.J., Fallin, M.D., Liang, K.nsCL/P., et al. (2011). Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. Genet Epidemiol 35, 469-478.
11. Marazita, M.L., Murray, J.C., Lidral, A.C., Arcos-Burgos, M., Cooper, M.E., Goldstein, T., Maher, B.S., Daack-Hirsch, S., Schultz, R., Mansilla, M.A., et al. (2004). Meta-analysis of 13 genome

scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32-35. Am J Hum Genet 75, 161-173.
12. Marazita, M.L., Lidral, A.C., Murray, J.C., Field, L.L., Maher, B.S., Goldstein McHenry, T., Cooper, M.E., Govil, M., Daack-Hirsch, S., Riley, B., et al. (2009). Genome scan, fine-mapping, and candidate gene analysis of non-syndromic cleft lip with or without cleft palate reveals phenotype-specific differences in linkage and association results. Hum Hered 68, 151-170.
13. Moreno, L.M., Mansilla, M.A., Bullard, S.A., Cooper, M.E., Busch, T.D., Machida, J., Johnson, M.K., Brauer, D., Krahn, K., Daack-Hirsch, S., et al. (2009). FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. Hum Mol Genet 18, 4879-4896.
14. Ludwig, K.U., Böhmer, A.C., Rubini, M., Mossey, P.A., Herms, S., Nowak, S., Reutter, nsCL/P., Alblas, M.A., Lippke, B., Barth, S., et al. (2014). Strong association of variants around FOXE1 and orofacial clefting. J Dent Res 93, 376-381.
15. Peyrard-Janvid, M., Leslie, E.J., Kousa, nsCL/P.A., Smith, T.L., Dunnwald, M., Magnusson, M., Lentz, B.A., Unneberg, P., Fransson, I., Koillinen, nsCL/P.K., et al. (2014). Dominant mutations in GRHL3 cause Van der Woude Syndrome and disrupt oral periderm development. Am J Hum Genet 94, 23-32.
16. Birnbaum, S., Reutter, nsCL/P., Lauster, C., Scheer, M., Schmidt, G., Saffar, M., Martini, M., Hemprich, A., Henschke, nsCL/P., Kramer, F.J., et al. (2008). Mutation screening in the IRF6-gene in patients with apparently nonsyndromic orofacial clefts and a positive family history suggestive of autosomal-dominant inheritance. Am J Med Genet A 146A, 787-790.
17. Leslie, E.J., Koboldt, D.C., Kang, C.J., Ma, L., Hecht, J.T., Wehby, G.L., Christensen, K., Czeizel, A.E., Deleyiannis, F.nsCL/P., Fulton, R.S., et al. (2015). IRF6 mutation screening in non-syndromic orofacial clefting: analysis of 1521 families. Clin Genet.
18. Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fen- nell, T., O'Donnell-Luria, A., Ware, J., Hill, A., Cummings, B., et al. (2015). Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv. http://dx.doi.org/10.1101/030338.
19. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature 526, 68–74.
20. Francioli, L.C., Menelaou, A., Pulit, S.L., Van Dijk, F., Palamara, P.F., Elbers, C.C., Neerincx, P.B.T., Ye, K., Guryev, V., Kloosterman, W.P., et al.; Genome of the Netherlands Consortium (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. 46, 818–825.
21. Mathieson, I., and McVean, G. (2012). Differential confounding of rare and common variants in spatially structured populations. Nat. Genet. 44, 243–246.
22. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191.
23. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529.
24. Cobat, A., Abel, L., Alcaïs, A., and Schurr, E. (2014). A general efficient and flexible approach for genome-wide association analyses of imputed genotypes in family-based designs. Genet. Epidemiol. 38, 560–571.
25. Jackson, M.R., Genin, E., Knapp, M., and Escary, J.L. (2002). Accurate power approximations for chi2-tests in case-control association studies of complex disease genes. Ann. Hum. Genet. 66, 307–321.
26. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252–W258.

27. Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 41, W349–W357.
28. Capriotti, E., Fariselli, P., and Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 33, W306–W310.
29. Jones, D.T., and Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31, 857–863.
30. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337. 31. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 22,1790–1797.
32. Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. J. Comput. Biol. 4,311–323.
33. Rintala, A.E., and Ranta, R. (1981). Lower lip sinuses: I. Epidemiology, microforms and transverse sulci. Br. J. Plast. Surg. 34,26–30.
34. Grosen, D., Chevrier, C., Skytthe, A., Bille, C., Mølsted, K., Sivertsen, A., Murray, J.C., and Christensen, K. (2010). A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. J. Med. Genet. 47, 162–168.
36. den Hollander, A.I., Koenekoop, R.K., Yzer, S., Lopez, I., Arends, M.L., Voesenek, K.E., Zonneveld, M.N., Strom, T.M., Meitinger, T., Brunner, H.G., et al. (2006). Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. Am. J. Hum. Genet. 79, 556–561.
37. Mangold, E., Reutter, H., Birnbaum, S., Walier, M., Mattheisen, M., Henschke, H., Lauster, C., Schmidt, G., Schiefke, F., Reich, R.H., et al. (2009). Genome-wide linkage scan of nonsyndromic orofacial clefting in 91 families of central European origin. Am. J. Med. Genet. A. 149A, 2680–2694.
38. Aldhorae, K.A., Bohmer, A.C., Ludwig, K.U., Esmail, A.H., Al- Hebshi, N.N., Lippke, B., Golz, L., Nothen, M.M., Daratsianos, N., Knapp, M., et al. (2014). Nonsyndromic cleft lip with or without cleft palate in arab populations: genetic analysis of 15 risk loci in a novel case-control sample recruited in Yemen. Birth Defects Res. A Clin. Mol. Teratol. 100, 307–313.
39. Nikopensius, T., Jagomagi, T., Krjutskov, K., Tammekivi, V., Saag, M., Prane, I., Piekuse, L., Akota, I., Barkane, B., Krumina, A., et al. (2010). Genetic variants in COL2A1, COL11A2, and IRF6 contribute risk to nonsyndromic cleft palate. Birth Defects Res. A Clin. Mol. Teratol. 88, 748–756.
40. Apostolidou, S., Abu-Amero, S., O'Donoghue, K., Frost, J., Olafsdottir, O., Chavele, K.M., Whittaker, J.C., Loughna, P., Stanier, P., and Moore, G.E. (2007). Elevated placental expression of the imprinted PHLDA2 gene is associated with low birth weight. J. Mol. Med. 85, 379–387.

FIGURE LEGENDS

Figure 1: Analysis of the effect of rs41268753 on the disorganized protein state for GRHL3

Using DISOPRED3[24], the intrinsic disorder profile for GRHL3 was generated for the wildtype (A) and the variant version p.Thr454Met (B). Protein-binding sites are also indicated within disordered regions. The black line indicates the variant position around which a set of 19 consecutive amino acids (positions 443 to 461) are newly annotated as "disordered" or "disordered protein binding" suggesting an increase of binding capacity for that region.

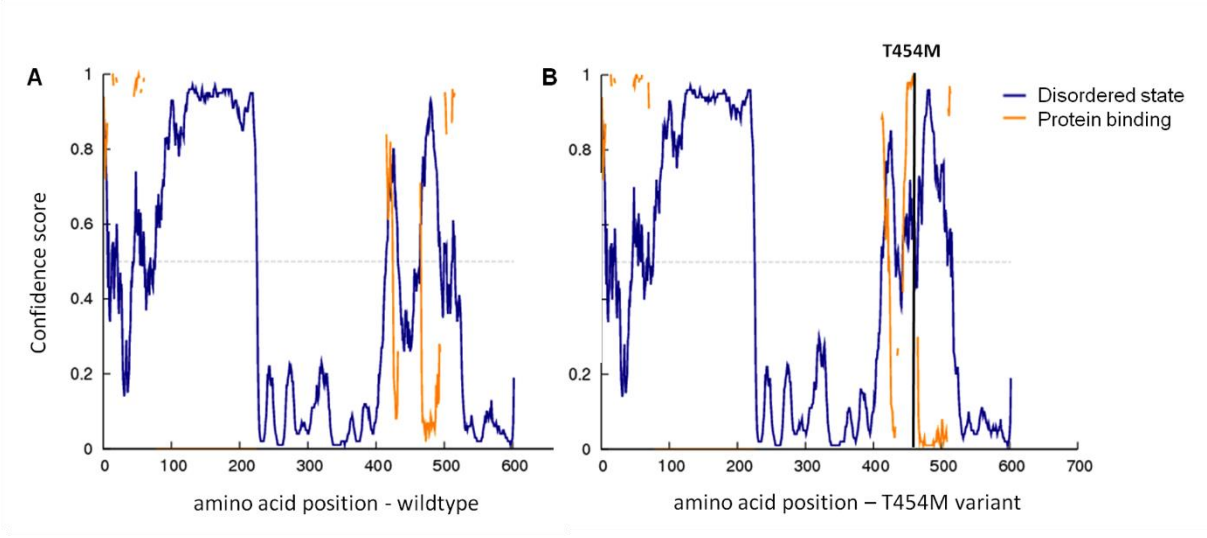
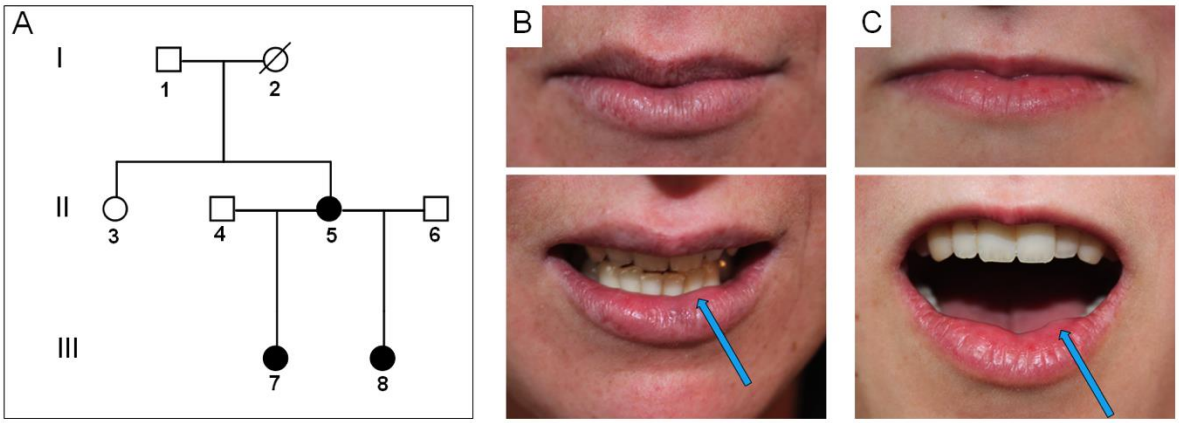


Figure 2. A *GRHL3* mutation in a family with three individuals affected by CPO, two of them with asymmetrical lower lips, which may represent a subtle VWS symptom

(A) In BN-813-III-7 we identified c.1285+2delT, predicted to affect a splice donor site. This variant was also identified in the half-sister (BN-813-III-8) and in the mother (BN-813-II-5), but was not present in the half-sister's father BN-813-II-6). No DNA from BN-813-I-1 or BN-813-I-2 was available for analysis. Both half-sisters had a complete cleft of the hard and soft palate. Of note, the mother was initially reported as unaffected, but on at a clinical re-examination was found to have a broad uvula and had nasal speech that was strongly suggestive of a submucous cleft palate. None of the three affected individuals showed the VWS typical lip pits. (B and C) Both half-sisters, BN-813-III-7 (B) and BN-813-III-8 (C), showed slightly asymmetrical lower lips with elevations on the left side when opening their mouths. Given their molecular diagnosis, these elevations might a posteriori be interpreted as a very subtle form of either lower lip pits or transverse sulci of the lips, symptoms occasionally described in VWS.



Tables

Table 1: Single marker association results of rs41268753 in different replication cohorts and combined analysis.

sample	cases		controls		test statistics	P_{trend}	allelic RR (95%CI)	genotyping method	sample description
	genotype distribution	MAF	genotype distribution	MAF					
Central Europeans	0/19/77	0.099	1/24/242	0.049	6.26	1.24×10^{-2}	2.15 (1.16-3.97)	KASP assay on demand	cases: Mangold et al. 2009[37] controls: Birnbaum et al. 2009[4]
Yemen	0/6/41	0.064	0/10/177	0.027	3.25	7.16×10^{-2}	2.48 (0.88-7.01)	KASP assay on demand	Aldhorae et al. 2014[38]
Latvia	0/10/41	0.098	0/3/91	0.016	10.92	9.53×10^{-4}	6.70 (1.80-24.95)	TaqMan Genotyping	Nikopensius el al. 2009[39]
White_London	1/11/82	0.069	1/10/166	0.034	3.098	7.84×10^{-2}	2.12 (0.95-4.74)	Sanger Sequencing	cases: see footnote [a] controls: Apostolidou et al. 2007[40]
Combined	-	-	-	-	-	2.63×10^{-5}	2.46 (1.62-3.74)		
In combination with European trios (dbGaP)[b]	-	-	-	-	-	2.73×10^{-9}		GWAS	Beaty et al 2011[10]

Abbreviations are as follows: MAF, minor allele frequency; RR, relative risk; CI, confidence interval.

[a]The patient cohort consisted of 136 individuals with nonsyndromic CPO (including 94 European individuals), who were clinically evaluated and treated at the North Thames Regional Cleft Centre based at Great Ormond Street Children's Hospital, London.

[b]The datasets used for the analyses described in this manuscript were obtained from dbGaP: phs000094.v1.p1. Genotypes were imputed using IMPUTE2 and analyzed using the FBATdosage method in which the mean of transmitted allele dosages is compared to the mean of non-transmitted allele dosages. The mean of transmitted allele dosages estimates the allele frequency in affected children (for rs41268753; 0.093) whereas the mean of non-transmitted allele dosages estimates the frequency in non-transmitted parental alleles (rs41267853: 0.036). For rs41268753, this reveals p 1⁄4 1.18 3 10 ␣4 in the European nsCPO trios.

Table 2: Overview of *GRHL3* truncating mutations in individuals with nsCPO

Genomic Position (hg19)	cDNA change	Location	Protein amino acid change	Functional effect	Status
24.664.177	c.738C>T	Exon 6	p.Gly246fs*10	introduces alternative splice donor site[a]	de novo
24.664.280	c.840+1G>T	Intron 6	p.?	splice site mutation	de novo
24.664.558	c.916dupC	Exon 7	p.Arg306Profs*11	frameshift	complete co-segregation
24.669.264	c.1285+2delT	Intron 10	p.?	splice site mutation	complete co-segregation

None of the truncating variants was observed in any of the following databases: European population of 1000 genomes phase 3, European variant server (European Americans), Non-Finnish Europeans of the Exome Aggregation consortium, Genome of the Netherlands. Positions given for NM_198174.2 / NP_937817.3. Pedigree information can be found in Figure S3.

[a]Effect at RNA level confirmed by cDNA sequencing.