



UNIVERSIDAD
NACIONAL
DE COLOMBIA

**Caracterización funcional de
variantes (SNV, Indels, CNVs)
analizadas por medios
computacionales en 5 familias
dominicanas con labio y/o paladar
hendido no sindrómico.**

Sandra Jhoana Ocampo Mahecha

Universidad Nacional de Colombia
Facultad de Medicina, Departamento de Morfología, Maestría en Genética
Humana
Bogotá, Colombia
2023

Caracterización funcional de variantes (SNV, Indels, CNVs) analizadas por medios computacionales en 5 familias dominicanas con labio y/o paladar hendido no sindrómico.

Sandra Jhoana Ocampo Mahecha

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

Magister en Genética Humana

Director (a):

M.Sc., director (a): Catalina María Arévalo Caro

PHD, codirector(a): Andrés Mauricio Pinzón Velasco

Línea de Investigación:

Línea de investigación en genética del crecimiento y malformaciones

Grupo de Investigación:

Grupo de Investigación en Crecimiento y Desarrollo Craneofacial

Universidad Nacional de Colombia

Facultad de Medicina, Departamento de Morfología, Maestría en Genética Humana

Bogotá, Colombia

2023

Cada uno de nosotros es una preciosidad, en una perspectiva cósmica. Si alguien discrepa de tus opiniones, déjalo vivir. En un trillón de galaxias, no hallarías otro igual. Debemos darnos cuenta del milagro que significa poder experimentar la vida.

Carl Sagan

Dedico mi trabajo en forma especial a mi madre Martha Yanet Mahecha, a mi abuela María Priscila Roa, a mi pareja Jerson Andrés Valencia y a mis hijas Tsuky y Pecas, quienes me han brindado apoyo, paciencia y mucho cariño.

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respecto de los derechos de autor. Este estudio es original, ha tomado como referencia ideas, figuras, protocolos y palabras de algunos autores, pero se ha reconocido su autoría en las referencias bibliográficas, aplicando las citaciones con normas Vancouver.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Sandra Jhoana Ocampo Mahecha

Fecha: 03.02.2023

Agradecimientos

Agradezco el apoyo y la guía en la construcción de este lindo estudio a mis asesores, la Dra. Catalina María Arévalo Caro y al Dr. Andrés Mauricio Pinzón Velasco, sin ellos no habría sido posible cumplir los objetivos.

Dra. Catalina, de nuevo quiero agradecerle por todo, siempre ha estado para lo que he necesitado, dando consejos de vida e impulsando a hacer cosas como presentar eventos en inglés, entre otras cosas, muchas gracias, siempre la voy a recordar y muy seguro estaré presta para lo necesite.

Reconozco el valor de los aportes de la Dra. María Guadalupe Silva de Reyes, docente de la Universidad Nacional Pedro Henríquez Ureña de República Dominicana, la Dra. Clara Eugenia Arteaga Díaz, docente de la Universidad Nacional de Colombia, la Dra. Janneth González Santos, docente de la Pontificia Universidad Javeriana, al Dr. William Usaquén y al grupo de genética de poblaciones de la Universidad Nacional de Colombia. Agradezco al Fondo Nacional de Innovación y Desarrollo Científico y Tecnológico (FONDOCyT) de República Dominicana. A los profesores María Hilde Torres Rivas y Rogelio Cordero por ser los co-creadores del acuerdo interinstitucional, a los pacientes, familiares y personal de apoyo de la “Fundación Niños que Ríen”.

Reconozco el apoyo durante la maestría de mis compañeros Diego Alejandro Saldaña y Ana Lorena Montealegre, les doy infinitas gracias.

Agradezco también la colaboración de la secretaria de la maestría Mireya y al profesor Mauricio Rey y a mis profesoras de clínica en especial a las Dras. Luz Yaqueline Ladino Cortes y Sandra Yaneth Ospina Lagos.

Resumen

Caracterización funcional de variantes (SNV, Indels, CNVs) analizadas por medios computacionales en 5 familias dominicanas con labio y/o paladar hendido no sindrómico.

Introducción: El labio y/o paladar hendido es una anomalía congénita, la más frecuente a nivel facial. Puede tener una etiología multifactorial o genética. En República Dominicana y en particular la región de Cibao, se identifica una alta frecuencia diagnóstica de hendiduras orofaciales; sin embargo, son pocos los estudios alrededor de la anomalía.

Objetivo: Determinar la estructura y funcionalidad de variantes (SNV, Indels, CNVs) asociadas con labio y/o paladar hendido no sindrómico en cinco familias dominicanas con herencia autosómica dominante, a través de secuenciación masiva y análisis computacional.

Metodología: Se reclutaron cinco familias dominicanas mediante la base de datos de la Fundación Niños que Ríen, a quienes se les realizó árboles genealógicos con datos clínicos y poblacionales sobre tres generaciones en cada familia. Se hizo una revisión clínica sistémica y orofacial, análisis de genoma completo en trío mediante el uso de tecnología Illumina por medio de la plataforma HiSeqX con profundidad promedio de 30X. Se tuvieron en cuenta elementos de la historia clínica y los criterios de clasificación de variantes para la definición final. Se hizo un análisis integral mediante herramientas bioinformáticas a las variantes relacionadas a labio y/o paladar hendido no sindrómico con el fin de entender sus mecanismos biológicos con relación a su funcionamiento y al desarrollo de fisuras orofaciales en los seres humanos, mediante modelamiento de proteínas, construcción de red de genes y acoplamiento molecular.

Resultados: Se encontraron ocho variantes relacionadas con riesgo de enfermedades; solo una de ellas fue asociada al desarrollo de fisuras orofaciales: una variante tipo CNV en el gen del factor 6 regulador de interferón (IRF6) con clasificación patogénica identificada en una familia afectada. Los factores ambientales identificados no fueron concluyentes. Se crearon tres modelos de IRF6, uno del dominio a unión a ADN sin cambios patogénicos, otro de la variante con pérdida de los primero tres exones y un último de la proteína completa. Mediante herramientas bioinformáticas se obtuvo un modelo de acoplamiento ADN-Dominio unión ADN de IRF6, no publicado en la literatura previamente. Se construyó una red de genes relacionados a IRF6 y el desarrollo de labio y/o paladar hendido no sindrómico.

Conclusiones: Se detectó una variante tipo CNV consistente en una deleción de una copia de los exones 1-3 del gen IRF6 asociada a fisuras orofaciales no reportada previamente en la literatura. Se creó un modelo de acoplamiento de ADN-proteína con IRF6 no publicado en la literatura científica. Se hizo un modelo nuevo de red de genes relacionado a IRF6 y labio y/o paladar hendido no sindrómico. Por otro lado, se resalta la importancia de la realización de pruebas moleculares para ofrecer asesoramiento genético a las familias y pacientes.

Palabras Claves: labio hendido, paladar hendido, no sindrómico, secuenciación del genoma completo, IRF6, simulación computacional, simulación acoplamiento molecular, red de genes.

Abstract

Functional characterization of variants (SNV, Indels, CNVs) analyzed by computational means in 5 Dominican families with non-syndromic cleft lip and/or palate.

Introduction: Cleft lip and/or palate is a congenital anomaly, the most frequent at the facial level. It may have a multifactorial or genetic etiology. In the Dominican Republic and in particular the Cibao region, a high diagnostic frequency of orofacial clefts is identified; however, there are few studies around the anomaly.

Objective: To determine the structure and functionality of variants (SNV, Indels, CNVs) associated with non-syndromic cleft lip and/or palate in five Dominican families with autosomal dominant inheritance, through massive sequencing and computational analysis.

Methodology: Five Dominican families were recruited through the database of the Niños que Ríen Foundation, to whom family trees were made with clinical and population data on three generations in each family. A systemic and orofacial clinical review was made, complete genome analysis in trio using Illumina technology through the HiSeqX platform with average depth of 30X. Elements of the clinical history and the variant classification criteria were taken into account for the final definition. A comprehensive analysis was made using bioinformatic tools of the variants related to non-syndromic cleft lip and/or palate in order to understand their biological mechanisms in relation to their functioning and the development of orofacial clefts in humans, through protein modeling. Gene network construction and molecular docking.

Results: Eight variants related to disease risk were found; only one of them was associated with the development of orofacial clefts: a CNV-type variant in the Interferon Regulatory Factor 6 (IRF6) gene with a pathogenic classification identified in an affected family. The environmental factors identified were not conclusive. Three models of IRF6 were created, one of the DNA-binding domain without pathogenic changes, another of the variant with loss of the first three exons, and a last one of the complete protein. Using bioinformatic tools, a DNA-DNA binding domain coupling model of IRF6 was obtained, not previously published in the literature. A network of genes related to IRF6 and the development of non-syndromic cleft lip and/or palate was constructed.

Conclusions: A CNV-type variant consisting of a deletion of one copy of exons 1-3 of the IRF6 gene associated with orofacial clefts not previously reported in the literature was detected. A model of DNA-protein docking with IRF6 not published in the scientific literature was created. A new model of gene network related to IRF6 and non-syndromic cleft lip and/or palate was made. On the other hand, the importance of performing molecular tests to offer genetic counseling to families and patients is highlighted.

Keywords: cleft lip, cleft palate, non-syndromic, whole genome sequencing, IRF6, computer simulation, molecular docking simulation, gene network.

Contenido

	Pág.
Resumen	VII
Lista de figuras	XIV
Lista de tablas	XVI
Lista de símbolos y abreviaturas	XVII
Planteamiento del problema	XX
Justificación	XXII
1. Objetivo general	XXV
1.1 Objetivos específicos.....	XXV
2. Marco teórico	28
2.1. Generalidades de las hendiduras orofaciales.....	28
2.2. Microformas.....	30
2.3. Guías usadas en la identificación de hendiduras orofaciales y malformaciones.....	33
2.4. Epidemiología del labio y/o paladar hendido.....	34
2.5. Etiología de las hendiduras orofaciales.....	34
2.6. Mecanismos embriológicos y genéticos relacionados con labio y/o paladar hendido	36
2.7. Genes relacionados con labio y/o paladar hendido	39
2.8. Factor 6 regulador del interferón.....	41
2.9. Estudios moleculares relacionados con labio y/o paladar hendido.....	44
2.10. Aproximaciones computacionales.....	46
2.10.1Análisis de genoma completo.....	46
2.10.2 Herramientas bioinformáticas para análisis de red de genes.....	49
2.10.3 Herramientas para análisis estructural de proteínas.....	50
2.10.4. Docking molecular.....	56
3. Metodología	62
3.1. Descripción general.....	62
3.2. Fases del estudio.....	63
3.2.1. Fase 1 preparatoria.....	63
3.2.2. Fase 2 identificación de familias y toma de muestras.....	63
3.2.3 Fase 3 Secuenciación e identificación de variantes.....	66
3.2.4 Fase 4 Análisis e interpretación de variantes identificadas.....	69
3.2.4.1 Ensamblaje de genoma completo y llamado de variantes.....	69

3.2.4.2 Análisis estructural y funcional de variantes encontradas en LPH-NS.....	70
3.2.4.2.1 Modelamiento de proteínas mediante herramientas computacionales.....	70
3.2.4.2.2 Acoplamiento molecular.....	71
3.2.4.2.3 Red de genes relacionada a IRF6 y al desarrollo de fisuras orofaciales no sindrómicas.....	71
3.3 Análisis estadístico	73
4. Resultados	79
4.1 Caracterización de la muestra.....	79
4.1.1 Componente sociodemográfico.....	79
4.1.2 Componente ambiental.....	80
4.1.3 Componente clínico.....	80
4.2 Microformas.....	82
4.3 Descripción general de cada familia con su respectiva genealogía informacional.	84
4.4 Caracterización molecular	97
4.5 Análisis estructural y funcional de variantes encontradas en L/PH-NS... 104	
4.5.1 Modelo unión a ADN de IRF6.....	104
4.5.1.1 Verificación modelo de dominio de unión a ADN en IRF6.....	106
4.5.2 Modelo dominio de unión a ADN de IRF6 con pérdida de los exones 1-3.....	111
4.5.3 Modelo ISRE.....	112
4.5.4 Modelo de proteína completa de IRF6.....	113
4.5.4.1 Verificación del modelo de la proteína IRF6.....	115
4.6 Ejecución docking.....	119
4.6.1 Resumen modelos docking.....	120
4.6.2 Resumen de interacciones modelos docking.....	123
4.6.3 Comparación experimentos previos IRF interacción dominio de unión a ADN y ADN.....	130
4.7 Red de genes.....	135
4.8 Segregación de variantes asociadas con L/PH-NS.....	140
5. Discusión	141
5.1 Caracterización de las familias.....	141
5.2 Variante patogénica en IRF6.....	143
5.3 Expresividad variable en el gen IRF6	147
5.4 Correlación entre microformas con L/PH-NS.....	150
5.5 La red de genes propuesta.....	152
5.6 Secuenciación de nueva generación.....	160
5.7 Limitaciones.....	160
6. Proceso de divulgación de resultados	162
7. Recomendaciones	163
8. Conclusiones	165
8.1 Financiación	165
9. Bibliografía	166

Lista de figuras

	Pág.
Figura 2-1: Estructura IRF6.....	41
Figura 2-2: Diagrama esquemático de ColabFold.....	55
Figura 2-3: Clasificación del software de acoplamiento molecular.....	56
Figura 3-1: Resumen metodológico del estudio.....	82
Figura 4-1: Percepción de estrato socioeconómico.....	80
Figura 4-2: Frecuencia de presencia o ausencia de hendiduras orofaciales.....	81
Figura 4-3: Genealogía familia 1.....	86
Figura 4-4: Información genealógica poblacional de la familia 1.....	86
Figura 4-5: Genealogía familia 2.....	88
Figura 4-6: Información genealógica poblacional familia 2.....	89
Figura 4-7: Genealogía familia 3.....	90
Figura 4-8: Información genealógica poblacional familia 3.....	91
Figura 4-9: Genealogía familia 4.....	92
Figura 4-10: Información genealógica poblacional familia 4.....	93
Figura 4-11: Fenotipo madre de la probando de la familia 5.....	94
Figura 4-12: Fenotipo de la probando de la familia 5.....	95
Figura 4-13: Genealogía familia 5.....	96
Figura 4-14: Información genealógica poblacional familia 5.....	97
Figura 4-15: Modelo dominio unión ADN de IRF6 obtenido por AlphaFold2.....	105
Figura 4-16: Topología del dominio de unión a ADN.....	106
Figura 4-17: Gráfico de Ramachandran.....	108
Figura 4-18: Gráfica ERRAT.....	109
Figura 4-19: Superposición estructural. 7OOT/IRF6.....	110
Figura 4-20: Gráfica pLDDT de modelado de la proteína IRF6 en AlphaFold2.....	111
Figura 4-21: Modelo de dominio de unión a ADN con pérdida del exón 3 en IRF6.....	112
Figura 4-22: Región ADN denominada ISRE.....	113
Figura 4-23: Modelo de IRF6 obtenido por AlphaFold2.....	114
Figura 4-24: Gráfica de Ramachandran del modelo de la proteína IRF6.....	116
Figura 4-25: Gráfica ERRAT.....	117

Figura 4-26	Descripción topológica de la proteína IRF6.....	118
Figura 4-27	Modelo de la proteína IRF6.....	119
Figura 4-28	Mapa de contacto de residuos de las predicciones complejo ADN-Proteína de las cadenas A y B de IRF6.....	121
Figura 4-29	Mapa de contacto de residuos de las predicciones complejo ADN-Proteína de la cadena con la delección del exón 3.....	122
Figura 4-30	Interacción de residuos Lys 109 y Trp40 (B), complejo dominio de unión a ADN de IRF6 y ADN.....	124
Figura 4-31	Interacción de residuos Cys 85 y Lys 89 (B), complejo dominio de unión a ADN de IRF6 y ADN.....	125
Figura 4-32	Interacción de residuos Asn88 (B), complejo dominio de unión a ADN de IRF6 y ADN.....	126
Figura 4-33	Interacción de residuos Asn88, Arg 84 y Lys 109 (A), complejo dominio de unión a ADN de IRF6 y ADN.....	127
Figura 4-34	Interacción de residuos Asn88 (A), complejo dominio de unión a ADN de IRF6 y ADN.....	128
Figura 4-35	Interacciones hidrofóbicas de algunos residuos como Lys89 y Cys85(B).....	129
Figura 4-36	Interacciones en el surco menor.....	130
Figura 4-37	Mapas de contacto de residuos de los modelos experimentales por cristalografía de rayos X, complejo proteína-ISRE (IRF2-IRF 3).....	133
Figura 4-38	Mapas de contacto de residuos de los modelos experimentales por cristalografía de rayos X, complejo proteína-ISRE (IRF4).....	134
Figura 4-39	Red de genes IRF6- L/PH-NS.....	136
Figura 4-40	Resumen anotación de genes relacionada a enfermedad, DAVID Data Base.....	137
Figura 4-41	Resumen anotación de componentes celulares y función molecular, DAVID Data Base	138

Lista de tablas

	Pág.
Tabla 2-1: Resumen características de algunas microformas.....	30
Tabla 2-2: Resumen herramientas computacionales empleadas para acoplamiento macromolécula- macromolécula.....	57
Tabla 3-1: Ejemplo de anonimización de datos.....	75
Tabla 3-2 Operacionalización de las variables del estudio.....	78
Tabla 4-1 Frecuencias tipo de microforma.....	82
Tabla 4-2 Resumen datos sociodemográficos, ambientales y clínicos de las cinco familias.....	82
Tabla 4-3 Descripción de variantes encontradas en el estudio.....	98
Tabla 4-4 Reclasificación de variantes encontradas no relacionadas con fisuras orofaciales.....	100
Tabla 4-5 Evaluación de la estructura por MolProbity de la estructura ISRE.....	113
Tabla 4-6 Resumen plantilla empleada para el acoplamiento HDOCK.....	120
Tabla 4-7 Resumen métricas de evaluación HDOCK de modelos ejecutados...	121
Tabla 4-8 Resumen experimentos previos estructuras por cristalografía de algunos miembros de la familia IRF.....	131
Tabla 4-9 Resumen anotación tributo de genes enlistados, DAVID Data Base..	139

Lista de Símbolos y abreviaturas

Símbolos

Símbolo	Término
A	Alanina
ADA2	Adenosina desaminasa 2
ANKRD62P1-PARP4P3	Lectura completa de ANKRD62P1-PARP4P3, pseudogen transcrito
BARX1	homeobox similar 1
B362CT	Beta 3-glucosiltransferasa
BMP	Proteína morfogenética ósea
C	Cisteína
CCT8L2	Chaperonin Containing TCP1 Subunit 8 Like 2
CECR	Síndrome del ojo de gato Región cromosómica
CHD7	Cromodominio Helicasa Proteína de unión a ADN 7
CTNND1	Catenin Delta 1
CYP21A1	Cytochrome P450 Family 21 Subfamily A Member 2
D	Ácido Aspártico
dA	Desoxiadenosina
dC	Desoxicidina
dG	Desoxiguanosina
Dkk	Proteína relacionada con Dickkopf
dT	Timidina
DUOX2	Oxidasa dual 2
FGF	Factor de crecimiento de fibroblastos
Foxf	Forkhead-Related Transcription Factor
GAB4	Miembro de la familia de proteínas de unión asociadas a GRB2 4
GRHL3	Grainyhead Like Transcription Factor
H	Histidina
HDHD5	Haloácido deshalogenasa como dominio de hidrolasa que contiene 5
HBB	Hemoglobina Subunidad Beta

HSFY1P1	Región cromosómica del síndrome del ojo de gato, candidato 8
IKKA	Inhibitor of Nuclear Factor Kappa-B Kinase Subunit Alpha
ICS	Secuencia de consenso de interferón
IL17RA	Interleukin 17 Receptor A
IRF	Factor regulador de interferón
IRF-E	Elemento de unión a IRF
ISRE	eElemento de respuesta estimulado por IFN
JAG2	Jagged Canonical Notch Ligand 2
LINC0	Interleukin 17 Receptor A
M	Metionina
ml	Milímetro
MN1	Probable proteína supresora de tumores MN1
MSX	Homeobox Protein MSX
MTHFR	Metilentetrahidrofolato reductasa
NOTCH	Notch Receptor
Osr	Odd-Skipped Related Transcription Factor
P	Prolina
P21	Inhibidor de quinasa dependiente de ciclina
PAX	Paired Box Protein Pax
Q	Glutamina
R	Arginina
S	Serina
SFN	Estratifina
SHH	Molécula de señalización Sonic Hedgehog
SHOX2	Short Stature Homeobox
SMPD1	Esfingomielina Fosfodiesterasa 1
SNAI2	Represor transcripcional de la familia del caracol 2
TFAP2A	Transcription Factor AP-2 Alpha
TMEM121B	Transmembrane Protein 121B
TPTEP1	Endogenous Retroviral Protease Pseudogene
Spry	Sprouty RTK Signaling Antagonist
TBX22	T-Box Transcription Factor
TGFB	Factor transformante Beta
TP63	Proteína turmo P63
µL	Microlitro
ZFHX4	Zinc Finger Homeobox 4
ZO-1	Proteína de unión estrecha 1
W	Triptófano

TGFA	Factor de crecimiento transformante alfa
VUS	Variantes de significado incierto
SVW	síndrome de Van Der Woude
WGS	Secuenciación de genoma completo

Planteamiento del problema

Las anomalías congénitas representan un problema de salud a nivel mundial, la OMS reporta 270.000 recién nacidos que mueren cada año durante las primeras semanas de vida por alguna de estas causas.¹ El labio y/o paladar hendido (L/PH) es una anomalía congénita frecuente, tiene una incidencia mundial de 1 en 700 nacidos vivos (NV),² la prevalencia varía según la etnia, por ejemplo, es mayor en asiáticos y nativos americanos (1 de cada 500 NV), seguida por la población europea e hispana (1 en 1000 NV) y más baja en africanos, (1 de cada 2500 NV).³

Con relación a la etiología del L/PH son varias las causas señaladas, estas incluyen factores ambientales, infecciosos, epigenéticos y genéticos.⁴⁻⁶ La Asamblea Mundial de la Salud adoptó una resolución en el 2010 que integró el fortalecimiento en registro y vigilancia de los defectos congénitos, además de robustecer los conocimientos y abrir campo en la investigación sobre estos temas;⁷ sin embargo, en algunos países aún no se tiene un registro epidemiológico ni programas fuertes en anomalías congénitas como es el caso de República Dominicana, donde se llevó a cabo el estudio, no cuenta con estadísticas claras sobre L/PH, lo que evidencia la necesidad de ahondar más sobre las hendiduras orofaciales.

Existen algunos reportes sobre L/PH asociados a misiones humanitarias en República Dominicana, con referencia a protocolos de cirugía y a la descripción de procesos en dicha anomalía, sin datos estadísticos ni epidemiológicos relevantes;^{8,9} así mismo, otros estudios divulgados en revistas dominicanas de carácter descriptivo,¹⁰ pero no cuenta con estudios publicados relacionados al campo molecular.

En el 2015 y 2016 se realizó un trabajo descriptivo en la región de Cibao (región del norte de República Dominicana), donde se encontró que las familias de los pacientes con labio y/o paladar hendido no sindrómico (L/PH-NS) pertenecían a clases sociales vulnerables, con bajo nivel educativo e imposición en estos territorios para el uso de productos agro-tóxicos como pesticidas, elementos teratogénicos que ponen en riesgo a las poblaciones en contacto. ¹¹

Por las razones señaladas se desarrolló este estudio en las regiones del Cibao y Santo Domingo, el cual se enfocó en analizar el genoma completo en trío, en cinco familias dominicanas con L/PH-NS con herencia autosómica dominante, analizar por medios computacionales las proteínas, las vías de señalización biológicas y metabolómicas en las variantes encontradas en las familias dominicanas afectadas con dicha anomalía, todo ello con la finalidad de sumar conocimientos relacionados a L/PH en la región y en el país.

Justificación

El L/PH es una anomalía que requiere gran atención en salud por las consecuencias y secuelas que tiene a largo plazo, tanto físicas como psicológicas, en especial si no se trata de forma adecuada y temprana, por ello la importancia de entender sobre dicha patología y en especial dilucidar la etiología en aquellas regiones donde se ha reportado de forma frecuente. El bajo reporte de estudios y la alta incidencia de la malformación en la región del Cibao y Santo Domingo, en República Dominicana ha motivado al desarrollo de esta investigación.

La Fundación Niños que Ríen ubicada en la región del Cibao atiende de forma anual a una gran cantidad de pacientes con la anomalía, lo que hace evidente la importancia de entender sobre dicha entidad. En el 2017 se realizó un estudio descriptivo en pacientes con L/PH a través de encuestas y construcción de árboles genealógicos en el que se encontraron varios patrones de herencia y una prevalencia similar de la malformación con referencia a Centroamérica, con ello aparece un proyecto enfocado en el campo de la genética, en el cual identificaron algunas familias con la enfermedad mencionada con herencia autosómica dominante,¹² por lo que en el 2018 se escogió una de ellas y se realizó un estudio de exoma completo sin resultados positivos asociados con la enfermedad (Arteaga

C., et al. Limitations of the whole exome sequencing in the genetic characterization of nonsyndromic cleft lip with or without cleft palate. República Dominicana.2021 [no publicado]); para el 2019-2020 nació la propuesta de análisis de genomas completos en familias dominicanas afectadas con dicha malformación, con la meta de determinar las variantes (SNV, Indels, CNVs) asociadas a L/PH-NS y entender el funcionamiento de dichas variantes en los diferentes procesos biológicos implicados en el desarrollo de las malformaciones craneofaciales específicamente de las hendiduras orofaciales de carácter no sindrómico.

Las pruebas de secuenciación de nueva generación como el genoma completo han revolucionado el estudio genómico, permitiendo obtener resultados a un menor tiempo y costo, así como una gran cantidad de información,¹³ todo ello se ha dado gracias a los avances tecnológicos en otras áreas como la bioinformática, la bioquímica y la física que de forma coordinada han contribuido a su progreso.¹⁴ Con este estudio se quiso ampliar el conocimiento sobre los mecanismos biológicos relacionados al desarrollo de L/PH-NS, mediante el uso de herramientas computacionales para comprender el componente estructural y funcional de las variantes encontradas relacionadas con la anomalía.¹⁵

El proyecto tenía, además, el propósito de visibilizar la importancia del campo genómico y genético con relación a L/PH-NS, mostrando la necesidad de generar información científica sobre esta anomalía que amplíe la visión y la comprensión sobre ella en el país, que pueda contribuir en un futuro a la creación de programas

en salud pública de promoción y prevención, relacionados a enfermedades genéticas y congénitas.

Sumado a ello, es poca la evidencia científica alrededor de L/PH en países centroamericanos. Este estudio pretende liderar estas temáticas y aportar al conocimiento con el uso de nuevas tecnologías genómicas, así como también contribuir al entendimiento de las causas de la patología en la región.

Debido a lo anterior, los resultados de este proyecto pueden contribuir al Plan Estratégico de Ciencia y Tecnología e Innovación que se está iniciando en la región que tiene como propósito promover componentes científicos, éticos, sociales, económicos y ambientales para impactar de forma positiva en la calidad de vida del pueblo dominicano.¹⁶

1. Objetivo general

Determinar la estructura y funcionalidad de variantes (SNV, Indels, CNVs) asociadas con L/PH-NS en cinco familias dominicanas con herencia autosómica dominante, a través de secuenciación masiva y análisis computacional.

1.1 Objetivos específicos:

1. Identificar cinco familias con herencia autosómica dominante con L/PH-NS, que cumplan con los criterios de inclusión.
2. Determinar variantes (SNV, Indels, CNVs) asociadas con L/PH-NS, en cinco familias dominicanas que presenten la malformación con patrón de herencia autosómico dominante.
3. Analizar por medios computacionales las implicaciones funcionales de las variantes identificadas en procesos biológicos previamente relacionados con L/PH-NS.
4. Determinar la segregación de variantes asociadas con L/PH-NS mediante el análisis de tríos en familias dominicanas con la malformación.

2. Marco Teórico

2.1. Generalidades de las hendiduras orofaciales

Las anomalías congénitas son trastornos que afectan a la población en general. Según la OMS, “son definidas como cualquier alteración estructural anatómica, morfológica o bioquímica que se desarrolla durante la gestación y se identifica en el nacimiento o después de este”.¹⁷ Se pueden clasificar en anomalías mayores cuando hay repercusión severa en la morbimortalidad y en la expectativa de vida, y, en anomalías menores cuando hay una afectación leve sin repercusión grave, generalmente de carácter estético. Estas anomalías pueden presentarse de forma aislada como un apéndice periauricular o estar representadas en una forma sindrómica donde hay varios hallazgos clínicos que se asocian en una patología.¹⁸ La mayoría de las hendiduras orofaciales son de aparición aislada es decir no sindrómicas que corresponde aproximadamente al 70% en comparación a las sindrómicas que equivalen a un 30% que se asocian a otro tipo de anomalías.¹⁹

El L/PH es una anomalía congénita, la más común a nivel craneofacial, puede afectar solo el labio o el paladar, también pueden afectarse ambos de forma unilateral o bilateral. Se han encontrado otros subfenotipos como paladar hendido

submucoso, insuficiencia velofaríngea y unas formas más leves a las que se les denomina microformas. ² La Academia Americana de Anomalías Craneofaciales (ACPA) ²⁰ instauró pautas para definir y clasificar las fisuras orofaciales, según la parte anatómica comprometida (labio, paladar), la ubicación (unilateral, bilateral, aislado), la profundidad (completa o incompleta) y la asociación (sindrómica y no sindrómica).²⁰

Se han propuesto una gran cantidad de clasificaciones para las hendiduras orofaciales, una de ellas se planteó en el año 1922 por Davis et al, ²¹ los autores indican tres grupos en donde usaban el proceso alveolar como línea divisora. Un primer grupo incluía las fisuras que afectan el labio y este a su vez las dividía en unilateral, bilateral y mediana. Un segundo grupo donde la estructura comprometida es el paladar y lo subdividía si se presentaba en el paladar blando o duro y finalmente un tercero que involucra el proceso alveolar, para este caso podría ser unilateral, bilateral y mediano. ²² Otra proposición de clasificación fue dada por Spina ²³ en 1973, en la que utiliza al foramen incisivo como referencia para la división en cuatro grupos, el grupo I incluía el preforamen, el II el transforamen, el III el posforamen y el grupo IV con raras hendiduras de la cara no relacionadas con el foramen incisivo. ²⁴

2.2. Microformas

Las microformas son las formas leves de las hendiduras orofaciales y son el resultado de la expresividad variable de algunos genes, ²⁵ por lo que se pueden

manifestar con anomalías orales o faciales menores, su presencia en una familia aumenta el riesgo de aparición de L/PH-NS en los miembros ²⁶. Se han reportado casos con los genes: la proteína morfogenética ósea 4 (BMP4) y factor 6 regulador del interferón (IRF6). ²⁷ Dentro de las microformas se pueden encontrar: agenesia de los dientes incisivos laterales, labio fisurado microformado, úvula bífida, paladar hendido submucoso, arco maxilar en forma de V, anomalías dentales como oligodoncia, entre otros, en la tabla 2-1 se explican las características generales de algunas de ellas.

Tabla 2-1 Resumen características de algunas microformas.

Microforma	Foto	Definición	Referencia
<p>Torus Palatino</p>		<p>Protuberancia o exostosis que se encuentra en forma localizada en la región bucal.</p>	<p>Antoniades et al. ²⁸</p>
<p>Úvula bífida</p>		<p>Muesca posterior de la úvula hasta una duplicación completa de dicha estructura.</p>	<p>Warthon et al. ²⁹</p>

<p>Anomalías dentales (Agenesia del incisivo lateral superior)</p>		<p>Ausencia de incisivo lateral superior debido a un desorden durante el proceso de formación de la lámina dentaria que la causa.</p>	<p>Emidio et al. ¹⁹</p>
<p>Anomalías dentales (Dientes supernumerarios).</p>		<p>Hiperdoncia o desarrollo de dientes extras de la serie normal.</p>	<p>Oropeza et al. ³⁰</p>
<p>Labio fisurado microformado</p>		<p>Hendidura labial leve que implica una discontinuidad en la unión del bermellón-cutáneo.</p>	<p>Amarillas et al. ³¹</p>
<p>Bóveda palatina profunda.</p>		<p>Bóvedas palatinas altas u ojivales en forma de "U".</p>	<p>Solano et al. ³²</p>

Arco maxilar en forma de V.



Compresión maxilar con forma de arco en V, asociado a mordida cruzada posterior bilateral y desviación de la línea media dentaria inferior.

Agurto et al. ³³

Morfología de los incisivos laterales superiores (cónicos).



Anomalía de la forma de los dientes con aspecto cónico.

Carmona et al. ³⁴

Paladar submucoso



Es una membrana mucosa, formada en los músculos del paladar blando.

Angeles et al. ³⁵

* Es de autoría propia, se emplearon fotografías de los pacientes que hacen parte del trabajo de investigación "Caracterización de microformas asociadas a labio y/o paladar hendido no sindrómico en familias que acuden a la Fundación Niños que ríen, Moca, República Dominicana". ³⁶

2.3 Guías usadas en la identificación de hendiduras orofaciales y malformaciones

Las malformaciones congénitas son alteraciones estructurales o funcionales que se desarrollan durante el embarazo y son evidentes en el nacimiento e incluso muchas veces de forma prenatal, la identificación de las anomalías son clave para el diagnóstico de muchas enfermedades que pueden aparecer de forma aislada o en conjunto con otras anomalías. Para la identificación de signos y síntomas que orienten hacia una anomalía congénita, es importante tomar en cuenta guías, protocolos e incluso libros con definiciones en relación a la clínica y componente dismorfológico, por ejemplo, el libro *The Bedside Dysmorphologist*, que permite visualizar diferentes elementos clínicos y conceptos importantes sobre malformaciones congénitas,³⁷ se han empleado guías, tales como la guía de detección de anomalías congénitas del recién nacido del Ministerio de Salud y Protección Social – Colciencias 2013 de Colombia,³⁸ y el manual para programas administrativos de vigilancia de defectos de nacimiento de la OMS 2015,³⁹ de donde se toman en cuenta definiciones y parámetros para el reconocimiento adecuado de anomalías congénitas, que cobra en conjunto un papel relevante en la determinación de pacientes con L/PH-NS.

2.4 Epidemiología del labio y/o paladar hendido

La prevalencia mundial del L/PH es de 1/700 a 1/1000 nacidos vivos,² siendo muy baja en afrodescendientes, intermedia en caucásicos, latinos y alta en asiáticos y nativos americanos; un estudio epidemiológico en fisuras orofaciales publicado en el 2015 lo confirma: se encontró para asiáticos (1,57), norteamericanos (1,56), europeos (1,55), oceánicos (1,33), suramericanos (0,99) y africanos (0.57) por

1,000 NV, respectivamente ⁴⁰. No se cuentan con datos epidemiológicos estadísticamente significativos en la región de Centroamérica ni en República Dominicana sobre L/PH.

2.5 Etiología de las hendiduras orofaciales

Con respecto al desarrollo craneofacial se debe tener en cuenta una compleja red de genes y factores de transcripción que participan en la migración, proliferación del mesénquima y progreso de cada estructura en la cara; el proceso clave inicia durante la embriogénesis a partir de la cuarta semana hasta las doce semanas. Luego de este periodo los cambios en el labio y el paladar son menores y dependen del crecimiento de la cabeza. ⁴¹ Durante la etapa de la morfogénesis facial es fundamental la atribución de los factores ambientales como cigarrillo, alcohol, plaguicidas, mala nutrición, medicamentos teratógenos, infecciones virales, los cuales repercuten en el código genético con la correspondiente aparición de variantes patogénicas, es decir, traducido en cambios puntuales, pérdidas o ganancias de datos en el ADN causantes de enfermedades. ⁴²

En las fisuras orofaciales se han descrito varios patrones de herencia, entre ellos el multifactorial que es el más relacionado, involucra tanto factores genéticos y ambientales. ⁴³ Las enfermedades monogénicas se asocian a otros patrones de herencia, con una menor frecuencia de aparición en relación al L/PH-NS, como es el caso de la herencia autosómica dominante y recesiva. ⁴⁴

A pesar de los múltiples estudios realizados en el campo molecular y epidemiológico son muchas las incógnitas relacionadas a este tema. ⁴⁵ La heredabilidad en las fisuras orofaciales ha sido difícil de determinar, ya que son varios los factores inmersos en su explicación, por lo que en este caso se podría inferir un concepto de heredabilidad perdida. ⁴³ En Colombia se publicó un artículo en 1998 en donde mediante el análisis de segregación, asociaciones de ligamiento genético y cálculos de heredabilidad en 60 individuos afectados con L/PH-NS, se encontró una heredabilidad del 96%, demostrando la presencia de un gen mayor, las familias estudiadas fueron relacionadas a genealogías con una herencia autosómica dominante. ⁴⁶

En estudios de casos y controles realizados en diferentes regiones del mundo han identificado los factores de riesgo y de protección relacionados con la aparición de las fisuras orofaciales; ejemplo de ello, un estudio publicado en la India en el 2017 evidencia el cigarrillo (OR 7,76, p 0,000, CI 95%) y el estrato socioeconómico bajo (OR 1,97, CL 95%, p 0,000) como factor de riesgo y el consumo de suplementos vitamínicos (OR 0,19, CI 95% p 0,003) como protector para el desarrollo de la malformación. ⁴⁷ Otro estudio llevado a cabo en Brasil en el 2020, encontró que el no consumo de ácido fólico en la gestación era un aspecto de riesgo (OR 3,94, CI 95% p 0,001) y la ingesta de analgésicos durante el embarazo era un factor protector (OR 0,39, CI 95%, p 0,001).⁴⁸ Otro artículo diseñado en China en el 2015

identificó varios factores con valores estadísticamente significativos, CI 95% y $p < 0,001$, describieron como elementos de riesgo durante el embarazo el alcohol (OR 9,39), el cigarrillo (OR 15,36), ser fumador pasivo (OR 4,05), la exposición a pesticidas (OR 7,84), el consumo de vitamina A (OR 16,46) y como factores protectores el consumo de ácido fólico (OR 0,18) y el de multivitamínicos (OR 0,18).⁴⁹

2.6 Mecanismos embriológicos y genéticos relacionados con labio y/o paladar hendido

Los estudios experimentales en embriones de animales como ratones y peces cebras han permitido aclarar el conocimiento sobre la formación del labio, el paladar primario y secundario en los seres humanos. El labio y el paladar primario provienen de las células de la cresta neural craneal y al mismo tiempo proceden del borde dorsal del tubo neural para formar durante la cuarta semana las prominencias frontonasales, pares de prominencias nasales mediales y laterales y de las prominencias maxilares.^{50,51} Las prominencias maxilares crecen de forma medial comprimiendo las prominencias nasales mediales hacia la línea media, hay un espacio que las separa, este se pierde a medida que aumentan de tamaño al punto de fusionarse y crear el segmento intermaxilar que da lugar al filtrum, al labio superior y al paladar primario, todo ello finaliza a la sexta semana de gestación. El paladar primario proviene del primer arco faríngeo que a la vez resulta de la migración de células de la cresta neural compuesta por los primeros dos rombómeros y algunas células del mesencéfalo.^{52,53} Este primer arco faríngeo

forma las prominencias mandibulares y maxilares, aproximadamente a la sexta semana de gestación, las cuales se crean a partir de las prominencias maxilares dos evaginaciones que crean las crestas palatinas. Estas crecen hacia abajo de lado a la lengua, finalmente toman una posición horizontal respecto a ella, las crestas constituyen los procesos palatinos laterales que pasan a un proceso de apoptosis a nivel de los bordes mediales con una posterior fusión, estableciendo en esta última etapa el paladar secundario. ^{50,51}

Las áreas de la biología del desarrollo y molecular han identificado el papel crucial de algunos morfógenos para la formación craneofacial, por ejemplo, el protooncogén Wnt (WNT), BMP y factor de crecimiento de fibroblastos (FGF) tiene un papel importante en la migración de las células de la cresta neural y en la formación de las prominencias faciales, demostrado en experimentos en ranas, peces y murinos. ^{54,55} No se cuenta con datos claros sobre genes y factores de transcripción relacionados al desarrollo de segmento premaxilar y paladar primario, pero sí sobre la formación del paladar secundario. El Sonic hedgehog (SHH) es clave en la señalización de los procesos palatinos laterales, inicialmente se expresa de forma uniforme y posteriormente se localiza a nivel de las rugosidades y las crestas epiteliales de los procesos palatinos. ⁵⁶ El SHH es regulado por los factores de crecimiento de fibroblasto, el FGF10 se expresa a nivel oral e induce la expresión del SHH, mientras el FGF7 ubicado en la región nasal lo reprime. El SHH regula la expresión de algunos factores de transcripción a nivel del mesénquima en los procesos palatinos laterales como Forkhead Box G 1 (Foxf1), Foxf2 y Odd-Skipped

Related Transcription Factor 2 (Osr2), los dos primeros genes juegan un papel importante en el crecimiento de las crestas palatinas y este último gen es clave como regulador en estas estructuras.⁵⁷ En la región anterior de los procesos palatinos laterales, se expresan algunos factores de transcripción como el homeobox Protein MSX-1 (MSX1), el short Stature Homeobox 2 (SHOX2). En la región posterior de los procesos laterales se expresan genes como: paired Box 9 (PAX9), homeobox Protein BarH-Like 1 (BARX1), T-Box Protein 22 (TBX22) y el MN1 Proto-Oncogene, Transcriptional Regulator (MN1); cuando hay alteraciones en estos patrones de expresión durante el desarrollo del paladar se correlacionan con defectos del desarrollo del paladar secundario, evidenciado en experimentos in vivo en ratones.^{56,58,59}

Para la palatogénesis es importante la diferenciación del epitelio, sus mecanismos celulares subyacentes han sido relacionados con la aparición de fisuras orofaciales y adherencias epiteliales orales. En dicho proceso se han implicado algunos genes como Jagged Canonical Notch Ligand 2 (JAG2), Notch Receptor 1 (NOTCH1), Grainyhead Like Transcription Factor 3 (GRHL3), Inhibitor Of Nuclear Factor Kappa-B Kinase Subunit Alpha (IKKA), Estratífina (SFN) y IRF6 en la diferenciación de las células del epitelio.⁶⁰ IRF6 y NOTCH son genes reguladores del factor de transcripción Tumor Protein P63 (TP63), que también promueve la diferenciación de esta estructura.⁵⁶ Los trastornos a nivel de la fusión de los procesos palatinos laterales se asocian al desarrollo de hendiduras orofaciales, en particular los pasos de ruptura del epitelio a nivel del borde medial de los procesos palatinos (apoptosis)

y la adhesión, estudios en murinos han demostrado el papel que tiene factor de crecimiento transformante Beta 3 (TGFB3) sobre el IRF6 para la fusión de los procesos palatinos laterales. El IRF6 está implicado en la represión de TP63 y en la activación indirecta del Inhibidor de quinasa dependiente de ciclina 1 (p21), todos ellos relacionados con la degeneración del borde epitelial de la línea media (MES) de los procesos palatinos, paso clave para la formación del paladar secundario.^{61,62}

2.7. Genes relacionados con labio y/o paladar hendido

Se han identificado una variedad de genes asociados al desarrollo de L/PH, gracias a estudios en el campo de la biología del desarrollo y molecular en animales que ha permitido extrapolar información en los seres humanos y entender mejor la etiología de esta anomalía, a continuación, se señalan algunos genes asociados en dicho proceso biológico. El IRF6 codifica proteínas asociadas a la formación de tejido conectivo, es un mediador del TGFB en la regulación de la transición mesenquimatosa epitelial y la fusión de los procesos palatinos laterales, también se relaciona con otras proteínas como el Snail Family Transcriptional Repressor 2 (SNAI2), JAG2 y diversos marcadores epiteliales como E-cadherina, plakofilina y Tight Junction Protein 1 (ZO-1),⁶³ que juegan un papel importante en el desarrollo del paladar en especial del paladar posterior.

La familia WNT codifica proteínas de señalización que participan en la regulación de la migración celular durante la embriogénesis, se ha señalado su conexión con

otros genes reguladores de la palatogénesis como el BMP, SHH, TGFB, FGF, Sprouty RTK Signaling Antagonist (SPRY), MSX1, MSX2 entre otros, todos interactúan como una red de genes de forma directa o indirecta, inhibiendo o estimulando en la expresión de sus proteínas codificantes. La WNT activa la señalización FGF, que a su vez modula al PAX9 y este último reprime la proteína Dickkopf-Related Protein (Dkk), un ligando inhibidor de LDL Receptor Related Protein (Lrp6) en la vía WNT, formando un bucle regulador de retroalimentación positiva durante el desarrollo orofacial; en caso contrario, el SPRY inhibe la señalización de FGF. ^{64,65} El MSX1, perteneciente a la familia de genes homeobox, codifica una proteína implicada en la represión transcripcional durante la embriogénesis a través de la interacción con otras homeoproteínas, participando en la regulación morfogénica del paladar, la dentición y el desarrollo craneofacial. ⁶⁶

Se ha propuesto que la Metilentetrahidrofolato reductasa (MTHFR) lleva a la disminución de los niveles de folato y por consiguiente altera la síntesis de ácidos nucleicos y con ello el desarrollo de fisuras orofaciales. ⁶⁷ El TBX22 es un gen que codifica una proteína que interviene en procesos durante la embriogénesis a nivel del mesodermo precordial, por ello participa en el desarrollo de tejidos especializados óseos y cartilagosos y ha sido identificado a nivel posterior de los procesos palatinos donde juega un papel importante en su desarrollo. ⁶⁸

2.8. El factor 6 regulador del interferón

El IRF6 hace parte de la familia de los factores reguladores del interferón, la cual tiene regiones proteicas similares como el dominio de unión a ADN con estructura hélice-giro-hélice N-terminal altamente conservado, constituido aproximadamente entre 100 a 120 aminoácidos y un dominio de unión a proteína C-terminal menos conservado, como se observa en la figura 2-1.⁶⁹

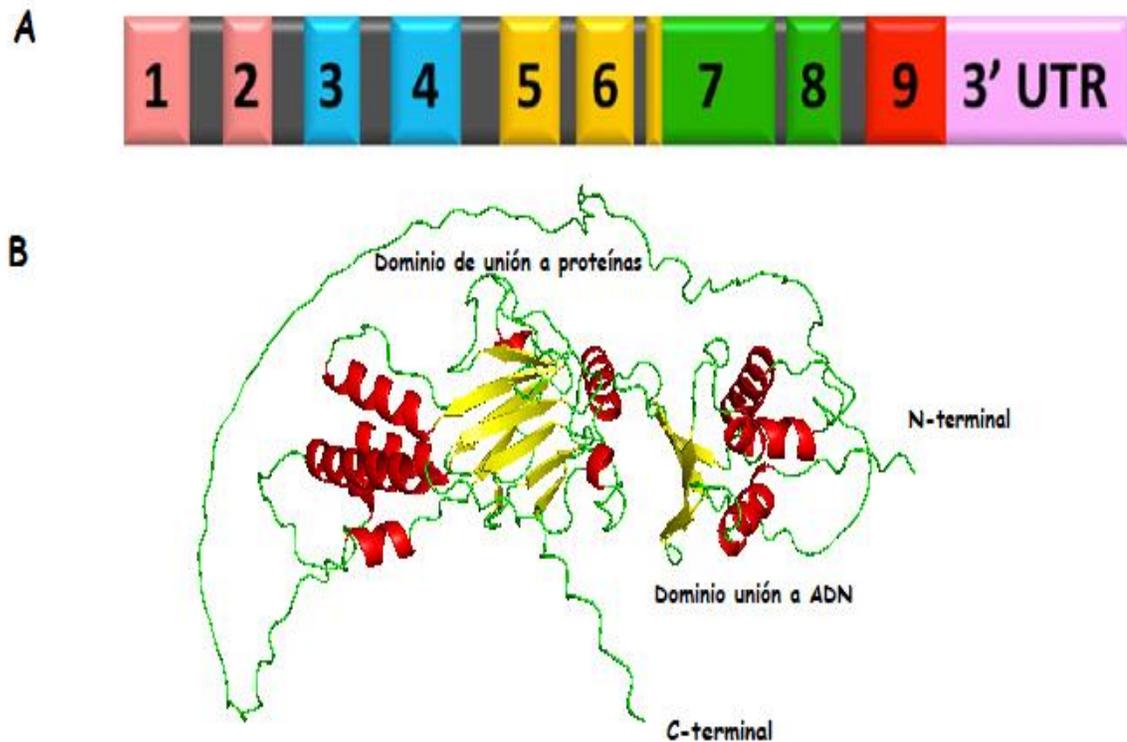


Figura 2-1. Estructura IRF6. Construida a partir de dos elementos, una figura del artículo “Toward an Orofacial Gene Regulatory Network”, elaborado por Kousa et al.⁷⁰ y de la estructura 3D de IRF6 obtenida por medios computacionales por AlphaFold y almacenada en la base de datos AlphaFold.⁷¹ **A.** Transcrito primario de IRF6 representado en una barra horizontal, cada número equivale a los nueve exones que lo conforman, cada color simboliza una región funcional de la proteína, los

exones 1 y 2 de color rosado constituyen la región no traducida 5' UTR. Los exones 3 y 4 son el dominio de unión al ADN en azul. En amarillo los exones 5 y 6 la región menos conservada. La mayoría del exón 7 y el exón 8 forman parte del dominio de unión a proteínas en verde y en rojo el exón 9 que contiene el dominio C-terminal.

⁷⁰ **B.** Estructura de predicción 3D de IRF6 obtenida por herramientas computacionales (AlphaFold), visualizada en PyMOL, ⁷² en rojo se representan las alfa hélices, en amarillo las láminas beta y en verde los bucles, se identifican los dominios de unión a ADN y proteínas.

Análisis previos en cristalografía de rayos X sobre la región de unión al ADN de IRF1 identificaron una arquitectura que contiene tres α -hélices antiparalelas de cuatro hojas plegadas β y tres bucles largos es una estructura similar a los dominios de unión al ADN que tiene los motivos hélice-giro-hélice señalados como reguladores de la expresión génica. ⁷³ Experimentos en IRF1 e IRF2 han evidenciado una interacción con el ADN a nivel de la tercera alfa hélice del motivo de unión a ADN en la secuencia consenso GAAA en el ADN. ⁷⁴ Esta secuencia consenso GAAA es clave en el reconocimiento y la unión del ADN con los miembros IRF, dicha secuencia se puede repetir tres, cuatro veces e incluso en algunas solo estar presente una vez. Las regiones en el ADN con estas secuencias de reconocimiento para los IRF se les ha denominado: elemento de unión a IRF (IRF-E), elemento de respuesta estimulada por IFN (ISRE) y secuencia de consenso de interferón (ICS). ^{73,74} Otro hallazgo compartido en los IRF son cinco residuos de triptófano conservados, de los cuales tres participan en el reconocimiento del ADN

a nivel del surco mayor, sin embargo, no es claro su mecanismo en dicha interacción.⁷⁵

2.9 Estudios moleculares relacionados con labio y/o paladar hendido

En el año 2012 Marazita⁷⁶ publicó el estado del arte de los estudios genéticos en L/PH, refiriendo que las divulgaciones científicas durante el siglo XIX alrededor de la anomalía se relacionaban a descripciones familiares, evidenció que para el siglo XX hubo un aumento de los estudios descriptivos y observacionales de las genealogías y patrones familiares de individuos afectados con L/PH, lo que llevó a cuestionar para ese tiempo una posible etiología genética, al identificar en ellas segregación familiar. Esto abrió camino a investigaciones en el campo de la genética, una publicación en el año 1989 por Ardinger et al.⁷⁷ en donde a partir de datos encontrados sobre palatogénesis en ratones, debatieron el papel de algunos genes en el desarrollo de L/PH en los humanos, por lo que estudiaron genes como el factor de crecimiento transformante alfa (TGFA) y el factor de crecimiento epidérmico e hicieron una comparación de casos de L/PH y controles con las frecuencias de doce fragmentos de restricción de longitud polimórfica, hallando asociación con el TGFA. Para el siglo XX fueron importantes las investigaciones en marcadores genéticos en genes candidatos, experimentando con microsatélites y polimorfismos; se dio importancia a estudios de ligamiento, PCR; finalmente surgieron nuevas pruebas como Sanger que dieron paso a la secuenciación de nueva generación y esta a su vez llevó a la aparición de pruebas genéticas más profundas y precisas como el exoma y el genoma completo para el siglo XXI.⁷⁸

En el 2009 se publicó el primer GWAS relacionado a L/PH-NS, en una cohorte pediátrica de ascendencia europea de casos y controles, identificando al locus 8q24 e infiriendo su papel en la palatogénesis.⁷⁹ Los GWAS realizados en los últimos 10 años han identificado múltiples genes y variantes relacionadas, por ejemplo, en el 2016 publicaron un GWAS multiétnico donde se encontraron variantes en siete genes y una región, la 8q24, asociados a la anomalía congénita;⁸⁰ en poblaciones latina, americana y europea, se halló un locus nuevo en el cromosoma 21 (21q22.2);⁸¹ en el 2008 se encontró variación en los genes WNT asociada a la malformación de carácter no sindrómico en EE. UU.⁸² También se han publicado estudios tipo exoma como el escrito por Basha et al² en el 2017 realizado en Bélgica, quienes identificaron variantes patogénicas en cuatro genes TP63, TBX1, LRP6 y GRLH3, en familias con fisuras orofaciales no sindrómicas con herencia autosómica dominante y en particular genes que han sido asociados a causa sindrómicas.

En el año 2020 fue publicado un artículo en la revista Americana de Genética Humana, en la cual se proyectó un estudio de secuenciación del genoma completo (WGS) en tríos en 756 familias de ascendencia europea, colombiana y taiwanesa, identificaron genes candidatos como IRF6, CHD7, CTNND1, TFAP2A y ZFH4, relacionados con L/PH-NS con herencia autosómica dominante. También plantearon una hipótesis, en donde los genes de novo se podían identificar a partir de un gen vecino, pero todo ello relacionado al pico generado en los GWAS, esto podría proporcionar evidencia para nominar un gen como verdaderamente causal.

En un estudio del 2019, Yan et al. ⁸⁴ identificaron algunos factores de transcripción y mRNA implicados en la regulación del sistema celular, intentaron construir una red de co-reguladores encontrando nueve genes con papel clave en la patología de las anomalías orofaciales. En un GWAS publicado en el 2020, Lappalainen et al. ⁸⁵ hallaron 27 genes con significancia nominal y en el análisis de enriquecimiento funcional demostraron que estos estaban involucrados en la adhesión celular, la localización de proteínas en la membrana plasmática, en la regulación de la vía de señalización apoptótica y otras condiciones patológicas asociadas a L/PH-NS.

La información brindada por los diversos estudios ha permitido identificar variantes en genes relacionados al L/PH-NS; sin embargo, son menos los estudios que intentan resolver la funcionalidad en diferentes procesos biológicos en la especie humana que lleven a una mejor comprensión del desarrollo de esta malformación.

2. Aproximaciones computacionales

2.10.1 Análisis de genoma completo

El análisis de genoma completo hace referencia al estudio exhaustivo y comprensivo de la totalidad del contenido de ADN de un organismo dado. De manera general, esta prueba molecular (denominada Secuenciación Genómica) permite identificar la disposición de cada una de las bases nucleotídicas del genoma

bajo estudio, lo cual admite, entre otros, realizar el análisis de variantes en las regiones codificantes y no codificantes del genoma, alcanzando el desarrollo de nuevos conocimientos en genómica, así como en diversas aplicaciones clínicas. ⁸⁶

El flujo de trabajo para la secuenciación de genoma completo (o WGS por sus siglas en inglés) sigue pasos ordenados que van desde la preparación de los datos, la alineación y el ensamblaje de los mismos, el llamado de variantes y finalmente la anotación y análisis de la información encontrada. Para su ejecución existen varias plataformas como: Illumina, Pacific Biosciences, Oxford Nanopore, Ion Torrent entre otras. ⁸⁷

Una de las más usadas es la tecnología Illumina, la cual emplea lecturas cortas (150 pb), generando millones de fragmentos cortos que representan el conjunto de bases nucleotídicas del genoma. Esta lleva a que dichos fragmentos deban ser ensamblados unos con otros, con el fin de obtener una única secuencia por genoma o cromosoma, dado el caso y no millones de fragmentos.

En Illumina el proceso de ensamblaje se puede efectuar utilizando un genoma de referencia o realizando un ensamblaje de novo. Para su flujo de trabajo, inicialmente, genera clústeres para la amplificación de fragmentos de ADN a través de una PCR en puente. Luego proceden a secuenciar base por base usando cuatro nucleótidos marcados con fluorescencia para secuenciar las decenas de millones de grupos en la superficie de la celda de flujo en paralelo, esto permite una alta

precisión y elimina los errores específicos del contexto de la secuencia.⁸⁸ Prosigue la recopilación, procesamiento y análisis de datos, mediante el uso de herramientas y algoritmos bioinformáticos, para esto se requiere un flujo de trabajo integral compuesto por varios pasos como: control de calidad, ensamblaje, refinamiento del ensamblaje (polishing) y por último, el análisis del genoma ensamblado dentro del cual se cuenta el descubrimiento de variantes, que es de particular interés en este estudio.

A continuación, se lleva a cabo la filtración de las variantes identificadas, su anotación e interpretación mediante protocolos que clasifican las variantes como es el caso del propuesto por el Colegio Americano de Genética Médica y Genómica (ACMG) que las clasifica en patogénicas, probablemente patogénicas, probablemente benignas, benignas y de significado incierto.^{89,90}

Para la priorización de variantes se tienen en cuenta las bases de datos poblacionales como EXAC, 1000 genomas, generalmente, con referencia a un MAF <1% en la población, lo que le da un mayor valor de patogenicidad por una menor prevalencia, para completar criterios y clasificar las variantes se puede hacer uso de bases de datos como ClinVar⁹¹ y dbSNP⁹², Varsome, que siguen protocolos para la clasificación de variantes como es el caso de la ACMG,⁹⁰ luego se completa el análisis de las variantes a través de herramientas de predicción en el cambio de aminoácidos en proteínas como Sift, Polyphen, Mutation taster.⁹³ Para la anotación

se pueden emplear herramientas como ANNOVAR⁹⁴ y métodos específicos dentro del flujo de trabajo de GATK.

2.10.2 Herramientas bioinformáticas para análisis de red de genes

La construcción de red de genes se ha convertido en una herramienta útil para entender el funcionamiento de muchas proteínas, pues esta representación define cómo interactúan directa o indirectamente con otras proteínas y al unísono cumplen tareas específicas en un organismo. Por ejemplo, una proteína específica podría actuar en una ruta metabólica en el caso de una enzima o ser un intermediario efector de la señalización siendo un receptor o parte de la maquinaria de transcripción e incluso un componente del citoesqueleto. En general, estas redes son esenciales y altamente reguladas para una adecuada homeostasis en los organismos vivos.⁹⁵ Las redes construidas a partir de genes identificados y relacionados a enfermedades específicas en estudios permiten un mejor entendimiento de los mecanismos biológicos que subyacen un fenotipo dado.

Para la identificación de la arquitectura de una red de genes es fundamental el uso de bases de datos especializadas, ejemplo de ello es STRING, una herramienta que posibilita cargar datos relacionados a redes de proteína-proteína personalizables y con caracterización funcional de genes.⁹⁶ GENEMANIA es otro instrumento web que comparte información sobre hipótesis de funciones e interacciones de una gran lista de genes.⁹⁷ KEEG es una enciclopedia de información de genes, que involucra

asociaciones moleculares y redes de interacciones en forma de mapas de rutas, jerarquías y módulos involucrados en sus algoritmos. ⁹⁸ DAVID es una base de datos para la anotación funcional que permite entender el significado biológico de una gran cantidad de genes. ⁹⁹ Las bases de datos descritas y muchas otras desempeñan funciones esenciales en el área de investigación bioinformática y en los saberes relacionados a procesos biológicos en los seres vivos.

2.10.3 Herramientas para análisis estructural de proteínas

Las nuevas tecnologías permiten avanzar en el conocimiento sobre la estructura de las proteínas, en cómo predecir y diseñar arquitecturas tridimensionales, en la extrapolación de datos biomoleculares y de sus propiedades, todo ello facilita el entendimiento de las interacciones con otras proteínas y cofactores de sus plegamientos y su funcionamiento dinámico en muchas redes de vías biológicas. ¹⁰⁰

El discernimiento de la estructura de las proteínas comprende el funcionamiento de las mismas y su interacción con procesos biológicos, para ello se emplean métodos experimentales y computacionales, el primero es más preciso, pero costoso y requiere mayor tiempo para su ejecución, entre sus metodologías se encuentran: la cristalografía de rayos X, la resonancia magnética nuclear, en donde se emplean las propiedades magnéticas de spin de los núcleos atómicos para obtener las distancias entre los átomos de la molécula. ¹⁰¹ Otra técnica es la microscopía electrónica tridimensional que usa algoritmos para la reconstrucción y la proyección

de las moléculas mediante microscopio electrónico de transmisión (TEM).¹⁰² El segundo método es rápido y emplea enfoques e instrumentos computacionales. Uno de los procedimientos bioinformáticos usados en la predicción de estructuras proteicas es el de homología, que aplica un modelo comparativo, en el que toma como referencia plantillas estructurales de proteínas obtenidas previamente por técnicas experimentales y almacenadas en bases de datos como la Protein Data Bank (PDB), y a partir de ello se modela una estructura de interés, generalmente, esa plantilla al menos debe guardar una similitud del 40% con la proteína diana para obtener una estructura de calidad y si la semejanza es baja se puede usar otro tipo de técnicas como el Ab Initio que predice la estructura a partir de la secuencia de aminoácidos, de principios físicos de plegamiento de proteínas sin referencia de alguna plantilla, buscando la estructura con la energía libre más baja.¹⁰³

Para la evaluación de la calidad del modelo, se emplean herramientas que cuantifican el error espacial en las posiciones de los aminoácidos y determinan si la cadena adopta o no el pliegue correcto. Entre las herramientas a aprovechar se encuentra el ClustalW¹⁰⁴ utilizada para la realización de múltiples alineaciones de secuencias de proteínas. Otras aplicaciones son los meta servidores que exploran las plantillas y la alineación secuencia-estructura como es el caso del 3D-SHOTGUN y 3D-JURY.¹⁰⁵ ERRAT es un algoritmo que señala las regiones que se encuentran correctamente determinadas en una estructura de proteína, basado en la interacción atómica.¹⁰⁶ VERIFY 3D halla la relación de secuencia, lo infiere del pliegue que adopta la estructura de la proteína desconocida y deduce si es similar

a una proteína de estructura conocida.¹⁰⁷ PROCHECK proporciona un detallado análisis sobre la estereoquímica de una estructura proteica.¹⁰⁸ SWISSMODEL tiene una herramienta denominada Structure Assessment que permite el análisis de estructuras 3D, proporciona gráficas como Ramachandran para la evaluación estereoquímica, accede al uso del servidor web MolProbity que valida estructuras a nivel global y local para proteínas y ácidos nucleicos, además, admite una estimación de calidad usando QMEAN.¹⁰⁸

Los servidores y software de modelado de proteínas son amplios y como se describió anteriormente, hay diferentes metodologías para la realización de predicción de estructuras, por ejemplo, MODELLER y ROSETTA usan un enfoque de novo, en cuanto al primero, se funda en la alineación con una o más plantillas de proteínas de estructura conocida¹⁰⁹ y el segundo ensambla fragmentos cortos de proteínas conocidas mediante una estrategia de Monte Carlo para producir conformaciones de proteínas similares a la original¹¹⁰.

SWISS-MODEL es una plataforma empleada para el modelado de estructuras proteicas en 3D por homología mediante un enfoque comparativo. Maneja un flujo de trabajo que inicia con una secuencia de aminoácidos de una proteína a estudio en formato que puede ser FASTA, extrapolando la información experimental de estructuras de proteína relacionadas evolutivamente que van a servir como plantillas. Este proceso se realiza a través de bases de datos propias y externas y

herramientas de búsqueda como UniprotKB, SWISS-MODEL SMTL, BLAST y HHblits.¹¹¹ Las mejores plantillas son seleccionadas bajo criterios de estimación de calidad global y de estructura,¹¹² luego se ejecuta el servidor y se crea para cada plantilla un modelo de proteína 3D, todo ello en un marco de biología estructural computacional OpenStructure y ProMod3.¹¹³ La estimación de la calidad de los modelos obtenidos al final se fundamenta en la función de puntuación QMEAN que utiliza potenciales estadísticos de fuerza media para generar estimaciones de calidad global y por residuo.^{111,114}

AlphaFold es una herramienta de inteligencia artificial que predice estructuras proteicas mediante el uso de redes neuronales y procedimientos basados en limitaciones geométricas y evolutivas. Además, incorpora múltiples alineaciones de secuencias (MSA) y características por pares que permiten una predicción más precisa de la estructura 3D.^{71,115} La red AlphaFold predice directamente las coordenadas 3D de los átomos para una proteína diana utilizando la secuencia de aminoácidos primaria y las secuencias alineadas y plantillas de homólogos como entradas, usa así mismo un sesgo inductivo físico y geométrico para construir componentes a partir de los datos PDB, de esta manera, produce predicciones mediante unidades de procesamiento de gráficos.⁷¹ Al final entrega cinco modelos, teniendo en cuenta medidas de confianza bien calibradas y resueltas en secuencia. Este servidor web emplea una métrica de confianza por residuo llamada prueba de diferencia de distancia local predicha¹¹⁶ que maneja una escala de 0 a

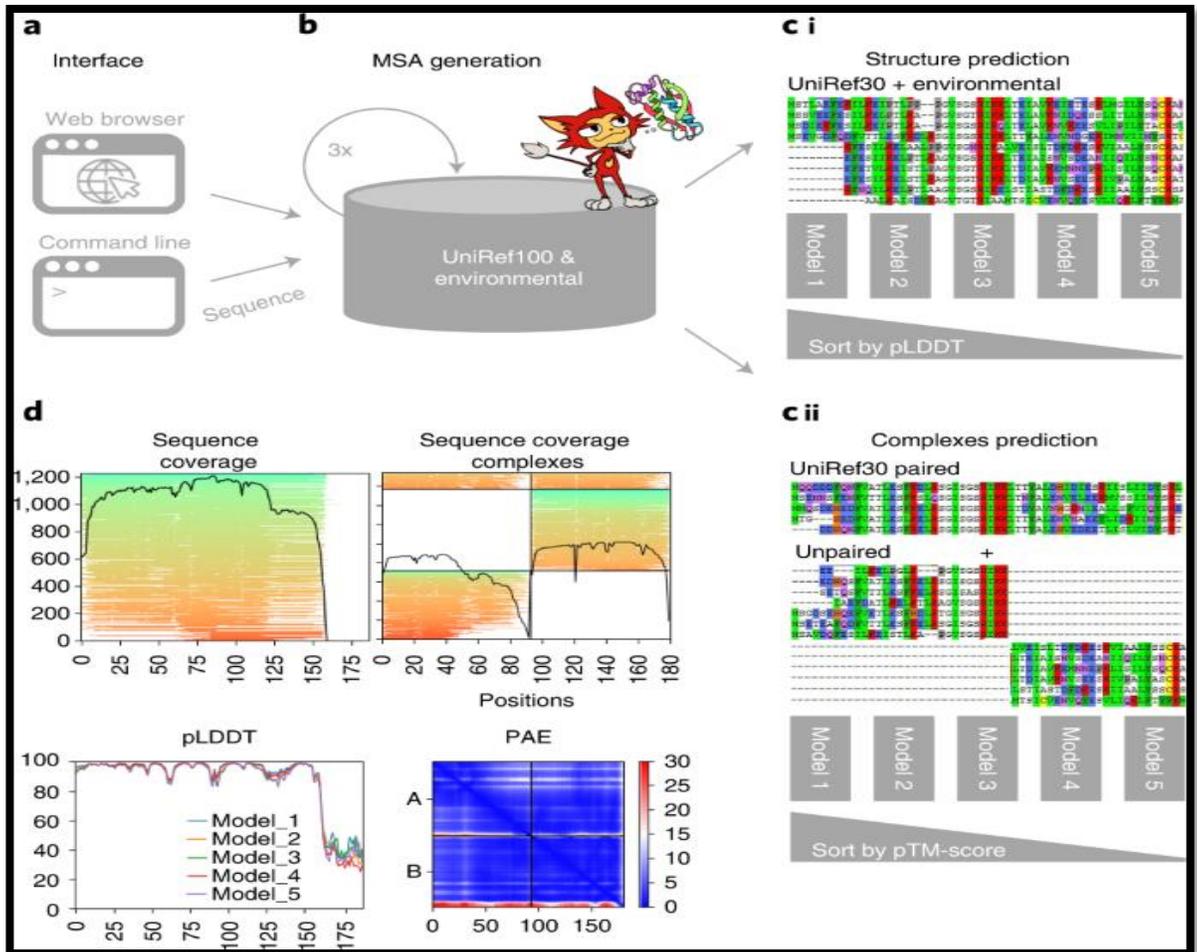
100, lo que hace es estimar la concordancia de la predicción con una estructura experimental.¹¹⁷

Alphafold puede ser ejecutado mediante la plataforma de Google Colab usando computación en la nube y una interfaz web simple a través de un entorno jerárquico basado en Python.¹¹⁸ ColabFold es empleada para la predicción de estructuras y complejos proteicos, para ello combinan la búsqueda rápida de homología de MMseqs2 (búsqueda de secuencia de muchos contra muchos) con AlphaFold2 o RoseTTAFold.¹¹⁹ Esta herramienta consta de tres elementos, el primero es un servidor de búsqueda por homología basado en MMseqs2, el segundo es una biblioteca de Python que prepara las características de entrada y visualiza los resultados y la tercera consta de los cuadernos Jupyter para uso básico, avanzado y por lotes. El algoritmo AlphaFold combina la autoatención en filas, columnas y triángulos para inferir la distancia de los residuos y la información evolutiva a partir de múltiples alineaciones de secuencias (MSA), basándose en trabajos anteriores.

120

La metodología de ColabFold se ejecuta en un conjunto de cuatro cuadernos Jupyter mediante Google Colaboratory o en las máquinas locales de los usuarios. Estos cuadernos están constituidos primero por AlphaFold2_mmseqs2, segundo por AlphaFold2_advanced, el tercero por AlphaFold2_batch y el último cuaderno por

RoseTTAFold. En la siguiente figura (Figura 2-2) tomada del artículo base de ColabFold se muestra el diagrama esquemático: ¹¹⁹



a , b , ColabFold tiene una web y una interfaz de línea de comandos (**a**) que envía secuencias de entrada FASTA a un servidor MMseqs2 (**b**) buscando en dos bases de datos, UniRef100 y una base de datos de secuencias ambientales, con tres iteraciones de búsqueda de perfil cada una . La segunda base de datos se busca utilizando un perfil de secuencia generado a partir de la búsqueda UniRef100 como entrada. El servidor genera dos MSA en formato A3M que contienen todas las secuencias detectadas. **C**, Para predicciones de estructuras individuales (i) filtramos ambos A3M usando un filtro consciente de la diversidad y devolvemos esto para que se proporcione como la característica de entrada de MSA a los modelos AlphaFold2. Para predicciones de complejos (ii) emparejamos los principales éxitos dentro de la misma especie para resolver los contactos entre cadenas y, además, agregamos dos MSA no emparejados (igual que i) para guiar la predicción de la estructura. Las predicciones de una sola cadena se clasifican por pLDDT y las complejas por la puntuación TM predicha. **d** , Para ayudar a los investigadores a juzgar la calidad de la predicción, visualizamos la profundidad y diversidad de MSA y mostramos las medidas de confianza de AlphaFold2 (pLDDT y PAE).

Figura 2-2 Diagrama esquemático de ColabFold. Fue extraído del artículo “ColabFold: making protein folding accessible to all” y elaborado por Mirdita et al. ¹¹⁹

Para el modelado de complejos proteína-proteína ColabFold utiliza dos formas de predicción: una a través de AlphaFold-multimer ¹²¹ y otra mediante la manipulación del índice de residuos en el modelo AlphaFold2. ¹²² Este usa una codificación posicional relativa con un límite en $|i-j| \geq 32$, al compensar el índice de residuos entre dos proteínas para que sea > 32 , las trata como cadenas polipeptídicas separadas.

119

2.10.4. Docking molecular

El acoplamiento molecular es un método computacional basado en mecánica molecular que predice las posibles interacciones entre dos moléculas que pueden ser proteína-ligando, proteína-proteína, proteína-ácido nucleico, entre otros modelos de complejo unido, ha sido empleado en muchos campos como en el diseño de fármacos, investigación y medicina. La mecánica molecular implica la descripción de un sistema poliatómico que utiliza la física clásica, para ello analiza datos de cargas, ángulos de torsión y geométricos de las moléculas implicadas para simular así la conformación más óptima. ¹²³

Hay varias metodologías para la realización de acoplamiento, una de ellas consiste en un tipo rígido donde hay un efecto de llave y candado, para encontrar la orientación correcta, esta se basa en la complementariedad geométrica. ¹²⁴ Otra técnica consiste en un ajuste inducido, donde la unión del ligando y la proteína es

flexible, ese cambio conformacional permite el encaje adecuado, este se fundamenta en complementariedad geométrica y energética.¹²⁵ Otros métodos de acoplamiento incluyen modificar la solvatación, el pH y acoplar en presencia o ausencia de agua.¹²⁶ A continuación, se describe la clasificación del software de acoplamiento molecular, figura tomada de una reseña científica (figura 2-3).¹²⁷

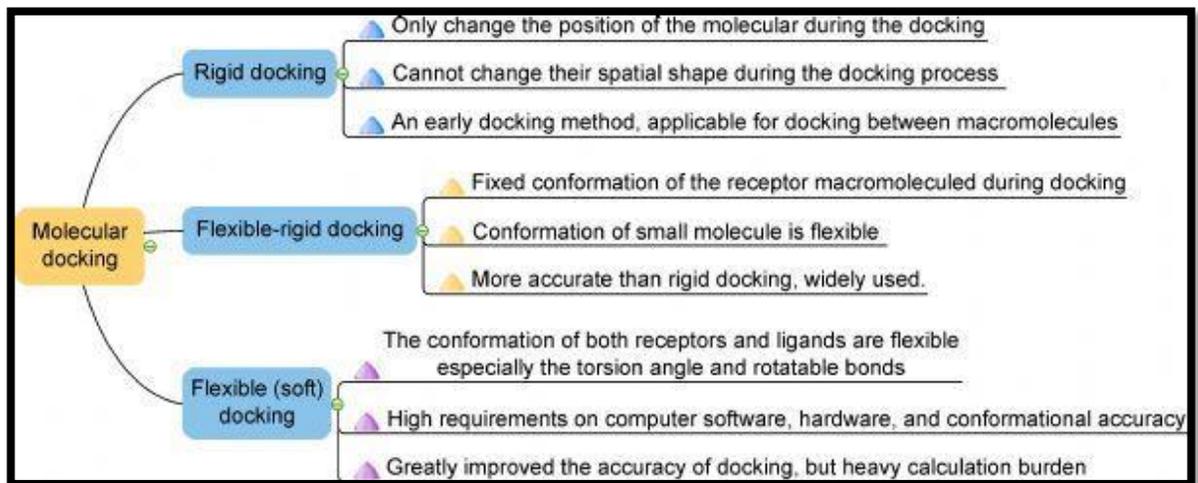


Figura 2-3 Clasificación del software de acoplamiento molecular. Fue extraído del artículo “Progress in molecular docking” y elaborado por Fan et al.¹²⁷

Existen diversos programas que median la simulación del acoplamiento molecular a través de algoritmos específicos y métodos de evaluación diferentes, enseguida se mencionan algunas herramientas computacionales empleadas para la ejecución de docking macromolécula-macromolécula, debido a que este estudio desarrolla este procedimiento y no acoplamiento proteína péptido o proteína-ligando, para tal efecto se muestra la tabla 2-2.

Tabla 2-2. Resumen herramientas computacionales empleadas para acoplamiento macromolécula-macromolécula.

Herramienta	Algoritmo	Tipo	Dirección	Referencia
AutoDock	Algoritmo genético	Independiente	http://autodock.scripps.edu/	128
ClusPro	FFT	En línea	https://cluspro.org	129
DOCK/PIERR	FFT	En línea	http://clsb.ices.utexas.edu/web/dock.html	130
DOT2	FFT	Independiente	http://www.sdsc.edu/CCMS/DOT/	131
FireDock	Monte-Carlo	En línea	http://bioinfo3d.cs.tau.ac.il/FireDock/	132
FRODOCK	FFT	En línea	http://frodock.chaconlab.org/	133
GRAMM	FFT	Online e independiente	http://vakser.compbio.ku.edu/main/resources_gramm.php	134
HADDOCK	Simulación annealing	En línea e independiente	http://www.bonvinlab.org/software/haddock2.2/	135
Hex	Correlaciones polares esféricas de Fourier	En línea e independiente	http://hexserver.loria.fr/	136
ICM-Pro	MC	Independiente	http://www.molsoft.com/icm_pro.html	137
LightDock	Optimización Glowworm Swarm	Independiente	https://life.bsc.es/pid/lightdock/	138
MEGADOCK	FFT	Independiente	http://www.bi.cs.titech.ac.jp/megadock/	139
PatchDock	Hash geométrico	En línea	http://bioinfo3d.cs.tau.ac.il/PatchDock/	140
PROBE	MC	En línea	http://pallab.serc.iisc.ernet.in/probe/	141
pyDock	FFT	Independiente	https://life.bsc.es/pid/pydock/	142
RosettaDock	MC	En línea	http://rosie.rosettacommons.org/docking2/su_bmit	143
ZDOCK	FFT	Independiente o en línea	https://zlab.umassmed.edu/zdock/http://zdock.umassmed.edu/	144

*Realizado empleando datos de la literatura como: “Evolution of In Silico Strategies for Protein-Protein Interaction Drug Discovery”, elaborado por Macalino et al. ¹⁴⁵

HADDOCK (High Ambiguity Driven protein-protein Docking) es un servidor web, usado para el estudio del acoplamiento de complejos de proteínas, se apoya en datos de interacción bioquímica y/o biofísica, igualmente de información

experimental de resonancia magnética nuclear (RMN) o mutagénesis. Así mismo, utiliza antecedentes de restricciones de interacción ambigua (AIR) descritos como la distancia ambigua entre los residuos involucrados en una interacción, todo ello para impulsar el proceso de acoplamiento.¹³⁵ Esta herramienta admite la formación de uniones mixtas entre macromoléculas, es decir, permite realizar predicciones de acoplamiento entre proteínas, entre proteína y ácidos nucleicos, entre los ácidos nucleicos, entre anticuerpos, todo ello bajo los protocolos de acoplamiento multicuerpo y solvado. HADDOCK presta sus clústeres locales para la ejecución de estas simulaciones. Tiene varias interfaces de acceso, una de ellas maneja un nivel básico llamada Easy, la cual permite el desarrollo de acoplamientos entre dos moléculas con formato PDB; otra interfaz es Expert, en este caso permite configurar características como el estado de protonación de cada residuo de histidina en las proteínas, definir el estado de carga de los extremos de las proteínas y ofrecer control sobre las regiones de las moléculas entre otros elementos de edición de la interacción.¹⁴⁶ La interfaz Guru aprueba el arreglo de casi todos los parámetros disponibles en el servidor web, aproximadamente 500. Las cuatro interfaces restantes consisten en la interfaz de predicción, la interfaz de refinamiento, la interfaz de multicuerpo y la interfaz de carga de archivos. Los resultados y el progreso de la simulación docking son enviados mediante un link al correo electrónico adscrito a la herramienta que dirige al usuario a una página de HADDOCK que muestra los diferentes modelos refinados creados y los enumera según el orden de puntuación.^{146,147}

ClustaPro es un servidor web para elaborar docking proteína-proteína mediante el método de acoplamiento de cuerpos rígidos para ejecutarlo requiere adjuntar los archivos en formato PDB del receptor y el ligando, este ofrece opciones avanzadas como: la eliminación de regiones de proteínas no estructuradas, la consideración de datos de dispersión de rayos X de ángulo pequeño (SAXS), encontrar sitios de unión a heparina, la aplicación de atracción o repulsión, la consideración de restricciones de distancia por pares y la construcción de homomultímeros.¹²⁹ Al final de cada ejecución de acoplamiento la herramienta proporciona diez modelos, los procesos pueden tardar hasta 4 horas.

Dentro de sus algoritmos computacionales, incluye el método de transformada rápida de Fourier (FFT) que utiliza un potencial de interacción por pares en su función de puntuación y aprueba acoplar péptidos flexibles en motivos de secuencia conocidos. En ClustaPro una de las proteínas es colocada en una cuadrícula fija, mientras que la otra se ubica en una cuadrícula móvil y al final analizan la energía de interacción sumando algunas de las funciones de correlación.¹⁴⁸ El algoritmo FFT permite realizar un proceso acelerado sin embargo posee dos limitaciones que consisten: una la aproximación de cuerpo rígido debido a las superposiciones que se pueden presentar y la otra es la expresión de energía a través de la suma de funciones de correlación.¹⁴⁹ Es una de las herramientas más usadas para docking proteína-proteína, es eficiente y una buena opción a la hora de realizar este tipo de procedimiento de química computacional.

HDOCK es otro servidor para ejecutar procesos docking para macromoléculas proteína-proteína y proteína-ácidos nucleicos, algo particular es que emplea un algoritmo basado en plantillas, situación que mejora su poder predictivo, claramente cuando existen estructuras en las bases de datos relacionados a la predicción a realizar, ya que cuando no hay una similitud o la identidad es menor al 30 % del complejo, puede proporcionar resultados engañosos, ¹⁵⁰ en este punto es donde el servidor también permite el análisis de complejos de novo sin el uso de plantillas. Otra ventaja son las opciones avanzadas que permiten adicionar datos claves para la realización del docking como agregar los residuos del receptor y ligando relacionados al contacto entre las macromoléculas, las restricciones de distancia entre residuos e incluso la utilización de archivos tipo SAXS. Sin embargo, tiene una desventaja es que solo admite el uso de dos macromoléculas para el desarrollo de sus acoplamiento, una gran diferencia con HADDOCK. ¹⁵¹

Al final de cada acoplamiento, HDOCK provee 100 modelos del complejo de predicción, proporciona información sobre la plantilla empleada y un resumen de los 10 mejores acoplamiento que incluye dos puntajes, un puntaje de acoplamiento que debe alcanzar -200 o una puntuación más negativa para referirse a un mejor

resultado y un score de confianza que debe ser mayor a 0,7 que confiere la probabilidad de unión entre las dos macromoléculas participantes en el docking.¹⁵²

HDOCK es un servidor práctico y clave a la hora de realizar procesos de acoplamiento molecular, en especial cuando sus ejecuciones tienen plantillas experimentales que dan como resultados predicciones de mayor confianza.

3. Metodología

3.1. Descripción general

Tipo de estudio: Observacional analítico de corte transversal.

Población: Hombres y mujeres dominicanos(as) con L/PH-NS con herencia autosómica dominante y sus padres (tríos) que asisten a la Fundación Niños que Ríen en República Dominicana que consintieron la participación en el estudio.

Muestra: Se escogieron por conveniencia cinco familias dominicanas en trío, de individuos afectados con L/PH-NS con herencia autosómica dominante que cumplían con los criterios de inclusión.

Criterios de selección:

Criterios de inclusión:

- Hombres y mujeres dominicanos (as) de todas las edades con L/PH-NS con herencia autosómica dominante.
- Origen dominicano conocido de al menos tres generaciones.
- Aceptación de participación en la investigación con posterior firma de un consentimiento informado. En caso de menores de edad, este

consentimiento debe estar firmado por sus padres o cuidadores legalmente constituidos.

Criterios de exclusión:

- Conocimiento acerca de familiares de primer, segundo y tercer grado de origen no dominicano.
- Individuos con L/PH sindrómico.

3.2 Fases del estudio

3.2.1 Fase 1. Preparatoria:

- Elaboración de instrumento de recolección de información.
- Capacitaciones: Identificación de microformas, realización de árboles genealógicos.¹⁵³
- Presentación del proyecto a los comités de ética de las instituciones participantes.
- Elaboración del documento de consentimiento informado.

3.2.2 Fase 2. Identificación de familias y toma de muestras

Selección y recolección de Muestras:

Para la selección de los pacientes elegidos se usó la base de datos de la Fundación Niños que Ríen, con diagnóstico de L/PH mediante el trabajo coordinado con los estudiantes y profesores de odontología de las universidades participantes (Universidad Nacional de Colombia y la Universidad Nacional Pedro Henríquez Ureña), se ubicaron los caso con L/PH en condición no sindrómica y herencia

autosómica dominante. Posterior a ello, por llamada telefónica se contactó a la persona a cargo del paciente para invitarla a participar del estudio, se citó para diligenciar el instrumento y verificar que cumpliera con todos los criterios de inclusión, además se tomaron datos clínicos, antecedentes familiares, personales, obstétricos, se confirmó el árbol genealógico ¹⁵³ en los casos de los pacientes seleccionados.

Una vez identificados los probandos y sus padres se recolectó la información mediante el instrumento, se validó la información y el cumplimiento de criterios de la inclusión, se tomaron las muestras bajo los protocolos del laboratorio Centogene, se rotularon bajo un registro único por familia y fueron enviadas las muestras al laboratorio, a continuación, se explica el paso a paso del proceso que se realizó en la toma de muestras:

- Se tuvo en cuenta el consentimiento informado para participantes; para menores de edad, se requirió autorización por parte de un representante legal.
- La extracción de sangre total periférica se hizo en tubo con EDTA (tubo tapa lila) para cada paciente (8 ml por cada paciente (2 tubos)).
- Se indicó el no requerimiento de ayuno.
- Las muestras de sangre fueron colocadas en una tarjeta de papel filtro, específicamente 50 μ L de sangre en cada uno de los 10 pozos de la Centocard, esta última fue proporcionada por el laboratorio Centogene para cada paciente.

- Una vez fueron ubicadas las muestras en la Centocard se dejaron secar a temperatura ambiente por dos horas.
- Las muestras permanecieron a temperatura ambiente.
- Se llenaron los datos de los probandos y sus padres, con el correspondiente antecedente familiar de L/PH-NS.
- Se archivaron en un sobre la Centocard, los datos requeridos y se enviaron al Laboratorio Centogene en Alemania con servicio de envío internacional.

Anonimización de datos: Se realizó un proceso de anonimización de datos, se dio un código en letras y números para proporcionar anonimato de la información de los participantes con el fin de evitar la identificación de estos (tabla-3-1). Fue codificado cada expediente, tubo de toma de muestra y tarjeta de laboratorio. Para el proceso bioinformático se continuó con la misma identificación en código de los participantes.

Tabla 3-1 Ejemplo de anonimización de datos.

Familia	FAM
Número de familia	1, 2, 3, 4 y 5
Probando	1
Madre	2
Padre	3

*De autoría propia.

Quedando de la siguiente forma:

- Para la familia #1: FAM11 (probando), FAM12 (madre), FAM13 (padre)
- Para la familia #2: FAM21 (probando), FAM22 (madre), FAM23 (padre)
- Para la familia #3: FAM31 (probando), FAM32 (madre), FAM33 (padre)
- Para la familia #4: FAM41 (probando), FAM42 (madre), FAM43 (padre)
- Para la familia #5: FAM51 (probando), FAM52 (madre), FAM53 (padre)

Resguardo y archivo de expedientes físicos y datos digitales: Los expedientes físicos fueron custodiados en archivo vertical con llave de la línea de investigación de L/PH en la oficina de Coordinación de Investigación de la Escuela de Odontología de la Universidad Nacional Pedro Henríquez Ureña (República Dominicana), con solo acceso al personal de investigación. Los datos digitales fueron manejados en computadora portátil por la línea y solo por los investigadores autorizados.

3.2.3 Fase 3. Secuenciación e identificación de variantes:

El análisis de genomas trío se realizó según los protocolos del laboratorio:

Centogene

El ADN genómico se fragmentó por sonicación, los fragmentos generados fueron ligados por adaptadores de Illumina, usados para la secuenciación

en la plataforma HiSeqX (Illumina) con una cobertura promedio de menos 30 x.

Se implementó un flujo de trabajo que incluyó el llamado de bases, multiplexación, alineación con referencia del genoma humano hg19, se realizó limpieza y calidad de los datos con identificación de lecturas de baja calidad y artefactos probables.¹⁵⁴

Para el llamado de variantes, se utilizaron los algoritmos de HiSeq del software v2.0 (Illumina, Inc., San Diego, CA), para la identificación de variantes tipo SNV e Indels se usaron los algoritmos de Starling Small Variant,¹⁵⁵ y para detectar SV y CNV se emplearon Canvas¹⁵⁶ y Manta.¹⁵⁷ Las lecturas cortas se alinearon usando Isaac.¹⁵⁵ La anotación de variantes se realizó con SnpEff¹⁵⁸ y herramientas bioinformáticas internas de ad hoc. Se aplicaron equipos de predicción in silico

para evaluar la conservación y el posible efecto de las variantes detectadas: FATHMM, PROVEAN, SIFT, PolyPhen2-HDIV, ADA_score, RF_score, MutationTaster, VEST3, LRT, MutationAssessor, MetaSVM, MetaLR, MCAP, REVEL, MutPred, CADD, DANN, GERP ++ NR, GERP ++ RS, phyloP100way Vertebrate, phyloP20way mammalian, phastCons100way Vertebrate. Todos los valores fueron obtenidos de la base de datos dbNSFP.^{154,159}

La información clínica e historia familiar se tomaron en cuenta para evaluar las variantes identificadas con respecto a su patogenicidad y posible causalidad, las variantes candidatas seleccionadas se clasificaron según las guías publicadas del ACMG como patogénicas, probablemente patogénicas y variantes de significado

incierto (VUS), en cuanto a enfermedad
y patrón de herencia. ¹⁵⁴

3.2.4 Fase 4. Análisis e interpretación de variantes identificadas.

3.2.4.1 Ensamblaje de genoma completo y llamado de variantes

Centogene envió dos datos provenientes del análisis de los cinco genomas trío, primero la información cruda en formato Fastq que se usó como ejercicio y entrenamiento para la realización del ensamblaje y llamado de las variantes de los quince genomas y segundo los datos de las variantes relacionadas a enfermedades. Con el uso de herramientas computacionales online se hizo la reclasificación de variantes encontradas asociadas a enfermedades incluido L/PH, se tuvieron en cuenta dos guías: la del ACMG ⁹⁰ y la de ClinGen. ¹⁶⁰ siguiendo los criterios de clasificación de variantes del ACMG. Para ello se usaron las bases de datos poblacionales como: GnomAD, 1000 genomas y ExAc, algoritmos de apoyo para clasificación de variantes: ClinVar, Varsome y InterVar. Se emplearon las herramientas de Genome Browser de la “University of California Santa Cruz”, en especial el Variant Annotation Integrator ¹⁶¹, donde se aplicaron algunos softwares de predicción: Mutation Taster, SIFT y PolyPhen-2, además, de la base de datos Database of Genomic Variants, (catálogo de variación estructural genómica humana) ¹⁶² y del NCBI Genome Data Viewer. ¹⁶³

3.2.4.2 Análisis estructural y funcional de variantes encontradas en L/PH-NS

3.2.4.2.1 Modelamiento de proteínas mediante herramientas computacionales

Para mapear los posibles cambios conformacionales causados por las variantes relacionadas a L/PH e identificar si las variaciones se encuentran en regiones clave que afecten la funcionalidad de las proteínas, primero se hizo la visualización de estas últimas, mediante la herramienta de PyMOL y para aquellas que no cuentan con plantilla en la Protein Data Bank (PDB), ni estructuras obtenidas por métodos experimentales en otras bases de datos, se emplearon programas para predicción de estructuras 3D como: SWISS-MODEL ¹¹¹ y ColabFold empleando AlphaFold2.

119

Para la evaluación de los modelos de predicción estructural se utilizaron: la herramienta "Structure Assessment de SWISS-MODEL, ¹⁶⁴ y el servidor de verificación y análisis de estructuras de proteínas (SAVES), ¹⁶⁵ usando las herramientas ERRAT, PROVE y PROCHECK. ¹⁰⁸

Se hizo, además, análisis por meta servidores para evaluar la estructura 2D de las estructuras 3D obtenidas.

3.2.4.2.2 Acoplamiento molecular

Para entender mejor el mecanismo biológico de daño de las variantes encontradas y describir las interacciones con otras estructuras proteicas se realizaron

ejecuciones de acoplamiento molecular, se tomaron en cuenta dos algoritmos: el servidor HDOCK y el HADDOCK, ambos ejecutan docking de proteína-proteína y proteína-ácidos nucleicos, con algoritmos diferentes.

Con el fin de visualizar los modelos de acoplamiento se usaron: la base de datos DNAProDB una herramienta online empleada para el análisis estructural de complejos ADN-proteína ¹⁶⁶ y el algoritmo Ligplot. ¹⁶⁷

3.2.4.2.3 Red de genes relacionada a IRF6 y al desarrollo de fisuras orofaciales no sindrómicas.

Se realizó una red de genes relacionada a las variantes identificadas y al desarrollo de fisuras orofaciales de carácter no sindrómico, se tomaron los recursos bibliográficos como: PubMd, Science Direct, EMBASE, Scopus, Elsevier, Biblioteca Virtual en Salud y Plos One, teniendo en cuenta los siguientes términos MeSH: cleft lip, cleft palate, Cleft Lip with or without Cleft Palate, Nonsyndromic, Orofacial Cleft, networks, gene, embryonic development, maxillofacial development, Palate and Lip. Se escogieron los artículos experimentales y con datos estadísticamente significativos con p valor < 0.05. Además, se hizo una revisión en bases de datos asociadas a redes reguladoras de genes: PathCards Pathway Unification Database , ¹⁶⁸ BioGRID, ¹⁶⁹ Genemania, ⁹⁷ STRING, ¹⁷⁰ NDEX, ¹⁷¹ Wikipathways, ¹⁷² TRRUST, ¹⁷³ PathDIP, ¹⁷⁴ e IntAct. ¹⁷⁵

Con el listado de genes obtenidos en el proceso anterior relacionados al desarrollo craneofacial y al L/PH, se construyó una red de genes a través de Excel, en las filas se tuvieron en cuenta las siguientes ubicaciones: en las dos primeras fueron situados los dos nodos (nodo1 y nodo2), seguido por el tipo de interacción identificado (expresión, inhibición, potenciador, co-expresión y desconocida relacionada a datos obtenidos por textmining y estadística), después, en la cuarta fila el tipo de evidencia (experimental y extracción de texto), en la quinta, la metodología empleada en los estudios encontrados asociados a la evidencia, en la siguiente fila se ubicó la especie en la que se realizaron las investigaciones, luego los resultados de estos mismos y en la última fila se dispusieron las referencias bibliográficas. Finalmente usando la herramienta Cytoscape, ¹⁷⁶ se creó el grafo dirigido.

Para el análisis de los datos obtenidos por la red de genes, se empleó la base de datos DAVID ⁹⁹ la cual ejecutó un estudio sistemático e integrador de todos los posibles procesos biológicos implicados, se tomó en cuenta el protocolo empleado en un estudio de red de genes publicado en el 2015. ¹⁷⁷ Para su realización primero se agregó el listado de genes de L/PH-NS, posteriormente, se seleccionó el tipo de identificador para los genes en este caso “oficial gene symbol”, después se eligió la especie en la que se deseaba realizar la anotación funcional (Homo sapiens) luego se escogió el tipo de lista, donde se optó por la lista de genes como alternativa y se dio inicio al proceso, en un segundo paso se tuvo acceso a los módulos analíticos de DAVID, que incluyeron: Las herramientas de anotación funcional, la agrupación

de anotaciones funcionales, el gráfico de anotación funcional y una variedad de tablas de anotaciones funcionales. ¹⁷⁸

3.3 Análisis estadístico

Se llevó a cabo una estadística descriptiva mediante el programa de Jamovi 2.2.5 ¹⁷⁹ se obtuvieron tablas de frecuencia con conteos y porcentajes, diagramas de barra para las variables nominales y categóricas de los datos sociodemográficos y clínicos. Para las variables continuas como la edad, se adquirieron datos de rangos, desviación estándar y frecuencias. A continuación, se evidencia la operacionalización de las variables:

Tabla 3-2 Operacionalización de las variables del estudio.

Operacionalización de las variables					
Variable	Definición	Posibles respuestas	Nivel de medición	Naturaleza	
Familia	Número de familia.	Familia 1 Familia 2 Familia 3 Familia 4 Familia 5	Nominal	Cualitativa	
Tipo de pariente	Tipo de familiar en la familia trío	1. Probando Madre 2. Padre	Nominal	Cualitativa	
Sexo	Condición biológica que distingue a los hombres de las mujeres.	1. Masculino 2. femenino	Nominal	Cualitativa	
Edad	Años cumplidos que ha vivido un individuo desde el nacimiento.	0-80	Razón	Cualitativa discreta	

Presencia de fisura orofacial	El individuo presenta o no labio y/o paladar hendido.	1. Afectado 2. No afectado	Nominal	Cualitativa
Ubicación de residencia	Sitio donde reside el individuo.	1. Urbana Rural 3. Peri-urbana	Nominal	Cualitativa
Procedencia	Provincia donde reside el individuo.	Provincias de República Dominicana	Nominal	Cualitativa
Percepción de estrato socioeconómico	Percepción de cada individuo de su estrato socioeconómico.	1. Alto Medio Medio-bajo Bajo	Nominal	Cualitativa
Edad de diagnóstico	Tiempo en el que se realiza diagnóstico de labio y/o paladar hendido.	1. Prenatal Natal 3 Posnatal	Nominal	Cualitativa
Nexo sanguíneo entre familiares cercanos	Parentesco natural de una persona con otra u otras que descienden de los mismos antepasados, hasta una tercera generación.	1. Si 2. No	Nominal	Cualitativa
Contacto con plaguicidas	Madre y/o padre estuvo en contacto con plaguicidas durante la gestación del probando.	1. Si 2. No 3. No sabe	Nominal	Cualitativa
Contacto con químicos	Madre y/o padre estuvo en contacto con químicos durante la gestación del probando.	1. Si 2. No 3. No sabe	Nominal	Cualitativa

Consumió ácido fólico 3 meses antes de la concepción	La madre del probando consumió previo a la concepción por tres meses ácido fólico.	1. Si sabe 2. No 3. No	Nominal	Cualitativa
Consumió ácido fólico durante los tres primeros meses del embarazo	La madre del probando consumió ácido fólico durante los primeros tres meses del embarazo.	1. Si sabe 2. No 3. No	Nominal	Cualitativa
Consumió la madre medicamento durante el embarazo	La madre del probando tomó medicamentos diferentes a los micronutrientes durante el embarazo.	1. Si sabe 2. No 3. No	Nominal	Cualitativa
Consumió licor durante el embarazo	La madre del probando consumió licor durante el embarazo.	1. Si sabe 2. No 3. No	Nominal	Cualitativa
Consumió cigarrillo durante el embarazo	La madre del probando consumió cigarrillo durante el embarazo.	1. Si sabe 2. No 3. No	Nominal	Cualitativa
Fue fumadora pasiva durante el embarazo	La madre del probando estuvo en contacto cercano de algún individuo fumador al menos 5 días a la semana.	1. Si sabe 2. No 3. No	Nominal	Cualitativa
Estuvo en contacto radiación durante el embarazo	La madre del probando estuvo en contacto con radiación ionizante en el embarazo.	1. Si sabe 2. No 3. No	Nominal	Cualitativa

Presencia de microformas	Formas leves de las hendiduras orofaciales.	1. Torus palatino 2. Úvula bífida 3. Agenesia del incisivo lateral superior 4. Dientes supernumerarios 5. Labio fisurado microformado 6. Bóveda palatina profunda 7. Arco maxilar en forma de V 8. Morfología de los incisivos laterales superiores (cónicos) 9. Paladar submucoso 10. Ninguna	Nominal	Cualitativa
Subfenotipo labio y/o paladar hendido	Conjunto de caracteres visibles que un individuo presenta relacionados a labio y/o paladar hendido.	1.Labio y paladar hendido bilateral 2.Labio y paladar hendido unilateral 3.Labio hendido unilateral 4.Labio hendido bilateral 5.Paladar hendido completo 6.Paladar hendido unilateral 7.Sano 8. Labio hendido unilateral y paladar hendido bilateral 9.Labio hendido bilateral y paladar hendido unilateral	Nominal	Cualitativa
Variantes incidentales	Documentación de variantes incidentales, patogénicas,	1. Positivo 2. Negativo	Nominal	Cualitativa

	probablemente patogénicas o de significado incierto en el análisis bioinformático del genoma.			
Variante relacionada a fisuras orofaciales	Cambios en la secuencia de ADN, patogénicos, probablemente patogénicos relacionados al fenotipo del paciente.	1. Si 2 No	Nominal	Cualitativa

*Autoría propia.

Diagrama de metodología



Figura 3-1 Resumen metodológico del estudio. Construcción propia, mediante Canva. 180

4. Resultados

4.1 Caracterización de la muestra

Mediante la base de datos de la Fundación Niños que Ríen, ubicada en la provincia de Espaillat en República Dominicana, se lograron identificar cinco familias que cumplieron con los criterios de inclusión mencionados en el apartado metodológico, a continuación, se describe la información relevante que se halló en cada una de las familias.

4.1.1 Componente sociodemográfico

De las cinco familias, tres tuvieron una percepción de estrato socioeconómico medio-bajo y dos de nivel medio (Figura 4-1). En cuanto a la ubicación de residencia, tres familias se localizaron en zonas rurales y dos en regiones peri-urbanas.

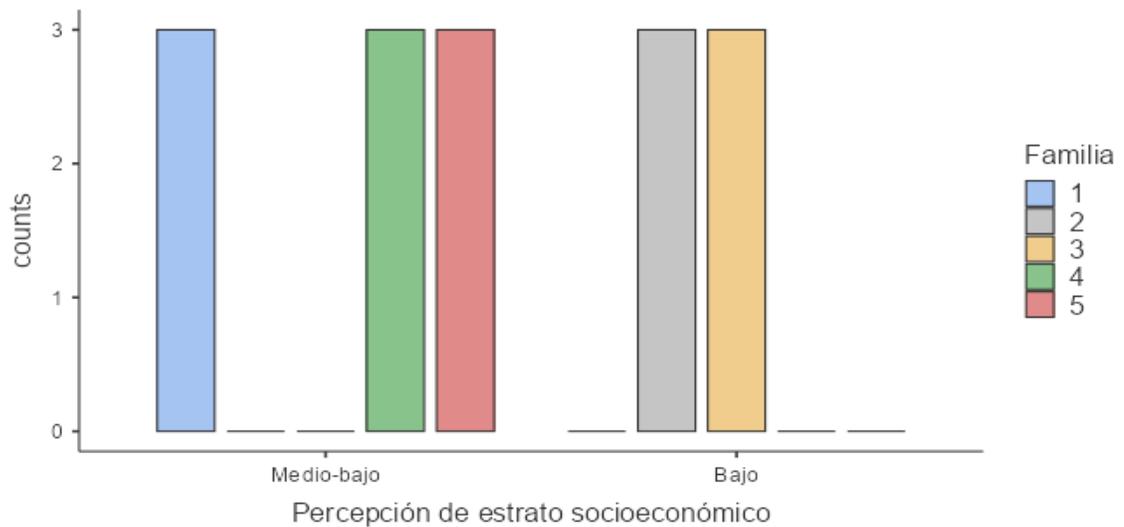


Figura 4-1 Frecuencia de percepción de estrato socioeconómico. De autoría propia, realizada mediante “The jamovi project (2022). jamovi (Version 2.2.5).⁷²

4.1.2 Componente ambiental

Con relación a los datos ambientales no hubo contacto de radiación ni consumo de alcohol por parte de las madres durante el embarazo. Las cinco madres no consumieron ácido fólico de manera preconcepcional (antes de los tres meses de concepción) y solo dos reportaron consumo de ácido fólico durante los primeros tres meses de gestación. En dos familias, los padres indicaron consumo de cigarrillo durante la gestación, dos padres y una madre tuvieron contacto con plaguicidas y químicos. Dos madres fueron fumadoras pasivas.

4.1.3 Componente clínico

Los quince participantes fueron caracterizados, tenían una edad media de 39.9 años, una desviación estándar de 14.4, con un valor mínimo de 16 años y un máximo de 63. Ocho presentaron L/PH-NS, cinco eran mujeres y tres hombres; de los siete no afectados cuatro fueron hombres y dos mujeres (Figura 4-2). Con respecto a los fenotipos de las fisuras, cuatro correspondían a labio y/o paladar hendido bilateral (LPHB), dos a labio y/o paladar hendido unilateral (LPHU), un individuo con paladar hendido (PH) y otro con labio hendido bilateral (LHB) y paladar hendido unilateral (PHU). La edad del diagnóstico del L/PH-NS para los cinco probandos fue de forma natal, es decir, al nacimiento.

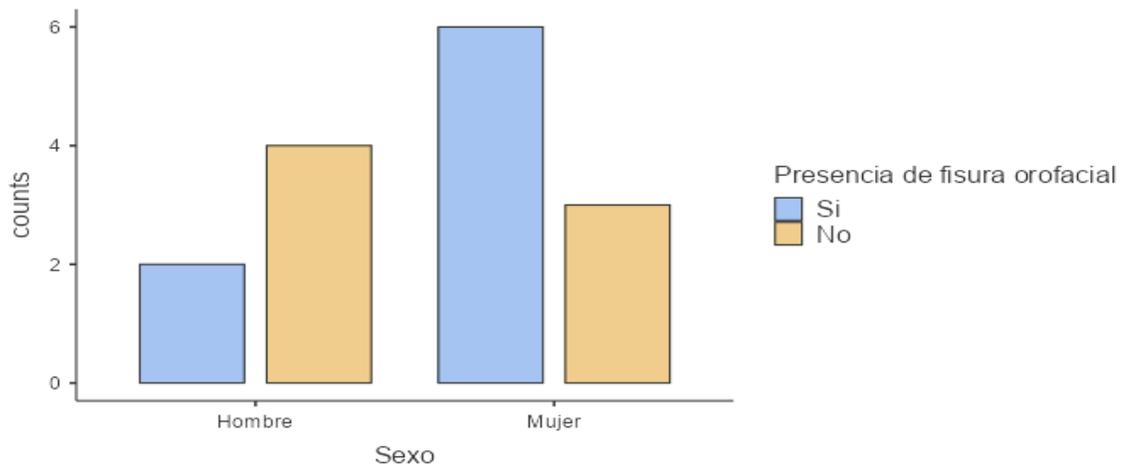


Figura 4-2 Frecuencia de presencia o ausencia de hendiduras orofaciales. De autoría propia, realizada mediante “The jamovi project (2022). jamovi (Version 2.2.5).⁷²

4.2 Microformas

Con relación a las microformas encontradas en las cinco familias reclutadas se identificaron los siguiente datos: (Tabla 4-1)

Tabla 4-1 Frecuencias tipo de microforma

Niveles	Conteo	% Total	Acumulativo%
Ausencia de microforma	4	12.1 %	12.1 %
Bóveda palatina profunda	4	12.1 %	24.2 %
Asimetría nasal	9	27.3 %	51.5 %
Agenesia de incisivo lateral superior	1	3.0 %	54.5 %
Úvula bífida	2	6.1 %	60.6 %
Arco maxilar en forma de V	5	15.2 %	75.8 %
Paladar submucoso	1	3.0 %	78.8 %
Incisivos laterales cónicos	1	3.0 %	81.8 %
Torus palatino	3	9.1 %	90.9 %
Ausencia congénita de dientes anteriores	2	6.1 %	97.0 %
Dientes supernumerarios	1	3.0 %	100.0 %

*De autoría propia, realizada mediante “The jamovi project (2022). jamovi (Version 2.2.5).⁷²

A continuación, se muestra una tabla resumen con la información más relevante identificada en las cinco familias. (Tabla 4-2)

Tabla 4-2 Resumen datos sociodemográficos, ambientales y clínicos de las cinco familias

Datos sociodemográficos		Familia 1	Familia 2	Familia 3	Familia 4	Familia 5
Sexo	Masculinos	1	0	0	1	0
	Femeninos	1	2	1	1	1
	Masculinos No afectados	1	1	1	0	1

Femenino	No afectados	0	0	1	1	1
Edad en años	Probando	17	34	27	23	16
	Madre	37	55	43	58	43
	Padre	48	63	51	43	41
Procedencia n	Probando	Provincia Puerto Plata	Provincia de Valverde	Provincia Santo Domingo	Provincia de La Vega	Provincia Espaillat
	Madre	Provincia Puerto Plata	Provincia de Valverde	Provincia de San Juan	Provincia de La Vega	Provincia Espaillat
	Padre	Provincia Dajabón	Provincia de Santiago Rodríguez	Provincia Puerto Plata	Provincia Espaillat	Provincia Espaillat
Percepción de estrato socioeconómico	Medio-bajo	1	0	0	1	1
	Medio	0	1	1	0	0
Datos ambientales		Familia 1	Familia 2	Familia 3	Familia 4	Familia 5
Consumo de cigarrillo	Madre	0	0	0	0	0
	Padre	0	1	1	0	1
Contacto plaguicidas padre	Madre	0	0	0	1	0
	Padre	0	1	0	1	0
Consumo de alcohol padre	Madre	0	0	0	0	0
	Padre	1	0	0	0	1
Contacto con radiación	Madre	0	0	0	0	0
	Padre	0	0	0	0	0
Consumo de ácido fólico	Antes de los 3 meses de embarazo	0	0	0	0	0

	Durante los primeros 3 meses de gestación	0	1	0	0	1
Datos Clínicos		Familia 1	Familia 2	Familia 3	Familia 4	Familia 5
Edad de diagnóstico	Prenatal	0	0	0	0	0
	Natal	1	1	1	1	1
	Posnatal	0	0	0	0	0
Tipo de fisura orofacial	Probando	LPHB	PH	LHB-PHI	LPHB	LHI-PH
	Madre	LPHB	LHI-PH	Sin fisura	Sin fisura	Sin fisura
	Padre	Sin fisura	Sin fisura	Sin fisura	LPHB	Sin fisura

*De autoría propia. Labio y paladar hendido bilateral (LPHB), paladar hendido (PH), labio hendido bilateral (LHB), labio hendido izquierdo (LHI), Paladar hendido izquierdo (PHI).

4.3 Descripción general de cada familia con su respectiva genealogía e información poblacional.

Familia 1

Es una familia que procede de Luperón, municipio de la provincia de Puerto Plata, ubicado al norte de República Dominicana, el conjunto escogido estuvo constituido por un probando masculino, la madre y el padre, en donde madre e hijo presentaban LPHB, sin ninguna otra asociación de anomalía o malformación congénita en algún otro sistema; el padre no presenta hendidura orofacial ni ningún otro compromiso relacionado a una enfermedad sindrómica. Por la línea materna, las cuatro últimas

generaciones proceden y residen en Luperón y como se muestra en el siguiente árbol genealógico del probando (Figura 4-4), hay dos generaciones afectadas con fisuras orofaciales, con dos diferentes fenotipos: paciente, madre y prima con LPHB y otra prima con PH, lo que constituye una genealogía con una herencia autosómica dominante. La línea paterna, abuelos y bisabuelos, proceden de Dajabón y del Partido, municipios que hacen parte de la Provincia de Dajabón al noroeste, no se evidenciaron familiares afectados con fisuras orofaciales o alguna condición sindrómica en este grupo (Figura 4-3).

Pertenecen a un grupo socioeconómico medio bajo, el probando es estudiante, la madre es ama de casa y el padre es pintor automotriz. El paciente fue fruto de primera gestación no planeada, pero deseada, inicio temprano de controles prenatales, primer trimestre, no consumo de ácido fólico antes y durante los primeros tres meses de la gestación, la madre negó contacto con plaguicidas, químicos, radiación cigarrillo y tabaco. El padre consumió alcohol en el primer trimestre, un día a la semana.

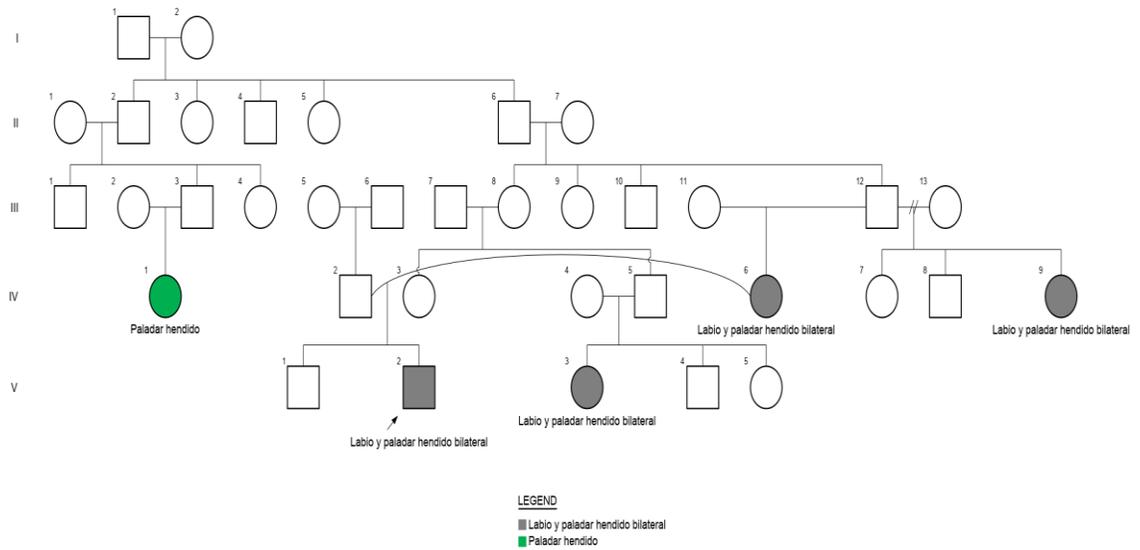


Figura 4-3 Genealogía familia 1. Realizado con la herramienta “Family history tool” de Invitae. ¹⁸¹

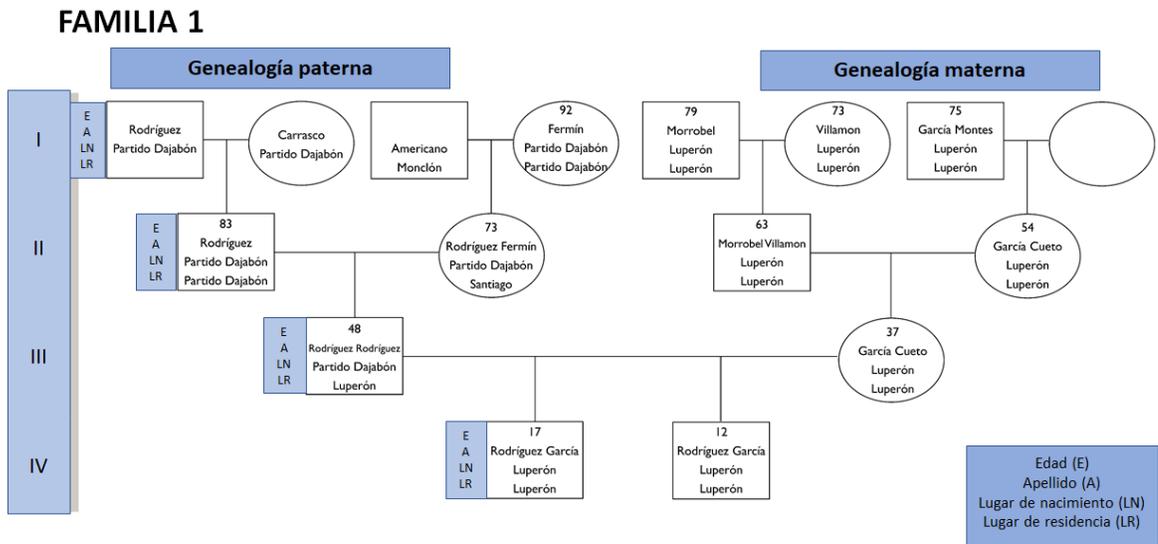


Figura 4-4 Información genealógica poblacional familia 1. Construcción propia, usando como base la plantilla de árbol genealógico empleado por el grupo de genética de poblaciones de la Universidad Nacional de Colombia. ¹⁸²

Familia 2

La segunda familia procede del Noroeste de República Dominicana en la provincia de Valverde y de Santiago de Rodríguez, constituida por una probando femenina, la madre y el padre, la primera mencionada nació en el pueblo de la Yagua localizado en la provincia de Valverde y reside en Jaibón de Pueblo Nuevo que es un distrito municipal referente al municipio de Mao en la Provincia de Valverde. La línea materna procede de Mao y algunos miembros como la abuela y bisabuela de Monción que es un municipio situado en la provincia de Santiago Rodríguez, limítrofe con Mao al norte. La línea paterna procede de Monción, el cual está ubicado al noroeste de República Dominicana y limítrofe a Mao (Figura 4-6). En esta familia hay dos generaciones afectadas por línea materna, madre y probando con fenotipo diferentes de fisuras orofaciales, la primera con PH y la segunda con LPHU, constituye una genealogía con herencia autosómica dominante, no asociada a alguna otra anomalía congénita, por lo que es una fisura no sindrómica (Figura 4-5).

La familia pertenece a un grupo socioeconómico medio, la probando y madre son amas de casa, el padre agricultor. La gestación de la probando fue planificada, producto del cuarto embarazo, la madre tuvo dos pérdidas sin etiología conocida,

hizo controles prenatales, no inició en el primer trimestre, no consumió ácido fólico antes de la gestación, pero sí en los primeros 3 meses, madre fue fumadora pasiva, padre fumaba 5 cigarrillo al día. Madre niega el consumo de alcohol, tabaco y cigarrillo. El padre tuvo contacto con plaguicidas y químicos como Cipemetrina y Biosin.

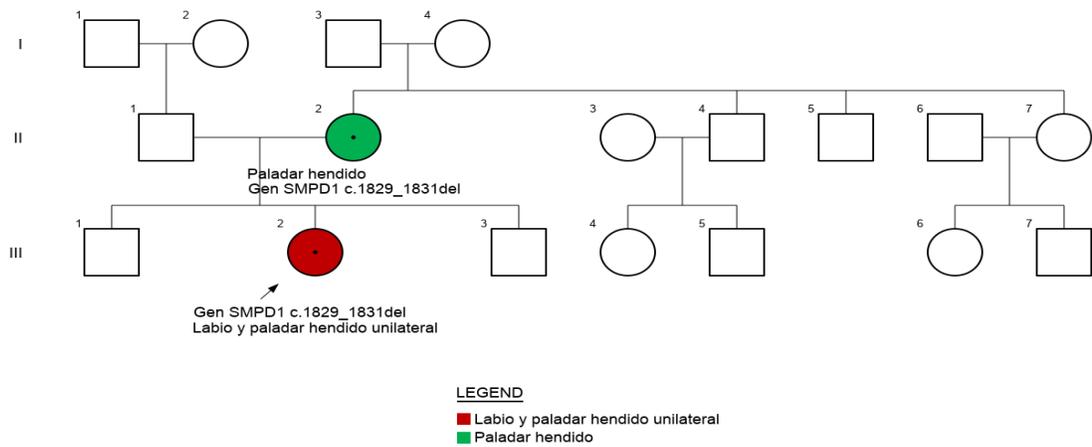


Figura 4-5 Genealogía familia 2. Realizado con la herramienta “Family history tool” de Invitae.¹⁸¹

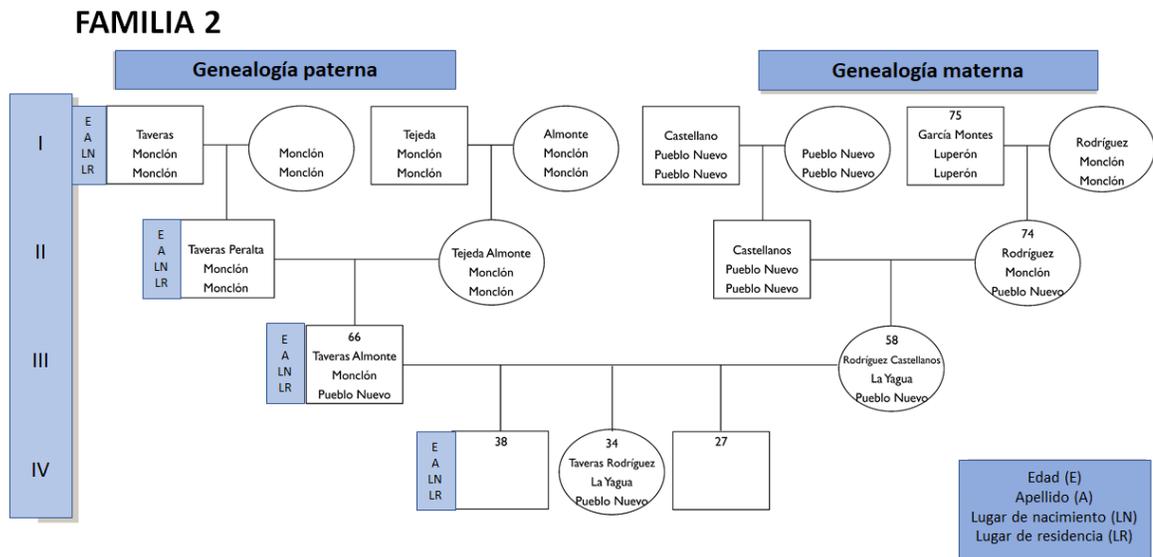


Figura 4-6 Información genealógica poblacional familia 2. Construcción propia, usando como base la plantilla de árbol genealógico empleado por el grupo de genética de poblaciones de la Universidad Nacional de Colombia. ¹⁸²

Familia 3

La tercera familia trío se conforma por una probando femenina, la madre y el padre. La probando procede de Santo Domingo, la Capital de República Dominicana, localizada al sur del país, la línea materna procede de la provincia de San Juan localizada en el centro-oeste. Esta limita al norte con las provincias de Santiago y Santiago Rodríguez, al este con la provincia Azúa y parcialmente con la provincia de La Vega en el noreste; la línea paterna procede de la provincia de Puerto Plata ubicada en el noroeste de República Dominicana (figura 4-9). Con relación al árbol genealógico de la probando, se encontraron afectadas tres generaciones con

fisuras orofaciales no sindrómicas relacionadas a la línea materna, dos tíos abuelos de género masculino y una tía LPHB y una probando con LHB y PHU (figura 4-8).

Paciente fruto de cuarta gestación, embarazo deseado no planeado, inicio de controles prenatales en primer trimestre, la madre niega consumo de licor, cigarrillo, tabaco y radiación. Los padres de este trío no tuvieron contacto con plaguicidas y químicos durante la gestación. La madre no consumió ácido fólico de forma preconcepcional ni en los tres primeros meses de gestación, además fue fumadora pasiva.

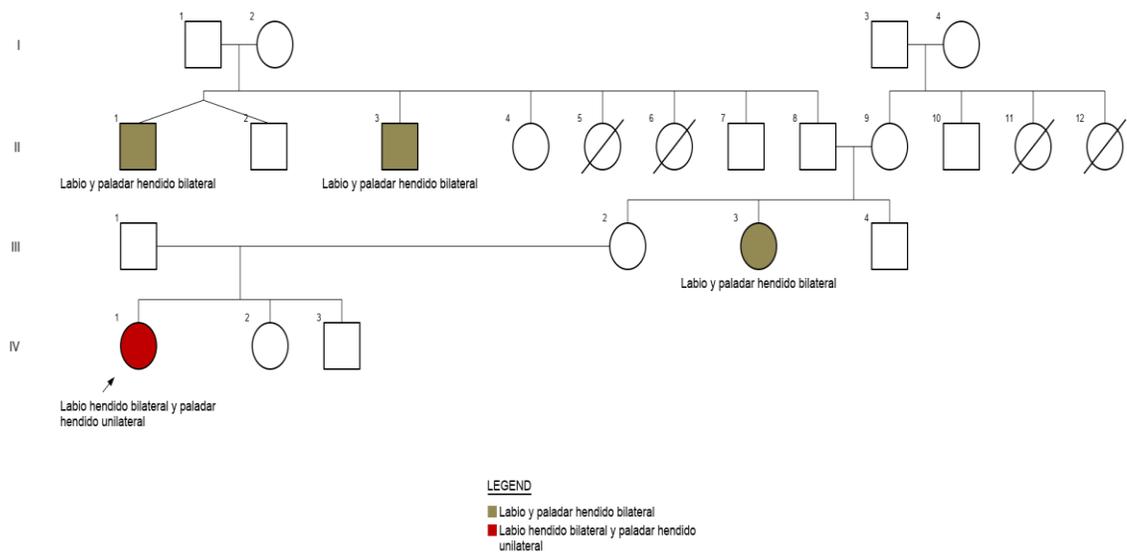


Figura 4-7. Genealogía familia 3. Realizado con la herramienta “Family history tool” de Invitae. ¹⁸¹

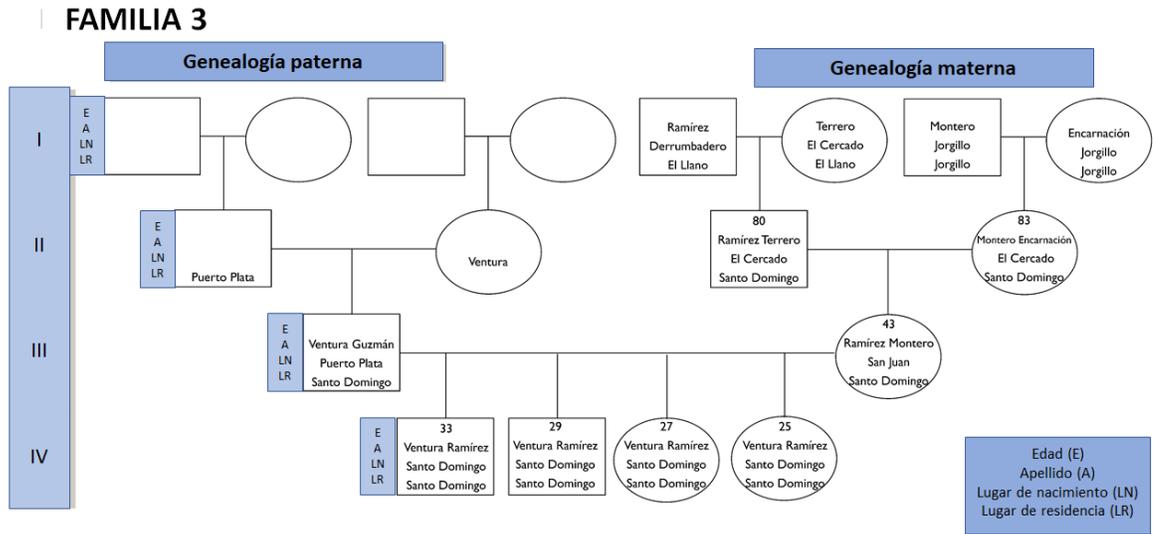


Figura 4-8. Información genealógica poblacional familia 3. Construcción propia, usando como base la plantilla de árbol genealógico empleado por el grupo de genética de poblaciones de la Universidad Nacional de Colombia. ¹⁸²

Familia 4

La cuarta familia procede de la provincia de La Vega, ubicada en el centro norte de República Dominicana, la estirpe trío escogida procede de los pueblos de Río Verde Abajo y La Vega hacen parte de la provincia ya mencionada, conformada por una probando femenina, la madre y el padre (Figura 4-10). El árbol genealógico de la probando presenta por la línea materna dos generaciones afectadas con fisuras orofaciales, madre, primo y probando con LPHB (Figura 4-9).

La familia pertenece a un grupo socioeconómico medio-bajo, probando es fruto de la tercera gestación, no inició de forma temprana controles prenatales, es decir, durante el primer trimestre de gestación, madre y padre refirieron contacto con plaguicidas y químicos, no relación en la gestación con alcohol y cigarrillo. No consumo de ácido fólico antes y durante el embarazo.

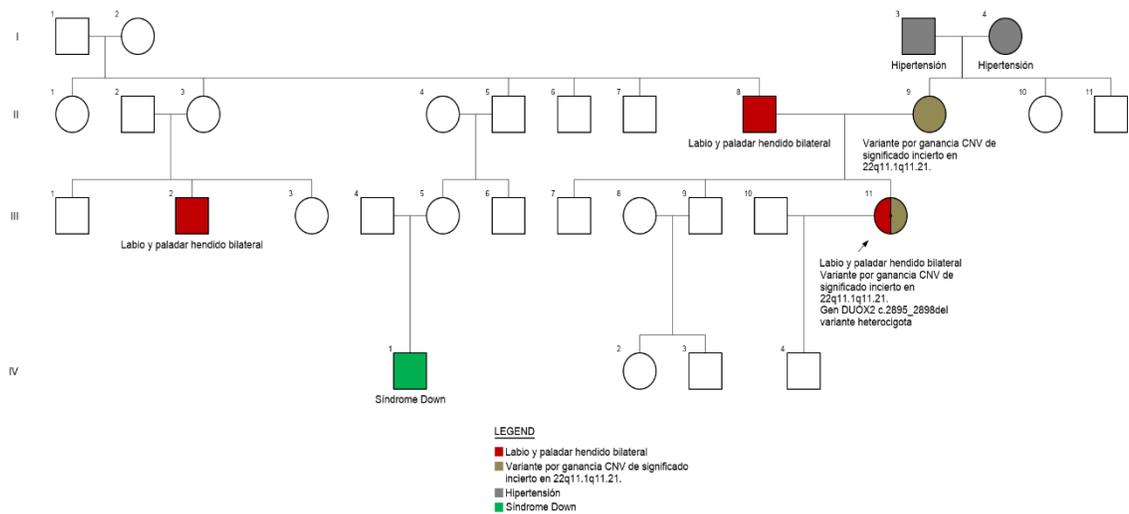


Figura 4-9 Genealogía familia 4. Realizado con la herramienta “Family history tool” de Invitae. ¹⁸¹

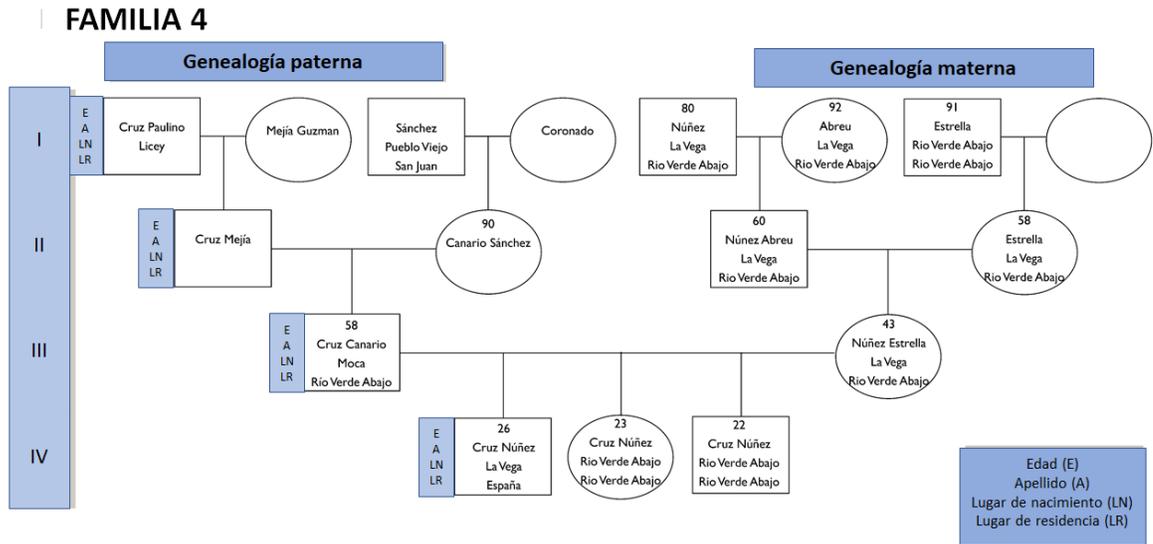


Figura 4-10 Información genealógica poblacional familia 4. Construcción propia, usando como base la plantilla de árbol genealógico empleado por el grupo de genética de poblaciones de la Universidad Nacional de Colombia. ¹⁸²

Familia 5

La quinta familia que procede de Moca, un municipio de la provincia de Espaillat, localizada en el centro norte de República Dominicana, tanto línea materna y paterna proceden de Moca, niegan consanguinidad, pero existe coterraniedad. El tríó seleccionado se constituye por probando femenina, la madre y el padre. El árbol genealógico de la probando (Figura 4-14) evidenció cuatro generaciones afectadas con fisuras orofaciales por la línea materna, bisabuela y primo con LPHU, madre con agenesia de incisivos laterales superiores y presunta microforma torus palatino bilobulado como se muestra en la figura 4-11; hermana, sobrino y primo con PH,

prima con LPHB y probando con LHI-PH , (Figura 4-13) ninguno con presencia de hoyuelos o fositas en el labio inferior a excepción de la probando que muestra una fístula unilateral en la región central de la mucosa oral cerca al labio inferior, como se evidencia en la figura 4-12.



Figura 4-11 Fenotipo madre de la probando de la familia 5. Obtenida durante la evaluación de los pacientes del estudio.



Figura 4-12 Fenotipo de la probando de la familia 5. Obtenida durante la evaluación de los pacientes del estudio.

La familia corresponde a un estrato socioeconómico medio-bajo. Probando fruto de primera gestación, embarazo deseado no planeado, inicia de forma temprana controles prenatales durante el primer trimestre, madre niega contacto directo con químicos, plaguicidas, no consumo de licor o cigarrillo. El padre consumió cigarrillo y licor en la gestación. Refirieron que viven cerca de un matadero y finca de cultivo de plátano y yuca.

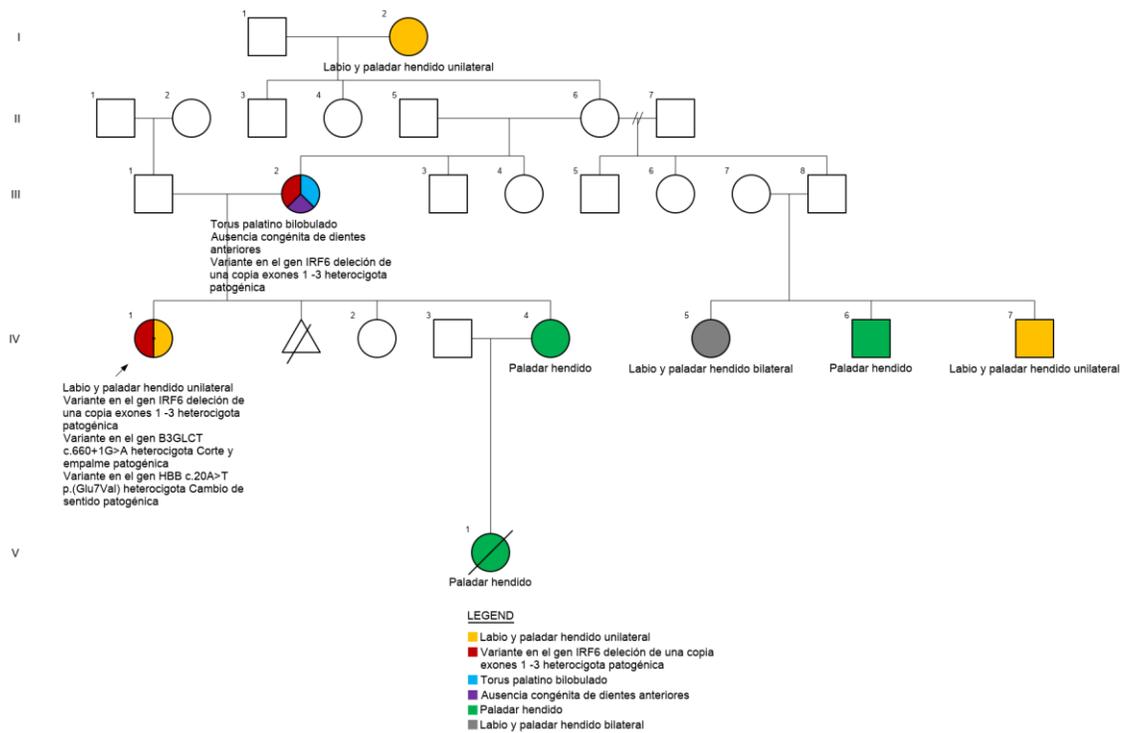


Figura 4-13 Genealogía familia 5. Realizado con la herramienta “Family history tool” de Invitae. ¹⁸¹

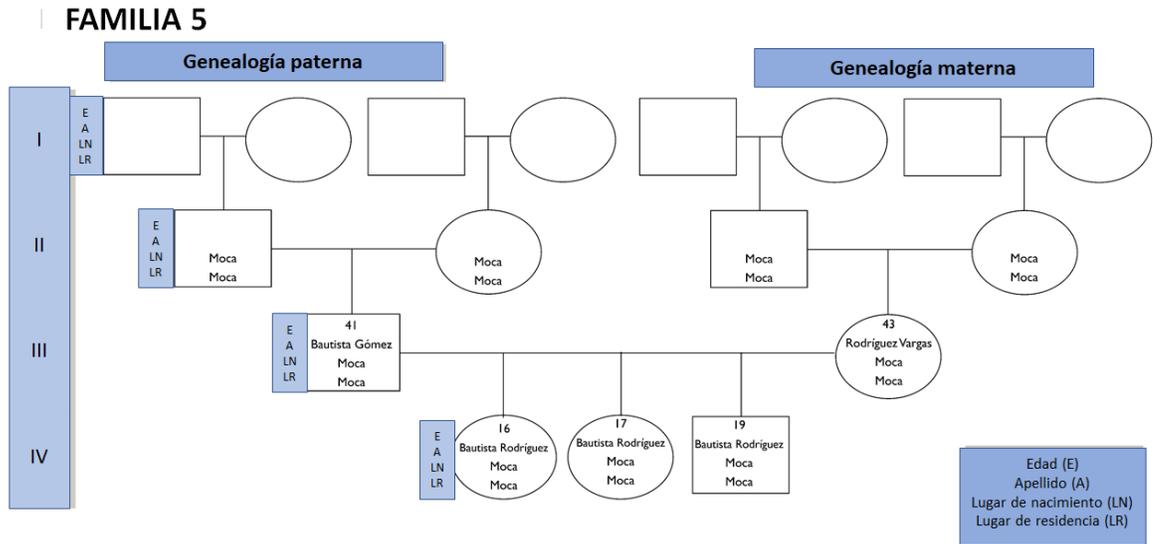


Figura 4-14 Información genealógica poblacional familia 5. Construcción propia, usando como base la plantilla de árbol genealógico empleado por el grupo de genética de poblaciones de la Universidad Nacional de Colombia. ¹⁸²

4.4 Caracterización molecular

En el análisis realizado de los genomas trío en las cinco familias de República Dominicana se identificaron ocho variantes relacionadas a enfermedades dos CNVs, dos Indels y cuatro SNVs. Al usar los criterios del ACMG se encontraron cinco variantes con clasificación patogénica y tres variantes de significado incierto (ver Tabla 4-3). Que incluyeron cambios en 23 genes IRF6, B362CT, HBB, CYP21A1, SMPD1, DUOX2, CCT8L2, TPTEP1, ANKRD62P1-PARP4P3, LINC01665, XKR3, HSFY1P1, GAB4, CECR7, IL17RA, TMEM121B, LINC01664, HDHD5, HDHD5-AS1, ADA2, CECR3, CECR2.

Tabla 4-3 Descripción de variantes encontradas en el estudio.

Gen o genes	Coordenadas de la variante	Cambio de Aminoácido	Cigotidad	Tipo y Clasificación	Familia y pariente
CYP21A2	NM_000500.5: c.-103A>G		Heterocigota	SNV Variante de significado incierto (Región promotora pseudogén efecto desconocido)	Familia dos probando y madre
CYP21A2	NM_000500.7: c.-113G>A		Heterocigota	SNV Variante de significado incierto (Región promotora pseudogén , efecto desconocido)	Familia dos probando y madre
SMPD1	NM_000543.4:c.1829_1831del	p.(Arg610del)	Heterocigota	Indel (In-frame) Patogénica	Familia dos Probando
CCT8L2, TPTEP1, ANKRD6 2P1-PARP4P 3, LINC016 65, XKR3, HSFY1P 1, GAB4, CECR7, IL17RA, TMEM12	arr[GRCh37] 22q11.1q11.2 1(16888900_17938906)x3			CNV Variante de significado incierto	Familia tres probando y madre

1B, LINC016 64, HDHD5, HDHD5- AS1, ADA2, CECR3, CECR2	DUOX2	NM_014080. 4: c.2895_2898 del	p.(Phe966 Serfs*29)	Heterocig ota	Indel (Frameshift) Patogénica	Familia tres Probando
	IRF6	seq[GRCh37] chr1:2099699 54- 209980303x1		Heterocig ota	CNV, delección de los exones 1 - 3 Patogénica	Familia cinco probando y madre
	B3GLCT	NM_194318. 3:c.660+1G> A		Heterocig ota	SNV (Corte y empalme) Patogénica	Familia cinco probando
	HBB	NM_000518. 4: c.20A>T	p.(Glu7Val)	Heterocig ota	SNV (Missense) Patogénica	Familia cinco probando

*De autoría propia, información tomada de los reportes de Centogene.

De las ocho variantes identificadas, solo una tuvo relación al desarrollo de L/PH-NS, encontrada en la familia cinco, una variante en el gen IRF6 en la probando (LHB y PH) y la madre (con ausencia de incisivos laterales superiores y torus palatino bilobulado). La variante consistió en una pérdida de una copia de los exones 1-3, se encontró en las coordenadas seq[GRCh37] chr1:209969954-209980303, que corresponde a una pérdida de 1 copia de 10.349 kb, este hallazgo fue confirmado

por análisis de MLPA. Esta variante es una CNV con clasificación patogénica según las recomendaciones de Centogene y ACMG. ⁹⁰

En la tabla 4-4 se muestra el resumen del análisis de reclasificación de las variantes encontradas no relacionadas con fisuras orofaciales:

Tabla 4-4 Reclasificación de variantes encontradas no relacionadas con fisuras orofaciales.

Familia	Variantes incidentales encontradas	Relación genotipo fenotipo	Criterios ACMG
Dos	CYP21A2: NM_000500.5:c.-103A>G heterocigota de significado incierto	Este gen codifica la enzima 21-hidroxilasa, esta última se involucra en muchos procesos relacionados con la producción de hormonas como mineralocorticoides, andrógenos y cortisol, a nivel de las glándulas suprarrenales, la deficiencia 21-hidroxilasa se relaciona con la aparición de hiperplasia suprarrenal congénita que tiene una herencia autosómica recesiva.	BS1, BP4 y BP6. Probablemente benigna.
	CYP21A2: NM_000500.7:c.-113G>A heterocigota de significado incierto	¹⁸³ Al realizar la evaluación del cuadro clínico de los individuos portadores, estos no manifiestan síntomas relacionados a la entidad descrita. Estas variantes se encuentran a nivel de la región promotora del pseudogén, sin embargo, estos dos cambios se han identificado que pasan al gen por conversión génica y afectan la función de la proteína que codifica el gen CYP21A2. ¹⁸⁴ El promotor del pseudogén es un 80 % menos	BS1, PP5. Variante de significado incierto.

activo que el CYP21A2 debido a la presencia de mutaciones -126C>T, -113G>A, -110T>C y -103A>G en estas cuatro variantes identificadas.¹⁸⁵

La variante c.-113G>A, tiene una frecuencia casi tres veces a la frecuencia alélica máxima esperada para una variante patogénica en el gen que causa enfermedad (0,002), lo que la cataloga en las bases poblacionales como una variante benigna. Los estudios de expresión in vitro del alelo aún no muestran resultados claros en la afectación funcional.¹⁸⁶ Sin embargo, en la literatura se encuentran reportados casos de pacientes con hiperplasia suprarrenal clásica y no clásica, pero asociada variantes heterocigotas compuestas, de una variante en región codificante del gen y otra dada en la no codificante como es el caso de la variante referida previamente.^{187,188} Con relación a la variante c.103 A>G, las bases poblacionales, evidencian a la frecuencia alélica máxima esperada para una variante patogénica en el gen que causa enfermedad en población africana (0,016) GnomAD.¹⁸⁹ En herramientas bioinformáticas para predecir daño la identifican como variante benigna y estudios en la literatura establecen la variante como benigna y en bases de datos como Global Variome shared LOVD indican casos reportados como de significado incierto (VUS).¹⁹⁰

	SMPD1: NM_000543.4: c.1829_1831del (p.Arg610del) heterocigota in frame	No se reportó esta variante en ninguno de los padres de la probando, por lo cual se puede considerar como una mutación nueva. En forma homocigota esta variante ha sido asociada con síndrome de Niemann Pick, ¹⁹¹ por lo que no se relaciona al fenotipo de la paciente.	PS3, PS4, PM4, PP3. Patogénica.
Cuatro	DUOX2: NM_014080.4: c.2895_2898del en forma heterocigota	Este gen codifica para una enzima relacionada con la formación de la hormona tiroidea y en forma homocigota puede ocasionar hipotiroidismo congénito. No está asociado con los antecedentes de la probando. ¹⁹²	PVS1, PM2, PP5. Patogénica.
	Ganancia en el número de copias (CNV) de significado incierto en 22q11.1q11.21	Se identificó una variante por ganancia en el número de copias (CNV) de significado incierto en el 22q11.1q11.21 de aproximadamente 1 Mb, la cual incluye 16 genes. Las duplicaciones en esta región son asociadas con el síndrome cat eye. Siguiendo las normas técnicas para la interpretación y el informe de variantes constitucionales del número de copias: una recomendación de consenso conjunto del ACMG y Clinical Genome Resource (ClinGen), ¹⁶⁰ . basados en esta información, la variante es de significado incierto, puntaje total de -0.60 (puntaje entre 0,89 a -0,89 puntos). Se confirma que esta variante es de origen materno, la madre es sana. Por lo anterior, se realizó un examen a la probando y a la madre en busca de los hallazgos clínicos relacionados con el síndrome de cat eye, no encontrando ni colobomas de iris ni antecedentes de alteraciones	2B, 3A, 4D, 5C. Significado incierto.

		anorectales. Igualmente, dado que el progenitor afectado por hendidura orofacial es el padre y la CNV se encontró en la madre, no se asoció este hallazgo con la hendidura orofacial. Los diferentes trastornos genómicos relacionados con el aumento de la dosis de 22q11.2 presentan varias características clínicas comunes, aunque tienen una penetrancia y expresividad variables, que van desde individuos aparentemente normales hasta individuos gravemente afectados. ¹⁹³	
Cinco	B3GLCT: NM_194318.3: c.660+1G>A en forma heterocigota	Esta variante patogénica tipo SNV de corte y empalme, el gen en particular codifica la enzima beta 3-glucosyltransferasa y está asociada a una entidad de carácter recesivo, es decir, requiere de dos variantes en los dos alelos del gen para que se presente la enfermedad, en este caso puede causar el síndrome de Peter plus, caracterizado por una afectación sistémica, ¹⁹⁴ por lo que el individuo se considera como portador, pero no afectado.	PVS1, PM2, PP5. Patogénica.
	HBB: NM_000518.4: c.20A>T heterocigota, con clasificación patogénica	Esta variante se transmite a través de la línea materna y con herencia autosómica recesiva, requiere una forma homocigota para que cause enfermedad, en este caso, anemia de células falciformes, ¹⁹⁵ por lo cual el individuo es portador, pero no afectado.	PS3, PS4, PM1, PM5. Patogénica.

4.5 Análisis estructural y funcional de variantes encontradas en L/PH-NS

4.5.1 Modelo unión a ADN de IRF6

Teniendo en cuenta los servidores web para evaluación de estructuras 3D de proteínas, se seleccionó el mejor modelo relacionado a la estructura de IRF6, se incluyeron parámetros de análisis como: puntuaciones de restricciones de distancia, combinaciones de ángulos diédricos Ψ (psi) y Φ (phi), las características fisicoquímicas y la estructura más acorde para evaluar el objetivo de analizar las implicaciones funcionales de la variante identificada en procesos biológicos relacionados a fisuras orofaciales. Se evaluaron dos programas (SWISS-MODEL y ColabFold: AlphaFold2) y se escogió el modelo obtenido por ColabFold-AlphaFold2.

Para la ejecución de AlphaFold2 se tuvo en cuenta:

- El uso de Amber.¹⁹⁶
- La selección de plantillas por homología de IRF6 (3DSH, 5JEL, 5JEO, 5JER, 7JFL, 1ZOQ, 5OEM, 5BVI, 7O56, 1IRG, 2DLL, 2IRF, 7JM4, 1T2K, 3QU3, 2O61, 6TD4).
- La elección de MMseqs2 (UniRef+Ambiental) para obtener el listado de secuencias relacionadas a la proteína diana.
- Las opciones avanzadas de AlphaFold-multimer-v[2].

- La elección de tres ciclos y de una resolución de 300 PPP (puntos por pulgada) para las imágenes resultantes.

Con dichos parámetros se ejecutó el algoritmo.^{118,119} Al modelo final se le hizo un refinamiento con el programa GalaxyRefine¹⁹⁷ y se obtuvo el siguiente modelo de predicción de estructura 3D del dominio de unión a ADN (Figura 4-15).

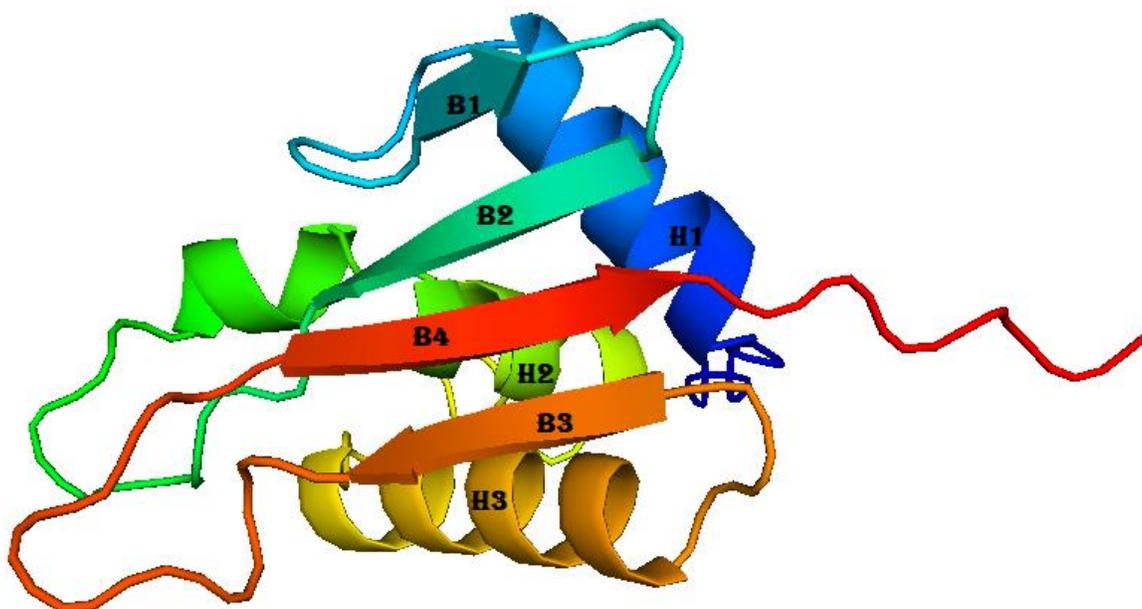


Figura 4-15 Modelo dominio unión ADN de IRF6 obtenido por AlphaFold2. Visualizado por PyMOL. Están presentes cuatro hélices alfa: H1 (residuos 10-20), H2 (residuos 49-64), y H3 (residuos 76-89), cuatro láminas beta β 1 (residuos 27-30), β 2 (residuos 35-39), β 3 (residuos 93-97) y β 4 (residuos 108-113), además, ocho bucles, tres de ellos largos.

Se prevé que H3 interactúe con el surco principal del ADN, H2 y H3 hacen la hélice giro hélice (HTH) motivo que forma el dominio de unión al ADN como se ha visto en estructuras obtenidas por métodos experimentales previos con otros IRF. Cabe aclarar que se hizo un análisis en cinco meta servidores de predicción de estructura secundaria identificando a la H2 con inicio en el residuo 49 y culminando en el 64, hay dos aminoácidos el N54 y el T55 relacionados a desorden, hallazgo en dos algoritmos de predicción posiblemente por ello se forma un pequeño loop en dicha región 3D. ^{73,75}

4.5.1.1 Verificación modelo de dominio de unión a ADN en IRF6

Se evaluó la calidad de la estructura 3D obtenida del dominio de unión a ADN de IRF6 con los siguientes análisis (figuras 4-16, 4-15, 4-16, 4-17, 4-20).

1. Descripción topológica

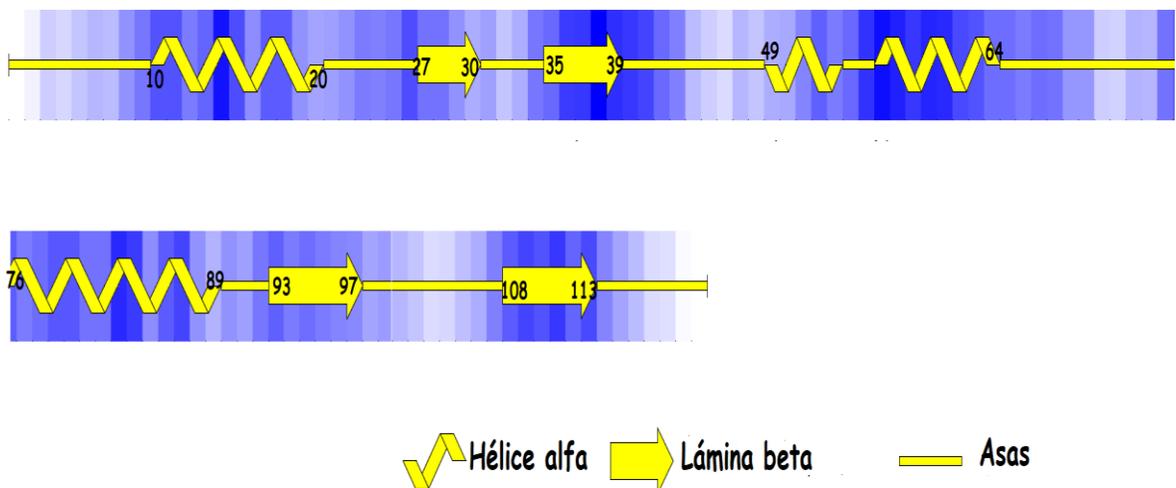


Figura 4-16. Topología del dominio de unión a ADN. Obtenida en SAVE. ¹⁶⁵ En la

gráfica se evidencia la estructura 2D del modelo de predicción 3D obtenido por medios computacionales, se identifican tres alfa hélices, cuatro láminas beta y ocho bucles, los residuos que representan cada arquitectura secundaria corresponden a la reconocida en el modelo 3D ya descrito.

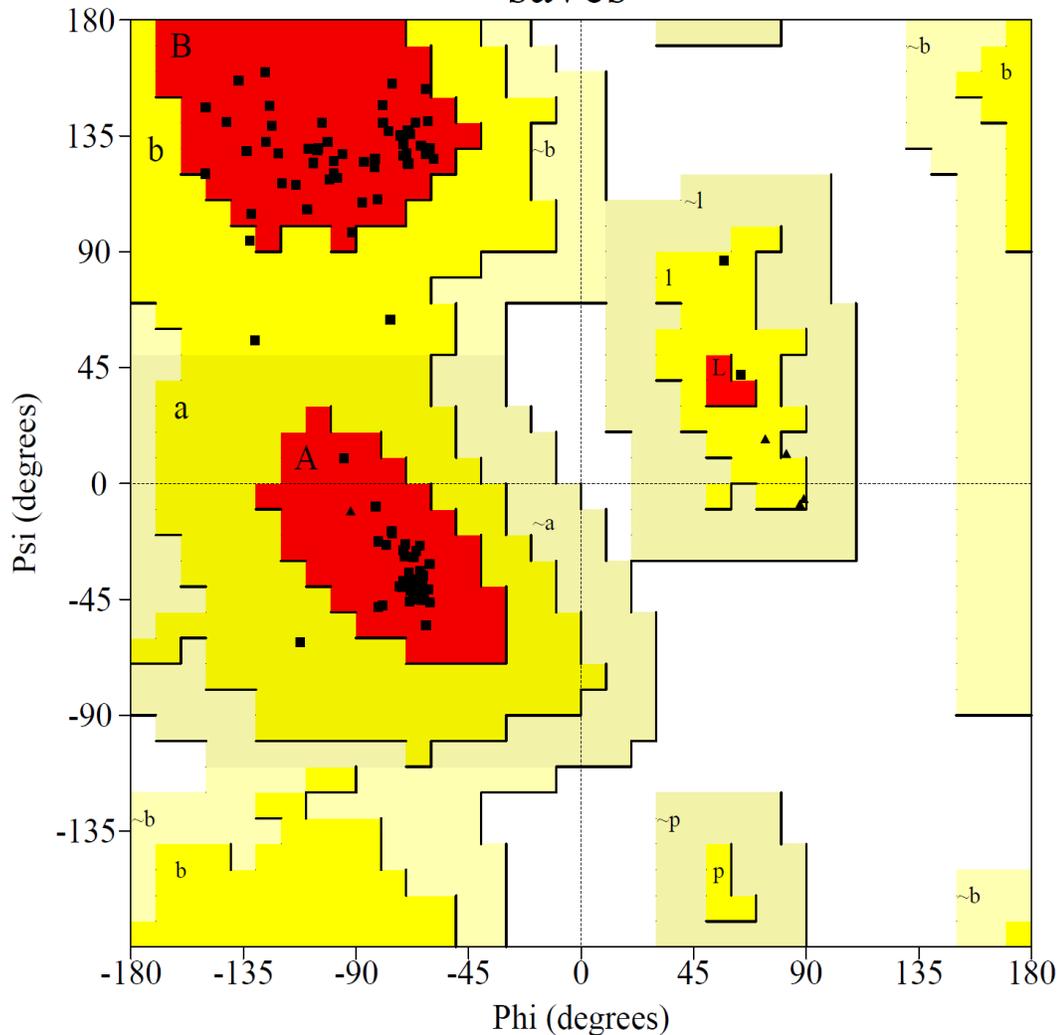
3. Validación con herramientas SAVE

Se hizo validación de la estructura obtenida del dominio de Unión a ADN de IRF6 mediante las herramientas SAVE ¹⁶⁵. A continuación, se muestran los resultados obtenidos por PROCHECK mediante un gráfico de Ramachandran (Ver figura 4), en él se evidenció que el 96,1% de los residuos de la estructura 3D se ubicaron en las regiones más favorables identificadas en color rojo, donde se sitúan los ángulos phi y psi más estables de las estructuras hélice alfa y lámina beta, corroborado por la densidad de los puntos localizados en la región roja. Mostró, además, que la proteína tiene una estructura en hélice alfa dextrógira mayor en comparación con la lámina beta y la hélice alfa levógira. El 3,9% de los aminoácidos se encontraron en la región favorable en color amarillo y no hay residuos en las zonas menos favorecidas que corresponden a los colores abano y blanco. Se identificaron, además cinco residuos de glicina que se reconocen como triángulos y once de prolina.

PROCHECK

Ramachandran Plot

saves



Plot statistics

Residues in most favoured regions [A,B,L]	98	96.1%
Residues in additional allowed regions [a,b,l,p]	4	3.9%
Residues in generously allowed regions [-a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%

Number of non-glycine and non-proline residues	102	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	5	
Number of proline residues	11	

Total number of residues	120	

Figura 4-17 Gráfico de Ramachandran. Obtenido por análisis PROCHECK. 165

En el siguiente parámetro se empleó el servidor ERRAT que traza dos errores en los aminoácidos 25 y 103, tiene un nivel de confianza del 98, 165%, esto determina la resolución de la estructura de predicción 3D (Figura 4-22).

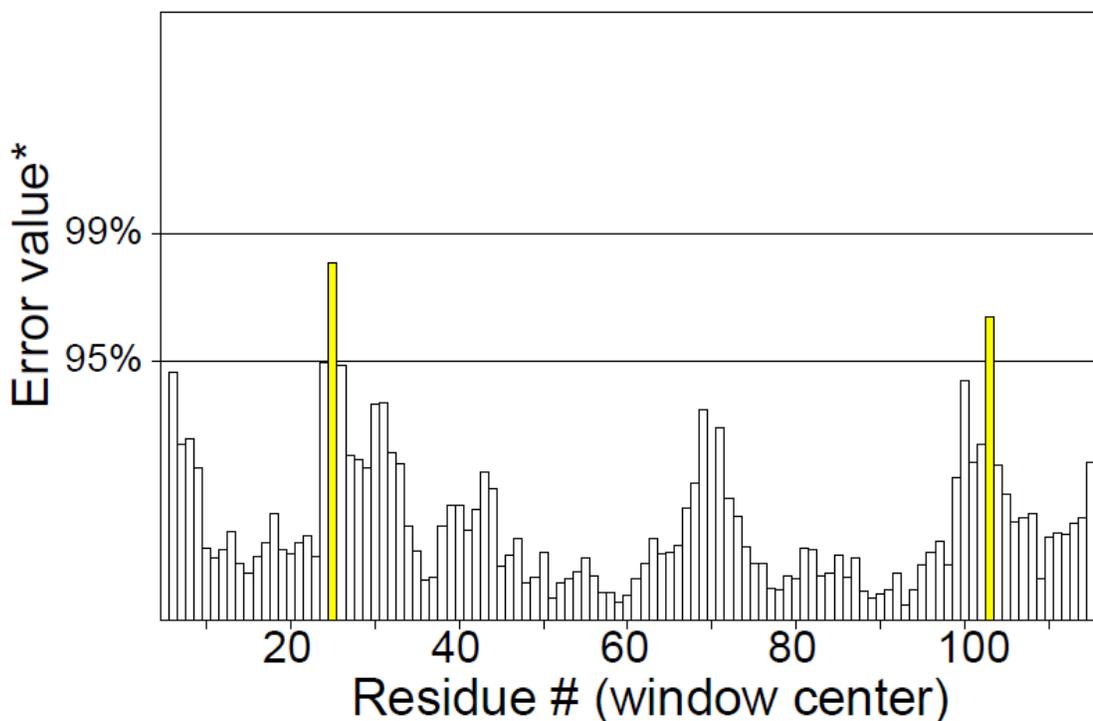


Figura 4-18 Gráfica ERRAT. Obtenido del análisis de ERRAT. ¹⁶⁵

3.Cálculo de la Raíz de la desviación cuadrática media (RMSD)

Usando la herramienta PyMOL, se realizó el cálculo de las distancias entre dos estructuras de dominio de unión a ADN, IRF4 e IRF6. En este estudio se empleó la plantilla 7OOT del PDB del IRF4 y la estructura 3D obtenida de IRF6. Se identificó un valor de 0.585 (Figura 4-19).

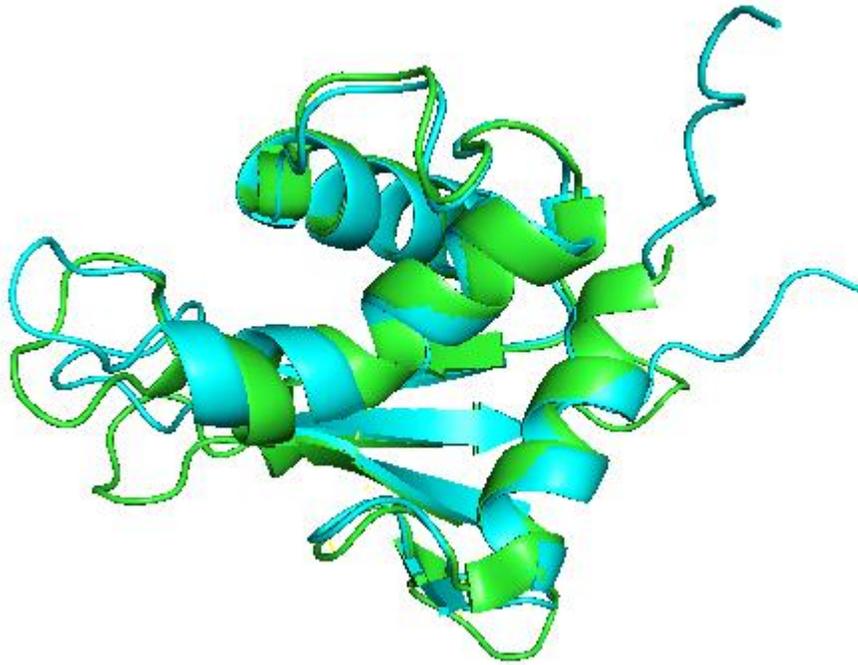


Figura 4-19 Superposición estructural. RMSD entre 7OOT (IRF4) en verde y IRF6 en azul. Obtenido por PyMOL.

4. Escalada de confianza por residuo determinación AlphaFold2

Utiliza una estimación de confianza por residuo que va de 0 a 100, las regiones con pLDDT > 90 representa alta precisión, entre 70 y 90 con buena precisión y las regiones entre 50 y 70 son de confianza baja y deben tratarse con precaución.

Como se observa en la gráfica (figura 4-) los 2 primeros residuos, los aminoácidos del 42 al 53 y del 100 al 105 tienen una calidad de 60 de baja confianza. (Figura 4.), los demás se encuentran en regiones de alta precisión.

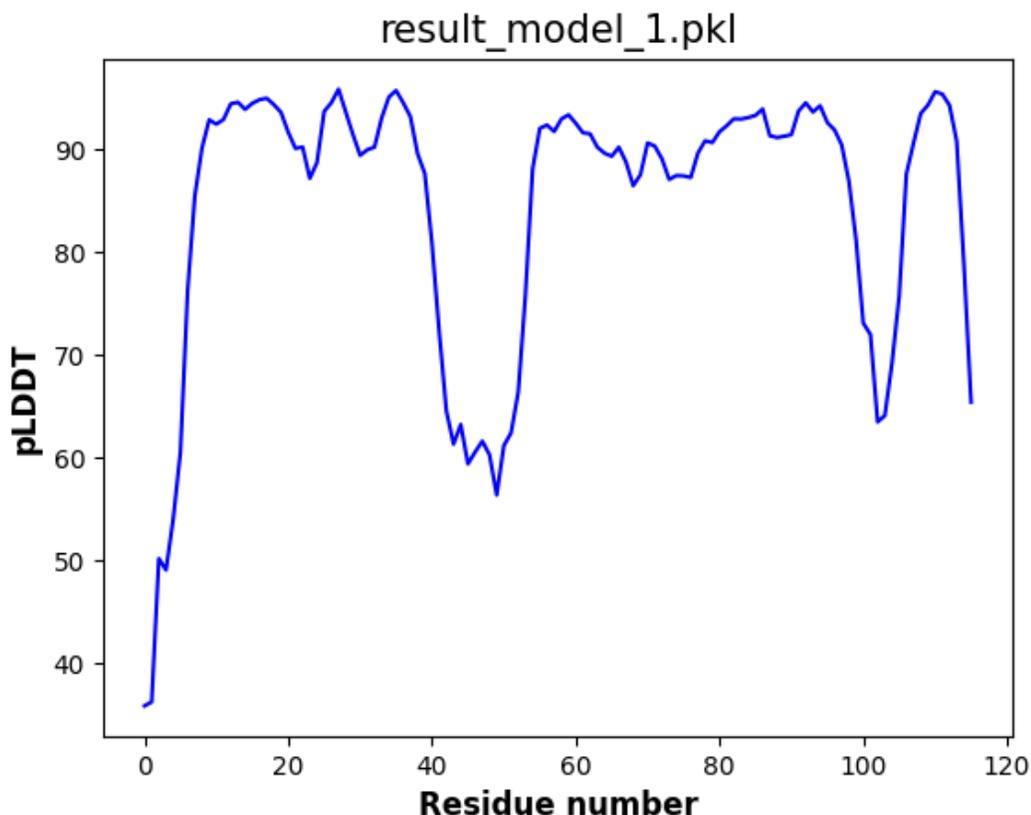


Figura 4-20 Gráfica pLDDT de modelado de la proteína IRF6 en AlphaFold2.

4.5.2 Modelo dominio de unión a ADN de IRF6 con pérdida de los exones 1-3

Se muestra el modelo de estructura del dominio de unión a ADN de IRF6 con la pérdida de los exones 1-3, empleando el modelo de predicción descrito, para la construcción se emplearon bases de datos de variantes estructurales como database of genomic variants y Genome Data Viewer, donde se verificó que la deleción relatada corresponde a la remoción de los primeros 58 aminoácidos que hacen parte del exón 3. Se visualiza el modelo en la figura 4-21.

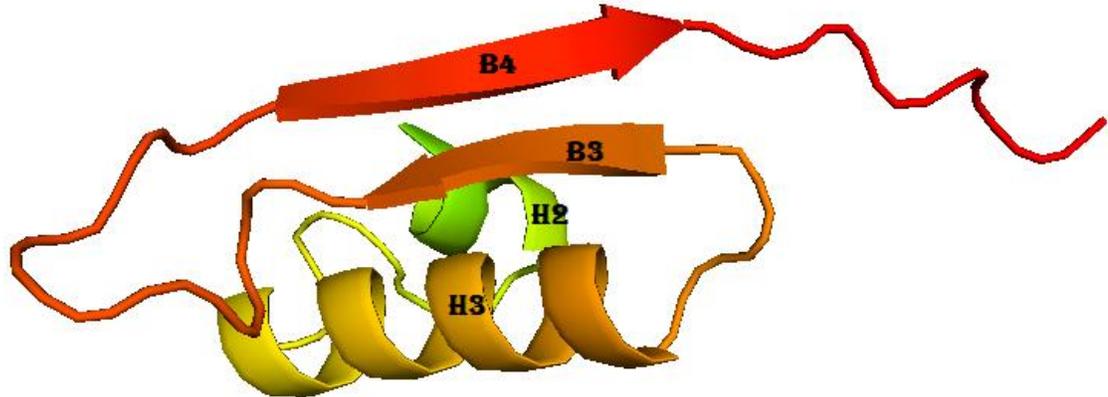


Figura 4-21 Modelo de dominio de unión a ADN con pérdida del exón 3 en IRF6. Visualizado por PyMOL. Como se ilustra en la gráfica la pérdida del exón 3 en el dominio de unión a ADN del IRF6 incluye la pérdida de la hélice alfa H1 y gran parte de la hélice H2 y de las láminas beta $\beta 1$ y $\beta 2$, la H2 es clave en el reconocimiento e interacción con el ADN.

4.5.3 Modelo ISRE

La estructura de la región ISRE del ADN que interactúa con los miembros IRF fue obtenida a partir de plantillas de experimentos previos mediante cristalografía por rayos X con integrantes de la familia IRF, en este caso se tomó el PDB 700T en el que se encontraron tres estructuras dos cadenas del dominio de unión a ADN de IRF4 y una región del ADN denominada ISRE. Mediante PyMOL se editó el formato PDB, se eliminaron las cadenas del IRF4 y creó un archivo nuevo constituido solo por la secuencia de ADN figura 4-22. La evaluación de la estructura obtenida fue realizada por MolProbity en la tabla 4-5.

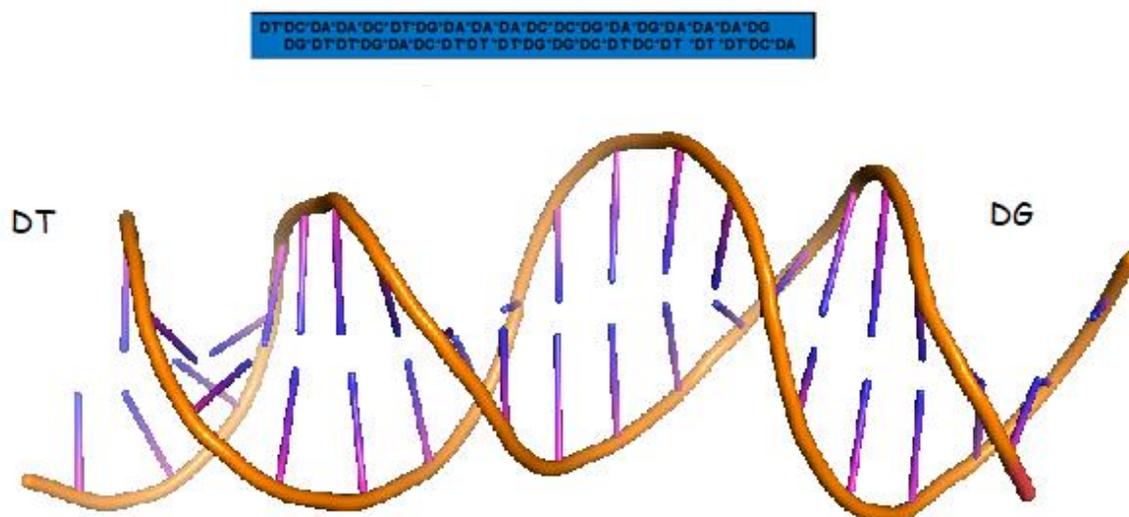


Figura 4-22 Región ADN denominada ISRE. Visualizada en PyMOL. En el cuadro azul se muestra la secuencia de nucleótidos.

Tabla 4-5 Evaluación de la estructura por MolProbity de la estructura ISRE.

	Malos lazos:	0 / 918	0,00%	Meta: 0%
Geometría de ácido nucleico	Malos ángulos:	0 / 1412	0,00%	Meta: <0.1%
Validaciones adicionales	Valores atípicos de volumen quiral	0/160		

*Resultado obtenido de MolProbity, base de referencia de evaluación de estructuras. ¹⁹⁸ No hay evidencia de malos ángulos ni valores atípicos con relación al volumen quiral en la estructura de ADN obtenida.

4.5.4 Modelo de proteína completa de IRF6

Para evaluar la interacción de IRF6 con otras proteínas relacionadas al desarrollo de fisuras orofaciales se hizo el modelamiento de la estructura completa de IRF6

mediante AlphaFold2, se eligió el modelo con mejores características estereoquímicas, se empleó refinamiento con el programa GalaxyRefine.¹⁹⁷ A continuación, se muestra la estructura 3D de la proteína. (Figura 4-23)

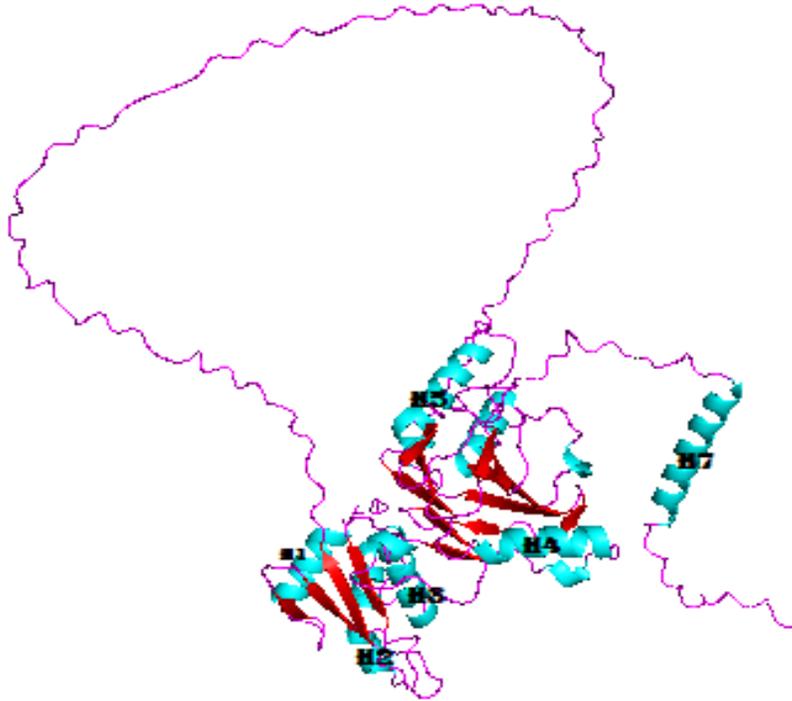


Figura 4-23 Modelo de IRF6 obtenido por AlphaFold2. Visualizada en PyMOL. Están presentes: H1 (residuos 10-20), β 1 (residuos 27-30), β 2 (residuos 35-39), H2 (residuos 49-64), H3 (residuos 76-89), cuatro β 3 (residuos 93-97) y β 4 (residuos 108-113), β 5 (residuos 226-232), β 6 (residuos 241-243), β 7 (residuos 248-252), β 8 (residuos 272-275), H4 (residuos 285-297), β 9 (residuos 302-307), β 10 (residuos 310-315), β 11 (residuos 321-324), β 12 (residuos 342-347), H5 (residuos 348-359), β 13 (residuos 371-375), β 14 (residuos 390-396), H6 (residuos 397-407), H7 (residuos 428-447). Las alfa hélices se representan con el color azul, las láminas beta con el rojo y los bucles en violeta.

4.5.4.1 Verificación del modelo de la proteína IRF6

Se realizó la evaluación de la estructura 3D obtenida de IRF6 con las siguientes estrategias:

1.Evaluación de modelo con herramientas SAVE

Usando las herramientas del servidor SAVE, se ejecutaron los análisis en PROCHECK y ERRAT. El gráfico de Ramachandran (Ver figura 4-24) mostró que el 95,2% de los aminoácidos de la estructura 3D se localizaron en la región más favorable, donde los ángulos phi y psi se encuentran más estables y hacen parte de las hélices alfa y laminas beta, identificados en color rojo. La densidad de los puntos reconoce una mayor conformación de la hélice alfa dextrógira, seguida por lámina beta y la hélice alfa levógira. El 4,8% de los residuos se encontraron en la región favorable en color amarillo y no hay residuos en las zonas menos favorecidas que corresponden a los colores abano y blanco, exceptuando los residuos de glicina que se identificaron con forma de triángulo, se evidencian según las estadísticas 26 glicinas y 47 prolinas.

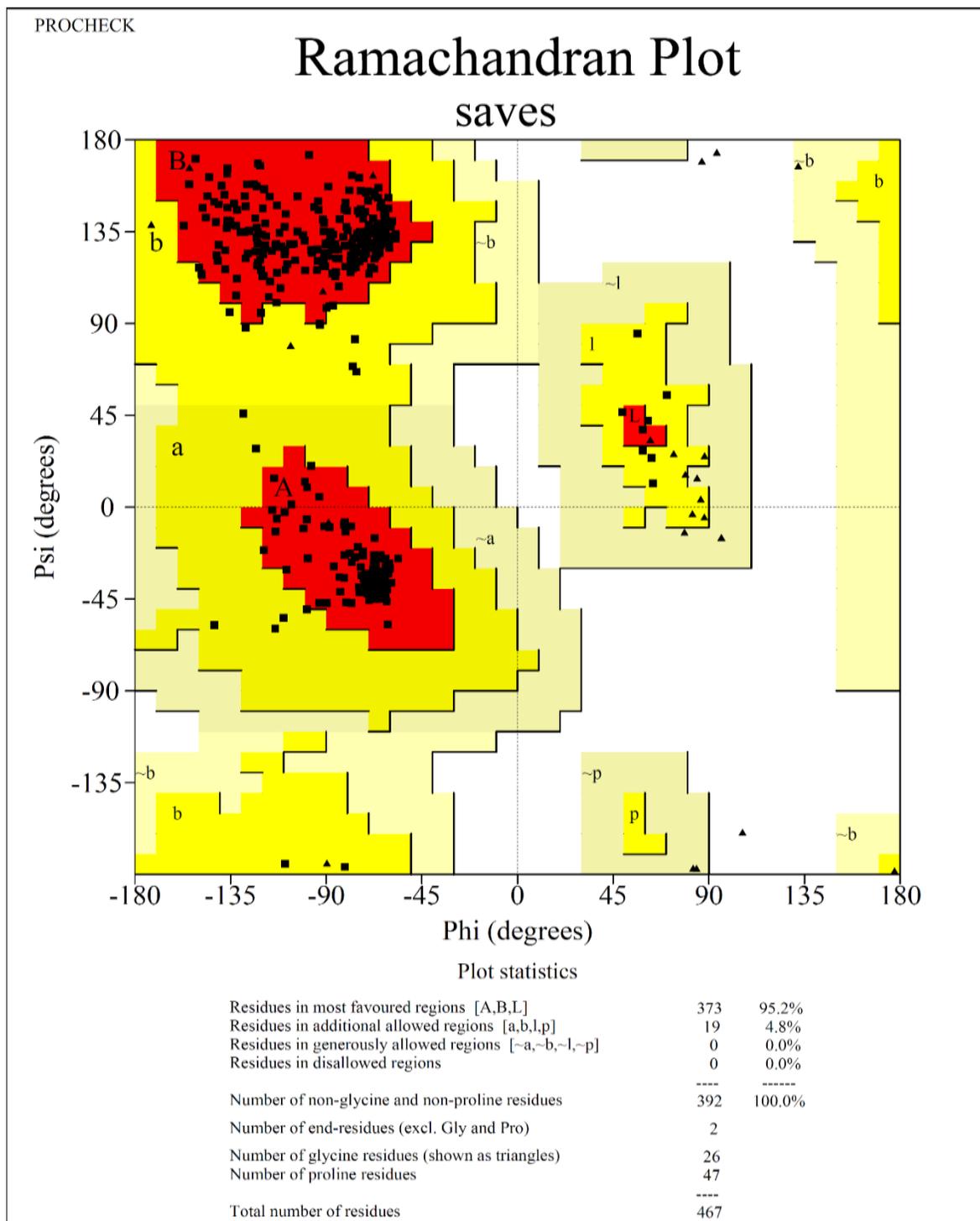


Figura 4-24 Gráfica de Ramachandran del modelo de la proteína IRF6. Obtenida de resultados PROCHECK. 165

Con el servidor ERRAT se evidenciaron los siguientes hallazgos: (Figura 4-25) se trazan errores en los aminoácidos 42-43, 207-210, 262, 227-229, 233, 360-362, 364-366, 375, 378-379, 457, 458 y 460. Teniendo en cuenta el porcentaje total de todos los residuos alcanzó un nivel de confianza del 94.366%.

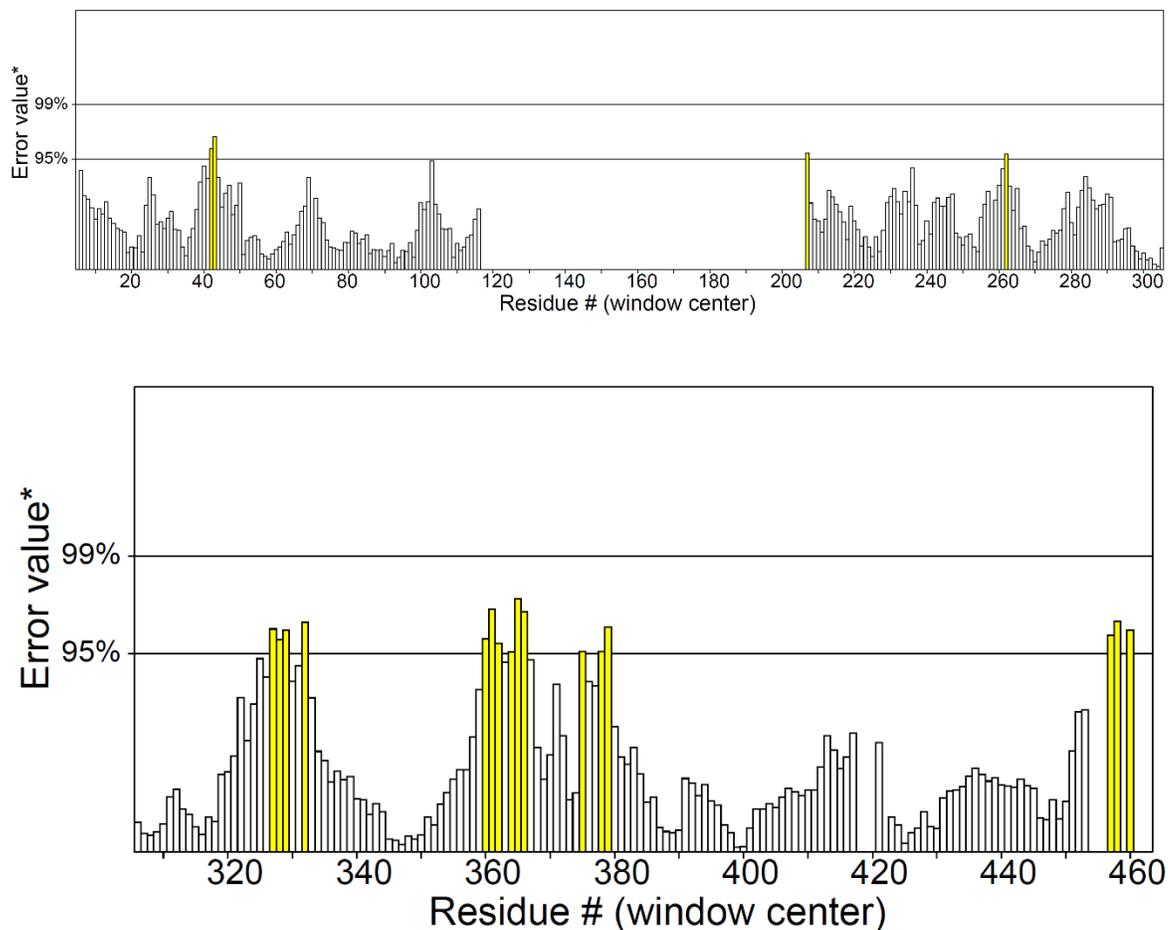


Figura 4-25 Gráfica ERRAT. Obtenida del análisis ERRAT. ¹⁶⁵

4. Descripción topológica

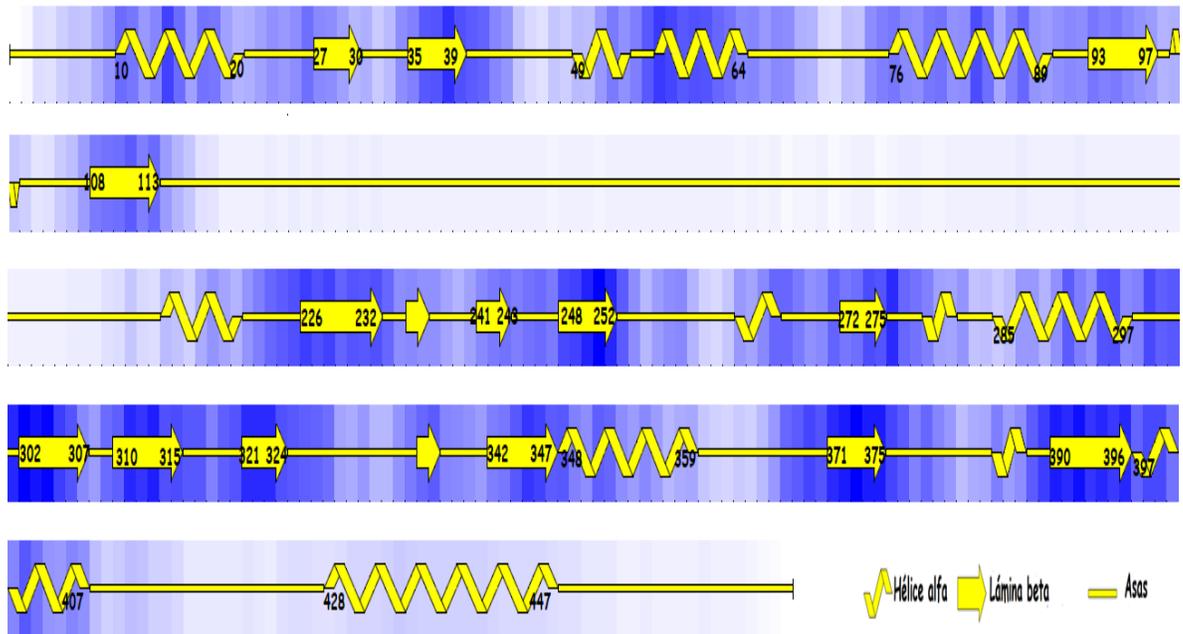


Figura 4-26 Descripción topológica de la proteína IRF6. Obtenida en SAVE.¹⁶⁵ Las estructuras no enumeradas consisten en aquellas que, al ejecutar los algoritmos de predicción de estructura secundaria, no eran definitorias para una hélice alfa o una lámina beta.

3. Cálculo de la RMSD

Usando la herramienta PyMOL, se analizó el RMSD entre la plantilla del PDB 3dsh del IRF5 que guarda mayor similitud con IRF6 en un 60,15% y la estructura 3D que se obtuvo de IRF6. Se encontró un valor de 0.793, se debe tener en cuenta que la plantilla empleada sólo contiene el dominio de unión a proteína y tiene ausente el dominio de unión a ADN y las regiones que corresponden a los exones 5 y 6, no hay en el PDB un modelo de proteína completa de ningún miembro de los IRF. (Figura 4-27)

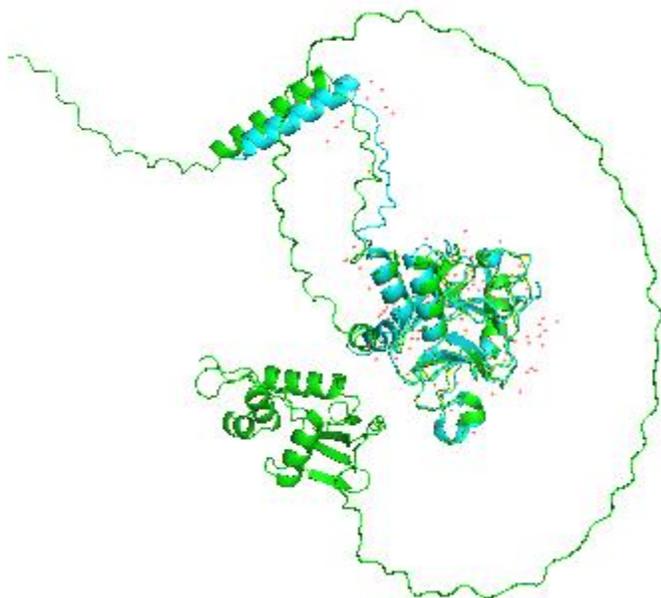


Figura 4-27 Modelo de la proteína IRF6. Obtenido por AlphaFold2.

4.6 Ejecución docking

Para dar mayor claridad de la importancia en el funcionamiento biológico del dominio de unión a ADN de la proteína IRF6 se realizaron dos ejecuciones docking entre proteína y ácido nucleico mediante HADDOCK y HDOCK, al final se seleccionó el modelo obtenido por la herramienta HDOCK, se usó el modelo IRF6 del dominio de unión a ADN y la secuencia de ADN ya descritos. Se empleó la plantilla PDB 7JM4 ¹⁹⁹ obtenida del experimento por cristalografía de rayos X entre el complejo homodímero de IRF4 e ISRE. Para el docking en HDOCK se hicieron tres ejecuciones, la primera con la cadena A del dominio de unión de ADN de IRF6, la segunda con la cadena B del dominio de unión a ADN y la tercera con la cadena del dominio de unión de ADN con la delección del exón 3 y todos acoplados a la región del ADN denominada ISRE. Cuando se ejecutó HDOCK, ¹⁵² se tuvo en

cuenta: el uso de las plantillas del ligando y el receptor en formato PDB ya mencionadas, la adición de los residuos del sitio de unión al receptor (para la cadena A: 82:A ,84:A, 85:A, 88:A, 90:A y la cadena B: 42:B, 82:B, 84:B, 85:B, 88:B, 90:B) y el ligando (Para la cadena A: 5:C, 6:C, 7:C, 8:C, 10:C, 10:E, 11:E, 12:E 13:E, 14:E y la cadena B: 12:C, 13:C, 14:C, 15:C, 16:C, 5:E, 6:E, 7:E), además de algunas restricciones de distancia de residuo obtenidas en HADDOCK, ¹⁴⁶ así se ejecutó el docking por HDOCK. (Tabla 4-6) A continuación se muestra los datos obtenidos.

4.6.1 Resumen modelos docking

Tabla 4-6 Resumen plantilla empleada para los acoplamiento HDOCK.

Molécula	PDB ID	Cadena ID	Longitud de alineamiento	Cobertura	Secuencia similaridad (%)
Receptor	<u>7JM4</u>	A	105	0.963	45.7
Ligando	<u>7JM4</u>	D	18	0.450	100.0

*Resultados ejecución HDOCK. ¹⁵⁰ Se empleó la misma plantilla para las ejecución de los tres modelos de acoplamiento molecular, en la tabla se muestra la similitud con el receptor y el ligando de la plantilla 7JM4 y los modelos obtenidos por medios computacionales.

De los 100 modelos proporcionados por HDOCK se seleccionó uno para cada ejecución, es decir, el de mejor puntaje, acoplamiento y confianza según los parámetros del servidor. Enseguida se muestran los hallazgos encontrados. (Tabla 4-7), (Figuras 4-28,4-29)

Tabla 4-7 Resumen métricas de evaluación HDOCK de modelos ejecutados.

Rango	Modelo	Modelos	Modelo	Modelo
	cadena A	Cadena B	exón 3	delección
Puntaje de acoplamiento	-284.35	-253.06	-156.57	
Puntuación de confianza	0.9363	0.8871	0.5328	

*Obtenido de HDOCK.¹⁵⁰

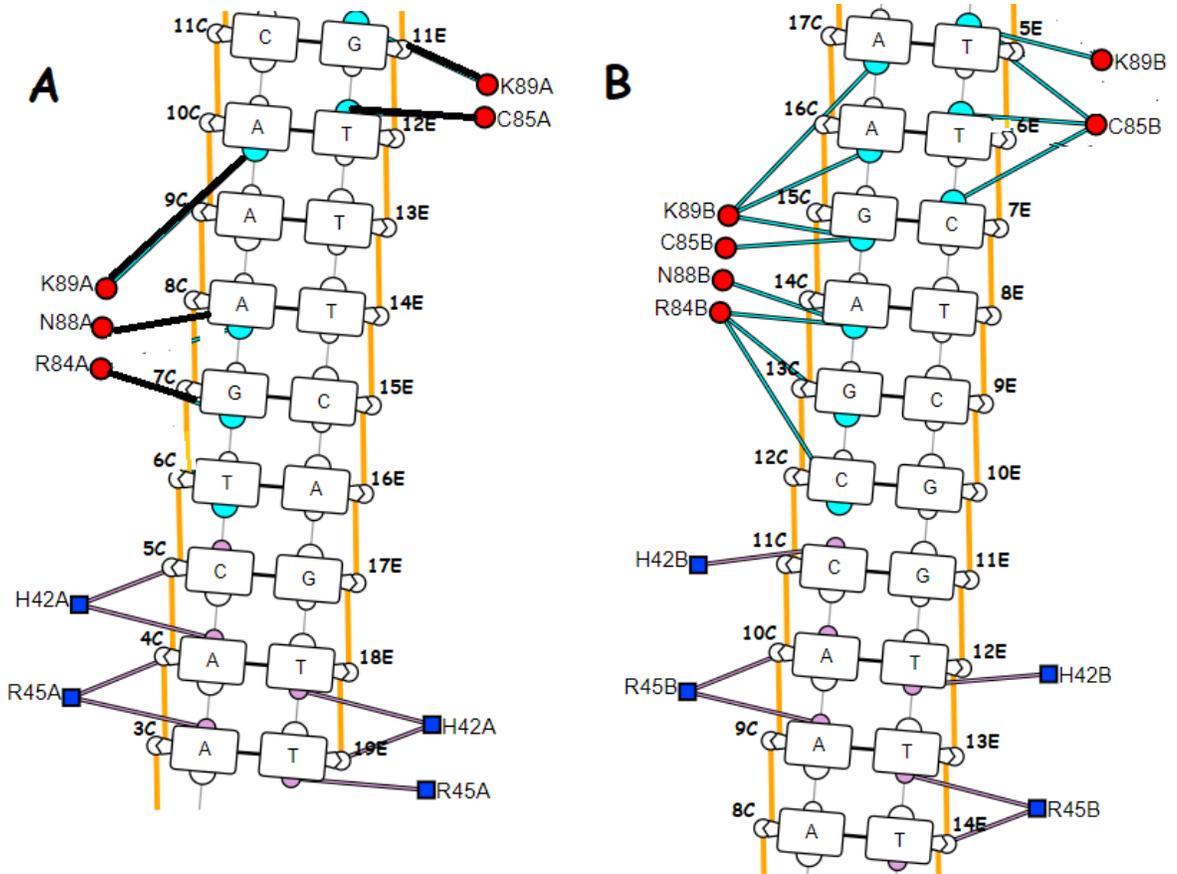


Figura 4-28 Mapa de contacto de residuos de las predicciones complejo ADN-Proteína de las cadenas A y B de IRF6. Obtenidos con HDOCK. ¹⁵⁰ Realizado usando DNAproDB. ¹⁶⁶ **A.** Mapa de contacto entre la cadena A del dominio de unión a ADN e ISRE, interacción de los residuos de la helice alfa 3: K89, C85, N88 y R84 que interactúan con las cadenas C y E del surco mayor. Residuos del loop 4: H42 y R45 se unen al surco menor. **B.** Mapa de contacto entre la cadena B del dominio de unión a ADN e ISRE, interacción de los residuos de la helice alfa 3: K89, C85, N88 y R84 que interactúan con las cadenas C y E del surco mayor. Residuos del loop 4: H42 y R45 se unen al surco menor.

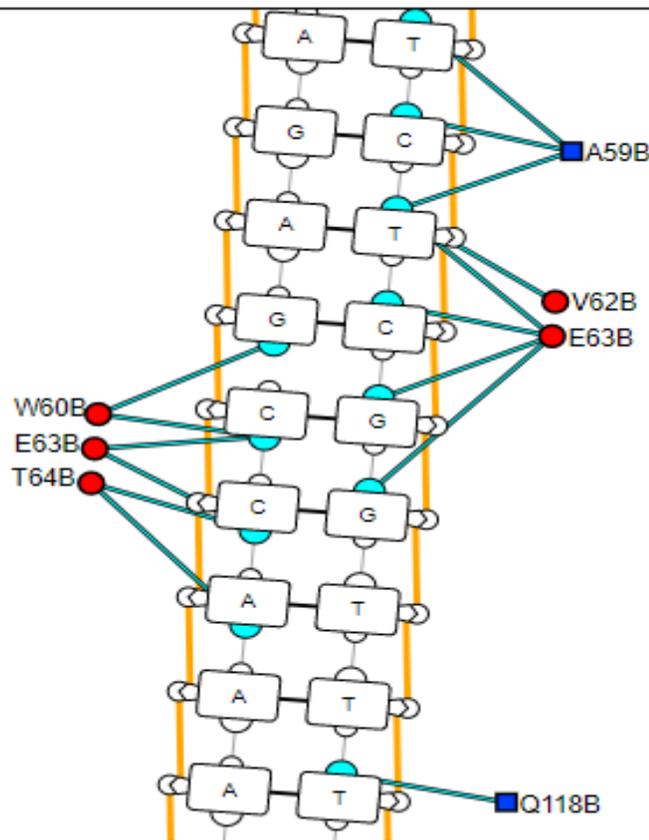


Figura 4-29 Mapa de contacto de residuos de las predicciones complejo ADN-Proteína de la cadena con la delección del exón 3. Los residuos que hacen contacto con el surco mayor pertenecen a la helice alfa 2: A59, W60, E63, T84 V62 y Q118.

4.6.2 Resumen de interacciones modelos docking

Los principales residuos de IRF6 identificados en este estudio que median el contacto con el surco mayor del ISRE fueron: K89, N88, C85 y R84 ubicados en la hélice alfa 3. También se identificaron otros aminoácidos como el K109, Trp40, Ala43, Lys 66, Lys 78 y la Lys 80, como se muestra en las figuras 4-30 a 4-35. La Lys 89 (B) forma un enlace de hidrógeno entre su grupo amino y el oxígeno 4 de la timina de la timidina 5 de la cadena E (dT5(E)), además tiene interacciones hidrofóbicas sobre la dA16(C) que hace parte de la segunda posición de la secuencia central (GAAA), y la Lys89 (A) ejerce fuerzas hidrofóbicas sobre dG10(E) y dA10((C). La Asn 88 (B) interactúa por puente de hidrógeno mediante su grupo amino y el oxígeno 2 del fosfato de la dA14(C) y la Asn 88 (A) se une por enlace de hidrógeno a través de su grupo amino al oxígeno 2 del fosfato de la dA8(C) que pertenece a la segunda posición de la adenina de la secuencia central (GAAA). La Cys 85 (B) a través de su grupo sulfhidrilo forma un enlace de hidrogeno con el oxígeno 4 de la timina de la dT6(E) y ejerce interacciones hidrofóbicas sobre dT5(E) y dT6(E). La Arg 84 (B) crea dos enlaces de hidrógeno con su grupo guanidino con el oxígeno 2 del fosfato de la dG13(C) y la Arg 84 (A) interacciona con dos enlaces covalentes externos dados por el NE y el CD sobre el oxígeno 2 del fosfato de la

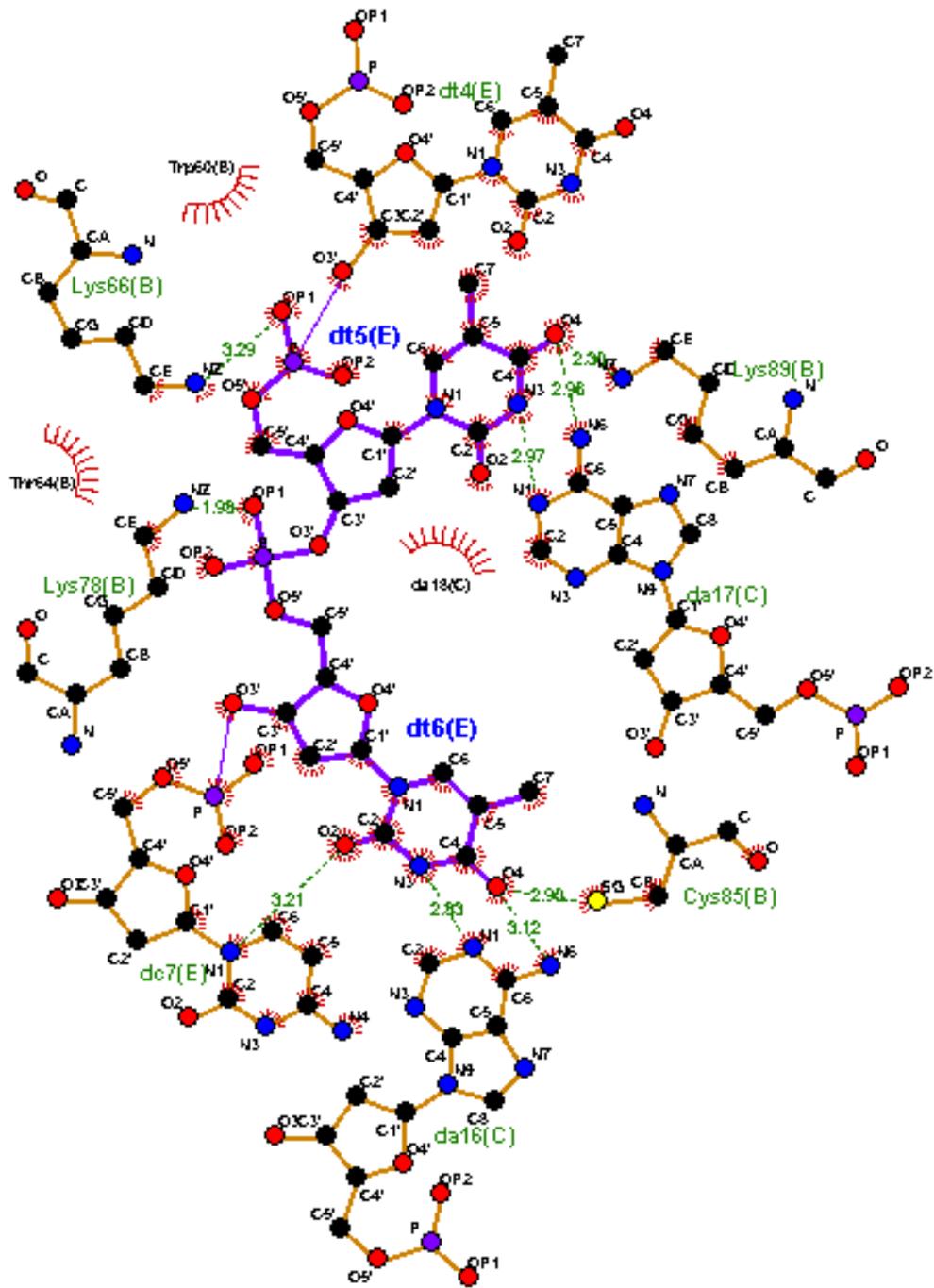


Figura 4-31 Interacción de residuos Cys 85 y Lys 89 (B), complejo dominio de unión a ADN de IRF6 y ADN. Visualizado por Ligplot.²⁰⁰ Se evidencia la unión mediante

puentes de hidrógeno de las Lys 66 y 78 con la dT5 y dT6 respectivamente con los oxígenos de sus grupos fosfato. Además se muestra las interacciones hidrofóbicas del Thr 64 y la Trp 60 con el dT5(E).

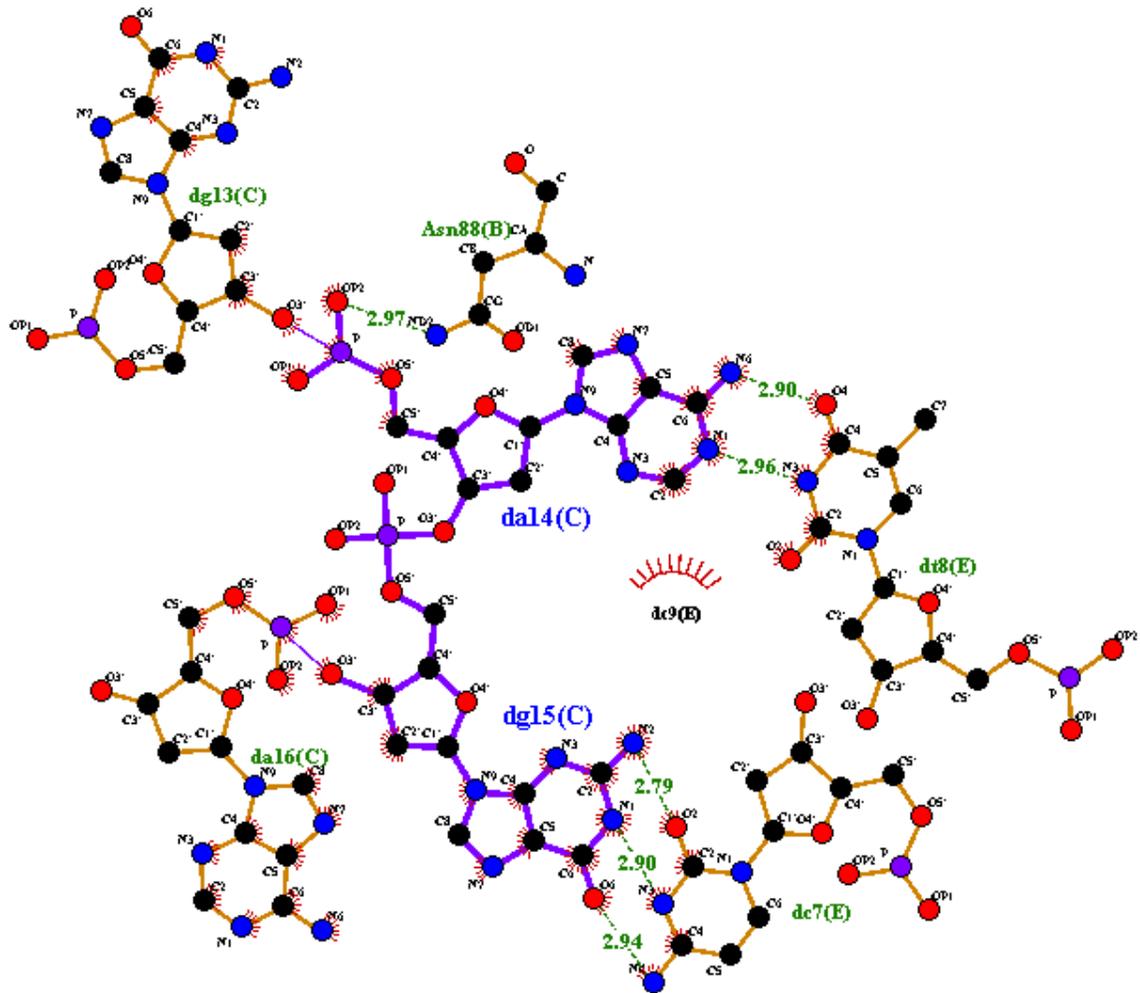


Figura 4-32 Interacción de residuos Asn88 (B), complejo dominio de unión a ADN de IRF6 y ADN. Visualizado por Ligplot.²⁰⁰ Se muestra la interacción entre Asn 88 e dA14(C).

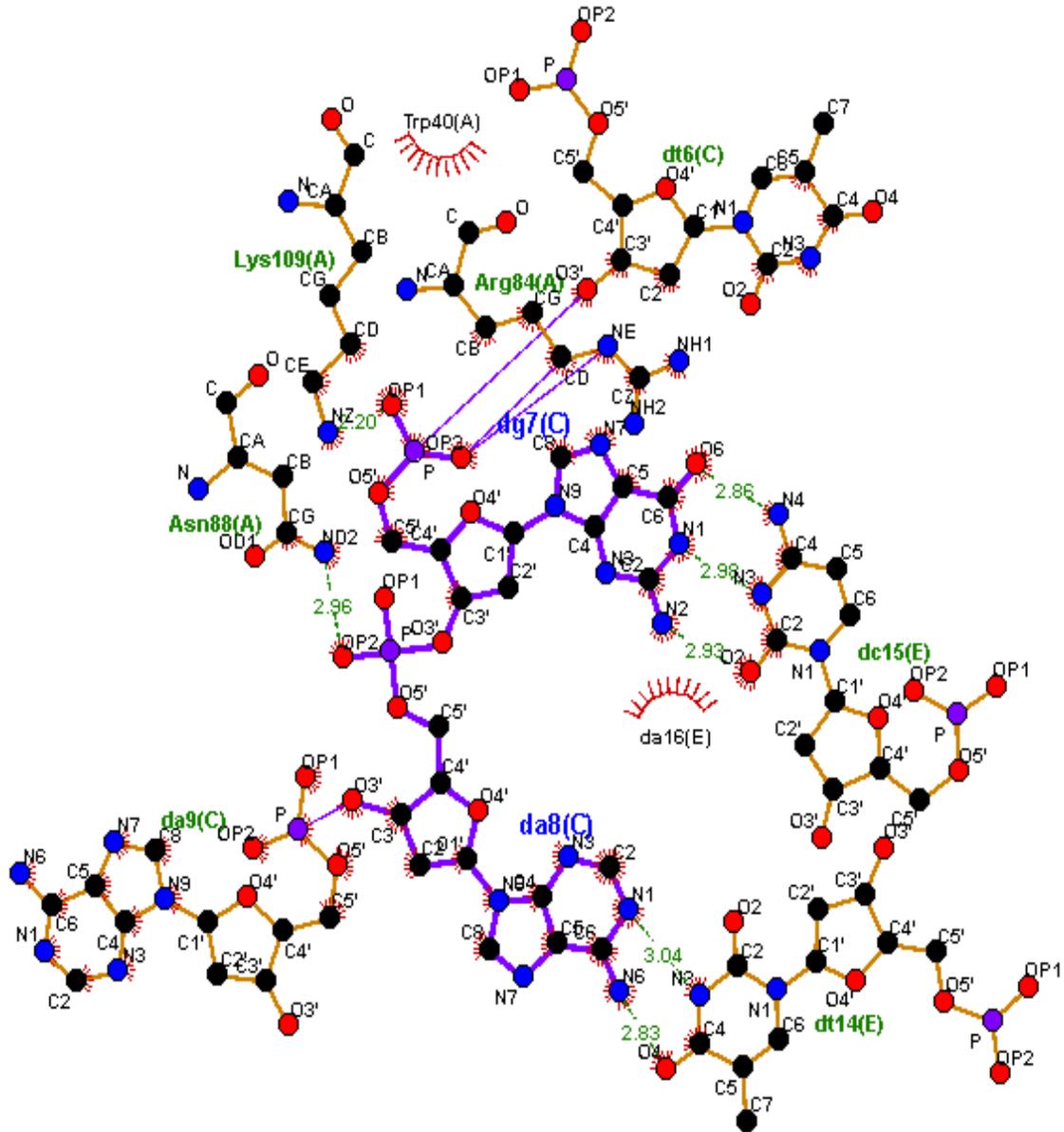


Figura 4-33 Interacción de residuos Asn88, Arg84 y Lys109 (A), complejo dominio de unión a ADN de IRF6 y ADN. Visualizado por Ligplot.²⁰⁰ Se muestra la interacción entre la Lys109(A) mediante puente de hidrógeno con el oxígeno del fosfato de la dG7(C) que hace parte de la primera posición de la secuencia GAAA y de la ASN88(A) con la dA8(C) a través de puente de hidrógeno del amino ND2 y

el oxígeno del fosfato, que se ubica en la segunda posición de la secuencia central ya mencionada.

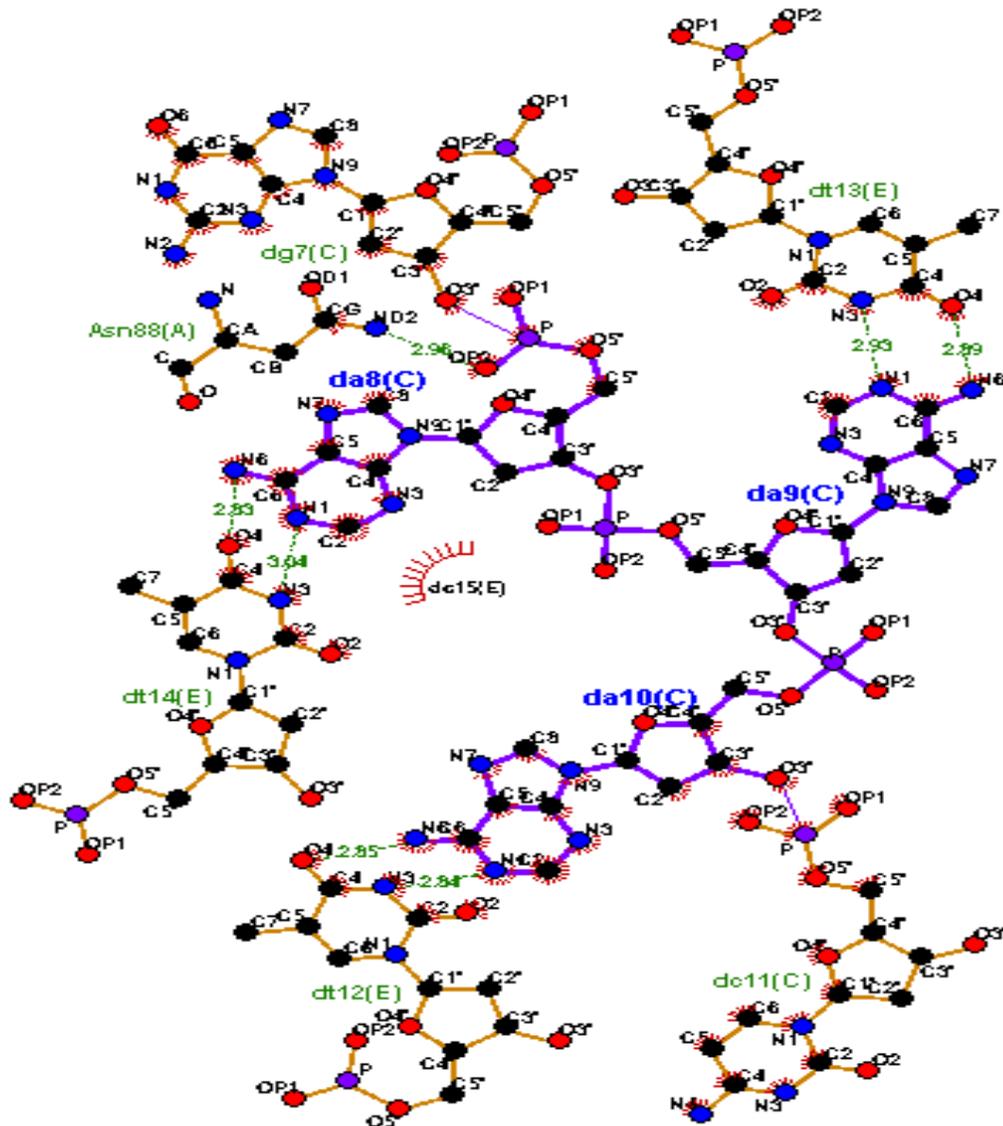


Figura 4-34 Interacción de residuos Asn88 (A), complejo dominio de unión a ADN de IRF6 y ADN. Visualizado por Ligplot. ²⁰⁰

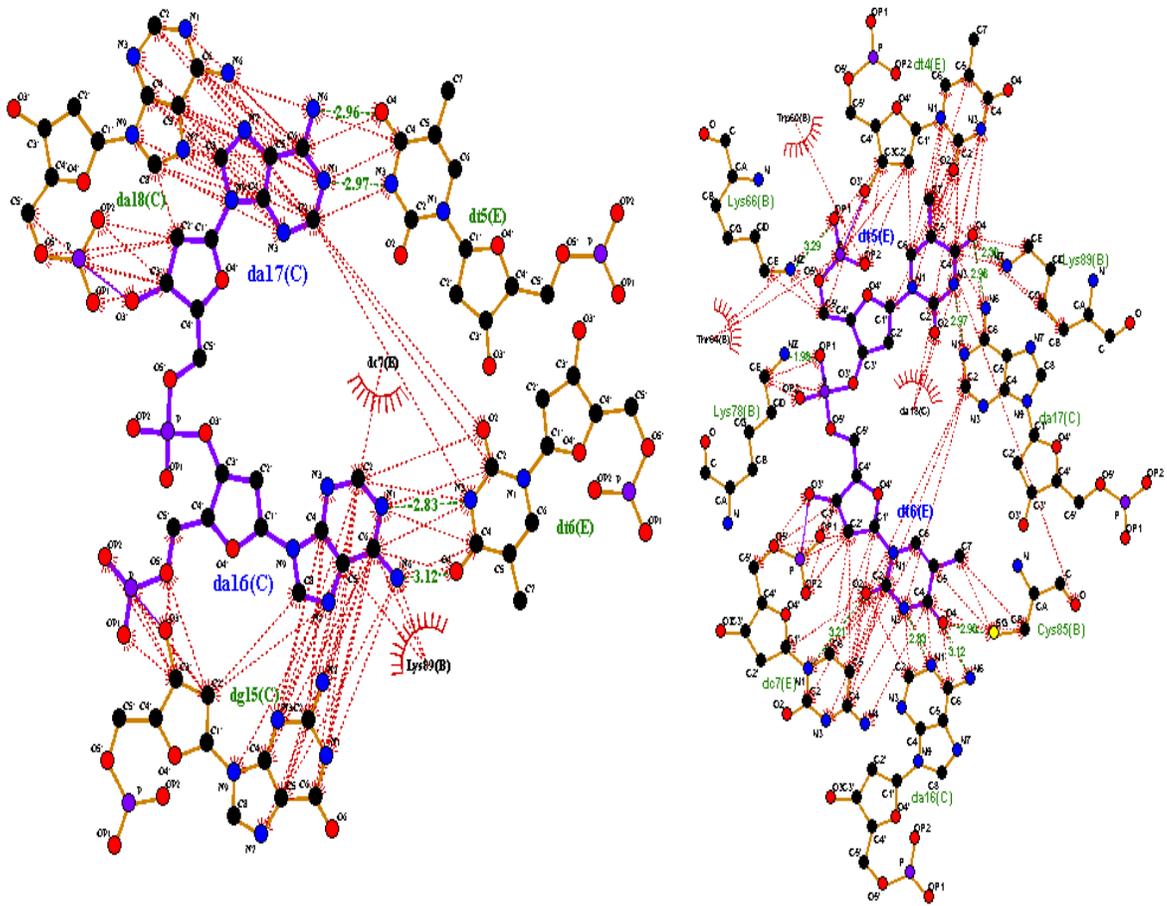


Figura 4-35 Interacciones hidrofóbicas de algunos residuos como Lys89 y Cys85(B). Visualizado por Ligplot. ²⁰⁰

Sobre el surco menor hay interacciones importantes de los residuos His40 Y Ala43 que hacen parte del bucle 4 de IRF6 como se evidencia en la figura 4-36.

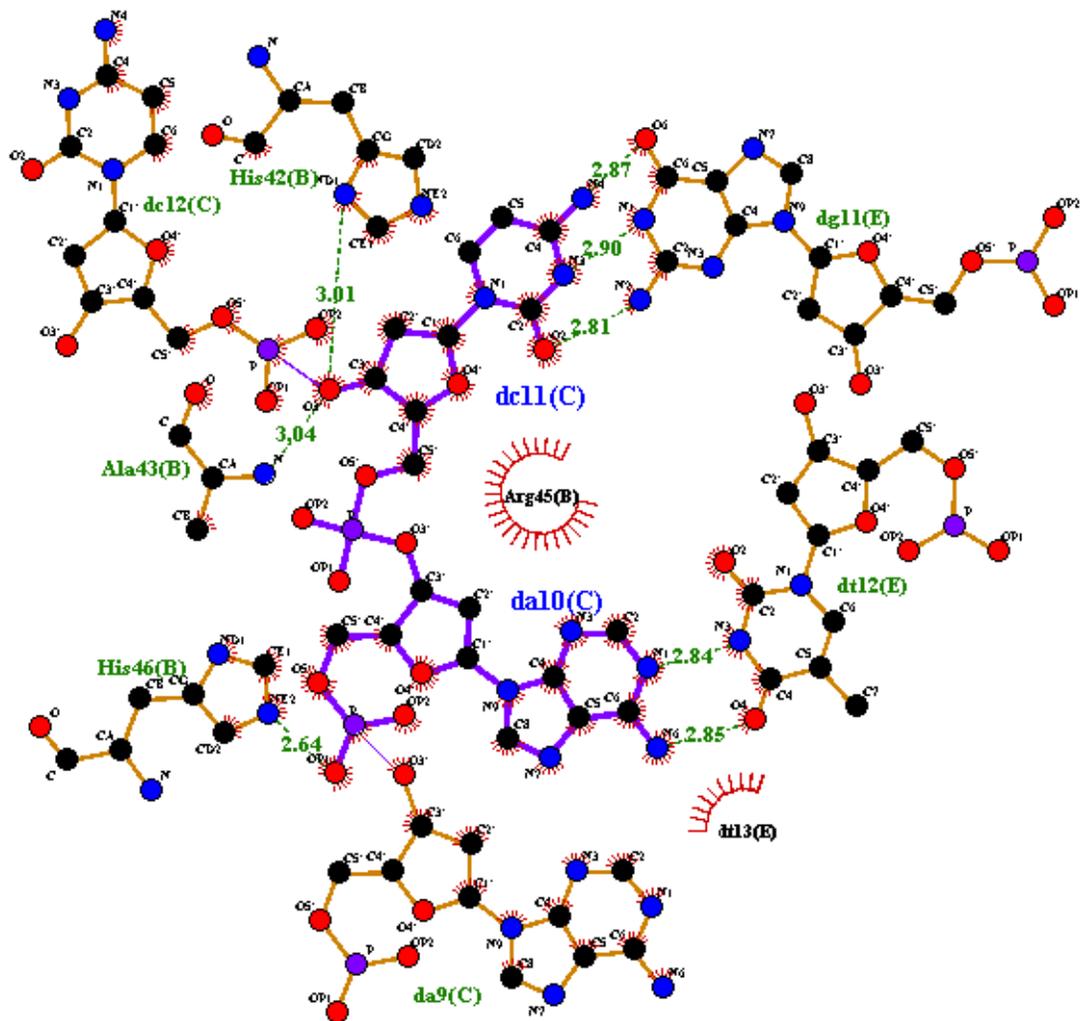


Figura 4-36 Interacciones en el surco menor. Visualizado por Ligplot. ²⁰⁰ Se muestra la interacción de algunos residuos como la His42 y la Ala43 con el dC12(C) del surco menor. También se identifica la interacción de la His46 con la dA10(C), que hace parte de la secuencia central GAAA.

4.6.3 Comparación experimentos previos IRF interacción dominio de unión a ADN y ADN

Se han publicado cuatro estudios de estructura 3D experimentales obtenidos por cristalografía de rayos X, de tres miembros de la familia de los factores del interferón

que evidencian la interacción del dominio de unión a ADN con una región en el ADN denominado elementos de respuesta estimulados por IFN (ISRE) y en los que se han encontrado una interacción de residuos común, el reconocimiento de la secuencia GAAA por parte de la hélice alfa 3 y la presencia de algunos aminoácidos de triptófano en el sitio del contacto es decir en los surcos del ADN, en la tabla 4-8 se resumen los hallazgos identificados en la relación IRF-ISRE y en las figuras 4-37 y 4-38 los mapas de residuos de las estructuras experimentales de IRF1, IRF2 y IRF4.

Tabla 4-8 Resumen experimentos previos estructuras por cristalografía de algunos miembros de la familia IRF.

Modelo IRF/ADN	Método	Principales residuos implicados en la interacción	Referencia
Estructura cristalina del complejo del IRF1-ADN. PDB: 1IF1	Difracción de rayos X	Ser87, Asn86, Arg82, Cys83, Asn80, Lys78, Lys75.	PDBDOI: <u>10.2210/pdb1IF1/pdb</u> ⁷³
Estructura cristalina del complejo IRF-2-ADN. PDB: 2IRF	Difracción de rayos X	Asn86, Cys83, Arg82, Asn80, His40	PD DOI: <u>10.2210/pdb2IRF/pdb</u> ⁷⁵
Estructura de rayos X del dominio de unión al ADN del IRF4 unido a un elemento de respuesta	Difracción de rayos X	Lys103, Asn102, Cys99, Arg98, His56	PDB: DOI: <u>10.2210/pdb7OOT/pdb</u> ²⁰¹

**estimulado por
interferón.
PDB: 7OOT**

Estructura de rayos X del dominio de unión al ADN del factor regulador de interferón 4 unido a un elemento de respuesta estimulado por interferón. PDB: 7JW4	Difracción de rayos X	Lys103, Asn102, Cys99, Arg98, His56	PDB DOI: 10.2210/pdb7JM4/pdb 199
---	--------------------------	--	--

*Construcción propia.

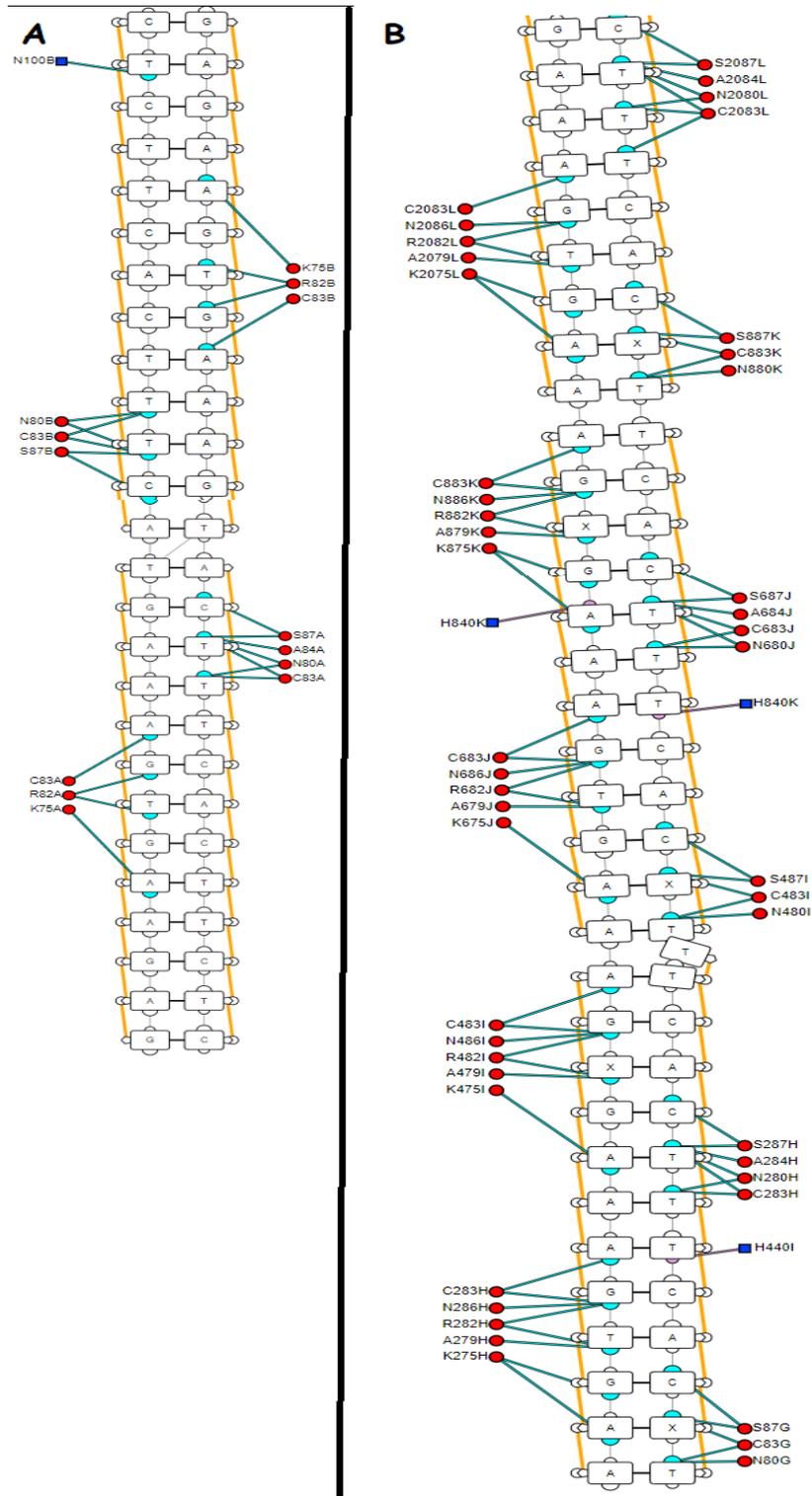


Figura 4-37 Mapas de contacto de residuos de las estructuras experimentales por cristalografía de rayos X, complejo proteína-ISRE (IRF2-IRF 3). Visualizado por

DNAProDB. ¹⁶⁶ A. Estructura IRF1 (1IF1). B. Estructura IRF2 (2IRF).

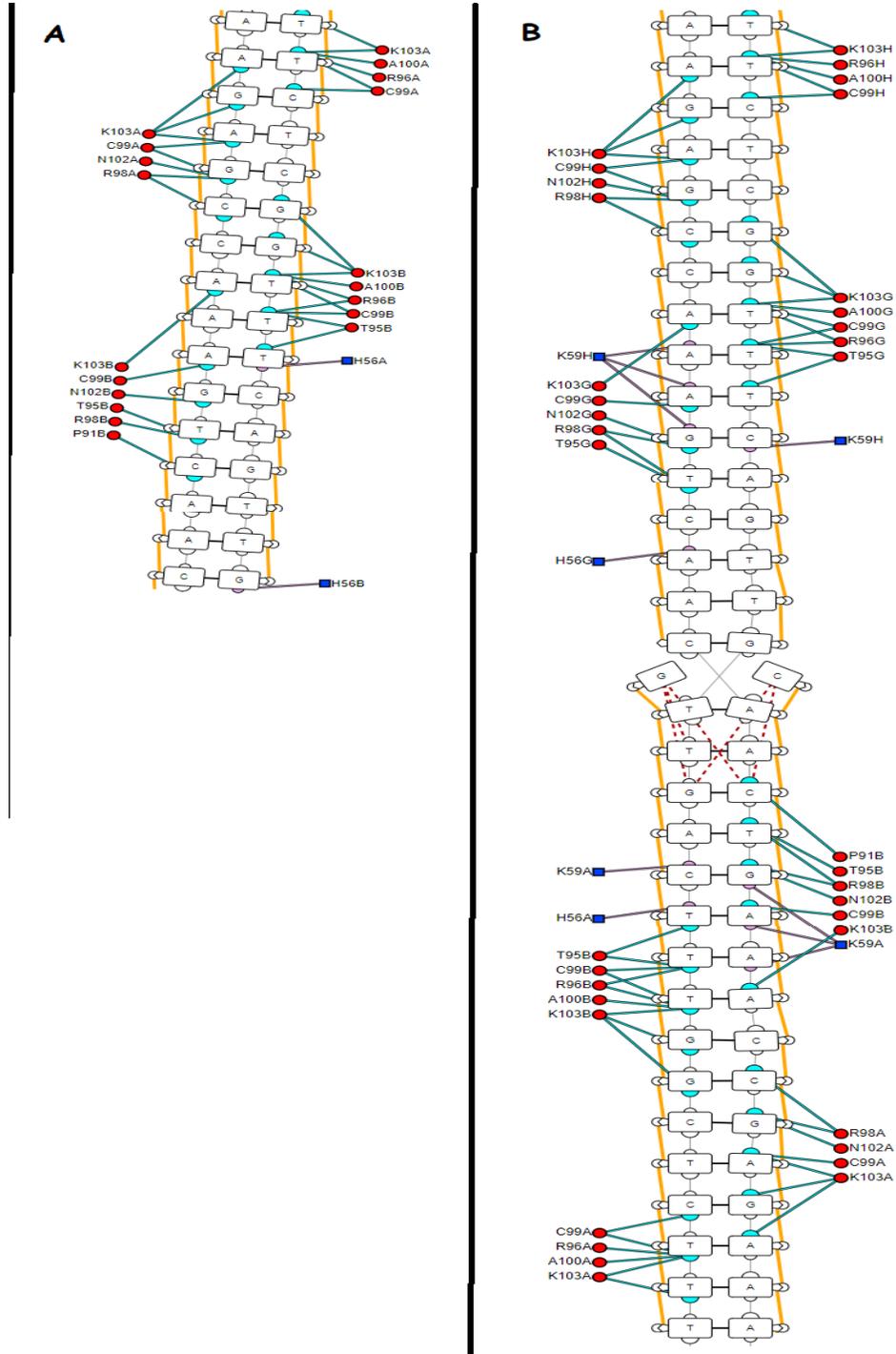


Figura 4-38. Mapas de contacto de residuos de las estructuras experimentales por cristalografía de rayos X, complejo proteína-ISRE (IRF4). Visualizado por DNAProDB. ¹⁶⁶ A. Estructura IRF4 (7OOT). B. Estructura IRF4 (7JW4).

4.7 Red de genes

En este estudio se obtuvieron 28 genes, el consumo de ácido fólico (Vit B9) y dos regiones en el ADN humano asociados a L/PH-NS y al gen IRF6, que constituyen la red de genes creada por este estudio. Todo ello fue resultado de la búsqueda y revisión de bases de datos científicas, de anotación, de vías de señalización, metabólicas y para la predicción de la interacción proteína-proteína en la que se incluyó información de estudios experimentales in vitro, antecedentes estadísticos con un p valor significativo < 0.05 y evidencia de extracción de texto con valores estadísticos significativos y bases de datos curadas, se consiguió la siguiente red de genes, empleando la herramienta Cytoscape ²⁰² (figura 4-39).

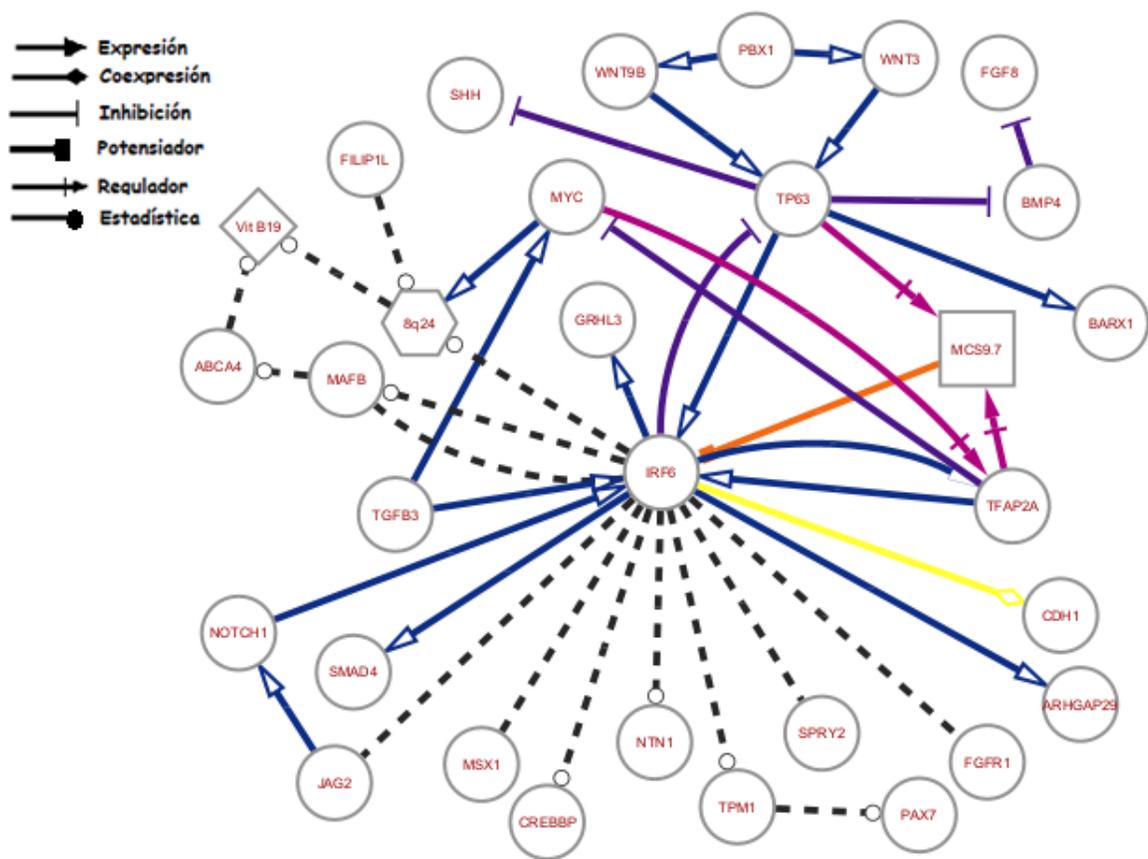


Figura 4-39 Red de genes IRF6- L/PH-NS. Autoría propia, realizado por Cytoscape.

203 El círculo simboliza a los genes, el cuadrado a los potenciadores, el hexágono a regiones en el ADN y los diamantes a los factores ambientales. La coexpresión se representa por el color amarillo, la expresión por el azul, la inhibición por el morado, el regulador por fucsia, el potenciador por el naranja y los datos estadísticos y extracción de texto en gris. Los datos de estudios experimentales se figuran en líneas sólidas y la información no experimental por líneas discontinuas. La forma de la flecha de destino también guarda significado en el grafo, como se evidencia en el recuadro superior al lado derecho.

Empleando la base de datos DAVID de anotación funcional,⁹⁹ se realizó el análisis de los 28 genes involucrados en la red expuesta previamente. En las siguientes figuras y tablas se muestran las anotaciones de atributos en Homo Sapiens como: procesos biológicos, enfermedades y ontología genética. (figuras 4-40 y 4-41).

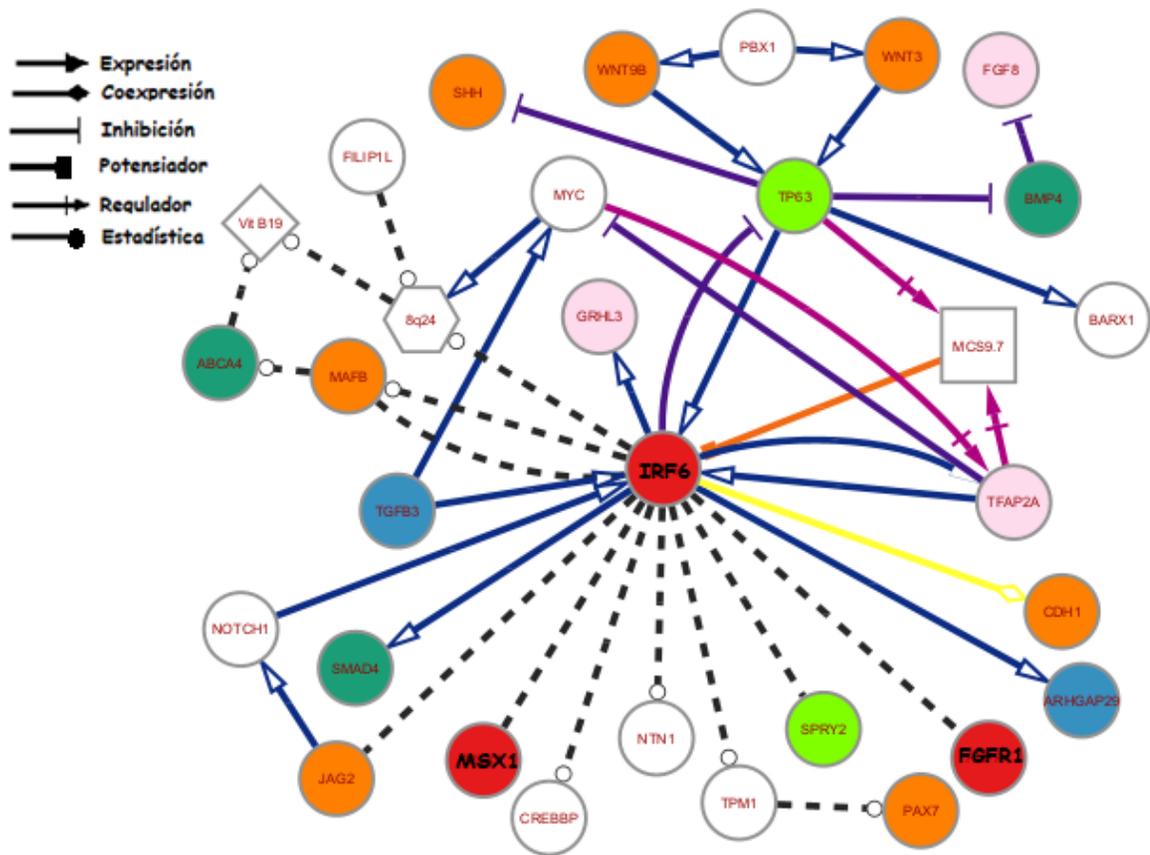


Figura 4-40 Resumen anotación de genes relacionada a enfermedad, DAVID Data Base. Cada color representa un fenotipo relacionado a hendiduras orofaciales, el rojo simboliza los genes que han sido asociados a cuatro fenotipos: LH, PH, LPH e

hipodondia. En verde limón: LPH, LH y PH. En azul: LPH y PH. En verde: LH y LPH. En Morado: LH y PH. En naranja: LPH. En rosado: LH.

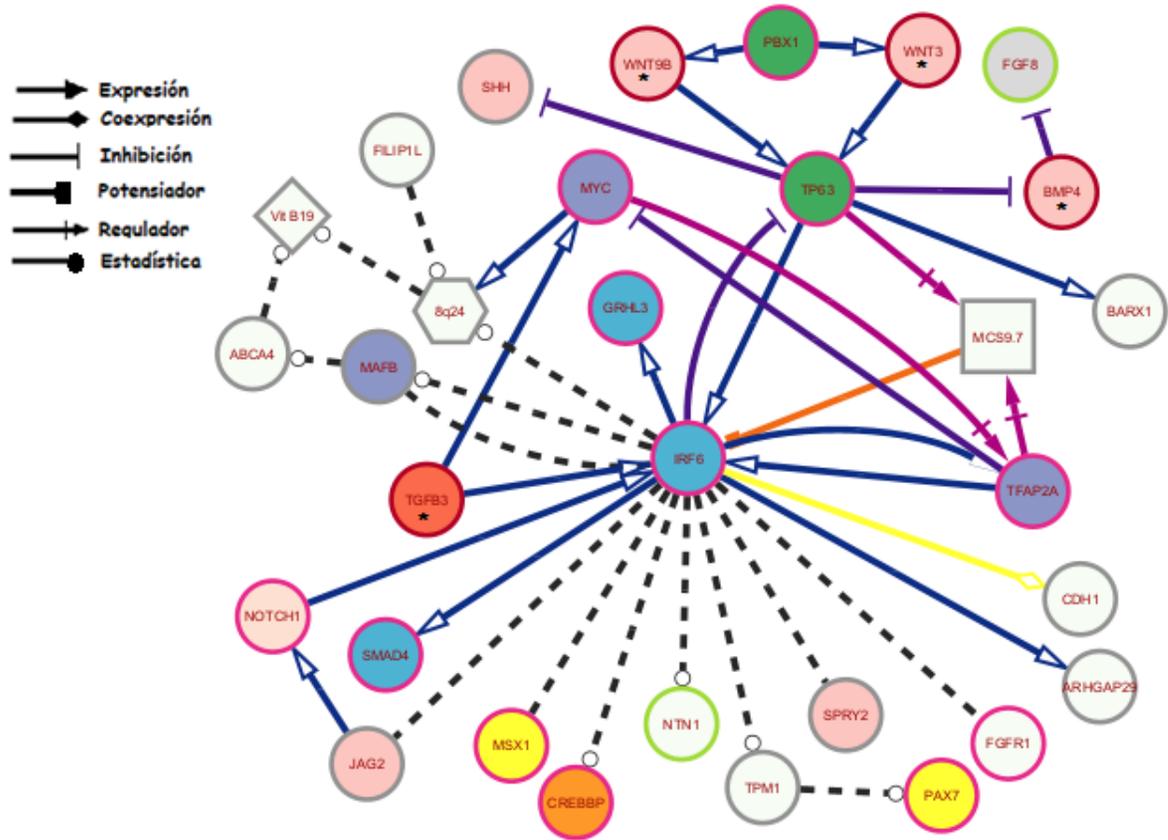


Figura 4-41 Resumen anotación de componentes celulares y función molecular, DAVID Data Base. Algunos genes de la red cumplen funciones moleculares, en este caso, el color del nodo representa un significado, el rosado está asociado al desarrollo de proteínas, el rojo a la activación, el azul con la unión al ADN y el gris a factores de crecimiento. Se han dado otros colores a los genes que comparten funciones como es el caso de los nodos de color verde, involucrados en la activación, el desarrollo de proteínas y la unión a ADN. El violeta representa dos acciones la unión a ADN y la activación. El amarillo al desarrollo de proteínas y a la

unión a ADN. El Vinotinto es un nodo involucrado en tres funciones: desarrollo de proteínas, factor de crecimiento y mitógeno. Con referencia a los bordes de los nodos, indica el componente celular al que pertenece, el fucsia se relaciona a nivel nuclear, el vinotinto a la matriz extracelular y el verde limón a los que son secretados. * Los nodos que tiene el asterisco indican que comparten dos relaciones en los componentes celulares la matriz celular y en los secretados.

Tabla 4-9 Resumen anotación tributo de genes enlistados, DAVID Data Base.

Recurso de ontología genética	Función	Genes	%	P valor
GOTERM_BP_1	Procesos del desarrollo	CREBBP, MYC, PBX1, MAFB, SMAD4, WNT3, WNT9B, CDH1, BMP4, FGF8, FGFR1, GRHL3, IRF6, SPRY2, JAG2, PAX7, SHH, TFAP2A, TP63, MSX1, NTN1, NOTCH1, TGFB3, TPM1, VAX1.	92,6	1,5E-9
GOTERM_BP_1	Señalización	ABCA4, CREBBP, ARHGAP29, MYC, SMAD4, WNT3, WNT9B, CDH1, BMP4, FGF8, FGFR1, GRHL3, SPRY2, JAG2, SHH, TP63, MSX1, NTN1, NOTCH1, TGFB3.	74,1	7,9E-5
GOTERM_BP_1	Regulación biológica	ABCA4, CREBBP, MYC, PBX1, MAFB, ARHGAP29, SMAD4, WNT3, WNT9B, CDH1, BMP4, FGF8, FGFR1, GRHL3, IRF6, SPRY2, JAG2, PAX7, SHH, TFAP2A, TP63, MSX1, NTN1, NOTCH1, TGFB3, TPM1, VAX1.	100,0	9,0E-5
GOTERM_BP_1	Respuesta de estímulo	ABCA4, CREBBP, MYC, ARHGAP29, SMAD4, WNT3, WNT9B, CDH1, BMP4, FGF8, FGFR1, GRHL3, SPRY2, JAG2, SHH, TFAP2A, TP63, MSX1, NTN1, NOTCH1, TGFB3, TPM1, VAX1	85,2	2,3E-4
GOTERM_BP_1	Organización de los componentes celulares	ABCA4, CREBBP, MYC, MAFB, SMAD4, WNT3, CDH1, BMP4, FGF8, GRHL3, SPRY2, SHH, TP63, MSX1, NOTCH1, TGFB3, TPM1, VAX1, NTN1.	70,4	9,6E-4
GOTERM_BP_1	Adhesión biológica	BMP4, CDH1, NTN1, NOTCH1, SHH, TPM1.	22,2	5,8E-2

4.8 Segregación de variantes asociadas con L/PH-NS

La variante identificada asociada a hendiduras orofaciales se identificó en el trío de la familia cinco, dicha variante ubicada en las siguientes coordenadas genómicas seq[GRCh37]chr1:209969954-209980303x1 con clasificación patogénica, se segregó de madre a hija, cada una con fenotipos diferentes, madre con dos microformas, torus palatino y agenesia de incisivos laterales superiores, e hija con LHI-PH, esta expresividad variable es una característica conocida del gen.

5. Discusión:

Este es un proyecto de impacto, con repercusiones positivas en comunidades vulnerables, como es el caso de la región de Cibao, al norte de República Dominicana, la cual presenta alta incidencia de L/PH. Adicionalmente, ha permitido sumar información científica relevante relacionada con la anomalía, resaltando la importancia de realizar este tipo de investigaciones, en donde observamos la trascendencia de practicar un examen clínico minucioso con posteriores pruebas moleculares pertinentes para determinar diagnósticos adecuados, programar tratamientos certeros y brindar asesoramientos genéticos con efectos favorables en las familias y la sociedad en general. El estudio tuvo como propósito, además, determinar la estructura y funcionalidad de variantes (SNV, Indels, CNVs) asociadas con L/PH-NS en cinco familias dominicanas con herencia autosómica dominante, a través de secuenciación masiva y análisis computacional.

5.1 Caracterización de las familias

En el estudio se identificaron cinco familias con L/PH-NS con herencia autosómica dominante. Mediante un instrumento de recolección de datos se recaudó información relacionada con factores ambientales, encontrando componentes en su mayoría por la línea paterna que corresponde al consumo de cigarrillo, alcohol y

contacto con plaguicidas, sin embargo, son factores que no han sido comprobados como riesgo para el desarrollo de L/PH. No obstante, dos elementos en las madres como ser fumadora pasiva y el no consumo de ácido fólico en la gestación sí se han relacionado al desarrollo de fisuras orofaciales. En un estudio publicado en el 2015 en China se identificaron varios factores con valores estadísticamente significativos, CI 95%, $p < 0,001$, se describieron como elementos de riesgo durante el embarazo el cigarrillo (OR 15,36), ser fumador pasivo (OR 4,05) y la exposición a pesticidas (OR 7,84) en la madre.⁴⁹ Los datos del actual estudio podrían no ser concluyentes, teniendo en cuenta la baja cantidad de individuos estudiados, pero sí se pueden sumar a investigaciones en esta región relacionadas con factores sociodemográficos como es el caso de un estudio recién publicado por Arévalo et al.¹² donde identificaron que la población más afectada con L/PH-NS hace parte de familias con clase socioeconómica baja y vulnerables al uso de agroquímicos; de los cinco grupos familiares de la presente investigación, tres tenían una percepción de estrato socioeconómico medio-bajo y dos familias estuvieron en contacto con plaguicidas durante la gestación. En el estudio mencionado, indicaron que el consumo de alcohol fue el elemento de riesgo más frecuentemente relacionado en los pacientes con L/PH-NS con herencia multifactorial; sin embargo, no se reportaron factores ambientales comúnmente asociados con L/PH en las familias con herencia autosómica dominante, por lo cual se postuló para este tipo de herencia una mayor influencia genética.¹²

La variante asociada a L/PH-NS fue identificada en la familia cinco, en la madre y la probando del trío, consiste en la delección de una copia de los exones 1-3 del gen IRF6 con clasificación patogénica, no reportada previamente en la literatura científica, un gen asociado a entidades sindrómicas y no sindrómicas, además al desarrollo de fisuras orofaciales en múltiples estudios.^{171,200}

En la familia cinco se obtuvo información de cinco generaciones del árbol genealógico, de las cuales cuatro tienen fisuras orofaciales con fenotipos diferentes, sin presencia de fositas en labio inferior en el núcleo familiar, la genealogía evidencia una expresividad variable, penetrancia incompleta y una herencia autosómica dominante, muy característica de variantes asociadas a IRF6. Cabe recordar que se han reportado variantes patogénicas en pacientes con oligodoncia aislada en IRF6,²⁰⁶ que podría ser una forma de expresión de la variante encontrada en esta paciente. Ante el resultado de la prueba molecular se hizo una segunda revisión a la familia, debido al hallazgo en el gen IRF6 y su relación a entidades sindrómicas como el Síndrome de Van Der Woude (SVW), no se encontraron fositas en el labio inferior de los pacientes, sin embargo, se identificó una lesión tipo fistula unilateral, medial en la mucosa oral cerca al labio inferior de la probando, situación que llevó a cuestionar si se trataba de una paciente con SVW o L/PH-NS.

5.2 Variante patogénica en IRF6

El IRF6 pertenece a la familia de los factores reguladores del interferón, la cual comparte regiones similares como el dominio de unión a ADN con una estructura

hélice-giro-hélice N-terminal muy conservada y un dominio de unión a proteína C-terminal menos conservado. El gen está constituido por 9 exones y 8 intrones, los exones 3 y 4 forman el dominio de unión al ADN, los exones 7 y 8 forman el dominio de unión a proteínas; donde se encuentran la mayoría de variantes asociadas con L/PH sindrómicas y no sindrómicas.^{69,207} Los exones 1 y 2 no codificantes forman la región 5' no traducida de este gen. La variante en IRF6 encontrada en esta investigación consiste en una CNV con delección de una copia de los exones 1-3, este último exón tiene un papel importante para el motivo más conservado de esta proteína y es fundamental para unirse al ADN.⁷⁰

La información sobre la frecuencia de esta variante no se encuentra en bases datos poblaciones como en gnomAD, ExAC, 1000 Genomas, y dbvar. Por la pérdida de los primeros exones se espera que interrumpa el empalme del ARN y esto puede conducir a una pérdida de la función de la proteína. Esta variante no ha sido reportada en la literatura previamente; sin embargo, hay casos reportados de variantes tipo missense, nonsense, splicing e Indels en los exones 1, 2 y 3 del gen IRF6 relacionadas a hendiduras orofaciales.²⁰⁷ Se ha informado previamente sobre variantes patogénicas en la región 5' no traducida (exones 1 y 2) relacionadas con SVW, sugieren que dicha ubicación puede desencadenar un codón de inicio alternativo ascendente que lleva a crear proteínas truncadas.²⁰⁸ Otro hallazgo reconocido por Leslie et al.²⁰⁸ fue la relación genotipo fenotipo donde propone que las variantes patogénicas en IRF6 localizadas en los exones 3 y 4 que forman el dominio de unión a ADN contribuyen a la aparición del síndrome de SVW o el

síndrome de pterigión poplíteo (SPP), encontrando frecuencias con significancia estadística (p valor de 0.0001), situación que se ha reportado en otros estudios, aunque, cabe aclarar que en ellos solo se han valorado familias e individuos con SVW o el SPP y sin incluir participantes con L/PH-NS.^{209,210} No hay estudios que relacionen el genotipo-fenotipo de IRF6 con L/PH-NS, pero sí hay casos de familias reportadas en la literatura con variantes puntuales en los exones 2 y 3.^{211–213}

La variante reportada en IRF6 es sensible a dosis, intolerante a las mutaciones con pérdida de función, las métricas que evalúan la probabilidad de haploinsuficiencia como pHaplo, pLI y el mapa de sensibilidad de dosis de ClinGen lo confirman.²¹⁴

Los estudios experimentales y los algoritmos de predicción no están disponibles para esta variante, no obstante, con la pérdida de los 3 primeros exones en IRF6 que incluye la pérdida del exón 3, el cual forma parte del dominio de unión a ADN, el más conservado evolutivamente de dicha proteína, denota que su ausencia conlleva a pérdida de su función. Para afirmar que la variante identificada en el gen IRF6 es dañina para el funcionamiento normal de la proteína, en este estudio se realizó un modelamiento del dominio de unión a ADN en IRF6 y de la variante patogénica ya descrita.

El dominio de unión a ADN de los miembros IRF contiene cinco triptófanos conservados y tiene una arquitectura compuesta por: cuatro láminas beta, tres loops grandes y tres hélices alfa (H1, H2 y H3), donde la H2 y H3 son importantes para el

reconocimiento al ADN, en especial, la última que identifica una secuencia de repetición GAAA en el surco mayor de ADN presente en los elementos de respuesta de los IRF.⁷³ Estudios previos experimentales realizados mediante cristalografía de rayos X han identificado el complejo de unión proteína-ADN en algunos miembros IRF: en IRF1, IRF2 e IRF4 en los que se han reconocido aminoácidos clave para el contacto con el ADN como es el caso de IRF1 con Arg 82, Cys 83, Asn 86 y Ser 87,⁷³ en IRF2 con Asn86, Cys83, Arg82, Asn80,⁷⁵ en IRF4 con la Lys103, Asn102, Cys99, Arg98, H56^{199,201} y en el modelo obtenido en este estudio por herramientas bioinformáticas en IRF6 mediante HDOCK, previamente no publicado en la literatura ni en bases de datos, se hallaron los siguientes aminoácidos: Arg 84, Cys 85 Asn 88 , His 42 y Lys89; este último no se encuentra en IRF2 ni en IRF1 pero sí en IRF4 y en los dos modelos planteados 7OOT y 7JW4 que justifican su papel en la interacción con el ADN, lo que evidencia la validación del modelo de acoplamiento resultado de IRF6-ISRE planteado.

Al realizar el modelo con la variante patogénica identificada en IRF6 con la delección del exón 3, hubo una inadecuada unión entre el IRF6 y el ISRE observando cómo se pierde el eje casi paralelo que se crea en la H3 con el surco mayor del ADN, además, se identificaron aminoácidos diferentes en el complejo de acoplamiento, en ese caso A59, W60, E63, T84, V62 y Q118 en una conformación diferente. Esto demuestra que hay una pérdida funcional de la unión del IRF6 con el ADN en la variante ya mencionada, teniendo en cuenta que este gen presenta haploinsuficiencia.^{215,216} Sumado a ello, mediante algunas herramientas

bioinformáticas como COACH ²¹⁷ y ejecuciones de docking entre IRF6 y proteínas relacionadas como TGTB3 indican de forma hipotética que el exón 3 del IRF6 es un sitio potencial de unión a proteínas, admitiendo la importancia funcional de esta región.

5.3 Expresividad variable en el gen IRF6

Los estudios en individuos y familias con variantes patogénicas en IRF6 han evidenciado expresividad variable en sus genealogías, con un espectro que va desde agenesia de un incisivo aislado, ²⁰⁶ L/PH-NS hasta fenotipos sindrómicos tales como: SVW y el SPP. El primer síndrome presenta generalmente fositas paramedianas bilaterales en el labio inferior asociadas o no a L/PH y puede incluir hipodoncia. ²⁰⁷ Se debe tener en cuenta que en el SVW un 15% de los individuos no tienen fositas, además, se presenta de forma diferente con otro tipo de lesiones en el labio inferior como fístulas o montículos, descritas en algunas publicaciones como en Rizos et al.²¹⁸ y Desmyter et al.²¹⁹ El segundo síndrome es una condición que implica anomalías craneofaciales, malformaciones de las extremidades y genitourinarias. ²²⁰

Se han reportado familias con SVW y SPP en una sola genealogía, en el 2016 Busche et al. ²²¹ observaron una familia de tres generaciones con fenotipos diferentes; probando con SPP y madre y abuelo por línea materna con SVW, quienes antes habían sido diagnosticados con L/PH-NS y sumado a ello presentaban la misma variante patogénica en el exón 4 de IRF6, (c.265A>G). Matsuzawa et al. ²²² estudiaron dos familias con SPP, lo interesante estaba en que

una de ellas, conformada por cuatro generaciones en donde el probando presentaba signos característicos de SPP, el padre tuvo LPHB sin otra anomalía asociada, se observaron microformas en dos familiares por línea paterna: una tía y un abuelo con úvula bífida y PH sin fosas labiales, de acuerdo a lo anterior, sin la presencia del paciente índice se podría estar evidenciando una genealogía característica de L/PH-NS.

Barger et al.²²³ exponen dos casos, el primero consiste en una familia con un probando masculino con paladar hendido submucoso sin otras anomalías, abuela materna con LH sin otros antecedentes familiares de hendiduras orofaciales ni fositas labiales, con variante patogénica missense heterocigota 1106 T>C en IRF6, lo llamativo en este caso es que esta misma variante ha sido publicada previamente en individuos con SVW.²⁰⁷ En la segunda familia se describe una probando femenina con LPHB y úvula bífida sin otra anomalía, con una variante en el exón 3 Met1Leu, 1A > C en IRF6, bis abuelo y tía abuela por línea paterna con LH sin otras anomalías, una familia con L/PH-NS,²²³ pero con una variante también notificada en SVW.²⁰⁷

Desmyter et al.²¹⁹ hace referencia a una hipótesis en la que se debe considerar a los pacientes con L/PH-NS como portadores de variantes patogénicas en IRF6 y que algunos clasificados como no sindrómicos podrían ser sindrómicos. En este estudio de las 95 familias consideradas al inicio con L/PH-NS, posterior a un examen exhaustivo, tres de ellas pasaron a ser clasificadas como sindrómicas, algo notable

en esta publicación fue la identificación de la lesión que describe dos nódulos debajo del labio inferior en el lado externo que catalogaron como un fenotipo variable de las fositas en el SVW.

Lo descrito anteriormente hace importante la minuciosa revisión clínica y odontológica de los individuos con variantes en IRF6 debido a los múltiples fenotipos relacionados, cada uno presenta un riesgo de recurrencia diferente para los familiares de primer grado, el L/PH-NS tiene un riesgo aproximado del 3 al 5 % mientras en el SVW es del 35 al 50%,²²⁴ clave al momento de realizar el asesoramiento individual y de la familias; en este caso para la familia cinco, que teniendo en cuenta los hallazgos clínicos y posterior a los resultados moleculares existe un diagnóstico diferencial de L/PH-NS y SVW.

Esta variabilidad puede ser causada por la gran cantidad de interacciones y reguladores corriente arriba que tiene IRF6, por lo que se ha propuesto en algunos estudios el papel de otros factores genéticos sobre el gen, como causantes de la expresividad variable, como es el caso de un elemento potenciador conocido como MCS9.7²²⁵. Lo descrito anteriormente refleja la falta de estudios en el campo molecular en familias con fenotipos de hendiduras mixtas con hallazgos asociados a síndromes que permitan ampliar el conocimiento sobre las raíces genéticas e incluso epigenéticas de la expresividad variable relacionada al gen IRF6.

5.4. Correlación entre microformas con L/PH-NS

Las microformas son manifestaciones leves de las hendiduras orofaciales, pueden incluirse: la agenesia dental, los dientes cónicos, la úvula bífida, la bóveda palatina y algunas otras. En las cinco familias de este estudio, se encontró que la asimetría nasal fue la microforma más frecuente en los participantes, seguido del arco maxilar en forma de V, la bóveda palatina y el torus palatino, hallazgos similares en otra investigación realizada en República Dominicana.³⁶

Otro hallazgo relacionado a microformas en este estudio fue el fenotipo de la madre de la probando de la familia cinco que es portadora de la variante patogénica en IRF6, presenta agenesia de incisivos laterales superiores, signo que ha sido asociado al SVW y también a L/PH-NS. Las alteraciones dentales hacen parte de las microformas y la agenesia dental no sindrómica es una de las más comunes, se define como la ausencia congénita de al menos un diente permanente. En un estudio publicado por Biedziak et al.²²⁶ identificaron genes relacionados a hipodoncia y oligodoncia, entre ellos el IRF6, MSX1, CREBBP, TBX22, TP63 y PAX9, que asimismo son asociados a L/PH-NS. Otros estudios han identificado la asociación de IRF6 con las agenesias dentales aisladas mediante estudios de desequilibrio de ligamiento, como es el caso de Vieira et al.²²⁷ donde reconocieron que algunas variantes en la hipodoncia estaban relacionadas previamente a hendiduras orofaciales aisladas (rs861019, $p = 0,058$; rs17015215-V274I, $p = 0,0006$; rs7802, $p = 0,004$).²²⁷

Así mismo, la madre de la probando de la familia cinco presento torus palatino bilobulado, el cual de manera controversial en la literatura científica es señalado como microforma. El torus palatino tiene una herencia multifactorial,²²⁸ donde se involucran factores ambientales como la edad, el hiperparatiroidismo entre otros y en algunos estudios se asocian a factores genéticos, ya que se describen en la literatura múltiples casos de familias con torus palatino, al parecer con una genealogía que muestra una herencia autosómica dominante. Los estudios moleculares no han descrito aún una relación con el gen IRF6, pero sí con otros genes como es el caso de CAPS2 encontrado en un GWAS publicado en el 2021 realizado en familias filipinas.²²⁸

En el 2020 se realizó en la Universidad de Iowa un estudio que evaluó las posibles conexiones entre el torus palatino y el L/PH. Encontraron que la presencia de torus palatino fue más común en los familiares de casos con L/PH que en los controles, con riesgo dos veces mayor de presentar fisuras orofaciales.²²⁹ De igual manera, estudios en familiares y gemelos han mostrado la transmisión vertical del torus palatino, siendo evidente en gemelos monocigóticos.²³⁰ Esto muestra el peso posible del componente genético en la etiología de esta anomalía, no obstante, faltan más estudios que amplíen el conocimiento con relación al componente molecular del torus palatino y sobre su relación con el desarrollo de L/PH.

5.5 La red de genes propuesta

En este estudio se propuso una red de genes relacionada a IRF6 y a L/PH-NS, teniendo en cuenta estudios experimentales y algunos hallazgos estadísticos significativos, se trató de unificar conceptos sobre el conjunto de genes, regiones en el ADN y factores en el ambiente que tienen implicaciones sobre el desarrollo de las fisuras orofaciales. Estudios experimentales sobre IRF6 en animales han identificado el papel de otros genes que de forma directa o indirecta participan en procesos biológicos como: la proliferación celular, el ciclo celular, la transcripción, la formación del epitelio y el ectodermo, todos importantes para el desarrollo normal del paladar y el labio, ¹⁷⁷ hallazgo también evidente desde el análisis de anotación funcional realizado mediante la base de datos DAVID ¹⁷⁸ con los genes obtenidos de la red de genes creada en el presente estudio. Una de las vías estudiadas en la palatogénesis involucra al TGFB3 señalado como un gen regulador en este proceso, interactúa corriente arriba de IRF6, estimulando su expresión en sitios clave como el borde medial de los procesos palatinos importante para la fusión de los mismos, esto se ha evidenciado en investigaciones en embriones de ratones mediante ensayos con inmunotinción y RT-PCR. ²³¹

El IRF6 regula corriente abajo a otros genes, es el caso de SNAI2 que aumenta su expresión y este a su vez inhibe la expresión de las proteínas PKP1, ZO2 y la E-Cadherina a nivel del borde medial de los procesos palatinos, todo en conjunto

participa en la degeneración del borde epitelial de la línea media (MES), que implica la migración celular, la apoptosis y la transición mesenquimatosa epitelial (EMT) necesarias para una correcta fusión de los procesos palatinos.^{63,232,233} SNAI2, PKP1, ZO1 no se incluyeron en la red de genes, los dos primeros se relacionan a entidades sindrómicas y en general los tres no han sido identificados en estudios en humanos con L/PH-NS. La E-cadherina codificada por el gen CDH1 a diferencia de la PKP1 y ZO1 que también son regulados de forma negativa por SNAI2, sí ha sido relacionada a pacientes con L/PH-NS en China.²³⁴

El papel que juega TGF β corriente arriba sobre IRF6 se realiza a través de dos vías de señalización, una dependiente de SMAD y la otra de MAPK; en estudios de expresión génica en ratones se ha demostrado que las mutaciones en estos dos genes pueden disminuir hasta en un 50% la expresión de IRF6 y llevar a la aparición de trastornos en el paladar.⁶² Se han identificado variantes patogénicas en SMAD4 asociadas a L/PH-NS, caso diferente con el gen MAPK que no tiene estudios que evidencien variantes afines con L/PH-NS. Hay otras vías de señalización en la palatogénesis en donde participa IRF6 como es el caso de JAG2 y NOTCH1, importantes para la formación y el mantenimiento del epitelio oral y clave para la fusión de los procesos palatinos. En algunos estudios se ha planteado la hipótesis de que IRF6 y JAG2 convergen y tienen efectos combinados sobre la formación, el mantenimiento y la función del epitelio que se clarificó en una investigación de

expresión en ratones publicada en el 2009, ²³⁵ sumado a ello se han hallado variantes relacionadas a L/PH-NS. ²³⁶

Otra vía de señalización en la palatogénesis involucra al TP63, al parecer este aumenta la expresión de IRF6 y en forma de retroalimentación negativa este último inhibe la expresión de TP63, el cual a su vez inhibe la expresión de p21 y en conjunto llevan a la apoptosis de los MES y MEE en los bordes mediales de los procesos palatinos, permitiendo la fusión de los mismos y la formación normal del paladar secundario. ^{55,211} Las variantes en TP63 han sido relacionadas a siete fenotipos, cinco sindrómicos y dos no sindrómicos, entre ellos el L/PH-NS. ²³⁸ Otra vía de señalización relacionada a la morfogénesis facial involucra a PBX1, WNT3 y WNT9B, TP63 e IRF6, genes asociados con L/PH-NS en seres humanos. ²³⁹ En un estudio llevado a cabo en el 2011 mediante qRT-PCR, análisis histológico, inmunohistoquímica, inmunofluorescencia y ensayos transcripcionales in vitro en embriones de ratón, se evidenció el papel de PBX en el desarrollo del ectodermo mediofacial, regulando la expresión de Wnt9b-Wnt3 y Tp63 e indirectamente al Fgf8.

240

En múltiples estudios tipo GWAS realizados en población europea, asiática, hispana y multiétnica se ha reconocido el papel de la región 8q24 en el desarrollo de LPH-NS; ^{80,241} encontrando que una de sus regiones intergénicas tiene un potenciador que funciona en cis, el cual regula positivamente la expresión de MYC, un gen relacionado con la morfogénesis facial en ratones. Esto fue investigado en

embriones de ratones a través de pruebas de inmunoprecipitación de cromatina (ChIP), qPCR e hibridación in situ ²⁴². Se han identificado variantes patogénicas en la región 8q24 relacionadas con L/PH-NS en humanos. ^{241,243} El gen *OVOL1* se asocia a *IRF6* como represor de la transcripción de *MYC* a nivel de los queratinocitos, donde regula su proliferación y diferenciación, esto se planteó en un análisis que se hizo del genoma, combinando la secuenciación de ChIP para los sitios de unión de *IRF6* y el perfil de expresión génica en queratinocitos humanos primarios, ²⁴⁴ adicional a ello, en estudios de expresión, se encontró que el estímulo de *IRF6* se relaciona a un aumento en la expresión de *OVOL1*; ²⁴⁵ no obstante, aún no hay hallazgos en la literatura con relación a L/PH-NS en seres humanos con este gen.

Otra región llamativa en el ADN y que se ha asociado a L/PH-NS es una secuencia corriente arriba de *IRF6* denominada MCS9.7, la cual, en varios estudios se determina como un potenciador para la expresión de *IRF6* y además posee un sitio de unión a TP63 que transactiva la expresión de *IRF6*, esto fue investigado en tejidos de embriones murinos mediante estudios de expresión, ²⁴⁶ como ya se ha planteado no es la única vía de activación de *IRF6*, pero es un punto importante en su regulación. Otro gen relacionado en esta red es el *TFAP2A*, algunos estudios en ratones sugieren una relación con la secuencia potenciadora, una especie de retroalimentación con MCS9.7 y al parecer podría estar implicado en la expresión corriente arriba de *IRF6* a nivel de la epidermis y en el epitelio oral y asociarse con

el riesgo de hendiduras orofaciales, ²²⁵ mecanismo apoyado por un ensayo de inmunoprecipitación de cromatina (ChIP) y qPCR realizado en embriones de ratón.

²⁴⁷

El GRHL3 es regulado corriente arriba por IRF6, disminuyendo su expresión y actuando como efector en la diferenciación del epitelio, situación expuesta en experimentos realizados en embriones de peces cebras y ranas, usando micromatrices. ²⁴⁸ ARHGAP29 es una proteína activadora de Rho GTPasa y ha sido señalada como causante de L/PH-NS, además, reconocida en una red hipotética en la morfogénesis craneofacial en varias bases de datos curadas como Reactome. ^{249,250} Estudios de expresión en tejido de ratón han demostrado la posible interacción de IRF6 corriente arriba en la expresión positiva de ARHGAP29, demostrado en la comparación entre murinos deficientes para IRF6 y los controles con IRF6 normal, ²⁵¹ confirmado en otro estudio usando ChIP-seq en queratinocitos humanos, reconociendo el rol de IRF6 corriente arriba en la regulación hacia ARHGAP29 con relación a su expresión. ²⁴⁴ Todo ello sugiere una nueva vía en la que la red reguladora del gen IRF6 conecta con la vía Rho a través de ARHGAP29 y este a su vez ha sido relacionado con MYC y con la región ya previamente mencionada con 8q24, adicionalmente también se conecta con el gen regulador PBX de los genes WNT, donde Rho actúa corriente abajo de la señalización WNT. ^{252,253}

Un estudio publicado en el 2016 analizó las asociaciones de 24 SNVs y factores de riesgo y protección relacionados a L/PH-NS, mediante un método de reducción de dimensionalidad multifactorial (MDR) en una población mexicana, demostrando

algunas interacciones con valores estadísticamente significativos, se tomaron en cuenta aquellas relacionadas de forma indirecta y directa con IRF6 como: NTN1/IRF6 ($p = 0.0002$), TPM1/IRF6 ($p = 0.0001$), ABCA4/MAFB ($p = 0.0006$), ABCA4/consumo de ácido fólico ($p < 0.0001$), ARHGAP29/PAX7 ($p = 0.0004$) y PAX7/TPM1 ($p = 0.0011$).²⁵⁴ El NTN1 codifica la netrin1, una proteína señalada en estudios en peces cebra que se expresa a nivel orofacial, es ortólogo del NTN1 humano con una similitud del 86% e identificado en estudios GWAS a L/PH-NS.^{255,256} No es claro y no ha sido estudiado el gen TPM1 en la morfogénesis craneofacial; sin embargo, cumple un papel en la organización del citoesqueleto y han sido identificadas variantes relacionadas a L/PH-NS.²⁵⁷ El ABCA4 cuenta con múltiples estudios donde reconocen variantes patogénicas asociadas a L/PH-NS; no obstante, su papel en la morfogénesis craneofacial aún se desconoce, la proteína se expresa principalmente en la retina.²⁵⁸ El PAX7 hace parte de la familia de los genes Paired Box Homeotic, en humanos y en animales se ha reconocido su función en el desarrollo craneofacial.^{259,260}

En otra publicación del 2014 se identificó una interacción con un p valor significativo entre el gen FILIP1L y la región 8q24 ($p = 0.00011$) a través de regresión logística y aprendizaje automático para probar interacciones entre genes identificados por un GWAS en tríos caso-padre para L/PH-NS,²⁶¹ llama la atención el gen FILIP1L, el cual, tiene variantes en cuanto a L/PH-NS, pero su función en el desarrollo facial no

es claro, solo se sabe que cumple un papel en la proliferación, migración celular y apoptosis en las células endoteliales. ²⁶²

Entre los factores protectores estadísticamente significativos encontrados en la literatura relacionada al desarrollo de fisuras orofaciales, se identifica el consumo de ácido fólico antes y durante el embarazo. En un artículo realizado en población mexicana en el año 2012, se observaron interacciones significativas entre el consumo de ácido fólico y la variante 8q24 ($p < 0.0001$). ²⁶³ En una tesis publicada en el 2015 se determinó la interacción entre IRF6 y MAFB ($p = 0.02$) relacionada a fisuras orofaciales, empleando dos pruebas, una de razón de verosimilitud (LRT) y otra de 4df. ²⁶⁴ El gen MAFB se expresa en la región craneofacial y al parecer tiene un papel en el desarrollo del ectodermo, no es claro el mecanismo en el desarrollo del paladar; sin embargo, se han reportado variantes relacionadas a L/PH-NS. ^{265,266} La red tomó en cuenta algunos genes propuestos en vías de señalización como STRING y las cruzó con resultados dados en un estudio tipo GWAS en L/PH-NS, se adicionaron los siguientes genes, todas funciones conocidas para el desarrollo del paladar y el labio: MSX1, SHH, SPRY2 y FGFR1. ²⁴¹ Msx1, Pax9 y Shh han sido genes implicados en fisuras orofaciales en modelos de ratón manipulados genéticamente y en estudios de tejidos humanos; señalados en el crecimiento de los procesos palatinos laterales en la región anterior durante la palatogénesis, los factores de crecimiento y sus receptores regulan la proliferación y supervivencia celular en los procesos palatinos. ²⁶⁷ El Spry2 juega un papel importante en la

elevación y proliferación de los procesos palatinos. ^{o268} El gen VAX1 codifica un regulador transcripcional que se expresa en tejidos del labio y el paladar de embriones de ratones y al parecer es regulador en el desarrollo craneofacial. ^{269,270} Se han publicado otras redes relacionadas a IRF6 y fisuras orofaciales, una de ellas fue publicada en el 2015, donde plantearon una red reguladora mediante un análisis sistemático bioinformático que señaló a ocho genes con evidencia significativa para el desarrollo de fisuras orofaciales: Msx1, Pvr11, Pax9, Irf6, Jag2, Tgfb3, Rara, Gli2 y Tgfb2. ¹⁷⁷ Otra propuesta de red relacionada a una vía de interacción basada en IRF6 e involucrada en la etiología del L/PH-NS humano, incluye seis genes con interacciones directas relevantes: TP63, NTN1, TFAP2A, TPM1, VAX1 y GRHL3; sin embargo, no es muy claro el mecanismo biológico directo de VAX1 sobre IRF6; la red de esta investigación resalta la importancia de los datos estadísticos encontrados en su estudio y en la literatura. ²⁵⁴ En la propuesta de red en IRF6 de esta investigación se da mayor significancia a los datos obtenidos en experimentos y en siguiente lugar a la estadística con valores significativos.

Se han logrado avances con relación a los mecanismos moleculares asociados al desarrollo del labio y paladar especialmente en ratones, permitiendo entender y extrapolar en la especie humana algunos datos; sin embargo, hay información en la que no se tiene claridad en cuanto a las vías de señalización y procesos biológicos en la morfogénesis del paladar, ejemplo de ello son NTN1 y FILIP1L, se requiere

por tanto indagar en el campo de la biología del desarrollo y dilucidar conceptos en esta área.

5.6 Secuenciación de nueva generación

En República Dominicana se realizó en el año 2018 un estudio de secuenciación de exoma completo (WES) con resultado negativo para la familia que en esta investigación con WGS se identificó una variante positiva en IRF6. Con relación al análisis de regiones codificantes en el ADN, ambas tecnologías mencionadas son adecuadas; sin embargo, las plataformas de WES basan su eficiencia en la profundidad durante la secuenciación, aun así, WGS supera a WES en la detección de variantes debido a su mejor cobertura en las regiones codificantes, como es el caso de las variantes tipo CNV. Un estudio publicado en el 2015 puso en evidencia que el WES tiene un rendimiento comparable a WGS en términos de bases cubiertas, si realiza la secuenciación con una cobertura de dos a tres veces mayor al WGS.²⁷¹ Al momento de escoger qué tecnología utilizar para un estudio de enfermedades genéticas, es importante tener en cuenta el tipo de estudio molecular, además de la cobertura y profundidad de cada una, ya que estos parámetros permiten una mejor calidad en los resultados.

5.7 Limitaciones

La familia cinco involucrada en el estudio participó hasta el último asesoramiento genético y posterior a ello se negó a seguir siendo parte del proyecto realizado, por

lo tanto, esto generó una limitante para adquirir información adicional y la realización de otros estudios como MLPA en los demás miembros afectados en esta genealogía.

La pandemia hizo que el cronograma se extendiera y retrasara varios procesos en el desarrollo de este estudio. Además, limitó la realización de otros procesos como el de dinámica molecular en las ejecuciones de docking.

6. Proceso de divulgación de resultados

Se reportaron los resultados a las cinco familias dominicanas, se entregó un informe escrito de la variante encontrada y de los hallazgos incidentales, fue realizado por un médico genetista y trabajadores de la salud capacitados. Posterior a ello, se efectuó la divulgación en la comunidad de la región de Cibao y Santo Domingo y parte del área científica en República Dominicana. Se participó en eventos de carácter académico como el Ceicyt- RD 2022 en República Dominicana, en la competencia Concurso Hatton-Unilever, Categoría Senior, en representación de la Universidad Nacional de Colombia. Este concurso se llevó a cabo en el XVI Encuentro de la IADR – División Colombia. Además, en el IV Congreso LATICFA (Latin American CranioFacial Association) y AMLPHAC (Asociación Mexicana de Labio, Paladar Hendido y Anomalías Craneofaciales) 2022, con presentación de una ponencia.

7. Recomendaciones

En el presente estudio se identificaron ocho variantes, siete de ellas no asociadas a L/PH-NS consideradas como hallazgos incidentales en los genes CYP21A2 c.-103A>G (probablemente benigna), CYP21A2 -113G>A (significado incierto), SMPD1(patogénica), DUOX2(patogénica), ganancia CNV 22q11.1q11.21(significado incierto), B3GLCT (patogénica) y HBB (patogénica), de las cuales las patogénicas presentan una herencia recesiva que hace a los individuos portadores, teniendo en cuenta que República Dominicana es una isla se genera mayor riesgo de endogamia y de presentar enfermedades de origen genético, por lo anterior, se resalta la importancia de promover pruebas moleculares en la población, además de campañas de prevención por parte de un equipo multidisciplinario.

Los estudios en genoma completo se benefician de una cobertura de 40X o más con el fin de encontrar mayor cantidad de información relacionada a variantes patogénicas.

República Dominicana se beneficiaría de proyectar estudios a futuro tipo panel de genes en sus pacientes con L/PH, siendo costo efectivo, para el campo clínico e investigativo.

8. Conclusiones

Se detectó una variante tipo CNV patogénica consistente en una delección de una copia de los exones 1-3 del gen IRF6, asociada a fisuras orofaciales no reportada previamente en la literatura.

Se construyó un modelo 3D de IRF6 mediante herramientas computacionales que evidenció similitud con meta servidores relacionados a evaluación de estructura 2D y guarda correspondencia con los otros miembros IRF, especialmente, en el dominio de unión a ADN y unión a proteína.

Se creó un modelo de acoplamiento de ADN-proteína con IRF6 no publicado en la literatura científica, en el que se encontraron elementos similares a las estructuras obtenidas por métodos experimentales de otros miembros de los IRF. Se hizo un nuevo grafo dirigido de genes relacionados a IRF6 y L/PH-NS.

En pacientes con L/PH-NS y herencia autosómica dominante, vale la pena estudiar genes relacionados a entidades sindrómicas; como se ha mostrado en este estudio, pues hay genes que están dentro del espectro de anomalías no sindrómicas y enfermedades sindrómicas, tal es el caso de IRF6 con expresividad variable. En la

literatura se identifican variantes patogénicas con manifestaciones diferentes en una misma genealogía que podría evidenciar el papel que juegan otros genes y elementos reguladores en el ADN.

Teniendo en cuenta la expresividad variable de algunos genes como el caso de IRF6, con fenotipos que incluyen entidades sindrómicas y no sindrómicas, es necesaria una evaluación exhaustiva clínica sistémica y oral con el fin de llegar a diagnósticos certeros y a realizar el asesoramiento genético individual y familiar adecuado.

8.1 Financiación

El proyecto de investigación “Caracterización funcional de variantes (SNV, Indels, CNVs) analizadas por medios computacionales en 5 familias dominicanas con labio y/o paladar hendido no sindrómico.” El proyecto fue financiado por el Fondo Nacional de Innovación y Desarrollo Científico y Tecnológico (FONDOCyT No. 2018-19-2A5-236) de República Dominicana, por la Universidad Nacional de Colombia y la Universidad Nacional Pedro Henríquez Ureña.

La Universidad Nacional de Colombia, permitió el uso de laboratorio de Bioinformática y Biología de Sistemas para el desarrollo de algunos procesos computacionales.

9. Bibliografía

1. Ospina Ramirez JJ, Castro David MI, Hoyos Ortiz LK, Montoya Martinez JJ, Porras Hurtado GL. Factores asociados a malformaciones congénitas: En un centro de tercer nivel región centro occidental - Colombia (ECLAMC). *Rev Médica Risaralda*. 2018;24(1):15.
2. Basha M, Demeer B, Revencu N, Helaers R, Theys S, Bou Saba S, et al. Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes. *J Med Genet*. 2018;55(7):449–458.
3. Mbuyi-musanzayi S, Kayembe TJ, Kashaal MK, Lukusa PT, Kalenga PM, Tshilombo FK, et al. Non-syndromic cleft lip and/or cleft palate: Epidemiology and risk factors in Lubumbashi (DR Congo), a case-control study. *J Cranio-Maxillofacial Surg*. 2018;46(7):1051–1058.
4. Chen H. Cleft Lip and/or Cleft Palate. In: Chen H, editor. *Atlas of Genetic Diagnosis and Counseling*. Third edi. New York: Springer; 2017. p. 475-484.
5. Murray JC. Gene/environment causes of cleft lip and/or palate. *Clin Genet*. 2002;61(4):248–256.

6. Conte F, Oti M, Dixon J, Carels CEL, Rubini M, Zhou H. Systematic analysis of copy number variants of a large cohort of orofacial cleft patients identifies candidate genes for orofacial clefts. *Hum Genet.* 2016;135(1):41–59.
7. Porras-Hurtado GL, León-castañeda OM, Molano-hurtado J. Prevalencia de defectos congénitos en Risaralda, 2010-2013. *Biomédica.* 2016;36:556–563.
8. Madsen C, Lough D, Lim A, Harshbarger RJ, Kumar AR. Cleft and craniofacial care during military pediatric plastic surgery humanitarian missions. *J Craniofac Surg.* 2015;26(4):1097–1101.
9. De La Cruz-Acosta F. Operación Sonrisa República Dominicana: 8 años de una intensa y positiva experiencia. *Cir Plast Ibero-Latinoamericana.* 2016;42(1):93–101.
10. García-Godoy F. Cleft lip and cleft palate in Santo Domingo. *Community Dent Oral Epidemiol.* 1980;8(2):89–91.
11. Maldonado Maldonado LA, Morales Borrero MC, Silvia-Vetri MG. Derecho a la atención sanitaria para madre y menores de 18 años afectados con labio y/o paladar hendido en República Dominicana. *Acta Odontol Colomb.* 2020;10(2):68–81.
12. Arévalo-Caro C, Silva-Vetri MG, Arteaga Díaz CA, Villadiego Pineda S, Ramirez Martinez A, Federico BJ. Clinical and socioenvironmental characterizations associated with the genealogical analysis of a group of Dominican patients with nonsyndromic cleft lip with or without cleft palate. *Oral Sci Int.* 2022;19(3):160-166.
13. Yohe S, Thyagarajan B. Review of clinical next-generation sequencing. *Arch Pathol Lab Med.* 2017;141(11):1544–1557.

14. Gutiérrez-Nava A, Mayorga-Reyes L. La era post-genómica en biomedicina. *Rev Mex Ciencias Farm.* 2011;42(2):7–13.
15. Leslie EJ, Taub MA, Liu H, Steinberg KM, Koboldt DC, Zhang Q, et al. Identification of functional variants for cleft lip with or without cleft palate in or near PAX7, FGFR2, and NOG by targeted sequencing of GWAS loci. *Am J Hum Genet.* 2015;96(3):397–411.
16. República Dominicana. Consejo para la Innovación y el Desarrollo Tecnológico. Secretaría de Estado de Educación Superior, Ciencia y Tecnología; Consejo para la Innovación y el Desarrollo Tecnológico. Plan estratégico de ciencia, tecnología e innovación 2008-2018. Construyendo la economía del conocimiento y la innovación en República Dominicana. Consejo para la innovación y el desarrollo tecnológico. [Internet]. (Agosto. 1, 2018). Available from: https://siteal.iiep.unesco.org/sites/default/files/sit_accion_files/do_5040.pdf
17. World Health Organization. WHO. 63 Asamblea Mundial De La Salud. Punto 11.7 del orden del día DEFECTOS CONGENITOS. 2010.
18. Rojas M, Walker L. Malformaciones congénitas: Aspectos generales y genéticos. *Int J Morphol.* 2012;30(4):1256–1265.
19. Emidio D', Toboso L, Sánchez F. Agenesia de incisivos laterales ¿Cerrar o abrir espacio? *RCOE.* 2017;22(4):197–208.
20. Wornom I, Will LA, Burdi A, Berkowitz S, Breen M, Clarke-Sheehan N, et al. Interdisciplinary team care, classification, airway, and feeding. In: Berkowitz S,

editor. Core Curriculum for Cleft Lip/Palate and other Craniofacial Anomalies.

Second Edi. Heidelberg:Springer; 2006. p. 285-288.

21. Davids JS, Ritchie HP. Classification of congenital clefts of the lip and palate with a suggestion for recording these cases. *JAMA*. 1922;79(16):1323–1327.
22. Allori AC, Mulliken JB, Meara JG, Shusterman S, Marcus JR. Classification of cleft lip/palate: Then and now. *Cleft Palate-Craniofacial J*. 2017;54(2):175–88.
23. Spina V. A proposed modification for the classification of cleft lip and cleft palate. *Cleft Palate J*. 1973;10:251–2.
24. Rodrigues R, Fernandes MH, Monteiro AB, Furfuro R, Sequeira T, Silva CC, et al. SPINA classification of cleft lip and palate: A suggestion for a complement. *Arch Pediatr*. 2018;25(7):439–441.
25. Sharma P, Kharbanda P. Role of Programmed Cell Death in Dental Anomalies Associated with Cleft Lip and Palate. *Med Hypotheses*. 1991;36(1):98–100.
26. Howe BJ, Cooper ME, Vieira AR, Weinberg SM, Resick JM, Nidey NL, et al. Spectrum of dental phenotypes in nonsyndromic orofacial clefting. *J Dent Res*. 2015;94(7):905–12.
27. Suzuki S, Marazita ML, Cooper ME, Miwa N, Hing A, Jugessur A, et al. Mutations in BMP4 Are Associated with Subepithelial, Microform, and Overt Cleft Lip. *Am J Hum Genet* [Internet]. 2009;84(3):406–11. Available from: <http://dx.doi.org/10.1016/j.ajhg.2009.02.002>
28. Antoniadis DZ, Belazi M, Papanayiotou P. Concurrence of torus palatinus with palatal and buccal exostoses. *Oral Surgery, Oral Med Oral Pathol Oral Radiol Endodontology*. 1998;85(5):552–557.

29. Wharton P, Mowrer DE, Cohn ER. Prevalence of cleft uvula among school children in kindergarten through grade five. *Cleft Palate Craniofac J.* 1992;29(1):10-14
30. Patricia M, Murillo O. Dientes supernumerarios. Reporte de un caso clínico. *Rev Odont Mex.* 2013;17(2):91–96.
31. Amarillas ED; Metlich MA. Microforma de fisura labial superior incompleta unilateral. Reporte de un caso. *Rev ADM [Internet].* 2018;75(5):278–82. Available from: www.medigraphic.com/adm
32. Solano Mendoza P; Bascones Martinez A. Consideraciones anatómicas durante la cirugía periodontal. *Av Periodon Implant.* 2014;26(1):11–17.
33. Agurto SP; Sandoval VP. Morfología del Arco Maxilar y Mandibular en Niños de Ascendencia Mapuche y no Mapuche. *Int. J. Morphol.* 2011 p. 1104-1108.
34. Carmona Marín LM. Diente cónico: presentación de dos casos. *Rev Médica Risaralda.* 2014;20(2):125-128.
35. De la Teja Ángeles E, Madrigal GE, Gutiérrez AD. Diagnóstico de paladar hendido submucoso. Características clínicas e informe de un caso. *Acta Pediatr Mex.* 2006;27(1):19–23.
36. Cabral Patrocinio M, De Los Santos Calderón PJ. Caracterización de microformas asociadas a labio y/o paladar hendido no sindrómico en familias que acuden a la Fundación Niños que ríen, Moca, República Dominicana Febrero-Julio 2021. [Tesis Doctor en Odontología] Santo Domingo: Facultad de Ciencias de la Salud, Escuela de Odontología, Universidad Nacional Pedro Henríquez Ureña; 2021.

37. Reardon W. The bedside dysmorphologist: Classic clinical signs in human malformation syndromes and their diagnostic significance. Second edi. Reardon W, editor. New York:Oxford; 2015. p. 1–305.
38. Colombia. Ministerio de Salud y Protección Social - Colciencias. Guía de práctica clínica Detección de anomalías congénitas en el recién nacido -2013 Guía No. 03. (Abril del 2013).
39. Suiza. Organización Mundial de la Salud, Centers for Disease Control and Prevention e International Clearinghouse for Birth Defects Surveillance and Research. [Internet]. Vigilancia de anomalías congénitas: manual para gestores de programas. (2015). Available from: <https://apps.who.int/iris/handle/10665/177241>
40. Panamonta V, Pradubwong S, Panamonta M, Chowchuen B. Global Birth Prevalence of Orofacial Clefts: A Systematic Review. *J Med Assoc Thai.* 2015;98(Suppl 7):S11–S21.
41. Twigg SRF, Wilkie AOM. New insights into craniofacial malformations. *Hum Mol Genet.* 2015;24(R1):R50–R59.
42. Yu W, Serrano M, Miguel SS, Ruest LB, Svoboda KK. Cleft lip and palate genetics and application in early embryological development. *Indian J Plast Surg.* 2009;42(suppl):S35-S50.
43. Pandey AS. Genetics in medicine. *J Kathmandu Med Coll.* 2017;6(1):1–2.
44. Lowry RB, Johnson CY, Gagnon F, Little J. Segregation analysis of cleft lip with or without cleft palate in the First Nations (Amerindian) people of British Columbia and review of isolated cleft palate etiologies. *Birth Defects Res Part A Clin Mol Teratol.* 2009;85(6):568–573.

45. Escobar LM, Prada-Arismendy J, Téllez C, Castellanos J. Genetic basis of orofacial cleft formation in humans. *Rev CES Odont.* 2013;26(1):57–67.
46. Moreno L, Bravo M, Valencia C, Jailler G, Villegas L, Lopez O AM. Evaluación epidemiológica genética de labio hendido con o sin paladar hendido en genealogías extendidas multigeneracionales pesquisadas en Colombia. *Acta Med Colomb.* 1998;23(4):156-61.
47. Goveas SR, Savitha NS. Role of Environmental Factors in the Etiology of Non-syndromic Cleft Lip Palate. *Int J Sci Stud.* 2017;4(12):21-26.
48. Regina Altoé S, Borges ÁH, Neves ATSC, Aranha AMF, Borba AM, Espinosa MM, et al. Influence of Parental Exposure to Risk Factors in the Occurrence of Oral Clefts. *J Dent.* 2020;21(2):119–126.
49. Xu LF, Zhou XL, Wang Q, Zhou JL, Liu YP, Ju Q, et al. A case-control study of environmental risk factors for nonsyndromic cleft of the lip and/or palate in Xuzhou, China. *Biomed Environ Sci.* 2015;28(7):535–538.
50. Sadler TW. Cabeza y cuello. In: Mendoza C, Segura Florez C, editors. *Embriología médica de Langman*. Fourth edi. Philadelphia: Wolters Kluwer; 2019. p. 454-494.
51. Ji Y, Garland MA, Sun B, Zhang S, Reynolds K, McMahon M, et al. Cellular and developmental basis of orofacial clefts. *Birth Defects Res.* 2020;112(19):1558-1587.
52. Carlson BM. Cabeza y cuello. In: Peña Melián AL, Tirado FV, editors. *Embriología humana y biología del desarrollo*. Fifth Edi. Barcelona: Elsevier, 2014. p. 309-349.

53. Moore KL, Persaud TVN, Torchia GM. Aparato faríngeo, cara y cuello. In: Álvarez Martínez C, editors. Embriología clínica. Eleventh Edi. Barcelona: Elsevier, 2013. p. 178-215.
54. Weng M, Chen Z, Xiao Q, Li R, Chen Z. A review of FGF signaling in palate development. *Biomed Pharmacother.* 2018;103:240-247.
55. Ji Y, Hao H, Reynolds K, McMahon M, Zhou CJ. Wnt Signaling in neural crest ontogenesis and oncogenesis. *Cells.* 2019;8(10):1173.
56. Lan Y, Xu J, Jiang R. Cellular and Molecular Mechanisms of Palatogenesis. *Curr Top Dev Biol.* 2015;115:59-84.
57. Lan Y, Ovitt CE, Cho ES, Maltby KM, Wang Q, Jiang R. Odd-skipped related 2 (Osr2) encodes a key intrinsic regulator of secondary palate growth and morphogenesis. *J. Dev.* 2004;131(13):3207-3216.
58. Yu L, Gu S, Alappat S, Song Y, Yan M, Zhang X, et al. Shox2-deficient mice exhibit a rare type of incomplete clefting of the secondary palate. *J. Dev.* 2005;132(19):4397–406.
59. Pauws E, Hoshino A, Bentley L, Prajapati S, Keller C, Hammond P, et al. Tbx22null mice have a submucous cleft palate due to reduced palatal bone formation and also display ankyloglossia and choanal atresia phenotypes. *Hum Mol Genet.* 2009;18(21):4171-4179.
60. Richardson RJ, Hammond NL, Coulombe PA, Saloranta C, Nousiainen HO, Salonen R, et al. Periderm prevents pathological epithelial adhesions during embryogenesis. *J Clin Invest.* 2014;124(9):3891-3900.

61. Moretti F, Marinari B, Lo Iacono N, Botti E, Giunta A, Spallone G, et al. A regulatory feedback loop involving p63 and IRF6 links the pathogenesis of 2 genetically different human ectodermal dysplasias. *J Clin Invest.* 2010;120(5):1570-1577.
62. Iwata J ichi, Suzuki A, Pelikan RC, Ho TV, Sanchez-Lara PA, Urata M, et al. Smad4-irf6 genetic interaction and TGF β -mediated IRF6 signaling cascade are crucial for palatal fusion in mice. *J. Dev.* 2013;140(6):1220-1230.
63. Ke CY, Xiao WL, Chen CM, Lo LJ, Wong FH. IRF6 is the mediator of TGF β 3 during regulation of the epithelial mesenchymal transition and palatal fusion. *Sci Rep.* 2015;5:12791.
64. Reynolds K, Kumari P, Rincon LS, Gu R, Ji Y, Kumar S, et al. Wnt signaling in orofacial clefts: Crosstalk, pathogenesis and models. *Dis Model Mech.* 2019;12(2):dmm037051.
65. Janečková E, Feng J, Li J, Rodriguez G, Chai Y. Dynamic activation of Wnt, Fgf, and Hh signaling during soft palate development. *PLoS One.* 2019;14(10):1–16.
66. Nakatomi M, Ludwig KU, Knapp M, Kist R, Lisgo S, Ohshima H, et al. Msx1 deficiency interacts with hypoxia and induces a morphogenetic regulation during mouse lip development. *J. Dev.* 2020;147(21):dev189175.
67. Pan X, Wang P, Yin X, Liu X, Li D, Li X, et al. Association between maternal MTHFR polymorphisms and nonsyndromic cleft lip with or without cleft palate in offspring, a meta-analysis based on 15 case-control studies. *Int J Fertil Steril.* 2015;8(4):463-480.

68. Tirado Amador LR, Anaya MVM, González Martínez FD. Genetic and epigenetic interactions related to non-syndromic cleft lip and palate. *Av Odontoestomatol.* 2016;32(1):21–34.
69. Gowans LJJ, Busch TD, Mossey PA, Eshete MA, Adeyemo WL, Aregbesola B, et al. The prevalence, penetrance, and expressivity of etiologic IRF6 variants in orofacial clefts patients from sub-Saharan Africa. *Mol Genet Genomic Med.* 2017;5(2):164-171.
70. Kousa YA, Schutte BC. Toward an orofacial gene regulatory network. *Dev Dyn.* 2016;245(3):220–232.
71. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–589.
72. Rigsby RE, Parker AB. Using the PyMOL application to reinforce visual understanding of protein structure. *Biochem Mol Biol Educ.* 2016;44(5):433–437.
73. Escalante CR, Yie J, Thanos D, Aggarwal AK. Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. *Nature.* 1998;391(6662):103-106.
74. Taniguchi T, Ogasawara K, Takaoka A, Tanaka N. IRF Family of transcription factors as regulators of host defense. *Annu Rev Immunol.* 2001;19:623-655.
75. Fujii Y, Shimizu T, Kusumoto M, Kyogoku Y, Taniguchi T, Hakoshima T. Crystal structure of an IRF-DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences. *EMBO J.* 1999;18(18):5028-5041.

76. Marazita ML. The evolution of human genetic studies of cleft lip and cleft palate. *Annu Rev Genomics Hum Genet.* 2012;13:263-283.
77. Ardinger HH, Buetow KH, Bell GI, Bardach J, VanDemark DR, Murray JC. Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. *Am J Hum Genet.* 1989;45(3):348–353.
78. Marazita ML, Leslie EJ. Genetics of Nonsyndromic Orofacial Clefting. In: Losee JE, Kirschner RE, editors. *Comprehensive Cleft Care, Volume 2. Second Edi.* Stuttgart: Georg Thieme Verlag KG; 2016. p. 207–224.
79. Grant SFA, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE, et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J Pediatr.* 2009;155(6):909–913.
80. Leslie EJ, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, et al. A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p 24.2, 17q23 and 19q13. *Hum Mol Genet.* 2016;25(13):2862–2872.
81. 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. *Hum Genet.* 2020;139(2):215–226.
82. Chiquet BT, Blanton SH, Burt A, Ma D, Stal S, Mulliken JB, et al. Variation in WNT genes is associated with non-syndromic cleft lip with or without cleft palate. *Hum Mol Genet.* 2008;17(14):2212–2218.

83. 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. *Am J Hum Genet.* 2020;107(1):124–136.
84. Li A, Qin G, Suzuki A, Gajera M, Iwata J, Jia P, et al. Network-based identification of critical regulators as putative drivers of human cleft lip. *BMC Med Genomics.* 2019;12(Suppl 1):16.
85. Yan F, Dai Y, Iwata J, Zhao Z, Jia P. An integrative, genomic, transcriptomic and network-assisted study to identify genes associated with human cleft lip with or without cleft palate. *BMC Med Genomics.* 2020;13(Suppl 5):39.
86. Haworth A, Savage H, Lench N. Diagnostic Genomics and Clinical Bioinformatics. In: Kumar D, Antonarakis S, editors. *Medical and Health Genomics* [Internet]. First edi. Hinxton: Elsevier Inc.; 2016. p. 37–50. Available from: <http://dx.doi.org/10.1016/B978-0-12-420196-5.00004-6>
87. Yin R, Kwok CK, Zheng J. Whole Genome Sequencing Analysis: Computational Pipelines and Workflows in Bioinformatics. In: Ranganathan S, Nakai K, Schönbach C, Gribskov M, editors. *Encyclopedia of Bioinformatics and Computational Biology*, vol. 3. Oxford: Elsevier; 2019. p. 176–183.
88. Technology Spotlight: Illumina Sequencing Technology. [Internet]. San Diego: Illumina, Inc; [cited 2020 Oct 23]. Available from: https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.

89. Ugur Sezerman O, Ulgen E, Seymen N, Melis Durasi I. Bioinformatics Workflows for Genomic Variant Discovery, Interpretation and Prioritization. Bioinformatics Tools for Detection and Clinical Interpretation of Genomic Variations [Internet]. 2019; Available from: <http://dx.doi.org/10.5772/intechopen.85524>
90. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405–424.
91. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(Database issue):D980-D985.
92. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308–311.
93. Pengelly RJ, Arias L, Martinez J, Upstill-Goddard R, Seaby EG, Gibson J, et al. Deleterious coding variants in multi-case families with non-syndromic cleft lip and/or palate phenotypes. *Sci Rep*. 2016;6:30457.
94. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
95. Zhang M, Li Q, Yu D, Yao B, Guo W, Xie Y, et al. GeNeCK: A web server for gene network construction and visualization. *BMC Bioinformatics*. 2019;20(1):12.

96. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49(D1): D605-D612.
97. Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, et al. GeneMANIA update 2018. *Nucleic Acids Res.* 2018;46(W1):W60-W64.
98. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353– D361.
99. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 2022;50(W1):W216–221.
100. Lushington GH. Comparative Modeling of Proteins. In: Kukol A, editors. *Molecular Modeling of Proteins. Second Edi.* New York: Springer; 2015. p. 309-330.
101. Vranken WF, Vuister GW, Bonvin AMJJ. NMR-Based modeling and refinement of protein 3D structures. In: Kukol A, editor. *Molecular Modeling of Proteins. Second edi.* New York: Humana Press; 2014. p. 351–380.
102. Universidad Internacional de Andalucía. *Bioinformática II : Estructura de Proteínas.* Sevilla; 2010.
103. Mullins JGL. Structural modelling pipelines in next generation sequencing projects. In: Donev R, editor. *Advances in Protein Chemistry and Structural Biology* [Internet]. First Edi. San Diego: Elsevier Inc.; 2012. p. 117–167. Available from: <http://dx.doi.org/10.1016/B978-0-12-394287-6.00005-7>

104. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting , position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22(22):4673–4680.
105. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: A simple approach to improve protein structure predictions. *Bioinformatics.* 2003;19(8):1015–1018.
106. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* 1993;2(9):1511–1519.
107. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Res Artic.* 1991;253(5016):164-170.
108. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993;26(2):283–291.
109. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinforma.* 2016;54:5.6.1-5.6.37.
110. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* 2004;32(web Server issue):W526–W531.
111. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296–W303.

112. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014;42(Web Server issue):252–W58.
113. Biasini M, Schmidt T, Bienert S, Mariani V, Studer G, Haas J, et al. OpenStructure: An integrated software framework for computational structural biology. *Acta Crystallogr D Biol Crystallogr.* 2013;69(5):701–709.
114. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 2011;27(3):343–350.
115. AlphaFold Protein Structure Database [Internet]. Hinxton: EMBL-EBI; [cited 2022 Nov 15]. Available from: <https://alphafold.ebi.ac.uk/about>
116. Mariani V, Biasini M, Barbato A, Schwede T. IDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics.* 2013;29(21):2722–2728.
117. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature.* 2021;596(7873):590–596.
118. Running AlphaFold2 in Google CoLab [Internet]. Massachusetts: python-based hierarchical environment for integrated xtallography; [cited 2022 Nov 15]. Available from: https://phenix-online.org/version_docs/1.20rc1-4398/reference/alphafold_in_colab.html
119. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods.* 2022;19(6):679–682.

120. Yin R, Feng BY, Varshney A, Pierce BG. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* 2022;31(8):e4379.
121. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv.* 2021. DOI: 10.1101/2021.10.04.463034.
122. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* 2021;373(6557):871–876.
123. Prieto-Martínez FD, Arciniega M, Medina-Franco JL. Acoplamiento Molecular: Avances Recientes y Retos. *Tip rev. espec. cienc. quím.-biol.* 2018;21(suppl 1):65–87.
124. Morrison JL, Breitling R, Higham DJ, Gilbert DR. A lock-and-key model for protein-protein interactions. *Bioinformatics.* 2006;22(16):2012–2019.
125. Raval K, Ganatra T. Basics, types and applications of molecular docking: A review. *IP Int J Compr Adv Pharmacol.* 2022;7(1):12–16.
126. Chen YC. Beware of docking!. *Trends Pharmacol Sci.* 2015;36(2):78–95.
127. Fan J, Fu A, Zhang L. Progress in molecular docking. *Quant Biol.* 2019;7(2):83–89.
128. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem.* 2009;30(16):2785–2791.

129. Kozakov D, Hallc DR, Xia B, Porter KA, Padhornya D, Yueh C, et al. The ClusPro web server for protein-protein docking. *Nat Protoc.* 2017;12(2):255–278.
130. Viswanath S, Ravikant DV, Elber R. DOCK / PIERR : Web Server for Structure Prediction of Protein – Protein Complexes. *Methods Mol Biol.* 2014;1137:199-207.
131. Roberts VA, Thompson EE, Piques ME, Perez MS, Ten Eyck LF. DOT2: Macromolecular Docking With Improved Biophysical Models. *J Comput Chem.* 2013;34(20):1743-1758.
132. Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ. FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res.* 2008;36(Web Server issue):W229–W232.
133. Garzon JI, Lopéz-Blanco JR, Pons C, Kovacs J, Abagyan R, Fernandez-Recio J, et al. FRODOCK: A new approach for fast rotational protein-protein docking. *Bioinformatics.* 2009;25(19):2544–2551.
134. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.* 2006;34(Web Server issue):W310–W314.
135. Dominguez C, Boelens R, Bonvin AM. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc.* 2003;125(7):1731–1737.
136. Ritchie DW. Evaluation of Protein Docking Predictions Using Hex 3.1 in CAPRI Rounds 1 and 2. *Proteins.* 2003;52(1):98–106.
137. Fernández-Recio J, Totrov M, Abagyan R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins.* 2003;52(1):113–117.

138. Jiménez-García B, Roel-Touris J, Romero-Durana M, Vidal M, Jiménez-González D, Fernández-Recio J. LightDock: A new multi-scale approach to protein-protein docking. *Bioinformatics*. 2018;34(1):49–55.
139. Ohue M, Shimoda T, Suzuki S, Matsuzaki Y, Ishida T, Akiyama Y. MEGADOCK 4.0: An ultra-high-performance protein-protein docking software for heterogeneous supercomputers. *Bioinformatics*. 2014;30(22):3281–3283.
140. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res*. 2005;33(Web Server issue):W363–W367.
141. Mitra P, Pal D. PRUNE and PROBE - Two modular web services for protein-protein docking. *Nucleic Acids Res*. 2011;39(Web Server issue):W229–W234.
142. Man-Kuang TC, Blundell TL, Fernandez-Recio J. pyDock: Electrostatics and Desolvation for Effective Scoring of Rigid-Body Protein–Protein Docking. *Proteins: Struct. Funct. Genet*. 2007;68(2):503–515.
143. Lyskov S, Gray JJ. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res*. 2008;36(Web Server issue):W233–W238.
144. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*. 2014;30(12):1771–1773.
145. Macalino SJY, Basith S, Clavio NAB, Chang H, Kang S, Choi S. Evolution of in silico strategies for protein-protein interaction drug discovery. *Molecules*. 2018;23(8):1963.

146. Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastiris PL, Karaca E, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol.* 2016;428(4):720–725.
147. Honorato RV, Koukos PI, Jiménez-García B, Tsaregorodtsev A, Verlato M, Giachetti A, et al. Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front Mol Biosci.* 2021;8:729513.
148. Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, et al. How Good is Automated Protein Docking?. *Proteins.* 2013;81(12):2159-2166.
149. Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure.* 2020;28(9):1071-1081.e3.
150. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDock: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* 2017;45(W1):W365–W373.
151. Yan Y, Tao H, He J, Huang SY. The HDock server for integrated protein–protein docking. *Nat Protoc.* 2020;15(5):1829–1852
152. Yan Y, Huang SY. Modeling Protein–Protein or Protein–DNA/RNA Complexes Using the HDock Webserver. In: Daisuke Kihara, editor. *Protein Structure Prediction [Internet]. Fourth Edi.* New York: Springer Protocols; 2020. p. 223-235. Available from: <http://www.springer.com/series/7651>
153. Bennett RL, French KS, Resta RG, Doyle DL. Standardized human pedigree nomenclature: Update and assessment of the recommendations of the National Society of Genetic Counselors. *J Genet Couns.* 2008;17(5):424–433.

154. Bertoli-Avella AM, Beetz C, Ameziane N, Rocha ME, Guatibonza P, Pereira C, et al. Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort. *Eur J Hum Genet.* 2021;29(1):141-153.
155. Racz C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: Ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics.* 2013;29(16):2041–2043.
156. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: Versatile and scalable detection of copy number variants. *Bioinformatics.* 2016;32(15):2375–2377.
157. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics.* 2016;32(8):1220–1222.
158. Agrovskii BS, Vorob'ev VV, Gurvich AS, Pokasov VV, Ushakov AN. Intensity Fluctuations of Pulsed Laser Radiation During Thermal Self-Interaction in a Turbulent Medium. *Sov J quantum Electron.* 1980;10(3):308–312.
159. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: A database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat.* 2013;34(9):E2393-E2402.
160. Riggs RR, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy number variants: a joint consensus recommendation of the American College of

- Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med.* 2020;22(2):245–257.
161. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. *Genome Res.* 2002;12(6):996–1006.
162. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: A curated collection of structural variation in the human genome. *Nucleic Acids Res.* 2014;42(Database issue):D986-D992.162.
163. Rangwala SH, Kuznetsov A, Ananiev V, Asztalos A, Borodin E, Evgeniev V, et al. Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome Res.* 2021;31(1):159–169.
164. SWISS-MODEL, Structure Assessment [Internet]. Switzerland: University of Basel[cited 2022 Oct 16]. Available from: <https://swissmodel.expasy.org/assess>
165. SAVESv6.0 - Structure Validation Server [Internet]. Los Angeles: The UCLA-DOE Institute for Genomics and Proteomics; [cited 2022 Nov 18]. Available from: <https://saves.mbi.ucla.edu/>
166. Sagendorf JM, Markarian N, Berman HM, Rohs R. DNAProDB: An expanded database and web-based tool for structural analysis of DNA-protein complexes. *Nucleic Acids Res.* 2020;48(D1):D277–D287.
167. Laskowski RA, Swindells MB. LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery. *J Chem Inf Model.* 2011;51:2778–2786.
168. Belinky F, Nativ N, Stelzer G, Zimmerman S, Stein TI, Safran M, et al. PathCards: Multi-source consolidation of human biological pathways. *Database.* 2015; 2015:bav006.

169. Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 2021;30(1):187–200.
170. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–D613.
171. Pratt D, Chen J, Welker D, Rivas R, Pillich R, Rynkov V, et al. NDEX, the Network Data Exchange. *Physiol Behav.* 2015;1(4):302-305.
172. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, et al. WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* 2018;46(D1):D661–D667.
173. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 2018;46(D1):D380–D386.
174. Rahmati S, Abovsky M, Pastrello C, Kotlyar M, Lu R, Cumbaa CA, et al. PathDIP 4: An extended pathway annotations and enrichment analysis resource for human, model organisms and domesticated species. *Nucleic Acids Res.* 2020;48(D1):D479–D488.
175. Del Toro N, Shrivastava A, Ragueneau E, Meldal B, Combe C, Barrera E, et al. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* 2022;50(D1):D648–D653.

176. Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics*. 2014;47:8.13.1-8.13.24.
177. Dai J, Yu H, Si J, Fang B, Shen SG. Irf6-related gene regulatory network involved in palate and lip development. *J Craniofac Surg*. 2015;26(5):1600–1605.
178. David Bioinformatics Database: Functional Annotation Tools [Internet]. Maryland:DAVID Bioinformatics Resources Laboratory of Human Retrovirology and Immunoinformatics; [cited 2022 Jan 6]. Available from: <https://david.ncifcrf.gov/tools.jsp>
179. Lumley J, Watson L, Watson M, Bower C. Periconceptional supplementation with folate and/or multivitamins for preventing neural tube defects. *Cochrane Database Syst Rev*. 2001;(3):CD001056.
180. Yadira Inza. Canva. [Internet]. España:Canva; [cited 2023 Jan 12]. Available from: <https://www.canva.com/p/yadirainza/>
181. Invitae. Invitae Family History Tool [Internet]. United States:Invitae; [cited 2022 Jan 11]. Available from: <https://familyhistory.invitae.com/login/?next=/>
182. Rojas MY, Alonso LA, Sarmiento VA, Eljach LY, Usaquén W. Structure analysis of the la Guajira-Colombia population: A genetic, demographic and genealogical overview. *Ann Hum Biol*. 2013;40(2):119–31.
183. Ignacio J, Aizpún L, Muñoz ADA, Lozano MF, Endocrinología U De, Pediatría S De, et al. Hiperplasia suprarrenal congénita. 2019;141–56.
184. Bruque CD. Análisis de variantes génicas en el gen CYP21A2. [Tesis doctoral en Ciencias Biológicas] Buenos Aires; Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires; 2019.

185. Araújo RS, Mendonca BB, Barbosa AS, Lin CJ, Marcondes JAM, Billerbeck AEC, et al. Microconversion between CYP21A2 and CYP21A1P promoter regions causes the nonclassical form of 21-hydroxylase deficiency. *J Clin Endocrinol Metab.* 2007;92(10):4028–4034.
186. Blanché H, Vexiau P, Clauin S, Gall IL, Fiet J, Mornet E, et al. Exhaustive screening of the 21-hydroxylase gene in a population of hyperandrogenic women. *Hum Genet.* 1997;101(1):56–60.
187. Chi DV, Tran TH, Nguyen DH, Luong LH, Le PT, Ta MH, et al. Novel variants of CYP21A2 in Vietnamese patients with congenital adrenal hyperplasia. *Mol Genet Genomic Med.* 2019;7(5):e623.
188. Simonetti L, Bruque CD, Fernández CS, Benavides-Mori B, Delea M, Kolomenski JE, et al. CYP21A2 mutation update: Comprehensive analysis of databases and published genetic variants. *Hum Mutat.* 2018;39(1):5–22.
189. gnomAD, Genome Aggregation Database [Internet]. Cambridge: Broad Institute; [cited 2022 Oct 13]. Available from: <https://gnomad.broadinstitute.org/>
190. Full data view for gene CYP21A2 - Global Variome shared LOVD [Internet]. Netherlands:LOVD v.3.0 - Leiden Open Variation Database; [cited 2022 Oct 13]. Available from: <https://databases.lovd.nl/shared/view/CYP21A2>
191. Kornhuber J, Rhein C, Müller CP, Mühle C. Secretory sphingomyelinase in health and disease. *Biol Chem.* 2015;396(6–7):707–736.
192. De Marco G, Agretti P, Montanelli L, Cosmo CD, Bagattini B, De Servi M, et al. Identification and functional analysis of novel dual oxidase 2 (DUOX2) mutations in

- children with congenital or subclinical hypothyroidism. *J Clin Endocrinol Metab.* 2011;96(8):E1335–E1339.
193. Morales C, Soler A, Margarit E, Madrigal I, Sánchez A. Trisomy of 19.4 Mb region of chromosome 22 and subtelomeric 17p identified in a male without clinical affection. *Am J Med Genet A.* 2007;143A(20):2423-2429.
194. Wang YE, Ramirez DA, Chang TC, Berrocal A. Peters plus syndrome and Chorioretinal findings associated with B3GLCT gene mutation-a case report. *BMC Ophthalmol.* 2020;20(1):118.
195. Domingos IF, Falcão DA, Hatzlhofer BL, Cunha AF, Santos MN, Albuquerque DM, et al. Influence of the β s haplotype and α -thalassemia on stroke development in a Brazilian population with sickle cell anaemia. *Ann Hematol.* 2014;93(7):1123–1129.
196. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz Jr KM, et al. The Amber biomolecular simulation programs. *J Comput Chem.* 2005;26(16):1668–1688.
197. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* 2013;41(Web Server issue 1):W384–W338.
198. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 2007;35(Web Server issue):W375–W383.
199. Sundararaj S, Williams SJ, Casarotto MG. 7JM4: IRF Transcription Factor [Internet]. United Kingdom:RCSB PDB; [cited 2022 Dec 14]. Available from: <https://www.rcsb.org/structure/7JM4>

200. Wallace AC, Laskowski RA, Thornton JM. Ligplot: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 1995;8(2):127–134.
201. Agnarelli A, Omari EK, Mancini EJ. 7OOT: X-ray Structure of Interferon Regulatory Factor 4 DNA binding domain bound to an interferon-stimulated response element [Internet]. United Kingdom: RCSB PDB; 2021 [cited 2022 Nov 24]. Available from: <https://www.rcsb.org/structure/7OOT>
202. Piñero J, Saüch J, Sanz F, Furlong LI. The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. *Comput Struct Biotechnol J.* 2021;19:2960–2967.
203. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of biomolecular interaction networks. *Genome Res.* 2013;13(11):2498-2504.
204. Neves LT, Dionísio TJ, Garbieri TF, Parisi VA, Oliveira FV, Oliveira TM, et al. Novel rare variations in IRF6 in subjects with non-syndromic cleft lip and palate and dental agenesis. *Oral Dis.* 2019;25(1):223–233.
205. Zuccherro TM, Cooper ME, Maher BS, Daack-Hirsch S, Nepomuceno B, Ribeiro L, et al. Interferon Regulatory Factor 6 (IRF6) Gene Variants and the Risk of Isolated Cleft Lip or Palate. *New Engl J Med.* 2004; 351(8):769-780.
206. Vieira AR, Seymen F, Patir A, Menezes R. Evidence of linkage disequilibrium between polymorphisms at the IRF6 locus and isolate tooth agenesis, in a Turkish population. *Arch Oral Biol.* 2008;53(8):780–784.

207. De Lima RL, Hoper SA, Ghassibe M, Cooper ME, Rorick NK, Kondo S, et al. Prevalence and nonrandom distribution of exonic mutations in interferon regulatory factor 6 in 307 families with Van der Woude syndrome and 37 families with popliteal pterygium syndrome. *Genet Med*. 2009;11(4):241–247.
208. Leslie EJ, Standley J, Compton J, Bale S, Schutte BC, Murray JC. Comparative analysis of IRF6 variants in families with Van der Woude syndrome and popliteal pterygium syndrome using public whole-exome databases. *Genet Med*. 2013;15(5):338–344.
209. Charzewska A, Obersztyn E, Hoffman-Zacharska D, Lenart J, Poznański J, Bal J. Novel mutations in the IRF6 gene on the background of known polymorphisms in Polish patients with orofacial clefting. *Cleft Palate-Craniofacial J*. 2015;52(5):e161–167.
210. Alade AA, Buxo-Martinez CJ, Mossey PA, Gowans LJJ, Eshete MA, Adeyemo WL, et al. Non-random distribution of deleterious mutations in the DNA and protein-binding domains of IRF6 are associated with Van Der Woude syndrome. *Mol Genet Genomic Med*. 2020;8(8):e1355.
211. Zhao H, Zhang M, Zhong W, Zhang J, Huang W, Zhang Y, et al. A novel IRF6 mutation causing non-syndromic cleft lip with or without cleft palate in a pedigree. *Mutagenesis*. 2018;33(3):195–202.
212. Koillinen H, Lahermo P, Rautio J, Hukki J, Peyrard-Janvid M, Kere J. A genome-wide scan of non-syndromic cleft palate only (CPO) in Finnish multiplex families. *J Med Genet*. 2005;42(2):177–184.

213. Pegelow M, Peyrard-Janvid M, Zucchelli M, Fransson I, Larson O, Kere J, et al. Familial non-syndromic cleft lip and palate - Analysis of the IRF6 gene and clinical phenotypes. *Eur J Orthod.* 2008;30(2):169–175.
214. Swaminathan GJ, Bragin E, Chatzimichali EA, Corpas M, Bevan AP, Wright CF, et al. Decipher: Web-based, community resource for clinical interpretation of rare variants in developmental disorders. *Hum Mol Genet.* 2012;21(R1):37–44.
215. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, et al. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat Genet.* 2002;32(2):285–289.
216. Schutte BC, Murray JC. The many faces and factors of orofacial clefts. *Hum Mol Genet.* 1999;8(10):1853–1859.
217. Wu Q, Peng Z, Zhang Y, Yang J. COACH-D: Improved protein-ligand binding sites prediction with refined ligand-binding poses through molecular docking. *Nucleic Acids Res.* 2018;46(W1):W438–W442.
218. Rizos M, Spyropoulos MN. Van der Woude syndrome: A review. Cardinal signs, epidemiology, associated features, differential diagnosis, expressivity, genetic counselling and treatment. *Eur J Orthod.* 2004;26(1):17–24.
219. Desmyter L, Ghassibe M, Revencu N, Boute O, Lees M, François G, et al. IRF6 screening of syndromic and a priori non-syndromic cleft lip and palate patients: Identification of a new type of minor VWS sign. *Mol Syndromol.* 2010;1(2):67–74.

220. Spencer LSB, Gondim DD, Alves RV, Silva RBHC, Lopes VDF. Popliteal pterygium syndrome: case report and literature review. *Rev Bras Cir Plást.* 2012;27(3):482–486.
221. Busche A, Hehr U, Sieg P, Gillessen-Kaesbach G. Van der Woude and Popliteal Pterygium Syndromes: Broad intrafamilial variability in a three generation family with mutation in IRF6. *Am J Med Genet Part A.* 2016;170(9):2404–2407.
222. Matsuzawa N, Kondo S, Shimozato K, Nagao T, Nakano M, Tsuda M, et al. Two missense mutations of the IRF6 gene in two Japanese families with popliteal pterygium syndrome. *Am J Med Genet Part A.* 2010;152A(9):2262–2267.
223. Rutledge KD, Barger C, Grant JH, Robin NH. IRF6 mutations in mixed isolated familial clefting. *Am J Med Genet Part A.* 2010;152A(12):3107–3109.
224. Leslie EJ, Koboldt DC, Kang CJ, Ma L, Hecht JT, Wehby GL, et al. IRF6 mutation screening in non-syndromic orofacial clefting: analysis of 1521 families. *Clin Genet.* 2016;90(1):28–34.
225. Kousa YA, Fuller E, Schutte BC. IRF6 and AP2A Interaction Regulates Epidermal Development. *J Invest Dermatol.* 2018;138(12):2578–2588.
226. Biedziak B, Firlej E, Dąbrowska J, Bogdanowicz A, Zadurska M, Mostowska A. Novel Candidate Genes for Non-Syndromic Tooth Agenesis Identified Using Targeted Next-Generation Sequencing. *J Clin Med.* 2022;11(20):6089.
227. Vieira AR, Modesto A, Meira R, Barbosa AR, Lidral AC, Murray JC. Interferon regulatory factor 6 (IRF6) and fibroblast growth factor receptor 1 (FGFR1) contribute to human tooth agenesis. *Am J Med Genet Part A.* 2007;143(6):538–545.

228. Bezamat M, Zhou Y, Park T, Vieira AR. Genome-wide family-based study in torus palatinus affected individuals. *Arch Oral Biol.* 2021;130:105221.
229. Obniski M. An Investigation of Palatal Rugae Patterns and Torus Palatinus in Unaffected Relatives of Individuals with Orofacial Clefts [Thesis Master of Science degree in Orthodontics] Iowa: The University of Iowa; 2020.
230. Auškalnis A, Bernhardt O, Putniene E, Šidlauskas A, Andriuškevičiute I, Basevičiene N. Oral bony outgrowths: Prevalence and genetic factor influence. Study of twins. *Med.* 2015;51(4):228–232.
231. Ke CY, Xiao WL, Chen CM, Lo LJ, Wong FH. IRF6 is the mediator of TGF β 3 during regulation of the epithelial mesenchymal transition and palatal fusion. *Sci Rep.* 2015;5:12791.
232. Cuervo R, Covarrubias L. Death is the major fate of medial edge epithelial cells and the cause of basal lamina degradation during palatogenesis. *Development.* 2004;131(1):15–24.
233. Jin JZ, Ding J. Analysis of cell migration, transdifferentiation and apoptosis during mouse secondary palate fusion. *Development.* 2006;133(17):3341–3347.
234. Song H, Wang X, Yan J, Mi N, Jiao X, Hao Y, et al. Association of single-nucleotide polymorphisms of CDH1 with nonsyndromic cleft lip with or without cleft palate in a northern Chinese Han population. *Med.* 2017;96(5):e5574.
235. Richardson RJ, Dixon J, Jiang R, Dixon MJ. Integration of IRF6 and Jagged2 signalling is essential for controlling palatal adhesion and fusion competence. *Hum Mol Genet.* 2009;18(14):2632–2642.

236. Martinelli M, Palmieri A, Carinci F, Scapoli L. Non-syndromic Cleft Palate: An Overview on Human Genetic and Environmental Risk Factors. *Front Cell Dev Biol.* 2020;8:592271.
237. McDade SS, Henry AE, Pivato GP, Kozarewa I, Mitsopoulos C, Fenwick K, et al. Genome-wide analysis of p63 binding sites identifies AP-2 factors as co-regulators of epidermal differentiation. *Nucleic Acids Res.* 2012;40(15):7190–7206.
238. Xu T, Du M, Bu X, Yuan D, Gu Z, Yu P, et al. Identification of a novel TP63 mutation causing nonsyndromic cleft lip with or without cleft palate. *BMC Med Genomics.* 2021;14(1):53.
239. Maili L, Letra A, Silva R, Buchanan EP, Mulliken JB, Greives MR, et al. PBX-WNT-P63-IRF6 pathway in nonsyndromic cleft lip and palate. *Birth Defects Res.* 2020;112(3):234–244.
240. Ferretti E, Li B, Zewdu R, Wells V, Hebert JM, Karner C, et al. A Conserved Pbx-Wnt-p63-Irf6 Regulatory Module Controls Face Morphogenesis by Promoting Epithelial Apoptosis. *Dev Cell.* 2011;21(4):627–641.
241. Yu Y, Zuo X, He M, Gao J, Fu Y, Qin C, et al. Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat Commun.* 2017;8:14364.
242. Uslu VV, Petretich M, Ruf S, Langenfeld K, Fonseca NA, Marioni JC, et al. Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat Genet.* 2014;46(7):753–758.
243. Dunkhase E, Ludwig KU, Knapp M, Skibola CF, Figueiredo JC, Hosking FJ, et al. Nonsyndromic cleft lip with or without cleft palate and cancer: Evaluation of a

- possible common genetic background through the analysis of GWAS data.
Genomics Data. 2016;10:22–29.
244. Botti E, Spallone G, Moretti F, Marinari B, Pinetti V, Galanti S, et al. Developmental factor IRF6 exhibits tumor suppressor activity in squamous cell carcinomas. *Proc Natl Acad Sci U S A*. 2011;108(33):13710–13715.
245. Richardson RJ, Dixon J, Malhotra S, Hardman MJ, Knowles L, Boot-Handford RP, et al. *Irf6* is a key determinant of the keratinocyte proliferation-differentiation switch. *Nat Genet*. 2006;38(11):1329–1334.
246. Fakhouri WD, Rhea L, Du T, Sweezer E, Morrison H, Fitzpatrick D, et al. MCS9.7 enhancer activity is highly, but not completely, associated with expression of *Irf6* and p63. *Dev Dyn*. 2012;241(2):340–349.
247. Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, et al. Disruption of an AP-2 α binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet*. 2008;40(11):1341–1347.
248. De La Garza G, Schleiffarth JR, Dunnwald M, Mankad A, Weirather JL, Bonde G, et al. Interferon regulatory factor 6 promotes differentiation of the periderm by activating expression of grainyhead-like 3. *J Invest Dermatol*. 2013;133(1):68–77.
249. Li MJ, Shi JY, Zhang BH, Chen QM, Shi B, Jia ZL. Targeted re-sequencing on 1p22 among non-syndromic orofacial clefts from Han Chinese population. *Front Genet*. 2022;13:947126.

250. Yu Q, Deng Q, Fu F, Li R, Zhang W, Wan J, et al. A novel splicing mutation of ARHGAP29 is associated with nonsyndromic cleft lip with or without cleft palate. *J Matern Fetal Neonatal Med.* 2022;35(13):2499-2506.
251. Leslie EJ, Mansilla MA, Biggs LC, Schuette K, Bullard S, Cooper M, et al. Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome-wide association on chromosome 1p22. *Birth Defects Res Part A - Clin Mol Teratol.* 2012;94(11):934–942.
252. Schlessinger K, Hall A, Tolwinski N. Wnt signaling pathways meet Rho GTPases. *Genes Dev.* 2009;23(3):265–277.
253. Ferretti E, Li B, Zewdu R, Wells V, Hebert JM, Karner C, et al. A Conserved Pbx-Wnt-p63-Irf6 Regulatory Module Controls Face Morphogenesis by Promoting Epithelial Apoptosis. *Dev Cell.* 2011;21(4):627–641.
254. Velázquez-Aragón JA, Alcántara-Ortigoza MA, Estandia-Ortega B, Reyna-Fabián ME, Méndez-Adame CD, González-Del Angel A. Gene interactions provide evidence for signaling pathways involved in cleft lip/palate in humans. *J Dent Res.* 2016;95(11):1257–1264.
255. Park KW, Urness LD, Senchuk MM, Colvin CJ, Wythe JD, Chien CB, et al. Identification of new netrin family members in zebrafish: Developmental expression of netrin2 and netrin4. *Dev Dyn.* 2005;234(3):726–731.
256. Jiang SY, Shi JY, Lin YS, Duan SJ, Chen X, Jiao JJ, et al. NTN1 gene was risk to non-syndromic cleft lip only among Han Chinese population. *Oral Dis.* 2019;25(2):535–542.

257. Qian Y, Li D, Ma L, Zhang H, Gong M, Li S, et al. TPM1 polymorphisms and nonsyndromic orofacial clefts susceptibility in a Chinese Han population. *Am J Med Genet Part A*. 2016;170(5):1208–1215.
258. Fontoura C, Silva RM, Granjeiro JM, Letra A. Further evidence of association of ABCAA gene with cleft lip/palate. *Eur J Oral Sci*. 2012;120(6):553–557.
259. Nogueira JM, Hawrot K, Sharpe C, Noble A, Wood W, Jorge EC, et al. The emergence of Pax7-expressing muscle stem cells during vertebrate head muscle development. *Front Aging Neurosci*. 2015;7:62.
260. Khan MI, CS P, Srinath N. Role of PAX7 Gene rs766325 and rs4920520 Polymorphisms in the Etiology of Non-syndromic Cleft Lip and Palate: A Genetic Study. *Glob Med Genet*. 2022;09(03):208–211.
261. Li Q, Kim Y, Suktitipat B, Hetmanski JB, Marazita ML, Duggal P, et al. Gene-Gene Interaction Among WNT Genes for Oral Cleft in Trios. *Genet Epidemiol*. 2015;39(5):385–394.
262. Kwon M, Hanna E, Lorang D, He M, Quick JS, Adem A, et al. Functional characterization of Filamin A interacting protein 1-like, a novel candidate for antivascular cancer therapy. *Cancer Res*. 2008;68(18):7332–7341.
263. Velázquez-Aragón JA, Alcántara-Ortigoza MA, Estandia-Ortega B, Reyna-Fabián ME, Cruz-Fuentes C, Villagómez S, et al. Association of interactions among the IRF6 gene, the 8q24 region, and maternal folic acid intake with non-syndromic cleft lip/palate in Mexican Mestizos. *Am J Med Genet Part A*. 2012;158 A(12):3207–3210.

264. Xiao Y. Detecting Gene-Gene Interactions for Cleft Lip with / without Cleft Palate in Targeted Sequencing Data. [Thesis Master of Health Science, Genetic Epidemiology] Baltimore: Johns Hopkins University; 2015.
265. Paul BJ, Palmer KJ, Rhea L, Carlson M, Sharp JC, Pratt CH, et al. The *Mafb* cleft-associated variant H131Q is not required for palatogenesis in the mouse. *Dev Dyn*. 2021;250(10):1463–1476.
266. Mi N, Hao Y, Jiao X, Zheng X, Song T, Shi J, et al. Association study of single nucleotide polymorphisms of *MAFB* with non-syndromic cleft lip with or without cleft palate in a population in Heilongjiang Province, northern China. *Br J Oral Maxillofac Surg*. 2014;52(8):746–750.
267. Bush JO, Jiang R. Palatogenesis: Morphogenetic and molecular mechanisms of secondary palate development(*Development*, 139, (231-243)). *Development*. 2012;139(4):828.
268. Matsumura K, Taketomi T, Yoshizaki K, Arai S, Sanui T, Yoshiga D, et al. *Sprouty2* controls proliferation of palate mesenchymal cells via fibroblast growth factor signaling. *Biochem Biophys Res Commun*. 2011;404(4):1076–1082.
269. Zhang BH, Shi JY, Lin YS, Shi B, Jia ZL. *VAX1* gene associated non-syndromic cleft lip with or without palate in Western Han Chinese. *Arch Oral Biol*. 2018;95:40–43.
270. Murphy P, Armit C, Hill B, Venkataraman S, Frankel P, Baldock RA, et al. Integrated analysis of Wnt signalling system component gene expression. *Development*. 2022;149(16):dev200312.

271. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Hum Mutat.* 2015;36(8):815–822.