



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MÉDICAS,
ODONTOLÓGICAS Y DE LA SALUD
BIOQUÍMICA CLÍNICA

Identificación de genes candidatos en pacientes con síndrome Patterson-
Stevenson-Fontaine y síndrome de Roifman-Chitayat

TESIS

QUE PARA OPTAR POR EL GRADO DE:
MAESTRÍA EN CIENCIAS

PRESENTA

GUADALUPE FERNANDA GODÍNEZ ZAMORA

TUTOR

DR. JESÚS AGUIRRE HERNÁNDEZ
HOSPITAL INFANTIL DE MÉXICO FEDERICO GÓMEZ

CIUDAD UNIVERSITARIA, CD. MX. SEPTIEMBRE, 2019



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.

Jurado de examen

Dra. Verónica Fabiola Morán Barroso
Hospital General de México "Eduardo Liceaga"

Dr. Jesús Aguirre Hernández
Hospital Infantil de México "Federico Gómez"

Dr. Francisco Javier Mena Estrada Instituto de Oftalmología "Fundación Conde de Valenciana I.A.P"

Dr. Javier Tadeo Granados Riverón
Hospital Infantil de México "Federico Gómez"

Dr. Miguel Ángel Alcántara Ortigoza
Intituto Nacional de Pediatría

Agradecimientos

A mi familia; mis padres y abuela por estar en cada momento, por su apoyo, amor y comprensión. Sin ustedes nada de esto habría sido posible.

A mi tutor por haber aceptado ser mi guía en este proceso, por su apoyo y por toda su paciencia.

A los miembros del Laboratorio de Genómica, Genética y Bioinformática del HIMFG por los conocimientos que me compartieron, por su ayuda y continua disposición, por ayudarme a crecer profesional y personalmente, y por haber me acompañado en esta etapa.

M. en C. Patricia Baeza Capetillo

Ma. Ángeles Telles Rangel

Pablo M. Valencia Segura

Al servicio de Genética del HIMFG:

Dra. América Villaseñor Domínguez y Dra. Cristina Armijos por toda su ayuda Clínica y por su amistad.

Dra. Verónica Fabiola Morán Barroso por su apoyo.

Resumen

Introducción: Las pacientes estudiadas en este trabajo tienen impresión diagnóstica de dos enfermedades clasificadas dentro del grupo de las displacias esqueléticas: el síndrome de Patterson-Stevenson-Fontaine (PSF) enfermedad de la cual se desconoce la causa genética, y el síndrome de Roifman-Chitayat (RC). Estas son enfermedades raras de las que se tienen muy pocos reportes de casos. Por este motivo se decidió estudiar a las afectadas y a sus padres para buscar variantes que puedan estar relacionadas con las enfermedades.

Objetivo: Identificar la causa genética de dos enfermedades que se sospecha tienen origen monogénico en pacientes del Hospital Infantil de México Federico Gómez (HIMFG).

Métodos: Se estudiaron los exomas completos de las pacientes y sus padres para buscar las variantes causantes de ambos fenotipos. Durante el análisis de variantes se revisaron principalmente variantes de un solo nucleótido. Las variantes priorizadas a lo largo del estudio fueron comparadas con reportes y frecuencias en bases de datos públicas y con la información de la base de datos del laboratorio de genómica, genética y bioinformática del Hospital.

Resultados: Durante el análisis de variantes para la familia de la paciente con diagnóstico de síndrome PSF se encontraron variantes candidatas, sin embargo, todas fueron descartadas debido a su alta frecuencia y a la función de los genes. En la familia de la paciente con impresión diagnóstica de síndrome RC se encontró una variante nueva en el gen *NBAS*, mutaciones homocigotas en este gen han sido señaladas como causantes de dos enfermedades; el síndrome SOPH y falla hepática aguda. La revisión de las características clínicas de la afectada mostró que en realidad la paciente presenta síndrome SOPH acompañado de falla hepática aguda.

Conclusión: La secuenciación de exomas completos por técnicas de segunda generación es una herramienta muy útil para el diagnóstico de enfermedades genéticas raras. Especialmente cuando se trata de variantes pequeñas como la mutación puntual encontrada en *NBAS*. Gracias a esta herramienta ha sido posible conocer la base genética de muchas enfermedades con herencia mendeliana en los últimos años. La técnica tiene ciertas limitaciones debido a que se secuencian fragmentos cortos y solo regiones codificantes. Es posible que la variante responsable del fenotipo de la paciente diagnosticada con síndrome PSF este en alguna región no codificante o que sea de una longitud mayor a la que puede ser detectada por la técnica.

Índice

Índice	5
Introducción	8
Síndrome de Patterson-Stevenson-Fontaine	8
Síndrome de Roifman-Chitayat.....	10
Búsqueda de genes candidatos y variantes patogénicas	11
Secuenciación de segunda generación	11
Enfermedades a estudiar	12
Planteamiento del problema.....	13
Pregunta de investigación	13
Justificación	13
Objetivo general	13
Objetivos particulares	13
Hipótesis.....	13
Tipo de estudio	14
Metodología	14
Métodos experimentales	14
1.1 Extracción de DNA	14
1.2 Evaluación y cuantificación de DNA	14
2. Síntesis de librerías.....	15
2.1 Tagmentación.....	15
2.2 Limpieza del DNA tagmentado.....	15
2.3 Adición de índices y amplificación del producto de la tagmentación	16
2.4 Limpieza del DNA amplificado	16
2.5 Hibridación de sondas	16
2.6 Captura de las sondas hibridadas.....	17
2.7 Segunda hibridación	17
2.8 Segunda captura.....	17
2.9 Lavado de los productos de la segunda captura	17

2.10 Amplificación de la librería enriquecida	18
2.11 Limpieza de la amplificación de la librería enriquecida	18
2.12 Preparación de la bandeja de reactivos.....	18
Métodos bioinformáticos	19
1. Demultiplexing y Alineamiento	19
2. Filtrado y Anotación	19
3. Priorización	19
4. Análisis de variantes	20
5. Análisis de variantes con Exomiser.....	20
6. Análisis de variantes sinónimas y análisis por segregación.	20
7. Análisis de variantes estructurales.....	20
Resultados	21
Datos clínicos de las familias estudiadas.....	21
Familia 1. Probable síndrome de Patterson-Stevenson-Fontaine	21
Familia 2. Probable síndrome de Roifman-Chitayat	24
Familia 1. Síndrome de Patterson-Stevenson-Fontaine.....	25
1.1 Datos generales de la secuenciación.....	25
1.2 Cálculo de frecuencia alélica máxima esperada en una población de individuos sanos....	26
1.3 Primer análisis de las variantes	27
1.4 Análisis de variantes con Exomiser.....	29
1.5 Análisis de variantes por trío.....	30
1.6 Análisis de variantes sinónimas.....	31
1.7 Análisis de variantes estructurales.....	32
Familia 2. Síndrome de Roifman-Chitayat.....	33
1.1 Datos generales de la secuenciación	33
1.2 Cálculo de frecuencia alélica máxima esperada en una población de individuos sanos	34
1.3 Análisis de las variantes	35
1.2 Análisis de la variante en Exomiser	39
1.2 Análisis <i>in silico</i> de la variante en el sitio aceptor de splicing	40
Discusión	43
Familia 1: Síndrome de Patterson-Stevenson-Fontaine.....	43
Análisis de variantes en Exomiser	44

Análisis de variantes por trio	44
Análisis de variantes sinónimas	44
Variantes estructurales (SV)	45
Otras posibilidades.....	45
Familia 2: Síndrome Roifman-Chitayat.....	47
Síndrome SOPH.....	49
Conclusiones.....	53
Perspectivas.....	53
Anexo I Grupos de displasias esqueléticas.....	54
Anexo II Cartas de consentimiento informado.....	56
Anexo III. Carta de aprobación del comité de bioseguridad	57
Anexo IV. Carta de aprobación consentimiento informado para toma de fotografía.....	58
Referencias.....	59

Introducción

Las enfermedades monogénicas son enfermedades de herencia mendeliana causadas por mutaciones en un solo gen. Generalmente las mutaciones que causan estas enfermedades se encuentran dentro de la secuencia de genes que codifican proteínas, con funciones fisiológicas importantes, causando que estas enfermedades tengan implicaciones multisistémicas (Oud *et al.* 2017). Las enfermedades monogénicas son poco frecuentes, pero comprenden una amplia gama de padecimientos, entre estos las displasias esqueléticas. Se calcula que existen 7000 enfermedades monogénicas de las cuales 436 son displasias esqueléticas (Hamosh *et al.* 2000; Bonafe *et al.* 2015).

Se estima que las displasias esqueléticas se presentan en 1/5 000 recién nacidos (Geister & Camper 2015). Las displasias esqueléticas son un grupo amplio y heterogéneo de enfermedades que pueden causar algunas anomalías que no comprometen la vida, pero también pueden ser letales. Hasta el momento de la publicación de la última nosología se conocía la base genética y mecanismos moleculares de 56 de estas enfermedades y se habían identificado 364 genes cuyas mutaciones son causa de alguna displasia esquelética (Bonafe *et al.* 2015). Las mutaciones en los genes involucrados causan diversas anormalidades óseas, como osteopenia y edad ósea retrasada, y/o malformaciones como displasia metaepifisiaria o espondiloepifisiaria en huesos de las extremidades, cráneo, tórax, etc. (Firth & Hurst 2017).

Las displasias esqueléticas se clasifican en 42 grupos de acuerdo a los mecanismos moleculares, evidencias radiológicas y fenotipo de cada enfermedad. Los 42 grupos de acuerdo a la clasificación de 2015 (Bonafe *et al.* 2015) se muestran en el anexo I. Las displasias esqueléticas que aquí se describen pertenecen a grupos distintos. El síndrome de Patterson-Stevenson-Fontaine forma parte del grupo de las disostosis acrofaciales. El síndrome de Roifman-Chitayat no se encuentra incluido en la nosología debido a los pocos casos que se han reportado, pero podría estar dentro del grupo de las displasias metaepifisiarias por las características radiológicas de las pacientes reportadas.

Síndrome de Patterson-Stevenson-Fontaine

El síndrome de Patterson-Stevenson-Fontaine (OMIM, 183700) es una disostosis facial (anomalía cráneo facial). Este grupo de enfermedades se caracterizan por la presencia de malformaciones en los huesos mandibulares. Las disostosis faciales se clasifican en 2 grupos; las disostosis mandibulofaciales que solo afectan los huesos cráneo faciales y las disostosis acrofaciales que se presentan con anormalidades cráneo faciales y en las extremidades (Wieczorek 2013; Terrazas *et al.* 2017).

El síndrome de Patterson-Stevenson-Fontaine se clasifica dentro del grupo de las disostosis acrofaciales, de las cuales se han descrito más de 18 entre las que se encuentran el síndrome de Miller (OMIM263750), el síndrome de Nager (OMIM154400) y la disostosis acrofacial tipo Cincinnati (OMIM616462) que han

sido caracterizados clínicamente y molecularmente (Bonafe *et al.* 2015; Terrazas *et al.* 2017).

El síndrome de Patterson-Stevenson-Fontaine fue descrito por primera vez en 1964. En el reporte del caso se describen las características de padre e hijo de una familia inglesa en la que ambos presentaban malformaciones mandibulares, y anomalías en dedos de manos y pies (Patterson & Stevenson 1964; Wilkie & Goodacre 1997). El estudio del caso comenzó ya que un niño de aproximadamente 8 meses fue referido al servicio de cirugía plástica del Hospital Churchill para ser intervenido debido a una malformación (pie dividido) en el pie izquierdo. Los padres no estaban emparentados.

Sin embargo, el padre tenía anomalías en los pies, tres dedos pequeños en el pie derecho le habían sido amputados durante la infancia. Presentaba marcada micrognatia y mandíbula asimétrica, así como paladar alto y dientes con mala oclusión. También mostraba malformaciones en el oído medio e interno causando hipoacusia severa por lo que usaba un aparato auditivo de lado izquierdo. Las orejas parecían ser muy largas y con lóbulos pequeños. Ninguno de los dos presentaba alteraciones oftalmológicas como se observa en otras disostosis.

Durante el seguimiento del paciente en la edad adulta se observó que el presentaba fenotipo similar al de su padre pero con audición normal. Lo cual podría indicar que la enfermedad tiene expresividad variable. Posterior a lo reportado por Patterson y Stevenson otros casos de esta enfermedad han sido descritos clínicamente, pero aún no se conoce la base genética de esta disostosis acrofacial (Wilkie & Goodacre 1997).

La enfermedad está bien caracterizada clínicamente, el diagnóstico clínico se basa en criterios como la presencia de anomalías en los huesos mandibulares (Figura 1A), en las extremidades y anomalías en dedos de manos y pies (Figura 1C) como braquidactilia, sindactilia y/o ectrodactilia (Wieczorek 2013). En algunas ocasiones el fenotipo incluye malformaciones del oído externo (Figura 1B), medio o interno, causando hipoacusia y problemas de lenguaje (Wilkie & Goodacre 1997).

De acuerdo con la literatura se sospecha que la enfermedad presenta patrón de herencia autosómico dominante y expresividad variable.

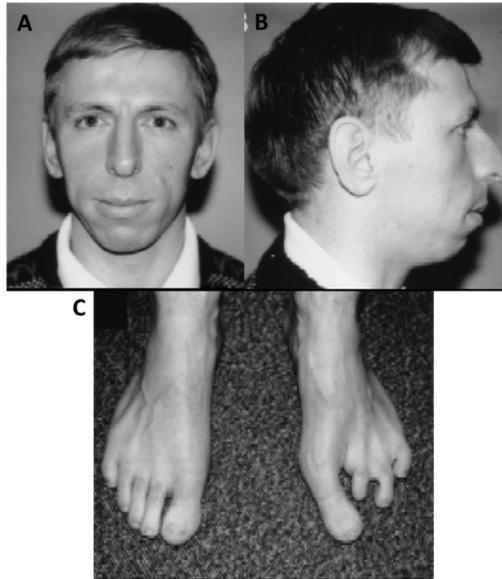


Figura 1. **Fenotipo de los pacientes con síndrome Patterson-Stevenson-Fontaine.** A) Características faciales. B) Malformaciones en el oído externo y retrognatia. C) Anomalías en pie izquierdo. Imágenes tomadas de Wilkie and Goodacre (1997).

Síndrome de Roifman-Chitayat

El síndrome de Roifman–Chitayat (OMIM, 613328) fue reportado por primera vez en 2009, en un par de hermanas hijas de padres consanguíneos de ascendencia inglesa (Roifman & Chitayat 2009). Al nacer ambas niñas presentaron hipotonía y dificultad para la alimentación. Una de las afectadas nació con un solo riñón. Ambas presentaron infecciones invasivas y recurrentes desde los 5 meses debido a una inmunodeficiencia combinada. Desde el primer año mostraron retraso global del desarrollo. Posteriormente ambas niñas comenzaron a tener convulsiones mioclónicas, hipoplasia y atrofia bilateral del nervio óptico, estenosis cervical y lumbar. La displasia metaepifisiaria (Figura 2B) (anormalidades en la metáfisis de los huesos largos) se volvió más evidente después de los 7 años al igual que los bordes irregulares de las cervicales.

Las dos afectadas mostraban dimorfismos faciales como puente nasal deprimido, labio inferior delgado, hipertelorismo, y frente amplia y ancha (Figura 2A).

De acuerdo con las características del caso se sospecha que la enfermedad tiene patrón de herencia autosómico recesivo y penetrancia completa (Roifman & Chitayat 2009). Para estudiar a las pacientes se realizaron algunas pruebas para identificar los sitios donde se encontraba la base molecular de la enfermedad. Los autores reportaron dos locus candidatos. Años más tarde en 2018, los autores reportaron datos del exoma de estas pacientes. No se han publicado nuevos reportes de casos de pacientes con la enfermedad.

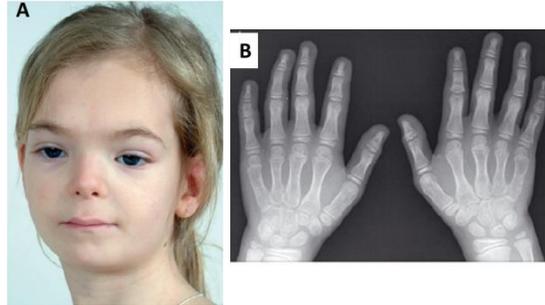


Figura 2. **Fenotipo de los pacientes con síndrome Roifman-Chitayat.** A) Características faciales de una de las afectadas. B) Se observa displasia metaepifisiaria en las falanges. Imágenes tomadas de Roifman and Chitayat (2009).

Búsqueda de genes candidatos y variantes patogénicas

Algunas metodologías como la búsqueda de genes candidatos y el análisis de ligamiento son utilizadas comúnmente en el estudio de polimorfismos, pero una de las desventajas que las hacen poco viables para el estudio de enfermedades monogénicas es que para tener éxito es necesario contar con múltiples pacientes con la enfermedad. En el caso de las enfermedades monogénicas esto no siempre es posible ya que en ocasiones pueden transcurrir varios años antes de encontrar a otro paciente con las mismas características (Kulkarni & Pfeifer ; Boycott *et al.* 2017).

Secuenciación de segunda generación

La secuenciación de segunda generación (NGS, por **N**ext **G**eneration **S**equencing) es una herramienta que permite conocer con detalle datos específicos de la secuencia de genomas completos o regiones amplias de los mismos, su organización, y variantes (cambios de una o más bases con respecto al genoma de referencia) (Shendure & Ji 2008). El desarrollo de estas nuevas técnicas de secuenciación ha hecho posible el descubrimiento de mutaciones patogénicas y el estudio de la etiología de enfermedades genéticas.

A diferencia de otras metodologías, la secuenciación de segunda generación permite estudiar casos aislados o familias pequeñas. Los métodos más usados para el descubrimiento de genes relacionados con enfermedades monogénicas son la comparación de información de varios individuos afectados, no relacionados o el estudio de tríos familiares (afectado y sus padres). Para esto han sido desarrollados algunos paneles, uno de ellos es el exoma completo, que abarca aproximadamente el 2% del genoma humano (45 millones de pb). Permite secuenciar específicamente los genes que codifican proteínas, que se piensa son los causantes de la mayoría de las enfermedades genéticas. La secuenciación de

exoma completo ha permitido el descubrimiento de numerosos genes asociados a enfermedades raras (Boycott *et al.* 2017).

Hay varias técnicas (o “plataformas químicas”) de secuenciación de segunda generación, en este estudio la secuenciación de los exomas se realizó mediante la técnica de secuenciación por síntesis de puentes de Illumina.

Un exoma puede tener más de 30,000 variantes después de haber sido alineado con el genoma de referencia, y más de 10,000 pueden tener efectos diversos como alteraciones en sitios de splicing, ganancia o pérdida de codones de paro, pequeñas indels, variantes en sentido equivocado entre otras (Robinson *et al.* 2014). Estas variantes son analizadas bioinformáticamente para así lograr identificar variantes en genes candidatos elegidos mediante diversas estrategias, como la frecuencia, comparación de los resultados del exoma de múltiples individuos afectados por la misma enfermedad, relación familiar o restringiendo la lista de genes candidatos a genes de una sola vía.

La conclusión de proyectos que colocan sus datos en bases de datos públicas como el proyecto de los 1000 genomas, ExAc, gnomAD, y otros más, han apoyado la implementación de estrategias de análisis y el descubrimiento de variantes en genes relacionados con enfermedades (Consortium 2004). Ya que la información de un número considerable de datos de individuos de diversas ascendencias permiten identificar variantes comunes en las poblaciones representadas y facilitan la identificación de variantes raras.

Por otro lado la técnica también tiene algunas limitaciones, se ha observado que en algunos casos las variantes causantes de la enfermedad no se encuentran dentro de regiones codificantes, o podrían estar en regiones pequeñas que no han sido incluidas en el panel, las variantes también podrían estar en regiones difíciles de secuenciar como segmentos ricos en GC o regiones con repetidos.

Enfermedades a estudiar

En este trabajo de tesis se estudiaron dos síndromes: el síndrome de Patterson-Stevenson-Fontaine cuya causa genética no se conoce, y el síndrome de Roifman-Chitayat. Ambas enfermedades pertenecen a las llamadas displasias óseas. El primer síndrome ha sido reportado en menos de 10 individuos (ORPHA:2439), y se considera que se hereda de manera autosómica dominante. Por su parte, el síndrome de Roifman-Chitayat ha sido reportado en un par de pacientes y la información limitada sugiere que segrega de manera autosómica recesiva.

Planteamiento del problema

El síndrome de Patterson-Stevenson-Fontaine y de Roifman-Chitayat son enfermedades genéticas raras cuya base genética no se conoce.

Para identificar los genes causantes de las enfermedades se estudiaron los 22 mil genes codificantes en los individuos afectados y en sus progenitores sanos mediante secuenciación de segunda generación.

Pregunta de investigación

¿Cuáles son los genes cuyas variantes son responsables de los fenotipos de los pacientes a estudiar?

Justificación

Identificar los genes causantes de las enfermedades y sus variantes presenta una oportunidad para aportar información nueva de la causa genética de estas dos enfermedades raras, o en su caso la acusa de los fenotipos de los pacientes estudiados, así como la comprensión de las funciones de los genes que resulten involucrados.

Objetivo general

- Identificar genes candidatos, y sus variantes, asociadas con las patologías de los pacientes a estudiar.

Objetivos particulares

- Identificar mutaciones en las secuencias exónicas de pacientes con diagnóstico de probable síndrome Patterson-Stevenson-Fontaine y Roifman-Chitayat.
- Conocer el efecto de las variantes encontradas y su implicación en los procesos celulares para comprender la etiología de las enfermedades.
- Asociar las variantes candidatas con los rasgos fenotípicos de los pacientes.

Hipótesis

Al secuenciar el exoma completo, de los pacientes y sus padres, se obtendrán datos de utilidad para identificar mutaciones en los genes causantes de las enfermedades de los pacientes a estudiar.

Tipo de estudio

Este es un estudio: observacional, descriptivo y transversal.

Fue realizado en el Laboratorio de Genómica, Genética y Bioinformática del Hospital Infantil de México Federico Gómez.

En un periodo de 2 años.

Metodología

Métodos experimentales

1.1 Extracción de DNA

Se obtuvieron muestras de sangre periférica, de los pacientes y de sus padres, en tubos con EDTA de sodio. Las muestras se almacenaron a 4°C hasta el momento de ser procesadas. A partir de las muestras se realizó la extracción de DNA mediante columnas Quick DNA Universal Kit (Zymo Research, Irvine, CA, E.E.U.U.). Esta se llevó a cabo a partir de 200 µL de sangre, mezclándola con 200 µL de solución de lisis y 20 µL de proteinasa K. La mezcla se incubó a 55°C por 10 min. Al finalizar la incubación, se agregaron 420 µL de amortiguador de unión de DNA, y se transfirió toda la solución a la columna. Esta fue colocada en un tubo de recolección, fue centrifugada 1 min a 12,000 x g, y se desechó el eluyente del tubo de recolección. Se transfirió la columna a un nuevo tubo recolector y a la columna se le agregaron 400 µL de amortiguador de prelavado de DNA. Se centrifugó a 12,000 x g durante 1 min y se desechó el contenido del tubo recolector.

A la columna se le añadieron 700 µL de amortiguador de lavado de DNA genómico y se centrifugó durante 1 min a 12,000 x g. Nuevamente se desechó el contenido del tubo de recolección. A la columna se le agregaron 200 µL de amortiguador de lavado de DNA genómico, se centrifugó 1 min a 12,000 x g y se desechó el tubo de recolección, junto con su contenido.

La columna se transfirió a un tubo nuevo para microcentrífuga y se le añadieron 50 µL de amortiguador de elución de DNA. Se incubó 5 min a temperatura ambiente y después se centrifugó a 12,000 x g un minuto, para eluir el DNA.

1.2 Evaluación y cuantificación de DNA

Para verificar la integridad del DNA se revisaron 3 µL en un gel de agarosa al 1%, observándose DNA de alto peso molecular.

El DNA fue cuantificado por fluorimetría, con un instrumento Qubit v. 3.0 (Thermo Fisher Scientific, CDMEX), de acuerdo con las indicaciones del manual (MAN0010866, revisión A.0). Por cada muestra a cuantificar, y por cada estándar de referencia, se preparó una solución de trabajo diluyendo 1 µL de fluoróforo intercalante de DNA de cadena doble y de alta sensibilidad (Qubit dsDNA HS, ThermoFisher Scientific) en 199 µL de amortiguador Qubit. De la mezcla se tomaron 198 µL que fueron transferidos a un tubo de 0.5 mL de pared clara y delgada. A este tubo se le agregaron 2µL de la muestra a cuantificar. Este

procedimiento se siguió para cada una de las muestras a cuantificar. Para la preparación de cada una de las 2 soluciones de referencia (a 0 ng/uL y a 60 ng/uL) se usaron 190 uL de la solución de trabajo, y se agregaron 10µL de solución de referencia.

2. Síntesis de librerías

La síntesis de la librería se realizó usando el kit Nextera DNA Exome de Illumina (San Diego, CA, E.E.U.U.), siguiendo las indicaciones del procedimiento descrito en la guía de referencia (guía 1000000039018 v00, oct. 2017).

2.1 Tagmentación

Se realizó la tagmentación del DNA genómico con ayuda de una transposasa que corta el DNA genómico en fragmentos de aproximadamente 350 pares de bases (pb) y al mismo tiempo añade adaptadores a los extremos libres de cada segmento. Estos adaptadores tienen secuencias complementarias a los oligos de la celda de flujo, lo que permite que los fragmentos generados por la tagmentación hibriden con los oligos de la celda de flujo.

Para la tagmentación, cada muestra de DNA genómico fue diluida con Tris HCl 10 mM pH 8.5 hasta obtener una solución a 5 ng/µL. En una placa de pozos profundos se agregaron, en el siguiente orden, 10 µL de DNA a 5 ng/uL, 25 µL de amortiguador de tagmentación, y 15 µL de enzima de tagmentación. Se agitó por 1 min y se centrifugó durante 1 min a 280 x g.

Las 9 muestras (5 de la familia con síndrome de Patterson-Stevenson-Fontaine, más el trío de la familia con síndrome de Roifman-Chitayat) se incubaron a 58°C durante 10 min. Para detener la acción de la transposasa, se añadieron 15 µL de amortiguador de paro de tagmentación. Se agitó y se centrifugó 1 min a 280 x g. Finalmente, se incubó a temperatura ambiente durante 4 min.

2.2 Limpieza del DNA tagmentado

A los 65 µL de la mezcla de la tagmentación de cada muestra se añadieron 65 µL de perlas magnéticas de purificación. Se agitó durante 1 min y se incubó a temperatura ambiente por 8 min. Posteriormente se centrifugó a 280 x g por un min y la placa se colocó sobre una gradilla magnética hasta que las perlas formaron una pastilla y el líquido quedó transparente (aproximadamente después de 5 min). Se desechó el sobrenadante transparente.

Sin quitar la placa de la gradilla magnética, las pastillas se lavaron 2 veces con 200 µL de etanol al 80%, recién preparado. Se incubó sobre la gradilla magnética durante 30 s y se desechó el sobrenadante. Este lavado se repitió una vez más y después los residuos de etanol se eliminaron con una pipeta de 20 µL, dejando secar las muestras por 10 min. La placa se removió de la gradilla magnética y las pastillas se resuspendieron con 22.5 µL de amortiguador de resuspensión (RSB); se agitó por 1 min y se incubó a temperatura ambiente durante 2 min para resuspender los fragmentos de DNA limpios.

Se centrifugó a 280 x g por 1 min y la placa con las muestras se colocó sobre la gradilla magnética hasta que el líquido mostró una apariencia clara y transparente (2 a 5 min). Se transfirieron 20 µL de sobrenadante a tubos de PCR.

2.3 Adición de índices y amplificación del producto de la tagmentación

Después de la tagmentación se agregaron índices a los fragmentos de DNA mediante una PCR. Los índices son secuencias conocidas de 8 pb que funcionan como una etiqueta que permite distinguir las muestras de cada individuo. Esto permite que en una etapa posterior las 9 muestras del estudio puedan ser mezcladas en un solo tubo (multiplexing).

A cada muestra, por separado, se le agregaron 5 uL de un índice 1 (i7) y 5 uL un índice 2 (i5), de tal manera que cada muestra tuviera una combinación única de ambos índices. Se añadieron 20 µL de mezcla de amplificación de librería, se agitó por 1 min y se centrifugó a 280 x g durante 1 min. Las muestras se llevaron al termociclador con el siguiente programa: 72°C por 3 min, 98°C, durante 30 s, y después 10 ciclos de 98°C por 10 s, 60°C por 30 s, 72°C durante 30 s; para finalizar, las muestras se incubaron a 72°C por 5 min.

2.4 Limpieza del DNA amplificado

Terminada la amplificación, las muestras se centrifugaron a 280 x g durante 1 min y se transfirieron a pozos limpios de la placa de pozo profundo. Se agregaron 90 µL de perlas magnéticas de purificación. Se agitó por 1 min y se incubó a temperatura ambiente por 10 min. Después se centrifugó la placa a 280 x g por 1 min, se colocó sobre la gradilla magnética hasta que las perlas formaron una pastilla y el sobrenadante quedó transparente (2 a 5 min). Este último se desechó.

Cada muestra se lavó 2 veces como se describió en la sección 2.2 Limpieza del DNA tagmentado (p. 15). Una vez secas las pastillas, se agregaron 27 µL de RSB a cada una, se agitaron durante 1 min, se incubaron a temperatura ambiente por 2 min y se centrifugaron a 280 x g por 1 min. La placa se colocó sobre la gradilla magnética hasta que el sobrenadante quedó transparente, y se transfirieron 25 µL del mismo a un tubo de 0.6 mL.

2.5 Hibridación de sondas

Todas las muestras fueron cuantificadas por duplicado con el fluorómetro Qubit v 3.0, como se describió en la sección 1.2 Evaluación y cuantificación de DNA (p. 14). Se combinaron 500 ng de cada muestra en un solo tubo (multiplexing).

Para efectuar el enriquecimiento de regiones exónicas, se agregaron en el siguiente orden, en un tubo para PCR, 40 µL de la solución multiplex, 50 µL de amortiguador de enriquecimiento de hibridación y 10 µL de sondas para exoma (CEX). Se agitó por 1 min, se centrifugó a 280 x g durante 1 min y el tubo se colocó en un termociclador. La hibridación se llevó a cabo con el programa NRC HYB (95°C 10 min; seguido de 94°C, 1 min; 92°C, 1 min; 90°C, 1 min; y continuando con el descenso de 2°C en la temperatura de hibridación hasta llegar a los 58°C. Finalmente, se incubó a 58°C durante 2 h.

2.6 Captura de las sondas hibridadas

En este paso se eliminaron de la librería los fragmentos que no hibridaron con las sondas.

El tubo de PCR se centrifugó a 280 x g durante 1 min y se transfirió el contenido (100 µL) a un pozo limpio de la placa de pozo profundo.

Se agregaron 250 µL de perlas magnéticas con estreptavidina, se agitó por 5 min, se incubó a temperatura ambiente por 25 min y se centrifugó a 280 x g por 1 min. La placa se colocó sobre la gradilla magnética hasta que el sobrenadante quedó transparente, y éste se desechó. Se retiró la placa de la gradilla magnética y se realizó el primer lavado con 200 µL de solución de lavado para enriquecimiento. Se agitó por 4 min y se resuspendió con micropipeta. Se colocó la placa en el sistema de incubación Hybex (SciGene, Sunnyvale, CA, E.E.U.U) precalentado a 50°C, y se incubó durante 30 min. Inmediatamente después de transcurrido este tiempo, se transfirió la placa a la gradilla magnética y se esperó a que el sobrenadante estuviera transparente para desecharlo. Se retiró la placa de la gradilla. Este lavado se repitió una vez más, comenzando con la adición de 200 µL de solución de lavado para enriquecimiento.

Para la elución se mezclaron, en un tubo de 1.6 mL para microcentrífuga, 28.5 µL amortiguador de elución y 1.5 µL de NaOH 2N. Se añadieron 23 µL de esta mezcla a la librería, se agitó por 2 min, se incubó a temperatura ambiente 2 min y se centrifugó a 280 x g durante 1 min. La placa se colocó sobre la gradilla magnética hasta que el sobrenadante quedó transparente. Se transfirieron 21 µL de éste a un pozo limpio de la placa. Se añadieron 4 µL de solución ET2, se agitó 1 min y se agregaron 5 µL de amortiguador de resuspensión, seguido de agitación por 1 min y centrifugación a 280 x g por 1 min.

2.7 Segunda hibridación

Se agregaron 15 µL de amortiguador de resuspensión, 50 µL de amortiguador de enriquecimiento de hibridación), y 10 µL de sondas para exoma (CEX). Se agitó 1 min y se centrifugó a 280 x g por 1 min. La mezcla se transfirió a un tubo para PCR y se llevó al termociclador para efectuar el segundo ciclo de enriquecimiento, La hibridación se llevó a cabo como se describió en la sección 2.5 Hibridación de sondas (p. 16), excepto que ahora la incubación final a 58°C duró 19 h.

2.8 Segunda captura

Terminada la segunda hibridación, el tubo de PCR se centrifugó a 280 x g durante 1 min, y se transfirió el contenido (100 µL) a un pozo limpio de la placa de pozo profundo. Se agregaron 250 µL de perlas magnéticas con estreptavidina para capturar las sondas hibridadas y se prosiguió tal como se describió en la sección 2.6 Captura de las sondas hibridadas (p. 17).

2.9 Lavado de los productos de la segunda captura

Se agregaron 45 µL de perlas magnéticas de purificación, se agitó 1 min y se incubó a temperatura ambiente 10 min. Se centrifugó a 280 x g por 1 min. La placa se colocó sobre la gradilla magnética hasta que el sobrenadante quedó

transparente. Éste se desechó. Se lavó dos veces con 200 μ L de etanol al 80% fresco, como se describió anteriormente (sección 2.2 Limpieza del DNA tagmentado; p. 15). Para resuspender, se agregaron 27.5 μ l de amortiguador de resuspensión, se agitó 1 min y se incubó a temperatura ambiente 2 min. Se centrifugó a 280 x g por 1 min, se colocó la placa sobre la gradilla magnética y, cuando el líquido quedó transparente, se transfirieron 25 μ L del sobrenadante a un tubo de PCR.

2.10 Amplificación de la librería enriquecida

Al tubo de PCR con la librería se le agregaron 5 μ L de oligos para PCR y 20 μ L de la mezcla para amplificación. Se agitó 1 min y se centrifugó a 280 x g durante 1 min. Se puso en el termociclador a 98°C por 30 s, para continuar con 10 ciclos de 98°C por 10 s, 60°C por 30 s y 72°C durante 30 s. Para finalizar, se incubó a 72°C por 5 min, y después se conservó a 10°C hasta retirar el tubo del termociclador.

2.11 Limpieza de la amplificación de la librería enriquecida

El tubo con la librería amplificada se centrifugó a 280 x g durante 1 min. Se transfirieron, a un pozo limpio de la placa de pozo profundo, 50 μ l de la librería enriquecida. Se agregaron 90 μ l de perlas magnéticas de purificación, se agitó 1 min, y se incubó a temperatura ambiente 10 min. Se centrifugo a 280 x g por 1 min. Posteriormente la placa se colocó en la gradilla magnética hasta ver el líquido transparente y se desechó el sobrenadante. El botón se lavó 2 veces con etanol al 80% recién preparado, tal como se describió anteriormente (2.2 Limpieza del DNA tagmentado; p. 15).

Se retiró la placa de la gradilla magnética y se agregaron 32 μ l de amortiguador de resuspensión. Se agitó 1 min y se incubó a temperatura ambiente 2 min. Se centrifugó a 280 x g por 1 min. Se colocó la placa sobre la gradilla magnética y, cuando el sobrenadante quedó transparente, se transfirieron 30 μ l a un tubo de 0.6 uL para microcentrífuga.

La calidad de la librería, y la distribución de los fragmentos, se revisaron en un gel de agarosa al 4%. Para la cuantificación, que se hizo por triplicado, se tomaron 2 μ l y se utilizó el fluorómetro Qubit v3.0 (véase p. 14). Se obtuvo una concentración 57 nM y a partir de este valor se llevaron a cabo diluciones hasta obtener una librería a 1.3 pM, de acuerdo con las instrucciones del manual NextSeq System Denature and Dilute Libraries Guide (15048776 Rev. E, mayo 2015).

2.12 Preparación de la bandeja de reactivos

La bandeja de reactivos para secuenciar se retiró del congelador y se descongeló en un refrigerador durante 2 horas, aproximadamente. Posteriormente el cartucho se terminó de descongelar colocándolo en una bandeja con agua destilada a temperatura ambiente. La librería, con una concentración final de 1.3 pM, se inyectó en el pozo número 10 de la bandeja de reactivos, y ésta se introdujo en el secuenciador NextSeq 500 (Illumina).

Métodos bioinformáticos

1. Demultiplexing y Alineamiento

Al término de las 29 horas de secuenciación se inició el proceso de demultiplexing, en el cual se separan las lecturas correspondientes a cada muestra, con base en la secuencia de los índices. Las lecturas de cada muestra se almacenan en archivos FASTQ separados. Estos archivos contienen la secuencia de las lecturas, junto con los valores de calidad correspondientes a cada base. Para el alineamiento se utilizó el genoma de referencia hg19.

El alineamiento y la identificación de variantes se realizaron con el software IssacEnrichment (Illumina), que incorpora versiones modificadas del programa Burrows-Wheeler Aligner (BWA), para el alineamiento (Li & Durbin 2009), y del programa GATK, para la identificación de variantes (McKenna *et al.* 2010; DePristo *et al.* 2011; Van der Auwera *et al.* 2013). Para cada muestra, se obtuvieron archivos bam con las lecturas alineadas (Li *et al.* 2009), y VCF (VCFv4.1) con la lista de variantes (Danecek *et al.* 2011).

2. Filtrado y Anotación

Las variantes encontradas, se analizaron mediante VariantStudio (v.3.0, Illumina) y bcftools. Las variantes de cada individuo fueron filtradas por calidad, profundidad ($\geq 10x$), frecuencia del alelo alternativo (≥ 0.19), efecto de la variante (variante de sentido equivocado, sin sentido, variante en sitio de splicing, pérdida o ganancia de codones de paro, etc.). También se filtraron por frecuencia, reteniéndose únicamente aquellas variantes con una frecuencia menor 0.01 en bases de datos públicas como ExAC (Exome Agregation Consortium) (Lek *et al.* 2016) y el Proyecto de los 1000 genomas (Abecasis *et al.* 2012; Auton *et al.* 2015).

Se aplicaron filtros de segregación en cada familia de acuerdo con el tipo de herencia esperado. En la familia con síndrome de Patterson-Stevenson-Fontaine se asumió una variante *de novo* en la madre de la propósita, y herencia autosómica dominante en la transmisión de la variante de la madre afectada a la propósita. En la familia con síndrome de Roifman-Chitayat se asumió herencia autosómica recesiva debido los datos de la enfermedad encontrados en la literatura y por la consanguinidad de los padres.

Las variantes obtenidas con VariantStudio (v3.0) fueron reanotadas con Variant Effect Predictor (VEP) en Ensembl (McLaren *et al.* 2016; Yates *et al.* 2016), en las dos versiones del genoma humano (hg19 y hg38). Para la conversión de las coordenadas de hg19 a hg38 se utilizó el convertidor de Ensembl. En esta tesis, todas las coordenadas mostradas corresponden a hg19, a menos de que se indique otra cosa.

3. Priorización

Para la priorización de variantes se usó la base de datos del Laboratorio de Genómica, Genética y Bioinformática del HIMFG que hasta el momento del análisis constaba de 124 muestras (1 genoma, 103 exomas y 20 muestras con la secuencia codificadora de 4,813 genes). Se priorizaron genes candidato de

acuerdo con la frecuencia y la distribución de las variantes en la base de datos del hospital mediante un programa desarrollado en laboratorio.

En el caso de la familia con síndrome de Patterson-Stevenson-Fontaine se realizó un segundo proceso de priorización con el software bcftools mediante filtros de segregación *de novo* y dominante en ambos tríos de la familia respectivamente.

4. Análisis de variantes

Se revisó la función y estructura de cada gen candidato. Se corroboró la frecuencia de cada variante en bases de datos públicas (ExAc, gnomAD y el proyecto de los 1000 genomas), y se revisó el genotipo. Se revisó también si los genes habían sido asociados con alguna enfermedad o fenotipo particular en bases de datos como OMIM (Amberger *et al.* 2015), ClinVar (Landrum *et al.* 2014), MalaCards (Rappaport *et al.* 2017) y Orphanet (Ayme 2003).

5. Análisis de variantes con Exomiser

Se usó el software Exomiser (Robinson *et al.* 2014b) que consiste en introducir los archivos VCF de la familia a analizar indicando quienes son los afectados y describiendo el fenotipo de los pacientes con términos HPO (por **H**uman **P**henotype **O**ntology) (Kohler *et al.* 2017). Para la priorización se utilizó información de fenotipos y genotipos de ratón, pez cebra (*Danio rerio*), humano, y redes de interacciones de proteínas. El algoritmo prioriza las variantes teniendo en cuenta la coincidencia con el fenotipo, el efecto de la variante, la frecuencia de esta, y la información que se encuentra en otras bases de datos (v. gr., STRING (Szklarczyk *et al.* 2015).

6. Análisis de variantes sinónimas y análisis por segregación.

Para la familia con síndrome de Patterson-Stevenson-Fontaine se realizó un segundo análisis en el cual se incluyeron todas las variantes de los 5 integrantes de la familia, incluyendo variantes sinónimas. Estas variantes fueron obtenidas después de un filtro de calidad estricto. Posteriormente se filtraron las variantes de acuerdo con la segregación esperada. Este análisis se llevó a cabo con el software bcftools, buscando una variante *de novo* en el trío formado por la madre afectada y los abuelos, y una variante con herencia autosómica dominante en el trío formado por la niña afectada y sus padres. De estas dos listas, se retuvieron únicamente las variantes presentes en la paciente y su madre.

7. Análisis de variantes estructurales

En la familia con síndrome de Patterson-Stevenson-Fontaine se realizó un análisis de variantes estructurales. A partir de los archivos VCF con información de las variantes estructurales se aplicaron filtros de calidad y de segregación para cada uno de los tríos en los que se dividió la familia para su análisis (*de novo* para la madre afectada y los abuelos maternos; herencia dominante para la transmisión de la madre a la hija afectada). Se retuvieron las variantes que pasaron el filtro de calidad y con la segregación esperada, y que se hallaron únicamente en las dos afectadas.

Resultados

Datos clínicos de las familias estudiadas

A continuación, se presentan las características y datos clínicos de los pacientes estudiados. La familia 1 se estudió porque la propósa fue diagnosticada con probable síndrome de Patterson-Stevenson-Fontaine y comparte algunas características con su madre. No se conoce el gen asociado con este síndrome. En la familia 2 se estudió una paciente con impresión diagnóstica de síndrome de Roifman-Chitayat.

Familia 1. Probable síndrome de Patterson-Stevenson-Fontaine

Paciente femenina de 3 años 2 meses de edad. Hija de padres jóvenes, no consanguíneos. Residentes de Chilpancingo, Guerrero. Obtenida vía vaginal a las 37 semanas de gestación, sin complicaciones. Presencia de dismorfias faciales como retrognatía (Figura 4A). Sindactilia en la mano y pie derechos, ectrodactilia en el pie izquierdo (Figura 4D y E), malformaciones de oído externo (Figura 4C), intelecto normal, audición normal y leve retraso del lenguaje.

En la madre de la paciente se observó oligodactilia y camptodactilia en la mano izquierda (Figura 5C y D), malformaciones de oído externo (Figura 5B), hipoacusia y alteración del lenguaje; con fenotipo similar al de la paciente. Padre aparentemente sano. Abuelos maternos aparentemente sanos. Por la segregación del fenotipo, se consideró que podría haber una mutación *de novo* en la madre de la propósa. Se muestra el árbol genealógico de la familia en la Figura 3.

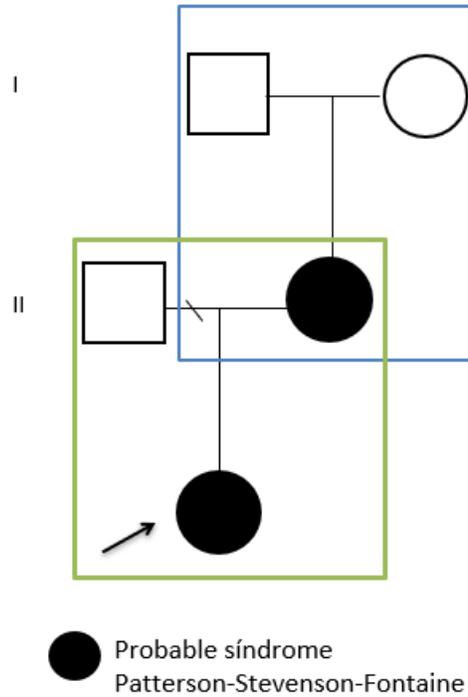


Figura 3. **Árbol genealógico de la familia 1. Paciente con diagnóstico de síndrome de Patterson-Stevenson-Fontaine.** Para el análisis de las variantes se dividió el árbol genealógico en dos tríos. El primero, en el recuadro azul, formado por la madre afectada y los abuelos maternos sanos. En este trío se asumió que la madre tenía una variante *de novo*. El segundo trío, en el recuadro verde, se incluyó a la propósita afectada, a la madre afectada y al padre sano. En este segundo trío se consideró herencia dominante.



Figura 4. **Paciente con probable síndrome de Patterson-Stevenson-Fontaine.** A) Dismorfias faciales; micro-retrognatia. B) No presentó paladar hendido. C) Anomalías en el pabellón auricular. D) Sindactilia en mano derecha. E) Sindactilia en pie derecho y ectrodactilia en el pie izquierdo.



Figura 5. **Madre de la propósita.** A) Microretrognatia. B) Malformaciones de oído externo. C y D) Oligodactilia y camptodactilia en mano izquierda.

Familia 2. Probable síndrome de Roifman-Chitayat

Paciente femenina de 3 años 11 meses de edad. Hija de padres jóvenes aparentemente sanos, consanguíneos (primos segundos), originarios y residentes de Pochutla, Oaxaca. Se muestra el árbol genealógico de la familia en la Figura 6. Obtenida vía vaginal a las 38 semanas de gestación, sin complicaciones.

Desde los primeros meses de vida la paciente presentó infecciones recurrentes principalmente en vías aéreas, deshidratación, fiebre, vómito y, durante las crisis infecciosas, niveles altos de enzimas hepáticas –alanina aminotransferasa (ALT) y aspartato aminotransferasa (AST)– hasta 8 veces más elevados que los valores de referencia.

También presentó retraso global en el desarrollo, hipotonía, desnutrición crónica, hidrocefalia y una inmunodeficiencia posiblemente primaria que no pudo ser confirmada debido a que la paciente falleció por una crisis infecciosa.

En la evaluación oftalmológica se reportó atrofia bilateral del nervio óptico. Las características faciales de la paciente incluyeron puente nasal deprimido, labio inferior delgado, frente amplia y ancha, exoftalmos, pestañas largas y rizadas (Figura 7A), y piezas dentales con coloración ámbar y opalescente (Figura 7B).

Dentro de las anomalías óseas que fueron encontradas, la paciente presentó edad cronológica ósea retrasada, de al menos 1 año, en huesos del carpo. También se observaron pseudoepífisis en las falanges distales (Figura 7D), presencia de huesos wormianos (Figura 7C), osteopenia (Figura 7E y F) y compresión del canal medular C7 a T5.

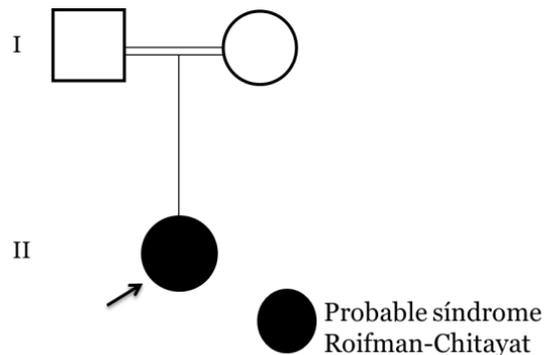


Figura 6. **Árbol genealógico de la familia 2, paciente con probable síndrome de Roifman-Chitayat.**

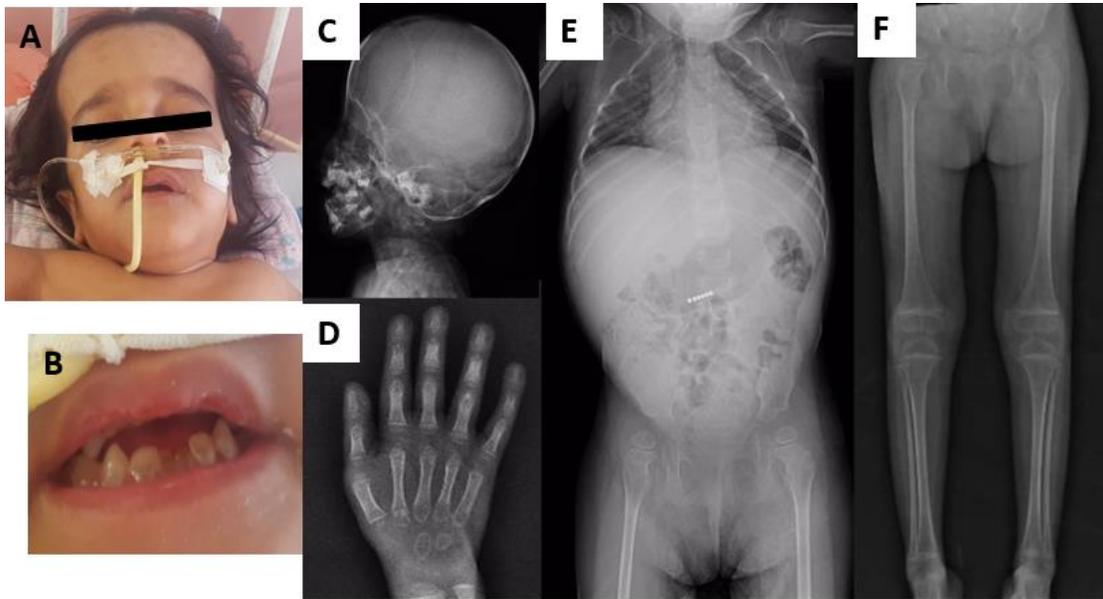


Figura 7. **Paciente con diagnóstico probable de síndrome Roifman-Chitayat.** A) Rasgos faciales de la paciente. B) Dentina con coloración ámbar y opalescente. C) Huesos Wormianos. D) Edad cronológica ósea retrasada en huesos del carpo y presencia de pseudoepífisis en las falanges distales. E) Osteopenia en costillas, e hígado crecido. F) Osteopenia en el fémur y huesos de las extremidades inferiores especialmente en metafisis y epífisis.

Se obtuvieron muestras de sangre periférica de los individuos descritos de ambas familias. Todas las muestras fueron procesadas en el mismo experimento de secuenciación, añadiéndose diferentes combinaciones de índices (secuencias de 8 pares de bases) para poder mezclarlas en un solo tubo y hacer los dos ciclos de enriquecimiento con la mezcla de las muestras. Una vez terminada la secuenciación, los índices permitieron separar las lecturas correspondientes a cada individuo, dirigiendo las lecturas de cada uno a archivos FASTQ separados.

Familia 1. Síndrome de Patterson-Stevenson-Fontaine

1.1 Datos generales de la secuenciación.

La Tabla 1 muestra las estadísticas correspondientes a los cinco individuos secuenciados de esta familia. Se obtuvo una cobertura promedio de 117x en la propósita y la distribución de la cobertura para esta muestra se ilustra en la Figura 8.

Tabla 1. Resumen de datos obtenidos en la secuenciación de los exomas de la familia con síndrome de Patterson-Stevenson-Fontaine.

	Abuelo	Abuela	Padre	Madre	Paciente
Total de lecturas	126,442,718	124,236,666	100,306,844	113,219,118	117,482,970
Porcentaje de bases con $\geq Q30$	76.15%	77.14%	77.91%	76.36%	77.37%
Longitud promedio de las lecturas	128 pb	123 pb	110 pb	127 pb	124 pb
Porcentaje de lecturas alineadas	89.64%	90.29%	88.61%	89.75%	89.98%
Regiones blanco cubiertas	98.24%	97.15%	95.86%	97.70%	97.46%
Profundidad promedio	131x	124x	95x	110x	117x

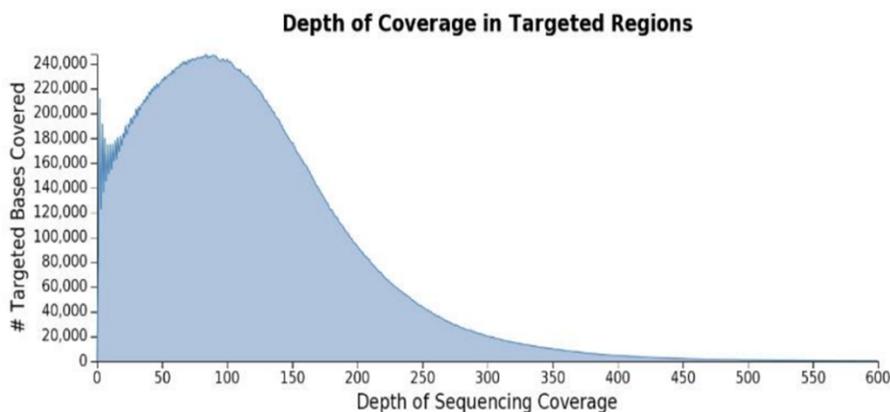


Figura 8. **Distribución de la profundidad en la propósita.** La profundidad promedio fue de 117x.

1.2 Cálculo de frecuencia alélica máxima esperada en una población de individuos sanos

El síndrome de Patterson-Stevenson-Fontaine es muy poco frecuente, lo mismo que todas las enfermedades raras o monogénicas (menos de 5 individuos afectados por cada 10,000, de acuerdo con la definición europea) (Forman *et al.* 2012). Consecuentemente, se espera que las variantes asociadas con este síndrome sean mucho muy raras en la población en general. Como se explica en la siguiente sección, durante el análisis de las variantes en esta familia se excluyeron aquellas que tuvieran una frecuencia mayor a 0.01 en bases de datos de individuos sanos, como ExAC y gnomAD (Lek *et al.* 2016). Sin embargo, para

Llevar a cabo un análisis más riguroso de la frecuencia máxima admisible de una variante patogénica asociada con el síndrome en bases de datos, se llevó a cabo el cálculo con FrequencyFilter (Whiffin *et al.* 2017). Los parámetros utilizados para el cálculo se muestran en la Tabla 2. En gnomAD podría haber, cuando mucho, una sola copia de una variante asociada con el síndrome de Patterson-Stevenson-Fontaine.

Tabla 2. Valores usados para realizar el cálculo, y resultados obtenidos de la frecuencia alélica máxima permitida, para variantes patogénicas en el síndrome de Patterson-Stevenson-Fontaine.

Características	Valor usado
Herencia	Monoalélica
Prevalencia	1 en 1 000 000
Heterogeneidad alélica	0.30
Heterogeneidad genética	0.30
Penetrancia	0.40
Grado de confianza	0.99
Resultados	
Máxima frecuencia alélica esperada para la variante en la base de datos gnomAD	0.000000112
Número máximo de alelos esperados en la base de datos gnomAD	1

1.3 Primer análisis de las variantes

La anotación de las variantes y el análisis inicial de los datos de esta familia se realizaron con el software VariantStudio v3.0 (Illumina). Se analizaron las variantes subdividiendo a los individuos en dos tríos (Figura 3, p.22). Uno conformado por la madre afectada y los dos abuelos sanos. En este trío se asumió la presencia de una variante *de novo* en la madre. El segundo trío comprendió a la propósa afectada, junto con su madre afectada y su padre sano. En este trío se consideró una herencia dominante.

En el primer trío (madre afectada y abuelos maternos sanos) se partió de 93,235 variantes en la madre, distribuidas en 20,377 genes. Después de aplicar los filtros de calidad y frecuencia (valor de "PASS" en la columna de filtros del archivo VCF, profundidad mínima de 10x, frecuencia mínima de la variante alternativa 0.2, variantes que alteran la secuencia codificadora o el splicing, y frecuencia de la

variante menor a 0.01 en bases de datos), quedaron 683 variantes en 589 genes. Se llevó a cabo el análisis de segregación de estas variantes en el trío, asumiendo la presencia de una variante *de novo* en la madre, y quedaron 186 variantes en 182 genes.

En el segundo trío, conformado por la paciente y sus padres, se inició con 88,786 variantes en 20,152 genes. Al aplicar los filtros de calidad y frecuencia, descritos en el párrafo anterior, quedaron 678 variantes en 563 genes. Se llevó a cabo el análisis de segregación de las variantes en este trío, considerando aquí herencia dominante debido a que la niña y su madre comparten el síndrome. Después del análisis de segregación, quedaron 182 variantes en 174 genes.

Las variantes de interés son aquellas que se presentan en la niña afectada y en su madre, pero no en el resto de los integrantes de la familia, ni en otros individuos sin el diagnóstico de este síndrome. Para hallar estas variantes de interés se analizó la distribución de las 368 variantes (186 del primer trío más 182 del segundo) tanto en la familia estudiada como en la base de datos del laboratorio. Este análisis se llevó a cabo con un programa propio del laboratorio, que prioriza las variantes en función de su distribución en la familia de interés y teniendo en cuenta la frecuencia de las variantes en toda la base de datos. Con este análisis se obtuvieron 2 variantes candidatas en 2 genes, una en *SPAG16* y otra en *HEPH*. Las variantes se describen en la Tabla 3. Ambas muestran una segregación consistente con la esperada en la familia ya que se encuentran en estado heterocigoto en la paciente y su madre, y dichas variantes no se encontraron en miembros no afectados de la familia.

Estas 2 variantes fueron reanotadas con VEP (v96), en las dos versiones del genoma humano, hg19 y hg38 (las versiones más recientes de algunas bases de datos, como GENCODE, HGMD-PUBLIC y ClinVar, solamente están disponibles en VEP para hg38) (McLaren *et al.* 2016).

La variante en *SPAG16* está en un sitio donador de splicing y en VEP el impacto de la variante es considerado elevado. En gnomAD esta variante está representada 6 veces (en un total de 251,162 alelos registrados en esa posición). Los 6 alelos se presentan en genotipos heterocigotos, incluyendo dos individuos de ascendencia latinoamericana (un hombre y una mujer). Esta frecuencia es superior a la esperada, de acuerdo con el cálculo descrito en la sección anterior (sección 1.2 Cálculo de frecuencia alélica; p. 26).

Por otra parte, la variante en *HEPH* (localizado en el cromosoma X) tampoco se encontró en otros individuos en la base de datos del laboratorio. Sin embargo, en gnomAD la variante aparece en 746 alelos de un total de 201,272 alelos en esa posición. En esta base de datos, 5 individuos femeninos son homocigotos para la variante. El resto de los individuos femeninos son heterocigotos para la misma. Todos los individuos masculinos con esta variante serían hemocigotos. En esa base de datos de individuos sanos, la frecuencia de la variante es muy elevada de acuerdo con los cálculos realizados en la sección anterior para la frecuencia máxima de una variante patogénica para este síndrome. Más aún, SIFT (Kumar *et*

al. 2009) y PolyPhen (Adzhubei *et al.* 2010) predicen que el efecto de la variante en la proteína es tolerable (valor 0.76) y benigno (valor 0), respectivamente. VEP considera que el impacto de la variante es moderado.

Tabla 3. Variantes de la familia con síndrome de Patterson-Stevenson-Fontaine priorizadas con la base de datos del laboratorio.

Gen	SPAG16	HEPH
HGNC ID	23225	4866
Coordenada (hg19)	2:214727354	X:65486284
Coordenada (hg38)	2:213862630	X:66266442
Cadena	Positiva	Positiva
Variante	ENST00000331683.10:c.1214+2T>A	ENST00000519389.5:c.3409G>T ENSP00000430620.1:p.Val1137Leu
Localización	Intrón 11/15	Exón 21/21
dbSNP	rs756174280	rs35700738
Consecuencia	Alteración en sitio donador de splicing	Variante de sentido equivocado
SIFT	-	Tolerada
PolyPhen	-	Benigna
Genotipo	Heterocigoto	Heterocigoto
(Núm. de alelos con la variante/núm tot de alelos) en gnomAD	6/251,162	746/201,272
Frecuencia alélica gnomAD	0.00002389	0.003706

Las 366 variantes restantes fueron revisadas manualmente para verificar que efectivamente cumplieran con las restricciones de segregación esperadas (variante *de novo* en el trío madre-abuelos maternos, y herencia dominante en el trío propósito-padres). Asimismo, se determinó que las variantes eran más frecuentes de los esperado en la base de datos del laboratorio, y estaban presentes en individuos sin este síndrome, o que las variantes tenían una frecuencia superior a la esperada en bases de datos públicas como gnomAD.

1.4 Análisis de variantes con Exomiser

Para descartar la posibilidad de que hubiera falsos negativos se reanalizaron las 368 variantes con Exomiser. Este programa prioriza las variantes en función del efecto predicho para la variante, su frecuencia en bases de datos públicas, y de las anotaciones asociadas al gen en el cual se encuentra la variante. Se llevó a cabo un análisis familiar con un archivo VCF conteniendo las variantes de los cinco individuos de la familia, más un archivo en formato ped conteniendo los fenotipos de los cinco individuos junto con su relación de parentesco. Las primeras 7 variantes priorizadas correspondieron a los genes *PLEC*, *PTK2*, *NARFL*, *ADPRH*, *YRDC*, *HYDIN* y *DPP9*. La Tabla 4 muestra los detalles de los datos obtenidos. A pesar de que se suministró la información fenotípica de los 5

individuos de esta familia, las variantes de estos 7 genes incumplen con el patrón de segregación esperado debido a que el programa no aplica un análisis estricto de la segregación. Más aún, estas variantes se hallaron en otros individuos en la base de datos del laboratorio, razón por la cual no se consideraron variantes candidatas para el síndrome en estudio. En la familia se observa solamente un heterocigoto para cada variante; sin embargo, esperaríamos que la variante estuviera presente en los 2 miembros afectados de la familia (la paciente y su madre).

Exomiser asigna valores de prioridad entre 0 y 1 a las variantes que analiza. El valor máximo observado fue de 0.80 (Tabla 4). Estos valores se obtienen de un promedio ponderado del valor (score) asignado a la variante y del valor asignado a las características asociadas con el gen y con el paciente.

Tabla 4. Información de las variantes priorizadas por Exomiser.

Gen	Exomiser score	Phenotype score	Variant score	Efecto	Función del gen	Segregación en la familia
PLEC	0.80	0.57	0.95	Codón de paro	Arquitectura del citoesqueleto y forma celular	1 heterocigoto
PTK2	0.76	0.50	1.00	Sentido equivocado	Tirosin cinasa presenten en sitios focales de adhesión celular	1 heterocigoto
NARFL	0.65	0.50	0.95	Codón de paro	Proteína nuclear asociada con la prelamina A	1 heterocigoto
ADPRH	0.54	0.50	0.90	Variante en sitio de splicing	Ribosil-arginin hidrolasa encima catalizadora en el ciclo de ribocilación de ADP	1 heterocigoto
YRDC	0.11	0.50	0.66	Sentido equivocado	Receptor de dopamina	1 heterocigoto
HYDYN	0.28	0.25	0.77	Sentido equivocado	Movilidad ciliar	2 heterocigotos, 1 no afectado
DPP9	0.07	0	0.90	Variante en sitio de splicing	Peptidil peptidasa encargada de degradar incretinas	1 heterocigoto

1.5 Análisis de variantes por trío

Finalmente, para descartar la posibilidad de que hubiera algún falso negativo causado por la metodología seguida en los análisis anteriores, se llevó a cabo un tercer análisis en la línea de comando utilizando los archivos VCF y el programa bcftools. Como parte del procesamiento automatizado en el laboratorio, los archivos VCF son filtrados para eliminar variantes que no pasan los filtros de calidad descritos anteriormente (profundidad mínima de 10x, frecuencia mínima de

0.2 del alelo alternativo en posiciones heterocigotas, y valor distinto de PASS en la columna de filtro). Como parte del procesamiento automatizado de los archivos, también se eliminan variantes con una frecuencia muy elevada en la base de datos local. Al final de este procedimiento se obtienen archivos depurados que pueden ser analizados más fácilmente por tener un menor número de variantes y por estar libres de un gran número de datos que no cumplen con los criterios de calidad. Con estos archivos se utilizó bcftools para analizar las variantes en los dos tríos en los que se dividió la familia, buscando variantes *de novo* en la madre, y variantes con herencia dominante en el trío de la propósita (Figura 3, p. 22). Se obtuvieron 2 variantes candidatas en los genes *WDR7* y *ADRBK2* en el trío conformado por los abuelos maternos y la madre de la paciente. Se encontraron también 2 variantes en el trío conformado por la paciente y sus padres. Sin embargo, en este segundo trío las variantes se encontraron en los genes *PLXNC1* y *PREX1*. Información de las 4 variantes se muestra en la Tabla 5. Ninguna de estas variantes es compartida por la propósita y su madre.

Tabla 5. Variantes encontradas en el análisis de segregación por tríos utilizando la línea de comando y archivos VCF con variantes previamente filtradas por calidad.

Gen	<i>WDR7</i>	<i>ADRBK2</i>	<i>PLXNC1</i>	<i>PREX1</i>
HGNC	13490	290	9106	32594
Coordenada hg19	18:54362428	22:26091074	12:94676075	20:47282822
Coordenada hg38	18:56695197	22:25695107	12:94282299	20:48666284
Cadena	Positiva	Positiva	Positiva	Negativa
Variante	ENST00000254442 :c.1356A>G	ENST00000324198 :c.1356A>G	ENST00000258526 :c.3777G>A	ENST00000371941 :c.1737C>T
Localización	Exón 11/28	Exón 13/21	Exón 23/31	Exón 15/40
dbSNP	-	-	rs555676791	rs374180744
Consecuencia	Sitio de splicing	Sitio de splicing	Sitio de splicing	Sitio de splicing
SIFT	-	-	-	-
Polyphen	-	-	-	-
Genotipo	Heterocigoto	Heterocigoto	Heterocigoto	Heterocigoto
Filtro de segregación aplicado	De novo	De novo	Dominante	Dominante
Num. De alelos con la variante/total de alelos en gnomAD	-	-	8/281230	29/208106
Frecuencia alélica en gnomAD	0	0	0.00002845	0.0001394

1.6 Análisis de variantes sinónimas

Para los análisis anteriores se consideraron únicamente variantes que pudieran afectar la secuencia de aminoácidos de la proteína. Sin embargo, al no encontrarse variantes candidatas, se decidió realizar un análisis con todas las variantes que pasaran el filtro de calidad, incluyendo variantes sinónimas, en virtud de que este tipo de variantes pueden llegar a asociarse con enfermedades monogénicas (Sauna & Kimchi-Sarfaty 2011; Hunt *et al.* 2014). A las variantes de

cada trío se les aplicaron filtros de segregación *de novo* y dominante, como se ha descrito anteriormente, mediante bcftools. Las variantes obtenidas del análisis de cada trío se juntaron en una sola lista de 172 variantes. Estas variantes fueron filtradas y priorizadas con la base de datos del laboratorio quedando 4 variantes (Tabla 6). Ninguna de estas puede ser considerada candidata para causar el síndrome debido a su alta frecuencia en bases de datos públicas.

Tabla 6. Variantes del análisis, incluyendo variantes sinónimas, priorizadas con la base de datos del laboratorio.

Gen	ZNF141	ZNF107	GPR20	LRIT1
HGNC	12926	12887	4475	23404
Coordenada hg19	4:367645	7:64167949	8:142366986	10:85992632
Coordenada hg38	4:373856	7:64707571	8:141356886	10:84232876
Cadena	Positiva	Positiva	Negativa	Negativa
Variante	ENST00000240499	ENST00000620827	ENST00000377741	ENST00000372105 ENSP00000361177
Localización	Exón 4/4	Exón 4/4	Exón 2/2	Exón 4/4
dbSNP	rs137931151	rs375727539	rs143127247	rs75589469
Consecuencia	Variante sinónima	Variante sinónima	Variante sinónima	Sentido equivocado
SIFT	-	-	-	-
Polyphen	-	-	-	-
Genotipo	Heterocigoto	Heterocigoto	Heterocigoto	Heterocigoto
Num. De alelos con la variante/total de alelos en gnomAD	715/272120	4/248872	1627/279636	359/275060
Frecuencia alélica en gnomAD	0.002628	0.00001607	0.005818	0.001305

1.7 Análisis de variantes estructurales

Al no encontrarse variantes puntuales, o inserciones y deleciones pequeñas que pudieran estar asociadas con el síndrome en estudio, se procedió al análisis de las variantes estructurales en la familia. Como en los análisis anteriores, las variantes fueron analizadas en dos tríos (abuelos maternos y madre afectada, con una variante *de novo*; padres y propósito, con herencia dominante de madre a hija). El análisis se llevó a cabo con bcftools y se encontraron dos deleciones en el gen *PRKRA*. La primera de 3,890 pb y la segunda de 1,037 pb (Tabla 7).

Tabla 7. Variantes estructurales halladas en el gen *PRKRA* en la familia con síndrome de Patterson-Stevenson-Fontaine.

Coordenada hg19	Tipo de variante	Longitud de la variante	Genotipo
2:179 296 981-179 300 871	Delecion	3,890 pb	Heterocigoto
2:179 308 110-179 309 147	Delecion	1,037 pb	Heterocigoto

Las variantes se examinaron visualmente con el software Integrative Genomics Viewer (IGV). Las 2 variantes encontradas están presentes en 3 de los 6 integrantes de la familia, lo cual es inconsistente con la segregación del fenotipo en estudio. Al comparar la región donde fueron encontradas las delecciones en la familia con otras muestras de la base de datos se observó que la mayoría de las muestras presentan patrones de delecciones iguales o muy similares en esa región.

Familia 2. Síndrome de Roifman-Chitayat

1.1 Datos generales de la secuenciación

Para la familia 2, con la paciente diagnosticada clínicamente con síndrome de Roifman-Chitayat, el resumen de los datos de la secuenciación se muestra en la Tabla 8. Más del 70% de las lecturas de las muestras de la familia cumplen con los parámetros de calidad Q30, más del 97% de las regiones blanco fueron secuenciadas y casi el 90% de las lecturas fueron alineadas con las regiones blanco. En la familia se obtuvo una cobertura promedio de 100x (Figura 9).

Tabla 8. Resumen de datos obtenidos de la secuenciación de los exomas de la familia 2.

	Padre	Madre	Paciente
Total de lecturas	106,244,148	102,377,258	120,417,076
Porcentaje de bases \geqQ30	76.82%	78.18%	76.29%
Longitud promedio de las lecturas	125 pb	103 pb	125 pb
Porcentaje de lecturas alineadas	89.75%	88.77%	89.52%
Regiones blanco cubiertas	97.57%	95.25%	97.69%
Profundidad promedio	109x	99x	122x

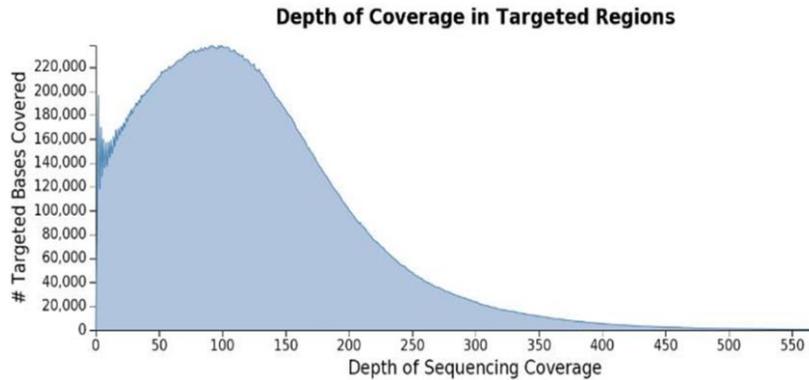


Figura 9. Distribución de la profundidad en la propósita. La profundidad promedio de 122x.

1.2 Cálculo de frecuencia alélica máxima esperada en una población de individuos sanos

El síndrome Roifman-Chitayat es muy raro, habiéndose reportado en una sola familia (Roifman & Chitayat 2009). Por lo tanto, cualquier variante asociada con este síndrome debe tener una frecuencia mucho muy baja en las bases de datos de alelos de individuos sanos. El cálculo de la frecuencia alélica máxima admisible en gnomAD, de acuerdo con FrequencyFilter (Whiffin *et al.* 2017), se muestra en la Tabla 9.

Tabla 9. Valores usados para realizar el cálculo y resultados de la frecuencia alélica máxima permitida para variantes patogénicas en el síndrome de Roifman-Chitayat.

Características	Valor
Herencia	bialélica
Prevalencia	1 en 1 000 000
Heterogeneidad alélica	0.10
Heterogeneidad genética	0.10
Penetrancia	0.80
Grado de confianza	0.95
Resultados	
Máxima frecuencia alélica esperada para la variante en la base de datos genomeAD	0.00000354
Número máximo de alelos esperados en la base de datos gnomAD	2

1.3 Análisis de las variantes

El análisis de los datos de esta familia se inició con el software VariantStudio (v.3). El archivo VCF de la propósita contenía 90,306 variantes en 20,266 genes. Al aplicar el filtro de calidad (filter=PASS, profundidad mayor a 9x y frecuencia mínima de la variante alternativa mayor a 0.19), quedaron 25,685 variantes en 10,859 genes. Al eliminar aquellas variantes que no afectarían la secuencia de las proteínas quedaron 10,554 variantes en 6,026 genes. A continuación, se eliminaron aquellas variantes que estuvieran presentes en bases de datos (ExAc, Proyecto de los 1000 Genomas) con una frecuencia de 0.01 o mayor, quedando 685 variantes en 577 genes. Se analizó la segregación de estas variantes en la familia, considerando una herencia autosómica recesiva, y se obtuvieron 114 variantes en 35 genes.

Estas 114 variantes fueron filtradas y priorizadas con la base de datos del laboratorio y quedaron 10 variantes en 6 genes candidatos: *NBAS*, *XIRP1*, *STAB1*, *PKD1L1*, *NKX3-1* y *CUBN*. Los datos de estas variantes se presentan en la Tabla 10. Se descartaron 5 de estos genes como candidatos *XIRP1*, *STAB1*, *PKD1L1*, *NKX3-1*, y *CUBN*. La variante en el gen *NBAS* se convirtió en el candidato más fuerte, ya que se encontró en los 3 miembros de la familia: en estado heterocigoto en los 2 padres y de forma homocigota en la paciente. En los demás genes las variantes fueron heterocigotas compuestas, con al menos una de las variantes en

cada gen teniendo una frecuencia elevada en bases de datos de individuos sanos. Adicionalmente, ninguno de los 5 genes eliminados ha sido asociado con enfermedades que incluyan característica observadas en la paciente.

Tabla 10. Variantes de la familia con síndrome de Roifman-Chitayat priorizadas con la base de datos del laboratorio.

Gen	Variante	Tipo de variante	dbSNP	Variante heredada por	Frec. Alélica en gnomAD	Genotipo de la paciente
NBAS	ENST00000281513 .9:c.1342-1G>C	Sitio de splicing		Ambos padres	0	Homocigoto
XIRP1	ENST00000340369 c.3433G>A, ENSP00000343140:p.Ala1145Thr	Sentido equivocado	rs147374766	Mamá	0.0007876	Heterocigoto
STAB1	ENST00000321725 c.5413-3 C>T	Sitio de splicing	rs199636230	Papá	0.000451	Heterocigoto compuesto
	ENST00000321725 c.99T>G, ENSP00000312946 :p.Asp33Glu	Sentido equivocado		Mamá	0	
PKD1L1	c.1208T>C, :p.Leu403Pro	Sentido equivocado	rs146750610	Papá	0.00753	Heterocigoto compuesto
	c.3085G>C, :p.Gly1029Arg			Mamá	0	
NKX3-1	c.144C>G, :p.Ser48Arg	Sentido equivocado	rs767428389	Papá	0.000024	Heterocigoto compuesto
	c.136A>G, :p.Thr46Ala			Mamá	0	
CUBN	ENST00000377833 c.3868A>G, ENSP00000367064:p.Ile1290Val	Sentido equivocado	rs751350167	Papá	0.000003	Heterocigoto compuesto
	ENST00000377833 c.722C>T: ENSP00000367064 p.Pro241Leu.		-	Mamá	0	

Al indagar en las funciones de *NBAS*, y su relación con fenotipos y patologías, se encontró que ha sido asociado con varios fenotipos entre ellos el síndrome SOPH (OMIM 614800) llamado así por las siglas en inglés de 3 particularidades que distinguen a la enfermedad: **s**hort stature, **o**ptic atrophy, **P**elger-Huet anomaly syndrome. Estas 3 características (estatura baja, atrofia del nervio óptico y anomalía de Pelger-Huet) son distintivas del síndrome. La paciente estudiada presenta dos de estas características: estatura baja y atrofia del nervio óptico. El síndrome SOPH tiene herencia autosómica recesiva, lo cual es también consistente con lo observado en la paciente.

En virtud del hallazgo de la variante en *NBAS*, el caso fue revisado para confirmar que clínicamente la paciente tuviera rasgos del síndrome SOPH, y para determinar las similitudes o diferencias entre el diagnóstico de probable síndrome de Roifman-Chitayat y el síndrome SOPH. Las características clínicas de la paciente se presentan en la Tabla 11. Como se observa, la paciente tiene las características del síndrome SOPH acompañado de falla hepática aguda. De hecho, hay una segunda enfermedad asociada con alteraciones en *NBAS* y ésta es el síndrome de falla hepática infantil 2 (OMIM 616483).

Tabla 11. Cuadro comparativo de los datos clínicos de la paciente estudiada, del síndrome de Roifman-Chitayat y del síndrome SOPH.

Característica	Síndrome SOPH	Síndrome Roifman-Chitayat	Paciente
Patrón de herencia	AR	AR	AR
Talla baja	Retraso en el crecimiento	Retraso global del desarrollo	Retraso global del desarrollo
Atrofia bilateral del nervio óptico	Atrofia del nervio óptico con pérdida de la agudeza visual y visión a color.	Atrofia /hipoplasia del nervio óptico	+
Hipotonía	+	-	+
Anomalía de Pelger-Huet	+	-	No evaluado
Cuello corto	+	+	-
Dismorfias faciales			
Labio inferior delgado	+	+	+
Epicanto	+	+	+
Características inmunológicas			
Inmunodeficiencia	Inmunodeficiencia primaria	Inmunodeficiencia combinada	Inmunodeficiencia posiblemente primaria (no confirmada)
Infecciones frecuentes	+	+	+
Características óseas			
Edad cronológica ósea retrasada	+	-	+
Osteopenia	+	-	+
Displasias esqueléticas	+	+	+
Anomalías en la columna	+	+	+
Características hepáticas			
Falla hepática aguda	+	-	+
Hepatomegalia	+	-	+
Enzimas hepáticas elevadas (ALT y AST)	+	-	+

AR autosómico recesivo, ALT alanino amino transferasa, AST aspartato amino transferasa, + presente en la paciente, - no se encontró en la paciente.

La variante encontrada en el gen *NBAS* es una transversión G>C en el sitio aceptor de splicing, en la base inmediatamente anterior al inicio del exón 15. Esta variante no ha sido reportada anteriormente. Sin embargo, como se observa en la Figura 6 ya se han reportado alrededor de 15 variantes patogénicas en el gen relacionadas con síndrome SOPH y/o falla hepática aguda. En la Figura 10 se muestra la nueva variante encontrada en la paciente estudiada en el presente trabajo.

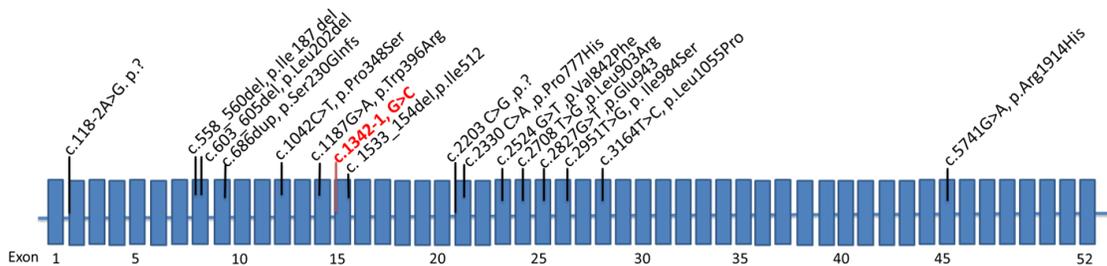


Figura 10. **Variantes que han sido encontradas en *NBAS* en individuos con síndrome SOPH o con falla hepática aguda.** En rojo se observa la nueva variante encontrada en el gen. Imagen modificada de Segarra *et al.* (2015).

En la Figura 11 se muestra la variante con el programa IGV, en la paciente y sus padres. En la parte superior de la imagen se muestra parte del cromosoma 2, y la barra roja en el brazo corto indica la región que aparece en detalle en la parte inferior de la imagen. La parte inferior de cada imagen muestra la secuencia de referencia del genoma humano y debajo de ella la barra horizontal azul representa parte del exón 15 de *NBAS* (el gen se encuentra en la cadena negativa). La variante se observa en la base inmediatamente anterior al exón 15. Las dos figuras de la parte superior corresponden a las lecturas alineadas del padre y de la madre, que son heterocigotos para la variante. La figura inferior corresponde a la paciente, que es homocigota para la alteración.

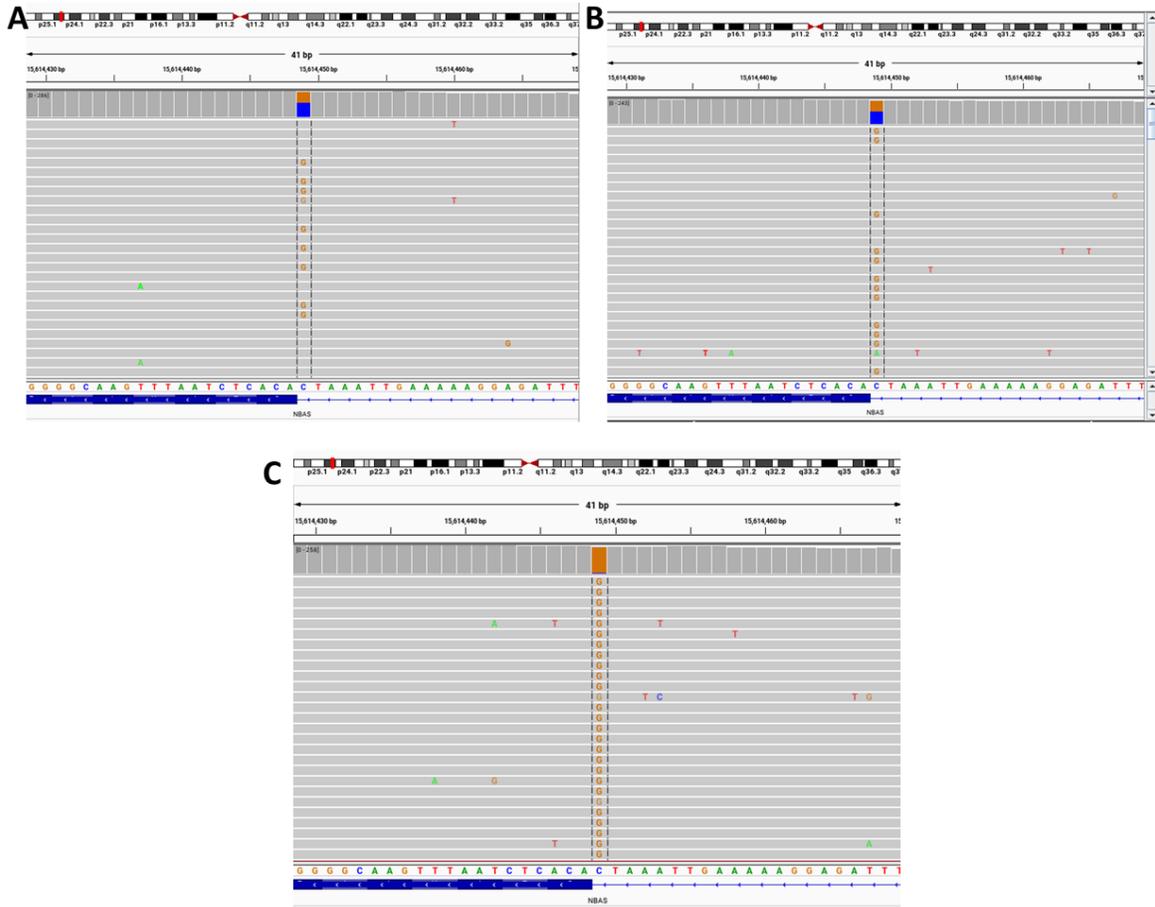


Figura 11. **Segregación de la variante en la familia 2** vista con el programa Integrative Genomics Viewer (IGV). A y B) La variante encontrada de forma heterocigota en ambos progenitores (padre y madre respectivamente). La mitad de las lecturas marcadas en azul corresponden a la base del genoma de referencia, y la otra mitad en naranja presenta el cambio. C) En esta imagen se observa que la propórita es homocigota para la variante ENST00000281513.5:c.1342-1G>C. Todas las lecturas de la paciente tienen esta variante.

1.2 Análisis de la variante en Exomiser

Para determinar que no hubiera otras variantes de interés, que hubieran sido pasadas por alto con el procedimiento anterior, se filtraron y priorizaron las 114 variantes con Exomiser. La variante con el score más elevado fue *NBAS*, tal como en el análisis previo (Tabla 12). El score del fenotipo fue 0.92 y el de la variante 0.90 (este último valor se asigna por defecto a variantes en sitios de splicing). El score global, que es un promedio ponderado de los anteriores, fue de 0.90.

Tabla 12. Valores de Exomiser para la variante en *NBAS*.

Gen	Exomiser score	Phen score	Variant score	Efecto de la variante	Segregación en la familia	Frecuencia en bases de datos
<i>NBAS</i>	0.90	0.92	0.90	Variante en el sitio aceptor de splicing	Progenitores portadores Paciente homocigota	0

Exomiser score: score global asignado por Exomiser. Phen score: score del fenotipo. Variant score: score de la variante.

1.2 Análisis *in silico* de la variante en el sitio aceptor de splicing

Para comprender mejor el efecto de la variante hallada en *NBAS* en el sitio aceptor de splicing del intrón 14, se realizó un análisis *in silico* con tres programas distintos. El análisis incluyó la variante junto con la secuencia de 231 nucleótidos en cada flanco. Los programas de predicción de sitios de splicing utilizados fueron FSPLICE (Xiong *et al.* 2009), NetGene2 (Brunak *et al.* 1991) y NNSplice (Reese *et al.* 1997). Los tres se basan en modelos Markov y son recomendados por el manual de análisis de variantes del NIH (versión 2013).

FSPLICE indica que no hay otros sitios que puedan ser usados como aceptores de splicing en el segmento analizado, y que puedan sustituir al sitio que se perdió. NetGene2 propone que podría haber 2 sitios aceptores de splicing pero les confiere valores de confianza menores al 43%. Finalmente, NNSplice tampoco encuentra probables sitios aceptores de splicing en el segmento analizado.

Adicionalmente, con el programa Human Splicing Finder se realizó la comparación de la secuencia de referencia de Ensembl y la secuencia con la variante (Desmet *et al.* 2009). Este programa predice que la alteración muy probablemente afecte el splicing (Figura 12A). Este programa sugiere que existen 3 posibles sitios que podrían ser usados como sitios aceptores de splicing cerca del exón 15 (Figura 12B).

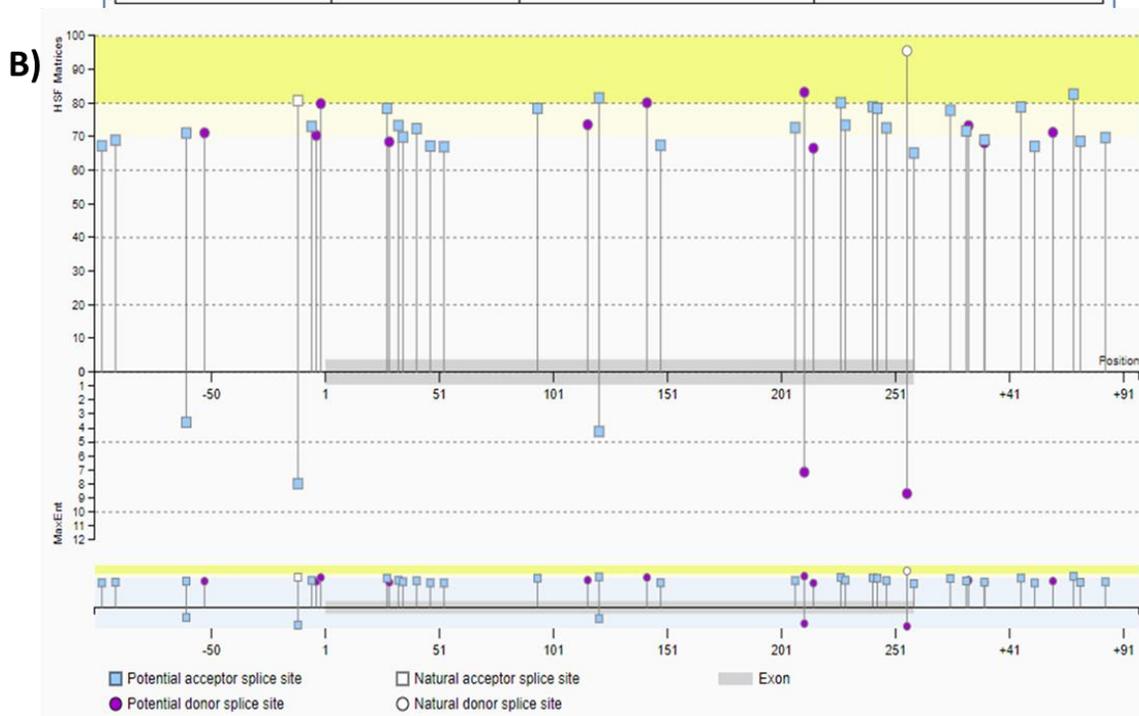
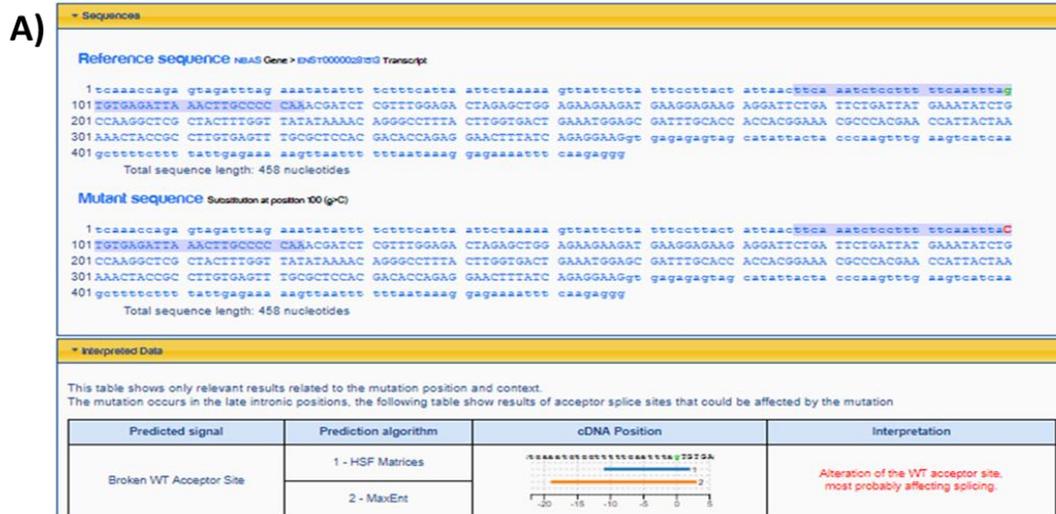


Figura 12. **Resultados obtenidos con el software Human Splicing Finder.** A) Se observa la secuencia de referencia de Ensembl y la secuencia con la variante encontrada, así como la interpretación del posible efecto de la variante (letras en rojo), muy probablemente la variante altera el sitio de splicing. B) Se observan los sitios dónde se encontraron potenciales sitios de aceptores de splicing (cuadrados azules) a lo largo de la secuencia analizada. En el eje vertical de la gráfica se observan los valores consenso (HSF matrices) que el programa otorga a cada sitio potencial, expresados en una escala del 1 al 100. En el eje horizontal observamos una escala numerada de las posiciones en la secuencia analizada. Finalmente, en la parte inferior de la gráfica se observan los valores de MaxEnt que son los valores máximos de entropía para señales de splicing en RNA.

Los resultados obtenidos con estas 4 herramientas sugieren que la variante altera el splicing del transcrito.

Discusión

En este trabajo se estudiaron 2 familias con individuos diagnosticados clínicamente con enfermedades genéticas que se piensa tienen una arquitectura monogénica: el síndrome de Patterson-Stevenson-Fontaine y el síndrome de Roifman-Chitayat. En ambos casos, el estudio fue de tipo familiar con objeto de poder analizar la cosegregación de las variantes con el fenotipo afectado. De este modo, en la primera familia, con diagnóstico de Patterson-Stevenson-Fontaine, fue posible estudiar tres generaciones. En esta, el análisis de las variantes se llevó a cabo subdividiendo la familia en dos tríos. Uno conformado por los abuelos maternos no afectados, más la madre afectada de la propósita. En la madre se asume que habría una variante *de novo* asociada con el síndrome. El segundo trío en esta familia estuvo conformado por la paciente, más la madre afectada y el padre sano. En este trío habría habido herencia de tipo dominante. Es decir, en esta familia la variante asociada con el síndrome habría aparecido *de novo* en la madre, quien la habría transmitido a la hija.

En el trío con la paciente con posible síndrome de Roifman-Chitayat se asumió herencia autosómica recesiva, con base en el trabajo en el que se describió esta enfermedad (Roifman & Chitayat 2009). Para ambas familias de las afectadas con los síndromes, se llevó a cabo secuenciación de exoma, se obtuvieron las variantes, y se analizaron las posibles consecuencias de estas con relación a las patologías estudiadas.

Familia 1: Síndrome de Patterson-Stevenson-Fontaine

En la familia 1, de la paciente diagnosticada con síndrome de Patterson-Stevenson-Fontaine, se llevó a cabo la secuenciación del exoma completo de tres generaciones. La primera generación comprendió a los abuelos maternos, sanos, de la propósita. La segunda generación incluyó a la madre afectada y al padre sano de la propósita, y la tercera generación incluyó solamente a la paciente. El análisis se llevó a cabo dividiendo a la familia en 2 tríos. Uno formado por los abuelos maternos y la madre afectada. Se asume que la alteración molecular apareció *de novo* en esta última. En el segundo trío, formado por la propósita y sus progenitores, habría herencia dominante, con transmisión de la variante de madre a hija. Se asume que el genotipo relevante de estas dos sería heterocigoto.

Las variantes que pasaron los filtros de calidad, profundidad y frecuencia se analizaron desde el punto de vista de su segregación en la familia y de su distribución en la base de datos del laboratorio. Las variantes que quedaron en este análisis (en *HEPH* y *SPAG16*) fueron consultadas en bases de datos de individuos presuntamente sanos, y se determinó que su frecuencia es incompatible con la prevalencia del síndrome y con el tipo de herencia del mismo.

Si se trabaja con una prevalencia de 1 en 1 000 000 para el síndrome de Patterson-Stevenson-Fontaine, la frecuencia alélica máxima esperada para una variante en la base de datos de gnomAD sería de 0.000000112 y un máximo de alelos de uno. Es decir, solamente podría haber un individuo afectado. Es importante notar que este es el número mínimo de individuos que arroja el cálculo,

por lo que el resultado no cambiaría aún si se considerara una prevalencia menor a la señalada. Las frecuencias de las variantes obtenidas en el estudio son muy superiores a este resultado, como se muestra en los resultados. Lo que se traduce en un individuo afectado.

Análisis de variantes en Exomiser

Con objeto de verificar los resultados anteriores se llevó a cabo un segundo análisis de las variantes, filtrándolas y priorizándolas con el programa Exomiser. También se incluyó en este análisis el criterio de cosegregación de la variante y datos del fenotipo de las afectadas. Sin embargo, una limitación de ese programa es que el criterio de cosegregación no es aplicado de manera estricta. El resultado fue que el programa entregó una lista de variantes candidatas que, al ser analizadas manualmente para determinar la cosegregación con el fenotipo, fueron excluidas. Por esta razón, al igual que con el método inicial, no quedaron variantes candidatas que pudieran asociarse con el síndrome en estudio.

Análisis de variantes por trio

Finalmente, para confirmar los resultados anteriores, se hizo un análisis en la línea de comando, con bcftools, filtrando las variantes únicamente por calidad y segregación en las tres generaciones de la familia. Esto con el propósito de lograr identificar variantes *de novo* en la madre de la paciente y poder revisar variantes heredadas de la madre a la propósita. Al igual que en análisis anteriores, se dividieron las tres generaciones en dos tríos. Se obtuvieron 2 variantes candidatas *de novo* en la madre (en *WDR7* y *ADRBK2*), ninguna de ellas presentes en bases de datos públicas o en la base de datos del laboratorio. En cuanto a las variantes transmitidas de la madre afectada a su hija, también quedaron dos (*PLXNC1* y *PREX1*), ambas con genotipos heterocigotos. Sin embargo, ambas variantes tienen una frecuencia mayor a la esperada en bases de datos de individuos presuntamente sanos, incluyendo individuos de ascendencia europea y latinoamericana. Más aún, ninguna de las variantes *de novo* en la madre corresponde a las variantes heredadas de la madre a la paciente, por lo que, al igual que en los análisis anteriores, al finalizar el procedimiento no quedaron variantes candidatas asociadas con el síndrome en estudio.

Análisis de variantes sinónimas

En los análisis descritos arriba se estudiaron variantes que pudieran alterar la secuencia de aminoácidos de la proteína. Esto incluye no solamente variantes de sentido equivocado, sino también variantes sin sentido, alteraciones en los codones de inicio y de término, cambios en los sitios de splicing y en algunas bases intrónicas adyacentes a estos sitios de splicing. También se analizaron inserciones y deleciones que pudieran o no alterar el marco de lectura. Sin embargo, no se obtuvo ninguna variante que pudiera asociarse con el síndrome, una vez considerados factores como la calidad de la base llamada, la segregación de la variante en la familia, y la frecuencia de la variante tanto en bases de datos públicas como en la base de datos del laboratorio. Por esta razón, se decidió considerar otras posibilidades. Por ejemplo, aunque normalmente se excluyen variantes sinónimas, debido a que no alteran la secuencia de aminoácidos de la

proteína, se sabe que pueden estar asociadas con enfermedades genéticas, lo que puede explicarse por la diferencia en la abundancia de los tRNA asociados con distintos tripletes que codifican para el mismo aminoácido (Zuben E. & Kimchi-Sarfaty 2011).

Al analizar las variantes sinónimas, se hallaron cuatro. Sin embargo, todas ellas presentaron frecuencias alélicas que rebasan los valores esperados en bases de datos de individuos sanos. Algunas de esas variantes fueron halladas, incluso, en genotipos homocigotos en individuos de bases de datos públicas.

Variantes estructurales (SV)

En el análisis de SV se encontraron 2 deleciones en el gen *PRKRA* en la madre y en la hija, pero no en los demás integrantes de la familia. Sin embargo, la deleción se encontró en otros individuos en la base de datos del laboratorio, lo que llevó a eliminarla como candidata a causar la enfermedad. Se sabe que el gen media el efecto del interferón contra infecciones virales e interactúa con RNA viral de doble cadena. De acuerdo con la literatura este gen ha sido señalado como el causante de distonia 16 (Dos Santos *et al.* 2018).

Otras posibilidades

Para identificar el gen asociado con el síndrome de Patterson-Stevenson-Fontaine se llevó a cabo la secuenciación del exoma en una familia en la cual dos generaciones presentan el síndrome. La variante habría aparecido *de novo* en la madre de la paciente, y después habría sido heredada de la madre a la hija. Sin embargo, análisis detallados de las variantes, incluyendo variantes sinónimas y variantes estructurales no condujeron a ninguna variante candidata. Hay varias explicaciones posibles.

Por un lado, debe considerarse que en estudios de secuenciación de exoma, no son secuenciadas todas las regiones codificadoras de los, aproximadamente, 20 mil genes codificantes de proteínas. Para obtener la librería del exoma es necesario separar éste del 98% restante del genoma humano que no va a ser secuenciado. La captura de las secuencias exónicas se lleva a cabo mediante sondas que hibridan los fragmentos exónicos de interés. Sin embargo, algunas de estas regiones exónicas son difíciles de capturar y no llegan a ser secuenciadas. Esto ocurre, en particular, con secuencias muy ricas en GC. Si la alteración molecular estuviera en alguna de estas regiones, no sería secuenciada.

Adicionalmente, la secuenciación del exoma abarca los exones codificadores por lo que alteraciones en regiones reguladoras del gen (incluyendo regiones promotoras, y enhancers, dentro y fuera del gen) no son estudiadas. Se sabe que en algunas enfermedades las alteraciones pueden estar en estas regiones reguladoras, que pueden estar en regiones intergénicas, a varios miles de bases de distancia del gen, o en secuencias intrónicas (Carss *et al.* 2017; Short *et al.* 2018). Las regiones no traducidas (UTR-3' y UTR-5') tampoco son secuenciadas.

Por otro lado, las secuencias que se desvían mucho (en cualquier dirección) de una proporción del 50% de contenido GC, tienden a tener valores de calidad bajos, así que pueden ser eliminadas al aplicar filtros de calidad.

En la secuenciación del exoma, únicamente pueden verse alteraciones estructurales (inserciones y deleciones grandes, transposiciones y translocaciones) que alteren secuencias exónicas. Si los puntos de ruptura de esas alteraciones se encuentran en intrones o regiones intergénicas, no pueden ser observados.

Un tipo de alteración que no se analizó en este trabajo es el cambio en el número de copias (CNV). Estas son variantes que conducen a la pérdida de una copia de la secuencia, o que llevan al incremento de la misma (Retterer *et al.* 2015). Debido a que la secuenciación del exoma involucra la captura de las secuencias codificadoras, la eficiencia de la captura, y las variaciones experimentales entre una muestra y otra pueden llevar a observar variaciones en la profundidad de las secuencias. Para determinar la presencia de CNV es necesario distinguir entre este “ruido” experimental, y la señal verdadera causada por la alteración en el número de copias (Fromer *et al.* 2012; Poultney *et al.* 2013). Se han diseñado varias herramientas bioinformáticas para llevar a cabo este análisis, aunque los resultados tienden a diferir entre ellas (Samarakoon *et al.* 2014; Tan *et al.* 2014). El estudio de este tipo de variantes es un análisis que valdría la pena efectuar (Ruderfer *et al.* 2016).

Las limitaciones anteriormente mencionadas son de tipo experimental y algunas de ellas podrían ser resueltas si se hiciera la secuenciación del genoma completo. No obstante, aun secuenciando todo el genoma, seguiría habiendo algunas limitaciones experimentales. De hecho, el genoma humano de referencia tiene, todavía, algunas pequeñas lagunas, especialmente en regiones centroméricas y teloméricas, que tienen secuencias repetidas de baja complejidad. El empleo de tecnologías de secuenciación de tercera o cuarta generación podría permitir superar, al menos parcialmente, algunas de las limitaciones señaladas (Chaisson *et al.* 2015). No obstante, este tipo de tecnologías tienden a tener una tasa de error superior a la secuenciación de segunda generación (Slatko *et al.* 2018).

Además de las limitaciones experimentales señaladas, también debe tenerse en mente que existen limitaciones bioinformáticas. En este sentido, la presencia de múltiples secuencias repetidas y de familias multigénicas en nuestro genoma dificultan el mapeo correcto de las lecturas pequeñas obtenidas mediante secuenciación de segunda generación. El mapeo correcto de las lecturas de secuencias repetidas es posible únicamente cuando en la lectura existe al menos una región de secuencia única en el genoma, que permite asignar la secuencia al sitio correcto. Cuando se tienen secuencias repetidas en tándem, y el número de repeticiones de la unidad rebasa la longitud de la lectura, no es posible determinar si hay expansiones o contracciones en el número de repeticiones. Sin embargo, se sabe que en algunas enfermedades, incluyendo la enfermedad de Huntington y algunas ataxias, el incremento en el número de unidades está asociado con la presencia de la enfermedad. Este tipo de limitaciones no pueden resolverse

mediante secuenciación de genoma completo, o utilizando lecturas ligadas artificialmente, como las desarrolladas por la compañía 10X Genomics. En este procedimiento se corta el DNA genómico en fragmentos largos, y cada uno de estos es cortado en fragmentos más pequeños a los cuales se les agrega una misma secuencia, como código de barras, que indican que provienen de un solo fragmento largo. Esto reduce el número de mapeos incorrectos de las lecturas cortas, debido a que puede identificarse la región de la cual provienen. No obstante, no resuelven la dificultad creada por la variación en el número de repeticiones en tándem.

La imposibilidad de mapear correctamente algunas secuencias repetidas, o que forman parte de familias multigénicas, podría ser resuelta por medio de la secuenciación de tercera o de cuarta generación (Chaisson *et al.* 2015). En ambos enfoques se lleva a cabo la secuenciación de moléculas individuales y es posible obtener lecturas de varios miles de pares de bases. Esto puede ser suficiente para abarcar toda la secuencia repetida, incluyendo repeticiones en tándem, más las secuencias únicas que las flanquean y que permitirían el mapeo correcto.

Las limitaciones experimentales y bioinformáticas de la técnica no son la única posibilidad. Se sabe que las marcas epigenéticas son importantes durante la diferenciación y desarrollo embrionarios debido a la implicación de estas modificaciones en la expresión de genes. Cada tipo celular adopta su propio epigenoma para hacer posible la expresión de genes propios de la línea celular a la que pertenezcan. Cambios de esta naturaleza en nuestra especie están relacionados con malformaciones cráneo faciales. Recientemente en un esfuerzo por identificar perfiles epigenéticos relacionados con el desarrollo embrionario cráneo facial en humanos y su relación con malformaciones, Wilderman (2018) y sus colaboradores realizaron un perfil de modificación de histonas, la mayoría de las modificaciones que encontraron se localizan en promotores, y en regiones intrónicas e intergénicas asociadas con represión de la expresión. Cambios de este tipo podrían ser la causa de las malformaciones de la paciente y de su madre, y para estudiarlo sería necesario utilizar una metodología diferente.

Familia 2: Síndrome Roifman-Chitayat

En la familia de la paciente con probable síndrome de Roifman-Chitayat se secuenció el exoma del trío formado por la propósa y sus padres sanos. Se cubrió casi el 98% del exoma y se obtuvo una profundidad promedio de 100x. Al término del análisis y de la priorización de variantes se obtuvieron variantes en 6 genes. Cinco de estas variantes, en *XIRP1*, *STAB1*, *PKD1L1*, *NKX3-1* y *CUBN*, fueron eliminadas en virtud de su alta frecuencia en bases de datos de variantes de personas sanas y, también, porque la función de los mismos no permite relacionarlos con las características fenotípicas de la paciente. Por ejemplo, la función de *XIRP1* es proteger los filamentos de actina durante los procesos de depolimerización (Otten *et al.* 2012). La variante en sentido equivocado que se encontró en la paciente, y en su madre de forma heterocigota se ha observado con frecuencia alta en las bases de datos, principalmente en población de ascendencia latinoamericana y europea. Este gen no ha sido asociado con

ninguna enfermedad. Más aún, el genotipo en esa posición es igual en la propósa y en su madre no afectada, por lo que no concuerda con la segregación del fenotipo.

Otro gen que fue eliminado fue *STAB1*, que participa en los procesos de angiogénesis y adhesión celular. Se expresa en el endotelio de tejidos como el bazo, también se encuentra en macrófagos y diferentes tipos de células endoteliales; su expresión es inducida durante procesos inflamatorios. Se ha visto que está involucrada en 2 vías de tráfico vesicular endocitosis y reciclaje en el aparato de Golgi (Kzhyshkowska *et al.* 2006). En este gen se encontraron 2 variantes en el trío; la primera es de sentido equivocado (c.99T>G :p.Asp33Glu) heredada por la madre, y la segunda es en un sitio de splicing (c.5413-3C>T) heredada por el padre. La paciente es heterocigota compuesta para estas variantes. El cambio encontrado en el padre tiene alta frecuencia en gnomAD y no se ha señalado como causante de ninguna enfermedad. La variante heredada por la madre no ha sido reportada en ninguna base de datos. Si esta última variante fuera relevante, se esperaría que la madre estuviera afectada.

PKD1L1 participa en el transporte de calcio y se sabe que ciertas mutaciones en el gen son causa de algunas malformaciones cardiacas (Landrum *et al.* 2014). Ninguna de las 2 variantes encontradas aquí ha sido reportada en pacientes. Una variante es de sentido equivocado (c.3085G>C, p.Gly1029Arg), está presente en la madre y no ha sido reportada en bases de datos. La segunda variante (c.1208T>C, p.Leu403Pro) también es de sentido equivocado y fue aportada por el padre, es muy común en todas las poblaciones en bases de datos.

NKX3-1 es un factor de transcripción que funciona como regulador negativo del crecimiento de células epiteliales en próstata (Xie & Wang 2017). En este gen se encontraron 2 variantes de sentido equivocado. Una heredada del padre (c.144C>G, p.Ser48Arg) y otra heredada de la madre (c.136A>G, p.Thr46Ala). Las variantes encontradas no han sido reportadas como causantes de ninguna patología. De la variante encontrada en la madre no existen datos de frecuencia y de la variante en el padre se observa una frecuencia alélica alta en gnomAD.

CUBN ha sido relacionado con diversos fenotipos y se han reportado polimorfismos de predisposición a enfermedades cardiovasculares y albuminuria (Aminoff *et al.* 1999). La proteína transporta vitamina B12 y se han encontrado mutaciones causantes de anemia megaloblástica (Collin *et al.* 2011). En este gen fueron encontradas 2 variantes de sentido equivocado; una heredada del padre (c.3868A>G, p.Ile1290Val), y otra de la madre (c.722C>T, p.Pro241Leu). El cambio heredado por el padre es muy raro; en la base de datos de gnomAD solo se observa un individuo heterocigoto con la variante y para la variante heredada por la madre no se tienen datos de frecuencia ni de reportes de la variante. Estas 2 mutaciones no parecen tener relación con la enfermedad de la paciente debido a la función del gen.

Finalmente, en el gen *NBAS* la variante encontrada fue una transversión en el sitio aceptor de splicing inmediatamente antes del exón 15 (c.1342-1G>C). Como se

muestra en la Figura 411 la variante está en ambos padres de forma heterocigota y la propósa es homocigota. Esto concuerda con una herencia autosómica recesiva y es consistente con la consanguinidad de los padres. La variante no se encuentra en ninguna base de datos.

Al introducir las variantes del trío en una herramienta de priorización de variantes (Exomiser), este último gen apareció en primer lugar en la lista, con valores elevados tanto para la variante, como para las características fenotípicas asociadas con el gen, reforzando la conclusión de que el genotipo en la propósa es patogénico y está asociado con las características observadas en la paciente.

El gen *NBAS* ha sido asociado con 2 enfermedades monogénicas: el síndrome SOPH (MIM 614800) y el síndrome de falla hepática infantil 2 (MIM 616483). El síndrome SOPH fue descrito por primera vez en el grupo étnico Yakut, una pequeña población que habita en el noreste de Siberia. Maksimova *et al.* (2010) estudiaron 34 individuos afectados de esta población, 22 mujeres y 12 hombres, que compartían diversas características fenotípicas, esqueléticas y oftalmológicas que definen el cuadro clínico de la enfermedad. Entre las características observadas destacan estatura baja, atrofia bilateral del nervio óptico, y anomalía de Pelger-Huet (Maksimova *et al.* 2010). La falla hepática aguda causada por mutaciones en el gen se ha observado acompañada de síndrome SOPH, pero también puede presentarse de manera independiente (Haack *et al.* 2015; Staufner *et al.* 2016).

La variante homocigota en *NBAS* llevó a reanalizar las características de la paciente, con objeto de determinar si su enfermedad pudiera corresponder al síndrome SOPH.

Síndrome SOPH

El síndrome SOPH (OMIM 614800) es una enfermedad de herencia autosómica recesiva, descrita inicialmente por Maksimova *et al.* (2010). La enfermedad se denomina síndrome SOPH por las iniciales de tres características principales de la enfermedad en inglés (short stature, optic nerve atrophy, and Pelger-Huet anomaly). El estudio inicial de esta enfermedad fue realizado en individuos afectados de 31 familias Yakut; una población aislada ubicada en el noreste de Siberia. Esta enfermedad tiene una incidencia de 1 en cada 10 000 en esa población (Maksimova *et al.* 2010).

Además de las tres características señaladas anteriormente, el síndrome SOPH también incluye hipotonía, retraso cronológico de la edad ósea, displasia esquelética -principalmente en segmentos metaepifisarios de huesos de las extremidades-, osteopenia, y anomalías en la columna. Adicionalmente, se observan diversos rasgos faciales característicos que incluyen cuello corto, puente nasal deprimido, frente amplia y ancha, pliegue epicántico, pestañas largas y rizadas, labio inferior delgado y exoftalmos. Otras características que han sido reportadas en algunos pacientes incluyen estrabismo, falla renal, falla hepática aguda, hepatomegalia, e inmunodeficiencias (Maksimova *et al.* 2010; Balasubramanian *et al.* 2017).

En el estudio de Maksimova *et al.* (2010) se reportó que la enfermedad es causada por mutaciones homocigotas en el gen *NBAS*. Se identificaron 6 variantes (Tabla 13) poco comunes en el gen, 5 de estas variantes son polimorfismos en la población estudiada, mientras que la variante encontrada en el exón 45 es la causante del síndrome y se encontró en forma homocigota en 33 de los 34 individuos afectados (Maksimova *et al.* 2010).

Tabla 13. Variantes en el gen *NBAS* observadas en la población Yakut (Maksimova *et al.* 2010).

Exón	Variante	Consecuencia	Tipo de variante
2	c.130 C>G	Variante de sentido equivocado	Polimorfismo
16	c.1611 A>G	Variante sinónima	Polimorfismo
25	c. 2775 T>C	Variante sinónima	Polimorfismo
25	c. 2845 G>C	Variante de sentido equivocado	Polimorfismo
26	c. 3026 G>C	Variante de sentido equivocado	Polimorfismo
45	c. 5741 G>A	Variante de sentido equivocado	Variante patogénica

Después de que la enfermedad fue descrita, otros casos más fueron reportados en pacientes de ascendencia europea y también otras variantes causantes de la enfermedad fueron encontradas a lo largo del gen *NBAS*. En algunos de estos casos se reporta que los pacientes presentan un cuadro claro de síndrome SOPH acompañado de falla hepática aguda con niveles enzimáticos altos de ALT, AST (normal 3-50 U/L; 5/80 U/L), que se presenta especialmente durante episodios infecciosos que son frecuentes y persistentes debido a inmunodeficiencias. Sin embargo, autores como Haack *et al.* (2015) también han descrito casos en los que mutaciones homocigotas en *NBAS* dan lugar a casos de falla hepática aguda sin que los pacientes presenten características del síndrome SOPH. Algunos autores (Segarra *et al.* 2015; Staufner *et al.* 2016; Kortum *et al.* 2017) consideran que estos fenotipos comprenden el espectro de una misma enfermedad.

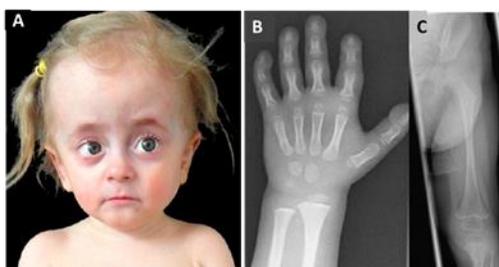


Figura 13. **Fenotipo de paciente con síndrome SOPH y falla hepática aguda.** A) Características y rasgos faciales de una paciente de 3 años con síndrome SOPH y falla hepática aguda. B) Edad ósea retrasada en huesos del carpo, metacarpo y falanges. C) Pierna izquierda con retraso en la maduración del

segmento epifisiario del hueso y osteopenia. Imágenes tomadas de Kortum *et al.* (2017).

La variante homocigota en *NBAS* encontrada en la paciente se localiza en un sitio aceptor de splicing, por lo que se recurrió a herramientas bioinformáticas, basadas en modelos Markov, para tratar de comprender el efecto de la variante en la formación del RNA mensajero maduro.

Se observa en la Figura 312, obtenida con el programa Human Splicing Finder, que existen 11 probables sitios de splicing cercanos al sitio natural que tienen valores de confianza menores a 0.20, lo que sugiere que la variante no puede ser sustituida por alguno de estos sitios y muy probablemente el transcrito maduro se verá afectado. Este tipo de programas de predicción son muy útiles, pero es importante tomar los resultados con cuidado ya que se ha señalado que las predicciones son acertadas solamente en el 50% de las variantes estudiadas (Wappenschmidt *et al.* 2012). Por este motivo, para conocer cómo es que esta variante da lugar al síndrome SOPH es necesario realizar estudios funcionales.

El análisis de las variantes en el trío llevó a concluir que solamente la variante en *NBAS* está asociada con las características de la paciente. Sin embargo, esto implica que el síndrome de Roifman-Chitayat, que corresponde al diagnóstico clínico de la paciente, es también causado por alteraciones bialélicas en *NBAS*, o bien que la paciente en realidad tiene síndrome SOPH, y no Roifman-Chitayat. Para determinar esto, se compararon las características de estos dos síndromes y las características de la paciente estudiada, tal como se muestra en la Tabla 11. La revisión de los datos clínicos, junto con las características moleculares halladas en el estudio, reafirman el cambio en el diagnóstico, así que la paciente tendría, en realidad, síndrome SOPH. Además de las características principales de este síndrome, en la paciente se observaron niveles altos de enzimas hepáticas y falla hepática aguda. Estas características han sido también reportadas en otros pacientes con este síndrome, aunque pueden presentarse por separado conformando el síndrome de falla hepática infantil aguda 2 (Haack *et al.* 2015; Segarra *et al.* 2015; Kortum *et al.* 2017).

La falla hepática aguda podría ser causada específicamente por mutaciones en la secuencia que abarca el dominio SEC39 que se encuentra entre los residuos 725 y 1376 (Staufner *et al.* 2016). Las mutaciones identificadas en la población Yakut se han encontrado casi al final de la secuencia del gen, en los exones 45 a 52, más allá del dominio mencionado, y esta podría ser la razón por la cual no presentan falla hepática aguda (Segarra *et al.* 2015). La mutación de nuestra paciente se encuentra antes del residuo número 533, por lo que el dominio SEC39 estaría afectado.

La proteína *NBAS* forma parte del complejo de la sintaxina 18, un complejo multimérico que se localiza en el retículo endoplásmico. *NBAS* forma un subcomplejo con 3 proteínas periféricas (p31, ZW10 y RINT-1) y tiene un papel importante en el transporte retrogrado del aparato de Golgi al retículo endoplásmico. En nuestra especie, la mayoría de los genes regulados por *NBAS*

están relacionados con el tráfico y modificación de proteínas acopladas al retículo endoplásmico. La deficiencia de *NBAS* no produce cambios en la morfología del aparato de Golgi o el retículo endoplásmico, pero sí causa problemas con el reciclaje y origina defectos en los procesos de glicosilación de proteínas, y da lugar a una respuesta de estrés en el retículo endoplásmico (Aoki *et al.* 2009).

Además de unirse a las proteínas del subcomplejo mencionado, *NBAS* se une a otras proteínas de membrana del retículo endoplásmico. Esto podría indicar que *NBAS* está involucrada en otros procesos o vías de señalización. Se ha visto que *NBAS* participa en los procesos de degradación del mRNA conocidos como NMD (por **n**onsense **m**ediated **d**ecay). Estos procesos consisten en la degradación de mRNAs que presentan codones de paro prematuros que daría lugar a proteínas truncas (Neu-Yilik *et al.* 2004; Conti & Izaurralde 2005; Maquat 2005; Amrani *et al.* 2006; Chang *et al.* 2007). Por otro lado, se ha reportado que una disminución en la expresión de *NBAS* causa una disminución concomitante en los niveles de expresión de aproximadamente 1440 genes (Longman *et al.* 2013).

En nuestra especie la mayoría de los genes relacionados con la vía de *NBAS* tiene ontologías relacionadas con mineralización ósea, y desarrollo y biosíntesis de colesterol. En ausencia de *NBAS*, la proteína MGP se presenta en niveles anormales; se sabe que esta proteína actúa como un inhibidor de la formación de huesos. Alteraciones en *MGP* se han relacionado con el síndrome de Keutel (MIM 245150), en el cual hay calcificación anormal de los cartílagos (Longman *et al.* 2013; Segarra *et al.* 2015). Todavía es necesario conocer más sobre las funciones de *NBAS* en otros procesos, y encontrar la vía mediante la cual este gen se relaciona con el desarrollo óseo.

El síndrome de Roifman-Chitayat, que corresponde al diagnóstico clínico inicial de la paciente estudiada, solamente ha sido reportado una vez (Roifman & Chitayat 2009), y algunas características de los individuos descritos en ese trabajo son similares a las de los pacientes con síndrome SOPH. El síndrome de Roifman-Chitayat fue mapeado, por análisis de ligamiento, a las regiones 1p36.23-p33 o 15q11-q21.1, los lod-scores estuvieron distantes de alcanzar el nivel de significancia de 3.3 convencionalmente aceptado en este tipo de estudios de mapeo. Sin embargo, un análisis posterior de esas regiones en los individuos afectados mostró variantes homocigotas en dos genes, *PIK3CD* y *KNSTRN*, que serían responsables de las anomalías en el desarrollo y de la inmunodeficiencia de las pacientes, respectivamente (Sharfe *et al.* 2018). Esto permite concluir que el síndrome de Roifman-Chitayat y el síndrome SOPH son dos entidades clínicas diferentes.

Conclusiones

El análisis de los exomas de estas pacientes y sus padres se realizaron con el propósito de obtener evidencias genéticas asociadas con las enfermedades que padecen las pacientes.

Durante los diferentes análisis bioinformáticos realizados a la familia 1 con diagnóstico de síndrome de Patterson-Stevenson-Fontaine se observaron diversos cambios genéticos, sin embargo, ninguna de las variantes aquí descritas es la causante del fenotipo de la paciente.

Las variantes encontradas en la familia con diagnóstico de síndrome Patterson-Stevenson-Fontaine tienen frecuencias elevadas en relación con la incidencia de la enfermedad, por lo que estas fueron descartadas como causa de la enfermedad.

Es necesario considerar que la alteración podría ser muy grande y no fue detectable por medio de esta técnica, también es preciso buscar mutaciones en regiones que este panel no contempla como regiones reguladoras, secuencias intergénicas, o regiones con repetidos.

En el caso de la familia 2 el diagnóstico clínico de síndrome Roifman-Chitayat se modificó al realizar los análisis de las variantes en la paciente y sus padres. Se encontró que en realidad la paciente cursaba con dos enfermedades, el síndrome SOPH acompañado de falla hepática aguda o como algunos autores lo refieren; un fenotipo intermedio del espectro causado por mutaciones homocigotas en el gen *NBAS*. La mutación encontrada en esta paciente no había sido reportada anteriormente.

En este caso el haber llegado al diagnóstico correcto mejora el manejo de la paciente principalmente de manera farmacológica para evitar exacerbar el problema hepático.

Este resultado muestra la relevancia de este tipo de estudios y resalta la importancia de que la medicina traslacional sea integrada al diagnóstico clínico.

Perspectivas

Para la familia 1 síndrome PSF es necesario ampliar las opciones de búsqueda. Posiblemente la causa de la enfermedad se encuentra en una región no analizada en este trabajo o no se pudo identificar debido a las limitaciones de la técnica.

En la familia de la paciente con síndrome SOPH sería importante poder corroborar cual es el efecto de la variante sobre el transcrito y más importante aún sería averiguar cuál es la función específica de *NBAS* en la sintaxina 18 y su relación con la calcificación ósea.

Anexo I Grupos de displasias esqueléticas

1. Condrodisplasias FGFR2
2. Grupo colágenas tipo 2
3. Grupo colágenas tipo 11
4. Trastornos de sulfatación
5. Grupo perlecan
6. Grupo arlecan
7. Filamina y enfermedades relacionadas
8. Grupo TRPV4
9. Ciliopatías con presentación esquelética
10. Displacia epifisaria múltiple y pseudocondroplasia
11. Displasias metaepifisarias
12. Displasias espondiloepifisarias
13. Displasias espónidloepimetafisarias
14. Displasias espondilodisplásicas severas
15. Displasias acromélicas
16. Displasias acromesomélicas
17. Displasias mesomélicas y rizomélicas
18. Displasia campomélica y enfermedades relacionadas
19. Displasias de huesos delgados
20. Displasias con múltiples luxaciones
21. Condrodisplasia punctata
22. Displasias neonatales osteoescleróticas
23. Osteopetrosis y enfermedades relacionadas
24. Otras enfermedades óseas esclerosantes
25. Osteogénesis imperfectas
26. Mineralización anormal

27. Enfermedades de depósito lisosomal y compromiso esquelético
28. Osteólisis
29. Desarrollo esquelético desorganizado
30. Síndromes de sobrecrecimiento y compromiso esquelético
31. Osteoartropatías genéticas inflamatorias
32. Displasia cleidocranial y defectos de osificación
33. Síndromes de craniosinostosis
34. Disostosis de compromiso craneofacial
35. Disostosis de compromiso vertebral y costal
36. Disostosis rotulianas
37. Braquidactilias sin manifestaciones esqueléticas
38. Braquidactilias con manifestaciones esqueléticas
39. Defectos de reducción
40. Ectrodactilia con otras manifestaciones
41. Polidactilia, sindactilia y trifalangismo
42. Defectos en la formación articular y sonostosis

Anexo II Cartas de consentimiento informado

 <p>HOSPITAL INFANTIL DE MÉXICO FEDERICO GÓMEZ INSTITUTO NACIONAL DE SALUD LABORATORIO DE GENÉTICA, GENÉTICA Y ASESORAMIENTO</p> 	 <p>HOSPITAL INFANTIL DE MÉXICO FEDERICO GÓMEZ INSTITUTO NACIONAL DE SALUD LABORATORIO DE GENÉTICA, GENÉTICA Y ASESORAMIENTO</p> 
<p align="center">Carta de consentimiento informado</p> <p align="center"><u>DESCUBRIMIENTO DE GENES DE ENFERMEDADES RARAS Y ULTRA-RARAS</u></p> <p>Estimado madre y/o padre:</p> <p>Por medio de esta carta, lo invitamos a participar en el estudio mencionado arriba. Usted debe saber que participar en esta investigación es totalmente voluntario, de manera que usted puede decidir no participar en el estudio y en este caso no perdiera ninguna prestación a la que tiene derecho.</p> <p>Debido a que la información que se recibe es absolutamente confidencial y con fines de investigación, usted debe saber que no daremos a conocer ninguna información acerca de su hijo(a), sin su consentimiento.</p> <p>JUSTIFICACIÓN DEL ESTUDIO La presente investigación busca determinar la causa genética de la enfermedad que tiene su hijo(a). En algunos casos se estudian todos los genes, y en otros todo el material genético, con el objeto de hallar la alteración genética. Esta información contribuye a comprender el funcionamiento del material hereditario y la manera en la cual las alteraciones que presenta dan el origen a enfermedades específicas. En algunos casos, esta información permite brindar un asesoramiento genético más amplio. Asimismo, en algunos casos, pero ciertamente no en todos, la información podrá ser de utilidad para que médico oriente el tratamiento o el manejo del paciente.</p> <p>OBJETIVO DEL ESTUDIO Su hijo(a) tiene una enfermedad con una base genética. El propósito de este estudio es la búsqueda de alteraciones en el material genético, que se extrae de la muestra sanguínea. La identificación de dicha alteración puede</p> <p>PROCEDIMIENTO DE ESTUDIO Para este estudio requerimos una muestra de sangre de 1-3 ml, mediante el siguiente procedimiento:</p> <ul style="list-style-type: none">➤ Se explica detalladamente el procedimiento al paciente y/o familiar.➤ Bajo las condiciones higiénicas óptimas se coloca el brazo extendido y se selecciona la vena por palpación, preferentemente en el antebrazo.➤ Mediante tubo S-Monovette® y dispositivo iVette®-Set para toma de muestra sanguínea se toman de 1-3 ml de sangre, aproximadamente.➤ La muestra sanguínea será almacenada a -04°C hasta el momento de extraer el material genético.	<p>RIESGOS Y MOLESTIAS Los posibles riesgos de la toma de muestra sanguínea son molestia leve en el sitio donde se toma la muestra, formación de moretones, hemorragia o infección.</p> <p>BENEFICIOS Se les proporcionará el asesoramiento genético con la información que llegue a integrarse del estudio al cual lo estamos invitando.</p> <p>En ciertos casos, pero no en todos, esta información puede ser de utilidad para orientar el manejo del paciente.</p> <p>Además, en caso de encontrarse otros hallazgos se informará lo aceptado internamente, siempre y cuando Usted desee conocer esta información.</p> <p>Su colaboración permite obtener conocimientos que tal vez ayuden a otras familias que presenten problemas similares en un futuro.</p> <p>COSTOS Las pruebas realizadas en la muestra sanguínea así como cualquier consulta que usted tenga con el médico en relación al estudio de investigación, no tendrán ningún costo. Los gastos adicionales generados por esta investigación serán absorbidos por la institución.</p> <p>DUDAS Y ACLARACIONES Usted tiene la garantía de recibir respuesta a sus preguntas y aclaración a cualquier duda acerca de los procedimientos, riesgos y beneficios y otros asuntos relacionados con la investigación y tratamiento del sujeto. Si este fuera el caso, deberá comunicarse con los médicos responsables de la investigación: Dr. Jesús Aguirre Hernández y Dra. Verónica Fabiola Moran Gamboa del Departamento de Genética: 52 25 99 17 ext. 4429 y/o la Dirección de Investigación del HILFQ: Ext. 2361</p> <p>Usted tiene la libertad de retirar este consentimiento en cualquier momento, y dejar de participar en el estudio sin que esto afecte los cuidados y el tratamiento que la institución otorga.</p> <p>La identidad del sujeto del estudio será confidencial, y la información relacionada será manejada de acuerdo con las políticas de privacidad vigentes.</p> <p>Le sugerimos que conserve una copia de este documento para consultarla si es necesario.</p>
 <p>INSTITUTO NACIONAL DE SALUD AFILIADO A LA OMS Tel.: México 555 2441, Telefax: México 555 2441, Ciudad México, C.P. 06702 México D.F., Carretera México-Toluca 3500-400 km. 1, 06000 y 060 00 www.hilfgo.gob.mx</p>	 <p>INSTITUTO NACIONAL DE SALUD AFILIADO A LA OMS Tel.: México 555 2441, Telefax: México 555 2441, Ciudad México, C.P. 06702 México D.F., Carretera México-Toluca 3500-400 km. 1, 06000 y 060 00 www.hilfgo.gob.mx</p>

Anexo III. Carta de aprobación del comité de bioseguridad

SALUD INSTITUTO NACIONAL DE SALUD AFILIADO A LA UNAM
 INSTITUTO NACIONAL DE SALUD AFILIADO A LA UNAM
 COMISIÓN DE BIOSEGURIDAD

México, D.F., a 23 de enero de 2017 DI 4000/ 009 /2017

Dr. Jesús Aguirre Hernández
 Laboratorio de Genética, Genética y Bioinformática
 Presente

Informo a usted, que los Comités de Investigación, Ética y Bioseguridad, después de revisar el protocolo **HIM 2017-018** "Descubrimiento por secuenciación de segunda generación, de variantes asociadas a distasias espinéticas de causa genética desconocida". Al respecto me permito informarle que los Comités han decidido emitir un dictamen favorable al mismo.

No omito recordarle que la autorización para que este estudio se lleve a cabo será otorgada por la Dirección General de nuestra institución.

Como investigador responsable de este protocolo será su obligación asegurarse de que sea revisado a cabalidad y como fue enviado a los Comités antes citados, cualquier modificación al protocolo o a la carta de consentimiento informado requiere ser notificada para ser evaluada por los mismos. Deberá ajustarse a las normas éticas vigentes y deberá emitir cualquier información que le sea solicitada a este Comité sobre su desarrollo. Cabe mencionar que para el registro de su protocolo es importante entregar invariablemente, al Departamento de Control y Gestión a Protocolos de Investigación el Formulario PIG4 en un plazo no mayor a 10 días naturales a partir de la recepción de este documento.

Atentamente

Dr. Juan Carlos Rodríguez
Presidente del Comité de Investigación
Dr. Juan Carlos Rodríguez
Presidente del Comité de Ética en Investigación

Dra. María Inés García
Presidenta del Comité de Bioseguridad

Con copia:
 Dr. José Alberto García Aranda Director General
 Lic. Martha Reynoso Robles Jefa del Departamento de Control y Gestión a protocolos de Investigación
 JAG/IGI/JCA/IR

INSTITUTO NACIONAL DE SALUD AFILIADO A LA UNAM
 DR. MÁRQUEZ INÉS, COK, DOCTORAS DEL QUADRIÉNIO, C.P. INT31 MEXICO D.F.
 COMISIÓN DE BIOSEGURIDAD, 5228-9927 ext. 4303 y 4300
 www.inhs@inhs.mx

SALUD INSTITUTO NACIONAL DE SALUD AFILIADO A LA UNAM
 INSTITUTO NACIONAL DE SALUD AFILIADO A LA UNAM
 COMISIÓN DE BIOSEGURIDAD

México, D.F., a 23 de enero de 2017 DG/1000/ 054 /2017

Dr. Jesús Aguirre Hernández
 Laboratorio de Genética, Genética y Bioinformática
 Presente

Informo a usted, que los Comités de Investigación, Ética y Bioseguridad, después de haber revisado su protocolo **HIM 2017-018** "Descubrimiento, por secuenciación de segunda generación, de variantes asociadas a distasias espinéticas de causa genética desconocida", han emitido el dictamen de:

APROBADO

En los términos y condiciones señalados por dichos Comités. Por lo anterior, se autoriza su desarrollo.

Atentamente

Dr. José Alberto García Aranda
Director General

Con copia:
 Lic. Martha Reynoso Robles Jefa del Departamento de Control y Gestión a protocolos de Investigación
 JAGA/IGI/JCA/IR

INSTITUTO NACIONAL DE SALUD AFILIADO A LA UNAM
 DR. MÁRQUEZ INÉS, COK, DOCTORAS DEL QUADRIÉNIO, C.P. INT31 MEXICO D.F.
 COMISIÓN DE BIOSEGURIDAD, 5228-9927 ext. 4303 y 4300
 www.inhs@inhs.mx

Anexo IV. Carta de aprobación consentimiento informado para toma de fotografía



**HOSPITAL INFANTIL DE MEXICO
FEDERICO GOMEZ**
INSTITUTO NACIONAL DE SALUD
DR. MÁRQUEZ NO. 162 COL. DOCTORES,
DELEGACIÓN CUAUHTÉMOC, C.P. 06720 MÉXICO, D.F.
INSTITUCIÓN DE SERVICIO MÉDICO, ENSEÑANZA E INVESTIGACIÓN
AFILIADO A LA UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

5228-9917

México, D.F. a ____ de ____ de 20__

A QUIEN CORRESPONDA:

Por medio de la presente deseo manifestar que estoy de acuerdo con la invitación que me ha/han hecho el Doctor/Doctores

del Departamento de Genética de Hospital Infantil de México Federico Gómez para que sean publicadas y/o presentadas en congresos científicos nacionales e internacionales las características clínicas de mi caso y/o el de mi hijo, en forma escrita o en fotografías.

Se me ha explicado que ello obedece al interés científico de compartir conocimiento con la comunidad médica y científica internacional que pudiese en un futuro redundar en un mejor conocimiento y entendimiento de la enfermedad de que se trate además del que actualmente se tiene. Se me ha explicado que de realizarse dicha publicación o presentación en congreso, esta será un medio científico, reconocido nacional e internacionalmente, que se tomarán todas las medidas necesarias para preservar la confidencialidad del caso y todo aquello que pudiese permitir la identificación de la persona de quien se trate, incluyendo la modificación de las fotografías clínicas para preservar el anonimato.

Por lo anterior establezco aquí que he tomado esta decisión libremente y que se me ha explicado que no es obligatorio el participar y que no percibiré por ello ningún beneficio en especie o económico de ninguna naturaleza y de la misma manera que de negarme a participar ello no influiría en el trato y prestaciones que recibo por parte del Hospital Infantil de México Federico Gómez, obedeciendo a los derechos que como paciente tengo y al respeto a mi toma de decisiones y de mi dignidad.

Por lo anterior firmo este documento de conformidad y recibo una copia del documento.

Nombre y firma _____

Paciente/tutor
Parentesco con el paciente _____

Testigo 1 _____ Testigo 2 _____

Referencias

- Abecasis G.R., Auton A., Brooks L.D., DePristo M.A., Durbin R.M., Handsaker R.E., Kang H.M., Marth G.T. & McVean G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65.
- Adzhubei I.A., Schmidt S., Peshkin L., Ramensky V.E., Gerasimova A., Bork P., Kondrashov A.S. & Sunyaev S.R. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-9.
- Amberger J.S., Bocchini C.A., Schiettecatte F., Scott A.F. & Hamosh A. (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* **43**, D789-98.
- Aminoff M., Carter J.E., Chadwick R.B., Johnson C., Grasbeck R., Abdelaal M.A., Broch H., Jenner L.B., Verroust P.J., Moestrup S.K., de la Chapelle A. & Krahe R. (1999) Mutations in CUBN, encoding the intrinsic factor-vitamin B12 receptor, cubilin, cause hereditary megaloblastic anaemia 1. *Nat Genet* **21**, 309-13.
- Amrani N., Dong S., He F., Ganesan R., Ghosh S., Kervestin S., Li C., Mangus D.A., Spatrack P. & Jacobson A. (2006) Aberrant termination triggers nonsense-mediated mRNA decay. *Biochem Soc Trans* **34**, 39-42.
- Aoki T., Ichimura S., Itoh A., Kuramoto M., Shinkawa T., Isobe T. & Tagaya M. (2009) Identification of the neuroblastoma-amplified gene product as a component of the syntaxin 18 complex implicated in Golgi-to-endoplasmic reticulum retrograde transport. *Mol Biol Cell* **20**, 2639-49.
- Auton A., Brooks L.D., Durbin R.M., Garrison E.P., Kang H.M., Korbel J.O., Marchini J.L., McCarthy S., McVean G.A. & Abecasis G.R. (2015) A global reference for human genetic variation. *Nature* **526**, 68-74.
- Ayme S. (2003) [Orphanet, an information site on rare diseases]. *Soins*, 46-7.
- Balasubramanian M., Hurst J., Brown S., Bishop N.J., Arundel P., DeVile C., Pollitt R.C., Crooks L., Longman D., Caceres J.F., Shackley F., Connolly S., Payne J.H., Offiah A.C., Hughes D., Study D.D.D., Parker M.J., Hide W. & Skerry T.M. (2017) Compound heterozygous variants in NBAS as a cause of atypical osteogenesis imperfecta. *Bone* **94**, 65-74.
- Bonafe L., Cormier-Daire V., Hall C., Lachman R., Mortier G., Mundlos S., Nishimura G., Sangiorgi L., Savarirayan R., Sillence D., Spranger J., Superti-Furga A., Warman M. & Unger S. (2015) Nosology and classification of genetic skeletal disorders: 2015 revision. *Am J Med Genet A* **167**, 2869-92.
- Boycott K.M., Rath A., Chong J.X., Hartley T., Alkuraya F.S., Baynam G., Brookes A.J., Brudno M., Carracedo A., den Dunnen J.T., Dyke S.O.M., Estivill X., Goldblatt J., Gonthier C., Groft S.C., Gut I., Hamosh A., Hieter P., Hohn S., Hurles M.E., Kaufmann P., Knoppers B.M., Krischer J.P., Macek M., Jr., Matthijs G., Olry A., Parker S., Paschall J., Philippakis A.A., Rehm H.L., Robinson P.N., Sham P.C., Stefanov R., Taruscio D., Unni D., Vanstone M.R., Zhang F., Brunner H., Bamshad M.J. & Lochmuller H. (2017) International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet* **100**, 695-705.
- Brunak S., Engelbrecht J. & Knudsen S. (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol* **220**, 49-65.
- Carss K.J., Arno G., Erwood M., Stephens J., Sanchis-Juan A., Hull S., Megy K., Grozeva D., Dewhurst E., Malka S., Plagnol V., Penkett C., Stirrups K., Rizzo R., Wright G., Josifova D., Bitner-Glindzicz M., Scott R.H., Clement E., Allen L., Armstrong R., Brady A.F., Carmichael J., Chitre M., Henderson R.H., Hurst J., MacLaren R.E., Murphy E., Paterson J., Rosser E., Thompson D.A., Wakeling E., Ouwehand W.H., Michaelides M., Moore A.T., Webster A.R. & Raymond

- F.L. (2017) Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. *Am J Hum Genet* **100**, 75-90.
- Chaisson M.J., Huddleston J., Dennis M.Y., Sudmant P.H., Malig M., Hormozdiari F., Antonacci F., Surti U., Sandstrom R., Boitano M., Landolin J.M., Stamatoyannopoulos J.A., Hunkapiller M.W., Korf J. & Eichler E.E. (2015) Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517**, 608-11.
- Chang Y.F., Imam J.S. & Wilkinson M.F. (2007) The Nonsense-Mediated Decay RNA Surveillance Pathway. *Annu Rev Biochem*.
- Collin S.M., Metcalfe C., Palmer T.M., Refsum H., Lewis S.J., Smith G.D., Cox A., Davis M., Marsden G., Johnston C., Lane J.A., Donovan J.L., Neal D.E., Hamdy F.C., Smith A.D. & Martin R.M. (2011) The causal roles of vitamin B(12) and transcobalamin in prostate cancer: can Mendelian randomization analysis provide definitive answers? *Int J Mol Epidemiol Genet* **2**, 316-27.
- Consortium E.P. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-40.
- Conti E. & Izaurralde E. (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr Opin Cell Biol* **17**, 316-25.
- Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G. & Durbin R. (2011) The variant call format and vcftools. *Bioinformatics* **27**, 2156-8.
- DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R., Hartl C., Philippakis A.A., del Angel G., Rivas M.A., Hanna M., McKenna A., Fennell T.J., Kernytsky A.M., Sivachenko A.Y., Cibulskis K., Gabriel S.B., Altshuler D. & Daly M.J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8.
- Desmet F.O., Hamroun D., Lalande M., Collod-Beroud G., Claustres M. & Beroud C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37**, e67.
- Dos Santos C.O., da Silva-Junior F.P., Puga R.D., Barbosa E.R., Azevedo Silva S.M.C., Borges V., Limongi J.C.P., Rocha M.S.G., Ferraz H.B. & de Carvalho Aguiar P. (2018) The prevalence of PRKRA mutations in idiopathic dystonia. *Parkinsonism Relat Disord* **48**, 93-6.
- Firth H.V.a. & Hurst J.A.a. *Clinical genetics and genomics*. Oxford press. UK 2017.
- Forman J., Taruscio D., Llera V.A., Barrera L.A., Cote T.R., Edfjall C., Gavhed D., Haffner M.E., Nishimura Y., Posada M., Tambuyzer E., Groft S.C. & Henter J.I. (2012) The need for worldwide policy and action plans for rare diseases. *Acta Paediatr* **101**, 805-7.
- Fromer M., Moran J.L., Chambert K., Banks E., Bergen S.E., Ruderfer D.M., Handsaker R.E., McCarroll S.A., O'Donovan M.C., Owen M.J., Kirov G., Sullivan P.F., Hultman C.M., Sklar P. & Purcell S.M. (2012) Discovery and Statistical Genotyping of Copy-Number Variation from Whole-Exome Sequencing Depth. *Am J Hum Genet* **91**, 597-607.
- Geister K.A. & Camper S.A. (2015) Advances in Skeletal Dysplasia Genetics. *Annu Rev Genomics Hum Genet* **16**, 199-227.
- Haack T.B., Stauffer C., Kopke M.G., Straub B.K., Kolker S., Thiel C., Freisinger P., Baric I., McKiernan P.J., Dikow N., Harting I., Beisse F., Burgard P., Kotzaeridou U., Kuhr J., Himbert U., Taylor R.W., Distelmaier F., Vockley J., Ghaloul-Gonzalez L., Zschocke J., Kremer L.S., Graf E., Schwarzmayr T., Bader D.M., Gagneur J., Wieland T., Terrile C., Strom T.M.,

- Meitinger T., Hoffmann G.F. & Prokisch H. (2015) Biallelic Mutations in NBAS Cause Recurrent Acute Liver Failure with Onset in Infancy. *Am J Hum Genet* **97**, 163-9.
- Hamosh A., Scott A.F., Amberger J., Valle D. & McKusick V.A. (2000) Online Mendelian Inheritance in Man (OMIM). *Hum Mutat* **15**, 57-61.
- Hunt R.C., Simhadri V.L., Iandoli M., Sauna Z.E. & Kimchi-Sarfaty C. (2014) Exposing synonymous mutations. *Trends Genet* **30**, 308-21.
- Kohler S., Vasilevsky N.A., Engelstad M., Foster E., McMurry J., Ayme S., Baynam G., Bello S.M., Boerkoel C.F., Boycott K.M., Brudno M., Buske O.J., Chinnery P.F., Cipriani V., Connell L.E., Dawkins H.J., DeMare L.E., Devereau A.D., de Vries B.B., Firth H.V., Freson K., Greene D., Hamosh A., Helbig I., Hum C., Jahn J.A., James R., Krause R., SJ F.L., Lochmuller H., Lyon G.J., Ogishima S., Olry A., Ouwehand W.H., Pontikos N., Rath A., Schaefer F., Scott R.H., Segal M., Sergouniotis P.I., Sever R., Smith C.L., Straub V., Thompson R., Turner C., Turro E., Veltman M.W., Vulliamy T., Yu J., von Ziegenweid J., Zankl A., Zuchner S., Zemojtel T., Jacobsen J.O., Groza T., Smedley D., Mungall C.J., Haendel M. & Robinson P.N. (2017) The Human Phenotype Ontology in 2017. *Nucleic Acids Res* **45**, D865-D76.
- Kortum F., Marquardt I., Alawi M., Korenke G.C., Spranger S., Meinecke P. & Kutsche K. (2017) Acute Liver Failure Meets SOPH Syndrome: A Case Report on an Intermediate Phenotype. *Pediatrics* **139**.
- Kulkarni S.e. & Pfeifer J.D.e. *CHECK: Clinical genomics*.
- Kumar P., Henikoff S. & Ng P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-81.
- Kzhyshkowska J., Gratchev A. & Goerd S. (2006) Stabilin-1, a homeostatic scavenger receptor with multiple functions. *J Cell Mol Med* **10**, 635-49.
- Landrum M.J., Lee J.M., Riley G.R., Jang W., Rubinstein W.S., Church D.M. & Maglott D.R. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-5.
- Lek M., Karczewski K.J., Minikel E.V., Samocha K.E., Banks E., Fennell T., O'Donnell-Luria A.H., Ware J.S., Hill A.J., Cummings B.B., Tukiainen T., Birnbaum D.P., Kosmicki J.A., Duncan L.E., Estrada K., Zhao F., Zou J., Pierce-Hoffman E., Berghout J., Cooper D.N., Deflaux N., DePristo M., Do R., Flannick J., Fromer M., Gauthier L., Goldstein J., Gupta N., Howrigan D., Kiezun A., Kurki M.I., Moonshine A.L., Natarajan P., Orozco L., Peloso G.M., Poplin R., Rivas M.A., Ruano-Rubio V., Rose S.A., Ruderfer D.M., Shakir K., Stenson P.D., Stevens C., Thomas B.P., Tiao G., Tusie-Luna M.T., Weisburd B., Won H.H., Yu D., Altshuler D.M., Ardissino D., Boehnke M., Danesh J., Donnelly S., Elosua R., Florez J.C., Gabriel S.B., Getz G., Glatt S.J., Hultman C.M., Kathiresan S., Laakso M., McCarroll S., McCarthy M.I., McGovern D., McPherson R., Neale B.M., Palotie A., Purcell S.M., Saleheen D., Scharf J.M., Sklar P., Sullivan P.F., Tuomilehto J., Tsuang M.T., Watkins H.C., Wilson J.G., Daly M.J. & MacArthur D.G. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91.
- Li H. & Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60.
- Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. & Durbin R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9.
- Longman D., Hug N., Keith M., Anastasaki C., Patton E.E., Grimes G. & Caceres J.F. (2013) DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and *Caenorhabditis elegans*. *Nucleic Acids Res* **41**, 8319-31.

- Maksimova N., Hara K., Nikolaeva I., Chun-Feng T., Usui T., Takagi M., Nishihira Y., Miyashita A., Fujiwara H., Oyama T., Nogovicina A., Sukhomyasova A., Potapova S., Kuwano R., Takahashi H., Nishizawa M. & Onodera O. (2010) Neuroblastoma amplified sequence gene is associated with a novel short stature syndrome characterised by optic nerve atrophy and Pelger-Huet anomaly. *J Med Genet* **47**, 538-48.
- Maquat L.E. (2005) Nonsense-mediated mRNA decay in mammals. *J Cell Sci* **118**, 1773-6.
- McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M. & DePristo M.A. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303.
- McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R., Thormann A., Flicek P. & Cunningham F. (2016) The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122.
- Neu-Yilik G., Gehring N.H., Hentze M.W. & Kulozik A.E. (2004) Nonsense-mediated mRNA decay: from vacuum cleaner to Swiss army knife. *Genome Biol* **5**, 218.
- Otten C., van der Ven P.F., Lewrenz I., Paul S., Steinhagen A., Busch-Nentwich E., Eichhorst J., Wiesner B., Stemple D., Strahle U., Furst D.O. & Abdelilah-Seyfried S. (2012) Xirp proteins mark injured skeletal muscle in zebrafish. *PLoS One* **7**, e31041.
- Oud M.M., Tuijnenburg P., Hempel M., van Vlies N., Ren Z., Ferdinandusse S., Jansen M.H., Santer R., Johannsen J., Bacchelli C., Alders M., Li R., Davies R., Dupuis L., Cale C.M., Wanders R.J.A., Pals S.T., Ocaña L., James C., Muller I., Lehmeberg K., Strom T., Engels H., Williams H.J., Beales P., Roepman R., Dias P., Brunner H.G., Cobben J.M., Hall C., Hartley T., Le Quesne Stabej P., Mendoza-Londono R., Davies E.G., de Sousa S.B., Lessel D., Arts H.H. & Kuijpers T.W. (2017) Mutations in EXTL3 Cause Neuro-immuno-skeletal Dysplasia Syndrome. *Am J Hum Genet* **100**, 281-96.
- Patterson T.J. & Stevenson A.C. (1964) CHANGE REF: Cranio-Facial Dysostosis and Malformations of Feet. *J Med Genet* **1**, 112-4.
- Poultney C.S., Goldberg A.P., Drapeau E., Kou Y., Harony-Nicolas H., Kajiwaru Y., De Rubeis S., Durand S., Stevens C., Rehnstrom K., Palotie A., Daly M.J., Ma'ayan A., Fromer M. & Buxbaum J.D. (2013) Identification of small exonic CNV from whole-exome sequence data and application to autism spectrum disorder. *Am J Hum Genet* **93**, 607-19.
- Rappaport N., Twik M., Plaschkes I., Nudel R., Iny Stein T., Levitt J., Gershoni M., Morrey C.P., Safran M. & Lancet D. (2017) MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res* **45**, D877-D87.
- Reese M.G., Eeckman F.H., Kulp D. & Haussler D. (1997) Improved splice site detection in Genie. *J Comput Biol* **4**, 311-23.
- Retterer K., Scuffins J., Schmidt D., Lewis R., Pineda-Alvarez D., Stafford A., Schmidt L., Warren S., Gibellini F., Kondakova A., Blair A., Bale S., Matyakhina L., Meck J., Aradhya S. & Haverfield E. (2015) Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. *Genet Med* **17**, 623-9.
- Robinson P.N., Kohler S., Oellrich A., Sanger Mouse Genetics P., Wang K., Mungall C.J., Lewis S.E., Washington N., Bauer S., Seelow D., Krawitz P., Gilissen C., Haendel M. & Smedley D. (2014a) Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res* **24**, 340-8.
- Robinson P.N., Kohler S., Oellrich A., Wang K., Mungall C.J., Lewis S.E., Washington N., Bauer S., Seelow D., Krawitz P., Gilissen C., Haendel M. & Smedley D. (2014b) Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res* **24**, 340-8.

- Roifman C.M. & Chitayat D. (2009) Combined immunodeficiency, facial dysmorphism, optic nerve atrophy, skeletal anomalies and developmental delay: a new syndrome. *Clin Genet* **76**, 449-57.
- Ruderfer D.M., Hamamsy T., Lek M., Karczewski K.J., Kavanagh D., Samocha K.E., Daly M.J., MacArthur D.G., Fromer M. & Purcell S.M. (2016) Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat Genet* **48**, 1107-11.
- Samarakoon P.S., Sorte H.S., Kristiansen B.E., Skodje T., Sheng Y., Tjonnfjord G.E., Stadheim B., Stray-Pedersen A., Rodningen O.K. & Lyle R. (2014) Identification of copy number variants from exome sequence data. *BMC Genomics* **15**, 661.
- Sauna Z.E. & Kimchi-Sarfaty C. (2011) Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* **12**, 683-91.
- Segarra N.G., Ballhausen D., Crawford H., Perreau M., Campos-Xavier B., van Spaendonck-Zwarts K., Vermeer C., Russo M., Zambelli P.Y., Stevenson B., Royer-Bertrand B., Rivolta C., Candotti F., Unger S., Munier F.L., Superti-Furga A. & Bonafe L. (2015) NBAS mutations cause a multisystem disorder involving bone, connective tissue, liver, immune system, and retina. *Am J Med Genet A* **167A**, 2902-12.
- Sharfe N., Karanxha A., Dadi H., Merico D., Chitayat D., Herbrick J.A., Freeman S., Grinstein S. & Roifman C.M. (2018) Dual loss of p110delta PI3-kinase and SKAP (KNSTRN) expression leads to combined immunodeficiency and multisystem syndromic features. *J Allergy Clin Immunol* **142**, 618-29.
- Shendure J. & Ji H. (2008) Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-45.
- Short P.J., McRae J.F., Gallone G., Sifrim A., Won H., Geschwind D.H., Wright C.F., Firth H.V., FitzPatrick D.R., Barrett J.C. & Hurles M.E. (2018) De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature*.
- Slatko B.E., Gardner A.F. & Ausubel F.M. (2018) Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol* **122**, e59.
- Staufner C., Haack T.B., Kopke M.G., Straub B.K., Kolker S., Thiel C., Freisinger P., Baric I., McKiernan P.J., Dikow N., Harting I., Beisse F., Burgard P., Kotzaeridou U., Lenz D., Kuhr J., Himbert U., Taylor R.W., Distelmaier F., Vockley J., Ghaloul-Gonzalez L., Ozolek J.A., Zschocke J., Kuster A., Dick A., Das A.M., Wieland T., Terrile C., Strom T.M., Meitinger T., Prokisch H. & Hoffmann G.F. (2016) Recurrent acute liver failure due to NBAS deficiency: phenotypic spectrum, disease mechanisms, and therapeutic concepts. *J Inherit Metab Dis* **39**, 3-16.
- Szklarczyk D., Franceschini A., Wyder S., Forslund K., Heller D., Huerta-Cepas J., Simonovic M., Roth A., Santos A., Tsafou K.P., Kuhn M., Bork P., Jensen L.J. & von Mering C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447-52.
- Tan R., Wang Y., Kleinstein S.E., Liu Y., Zhu X., Guo H., Jiang Q., Allen A.S. & Zhu M. (2014) An evaluation of copy number variation detection tools from whole-exome sequencing data. *Hum Mutat* **35**, 899-907.
- Terrazas K., Dixon J., Trainor P.A. & Dixon M.J. (2017) Rare syndromes of the head and face: mandibulofacial and acrofacial dysostoses. *Wiley Interdiscip Rev Dev Biol* **6**.
- Van der Auwera G.A., Carneiro M.O., Hartl C., Poplin R., Del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella K.V., Altshuler D., Gabriel S. & DePristo M.A. (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **11**, 11 0 1- 0 33.
- Wappenschmidt B., Becker A.A., Hauke J., Weber U., Engert S., Kohler J., Kast K., Arnold N., Rhiem K., Hahnen E., Meindl A. & Schmutzler R.K. (2012) Analysis of 30 putative BRCA1 splicing

- mutations in hereditary breast and ovarian cancer families identifies exonic splice site mutations that escape in silico prediction. *PLoS One* **7**, e50800.
- Whiffin N., Minikel E., Walsh R., O'Donnell-Luria A.H., Karczewski K., Ing A.Y., Barton P.J.R., Funke B., Cook S.A., MacArthur D. & Ware J.S. (2017) Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med* **19**, 1151-8.
- Wieczorek D. (2013) Human facial dysostoses. *Clin Genet* **83**, 499-510.
- Wilderman A., VanOudenhove J., Kron P.J., Noonan J. & Cotney J. (2018) High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Reports* **23**, 1581–1597.
- Wilkie A.O. & Goodacre T.E. (1997) Patterson-Stevenson-Fontaine syndrome: 30-year follow-up and clinical details of a further affected case. *Am J Med Genet* **69**, 433-4.
- Xie Q. & Wang Z.A. (2017) Transcriptional regulation of the Nkx3.1 gene in prostate luminal stem cell specification and cancer initiation via its 3' genomic region. *J Biol Chem* **292**, 13521-30.
- Xiong F., Gao J., Li J., Liu Y., Feng G., Fang W., Chang H., Xie J., Zheng H., Li T. & He L. (2009) Noncanonical and canonical splice sites: a novel mutation at the rare noncanonical splice-donor cut site (IVS4+1A>G) of SEDL causes variable splicing isoforms in X-linked spondyloepiphyseal dysplasia tarda. *Eur J Hum Genet* **17**, 510-6.
- Yates A., Akanni W., Amode M.R., Barrell D., Billis K., Carvalho-Silva D., Cummins C., Clapham P., Fitzgerald S., Gil L., Giron C.G., Gordon L., Hourlier T., Hunt S.E., Janacek S.H., Johnson N., Juettemann T., Keenan S., Lavidas I., Martin F.J., Maurel T., McLaren W., Murphy D.N., Nag R., Nuhn M., Parker A., Patricio M., Pignatelli M., Rahtz M., Riat H.S., Sheppard D., Taylor K., Thormann A., Vullo A., Wilder S.P., Zadissa A., Birney E., Harrow J., Muffato M., Perry E., Ruffier M., Spudich G., Trevanion S.J., Cunningham F., Aken B.L., Zerbino D.R. & Flicek P. (2016) Ensembl 2016. *Nucleic Acids Res* **44**, D710-6.
- Zuben E. & Kimchi-Sarfaty. (2011) Understanding the contribution of synonymous mutations to human disease. *Nature* **12**, 683-691.