



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Programa de Maestría y Doctorado en Ciencias Médicas, Odontológicas y de la Salud
División de Estudios de Posgrado e Investigación
Laboratorio de Genómica Clínica

**Estudio piloto para la detección de mutaciones
germinales a partir de muestras de saliva de
pacientes con cáncer pediátrico**

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Doctor en Ciencias

PRESENTA:

Oscar Alonso Luna

TUTORA:

Dra. Gabriela Elisa Mercado Celis
Laboratorio de Genómica Clínica, DEPeI-UNAM

MIEMBROS DEL COMITÉ TUTOR

Dr. Jorge Meléndez Zajgla
Instituto Nacional de Medicina Genómica

Dra. Socorro Aida Borges Yañez
Departamento de Salud Pública Bucal, DEPeI-UNAM

Ciudad Universitaria, CDMX, ENERO 2024



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

CONTENIDO

1.- Antecedentes.....	5
1.1 Epidemiología del cáncer en niños y adolescentes.....	5
1.2 Bases moleculares del cáncer pediátrico vs adultos.....	6
1.3 Predisposición genética en cáncer pediátrico.....	8
1.4 Genes de predisposición a cáncer pediátrico y tumores asociados.....	10
1.5 Factores de riesgo en cáncer pediátrico	12
1.5.1 Interacción gen-ambiente (GxE)	13
1.6 Estrategias genómicas en cáncer pediátrico.....	14
1.7 Nuevos abordajes del cáncer pediátrico.....	17
2.- Planteamiento del problema	18
3.- Justificación	19
4.- Objetivos.....	19
5.- Hipótesis	19
6.- Metodología	20
Saliva.....	21
Sangre.....	21
Controles de calidad	21
7.-Resultados	23
9.- Discusión.....	59
10.- Conclusiones.....	61
11.- Referencias	61
12.- Anexos	68

Tablas

TABLA 1. CRITERIOS DE SOSPECHA DE SÍNDROMES DE PREDISPOSICIÓN A CÁNCER EN NIÑOS.....	9
TABLA 2. TUMORES PEDIÁTRICOS ASOCIADOS CON SÍNDROMES DE PREDISPOSICIÓN.....	11
TABLA 3. DISEÑOS PARA EL ESTUDIO DE INTERACCIONES GxE.....	14
TABLA 4. ESTRATEGIAS DE SECUENCIACIÓN EN CÁNCER PEDIÁTRICO.....	14
TABLA 5. APLICACIONES DE LOS ESTUDIOS DE SECUENCIACIÓN EN TRÍOS.....	15
TABLA 6. ESTADÍSTICA DESCRIPTIVA DE LOS PACIENTES SECUENCIADOS.....	25
TABLA 7. HISTORIA FAMILIAR DE CANCER (N=24).....	28
TABLA 8. CRITERIOS DE SOSPECHA A CÁNCER HEREDITARIO (EUR J MED GENET . 2016 MAR;59(3):116-25).....	29
TABLA 9. CONCENTRACIONES DE ADNG OBTENIDO A PARTIR DE 3 DIFERENTES TOMAS DE MUESTRA Y ANÁLISIS DE VARIANZA.....	31
TABLA 10. FRACCIÓN DE BASES CUBIERTAS A 20X. LOS DATOS SE PRESENTAN EN ORDEN DESCENDENTE.....	34
TABLA 11. FRACCIÓN DE BASES CUBIERTAS A 20X DETERMINADA A TRAVÉS DEL NUEVO ARCHIVO DE INTERVALOS.....	35
TABLA 12. LISTA DE VARIANTES CON ALTO ÍNDICE DE PATOGENIDAD IDENTIFICADAS EN LOS INDIVIDUOS SECUENCIADOS CON SEGÚN SU CLASIFICACIÓN EN CLINVAR.....	41
TABLA 13. LISTA DE VARIANTES FILTRADAS MEDIANTE UNA SEGUNDA ESTRATEGIA DE FILTRADO DE 5 NIVELES.....	44
TABLA 14. LISTA COMPLETA DE VARIANTES FILTRADAS MEDIANTE LA TERCERA ESTRATEGIA Y CLASIFICACIÓN.....	49
TABLA 15. RELACIÓN DE CASOS ÍNDICE PORTADORES DE MUTACIONES GERMINALES O VARIANTES CANDIDATAS CON DATOS DE SECUENCIACIÓN EN TEJIDO.....	54
TABLA 16. ESTIMACIONES DE PARENTESCO ENTRE LOS INDIVIDUOS SECUENCIADOS.....	56
TABLA 17. PACIENTES PORTADORES DE MUTACIONES GERMINALES E INFORMACIÓN CLÍNICA.....	58
TABLA 18. VARIANTES CANDIDATAS Y PUNTAJES DE PREDICCIÓN.....	59

Figuras

FIGURA 1. DISTRIBUCIÓN GLOBAL DEL CÁNCER EN NIÑOS Y ADOLESCENTES POR REGIÓN GEOGRÁFICA.(HTTP://DATA.WORLDBANCK.ORG/COUNTRY	5
FIGURA 2. MUTACIONES SOMÁTICAS DEL ESTUDIO PAN-CÁNCER PEDIÁTRICO VS TCGA DE CÁNCER EN ADULTOS. HTTPS://PECAN.STJUDE.CLOUD/PROTEINPAINT/STUDY/PAN-TARGET	7
FIGURA 3. FLUJO DE TRABAJO DE LAS BUENAS PRÁCTICAS DEL GENOME ANALYSIS TOOLKIT, GATK.....	22
FIGURA 4. PACIENTES RECLUTADOS, EN PROCESO DE SECUENCIAR Y EN PROCESO DE ANÁLISIS. LAS CARACTERÍSTICAS CLÍNICAS Y DEMOGRÁFICAS SE PRESENTAN EN LA ESTADÍSTICA DESCRIPTIVA.....	23
FIGURA 5. FRECUENCIA Y DISTRIBUCIÓN DE LOS 277 CASOS ÍNDICE RECLUTADOS.....	24
FIGURA 6. GRÁFICA DE KAPLAN-MEIER DE SUPERVIVENCIA GLOBAL DEL GRUPO DE PACIENTES SECUENCIADOS.....	27
FIGURA 7. CÁLCULO DE VALORES DE PREDICCIÓN, SENSIBILIDAD Y ESPECIFICIDAD.....	30
FIGURA 8. ANÁLISIS CUALITATIVO DE ADNG. LOS PANELES A Y B MUESTRAN LAS BANDAS DE ADNG PURIFICADO A PARTIR DE TRES DIFERENTES TIPOS DE TOMA DE MUESTRA. A PARTIR DE 50 NG TOTALES DE ADNG TOTAL, SE PUEDE VISUALIZAR LA BANDA DE MÁS DE 10K PB. (F, TUBO FALCON; O, ORAGENE; S, STRATEC).....	33
FIGURA 9. NÚMERO DE GENES POR PANEL, INTERSECCIÓN Y DISYUNCIÓN. EN EL PANEL B SE ENLISTAN ALGUNOS GENES REPRESENTATIVOS.....	35
FIGURA 10. HISTOGRAMAS DE COBERTURA. EL PANEL A MUESTRA EL PRIMER HISTOGRAMA DE LA MUESTRA NA12891.EXOME.BAM OBTENIDO CON LOS COMANDOS USADOS CON LOS DATOS EN PROCESO DE ANÁLISIS, MOSTRANDO UN PATRÓN SIMILAR NO ESPERADO. POSTERIOR AL ORDENADO POR CROMOSOMA, SE MEJORA EL CONSUMO DE MEMORIA Y SE LOGRA EL HISTOGRAMA ESPERADO (PANEL B).....	37
FIGURA 11. MÉTRICAS DE SECUENCIACIÓN DE MUESTRAS DE SALIVA. SE REALIZARON GRÁFICAS DE CAJAS Y BIGOTES PARA VISUALIZAR EL DESEMPEÑO DE SECUENCIACIÓN A PARTIR DE ADNG PURIFICADO DE MUESTRAS DE SALIVA. SE OBSERVA QUE LA FRACCIÓN DE LECTURAS ALINEADAS AL GENOMA DE REFERENCIA HG19 (A) ES MAYOR AL 95%, MIENTRAS QUE LA FRACCIÓN DE BASES A 20X (B) ES MAYOR AL 77%.....	38
FIGURA 12. GRÁFICOS DE AGRUPACIÓN DE VARIANTES Y TRAMOS. EN EL PANEL A Y B SE MUESTRAN LOS GRÁFICOS OBTENIDOS EN EL PRIMER INTENTO, EN LOS CUALES, SE OBSERVA UN PREDOMINIO DE VARIANTES CON BAJOS PUNTAJES DE CALIDAD Y ALTA PROBABILIDAD DE FALSOS POSITIVOS. TRAS EL REPROCESO DEL GENOTIPADO CON EL ARCHIVO DE INTERVALOS ADECUADO, SE OBSERVA UNA AGRUPACIÓN DE VARIANTES DE ALTA CALIDAD EN LOS RANGOS DE LOS FILTROS DE ANOTACIÓN.....	

CORRESPONDIENTES (PANEL C), ADEMÁS, EL GRÁFICO DE TRAMOS YA SE VISUALIZAN VARIANTES VERDADERAS POSITIVAS (PANEL D)	39
FIGURA 13. ESTADISTICA DE VARIANTES PROCESADAS EN LA POBLACIÓN DE ESTUDIO. POSTERIOR A LA ANOTACIÓN DE CADA VCF INDIVIDUAL, SE DETERMINÓ LA CANTIDAD Y PORCENTAJE DE LAS VARIANTES PROCESADAS PARA SUBSECUENTEMENTE, OBTENER UNA ESTADISTICA GLOBAL (PANEL A Y B). ADEMÁS, SE DETERMINÓ EL TIPO Y PORCENTAJE DE TODAS LAS VARIANTES EN REGIONES CODIFICANTES (PANEL C Y D). LOS ANTERIORES RESULTADOS DEMUESTRAN UNA MAYOR ABUNDANCIA DE VARIANTES SINÓNIMAS Y CON MENOR FRECUENCIA LAS CODÓN DE PARO PREMATURO Y DELECCIONES INSERCIONES EN MARCO; ESTAS ÚLTIMAS ASOCIADAS CON EFECTO DELETÉREO.	40
FIGURA 14. ESTRATEGIA DE FILTRADO PARA LA IDENTIFICACIÓN DE MÁS VARIANTES CON ALTO ÍNDICE DE PATOGENICIDAD EN LOS CASOS ÍNDICE (PANEL A). EL PORCENTAJE PROMEDIO DE VARIANTES RETENIDAS Y ELIMINADAS EN CADA FILTRO, SE MUESTRA EN EL PANEL B.	43
FIGURA 15. ESTRATEGIA DE FILTRADO PARA BÚSQUEDA DE MUTACIONES PATOGENICAS O VARIANTES CANDIDATAS. MEDIANTE LA UNIFICACIÓN DE 2 LISTAS DE GENES DE PREDISPOSICIÓN A CÁNCER (PANEL A), SE OBTUVO UNA LISTA FINAL DE GENES (PANEL B) PARA ENFOCAR EL FILTRADO DE VARIANTES	48
ILUSTRACIÓN 16. TERCERA ESTRATEGIA DE FILTRADO PARA LA IDENTIFICACIÓN DE VARIANTES CON ALTO ÍNDICE DE PATOGENICIDAD EN LOS CASOS ÍNDICE (PANEL A). EL PORCENTAJE PROMEDIO DE VARIANTES RETENIDAS Y ELIMINADAS EN CADA FILTRO, SE MUESTRA EN EL PANEL B.	49
FIGURA 17. VERIFICACIÓN DE VARIANTES EN LOS GENES NF1 Y MSH6 EN EL ARCHIVO BAM DEL CASO ÍNDICE CPA_L002. EL PANEL A MUESTRA LA IMAGEN DE LAS LECTURAS ALINEADAS EN LA REGIÓN 66,459,248 – 66,459,293 DEL GEN NF1, DONDE SE OBSERVA LA VARIANTE HETEROCIGOTA (NF1):c.2953C>T (p.GLN985TER). EL MISMO INDIVIDUO FUE PORTADOR DE LA COVARIANTE HETEROCIGOTA (MSH6):c.114DEL (p.ALA40FS) EN EL GEN MSH6 (PANEL B).	53
FIGURA 18. CONFIRMACIÓN DE LA VARIANTE DICER1::c.5017DUP (p.ILE1673FS) EN LA CASO ÍNDICE 521010310090. LA REVISIÓN DE LOS ALINEAMIENTOS SOBRE LOS ARCHIVOS BAM RECALIBRADOS PARA SANGRE PERIFÉRICA () Y TEJIDO () DEMOSTRARON UNA FRECUENCIA ALÉLICA DE LA VARIANTE (VAF) DE 0.42 Y 0.99, RESPECTIVAMENTE.	55
FIGURA 19. GRÁFICO DE ANCESTRÍA. EL COMPONENTE MÁS DESTACADO QUE SE IDENTIFICÓ CORRESPONDE A MEZCLA DE NATIVOS AMERICANOS, EN EL CUAL SE INCLUYEN GRUPOS ÉTNICOS COMO HUICHOL, MAYA, MIXTECO Y KARITIANA.....	57

1.- Antecedentes

1.1 Epidemiología del cáncer en niños y adolescentes

El cáncer es la primera causa de muerte por enfermedad en niños y adolescentes a nivel mundial. Las últimas estimaciones de la *International Agency of Research of Cancer (IACR)*, obtenidas a través de más de 153 registros de cáncer en 62 países, demostraron que la incidencia mundial del cáncer en la niñez es mayor a la evaluada previamente. La incidencia en niños de 0 a 14 años fue de 140.6 por millón de personas/año, y en los de 15 a 19 años fue de 185.3 por millón de personas/año, mostrando un aumento del 0.5 al 1% anual en países de altos ingresos (1, 2). Sin embargo, la falta de registros de cáncer en países con ingresos medios y bajos, limitan las estimaciones globales reales (Figura 1).

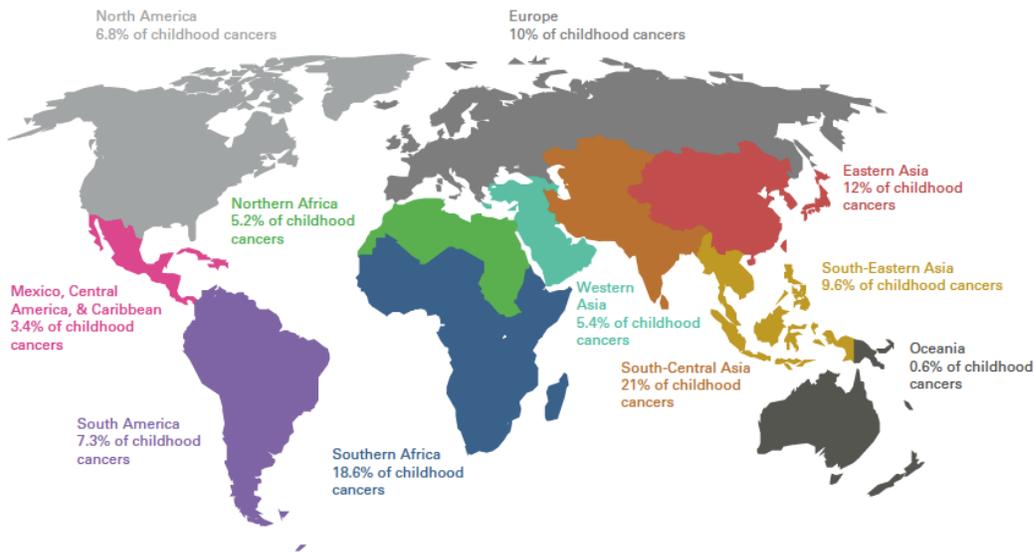


Figura 1. Distribución global del cáncer en niños y adolescentes por región geográfica. <http://data.worldbank.org/country>

También se han reportado diferencias en las tasas de respuestas entre poblaciones. Los niños y adolescentes que viven en países con altos ingresos presentan mejores tasas de supervivencia, donde la supervivencia a 5 años alcanza > 80%, a diferencia de países con ingresos medios o bajos donde el panorama es menos favorable, alcanzando supervivencias a 5 años entre el 5 y 60% (3). En México, las últimas estadísticas del Registro Nacional del Cáncer demostraron una tasa de incidencia similar a los países de altos ingresos, sin embargo, la supervivencia global a 5 años permanece baja (67%) (4, 5). Además, se han reportado disparidades raciales/étnicas en las tasas de supervivencia asociadas principalmente a factores socio-demográficos como deficiencia en los programas de atención primaria, falta de acceso a sistemas de salud, infraestructura sanitaria débil, personal de salud escaso o poco capacitado, diagnósticos tardíos o erróneos, nivel socio-

cultural y abandono de tratamientos (6, 7). Datos del *Surveillance Epidemiology and End Results (SEER)* demostraron diferencias de tasas de supervivencia en Estados Unidos entre niños blancos y los niños de etnias latina y afroamericana, principalmente para Leucemia Linfoblástica Aguda (LLA), tumores del Sistema Nervioso Central (astrocitoma), y osteosarcoma, por lo que características genéticas propias de cada población también pueden tener un papel importante en dichas diferencias de supervivencia, aunque hasta el momento no se han realizado estudios genéticos poblacionales que confirmen su papel en dichas disparidades (8, 9).

En el 2017, el *St Jude Children's Research Hospital* y el *Institute for Health Metrics and Evaluation* de la Universidad de Washington, unieron esfuerzos para mejorar las estimaciones globales de la carga de cáncer en niños y adolescentes. El estudio reiteró cómo el cáncer pediátrico afecta de manera desproporcional a las poblaciones de bajos ingresos, donde se observó una mayor tasa de mortalidad (10). Un dato muy interesante e inesperado fue la asociación entre las tasas de incidencia y los índices sociodemográficos como edad, género y región geográfica, debido a que hay pocos factores de riesgo ambiental establecidos para la mayoría de los tumores pediátricos y la evidencia actual que sugiere que mutaciones germinales en genes de predisposición a cáncer afecta a menos del 10% de la población con cáncer pediátrico (11), sin embargo, las variaciones étnicas y regionales en la incidencia de los diferentes tipos de tumores en niños y adolescentes, sugieren interacciones únicas entre factores ambientales y genéticos que pueden proporcionar oportunidades para la investigación de la etiología de la enfermedad (12).

1.2 Bases moleculares del cáncer pediátrico vs adultos

Aunque los tumores que ocurren en la infancia y adolescencia presentan características comunes, sin embargo, las causas, el diagnóstico, tratamiento y pronóstico difieren considerablemente de los que ocurren en edades más avanzadas (13).

Se hipotetiza que los tumores pediátricos surgen como "errores" durante el crecimiento, ya que, los tejidos aún en desarrollo se encuentran bajo una expansión sustancial durante la formación, crecimiento y maduración temprana de los órganos, por lo que un cambio o mutación esporádica en una célula totipotencial será transmitida en las demás células especializadas, lo que suele ser la causa del cáncer infantil (14). Mientras que en adultos los tumores más frecuentes son los carcinomas de pulmón, mama y colon (15), en niños, los tumores más frecuentes son las leucemias, linfomas y tumores del Sistema Nervioso Central (SNC), mientras que en adolescentes, son más comunes los linfomas, melanomas, tumores de células germinales y algunos carcinomas como el de tiroides (16), por lo que se ha establecido que para los niños, la clasificación de los tumores debe basarse en la morfología en lugar de el sitio primario de origen como en los adultos (17).

En adultos, entre el 90 y el 95% de los tumores son causados por la acumulación de mutaciones genéticas a través del tiempo, a partir de la exposición de factores ambientales

y hábitos nocivos como tabaquismo, alcoholismo, obesidad, entre otros. En contraparte, en los tumores pediátricos el tiempo de exposición a factores ambientales es muy corto para la edad al diagnóstico (muy temprana) y presentan muy pocas mutaciones (18).

Se ha demostrado que los tumores pediátricos, muestran 14 veces menos mutaciones somáticas que los tumores en adultos (0.13 vs 1.8 mutaciones por Mb); sin embargo, existen tumores pediátricos clasificados como “altamente mutados” (2 -10 mutaciones por Mb) como los gliomas de alto grado con mutaciones germinales monoalélicas en los genes PMS2, MSH2 y MSH6, tumores “hipermutados” (más de 10 mutaciones por Mb), como los gliomas de alto grado con mutaciones germinales bialélicas en los genes MSH6 y PMS2 (Figura 2) (19, 20).

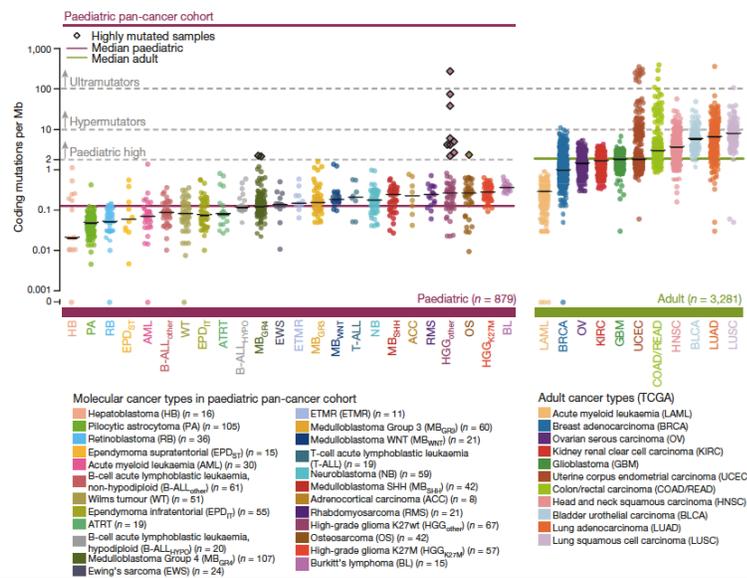


Figura 2. Mutaciones somáticas del estudio PAN-Cáncer pediátrico vs TCGA de cáncer en adultos. <https://pecan.stjude.cloud/proteinpaint/study/pan-target>

Debido a la edad temprana al diagnóstico, corto tiempo de exposición a factores ambientales, riesgo elevado a cáncer en familiares de primer grado de niños con cáncer y evidencia sobre la base genética de las disparidades poblacionales de incidencia y mortalidad del cáncer, se ha propuesto que el cáncer pediátrico tiene un componente hereditario causal mayor en comparación con el cáncer de adultos (21). Lo anterior, puede explicarse por la presencia de mutaciones en genes raros y/o alelos en genes específicos que confieren un mayor riesgo. Además, dichas diferencias étnicas pueden derivarse también por el efecto de mutaciones fundadoras, las cuales, se originan a partir de una comunidad pequeña y una vez que dicha población se expande, la mutación se vuelve más prevalente en una mayor proporción de individuos. Algunos ejemplos de efectos fundadores reportados incluyen a los Judíos Ashkenazi con mutaciones en el gen APC (22), Finlandeses con mutaciones en hMLH1 (23), Alemanes con mutaciones en el gen VHL (24) y Mexicanos con mutaciones en BRCA1 (25).

En cáncer pediátrico no existen estrategias de prevención y/o reducción de riesgo, y debido a que los tumores son silenciosos al inicio, el diagnóstico se realiza en etapas más avanzadas que en adultos; en contraparte, las tasas de respuesta en niños con cáncer vs adultos son mejores. Sin embargo, los efectos secundarios no deseados son mayores en niños (cardiopatías, disfunción cognitiva, endocrinopatías, etc.), y la posibilidad de desarrollar segundas neoplasias aumenta debido a que aún se encuentran en periodo de crecimiento (26).

Por tal motivo, la investigación de los factores genéticos y ambientales, así como su interacción en etapas preconcepción, gestación y postnatal, es de suma importancia para entender la biología tumoral y causas del cáncer pediátrico.

1.3 Predisposición genética en cáncer pediátrico

El riesgo a cáncer es directamente proporcional a la edad de las personas y, por ende, poco frecuente en niños. Si el cáncer se diagnostica en edades muy cortas, puede existir un aumento en la susceptibilidad para desarrollar cáncer cuya causa más frecuente es cambio genético (mutación) que puede ser heredado (27).

El cáncer que se origina a partir de mutaciones germinales presenta características clínicas diferentes a los tumores que se originan de mutaciones somáticas, principalmente: diagnóstico a una edad más temprana (<45 años), tumores sincrónicos o bilaterales y existencia de tumores relacionados entre miembros de la familia (cáncer de mama, colon, osteosarcomas, etc.), resultando en Síndrome de predisposición a cáncer, donde el tipo de tumor(es) depende del gen mutado (28, 29).

En adultos, se han establecido algunos criterios de sospecha a cáncer hereditario que permiten identificar a individuos en riesgo, confirmación a través de una prueba genética y establecimiento de plan de reducción de riesgo; sin embargo, en niños ha sido complicado utilizar dichos criterios debido que la mayoría de los síndromes genéticos carecen de historia familiar de cáncer, familias más pequeñas (menos individuos) y las mutaciones en genes de predisposición no necesariamente resultan en un fenotipo clínico reconocible (por ejemplo, síndromes autosómicos recesivos). Marjolijn C.J. Jongmans y cols desarrollaron una herramienta de selección para identificar niños con cáncer con riesgo elevado de predisposición genética.(30) A través de una revisión sistemática, los autores definieron 5 criterios principales de sospecha a síndromes de predisposición específicos para niños, los cuales se enlistan la siguiente tabla 1:

Tabla 1. Criterios de sospecha de síndromes de predisposición a cáncer en niños

1. Historia familiar de cáncer en niños		
<ul style="list-style-type: none">• ≥ 2 tumores en la infancia• Un familiar de primer grado con cáncer <45 años• ≥ 2 familiares de segundo grado con cáncer <45 años del mismo lado familiar• Consanguinidad entre padres		
2. Una persona con diagnóstico de cáncer en la infancia		
<ul style="list-style-type: none">• Carcinoma adrenocortical• Tumor rabdoide teratoide atípico• Gangliocitoma cerebral• Carcinoma del plexo coroideo• Tumores de saco endolinfático• Hemangioblastoma• Hepatoblastoma	<ul style="list-style-type: none">• LMMJ• LLA hipodiploide baja• Tumor de la vaina nerviosa periférica maligno• Carcinoma medular de tiroides• Glioma óptico• Tumor de células de Sertoli-Leydig de ovario	<ul style="list-style-type: none">• Blastoma pleuropulmonar• Blastoma pituitario• Pinealoblastoma• Retinoblastoma• Schwannoma• Tumor subependimario de células gigantes
3. Un niño con 2 tumores primarios con inicio ≤ 18 años (Excepto tumores cuyo tiempo y tejido se asocien con el tratamiento)		
4. Un niño con cáncer y cualquier anomalía congénita y otros síntomas específicos		
<ul style="list-style-type: none">• Anomalías congénitas• Dismorfismo facial• Discapacidad intelectual• Aberraciones del crecimiento• Aberraciones de la piel• Enfermedades hematológicas• Inmunodeficiencia		
5. Un niño con excesiva toxicidad al tratamiento		

Los pacientes que cumplieran con al menos un criterio pueden referirse con el médico genetista para una evaluación más estandarizada y probable análisis genético-molecular.(30) Meses después, el grupo de expertos del *Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology* agregaron un sexto criterio, que el análisis genómico tumoral identificara alteraciones con altas probabilidades de origen germinal (11). Aunque esta herramienta no está recomendada en niños sanos, su aplicación en oncología pediátrica dependerá de los resultados que se obtengan a través de la validación a gran escala y en diferentes poblaciones.

1.4 Genes de predisposición a cáncer pediátrico y tumores asociados

Las leyes de la herencia permitieron establecer las bases de las enfermedades monogénicas, donde los primeros genes asociados con la predisposición hereditaria a cáncer fueron caracterizados en familias afectadas mediante el análisis de ligamiento genético. A partir de entonces, múltiples genotipos han sido identificados por ser causales de fenotipos malignos.

Gracias a las tecnologías de secuenciación masiva paralela (*NGS* por sus siglas en inglés) y las mejoras en el desarrollo de análisis genómicos exhaustivos, la investigación y conocimiento sobre la contribución de los factores hereditarios en el desarrollo de cáncer en niños se ha acelerado (31). Los resultados indican que una considerable proporción de tumores en niños y adolescentes, pueden originarse a partir de mutaciones germinales o mosaicos en genes de predisposición a cáncer (11, 19, 20).

A principios de los 90s, se estimó que entre el 1 – 10% de los tumores infantiles estaban asociados con un patrón hereditario. La secuenciación germinal en 1,201 individuos demostró que el 21% de los pacientes con cáncer de inicio temprano habían heredado una mutación genética, en comparación con el 13% de los pacientes con cáncer de adultos jóvenes (32). Actualmente, se ha postulado que la predisposición genética puede ser causante hasta en un 30% de todos los casos de cáncer en niños y adolescentes (33).

Un estudio realizado en el 2015 por Zhang y colaboradores (*Pediatric Cancer Genome Project/St. Jude Children's Research Hospital*) que incluyó 1,120 pacientes con cáncer menores de 20 años, demostró que el 8.5% de los pacientes eran portadores de mutaciones patogénicas en genes como TP53, APC, BRCA2, NF1, PMS2, RB1 y RUNX3, principalmente en tumores sólidos (16%), tumores cerebrales (8.6%) y leucemias (3.9%); dicho porcentaje fue mayor comparado con el 1.1% identificado en la población control del proyecto 1000 Genomas. Un dato a resaltar en este estudio, es que el 60% de los pacientes portadores de mutaciones no reportó datos de historia familiar de cáncer, por lo que puede explicar un probable sub-diagnóstico de casos relacionados con síndrome de cáncer hereditario o que la enfermedad en niño es tan temprana que aún no se han manifestado fenotipos malignos en familiares de primer o segundo grado (34). Los resultados de Zhang son consistentes con los publicados en 2018 en la cohorte del estudio Pan-Cancer que incluyó a 914 pacientes y se demostró que en el 7.6% de los pacientes existía una variante de línea germinal patogénica en genes como MSH2, MSH6, PMS2, TP53, BRCA2 y CHEK2 (19). El porcentaje más alto de mutaciones germinales predisponentes en niños con cáncer a la fecha fue reportada por Oberg y colaboradores, donde a través del análisis de secuenciación de exoma humano completo se demostró que el 14% de los individuos era portador de una mutación germinal en genes de predisposición a cáncer (35).

Aunque los estudios anteriores incluyeron un mayor número de individuos secuenciados, algunas de las principales limitaciones son: primero, no se incluyeron todos los tipos de tumores y algunos de ellos han mostrado una asociación sólida con mutaciones de la línea germinal (por ejemplo, hasta el 55% de los pacientes con sarcoma albergan al menos una

mutación) (36); segundo, la herencia y el origen de las mutaciones no se pueden establecer porque las pruebas de secuenciación no fueron realizadas a los padres o familiares y, por último, los factores ambientales no fueron evaluados. Por estas razones, es necesario más estudios de secuenciación de la línea germinal basada en la familia para establecer el origen de las mutaciones (padre, la madre, ambos o mutación *de novo*), las implicaciones hereditarias, la interacción entre el medio ambiente y la evaluación de las diferencias en la penetrancia o la expresividad de los genotipos. Algunos síndromes de predisposición a cáncer en menores de 18 años y tumores asociados se resumen en la tabla 2.

Tabla 2. Tumores pediátricos asociados con síndromes de predisposición

Tipo de tumor	Síndromes de predisposición	Gen(es) asociados
Leucemias	Anemia de Fanconi, NF-1, ATM, Li-Fraumeni, Nijmegen, retinoblastoma hereditario	FANCA, FANCB, BRCA, NF1, ATM, TP53, RB1
Linfomas	NF-1, ATM, Wiskott-Adrich, Bloom, Nijmegen, Li-Fraumeni, Beckwith-Wiedemann	NF1, ATM, WASP, BLM, NBS1, TP53, BRCA, (CDKN1C, H19, IGF2, KCNQ1OT1).
Tumores del SNC	Tumor rabdoide familiar, Li-Fraumeni, Noonan, MEN-1, NF-2, ATM, Beckwith-Wiedemann, Li-Fraumeni, Von Hippel Lindau, DICER1	SMARCB1, TP53, RAS, MEN1, NF2, ATM, (CDKN1C, H19, IGF2, KCNQ1OT1), TP53, VHL, DICER1
Neuroblastoma	NF1, Li-Fraumeni, Beckwith-Wiedemann, Noonan	NF1, TP53, (CDKN1C, H19, IGF2, KCNQ1OT1), RAS
Retinoblastoma	Retinoblastoma hereditario	RB1
Tumores renales	Beckwith-Wiedemann, Li-Fraumeni, Denysh Drash	(CDKN1C, H19, IGF2, KCNQ1OT1), TP53, WT1
Tumores hepáticos	Beckwith-Wiedemann, Li-Fraumeni, Anemia de Fanconi	(CDKN1C, H19, IGF2, KCNQ1OT1), TP53, FANCA, FANCB, BRCA
Sarcomas óseos	Li-Fraumeni, retinoblastoma hereditario, Beckwith-Wiedemann,	TP53, RB1, (CDKN1C, H19, IGF2, KCNQ1OT1)
Sarcomas de tejidos blandos	Li-Fraumeni, retinoblastoma hereditario, NF-1, DICER1, Beckwith-Wiedemann	TP53, RB1, NF1, DICER1
Tumores germinales	Denysh Drash, Cowden, Noonan, Peutz-Jeghers	WT1, PTEN, RAS, STK11
Tumores epiteliales	Li-Fraumeni, Beckwith-Wiedemann, XP, melanoma cutáneo maligno hereditario	TP53, (CDKN1C, H19, IGF2, KCNQ1OT1), (XPA – XPG y ERCC5), CDKN2A

También es importante resaltar que, por sí mismas, las mutaciones germinales en genes de predisposición a cáncer no son suficientes para iniciar la oncogénesis; las células de dichos pacientes requieren adquirir una o más mutaciones posteriores (teoría del doble golpe de Knudson) (37). Es por ello que, para algunos síndromes, no todos los que heredaron una mutación desarrollarán cáncer (debido a que no adquieren una segunda mutación) (38, 39).

La investigación sigue en curso, y es probable que el porcentaje de pacientes portadores de mutaciones germinales de predisposición a cáncer aumente a medida que se caractericen nuevas variantes en genes aún no identificados; por lo que el beneficio será directo tanto para los niños con cáncer y sus familiares, a través de la identificación de individuos con mayor riesgo, mejoras en la prevención, vigilancia, tratamiento y seguimiento de la enfermedad.

1.5 Factores de riesgo en cáncer pediátrico

Mientras que, en adultos, los factores de riesgo relacionados con el ambiente y estilo de vida como hábitos nocivos, dieta, actividad física, entre otros, son determinantes para el desarrollo de cáncer, en niños, hasta el momento no se ha demostrado un papel similar debido al tiempo corto de exposición y edad temprana al diagnóstico.

La radiación ionizante y quimioterapia previa han sido los pocos factores ambientales asociados por aumentar el riesgo a cáncer en la infancia (riesgo relativo, RR > 2). Los resultados de estudios epidemiológicos han demostrado que la exposición a radiación ionizante durante la infancia puede resultar en un mayor riesgo a cáncer en comparación con adultos (principalmente a leucemias), reflejando una mayor sensibilidad en niños, probablemente porque los órganos y tejidos aún se encuentran en desarrollo (40). Los estudios posteriores al accidente nuclear en Chernóbil demostraron un aumento de cáncer de tiroides en niños y adultos ante la exposición de radiación de más de 1 Gy. Sin embargo, factores como la exposición parental a pesticidas (RR 1.40 – 5), bencenos (RR 1.48 – 1.70) y tabaquismo en padres (RR 1.06 – 1.24) también se han asociado con un mayor riesgo a cáncer pediátrico (41, 42); sin embargo el riesgo asociado a tabaquismo en madres se limitó únicamente al origen hispanico (RR 1.20 – 3.61) (43). Otros factores como el peso al nacer (macrosomía) y edad avanzada en padres también han sido demostrados de manera consistente en varios estudios, mientras que los resultados sobre el consumo de alcohol, tabaco, café, vitaminas y exposición a campos electromagnéticos en periodos pre y postnatal y el riesgo a cáncer en niños han sido inconsistentes (18).

Algunos factores de riesgo descritos recientemente incluyen la transmisión del cáncer de madre a hijo en niños menores de 2 años, presumiblemente por vía transplacentaria. Un reporte reciente documentó dos casos de cáncer de pulmón en niños de 2 y 6 años causado por la transmisión de tumores de cuello uterino de las madres de los niños. Los datos de la revisión histopatológica y la secuenciación de ambos tumores confirmaron la presencia del mismo fenotipo y genotipo tumoral, así como su origen a partir de la aspiración de fluidos contaminados con células tumorales durante el nacimiento, además de metástasis asociadas con una probable diseminación hematogena a partir de la placenta (44).

Un enfoque que puede ayudar a entender las bases biológicas del cáncer pediátrico y cómo el daño a las células puede manifestarse en un periodo de tiempo tan corto, es la interacción entre los factores genéticos y ambientales.

1.5.1 Interacción gen-ambiente (GxE)

El cáncer se considera una enfermedad compleja, donde el fenotipo resulta de la interacción entre factores genéticos y ambientales. En algunas situaciones (menos comunes), dichos factores genéticos penden de un solo gen, sin embargo, en la gran mayoría depende de la interacción entre varios genes. De hecho, las estimaciones altas de heredabilidad del cáncer obtenidas a partir de estudios en gemelos homocigotos sugieren que gran parte de los tumores tienen una base poligénica, donde la herencia de múltiples variantes genéticas (con efecto individual modesto), puede explicar una proporción de riesgo poblacional más amplia. Lo anterior explica el papel que tienen los factores ambientales y el estilo de vida para el desarrollo de cáncer (45).

Entender cómo factores genéticos y ambientales influyen en el desarrollo de cáncer, ha sido un gran reto para la ciencia. Sin embargo, el estudio de las interacciones gen-ambiente (*GxE* por sus siglas en inglés) es una estrategia para mejorar el descubrimiento de variantes genéticas asociadas con cáncer y los efectos ambientales en individuos genéticamente susceptibles. Existen 3 principales diseños para el estudio de las interacciones *GxE*, cuyas ventajas y desventajas se describen en la tabla 3.

El estudio de las interacciones *GxE* han sido muy importantes en salud pública, ya que permiten mejorar la normativa ambiental, identificar personas con riesgo elevado y establecer medidas preventivas, por lo que su estudio permitirá mejorar la precisión y exactitud en la evaluación tanto de influencias genéticas como ambientales, ya que los efectos ambientales son evidentes en personas genéticamente susceptibles y las influencias genéticas son claras en personas con antecedentes de exposición apropiados. Hasta el momento no es posible reparar el daño en los genes, pero si modificar el entorno para evitar o retrasar el desarrollo de un tumor.

Existen pocos estudios que han demostrado posibles interacciones de *GxE* en el cáncer infantil. Aurora Medina-Sanson y colaboradores realizaron un estudio de casos y controles de 478 niños con LLA y 248 controles, donde se demostró que el polimorfismo rs1799929-TT en el gen NAT2 confiere un riesgo elevado a LLA ante la exposición a fertilizantes, insecticidas, derivados de hidrocarburos y tabaquismo de los padres (RM 1,96; IC del 95%: 1,55-2,49) (46). Aunque este estudio se realizó en población mexicana mestiza, el polimorfismo y riesgo asociado puede variar en otras poblaciones.

Tabla 3. Diseños para el estudio de interacciones GxE

Tipo	Ventajas	Desventajas
Estudio basado en familias	Permiten la evaluación de la transmisión Mendeliana de alelos Disminuyen el sesgo relacionado con la estratificación poblacional	A veces es complicada la recolección de información genética en padres (especialmente cuando la enfermedad se presenta a edades avanzadas) o encontrar controles adecuados entre hermanos
Casos y controles	Ideal para el estudio de enfermedades raras con exposición a factores ambientales comunes	Sesgo de selección (i.e., los controles pueden no representar la población en la que ocurrieron los casos)
Estudios de cohorte	Permiten obtener información sobre la exposición antes de que se manifieste la enfermedad.	Requiere tamaños de muestra muy grandes y el seguimiento es muy extenso Son muy caros y largos

1.6 Estrategias genómicas en cáncer pediátrico

La revolución tecnológica en los últimos 10 años ha permitido el desarrollo de la secuenciación NGS, la cual, es ampliamente utilizada en el estudio y diagnóstico de enfermedades como el cáncer. De hecho, el descubrimiento de variantes causales ha sido acelerado gracias a la implementación de esta tecnología, costo-efectividad, velocidad de análisis y disponibilidad. Aunque uno de sus principales retos es el análisis bioinformático y la interpretación de la información, existen esfuerzos para generar bases de datos genómicas agregadas, buenas prácticas de análisis y guías para el análisis de datos (47-49).

Gracias a su escalabilidad, la secuenciación NGS es utilizada en diferentes estrategias de análisis en cáncer pediátrico (Tabla 4) y diseños metodológicos (secuenciación de casos, tríos o casos y controles).

Tabla 4. Estrategias de secuenciación en cáncer pediátrico

	Genoma humano completo	Exoma humano completo	Paneles
Principal aplicación	Estudios de genómica de poblaciones	Identificación de nuevos genes causantes de enfermedad	Análisis de genes conocidos asociados a condiciones específicas
Porcentaje analizado del genoma	95-98%	1.5-2.0%	>0.1%
Número de variantes detectadas	4,000,000	20,000	Variable, depende del tamaño del panel
Regiones analizadas	Regiones exónicas, intrónicas y genoma mitocondrial	Regiones exónicas y uniones exón-intrón	Regiones de interés (hotspots)

Estrategia para generación de bibliotecas	Ninguna (Shutgun)	Selección (captura)	Selección (captura o amplificación)
Profundidad de secuenciación	>30X	50 – 100X	>500X
Tipos de variantes analizadas	Variantes de nucleótido único (SNVs) Inserciones/delecciones (INDELs) Variantes estructurales (SVs) Variantes de número de copias (CNVs) Expansiones de repetición ADN mitocondial (mtDNA)	Variantes de nucleótido único (SNVs), inserciones/delecciones (INDELs) y variantes de número de copias (CNVs) en regiones codificantes	Depende del tipo de variantes de interés
Ventajas	Identificación de una mayor cantidad y tipo de variantes nuevas en regiones codificantes, no codificantes mitocondriales Mejor uniformidad de secuenciación	Identificación de variantes nuevas en regiones codificantes Cubre el 85% de las variantes de relevancia clínica Costo mas accesible en comparación con genoma completo	Menor costo Adaptable a necesidades específicas
Limitaciones	Costos altos Análisis computacional más complejo Difícil anotación e interpretación de variantes Requiere alta capacidad de almacenamiento de datos	No identifica CNVs en regiones no codificantes La cobertura de secuenciación heterogénea	Uso para análisis y confirmación de genotipos específicos

La estrategia predilecta de análisis genético en la investigación de síndromes de predisposición a cáncer ha sido la secuenciación de individuos afectados (casos), cuyo uso se ha extendido hasta la aplicación clínica. Una de las principales limitaciones que tiene dicha estrategia en la investigación del cáncer pediátrico, es que no permite conocer la proporción de mutaciones heredadas vs las mutaciones *de novo*. La secuenciación en tríos es una estrategia que permite subsanar esta y otras limitaciones, como la mejora en el entendimiento de los patrones de herencia y caracterización de mutaciones de penetrancia incompleta, ya que, la “ausencia” de historia familiar de cáncer en familiares de primero o segundo grado, puede deberse más bien a que la enfermedad no ha tenido el tiempo suficiente para manifestarse (50-52). Algunas de las ventajas de los estudios de secuenciación en tríos, se enlistan en la tabla 5.

Tabla 5. Aplicaciones de los estudios de secuenciación en tríos

Ventajas	Aplicaciones
Determinar patrones de herencia o mutaciones de <i>novo</i>	Identificación de riesgo a cáncer ligado al sexo Identificación de al menos 2 mutaciones germinales monoalélicas en diferentes genes de vías relacionadas. Definir si las mutaciones fueron heredadas por el padre, madre o ambos; o si una fue heredada por alguno de los progenitores y otra se originó <i>de novo</i> .

	<p>Pacientes con fenotipos de síndromes de predisposición con ausencia de mutaciones clásicas, como el fenotipo tipo Li-Fraumeni asociado a mutaciones en ATM y CHEK2</p> <p>Presencia de genotipos concomitantes de baja penetrancia en pacientes sin historia familiar de cáncer</p> <p>Identificación de casos índice con fenotipos portadores de mutaciones bialélicas, cada una heredada por un progenitor monoalélico o consanguíneo</p> <p>Confirmación de mosaicos genéticos</p> <p>Definir el origen <i>cis</i> o <i>trans</i> de mutaciones concomitantes</p> <p>Estimar la proporción de tumores originados a partir de mutaciones germinales</p> <p>Definir las implicaciones de variantes de significado incierto</p>
Evaluar efectos de la exposición ambiental	<p>Identificar el impacto de factores ambientales en pacientes con mutaciones <i>de novo</i></p> <p>Relacionar genotipos de baja penetrancia sensibles a factores ambientales</p> <p>Implicaciones genéticas y ambientales del “doble golpe”</p> <p>Impacto de la heredabilidad y factores ambientales en gemelos homocigotos</p> <p>Influencias de los factores ambientales sobre la penetrancia mutacional en diferentes generaciones</p>
Vigilancia e identificación de familiares con riesgo	<p>Detección temprana en individuos con riesgo</p> <p>Identificación de pacientes susceptibles a tumores metacrónicos o sincrónicos</p> <p>Establecer planes de reducción de riesgo en casos índice (evitar radioterapia en pacientes con mutaciones germinales en TP53) y probables familiares afectados (dosis bajas de aspirina en pacientes con síndrome de Lynch, quimioprevención, ultrasonido abdominal o IRM de cuerpo completo en pacientes con síndrome de Li-Fraumeni)</p>

La obtención de ADN de alta calidad para análisis genómicos es un elemento clave para la adecuada caracterización de variantes germinales con relevancia clínica. El ADN genómico (ADNg) purificado a partir de muestras de sangre periférica, ha sido el estándar de oro para análisis genéticos, sin embargo, su obtención es invasiva, requiere personal de salud altamente capacitado para su toma, transporte y almacenamiento en cadena fría y tiempo corto de estabilidad (máximo 7 días).

La saliva es un biofluido constituido por agua en el 99%, componentes orgánicos 0.3% (mucina, amilasa, proteínas plasmáticas) y sustancias inorgánicas en el 0.2% restante (53). Se postula que las moléculas derivadas de la sangre ingresan a las glándulas salivales a través de 4 mecanismos: difusión intercelular (moléculas de un peso inferior a 1900 D), filtración a través de poros de membranas (agua, electrolitos y moléculas inferiores a 400 D), ruta transmembranal por difusión pasiva (moléculas lipofílicas) o transporte activo por canales proteicos (péptidos) y pinocitosis. Una vez secretada a la cavidad bucal, se mezcla con líquido crevicular, el cual contiene leucocitos, células epiteliales, microorganismos y detritus alimenticio (54).

La fuente de ADNg en saliva son los leucocitos (principalmente los PMNs) y células descamadas del epitelio bucal. En algunos estudios genómicos y genéticos, las tomas de muestra de saliva han sido una excelente alternativa debido a su nula invasividad, recolección por los mismos pacientes, fácil manejo a través de correo ya que no requiere de cadena fría y estabilidad a largo plazo a temperatura ambiente (55, 56).

Cada vez más reportes ponen en evidencia las similitudes de rendimiento en calidad y cantidad del ADNg de saliva vs sangre, así como la concordancia y consistencia de resultados obtenidos principalmente en estudios de genotipificación, chips de arreglos de genoma completo o secuenciación dirigida. Una de las principales limitaciones observadas en las muestras de saliva, es la contaminación con ADN bacteriano, sin embargo, el uso de dispositivos con agentes estabilizadores ayuda a disminuir hasta el 11.8% de ADN bacteriano, asegurando un contenido de ADNg humano del 88%. Algunos reportes han demostrado

similitudes en el porcentaje de lecturas no alineadas al genoma de referencia humano entre ambos tipos de muestra (5.3% vs 4.4%). Otros estudios de secuenciación de exomas humanos no han reportado diferencias significativas en las tasas de errores, así como una concordancia del 100% de las variantes detectadas entre ambos tipos de muestra (57, 58).

1.7 Nuevos abordajes del cáncer pediátrico

En los últimos años, la introducción de la secuenciación NGS en la práctica clínica ha sido fundamental en el diagnóstico, el manejo y el tratamiento de enfermedades humanas.

Inicialmente, las pruebas genéticas en menores de edad no eran realizadas debido a la falta de recomendaciones preventivas y de vigilancia disponibles para guiar el manejo clínico de los pacientes portadores de mutaciones; sin embargo, organizaciones como *The American College of Medical Genetics and Genomics (ACMG)*, *The American Association of Pediatrics (AAP)*, *The American Society of Human Genetics (ASHG)* and *The American Society of Clinical Oncology (ASCO)* han establecido recomendaciones con fines diagnósticos (confirmación de síndrome genético) y recomendaciones de tamizaje (individuos sin cáncer con parientes cercanos con una condición de predisposición conocida) en población pediátrica, siempre y cuando la prueba genética (generalmente panel de genes específicos) contribuya con su manejo médico y vigilancia (59).

Se reportado que el 75% de los niños que cumplen con los criterios clínicos de síndromes de cáncer genético son portadores de mutaciones patógenas de la línea germinal,(60) por lo que el asesoramiento genético es fundamental para identificar a los individuos de alto riesgo, implementación de estrategias de reducción de riesgo y toma de decisiones. Además, el asesoramiento previo y posterior a la prueba permite aclarar los posibles riesgos y beneficios, ya que algunas pruebas de NGS como la secuenciación de exoma humano completo (*WES*) pueden revelar muchas variantes de significado incierto (*VUS*) o hallazgos incidentales como el riesgo a una enfermedad no oncológica.

Para su aplicación clínica, las pruebas genéticas deben realizarse en laboratorios que cumplan con estándares de alta calidad, personal capacitado y buenas prácticas de laboratorio. Todos los paneles deben someterse a protocolos de estandarización y validación rigurosos que muestren la precisión y las limitaciones metodológicas (61).

Algunos paneles comerciales con validación clínica que están disponibles para pruebas genéticas de cáncer infantil son: Panel de cáncer pediátrico hereditario (*Blueprint Genetics*) que detecta variantes de nucleótido único (*SNV's*), deleciones/inserciones (*INDEL's*) y variantes de número de copias (*CNV's*) en 71 genes (62); Paneles de tumores sólidos pediátricos, tumores del sistema nervioso/cerebrales pediátricos y neoplasias malignas pediátricas y hematológicas (*INVITAE*) que incluye secuenciación de exoma completo (+/- 20pb de las uniones exón-intrón) en 53, 34 y 16 genes respectivamente (63). Instituciones como *Memorial Sloan Kettering* a través del Servicio de Genética Clínica para Niños y el Programa de Detección de Predisposición al Cáncer Pediátrico (64) y el Programa de Riesgo

Genético del Cáncer Pediátrico del Centro Infantil de *Dana-Farber Boston* ofrecen pruebas genéticas para varios síndromes genéticos (65).

Aunque las variantes genéticas heredadas identificadas han proporcionado información importante sobre las causas del cáncer infantil, aún existen brechas críticas en el conocimiento de esta patología. La mayoría de las variantes genéticas que causan enfermedades siguen sin identificarse debido a las limitaciones propias de la secuenciación en regiones genómicas como regiones intragénicos/intergénicos u homopoliméricas y falta evidencia clínica de causalidad a cáncer, predicción computacional y ensayos funcionales.

De hecho, uno de los mayores desafíos en el asesoramiento genético es la discusión de los resultados con variantes VUS, que representan una gran proporción de las variantes identificadas por las tecnologías NGS. De hecho, se recomienda el seguimiento semestral de las variantes VUS, con el fin de comprobar cambios en su clasificación (en casos poco frecuentes, hacia la patogenicidad). El establecimiento y validación de factores de riesgo infantil en estudios poblacionales a gran escala, la secuenciación familiar, la discusión de variantes en grupos multidisciplinarios de expertos y la evaluación de factores de riesgo en el período prenatal, embarazo y posnatal, mejorarán el conocimiento de la base genética y el origen de tumores infantiles.

Como parte de los esfuerzos internacionales para el desarrollo de estudios a gran escala, el presente estudio piloto tiene el objetivo de establecer las bases para desarrollo del proyecto genómico “1000 genomas: Factores de predisposición genética y ambiental para cáncer pediátrico”, con el fin de describir la frecuencia y distribución de mutaciones germinales en niños con cáncer de población mexicana.

2.- Planteamiento del problema

El cáncer pediátrico es la primera causa de muerte por enfermedad en niños de 5 a 14 años, para el cual, no existen estrategias de prevención ni reducción de riesgo. En nuestro país, el porcentaje de supervivencia global a 5 años permanece bajo en comparación con Estados Unidos y Canadá, disparidad que puede no estar exclusivamente asociada al nivel socioeconómico. Debido a la edad al diagnóstico y corto tiempo de exposición a factores ambientales, mutaciones germinales en genes de predisposición a cáncer pueden tener un papel importante en el desarrollo de esta enfermedad, lo cual, ha sido comprobado en estudios con otras poblaciones; sin embargo, no existen estudios genéticos y de factores ambientales en población mexicana que demuestren la frecuencia y distribución de este tipo de mutaciones

3.- Justificación

La investigación sobre las bases genéticas del cáncer en niños mexicanos a partir de una estrategia costo-efectiva como la secuenciación de exomas en tríos, permitirá el esclarecimiento de la herencia del cáncer pediátrico a nivel local y global, así como su interacción con factores ambientales.

Lo anterior, contribuirá a mejoras en la identificación de niños portadores de este tipo de mutaciones y subsecuente implementación de estrategias de reducción de riesgo, beneficio para pacientes y familiares.

4.- Objetivos

a. General

Describir la distribución y espectro de mutaciones germinales mediante secuenciación de exomas en pacientes pediátricos oncológicos sin criterios de sospecha a cáncer hereditario a partir de muestras de saliva

b. Específicos

1. Reclutamiento de pacientes y/o familias con cáncer pediátrico
2. Elaboración de familogramas y aplicación de cuestionario de riesgo a cáncer pediátrico
3. Estandarizar un protocolo de recolección de saliva y purificación de ADN para análisis de exomas.
4. Secuenciación de nueva generación (NGS) de exomas de niños con cáncer y sus progenitores
5. Estandarizar el flujo de análisis bioinformático para el análisis de exoma humano
6. Verificar la cobertura de bases alineadas al genoma de referencia entre las tomas de muestra de sangre y saliva
7. Identificar mutaciones conocidas y relacionadas con síndromes de cáncer hereditario
8. Validación las variantes identificadas
9. Correlación de las mutaciones identificadas con la historia personal y familiar de cáncer
10. Establecer las bases para el proyecto "1000 familias"

5.- Hipótesis

La prevalencia de mutaciones germinales identificadas mediante secuenciación de exomas de niños con cáncer será mayor del 8% en población de estudio.

6.- Metodología

6.1 Tipo de estudio

El presente proyecto de investigación es un estudio piloto, observacional y descriptivo que fue aprobado por el comité de ética e investigación de la División de Estudios de Posgrado e Investigación de la Facultad de Odontología, UNAM (CIE/0103/08/2018).

6.2 Población de estudio

El objetivo de este estudio fue reclutar tríos de pacientes con diagnóstico de leucemia, linfoma, tumores del sistema nervioso central (SNC), sarcomas de tejidos blandos, sarcomas óseos y tumores hepáticos entre los 0 y 18 años, de cualquier género y estadio que aceptaran participar en el proyecto.

Las muestras de los pacientes reclutados hasta el momento provienen de 2 fuentes, la primera corresponde a voluntarios que participaron en el proyecto del Genoma del Cáncer Pediátrico realizado en colaboración con el Hospital Infantil de México “Federico Gómez” en el año 2011, las cuales, fueron procesadas en el *Broad Institute* como parte de la fase inaugural de la *Slim Initiative in Genomic Medicine for the Americas, SIGMA*; además, se han obtenido muestras del Hospital de Pediatría del CMNSXXI y del Hospital del niño DIF-Hidalgo, la cuales, fueron procesadas y secuenciadas en el INMEGEN, México.

6.3 Criterios de selección

Criterios de inclusión

- Pacientes con diagnóstico sugestivo de sarcomas de hueso y tejidos blandos, tumores de sistema nervioso central, y cánceres hematológicos
- Que aceptaran participar y firmar la carta de consentimiento informado
- Pacientes sin tratamiento previo

Criterios de Exclusión

- Pacientes con tratamiento previo

Criterios de Eliminación

- Pacientes cuyo diagnóstico histopatológico postquirúrgico excluya cáncer
- Pacientes con muestra insuficiente o de mala calidad
 - Muestra con menos de 80% de células tumorales
 - Menos de 15 µg de ADN total
 - Ácidos nucleicos degradados

Se elaboraron familiogramas y se revisaron los expedientes de los pacientes secuenciados para obtener información respecto a las características del tumor, antecedentes patológicos y no patológicos personales e historia familiar. Además, el riesgo a cáncer pediátrico fue determinado mediante la herramienta desarrollada por Jongmans Marjolin (30). La información clínica fue correlacionada con las mutaciones identificadas en los pacientes y tipos de tumores.

6.4 Tomas de muestra, extracción y purificación de ADN

Saliva

Se realizó un protocolo para estandarización de tomas de muestra de saliva, en el cual, se evaluará el desempeño de los kits *Oragene DNA (DNAgenotek)* y *SalivaGene Collection (Stratec)*, así como tomas de muestra convencional recolectadas en tubos tipo Falcon. Las condiciones de toma de muestra fueron determinadas mediante las recomendaciones del proveedor y revisión bibliográfica de guías, manuales, recomendaciones y buenas prácticas de tomas de muestra de saliva.

La extracción y purificación de ADNg se realizó con el kit *Gentra PureGene (Qiagen)* siguiendo las buenas prácticas de laboratorio. En breve, 500 μ L de saliva previamente concentrada por centrifugación fueron utilizados para lisis celular durante 15 minutos a 65°C, posterior a la precipitación de proteínas, el sobrenadante fue utilizado para subsecuente purificación de ácidos nucleicos.

Sangre

Las muestras de sangre fueron colectadas a partir de una punción venosa convencional. El ADN fue extraído a partir de *buffy coats* utilizando el kit *Gentra PureGene Blood (Qiagen)*, siguiendo las instrucciones del fabricante

Controles de calidad

El ADNg purificado fue cuantificado mediante NanoDrop 1000 (relación de absorbancia 260/280 nm) y fluorimetría con el kit *Qubit dsDNA HS Assay (Life Technologies)*, además su integridad fue evaluada mediante geles de agarosa al 1% en solución amortiguadora Tris-acetato EDTA (TAE) 100V fijos durante una hora. Las bandas de ácidos nucleicos fueron visualizadas a través de la incubación de geles de agarosa con GelRed 1X. El ADN fue congelado a -80 C hasta su uso.

6.5 Preparación de bibliotecas y secuenciación de exomas humanos

3 μ g totales de ADNg fueron utilizados para la preparación de bibliotecas mediante el kit *Sure-Select Human All Exon V4 (Agilent Technologies)* para su subsecuente secuenciación de acuerdo con el protocolo utilizado por el *Broad Institute* (Figura 3) (66); además, algunas de las muestras fueron procesadas en el INMEGEN de acuerdo a sus

protocolos con TrueSeq Exome kit (*Illumina*). En breve, el ADNg fue fragmentado con energía ultrasónica (*Covaris System*), para subsecuente ligación de adaptadores y amplificación. Los fragmentos amplificados fueron hibridados con las sondas para posterior captura mediante perlas magnéticas cubiertas con estreptavidina y ligación de índices. Las librerías purificadas fueron bioanalizadas para determinar su longitud (250-350 pb) y concentración para finalmente ser multiplexadas. 8-12 pM de librerías fueron utilizadas para secuenciar en el equipo *HiSeq 2000* (*Illumina Inc*).

6.6 Análisis bioinformático y clasificación de variantes

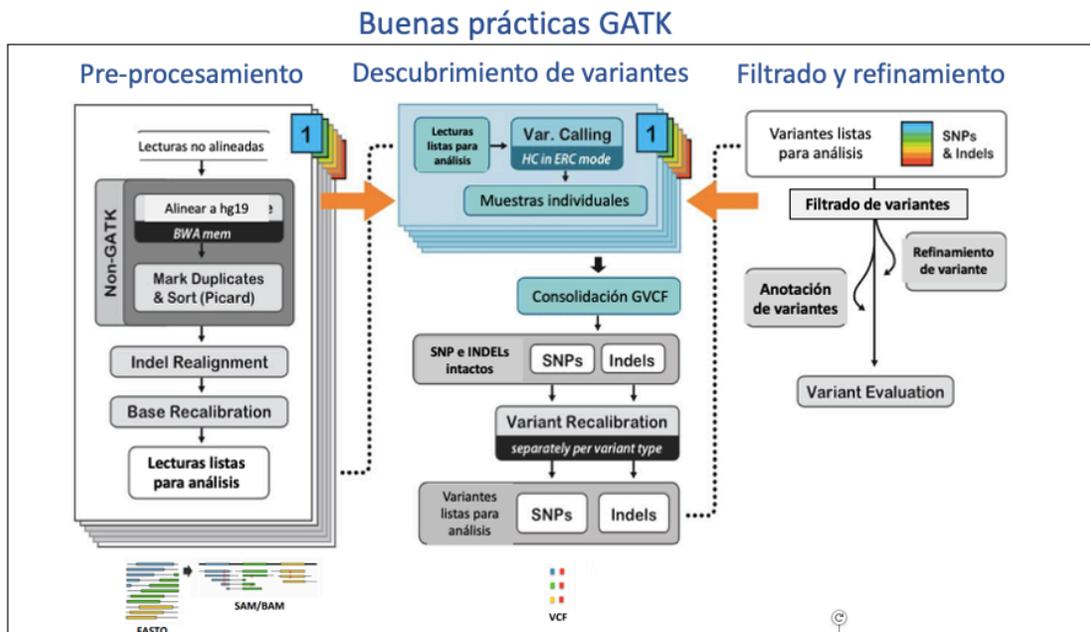


Figura 3. Flujo de trabajo de las buenas prácticas del Genome Analysis ToolKit, GATK

Las lecturas fueron mapeadas al genoma de referencia hg19 mediante el programa *Burrows-Wheeler Aligner* (*BWA*) para posterior generación de archivos BAM con *Picard tools* 1.110 (67). Dichas lecturas fueron procesadas siguiendo las buenas prácticas del *Genome Analysis Tool Kit*, *GATK* (*Broad Institute*) con la versión 4.1.7.0 (68). Los archivos BAMs cuya fracción de bases a 20X era <0.6 fueron eliminados para el llamado de variantes. Los SNPs e INDELS en cada muestra fueron llamados con *HaplotypeCaller* (modo *GVCF*). Los archivos VCF de cada paciente fueron unificados para posterior llamado en conjunto con la herramienta *GenotypeGVCF*. Las variantes fueron filtradas para reducir la cantidad de falsos positivos a través de la recalibración del puntaje de calidad de la variante (*VQSR*) mediante un modelo de entrenamiento a partir de archivos con variantes de referencia como dbSNP, HapMap 3.3, 1000G_phase1 high confidence (snps e indels), Omni 2.5 y, Mills_1000G gold standard indels.

Las variantes recalibradas fueron anotadas utilizando cadenas de filtrado por niveles. Primero, todas las variantes con una frecuencia alélica menor $>1\%$ fueron descartadas ($MAF > 0.1$); segundo, exclusión de variantes sinónimas; tercero, las variantes *missense* (*en sentido erróneo*), *nonsense* (*sin sentido*), *frameshift* (*cambio del marco de lectura*) SNP (variantes de nucleótido único) e INDELS (deleciones/inserciones) y de sitio *splicing* (*corte y empalme*) fueron incluidas para subsecuente evaluación mediante algoritmos de predicción como *SIFT*, *PolyPhen*, *FATHAMM* y *CADD*. La interpretación clínica de las variantes se realizó mediante la revisión en bases de datos *ClinVar*, *HGMD* y *ClinGen*.

Las variantes filtradas fueron clasificadas con base en las recomendaciones del *American College of Medical Genetics and Genomics* (ACMG) de acuerdo con el índice de patogenicidad y evidencia científica disponible (48).

7.-Resultados

7.1 Reclutamiento de pacientes y/o familias

El proyecto del Genoma del Cáncer Pediátrico reclutó 263 niños con diagnóstico de cáncer, de los cuales, se cuentan con tejido tumoral y no tumoral, además de muestras de sangre de padres biológicos. De las 263 muestras, se cuenta con 105 archivos BAM de análisis germinal y somático secuenciados en el Broad Institute. Para este estudio piloto, se incluyeron 69 archivos BAMs del proyecto antes mencionado y 9 individuos secuenciados en el INMEGEN, dando un total de 78 individuos. Tres individuos fueron eliminados previo al llamado de variantes debido a una baja fracción de bases a 20X y otro fue eliminado por duplicidad. El universo de pacientes y la población de estudio se presentan en la figura 4.

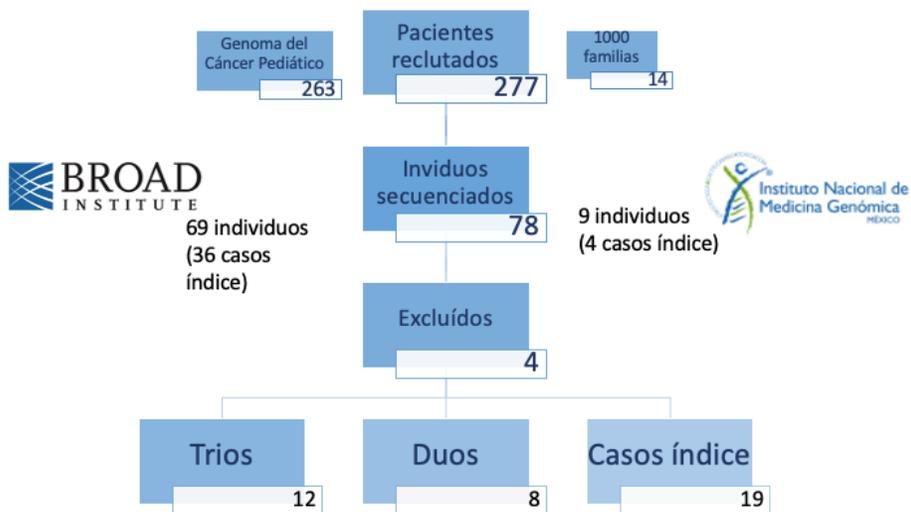


Figura 4. Pacientes reclutados, en proceso de secuenciar y en proceso de análisis. Las características clínicas y demográficas se presentan en la estadística descriptiva

7.1.1 Estadística descriptiva

Casos índices reclutados

Hasta el momento, se cuentan con 277 casos índice reclutados, de los cuales, el 58% corresponde a tumores sólidos y el 42% a neoplasias hematológicas con un promedio de edad de 110 meses y un rango entre 5 y 205 meses. De acuerdo con la *International Classification of Childhood Cancer*, 3ra edición (ICCC-3) (17), la frecuencia de tumores se describe en la figura 5.

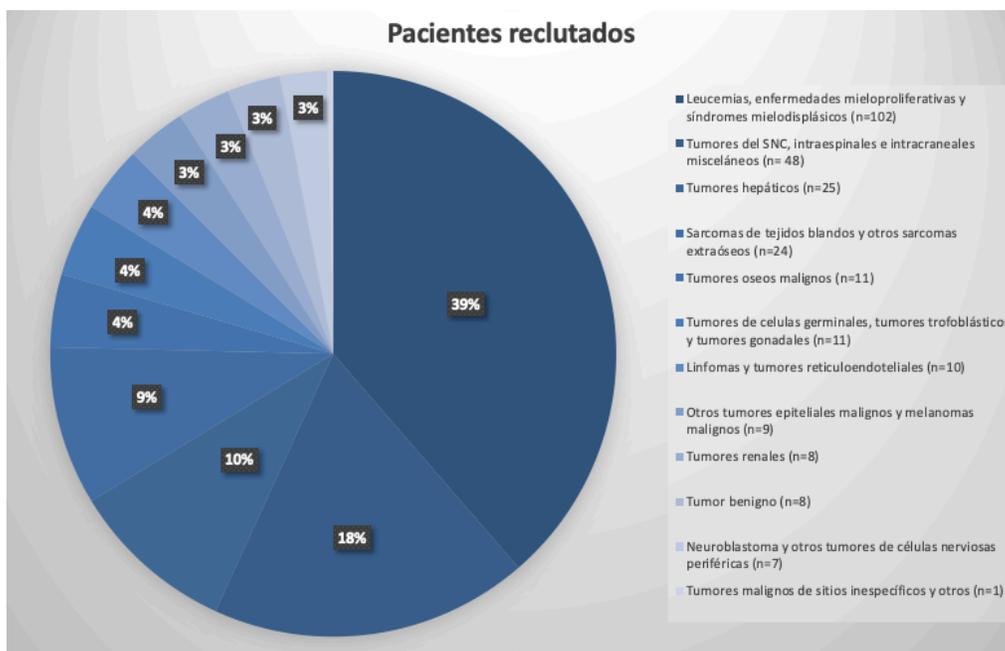


Figura 5. Frecuencia y distribución de los 277 casos índice reclutados.

Nuestra cohorte de pacientes presenta una mayor frecuencia de tumores del SNC y hepáticos en comparación con los últimos datos epidemiológicos de la SEER (69), aunque 13 pacientes no cuentan con diagnóstico confirmado.

Casos índices secuenciados

Fueron secuenciados 78 individuos, de los cuales 40 corresponden a casos índice. El 52% de los pacientes se encuentran en trio, 20% dúos y 28% solo caso índice. Las características clínicas y sociodemográficas se resumen en la tabla 6.

Tabla 6. Estadística descriptiva de los pacientes secuenciados

Información demográfica		
n = 40	Frecuencia	%
Sexo		
Femenino	24	60.0
Masculino	16	40.0
Edad al diagnóstico (años)		
Promedio (rango) \pm DE	6.5 (0 - 17) \pm 4.91	
Lugar de Nacimiento		
Ciudad de México	12	30.0
Estado de México	11	27.5
Guanajuato	3	7.5
Guerrero	3	7.5
Veracruz de Ignacio de la Llave	3	7.5
Hidalgo	2	5.0
Chiapas	1	2.5
Michoacán de Ocampo	1	2.5
Oaxaca	1	2.5
Sinaloa	1	2.5
Puebla	2	5.0
Antecedentes personales patológicos		
No	26	65
Varicela	8	20
Sarampion	1	2.5
Sin dato	5	12.5
Tipo de tumor (International Classification of Childhood Cancer, Third edition)		
n = 40	Frecuencia	%
Tumores hepáticos		
Hepatoblastoma	12	32.5
Carcinomas hepáticos	1	
Leucemias, enfermedades mieloproliferativas y síndromes mielodisplásicos		
Leucemias linfoides	9	22.5
Tumores del SNC, intraespinales e intracraneales misceláneos		
Ependimomas y tumores del plexo coroideo	5	20.0
Tumores intracraneales y embrionarios intraespinales	3	
Tumores oseos malignos		
Osteosarcomas	4	12.5
Tumor de células gigantes en columna	1	
Linfomas y tumores reticuloendoteliales		
Linfoma Hodgkin	1	2.5
Tumores renales	2	
Nefroblastoma y otros tumores renales no epiteliales	2	5.0
Sarcomas de tejidos blandos y otros sarcomas extraóseos		
Fibrosarcomas, tumores periféricos de la vaina y otras neoplasias fibrosas	1	2.5

Tumores de células germinales, tumores trofoblásticos y tumores gonadales	1	2.5
Información clínica		
n = 40	Frecuencia	%
Estado al diagnóstico		
Diagnóstico de primera vez	38	95
Recurrencia	2	5
Tratamiento previo		
Si	27	67.5
No	10	25.0
No confirmado	3	7.5
Historia Familiar de cancer		
Si	25	62.5
No	15	37.5
Metástasis (Solo tumores sólidos, n = 30)		
Si	14	46.66
No	12	40
No confirmado	4	13.33
Quimioterapia		
Si	33	82.5
No	4	10
Sin dato	3	7.5
Radioterapia		
No	21	52.5
Si	15	37.5
Sin dato	4	10
Estado		
Vigilancia	19	47.5
Muerto	10	25
Desconocido	11	27.5

Más del 50% de estos pacientes nacieron en el centro del país, debido probablemente por el lugar de atención. El promedio de edad de los pacientes es de 6.5 años, con un rango de 0 a 17 años. Existe un predominio por el género femenino en comparación con el masculino (60% vs 40%). Cabe destacar que la edad promedio de los padres de dichos pacientes es de 35 años, lo que podría explicar la ausencia de fenotipos malignos en familiares de primero o segundo grado en algunos casos.

El 75% de los casos índice tienen diagnóstico de algún tumor sólido, principalmente hepatoblastomas, osteosarcomas y ependimomas/tumores del plexo coroides, mientras que del 25% restante, 9 casos corresponden a leucemia linfoblástica aguda y 1 caso a linfoma de Hodgkin.

Respecto a las características clínicas, al momento de la toma de muestra el 95% de los pacientes recibieron por primera vez diagnóstico de cáncer, solo 2 pacientes presentaban recaída, uno con ependimoma y otro paciente con un segundo tumor primario (Linfoma de Hodgkin y tumor de Askin; de estos pacientes, el 67.5% ya habían recibido algún tipo de tratamiento. La presencia de metástasis fue evaluada en tumores sólidos (30 casos), de los

cuales el 46.7% cursaba con enfermedad metastásica, principalmente en pacientes con hepatoblastoma y osteosarcoma.

Respecto a los antecedentes personales patológicos, solo 9 niños refirieron exantemáticas como varicela (8 casos) y sarampión (1 caso); 1 paciente reportó alergia al Trimetropim/Sulfamentoazol y solo un caso refirió hepatitis a los 6 años. Este último caso correlaciona con el diagnóstico de carcinoma hepatocelular.

Para conocer el estado actual de los pacientes, respuesta al tratamiento y supervivencia global, se propuso realizar la actualización de la información a través de la revisión de expedientes clínicos mediante visitas al Hospital Infantil de México “Federico Gómez” y el Instituto Nacional de Pediatría.

Los datos actualizados demuestran una supervivencia global promedio de 61 meses (1 – 145 meses), con 10 defunciones documentadas, principalmente en diagnósticos como osteosarcomas, hepatoblastomas y una niña con síndrome de Down y leucemia linfoblástica aguda. Hasta el momento de la revisión de expedientes, la mayoría de los pacientes siguen en vigilancia o no existen más registros que documenten su defunción. Con los datos actualizados de los pacientes con registros de vigilancia o defunción, la supervivencia global >5 años es del 70% (figura 6).

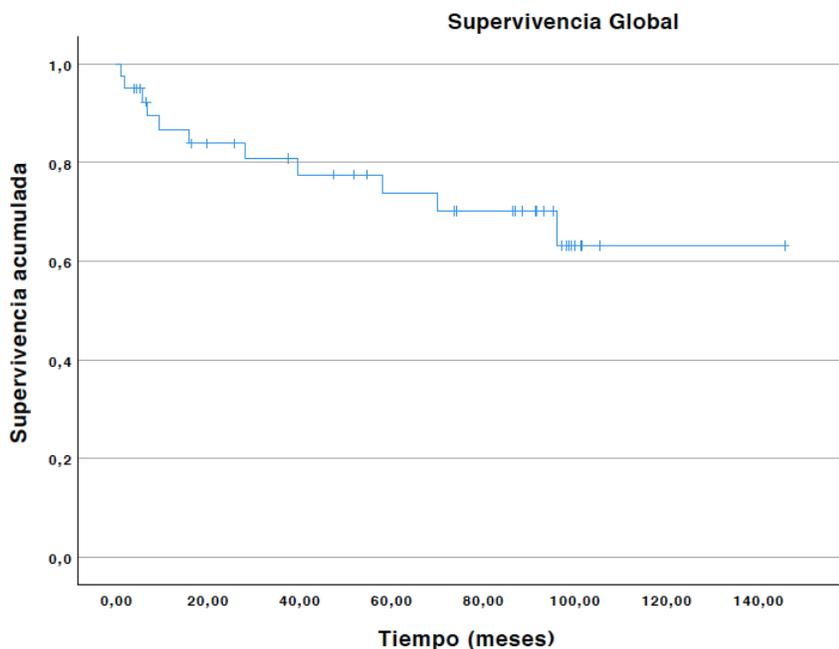


Figura 6. Gráfica de Kaplan-Meier de supervivencia global del grupo de pacientes secuenciados.

7.2 Elaboración de familiogramas y aplicación del cuestionario de riesgo.

7.2.1 Historia familiar de cáncer

Uno de los principales criterios de sospecha a cáncer hereditario es la historia familiar en individuos <45 años, sin embargo, los datos de Zhang y colaboradores demostraron que el 60% de los niños portadores de mutaciones germinales carecían de historia familiar de cáncer a cualquier edad. Por tal motivo, en este estudio piloto se evaluó la historia familiar en los individuos secuenciados, la cual, fue considerada como positiva si cumplían >1 criterio. Con el objetivo de precisar criterios de sospecha a cáncer hereditario, los 24 pacientes con al menos un familiar afectado fueron segmentados en 3 grupos de edad al diagnóstico: infancia y adolescencia, adultos jóvenes y mayores de 45 años (tabla 7).

Tabla 7. Historia familiar de cancer (n=24)

Criterio	Frecuencia	%
Familiares con cáncer antes de los 18 años	4	16.0
Familiares con cáncer entre 18-45 años	10	40.0
Familiares con cáncer despues de los 45 años	7	28.0
>2 familiares de primer o segundo grado del mismo lado de la familia con diagnóstico de cáncer antes de los 45 años	1	4.0
Algun familiar con >2 tumores	1	4.0
Diagnóstico o edad no confirmada	2	8.0

Pese a la edad al diagnóstico, el grupo de pacientes analizados reportaron una elevada historia familiar de cáncer, donde el 56% de los casos se relaciona con diagnóstico antes de los 45 años. Además, 2 pacientes presentaron un segundo tumor primario, una paciente con fibrosarcoma ovárico y tumor contralateral de células de la granulosa y un paciente con Linfoma de Hodgkin y sarcoma de Ewing (Ver Familiogramas en el anexo 3). Estos datos demuestran un porcentaje elevado de pacientes con historia familiar de cáncer, que, hasta el momento, no ha sido contemplada como criterio de sospecha de individuos con síndromes de predisposición subyacente.

7.2.1 Evaluación de cuestionario de riesgo a cáncer pediátrico.

Los criterios de sospecha a cáncer hereditario en adultos están definidos para algunos tipos de tumor; sin embargo, no son específicos para población pediátrica. La definición de criterios a través de un cuestionario de riesgo puede ayudar a mejorar la identificación de niños con cáncer que puedan referirse a consultoría genética y subsecuente análisis molecular. Hasta el momento, un cuestionario realizado a través de una revisión sistemática ha sido propuesto para evaluación en cohortes independientes y poblaciones diversas, por tal motivo, decidimos utilizar la herramienta propuesta por Jongmans Marjolin y colaboradores para evaluar su desempeño (tabla 8). El 37.5% de los pacientes (15/40) cumplió con al menos un criterio de sospecha, siendo un tumor en edad adulta en algún familiar el más frecuente. Tres individuos tienen al menos un familiar de primer grado con diagnóstico de cáncer antes de los 45 años, 2 individuos presentaron 2 tumores no asociados al tratamiento y sólo una paciente presentó anomalías congénitas (síndrome de Down). De todos los pacientes con algún criterio de sospecha, el 92% solo presentó uno y 8% 2 criterios.

Tabla 8. Criterios de sospecha a cáncer hereditario (Eur J Med Genet . 2016 Mar;59(3):116-25)

Criterios de sospecha (n = 40)	Frecuencia
1 Historia familiar de un niño con cáncer	3
a. ≥ 2 tumores en la infancia	
b. Un familiar de primer grado (padres o hermanos) con cáncer antes de los 45 años	3
c. ≥ 2 familiares de 2 grado con cáncer <45 años del mismo lado familiar	
d. Consanguinidad de los padres biológicos	
2 Un familiar con alguno de los siguientes tumores infantiles:	12
Anomalías congénitas	
Dismorfismos faciales	
Discapacidad intelectual	
Crecimiento aberrante	
Anomalías cutáneas	
Enfermedades hematológicas	
Deficiencia inmune	
-LMMJ	2
-LAL hipodiploide baja	
-Tumor de la vaina del nervio periférico maligno	
-Carcinoma medular de tiroides	
-Meduloblastoma	
-Glioma óptico	
-Tumor de células de Leyding-Sertoli ovárico	
-Blastoma pleuropulmonar	
-Blastoma pituitario	
-Pineoblastoma	
-Retinoblastoma	
-Schwannoma	
-Tumor subependimario de células gigantes	
O, un tumor en la edad adulta (colorrectal, mama, ovario, carcinoma de células basales, etc.)	10
3 Un niño con dos tumores antes de los 18 años (Excepto tumores secundarios al tratamiento consistentes con el tiempo y tipo de tejido)	2
4 Un niño con cáncer, enfermedades congénitas u otros síntomas específicos	1
Anomalías congénitas	1
Dismorfismos faciales	
Discapacidad intelectual	
Crecimiento aberrante	
Anomalías cutáneas	
Enfermedades hematológicas	
Deficiencia inmune	
5 Un niño con toxicidad excesiva al tratamiento	No evaluada

Como objetivo secundario y con base en los resultados de secuenciación (estándar de oro), se decidió evaluar los valores de predicción, sensibilidad y especificidad (Figura 7). Los resultados demostraron una sensibilidad del 100% y un porcentaje de verdaderos negativos del 100%. En contraste, se obtuvo un valor predictivo positivo del 40% y una especificidad del 75%, lo que puede deberse a la baja prevalencia de mutaciones, además que, en algunos individuos no pudo confirmarse el diagnóstico específico del tumor o, pudieran existir algunas mutaciones aún no clasificadas como patogénicas o en regiones no analizadas (i.e., regiones intrónicas profundas). Aunque estos resultados son prometedores, se requiere la aplicación en más individuos (incluyendo más fenotipos y adultos jóvenes) para confirmar su utilidad en la identificación de niños con cáncer con altas posibilidades de padecer un síndrome de predisposición subyacente, los cuales, puedan ser referidos por el personal de primer contacto para evaluación más detallada con el médico genetista.

		Estándar de oro (Secuenciación)				
		+	-			
Resultado cuestionario	+	6	9	15	VPP	40%
	-	0	25	25	VPN	100%
		6	34	40		
		Sensibilidad		Especificidad		
		100%		74%		

Figura 7. Cálculo de valores de predicción, sensibilidad y especificidad. **VPP (Valor predictivo positivo)**: Probabilidad de que individuos positivos en el cuestionario sean portadores de mutaciones germinales. **Especificidad: (Tasa de negatividad)**. Probabilidad del cuestionario de identificar a los individuos sin mutaciones germinales. **VPN (Valor predictivo negativo)**: Probabilidad de que individuos negativos en el cuestionario no porten alguna mutación germinal. **Sensibilidad: (Tasa de positividad)**: Probabilidad del cuestionario de identificar a todos los individuos mutados.

7.3 Estandarización de toma de muestra de saliva y extracción y purificación de ADN genómico (ADNg)

Fuentes alternativas de ADNg mediante tomas de muestra sencillas, accesibles y no invasivas, son necesarias para estudios de epidemiología genética. La saliva es un biofluido producido por las células acinares mediante ultrafiltración del plasma, una vez liberado hacia la cavidad bucal, es mezclada con células epiteliales de descamación y células inmunes, líquido crevicular gingival, detritus alimenticio y microorganismos. Cada vez es más utilizada como fuente de ADNg en diferentes estudios genéticos, principalmente en genotipado de SNPs e identificación de mutaciones a través de secuenciación dirigida.

Con el objetivo de demostrar el desempeño de la saliva en la detección de mutaciones germinales a través de secuenciación de exomas humanos, realizamos una revisión

bibliográfica a través de artículos, manuales, guías y buenas practicas que empleaban saliva para el descubrimiento de biomarcadores de enfermedades locales y sistémicas. Variables como: saliva no estimulada vs estimulada, estabilidad de la muestra, fuentes de contaminación y uso de agentes estabilizadores fueron contempladas para la estandarización y desarrollo del protocolo de toma de muestra.

Se realizó un análisis comparativo de rendimiento entre 3 tipos de toma de muestra en 2 familias (6 pacientes, niño con cáncer, padre y madre), a través de:

1. Toma de muestra sin agente estabilizador en tubo tipo Falcon en hielo
2. Toma de muestra con el kit Oragene DNA (*DNA Genotek*)
3. Toma de muestra con el kit SalivaGene Collection (*Stratec*)

Las tomas se realizaron el mismo día y bajo las mismas condiciones, para las cuales, se solicitó a los participantes recolectar 10 mL de saliva en tubo Falcon y 5 mL en cada dispositivo. Los resultados se muestran a continuación.

7.3.1 Rendimientos de concentración y pureza

Se realizó extracción y purificación de ADNg en paralelo para cada condición, siguiendo el protocolo descrito en la metodología. Los resultados de concentración de dsDNA determinados por NanoDrop y Qubit, así como la razón 260/280 se presentan en la tabla 9. El promedio de concentración determinada tanto por NanoDrop como fluorimetría fue superior con Oragene comparado con las otras metodologías. Para confirmar si existen diferencias significativas en las concentraciones obtenidas en cada individuo y toma de muestra, se realizó un análisis de varianza (ANOVA), la cual, no demostró variabilidad estadísticamente significativa entre las concentraciones ($p=0.26$ para concentraciones obtenidas con Nanodrop y $p=0.09$ para concentraciones obtenidas con Qubit) y razones de pureza ($p=0.25$).

Tabla 9. Concentraciones de ADNg obtenido a partir de 3 diferentes tomas de muestra y análisis de varianza

Familia	ID	Familiar	Características generales		NanoDrop		Qubit
			Tipo de toma de muestra	[ng/μl]	260/280	[ng/μl]	
275	103275001	Px	Tubo falcon sin estabilizador	8.2	1.93	0.46	
			Oragene	17.1	1.98	5.55	
			Stratec	10.5	2.02	1.57	
	103275002	Mamá	Tubo falcon sin estabilizador	7.8	2.16	1.27	
			Oragene	8.1	1.96	1.69	
			Stratec	9.9	2.13	0.982	
	103275003	Papá	Tubo falcon sin estabilizador	69.1	1.82	15.4	
			Oragene	194.6	1.82	43.5	
			Stratec	6.2	2.04	2.95	
276	103276001	Px	Tubo falcon sin estabilizador	1.6	0.92	0.09	
			Oragene	26.8	1.97	9.87	
			Stratec	9.8	2.32	1.06	
	103276002	Mamá	Tubo falcon sin estabilizador	21.5	1.89	2.74	
			Oragene	38.6	2	14.4	

		Stratec	12.2	2.24	3.45
103276003	Papá	Tubo falcon sin estabilizador	17.3	2.25	2.71
		Oragene	34.8	1.95	22.1
		Stratec	32.4	1.94	18.4

ANÁLISIS DE VARIANZA (Concentración determinada con Nanodrop)

<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
Entre grupos	5385.083333	2	2692.541667	1.43885944	0.26812892	3.682320344
Dentro de los grupos	28069.54167	15	1871.302778			
Total	33454.625	17				

ANÁLISIS DE VARIANZA (Concentración determinada con Qubit)

<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
Entre grupos	571.8721871	2	285.9360936	2.776550921	0.09421155	3.682320344
Dentro de los grupos	1544.737167	15	102.9824778			
Total	2116.609354	17				

ANÁLISIS DE VARIANZA (Pureza 260/280)

<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
Entre grupos	0.249033333	2	0.124516667	1.493045378	0.25624688	3.682320344
Dentro de los grupos	1.250966667	15	0.083397778			
Total	1.5	17				

7.3.2 Geles de integridad

La visualización de las bandas de ADNg fue realizada mediante geles de agarosa al 1% siguiendo el protocolo descrito en la metodología. Inicialmente, en cada pozo se incluyeron 2 mL de muestra sin concentración estándar de ADNg, el grosor de las bandas visualizadas fue directamente proporcional su concentración, además, todas las bandas incluían fragmentos de más de 10K pb. (Figura 8a y b).

Para mejorar la visualización de bandas en subsecuentes muestras y mitigar el barrido de fondo, se realizó una curva que incluyó concentraciones de 25, 50, 100, 150 y 200 ng de ADNg total. Las bandas fueron visibles a partir de 50 ng totales bajo las condiciones de tinción con GelRed, por lo que se decide en subsecuentes muestras, estandarizar a dicha concentración. (Figura 8c).

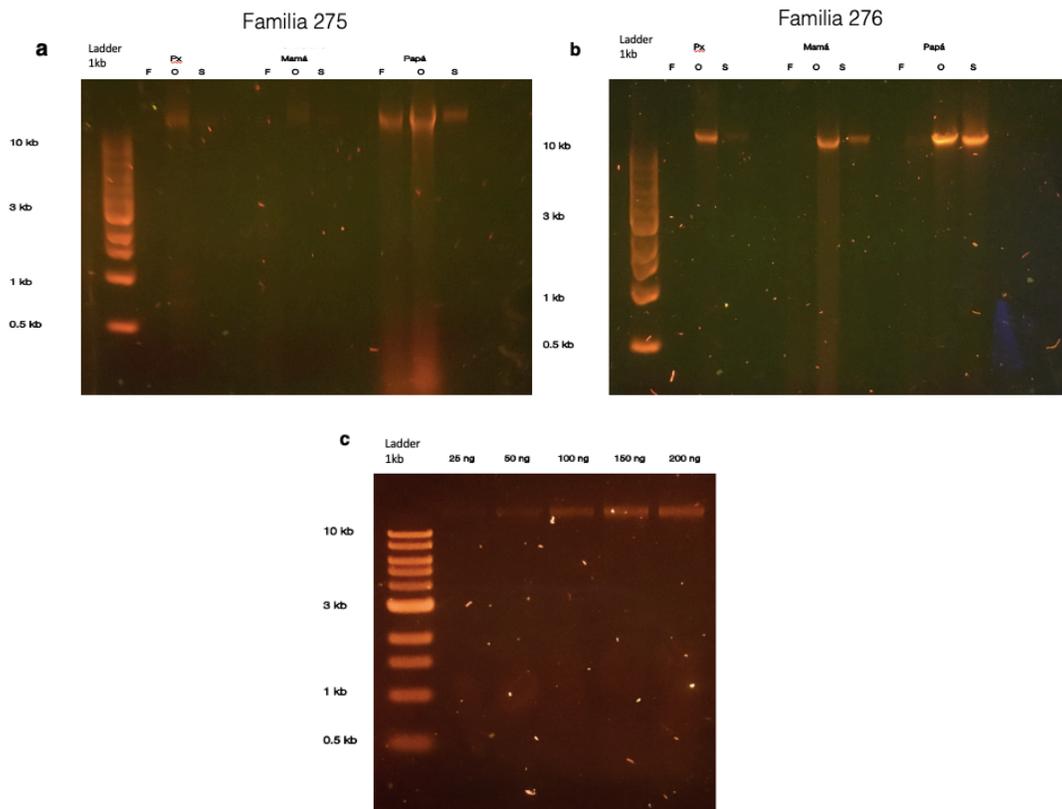


Figura 8. Análisis cualitativo de ADNg. Los paneles a y b muestran las bandas de ADNg purificado a partir de tres diferentes tipos de toma de muestra. A partir de 50 ng totales de ADNg total, se puede visualizar la banda de más de 10k pb. (F, tubo Falcon; O, Oragene; S, Stratec)

7.4 Secuenciación de siguiente generación (NGS) de exomas de pacientes con cáncer y progenitores

[La secuenciación se realizó en 2 centros, la primera parte fue secuenciada en el Broad Institute \(MIT-Harvard\) y la segunda parte en el Instituto Nacional de Medicina Genómica, INMEGEN-México. Los resultados de las métricas se muestran a continuación.](#)

7.5 Estandarización del flujo de análisis bioinformático

Los datos obtenidos a partir de la secuenciación de exomas fueron analizados con las buenas prácticas del GATK descritas en la metodología. Todos los comandos utilizados para cada fase se pueden consultar en el anexo 1 y los resultados obtenidos para cada fase se muestran a continuación.

7.5.1 Fase 1. Pre-procesamiento de datos

El archivo de salida .txt contiene diferentes parámetros de métricas de calidad, como: total de lecturas, número de lecturas no identificadas como duplicados, fracción de bases alineadas al genoma de referencia, fracción de bases *on y off target*, promedio de cobertura, y fracción de bases a diferentes profundidades de secuenciación (desde 1X hasta 100X). Los resultados respecto a la fracción de bases cubiertas a 20X para cada uno de los 78 archivos BAM, se presentan en la tabla 10.

Tabla 10. Fracción de bases cubiertas a 20X. Los datos se presentan en orden descendente.

SAMPLE_ID	20X	SAMPLE_ID	20X
1030010001_sorted_MD.RG.IND.RC_hs_metrics.txt	0.93	520000000346_sorted_MD.RG.IND.RC_hs_metrics.txt	0.88
1030010002_sorted_MD.RG.IND.RC_hs_metrics.txt	0.93	521010310213_sorted_MD.RG.IND.RC_hs_metrics.txt	0.70
520000000395_sorted_MD.RG.IND.RC_hs_metrics.txt	0.93	521010310153_sorted_MD.RG.IND.RC_hs_metrics.txt	0.70
521010310292_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77	521010310073-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.69
521010310328_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76	521010310149_sorted_MD.RG.IND.RC_hs_metrics.txt	0.70
521010310120_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76	521010310086-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.69
520000000394_sorted_MD.RG.IND.RC_hs_metrics.txt	0.92	521010310112_sorted_MD.RG.IND.RC_hs_metrics.txt	0.69
521010310254_sorted_MD.RG.IND.RC_hs_metrics.txt	0.75	521010310063-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.69
521010310302_sorted_MD.RG.IND.RC_hs_metrics.txt	0.75	521010310085-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.68
521010310284_sorted_MD.RG.IND.RC_hs_metrics.txt	0.75	521010310168_sorted_MD.RG.IND.RC_hs_metrics.txt	0.69
521010310314_sorted_MD.RG.IND.RC_hs_metrics.txt	0.75	521010310167_sorted_MD.RG.IND.RC_hs_metrics.txt	0.69
521010310249_sorted_MD.RG.IND.RC_hs_metrics.txt	0.74	521100310123-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.68
1030011002_sorted_MD.RG.IND.RC_hs_metrics.txt	0.91	521010310152_sorted_MD.RG.IND.RC_hs_metrics.txt	0.68
521010310033_sorted_MD.RG.IND.RC_hs_metrics.txt	0.74	521010310092-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.68
1030011003_sorted_MD.RG.IND.RC_hs_metrics.txt	0.91	521010310175_sorted_MD.RG.IND.RC_hs_metrics.txt	0.68
521010310037_sorted_MD.RG.IND.RC_hs_metrics.txt	0.74	521010310166_sorted_MD.RG.IND.RC_hs_metrics.txt	0.68
521010310184_sorted_MD.RG.IND.RC_hs_metrics.txt	0.73	521010310089-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.68
521010310185_sorted_MD.RG.IND.RC_hs_metrics.txt	0.73	521010310165_sorted_MD.RG.IND.RC_hs_metrics.txt	0.68
1030011001_sorted_MD.RG.IND.RC_hs_metrics.txt	0.90	521010310164_sorted_MD.RG.IND.RC_hs_metrics.txt	0.68
521010310283_sorted_MD.RG.IND.RC_hs_metrics.txt	0.73	521010310106-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.67
521010320129_sorted_MD.RG.IND.RC_hs_metrics.txt	0.73	521010310105-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.67
521010310076-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.72	521010310151_sorted_MD.RG.IND.RC_hs_metrics.txt	0.67
521010310066-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.72	521010310090-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.66
521010310084-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.72	521010310161_sorted_MD.RG.IND.RC_hs_metrics.txt	0.67
521010310219_sorted_MD.RG.IND.RC_hs_metrics.txt	0.71	521010310162_sorted_MD.RG.IND.RC_hs_metrics.txt	0.66
521010310214_sorted_MD.RG.IND.RC_hs_metrics.txt	0.72	521010310150_sorted_MD.RG.IND.RC_hs_metrics.txt	0.65
521010320218_sorted_MD.RG.IND.RC_hs_metrics.txt	0.71	521010310099-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.65
521010310048-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.71	521010320114-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.65
521010310172_sorted_MD.RG.IND.RC_hs_metrics.txt	0.71	521010310203_sorted_MD.RG.IND.RC_hs_metrics.txt	0.65
521010310220_sorted_MD.RG.IND.RC_hs_metrics.txt	0.71	521010310171_sorted_MD.RG.IND.RC_hs_metrics.txt	0.64
521010320076_sorted_MD.RG.IND.RC_hs_metrics.txt	0.71	521010310091-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.64
521010310051-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.71	521010310143-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.59
521010310047_sorted_MD.RG.IND.RC_hs_metrics.txt	0.71	521010310126-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.58
521010310103-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.70	521010310107-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.56
521010310209_sorted_MD.RG.IND.RC_hs_metrics.txt	0.70	521010310137-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.56
521010310069-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.70	521010310119-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.53
521010310222_sorted_MD.RG.IND.RC_hs_metrics.txt	0.70	521010310215_sorted_MD.RG.IND.RC_hs_metrics.txt	0.14
521010310078-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.70	521010310216_sorted_MD.RG.IND.RC_hs_metrics.txt	0.13
521010310053-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.70	521010310217_sorted_MD.RG.IND.RC_hs_metrics.txt	0.09

Con base en los resultados obtenidos, se observó que sólo 8 archivos presentaron un porcentaje >80% de las bases secuenciadas a 20X. Con el objetivo de excluir la menor cantidad de archivos, se realizó una evaluación de porcentaje de bases cubiertas a 20X específicamente en genes de interés primaria. Por tal motivo, se realizó una revisión bibliográfica para definir y seleccionar genes con relevancia clínica con cáncer pediátrico. El último reporte de genes de predisposición a cáncer pediátrico desarrollado por investigadores del *St Jude Children's Hospital*, seleccionó a 565 genes para análisis con base

en una revisión del *American College of Medical Genetics (ACMG)*, de los cuales 60 genes corresponden a genes de predisposición autosómico dominante y 29 autosómico recesivo, 23 genes de tirosina-cinasa, 58 genes supresores de tumores y otros 365 genes (34).

Además, se contemplaron 2 paneles clínicos de cáncer pediátrico, Blueprint Genetics (71 genes) y un panel de UCSF (198 genes). Con el objetivo de seleccionar los genes más representativos entre los paneles, se realizó un diagrama de Venn a partir del cual, se seleccionaron un total de 572 genes, incluyendo los 565 genes del estudio de mutaciones germinales en cáncer pediátrico, 7 genes adicionales del Blueprint y un gen del panel UCSF. Los genes excluidos en el archivo de intervalos fueron debido a las implicaciones somáticas más que germinales. (Figura 9) Los genes seleccionados fueron incluidos en un nuevo archivo de intervalos utilizado para la generación de nuevas métricas.

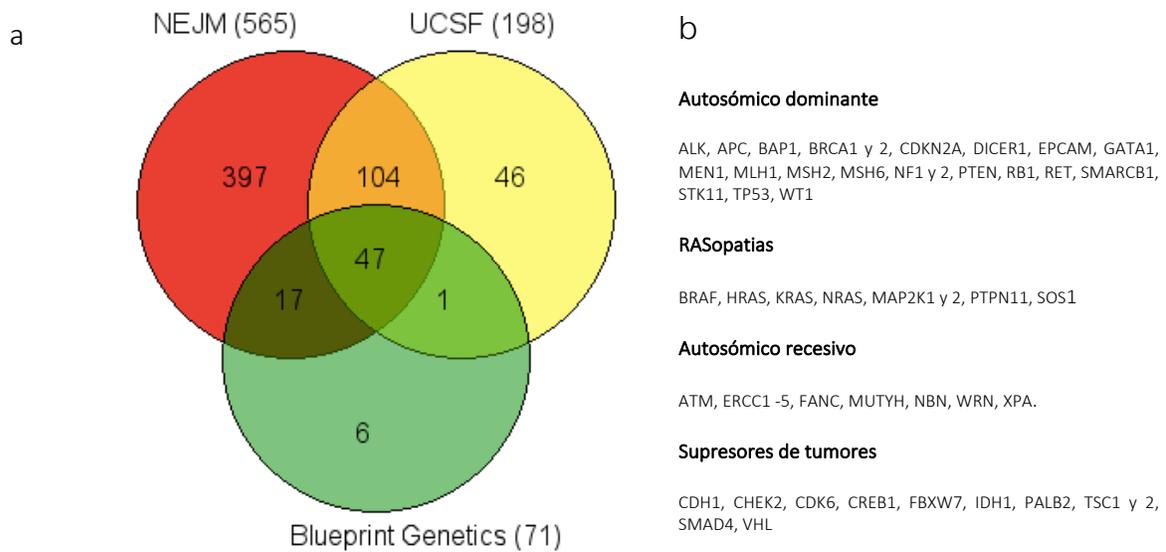


Figura 9. Número de genes por panel, intersección y disyunción. En el panel b se enlistan algunos genes representativos

Con 572 genes seleccionados, se generó un nuevo archivo .bed con la herramienta *bedtools v2*, para el cual se incluyeron 3 campos: chrom, start y end. El nuevo archivo bed fue convertido en archivo de intervalos (*intervals_for_qc_large_list*) con la opción *BedToInterval* (*Picard tools*). Se realizó la recolección de métricas con el archivo de intervalos de genes de interés primario, los resultados se muestran en la tabla 11.

Tabla 11. Fracción de bases cubiertas a 20X determinada a través del nuevo archivo de intervalos

SAMPLE_ID	20X	SAMPLE_ID	20X
1030010001_sorted_MD.RG.IND.RC_hs_metrics.txt	0.85	520000000346_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77
1030010002_sorted_MD.RG.IND.RC_hs_metrics.txt	0.85	521010310213_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77
520000000395_sorted_MD.RG.IND.RC_hs_metrics.txt	0.84	521010310153_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77
521010310292_sorted_MD.RG.IND.RC_hs_metrics.txt	0.84	521010310073-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77
521010310328_sorted_MD.RG.IND.RC_hs_metrics.txt	0.84	521010310149_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77
521010310120_sorted_MD.RG.IND.RC_hs_metrics.txt	0.83	521010310086-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77
520000000394_sorted_MD.RG.IND.RC_hs_metrics.txt	0.83	521010310112_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77
521010310254_sorted_MD.RG.IND.RC_hs_metrics.txt	0.83	521010310063-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77
521010310302_sorted_MD.RG.IND.RC_hs_metrics.txt	0.83	521010310085-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.77
521010310284_sorted_MD.RG.IND.RC_hs_metrics.txt	0.82	521010310168_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76

521010310314_sorted_MD.RG.IND.RC_hs_metrics.txt	0.82	521010310167_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76
521010310249_sorted_MD.RG.IND.RC_hs_metrics.txt	0.82	521100310123-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76
1030011002_sorted_MD.RG.IND.RC_hs_metrics.txt	0.82	521010310152_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76
521010310033_sorted_MD.RG.IND.RC_hs_metrics.txt	0.82	521010310092-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76
1030011003_sorted_MD.RG.IND.RC_hs_metrics.txt	0.82	521010310175_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76
521010310037_sorted_MD.RG.IND.RC_hs_metrics.txt	0.81	521010310166_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76
521010310184_sorted_MD.RG.IND.RC_hs_metrics.txt	0.81	521010310089-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.76
521010310185_sorted_MD.RG.IND.RC_hs_metrics.txt	0.80	521010310165_sorted_MD.RG.IND.RC_hs_metrics.txt	0.75
1030011001_sorted_MD.RG.IND.RC_hs_metrics.txt	0.80	521010310164_sorted_MD.RG.IND.RC_hs_metrics.txt	0.75
521010310283_sorted_MD.RG.IND.RC_hs_metrics.txt	0.80	521010310106-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.75
521010320129_sorted_MD.RG.IND.RC_hs_metrics.txt	0.80	521010310105-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.74
521010310076-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.80	521010310151_sorted_MD.RG.IND.RC_hs_metrics.txt	0.74
521010310066-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.80	521010310090-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.74
521010310084-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.79	521010310161_sorted_MD.RG.IND.RC_hs_metrics.txt	0.74
521010310219_sorted_MD.RG.IND.RC_hs_metrics.txt	0.79	521010310162_sorted_MD.RG.IND.RC_hs_metrics.txt	0.74
521010310214_sorted_MD.RG.IND.RC_hs_metrics.txt	0.79	521010310150_sorted_MD.RG.IND.RC_hs_metrics.txt	0.73
521010320218_sorted_MD.RG.IND.RC_hs_metrics.txt	0.79	521010310099-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.73
521010310048-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.79	521010320114-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.72
521010310172_sorted_MD.RG.IND.RC_hs_metrics.txt	0.79	521010310203_sorted_MD.RG.IND.RC_hs_metrics.txt	0.72
521010310220_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310171_sorted_MD.RG.IND.RC_hs_metrics.txt	0.72
521010320076_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310091-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.72
521010310051-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310143-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.67
521010310047_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310126-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.66
521010310103-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310107-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.63
521010310209_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310137-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.63
521010310069-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310119-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.60
521010310222_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310215_sorted_MD.RG.IND.RC_hs_metrics.txt	0.17
521010310078-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310216_sorted_MD.RG.IND.RC_hs_metrics.txt	0.15
521010310053-PB_sorted_MD.RG.IND.RC_hs_metrics.txt	0.78	521010310217_sorted_MD.RG.IND.RC_hs_metrics.txt	0.11

Con el nuevo archivo de intervalos, se observó un aumento de 8 a 23 archivos BAM con un 80% de bases a 20X; 70 archivos con 80% de bases a 10X y 2 con 79% de bases a 20X. Algunos factores que influyen en el porcentaje de lecturas on target y su uniformidad son: kit para preparación de librerías, desarrollo metodológico, número de muestras secuenciadas en paralelo, longitud de lecturas, complejidad del genoma (regiones homopoliméricas, pseudogenes) y tasas de error del secuenciador, principalmente. El kit utilizado para secuenciar las muestras (SureSelect v4) asegura un ~75% on target y uniformidad de 75% a 20X, el cual, se alcanzó en el 80% de los archivos BAM.

Tomando en cuenta los anteriores factores, la distribución de lecturas de manera aleatoria en las regiones secuenciadas (la fracción de bases a 20X es un promedio, mas no es uniforme), los puntajes de calidad y mapeo de las bases; el marcado de lecturas duplicadas, realineamiento de INDELS y recalibración de bases, así como los criterios utilizados en proyectos genómicos de línea germinal que no demandan una alta profundidad de secuenciación (70-73), se decide incluir 75 archivos con >79% de las bases a 10X.

7.5.1.1 Histogramas de cobertura (Picard tools)

Aunado a las métricas previamente obtenidas, se generaron histogramas (análisis cualitativo) para visualizar la profundidad de secuenciación y fracción de bases con la opción CoverageBed (*bedtools v2.27*) y *Rstudio*.

Durante este proceso, nos enfrentamos a varios problemas, ya que, los histogramas obtenidos no eran los esperados (Figura 10a), se establecieron causas raíz de problema como capacidad limitada del servidor (consumo de toda la memoria RAM), problemas de sintaxis con el comando para generar gráficos en R o problemas con pasos inmediatos. Para cada

causa raíz, se estableció una estrategia. Se corrió un control (muestra NA12891.exome.bam del proyecto 1KG) y se concluyó que el problema se debía a que en archivos de gran tamaño como los analizados, se debe hacer un pre-ordenado por cromosoma y posición de inicio, lo cual, promueve el uso de un algoritmo de eficiencia de memoria.

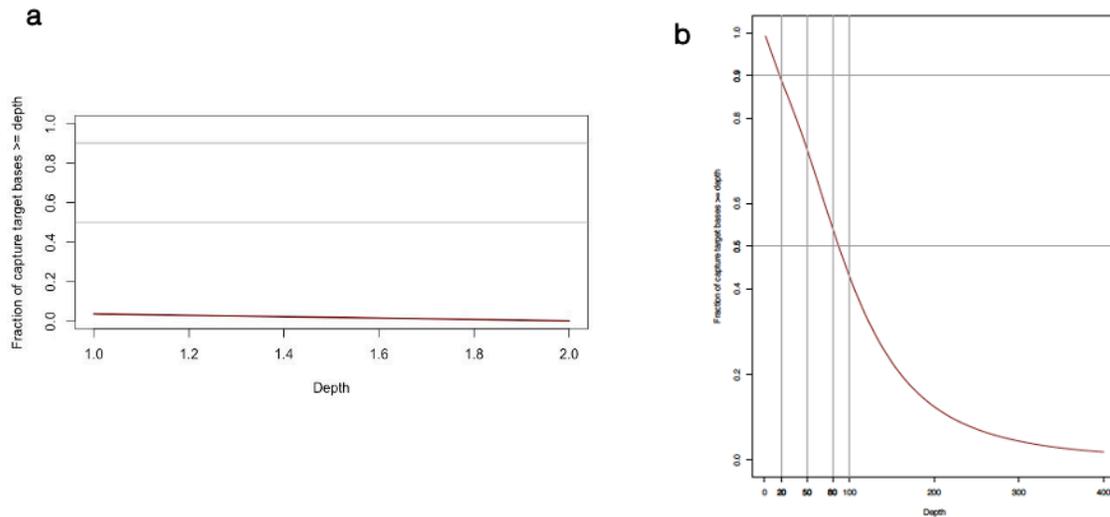


Figura 10. Histogramas de cobertura. El panel a muestra el primer histograma de la muestra NA12891.exome.bam obtenido con los comandos usados con los datos en proceso de análisis, mostrando un patrón similar no esperado. Posterior al ordenado por cromosoma, se mejora el consumo de memoria y se logra el histograma esperado (panel b).

Una vez estandarizados los comandos, se realizó el histograma de cobertura para cada archivo BAM.

7.5.1.2 Comparación de métricas de secuenciación e histogramas de cobertura entre muestras de sangre y saliva

Uno de los principales objetivos del proyecto es determinar si el uso de saliva es una alternativa viable para el análisis por secuenciación de exomas. Los criterios preanalíticos demostraron un adecuado rendimiento y calidad de ADNg purificado de saliva, sin embargo, decidimos evaluar el porcentaje de bases alineadas al genoma de referencia y fracción de bases a 20X entre muestras de sangre y saliva (Figura 11).

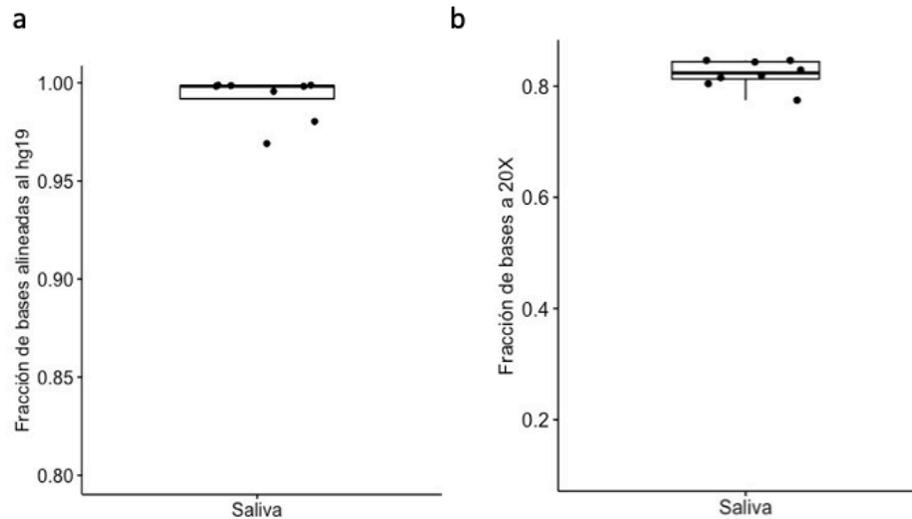


Figura 11. Métricas de secuenciación de muestras de saliva. Se realizaron gráficas de cajas y bigotes para visualizar el desempeño de secuenciación a partir de ADN_g purificado de muestras de saliva. Se observa que la fracción de lecturas alineadas al genoma de referencia hg19 (a) es mayor al 95%, mientras que la fracción de bases a 20X (b) es mayor al 77%

Los resultados de métricas de secuenciación entre muestras de saliva y sangre, así como sus respectivos histogramas de cobertura, demuestran una similitud entre la fracción de bases a la mínima profundidad de secuenciación (20X). Además, la fracción de lecturas alineadas al genoma de referencia en las muestras de saliva, confirma la baja contaminación con ADN bacteriano lograda con el kit Oragene DNA, aunque el número de muestras secuenciadas a partir de tomas de saliva es limitado.

7.6 Fase 2. Descubrimiento de variantes

75 archivos BAM realineados y recalibrados fueron incluidos para subsecuente descubrimiento de variantes. Las buenas prácticas del GATK para el descubrimiento de SNPs y INDELS germinales, recomiendan realizar el descubrimiento de variantes en varias muestras a la vez (*joint genotyping*), mejorando la detección de variantes con una mayor sensibilidad y precisión posible. Esta fase involucra 5 pasos principales descritos a continuación.

En este proceso, se realizaron varios intentos debido a que, en los primeros gráficos de distribución de variantes, no eran tan evidentes las agrupaciones de variantes de alta calidad, existiendo un predominio por resultados falsos positivos; además, los gráficos de tramos no incluían verdaderos positivos y el Ti/Tv fue mucho menor al esperado para exomas (Figura 11^a y b). Algunas probables causas raíz establecidas fueron errores de sintaxis, específicamente en los parámetros de los archivos de referencia, baja profundidad de secuenciación en algunos archivos BAM, error en pasos anteriores (como la unificación de archivos .g.vcf) o la variabilidad propia de la población de estudio (Mexicana). Para cada una de las probables causa raíz, se estableció una estrategia y pruebas, resaltando la inclusión de

únicamente 30 archivos BAM con mayor fracción de bases a 20X, obteniendo resultados similares.

Hasta el momento, la estrategia que contribuyó a obtener mejores agrupaciones de variantes y visualización de SNPs verdaderas positivas, fue el uso adecuado de los intervalos durante el genotipado de VCFs, en el cual, se había integrado todas las regiones por cromosoma (L1, L2 ...), por lo que este paso fue actualizado con el archivo de intervalos de genes de interés primario (intervals_for_qc_large_list). Los resultados hasta el momento, se presentan en la figura 12c y d

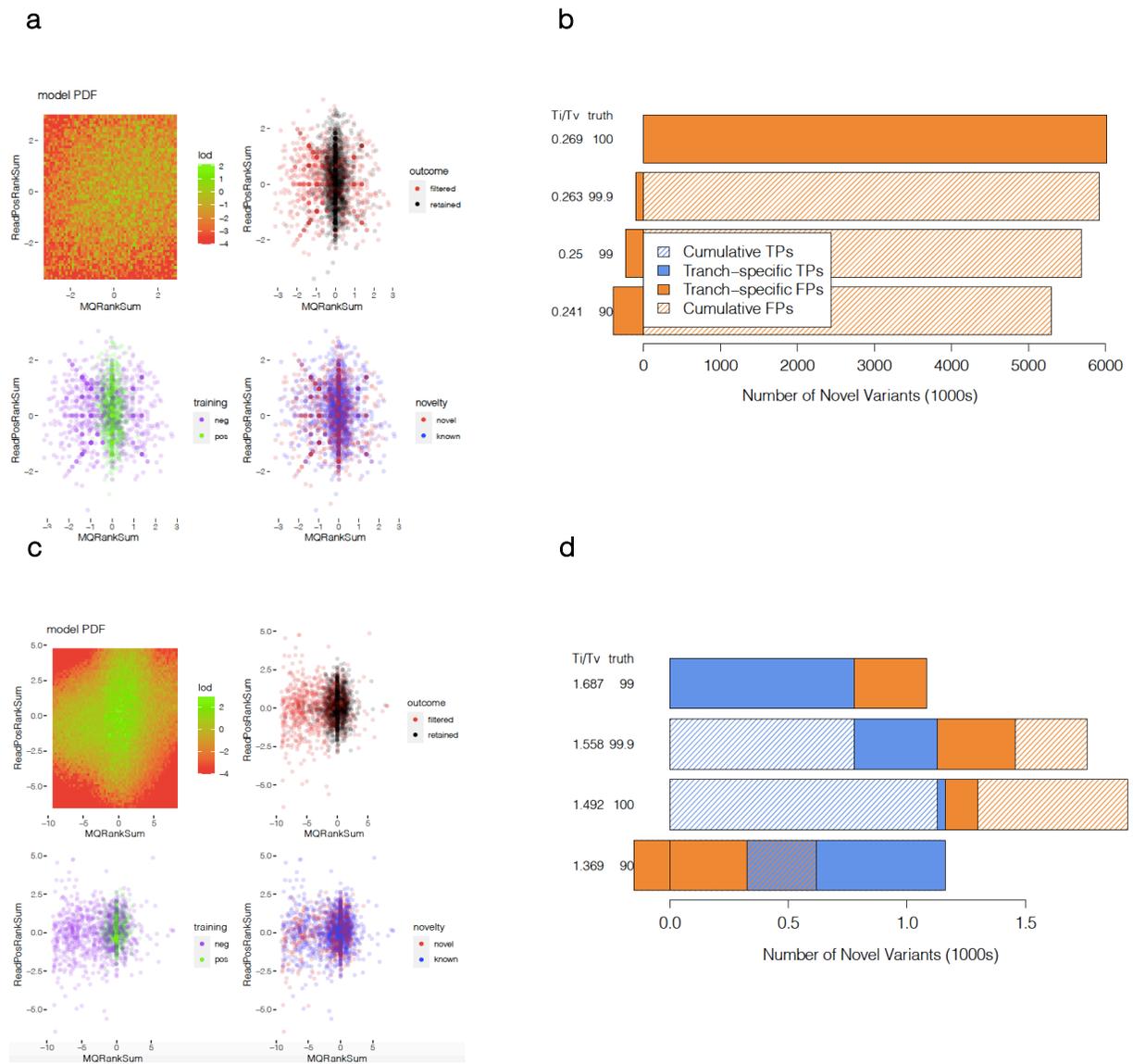


Figura 12. Gráficos de agrupación de variantes y tramos. En el panel a y b se muestran los gráficos obtenidos en el primer intento, en los cuales, se observa un predominio de variantes con bajos puntajes de calidad y alta probabilidad de falsos positivos. Tras el reproceso del genotipado con el archivo de intervalos adecuado, se observa una agrupación de variantes de alta calidad en los rangos de los filtros de anotación correspondientes (panel c), además, el gráfico de tramos ya se visualizan variantes verdaderas positivas (panel d).

7.7 Fase 3. Anotación y filtrado de variantes

El archivo de salida *Piloto_Dic2020-recalibratedSNPINDEL.vcf*, corresponde al archivo unificado y recalibrado de variantes de los 75 individuos en proceso de análisis en modo *Joint Genotyping*. La anotación y filtrado de variantes es la fase en la cual, a cada variante contenida en un archivo *vcf*, le es asignada información biológica, funcional, poblacional y clínica. Dicha información es indispensable para poder realizar subsecuente filtrado e identificación de variantes de relevancia clínica y asociación con fenotipo. El archivo de salida, mencionado previamente, es utilizado como archivo de entrada para la fase 3.

Cada archivo *vcf* individual, fue utilizado como archivo de entrada para anotación de variantes con el programa *Variant Effect Predictor, VEP (Ensamble)* (74). La estadística del número y tipo de variantes para la población de estudio se presenta en la figura 13

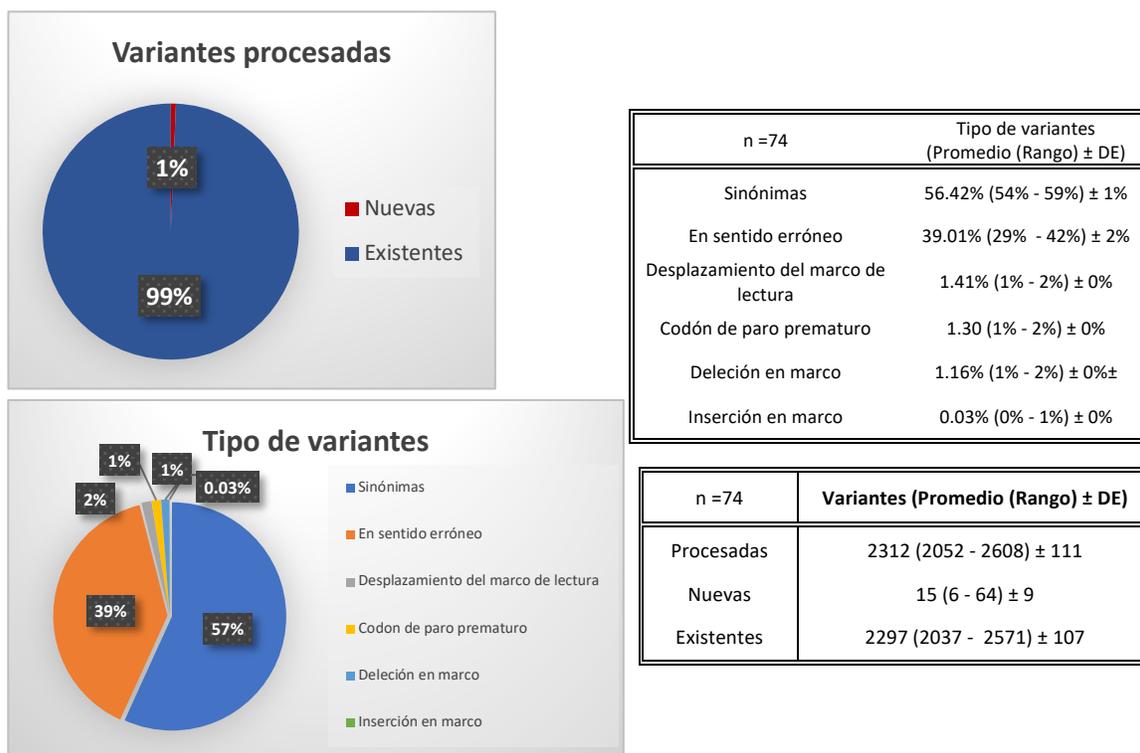


Figura 13. Estadística de variantes procesadas en la población de estudio. Posterior a la anotación de cada *vcf* individual, se determinó la cantidad y porcentaje de las variantes procesadas para subsecuentemente, obtener una estadística global (panel a y b). Además, se determinó el tipo y porcentaje de todas las variantes en regiones codificantes (panel c y d). Los anteriores resultados demuestran una mayor abundancia de variantes sinónimas y con menor frecuencia las codón de paro prematuro y delecciones inserciones en marco; estas últimas asociadas con efecto deletéreo.

7.7.1 Filtrado de variantes e identificación de variantes relacionadas con síndromes de predisposición a cáncer

- **Primer filtro**

La primera estrategia de filtrado que se utilizó para la identificación de variantes con alto grado de patogenicidad fue con base en la variable “*Clinical significance*” de VEP, la cual, integra la clasificación asignada por la base de datos ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) (75). La lista de variantes filtradas en al menos uno de los tríos, dúos o casos índice se muestra en la tabla 12. Las variantes en rojo están clasificadas en ClinVar como patogénicas o probablemente patogénicas, mientras las variantes en naranja se mantienen con conflictos hacia la patogenicidad.

Tabla 12. Lista de variantes con alto índice de patogenicidad identificadas en los individuos secenciados con según su clasificación en ClinVar

No de trio	ID	Individuo	Dianóstico	Locus	Genotipo (primer filtro)	VAF	MAF	Profundidad
1	521100310123-PB	Caso índice	Hepatoblastoma	8:128750680	(MYC):c.217A>C (p.Thr73Pro)	0.25	0.027	27x
	521010310165	Padre		N/A	No identificada	N/A	N/A	N/A
	521010310164	Madre		N/A	No identificada	N/A	N/A	N/A
2	521010310106-PB	Caso índice	Hepatoblastoma	N/A	No identificada	N/A	N/A	N/A
	521010310150	Madre		8:90955519	(NBN):c.2146A>G (p.Asn716Asp)	0.58	0.0130	76x
	521010310151	Padre		N/A	No identificada	N/A	N/A	N/A
3	521010310048-PB	Caso índice	Hepatoblastoma	5:1254594	(TERT):c.3184G>A (p.Ala1062Thr)	0.44	0.02	39x
	521010310161	Padre		5:1254594	(TERT):c.3184G>A (p.Ala1062Thr)	0.51	0.02	35x
	521010310162	Madre		N/A	No identificada	N/A	N/A	N/A
4	521010310249	Caso índice	Ependimoma	N/A	No identificada	N/A	N/A	N/A
	521010310314	Padre		3:37035011	(MLH1):c.-28A>G	0.49	0.0066	95x
	521010310328	Madre		N/A	No identificada	N/A	N/A	N/A
5	521010310143-PB	Caso índice	Ependimoma	11:64575505	(MEN1):c.512G>A (p.Arg171Gln)	0.48	0.02	35x
	521010310254	Madre		11:64575505	(MEN1):c.512G>A (p.Arg171Gln)	1.00	0.02	31x
	521010310302	Padre		X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	0.16	0.0490	12x
6	SHL_L002	Caso índice	Linfoma Hodgkin/sarcoma de Ewing	7:66459273	(SBDS):c.184A>T (p.Lys62Ter)	0.20	0.0066	39x
	HVL_L002	Madre		N/A	No identificada	N/A	N/A	N/A
	SVL_L002	Padre		N/A	No identificada	N/A	N/A	N/A
7	346_L002	Caso índice	Tumor germinal	N/A	No identificada	N/A	N/A	N/A
	394_L001	Padre		N/A	No identificada	N/A	N/A	N/A
	395_L001	Madre		2:48010481	(MSH6):c.114del (p.Ala40fs)	0.47	0.020	169x
No de duo	ID	Individuo	Dianóstico	Locus	Genotipo (primer filtro)	VAF	MAF	Profundidad

1	521010310086-PB	Caso índice	Hepatoblastoma	9:98231294	(PTCH1):c.1989G>C (p.Gln663His)	0.45	0.013	175x
	521010310171	Madre		9:98231294	(PTCH1):c.1989G>C (p.Gln663His)	0.38	0.013	144x
				X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	0.1	0.0490	20x
2	521010310047	Caso índice	Hepatoblastoma	8:128750680	(MYC):c.217A>C (p.Thr73Pro)	0.44	0.027	62x
	521010310222	Padre		17:42148542	(G6PC3):c.210del (p.Phe71fs)	0.26	0.0066	27x
3	521010320114-PB	Caso índice	Osteosarcoma	X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	0.13	0.0490	8x
	521010310203	Madre		N/A	No identificada	N/A	N/A	N/A
4	521010320129	Caso índice	Hepatoblastoma	10:43609994	(RET):c.1946C>T (p.Ser649Leu)	0.46	0.013	129x
				8:128750680	(MYC):c.217A>C (p.Thr73Pro)	0.24	0.027	57x
	521010320218	Madre		10:43609994	(RET):c.1946C>T (p.Ser649Leu)	0.44	0.013	111x
No de caso índice	ID	Individuo	Dianóstico	Locus	Genotipo (primer filtro)	VAF	MAF	Profundidad
1	CPA_L002	Caso índice	Osteosarcoma	17:29556955	(NF1):c.2953C>T (p.Gln985Ter)	0.50	0.0140	4x
				2:48010481	(MSH6):c.114del (p.Ala40fs)	0.45	0.020	177x
2	CPM_L002	Caso índice	Tumor de células gigantes	17:29556955	(NF1):c.2953C>T (p.Gln985Ter)	0.50	0.0140	4x
				2:48010481	(MSH6):c.114del (p.Ala40fs)	0.45	0.020	177x
3	521010310033	Caso índice	Ependimoma	11:108160350	(ATM):c.4258C>T (p.Leu1420Phe)	0.58	0.0066	36x
4	521010310069-PB	Caso índice	LLA	13:32911971	(BRCA2):c.3479G>A (p.Arg1160Lys)	0.37	0.0066	156x
5	521010310078-PB	Caso índice	LLA	8:90983460	(NBN):c.643C>T (p.Arg215Trp)	0.42	0.0066	73x
6	521010310084-PB	Caso índice	Tumor neuroectodérmico primitivo	1:94480241	(ABCA4):c.5318C>T (p.Ala1773Val)	0.43	0.0066	201x
				8:128750680	(MYC):c.217A>C (p.Thr73Pro)	0.40	0.027	55x
7	521010310089-PB	Caso índice	Tumor rabdoide teratoide atípico	12:133219838	(POLE):c.4523G>A (p.Arg1508His)	0.39	0.0066	61x
				X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	0.11	0.0490	42x
8	521010310090-PB	Caso índice	Fibrosarcoma	22:29107982	(CHEK2):c.707T>C (p.Leu236Pro)	0.49	0.0066	87x
9	521010310099-PB	Caso índice	Linfoma de Hodgkin	5:1254594	(TERT):c.3184G>A (p.Ala1062Thr)	0.39	0.02	28x
10	521010310103-PB	Caso índice	LLA	X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	0.11	0.0490	36x
				8:90955519	(NBN):c.2146A>G (p.Asn716Asp)	0.52	0.0130	326x
11	521010310107-PB	Caso índice	Osteosarcoma	1:45797507	(MUTYH):c.970C>T (p.Gln324Ter)	0.57	0.0066	7x
				9:21974681	(CDKN2A):c.146T>C (p.Ile49Thr)	0.98	0.0130	168x
				X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	0.22	0.0490	9x
12	521010310119-PB	Caso índice	LLA	9:98218623	(PTCH1):c.3241G>A (p.Val1081Met)	0.56	0.0066	53x
13	521010310126-PB	Caso índice	LLA	5:112178327	(APC):c.7036C>T (p.Pro2346Ser)	0.35	0.0066	80x

Con esta estrategia de filtrado, se identificaron variantes patogénicas o probablemente patogénicas en 7/41 (22.58%) casos índice, de los cuales, 4/41 (9.76%) casos índice presentaron variantes en genes de predisposición a cáncer, principalmente en los genes NF1, MSH6, CHEK2, MUTYH, CDKN2A y SBDS. Tres casos índice presentaron la misma variante

c.217A>C (p.Thr73Pro) en el gen MYC, la cual, fue clasificada como probablemente patogénica con origen somático. En 13/41 (31.7%) casos índice, se filtraron variantes con conflicto de interpretaciones hacia la patogenicidad, la mayoría de ellas asociadas a síndromes de predisposición a cáncer.

- **Segundo filtro**

Se estableció una segunda estrategia para verificar la presencia de más variantes con alto índice de patogenicidad, cuyos criterios de filtrado se establecieron con base en las recomendaciones para el filtrado y clasificación de variantes germinales y las variables disponibles en el anotador VEP (figura 14) (48, 76, 77).

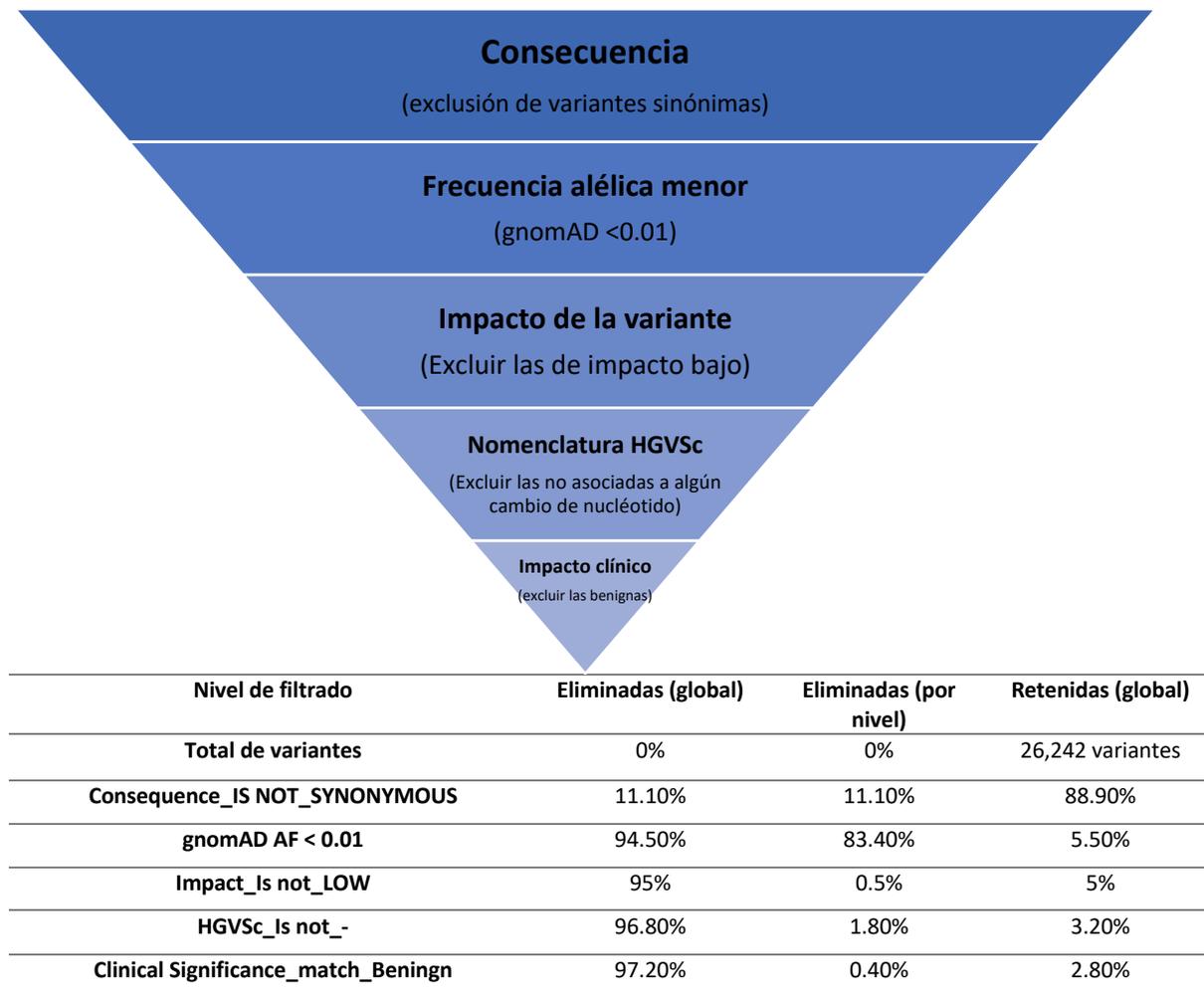


Figura 14. Estrategia de filtrado para la identificación de más variantes con alto índice de patogenicidad en los casos índice (panel a). El porcentaje promedio de variantes retenidas y eliminadas en cada filtro, se muestra en el panel b.

En la segunda estrategia de filtrado, se identificó la variante probablemente patogénica c.2639G>A (p.Arg880Gln) en el gen FANCA, la cual, no fue filtrada inicialmente debido a que estaba clasificada como variante de significado incierto. También se detectó la variante patogénica c.*1093G>C en el gen TPR, sin embargo, hasta el momento este gen ha sido asociado con el síndrome de camptodactilia-artropatía-coxa vara-pericarditis, y no con predisposición a cáncer. En contraparte, algunas variantes identificadas con la estrategia inicial, no fueron filtradas en ésta segunda estrategia debido a su exclusión por una frecuencia poblacional mayor al 0.01%, como las variantes probablemente patogénicas c.217A>C (p.Thr73Pro) en el gen MYC indentificadas en 4 individuos (gnomAD 0.01138); similar a la anterior, las variantes con conflictos de interpretaciones hacia la patogenicidad (TERT):c.3184G>A (p.Ala1062Thr) y (MEN1):c.512G>A (p.Arg171Gln). Una mayor cantidad de variantes con conflicto de interpretaciones hacia la patogenicidad, fueron adicionalmente identificadas en este filtro para casi todos los casos índice. La lista de variantes identificadas con este filtro se muestra en la tabla 13.

Tabla 13. Lista de variantes filtradas mediante una segunda estrategia de filtrado de 5 niveles

#	ID	Dianóstico	Locus	Genotipo (segundo filtro)	Patrón de herencia	VAF	MAF	Profundidad
1	521100310123-PB	Hepatoblastoma	13:32954019	(BRCA2):c.9086C>T (p.Ala3029Val)	AD	0.64	0.013	17x
			14:45644816	(FANCM):c.2859A>C (p.Lys953Asn)	AR	0.51	0.013	41x
2	521010310048-PB		16:68849457	(CDH1):c.1360G>A (p.Val454Ile)	AD	0.38	0.013	65x
3	521010310143-PB		3:14219965-14219968	(XPC):c.90GGA[4] (p.Glu34del)	AR	0.17	0.33	45x
4	SHL_L002	Linfoma Hodgkin sarcoma de Ewing	7:66459273	(SBD5):c.184A>T (p.Lys62Ter)	AR	0.20	0.0066	39x
5	521010310086-PB	Hepatoblastoma	9:98231294	(PTCH1):c.1989G>C (p.Gln663His)	AD	0.45	0.013	175x
			1:241672025	(FH):c.616G>A (p.Val206Ile)	AR	0.39	0.013	62x
			X:70360679	(MED12):c.6241CAG[5] (p.Gln2086del)	AD	0.13	0.33	15x
6	521010310047		16:23640970	(PALB2):c.2505C>T (p.Ser835=)	AD	0.56	0.0066	55x
			17:63534340	(AXIN2):c.1181G>A (p.Arg394His)	AD	0.60	0.02	64x
			3:30691871-30691872	(TGFB2):c.383dup (p.Pro129fs)	AD	0.16	0.02	71x
7	521010310076-PB	Carcinoma hepatocelular	13:32893197-32893198	(BRCA2):c.68-7del	AD	0.14	0.033	43x
			17:63554000	(AXIN2):c.739G>C (p.Val247Leu)	AD	0.41	0.02	118x
8	521010320114-PB	Osteosarcoma	X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	AR	0.13	0.0490	8x
			7:91682040	(AKAP9):c.5369T>G (p.Val1790Gly)	AD	0.56	0.0066	25x
9	521010310037	Ependimoma	19:17954215	(JAK3):c.394C>A (p.Pro132Thr)	AR	0.50	0.023	12x
10	521010320129	Hepatoblastoma	10:43609994	(RET):c.1946C>T (p.Ser649Leu)	AD	0.46	0.013	129x
			3:14219965-14219968	(XPC):c.90GGA[4] (p.Glu34del)	AR	0.09	0.33	84x
11	521010310209	Hepatoblastoma	22:24143134	(SMARCB1):c.366A>T (p.Glu122Asp)	AD	0.44	0.0066	117x
			22:29130695	(CHEK2):c.15G>A (p.Ser5=)	AD	0.36	0.0066	92x
12	CPM_L002	Tumor de células gigantes	7:142457375	(PRSS1):c.40C>G (p.Leu14Val)	AD	0.09	0.04	1093x

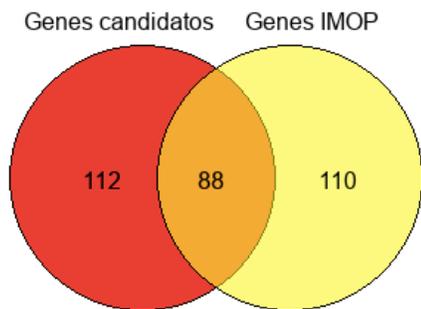
13	521010310033	Ependimoma	9:135786099-135786101	TSC1:c.1142-22_1142-21del	AD	0.48	0.0066	198x
			11:47254419	(DDB2):c.511C>G (p.Gln171Glu)	AR	0.45	0.0066	191x
14	521010310063-PB	LLA	17:41258553	(BRCA1):c.135-3T>C	AD	0.58	0.0066	36x
			19:42383644	(CD79A):c.419C>A (p.Thr140Asn)	AR	0.41	0.0066	29x
15	521010310066-PB	Nfibroblastoma	16:89815118	(FANCA):c.3297G>C (p.Gln1099His)	AR	0.44	0.0066	61x
			8:30933799	(WRN):c.835C>T (p.Arg279Trp)	AR	0.34	0.0066	52x
			13:32911971	(BRCA2):c.3479G>A (p.Arg1160Lys)	AD	0.37	0.0066	156x
16	521010310069-PB	LLA	13:32893197-32893198	(BRCA2):c.68-7del	AD	0.17	0.033	36x
			22:29091840	(CHEK2):c.1117A>G (p.Lys373Glu)	AD	0.10	0.0066	70x
			8:90983460	(NBN):c.643C>T (p.Arg215Trp)	AR	0.42	0.0066	73x
			X:70360679	(MED12):c.6241CAG[5] (p.Gln2086del)	AD	0.11	0.33	64x
17	521010310078-PB	LLA	5:236676	(SDHA):c.1394G>A (p.Arg465Gln)	AD	0.32	0.068	87x
			12:49431305-49431308	(KMT2D):c.9822GCA[5] (p.Gln3282dup)	AD	0.18	0.0066	11x
			2:223158900	(PAX3):c.572T>C (p.Ile191Thr)	AD	0.46	0.0066	196x
			1:94480241	(ABCA4):c.5318C>T (p.Ala1773Val)	AR	0.43	0.0066	201x
18	521010310084-PB	Tumor neuroectodérmico primitivo	1:17355211	(SDHB):c.307A>G (p.Met103Val)	AR	0.45	0.0066	159x
			5:131931468	(RAD50):c.2173C>T (p.Arg725Trp)	¿AR o AD?	0.40	0.0066	90x
			5:131940701	(RAD50):c.2718+10T>G	¿AR o AD?	0.75	0.0066	24x
			11:32421548	(WT1):c.1059C>T (p.Ile353=)	AD	0.37	0.0066	164x
19	521010310085-PB	Nfibroblastoma	2:47703623	(MSH2):c.2123T>C (p.Ile708Thr)	AD	0.44	0.0066	124x
			5:236676	(SDHA):c.1394G>A (p.Arg465Gln)	AD	0.12	0.068	128x
			12:133219838	(POLE):c.4523G>A (p.Arg1508His)	AD	0.39	0.0066	61x
20	521010310089-PB	Tumor rabdoide teratoide atípico	X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	AR	0.11	0.0490	42x
			10:112724116	(SHOC2):c.-1C>T	AD	0.40	0.0066	20x
			11:64572021	(MEN1):c.1618C>T (p.Pro540Ser)	AD	0.24	0.0066	34x
			22:29107982	(CHEK2):c.707T>C (p.Leu236Pro)	AD	0.49	0.0066	87x
			3:14219965-14219968	(XPC):c.90GGA[4] (p.Glu34del)	AR	0.08	0.33	59x
			3:30691871-30691872	(TGFBR2):c.383del (p.Lys128fs)	AD	0.18	0.013	53x
21	521010310090-PB	Fibrosarcoma	X:70360679	(MED12):c.6241CAG[5] (p.Gln2086del)	AD	0.07	0.33	73x
			5:236676	(SDHA):c.1394G>A (p.Arg465Gln)	AD	0.19	0.068	101x
			4:55569954	(KIT):c.821C>T (p.Thr274Met)	AD	0.47	0.0066	131x
			5:35876215	(IL7R):c.1007A>G (p.Glu336Gly)	AR	0.44	0.0066	71x
			19:17945696	(JAK3):c.2164G>A (p.Val722Ile)	AR	0.45	0.0066	95x
22	521010310091-PB	LLA	14:95574250	(DICER1):c.2617G>C (p.Asp873His)	AD	0.39	0.0066	59x
			3:30691871-30691872	(TGFBR2):c.383del (p.Lys128fs)	AD	0.14	0.013	50x
			5:236676	(SDHA):c.1394G>A (p.Arg465Gln)	AD	0.09	0.068	87x
23	521010310092-PB	LLA	12:49426905-49426908	(KMT2D):c.11568GCA[6] (p.Gln3863dup)	AD	0.25	0.0066	8x
			13:32893197-32893198	(BRCA2):c.68-7del	AD	0.13	0.033	23x

			16:89838165	(FANCA):c.2072A>G (p.Asn691Ser)	AR	0.61	0.0066	131x
			22:28195602-28195605	(MN1):c.912GCA[5] (p.Gln309del)	Sin dato	0.40	0.023	10x
			11:108123551	(ATM):c.1810C>T (p.Pro604Ser)	AR	0.33	0.0066	18x
			11:108141839	(ATM):c.2887A>G (p.Met963Val)	AR	0.41	0.0066	37x
24	521010310099-PB	Linfoma de Hodgkin	1:65321324	(JAK1):c.1516C>T (p.Arg506Cys)	AD	0.40	0.0066	42x
			16:89831437	(FANCA):c.2639G>A (p.Arg880Gln)	AR	0.52	0.0066	56x
			22:28195602-28195605	(MN1):c.912GCA[5] (p.Gln309del)	Sin dato	0.22	0.023	9x
			X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	AR	0.11	0.0490	36x
			8:90955519	(NBN):c.2146A>G (p.Asn716Asp)	AR	0.52	0.0130	326x
25	521010310103-PB	LLA	1:94520801	(ABCA4):c.2453G>A (p.Gly818Glu)	AR	0.44	0.0066	140x
			2:167129240	(SCN9A):c.3020G>A (p.Arg1007His)	AD	0.47	0.0066	34x
			4:55129965	(PDGFRA):c.499G>A (p.Val167Met)	AD	0.50	0.0066	115x
			9:98224170	(PTCH1):c.2671G>A (p.Gly891Ser)	AD	0.47	0.0066	75x
26	521010310105-PB	LLA	19:17954215	(JAK3):c.394C>A (p.Pro132Thr)	AR	0.59	0.023	27x
			5:131923710	(RAD50):c.980G>A (p.Arg327His)	¿AR o AD?	0.43	0.0066	65x
			1:45797507	(MUTYH):c.970C>T (p.Gln324Ter)	AR	0.57	0.0066	7x
			9:21974681	(CDKN2A):c.146T>C (p.Ile49Thr)	AD	0.98	0.0130	168x
			X:153008476	(ABCD1):c.1816T>C (p.Ser606Pro)	AR	0.22	0.0490	9x
27	521010310107-PB	Osteosarcoma	11:119149358	(CBL):c.1366G>T (p.Asp456Tyr)	AD	0.50	0.0066	46x
			13:32893197-32893198	(BRCA2):c.68-7del	AD	0.12	0.033	43x
			22:28195602-28195605	(MN1):c.912GCA[5] (p.Gln309del)	Sin dato	0.22	0.023	9x
			5:236676	(SDHA):c.1394G>A (p.Arg465Gln)	AD	0.08	0.068	51x
			9:98218623	(PTCH1):c.3241G>A (p.Val1081Met)	AD	0.57	0.0066	53x
28	521010310119-PB	LLA	2:29430133	(ALK):c.3842G>A (p.Ser1281Asn)	AD	0.50	0.0066	40x
			20:62292756	(RTEL1):c.208C>T (p.Arg70Cys)	¿AR o AD?	0.38	0.0066	37x
			3:30691871-30691872	(TGFB2):c.383del (p.Lys128fs)	AD	0.17	0.02	53x
			5:112178327	(APC):c.7036C>T (p.Pro2346Ser)	AD	0.35	0.0066	80x
			3:14199844	(XPC):c.1539G>A (p.Met513Ile)	AR	0.56	0.0066	64x
29	521010310126-PB	LLA	1:186282010	(TPR):c.*1093G>C	AR	0.47	0.0066	47x
			17:41242938	(BRCA1):c.4185+11TG[7]	AD	0.25	0.0230	8x
			7:128851596	(SMO):c.1921C>G (p.Pro641Ala)	Sin dato	0.50	0.0066	22x
			13:103515070	(ERCC5):c.1571C>T (p.Pro524Leu)	Sin dato	0.24	0.0066	21x
			17:37682375	(CDK12):c.3566T>A (p.Leu1189Gln)	Sin dato	0.47	0.0066	32x
30	521010310137-PB	Tumor neuroectodérmico primitivo	3:14219965-14219968	(XPC):c.90GGA[4] (p.Glu34del)	AR	0.13	0.33	38x
			5:236676	(SDHA):c.1394G>A (p.Arg465Gln)	AD	0.17	0.068	36x
			9:139409089	(NOTCH1):c.2080G>A (p.Glu694Lys)	AD	0.47	0.0066	17x
31	521010320076	Carcinoma hepatocelular	17:63554000	(AXIN2):c.739G>C (p.Val247Leu)	AD	0.40	0.02	97x
			13:32893197-32893198	(BRCA2):c.68-7del	AD	0.13	0.033	24x

32	521010310175	Hepatoblastoma	10:72360496	(PRF1):c.163C>T (p.Arg55Cys)	AR	0.50	0.013	63x
			13:32893197-32893198	(BRCA2):c.68-7del	AD	0.13	0.033	24x
33	521010310106-PB	Hepatoblastoma	2:167141318	(SCN9A):c.1619G>A (p.Arg540His)	AD	0.50	0.02	18x
34	521010310051-PB	Hepatoblastoma	11:22647387	(FANCF):c.-31T>C	AR	0.46	0.013	46x
			13:32890681	(BRCA2):c.67+17T>C	AD	0.50	0.013	22x
35	521010310112	Hepatoblastoma	17:63534340	(AXIN2):c.1181G>A (p.Arg394His)	AD	0.59	0.02	68x
			2:25457211	(DNMT3A):c.2676A>G (p.Ser892=)	AD	0.40	0.013	62x
			20:62321453	(RTEL1):c.2227G>A (p.Asp743Asn)	¿AR o AD?	0.41	0.013	130x
36	521010310053-PB	Hepatoblastoma	17:17131318	(FLCN):c.134C>G (p.Ala45Gly)	AD	0.50	0.013	48x
			8:145737877	(RECQL4):c.2953G>A (p.Val985Met)	AR	0.45	0.013	31x
			X:70360679	(MED12):c.6241CAG[5] (p.Gln2086del)	AD	0.11	0.33	28x
37	521010310073-PB	Osteosarcoma	7:152357857	(XRCC2):c.50G>A (p.Arg17Gln)	AR	0.45	0.013	82x
			8:145741761	(RECQL4):c.742C>T (p.Arg248Cys)	AR	0.43	0.013	30x
38	521010310120	Subependimoma	17:29676104-29676108	(NF1):c.7190-33TTGT[7]	AD	0.60	0.013	5x
			9:139417343	(NOTCH1):c.701G>A (p.Arg234His)	AD	0.27	0.013	11x
39	521010310249	Ependimoma	17:73831566	(UNC13D):c.1772C>T (p.Pro591Leu)	AR	0.50	0.013	32x
			5:236676	(SDHA):c.1394G>A (p.Arg465Gln)	AD	0.11	0.068	107x
40	346_L002	Tumor germinal	7:142457375	(PRSS1):c.40C>G (p.Leu14Val)	AD	0.09	0.04	662x

- Tercer filtro

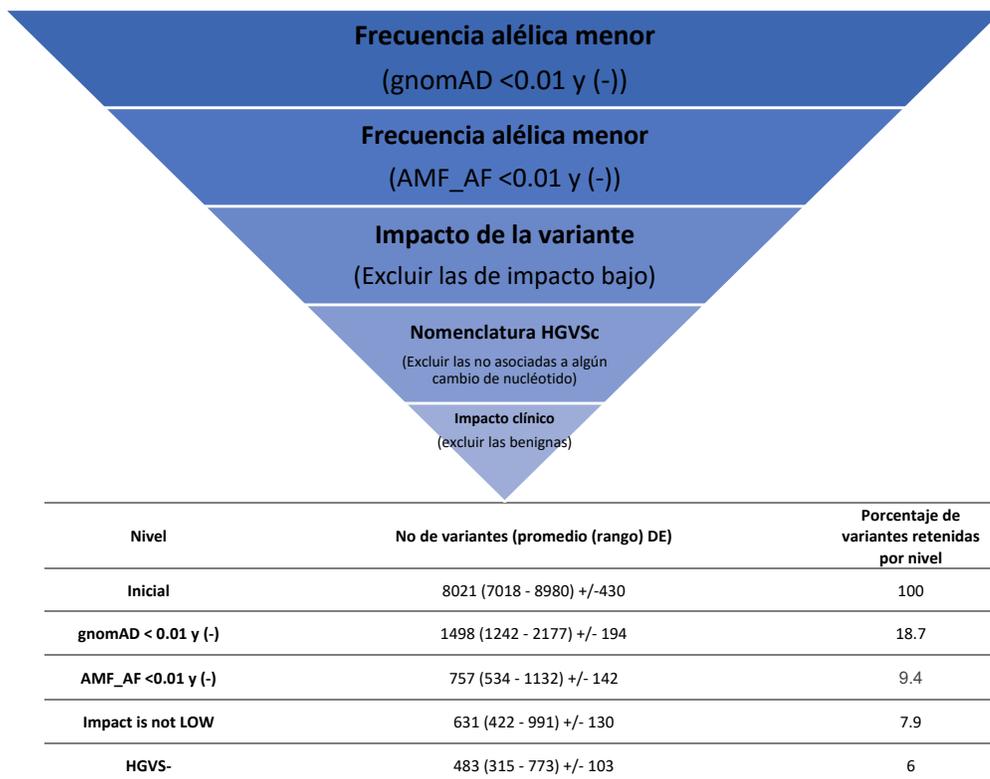
Las anteriores estrategias de filtrado permitieron identificar variantes germinales patogénicas/pb patogénicas en 10/40 casos índice (25%), de los cuales, 5/40 (12.5%) están asociadas con síndromes de predisposición a cáncer en ClinVar. Aunque el anterior porcentaje confirma la hipótesis planteada, con el objetivo de verificar si existen mutaciones adicionales o variantes candidatas, se realizó la búsqueda acotada en genes y síndromes de predisposición a cáncer descritos en niños y adultos a través de revisión bibliográfica (200 genes); además, se integraron genes utilizados en la IMOP (198). La lista final incluyó 310 genes que fueron depositados en un archivo txt, la lista se muestra en la figura 15.



Lista de genes
ABC11 ABL1 ABL2 ACD ACVR1 AGL AKT1 ALK ANKRD26 APC ARID1A ARID2 ASXL1 ASXL2
ATF1 ATG2B ATM ATRX AXIN2 BAP1 BARD1 BCOR BCORL1 BLM BMPR1A BRAF BRCA1
BRCA2 BRIP1 BTG1 BUB1B C19MC CBF1 CBL CCND3 CDC73 CDH1 CDK4 CDKN1B CDKN1C
CDKN2A CDKN2B CEBPA CHEK2 CIC COL7A1 CREB1 CREBBP CRLF2 CSF1R CSF3R CTC1 CTCF
CTNNA1 CTR9 CXCR4 CYLD DDB2 DDX3X DDX41 DHCR7 DICER1 DIS3L2 DKC1 DNAB1
DNMT3A EBF1 EGFR ELANE EML4 EP300 EPCAM EPOR ERCC1 ERCC2 ERCC3 ERCC4 ERCC5
ERG ETV6 EWSR1 EXT1 EXT2 EZH2 FAH FANCA FANCA1 FANCA2 FANCAV FANCC FANCE
FANCF FANCG FANCI FANCL FANCM FAS FBXW7 FGFR1 FGFR2 FGFR3 FH FLCN FLT3 FOXE1
FOXO1 FUS G6PC GATA1 GATA2 GJB2 GNAQ GNAS GPC3 GPC4 GRIA1 GSKIP H3F3A HABP2
HFE HIST1H3B HMBS HNF1A HRAS IDH1 IDH2 IKZF1 IL7R IPMK ITK JAK1 JAK2 JAK3 KDM6A
KIAA1549 KIT KMT2A KMT2D KRAS L2HGDH LEF1 LMO1 LZTR1 MAP2K1 MAX MC1R
MDM2 MED12 MEN1 MET MGA MITF MLH1 MPL MRE11A MSH2 MSH3 MSH6 MTHFR
MTOR MUTYH MYB MYBL1 MYC MYCN NBN NCOR1 NF1 NF2 NHP2 NIPBL NX2-1 NOP10
NOTCH1 NOTCH3 NPM1 NRAS NSD1 NTHL1 NTRK1 NTRK2 NTRK3 NUDT15 NUTM1 PALB2
PARN PAX3 PAX5 PAX6 PAX7 PAX8 PBRM1 PDGFB PDGFRB PDGFRB PHF6 PHOX2B PIK3CA
PIK3R1 PMS1 PMS2 POLD1 POLE POLH POT1 PPM1D PPOX PRF1 PRKACA PRKAR1A PRKCA
PRSS1 PTCH1 PTCH2 PTEN PTPN11 PTPRD RAD50 RAD51B RAD51C RAD51D RAF1 RARG
RB1 RECQL4 RELB REST RET RHOA ROS1 RPL11 RPL15 RPL26 RPL27 RPL31 RPL35A RPL5
RPS10 RPS17 RPS19 RPS24 RPS26 RPS27 RPS28 RPS29 RPS7 RTEL1 RUNX1 SAMD9
SAMD9L SDBS SDHA SDHAF2 SDHB SDHC SDHD SETBP1 SETD2 SF3B1 SH2B3 SHOC2
SLC28A3 SLX4 SMAD4 SMARCA4 SMARCB1 SMARCE1 SMO SOFU SOS1 SOS2 SPRTN SRP72
SS18 STAG2 STAT6 STK11 SUFU TERC TERT TET2 TFE3 TINF2 TMEM127 TP53 TPMT
TRIM37 TSC1 TSC2 TSR2 TTYH1 TYK2 U2AF1 UGT1A1 UGT1A8 USP6 VEGFA VHL WHSC1
WRAP53 WRN WT1 XPA XPC YAP1 YWHAE ZEB2 ZMIZ1 ZMYM3 ZNF217

Figura 15. Estrategia de filtrado para búsqueda de mutaciones patogénicas o variantes candidatas. Mediante la unificación de 2 listas de genes de predisposición a cáncer (panel a), se obtuvo una lista final de genes (panel b) para enfocar el filtrado de variantes

Los 310 genes fueron incluidos en un archivo de intervalos y subsecuentemente utilizados para acotar la anotación de variantes con VEP (anexo 1, fase 3). Se realizaron ajustes a los criterios de filtrado aplicados en esta tercera estrategia, los cuales se muestran en la figura 16.



Is not BENIGN	376 (230 - 664) +/-94	4.7
Finales	48 (31 - 102) +/-15	0.6

Ilustración 16. Tercera estrategia de filtrado para la identificación de variantes con alto índice de patogenicidad en los casos índice (panel a). El porcentaje promedio de variantes retenidas y eliminadas en cada filtro, se muestra en el panel b.

En filtros anteriores, se observó que algunas variantes carecían de información poblacional, en la cual, VEP lo especificaba como (-), por tal motivo, se incluye dicho criterio porque pueden existir variantes patogénicas sin información poblacional, como es el caso de la variante en el gen DICER1 c.5017dup (p.Ile1673AsnfsTer). Esta estrategia permitió la identificación de todas las variantes patogénicas en genes de predisposición identificadas en los dos filtros anteriores de una manera más sencilla y todas las variantes de impacto bajo fueron excluidas con esta estrategia; además, se identificaron 4 variantes candidatas que requieren estudios adicionales para confirmar su efecto biológico y causal. La lista completa de variantes identificadas en los casos índice se muestra en la tabla 14.

Con estos resultados, se concluye la estandarización del flujo de análisis bioinformático y estrategia de filtrado de variantes de manera satisfactoria.

Tabla 14. Lista completa de variantes filtradas mediante la tercera estrategia y clasificación

ID	Localización	Gen	HGVSc	HGVSp	ClinVar (Confirmación manual)
SHL_L002	7:66459273-66459273	SBDS	NM_016038.4:c.184A>T	NP_057122.2:p.Lys62Ter	Patogénica
	22:21351503-21351503	LZTR1	NM_006767.4:c.2407-18G>A	-	VUS
CPM_L002	17:29556955-29556955	NF1	NM_001042492.3:c.2953C>T	NP_001035957.1:p.Gln985Ter	Patogénica
	2:48010481-48010482	MSH6	NM_001281492.1:c.114del	NP_001268421.1:p.Ala40ProfsTer41	Patogénica
	4:57796771-57796771	REST	NM_005612.5:c.1747A>G	NP_005603.3:p.Thr583Ala	VUS
521010310 107-PB	1:45797507-45797507	MUTYH	NM_001048171.1:c.970C>T	NP_001041636.1:p.Gln324Ter	Patogénica
	9:21974681-21974681	CDKN2A	NM_058197.4:c.146T>C	NP_478104.2:p.Ile49Thr	Probablemente patogénica
	11:119149358-119149358	CBL	NM_005188.4:c.1366G>T	NP_005179.2:p.Asp456Tyr	VUS
521010310 099-PB	16:89831437-89831437	FANCA	NM_001286167.3:c.2639G>A	NP_001273096.1:p.Arg880Gln	Probablemente patogénica
	11:108123551-108123551	ATM	NM_001351834.2:c.1810C>T	NP_001338763.1:p.Pro604Ser	Benigna (12), Pb benigna (5), VUS (3)
	11:108141839-108141839	ATM	NM_001351834.2:c.2887A>G	NP_001338763.1:p.Met963Val	VUS
	1:65321324-65321324	JAK1	NM_002227.4:c.1516C>T	NP_002218.2:p.Arg506Cys	Not provided, alelo altamente penetrante
521010310 090-PB	22:29107982-29107982	CHEK2	NM_145862.2:c.707T>C	NP_665861.1:p.Leu236Pro	Probablemente patogénica

	14:95562239-95562239	DICER1	NM_177438.2:c.5017dup	NP_803187.1:p.Ile1673AsnfsTer8	Patogénica
	4:55569954-55569954	KIT	NM_001093772.1:c.821C>T	NP_001087241.1:p.Thr274Met	VUS (4), benigna (2)
521010310 103-PB	1:94520801-94520801	ABCA4	NM_000350.3:c.2453G>A	NP_000341.2:p.Gly818Glu	Patogénica
	4:55129965-55129965	PDGFRA	NM_006206.6:c.499G>A	NP_006197.1:p.Val167Met	VUS
	9:98224170-98224170	PTCH1	ENST00000331920.6:c.2671G>A	ENSP00000332353.6:p.Gly891Ser	Pb benigna (2), VUS (1)
	2:58390205-58390205	FANCL	NM_018062.3:c.779dup	NP_060532.2:p.Pro262ThrfsTer5	VARIANTE CANDIDATA
	1:94480241-94480241	ABCA4	NM_000350.3:c.5318C>T	NP_000341.2:p.Ala1773Val	Patogénica
521010310 084-PB	1:17355211-17355211	SDHB	NM_003000.3:c.307A>G	NP_002991.2:p.Met103Val	VUS
	5:131940701-131940701	RAD50	NM_005732.4:c.2718+10T>G	-	VUS (1), Pb benigna (1)
	9:135772705-135772705	TSC1	NM_000368.5:c.2841G>T	NP_000359.1:p.Arg947Ser	VUS
521010310 143-PB	19:11145715-11145718	SMARCA4	NM_003072.4:c.4090_4092del	NP_003063.2:p.Glu1364del	VUS
	1:27057722-27057722	ARID1A	NM_139135.4:c.1430A>G	NP_624361.1:p.His477Arg	VUS
521010310 105-PB	5:131923710-131923710	RAD50	NM_005732.4:c.980G>A	NP_005723.2:p.Arg327His	Binigna (4), Pb benigna (2), VUS (2)
	22:29695789-29695789	EWSR1	NM_013986.4:c.1894G>T	NP_053733.2:p.Glu632Ter	VARIANTE CANDIDATA
521010310 086-PB	1:241672025-241672025	FH	NM_000143.3:c.616G>A	NP_000134.2:p.Val206Ile	VUS
	14:45644342-45644342	FANCM	NM_020937.4:c.2385T>G	NP_065988.1:p.Ile795Met	VUS
	9:98231294-98231294	PTCH1	NM_000264.5:c.1989G>C	NP_000255.2:p.Gln663His	VUS (3), Pb benigna (2)
	2:47596763-47596763	EPCAM	ENST00000419334.1:c.122_123insCGGGG	ENSP00000389028.1:p.Ser43GlyfsTer14	VARIANTE CANDIDATA
	11:533518-533518	HRAS	NM_176795.4:c.385C>T	NP_789765.1:p.Gln129Ter	VARIANTE CANDIDATA
521100310 123-PB	13:32954019-32954019	BRCA2	NM_000059.3:c.9086C>T	NP_000050.2:p.Ala3029Val	VUS
	14:45644816-45644816	FANCM	NM_020937.4:c.2859A>C	NP_065988.1:p.Lys953Asn	VUS
	15:89860130-89860130	POLG	NM_002693.2:c.3644-72C>A	-	VUS (2), benigna (1)
521010320 129	10:43609994-43609994	RET	NM_020975.6:c.1946C>T	NP_066124.1:p.Ser649Leu	Pb benigna (8), VUS (5), benigna (2)
	7:138554323-138554323	KIAA1549	NM_020910.3:c.4736C>A	NP_065961.2:p.Thr1579Lys	VUS
521010320 076	17:63554000-63554000	AXIN2	NM_004655.4:c.739G>C	NP_004646.3:p.Val247Leu	VUS
521010310 175	10:72360496-72360496	PRF1	NM_005041.5:c.163C>T	NP_005032.2:p.Arg55Cys	VUS
	19:17943595-17943595	JAK3	ENST00000527031.1:n.2278+3941A>C	-	VUS
521010310 137-PB	9:139409089-139409089	NOTCH1	NM_017617.5:c.2080G>A	NP_060087.3:p.Glu694Lys	Pb benigna (4), benigna (2), VUS (1)
	13:103515070-103515070	ERCC5	NM_000123.3:c.1571C>T	NP_000114.2:p.Pro524Leu	Not provided, alelo altamente penetrante
521010310 126-PB	3:14199844-14199844	XPC	NM_004628.4:c.1539G>A	NP_004619.3:p.Met513Ile	VUS

	5:112178327-112178327	APC	NM_000038.6:c.7036C>T	NP_000029.2:p.Pro2346Ser	Binigna (3), Pb benigna (3), VUS (3)
	7:128851596-128851596	SMO	NM_005631.5:c.1921C>G	NP_005622.1:p.Pro641Ala	Not provided, alelo altamente penetrante
521010310 120	9:139417343-139417343	NOTCH1	NM_017617.5:c.701G>A	NP_060087.3:p.Arg234His	Pb benigna (3), VUS (2)
	2:29430133-29430133	ALK	NM_004304.5:c.3842G>A	NP_004295.2:p.Ser1281Asn	VUS
521010310 119-PB	20:62292756-62292756	RTEL1	NM_032957.5:c.208C>T	NP_116575.3:p.Arg70Cys	VUS
	9:98218623-98218623	PTCH1	NM_001354918.2:c.3085G>A	NP_001341847.1:p.Val1029Met	Pb benigna (2), pb patogénica (1), VUS (1)
521010310 112	20:62321453-62321453	RTEL1	NM_032957.5:c.2227G>A	NP_116575.3:p.Asp743Asn	VUS
521010310 092-PB	16:89838165-89838165	FANCA	NM_001286167.3:c.2072A>G	NP_001273096.1:p.Asn691Ser	VUS
521010310 091-PB	14:45620727-45620727	FANCM	NM_020937.4:c.1046T>C	NP_065988.1:p.Ile349Thr	VUS
	14:95574250-95574250	DICER1	NM_177438.2:c.2617G>C	NP_803187.1:p.Asp873His	Pb benigna (1), VUS (1)
521010310 089-PB	10:112724116-112724116	SHOC2	NM_007373.3:c.-1C>T	-	VUS
	12:133219838-133219838	POLE	NM_006231.4:c.4523G>A	NP_006222.2:p.Arg1508His	VUS (5), Pb benigna (3), benigna (1)
521010310 085-PB	11:32421548-32421548	WT1	NM_001367854.1:c.-130C>T	-	Pb benigna (1), VUS (1)
	2:47703623-47703623	MSH2	NM_000251.3:c.2123T>C	NP_000242.1:p.Ile708Thr	VUS
521010310 078-PB	2:223158900-223158900	PAX3	NM_181461.4:c.572T>C	NP_852126.1:p.Ile191Thr	Pb benigna (2), VUS (1)
	7:6027011-6027011	PMS2	NM_001322014.2:c.1385T>C	NP_001308943.1:p.Ile462Thr	VUS
	8:90983460-90983460	NBN	NM_002485.4:c.643C>T	NP_002476.2:p.Arg215Trp	VUS (5), Pb benigna (6), benigna (3), patogénica (2)
521010310 073-PB	8:145741761-145741761	RECQL4	ENST00000428558.2:c.742C>T	ENSP00000475456.1:p.Arg248Cys	VUS
	7:152357857-152357857	XRCC2	NM_005431.2:c.50G>A	NP_005422.1:p.Arg17Gln	VUS
521010310 069-PB	22:29091840-29091840	CHEK2	NM_007194.4:c.1117A>G	NP_009125.1:p.Lys373Glu	VUS
	13:32911971-32911971	BRCA2	ENST00000380152.3:c.3479G>A	ENSP00000369497.3:p.Arg116OLys	VUS
521010310 066-PB	10:89653825-89653825	PTEN	NM_001304717.5:c.642A>T	NP_001291646.4:p.Arg214Ser	VUS
	16:89815118-89815118	FANCA	NM_001286167.3:c.3297G>C	NP_001273096.1:p.Gln1099His	VUS
	8:30933799-30933799	WRN	NM_000553.6:c.835C>T	NP_000544.2:p.Arg279Trp	VUS
521010310 063-PB	11:47254419-47254419	DDB2	NM_000107.2:c.511C>G	NP_000098.1:p.Gln171Glu	Pb benigna (1), VUS (1)
	16:14028081-14028081	ERCC4	NM_005236.3:c.1135C>T	NP_005227.1:p.Pro379Ser	Pb benigna (2), benigna (1), VUS (1)
	17:41258553-41258553	BRCA1	ENST00000489037.1:c.135-1580T>C	-	VUS

	3:52668822-52668822	PBRM1	NM_181042.5:c.1097A>G	NP_851385.1:p.Tyr366Cys	Not provided, alelo altamente penetrante
	12:133220529-133220529	POLE	NM_006231.4:c.4184A>G	NP_006222.2:p.Tyr1395Cys	Benigna (4), Pb benigna (1), VUS (1)
521010310 053-PB	17:17131318-17131318	FLCN	NM_144997.7:c.134C>G	NP_659434.2:p.Ala45Gly	Benigna (1), Pb benigna (1), VUS (1)
	8:145737877-145737877	RECQL4	NM_004260.3:c.2953G>A	NP_004251.3:p.Val985Met	Benigna (1), Pb benigna (1), VUS (1)
521010310 051-PB	11:22647387-22647387	FANCF	ENST00000327470.3:c.-31T>C	-	VUS
	13:32890681-32890681	BRCA2	NM_000059.3:c.67+17T>C	-	Pb benigna (3), benigna (2), VUS (1)
521010310 048-PB	16:68849457-68849457	CDH1	NM_004360.5:c.1360G>A	NP_004351.1:p.Val454Ile	VUS
521010310 047	16:23640970-23640970	PALB2	ENST00000568219.1:c.*2380C>T	-	Pb benigna (4), VUS (1)
521010310 037	8:145742534-145742534	RECQL4	NM_004260.3:c.254C>T	NP_004251.3:p.Ala85Val	VUS
521010310 033	5:112174677-112174677	APC	NM_001354895.2:c.3386T>C	NP_001341824.1:p.Leu1129Ser	Benigna (7), Pb benigna (5), VUS (1)
521010310 249	17:73831566-73831566	UNC13D	NM_199242.2:c.1772C>T	NP_954712.1:p.Pro591Leu	VUS
521010310 209	22:24143134-24143134	SMARCB1	NM_003073.5:c.366A>T	NP_003064.2:p.Glu122Asp	VUS
	22:29130695-29130695	CHEK2	NM_001257387.2:c.-763G>A	-	Pb benigna (6), VUS (1)
521010310 076-PB	17:63554000-63554000	AXIN2	NM_004655.4:c.739G>C	NP_004646.3:p.Val247Leu	VUS
346_L002	N/A	N/A	N/A	N/A	N/A
521010320 114-PB	N/A	N/A	N/A	N/A	N/A
521010310 106-PB	N/A	N/A	N/A	N/A	N/A

7.8 Verificación de las variantes identificadas

7.8.1 Revisión manual de variantes

Todas las variantes patogénicas, probablemente patogénicas, variantes de significado incierto (VUS) y variantes candidatas fueron revisadas manualmente y confirmadas a través del archivo VCF y BAM con IGV. Las imágenes representativas (variantes con alto índice de patogenicidad) se muestran la figura 17, las demás se encuentran en el anexo 2.

a)



b)

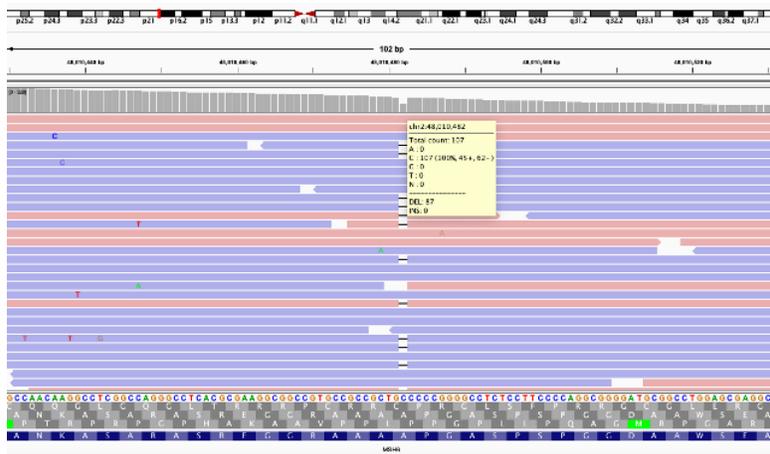


Figura 17. Verificación de variantes en los genes NF1 y MSH6 en el archivo BAM del caso índice CPA_L002. El panel A muestra la imagen de las lecturas alineadas en la región 66,459,248 – 66,459,293 del gen NF1, donde se observa la variante heterocigota (NF1):c.2953C>T (p.Gln985Ter). El mismo individuo fue portador de la covariante heterocigota (MSH6):c.114del (p.Ala40fs) en el gen MSH6 (panel b).

7.8.2 Prueba de concepto con programas comerciales para identificación de variantes germinales

Además, con el objetivo de validar el adecuado flujo de análisis bioinformático e identificación de variantes, se realizó una prueba de concepto con 2 programas para identificación de variantes germinales utilizados en contexto clínico, Sophia Genetics y Qiagen Clinical Insight (QCI) interpret en 2 y 3 casos índice, respectivamente.

Las variantes filtradas por Sophia Genetics permitieron la identificación de la variante MSH6:c.114del (p.Ala40fs) en el caso índice CPA_L002, la cual también fue identificada en nuestro flujo de trabajo; sin embargo, la variante (NF1):c.2953C>T (p.Gln985Ter) no fue filtrada. Para el caso SHL_L002, también se identificó la variante c.184A>T

(p.Lys62Ter) en el gen SBDS. En contraste, el QCI identificó la variante (NF1):c.2953C>T (p.Gln985Ter) del caso índice CPA_L002, pero no la variante MSH6:c.114del (p.Ala40fs); además, no se identificó las variantes en SBDS en el caso índice SHL_L002. Para el tercer caso índice (521010310073-PB), no se identificaron variantes patogénicas ni en QCI ni en nuestro flujo de análisis.

Estos resultados respaldan el correcto flujo de trabajo y filtrado de nuestro flujo de análisis bioinformático.

7.8.3 Validación de las variantes de relevancia clínica

La validación de las variantes patogénicas, probablemente patogénicas y candidatas, se realizó mediante la comparación y búsqueda de variantes en los archivos BAM de una muestra independiente (tejido) y los germinales (sangre o saliva). De los 6 individuos con mutaciones patogénicas y probablemente patogénicas identificadas, 3 contaron con datos de secuenciación en muestras de tejido; mientras que todos los individuos portadores de variantes candidatas contaron con datos de secuenciación en tejido (tabla 15).

Tabla 15. Relación de casos índice portadores de mutaciones germinales o variantes candidatas con datos de secuenciación en tejido

ID	Género	Diagnóstico / Edad al diagnóstico	Tipo de variante y número	¿Disponible BAM somático?
CPM_L002	Femenino	Tumor de células gigantes en columna, 16 años	Mutaciones (2)	No
CPA_L002	Masculino	Osteosarcoma, 14 años	Mutación (1)	No
521010310107-PB	Femenino	Osteosarcoma, 14 años	Mutaciones (2)	Si
521010310090-PB	Femenino	Fibrosarcoma de ovario, 7 años	Mutaciones (2)	Si
SHL_L002	Masculino	Sarcoma de Ewing y linfoma no Hodgkin, 8 años	Mutación (1)	No
521010310099-PB	Masculino	Linfoma de Hodgkin, edad desconocida	Mutación (1)	Si
521010310103-PB	Masculino	LLA	Variante candidata (1)	Si
521010310105-PB	Femenino	LLA	Variante candidata (1)	Si
521010310086-PB	Masculino	Hepatoblastoma	Variante candidata (1)	Si

Todas las mutaciones identificadas en los 3 casos índice con datos de secuenciación en tejido disponible fueron confirmadas (Figura 18). Las imágenes comparativas de todas las variantes confirmadas pueden consultarse en el anexo 2.

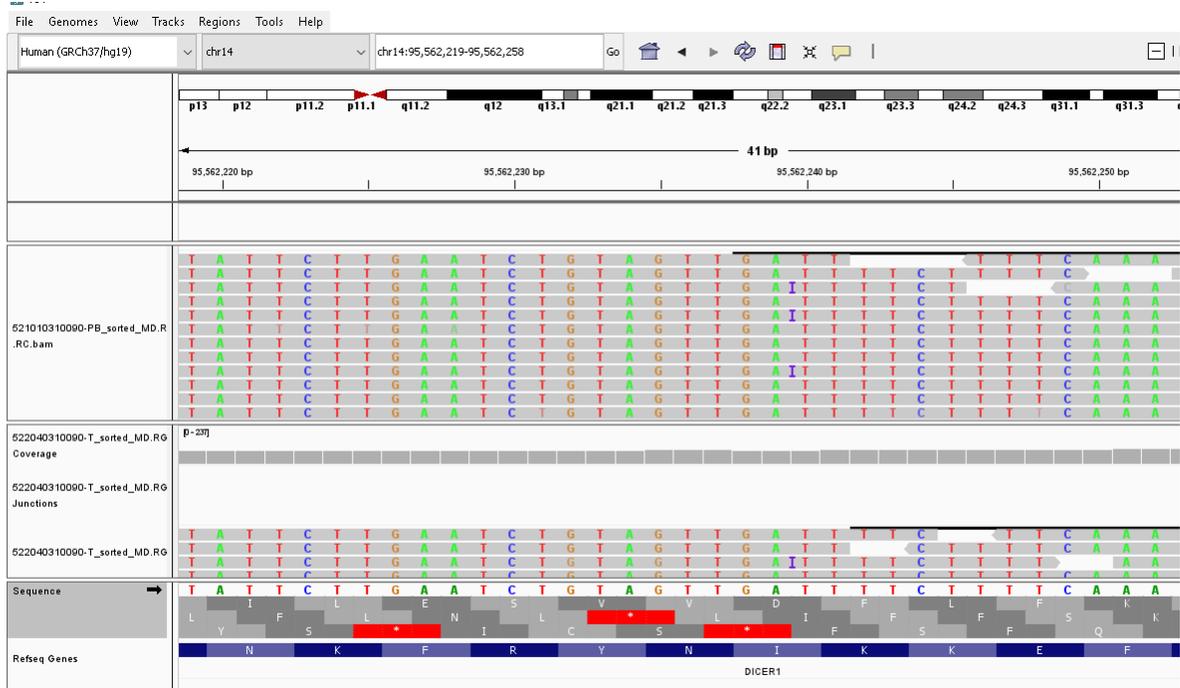


Figura 18. Confirmación de la variante *DICER1*::c.5017dup (p.Ile1673fs) en la caso índice 521010310090. La revisión de los alineamientos sobre los archivos BAM recalibrados para sangre periférica () y tejido () demostraron una frecuencia alélica de la variante (VAF) de 0.42 y 0.99, respectivamente.

7.8.4 Análisis de confirmación de sexo y relaciones genéticas

Uno de los principales controles de calidad en estudios de epidemiología genética, específicamente aquellos que agrupan individuos en tríos o dúos, es su confirmación de relaciones genéticas y sexo, con el objetivo de evitar resultados discordantes por intercambio de muestras, errores de identificación, etc. Los resultados de estos análisis permitieron confirmar el sexo (anexo 3) y parentesco de todos los individuos, excepto en 2 familias (Tabla 16).

La primer familia está conformada por 2 casos índice (CPM_L002 y CPA_L002) que corresponden a 2 casos índice hermanos, donde inicialmente fueron identificadas 2 mutaciones en los genes *NF1* y *MSH6* en CPM pero no en CPA. Interesantemente, la mutación en el gen *MSH6* fue identificada en una madre de otro caso índice no relacionado (395_L001). Tras el análisis de parentesco, se confirmó que el sexo de la madre es masculino (no femenino), aunado a que se validó el parentesco en primer grado con CPM_L002. Estos datos permitieron confirmar que el archivo identificado inicialmente como 395_L001 se trataba del caso índice CPA_L002.

Tabla 16. Estimaciones de parentesco entre los individuos secuenciados

ID	Puntaje
521010320076 - 521010310219	0.15489
1030011001 (SHL) - 1030011002 (HVL)	0.13932
1030011001 (SHL) - 1030011003 (SVL)	0.13984
1030011003 (SVL) - 1030011002 (HVL)	0.18582
521010320129 - 521010320218	0.14702
1030010001 (CPM) - 1030010001 (CPA)	0.14756
521010310112 - 521010310167	0.1503
521010310112 - 521010310166	0.15425
521010310120 - 521010310284	0.15139
521010310120 - 521010310283	0.15836
521010310076 - 521010310219	0.15188
520000000346 - 520000000394	0.15266
521010310047 - 521010310222	0.15302
521010310053-PB - 521010310213	0.15327
521010310053-PB - 521010310214	0.15869
521010310106 - 521010310151	0.1547
521010310106 - 521010310150	0.1573
521010310037 - 521010310292	0.15552
521010310143 - 521010310302	0.1572
521010310143 - 521010310254	0.16003
521010310175 - 521010310149	0.15849
521010310175 - 521010310172	0.15952
521010310249 - 521010310314	0.15911
521010310249 - 521010310328	0.16063
521010310086 - 521010310171	0.15948
521010310051 - 521010310153	0.16443
521010310051 - 521010310152	0.16454
521010310073 - 521010310185	0.17312
521010310073 - 521010310184	0.18057
521010310048 - 521010310161	0.16444
521010310048 - 521010310162	0.16753
521100310123 - 521010310165	0.17019
521100310123 - 521010310164	0.17202
521010320114-PB - 521010310203	0.18091

Así mismo, se confirmó consanguinidad de primer grado entre los padres biológicos del caso índice SHL_L002, quien tiene diagnóstico de doble tumor primario (Linfoma Hodgkin y sarcoma de Ewing). Dicha consanguinidad fue sospechada por el apellido de ambos padres biológicos, sin embargo, ellos negaron tener parentesco alguno.

Estos resultados resaltan la importancia de la confirmación de relaciones tanto para la adecuada agrupación de los individuos como para confirmar vínculos genéticos entre padres biológicos.

7.8.5 Análisis de ancestría

Se detectaron cinco componentes ancestrales tanto en los casos como en los controles pareados (Figura 19), los cuales, fueron consistentes con los modelos demográficos asumidos en mexicanos modernos de ascendencia mixta (78-82). En casos y controles, se identificó un componente de mezclas de nativos americanos como el más destacado (74,11 % \pm 15,72 % frente a 68,54 % \pm 19,73 %, respectivamente) seguido de un componente de mezclas europeas (22,29 % \pm 14,65 % vs 27,49% \pm 18,50%, respectivamente), mezclas africanas (1,55 % \pm 2,90 % frente a 1,23 % \pm 2,50 %, respectivamente), asiáticas orientales (1,71 % \pm 2,07 % frente a 1,95 % \pm 1,92 %, respectivamente) y habitantes de Oceanía (0,32 % \pm 0,69 % vs. 0,78% \pm 1,86%, respectivamente). Los grupos de casos y controles no difirieron significativamente en ninguno de los cinco componentes ancestrales encontrados para nuestro modelo, pero es importante señalar que el poder fue limitado debido al tamaño de la muestra.

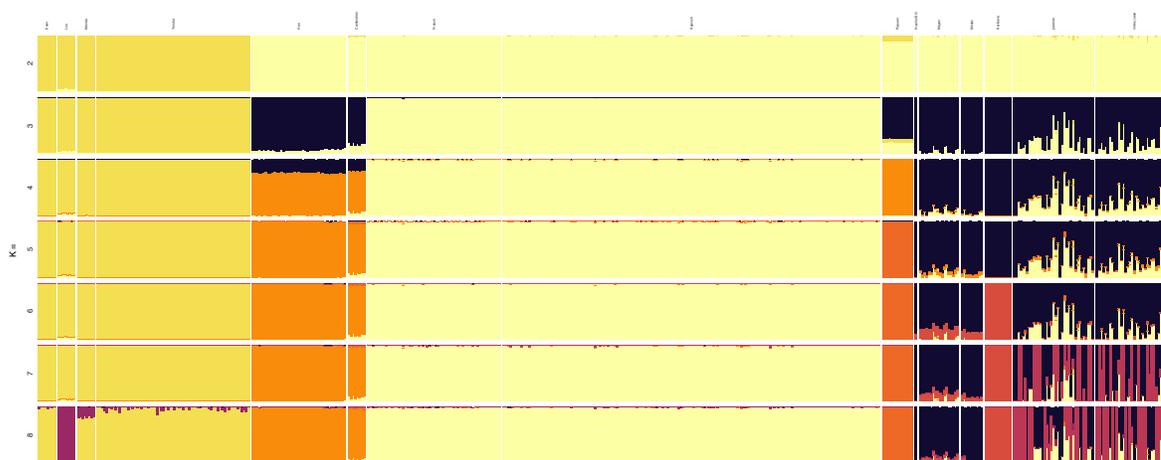


Figura 19. Gráfico de ancestría. El componente más destacado que se identificó corresponde a mezcla de nativos americanos, en el cual se incluyen grupos étnicos como Huichol, Maya, Mixteco y Karitiana.

7.9 Correlación clínico-biológica de las variantes identificadas

Una vez confirmadas las relaciones familiares y verificadas las variantes, en la tabla 17 se enlistan los casos índice portadores de variantes patogénicas o probablemente patogénicas, así como información clínica y familiar de cada paciente. Se identificaron nueve mutaciones patogénicas o probablemente patogénicas en genes de predisposición en 6/40 casos índice (15%). En 3 pacientes se identificaron 2 mutaciones coexistentes y en uno de ellos, fue portador tanto de mutaciones con patrón de herencia autosómico dominante (*CDKN2A*) y recesivo (*MUTYH*); solo 2 pacientes fueron portadores de mutaciones autosómicas recesivas (*FANCA* y *SBDS*).

Tabla 17. Pacientes portadores de mutaciones germinales e información clínica

Familia	Miembro	Sexo	Diagnóstico /Edad al diagnóstico	Family history of cancer	Supervivencia Global	Estado actual	Gen	Variante	Cigocidad	Clasificación	Patrón de herencia
1	III-1	F	Tumor de células gigantes en columna 16 años	Osteosarcoma (Hermano, 14 años)	54.5 meses	Vivo	<i>NFI</i> <i>MSH6</i>	c.2953C>T (p.Gln985Ter) c.114del (p.Ala40fs)	Heterocigoto Heterocigoto	Patogénica Patogénica	Heredada (<i>MSH6</i>)
	III-2	M	Osteosarcoma, 14 años	Tumor de células gigantes en columna 16 años	1.7 meses	Muerto	<i>MSH6</i>	c.114del (p.Ala40fs)	Heterocigoto	Patogénica	Heredada
2	III-1	F	Osteosarcoma, 14 años	Cáncer cervicouterino (Tía maternal y abuela materna, desconocida)	5.7 meses	Muerto	<i>MUTYH</i> <i>CDKN2A</i>	c.928C>T (p.Gln310Ter) c.146T>C (p.Ile49Thr)	Heterocigoto Homocigoto	Patogénica Probablemente patogénica	Desconocida
3	III-2	F	Fibrosarcoma de ovario (7 años) / Tumor de ovario contralateral de células de la granulosa (11 años)	Wilms (Hermano, 13 años) Tumor de ovario (Primo materno, desconocido)	99 meses	Vivo	<i>CHEK2</i> <i>DICER1</i>	c.707T>C (p.Leu236Pro)	Heterocigoto	Probablemente patogénica	Desconocida
								c.5017dup (p.Ile1673fs)	Heterocigoto	Patogénica	
4	III-1	M	Sarcoma de Ewing (5 años) / LH (7 años)	Negativo	81.6 meses	Vivo	<i>SBDS</i>	c.184A>T (p.Lys62Ter)	Heterocigoto	Patogénica	De novo
5	III-1	M	Hepatoblastoma / edad desconocida	Cáncer cervicouterino (Abuela materna, desconocida)	6.5 meses	Sin dato	<i>FANCA</i>	c.2639G>A (p.Arg880Gln)	Heterocigoto	Probablemente patogénica	Desconocida

Adicionalmente, filtraron 3 variantes sin sentido candidatas, para las cuales, se calcularon los puntajes de patogenicidad mediante los algoritmos FATHMM(83), MutPred2(84) y MuPro(85), los cuales, demostraron que el efecto sobre la estructura y función de cada proteína es deletéreo y/o patogénico (tabla 18).

Tabla 18. Variantes candidatas y puntajes de predicción

ID	Locus	Variant	VAF	Depth	FATHMM score	MutPred2 Score	MuPro Score
521010310103-PB	Chr2:58390205	(FANCL): c.779dup (p.Pro262Thrfs Ter5)	0.48	41x	Pathogenic (0.94)	Pathogenic (0.614)	- 0.81072694 (DECREASE stability)
521010310105-PB	Chr22:29695789	(EWSR1): c.1894G>T (p.Glu632Ter)	0.2	10x	Pathogenic (0.92313)	Pathogenic (0.5267)	- 0.76687274 (DECREASE stability)
521010310086-PB	Chr11:533518	(HRAS): c.385C>T (p.Gln129Ter)	0.5	115x	Pathogenic (0.98008)	Pathogenic (0.52)	- 1.278893 (DECREASE stability)

Tanto las variantes candidatas como las variantes de significado incierto, requieren un análisis más profundo y/o seguimiento con el objetivo de confirmar su impacto clínico, o cambios en la clasificación hacia la patogenicidad. Dichos cambios en su clasificación podrían aumentar el porcentaje de mutaciones germinales en genes de predisposición.

9.- Discusión

La cobertura de secuencia del 10X alcanzada en más del 80% fue suficiente para un genotipado preciso (70), además, el ADN purificado a partir de muestras de saliva demostró un adecuado desempeño en la secuenciación de exomas humanos, sin pérdida significativa de lecturas alineadas al genoma de referencia similar a otros reportes previos (57, 86).

Este estudio incluyó a 40 niños con cáncer, de los cuales, el 15% fueron portadores de mutaciones germinales en genes de predisposición a cáncer. Este porcentaje es mayor que los reportes previos (19, 34, 35) y podría explicar en parte las disparidades de prevalencia, severidad de la enfermedad y bajas tasas de sobrevida reportados en nuestra población (87, 88), aunque no puede excluirse un sesgo por muestreo, ya que la mayoría de los casos positivos fueron diagnosticados con sarcomas, tumor con fuerte asociación con síndromes de cáncer hereditario (89, 90).

En 32 casos índice se identificaron 40 variantes de significado incierto en genes de predisposición, la mayoría heredadas del padre, madre o ambos. De estas variantes, 35% se predicen deletéreas por al menos dos algoritmos *In Silico*, sin embargo, la caracterización adicional y el seguimiento son necesarios para confirmar su papel en predisposición a cáncer (77, 91). Además, identificamos 3 variantes candidatas sin sentido en genes de predisposición en 3 casos índice. Aunque no están completamente caracterizadas, la predicción *In Silico* sugiere que pueden tener impacto en la estructura y función de la proteína. Una de estas variantes en el gen FANCL, coexiste con una mutación patogénica en el gen ABDC4; interesantemente, este paciente fue diagnosticado con LLA a los 9 años y cuenta con 2 familiares afectados con cáncer (abuela materna con leucemia, edad desconocida y un primo con fenotipo desconocido a los 38 años).

Dos casos índices fueron portadores de mutaciones patogénicas en el gen ABCD4, las cuales, están asociadas con enfermedades oftalmológicas; sin embargo, existe 2 reportes de casos de niños afectados con retinoblastoma que fueron diagnosticados con enfermedad de

Stargardt (uno de ellos, bilateral), lo que demuestra una probable contribución de este gen en el desarrollo de cáncer que requiere ser explorado (92, 93).

Pese a que nuestros resultados confirmaron una fuerte historia familiar de cáncer, solo un caso índice fue previamente referida a consultoría genética por sospecha de síndrome de Beckwith-Wiedemann. Esta información es muy importante para la detección temprana de niños con predisposición a cáncer, sin embargo, en la práctica clínica, los síndromes subyacentes y los antecedentes familiares de cáncer no son identificados (94-96). Los resultados del cuestionario de Marjolijn C.J. Jongmans demostraron una alta especificidad y un valor predictivo negativo prometedores, por lo que puede ser una buena opción para cualquier personal de salud para el reconocimiento de niños con altas probabilidades de ser portadores de mutaciones germinales. El bajo valor predictivo positivo puede deberse a que la estrategia de secuenciación no incluye variantes en regiones no codificantes, puede haber variantes desconocidas hasta este momento y, en algunos familiares con cáncer, el diagnóstico histopatológico y/o edad al diagnóstico no pudo ser confirmado.

Un control de calidad importante en estudios de familias es asegurar el adecuado agrupamiento de individuos para evitar errores de diagnóstico por contaminación de muestras, etiquetado o intercambio de muestras (97). Interesantemente, posterior a la verificación de relaciones familiares confirmamos consanguinidad de segundo grado en los padres biológicos de un caso índice portador de una mutación en SBDS, lo que puede explicar el origen de la mutación y respaldo de la importancia de confirmar relaciones familiares entre padres y su asociación con cáncer reportados por otros autores (98-100).

Las poblaciones latinas son pobremente caracterizadas, ya que, la mayoría de las variantes han sido caracterizadas principalmente en poblaciones caucásicas (101). Nuestros resultados demostraron una proporción de componente ancestral de mezclas de nativos americanos, lo que puede ser respaldado por el lugar de nacimiento de la mayoría de los individuos reclutados. Una noción importante que surge de este trabajo es que aún se necesitan esfuerzos adicionales para superar la subrepresentación de pacientes hispanos en los estudios genéticos, lo que podría explicar las disparidades de alta incidencia, recurrencia y tasas de supervivencia más bajas reportadas en países latinoamericanos como México.

Este estudio presenta algunas limitaciones. Primero, no se incluyeron todos los tipos de cáncer y nuestra cohorte, incluye una mayor proporción de hepatoblastomas; sin embargo, ninguno de estos pacientes presentó mutaciones; segundo, no todos los casos índice pudieron ser agrupados y el origen de las variantes no pudo ser confirmado en todos los pacientes; tercero, algunas variantes no pudieron ser confirmadas por la falta de disponibilidad de muestra o porque no contaban con secuenciación en tejido y cuarto, aunque se ha reportado que hasta el 85% de las variantes con relevancia clínica se encuentran en regiones codificantes, la caracterización de variantes intrónicas profundas puede aumentar el número de individuos positivos.

10.- Conclusiones

Se encontraron mutaciones del gen de predisposición de la línea germinal en el 15% de los niños mexicanos con cáncer, aunque la confirmación de patogenicidad de VUS o variantes candidatas podría aumentar el porcentaje de individuos con mutaciones de la línea germinal; así mismo, todos los individuos portadores de mutaciones fueron identificados por La herramienta de selección Jongmans Marjolin, la cual podría ser una buena alternativa para la identificación de pacientes con cáncer infantil con síndromes de cáncer subyacentes.

El análisis genético poblacional confirmó una ascendencia mixta mexicana moderna que no ha sido incluida en otros reportes y su exploración, podría explicar las disparidades reportadas en nuestra población.

La identificación de portadores de mutaciones podría mejorar el diagnóstico temprano, ya que en México la mayoría de los diagnósticos de cáncer infantil se realizan en etapas avanzadas. Se necesitan más esfuerzos para evaluar el impacto de las variantes genéticas en las poblaciones latinas, incluida la evaluación de los factores ambientales y su interacción con las variantes genéticas.

11.- Referencias

1. Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al. International incidence of childhood cancer, 2001-10: a population-based registry study. *Lancet Oncol.* 2017;18(6):719-31.
2. (IACR) IAoRoC. International Childhood Cancer Day: Much remains to be done to fight childhood cancer World Health Organization; 2016.
3. Denburg A, Cuadrado C, Alexis C, Klusmann FA, Zamora JCB, Bodkyn C, et al. Improving childhood cancer care in Latin America and the Caribbean: a PAHO Childhood Cancer Working Group position statement. *Lancet Oncol.* 2017;18(6):709-11.
4. Rivera-Luna R, Zapata-Tarres M, Shalkow-Klincovstein J, Velasco-Hidalgo L, Olaya-Vargas A, Finkelstein-Mizrahi N, et al. The burden of childhood cancer in Mexico: Implications for low- and middle-income countries. *Pediatr Blood Cancer.* 2017;64(6).
5. Rivera-Luna R, Velasco-Hidalgo L, Zapata-Tarrés M, Cárdenas-Cardos R, Aguilar-Ortiz MR. Current outlook of childhood cancer epidemiology in a middle-income country under a public health insurance program. *Pediatr Hematol Oncol.* 2017;34(1):43-50.
6. Gelband H, Jha P, Sankaranarayanan R, Horton S. *Cancer: Disease Control Priorities, Third Edition (Volume 3).* 2015.
7. Ribeiro RC, Antillon F, Pedrosa F, Pui CH. Global Pediatric Oncology: Lessons From Partnerships Between High-Income Countries and Low- to Mid-Income Countries. *J Clin Oncol.* 2016;34(1):53-61.
8. Bhatia S. Disparities in cancer outcomes: lessons learned from children with cancer. *Pediatr Blood Cancer.* 2011;56(6):994-1002.

9. Linabery AM, Ross JA. Childhood and adolescent cancer survival in the US by race and ethnicity for the diagnostic period 1975-1999. *Cancer*. 2008;113(9):2575-96.
10. Collaborators GCC. The global burden of childhood and adolescent cancer in 2017: an analysis of the Global Burden of Disease Study 2017. *Lancet Oncol*. 2019;20(9):1211-25.
11. Ripperger T, Bielack SS, Borkhardt A, Brecht IB, Burkhardt B, Calaminus G, et al. Childhood cancer predisposition syndromes-A concise review and recommendations by the Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology. *Am J Med Genet A*. 2017;173(4):1017-37.
12. Rodriguez-Galindo C, Friedrich P, Alcasabas P, Antillon F, Banavali S, Castillo L, et al. Toward the Cure of All Children With Cancer Through Collaborative Efforts: Pediatric Oncology As a Global Challenge. *J Clin Oncol*. 2015;33(27):3065-73.
13. Murphy MF, Bithell JF, Stiller CA, Kendall GM, O'Neill KA. Childhood and adult cancers: contrasts and commonalities. *Maturitas*. 2013;76(1):95-8.
14. Downing JR, Wilson RK, Zhang J, Mardis ER, Pui CH, Ding L, et al. The Pediatric Cancer Genome Project. *Nat Genet*. 2012;44(6):619-22.
15. Globocan. Cancer today. <https://gco.iarc.fr/today>. 2020.
16. Barr RD, Ferrari A, Ries L, Whelan J, Bleyer WA. Cancer in Adolescents and Young Adults: A Narrative Review of the Current Status and a View of the Future. *JAMA Pediatr*. 2016;170(5):495-501.
17. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. *Cancer*. 2005;103(7):1457-67.
18. Spector LG, Pankratz N, Marcotte EL. Genetic and nongenetic risk factors for childhood cancer. *Pediatr Clin North Am*. 2015;62(1):11-25.
19. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555(7696):321-7.
20. Ma X, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, Zhou X, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*. 2018;555(7696):371-6.
21. Neuhausen SL. Ethnic differences in cancer risk resulting from genetic variation. *Cancer*. 1999;86(11 Suppl):2575-82.
22. Laken SJ, Petersen GM, Gruber SB, Oddoux C, Ostrer H, Giardiello FM, et al. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat Genet*. 1997;17(1):79-83.
23. Nyström-Lahti M, Wu Y, Moisio AL, Hofstra RM, Osinga J, Mecklin JP, et al. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. *Hum Mol Genet*. 1996;5(6):763-9.
24. Brauch H, Kishida T, Glavac D, Chen F, Pausch F, Höfler H, et al. Von Hippel-Lindau (VHL) disease with pheochromocytoma in the Black Forest region of Germany: evidence for a founder effect. *Hum Genet*. 1995;95(5):551-6.
25. Fragoso-Ontiveros V, Velázquez-Aragón JA, Nuñez-Martínez PM, de la Luz Mejía-Aguayo M, Vidal-Millán S, Pedroza-Torres A, et al. Mexican BRCA1 founder mutation: Shortening the gap in genetic assessment for hereditary breast and ovarian cancer patients. *PLoS One*. 2019;14(9):e0222709.

26. Late Effects of Treatment for Childhood Cancer (PDQ®): Health Professional Version. PDQ **Cancer** Information Summaries [Internet]. Bethesda (MD): National **Cancer** Institute (US); 2002-. 2020 Feb 4.: PDQ Pediatric Treatment Editorial Board . 2020.
27. Vogelstein B KK. The Genetic Basis of Human Cancer. 2nd ed ed. New York, NY2002.
28. Garber JE, Offit K. Hereditary cancer predisposition syndromes. *J Clin Oncol*. 2005;23(2):276-92.
29. Knapke S, Zelle K, Nichols KE, Kohlmann W, Schiffman JD. Identification, management, and evaluation of children with cancer-predisposition syndromes. *Am Soc Clin Oncol Educ Book*. 2012:576-84.
30. Jongmans MC, Loeffen JL, Waanders E, Hoogerbrugge PM, Ligtenberg MJ, Kuiper RP, et al. Recognition of genetic predisposition in pediatric cancer patients: An easy-to-use selection tool. *Eur J Med Genet*. 2016;59(3):116-25.
31. Kuhlen M, Borkhardt A. Cancer susceptibility syndromes in children in the area of broad clinical use of massive parallel sequencing. *Eur J Pediatr*. 2015;174(8):987-97.
32. Staler Z, Maio A, Padunan A, editors. Germline mutations prevalence in young adults with cancer. AACR Annual Meeting 2020; 2020 June 22, 2020: American Association of Cancer Research.
33. Walsh M, Wu G, Edmonson M, Gruber TA, Easton J, Yergeau D, et al. Incidence of Germline Mutations in Cancer-Predisposition Genes in Children with Hematologic Malignancies: a Report from the Pediatric Cancer Genome Project. *Blood*. 2014;124(21):127-.
34. Zhang J, Nichols KE, Downing JR. Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med*. 2016;374(14):1391.
35. Oberg JA, Glade Bender JL, Sulis ML, Pendrick D, Sireci AN, Hsiao SJ, et al. Implementation of next generation sequencing into pediatric hematology-oncology practice: moving beyond actionable alterations. *Genome Med*. 2016;8(1):133.
36. Ballinger ML, Goode DL, Ray-Coquard I, James PA, Mitchell G, Niedermayr E, et al. Monogenic and polygenic determinants of sarcoma risk: an international genetic study. *Lancet Oncol*. 2016;17(9):1261-71.
37. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971;68(4):820-3.
38. Foulkes WD. Inherited susceptibility to common cancers. *N Engl J Med*. 2008;359(20):2143-53.
39. Eng C, Ponder BA. The role of gene mutations in the genesis of familial cancers. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 1993;7(10):910-9.
40. Kutanzi KR, Lumen A, Koturbash I, Miousse IR. Pediatric Exposures to Ionizing Radiation: Carcinogenic Considerations. *Int J Environ Res Public Health*. 2016;13(11).
41. Patel DM, Jones RR, Booth BJ, Olsson AC, Kromhout H, Straif K, et al. Parental occupational exposure to pesticides, animals and organic dust and risk of childhood leukemia and central nervous system tumors: Findings from the International Childhood Cancer Cohort Consortium (I4C). *Int J Cancer*. 2020;146(4):943-52.

42. Carlos-Wallace FM, Zhang L, Smith MT, Rader G, Steinmaus C. Parental, In Utero, and Early-Life Exposure to Benzene and the Risk of Childhood Leukemia: A Meta-Analysis. *Am J Epidemiol*. 2016;183(1):1-14.
43. Metayer C, Petridou E, Arangur  JM, Roman E, Sch z J, Magnani C, et al. Parental Tobacco Smoking and Acute Myeloid Leukemia: The Childhood Leukemia International Consortium. *Am J Epidemiol*. 2016;184(4):261-73.
44. Arakawa A, Ichikawa H, Kubo T, Motoi N, Kumamoto T, Nakajima M, et al. Vaginal Transmission of Cancer from Mothers with Cervical Cancer to Infants. *N Engl J Med*. 2021;384(1):42-50.
45. Khoury MJ. Editorial: Emergence of Gene-Environment Interaction Analysis in Epidemiologic Research. *Am J Epidemiol*. 2017;186(7):751-2.
46. Medina-Sanson A, N n ez-Enr quez JC, Hurtado-Cordova E, P rez-Saldivar ML, Mart nez-Garc a A, Jim nez-Hern ndez E, et al. Genotype-Environment Interaction Analysis of NQO1, CYP2E1, and NAT2 Polymorphisms and the Risk of Childhood Acute Lymphoblastic Leukemia: A Report From the Mexican Interinstitutional Group for the Identification of the Causes of Childhood Leukemia. *Front Oncol*. 2020;10:571869.
47. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-303.
48. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-24.
49. Koch L. Exploring human genomic diversity with gnomAD. *Nat Rev Genet*. 2020;21(8):448.
50. Kuhlen M, Taeubner J, Brozou T, Wieczorek D, Siebert R, Borkhardt A. Family-based germline sequencing in children with cancer. *Oncogene*. 2019;38(9):1367-80.
51. Kuhlen M, Borkhardt A. Trio sequencing in pediatric cancer and clinical implications. *EMBO Mol Med*. 2018;10(4).
52. Narod SA, Goldgar D, Cannon-Albright L, Weber B, Moslehi R, Ives E, et al. Risk modifiers in carriers of BRCA1 mutations. *Int J Cancer*. 1995;64(6):394-8.
53. Humphrey SP, Williamson RT. A review of saliva: normal composition, flow, and function. *J Prosthet Dent*. 2001;85(2):162-9.
54. Vidovi  A, Vidovi  Juras D, Vu i evi  Boras V, Luka  J, Grubi i -Ili  M, Rak D, et al. Determination of leucocyte subsets in human saliva by flow cytometry. *Arch Oral Biol*. 2012;57(5):577-83.
55. Abraham JE, Maranian MJ, Spiteri I, Russell R, Ingle S, Luccarini C, et al. Saliva samples are a viable alternative to blood samples as a source of DNA for high throughput genotyping. *BMC Med Genomics*. 2012;5:19.
56. Garbieri TF, Brozoski DT, Dion sio TJ, Santos CF, Neves LT. Human DNA extraction from whole saliva that was fresh or stored for 3, 6 or 12 months using five different protocols. *J Appl Oral Sci*. 2017;25(2):147-58.

57. Zhu Q, Hu Q, Shepherd L, Wang J, Wei L, Morrison CD, et al. The impact of DNA input amount and DNA source on the performance of whole-exome sequencing in cancer epidemiology. *Cancer Epidemiol Biomarkers Prev.* 2015;24(8):1207-13.
58. Wall JD, Tang LF, Zerbe B, Kvale MN, Kwok PY, Schaefer C, et al. Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Res.* 2014;24(11):1734-9.
59. Johnson LM, Hamilton KV, Valdez JM, Knapp E, Baker JN, Nichols KE. Ethical considerations surrounding germline next-generation sequencing of children with cancer. *Expert Rev Mol Diagn.* 2017;17(5):523-34.
60. Coury SA, Schneider KA, Schiend J, Tan WH. Recognizing and Managing Children with a Pediatric Cancer Predisposition Syndrome: A Guide for the Pediatrician. *Pediatr Ann.* 2018;47(5):e204-e16.
61. Chen B, Gagnon M, Shahangian S, Anderson NL, Howerton DA, Boone DJ. Good Laboratory Practices for Molecular Genetic Testing for Heritable Diseases and Conditions: Centers for Disease Control and Prevention; 2009 [Available from: <https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5806a1.htm>].
62. Blueprint Genetics Hereditary Pediatric Cancer Panel [Available from: <https://blueprintgenetics.com/tests/panels/hereditary-cancer/hereditary-pediatric-cancer-panel/>].
63. Invitae Pediatric Genetics [Available from: <https://www.invitae.com/en/physician/category/CAT000016/#category-content-CAT000044>].
64. MSK Kids. Clinical Genetics Service and Pediatric Cancer Predisposition Screening Program [Available from: <https://www.mskcc.org/pediatrics/cancer-care/about/clinical-genetics-service-and-pediatric-cancer-predisposition-screening>].
65. Pediatric Cancer Genetic Risk Program of Dana-Farber Boston Children's Center [Available from: <http://www.danafarberbostonchildrens.org/innovative-approaches/pediatric-cancer-genetics-program.aspx>].
66. Melendez-Zajgla J, Mercado-Celis GE, Gaytan-Cervantes J, Torres A, Gabiño NB, Zapata-Tarres M, et al. Genomics of a pediatric ovarian fibrosarcoma. Association with the DICER1 syndrome. *Sci Rep.* 2018;8(1):3252.
67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
68. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43:11.0.1-0.33.
69. N H, A N, Krapcho M ea. SEER Cancer Statistics Review, 1975-2015.: National Cancer Institute; 2015.
70. Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagn.* 2018;20(1):4-27.
71. Kong SW, Lee IH, Liu X, Hirschhorn JN, Mandl KD. Measuring coverage and accuracy of whole-exome sequencing in clinical context. *Genet Med.* 2018;20(12):1617-26.

72. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, et al. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn.* 2017;19(3):341-65.
73. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061-73.
74. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122.
75. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. *Nucleic Acids Res.* 2020;48(D1):D835-D44.
76. Park S, Supek F, Lehner B. Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits. *Nat Commun.* 2018;9(1):2601.
77. Nykamp K, Anderson M, Powers M, Garcia J, Herrera B, Ho YY, et al. Sherlock: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet Med.* 2017;19(10):1105-17.
78. Price AL, Patterson N, Yu F, Cox DR, Waliszewska A, McDonald GJ, et al. A genomewide admixture map for Latino populations. *Am J Hum Genet.* 2007;80(6):1024-36.
79. Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, et al. Genetic variation and population structure in native Americans. *PLoS Genet.* 2007;3(11):e185.
80. Zúñiga J, Yu N, Barquera R, Alosco S, Ohashi M, Lebedeva T, et al. HLA class I and class II conserved extended haplotypes and their fragments or blocks in Mexicans: implications for the study of genetic diversity in admixed populations. *PLoS One.* 2013;8(9):e74442.
81. Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, et al. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* 2014;10(9):e1004572.
82. Barquera R, Hernández-Zaragoza DI, Bravo-Acevedo A, Arrieta-Bolaños E, Clayton S, Acuña-Alonzo V, et al. The immunogenetic diversity of the HLA system in Mexico correlates with underlying population genetic structure. *Hum Immunol.* 2020;81(9):461-74.
83. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat.* 2013;34(1):57-65.
84. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun.* 2020;11(1):5918.
85. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins.* 2006;62(4):1125-32.
86. Ibrahim O, Sutherland HG, Haupt LM, Griffiths LR. Saliva as a comparable-quality source of DNA for Whole Exome Sequencing on Ion platforms. *Genomics.* 2020;112(2):1437-43.

87. Rodriguez-Romo L, Olaya Vargas A, Gupta S, Shalkow-Klincovstein J, Vega-Vega L, Reyes-Lopez A, et al. Delivery of Pediatric Cancer Care in Mexico: A National Survey. *J Glob Oncol*. 2018;4:1-12.
88. Rivera-Luna R, Shalkow-Klincovstein J, Velasco-Hidalgo L, Cárdenas-Cardós R, Zapata-Tarrés M, Olaya-Vargas A, et al. Descriptive Epidemiology in Mexican children with cancer under an open national public health insurance program. *BMC Cancer*. 2014;14:790.
89. Farid M, Ngeow J. Sarcomas Associated With Genetic Cancer Predisposition Syndromes: A Review. *Oncologist*. 2016;21(8):1002-13.
90. Chan SH, Lim WK, Ishak NDB, Li ST, Goh WL, Tan GS, et al. Germline Mutations in Cancer Predisposition Genes are Frequent in Sporadic Sarcomas. *Sci Rep*. 2017;7(1):10660.
91. Moghadasi S, Eccles DM, Devilee P, Vreeswijk MP, van Asperen CJ. Classification and Clinical Management of Variants of Uncertain Significance in High Penetrance Cancer Predisposition Genes. *Hum Mutat*. 2016;37(4):331-6.
92. Margalit E, Sunness JS, Green WR, Kelman SE, Schachat AP, Fiergang D, et al. Stargardt disease in a patient with retinoblastoma. *Arch Ophthalmol*. 2003;121(11):1643-6.
93. Steinmetz RL, Garner A, Maguire JJ, Bird AC. Histopathology of incipient fundus flavimaculatus. *Ophthalmology*. 1991;98(6):953-6.
94. Wood ME, Kadlubek P, Pham TH, Wollins DS, Lu KH, Weitzel JN, et al. Quality of cancer family history and referral for genetic counseling and testing among oncology practices: a pilot test of quality measures as part of the American Society of Clinical Oncology Quality Oncology Practice Initiative. *J Clin Oncol*. 2014;32(8):824-9.
95. Lu KH, Wood ME, Daniels M, Burke C, Ford J, Kauff ND, et al. American Society of Clinical Oncology Expert Statement: collection and use of a cancer family history for oncology providers. *J Clin Oncol*. 2014;32(8):833-40.
96. Oncology ASOC. American Society of Clinical Oncology policy statement update: genetic testing for cancer susceptibility. *J Clin Oncol*. 2003;21(12):2397-406.
97. Pedersen BS, Quinlan AR. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. *Am J Hum Genet*. 2017;100(3):406-13.
98. Kakaje A, Alhalabi MM, Ghareeb A, Karam B, Mansour B, Zahra B, et al. Interactions of Consanguinity and Number of Siblings with Childhood Acute Lymphoblastic Leukemia. *Biomed Res Int*. 2020;2020:7919310.
99. Bittles A. Consanguinity and its relevance to clinical genetics. *Clin Genet*. 2001;60(2):89-98.
100. Jastaniah W, Aljefri A, Ayas M, Alharbi M, Alkhayat N, Al-Anzi F, et al. Prevalence of hereditary cancer susceptibility syndromes in children with cancer in a highly consanguineous population. *Cancer Epidemiol*. 2018;55:88-95.
101. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538(7624):161-4.

12.- Anexos

Anexo 1. Flujo de análisis bioinformático

Fase 1. Pre-procesamiento de datos

Ordenado de lecturas (Picard tools)

Posterior al alineamiento al genoma de referencia, las lecturas de secuencia son ordenadas de manera aleatoria. Este paso es necesario para acceder de manera más eficiente a la información contenida en los archivos BAM (archivos de alineamientos mapeados en formato binario). Todos los archivos fueron ordenados de acuerdo con las coordenadas de alineación en cada cromosoma (opción `SO=coordinate`).

- Estructura del comando:

```
java -jar picard.jar SortSam \  
I=file.bam \  
O=file_sorted.bam \  
SO=coordinate \  

```

Marcado de lecturas duplicadas (Picard tools)

Los archivos de lecturas de secuencia incluyen artefactos originados durante la amplificación de librerías o duplicados ópticos durante el proceso de secuenciación. Este paso permite la identificación y mitigación de dichos artefactos para evitar errores amplificados durante el llamado de variantes.

- Estructura del comando:

```
java -jar -Xmx4g picard.jar MarkDuplicates \  
I=file_sorted.bam \  
O=file_sorted_MD.bam METRICS_FILE=file_metrics.metrics
```

Agregar o reemplazar *ReadGroups*, RG (Picard tools)

En corrida de secuenciación multiplexada, se obtiene un subconjunto de lecturas para cada muestra originadas a partir de una librería amplificada de manera independiente y secuenciadas en líneas separadas. La información de dichos subconjuntos se encuentra en el *header* de cada archivo BAM, el cual incluye: el identificador del grupo de lectura (RGID), identificador de la librería (RGLB), información de la plataforma utilizada (RGPU), identificador de la muestra (RGSM) y el nombre del centro donde se realizó la secuenciación (RGCN). El objetivo de este paso es asignar un mismo identificador de grupo a todos los subconjuntos de lecturas en un mismo archivo.

Este paso es necesario para que, en pasos subsecuentes como la recalibración de bases, se puedan diferenciar características técnicas asociadas con artefactos.

El primer paso es visualizar el header del BAM para obtener la información de los subconjuntos de lectura, después, se asigna un mismo RG a todas las lecturas.

- Estructura de comandos

Paso 1. Visualización de headers (samtools)

```
samtools view -H file.bam | grep @RG
```

Paso 2. Asignación de un RG a cada archivo (Picard tools)

```
java -jar picard.jar AddOrReplaceReadGroups \  
I=file_sorted_MD.bam \  
O=file_sorted_MD.RG.bam \  
RGID=C2BTV.1 RGLB=Pond-273827 RGPL=illumina \  
RGPU=C2BTVACXX130914.1.AGGTCGCA-ATCCTTAA RGSM=521010310033 RGCN=BI
```

Indización de archivos BAM (Picard tools)

Una vez que los archivos BAM han sido ordenados, se crea un archivo complementario (.bai) que funciona como índice. Esto permite que las herramientas bioinformáticas puedan acceder más rápido a las lecturas alineadas en regiones genómicas específicas.

- Estructura del comando:

```
java -jar picard.jar BuildBamIndex \  
I=file_sorted_MD.RG.bam \  
O=file_sorted_MD.RG.bam.bai
```

Realineación de INDELS (GATK)

Aunque a partir del GATK v4.0 o superiores, el flujo de trabajo de llamado de variantes con HaplotypeCaller o MuTec2 omiten el paso de realineación de INDELS, persiste la recomendación para seguir realizando este paso. Por tal motivo, se instaló la versión 3.8.1 para poder realizar dicho proceso.

Este paso permite corregir errores de mapeo durante el alineamiento de lecturas, haciendo una realineación específica en sitios genómicos que contienen INDELS a través de un modelo de covariación mediante archivos con variantes ya conocidas (dbSNP). El proceso se realiza mediante 2 pasos: la identificación de regiones necesarias para la realineación y creación de archivo de intervalos (*RealignerTargetCreator*) y la realineación propiamente dicha (*IndelRealigner*).

- Estructura de comandos

Paso 1. RealignerTargetCreator (GATK)

```
java -Xmx2g -jar GenomeAnalysisTK.jar \  
-T RealignerTargetCreator \  
-R Homo_sapiens_assembly19.fasta \  
-I file_sorted_MD.RG.bam \  
-o file_sorted_MD.RG.intervals \  
-known 1000G_phase1.indels.b37.vcf
```

Paso 2. IndelRealigner (GATK)

```
java -Xmx4g -jar GenomeAnalysisTK.jar \  
-T IndelRealigner \  
-R Homo_sapiens_assembly19.fasta \  
-I file_sorted_MD.RG.bam \  
-targetIntervals file_sorted_MD.RG.intervals \  
-o file_sorted_MD.RG.IND.bam \  
-known 1000G_phase1.indels.b37.vcf
```

Recalibración de bases (GATK)

Los puntajes de calidad de base son estimaciones de errores para cada base llamada por los secuenciadores. Esto proporciona un puntaje de calidad en escala *Phred* a cada base (*Q score*) y su respectiva precisión (en porcentaje).

El Q score es una propiedad relacionada logaritmicamente con las posibilidades de error de llamado de bases (P).

$$Q = -10 \log_{10} P$$

Un valor Q30, refleja una probabilidad error de una base por cada 1000 y por ende, una exactitud del 99.9%. La estimación de puntajes de calidad es de suma importancia para el llamado de variantes, por lo que un Q score bajo puede incrementar la posibilidad de un resultado falso positivo.

Los puntajes de calidad asignados por lo secuenciadores están sujetos a errores sistemáticos, por lo que pueden existir valores sobre o subestimados. La recalibración de bases usa un algoritmo de *machine learning* para modelar dichos errores y ajustar los puntajes de calidad. El proceso de recalibración involucra 2 pasos: el primero crea un modelo de covariación entre los archivos en proceso de análisis y un conjunto de variantes ya conocidas (*BaseRecalibrator*); el segundo, a partir del archivo de recalibración generado en el primer

paso, se ajustan los puntajes de calidad (*ApplyBQSR*) y se reescriben las bases con puntajes de calidad ajustados (*PrintReads*).

- Estructura de comandos

Paso 1. BaseRecalibrator (GATK)

```
java -Xmx4g -jar GenomeAnalysisTK.jar \  
-T BaseRecalibrator \  
-R Homo_sapiens_assembly19.fasta \  
-I file_sorted_MD.RG.IND.bam \  
-knownSites:name,VCF 1000G_phase1.indels.b37.vcf \  
-knownSites:name,VCF dbsnp_138.b37.vcf \  
-o file_recalibration_report.grp
```

Paso 2. PrintReads (GATK)

```
java -Xmx4g -jar GenomeAnalysisTK.jar \  
-T PrintReads \  
-R Homo_sapiens_assembly19.fasta \  
-I file_sorted_MD.RG.IND.bam \  
-BQSR file_recalibration_report.grp \  
-o file_sorted_MD.RG.IND.RC.bam
```

- Tiempo aproximado por muestra: 300 minutos

Recolección de métricas (Picard tools)

Una vez que el archivo BAM está realineado y recalibrado, se realiza una recolección de métricas de los datos de secuencia a través de selección híbrida (*HS*). Esta técnica es la más usada para capturar secuencias específicas de exones y estimar el porcentaje de lecturas *on* y *off target* (cobertura).

- Estructura del comando:

```
java -jar picard.jar CollectHsMetrics \  
I=file_sorted_MD.RG.IND.RC.bam \  
O=file_sorted_MD.RG.IND.RC_hs_metrics.txt \  
R=Homo_sapiens_assembly19.fasta \  
METRIC_ACCUMULATION_LEVEL=ALL_READS BAIT_INTERVALS= targets_v4.bed \  
TARGET_INTERVALS=targets_v4.bed
```

Histogramas de cobertura (Picard tools)

Aunado a las métricas previamente obtenidas, se generaron histogramas (análisis cualitativo) para visualizar la profundidad de secuenciación y fracción de bases con la opción CoverageBed (*bedtools v2.27*) y *Rstudio*.

- Estructura del comando:

Paso 1. Ordenado de archivos por cromosoma (Samtools)

```
samtools sort \  
file_sorted_MD.RG.IND.RC.bam > file_sorted_MD.RG.IND.RC.sorted.bam
```

Paso 2. CoverageBed (bedtools)

```
coverageBed -hist \  
-abam intervals_for_bedtools.sorted.bed \  
-b file_sorted_MD.RG.IND.RC.sorted.bam > file.coverage.sort.hist.txt
```

Paso 3. Histogramas de cobertura (RStudio)

```
cov = read.table("521010310105-PB.coverage.sort.hist.txt", fill=T)  
gcov = cov[cov[,1] == 'all',]  
gcov_cumul = 1 - cumsum(gcov[,5])  
plot(gcov[2:401,2], gcov_cumul[1:400],  
     col='darkred', type='l', lwd=2,  
     xlab="Depth",  
     ylab="Fraction of capture target bases >= depth",  
     ylim=c(0,1.0))  
abline(v = 20, col = "gray60")  
abline(v = 50, col = "gray60")  
abline(v = 80, col = "gray60")  
abline(v = 100, col = "gray60")  
abline(h = 0.50, col = "gray60")  
abline(h = 0.90, col = "gray60")  
axis(1, at=c(20,50,80), labels=c(20,50,80))  
axis(2, at=c(0.90), labels=c(0.90))  
axis(2, at=c(0.50), labels=c(0.50))  
axis(1, at=c(20,50,80), labels=c(20,50,80))  
axis(2, at=c(0.90), labels=c(0.90))  
axis(2, at=c(0.50), labels=c(0.50))
```

Fase 2. Descubrimiento de variantes

75 archivos BAM realineados y recalibrados fueron incluidos para subsecuente descubrimiento de variantes. Las buenas prácticas del GATK para el descubrimiento de SNPs y INDELS germinales, recomiendan realizar el descubrimiento de variantes en varias muestras a la vez (*joint genotyping*), mejorando la detección de variantes con una mayor sensibilidad y precisión posible. Esta fase involucra 5 pasos principales descritos a continuación.

Llamado de variantes con *HaplotypeCaller* (GATK)

HaplotypeCaller es una herramienta del GATK que permite llamar SNP e INDELS germinales mediante un re-ensamble local de haplotipos en regiones con signos de variación. Para esto, el programa determina las regiones de operación (activas) con base en la presencia variación, para posteriormente, asignar el genotipo más probable. En este paso, el llamado de variantes primero se realiza de manera individual a través del modo GVCF

- Estructura del comando:

```
java -Xmx4g -jar gatk-package-4.1.7.0-local.jar HaplotypeCaller \  
-R Homo_sapiens_assembly19.fasta \  
-I file_sorted_MD.RG.IND.RC.bam \  
-O file.variants.g.vcf \  
-ERC GVCF
```

Consolidación de *g.vcf* con *GenomicsDBImport* (GATK)

Posterior al llamado de variantes, se generaron archivos *.g.vcf* (archivo que almacena variantes de secuencia y su calidad) individuales, los cuales, fueron subsecuentemente unificados. *GenomicsDBImport* es un sistema de almacenamiento y consulta de datos específicos para aplicaciones genómicas.

- Estructura del comando:

```
java -Xmx4g -jar gatk-package-4.1.7.0-local.jar GenomicsDBImport \  
-R Homo_sapiens_assembly19.fasta \  
-V sample1.variants.g.vcf \  
-V sample2.variants.g.vcf \  
-L targets_v4.bed  
--genomicsdb-update-workspace-path all-gvcfs_db
```

Genotipado en conjunto con *GenotypeGVCFs* (GATK)

El archivo consolidado con la información de todas las muestras, fue utilizado para realizar el genotipado en modo joint-genotyping, obteniendo un archivo .vcf

- Estructura del comando:

```
java -Xmx4g -jar gatk-package-4.1.7.0-local.jar GenotypeGVCFs \  
-R Homo_sapiens_assembly19.fasta \  
-V gendb://all-gvcfs_db \  
-O joint_genotyping.vcf.gz --tmp-dir=/tmp
```

Recalibración de variantes (GATK)

El archivo .vcf generado a partir del genotipado en conjunto, contiene una gran cantidad de variantes. La recalibración de variantes permite identificar y seleccionar solo aquellas con altos valores de anotación (propiedades de cada variante). Para este paso, se utilizó el algoritmo VQSR, un algoritmo muy estricto que permite identificar variantes con alta probabilidad de que sean reales a través de archivos de referencia que incluyen variantes genéticas reales (hapmap), falsos y verdaderos positivos (1000G_omni y 1000G_phase1) y de sitios conocidos (dbSNP).

Los filtros de anotación utilizados para el modelo de entrenamiento son:

- **QualbyDepth (QD<2.0)**
Normalización de calidad de variantes
- **FisherStrand (FS=0)**
Estimación de sesgo de cadena
Define si la variante está en cadena (+) o (-)
- **StrandOddRatio (SOR<3.0)**
Estimación de sesgo de cadena
Similar a FS pero aplicada al final de los exones
- **RMSMappingQuality (MQ≈60)**
Calidad de mapeo del promedio de todas las lecturas en un sitio específico para incluir D.E. de las calidades de mapeo
- **MappingQualityRankSumTest (MQRankSum -2.5 a 2.5, 0)**
Comparación de calidades de mapeo de lecturas en alelo alt y ref
- **ReadPosRankSumTest (ReadPosRankSum -1 3.5, 0)**
Compara si las posiciones del alelo ref y alt son diferentes en las lecturas

Este proceso se realizó a través de 2 pasos, primero la construcción del modelo “Gaussian Mixture” y gráficos de distribución y, finalmente la aplicación de filtros e impresión de variantes recalibradas. Cada uno de estos pasos fue realizado por separado para SNPs e INDELS, por lo que el archivo de salida .vcf recalibrado para SNPs, fue utilizado como archivo de entrada para poder recalibrar los INDELS.

- Estructura del comando:

Paso 1. *VariantRecalibrator* SNPs (GATK y Rstudio)

```
java -Xmx4g -jar gatk-package-4.1.7.0-local.jar VariantRecalibrator \  
-R Homo_sapiens_assembly19.fasta \  
-V joint_genotyping.vcf.gz \  
--resource:hapmap,known=false,training=true,truth=true,prior=15.0 hapmap_3.3.b37.vcf \  
--resource:omni,known=false,training=true,truth=false,prior=12.0 1000G_omni2.5.b37.vcf \  
--resource:1000G,known=false,training=true,truth=false,prior=10.0  
1000G_phase1.snps.high_confidence.b37.vcf \  
--resource:dbsnp,known=true,training=false,truth=false,prior=2.0 dbsnp_138.b37.vcf  
-tranche 100.0 -tranche 99.9 -tranche 99.0 -tranche 90.0 \  
-an QD -an FS -an MQRankSum -an ReadPosRankSum -an SOR -an MQ -mode SNP  
-O file.recal \  
--tranches-file file.tranches \  
--rscript-file file.plots.R
```

Paso 2. *VQSR* SNPs (GATK)

```
java -Xmx4g -jar gatk-package-4.1.7.0-local.jar ApplyVQSR \  
-R Homo_sapiens_assembly19.fasta \  
-V joint_genotyping.vcf.gz \  
-O file-recalibrated.vcf \  
--truth-sensitivity-filter-level 99.0 \  
--tranches-file file.tranches \  
--recal-file file.recal \  
-mode SNP
```

Paso 3. *VariantRecalibrator* INDELS (GATK y Rstudio)

```
java -Xmx4g -jar gatk-package-4.1.7.0-local.jar VariantRecalibrator \  
-R Homo_sapiens_assembly19.fasta \  
-V file-recalibrated.vcf \  
--resource:1000G,known=false,training=true,truth=false,prior=10.0 1000G_phase1.indels.b37.vcf \  
--resource:mills,known=false,training=true,truth=true,prior=12.0  
Mills_and_1000G_gold_standard.indels.b37.vcf \  
-an QD -an MQ -an MQRankSum -an ReadPosRankSum -an FS -an SOR  
--max-gaussians 3 \  
-mode INDEL \  
-O file-recalibratedSNPINDEL.recal \  
--tranches-file file-recalibratedSNPINDEL.tranches \  
--rscript-file file-recalibratedSNPINDEL.plots.R
```

Paso 4. VQSR INDELS (GATK)

```
java -Xmx4g -jar gatk-package-4.1.7.0-local.jar ApplyVQSR \  
-R Homo_sapiens_assembly19.fasta \  
-V file-recalibrated.vcf \  
-O file-recalibratedSNPINDEL.vcf \  
--truth-sensitivity-filter-level 99.0 \  
--tranches-file file-recalibratedSNPINDEL.tranches \  
--recal-file file-recalibratedSNPINDEL.recal \  
-mode INDEL
```

Fase 3. Anotación y filtrado de variantes

El archivo de salida `Piloto_Dic2020-recalibratedSNPINDEL.vcf`, corresponde al archivo unificado y recalibrado de variantes de los 75 individuos en proceso de análisis en modo *Joint Genotyping*. La anotación y filtrado de variantes es la fase en la cual, a cada variante contenida en un archivo `vcf`, le es asignada información biológica, funcional, poblacional y clínica. Dicha información es indispensable para poder realizar subsecuente filtrado e identificación de variantes de relevancia clínica y asociación con fenotipo. Dicho archivo de salida, es utilizado como archivo de entrada para la fase 3.

División del vcf unificado

La división del archivo `Piloto_Dic2020-recalibratedSNPINDEL.vcf` se realizó con el programa *vcftools 0.1.13* de acuerdo a la siguiente sintaxis para cada archivo `vcf` individual:

```
vcf-subset --exclude-ref -c sample1 in.vcf > out.vcf
```

De los 75 individuos, 41 archivos `vcf` correspondieron a casos índice y 34 archivos a padres biológicos. Posteriormente, se realizó la indización de cada `vcf` con la opción *IndexFeatureFile* del GATK siguiendo la estructura del siguiente comando:

```
gatk IndexFeatureFile \  
-I cohort.vcf.gz
```

Anotación de vcf individuales

Cada archivo `vcf` individual, fue utilizado como archivo de entrada para anotación de variantes con el programa *Variant Effect Predictor, VEP (Ensamble)*.(74)

Búsqueda de variantes en genes de predisposición

El proceso para obtener archivos txt con variantes anotadas en los genes de predisposición a cáncer se realizó mediante los siguientes pasos y comandos:

1. **Paso 1.** Filtrado de variantes en genes de predisposición a cáncer

```
grep -w -F -f\  
lista_de_genes.txt \  
archivo_variantes_annotado_#individuo.txt >  
archivo_variantes_annotado_#individuo_genes.txt
```

2. **Paso 2.** Agregar encabezado al archivo

```
echo -e \  
'head_1\thead2\... ' |  
cat - file1.txt >  
file2.txt
```

Anexo 2. Validación de variantes de relevancia clínica

Mutaciones patogénicas o probablemente patogénicas

521010310107 (MUTYH: c.928C>T (p.Gln310Ter))

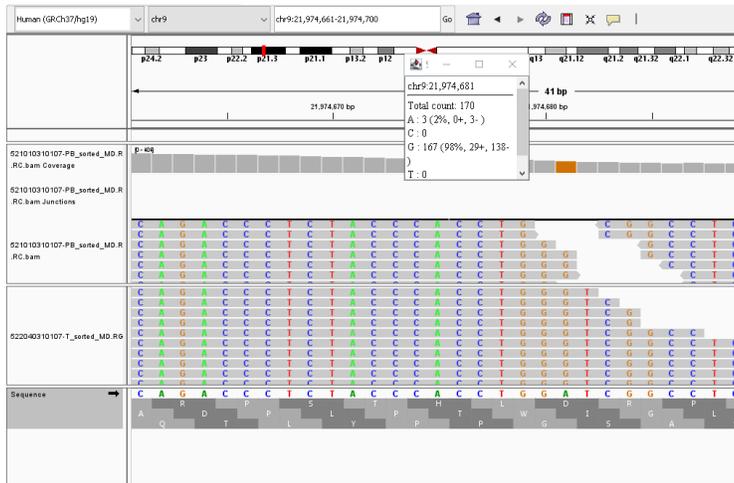


Germinal
Profundidad: 7x
VAF: 0.57

Somático
Profundidad: 8x
VAF: 0.25

5

521010310107 (CDKN2A:c.146T>C (p.Ile49Thr))

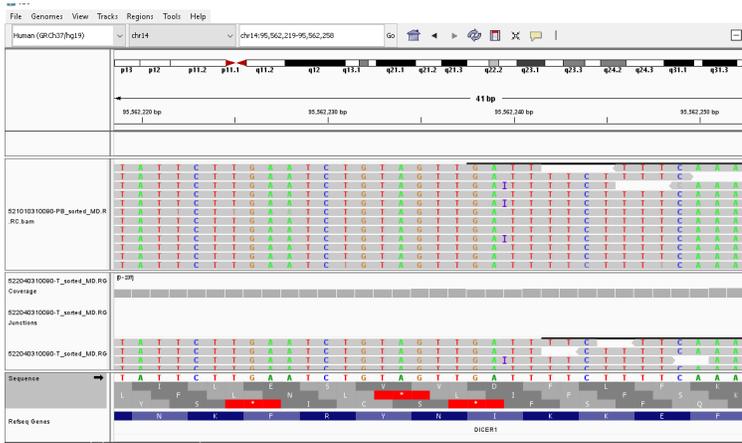


Germinal
Profundidad: 168x
VAF: 0.98

Somático
Profundidad: 292x
VAF: 0.99

6

521010310090 (DICER1::c.5017dup (p.Ile1673fs))



Germinal
Profundidad: 36x
VAF: 0.42

Somático
Profundidad: 82x
VAF: 0.99

7

521010310090 (CHEK2:c.707T>C (p.Leu236Pro))

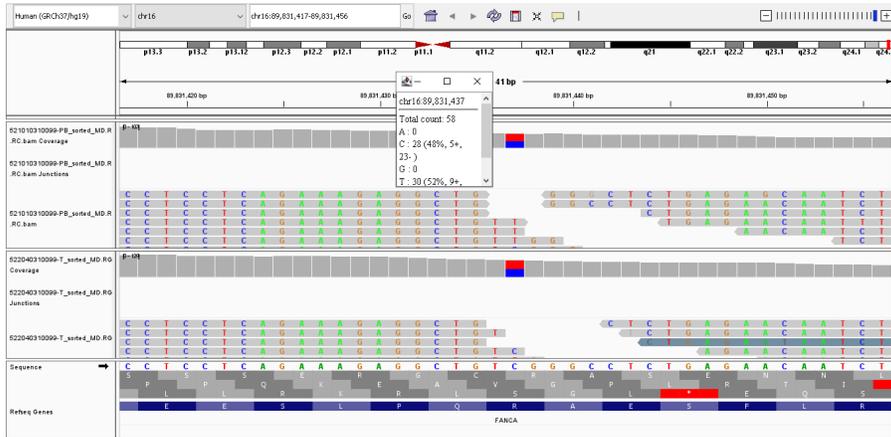


Germinal
Profundidad: 87x
VAF: 0.5

Somático
Profundidad: 104x
VAF: 0.45

8

521010310099 (FANCA:c.2639G>A (p.Arg880Gln))



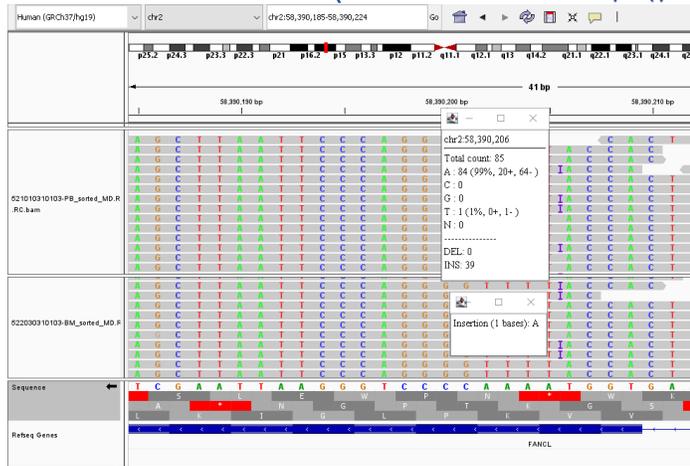
Germinal
Profundidad: 58x
VAF: 0.52

Somático
Profundidad: 84x
VAF: 0.54

9

Variantes candidatas

521010310103 (FANCL:c.779dup (p.Pro262ThrfsTer5))

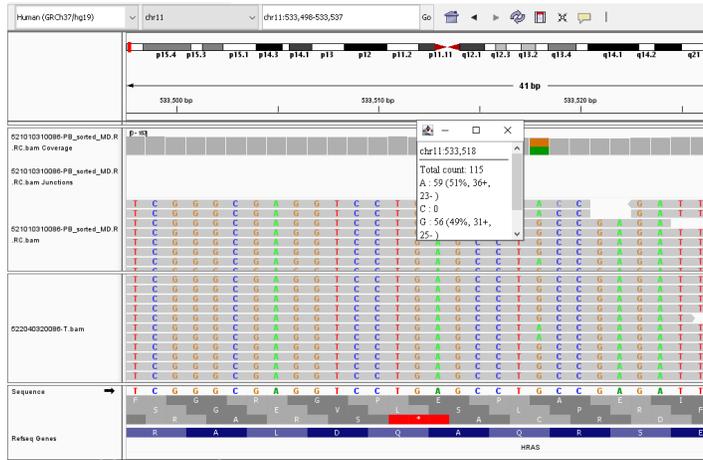


Germinal
Profundidad: 84x
VAF: 0.46

Somático
Profundidad: 119x
VAF: .49

15

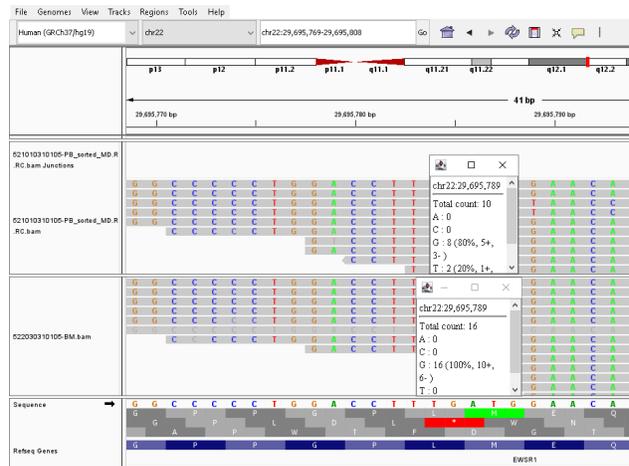
5210103100086 (HRAS: c.385C>T (p.Gln129Ter))



Germinal
 Profundidad: 115x
 VAF: 0.5

Somático
 Profundidad: 132x
 VAF: 0.11

521010310105 (EWSR1: c.1894G>T (p.Glu632Ter))



Germinal
 Profundidad: 10x
 VAF: 0.2

Somático
 Profundidad: 16x
 VAF: 0

NO CONFIRMADA

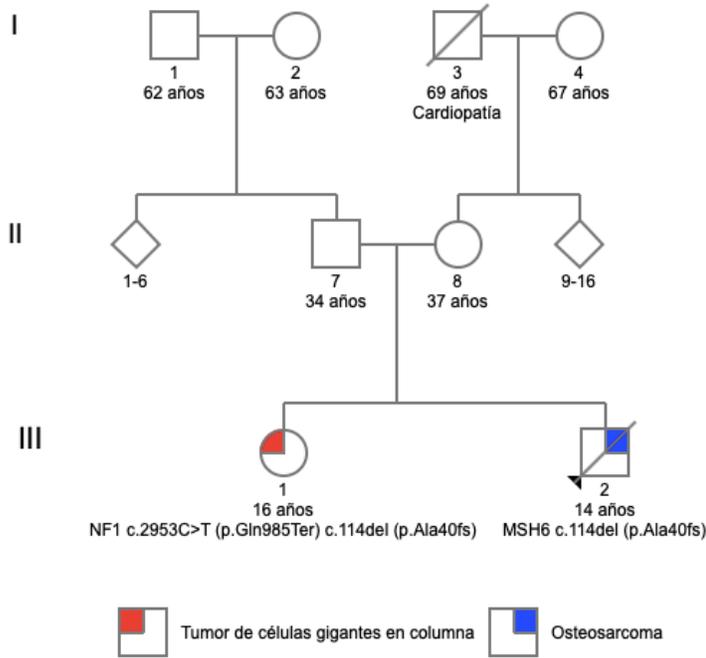
Anexo 3. Determinación de sexo

#Sample	#SnpsAut	#SNPsX	#SnpsY	NrAut	NrX	NrY	x-rate	y-rate	Err(x-rate)	Err(y-rate)	Sex
1030010001 / CPM	1150639	49704	32670	4375442	121476	5447	0.64271184	0.04384554	0.00186947	0.00059445	Female
1030011001 / SHL_L002	1150639	49704	32670	3640309	53136	27003	0.33790791	0.26125447	0.00147656	0.00159574	Male
1030011002 / HVL_L002	1150639	49704	32670	3932512	58260	28632	0.34296371	0.25643162	0.00143138	0.00152097	Male
1030011003 / SVL_L002	1150639	49704	32670	4322185	117413	3731	0.62886957	0.03040268	0.00186004	0.00049795	Female
52000000346	1150639	49704	32670	3283682	49069	26305	0.34593449	0.28214163	0.00157333	0.00174655	Male
52000000394	1150639	49704	32670	4545482	63845	33061	0.32515819	0.25616866	0.00129587	0.00141397	Male
52000000395	1150639	49704	32670	4407121	65168	34341	0.34231598	0.27444032	0.00135082	0.00148671	Male
521010310033	1150639	49704	32670	2485812	60861	41	0.56678527	0.00058091	0.00232542	9.07E-05	Female
521010310037	1150639	49704	32670	2528095	31716	12504	0.29042418	0.17419895	0.00164097	0.00156168	Male
521010310047	1150639	49704	32670	2324674	51532	238	0.51317169	0.00360583	0.00228552	0.00023374	Female
521010310048-PB	1150639	49704	32670	2278153	52688	448	0.53539779	0.00692604	0.00235931	0.00032726	Female
521010310051-PB	1150639	49704	32670	2261700	51896	58	0.53118602	0.0009032	0.00235834	0.0001186	Female
521010310053-PB	1150639	49704	32670	2302838	27737	10445	0.27883279	0.15974785	0.00168428	0.00156662	Male
521010310063-PB	1150639	49704	32670	3712658	81455	660	0.50790274	0.00626108	0.00179901	0.00024373	Female
521010310066-PB	1150639	49704	32670	3018916	72061	633	0.55258234	0.00738487	0.0020829	0.00029355	Female
521010310069-PB	1150639	49704	32670	3268094	72080	520	0.51058492	0.00560401	0.00192264	0.00024577	Female
521010310073-PB	1150639	49704	32670	3795454	43324	20785	0.26424852	0.19287514	0.00127677	0.00134149	Male
521010310076-PB	1150639	49704	32670	2551859	58248	283	0.5284113	0.00390589	0.00221428	0.00023219	Female
521010310084-PB	1150639	49704	32670	2678632	31167	13194	0.26935789	0.17348159	0.0015346	0.00151402	Male
521010310086-PB	1150639	49704	32670	2299705	26685	10832	0.26862277	0.1658924	0.00165392	0.00159769	Male
521010310089-PB	1150639	49704	32670	2626841	62461	242	0.55045636	0.00324468	0.00222854	0.00020859	Female
521010310090-PB	1150639	49704	32670	3267814	70246	489	0.49763626	0.00527037	0.00189767	0.00023835	Female
521010310091-PB	1150639	49704	32670	2304377	51416	342	0.51652638	0.00522712	0.00230322	0.00028267	Female
521010310105-PB	1150639	49704	32670	3177366	68866	600	0.50174767	0.0066508	0.00193259	0.00027154	Female
521010310106-PB	1150639	49704	32670	2063167	46450	60	0.52119361	0.00102425	0.00244535	0.00013223	Female
521010310107-PB	1150639	49704	32670	1668956	43831	147	0.60797293	0.00310215	0.00294187	0.00025587	Female
521010310112	1150639	49704	32670	2155847	49188	90	0.52818854	0.00147033	0.00240856	0.00015499	Female
521010310119-PB	1150639	49704	32670	1577414	20991	10113	0.30805991	0.22580017	0.00214037	0.00225254	Male
521010310120	1150639	49704	32670	2825985	33869	13433	0.27744715	0.16741452	0.00151658	0.00144789	Male
521010310126-PB	1150639	49704	32670	1953895	69351	387	0.82167345	0.00697589	0.00317502	0.00035464	Female
521010310137-PB	1150639	49704	32670	1663237	43736	178	0.60874117	0.00376926	0.00294883	0.00028253	Female
521010310143-PB	1150639	49704	32670	1918292	51236	252	0.61831282	0.00462675	0.00276786	0.00029148	Female
521010310149	1150639	49704	32670	2233701	26433	10573	0.27394865	0.16671057	0.00169493	0.00162514	Male
521010310150	1150639	49704	32670	1789832	41481	66	0.5365185	0.00129874	0.00266462	0.00015987	Female
521010310151	1150639	49704	32670	1981255	23166	9185	0.27068141	0.1632784	0.00178878	0.00170763	Male

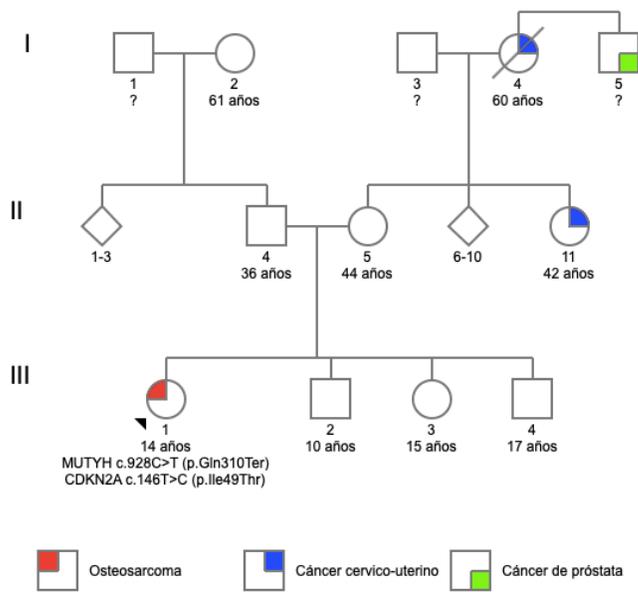
521010310152	1150639	49704	32670	2033291	46161	283	0.52556135	0.00490204	0.00247378	0.00029142	Female
521010310153	1150639	49704	32670	2279867	26835	10945	0.27248327	0.16908155	0.00167313	0.00162005	Male
521010310161	1150639	49704	32670	1967667	23662	9130	0.27838613	0.16342148	0.00182061	0.00171427	Male
521010310162	1150639	49704	32670	1911741	44272	164	0.53610251	0.00302138	0.00257724	0.00023594	Female
521010310164	1150639	49704	32670	2014576	46340	88	0.53250063	0.00153847	0.00250196	0.00016401	Female
521010310165	1150639	49704	32670	2123592	24732	9971	0.26960994	0.16537033	0.00172433	0.00165999	Male
521010310166	1150639	49704	32670	2046391	47124	84	0.53309091	0.00144571	0.00248384	0.00015774	Female
521010310167	1150639	49704	32670	2163693	25756	10194	0.27556911	0.16593536	0.00172727	0.00164736	Male
521010310168	1150639	49704	32670	2142219	25587	10161	0.27650517	0.16705618	0.00173889	0.0016612	Male
521010310171	1150639	49704	32670	1713196	39513	65	0.53392555	0.00133628	0.00271683	0.00016575	Female
521010310172	1150639	49704	32670	2341347	53427	300	0.52825395	0.00451279	0.00231133	0.00026056	Female
521010310175	1150639	49704	32670	2113577	25511	10079	0.27941979	0.1679536	0.00175994	0.00167693	Male
521010310184	1150639	49704	32670	2737754	32338	12900	0.27344279	0.16595305	0.00152954	0.00146457	Male
521010310185	1150639	49704	32670	2595972	59163	400	0.52759167	0.00542688	0.00219364	0.00027136	Female
521010310203	1150639	49704	32670	1721433	38924	65	0.52344986	0.00132988	0.00268301	0.00016495	Female
521010310209	1150639	49704	32670	2292181	26760	10440	0.27026198	0.16041373	0.00166174	0.00157354	Male
521010310213	1150639	49704	32670	2186530	51505	95	0.54530779	0.00153023	0.00243093	0.000157	Female
521010310214	1150639	49704	32670	2467193	29259	11552	0.27453904	0.16490887	0.00161449	0.00153791	Male
521010310215	1150639	49704	32670	67795	1573	15	0.53712925	0.00779262	0.0136992	0.00201227	Female
521010310216	1150639	49704	32670	75565	1053	407	0.32259337	0.1896984	0.01001028	0.00942829	Male
521010310217	1150639	49704	32670	48811	566	364	0.26843953	0.26264772	0.01134859	0.01381772	Male
521010310219	1150639	49704	32670	2360145	54085	147	0.53050062	0.00219366	0.00230711	0.00018094	Female
521010310220	1150639	49704	32670	2382078	52290	452	0.50817163	0.00668301	0.00224655	0.00031437	Female
521010310222	1150639	49704	32670	2314611	27505	10969	0.27509417	0.1669087	0.00166856	0.00159743	Male
521010310249	1150639	49704	32670	2528395	60741	54	0.55614081	0.00075221	0.00228349	0.00010236	Female
521010310254	1150639	49704	32670	2648223	64973	72	0.56797094	0.00095756	0.0022554	0.00011285	Female
521010310283	1150639	49704	32670	2356647	29283	11226	0.28765292	0.16777238	0.00169139	0.00158723	Male
521010310284	1150639	49704	32670	2609521	64637	382	0.57341381	0.00515576	0.00228318	0.00026381	Female
521010310292	1150639	49704	32670	2895082	70199	65	0.56132942	0.00079076	0.00214415	9.81E-05	Female
521010310302	1150639	49704	32670	2581911	32326	12754	0.28984009	0.17397831	0.00162212	0.00154434	Male
521010310314	1150639	49704	32670	2762683	33830	13137	0.28347756	0.16747697	0.00155064	0.00146466	Male
521010310328	1150639	49704	32670	2739627	68269	65	0.57687252	0.00083563	0.00223518	0.00010365	Female
521010320076	1150639	49704	32670	2295224	52117	363	0.52565655	0.00557021	0.00232856	0.00029238	Female
521010320114-PB	1150639	49704	32670	1781679	20890	8156	0.2714293	0.16122698	0.00188894	0.00178933	Male
521010320129	1150639	49704	32670	2452507	28526	10564	0.26926405	0.15170786	0.0016035	0.0014792	Male
521010320218	1150639	49704	32670	2408399	54045	288	0.51948718	0.00421167	0.00225952	0.00024819	Female
521100310123-PB	1150639	49704	32670	2149196	48156	56	0.51870703	0.0009177	0.00239006	0.00012263	Female

Anexo 3. Familiogramas

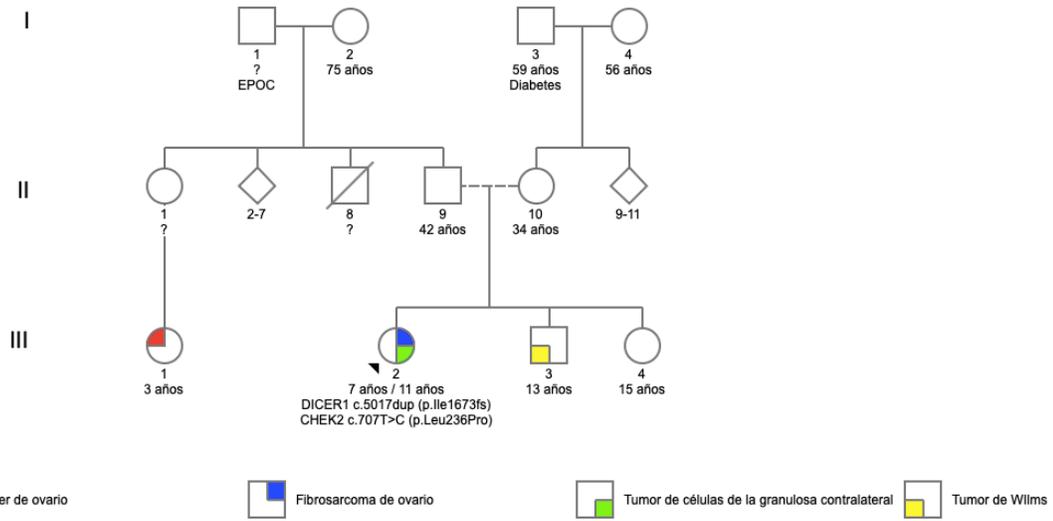
Familia 1



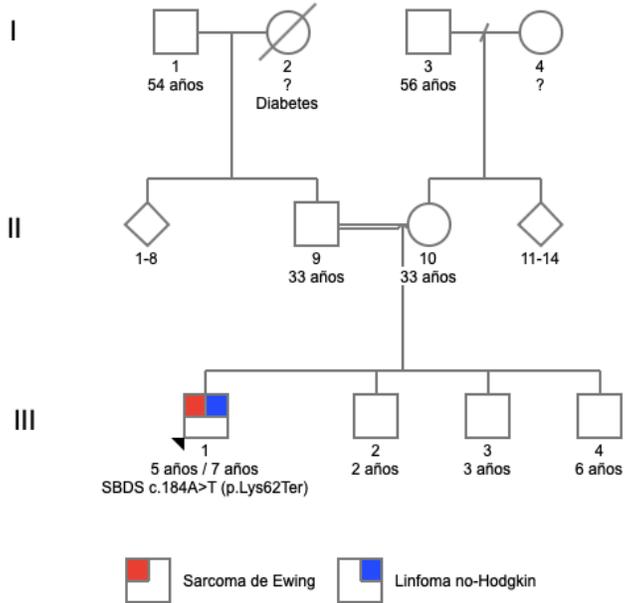
Familia 2



Familia 3



Familia 4



Familia 5

