



Universidad Católica de Santa María

Escuela de Postgrado

Doctorado en Ciencias de la Salud

Análisis molecular del gen humano P53 en pterigión primario

Tesis presentada por:

Manzaneda Murguía, Berioska

ORCID: 0000-0002-4474-893X

Para optar el Grado Académico de Doctor en Ciencias de la Salud

Asesora:

Dra. Pacheco Aranibar, Jani

ORCID: 0000-0002-7898-4412

Arequipa - Perú

2026

UNIVERSIDAD CATÓLICA DE SANTA MARÍA
ESCUELA DE POSTGRADO
DICTAMEN APROBACIÓN DE BORRADOR DE TESIS

Arequipa, 04 de Abril del 2025

Dictamen: 001300-C-EPG-2025

Visto el borrador del expediente 001300, presentado por:

2017009122 - MANZANEDA MURGUIA BERIOSKA

Titulado:

ANÁLISIS MOLECULAR DEL GEN HUMANO P53 EN PTERIGIÓN PRIMARIO

Nuestro dictamen es:

APROBADO

29285302 - GUTIERREZ ARANIBAR ROXANA JACQUELINE CANDELARIA
DICTAMINADOR



29201360 - VILLANUEVA SALAS JOSE ANTONIO
DICTAMINADOR



29228284 - DAVILA DEL CARPIO GONZALO HERMILIO
DICTAMINADOR



00476696 - BERNABE ORTIZ JULIO CESAR
DICTAMINADOR



16423061 - FERNANDEZ FERNANDEZ FERNANDO ALBERTO
DICTAMINADOR



Análisis molecular del gen humano P53 en pterigión primario

INFORME DE ORIGINALIDAD

18%

INDICE DE SIMILITUD

19%

FUENTES DE INTERNET

4%

PUBLICACIONES

11%

TRABAJS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1	Submitted to Universidad Católica de Santa María	7%
	Trabajo del estudiante	
2	cybertesis.unmsm.edu.pe	2%
	Fuente de Internet	
3	www.seqc.es	2%
	Fuente de Internet	
4	revistahematologia.com.ar	2%
	Fuente de Internet	
5	tesis.ucsm.edu.pe	1%
	Fuente de Internet	
6	Benjamín Rodríguez-Santiago, Lluís Armengol. "Tecnologías de secuenciación de nueva generación en diagnóstico genético pre- y postnatal", Diagnóstico Prenatal, 2012	1%
	Publicación	
7	uvadoc.uva.es	1%
	Fuente de Internet	

DEDICATORIA

Dedico esta tesis a mis hijos, inspiración y complemento de mi vida.



AGRADECIMIENTOS

Agradezco al Dr. Julio C. Bernabé Ortiz, por su apoyo en la elaboración de esta tesis y por ser ejemplo de dedicación en su labor como docente.



EPÍGRAFE

“Sé el cambio que quieres ver en el mundo”

Mahatma Gandhi



RESUMEN

La nueva tecnología de secuenciación de próxima generación (NGS) RNA-Seq ha permitido realizar estudios detallados de la genética del pterigion.

La presente tesis tiene como objetivo describir el perfil de expresión génica del pterigion en la población de Puno. Para esto, se trabajó con siete muestras de pterigion primario, obtenidas de pacientes sometidos a cirugía de plastia libre en el período 2019-2020 en el Hospital III Juliaca EsSalud. El ARN total fue aislado de las biopsias y sometidos a secuenciación masiva mediante RNA-Seq, siguiendo el protocolo de Illumina.

El análisis epidemiológico mostró un predominio del grupo etario de 30 a 49 años; el 57.14 % de los pacientes fueron hombres y el 42.86 % mujeres, con una edad promedio de 46.57 años.

En el análisis bioinformático de la secuenciación masiva (RNA Seq), realizado con el software IGV y utilizando como referencia el genoma GRCh38.p13, se identificaron 19 261 genes en cada muestra de pterigión. Los genes con mayor nivel de expresión fueron RPS29, AHNAK, MTRNR2L12, PCBD2.

Todas las muestras expresaron el gen p53, identificándose cinco mutaciones SNV (variantes de un solo nucleótido) según la base de datos del ensamblaje del genoma humano GRCh38.p13: SNV C > G del cromosoma: 17: 7676154 en el exón 4, SNV A > C del cromosoma: 17: 7669372 en el exón 8, SNV C > T del cromosoma: 17: 7669 124 en el exón 8, SNV G > T, A del cromosoma: 17: 7668836 en el exón 8, SNV C > G, T del cromosoma: 17: 7668799 en el exón 8.

Asimismo, el análisis bioinformático con el software Arriba (versión 1.2.0) evidenció la presencia de 13 genes de fusión. Algunos no presentaron características funcionales relevantes; sin embargo, los genes de fusión PSMC3 - NXF1 y WNK1 - ZNF79 conservaron dominios de proteína retenidos y fusiones en el marco, indicando que sintetizaron las proteínas potencialmente implicadas como potentes impulsores en la génesis del pterigion.

Las mutaciones identificadas en el gen p53, así como los genes hallados y las proteínas de fusión reportadas en nuestras muestras de pterigion, son reportados por primera vez en esta investigación. La relevancia de estos hallazgos radica en que podrían impulsar el desarrollo de nuevas estrategias diagnósticas y terapias dirigidas, abriendo el camino hacia una medicina personalizada para el tratamiento del pterigion.

Palabras clave: Pterigion, genes de fusión, p53.

ABSTRACT

The new RNA-Seq next-generation sequencing (NGS) technology has allowed us to carry out studies of the genetics of pterygium.

The present thesis aims to describe the expression profile of pterygium in the population of Puno. We worked with 7 samples of primary pterygium undergoing free plasty surgery in the period 2019-2020 at the Juliaca EsSalud Hospital III. Total RNA was isolated from the biopsies obtained from the patients and they were subjected to massive sequencing by RNA-Seq using the Illumina protocol.

The epidemiological analysis of the patients shows a predominance of the age group of 30-49 years, 57.14% were men and 42.86% women, the average age was 46.57 years.

In the bioinformatics analysis of massive sequencing (RNA Seq), using the IGV software and using the GRCh38.p13 genome, 19,261 genes were found in each pterygium sample, the most expressed genes were RPS29, AHNAK, MTRNR2L12, PCBD2. All pterygium samples expressed the p53 gene, 5 SNV (single nucleotide variant) mutations of p53 were found taking as reference the human genome assembly database GRCh38.p13: SNV C > G of chromosome: 17: 7676154 in exon 4, SNV A > C of chromosome: 17: 7669372 in exon 8, SNV C > T of chromosome: 17: 7669124 in exon 8, SNV G > T, A of chromosome: 17: 7668836 in exon 8, SNV C > G, T of chromosome: 17: 7668799 in exon 8.

The bioinformatic analysis of the sequences using the Arriba software (version 1.2.0) shows the presence of 13 fusion genes, some of which did not show any specific characteristics. functional, unlike the fusion genes PSMC3 - NXF1 and WNK1 - ZNF79 that have retained protein domains and in-frame fusions, meaning that they synthesize proteins that can act as potent drivers in pterygium genesis.

The mutations described in the p53 gene, the genes found and the fusion proteins reported in our pterygium samples, are described for the first time in the present thesis.

The importance of the findings will promote the use of new diagnostic alternatives, targeted therapies, opening the way towards individualized medicine in pterygium.

Keywords: Pterygium, fusion genes, p53.

ÍNDICE GENERAL

DEDICATORIA

AGRADECIMIENTOS

EPÍGRAFE

RESUMEN

ABSTRACT

INTRODUCCIÓN..... 1

HIPOTÉISIS DE LA INVESTIGACIÓN 2

OBJETIVOS..... 3

Objetivo general 3

Objetivos específicos..... 3

CAPÍTULO I MARCO TEÓRICO Y CONCEPTUAL 4

1.1 Definición de pterigion..... 4

1.2 Patogénesis 5

1.2.1 Radiación Ultravioleta..... 5

1.2.2 Infecciones Virales 7

1.2.3 Proteínas relacionadas a la apoptosis..... 7

1.2.4 Moléculas de adhesión celular..... 9

1.2.5 Proteínas relacionadas con la proliferación celular 9

1.2.6 Proteínas de shock térmico 9

1.2.7 Proteínas de unión estrecha 10

1.2.8 Proteínas de la matriz extracelular..... 10

1.2.9 Metaloproteinasas (MMP)..... 10

1.2.10 Interleuquinas 11

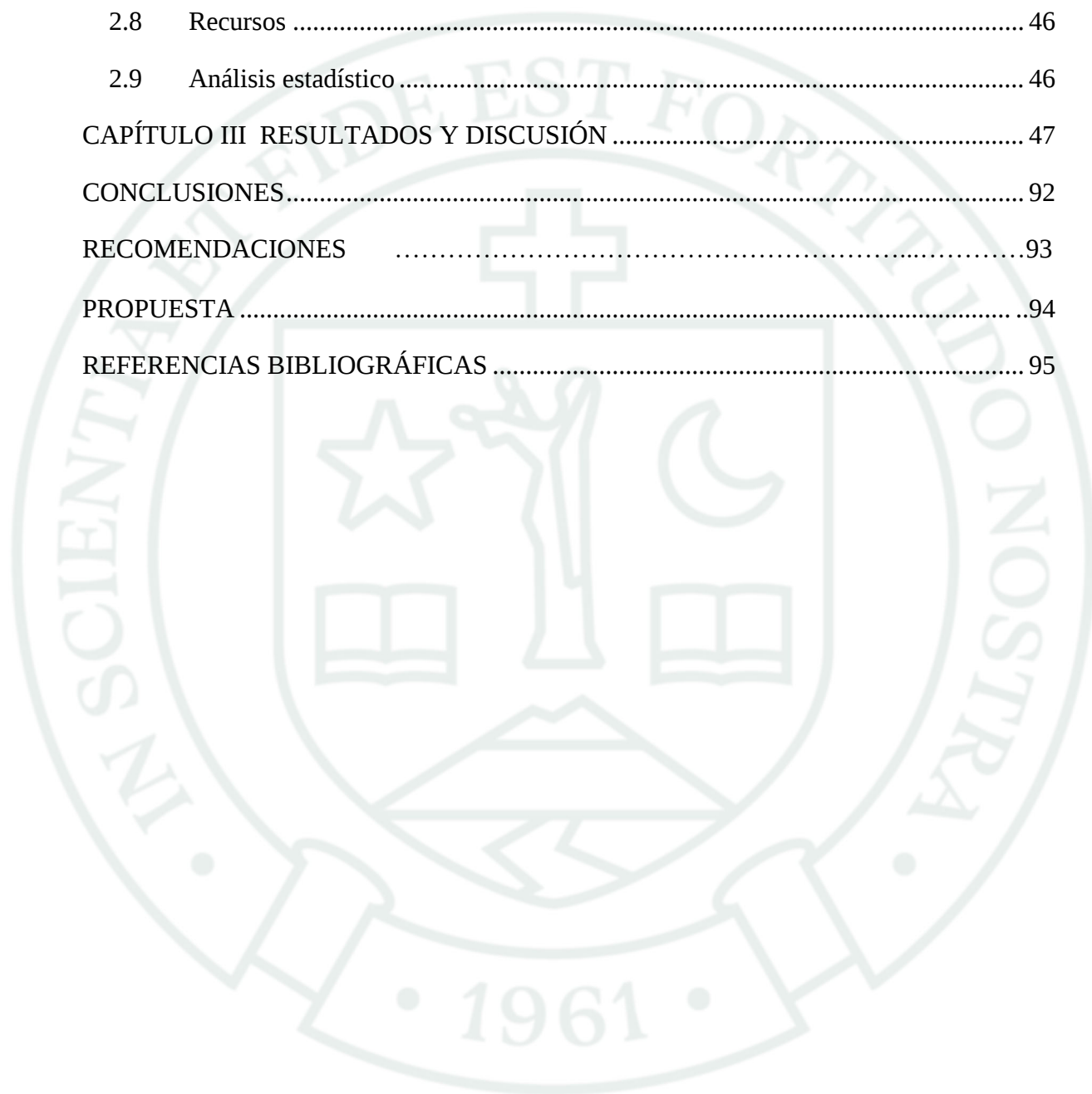
1.2.11 Factores de crecimiento 11

1.2.12 Genes supresores de tumores - Gen p53..... 12

1.3 Técnicas de secuenciación masiva (NGS)..... 13

1.3.1	Introducción.....	13
1.3.2	Técnicas de secuenciación.....	14
1.3.3	Preparación de bibliotecas	15
1.3.4	Enriquecimiento del blanco (<i>target enrichment</i>).....	16
1.3.5	Amplificación del templado y secuenciación masiva paralela.....	16
1.3.6	Procesamiento de datos	18
1.3.7	Plataformas de secuenciación de nueva generación	23
1.3.8	Utilidad de la secuenciación de nueva generación para detectar variantes genéticas	25
1.3.9	Aplicaciones de la secuenciación de nueva generación en la práctica clínica	26
1.4	Análisis de antecedentes investigativos.....	32
1.4.1	Expresión del gen supresor de tumores p53 en el pterigión	32
CAPÍTULO II METODOLOGÍA		35
2.1	Tipo y nivel de investigación.....	35
2.2	Técnicas e instrumentos.....	35
2.2.1	Técnica	35
2.2.2	Instrumento	35
2.2.3	Cuadro de coherencia	36
2.3	Campo de verificación.....	36
2.3.1	Ubicación espacial.....	36
2.3.2	Ubicación temporal.....	36
2.4	Unidades de estudio.....	37
2.5	Materiales, equipos y reactivos	37
2.5.1	Materiales	37
2.5.2	Equipos	38
2.5.3	Reactivos	38
2.6	Estrategia de recolección de datos.....	39

2.6.1	Toma de muestra	39
2.6.2	Análisis Molecular.....	39
2.6.3	Análisis Bioinformático.....	43
2.7	Organización.....	46
2.8	Recursos	46
2.9	Análisis estadístico	46
CAPÍTULO III RESULTADOS Y DISCUSIÓN		47
CONCLUSIONES.....		92
RECOMENDACIONES		93
PROPUESTA		94
REFERENCIAS BIBLIOGRÁFICAS		95



ÍNDICE DE TABLAS

Tabla 1. Comparación entre diferentes plataformas de secuenciación	15
Tabla 2. Características epidemiológicas de pacientes con pterigion primario.....	47
Tabla 3. Identificación de muestras y resultados de control de calidad	47
Tabla 4. Estadísticas generales MultiQC	48
Tabla 5. RNAseq: Conteo normalizado de los 50 genes más expresados en muestras 1,2,3 de pterigion (Parte 1 de 2)	57
Tabla 6. RNAseq: Conteo normalizado de los 50 genes más expresados en muestras 1.2.3 de pterigion (Parte 2 de 2)	58
Tabla 7. RNAseq: Conteo normalizado de los 50 genes más expresados en muestras 4,5,6,7 de pterigion (Parte 1 de 2)	61
Tabla 8. RNAseq: Conteo normalizado de los 50 genes más expresados en muestras 4,5,6,7 de pterigion (Parte 2 de 2)	62
Tabla 9. Análisis funcional de los genes mapeados	65
Tabla 10. Nuevos genes asociados a pterigion.....	66
Tabla 11. Ubicación de las mutaciones puntuales del gen P 53, comunes en los 7 casos de pterigion.....	67
Tabla 12. Genes componentes de las traslocaciones tipo fusión encontradas en 4 casos de las 7 muestras de pacientes con pterigion (Parte 1 de 2).....	73
Tabla 13. Genes componentes de las traslocaciones tipo fusión encontradas en 4 casos de las 7 muestras de pacientes con pterigion (Parte 2 de 2).....	74

ÍNDICE DE FIGURAS

Figura 1. <i>Secuenciación en plataformas Illumina.</i>	17
Figura 2. <i>Esquema básico de trabajo en estudios de secuenciación masiva.</i>	19
Figura 3. <i>Representación esquemática del procesamiento bioinformático típico de datos de secuencias de alto rendimiento para experimentos de RNA-seq ⁽⁷⁷⁾.</i>	31
Figura 4. <i>Funciones de conteo asignadas en porcentaje</i>	48
Figura 5. <i>STAR – Puntuaciones de alineación</i>	49
Figura 6. <i>Cutadapt: Longitudes de secuencias recortadas</i>	49
Figura 7. <i>FastQC. Histogramas de calidad de secuencia</i>	50
Figura 8. <i>FastQC. Puntuaciones de calidad por secuencia</i>	50
Figura 9. <i>FastQC. Por contenido de secuencia base-Muestra 1</i>	51
Figura 10. <i>FastQC. Por contenido de secuencia base-Muestra 2</i>	51
Figura 11. <i>FastQC. Por contenido de secuencia base-Muestra 3</i>	52
Figura 12. <i>FastQC. Por contenido de secuencia base-Muestra 4</i>	52
Figura 13. <i>FastQC. Por contenido de secuencia base-Muestra 5</i>	53
Figura 14. <i>FastQC. Por contenido de secuencia base-Muestra 6</i>	53
Figura 15. <i>FastQC. Contenido de ATCG por base secuenciada - Muestra 7</i>	54
Figura 16. <i>FastQC. Contenido de GC por secuencia</i>	54
Figura 17. <i>FastQC. Por contenido base N</i>	55
Figura 18. <i>FastQC. Distribución de longitud de secuencia</i>	55
Figura 19. <i>FastQC. Nivel de duplicación de secuencias</i>	56
Figura 20. <i>FastQC. Secuencias sobrerrepresentadas</i>	56
Figura 21. <i>10 Genes con mayor frecuencia de expresión en la muestra 1 de pterigión</i>	59
Figura 22. <i>10 Genes con mayor frecuencia de expresión en la muestra 2 de pterigión</i>	59
Figura 23. <i>10 Genes con mayor frecuencia de expresión en la muestra 3 de pterigión</i>	60

Figura 24. 10 Genes con mayor frecuencia de expresión en la muestra 4 de pterigion	63
Figura 25. 10 Genes con mayor frecuencia de expresión en la muestra 5 de pterigión	63
Figura 26. 10 Genes con mayor frecuencia de expresión en la muestra 6 de pterigion	64
Figura 27. 10 Genes con mayor frecuencia de expresión en la muestra 7 de pterigión	64
Figura 28. Mutaciones en gen TP 53-IG: Se observa el gen P53 ubicado en el brazo corto del cromosoma 17, región 13.....	67
Figura 29. Mutación en exón 4, ubicación 7 676 154 bp, brazo corto del cromosoma 17, región 13 (Gen P53).....	68
Figura 30. Mutación en exón 8, ubicación 7 669 372 bp, brazo corto del cromosoma 17, región 13 (Gen P53).....	69
Figura 31. Mutación en exón 8, ubicación 7 669 124 bp, brazo corto del cromosoma 17, región 13 (Gen P53).....	70
Figura 32. Mutación en exón 8, ubicación 7 668 836 bp, brazo corto del cromosoma 17, región 13 (Gen P53).....	71
Figura 33. Mutación en exón 8, ubicación 7 668 799 bp, brazo corto del cromosoma 17, región 13 (Gen P53).....	72
Figura 34. Gen de fusión CLDND1 – KIAA2026, en la muestra 1 de pterigión.....	75
Figura 35. Gen de Fusión RNU1-2 – ATP13A2	75
Figura 36. Gen de Fusión RNU1–2 – SDHB, en la muestra 3 de pterigión	76
Figura 37. Gen de Fusión WIZ – WHRN, en la muestra 3 de pterigión.....	76
Figura 38. Gen de fusión 19:40114634 - SNORD3A, en la muestra 6 de pterigión	77
Figura 39. Gen de fusión 6:164999470 - RNA5SP226, en la muestra 6 de pterigión	77
Figura 40. Gen de fusión PSMC3 - NXF1, en la muestra 6 de pterigión.....	78
Figura 41. Gen de fusión WNK1 - ZNF79, en la muestra 6 de pterigión.....	78
Figura 42. Gen de fusión DOCK8 - RF00100, en la muestra 6 de pterigión.....	79

Figura 43. *Gen de fusión 19:40114634 - SNORD3A, en la muestra 7 de pterigion* 79

Figura 44. *Gen de fusión AC093752.1 - RUFY3, en la muestra 7 de pterigion.....* 80

Figura 45. *Gen de fusión AC104151.1 - CNTNAP4, en la muestra 7 de pterigión.....* 80

Figura 46. *Gen de fusión AC104151.1 - CNTNAP4, en la muestra 7 de pterigion.....* 81



ÍNDICE DE ANEXOS

Anexo 1. <i>Consentimiento informado</i>	103
---	-----



LISTA DE ABREVIATURAS

ADN: Ácido desoxirribonucleico

ADNc: Ácido desoxirribonucleico complementario

ARN: Ácido ribonucleico

ARNm: Ácido ribonucleico mensajero

ARNnc: Ácido ribonucleico no codificante

BAM: Binary Alignment Map (mapa del alineamiento binario)

DAMP: Patrones moleculares asociados al daño

COX-2: Enzima ciclooxygenasa-2

HSP: Proteínas de choque térmico

ICAM 1: Molécula de adhesión intercelular -1

IAP: Inhibidores de la apoptosis

IHC: Inmunohistoquímica

IL: Interleuquinas

MMP: Metaloproteinasas

NGS: Next Generation Sequencing (Secuenciamiento de Próxima Generación)

MMC: Mitomicina C

PCNA: Antígeno nuclear de células proliferantes

PE: Paired End (final pareado)

RNA-seq: RNA sequencing

SAM: Sequence Alignment Map (Mapa del alineamiento de la secuencia)

SE: Single End (final singular)

TIMP: Inhibidores de las metaloproteinasas

TCGA: The Cancer Genome Atlas

MAPK: Mitogen-activated protein kinases

MMP: Prefijo Máximo Mapeable

PCBD2: Pterin-4 alfa-carbinolamina deshidratasa 2

LRMDA: Asociada a diferenciación de melanocitos ricos en leucina.

QC: Control de calidad

RIN: Número de integridad del RNA

RL: Radicales libres

ROS: Especies reactivas de oxígeno

STAR: Spliced Transcripts Alignment to a Reference

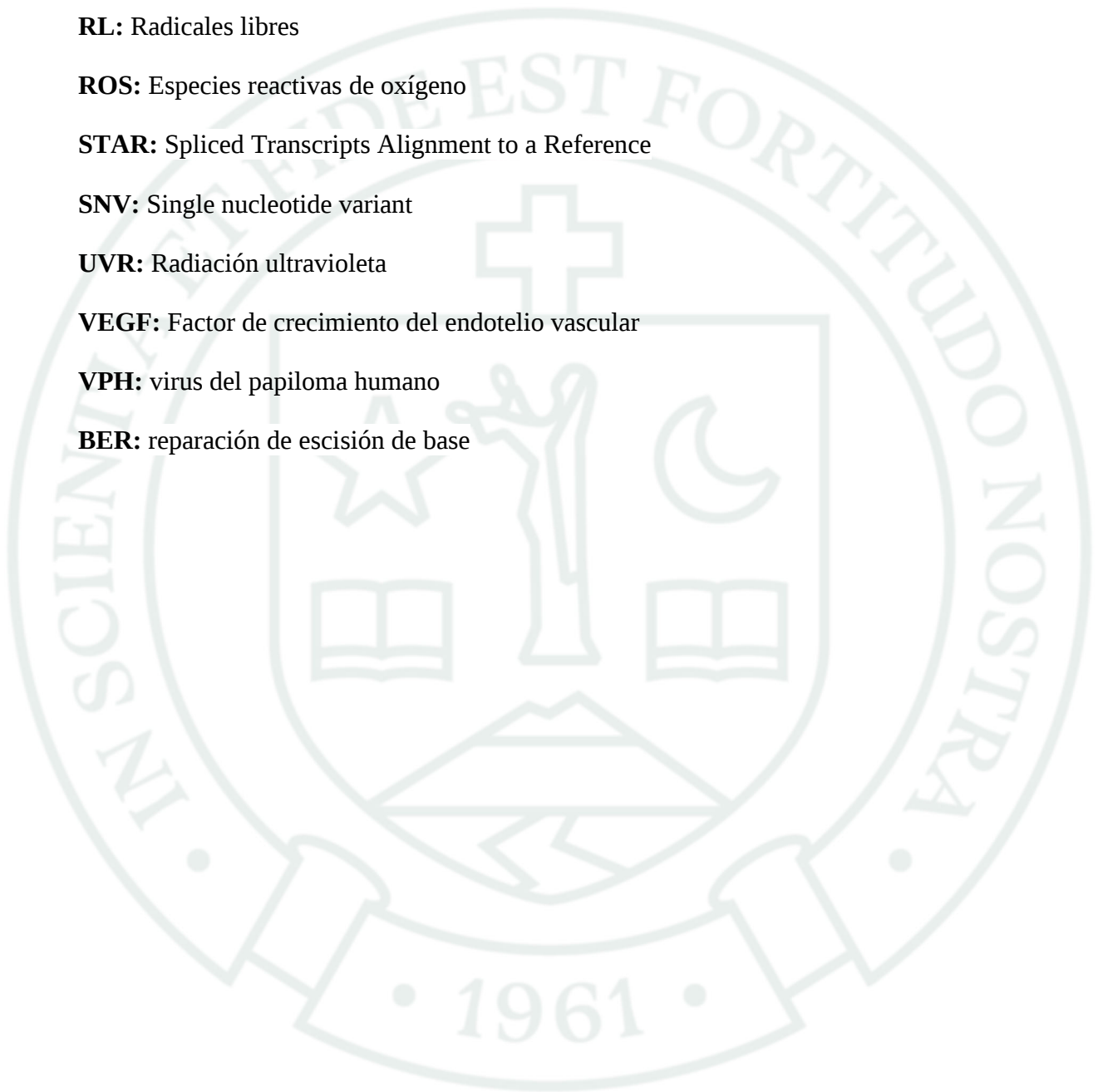
SNV: Single nucleotide variant

UVR: Radiación ultravioleta

VEGF: Factor de crecimiento del endotelio vascular

VPH: virus del papiloma humano

BER: reparación de escisión de base



INTRODUCCIÓN

El pterigión es una de las principales patologías oftalmológicas; sin embargo, los estudios existentes sobre su etiología y mecanismos moleculares son escasos. Su patogénesis involucra diversos procesos, entre ellos mecanismos antiapoptóticos, respuestas inmunológicas, acción de citocinas y factores de crecimiento, modulación de la matriz extracelular, alteraciones genéticas, infecciones virales y exposición a radiación ultravioleta ^(1,2,3,4,5).

Entre los factores genéticos, destaca el gen p53, un gen supresor tumoral que codifica la proteína p53, la cual desempeña un papel crucial en la regulación del ciclo celular, la apoptosis y la protección del genoma ⁽⁶⁾. Los resultados obtenidos en la presente investigación evidenciaron mutaciones en este gen, lo que sugiere que el pterigión podría originarse por una proliferación celular incontrolada, más que por un proceso puramente degenerativo.

El impacto del pterigión en la salud visual lo convierte en un problema de salud pública, ya que en estadios avanzados puede conducir a ceguera. Por esto, comprender sus mecanismos fisiopatológicos es esencial para prevenir su aparición y desarrollar tratamientos que inhiban su crecimiento y reduzcan la recurrencia postquirúrgica.

En América Latina, no se han realizado estudios que evalúen la expresión génica y las mutaciones del gen p53 en pterigión mediante la técnica de *Next Generation Sequencing* (NGS). La aplicación de esta metodología en el presente trabajo permitió obtener datos de alta resolución, fundamentales para la identificación y caracterización de mutaciones.

El uso de herramientas genómicas como NGS se encuentra en constante expansión, tanto en el ámbito académico como en el asistencial, impulsado por el desarrollo de la medicina de precisión. Este paradigma propone un abordaje individualizado, orientado a optimizar el diagnóstico y el tratamiento en cada paciente.

La relevancia de esta investigación radica en que sus hallazgos ofrecen nuevas dianas terapéuticas, con el potencial de revolucionar las estrategias de diagnóstico, estadiaje y tratamiento de enfermedades oculares ⁽⁷⁾.

HIPOTÉSIS DE LA INVESTIGACIÓN

La exposición a radiación ultravioleta y otras noxas ambientales puede inducir mutaciones en el gen humano p53 presente en la córnea y la conjuntiva ocular. Por lo tanto, es probable que en los pterigion primarios se detecte la presencia de este gen junto con mutaciones asociadas.



OBJETIVOS

Objetivo general

- Análisis molecular del gen humano p53 en pterigión primario.

Objetivos específicos

- Identificar mutaciones del gen humano p53 en muestras de pterigión primario.
- Analizar y describir mutaciones encontradas en el gen humano p53.
- Identificar el perfil de expresión génica en pterigión primario.
- Establecer las características clínicas - epidemiológicas asociadas a los pterigion primarios con presencia de mutaciones en el gen humano p53.
- Determinar la presencia de genes de fusión en las secuencias obtenidas por RNA Seq.

CAPÍTULO I

MARCO TEÓRICO Y CONCEPTUAL

1.1 Definición de pterigion

El pterigión es una de las patologías oftalmológicas más descritas debido a su carácter antiestético y progresivo, reconocido desde hace más de 3000 años. Etimológicamente, proviene del vocablo griego *pterygos*, que significa “ala de insecto” o “pequeña ala” (8,9,10,11).

Se localiza en la región interpalpebral y puede presentarse de forma unipolar o bipolar, siendo más frecuente en el lado nasal. También puede ser unilateral o bilateral. Su forma es generalmente triangular, con la base en la periferia y el ápex invadiendo la córnea. Se clasifica según la invasión corneal: grado 1, cuando el tejido fibrovascular alcanza el limbo; grado 2, cuando cubre aproximadamente 2 mm de la córnea; grado 3, cuando alcanza el margen pupilar; y grado 4, cuando lo sobrepasa. Asimismo, se clasifica por invasión limbar (L1–L4) y por su aspecto morfológico: atrófico, intermedio o carnosos (12,13,14).

En el estudio histopatológico, el pterigión presenta una masa de tejido subepitelial con degeneración elastótica. Las fibras colágenas degeneradas se tiñen de forma basofílica y son positivas para tinciones de tejido elástico, pero no sensibles a la digestión por elastasa. Estos haces fibrosos se disponen sobre un fondo de degeneración hialina, lo que indica daño actínico, con alteraciones en el colágeno, la sustancia propia y los fibrocitos, que forman un material elastótico anormal. En la córnea, la capa de Bowman puede observarse engrosada o destruida, y el epitelio que la recubre puede mostrar acantosis, disqueratosis y cambios displásicos. También es frecuente la neoformación de vasos sanguíneos junto a inflamación. En el estroma subyacente se observan proliferación fibroblástica, migración celular, infiltración inflamatoria, neovascularización y remodelación de la matriz extracelular (15,16,17,18,19,20).

La prevalencia del pterigión varía según factores geográficos y demográficos. Se han reportado cifras tan bajas como 0,07 % en Arabia Saudita y tan altas como 53 % en China (21). En un estudio realizado por Flores A. en Trujillo, Perú (2018), la prevalencia fue del 27,8 % en poblaciones a ≥ 3000 msnm y del 11,3 % en poblaciones a nivel del mar (22). La mayor incidencia se presenta en regiones cercanas al ecuador, con prevalencias

moderadas a altas en latitudes de hasta 35° por encima y por debajo de esta línea, lo que sugiere que la exposición prolongada a la radiación ultravioleta es un factor causal importante. No obstante, también se han implicado mecanismos inmunológicos, factores de crecimiento e influencia genética^(23,24,25,26,27,28).

La superficie ocular y la córnea, al proteger las estructuras internas del ojo, están expuestas a un alto estrés oxidativo ambiental. Los mecanismos mutagénicos, tanto de origen ambiental como endógeno, desempeñan un papel fundamental en las enfermedades oculares degenerativas^(29,30). Entre los factores de riesgo para el desarrollo del pterigión se encuentran el sexo masculino, la exposición al polvo, el calor, la inflamación o infección ocular, la residencia en áreas rurales, un bajo nivel educativo y la actividad al aire libre^(9,12,31). La predisposición hereditaria, con un patrón de herencia multifactorial, también es determinante para su aparición y persistencia⁽³²⁾.

El pterigión puede provocar astigmatismo y pérdida visual, siendo la cirugía el tratamiento definitivo. Sin embargo, la recurrencia postquirúrgica sigue siendo un desafío. En estadios avanzados, produce alteraciones en la refracción corneal esférica y cilíndrica, aumenta la irregularidad corneal y, en consecuencia, genera discapacidad visual. En estos casos, la escisión quirúrgica busca mejorar la topografía corneal y la función visual⁽¹²⁾.

Las opciones terapéuticas actuales incluyen el autoinjerto conjuntival, los colgajos conjuntivales deslizantes, el tratamiento con láser excímer y el trasplante de membrana amniótica^(7,8,33). No obstante, las complicaciones incluyen recurrencias que van del 15 % al 75 % tras una escisión simple, y alrededor del 6 % tras la cirugía con colgajo libre^(13,23). Como tratamiento complementario para prevenir la recurrencia postquirúrgica se han utilizado fármacos como la mitomicina C (MMC), el bevacizumab subconjuntival, la ciclosporina, la pirfenidona y el 5-fluorouracilo^(34,35).

1.2 Patogénesis

1.2.1 Radiación Ultravioleta

Los tejidos oculares cuentan con defensas antioxidantes efectivas para protegerlos frente a noxas asociadas con especies reactivas de oxígeno (ROS); sin embargo, múltiples factores pueden alterar esta homeostasis^(36,37,38,39).

El ozono, un gas compuesto por tres átomos de oxígeno, se encuentra principalmente en las capas superiores de la atmósfera y actúa como un potente oxidante. Su función

es esencial para la vida en la Tierra, ya que absorbe gran parte de la radiación ultravioleta (UVR). Aproximadamente el 90 % del ozono se localiza en la estratosfera (hasta 40 km de altura) y el resto en la troposfera. Estas capas bloquean casi todas las longitudes de onda inferiores a 290 nm y absorben entre el 70 % y el 90 % de la radiación UVB. El agotamiento de la capa de ozono ha incrementado la llegada de energía solar a la superficie terrestre.

La principal fuente natural de radiación ultravioleta es el sol. Los rayos UV afectan directamente la integridad del material genético celular y se clasifican en tres bandas según su longitud de onda: UVA (320–400 nm), UVB (290–320 nm) y UVC (<290 nm). La UVR genera efectos genotóxicos al inducir mutaciones y daños en el ADN mediante reacciones fotoquímicas cuya eficiencia depende de la longitud de onda y de la absorción directa por bases nitrogenadas, lípidos y proteínas. Asimismo, induce estrés oxidativo en las células irradiadas por la activación de fotosensibilizadores como riboflavina, triptófano y porfirina, que generan ROS capaces de activar procesos inflamatorios ⁽³⁹⁾.

En el ojo, la absorción de radiación UV depende de la longitud de onda. La córnea absorbe el 92 % de los rayos UV de 300 nm, mientras que longitudes de onda mayores pueden atravesarla y alcanzar el cristalino, donde más del 50 % de la radiación UVA de 360 nm es absorbida por el lente intraocular ^(40,41,42,43).

La exposición a radiación UV provoca la formación de dímeros de pirimidina — lesiones premutagénicas compuestas por timina o citosina— que alteran la estructura del ADN. Los dímeros de timina pueden repararse por fotorreactivación, pero, si permanecen sin reparar, resultan mutagénicos ⁽²⁶⁾. La radiación UV ejerce sus efectos tanto de forma directa, por fototoxicidad, como indirecta, a través de radicales libres (RL). Estos últimos producen daño oxidativo sobre proteínas, lípidos y ácidos nucleicos. Entre los RL más relevantes se incluyen el anión superóxido (O_2^-), el peróxido de hidrógeno (H_2O_2) y el peroxinitrito ($ONOO^-$), que son neutralizados por el sistema antioxidante enzimático compuesto por superóxido dismutasa (SOD), catalasa (CAT) y glutatión peroxidasa (GSH-Px), manteniendo un equilibrio homeostático ⁽³⁰⁾.

Entre los productos del daño fotooxidativo del ADN, la 8-hidroxideoxiguanosina (8-OHdG) es un biomarcador sensible y estable para evaluar el grado de daño genético

(29). En el pterigión, se han registrado niveles significativamente más altos de 8-OHdG en comparación con la conjuntiva adyacente normal, lo que respalda el papel del daño oxidativo del ADN en su desarrollo (37).

Otros marcadores moleculares también han sido asociados al pterigión. La ciclooxigenasa-2, implicada en la carcinogénesis cutánea, se sobreexpresa en el citoplasma de la capa epitelial —principalmente en la basal— del pterigión, en contraste con la ausencia de expresión en conjuntiva normal (44). La fosfolipasa D, involucrada en la inflamación y diferenciación celular, también se expresa en este tejido (3).

En cuanto a la respuesta inmune, se ha evidenciado un incremento de mastocitos en el pterigión frente a la conjuntiva normal. Estas células participan en la hipersensibilidad tipo I, liberando compuestos altamente tóxicos como la proteína básica mayor (PBM), la proteína catiónica eosinofílica (PCE), radicales libres de oxígeno, eicosanoides, citocinas Th2 y factores de crecimiento, que contribuyen al desprendimiento epitelial y al aumento de la permeabilidad vascular (17). Este mecanismo puede verse potenciado por irritantes exógenos como polvo o viento, que incrementan la producción de IgE en el pterigión.

Asimismo, la proteína endógena SERPINA3K, con funciones antiinflamatorias, antiangiogénicas y antioxidantes, inhibe proteasas de serina y suprime la proliferación y migración de células epiteliales en el pterigión. Por ello, se perfila como una posible diana terapéutica para prevenir su recurrencia postquirúrgica (19).

1.2.2 Infecciones Virales

Diversos estudios han evidenciado la presencia del virus herpes simple y del virus del papiloma humano (VPH) en muestras de pterigión. Estos virus codifican proteínas capaces de inactivar la p53, lo que genera inestabilidad cromosómica y aumenta la probabilidad de progresión del pterigión hacia variantes malignas.

1.2.3 Proteínas relacionadas a la apoptosis

El pterigion es un trastorno del crecimiento tisular de carácter similar a un tumor, asociado con una apoptosis defectuosa (29).

- **Survivina:** proteína codificada por el gen *BIRC5* en humanos, con funciones citoprotectoras y perteneciente a la familia de genes inhibidores de la apoptosis (IAP). Además, cumple un papel esencial en la división celular. Su sobreexpresión inhibe la apoptosis inducida por las vías extrínseca o intrínseca. El estrés oxidativo estimula significativamente la expresión de survivina, generando un estado hiperproliferativo (3,45). En el pterigión, su expresión podría contrarrestar la apoptosis inducida por radiación UV-B y, junto con la pérdida de p53, contribuir a la inhibición aberrante de la apoptosis, favoreciendo el desarrollo y potencial progresión neoplásica de la lesión (29,46).
- **Bcl-2:** proteína reguladora de la apoptosis, capaz de inducirla o inhibirla, codificada por el gen *Bcl-2* en humanos. Se ha observado su expresión en la capa basal epitelial de todas las células epiteliales del pterigión, mientras que en la conjuntiva normal no se detecta (9).
- **Complejo de rapamicina 1 (mTORC1):** regulador del crecimiento celular, síntesis de proteínas, autofagia y transcripción. Su señalización se encuentra altamente activada en el pterigión, provocando apoptosis aberrante y proliferación celular (30,9).
- **C-Myc:** oncoproteína que regula el crecimiento, la proliferación, la diferenciación y la apoptosis celular mediante la modulación de un amplio número de genes. Es uno de los cuatro factores de transcripción capaces de reprogramar células hacia un estado pluripotente y se sobreexpresa en la mayoría de los cánceres, lo que confirma su papel en la aceleración de la proliferación celular. En el pterigión, la sobreexpresión de C-Myc podría alterar el equilibrio entre apoptosis y proliferación, favoreciendo la génesis y progresión de la enfermedad (44).
- **K-Ras:** mutaciones en este gen se detectan en aproximadamente el 22 % de los tumores, desempeñando un papel clave en la progresión del cáncer colorrectal y en el 90 % de los adenocarcinomas pancreáticos. Las mutaciones más frecuentes se localizan en los codones 12 y 13, y en menor proporción en los codones 61 o 64. En el pterigión, se ha reportado una tasa de mutación del 10 % en el codón 12 en casos recurrentes y del 28 % en el codón 61 en casos primarios, sin mutaciones en la conjuntiva normal (47).
- **Piroptosis:** forma de muerte celular programada de carácter proinflamatorio, inducida por estímulos patológicos como infecciones microbianas, enfermedades cardiovasculares, trastornos del sistema nervioso central y tumores malignos. Marcadores como el receptor tipo Nod pirina-3 (NLRP3), caspasa-1, IL-18 e IL-1 β

son indicadores aceptados de piroptosis. Esta puede ser activada por patrones moleculares asociados a patógenos (PAMP) y a daño celular (DAMP). La radiación UV y el estrés oxidativo pueden generar DAMP y desencadenar este proceso en el pterigión. La evidencia sugiere que la piroptosis favorecería la transición de fibroblastos a miofibroblastos activados, lo que aceleraría la fibroproliferación y la acumulación de matriz extracelular, contribuyendo a la progresión de la lesión ⁽¹⁾.

1.2.4 Moléculas de adhesión celular

La molécula de adhesión intercelular-1 (ICAM-1) se expresa en el pterigión y está ausente en el epitelio de conjuntivas normales. Asimismo, se ha propuesto que la E-cadherina y la beta-catenina están presentes en el tejido del pterigión y podrían participar en la proliferación y adhesión epitelial ^(3,9).

1.2.5 Proteínas relacionadas con la proliferación celular

Diversas proteínas implicadas en la proliferación, como Ki-67, ciclina D1 y el antígeno nuclear de células proliferantes (PCNA), desempeñan un papel central en la regulación del ciclo celular. El PCNA es esencial para la síntesis de ADN y constituye un marcador útil de proliferación celular. Estudios han reportado niveles significativamente elevados de PCNA y Ki-67 en el pterigión, en comparación con la conjuntiva normal, confirmando su papel en la patogenia de esta lesión ⁽²⁵⁾. Por su parte, la ciclina D1, gen promotor de la progresión del ciclo celular, presenta una mayor expresión en pterigiones proliferativos que en los atróficos ⁽⁴⁸⁾.

1.2.6 Proteínas de shock térmico

Las proteínas de choque térmico (HSP) constituyen una familia de proteínas sintetizadas por las células en respuesta a condiciones de estrés, como la exposición a temperaturas extremas, radiación ultravioleta (UV) o durante procesos de cicatrización y remodelación tisular. Entre ellas, Hsp27, Hsp70, Hsp90 y el factor inducible por hipoxia 1 α (HIF-1 α) se han detectado en el tejido pterigial. Las variaciones en los niveles de HIF-1 α y HSP se asocian con mecanismos de supervivencia celular en entornos de estrés ⁽⁹⁾.

1.2.7 Proteínas de unión estrecha

Las claudinas, componentes clave de las uniones estrechas, desempeñan un papel esencial en el mantenimiento de la barrera funcional y estructural de los tejidos. Alteraciones en su expresión se han vinculado a diversas patologías, incluido el cáncer. En el pterigión se ha observado una disminución significativa de claudina-1 respecto a la conjuntiva normal, lo que sugiere su participación en la formación de esta lesión ⁽⁹⁾.

1.2.8 Proteínas de la matriz extracelular

La expresión aberrante de proteínas de la matriz extracelular se relaciona con el crecimiento proliferativo del pterigión. En particular, las queratinas K8, K16, K14 y AE3 se encuentran presentes en todo el espesor epitelial del pterigión, pero están ausentes en la conjuntiva ocular sana. El colágeno tipo II se detecta exclusivamente en el pterigión, mientras que los tipos I, III y IV aparecen tanto en pterigión como en conjuntiva normal ^(9,5).

El factor de crecimiento transformador beta 1 (TGF- β 1) inhibe la síntesis de enzimas proteolíticas de la matriz extracelular, evitando su degradación y favoreciendo la acumulación de colágeno y fibras de elastina, lo que conduce a proliferación anormal y desarrollo del pterigión. Por su parte, el factor de transcripción Sp1, implicado en procesos como el ciclo celular, crecimiento, proliferación, metabolismo y apoptosis, presenta niveles significativamente mayores en pterigiones primarios que en conjuntiva normal ⁽⁵⁾.

1.2.9 Metaloproteinasas (MMP)

Las metaloproteinasas (MMP) son endopeptidasas dependientes de zinc que forman una familia de más de 25 enzimas capaces de degradar componentes de la matriz extracelular. Participan en procesos fisiológicos y patológicos como la remodelación tisular, cicatrización, angiogénesis, invasión celular y diferenciación. En el pterigión, la sobreexpresión de MMP contribuye a la destrucción de la capa de Bowman y del estroma corneal.

Se han detectado MMP-1, MMP-2, MMP-3, TIMP-1 y TIMP-3 en mayores cantidades en pterigión que en conjuntiva normal ⁽³⁾. La expresión y actividad de MMP e

inhibidores tisulares (TIMP) varía según la etapa de la lesión: MMP-1 y MMP-3 presentan mayor concentración en la cabeza del pterigión que en el cuerpo o en fibroblastos subconjuntivales⁽⁴⁹⁾. El desequilibrio entre MMP y TIMP se considera un factor clave en la progresión y recurrencia del pterigión⁽³⁶⁾.

1.2.10 Interleuquinas

Las interleuquinas (IL) están estrechamente relacionadas con la respuesta inflamatoria en el pterigión. IL-1 α se encuentra altamente expresada tanto en pterigión primario como recurrente, mientras que IL-6, IL-8, IL-10 e IL-17 muestran una fuerte expresión en el epitelio pterigial comparado con córnea, conjuntiva y limbo normales. La exposición a radiación UVB incrementa significativamente la producción de IL-6 e IL-8⁽¹²⁾.

El estrés oxidativo genera especies reactivas de oxígeno que fosforilan lípidos de membrana, produciendo compuestos como la prostaglandina E-2 (PGE-2), detectada en el pterigión. Estos lípidos oxidados estimulan la síntesis de ciclooxigenasa-2 (COX-2) e IL-8, ambas moléculas proinflamatorias⁽³²⁾. Tras la exposición a UV-B, se observa un aumento en IL-1, MMP y otros mediadores inflamatorios⁽³⁾.

1.2.11 Factores de crecimiento

Los factores de crecimiento son moléculas capaces de estimular la proliferación, diferenciación y reparación celular⁽⁵⁰⁾. En la patogénesis del pterigión intervienen múltiples factores, entre ellos: factor de crecimiento endotelial vascular (VEGF), factor de crecimiento transformador beta (TGF- β), factor básico de crecimiento de fibroblastos (bFGF), factor de crecimiento similar a la insulina (IGF), factor de crecimiento nervioso (NGF), factor de crecimiento del tejido conectivo (CTGF), factor de crecimiento epidérmico de unión a heparina (HB-EGF) y factor de crecimiento derivado de plaquetas (PDGF).

Se ha demostrado que bFGF, VEGF y HB-EGF se regulan al alza tras la exposición a UV-B. Asimismo, el aumento de expresión de factores neurotróficos como CNTF, sustancia P, neurotrofina-4 y neurotrofina-5 también ha sido evidenciado en el tejido pterigial. La heterogeneidad en las mutaciones asociadas al pterigión podría explicar la variabilidad clínica de su crecimiento⁽³⁾.

En modelos murinos, la doxíciclin —antibiótico bacteriostático de bajo costo y uso seguro— inhibió de forma significativa la angiogénesis, efecto atribuido a su capacidad para suprimir la actividad catalítica de diversas MMP, independientemente de su acción antimicrobiana ⁽⁵¹⁾. El conocimiento de estos factores abre nuevas posibilidades para el desarrollo de estrategias que prevengan la aparición y progresión del pterigión, con el potencial de reducir la necesidad de intervenciones quirúrgicas ⁽⁵²⁾.

1.2.12 Genes supresores de tumores - Gen p53

El **gen p53** se localiza en el brazo corto del cromosoma 17 y cumple principalmente la función de gen supresor tumoral. Participa en la regulación del ciclo celular, la reparación y síntesis de ADN, la diferenciación celular y la apoptosis⁴³. Su acción se basa en detener el ciclo celular cuando se detecta daño genómico, permitiendo su reparación; si el daño es irreparable, induce apoptosis. Las mutaciones en el gen p53 inactivan o alteran estas funciones, constituyéndose en marcadores genéticos del crecimiento neoplásico en diversos tipos de tumores ^(6,53,54).

La proteína p53 normalmente se expresa en niveles bajos; sin embargo, frente a estímulos tumorigénicos como radiación ultravioleta (UV), radiación gamma, exposición a determinados compuestos químicos o sobreexpresión de oncogenes, se incrementa su concentración nuclear. Esto promueve la transcripción de genes implicados en respuestas protectoras, como la detención del ciclo celular, la reparación del ADN, la apoptosis y la disminución de la angiogénesis. Debido a este papel, p53 ha sido denominado “el guardián del genoma” ⁽⁴⁵⁾.

En aproximadamente la mitad de los tumores humanos, el gen TP53 presenta mutaciones sin sentido en su región central (exones 5–8), la cual codifica el dominio responsable de la unión al ADN. Esta unión es indispensable para la función transcripcional de p53. Dichas mutaciones originan una proteína aberrante con efecto dominante negativo, capaz de inhibir la acción de la proteína de tipo salvaje.

El aumento en los niveles de p53 se debe, en gran medida, a una mayor estabilidad proteica, regulada parcialmente por proteasas celulares y por la interacción diferencial con la oncoproteína MDM2 (murine double minute 2). La prevalencia de la expresión anormal de p53 varía según el tipo de tumor, oscilando entre 0 % y 60 % (56 % en carcinoma de pulmón, 50 % en carcinoma de colon, 44 % en tumores cutáneos, entre

otros) ^(55,56,57,58). También se ha documentado en tumores benignos, como el tumor filoides en biopsias mamarias ⁽⁵⁹⁾.

La radiación UV produce fosforilación de la serina 15 de p53, catalizada por la quinasa ATR (*ataxia telangiectasia and Rad3-related*). Una falla en este mecanismo podría impedir su activación en el pterigión. En particular, la radiación UVB induce mutaciones características en p53 (transiciones C→T y CC→TT), las cuales no suelen originarse por otros carcinógenos ⁽⁶⁰⁾. Asimismo, la radiación UV provoca acetilación de p53 en múltiples residuos de lisina, un proceso que regula su unión al ADN; defectos en esta acetilación pueden ocasionar fallos de activación.

La expresión aberrante de p53 favorece la proliferación celular y reduce la apoptosis, acelerando el desarrollo del pterigión ⁽⁵⁹⁾. Además de p53, otros genes supresores de tumores, como p63, p16 y p27, participan en la patogénesis de esta lesión. El gen p63 muestra una mayor expresión en la capa basal y parabasal del epitelio en pterigión primario, y en todo el espesor epitelial en pterigión recidivado.

Diversos estudios han reportado entre 7,9 % y 100 % de tinción positiva para p53 mediante inmunohistoquímica en el epitelio de pterigiones primarios y recurrentes ⁽⁴⁶⁾. Se ha propuesto que las mutaciones de p53 en células pterigiales constituyen un evento temprano en su desarrollo, probablemente inducido por la radiación UV, reconocida como factor de riesgo para múltiples tumores malignos, como el carcinoma de células escamosas, debido a su impacto sobre el material genético.

El esclarecimiento del mecanismo de disfunción de p53 en el pterigión reviste relevancia no solo para la oftalmología, sino también para la oncología, ya que puede aportar información valiosa sobre la inactivación de este gen supreso ^(61,62,63,64,65,66,67).

1.3 Técnicas de secuenciación masiva (NGS)

1.3.1 Introducción

Desde el descubrimiento del ADN hasta la actualidad, el estudio y conocimiento de esta molécula, así como de la genómica aplicada, han experimentado notables avances ⁽⁶⁸⁾.

La secuenciación de los nucleótidos que componen la molécula de ADN constituye el análisis más detallado de su estructura y una herramienta eficaz para identificar

variantes en el material genético. La secuenciación Sanger, desarrollada en la década de 1970, supuso una revolución en la genética humana y marcó el inicio de la era genómica.

En los últimos años, se han desarrollado nuevas plataformas de secuenciación denominadas de alto rendimiento o de nueva generación (*Next-Generation Sequencing*, NGS), capaces de generar de manera masiva y paralela millones de fragmentos de ADN en un único proceso. Este avance ha permitido aumentar significativamente el rendimiento a un coste menor, ofreciendo ventajas sustanciales respecto a los métodos convencionales ⁽⁶⁹⁾.

El uso de esta tecnología se encuentra en constante expansión en ámbitos académicos y clínicos, impulsado por el auge de la medicina de precisión, un nuevo paradigma de atención médica que busca maximizar la eficacia diagnóstica y terapéutica en función de las características genéticas de cada paciente ⁽⁶⁸⁾.

1.3.2 Técnicas de secuenciación

Entre las plataformas de NGS más utilizadas se encuentran el 454 GS Flex de Roche, Genome Analyzer o HiSeq de Illumina y SOLiD de Life Technologies (70). En años recientes han surgido versiones con menor rendimiento y capacidad de secuenciación, pero con mayor facilidad de uso y dirigidas a segmentos de mercado específicos. Algunos ejemplos son el 454 GS Junior de Roche, el MiSeq Personal Sequencer de Illumina y el Ion Torrent de Life Technologies (Tabla 1) ⁽⁶⁹⁾.

Tabla 1. Comparación entre diferentes plataformas de secuenciación

Plataforma	Tiempo carrera ^a	Reads/carrera (en millones)	Bases/read ^b	Rendimiento (Mb/carrera)
3730xl(capilares, no NGS)	2 h	0,000096	650	0.06
Ion Torrent (chip 314)	2 h	0,10	100	> 10
454 GS Jr. Titanium	10 h	0,10	400	50
Starlight ^g	?	~0,01	> 1.000	?
PacBio RS	0,5 – 2 h	0,01	860 – 1.100	5 – 10
454 FLX Titanium	10 h	1	400	500
454 FLX + ^c	18 – 20 h	1	700	900
Ion Torrent (chip 316)	2 h	1	> 100	> 100
Helicos ^d	N/A	800	35	28.000
Ion Torrent (chip 318)	2 h	4-8	> 100	> 1.000
Illumina MiSeq ^e	26 h	3,4	150 + 150	1.020
Illumina iScanSQ	8 días	250	100 + 100	50.000
Illumina GAllx	14 días	320	150 + 150	96.000
SOLiD - 4	12 días	>840 ^e	50 + 35	71.400
Illumina HiSeq 1000	8 días	500	100 + 100	100.000
Illumina HiSeq 2000	8 días	1000	100 + 100	200.000
SOLiD – 5500 (PI) ^g	8 días	> 700 ^e	75 + 35	77.000
SOLiD – 5500xl (4hq) ^g	8 días	> 1.410 ^e	75 + 35	155.100
Illumina HiSeq 2000 – v3 ^f	10 días	≤ 3.000	100 + 100	≤ 600.000

h; horas; Mb: megabases

~Valor probablemente derivado de información no publicada indisponible en mayo de 2011.

Adaptado de Glenn, 2011⁵⁹

^a Tiempo necesario en el instrumento para conseguir la longitud máxima de read.

^b Longitud promedio para los reads de alta calidad.

^c Actualización del instrumento FLX, verano 2011.

^d Instrumentos y reactivos no se pueden comprar ya, solo ofrecen servicios.

^e Reads alineables (número crudo de reads de alta calidad).

^f Reactivos y software TruSeq v3 anunciados, reads y rendimiento son la mitad que el HiSeq1000.

^g Información basada solamente en datos de la compañía (datos independiente todavía no disponibles).

Nota. Rodríguez y Armengol ⁽⁶⁹⁾.

1.3.3 Preparación de bibliotecas

El objetivo de esta etapa es obtener fragmentos cortos de ADN, cuyo tamaño varía según la plataforma utilizada, aunque generalmente oscila entre 150 y 400 pares de bases. Dependiendo de la estrategia, la preparación de las bibliotecas puede realizarse antes o después del proceso de enriquecimiento.

La preparación consiste en añadir adaptadores a ambos extremos de cada uno de los millones de fragmentos de ADN que serán secuenciados (denominados insertos). En esta fase también es posible incorporar etiquetas —secuencias específicas de ADN denominadas *sample barcodes*— que permiten diferenciar cada muestra, posibilitando así la combinación de múltiples muestras en un mismo ensayo sin riesgo de entrecruzamiento ⁽⁶⁸⁾.

Los adaptadores son específicos para cada plataforma de secuenciación y corresponden a fragmentos de ADN de doble cadena con secuencias consenso, en las que se alinean los cebadores para las etapas de amplificación del templado (ya sea por PCR en puente o PCR en emulsión, según la plataforma) y los cebadores para la secuenciación propiamente dicha ⁽⁶⁸⁾. La obtención de fragmentos flanqueados por adaptadores con secuencias específicas hace posible realizar millones de reacciones de manera simultánea, obteniendo la secuencia de cada inserto.

1.3.4 Enriquecimiento del blanco (*target enrichment*)

En los protocolos de secuenciación dirigida, es necesario enriquecer las regiones de interés y separar el resto del ADN genómico. Para ello, se emplean principalmente dos enfoques: enriquecimiento por captura y enriquecimiento por PCR.

En el enriquecimiento por captura, el ADN se fragmenta para obtener fragmentos cortos; posteriormente, se preparan bibliotecas con dichos fragmentos y se hibridan con sondas de ADN o ARN complementarias a las regiones objetivo. Inicialmente, estas sondas se inmovilizaban en micromatrices (*microarrays*) en fase sólida; sin embargo, actualmente se emplean con mayor frecuencia sondas en solución líquida ⁽⁶⁸⁾.

En cuanto al enriquecimiento basado en PCR multiplex, se utilizan múltiples cebadores específicos en una misma reacción para amplificar regiones objetivo y generar amplicones de tamaño reducido. Adicionalmente, se han desarrollado técnicas basadas en microfluídica asociada a PCR en emulsión, que permiten realizar millones de reacciones independientes y en paralelo, a diferencia de la PCR multiplex convencional ⁽⁶⁸⁾.

Un tercer enfoque corresponde a la captura por circularización, en el que destacan las sondas moleculares invertidas. Esta técnica combina una fase de hibridación con una posterior amplificación por PCR, dirigida al enriquecimiento de las regiones de interés.

1.3.5 Amplificación del templado y secuenciación masiva paralela

Tanto las plataformas Illumina como Ion Torrent utilizan las bibliotecas como templado para la secuenciación, requiriendo su amplificación para generar la cantidad de señal necesaria según la sensibilidad del método de detección.

En plataformas Illumina (Figura 1), las bibliotecas son cargadas en una celda de flujo (68,69).

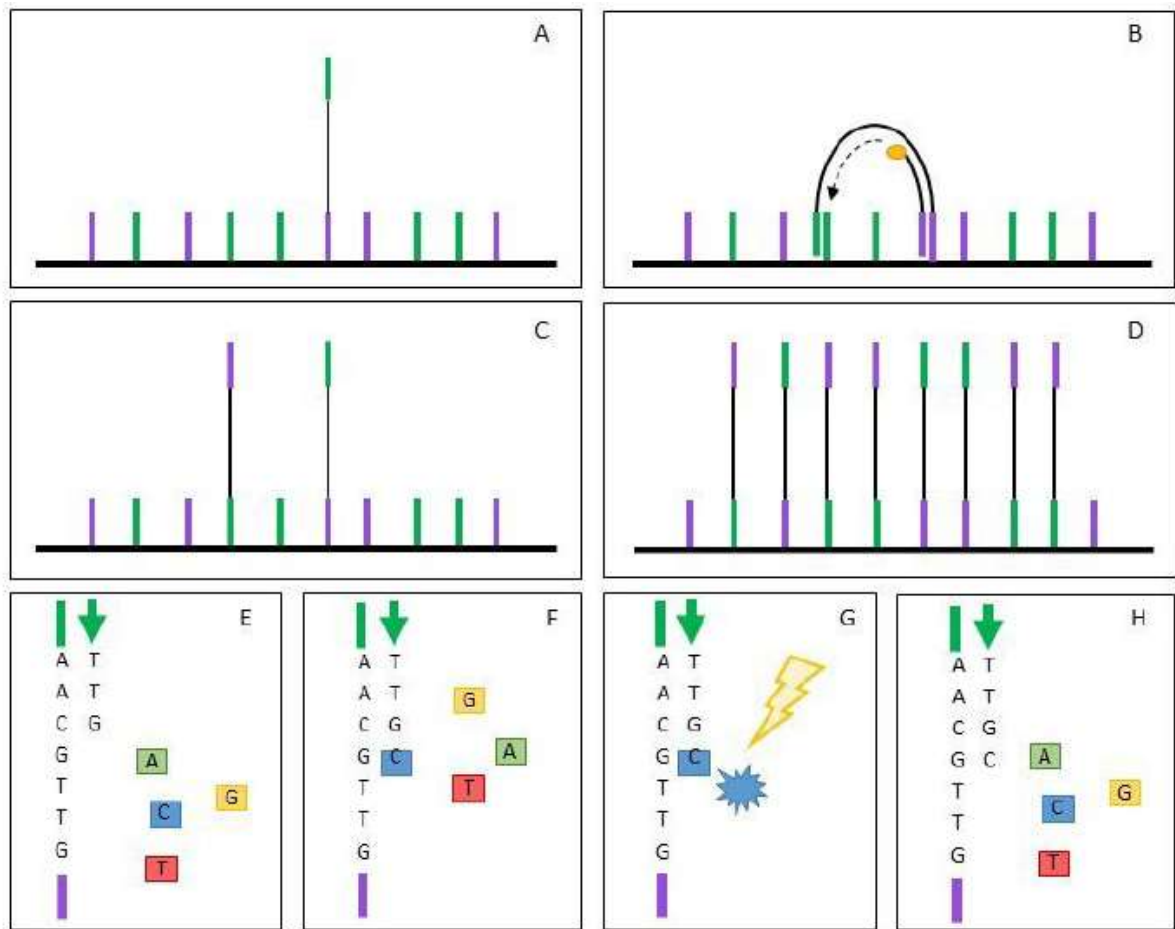


Figura 1. Secuenciación en plataformas Illumina.

Nota: Juk F ⁽⁶⁸⁾

Luego de la preparación de la biblioteca, se lleva a cabo la amplificación de grupos (clusters) en una celda de flujo (A a D). En este proceso, la biblioteca es diluida y cargada en la celda de flujo, de modo que cada hebra de ADN se une a ella mediante la hibridación de los adaptadores (verde y violeta) con oligonucleótidos complementarios fijados al soporte sólido. A partir de estos oligonucleótidos anclados, se genera una copia del inserto mediante una reacción de polimerización (A). Posteriormente, se inicia la PCR-puente: el fragmento de ADN se curva, y el adaptador del extremo libre se une a un oligonucleótido complementario igualmente anclado al soporte sólido. Una segunda reacción de polimerización (B) produce una hebra complementaria a la inicial (C). Este proceso se repite hasta formar un grupo (cluster) derivado de la hebra primaria, lo que facilita su posterior detección por fluorescencia.

La secuenciación por síntesis (E a H) comienza con cebadores específicos para los adaptadores (flecha verde). En cada ciclo, nucleótidos marcados con fluorescencia —terminadores reversibles— compiten por unirse a la cadena de ADN en elongación (E). Solo el nucleótido complementario a la hebra molde se incorpora (F), mientras que el resto es eliminado. Una fuente lumínica excita el fluoróforo del nucleótido incorporado (G) y la señal emitida es capturada por un dispositivo CCD para su análisis. El ciclo se repite con un nuevo flujo de nucleótidos para continuar la síntesis (H).

En términos generales, cada inserto de la biblioteca es amplificado clonalmente mediante PCR-puente, produciendo un cluster unido al soporte sólido que servirá como molde. La secuenciación por síntesis con terminadores reversibles asegura que solo se añada un nucleótido por ciclo. Tras su incorporación, el nucleótido es excitado, su señal óptica registrada y, posteriormente, el fluoróforo y el terminador son eliminados antes de iniciar un nuevo ciclo. Las imágenes obtenidas de cada cluster son procesadas para el llamado de bases (*base calling*), obteniendo así la secuencia de cada lectura.

El procedimiento se realiza en ambos sentidos, a partir de cada adaptador, lo que permite obtener lecturas de extremos pareados (*paired-end reads*). Tanto Illumina como Ion Torrent disponen de secuenciadores de distinta escala y capacidad, adaptados a diferentes volúmenes de procesamiento. Cada equipo es compatible con distintos tipos de celdas de flujo o chips semiconductores, lo que permite seleccionar la estrategia más adecuada según los objetivos del análisis ⁽⁶⁸⁾.

1.3.6 Procesamiento de datos

El procesamiento de datos comienza con la señal cruda obtenida por el secuenciador y finaliza con la interpretación de las variantes, de acuerdo con el objetivo del ensayo.

En el análisis de variantes puntuales (SNVs), inserciones o deleciones (indels) a nivel de ADN, este procedimiento se estructura en una serie de etapas concatenadas, conocidas como *tubería* o *pipeline* bioinformático.

El análisis primario consiste en convertir la señal cruda generada por el secuenciador en millones de secuencias cortas de ADN, denominadas lecturas (*reads*). Estas lecturas se convierten en la unidad fundamental de análisis bioinformático, correspondiendo cada una a un inserto de la biblioteca. El resultado de esta etapa se almacena, por lo

general, en un archivo FASTQ, el cual contiene tanto la secuencia de cada lectura obtenida como un valor de calidad asociado a cada nucleótido (*Phred score*) ^(68,69). (Figura 2).

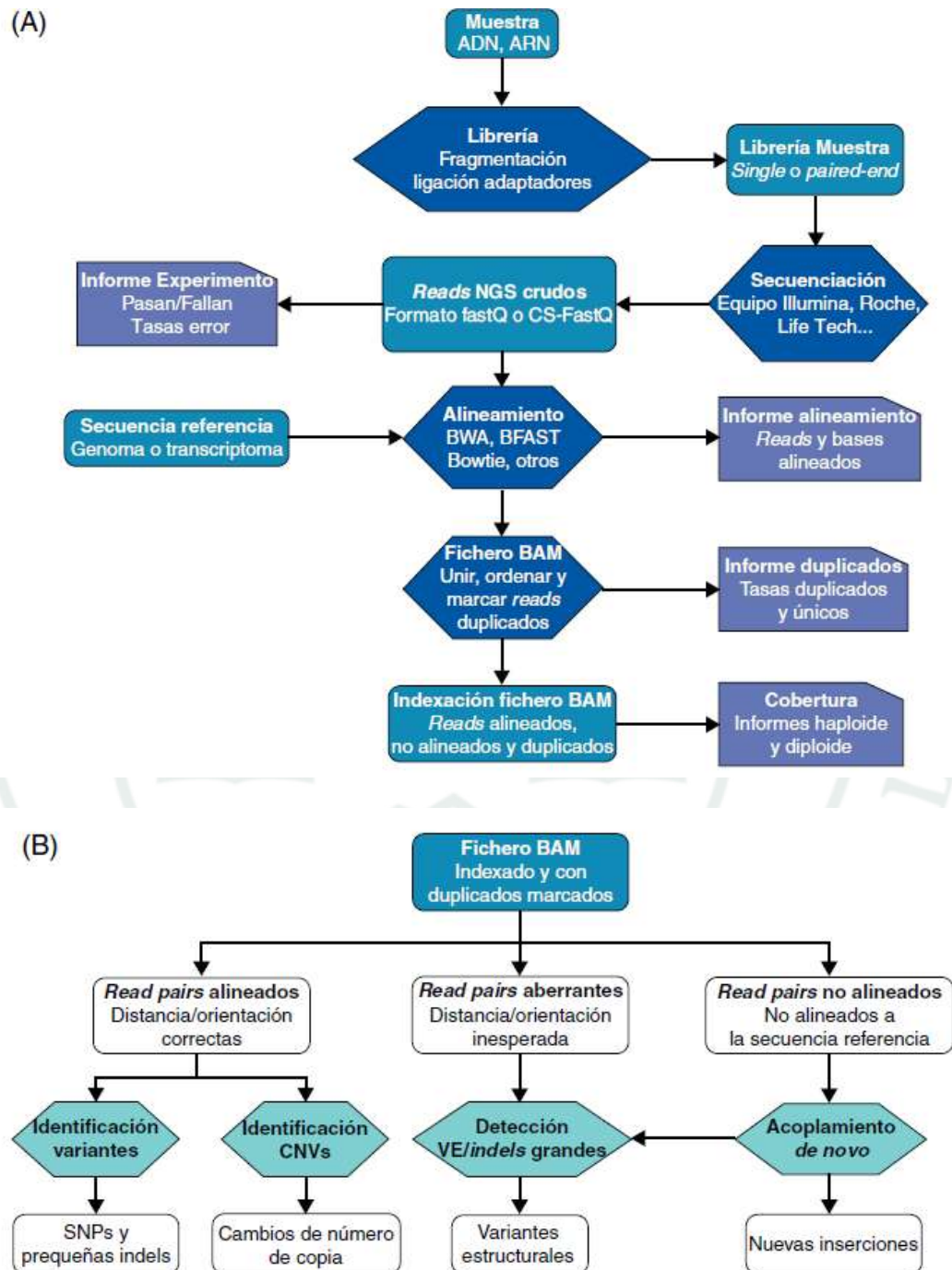


Figura 2. Esquema básico de trabajo en estudios de secuenciación masiva.

Nota: Rodríguez B, Armengol L ⁽⁶⁹⁾

Secuenciación y alineamiento

Las bibliotecas generadas con fragmentos de ADN o ARN se secuencian de forma masiva y paralela. Tras la captura de imágenes y su conversión en secuencias de nucleótidos, las lecturas (*reads*) resultantes se alinean contra una secuencia de referencia. El resultado se exporta en un archivo SAM (*Sequence Alignment/Map*) o en su versión comprimida BAM (*Binary Alignment/Map*), marcando las lecturas duplicadas.

En el análisis del archivo BAM, se indica el estado y la calidad de alineamiento de cada *read*.

- Lecturas correctamente alineadas: se utilizan para detectar SNP (single nucleotide polymorphisms), pequeñas inserciones y deleciones (indels) y para estimar el número de copias (copy number variants, CNV).
- Lecturas alineadas de forma aberrante: aquellas cuyo par presenta una distancia u orientación inesperada, son evaluadas como posibles indicadores de variantes estructurales (VE).
- Lecturas no alineadas: se ensamblan de novo para predecir variantes estructurales y nuevas inserciones.

Nota. Rodríguez y Armengol ⁽⁶⁹⁾

Análisis secundario

Incluye control de calidad, recorte o eliminación de lecturas de baja calidad, mapeo contra la referencia, alineamiento local y llamado de variantes. Los archivos SAM/BAM pueden visualizarse en programas como IGV (*Integrative Genomics Viewer*), lo que permite inspeccionar secuencias, mapeo y alineamiento ⁽⁶⁸⁾.

El llamado de variantes genera un archivo VCF (*Variant Call Format*), que incluye: genoma de referencia, coordenadas cromosómicas de la variante, tipo de cambio de secuencia y valores de calidad del llamado de variantes. En este archivo son consideradas todas las variantes que muestran alguna diferencia con el genoma de referencia, sin que esto signifique estrictamente patogenicidad ⁽⁶⁸⁾.

Análisis terciario

Se realiza sobre las variantes identificadas e implica su anotación y filtrado según el objetivo del estudio. La anotación utiliza bases de datos que aportan: información

genómica (gen, exón, variante en secuencia codificante y en la proteína, etc.), información funcional, frecuencia poblacional de la variante, predicciones bioinformáticas de patogenicidad, información fenotípica o clínica vinculada a la variante ⁽⁶⁸⁾.

Cobertura y profundidad

Cobertura (*coverage breadth*): proporción del genoma representada por las lecturas alineadas contra la referencia.

Profundidad (*coverage depth*): número de lecturas mapeadas que representan un mismo locus, indicando cuántas veces se secuenció esa región ^(68,69).

Procesamiento en estudios de ARN

En el procesamiento bioinformático de las lecturas, es posible detectar la presencia de:

Genes de fusión: también denominados genes quiméricos o híbridos, corresponden a la yuxtaposición de dos genes localizados en un mismo cromosoma o en cromosomas distintos. Se originan como consecuencia de variantes estructurales, definidas típicamente como grandes alteraciones en la organización del genoma humano ⁽⁷⁰⁾. Inicialmente identificados en la patogenia de neoplasias malignas hemato-linfoides, actualmente se reconocen como eventos impulsores clave en la génesis de múltiples neoplasias benignas y malignas, incluidos tumores de tejidos blandos y epiteliales ⁽⁷¹⁾.

Los genes de fusión recurrentes son relativamente infrecuentes en comparación con otras mutaciones somáticas, probablemente debido a la complejidad de sus mecanismos de generación. Los eventos de fusión patológica suelen ser altamente penetrantes y conllevan transformaciones celulares significativas ⁽⁷²⁾. En oftalmología, se han descrito genes de fusión en tumores de glándulas lagrimales, tumores primarios de tejidos blandos orbitarios y anexos oculares ⁽⁷¹⁾.

No todas las fusiones génicas están asociadas a procesos oncogénicos, ya que pueden hallarse en gran número incluso en muestras no tumorales. Sin embargo, desempeñan un papel central en la investigación del desarrollo del cáncer. Determinar la probabilidad de que las transcripciones derivadas de genes de fusión

desencadenen tumores constituye un reto científico aún no completamente resuelto (73).

• Mecanismos que forman genes de fusión

Generalmente, los genes de fusión surgen por alteraciones estructurales del genoma que implican disrupción del ADN, seguida de reintegración y replicación errónea (73). El reordenamiento genómico puede producirse dentro de un mismo gen o entre dos genes independientes, e involucra cuatro mecanismos principales: translocaciones, inversiones, duplicaciones en tándem y deleciones. Mientras que las inversiones, duplicaciones en tándem y deleciones suelen ocurrir dentro de uno o dos genes adyacentes, las translocaciones representan anomalías de mayor escala, derivadas de interacciones entre regiones distantes del genoma o dentro del mismo cromosoma, es decir, eventos inter o intracromosómicos (72).

Como consecuencia, se generan fusiones génicas que producen transcripciones interrumpidas o fusionadas, y proteínas anómalas o quiméricas con el potencial de activar, inhibir o eliminar funciones originales de las proteínas (74). Un ejemplo emblemático es la fusión BCR-ABL1, producto de una translocación intercromosómica característica —conocida como *cromosoma Filadelfia*— que une la porción 5' del gen BCR (en el cromosoma 22) con la porción 3' del gen ABL (en el cromosoma 9). Esta transcripción origina una tirosina quinasa anormal que impulsa el desarrollo de leucemia mieloide crónica (LMC).

El tratamiento dirigido con imatinib (Gleevec/Glivec) ha demostrado gran eficacia frente a la LMC, lo que incentivó la búsqueda de otros genes de fusión como biomarcadores tumorales específicos o dianas terapéuticas. El *cromosoma Filadelfia* fue el primer gen de fusión descrito (1973) en LMC; se encuentra en más del 95% de los pacientes y ejerce su efecto oncogénico mediante la codificación de una tirosina quinasa constitutivamente activa (ABL1) (74).

En tumores sólidos, el primer gen de fusión identificado fue **CTNNB1-PLAG1** en adenomas de glándulas salivales, generalmente benignos. Posteriormente, se descubrieron fusiones génicas en glioblastomas, melanomas y diversos cánceres (próstata, mama, ovario, pulmón, colorrectal, cabeza y cuello) (75). Entre 1982 y 1988 se identificaron 10 genes de fusión, cifra que aumentó a 162 en la década de 1990-1999. En enero de 2021, la *Base de datos Mitelman de aberraciones*

cromosómicas y genes de fusión en cáncer registró 32 618 genes de fusión, número que continuará en aumento gracias a las técnicas de secuenciación de alto rendimiento aplicadas a más muestras tumorales ⁽⁷⁵⁾.

1.3.7 Plataformas de secuenciación de nueva generación

Cada plataforma de secuenciación masiva se basa en principios químicos distintos, lo que genera diferencias tanto cuantitativas como cualitativas:

Secuenciación por síntesis o polimerización: sistemas dependientes de ADN polimerasa.

Secuenciador semiconductor: registra los cambios de pH producidos durante la incorporación de bases en la síntesis de ADN. Esta tecnología es utilizada por las plataformas de Ion Torrent™ (Ion PGM, Ion PROTON) (<https://www.thermofisher.com/es/es/home/brands/ion-torrent.html>) ⁽⁷⁴⁾.

Terminación reversible cíclica: utiliza nucleótidos con un grupo terminador marcado con moléculas fluorescentes, impidiendo la incorporación del siguiente nucleótido hasta que el terminador es retirado tras la lectura de la señal. Esta técnica es empleada por plataformas como Illumina (<https://www.illumina.com/>) y Helicos BioSciences (<http://seqll.com/>) ⁽⁷⁴⁾.

- **Secuenciación en Illumina**

La tecnología Illumina se caracteriza por dos procesos clave:

- a) La amplificación de fragmentos de ADN para la generación de clústeres (colonias clonales del mismo fragmento) mediante PCR en puente.
- b) La detección de bases a través de etiquetas fluorescentes ⁽⁷⁶⁾.

- **Generación del clúster**

Este proceso se realiza mediante amplificación en puente. Los fragmentos de ADN se disponen sobre una superficie sólida de vidrio dividida en carriles, recubiertos con oligonucleótidos complementarios a los adaptadores de cada fragmento. Esto permite el anclaje de los fragmentos a la celda de flujo. Una vez fijados, la polimerasa sintetiza una hebra complementaria, eliminándose posteriormente la hebra original. La hebra recién sintetizada se pliega y se une a un oligonucleótido complementario formando un “puente”. La polimerasa sintetiza entonces una copia

idéntica a la hebra original, generando dos copias clonales. Este ciclo se repite masivamente hasta producir millones de copias idénticas de cada fragmento ⁽⁷⁶⁾.

• **Secuenciación**

Finalizada la amplificación clonal, se eliminan las hebras reversas, quedando solo las hebras idénticas a las originales. Se introducen nucleótidos modificados con etiquetas fluorescentes específicas para cada base. Estos nucleótidos presentan un terminador reversible que impide la incorporación de más de una base por ciclo, reduciendo así los errores. Cada incorporación emite una fluorescencia característica que es registrada. Antes del siguiente ciclo, la etiqueta se retira para evitar señales superpuestas. Este proceso se repite en paralelo en todos los clústeres hasta completar la secuenciación ⁽⁷⁶⁾.

La exactitud de la secuenciación depende de la intensidad de la señal, y la longitud de lectura, del número de ciclos realizados. Actualmente, Illumina alcanza lecturas de hasta 300 pares de bases ⁽⁷⁶⁾.

• **Secuenciación en tiempo real**

Estas técnicas, desarrolladas por *Pacific Biosciences* (SMRT Technology), son rápidas porque no interrumpen la incorporación de nucleótidos para realizar la lectura ni requieren ciclos de lavado. No precisan amplificación y evitan problemas de desfase. Un dispositivo óptico detecta la fluorescencia emitida por la incorporación de nucleótidos en tiempo real ⁽⁷⁴⁾.

Otras técnicas de secuenciación:

- **Nanoporos:** identifican las bases midiendo variaciones en una señal eléctrica o óptica al atravesar el ADN un nanoporo anclado en una membrana. Desarrollada por Oxford Nanopore Systems (<https://nanoporetech.com/>) ⁽⁷⁴⁾.

- **Observación directa con microscopía:** desarrollada por compañías como ZS Genetics (<http://allseq.com/knowledge-bank/emerging-technologies/zs-genetics/>), permite leer el ADN mediante microscopía electrónica, sin necesidad de amplificación ⁽⁷⁴⁾.

1.3.8 Utilidad de la secuenciación de nueva generación para detectar variantes genéticas

–Variantes de nucleótido único

La detección de SNV es altamente precisa cuando se cuenta con una cobertura mínima de 10–15 lecturas por posición y una tasa de error de secuenciación aceptable. Los algoritmos empleados suelen basarse en modelos bayesianos, calculando la probabilidad condicional de cada nucleótido en función del número de lecturas que lo contienen, la calidad asignada a la base y otros parámetros. En Illumina, los errores tienden a acumularse hacia el final de la lectura; en la plataforma Roche/454, se concentran en secuencias de homopolímeros (regiones con ≥ 6 –7 bases idénticas consecutivas) ⁽⁶⁹⁾.

–Inserciones y deleciones (indels)

Las variantes que implican la ganancia o pérdida de una sola base suelen alinearse incorrectamente al genoma de referencia, aumentando el riesgo de falsos positivos. Un alineamiento *de novo* regional, aunque computacionalmente exigente, mejora la detección, pero su sensibilidad y especificidad aún son inferiores a las de la detección de SNV ⁽⁶⁹⁾.

–Variantes de número de copia e inversiones grandes

La detección inicial de variantes estructurales se realizó con datos de *paired-reads* y *mate-pairs* (pares de lecturas relacionadas, pero no adyacentes ni complementarias). Cuando la distancia entre lecturas o su orientación difiere significativamente de lo esperado, puede indicar la presencia de una variante estructural.

No obstante, un único *mate-pair* no es suficiente para confirmar estas variantes debido a:

1. Imprecisión en el tamaño real del inserto,
2. alineamientos incorrectos que simulan variantes, y
3. la existencia de pares quiméricos.

Un método alternativo consiste en analizar la profundidad de cobertura para identificar deleciones y duplicaciones ⁽⁶⁹⁾.

1.3.9 Aplicaciones de la secuenciación de nueva generación en la práctica clínica

– Diagnóstico molecular de enfermedades hereditarias:

- **Secuenciación dirigida:** es la estrategia más utilizada y consiste en el aislamiento, enriquecimiento y secuenciación de regiones específicas del genoma (regiones codificantes de interés y zonas intrónicas flanqueantes). Se desarrollan paneles genéticos de apoyo al diagnóstico que permiten identificar variantes, polimorfismos, reordenamientos y variantes somáticas de baja frecuencia en numerosas muestras de forma simultánea. Esta técnica es especialmente útil para estudiar enfermedades con componente hereditario en las que ya se han identificado genes o mutaciones asociadas a su etiología. También se aplica en la detección de mutaciones presentes en tumores, lo que permite evaluar si estos podrían responder a terapias con fármacos dirigidos ⁽⁷⁴⁾.

- **Secuenciación de exoma:** consiste en seleccionar las secuencias codificantes del genoma y secuenciarlas. Esta estrategia se emplea para identificar genes causantes de una patología cuando se desconoce la base genética de un fenotipo clínico específico o en enfermedades con alta heterogeneidad genética y fenotípica. No está indicada en casos en los que ya se conoce el gen causante de la patología o cuando se ha identificado la mutación responsable en otros miembros de la familia ⁽⁷⁴⁾.

- **Secuenciación de genoma completo:** el principal reto de esta estrategia es lograr una interpretación eficiente y fiable antes de su uso con fines diagnósticos en genética humana. Actualmente, se aplica en microbiología para obtener genomas completos de bacterias, virus y patógenos fúngicos ⁽⁷⁴⁾.

- **Diagnóstico prenatal de aneuploidías:** la secuenciación de ADN fetal presente en el plasma materno es un método no invasivo y eficaz utilizado para detectar las principales aneuploidías ⁽⁷⁴⁾.

- **Estudio del microtranscriptoma:** permite la identificación, cuantificación y caracterización de cientos de pequeñas moléculas de ARN no codificante (SmallRNAs: miRNAs, snoRNAs, piRNAs) cuya función puede ser clave para el funcionamiento celular ⁽⁷⁴⁾.

- **Identificación de zonas de interacción proteína-ADN (ChIP-seq):** se realiza por la técnica ChIP-Seq que combina la inmunoprecipitación de la cromatina con la secuenciación masiva ⁽⁷⁴⁾.
- **Identificación de patrones de Metilación (metiloma):** utilizada para el estudio de la regulación de procesos clave como la diferenciación celular, el desarrollo o la aparición de enfermedades ⁽⁷⁴⁾.
- **Análisis del transcriptoma (ARN-Seq - Transcriptoma completo):** La secuenciación masiva del ADN complementario (ADNc) genera información global sobre el contenido de ARN de una muestra, incluyendo ARN mensajero, ribosomal, de transferencia y otros ARN no codificantes. Proporciona una forma eficiente de medir niveles de expresión génica, identificar eventos de *splicing* alternativo, fusiones génicas y SNPs de forma simultánea (74).

La utilidad de RNA-seq se ha extendido más allá de la genómica y se ha convertido en una herramienta estándar en la investigación en ciencias de la vida. Dentro de nuestras células, existen diversas especies de ARN con funciones distintas:

- **ARNm:** codifica proteínas.
- **ARNt y ARNr:** participan en la traducción.
- **ARN reguladores:** incluyen ARN antisentido (asARN), microARN (miARN), piRNA, ARN de interferencia pequeña (ARNip), ARN de horquilla corta (ARNhc) y ARN no codificante largo (ARNlnc), todos implicados en la regulación de la expresión génica ⁽⁷⁷⁾.

En algunos casos, el ARNr debe eliminarse para obtener bibliotecas enriquecidas en transcripciones de ARNm. Esto puede lograrse mediante:

- Enriquecimiento de ARNm usando selección de poli (A), dirigida a la poliadenosina monofosfatos en 3'.
- Agotamiento del rRNA con sistemas como Ribo-Zero (Illumina, CA, EE. UU.) ⁽⁷⁷⁾.

El agotamiento de ARNr se utiliza cuando la muestra no proporciona suficiente cantidad o calidad de ARNm mediante selección de poli(A). En casos de material escaso, existen kits específicos de preparación de bibliotecas, como SMART-seq

(Takara Bio, CA, EE.UU.), que incluyen preamplificación y, en algunos casos, una segunda etapa de amplificación.

Para preparar una biblioteca para secuenciación, el ARN se fragmenta a un tamaño adecuado (por métodos físicos o enzimáticos), se sintetiza ADNc a partir de la primera hebra y se añaden adaptadores para permitir la secuenciación. Algunas plataformas (Illumina TruSeq, Takara Clontech SMARTer, PerkinElmer NEXTflex, KAPA Biosystems) incorporan adaptadores tras la síntesis de ADNc. También se añaden índices para identificar muestras y permitir la multiplexación.

La **profundidad de secuenciación** es crítica:

- Estudios indican que aumentar el número de lecturas no siempre aporta más relevancia biológica.
- Detectar ARN de baja abundancia requiere incrementar las lecturas ⁽⁷⁷⁾.
- Para expresión diferencial en humanos, entre 10 y 20 millones de lecturas por muestra suelen ser suficientes para la mayoría de genes.

Una vez generados los datos de secuenciación, es necesario alinearlos con el genoma o con secuencias de referencia del transcriptoma. En los casos en que se busca identificar transcripciones novedosas, se realiza la alineación al genoma seguida de un ensamblaje de transcripciones *de novo*. Posteriormente, se calculan y normalizan los recuentos, para luego identificar las características diferencialmente expresadas ⁽⁷⁷⁾.

La plataforma de secuenciación de Illumina produce archivos FASTQ sin procesar, los cuales representan la secuencia de la biblioteca analizada. El formato FASTQ es un archivo de texto que contiene las secuencias y las puntuaciones de calidad asociadas (Figura 3) ⁽⁷⁷⁾. El procesamiento inicial de estos archivos incluye una evaluación de la calidad mediante herramientas como FASTQC o FASTXToolkit, que generan resúmenes y representaciones gráficas, permitiendo la identificación visual de posibles problemas.

Tras la obtención de las lecturas limpias, se procede a su alineación con el genoma humano —por ejemplo, las compilaciones de Ensembl GRCh38 o NCBI hg38—, con el transcriptoma humano asociado, o mediante pseudoalineación con herramientas como Salmon, Kallisto y Sailfish. El mapeo de lecturas, que abarca

coordenadas de uniones de empalme, requiere algoritmos capaces de distribuir las lecturas en diferentes posiciones exónicas. Entre los programas más utilizados para este fin se encuentran HISAT2, SOAPsplice, TopHat2 y STAR, los cuales generan archivos de secuencias o binarios de mapas de alineación ⁽⁷⁷⁾. Estas alineaciones pueden visualizarse con herramientas como Integrative Genomics Viewer (IGV) o Genome Browse. Los niveles de expresión —ya sea a nivel génico o transcripcional— se estiman a partir de las lecturas mapeadas, generando archivos de anotación de características. Para manejar la variabilidad de los datos, se emplean modelos estadísticos basados en distribuciones binomiales negativas, como los implementados en edgeR, BaySeq, Cufflinks2 y DESeq2. Estas herramientas requieren matrices de recuentos por gen o transcripción, que pueden obtenerse con featureCounts o htseq-count.

Los recuentos sin procesar no son directamente comparables entre muestras debido a sesgos derivados de la longitud de los genes, el tamaño de las lecturas y la profundidad de secuenciación. Para corregir estos factores, se utilizan métricas de normalización como RPKM (*reads per kilobase of exon per million reads*), FPKM (*fragments per kilobase of exon per million reads*) y TPM (*transcripts per million*), siendo esta última la más recomendable para comparar muestras con distinto origen y composición, ya que normaliza primero por la longitud del gen y luego por la profundidad de secuenciación ⁽⁷⁷⁾.

El análisis de RNA-seq busca evaluar cambios en la expresión génica entre diferentes condiciones experimentales. Para este fin, se han desarrollado múltiples métodos de normalización y modelado a partir de recuentos sin procesar, a fin de evitar sesgos derivados de normalizaciones previas. Entre las herramientas más comunes para este análisis se incluyen DESeq2, Cufflinks2, NOISeq y edgeR ⁽⁷⁷⁾.

El resultado del análisis diferencial de expresión suele ser una lista de genes diferencialmente expresados (*differentially expressed genes*, DEG). Para explorar la estructura de los datos, el análisis de componentes principales (PCA) permite reducir la dimensionalidad a dos o tres componentes, facilitando su visualización e interpretación ⁽⁷⁷⁾.

A nivel transcripcional, el análisis diferencial puede detectar cambios en isoformas. El ensamblaje de transcripciones, mediante programas como Cuffdiff2

o StringTie, combina conocimiento previo con la identificación de nuevas isoformas, permitiendo su comparación posterior. El empalme alternativo, presente en el 90–95 % de los genes de mamíferos, puede analizarse a través de herramientas que evalúan el uso diferencial de exones o sitios de empalme, diferenciando entre enfoques centrados en características específicas y aquellos que consideran la expresión global del gen ⁽⁷⁷⁾.

La visualización de datos de RNA-seq se realiza mediante navegadores genómicos como IGV, UCSC Genome Browser y Genome Browser, que permiten observar alineaciones, cobertura y eventos de empalme alternativo. Asimismo, herramientas como MultiQC, QualiMap y RNA-SeQC proporcionan resúmenes estadísticos y métricas de alineación, facilitando la evaluación de la calidad de los datos ⁽⁷⁷⁾.

Mantener flujos de trabajo reproducibles y documentados es esencial para garantizar resultados precisos y trazables, así como para detallar cómo se han manipulado, filtrado y evaluado los datos. En el caso de las enfermedades oculares, el estudio del transcriptoma ofrece información valiosa sobre cómo las alteraciones génicas afectan el desarrollo y la función ocular, y tiene el potencial de identificar nuevos blancos terapéuticos.

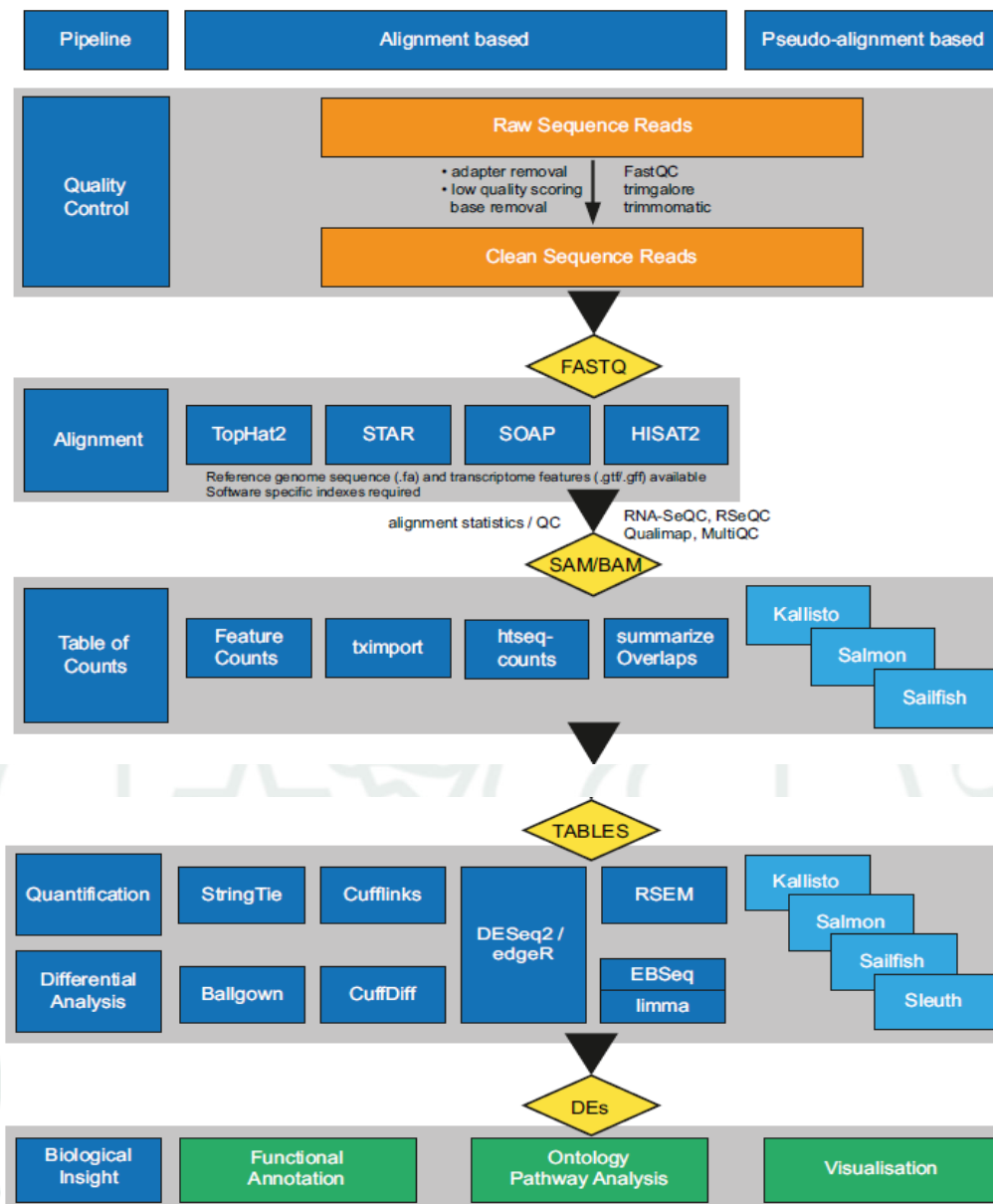


Figura 3. Representación esquemática del procesamiento bioinformático típico de datos de secuencias de alto rendimiento para experimentos de RNA-seq ⁽⁷⁷⁾.

Nota. Owen y Moosajee ⁽⁷⁸⁾

Las lecturas sin procesar generadas por la plataforma de secuenciación (FASTQ) están sujetas a control de calidad y evaluación. Cuando se dispone de un genoma de referencia y una anotación de alta calidad, estas lecturas pueden emplearse en procesos basados en alineación o pseudoalineación.

En el enfoque basado en alineación, las lecturas se asignan al genoma y al transcriptoma de manera consciente del empalme. Las alineaciones resultantes (SAM/BAM/CRAM) se evalúan para determinar la calidad del mapeo y realizar el

recuento de características (genes, transcritos o exones). Estos conteos se modelan para la cuantificación y el análisis diferencial mediante diversos métodos, obteniendo como resultado listas de características diferenciales.

En los métodos basados en pseudoalineación, las lecturas limpias se asignan directamente al transcriptoma, lo que permite la cuantificación inmediata de las características, adecuada para el análisis diferencial. Los resultados de ambos enfoques pueden enriquecerse mediante análisis de ontología génica (GSEA/GO ORA), análisis de rutas (Panther, KEGG, DAVID) y herramientas de visualización (IGV, GenomeBrowse, Bioconductor), con el fin de generar informes integrales.

DAVID: Base de Datos para Anotación, Visualización y Descubrimiento Integrado; GO: Ontología Génica; GSEA: Análisis de Enriquecimiento del Conjunto de Genes; IGV: Visor de Genómica Integrativa; KEGG: Enciclopedia de Genes y Genomas de Kioto; ORA: Análisis de Sobrerrepresentación; Panther: Análisis de Proteínas a través de Relaciones Evolutivas; SAM/BAM: Mapa de Alineamiento de Secuencia/Binario.

1.4 Análisis de antecedentes investigativos

1.4.1 Expresión del gen supresor de tumores p53 en el pterigión

En el año 1997, Tan, D. et al., estudiaron la expresión del gen p53 en la conjuntiva de pacientes con pterigión para determinar si esta lesión corresponde a un trastorno de crecimiento anormal. Utilizando inmunotinción con el anticuerpo monoclonal de ratón pAb 240, analizaron seis casos de pterigión primario y dos de pterigión recurrente. En tres de las ocho muestras se detectó expresión anormal de p53 tanto en el epitelio del pterigión como en la conjuntiva bulbar superior. Los autores concluyeron que la expresión anormal de p53 en el pterigión primario y recurrente sugiere un crecimiento desordenado, más que un proceso degenerativo ⁽⁷⁹⁾.

En 2006, se evaluó si la acumulación de proteína p53 en el pterigión se debía a mutaciones sin sentido en los exones 5–8 del gen p53, como ocurre en tumores malignos con altos niveles de p53 anormal. En 11 tejidos epiteliales de pterigión, la cantidad de proteína p53 detectada por western blot fue mayor que en conjuntiva normal y epitelio límbico corneal. No se identificaron mutaciones en el gen p53, lo que indica que el aumento de la proteína se debe a un mecanismo distinto al

observado en tumores malignos. Además, esta acumulación no indujo apoptosis ni bloqueó la proliferación celular, lo que sugiere una inactivación de las funciones normales de p53 en las células pterigiales ⁽⁸⁰⁾.

En 1998, se analizó la expresión inmunohistoquímica de p53 en tejidos de pterigión. De 38 muestras estudiadas, 35 (92,1 %) no presentaron tinción nuclear específica, y solo tres (7,9 %) mostraron positividad parcial. Los autores concluyeron que la disfunción del gen p53 probablemente no desempeña un papel relevante en la patogénesis del pterigión ⁽⁸¹⁾.

En 2002, se estudió la sobreexpresión del p53 mediante inmunohistoquímica con el anticuerpo monoclonal DO-7 en 13 casos de pterigión (4 primarios, 11 recurrentes) y dos controles de conjuntiva normal. Se halló expresión anormal en el 50 % de los pterigiones primarios y en el 55,5 % de los recurrentes, mientras que los controles fueron negativos. Se concluyó que el pterigión podría resultar de una proliferación celular incontrolada más que de un proceso degenerativo, sin relación aparente entre la expresión anormal de p53 y la recurrencia ⁽⁵⁹⁾.

En 2005, se examinó el espectro de mutación del gen p53 y su relación con los niveles de proteína en el pterigión. Se analizaron 51 muestras de pterigión y 10 de conjuntiva normal, secuenciando los exones 4–8 del gen. Se detectaron mutaciones puntuales en 8 casos (15,7 %): 6 sustituciones y 2 deleciones, localizadas principalmente en los exones 6 y 7. La inmunohistoquímica mostró positividad en el 39,2 % de las muestras; las deleciones fueron negativas para p53, mientras que las sustituciones fueron positivas. No se observó correlación entre la presencia de mutaciones y los niveles de proteína detectados ⁽⁸²⁾.

En 2020, se evaluó la expresión de p53, Bcl-2 y Ki-67 en 35 casos de pterigión primario, 32 recurrentes y 30 controles de conjuntiva normal. Los tres marcadores presentaron niveles significativamente más altos en pterigión que en el grupo control ($p < 0,001$ para todos). No se hallaron diferencias significativas en la expresión de p53 y Ki-67 entre pterigión primario y recurrente ($p = 0,321$ y $p = 0,215$, respectivamente) ⁽⁸³⁾.

Expresión de las proteínas p53 y Ki-67 en pacientes con pterigion de mayor severidad y duración (2021). Se analizaron 43 muestras de pacientes indios pertenecientes a diferentes grupos etarios. Las muestras se clasificaron según la gravedad del pterigion

(leve, moderado y grave) y según su tiempo de evolución (<4 años y >4 años). El análisis se realizó mediante inmunohistoquímica, utilizando anticuerpos dirigidos contra las proteínas p53 y Ki-67, considerando significativa una expresión superior al 5 %.

En total, la expresión de p53 y Ki-67 fue positiva en 33 casos. La positividad de p53 se registró en el 33,3 % de los casos leves, el 78,4 % de los moderados y el 100 % de los graves. Asimismo, se observó un incremento en la expresión de p53 según la duración de la enfermedad: 79,3 % de positividad en casos con menos de 4 años y 92,9 % en aquellos con más de cuatro años de evolución.

Este estudio evidenció que el aumento tanto en la duración como en la gravedad del pterigion se asocia con una mayor expresión de p53. En el caso de Ki-67, su expresión se incrementó con la duración de la enfermedad, aunque no mostró relación con la gravedad ⁽⁸⁴⁾.

CAPÍTULO II

METODOLOGÍA

2.1 Tipo y nivel de investigación

Esta investigación es de tipo observacional, transversal. Respecto al nivel de investigación, el presente es un estudio explicativo.

Es de tipo observacional porque se recopilaban datos, por medio de la revisión de los archivos que contienen la información básica de los pacientes (diagnóstico clínico, exámenes auxiliares), y el genotipado (diagnóstico molecular) mediante técnicas de biología molecular (NGS).

Es transversal porque los datos de los pacientes y las muestras recolectadas fueron obtenidas en un solo espacio de tiempo. Las pruebas moleculares se llevaron a cabo con las muestras obtenidas en la cirugía.

Este estudio es explicativo porque comenzó con una idea general y utilizó la investigación como una herramienta para encontrar nuevos conocimientos que no se estudiaron antes.

El presente trabajo corresponde al área de ciencias médicas y de salud y la línea de investigación corresponde a biología molecular oftalmológica

Su objetivo es proporcionar detalles donde existe una pequeña cantidad de información sobre biología molecular del pterigion.

2.2 Técnicas e instrumentos

2.2.1 Técnica

- Extracción de RNA de muestra biológica
- Secuenciación por NGS
- Observación documental

2.2.1 Instrumento

- Ficha de recolección de datos y bitácora de laboratorio

2.2.2 Cuadro de coherencia

Variable	Indicadores	Sub indicadores	Técnicas
Exposición a radiación ultravioleta y otras noxas ambientales en pacientes con pterigión primario	Edad Sexo Trabajo habitual Ojo afectado Grado de pterigión	18-39 años 40-59 años >60 años Masculino Femenino Ama de casa Agricultor Trabajo en oficina Derecho Izquierdo Primer grado Segundo grado Tercer grado Cuarto grado	Ficha de recolección de datos e historia clínica
Presencia de mutaciones del gen humano p53 en pacientes con pterigión primario	Secuenciación del gen humano P53	Fichero BAM	Extracción de ARN Secuenciación por NGS

2.3 Campo de verificación

2.3.1 Ubicación espacial

- Servicio de oftalmología del Hospital III Juliaca EsSalud
- Laboratorio de biología molecular de la Universidad Católica de Santa María

2.3.2 Ubicación temporal

- Meses: agosto 2019 a diciembre 2020.

2.4 Unidades de estudio

El universo estuvo constituido por todos los pacientes atendidos en el servicio de oftalmología del Hospital III Juliaca EsSalud.

La población de estudio la conformaron siete muestras de pterigion de siete pacientes atendidos en el servicio de oftalmología del Hospital III Juliaca EsSalud, que cumplieron con los criterios de inclusión y exclusión.

- **Criterios de inclusión**

Pacientes operados de escisión de pterigión con autoinjerto conjuntival en el servicio de oftalmología del Hospital III Juliaca durante el período agosto 2019-febrero 2020.

- Pacientes cuyo diagnóstico preoperatorio sea claro y determinante para realizar la cirugía.
- Pacientes con seguimiento en el posoperatorio no menor de 6 meses que tengan los datos en la ficha clínica correctamente anotados.
- Pacientes cuyos antecedentes y ficha clínica se encuentren en los archivos de la institución y cumplan con el completo llenado de la información requerida, cualquier falta de información será observada para su inclusión.
- Pacientes que voluntariamente firmen el consentimiento informado.

- **Criterios de exclusión**

- Pacientes con cirugías oculares previas.
- Antecedentes de enfermedades oculares: glaucoma, uveítis, ojo seco moderado o severo
- Pacientes con enfermedades sistémicas. diabetes o artritis reumatoide.
- Pterigión de tipo atrófico.
- Pacientes que voluntariamente decidieron no participar en el estudio.

2.5 Materiales, equipos y reactivos

2.5.1 Materiales

- Frascos de cultivo celular de 15 y 75 ml
- Tubos Falcon de 15 y 50 ml
- Tubos Eppendorf

- Pipeta automática de 0.5 µl a 10 µl, 20 µl a 200 µl, 100 µl a 1000 µl
- Tips estériles para micropipetas de 1000, 200 y 20 µl
- Vaso precipitado
- Tijera quirúrgica Vannas
- Mango de bisturi
- Hoja de bisturi número 15

2.5.2 Equipos

- Cámara de Bioseguridad
- Campana de flujo laminar
- Centrifuga para tubos Falcon de 15mL
- Centrifuga refrigerada
- Termociclador para PCR en tiempo real
- Equipo para Electroforesis
- Vortex
- Congelador - 20 °C, - 80 °C
- Software IGV

2.5.3 Reactivos

- SBF – Gibco™ suero fetal de bovino
- Medio DMEM - Gibco™
- PBS (buffer fosfato salino) - Merck
- Etanol 96° y 70°
- Agarosa - TrackIt™
- Buffer tris-acetato-EDTA 50x (TAE)
- Buffer Loading
- Tripsina - Gibco™
- Trizol – TRI Reagent – Merck - 100mL
- F12 suplemento
- Cloroformo
- Colagenasa tipo I – SIGMA

2.6 Estrategia de recolección de datos

2.6.1 Toma de muestra

Se obtuvo la muestra en sala de operaciones del Hospital III Juliaca, mediante técnica de exéresis y plastia libre, realizados por una sola cirujana utilizando la misma técnica en todos los pacientes, se aplicó asepsia y antisepsia de la zona operatoria, anestesia tópica y 1 ml de lidocaína al 2% con epinefrina en concentración 1:10,000 en el tejido subconjuntival circundante al pterigión, la cirugía se realizó con el microscopio Zeiss opmi Lumera 300, se procedió con la disección del ápex del pterigión desde la córnea al limbo con bisturí número 15, el cuerpo del pterigión se cortó con tijera de Wescott, los tejidos subconjuntivales fueron removidos prolijamente, teniendo precaución de no dañar los músculos adyacentes, luego se colocó mitomicina C al 0,2% durante 3 minutos en torundas de algodón en el tejido subconjuntival y posterior lavado ocular con 50 cc de cloruro de sodio al 0,9%, la esclera se cubrió mediante autoinjerto de conjuntiva de la región temporal superior y se suturó con puntos separado de nylon 10/00.

Los especímenes frescos de pterigión fueron lavados con 50 cc de agua destilada y se dividieron en dos realizando una incisión longitudinal a lo largo del cuerpo y la cabeza del pterigión, se procedió a colocarlos en 500 ul de medio de cultivo DMEM (Dulbecco Modified Eagle Medium) a 4 °C y en otro tubo eppendorf con 500 ul de Trizol Reagent a -20°C, y fueron transportados al laboratorio.

2.6.2 Análisis Molecular

Las muestras fueron procesadas en el laboratorio de Biología Molecular de la Universidad Católica de Santa María de la ciudad de Arequipa- Perú, se utilizó técnica estéril para manejo de los tejidos, que fueron procesados en la cabina de flujo laminar.

2.6.2.1 Extracción de ARN

- a) La muestra, se lavó dos veces con PBS estéril y se realizó el trozado de las muestras con tijera Vannas de capsulotomía Rumex y fórceps recto Katena, obteniendo cubos de aproximadamente 1 x 1 mm.
- b) La muestra se colocó en tubo eppendorf y se incorporó 500 ul de cloroformo
- c) Se centrifugó a 12000 rpm por 10 minutos

- d) Se transfirió la fase acuosa a un nuevo eppendorf (aproximadamente 300 - 400µL de muestra) sin tocar la interfase.
- e) Se repitió el mismo paso 2 veces más.
- f) Agregamos 500 ul de isopropanol
- g) Se centrifugó a 12000 rpm por 10 minutos.
- h) Eliminamos el isopropanol por inversión en un papel de toalla una sola vez y aislamos el pellet.
- i) Suspendimos el pellet de DNA con 500µL de etanol al 70-75%.
- j) Centrifugamos a 12000 rpm por 10 minutos.
- k) Eliminamos el sobrenadante por inversión en papel toalla una sola vez y aislamos el pellet.
- l) Dejamos secar 30 segundos aproximadamente, luego cerramos la tapa del eppendorf.
- m) Resuspendimos el pellet con 50 µL de agua para inyección.
- n) Analizamos la integridad y la cantidad de RNA por electroforesis y espectrofotometría respectivamente.
- o) Preservamos el RNA a -20°C.
- p) Los 7 pellet preparados suspendidos en 1 cc de agua libre de RNasa, usando como conservante hielo seco, fueron remitidos al laboratorio de NOVOGEN y Admera Health Biofarma Services en Estados Unidos de Norteamérica (USA)

2.6.2.2 Extracción de ARN Total

Las muestras de ARN enviadas al laboratorio Admera en el estado de New Jersey, Estados Unidos, para el secuenciamiento por la tecnología NGS, fueron sometidas a controles de calidad utilizando el Kit de inicio de cuantificación invitrogen fluorómetro Qubit 4, que permite medir con precisión la cantidad, calidad e integridad de ARN.

Se verificó la extracción efectiva del ARN, para ello se corrió lo extraído en gel de agarosa teñido con el reactivo fluorescente no mutagénico (CSL-RUNSAFE), leído al transiluminador LED de luz azul (290). También, se determinó la cantidad de ARN en la plataforma Picogreen usando el método fluométrico, mientras que

el número de integridad del RNA (RIN) fue obtenido mediante el Bioanalizador Agilent Technologies 2100 ⁽⁸⁵⁾.

Para efectos de obtener un ARN sin contaminantes, el ADN es digerido por las DNasas que contiene el reactivo utilizado en el protocolo y, además, previa elaboración de la librería, si existe algún remanente, es removido durante la purificación del ARN mensajero ⁽⁸⁵⁾.

Los valores de RIN encontrados en las muestras tuvieron cifras por debajo de 7, por lo que, según los estándares de calidad, se optó por el método de la deplección de RNA ribosomal de Illumina, el cual necesita RIN: ≥ 2 , para la construcción de la librería de RNAseq.

2.6.2.3 Secuenciamiento del Transcriptoma (Protocolo Illumina)

Los ARNs que fueron extraídos cumplieron con los requisitos necesarios (Número de integridad de ARN y concentración de ARN) fueron sometidas a la preparación de la librería (RNA-seq).

Para poder realizar este procedimiento, el ARN total tuvo que ser convertido a ADN complementario, pasando por el protocolo indicado por la guía de preparación de la muestra: “TruSeq Stranded Total RNA” para la preparación de la librería ⁽⁸⁵⁾.

La plataforma Next Generation Sequencing utilizada para el secuenciamiento fue la “NovaSeq 6000”.

El sistema NovaSeq 6000 representa la plataforma de secuenciación de la tecnología de Illumina de alto rendimiento, el instrumento, se basa en la secuenciación por síntesis (sequencing by synthesis), en la que los fragmentos de ARN segmentados se copian en el ADNc de la primera cadena mediante cebadores aleatorios y transcriptasa inversa y seguidamente, se produce la síntesis de ADNc de la segunda cadena mediante ADN polimerasa I y ARNasa H (se incorpora dUTP en lugar de dTTP para generar ADNc bicatenario, obteniendo ADNc con extremos romos) ⁽⁸⁵⁾.

Seguidamente, se añade una base única “A” en estos fragmentos de ADNc y se produce la posterior ligadura del adaptador ⁽⁸⁵⁾.

Los productos se purifican y se enriquecen con PCR para crear la biblioteca definitiva de ADNc ⁽⁸⁵⁾.

El método Truseq stranded de ARN total con el kit Ribo-Zero fue empleado para la preparación de la librería tipo Paired-end (PE) ⁽⁸⁵⁾.

Se usó el kit Ribo-Zero, porque mejora los resultados de RNA-seq al eliminar el mtrRNA (RNA mitocondrial) y el rRNA (RNA ribosomal) citoplásmico, e incluso de muestras de ARN degradadas o archivadas, lo que permite un mejor mapeo de las lecturas exómicas, maximizando la cantidad de lecturas útiles de ARN-seq del ARN total derivado de células de mamíferos ⁽⁸⁵⁾.

El procedimiento para el secuenciamiento por la tecnología NGS de Illumina incluye 4 pasos:

- a) **Preparación de la muestra:** el ARN fue extraído de la muestra; después de aplicar el control de calidad (QC), las muestras que calificaron fueron incluidas para la construcción de la librería
- b) **Construcción de la librería:** la librería fue preparada a partir de la fragmentación aleatoria de la muestra de ADN complementario, seguido por la ligación 5' y 3'. Alternativamente, la “tagmentación” combinó las reacciones de fragmentación y ligación en un solo paso que incrementó enormemente la eficiencia del proceso de la preparación de la librería. Los fragmentos ligados al adaptador se amplificaron por PCR y se purificaron en gel.
- c) **Secuenciamiento:** para la generación de agrupamientos, la biblioteca se cargó en una celda de flujo donde los fragmentos se capturaron en un campo de oligos unidos a la superficie y que son complementarios a los adaptadores de la librería. Luego, cada fragmento se amplificó por grupos, cuando se completó la generación del clúster, las plantillas estuvieron listas para la secuenciación propiamente dicha en el equipo. El número de lecturas fue de 40 a 60 millones por muestra durante el secuenciamiento.
- d) **Datos sin procesar (raw data):** la data resultado del secuenciamiento fue procesada por el software y generó un archivo de tipo FastQ.

2.6.3 Análisis Bioinformático

La raw data fue recibida en formato FastQ, por cada muestra se generaron dos archivos ya que el tipo de corrida utilizado para la librería fue paired – end, en la que se obtuvo lecturas de las dos terminaciones, tanto en el sentido ‘forward’ como en el sentido ‘reverse’.

El primer tipo de secuencias que surgió de la plataforma NGS, es el diagrama de flujo o cromatograma ⁽⁸⁵⁾.

Estas secuencias están representadas por una serie de picos a lo largo del tiempo donde cada pico es la intensidad de la señal y el tiempo es el orden de las bases dentro de la secuencia de ADN. Para transformar estos picos se usó un “algoritmo de llamada de base” para designar un nucleótido a cada pico presente en el archivo sin procesar.⁸⁸

La raw data generada en formato FastQ, fue analizada por programas bioestadísticos considerando las siguientes etapas:

2.6.3.1 Control de Calidad de las lecturas secuenciadas

Los parámetros de calidad de los transcritos se evaluaron con el programa FASTQC, los resultados son expresados en los siguientes figuras:

- a) **Estadística Básica:** el programa permitió una composición estadística de los archivos analizados. En este figura se encuentra información como la longitud de las secuencias, el total de las secuencias procesadas, codificación ASCII, entre otros⁽⁸⁵⁾.
- b) **Calidad de la Secuencia por Base:** para cada posición de lectura se dibujó un figura de tipo cajas. La línea central roja en cada caja es el valor de la mediana, la caja amarilla representa el rango de los inter-cuartiles (25-75%), los bigotes arriba y debajo de la caja amarilla simbolizan el 10% y 90% de puntos, y finalmente la línea azul simboliza la calidad ⁽⁸⁵⁾.
- c) **Contenido por base de la secuencia:** este figura señala la proporción porcentual de cada base en cuanto a su posición en el archivo para lo cual cada una de las cuatro bases de ADN está expresada por colores ⁽⁸⁵⁾.
- d) **Duplicado de secuencias:** este módulo cuenta el grado de duplicación de cada

secuencia y creó una figura que muestra el número relativo de secuencias con diferentes grados de duplicación ⁽⁸⁵⁾.

2.6.3.2 Alineamiento de las lecturas: STAR

Las lecturas secuenciadas son comúnmente mapeadas teniendo de base un genoma o un transcriptoma. Este paso del análisis se llevó a cabo con el software de alineación de transcripciones empalmadas a una referencia (STAR).

STAR es un alineador RNA-seq universal ultrarrápido, que se diseñó para abordar específicamente muchos de los desafíos del mapeo de datos de RNA-seq, y utiliza una estrategia novedosa para alineaciones empalmadas.

2.6.3.3 Ensamblaje de las lecturas: Cufflinks

Para poder efectuar el ensamblaje se utilizó el software libre Cufflinks. Este se encargó de ensamblar los transcritos individuales a partir de las lecturas de RNA-seq que han sido alineadas en relación con el genoma de referencia ⁽⁸⁵⁾.

2.6.3.4 Análisis de la expresión génica diferencial: Cuffdiff

Cuffdiff, calculó los niveles de expresión en dos o más muestras y probó la relevancia estadística de los cambios detectados en la expresión entre dichas muestras, identificó genes que están regulados diferencialmente a nivel transcripcional o postranscripcional ⁽⁸⁵⁾.

Tomó las lecturas alineadas de dos o más condiciones e informó genes y transcripciones que se expresan diferencialmente mediante un riguroso análisis.

Este software reportó el análisis diferencial de las muestras en cuestión y se obtuvieron los archivos de los transcritos ensamblados (BAM), que se generaron a partir del software Cufflinks.

Para la visualización de los resultados de RNA-seq de p53, se utilizó IGV_2.9.4 (Integrative Genomic Viewer), los archivos BAM fueron leídos por la plataforma IGV, donde se reportaron las mutaciones en el gen p53, la ventaja de usar esta plataforma es que nos permitió obtener la posición exacta de la localización de la mutación, la que debe ser evaluada en la plataforma de SNP (polimorfismos de un

solo nucleótido), para verificar las variantes genéticas en p53 y correlacionar con las mutaciones encontradas en el presente estudio.

2.6.3.5 Análisis de rutas las metabólicas y funciones fisiológicas de los genes diferencialmente expresados (upregulated y downregulated): KEGG-DAVID

El archivo Excel generado por Cuffdiff contiene el nombre o denominación de los genes mapeados, este archivo tiene la métrica TPM (transcripts per million),

Luego de obtener esta información se realizó la búsqueda en la Enciclopedia Kyoto de Genes y Genomas (KEGG según siglas en inglés) y la búsqueda de las rutas metabólicas en la plataforma DAVID. Esta búsqueda se realizó a partir de los genes expresados diferencialmente (sub-regulados o sobre-regulados).

2.6.3.6 Detección de genes de fusión

La información obtenida de RNA-seq, fue analizada por Arriba (versión 1.2.0) una novedosa herramienta computacional, con alta sensibilidad y corto tiempo de ejecución para detectar y visualizar fusiones de genes a partir de datos de RNA-Seq con parámetros predeterminados ⁽⁸⁶⁾.

El rendimiento de Arriba, proporcionó la secuencia de transcripción que flanquea el sitio de unión, que ayuda con el diseño de cebadores para la validación mediante secuenciación Sanger. También calculó la secuencia de péptidos resultante de la transcripción quimérica, que puede servir como base para la predicción de neoepítopos derivados de la fusión, ofrece una serie de funciones útiles que van más allá de la mera predicción de puntos de ruptura de fusión ⁽⁸⁶⁾.

Las lecturas de apoyo son mínimas, el flujo de trabajo de Arriba requirió de un solo paso de alineación seguido de un proceso de filtrado.

La fracción de errores de coincidencia en las lecturas de apoyo fue inferior a 0,8, los genes con más del 30% de identidad de secuencia se consideraron homólogos y se eliminaron mediante el filtro de "homólogos" ⁽⁸⁶⁾.

Una vez que se recopilaron todas las alineaciones candidatas, se aplicó un conjunto de filtros para eliminar artefactos y enriquecer los resultados con un nivel elevado de confianza ⁽⁸⁶⁾.

2.7 Organización

- Autorización del jefe de servicio de cirugía especialidades del Hospital III Juliaca
- Aspectos éticos: El comité de ética del Hospital III Juliaca (código de protocolo:1299-2020-494) aprobó el desarrollo del estudio.
- La investigación se realizó cumpliendo los principios básicos y operacionales de las normas éticas internacionalmente exigidas, como el respeto al individuo, su derecho a autodeterminación, consentimiento informado, poniendo en primer plano la salud e integridad psicofísica del paciente

2.8 Recursos

- a) Recursos humanos
 - Investigadora: Mgter. Berioska Manzaneda Murguía
 - Asesora: Dra. Pacheco Aranibar Jani
- b) Recursos económicos
 - Universidad Católica de Santa María
 - Investigadora
- c) Recursos institucionales
 - Hospital III Juliaca EsSalud
 - Universidad Católica de Santa María

2.9 Análisis estadístico

El análisis y procesamiento de datos, se realizó utilizando el paquete SPSS (20.0 para Windows, SPSS Inc., Chicago, IL, EE. UU.). Se aplicó estadística descriptiva, utilizando tablas de distribución de frecuencias en la que los datos categóricos se presentaron como frecuencia y porcentaje.

CAPÍTULO III

RESULTADOS Y DISCUSIÓN

Tabla 2. *Características epidemiológicas de pacientes con pterigion primario*

Muestra	Edad	Sexo	Ojo operado	Grado PTG	Trabajo
1	47	F	Izquierdo	3	Agricultor
2	36	M	Izquierdo	1	Docente
3	33	F	Derecho	3	Su casa
4	49	M	Izquierdo	3	Docente
5	56	F	Derecho	3	Su casa
6	66	M	Izquierdo	3	Agricultor
7	39	M	Izquierdo	3	Docente

Nota. Elaboración propia.

La edad promedio de los pacientes fue de 46.57 años, el 57.14 % de sexo masculino. La presentación más frecuente fue pterigion en ojo izquierdo; al momento de la cirugía, el 85.71 % presentaban tercer grado de pterigion, y el 42.86 % labora como docente.

Tabla 3. *Identificación de muestras y resultados de control de calidad*

N° muestra	Tipo	Volumen (ul)	Admera Health concentración (ng/ul)	Admera Health cantidad total (ng)	RIN	DV200 (%)
1	RNA	50	34.0	1700.0	3.2	74.87
2	RNA	50	44.0	2200.0	2.1	49.05
3	RNA	50	432.0	21600.0	3.4	62.84
4	RNA	50	21.20	1060.0	3.5	99.62
5	RNA	50	0.20	10.2	2.8	73.26
6	RNA	50	60.40	3020.0	2.8	59.75
7	RNA	50	17.80	890.0	2.2	49.47

Nota. Elaboración propia.

Las muestras fueron procesadas en el laboratorio Admera, el volumen requerido por el laboratorio para realizar RNAseq es ≥ 20 ul, y todas las muestras tuvieron 50 ul, sin embargo, el laboratorio permite muestras mayores de 1 ug, y, en este caso, las siete muestras tuvieron cantidades totales superiores a las solicitadas. En cuanto a la concentración de ARN, esta fue de entre 200 ng – 1 ug ; el número de integridad del ARN (RIN) de las siete muestras estuvo en un rango entre 3 y 4 que funciona para realizar amplificación de pequeñas regiones génicas, el porcentaje de fragmentos de ARN > 200 nucleótidos (DV 200) en la mayoría de muestras presenta cifras superiores a 66.1 %.

Tabla 4. Estadísticas generales MultiQC

N° Muestra	% Asignado	M Asignado	% Alineado	M Alineado	% Recortado	% GC	Largo	M Secuencias
1	30.6%	13.4	77.3%	27.1	8.1%	46%	142 bp	35.0
2	34.2%	13.9	83.7%	29.3	6.1%	44%	145 bp	35.0
3	26.3%	16.2	69.1%	32.4	11.8%	49%	136 bp	46.8
4	28.8%	15.3	63.3%	22.6	11.9%	51%	135 bp	35.7
5	27.6%	16.4	64.8%	23.1	10.1%	53%	104 bp	36.1
6	17.7%	9.0	52.8%	17.1	14.0%	53%	131 bp	32.4
7	32.9%	18.4	82.3%	39.2	6.8%	44%	144 bp	47.6

Nota. Elaboración propia.

La media de las lecturas asignadas es de 28.3 %, y la media de las lecturas mapeadas de forma única es de 70.5 %.

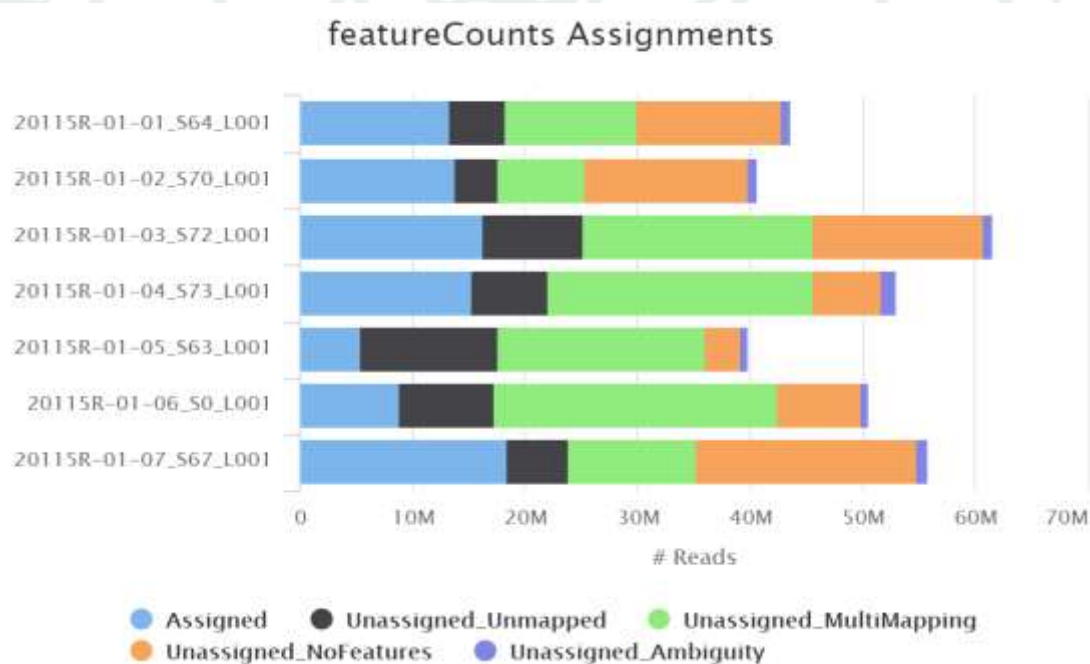


Figura 4. Funciones de conteo asignadas en porcentaje

Nota. Admera health.

En la figura se observa que la media de las lecturas asignadas es de 28.3 %, las no mapeadas son de 15.04 %, las multimapeadas marcan un 34.26 %, las lecturas no asignadas sin características son de 22.63 %, y de las ambiguas no asignadas 1.75 %.

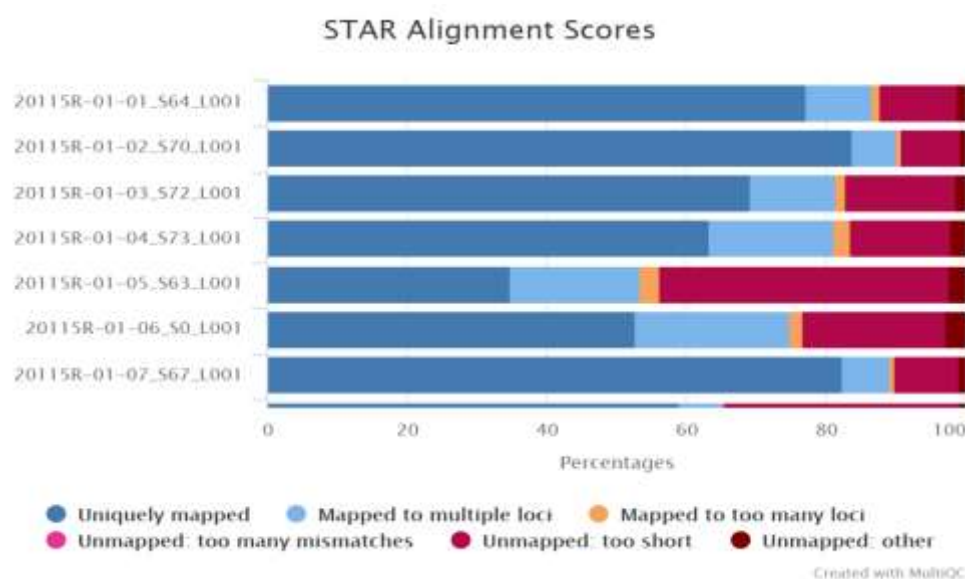


Figura 5. STAR – Puntuaciones de alineación

Nota. Admera health.

STAR brinda información relacionada con la media de las lecturas mapeadas en múltiples loci, la cual marca un 13.29 %. En cuanto a las lecturas mapeadas en demasiados loci señala un 1.53 %, y las lecturas no mapeadas por demasiados desajustes un 17.33 %.

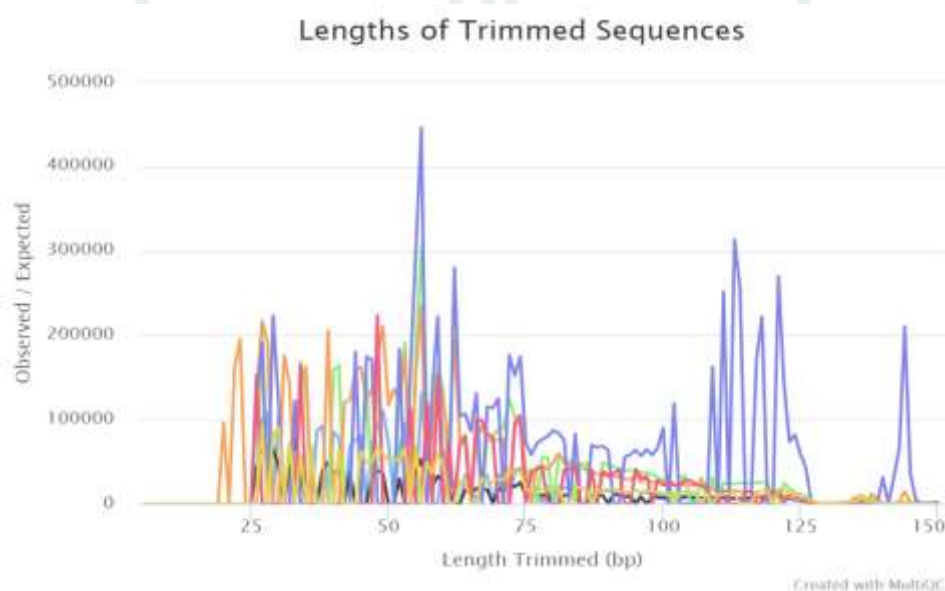


Figura 6. Cutadapt: Longitudes de secuencias recortadas

Nota. Admera health.

Esta figura evidencia el número de lecturas con longitudes de adaptador recortadas. Observed/Expected muestra los recuentos sin procesar, divididos por el número esperado debido a errores de secuenciación.



Figura 7. *FastQC. Histogramas de calidad de secuencia*

Nota. Admera health.

El histograma expone una descripción general del rango de valores de calidad en todas las bases en cada posición en el archivo FastQC. El eje Y representa la calidad de secuencia media (phred score) y el eje X las bases de las lecturas (bp). En lo relacionado con la calidad de secuencia de las siete muestras, es preciso señalar que es muy buena debido a su posición en la zona verde y no presenta dispersión.



Figura 8. *FastQC. Puntuaciones de calidad por secuencia*

Nota. Admera health.

La figura muestra el número de lecturas con puntajes de calidad promedio. El eje X representa la calidad de secuencia media (phred score) y el eje Y el conteo de las lecturas. Todas las muestras presentan puntuaciones de alta calidad y se ubican en el rango de calidad entre 30 y 40.

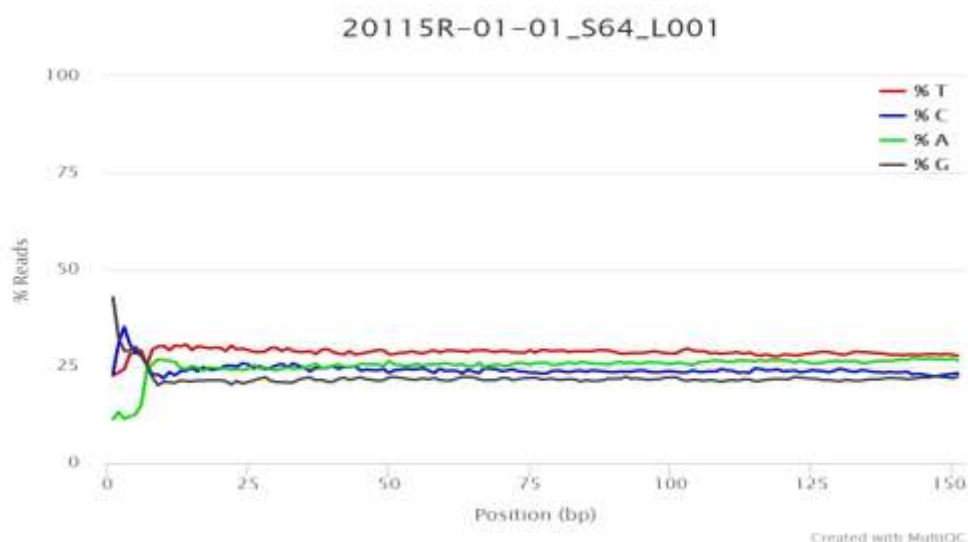


Figura 9. FastQC. Por contenido de secuencia base-Muestra 1

Nota. Admera health.

Esta figura señala la proporción de la posición de cada base según cada una de las cuatro bases que forman el ADN en la muestra 1. Las líneas de las bases timina (T), citosina (C), adenina (A), guanina (G), corren paralelas entre sí.

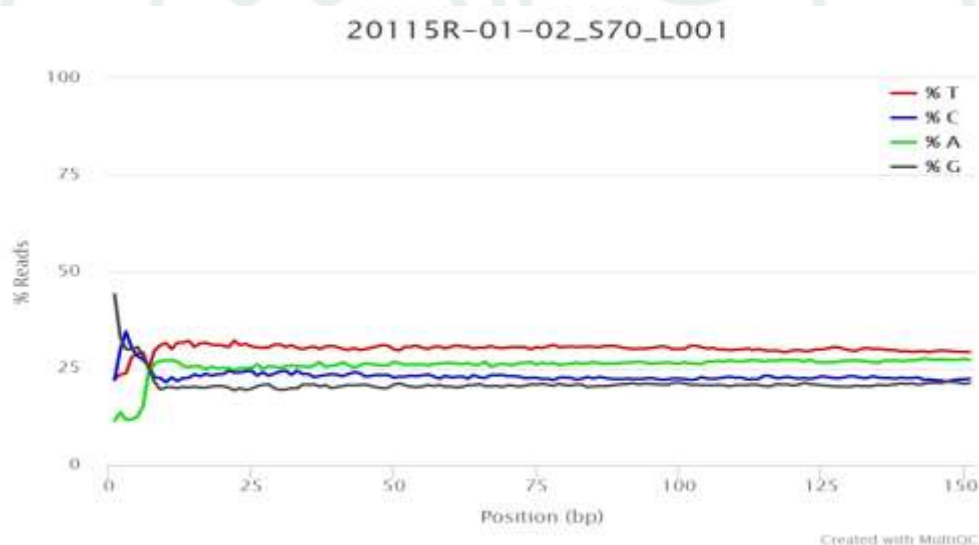


Figura 10. FastQC. Por contenido de secuencia base-Muestra 2

Nota. Admera health.

En la muestra 2, las líneas de las cuatro bases que forman el ADN en la muestra 2, timina (T), citosina (C), adenina (A), guanina (G), se encuentran paralelas entre sí.

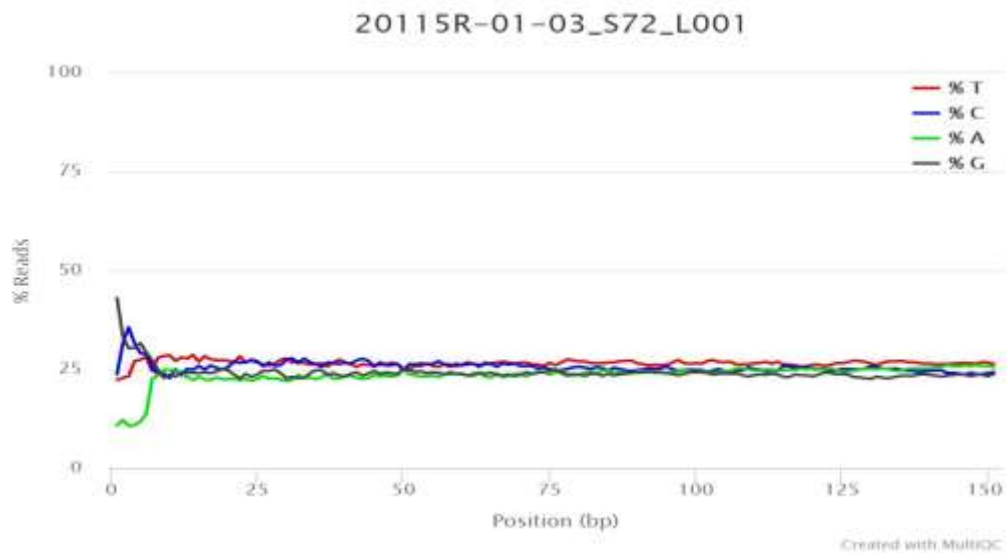


Figura 11. FastQC. Por contenido de secuencia base-Muestra 3

Nota. Admera health.

En esta se señala la proporción de la posición de cada base según cada una de las cuatro que forman el ADN, las cuales se encuentran paralelas entre sí en la muestra 3.

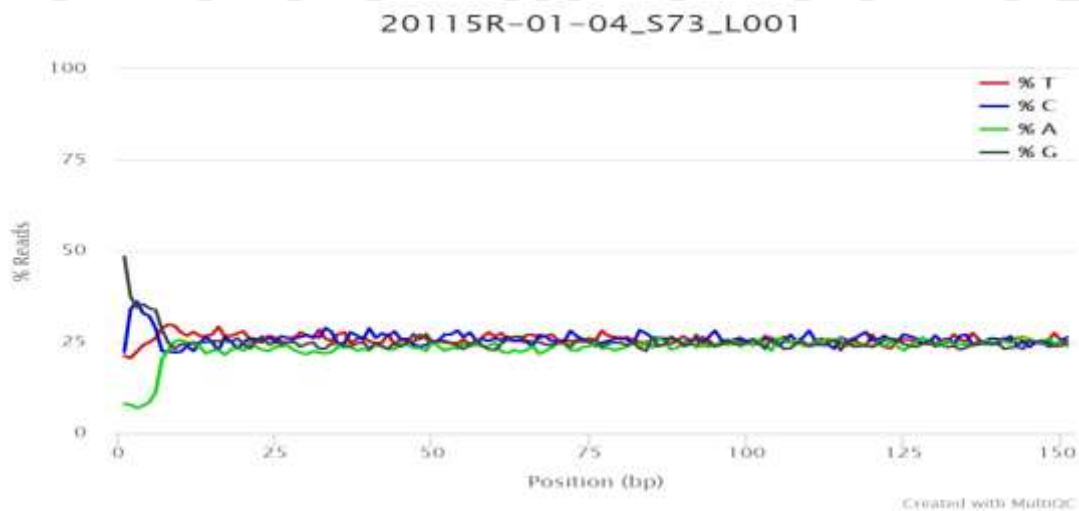


Figura 12. FastQC. Por contenido de secuencia base-Muestra 4

Nota. Admera health.

En la muestra 4, se puede observar la existencia de diferencias entre unas bases y otras cuando la cantidad de A debería ser similar que la de T, así como con G y C.

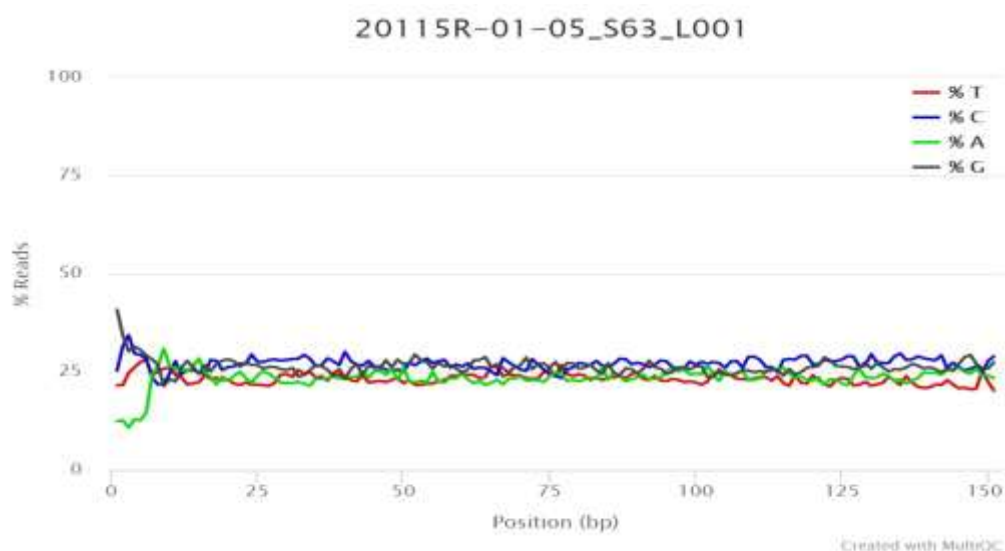


Figura 13. FastQC. Por contenido de secuencia base-Muestra 5

Nota. Admera health

En la muestra 5, las cuatro bases que forman el ADN no corren paralelas entre sí, puesto que existen desviaciones que predominan al principio debido al uso de primers.

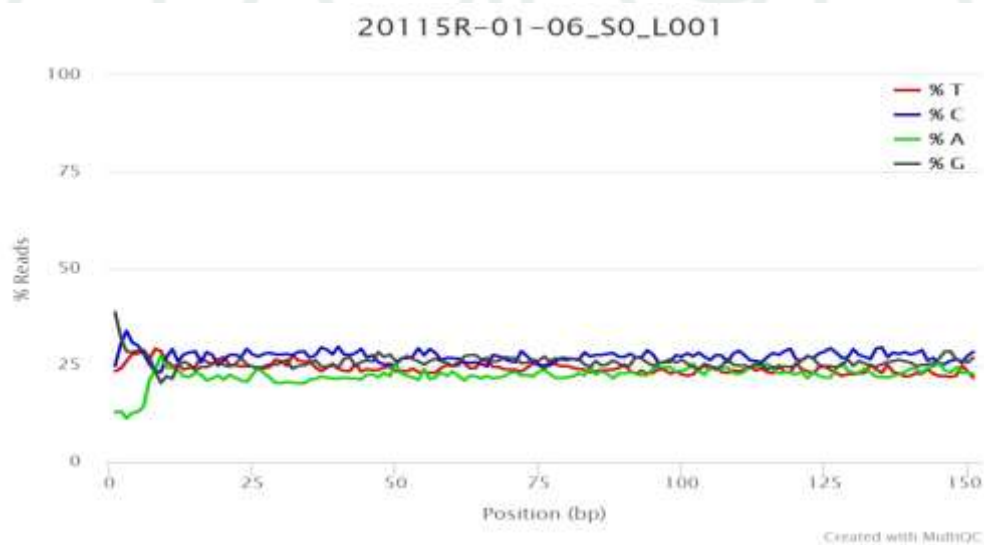


Figura 14. FastQC. Por contenido de secuencia base-Muestra 6

Nota. Admera health

En la muestra 6, la diferencia entre las bases no va más allá del 10 % en cualquier posición.

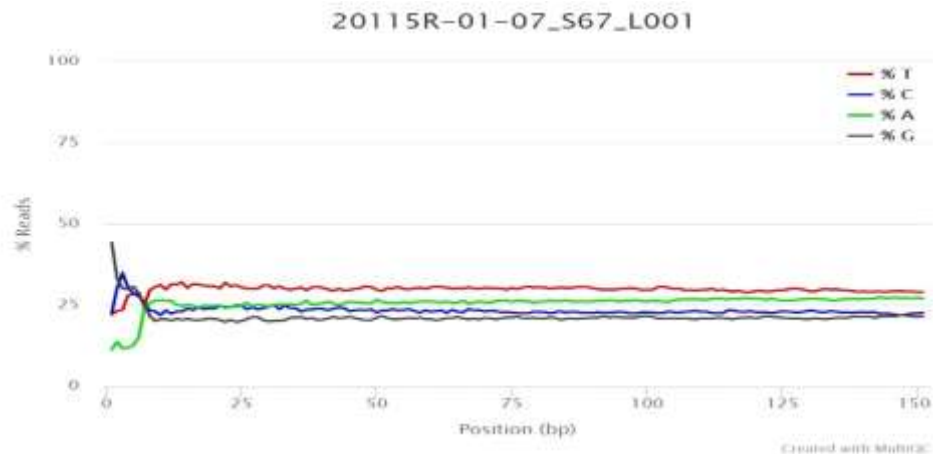


Figura 15. *FastQC. Contenido de ATCG por base secuenciada - Muestra 7*

Nota. Admera health

En la muestra 7, las líneas de cada base corren paralelas entre sí.

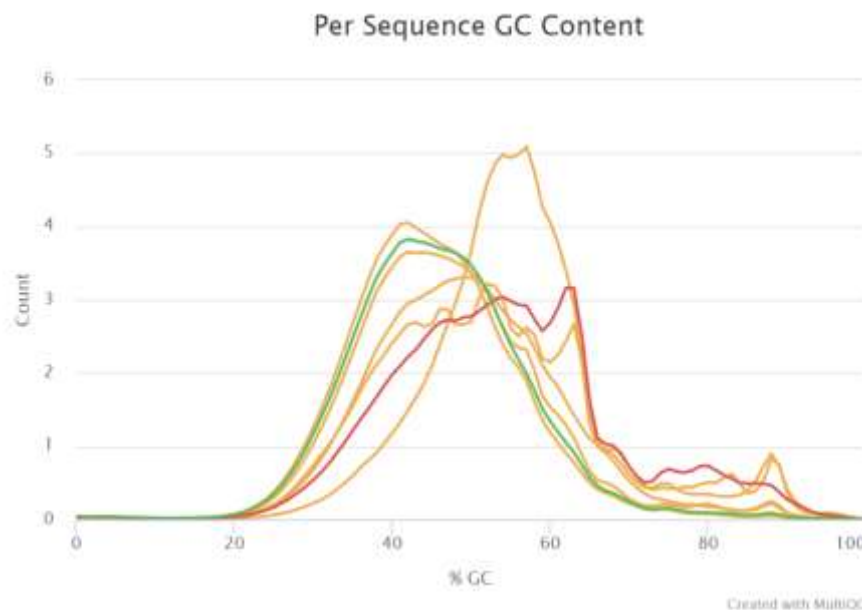


Figura 16. *FastQC. Contenido de GC por secuencia*

Nota. Admera health

Es posible observar el contenido de GC en toda la longitud de cada secuencia en las siete muestras. En el eje X se muestra el porcentaje medio del contenido de G y C, y en el Y se evidencia el número de lecturas. Cabe mencionar que la mayoría de las muestras presenta una curva de Gauss con distribución normal.

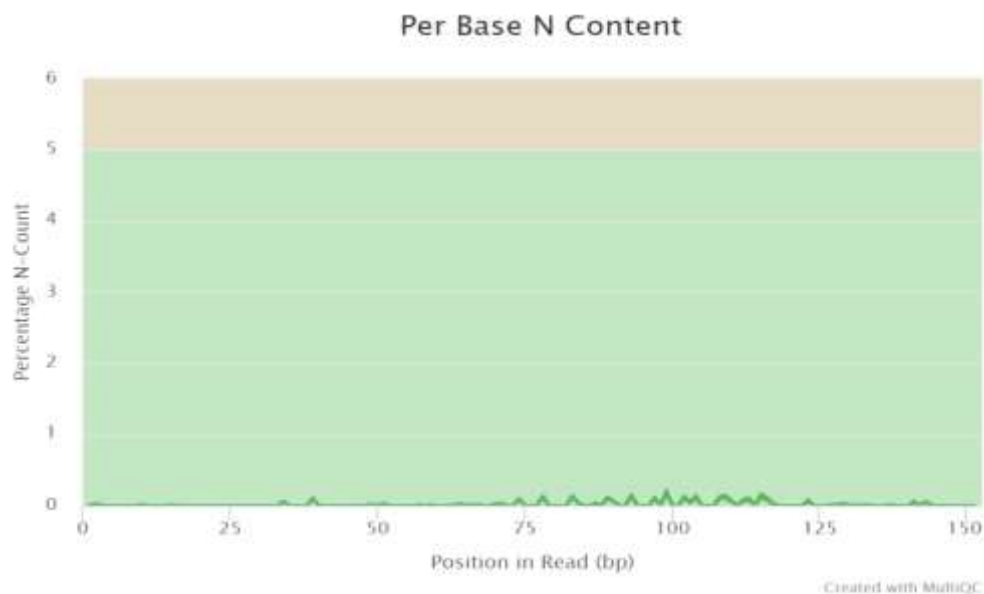


Figura 17. FastQC. Por contenido base N

Nota. Admera health

Este figura indica el contenido de bases que tienen una N (base no asignada), en las muestras se observó que el contenido de N es menor al 5 %.

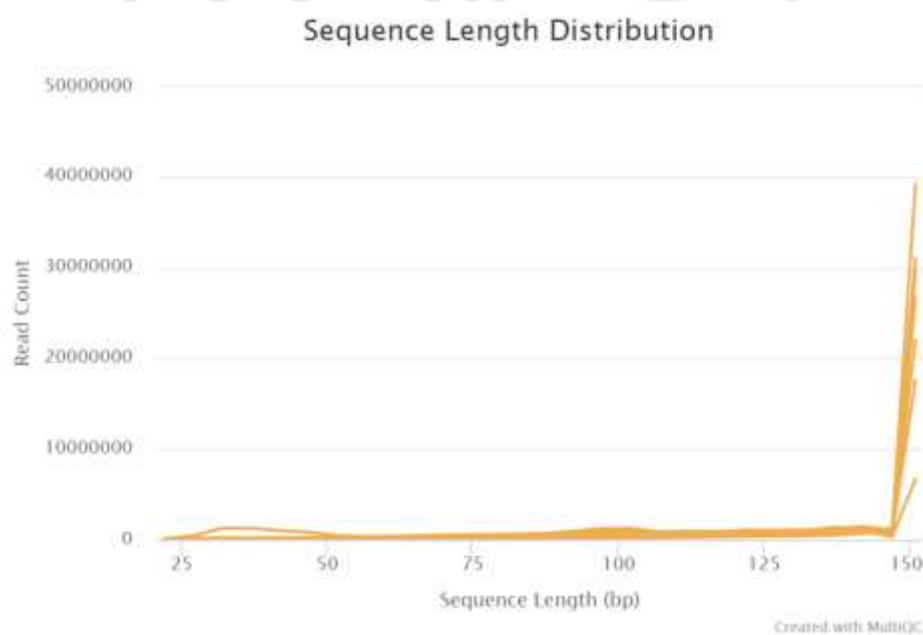


Figura 18. FastQC. Distribución de longitud de secuencia

Nota. Admera health

La figura muestra la distribución de longitud uniforme de los fragmentos de las siete muestras analizadas. A raíz de esto se observó que las secuencias son homogéneas las 25000 y presentan aproximadamente 150 pares de bases.

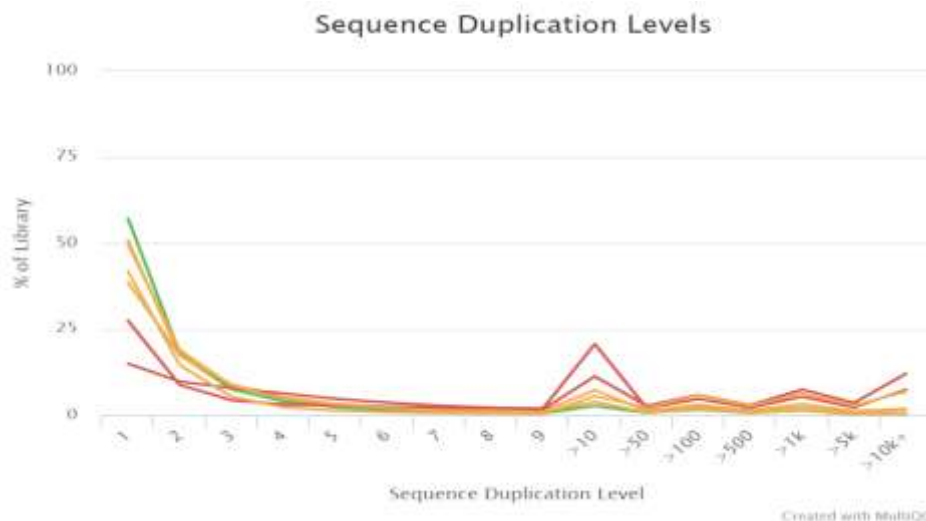


Figura 19. FastQC. Nivel de duplicación de secuencias

Nota. Admera health

La figura evidencia el número relativo de secuencias con distintos grados de duplicación.

En el caso de la secuencia completa de las siete muestras se puede observar 3 picos:

>10: en este caso entre un 5-22 % de secuencias tienen el mismo fragmento desde el principio hasta el final 10 veces.

>100: un 5 % de ellas tienen el mismo fragmento desde el principio hasta el final 100 veces.

>1K: un 7.5 % de secuencias repetidas, es decir, tienen el mismo fragmento desde el principio hasta el final 1 000 veces.

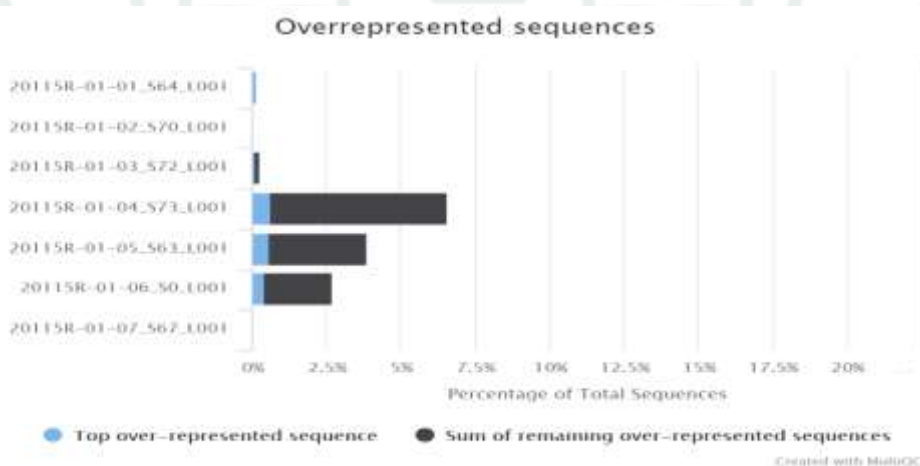


Figura 20. FastQC. Secuencias sobrerrepresentadas

Nota. Admera health

Este figura manifiesta la valoración de la cantidad de secuencias sobrerrepresentadas que se encontraron al mapear. En las muestras 4, 5 y 6 se observan que las secuencias sobrerrepresentadas exceden el 1 % del porcentaje del total de secuencias.

Tabla 5. *RNAseq: Conteo normalizado de los 50 genes más expresados en muestras 1,2,3 de pterigion (Parte 1 de 2)*

MUESTRA 1			MUESTRA 2		MUESTRA 3	
	CONTEO	GEN	CONTEO	GEN	CONTEO	GEN
1	54140.5968	RPS29	22583.45504	PCBD2	42602.62998	RPS29
2	34422.24981	EPHB1	19416.44463	RPS29	34517.00716	EPHB1
3	28731.30356	AHNAK	19322.70647	RBFOX1	30830.22883	KRT13
4	22359.1409	MTRNR2L12	18306.46588	KAZN	22833.37479	PCBD2
5	21064.27974	NFYB	17878.69256	DLG2	21974.42505	KAZN
6	20186.63415	KAZN	17861.58163	MACROD2	20286.54091	AHNAK
7	17293.62516	RBFOX1	16208.51677	PTPRD	20003.63032	KRT4
8	17150.62472	PDE4D	16140.81699	PDE4D	18682.95518	MTRNR2L12
9	15735.54889	MACROD2	16137.84117	CNTNAP2	16590.82183	RBFOX1
10	15651.4772	PTPRD	15630.46483	CSMD1	16563.36098	KRT5
11	15044.11817	DLG2	14308.45931	NRXN3	12897.65725	DLG2
12	14520.04512	CNTNAP2	13978.14392	MTRNR2L12	12416.13449	MACROD2
13	13879.68598	CSMD1	13418.69082	FHIT	11988.89481	PTPRD
14	13368.97011	NRXN3	13356.94267	KRT12	11797.94613	PDE4D
15	13225.96967	MTRNR2L8	13320.48895	AHNAK	11584.6456	CSMD1
16	12982.39748	PCBD2	13104.74241	LRP1B	11548.88264	CNTNAP2
17	12663.39649	KRT13	13055.64147	WWOX	11446.70274	MUC16
18	12471.68161	WDR74	12666.55375	KRT5	11439.67787	DMD
19	12156.6092	DMD	12646.467	EYS	10976.67522	NRXN3
20	12010.46589	FHIT	12601.08583	DAB1	10928.7784	RORA
21	11850.17968	MUC16	12508.09163	ROBO2	10756.34982	FAM234B
22	11734.67932	LRP1B	11825.14224	CTNNA2	10171.36993	WWOX
23	11662.39338	ROBO2	11659.98454	CDH13	10004.68897	DAB1
24	11637.25044	DAB1	11593.02872	AGBL4	9892.291087	CROCC
25	11183.89189	EYS	11401.0887	RORA	9661.747699	PLEC

Nota. elaboración propia.

En la parte 1 de la Tabla 5 se observan las frecuencias de transcripción de los genes del puesto 1 al puesto 25 expresados en las muestras 1, 2 y 3. Se puede ver que el gen RPS29 es el más expresado en las muestras 1 y 3. *Nota:* elaboración propia.

Tabla 6. RNAseq: Conteo normalizado de los 50 genes más expresados en muestras 1,2,3 de pterigion (Parte 2 de 2)

MUESTRA 1			MUESTRA 2		MUESTRA 3	
	CONTEO	GEN	CONTEO	GEN	CONTEO	GEN
26	11133.60602	RORA	11125.08192	PCDH15	9622.791614	CDH13
27	10964.67693	CTNNA2	11108.71494	LRMDA	9600.439762	RUNX1
28	10952.10546	ZBTB20	11049.19865	PRKN	9586.390027	FHIT
29	10831.10508	KRT5	11028.36795	CTNNA3	9456.11066	RCC1
30	10708.53327	LRMDA	10858.74653	MAGI2	9346.905898	KIAA1217
31	10690.46179	WWOX	10457.0116	GRID2	9240.894257	ROBO2
32	10240.2461	AGBL4	10386.33601	LSAMP	9015.459863	AGBL4
33	9872.530676	CDH13	10276.23088	GPC5	8935.63182	FOXN3
34	9848.173457	PCDH15	10202.57947	ZBTB20	8896.675735	LRP1B
35	9825.387672	SERHL2	10073.8755	TBC1D5	8869.214888	LRMDA
36	9810.459054	RUNX1	10067.92387	RUNX1	8855.803777	AUTS2
37	9798.673304	PRKN	10038.90968	ANKS1B	8779.80748	EYS
38	9742.101699	DST	10015.84712	DST	8743.405893	CTNNA2
39	9742.101699	MAGI2	9960.050601	AUTS2	8581.195309	PRKN
40	9702.030146	KIAA1217	9845.481749	KIAA1217	8514.139753	DSP
41	9559.029702	ANK3	9820.931281	NAALADL2	8286.150862	ZBTB20
42	9538.601067	TBC1D5	9811.259884	DPP10	8219.095306	ASIC2
43	9496.95808	LSAMP	9657.261492	GPC6	8014.735516	PARD3
44	9478.100879	PARD3B	9654.285678	PARD3B	7875.515409	EXT1
45	9453.74366	EEF1A1	9579.146366	DSP	7804.628107	PARD3B
46	9441.957909	CTNNA3	9572.450783	CNTN4	7798.241864	SLIT3
47	9390.886322	ERBB4	9535.997058	CCSER1	7784.192128	MAGI2
48	9390.886322	ADARB2	9535.997058	KCNIP4	7740.127048	TBC1D5
49	9390.100605	NAALADL2	9518.886125	ASIC2	7699.893715	ZFHX3
50	9300.528899	GRID2	9518.886125	LRRC4C	7521.078898	OPCML

Nota. elaboración propia.

En la parte 2 se evidencia las frecuencias de transcripción de los genes del puesto 26 al puesto 50 en las muestras 1, 2 y 3. Se puede observar que la expresión de genes es muy similar en las tres muestras.

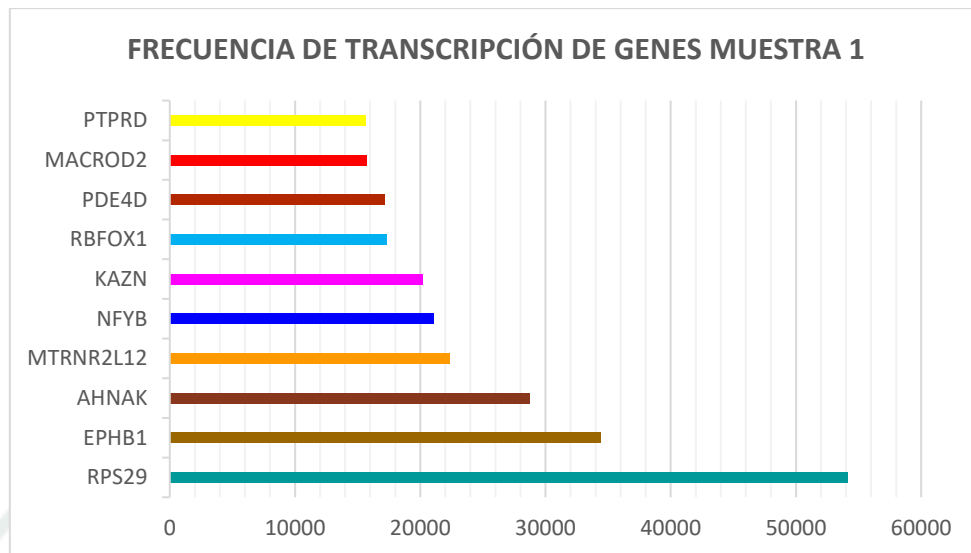


Figura 21. 10 Genes con mayor frecuencia de expresión en la muestra 1 de pterigión

Nota. Elaboración propia.

En la figura se observa que el gen que se expresa con mayor frecuencia es el RPS29.

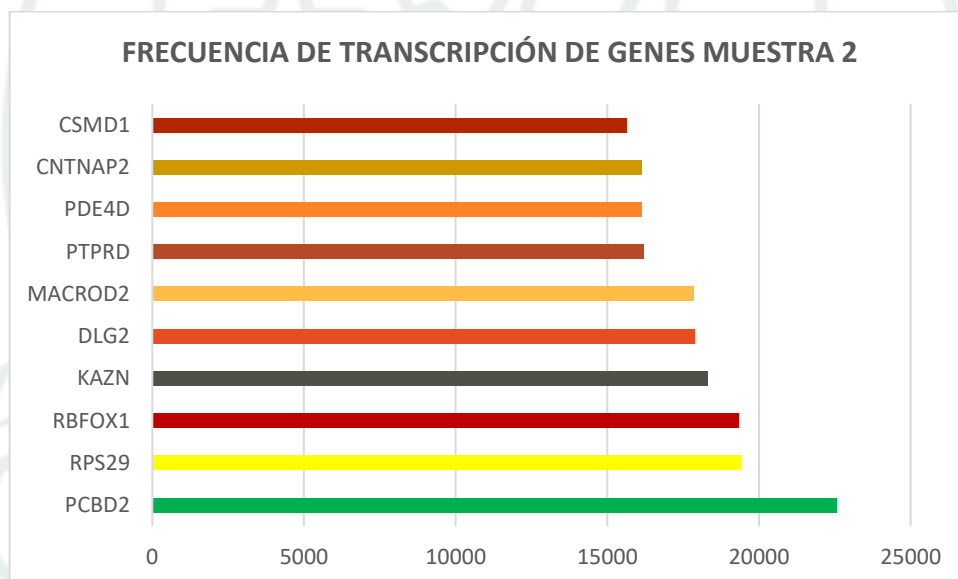


Figura 22. 10 Genes con mayor frecuencia de expresión en la muestra 2 de pterigión

Nota. Elaboración propia.

Se evidencia que el gen que se expresa con mayor frecuencia es el gen PCBD2

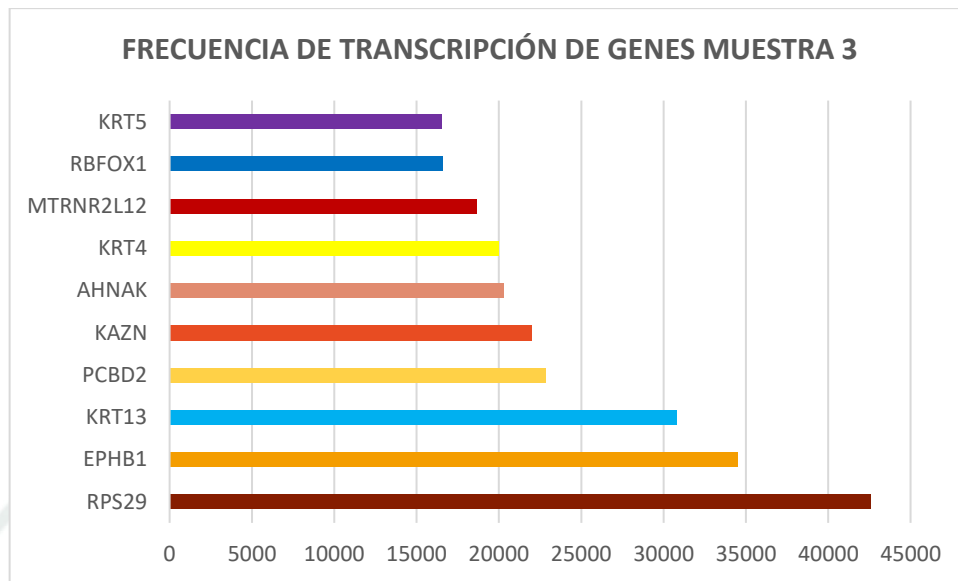


Figura 23. 10 Genes con mayor frecuencia de expresión en la muestra 3 de pterigión

Nota. Elaboración propia.

Se observa que el gen que se expresa más frecuentemente en la muestra 3 es el gen RPS29

Tabla 7. RNAseq: Conteo normalizado de los 50 genes más expresados en muestras 4,5,6,7 de pterigion (Parte 1 de 2)

MUESTRA 4			MUESTRA 5		MUESTRA 6		MUESTRA 7	
	CONTEO	GEN	CONTEO	GEN	CONTEO	GEN	CONTEO	GEN
1	225932.735	RPS29	660016.7799	RPS29	985607.0949	RPS29	34665.6478	RPS29
2	176097.724	EPHB1	616581.283	EPHB1	375320.2599	MTRNR2L12	20999.5682	RBFOX1
3	143276.406	MTRNR2L12	219475.039	NFYB	320258.2038	EPHB1	18477.2423	MTRNR2L12
4	124596.54	PCBD2	169149.6045	MTRNR2L12	213984.0771	PCBD2	17040.9624	PCBD2
5	80347.2049	AHNAK	115913.3893	CROCC	116102.3313	CROCC	14858.1598	MACROD2
6	56758.948	IGHG1	80831.39663	WDR74	106031.4741	ANKRD28	14853.8756	KAZN
7	41232.0224	DHFR	76233.56245	PCBD2	104447.277	MTRNR2L8	14817.9953	DLG2
8	40833.72	CHD8	64503.22557	RCC1	75492.78982	DHFR	14589.3259	CSMD1
9	37524.5482	SERHL2	61401.66345	MTRNR2L8	73010.62337	PTPRD	14564.1562	PTPRD
10	36886.7494	MUC16	57166.31412	DHFR	66292.8191	SNX6	14458.1221	CNTNAP2
11	33990.473	MTRNR2L8	55912.60712	AHNAK	58436.00963	RCC1	13702.4954	PDE4D
12	30712.204	ARMH4	55552.84771	SERHL2	51494.9078	AHNAK	13494.1759	AHNAK
13	29855.5105	IGKC	54525.35306	ARMH4	50252.27901	WDR74	13278.359	NRXN3
14	27597.8914	ANKRD28	53576.89645	SRP54	41243.22032	NEMF	11943.2935	DAB1
15	27297.4478	ZBTB20	37943.71515	EIF4A2	34850.79164	EEF1A1	11124.4748	CROCC
16	25025.2357	DST	35670.68983	KRT12	26974.6271	NBPF1	11008.2657	FHIT
17	24033.7718	EEF1A1	34842.15303	PTPRD	26163.20905	IGHA1	10916.6908	LRP1B
18	23094.6709	KAZN	32133.05571	RPS3	25645.44706	EIF4A2	10896.3408	ROBO2
19	22890.3693	WDR74	30525.0402	KRT5	24688.74654	TPT1	10826.7225	AGBL4
20	22282.6148	RCC1	30473.25665	MED12L	24571.28411	EGLN3	10791.3778	CTNNA2
21	22109.2159	EIF4A2	27832.29558	KRT13	21524.98892	RPL13A	10682.1306	EYS
22	20886.8396	RPLP0	27535.22153	SHF	21415.25429	KRT13	10550.9268	CDH13
23	20121.996	TCP1	26984.68063	EGLN3	21144.78161	RPS3	10524.1505	ASIC2
24	18448.0959	CD74	26540.43227	IGHA1	19979.43073	KRT5	10305.1205	WWOX
25	18094.4309	B2M	26077.10577	MUC16	18905.26779	KIF5C	10252.639	PRKN

Nota. Elaboración propia.

En la parte 1 de la Tabla 6, se exponen las frecuencias de transcripción de los primeros 25 genes expresados en las muestras 4, 5, 6 y 7 de pterigion. El gen RPS29 es el de mayor cantidad de transcripciones en dichas muestras. *Nota:* elaboración propia.

Tabla 8. RNAseq: Conteo normalizado de los 50 genes más expresados en muestras 4,5,6,7 de pterigion (Parte 2 de 2)

MUESTRA 4			MUESTRA 5		MUESTRA 6		MUESTRA 7	
	CONTEO	GEN	CONTEO	GEN	CONTEO	GEN	CONTEO	GEN
26	17898.7133	SPAG8	25368.48877	DSP	18211.31216	EAPP	10210.3325	LRMDA
27	17811.1555	ERCC1	25215.86357	NBPF1	18166.49097	SRP54	10036.8221	KRT13
28	17797.4209	CCIN	25085.04196	ANKRD28	17947.02171	SERHL2	10002.013	RUNX1
29	17222.286	KIF5C	24877.90776	TPT1	17939.29392	MTRNR2L1	9827.96713	AUTS2
30	16525.2569	KRT4	24847.92781	DST	17368.98295	KRT4	9509.32935	OPCML
31	16429.9733	ACTB	24128.40901	CLU	16715.21183	KRT14	9400.61764	PCDH15
32	16267.7338	CFH	23294.4213	KIF5C	16600.84053	RPS2	9372.23478	CTNNA3
33	15370.695	RORA	23250.8141	PARP16	15702.87122	KIAA0825	9321.89537	MAGI2
34	15333.7833	IGFBP5	23239.9123	CDH1	14471.06134	IGKC	9291.90593	NTM
35	15281.4203	RPL13A	23231.73595	EEF1A1	14015.12167	HBB	9127.49955	LSAMP
36	15224.7652	PDE4D	23019.15085	ERCC1	13889.93146	IGFBP5	9111.96931	PTPRT
37	15012.7379	CLU	22986.44545	FAM234B	13452.53849	CLU	9056.81016	ANKS1B
38	14107.115	UTRN	22239.67215	IGKC	13291.80044	PXN	8939.53004	ZFH3
39	13854.7424	KRT13	21806.32559	IGHG1	13237.7059	VIM	8921.85769	KCNIP4
40	13782.6359	MACF1	20773.38004	RPL13A	13100.15122	ACTB	8768.16182	GPC5
41	13399.7849	SPTBN1	18911.89767	ACTB	12438.65232	RPL3	8716.21583	RORA
42	13369.7406	DSP	18279.59327	PXN	12127.99512	SPTBN1	8686.2264	TBC1D5
43	13324.2448	ANK3	18143.32077	IPO7	12010.5327	DSP	8630.53173	GRID2
44	13252.9968	MAML2	18058.83182	FOXN3	12004.35047	RPL19	8553.95156	DPP10
45	12814.3491	RPL23A	17987.97012	TCP1	11653.50876	CDH1	8407.21752	NAALADL2
46	12680.4371	RPS3	17824.44312	TGFBI	11599.41422	RPS6	8342.95444	PARD3B
47	12524.2064	MED12L	17491.93821	KAZN	11580.86752	CHD8	8338.67024	ROBO1
48	12475.277	ASPH	16723.36131	RPS2	11333.57821	KAZN	8328.49525	ASTN2
49	12100.1517	CROCC	16622.51966	KRT14	11290.30258	PARP16	8299.57686	KIAA1217
50	12082.9835	NFIA	16461.71811	BAZ1A	11239.29916	KRT19	8232.63615	ZBTB20

Nota. Elaboración propia.

En la parte 2 de la Tabla 6, se observan las frecuencias de transcripción de los primeros 25 genes expresados en las muestras 4, 5, 6 y 7 de pterigion. Esto manifiesta que en las cuatro muestras es muy similar la expresión de genes.

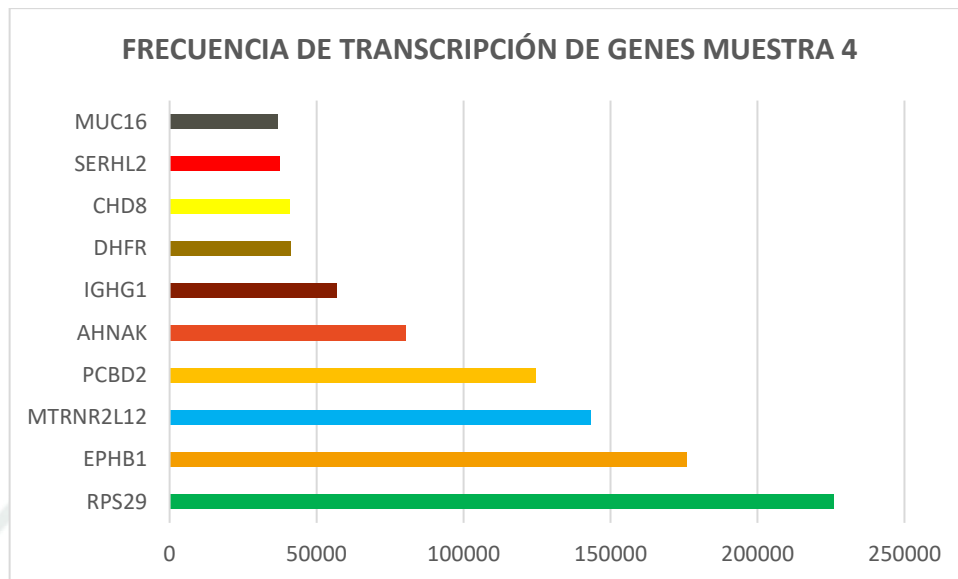


Figura 24. 10 Genes con mayor frecuencia de expresión en la muestra 4 de pterigion

Nota. Elaboración propia.

En la figura se observa que el gen que se expresa con mayor frecuencia es el RPS29.

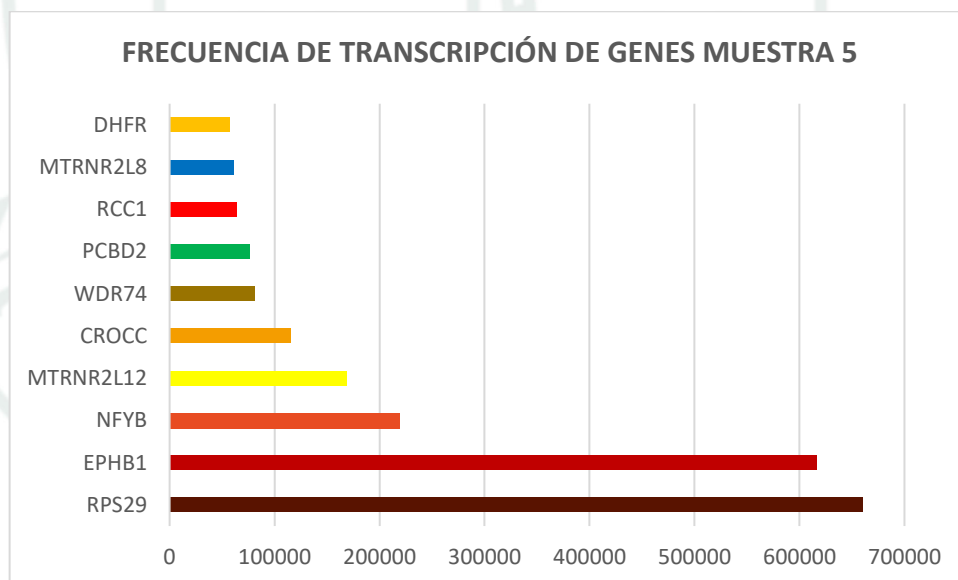


Figura 25. 10 Genes con mayor frecuencia de expresión en la muestra 5 de pterigión

Nota. Elaboración propia.

En la figura se muestra que el gen que tiene mayor frecuencia de expresión es el RPS 29.

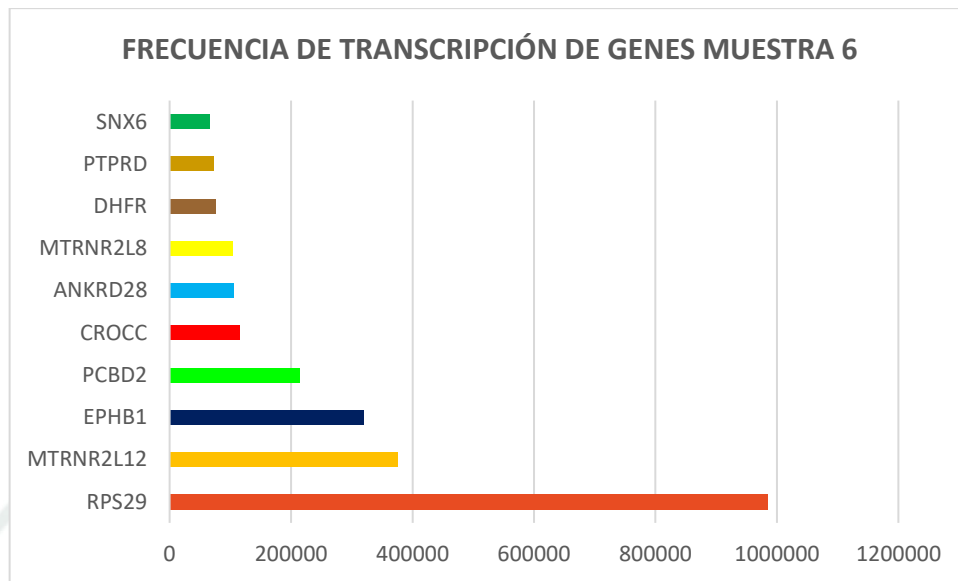


Figura 26. 10 Genes con mayor frecuencia de expresión en la muestra 6 de pterigión

Nota. Elaboración propia.

El gen RPS29 es el que más se expresa en la muestra 6 de pacientes con pterigión.

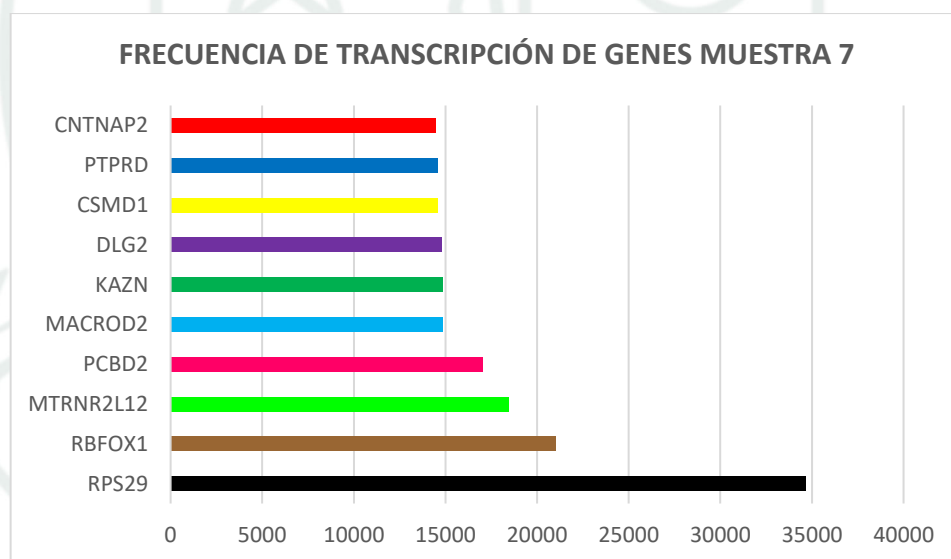


Figura 27. 10 Genes con mayor frecuencia de expresión en la muestra 7 de pterigión

Nota. Elaboración propia.

Se observa en la figura que el gen con más frecuencia se expresa en la muestra 7 de pterigión es el gen RPS 29.

Tabla 9. *Análisis funcional de los genes mapeados*

Término	Conteo	%
Variante de empalme	18	62,1
Constituyente estructural del citoesqueleto	3	10,3
Unión al ADN	3	10,3
Unión de iones de zinc	3	10,3
Unión de iones metálicos	3	10,3
Guía del axón	3	10,3
Desarrollo cerebral	3	10,3
Angiogénesis	3	10,3
Región de composición sesgada: rica en gly	3	10,3
Proteína de unión al factor de crecimiento similar a la insulina, N-terminal	3	10,3
Regulación positiva de la transcripción, plantilla de ADN	3	10,3
Adhesión celular	4	13,8
Enlace proteico	19	65,5
Organización del citoesqueleto	4	13,8
Sitio de mutagénesis	6	20,7
Transcripción, plantilla de ADN	4	13,8
Regulación de la transcripción, plantilla de ADN	4	13,8

Nota. Elaboración propia.

Las funciones de los genes mapeados y evaluados en la base de datos DAVID son diversas como se puede observar en la tabla. Estos roles van desde variantes de empalme hasta regulación en la transcripción.

Tabla 10. *Nuevos genes asociados a pterigion*

Símbolo	N°	Especie	Nombre
EIF4A2	1974	Homo sapiens	eukaryotic translation initiation factor 4A2(EIF4A2)
ROBO2	6092	Homo sapiens	roundabout guidance receptor 2(ROBO2)
MTRNR2L12	100463498	Homo sapiens	MT-RNR2-like 12(MTRNR2L12)
PCDH15	65217	Homo sapiens	protocadherin related 15(PCDH15)
MTRNR2L8	100463486	Homo sapiens	MT-RNR2-like 8(MTRNR2L8)
MTRNR2L1	100462977	Homo sapiens	MT-RNR2-like 1(MTRNR2L1)
NBPF1	55672	Homo sapiens	neuroblastoma breakpoint family member 1(NBPF1)
MED12L	116931	Homo sapiens	mediator complex subunit 12 like(MED12L)
RUNX1	861	Homo sapiens	runt related transcription factor 1(RUNX1)
KIAA1217	56243	Homo sapiens	KIAA1217(KIAA1217)
SHF	90525	Homo sapiens	Src homology 2 domain containing F(SHF)
AGBL4	84871	Homo sapiens	ATP/GTP binding protein like 4(AGBL4)
IGHG1	3500	Homo sapiens	immunoglobulin heavy constant gamma 1 (G1m marker)(IGHG1)
TCP1	6950	Homo sapiens	t-complex 1(TCP1)
PCBD2	84105	Homo sapiens	pterin-4 alpha-carbinolamine dehydratase 2(PCBD2)
IGHA1	3493	Homo sapiens	immunoglobulin heavy constant alpha 1(IGHA1)
CSMD1	64478	Homo sapiens	CUB and Sushi multiple domains 1(CSMD1)
TPT1	7178	Homo sapiens	tumor protein, translationally-controlled 1(TPT1)
EYS	346007	Homo sapiens	eyes shut homolog (Drosophila)(EYS)

Nota. Elaboración propia.

Los 19 genes de la tabla se reportan por primera vez relacionados con pterigion, no existen precedentes en la literatura al respecto.

Tabla 11. Ubicación de las mutaciones puntuales del gen P 53, comunes en los 7 casos de pterigion

Mutación	Referencia Bibliográfica
Tipo de variante: SNV Alelos: C > G Cromosoma: 17: 7676154 (GRCh38) Exón 4	Mutación de novo
Tipo de variante: SNV Alelos: A > C Cromosoma: 17: 7669372 (GRCh38) Exón 8	Mutación de novo
Tipo de variante: SNV Alelos: C > T Cromosoma: 17: 7669 124 (GRCh38) Exón 8	Mutación de novo
Tipo de variante: SNV Alelos: G > T, A Cromosoma: 17: 7668836 (GRCh38) Exón 8	Mutación de novo
Tipo de variante: SNV Alelos: C > G, T Cromosoma: 17: 7668799 bp (GRCh38) Exón 8	Mutación de novo

Nota. Elaboración propia.

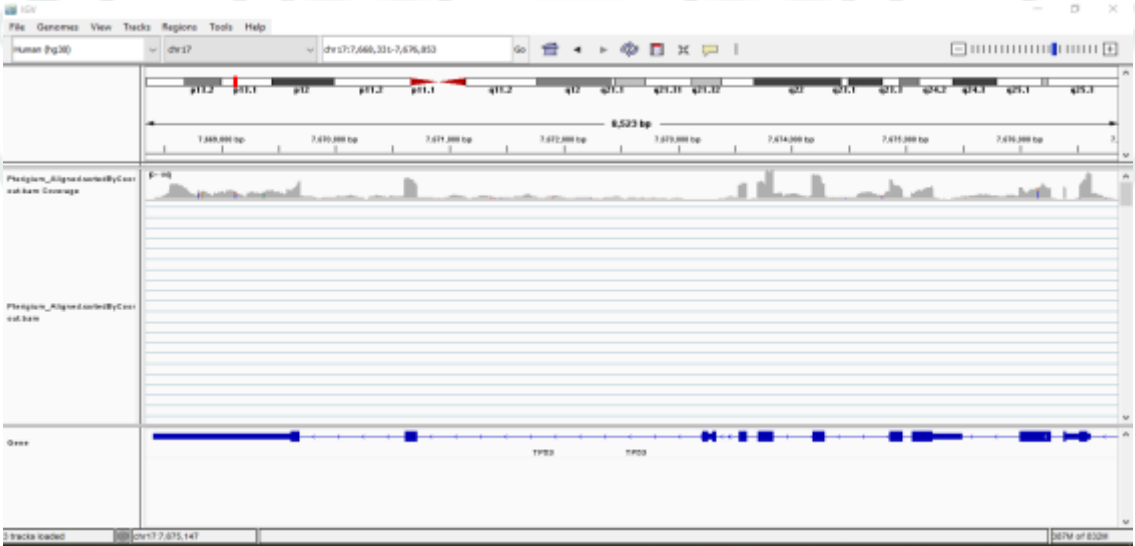
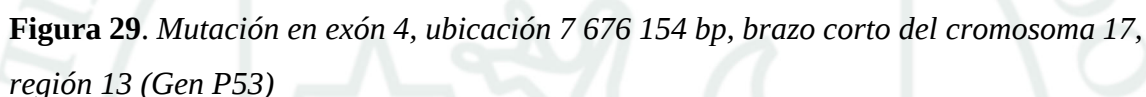


Figura 28. Mutaciones en gen TP 53-IG: Se observa el gen P53 ubicado en el brazo corto del cromosoma 17, región 13.

Nota. Elaboración propia.



Se evidencia la pista de cobertura predeterminada para un archivo de alineación BAM, empleando el *software* IGV (Integrative Genomics Viewer) de la combinación de los 7 pacientes de la investigación. Si un nucleótido no está vinculado con la secuencia de referencia en más del 20 % de las lecturas ponderadas de calidad, el IGV colorea la barra en proporción al recuento de lecturas de cada base (A, C, G, T). Se encontró el polimorfismo de nucleótido único (SNP) Chr 17: 7 676 154, SNV (variación en la secuencia de ADN que ocurre cuando se altera un solo nucleótido) C > G.

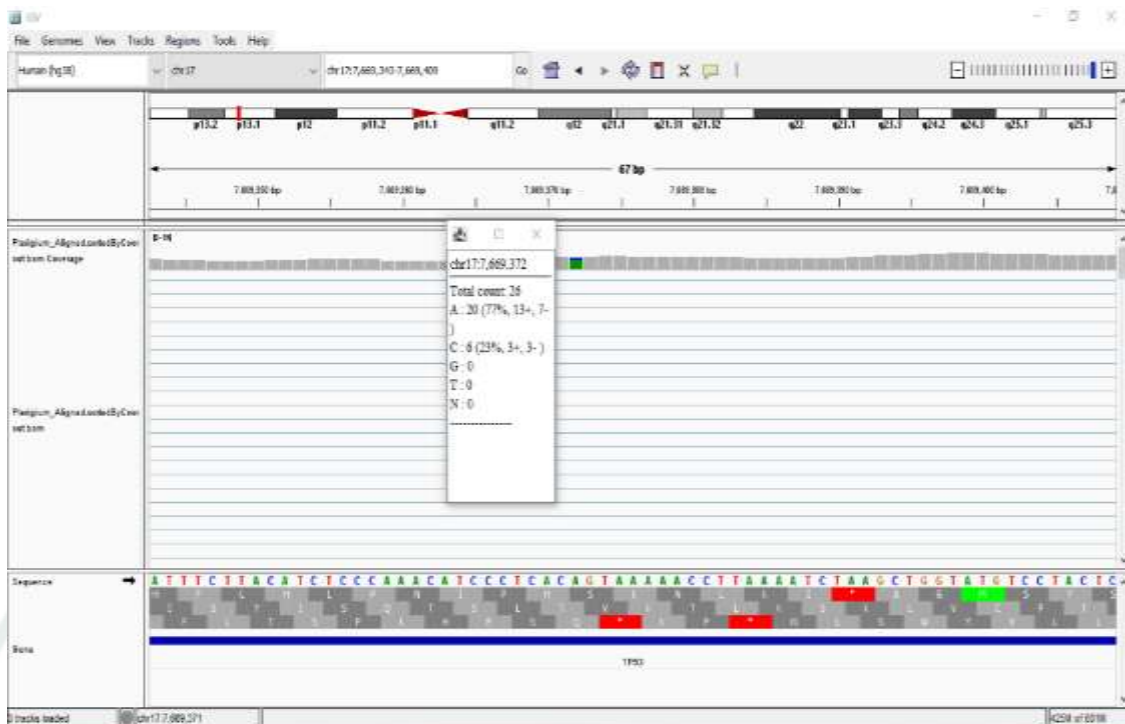


Figura 30. Mutación en exón 8, ubicación 7 669 372 bp, brazo corto del cromosoma 17, región 13 (Gen P53)

Nota. Elaboración propia.

Se evidencia la pista de cobertura preestablecida para un archivo de alineación BAM, empleando el *software* IGV de la combinación de los 7 pacientes del estudio. Si un nucleótido no está sujeto a la secuencia de referencia en más del 20 % de las lecturas ponderadas de calidad, el IGV colorea la barra en proporción al recuento de lecturas de cada base (A, C, G, T). A raíz de esto, se encontró el polimorfismo de nucleótido único (SNP) Chr 17:7 669 372, SNV (variación en la secuencia de ADN que ocurre cuando se altera un solo nucleótido) A > C.

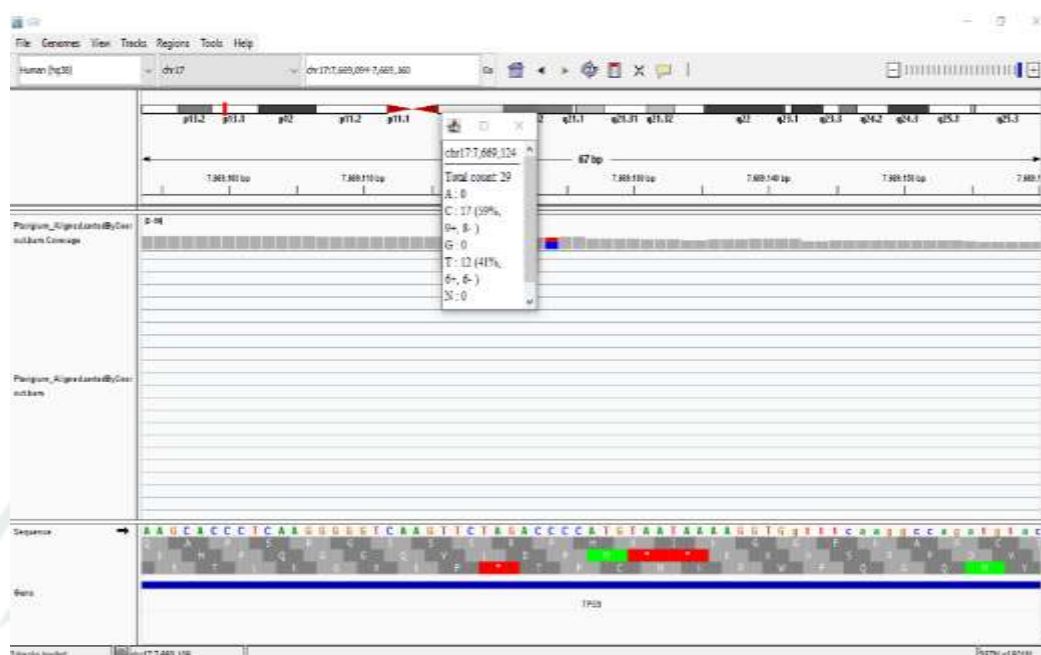


Figura 31. Mutación en exón 8, ubicación 7 669 124 bp, brazo corto del cromosoma 17, región 13 (Gen P53)

Nota. Elaboración propia.

Se observa la pista de cobertura predeterminada para un archivo de alineación BAM, utilizando el *software* IGV de la combinación de los 7 pacientes de la investigación. Si un nucleótido difiere de la secuencia de referencia en más del 20 % de las lecturas ponderadas de calidad, el IGV colorea la barra en proporción al recuento de lecturas de cada base (A, C, G, T). Por lo tanto, se encontró el polimorfismo de nucleótido único (SNP) Chr 17: 7 669 124, SNV (variación en la secuencia de ADN que ocurre cuando se altera un solo nucleótido) C > T.

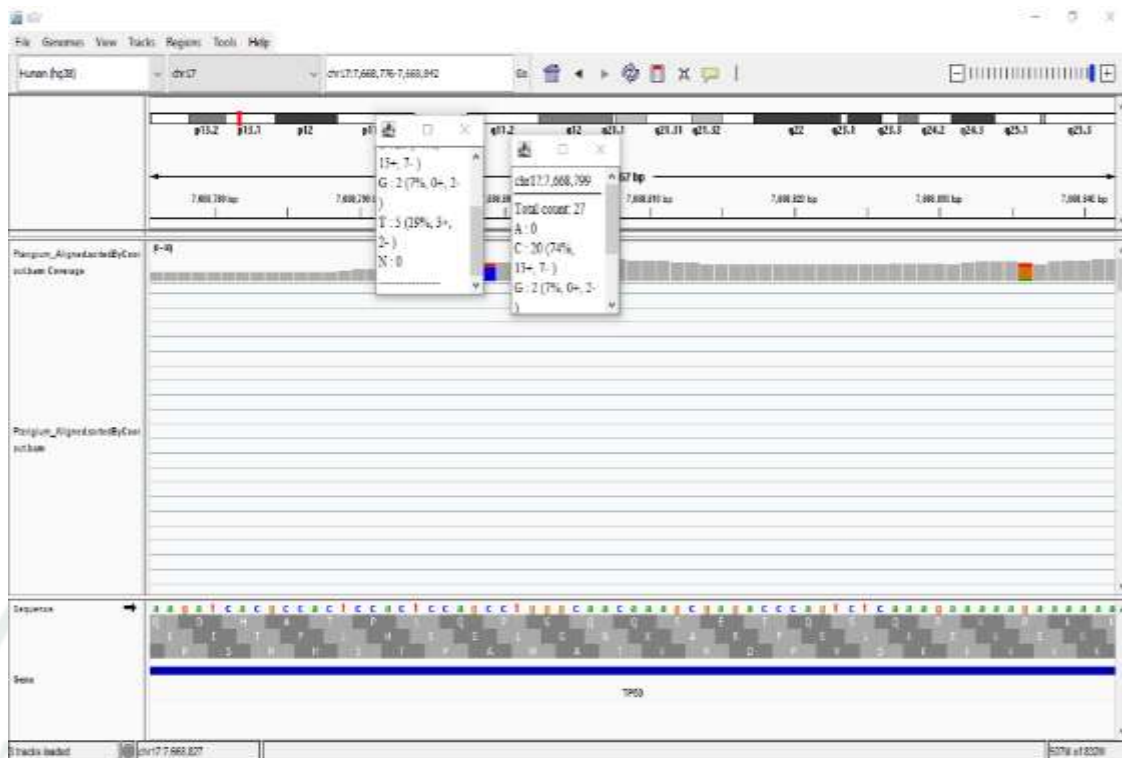


Figura 33. Mutación en exón 8, ubicación 7 668 799 bp, brazo corto del cromosoma 17, región 13 (Gen P53)

Nota. Elaboración propia.

Se observa la pista de cobertura preestablecida para un archivo de alineación BAM, empleando el *software* IGV de la combinación de los 11 pacientes del estudio. Si un nucleótido no guarda relación con la secuencia de referencia en más del 20 % de las lecturas ponderadas de calidad, el IGV colorea la barra en proporción al recuento de lecturas de cada base (A, C, G, T). Así, se encontró el polimorfismo de nucleótido único (SNP) BRAF 7: 7 668 799, SNV (variación en la secuencia de ADN que ocurre cuando se altera un solo nucleótido) C > G, T.

Tabla 12. Genes componentes de las traslocaciones tipo fusión encontradas en 4 casos de las 7 muestras de pacientes con pterigion (Parte 1 de 2)

Gen	Función y patologías asociadas	Referencia
CLDND1	Proteína transmembrana que se expresa en el sistema nervioso central de manera predominante, se asocia a leucodistrofia similar a la de Pelizaeus-Merzbacher.	(90)
KIAA2026	Se expresa en el citoplasma en la mayoría de los tejidos, como la retina y testículos.	(90)
RNU1-2	Es un gen snRNA. Entre sus vías relacionadas se halla el ciclo de empalme espliceosómico. Se expresa en el cuerpo calloso, el tejido suprarrenal, el nervio sural y otros 95 tipos de células o tejidos.	(90)
ATP13A2	Participa en el transporte de cationes de metales de transición divalentes. La deficiencia se vincula con paraplejía espástica y síndrome de Kufor-Rakeb.	(90)
SDHB	Es succinato deshidrogenasa [ubiquinona] subunidad de hierro y azufre, se relaciona con la Feocromocitoma-paraganglioma hereditario, Síndrome de Cowden, Síndrome de Carney-Stratakis.	(90)
WIZ	Participa en la regulación positiva de la replicación del ADN del ciclo celular nuclear y la estabilización de proteínas. Se relaciona con el Síndrome de microduplicación, tumor del estroma de los cordones sexuales ováricos.	(90)
WHRN	Participa en la organización y estabilización de la elongación de los esterocilios y el ensamblaje citoesquelético de actina. Las mutaciones se han relacionado con la sordera autosómica recesiva no sindrómica y el síndrome de Usher.	(90)
19:40114634	Gen de novo.	(90)
SNORD3A	Es un ARN nucleolar pequeño y abundante (snoRNA), desarrolla un rol en el procesamiento de los precursores del ARN ribosómico.	(90)
6:164999470	Gen de novo.	
RNA5SP226	Pseudogén. Se expresa en el nervio sural, la sangre, el tendón calcáneo y otros 65 tipos de células o tejidos.	(90)
PSMC3	Es un gen codificador de proteínas. Entre las enfermedades relacionadas se encuentran la sordera, cataratas, deterioro del desarrollo intelectual y polineuropatía y displasia del desarrollo de la cadera.	(90)

Nota. Elaboración propia.

Se muestra la función de 12 genes componentes de las traslocaciones tipo fusión encontradas en 4 casos de los 7 pacientes investigados con pterigion.

Tabla 13. Genes componentes de las traslocaciones tipo fusión encontradas en 4 casos de las 7 muestras de pacientes con pterigion (Parte 2 de 2)

Gen	Función y patologías asociadas	Referencia
NXF1	Las anotaciones de ontología génica (GO) relacionadas con este gen incluyen la unión de ARN y la unión de nucleótidos. Las enfermedades asociadas con NXF1 incluyen el síndrome de Torch y la tuberculosis hepática.	(90)
WNK1	Interviene en la homeostasis iónica y modifica la presión arterial, transportando iones dentro y fuera de la célula. Se ha relacionado con la hipertensión hiperkalémica familiar.	(90)
ZNF79	Es un gen codificador de proteínas. Entre las enfermedades relacionadas se encuentran la distrofia macular con patrón 2 y la anosognosia.	(90)
DOCK8	Es una proteína adaptadora en la señalización del receptor tipo Toll 9 (TLR9) y en la vía de señalización del receptor de linfocitos T y en la reorganización del citoesqueleto. Las personas con deficiencia de DOCK8 tienen una cantidad de células inmunitarias inferior a la normal.	(90)
RF00100	Es un snRNA, no se relaciona con enfermedades.	(90)
19:40114634	Gen de novo.	(90)
SNORD3A	Desempeña un papel en el procesamiento de los precursores del ARN ribosómico. Las enfermedades asociadas son el síndrome tricohepatoentérico 1 y el síndrome tricohepatoentérico 2.	(90)
AC093752.1	Gen de novo.	(90)
RUFY3	Codifica una proteína necesaria para mantener la polaridad neuronal, se la ha implicado en la mediación de la migración e invasión de células de cáncer gástrico por medio de la interacción con la quinasa-1 activada por P21, que fomenta su expresión.	(90)
AC104151.1	Gen de novo.	(90)
CNTNAP4	Pertenece a la familia de las neurexinas, que tiene miembros que funcionan en el sistema nervioso de los vertebrados como moléculas de adhesión celular y receptores. Las mutaciones en este gen pueden estar relacionadas con ciertas enfermedades psiquiátricas.	(90)

Nota. Elaboración propia.

Se muestra la función de 11 genes componentes de las traslocaciones tipo fusión encontradas en 4 casos de los 7 pacientes.

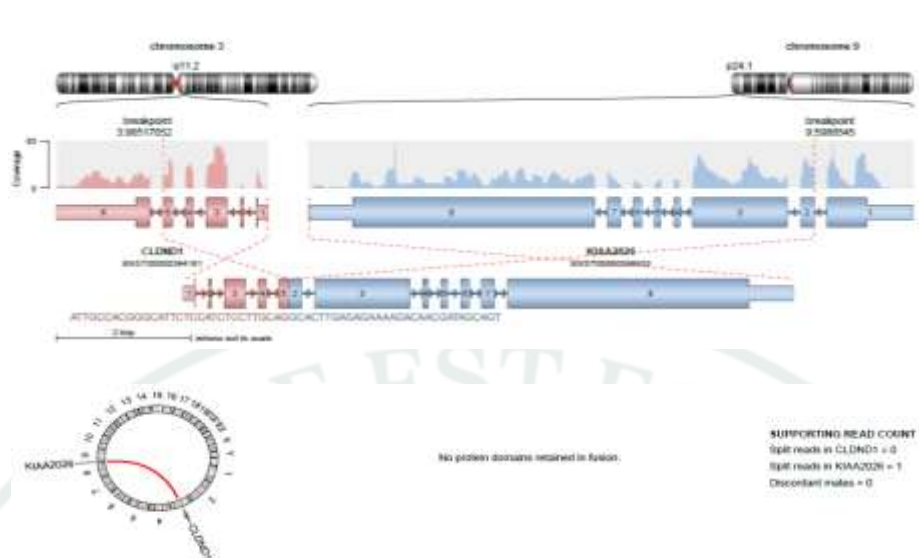


Figura 34. Gen de fusión CLDND1 – KIAA2026, en la muestra 1 de pterigión

Nota. Admera health.

Se puede ver un punto de ruptura a nivel del exón 5 del gen CLDND1 que se encuentra en el cromosoma 3 y ruptura en el exón 2 del gen KIAA2026 que se halla en el cromosoma 9, el cual resulta en el gen de fusión que no retiene dominios proteicos.

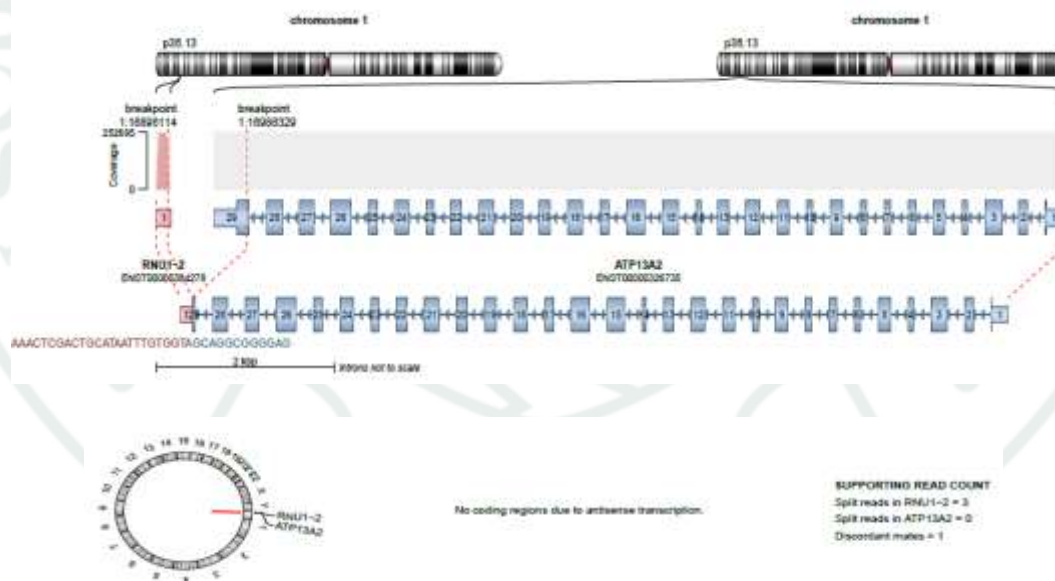


Figura 35. Gen de Fusión RNU1-2 – ATP13A2

Nota. Admera health.

En la muestra 3 de pterigión, el gen de fusión se forma en el cromosoma 1. Cabe mencionar que existe un punto de ruptura en parte del exón 29 de la proteína ATP13A2 que se une con el exón 1 del gen RNU1-2, formando el gen de fusión RNU1-2 - ATP13A2, que no presenta regiones codificantes debido a la transcripción anti sentido.

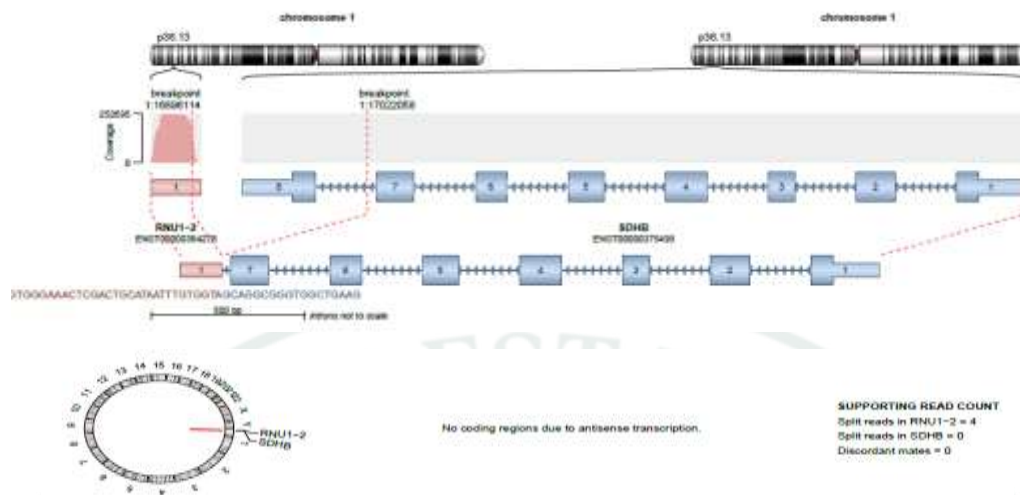


Figura 36. Gen de Fusión RNU1-2 – SDHB, en la muestra 3 de pterigión

Nota. Admera health.

El gen de fusión se forma en el cromosoma 1, por la ruptura en el exón 7 de la proteína SDHB que se une con el exón 1 del gen RNU1-2 y se forma el gen de fusión RNU1-2 – SDHB, el cual no presenta regiones codificantes debido a la transcripción anti sentido.

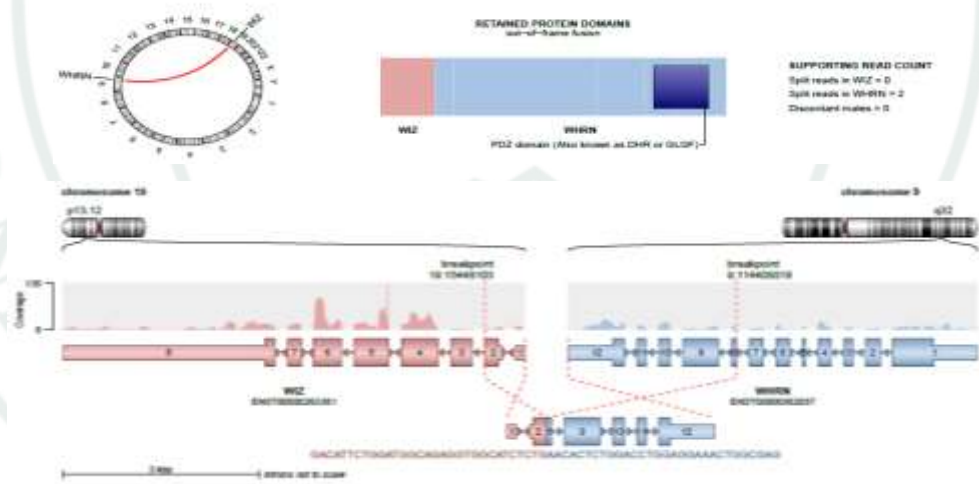


Figura 37. Gen de Fusión WIZ – WHRN, en la muestra 3 de pterigión

Nota. Admera health.

A nivel de la sub sub banda 2, sub banda 1, de la banda 3, y de la región 1 en el brazo corto del cromosoma 19, se presenta ruptura en el exón 2 del gen WIZ, el cual se une al fragmento producido como resultado de la ruptura en el exón 8 del gen WHRN a nivel de la banda 2, región 3 del brazo largo del cromosoma 9, obteniendo un gen de fusión con dominios de proteína retenidos y fusión fuera del marco.

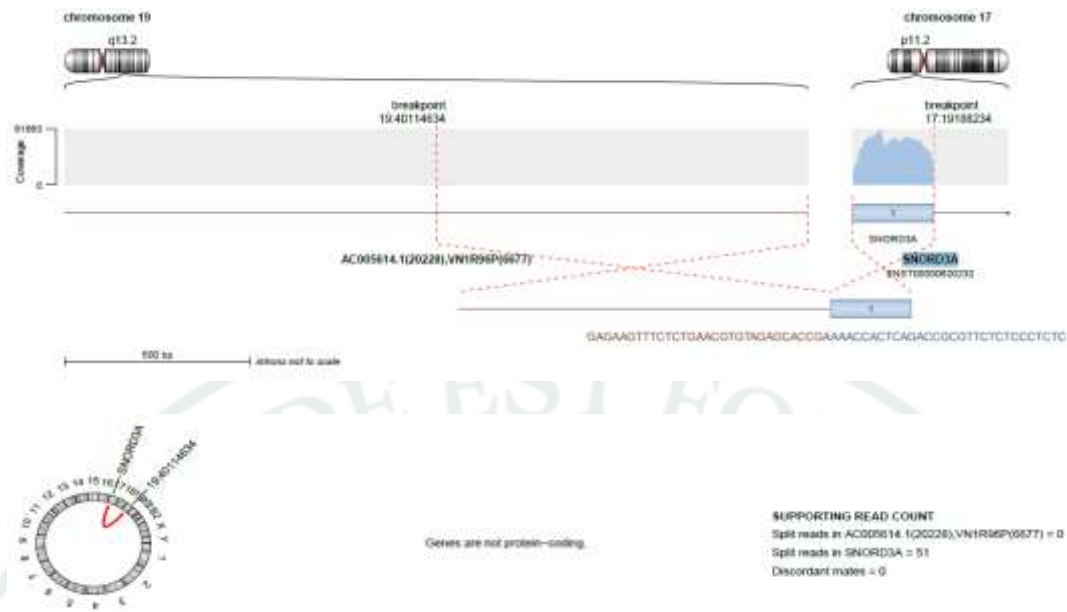


Figura 38. Gen de fusión 19:40114634 - SNORD3A, en la muestra 6 de pterigión

Nota. Admera health.

El gen de fusión se forma por la unión del exoma 1 del gen SNORD3A ubicado en el cromosoma 17, con la proteína derivada de la ruptura a nivel de la sub banda 2, banda 3, región 1 del brazo largo del cromosoma 19. Este gen no codifica proteínas.

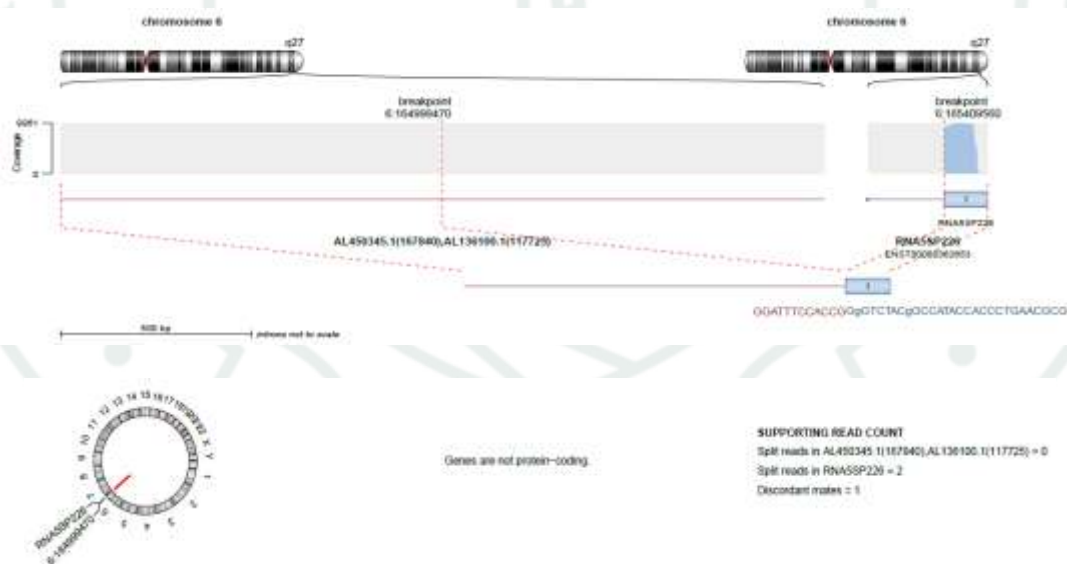


Figura 39. Gen de fusión 6:164999470 - RNA5SP226, en la muestra 6 de pterigión

Nota. Admera health.

El gen de fusión se construye en el cromosoma 6, por la unión del exón 1 del gen RNA5SP226, a partir de las proteínas localizadas en la banda 7, región 2 del brazo largo del cromosoma 6 y no codifican proteínas.

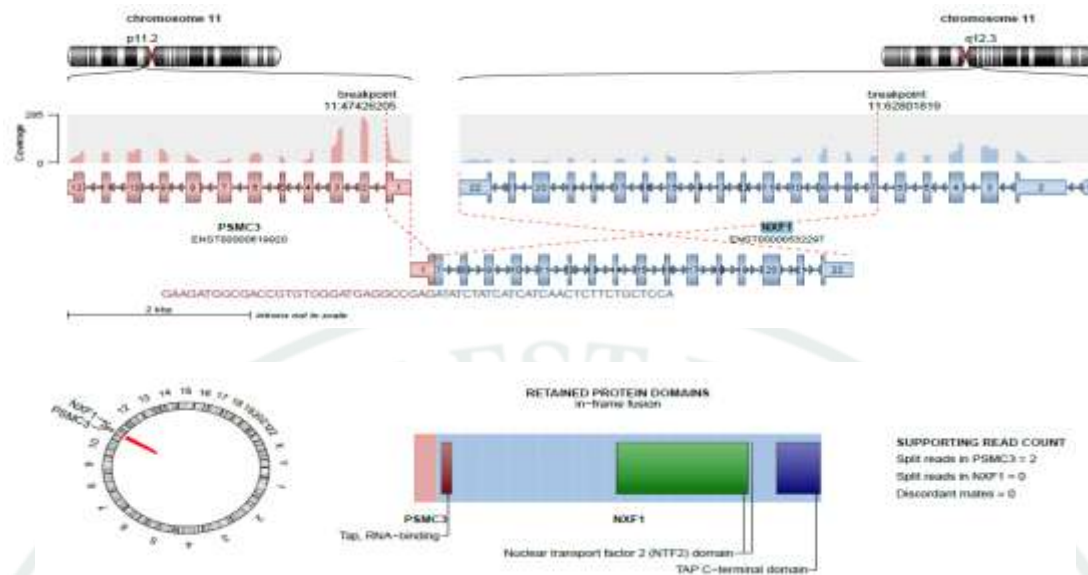


Figura 40. Gen de fusión PSMC3 - NXF1, en la muestra 6 de pterigión

Nota. Admera health.

El gen de fusión PSMC3 - NXF1 se forma en el cromosoma 11, y existe un punto de ruptura en el exón 7 del gen NXF1, el cual se une al segmento que surge a raíz de la ruptura en el exón 1 del gen PSMC3. Este presenta dominios de proteína retenidos y fusión en el marco.

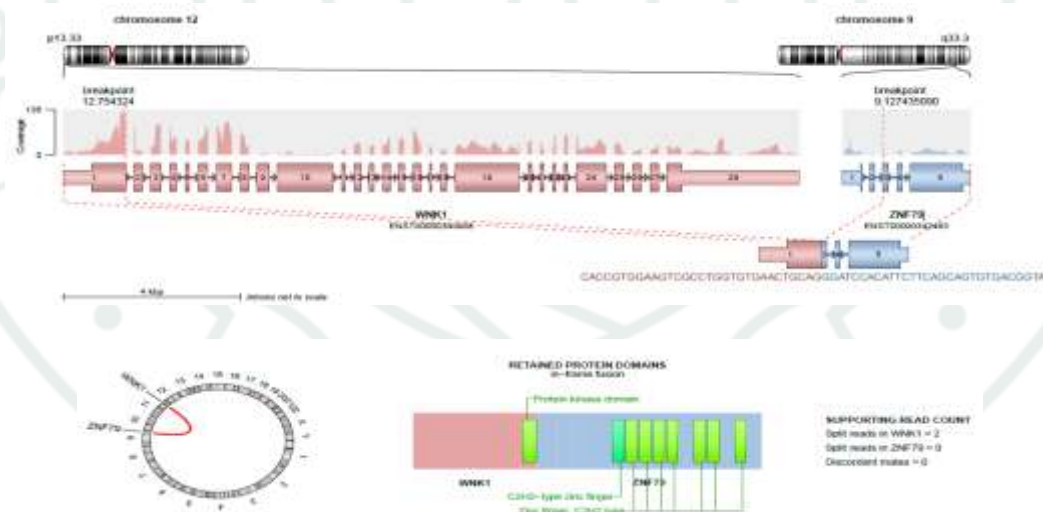


Figura 41. Gen de fusión WNK1 - ZNF79, en la muestra 6 de pterigión

Nota. Admera health.

En el cromosoma 12, el gen presenta una ruptura en el exón 1. Este fragmento se une al que se originó a raíz de un punto de ruptura a nivel del exón 3 del gen ZNF79 en el cromosoma 9. El gen de fusión que se forma presenta dominios de proteína retenidos y fusión en el marco.

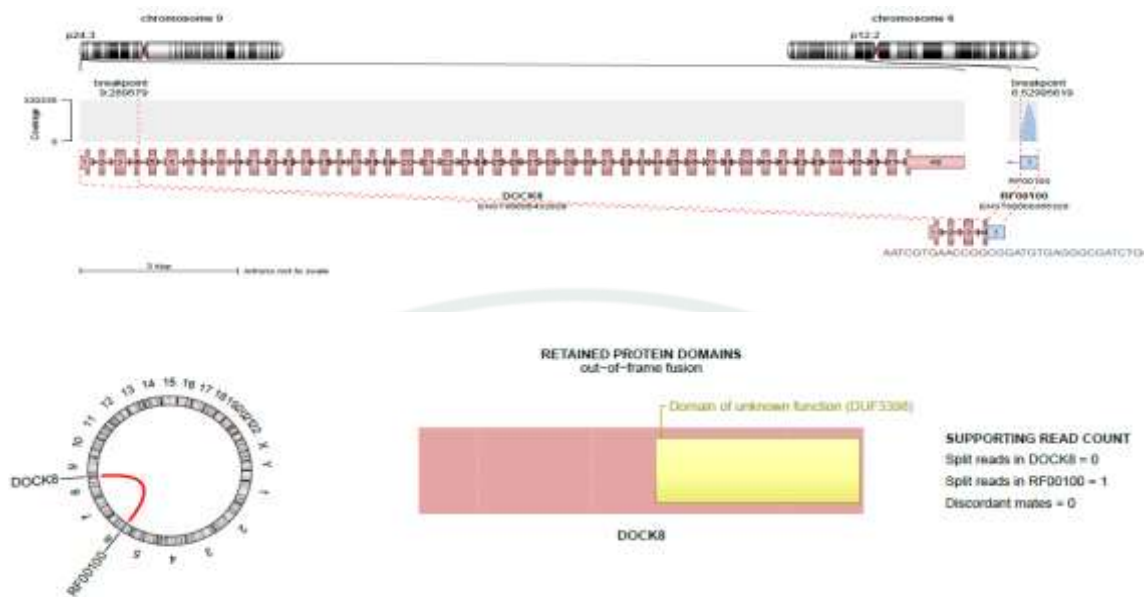


Figura 42. Gen de fusión DOCK8 - RF00100, en la muestra 6 de pterigión

Nota. Admera health.

El gen de fusión DOCK8 - RF00100 se genera a raíz de la unión del exón 1 del gen RF00100 y el segmento formado por una ruptura en el exón 4 de la proteína DOCK8. Esta unión expone dominios de proteína retenidos y fusión fuera del marco.

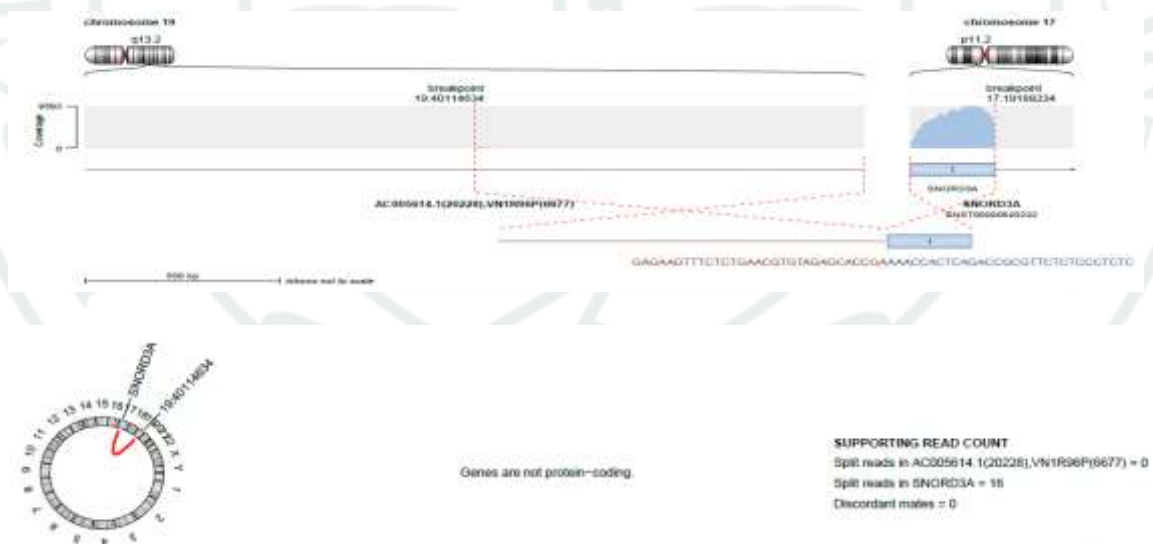


Figura 43. Gen de fusión 19:40114634 - SNORD3A, en la muestra 7 de pterigión

Nota. Admera health.

El gen de fusión se genera a raíz de la unión del exón 1 de SNORD3A, el cual se sitúa en el cromosoma 17 y las proteínas ubicadas en la sub banda 2, banda 3, región 1 del brazo largo del cromosoma 19. Este gen no codifica proteínas.

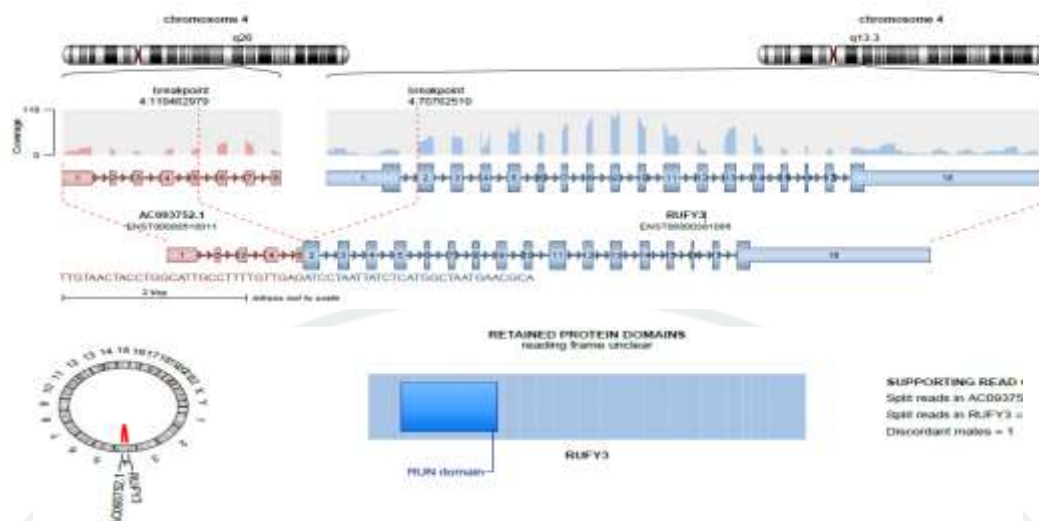


Figura 44. Gen de fusión AC093752.1 - RUFY3, en la muestra 7 de pterigion

Nota. Admera health.

El gen de fusión se formó en el cromosoma 4. En el gen AC093752.1 se presentó un punto de ruptura en el exón 5. Este fragmento se une al formado por la ruptura en el exón 2 del gen RUFY3. Dicha variación estructural presenta dominios de proteína retenidos y en lo relacionado con el marco de lectura, es preciso señalar que es poco claro.

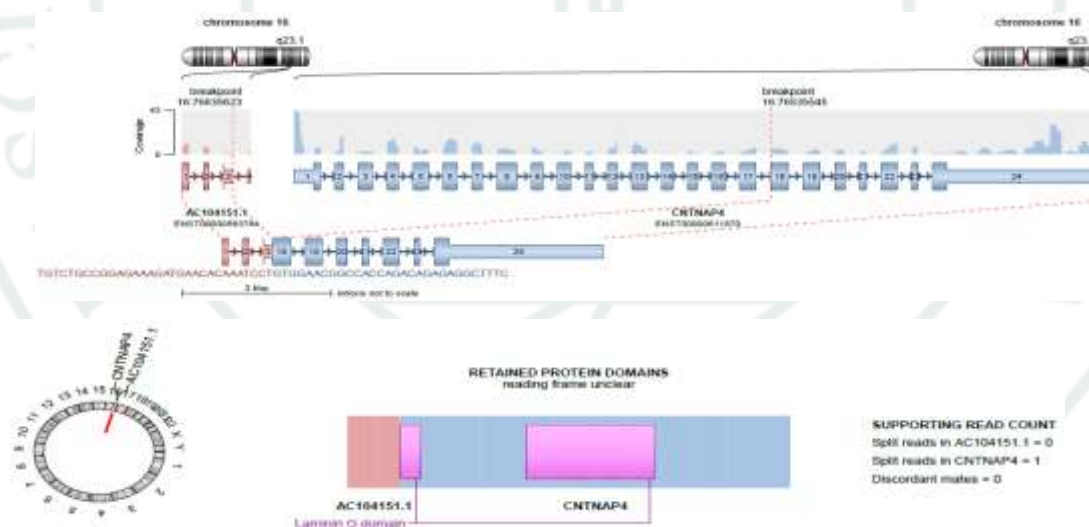


Figura 45. Gen de fusión AC104151.1 - CNTNAP4, en la muestra 7 de pterigión

Nota. Admera health.

El gen de fusión se forma en el cromosoma 16. En el exón 3 del gen AC104151.1 existe un punto de ruptura que permite unirse al fragmento que se genera a raíz de la ruptura en el exón 18 del gen CNTNAP4. Este gen de fusión presenta dominios de proteína retenidos y el marco de lectura es poco claro.

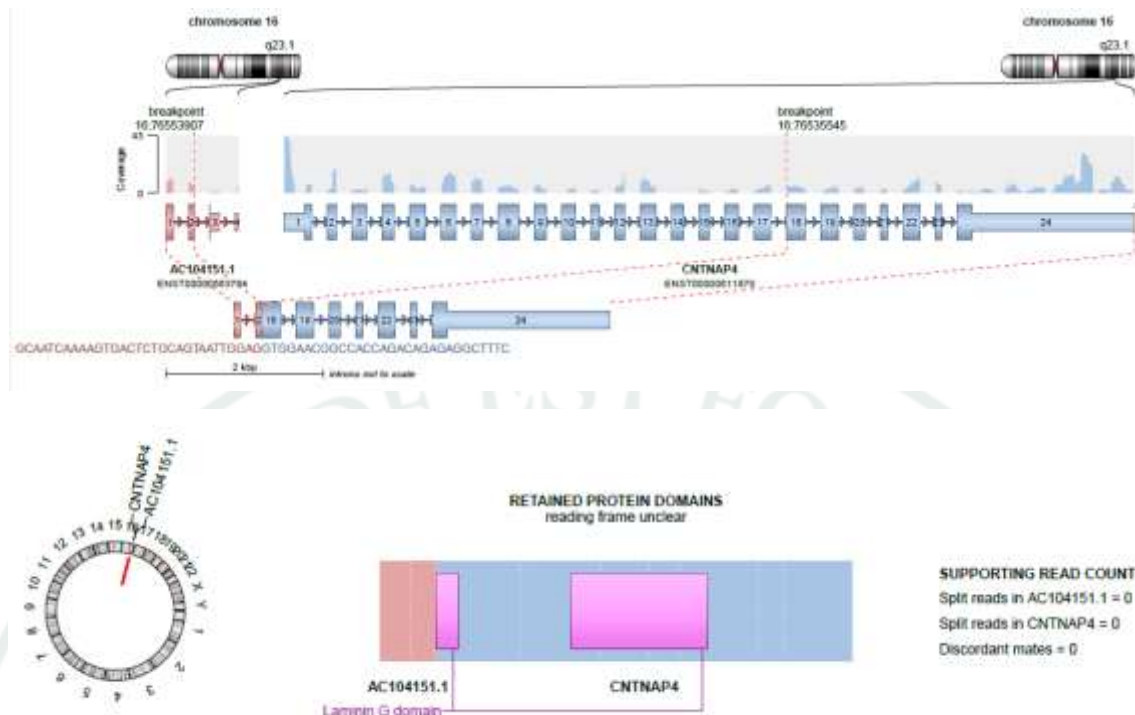


Figura 46. Gen de fusión AC104151.1 - CNTNAP4, en la muestra 7 de pterigion

Nota. Admera health.

El gen de fusión se forma en el cromosoma 16. En el gen AC104151.1 existe un punto de ruptura a nivel del exón 2 y se une al fragmento originado por la ruptura en el exón 18 del gen CNTNAP4. Este gen de fusión evidencia dominios de proteína retenidos y el marco de lectura es poco claro.

DISCUSIÓN

El presente estudio se llevó a cabo con el propósito de identificar cambios del gen humano p53 y hallar genes de fusión en muestras de pterigión primario por la técnica RNA-Seq en pacientes del Hospital III Juliaca EsSalud, 2019-2020.

Esta investigación se desarrolló por la alta prevalencia de pterigion en el sur del Perú, y debido a que, en los últimos años, la incidencia de nuevos casos en pacientes cada vez más jóvenes es preocupante debido a lo invalidante de la enfermedad. Una de las hipótesis es la mutación del gen P 53.

Por esta razón, una de las estrategias para contrarrestar esta enfermedad es la caracterización molecular tanto para el diagnóstico como para las terapias individualizadas de los pacientes portadores de esta enfermedad. Esta información se puede adquirir a raíz de la técnica RNA-Seq; lo cual justifica la realización de este proyecto en pro de la búsqueda de un perfil de mutaciones autóctonas de pacientes con pterigion que se atienden en el Hospital III Juliaca EsSalud Juliaca.

Para llevarlo a cabo, se obtuvo la aprobación del comité de ética del Hospital III Juliaca EsSalud. Además, se revisaron las historias clínicas de 7 pacientes operados de pterigion en el periodo 2019-2020, donde las muestras se obtuvieron mediante cirugía de exéresis de pterigion, las cuales se colocaron de forma inmediata en un tubo estéril eppendorf con hielo seco, transportadas y almacenadas en cadena de frío en el laboratorio de biología molecular de la Universidad Católica de Santa María Arequipa, donde fueron procesadas para aislamiento de ADN, y enviadas al laboratorio NOVOGEN y Admera Health Biopharma Services en USA para secuenciación mediante técnica RNA-Seq. A su vez, se contó con una segunda opinión de la lectura bioinformática en el laboratorio del Instituto Pasteur de París, Francia. Con esta información se pudo recolectar la información en formato BAM, sobre el perfil de expresión génica y mutacional con ayuda del *software* IGV.

En la **Tabla 2** se observaron las características epidemiológicas de los pacientes con el diagnóstico de pterigion. El 57.14% fueron varones y 42.86% mujeres, con edad predominante entre los 30 y 49 años. La edad promedio fue de 46.57 años, y el 42.86 % laboraba como docente.

Los resultados de este estudio difieren de los encontrados por Moran DJ, Hollows FC⁴¹, quienes analizaron pacientes aborígenes de Australia. Este último no encontró diferencias en la presentación en varones versus mujeres.

Por otro lado, en el estudio realizado por Saw et al. ⁽¹⁰⁾ en Singapore, el 80.3 % fueron varones, lo cual es equiparable con el presente estudio. La edad promedio fue de 54.2 años, difiriendo de los resultados y la mayoría de pacientes que presentaban pterigion, tuvieron antecedentes de exposición prolongada a la radiación ultravioleta, lo cual coincide con este estudio.

En la **Tabla 3** se identificaron las muestras, y los hallazgos de control de calidad cumplieron con los requerimientos del laboratorio Admera.⁹¹ El volumen para realizar RNAseq es ≥ 20 ul, todas las muestras presentadas tuvieron 50 ul. En lo relacionado con la cantidad requerida, el laboratorio permite muestras mayores de 1 ug y, en este caso, las siete muestras tuvieron cantidades totales superiores a las solicitadas en una concentración de ARN entre 200 ng – 1 ug.

Por otra parte, los valores de RIN se encontraron por debajo de 7 en todas las muestras, por lo que, según los estándares de calidad, se optó por el método de la depleción de RNA ribosomal de Illumina, el cual necesita RIN: ≥ 2 . Para la construcción de la librería de RNAseq, el número de integridad del ARN (RIN) de las siete muestras estuvo en un rango entre 3 y 4, lo cual funcionó para realizar amplificación de pequeñas regiones génicas, el porcentaje de fragmentos de ARN > 200 nucleótidos (DV 200); en la mayoría de las muestras presenta cifras superiores a 66.1%.

En la **Tabla y Figura 4** se mostraron las Estadísticas generales MultiQC. En los diferentes parámetros cumple con lo requerido por el laboratorio Admera⁹¹, el contenido promedio de GC de las lecturas estudiadas se encontró en 49 % (los rangos de normalidad en el genoma humano varían entre el 35-60 %).

Los porcentajes de alineaciones se encontraron mayores al 70.5 % (considerándose óptimos mayores a 70 %). La distribución del porcentaje de bases primarias que se alinean con regiones en el genoma de referencia, evidencian mayores porcentajes de alineaciones de lectura con las áreas intrónicas e intergénicas, lo cual se espera como resultado del estudio con la metodología de la depleción de RNA ribosomal ⁽⁸⁷⁾.

La **Figura 5** brindó los datos sobre STAR (Spliced Transcripts Alignment to a Reference) que es un mapeador rápido de lecturas de secuenciación de ARN, con soporte para detección de lecturas de fusión y empalme.

Asimismo, alinea las lecturas al encontrar los puntos de coincidencia del Prefijo Máximo Mapeable (MMP) entre las lecturas (o pares de lecturas) y el genoma, empleando un índice de matriz de sufijos⁽⁸⁸⁾.

En el presente trabajo se encontró 13.29 %, de las lecturas mapeadas en múltiples loci, lecturas mapeadas en demasiados loci 1.53 %, y lecturas no mapeadas por demasiados desajustes 17.33 %, de acuerdo con lo requerido por el laboratorio Admera.

La **Figura 6** evidenció las longitudes de secuencias recortadas, utilizando Cutadapt, la cual es una herramienta que puede eliminar secuencias adaptadoras de los datos de secuenciación de ADN. Esta permite la eliminación de múltiples adaptadores tanto de los extremos 3' como 5'.

El algoritmo únicamente elimina bases de baja calidad. En la figura, las longitudes recortadas en las siete muestras son variables, garantizando muestras de buena calidad.

En las **Figuras 7 y 8** se expusieron los resultados obtenidos por FastQC, que es una herramienta que realiza controles de calidad de datos de secuencias sin procesar, en el Histograma de calidad de secuencia y en las Puntuaciones de calidad por secuencia. Estas puntuaciones son superiores a 30, lo que se considera generalmente muy buena y se la conoce comúnmente como Q30, asegurando la calidad de muestra⁽⁸⁹⁾.

Las **Figuras 9-15**, mostraron los resultados del Contenido de ATCG por base secuenciada de las muestras 1 – 7. A raíz de esto, se encontró que las cuatro líneas que representan las proporciones de las bases en cada posición son paralelas y están juntas en la mayoría de las muestras⁽⁸⁹⁾.

Además, la proporción de bases A/T o G/C no difiere en más del 10 % en cualquier posición, lo que garantiza muestras de secuenciación de buena calidad.

En la **Figura 16**, Contenido de GC por secuencia, la mayoría de las muestras presenta una curva de Gauss con distribución normal del contenido de GC y se emplea a menudo como distribución de referencia para examinar la calidad de las bibliotecas de secuenciación.

En la **Figura 17**, FastQC, por contenido base N, se pudo observar que el contenido de N es menor al 5 %. La proporción N representa el porcentaje de posiciones en las lecturas de secuenciación en las que no se puede determinar la base con seguridad, y suele ser menor al 5 %, lo cual coincide con las muestras del presente proyecto, garantizando precisión de las lecturas de secuenciación.

En la **Figura 18** se observó la distribución de longitud de secuencia, además, muestra la distribución de longitud uniforme de los fragmentos de las siete muestras analizadas. Esto se observa debido a que la biblioteca se generó a partir de la plataforma Illumina, la cual permitió un correcto alineamiento y el ensamblaje.

La **Figura 19** mostró el nivel de duplicación de secuencias. La mayoría de las muestras no supera el 20 %, sin embargo, esto se esperaba debido a que, cuando la profundidad de secuenciación es alta pero cuenta con demasiada duplicación, puede indicar problemas técnicos o contaminación de la muestra.

En la **Figura 20** se evidenciaron las secuencias sobrerrepresentadas. La mayoría de las muestras está sobre representadas en menos del 0,1 %. Es importante tener en cuenta que el análisis se lleva a cabo en las primeras 200 000 lecturas para facilitar el cálculo. Por lo tanto, las secuencias sobrerrepresentadas que se produzcan después de este punto, pueden no incluirse en el análisis ⁽⁸⁹⁾.

En la **Tabla 5, 6 (parte 1-2)** y 8, se mostraron los 50 genes con mayor frecuencia de expresión en cada una de las siete muestras y, en las **Figuras 21-26**, se grafican, por criterio jerárquico, los 10 primeros de estos genes anteriormente mencionados.

En el presente estudio el método de secuenciación de RNA Seq NGS, permitió identificar los genes con mayor frecuencia de transcripción, comunes y repetitivos en los 7 casos de estudio. Ante esto, los 5 primeros genes con mayor frecuencia de expresión en la muestra 1 los siguientes: RPS29, EPHB1, AHNAK, MTRNR2L12, NFYB.

Es importante mencionar que ninguno de estos genes se ha descrito en la literatura relacionados con pterigion, los genes RPS29, EPHB1, AHNAK, se han relacionado con neoplasias colorrectales, gliomas, y neuroblastomas respectivamente. Y el gen MTRNR2L12 regula de manera negativa la apoptosis, la cual respalda la teoría de que el pterigion es una neoplasia benigna.

En la muestra 2, se expresan más los 5 genes: PCBD2, RPS29, RBFOX1, KAZN, DLG2. Estos genes no se han reportado vinculados con pterigion y se relacionan con alteraciones en el desarrollo y neurodesarrollo.

En la muestra 3, los 5 genes expresados en orden jerárquico son los siguientes: RPS29, EPHB1, KRT13, PCBD2, KAZN. Estos se relacionan con cáncer de pulmón, mama, cerebro, hueso, colorectal. Adicionalmente, varios de estos genes están involucrados en el desarrollo del cito esqueleto, al expresarse en las muestras de pterigion se evidencia que presentan genes muy importantes vinculados con neoplasias en su estructura.

En la muestra 4, se expresan más los genes: RPS29, EPHB1, MTRNR2L12, PCBD2, AHNAK, y estos se relacionan con cáncer colorrectal, cánceres hematológicos, gliomas, participan en la regulación negativa de la fase de ejecución de la apoptosis, en el metabolismo de proteínas, formación de la barrera hematoencefálica, en la migración celular, la regulación del canal de calcio cardíaco y la metástasis tumoral, respectivamente. No se han reportado asociados a pterigion.

La muestra 5 expresa con más frecuencia los genes: RPS29, EPHB1, NFYB, MTRNR2L12, CROCC, los cuales se manifiestan en el cáncer colorrectal, cánceres hematológicos, gliomas, y el metabolismo de esteroides. Estos hacen parte de la regulación negativa de la fase de ejecución de la apoptosis, y se vinculan con moléculas estructurales y la unión, respectivamente.

Los 5 genes más expresados en la muestra 6 fueron los siguientes: RPS29, MTRNR2L12, EPHB1, PCBD2, CROCC, los cuales se asocian a cáncer colorrectal, cánceres hematológicos, gliomas, y participan en la regulación negativa de la fase de ejecución de la apoptosis. Además, forman parte del metabolismo de proteínas, y se relacionan con moléculas estructurales y de unión, respectivamente.

En la muestra 7, los 5 más frecuentes son los que se mencionan aquí: RPS29, RBFOX1, MTRNR2L12, PCBD2, MACROD2. Estos se relacionan con cáncer colorrectal, enfermedades neurodegenerativas familiares, y participan en la regulación negativa de la fase de ejecución de la apoptosis, codifican proteínas, respectivamente.

En las siete muestras estudiadas, ninguno de los genes mencionados se reportan en la literatura asociados a pterigion. El gen que se expresó con más frecuencia es el gen RPS29, el cual se encuentra en el cromosoma 14. Este codifica proteínas ribosómicas y participa en la actividad supresora de tumores. Al respecto, se ha observado una expresión variable

de este gen en cánceres colorrectales si se comparan con tejidos normales adyacentes. Los hallazgos adquiridos suman desde la nueva información científica y permite un mayor conocimiento de la genética del pterigion, demostrando que en la estructura de pterigion están involucrados diversos genes relacionados con neoplasias y con la apoptosis. A raíz de esto, es preciso aseverar que el pterigion es una neoplasia benigna.

En el presente estudio, el 100 % de las muestras de pterigion, expresaron el gen p53.

En la **Tabla 9 y Figuras 28-33** se mostró esquemáticamente la ubicación cromosómica de las mutaciones puntuales del gen p53 comunes encontradas en los 7 pacientes de la población de pacientes del estudio, portadores de pterigion.

Asimismo, se hallaron 5 mutaciones SNV (variante de un solo nucleótido) de p53, tomando como referencia la base de datos del ensamblaje del genoma humano GRCh38.p13 (versión 13 del Consorcio de investigación del genoma humano 38), que se menciona a continuación:

SNV C > G del cromosoma: 17: 7676154 en el exón 4, SNV A > C del cromosoma: 17: 7669372 en el exón 8, SNV C > T del cromosoma: 17: 7669 124 en el exón 8, SNV G > T, A del cromosoma: 17: 7668836 en el exón 8, SNV C > G, T del cromosoma: 17: 7668799 en el exón 8.

Cabe mencionar que la expresión de p53 en pterigión se ha estudiado en pocos trabajos de investigación. La mayoría de estos estudios utilizaron IHC para examinar los niveles de proteína p53, solo dos estudios emplearon el análisis de transferencia Western y ensayos inmunoabsorbentes ligados a enzimas. Otro estudio utilizó muestras de ADN para analizar la mutación de p53 de células epiteliales y se sometieron a secuenciación de ADN para estudiar mutaciones en los exones 4, 5, 6, 7 y 8 del gen p53.

En los estudios de IHC, la tasa positiva de p53 informada, o la prevalencia del gen p53 varió de 7,9 a 100 %. ^{58,59,63.66.82}

La IHC refleja, simplemente, los niveles de proteína disponibles para la inmunorreactividad en las células, por lo tanto, las mutaciones en un gen pueden no tener ningún efecto negativo en la inmunorreactividad de la proteína.

En el estudio de secuenciación de ADN realizado por Tsai et al.¹ se detectaron mutaciones en el gen p53 en 8 muestras pterigiales (15,7 %), y solo se encontró una mutación en cada muestra. Todas las mutaciones observadas fueron puntuales, entre las cuales 6 fueron

sustituciones y 2 deleciones. Se identificaron tres mutaciones en el exón 6, dos en el exón 7 y una única mutación en cada uno de los exones 4, 5 y 8 ⁽⁸²⁾.

Además, Reisman et al. ⁽⁹⁰⁾ encontraron que p53 había sufrido una delección monoalélica y el alelo restante permaneció en estado salvaje en 9 pacientes estadounidenses. Asimismo, en un estudio de Spandidos et al. ⁽⁹¹⁾ el 60 % de los pterigion exhibieron varias mutaciones de ADN, incluyendo pérdida de heterocigosidad e inestabilidad de microsatélites, que terminan siendo alteraciones de las vías de reparación del ADN. Shimmura et al. ⁽⁹²⁾ informaron que no se encontró ninguna mutación en los exones 5 a 8 en 6 japoneses; estas incongruencias podrían deberse a factores étnicos.

A diferencia de los resultados mencionados, en este estudio se empleó la técnica RNA seq NGS. Al respecto, cabe mencionar que en la literatura no se reportan estudios utilizando dicho protocolo, lo que permite contribuir al presente trabajo con nuevos conocimientos en el campo de la genética oftalmológica.

Las mutaciones halladas en los casos expuestos fueron mutaciones puntuales; cuatro en el exón 8 y una mutación en el exón 4. Las células con mutaciones fuera de los exones 5 a 8 son raras. Se ha informado que el exón 4 abarca los codones 32-125 y contienen dominios involucrados en la inducción de apoptosis, que puede estar relacionado con el comportamiento del tumor ⁽⁶¹⁾.

En humanos, los daños en el ADN son subsanados por diferentes Sistemas de reparación de ADN. Los cambios de base única que son los que se observan en los resultados del presente proyecto, son normalmente revertidos por el sistema de reparación de escisión de base (BER). Por lo tanto, la presencia de un cambio de base única de la transición de C a T que se evidencia en el cromosoma: 17: 7669 124 (GRCh38) exón 8, se reporta que debe estar vinculado con perturbaciones en las diferentes reparaciones del ADN. Esto sugiere que los pacientes con pterigión tienen sistemas BER defectuosos.

Antes, el pterigion se clasificaba como una enfermedad degenerativa crónica. Sin embargo, a partir de los resultados obtenidos es posible considerar que el pterigión es una proliferación celular descontrolada secundaria a una expresión anormal de la proteína p53 dentro del epitelio conjuntival ⁽⁵⁹⁾.

En la **Tabla 9 (parte 1 y 2) y las Figuras 34 a 46** se mostraron las traslocaciones de genes tipo fusión, que se presentan por cambios estructurales genómicos resultantes de la disrupción del ADN, seguida de reincorporación y reproducciones erróneas ^(72,93).

En la **Figura 34**, se observó el gen de fusión CLDND1 – KIAA2026 que se expresó en la muestra 1 de pterigion. El gen CLDND1 es una proteína transmembrana que predomina en el sistema nervioso central, y se relaciona con leucodistrofia similar a la de Pelizaeus-Merzbacher. KIAA2026 se expresa en el citoplasma en la mayoría de los tejidos, como la retina y testículos, este gen de fusión que se encuentra en el cromosoma 9, y no retiene dominios proteicos.

En la **Figura 35**, se reportó el gen de fusión RNU1-2 – ATP13A2, en la muestra 3 de pterigion, RNU1-2 es un gen snRNA. Entre sus vías relacionadas se encuentra el ciclo de empalme espliceosómico y ATP13A2 participa en el transporte de cationes de metales de transición divalentes. La deficiencia se relaciona con paraplejía espástica y síndrome de Kufor-Rakeb. Asimismo, el gen de fusión RNU1-2 - ATP13A2, no presenta regiones codificantes debido a la transcripción anti sentido.

En la **Figura 36** se graficó el gen de fusión RNU1-2 – SDHB. En la muestra 3 de pterigion, RNU1-2 es un gen snRNA, se relaciona con el ciclo de empalme espliceosómico, SDHB es un gen supresor de tumores que ayuda con la protección contra tumores benignos y malignos. Además, el gen de fusión RNU1-2 – SDHB, no presenta regiones codificantes debido a la transcripción anti sentido.

En la **Figura 37** se expresa el gen de fusión WIZ – WHRN, el gen WIZ participa en la regulación positiva de la replicación del ADN del ciclo celular nuclear y la estabilización de proteínas. Por otra parte, el gen WHRN está involucrado en la organización, estabilización de la elongación de los esterocilios y el ensamblaje citoesquelético de actina, WIZ – WHRN es un gen de fusión con dominios de proteína retenidos y fusión fuera del marco.

En la **Figura 38**, el gen de fusión 19:40114634 - SNORD3A, involucra el gen SNORD3A que es un ARN nucleolar pequeño. Este desempeña un rol en el procesamiento de los precursores del ARN ribosómico. Y por último, el gen 19:40114634 que es uno nuevo ubicado en el cromosoma 19; este nuevo gen de fusión no codifica proteínas.

En la **Figura 39** se evidenció el gen de fusión 6:164999470 - RNA5SP226, el gen RNA5SP226 es un pseudogén, se expresa múltiples tejidos. Y el gen 6:164999470 ubicado en el cromosoma 6, es un nuevo gen, el gen de fusión 6:164999470 - RNA5SP226 no codifica proteínas.

En la **Figura 40** el gen de fusión PSMC3 - NXF1 incluye al gen PSMC3, el cual es codificador de proteínas. Las enfermedades relacionadas con este son la sordera, cataratas, deterioro del desarrollo intelectual y polineuropatía y displasia del desarrollo de la cadera. En lo concerniente al gen NXF1, es preciso señalar que está relacionado con la unión de ARN y la unión de nucleótidos. El gen de fusión PSMC3 - NXF1, presenta dominios de proteína retenidos y fusión en el marco.

La **Figura 41** graficó el gen de fusión WNK1 - ZNF79. En este caso, el gen WNK1 interviene en la homeostasis iónica y modula la presión arterial, transportando iones dentro y fuera de la célula. ZNF79 es un gen codificador de proteínas, y presenta dominios de proteína retenidos y fusión en el marco.

La **Figura 42** presentó el gen de fusión DOCK8 - RF00100. El primero es una proteína adaptadora en la señalización del receptor tipo Toll 9 (TLR9), involucrada en la vía de señalización del receptor de linfocitos T y en la reorganización del citoesqueleto, RF00100 es un snRNA. Actualmente no se ha reportado su asociación a enfermedades. El gen de fusión DOCK8 - RF00100 presenta dominios de proteína retenidos y fusión fuera del marco.

La **Figura 43** indicó el gen de fusión 19:40114634 - SNORD3A. El gen SNORD3A desempeña un papel en el procesamiento de los precursores del ARN ribosómico. Así, 19:40114634 es un nuevo gen que se reporta en este estudio, el gen de fusión 19:40114634 - SNORD3A no codifica proteínas.

En la **Figura 44** se puede observar el gen de fusión AC093752.1 - RUFY3. El AC093752.1 es un nuevo gen que se reporta en este estudio de investigación,. Por otra parte, el gen RUFY3 codifica proteínas necesarias para la conservación de la polaridad neuronal. Por lo tanto, el gen de fusión AC093752.1 - RUFY3 presenta dominios de proteína retenidos y el marco de lectura es poco claro.

La **Figura 45** presentó el gen de fusión AC104151.1 - CNTNAP4. AC104151.1 es un nuevo gen que se reporta en el presente trabajo de investigación. Por otro lado, CNTNAP4 está involucrado en con la adhesión celular, este gen de fusión presenta dominios de proteína retenidos y el marco de lectura es poco claro.

En la **Figura 46** se evidenció el gen de fusión AC104151.1 - CNTNAP4. Por un lado, el AC104151.1 es un nuevo gen que se reporta en la literatura y CNTNAP4 está involucrado con la adhesión celular. Por lo tanto, este gen de fusión se forma en el cromosoma 16. En

el gen AC104151.1 existe un punto de ruptura a nivel del exón 2 y se une al fragmento formado por la ruptura en el exón 18 del gen CNTNAP4. Cabe mencionar que este gen de fusión presenta dominios de proteína retenidos y el marco de lectura es poco claro.

Los genes de fusión PSMC3 - NXF1 y WNK1 - ZNF79. Estos tienen dominios de proteína retenidos y fusiones en el marco, sintetizan proteínas que pueden actuar como potentes impulsores en la génesis del pterigion. Los otros genes de fusión mencionados no codifican proteínas, lo cual señala que estos no son funcionales ^(94,95).

Los genes de fusión que se reportan en el presente estudio no se hallan en la base de datos Mitelman de aberraciones cromosómicas y genes de fusión en cáncer. Esto significa que se han encontrado nuevos genes de fusión.

El descubrimiento de nuevos genes, mutaciones y genes de fusión en pterigion va a brindar una mayor comprensión de las transformaciones moleculares en el tejido proliferativo y permitirá mejorar el enfoque para el diagnóstico, pronóstico y terapia en pacientes con pterigion.

La importancia funcional de los hallazgos del trabajo de investigación, promoverá el uso de terapias dirigidas, allanando el camino hacia la medicina individualizada.

CONCLUSIONES

Primera. Se identificaron mutaciones del gen humano p53 en las muestras de pterigión primario.

Segunda. La expresión génica de las biopsias analizadas muestra un perfil nunca antes reportado en esta patología.

Tercera. Se han encontrado 5 nuevas mutaciones en el gen p53: SNV C > G del cromosoma: 17: 7676154 en el exón 4, SNV A > C del cromosoma: 17: 7669372 en el exón 8, SNV C > T del cromosoma: 17: 7669 124 en el exón 8, SNV G > T, A del cromosoma: 17: 7668836 en el exón 8, SNV C > G, T del cromosoma: 17: 7668799 en el exón 8.

Cuarta. Se encontraron 19 261 genes en cada muestra de pterigión, los genes más expresados fueron RPS29, AHNAK, MTRNR2L12, PCBD2

Quinta. Se hallaron 13 proteínas de fusión utilizando el *software* “Arriba (versión 1.2.0)”: CLDND1 – KIAA2026, RNV1-2 – ATP13A2, RNU1-2 – SDHB, WIZ – WHRN, 19:40114634 - SNORD3A, 6:164999470 - RNA5SP226, PSMC3 - NXF1, WNK1 - ZNF79, DOCK8 - RF00100, 19:40114634 - SNORD3A, ACO93752.1 - RUFY3, AC104151.1 - CNTNAP4, AC104151.1 - CNTNAP4, que por primera vez son descritas en esta tesis.

Sexta. Este estudio confirma que el pterigión está relacionado con factores oncogénicos, angiogénicos, fibrogénicos e inflamatorios y, además, brinda evidencia sobre nuevos genes, los cuales pueden contribuir a la patogenia.

RECOMENDACIONES

1) Los resultados obtenidos, brindan información que permitirá mejorar el enfoque diagnóstico, terapéutico del pterigion y también evidencia lo importante de realizar investigación genética de patologías oftalmológicas.

Por lo mencionado, se recomienda llevar a cabo estudios de enfermedades muy prevalentes en el sur de Perú como pterigion, queratocono, degeneración macular relacionada con la edad, glaucoma. Todo esto, a través de la aplicación de técnicas de RNA-Seq para la pesquisa de blancos terapéuticos y pronósticos a partir de los hallazgos de perfil de expresión génica autóctono de cada población, que nos va a permitir tratamientos individualizados.

2) El acceso a técnicas de diagnóstico genético es muy limitada en el sistema sanitario de nuestro país. Por esta razón, se debería permitir el acceso al uso de técnicas de RNA-Seq en los diferentes hospitales y clínicas del Perú. Estos avances van a brindar un diagnóstico más preciso y, por consiguiente, tratamientos oportunos y eficientes.

3) Las nuevas mutaciones reportadas y las proteínas de fusión deben ser estudiadas en mayor cantidad de muestras, con el fin de desarrollar alternativas diagnósticas y terapéuticas innovadoras.

PROPUESTA

1) La Universidad Católica de Santa María, tiene el potencial humano y logístico que permitiría desarrollar convenios prestacionales entre la Universidad Católica de Santa María y Ministerio de Salud, EsSalud, Fuerzas armadas e instituciones privadas del sur del Perú para la ejecución de investigación científica, aplicando técnicas de RNA-Seq por el laboratorio de biología molecular de la Universidad, para conocer el perfil genómico de las enfermedades más prevalentes del sur del Perú, que va a permitir optimizar la calidad de atención en salud, mejorar los diagnósticos e incluir tratamientos innovadores.



REFERENCIAS BIBLIOGRÁFICAS

1. Sun N, Zhang H. Pyroptosis in pterygium pathogenesis. *Bioscience reports*. 2018; 38(3).
2. Young AL, Cao D, Chu WK, Ng TK, Yip Y, Jhanji V, et al. The Evolving Story of Pterygium. *Cornea*. 2018; 37(1).
3. Bradley JC, Yang W, Bradley RH, Reid TW, Schwab IR. The science of pterygia. *The British journal of ophthalmology*. 2010; 94(7).
4. Naib-Majani W, Eltohami I, Wernert N, Watts W, Tschesche H, Pleyer U, et al. Distribution of extracellular matrix proteins in pterygia: an immunohistochemical study. *Graefes Arch Clin Exp Oftalmol*. 2004; 242(4).
5. Jiang B, Wang Y, Zhu ZC, Wu ZY. Expression of specificity protein 1 and collagen I in primary pterygial tissues. *Arquivos brasileiros de oftalmologia*. 2019; 82(5).
6. A M, Contreras E, Figueroa E, Velasquez M, Descailleaux E. Mutations in TP53: A short review.. 2016. *Revista Peruana de Divulgación Científica en Genética y Biología Molecular*.
7. Bahar I, Kaiserman I, Lange AP, Slomovic A, Levinger E, Sansanayudh W, et al. The Effect of Mitomycin C on Corneal Endothelium in Pterygium Surgery. *American Journal of Ophthalmology*. 2009; 147(3).
8. Kaufman SC, Jacobs DS, Lee WB, Deng SX, Rosenblatt MI, Shtein RM. Options and Adjuvants in Surgery for Pterygium. *Ophthalmology*. 2013; 12(1).
9. Wanzeler A, Barbosa I, Duarte B, Borges D, Barbosa EB, Kamiji D, et al. Mechanisms and biomarker candidates in pterygium development. *Arquivos brasileiros de oftalmologia*, 82(6),. 2019; 82(6).
10. Saw SM, Banerjee K, Tan D. Risk factors for the development of pterygium in Singapore: A hospital-based case-control study. *Acta Ophthalmologica Scandinavica*. 2000; 78(2).

11. Safi H, Kheirkhah A, Mahbod M, Molaei S, Hashemi H, Jabbarvand M. Correlations Between Histopathologic Changes and Clinical Features in Pterygia. *Journal of ophthalmic & vision research*. 2016; 11(2).
12. Miyata K, Minami K, Otani A, Tokunaga T, Tokuda S, Amano S. Proposal for a Novel Severity Grading System for Pterygia Based on Corneal Topographic Data. *Cornea*. 2017; 36(7).
13. Reda AM, Shaaban Y, Saad El-Din SA. Histopathological Parameters in Pterygia and Significant Clinical Correlations. *Journal of ophthalmic & vision research*. 2018; 13(2).
14. Hilgers JHC. Pterygium: Its Incidence, Heredity and Etiology. *American Journal of Ophthalmology*. 1960; 50(4).
15. Reid TW, Dushku N. What a study of pterygia teaches us about the cornea? Molecular mechanisms of formation. *Eye Contact Lens*. 2010 ; 36(5).
16. Detorakis E, Drakonaki E, Spandidos D. Molecular genetic alterations and viral presence in ophthalmic pterygium. *International Journal of Molecular Medicine*. 2000; 6.
17. García K, Romero M, Rodríguez M, Tenorio G. Correlación morfológica del pterigión y su evolución clínica. *Rev Med Hosp Gen Mex*. 2006; 69(4).
18. Fry J, Grau A, Herrera J, Zoroquiain JP. Evaluación histopatológica en las muestras de cirugía de pterigión ¿Es necesario? *ARS Medica Revista de Ciencias Médicas*. 2019; 44(2).
19. García A, Casellas M, Rodríguez M, González N, Acevedo L. Caracterización clínica e histopatológica del pterigion primario. *Rev. Arch Med Camagüey*. 2018; 22(2).
20. Rojas E, Pérez A, González J. Aspectos anatomopatológicos del pterigion primario. X Congreso Virtual Hispanoamericano de Anatomía Patológica [Internet]. [Online].; 2009. Available from: HYPERLINK http://www.conganat.org/10congreso/trabajo.asp?id_trabajo=1821&tipo=1.

21. Rezvan F, Khabazkhoob M, Hooshmand E, Yekta A, Saatchi M, Hashemi H. Prevalence and risk factors of pterygium: a systematic review and meta-analysis. *Survey of Ophthalmology*. 2018; 63 (5).
22. Flores A. Altitud geografica como factor asociado al desarrollo de pterigión. [Universidad Privada Antenor Orrego – UPAO]. [Online].; 2018. Available from: HYPERLINK <http://repositorio.upao.edu.pe/handle/upaorep/3889>.
23. Tan DT, Chee SP, Dear KB, Lim AS. Effect of pterygium morphology on pterygium recurrence in a controlled trial comparing conjunctival autografting with bare sclera excision. *Archives of ophthalmology (Chicago, Ill. : 1960)*. 1997; 115(10).
24. Chiang CC, Tsai YY, Bau DT, Cheng YW, Tseng SH, Wang RF, et al. Pterygium and genetic polymorphisms of the DNA repair enzymes XRCC1, XPA, and XPD. *Molecular Vision*. 2010; 16.
25. Liang K, Jiang Z, Ding BQ, Cheng P, Huang DK, Tao LM. Expression of cell proliferation and apoptosis biomarkers in pterygia and normal conjunctiva. *Molecular Visio*. 2011; 17.
26. Cimpean AM, Sava MP, Raica M. DNA damage in human pterygium: one-shot multiple targets. *Molecular Vision*. 2013; 19.
27. Ramirez CA, Pérez-Martinot M, Gil-Huayanay D, Urrunaga-Pastor D, Benites-Zapata VA. Ocular Exposure to Particulate Matter and Development of Pterygium: A Case-Control Study. *The international journal of occupational and environm. The international journal of occupational and environmental medicine*. 2018; 9(4).
28. Hu X, Tandra N, Zhang Z, Gong A, Chen J, Li Y, et al. Identification and differentiation therapy strategy of pterygium in vitro. *American Journal of Translational Research*. 2018; 10(8).
29. Maxia C, Perra MT, Demurtas P, Minerba L, Murtas D, Piras F, et al. Expression of survivin protein in pterygium and relationship with oxidative DNA damage. *Journal of Cellular and Molecular Medicine*. 2008 Dec; 6A.
30. Balci M, Sahin S, Mutlu FM, Yağci R, Karanci P, Yildiz M. Investigation of oxidative stress in pterygium tissue. *Molecular Vision*. 2011; 17.

31. Ochoa T. Génesis del pterigión. Una aproximación desde la biología molecular. *Rev Mex Oftalmol*. 2006; 80(6).
32. Anguria P, Kitinya J, Ntuli S, Carmichael T. The role of heredity in pterygium development. *International journal of ophthalmology*. 2014; 7 (3).
33. Fang C, Illingworth CD, Qian L, Wormstone IM. Serum deprivation can suppress receptor-mediated calcium signaling in pterygial-derived fibroblasts. *Investigative Ophthalmology & Visual Science*. 2013; 54(7).
34. Kim HK, Choi JY, Park SM, Rho CR, Cho KJ, Jo SA. Tyrosine Kinase Inhibitor, Vatalanib, Inhibits Proliferation and Migration of Human Pterygial Fibroblasts. *Cornea*. 2017; 36(9).
35. Gum SI, Kim YH, Jung JC. Cyclosporine A inhibits TGF- β 2-induced myofibroblasts of primary cultured human pterygium fibroblasts. *Biochem Biophys Res Commun*. 2017; 482(4).
36. Tsai YY, Cheng YW, Lee H, Tsai FJ, Tseng SH, Lin CL, et al. Oxidative DNA damage in pterygium. *Molecular Vision*. 2005; 11.
37. Kau HC, Tsai CC, Lee CF, Kao SC, Hsu WM, Liu JH, et al. Increased oxidative DNA damage, 8-hydroxydeoxy- guanosine, in human pterygium. *Eye*. 2005; 20(7).
38. Perra MT, Maxia C, Corbu A, Minerba L, Demurtas P, Colombari R, et al. Oxidative stress in pterygium: relationship between p53 and 8-hydroxydeoxyguanosine. *Molecular Vision*. 2006; 12.
39. Saccà SC, Roszkowska AM, Izzotti A. Environmental light and endogenous antioxidants as the main determinants of non-cancer ocular diseases. *Mutation Research/Reviews in Mutation Research*. 2013; 752(2).
40. Kilic-Toprak E, Toprak I, Caliskan S, Ozdemir Y, Demirtas O, Altintas F, et al. Oxidative Stress and Genotoxicity in Pterygium. *Eye & Contact Lens: Science & Clinical Practice*. 2019; 45(6).
41. Moran DJ, Hollows FC. Pterygium and ultraviolet radiation: a positive correlation. *The British journal of ophthalmology*. 1984; 68(5).

42. Coroneo MT. Pterygium as an early indicator of ultraviolet insolation: a hypothesis. *The British journal of ophthalmology*. 1993; 77 (11).
43. Tasheva ES, Conrad GW. The UV responsive elements in the human mimecan promoter: a functional characterization. *Molecular Vision*. 2003; 9.
44. Cui D, Pan Z, Zhang S, Zheng J, Huang Q, Wu K. Downregulation of c-Myc in pterygium and cultured pterygial cells. *Clinical & Experimental Ophthalmology*. 2011; 39(8).
45. Zhu C, Pan F, Ge L, Zhou J, Chen L, Zhou T, et al. SERPINA3K plays antioxidant roles in cultured pterygial epithelial cells through regulating ROS system. *PloS one*. 2014; 9(10).
46. Zhang LW, Chen BH, Xi XH, Han QQ, Tang LS. Survivin and p53 expression in primary and recurrent pterygium in Chinese patients. *International Journal of Ophthalmology*. 2011; 4(4).
47. Ozturk BT, Yildirim MS, Zamani A, Bozkurt B. K-ras oncogene mutation in pterygium. *Eye*, 31(3),. 2016; 31(3).
48. Ueda Y, Kanazawa S, Kitaoka T, Dake Y, Ohira A, Ouertani AM, et al. Immunohistochemical study of p53, p21 and PCNA in pterygium. *Acta histochemica*. 2001; 103 (2).
49. Li DQ, Lee SB, Gunja-Smith Z, Liu Y, Solomon A, Meller D, et al. Overexpression of collagenase (MMP-1) and stromelysin (MMP-3) by pterygium head fibroblasts. *Archives of ophthalmology (Chicago, Ill: 1960)*. 2001; 119(1).
50. Kria L, Ohira A, Amemiya T. Growth factors in cultured pterygium fibroblasts: immunohistochemical and ELISA analysis. *Graefes Arch Clin Exp Oftalmol*. 1998; 236(9).
51. Cox CA, Amaral J, Salloum R, Guedez L, Reid TW, Jaworski C, et al. Doxycycline's effect on ocular angiogenesis: an in vivo analysis. *Ophthalmology*. 2010; 117 (9).
52. Khalfaoui T, Mkannez G, Colin D, Imen A, Zbiba W, Errais K, et al. Immunohistochemical analysis of vascular endothelial growth factor (VEGF) and

- p53 expression in pterygium from Tunisian patients. *Pathologie Biologie*. 2011; 59(3).
53. Sjogren S, Inganas M, Norberg T, Lindgren A, Nordgren H, Holmberg L, et al. The p53 Gene in Breast Cancer: Prognostic Value of Complementary DNA Sequencing Versus Immunohistochemistry. *JNCI Journal of the National Cancer Institut*. 1996; 88(3-4).
54. Sezaki T, Tomiyama L, Kimura Y, Ueda K, Kioka N. Dlg5 interacts with the TGF- β receptor and promotes its degradation. *FEBS Letters*. 2013; 587.
55. Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS. Why are there hotspot mutations in the TP53 gene in human cancers? *Cell death and differentiation*. 2018; 25(1).
56. Bae SC, Kolinjivadi AM, Ito Y. Functional relationship between p53 and RUNX proteins. *Journal of molecular cell biology*. 2019; 11(3).
57. Bowen ME, Attardi LD. The role of p53 in developmental syndromes. *Journal of Molecular Cell Biology*. 2019; 11(3).
58. Dushku N, Reid TW. P53 expression in altered limbal basal cells of pingueculae, pterygia, and limbal tumors. *Current Eye Research*. 1997; 16 (12).
59. Weinstein O, Rosenthal G, Zirkin H, Monos T, Lifshitz T, Argov S. Overexpression of p53 tumor suppressor gene in pterygia. *Eye*. 2002; 16 (5).
60. Tomas D. Apoptoza, UV-zracenje, prekanceroze i tumorikoze [Apoptosis, UV-radiation, precancerosis and skin tumors]. *Acta Medica Croatica : Casopis Hrvatske Akademije Medicinskih Znanosti*. 2009; 63(2).
61. Tsai YY, Chang KC, Lin CL, Lee H, Tsai FJ, Cheng YW, et al. p53 Expression in Pterygium by Immunohistochemical Analysis. *Cornea*. 2005; 24(5).
62. Tsai YY, Chang KC, Lee H, Cheng YW, Tsai FJ, Tseng SH, et al. Laboratory Science. Effect of p53 codon 72 polymorphism on p53 protein expression in pterygium. *Clinical and Experimental Ophthalmology*. 2005; 33(1).

63. Lai TJ, Tsai YY, Cheng YW, Chiang CC, Lee H, Chou MC, et al. An association between BPDE-like DNA adduct levels and P53 gene mutation in pterygium. *Molecular Vision*. 2006; 12.
64. Rodrigues FW, Arruda JT, Silva RE, Moura KK. TP53 gene expression, codon 72 polymorphism and human papillomavirus DNA associated with pterygium. *Genetics and Molecular Research: GMR*. 2008; 7(4).
65. Nandhakumar S, Parasuraman S, Shanmugam MM, Rao KR, Chand P, Bhat BV. Evaluation of DNA damage using single-cell gel electrophoresis (Comet Assay). *Journal of pharmacology & pharmacotherapeutics*. 2011; 2(2).
66. Wu CW, Peng ML, Yeh KT, Tsai YY, Chiang CC, Cheng YW. Inactivation of p53 in pterygium influence miR-200a expression resulting in ZEB1/ZEB2 up-regulation and EMT processing. *Experimental Eye Research*. 2016; 146.
67. Tomas D. Apoptoza, UV-zracenje, prekanceroze i tumorikoze [Apoptosis, UV-radiation, precancerosis and skin tumors]. *Acta medica Croatica : casopis Hrvatske akademije medicinskih znanosti*. 2009; 63(2).
68. Juk F. Secuenciación masiva paralela (NGS): conceptos básicos y aplicaciones. *Revista Hematología*. 2019 ; 23(Extra 1).
69. Rodríguez B, Armengol L. Tecnologías de secuenciación de nueva generación en diagnóstico genético pre y postnatal. *Diagnóstico Prenatal*. 2012; 23(2).
70. Wang Q, Xia J, Jia P, Pao W, Zhao Z. Application of next generation sequencing to human gene fusion detection: computational tools, features and perspectives. *Brief. Bioinformatics*. 2013; 14(4).
71. Milman T, Ida C, Zhang PJL, Eagle RC. Gene fusions in ocular adnexal tumors. *Am. J. Ophthalmol.* ; 221. 2021; 221.
72. Dai X, Theobard R, Cheng H, Xing M, Zhang J. Fusion genes: a promising tool combating against cancer. *Reviews on Cancer*. 2018; 1869(2).
73. Lovino M, Urgese G, Macii E, Di Cataldo S, Ficarra EA. deep learning approach to the screening of oncogenic gene fusions in humans. *Int. J. Mol. Sci*. 2019; 20(7).

74. Santamaría M, Lezana J. Técnicas de secuenciación masiva (NGS). Ed. Cont. Lab. Clin. 2018; 37.
75. Taniue K, Akimitsu N. Fusion genes and RNAs in cancer development. Noncoding RNA. 2021; 7(1).
76. Rubio S, Pacheco-Orozco RA, Gómez AM, Perdomo S, García-Robles R. Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica.. Univ. Med. 2020; 61(2).
77. Owen N, Moosajee M. RNA-sequencing in ophthalmology research: considerations for experimental design and analysis. Therapeutic Advances in Ophthalmology. 2019; 11.
78. Owen N, Moosajee M. RNA-sequencing in ophthalmology research: considerations for experimental design and analysis. Therapeutic Advances in Ophthalmology. 2019; 11.
79. Tan DTH, Lim ASM, Goh HS, Smith DR. Abnormal Expression of the p53 Tumor Suppressor Gene in the Conjunctiva of Patients With Pterygium. American Journal of Ophthalmology. 1997; 123(3).
80. Schneider BG, John-Aryankalayil M, Rowsey JJ, Dushku N, Reid TW. Accumulation of p53 protein in pterygia is not accompanied by TP53 gene mutation. Experimental Eye Research. 2006; 82(1).
81. Onur C, Orhan D, Orhan M, Dizbay Sak S, Tulunay O, Irkeç M. Expression of p53 protein in pterygium. European journal of ophthalmology. 1998; 8(3).
82. Tsai YY, Cheng YW, Lee H, Tsai FJ, Tseng SH, Chang KC. P53 gene mutation spectrum and the relationship between gene mutation and protein levels in pterygium. Molecular Vision. 2005; 11.
83. Turan M, Turan G. Bcl-2, p53, and Ki-67 expression in pterygium and normal conjunctiva and their relationship with pterygium recurrence. European journal of ophthalmology. 2020; 30(6).

84. Mahesh M, Mittal SK, Kishore S, Singh A, Gupta N, Rana R. Expression of p53 and Ki-67 proteins in patients with increasing severity and duration of pterygium. Indian journal of ophthalmology. 2021; 69 (4).
85. Vega D. Secuenciamiento masivo (RNA-seq) y bioinformática del transcriptoma de tejido graso de Gallus gallus procedentes de centros de venta de Lima [Tesis Pregrado]. : Universidad Nacional Mayor de San Marcos; 2019.
86. Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res. 2021 ; 31(3).
87. Cárdenas C. Perfil de expresión génica por RNA-Seq del cáncer diferenciado de tiroides diagnosticado en el HNCASE EsSalud, Arequipa 2019-2020 [Universidad Católica de Santa María]. [Online].; 2022. Available from: HYPERLINK <https://repositorio.ucsm.edu.pe/items/e1eb56ee-f9d7-44e1-bc9c-eac38274421d>.
88. Illumina. Centro de soporte de Illumina. [Online].; s.f.. Available from: HYPERLINK <https://www.support.illumina.com/> .
89. CD Genomics. Home. [Online].; s.f.. Available from: HYPERLINK "https://bioinfo.cd-genomics.com/" <https://bioinfo.cd-genomics.com/> .
90. Reisman D, McFadden JW, Lu G. Loss of heterozygosity and p53 expression in Pterygium. Cancer Letters. 2004; 206(1).
91. Spandidos DASG;KH;TJ. Inestabilidad de microsatélites y pérdida de heterocigosidad en pterigión humano. Br. J. Oftalmol. 1997; 81.
92. Shimmura S, Ishioka M, Hanada K, Shimazaki J, Tsubota K. Telomerase activity and p53 expression in pterygia. Investigative Ophthalmology & Visual Science. 2000; 41(6).
93. Kim P, Zhou X. FusionGDB: fusion gene annotation DataBase. Nucleic Acids Res. 2019; 47.
94. Findley MK, Koval M. Regulation and roles for claudin-family tight junction proteins. IUBMB Life. 2009; 61(4).

95. Kim P, Yiya K, Zhou X. FGviewer: an online visualization tool for functional features of human fusion genes. *Nucleic Acids Res.* 2020; 48.
96. Dai X, Theobard R, Cheng H, Xing M, Zhang J. Fusion genes: a promising tool combating against cancer. *Reviews on Cancer.* ; (2):. 2018; 1869(2).



ANEXOS

ANEXO 1

CONSENTIMIENTO INFORMADO

Mediante la presente se le invita a usted participar en un estudio de investigación, que tiene como propósito conocer: **Análisis molecular del gen humano p53 en pterigión primario.**

El estudio está a cargo de la investigadora Berioska Manzaneda Murguía de la Universidad Católica de Santa María del doctorado en salud. El proyecto es financiado por la Universidad Católica de Santa María.

Participación voluntaria

Luego de terminar de revisar este consentimiento, usted es el único que decide si desea participar en el estudio. Su participación es completamente voluntaria.

Procedimientos

1. Se tomará una encuesta con preguntas sobre datos socio demográficos
2. Se utilizará el tejido extraído en su cirugía de plastia libre por pterigión primario
3. Se evaluarán resultados obtenidos en la encuesta y en las muestras de pterigión
4. Colocar y explicar cada procedimiento en palabras sencillas

Riesgos y procedimientos para minimizar los riesgos

No existen riesgos en esta investigación.

Costos

Usted no deberá asumir ningún costo económico para la participación en este estudio, el costo de las pruebas de secuenciación por NGS de pterigión serán asumidos por la investigadora.

Beneficios

Puede ser que no haya un beneficio inmediato para usted por participar de este estudio. Sin embargo, usted estará contribuyendo decisivamente a ampliar los conocimientos fisiopatológicos del pterigión primario.

Confidencialidad

Como hemos referido todos sus resultados que se generen serán tratados con la más estricta confidencialidad.

Contacto

Si usted tiene alguna pregunta acerca de este estudio el personal asignado a coordinar este estudio puede responder preguntas adicionales sobre la encuesta, o cualquier procedimiento.

Cualquier duda comunicarse a: berioska999@gmail.com

Celular 959541189, MO. Berioska Manzaneda Murguía (Investigadora)

DECLARACIÓN DE CONSENTIMIENTO INFORMADO/AUTORIZACION

Yo he tenido oportunidad de hacer preguntas, y siento que todas mis preguntas han sido contestadas.

He comprendido que mi participación es voluntaria y que puedo retirarme del estudio en cualquier momento. Además, entiendo que estando de acuerdo de participar en este estudio estoy dando permiso para se procese la información que he respondido en los cuestionarios.

En base a la información que me han dado, estoy de acuerdo en participar en este estudio.

Nombre de participante que da su autorización DNI:	
--	--

Firma y/o huella digital del participante que da su autorización	
---	--

Nombre del Investigador principal Documento de identidad	
---	--

Fecha: