ORIGINAL ARTICLE

Volume 20 Issue 2 2025

DOI: 10.21315/aos2025.2002.OA04

ARTICLE INFO

Submitted: 02/05/2025
Accepted: 09/11/2025
Online: 28/12/2025

The Association of *TOX3* Copy Number Variation with Gene Expression and Susceptibility to Nonsyndromic Cleft Lip and/or Palate in a Malay Cohort

Noor Areefa Ameera Mohd Ma'amor[a], Nurul Syazana Mohamad Shah[a]*, Sarina Sulong[b], Nazia Abdul Majid[c], Izzeddin Jamil Abualjubain[a], Wan Azman Wan Sulaiman[a]

[a]*Reconstructive Sciences Unit, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, 16150 Kubang Kerian, Kelantan, Malaysia*

[b]*Human Genome Center, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, 16150 Kubang Kerian, Kelantan, Malaysia*

[c]*Institute of Biological Sciences, Faculty of Sciences, Universiti Malaya, 50603 Kuala Lumpur, Malaysia*

*Corresponding author: syazanashah@usm.my

To cite this article: Ma'amor NAAM, Shah NSM, Sulong S, Majid NA, Abualjubain IJ, Sulaiman WAW (2025). The association of *TOX3* copy number variation with gene expression and susceptibility to nonsyndromic cleft lip and/or palate in a Malay cohort. *Arch Orofac Sci*, 20(2): 151–164. https://doi.org/10.21315/aos2025.2002.OA04

To link to this article: https://doi.org/10.21315/aos2025.2002.OA04

ABSTRACT

Nonsyndromic cleft lip and/or palate (NSCL/P) is a common congenital malformation with genetic influences. While Thymocyte selection-associated high mobility group box 3 (*TOX3*) is involved in other developmental processes, its role in NSCL/P remained unexplored. This study investigated the association between *TOX3* copy number, expression, and NSCL/P in 64 Malay NSCL/P cases and 64 normal controls. Samples from patients undergoing cleft repair surgery and eligible volunteers for the control group were quantified via quantitative polymerase chain reactions (qPCR). A higher mean of *TOX3* copy number was found in cases (2.195 ± 0.689) compared to controls (1.962 ± 0.558; $p < 0.05$). Similarly, a higher *TOX3* expression was observed in cases (0.014 [IQR 0.024]) compared to controls (0.006 [IQR 0.019]; $p < 0.001$). Unadjusted analyses showed higher *TOX3* copy number (OR = 1.850; $p < 0.05$) and its expression associated with NSCL/P. However, these associations were nullified after adjusting for sex and age ($p > 0.05$). Instead, male sex emerged as a significant independent predictor for NSCL/P (adjusted OR = 4.03; $p < 0.001$). Besides, an inverse, weak correlation was observed between *TOX3* copy number and expression in NSCL/P patients ($\rho = -0.285$; $p < 0.05$) indicating the potential role of epigenetics in this condition. While male sex strongly contributed to the NSCL/P condition, our results suggest that *TOX3* is not an independent genetic risk factor for NSCL/P in this population. These results highlight sex as a primary demographic risk factor and underscore the importance of considering demographic context in genetic association studies.

Keywords: *Cleft lip; cleft palate; DNA copy number variations; gene expression; TOX3*

INTRODUCTION

Orofacial clefts, including nonsyndromic cleft lip and/or palate (NSCL/P), are common congenital defects among newborns (Worley *et al.*, 2018). Recently, it has been estimated that the prevalence of orofacial clefts affects approximately 10.8 million people, with 94.1% of the disease burden experienced by low- and middle-income countries (Massenburg *et al.*, 2021). A systematic review of the global prevalence of orofacial clefts reported that the highest prevalence was among Asians, followed by North America and Europe (Panamonta *et al.*, 2015; Wang *et al.*, 2023).

The prevalence of cleft cases varies among populations due to ethnic, racial, or geographical differences (Nabavizadeh *et al.*, 2024). In Malaysia, the birth prevalence of cleft lip and/or palate was 1 in 591 births (Thong *et al.*, 2005). Generally, the Malay ethnic group was the most prevalent, followed by the Chinese and Indians (Shah *et al.*, 2018). Meanwhile, most of the cases affected among Malaysians were cleft lip and palate (CLP), followed by cleft palate (CP), and cleft lip (CL) was the least common (Cheng *et al.*, 2013; Shah *et al.*, 2018). Concerning the sex prevalence, most studies reported that males predominate in the prevalence of NSCL/P compared to females. However, the outcome varies depending on other factors, such as the type of cleft and demographic factors (Yow *et al.*, 2021; Zhu *et al.*, 2021).

NSCL/P occurs due to disruptions in craniofacial development during the early weeks of human gestation, as early as week 4 until week 9 (Reynolds *et al.*, 2019). This can be observed through the incomplete formation of the mouth and nasal structures. Clinically, NSCL/P can be classified as complete or incomplete and unilateral or bilateral (Shkoukani *et al.*, 2014). The most common types of orofacial cleft are unilateral and bilateral CL, unilateral and bilateral CLP and CP only (Yow *et al.*, 2021).

Various problems arising from NSCL/P malformation can affect patients, including difficulties with feeding, speaking, and hearing, as well as the productivity of their daily lives, which can alleviate psychological distress among family members (de Vries *et al.*, 2014). Proper treatment and psychological support from experts can help affected families cope with these challenges (Sreejith *et al.*, 2018; Maximino *et al.*, 2022). Several studies have reported NSCL/P with a wide range of contributing risk factors, encompassing both environmental and genetic factors, as well as their interactional impact (Martinelli *et al.*, 2020).

Copy number variation is a type of structural variation in the genome segmentation that can affect the genomic dosage. These variations involve deoxyribonucleic acid (DNA) segments that range in size from a few kilobases (kb) to several megabases (Mb). The processes result from the insertions, deletions, duplications, inversions, and translocations processes (Shaikh, 2017). Some of these variations could implicate the phenotypic effect, resulting in diseases related to the immune system, syndromic congenital malformations, cancers, and neurodegenerative disorders, depending on the regulatory element mechanisms and the impact of dosage-sensitive genes (Almal & Padh, 2012).

Thymocyte selection-associated high mobility group box 3 (*TOX3*) belongs to the high mobility group box (HMG-box) superfamily and is also known as trinucleotide repeat containing 9 (TNRC9) (Jiang *et al.*, 2019). It contains a HMG-box domain and a C-terminal poly-glutamine stretch (Sahu *et al.*, 2016). In previous work, *TOX3* was identified as a calcium-dependent transactivator interacting with the cAMP response element (CRE)-binding protein (CREB) (Yuan *et al.*, 2009). Besides, several genome-wide association studies (GWAS) have identified single-nucleotide polymorphisms (SNPs) of *TOX3* associated with breast cancer susceptibility

across different populations (Palomba *et al.*, 2015; Hsieh *et al.*, 2017; Shi *et al.*, 2017). Moreover, these variants have also been implicated in neurological diseases such as restless leg syndrome (RLS) and Parkinson's disease (Mohtashami *et al.*, 2018).

Prior research has revealed a compelling link between *TOX3* and NSCL/P. Specifically, they found high linkage of *TOX3* in an extended family with members affected with NSCL/P through genome-wide linkage analysis (Mohamad Shah *et al.*, 2016). This finding was supported by a validation study, which showed that the copy number of *TOX3* in the affected family was significantly lower than that of the normal group (Mohamad Shah *et al.*, 2019). From these findings, *TOX3* might play an important role in the aetiology of NSCL/P.

Whilst these studies provide strong preliminary evidence, the scope was restricted to a single-family cohort only. This limitation arises from the issue of whether this association is a family-specific risk factor or can be generalised in a population, leading to a critical knowledge gap on the role of *TOX3* in NSCL/P. Therefore, this study aimed to address these gaps by investigating the association of *TOX3* copy number variation and gene expression with NSCL/P in a larger cohort, while controlling for the confounding effects, thereby providing crucial insights into the generalisability of this genetic mechanism underlying this condition.

MATERIAL AND METHODS

Study Population

This study was a case-control study. The procedures have been approved by Human Research Ethics Committee, Universiti Sains Malaysia (code: USM/JEPeM/21010010), in accordance with principles of the Declaration of Helsinki. Subjects' details obtained from the consent form were kept for research purposes only.

All participants were recruited from Universiti Sains Malaysia Specialist Hospital. Both consented patients and normal individuals were screened by physicians for eligibility criteria before being included in this study. The eligibility criteria for patients with NSCL/P were those who were not associated with any syndromes or other significant anomalies. Normal individuals were included once they agreed and fulfilled the criteria as healthy persons with no family history of NSCL/P. The exclusion criteria for patients with NSCL/P were those diagnosed with syndromic disease and significant abnormalities. Meanwhile, individuals who refused to allow blood withdrawal for the study were excluded.

Sample Collection

A total of 1 mL to 3 mL of blood samples was withdrawn from the included subjects. DNA and RNA samples were extracted using QIAamp® DNA Blood Mini Kit (Qiagen, USA) and QIAamp® RNA Blood Mini Kit (Qiagen, USA). The purity and concentration of DNA samples were measured using Thermo Scientific™ Multiskan Skyhigh Microplate Reader (Thermo Fisher, USA). Samples of 1.8 to 2.0 purity at A260/280 were chosen for downstream analysis.

TOX3 Copy Number Variation Assay

The copy number of *TOX3* was quantified using TaqMan™ Copy Number Assays (Assay ID: Hs03924205; Applied Biosystems, USA). Ribonuclease P (*RNAse P*) was used as a reference assay. It was conducted following the kit protocol using the quantitative polymerase chain reaction (qPCR) method. The qPCR reaction was conducted utilising the StepOnePlus™ Real-Time PCR System from Applied Biosystems. Cycle threshold (Cq) values obtained were exported to CopyCaller™ Software (Applied Biosystems, USA). The calculated copy number was included in the copy number analysis. Data were presented as mean ± standard deviation (SD).

Reverse Transcription Polymerase Chain Reaction (RT-PCR)

Reverse transcription process

Reverse transcription of RNA was conducted using the qPCR cDNA Synthesis Kit (PCR Biosystems, UK), following the protocol instructions. The cDNA was generated with oligo(dT) from 1 μg of RNA in a final volume of 20 μL. cDNA was diluted to 20 ng/μL prior to the qPCR assay.

TOX3 gene expression

Primer sequences for target genes *TOX3* and reference gene glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) were identified using the National Centre for Biotechnology Information (NCBI) database and compared with the human genome using the Basic Local Alignment Search Tool (BLAST) program (https://blast.ncbi.nlm.nih.gov/). The primer pairs obtained were cross-checked using uMelt software (https://dna-utah.org/) for melting curve prediction and OligoAnalyzer Tool (USA) for oligonucleotide property identification. The list of primer pair sequences used is listed in Table 1. The qPCR was carried out using the qPCRBIO SyGreen Mix Separate-ROX kit (PCR Biosystems, UK) according to the kit protocol. Gene expression of *TOX3* was calculated in comparison to the reference gene, *GAPDH*, using the formula $2^{-\Delta Cq}$. Data were presented as median (interquartile range [IQR]).

Statistical Analysis

We estimated the statistical power for our study using the PS Power and Sample Size Calculation software. Fisher's exact test was carried out to determine the association between sex distribution and the participant groups. The Mann-Whitney test was used to assess the median age and the gene expression between the case and control groups. An independent t-test was used to compare the *TOX3* copy number between patients and the normal group. Simple and multiple logistic regression were applied to evaluate the independent and adjusted associations of sex and *TOX3* copy number with NSCL/P. In this model, *TOX3* calculated copy number was treated as a continuous variable. We validated this approach through a visual inspection of the empirical log-odds across data quartiles, which showed a linear relationship between continuous copy number and the log-odds of the outcome. Thus, confirming the appropriateness of this approach. Besides, treating copy number variation as a continuous variable in analysis was less challenging compared to inferring it in the accurate copy number integer states (Sudmant *et al.*, 2010; Gamazon & Stranger, 2015). Both sex and *TOX3* copy number were included in a multiple regression model to control for potential confounding. Meanwhile, a non-parametric analysis of covariance (ANCOVA) on ranks was performed to compare the expression difference of *TOX3* between cases and controls while adjusting for age. The analysis was performed by converting the *TOX3* expression and age data into ranks. Then, a multiple linear regression model was applied with the ranks of *TOX3* expression as the dependent variable and group status and the ranks of age as independent variables. Correlation analysis was used to assess the correlation between *TOX3* copy number and gene expression in patients with NSCL/P and the normal group. Statistical analysis was performed using GraphPad Prism version 9.0 and SPSS software.

Table 1 Primer sequences of *GAPDH* and *TOX3* gene expression in qPCR

Genes	Accession number	Primers	Sequences (5′ to 3′)
TOX3	XM_054380040.1	Forward	ATTCCACCAATCACGCCTCC
		Reverse	GGATCGCTGAGGGCTTGAAA
GAPDH	NM_002046.4	Forward	GTCTCCTCTGACTTCAACAGCG
		Reverse	ACCACCCTGTTGCTGTAGCCAA

RESULTS

A total of 128 individuals were included in this study, comprising 64 NSCL/P cases and 64 normal controls. Patients with NSCL/P consisted of 31 males and 33 females, while the normal group consisted of 13 males and 51 females. A significant sex imbalance was observed between the groups ($p < 0.05$), with a higher proportion of females in the control group. In addition, a significant age difference was noted between cases and controls, with a median age of 1.00 (IQR: 4.83) years for cases and 30.00 (IQR: 10.50) years for controls ($p < 0.0001$) (Table 2).

Within the NSCL/P cohort, patients with CLP exhibited the most prevalent phenotype compared to isolated CL and CP (Table 2). CLP also demonstrated a significant sex-specific difference in the male cohort, covering over half of the group at 56.8% (95% CI = 42.2% to 70.3%), while only Within the NSCL/P cohort, patients with CLP exhibited the most prevalent phenotype compared to isolated CL and CP (Table 2). CLP also demonstrated a significant sex-specific difference in the male cohort, covering over half of the group at 56.8% (95% CI = 42.2% to 70.3%), while only 23.8% of the cases were found in the female cohort (95% CI = 16.0% to 33.9%). The statistically significant difference was confirmed from the non-overlapping 95% confidence intervals.

In contrast, the number of individuals by sex between CL and CP was slightly different. CL was 6.8% in males and 4.8% in females. Meanwhile, 6.8% of males and 7.1% of females had the CP phenotype. Broad and overlapping confidence intervals for both cleft types revealed no statistically significant difference in the prevalence between sexes within the cohort. The data on the type of cleft were unavailable for three female subjects (Table 2).

Table 2 Demographic and clinical characteristics of the study participants (N = 128)

Participants	Male (n = 44)[a]	Female (n = 84)	Total (N = 128)[b]	*p*-value
Normal	13 (29.5% [18.2–44.2][c])	51 (60.7% [50.0–70.5])	64 (50.0%)	< 0.05*
NSCL/P[d]				
Cleft lip	3 (6.8% [2.3–18.2])	4 (4.8% [1.9–11.6])	7 (5.5%)	
Cleft palate	3 (6.8% [2.3–18.2])	6 (7.1% [3.3–14.7])	9 (7.0%)	
Cleft lip and palate	25 (56.8% [42.2–70.3])	20 (23.8% [16.0–33.9])	45 (35.2%)	
N/A[e]	0	3 (3.6% [1.0–10.0])	3 (2.3%)	
Age (year)				
Normal		30.00 (10.50)[f]		< 0.0001
NSCL/P		1.00 (4.83)		

Notes: [a] n = number of participants; [b] N = total number of participants; [c] 95% confidence interval (%); [d] NSCL/P = nonsyndromic cleft lip and/or palate; [e] N/A = not available; [f] Data were presented as median (IQR); * *p*-value < 0.05 is significant

Analysis of the calculated *TOX3* copy number revealed a significantly higher mean in patients with NSCL/P (2.195 ± 0.689) compared to the normal group (1.962 ± 0.558; 95% CI = –0.453 to –0.014; $p < 0.038$). Meanwhile, Mann-Whitney test analysis revealed that the median of *TOX3* expression in a group of patients with NSCL/P was 0.014 (IQR: 0.024), significantly higher compared to the normal group, which was 0.006 (IQR: 0.019) (95% CI = –0.010 to –0.001; $p < 0.001$) (Fig. 1).

Initially, we evaluated the association of sex and *TOX3* copy number with NSCL/P using simple logistic regression analysis. The unadjusted association analysis demonstrated that both male sex (OR = 4.07; 95% CI = 1.836 to 9.027; $p = 0.001$) and *TOX3* copy number (OR = 1.850; 95% CI = 1.021 to 3.352; $p = 0.043$) showed a significant association with NSCL/P (Table 3).

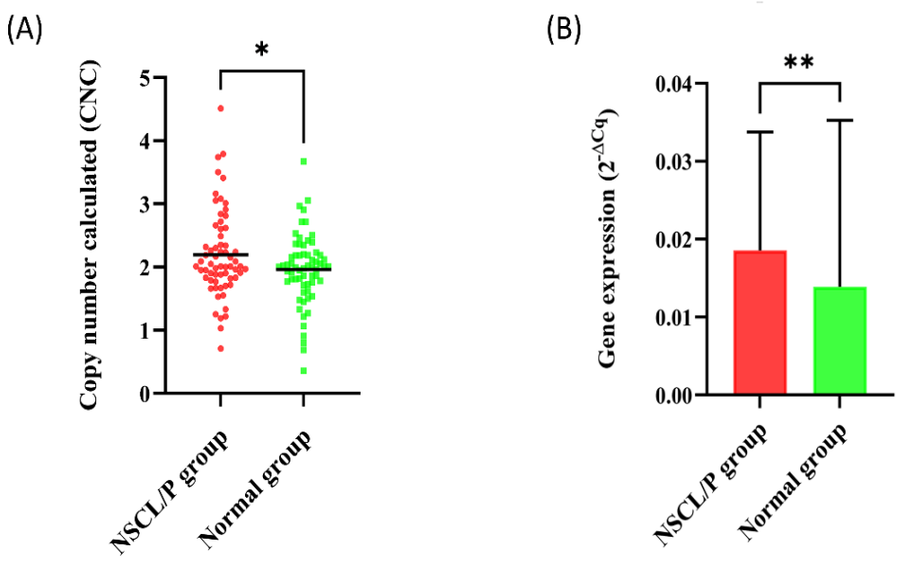

Fig. 1 (A) The distribution of *TOX3* copy number between patients with NSCL/P and normal group. The mean *TOX3* calculated copy number was higher in patients with NSCL/P compared to the normal group ($p < 0.05$, N = 128). Statistical analysis was performed using an independent *t*-test. Both data were presented as mean ± standard deviation. (B) Comparison of *TOX3* expression between patients with NSCL/P and normal group. The expression of *TOX3* was higher in patients with NSCL/P compared to the normal group ($p < 0.05$, N = 128). Statistical analysis was conducted using Mann-Whitney test. The data were presented as medians and interquartile range.

Table 3 Analysis of logistic regression on sex and *TOX3* calculated copy number with NSCL/P (N = 128)

Variable	Simple logistic regression			Multiple logistic regression		
	Regression coefficient (β)	Crude odds ratio (95% CI)[a]	*p*-value	Regression coefficient (β)	Adjusted odds ratio[b] (95% CI)	*p*-value
Sex						
Male	1.400	4.07 (1.836, 9.027)	< 0.001*	1.395	4.034 (1.793, 9.076)	< 0.001*
Female (Reference)	0	1		0	1	
Calculated copy number						
TOX3	0.615	1.850 (1.021, 3.352)	0.043*	0.619	1.858 (0.972, 3.552)	0.061

Notes: [a] 95% CI = 95% confidence interval; [b] Odds ratio (OR) of multiple logistic regression was adjusted with sex; * *p*-value < 0.05 is significant

When adjusting for sex in a multiple logistic regression model, the association between *TOX3* calculated copy number and NSCL/P was no longer significant (adjusted OR = 1.86; 95% CI = 0.97 to 3.55; $p = 0.056$). In this adjusted model, male sex remained a strong and independent predictor of NSCL/P (adjusted OR = 4.034; 95% CI = 1.793 to 9.076; $p < 0.001$) (Table 3).

To determine the influence of the age difference between cases and controls, a non-parametric analysis of covariance (ANCOVA) on ranks was performed for *TOX3* expression. After adjusting for age, we found no significant difference in *TOX3* expression between NSCL/P and normal groups ($p = 0.057$) (Table 4).

Finally, we determined the correlation between *TOX3* copy number and gene expression within each group. Our analysis revealed a significant, weak, and inverse correlation between the two variables in patients with NSCL/P ($\rho = -0.285$; 95% CI = –0.501 to –0.034; $p < 0.05$). Meanwhile, no significant correlation was found between *TOX3* copy number and gene expression in the normal group ($\rho = -0.051$; 95% CI = –0.230 to 0.205; $p > 0.05$) (Fig. 2).

Table 4 Analysis of non-parametric ANCOVA on ranks for the association of group status and age with *TOX3* expression

Variable	Regression coefficient (β)[a]	95% CI[b]	*p*-value
Group (Patient vs Control)[c]	21.620	–0.667 to 43.910	0.057
Age (year)	–0.033	–0.380 to 0.315	0.852

Notes: [a] Coefficients (β) represent the estimated change in the mean rank of TOX3 expression; [b] 95% CI = 95% confidence interval; [c] The coefficient for 'Group' represents the adjusted difference in the mean rank for the Patient group compared to the Normal group, which served as the reference category.

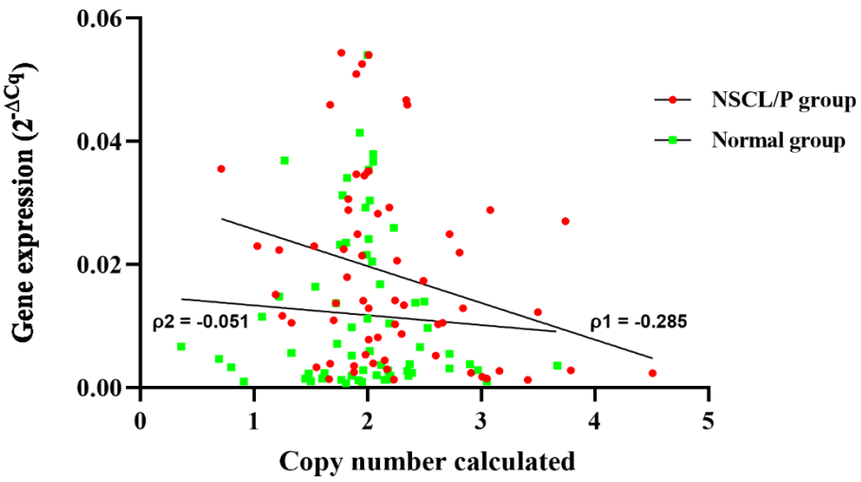

Fig. 2: Correlation analysis between TOX3 copy number calculated and gene expression. The analysis was conducted on 64 subjects in each group of patients with NSCL/P and normal group. A significant weak, inverse correlation was observed in TOX3 calculated copy number and gene expression of patients with NSCL/P ($\rho1 = -0.285$, $p < 0.05$). Meanwhile, the correlation between TOX3 calculated copy number and gene expression in normal group was not significant ($\rho2 = -0.051$, $p > 0.05$).

DISCUSSION

This study was conducted from 2022 to 2024 at the Universiti Sains Malaysia Specialist Hospital in Kubang Kerian, Kelantan, Malaysia. The current study sample reported that the highest case was observed in CLP at 70.31%, followed by CP (14.06%), and the least was CL (10.94%). CLP is the most common in the population compared to the other types of clefts globally (Chowchuen *et al.*, 2016; Nahas *et al.*, 2021; Yow *et al.*, 2021). Some literature also reports CP as the second most common condition found among patients, with CL being the least common (Pavri & Forrest, 2013; Malic *et al.*, 2020). On the other hand, some studies revealed that CL and CP were the most prevalent in their study. The divergence perceived in the cleft type of distribution is potentially reflected by the different ethnicities and regions (Abdulhameed *et al.*, 2014; Ryu *et al.*, 2022).

Overall, we found that the number of females was higher in NSCL/P than in males. Females also dominated the CL and CP categories. Meanwhile, males were more prevalent in CLP than females. Although our sample size was smaller compared to other prevalence studies, our results on sex in CP and CLP were consistent with most of the studies worldwide (Eshete, 2021; Sakran *et al.*, 2021; Zhu *et al.*, 2021).

Meanwhile, in the CL case, our result on the higher females compared to the males was consistent with previous retrospective studies (Bartzela *et al.*, 2021; Errari-Piloni *et al.*, 2021). In contrast, most regional studies reported males having a higher prevalence of CL compared to females (Galeh *et al.*, 2021; Belachew *et al.*, 2022; Ryu *et al.*, 2022).

In addition, our findings on the strong association between males and NSCL/P were supported by a cross-sectional study performed in Brazil (Martelli *et al.*, 2015). Besides the effect of genetic and environmental factors, such as smoking and alcohol exposure during pregnancy (Babai & Irving, 2023), the differences in cleft occurrence during embryogenesis depend

on sex. Clefts in females were more likely to occur during the late embryonic period, which, in most cases, involved the abnormal fusion of the secondary palate (Pool *et al.*, 2021). Meanwhile, in a separate study, a higher occurrence of CP in females might be due to the longer time of palatal fusion in the female embryo compared to the male. This condition could lead to the possibility of the pregnant woman being exposed to teratogenic factors within the critical period (Burdi & Silvey, 1969).

On the other hand, clefts occur more frequently in both early and late embryonic periods for male infants, specifically during the differentiation and fusion phases, which affect the primary and secondary palate (Pool *et al.*, 2021). Parents should be encouraged to attend early maternity check-ups during pregnancy, enabling early detection of any malformation such as CL and/or palate. Thus, allowing parental education and procedures to cater to the affected children after birth (Vyas *et al.*, 2020).

Initially, we observed a higher *TOX3* copy number in our NSCL/P cohort, which contrasts with the findings of Mohamad Shah *et al.* (2019), who reported a lower copy number. This discrepancy might be due to the different study designs-a family-based and a population-based genetic study.

Previous work by Mohamad Shah *et al.* (2019) employed a classic family-based linkage study, which found that a lower copy number of *TOX3* reflected the specific family's genetic background. On the other hand, our study included a larger and genetically diverse population, in which we expected the genetic causes to be more heterogeneous. The higher *TOX3* copy number in our cohort study suggests that the genetic architecture may vary across human populations. Such differences often arise from the occurrence of natural selection, genetic drift, and population expansion over time (Abd Rahim *et al.*, 2017).

While the unadjusted analyses suggested that high *TOX3* copy number and its expression significantly contribute as risk factors, these results were no longer significant after adjustment for sex and age ($p = 0.056$ for copy number association; $p = 0.057$ for expression difference after age adjustment). Instead, the male sex withstood the advanced statistical adjustment, reaffirming its critical role in NSCL/P aetiology. These results strongly suggest that *TOX3* is not an independent genetic risk factor for NSCL/P in this population, and its apparent associations in univariate analyses were likely driven by confounding demographic factors. These potential weak trends or hints of association would require confirmation covering larger region and well-matched cohorts to assess the independent contribution of *TOX3* in NSCL/P population.

In addition to the weak inverse correlation between TOX3 copy number and gene expression observed in this study, other factors might influence the changes in either copy number or gene expression. Inherited copy number from the parental origin with various haplotype combinations, which might be presented due to the linkage disequilibrium, resulting in allele-specific differences in gene expression. Furthermore, in this study, post-transcriptional regulation might be responsible for this condition. As such, *TOX3* expression might be altered, opposing the gene-dosage effect, potentially through the involvement of trans-regulatory elements that bind to the DNA sequence (Gamazon & Stranger, 2015). In addition, this inverse correlation also suggests that epigenetic changes occur throughout the individual's lifespan and may potentially differ between males and females (Shealy *et al.*, 2025). In this study, we found that age disparity and sex variable proved to be critical confounders, as the effect of *TOX3* copy number and its expression became less significant in our adjusted models.

Rather than demonstrating a simple gene dosage law, the result is likely due to the composite signal formed by these identified confounders. This context highlights the importance of considering age and sex in future genetic studies with larger and stratified cohorts, as disease development also relies on the combinations and interactions of other genetic variations and environmental factors (Almal & Padh, 2012).

This study had some limitations and challenges. A primary challenge that we faced in this study was the attempt to collect an age-matched control group. Our patient samples were collected postnatally during their scheduled CLP repair surgeries. While most of them were around 1 year old, most of the parents/guardians allowed the blood collection during the surgery. Recruiting age-matched infant controls would be ideal but was not feasible due to the invasive blood draw procedure that could bring emotional hurdles for the patients and parents. Alternatively, our controls were recruited from eligible consented control adults, who are 18 years old and above, to minimise procedural anxiety, leading to an unavoidable age disparity between the case and control groups.

In addition to the age disparity, a significant sex imbalance was observed in our sampling population. While sex and age imbalances may naturally differ in population-based settings, we rigorously addressed these challenges using multiple logistic regression for sex and ANCOVA on ranks for age adjustment in expression analysis. We acknowledge that alternative approaches, such as prospective matching and stratified analysis, could have been considered from the outset to mitigate these imbalances. However, given the retrospective nature of some sample collection and the practical challenges of precisely matching for multiple demographic and genetic factors in this specific population, a post-hoc statistical adjustment strategy was deemed most appropriate and feasible. The finding that *TOX3* associations were nullified after these adjustments, while male sex emerged as a strong independent predictor, underscores the effectiveness of our approach in disentangling the effects of demographic confounders from genetic associations. Despite our careful statistical adjustments, we recognise that inherent differences between unmatched groups could still influence subtle findings. Future studies with prospectively matched cohorts or larger sample sizes, allowing for more robust stratified analyses, would further strengthen the investigation into gene-sex interactions in NSCL/P.

Another key consideration is the biological material chosen in this study. We acknowledged that lip or palate tissues are more relevant for studying NSCL/P; however, collecting tissue samples is often infeasible for a large-scale population study. Accordingly, we utilised the peripheral blood, as it is widely accepted and easily accessible. This methodology was supported by Sharp *et al*. (2017) study, which reported that DNA methylation profiles at specific genomic sites in blood are highly correlated with those in orofacial tissues, proving that blood-based samples can provide insights into the underlying genetic and epigenetic aspects of NSCL/P.

Another limitation of this study is the absence of data on environmental and familial risk factors associated with NSCL/P, such as maternal folic acid intake, maternal smoking, family history, and maternal age. Lacking this information prevented us from exploring the possibility of gene-environment interactions, which are crucial in understanding disease complexity. As such, the probability of environmental factors influencing the biological effects of *TOX3*, whether they amplify or mitigate these effects in NSCL/P, remains unknown.

CONCLUSION

In conclusion, our study did not find a direct, independent association between *TOX3* and NSCL/P status in this population, whereas male sex emerged as a substantial contributing factor. The weak and inverse correlation between *TOX3* copy number and its expression suggests the role of epigenetic factors underlying this condition. We suggest confirming the role of *TOX3* in NSCL/P in a larger cohort with a precisely matched case-control group. Additionally, accounting for environmental factors in future studies is also crucial to exploring the potential of genetic-environmental interactions. In addition to this, functional validation at the cellular level using craniofacial precursor cells or genetically modified animal models could explore the effect of *TOX3* dysregulation on NSCL/P at the molecular level. Integrating multi-omics analysis, including transcriptomics, epigenomics, and proteomics, would enlighten the regulatory networks involving *TOX3* in craniofacial development and NSCL/P pathogenesis.

ACKNOWLEDGEMENTS

We are deeply grateful to all the participants for their willingness to participate in this study. We also appreciate the Centre for Research Laboratory staff, medical officers, Ms. Nadiah binti Sa'at, and Dr. Noor Fatmawati Mokhtar, for their assistance throughout this research. This work was supported by the Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education Malaysia; reference code: FRGS/1/2020/SKK0/USM/02/14.

REFERENCES

Abd Rahim Z, Bakar SA, KQueen CY, Khan FAA AHA, Tajuddin M (2017). A preliminary study on the distribution of beta defensins copy number variable gene in different ethnics of Sarawak, Malaysian Borneo. *J Sustain Sci Manag*, **12**(1): 102–113.

Abdulhameed FD, Sabbagh HJ, Hummaida TI, Alamoudi NM (2014). Epidemiology of non-syndromic orofacial cleft (NSOFC) in Medina, Saudi Arabia. *Exp Clin Cardiol*, **20**(7): 505–516.

Almal SH, Padh H (2012). Implications of gene copy-number variation in health and diseases. *J Hum Genet*, **57**(1): 6–13. https://doi.org/10.1038/jhg.2011.108

Babai A, Irving M (2023). Orofacial clefts: Genetics of cleft lip and palate. *Genes*, **14**(8): 1603. https://doi.org/10.3390/genes14081603

Bartzela T, Theuerkauf B, Reichardt E, Spielmann M, Opitz C (2021). Clinical characterization of 266 patients and family members with cleft lip and/or palate with associated malformations and syndromes. *Clin Oral Investig*, **25**(9): 5531–5540. https://doi.org/10.1007/s00784-021-03863-2

Belachew FK, Gerbu DG, Weldesenbet EB, Abay ES, Maswime S, Eshete M (2022). Clinical profiles of children born with orofacial clefts: Results from fourteen East African countries [preprint]. *medRxiv*. https://doi.org/10.1101/2022.11.09.22282144

Burdi AR, Silvey RG (1969). Sexual differences in closure of the human palatal shelves. *Cleft Palate J*, **6**(1): 1–7.

Cheng CS, Jimeno SKL, Sasidaran R, Sergius A (2013). Pilot epidemiological study of cleft lip and/or palate in Kota Kinabalu, Sabah. *Asian J Med Sci*, **4**(3): 86–91.

Chowchuen B, Thanaviratananich S, Chichareon V, Kamolnate A, Uewichitrapochana C, Godfrey K (2016). A multisite study of oral clefts and associated abnormalities in Thailand: the epidemiologic data. *Plast Reconstr Surg Glob Open*, **3**(12): e583. https://doi.org/10.1097/GOX.0000000000000570

de Vries IA, Breugem CC, van der Heul AM, Eijkemans MJ, Kon M, Mink van der Molen AB (2014). Prevalence of feeding disorders in children with cleft palate only: A retrospective study. *Clin Oral Investig*, **18**(5): 1507–1515. https://doi.org/10.1007/s00784-013-1117-x

Errari-Piloni C, Barros LAN, Jesuíno FAS, Valladares-Neto J (2021). Prevalence of cleft lip and palate and associated factors in Brazil's Midwest: A single-center study. *Braz Oral Res*, **35**: e039. https://doi.org/10.1590/1807-3107bor-2021.vol35.0039

Eshete M (2021). Pattern of orofacial clefts at a tertiary care hospital in Ethiopia. *Ethiop J Health Sci*, **31**(6): 1175–1184. https://doi.org/10.4314/ejhs.v31i6.12

Galeh SD, Nouri-Vaskeh M, Alipour M, Fakhim SA (2021). Clinical and demographical characteristics of cleft lip and/or palate in the northwest of Iran: An analysis of 1500 patients. *Cleft Palate Craniofac J*, **58**(10): 1281–1286. https://doi.org/10.1177/1055665620980633

Gamazon ER, Stranger BE (2015). The impact of human copy number variation on gene expression. *Brief Funct Genomics*, **14**(5): 352–357. https://doi.org/10.1093/bfgp/elv017

Hsieh YC, Tu SH, Su CT, Cho EC, Wu CH, Hsieh MC *et al.* (2017). A polygenic risk score for breast cancer risk in a Taiwanese population. *Breast Cancer Res Treat*, **163**(1): 131–138. https://doi.org/10.1007/s10549-017-4144-5

Jiang B, Chen W, Qin H, Diao W, Li B, Cao W *et al.* (2019). *TOX3* inhibits cancer cell migration and invasion via transcriptional regulation of SNAI1 and SNAI2 in clear cell renal cell carcinoma. *Cancer Lett*, **449**: 76–86. https://doi.org/10.1016/j.canlet.2019.02.020

Malic CC, Lam M, Donelle J, Richard L, Vigod SN, Benchimol EI (2020). Incidence, risk factors, and mortality associated with orofacial cleft among children in Ontario, Canada. *JAMA Netw Open*, **3**(2): e1921036–e1921036. https://doi.org/10.1001/jamanetworkopen.2019.21036

Martelli DR, Coletta RD, Oliveira EA, Swerts MS, Rodrigues LA, Oliveira MC *et al.* (2015). Association between maternal smoking, gender, and cleft lip and palate. *Braz J Otorhinolaryngol*, **81**(5): 514–519. https://doi.org/10.1016/j.bjorl.2015.07.011

Martinelli M, Palmieri A, Carinci F, Scapoli L (2020). Non-syndromic cleft palate: An overview on human genetic and environmental risk factors. *Front Cell Dev Biol*, **8**: 592271. https://doi.org/10.3389/fcell.2020.592271

Massenburg BB, Hopper RA, Crowe CS, Morrison SD, Alonso N, Calis M *et al.* (2021). Global burden of orofacial clefts and the world surgical workforce. *Plast Reconstr Surg*, **148**(4): 568e–580e. https://doi.org/10.1097/prs.0000000000008334

Maximino LP, Marcelino FC, Cavalheiro MG, Abramides DVM, Caldana ML, Corrêa CC *et al.* (2022). Auditory and language skills in children with cleft lip and palate. *Acta Otorrinolaringol Esp (Engl Ed)*, **73**(3): 157–163. https://doi.org/10.1016/j.otoeng.2020.11.005

Mohamad Shah NS, Salahshourifar I, Sulong S, Wan Sulaiman WA, Halim AS (2016). Discovery of candidate genes for nonsyndromic cleft lip palate through genome-wide linkage analysis of large extended families in the Malay population. *BMC Genet*, **17**: 39. https://doi.org/10.1186/s12863-016-0345-x

Mohamad Shah NS, Sulong S, Wan Sulaiman WA, Halim AS (2019). Two novel genes *TOX3* and *COL21A1* in large extended Malay families with nonsyndromic cleft lip and/or palate. *Mol Genet Genomic Med*, **7**(5): e635. https://doi.org/10.1002/mgg3.635

Mohtashami S, He Q, Ruskey JA, Zhou S, Dion PA, Allen RP *et al.* (2018). *TOX3* variants are involved in restless legs syndrome and Parkinson's disease with opposite effects. *J Mol Neurosci*, **64**(3): 341–345. https://doi.org/10.1007/s12031-018-1031-4

Nabavizadeh SS, Mootz JJ, Nadjmi N, Massenburg BB, Khoshnood K, Shojaeefard E *et al.* (2024). Gender inequality and burden of orofacial clefts in the Eastern Mediterranean region: Findings from global burden of disease study 1990–2019. *BMC Pediatr*, **24**(1): 76. https://doi.org/10.1186/s12887-024-04569-6

Nahas LD, Alzamel O, Dali MY, Alsawah R, Hamsho A, Sulman R *et al.* (2021). Distribution and risk factors of cleft lip and palate on patients from a sample of Damascus hospitals - A case-control study. *Heliyon*, 7(9): e07957. https://doi.org/10.1016/j.heliyon.2021.e07957

Palomba G, Loi A, Porcu E, Cossu A, Zara I, Budroni M *et al.* (2015). Genome-wide association study of susceptibility loci for breast cancer in Sardinian population. *BMC Cancer*, **15**(1): 383. https://doi.org/10.1186/s12885-015-1392-9

Panamonta V, Pradubwong S, Panamonta M, Chowchuen B (2015). Global birth prevalence of orofacial clefts: A systematic review. *J Med Assoc Thai*, **98**(Suppl 7): S11–S21.

Pavri S, Forrest C (2013). Demographics of orofacial clefts in Canada from 2002 to 2008. *Cleft Palate Craniofac J*, **50**(2): 224–230. https://doi.org/10.1597/10-223

Pool SMW, der Lek LMV, de Jong K, Vermeij-Keers C, Mouës-Vink CM (2021). Embryologically based classification specifies gender differences in the prevalence of orofacial cleft subphenotypes. *Cleft Palate Craniofac J*, **58**(1): 54–60. https://doi.org/10.1177/1055665620935363

Reynolds K, Kumari P, Sepulveda Rincon L, Gu R, Ji Y, Kumar S *et al.* (2019). Wnt signaling in orofacial clefts: Crosstalk, pathogenesis and models. *Dis Model Mech*, **12**(2): dmm037051. https://doi.org/10.1242/dmm.037051

Ryu JY, Park TH, Cho BC, Choi KY (2022). The prevalence, risk of premature births, mortality and causes of death of cleft lip with or without palate in South Korea: A nationwide population-based cohort study. *Int J Epidemiol*, **51**(3): 974–983. https://doi.org/10.1093/ije/dyac019

Sahu SK, Fritz A, Tiwari N, Kovacs Z, Pouya A, Wüllner V *et al.* (2016). *TOX3* regulates neural progenitor identity. *Biochim Biophys Acta*, **1859**(7): 833–840. https://doi.org/https://doi.org/10.1016/j.bbagrm.2016.04.005

Sakran KA, Mashrah MA, Al-Rokhami RK, Hsieh TY, Huang H, Alkebsi K *et al.* (2021). Nonsyndromic oral clefts and associated risk factors in Gansu Province, Northwest of China. *J Oral Maxillofac Surg Med Pathol*, **33**(5): 494–499. https://doi.org/10.1016/j.ajoms.2021.02.012

Shah SYA, Mirani SA, Sahito MA (2018). Evaluating occurrence of variable cleft lip and palate types among ethnic groups of Malaysia. *J Pak Dent Assoc*, **27**(1): 9–12. https://doi.org/10.25301/JPDA.271.9

Shaikh TH (2017). Copy number variation disorders. *Curr Genet Med Rep*, **5**(4): 183–190. https://doi.org/10.1007/s40142-017-0129-2

Sharp GC, Ho K, Davies A, Stergiakouli E, Humphries K, McArdle W *et al.* (2017). Distinct DNA methylation profiles in subtypes of orofacial cleft. *Clin Epigenetics*, **9**(1): 63. https://doi.org/10.1186/s13148-017-0362-2

Shealy EP, Schwartz TS, Cox RM, Reedy AM, Parrott BB (2025). DNA methylation–based age prediction and sex-specific epigenetic aging in a lizard with female-biased longevity. *Sci Adv*, **11**(5): eadq3589. https://doi.org/10.1126/sciadv.adq3589

Shi M, O'Brien KM, Sandler DP, Taylor JA, Zaykin DV, Weinberg CR (2017). Previous GWAS hits in relation to young-onset breast cancer. *Breast Cancer Res Treat*, **161**(2): 333–344. https://doi.org/10.1007/s10549-016-4053-z

Shkoukani MA, Lawrence LA, Liebertz DJ, Svider PF (2014). Cleft palate: A clinical review . *Birth Defects Res C Embryo Today*, **102**(4): 333–342. https://doi.org/10.1002/bdrc.21083

Sreejith VP, Arun V, Devarajan AP, Gopinath A, Sunil M (2018). Psychological effect of prenatal diagnosis of cleft lip and palate: A systematic review. *Contemp Clin Dent*, **9**(2): 304–308. https://doi.org/10.4103/ccd.ccd_673_17

Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A *et al.* (2010). Diversity of human copy number variation and multicopy genes. *Science*, **330**(6004): 641–646. https://doi.org/10.1126/science.1197005

Thong MK, Ho JJ, Khatijah NN (2005). A population-based study of birth defects in Malaysia. *Ann Hum Biol*, **32**(2): 180–187. https://doi.org/10.1080/03014460500075332

Vyas T, Gupta P, Kumar S, Gupta R, Gupta T, Singh HP (2020). Cleft of lip and palate: A review. *J Fam Med Prim Care*, **9**(6): 2621–2625. https://doi.org/10.4103/jfmpc.jfmpc_472_20

Wang D, Zhang B, Zhang Q, Wu Y (2023). Global, regional and national burden of orofacial clefts from 1990 to 2019: An analysis of the Global Burden of Disease Study 2019. *Ann Med*, **55**(1): 2215540 https://doi.org/10.1080/07853890.2023.2215540

Worley ML, Patel KG, Kilpatrick LA (2018). Cleft lip and palate. *Clin Perinatol*, **45**(4): 661–678. https://doi.org/10.1016/j.clp.2018.07.006

Yow M, Jin A, Yeo GSH (2021). Epidemiologic trends of infants with orofacial clefts in a multiethnic country: A retrospective population-based study. *Sci Rep*, **11**(1): 7556. https://doi.org/10.1038/s41598-021-87229-4

Yuan SH, Qiu Z, Ghosh A (2009). *TOX3* regulates calcium-dependent transcription in neurons. *Proc Natl Acad Sci U S A*, **106**(8): 2909–2914. https://doi.org/10.1073/pnas.0805555106

Zhu Y, Miao H, Zeng Q, Li B, Wang D, Yu X *et al.* (2021). Prevalence of cleft lip and/or cleft palate in Guangdong province, China, 2015-2018: A spatio-temporal descriptive analysis. *BMJ Open*, **11**(8): e046430. https://doi.org/10.1136/bmjopen-2020-046430