



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Biológicas

Escuela Profesional de Genética y Biotecnología

**Detección de fusiones génicas en datos de secuenciación de
RNA en muestras de pacientes con melanoma lentiginoso
acról del INCan-México**

TESIS

Para optar el Título Profesional de Bióloga Genetista
Biotecnóloga

AUTOR

Ana Cecilia ROMANI VASQUEZ

ASESOR

Mónica PAREDES ANAYA

Daniela ROBLES ESPINOZA

Obed RAMÍREZ SÁNCHEZ

Lima, Perú

2023



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Romani, A. (2023). *Detección de fusiones génicas en datos de secuenciación de RNA en muestras de pacientes con melanoma lentiginoso acral del INCan-México*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Biológicas, Escuela Profesional de Genética y Biotecnología]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Ana Cecilia Romani Vasquez
Tipo de documento de identidad	DNI
Número de documento de identidad	76193612
URL de ORCID	https://orcid.org/0000-0002-5889-2721
Datos de asesor	
Nombres y apellidos	Monica Yolanda Paredes Anaya
Tipo de documento de identidad	DNI
Número de documento de identidad	07901815
URL de ORCID	https://orcid.org/0000-0002-9038-2270
Datos de coasesor 1	
Nombres y apellidos	Carla Daniela Robles Espinoza
Tipo de documento de identidad	Pasaporte
Número de documento de identidad	MX / G25940886
URL de ORCID	https://orcid.org/0000-0003-3277-7466
Datos de coasesor 2	
Nombres y apellidos	Obed Ramírez Sánchez
Tipo de documento de identidad	Cédula de identidad
Número de documento de identidad	MX / RASO850320HPLMNB08

URL de ORCID	https://orcid.org/0000-0003-3156-1209
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Gustavo Adolfo Sandoval Peña
Tipo de documento	DNI
Número de documento de identidad	41020762
Miembro del jurado 1	
Nombres y apellidos	Leyda Geraldine Cabrera Córdova
Tipo de documento	DNI
Número de documento de identidad	09156694
Miembro del jurado 2	
Nombres y apellidos	Daniel Saul Ore Chavez
Tipo de documento	DNI
Número de documento de identidad	10093526
Datos de investigación	
Línea de investigación	B.2.5.4. Tecnologías ómicas y bioinformática aplicadas en salud
Grupo de investigación	Laboratorio Internacional de Investigación sobre el Genoma Humano - LIIGH
Agencia de financiamiento	Sin financiamiento.
Ubicación geográfica de la investigación	Edificio: Facultad de Ciencias Biológicas – Universidad Nacional Mayor de San Marcos País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Calle: Germán Amezaga N° 375 Latitud: -12.059438 Longitud: -77.082136
Año o rango de años en que se realizó la investigación	Noviembre 2020 - Febrero 2022
URL de disciplinas OCDE	Bioinformática https://purl.org/pe-repo/ocde/ford#1.02.03

URL de disciplinas OCDE

Genética, Herencia

<https://purl.org/pe-repo/ocde/ford#1.06.07>



Universidad Nacional Mayor de San Marcos
(Universidad del Perú, Decana de América)

FACULTAD DE CIENCIAS BIOLÓGICAS

**ACTA DE SESIÓN PARA OPTAR AL TÍTULO PROFESIONAL DE BIÓLOGA
GENETISTA BIOTECNÓLOGA
(MODALIDAD: SUSTENTACIÓN VIRTUAL DE TESIS)**

Siendo las 08:30 horas del 24 de abril de 2023, en el Salón de Grados Virtual de la Facultad de Ciencias Biológicas cuya dirección electrónica fue <https://meet.google.com/mmb-efuz-cfy>, y en presencia del Jurado formado por los profesores que suscriben, se inició la sesión para optar al **Título Profesional de Bióloga Genetista Biotecnóloga** de ANA CECILIA ROMANI VÁSQUEZ.

Luego de dar lectura y conformidad al expediente N° UNMSM-20220092007, la titulando expuso su Tesis: **DETECCIÓN DE FUSIONES GÉNICAS EN DATOS DE SECUENCIACIÓN DE RNA EN MUESTRAS DE PACIENTES CON MELANOMA LENTIGINOSO ACRAL DEL INCan-MÉXICO**, y el Jurado efectuó las preguntas del caso calificando la exposición con la nota 19, calificativo: Aprobado con máximos honores.

Finalmente, el expediente será enviado a la Escuela Profesional de Genética y Biotecnología y al Consejo de Facultad para que se apruebe otorgar el **Título Profesional de Bióloga Genetista Biotecnóloga** a ANA CECILIA ROMANI VÁSQUEZ y se eleve lo actuado al Rectorado para conferir el respectivo grado, conforme a ley.

Siendo las 10:10horas se levantó la sesión.

Ciudad Universitaria, 24 de abril de 2023.

Mg. GUSTAVO SANDOVAL PEÑA
(PRESIDENTE)

Dra. MONICA PAREDES ANAYA
(ASESORA)

Mg. LEYDA CABRERA CORDOVA
(MIEMBRO)

Blgo. DANIEL ORE CHAVEZ
(MIEMBRO)



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
(Universidad del Perú, DECANA DE AMÉRICA)
FACULTAD DE CIENCIAS BIOLÓGICAS

INFORME DE EVALUACION DE ORIGINALIDAD

ESCUELA PROFESIONAL : GENÉTICA Y BIOTECNOLOGÍA

DIRECTOR DE LA ESCUELA PROFESIONAL : Dra. MÓNICA PAREDES ANAYA

OPERADOR DEL PROGRAMA INFORMÁTICO DE SIMILITUDES: Dr. DAN ERICK VIVAS RUIZ

DOCUMENTO EVALUADO : Tesis para optar el Título Profesional

AUTOR DEL DOCUMENTO : **ROMANI VASQUEZ, ANA CECILIA**

TÍTULO DEL DOCUMENTO : Detección de fusiones génicas en datos de secuenciación de RNA en muestras de pacientes con melanoma lentiginoso acral del INCan-México

FECHA DE RECEPCIÓN DEL DOCUMENTO : 13/03/23

FECHA DE APLICACIÓN DEL SISTEMA INFORMÁTICO : 13/03/23

SOFTWARE UTILIZADO :

TURNITIN (X)

ITHENTICATE ()

OTRO

CONFIGURACION DEL PROGRAMA DE SIMILITUDES

Excluye textos entrecomillados (X)

Excluye bibliografía (X)

Excluye cadenas menores a 40 palabras (X)

Otro criterio () (especificar)

PORCENTAJE DE SIMILITUDES : **tres por ciento (3%)**

FUENTES ORIGINALES DE LAS SIMILITUDES ENCONTRADAS : UNMSM (1%)
hdl.handle.net (1%)
saber.ucv.ve (1 %)
Otros (< 1%)

OBSERVACIONES:

Ninguna

CALIFICACIÓN DE ORIGINALIDAD:

Documento cumple criterios de originalidad sin observaciones (X)

Documento cumple criterios de originalidad con observaciones ()

Documento no cumple criterios de originalidad ()

FECHA DEL INFORME : 14/03/23

FIRMA DEL EVALUADOR

* Tesis para grado o título de pregrado

** Apellidos y nombres completos

*** En letras y números

**** Indicarlas en orden decreciente y su respectivo porcentaje

FIRMA DEL DIRECTOR EP

Dedicatoria

Le dedico este trabajo a toda mi familia, por siempre haberme dado ánimo, comprensión y amor durante el transcurso de mi carrera profesional.

A mi madre Eugenia Vasquez y mi padre Joaquín Romani, que han estado a mi lado apoyándome en todos los aspectos posibles para lograr mis objetivos profesionales. Les agradezco mucho el haberme permitido estudiar una carrera científica. Los amo mucho.

A mis hermanas Sara, Gina, Lorena y Roxana, que han crecido junto a mí y las que forman parte de muchos recuerdos y experiencias inolvidables.

A mis amigos del colegio y la universidad con los que compartí experiencias de aprendizaje.

Espero poder seguir aprendiendo con ustedes. También, agradezco a mis mentores David Requena, Luis Soto, Daniela Robles, Obed Ramírez que me ayudaron en el aspecto académico y profesional y me siguen impulsando a continuar con mi carrera.

Agradecimientos

Quiero agradecer a mis asesores Mónica Paredes, Daniela Robles y Obed Ramírez, por el apoyo y asesoramiento durante la realización de la tesis. También por la paciencia y atención que tuvieron conmigo.

El presente trabajo de tesis fue desarrollado como parte de las actividades del International Laboratory for Human Genome Research de la Universidad Nacional Autónoma de México (UNAM).

Para la realización de este trabajo se contó con el apoyo de Luis A. Aguilar, Alejandro de León y Jair García del Laboratorio Nacional de Visualización Científica Avanzada.

Lista de abreviaturas

- ACACA: Gen de la acetil-CoA carboxilasa 1
- ADYC2: Adenilato ciclasa 2
- ALM: Melanoma lentiginoso acral
- AS: Empalme alternativo
- BRAF: serina/treonina-proteína quinasa B-Raf
- BRX1: gen del homólogo de la proteína de biogénesis del ribosoma
- CM: Melanoma cutáneo
- DLG2: gen homólogo 2 de discos grandes
- DNA: Ácido Desoxirribonucleico
- EIF2B3: Gen de la subunidad gamma del factor de iniciación de la traducción eucariótico 2B
- GOLGA4: Golgina A4
- GPNMB: Glicoproteína transmembrana NMB
- GRCh37: Consorcio de referencia del genoma Estructura humana 37
- GSE1: Elemento supresor genético 1
- LncRNA: Ácido Ribonucleico largo no codificante
- MAP2K5: Proteína quinasa quinasa 5 activada por mitógeno de doble especificidad
- MAPK: Proteína quinasa activadora de mitógenos
- MAST2: Serina/treonina proteína quinasa asociada a los microtúbulos 2
- NAV3: Navegador neuronal 3
- NEK1: Quinasa 1 relacionada a NIMA
- NLK: Quinasa similar a la proteína nemo
- PICALM: Proteína de ensamblaje de clatrina de unión a fosfatidilinositol
- RNA: Ácido Ribonucleico

- RNAm: Ácido Ribonucleico mensajero
- ROCK: Quinasa asociada a Rho
- RP1: Proteína de retinitis pigmentosa 1
- RP11: Proteína de retinitis pigmentosa 11
- TPRX1: Tetrapéptido repetido homeobox 1
- TS: Genes supresores

Índice general

Dedicatoria	1
Agradecimientos	2
Lista de abreviaturas	3
Resumen	8
Abstract	9
Introducción	10
Marco teórico	13
2.1. Definición de cáncer y melanoma	13
2.1.1 Definición del melanoma lentiginoso acral	14
2.2. Recombinaciones cromosómicas.....	15
2.3. Empalme del RNA.....	17
2.3.1 Empalme alternativo	17
2.4. Origen del cáncer y su relación con la recombinación cromosómica y AS.....	18
2.5. Marcadores tumorales	20
2.6. Fusiones génicas	21
2.6.1 Fusiones génicas en melanoma.....	21
2.7. Proteína quinasas	22
2.7.1 Mutaciones de quinasa en melanoma y ALM.....	23
2.7.2 Fusiones génicas quinasa en cáncer	24
2.8. Detección de fusiones a nivel de DNA y RNA	24

2.8.1 Principio teórico de detección de fusiones génicas en datos de secuenciación de RNA mediante las herramientas bioinformáticas	25
Objetivos	27
Hipótesis	27
Materiales y métodos	28
5.1. Técnicas e Instrumentos:	28
5.2. Análisis de calidad y preprocesamiento de los datos crudos.....	29
5.3. Detección de las fusiones génicas	29
5.4. Análisis exploratorio (A.E) de los resultados de Fuseq y STAR-Fusion.....	30
5.5. Comparación de los resultados en Fuseq y STAR-Fusion	31
5.6. Búsqueda de genes proteína quinasa.....	32
5.7. Validación <i>in silico</i> de las fusiones génicas quinasa.....	32
5.8. Predicción de dominios funcionales de las fusiones génicas quinasa.....	32
5.9. Disponibilidad de códigos	33
Resultados	34
6.1. Análisis exploratorio de las características de los archivos fastq.....	34
6.2. Análisis exploratorio de los resultados de los programas de Fuseq y STAR-Fusion.....	34
6.3. Detección de proteína quinasas del genoma humano	46
6.4. Validación <i>in silico</i> de las fusiones génicas quinasa	49
6.5. Predicción de los dominios funcionales de las fusiones génicas quinasa.....	52

Discusión	54
Conclusiones	61
Recomendaciones	61
Bibliografía	62
Anexos	75

Resumen

Las fusiones génicas son genes híbridos producto de recombinaciones cromosómicas o empalme alternativo aberrante. Las fusiones génicas de quinasas activadas se han encontrado en múltiples tipos de cáncer. Existen varias proteínas quinasas usadas como marcadores tumorales debido a su alta frecuencia mutacional en pacientes con cáncer. Además, se ha encontrado fusiones génicas quinasa relacionadas intrínsecamente a la alteración de procesos como la regulación del crecimiento y proliferación celular. El principal elemento para la funcionalidad de una proteína quinasa es el dominio quinasa, su función se controla comúnmente con un dominio regulador de la propia proteína quinasa. El melanoma acral lentiginoso (ALM) es un subtipo de melanoma cutáneo poco estudiado. Hasta la fecha el número de fusiones génicas encontradas en ALM es relativamente bajo y se desconoce su asociación a este tipo de cáncer. Por lo tanto, en el presente estudio se analizaron datos de secuenciación de RNA de 112 muestras de biopsias de pacientes mexicanos con ALM para la detección e identificación de fusiones génicas. Luego, se analizaron los dominios proteicos de las fusiones génicas quinasa. Los resultados mostraron 337 fusiones génicas en 80 muestras, entre las que se encontraron fusiones génicas quinasa con presencia de MAP quinasas. Se identificaron 21 fusiones génicas quinasa en 19 muestras tumorales, las cuales no han sido reportadas previamente en ALM. La predicción de los dominios a partir de la traducción de las fusiones génicas quinasa evidenció 6 fusiones proteína quinasa con dominios proteína quinasa completos, indicando una alta probabilidad del funcionamiento de la actividad quinasa. En conclusión, el estudio muestra la existencia de fusiones génicas que son potenciales marcadores para ALM y fusiones génicas que podrían estar involucradas en el desarrollo de ALM.

Palabras clave: cáncer, fusión génica, melanoma acral lentiginoso, proteína quinasa, secuenciación de RNA

Abstract

Gene fusions are hybrid genes produced by chromosomal recombination or aberrant alternative splicing. Activated kinase fusion genes have been found in multiple types of cancer. There are several protein kinases used as tumoral biomarkers due to their high mutation frequency in cancer patients. Furthermore, these kinase fusion genes have been found to be intrinsically related to processes such as growth regulation and cell replacement. The main element of the protein kinase function is the kinase domain which is normally controlled by a regulatory domain of the protein kinase itself. Acral lentiginous melanoma (ALM) is a poorly studied subtype of cutaneous melanoma. Currently, the number of fusion genes found in ALM is relatively low and their association with this type of cancer is unknown. Therefore, in the present study, RNA sequencing data from 112 samples from Mexican patients with ALM were analyzed for the detection and identification of gene fusions. Then, protein domains of the protein kinase gene fusions found were analyzed. The results showed 337 gene fusions in 80 samples, among which protein kinase gene fusions with the presence of MAP kinases were found. We identified 21 protein kinase gene fusions in 19 tumor samples which have not been previously reported in ALM. Domain prediction from translation of the protein kinase gene fusions showed 6 protein kinase fusions with complete protein kinase domain, indicating a high probability of kinase activity. In conclusion, the study shows the existence of gene fusions that are potential markers for ALM and gene fusions that could be involved in the development of ALM.

Keywords: cancer, gene fusion, acral lentiginous melanoma, protein kinase, RNA-seq

Introducción

El melanoma acral lentiginoso (ALM) es un subtipo de melanoma que se caracteriza por crecer en la piel de las palmas de las manos, las plantas de los pies y los dedos (Basurto-Lozada *et al.*, 2021). ALM presenta una de las peores prognosis con respecto a los otros tipos de melanoma cutáneo (Kuchelmeister *et al.*, 2000), siendo una de las principales causas el diagnóstico tardío (Susok & Gambichler, 2022). Los casos de ALM representan alrededor del 3% de los nuevos casos de melanoma en USA (Huang *et al.*, 2020) y es el menos común de los melanomas cutáneos diagnosticado entre los individuos de ascendencia europea, comprendiendo solo aproximadamente un 2.5% del total de casos (Susok & Gambichler, 2022). Sin embargo, en países con una historia demográfica compleja como México y Perú, el ALM representa entre un 35 a 44% de los melanomas diagnosticados en estudios de más de 400 casos (Coras *et al.*, 2013; Lino-Silva *et al.*, 2016), presentando una mayor proporción de melanoma acral lentiginoso en comparación con las poblaciones de USA y Europa. No obstante, la mayor parte de estos estudios a gran escala en Latinoamérica está basada en diagnósticos morfológicos e histopatológicos sin uso de marcadores genéticos. En el caso del ALM, el uso de marcadores genéticos podría permitir un diagnóstico más temprano y un menor número de diagnósticos errados (Basurto-Lozada *et al.*, 2021).

Las fusiones génicas son dos genes unidos producto de recombinaciones cromosómicas o empalme alternativo que pueden generar proteínas con cambios en su funcionalidad (Taniue & Akimitsu, 2021). Entre estas se encuentran las fusiones génicas oncogénicas, las cuales son potenciales marcadores de diagnóstico y pronóstico en cáncer por su prevalencia en pacientes con tumores (Chang *et al.*, 2019; Schram *et al.*, 2017). Las

fusiones génicas oncogénicas han sido estudiadas desde la década del 60, siendo las primeras fusiones génicas detectadas mediante cariotipos (Schram *et al.*, 2017). Por otro lado, las fusiones génicas detectadas a partir de datos de secuenciación de RNA son estrictamente fusiones a nivel de RNA, las cuales se pueden originar tanto en eventos de recombinación cromosómica (fusiones génicas), como también en el empalme (transcritos fusión).

En años recientes, los estudios de detección de fusiones génicas se han realizado a partir de los datos de secuenciación de RNA (Wang *et al.*, 2013). Actualmente existe un gran número de herramientas bioinformáticas usadas para la detección de fusiones génicas en secuenciación de RNA, tales como Fuseq (Vu *et al.*, 2018) y STAR-Fusion (Haas *et al.*, 2017). Además, existen también bases de datos como FusionGDB, la cual presenta fusiones génicas predichas mediante algoritmos de *deep learning* (Kim & Zhou, 2019) y, COSMIC, la cual contiene fusiones génicas que fueron manualmente curadas a partir de publicaciones anteriores (Tate *et al.*, 2019).

Las fusiones génicas oncogénicas que han tenido un mayor impacto en el diagnóstico y desarrollo de terapias son las fusiones génicas quinasa (Mitelman *et al.*, 2007; Niu *et al.*, 2013; Shaw *et al.*, 2013; Taniue & Akimitsu, 2021). Las proteína quinasas son enzimas involucradas en un gran número de procesos de regulación y activación mediante cascadas de fosforilación, tales como la vía de la proteína quinasa activada por mitógenos (MAPK) (Manning *et al.*, 2002). Su actividad catalítica se debe principalmente al dominio quinasa presente en la enzima, siendo el núcleo principal de los procesos de fosforilación (Arter *et al.*, 2022). Frecuentemente el dominio quinasa se encuentra acompañado de dominios de regulación encargados de regular su actividad (Arter *et al.*, 2022; Manning *et al.*, 2002). Por lo tanto, si el dominio quinasa o sus reguladores adquieren funciones transformadas por mutaciones, podría generar la sobreexpresión de la actividad quinasa guiando a la malignidad (Paul & Mukhopadhyay, 2004).

Distintos tipos de fusiones génicas quinasa se han observado con una alta frecuencia en pacientes con cáncer. Por ejemplo, la fusión génica quinasa *BCR-ABL* (*ABL* pertenece a la familia de las tirosina quinasas) se encontró en pacientes con leucemia mieloide crónica (CML) (Taniue & Akimitsu, 2021) o translocaciones de *BRAF/CRAF* (serina/treonina quinasas) en tumores sólidos (Schram *et al.*, 2017). En melanoma cutáneo se han detectado fusiones con quinasas como *CRAF* (Schram *et al.*, 2017). En ALM, se han identificado fusiones génicas con *PAK-1*, perteneciente a las serina/treonina quinasas, en una muestra de 35 pacientes (Hayward *et al.*, 2017). Con base en lo anterior, la detección de fusiones génicas en pacientes con ALM podría conllevar a la obtención de distintos tipos de marcadores tumorales, principalmente de terapia y diagnóstico.

En resumen, el estudio de fusiones génicas en muestras de ALM de pacientes de países con una alta proporción de casos como México y Perú, donde la mayor parte de estudios a gran escala son basados en diagnósticos morfológicos e histopatológicos, aportaría información novedosa sobre potenciales marcadores genéticos en ALM. En el presente trabajo, el enfoque se centró en la detección de fusiones génicas mediante programas bioinformáticos en datos de secuenciación de RNA en una cohorte de pacientes con ALM (n = 112) del Instituto Nacional de Cancerología (INCan) en México. Las fusiones génicas principalmente analizadas fueron las que presentan secuencias de genes proteína quinasa, debido a su importancia funcional y clínica, lo que proporcionaría potenciales marcadores de diagnóstico de ALM que en un futuro podrían llegar a ser usados como objetivos terapéuticos.

Marco teórico

2.1. Definición de cáncer y melanoma

El cáncer es un gran conjunto de enfermedades caracterizado por el desarrollo de células anormales que se dividen de manera incontrolable y presentan la habilidad de infiltrarse y destruir tejido conformado por células normales (Fior & Zilhão, 2019). Las células anormales que generan el cáncer presentan mutaciones que desencadenan el fenotipo de crecimiento ilimitado y de propagación en tejidos foráneos (Fior & Zilhão, 2019). Estas mutaciones se encuentran principalmente en proteínas involucradas en mecanismos regulatorios de replicación del DNA y proteínas relacionadas con el crecimiento y continuidad del ciclo celular (Fior & Zilhão, 2019).

El melanoma es un tumor maligno que se origina en los melanocitos. Se caracteriza por presentar metástasis desde estadios muy tempranos del desarrollo del tumor (Oba & Woodman, 2021). El melanoma es clasificado tradicionalmente por la localización del tumor: 1) el melanoma cutáneo (CM) típicamente crece de los melanocitos centro de la capa epidérmica, y muy poco frecuentemente de la dermis (Rabbie *et al.*, 2019), 2) el melanoma mucoso (MM) surge de los melanocitos en el revestimiento mucoso de los tejidos internos, como los tractos sinonasal, anorrectal y vulvovaginal (Rabbie *et al.*, 2019), y 3) el melanoma ocular, el cual está dividido en melanoma uveal (UM) y el melanoma conjuntival (CjM) (Rabbie *et al.*, 2019; van Poppelen *et al.*, 2021). Actualmente existe una nueva clasificación de un subtipo de melanoma denominado “Spitz melanoma” que se caracteriza por los cambios histológicos similares a los nevos Spitz y una fusión génica oncogénica que presenta proteína quinasa (Elder *et al.*, 2020).

Dentro de CM, la clasificación clínico-patológica comúnmente usada comprende cuatro subtipos mayores: El melanoma de extensión superficial (SSM), el maligno lentigo

melanoma (LMM), el melanoma nodular (NM) y el melanoma acral lentiginoso (ALM) (Kuchelmeister *et al.*, 2000). Sin embargo, también existen otros subtipos en una frecuencia mucho menor, por ejemplo, una variante histológica relativamente rara de CM es el melanoma desmoplásico (DM) que consiste en células tumorales fusiformes rodeadas por una matriz de colágeno fibroso grueso (Nicolson & Han, 2019; Rabbie *et al.*, 2019). Algunos autores consideran al melanoma acral (AM) fuera de CM debido a su origen en la piel glabra (sin pelos) de la palma, plantas y lechos ungueales (Oba & Woodman, 2021), sin embargo, en esta tesis lo consideramos dentro de CM ya que surge de la piel.

Distintos análisis a escala del genoma y transcriptoma han mostrado diferencias moleculares en los melanomas tradicionalmente clasificados por localización anatómica e histopatología (Oba & Woodman, 2021). En el 2018, la organización mundial de la Salud (WHO) generó una nueva clasificación basada en localización, histopatología y características genéticas. Esta nueva clasificación fue agrupada en causada por daño solar (CSD) y no causada por daño solar (no CSD), siendo DM y maligno lentigo melanoma (LMM) incluidos en el subgrupo de alto CSD. En el subgrupo de bajo CSD se encuentra el melanoma de extensión superficial (SSM). Por otro lado, en el grupo no CSD se encuentran AM, MM, UM, y los Spitz melanomas (SMs).

Con respecto al CM, su origen se basa principalmente en las interacciones entre factores genéticos y ambientales (McMeniman *et al.*, 2020). Los principales factores ambientales son la exposición solar agudo o crónica (McMeniman *et al.*, 2020). Con respecto a los factores genéticos, los principales genes reconocidos son *CDKN2A* y el *CDK4* (Goldstein *et al.*, 2000). Se estima que entre 15% y 25% de casos familiares presentan mutaciones en *CDKN2A* (Goldstein *et al.*, 2000).

2.1.1 Definición del melanoma lentiginoso acral

El melanoma acral lentiginoso (ALM) es un subtipo de melanoma cutáneo (CM) que crece en la piel de las palmas de las manos, las plantas de los pies y los dedos. La

denominación 'lentiginoso' es debido a su crecimiento radial, que aparece antes de que el tumor empiece a invadir zonas de la dermis (Basurto-Lozada *et al.*, 2021). En algunos casos, los autores usan como término intercambiable a melanoma acral (AM) y melanoma lentiginoso acral (ALM) (Basurto-Lozada *et al.*, 2021), sin embargo, se ha reconocido que el ALM tiene características histológicas específicas.

Con respecto a su etiología, ésta presenta controversias en las hipótesis planteadas. Por ejemplo, la susceptibilidad de heridas mecánicas en las zonas donde se desarrolla ALM se ha hipotetizado como un posible desencadenante para su formación. Existen estudios que indican una asociación entre trauma previo y el desarrollo de enfermedad en el sitio del tumor (Möhrle & Häfner, 2002; Troyanova, 2002; Zhang *et al.*, 2014). Sin embargo, otros estudios no han encontrado esta asociación (Costello *et al.*, 2017; Kaskel *et al.*, 2000; Murph, 2015). Por otro lado, los estudios de secuenciación en ALM han identificado mutaciones, promotoras del desarrollo tumoral, en un gran número de genes. Por ejemplo, se detectó en muestras de tumores de ALM ganancias de ciclina dependiente de quinasa 4 (*CDK4*) y ciclina D1 (*CCND1*) o pérdidas de *CDKN2A* (Kong *et al.*, 2017), generando una desregulación en la vía CDK4. También, se han observado aberraciones en la telomerasa transcriptasa inversa (*TERT*) de pacientes con ALM (Liang *et al.*, 2017). De esta manera, su etiología es aún controversial y con respecto a estudios anteriores sería multifactorial.

2.2. Recombinaciones cromosómicas

La recombinación cromosómica es un aspecto importante en la estabilidad del genoma y reparación del daño al DNA en la mitosis (Aguilera & Rothstein, 2007). Estas recombinaciones se dan de manera natural entre cromosomas homólogos cuando suceden rupturas de doble cadena. Para reparar esta ruptura se requiere un referente, el cual es provisto por el cromosoma homólogo (Aguilera & Rothstein, 2007).

Las recombinaciones cromosómicas no homólogas pueden generar cambios funcionales en la célula e inestabilidad genómica. Este tipo de recombinación cromosómica también es llamada translocación (Figura 1A). Las translocaciones consisten en un intercambio de partes entre dos cromosomas no homólogos. Las translocaciones pueden ser recíprocas o no recíprocas (Dai *et al.*, 2018). Las recíprocas comprenden un intercambio mutuo entre dos cromosomas no homólogos, mientras las no recíprocas comprenden solo la reubicación de una de las partes de un cromosoma no homólogo (Dai *et al.*, 2018). Entre todos los reordenamientos, las translocaciones presentan una mayor escala de anomalías debido a que pueden generarse entre regiones distantes del genoma o dentro del mismo cromosoma, a diferencia de las inversiones, duplicaciones y deleciones las cuales ocurren solo dentro de un mismo cromosoma (Dai *et al.*, 2018).

Las deleciones, duplicaciones e inversiones son rearrreglos intracromosómicos (Figura 1B). Las deleciones consisten en la escisión de partes de secuencias de un gen. Mientras, las duplicaciones en tándem consisten en un número de copias de exones dentro del mismo gen, generando exones subsecuentes. Las inversiones son reordenamientos en la ubicación respectiva de la secuencia, colocando la secuencia en un orden reverso, los dos tipos de inversión son pericéntrica y paracéntrica. Estas últimas no incluyen al centrómero en los eventos de ruptura, solo ocurriendo en un solo brazo del cromosoma. Mientras, las pericéntricas incluyen al centrómero en el evento de ruptura y una ruptura en cada brazo del cromosoma.

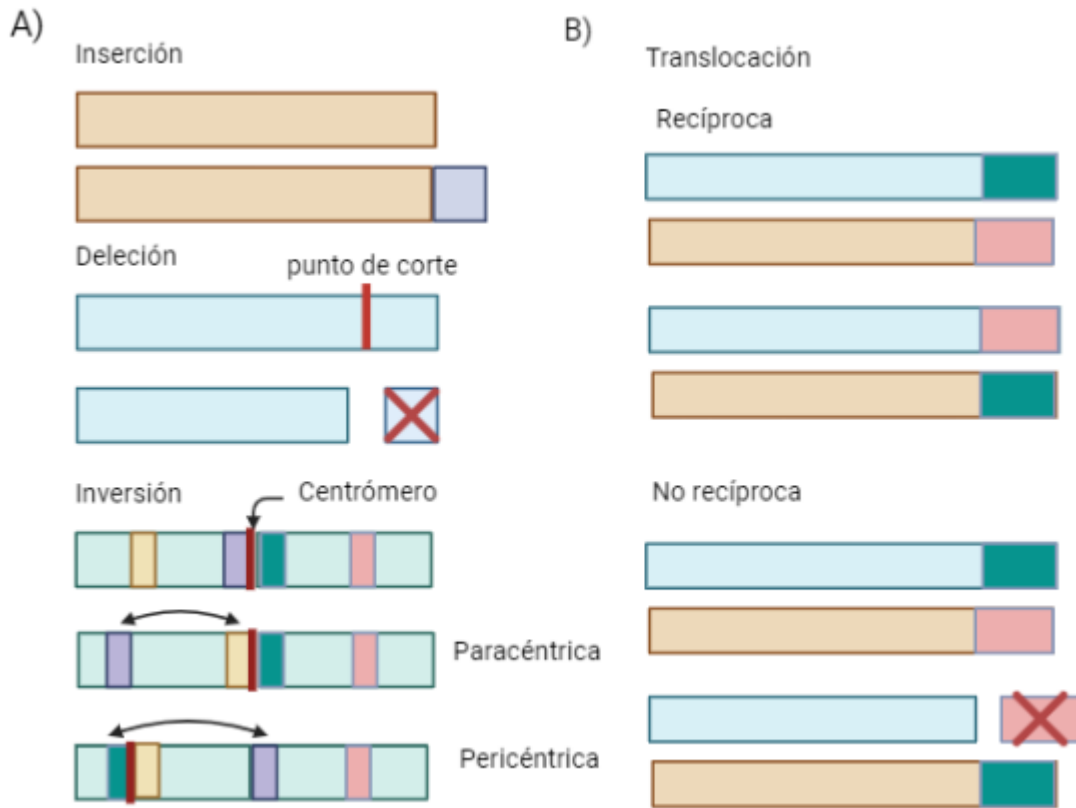


Figura 1: Tipos de recombinaciones cromosómicas. (A) Cambios intracromosómicos. (B) Cambios intercromosómicos. Fuente: Autoría propia.

2.3. Empalme del RNA

Después de la transcripción, la cual permite el paso de información de DNA a RNA, procede al empalme del RNA. Este proceso consiste en la escisión de los intrones del preRNAm por medio del espliceosoma y ligación de los exones de la cadena (Figura 2). En promedio, un gen humano presenta de 8 a 10 exones, los cuales están separados por intrones con un tamaño de 10 a 100 veces mayor que los exones.

2.3.1 Empalme alternativo

Se denomina empalme alternativo (AS, alternative splicing en inglés) a las distintas combinaciones que generan transcritos de RNAm alternativos a partir de la retención o exclusión de diferentes exones.

Los diferentes mecanismos de corte y empalme incluyen la retención de intrones, omisión de exones, sitios de empalme 5' o 3' alternativos, y exones mutuamente excluyentes (B. Wang & Lee, 2018). Además, existen otros mecanismos que están más relacionados al proceso de transcripción, dependiendo de los factores de transcripción (TF) y no al proceso de corte y empalme como la selección de promotores alternativos (Wang & Lee, 2018).

El AS genera una mayor complejidad para los estudios en células eucariotas, debido a la expansión del repertorio de codificación de proteínas, generando isoformas que pueden tener actividades biológicas estructurales distintas (Lodish *et al.*, 2016; Pan *et al.*, 2008). Además, la regulación coordinada de AS presenta un papel clave en la diferenciación y especialización celular (Wang & Lee, 2018) indicando que no es un proceso generado de manera estocástica. Según estudios de micromatriz 'sensible al empalme' y secuenciación profunda, el 95% de los genes humanos transcritos se someten a AS (Pan *et al.*, 2008); por lo que es un mecanismo de alta frecuencia.

2.4. Origen del cáncer y su relación con la recombinación cromosómica y AS

La pérdida de regulación celular que origina el cáncer se debe a daño genético, principalmente por mutaciones en oncogenes y genes supresores de tumores (TS) (Lodish *et al.*, 2016; Fior & Zilhão, 2019). Los primeros son genes involucrados con el crecimiento y división celular, al ser alterados pueden presentar cambios que generan un exceso de su función (Lodish *et al.*, 2016; Fior & Zilhão, 2019). Mientras, los TS limitan el crecimiento, por lo que mutaciones que generen pérdida de funcionalidad en estos genes producen una división celular inapropiada (Lodish *et al.*, 2016; Levitt & Hickson, 2002). Los TS se dividen en dos categorías, los puntos de control y los de mantenimiento (Levitt & Hickson, 2002). Los puntos de control son los que supervisan si la célula puede seguir proliferando, en

cambio, los genes cuidadores o de mantenimiento se encargan de la protección de la integridad del genoma mediante mecanismos de reparación (Levitt & Hickson, 2002).

La plasticidad celular ofrecida por las recombinaciones y AS permite la producción de isoformas, proteínas aberrantes o proteínas quimeras que pueden favorecer el crecimiento celular o la propagación (Dai *et al.*, 2018). Consecuentemente, estos mecanismos están en una intrínseca relación con el desarrollo y caracterización del cáncer. Estudios en cariogramas de distintos tipos de células cancerosas han mostrado una mayor frecuencia de aberraciones cromosómicas a diferencia de células normales (Gisselsson *et al.*, 2002; Ma & Wan, 2000; Rajagopalan & Lengauer, 2004). El estudio de Boffetta *et al.* (Boffetta *et al.*, 2007) propuso que un alto nivel de aberraciones cromosómicas podría ser un marcador de riesgo de cáncer al analizar el número de aberraciones ocurridas en muestras de pacientes de distintos tipos de cáncer. También, se ha reportado empalmes aberrantes y la alteración en la proporción de los transcritos alternativos en neoplasias malignas (Dong & Chen, 2020).

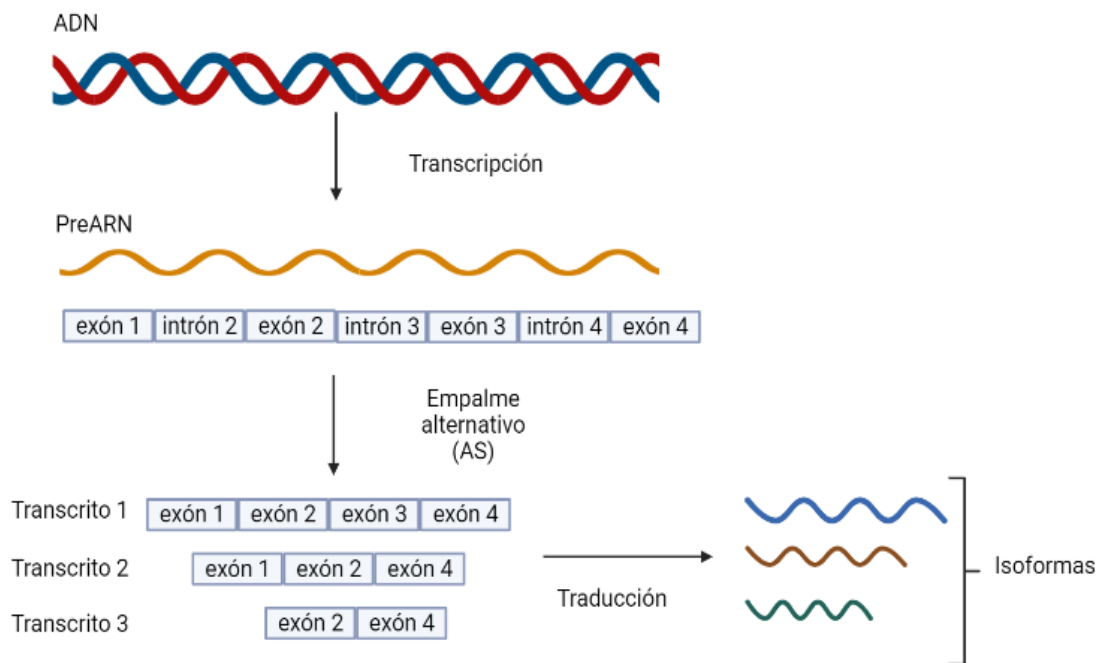


Figura 2: Proceso de transcripción, empalme y traducción. Fuente: Autoría propia.

2.5. Marcadores tumorales

Un marcador tumoral es cualquier indicio que está presente o es producido por las células cancerosas u otras células en respuesta al cáncer (*Tumor Markers - NCI, 2021*). Los marcadores tumorales han sido tradicionalmente proteínas o sustancias que se han producido en altas cantidades en células cancerosas a diferencia de células normales (*Tumor Markers - NCI, 2021*). No obstante, estos incluyen un variado tipo de sustancias como antígenos de la superficie celular, proteínas citoplasmáticas, enzimas, hormonas, oncogenes, entre otros (Sharma, 2009).

Una de las primeras características que se toma en cuenta para la selección de un marcador es su frecuencia en la población estudiada (Vine, 1994; Wang *et al.*, 2020). En cáncer, el gen BCR-ABL, descubierto en 1960, fue considerado de gran interés como marcador en CML debido a su alta frecuencia (95 %) en los pacientes (Druker, 2008). Actualmente, se propone genes como marcadores tumorales para un tipo o subtipo de cáncer principalmente basándose en la frecuencia de sus patrones mutacionales y su expresión diferencial con personas sanas. Para su uso clínico, un marcador tumoral debe presentar principalmente una alta especificidad y sensibilidad, además de estar presentes de manera continua en los diagnósticos.

2.6. Fusiones génicas

Las fusiones génicas son dos genes los cuales han sido transcritos dentro de una sola secuencia de RNAm (Taniue & Akimitsu, 2021). Su origen se debe principalmente a las distintas translocaciones intra e inter cromosómicas. No obstante, los otros tipos de recombinaciones cromosómicas también generan fusiones génicas. Además, a nivel de RNAm, el empalme alternativo (AS) aberrante puede generar transcritos fusionados de distintos genes (B. Wang & Lee, 2018).

Las fusiones génicas presentan distintas características y funciones dependiendo de la localización de la fusión, por lo que estas no necesariamente llevan al desarrollo del tumor. Este es el caso de las fusiones que se generan en regiones no codificantes o secuencias codificantes fuera del marco de lectura, las cuales es muy probable que no sean funcionales (Kim & Zhou, 2019; Taniue & Akimitsu, 2021). En cambio, un transcrito formado por fusiones génicas dentro del marco de lectura puede convertirse en una proteína fusión que actúe como un potente controlador (driver) oncogénico.

La estimación de morbilidad causada por fusiones génicas en cáncer es del 20%, aunque los estudios presentan distintos porcentajes de presencia de fusiones génicas con respecto al tejido de desarrollo de la tumorigénesis (Mitelman *et al.*, 2007). Por ejemplo, en CML, se ha encontrado el cromosoma Filadelfia (*BCRL-ABL 1*) en el 95% de los pacientes (Druker, 2008), las fusiones de *RET* en carcinoma papilar (PTC) se han encontrado en 80% de tumores asociados a radiación (Nikiforov, 2002; Ullmann *et al.*, 2022), y la fusión génica *KIAA1549-BRAF* se ha encontrado en un 66% de los casos de astrocitoma pilocítico (Faulkner *et al.*, 2015).

2.6.1 Fusiones génicas en melanoma

Las fusiones génicas en melanoma que se han observado con mayor frecuencia son las que presentan un gen proteína quinasa como *RAF1* (Schram *et al.*, 2017). La estimación

de ocurrencia de las fusiones *BRAF* en melanoma son entre el 2.6% al 6.7% de casos (Botton *et al.*, 2019). Además, en el subtipo Spitz melanoma se ha encontrado fusiones génicas con proteína quinasas como *ROS1*, *MAPK*, *NTRK1* y *NTRK3* (Quan *et al.*, 2020). También, un estudio entre distintos tipos de melanoma mostró que hay fusiones génicas que presentan *RAF1*, *RAC1*, *MAP2K2*, *GRM3*, *MAP3K9*, *TRRAP38* y *PLCB4* y genes supresores de tumores como *NF1*, *PARK2*, *TP53*, *PTEN* y *PPP6C* (Hayward *et al.*, 2017). Estos descubrimientos indicarían una posible asociación del desarrollo de melanoma con las fusiones génicas, en especial con las que presentan genes proteína quinasa.

2.7. Proteína quinasas

Las quinasas son enzimas encargadas de catalizar de la transferencia de grupos fosfato de moléculas donadoras de fosfato de alta energía a substratos específicos, por lo que permiten la activación o desactivación de distintos procesos (Lodish *et al.*, 2016; Manning *et al.*, 2002). Dentro del genoma humano, las quinasas se han clasificado en 44 familias (Manning *et al.*, 2002). Una de las más grandes superfamilias de quinasas son las proteína quinasas, las cuales tienen como substrato específico a las proteínas (Manning *et al.*, 2002; Zhao *et al.*, 2014). Los genes de estas proteínas se encuentran con una alta conservación en los organismos eucariotas debido a su rol crítico en la activación enzimática y señalización intracelular que cumplen junto a las fosfatasas (Zhao *et al.*, 2014). En el humano, se han identificado 518 genes de proteína quinasas (Manning *et al.*, 2002).

Familias largamente estudiadas dentro de las proteína quinasas son la familia tirosina quinasa (Paul & Mukhopadhyay, 2004; Robinson *et al.*, 2000) y la familia serina/treonina quinasa (Kumar *et al.*, 2017; Roskoski, 2019). La familia tirosina quinasa tiene funciones principales en la regulación de múltiples aspectos del organismo como las señales de crecimiento entre células, diferenciación, adhesión, motilidad y muerte (Robinson *et al.*, 2000). En contraste, la familia serina/treonina quinasa presenta proteína quinasas

dependientes de ciclina (CDK) y las MAP quinasas (MAPK), las cuales son fundamentales para la regulación y expresión de un gran número de componentes en las etapas del ciclo celular (Roskoski, 2019). Además, las CDK presentan miembros que participan en actividades como la transcripción, metabolismo y función neuronal (Roskoski, 2019).

2.7.1 Mutaciones de quinasa en melanoma y ALM

Los genes de las proteína quinasas, las cuales cumplen un amplio rango de funciones en el humano, pueden ser alterados por mutaciones que ocasionan el desarrollo de múltiples enfermedades, incluyendo al cáncer (Robinson *et al.*, 2000; Roskoski, 2019). Una proteína de gran relevancia en cáncer es BRAF, la cual es una serina/treonina quinasa, miembro de la familia RAF quinasa, la cual es parte de la cascada de señalización de las serina/treonina quinasa RAS/RAF/MEK/ERK, también denominada vía MAPK. Esta vía regula el crecimiento celular, la supervivencia y diferenciación de las células mediante receptores tirosina quinasa y receptores acoplados a proteínas G (Lodish *et al.*, 2016; Manning *et al.*, 2002).

La mutación BRAF^{V600E} es diagnosticada aproximadamente en la mitad de los pacientes con melanoma metastásico (Ascierto *et al.*, 2012), por lo que es utilizada como objetivo terapéutico. Al administrar inhibidores BRAF a los pacientes con dicha mutación se logra la recesión parcial o completa del tumor en la mayoría de estos pacientes. No obstante, en algunos casos se observa una respuesta inicial, pero eventualmente se desarrolla resistencia a los inhibidores de BRAF (Solit & Rosen, 2011). Sin embargo, en ALM, las mutaciones de BRAF tienen una frecuencia menor (9.5%) comparadas con respecto a otros subtipos de melanomas (Saldanha *et al.*, 2006). En ALM, se han encontrado mutaciones de genes quinasa como *KIT*, amplificaciones en *CDK4* e inactivaciones de *CDKN2A* (Yeh & Bastian, 2021).

2.7.2 Fusiones génicas quinasa en cáncer

Las fusiones génicas que presentan transcritos de gen proteína quinasa se han reportado frecuentemente en varios tipos de cáncer (Dai *et al.*, 2018; Schram *et al.*, 2017; Shaw *et al.*, 2013), incluyendo a melanoma (Botton *et al.*, 2019; Turner *et al.*, 2017). Las fusiones génicas quinasa pueden actuar como drivers oncogénicos, uno de los casos más conocidos es el de *BCR-ABL1* (*ABL* pertenece a las tirosina quinasas), donde la substitución de la secuencia del N-terminal del gen *ABL1* genera una fusión génica con una actividad quinasa constitutiva llevando a la estimulación de la proliferación celular (Paul & Mukhopadhyay, 2004). En CM, el estudio de Hayward menciona que las fusiones con la quinasa CRAF generan un incremento de la actividad de fosforilación en la vía MAPK (Hayward *et al.*, 2017).

Con respecto a ALM, se ha encontrado fusiones génicas que presentan un gen de proteína quinasa como *PAK1* (Hayward *et al.*, 2017), *NTRK1* (Turner *et al.*, 2017), *BRAF*, *NTRK3*, *ALK* y *PRKCA* (Yeh *et al.*, 2019). La fusión con *PAK1* se presentó fuera del marco de lectura, indicando que no sería funcional (Hayward *et al.*, 2017). En el caso de *NTRK1* el hallazgo se realizó a partir de la técnica FISH, no obstante, no se realizó análisis de secuenciación de RNA, por lo que no se podría confirmar su expresión y funcionalidad (Turner *et al.*, 2017). Mientras tanto, la fusión génica quinasa en *BRAF* presentaba su dominio quinasa (Yeh *et al.*, 2019). Estos resultados muestran que ALM presenta fusiones quinasa que podrían generar alteraciones en la funcionalidad de las proteína quinasas que llevarían al desarrollo tumoral.

2.8. Detección de fusiones a nivel de DNA y RNA

A través de los años, la identificación de fusiones génicas ha sido efectuada con distintas metodologías, desde el cariotipo hasta tecnologías de secuenciación (Schram *et al.*, 2017).

Los primeros hallazgos de fusiones génicas oncogénicas se descubrieron en neoplasias malignas hematológicas mediante el cariotipo. Más adelante, las técnicas de fluorescencia como FISH, permitirían mejorar la localización de los puntos de corte en los cariotipos (Schram *et al.*, 2017). Posteriormente, la llegada de la tecnología de microarreglos permitió un perfil de expresión a gran escala, generando nuevos descubrimientos. Actualmente, la detección de fusiones génicas se realiza principalmente mediante herramientas bioinformáticas en base a los datos generados por los métodos de secuenciación (Schram *et al.*, 2017). Estas herramientas han permitido una búsqueda en todo el genoma y transcriptoma secuenciado, lo que ha facilitado la detección de las fusiones, sin necesidad de información biológica y estructural (Schram *et al.*, 2017).

2.8.1 Principio teórico de detección de fusiones génicas en datos de secuenciación de RNA mediante las herramientas bioinformáticas

Las herramientas actuales de detección de fusiones génicas en datos de secuenciación de RNA comprenden dos metodologías: alineamiento y ensamble *de novo* (Haas *et al.*, 2013). La primera se basa en el alineamiento con un genoma de referencia para encontrar discordancia en las lecturas mapeadas, indicando un posible reordenamiento (Figura 3A). La evidencia de las fusiones predichas consiste en la cantidad de lecturas de soporte. Estas lecturas se dividen en lecturas de unión y lecturas de extensión (Vu *et al.*, 2018). Las primeras consisten en la superposición directa con la unión quimérica, es decir, una sección de la lectura es alineada en un gen y la otra en otro gen (Vu *et al.*, 2018). Por otra parte, en las lecturas de extensión (Figura 3B) cada par de lecturas se alinea en diferentes genes (Vu *et al.*, 2018).

La segunda metodología se basa en el ensamble *de novo* que directamente ensambla las lecturas en transcritos más largos siguiendo con la identificación de transcritos consistentes con reordenamientos cromosómicos (Haas *et al.*, 2013). Las principales

ventajas del ensamble *de novo* es la predicción de las fusiones sin la necesidad de un genoma de referencia y la identificación de isoformas mediante los transcritos producidos (Haas *et al.*, 2013). Los programas de detección de fusiones génicas pueden usar ambas tecnologías.

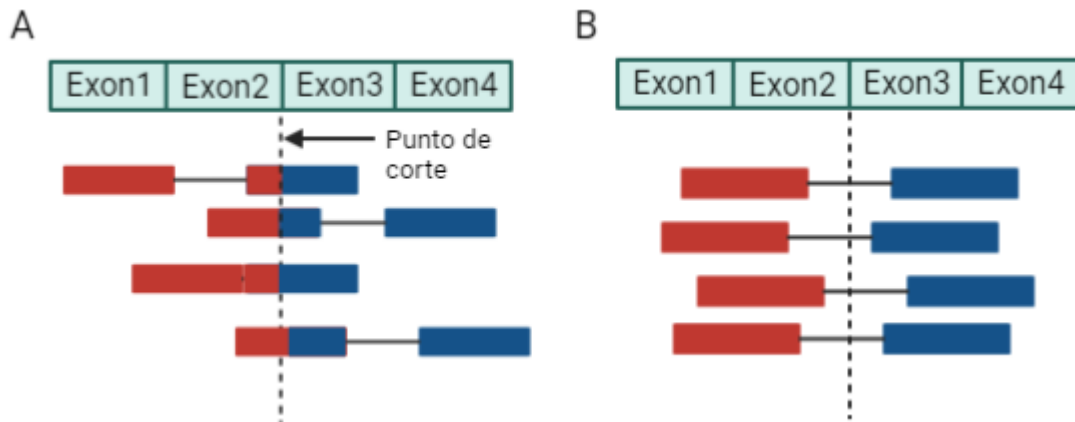


Figura 3: Tipos de lecturas de soporte. (A) Lecturas de unión (B) Lecturas de extensión.

Fuente: Autoría propia.

Objetivos

3.1. Objetivo general:

- Detectar e identificar fusiones génicas con potencial de ser marcadores tumorales en tejido tumoral de pacientes con melanoma lentiginoso acral utilizando herramientas bioinformáticas.

3.2. Objetivos específicos:

- Detectar e identificar las fusiones génicas en datos de RNA-seq. Asimismo, determinar las características generales de todas las fusiones génicas como la distribución cromosómica y su clasificación en eventos inter e intracromosómicos.
- Identificar fusiones génicas con potencial de ser marcadores tumorales mediante la determinación de la frecuencia en las muestras y la promiscuidad de los genes que las integran.
- Buscar fusiones génicas que presentan secuencias génicas de proteína quinasa y cuyos transcritos están dentro del marco de lectura para su traducción.
- Identificar fusiones génicas quinasa que presentan dominio proteína quinasa catalítico completo.

Hipótesis

Hipótesis de trabajo: Existen fusiones génicas con potencial de ser marcadores tumorales en pacientes mexicanos con melanoma acral lentiginoso.

Materiales y métodos

- Muestras biológicas: Transcriptomas de 112 muestras de tejido tumoral de pacientes que fueron atendidos en el Instituto Nacional de Cancerología (INCan) entre los meses de julio y agosto del 2019. Los pacientes se diagnosticaron con ALM por su médico dermatólogo y siguiendo los criterios histopatológicos establecidos. El permiso para el análisis de los datos de los pacientes se encuentra en el anexo 1.

➤ *Procesamiento de las muestras de los pacientes (INCan 2019) hasta la obtención de los archivos bam:*

Las muestras fueron analizadas por un hispatólogo experto, para determinar el tejido que presenta un mayor porcentaje de células tumorales. A partir de este tejido, personal del laboratorio realizó la extracción de RNA.

A continuación, se realizó la secuenciación en el Instituto Wellcome Sanger, en Cambridge, Inglaterra, con el protocolo RNA capture utilizando la tecnología Illumina.

La secuenciación de las muestras fue pareada, es decir, se tomó como molde tanto a la cadena original como a la complementaria, por lo que se realizó el proceso con dos archivos por cada muestra.

El alineamiento se realizó con la herramienta STAR y utilizando el genoma de referencia GRCh37.

5.1. Técnicas e Instrumentos

- Clúster de supercómputo LAVIS: Es un grupo de computadoras unidas que permite realizar procesos con una alta demanda computacional. Se encuentra físicamente en el campus Juriquilla de la Universidad Nacional Autónoma de México. Sus características principales son:

➤ 20 nodos

- 1000 núcleos
- 4 TB de memoria RAM
- 1 PB de almacenamiento
- HP Laptop: Dispositivo usado para el análisis y visualización de resultados mediante paquetes en lenguaje de programación R
 - Procesador AMD Ryzen 3 2200U con Radeon Vega Mobile Gfx -2.50 GHz
 - 8 GB de RAM
 - 200GB disponibles de almacenamiento

5.2. Análisis de calidad y preprocesamiento de los datos crudos

El análisis de calidad de los datos crudos se realizó con MultiQC (Ewels *et al.*, 2016) para información de la distribución de la longitud de las lecturas, medidas centrales y número de lecturas por muestra. Estas medidas nos permiten saber si la calidad de los datos es aceptable para continuar con el proceso de detección de fusiones génicas.

Los programas de detección de fusiones génicas requerían como input a archivos con el formato fastq. Por lo tanto, se realizó el preprocesamiento de los datos crudos en formato bam. La conversión de formatos se realizó a través del programa bedtools versión 2.27.1 (Quinlan & Hall, 2010) y la automatización de la generación de los archivos mediante un script en Linux.

5.3. Detección de las fusiones génicas

La detección de las fusiones génicas se realizó mediante Fuseq version 1.1.4 (Vu *et al.*, 2018) y STAR-Fusion (Haas *et al.*, 2017). El programa Fuseq v1.1.4 presenta una mayor velocidad y menor demanda computacional en comparación con programas como FusionMap, TRUP, TopHat-Fusion, SOAPfuse y JAFFA (Vu *et al.*, 2018). Fuseq v1.1.4 usa un quasimapeo denominado RapMap (Srivastava *et al.*, 2016), que permite el rápido

reconocimiento de los genes mediante las lecturas. Después, el programa presenta una serie de filtros distintos para las clases de lecturas que han permitido la detección de la fusión génica.

La generación del índice del transcriptoma de referencia con Fuseq v1.1.4 se hizo con base en el tamaño promedio de las lecturas contenidas en los archivos input. Se realizó este proceso en base de la documentación del programa, en el cual se recomienda que los k-meros presenten un tamaño menor a las lecturas.

Después, se automatizó el proceso de detección por cada muestra a través de un script, usando los comandos indicados en la documentación de Fuseq v1.1.4, los parámetros usados en todas las muestras fueron por defecto y se extrajo las lecturas soportes de cada muestra. La versión de la anotación y el genoma de referencia fue GRCh37.75.

El segundo programa usado para la detección de fusiones génicas fue STAR-fusion versión 1.10.1 (Haas *et al.*, 2017). El programa usa el alineamiento STAR para la identificación de transcritos fusión soportados por las lecturas respectivas. La librería y el índice del genoma humano de referencia (GRCh37) se implementan junto con la descarga e instalación del programa. También se generó un script para la automatización de los comandos con cada muestra, los parámetros usados en todas las muestras fueron por defecto, se utilizó la misma anotación de referencia que Fuseq.

5.4. Análisis exploratorio (A.E) de los resultados de Fuseq y

STAR-Fusion

Todos los resultados se analizaron con el uso del lenguaje de programación R v3.6.2 (R Core Team, 2013) dentro del ambiente de desarrollo Rstudio (RStudio Team, 2021).

Las tablas resultantes de la detección de fusiones génicas se agruparon para obtener una tabla total de resultados, la cual se usó en los posteriores análisis. Se realizaron tablas

de frecuencia del número de fusiones génicas de todas las muestras para saber qué fusiones génicas predominan en el conjunto de muestras. También, se realizó una tabla de frecuencias de los genes que componen las fusiones génicas para determinar si hay genes que se presentan en más de una fusión génica, los cuales fueron indicados como genes compañeros promiscuos. Para la generación de las tablas se usaron los paquetes tidyverse 1.3.1 (Wickham *et al.*, 2019) y dplyr 1.0.7 (Wickham *et al.*, 2021). Con respecto a los gráficos principales, se realizaron gráficos de barras con respecto a la frecuencia de cada fusión génica de todas las muestras y la clasificación de las fusiones génicas en eventos intracromosómicos e intercromosómicos. Además, se realizó un gráfico de densidad de la frecuencia del número de compañeros de cada gen que compone una fusión génica. Para la generación de los gráficos se usó el paquete ggplot2 3.3.5 (Wickham, 2016).

Estos análisis nos permitieron tener una vista general de los resultados y la cobertura de las fusiones génicas en las muestras, además de un panorama previo para la comparación entre los resultados de ambos programas.

5.5. Comparación de los resultados en Fuseq y STAR-Fusion

La determinación objetiva de la similitud de los resultados es importante para la toma de decisiones en los procedimientos posteriores. Por lo tanto, se realizó un diagrama de Venn con las fusiones génicas que determina Fuseq y STAR-fusion para observar cuántas fusiones génicas se obtuvieron por ambos programas. Luego de la visualización general de genes compartidos, se generó un conteo de los resultados coincidentes con ambos programas de cada gen por muestra. Este proceso funcionó como filtro para obtener el menor número de falsos positivos, priorizando la especificidad. Además, se realizó una búsqueda en las bases de datos FusionGDB (Kim & Zhou, 2019) y COSMIC-Fusion (Tate *et al.*, 2019) de los resultados congruentes en ambos programas.

5.6. Búsqueda de genes proteína quinasa.

La búsqueda de genes quinasa se hizo a partir de los genes filtrados. Se seleccionaron los resultados que estén compuestos por al menos un gen quinasa. Se generó tablas de los genes que integraban las fusiones génicas de todas las muestras y se procedió a filtrarlos a partir de sus respectivos nombres. La selección se basó en la tabla de genes de proteína quinasas reportados en el genoma humano usando secuencias genómicas, DNA complementario y etiquetas de secuencias expresadas (ETS) (Manning *et al.*, 2002).

5.7. Validación *in silico* de las fusiones génicas quinasa

Para una mayor robustez de los resultados de las fusiones génicas quinasa se realizó la reevaluación del alineamiento de las lecturas de las fusiones génicas quinasa mediante FusionInspector version 2.6.0 (Haas *et al.*, 2021), Además, se usó la opción de ensamble *de novo* mediante Trinity (Grabherr *et al.*, 2011), la cual permite obtener transcritos de los genes, los cuales fueron soportados por las lecturas, sin tomar en cuenta el genoma de referencia. Por otro lado, se realizó el alineamiento de las lecturas de soporte en el transcrito mediante el paquete chimeraviz version 1.12.3 (Lågstad *et al.*, 2017) de Biomanager, sirviendo como una segunda validación *in silico* de las fusiones génicas quinasa. Los genes quinasa validados se usaron en los análisis posteriores.

5.8. Predicción de dominios funcionales de las fusiones génicas quinasa

La información sobre las fusiones génicas quinasa con secuencias codificantes se obtuvo mediante FusionInspector version 2.6.0 (Haas *et al.*, 2021). Los dominios contenidos en la secuencia fueron predichos mediante la base de datos Pfam (Mistry *et al.*, 2021). Se buscó entre los genes quinasa el dominio quinasa completo, porque proporciona una mayor seguridad con respecto a la funcionalidad de la fusión génica

quinasa. Además, se hizo una segunda predicción de los dominios mediante el sitio web (<https://prosite.expasy.org/>) de ScanProsite (De Castro *et al.*, 2006).

5.9. Disponibilidad de códigos

Los códigos utilizados en el desarrollo de esta investigación se encuentran disponibles en el repositorio: https://github.com/ana-45699/Tesis_licenciatura_UNSM

Resultados

6.1. Análisis exploratorio de las características de los archivos

fastq

El número de lecturas de todas las muestras presenta una densidad similar a la distribución normal (Figura 4), siendo el máximo y mínimo valor de 279.9 y 2.5 millones de lecturas, respectivamente. La media del tamaño de las lecturas en todas las muestras fue 75 pb y su porcentaje medio de duplicados fue 27.21%. Los datos de calidad por muestra se encuentran en el Anexo 2.

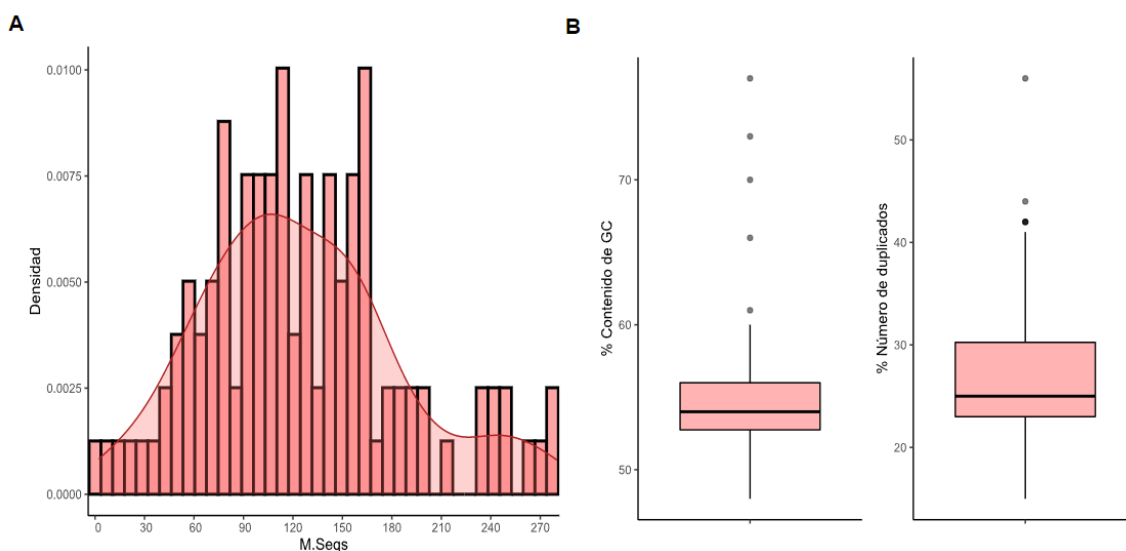


Figura 4: Análisis de calidad. A) Distribución del número de secuencias en cada muestra. B) Diagramas de cajas del contenido de GC y duplicaciones en las muestras.

6.2. Análisis exploratorio de los resultados de los programas de

Fuseq y STAR-Fusion

En el análisis con Fuseq, se detectaron 672 fusiones génicas distintas. La cobertura en el total de muestras fue 92.85% (104 muestras). La fusión génica con mayor

frecuencia fue *TPRX1--GPNMB*, se registró en 53 muestras, indicando su presencia en 47% de las muestras que presentaban alguna fusión génica (Figura 5).

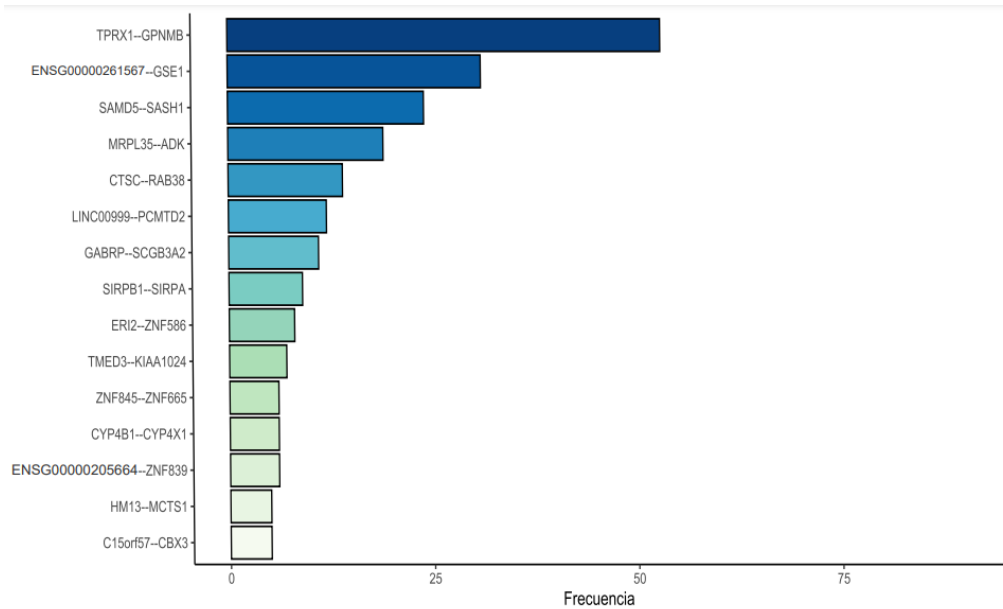


Figura 5: Las 15 fusiones génicas más frecuentes detectadas por FusedSeq.

Los genes que conforman las fusiones génicas pueden estar presentes en más de una fusión génica. Por ello se procedió a hacer el conteo de genes en forma individual. Luego se realizó un conteo de genes compañeros de cada gen, el cual debe estar presente en al menos una fusión génica. La mayor parte de genes que conforman las fusiones génicas presentan un solo compañero (Figura 6). El gen homólogo 2 de discos grandes (*DLG2*) presentó 9 genes compañeros diferentes, siendo el mayor número de genes compañeros diferentes (Anexo 4).

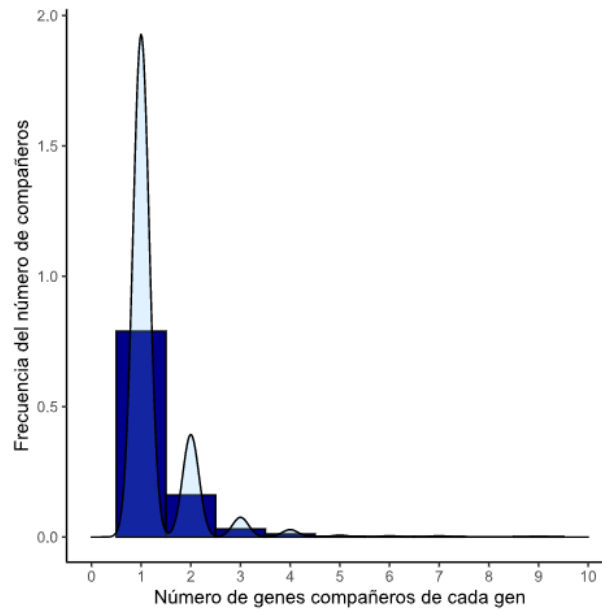


Figura 6: Distribución del número de genes compañeros de Fuseq.

En el análisis con STAR-Fusion, se detectaron 693 fusiones génicas distintas. La cobertura en las muestras total fue 91.07% (102 muestras). La fusión génica de un transcrito de la proteína de retinitis pigmentosa 11 y Navegador neuronal 3 (*RP1-34H18.1-NAV3*) se presentó con una frecuencia de 83 muestras (Anexo 5), indicando una cobertura del 73% de las muestras con presencia de fusiones génicas (Figura 7).

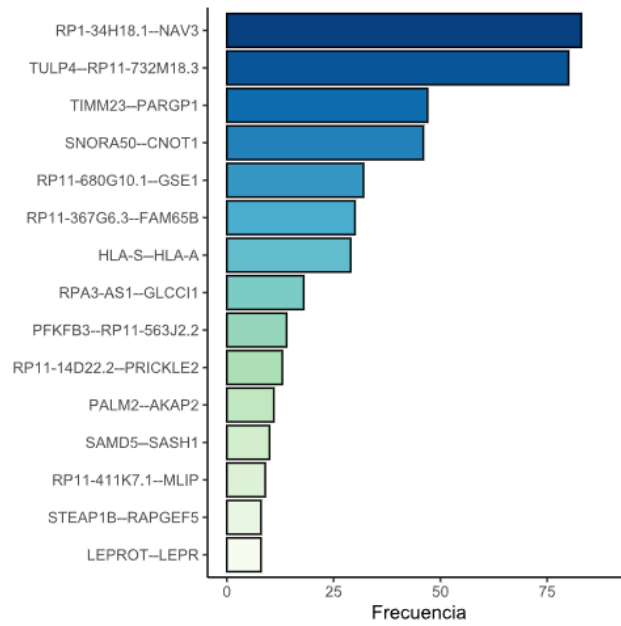


Figura 7: Las 15 fusiones génicas más frecuentes detectadas por STAR-Fusion.

La mayor parte de genes que conforman las fusiones génicas presentan un solo compañero (Figura 8). El gen que presentó un mayor número de genes compañeros fue *ADYC2* (Anexo 6).

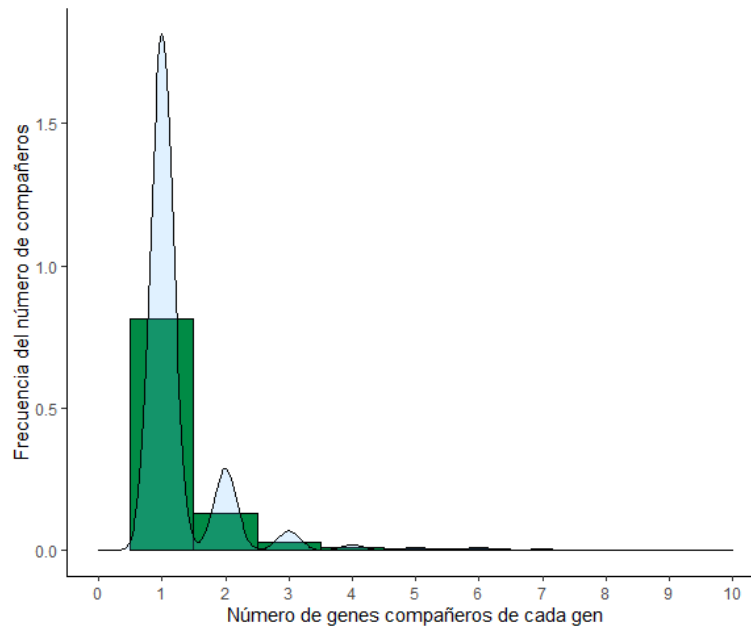


Figura 8: Distribución del número de genes compañeros de STAR-Fusion.

Con respecto a los eventos intracromosómicos e intercromosómicos de ambos programas, se encontró que los eventos intracromosómicos son los más frecuentes. No obstante, solo hubo diferencias significativas (Prueba t de Student, $p\text{-value}=1.1\text{e-}12$) entre la media de los eventos intracromosómicos e intercromosómicos por muestra en los resultados de STAR-Fusión (Figura 9).

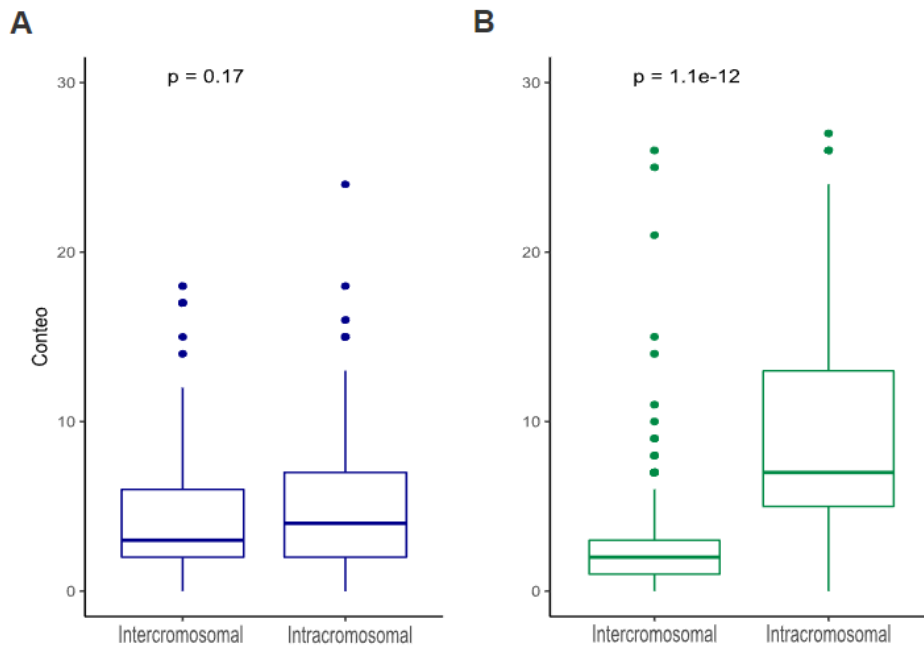


Figura 9: Diagrama de cajas de las fusiones génicas clasificadas por eventos intracromosómicos e intercromosómicos de Fuseq (A) y STAR-Fusion (B).

La distribución de las fusiones génicas en los cromosomas, sin contar las isoformas, presentó una densidad similar, teniendo una muy baja frecuencia el cromosoma X, en los resultados de Fuseq solo se presentó una fusión génica dentro del cromosoma Y (Figura 10A), mientras en STAR-Fusion no hubo fusiones génicas en el cromosoma Y y presentó una fusión génica en el genoma mitocondrial. No obstante, se observaron disimilitudes de gran magnitud como la cantidad de fusiones génicas en los cromosomas

6 y 19 (Figura 10 A y B). En el cromosoma 19 se encuentra uno de los genes (*TPRX1*) que conforma la fusión génica con mayor frecuencia los resultados de Fuseq (*TPRX1—GPNMB*).

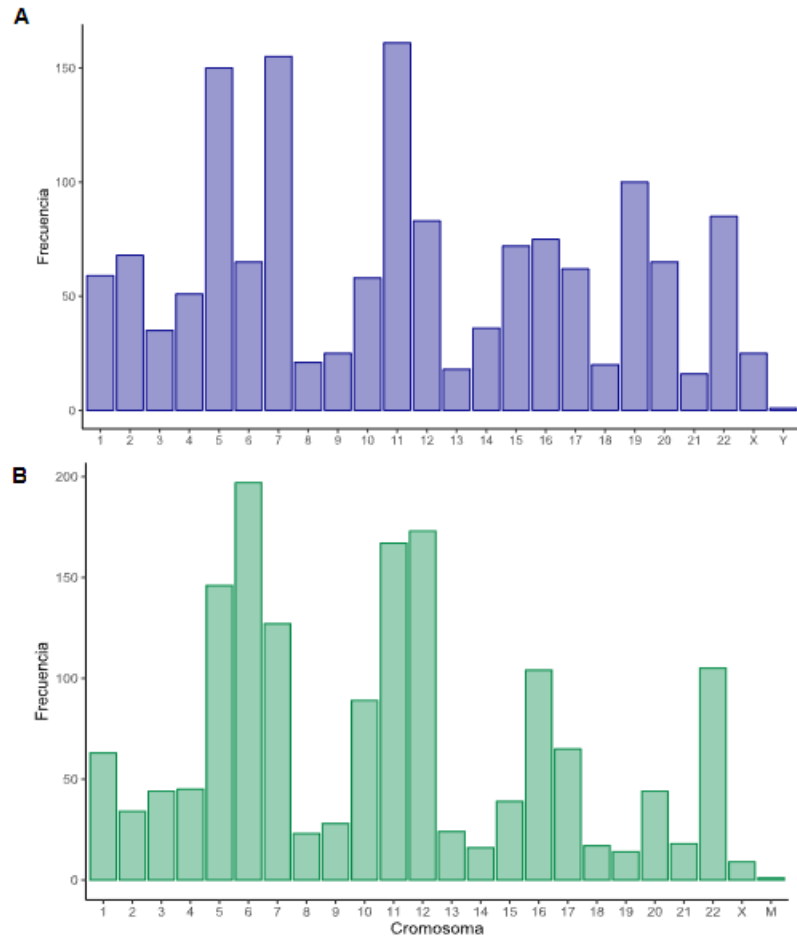


Figura 10: Conteo de las fusiones génicas por cromosoma de Fuseq (A) y STAR-Fusion (B).

Los resultados de Fuseq y STAR-Fusion cubrieron más del 90% de las muestras (Fuseq=104 y STAR-Fusion=102), indicando que la mayor parte de las muestras presentaron al menos una fusión génica. Por otro lado, debido a las diferencias en los resultados por cada programa usado, se procedió a generar dos tablas de resultados basadas en los genes compartidos de ambos programas. Se encontraron en total 341

genes compartidos entre Fuseq y STAR-Fusion (Figura 11), entre los que se presentaron 337 fusiones génicas compartidas y concordantes en cada muestra, representando aproximadamente el 50% de los resultados de ambos programas. Las fusiones génicas concordantes se hallaron en 80 muestras.

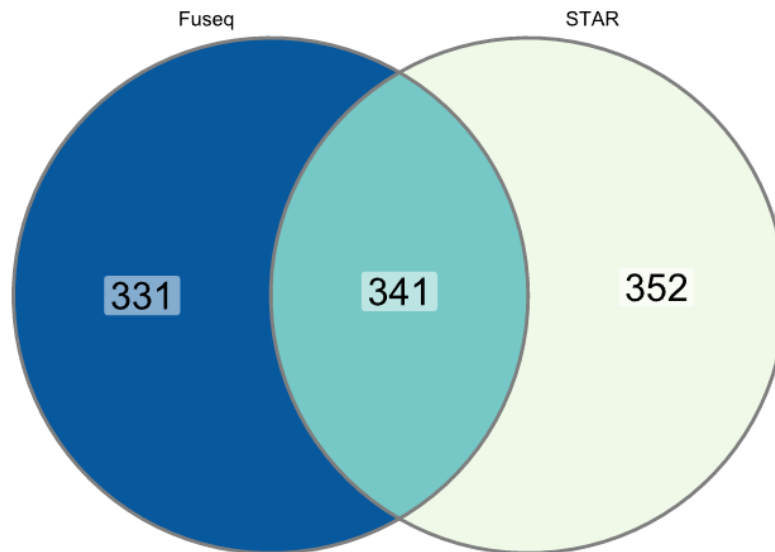


Figura 11: Diagrama de Venn de los genes de STAR-Fusion y Fuseq

La fusión génica con mayor cobertura entre las muestras fue un transcrito de la proteína de retinitis pigmentosa 11 y Elemento supresor genético 1 (*RP11-680G10.1--GSE1*) con presencia en 17 muestras. La segunda fusión génica más frecuente fue *SAMD5-SASH1*, la cual se encontró en 9 muestras. Mientras que las otras fusiones génicas dentro de las 15 fusiones génicas más frecuentes estaban entre 4 y 2 muestras (Figura 12).

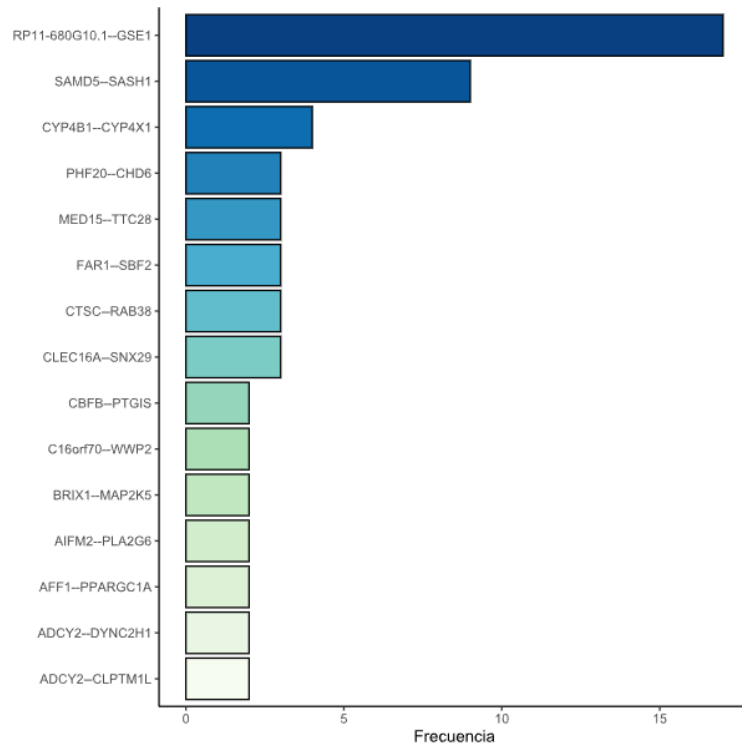


Figura 12: Las 15 fusiones génicas con mayor frecuencia

Con respecto al tipo de evento cromosómico, se observó un mayor número de eventos intracromosómicos en comparación con los intercromosómicos con una diferencia significativa mediante la prueba t de Student (p-value: 0.004).

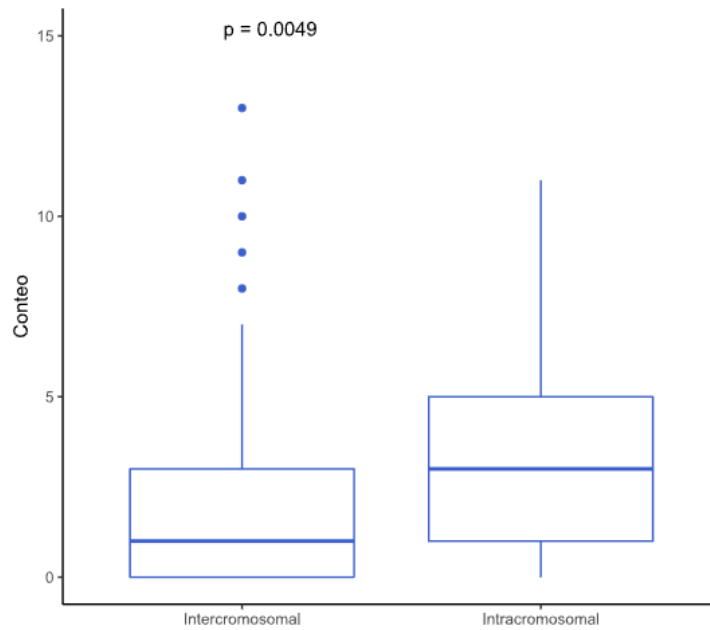


Figura 13: Distribución de las fusiones génicas por tipo de evento intracromosómico e intercromosómico.

La mayor parte de genes que integra las fusiones génicas pertenece al cromosoma 11, además la frecuencia de genes pertenecientes al cromosoma 6 en los resultados de STAR-Fusion ha disminuido casi hasta el 10% (Figura 14).

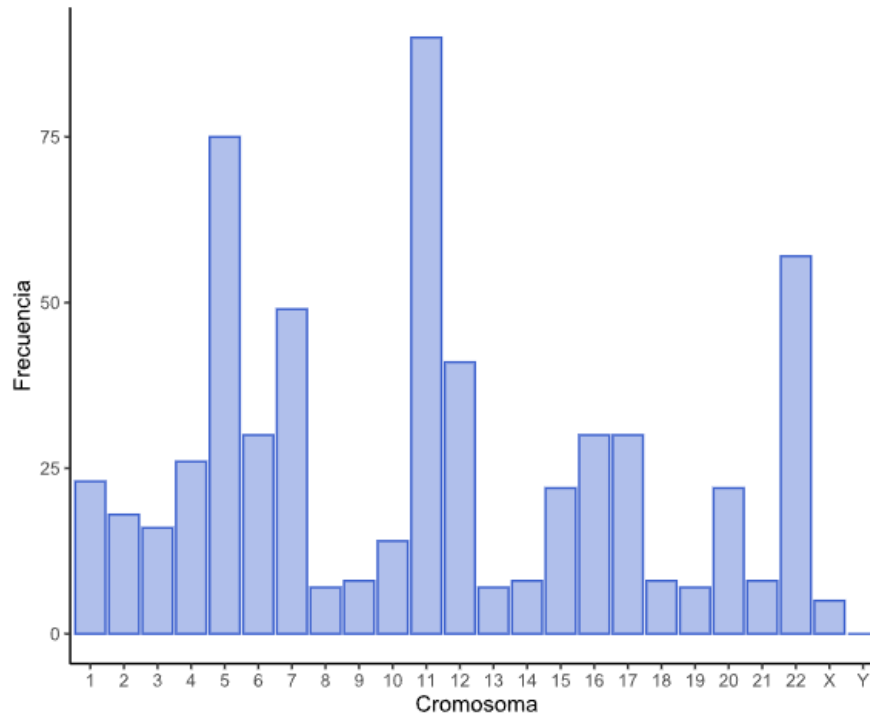


Figura 14: Distribución de las fusiones génicas en los cromosomas

Los genes comunes en ambos resultados se buscaron en bases de datos específicas para fusiones génicas. Estas bases de datos presentan información sobre los genes que conforman la fusión génica, el punto de corte y, en algunos casos, funcionalidad. Por lo tanto, las bases de datos de fusiones génicas podrían proporcionar información sobre la funcionalidad de las fusiones génicas detectadas. Se encontraron 16 genes que componían a las fusiones génicas en las bases de datos FusionGDB y COSMIC-Fusion (Figura 15). No obstante, las fusiones génicas que incluían estos genes no eran las mismas, es decir, los genes compañeros eran distintos a los presentados en la base de datos.

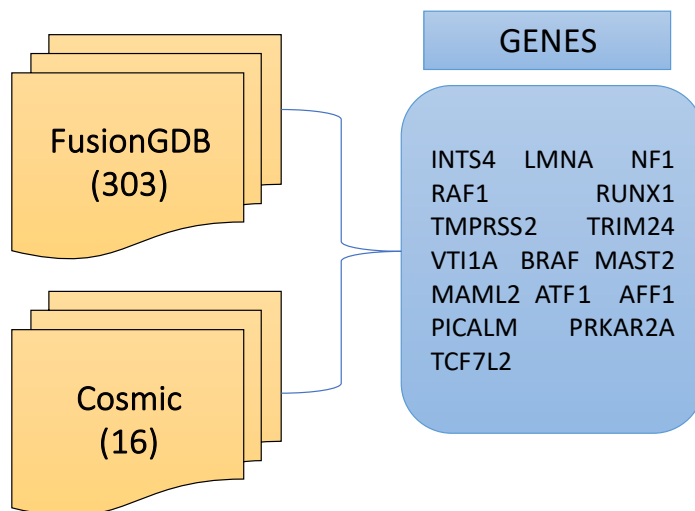


Figura 15: Genes que conforman fusiones génicas en los resultados compartidos de Fuseq y STAR-Fusion dentro de las bases de datos de FusionGDB y COSMIC-Fusion.

Asimismo, se evaluó la promiscuidad de los genes compañeros en los resultados compartidos, siendo el gen compañero más promiscuo el homólogo 2 de discos grandes (*DLG2*), el cual conforma 7 fusiones génicas diferentes (Tabla 1). La mayor parte de los genes solo presentaban un gen compañero, indicando que la mayor parte de las fusiones génicas tienen genes únicos.

Gen	Proteína	Gen_ID	Número de genes compañeros
<i>DLG2</i>	Homólogo 2 de discos grandes	ENSG00000150672	7
<i>ADCY2</i>	Adenilato ciclasa 2	ENSG00000078295	5
<i>EP300</i>	Histona acetiltransferasa p300	ENSG00000100393	4
<i>GAB2</i>	Proteína 2 de unión asociada a GRB2	ENSG00000033327	4
<i>GOLPH3</i>	Fosfoproteína 3 de Golgi	ENSG00000113384	4
<i>LRBA</i>	Proteína de anclaje tipo beige sensible a LPS	ENSG00000198589	4
<i>TENM4</i>	Teneurina transmembrana 4	ENSG00000149256	4
<i>UVRAG</i>	Proteína génica asociada a la resistencia a la radiación UV	ENSG00000198382	4

Tabla 1: Genes promiscuos que forman fusiones génicas con más de 3 genes
compañeros

6.3. Detección de proteína quinasas del genoma humano

Se realizó un filtro de fusiones génicas que presentaran un gen proteína quinasa en cualquier posición de la fusión génica (3' o 5'). Se encontraron 21 fusiones génicas quinasa con 17 genes proteína quinasa, entre las cuales se obtuvieron genes que codifican proteína quinasas pertenecientes a la cascada de señalización MAPK (*ROCK2*, *BRAF*, *MAP3K3*, *MAP2K5*, *RAF1*). Asimismo, todos los genes proteína quinasa que componen las fusiones génicas detectadas en el análisis tienen un gen compañero único, excepto *NLK*, el cual conforma 3 fusiones génicas distintas (*NLK--ACACA*, *NLK--BCAS3*, *NLK--HLF*) como se detalla en la Tabla 2. Las fusiones génicas proteína quinasa presentaron una cobertura total de 19 muestras tumorales, indicando que hay muestras que presentan más de una fusión génica proteína quinasa (Tablas 2 y 3). Por otro lado, la mayor frecuencia individual de una fusión génica proteína quinasa es de 2 muestras, observándose que solo el 19% de las fusiones génicas quinasa estaban presentes en dos muestras distintas (Figura 16).

N°	Muestra	Fusión génica
1	2260_PR40969d	BRIX1--MAP2K5
2	2260_PR40970a	AP1S2--CDKL5, CDKL5--AP1S2
3	2260_PR40971a	DLG2--PAK1
4	2260_PR40972a	BRIX1--MAP2K5
5	2260_PR40983d	BRD4--MAST3
6	2260_PR40985a	AKT3--DKK2
7	2260_PR40986a	KIF5B--RAF1
8	2260_PR40986d	KIF5B--RAF1, ROCK2--CACNA2D3
9	2260_PR40989a	SMARCA5--NEK1, PASK--LMBR1, ATP13A3--MARK2
10	2260_PR41020d	MAP3K3--RAD51C
11	2260_PR41025a	EIF2B3--MAST2
12	2260_PR41039a	HIPK3--INSC
13	2260_PR41039d	HIPK3--INSC
14	2260_PR41901a	NLK--ACACA, NLK--BCAS3
15	2260_PR41907a	GOLGA4--BRAF
16	2260_PR41910a	ROCK2--PICALM
17	2260_PR41911a	DYRK2--ELK3, NLK--HLF
18	2260_PR41920a	TENM4--ATM
19	2260_PR41920c	TENM4--ATM

Tabla 2: Fusiones génicas proteína quinasa halladas en los resultados compartidos de STAR-Fusion y Fuseq con respecto a las muestras en las que se presentan. El gen proteína quinasa de cada fusión génica está indicado en negrita.

Fusión quinasa	Gen 5'	Gen 3'	Frecuencia
<i>BRIX1--MAP2K5</i>	ENSG00000113460	ENSG00000137764	2
<i>KIF5B--RAF1</i>	ENSG00000170759	ENSG00000132155	2
<i>HIPK3--INSC</i>	ENSG00000110422	ENSG00000188487	2
<i>TENM4--ATM</i>	ENSG00000149256	ENSG00000149311	2

Tabla 3: Fusiones génicas proteína quinasa presentes en más de una muestra. El gen proteína quinasa está indicado en negritas en la primera columna, el código de los genes 5' y 3' están indicados en segunda y tercera columna, respectivamente.

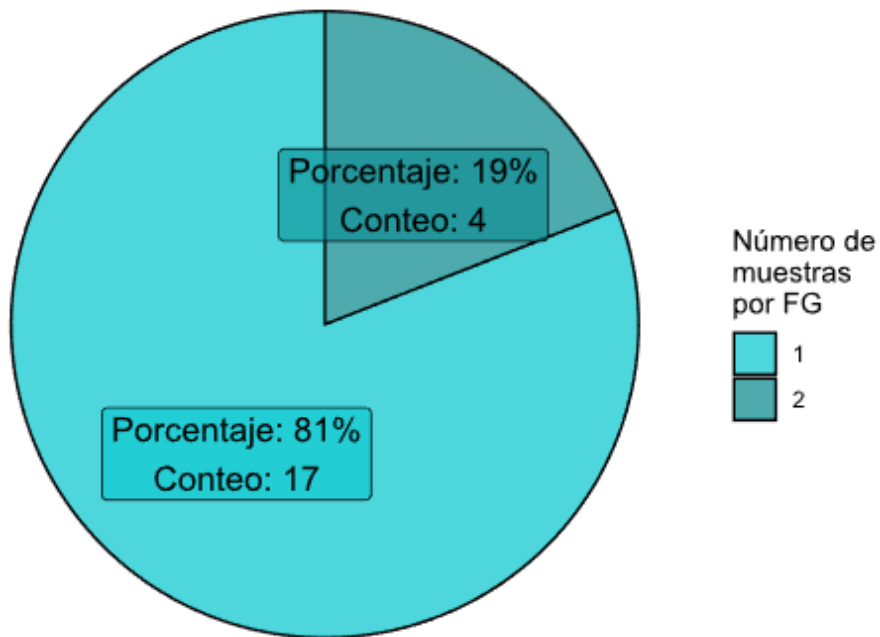
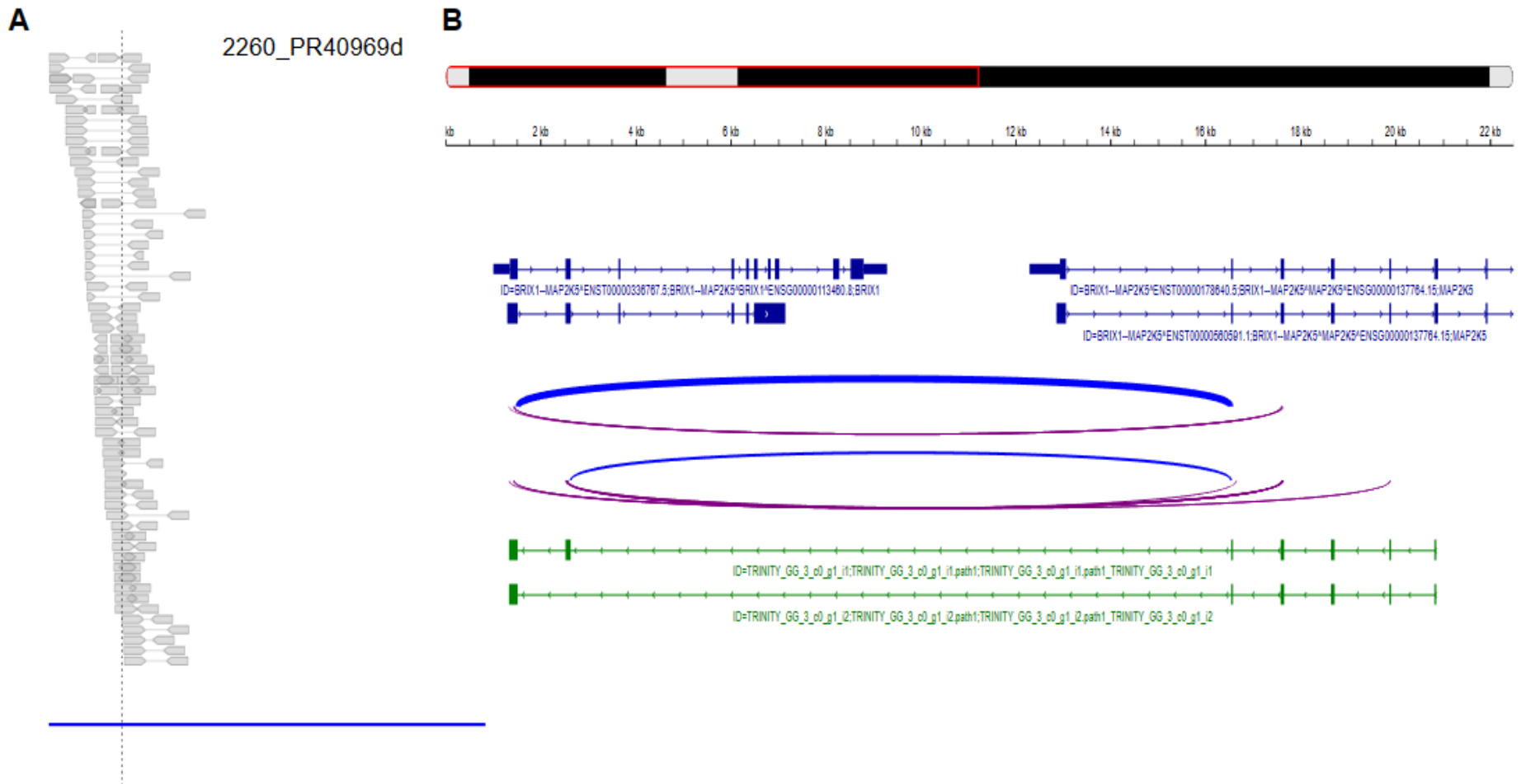


Figura 16: Gráfico de torta del porcentaje de fusiones proteína quinasa por el número de muestras de cada fusión génica (FG) que se presentan por identificación génica.

6.4. Validación *in silico* de las fusiones génicas quinasa

El 66.67% (14/21) de las fusiones génicas quinasa se detectaron en la validación (Anexo 8). Las lecturas de soporte de los genes quinasa se distribuyeron entre los transcritos de los genes que conforman la fusión génica (Figura 17A). Las lecturas de unión se observan aglomeradas en el punto de corte. Se formaron distintos transcritos generados por el *ensamblaje de novo* indicando distintas isoformas de la fusión génica (Figura 17B).

Figura 17: Validación de la fusión génica BRIX1—MAP2K5. Alineamiento de las lecturas de soporte de la fusión génica BRIX1—MAP2K5 (A). Reporte de FusionInspector de la fusión génica BRIX1—MAP2K5 (B).



La clasificación con respecto a las secuencias codificantes mostró que 52.40 % de las traducciones de las fusiones génicas se encontraban dentro del marco de lectura, mientras 47.60% estaba fuera del marco (Figura 18A). Las fusiones génicas dentro del marco de lectura presentaron distintos valores de fragmentos fusión del total de fragmentos por millón (FFPM) en sus respectivas muestras, siendo el valor de FFPM más alto de la fusión génica *NLK--ACACA* y el más bajo es de *EIF2B3--MAST2* (Figura 18B),

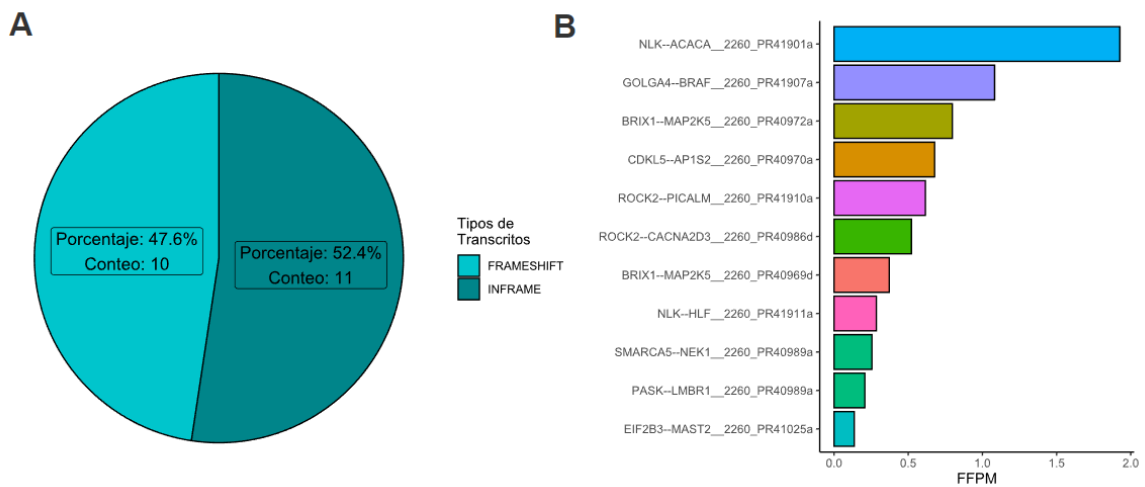


Figura 18: Proteína quinasas dentro del marco de lectura y su expresión. Gráfico de torta del porcentaje de quinasas por el tipo de secuencia codificante, las que se encuentran dentro del marco de lectura están en la categoría INFRAME, mientras las que se encuentran fuera del marco de lectura son FRAMESHIFT(A). Gráfico de barras de los valores de expresión de los transcritos fusión, donde FFPM significa fragmentos fusión del total de fragmentos por millón (B).

6.5. Predicción de los dominios funcionales de las fusiones génicas quinasa

Debido a la mayor importancia funcional de las secuencias codificantes dentro del marco de lectura, se extrajeron las fusiones génicas quinasa que estaban dentro del marco de lectura (*CDKL5--AP1S2*, *NLK--ACACA*, *NLK--HLF*, *PASK--LMBR1*, *ROCK2--CACNA2D3*, *ROCK2--PICALM*, *BRIX1--MAP2K5*, *EIF2B3--MAST2*, *GOLGA4--BRAF*, *SMARCA5--NEK1*) y se procedió a analizar si presentaban el dominio fusión quinasa (Anexo 10 y 11). Adicionalmente, para corroborar la traducción de los transcritos generados se realizó el alineamiento con la base de datos del proteoma humano en Ensembl, mostrando que los resultados de los programas con respecto a las secuencias codificantes eran coincidentes (Anexo 9).

Por otra parte, las predicciones de PfamScan mostraron dominios proteína quinasa de virus y hongos (Anexo 10) teniendo valores de p-value de $10e-5$. En cambio, las predicciones en Prosite no mostraron estos dominios (Anexo 11). La predicción de Prosite presentó ubicaciones del dominio proteína quinasa similares a PfamScan (Anexo 10), a excepción de algunos dominios incompletos, los cuales no pudieron ser predichos por Prosite (Anexo 11). La mayoría de las fusiones génicas quinasa presentaron otros dominios aparte del dominio proteína quinasa, tres de las fusiones quinasa (*BRIX—MAP2K5*, *NLK—ACACA*, *ROCK2—CACNA2D3*) presentaron dominios en la secuencia del gen que no era proteína quinasa (Figura 19). Las fusiones proteína quinasa con dominios completos fueron serina/treonina quinasa.

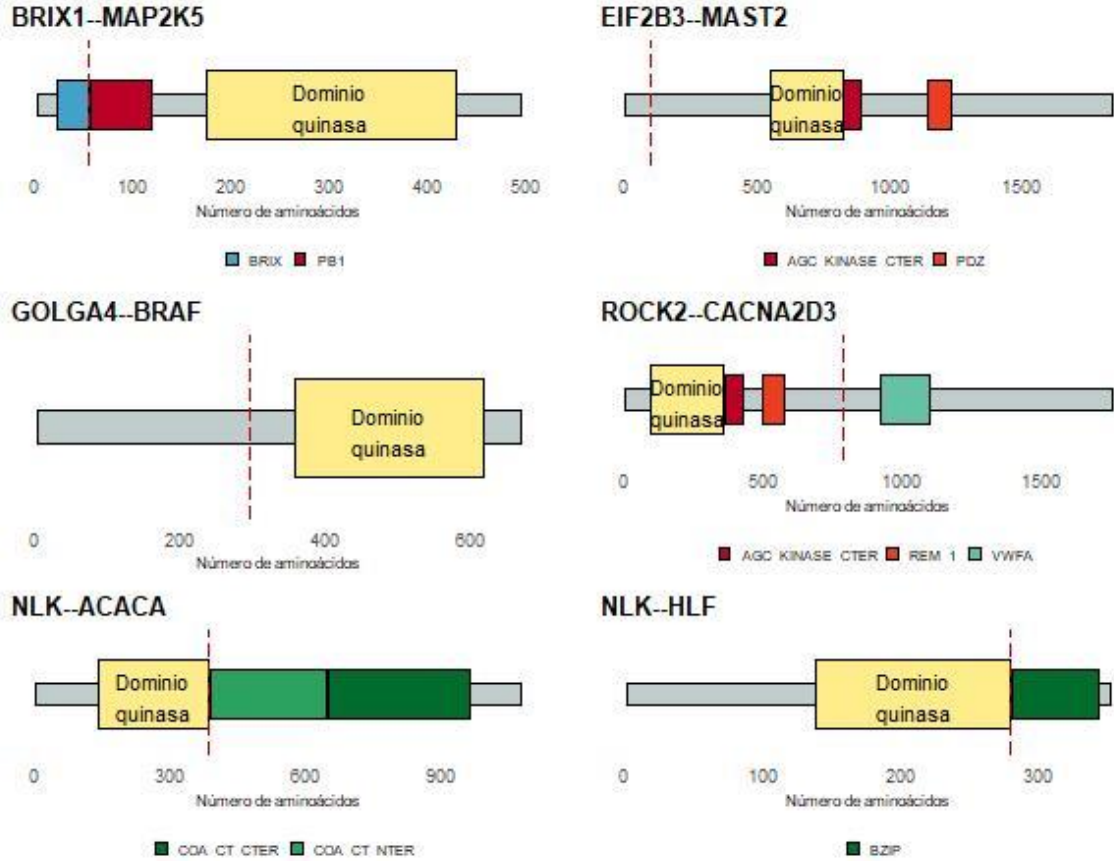


Figura 19: Secuencia de aminoácidos de las fusiones génicas quinasa. La secuencia de aminoácidos de proteína quinasa presenta el dominio quinasa catalítico completo (amarillo) y sus dominios complementarios (rojo y anaranjado). Los dominios del gen compañero 5' o 3' se representan con tonalidades de azul y verde, respectivamente. El punto de corte se representa como una línea roja punteada.

Discusión

Siendo ALM un cáncer poco estudiado, el presente trabajo presenta el estudio genómico de una de las cohortes más grandes de ALM hasta la actualidad. Anteriormente, también se realizó un estudio de 112 muestras de ALM (Kuchelmeister *et al.*, 2000). Sin embargo, solo se realizó un análisis de sus características histopatológicas, además las muestras eran de pacientes con únicamente ascendencia europea (Kuchelmeister *et al.*, 2000). El presente estudio se centra en datos de secuenciación de RNA de pacientes mexicanos tratados en el Instituto Nacional de Cancerología de México.

El análisis exploratorio de las fusiones génicas es importante para observar todo el panorama de opciones y sus características para la elección de las fusiones génicas que podrían ser aptas para el diagnóstico y terapia en ALM. El enfoque en el potencial de las fusiones génicas para el diagnóstico se basó principalmente en la frecuencia de la fusión génica dentro de las muestras y la presencia de genes proteína quinasa. Además, se analizó la promiscuidad, es decir presencia de distintos genes compañeros, de los genes que conformaban las fusiones génicas, indicando una posible tendencia de estos genes a formar fusiones independientemente de sus genes compañeros. Acerca de las fusiones génicas quinasa, se realizó un análisis adicional que consistía en la validación *in silico* y funcionalidad de estas.

En los resultados compartidos por STAR-Fusion y Fuseq se presentó una diferencia significativa entre los tipos de eventos cromosómicos, siendo las translocaciones intracromosómicos más frecuentes que las intercromosómicos. Esto podría ser explicado por los distintos territorios nucleares que ocupan los cromosomas, por lo que la posición no aleatoria de los genes nucleares y cromosomas conduciría a que la proximidad espacial entre dos secuencias promovería su translocación preferencial (Cremer & Cremer, 2010).

Anteriormente, estudios citogenéticos han mostrado mayor frecuencia de las translocaciones con una relativa cercanía. En modelos murinos expuestos a radiación ionizante (IR), la observación de la organización del genoma y las translocaciones del DNA nuclear mediante análisis de capturas conformacionales de la cromatina (Hi-C) muestra que las translocaciones más frecuentes son las intracromosómicas (Zhang *et al.*, 2012). Además, pacientes con cáncer de mama primario han presentado una mayor frecuencia de translocaciones intracromosómicas (Parl *et al.*, 2019). Por último, en melanoma se ha reportado 37 rearrreglos somáticos, donde 3 fueron intercromosómicos y 34 fueron intracromosómicos, incluyendo deleciones inversiones y duplicaciones (Pleasance *et al.*, 2010).

La fusión génica con una mayor frecuencia es *RP11-680G10.1--GSE1*. Esta fusión ha sido predicha hasta en un 50% en cáncer endometrial por la metodología de STAR-Fusion en 28 pacientes (Yao *et al.*, 2019). *RP11-680G10.1* es un lncRNA ubicado en el cromosoma 16. Estudios previos han presentado que *GSE1* puede mediar en la progresión del cáncer in vitro (Chai *et al.*, 2016). Por otro lado, *SAMD5-SASH1* (segunda fusión génica más frecuente, Anexo 7) se ha presentado en distintas muestras de la base de datos del Atlas del genoma del cáncer (TCGA) tanto de muestras no cancerosas y diferentes tipos de cáncer, incluido melanoma cutáneo (CM).

Además de la frecuencia, también se analizó la promiscuidad de los genes que conformaban las fusiones génicas. Los genes *DLG2* y *ADCY2* comprendían un mayor número de genes compañeros, se presentaron en 74 y 22 fusiones génicas, respectivamente dentro de la base de datos FusionGDB. Entre estas fusiones génicas se encuentra anotado *ADYC2-CPTML1*, presente en los resultados, el cual está curado y ha sido identificado en una muestra de CM en la base de datos del TCGA (TCGA-FS-A4F4-

06A). El gen homólogo 2 de discos largos (*DLG2*) presenta funciones como supresor de tumor, por lo que recombinaciones cromosómicas podrían alterar o eliminar estas funciones (Keane *et al.*, 2020). Adicionalmente, translocaciones de este gen se han encontrado frecuentemente en neuroblastoma (Lopez *et al.*, 2019) y su baja expresión ha sido asociada con cáncer de ovario y osteosarcoma (Keane *et al.*, 2020).

Con respecto a las fusiones génicas quinasa, el 52.4% de las proteína quinasa encontradas se presentó dentro del marco de lectura, de las cuales se procedió a analizar las secuencias traducidas. La predicción de dominios mostró resultados por parte de PfamScan de especies no pertenecientes al humano, siendo los dominios proteína quinasa predichos también de otros organismos como hongos, mostrando la conservación de estos dominios en diferentes especies (Zhao *et al.*, 2014).

Se identificó fusiones génicas quinasa que presentan genes de la cascada de MAPK (*PAK1*, *MAP2K5*, *NLK* y *BRAF*), la cual es una vía generalmente alterada en el melanoma (Inamdar *et al.*, 2010). Estas fusiones génicas quinasa presentaron la parte activa de dominios quinasa (Figura 10), a excepción de *DLG2-PAK1*. Los compañeros de estos genes presentan dominios con diferentes funciones. La fusión *BRIX1-MAP2K5* presenta un dominio Brix, que es un dominio conservado en varias proteínas involucradas en la biogénesis de los ribosomas y se asoció con el crecimiento y desarrollo de organismos (Maekawa *et al.*, 2018). *BRIX1* está respaldado como un gen impulsor del cáncer a partir de experimentos de mutagénesis por inserción en ratones (Abbott *et al.*, 2015). Por lo tanto, se podría suponer que el dominio Brix asociado al dominio quinasa de *MAP2K5* generaría alteraciones funcionales involucradas en el desarrollo de ALM.

Otra fusión génica quinasa encontrada es *GOLGA4-BRAF*. Solo se identificó el dominio quinasa perteneciente a la secuencia de *BRAF* (Anexo 10 y 11). El gen *BRAF* es un gen altamente estudiado en CM. Su intenso estudio es debido a la alta frecuencia de mutaciones

este gen en pacientes CM (Ascierto *et al.*, 2012; Botton *et al.*, 2019; Solit & Rosen, 2011). Además, se ha sugerido la clasificación de subtipos de CM con respecto al patrón de mutaciones de este gen en combinación con otros, como NRAS (Hayward *et al.*, 2017; Puig-Butillé *et al.*, 2013; Saldanha *et al.*, 2006). Su gen asociado, *GOLGA4*, ha sido reportado como socio con diferentes quinasas. Se ha reportado en fusiones génicas como compañero de quinasas de la familia JAK en leucemia linfoblástica aguda de células B (Downes *et al.*, 2022) y en melanoma (Hayward *et al.*, 2017). Otro estudio reportó la fusión de *GOLGA4-RAF1* en melanoma metastásico (McEvoy *et al.*, 2019).

Por otro lado, se obtuvieron fusiones con NLK (NLK-ACACA, NLK-HLF), el cual es una serina/treonina quinasa que ha presentado importancia en otros tipos de cáncer. El estudio de Yasuda *et al.* (Yasuda *et al.*, 2003) propone que actúa como un supresor de tumores induciendo a la apoptosis en cáncer colorrectal; mientras en hepatocarcinoma la inhibición de la expresión de *NLK* detiene la proliferación de células cancerosas (Jung *et al.*, 2010).

Además de las quinasas de la vía principal relacionada a melanoma como las MAPK, se encontraron otros genes proteína quinasa relacionados a cáncer en las fusiones génicas quinasa. La fusión *ROCK2-CACNA2D3* tenía un dominio quinasa y un dominio VWA. El gen *ROCK* (serina/treonina quinasa) tiene un papel en la citocinesis mitótica, organización de la estructura de la actomiosina. Se ha reportado su mayor expresión en otros tipos de cáncer como el gástrico (Li *et al.*, 2017). Además, *ROCK2* se ha relacionado con el proceso de metástasis en el carcinoma hepatocelular (Huang *et al.*, 2014). El gen asociado (*CACNA2D3*) participa en el transporte transmembrana de iones de calcio y el dominio VWA está involucrado en el reconocimiento de una amplia gama de proteínas.

La fusión *EIF2B3-MAST2* presenta a *MAST2* (serina/treonina quinasa) con su dominio quinasa completo y un dominio nucleotidil transferasa. Los genes que integran esta fusión se encontraron en FusionGDB con diferentes socios, mientras en la base de datos de Fusion COSMIC solo se encontró a *MAST2*. El gen *MAST2* ha sido asociado principalmente

al desarrollo de cáncer de mamá. El estudio de Robinson et al (Robinson *et al.*, 2011) reportó una alta frecuencia de translocaciones de los genes de la familia MAST.

La incongruencia entre los resultados de ambos programas con respecto a la predicción de las fusiones génicas se atribuye principalmente a los transcritos referencia y filtros en cada programa. A pesar de que los transcritos son con respecto al genoma GRCh37 y la anotación de los genes con respecto a su ID es coincidente al ser de Ensembl, las secuencias referencia de los transcritos de algunos genes no son únicas, como la del *TPRX1*, generando un alineamiento múltiple que podría atribuirse a un único gen. Por ejemplo, se observó un gran número de eventos inter cromosómicos en el cromosoma 19 en los resultados de Fuseq a diferencia de STAR-Fusion (Figura 10), esto debido a que el gen con mayor frecuencia de sus resultados presentaba *TPRX1* (cromosoma 19), no obstante, al realizarse un BLAST a las lecturas de soporte se encontró que las lecturas de la fusión génica (*TPRX1-- GPNMB*) presentaban hits hasta en 100 partes distintas del genoma (Anexo 12).

Por otro lado, la fusión génica con mayor frecuencia de STAR-Fusion (*RP1-34H18.1-- NAV3*), sus genes asociados presentes en el cromosoma 20 y 12, se alinearon en posiciones diferentes a los resultados correspondientes (Anexo 13). Con respecto a los filtros, Fuseq presenta un filtro en donde la distancia es la distancia entre los breakpoints de los genes que conforman la fusión génica tienen que ser mayor a 100000 pb (distancia (*geneA*, *geneB*) > 100000) en eventos intracromosómicos. Al colocar este filtro en los resultados de STAR-Fusion, se descartaron 154 eventos intracromosómicos. Por lo tanto, el uso de los resultados congruentes de ambos programas permite enfocarse en los resultados predictivos con mayor soporte al estar predichos por distintas formas de alineamiento.

Con respecto a los objetivos planteados, estos se pudieron realizar detectando fusiones génicas dentro de las 112 muestras e identificando fusiones génicas quinasa con dominios activos. Las fusiones génicas candidatas a marcadores de diagnóstico potenciales debido a su frecuencia presentan genes que han sido estudiados anteriormente. Por ejemplo, uno de los compañeros que presenta la fusión génica más frecuente (*GSE1*) ha estado presente en estudios involucrados de cáncer debido a su funcionamiento en distintos cánceres (Bamodu *et al.*, 2021; Chai *et al.*, 2016), colocándose como un objetivo terapéutico atractivo. En este sentido, se realizaron pruebas de inhibición con Kaempferol, sin embargo, no mostró una inhibición significativa (Song *et al.*, 2015).

Los genes con un mayor número de compañeros dentro de las fusiones génicas se han detectado conformando fusiones génicas en otros tipos de cánceres, además algunos presentan funciones relacionadas al desarrollo tumoral como el gen *DLG2*, por lo que su estudio podría involucrar un mayor conocimiento de su funcionalidad en ALM. Por lo tanto, estos genes con alta promiscuidad y frecuencia presentan potencial para ser futuros marcadores de diagnóstico y objetivos terapéuticos en ALM.

Finalmente, las fusiones génicas quinasa encontradas presentan similitudes con anteriores estudios como el de Hayward *et al.*, 2017 en donde encontraron fusiones quinasa PAK-1 no funcionales en ALM. En nuestros resultados se encontró *DLG2—PAK-1*, la cual estaba fuera del marco de lectura indicando pocas probabilidades de ser funcional. También, se encontró la fusión *GOLGA4—BRAF* siendo muy similar a la fusión quinasa encontrada en el trabajo de Hayward *RAF1—GOLGA4* de melanoma cutáneo sin causas por UV, estando ambas dentro del marco de lectura, indicando una alta probabilidad de ser funcionales.

Estos resultados muestran que las fusiones proteína quinasa se presentan una baja frecuencia de manera individual, señalando una reducción en su potencial de ser un marcador en ALM. No obstante, existe la posibilidad de que la presencia de fusiones

génicas quinasa se presente en un subgrupo de ALM. La categorización en pacientes por datos genómicos en ALM aún no está completamente consolidada; sin embargo, estudios han diferenciado subtipos de ALM en base a análisis ómicos. El estudio de Puig-Butillé et al. (Puig-Butillé *et al.*, 2013) sugiere la existencia de subgrupos de ALM basado en la distribución de alteraciones de oncogenes dentro de las muestras. Además, el estudio de Dinesh et al. (Pradhan *et al.*, 2019) encontró patrones diferenciados de metilación entre ALM primario y ALM metastásico. Referente a las fusiones génicas quinasa, el estudio de Turner et al. sugiere que estas fusiones podrían ser parte del análisis clínico al encontrarse en un subgrupo de pacientes ALM sin mutaciones en los drivers comúnmente usados (Turner *et al.*, 2017).

En resumen, el análisis exploratorio de las fusiones génicas en ALM permitirá el desarrollo de estudios más específicos sobre las fusiones con mayor frecuencia, promiscuidad y contenido de genes quinasa detectadas en las muestras. Por ejemplo, los resultados presentaron fusiones génicas con ROCK2 y BRAF los cuales presentan inhibidores como Y27632 (Rath & Olson, 2012) y vemurafenib (Botton *et al.*, 2019). Este resultado permitirá el desarrollo de estudios enfocados en estas fusiones génicas quinasa como objetivo terapéutico, lo cual podría tener relevancia en la elección de terapias. Adicionalmente, el descubrimiento de estas fusiones génicas quinasa podría estar vinculado al desarrollo de ALM debido a las funciones que presentan sus dominios, sin embargo, esta asociación debe ser probada en estudios posteriores.

Conclusiones

- Se detectaron 337 fusiones génicas en 80 muestras de pacientes ALM de las 112 muestras analizadas. Las fusiones detectadas no se han encontrado anteriormente en bases de datos como FusionGDB y COSMIC.
- Las fusiones génicas con una mayor frecuencia (8-15%) en las muestras son potenciales candidatos a ser marcadores en ALM. Los genes (*RP11-680G10.1*, *GSE1*, *SAMD5*, *SASH1*) que las integran se han encontrado en estudios anteriores de ALM.
- Las recombinaciones cromosómicas de los genes que integran distintas fusiones génicas halladas en este estudio están asociados a distintos tipos de cáncer.
- Las fusiones génicas quinasa presentaron una cobertura total de 16.96% (19/112) de muestras tumorales y una frecuencia individual de 0.89% y 1.79%, para fusiones génicas proteína quinasa presentes en una y dos muestras, respectivamente. Por ello su uso potencial como marcadores tumorales se ve reducido. Los genes y sus respectivos dominios proteicos están involucrados en procesos oncogénicos.

Recomendaciones

- Se detectaron fusiones génicas con una frecuencia mayor al 10%, teniendo un potencial de ser usadas como marcadores en ALM, no obstante, se necesita realizar un mayor número de estudios para probar si son marcadores ideales.
- Las fusiones génicas quinasa se encontraron en una frecuencia baja, sin embargo, no se puede descartar la posibilidad de ser parte de un subgrupo de ALM y que el estudio de su funcionalidad podría ser relevante para entender el desarrollo de ALM.

Bibliografía

- Abbott, K. L., Nyre, E. T., Abrahante, J., Ho, Y.-Y., Isaksson Vogel, R., & Starr, T. K. (2015). The Candidate Cancer Gene Database: A database of cancer driver genes from forward genetic screens in mice. *Nucleic Acids Research*, *43*(Database issue), D844–D848.
- Aguilera, A., & Rothstein, R. (Eds.). (2007). *Molecular genetics of recombination*. Springer.
- Arter, C., Trask, L., Ward, S., Yeoh, S., & Bayliss, R. (2022). Structural features of the protein kinase domain and targeted binding by small-molecule inhibitors. *Journal of Biological Chemistry*, *298*(8).
- Ascierto, P. A., Kirkwood, J. M., Grob, J.-J., Simeone, E., Grimaldi, A. M., Maio, M., Palmieri, G., Testori, A., Marincola, F. M., & Mozzillo, N. (2012). The role of BRAF V600 mutation in melanoma. *Journal of Translational Medicine*, *10*(1), 85.
- Bamodu, O. A., Wang, Y.-H., Ho, C.-H., Hu, S.-W., Lin, C.-D., Tzou, K.-Y., Wu, W.-L., Chen, K.-C., & Wu, C.-C. (2021). Genetic Suppressor Element 1 (GSE1) Promotes the Oncogenic and Recurrent Phenotypes of Castration-Resistant Prostate Cancer by Targeting Tumor-Associated Calcium Signal Transducer 2 (TACSTD2). *Cancers*, *13*(16), Article 16.
- Basurto-Lozada, P., Molina-Aguilar, C., Castaneda-Garcia, C., Vázquez-Cruz, M. E., Garcia-Salinas, O. I., Álvarez-Cano, A., Martínez-Said, H., Roldán-Marín, R., Adams, D. J., Possik, P. A., & Robles-Espinoza, C. D. (2021). Acral lentiginous melanoma: Basic facts, biological characteristics and research perspectives of an understudied disease. *Pigment Cell & Melanoma Research*, *34*(1), 59–71.
- Boffetta, P., van der Hel, O., Norppa, H., Fabianova, E., Fucic, A., Gundy, S., Lazutka, J., Cebulska-Wasilewska, A., Puskaierova, D., Znaor, A., Kelecsenyi, Z., Kurtinaitis,

- J., Rachtan, J., Forni, A., Vermeulen, R., & Bonassi, S. (2007). Chromosomal Aberrations and Cancer Risk: Results of a Cohort Study from Central Europe. *American Journal of Epidemiology*, 165(1), 36–43.
- Botton, T., Talevich, E., Mishra, V. K., Zhang, T., Shain, A. H., Berquet, C., Gagnon, A., Judson, R. L., Ballotti, R., Ribas, A., Herlyn, M., Rocchi, S., Brown, K. M., Hayward, N. K., Yeh, I., & Bastian, B. C. (2019). Genetic Heterogeneity of BRAF Fusion Kinases in Melanoma Affects Drug Responses. *Cell Reports*, 29(3), 573-588.e7.
- Chai, P., Tian, J., Zhao, D., Zhang, H., Cui, J., Ding, K., & Liu, B. (2016). GSE1 negative regulation by miR-489-5p promotes breast cancer cell proliferation and invasion. *Biochemical and Biophysical Research Communications*, 471(1), 123–128.
- Chang, F., Lin, F., Cao, K., Surrey, L. F., Aplenc, R., Bagatell, R., Resnick, A. C., Santi, M., Storm, P. B., Tasian, S. K., Waanders, A. J., Hunger, S. P., & Li, M. M. (2019). Development and Clinical Validation of a Large Fusion Gene Panel for Pediatric Cancers. *The Journal of Molecular Diagnostics : JMD*, 21(5), 873–883.
- Coras, N., Morales, D., Yabar, A., & Beltran, B. E. (2013). Prognosis of melanoma in Peru: An analysis of 410 cases. *Journal of Clinical Oncology*, 31(15_suppl), e20023–e20023.
- Costello, C. M., Pittelkow, M. R., & Mangold, A. R. (2017). Acral Melanoma and Mechanical Stress on the Plantar Surface of the Foot. *New England Journal of Medicine*, 377(4), 403–405.
- Cremer, T., & Cremer, M. (2010). Chromosome Territories. *Cold Spring Harbor Perspectives in Biology*, 2(3), a003889.
- Dai, X., Theobard, R., Cheng, H., Xing, M., & Zhang, J. (2018). Fusion genes: A promising tool combating against cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1869(2), 149–160.

- De Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., & Hulo, N. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, *34*, W362-365.
- Dong, X., & Chen, R. (2020). Understanding aberrant RNA splicing to facilitate cancer diagnosis and therapy. *Oncogene*, *39*(11), Article 11.
- Downes, C. E. J., Rehn, J., Heatley, S. L., Yeung, D., McClure, B. J., & White, D. L. (2022). Identification of a novel GOLGA4–JAK2 fusion gene in B-cell acute lymphoblastic leukaemia. *British Journal of Haematology*, *196*(3), 700–705. <https://doi.org/10.1111/bjh.17910>
- Druker, B. J. (2008). Translation of the Philadelphia chromosome into therapy for CML. *Blood*, *112*(13), 4808–4817. <https://doi.org/10.1182/blood-2008-07-077958>
- Elder, D. E., Bastian, B. C., Cree, I. A., Massi, D., & Scoyler, R. A. (2020). *The 2018 World Health Organization Classification of Cutaneous, Mucosal, and Uveal Melanoma: Detailed Analysis of 9 Distinct Subtypes Defined by Their Evolutionary Pathway | Archives of Pathology & Laboratory Medicine*. *4*. <https://meridian.allenpress.com/aplm/article/144/4/500/427497/The-2018-World-Health-Organization-Classification>
- Ewels, P., Magnusson, M., Lundin, S., & Källér, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048.
- Faulkner, C., Ellis, H. P., Shaw, A., Penman, C., Palmer, A., Wragg, C., Greenslade, M., Haynes, H. R., Williams, H., Lewis, S., White, P., Williams, M., Capper, D., & Kurian, K. M. (2015). BRAF Fusion Analysis in Pilocytic Astrocytomas: KIAA1549-BRAF 15-9 Fusions Are More Frequent in the Midline Than Within the Cerebellum. *Journal of Neuropathology and Experimental Neurology*, *74*(9), 867–872.

- Fior, R., & Zilhão, R. (2019). *Molecular and Cell Biology of Cancer* (Springer).
<https://link.springer.com/book/10.1007/978-3-030-11812-9>
- Gisselsson, D., Jonson, T., Yu, C., Martins, C., Mandahl, N., Wiegant, J., Jin, Y., Mertens, F., & Jin, C. (2002). Centrosomal abnormalities, multipolar mitoses, and chromosomal instability in head and neck tumours with dysfunctional telomeres. *British Journal of Cancer*, *87*(2), 202–207. <https://doi.org/10.1038/sj.bjc.6600438>
- Goldstein, A. M., Struewing, J. P., Chidambaram, A., Fraser, M. C., & Tucker, M. A. (2000). Genotype–Phenotype Relationships in U.S. Melanoma-Prone Families With CDKN2A and CDK4 Mutations. *JNCI: Journal of the National Cancer Institute*, *92*(12), 1006–1010.
- Haas, B. J., Dobin, A., Ghandi, M., Arsdale, A. V., Tickle, T., Robinson, J. T., Gillani, R., Kasif, S., & Regev, A. (2021). *Targeted in silico characterization of fusion transcripts in tumor and normal tissues via FusionInspector* (p. 2021.08.02.454639). bioRxiv.
- Haas, B. J., Dobin, A., Stransky, N., Li, B., Yang, X., Tickle, T., Bankapur, A., Ganote, C., Doak, T. G., Pochet, N., Sun, J., Wu, C. J., Gingeras, T. R., & Regev, A. (2017). *STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq* (p. 120295). bioRxiv.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), Article 8. <https://doi.org/10.1038/nprot.2013.084>
- Hayward, N. K., Wilmott, J. S., Waddell, N., Johansson, P. A., Field, M. A., Nones, K., Patch, A.-M., Kakavand, H., Alexandrov, L. B., Burke, H., Jakrot, V., Kazakoff, S.,

- Holmes, O., Leonard, C., Sabarinathan, R., Mularoni, L., Wood, S., Xu, Q., Waddell, N., ... Mann, G. J. (2017). Whole-genome landscapes of major melanoma subtypes. *Nature*, *545*(7653), Article 7653. <https://doi.org/10.1038/nature22071>
- Huang, K., Fan, J., & Misra, S. (2020). Acral Lentiginous Melanoma: Incidence and Survival in the United States, 2006-2015, an Analysis of the SEER Registry. *Journal of Surgical Research*, *251*, 329–339. <https://doi.org/10.1016/j.jss.2020.02.010>
- Inamdar, G. S., Madhunapantula, S. V., & Robertson, G. P. (2010). Targeting the MAPK pathway in melanoma: Why some approaches succeed and other fail. *Biochemical Pharmacology*, *80*(5), 624–637. <https://doi.org/10.1016/j.bcp.2010.04.029>
- Jung, K. H., Kim, J. K., Noh, J. H., Eun, J. W., Bae, H. J., Xie, H. J., Ahn, Y. M., Park, W. S., Lee, J. Y., & Nam, S. W. (2010). Targeted disruption of Nemo-like kinase inhibits tumor cell growth by simultaneous suppression of cyclin D1 and CDK2 in human hepatocellular carcinoma. *Journal of Cellular Biochemistry*, *110*(3), 687–696. <https://doi.org/10.1002/jcb.22579>
- Kaskel, P., Kind, P., Sander, S., Peter, R. U., & Krähn, G. (2000). Trauma and melanoma formation: A true association? *The British Journal of Dermatology*, *143*(4), 749–753.
- Keane, S., Améen, S., Lindlöf, A., & Ejeskär, K. (2020). Low DLG2 gene expression, a link between 11q-deleted and MYCN-amplified neuroblastoma, causes forced cell cycle progression, and predicts poor patient survival. *Cell Communication and Signaling*, *18*(1), 65. <https://doi.org/10.1186/s12964-020-00553-6>
- Kim, P., & Zhou, X. (2019). FusionGDB: Fusion gene annotation DataBase. *Nucleic Acids Research*, *47*(D1), D994–D1004.
- Kong, Y., Sheng, X., Wu, X., Yan, J., Ma, M., Yu, J., Si, L., Chi, Z., Cui, C., Dai, J., Li, Y., Yu, H., Xu, T., Tang, H., Tang, B., Mao, L., Lian, B., Wang, X., Yan, X., ... Guo, J.

- (2017). Frequent Genetic Aberrations in the CDK4 Pathway in Acral Melanoma Indicate the Potential for CDK4/6 Inhibitors in Targeted Therapy. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 23(22), 6946–6957.
- Kuchelmeister, C., Schaumburg-Lever, G., & Garbe, C. (2000). Acral cutaneous melanoma in caucasians: Clinical features, histopathology and prognosis in 112 patients. *British Journal of Dermatology*, 143(2), 275–280.
- Kumar, R., Sanawar, R., Li, X., & Li, F. (2017). Structure, Biochemistry, and Biology of PAK Kinases. *Gene*, 605, 20–31.
- Levitt, N. C., & Hickson, I. D. (2002). Caretaker tumour suppressor genes that defend genome integrity. *Trends in Molecular Medicine*, 8(4), 179–186.
- Li, M., Ke, J., Wang, Q., Qian, H., Yang, L., Zhang, X., Xiao, J., Ding, H., Shan, X., Liu, Q., Xiao, Y., Bao, B., & Huang, H. (2017). Upregulation of ROCK2 in gastric cancer cell promotes tumor cell proliferation, metastasis and invasion. *Clinical and Experimental Medicine*, 17(4), 519–529.
- Liang, W. S., Hendricks, W., Kiefer, J., Schmidt, J., Sekar, S., Carpten, J., Craig, D. W., Adkins, J., Cuyugan, L., Manojlovic, Z., Halperin, R. F., Helland, A., Nasser, S., Legendre, C., Hurley, L. H., Sivaprakasam, K., Johnson, D. B., Crandall, et al. (2017). Integrated genomic analyses reveal frequent TERT aberrations in acral melanoma. *Genome Research*, 27(4), 524–532.
- Lino-Silva, L. S., Domínguez-Rodríguez, J. A., Aguilar-Romero, J. M., Martínez-Said, H., Salcedo-Hernández, R. A., García-Pérez, L., Herrera-Gómez, Á., & Cuellar-Hubbe, M. (2016). Melanoma in Mexico: Clinicopathologic Features in a Population with Predominance of Acral Lentiginous Subtype. *Annals of Surgical Oncology*, 23(13), 4189–4194.

- Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Bretscher, A., Ploegh, H., & Amon, A. (2016). *Biologia Celular y Molecular* (7th ed.). Panamericana.
- Lopez, G., Conkrite, K. L., Doepner, M., Rathi, K. S., Modi, A., Vaksman, Z., Farra, L. M., Hyson, E., Nouredine, M., Wei, J. S., Smith, M. A., Asgharzadeh, S., Seeger, R. C., Khan, J., Auvil, J. G., Gerhard, D. S., Maris, J. M., & Diskin, S. J. (2019). *Somatic structural variation targets neurodevelopmental genes and identifies SHANK2 as a tumor suppressor in neuroblastoma* (p. 572248). bioRxiv.
- Ma, S. K., & Wan, T. S. K. (2000). Single Autosomal Trisomy in Acute Myeloid Leukemia and Myelodysplastic Syndrome. *Current Genomics*, 1(2), 153–173.
- Maekawa, S., Ueda, Y., & Yanagisawa, S. (2018). Overexpression of a Brix Domain-Containing Ribosome Biogenesis Factor ARPF2 and its Interactor ARRS1 Causes Morphological Changes and Lifespan Extension in Arabidopsis thaliana. *Frontiers in Plant Science*, 9.
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The Protein Kinase Complement of the Human Genome. *Science*, 298(5600), 1912–1934.
- McEvoy, C. R., Xu, H., Smith, K., Etemadmoghadam, D., San Leong, H., Choong, D. Y., Byrne, D. J., Irvani, A., Beck, S., Mileshekin, L., Tothill, R. W., Bowtell, D. D., Bates, B. M., Nasteovski, V., Browning, J., Bell, A. H., Khoo, C., Desai, J., Fellowes, et al. (2019). Profound MEK inhibitor response in a cutaneous melanoma harboring a GOLGA4-RAF1 fusion. *The Journal of Clinical Investigation*, 129(5), 1940–1945.
- McMeniman, E. K., Duffy, D. L., Jagirdar, K., Lee, K. J., Peach, E., McLnerney-Leo, A. M., De'Ambrosis, B., Rayner, J. E., Smithers, B. M., Soyer, H. P., & Sturm, R. A. (2020). The interplay of sun damage and genetic risk in Australian multiple and single primary melanoma cases and controls. *The British Journal of Dermatology*, 183(2), 357–366.

- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1), D412–D419.
- Mitelman, F., Johansson, B., & Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*, 7(4), Article 4.
- Möhrle, M., & Häfner, H. M. (2002). Is subungual melanoma related to trauma? *Dermatology (Basel, Switzerland)*, 204(4), 259–261.
- Murph, M. (2015). *Melanoma: Current Clinical Management and Future Therapeutics* (31-78). BoD – Books on Demand.
- Nicolson, N. G., & Han, D. (2019). Desmoplastic melanoma. *Journal of Surgical Oncology*, 119(2), 208–215.
- Nikiforov, Y. E. (2002). RET/PTC rearrangement in thyroid tumors. *Endocrine Pathology*, 13(1), 3–16.
- Niu, H.-T., Zhou, Q.-M., Wang, F., Shao, Q., Guan, Y.-X., Wen, X.-Z., Chen, L.-Z., Feng, Q.-S., Li, W., Zeng, Y.-X., & Zhang, X.-S. (2013). Identification of anaplastic lymphoma kinase break points and oncogenic mutation profiles in acral/mucosal melanomas. *Pigment Cell & Melanoma Research*, 26(5), 646–653.
- Oba, J., & Woodman, S. E. (2021). The genetic and epigenetic basis of distinct melanoma types. *The Journal of Dermatology*, 48(7), 925–939.
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40(12), Article 12.
- Parl, F. F., Dupont, W. D., & Crooke, P. S. (2019). Interchromosomal Translocations as a Means to Map Chromosome Territories in Breast Cancer. *Cancer Informatics*, 18.

- Paul, M. K., & Mukhopadhyay, A. K. (2004). Tyrosine kinase – Role and significance in Cancer. *International Journal of Medical Sciences*, 1(2), 101–115.
- Pleasance, E. D., Cheetham, R. K., Stephens, P. J., McBride, D. J., Humphray, S. J., Greenman, C. D., Varela, I., Lin, M.-L., Ordóñez, G. R., Bignell, G. R., Ye, K., Alipaz, J., Bauer, M. J., Beare, D., Butler, A., Carter, R. J., Chen, L., Cox, A. J., Edkins, S., ... Stratton, M. R. (2010). A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*, 463(7278), Article 7278.
- Pradhan, D., Jour, G., Milton, D., Vasudevaraja, V., Tetzlaff, M. T., Nagarajan, P., Curry, J. L., Ivan, D., Long, L., Ding, Y., Ezhilarasan, R., Sulman, E. P., Diab, A., Hwu, W.-J., Prieto, V. G., Torres-Cabala, C. A., & Aung, P. P. (2019). Aberrant DNA Methylation Predicts Melanoma-Specific Survival in Patients with Acral Melanoma. *Cancers*, 11(12), Article 12.
- Puig-Butillé, J. A., Badenas, C., Ogbah, Z., Carrera, C., Aguilera, P., Malveyh, J., & Puig, S. (2013). Genetic alterations in RAS-regulated pathway in acral lentiginous melanoma. *Experimental Dermatology*, 22(2), 148–150.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), 841–842.
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. <http://www.R-project.org/>
- Rabbie, R., Ferguson, P., Molina-Aguilar, C., Adams, D. J., & Robles-Espinoza, C. D. (2019). Melanoma subtypes: Genomic profiles, prognostic molecular markers and therapeutic possibilities. *The Journal of Pathology*, 247(5), 539–551.
- Rajagopalan, H., & Lengauer, C. (2004). Aneuploidy and cancer. *Nature*, 432(7015), Article 7015.
- Rath, N., & Olson, M. F. (2012). Rho-associated kinases in tumorigenesis: Re-considering ROCK inhibition for cancer therapy. *EMBO Reports*, 13(10), 900–908.

- Robinson, D. R., Kalyana-Sundaram, S., Wu, Y.-M., Shankar, S., Cao, X., Ateeq, B., Asangani, I. A., Iyer, M., Maher, C. A., Grasso, C. S., Lonigro, R. J., Quist, M., Siddiqui, J., Mehra, R., Jing, X., Giordano, T. J., Sabel, M. S., Kleer, C. G., Palanisamy, N., ... Chinnaiyan, A. M. (2011). Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nature Medicine*, *17*(12), Article 12.
- Robinson, D. R., Wu, Y.-M., & Lin, S.-F. (2000). The protein tyrosine kinase family of the human genome. *Oncogene*, *19*(49), Article 49.
- Roskoski, R. (2019). Cyclin-dependent protein serine/threonine kinase inhibitors as anticancer drugs. *Pharmacological Research*, *139*, 471–488.
- RStudio Team. (2021). RStudio: Integrated Development for R. *RStudio*, PBC.
<http://www.rstudio.com/>
- Saldanha, G., Potter, L., DaForno, P., & Pringle, J. H. (2006). Cutaneous Melanoma Subtypes Show Different BRAF and NRAS Mutation Frequencies. *Clinical Cancer Research*, *12*(15), 4499–4505.
- Schram, A. M., Chang, M. T., Jonsson, P., & Drilon, A. (2017). Fusions in solid tumours: Diagnostic strategies, targeted therapy, and acquired resistance. *Nature Reviews Clinical Oncology*, *14*(12), Article 12.
- Sharma, S. (2009). Tumor markers in clinical practice: General principles and guidelines. *Indian Journal of Medical and Paediatric Oncology : Official Journal of Indian Society of Medical & Paediatric Oncology*, *30*(1), 1–8.
- Shaw, A. T., Hsu, P. P., Awad, M. M., & Engelman, J. A. (2013). Tyrosine kinase gene rearrangements in epithelial malignancies. *Nature Reviews. Cancer*, *13*(11), 772–787.
- Solit, D. B., & Rosen, N. (2011). Resistance to BRAF Inhibition in Melanomas. *New England Journal of Medicine*, *364*(8), 772–774.

- Song, H., Bao, J., Wei, Y., Chen, Y., Mao, X., Li, J., Yang, Z., & Xue, Y. (2015). Kaempferol inhibits gastric cancer tumor growth: An in vitro and in vivo study. *Oncology Reports*, 33(2), 868–874.
- Srivastava, A., Sarkar, H., Gupta, N., & Patro, R. (2016). RapMap: A rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. *Bioinformatics*, 32(12), i192–i200.
- Susok, L., & Gambichler, T. (2022). Caucasians with acral lentiginous melanoma have the same outcome as patients with stage- and limb-matched superficial spreading melanoma. *Journal of Cancer Research and Clinical Oncology*, 148(2), 497–502.
- Taniue, K., & Akimitsu, N. (2021). Fusion Genes and RNAs in Cancer Development. *Non-Coding RNA*, 7(1), 10.
- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., ... Forbes, S. A. (2019). COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, 47(D1), D941–D947.
- Troyanova, P. (2002). The role of trauma in the melanoma formation. *Journal of B.U.ON.: Official Journal of the Balkan Union of Oncology*, 7(4), 347–350.
- Tumor Markers—NCI (nciglobal,ncicenterprise). (2021, May 27). <https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-markers-fact-sheet>
- Turner, J., Coutts, K., Sheren, J., Saichaemchan, S., Ariyawutyakorn, W., Avolio, I., Cabral, E., Glogowska, M., Amato, C., Robinson, S., Hintzsche, J., Applegate, A., Seelenfreund, E., Gonzalez, R., Wells, K., Bagby, S., Tentler, J., Tan, A.-C., et al. (2017). Kinase gene fusions in defined subsets of melanoma. *Pigment Cell & Melanoma Research*, 30(1), 53–62.

- Ullmann, T. M., Thiesmeyer, J. W., Lee, Y. J., Beg, S., Mosquera, J. M., Elemento, O., Fahey, T. J., Scognamiglio, T., & Houvras, Y. (2022). RET Fusion-Positive Papillary Thyroid Cancers are Associated with a More Aggressive Phenotype. *Annals of Surgical Oncology*, 29(7), 4266–4273.
- Van Poppelen, N. M., de Bruyn, D. P., Bicer, T., Verdijk, R., Naus, N., Mensink, H., Paridaens, D., de Klein, A., Brosens, E., & Kiliç, E. (2021). Genetics of Ocular Melanoma: Insights into Genetics, Inheritance and Testing. *International Journal of Molecular Sciences*, 22(1), Article 1.
- Vine, M. F. (1994). Biological Markers: Their Use in Quantitative Assessments. *Advances in Dental Research*, 8(1), 92–99.
- Vu, T. N., Deng, W., Trac, Q. T., Calza, S., Hwang, W., & Pawitan, Y. (2018). A fast detection of fusion genes from paired-end RNA-seq data. *BMC Genomics*, 19(1), 786.
- Wang, B., & Lee, N. H. (2018). Aberrant RNA Splicing in Cancer and Drug Resistance. *Cancers*, 10(11), Article 11.
- Wang, Q., Xia, J., Jia, P., Pao, W., & Zhao, Z. (2013). Application of next generation sequencing to human gene fusion detection: Computational tools, features and perspectives. *Briefings in Bioinformatics*, 14(4), 506–519.
- Wang, X., Ruan, W., Zhou, S., Feng, X., Yan, H., & Huang, F. (2020). Prevalence of molecular markers associated with drug resistance of Plasmodium vivax isolates in Western Yunnan Province, China. *BMC Infectious Diseases*, 20(1), 307.
- Wickham. (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.
<https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., & D'Agostino McGowan, L. (2019). Welcome to the {tidyverse}. *Journal of Open Source Software*, 4(43), 1686.

- Wickham, H., François, R., Lionel, H., & Kirill, M. (2021). *dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>
- Yasuda, J., Tsuchiya, A., Yamada, T., Sakamoto, M., Sekiya, T., & Hirohashi, S. (2003). Nemo-like kinase induces apoptosis in DLD-1 human colon cancer cells. *Biochemical and Biophysical Research Communications*, *308*(2), 227–233.
- Yeh, I., & Bastian, B. c. (2021). Melanoma pathology: New approaches and classification*. *British Journal of Dermatology*, *185*(2), 282–293.
- Yeh, I., Jorgenson, E., Shen, L., Xu, M., North, J. P., Shain, A. H., Reuss, D., Wu, H., Robinson, W. A., Olshen, A., von Deimling, A., Kwok, P.-Y., Bastian, B. C., & Asgari, M. M. (2019). Targeted Genomic Profiling of Acral Melanoma. *JNCI: Journal of the National Cancer Institute*, *111*(10), 1068–1077.
- Zhang, N., Wang, L., Zhu, G. n., Sun, D. j., He, H., Luan, Q., Liu, L., Hao, F., Li, C. y., & Gao, T. w. (2014). The association between trauma and melanoma in the Chinese population: A retrospective study. *Journal of the European Academy of Dermatology and Venereology*, *28*(5), 597–603.
- Zhang, Y., McCord, R. P., Ho, Y.-J., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., Becker, M. S., Alt, F. W., & Dekker, J. (2012). Chromosomal translocations are guided by the spatial organization of the genome. *Cell*, *148*(5), 908–921.
- Zhao, Z., Jin, Q., Xu, J.-R., & Liu, H. (2014). Identification of a Fungi-Specific Lineage of Protein Kinases Closely Related to Tyrosine Kinases. *PLOS ONE*, *9*(2), e89813.

Anexos

Anexo 1: Permiso de acceso a los datos de pacientes

Permiso de acceso

Yo, Carla Daniela Robles Espinoza, investigadora principal del laboratorio internacional para la investigación del genoma humano (LIGH-UNAM), por medio del presente documento autorizo el acceso de Ana Cecilia Romani Vasquez, bachiller de genética y Biotecnología de la Universidad Nacional Mayor de San Marcos (UNMSM), a los datos de los pacientes del INCan de la cohorte 2019, como parte integral de nuestro grupo de investigación y a través del clúster de supercómputo del LAVIS (UNAM).



Carla Daniela Robles Espinoza

Ana Cecilia Romani Vasquez

Nombre y firma de usuario

Anexo 2: Datos por muestra del análisis de calidad de los archivos sin procesar

Número	Muestra	Duplicado (%)	Cont_GC (%)	Longitud (bp)	Secuencias por millón
1	PR40985a	31.7	55	75	279.9
2	PR41043d	29.4	56	75	276.1
3	PR40968a	30.5	54	75	269
4	PR41921a	28.8	54	75	261.9
5	PR40989a	33.7	54	75	251
6	PR40963a	33.7	61	75	245.9
7	PR41013a	37.8	52	75	243
8	PR40966d	26.3	54	75	238.6
9	PR41025a	23.8	56	75	238.1
10	PR40962a	41.3	55	75	232
11	PR41035a	29.1	52	75	211.2
12	PR40986a	38.5	58	75	202.2
13	PR41919a	42.2	51	75	200.7
14	PR40960a	28	54	75	195.2
15	PR41902a	44.7	51	75	191.4
16	PR40994a	26.1	53	75	187.9
17	PR41017a	29.1	56	75	182.4
18	PR40966e	24.9	54	75	179.5

19	PR40988a	25.2	56	75	176.4
20	PR41044a	26.3	55	75	169.5
21	PR41917a	28.8	51	75	164
22	PR41896c	23.6	53	75	163.7
23	PR41046e	26	56	75	163.2
24	PR40961e	25	53	75	163.1
25	PR41029e	26.5	58	75	162.9
26	PR41905a	24.7	53	75	162.1
27	PR41916a	24.8	56	75	161.3
28	PR40956d	25.4	54	75	161.3
29	PR40997a	30.3	54	75	159.6
30	PR40970a	24	54	75	159.5
31	PR41913a	27.9	53	75	155.2
32	PR41020a	35.9	53	75	154.8
33	PR41020d	41.4	48	75	153.3
34	PR40984a	24.8	53	75	153
35	PR41912a	40.7	59	75	152.5
36	PR40978d	24.4	56	75	151.3
37	PR40969a	28.8	52	75	149.1
38	PR41897a	27.6	53	75	147.1
39	PR41922d	23.1	52	75	145.8
40	PR41011a	33.8	50	75	145.8
41	PR41037a	25.6	54	75	141.8
42	PR41915a	30.2	51	75	141.8
43	PR40967a	22.3	60	75	140.4
44	PR41029g	24.7	56	75	139.2
45	PR40966a	22	55	75	136.8
46	PR40974a	34.9	60	75	132.5
47	PR40983a	40	58	75	130.2
48	PR40971a	27.7	53	75	129.5
49	PR41026d	26.4	54	75	128.6
50	PR41003a	28.3	52	75	128.5
51	PR41025d	23.7	53	75	128.1
52	PR40996a	24	54	75	127.2
53	PR41030a	30.3	53	75	119.7
54	PR41911a	34.7	52	75	119.4
55	PR40965f	24.9	51	75	119.2
56	PR41920c	26.2	54	75	116.5

57	PR41002a	25.4	54	75	116.1
58	PR41908a	31.7	50	75	115.4
59	PR41039d	24	54	75	114.5
60	PR41900a	21.9	57	75	114.1
61	PR41901a	20.6	55	75	111.2
62	PR40990a	20.6	54	75	111.1
63	PR41922c	23.1	51	75	111.1
64	PR41019a	24.6	55	75	108.8
65	PR40972a	21.6	54	75	107.4
66	PR41005a	31	51	75	106.3
67	PR41002e	26.9	52	75	106
68	PR40961d	22.8	53	75	104.2
69	PR41920a	24.7	55	75	103.9
70	PR41034a	19.7	55	75	102.1
71	PR41006a	42.7	58	75	98.7
72	PR41039a	23.4	56	75	97.9
73	PR40961a	24.2	51	75	97.9
74	PR41032a	22.1	54	75	97.7
75	PR41043a	23.1	58	75	97.2
76	PR41907a	23.6	52	75	95.9
77	PR40971d	20.3	54	75	94.8
78	PR41001a	42.3	52	75	93.1
79	PR40969d	19.3	54	75	91.9
80	PR41038a	22.4	60	75	90.9
81	PR41036a	18.2	53	75	90.6
82	PR41909a	24.4	55	75	86.4
83	PR40976a	20.8	53	75	84.3
84	PR41915c	23.8	53	75	80.5
85	PR41000a	27.1	52	75	78.9
86	PR41895a	23.8	53	75	78.3
87	PR41027a	17.2	54	75	77.7
88	PR41035d	25.5	51	75	76.2
89	PR41007a	34	56	75	76.1
90	PR40980a	21.8	52	75	75.2
91	PR40982a	22.3	54	75	73.8
92	PR40987a	23.7	53	75	72.1
93	PR40983d	21.6	51	75	71.1
94	PR41042a	35	49	75	69.1

95	PR41033a	35.7	77	75	66.7
96	PR41900c	27.1	59	75	64.1
97	PR41021d	15.5	55	75	62.6
98	PR40987d	21.4	53	75	59
99	PR41910a	30.3	57	75	58.6
100	PR41906a	24	59	75	54.9
101	PR41910c	18.8	58	75	54.4
102	PR40964a	20	55	75	51.6
103	PR41046a	35.7	70	75	49.5
104	PR40980d	20.8	56	75	48.8
105	PR40973a	18.6	54	75	45.6
106	PR40983e	39.2	48	75	40.9
107	PR40986d	15.5	53	75	34.6
108	PR40978a	41.8	58	75	29.7
109	PR40956a	34.4	73	75	20
110	PR40952a	16.5	66	75	17.5
111	PR41907c	41.5	48	75	7.2
112	PR41014a	56	48	75	2.5

Anexo 3: Frecuencia de las 10 primeras fusiones génicas de Fuseq. Se indica la frecuencia total de cada gen que conforma la fusión génica en las muestras tumorales.

Fusión génica	Frecuencia	Gen	Frecuencia
<i>TPRX1--GPNMB</i>	53	<i>TPRX1</i>	53
		<i>GPNMB</i>	53
<i>RP11-680G10.1--GSE1</i>	31	<i>RP11-680G10.1</i>	32
		<i>GSE1</i>	32
<i>SAMD5--SASH1</i>	24	<i>SAMD5</i>	24
		<i>SASH1</i>	24
<i>MRPL35--ADK</i>	19	<i>MRPL35</i>	19
		<i>ADK</i>	19
<i>CTSC--RAB38</i>	14	<i>CTSC</i>	14
		<i>RAB38</i>	15
<i>LINC00999--PCMTD2</i>	12	<i>LINC00999</i>	12
		<i>PCMTD2</i>	12
<i>GABRP--SCGB3A2</i>	11	<i>GABRP</i>	15
		<i>SCGB3A2</i>	15
<i>SIRPB1--SIRPA</i>	9	<i>SIRPB1</i>	9
		<i>SIRPA</i>	10
<i>ERI2--ZNF586</i>	8	<i>ERI2</i>	9
		<i>ZNF586</i>	11
<i>TMED3--KIAA1024</i>	7	<i>TMED3</i>	7
		<i>KIAA1024</i>	7

Anexo 4: Los 9 primeros genes que presentan un mayor número de genes compañeros de Fuseq.

Gen	Gen_ID	N° de genes compañeros
<i>DLG2</i>	ENSG00000150672	9
<i>ADCY2</i>	ENSG00000078295	7
<i>UVRAG</i>	ENSG00000198382	7
<i>GAB2</i>	ENSG00000033327	6
<i>MYH9</i>	ENSG00000100345	6
<i>GOLPH3</i>	ENSG00000113384	5
<i>TENM4</i>	ENSG00000149256	5
<i>DYNC2H1</i>	ENSG00000187240	5
<i>WDR70</i>	ENSG00000082068	4

Anexo 5: Frecuencia de las 10 primeras fusiones génicas de STAR-Fusion. Presenta la frecuencia total de cada gen que conforma la fusión.

Fusión génica	Frecuencia	Gen	Frecuencia
<i>RP1-34H18.1--NAV3</i>	83	<i>RP1-34H18.1</i>	83
		<i>NAV3</i>	83
<i>TULP4--RP11-732M18.3</i>	80	<i>TULP4</i>	80
		<i>RP11-732M18.3</i>	80
<i>TIMM23--PARGP1</i>	47	<i>TIMM23</i>	47
		<i>PARGP1</i>	47
<i>SNORA50--CNOT1</i>	46	<i>SNORA50</i>	46
		<i>CNOT1</i>	47
<i>RP11-680G10.1--GSE1</i>	32	<i>RP11-680G10.1</i>	32
		<i>GSE1</i>	32

<i>RP11-367G6-3--FAM65B</i>	30	<i>RP11-367G6-3</i>	30
		<i>FAM65B</i>	30
<i>HLA-S--HLA-A</i>	29	<i>HLA-S</i>	29
		<i>HLA-A</i>	29
<i>RPA3-AS1--GLCCI1</i>	18	<i>RPA3-AS1</i>	18
		<i>GLCCI1</i>	18
<i>PFKB3--RP11-563J2.2</i>	14	<i>PFKB3</i>	14
		<i>RP11-563J2.2</i>	14
<i>RP11-14D22.2--PRICKLE2</i>	13	<i>RP11-14D22.2</i>	13
		<i>PRICKLE2</i>	13

Anexo 6: Los 9 primeros genes que presentan un mayor número de genes compañeros de STAR-Fusion

Gen	Gen_ID	N° de genes compañeros
<i>ADCY2</i>	ENSG00000078295.11	9
<i>CHCHD3</i>	ENSG00000106554.7	7
<i>DLG2</i>	ENSG00000150672.12	7
<i>UVRAG</i>	ENSG00000198382.4	7
<i>DIP2B</i>	ENSG00000066084.8	6
<i>FCHSD2</i>	ENSG00000137478.10	6
<i>GAB2</i>	ENSG00000033327.8	6
<i>LRBA</i>	ENSG00000198589.6	6
<i>NUP155</i>	ENSG00000113569.11	6

Anexo 7: Frecuencia de las 10 primeras fusiones génicas de Fuseq y STAR-Fusion. Se indica su respectiva medida de expresión media y la frecuencia total de los genes que los conforman.

Fusión génica	Frecuencia	Media de los fragmentos fusión por millón	Gen	Frecuencia
<i>RP11-680G10.1--GSE1</i>	17	0.1860	<i>RP11-680G10.1</i>	17
			<i>GSE1</i>	17
<i>SAMD5--SASH1</i>	9	0.1299	<i>SAMD5</i>	9
			<i>SASH1</i>	9
<i>CYP4B1--CYP4X1</i>	4	0.2682	<i>CYP4B1</i>	4
			<i>CYP4X1</i>	4
<i>PHF20--CHD6</i>	3	0.4135	<i>PHF20</i>	3
			<i>CHD6</i>	3
<i>CLEC16A--SNX29</i>	3	0.1953	<i>CLEC16A</i>	3
			<i>SNX29</i>	3
<i>MED15--TTC28</i>	3	0.5453	<i>MED15</i>	5
			<i>TTC28</i>	4
<i>CTSC--RAB38</i>	3	0.1279	<i>CTSC</i>	3
			<i>RAB38</i>	3
<i>FAR1--SBF2</i>	3	0.3388	<i>FAR1</i>	3

			<i>SBF2</i>	3
<i>AIFM2-- PLA2G6</i>	2	0.4133	<i>AIFM2</i>	2
			<i>PLA2G6</i>	2
<i>CBFB--PTGIS</i>	2	0.1279	<i>CBFB</i>	2
			<i>PTGIS</i>	2

Anexo 8: Validación de fusiones génicas quinasas. La segunda y tercera fila indican el número de lecturas de soporte, las cuales han sido clasificadas en lecturas de unión (JunctionReadCount) y extensión (SpanningFragCount). La cuarta y quinta columna indican el punto de corte en la secuencia del gen 5' y gen 3', respectivamente. El término FFPM se refiere a los fragmentos de las fusiones génicas del total fragmentos por millón. La séptima columna indica el tipo de fusión con respecto a si se presenta dentro del marco de lectura (INFRAME) o fuera de este (FRAMESHIFT).

Fusion gene	JunctionReadCount	SpanningFragCount	LeftBreakpoint	RightBreakpoint	FFPM	PROT_FUSION_TYPE	Sample
<i>CDKL5--AP1S2</i>	48	6	chrX:18528974:+	chrX:15870647:-	0.678	INFRAME	2260_PR40970a
<i>MAP3K3--RAD51C</i>	13	0	chr17:61735267:+	chr17:56772292:+	0.17	FRAMESHIFT	2260_PR41020d
<i>NLK--ACACA</i>	90	17	chr17:26499644:+	chr17:35518964:-	1.9265	INFRAME	2260_PR41901a
<i>NLK--BCAS3</i>	17	3	chr17:26370357:+	chr17:58885365:+	0.3601	FRAMESHIFT	2260_PR41901a
<i>NLK--HLF</i>	16	1	chr17:26490654:+	chr17:53398025:+	0.2854	INFRAME	2260_PR41911a
<i>PASK--LMBR1</i>	18	8	chr2:242062147:-	chr7:156629579:-	0.2075	INFRAME	2260_PR40989a
<i>ROCK2--CACNA2D3</i>	7	2	chr2:11348413:-	chr3:54537519:+	0.5215	INFRAME	2260_PR40986d
<i>ROCK2--PICALM</i>	18	0	chr2:11484122:-	chr11:85670103:-	0.6157	INFRAME	2260_PR41910a
<i>BRIX1--MAP2K5</i>	35	12	chr5:34918580:+	chr15:67842372:+	1.0001	FRAMESHIFT	2260_PR40969d
<i>BRIX1--MAP2K5</i>	16	4	chr5:34916002:+	chr15:67842372:+	0.372	INFRAME	2260_PR40969d
<i>BRIX1--</i>	20	9	chr5:34918580:+	chr15:67842372:+	0.4704	FRAMESHIFT	2260_PR40972a

<i>MAP2K5</i>							
<i>BRX1-- MAP2K5</i>	39	6	chr5:34916002:+	chr15:67842372:+	0.7974	INFRAME	2260_PR40972a
<i>DLG2--PAK1</i>	9	1	chr11:85309701:-	chr11:77070062:-	0.1547	FRAMESHIFT	2260_PR40971a
<i>EIF2B3-- MAST2</i>	9	8	chr1:45443987:-	chr1:46290105:+	0.1362	INFRAME	2260_PR41025a
<i>GOLGA4-- BRAF</i>	44	8	chr3:37337710:+	chr7:140482957:-	1.0821	INFRAME	2260_PR41907a
<i>SMARCA5-- NEK1</i>	25	7	chr4:144467200:+	chr4:170506700:-	0.2554	INFRAME	2260_PR40989a
<i>TENM4-- ATM</i>	5	3	chr11:78419406:-	chr11:108099905:+	0.1541	FRAMESHIFT	2260_PR41920a
<i>TENM4-- ATM</i>	17	6	chr11:78419406:-	chr11:108099905:+	0.3948	FRAMESHIFT	2260_PR41920c
<i>TENM4-- ATM</i>	3	0	chr11:78419406:-	chr11:108114680:+	0.0516	FRAMESHIFT	2260_PR41920c
<i>TENM4-- ATM</i>	1	1	chr11:78419406:-	chr11:108106397:+	0.0179	FRAMESHIFT	2260_PR41920c
<i>TENM4-- ATM</i>	3	0	chr11:78419406:-	chr11:108114702:+	0.0516	FRAMESHIFT	2260_PR41920c

Anexo 9: Alineamiento en BLAT Ensembl. La segunda columna indica la anotación proteica que BLAT Ensembl asigna al resultado. La tercera columna indica el gen de donde proviene la secuencia con la que se realizó el alineamiento. La cuarta y quinta columna presenta el inicio y final del alineamiento en la secuencia proteica del gen de referencia, respectivamente. La sexta columna indica la localización genómica del gen referencia. La séptima y octava columna indica el inicio y final del alineamiento de la secuencia proteica objetivo, respectivamente. Mientras, la novena columna seria la longitud del alineamiento. El E-val indica un valor estadístico que determina si el resultado es significativo y no un resultado aleatorio. El valor ID indica el porcentaje de identidad de la secuencia proteica objetivo con respecto a la referencial.

Fusion gene	Protein annotation	Gene hit	Subject start	Subject end	Genomic Location	Query start	Query end	Length	E-val	ID
<i>SMARCA5--NEK1</i>	ENSP00000424757	NEK1	203	1286	4:169394513-169585549	841	1924	1084	0	100
	ENSP00000283131	SMARCA5	58	820	4:143514096-143545987	58	820	763	0	100
<i>ROCK2--PICALM</i>	ENSP00000317985	ROCK2	1	47	2:11343996-11344136	1	47	47	3.16E-27	100
	ENSP00000261535	ROCK2	1	47	2:11343996-11344136	1	47	47	4.39E-27	100
	ENSP00000513431	ROCK2	1	47	2:11343996-11344136	1	47	47	4.48E-27	100
<i>ROCK2--CACNA2D3</i>	ENSP00000288197	CACNA2D3	128	1091	3:54503492-55074203	789	1752	964	0	100
	ENSP00000419101	CACNA2D3	128	1091	3:54503492-55074203	789	1752	964	0	100
	ENSP00000389506	CACNA2D3	128	1091	3:54503492-55074203	789	1752	964	0	100

	ENSP0000041 7279	CACNA2 D3	34	997	3:54503492-55074203	789	1752	964	0	100
	ENSP0000041 8028	CACNA2 D3	34	126	3:54503492-54562915	789	881	93	3.77E- 59	100
	ENSP0000039 5957	ROCK2	12	107	2:11249661-11287703	59	154	96	9.31E- 57	100
<i>PASK-- LMBR1</i>	ENSP0000039 5672	PASK	16	193	2:241137007-241138794	201	378	178	1.42E- 107	100
	ENSP0000040 0734	PASK	1	140	2:241137073-241138746	217	356	140	3.53E- 81	100
	ENSP0000039 7922	PASK	1	102	2:241140644-241143032	1	102	102	2.30E- 62	100
	ENSP0000048 6149	PASK	1	65	2:241142838-241143032	1	65	65	7.28E- 36	100
	ENSP0000040 8506	PASK	1	65	2:241142838-241143032	1	65	65	7.28E- 36	100
	ENSP0000040 3052	LMBR1	1	35	7:156796389-156796493	1109	1143	35	1.49E- 17	100
<i>NLK--HLF</i>	ENSP0000038 4625	NLK	1	279	17:28042874-28163628	1	279	279	5.31E- 122	100
	ENSP0000043 3117	NLK	4	227	17:28043039-28163628	56	279	224	7.82E- 118	100
	ENSP0000046 0572	HLF	138	210	17:55315442-55320876	278	350	73	4.78E- 32	100
	ENSP0000046 0296	HLF	138	210	17:55315442-55320876	278	350	73	4.78E- 32	100
	ENSP0000040 2496	HLF	138	210	17:55315442-55320876	278	350	73	4.78E- 32	100
	ENSP0000022 6067	HLF	223	295	17:55315442-55320876	278	350	73	3.62E- 31	100
<i>NLK--</i>	ENSP0000048	ACACA	1584	2288	17:37087319-37179304	370	1073	705	0	98.

ACACA	2269									44
	ENSP0000048 3300	ACACA	1679	2383	17:37087319-37179304	370	1073	705	0	98.44
	ENSP0000047 9901	ACACA	1584	2288	CHR_HSCHR17_7_CTG4:37088787 -37180778	370	1073	705	0	98.44
	ENSP0000047 8577	ACACA	1679	2383	CHR_HSCHR17_7_CTG4:37088787 -37180778	370	1073	705	0	98.44
	ENSP0000048 2368	ACACA	1564	2268	17:37087319-37179304	370	1073	705	0	98.44
	ENSP0000047 8547	ACACA	1642	2346	17:37087319-37179304	370	1073	705	0	98.44
	ENSP0000048 0031	ACACA	1564	2268	CHR_HSCHR17_7_CTG4:37088787 -37180778	370	1073	705	0	98.44
	ENSP0000048 3674	ACACA	1642	2346	CHR_HSCHR17_7_CTG4:37088787 -37180778	370	1073	705	0	98.44
	ENSP0000038 4625	NLK	1	383	17:28042874-28172618	1	383	383	0	100
	ENSP0000043 3117	NLK	4	331	17:28043039-28172618	56	383	328	0	100
GOLGA4-- BRAF	ENSP0000041 9060	BRAF	393	767	7:140726497-140787548	294	668	375	0	100
	ENSP0000049 3783	BRAF	92	363	7:140749287-140787548	294	565	272	0	100
	ENSP0000051 4745	GOLGA4	42	293	3:37251446-37296218	42	293	252	2.66E-164	100
	ENSP0000051 4746	GOLGA4	42	293	3:37251446-37296218	42	293	252	3.01E-164	100
	ENSP0000051 4744	GOLGA4	42	293	3:37251446-37296218	42	293	252	3.01E-164	100
	ENSP0000034 9305	GOLGA4	42	293	3:37251446-37296218	42	293	252	3.56E-164	100

	ENSP0000040 5842	GOLGA4	42	293	3:37251446-37296218	42	293	252	4.37E- 164	100
<i>CDKL5-- AP1S2</i>	ENSP0000036 9645	AP1S2	1	160	X:15832954-15852524	34	193	160	1.29E- 115	100
	ENSP0000050 0066	AP1S2	1	142	X:15845379-15852524	34	175	142	4.50E- 101	100
	ENSP0000050 0850	AP1S2	1	142	X:15845379-15852524	34	175	142	4.50E- 101	100
	ENSP0000050 0737	AP1S2	1	142	X:15845379-15852524	34	175	142	1.10E- 100	100
	ENSP0000050 0798	AP1S2	1	142	X:15845379-15852524	34	175	142	1.10E- 100	100
	ENSP0000040 3498	AP1S2	2	139	X:15845379-15852512	38	175	138	3.06E- 97	100
	ENSP0000048 5359	CDKL5	1	33	X:18507097-18510854	1	33	33	3.18E- 17	100
	ENSP0000048 5581	CDKL5	1	33	X:18507097-18510854	1	33	33	3.90E- 17	100
	ENSP0000048 9879	CDKL5	1	33	X:18507097-18510854	1	33	33	1.69E- 15	100
	ENSP0000048 5625	CDKL5	1	17	X:18507097-18507147	1	17	17	9.02E- 05	100
<i>MAST2-- EIF2B3</i>	ENSP0000036 1078	MAST2	53	787	1:45959387-46030759	207	941	735	0	100
	ENSP0000048 5842	EIF2B3	1	98	1:44978315-44981168	1	98	98	3.92E- 61	100
<i>BRIX1-- MAP2K5</i>	ENSP0000034 6493	MAP2K 5	10	412	15:67550034-67806747	54	456	403	0	100
	ENSP0000017 8640	MAP2K 5	46	448	15:67550034-67806747	54	456	403	0	100
	ENSP0000048	MAP2K	2	175	15:67585892-67664644	117	290	174	5.54E-	100

	0580	5							125	
	ENSP0000039 0196	MAP2K 5	1	128	15:67563300-67630927	76	203	128	1.82E- 88	100

Anexo 10: Predicciones generadas por PfamScan de las secuencias proteicas traducidas a partir de los transcritos de los genes que conforman las fusiones génicas proteína quinasa. La primera columna indica la fusión génica. La segunda y tercera columna indica las predicciones de dominios pertenecientes a la secuencia proteica del gen 5' (PFAM_LEFT) y gen 3' (PFAM_RIGHT), respectivamente.

FusionName	PFAM_LEFT	PFAM_RIGHT
CDKL5-- AP1S2	FTA2 11-55 0.0023^Pkinase-PARTIAL 14-99~ 7.9e-78^PK_Tyr_Ser-Thr-PARTIAL 17-99~ 1.5e-37^RIO1-PARTIAL 43-99~ 3.7e-06	Clat_adaptor_s 1-139 3.2e-64^PulA_N1 70-115 1.7e-05
MAP3K3-- RAD51C	PB1 76-151 8.8e-16^Kinase-like 166-193 1.7^Kinase-like 390-445 0.07^Pkinase-PARTIAL 391-474~ 2.2e-71^PK_Tyr_Ser-Thr-PARTIAL 392-474~ 2.1e-44	Rad51-PARTIAL ~146-348 2.1e-47^AAA_25-PARTIAL ~146-283 6.8e-11^RecA-PARTIAL ~146-173 1.5e-06^AAA_18-PARTIAL ~146-254 1.3e-05^AAA-PARTIAL ~146-172 0.00022^AAA_24-PARTIAL ~146-180 2.6e-05^DnaB_C-PARTIAL ~146-291 0.00018^cobW-PARTIAL ~146-180 0.0046^HHH_5 169-181 6.5^RecA 254-290 0.089^AAA 257-292 0.24^AAA_33 261-291 1.5^cobW 262-293 0.02
NLK-- ACACA	Pkinase 141-427 2.5e-67^PK_Tyr_Ser-Thr 141-335 2.3e-34^Kinase-like 197-314 1.1e-05^Pkinase_fungal 245-318 3.7e-08^FTA2 245-285 1.4e-05	DUF445-PARTIAL ~925-960 0.00035
NLK--	Pkinase 141-427 2.5e-67^PK_Tyr_Ser-Thr 141-	WD40-PARTIAL ~404-423 0.0042^BCAS3 577-792 5.1e-30

BCAS3	335 2.3e-34^Kinase-like 197-314 1.1e-05^Pkinase_fungal 245-318 3.7e-08^FTA2 245-285 1.4e-05	
NLK--HLF	Pkinase 141-427 2.5e-67^PK_Tyr_Ser-Thr 141-335 2.3e-34^Kinase-like 197-314 1.1e-05^Pkinase_fungal 245-318 3.7e-08^FTA2 245-285 1.4e-05	.
PASK-- LMBR1	PAS 14-70 3.8^PAS 133-230 6.8e-10^PAS_8 133-170 5.5e-05^PAS_9 134-224 7.1e-12^PAS_2 134-183 0.00026^PAS_4 143-184 0.012^PAS_4 242-282 1.1^PAS_9 245-310 0.00067^PAS 246-280 0.05^PAS_8 248-272 9^PAS_9 348-385 8.4e-06^PAS_4 348-380 0.073^PAS 350-400 0.0034^PAS_8 365-395 0.18^PAS_2 726-747 5^Kinase-like 994-1029 3.5^Pkinase 999-1251 8.3e-60^PK_Tyr_Ser-Thr 1002-1246 7.2e-27^Kinase-like 1106-1238 7.1e-06^Pkinase_fungal 1119-1177 8.4e-06	LMBR1-PARTIAL ~67-491 1.3e-96^PrsW-protease 97-128 4.9e-05^Orf78 113-130 2.2^Sigma_reg_N 139-192 3.3e-05^Orf78 149-165 2.5^CrgA 150-182 0.00086^PrsW-protease 155-174 1.8^Orf78 315-362 0.13^CrgA 320-344 0.35^CrgA 385-403 2.9^Sigma_reg_N 393-425 0.25^Sigma_reg_N 463-523 0.068^PrsW-protease 489-506 0.036
ROCK2-- CACNA2D3	Pkinase 92-354 1.7e-65^PK_Tyr_Ser-Thr 95-317 8.8e-33^APH 154-179 0.1^APH 207-240 7.5e-06^Pkinase_C 376-415 0.0014^Rho_Binding 440-490 3^Rho_Binding 453-500 3^Rho_Binding 529-573 0.91^APH 552-552 4.5^Rho_Binding 555-	VWA_3-PARTIAL ~382-417 9.6e-17^VWA-PARTIAL ~382-428 3.7e-14^dCache_1 462-618 2.8e-11^Cache_3-Cache_2 497-600 2.8e-05^dCache_2 507-542 2.8e-06^VGCC_alpha2 654-1068 1.4e-41^dCache_1 769-835 0.087^Fer4_21 843-870 0.0042^Fer4_21 983-

	612 6.7^PK_Tyr_Ser-Thr 580-612 0.33^Rho_Binding 664-705 3^Rho_Binding 674-710 1.7^Rho_Binding 700-749 0.23^Pkinase 721-743 1.9^PK_Tyr_Ser-Thr 731-777 1.1^Rho_Binding 785-831 2.3^Rho_Binding 841-888 7^Rho_Binding 885-908 2.7^Rho_Binding 916-965 1.1e-05^PK_Tyr_Ser-Thr 932-958 1.3^Pkinase 933-959 1.7^Rho_Binding 978-1046 1.7e-27^Pkinase 1038-1084 1.1^Rho_Binding 1055-1098 0.18^Rho_Binding 1106-1125 2.9^PH 1153-1234 7e-08^C1_1 1261-1303 8.2e-08^PH 1313-1346 0.0071	1025 0.083
ROCK2-- PICALM	Pkinase-PARTIAL 92-141~ 1.7e-65^PK_Tyr_Ser-Thr-PARTIAL 95-141~ 8.8e-33	.
BRIX1-- MAP2K5	Brix 67-242 2.7e-43	PB1-PARTIAL ~28-55 3.4e-11^PK_Tyr_Ser-Thr 132-377 4.4e-42^Pkinase 134-383 4e-67^Pkinase_fungal 140-306 6.6e-09^APH 216-261 0.00011^Kinase-like 227-326 3.1e-07^Seadorna_VP7 232-281 6.1e-05
BRIX1-- MAP2K5	Brix-PARTIAL 67-159~ 2.7e-43	PB1-PARTIAL ~28-55 3.4e-11^PK_Tyr_Ser-Thr 132-377 4.4e-42^Pkinase 134-383 4e-67^Pkinase_fungal 140-306 6.6e-09^APH 216-261 0.00011^Kinase-like 227-326 3.1e-07^Seadorna_VP7 232-281 6.1e-05

BRIX1-- MAP2K5	Brix 67-242 2.7e-43	PB1-PARTIAL ~28-55 3.4e-11^PK_Tyr_Ser-Thr 132-377 4.4e-42^Pkinase 134-383 4e-67^Pkinase_fungal 140-306 6.6e-09^APH 216-261 0.00011^Kinase-like 227-326 3.1e-07^Seadorna_VP7 232-281 6.1e-05
BRIX1-- MAP2K5	Brix-PARTIAL 67-159~ 2.7e-43	PB1-PARTIAL ~28-55 3.4e-11^PK_Tyr_Ser-Thr 132-377 4.4e-42^Pkinase 134-383 4e-67^Pkinase_fungal 140-306 6.6e-09^APH 216-261 0.00011^Kinase-like 227-326 3.1e-07^Seadorna_VP7 232-281 6.1e-05
DLG2-- PAK1	L27_1 1-25 8.4e-14^Chordopox_A33R-PARTIAL 27-40~ 0.00013	Pkinase-PARTIAL ~184-419 1.8e-74^PK_Tyr_Ser-Thr-PARTIAL ~184-419 2.5e-50^Kinase-like-PARTIAL ~184-212 0.00058^APH-PARTIAL ~184-284 0.003^PBD 199-216 7.2^Kdo 199-224 1.5^YrbL-PhoP_reg 200-220 2.8^DSHCT 239-280 1e-05^Kdo 245-321 4.3e-05^YrbL-PhoP_reg 252-304 0.00018^Pkinase_fungal 267-340 6.1e-09^APH 270-313 0.00014^DUF3741 272-285 0.0021^Kinase-like 280-359 1.6e-06
EIF2B3-- MAST2	NTP_transferase 5-130 1.5e-19^NTP_transf_3 5-142 1.3e-19^CTP_transf_3 25-68 1.5e-05^CTP_transf_3 215-270 0.23	DUF1908 201-475 6e-142^Pkinase 512-785 7.2e-65^PK_Tyr_Ser-Thr 514-770 2.7e-28^APH 515-599 0.0053^Kinase-like 515-547 1.1^RIO1 580-654 0.00011^Kinase-like 624-660 3.6e-06^APH 626-661 4.8e-05^PDZ 1131-1186 9e-11^PDZ_2 1136-1172 2.1e-09^Tricorn_PDZ 1137-1186 0.00013^PDZ_6 1140-1183 4.2e-12^DUF1908 1276-

		1284 10
GOLGA4-- BRAF	.	PK_Tyr_Ser-Thr 65-319 6.9e-68^Pkinase 66-319 9.6e-62^Pkinase_fungal 160-249 7.4e-07^Kdo 163-223 2.1e-05^APH 170-206 2.3e-05^Kinase-like 171-311 4.3e-12
SMARCA5-- NEK1	DBINO 70-107 4.4e-08^DBINO 143-167 6.4^ResIII 179-339 1.6e-13^SNF2_N 193-462 2.8e-82^DEAD 197-339 6.2e-09^Helicase_C 284-332 1.4^HAND 291-318 5.9^HDA2-3 417-615 4.5e-08^Helicase_C 484-597 2e-22^SNF2_N 593-621 0.69^DEAD 593-641 1.4^HAND 743-841 3.7e-33^Myb_DNA-binding 844-882 0.0012^SLIDE 898-1011 5.4e-49^Myb_DNA-binding 946-965 0.00021^HAND 1013-1024 1.1^DBINO 1016-1038 0.8	Pkinase_fungal 711-711 5
TENM4-- ATM	Ten_N 11-168 1.5e-83^Ten_N 166-340 2.3e-74^EGF_2 566-592 1.8e-05^I-EGF_1 568-583 0.37^hEGF 570-583 0.12^EGF_2 600-623 0.00041^NAGLU_N 600-627 0.00084^I-EGF_1 600-618 2e-05^hEGF 601-618 0.26^EGF_2 632-657 1.6e-08^I-EGF_1 634-652 0.00011^hEGF 635-652 1.9e-06^EGF_2 662-689 2.6e-06^I-EGF_1 665-684 0.16^hEGF 667-	ANAPC5 211-241 0.00025^TAN 734-791 0.44^PI3_PI4_kinase 975-993 3.6^HEAT 1037-1051 0.25^HEAT 1074-1097 0.011^TAN 1722-1744 4.2^HEAT 1726-1741 0.0081^FAT 2097-2488 5.3e-49^TAN 2130-2179 0.48^PI3_PI4_kinase 2715-2961 2.7e-54^HEAT 2928-2952 5.6^FATC 3026-3055 3.1e-15

	<p>684 0.0055^NAGLU_N 677-693 0.14^EGF_2 699-724 1.2e-05^I-EGF_1 699-719 8.4e-06^hEGF 700-717 0.041^NAGLU_N 712-725 6.1^EGF_2 729-755 2.7e-08^I-EGF_1 730-750 0.00085^hEGF 733-750 0.0057^EGF_2 760-786 2.5e-06^I-EGF_1 762-781 0.041^hEGF 764-781 3.2e-05^EGF_2 800-830 0.00081^hEGF 804-825 0.0033^I-EGF_1 804-825 3.1^CarboxypepD_reg 924-969 2.2e-05^NHL 1238-1249 0.38^NHL 1347-1362 0.51^NHL 1408-1440 0.0021^NHL 1472-1486 2.2^EGF_2 1515-1526 3.8^hEGF 1521-1523 9^NHL 1536-1560 3.7e-06^CarboxypepD_reg 1599-1620 4.3^RHS_repeat 1612-1629 0.3^RHS_repeat 1680-1712 2.6e-07^CarboxypepD_reg 1723-1749 2.6^RHS_repeat 1907-1928 0.00014^RHS_repeat 1947-1957 5.7^RHS_repeat 2021-2036 1.2^RHS_repeat 2083-2093 9^RHS_repeat 2169-2178 0.6^RHS_repeat 2187-2205 0.00089^RHS_repeat 2196-2223 2.1e-</p>	
--	---	--

	05^RHS_repeat 2217-2253 0.13^RHS_repeat 2271-2301 1e-06^Tox-GHH 2685-2762 3.5e-33	
TENM4-- ATM	Ten_N 11-168 1.5e-83^Ten_N 166-340 2.3e-74^EGF_2 566-592 1.8e-05^I-EGF_1 568-583 0.37^hEGF 570-583 0.12^EGF_2 600-623 0.00041^NAGLU_N 600-627 0.00084^I-EGF_1 600-618 2e-05^hEGF 601-618 0.26^EGF_2 632-657 1.6e-08^I-EGF_1 634-652 0.00011^hEGF 635-652 1.9e-06^EGF_2 662-689 2.6e-06^I-EGF_1 665-684 0.16^hEGF 667-684 0.0055^NAGLU_N 677-693 0.14^EGF_2 699-724 1.2e-05^I-EGF_1 699-719 8.4e-06^hEGF 700-717 0.041^NAGLU_N 712-725 6.1^EGF_2 729-755 2.7e-08^I-EGF_1 730-750 0.00085^hEGF 733-750 0.0057^EGF_2 760-786 2.5e-06^I-EGF_1 762-781 0.041^hEGF 764-781 3.2e-05^EGF_2 800-830 0.00081^hEGF 804-825 0.0033^I-EGF_1 804-825 3.1^CarboxypepD_reg 924-969 2.2e-05^NHL 1238-1249 0.38^NHL 1347-1362 0.51^NHL 1408-1440 0.0021^NHL 1472-1486 2.2^EGF_2 1515-1526 3.8^hEGF 1521-1523 9^NHL 1536-1560 3.7e-06^CarboxypepD_reg 1599-	ANAPC5 211-241 0.00025^TAN 734-791 0.44^PI3_PI4_kinase 975-993 3.6^HEAT 1037-1051 0.25^HEAT 1074-1097 0.011^TAN 1722-1744 4.2^HEAT 1726-1741 0.0081^FAT 2097-2488 5.3e-49^TAN 2130-2179 0.48^PI3_PI4_kinase 2715-2961 2.7e-54^HEAT 2928-2952 5.6^FATC 3026-3055 3.1e-15

	<p>1620 4.3^RHS_repeat 1612-1629 0.3^RHS_repeat 1680-1712 2.6e-07^CarboxypepD_reg 1723-1749 2.6^RHS_repeat 1907-1928 0.00014^RHS_repeat 1947-1957 5.7^RHS_repeat 2021-2036 1.2^RHS_repeat 2083-2093 9^RHS_repeat 2169-2178 0.6^RHS_repeat 2187-2205 0.00089^RHS_repeat 2196-2223 2.1e-05^RHS_repeat 2217-2253 0.13^RHS_repeat 2271-2301 1e-06^Tox-GHH 2685-2762 3.5e-33</p>	
TENM4-- ATM	<p>Ten_N 11-168 1.5e-83^Ten_N 166-340 2.3e-74^EGF_2 566-592 1.8e-05^I-EGF_1 568-583 0.37^hEGF 570-583 0.12^EGF_2 600-623 0.00041^NAGLU_N 600-627 0.00084^I-EGF_1 600-618 2e-05^hEGF 601-618 0.26^EGF_2 632-657 1.6e-08^I-EGF_1 634-652 0.00011^hEGF 635-652 1.9e-06^EGF_2 662-689 2.6e-06^I-EGF_1 665-684 0.16^hEGF 667-684 0.0055^NAGLU_N 677-693 0.14^EGF_2 699-724 1.2e-05^I-EGF_1 699-719 8.4e-06^hEGF 700-717 0.041^NAGLU_N 712-725 6.1^EGF_2 729-</p>	<p>TAN 734-791 0.44^PI3_PI4_kinase 975-993 3.6^HEAT 1037-1051 0.25^HEAT 1074-1097 0.011^TAN 1722-1744 4.2^HEAT 1726-1741 0.0081^FAT 2097-2488 5.3e-49^TAN 2130-2179 0.48^PI3_PI4_kinase 2715-2961 2.7e-54^HEAT 2928-2952 5.6^FATC 3026-3055 3.1e-15</p>

	755 2.7e-08^I-EGF_1 730-750 0.00085^hEGF 733-750 0.0057^EGF_2 760-786 2.5e-06^I-EGF_1 762-781 0.041^hEGF 764-781 3.2e-05^EGF_2 800-830 0.00081^hEGF 804-825 0.0033^I-EGF_1 804-825 3.1^CarboxypepD_reg 924-969 2.2e-05^NHL 1238-1249 0.38^NHL 1347-1362 0.51^NHL 1408-1440 0.0021^NHL 1472-1486 2.2^EGF_2 1515-1526 3.8^hEGF 1521-1523 9^NHL 1536-1560 3.7e-06^CarboxypepD_reg 1599-1620 4.3^RHS_repeat 1612-1629 0.3^RHS_repeat 1680-1712 2.6e-07^CarboxypepD_reg 1723-1749 2.6^RHS_repeat 1907-1928 0.00014^RHS_repeat 1947-1957 5.7^RHS_repeat 2021-2036 1.2^RHS_repeat 2083-2093 9^RHS_repeat 2169-2178 0.6^RHS_repeat 2187-2205 0.00089^RHS_repeat 2196-2223 2.1e-05^RHS_repeat 2217-2253 0.13^RHS_repeat 2271-2301 1e-06^Tox-GHH 2685-2762 3.5e-33	
TENM4--	Ten_N 11-168 1.5e-83^Ten_N 166-340 2.3e-	TAN 734-791 0.44^PI3_PI4_kinase 975-

ATM	<p>74^EGF_2 566-592 1.8e-05^I-EGF_1 568-583 0.37^hEGF 570-583 0.12^EGF_2 600-623 0.00041^NAGLU_N 600-627 0.00084^I-EGF_1 600-618 2e-05^hEGF 601-618 0.26^EGF_2 632-657 1.6e-08^I-EGF_1 634-652 0.00011^hEGF 635-652 1.9e-06^EGF_2 662-689 2.6e-06^I-EGF_1 665-684 0.16^hEGF 667-684 0.0055^NAGLU_N 677-693 0.14^EGF_2 699-724 1.2e-05^I-EGF_1 699-719 8.4e-06^hEGF 700-717 0.041^NAGLU_N 712-725 6.1^EGF_2 729-755 2.7e-08^I-EGF_1 730-750 0.00085^hEGF 733-750 0.0057^EGF_2 760-786 2.5e-06^I-EGF_1 762-781 0.041^hEGF 764-781 3.2e-05^EGF_2 800-830 0.00081^hEGF 804-825 0.0033^I-EGF_1 804-825 3.1^CarboxypepD_reg 924-969 2.2e-05^NHL 1238-1249 0.38^NHL 1347-1362 0.51^NHL 1408-1440 0.0021^NHL 1472-1486 2.2^EGF_2 1515-1526 3.8^hEGF 1521-1523 9^NHL 1536-1560 3.7e-06^CarboxypepD_reg 1599-1620 4.3^RHS_repeat 1612-1629 0.3^RHS_repeat 1680-1712 2.6e-07^CarboxypepD_reg 1723-</p>	<p>993 3.6^HEAT 1037-1051 0.25^HEAT 1074-1097 0.011^TAN 1722-1744 4.2^HEAT 1726-1741 0.0081^FAT 2097-2488 5.3e-49^TAN 2130-2179 0.48^PI3_PI4_kinase 2715-2961 2.7e-54^HEAT 2928-2952 5.6^FATC 3026-3055 3.1e-15</p>
-----	---	---

	<p>1749 2.6^RHS_repeat 1907-1928 0.00014^RHS_repeat 1947-1957 5.7^RHS_repeat 2021-2036 1.2^RHS_repeat 2083-2093 9^RHS_repeat 2169-2178 0.6^RHS_repeat 2187-2205 0.00089^RHS_repeat 2196-2223 2.1e-05^RHS_repeat 2217-2253 0.13^RHS_repeat 2271-2301 1e-06^Tox-GHH 2685-2762 3.5e-33</p>	
TENM4-- ATM	<p>Ten_N 11-168 1.5e-83^Ten_N 166-340 2.3e-74^EGF_2 566-592 1.8e-05^EGF_1 568-583 0.37^hEGF 570-583 0.12^EGF_2 600-623 0.00041^NAGLU_N 600-627 0.00084^EGF_1 600-618 2e-05^hEGF 601-618 0.26^EGF_2 632-657 1.6e-08^EGF_1 634-652 0.00011^hEGF 635-652 1.9e-06^EGF_2 662-689 2.6e-06^EGF_1 665-684 0.16^hEGF 667-684 0.0055^NAGLU_N 677-693 0.14^EGF_2 699-724 1.2e-05^EGF_1 699-719 8.4e-06^hEGF 700-717 0.041^NAGLU_N 712-725 6.1^EGF_2 729-755 2.7e-08^EGF_1 730-750 0.00085^hEGF 733-750 0.0057^EGF_2 760-786 2.5e-06^EGF_1 762-781 0.041^hEGF 764-781 3.2e-05^EGF_2 800-</p>	<p>TAN 734-791 0.44^PI3_PI4_kinase 975-993 3.6^HEAT 1037-1051 0.25^HEAT 1074-1097 0.011^TAN 1722-1744 4.2^HEAT 1726-1741 0.0081^FAT 2097-2488 5.3e-49^TAN 2130-2179 0.48^PI3_PI4_kinase 2715-2961 2.7e-54^HEAT 2928-2952 5.6^FATC 3026-3055 3.1e-15</p>

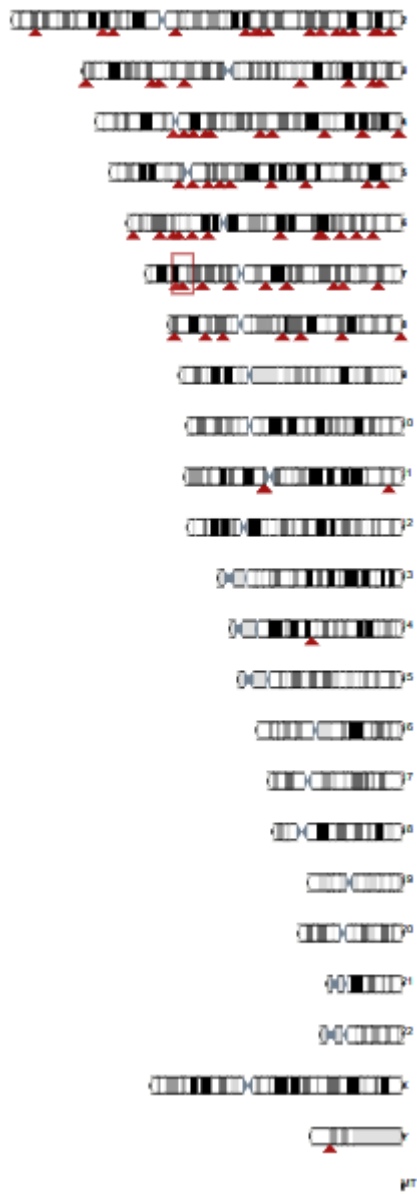
	<p>830 0.00081^hEGF 804-825 0.0033^I-EGF_1 804-825 3.1^CarboxypepD_reg 924-969 2.2e-05^NHL 1238-1249 0.38^NHL 1347-1362 0.51^NHL 1408-1440 0.0021^NHL 1472-1486 2.2^EGF_2 1515-1526 3.8^hEGF 1521-1523 9^NHL 1536-1560 3.7e-06^CarboxypepD_reg 1599-1620 4.3^RHS_repeat 1612-1629 0.3^RHS_repeat 1680-1712 2.6e-07^CarboxypepD_reg 1723-1749 2.6^RHS_repeat 1907-1928 0.00014^RHS_repeat 1947-1957 5.7^RHS_repeat 2021-2036 1.2^RHS_repeat 2083-2093 9^RHS_repeat 2169-2178 0.6^RHS_repeat 2187-2205 0.00089^RHS_repeat 2196-2223 2.1e-05^RHS_repeat 2217-2253 0.13^RHS_repeat 2271-2301 1e-06^Tox-GHH 2685-2762 3.5e-33</p>	
--	--	--

Anexo 11: Predicciones de dominios proteicos por Prosite a partir de los transcritos traducidos dentro del marco de lectura. Se indica todos los dominios predichos dentro de la secuencia aminoacídica de la fusión génica correspondiente.

Fusión génica	ID_ProSITE	Nombre_ProSITE	Puntaje de dominio	Inicio	Final
CDKL5-- AP1S2	PS00989	CLAT_ADAPTOR_S	.	89	99
NLK-- ACACA	PS50011	PROTEIN_KINASE_DOM	39.283	138	462
	PS00107	PROTEIN_KINASE_ATP	.	144	167
	PS01351	MAPK	.	173	276
	PS00108	PROTEIN_KINASE_ST	.	260	272
	PS50980	COA_CT_NTER	40.874	385	641
	PS50989	COA_CT_CTER	62.042	645	961
NLK--HLF	PS50011	PROTEIN_KINASE_DOM	24.174	138	350
	PS00107	PROTEIN_KINASE_ATP	.	144	167
	PS01351	MAPK	.	173	276
	PS00108	PROTEIN_KINASE_ST	.	260	272
	PS50217	BZIP	11.691	280	343
PASK-- LMBR1	PS50112	PAS	12.164	134	190
ROCK2-- CACNA2D3	PS50011	PROTEIN_KINASE_DOM	45.94	92	354
	PS00107	PROTEIN_KINASE_ATP	.	98	121
	PS00108	PROTEIN_KINASE_ST	.	210	222
	PS51285	AGC_KINASE_CTER	11.931	357	425
	PS51860	REM_1	22.734	497	573
	PS50234	VWFA	15.272	917	1099
BRX1-- MAP2K5	PS51745	PB1	13.413	13	117
	PS50011	PROTEIN_KINASE_DOM	42.591	174	427
	PS00107	PROTEIN_KINASE_ATP	.	180	203
	PS00108	PROTEIN_KINASE_ST	.	287	299
EIF2B3-- MAST2	PS50011	PROTEIN_KINASE_DOM	46.534	551	824
	PS00108	PROTEIN_KINASE_ST	.	670	682

	PS51285	AGC_KINASE_CTER	9.003	825	893
	PS50106	PDZ	14.576	1143	1231
GOLGA4-- BRAF	PS50011	PROTEIN_KINASE_DOM	45.205	358	618
	PS00107	PROTEIN_KINASE_ATP	.	364	384
	PS00108	PROTEIN_KINASE_ST	.	473	485
SMARCA5-- NEK1	PS51192	HELICASE_ATP_BIND_1	26.548	192	357
	PS51194	HELICASE_CTER	20.443	487	638

Anexo 12: Alineamiento de lecturas de soporte de TPRX1--GNPBPM en el genoma humano en Ensembl.



Anexo 13: Alineamiento de lecturas de soporte de RP1-34H18.1--NAV3 en el genoma humano en Ensembl.

