Genetic alterations associated with cleft lip and palate patients as revealed by whole-exome sequencing and bioinformatics

Rashidov Rustam Abdurasul ugli

The Graduate School

Yonsei University

Department of Dentistry

Genetic alterations associated with cleft lip and palate patients as revealed by whole-exome sequencing and bioinformatics

Directed by Professor Jae Hoon Lee, D.D.S., M.S.D., Ph.D.

The Doctoral Dissertation

submitted to the Department of Dentistry

and the Graduate School of Yonsei University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Dental Science

Rashidov Rustam Abdurasul ugli

June 2021

This certifies that the Doctoral Dissertation
of Rashidov Rustam Abdurasul ugli is approved.

_____

Thesis supervisor: Jae Hoon Lee

_____

Thesis committee: Sung Won Cho

_____

Thesis committee: Dong Wook Kim

_____

Thesis committee: Sanguk Kim

_____

Thesis committee: Ji Hyun Lee

The Graduate School
Yonsei University
June 2021

ACKNOWLEDGEMENT

First and foremost, praise and thanks to Allah for His showers of blessings throughout my research to complete the study successfully.

I want to express my deep and sincere gratitude to my supervisor, Professor Jae Hoon Lee D.D.S., M.S.D., Ph.D. Professor of the prosthodontics department, Yonsei University, College of dentistry, guided me to do my research successfully and provided invaluable guidance throughout this research. His dynamism, foresight, honesty, and motivation have deeply inspired me. He has taught me to know myself deeper and look forward to the future as a researcher and dentist, at the same time to be beneficial for society, always search for the truth as a human being. It was a great opportunity and honor to work and study under his guidance.

Besides my supervisor, I would like to thank rest of my thesis committee Prof. Sung Won Cho, Prof. Dong Wook Kim, Prof. Sanguk Kim, Prof. Ji Hyun Lee, for their encouragement, insightful recomedations.

I appreciate the fiancial support of the “El-yurt umidi” foundation, which covered all my expenses for three years.

My sincerence thanks also goes to professors of Dental College, especially, Department of Prosthodontics, Prof. Keun Woo Lee, Prof. Hong Seok Moon, Prof. June Sung Shim, Prof. Jee Hwan Kim, Prof. Jong Eun Kim, Prof. Ji-Man Park, Prof. Kyung Chul Oh who gave me their

advices to improve my knowlegde, and Prof. Khabilov Nigman Lukmanovich (Tashkent state dental institute), who teached me for a long time and helped to become a prosthodontics. To Prof. Babaev Abdukadyr (Research institute of hematology and blood transfusion) who support in laboratory work in Uzbekistan. Dr. Lee, Kwanghwan, who support our research project.

I would like to say thank you to my friends who provided stimulating discussions as well as happy distractions to rest my mind outside of my research Dr Ami, Dr Enkhjargal, Dr. Farkhod, Dr. Bakhtiyor, Dr.Begzod and of course my roommate Dr. Kamoliddin.

Finally, I could not have completed this dissertation without the support of my family, I did it with their love, prayers, caring and supporting me spiritually throughout mylife.

June, 2021

Rustam Rashidov

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Genetic alterations associated with cleft lip and palate patients as revealed by whole-exome sequencing and bioinformatics

Rashidov Rustam Abdurasul ugli D.D.S., M.S.D.

Department of Dentistry

The Graduate School, Yonsei University

(Directed by Professor Jae Hoon Lee, D.D.S., M.S.D., Ph.D.)

Purpose: Cleft lip and/or palate (collectively referred to as orofacial cleft) is one of the most common orofacial birth defects. Extensive research has shown that orofacial cleft is a multigenic disease and involves both genetic and environmental factors. The primary purpose of this study was to determine candidate gene variants associated with orofacial clefts in the Uzbekistan population by whole-exome sequencing (WES) and statistical analyses.

Materials & Methods: Eight members (two affected and six unaffected) of a single family were included in the study. Saliva was collected from each family member

for WES (Agilent SureSelect Human All Exon 50 Mb). Saliva samples were collected for massive parallel sequencing, and the SnpEff from the 1000 Genomes Project was used for the filtering and annotation of samples from unaffected family members.

Results: A total of 47,290 variants were detected. Using a bioinformatics approach, we identified candidate gene variants related to cleft lip and palate in the Uzbekistan population. In the two affected members, 19 candidate genes (25 gene variants) were identified. A STRING network analysis revealed gene interactions involving candidate genes related to cleft lip and palate, including *HLA-DQA1* and *IRF6*.

Conclusion: This is the first application of WES and a bioinformatics approach to determine the genetic etiology of cleft lip and palate in the Uzbekistan population. This was a pilot study involving a single family aimed at clarifying the genetic factors contributing to the pathophysiology of orofacial disease. The candidate genes and interactions discovered in our study improve our understanding of cleft lip and palate in the Uzbekistan population.

Keywords: cleft lip and palate, whole-exome sequencing, bioinformatics, candidate genes, gene ontology

Genetic alterations associated with cleft lip and palate patients as revealed by whole-exome sequencing and bioinformatics

Rashidov Rustam Abdurasul ugli D.D.S., M.S.D.

Department of Dentistry
The Graduate School,
Yonsei University

(Directed by Professor Jae Hoon Lee, D.D.S., M.S.D., Ph.D.)

I. INTRODUCTION

Orofacial clefts (OFCs) form a group of structural mutations in the oral cavity. They can extend to the face, resulting in oral and facial deformations (Yaqoob et al., 2013). OFCs are the leading congenital disorders in Uzbekistan, with a frequency of approximately 0.28–1.32 per 1000 live births (Inoyatov, 2014).

They are divided into cleft lip and/or cleft palate (CL/P). CL/P is associated with complications affecting feeding, discourse, hearing, and psychological development (Stanier and Moore, 2004).

Different classification systems for OFCs have been suggested in the surgical and dental literature. These systems are primarily divided into anatomical systems, which are useful for surgeons, and embryology-based systems, which are useful for genetic counseling and research. The distinct OFC data types utilized in surgery, genetic counseling, and research limit the development of a universally acceptable and useful classification system.

The etiology of CL/P has been studied in industrialized countries and is heterogeneous, with increasing evidence that these anomalies are multifactorial. Although genetic and environmental triggers are vital for syndromic CL/P, the etiology of the more common non-syndromic (isolated) forms remains poorly understood (Yaqoob et al., 2013).

To understand the molecular and developmental mechanism underlying OFCs, a comprehensive review of the literature is essential. Causative factors can be broadly grouped into environmental and genetic factors (Seto-Salvia and Stanier, 2014). Thorough scientific reviews accompanied by using relevant diagnostic investigations have identified numerous causes of CL/P inside the gift examination. The presence of 18% of Genetic or environmental causes have been reported in

18% of cases in studies of Europe and the East, while the cause of CL/P has not been identified in 82% of cases, consistent with other estimates (Seto-Salvia and Stanier, 2014). The set of environmental factors influencing cranial development differ depending on several factors, such as maternal age, usage of medicines (such as anti-epileptic pharmaceuticals or corticosteroids), or cigarette and alcohol consumption during pregnancy. Exogenous factors together with drugs, cigarette smoking, maternal diabetes, alcohol consumption at some point during pregnancy, and hyperthermia at some point in early pregnancy, substantially increase the chance of cleft formation in developing embryos (Yaqoob et al., 2013).

With the emergence of genomics, there has been significant progress in the identification of the causative hereditary transformations in basic syndromic forms of CLP (Dixon et al., 2011).

The detection of genes associated with CL/P has been the focus of extensive research using various techniques (e.g., linkage analyses, genomic rearrangements, candidate gene approaches, and genome-wide association studies) (Leslie and Marazita, 2013).

Recent advances in genomic technologies provide an opportunity to sequence every exon by whole exome capture and massively parallel DNA sequencing, referred to as whole-exome sequencing (WES). Considering that protein-coding regions contain 85% of known mutations associated with diseases

but account for only 1% of the human genome, WES is an effective tool for detecting pathogenic mutations (Aylward et al., 2016). WES detects an underlying genetic cause in 25–35% of children with an unexplained suspected genetic disorder (e.g., birth defect) after negative karyotype and chromosomal microarray test results (Petrovski et al., 2019). In this study, WES was used to identify candidate gene variants associated with OFCs in the Uzbekistan population.

II. MATERIALS AND METHODS

1. Ethics Statement

All research involving human subjects or human data was approved by the Institutional Review Board of the Ethics Committee of Ministry of Healthcare, Republic of Uzbekistan, (protocol number “5” 575.113.3:611.716.2:616.315-007.254). All clinical investigations were performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants.

2. Patient Selection

A family was selected based on the observation of more than one affected member with a cleft lip and palate. Affected members were observed and underwent surgical procedures at the Tashkent State Dental Institute’s Clinic, Pediatric Maxillofacial Surgery Department. Two members (Fig. 1,2) of the family were affected by cleft lip and palate, and six members were not affected.

Fig. 1: Affected patient (3-3)
(third generation in the family pedigree)

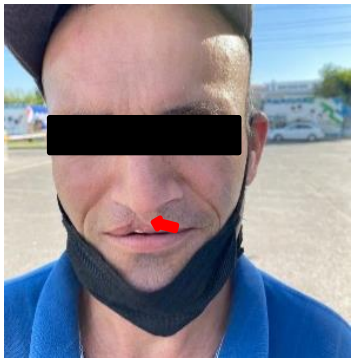

Fig. 2: Affected patient (2-3)
(second generation in the family pedigree)

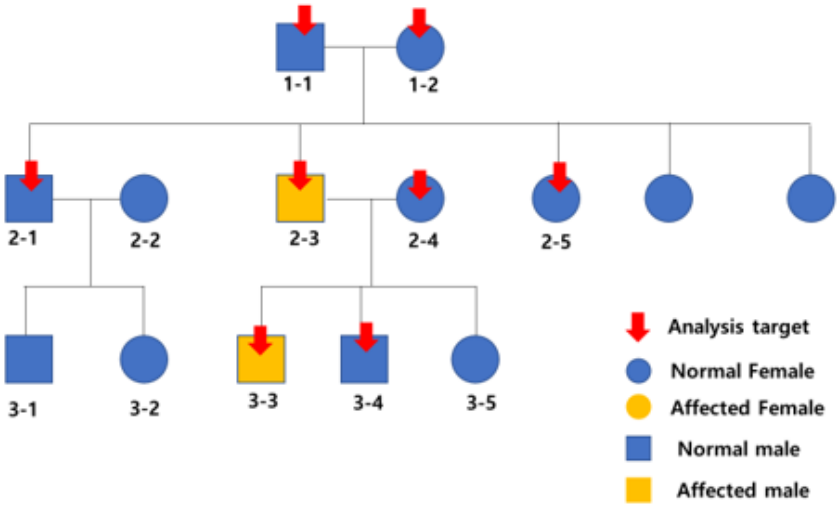

Table I. Patient pedigree. The analysis targets (indicated by red arrows) span three generations. The two affected members are in the second and third generations in the pedigree.

3. Comparing Data Sets

dbSNP138, GRCh37/hg19, and the 1000 Genomes Project (population ALL) were used for comparison with data for unaffected family members. Common variants derived from all data sets were used to filter out common variants from the WES data for the case population. This group had no history of diseases known to affect CL/P.

4. Sample Collection

To obtain DNA data, 2 mL of saliva was collected from all family members who participated in the study using the Oragene DNA Self-Collection Kit (DNA GenoTek, Ottawa, Ontario, Canada; Cat. #OG-500) in a tube containing 2 mL of DNA-preserving solution. After saliva collection, the lid was closed to release the DNA-preserving solution into the saliva sample. The samples in Uzbekistan was delivered to the DNA Link Laboratory (Seoul, South Korea) for DNA extraction and processing.

5. Whole-exome sequencing using the HiSeq 2500 platform

To ensure that the DNA was of sufficient quality, 1% agarose gel electrophoresis and the PicoGreen® dsDNA Assay (Invitrogen, Carlsbad, CA, USA) were employed. A OD260/280 ratio of between 1.8 and 2.2 indicated that the DNA had a high purity.

A Bravo Automated Liquid Handler and SureSelect sequencing libraries were used following the manufacturer's recommendations. In 120 mL of EB buffer, 1 μg of

genomic DNA was fragmented to a median size of 150 bp using a Covaris-S2 (Covaris, Woburn, MA, USA) with the following settings: duty cycle set to 10%, intensity at 5, cycles per burst set to 200, and frequency sweeping with a voltage amplitude of 4 for 360 s. To verify the effectiveness of fragmentation, capillary electrophoresis on DNA 100 chips was used (Bioanalyzer; Agilent, Santa Clara, CA, USA). Following the manufacturer's instructions (Agilent) and with the assistance of a secondary adapter, adapters were ligated to the DNA fragments, followed by PCR amplification. To confirm the integrity of the PCR products, capillary electrophoresis (Bioanalyzer, Agilent) was performed. Briefly, Reagents #1, #2, #3, and #4 (Agilent) were combined to obtain the hybridization buffer. They were concentrated in a solution of 500 ng in 3.4 μL. SureSelect block reagents #1, #2, and #3 (Agilent) were combined with 500 ng of DNA and incubated for 5 min at 65℃. For extension, the hybridization buffer and DNA blocker mix were heated from 65℃ to 95℃. The samples were incubated at 95℃ for 5 min. For a faster extension, the samples were incubated at 65℃ for 10 min. In addition, a blocking reagent (RNase) was added to the SureSelect oligo capture library (Agilent), followed by incubation for 2 min at 65℃. After incubation for 24 h at 65℃, the capture library was mixed with the hybridization buffer and the DNA blocker solution and incubated for an additional 4 h at 65℃ in a thermal cycler. A solution of 200 mL of SureSelect binding buffer (Agilent) was used to wash 50 μL of streptavidin coating the Dynal MyOne Streptavidin T1 (Invitrogen) three times. Following a 30-min incubation, the hybridization mixture was added to the bead suspension for an additional 30 min, and the mixture was incubated for an additional 30

min. Following the recommended instructions, the beads were washed with water. The beads were washed with 500 mL of SureSelect Wash Buffer #1 (Agilent) for 15 min at room temperature, followed by three washes with 500 mL of SureSelect Buffer #2 (Agilent) for 10 min at 65℃. Following elution, the sample was allowed to sit in 50 mL of SureSelect Elution Buffer (Agilent) at room temperature for 10 min. Agilent SureSelect Neutralizing Buffer (50 mL) was added to the eluted DNA. The reaction product was purified using AMPure XP beads (Beckman, Brea, CA, USA). To create the capture library, the library was amplified using the Agilent Bioanalyzer and Hercules II fused DNA polymerase was used to check the quality of the amplified libraries (Finnzymes, Espoo, Finland). The six libraries were mixed into a single equimolar mixture using SYBR Green PCR Master Mix and were amplified by PCR in equimolar quantities (Applied Biosystems, Waltham, MA, USA). Using the cBot automated cluster formation system (Illumina), a cluster creation process was automatically started in the flow, and the flow cell was loaded on the HiSeq 2500 sequencing system (Illumina) for sequencing with a 2 × 101 bp read length.

6. Whole-exome sequencing and variant analysis

Variants were identified based on WES results. Only variants with call rates greater than 80% were considered. A pedigree analysis of the family suggested an autosomal recessive inheritance pattern. Variants that were homozygous in the affected family members but absent or heterozygous in unaffected members were identified.

Additional filters were applied to exclude variants annotated as LOW and MODIFIER with respect to impact using SnpEff. In total, 25 variants were retained for subsequent analyses.

7. Functional analysis of gene variants using ClueGO

To clarify the connection between the candidate gene variants and disease, a systematic and integrative functional enrichment analysis was performed using Cytoscape v3.8.2 and ClueGO v2.5.7 (a Cytoscape pug-in). ClueGO v2.5.7 can be used to identify and functionally categorize significant Gene Ontology (GO) terms, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways, and BioCarta pathways for gene lists. A feature of ClueGO is the ability to combine GO terms and KEGG/BioCarta pathways to generate a functional network. It can be used to analyze a single list of genes or to compare two lists of genes and to thoroughly visualize functionally related terms (Bindea et al., 2009)

III. RESULTS

Eight samples were evaluated by WES, including samples from three females and five males, two of whom had CL/P (see Table I for a pedigree). The pedigree suggests the possibility of an autosomal recessive pattern of inheritance. An average of 46,852 reads and 5.87 megabases of data were obtained by WES for eight individuals. A total of 47,290 variants were detected. SnpEff was used to eliminate variants that did not alter the protein sequence as well as to predict the effects of each variant on gene function, has done segregation analysis, f162ollowed by searches for variants that are homozygous in affected members and heterozygous in unaffected family members, with no homozygosity observed in unaffected members. Filtering was then performed using SnpEff, where genes predicted to have weak effects were eliminated (after classification as high, moderate, low, or modifier). Genes with high and moderate impacts were retained after SnpEff filtering. After filtering, 19 genes (25 gene variants) were included in subsequent analyses, as listed in Table II.

Gene	SNP	CHR	REF	ALT	Transcript ID	AA change	1000G Maf
C1orf167	rs4845880	1	G	A	ENST00000433342	p.Arg544Gln	0.6708
	rs4845881	1	G	A	ENST00000433342	p.Arg571Gln	0.6735
	rs6697244	1	G	T	ENST00000433342	p.Ser848Ile	0.6433
	rs4846043	1	G	A	ENST00000433342	p.Arg944His	0.6438
LGR6	rs788795	1	T	C	ENST00000367278	p.Val592Ala	0.6154
LRRN2	rs11588857	1	G	A	ENST00000367175	p.Pro692Ser	0.1914
	rs3747631	1	G	C	ENST00000367175	p.Leu518Val	0.1914
	rs3789044	1	G	A	ENST00000367175	p.Pro7Leu	0.1900
TSSC1	rs7595702	2	C	A	ENST00000443925	p.Gly181* (stop gain)	0.5343
TRAPPC12	rs4971514	2	G	C	ENST00000416918	p.Ala7Pro	0.4267
ATXN1	rs16885	6	G	A	ENST00000244769	p.Pro753Ser	0.1200
SLC17A2	rs2071299	6	G	A	ENST00000265425	p.Pro437Ser	0.4258
BTN3A2	rs9358936	6	A	G	ENST00000356386	p.Asn181Asp	0.0723

	rs2072803	6	G	C	ENST00000432533	p.Ala255Pro	0.0852
HLA-DQA1	rs12722051	6	A	T	ENST00000343139	p.Tyr48Phe	0.1896
CLPSL2	rs2478467	6	C	T	ENST00000360454	p.Arg79Cys	0.6983
VNN3	rs764263	6	G	C	ENST00000417437	p.Pro74Ala	
SYTL3	rs3123101	6	T	A	ENST00000297239	p.Leu587Gln	0.5325
NOD1	rs2075820	7	C	T	ENST00000222823	p.Glu266Lys	0.2972
SLC18A1	rs2270637	8	C	G	ENST00000276373	p.Ser98Thr	0.1891
PLXDC2	rs3817405	10	G	A	ENST00000377252	p.Val396Ile	0.5577
ANKRD26	rs2274741	10	A	T	ENST00000436985	p.Phe1530Leu	0.3022
	rs10829163	10	C	T	ENST00000436985	p.Val1321Ile	0.3022
PMFBP1	rs16973716	16	T	G	ENST00000537465	p.Lys918Asn	0.4299
VSX1	rs6138482	20	C	T	ENST00000376707	p.Arg217His	0.2647

Table II. Autosomal recessive candidate genes and variants. Columns show the following: SNP, single nucleotide polymorphism; CHROM, chromosome; POS, base-pair position; REF, reference allele, ALT, alternative allele; AA, amino acid change; MAF, minor allele frequency based on the 1000 Genome Project

Protein–protein interactions involving HLA class II histocompatibility antigen, DQ alpha 1 chain (HLA-DQA1) and interferon regulatory factor 6 (IRF6) were detected, as shown in Fig. 3, using the STRING and Cytoscape databases.

Irf6 was originally linked to Van der Woude syndrome; later genome-wide association studies demonstrated that *Irf6* mutations are associated with non-syndromic CL/P (Dai et al., 2015).

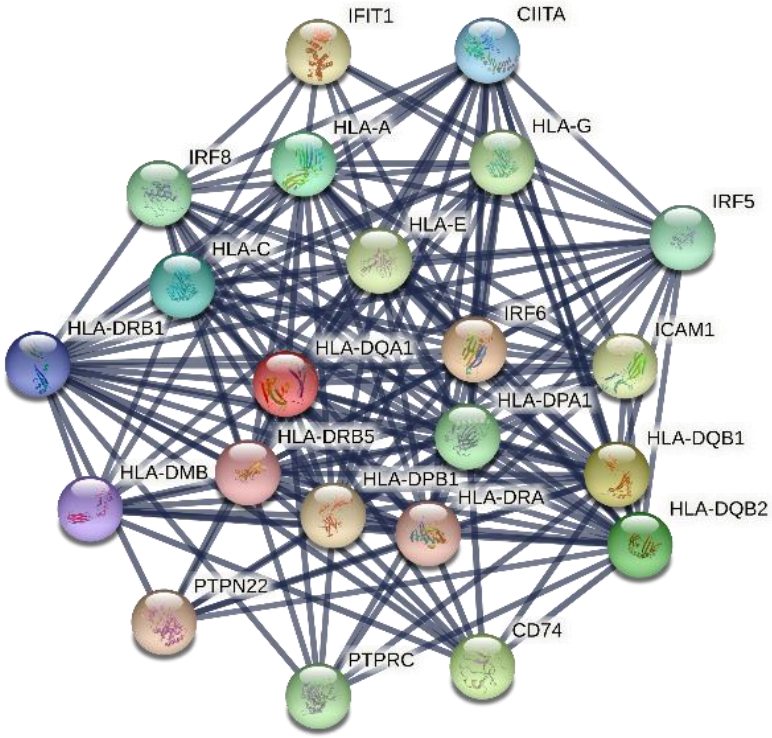

Fig. 3. STRING network for HLA-DQA1 and IRF6.

We conducted a systematic and integrative analysis of the functions of candidate genes using Cytoscape v3.8.2 and ClueGO v2.5.7. Table III summarizes the results of the ClueGO analysis. Similarity in GO terms among genes was evaluated by kappa scores. In total, 162 functional terms involved in KEGG and BioCarta pathways were identified.

GO terms	162
Ontology source:	GO: biological process GO: molecular function GO: cellular component Kegg pathway
Average P-value across terms	1.49E-05
Average Bonferroni-corrected P-value across terms	6.47982E-05
Average P-value across groups	2.94E-06
Average Bonferroni-corrected P-value across groups	4.74923E-06
GO groups	25
Genes with associations (%)	14.22

Table III. Summary of ClueGO plug-in results.

The most significant GO term was positive regulation of MAPK cascade (GO:0043410; P = 0.000661617, Bonferroni-corrected P = 0.000661617). In total, 15 genes were associated with this GO term (5.08%).

GO ID	GO Term	P-value	Bonferroni-corrected P-value
GO:0043410	positive regulation of MAPK cascade	0.000661617	0.000661617
GO:0043410	positive regulation of MAPK cascade	0.000661617	0.000661617
GO:0000226	microtubule cytoskeleton organization	0.000301141	0.000903422
GO:0019787	ubiquitin-like protein transferase activity	0.000240921	0.000963683
GO:0010942	positive regulation of cell death	6.2789E-05	0.000502312
GO:0043068	positive regulation of programmed cell death	5.74057E-05	0.000516651
GO:0023061	signal release	5.42087E-05	0.000542087
GO:0046903	Secretion	4.98296E-05	0.000548125
GO:0043065	positive regulation of apoptotic process	4.92206E-05	0.000590647

Table IV. List of the 10 most significant GO terms

There were 25 GO groups. Group 7 was the most significant (P = 0.00024092, Bonferroni-corrected P = 0.000240921). This group involved the GO term ubiquitin-like protein transferase activity (GO: 0019787), and 15 genes were associated with this GO term (5.60%).

GO ID	GO Term	GO Group	P-Value	Bonferroni-corrected P-Value
GO:0019787	ubiquitin-like protein transferase activity	Group 07	0.00024092	0.000240921
GO:0010942	positive regulation of cell death	Group 14	6.2789E-05	0.000125578
GO:0043068	positive regulation of programmed cell death			
GO:0043065	positive regulation of apoptotic process			
GO:0000226	microtubule cytoskeleton organization	Group 11	1.6601E-05	4.9803E-05
GO:0005819	spindle			
GO:0005773	vacuole	Group 15	3.248E-06	1.2992E-05
GO:0000323	lytic vacuole			
GO:0005764	lysosome			

Table V. List of the 10 most significant GO groups

To better understand known and predicted protein–protein interactions, we obtained 19 candidate genes (25 gene variants) and 325 nearest neighbor gene variants in the STRING network (344 total genes). The results of the functional analysis are shown in Figure 4.

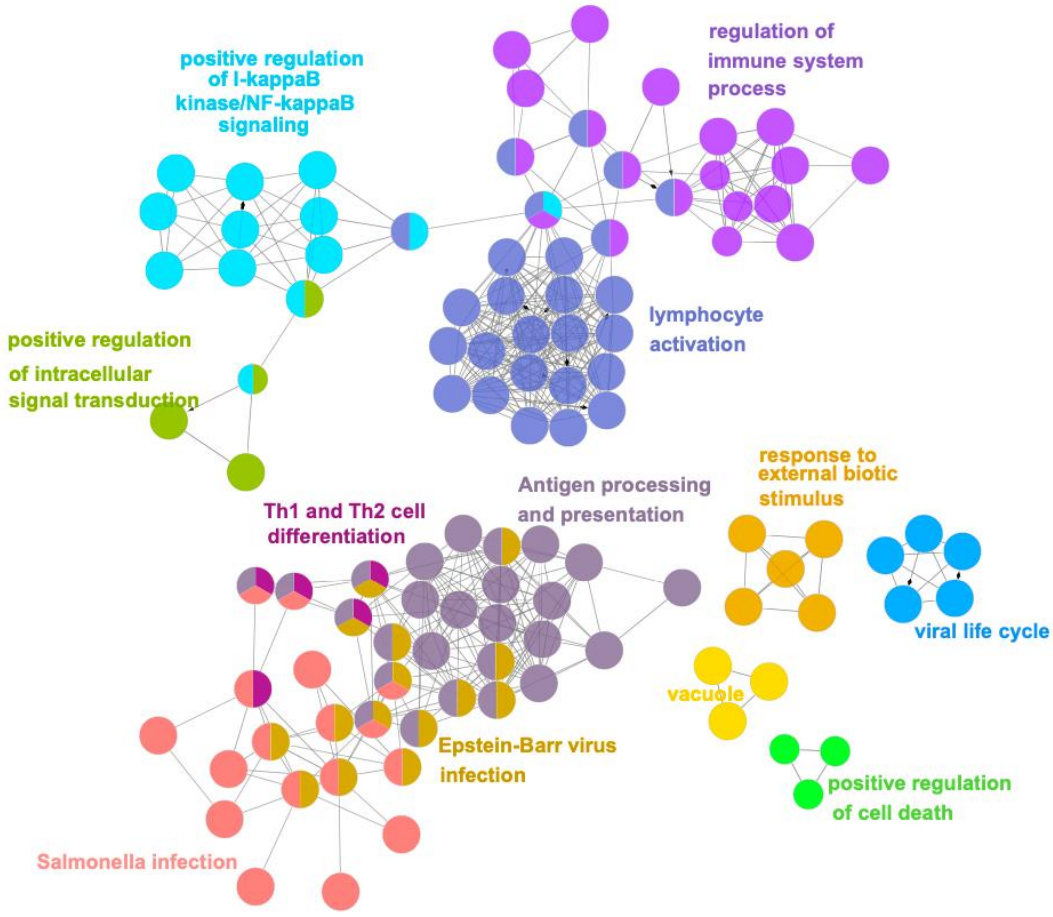

Fig. 4. Network view of functional interactions. In total, 344 genes were included in the functional analysis.

IV. DISCUSSION

This is the first application of WES and bioinformatics analyses to determine the molecular basis of CL/P in the Uzbekistan population. Using a single family, 25 candidate gene variants were identified by WES, a segregation analysis, and variant filtering.

Several variants were related to CL/P based on previous studies, including *LGR6* and *C1orf167*. In previous study of *LGR6* (Leucine-rich repeat-containing G-protein coupled receptor 6), a candidate gene expressed in saliva glands, disease samples collected from patients with CL/P deformities had the same *LGR6* gene variants as those in affected members of the family (Zhang et al., 2017). *C1orf167* (Uncharacterized protein C1orf167) is involved in mandibular and maxillary development (Genno et al., 2019). It is hypothesized that *LGR6* and *C1orf167* gene variants affect CL/P in the Uzbekistan population.

CL/P is a genetic disorder caused by multiple genetic and environmental factors. A mutation in one or more candidate genes may disrupt orofacial development, resulting in CL/P. OFCs are among the most common birth defects, and the biological mechanisms underlying this multifactorial disease are still unknown, especially in the Uzbekistan population. Various genetic tools, including linkage analyses and genome-wide association analyses, can be used to identify genomic susceptibility regions, including multiple causative genes and chromosomal loci responsible for the pathogenesis of the disease. It is possible that interactions between environmental and genetic variables during pregnancy result in clefts. This also provides a reasonable explanation for why in many families with

numerous afflicted members across many generations, transmission is more typically the result of genetic "carriers." Here, the carrier harbors the same causative mutation; however, the trait does not have a corresponding phenotypic effect, either due to a better uterine environment or a protective genetic background. Accordingly, the mechanism underlying CL/P is complex and hence the probability that it is passed to offspring is difficult to determine. Therefore, we cannot pinpoint a small number of gene variants as the exact cause of the disease. However, functional enrichment analyses provide insight into related genes.

The expression of a single gene may be associated with several parameters during the physiological or pathological processes. Conventional biology research has typically focused on one or a few genes at a time (Dai et al., 2014). Rather than focus on single genes, we considered candidate gene regulators associated with diseases, biological processes, or signaling pathways to better understand the link between complex gene networks and complicated hereditary diseases.

Relational protein–protein interactions and protein–DNA interactions may both exist in the same gene regulatory network (Dai et al., 2014). In this study, we utilized the STRING database to predict protein–protein interactions involving known genes and other possible network components. The STRING network revealed interactions between the candidate gene *HLA-DQA1* (major histocompatibility complex, class II, DQ alpha 1) with *IRF6* (interferon regulatory factor 6). A large number of genes, including *IRF6*, have established roles in the formation of CL/P (Dai et al., 2015).

Owing to the vast number of genes that may be assayed in high-throughput studies, beyond our ability to process, numerous methods have been used to pool data and generate biological information in a systematic manner (Bindea et al., 2009).

For a systematic analysis, Cytoscape was used in this study. Our functional enrichment analysis of candidate genes demonstrated a connection with the bifunctional enzymes TH1 and TH2. Previous studies have suggested a change towards the Th1 differentiation pathway activation in CL/P. Certain underlying processes in the cleft are correlated with Th1 cell development and the constitutive production of IL-2, 6, 13, and TNF-α. Additionally, the transition from Th1 to Th2 cell differentiation is a necessary for a healthy pregnancy. Unbalanced placental and embryonic development, with significant abnormalities, may result from the preferential activation of the Th1 pathway (Pilmane et al., 2021).

Our study was limited by the lack of availability of a genome database for the Uzbekistan population for comparison.

V. CONCLUSION

WES and bioinformatics approaches were used for the first time to evaluate the molecular basis for OFCs in the Uzbekistan population based on a single family. Previous studies have frequently characterized genetic disorders by karyotyping affected patients. In this study, 19 genes (25 gene variants) were identified as novel variants associated with OFCs in the Uzbekistan population. Two candidate genes, *LGR6* and *C1orf167*, matched previously reported genes. STRING analyses showed a relationship between the *HLA-DQA1* candidate gene and the previously identified *IRF6*. These data provide basic information related to the pathogenesis of CL/P and may guide precision medicine.

REFERENCES

Aylward A, Cai Y, Lee A, Blue E, Rabinowitz D, Haddad J, Jr., et al. (2016). Using Whole Exome Sequencing to Identify Candidate Genes With Rare Variants In Nonsyndromic Cleft Lip and Palate. *Genet Epidemiol* 40(5): 432-441.

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25(8): 1091-1093.

Dai J, Mou Z, Shen S, Dong Y, Yang T, Shen SG (2014). Bioinformatic analysis of Msx1 and Msx2 involved in craniofacial development. *J Craniofac Surg* 25(1): 129-134.

Dai J, Yu H, Si J, Fang B, Shen SG (2015). Irf6-Related Gene Regulatory Network Involved in Palate and Lip Development. *J Craniofac Surg* 26(5): 1600-1605.

Dixon MJ, Marazita ML, Beaty TH, Murray JC (2011). Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet* 12(3): 167-178.

Genno PG, Nemer GM, Zein Eddine SB, Macari AT, Ghafari JG (2019). Three novel genes tied to mandibular prognathism in eastern Mediterranean families. *Am J Orthod Dentofacial Orthop* 156(1): 104-112 e103.

Leslie EJ, Marazita ML (2013). Genetics of cleft lip and cleft palate. *Am J Med Genet C Semin Med Genet* 163C(4): 246-258.

Petrovski S, Aggarwal V, Giordano JL, Stosic M, Wou K, Bier L, et al. (2019). Whole-exome sequencing in the evaluation of fetal structural anomalies: a prospective cohort study. *Lancet* 393(10173): 758-767.

Pilmane M, Jain N, Jain S, Akota I, Kroiča J (2021). Quantification of Cytokines in Lip Tissue from Infants Affected by Congenital Cleft Lip and Palate. *Children (Basel)* 8(2).

Seto-Salvia N, Stanier P (2014). Genetics of cleft lip and/or cleft palate: association with other common anomalies. *Eur J Med Genet* 57(8): 381-393.

Stanier P, Moore GE (2004). Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Hum Mol Genet* 13 Spec No 1: R73-81.

Yaqoob M, Mahmood F, Hanif G, Bugvi SM, Sheikh MA (2013). Etiology and genetic factors in clefts of lip and/or palate reported at children's hospital, Lahore, Pakistan. *Indian J Hum Genet* 19(2): 136-143.

Zhang C, Li Y, Zhang XY, Liu L, Tong HZ, Han TL, et al. (2017). Therapeutic potential of human minor salivary gland epithelial progenitor cells in liver regeneration. *Sci Rep* 7(1): 12707.

요약

WES 및 생물 정보학에 의해 밝혀진 구순열 및 구개열 환자와 관련된 유전적 변이

Rashidov Rustam Abdurasul ugli, D.D.S., M.S.D.

(이재훈 교수, D.D.S., M.S.D., Ph.D.)

목적 : 구순열 및 / 또는 구개열은 가장 흔한 구강 안면 선천적 결함 중 하나다. Ora - facial clefts 는 cleft lip only, cleft lip and palate or cleft plate only 를 포함한다. 오랜 시간 동안 구개열 및 구순열의 원인에 대해 연구되어 왔으며 다유전자성 질환은 환경과 동시에 유전적 요인에 의해 영향을 받을 수 있다. 이 연구의 주요 목적은 WES 를 통해 후보 유전자 변이를 찾은 다음 우즈베키스탄 인구에 대한 통계 분석을 하는 것이다.

재료 및 방법 : 한가족 중 8 명 (2 명은 질환자 6 명은 무질환자)을 연구 대상으로 선정하고 엑솜 시퀀싱 (WES) (Agilent SureSelect Human All Exon 50Mb 키트) 을 위해 OG-500 (DNA Genotek, Ottawa, Ontario, Canada)을 사용하여 각 가족구성원의 타액을 수집했다. 대규모 스크리닝을 위해 타액샘플을 채취했으며

SNPeff 1000 게놈 프로젝트, 모집단 ALL, 필터링을 위해 무질환 가족 구성원의 샘플을 사용했다.

결과 : 총 47,290 종의 변종이 검출되었다. 생물정보학 접근방식을 사용하여 우즈베키스탄 인구의 구순열과 구개열과 관련된 후보 유전자 변종을 식별했다. 영향을 받은 두 멤버에서 19 개 후보 유전자(유전자 변종 25 개)가 확인됐다. STRING 네트워크 분석 결과 HLA-DQA1, IRF6 등 구순구개열과 관련된 후보 유전자의 유전자 상호작용이 확인됐다.

결론: 우즈베키스탄에서 구순열과 구개열의 유전학적 특성을 파악하기 위한 전장엑솜분석과 생물정보학 접근방식이 처음으로 적용된 연구이. 이것은 한 가족이 참여한 파일럿 연구로, 치과 질환의 병리생리학에 기여하는 유전적 요인을 규명하는 것을 목표로 했다. 본 연구에서 발견된 후보 유전자와 상호작용은 우즈베키스탄 인구의 구순열과 구개열에 대한 이해를 향상시킬 것이다