# Integrative Multi-omics Analysis Identifies Genetic Variants Contributing to Non-syndromic Cleft Lip with or without Cleft Palate

Shu LOU[1,2,3,4,#], Jing YANG[1,#], Gui Rong ZHU[1,#], Dan Dan LI[1,2,3,4], Lan MA[1,3,4], Lin WANG[1,2,3,4], Yong Chu PAN[1,2,3,4]

**Objective:** *To provide novel insights into the aetiology of non-syndromic cleft lip with or without cleft palate (NSCL/P) by integrating multi-omics data and exploring susceptibility genes associated with NSCL/P.*

**Methods:** *A two-stage genome-wide association study (GWAS) of NSCL/P was performed, involving a total of 1,069 cases and 1,724 controls. Using promoter capture Hi-C (pCHi-C) datasets in human embryonic stem cells (hESC) and chromatin immunoprecipitation sequencing (ChIP-seq) in craniofacial tissues, we filtered out single nucleotide polymorphisms (SNPs) with active cis-regulation and their target genes. Additionally, we employed expression quantitative trait loci (eQTL) analysis to identify candidate genes.*

**Results:** *Thirteen SNPs were identified as cis-regulation units associated with the risk of NSCL/P. Five of these were proven to be active in chromatin states in early human craniofacial development (rs7218002: odds ratio [OR] 1.50, P = 8.14E-08; rs835367: OR 0.78, P = 3.48E-05; rs77022994: OR 0.55, P = 1.05E-04; rs961470: OR 0.73, P = 1.38E-04; rs17314727: OR 0.73, P = 1.85E-04). Additionally, pCHi-C and eQTL analysis prioritised three candidate genes (rs7218002: NTN1, rs835367: FGGY, LINC01135). NTN1 and FGGY were expressed in mouse orofacial development. Deficiencies in NTN1, FGGY and LINC01135 were associated with cleft palate and cleft lip, abnormal facial shape and bifid uvula, and abnormality of the face, respectively.*

**Conclusion:** *Our study identified five SNPs (rs7218002, rs835367, rs77022994, rs961470 and rs17314727) and three susceptibility genes (NTN1, FGGY and LINC01135) associated with NSCL/P. These findings contribute to a better understanding of the genetic factors involved.*

**Keywords:** *ChIP-seq, expression quantitative trait loci, non-syndromic cleft lip, pCHi-C*

*Chin J Dent Res 2024;27(1):65–73; doi: 10.3290/j.cjdr.b5136745*

Non-syndromic cleft lip with or without cleft palate (NSCL/P) is one of the most common human birth defects, occurring in 1 in 700 live births and imposing heavy health and financial burdens on individuals.[1]

1 Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, P.R. China.
2 Department of Orthodontics, Affiliated Hospital of Stomatology, Nanjing Medical University, Nanjing, P.R. China.
3 State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, P.R. China.
4 Jiangsu Province Engineering Research Center of Stomatological Translational Medicine, Nanjing Medical University, Nanjing, P.R. China.

# These authors contributed equally to this study.

**Corresponding authors:** Dr Yong Chu PAN and Dr Lin WANG, Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, 136 Hanzhong Road, Nanjing 210029, P.R. China. Tel: 86-25-86862025. Email: panyongchu@njmu.edu.cn; lw603@njmu.edu.cn

This work was supported by the National Natural Science Foundation of China (82270496, 81830031, 81970969, 82001088, 82101054), the Natural Science Foundation of Jiangsu Province (BK20220309), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (22KJB320003, 22KJA320002), the Chinese Postdoctoral Science Foundation (2022M721677), the Young Talents Project of the Orthodontics Committee of the Chinese Stomatological Association (COS-B2021-09), Jiangsu Province Capability Improvement Project through Science, Technology and Education-Jiangsu Provincial Research Hospital Cultivation Unit (YJXYYJSDW4), and Jiangsu Provincial Medical Innovation Center (CXZX202227).

These defects arise during the embryonic period, specifically between the sixth and twelfth weeks, when there is a failure of fusion between the maxillary and medial nasal prominences, followed by the primary palate with the two lateral palatal shelves.[2]

In recent years, numerous genome-wide association studies (GWASs) have identified thousands of genetic variants associated with various phenotypes and diseases.[1,3-5] So far, at least 190 single nucleotide polymorphisms (SNPs) have been discovered to be related to NSCL/P according to the GWAS catalogue. Nevertheless, caution must be exercised when interpreting these associations, as GWAS findings often include both true driver variants that contribute to disease and passenger variants that are merely correlated with the SNPs susceptible to disease.[6,7] To gain a deeper understanding of the functional mechanisms underlying these associations, it is crucial to integrate genetic, transcriptional, epigenetic and other datasets systematically.

Promoter capture Hi-C (pCHi-C) is a recent technique that utilises sequence capture to enrich interactions between gene promoters and long-range cis-regulatory elements, such as enhancers, thereby providing robust evidence for interactions between distal regulatory elements and target genes.[8-11] By combing pCHi-C with GWAS data, it becomes possible to elucidate the functions of non-coding variants and establish potential links between SNPs and disease-related genes. For instance, Montefiori et al[12] employed pCHi-C maps of cardiomyocytes to associate 1,999 SNPs related to cardiovascular diseases with 347 target genes. Li et al[13] integrated pCHi-C data with GWAS data on systemic sclerosis and successfully identified four susceptibility loci interacting with differentially methylated CpG sites.

Chromatin immunoprecipitation sequencing (ChIP-seq) is another widely used approach to investigate genome-wide protein-DNA interactions and predict gene regulatory mechanisms.[14] Combining ChIP-seq with pCHi-C data can help shed light on the underlying mechanisms by identifying which regulatory elements mediate these interactions.[9] Deoxyribonuclease I (DNase I)-hypersensitive sites (DNase) serve as markers of chromatin accessibility and play a critical role in discovering various types of cis-regulatory elements, including enhancers, promoters and silencers.[15] Histone modifications, such as H3K4me1 and H3K27ac, are associated with distal enhancers, with H3K27ac specifically indicating active enhancers that also possess H3K4me1 modifications.[16] Additionally, H3K4me3 is typically found at gene promoter regions with active transcriptional activity.[17]

In the present study, to gain comprehensive insights into the mechanisms underlying NSCL/P, we integrated GWAS data with pCHi-C data obtained from human embryonic stem cells (hESCs) and incorporated chromatin modifications of craniofacial tissues (CTs) at different developmental stages to identify potential risk SNPs. By utilising pCHi-C and expression quantitative trait loci (eQTL) analysis, we prioritised candidate target genes associated with NSCL/P development.

## Materials and methods

### *Study populations*

In the present study, we carried out a two-stage GWAS of NSCL/P. The discovery stage included 504 NSCL/P cases and 455 controls in Stage I, and the replication stage was performed in an additional 565 unrelated NSCL/P cases and 1,269 controls in Stage II. All participants were recruited from the Chinese population and were examined by two experienced oral surgeons. Patients with an underlying syndrome were excluded. Written consent was collected from all participants. Basic information on the age and sex of the subjects was collected, and peripheral blood samples were obtained in the clinical laboratory. The study was approved by the institutional ethics committee of Nanjing Medical University (NJMUERC [2008] No.20).

### *DNA extraction*

From each subject, 2 ml whole blood was collected using an anticoagulant tube. The blood samples were centrifuged at 3,000 rpm at 4°C for 5 minutes to separate them into plasma and cell parts. TIANamp Genomic DNA Kit (Tiangen Biotech, Beijing, China) was used to extract partial genomic DNA of cells following the manufacturer's instructions.

### *Genotyping, quality control and imputation of GWAS data*

Genotyping was used by Affymetrix Axiom Genome-Wide CHB1 and CHB2 Array Plates in Stage I and by Illumina Infinium Asian Screening Array (ASA) v1.0 in Stage II. SNPs and samples that did not meet the criteria were excluded before imputation as previous described.[18] SNPs with a call rate < 95%, minor allele frequency (MAF) < 0.05 or $P \leq$ 1E-04 for Hardy-Weinberg equilibrium (HWE) were excluded. Samples with a call rate < 95%, sex discrepancy and extreme heterozygosity

(> 6 standard deviation from the mean) were excluded. The 1,000 Genomes Project (https://www.1000genomes.org, phase1, release3) was used as a reference and imputation was performed based on SHAPEIT (SHAPEIT_v2) and IMPUTE2 (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html, imputation step). SNPs with a call rate < 0.95, imputation quality info ≤ 0.8, MAF < 0.05 and HWE in controls ≤ 1E-04 were excluded from the analysis.

### *Selection of SNPs with cis-regulation*

We selected SNPs with $P < 0.01$ in both GWAS studies, along with association direction consistency in both GWAS datasets for subsequent analysis. SNPs in cis-regulatory units of genes in hESC sequenced by pCHi-C maps in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo, acc=GSE86821) and without linkage disequilibrium (LD, r2 < 0.4) were identified for study. To further identify regulatory mechanisms of SNPs for human craniofacial development, we utilised ChIP-seq of four histone modifications across multiple stages (CS13, CS14, CS15, CS17) in CTs in the GEO database (acc = GSE97752) and eliminated SNPs in quiescent states.[19]

### *In silico functional annotation of SNPs and genes*

After identifying these susceptibility loci, several annotation approaches were used, including HaploReg v4 (http://compbio.mit.edu/HaploReg) and RegulomeDB (https://regulome.stanford.edu/regulome-search) for functional annotation. The authors then integrated pCHi-C data and eQTL analysis to highlight potential candidate genes, and based on 3DSNP (http://biotech.bmi.ac.cn/3dsnp/), a database for linking SNPs to their interacting genes in three dimensions and the 3D Genome Browser (http://3dgenome.org), a database for visualising chromatin interaction data by Hi-C to verify functional interactions between SNPs and these candidate genes.[20,21] To observe their expression in orofacial development, the GEO database (acc = GSE67985) and FaceBase (https://www.facebase.org/) were analysed among E10.5 (acc = FB00000662.01), E11.5 (acc = FB00000663.01), E12.5 (acc = FB00000664.01), E13.5 (acc = FB00000665.01) and E14.5 (acc = FB00000666.01). The phenotypes of the candidate genes in humans were assessed using the Database of Genomic Variation and Phenotype in Humans using Ensemble Resources (DECIPHER, https://decipher.sanger.ac.uk/).[22]

### *Statistical analysis*

GWASs were analysed using SNPTEST (version 2.5.6), PLINK (version 1.09) and R software (version 3.6.2).[23,24] In two-stage GWAS, the association between each SNP and NSCL/P risk was evaluated using odds ratios (ORs) and 95% confidence intervals (CIs) in logistic regression analysis under an additive model. The cumulative effects of five risk loci on NSCL/P susceptibility were assessed by calculating the number of risk alleles per subject and categorising pooled variables in unconditional logistic regression analysis (0 to 4 as a reference, 5 to 6, 7, 8 and 9 to 10).[25] Epigenetic annotation for SNPs was completed in WashU Epigenome Browser.[25,26] Statistical figures were performed using GraphPad Prism5 (GraphPad Software, San Diego, CA, USA).

## Results

### *Genome-wide association of SNPs with NSCL/P risk*

The workflow for identifying risk SNPs associated with NSCL/P is shown in Fig 1. To find susceptibility variants for NSCL/P in the Chinese population, we conducted a two-stage analysis, including 1,069 cases and 1,724 controls (Supplementary Table 1, provided on request). A total of 6,206,820 SNPs were identified in GWAS Stage I and 3,806,365 were filtered out in GWAS Stage II after quality control. A Manhattan plot displaying the results of the genome-wide association analysis is presented in Fig 2. Ultimately, 411 SNPs with $P < 0.01$ and consistent association direction in both GWAS datasets were extracted.

### *SNPs with active cis regulation associated with NSCL/P*

After combing pCHi-C maps in hESC (Supplementary Table 2, provided on request) with two-stage GWAS, 13 SNPs located in these units were identified after excluding strong linkage disequilibrium. Five of these were identified as having different histone modifications and were associated with active chromatin states in human craniofacial tissues at various developmental stages (Table 1 and Supplementary Table 2). Rs7218002 (OR 1.50, 95% CI 1.35–1.64, $P_{meta}$ = 8.14E-08) was in a bivalent promoter or repressed polycomb state. Rs835367 (OR 0.78, 95% CI 0.71–0.85, $P_{meta}$ = 3.48E-05) and rs961470 (OR 0.73, 95% CI 0.58-0.89, $P_{meta}$ = 1.38E-04) were in active TSS state, whereas rs77022994 (OR 0.55, 95% CI 0.35–0.85, $P_{meta}$ = 1.05E-04) was in transcription state

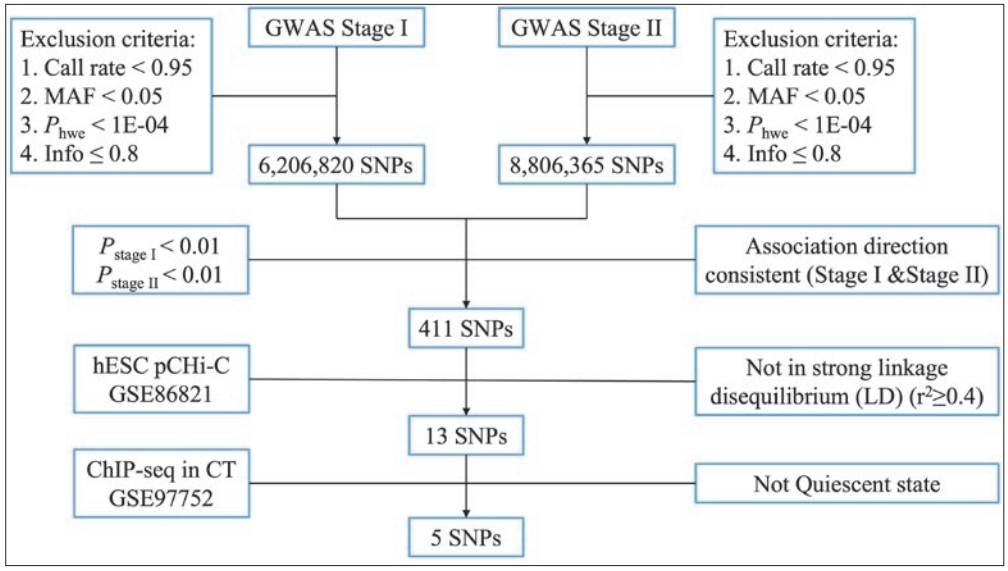


**Fig 1** Workflow for identifying the susceptible SNPs with active cis-regulatory elements involved in non-syndromic orofacial cleft. ChIP-seq, chromatin immunoprecipitation sequence; CTs, craniofacial tissues; GWAS, genome-wide association studies; hESC, human embryonic stem cell; HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency; pCHi-C, promoter capture Hi-C.

**Table 1** Five SNPs with active cis-regulatory elements involved in non-syndromic orofacial cleft.

| SNP | CHR | BP (hg19) | Alleles | MAF[a] | Stage I | | | Stage II | | | Meta analysis | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | MAF (control/case) | OR (95% CI) | *P*[b] | MAF (control/case) | OR (95% CI) | *P*[b] | OR (95% CI) | *P* |
| rs7218002 | 17 | 8923718 | T/A | 0.19 | 0.18/0.26 | 1.62 (1.38–1.86) | 7.46E-05 | 0.19/0.24 | 1.42 (1.24–1.61) | 2.05E-04 | 1.50 (1.35–1.64) | 8.14E-08 |
| rs835367 | 1 | 59762468 | A/G | 0.44 | 0.48/0.40 | 0.71 (0.62–0.83) | 8.44E-04 | 0.42/0.38 | 0.81 (0.72–0.93) | 7.74E-03 | 0.78 (0.71–0.85) | 3.48E-05 |
| rs77022994 | 11 | 125255402 | T/C | 0.03 | 0.06/0.03 | 0.51 (0.04–0.98) | 5.35E-03 | 0.05/0.03 | 0.59 (0.20–0.97) | 6.26E-03 | 0.55 (0.35–0.85) | 1.05E-04 |
| rs961470 | 1 | 119530133 | G/A | 0.18 | 0.22/0.16 | 0.71 (0.47–0.96) | 6.68E-03 | 0.17/0.13 | 0.75 (0.54–0.96) | 7.01E-03 | 0.73 (0.58–0.89) | 1.38E-04 |
| rs17314727 | 5 | 93228046 | C/T | 0.13 | 0.18/0.15 | 0.69 (0.43–0.95) | 5.90E-03 | 0.15/0.12 | 0.75 (0.54–0.97) | 9.91E-03 | 0.73 (0.56–0.89) | 1.85E-04 |

Alleles, minor allele/major allele; BP, base-pair position; CHR, chromosome.

[a]Minor allele frequency in 1000 Genomes Project (CHB+JPT);

[b]*P* value was adjusted by logistic regression in SNPTEST including sex and PCA as covariates.

and rs17314727 (OR 0.73, 95% CI 0.56–0.89, $P_{meta}$ = 1.85E-04) was in a weak enhancer state. We demonstrated that rs835367 was in DNase chromatin and had abundant H3K4me3 and H3K27ac. Rs961470 was also present in DNase chromatin and was marked by H3K4me1, H3K4me3 and H3K27ac. Additionally, rs17314727 was surrounded by H3K4me3 and H3K27ac marks (Fig 3a).

### *Cumulative effects of risk loci on NSCL/P subgroups*

The cumulative effects of five significant SNPs according to the risk alleles (rs7218002 T, rs835367 G, rs77022994 C, rs961470 A and rs17314727 T) are assessed in Table 2. Compared with individuals with 0 to 4 risk alleles, those carrying 7, 8 or 9 to 10 risk alleles had ORs of 1.53 (95% CI 0.79–2.95), 2.12 (95% CI 1.07–4.20) or 4.28 (95% CI 1.89–9.74) in Stage I. We found individuals with mul-

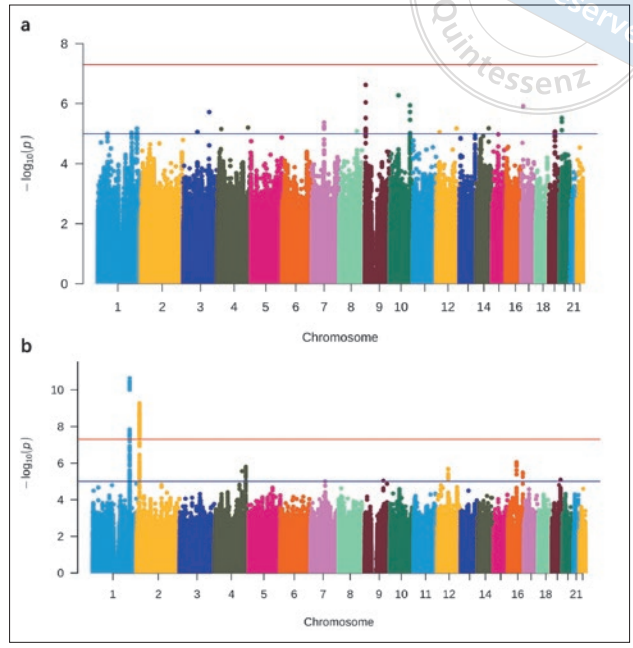


**Fig 2** Manhattan plot of genome-wide association analysis illustrating the level of statistical significance (y-axis), as measured by the negative log of the corresponding *P* value in Stage I **(a)** and Stage II **(b)**, for each single nucleotide polymorphism (SNP). Each typed SNP is indicated by a grey or blue dot. SNPs are arranged by chromosomal location (x-axis). The red line represents the genome-wide level (*P* = 5E-08) and the blue line represents the suggestive significant level (*P* = 1E-05).

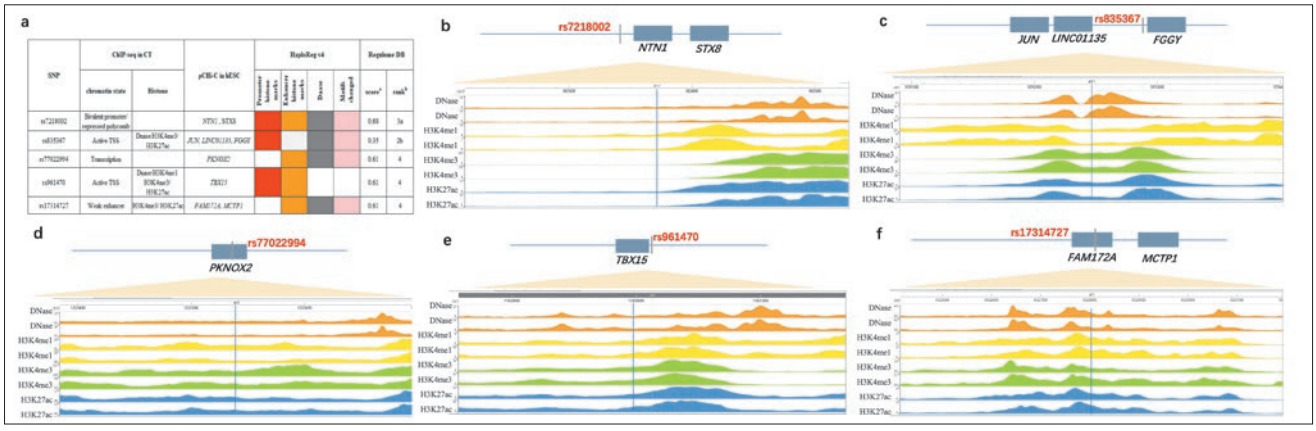


**Fig 3** Functional annotation of five risk SNPs with active cis-regulatory elements involved in non-syndromic cleft lip with or without cleft palate. Annotation of five SNPs with active cis-regulatory elements involved in non-syndromic orofacial cleft according to HaploReg v4 and Regulome DB **(a)**. Annotation of five SNPs with active cis-regulatory elements involved in non-syndromic orofacial cleft according to ChIP-seq in CT and pCHi-C in hESC **(b to f)**. The data were obtained from GEO Project (GSE97752), including active histone modification (H3k4me1, H3K4me3, H3k27ac) as well as DNase. Orange represented DNase chromatin accessibility, yellow indicated histone H3k4me1, green represented histone H3K3me3 and blue indicated histone H3K27ac. The nearest gene and interaction genes of five SNPs were shown in grey rectangles from pCHi-C datasets (GSE86821). ChIP-seq, chromatin immunoprecipitation sequence; CTs, craniofacial tissues; hESC, human embryonic stem cell; pCHi-C, promoter capture Hi-C. [a]The Regulome DB probability score ranges from 0 to 1 and 1 is most likely to be a regulatory variant. [b]2b means TF binding + any motif + DNase Footprint + DNase peak, 3a means TF binding + any motif + DNase peak, and 4 means TF binding + DNase peak.

tiple risk alleles had a higher risk of NSCL/P, and similar results were also observed in Stage II. We stratified all NSCL/P cases into cleft lip only (CLO) and cleft lip and cleft palate (CLP) subgroups and observed that the risk increased in a dose-dependent manner with the number of risk variant alleles.

### *Annotation and functional assessment of genetic variants*

HaploReg v4.1 and RegulomeDB were used to determine the function of the five significant SNPs. With prediction by HaploReg V4.1, functional annotation of rs7218002

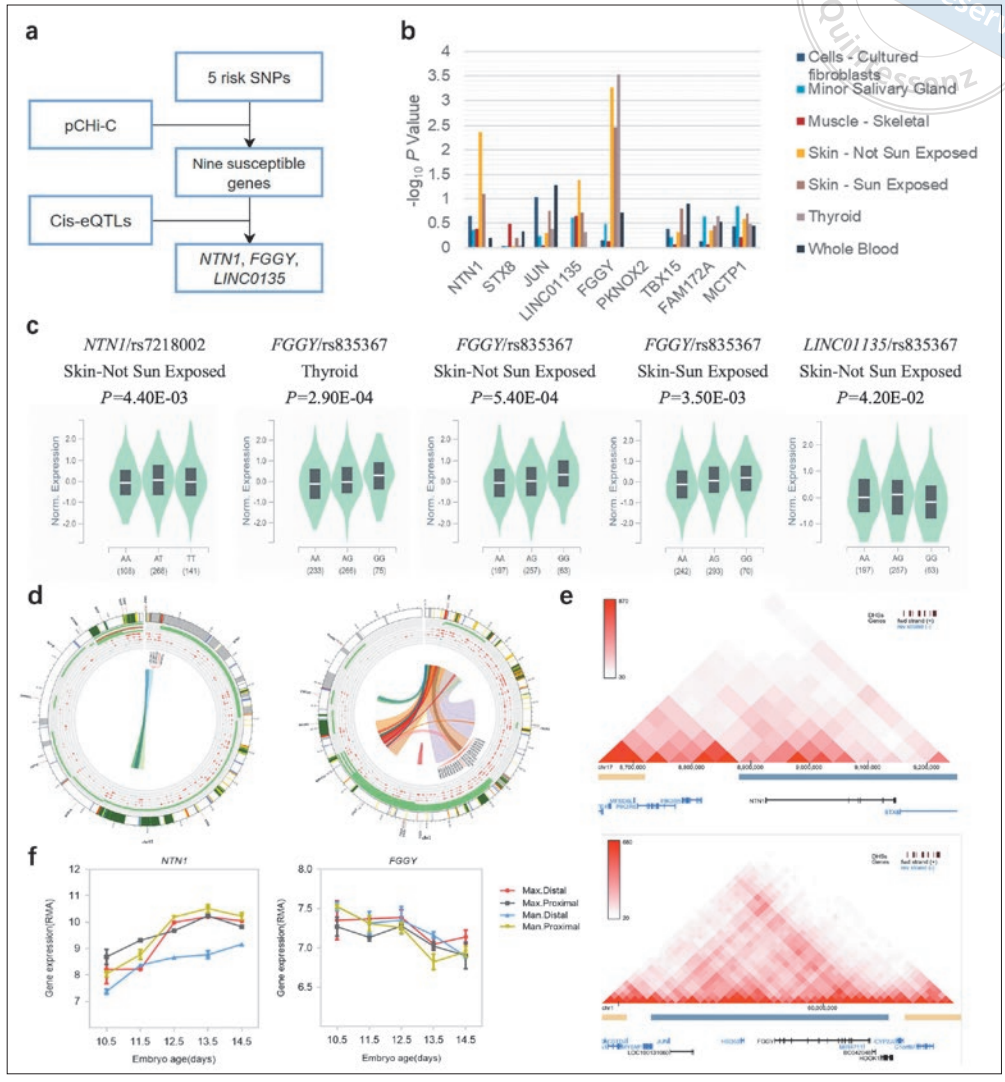


**Fig 4** Identification of susceptible genes associated with non-syndromic orofacial cleft. A simplified diagram of the gene screening approach **(a)**. eQTL analysis of five SNPs and correlated genes in different tissues **(b and c)**. 3D chromatin looping of rs7218002 (left) and rs835367 (right) to their interacting genes **(d)**. A region of the Hi-C interaction map in hESC on rs7218002 (left) and rs835367 (right). The alternating yellow and blue bars were predicted topologically associating domains (TADs) **(e)**. Median expression levels of NTN1 and FGGY in proximal and distal locations of maxilla and mandible during the 10.5th day to 14.5th day of mouse embryonic stage based on GEO datasets (GSE67985) **(f)**. Man. Dis, mandibular distal location; Man.Pro, mandibular proximal location; Max.Dis, maxillary distal location; Max.Pro, maxillary proximal location.

included promoter histone marks, enhancer histone marks, DNase and motifs changed. Rs835367 might lead to promoter histone marks, DNase and motifs changed. Functional annotation of the other three SNPs is shown in Fig 3a. rs7218002 was in moderate linkage disequilibrium with rs4791774 ($r^2$ = 0.51) and rs2944377($r^2$ = 0.55), which were previously reported as having a risk of NSCL/P.[27]

Based on RegulomeDB scores, rs7218002 received a score of 3a, indicating a lesser likelihood of affecting transcription factor (TF) binding. On the other hand, rs835367 obtained a higher score of 2b, suggesting a greater potential for affecting TF binding. Rs7702294, rs961470 and rs17314724 received a RegulomeDB score of 4, indicating minimal binding ability for these SNPs.[28]

### *Bioinformatic predictions of risk gene*

A simplified diagram of the approach to identify the susceptible genes associated with NSCL/P is shown in Fig 4a. With prediction of pCHi-C in hESC, we examined the interaction gene of five SNPs and found that five SNPs could link to nine genes (rs7218002: *NTN1*, *STX8*; rs835367: *FGGY*, *JUN*, *LINC01135*; rs77022994: *PKNOX2*; rs961470: *TBX15*; rs17314727: *FAM172A*, *MCTP1*) (Fig 3).

Further eQTL analysis was conducted using the GTEx database. rs7218002 was found to exhibit a significant association with expression of *NTN1* (*P* = 4.40E-03) in skin not exposed to the sun. *FGGY* was associated with rs835367 in the thyroid (*P* = 2.90E-04), skin not exposed to the sun (*P* = 5.40E-04) and skin exposed to the sun (*P* = 3.50E-03). *LINC01135* showed an association with

**Table 2** Cumulative effect of five risk SNP in non-syndromic orofacial cleft and subgroups.

| GWAS | No. of risk allele | Control | Total case | OR (95% CI) | CLO | OR (95% CI) | CLP | OR (95% CI) |
|---|---|---|---|---|---|---|---|---|
| Stage I | 0–4 | 23 | 18 | Reference | 9 | Reference | 9 | Reference |
| | 5–6 | 208 | 149 | 0.61 (0.32–1.18) | 66 | 0.81 (0.36–1.84) | 83 | 1.02 (0.45–2.30) |
| | 7 | 137 | 164 | 1.53 (0.79–2.95) | 65 | 1.21 (0.53–2.77) | 99 | 1.85 (0.82–4.16) |
| | 8 | 70 | 116 | 2.12 (1.07–4.20) | 45 | 1.64 (0.70–3.87) | 71 | 2.59 (1.12–5.99) |
| | 9–10 | 17 | 57 | 4.28 (1.89–9.74) | 20 | 3.01 (1.10–8.22) | 37 | 5.56 (2.13–14.54) |
| Stage II | 0–4 | 55 | 10 | Reference | 2 | Reference | 5 | Reference |
| | 5–6 | 454 | 202 | 2.45 (1.22–4.90) | 62 | 3.76 (0.89–15.78) | 106 | 6.42 (1.54–26.74) |
| | 7 | 404 | 256 | 3.49 (1.75–6.96) | 88 | 5.99 (1.43–25.02) | 121 | 8.23 (1.98–34.26) |
| | 8 | 278 | 203 | 4.02 (2.00–8.07) | 65 | 6.43 (1.53–27.04) | 103 | 10.19 (2.44–42.53) |
| | 9–10 | 78 | 70 | 6.13 (2.90–12.94) | 25 | 8.81 (2.00–38.76) | 35 | 12.34 (2.85–53.46) |

CLO, cleft lip only; CLP, cleft lip and palate.

rs835367 in skin not exposed to the sun ($P$ = 4.20E-02) (Fig 4b and c).

The 3D chromatin looping data demonstrated that rs7218002 interacted with *NTN1* and rs835367 could interact with *FGGY* and *LINC01135* (Fig 4d). According to the Hi-C map in hESC by 3D Genome Browser, rs7218002 and *NTN1* were located within a topologically associated domain (TAD) structure and so too were rs835367, *FGGY* and *LINC01135* (Fig 4e).

### *Gene expression in mouse craniofacial structures and phenotype in humans*

Both *NTN1* and *FGGY* were expressed in the maxilla and mandible over five periods in mouse orofacial development (Fig 4f). A lack of considerable homology was detected for LncRNA *LINC01135* between humans and mice. According to DECIPHER, we found that individuals with an *NTN1* deficiency might present with multiple phenotypes like cleft palate and cleft lip. Individuals with an *FGGY* deficiency might show traits like thick lower lip vermilion, thick upper lip vermilion, abnormal facial shape and bifid uvula, and those with an *LINC01135* deficiency might present abnormalities in facial features, thick lower lip vermilion and thick upper lip vermilion.

## Discussion

The present study investigated systematic insights into mechanisms of NSCL/P by using previously published GWAS data, pCHi-C data of hESC and ChIP-seq of craniofacial tissues. The function of SNPs was assessed using HaploReg v4.1 and RegulomeDB. Candidate genes were identified based on pCHi-C data and eQTL analysis, supported by 3DSNP and 3D Genome Browser. The functional roles of these genes during embryogenesis were evaluated using FaceBase in mouse orofacial development and DECIPHER in humans.

The study identified five SNPs that may confer risk for NSCL/P, with rs7218002 exhibiting moderate linkage disequilibrium with rs4791774 and rs2944377. These two linked SNPs were reported in previous GWAS studies of NSCL/P according to the GWAS Catalogue.[29] By integrating various approaches, the authors found that rs7218002 could influence the gene expression of *NTN1*. Abnormal expression of *NTN1* has been implicated in the development of non-syndromic cleft lip with or without cleft palate in previous studies.[27] The present group generated *NTN1* knockout zebrafish models by CRISPR/Cas9 and observed relatively wider maxillomandibular fissures.[30] Complementary transcriptomic profiling was performed in embryonic chicks and demonstrated that *NTN1* was precisely expressed in the chick palate fissure margin during palate fusion.[31]

We identified rs835367, which lied in the promoter of *FGGY*, was associated with NSCL/P by two GWAS datasets. Our study also provided evidence that *FGGY* and *LINC01135* could be regulated by rs835367. The expression of *FGGY* decreased greatly from 10.5 to 14.5 days based on RNA-seq data of embryonic mouse tissues. Notably, previous meQTL analysis in 4,170 whole blood tissues suggested that rs835367 could regulate cg18108087, which is located near the promoter region of *FGGY* ($P$ = 7.31E-24).[32] Individuals with an *FGGY* deficiency were reported to exhibit an abnormal facial shape and bifid uvula according to DECIPHER. A bifid uvula is an anatomical variation that can be predictive of sub-mucous cleft palate, indicating that *FGGY* may contribute to the risk of NSCL/P to a certain degree.[33] Individuals with a *LINC01135* deficiency may exhibit abnormalities in facial features, thick lower lip vermilion and thick upper lip vermilion, suggesting that *LINC01135* may be a shared susceptibility factor for

NSCL/P. Taken together, these findings indicate that *FGGY* and *LINC01135* could be novel genes associated with NSCL/P.

## Conclusion

In the present study, a multi-omics approach was employed to provide novel insights into the aetiology of NSCL/P and identify five risk SNPs and three susceptibility genes for NSCL/P; however, further experimental studies are required to validate these results.

## Conflicts of interest

The authors declare no conflicts of interest related to this study.

## Author contribution

Drs Shu LOU and Jin YANG analysed the data and drafted the manuscript; Drs Gui Rong ZHU, Dan Dan LI and Lan MA performed the statistical analysis and revised the manuscript; Profs Lin WANG and Yong Chu PAN designed and directed the study and critically revised the manuscript.

(Received Jul 06, 2023; accepted Nov 02, 2023)

## References


1. Mbuyi-Musanzayi S, Kayembe TJ, Kashal MK, et al. Non-syndromic cleft lip and/or cleft palate: Epidemiology and risk factors in Lubumbashi (DR Congo), a case-control study. J Craniomaxillofac Surg 2018;46:1051–1058.
2. Hosseini HR, Kaklamanos EG, Athanasiou AE. Treatment outcomes of pre-surgical infant orthopedics in patients with non-syndromic cleft lip and/or palate: A systematic review and meta-analysis of randomized controlled trials. PLoS One 2017;12:e0181768.
3. Wu MC, Kraft P, Epstein MP, et al. Powerful SNP-set analysis for case-control genome-wide association studies. Am J Hum Genet 2010;86:929–942.
4. Horwitz T, Lam K, Chen Y, Xia Y, Liu C. A decade in psychiatric GWAS research. Mol Psychiatry 2019;24:378–389.
5. Pathak GA, Zhou Z, Silzer TK, Barber RC, Phillips NR; Alzheimer's Disease Neuroimaging Initiative B, Prostate Cancer Cohort C, Alzheimer's Disease Genetics C. Two-stage Bayesian GWAS of 9576 individuals identifies SNP regions that are targeted by miRNAs inversely expressed in Alzheimer's and cancer. Alzheimers Dement 2020;16:162–177.
6. Dogruluk T, Tsang YH, Espitia M, et al. Identification of variant-specific functions of PIK3CA by rapid phenotyping of rare mutations. Cancer Res 2015;75:5341–5354.
7. Raimondi D, Passemiers A, Fariselli P, Moreau Y. Current cancer driver variant predictors learn to recognize driver genes instead of functional variants. BMC Biol 2021;19:3.
8. Choy MK, Javierre BM, Williams SG, et al. Promoter interactome of human embryonic stem cell-derived cardiomyocytes connects GWAS regions to cardiac gene networks. Nat Commun 2018;9:2526.
9. Schoenfelder S, Javierre BM, Furlan-Magaril M, Wingett SW, Fraser P. Promoter capture Hi-C: High-resolution, genome-wide profiling of promoter interactions. J Vis Exp 2018:57320.
10. Zhang N, Mendieta-Esteban J, Magli A, et al. Muscle progenitor specification and myogenic differentiation are associated with changes in chromatin topology. Nat Commun 2020;11:6222.
11. Sun F, Sun T, Kronenberg M, Tan X, Huang C, Carey MF. The Pol II preinitiation complex (PIC) influences Mediator binding but not promoter-enhancer looping. Genes Dev 2021;35:1175–1189.
12. Montefiori LE, Sobreira DR, Sakabe NJ, et al. A promoter interaction map for cardiovascular disease genetics. Elife 2018;7:e35788.
13. Li T, Ortiz-Fernández L, Andrés-León E, et al. Epigenomics and transcriptomics of systemic sclerosis CD4+ T cells reveal long-range dysregulation of key inflammatory pathways mediated by disease-associated susceptibility loci. Genome Med 2020;12:81.
14. Jaini S, Lyubetskaya A, Gomes A, et al. Transcription factor binding site mapping Using ChIP-Seq. Microbiol Spectr 2014;2.
15. Liu Y, Fu L, Kaufmann K, Chen D, Chen M. A practical guide for DNase-seq data analysis: From data management to common applications. Brief Bioinform 2019;20:1865–1877.
16. Local A, Huang H, Albuquerque CP, et al. Identification of H3K4me1-associated proteins at mammalian enhancers. Nat Genet 2018;50:73–82.
17. Zhang B, Zheng H, Huang B, et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature 2016;537:553–557.
18. Sun Y, Huang Y, Yin A, et al. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. Nat Commun 2015;6:6414.
19. Wilderman A, VanOudenhove J, Kron J, Noonan JP, Cotney J. High-resolution epigenomic atlas of human embryonic craniofacial development. Cell Rep 2018;23:1581–1597.
20. Lu Y, Quan C, Chen H, Bo X, Zhang C. 3DSNP: A database for linking human noncoding SNPs to their three-dimensional interacting genes. Nucleic Acids Res 2017;45:D643–D649.
21. Wang Y, Song F, Zhang B, et al. The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol 2018;19:151.
22. Swaminathan GJ, Bragin E, Chatzimichali EA, et al. DECIPHER: Web-based, community resource for clinical interpretation of rare variants in developmental disorders. Hum Mol Genet 2012;21:R37–R44.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
24. Galesloot TE, van Steen K, Kiemeney LA, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. PLoS One 2014;9:e95923.
25. Pan Y, Han Y, Zhang H, et al. Association and cumulative effects of GWAS-identified genetic variants for nonsyndromic orofacial clefts in a Chinese population. Environ Mol Mutagen 2013;54:261–267.
26. Li D, Hsu S, Purushotham D, Sears RL, Wang T. WashU Epigenome Browser update 2019. Nucleic Acids Res 2019;47:W158–W165.

27. Jiang S, Shi JY, Lin YS, et al. NTN1 gene was risk to non-syndromic cleft lip only among Han Chinese population. Oral Dis 2019;25:535–542.
28. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012;22:1790–1797.
29. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019;47:D1005–D1012.
30. Li D, Zhu G, Lou S, et al. The functional variant of NTN1 contributes to the risk of nonsyndromic cleft lip with or without cleft palate. Eur J Hum Genet 2020;28:453–460.
31. Hardy H, Prendergast JG, Patel A, et al. Detailed analysis of chick optic fissure closure reveals Netrin-1 as an essential mediator of epithelial fusion. Elife 2019;8:e43877.
32. Huan T, Joehanes R, Song C, et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. Nat Commun 2019;10:4267.
33. Feka P, Banon J, Leuchter I, La Scala GC. Prevalence of bifid uvula in primary school children. Int J Pediatr Otorhinolaryngol 2019;116:88–91.