medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Multi-omic analyses of a twin pair with mirror image cleft lip identifies pathogenic variant in *FGF20* modified by differential methylation upstream of *ZFP57*

Waheed Awotoye[1], Ligiane A Machado-Paula[1], Luke Hovey[1], Henry Keen[2], Michael Chimenti[2], Benjamin Darbro[2], Shareef Dabdoub[1], James C Thomas[1], Jeff Murray[2], Shankar Rengasamy Venugopalan[3], Lina Moreno-Uribe[1], Aline L Petrin[1]

1. College of Dentistry and Dental Clinics, University of Iowa, Iowa City, IA, 52240
2. Carver College of Medicine, University of Iowa, Iowa City, IA, 52240
3. Tufts University School of Dental Medicine, Boston, MA, USA

Corresponding author:
A.L. Petrin, College of Dentistry and Dental Clinics, University of Iowa, N405, DSB, 801 Newton Road, Iowa City, IA 52242, USA. Email: aline-petrin@uiowa.edu Tel: 319-335-6523. **Twitter:** @PetrinLab

Funding: NIH/NIDCR K01DE027995

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background: Disturbances in the intricate processes that control craniofacial morphogenesis can result in birth defects, most common of which are orofacial clefts (OFCs). Nonsyndromic cleft lip (nsCL), one of the phenotypic forms amongst OFCs, has a non-random laterality presentation with the left side being affected twice as often compared to the right side. This study investigates the etiology of nsCL and the factors contributing to its laterality using a pair of monozygotic twins with mirror-image cleft lip.

Methods: We conducted whole-genome sequencing (WGS) analyses in a female twin pair with mirror image nsCL, their affected mother and unaffected father to identify etiopathogenic variants. Additionally, to identify possible cleft lip laterality modifiers, DNA-methylome analysis was conducted to test for differential methylation patterns between the mirror twins. Lastly, DNA methylation patterns were also analyzed on an independent cohort of female cases with unilateral cleft lip (left=22; right=17) for replication purposes.

Results: We identified a protein-altering variant in *FGF20* (p.Ile79Val) within the fibroblast growth factor interacting family domain segregating with the nsCL in this family. Concurrently, DNA-methylome analysis identified differential methylation regions (DMRs) upstream of Zinc-finger transcription factor *ZFP57* ($\Delta\beta > 5\%$). Replication of these results on an independent cohort, confirmed these DMRs, emphasizing their biological significance ($p<0.05$). Enrichment analysis indicated that these DMRs are involved in DNA methylation during early embryo development (FDR adjusted p-value = 1.3241E-13). Further bioinformatics analyses showed one of these DMRs acting as a binding site for transcription factor *AP2A* (*TFAP2A*), a key player in craniofacial development. Interactome analysis also suggested a potential role for *ZFP57* in left/right axis specification, thus emphasizing its significance in cleft laterality.

Conclusion: This study provides novel insights into the etiology of nsCL and its laterality, suggesting an interplay between etiopathogenic variants and DNA methylation in cleft laterality. Our findings elucidate the intricate mechanisms underlying OFCs development. Understanding these factors may offer new tools for prevention and management of OFCs, alleviating the burden on affected individuals, their families and global health.

Keywords: Genetics, Fibroblast growth factor, Epigenetics, DNA methylation, Orofacial cleft

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

OFCs are birth defects that occur due to disturbances during embryonic development which results in the failure of the fusion of the upper lip (cleft lip), palate (cleft palate) or both (cleft lip and palate)(1). In addition to these structural birth defect, some cases may have additional anomalies, thus called syndromic OFCs, which account for about 30% of all OFCs(2, 3). Though variable based on ethnicity, the global incidence of OFCs averages 1 in 700 live births(1). The accompanying feeding problems, malocclusion, speech defects, esthetics problems, psychosocial impact on those affected and their families, as well as the financial burden due to the complex rehabilitation needed, contribute to the huge burden on the public health(1). Thus, efforts to understand the causes to successfully prevent new cases reducing the incidence of the clefts are warranted.

Several environmental and genetic risk factors have been implicated in OFCs developmental pathogenesis(1). Environmental risk factors impact the molecular signaling involved in craniofacial development, thus resulting in clefting via epigenetic mechanisms(4-6). One of the epigenetic factors that have been implicated in the etiology of cleft is abnormal DNA methylation(5, 7-17). DNA methylation can affect a phenotypic outcome via regulation of gene expression(5, 18, 19) by the addition of a methyl group to cytosine residues located within the CpG sites, typically found in regulatory regions of the genome, including promoters (18-20). Promoters serve as the binding sites for transcription factors and this binding activates the expression of a gene, a critical step in gene function. Methylation of CpG sites within promoter regions affect the transcription of their target genes. Thus, DNA methylation is regarded as a gene repression mechanism(21-24). This mechanism of gene repression has been reported in the developmental pathogenesis of clefting and its subtypes(7).

Monozygotic twins are invaluable to study the contribution of genetic and environmental factors to the etiology of birth defects as they share almost identical genetic compositions(25). Therefore, differences in their traits, which are regarded as discordant traits, are highly suggestive of an environmental etiology(25). Monozygotic twins arise from a single fertilized ovum that undergoes a twinning event post-fertilization. Epigenetic changes such as DNA methylation may preferentially occur in one of the twins thus contributing to the discordant phenotype(26). Differential methylation in monozygotic twin pairs has been investigated in the etiology of discordant phenotypes(26-31). A common type of monozygotic twins with nonconforming traits is the Mirror twins. Mirror twins are described as those monozygotic twins with discernable traits on contralateral sides of the body plane. About 1 in every 4 monozygotic twins show traits on contralateral sides such that one of the pair shows a trait on the right side while the other on the left side(32). Some of these traits include hair whirls, birth marks and more severely birth defects. Unilateral clefting with laterality discordance is one the birth defects that have been reported in mirror twins where one of the monozygotic pair has the cleft on the right side

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

of the lip and the other on the left(33). The mirror cleft twins are significant in understanding the mechanisms determining cleft laterality. Previous studies that examined genetic factors that contribute to nonconforming laterality in mirror twins with OFC found no discordant genetic variants that may explain the side difference(34); instead, the role of epigenetic factors has been suggested(35). Thus, we explored the genomic and epigenomic landscape in a mirror cleft twin female pair to identify the genetic etiology and the factors contributing to the laterality difference of this common birth defect.

Materials and methods

Subjects, Sample collections and DNA extraction: In this study, we recruited a case-parent quartet consisting of twin pair and their parents. The twin pair were females with mirror-image nonsyndromic cleft lip only (nsCLO) phenotype and born to Filipino parents with only the mother diagnosed with the microform cleft lip. The father did not present with any structural abnormalities. Blood samples were collected from the probands and parents. The study was approved by the respective institutional review boards and the participants signed informed consents prior to collection of clinical data and biological samples.

Next-generation sequencing analyses: We conducted whole-genome sequencing (WGS) analyses of the quartet to identify pathogenic variants that contribute to the risk of the unilateral nsCLO subtypes (right-sided vs left-sided) in the mirror-twin family pedigree. Although, a previous report found no discordant genetic variants in mirror twins with different OFC subtypes (36), we investigated this within our cohort as well. Here, our research question was designed to identify the genetic variants that contribute to the right-sided nsCL and left-sided nsCL specifically.

Based on deep phenotyping of the pedigree (Figure 1A), we first analyzed the entire genome of this pedigree for discordant pathogenic protein-altering genetic variants. This analytical pipeline (Figure 1B) is based on the hypothesis that pathogenic variant(s) unique to each of these MZ-twin contribute to the specific and discordant nsCL subtypes (the laterality differences). For the MZ-twin with left-sided cleft lip (LCL), we screened for high confidence rare/novel protein-altering variants that are unique to this twin (absent in the other MZ-twin pair) and shared with the mother (with microform LCL). This analysis to identify the unique high confidence rare/novel protein-altering variants was repeated

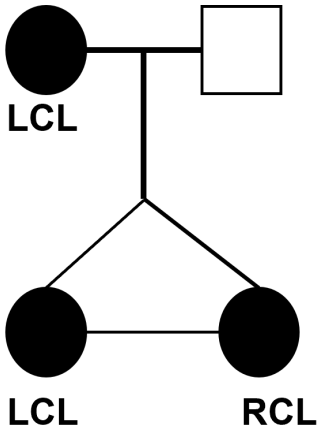


Figure 1A: Pedigree showing the inheritance pattern of clefting in the family. RCL- Right cleft lip; LCL- Left cleft lip

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

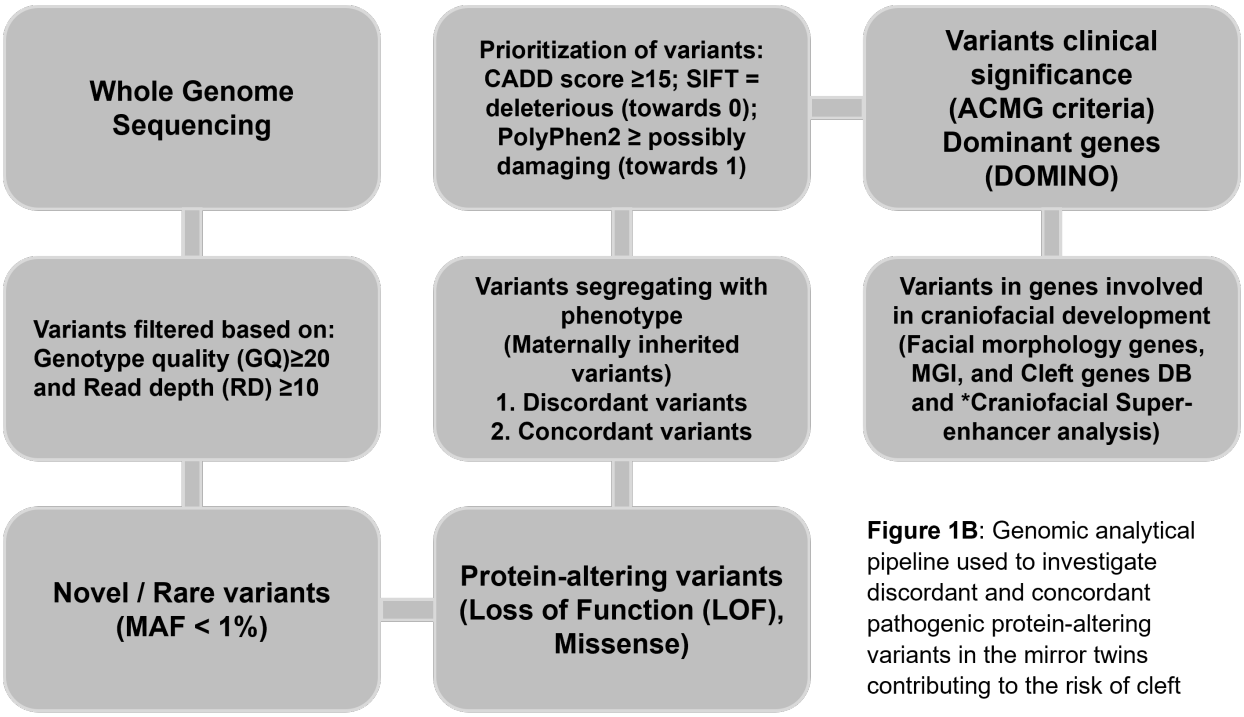


Figure 1B: Genomic analytical pipeline used to investigate discordant and concordant pathogenic protein-altering variants in the mirror twins contributing to the risk of cleft

for the MZ-twin with RCL. These unique variants in the MZ-twin with RCL were then screened to eliminate maternally inherited genetic variants.

As an alternate hypothesis, we investigated those maternally inherited shared protein-altering variants shared amongst all 3 cases but not present in the unaffected father. This is in accordance with previous studies that found no genetic basis for the laterality difference in mirror-twins with nsCL (34, 37). For each of the hypotheses, we used *in silico* tools to prioritize the variants based on their pathogenicity and on their location on genes with established roles in craniofacial development.

Genomic Analysis workflow: For both hypotheses, the entire genomic region of the parents and mirror twins were sequenced at an average coverage depth of 30x. The sequence data were aligned to the human genome assembly GRCh38 (Hg38) and alternate alleles at each genomic loci were called using the Dynamic Read Analysis for GENomics (DRAGEN). These alternate alleles include single nucleotide variants (SNVs), insertions and deletions (InDels).

We then analyzed the genomes for novel/rare protein-altering genetic variants that contribute to the risk of clefting in the twin pair. Prior to analyzing the genome for protein-altering variants, we selected the high confidence variants which are those variants with a genotype quality (GQ) of at least 20 and a read depth (RD) of at least 10. Using the genomic population database, gnomAD (https://gnomad.broadinstitute.org/), we selected those variants with minor allele frequency (MAF) less than 1% (0.01).

Next, we screened for those variants that alter the protein (missense and LOF variants), those that segregate with the phenotype within the pedigree and subsequently

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

used *in silico* tools including Sorting Intolerant From Tolerant, SIFT (http://sift.jcvi.org/), Polymorphism Phenotyping, PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/) and Combined Annotation Dependent Depletion, CADD (https://cadd.gs.washington.edu/) to predict the pathogenicity of these variants. Combinatorially, we used the American College of Medical Genetics and Genomics (ACMG) classification system to categorize these variants based on a benign or pathogenic scale. In addition, we used the web-based machine-learning algorithm, DOMINO (https://domino.iob.ch/) to assess the probability that the variants identified in these genes are dominant. Thereafter, we prioritized those variants in genes that contribute to craniofacial development. The prioritization of these genetic variants is based on the knowledge that pathogenic variants in craniofacial genes contribute to clefting.

DNA-methylome analyses: To identify factors that modify the effect of the etiopathogenic variants, thus resulting in the differential phenotypic expression in the MZ-twin (laterality difference of nsCLO), we investigated the DNA methylation patterns in each of the twins. The genome-wide DNA methylation profiles were generated using the EPIC BeadChip assay (EPIC array, Illumina, San Diego, CA, United States) which contains over 850,000 probes. Thus, this EPIC array was used to evaluate the methylation profiles of over 850K CpG sites in the mirror twins. Importantly, these sites cover annotated regions of the human genome designated as CpG islands, RefSeq genes, ENCODE open chromatin, ENCODE transcription factor-binding sites and FANTOM5 enhancers. Details of the data generation, controls and cleaning have been previously published (38).

Following data cleaning and preprocessing as previously described (38), the methylation profiles were estimated as Beta (β) values (ranging from 0 to 1), which is the ratio of methylated signals to the total sum of unmethylated and methylated signal within a CpG site. Percent composition for 12 major cell types in the blood was determined using the EpiDISH package (version 2.16.0) (39), which was used to normalize the beta-values to remove variance due to differences in cell type composition among the samples. The normalized beta-values were used for the downstream analysis. Additionally, we evaluated the methylation profiles of gender-matched controls using the same array to identify and remove highly variable CpG sites with coefficient of variation greater than 20% which are unlikely to be related to the OFC phenotypes. These nuisance CpG sites were not included in our downstream analysis.

Thereafter, we estimated the absolute methylation differences between the mirror twins ($\Delta\beta = |\text{Left CL } \beta - \text{Right CL } \beta|$) for each CpG sites remaining. CpG sites with $\Delta\beta \geq 5\%$ encompassed the identified differentially methylated positions (DMPs) for subsequent analyses below.

Subsequently, we performed gene ontology and enrichment analysis to identify the cellular processes as well as those enriched among the CpG sites within the identified DMPs.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

To demonstrate the reproducibility of these DMPs, we conducted a replication study using the CpG sites that showed a Δβ ≥ 5%. Here, we selected an independent female cohort with methylation profiles available for such analysis. This independent cohort consisted of female individuals (n=39) with unilateral clefting phenotype (nsCL/P). Of these, 22 individuals had left-sided nsCL and 17 individuals had right-sided nsCL.

Results

WGS analyses identify shared protein-altering variants in Craniofacial genes.

The analysis of discordant protein-altering variants amongst the MZ-twins that would explain the etiology of the difference in laterality affection identified a high confidence heterozygous variant in *KRT6B* present in the MZ-twin with LCL only. This variant was a paternally inherited missense variant predicted to be among the top 1% most deleterious mutations in the human genome (CADD score = 22.7) and consistently deleterious by other *in silico* tools except for ClinVar. Although *KRT6B* has not been associated with orofacial clefts, the knockout mouse displayed an abnormal hard palate morphology. Despite the identification of the damaging protein-altering variant in this gene, we found that the phenotypic effect of this gene occurs in homozygous state as predicted by the machine learning tool, DOMINO. Thus, based on the occurrence of this damaging *KRT6B* variant in a heterozygous state, the presence of this variant in an unaffected father and a ClinVar prediction of a benign effect, we concluded that this variant is unlikely to be causal for the nonconforming laterality of the cleft lip phenotype. We also found a pathogenic splice site *de novo* variant in *ACLY* gene as supported by ClinVar classification. Although, no evidence suggests a role for this gene in craniofacial development, the knockout mouse showed an early embryonic stage lethality as the mice did not survive beyond E7.

Results of this analysis in the MZ-twin with LCL did not find any variants of clinical significance in genes with established roles in craniofacial development. In fact, we

identified a protein-altering variant of uncertain significance in *MUC3A* which was maternally-inherited (heterozygous). Based on the *in silico* prediction of the *MUC3A* variant and the prediction that the gene itself is likely to contribute to a recessive phenotype, we concluded that this heterozygous maternally-transmitted variant was highly unlikely to be the risk variant. Thus, similarly to previous reports we did not find any discordant pathogenic variants that could explain the laterality affection difference amongst the members of this twin pair.

Investigation of maternally transmitted shared variants for those potentially pathogenic variants (detailed result shown in Table 1) that would explain the developmental pathogenesis of the cleft lip phenotype identified 16 variants in potential gene candidates (Table 2).

These variants are rare, or novel based on the MAF in population control databases, alter the protein sequences and ranked among the top 1% most deleterious mutations in the human genome (based on CADD scores ≥ 15). Additionally, other *in silico* tools

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1: Number of variants (CADD >15) at each analysis steps towards prioritization of pathogenic variants

Filter	Number of Variants
Protein-altering variants with MAF < 1%	955 (MZ-1); 939 (MZ-2)
Maternally inherited protein-altering variants with MAF < 1%	393 (MZ-1); 416 (MZ-2)
Shared maternally inherited protein-altering variants with MAF < 1%	387
Variants with CADD≥15	175
Variants with CADD≥15 + deleterious (SIFT)	90
Variants with CADD≥15 + damaging (Polyphen2)	82
Variants with CADD≥15 + deleterious (SIFT) + damaging (Polyphen2)	59
Variants with CADD≥15 in Craniofacial genes	16
Variants with CADD≥15 + deleterious (SIFT) and/or damaging (Polyphen2) in craniofacial genes	7

(SIFT and Polyphen2) were used to predict the deleteriousness and damaging effects of the amino acid changes on the protein (Table 1B).

These 16 protein-altering variants are in genes that contribute to craniofacial development based on evidence from the facial gene scan, mouse genome informatics and cleft genes databases. The facial genes scan consisted of list of genes that previously detected through association studies of 3D soft tissue facial morphogenesis(40). Some of these variants are in genes whose genetic manipulation in murines resulted in

Table 2: Maternally-transmitted shared variants with evidence supporting the roles of the genes in craniofacial development.

S/N	Chrom:Pos	Ref/ Alt	Gene	HGVS p.	SIFT	PolyPhen	CADD	Facial genes	Craniofacial phenotype	Cleft genes (DB)	ACMG cat	Inheritance
1	7:7452365	G/C	*COL28A1*	NP_001032852.2: p.Pro488Arg	tolerated	Possibly damaging	22.9	Mid forehead	Nil	Nil	-13 (B)	LR
2	9:126503192	C/T	*MVB12B*	NP_258257.1: p.Arg297Cys	tolerated	Probably damaging	26.4	Self-reported chin dimples	Nil	Nil	0 (VUS)	LR
3	5:113028959	C/T	*MCC*	NP_001078846.2: p.Val952Met	deleterious	Probably damaging	23.8	Upper lip center	Nil	Nil	0 (VUS)	VLR
4	13:37590506	C/T	*POSTN*	NP_006466.2: p.Gly103Ser	tolerated	Probably damaging	24.9	Nil	Abnormal ameloblast morphology; Abnormal	Nil	0 (VUS)	VLR

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

									tooth eruption			
5	21:40296147	C/T	*DSCAM*	NP_001380.2: p.Arg697Gln	tolerated	benign	20.9	Nil	Abnormal cranial cavity morphology; dome cranium; small ears	Nil	-11 (B)	VLD
6	4:2833015	C/G	*SH3BP2*	NP_003014.3: p.Ser505Cys	tolerated	benign	15.58	Nil	Abnormal craniofacial bone morphology	Nil	-14 (B)	VLR
7	15:89291615	A/C	*FANCI*	NP_001106849.1: p.Leu631Phe	tolerated	benign	15.43	Nil	Abnormal craniofacial bone morphology; Cleft palate	Nil	-21 (B)	VLR
8	20:51790193	A/G	*SALL4*	NP_065169.1: p.Ser764Pro	tolerated	benign	21	Nil	Abnormal craniofacial bone morphology; Cleft palate	Nil	-21 (B)	VLD
9	16:2100458	C/T	*PKD1*	NP_001009944.3: p.Arg3169Gln	tolerated	benign	16.98	Nil	Abnormal craniofacial bone morphology; short maxilla	Nil	-2 (LB)	VLD
10	16:81044059	G/A	*ATMIN*	NP_056066.2: p.Val521Ile	tolerated	benign	21.6	Nil	Abnormal craniofacial morphology; micrognathia; Thick upper lip	Nil	-12 (B)	VLR
11	5:84064787	T/C	*EDIL3*	NP_005702.3: p.Ile289Val	tolerated	benign	22.7	Nil	Abnormal pinna cartilage morphology; floppy ears	Nil	-5 (LB)	LR
12	8:17001798	T/C	*FGF20*	NP_062825.1: p.Ile79Val	deleterious	Possibly damaging	26.8	Nil	Abnormal tooth morphology	Nil	-15 (B)	VLD
13	8:117835473	C/T	*EXT1*	NP_000118.2: p.Val379Ile	tolerated	benign	16.39	Nil	Cleft of Secondary palate	Yes	0(VUS)	VLD

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

14	5:37180175	G/A	*CPLANE 1*	NP_075561.3: p.Pro1860Leu	deleterious	benign	22.6	Nil	Cleft of upper lip; Cleft palate (Joubert syndrome)	Yes	-7 (B)	VLR
15	14:39180971	C/T	*PNN*	NP_002678.3: p.Ala421Val	tolerated	benign	17.38	Nil	Cleft palate	Yes	-3 (LB)	VLR
16	15:74722964	C/T	*CYP1A1*	NP_001306146.1: p.Gly45Asp	deleterious	Possibly damaging	22.8	Nil	Nil	Yes	-14 (B)	LR

*LR: likely recessive; VLR: very likely recessive; VLD: very likely dominant

*Nil: not in the database investigated; Yes: Present in the investigated database

craniofacial disorders. Indeed, *FANCI, SALL4, EXT1, CPLANE1,* and *PNN* resulted in cleft phenotypes (Table 1B). Additionally, *EXT1, CPLANE1, PNN* and *CYP1A1* are among those genes within the cleft gene database (CleftGeneDB; https://bioinfo.uth.edu/CleftGeneDB)(41) supporting their roles in clefting.

The machine learning tool DOMINO makes robust and reliable inference of the inheritance patterns of different genes(42). Through these inferences of the genes' inheritance patterns, we assessed the likelihood of the genetic variants to result in dominant traits(42). Among those potentially pathogenic protein-altering variants in craniofacial genes, only *FGF20* was predicted to follow a dominant inheritance pattern, all others that had machine learning predictions follow a recessive mode of inheritance. Sanger validation confirmed the presence of this *FGF20* pathogenic mutation in the twins and the affected mother, but not the unaffected father.

Protein-function analysis showed that the isoleucine amino acid residue resides within the fibroblast growth factor family domain (IPR002209). This domain is critical for the binding of *FGF20* with other molecules and in contact with another part of the protein which contribute to *FGF20* receptor binding activity(43, 44). The wild-type residue is highly conserved at this position. Notably, the mutant residue, valine is not found in homologs, suggesting that the mutation is damaging. These analyses suggest that the damaging variants affects the interaction of *FGF20* with other molecules thereby disturbing the signal transduction function of the *FGF20* (44).

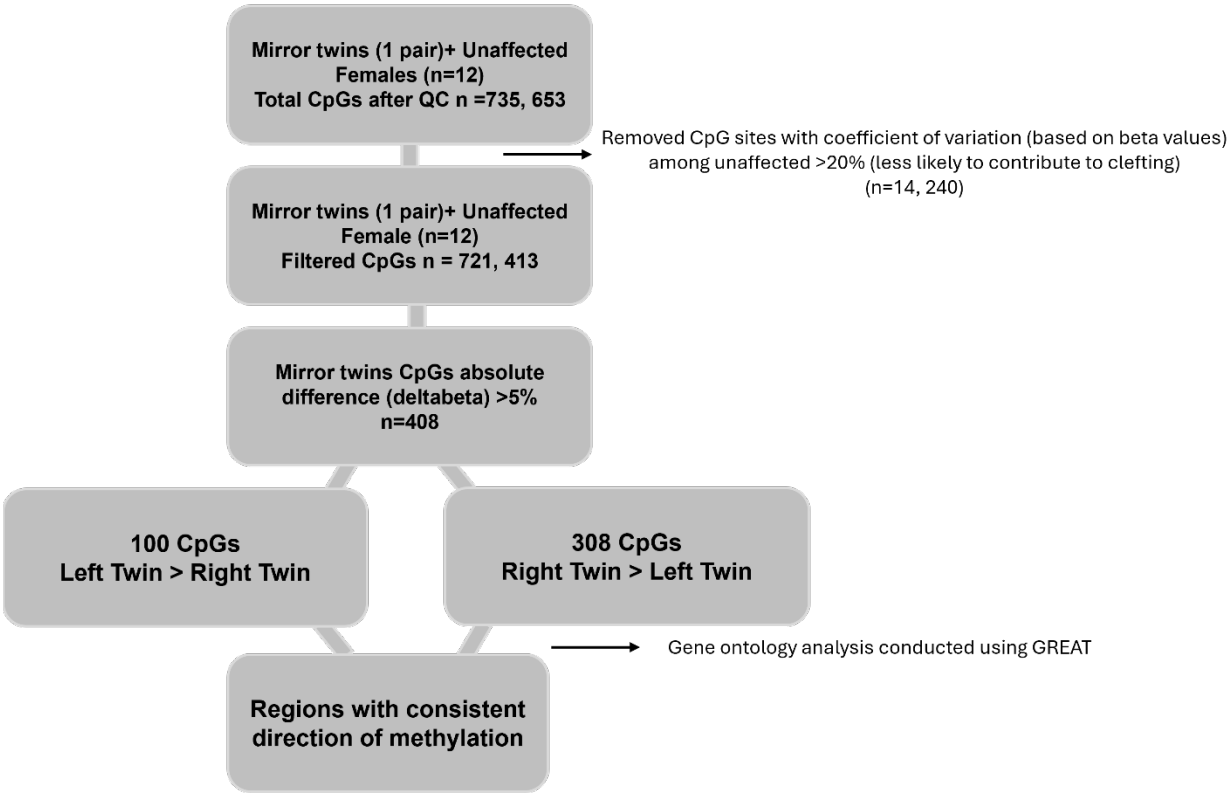
DNA-methylome Analyses Identify DMR in Zinc-finger transcription factor expressed in early embryonic development.

In the analyses of the contribution of DNA methylation to the discordant laterality affectation of cleft lip in this MZ-twin pair, we conducted a genome-wide methylation analysis using the 850k EPIC array (Figure 2). We investigated the genome-wide methylation profile of the CpG sites from the MZ-twin with nonsyndromic cleft lip as well as gender-matched unrelated controls. Methylation profiles were estimated as beta values and those CpG sites with coefficient of variation (CoV) in the controls > 20% were filtered out of this analysis as they are less likely to contribute to clefting.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

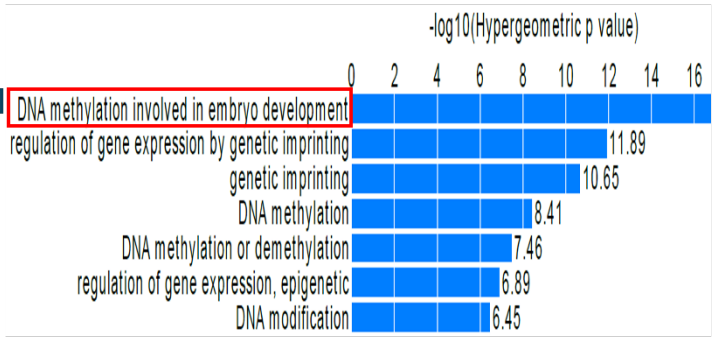
Figure 2: DNA Methylome analytical pipeline.



For remaining CpG sites, we evaluated the absolute difference of the beta values (Delta beta). This delta beta value is estimated as the difference between the absolute beta value between the mirror-twin with LCL and that with RCL ($\Delta\beta = |LCL\ \beta - RCL\ \beta|$). We thereafter selected those sites with absolute delta beta ($|\Delta\beta|$) > 5%. We identified 408 CpG site with absolute delta beta ($|\Delta\beta|$) > 5%. Of these, 100 presented higher methylation in the MZ-twin with LCL while 308 CpG sites had higher methylation values in the RCL affected twin.

To identify those CpG sites with biological

Figure 3: Enrichment analysis of DMRs.



S/N	CpG	*Genomic location	GREAT Annotation	Distance from TSS	Δβ (|LCL β – RCL β|)
1	cg00588198	6:29648452	*ZFP57*	-3522	0.0729
2	cg06032337	6:29648468	*ZFP57*	-3538	0.123
3	cg20228636	6:29648525	*ZFP57*	-3595	0.0704
4	cg03198009	6:29648604	*ZFP57*	-3674	0.0761
5	cg15570656	6:29648628	*ZFP57*	-3698	0.0729
6	cg13835168	6:29648756	*ZFP57*	-3826	0.0955
7	cg24100841	6:29649024	*ZFP57*	-4094	0.0862
8	cg19636627	6:29649084	*ZFP57*	-4154	0.0781
9	cg20961245	6:29649092	*ZFP57*	-4162	0.076

*Human genome build Hg19; TSS: Transcription Start Site

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

significance, we selected those with consistency in direction that are within 5kb of the transcription start site of a gene. Enrichment analysis showed a significant enrichment of the biological process with the GO term “DNA methylation involved in embryo development” (adjusted FDR p-value = 1.3241E-13) (Figure 3).

Further evaluation of these DMRs led to the identification of a cluster of CpG sites upstream of the transcription start site (TSS) of the *ZFP57* gene (Figure 4). Notably, these CpG sites upstream *ZFP57*, herein after termed as differentially methylated regions (DMRs), contributed significantly to the most enriched biological process. These DMRs showed a higher methylation value in the MZ-twin with LCL than that with RCL with Δβ ranging from 7% to 12%.

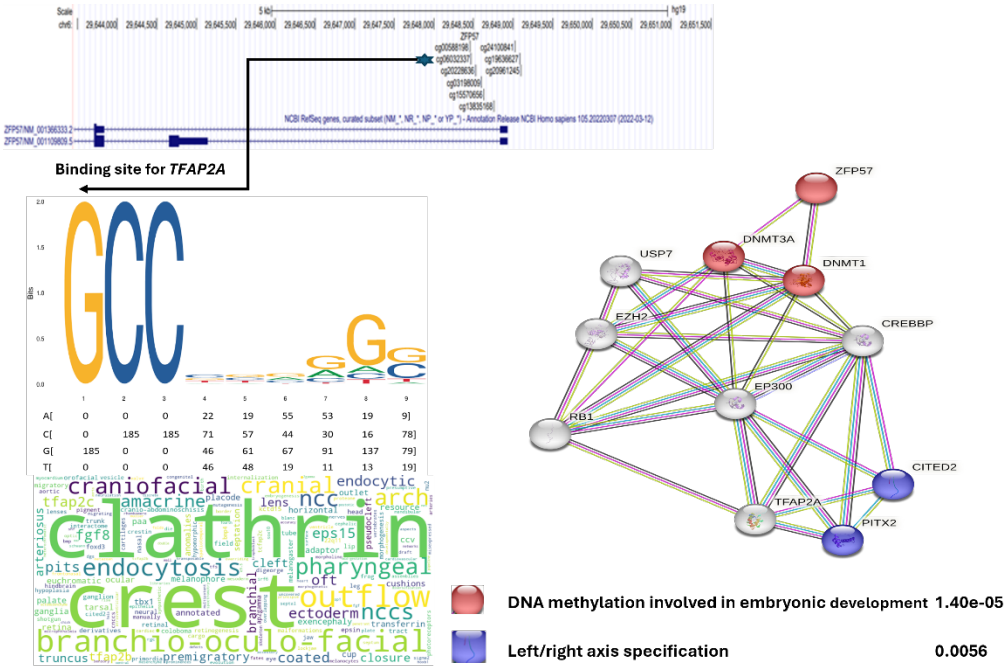


Figure 4: CpG sites upstream *ZFP57* predicted to be binding site for *TFAP2A*, an interaction that contributes to craniofacial development. Protein interactome analysis also confirms the interaction with TFAP2A. Interactome analysis also showed significant contribution of this network to methylation program involved in embryonic development and laterality (left/right axis) determination.

<u>CpG site upstream *ZFP57* replicated in independent cohort and predicted to act as *TFAP2A* binding site.</u>

Among the CpG sites upstream *ZFP57,* the cg06032337 site showed a significant difference (p-value = 0.04) in methylation between the left-sided nsCL and right-sided nsCL independent female cohort (Figure 4).Following this finding, we investigated the likelihood of this promoter region acting as a transcription faction binding site to genes involved in craniofacial development. We used the Jaspar bioinformatic analysis tool to investigate possible genes that interact with this promoter region. This CpG site is predicted to act as binding site for transcription factor *AP2A,* an interaction that is involved in craniofacial development (Figure 4).

Furthermore, the interaction between the *ZFP57* and *TFAP2A* was investigated at the protein level with *in silico* tools. We used String-db (https://string-db.org/) to investigate possible protein-protein interaction networks involving *ZFP57* and *TFAP2A* gene-

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

products as well as the functional enrichment of the resulting networks. We found significant interactions between *ZFP57* and *TFAP2A* as well as other proteins (Figure 4). Functional enrichment analysis showed that the interaction network is significantly associated with DNA methylation involved in embryonic development and left/right axis specification processes. This finding suggests that the interaction between *ZFP57* and *TFAP2A* through the CpG site (cg06032337) plays a significant role in the laterality of nonsyndromic cleft lip.

Discussion

This study was conducted in twins which are highly valuable in the genomics as well as epigenomics studies of phenotypes. Monozygotic twins share similar genetic makeup thus, phenotypic differences may be explained by epigenetic dissimilarities amongst them. Our current analysis focuses on the genomic etiology of nonsyndromic cleft lip (nsCL) and the epigenomic factors that contribute to the side on which the nsCL appeared. We therefore conducted whole-genome sequencing analyses (WGS) and DNA-methylome analyses to explore the pathogenic variants and epigenetic patterns of this unique cohort.

Whole-genome sequencing analysis have been used to identify pathogenic variants playing roles in the etiology of OFCs(45, 46). These next-generation sequencing analyses have successfully identified novel risk loci and variants associated with this complex trait(45-48). We identified protein-altering potentially pathogenic mutation in fibroblast growth factor 20 (*FGF20*) gene which is on chromosome 8p22, one of the loci associated with OFC(49, 50). To the best of our knowledge, the discovery of dominant *FGF20* mutation in these mirror-image twins' pedigree with nsCL provides for the first time, evidence for the role of rare coding *FGF20* variants in the etiopathogenesis of clefting in humans.

FGF20 is one of the members of the fibroblast growth factor family and belongs to the factor 9 (*FGF9*) subfamily which binds to the fibroblast growth factor receptors 2 and 4 (*FGFR2* and *FGFR4*) activating cellular processes that drive embryonic development(51). Expression studies have reported co-expression of *FGF20* and *FGFR2* in the developing mice palatal epithelium and the contribution of the *FGFR2* to palatogenesis has been reported(52, 53) suggesting a possible role of the gene in craniofacial development.

Due to the similar genetic makeup in monozygotic twins, we hypothesize that given that the *FGF20* pathogenic variant is present in both twins and is inherited from the affected mom, epigenetic changes are more likely to contribute to the laterality difference in the phenotypic expression. Thus, we investigated DNA methylation patterns in the mirror twins to elucidate their role in the laterality of nsCL. The DNA-methylome results identified differentially methylated regions (DMRs), with a notable enrichment in the GO term "DNA methylation involved in embryo development." Specifically, a cluster of CpG sites within 5kb upstream of the transcription start site (TSS) of the *ZFP57* gene exhibited significant methylation differences between the twins. These DMRs are associated with

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

the DNA methylation processes controlling gene expression as well as embryonic development. Although this study is the first to report the role of *ZFP57* in the developmental pathogenesis of nsCL, a previous study had reported an association of the gene with neurodevelopmental disorder(54). There is a strong correlation between these disorders suggesting pleiotropic roles of genes involved in both processes. Additionally, these suggest the early expression of *ZFP57* during embryonic development.

Further exploration of these DMRs showed that one of them: cg06032337, predictably acts as binding site for the transcription factor AP2A: a member of the transcription factor family that controls the neural crest gene regulatory network(55). Mutations in this gene have been implicated in the etiopathogenesis of clefting(56-58). Protein interactome analysis showed that the *ZFP57* protein is in a network with other proteins such that one of the sub-networks is associated with left-right axis specification during embryogenesis.

One limitation in our analysis is the sample size of monozygotic twins with nsOFCs. Albeit we detected a nominal significant association between one of the CpGs and cleft laterality, the p-value would not have withstood multiple testing. Thus, the need to replicate this study in a larger cohort to increase our power to detect significant associations after correction for multiple testing. In conclusion, this integrative OMICs analysis identified a pathogenic mutation in *FGF20* whose phenotypic expression of unilateral nsCL (left vs right) is modified by differential methylation of CpG sites upstream *ZFP57* whose gene-product is predicted to interact with other proteins in a network important for embryonic development and left-right axis specification.

References

1. Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. Lancet. 2009;374(9703):1773-85.
2. Gowans LJJ, Al Dhaheri N, Li M, Busch T, Obiri-Yeboah S, Oti AA, et al. Co-occurrence of orofacial clefts and clubfoot phenotypes in a sub-Saharan African cohort: Whole-exome sequencing implicates multiple syndromes and genes. Mol Genet Genomic Med. 2021;9(4):e1655.
3. Awotoye W, Mossey PA, Hetmanski JB, Gowans LJJ, Eshete MA, Adeyemo WL, et al. Whole-genome sequencing reveals de-novo mutations associated with nonsyndromic cleft lip/palate. Sci Rep. 2022;12(1):11743.
4. Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. Nat Rev Genet. 2011;12(3):167-78.
5. Garland MA, Sun B, Zhang S, Reynolds K, Ji Y, Zhou CJ. Role of epigenetics and miRNAs in orofacial clefts. Birth defects research. 2020;112(19):1635-59.
6. SUTHERLAND JE, COSTA M. Epigenetics and the Environment. Annals of the New York Academy of Sciences. 2003;983(1):151-60.
7. Sharp GC, Ho K, Davies A, Stergiakouli E, Humphries K, McArdle W, et al. Distinct DNA methylation profiles in subtypes of orofacial cleft. Clin Epigenetics. 2017;9:63.
8. Ulschmid CM, Sun MR, Jabbarpour CR, Steward AC, Rivera-González KS, Cao J, et al. Disruption of DNA methylation-mediated cranial neural crest proliferation and differentiation causes orofacial clefts in mice. Proc Natl Acad Sci U S A. 2024;121(3):e2317668121.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

9. Zhang B, Zhang Y, Wu S, Ma D, Ma J. DNA methylation profile of lip tissue from congenital nonsyndromic cleft lip and palate patients by whole-genome bisulfite sequencing. Birth defects research. 2023;115(2):205-17.
10. Charoenvicha C, Sirimaharaj W, Khwanngern K, Chattipakorn N, Chattipakorn SC. Alterations in DNA Methylation in Orofacial Clefts. Int J Mol Sci. 2022;23(21).
11. Alvizi L, Brito LA, Kobayashi GS, Bischain B, da Silva CBF, Ramos SLG, et al. mir152 hypomethylation as a mechanism for non-syndromic cleft lip and palate. Epigenetics. 2022:1-18.
12. Young JI, Slifer S, Hecht JT, Blanton SH. DNA Methylation Variation Is Identified in Monozygotic Twins Discordant for Non-syndromic Cleft Lip and Palate. Frontiers in Cell and Developmental Biology. 2021;9(860).
13. Xu Z, Lie RT, Wilcox AJ, Saugstad OD, Taylor JA. A comparison of DNA methylation in newborn blood samples from infants with and without orofacial clefts. Clinical epigenetics. 2019;11(1):40.
14. Howe LJ, Richardson TG, Arathimos R, Alvizi L, Passos-Bueno MR, Stanier P, et al. Evidence for DNA methylation mediating genetic liability to non-syndromic cleft lip/palate. Epigenomics. 2019;11(2):133-45.
15. Xu XY, Wei XW, Ma W, Gu H, Liu D, Yuan ZW. Genome-Wide Screening of Aberrant Methylation Loci for Nonsyndromic Cleft Lip. Chin Med J (Engl). 2018;131(17):2055-62.
16. Howe LJ, Richardson TG, Arathimos R, Alvizi L, Passos-Bueno M-R, Stanier P, et al. DNA methylation mediates genetic liability to non-syndromic cleft lip/palate. bioRxiv. 2018.
17. Alvizi L, Ke X, Brito LA, Seselgyte R, Moore GE, Stanier P, et al. Differential methylation is associated with non-syndromic cleft lip and palate and contributes to penetrance effects. Scientific reports. 2017;7(1):2441.
18. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975;187(4173):226-32.
19. Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet. 1975;14(1):9-25.
20. Levenson JM, Sweatt JD. Epigenetic mechanisms in memory formation. Nat Rev Neurosci. 2005;6(2):108-18.
21. Kiefer JC. Epigenetics in development. Dev Dyn. 2007;236(4):1144-56.
22. Iguchi-Ariga SM, Schaffner W. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes Dev. 1989;3(5):612-9.
23. Nan X, Campoy FJ, Bird A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell. 1997;88(4):471-81.
24. Boyes J, Bird A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell. 1991;64(6):1123-34.
25. Sahu M, Prasuna JG. Twin Studies: A Unique Epidemiological Tool. Indian J Community Med. 2016;41(3):177-82.
26. Bell JT, Spector TD. A twin approach to unraveling epigenetics. Trends in genetics : TIG. 2011;27(3):116-25.
27. Castillo-Fernandez JE, Spector TD, Bell JT. Epigenetics of discordant monozygotic twins: implications for disease. Genome medicine. 2014;6(7):60.
28. Bell JT, Saffery R. The value of twins in epigenetic epidemiology. International journal of epidemiology. 2012;41(1):140-50.
29. Bell JT, Spector TD. DNA methylation studies using twins: what are they telling us? Genome biology. 2012;13(10):172-.
30. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet. 2002;3(11):872-82.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

31. Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GH, Wong AH, et al. DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet. 2009;41(2):240-5.
32. Stanković-Babić G, Vujanović M, Cekić S. Identical twins with "mirror image" anisometropia and esotropia. Srp Arh Celok Lek. 2011;139(9-10):661-5.
33. Youngleson J. Discordant deformity of cleft lip and cleft palate in monozygotic twins-a case report and discussion. The Cleft Palate Journal. 1971;8(1):77-83.
34. Takahashi M, Hosomichi K, Yamaguchi T, Nagahama R, Yoshida H, Marazita ML, et al. Exploration of genetic factors determining cleft side in a pair of monozygotic twins with mirror-image cleft lip and palate using whole-genome sequencing and comparison of craniofacial morphology. Arch Oral Biol. 2018;96:33-8.
35. Wang L, Liu Z, Lin H, Ma D, Tao Q, Liu F. Epigenetic regulation of left–right asymmetry by DNA methylation. The EMBO journal. 2017;36(20):2987-97.
36. Takahashi M, Hosomichi K, Yamaguchi T, Nagahama R, Yoshida H, Marazita ML, et al. Exploration of genetic factors determining cleft side in a pair of monozygotic twins with mirror-image cleft lip and palate using whole-genome sequencing and comparison of craniofacial morphology. Archives of Oral Biology. 2018;96:33-8.
37. Takahashi M, Hosomichi K, Yamaguchi T, Nagahama R, Yoshida H, Maki K, et al. Whole-genome sequencing in a pair of monozygotic twins with discordant cleft lip and palate subtypes. Oral diseases. 2018;24(7):1303-9.
38. Petrin AL, Zeng E, Thomas MA, Moretti-Ferreira D, Marazita ML, Xie XJ, et al. DNA methylation differences in monozygotic twins with Van der Woude syndrome. Front Dent Med. 2023;4.
39. Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. BMC Bioinformatics. 2017;18(1):105.
40. Shaffer JR, Orlova E, Lee MK, Leslie EJ, Raffensperger ZD, Heike CL, et al. Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. PLoS Genet. 2016;12(8):e1006149.
41. Suzuki A, Jun G, Abdallah N, Gajera M, Iwata J. Gene datasets associated with mouse cleft palate. Data Brief. 2018;18:655-73.
42. Quinodoz M, Royer-Bertrand B, Cisarova K, Di Gioia SA, Superti-Furga A, Rivolta C. DOMINO: Using Machine Learning to Predict Genes Associated with Dominant Disorders. Am J Hum Genet. 2017;101(4):623-9.
43. Ohmachi S, Mikami T, Konishi M, Miyake A, Itoh N. Preferential neurotrophic activity of fibroblast growth factor-20 for dopaminergic neurons through fibroblast growth factor receptor-1c. J Neurosci Res. 2003;72(4):436-43.
44. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics. 2010;11:548.
45. Awotoye W, Mossey PA, Hetmanski JB, Gowans LJJ, Eshete MA, Adeyemo WL, et al. Whole-genome sequencing reveals de-novo mutations associated with nonsyndromic cleft lip/palate. Scientific Reports. 2022;12(1):11743.
46. Lace B, Pajusalu S, Livcane D, Grinfelde I, Akota I, Mauliņa I, et al. Monogenic Versus Multifactorial Inheritance in the Development of Isolated Cleft Palate: A Whole Genome Sequencing Study. Front Genet. 2022;13:828534.
47. Bishop MR, Diaz Perez KK, Sun M, Ho S, Chopra P, Mukhopadhyay N, et al. Genome-wide Enrichment of De Novo Coding Mutations in Orofacial Cleft Trios. The American Journal of Human Genetics. 2020;107(1):124-36.

medRxiv preprint doi: https://doi.org/10.1101/2024.11.16.24317351; this version posted November 18, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

48. Mukhopadhyay N, Bishop M, Mortillo M, Chopra P, Hetmanski JB, Taub MA, et al. Whole genome sequencing of orofacial cleft trios from the Gabriella Miller Kids First Pediatric Research Consortium identifies a new locus on chromosome 21. Human Genetics. 2020;139(2):215-26.

49. Awotoye W, Comnick C, Pendleton C, Zeng E, Alade A, Mossey PA, et al. Genome-wide Gene-by-Sex Interaction Studies Identify Novel Nonsyndromic Orofacial Clefts Risk Locus. Journal of Dental Research. 2021;101(4):465-72.

50. Hoffman JD, Zhang Y, Greshock J, Ciprero KL, Emanuel BS, Zackai EH, et al. Array based CGH and FISH fail to confirm duplication of 8p22-p23.1 in association with Kabuki syndrome. J Med Genet. 2005;42(1):49-53.

51. Weng M, Chen Z, Xiao Q, Li R, Chen Z. A review of FGF signaling in palate development. Biomed Pharmacother. 2018;103:240-7.

52. Lee S, Crisera CA, Erfani S, Maldonado TS, Lee JJ, Alkasab SL, et al. Immunolocalization of fibroblast growth factor receptors 1 and 2 in mouse palate development. Plast Reconstr Surg. 2001;107(7):1776-84; discussion 85-6.

53. Hosokawa R, Deng X, Takamori K, Xu X, Urata M, Bringas P, Jr., et al. Epithelial-specific requirement of *FGFR2* signaling during tooth and palate development. J Exp Zool B Mol Dev Evol. 2009;312b(4):343-50.

54. Aspra Q, Cabrera-Mendoza B, Morales-Marín ME, Márquez C, Chicalote C, Ballesteros A, et al. Epigenome-Wide Analysis Reveals DNA Methylation Alteration in *ZFP57* and Its Target RASGFR2 in a Mexican Population Cohort with Autism. Children (Basel). 2022;9(4).

55. de Crozé N, Maczkowiak F, Monsoro-Burq AH. Reiterative AP2a activity controls sequential steps in the neural crest gene regulatory network. Proc Natl Acad Sci U S A. 2011;108(1):155-60.

56. Schorle H, Meier P, Buchert M, Jaenisch R, Mitchell PJ. Transcription factor AP-2 essential for cranial closure and craniofacial development. Nature. 1996;381(6579):235-8.

57. Li H, Sheridan R, Williams T. Analysis of *TFAP2A* mutations in Branchio-Oculo-Facial Syndrome indicates functional complexity within the AP-2α DNA-binding domain. Hum Mol Genet. 2013;22(16):3195-206.

58. Van Otterloo E, Milanda I, Pike H, Thompson JA, Li H, Jones KL, et al. AP-2α and AP-2β cooperatively function in the craniofacial surface ectoderm to regulate chromatin and gene expression dynamics during facial development. eLife. 2022;11:e70511.