Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Epigenomics

Variants in miRNA regulome and their association with the risk of nonsyndromic orofacial clefts

Guirong Zhu[‡,1,2], Chi Zhang[‡,1,2], Yuting Wang[‡,1,2], Yuli Wang[‡,1,2], Dandan Li[1,2], Xin Yu[1,2], Weihao Zhu[1,2], Chengyi Fu[1,2], Shu Lou[1,2], Liwen Fan[1,2], Lan Ma[1], Lin Wang[1,2] & Yongchu Pan*[,1,2]

[1]Institute of Stomatology, Nanjing Medical University, Nanjing, PR China

[2]Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, PR China

*Author for correspondence: Tel.: +86 25 86862025; Fax: +86 25 86862823; panyongchu@njmu.edu.cn

[‡]Authors contributed equally

Aim: To investigate the associations between single nucleotide polymorphisms (SNPs) in miRNA regulome and nonsyndromic orofacial clefts. **Materials & methods:** The associations were evaluated by logistic regression model in stage I (504 cases and 455 controls) and stage II (1500 cases and 1386 controls). Functional experiments including luciferase activity assay, cell apoptosis and proliferation, serum miRNA expression, and mouse embryo RNA sequencing were performed. **Results:** Rs3830766 in the enhancer of hsa-miR-4260 was associated with cleft lip only (CLO) and enhancer activity. Hsa-miR-4260 expression decreased in the serum of CLO. Overexpression of miR-4260 inhibited cell proliferation and promoted cell apoptosis. *UBB* was the target gene of hsa-miR-4260. **Conclusion:** Rs3830766 in the hsa-miR-4260 enhancer that can interact with *UBB* was relevant to CLO.

First draft submitted: 31 March 2020; Accepted for publication: 28 April 2020; Published online: 15 May 2020

Keywords: CLO • miRNA regulome • NSOFC • SNPs

Nonsyndromic orofacial clefts (NSOFC) characterized by failed fusion of normal facial processes [1], have a general incidence of one in 700 births, with varying prevalence depending on geographic origin and ethnicity [2,3]. NSOFC always result in a series of problems on eating, pronunciation, facial beauty and mental health, and cause psychological and economic burden [4]. Quite a few genes/loci including *IRF6*, *MAFB*, *VAX1* and *MSX1* were associated with the development of NSOFC [5]. Meanwhile, environmental exposures, such as maternal cigarette smoking, alcohol consumption and insufficient folic acid supplementation were also identified as risk factors [6].

NSOFC can be classified into three main types: cleft lip only (CLO), cleft lip and palate (CLP) and cleft palate only (CPO) [7,8]. CLO and CLP are traditionally grouped together and termed as cleft lip with or without palate (CL/P) given their similar etiology. However, an increasing number of studies have shown that these three subgroups are caused by different genetic susceptive factors and must be investigated separately [9]. For instance, compared with CLO and CLP, CPO was more susceptible to genetic factors [10]. In addition, CLO and CLP were often caused by common variants, while CPO might be influenced by rare variants [11].

miRNAs are 20–24 nucleotides long, small noncoding RNAs that function in the RNA silencing and post-transcriptional regulation of specific genes, mainly through suppressing the target mRNAs with complementary sequences via base-pairing [12]. MiRNA regulome (including promoters and enhancers), by definition, is composed of regulatory elements that regulate miRNA expression or are regulated by miRNA activity, which is critical to human complex phenotypes [13,14]. Promoters are 100–1000 base pair (bp) long sequences upstream of the genes that can initiate the transcription of a specific gene [15], while enhancers are 50–1500 bp long protein-bound regions that can increase the likelihood of transcription in a particular gene [16]. It is reported that nucleosomes with H3 lysine 4 trimethylation (H3K4me3) modification, together with RNA polymerase II, occupy the promoters of most protein-coding genes in human [17]. H3 lysine 4 monomethylation (H3K4me1) is a hallmark of promoter-distal

cis-regulatory elements (i.e., enhancers) [18,19]. Furthermore, deoxyribonuclease I (DNase I) hypersensitive sites (DHSs) is extensively used to map regulatory DNA regions (enhancers, promoters, silencers, insulators and locus control regions) in diverse organisms [20]. Thus, the availability of the human genome sequence reveals that the H3K4me3 and H3K4me1 histone markers can respectively represent the active promoter and enhancer elements, while DHSs are functionally related with transcriptional activity of promoters and enhancers.

In recent years, the genetic variants in miRNA promoters and enhancers have attracted increasing attention because of their potential associations with various human diseases. For instance, Xie *et al.* found that the functional single nucleotide polymorphism (SNP) rs12740674 in miR-1262 enhancer was significantly associated with the development of lung cancer. Furthermore, the risk allele rs12740674-T could reduce the miR-1262 expression in lung tissue through chromosomal looping and consequently influence the expression of *ULK1* and *RAB3D* [14]. In addition, rs4938723 in the pri-miR-34b/c promoter region was associated with the susceptibility to hepatocellular carcinoma in Chinese population [21]. A variant in the promoter of the miRNA let-7 was also found to be related to human height variation [13]. Collectively, these results demonstrated the significance of genetic variants in the miRNA promoters and enhancers to the occurrence of human diseases. It has been well-known that miRNAs are involved in the pathophysiology of NSOFC [22–24]. For example, our team has confirmed that rs2910164 in miR-146a could interact with the target gene *TRAF6*, leading to the occurrence of NSOFC [23]. However, the association between genetic variants in the miRNA regulome and the risk of developing NSOFC remains unclear.

Recently, the status of six post-translational histone modifications (H3K4me1, H3K4me2, H3K4me3, H3K36me3, H3K27ac and H3K27me3) across multiple Carnegie stages (CS13, CS14, CS15, CS17 and CS20) of early human craniofacial development were successfully identified using chromatin immunoprecipitation sequencing (ChIP-seq) [25]. Consequently, the epigenomic annotations of human embryonic craniofacial tissues offer an opportunity to study the miRNA regulome during craniofacial development.

In this study, the genetic variants among miRNA regulome (promoters and enhancers) were firstly identified from the human miRBase v22, UCSC common SNPs (150) and GEO (GSE97752) databases. Then, we validated the associations between SNPs and NSOFC in a two-stage case–control study including 2004 NSOFC cases and 1891 healthy controls. Moreover, we performed functional studies to elucidate the role of promising miRNAs and target genes in NSOFC.

Materials & methods

Identification of SNPs in miRNA promoters & enhancers

The flow chart of our strategy to detect SNPs in miRNA promoters and enhancers that are associated with NSOFC is shown in Figure 1.

The miRNA-encoding sequences were first retrieved from the human miRBase v22 (www.mirbase.org/) [26] and UCSC common SNPs (150) (http://genome.ucsc.edu/index.html) [27] databases. Next, we annotated the SNPs in miRNA promoters (50-kb upstream of miRNAs) and enhancers (100-kb upstream or downstream of miRNAs) using histone ChIP-seq (H3K4me3 and H3K4me1) peaks in CS17 craniofacial tissue samples of human embryos (equivalent to 6 weeks postconception when the cleft lip happens) from the GEO (GSE97752) database (www.ncbi.nlm.nih.gov/geo/) [28]. Accordingly, all the potentially functional SNPs must be simultaneously located in the DHSs of the same tissue.

Study population

The information on the study populations is shown in Table 1. Stage I included 504 CL/P cases and 455 controls while stage II composed of 1500 NSOFC cases and 1386 controls. The cases were divided into three subgroups according to the clinical manifestation: 494 CLO cases (32.93%), 748 CLP cases (49.87%) and 258 CPO cases (17.20%).

All individuals were confirmed to be Han Chinese and genetically independent without other congenital deformities or syndromic orofacial clefts. Signed informed consents were obtained from participants or their legal guardians. This study is approved by the Institutional Review Board of Nanjing Medical University (NJMUERC [2008] no. 20).

DNA extraction

Peripheral venous blood (~2 ml) was acquired from each individual using an anticoagulant tube. The serum and cellular fractions of blood samples were separated by centrifugation at 3000 rpm/min for 4 min at 4°C.

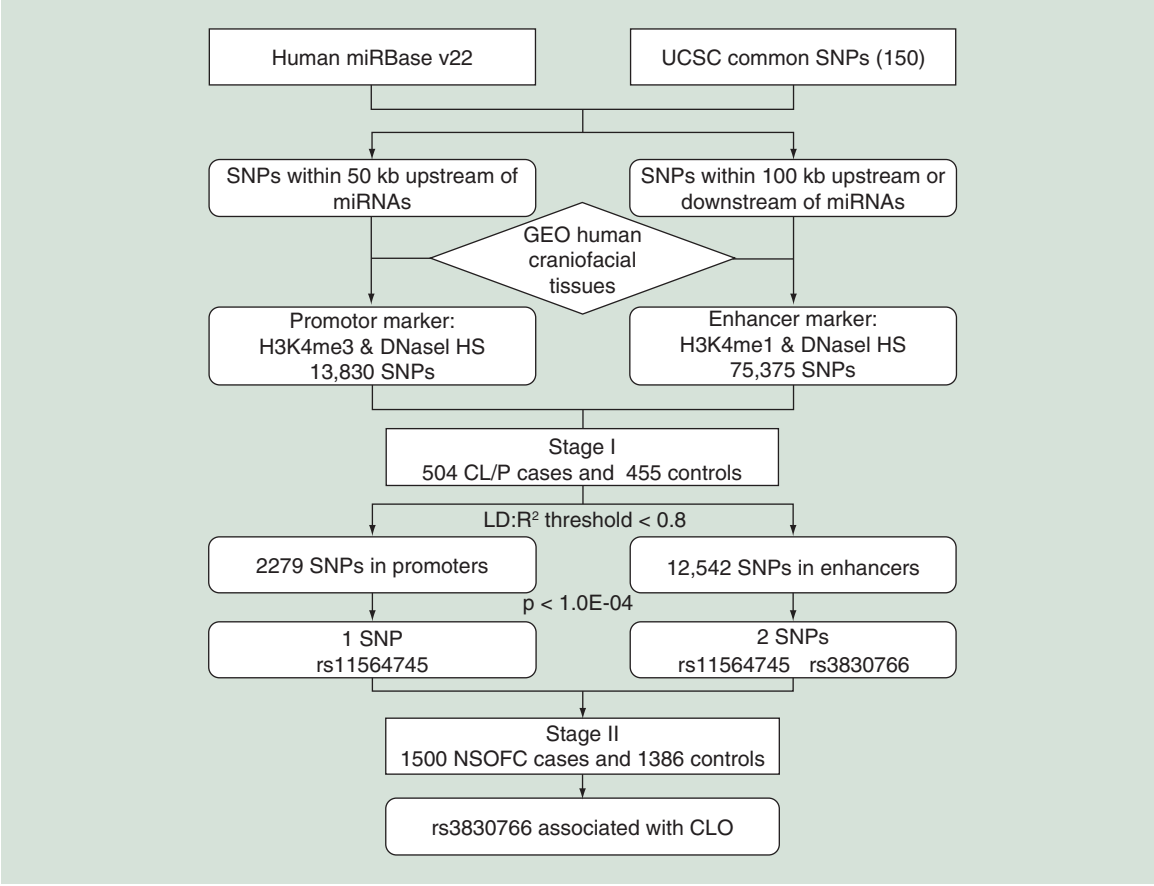

Figure 1. Identification of SNPs in miRNA regulome that are associated with nonsyndromic orofacial clefts. This flow chart shows our approach to identify SNPs that are in miRNA regulome and their associations with the risks of NSOFC. CL/P: Cleft lip with or without palate; DNaseI HS: Deoxyribonuclease I hypersensitive sites; GEO: Gene Expression Omnibus; LD: Linkage disequilibrium; NSOFC: Nonsyndromic orofacial cleft; SNP: Single nucleotide polymorphism; UCSC: University of California, Santa Cruz.

Table 1. Characteristics of the controls and nonsyndromic orofacial cleft cases.

	Stage I		Stage II	
	CL/P cases (%) n = 504	Controls (%) n = 455	NSOFC cases (%) n = 1500	Controls (%) n = 1386
Age (years)				
(Mean ± SD)	1.51 ± 0.51	0.00 ± 0.00	4.56 ± 6.86	10.80 ± 2.20
Gender				
Males	308 (61.11)	236 (51.87)	921 (61.40)	852 (61.47)
Females	196 (38.89)	219 (48.13)	579 (38.60)	534 (38.53)
Subgroups				
CL/P	504 (100.00)		1242 (82.80)	
CLO	–		494 (32.93)	
CLP	–		748 (49.87)	
CPO	–		258 (17.20)	

0.00 ± 0.00 indicates that all samples involved are new-born infants.
CL/P: Cleft lip with or without palate; CLO: Cleft lip only; CLP: Cleft lip and palate; CPO: Cleft palate only; NSOFC: Nonsyndromic orofacial cleft.

10.2217/epi-2020-0124

Genomic DNA was extracted using the conventional phenol-chloroform method or TIANamp Genomic DNA Kit (TIANGEN, Beijing, China) and Qiagen Blood Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.

SNPs genotyping

Stage I

Genotyping was conducted via Affymetrix Axiom Genome-Wide CHB1 and CHB2 array. To screen more variants across the genome, imputation of genotypes was performed. Briefly, the genotype data were phased using SHAPEIT v2 [29]. Imputation of ungenotyped SNPs was performed using IMPUTE2 [30] based on variant positions from the 1000 Genomes Project (the Phase I integrated variant set release, v3), which included 286 Asians. First, a basic quality control was performed by removing the SNPs that did not map to autosomal chromosomes, had a call rate <95%, or had a genotype distribution in the controls that deviated from Hardy–Weinberg equilibrium (HWE; $p < 1.0 \times 10^{-5}$). In addition, we excluded the samples with overall successful genotyping call rates <95%, sex discrepancies between the records and the genetically inferred data, or outlying autosomal heterozygosity rates (>6 SD from the mean). Second, we obtained the imputed SNPs with quality info ≥0.8 and MAF ≥0.05. Finally, 2279 SNPs in the promoter and 12,542 SNPs in the enhancer elements were remained. We extracted the two most promising SNPs with the lowest p-value to validate in stage II.

Stage II

The SNPs were genotyped using TaqMan® assays with 5 μl final reaction volumes and ABI Prism® 7900HT Real-Time PCR System (Applied Biosystems, CA, USA). The information on the designed TaqMan® primers and probes are presented in Supplementary Table 1. The genotyping results were analyzed using SDS v.2.4.1 software (Applied Biosystems) and independently reviewed (blind) by two researchers. Random samples (5%) were tested for repeat analysis to ensure 100% consistency.

Functional annotation

We annotated the SNPs using the HaploReg v4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) [31] and RegulomeDB (www.regulomedb.org/) [32] databases. HaploReg v4.1 contained the information on promoter and enhancer histone marks and expression quantitative trait loci (eQTL) for all SNPs in linkage disequilibrium (LD; $R^2 \geq 0.8$). RegulomeDB was used to predict regulatory DNA elements that had been biochemically characterized to regulate transcription. The UCSC Genome Browser (http://genome.ucsc.edu/) was then used to visualize and compare the ChIP-seq tracks in GSE97752 [25,27]. We also used 3D chromatin looping data (http://cbportal.org/3dsnp/) to link the promising SNPs with their interacting genes in the blood and in other tissues [33].

MiRNA target gene prediction

We predicted the miRNA target genes using Targetscan, miRDB and miRTarBase databases (www.targetscan.org/vert_72/, http://mirdb.org/, http://mirtarbase.mbc.nctu.edu.tw/php/index.php) [34–36]. The target genes were further investigated by detecting their expression in the lip and palate of embryonic mouse at E10.5-E15.5 based on RNA sequencing (RNA-seq) count data. In addition, we checked the expression patterns of targeted genes in dental pulp stem cells of seven CL/P and six control samples from the GEO (GSE42589) database.

Cell culture & clone construction

Human embryonic palatal mesenchyme (HEPM) cells were cultured in Eagle's minimum essential medium supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin solution in 5% CO_2 at 37°C, while human oral epithelial cells (HOECs) were maintained in Dulbecco's modified eagle medium/high glucose supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin solution in 5% CO_2 at 37°C.

Sequences containing the reference or alternative allele of the variants upstream of the minimal promoter of firefly luciferase vector, pGL3 (Promega, WI, USA), were cloned using *KpnI* and *SacI* restriction enzymes to test the enhancer activity. All miRNA mimics were synthesized by GenePharma (Shanghai, China) and 3′-untranslated region (3′-UTR) luciferase reporter plasmids were purchased from Promega.

Transfection & luciferase reporter assays

HEPM cells and HOECs (1×10^6) were seeded into 12-well culture plates and co-transfected with 400 ng pGL3, which contains the reference or alternative allele in either orientation or an empty vector, and 50 ng pRL-SV40 vector (Promega) using Lipofectamine 2000 (Invitrogen, CA, USA). The two cell lines were also transfected with the target gene 3′-UTR luciferase reporter plasmids (wild-type or mutation-type) and miRNA mimics using Lipofectamine 2000 (Invitrogen), according to the manufacturer's protocol. The vectors without inserts and random miRNA sequence mimics (NC mimics; GenePharma) served as the negative control. All transfection experiments were performed in triplicate. The cells were obtained 48 h after transfection and assayed for the firefly luciferase to renilla ratio using the Dual-Luciferase® Reporter System (Promega).

Cell proliferation & apoptosis assays

The human miRNA mimics or NC mimics were transfected into HEPM cells and HOECs in the 96-well plates using Lipofectamine 2000 (Invitrogen), according to the manufacturer's instructions. Each well containing 100 µl culture medium was added with 10 µl CCK-8 (Dojindo Laboratories, Kumamoto, Japan) at different time periods (24, 36, 48 and 60 h) after transfection. The cells were further incubated at 37°C in 5% CO_2 for 2 h. The absorbance values were subsequently measured at 450 nm using SpectraMax 190 spectrophotometer (Molecular Devices, CA, USA). Each proliferation assay was performed in triplicate.

To analyze the effect of the miRNAs on cell apoptosis, HEPM cells and HOECs in six-well plates were transfected with miRNA mimics or NC mimics. After 48 h, the cells were collected and stained with Annexin V-FITC/Propidium Iodide Kit (KeyGEN Biotech, Nanjing, China). The cells were incubated in the dark at 20–25°C for 10 min. The apoptotic cells were analyzed using flow cytometry and FlowJo™ v10 software (FlowJo LLC, OR, USA). Each experiment had three technical replicates.

RNA extraction & quantitative real-time PCR in the serum of human subjects

To assess the miRNA expression levels in the CLO cases and controls, total RNA was extracted from the serum (500 µl) of five CLO cases (two males and three females; mean age: 13.6 ± 4.8) and five controls (two males and three females; mean age: 8.2 ± 12.4) using TRIzol reagent (Invitrogen). The bulge-loop reverse transcription (RT) and quantitative real-time PCR primers were designed and synthesized by RiboBio (Guangzhou, China). The total RNA was reverse-transcribed for cDNA synthesis according to the manufacturer's protocol (RiboBio, Guangzhou, China). We then determined the miRNA expression using SYBR Mastermix on ABI Prism® 7900HT Real-Time PCR System (Applied Biosystems). The miRNA expression levels were normalized against the endogenous *U6* RNA control and analyzed using the $2^{-\Delta\Delta Ct}$ method [37].

Mouse embryo collection & RNA sequencing (RNA-seq)

We purchased the C57BL/6 mice from the Animal Center of Yangzhou University. Two female mice and one male mouse were placed in the same cage at 8 PM and separated at 8 AM the next day. The embryos were considered as E0.5 on the morning of vaginal plug discovery. The lip and palate tissues of E11.5, E12.5, E13.5, E14.5 and E15.5 embryos were subsequently collected.

Total RNA was extracted from these mice embryonic tissues using TRIzol reagent (Invitrogen). The RNA quantity and quality were determined using NanoDrop™ spectrophotometer (Thermo Fisher Scientific, MA, USA) and 1% agarose gel electrophoresis. RNA-seq was subsequently performed using HiSeq 3000 platform (Illumina, CA, USA) at 10-M and 6-G depth, respectively. The raw reads were aligned to the mouse genome (MM10) using gSNAP [38] and the expression values (FPKM) were derived using Cufflinks [39].

Statistical analysis

All statistical analyses were performed using PLINK v1.07 [40] and GraphPad Prism® 7 (GraphPad Software, CA, USA). The controls were tested for HWE using Fisher's exact test. Logistic regression analysis was performed to compare the allele and genotype frequencies between the cases and controls. The odds ratios (ORs) and 95% CIs of each genotype were calculated to estimate its effect on the occurrence of NSOFC and its subgroups. p-value less than 0.05 was considered as nominal, whereas the significant level after Bonferroni-correction was 0.00833.

Table 2. Associations between two SNPs and nonsyndromic orofacial cleft under different genetic models.

SNP	Controls n = 1358	NSOFC cases n = 1435	Models		p-value†	OR (95% CI)‡
rs11564745 (C>T)						
Genotype	594/606/158	590/674/171	Genotype	CC/CT/TT	0.365	–
CC	594 (43.74)	590 (41.11)	Heterozygote	CT/CC	0.161	1.12 (0.96–1.31)
CT	606 (44.62)	674 (46.97)	Homozygote	TT/CC	0.491	1.09 (0.85–1.39)
TT	158 (11.64)	171 (11.92)	Dominant	TT+CT/CC	0.161	1.11 (0.96–1.29)
C	1794 (66.05)	1854 (64.60)	Recessive	TT/CC+CT	0.818	1.03 (0.82–1.29)
T	922 (33.95)	1016 (35.40)	Allele	C/T	0.254	1.07 (0.96–1.19)
rs3830766 (D>G)						
Genotype	450/663/245	455/729/251	Genotype	DD/DG/GG	0.575	–
DD	450 (33.14)	455 (31.71)	Heterozygote	DG/DD	0.326	1.09 (0.92–1.29)
DG	663 (48.82)	729 (50.80)	Homozygote	GG/DD	0.906	1.01 (0.81–1.26)
GG	245 (18.04)	251 (17.49)	Dominant	GG+DG/DD	0.420	1.07 (0.91–1.25)
D	1563 (57.55)	1639 (57.11)	Recessive	GG/DD+DG	0.704	0.96 (0.79–1.17)
G	1153 (42.45)	1231 (42.89)	Allele	D/G	0.740	1.02 (0.92–1.13)

†Logistic regression analysis.
‡OR: Odds ratio; 95% CI: 95% confidence interval.
D: Deletion; DD: Both alleles are Deletion; NSOFC: Nonsyndromic orofacial cleft; SNP: Single nucleotide polymorphism.

Results

Screening of candidate SNPs in miRNA promoters & enhancers

A total of 13,830 and 75,375 SNPs were identified in the miRNA promoters and enhancers, respectively, after the annotation of regulatory regions based on DHSs and ChIP-seq (H3K4me3 and H3K4me1) peaks in human embryonic craniofacial tissues (GSE97752). After screening in stage I, we identified 2279 and 12,542 SNPs in the promoter and enhancer elements, respectively (Figure 1). We selected the two most promising SNPs with the lowest p-values, specifically rs11564745 that is 6784-bp upstream of hsa-miR-675 in both enhancer and promoter elements (p = 8.37E-05) and rs3830766 that is 52,611-bp downstream of hsa-miR-4260 in the enhancer element (p = 9.52E-05), for further validation (Supplementary Table 2).

Association of SNPs with the risks of NSOFC & its subgroups

In stage II, 1435 (95.67%) cases and 1358 (97.98%) controls were successfully genotyped. The two promising SNPs conformed to HWE ($p > 0.05$) in the control groups (Supplementary Table 2). The analysis of associations between the two SNPs and the risks of NSOFC are shown in Table 2. However, no significant association was detected in any genotype or allele test models. For instance, the distributions of CC, CT and TT genotypes in rs11564745 were 594 (43.74%), 606 (44.62%) and 158 (11.64%) among the controls and 590 (41.11%), 674 (46.97%) and 171 (11.92%) among the NSOFC cases; these were not statistically significant (p = 0.365). The ORs of the heterozygote, homozygote, dominant and recessive comparison models were 1.12, 1.09, 1.11 and 1.03, while the 95% CIs were 0.96–1.31, 0.85–1.39, 0.96–1.29 and 0.82–1.29, respectively. Similarly, the allele comparison between the C and T alleles was not statistically significant (p = 0.254).

We further classified the NSOFC cases into the CLO, CLP and CPO subgroups and found that rs3830766 was nominally associated with CLO under the homozygote model (GG/DD: p = 0.018; OR = 0.67; 95% CI: 0.48–0.93; D referred to the deletion allele) and significantly associated with CLO under the recessive model (GG/DD+DG: p = 0.007; OR = 0.66; 95% CI: 0.49–0.89).The GG carrier, compared with DD and DD+DG, was associated with 33 and 34% decreased risk of CLO, respectively (Table 3).

In contrast, no significant associations between the two SNPs and the risks of CLP (Supplementary Table 3) or CPO (Supplementary Table 4) were observed.

In silico functional annotation of selected SNPs

Both rs11564745 and rs3830766 were located in vital regulatory elements (Supplementary Table 5). Rs11564745 was predicted to affect the binding motifs of three transcription factors (TFs; *ELF1*, *NRSF* and *Zfx*) and overlap with seven eQTL tissue associations, rs3830766 could potentially alter ten TF-binding motifs. The RegulomeDB

Table 3. Associations between two SNPs and cleft lip only under different genetic models.

SNP	Controls n = 1358	CLO cases n = 473	Models		p-value†	OR (95% CI)‡
rs11564745 (C>T)						
Genotype	594/606/158	196/212/65	Genotype	CC/CT/TT	0.425	
CC	594 (43.74)	196 (41.44)	Heterozygote	CT/CC	0.610	1.06 (0.85–1.33)
CT	606 (44.62)	212 (44.82)	Homozygote	TT/CC	0.191	1.25 (0.90–1.74)
TT	158 (11.64)	65 (13.74)	Dominant	TT+CT/CC	0.384	1.10 (0.89–1.36)
C	1794 (66.05)	604 (63.85)	Recessive	TT/CC+CT	0.228	1.21 (0.89–1.65)
T	922 (33.95)	342 (36.15)	Allele	C/T	0.219	1.10 (0.94–1.28)
rs3830766 (D>G)						
Genotype	450/663/245	165/248/60	Genotype	DD/DG/GG	**0.027**	
DD	450 (33.14)	165 (34.88)	Heterozygote	DG/DD	0.865	1.02 (0.81–1.29)
DG	663 (48.82)	248 (52.43)	Homozygote	GG/DD	**0.018**	0.67 (0.48–0.93)
GG	245 (18.04)	60 (12.68)	Dominant	GG+DG/DD	0.489	0.93 (0.74–1.15)
D	1563 (57.55)	578 (61.10)	Recessive	GG/DD+DG	**0.007**	0.66 (0.49–0.89)
G	1153 (42.45)	368 (38.90)	Allele	D/G	0.056	0.86 (0.74–1.00)

†Logistic regression analysis.
‡OR: Odds ratio; 95% CI: 95% confidence interval.
Note: Bold values, represent the association with nominal or significant values.
CLO: Cleft lip only; D: Deletion; DD: Both alleles are Deletion; SNP: Single nucleotide polymorphism.

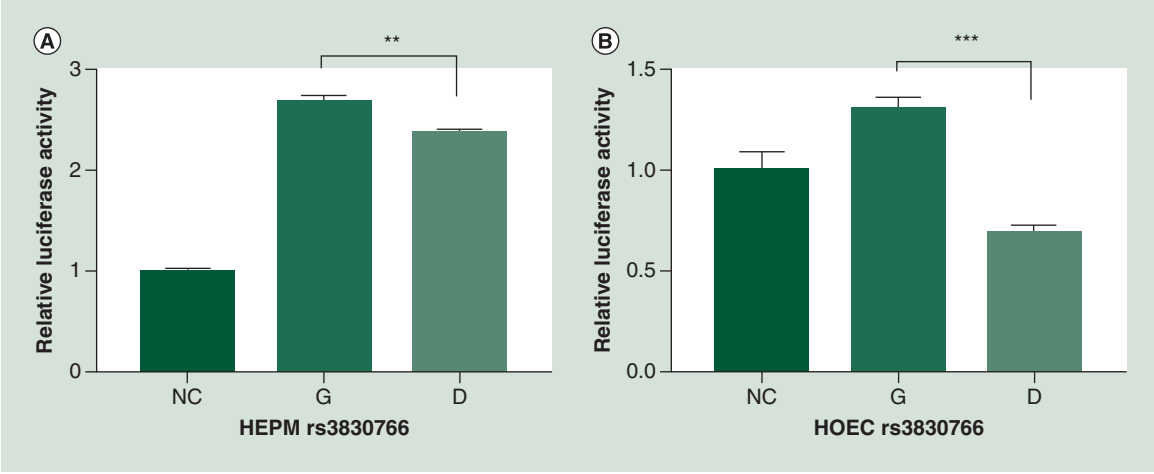

Figure 2. Allelic effects of rs3830766 on enhancer activity. rs3830766 has allelic effects on enhancer activity in **(A)** HEPM cells and **(B)** HOECs (n = 3), where the reference (ref) allele G has higher activity than the alternate allele D (Deletion). Values represent mean and SD.
**p < 0.01; ***p < 0.001.
HEPM: Human embryonic palatal mesenchyme; HOEC: Human oral epithelial cell.

scores of rs11564745 and rs3830766 were 2b and 3a, suggesting that the two SNPs were located in the DNase peak region and can potentially affect TF binding. Furthermore, The UCSC showed that rs11564745 was located within a craniofacial super-enhancer region, while rs3830766 was located within a bivalent promoter state region (Supplementary Figure 1).

Allele-specific enhancer activity of rs3830766

Using luciferase reporter assay in HEPM cells and HOECs, we validated that the rs3830766 with the G allele had a higher enhancer activity than the one with the D allele (P_{HEPM} = 0.003, P_{HOEC} = 6.45E-05) (Figure 2).

10.2217/epi-2020-0124

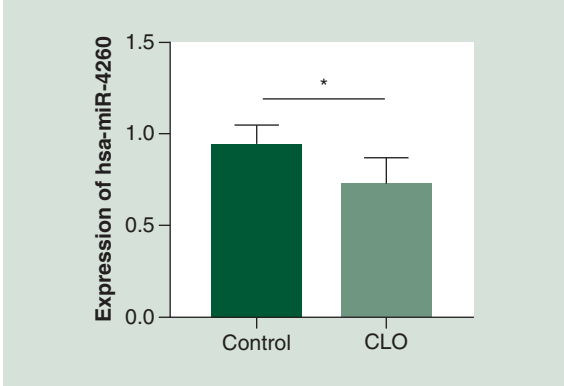

Figure 3. Expression of hsa-miR-4260 in serum.
Hsa-miR-4260 expression is significantly reduced in serum of CLO cases compared with the controls.
*$p < 0.05$.
CLO: Cleft lip only.

Expression pattern of hsa-miR-4260 in the serum of human subjects

The hsa-miR-4260 expression level was significantly lower in the serum of CLO cases than in normal controls ($p = 0.026$; Figure 3). In addition, the 3D chromatin looping data revealed that hsa-miR-4260 can interact with rs3830766 in the blood and in other tissues (Supplementary Figure 2). Therefore, rs3830766 had an effect on hsa-miR-4260 through chromatin looping and hsa-miR-4260 played a defensive role in CLO cases.

The role of miR-4260 mimics in cell proliferation & apoptosis

The assay results showed that the HEPM cells and HOECs transfected with the miR-4260 mimics had suppressed proliferation and increased apoptosis. The ectopic expression of miR-4260 mimics significantly inhibited the cell proliferation 36–60 h after transfection (Supplementary Figure 3A). Flow cytometric analysis demonstrated that the overexpression of miR-4260 mimics in HEPM cells and HOECs resulted in a significantly increased percentage of FITC-A^+ cells (Supplementary Figure 3A).

Target genes of hsa-miR-4260

According to three bioinformatics algorithms, five genes, including *ZNF215, FXR2, UBB, SKI* and *DYNAP*, were predicted as hsa-miR-4260 target genes (Supplementary Table 6). Among these, *FXR2* and *UBB* were detected in the lip and palate tissues during mouse embryonic (E10.5-E15.5) development (Supplementary Figure 4A & B). In addition, we also observed the differential expression of *UBB* ($p = 0.011$) and *FXR2* ($p = 3.60E-04$) in the dental pulp stem cells of CL/P patients and normal controls (Supplementary Figure 4C & D). Hence, we selected *FXR2* and *UBB* to verify their interactions with miR-4260 mimics using dual-luciferase reporter assays. The miR-4260 mimics significantly reduced the luciferase activity of the wild-type *UBB* 3′-UTR in HEPM cells and HOECs. In contrast, this effect was not observed in the wild-type *FXR2* 3′-UTR. Thus, *UBB* might be the actual hsa-miR-4260 target gene that was associated with CLO (Figure 4).

Discussion

Proper craniofacial development is spatiotemporally controlled by a number of regulatory mechanisms. MiRNAs have well-defined roles in vertebrate facial morphogenesis, mainly through repressing gene expression by binding to the recognition sequences located in the 3′-UTR of target mRNAs [41]. For instance, miR-96-5p and *TBX1* worked in a regulatory loop to regulate craniofacial development and the proliferation of dental progenitor cells [42]. Additionally, Mir23b and Mir133b were discovered to participate during midface development [43].

In contrast, aberrant miRNA expression results in various facial birth defects, including NSOFC [22]. Furthermore, SNPs in miRNAs may be associated with miRNA expression and the susceptibility to NSOFC. For example, the CA/AA genotypes in rs7205289 and passive smoking in the first 3 months of pregnancy reduced miR-140 expression and resulted in NSOFC [44,45]. Currently, there are no studies that focus on investigating the associations between the SNPs in the miRNA regulome and the risks of NSOFC. In the present study, we systematically filtered the SNPs in the miRNA enhancers and identified hsa-miR-4260/rs3830766 that interacted with *UBB*, which was potentially associated with the susceptibility to CLO. One limitation of our study was that CLO and CLP cases (CL/P cases) were recruited in stage I, but we did not distinguish CLO from CLP when collecting these samples, so we were unable to analyze them separately in stage I. In order to verify which subtype was related to these SNPs,

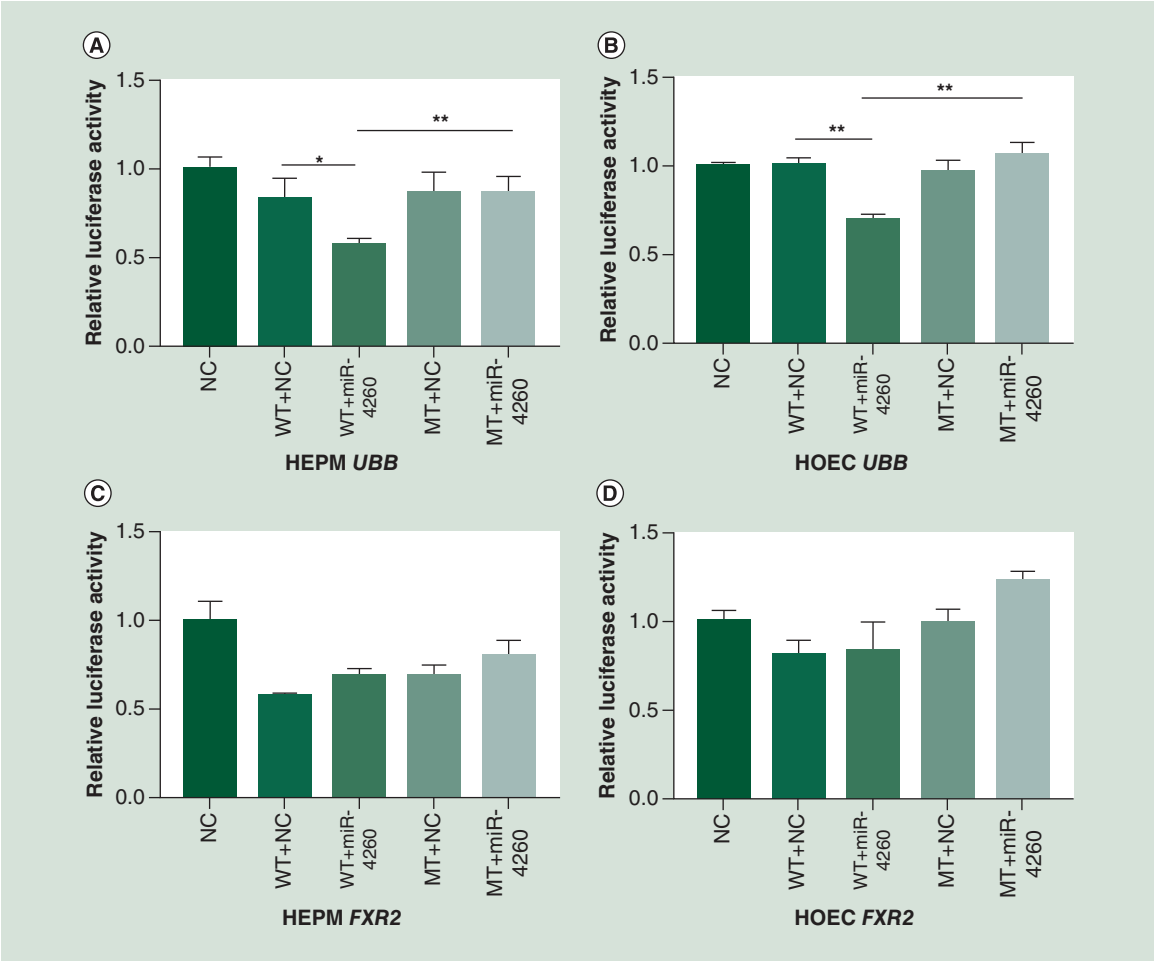

Figure 4. Binding ability assay of plasmids constructed with the 3′-untranslated region fragments of *UBB* and *FXR2* and miR-4260 mimics in human embryonic palatal mesenchyme cells and human oral epithelial cells. MiR-4260 mimics can suppress luciferase activity driven by the wild-type *UBB* 3′-UTR **(A)**, but not by the wild-type *FXR2* 3′-UTR **(B)**.
*$p < 0.05$; **$p < 0.01$.
WT: Wild-type; MT: Mutation-type; NC: Negative control.
HEPM: Human embryonic palatal mesenchyme; HOEC: Human oral epithelial cell.

we validated SNPs in three subtypes separately in stage II. However, rs3830766 was not associated with CLP and CPO, which confirmed the different genetic etiologies of the three NSOFC subgroups [46,47].

Rs3830766 was located within a bivalent promoter site that was 52,611-bp downstream of hsa-miR-4260. The bivalent promoter annotation identified the vital developmental TFs showing extensive activation during early craniofacial development [25]. Based on HaploReg v.4.1, the D allele of rs3830766 could partially inhibit *ATF3* motif binding from -4.7 to -8.7, which was important in human embryonic development during the transition from the zygote to the morula stage [48]. Furthermore, the G-to-D change in rs3830766 significantly decreased *PAX-1* motif binding from -24.5 to -50.1. The *PAX* genes were reported as the candidate genes for CL/P through maternal effects [49]. In addition, the luciferase experiments verified that rs3830766 had allelic effects on the enhancer activity. Results also suggested that rs3830766 may interact with hsa-miR-4260 through chromatin looping.

The upper lip develops according to strict procedures and programmed cell migration, proliferation and apoptosis constitute an integral part of it [50]. In this study, the miR-4260 mimics inhibited the proliferation and increased the apoptosis of HEPM cells and HOECs, which might be responsible for the CLO phenotype. Similarly, hsa-miR-4260 had increased expression in human colorectal cancer tissues and could be a potential therapeutic target for colorectal cancer [51]. It is well-known that colorectal cancer and NSOFC share some common genetic contributions, such as *MYH9* [52,53] and *TPM1* [54,55]. Interestingly, our study revealed that the CLO cases were associated with lower

hsa-miR-4260 expression in serum, suggesting that hsa-miR-4260 could be potentially used as a molecular marker for CLO. MiRNAs in the serum within the same species have the characteristics of good stability, repeatability and consistency [56]. Coenen-Stass *et al.* found that tissue-specific miRNAs can be selectively released to serum during neonatal development [57], which can potentially serve as biomarkers of diseases [58]. However, the sample size in the current study is still limited and future studies with larger sample sizes are required to replicate this finding.

One of the target genes of hsa-miR-4260 was *UBB*, which was expressed in specific NSOFC-related tissues, including lip and palate, during mouse embryonic (E10.5-E14.5) development. We queried *UBB* against the GeneCards (www.genecards.org/) [59] and DECIPHER (https://decipher.sanger.ac.uk/) [60] databases and discovered that *UBB* was associated with the cleft lip phenotype.

In summary, we identified a novel CLO risk locus rs3830766 in the hsa-miR-4260 enhancer through a two-stage case–control study. We also observed that rs3830766 was significantly associated with enhancer activity and overexpression of miR-4260 mimics suppressed cell proliferation and increased cell apoptosis. Finally, *UBB* was verified to be the target gene of hsa-miR-4260 that might be associated with CLO.

Conclusion

In this study, we identified a miRNA regulatory SNP, hsa-miR-4260/rs3830766, that could interact with *UBB* and was associated with the susceptibility to CLO.

Future perspective

Our findings provided additional evidence that CLO, CLP and CPO might be caused by different genetic contributions. The present study helped to elucidate the role of miRNAs and miRNA regulome in the development of NSOFC, as well as provided new insights that might be useful in the prevention and treatment of NSOFC. Further studies are supposed to explore the function of SNPs, miRNAs and genes *in vivo*.

Summary points

- Increasing evidence has shown that variants in miRNA promoters and enhancers are associated with various human diseases.
- Rs3830766 in the hsa-miR-4260 enhancer was associated with cleft lip only (CLO).
- Rs3830766 with the G allele had a higher enhancer activity than the one with the Deletion allele.
- Rs3830766 had a potential effect on hsa-miR-4260 through chromatin looping.
- CLO cases were associated with lower hsa-miR-4260 expression in serum compared with controls.
- Overexpression of miR-4260 mimics suppressed cell proliferation and increased cell apoptosis.
- *UBB* was expressed in lip and palate tissues and was verified to be the target gene of hsa-miR-4260.
- CLO, cleft lip and palate (CLP) and cleft palate only (CPO) might be caused by different genetic contributions.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/epi-2020-0124

Acknowledgments

We thank all the participants, research staff and students who contributed to this study. We thank Wiley Editing Services (https://wileyeditingservices.com/cn/) for providing the English writing assistance.

Financial & competing interests disclosure

This work was supported by the National Natural Science Foundation of China (81830031, 81970969, 81570959), the State Key Lab of Reproductive Medicine of Nanjing Medical University (JX116GSP20171416), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD-2018-87), the Natural Science Foundation of Jiangsu Province (BL2014073 and 15KJA320002 to L Wang), the Qing Lan Project (Y Pan), Six Distinguished Talent (2016-WSW-008) and Jiangsu Provincial Medical Talent (ZDRCC2016023 Y Pan), the Jiangsu Provincial Key Medical Discipline (zdxka2016026), the Natural Science Foundation of Jiangsu Province (BK20180667 to D Li), and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (18KJB320011). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Writing assistance was utilized in the production of this manuscript and funded by the National Natural Science Foundation of China (81570959).

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. *Lancet* 374(9703), 1773–1785 (2009).
2. Bender PL. Genetics of cleft lip and palate. *J. Pediatr. Nurs.* 15(4), 242–249 (2000).
3. Panamonta V, Pradubwong S, Panamonta M, Chowchuen B. Global birth prevalence of orofacial clefts: a systematic review. *J. Med. Assoc. Thai.* 98(Suppl. 7), S11–S21 (2015).
4. Tillman KK, Hakelius M, Hoijer J *et al.* Increased risk for neurodevelopmental disorders in children with orofacial clefts. *J. Am. Acad. Child Adolesc. Psychiatry* 57(11), 876–883 (2018).
5. Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. *Nat. Rev. Genet.* 12(3), 167–178 (2011).
6. Haaland OA, Lie RT, Romanowska J, Gjerdevik M, Gjessing HK, Jugessur A. A genome-wide search for gene-environment effects in isolated cleft lip with or without cleft palate triads points to an interaction between maternal periconceptional vitamin use and variants in ESRRG. *Front. Genet.* 9, 60 (2018).
7. Kosowski TR, Weathers WM, Wolfswinkel EM, Ridgway EB. Cleft palate. *Semin. Plast. Surg.* 26(4), 164–169 (2012).
8. Harville EW, Wilcox AJ, Lie RT, Vindenes H, Abyholm F. Cleft lip and palate versus cleft lip only: are they distinct defects? *Am. J. Epidemiol.* 162(5), 448–453 (2005).
9. Huang L, Jia Z, Shi Y *et al.* Genetic factors define CPO and CLO subtypes of nonsyndromicorofacial cleft. *PLoS Genet.* 15(10), e1008357 (2019).

• **The literature emphasizes that specific subtypes of nonsyndromic orofacial clefts could have been diluted by other subtypes.**

10. Mangold E, Bohmer AC, Ishorst N *et al.* Sequencing the GRHL3 coding region reveals rare truncating mutations and a common susceptibility variant for nonsyndromic cleft palate. *Am. J. Hum. Genet.* 98(4), 755–762 (2016).
11. Ludwig KU, Bohmer AC, Bowes J *et al.* 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(4), 829–842 (2017).
12. Bartel DP. Metazoan microRNAs. *Cell* 173(1), 20–51 (2018).
13. Bulik-Sullivan B, Selitsky S, Sethupathy P. Prioritization of genetic variants in the microRNA regulome as functional candidates in genome-wide association studies. *Hum. Mutat.* 34(8), 1049–1056 (2013).
14. Xie K, Chen M, Zhu M *et al.* A polymorphism in miR-1262 regulatory region confers the risk of lung cancer in Chinese population. *Int. J. Cancer* 141(5), 958–966 (2017).

•• **rs12740674 in miR-1262 enhancer was significantly associated with lung cancer and we were enlightened by it in the conception process of our project.**

15. Li J, Zhang Y. Relationship between promoter sequence and its strength in gene expression. *Eur. Phys. J. E Soft Matter* 37(9), 44 (2014).
16. Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. *Science* 281(5373), 60–63 (1998).
17. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130(1), 77–88 (2007).
18. Heintzman ND, Stuart RK, Hon G *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39(3), 311–318 (2007).
19. Pekowska A, Benoukraf T, Zacarias-Cabeza J *et al.* H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J.* 30(20), 4198–4210 (2011).
20. Thurman RE, Rynes E, Humbert R *et al.* The accessible chromatin landscape of the human genome. *Nature* 489(7414), 75–82 (2012).
21. Xu Y, Liu L, Liu J *et al.* A potentially functional polymorphism in the promoter region of miR-34b/c is associated with an increased risk for primary hepatocellular carcinoma. *Int. J. Cancer* 128(2), 412–417 (2011).
22. Radhakrishna U. Small players with a big role: microRNAs in pathophysiology of cleft lip and palate. *Indian J. Hum. Genet.* 18(3), 272–273 (2012).
23. Pan Y, Li D, Lou S *et al.* A functional polymorphism in the pre-miR-146a gene is associated with the risk of nonsyndromic orofacial cleft. *Hum. Mutat.* 39(5), 742–750 (2018).

24. Zou J, Li J, Li J, Ji C, Li Q, Guo X. Expression profile of plasma microRNAs in nonsyndromic cleft lip and their clinical significance as biomarkers. *Biomed. Pharmacother.* 82, 459–466 (2016).
25. Wilderman A, Van Oudenhove J, Kron J, Noonan JP, Cotney J. High-resolution epigenomic atlas of human embryonic craniofacial development. *Cell Rep.* 23(5), 1581–1597 (2018).

•• **This literature describes comprehensive epigenomic annotations from human embryonic craniofacial tissues.**

26. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences targets and gene nomenclature. *Nucleic Acids Res.* 34(Database issue), D140–D144 (2006).
27. Raney BJ, Dreszer TR, Barber GP *et al.* Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 30(7), 1003–1005 (2014).
28. Barrett T, Wilhite SE, Ledoux P *et al.* NCBI GEO: archive for functional genomics data sets – update. *Nucleic Acids Res.* 41(Database issue), D991–D995 (2013).
29. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* 10(1), 5–6 (2013).
30. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5(6), e1000529 (2009).
31. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* 44(D1), D877–D881 (2016).
32. Boyle AP, Hong EL, Hariharan M *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22(9), 1790–1797 (2012).
33. 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.* 45(D1), D643–D649 (2017).
34. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120(1), 15–20 (2005).
35. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 48(D1), D127–D131 (2020).
36. Huang HY, Lin YC, Li J *et al.* miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* 48(D1), D148–D154 (2020).
37. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25(4), 402–408 (2001).
38. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26(7), 873–881 (2010).
39. Trapnell C, Williams BA, Pertea G *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28(5), 511–515 (2010).
40. 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.* 81(3), 559–575 (2007).
41. Tavares ALP, Artinger KB, Clouthier DE. Regulating craniofacial development at the 30 end: microRNAs and their function in facial morphogenesis. *Curr. Top. Dev. Biol.* 115, 335–375 (2015).
42. Gao S, Moreno M, Eliason S *et al.* TBX1 protein interactions and microRNA-96-5p regulation controls cell proliferation during craniofacial and dental development: implications for 22q11.2 deletion syndrome. *Hum. Mol. Genet.* 24(8), 2330–2348 (2015).
43. Ding H, Hooper JE, Batzel P *et al.* MicroRNA profiling during craniofacial development: potential roles for Mir23b and Mir133b. *Front. Physiol.* 7, 1–16 (2016).
44. Li L, Meng T, Jia Z, Zhu G, Shi B. Single nucleotide polymorphism associated with nonsyndromic cleft palate influences the processing of miR-140. *Am. J. Med. Genet. A* 152A(4), 856–862 (2010).
45. Li L, Zhu GQ, Meng T *et al.* Biological and epidemiological evidence of interaction of infant genotypes at rs7205289 and maternal passive smoking in cleft palate. *Am. J. Med. Genet. A* 155A(12), 2940–2948 (2011).
46. Yu Y, Zuo X, He M *et al.* Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat. Commun.* 8(2), 14364 (2017).
47. Huang L, Jia Z, Shi Y *et al.* Genetic factors define CPO and CLO subtypes of nonsyndromicorofacial cleft. *PLoS Genet.* 15(10), e1008357 (2019).
48. Godini R, Fallahi H. Dynamics changes in the transcription factors during early human embryonic development. *J. Cell. Physiol.* 234(5), 6489–6502 (2019).
49. Sull JW, Liang KY, Hetmanski JB *et al.* Maternal transmission effects of the PAX genes among cleft case-parent trios from four populations. *Eur. J. Hum. Genet.* 17(6), 831–839 (2009).
50. Jiang R, Bush JO, Lidral AC. Development of the upper lip: morphogenetic and molecular mechanisms. *Dev. Dyn.* 235(5), 1152–1166 (2006).

51. Xiao J, Lv D, Zhou J *et al.* Therapeutic inhibition of miR-4260 suppresses colorectal cancer via targeting MCC and SMAD4. *Theranostics* 7(7), 1901–1913 (2017).
52. Wang Y, Li D, Xu Y *et al.* Functional effects of SNPs in MYH9 and risks of nonsyndromic orofacial clefts. *J. Dent. Res.* 97(4), 388–394 (2018).
53. Wang B, Qi X, Liu J *et al.* MYH9 promotes growth and metastasis via activation of MAPK/AKT signaling in colorectal cancer. *J. Cancer* 10(4), 874–884 (2019).
54. Mlakar V, Berginc G, Volavsek M, Stor Z, Rems M, Glavac D. Presence of activating KRAS mutations correlates significantly with expression of tumour suppressor genes DCN and TPM1 in colorectal cancer. *BMC Cancer* 9, 282 (2009).
55. Qian Y, Li D, Ma L *et al.* TPM1 polymorphisms and nonsyndromic orofacial clefts susceptibility in a Chinese Han population. *Am. J. Med. Genet. A* 170A(5), 1208–1215 (2016).
56. Chen X, Ba Y, Ma L *et al.* Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* 18(10), 997–1006 (2008).
57. Coenen-Stass AM, Betts CA, Lee YF *et al.* Selective release of muscle-specific, extracellular microRNAs during myogenic differentiation. *Hum. Mol. Genet.* 25(18), 3960–3974 (2016).
58. Pigati L, Yaddanapudi SC, Iyengar R *et al.* Selective release of microRNA species from normal and malignant mammary epithelial cells. *PLoS ONE* 5(10), e13515 (2010).
59. Safran M, Dalah I, Alexander J *et al.* GeneCards version 3: the human gene integrator. *Database (Oxford)* 2010, q20 (2010).
60. Firth HV, Richards SM, Bevan AP *et al.* DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am. J. Hum. Genet.* 84(4), 524–533 (2009).