High incidence and regional distribution of cleft palate in Finns are associated with a functional variant in an IRF6 enhancer

Fedik Rahimov  
AbbVie  https://orcid.org/0000-0001-6519-7404

Pekka Nieminen  
University of Helsinki

Priyanka Kumari  
University of Iowa  https://orcid.org/0000-0002-7210-172X

Emma Juuri  
University of Helsinki and Helsinki University Hospital

Tiit Nikopensius  
University of Tartu

Juha Karjalainen  
Broad Institute of Harvard and MIT

Mitja Kurki  
The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

Aarno Palotie  
University of Helsinki  https://orcid.org/0000-0002-2527-5874

FinnGen NA  
NA

Arja Heliövaara  
University of Helsinki and Helsinki University Hospital  https://orcid.org/0000-0001-8064-3484

Kitt Paraiso  
Lawrence Berkeley National Laboratory  https://orcid.org/0000-0002-6489-810X

Axel Visel  
Lawrence Berkeley National Laboratory  https://orcid.org/0000-0002-4130-7784

Walid Fakhouri  
University of Texas Health Science Center at Houston

Brian Schutte  
Michigan State University

Robert Cornell  
University of Iowa

David Rice ( david.rice@helsinki.fi )
Genetics Article

Keywords: cleft palate, cleft lip, CRISPR-Cas9, genome study

Posted Date: October 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-941741/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
High incidence and regional distribution of cleft palate in Finns are associated with a functional variant in an IRF6 enhancer

Fedik Rahimov\textsuperscript{1,16}, Pekka Nieminen\textsuperscript{2,16}, Priyanka Kumar\textsuperscript{3,16}, Emma Juuri\textsuperscript{4}, Tiit Nikopensius\textsuperscript{5}, Juha Karjalainen\textsuperscript{6,7,8}, Mitja Kurk\textsuperscript{6,7,8}, Aarno Palotie\textsuperscript{6,7,8}, FinnGen, Arja Heliövaara\textsuperscript{9}, Kitt Paraiso\textsuperscript{10}, Axel Visel\textsuperscript{10,11,12}, Walid D. Fakhouri\textsuperscript{13,14}, Brian C. Schutte\textsuperscript{15,*}, Robert A. Cornell\textsuperscript{3,*}, David Rice\textsuperscript{4,*}

\textsuperscript{1}Department of Human Genetics, Genomics Research Center, AbbVie, Inc., North Chicago, IL 60064, USA; \textsuperscript{2}Department of Orthodontics, University of Helsinki, Helsinki 00014, Finland; \textsuperscript{3}Department of Anatomy and Cell Biology, University of Iowa, Iowa City, IA 52242, USA; \textsuperscript{4}Department of Orthodontics, University of Helsinki and Helsinki University Hospital, Helsinki 00014, Finland; \textsuperscript{5}Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu 51010, Estonia; \textsuperscript{6}Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; \textsuperscript{7}Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; \textsuperscript{8}Institute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of Helsinki, Helsinki 00014, Finland; \textsuperscript{9}Cleft Palate and Craniofacial Center, Department of Plastic Surgery, University of Helsinki and Helsinki University Hospital, 00029 HUS, Finland; \textsuperscript{10}Environmental Genomics and Systems Biology Division, Lawrence Berkeley Laboratories, Berkeley, CA 94720, USA; \textsuperscript{11}U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley Laboratories, Berkeley, CA 94720, USA; \textsuperscript{12}University of California, Merced, CA 95343, USA; \textsuperscript{13}Department of Diagnostic and Biomedical Sciences, School of Dentistry, University of Texas Health Science Center at Houston, Houston, TX 77054, USA; \textsuperscript{14}Department of Pediatrics, McGovern Medical School, University of Texas Health Science Center at Houston, TX, 77030; \textsuperscript{15}Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA.

\textsuperscript{16}These authors contributed equally to this work

*Correspondence: schutteb@msu.edu (B.C.S), robert-cornell@uiowa.edu (R.A.C.), david.rice@helsinki.fi (D.R)
Summary

In Finland the frequency of isolated cleft palate (CP) is higher than that of isolated cleft lip (CL) or cleft lip with cleft palate (CLP). This trend contrasts to that in other countries but its genetic underpinnings are unknown. We performed a genome-wide association study for orofacial cleft, which includes CL, CLP, and CP in the Finnish population. We identified rs570516915, a single nucleotide polymorphism that is virtually specific to Finns and Estonians, as being strongly associated with orofacial cleft, and predominantly with CP \( (p = 5.25 \times 10^{-34}, \text{OR} = 8.65, 95\% \text{ CI} 6.11-12.25) \). The risk allele frequency for rs570516915 parallels the regional variation of CP prevalence in Finland, and the association was replicated in an independent sample of CP cases from Estonia \( (p < 1.3 \times 10^{-11}) \). The rs570516915 variant lies in an \textit{IRF6} (Interferon Regulatory Factor 6) enhancer active in ectodermal and oral epithelial cells. Other variants in the \textit{IRF6} locus, including those in the same enhancer, are associated with orofacial cleft, but predominantly with CL or CLP, suggesting shared mechanisms in the distinct processes of lip and palate development. By luciferase and transgenic mouse reporter assays we found that the risk allele of rs570516915 diminishes the activity of the enhancer. CRISPR-Cas9 edited oral epithelial cells demonstrate that rs570516915 disrupts an IRF6 binding site and decreases the expression of \textit{IRF6}, suggesting impaired \textit{IRF6} autoregulation as the molecular mechanism underlying risk for CP.
Introduction

Orofacial clefts (OFCs) are common congenital malformations that affect approximately 1 in 700 newborns globally.\(^1\) OFCs may be classified as cleft lip only (CL), cleft lip with cleft palate (CLP) or cleft palate only (CP). Epidemiologically, CL and CLP are usually combined because it is unclear whether the cleft palate in individual cases of CLP is caused by the cleft of the lip which occurs prior to palatogenesis. The incidence of OFCs tends to vary between countries, racial and ethnic groups, and based on the socioeconomic status of affected families. In general, the incidences of CL and CLP are higher in the European populations than the incidence of CP. However, Finland has one of the highest rates of nonsyndromic (or isolated) CP in the world and the ratio of incidence of CL and CLP versus CP is reversed, with CP being more common.\(^2\)

Notably, even among the five Nordic countries where the total incidence of clefts is very similar, Finland stands out for having higher CP incidence than CLP.\(^2\) In Finland the prevalence of CL and CLP between 1993 and 2011 was recorded at 0.95 per 1,000 live births, whereas the prevalence of CP was 1.39 per 1,000 live births.\(^3\) Furthermore, there is variation in the incidence rates of different types of clefts among regions within the country. CP is more common in northern and eastern Finland, whereas CLP is more common in southern and western parts of the country. For example, the incidence of CP in the Oulu University Hospital between 1993 and 2011 was recorded at 1.78 per 1,000 live births and selective termination of pregnancies. In contrast, the incidence of CLP was 1 in 1,000.\(^3\) Moreover, one-fifth of OFC cases from the Northern Finland also have a family history, among whom the highest incidence (76.7%) was documented in cases with CP.\(^4\) Thus, as previously hypothesized,\(^2\) the epidemiological observations suggest the presence of genetic risk factor(s) that contribute to this exceptionally high incidence and regional variation of CP rate in the Finnish population.

OFCs exhibit a multifactorial etiology that results from a complex interaction of various environmental exposures and genetic predisposition.\(^5\) Before genome-wide association studies (GWAS), linkage and candidate gene association studies yielded inconsistent evidence of association for many genes in different populations except for a handful of genes. GWASs identified over two dozen genes or loci achieving genome-
wide significance, many of which have been independently replicated. Most of the GWASs have been conducted in CL and CLP, and GWAS loci from these studies tested in CP did not always replicate. Few GWAS have been carried out for CP, resulting in only a few replicated and functionally characterized association signals. For example, a common missense variant in the Grainyhead-like transcription factor 3 (GRHL3) gene was associated with CP in multiple cohorts, but the GHRL3 gene locus has not been associated with CL or CLP, to date.

Epidemiological and embryological studies suggest that the development of lip and secondary palate are distinct processes, perhaps partially explaining the nonoverlap of GWAS signals between CLP and CP. However, there are a few syndromic forms of OFCs where the CL or CLP and CP may be observed in the same family. The classic example of this “mixed” orofacial cleft syndrome is Van der Woude syndrome (VWS [MIM: 119300]). VWS can be caused by rare pathogenic mutations in either IRF6 or GRHL3. Conversely with GRHL3, DNA variants in the IRF6 locus have been associated, with rare exception, only with nonsyndromic CL and CLP in both candidate gene studies and GWAS. Interestingly, GRHL3 has been shown to be a downstream effector of IRF6, suggesting that this pathway might be involved in pathogenesis of isolated CP.

Here we conducted a GWAS in the Finnish population and report a strong association between CP and the IRF6 locus, a low-frequency, non-coding single nucleotide polymorphism (SNP) located in an enhancer for the IRF6 gene. Furthermore, through cell-based and animal studies, we validate the regulatory effect of the variant on IRF6 expression.

Material and Methods

FinnGen cohort
The FinnGen study is a population-based cohort study launched in 2017 as a public private partnership aiming to identify genotype-phenotype relationships in the Finnish founder population to support drug discovery programs (see Web Resources for more information). Three national and six regional biobanks provide samples to the FinnGen
study. Individuals from several previously established prospective epidemiological and
disease cohort studies are included in FinnGen as well. Patients and control subjects in
FinnGen provided informed consent for biobank research, based on the Finnish
Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish
Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017),
were collected based on study-specific consents and later transferred to the Finnish
biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory
Authority for Welfare and Health. Recruitment protocols followed the biobank protocols
approved by Fimea. The Coordinating Ethics Committee of the Hospital District of
Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr
HUS/990/2017. The FinnGen study is approved by Finnish Institute for Health and
Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017,
THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019,
THL/1721/5.05.00/2019, THL/1524/5.05.00/2020, and THL/2364/14.02/2020), Digital
and population data service agency (permit numbers: VRK43431/2017-3,
VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers:
and Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020
(earlier TK-53-90-20). The Biobank Access Decisions for FinnGen samples and data
utilized in FinnGen Data Freeze 7 include: THL Biobank BB2017_55, BB2017_111,
BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8,
Biobank HUS/359/2017, Auria Biobank AB17-5154 and amendment #1 (August 17
2020), Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland
1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and
amendments (21.02.2020 & 06.10.2020), Central Finland Biobank 1-2017, and
Terveystalo Biobank STB 2018001. The present GWAS was conducted in 309,154
study participants included in Data Freeze 7 and had a Finnish genetic ancestry based
on a principal component analysis of the genotypes.
Estonian cohort

The Estonian Biobank is described broadly elsewhere.\textsuperscript{27,28} The Estonian cohort consisted of 73 CP cases and 152,284 population controls. Ethical approval for this study was obtained from the Research Ethics Review Committee of the University of Tartu. All participants signed informed consent.

Phenotype definitions

Case subjects with nonsyndromic forms of OFCs were defined from electronic health records using diagnosis codes of the World Health Organization’s International Classification of Diseases and Related Health Problems (ICD). Codes of versions ICD-10, ICD-9 and ICD-8 were available since 1996, between 1987-1995 and between 1967-1986, respectively. Syndromic cases were excluded. Cases with CP only \((n = 228; 157\) females and 71 males) were defined based on the ICD-10 code Q35, ICD-9 code 749.0, and ICD-8 code 749.00. Cases with CLP \((n = 121; 62\) females and 59 males) were defined using the ICD-10 code Q37, ICD9 code 7492[ABCDE], and ICD8 codes 74920 and 74924. Cases with CL \((n = 54; 28\) females and 26 males) were defined using ICD10 code Q36, ICD9 codes 7491[ABCDEFG] and ICD8 codes 74910 and 7491[123], but due to the small number of cases, CL cases were combined with CLP cases in GWAS. CP cases \((n = 73, 56\) females and 17 males) in the Estonian biobank were also ascertained using the ICD-10 code Q35 through a questionnaire implemented by the biobank. The rest of the study participants were used as controls.

Genotyping and imputation

Samples collected and genotyped before the launch of the FinnGen study were genotyped with a mix of Illumina (Illumina) and Affymetrix arrays (Thermo Fisher Scientific). However, recently collected samples were genotyped on two custom-designed Affymetrix arrays referred to as FinnGen1 and FinnGen2 (655,973 markers). Genotype calls were made with GenCall and zCall algorithms for Illumina, and AxiomGT1 algorithm for Affymetrix data. Individuals with gender mismatches, high genotype missingness (>5%), excess heterozygosity (±4SD) and non-Finnish ancestry were removed from the study. Variants with high missingness (>2%), deviation from
Hardy-Weinberg equilibrium (p < 1 x 10^{-6}) and minor allele count (MAC < 3) were excluded. Chip-genotyped samples were pre-phased chip-wise with Eagle (v2.3.5).\textsuperscript{29} Genotype imputation was performed with Beagle (v4.1)\textsuperscript{30} by using a Finnish population-specific reference panel built from 3,775 high-coverage (25-30x) whole-genome sequences (WGS) in Finns. A total of 17M variants were imputed. Variants with an imputation INFO score < 0.8 or MAF < 0.0001 were excluded, resulting in a combination of 15M directly genotyped and imputed variants included in association tests. In total, 309,154 individuals (173,746 females and 135,408 males) passed QC and were included in association analyses. Samples from the Estonian cohort were genotyped using the Infinium Global Screening Array (Illumina). Directly genotyped markers were imputed with the IMPUTE (v.2.2.2) program using an Estonian population specific reference panel of WGS data from 2,240 Estonian individuals described previously.\textsuperscript{31}

**Association analysis**

Because biobank-based cohort studies tend to contain biologically related individuals, an association of directly genotyped and imputed variants with OFC phenotypes were tested using the mixed-effects logistic regression method implemented in the REGENIE package\textsuperscript{32}. We used the code version 2.0.2 of the REGENIE program which was modified to include dosage-based calculation of allele frequencies in cases and controls. We corrected for within-sample relatedness using a genetic relationship matrix (GRM) that was constructed from a linkage disequilibrium (LD)-pruned set of 55,139 well-imputed, high quality (INFO > 0.95 and MAF > 0.01) SNPs. LD pruning was performed with PLINK (v2.00a2.3LM).\textsuperscript{33} In addition, sex as determined from genotypes, age, genotyping batch, and top 10 principal components (PCs) were used as covariates. We used the approximate Firth test for variants with an initial $p$ value of less than 0.01 and computed the standard error based on effect size and likelihood ratio test $p$ value. Replication SNPs from the Estonian Biobank were tested using a standard 1-df chi-square allelic test. Because we tested a limited number of markers for replication, we could not correct for population stratification in the Estonian cohort. Manhattan plots were generated using the *ggplot2* R package.\textsuperscript{34} Regional association plots were
generated with the stand-alone version of LocusZoom (v1.3),\textsuperscript{35} using the Finnish population-specific LD structure estimated in 3,775 whole-genome sequences of Finns.

**Conditional analysis**

Conditional analysis was performed using the COJO\textsuperscript{36} approach (\texttt{--cojo-cond} command) implemented in the GCTA software tool (v1.93.2)\textsuperscript{37}. For this analysis, we used summary statistics output from the REGENIE GWAS and imputed genotypes from the entire FinnGen cohort for LD estimation within 1 Mb to each direction from the conditioning SNP.

**Geographic variation**

Geographic distribution of allele frequencies was calculated and plotted using region-level, birthplace data from 320,471 FinnGen study participants (obtained from Statistics Finland) with the \texttt{ggplot2}, \texttt{ggrepel} and \texttt{sf}\textsuperscript{38} R packages. Spatial polygon data (v3.6) for the Finnish map were downloaded from the database of Global Administrative Areas (GADM, see Web Resources). Finnish enrichment is calculated as \(\text{AF}(\text{FIN}) / \text{AF}(\text{NFSEE})\) in gnomAD 2.1.1 (see Web Resources), where NFSEE stands for non-Finnish-non-Swedish-non-Estonian European. Correlation between the regional prevalence of CP and allele frequency was calculated and plotted using Pearson’s product-moment correlation in R (v3.6.1).

**Population attributable risk**

The population attributable risk (PAR) is the proportion of cases that would be prevented if a risk factor were eliminated from the population. PAR of individual causal variants was calculated as described previously,\textsuperscript{39}

\[
PAR = 1 - \frac{1}{(1 - p)^2 + 2p(1 - p)e^\beta + p^2e^{2\beta}}
\]
where \( p \) is the risk allele frequency in controls in the FinnGen cohort and \( \beta \) is the beta coefficient of the variant from the REGENIE regression analysis, which indicates the effect size of the variant on phenotype.

**Bioinformatic analysis**

SNP overlap with open-chromatin peaks, chromatin-immunoprecipitation-sequencing (ChIP-Seq) peaks of histone modification (H3K27ac) from HOIEC oral epithelium cells and publicly available chromatin mark datasets from 9 principal cell types from ENCODE Project Consortium\(^{40}\) and from different stages of embryonic facial explants from Cotney's lab\(^{41}\) were visualized using the UCSC Genome Browser. H3K27Ac ChIP-Seq data from HOIEC oral epithelium cells and ATAC-Seq (Assay for Transposase-Accessible Chromatin using sequencing) data from the HOIEC oral epithelium cells and the HEPM oral mesenchyme cells were generated previously.\(^{42}\) **In silico** transcription factor binding activities at the risk variant site were determined using HOMER\(^{43}\) and the Transcription factor Affinity Prediction (TRAP) tool\(^{44}\) (see Web Resources). Position Weight Matrix of the consensus IRF6 binding sequence was obtained from the JASPAR database.\(^{45}\)

**Cell culture**

The immortalized human embryonic oral epithelial cell line (GMSM-K)\(^{46}\) (a kind gift from Dr. Daniel Grenier) was maintained in keratinocyte serum-free medium (Thermo Fisher Scientific) supplemented with human recombinant epidermal growth factor (EGF-30 \( \mu \)g/ml) and bovine pituitary extract (BPE-0.2 ng/ml) (Thermo Fisher Scientific).

**Electroporation and dual luciferase reporter assay**

For dual luciferase assays, each firefly reporter construct (Supplemental Material and Methods) was co-transfected with a constitutively driven Renilla luciferase plasmid. Briefly, GMSM-K cells were electroporated with plasmid using the Amaxa Cell Line Nucleofector Kit (Lonza) and the Nucleofector II instrument (Lonza) program T-020. We used a dual-luciferase reporter assay system (Promega) and FB12 Luminometer (Berthold Detection Systems) to evaluate the luciferase activity following manufacturer's
instructions. Relative luciferase activity was calculated as the ratio between the value for the Firefly and Renilla enzymes. Three independent measurements were performed for each transfection group, and three biological replicates were performed. All results are presented as mean ± standard deviation. Statistical significance was determined using the Student’s *t*-test.

**CRISPR-Cas9 mediated genome editing**

Engineering of GMSM-K cells was achieved using CRISPR-Cas9 mediated homology-directed repair (HDR) using the recommended protocol (Integrated DNA Technologies [IDT]). IDT’s CRISPR-Cas9 guide RNA (gRNA) design checker tool (see Web Resources) was used to design specific CRISPR RNA (crRNA) harboring the rs570516915 in the seed region to avoid the re-cleavage by CRISPR-Cas9 (Figure S1). HDR template including the desired mutation was designed to have a minimum of 30-35nt long homology arms (Figure S1). Briefly, equimolar concentrations of crRNA and trans-activating crRNA (tracrRNA, IDT: 1072532) were annealed at 95°C for 5 min followed by cooling at room temperature (RT) to form functional gRNA duplexes. Further, the ribonucleoprotein (RNP) complex was prepared 10 minutes before transfection by mixing gRNAs with Cas9 protein (IDT: 1081058) at RT. RNP complexes together with the repair template were transfected into approximately 1 million GMSM-K cells with Amaxa Cell Line Nucleofector Kit V (Lonza) using Nucleofector II (Lonza) (program: T-020). After 48-72 hours of growth, cells were re-plated in a 150 mm dish to get single cell colonies. Single cell colonies were screened by PCR and sequenced using primers flanking rs570516915 (Primer sequences are provided in Table S1). Clones of cells from the same transfection but that were not edited at rs570516915 were used as a negative control (non-risk) such that off-target changes might be shared.

**Quantitative RT-PCR on CRISPR edited cells**

RNA was extracted from five engineered clones and four non-edited clones using Quick-DNA/RNA Miniprep kit (Zymo Research) followed by DNase treatment using manufacturer’s instructions (Thermo Fisher Scientific). Reverse transcription was
performed using High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). Real-time PCR was performed using SYBR Green qPCR mix (Bio-Rad) in Bio-Rad CFX Connect Real-Time System. Expression level of the ACTB (beta-actin) gene was used as the internal control. Quantitative PCR reaction for each sample was performed in triplicate. Fold change was calculated using the $2^{-\Delta\Delta C_T}$ analysis method. Primer sequences used in the study for qRT-PCR are provided in Table S1.

**Chromatin immunoprecipitation qPCR**

GMSM-K edited (harboring risk allele) and non-edited (harboring non-risk allele) cells were seeded at 5 x $10^6$ cells per 150 mm plate (one plate per biological replicate), grown to 90-100% confluency. Cells were fixed in 1% formaldehyde for 10 min at RT. Fixation was stopped with 125 mM glycine for 5 min at RT. Cells were washed twice and collected with a cell scraper in ice-cold PBS followed by centrifugation at 1000 rpm for 10 min. Cells were sonicated for 5 min with 30 sec ‘on’ and 40 sec “off” (VirSonic 600) to obtain an average size between 300 and 600 bp of genomic DNA. The sonicated cell lysates were subjected to chromatin immunoprecipitation with 4 μg of specific antibody (Rabbit anti-IRF6 from Akira Kinoshita, Nagasaki University as previously described and anti-H3K27Ac from Millipore-Sigma). DNA was isolated from immune complexes using the phenol:chloroform:isoamyl alcohol extraction method. Quantitative PCR was performed in triplicate and the fold enrichment was normalized over IgG (primer sequences are shown in Table S2) from two replicates of anti-H3K27Ac ChIP and one of anti-IRF6 ChIP. A negative control set of primers was designed to a sequence that lacked IRF6 transcription factor binding and did not harbor active elements identified from ATAC-Seq and H3K27Ac in HOIEC and NHEK cells.

**Transgenic mouse reporter assays**

MCS-9.7-lacZ transgenic reporter constructs were generated with the risk and non-risk alleles of rs570516915 on the Finnish haplotype background with a mouse Shh promoter (Supplemental Material and Methods). These constructs were co-injected with Cas9 protein and gRNA targeting the H11 locus into the mouse embryos, and embryos were harvested at E13.5 and processed for X-gal staining as described.
Plasmid copy number integration for transgenic embryos was estimated by RT-qPCR with probes targeting the mouse Shh promoter region (Thermo Fisher Cat#4400291 Mm00560391_cn). This probe targets both endogenous and plasmid copies. The Tfrc gene (Thermo Fisher Cat#4458366) was used as a reference locus. Three non-transgenic mouse DNA samples were used as normalization controls. To calculate an estimated plasmid integration copy number, the $2^{-\Delta\Delta C_p}$ approach was used. Mouse reporter experiments for LacZ reporter transgenic animal work were performed at the Lawrence Berkeley National Laboratory (LBNL) and were reviewed and approved by the LBNL Animal Welfare and Research Committee.

Results

A novel SNP in the IRF6 enhancer MCS-9.7 is associated with CP

We conducted a GWAS of nonsyndromic OFC in the Finnish population using the FinnGen study cohort with a mixture of 355 CL, CLP, and CP cases and 308,799 ethnicity-matched population controls. We identified genome-wide significant associations with two regions located on the long and short arms of chromosome 1 (Figure S2A). Upon dividing the case group into subtypes of clefts, we discovered that the association signals at both loci were entirely driven by the subgroup of 228 cases with CP (Figure 1). GWAS of a combined group of 151 cases with CL and CLP did not yield any significant associations (Figure S2B). Due to small sample sizes, we did not conduct independent analyses in the CL and CLP subsets. Variants reaching p values less than $1.0 \times 10^{-5}$ in all three association analyses are listed in Table S3, S4 and S5.

In the GWAS of CP, ten SNPs reached genome-wide significance on chromosome 1p36.1, including the lead SNP rs113965554 ($p = 2.32 \times 10^{-11}$, OR = 2.73) and the previously described etiologic missense variant p.Thr454Met (rs41268753, $p = 1.39 \times 10^{-10}$, OR = 2.66) in the GRHL3 gene (Table 1). The top five variants with similar p values are in strong LD with one another (Figure 2A). These results indicate that the 1p36.1 signal is likely driven by the GRHL3 missense variant, further replicating the association between this common missense SNP and CP in the Finnish population.
Most importantly, we observed highly significant association signals on chromosome 1q32.2, where a total of 251 SNPs reached genome-wide significance. These consisted of two independent association signals in the region covering the *IRF6* gene (Figure 2B). One signal comprised a 62 kb-long haplotype of SNPs in strong LD with each other upstream of *IRF6*, with the lead SNP rs7546104 reaching a p value of 9.53 x 10^{-14} (Table 1, Figures S3A and S3B). One of the SNPs within this haplotype, rs72741048 (p = 1.15 x 10^{-11}, Figures S3B), has previously been linked to OFC, including CP, in a large GWAS in the Chinese population. The second association signal spanned a genomic region of over 7 Mb (Figure 2B). The lead SNP rs570516915 (p = 5.25 x 10^{-34}, OR = 8.65, 95% CI 6.11-12.25) is a novel association with CP and changes a highly-conserved nucleotide (Figure S4) in a previously-described cis-regulatory element 9.7 kb upstream of *IRF6* transcription start site (i.e., multi-species conserved sequence, 9.7 kb, “*IRF6* MCS-9.7”). After conditioning on rs570516915, the rs7546104 SNP showed residual evidence of association exceeding genome-wide significance (r^2 = 0.035, OR\textsubscript{cond} = 1.75, p\textsubscript{cond} = 3.98 x 10^{-10}), supporting other evidence that there are two independent association signals at this locus (Figures S3C and S3D).

Given the location of the rs570516915 variant in a well-characterized enhancer, in which two other variants have previously been associated with OFCs, we focused on this variant as the likely causal variant for further inquiry. First, we sought to replicate the association of this variant in an independent cohort of CP cases. This was a challenge because the minor allele frequency (MAF) of rs570516915 (1.2%) is 19-fold higher in the Finnish population than in other Europeans and extremely rare in all examined populations outside Finland or Estonia (Table 1). However, the Estonian population is geographically and linguistically close to the Finnish population, and previous genetic associations in the Finnish population were replicated with the Estonian Biobank. The MAF for rs570516915 in the Estonian population was 0.3%, four times lower than in Finland. Nonetheless, in 73 CP cases and 152,284 population controls from the Estonian biobank the rs570516915 variant showed a significant pairwise association with CP (p < 1.3 x 10^{-11}). In summary, the association of rs570516915 with CP was observed in two independent cohorts.
The risk allele (G) of rs570516915 is tightly linked to the non-risk allele (also G) of the previously described common risk variant rs642961 in the same enhancer\textsuperscript{50}, suggesting that the association signals are independent of one another. Supporting this conclusion, as mentioned above, the rs570516915 variant shows specific association with CP alone while rs642961 is strongly associated with CL and modestly associated with CLP.\textsuperscript{50} The rs642961 variant, which has a minor allele frequency of 22\% in the Finnish population, did not reach statistical significance in the Finnish cohort of CLP cases evaluated here, possibly because of the small sample size.

**Correlation of regional rs570516915 allele frequency with regional CP prevalence**

Global epidemiological studies reveal a much higher incidence of CP in Finland in comparison to other populations; for instance, the incidence is 1.8-fold that in neighboring Sweden.\textsuperscript{2} Moreover, epidemiological studies from Finland have described a regional distribution of the incidence of different types of OFCs.\textsuperscript{54} Although Finns are a relatively homogenous population, genetic variation in Finland broadly reflects geographic birthplace.\textsuperscript{55} To test whether the distribution of the risk allele tracked with CP across the 19 regions of Finland, we used region-level birthplace registry data of 320,471 FinnGen study participants (~5.8\% of Finland’s total population). We observed that the risk allele shows higher allele frequency in the central and northern regions of the country that reaches up to 2.7\% (Figure 3A and Table S6), where the incidence of CP is also known to be higher compared to the southern regions of the country.\textsuperscript{4} The frequency of the risk allele in southern regions (i.e., 0.9\%) is lower than the average allele frequency of in entire Finland (i.e., 1.2\%). Interestingly, there is also a statistically significant correlation (Pearson’s $r = 0.5$, $p < 0.02$) between the regional distribution of the risk allele and CP prevalence in our cohort of CP cases (Figure 3B). On the other hand, the frequency of the risk allele for the other two associated SNPs, rs41268753 (p.Thr454Met in \textit{GRHL3}) and rs7546104 (\textit{IRF6} locus), did not vary similarly across the 19 regions in Finland. Therefore, of these three associated SNPs, only rs570516915 appears to contribute to the regional distribution pattern of CP in Finland.

We also compared the overall impact of these three variants on CP in Finland by estimating the population attributable risk (PAR) for each. Even though rs570516915
has a much lower MAF than the other two variants in Finland (Table 1), it has a comparable PAR (16.2% versus 15.0% and 34.1%, respectively). The reason for the strong PAR by rs570516915 is due to its high effect size, as evidenced by its large odds ratio (8.65). While the PAR for rs7546104 appears much larger than the other two SNPs, it is important to keep in mind that the frequency of the causal allele tagged by this SNP is unknown. In comparison with the rest of the world, we observed that the frequency of the associated alleles for these three SNPs is higher in Finland (Table 1). Assuming the effect size for each SNP is the same in different countries, then the higher allele frequencies could help explain the higher incidence of CP in Finland compared to other countries.

**rs570516915 disrupts regulatory activity of the IRF6 enhancer MCS-9.7**

An additional six SNPs were in strong LD ($r^2 > 0.8$) with the lead SNP rs570516915. To rule out the other *cis*-regulatory regions that may contain the causal variant in LD with rs570516915, we examined relevant chromatin marks at each of the top seven SNPs including: i) ATAC-seq peaks, which reveal open chromatin, from the HOIEC oral epithelium cells and the HEPM oral mesenchyme cells,\(^42\) ii) H3K27Ac peaks, revealing active enhancers and promoters, from the HOIEC cells,\(^42\) and iii) aggregate chromatin mark data from 9 principal cell types from the ENCODE project\(^40\) and from human embryonic facial enhancers datasets\(^41\). Of these seven SNPs, two SNPs, rs570516915 and rs556188853, fall into regions with chromatin marks consistent with enhancer activity in epithelial cells (HOIEC and NHEK) (Figure 4A and Figure S5C). None of the other SNPs lie in both open chromatin and apparent enhancers of any of the 9 ENCODE cell types (Figure S5). We next engineered 701-bp DNA fragments centered on the ATAC-seq peak in HOIEC cells containing rs556188853 or rs570516915 and harboring either the risk-associated or non-risk associated allele of the SNP into a luciferase-based reporter vector (Figure S6). We separately transfected these constructs into the GMSM-K basal oral epithelial cells.\(^46\) The vectors harboring rs570516915 drove luciferase levels above background, replicating previous results.\(^50\) However, the vector harboring *IRF6* MCS-9.7 with the derived risk allele of rs570516915 showed significantly reduced activity from the ancestral, non-risk allele (p
By contrast, the vectors harboring the region surrounding rs556188853 did not drive luciferase levels above background, and there was no significant difference in luciferase levels between the two alleles (Figure 4B). These results provide evidence that the risk allele at rs570516915 is functional by reducing the activity of the \textit{IRF6} MCS-9.7 enhancer.

To determine the effect of rs570516915 in its native chromatin context on the expression of \textit{IRF6}, we also generated isogenic oral epithelial cells that were homozygous for the risk allele G. Sequencing from GMSM-K cells showed this cell line is homozygous for the non-risk allele, T (i.e., TT). We transfected GMSM-K cells with appropriate CRISPR-Cas9 reagents to render them homozygous for the risk allele (i.e., GG) and isolated four and five independent clones of TT and GG genotype, respectively (Figure 4C). Quantitative RT-PCR revealed that the amount of \textit{IRF6} mRNA was significantly lower in cells homozygous for the risk allele compared with the cells homozygous for the non-risk allele (p < 0.0001) (Figure 4D). These findings provide evidence that the risk allele at rs570516915 directly affects risk for CP by reducing steady state levels of \textit{IRF6} mRNA.

\textbf{rs570516915 disrupts the autoregulatory activity of the \textit{IRF6} gene}

Motif analysis of the conserved sequences surrounding rs570516915 revealed a sequence matching the consensus binding site for Interferon Regulatory Factor 6, the transcription factor encoded by the \textit{IRF6} gene (Figure 5A). The risk-associated allele at rs570516915 is predicted to disrupt this binding site (Figure S7), suggesting that IRF6 autoregulates its expression through a positive feedback loop. To test this possibility, we performed chromatin immunoprecipitation and quantitative PCR (ChIP-qPCR) with anti-IRF6 antibody in the CRISPR-Cas9 engineered GMSM-K cells. We observed binding by IRF6 at \textit{IRF6} MCS-9.7 in GMSM-K cells compared with a negative control locus 103.7 kb from \textit{IRF6} that lacked an IRF6 binding site (i.e., \textit{IRF6}-103.7, Figure 5B). Binding by IRF6 was significantly decreased at \textit{IRF6} MCS-9.7 in the mutated GMSM-K cells that were homozygous for the risk allele (Figure 5B). To determine the impact on chromatin state by rs570516915 we performed parallel ChIP-qPCR experiments with anti-H3K27Ac. The anti-H3K27Ac ChIP-qPCR showed almost 3-fold reduction in
H3K27Ac binding in presence of risk allele as compared to the non-risk allele (Figure 5C).

To determine the \textit{in vivo} effect of rs570516915, we performed a transgenic F0 embryo assay with the \textit{IRF6} MCS-9.7 enhancer fused to the \textit{lacZ} reporter gene (Figure S8A).\textsuperscript{49} At E13.5, three of three embryos carrying the \textit{IRF6} MCS-9.7-\textit{lacZ} transgene with the non-risk allele showed strong staining in the ectoderm of face, torso, limbs, eye and pinna (Figure 5D, Figure S8B), consistent with previous staining patterns.\textsuperscript{50,56} The five embryos that carried the \textit{IRF6} MCS-9.7-\textit{lacZ} transgene with the risk allele of rs570516915 displayed two different staining patterns. Two embryos showed a pattern that appeared qualitatively similar, but quantitatively reduced, compared with the embryos that carried the reference sequence (Figure 5E, Figure S8C and Figure S8D), and three embryos showed little to no staining (Figure 5F, Figure S8C and Figure S8D). These data are consistent with the \textit{in vitro} data, showing that the risk allele for rs570516915 reduces \textit{IRF6} MCS-9.7 enhancer activity \textit{in vivo}.

The \textit{IRF6} MCS-9.7 enhancer recapitulates much of the expression pattern of \textit{IRF6}.\textsuperscript{56} Crossing of a mouse strain stably transgenic for the \textit{IRF6} MCS-9.7-\textit{lacZ} transgene\textsuperscript{56} with an \textit{Irf6} null mutant\textsuperscript{57} previously showed intense staining for the β-galactosidase activity in the skin, ear pinna, vibrissae follicles, and submandibular salivary glands of wild-type embryos.\textsuperscript{56,58} By contrast, in homozygous \textit{Irf6} mutant embryos, blue staining was absent in the skin, ear pinna, and vibrissae follicles, but strong blue staining was present in the submandibular salivary gland.\textsuperscript{58} These data show that IRF6 is required for the activity of the \textit{IRF6} MCS-9.7 enhancer during the development of ectoderm and its derived tissues (Figure 5G). Altogether, our data are consistent with a mechanism whereby rs570516915 disrupts binding of IRF6 at a critical site in \textit{IRF6} MCS-9.7 to destabilize a positive feedback loop for \textit{IRF6} expression that is required for palate development.

\textbf{Discussion}

Since the multiple waves of settlement of Finland after the ice age, several genetic founder and bottleneck events have occurred, and despite the considerable expansion
of the population since the early 1800s, regions of genetic isolation have persisted.\textsuperscript{59,60} Consequently, due to strong genetic drift individuals of Finnish ancestry have accumulated numerous deleterious alleles of clinical significance at a relatively high frequency compared to non-Finnish Europeans.\textsuperscript{61} This is well-exemplified by a group of 36 rare monogenic diseases, also known as the Finnish disease heritage, that exists at a higher frequency in Finland than anywhere else.\textsuperscript{62} There is also a strong regional concentration of complex diseases. For instance, schizophrenia, cognitive impairment, and familial hypercholesterolemia are more prevalent in the northeastern region of the country than in other regions and show association with regionally enriched genetic variants.\textsuperscript{63,64} The rs570516915 variant is another clear example of a functional variant with increased allele frequency in Finns (likely due to genetic drift) contributing to risk for a common, complex disease. It is, to our knowledge, the first non-coding functional variant that shows a strong association with a structural birth defect in the Finnish population. It was postulated before the genomics era that the relatively high incidence of CP in Finland compared to in other Scandinavian countries, which all have similar living standards, climate, infectious diseases, diet, cultural habits, and total incidence of clefts as Finland, is likely to be the distinct components of genetic heritage of Finns.\textsuperscript{2} Indeed, association of three independent loci that are driven by variants with high allele frequencies in Finland compared to in other populations explains the high incidence rate of CP in Finland. This finding has implications for genetic counseling and public health in Finland because the treatment needs of CP patients are typically different from those of CL or CLP patients.

The \textit{IRF6} gene, that encodes the transcription factor Interferon Regulatory Factor 6, was among the first genetic risk factors that consistently showed a strong association with the nonsyndromic form of OFCs.\textsuperscript{23-25,50} This is not surprising because haploinsufficiency of \textit{IRF6} causes VWS, an autosomal dominant form of CLP with lip pits. Moreover 15\% of VWS cases lack lip pits and such cases mimic nonsyndromic forms of OFC.\textsuperscript{19} Subsequent studies attributed some of the association signals at this locus to a common SNP (rs642961) in \textit{IRF6} MCS-9.7, an evolutionarily conserved enhancer located 9.7kb upstream the transcription start site of \textit{IRF6}.\textsuperscript{50} The rs642961 variant showed strong association with CL, but no effect with CP. Moreover, the risk-
associated allele of rs64296 disrupts binding of the transcription factor AP2-α in an
electrophoretic mobility shift assay, suggesting it is functional.\textsuperscript{50} Furthermore, we
identified a rare, single nucleotide insertional mutation (350dupA) in \textit{IRF6} MCS-9.7 that
showed incomplete penetrance in a Brazilian family affected with VWS.\textsuperscript{52} In the present
study, we report another highly conserved, low-frequency SNP located only 153 bp
away from the previously reported common risk variant which shows remarkably
particular association for CP alone. This SNP appears to be functional as it reduces
\textit{IRF6} expression in oral epithelial cells apparently by blocking the binding of IRF6. In
summary, this study identifies the third OFC-associated and potentially functional
variant in the same enhancer for \textit{IRF6}.

Human lip and palate development starts in the 4\textsuperscript{th} and 5\textsuperscript{th} gestational week,
respectively, with the coordinated growth and morphogenesis of the facial processes.
These facial processes make contact, adhere, and eventually fuse in a synchronized
manner which if disrupted can result in an orofacial cleft.\textsuperscript{5,65} \textit{IRF6} could affect several
stages of development, but one area that has received particular interest is the
adhesion, rearrangement and ultimately removal of the epithelial cells of the lip and
palatal processes at the sites of fusion.\textsuperscript{66} The known roles of \textit{IRF6} in regulating
epithelial cell migration, E-Cadherin recycling, and the delivery of E-Cadherin to the
plasma membrane can potentially explain the cellular processes involved in lip and
palatal epithelial seam obliteration.\textsuperscript{67}

CL or CLP and CP have distinct pathophysiologic mechanisms due to different
embryonic origins, timing, and epidemiology. The association of two common genetic
variants (rs642961 and rs570516915) within the same regulatory element suggests that
the tissue where this enhancer is active is likely necessary for both events. In the
palate, \textit{IRF6} MCS-9.7 is active in both periderm and the basal epithelial layer.\textsuperscript{56} While
both of these tissues are involved in secondary palate fusion\textsuperscript{68,69} and fusion at the
lambdoidal junction,\textsuperscript{70} we favor periderm as the relevant tissue because \textit{IRF6} and
\textit{GRHL3} are required for periderm formation and/or function.\textsuperscript{21,68} Failure of periderm
formation or function leads to intraoral epithelial adhesions that hinder development of
the palatal shelves and can cause cleft palate.\textsuperscript{68} Further studies are necessary to
understand better how a single ~550bp long regulatory region controls the expression of
IRF6 in proper development of both lip and palate and how abnormal IRF6 expression driven by the distinct variants somehow affect distinct phenotypes, i.e., CP or CLP. In conclusion, the IRF6 MCS-9.7 regulatory region represents a mutational “hot” spot for both rare and common genetic variation that cause and contribute risk to all three forms of OFCs.

Supplemental Data

Supplemental Data include eight figures and six tables.

Declaration of Interests


Acknowledgements

FinnGen study, https://www.finngen.fi/en
Online Mendelian Inheritance in Man (OMIM), https://www.omim.org
The Genome Aggregation Database (gnomAD), https://gnomad.broadinstitute.org
UCSC Genome Browser, https://genome.ucsc.edu
Database of Global Administrative Areas, https://gadm.org
JASPAR, http://jaspar.genereg.net
Integrated DNA Technologies CRISPR-Cas9 guide RNA design checker, https://www.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE
ENCODER, https://www.encodeproject.org/
References


Figure 1. Manhattan plot showing GWAS results of 228 cases affected with nonsyndromic CP and 308,799 population controls. Negative log_{10} p values are plotted for each variant, represented by dots, against their chromosomal position. The threshold for genome-wide significance (p = 5 x 10^{-8}) is indicated by a red dashed line.
**Figure 2. Regional association plots of the chromosome 1 risk loci for nonsyndromic CP.** Data are shown for association on chromosome (A) 1p36.1 (GRHL3) and (B) 1q32.2 (IRF6). Negative log₁₀ p values are shown for variants within a 1Mb region centered at the reference SNP. The reference SNP is marked with a purple diamond, and pairwise LD ($r^2$) between the reference SNP and other variants are indicated by color. The $r^2$ values were estimated from high-coverage whole-genome sequences of 3,775 Finns. Both directly genotyped and imputed SNPs are plotted. Genomic coordinates are shown according to the human genome build GRCh38/hg38.
Figure 3: Geographic distribution and correlation of the rs570516915 variant allele frequency with regional prevalence of nonsyndromic CP in the FinnGen cohort.

A) Distribution of the rs570516915 G allele frequencies in 19 regions of Finland. Gradients in allele frequency across different regions are shown by the intensity of the blue shading. Regions are labeled by their corresponding codes from Statistics Finland.

B) Correlation between the regional prevalence of CP and risk allele frequency of rs570516915 among 320,471 individuals from the FinnGen study. There were no CP cases from the Åland Islands.
Figure 4: The rs570516915 variant disrupts enhancer activity of MCS-9.7.

A) Browser view of the human genome, GRCh37/hg19, focused on the rs570516915 variant. Tracks from top to bottom: rs570516915; next two tracks represent ATAC-Seq detected open chromatin peaks from HOIEC cells and HEPM cells. Next track is H3K27Ac marks illustrating rs570516915 as a part of enhancer element. Color coded bars represent chromatin status revealed by ChIP-Seq to various chromatin marks from ENCODE Project cell lines (ENCODE Project Consortium) and facial explants from human embryos at Carnegie stage (CS) 13-20, encompassing the time when palate shelves fuse where golden and yellow bars represent the active and weak enhancer element respectively: GM12878, B-cell derived cell line; ESC, embryonic stem cells; K562, myelogenous leukemia; HepG2, liver cancer; HUVEC, human umbilical vein endothelial cells; HMEC, human mammary epithelial cells; HSMM, human skeletal muscle myoblasts; NHEK, normal human epidermal keratinocytes; NHLF, normal human lung fibroblasts, CS13-CS20 are facial explants from human embryos. B) Scattered dot plot of relative luciferase activity for non-risk and risk alleles of rs556188853 and rs570516915 in GMSM-K cells. C) Chromatograms illustrating the CRISPR-Cas9 edited oral epithelial cells for both (TT and GG) genotypes of rs570516915 in GMSM-K cells, generated by CRISPR-Cas9 mediated homology-directed repair. D) Scattered dot plot of relative levels of IRF6 mRNA in edited vs parental cells assessed by qRT-PCR. Expression levels of IRF6 are normalized to expression of ACTB (beta-actin). Statistical significance is determined by Student’s t-test and NS represents non-significant.
**Figure 5:** rs570516915 disrupts an IRF6 binding site and perturbs positive autoregulation of IRF6 expression via the IRF6 MCS-9.7 enhancer. A) Consensus IRF6 binding motif from the JASPAR database of transcription factor DNA-binding preferences (Matrix ID: PB0036.1) and alignment of the variant site in different species, which shows that the ancestral allele A is completely conserved in mammals and is critical for binding of IRF6. Fold enrichment identified by ChIP-qPCR for B) anti-IRF6 and C) anti-H3K27Ac in GMSM-K cells harboring TT and GG genotypes for rs570516915 using primers specific to the IRF6 MCS-9.7 enhancer site and a region 103.7 Kb upstream IRF6 transcription start site and devoid of a predicted IRF6 binding site. Error bars refer to 3 technical replicates each from two biological replicates for anti-H3K27Ac ChIP and one for anti-IRF6 ChIP. Statistical significance is determined by Student’s t-test and NS represents non-significant. Lateral views of three independent transgenic embryos at E13.5 expressing lacZ directed by the IRF6 MCS-9.7 enhancer with the D) non-risk and E-F) risk alleles of rs570516915. Number of estimated plasmid integration copy numbers are shown for each embryo. G) Schematic model depicting autoregulation of IRF6 through the IRF6 MCS-9.7 enhancer region.
Table 1. Top independent variants associated with CP in the Finnish population

<table>
<thead>
<tr>
<th>Locus</th>
<th>SNP</th>
<th>RA</th>
<th>OR</th>
<th>95% CI</th>
<th>p value</th>
<th>RAF (%)</th>
<th>AF (%) in gnomAD (3.1.1.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cases</td>
<td>Controls</td>
</tr>
<tr>
<td>1p36.1</td>
<td>rs113965554* (rs41268753)</td>
<td>A (T)</td>
<td>2.73 (2.66)</td>
<td>2.03 – 3.66 (1.97 – 3.58)</td>
<td>2.32 x 10^-11 (1.39 x 10^-10)</td>
<td>13.3 (12.8)</td>
<td>4.9 (4.9)</td>
</tr>
<tr>
<td>1q32.2</td>
<td>rs7546104</td>
<td>A</td>
<td>1.95</td>
<td>1.64 – 2.33</td>
<td>9.53 x 10^-14</td>
<td>41.1</td>
<td>24.1</td>
</tr>
<tr>
<td>1q32.2</td>
<td>rs570516915</td>
<td>G</td>
<td>8.65</td>
<td>6.11 – 12.25</td>
<td>5.25 x 10^-14</td>
<td>12.5</td>
<td>1.2</td>
</tr>
</tbody>
</table>

*rs113965554 and rs41268753 (the etiologic missense variant p.Thr454Met in GRHL3) are in nearly complete LD ($r^2 = 0.99$) with each other. **Besides Finnish and non-Finnish Europeans, only 3 alleles were found in African/African Americans in the gnomAD database. The rs113965554 variant was not found in East Asian, Middle Eastern, South Asian, Latino/Admixed American, Amish, and Ashkenazi Jewish individuals. RA – risk allele; OR – odds ratio; CI – confidence interval, RAF – risk allele frequency, AF – allele frequency, NFE – Non-Finnish European.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FinnGenCleftPalateGWASSupplemental.pdf