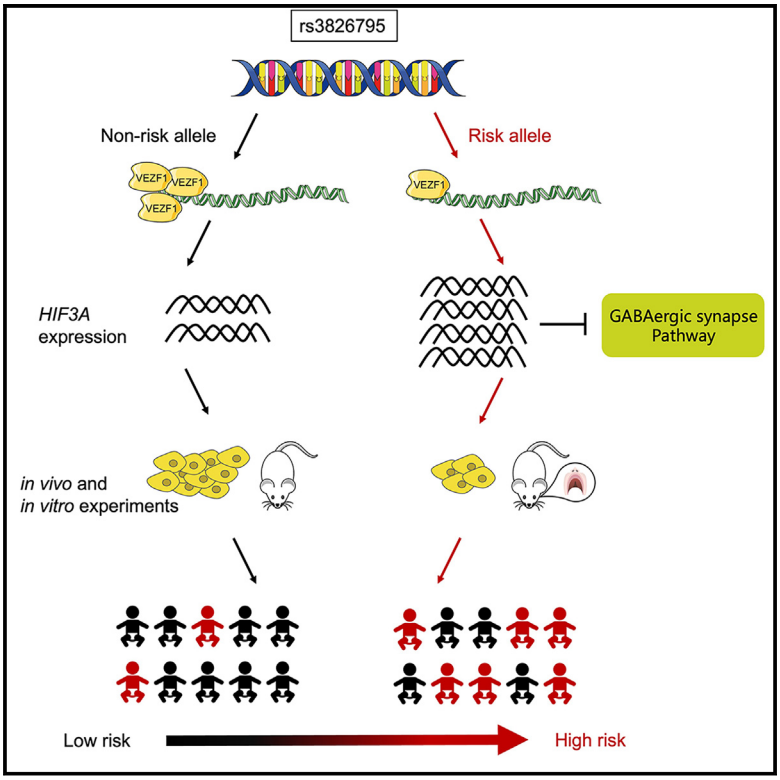
iScience

Article

Functional variant at 19q13.3 confers nonsyndromic cleft palate susceptibility by regulating *HIF3A*

Graphical abstract



Highlights

- Rs3826795 in *HIF3A* is significantly associated with the risk of NSCPO
- The A allele of rs3826795 enhances *HIF3A* expression by reducing VEZF1 binding
- *HIF3A* reduces cell proliferation and migration while promoting apoptosis
- *HIF3A* alters the GABAergic synapse pathway, providing insights into NSCPO biology

Authors

Shu Lou, Ziyue Miao, Xiaofeng Li, ..., Mulong Du, Lin Wang, Yongchu Pan

Correspondence

panyongchu@njmu.edu.cn (Y.P.), lw603@njmu.edu.cn (L.W.)

In brief

Quantitative genetics; Molecular interaction; Integrative aspects of cell biology; Association analysis; Transcriptomics

Lou et al., 2025, iScience *28*, 111829
February 21, 2025
https://doi.org/10.1016/j.isci.2025.111829

CellPress

iScience

CellPress
OPEN ACCESS

Article

Functional variant at 19q13.3 confers nonsyndromic cleft palate susceptibility by regulating *HIF3A*

Shu Lou,[1,2,3] Ziyue Miao,[1,2] Xiaofeng Li,[1,2] Lan Ma,[1,3] Dandan Li,[1,2,3] Mulong Du,[4] Lin Wang,[1,2,3,5,*] and Yongchu Pan[1,2,3,5,6,*]

[1]State Key Laboratory of Cultivation Base of Research, Prevention and Treatment for Oral Diseases, Nanjing Medical University, Nanjing 210029, China

[2]Department of Orthodontics, The Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing 210029, China

[3]Jiangsu Province Engineering Research Center of Stomatological Translational Medicine, Nanjing Medical University, Nanjing 210029, China

[4]Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China

[5]State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 211166, China

[6]Lead contact

*Correspondence: panyongchu@njmu.edu.cn (Y.P.), lw603@njmu.edu.cn (L.W.)

https://doi.org/10.1016/j.isci.2025.111829

SUMMARY

This study investigates the genetic factors associated with nonsyndromic cleft palate only (NSCPO) through a two-stage genome-wide association study (GWAS) and functional experiments. A total of 311 NSCPO cases and 1,136 controls were included, and we identified rs3826795 in *HIF3A* significantly associated with NSCPO risk ($p = 4.27E-09$). Functional assays showed that the A allele of rs3826795 reduced VEZF1 binding to the *HIF3A* promoter, thereby enhancing *HIF3A* expression. Additionally, we demonstrated that *HIF3A* significantly influenced cellular processes, including apoptosis, proliferation, and migration, in both *in vitro* and *in vivo* models. Pathway enrichment analysis indicated that *HIF3A* suppressed the GABAergic synapse pathway, which plays a crucial role in embryonic development. These findings provide valuable insights into the genetic mechanisms underlying NSCPO susceptibility and suggest potential targets for future therapeutic interventions.

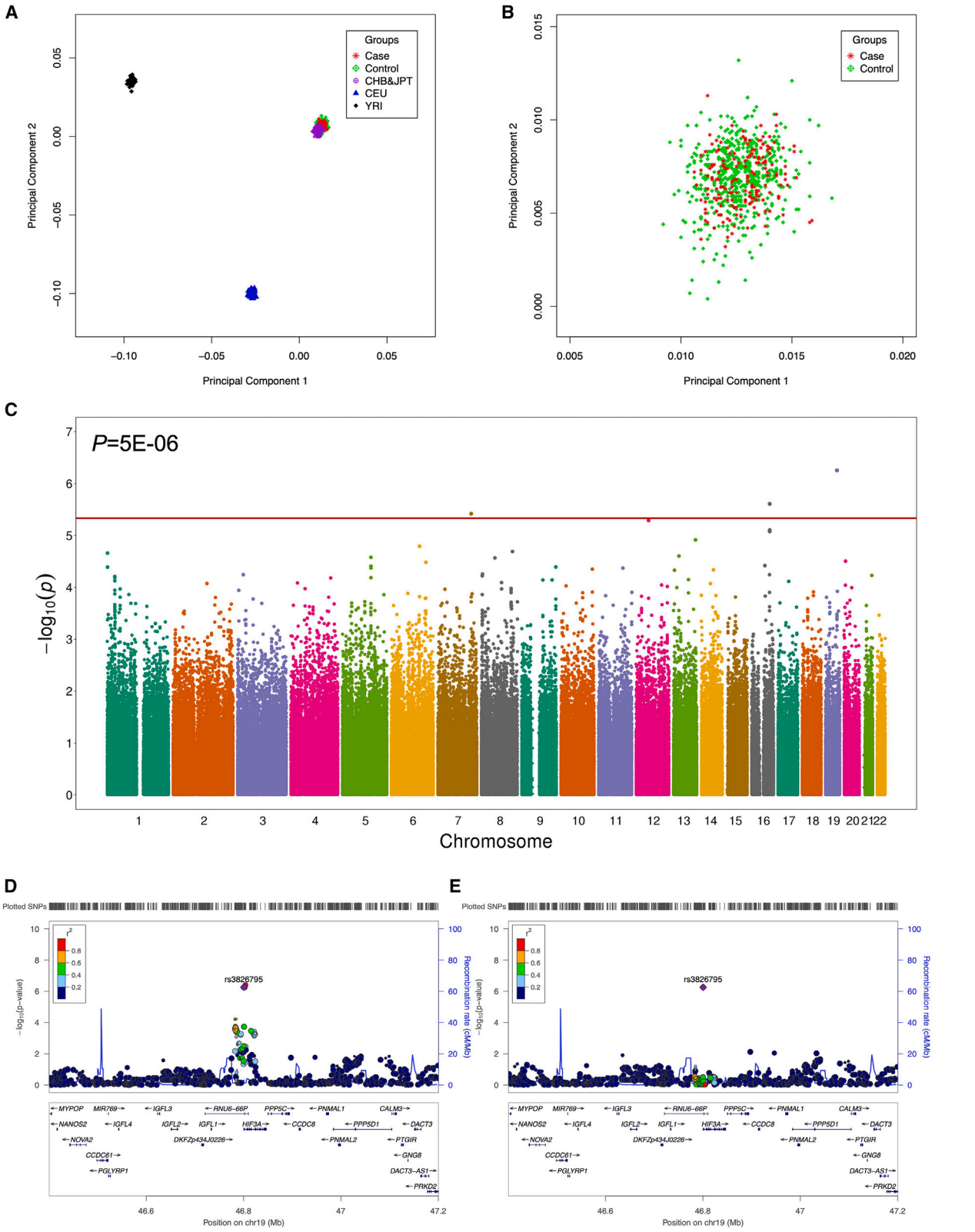
INTRODUCTION

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) and nonsyndromic cleft palate only (NSCPO), known as nonsyndromic orofacial cleft (NSOC), are some of the most common birth defects in humans, occurring in 1 of 700 live births worldwide.[1] NSOC occurs in the absence of other malformations or abnormalities and has a complex etiology involving multiple genetic and environmental risk factors.[2]

The substantial progress in identifying potential susceptibility genes for NSCL/P has been attributed to multiple genome-wide association studies (GWASs) over the past decades.[3–13] It has been conducted in the ancestry of European, Asian, American, and multi-ethnicity, and more than forty risk loci have been identified. However, the genetic risks associated with NSCPO are still not well understood, as only a limited number of GWAS have been conducted[13–18]; approximately a dozen susceptibility loci have been identified in the European, Asian, and African populations. Beaty et al. conducted a GWAS on 550 trios of European and Asian descent and found no single nucleotide polymorphisms (SNPs) reaching genome-wide significance.[14] However, a missense mutation in *GRHL3* (p.Thr454Met; rs41268753; $p = 4.08E-09$) was identified as being associated with NSCPO in the European population.[15] Huang et al. conducted a GWAS on NSCL/P and NSCPO in Asia, finding nine susceptibility variants associated with NSCPO, with *PAX9* identified as a key transcription factor.[16] In a recent study, researchers identified a novel locus (15q24.3) significantly associated with NSCPO in the Asian population.[17] Another GWAS on African NSCPO cases and controls identified new susceptibility loci on chromosomes 2 and 19.[18]

However, the elucidated susceptibility loci, mostly located in the intergenic regions without distinct evidence of functional signals, account for only a small fraction of the overall heritability of NSCPO. With the beginning of the post-GWAS era, current research focuses on exploring the biological mechanisms underlying specific susceptibility loci known as functional variants by applying novel and precise analytical methods to available GWAS data.[19,20] Accordingly, integrated molecular approaches involving both cell and animal models have been shown to be powerful for understanding the biological mechanisms of susceptibility loci in disease development.

To address the roles of genetic variants and genes involved in NSCPO susceptibility, here we performed a two-stage GWAS of NSCPO among 311 cases and 1,136 controls



(legend on next page)

recruited in a Chinese population. This approach enhances statistical power by initially screening a larger sample and then validating results in a focused group, thereby increasing the robustness and validity of our findings on genetic variants associated with NSCPO.[21] We identified a new locus at 19q13.3 (rs3826795 in gene encoding hypoxia inducible factor 3 subunit alpha (*HIF3A*); with p = 4.27E-09, odds ratio (OR) = 1.78, and 95% confidence interval (CI) = 1.47–2.16) associated with NSCPO risk. We conducted bioinformatics analysis and biological experiments were further conducted to explore the genetic mechanisms of the associated gene/region in the occurrence of NSCPO. A more comprehensive assessment of rs3836795 and *HIF3A* function enhanced our understanding of the underlying biological mechanisms in the susceptibility of NSCPO.

RESULTS

A novel independent nonsyndromic cleft palate only susceptibility variant was identified

An overview of the analysis strategy is presented in Figure S1. After quality control, 177 NSCPO cases and 513 controls were retained for the GWAS discovery stage, and a total of 701,516 autosomal SNPs were included in the genetic association analysis. The small genomic-control inflation factor (λ) of 1.04 indicated a low possibility of false-positive associations caused by population stratification in the quantile-quantile (Q-Q) plot (Figure S2). The principal components analysis (PCA) revealed that all of the analyzed GWAS participants were confirmed to be of Chinese ethnicity from the 1000 Genomes Project (Figure 1A); the cases and controls as genetically matched and homogeneous (Figure 1B). Manhattan plots from the GWAS discovery stage are shown in Figure 1C.

Genetic variants of the three regions were found to be associated with NSCPO risk at p values < 5E−06 in the discovery stage, including 7q33 (lead variant rs7805912: p = 3.80E−06), 16q22.3 (rs4641771: p = 2.46E−06) and 19q13.3 (rs3826795: p = 5.54E−07) (Table 1). Variants with p values < 5E−06 from each region were then genotyped with an additional 126 NSCPO cases and 621 control subjects of Chinese Han ethnicity. Rs3826795 showed significant associations in the same direction as that observed in the GWAS (p = 7.50E-04, OR = 1.57, 95% CI = 1.21–2.05). After combining the results from two stages, we confirmed that rs3826795 was significantly associated with NSCPO risk at genome-wide significance (P < 5E-08) without significant heterogeneity ($P_{heterogeneity}$ = 0.16, OR = 1.78, p = 4.27E-09, Table 1).

Fine-mapping revealed candidate functional variants

To identify the independent signal and increase the spectrum of variants tested for association in the significant region, we performed imputation with samples from the 1000 Genomes Project March 2012 release as the reference. After imputation across a 1-Mb region centered on the index SNP, we were able to test 18 SNPs in medium linkage disequilibrium (LD) ($r^2 \geq 0.5$) with the most significant genotyped SNP. Conditional logistic regression analysis of 19q13.3 revealed no secondary signals with independent effects from the top hits (Figures 1D and 1E; Table S1). In addition, we performed an in-silico analysis to annotate the functions of these 18 SNPs. A combination analysis with HaploReg and RegulomeDB showed that rs3826795 was predicted to be the most potential functional SNP among this region (Table S2).

To provide more candidate variants for downstream functional interpretations, we constructed credible sets that were 70% likely to contain the potentially functional SNPs based on posterior probability (PP) (Table S3). In this region, rs28654980 had the highest PP of 0.239, while the GWAS-discovered SNP rs3826795 ranked second with a PP of 0.196, indicating that both SNPs may be potential functional variants. Since the LD of these two SNPs was 1, a statistically significant difference was not observed between rs28654980 and the risk of NSCPO after correction for rs3826795. However, rs3826795 acted more biologically in functional prediction. Thus, rs3826795 was a potential functional variant in this region.

Gene-based analysis and expression quantitative trait locus analysis demonstrated hypoxia inducible factor 3 subunit alpha as a susceptibility gene

To identify the susceptibility genes that are regulated by rs3826795, we used the FUMA database to map candidate genes based on physical position, expression quantitative trait locus (eQTL) hits, and chromatin interaction mapping results (Figure S3). To evaluate the joint effects of common SNPs within gene units and enhance the detection of cumulative genetic effects, we further performed gene-based analysis using the sequence kernel association test (SKAT) and MAGMA methods to identify potentially associated genes and revealed that only *HIF3A* was significantly associated with risk of NSCPO both in two different methods (Table S4, p = 3.57E-02 via SKAT and p = 3.47E-05 via MAGMA).

We further performed an eQTL analysis based on the GTEx v8 database, which indicated that the expression of *HIF3A* expression was significantly up-regulated with the AA genotype of rs3826795 in four normal tissues and cells, including

Figure 1. Overview of NSCPO GWAS analysis. The PCA was performed to investigate

(A) the relevance between the GWAS subjects and 1000 Genomes Project population data from the HapMap project (Chinese and Japanese (CHB+JPT), European (CEU), African (YRI), and (B) the population structures between 185 cases and 515 controls on the discovery stage.

(C) Manhattan plot for genome-wide association study on discovery stage, where the y axis was defined as -$\log_{10}(p)$ and the genomic position of NCBI build 37 was represented along the x axis. The red horizontal line denotes a p value of 5E-06.

(D) Regional plots of association. Association results are shown for genotyped and imputed SNPs along with recombination rates showed using purple diamonds at 19q13.3.

(E) Association results are shown for genotyped and imputed SNPs after adjusting the lead SNP using conditional regression analyses at 19q13.3. For each plot, associations of individual variants are plotted as -$\log_{10}(p)$ against chromosomal position. The right y axis shows the recombination rate estimated based on the ASN 1000 Genomes Project Nov 2014 reference/hg19. GWAS, genome-wide association study; PCA, principal components analysis; SNP, single nucleotide polymorphism.

Table 1. The associations with NSCPO risk of the three variants in discovery and replication stages

Chr	SNP	Position (hg19)	Allele[a]	Gene	Stage	MAF Cases	MAF Controls	OR (95% CI)[b]	P^b	P_{het}[c]
7	rs7805912	134972466	G/A	*LOC107984123*	Discovery	0.260	0.327	0.48 (0.35–0.66)	3.80E-06	
					Replication	0.319	0.360	0.82 (0.61–1.11)	1.98E-01	
					Combined[d]			0.63 (0.37–1.06)	8.23E-02	1.57E-02
16	rs4641771	7363162	A/G	*C16orf47*	Discovery	0.157	0.097	2.76 (1.81–4.21)	2.46E-06	
					Replication	0.109	0.100	1.12 (0.71–1.77)	6.26E-01	
					Combined[d]			1.76 (0.73–4.27)	2.07E-01	4.42E-03
19	**rs3826795**	**46800433**	**A/G**	***HIF3A***	Discovery	0.597	0.472	2.07 (1.56–2.75)	5.54E-07	
					Replication	0.527	0.407	1.57 (1.21–2.05)	7.50E-04	
					Combined[d]			**1.78 (1.47–2.16)**	**4.27E-09**	**1.64E-01**

Significant results appear in boldface type.
SNP, single-nucleotide polymorphism; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.
[a]Minor allele/major allele.
[b]OR (95% CI) and *p* values were derived from logistic regression model with adjustment for sex, PC1, PC2 in discovery stage, and sex in replication stage under the assumption of an additive genetic model.
[c]*p* values were derived from Cochran's Q test of heterogeneity.
[d]Combined analysis was derived from discovery stage and replication stage using METAL under the assumption of a fixed model.

muscle-skeleton, cultured fibroblast cells, sun-exposed skin, and non-sun-exposed skin (Figures S4A–S4D). For further verification, we constructed the expression plasmids of rs3826795 [G] and [A], and transfected in the human embryonic palatal mesenchymal (HEPM) cell and the human oral keratinocytes (HOK), respectively. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) revealed that the transfection of the rs3826795 risk allele can significantly up-regulate the expression of *HIF3A* (Figures S4E and S4F).

Additionally, we found that the expression of *HIF3A* was increased in the dental pulp stem cells (DPSCs) of NSCL/P cases, compared with those in controls ($p = 2.18E-03$, Figure S5A). Moreover, *Hif3a* was continuously expressed in mouse lip and palate development during E10.5d to E15.5d as well as craniofacial structures from E10.5d to E14.5d (Figure S5B), implicating its essential role in the craniofacial development. Further analysis of the expression pattern of *Hif3a* demonstrated that *Hif3a* is significantly expressed in relevant facial structures associated with palate development compared to other craniofacial structures such as the brain and neural tube (Figures S5C and S5D).

Genetic variant rs3826795 regulated hypoxia inducible factor 3 subunit alpha as a promoter

To investigate the underlying mechanisms of rs3826795, we conducted a functional annotation analysis. A mammalian genome analysis from ENCODE data provided by the WashU Browser revealed that rs3826795 presented substantial enrichment of the H3K4me1, H3K4me3, and H3K27ac histone marks, indicating a potential promoter in this region based on chromatin immunoprecipitation sequencing (ChIP-seq) evidence from the human embryonic stem cells (hESC) and human keratinocyte cell lines. Enrichment of orofacial clefting and craniofacial measure associations in craniofacial enhancers and promoter annotations for early development enhancer state prediction (Figure S6).[22]

We thereby constructed promoter luciferase reporter vectors centered around rs3826795 and then tested the luciferase activity 48h after transfecting different plasmids in the human embryonic kidney 293 (HEK-293), HEPM, and HOK. The rs3826795 A allele significantly enhanced the promoter activity compared with that of the G allele (Figures 2A–2C).

To explore whether a base change of rs3826795 affects transcription factor binding preferences, we extracted the nuclear proteins from HEPM and HOK and performed an electrophoretic mobility shift assay (EMSA). The nuclear protein showed a weaker binding band with the A allele of rs3826795. A 50-fold excess of an unlabeled probe containing the G allele could completely abolish binding with the A allele, whereas the unlabeled A allele could not abolish binding with the G allele (Figure 2D). Thus, the A allele of rs3826795 weakened the binding ability of this region to transcription factors.

The A allele of rs3826795 decreased the binding of transcriptional repressor vascular endothelial zinc finger 1 with hypoxia inducible factor 3 subunit alpha promoter

Further *in silico* analysis revealed that rs3826795 was mapped within the binding motif of the transcription factor vascular endothelial zinc finger 1 (VEZF1) (Figure 2E). Based on the correlation analysis of Facebase (mouse embryo craniofacial tissue), mouse embryo lip and palate tissue, and mouse palate shelves,[23] the expression of *Vezf1* and *Hif3a* was found to be significantly negatively correlated ($p < 0.05$, Figure 2F). Inhibition of *VEZF1* expression with specific small interfering RNA (siRNA) duplexes resulted in an increased *HIF3A* expression (Figure S7). These findings suggest that VEZF1 may act as a potential transcriptional repressor of *HIF3A*.

Preferential binding of VEZF1 at the rs3826795 site was observed via chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR) (Figure 2G). VEZF1 antibody supershift assays further revealed that VEZF1 was an allele-preferential binder of rs3826795. The probe labeled with the A allele showed

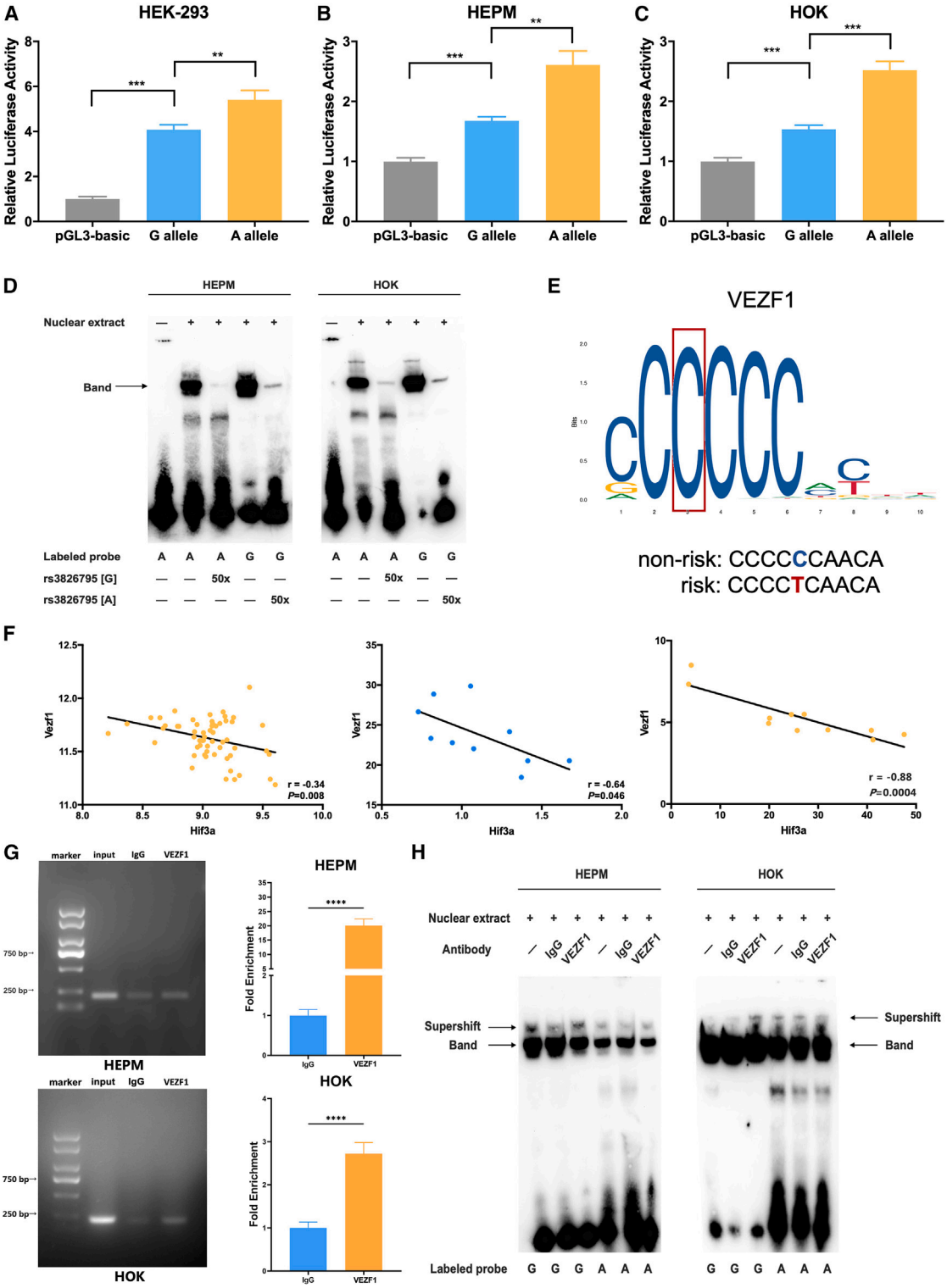


Figure 2. The variant rs3826795 at 19q13.3 regulates transcriptional activity of *HIF3A* by modulating the binding affinity of VEZF1 to the promoter

Allele-specific constructs containing the putative promoter sequence flanking rs3826795 were cloned into the pGL3-basic luciferase reporter vector and transfected into the (A) HEK-293, (B) HEPM, and (C) HOK. The results of luciferase activity were normalized to those of pGL3 basic (n = 6).

(legend continued on next page)

significantly weaker binding to VEZF1 compared to the probe labeled with the G allele (Figure 2H). Taken together, the A allele of rs3826795 weakened the binding affinity of VEZF1, a transcriptional repressor, to the promoter of *HIF3A*.

Hypoxia inducible factor 3 subunit alpha promoted apoptosis, and inhibited migration and proliferation *in vitro* and *in vivo*

To further address the function of *HIF3A*, we knocked down and overexpressed *HIF3A* in HEPM and HOK through the transfection of *HIF3A* siRNA-2 given its highest efficiency and plasmids (Figure S8). Flow cytometric analysis revealed a significantly decreased apoptosis rate in cells transfected with *HIF3A* siRNA and a significantly increased apoptosis rate in cells transfected with *HIF3A* overexpression plasmids (Figures 3A–3D). *HIF3A* knockdown caused increased migration levels, but opposite results were observed with *HIF3A* overexpression (Figures 3E–3G). Moreover, EdU assays revealed that *HIF3A* knockdown significantly promoted cell proliferation, whereas *HIF3A* overexpression significantly inhibited cell proliferation (Figure 3H).

Hif3a expression in the E13.5d-E15.5d mouse embryos was detected in the proximal maxillary process with a continuous expression in both the palatal epithelium and the mesenchyme, and the epithelial expression was stronger than that in the mesenchyme (Figure S9). To further verify the effects of *Hif3a*, we extracted palate shelves from mice embryos during the elevation and fusion stage of palate, and then transfected them with over-expressing *Hif3a* lentivirus (Figure 4A). The proliferation level of the *Hif3a* overexpression group was significantly decreased than that of the blank group and the control group, and the epithelial expression was particularly obvious. At this time, the apoptosis level of the palatal epithelium has increased significantly (Figures 4B–4E).

Hypoxia inducible factor 3 subunit alpha-related genes were enriched in the gamma-amino butyric acid ergic synapse pathway

To explore the potential pathways that were enriched in the *HIF3A*-related genes, we performed pathway analysis of RNA sequencing (RNA-seq) data of mouse embryonic palatal tissues overexpressing *Hif3a* and the control groups. The Pearson correlation coefficient was used to evaluate the correlation of each replicate, which revealed that five replicates in each group had a strong correlation (Figure 5A). Volcano plots of the differentially expressed genes in the three groups are shown in Figure 5B. The GABAergic synapse pathway was enriched as per both the Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene set enrichment analysis (GSEA) among the significantly different pathways (Figures 5C and 5D). We further verified the differentially expressed genes in HEPM and HOK, and found that among the 16 differentially expressed genes, the mRNA expression level of ten genes (*SLC6A1*, *GAD2*, *GNB3*, *GABRA5*, *CACNA1B*, *GABRB3*, *GABRG1*, *KCNJ6*, *GABRG2*, *SCL32A1*) decreased significantly when *HIF3A* was overexpressed (Figures 5E and 5F).

DISCUSSION

Previous studies have reported only a few susceptibility loci in NSCPO, which explains only a small fraction of NSCPO heritability in the Chinese population. Moreover, the molecular mechanisms of risk variants and their associated genes in NSCPO have rarely been comprehensively investigated. In the present study, we performed a two-stage case-control study and systematically evaluated the association between common variants and susceptibility to NSCPO. A novel variant (19q13.3) was identified that reached the genome-wide significance level. Further functional gene mapping and gene-based analysis supported *HIF3A* as a susceptibility gene for NSCPO.

HIF3A attenuates the ability of transcription factor *HIF1A* to bind to the hypoxia-response element (HRE) located within the enhancer/promoter of hypoxia-inducible target genes, thereby inhibiting HRE-driven transcriptional activation.[24] Severe congenital malformations, including heart defects, retinopathy, and brain damage, have been observed when mammalian embryos are cultured at ambient oxygen concentrations, which are mainly regulated by hypoxia-inducible factors (HIFs).[25] Both mouse and chicken embryos are in a low-oxygen environment during the early development process, which is essential for palatal shelf growth, elevation, and fusion.[26,27] *HIF3A* appears to be crucial for the initial adaptation of human embryonic stem cells to a hypoxic environment.[28] Moreover, according to the DECIPHER database, the duplication of *HIF3A* gene fragments can cause congenital developmental defects, such as cleft palate, growth retardation, microcephaly, earlobe hypoplasia, and abnormal cardiovascular system morphology.[29] Sharp et al. discovered that cg16672562, located in the first intron of *HIF3A* (1 kb away from rs3826795), is significantly hypermethylated in NSCPO compared to NSCLO in blood, suggesting that *HIF3A* methylation is involved in the occurrence of NSCPO.[30] The present study elucidated that rs3826795 G>A can enhance the transcriptional activity of the promoter region of *HIF3A* and up-regulate *HIF3A* expression. *In vivo* and *in vitro* experiments

(D) EMSA with biotin-labelled oligonucleotides containing the rs3826795[A] or rs3826795[G] allele and nuclear extracts from HEPM and HOK. Lanes 1 showed the mobility of the labeled oligonucleotides without nuclear extracts; lanes 2 and 4 showed the mobility of the labeled oligonucleotides with nuclear extracts in the absence of the competitor oligonucleotide and lanes 3 and 5 showed the mobility of the labeled oligonucleotides with nuclear extracts in the presence of unlabeled competitors. The arrow indicates a DNA–protein complex.

(E) *In silico* predicted preferential binding of VEZF1 to the non-risk allele C of rs3826795.

(F) The transcription factor VEZF1 was analyzed for correlation with *Hif3a* in mouse embryo craniofacial tissues, lip and palate tissues, palate shelves (from left to right).

(G) The ChIP enrichment of VEZF1 as determined by ChIP-qPCR. IgG was used as a negative control.

(H) Supershift assays using the VEZF1 antibody in HEPM and HOK. IgG was used as a negative control. The data are presented as the means ± SD, and the p values were calculated using Student's t test. ** indicates $p < 0.01$, *** indicates $p < 0.001$, and **** indicates $p < 0.0001$. EMSA, electrophoretic mobility shift assay; ChIP, chromatin immunoprecipitation. SD, standard deviation.

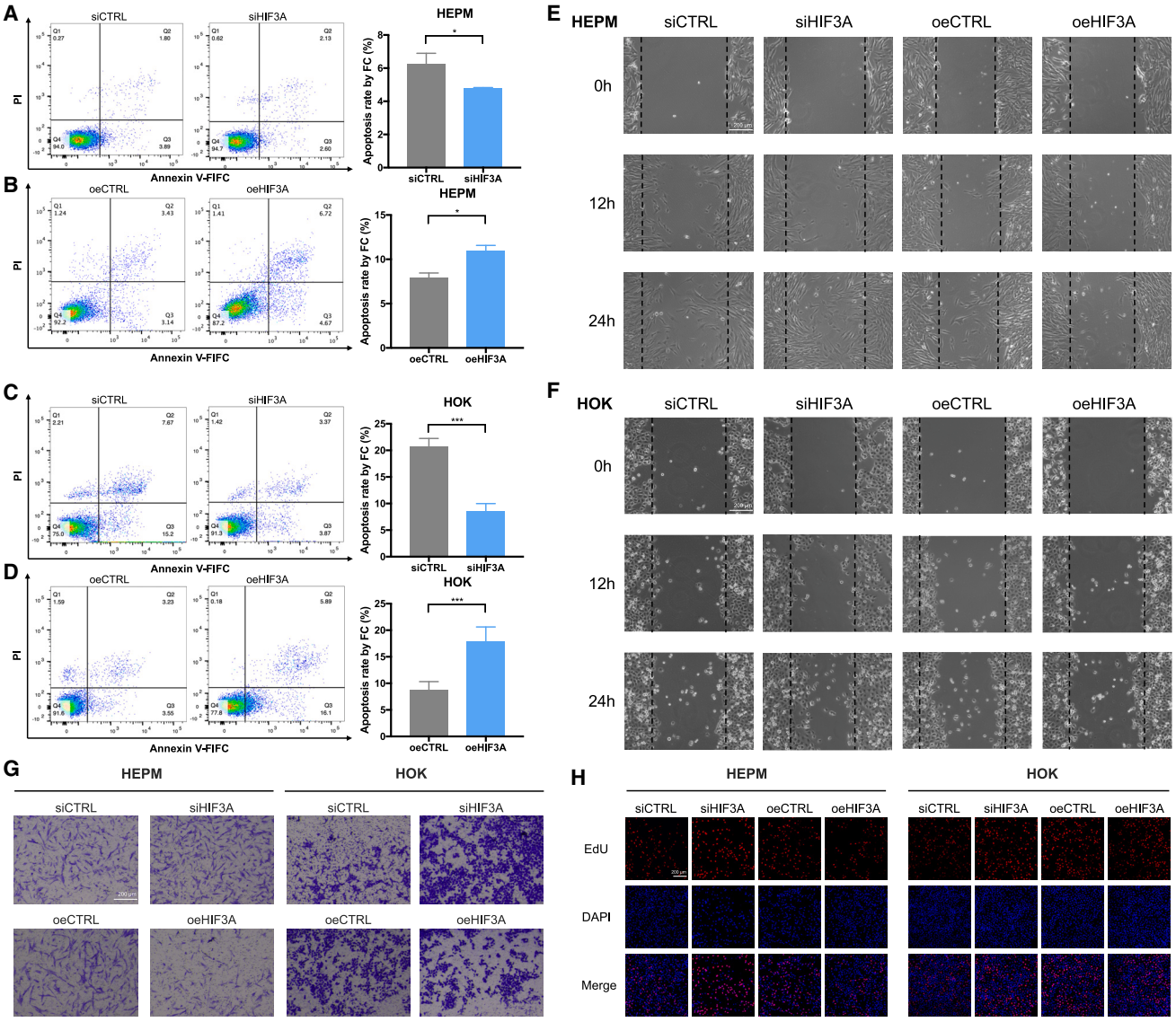


Figure 3. *In vitro* functional roles of *HIF3A* in HEPM and HOK

(A–D) Quantitative analysis of cell apoptosis by flow cytometry between the two groups ($n = 6$). The rate of apoptosis is lower with *HIF3A* knockdown and is higher than those in the corresponding negative controls in HEPM and HOK.

(E and F) Photos showing the migration of HEPM and HOK with knockdown and overexpression of *HIF3A* in 0h, 12h, and 24h after scratch. Overexpression of *HIF3A* significantly reduced the migration ability in HEPM and HOK (scale bar = 200 μm).

(G) Transwell assays was used to assess the migration of HEPM and HOK after transfection with *HIF3A* siRNAs or overexpression plasmids. The rate of migration is lower with *HIF3A* overexpression groups in HEPM and HOK (scale bar = 200 μm).

(H) EdU proliferation analysis before or after knockdown/overexpression of *HIF3A* in HEPM and HOK. High *HIF3A* expression significantly reduced the proliferation rate of HEPM and HOK (scale bar = 200 μm). The data are presented as the means ± SD, and the *p* values were calculated using Student's t test. * indicates $p < 0.05$, *** indicates $p < 0.001$.

revealed that *HIF3A* can inhibit proliferation and migration, promote apoptosis, and thus play an important role in the occurrence and development of the cleft palate.

VEZF1, an early developmental gene that encodes a zinc finger transcription factor, is expressed in endothelial cells during early embryonic development and is involved in regulating cell proliferation, migration, and network formation *in vitro*.[31,32] In a previous study, in *Vezf1* KO C57BL/6 mice, embryonic death and intra-embryonic hemorrhage were observed on E9.5d; on E16.5d, the surviving *Vezf1*$^{-/-}$ embryos exhibited a large number of developmental delays.[33] Our study suggests that the G allele of rs3826795 enhances the binding affinity of VEZF1, a transcriptional repressor, to the promoter of *HIF3A*.

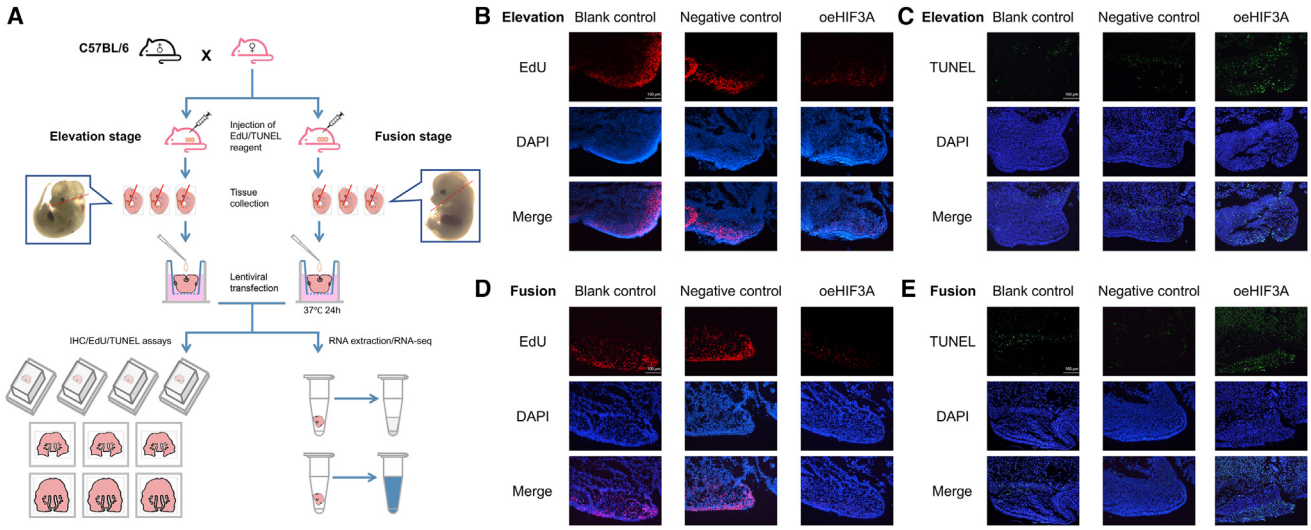


Figure 4. Effect of *Hif3a* overexpression on proliferation and apoptosis *in vivo*

(A) Schematic diagram of the mice experiment.

(B) Proliferation analysis before or after overexpressing of *Hif3a* in E13.5d mouse palate shelves (scale bar = 100 μm). High *Hif3a* expression significantly reduced the proliferation rate.

(C) Apoptosis analysis before or after overexpressing of *Hif3a* in E13.5d mouse palate shelves (scale bar = 100 μm). The rate of apoptosis is higher with *Hif3a* overexpression.

(D) Proliferation analysis before or after overexpressing of *Hif3a* in E14.5d mouse palate shelves (scale bar = 100 μm). High *Hif3a* expression significantly reduced the proliferation rate.

(E) Apoptosis analysis before or after overexpressing of *Hif3a* in E14.5d mouse palate shelves (scale bar = 100 μm). The rate of apoptosis is higher with *Hif3a* overexpression.

RNA-seq and pathway enrichment analyses were performed on the palate shelves of mice to identify the susceptible pathways. In the present study, the GABAergic synapse pathway was significantly enriched, and ten key genes were found to be downregulated in HEPM and HOK when *HIF3A* was overexpressed, indicating that *HIF3A* might participate in the pathogenesis of NSCPO through the GABAergic synapse pathway. The GABAergic synaptic pathway plays a crucial role in embryonic development.[34,35] During early pregnancy, GABA consumption can impair preimplantation embryo development via its B-type and endometrial receptivity, which disturbs early embryo implantation in mice.[36] Neuronal somata, growth cones, axons, and dendrites showed GABA-like immunoreactivity in developing *Xenopus laevis* embryos.[37] The GABAergic synapse pathway is potentially involved in palatogenesis.[38] A previous study reported regulated cholinergic signaling in human fibroblasts *in vitro* and differently regulated GABAergic signaling systems in NCPL/P fibroblasts and normal control cells.[39]

Therefore, the present study has advanced our understanding of the genetic architecture of NSCPO by revealing new genetic risk factors. We identified a functional variant, rs3826795, in *HIF3A* associated with NSCPO in a Chinese population. We found that rs3826795 G>A reduced the ability of the *HIF3A* promoter region to bind to the transcription repressor VEZF1, thereby upregulating the expression of *HIF3A*, leading to a specific biological phenotype.

Limitations of the study

The power of the present study is limited by the relatively small sample size, and larger cohorts are needed to validate our findings. Further sequencing and functional investigations are required to confirm the results. Specifically, future studies should focus on developing targeted mutation mouse models to assess the impact of the rs3826795 variant on phenotypic outcomes. Additionally, further exploration of how *HIF3A* influences the GABAergic synapse pathway is essential for gaining deeper insights into the genetic and functional mechanisms underlying NSCPO.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Yongchu Pan (panyongchu@njmu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The genotyping data reported in this article have been deposited in the Genome Variation Map. The RNA-seq data generated during the current study are available in NCBI. Accession numbers are listed in the key resources table.
- This article does not report original code.
- Any additional information required to reanalyze the data reported in this article is available from the lead contact upon request.

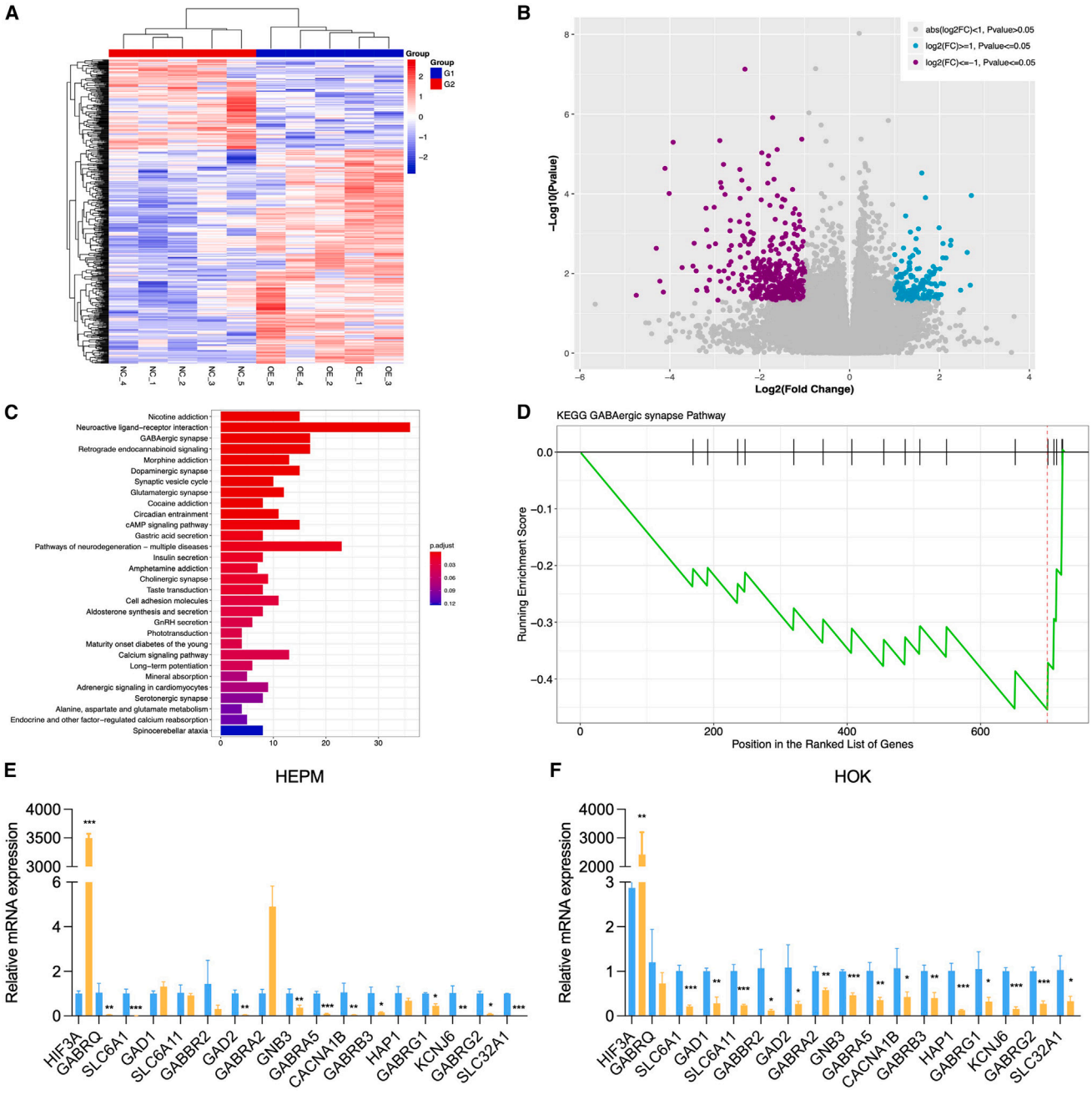


Figure 5. Pathway analysis with overexpression of *Hif3a* in mouse embryo

(A) Heatmap of differentially expressed genes between controls and overexpression of *Hif3a*.

(B) Volcano plot of differentially expressed genes between controls and overexpression of *Hif3a*.

(C) KEGG and (D) GSEA pathway enrichment analysis showed the association function of the genes reside on the susceptible region. The term/pathway were ranked by -$\log_{10}(P)$. The relative mRNA expression levels of all differentially expressed genes in GABAergic synapse pathway were assessed by qRT-PCR with (E) HEPM and (F) HOK. The data are presented as the means ± SD, and the *p* values were calculated using Student's t test. * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$. KEGG, Kyoto encyclopedia of genes and genomes; GSEA, gene set enrichment analysis.

ACKNOWLEDGMENTS

We thank all the study participants, research staff, and students who contributed to this study. We would like to thank Editage (www.editage.com) for English language editing. The graphical abstract was created in BioRender.com.

This work was supported by the National Natural Science Foundation of China (82301014, 82270946), the Key Research and Development Program of Jiangsu Province (BE2023831), the Natural Science Foundation of Jiangsu Province (BK20220309), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (22KJB320003, 22KJA320002), Chinese

Postdoctoral Science Foundation (2022M721677), Jiangsu Province Capability Improvement Project through Science, Technology and Education-Jiangsu Provincial Research Hospital Cultivation Unit (YJXYYJSDW4), and Jiangsu Provincial Medical Innovation Center (CXZX202227).

AUTHOR CONTRIBUTIONS

S.L.: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft, Writing – review and editing; Z.M. and X.L.: Investigation, Writing – review and editing; L.M. and D.L.: Writing – review and editing; M.D., L.W., and Y.P.: Conceptualization, Supervision, Writing – review and editing.

DECLARATION OF INTERESTS

The authors declare that they have no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Samples information
  - Cell culture
  - Mice
- METHOD DETAILS
  - Genotyping, quality control, and imputation
  - Fine-mapping analysis
  - Functional annotations
  - Gene-based analysis
  - Gene expression in human samples and during mouse craniofacial development
  - Luciferase activity assay
  - EMSA
  - ChIP assay
  - RNA extraction and real-time quantitative PCR
  - Cell apoptosis, cell migration, cell proliferation assay
  - Lentivirus infection, extraction of mouse embryo tissues and RNA-seq
  - Immunohistochemistry (IHC), EdU and TUNEL assays
  - Pathway enrichment analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.111829.

Received: January 30, 2024
Revised: August 8, 2024
Accepted: January 15, 2025
Published: January 16, 2025

REFERENCES

1. Mangold, E., Ludwig, K.U., and Nöthen, M.M. (2011). Breakthroughs in the genetics of orofacial clefting. Trends Mol. Med. *17*, 725–733. https://doi.org/10.1016/j.molmed.2011.07.007.
2. Leslie, E.J., and Marazita, M.L. (2013). Genetics of cleft lip and cleft palate. Am. J. Med. Genet. C Semin. Med. Genet. *163C*, 246–258. https://doi.org/10.1002/ajmg.c.31381.
3. Birnbaum, S., Ludwig, K.U., Reutter, H., 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. https://doi.org/10.1038/ng.333.
4. Grant, S.F.A., Wang, K., Zhang, H., Glaberson, W., Annaiah, K., Kim, C.E., Bradfield, J.P., Glessner, J.T., Thomas, K.A., Garris, M., et al. (2009). A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. J. Pediatr. *155*, 909–913. https://doi.org/10.1016/j.jpeds.2009.06.020.
5. Beaty, T.H., Murray, J.C., Marazita, M.L., Munger, R.G., Ruczinski, I., Hetmanski, J.B., Liang, K.Y., 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. https://doi.org/10.1038/ng.580.
6. Sun, Y., Huang, Y., Yin, A., Pan, Y., Wang, Y., Wang, C., Du, Y., 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. https://doi.org/10.1038/ncomms7414.
7. Mangold, E., Ludwig, K.U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., Reutter, H., 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. https://doi.org/10.1038/ng.506.
8. Yu, Y., Zuo, X., He, M., Gao, J., Fu, Y., Qin, C., Meng, L., Wang, W., Song, Y., Cheng, Y., et al. (2017). Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nat. Commun. *8*, 14364. https://doi.org/10.1038/ncomms14364.
9. Ludwig, K.U., Mangold, E., Herms, S., Nowak, S., Reutter, H., 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. https://doi.org/10.1038/ng.2360.
10. Ludwig, K.U., Ahmed, S.T., Böhmer, A.C., Sangani, N.B., Varghese, S., Klamt, J., Schuenke, H., Gültepe, P., Hofmann, A., Rubini, M., et al. (2016). Meta-analysis Reveals Genome-Wide Significance at 15q13 for Nonsyndromic Clefting of Both the Lip and the Palate, and Functional Analyses Implicate GREM1 As a Plausible Causative Gene. PLoS Genet. *12*, e1005914. https://doi.org/10.1371/journal.pgen.1005914.
11. Leslie, E.J., Carlson, J.C., Shaffer, J.R., Feingold, E., Wehby, G., Laurie, C.A., Jain, D., Laurie, C.C., Doheny, K.F., McHenry, T., et al. (2016). A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. Hum. Mol. Genet. *25*, 2862–2872. https://doi.org/10.1093/hmg/ddw104.
12. Leslie, E.J., Carlson, J.C., Shaffer, J.R., Butali, A., Buxó, C.J., Castilla, E.E., Christensen, K., Deleyiannis, F.W.B., Leigh Field, L., Hecht, J.T., et al. (2017). Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. Hum. Genet. *136*, 275–286. https://doi.org/10.1007/s00439-016-1754-7.
13. Ludwig, K.U., Bohmer, A.C., Bowes, J., Nikolic, M., Ishorst, N., Wyatt, N., Hammond, N.L., Golz, L., Thieme, F., Barth, S., et al. (2017). Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip +/- cleft palate and cleft palate only. Hum. Mol. Genet. *26*, 829–842. https://doi.org/10.1093/hmg/ddx012.
14. Beaty, T.H., Ruczinski, I., Murray, J.C., Marazita, M.L., Munger, R.G., Hetmanski, J.B., Murray, T., Redett, R.J., Fallin, M.D., Liang, K.Y., et al. (2011). Evidence for gene-environment interaction in a genome wide study of non-syndromic cleft palate. Genet. Epidemiol. *35*, 469–478. https://doi.org/10.1002/gepi.20595.
15. Leslie, E.J., Liu, H., Carlson, J.C., Shaffer, J.R., Feingold, E., Wehby, G., Laurie, C.A., Jain, D., Laurie, C.C., Doheny, K.F., et al. (2016). A Genome-wide Association Study of Nonsyndromic Cleft Palate Identifies

an Etiologic Missense Variant in GRHL3. Am. J. Hum. Genet. *98*, 744–754. https://doi.org/10.1016/j.ajhg.2016.02.014.

16. Huang, L., Jia, Z., Shi, Y., Du, Q., Shi, J., Wang, Z., Mou, Y., Wang, Q., Zhang, B., Wang, Q., et al. (2019). Genetic factors define CPO and CLO subtypes of nonsyndromicorofacial cleft. PLoS Genet. *15*, e1008357. https://doi.org/10.1371/journal.pgen.1008357.
17. He, M., Zuo, X., Liu, H., Wang, W., Zhang, Y., Fu, Y., Zhen, Q., Yu, Y., Pan, Y., Qin, C., et al. (2020). Genome-wide Analyses Identify a Novel Risk Locus for Nonsyndromic Cleft Palate. J. Dent. Res. *99*, 1461–1468. https://doi.org/10.1177/0022034520943867.
18. Butali, A., Mossey, P.A., Adeyemo, W.L., Eshete, M.A., Gowans, L.J.J., Busch, T.D., Jain, D., Yu, W., Huan, L., Laurie, C.A., et al. (2019). Genomic analyses in African populations identify novel risk loci for cleft palate. Hum. Mol. Genet. *28*, 1038–1051. https://doi.org/10.1093/hmg/ddy402.
19. Sud, A., Kinnersley, B., and Houlston, R.S. (2017). Genome-wide association studies of cancer: current insights and future perspectives. Nat. Rev. Cancer *17*, 692–704. https://doi.org/10.1038/nrc.2017.82.
20. Farashi, S., Kryza, T., Clements, J., and Batra, J. (2019). Post-GWAS in prostate cancer: from genetic association to biological contribution. Nat. Rev. Cancer *19*, 46–59. https://doi.org/10.1038/s41568-018-0087-3.
21. Wang, Y., Xiao, F., Zhao, Y., Mao, C.X., Yu, L.L., Wang, L.Y., Xiao, Q., Liu, R., Li, X., McLeod, H.L., et al. (2022). A two-stage genome-wide association study to identify novel genetic loci associated with acute radiotherapy toxicity in nasopharyngeal carcinoma. Mol. Cancer *21*, 169. https://doi.org/10.1186/s12943-022-01631-8.
22. Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J.P., and Cotney, J. (2018). High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. Cell Rep. *23*, 1581–1597. https://doi.org/10.1016/j.celrep.2018.03.129.
23. Huang, W., Zhong, W., He, Q., Xu, Y., Lin, J., Ding, Y., Zhao, H., Zheng, X., and Zheng, Y. (2023). Time-series expression profiles of mRNAs and lncRNAs during mammalian palatogenesis. Oral Dis. *29*, 2163–2176. https://doi.org/10.1111/odi.14237.
24. Depping, R., Steinhoff, A., Schindler, S.G., Friedrich, B., Fagerlund, R., Metzen, E., Hartmann, E., and Köhler, M. (2008). Nuclear translocation of hypoxia-inducible factors (HIFs): involvement of the classical importin alpha/beta pathway. Biochim. Biophys. Acta *1783*, 394–404. https://doi.org/10.1016/j.bbamcr.2007.12.006.
25. Semenza, G.L. (2001). Hypoxia-inducible factor 1: oxygen homeostasis and disease pathophysiology. Trends Mol. Med. *7*, 345–350.
26. Scully, D., Keane, E., Batt, E., Karunakaran, P., Higgins, D.F., and Itasaki, N. (2016). Hypoxia promotes production of neural crest cells in the embryonic head. Development *143*, 1742–1752. https://doi.org/10.1242/dev.131912.
27. Xu, J., Liu, H., Lan, Y., Aronow, B.J., Kalinichenko, V.V., and Jiang, R. (2016). A Shh-Foxf-Fgf18-Shh Molecular Circuit Regulating Palate Development. PLoS Genet. *12*, e1005769. https://doi.org/10.1371/journal.pgen.1005769.
28. Forristal, C.E., Wright, K.L., Hanley, N.A., Oreffo, R.O.C., and Houghton, F.D. (2010). Hypoxia inducible factors regulate pluripotency and proliferation in human embryonic stem cells cultured at reduced oxygen tensions. Reproduction *139*, 85–97. https://doi.org/10.1530/rep-09-0300.
29. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. *84*, 524–533. https://doi.org/10.1016/j.ajhg.2009.03.010.
30. Sharp, G.C., Ho, K., Davies, A., Stergiakouli, E., Humphries, K., McArdle, W., Sandy, J., Davey Smith, G., Lewis, S.J., and Relton, C.L. (2017). Distinct DNA methylation profiles in subtypes of orofacial cleft. Clin. Epigenet. *9*, 63. https://doi.org/10.1186/s13148-017-0362-2.
31. Xiong, J.W., Leahy, A., Lee, H.H., and Stuhlmann, H. (1999). Vezf1: A Zn finger transcription factor restricted to endothelial cells and their precursors. Dev. Biol. *206*, 123–141. https://doi.org/10.1006/dbio.1998.9144.
32. Miyashita, H., Kanemura, M., Yamazaki, T., Abe, M., and Sato, Y. (2004). Vascular endothelial zinc finger 1 is involved in the regulation of angiogenesis: possible contribution of stathmin/OP18 as a downstream target gene. Arterioscler. Thromb. Vasc. Biol. *24*, 878–884. https://doi.org/10.1161/01.ATV.0000126373.52450.32.
33. Kuhnert, F., Campagnolo, L., Xiong, J.W., Lemons, D., Fitch, M.J., Zou, Z., Kiosses, W.B., Gardner, H., and Stuhlmann, H. (2005). Dosage-dependent requirement for mouse Vezf1 in vascular system development. Dev. Biol. *283*, 140–156. https://doi.org/10.1016/j.ydbio.2005.04.003.
34. Kostović, I., and Judas, M. (2007). Transient patterns of cortical lamination during prenatal life: do they have implications for treatment? Neurosci. Biobehav. Rev. *31*, 1157–1168. https://doi.org/10.1016/j.neubiorev.2007.04.018.
35. Samarut, E., Swaminathan, A., Riche, R., Liao, M., Hassan-Abdi, R., Renault, S., Allard, M., Dufour, L., Cossette, P., Soussi-Yanicostas, N., and Drapeau, P. (2018). gamma-Aminobutyric acid receptor alpha 1 subunit loss of function causes genetic generalized epilepsy by impairing inhibitory network neurodevelopment. Epilepsia *59*, 2061–2074. https://doi.org/10.1111/epi.14576.
36. Tian, N., Liang, H., Luo, W., Wang, X., Cao, K., Zhang, Q., Tan, Y., and Tan, D. (2020). GABA consumption during early pregnancy impairs endometrial receptivity and embryo development in mice. J. Biochem. Mol. Toxicol. *34*, e22473. https://doi.org/10.1002/jbt.22473.
37. Roberts, A., Dale, N., Ottersen, O.P., and Storm-Mathisen, J. (1987). The early development of neurons with GABA immunoreactivity in the CNS of Xenopus laevis embryos. J. Comp. Neurol. *261*, 435–449. https://doi.org/10.1002/cne.902610308.
38. Baroni, T., Bellucci, C., Lilli, C., Pezzetti, F., Carinci, F., Becchetti, E., Carinci, P., Stabellini, G., Calvitti, M., Lumare, E., and Bodo, M. (2006). Retinoic acid, GABA-ergic, and TGF-beta signaling systems are involved in human cleft palate fibroblast phenotype. Mol. Med. *12*, 237–245. https://doi.org/10.2119/2006-00026.Baroni.
39. Baroni, T., Bellucci, C., Lilli, C., Pezzetti, F., Carinci, F., Lumare, E., Palmieri, A., Stabellini, G., and Bodo, M. (2010). Human cleft lip and palate fibroblasts and normal nicotine-treated fibroblasts show altered in vitro expressions of genes related to molecular signaling pathways and extracellular matrix metabolism. J. Cell. Physiol. *222*, 748–756. https://doi.org/10.1002/jcp.22006.
40. Yin, X., Ma, L., Li, Y., Xu, M., Wang, W., Wang, H., Yuan, H., Du, Y., Li, S., Ma, J., et al. (2017). Genetic variants of 20q12 contributed to non-syndromic orofacial clefts susceptibility. Oral Dis. *23*, 50–54. https://doi.org/10.1111/odi.12570.
41. Ma, L., Xu, M., Li, D., Han, Y., Wang, Z., Yuan, H., Ma, J., Zhang, W., Jiang, H., Pan, Y., and Wang, L. (2014). A miRNA-binding-site SNP of MSX1 is Associated with NSOC Susceptibility. J. Dent. Res. *93*, 559–564. https://doi.org/10.1177/0022034514527617.
42. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6. https://doi.org/10.1038/nmeth.2307.
43. 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. https://doi.org/10.1371/journal.pgen.1000529.
44. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. *8*, 1826. https://doi.org/10.1038/s41467-017-01261-5.

45. Timbers, T.A., Garland, S.J., Mohan, S., Flibotte, S., Edgley, M., Muncaster, Q., Au, V., Li-Leger, E., Rosell, F.I., Cai, J., et al. (2016). Accelerating Gene Discovery by Phenotyping Whole-Genome Sequenced Multi-mutation Strains and Using the Sequence Kernel Association Test (SKAT). PLoS Genet. *12*, e1006235. https://doi.org/10.1371/journal.pgen.1006235.

46. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219. https://doi.org/10.1371/journal.pcbi.1004219.

47. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191. https://doi.org/10.1093/bioinformatics/btq340.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-VEZF1	Santa	Cat# sc-365560 X
Anti-HIF3A	ThermoFisher Scientific	Cat# PA5-99079
Anti-IgG	Cell Signaling Technology	Cat# 2729
Bacterial and virus strains		
Lentivirus Hif3a overexpression	GENECHEM	N/A
pcDNA3.1-rs3826795 G/A	Nanjing Genebay Biotechnology	N/A
pGL3-basic-rs3826795 G/A	Nanjing Genebay Biotechnology	N/A
Biological samples		
NSCPO cases and healthy controls	Stomatological Hospital of Jiangsu Province, Nanjing Children's Hospital, Xuzhou First People's Hospital, and Huai'an First People's Hospital	N/A
Chemicals, peptides, and recombinant proteins		
Lipofectamine 2000	Invitrogen	Cat# 11668019
OPTI-MEM	GIBCO	Cat# 31985070
FBS	GIBCO	Cat# 16000044
RIPA buffer	Beyotime	Cat# p0013B
Trypsin	GIBCO	Cat# 25200072
Penicillin-Streptomycin	GIBCO	Cat# 15140122
TRIzol	Invitrogen	Cat# 10296010
PBS	Biosharp	Cat# BL302A
4% paraformaldehyde	Biosharp	Cat# 143174
Critical commercial assays		
PrimeScript™ RT Master Mix (Perfect Real Time)	TaKaRa	Cat# RR036A
BCA protein assay kit	Beyotime	Cat# P0012S
Dual Luciferase Reporter Gene Assay Kit	Promega	Cat# E1910
NE-PER Nuclear and Cytoplasmic Extraction Kit	Thermo	Cat# 78833
LightShift EMSA Optimization & Control Kit	Thermo	Cat# 20148X
Magna ChIP A/G Chromatin Immunoprecipitation Kit	Millipore	Cat# 17-10085
Cell-light™ EdU Apollo567 *in Vitro* Kit	RIBOBIO	Cat# C10310-1
TUNEL assay	Roche	Cat# 12156792910
Deposited data		
Genotyping data	This paper	GVM000806
RNA-seq data	This paper	PRJNA1132506
GEO database (GSE42589 and GSE67985)	Open-source	http://www.ncbi.nlm.nih.gov/geo/
GTEx database	Open-source	https://www.gtexportal.org/
Experimental models: Cell lines		
HEPM	American Type Culture Collection	ATCC® CRL-1486™
HOK	Shanghai Jining Industrial	N/A
HEK-293	American Type Culture Collection	ATCC® CRL-1573™

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
C57BL/6 mice	Animal Center of Yangzhou University	N/A
Oligonucleotides		
Probe Sequence for EMSA, see Table S6	This paper	N/A
Primers for qRT-PCR, see Table S7	This paper	N/A
siRNA sequence, see Table S8	This paper	N/A
Software and algorithms		
Plink 1.09	Open-source	https://www.cog-genomics.org/plink/1.9
METAL	Open-source	http://csg.sph.umich.edu/abecasis/metal/index.html
MAGMA	Open-source	https://ctg.cncr.nl/software/magma
FUMA	Open-source	https://fuma.ctglab.nl
R 3.4.2	Open-source	https://www.r-project.org
SHAPEIT	Open-source	https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
IMPUTE2	Open-source	https://mathgen.stats.ox.ac.uk/impute/impute_v2.html
LocusZoom 1.3	Open-source	http://locuszoom.sph.umich.edu
FlowJo	BD Biosciences	FlowJo v10
Prism 9.0	GraphPad	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Samples information

This study was approved by the Ethics Committee of Nanjing Medical University (NJMUERC [2008] No. 20). The discovery stage comprised 185 NSCPO cases and 515 controls recruited from Stomatological Hospital of Jiangsu Province, Nanjing Children's Hospital, Xuzhou First People's Hospital, and Huai'an First People's Hospital, as previously described.[40,41] The replication stage included 126 cases and 621 controls, who were recruited from Stomatological Hospital of Jiangsu Province (Table S5). The samples in the replication stage were independent of those in the discovery stage. All participants were interviewed and clinically assessed to ensure that individuals with other congenital anomalies, recognized CL/P syndromes, or developmental delays were excluded from the study. Venous blood samples were collected from all participants for genetic analysis. At recruitment, informed written consent was obtained from all the participants or their guardians.

Cell culture

The HEPM and the HEK-293 cells were purchased from American Type Culture Collection (ATCC, Manassas, VA, USA), and cultured in Eagle's Minimum Essential Medium (ATCC), supplemented with 10% fetal bovine serum (FBS, Gibco), 100 units/mL antibiotics at 37°C under 5% CO_2. The HOK were obtained from BeNa Culture Collection (Beijing, China), and cultured in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 10% FBS (Gibco) and 100 units/ml antibiotics and maintained at 37°C under 5% CO_2.

Mice

Six adult male and twelve adult female C57BL/6 mice were purchased from the Animal Center of Yangzhou University. The mice were housed in specific pathogen-free (SPF) conditions. Two female mice and one male mouse were placed together in the same cage at 8 p.m. and separated at 8 a.m. the following morning. Embryos were counted on the embryonic 0.5 days (E0.5d) in the morning, and the vaginal plug was checked. All animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC) of Nanjing Medical University (Approval Code: IACUC-2206010).

METHOD DETAILS

Genotyping, quality control, and imputation

All samples from the discovery stage were genotyped using the Illumina HumanOmniZhongHua-8 BeadChip (871,498 SNPs). As shown in Figure S1, quality control on discovery stage was performed on raw data using PLINK 1.09 to filter both unqualified samples and SNPs. Samples were excluded if they (i) had overall successful genotyping call rates<95%, (ii) had sex discrepancies between

the records and the genetically inferred data or (iii) were ancestry outliers or heterozygosity outliers by identify-by-decent analysis. SNPs were excluded if they (i) did not map to autosomal chromosomes, (ii) had a missing rate>5%, (iii) had a minor allele frequency (MAF) < 0.05 or (iv) significant deviation from Hardy–Weinberg equilibrium (HWE) in the controls (P < 1E-05). The Q-Q plot demonstrated distributions of the observed p values and those expected by chance. The genetic backgrounds of all cases and controls were assessed by PCA.

We selected variants with suggestive significance for genotyping during the replication stage. DNA extracted from venous blood samples were genotyped for all cases and controls. Whole exome sequencing was conducted for cases from replication stage using Agilent SureSelectXT Human All Exon KitV7. All controls from replication stage were genotyped by Illumina Infinium Global Screening Array (GSA) v1.0.

We analyzed genotyping data during replication stage from the 1000 Genomes Project (version 1) using the SHAPEIT software[42] and IMPUTE2 software[43] for prephasing and imputation, respectively. Only imputed high-quality genotypes (INFO score >0.8) were included in further association analysis.

Fine-mapping analysis

Stepwise conditional analysis was performed to explore independently effective loci in the 19q13.3 region in a forward stepwise manner. In each step, the most significant variant was considered as covariates in the next logistic model until no variants satisfied the significance threshold.

To define a potentially functional variant for each locus, we calculated the PPs using the approximate Bayesian factor (ABF) under the assumption of a single variant. We estimated the ABF for each variant as follows:

$$ABF = \sqrt{1 - r}e^{(r^2/2)},$$

where

$$r = \frac{0.04}{\left(SE^2 + 0.04\right)},$$

The SE are the corresponding standard error resulting from testing for association under a logistic regression model. The PP for each variant was obtained as follows:

$$Posterior\ Probability_i \frac{ABF_i}{T}$$

where ABF_i corresponds to the approximate Bayes' factor for the marker i and T represents the sum of all the ABF values from the candidate variants enclosed in the interval being evaluated.

Functional annotations

The SNPs were annotated for potential regulatory functions using RegulomeDB (http://www.regulomedb.org) and HaploReg v4.1 (http://www.broadinstitute.org/mammals/haploreg/haploreg.php). The regulatory features were obtained from the Encyclopedia of DNA Elements (ENCODE) and NIH Roadmap Epigenomics data using the UCSC Genome Browser (http://genome.ucsc.edu/). The SNPs were then mapped to genes based on the physical position of the genome, eQTL associations (all tissues because of lack of craniofacial-specific tissue), and chromatin interaction mapping using FUMA.[44]

Gene-based analysis

Apart from examining the separate effect of the single genetic variant, we aimed to identify NSCPO related genes through evaluating the overall effects of all the SNPs in the region. The quality control criteria for SNPs were set at MAF >0.05 and Hardy-Weinberg equilibrium P > 1E-06. The SKAT[45] and MAGMA tool[46] were used for gene-based analysis using discovery datasets in the present study with adjustment for sex, principal component (PC) 1, and PC2.

Gene expression in human samples and during mouse craniofacial development

Differential expression data for genes in both NSCL/P cases and controls of DPSCs were obtained from the Gene Expression Omnibus (GEO) repository (GSE42589). RNA-seq data were downloaded from the FaceBase consortium at FaceBase.org (https://www.facebase.org/, GSE67985) and the ENCODE database for mouse development. The count data were normalised using regularised logarithmic transformation in the DESeq2 R package.

Luciferase activity assay

To evaluate whether there was a difference in the promoter activity of *HIF3A* for different alleles, we performed the luciferase activity assay. A sequence of 1,000-bp containing rs3826795 G or A allele and *HIF3A* promoter region were synthesized and cloned into the *Nhe*I and *Xho*I restriction sites of the pGL3-basic vector (Promega). HEK-293, HEPM and HOK were seeded at in 24-well culture plates, and each well was transiently transfected with reporter plasmids using Lipofectamine 2000 (Invitrogen, Carlsbad, CA)

according to the manufacturer's instructions. All plasmids were co-transfected with 10 ng of pRL-SV40 containing the *Renilla* luciferase gene. Luciferase activity was measured 48 h after transfection, using a dual-luciferase reporter assay system (Promega). The ratio of firefly luciferase activity to *Renilla* luciferase activity was calculated. Transfection experiments were performed in triplicate.

EMSA

Nuclear extracts were prepared from HEPM and HOK using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Scientific, Waltham, MA, USA). DNA oligonucleotides for the variant were 5′-biotin-labelled and HPLC-purified (Sangon Biotech, Shanghai, China). Probes and nuclear extracts were incubated as per the instructions of the LightShift EMSA Optimization & Control Kit (Thermo Scientific) at 4°C for 20 min. For competition assays, unlabelled competitors with 50-fold more oligonucleotides were added to the reaction mixture 10 min before the addition of the labeled probes. For supershift assay, supershift antibody (VEZF1, Cell Signaling Technology, Danvers, MA, USA) or normal rabbit IgG (Millipore, Billerica, MA, USA) was incubated with nuclear extract before incubation with poly(dI:dC) at 4°C for 1 h. After incubation, binding was separated on a 6% polyacrylamide gel and transferred blots were developed using the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific). Probe sequence used for EMSA are shown in Table S6.

ChIP assay

The HEPM and HOK were cross-linked with 1% formaldehyde at room temperature for 10 min. The nuclear extracts were sonicated to generate chromatin fragments of 200-1,000 bp, which were immunoprecipitated using a ChIP assay kit (Upstate Biotechnology); the antibodies used were anti-VEZF1 (ab53570; Abcam) and anti-rabbit IgG (2729; Cell Signaling Technology). Immunoprecipitation was assessed using gel electrophoresis and qRT-PCR assays. Enrichment was expressed as the ratio of VEZF1 or IgG to that of the input control. Data points and error bars represent mean and standard deviation (SD), respectively, calculated from triplicate experiments.

RNA extraction and real-time quantitative PCR

To compare the gene expression in cells and tissues with different genotypes, we constructed plasmids pcDNA3.1-rs3826795 G and pcDNA3.1-rs3826795 A, which were transfected into cells and harvested after 48 h. Total RNA was extracted from the cells for reverse transcription and PCR using the PrimeScript RT-PCR kit (TaKaRa, Shiga, Japan). The relative mRNA expression levels of *HIF3A* and *GAPDH* (internal control) were assessed using an ABI 7900 Real-Time PCR system (Applied Biosystems). The primer sequences involved in qRT-PCR are listed in Table S7. All reactions were conducted in triplicate and data were analyzed using the $2^{-\Delta\Delta Ct}$ method. Data are shown as the mean ± SD.

Cell apoptosis, cell migration, cell proliferation assay

The cells were treated with trypsin-EDTA (Gibco) and resuspended in a single-cell suspension 48 h after transfection; they were stained using the Annexin V:PE Apoptosis Detection Kit (BD Biosciences, San Jose, CA, USA) for cell apoptosis analysis and analyzed using a fluorescence-activated cell sorting (FACS) system (BD Biosciences, San Jose, CA).

For cell migration, the cell suspension was added to the upper layer of the transwell chamber (Millipore) and 20% FBS medium was added to the lower chamber. The transwell chamber was incubated for 36 h. Thereafter, the chamber was removed and the cells were fixed with 4% paraformaldehyde (Biosharp, Hefei, China) for 30 min and stained with crystal violet. The invading cells on the lower surface were observed using an inverted microscope (Nikon, SMZ800N).

Cell proliferation was assessed using the EdU reagents to observe proliferation as per the manufacturer's instructions. The samples were analyzed using the fluorescence microscope (Leica DM4000). Independent experiments were performed in triplicate. All data are shown as mean ± SD. Table S8 lists the sense and antisense sequences for siRNAs.

Lentivirus infection, extraction of mouse embryo tissues and RNA-seq

The lentiviral *Hif3a* constructs were purchased from GENECHEM (Shanghai, China). A lentivirus with a non-targeting sequence was used as a negative control. The palate tissues of the embryos were collected. Total RNA was extracted with TRIzol reagent (Invitrogen, Carlsbad, CA, USA) from the embryonic tissues of mice; RNA quantity and quality was assessed using a Nanodrop and 1% agarose electrophoresis. RNA-seq was performed on the Hiseq3000 platform at 10 M and 6 G depths. RNA reads were aligned to the reference mouse genome (MM10) using gSNAP, and the average fragments per kilobase of exon model per million fragments mapped (FPKM) value of all samples was used to normalise mRNA expression.

Immunohistochemistry (IHC), EdU and TUNEL assays

Paraffin-embedded tissue blocks from embryos (E13.5d-E15.5d) were cut into 5-μm section. The sections were deparaffinized in xylene, hydrated with gradient ethanol, treated with 0.01 M sodium citrate for 20 min for antigen retrieval, immersed in 3% H_2O_2, and finally incubated with 10% goat serum. Tissue sections were incubated at 4°C overnight with the rabbit monoclonal anti-HIF3A antibody. Finally, the sections were stained with peroxidase-conjugated avidin, followed by treatment with 3,3-diaminobenzidine tetrahydrochloride (DAB). Haematoxylin blue was used for counterstaining.

In vivo proliferation levels were investigated using an EdU proliferation assay kit. EdU (0.1 mg) was diluted with 100 μL saline, injected intraperitoneally, washed with 2 mg/mL glycine solution, incubated for 10 min, and stored for photography. TUNEL assay was performed according to the manufacturer's instructions (Roche, #12156792910). Staining was performed with 4′,6-diamidino-2-phenylindole (DAPI; 1:1,000), and the samples were visualised using a laser confocal scanning microscope (Carl Zeiss, Heidenheim, Germany) and a Leica DM 4000 fluorescence system.

Pathway enrichment analysis

Differentially expressed genes were filtrated with $|logFC| > 1$ and p value <0.05 by R package DESeq2. Pathway enrichment analysis was conducted on the RNA-seq data using KEGG and GSEA by function GSEA in R package clusterProfiler (v 4.0.5).

QUANTIFICATION AND STATISTICAL ANALYSIS

We performed a two-stage case-control analysis of NSCPO in Chinese Han population. PLINK 1.09 and R 3.4.2 (http://www.r-project.org/) were used for statistical analysis. Population structure was evaluated by PCA using the EIGENSTRAT 3.0 software package. The OR, 95% CIs, and p values were calculated using an additive model in logistic regression analysis with adjustments for sex, PC1, and PC2 at the discovery stage. The heterogeneity of stages was tested using Cochran's Q statistic and I^2 using the R software. A combined analysis of the two stages was performed using a fixed-effect model and weighted by the estimated standard error (SE) in METAL.[47] A Manhattan plot of $-\log_{10}P$ was generated using the ggplot2 package in the R software. Regional plots were generated using the LocusZoom 1.3. For all graphs, statistical analyses were performed using a two-tailed unpaired Student's t test. Before t-test, the normal distribution of all data was checked using a normality test, and the equality of variances was checked using F-test. Data were considered statistically significant at $p < 0.05$. * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$, **** indicates $p < 0.0001$.