



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica Superior
d'Enginyeria Agrònoma i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica
y del Medio Natural

Caracterización de genes y mutaciones de trastornos del
movimiento

Trabajo Fin de Grado

Grado en Biotecnología

AUTOR/A: López Nieto, Marta

Tutor/a: Espinós Armero, Carmen Ángeles

Director/a Experimental: Martínez Rubio, María Dolores

CURSO ACADÉMICO: 2023/2024



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



VALENCIA BIOMEDICAL
RESEARCH FOUNDATION
CENTRO DE INVESTIGACIÓN PRÍNCIPE FELIPE

Caracterización de genes y mutaciones de trastornos del movimiento

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural

Trabajo Fin de Grado

Grado en Biotecnología

ALUMNA: Marta López Nieto

TUTORA UPV: Carmen Espinós Armero

TUTORA CIPF: Esther Martínez Martínez

DIRECTORA EXPERIMENTAL: María Dolores Martínez Rubio

Curso Académico: 2023-2024

VALENCIA, junio de 2024

Título/Títol	Caracterización de genes y mutaciones de trastornos del movimiento / Caracterització de gens i mutacions de trastorns del moviment
Autora	Dña.Marta López Nieto
Tutora CIPF	Dra. Esther Martínez Martínez
Dir. Experimental	Dña. María Dolores Martínez Rubio
Tutora UPV	Dra. Carmen Espinós Armero
Localidad y fecha	Valencia, junio de 2024

Resumen

Las enfermedades raras (EERR) se definen en la Unión Europea como aquellas que afectan a 1 persona de cada 2.000. Aunque una enfermedad rara afecta a muy pocos pacientes, se conocen más de 7.000 condiciones, por lo que el número total de enfermos es notable. Así, se estima que en España hay alrededor de 3.000.000 personas que padecen una enfermedad rara. La gran mayoría de ellas son monogénicas, pediátricas y discapacitantes. El desarrollo de las nuevas tecnologías de secuenciación, como la secuenciación de exoma (WES, *Whole Exome Sequencing*) o de genoma completo (WGS, *Whole Genome Sequencing*), ha permitido dar pasos de gigantes en el conocimiento de las bases genéticas de las EERR, permitiendo el diagnóstico y el consejo genético, y con ello, mejorando la atención que reciben los pacientes con EERR y sus familiares.

En el presente Trabajo Final de Grado se han analizado por filtrado y priorización de variantes de secuenciación de exoma completo (WES), dos familias con un paciente con trastorno de movimiento. El objetivo era hallar la causa genética de la enfermedad mediante WES del niño/a afecto/a y sus progenitores (WES-trío). Los resultados obtenidos a partir del ensayo de minigenes y del análisis de transcritos de sangre periférica de uno de los probandos aportan evidencia del carácter deletéreo de una variante hallada en homocigosis en el gen *SLC38A6*. Asimismo, muestran la necesidad de realizar más estudios sobre las variantes seleccionadas en el otro probando, reevaluando el exoma en el futuro cuando se conozca más información acerca de la función de dichos genes y de las proteínas codificadas por los mismos.

En paralelo, se ha investigado la patogenicidad de dos variantes genéticas identificadas en familias estudiadas en el laboratorio de Enfermedades Raras Neurodegenerativas. Se trata de cambios nucleotídicos de tipo *missense* que afectan a la proteína PTPN23. Variantes en el gen *PTPN23* se han asociado con un trastorno del neurodesarrollo caracterizado por retraso en el desarrollo y anomalías estructurales en el cerebro. Algunos autores atribuyen el mecanismo patogénico a la fosforilación y localización anormales de SMN, relacionada con PTPN23 por ser este un regulador de SMN, por lo que se ha cuantificado esta proteína y se han realizado estudios de localización subcelular en fibroblastos de paciente para determinar la posible relación con el fenotipo del probando. Los hallazgos experimentales no han permitido conocer la patogenicidad de las variantes estudiadas, lo cual ha evidenciado la urgencia de hacer más estudios, empleando diferentes técnicas y/o más réplicas y muestras.

Palabras clave: enfermedad rara; trastorno del movimiento; genética; secuenciación de exoma; *SLC38A6*; *PTPN23*.

Title	Characterization of genes and mutations associated to movement disorders
Author	Ms. Marta López Nieto
Tutor	Dr. Esther Martínez Martínez
Experimental director	Ms. María Dolores Martínez Rubio
Academic Tutor	Dr. Carmen Espinós Armero
Place and date	Valencia, June 2024

Abstract

Rare diseases (RDs) are defined in the European Union as those affecting 1 people in 2000. Although a certain rare disease affects very few patients, more than 7000 RDs are known nowadays, meaning that the total number of affected people is quite remarkable. Thus, it is estimated there are over 3 million people suffering a rare disease in Spain. Most of them are monogenic, pediatric, and disabling. The development of new sequencing technologies, such as whole exome sequencing (WES) or whole genome sequencing (WGS), has allowed to take huge steps in the knowledge of the genetic bases of RDs, enabling diagnosis and genetic counselling and, with it, improving the attention provided to patients suffering RDs and their relatives.

In the current bachelor's thesis, filtering and prioritization of exome sequencing variants has been performed, studying two families with a member affected by a movement disorder. The aim was to identify the genetic background of the disease by WES, with this purpose, the affected child was sequenced, together with his/her parents (WES-trio). Results shown by minigene assay and transcript analysis from the proband's peripheric blood demonstrate the deleterious nature of a variant in homozygosity found in the *SLC38A6* gene. Moreover, it is necessary to study more deeply the variants selected in the other patient, with further exome revaluation in the future when more information is known about the function of those genes and the proteins they encode.

In parallel, the pathogenicity of two genetic variants already identified in families studied by the Rare Neurodegenerative Diseases Laboratory has been explored. They are nucleotide *missense* changes with an impact in PTPN23 protein. Variants in *PTPN23* gene have been associated with a neurodevelopmental disorder characterized by developmental delay and structural brain anomalies. Some researchers attribute the pathogenicity of the disease to the abnormal phosphorylation and localization of SMN protein, which is linked to PTPN23 being the former a regulator of SMN, so we have quantified the named protein and performed subcellular localization studies in the patient's fibroblasts in order to determine the possible connection with the proband's phenotype. Experimental evidence does not allow to conclude anything about the pathogenicity of the studied variants, suggesting the need to perform more studies with different techniques and/or more replicates and samples.

Key words: rare disease; movement disorder; genetics; exome sequencing; *SLC38A6*; PTPN23.

Agradecimientos

En primer lugar, me gustaría agradecer a mi tutora, Carmen Espinós, por haberme abierto las puertas de su laboratorio y darme la oportunidad de adentrarme en el mundo de la investigación, aportando mi pequeño granito de arena a la investigación en enfermedades raras. También por todos sus consejos a lo largo de mi estancia en el laboratorio y por haber estado pendiente de mí en todo momento.

No puedo olvidarme de las chicas del I-54: Edna, Esther, Lola y Sofía; quienes me acogieron desde el primer día con una sonrisa, haciéndome sentir como en casa. Me gustaría agradecer especialmente a Esther y a Lola: ¡sin vosotras esto no habría sido posible! No tengo palabras de agradecimiento suficientes para expresar todo lo que habéis hecho por mí durante estos 6 meses. Gracias de corazón por enseñarme con una paciencia infinita, por acompañarme en todos los experimentos y animarme cuando no salían como esperábamos, por resolver mis numerosas dudas, y por sacar tiempo de donde no lo había para estar conmigo. A Lola, gracias por confiar tanto en mí, por enseñarme incansablemente dónde estaba cada cosa y todo lo necesario respecto a la parte de genética, y por preocuparte por mí. A Esther, gracias por todos los trucos de laboratorio, por enseñarme a desenvolverme en cultivos y con las múltiples inmunos que hicimos (casi podría decirse que soy una experta jajajaja), por ayudarme a entender cada cosa que hacíamos y estar siempre dispuesta a explicarme más, y por corregirme estos últimos días el TFG con una rapidez increíble. Como no, gracias por aportar el toque dulce al lab con esos deliciosos *muffins* de zanahoria y canela que preparabas, Esther. A Edna y a Sofía, gracias por atenderme tan amablemente siempre que os he preguntado algo, por aconsejarme en todos los aspectos, y por animar el laboratorio con vuestras historietas. También me gustaría mencionar a M Carmen y Christian, del I-56, por ser unos compis de lab estupendos; y especialmente a M Carmen por ayudarme en todo, con la clonación TOPO-TA, aunque al final no saliera (qué se le va a hacer), por tus consejos y por interesarte por mí y animarme a seguir cada día.

Por otra parte, mis amigos de la carrera han sido fundamentales a lo largo de este camino: Júlia, Lluna, Marta F, Marta M y Miguel. Por esas mañanas y tardes de biblio en tercero, que juntos eran más llevaderas; por esas quedadas después de los exámenes para despejarnos y coger fuerzas para seguir; por los trabajos juntos; por las fiestas biotec; por los viajes; por las anécdotas; y por todo lo que nos queda por vivir. Mención especial para Marta, mi llombaina favorita, por ser el alma de la fiesta, por buscar todos los planes habidos y por haber, por tener tanta energía y contagiarla, por esas cenas de shushi hablando de la vida, por acogernos en tu casa siempre, por tus conciertos de piano, por confiar plenamente en mí (incluso más que yo misma), y por querer tanto a tus mellis.

Por último, pero no por ello menos importante, a mi familia. A mis padres, por sufrir conmigo en el camino, por la paciencia con todo el estrés, risas y llantos durante estos 4 años (y siempre en realidad), por esos *tuppers* tan ricos que me hacían coger fuerzas para afrontar la tarde de prácticas, del laboratorio, de estudio...; por confiar en mí y apoyarme en mis decisiones; y por todos los esfuerzos que hacéis día a día. A mi hermano, Pablo, pieza fundamental del puzzle, por nuestra conexión, por entenderme como nadie, por nuestras risas (y a veces llantos), por ayudarme siempre con todas mis dudas, por hacer el camino conmigo, por cuidarme, por hacer de hermano mayor aunque no fuera el caso, simplemente gracias. A mi tía, por estar tan orgullosa de sus sobrinos, por sus paellas y cocas espectaculares, por venir a vernos siempre que puede. ¡Muchas gracias a todos!

Índice de contenido

Relación del trabajo con los ODS (Objetivos de Desarrollo Sostenible) Agenda 2030

1. INTRODUCCIÓN	1
1.1. ¿QUÉ SON LAS ENFERMEDADES RARAS?	1
1.1.1. Trastornos del movimiento	2
1.2. Uso de tecnologías NGS para el diagnóstico de enfermedades raras	3
1.2.1. Secuenciación de exoma completo (WES)	4
1.2.2. Filtrado y priorización de variantes	6
1.3. Familia de transportadores SLC: <i>SLC38A6</i>	7
1.4. PTPN23: funciones y relación con SMN1	8
2. HIPÓTESIS Y OBJETIVOS	9
3. MATERIALES Y MÉTODOS	10
3.1. Pacientes	10
3.2. Estudio genético: secuenciación de exoma	10
3.2.1. Filtrado y priorización de variantes	11
3.2.2. Validación de cambios candidatos	12
3.2.2.1. Diseño de cebadores	12
3.2.2.2. Amplificación por PCR	12
3.2.2.3. Análisis de las secuencias mediante secuenciación Sanger	13
3.2.3. Análisis de CNVs	14
3.3. Estudios funcionales para estudiar la patogenicidad de las variantes seleccionadas	14
3.3.1. Ensayo de minigenes del gen <i>SLC38A6</i>	14
3.3.1.1. Digestión doble del vector pSPL3 y del inserto	14
3.3.1.2. Ligación	16
3.3.1.3. Transformación en <i>Escherichia coli</i> DH5 α	16
3.3.1.4. PCR de colonias y extracción del vector en los clones con inserto	16
3.3.1.5. Transfección de células HeLa con el vector <i>wild type</i> y c.690+2T>G de <i>SLC38A6</i>	17
3.3.1.6. Extracción de ARN	18
3.3.1.7. Retrotranscripción a ADNc	19
3.3.1.8. Clonación del constructo WT en vector pcDNA3	20
3.3.2. Análisis de transcritos de sangre periférica de paciente de <i>SLC38A6</i>	20
3.3.2.1. Extracción de ARN del probando (MD-626)	20
3.3.2.2. Retrotranscripción a ADNc	21
3.3.3. Estudio de las proteínas PTPN23 y SMN1 en fibroblastos de paciente con variantes <i>missense</i> en PTPN23	21
3.3.3.1. Cultivo de fibroblastos primarios de paciente	21
3.3.3.2. Western blot	22

3.3.3.2.1.	Extracción de proteínas	22
3.3.3.2.2.	Cuantificación de proteínas mediante el método BCA	22
3.3.3.2.3.	Preparación de muestras para western blot	23
3.3.3.2.4.	Electroforesis de proteínas: SDS-PAGE	23
3.3.3.2.5.	Transferencia a la membrana de PVDF.....	24
3.3.3.2.6.	Inmunodetección de proteínas transferidas a la membrana.....	24
3.3.3.2.7.	Revelado mediante reacción de quimioluminiscencia	25
3.3.3.2.8.	Cuantificación y análisis estadístico	25
3.3.3.3.	Inmunofluorescencia (IFI).....	25
3.3.3.3.1.	Siembra y fijación de células en cristales.....	25
3.3.3.3.2.	Protocolo de inmunofluorescencia.....	25
3.3.3.3.3.	Montaje y observación de cristales por microscopía confocal	26
4.	RESULTADOS	26
4.1.	Validación de variantes candidatas en las familias	26
4.1.1.	Estudio de variaciones en el número de copias (CNVs).....	28
4.2.	Ensayos funcionales de la variante c.690+2T>G en el gen <i>SLC386</i>	30
4.2.1.	Ensayo de minigenes	30
4.2.2.	Análisis de transcritos.....	32
4.3.	Estudio de las proteínas PTPN23 y SMN1 en fibroblastos de paciente	33
4.3.1.	Western blot.....	33
4.3.2.	Inmunofluorescencia (IFI).....	34
5.	DISCUSIÓN	36
5.1.	Validación de cambios candidatos en las familias.....	36
5.2.	Caracterización de la variante c.690+2T>G en el gen <i>SLC386</i>	37
5.3.	Estudio de las proteínas PTPN23 y SMN1 en fibroblastos de paciente	38
5.3.1.	Estudio de la expresión de SMN1 mediante <i>western blot</i>	38
5.3.2.	Estudio de la localización subcelular de SMN1.....	38
6.	CONCLUSIONES	40
7.	BIBLIOGRAFÍA	41
8.	ANEXOS.....	47
8.1.	Anexo I: Figuras y tablas suplementarias.....	47
8.2.	Anexo II: Clonación de <i>SL38A6</i> en vector pcDNA3	53

Índice de Tablas

Tabla 1. Reactivos de la PCR Master Mix y programa del termociclador.

Tabla 2. Cebadores empleados para amplificar el inserto del gen *SLC38A6* para el ensayo de minigenes.

Tabla 3. Reactivos para la digestión del inserto *SLC38A6_E9* y del vector pSPL3 con XhoI y NheI.

Tabla 4. Reactivos para la ligación del inserto *SLC38A6_E9* y del vector pSPL3 digeridos.

Tabla 5. Cebadores empleados para la PCR de colonias del ensayo de minigenes.

Tabla 6. Reactivos de la PCR de colonias y programa del termociclador.

Tabla 7. Volúmenes a añadir de los reactivos para el protocolo de transfección para cada una de las mezclas.

Tabla 8. Volúmenes a añadir de los reactivos para la retrotranscripción.

Tabla 9. Cebadores empleados para la amplificación por PCR del ADNc del ensayo de minigenes.

Tabla 10. Volúmenes a añadir de los reactivos para la retrotranscripción.

Tabla 11. Cebadores empleados para la amplificación por PCR del ADNc de la sangre.

Tabla 12. Componentes del tampón de lisis celular para la extracción de proteínas.

Tabla 13. Preparación de los puntos de la curva patrón con BSA.

Tabla 14. Medidas de absorbancia de las muestras, recta patrón y preparación de muestras para la carga en el gel de poliacrilamida.

Tabla 15. Anticuerpos primarios y secundarios empleados en el protocolo de western blot.

Tabla 16. Anticuerpos primarios y secundarios empleados para cada experimento de inmunofluorescencia (IFI) y sus diluciones correspondientes.

Tabla 17. Variantes candidatas priorizadas en las familias analizadas por WES.

Índice de Figuras

Figura 1. Esquema del *pipeline* seguido en la secuenciación de exoma completo (WES).

Figura 2. Esquema del proceso de filtrado y priorización de variantes procedentes de secuenciación de exoma para enfermedades raras.

Figura 3. Pedigrís de las familias fMD-275 y fMD-279.

Figura 4. Esquema del vector pSPL3 y del inserto del gen *SLC38A6*.

Figura 5. Esquema del vector pcDNA3 y el constructo *SLC38A6* WT introducido.

Figura 6. Electroferograma de la secuenciación Sanger de la miniprep del vector pSPL3 de clones con el inserto del gen *SLC38A6*.

Figura 7. Resultados ensayo de minigenes.

Figura 8. Electroferograma de la secuenciación Sanger de la clonación en pcDNA3 del minigen *SLC38A6* WT.

Figura 9. Resultados del análisis de transcritos.

Figura 10. Resultados representativos del inmunoblot de SMN1 con extractos proteicos de NHF-F1A (sano) y MD-172 (paciente).

Figura 11. Cuantificación de los niveles de expresión de la proteína SMN1 en fibroblastos sanos (NHF-F1A) o de paciente (MD-172).

Figura 12. Imágenes representativas de los experimentos de inmunofluorescencia.

Acrónimos y Abreviaturas

AD: autosómico dominante

ADN: ácido desoxirribonucleico

ADNc: ácido desoxirribonucleico complementario, *complementary deoxyribonucleic acid (cDNA)*

AmpR: resistencia a la ampicilina

ARN: ácido ribonucleico

ARNm: ácido ribonucleico mensajero, *messenger ribonucleic acid (mRNA)*

AR: autosómico recesivo

AR-HetComp: autosómico recesivo en heterocigosis compuesta

ASTN1: astrotactina 1, *astrotactin 1*

ATAD3A: miembro 3A de la familia ATPasa que contiene dominios AAA, *ATPase family AAA domain containing 3A*

BCA: ácido bicineconínico, *bicinchoninic acid*

BSA: albúmina de suero bovino, *bovine serum albumin*

CACNA1A: subunidad alfa 1A del canal de calcio regulado por voltaje, *calcium voltage-gated channel subunit alpha1 A*

CACNA1G: subunidad alfa 1G del canal de calcio regulado por voltaje, *calcium voltage-gated channel subunit alpha1 G*

CBs: cuerpos de Cajal, *Cajal bodies*

CES: secuenciación de exoma clínico, *clinical exome sequencing*

CIBERER: Centro de Investigación Biomédica en Red de Enfermedades Raras

CMAs: microarrays cromosómicos, *chromosomal microarrays*

CNAG: Centro Nacional de Análisis Genómico

CNV: Variante en el número de copias, *Copy Number Variation*

Codón de STOP: codón de parada

CROCC: *ciliary rootlet coiled-coil (rootletin)*

DGV: Base de datos de variantes genómicas, *Database of Genomic Variations*

DMEM: *Dulbecco's Modified Eagle's Medium*

DMSO: dimetilsulfóxido

dNTP: desoxirribonucleósido trifosfato

EDTA: ácido etilendiaminotetraacético

EERR: enfermedades raras

FBS: suero bovino fetal, *fetal bovine serum*

FEDER: Federación Española de Enfermedades Raras

GAPDH: gliceraldehído-3-fosfato deshidrogenasa, *glyceraldehyde-3-phosphate dehydrogenase*

gnomAD: consorcio de agregación de genomas, *Genome Aggregation Consortium*

GRCh37/GRCh38: versión 37 o versión 38 del genoma de referencia humano

HGMD: base de datos de mutaciones de genes humanos, *Human Gene Mutation Database*

HRP: peroxidasa de rábano, *horseradish peroxidase*

HS: suero de caballo, *horse serum*

IFI: inmunofluorescencia

Indel: pequeñas inserciones y deleciones

ISCI: Instituto de Salud Carlos III

LB: medio Luria-Bertani

MAF: frecuencia del alelo menor, *minor allele frequency*

MD: trastorno del movimiento, *movement disorder*

MDS: sociedad del trastorno del movimiento, *Movement Disorder Society*

Membrana PVDF: membrana de fluoruro de polivinilideno, *polyvinylidene fluoride membrane*

MCBs: cuerpos de Cajal mitóticos, *mitotic Cajal bodies*

MVBs: cuerpos multivesiculares, *multivesicular bodies*

NCBI: Centro nacional para la información en biotecnología, *National Center of Biotechnology Information*

NGS: tecnologías de secuenciación de nueva generación, *Next Generation Sequencing*

OMIM: Herencia Mendeliana en el Hombre, *Online Mendelian Inheritance in Man*

OMS: Organización Mundial de la Salud

pb: pares de bases

PBS: tampón fosfato salino, *phosphate buffered saline*

PCR: reacción en cadena de la polimerasa, *polymerase chain reaction*

Promotor CMV: promotor del citomegalovirus

PTPN23: proteína tirosina fosfatasa tipo 23 no receptor, *protein tyrosine phosphatase non-receptor type 23*

SLITRK3: miembro 3 de la familia SLIT y NTRK, *SLIT and NTRK like family member 3*

SMN1: proteína de supervivencia motoneuronas 1, *survival of motor neuron 1*

SLC38A6: miembro 6 de la familia de transportadores de solutos 38, *solute carrier family 38 member 6*

RIPA: ensayo de radioinmunoprecipitación, *Radio Immunoprecipitation Assay*

SDS-PAGE: electroforesis en gel de poliacrilamida con dodecil sulfato sódico, *sodium dodecyl sulfate polyacrylamide gel electrophoresis*

SNV: variante de un único nucleótido, *Single Nucleotide Variant*

SV: variación estructural, *Structural Variation*

Tm: temperatura de hibridación, *melting temperature*

TAF4: factor asociado 4 a la proteína que se une a la secuencia consenso TATA de la región promotora, *TATA-box binding protein associated factor 4*

UCSC Genome Browser: visor de genomas de la Universidad de California, Santa Cruz; *University of California, Santa Cruz Genome Browser*

UsnRNPs: pequeñas proteínas nucleares ricas en uridina, *uridine-rich small nuclear ribonucleoproteins*

UTR: región no traducida, *untranslated region*

VUS: variante de significado incierto, *Variant of Unknown Significance*

WES: secuenciación de exoma completo, *whole exome sequencing*

WGS: secuenciación de genoma completo, *whole genome sequencing*

WT: tipo silvestre, *wild type*

RELACIÓN DEL TRABAJO CON LOS ODS (OBJETIVOS DE DESARROLLO SOSTENIBLE) AGENDA 2030

OBJETIVO DE DESARROLLO SOSTENIBLE	ALTO	MEDIO	BAJO	NO PROCEDE
ODS 1. Fin de la pobreza				X
ODS2. Hambre cero				X
ODS3. Salud y bienestar	X			
ODS4. Educación de calidad				X
ODS5. Igualdad de género				X
ODS6. Agua limpia y saneamiento				X
ODS7. Energía asequible y no contaminante				X
ODS8. Trabajo decente y crecimiento económico				X
ODS9. Industria, innovación e infraestructuras				X
ODS10. Reducción de las desigualdades	X			
ODS11. Ciudades y comunidades sostenibles				X
ODS12. Producción y consumo responsables				X
ODS13. Acción por el clima				X
ODS14. Vida submarina				X
ODS15. Vida de ecosistemas terrestres				X
ODS16. Paz, justicias e instituciones sólidas				X
ODS17. Alianzas para lograr objetivos				X

Como se menciona a lo largo del presente Trabajo Fin de Grado, aunque las enfermedades raras (EERR) individualmente afectan a pocos pacientes, en conjunto, más de 470 millones de personas a nivel mundial sufren alguna de ellas. La mayoría de pacientes con EERR tienen problemas para ser diagnosticados y tratados, debido a la falta de facultativos con experiencia en su enfermedad, de pruebas diagnósticas concretas o de tratamientos específicos, entre otros factores. Además, en muchas ocasiones, no se conoce la causa última de la enfermedad o, a pesar de conocerse, las empresas farmacéuticas no invierten en el descubrimiento y desarrollo de fármacos para tratarlas, por no ser rentables económicamente. En este sentido, la investigación básica y aplicada en EERR es crucial para ofrecer un cuidado global adecuado de los pacientes: tanto para el descubrimiento de qué variantes y en qué genes causan la enfermedad, el mecanismo molecular o patogénico, como para el desarrollo de test de diagnóstico y tratamientos.

Por los motivos expuestos anteriormente y desarrollados más adelante, este trabajo presenta un alto grado de relación con dos de los Objetivos de Desarrollo Sostenible (ODS), en concreto, se trata de los ODS número 3 y número 10.

En primer lugar, el ODS 3, “Salud y bienestar”, se centra en mejorar la salud de las personas en todos sus aspectos y garantizar el acceso igualitario a los recursos sanitarios, vacunas y tratamientos. En esta línea, el presente proyecto se basa en la investigación de las bases genéticas y moleculares de la enfermedad rara que presentan varios pacientes con el objetivo de proveerles con un diagnóstico genético preciso y ofrecerles la mejor opción terapéutica disponible. Más concretamente, se alinearía con la meta 3.8: “Lograr la cobertura sanitaria universal [...], el acceso a servicios de salud esenciales de calidad y el acceso a medicamentos y vacunas seguros, eficaces, asequibles y de calidad para todos”, lo cual incluye también a pacientes de enfermedades raras.

Por otra parte, en cuanto al ODS 10, “Reducción de las desigualdades” la relación con el proyecto es clara: la investigación de enfermedades raras contribuye a dar luz en cuanto al diagnóstico y tratamiento de una minoría (los pacientes de enfermedades raras y sus familias) que ha sido abandonada por la industria farmacéutica, por no resultar rentable el desarrollo de fármacos destinados a unos pocos pacientes. La odisea temporal que los pacientes pasan hasta

encontrar, con suerte, un diagnóstico debe reducirse mediante el empleo de técnicas de última generación, como las NGS, entre ellas el WES (*Whole Exome Sequencing*) utilizado en el presente trabajo. Así, con un correcto diagnóstico, se pueden barajar todas las opciones terapéuticas hasta dar con la adecuada. Asimismo, el hecho de dar nombre a la enfermedad ayuda a los pacientes a entender la patología, y es fundamental para dar visibilidad a realidades que existen, aunque pocos conoces. En este sentido, las asociaciones de pacientes juegan un papel clave, y desarrollan una actividad extraordinaria visualizando cada enfermedad, informando a los pacientes y familiares, y reuniendo fondos para la investigación. Por todo ello, el trabajo cumpliría con la meta 10.2 “[...] potenciar y promover la inclusión social [...] de todas las personas, independientemente de su edad, sexo, discapacidad, raza, etnia, origen, religión o situación económica u otra condición”.

1. INTRODUCCIÓN

1.1. ¿QUÉ SON LAS ENFERMEDADES RARAS?

Las enfermedades raras (EERR) son condiciones que afectan a la salud humana con una prevalencia baja (< 1 afectado por cada 2.000 personas, según la UE) en comparación con las enfermedades comunes. Se trata, normalmente, de enfermedades genéticas y, por tanto, crónicas, muchas de ellas pediátricas e incluso fatales (Abozaid et al., 2022). En conjunto, se estima que un 6% de la población mundial sufre alguna enfermedad rara, lo que equivale a más de 470 millones de personas afectadas (Abozaid et al., 2022; Haendel et al., 2020; Nguengang Wakap et al., 2020).

No obstante, no existe un consenso acerca de la definición de las enfermedades raras, sino que la entidad reguladora en materia de salud de cada país establece su propia definición. Así, por ejemplo, la Ley de Medicamentos Huérfanos en Estados Unidos (1983) considera que una enfermedad es rara cuando afecta a menos de 200.000 personas, mientras que la legislación análoga en la Unión Europea (UE), del año 2000, las define como aquellas enfermedades que afectan a menos de 1 persona de cada 2.000. El criterio empleado en Japón asume que las EERR afectan a 40 entre 100.000 personas, y, en el caso de Australia, se fija la cifra en 11 afectados por cada 100.000 personas. Y así un largo etcétera que lleva a que una enfermedad sea considerada rara o no en función de la localización geográfica del paciente, lo cual supone un problema en cuanto a su diagnóstico y tratamiento, pero también para la investigación sobre los mecanismos de la enfermedad y potenciales terapias. En un intento por unificar criterios y facilitar el intercambio de información y el desarrollo de medicamentos huérfanos, la OMS (Organización Mundial de la Salud) definió las EERR como aquellas con una prevalencia de 65 de cada 100.000 individuos (Abozaid et al., 2022; Lopes-Júnior et al., 2022).

Se calcula que existen entre 5.000 y 8.000 EERR (Haendel et al., 2020), aunque otros autores hablan de un número mucho mayor, cifrándolo en 10.000 o superior (Smith et al., 2022; Tesi et al., 2023). De nuevo, la falta de estandarización provoca la falta de consenso en los números. De las bases de datos existentes que recogen información sobre EERR, Orphanet (<https://www.orpha.net>) es la única centrada exclusivamente en enfermedades raras, y cuenta con un total de 6.313 EERR registradas (último acceso: 10/06/2024). Se trata de una plataforma que surgió en Francia y se ha extendido hasta formar un consorcio de 40 países, que recoge información revisada por expertos sobre el diagnóstico, cuidado, tratamiento, bases genético-moleculares, investigación y ensayos clínicos de las EERR. Igual de relevante es la base de datos OMIM (*Online Mendelian Inheritance in Man*), catálogo de enfermedades mendelianas que actualmente refiere 6.865 EERR en las que se conoce el cuadro clínico y el gen implicado (último acceso: 16/06/24). Si bien, no es una base de datos exclusiva de EERR, la mayoría de éstas son mendelianas y por tanto, la mayoría de condiciones incluidas en OMIM son EERR. OMIM es una plataforma creada por el Dr. Victor McKusick en los años 60, disponible en la web desde 1995 por iniciativa de NCBI (*National Center for Biotechnology Information*) estadounidense. Conjuntamente, las EERR incluyen un amplio espectro de desórdenes con diferentes mecanismos patogénicos. De ellas, más del 80% son monogénicas (Lee et al., 2020; Chung et al., 2021). Al contrario que las enfermedades comunes, causadas por numerosas variantes genéticas con un pequeño efecto sobre el fenotipo, e influenciadas por factores ambientales, las EERR se deben a una variante genética que conduce a una enfermedad (Haendel et al., 2020). A pesar de los avances en los últimos años en el conocimiento de las EERR y el desarrollo de las tecnologías en investigación, su diagnóstico sigue siendo un reto. La mayoría de pacientes se embarcan en la denominada “odisea diagnóstica”, pasando por años de visitas médicas, diferentes especialistas, con múltiples pruebas, hasta que, con suerte, logran ser diagnosticados. Este retraso en el diagnóstico (o incluso diagnóstico erróneo) aumenta notablemente el sufrimiento de los/las pacientes y sus familias, a la par que genera elevados costes en el sistema sanitario, llegando incluso

a empeorar los signos clínicos de la patología por la falta de tratamiento temprano (Benito-Lozano et al., 2022). En este sentido, el Consorcio Internacional de Investigación en Enfermedades Raras (IRDiRC, por sus siglas en inglés) tiene como meta para 2027 diagnosticar todas las EERR conocidas en un máximo de un año desde la primera consulta médica. Para ello, las tecnologías de secuenciación masiva se posicionan como herramienta indispensable, al ser capaces de disminuir el tiempo de diagnóstico, proporcionando gran cantidad de información en una sola prueba (Tesi et al., 2023).

Por otra parte, aunque se consiga un diagnóstico genético, en muchas ocasiones no existen terapias efectivas para la enfermedad, sino que se administra un tratamiento paliativo de los síntomas. En este sentido, dado que cada enfermedad rara tiene un número muy pequeño de pacientes, se han convertido en “huérfanas” de medicamentos, pues no resulta rentable para la industria farmacéutica ni la investigación ni el desarrollo de fármacos. De hecho, la prevalencia de las EERR fue establecida por las empresas farmacéuticas, como límite en que no es rentable fabricar fármacos. Así, surge el término *orphan drugs* o medicamentos huérfanos, para referirse a aquellos fármacos destinados a tratar o prevenir enfermedades raras (Abozaid et al., 2022). Estos suelen desarrollarse de la mano de iniciativas públicas, o mediante incentivos a las empresas farmacéuticas; y además suelen tener un precio elevado, lo que dificulta aún más el acceso a ellos (Lopes-Júnior et al., 2022).

En España, se calcula que aproximadamente 3 millones de personas sufren alguna EERR (Riera-Mestre, 2022). Aquí, encontramos organismos dedicados a la investigación (básica y traslacional) en enfermedades raras como el Instituto de Investigación en Enfermedades Raras (IIER), así como el CIBERER (Centro de Investigación Biomédica En Red de Enfermedades Raras), ambas iniciativas del Instituto de Salud Carlos III (ISCIII). El CIBERER es un consorcio público nacional creado en 2006 y formado por 75 grupos de investigación especializados en aspectos clínicos, genéticos, moleculares, bioquímicos y celulares de las EERR (Luque et al., 2022). Además de su colaboración con asociaciones de pacientes como FEDER (Federación Española de Enfermedades Raras), el CIBERER ha promovido la aprobación de 12 medicamentos huérfanos en los últimos 10 años para tratar enfermedades como la anemia de Fanconi, la deficiencia en piruvato quinasa o la cistinuria, entre otras (Luque et al., 2022). Sin embargo, España sigue siendo el único país de Europa sin especialidad médica en genética. Aunque se están dando pasos en dicha dirección, la falta de normativa al respecto retrasa su materialización, repercutiendo finalmente en los pacientes.

1.1.1. Trastornos del movimiento

Los trastornos del movimiento comprenden un grupo de alteraciones neurológicas que causan movimientos anormales, voluntarios o involuntarios (Rabbani et al., 2014), afectando a la capacidad de producir y controlar el movimiento debido a la disfunción de los ganglios basales o de estructuras que conectan con estos (Montaut et al., 2018). Se pueden clasificar en dos grupos: hipercinéticos (con exceso de movimiento, donde se incluyen tics, corea distonía, temblor y mioclonías) e hipocinéticos (con lentitud y reducción del movimiento, donde destacan los síndromes parkinsonianos) (Avilés-Olmos et al., 2023). Se trata de un conjunto de enfermedades de difícil diagnóstico debido a la complejidad de las manifestaciones clínicas, frecuentemente variables (Abdo et al., 2010). Algunos trastornos del movimiento son enfermedades raras (Jinnah et al., 2018), donde se incluyen, según la MDS (*Movement Disorder Society*), fenotipos de ataxia espinocerebelosa, corea, distonía, parkinsonismo, paraplejía espástica hereditaria y trastornos del movimiento paroxísticos.

Los **tics** son una categoría de movimientos repentinos y espasmódicos con movimientos recurrentes estereotipados (Abdo et al., 2010). Generalmente, se producen en la cara, brazos y cuello, y se dividen en simples (por ejemplo, guiño de ojos) y complejos (tocar u oler objetos, saltos).

La **corea** se define como un conjunto de movimientos involuntarios abruptos, impredecibles y no rítmicos, resultantes de un flujo continuo de contracciones musculares aleatorias (Thenganatt & Jankovic, 2019). A diferencia de la mioclonía, el patrón de movimientos cambia de una parte del cuerpo a otra, simulando una danza, afectando a las extremidades, tronco y cara. Sus causas son diversas, aunque tienen en común la disfunción del circuito motor de los ganglios basales (Avilés-Olmos et al., 2023). La **distonía** comprende posturas y movimientos anómalos producidos por contracción muscular sostenida que afecta a las extremidades, tronco, cara o boca. Suele ir acompañada de movimientos rítmicos, irregulares o paroxísticos. En ocasiones, puede presentarse junto con corea, constituyendo los denominados movimientos coreodistónicos móviles.

Por otra parte, el **temblor** es una oscilación rítmica involuntaria y sinusoidal de una parte del cuerpo, como puede ser la mandíbula, barbilla, cabeza, piernas o extremidades superiores, siendo estas últimas las principales afectadas. Está causado por la activación consecutiva de músculos agonistas y antagonistas, provocando una actividad oscilatoria a una frecuencia determinada (Avilés-Olmos et al., 2023). La **mioclonía** consiste en un conjunto de movimientos involuntarios, espontáneos, cortos y en forma de shock, que generalmente son positivos (causados por contracción muscular) pero también pueden ser negativos (debido a la pérdida o inhibición del tono muscular) (Abdo et al., 2010). En definitiva, se trata de sacudidas musculares bruscas, repentinas y de breve duración (<250 milisegundos) que ocurren tras una descarga neuronal (Avilés-Olmos et al., 2023).

Por último, el **parkinsonismo** se caracteriza por rigidez, bradiquinesia o lentitud en los movimientos y temblor en reposo (Di Fonzo et al., 2018). Además, la fatiga es un elemento común en los pacientes. La lentitud de movimiento se reconoce mediante el análisis de la disfunción cerebelar y piramidal (Abdo et al., 2010).

1.2. USO DE TECNOLOGÍAS NGS PARA EL DIAGNÓSTICO DE ENFERMEDADES RARAS

Los primeros diagnósticos de EERR se produjeron a principios del siglo XX (Garrod, 1902) (Vinkštel et al., 2021). Más adelante, en los años 60, las técnicas citogenéticas como el cariotipado eran el único método de detección de anomalías cromosómicas como aneuploidías (menor número de cromosomas) o grandes deleciones/duplicaciones (Ramos-Fuentes et al., 2020). Años más tarde, se desarrollaron los *microarrays* cromosómicos (CMAs) o cariotipado molecular, basado en el uso de fluorescencia para contar el número de alelos en las regiones cromosómicas, con mayor sensibilidad para detectar pequeñas variaciones en el número de copias (CNVs) (Schuler et al., 2022). El diagnóstico de una enfermedad debe empezar con el diagnóstico clínico, en base al cual se puede orientar el análisis genético. La secuenciación Sanger, que permite el análisis genético de genes exón a exón, ha sido una herramienta fundamental en las unidades de genética y es hoy en día, la herramienta *gold standard* para confirmación de variantes genéticas. El surgimiento de las aproximaciones NGS (*Next Generation Sequencing*) con paneles de genes, CES (*Clinical Exome Sequencing*) y WES (*Whole Exome Sequencing*), principalmente consisten en el análisis de cientos a miles de genes, llegando a los 22.000 genes estimados en el genoma humano, y ha supuesto un gran avance en el diagnóstico de las EERR. Así, tras la finalización del Proyecto Genoma Humano en 2003, las tecnologías de secuenciación masiva (NGS, por sus siglas en inglés) vieron su auge, abaratando cada vez más su coste, y mejorando su capacidad de trabajo (Straiton et al., 2019). Esta tecnología lee simultáneamente millones de secuencias cortas de ADN, permitiendo secuenciar diferentes genes en una misma reacción (Ramos-Fuentes et al., 2020). Esto las convierte en la herramienta perfecta para el diagnóstico de EERR, caracterizadas por la heterogeneidad genética.

Existen diferentes posibilidades a la hora de emplear NGS en el diagnóstico: desde paneles de genes, hasta secuenciación de exoma (WES) o de genoma completos (WGS) (Fernández-Marmiesse et al., 2018; Tesi et al., 2023). Los **paneles de genes** son estrategias dirigidas basadas en

la secuenciación de un grupo de genes que se sabe que albergan mutaciones que contribuyen a la patogénesis de ciertas enfermedades. Por tanto, solo se detectan variantes en genes seleccionados, generando un menor volumen de datos, de manera que la interpretación es más sencilla y se encuentran muchas menos variantes de significado incierto (VUS) que con el resto de tecnologías. Se requieren actualizaciones periódicas en los genes analizados y, al igual que la secuenciación de exoma completo, tiene limitaciones para detectar variantes estructurales y elementos repetitivos (Vinkštel et al., 2021). Por su parte, la **secuenciación de exoma completo (WES)** puede tener aproximaciones dirigidas o no dirigidas. La estrategia dirigida es la secuenciación exclusiva de aquellos genes implicados en enfermedad humana: CES (*Clinical Exome Sequencing*). La selección de estos genes se hace en base a la información disponible en OMIM habitualmente si bien, algunos CES comerciales utilizan la base de datos HGMD (*Human Gene Mutation Data Base*). La estrategia no dirigida consiste en la secuenciación de todo el exoma, es decir, de todos los exones codificantes, aproximadamente el 2% del genoma (Gonzaga-Jauregui, 2021). De esta forma, se pueden identificar variantes en genes no asociados previamente con enfermedad humana. Como toda tecnología, presenta limitaciones como la cobertura incompleta de ciertas regiones, así como la limitada detección de algunos tipos de variantes, mencionadas anteriormente. Por último, la **secuenciación de genoma completo (WGS)** abarca el genoma humano completo, superando las limitaciones del resto de tecnologías en cuanto a la secuenciación de regiones no codificantes (Vinkštel et al., 2021). Sin embargo, a pesar de su potencial por la mayor cantidad de datos generada, el aún limitado conocimiento de la patogenicidad de las variantes en regiones no codificantes ha retrasado su aplicación en clínica, aplicándose únicamente en casos concretos (Gonzaga-Jauregui & Zepeda, 2021, Vinkštel et al., 2021). Además, su elevado coste y baja disponibilidad de equipos magnifican sus desventajas (Vinkštel et al., 2021).

La tasa de diagnóstico de EERR empleando NGS varía entre un 25% y un 52% en función del tipo de enfermedad, la estrategia analítica y el diseño del estudio (probando versus trío) (Liu et al., 2019). En este sentido, el análisis del trío se refiere al estudio simultáneo del/de la probando y sus progenitores, teniendo mayor sensibilidad para detectar variantes que surgen *de novo* en el/la probando, y también para obtener variantes bialélicas con diferentes tipos de herencia (Liu et al., 2019; Tan et al., 2019). Así, las probabilidades de diagnóstico son el doble que con el análisis único del probando (Marwaha et al., 2022). Esta mayor eficiencia se debe a la reducción de entre 2 y 10 órdenes de magnitud en el número de variantes candidatas debido al uso de la información parental, asumiendo que los progenitores son sanos y, por tanto, los progenitores no son portadores de la variante causal en una herencia dominante o son portadores heterocigotos en una herencia recesiva (Tan et al., 2019; Wright et al., 2015). En el caso de que los progenitores estén afectados por la enfermedad, también resulta interesante incluirlos en el estudio para así buscar la variante causal que todos ellos comparten.

1.2.1. Secuenciación de exoma completo (WES)

La primera publicación que emplea la secuenciación de exoma para la identificación de una enfermedad mendeliana rara de causa desconocida se remonta a 2009 (Ng et al., 2009). El estudio consistió en cuatro pacientes de tres familias distintas, y en él se identificó un único gen candidato (*DHODH*) que resultó ser causante de la enfermedad que presentaban, el síndrome de Miller.

Como se ha mencionado, el WES se fundamenta en la secuenciación del exoma, es decir, de las regiones codificantes del genoma, que representa el 1-2% en humanos (30-65 Mb) (Veeramachaneni, 2020). Se estima que el exoma alberga el 85% de las variantes patogénicas en el ADN (Rabbani et al., 2014) y constituye una técnica más económica, con un análisis de datos más sencillo y, por tanto, más rápida que la secuenciación de genoma completo (Seaby et al., 2016). Al secuenciar los aproximadamente 180.000 exones del genoma, se pueden descubrir nuevas variantes y genes asociados a enfermedad. En comparación con WGS, WES presenta ciertos

inconvenientes, como la falta de uniformidad en la cobertura (sobre todo en los primeros exones o en regiones con alto contenido en GC/AT), de manera que se pierde información en dichas regiones. Además, se trata de una tecnología poco sensible a variantes estructurales (CNVs, inversiones, grandes inserciones/deleciones), así como a variantes en regiones reguladoras o intrónicas (Souche et al., 2022). Estas limitaciones son superadas por WGS, que cuenta con una tasa de falsos positivos en variantes de un único nucleótido (SNVs) de un 17%, frente al 78% del WES (Belkadi et al., 2016), lo que la convierte en una tecnología muy atractiva, aunque menos aplicada debido a sus elevados costes y tiempo requerido para el análisis de datos, así como por la dificultad en la interpretación de las variantes en regiones no codificantes ya que, aun secuenciado todo el genoma, se secuencia el mismo número de genes que en un WES.

A grandes rasgos, el proceso o *pipeline* realizado en WES se refleja en la **Figura 1**. En primer lugar, el ADN puede extraerse de diferentes tejidos, siendo el más común la sangre periférica. Un punto clave en este paso es el rendimiento y la calidad del ADN, muy dependientes de la técnica empleada. A continuación, la preparación de librerías consiste en la fragmentación del ADN genómico en fracciones de 200-400 pares de bases (pb), y la posterior ligación de adaptadores en sus extremos. Los adaptadores son oligonucleótidos que contienen secuencias para la unión de los cebadores (necesarios para la secuenciación), así como para la unión a la *flowcell* (soporte donde se amplificarán las muestras a analizar), y también identificadores únicos para cada muestra, conocidos como *barcodes*, que permiten secuenciar varias librerías diferentes en la misma *flowcell* (Gunter et al., 2022). Después, para la secuenciación de exoma, se deben enriquecer y capturar los exones mediante hibridación con sondas complementarias biotiniladas, posteriormente recogidas con bolas magnéticas conjugadas con estreptavidina. Así, cualquier fragmento no comprendido por las sondas será lavado y no capturado. En ocasiones, se realiza una amplificación de la librería antes de la secuenciación, siendo Illumina la plataforma más comúnmente utilizada. Esta amplificación se realiza en soporte sólido, donde se encuentran covalentemente unidos los cebadores complementarios a la región del ADN a amplificar. Así, se produce una amplificación por PCR puente, generando grupos de secuencias clonalmente amplificadas unidas al soporte. Tras la amplificación, tiene lugar la secuenciación por síntesis (SBS, *sequencing by synthesis*) empleando dideoxinucleótidos trifosfato (ddNTPs) marcados fluorescentemente y modificados químicamente. Por poner un ejemplo, una de las químicas empleadas en la SBS consiste en que en cada ciclo se añade un ddNTP complementario al ADN molde, se detecta la señal fluorescente del fluoróforo y se elimina el grupo terminador y el fluoróforo, regenerando el extremo 3'OH que permite continuar con la secuenciación. Por último, el análisis de datos consiste en la obtención de lecturas, su mapeado y alineamiento frente al genoma de referencia, la identificación de variantes, su anotación e interpretación (Gonzaga-Jauregui & Zepeda, 2021).

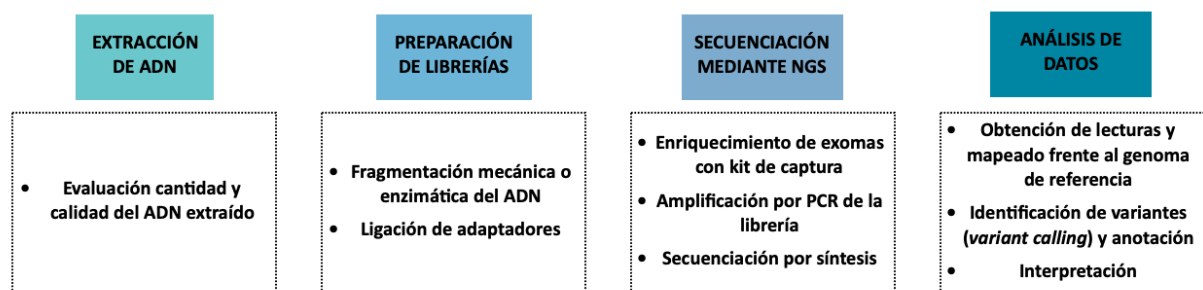


Figura 1. Esquema del *pipeline* seguido en la secuenciación de exoma completo (WES).

En caso de no encontrar variantes candidatas en un primer análisis de los datos obtenidos por WES, se recomienda reevaluar los datos años más tarde en el denominado WES-dinámico. Ello permite aumentar el rendimiento diagnóstico entre un 5 y un 26% (Marwaha et al., 2022), gracias a la incorporación de nuevas variantes patológicas en diferentes bases de datos, a la obtención de información sobre nuevos genes involucrados en enfermedad, y a la mejor anotación del genoma.

1.2.2. Filtrado y priorización de variantes

Antes de la interpretación, es necesario el filtrado de las variantes para reducir la cantidad de datos a analizar sin perder información relevante. Uno de los primeros filtros aplicados es la frecuencia de la variante en la población general, asumiendo que una enfermedad rara con herencia dominante estará causada por una variante que nadie o muy pocas personas presentarán; en el caso de herencias recesivas, sí es posible que la variante causal tenga una frecuencia más alta, incluso superior al 1% en población control (Gonzaga-Jauregui & Zepeda, 2021). Por tanto, esta hipótesis debe revisarse en cada caso, dependiendo del tipo de enfermedad, su gravedad, y la posibilidad de penetrancia incompleta. Si se dispone de la información del probando y sus progenitores (trío) se puede filtrar por diferentes modelos de herencia, incluyendo variantes *de novo* y en heterocigosis compuesta, que consiste en la presencia de dos variantes en el mismo gen, una en cada alelo (*trans*), procedentes de cada uno de los progenitores (Umlai et al., 2022). Otra opción es filtrar por el efecto funcional de la variante en la proteína. Por una parte, las variantes de alto impacto (*high impact*) se relacionan con la alteración de la proteína, al crear un codón de parada prematuro (variantes *nonsense*) o insertar/delecionar bases que cambian la secuencia (variantes indel *frameshit*) y producen un truncamiento temprano de la proteína. En cuanto a variantes de cambio de aminoácido (*missense*), se deben evaluar en función de la conservación evolutiva del residuo y del dominio funcional que se ve afectado. Asimismo, se puede tener en cuenta si el gen se ha asociado previamente a enfermedad, y también información como el patrón de expresión o la función predicha resultan útiles para poder relacionar las variantes encontradas con el fenotipo del probando (Gonzaga-Jauregui & Zepeda, 2021). Bases de datos como gnomAD (*genome aggregation database*), que contiene información de frecuencias de variantes en población sana de diferentes proyectos de secuenciación de exoma y genoma, contribuyen al filtrado de variantes, descartando las más frecuentes en la población general (MAF o *minor allele frequency* > 0,5%).

Una vez seleccionadas las variantes candidatas tras el filtrado, se accede a bases de datos especializadas en genes y variantes asociados a enfermedad para comprobar si la variante caracterizada se ha descrito previamente como patológica. Un ejemplo es OMIM (<https://www.omim.org>) que, como se ha mencionado anteriormente, es un catálogo de enfermedades monogénicas (construida en base a las publicaciones científicas disponibles en PubMed) donde es fácil averiguar si un gen está implicado en el desarrollo de un trastorno o no. También se analiza el cambio en plataformas como Varsome (<https://varsome.com>), que recoge información de varias bases de datos sobre la clasificación de la variante (patogénica, probablemente patogénica, benigna, probablemente benigna o de significado incierto, VUS), entre ellas, ClinVar, que es una base de datos pública curada por expertos que recoge variantes registradas por laboratorios de investigación y clínicos y su relación con el fenotipo humano. Para relacionarlas con el fenotipo del paciente, *Human Protein Atlas* (<https://www.proteinatlas.org>) informa acerca de los tejidos en que se expresa la proteína codificada por el gen de interés (Gonzaga-Jauregui & Zepeda, 2021). Por último, la información proporcionada por predictores *in silico* sobre el efecto de las variantes seleccionadas es de gran utilidad. Por ejemplo, algoritmos bioinformáticos como SIFT (*Sorting Intolerant From Tolerant*) o PolyPhen (*Polymorphism phenotyping*) predicen el efecto funcional del cambio de aminoácido; MutationTaster se centra en el impacto de cambios nucleotídicos (de un nucleótido o pequeños indels); CADD (*Combined annotation-dependent depletion*) evalúa la patogenicidad de variantes genómicas basándose en la información de otros algoritmos (Gonzaga-Jauregui & Zepeda, 2021). También existen predictores del efecto de variantes de *splicing* como SpliceAI o NNSplice (Veeramachaneni, 2020). El proceso de filtrado y priorización de variantes queda resumido en la **Figura 2**.

Así, integrando toda la información recogida, se deben seleccionar las variantes más interesantes para validar su existencia en el probando mediante secuenciación Sanger. De esta forma, se comprueba que el paciente presenta el cambio detectado, y no se trata de un artefacto de la tecnología WES. Finalmente, se realizan estudios funcionales de las variantes validadas y potencialmente relacionadas con el fenotipo del paciente. En función del tipo de variante, los ensayos serán diferentes, consistiendo en el análisis de transcritos o de minigenes para variantes que afecten al *splicing* (Abramowicz & Gos, 2018; Richards et al., 2015) y en estudios de localización subcelular, de actividad enzimática o de expresión de la proteína para variantes *missense* (Sunyaev, 2012).

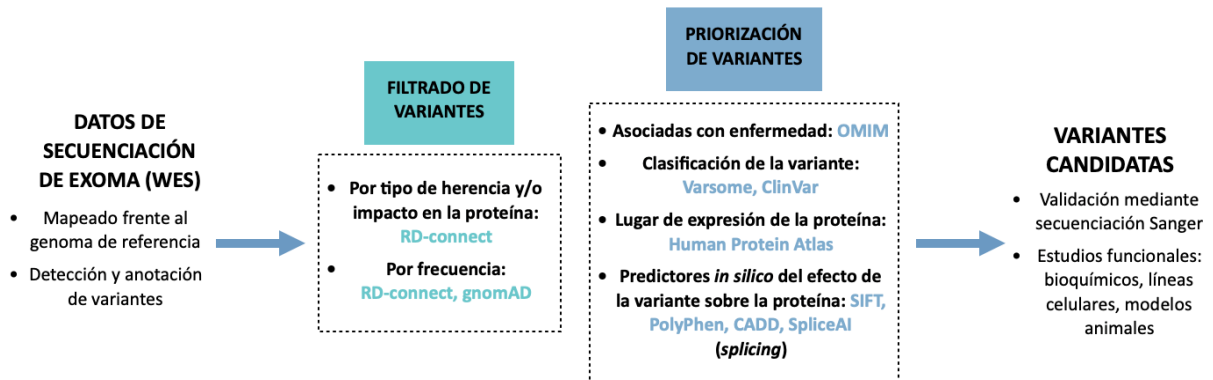


Figura 2. Esquema del proceso de filtrado y priorización de variantes procedentes de secuenciación de exoma para enfermedades raras.

1.3. FAMILIA DE TRANSPORTADORES SLC: *SLC38A6*

La superfamilia de transportadores SLC (*solute carrier transporters*) está formada por un grupo de proteínas de membrana que transportan solutos a través de la membrana celular. Se trata del grupo más numeroso de transportadores codificado por el genoma humano, que incluye más de 450 proteínas divididas en 66 familias (Hushmandi et al., 2024; Xia et al., 2024). En general, son transportadores secundarios que facilitan el transporte pasivo de moléculas, empleando para ello iones (Na^+ , Cl^- , K^+ , H^+) y funcionando como intercambiadores de solutos, tales como aminoácidos, nucleótidos, azúcares, iones inorgánicos y fármacos. En algunos casos, el mismo transportador SLC transporta diferentes sustratos, y en otros, actúan específicamente sobre una molécula. Cabe destacar que suelen existir múltiples isoformas del mismo transportador, que difieren en su extremo N- o C-terminal, con diferentes propiedades de interacción proteína-proteína, eficiencia de transporte o localización. En cuanto a su estructura, esta familia se caracteriza por su heterogeneidad y presenta diferentes mecanismos de transporte. Se localizan en la membrana celular o de orgánulos, y mutaciones en ciertos transportadores SLC se han relacionado con enfermedades neurológicas, metabólicas y cáncer (Hushmandi et al., 2024).

En concreto, la familia *SLC38A_x* transporta aminoácidos pequeños, polares y neutros de manera dependiente del ion sodio (Hushmandi et al., 2024), por lo que también son conocidos como SNATs (*sodium-coupled neutral amino acid transporters*) (Lindberg et al., 2022). La familia comprende 11 miembros y se dividen en dos grupos, según la forma de transporte: sistema A y sistema N. El transporte mediante el sistema A comprende un amplio rango de sustratos y es inhibido por el ácido N-metil aminoisobutírico (MeAIB). El sistema N se refiere al transporte específico de sustratos como glutamina, asparagina e histidina, cuyo mecanismo se basa en el cotransporte de sodio (Na^+) y antiporte de protones (H^+) (Bröer, 2014). La mayoría de SNATs juegan un papel clave en el ciclo de glutamato-glutamina en el cerebro (entre ellos, SNAT6 o *SLC38A6*) (Lindberg et al., 2022), un sistema fundamental en la regulación de la señalización sináptica excitatoria inducida por el glutamato, que evita la neurotoxicidad o excitotoxicidad (Gandasi et al., 2021). También tienen funciones en el hígado, músculo, intestino, riñón y placenta (Bröer, 2014).

SNAT6, codificado por el gen *SLC38A6* (*solute Carrier family member 6*) es el único miembro de la familia SLC38 que se expresa únicamente en las neuronas excitatorias del cerebro (Gandasi et al., 2021). Respecto a su estructura, se predice está compuesto por siete segmentos transmembrana, con una región intracelular N-terminal larga y C-terminal más corta (Gandasi et al., 2021). Aunque se conoce poco sobre su función, se cree que participa en el ciclo glutamato-glutamina junto con otros transportadores de la misma familia. El glutamato, a diferencia de otros neurotransmisores, es captado por un transportador presináptico (EAAT, *excitatory amino acid transporter*). El mecanismo principal de captación y eliminación de glutamato se da en células gliales (astrocitos), mediado por los transportadores de glutamato GLAST y GLT-1. Tras su absorción en las células de la glía, el glutamato se convierte en glutamina por acción de la glutamina sintetasa. La glutamina, que no es excitotóxica, es transportada al espacio extracelular, probablemente a través de SNAT3 y SNAT5. Finalmente, la glutamina es captada por la neurona glutamatérgica presináptica, donde se hipotetiza que SNAT6 desempeña su función (Gandasi et al., 2021), internalizando glutamina que será convertida en glutamato y amoníaco por diferentes enzimas. Aunque no se ha relacionado directamente con enfermedad, la homeostasis aminoacídica es de vital importancia en el sistema nervioso central, así como en todo el organismo (Lindberg et al., 2022), y se cree que la desregulación del ciclo glutamato-glutamina está relacionada con trastornos neurológicos y psiquiátricos (Andersen et al., 2021; Gandasi et al., 2021; Xia et al., 2024).

1.4. PTPN23: FUNCIONES Y RELACIÓN CON SMN1

El gen *PTPN23* codifica para una proteína de 1.636 aminoácidos, PTPN23 (*protein tyrosine phosphatase non-receptor type 23*), una tirosina fosfatasa sin actividad catalítica debido a la presencia de una alanina en una posición que requiere una serina para ejercer la catálisis (Gruss et al., 2017). Por tanto, se hipotetiza que una de sus funciones es la de mantener residuos de tirosina fosforilados en sus proteínas sustrato (Gruss et al., 2017; Smigiel et al., 2018). Se trata de una proteína ubicua con múltiples funciones celulares. Como parte del complejo ESCRT-I (*endosomal sorting complex requered for transport I*), contribuye a la distribución de proteínas ubiquitinadas en cuerpos multivesiculares (MVBs, *multivesicular bodies*). También se ha sugerido que *PTPN23* podría ser un gen supresor de tumores por su posible papel en la regulación negativa de la actividad mitogénica de Ras, a través de su dominio C-terminal catalíticamente inactivo PTP (*protein tyrosine phosphatase*) (Bend et al., 2020). Por último, se ha identificado como regulador del complejo SMN (*survival of motor neuron*), principalmente en su función de ensamblaje de las UsnRNPs (*uridine-rich small nuclear ribonucleoproteins*).

Las pequeñas proteínas nucleares ricas en uridina (UsnRNPs) son complejos formados por partes proteicas y partes ribonucleicas (ARN) que, junto con otros complejos, forman la maquinaria macromolecular o espliceosoma que media el *splicing* del pre-ARNm (Smigiel et al., 2018). El complejo SMN juega un papel clave en el ensamblaje de estas UsnRNPs en el citoplasma, su transporte al núcleo y acumulación en subestructuras nucleares conocidas como cuerpos de Cajal (CBs), donde tiene lugar la maduración de las UsnRNPs. A su vez, SMN interactúa con Coilin, la proteína principal de los CBs que determina el número y la integridad de los mismos (Gruss et al., 2017). Así, en primer lugar, SMN se encontraría en el citoplasma para el ensamblaje de las UsnRNPs, y migraría al núcleo en la etapa de maduración de estas proteínas. Se hipotetiza que SMN volvería al citoplasma tras llevar las UsnRNPs a los cuerpos de Cajal, aunque también tendría otras funciones nucleares, estando regulada su localización por modificaciones post-traduccionales (Gruss et al., 2017). De hecho, la forma citosólica de SMN está hiperfosforilada, en comparación con la nuclear, hipofosforilada. En este contexto, resulta curioso el papel PTPN23, una fosfatasa catalíticamente inactiva que estabiliza los residuos de tirosina fosforilados en SMN, cuya pérdida evita la acumulación de SMN en los cuerpos de Cajal, sugiriendo la necesidad de esta modificación post-traducciona para la localización y función nuclear de SMN (Gruss et al., 2017; Husedzinovic et al., 2015). Además, se demostró que la ausencia de PTPN23 reduce la presencia de SMN en los CBs

(Husedzinovic et al., 2015), y que PTPN23 es una proteína que se localiza mayoritariamente en el citoplasma, aunque también migra al núcleo ocasionalmente. De esta manera, PTPN23 regularía al complejo SMN en ambos compartimentos, o incluso su intercambio núcleo-citoplasma.

En resumen, la función de PTPN23 sería la de regular la localización de SMN a través de la estabilización tirosinas fosforiladas de SMN, permitiendo su transición al núcleo y acumulación en los cuerpos de Cajal, donde maduran las UsnRNPs. Variantes en *PTPN23* se han asociado con un trastorno del neurodesarrollo caracterizado por retraso en el desarrollo y anomalías estructurales en el cerebro (Bend et al., 2020). Algunos autores atribuyen el mecanismo patogénico a la fosforilación y localización anormales de SMN o a la maduración incorrecta de las UsnRNPs (Smigiel et al., 2018); mientras que otros creen que el mismo fenotipo puede estar producido por cualquiera de las diferentes vías en las que participa PTPN23 (Bend et al., 2020). Por ejemplo, se ha asociado la función incorrecta del complejo ESCRT (del que forma parte PTPN23) con la acumulación de proteínas ubiquitinadas en el cerebro, causando degeneración neuronal y microcefalia (Bend et al., 2020). Lo único que queda claro en la literatura es que se necesita mantener funcionalidad residual en PTPN23, ya que variantes bialélicas que truncan la proteína son potencialmente letales en humanos en fase embrionaria, hecho que se ha comprobado en ratones *knockout* (*PTPN23*^{-/-}) homocigotos (Bend et al., 2020; Smigiel et al., 2018).

2. HIPÓTESIS Y OBJETIVOS

La gran heterogeneidad genética y variabilidad fenotípica presentes en las enfermedades raras neurodegenerativas complican significativamente el diagnóstico de los/las pacientes. En los últimos años, el diagnóstico genético ha evolucionado desde el uso de paneles de genes, que comprenden grupos de genes asociados a enfermedad, hasta la secuenciación de exomas clínicos dirigidos a un conjunto de genes que podrían relacionarse con el fenotipo del paciente, gracias al abaratamiento de costes de las tecnologías de secuenciación masiva. Ninguna de las estrategias mencionadas permite el estudio de todos los genes, sino que se centran en el estudio de genes conocidos, previamente asociados a enfermedad. Por ello, postulamos que el uso de la secuenciación de exoma completo (WES) es una mejor aproximación para conseguir un diagnóstico genético definitivo en pacientes con EERR, al poder estudiar todas las regiones codificantes del genoma, incluyendo genes no considerados con otras estrategias.

Por otra parte, actualmente el reto asociado a la secuenciación de exomas consiste en la interpretación de la enorme cantidad de datos generada con esta tecnología. No basta con filtrar y priorizar variantes, sino que se necesitan estudios funcionales para ganar evidencia sobre la posible patogenicidad de los cambios identificados, pues, en muchas ocasiones, se encuentran variantes cuyas consecuencias sobre la salud aún no se conocen. Se trata de ensayos específicos para cada gen o, incluso, para cada variante, que solo pueden ser abarcados por laboratorios de investigación. Por todo esto, también planteamos en el presente trabajo que el correcto diseño de los ensayos funcionales permitirá dilucidar la implicación de los cambios estudiados en el fenotipo de los/las probandos.

Así, los objetivos específicos de este Trabajo de Fin de Grado son:

1. Caracterizar la causa genética de la enfermedad en dos probandos que cursan con trastornos del movimiento mediante el análisis de los datos de secuenciación de exoma.
2. Determinar la patogenicidad de las variantes *SLC38A6* c.690+2T>G realizando ensayo de minigenes y análisis de transcritos.

- Determinar la patogenicidad de las variantes p.V133M/p.R1345H en *PTPN23* realizando estudios de localización subcelular y cuantificación de la proteína.

3. MATERIALES Y MÉTODOS

3.1. PACIENTES

En el presente estudio se analizan 2 familias, con un diseño en trío (probando y los dos progenitores) con el objetivo de encontrar la causa genética hereditaria de la enfermedad que presenta el probando. A continuación, se muestran los pedigrís de las familias estudiadas (**Figura 3**).

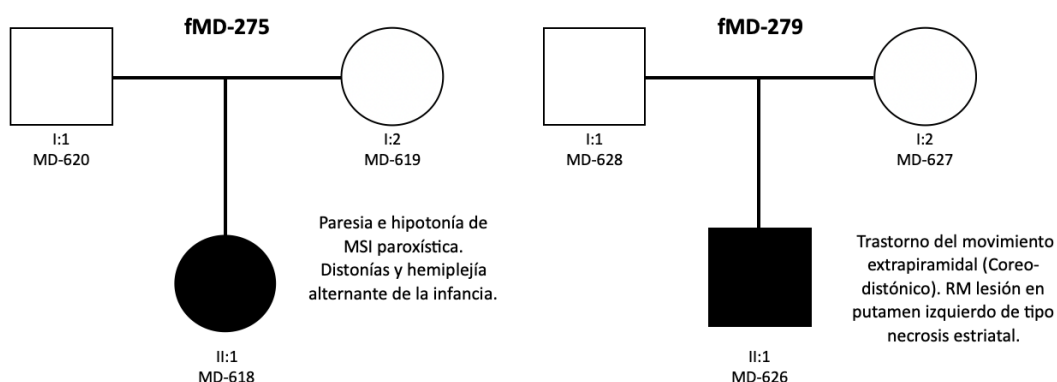


Figura 3. Pedigrís de las familias fMD-275 y fMD-279. En cada caso, se muestran tríos donde se analizan los progenitores y el probando (hijo/a afectado/a), con el fenotipo correspondiente de cada paciente. Cada individuo presenta un identificador asignado en el laboratorio (MD seguido de un número). Las formas en negro representan personas afectadas y las formas en blanco, personas sanas. Los cuadrados indican sexo masculino y los círculos sexo femenino.

En la familia fMD-275, el probando MD-618 presenta un fenotipo clínico de paresia e hipotonía de miembro superior izquierdo (MSI) paroxística, así como distonías y hemiplejía alternante de la infancia. En la familia fMD-279, el probando MD-626 cursa con un cuadro clínico de corea-distonía con discapacidad intelectual leve, observando en resonancia magnética (RM) una lesión en putamen izquierdo de tipo necrosis estriatal.

Por otro lado, estudiamos el caso de una paciente con dos variantes en heterocigosis compuesta en el gen *PTPN23*, que a nivel de proteína se traducen en p.V133M y p.R1345H. La paciente, previamente analizada a nivel genético en el laboratorio, presenta un cuadro clínico caracterizado por ataxia, temblor y calcificaciones en neuroimagen. Se realizaron estudios funcionales con muestras de biopsia cutánea de la paciente para corroborar la patogenicidad de las variantes que presenta.

3.2. ESTUDIO GENÉTICO: SECUENCIACIÓN DE EXOMA

Con el objetivo de identificar variantes causales del fenotipo patológico que presentan los probandos MD-618 y MD-626, se realizó una secuenciación del exoma completo (WES). Para ello, se aísla el ADN genómico procedente de sangre periférica de pacientes y familiares. A continuación, se cuantifica el ADN extraído y se evalúa su calidad mediante espectrofotometría UV/Visible en un equipo Nanodrop ND1000 (Thermo Scientific) para comprobar que se cumplen los estándares de calidad requeridos para la secuenciación (ratio de absorbancias a 260/280 nm entre 1,80-2,00 y ratio de absorbancias a 260/230 > 2,00). Además, se verifica la integridad del ADN genómico

mediante electroforesis en gel de agarosa, como se explica más adelante (**apartado 3.2.2.2**). Posteriormente, se envían las muestras para secuenciación en el CNAG (Centro Nacional de Análisis Genómicos, Barcelona). En esta etapa, se obtienen librerías a partir del ADN genómico, se enriquecen las regiones exónicas mediante un kit de captura (*KAPA HyperExome Probes*, Roche) y, finalmente, se secuencian las muestras con un secuenciador de *Illumina*.

Una vez finalizada la secuenciación, los resultados se envían al laboratorio a través de la plataforma RD-Connect, que dispone de un *pipeline* propio para la identificación de variantes de tipo *SNV* (*Single Nucleotide Variant*) o pequeños *indels*, mediante el mapeo de las lecturas o *reads* de la secuenciación contra la versión 37 del genoma de referencia humano (GRCh37). Asimismo, RD-Connect cuenta con herramientas bioinformáticas para la anotación, el filtrado de variantes y su interpretación (Laurie et al., 2022). El presente Trabajo Fin de Grado comienza con el filtrado y priorización de variantes, por lo que el ADN genómico de los tríos se extrajo y se secuenció con anterioridad.

3.2.1. Filtrado y priorización de variantes

La gran cantidad de datos generados tras la secuenciación es, en primera instancia, filtrada mediante diferentes estrategias para poder analizar las variantes más significativas. Se tienen en cuenta factores como el tipo de herencia, la frecuencia alélica en la población y el impacto de la variante en la proteína. El análisis continúa en Excel, tras descargar ficheros CSV de la plataforma RD-Connect, diferentes según el filtrado aplicado. A grandes rasgos, se consideran variantes con una profundidad de lectura mínima 10x y un valor de calidad ≥ 30 en la escala Phred.

En las dos familias (fMD-275 y fMD-279) se han filtrado cambios producidos *de novo* en los probandos (no presentes en los progenitores) y con distintos tipos de herencia: autosómica dominante (AD), autosómica recesiva (AR) y AR en heterocigosis compuesta (AR-HetComp). Además, en la familia fMD-279, se analizó también la herencia ligada al cromosoma X (recesiva).

Otro de los componentes clave del filtrado es la frecuencia alélica en la población control. En este sentido, únicamente se han considerado aquellas variantes con una MAF (frecuencia del alelo menos común) $< 0,5\%$ en población general en la versión 4.1.0 de gnomAD, en la base de datos 1000Genomes y en la propia base de datos interna de RD-Connect. Además, se priorizan aquellas variantes con una $MAF < 0,01\%$ por tratarse de enfermedades raras monogénicas que, en principio, estarán causadas por variantes muy poco frecuentes en población control, incluso, por variantes noveles. Se ha aplicado este valor de MAF en todos los casos, excepto en aquellos con un alto número de variantes (cambios *de novo*), donde se ha acotado la búsqueda aplicando un valor de $MAF < 0,0001\%$.

Finalmente, se ha analizado el impacto que la variante podría tener sobre la proteína, distinguiendo cuatro efectos: 1) *HIGH*, que incluye variantes con un impacto elevado en la proteína, resultando en su truncamiento, con variantes de tipo *nonsense* (que generan un codón de parada), *frameshift* (de cambio en la pauta de lectura), variantes que generan pérdida de codón de inicio o de parada o que generan nuevos sitios de *splicing*; 2) *MODERATE*, con cambios que reducen la actividad de la proteína, incluyendo variantes *missense* (cambio de aminoácido) o *inframe* (inserciones o deleciones que no modifican la pauta de lectura); 3) *LOW*, donde se agrupan cambios que es probable que no tengan un impacto relevante sobre la proteína, como variantes sinónimas o en sitios de *splicing* relativamente alejados de los exones; y 4) *MODIFIER*, que son variantes que se dan en regiones generalmente no codificantes, por ejemplo, en regiones 5'-UTR o 3'-UTR. Para los cambios *de novo*, en homocigosis y ligados al cromosoma X se estudiaron variantes *HIGH*, *MODERATE*, *LOW* y *MODIFIER IMPACT*, mientras que para heterocigosis compuesta se analizaron variantes *HIGH*, *MODERATE* y *LOW IMPACT*.

El siguiente paso consiste en priorizar los cambios descritos para seleccionar las variantes candidatas a ser causantes de la enfermedad, para lo cual se realizan búsquedas en bases de datos como PubMed, OMIM, *Gene* del NCBI (*National Center for Biotechnology Information*) o *Human Protein Atlas*. Con ello, se obtiene información acerca de la función de la proteína codificada por el gen seleccionado, si se ha asociado previamente con enfermedad, el tipo de herencia y fenotipo que causa, o su localización subcelular, entre otras. Además, para aquellas variantes que podrían afectar al *splicing*, generando o eliminando sitios aceptores o dadores de *splicing*, se emplean predictores *in silico*, como SpliceAI o NNSplice. Estas herramientas predicen la alteración de sitios consenso de reconocimiento de *splicing* en los extremos 5' y 3' del ARNm, del espliceosoma (maquinaria que participa en el proceso de *splicing*), o de elementos activadores/silenciadores del mismo. Por último, se consulta VarSome, un motor de búsqueda que incluye resultados de diferentes bases de datos (ClinVar, gnomAD) o predictores *in silico* de patogenicidad (CADD, SIFT, PolyPhen2), con el propósito de corroborar si la variante está registrada previamente en ClinVar o si es novel; así como para ver la clasificación otorgada por los predictores (*benign*, *likely benign*, *likely pathogenic*, *pathogenic* o *Variant of Unknown Significance* (VUS)).

3.2.2. Validación de cambios candidatos

Las variantes seleccionadas se validaron mediante secuenciación Sanger, previa amplificación por PCR del ADN genómico de los probandos. Además, se analizó la segregación de la variante amplificando y secuenciando el ADN genómico de los progenitores.

3.2.2.1. Diseño de cebadores

La amplificación de las regiones de interés comienza con el diseño de pares de cebadores (uno directo y otro reverso) que flanquean las regiones donde se encuentran los cambios a estudiar. La secuencia del gen a amplificar se consulta en la versión 38 del genoma humano (GRCh38) de la base de datos Ensembl. Para el diseño de cebadores se deben tener en cuenta una serie de características: 1) la longitud de cada cebador debe estar comprendida entre 18 y 25 pb para que sean suficientemente específicos; 2) la temperatura de *annealing* (hibridación con el ADN molde) o T_m debe estar alrededor de 55-65°C, y la diferencia de T_m entre la pareja de cebadores no debe ser superior a 2-4°C; 3) el porcentaje de GCs (guaninas y citosinas) debe estar en el rango de 40-60% para que la hibridación sea correcta; 4) es recomendable que la secuencia final del cebador (extremo 3') acabe en G, C o GC/CG; 5) la región amplificada (amplicón) debe medir 200-800 pb; 6) para que la amplificación se produzca correctamente, los cebadores no deben formar estructuras secundarias ni dímeros; y 7) siempre que sea posible, no es recomendable diseñar los cebadores en regiones con polimorfismos (SNPs) que tengan una elevada frecuencia en la población, pues es probable que se encuentren en la secuencia del ADN y produzcan un desapareamiento entre el cebador y el molde, resultando en una peor amplificación. Para evitar dichas regiones, se puede añadir un filtro en Ensembl que muestre las variantes con MAF <1%, consideradas variantes raras.

Por último, se obtienen los parámetros anteriormente mencionados (T_m, %GC, longitud...) con la herramienta OligoEvaluator de Sigma-Aldrich (<http://oligoevaluator.com/>). Además, se puede predecir la especificidad del par de cebadores *in silico* con *Primer-BLAST* del NCBI (<http://ncbi.nlm.nih.gov/tools/primer-blast>).

3.2.2.2. Amplificación por PCR

En primer lugar, se debe poner a punto la temperatura de hibridación (T_m) del par de cebadores. Para ello, se realiza una PCR con un ADN control y un gradiente de diferentes T_m a testar, partiendo de aquella predicha por OligoEvaluator. Si tras el análisis mediante electroforesis en gel de agarosa no se observa amplificación, se puede realizar otra PCR añadiendo DMSO (dimetilsulfóxido), MgCl₂ adicional o ambos para favorecer la reacción. El DMSO inhibe las

estructuras secundarias en zonas ricas en GC, tanto en el ADN como en los cebadores, aumentando mucho la especificidad de la reacción. El MgCl₂ es un cofactor de la Taq polimerasa que optimiza la reacción de PCR. Los termocicladores empleados para la amplificación fueron VeritiPro™ Thermal Cycler (Applied Biosystems) y GeneAmp™ Thermal Cycler (Applied Biosystems).

Las secuencias de los cebadores diseñados y las temperaturas de hibridación correspondientes se recogen en la **Tabla Suplementaria 1 (Anexo I)**. Las concentraciones de los reactivos para la reacción de PCR, empleadas con carácter general, así como el programa utilizado en el termociclador se encuentran en la **Tabla 1**.

Tabla 1. Reactivos de la PCR Master Mix y programa del termociclador.

Reactivo y concentración inicial	Concentración en la reacción de PCR	Volumen a añadir (V _F = 25 µL)	Etapas del programa	Temperatura	Tiempo	Ciclos
H ₂ O libre de ADNAsas y ARNAsas	-	Completar hasta 25 µL*	Desnaturalización inicial	95°C	5 min	1
Tampón Taq polimerasa 10X (con MgCl ₂)	1X	2,5 µL	Desnaturalización	95°C	30 s	35
dNTPs (20 mM)	0,4 mM	0,5 µL	Annealing o hibridación	T _m (Temperatura de melting)**	30 s	
Cebador directo, 5' (10 µM)	0,2 µM	0,5 µL	Extensión	72°C	Según el amplicón: 1kb/min. 30 o 40 s en general	
Cebador reverso, 3' (10 µM)	0,2 µM	0,5 µL	Extensión final	72°C	7 min	1
Taq ADN polimerasa, Biotools (1 U/µL)	0,02 U/µL	0,5 µL	Conservación	4°C	∞	
ADN molde (~ 100-200 ng/µL)	~ 4-8 ng/ µL	1 µL				

*En función de si la reacción requiere DMSO, MgCl₂, ambos o ninguno, el volumen de agua será diferente. En caso de no necesitar ninguno, 19 µL H₂O; con 5% DMSO (1,25 µL), 18,25 µL H₂O; con 1 µL MgCl₂ 20 mM, 18 µL H₂O; y con ambos, 17,25 µL H₂O. ** La T_m depende de la secuencia del par de cebadores (**Tabla Suplementaria 1**).

Tras la reacción de PCR, se corre un gel de agarosa con el producto amplificado, cargando 5 µL de producto de PCR junto con 1 µL de tampón de carga 6x (*Gel Loading Dye Purple (6X) no SDS*, New England BioLabs), para comprobar la correcta amplificación del fragmento. Se suele realizar un gel de agarosa al 1% (agarosa *Low EEO*, Fisher Scientific) en tampón TAE al 1% (40 mM Tris-HCL, 19 mM ácido acético, 1 mM EDTA, H₂O) con GelRed® (Biotium) como agente intercalante, para visualizar el ADN, el cual se puede añadir directamente a la agarosa o teñir fuera una vez ha corrido el gel. Como patrón de peso molecular se emplea *1 Kb Plus DNA Ladder* (Invitrogen). En general, la electroforesis se corre durante 30 minutos a 100V. Finalmente, el gel se visualiza en un transiluminador con luz ultravioleta que cuenta con una cámara fotográfica (*Molecular Imager® Gel Doc™ XR System*, Bio-Rad).

3.2.2.3. Análisis de las secuencias mediante secuenciación Sanger

Los fragmentos amplificados por PCR se secuencian en el Servicio de Genómica y Genética del CIPF en el secuenciador *3730xl DNA Analyzer* (Applied Biosystems). Para ello, se emplea uno de los dos cebadores (el directo o el reverso) a una concentración de 3,2µM.

Las secuencias, en formato FASTA, se alinean con la herramienta BLASTn (<https://blast.ncbi.nlm.nih.gov/>) contra la secuencia de referencia GRCh38 del amplicón, disponible en Ensembl. Además, se visualiza el electroferograma con el programa *4Peaks* v1.8 para evaluar la calidad de las secuencias e identificar las variantes candidatas previamente seleccionadas.

3.2.3. Análisis de CNVs

Las CNVs (*Copy Number Variations*) son cambios en el número de copias de los genes, debidos a inserciones o deleciones. Los algoritmos de detección de CNVs se basan en la normalización de la cobertura de cada región de interés de la muestra en base a la cobertura total de la misma, y la comparación con la cobertura normalizada de otras muestras procesadas al mismo tiempo, pudiendo establecer puntos de corte para determinar si hay o no una inserción/delección. Entre ellos destacan programas como CoNIFER (*Copy Number Inference From Exome Reads*), muy específico, pero menos sensible a deleciones que afectan a menos de 3 exones; y XHMM (*eXome Hidden Markov Model*), más sensible a fenómenos pequeños, pero menos específico (más falsos positivos). Tras el empleo de estos programas, se obtienen ficheros Excel con los genes que presentan posibles deleciones e inserciones. Estos ficheros se filtran y analizan en base a la función del gen y su posible relación con el fenotipo del paciente, así como por el número de muestras registradas con dicha inserción/delección en la base de datos interna del programa.

El siguiente paso es buscar en los ficheros de SNVs si existe otro cambio en el mismo gen en el que se ha detectado una CNV de interés. Asimismo, se emplea el visualizador *UCSC Genome Browser* para observar la región donde se encuentra la CNV, viendo si la delección o inserción afecta al gen completo o a ciertos exones. También se emplea para determinar si la región es polimórfica en cuanto a alteraciones de dosis génica y, por tanto, es común en la población, y con ello menos probable que la CNV sea patogénica.

3.3. ESTUDIOS FUNCIONALES PARA ESTUDIAR LA PATOGENICIDAD DE LAS VARIANTES SELECCIONADAS

Tras seleccionar las variantes candidatas, se realizan diferentes ensayos experimentales que aporten evidencias de la patogenicidad de las variantes.

3.3.1. Ensayo de minigenes del gen *SLC38A6*

El objetivo de este ensayo es determinar el efecto del cambio *splice donor variant* c.690+2T>G en el gen *SLC38A6* (NM_001172702.2) que el probando MD-626 presenta en homocigosis, estudiando el potencial *splicing* diferencial entre la versión WT (*wild type*) o canónica y mutada (c.690+2T>G) del gen. El estudio experimental consiste en la transfección de células HeLa (una línea celular adherente e inmortalizada de células epiteliales de cáncer de cervical) con un vector (pSPL3) que contiene parte del ADN del gen a analizar, incluyendo el exón 9 (el más cercano al cambio) y parte de los intrones flanqueantes (**Figura 4**).

3.3.1.1. Digestión doble del vector pSPL3 y del inserto

El inserto del gen *SLC38A6* (que comprende el exón 9 + 255 nucleótidos aguas arriba y 237 nucleótidos aguas abajo) se introduce entre los exones constitutivos SD6 y SA2 del vector pSPL3, tal como muestra la **Figura 4**. Previamente, se comprueba qué enzimas de restricción no cortan el inserto con la herramienta *Neb cutter* V2.0 (New England BioLabs) (<https://nc2.neb.com/NEBcutter2/>), siendo XhoI y NheI las enzimas elegidas.

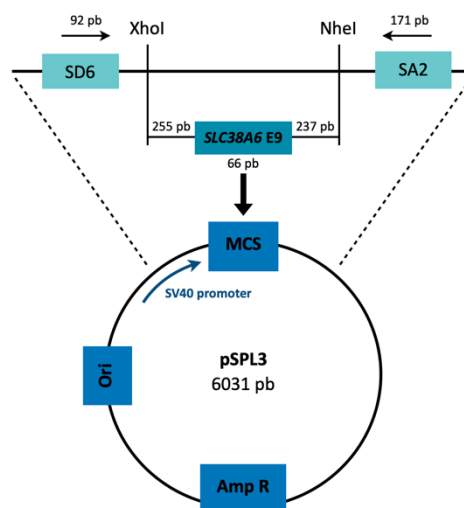


Figura 4. Esquema del vector pSPL3 y del inserto del gen SLC38A6. El vector contiene el origen de replicación bacteriano (Ori) que permite la replicación del plásmido en bacterias, secuencias de resistencia a antibióticos como la ampicilina (Amp R), empleado como agente de selección de bacterias que incorporen el vector. Además, tiene el promotor del virus SV40, que permite la expresión del inserto en células eucariotas de mamífero. Finalmente, el sitio de clonación múltiple (MCS) contiene secuencias de reconocimiento de diferentes enzimas de restricción que permiten cortar el vector, generando extremos cohesivos que dan direccionalidad al inserto. Las flechas negras indican el lugar de diseño de los cebadores para testar la inserción del fragmento, en los exones SD6 y SA2.

Los cebadores diseñados para amplificar el inserto del gen *SLC38A6* (E9) e incorporar las secuencias de restricción de las enzimas para XhoI y NheI se muestran en la **Tabla 2**.

Tabla 2. Cebadores empleados para amplificar el inserto del gen *SLC38A6* para el ensayo de minigenes. En azul se marcan las secuencias de reconocimiento de las enzimas de restricción, y por delante, en minúscula, cuatro nucleótidos que favorecen su corte.

Cebador	Secuencia (5'-3')	Temperatura hibridación	Tamaño amplicón
SLC38A6_E9_D_XhoI	ggaaCTCGAGTTGAAAGTTGACACAGGAG	58°C	558 pb
SLC38A6_E9_R_NheI	gcgcGCTAGCCTGTTTAAAGAGCATTGAGGG		

El inserto *SLC38A6_E9*, procedente del ADN de la madre (heterocigota, por tanto, presenta un alelo *WT* y otro mutante), se **amplifica por PCR** con los reactivos indicados en la **Tabla 1**, aunque se emplea como polimerasa la *Phusion™ High-Fidelity DNA Polymerase* (2 U/μL) (ThermoFisher Scientific), con su tampón correspondiente, para evitar la introducción de mutaciones por errores de la polimerasa. Esta polimerasa de alta fidelidad precisa una temperatura de desnaturalización de 98°C, empleando el siguiente programa de amplificación: desnaturalización inicial (30s a 98°C); 35 ciclos (desnaturalización 10s a 98°C, hibridación 30s a 62°C y extensión 30s a 72°C); y 7 minutos de extensión final a 72°C. A continuación, el producto de PCR se **purifica** con un kit comercial (*Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up XS* de Fisher Scientific). Posteriormente, se realiza la **digestión** del inserto y del vector pSPL3 con XhoI y NheI (ThermoFisher Scientific), con los reactivos mostrados en la **Tabla 3**. La digestión se incubó a 37°C *overnight*.

Una vez finalizada la digestión, tanto el inserto como el vector cortados deben ser **purificados** con el kit *Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up XS* (Fisher Scientific). En el caso del vector, se purifica cortando la banda correspondiente, con un tamaño de 5078pb, de un gel de agarosa al 1%.

Tabla 3. Reactivos para la digestión del inserto SLC38A6_E9 y del vector pSPL3 con XhoI y NheI.

Reactivo y concentración inicial	Concentración en la reacción		Volumen a añadir (V _F = 50 µL)	
	INSERTO	VECTOR	INSERTO	VECTOR
ADN (vector: 394ng/µL)	Amplificado por PCR	3000 ng en 50 µL = 59,97 ng/µL	20 µL	7,61 µL de vector + 12,4 µL H ₂ O = 20 µL
Tampón TANGO (10X)	1X		5 µL	5 µL
XhoI (10 U/µL)	0,8 U/µL		4 µL	4 µL
NheI (10 U/µL)	0,2 U/µL		1 µL	1 µL
H ₂ O libre de ADNasas y ARNasas	-		20 µL	20 µL

3.3.1.2. Ligación

La ligación del inserto *SLC38A6_E9* y el vector pSPL3 digeridos se realiza con un ratio molar 3:1 (inserto:vector, según la **Ecuación 1**), incubando la mezcla a 4°C *overnight*. La **Tabla 4** recoge las concentraciones y volúmenes empleados en la reacción de ligación.

$$ng \text{ inserto} = 50 \times \frac{\text{tamaño del inserto (pb)}}{\text{tamaño del vector (pb)}} \times \frac{3}{1} = 50 \times \frac{558}{5078} \times \frac{3}{1} = 16,48 \text{ ng} \text{ (Ecuación 1)}$$

Tabla 4. Reactivos para la ligación del inserto SLC38A6_E9 y del vector pSPL3 digeridos.

Reactivo y concentración inicial	Concentración en la reacción	Volumen a añadir (µL) (V _F = 20 µL)
pSPL3 digerido (50 ng/µL)	50 ng en 20 µL = 2,5 ng/µL	1 µL
SLC38A6_E9 digerido (34,5 ng/µL)	16,48 ng en 20 µL = 0,82 ng/µL	1 µL del inserto diluido 1:1 en H ₂ O
T4 ligasa (5 U/µL)	0,05 U/µL	0,2 µL
Tampón T4 ligasa (10X)	1X	2 µL
H ₂ O libre de ADNasas y ARNasas	-	15,8 µL

Se obtendrán dos constructos diferentes: uno *wild type* (WT), carente de la mutación (pSPL3-*SLC38A6*^{WT}); y uno mutante (MUT), que presenta la mutación c.690+2T>G (pSPL3-*SLC38A6*^{MUT}).

3.3.1.3. Transformación en Escherichia coli DH5α

Las construcciones obtenidas se introducen en bacterias *Escherichia coli DH5α* mediante transformación por choque térmico. La transformación comienza añadiendo 3 µL de ligación a 40 µL de bacterias competentes, sin pipetear para no romper las bacterias (mezclando a toques), e incubándolas 30 minutos en hielo. Tras esto, se realiza el choque térmico en un termobloque a 42°C durante 40 segundos, y se incuban en hielo 2 minutos. A continuación, las bacterias se incuban en 500 µL de medio LB líquido (Scharlab) sin antibióticos, durante 1h en agitación (190 rpm) a 37°C, para que las bacterias se recuperen y adquieran resistencia a ampicilina. Después, se centrifugan 30 segundos a 6000g y el pellet se resuspende en 200 µL de medio LB. Finalmente, se siembra una placa de LB-agar suplementado con ampicilina (50 µg/mL) para conseguir un crecimiento selectivo de las bacterias que han incorporado el plásmido, con 50-150 µL de bacterias transformadas. La placa se incubaba a 37°C *overnight*.

3.3.1.4. PCR de colonias y extracción del vector en los clones con inserto

Para testar los plásmidos generados, se realizaron reacciones de **PCR de las colonias bacterianas** seleccionadas de la placa. Para conservar la colonia para estudios posteriores, se hacen réplicas en una placa de LB-agar con ampicilina (50 µg/µL) antes de usarla como molde para la PCR. Los cebadores están diseñados entre los exones SD6 y SA2, y el inserto (**Figura Suplementaria 1**). El tamaño esperado en los clones que presentan el fragmento de interés es de 748pb (558pb inserto).

+ 190pb vector), mientras que si no hay inserto porque el vector se ha religado consigo mismo o ha permanecido sin cortar el tamaño será de 190pb y 1135pb, respectivamente.

La secuencia de los cebadores diseñados se encuentra en la **Tabla 5** y la reacción y programa de la PCR de colonias en la **Tabla 6**.

Tabla 5. Cebadores empleados para la PCR de colonias del ensayo de minigenes.

Cebador	Secuencia (5'-3')	Temperatura hibridación	Tamaño amplicón
pSPL3_seq_D	CATGCTCCTTGGGATGTTGATG	58°C	190pb vector religado; 748pb vector con <u>inserto</u> ; 1135pb vector <u>sin cortar</u>
pSPL3_seq_R	ACTGTGCGTTACAATTTCTGG		

Tras comprobar qué colonias tienen el tamaño adecuado (748pb) en un gel de agarosa, se **extrae el ADN plasmídico** de las mismas mediante un kit de *miniprep* (extracción a pequeña escala). Para ello, se realiza un cultivo de la colonia de interés en 4 mL de medio líquido LB Broth con (50 µg/mL) de ampicilina. Los tubos se incuban en agitación (175rpm) a 37°C durante 16h, en inclinación y con la tapa ligeramente abierta para que entre oxígeno, necesario para el crecimiento de las bacterias aerobias. Para la *miniprep* se emplea el kit comercial *Macherey-Nagel™ NucleoSpin Plasmid QuickPure™ Kit* (ThermoFisher Scientific) siguiendo las instrucciones del fabricante. Después de la purificación, las muestras se cuantifican en el Nanodrop, obteniendo también datos de la pureza con los ratios de absorbancia 260/280nm (A_{260}/A_{280} , indica una posible contaminación por proteínas si el ratio es menor a 1,8 para el ADN) y 260/230nm (A_{260}/A_{230} , indica una posible contaminación por otras macromoléculas, etanol, etc., si el ratio es menor a 2). Finalmente, las muestras purificadas se envían al Servicio de Genómica y Genética del CIPF para su secuenciación para determinar si se trata del constructo WT (pSPL3-*SLC38A6*^{WT}) o mutante (pSPL3-*SLC38A6*^{MUT}).

Tabla 6. Reactivos de la PCR de colonias y programa del termociclador.

Reactivo y concentración inicial	Concentración en la reacción de PCR	Volumen a añadir (V _F = 25 µL)	Etapas del programa	Temperatura	Tiempo	Ciclos
H ₂ O libre de ADNsas y ARNsas	-	19µL	Desnaturalización inicial	96°C	7 min	1
Tampón Taq polimerasa 10X (con MgCl ₂)	1X	2,5µL	Desnaturalización	96°C	1 min	35
dNTPs (20 mM)	0,4 mM	0,5µL	Annealing o hibridación	58°C	30 s	
Cebador directo, 5' (10 µM)	0,2 µM	0,5µL	Extensión	72°C	1 min	
Cebador reverso, 3' (10 µM)	0,2 µM	0,5µL	Extensión final	72°C	7 min	1
Taq ADN polimerasa, Biotools (1 U/µL)	0,02 U/µL	0,5µL	Conservación	4°C	∞	
MgCl ₂ (20mM)	0,4 mM	0,5µL				
ADN molde (~ 100-200 ng/µL)	~ 4-8 ng/ µL	1µL				

3.3.1.5. Transfección de células HeLa con el vector *wild type* y c.690+2T>G de *SLC38A6*

Se transfectan células HeLa con el vector pSPL3-*SLC38A6*^{WT}, pSPL3-*SLC38A6*^{MUT} o pSPL3-vacío (sin inserto), empleado como control positivo de salto de exón. Las células HeLa se cultivan en monocapa en *flasks* T75 de poliestireno (BioLite ThermoFisher Scientific) TC (tratados para

favorecer la adhesión), con medio DMEM *High Glucose* (*Dulbecco's Modified Eagle's Medium*, 4,5 mg/mL glucosa), suplementado con 2mM L-Glutamina, 10% (v/v) de suero fetal bovino (FBS) inactivado, 100 U/mL de penicilina y 100 µg/mL de estreptomycin, todos ellos de Sigma-Aldrich. Las células se mantienen en un incubador a 37°C con una atmósfera al 5% de CO₂ y oxígeno a saturación (98%).

Para la transfección, las células se cultivan en placas de 6 pocillos (ThermoFisher Scientific), sembrando 50.000 células/pocillo. Para ello, se quita el medio a las células sembradas en el *flask* y se lavan con PBS (*Phosphate buffered saline*, Sigma-Aldrich) para eliminar los restos de medio que pudieran quedar. Posteriormente, las células se levantan del *flask* con tripsina, incubándolas 3 minutos en el incubador. Después, se añade medio para neutralizar la tripsina, gracias al FBS presente en el mismo, y se recolectan las células en un tubo tipo Falcon de 15mL. Tras esto, se cuentan las células en cámara Neubauer y se deposita el volumen correspondiente en los diferentes pocillos para sembrar 50.000 células/pocillo, sin superar los 2mL totales (medio + células) por pocillo. Las células se incuban durante 24h antes de la transfección. Los reactivos utilizados para la siembra de células deben estar previamente atemperados a 37°C.

Al día siguiente, se realiza la transfección de las células HeLa con FuGENE® HD (Promega) con un ratio 5:2 (µL FuGENE:µg ADN), y empleando Opti-MEM™ (Gibco), un medio reducido en suero. Una vez preparadas las mezclas de transfección como se indica en la **Tabla 7**, se agitan y se incuban 15 minutos a temperatura ambiente. En este tiempo, se refresca el medio de las células con medio DMEM nuevo. Por último, se le da un *spin* a la mezcla de transfección y se añaden 50µL al pocillo correspondiente, mezclando en cruz y llevando la placa al incubador.

Tabla 7. Volúmenes a añadir de los reactivos para el protocolo de transfección para cada una de las mezclas.
C_i: concentración inicial; C_f: concentración final.

Reactivo	pSPL3- vacío (sin inserto) C _i = 1 µg/µL C _f = 2 µg	pSPL3-SLC38A6 ^{WT} C _i = 0,359 µg/µL C _f = 2 µg	pSPL3-SLC38A6 ^{MUT} C _i = 0,251 µg/µL C _f = 2 µg
ADN	2 µL	5,57 µL	7,97 µL
Opti-MEM™	100 µL		
FuGENE® HD	5 µL		

3.3.1.6. Extracción de ARN

A las 24 y 48 horas post-transfección se extrae el ARN de las células HeLa transfectadas. Para ello, se recoge el *pellet* de células, levantando las células de igual manera que se hizo en los *flask*. En resumen, se lava cada pocillo con de PBS, se levantan las células con tripsina, se recuperan las células 200 µL de medio de cultivo, y se transfieren a un tubo LoBind® de 1,5 mL (Eppendorf). Seguidamente, se centrifugan las células 5 minutos a 1500g y 4°C, obteniendo el *pellet* de células tras eliminar el sobrenadante. Se lava el *pellet* en 1 mL de PBS frío (4°C) para eliminar los posibles restos de medio y se vuelve a centrifugar 5 minutos a 400g y 4°C, eliminando el sobrenadante y resuspendiendo el *pellet* en 1mL de trizol (Thermo Fisher).

En este momento comienza la extracción de ARN con trizol-cloroformo. Es muy importante rociar con la solución *RNAse ZAP™* (Sigma) las superficies, pipetas, tubos y guantes para evitar la degradación del ARN por las ARNasas. Además, se debe trabajar en campana, dada la toxicidad de los reactivos empleados. Tras incubar los *pellets* con el trizol 5 minutos a temperatura ambiente, se añaden 200 µL de cloroformo (Scharlab) a cada tubo y se agitan fuertemente en el vórtex. Después, se incuban a temperatura ambiente durante 3 minutos y se centrifugan las muestras 15 minutos a 12.000g a 4°C. Tras la centrifugación, el ARN se encontrará en la fase acuosa, la cual se transfiere a un nuevo tubo LoBind®. Se le añaden 500 µL de isopropanol, mezclando por inversión e incubando los tubos 10 minutos a temperatura ambiente. Se centrifugan 10 minutos a 12.000g a 4°C, se

elimina el sobrenadante y se lava el *pellet* con 1 mL de etanol al 75%, centrifugando de nuevo a 4°C 5 minutos a 7.500g. Se descarta el etanol y se centrifugan los tubos en seco para eliminar las gotas restantes de etanol. Para asegurar la eliminación completa del etanol, se secan los tubos en campana durante 5-10 minutos con la tapa abierta. Por último, se añaden 30 µL de agua libre de ADNasas/ARNasas (Invitrogen) a cada tubo y se incuban durante 15 minutos a 55°C en agitación ligera. El ARN extraído se guarda a 4°C, para su posterior cuantificación espectrofotométrica en el Nanodrop, evaluando también su pureza con los ratios A_{260}/A_{280} y A_{260}/A_{230} .

3.3.1.7. Retrotranscripción a ADNc

Para analizar el patrón de *splicing*, se retrotranscribe el ARN a ADNc para poder amplificarlo. La retrotranscripción se realiza con el kit comercial *qScript cDNA SuperMix* (Quantabio), siguiendo las instrucciones del fabricante. Para retrotranscribir 250ng de ARN, se utilizaron las condiciones especificadas en la **Tabla 8**.

Tabla 8. Volúmenes a añadir de los reactivos para la retrotranscripción. V_F: volumen final.

Reactivo (V _F = 20µL)	ARN 24h			ARN 48h		
	VACÍO	WT	MUT	VACÍO	WT	MUT
ARN	2,10 µL	2,98 µL	3,09 µL	1,18 µL	4,9 0µL	1,32 µL
H ₂ O libre de ADNasas y ARNasas	13,90 µL	13,02 µL	12,91 µL	14,82 µL	11,10 µL	14,68 µL
SuperMix (5X)*	4 µL			4 µL		

*El SuperMix (5X) contiene dNTPs, MgCl₂, transcriptasa reversa, cebadores aleatorios, cebadores oligo(dT), inhibidor de ARNasas y estabilizadores.

Una vez realizadas las mezclas correspondientes, la retrotranscripción se lleva a cabo en el termociclador con el siguiente programa: 5 minutos a 25°C, 30 minutos a 32°C, 5 minutos a 85°C y un tiempo de conservación indefinido a 4°C.

Finalizada la retrotranscripción, se realiza una amplificación del ADNc mediante PCR con cebadores específicos y se testa en un gel de agarosa. Por una parte, se emplean los cebadores diseñados en los exones del vector pSPL3 (SD6 y SA2), entre los cuales está insertado el exón 9 del gen *SLC38A6*, y por otra parte, cebadores para amplificar el gen de control interno (*housekeeping gene*) *GAPDH*, que sirve para comprobar que la retrotranscripción ha funcionado correctamente. Los volúmenes empleados se indican en la **Tabla 6** (modificando la cantidad de molde a 1,5 µL de ADNc, por lo que el volumen de agua debe ser de 18,5 µL) y el programa del termociclador se encuentra en la **Tabla 1**, con una T_m de 55°C para ambas parejas de cebadores (SD6_F/SA2_R y GAPDH_F/GAPDH_R) y 30 segundos de extensión. La secuencia de los cebadores (Sigma) se muestra en la **Tabla 9**. Los productos de PCR se visualizan en gel de agarosa al 1,5% (100V, 45 minutos) y se secuencian.

Tabla 9. Cebadores empleados para la amplificación por PCR del ADNc del ensayo de minigenes.

Cebador	Secuencia (5'-3')	Temperatura hibridación	Tamaño amplicón
SD6_D	TCTGAGTCACCTGGACAACC	55°C	263pb (sin inserto o <i>exon skipping</i>) o 329pb (WT)
SA2_R	GCTCACAATAACCACTGAGAT		
GAPDH_D	CATTGACCTCAACTACATGG	55°C	395pb
GAPDH_R	CAAAGTTGTCATGGATGACC		

3.3.1.8. Clonación del constructo WT en vector pcDNA3

Para analizar el patrón de bandas observado en el constructo *wild type* (pSPL3-SLC38A6^{WT}), tras varios intentos fallidos de purificar banda y secuenciar debido a la cercanía de las bandas en el gel (ver **Figura 7, apartado 4.2.1**), se clonaron los productos de PCR en el vector pcDNA3 (GeneScript) (**Figura 5**) como se explica en el **apartado 3.3.1.(1-4)**. Para ello, se purificó con kit comercial (**apartado 3.3.1.1**) el ADNc del constructo WT previamente amplificado por PCR, y se digirieron el inserto y el vector con las restrictasas HindIII y BamHI (ThermoFischer Scientific) en dos reacciones de digestión consecutivas. Se purificaron las digestiones y se realizó la ligación con el ratio molar 3:1 (inserto:vector, Ecuación 2, **Anexo II**). Finalmente, se transformaron las bacterias *E.coli DH5α* con 3 µL de ligación como se detalla en el **apartado 3.3.1.3** y se realizó PCR de colonias para secuenciar los clones con inserto. Los volúmenes y reactivos empleados en las digestiones, ligación y PCR de colonias se encuentran en el **Anexo II**.

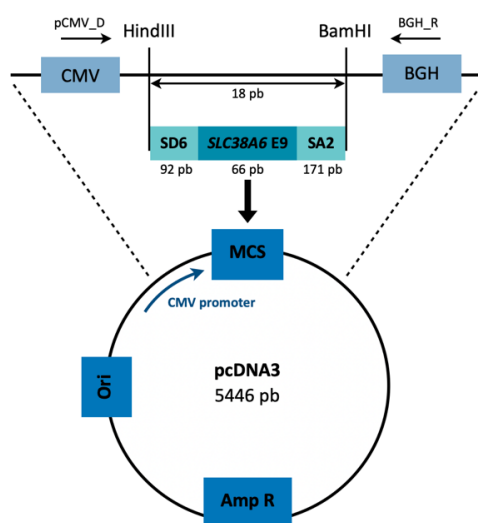


Figura 5. Esquema del vector pcDNA3 y el constructo SLC38A6^{WT}. El vector contiene el origen de replicación bacteriano (Ori) que permite la replicación del plásmido en bacterias y secuencias de resistencia a ampicilina (Amp R), empleado como agente de selección de bacterias que incorporen el vector. El MCS contiene secuencias de reconocimiento de enzimas de restricción que permiten cortar el vector, generando extremos cohesivos que dan direccionalidad al inserto. Se eliminan 18pb entre HindIII y BamHI, y se introducen 329pb (o menos, según el inserto) del constructo SLC38A6^{WT}. pCMV_D y BGH_R son los cebadores empleados en la PCR de colonias.

3.3.2. Análisis de transcritos de sangre periférica de paciente de SLC38A6

Además del ensayo de minigenes, se estudió el efecto de la variante c.690+2T>G sobre el *splicing* del gen SLC38A6 mediante análisis de transcritos obtenidos de sangre del paciente.

3.3.2.1 Extracción de ARN del probando (MD-626)

A partir de la sangre periférica del probando MD-626, conservada en tubos PAXgene para la preservación del ARN, se extrae el ARN para analizar los transcritos del gen SLC38A6, empleando el kit comercial *PAXgene Blood RNA kit* (Qiagen), siguiendo el protocolo del fabricante. Tras la extracción del ARN se cuantifica y estima su pureza (ratios A_{260}/A_{280} y A_{260}/A_{230}) en el Nanodrop.

3.3.2.2 Retrotranscripción a ADNc

El ARN extraído del paciente es retrotranscrito a ADNc tal como se describe en el **apartado 3.3.1.7**. Para un volumen final de 20µL, se deben añadir 250ng de ARN, lo cual corresponde a los volúmenes indicados en la **Tabla 10**.

Tabla 10. Volúmenes a añadir de los reactivos para la retrotranscripción. V_F: volumen final.

Reactivo (V _F = 20 µL)	ARN MD-626
ARN	4,90µL
H ₂ O libre de ADNAsas y ARNAsas	11,10µL
SuperMix (5X)*	4 µL

*El SuperMix (5X) contiene dNTPs, MgCl₂, transcriptasa reversa, cebadores aleatorios, cebadores oligo(dT), inhibidor de ARNAsas y estabilizadores.

Tras la retrotranscripción, se **amplifica mediante PCR** la región probablemente afectada por la variante. Para ello, se diseñan cebadores flanqueantes en la secuencia del ADNc de *SLC38A6*. Para evitar la amplificación eventual de una posible contaminación con ADN genómico, se diseñan cebadores solapantes entre el final de un exón y el principio del siguiente, siempre que sea posible. La secuencia de los cebadores diseñados y su T_m se muestran en la **Tabla 11**. Se realiza la amplificación del ADNc del probando y de un individuo control, siguiendo el protocolo explicado en el **apartado 3.3.1.7**.

Tabla 11. Cebadores empleados para la amplificación por PCR del ADNc de la sangre.

Cebador	Secuencia (5'-3')	Temperatura hibridación	Tamaño amplicón
SLC38A6_E6/7_D	AGTAGATATTGGTATCTTGATGGAC	60°C	476pb o 410pb (<i>exon skipping</i>)
SLC38A6_E13_R	CCTTTTAGTAATTCTGACTCCA		

3.3.3. Estudio de las proteínas PTPN23 y SMN1 en fibroblastos de paciente con variantes *missense* en PTPN23

Mediante WES de la paciente se detectaron dos variantes en heterocigosis en la proteína PTPN23: p.V133M y p.R1345H. Un estudio previo en fibroblastos de paciente con dos variantes en heterocigosis compuesta del gen *PTPN23*, una *missense* (p.N634K) y otra *nonsense* (p.L992Yfs*168) confirmó una deslocalización de la proteína SMN (*survival of motor neuron*) en los cuerpos de Cajal (CBs), supuestamente debido a la reducida funcionalidad de PTPN23 (Smigiel et al., 2018). Para comprobar si las mutaciones encontradas en nuestra paciente tienen un efecto similar, se realizaron estudios de localización subcelular de SMN1 mediante inmunofluorescencia (IFI) en fibroblastos de la paciente y de su estabilidad mediante *western blot* (WB).

3.3.3.1. Cultivo de fibroblastos primarios de paciente

Para realizar los estudios funcionales se utilizan fibroblastos primarios derivados de una biopsia de piel de la paciente (MD-172) y de un donante sano (NHF-F1A) como control. Estas células se cultivan como se ha explicado en el **apartado 3.3.1.5** para las células HeLa. Para obtener el *pellet* celular para el WB, se levantan las células (**3.3.1.5**) y, se recoge una parte en un tubo cónico de 15 mL (*pellet*) y otra parte se utiliza para mantener la línea celular. Se centrifugan las células en el tubo a 400g durante 5 minutos, se elimina el sobrenadante y se resuspende el *pellet* en PBS para lavar los posibles restos de medio, y transferir a un tubo Eppendorf de 1,5 mL. Se vuelve a centrifugar y se elimina el PBS, obteniendo así un *pellet* de células para la posterior extracción de proteínas. El *pellet* se congela a -80°C hasta su uso.

3.3.3.2. Western blot

Para comprobar si existe expresión diferencial de las proteínas PTPN23 y SMN1 entre las células de paciente (MD-172) y células sanas (NHF-F1A) se llevó a cabo un ensayo de *western blot*.

3.3.3.2.1. Extracción de proteínas

El primer paso consiste en realizar una **extracción de proteínas** a partir de los *pellets* de células congelados a -80°C . Para ello, se emplea RIPA como tampón de lisis (50 mM Tris-HCl pH 7,5; 1% (v/v) NP-40; 0,5% (v/v) ácido desoxicólico, 0,1% (v/v) SDS; y 150 mM NaCl). A este se le añade un cóctel de inhibidores de proteasas (*Complete™ Mini tablets*, Roche) e inhibidores de fosfatasa, según se indica en la **Tabla 12**, teniendo en cuenta que se van a extraer proteínas de 7 muestras a las que se debe añadir 100 μL de la mezcla. Las muestras se incuban con el tampón de lisis durante 10-15 minutos en hielo, moviendo de vez en cuando los tubos para favorecer la extracción. Seguidamente, se centrifuga a 4°C y 14.000g durante 15 minutos, tras lo cual encontraremos las proteínas en el sobrenadante, recogiendo con cuidado de no arrastrar parte del *pellet*, compuesto por *debris* y restos celulares.

Tabla 12. Componentes del tampón de lisis celular para la extracción de proteínas.

Componente		Concentración final	Volumen añadido ($V_F = 850 \mu\text{L}$)
RIPA 1X		1X	799 μL
Cóctel de inhibidor de proteasas 25X		1X	34 μL
Cóctel de inhibidor de fosfatasa	NaF (fluoruro de sodio) 0,5M	5 mM	8,5 μL
	Na_3VO_4 (ortovanadato sódico) 0,1M	0,5 mM	4,25 μL
	Na_2MoO_4 (molibdato sódico) 0,2M	1 mM	4,25 μL

3.3.3.2.2. Cuantificación de proteínas mediante el método BCA

Tras la extracción, se cuantifican las proteínas con el método BCA con el kit *Pierce BCA Protein assay kit* (Thermo Scientific). Este método se basa en la reducción del ion Cu^{2+} a Cu^+ por oxidación de las proteínas en un medio alcalino (reacción de Biuret) y la posterior reacción entre la sal del ácido bicinónico (BCA) y el ion Cu^+ , que forma un complejo de un intenso color púrpura que presenta absorbancia lineal a 562nm, de manera que la intensidad de color es proporcional a la concentración de proteína (Shen, 2023). Este ensayo permite cuantificar proteínas en un rango de concentraciones de 20-2000 $\mu\text{g}/\text{mL}$ mediante el uso de una recta patrón, que se construye con un estándar de BSA (*bovine serum albumin*), en este caso empleando 8 puntos (**Tabla 13**). Para cada punto de la recta patrón, así como para cada muestra se hacen duplicados, calculando posteriormente la media de absorbancia de ambos.

Tabla 13. Preparación de los puntos de la curva patrón con BSA.

Punto curva patrón	Volumen H_2O (μL)	Volumen BSA (μL)
2 $\mu\text{g}/\mu\text{L}$	-	300 μL de BSA 2 $\mu\text{g}/\mu\text{L}$
1,5 $\mu\text{g}/\mu\text{L}$	125	375 μL de BSA 2 $\mu\text{g}/\mu\text{L}$
1 $\mu\text{g}/\mu\text{L}$	325	325 μL de BSA 2 $\mu\text{g}/\mu\text{L}$
0,75 $\mu\text{g}/\mu\text{L}$	175	175 μL de BSA 1,25 $\mu\text{g}/\mu\text{L}$
0,5 $\mu\text{g}/\mu\text{L}$	325	325 μL de BSA 0,05 $\mu\text{g}/\mu\text{L}$
0,25 $\mu\text{g}/\mu\text{L}$	325	325 μL de BSA 0,025 $\mu\text{g}/\mu\text{L}$
0,125 $\mu\text{g}/\mu\text{L}$	325	325 μL de BSA 0,005 $\mu\text{g}/\mu\text{L}$
0 $\mu\text{g}/\mu\text{L}$	300	-

El kit cuenta con 2 reactivos (solución A y B), que se deben mezclar en una proporción 50:1, respectivamente, dando lugar a lo que llamaremos *mix BCA*. En una placa de 96 pocillos, se añade en cada pocillo 5 μ L de BSA o muestra, esta última a una dilución 1:2 (2,5 μ L de muestra y 2,5 μ L de tampón de lisis) para evitar que la concentración sea mayor que el punto más alto de la recta patrón. Se pone un blanco, que servirá para restar la absorbancia basal no debida a la proteína, compuesto por 5 μ L de tampón de lisis. Se añaden 100 μ L/pocillo de *mix BCA* y se incuba durante 30 minutos a 37°C en oscuridad. Posteriormente, se mide la absorbancia a 562 nm en un espectrofotómetro *Wallac 1420 Victor2™* (Perkin Elmer) y se calcula la concentración de proteína de las muestras a partir de la recta patrón. Para ello, se calcula la media de absorbancia de las medidas del BSA y se les resta la absorbancia su blanco (H₂O). Después, se ajustan los datos a una recta de regresión lineal de tipo $y=mx+n$, siendo “x” la concentración de proteína en μ g/ μ L e “y” la absorbancia. De esta manera, poniendo el valor medio de la absorbancia de la muestra (restado su blanco) y despejando la “x” de la ecuación obtenida, se obtienen las concentraciones de proteína de las muestras, multiplicando por la inversa de la dilución. La ecuación de la recta patrón obtenida y las concentraciones calculadas a partir de esta se muestran **Tabla 14**.

3.3.3.2.3. Preparación de muestras para western blot

Tras calcular la concentración de proteína de las muestras, se realizan los cálculos para cargar 25 μ g de proteína en un volumen de 31 μ L en un gel de poliacrilamida (PAGE, *polyacrilamide gel electrophoresis*). Las muestras se preparan en tampón de carga casero SB 6X (5% glicerol; 1,7% SDS y 0,5% (p/v) azul de bromofenol; 0.06M Tris pH 6.8 y DTT 60 mM), que se diluye con la proteína y H₂O mili-Q a 1X. Los volúmenes necesarios para cada muestra se muestran en la **Tabla 14**.

A continuación, se hierven las muestras 5 minutos a 95°C en el termobloque para desnaturalizar las proteínas, se les da un *spin* en la centrífuga y se guardan a -20°C hasta el momento la electroforesis en gel de poliacrilamida en condiciones desnaturalizantes (SDS-PAGE).

Tabla 14. Medidas de absorbancia de las muestras, recta patrón y preparación de muestras para la carga en el gel de poliacrilamida.

Muestra	Absorbancia media 562nm	Concentración de proteína (µg/µL)*	Volumen proteína (µL)**	Volumen SB 6X + DTT (µL)**	Volumen H2O mili-Q (µL)**
MD-172 (p8)	0,278	2,57	22,35	11,88	37,07
MD-172 (p6)	0,242	2,23	25,75		33,67
MD-172 (p5)	0,229	2,11	27,25		32,16
NHF-F1A (p11)1	0,204	1,87	30,72		28,70
NHF-F1A (p11)2	0,201	1,84	31,29		28,13
NHF-F1A (p7)1	0,112	0,998	57,61		1,81
NHF-F1A (p7)2	0,110	0,972	59,15		0,27
Ecuación de la recta patrón: $y = 0,2101x + 0,0075$; $R^2 = 0,9942$					

*Se multiplica el valor obtenido por 2 ya que se habían diluido las muestras en tampón RIPA (1:1). ** Los cálculos se han realizado contando con hacer 2,3 *western blot*.

3.3.3.2.4. Electroforesis de proteínas: SDS-PAGE

El siguiente paso consiste en realizar una electroforesis en gel de poliacrilamida en condiciones desnaturalizantes (SDS-PAGE), en geles comerciales (Bio-Rad) en gradiente (4-15%).

En primer lugar, se lava el gel con agua destilada para quitar los restos de tampón en el que viene y se coloca en un soporte (Bio-Rad). A continuación, se llena la cubeta de *running buffer* 1X o tampón TGS (tampón Tris-glicina pH 7,5 con un 0,1% de SDS, detergente aniónico que desnaturaliza

las proteínas y las dota con cargas negativas proporcionalmente a su masa) y se introduce en ella el soporte con el gel de poliacrilamida. Después, se cargan 31 μ L de muestra y 6 μ L patrón de peso molecular (*PageRuler PreStained 10-250kDa*, ThermoFisher Scientific), y se comienza a correr la electroforesis durante 10-15 minutos a 90V para que las muestras entren en el gel. Después, se aumenta el voltaje hasta 120V para separar las proteínas por su peso molecular, durante 90 minutos aproximadamente.

3.3.3.2.5. Transferencia a la membrana de PVDF

El protocolo del inmunoblot comienza con la transferencia de las proteínas separadas mediante SDS-PAGE a una membrana de fluoruro de polivinilideno (PVDF o *polyvinylidene fluoride*), previamente activada con metanol durante 5 minutos.

Primero, se monta un *sándwich* de la siguiente manera: se pone parte negra del soporte, se añade una almohadilla de fibra, después un papel de filtro Whatman remojado en tampón, a continuación el gel de poliacrilamida con las proteínas separadas, luego la membrana de PVDF activada (evitando la formación de burbujas), seguidamente otro papel Whatman y, finalmente, otra almohadilla de fibra y la parte transparente del soporte, cerrando el *sándwich*. El tampón en el que se mantienen todos los componentes del *sándwich* es el *Transfer Buffer* o tampón de transferencia, consistente en *buffer* Tris-Glicina con un 20% de metanol. El *sándwich* montado se introduce en un soporte que se introduce en la cubeta llena con *Transfer buffer*. En este proceso es muy importante mantener la cubeta en frío (hielo) para que no se sobrecaliente durante la transferencia. Además, se añade un enfriador dentro de ella y un pez para agitar con un agitador magnético, y se lleva a una cámara de 4°C, con la cubeta de transferencia a 100V durante 2 horas.

3.3.3.2.6. Inmunodetección de proteínas transferidas a la membrana

Acabada la transferencia, se lava la membrana durante 5 minutos en agua y después se incuba 1 hora en agitación con solución de bloqueo, la cual contiene 5% de leche en polvo diluida en TBST 1X (25 mM Tris-HCl, 150 mM NaCl pH 7,5 con 0,1% (v/v) Tween-20). Transcurrida una hora, se corta la membrana en diferentes trozos según los tamaños de las proteínas a estudiar para poder incubar cada uno con su anticuerpo primario correspondiente. Se realizan tres fragmentos para estudiar las proteínas PTPN23 (179kDa), SMN1 (35kDa) y LC3B (entre 14 y 16 kDa), empleando como control de carga la proteína α -tubulina (52kDa). Los anticuerpos primarios empleados se recogen en la **Tabla 15**. Las membranas se meten en bolsitas termoselladas con 2-3mL de anticuerpo primario, disuelto en solución de bloqueo, y se dejan a 4°C y en agitación ligera (50-100 rpm) *overnight*. Al día siguiente, se realizan 3 lavados en agitación (100 rpm) de 10 minutos cada uno con TBST 1X, añadiendo un volumen suficiente para cubrir las membranas. Seguidamente, se incuban las membranas con el anticuerpo secundario correspondiente (**Tabla 15**), conjugado con HRP (*horseradish peroxidase*) para el posterior revelado; a temperatura ambiente, y en agitación (100 rpm) durante 1 hora. Pasado este tiempo, se vuelven a hacer 3 lavados con TBST como se ha explicado anteriormente y se dejan las membranas en TBST hasta su revelado.

Tabla 15. Anticuerpos primarios y secundarios empleados en el protocolo de western blot.

Anticuerpo y dilución		Proteína diana	Casa comercial	Referencia
Primario	HD-PTP(F-4) (<i>mouse</i>) 1:500	PTPN23	Santa Cruz	#sc-398711
	anti-SMN1 (<i>mouse</i>) 1:500	SMN1	Adipogen	AG-20T-0003-C200)
	anti-LC3B (<i>rabbit</i>) 1:2000	LC3B	Cell Signalling Technology	#3868(1673868T)
	anti- α -tubulina (<i>mouse</i>) 1:2000	Tubulina	Sigma	#T8203
Secundario	<i>goat</i> anti- <i>mouse</i> HRP 1:10000	Mouse IgG	Invitrogen Thermo Fisher	#31430
	<i>goat</i> anti- <i>rabbit</i> HRP 1:10000	Rabbit IgG		#31460

3.3.3.2.7. Revelado mediante reacción de quimioluminiscencia

Finalmente, se procede al revelado por medio de una reacción de quimioluminiscencia con el *Amersham™ ECL™ Prime Western blotting detection reagent* (GE healthcare life sciences), el cual consta de una solución A que contine luminol y una solución B de peróxido de hidrógeno (H_2O_2). Los reactivos se mezclan en proporción 1:1 y, para que tenga lugar a reacción de quimioluminiscencia, se extiende por la zona de la membrana donde se encuentra la proteína de interés y se incuba durante 30s-1min en oscuridad. La reacción se fundamenta en la oxidación del luminol por parte de la HRP que lleva acoplado el anticuerpo secundario en presencia de H_2O_2 . Dicha oxidación produce 3-aminofalato, una molécula excitada que emite fotones a 425nm, los cuales son captados con dispositivos de imagen con cámara CCD que detectan la quimioluminiscencia (Cytiva Amersham ImageQuant 800).

3.3.3.2.8. Cuantificación y análisis estadístico

La intensidad de las bandas obtenidas en el inmunoblot se cuantifica con la herramienta ImageJ (<https://imagej.net>). Para ello, se trazan rectángulos individuales a cada banda y se obtiene el área bajo la curva mediante el uso de *wand (tracing) tool* tras seleccionar *Analyze>Gels>Plot Lanes* en el programa. A continuación, se relativiza cada muestra a la medida del área del gen constitutivo (tubulina, en este caso) y, finalmente, se exportan los datos a GraphPad v.10 (www.graphpad.com) para analizar si existen diferencias significativas de intensidad (y, por tanto, concentración) entre las muestras control (NHF-F1A) y problema (MD-172). Se realiza un test t de Student *unpaired*, asumiendo normalidad y homocedasticidad en las poblaciones.

3.3.3.3. Inmunofluorescencia (IFI)

Con el objetivo de comprobar si se produce deslocalización entre las proteínas SMN1 y Coilin, un marcador de los cuerpos de Cajal (CB) como se ha descrito en estudios anteriores (Smigiel et al., 2018), se realiza el protocolo de inmunofluorescencia.

3.3.3.3.1. Siembra y fijación de células en cristales

Las células cultivadas como se detalla en el **apartado 3.3.3.1** se levantan y cuentan en cámara Neubauer, calculando el volumen necesario para sembrar 20.000 células/pocillo en placas de 12 pocillos (p12) que contienen cristales para inmunofluorescencia. A las 24 horas, se fijan las células con paraformaldehído 4% (PFA) durante 15 minutos a temperatura ambiente y en oscuridad. Tras este tiempo, se hacen 5 lavados con PBS para eliminar los restos de PFA y se mantienen los cristales a 4°C con una solución de PBS y 0,02% de azida sódica, para evitar contaminaciones.

3.3.3.3.2. Protocolo de inmunofluorescencia

En primer lugar, los cristales se lavan con PBS para eliminar los restos de azida. Después, se procede al **bloqueo y permeabilización** de las células, eliminando los sitios de unión inespecíficos para los anticuerpos y permitiendo la entrada del anticuerpo primario al interior celular. Este paso se realiza en una sola etapa, empleando una disolución de *horse serum* (HS, para el bloqueo) al 5% preparada en PBS-Tx100 (PBS con Triton X-100 (0,25%), detergente que permeabiliza las células). Los cristales se incuban en esta solución durante 1 hora a temperatura ambiente en agitación (50 rpm) en una placa p12. A continuación, se pasan los cristales a una cámara húmeda, se añade el **anticuerpo primario** sobre su superficie y se incuba *overnight* a 4°C. Los anticuerpos primarios se preparan en *horse serum* al 1% en PBS-Tx100. Las diluciones empleadas se indican en la **Tabla 16**. Al día siguiente, se hacen 3 lavados con PBS en agitación (50 rpm) durante 10 minutos cada uno en

la placa p12 y se incuban los cristales con el **anticuerpo secundario (Tabla 16)**, disueltos en HS al 1%, durante 1 hora en agitación y oscuridad. A partir de este momento es muy importante mantener los cristales en oscuridad para evitar la pérdida de fluorescencia ya que los anticuerpos secundarios se encuentran conjugados a fluoróforos específicos. Después, se vuelven a realizar 3 lavados en PBS como se ha indicado y se incuban los cristales con DAPI (1 mg/mL, dilución 1:1000 en PBS) durante 15 minutos en agitación y oscuridad para teñir los núcleos celulares. Finalmente, se hacen 2 lavados más con PBS y se pasa al montaje de los cristales.

Tabla 16. Anticuerpos primarios y secundarios empleados para cada experimento de inmunofluorescencia (IFI) y sus diluciones correspondientes.

EXPERIMENTO		Anticuerpo primario	Anticuerpo secundario (1:1000)	
1	NHF	Anti-Coilin (Rb) 1:1000 + Anti-SMN1 (Ms) 1:100	Anti-Rb 488nm; Anti-Ms 633nm	
	MD			
	Neg Ctrl			
2	NHF	Anti-Coilin (Rb) 1:800 + Anti-SMN1 (Ms) 1:100	Anti-Rb 633nm; Anti-Ms 488nm	
	MD			
3	HeLa	Anti-Coilin (Rb) 1:800 + Anti-SMN1 (Ms) 1:100	Anti-Rb 633nm; Anti-Ms 488nm	Anti-Rb 488nm; Anti-Ms 633nm
	Neg Ctrl		Anti-Rb 633nm; Anti-Ms 488nm	
4*	NHF FGF2	Anti-Coilin (Rb) 1:800 + Anti-SMN1 (Ms) 1:100	Anti-Rb 633nm; Anti-Ms 488nm	
	NHF FBS			
	NHF E6E7			
	Neg Ctrl			

*Se probaron diferentes tratamientos en las células tras los resultados de experimentos anteriores, con el objetivo de ver si aumentaban los niveles de SMN1 en el núcleo: FGF2 (50 ng/μL) a tiempos 0, 3, 6 y 24h; FBS (20%) a tiempos 0, 3, 6 y 24h; así como una línea celular de fibroblastos inmortalizados (NHF E6E7) con el virus del papiloma humano (HPV). Neg Ctrl: control negativo (sin anticuerpo primario); Rb: *rabbit*; Ms: *mouse*. Las casas comerciales de los anticuerpos son: anti-Coilin (Cell Signalling Technology, #610456); anti-SMN1 (Bio-technique, #NB100-1936SS); los anticuerpos secundarios son todos de Invitrogen ThermoFisher (Rb 488 #A11008, Rb 633 #A21070, Ms #A11029, Ms #A21050).

3.3.3.3. Montaje y observación de cristales por microscopía confocal

El último paso en el protocolo de inmunofluorescencia consiste en el montaje de los cristales para su observación en el microscopio confocal. Para ello, se coge con pinzas el cristal de la placa p12 y se pasa por agua destilada para lavar las sales del PBS, ya que podrían afectar a la visualización de los cristales. Después, se coloca el cristal con las células hacia arriba sobre un portaobjetos (2 cristales/portaobjetos), se añade una gota de agente de montaje (*FluorSave™ Reagent*, Merck Millipore) sobre las células y, por último, se coloca un cubreobjetos grande encima de los cristales. Se deja secar el agente de montaje durante 30 minutos en oscuridad y se sella el cubreobjetos con laca de uñas, guardándolo a 4°C hasta su observación. Los cristales se visualizan en el microscopio confocal TCS SP8 de Leica Microsystems, con un objetivo 63x en aceite de inmersión. Las imágenes se obtienen y analizan con el *software* LasX de Leica.

4. RESULTADOS

4.1. VALIDACIÓN DE VARIANTES CANDIDATAS EN LAS FAMILIAS

Se han estudiado dos probandos (MD-618 y MD-626) pertenecientes a dos familias (fMD-275 y fMD-279, respectivamente) mediante secuenciación de exoma completo (WES) con el objetivo de encontrar la posible causa genética del fenotipo clínico que presentan. En base a los criterios de filtrado y priorización de variantes mencionados en el **apartado 3.2.1** se seleccionaron

las variantes candidatas recogidas en la **Tabla 17**. En la familia fMD-275 se priorizaron cinco cambios candidatos en cinco genes diferentes, mientras que en la familia fMD-279 se seleccionaron cuatro variantes candidatas en tres genes diferentes. Las variantes seleccionadas se validaron por secuenciación Sanger del ADN genómico, y se estudió la segregación en los progenitores (resultados en las **Figuras Suplementarias 2-7 del Anexo I**).

A continuación, se describen las funciones de los genes que poseen variantes candidatas, así como el motivo de su selección:

- **CACNA1G (NM_018896.5) c.4760-42A>T**: gen que codifica para un canal de calcio tipo T regulado por voltaje Ca(v)3.1. Su expresión es elevada en células de Purkinje y en los núcleos cerebelosos profundos. Juega un importante papel en la regulación del potencial de membrana y la entrada de calcio en las neuronas (Kunii et al., 2020). La variante identificada en el probando, **c.4760-42A>T**, tendría una herencia AD, siendo una variante *de novo*. El cambio afecta a una región intrónica que se encuentra a 42pb del exón 27, aunque según los predictores *in silico* (SpliceAI, NNSplice) no afecta al *splicing*. A pesar de ello, se trata de un gen asociado a enfermedad según la base de datos OMIM, en concreto, a una ataxia espinocerebelosa con herencia AD (MIM #616795). Además, por la baja frecuencia del cambio en población control sana (MAF<0,01%), así como por su posible relación con el fenotipo del probando, que sufre una canalopatía, es una variante que debe validarse. La secuenciación resultó en una validación negativa, confirmando que se trataba de un artefacto de la tecnología WES (**Figura Suplementaria 2**).

- **TAF4 (NM_003185.4) c.1018_1019insTTACT**: codifica para una subunidad del factor de transcripción IID (TFIID), un complejo requerido para la iniciación de la transcripción por parte de la ARN polimerasa II. TFIID está compuesto por TBP (*TATA-box binding protein*) y por TAFs (*TBP-associated factors*), un conjunto de proteínas evolutivamente conservadas. La unión aberrante de TAF a proteínas que contienen regiones de poliglutamina expandidas se ha relacionado con un grupo de trastornos neurodegenerativos conocidos como enfermedades de poliglutamina (Janssen et al., 2022). La variante **c.1018_1019insTTACT** cambia la pauta de lectura (*frameshift*) y se predice probablemente patogénica por la base de datos Varsome. Asimismo, el gen está asociado con un trastorno del desarrollo intelectual, con herencia AD en OMIM (MIM #620450). No obstante, dada la complejidad de la región a amplificar (por ser muy repetitiva), no fue posible validar el cambio.

- **CROCC (NM_014675.5) c.12G>A**: este gen codifica para la proteína rootletin, un componente estructural similar al citoesqueleto que se proyecta desde el extremo proximal del cuerpo basal de los cilios. Si bien no se conoce completamente su función, se sabe que se une a otras proteínas como kinesinas, APP (amyloid precursor protein) y presenilinas. Además, ayuda a estabilizar los cilios contra el flujo externo y regula el tráfico ciliar (Turn et al., 2021). Aunque no ha sido asociado con enfermedad, se trata de un gen aún algo desconocido, motivo por el cual resultaba interesante su validación, pudiendo quizá asociarlo con enfermedad con una variante novel. La validación mediante secuenciación Sanger confirmó en la existencia del cambio *de novo* en heterocigosis en el probando MD-618 (**Figura Suplementaria 3**).

- **CACNA1A (NM_023035.3) c.7266_7271delCAGCGG**: es un gen que codifica para un canal de calcio regulado por voltaje, que regula la entrada de iones calcio en células excitables. También se relaciona con otras funciones como la contracción muscular, la liberación de neurotransmisores o la expresión génica. Los canales de calcio son complejos formados por varias subunidades: alfa-1, beta, alfa-2/delta y gamma. CACNA1A constituye la subunidad alfa-1 y se expresa principalmente en tejido neuronal (Lipman et al., 2022). Las mutaciones en este gen se han asociado con trastornos neurológicos (MIM #617106), ataxia episódica (MIM #108500) y ataxia espinocerebelosa (MIM #183086). A pesar de que la frecuencia no es tan baja como cabría esperar para una enfermedad rara, el gen resulta interesante por su función ya que el paciente sufre una canalopatía. Tras la

validación, se corroboró la existencia de la delección en heterocigosis en el paciente (**Figura Suplementaria 4**).

- **ATAD3A (NM_018188.5) c.908T>A**: este gen codifica una proteína mitocondrial de membrana, expresada ubicuamente, relacionada con la dinámica mitocondrial, la organización del nucleóide, la traducción de proteínas y el crecimiento celular (Skopkova et al., 2023). *ATAD3A* es miembro de la familia génica ATPasa con dominios AAA. Se ha asociado con el síndrome de Harel-Yoon (MIM #617183), entre otras enfermedades. Se trata de un cambio en heterocigosis procedente de la madre. Se propone su validación dada su baja frecuencia en población control (MAF<0,01%) y la posible creación de un sitio aceptor de *splicing* predicha por SpliceAI. Sin embargo, aunque se pudo amplificar por PCR la región, no fue posible secuenciarla debido a su complejidad, con alto nivel de repeticiones de Gs y Cs.

- **SLC38A6 (NM_001172702.2) c.690+2T>G**: también conocido como *SNAT6*, *SLC38A6* es miembro de la familia de transportadores iónicos *SLC38* que participan en el ciclo glutamina-glutamato en el sistema nervioso central, permitiendo el transporte transmembrana de L-glutamina (Gandasi et al., 2021). Es el único miembro de la familia que se expresa exclusivamente en las neuronas excitatorias del cerebro. Poco más se conoce sobre su función, el cual resulta interesante como candidato del probando MD-626 por estar en homocigosis y producir, *in silico*, pérdida de sitios dadores y aceptores de *splicing* según SpliceAI. Tras la validación por secuenciación Sanger se confirmó la existencia del cambio en homocigosis en el paciente (**Figura Suplementaria 5**).

- **SLITRK3 (NM_001318810.2) c.2075G>T**: miembro de la familia *SLITRK*, formada por proteínas transmembrana relacionadas estructural y funcionalmente con el control del crecimiento de neuritas, un proceso fundamental en la diferenciación de las neuronas. Su región C-terminal es homóloga a los receptores neurotrofina (NTRK). En concreto, *SLITRK3* (Slit and Trk-like family member 3) es una molécula de adhesión de células sinápticas que se expresa en las sinapsis inhibitorias y promueve su formación (Efthymiou et al., 2024). El cambio se ha producido *de novo* en el probando MD-626, y no se ha descrito previamente la variante que presenta, un cambio *missense* de una arginina por una leucina. Dicho cambio se validó por secuenciación Sanger, siendo positivo en heterocigosis (**Figura Suplementaria 6**).

- **ASTN1 (NM_004319.3) c.3227-8C>G y c.3089C>T**: gen que codifica para la astrotactina, una molécula de adhesión neuronal requerida para la migración guiada por la glía de neuroblastos posmitóticos jóvenes en regiones corticales del cerebro en desarrollo (cerebro, hipocampo, cerebelo y bulbo olfatorio) (Chang, 2017). El probando MD-626 presenta dos cambios en este gen, en heterocigosis compuesta, heredando uno de los cambios de la madre y, otro, del padre. La frecuencia de ambos es relativamente baja, sobre todo del cambio procedente de la madre (MAF<0,01%), y no se predice que ninguno afecte al *splicing*, aunque resulta conveniente su validación por si pudieran asociarse a enfermedad. La validación confirmó la presencia de ambas variantes en el probando, heredadas de sus progenitores (**Figura Suplementaria 7**).

4.1.1. Estudio de variaciones en el número de copias (CNVs)

Como análisis complementario al estudio de las variantes de un único nucleótido (SNVs), se analizaron las variantes en el número de copias (CNVs), detectadas por los programas XHMM y CoNIFER en base a los datos de secuenciación de exoma completo. El análisis no permitió la identificación de ninguna CNV de interés en la familia fMD-279, al contrario que en la familia fMD-275, donde se observó una posible delección pequeña en el gen *CROCC* (NM_014675.5), que presentaba ya un SNV candidato en heterocigosis *de novo*. En concreto, se trataría de una delección de 26.932 pares de bases que afecta a los exones 1-19 del gen que, junto con el cambio c.12G>A, podría contribuir a producir la enfermedad del probando MD-618.

Tabla 17. Variantes candidatas priorizadas en las familias analizadas por WES.

Familia	Gen	Individuo	Genotipo	HGVS ADNc	HGVS proteína	Frecuencia en gnomAD	rs (dbSNP)	Bases de datos		
								ClinVar	Varsome	SpliceAI (predicador de <i>splicing in silico</i>)
fMD-275	CACNA1G (NM_018896.5)	MD-618	0/1	c.4760-42A>T	-	GRCh38 0.00004257	-	-	Likely benign	No afecta al <i>splicing in silico</i>
		MD-619	0/0							
		MD-620	0/0							
	TAF4 (NM_003185.4)	MD-618	0/1	c.1018_1019insTTACT	p.Pro340LeufsTer71	-	-	-	Likely pathogenic	-
		MD-619	0/0							
		MD-620	0/0							
	CROCC (NM_014675.5)	MD-618	0/1	c.12G>A	p.Gly4=	GRCh38 0.000	-	-	Likely benign	No afecta al <i>splicing in silico</i>
		MD-619	0/0							
		MD-620	0/0							
	CACNA1A (NM_023035.3)	MD-618	0/1	c.7266_7271delCAGCGG	p.Ser2423_Gly2424del	GRCh38 0.0003993	rs775428832	Conflicting interpretations of pathogenicity	VUS	No afecta al <i>splicing in silico</i>
		MD-619	0/1							
		MD-620	0/0							
	ATAD3A (NM_018188.5)	MD-618	0/1	c.908T>A	p.Val303Asp	GRCh38 0.00001316	rs539009318	-	VUS	Ganancia de sitio aceptor (0,42)
		MD-619	0/1							
		MD-620	0/0							
fMD-279	SLC38A6 (NM_001172702.2)	MD-626	1/1	c.690+2T>G	-	GRCh38 0.0000034	rs766211959	-	VUS	Pérdida de sitio dador (0,59) y de sitio aceptor (0,55)
		MD-627	0/1							
		MD-628	0/0							
	SLITRK3 (NM_001318810.2)	MD-626	0/1	c.2075G>T	p.Arg692Leu	-	-	-	VUS	-
		MD-627	0/0							
		MD-628	0/0							
	ASTN1 (NM_004319.3)	MD-626	0/1	c.3227-8C>G	-	GRCh38 0.000001859	-	-	Likely benign	No afecta al <i>splicing in silico</i>
		MD-627	0/1							
		MD-628	0/0							
		MD-626	0/1	c.3089C>T	p.Thr1030Met	GRCh38 0.0001290	rs143440206	Conflicting interpretations of pathogenicity	Likely benign	No afecta al <i>splicing in silico</i>
		MD-627	0/0							
		MD-628	0/1							

HGVS: Human Genome Variation Society; rs: identificador dbSNP; GRCh38: v4.1.0. de gnomAD que contiene información de variantes de WGS en la versión 38 del genoma humano; VUS: variant of uncertain significance

4.2. ENSAYOS FUNCIONALES DE LA VARIANTE C.690+2T>G EN EL GEN SLC386

Tras identificar el cambio c.690+2T>G en homocigosis del gen *SLC386* en el probando MD-626, del cual se predecía *in silico* afectaba al *splicing* al producir la pérdida de sitios de dadores y aceptores de *splicing* (SpliceAI), se realizaron dos ensayos experimentales (el ensayo de minigenes y el análisis de transcritos) para corroborar la patogenicidad de la variante.

4.2.1. Ensayo de minigenes

El estudio consistió en el diseño de una versión pequeña del gen *SLC38A6* dentro de un vector (pSPL3) (**Figura 4, apartado 3.3.1.1**), con el objetivo de analizar el potencial *splicing* diferencial entre la versión WT y mutada del gen. Para ello, se utilizó el ADN de la madre como molde por ser heterocigota, de manera que presentaba un alelo *wild type* (c.690+2T) y uno con la variante c.690+2T>G detectada en el probando, que aquí denominaremos “mutante”. Se introdujo el exón 9 de *SLC38A6* junto con ciertos nucleótidos flanqueantes entre los exones fuertes (SD6 y SA2) del vector. El vector se clonó en bacterias y se extrajo, obteniendo colonias bacterianas con el vector *wild type* (pSPL3-*SLC38A6*^{WT}) y mutante (pSPL3-*SLC38A6*^{MUT}), confirmados mediante secuenciación Sanger (**Figura 6**).

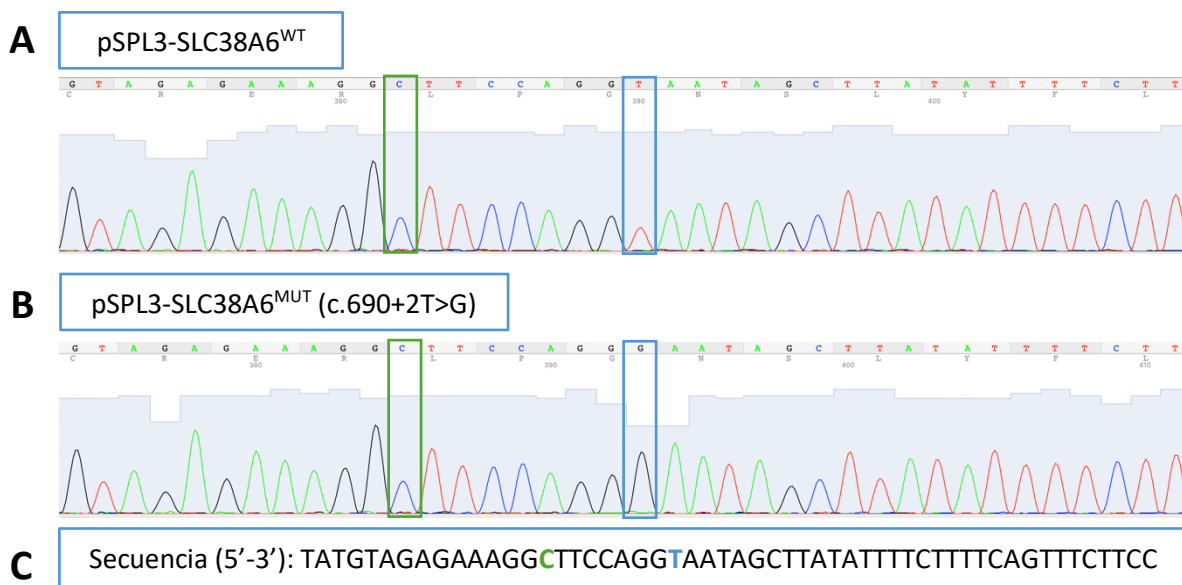


Figura 6. Electroferograma de la secuenciación Sanger de la miniprep del vector pSPL3 de clones con el inserto del gen *SLC38A6*. A) Clon wild type (sin el cambio) B) Clon mutante (con el cambio c.690+2T>G). C) Secuencia de referencia. Se marca en verde un polimorfismo que existe respecto a la secuencia del genoma de referencia (T); en azul, la base afectada por el cambio c.690+2T>G.

Tras la extracción de los vectores mediante Mini-Prep, se transfectaron células HeLa con pSPL3-*SLC38A6*^{WT}, pSPL3-*SLC38A6*^{MUT} y el vector pSPL3 vacío (sin inserto) como control de tamaño de los exones SD6-SA2, simulando el salto de exón que se produciría en la versión mutante del vector. Se extrajo el ARN de las células a tras 24 y 48 horas de la transfección, y se retrotranscribió a ADNc, el cual se amplificó por PCR con cebadores diseñados en los exones del vector pSPL3 (SD6-SA2), entre los cuales se encontraría el inserto del exón 9 del gen *SLC38A6* en la versión WT y, si se cumple la hipótesis planteada, solo los exones SD6 y SA2 en la versión mutante del vector por producirse salto de exón debido al cambio c.690+2T>G. Finalmente, se corrió un gel de agarosa con el producto de la amplificación del ADNc del minigen y del gen constitutivo *GAPDH* como control de la integridad del ADNc. En la **Figura 7** se muestra el resultado del ensayo. Las pequeñas diferencias de intensidad en *GAPDH* en el gel se deben a que el final se visualiza peor en el transiluminador y no a diferencias en la cantidad de ADN cargado en el gel. Se observaron tres

bandas en el ADNc WT, y una en las versiones mutante y vector vacío, correspondiente al tamaño de los exones SD6-SA2. Esto indica un posible salto de exón debido a la mutación en *SLC38A6*. Este resultado se corroboró en el análisis de transcritos.

Hipotetizamos que, del patrón de tres bandas observado en el ADNc del minigen WT (**Figura 7**), la banda de mayor tamaño (~400pb) corresponde a la versión *wild type* del minigen (formada por los exones constitutivos del vector SA2-SD6 y el exón 9 de *SLC38A6*). La banda de menor tamaño correspondería a una versión del minigen sin inserto, ya que su tamaño coincide con el de la versión mutante y vector vacío (constituida únicamente por los exones SD6-SA2). Probablemente, la banda intermedia sería un heterodúplex de ambas versiones del minigen (WT-vector vacío), como se ha descrito en otros ensayos de minigenes (Aparisi et al., 2013; Desviat et al., 2012). El heterodúplex es una molécula de ADN de doble cadena resultante de la recombinación de cadenas simples diferentes y parcialmente complementarias.

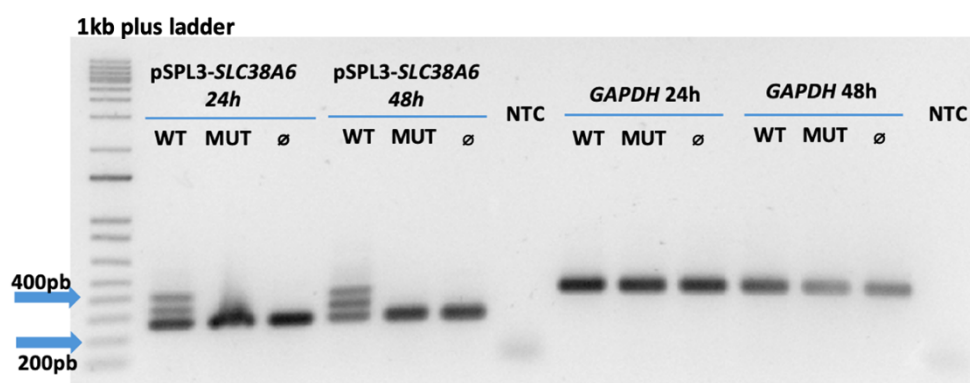


Figura 7. Resultados ensayo de minigenes. Electroforesis en gel de agarosa (1,5%) 90V, 45min donde se observa el salto de exón inducido por el cambio c.690+2T>G en el vector con el inserto mutado; WT: wild type; MUT: mutante (c.690+2T>G); ∅: vector pSPL3 vacío (sin inserto); GAPDH: gen constitutivo.

Para comprobar esta hipótesis, intentamos cortar cada una de las bandas y secuenciarlas, sin éxito debido a su cercanía en el gel. Por tanto, se procedió a la clonación del ADNc del minigen *wild type* en otro vector (pcDNA3) para poder secuenciar diferentes clones y ver qué inserto contenían. Para ello, se digirieron el vector pcDNA3 y el inserto *SLC38A6*^{WT} con las enzimas de restricción HindIII y BamHI, se llevó a cabo la ligación y se realizó la transformación bacterias *E.coli* DH5α competentes. Finalmente, se picaron diferentes colonias, se amplificaron por PCR, y se secuenciaron. Los resultados de secuenciación respaldaron la hipótesis planteada, ya que se observaron clones con insertos *wild type* (SD6-E9 *SLC38A6*-SA2) (**Figura 8A**) e insertos formados por los exones SD6-SA2 del vector pSPL3 vacío (**Figura 8B**). No hubo ningún clon con heterodúplex porque este solo se forma al correr el gel de agarosa.



Figura 8. Electroferograma de la secuenciación Sanger de la clonación en pcDNA3 del minigen *SLC38A6*^{WT}. Se observa un clon wild type (**A**) y uno que consiste en los exones constitutivos SD6-SA2 del vector pSPL3 (**B**).

4.2.2. Análisis de transcritos

El análisis de transcritos se realizó a partir del ARN extraído de sangre periférica del probando MD-626 y un control. La concentración de ARN obtenida en la extracción fue de 51 ng/ μ L, con un ratio $A_{260/280}$ de 2,12, indicando que se había extraído correctamente el ARN; y un ratio $A_{260/230}$ de 1,48, indicando una posible contaminación con tampones o alcoholes de la extracción. Aun así, se procedió a la retrotranscripción a ADNc y la posterior RT-PCR para la amplificación de los exones 7-13 del gen *SLC38A6*. Los tamaños esperados eran de 476 pb en la versión WT del gen (**Figura 9B-1**) y 410 pb en la versión mutada (**Figura 9B-2**) si se producía el salto del exón 9 predicho por SpliceAI y observado en el ensayo de minigenes.

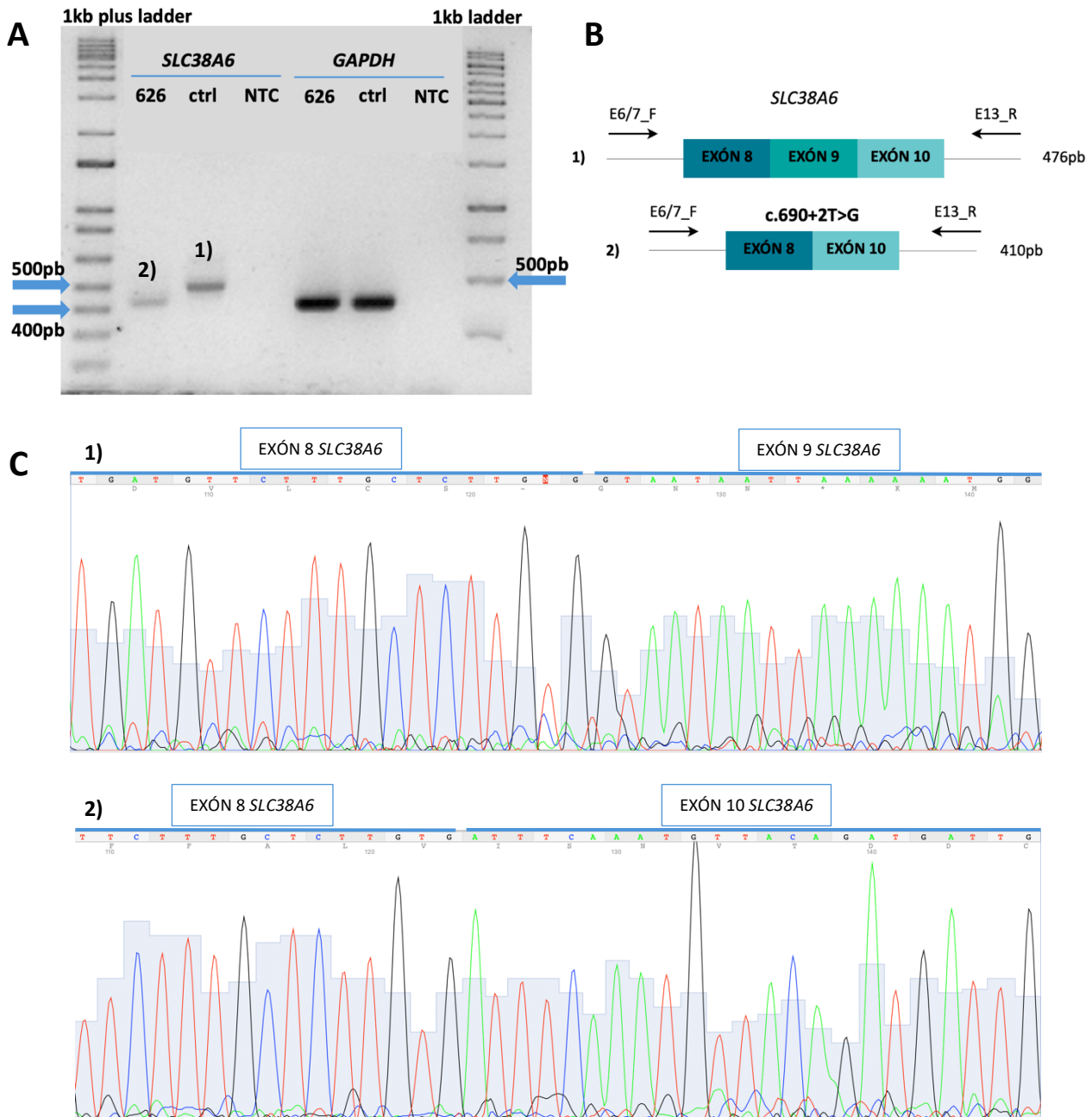


Figura 9. Resultados del análisis de transcritos. **A)** Electroforesis en gel de agarosa (1,5%) 90V, 45min que muestra una banda de menor tamaño en el individuo MD-626 que en el control. **B)** Esquema representativo de los fragmentos amplificados del gen *SLC38A6* en el individuo control (**1**) y en el probando MD-626 (**2**). **C)** Electroferograma de la secuenciación de 476pb del control (**1**) y de la banda de 410pb de MD-626 (**2**) donde se observa la pérdida del exón 9 en MD-626.

Los productos de PCR se cargaron en un gel de agarosa al 1,5% para su separación mediante electroforesis, encontrando bandas de diferente tamaño entre el probando y el control como muestra la **Figura 9A**. Esta diferencia de tamaños parecía indicar la pérdida de un exón del gen *SLC38A6* en el paciente MD-626. Para comprobarlo, se procedió a la secuenciación de los productos de PCR, la cual corroboró la pérdida completa del exón 9 en el paciente con el cambio c.690+2T>G en homocigosis, en comparación con el control, que presentaba todos los exones (**Figura 9C**).

4.3. ESTUDIO DE LAS PROTEÍNAS PTPN23 Y SMN1 EN FIBROBLASTOS DE PACIENTE

El estudio genético de la paciente MD-172 reveló la presencia de dos variantes *missense* en el gen *PTPN23* que se traducían a nivel proteico en los cambios p.V133M y p.R1345H. Dichas variantes implican el cambio a aminoácidos del mismo tipo (apolares y básicos, respectivamente), aunque de diferente tamaño. Con el objetivo de determinar la patogenicidad de las variantes, se realizaron ensayos que pudieran revelar cambios en la función o localización de la proteína codificada, como se había descrito anteriormente para otras variantes (Smigiel et al., 2018).

4.3.1. Western blot

Con el objetivo de ver si existía una expresión diferencial de las proteínas PTPN23 y SMN1 entre fibroblastos derivados de paciente (MD-172) y de donante sano (NHF-F1A) debido a los cambios nucleotídicos que la paciente presenta en el gen *PTPN23*, se realizó un *western blot*. Se quiso comprobar si aumentaba la expresión de PTPN23 en células de paciente por un mecanismo celular compensatorio, como sugieren Smigiel et al. (2018); y si se perdía una posible forma fosforilada, de mayor tamaño, de la proteína SMN1.

En cuanto a PTPN23, no logramos la banda correspondiente en la membrana (datos no mostrados), probablemente debido a que el anticuerpo no fue suficientemente específico. Sin embargo, se deberían probar otros anticuerpos o extractos de líneas celulares diferentes para verificar que se trata de un problema técnico. Con relación a SMN1, en la **Figura 10** se observa la aparición de dos bandas, similar a lo que describen Smigiel y colaboradores (2018), aunque no sabemos exactamente a qué corresponde cada una de ellas. Hipotetizamos que la banda inferior (35 kDa) debería ser SMN1, ya que coincide con su peso molecular predicho. La banda superior podría ser otra isoforma de la misma proteína o una forma modificada post-traduccionalmente (p.e., una forma fosforilada). A diferencia de Smigiel et al. (2018), que observaron una pérdida de la banda superior para SMN1 en los fibroblastos de paciente (p.N634K/p.L992Yfs*168), nosotros no observamos dicho patrón en la paciente MD-172 (p.V133M/p.R1345H).

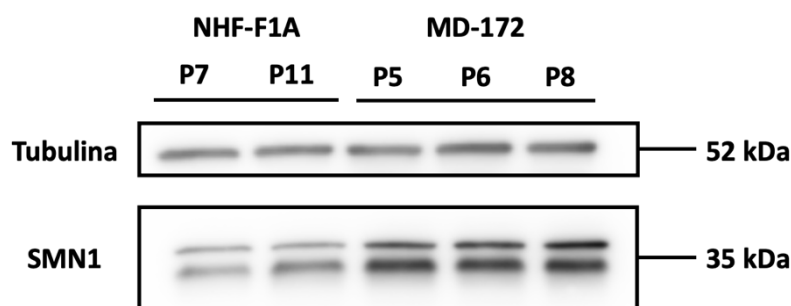


Figura 10. Resultados representativos del inmunoblot de SMN1 con extractos proteicos de NHF-F1A (sano) y MD-172 (paciente). Se emplea α -tubulina como control de carga (proteína constitutiva, cuya expresión no varía independientemente de la condición). P(5-11): número de pase en que se encontraban los fibroblastos de los que proceden los extractos proteicos.

Dado que se apreciaban diferencias de intensidad de SMN1 entre las muestras control (NHF-F1A) y las de paciente (MD-172), y no en la proteína constitutiva (α -tubulina), señalando que no se trataba de un error de carga; se cuantificó la intensidad de las bandas con el *software* ImageJ (NIH). Se encontraron diferencias estadísticamente significativas entre los grupos (**Figura 11**) tanto en la banda superior como inferior. Estas diferencias podrían deberse a diversos factores técnicos o biológicos, que se explicarán en el apartado de discusión.

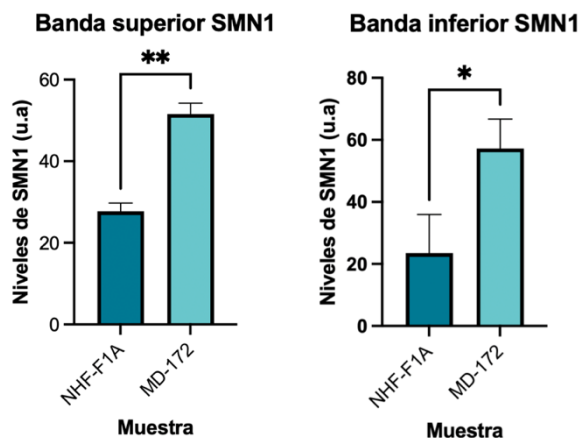


Figura 11. Cuantificación de los niveles de expresión de la proteína SMN1 en fibroblastos sanos (NHF-F1A) o de paciente (MD-172). Se cuantificó la intensidad de las bandas del inmunoblot obtenidas para SMN1 con el software ImageJ (NIH). Los valores de intensidad de SMN1 se relativizaron respecto a la intensidad de α -tubulina correspondiente. Los datos representados corresponden a la media $\pm$ desviación estándar de dos repeticiones biológicas para NHF-F1A (n = 2) y 3 repeticiones biológicas para MD-172 (n = 3). **p*-valor $\leq 0,05$ y ***p*-valor $\leq 0,01$ utilizando la prueba t de Student.

4.3.2. Inmunofluorescencia (IFI)

Por otra parte, Smigiel y colaboradores (2018) también estudiaron la actividad celular de SMN1 indirectamente a través de la acumulación de la proteína en los cuerpos de Cajal (CBs). Para ello, realizaron un protocolo de inmunofluorescencia, empleando un anticuerpo anti-Coilin como marcador de los cuerpos de Cajal, por ser la proteína principal de estas estructuras, así como un anticuerpo específico anti-SMN1. Los resultados publicados en dicho trabajo (Smigiel et al., 2018) muestran una reducción de la localización de SMN1 en los cuerpos de Cajal en las células de paciente frente a fibroblastos control. Además, la señal de SMN1 en fibroblastos de paciente que sí acumulaban SMN en los CBs era menos intensa, y aumentaba su señal nucleoplásmica. No obstante, la señal de Coilin era comparable entre las muestras, indicando que los cuerpos de Cajal no se veían afectados en las células de paciente. En definitiva, estos datos parecían mostrar que las variantes en heterocigosis compuesta en PTPN23 evitaban la correcta acumulación de SMN en los CBs.

Para comprobar si ocurría lo mismo con las variantes que presentaba la paciente MD-172, se llevó a cabo un protocolo de inmunofluorescencia similar al descrito por los autores (Smigiel et al., 2018). No obstante, no se pudieron observar correctamente los cuerpos de Cajal ni en fibroblastos control ni afectados, ya que el anticuerpo anti-Coilin no marcaba puntos definidos localizados en el núcleo, sino que más bien presentaba un patrón difuso. Lo mismo ocurría con SMN1, difuso tanto en el núcleo como en el citoplasma (**Figura 12A-B**). Para comprobar que no había habido ningún fallo técnico en el protocolo, se realizó otra IFI con los mismos anticuerpos empleando como control positivo células HeLa, dado que en bibliografía se corroboró la presencia de SMN1 y Coilin en el patrón esperado en dicha línea celular (Hebert et al., 2001). En este caso se pudieron observar las proteínas de interés marcadas como puntos localizados en el núcleo (**Figura 12C**), lo cual verificó que no se trataba de un fallo técnico y que los anticuerpos empleados funcionaban correctamente.

Esto nos llevó a plantearnos que quizás estas estructuras son más frecuentes en células con una elevada actividad proliferativa y transcripcional, características que no son inherentes a los fibroblastos primarios. Para comprobarlo, se crecieron células NHF suplementando el medio descrito con un 20% de suero bovino fetal (FBS), lo cual induce una aceleración del crecimiento celular gracias a los factores presentes en el mismo y potencialmente de la maquinaria transcripcional celular. Otra aproximación consistió en suplementar el medio con el factor de crecimiento fibroblástico 2 (FGF2, 50ng/μL) con el mismo objetivo. Se probaron diferentes tiempos de incubación (0, 3, 6 y 24 horas), observándose un incremento en la intensidad de Coilin únicamente tras 3h de exposición al FGF2 (**Figura Suplementaria 8**), aunque el patrón seguía siendo difuso y no punteado en el núcleo y no se vio colocalización de SMN1 y Coilin. También se estudió por inmunofluorescencia una línea celular de fibroblastos inmortalizada con el virus del papiloma humano (NHF_E6E7) para comprobar si su mayor capacidad replicativa afectaba a la localización de SMN1/Coilin. De nuevo, no se detectó ninguna diferencia respecto a experimentos anteriores. Los controles negativos y las imágenes de otros experimentos realizados se encuentran en la **Figura Suplementaria 8**.

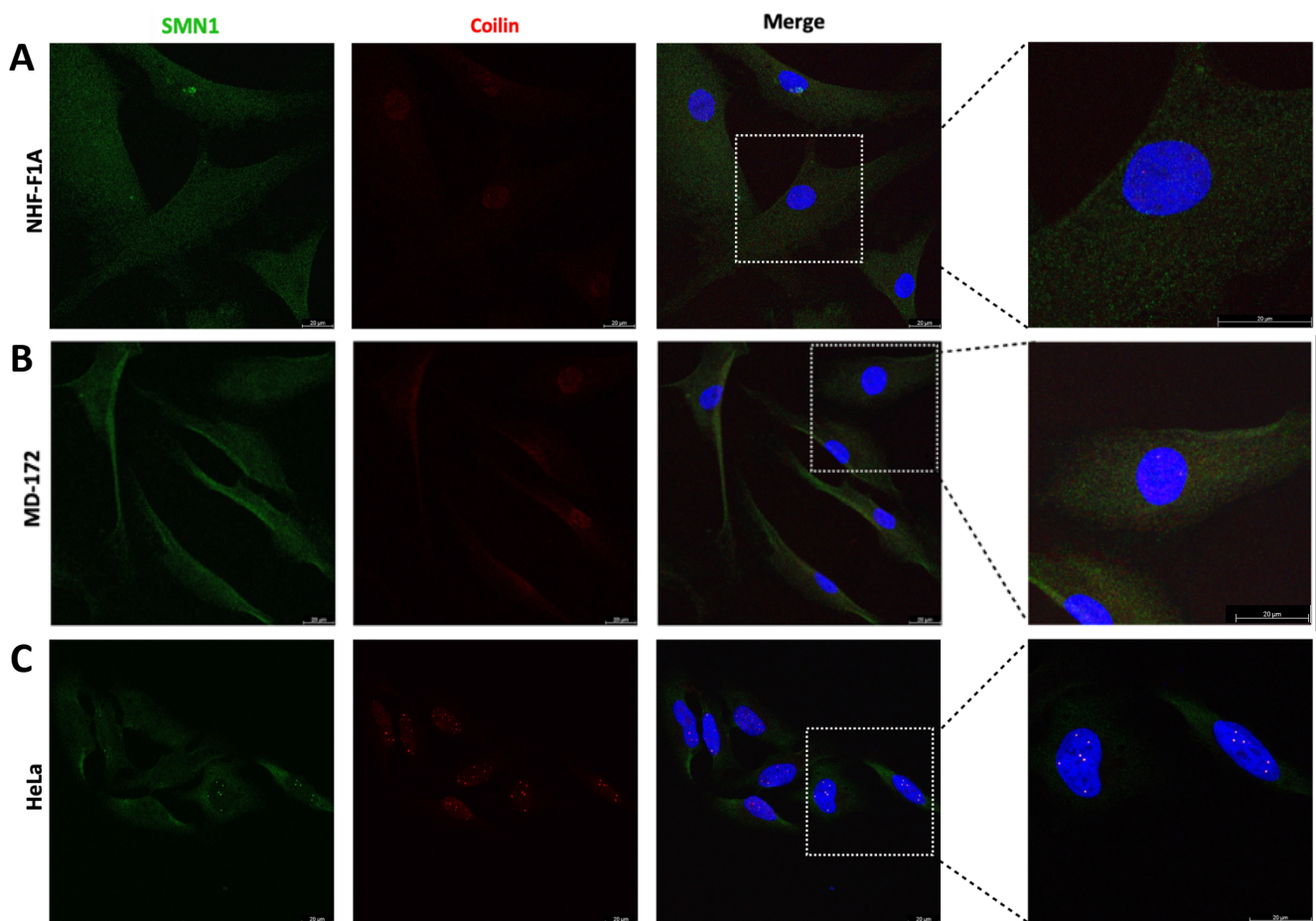


Figura 12. Imágenes representativas de los distintos experimentos de inmunofluorescencia. Se observan fibroblastos control (NHF-F1A) **(A)**, fibroblastos de paciente (MD-172) **(B)** y células HeLa **(C)**, con anticuerpos anti-SMN1 (verde) y anti-Coilin (rojo), este último empleado para marcar los cuerpos de Cajal (CBs) y detectar la posible acumulación de SMN1 en los mismos. Los núcleos celulares se tiñen con DAPI (azul). Barra de escala: 20μm. El recuadro indica la zona en la que se hace zoom.

5. DISCUSIÓN

5.1. VALIDACIÓN DE CAMBIOS CANDIDATOS EN LAS FAMILIAS

Uno de los objetivos de este estudio ha sido identificar las bases genéticas de la enfermedad que presentaba el probando afectado por un trastorno del movimiento en dos familias diferentes. Para ello, se analizaron los datos de secuenciación de exoma completo (WES) mediante el filtrado y priorización de variantes, considerando la información disponible en diferentes bases de datos y plataformas, como RD-Connect. Se analizó toda la región codificante del genoma (que supone alrededor del 2% del genoma completo) mediante un estudio mucho más amplio que el que permitiría realizar un panel de genes, pudiendo incluso asociar nuevos genes a una determinada enfermedad. Sin embargo, el coste en términos económicos y de procesamiento de datos es mayor.

Los pacientes estudiados en el presente Trabajo de Fin de Grado presentaban fenotipos clínicos de paresia-hipotonía, corea-distonía y ataxia, todos ellos agrupados dentro de los trastornos del movimiento, un conjunto de enfermedades raras muy heterogéneas por el elevado número de genes asociados a estos trastornos (Abdo et al., 2010).

En la **familia fMD275**, de los cambios encontrados en los genes *CACNA1A*, *CROCC*, *CACNA1G*, *TAF4* y *ATAD3A*, solo se validaron positivamente los dos primeros mientras que el resto no pudieron corroborarse. El cambio en *CACNA1A* resultó bastante interesante porque el gen codifica para un canal de calcio y el paciente sufre una canalopatía, pese a que la frecuencia de la variante es algo elevada para tratarse de una enfermedad rara y, además, la delección no cambia la pauta de lectura. Por otra parte, se trata de un gen ya asociado con enfermedad (ataxia episódica o espinocerebelosa), que se encuentra en varios paneles de genes relacionados con trastornos del movimiento, aunque bases de datos como Varsome o ClinVar clasifican a la variante del paciente (c.7266_7271delCAGCGG) como VUS (variante de significado incierto), por lo que se necesitarían estudios adicionales para relacionarla con el fenotipo del probando. *CROCC* es un gen algo desconocido para el cual se consiguió validar un cambio *de novo* en heterocigosis. La variante sinónima c.12G>A no parecía deletérea según los predictores empleados, aunque, considerando que el programa XHMM señalaba que el probando podría presentar una delección en el mismo gen que afectaría a los exones 1-19 en el otro alelo, no se podía descartar la patogenicidad de dichos cambios. Se debería validar la delección por otro método, como el MLPA (*Multiplex Ligation dependent Probe Amplification*), que es la técnica *gold standard* en estos casos. Al no existir sondas para el gen *CROCC* se podría optar por un array para la validación. No obstante, la delección predicha se encuentra en una región relativamente polimórfica donde ya se han descrito más inserciones/delecciones. En definitiva, los cambios se deberían reevaluar más adelante cuando se conozca más información sobre el gen. En cuanto al resto de genes, no se obtuvieron resultados que ligaran los cambios detectados con la enfermedad del probando en el proceso de validación. *CACNA1G* resultó ser un falso positivo, lo que significa que el paciente no presentaba el cambio detectado, sino que se trataba de un artefacto de la tecnología de secuenciación de exoma. *TAF4* y *ATAD3A* no pudieron amplificarse/secuenciarse por las características de la región, y deberían probarse condiciones de amplificación más restrictivas (aumentando la temperatura) y empleando una polimerasa de mayor fidelidad (Riet et al., 2017), al menos en el caso de *TAF4* ya que ni con DMSO, magnesio ni con ambos se consiguió amplificar la región. En el caso de *ATAD3A* se podrían diseñar nuevos cebadores, más cercanos a la zona que presenta la variante, para evitar la caída de la polimerasa al secuenciar el producto amplificado.

En la **familia fMD279** se consiguieron validar los tres cambios en los tres genes candidatos (*SLC38A6*, *SLITRK3* y *ASTN1*) con resultado positivo. De ellos, el más interesante fue el cambio c.690+2T>G en el gen *SLC38A6* que, aunque no se había asociado a enfermedad, al encontrarse en homocigosis y predecirse la afectación en el proceso de *splicing* con SpliceAI se decidió realizar

estudios funcionales para comprobar su patogenicidad. Asimismo, la proteína codificada por el gen participa en el ciclo glutamato-glutamina en el sistema nervioso central (Bagchi et al., 2014), una vía que se ve afectada en diversas enfermedades neurodegenerativas (Andersen et al., 2021), causando excitotoxicidad o toxicidad neuronal por la acumulación del neurotransmisor glutamato ante la disfunción de alguno de los componentes de la ruta. En relación con *SLITRK3* y *ASTN1*, son genes que codifican para moléculas importantes en el sistema nervioso central, cuya validación deja abierta la puerta a nuevos estudios experimentales tras reevaluar el exoma más adelante.

5.2. CARACTERIZACIÓN DE LA VARIANTE C.690+2T>G EN EL GEN *SLC386*

Tanto en el ensayo de minigenes como en el análisis de transcritos en sangre periférica se comprobó que la variante c.690+2T>G genera un salto de exón debido a la pérdida de un sitio dador de *splicing* (+2) que resulta en la pérdida completa del exón 9 del gen (66 nucleótidos) dentro de la pauta de lectura. A nivel proteico, supondría la pérdida de 22 aminoácidos (p.V209_Q230del), y coincide con el final de una de las regiones hélice transmembrana de la proteína y el inicio de una posible región intracelular o extracelular, según AlphaFold. Además, comprende una región que establece puentes disulfuro (C218-C238), de acuerdo con UniProt. La pérdida de la cisteína 218, junto el resto de aminoácidos hasta el 230, imposibilitaría la formación del mencionado puente disulfuro en la proteína, lo que podría desestabilizar completamente su estructura tridimensional, al tratarse un enlace covalente fundamental para mantener la estabilidad de la estructura terciaria de las proteínas (Karimi et al., 2016). De esta manera, la función del transportador de aminoácidos SNAT6 (codificado por el gen *SLC38A6*) podría verse alterada, comprometiendo su actividad transportadora de glutamina al interior de la neurona presináptica y disruptiendo la homeostasis del ciclo glutamato-glutamina. Además, canónicamente, existen dos isoformas de *SLC38A6*, una que codifica para una proteína de 456 aminoácidos (siendo esta la secuencia de referencia, validada en RefSeq) y otra de 521 aminoácidos; y ambas incluyen el exón 9. Esto podría indicar que la pérdida del exón 9 no es un evento que ocurre de manera fisiológica en nuestro organismo, apuntando a la potencial patogenicidad de esta variante. Se trata de un mecanismo mutacional que supone la eliminación de un exón tras perder un sitio canónico de *splicing*, el cual ha sido descrito y relacionado con enfermedad anteriormente (Daich Varela et al., 2022; Quental et al., 2023; Sewerin et al., 2024). En definitiva, los ensayos experimentales y la literatura disponible al respecto respaldan la naturaleza patogénica de la variante c.690+2T>G en el gen *SLC38A6*. Por un lado, se trata de un cambio nucleotídico muy interesante en base a los criterios de filtrado e interpretación de variantes: no está en la base de datos gnomAD, es decir, es novel, y está en homocigosis en el paciente. Además, podría producir la pérdida de función de la proteína SNAT6 debido al salto del exón 9 en el ARNm del gen *SLC38A6* (Richards et al., 2015).

Para confirmar la implicación del gen en la enfermedad del probando MD-626 sería necesario estudiar casos de familias con un fenotipo similar y variantes en *SLC38A6*. No obstante, esto supone un gran reto por tratarse de enfermedades raras puesto que, por definición, tienen muy pocos pacientes. Asimismo, se deberían realizar estudios adicionales sobre las células del paciente, idealmente, procedentes de tejido nervioso. Sin embargo, por la dificultad de acceso a dicho tejido, se podrían caracterizar otras células que pudieran verse afectadas, como fibroblastos, los cuales también podrían derivarse a células nerviosas. Al tratarse de un gen no asociado previamente a enfermedad, no bastaría con observar alteraciones o ausencia de la proteína en células, sino que se debería estudiar el efecto funcional de revertir el fenotipo del paciente en el modelo celular, por ejemplo, corrigiendo la mutación mediante la técnica CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*). Finalmente, para comprobar que los efectos observados se deben a los cambios nucleotídicos presentes en homocigosis (c.690+2T>G) en el gen *SLC38A6* se podrían establecer modelos animales (en *Caenorhabditis elegans*, *Drosophila melanogaster* o *Mus musculus*, entre otros) o celulares con el gen silenciado o mutado (Yang et al., 2022).

5.3. ESTUDIO DE LAS PROTEÍNAS PTPN23 Y SMN1 EN FIBROBLASTOS DE PACIENTE

Variantes *missense* en heterocigosis compuesta en *PTPN23*, como la que presenta la paciente MD-172, se han asociado con un trastorno del neurodesarrollo caracterizado por anomalías estructurales que cursa con o sin convulsiones y espasticidad (MIM #618890). Los resultados de los estudios funcionales realizados en este trabajo en las proteínas *PTPN23* y *SMN1* no permiten sacar conclusiones sobre la patogenicidad de las variantes que la paciente MD-172 presenta. Por tanto, se deberían realizar más estudios para ganar evidencia de la patogenicidad de los cambios p.V133M y p.R1345H en la proteína *PTPN23*.

5.3.1. Estudio de la expresión de SMN1 mediante *western blot*

Basándonos en un trabajo previo publicado por Smigiel et al. (2018), tratamos de corroborar si veíamos diferencias en la presencia de una forma de *SMN1* que migra más lento en SDS-PAGE, posiblemente una forma fosforilada, entre extractos proteicos de fibroblastos de pacientes y muestras control. Sin embargo, a pesar de que observamos una doble banda de *SMN1* similar a la que describen en dicha publicación, no detectamos diferencias tan marcadas como ellos en la intensidad de las bandas. Esta diferencia de resultados podría deberse a que las variantes presentadas por los pacientes no son comparables. En nuestro caso, se trata de dos cambios *missense*, mientras que en el estudio mencionado se trata de uno *missense* y otro *nonsense*, el cual se predice que producirá pérdida de función al generar una proteína truncada o al degradarse el ARNm mediante *nonsense-mediated decay* (NMD), un mecanismo que evita la síntesis de proteínas con codones de parada prematuros. Para determinar si los resultados obtenidos se deben a la falta de similitud de los genotipos se podría realizar un inmunoblot a partir de extractos proteicos de una línea celular *knockdown/knockout* en *PTPN23* o con el mismo tipo de variantes (*missense* y *nonsense*) en heterocigosis compuesta generada, por ejemplo, mediante la técnica CRISPR-Cas9.

Por otro lado, observamos que existe una mayor expresión de *SMN1* en MD-172 respecto a los fibroblastos control NHF-F1A. Esto podría explicarse por múltiples factores. En primer lugar, la técnica *western blot* no tiene una buena reproducibilidad, por lo que quizá podría ser artefactual. Además, el número de réplicas debería ser mucho mayor para poder sacar conclusiones, pues contamos con 2 réplicas para el control sano (NHF) y 3 para el paciente. Asimismo, podría deberse a que, debido a la mutación en *PTPN23*, el paciente presente mayor cantidad de *SMN1* en citoplasma que el control sano, provocando que la extracción de *SMN1* en control sea menos eficiente al encontrarse en núcleo. No obstante, se supone que el tampón de lisis empleado (RIPA) tiene buena eficiencia para la extracción de proteínas nucleares. Para corroborar esta hipótesis se podría hacer una separación de las proteínas nucleares y citoplásmicas mediante centrifugación diferencial, y comparar así la intensidad en cada fracción celular. Esta hipótesis está fundamentada en que se ha descrito que *PTPN23* estabiliza la fosforilación de tirosinas (Y109, Y126, Y130) en *SMN* y la pérdida de estas fosforilaciones impide la acumulación nuclear de *SMN* en los cuerpos de Cajal (Gruss et al., 2017). De manera que las mutaciones en *PTPN23* podrían producir la deslocalización de *SMN* del núcleo, aunque incrementaran su presencia en citoplasma.

5.3.2. Estudio de la localización subcelular de SMN1

De nuevo, basándonos en el trabajo de Smigiel et al. (2018) tratamos de estudiar si las mutaciones en *PTPN23* podían provocar la deslocalización de *SMN1* de los cuerpos de Cajal (CBs) en fibroblastos de paciente mediante inmunofluorescencia y microscopía confocal. Al contrario que los autores de la citada publicación, no fuimos capaces de observar, ya no solo una deslocalización, sino que tampoco llegamos a visualizar con claridad los cuerpos de Cajal (localizados a través de la proteína Coilin) en los fibroblastos, ni de paciente ni de donante sano. Sin embargo, logramos

detectar tanto los CBs como SMN1 y su colocación en el núcleo de células HeLa, lo que indicaba que la técnica y los anticuerpos contra Coilin y SMN1 funcionaban correctamente.

Se probaron varios tratamientos en células de donante sano (NHF) para intentar favorecer la localización de SMN1 y Coilin en el patrón punteado que correspondía en el núcleo, es decir, en la estructura de los cuerpos de Cajal; consistentes en incrementar su actividad replicativa y transcripcional. Sin embargo, ninguno de ellos consiguió dicho propósito. Por tanto, postulamos que en el caso de los fibroblastos quizás solo ciertas células, por su estado metabólico/replicativo presentarían dichas estructuras y podría resultar muy difícil observarlas. Esto queda respaldado por publicaciones anteriores (Carvalho et al., 1999; Sleeman et al., 2003), que describen que los cuerpos de Cajal solo se localizan en gran número en el núcleo entre el final de la telofase y la fase G1 temprana del ciclo celular, no así durante la mitosis (fase M); y SMN migra al núcleo en etapas más tardías de la fase G1. Entre la profase y la telofase, SMN y Coilin se encuentran en los denominados cuerpos de Cajal mitóticos (MCBs), que son estructuras pequeñas y poco numerosas (Sleeman et al., 2003). Por tanto, quizá las células observadas no estaban en la fase adecuada para observar los CBs. Para corroborarlo, se podría hacer una búsqueda más exhaustiva de células con localización nuclear de Coilin y SMN1, o bien se podría realizar un bloqueo con timidina para sincronizar las células en la interfase (G1/S/G2) (Wang R. & Wang Z., 2022) y ver si entonces Coilin y SMN se localizan en los cuerpos de Cajal.

Por otra parte, el modelo celular empleado puede ser uno de los problemas, puesto que los cuerpos de Cajal sí se observan en células HeLa, que tienen una mayor tasa proliferativa. De hecho, Carvalho et al. (1999) citando publicaciones anteriores mencionan que los fibroblastos primarios son células que normalmente presentan cuerpos de Cajal vacíos, aunque SMN sí se localiza en forma de *foci* en el núcleo. No obstante, Smigiel et al. (2018) sí fueron capaces de observar CBs, a través de Coilin, en fibroblastos sanos y de paciente. Se ha descrito (Novotný et al., 2015) que la inhibición de proteínas implicadas en el ensamblaje de las snRNP (*small nuclear ribonucleoproteins*) induce la formación de CBs en fibroblastos procedentes de tejido pulmonar (WI-38). Por tanto, se podríamos inhibir dichas proteínas en nuestros fibroblastos para ver si se induce la formación de cuerpos de Cajal, lo cual indicaría que el modelo celular del que disponemos no sería el más adecuado en este caso.

En definitiva, el mecanismo patogénico para las variantes que alteran la función de la proteína PTPN23 sugerido en el artículo de Smigiel et al. (2018), que implica una localización y fosforilación anormales de SMN1, no se ha podido comprobar en la paciente MD-172. No obstante, las múltiples vías en las que participa PTPN23 indican que el fenotipo clínico observado en la paciente puede darse por mecanismos diferentes a la vía de SMN1, aunque con los estudios realizados no se puede descartar la afectación de SMN1 por falta de resultados concluyentes.

6. CONCLUSIONES

1. El uso de WES permite el diagnóstico de pacientes con enfermedades raras carentes de diagnóstico por otros métodos, como los paneles de genes. A pesar de la gran complejidad que supone el análisis de los datos, la posterior identificación de variantes candidatas y consiguientes estudios funcionales, el resultado final puede ser la asociación de una variante con el fenotipo patológico del paciente.
2. Se han detectado variantes candidatas en las dos familias analizadas en los genes *ATAD3A*, *CACNA1A*, *CACNA1G*, *CROCC*, *TAF4*, *SLC38A6*, *SLITRK3* y *ASTN1*. De ellos, tras la validación por secuenciación Sanger, se ha descartado la variante en *CACNA1G* en la familia fMD275 por tratarse de un artefacto de la tecnología de secuenciación de exoma.
3. En el probando MD-626 se ha estudiado el efecto de la variante *SLC38A6* c.690+2T>G en homocigosis sobre el *splicing*.
4. En la probando MD-618 se han seleccionado cinco variantes candidatas, de las cuales una resultó ser un falso positivo y dos no pudieron validarse. De las tres variantes restantes no se hicieron estudios funcionales por falta de evidencia de patogenicidad y/o de relación con el fenotipo clínico de la paciente. No obstante, se debe continuar su estudio más adelante mediante WES-dinámico o secuenciación de genoma completo (WGS).
5. Se necesitarían hacer estudios funcionales sobre las variantes candidatas no estudiadas para determinar su patogenicidad y relación con el fenotipo de los probandos.
6. Mediante el ensayo de minigenes y el análisis de transcritos en sangre periférica se ha confirmado que la variante c.690+2T>G en el gen *SLC38A6*, presente en homocigosis en el probando MD-626, genera un patrón de *splicing* anómalo al producir el salto del exón 9 (66 nucleótidos) debido a la pérdida de un sitio dador de *splicing* canónico.
7. Se han realizado estudios funcionales de localización y expresión de la proteína SMN1 a partir de fibroblastos primarios de la paciente MD-172, que presenta dos cambios en heterocigosis compuesta en el gen que codifica para la proteína PTPN23 (p.V133M y p.R1345H). No se han observado diferencias entre células control y de paciente en cuanto a la localización celular de las proteínas estudiadas. Por ello, es necesario realizar más estudios para averiguar el mecanismo patogénico de las variantes en PTPN23.
8. Los motivos de las diferencias de expresión de SMN1 encontradas entre fibroblastos control sanos y del probando MD-172 no han podido determinarse. Son necesarios más estudios, empleando diferentes técnicas y/o más réplicas y muestras.

7. BIBLIOGRAFÍA

1. ABDO, W. F., VAN DE WARRENBURG, B. P. C., BURN, D. J., QUINN, N. P., & BLOEM, B. R. (2010). The clinical approach to movement disorders. *Nature Reviews Neurology*, 6(1), 29–37. <https://doi.org/10.1038/nrneurol.2009.196>
2. ABOZAID, G. M., KERR, K., MCKNIGHT, A., & AL-OMAR, H. A. (2022). Criteria to define rare diseases and orphan drugs: a systematic review protocol. *BMJ Open*, 12(7), e062126. <https://doi.org/10.1136/bmjopen-2022-062126>
3. ANDERSEN, J. V., MARKUSSEN, K. H., JAKOBSEN, E., SCHOUSBOE, A., WAAGEPETERSEN, H. S., ROSENBERG, P. A., & ALDANA, B. I. (2021). Glutamate metabolism and recycling at the excitatory synapse in health and neurodegeneration. *Neuropharmacology*, 196, 108719. <https://doi.org/10.1016/j.neuropharm.2021.108719>
4. ANNA, A., & MONIKA, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of Applied Genetics*, 59(3), 253–268. <https://doi.org/10.1007/s13353-018-0444-7>
5. APARISI, M. J., GARCÍA-GARCÍA, G., ALLER, E., SEQUEDO, M. D., MARTÍNEZ-FERNÁNDEZ DE LA CÁMARA, C., RODRIGO, R., ARMENGOT, M., CORTIJO, J., MILARA, J., DÍAZ-LLOPIS, M., JAIJO, T., & MILLÁN, J. M. (2013). Study of USH1 Splicing Variants through Minigenes and Transcript Analysis from Nasal Epithelial Cells. *PLoS ONE*, 8(2), e57506. <https://doi.org/10.1371/journal.pone.0057506>
6. AVILÉS-OLMOS, I., VILLINO, R., MARTÍN-BASTIDA, A., ESPINOZA, C., & RODRÍGUEZ-OROZ, M. C. (2023). Trastornos del movimiento. *Medicine - Programa de Formación Médica Continuada Acreditado*, 13(76), 4481–4490. <https://doi.org/10.1016/j.med.2023.03.019>
7. BAGCHI, S., BAOMAR, H. A., AL-WALAI, S., AL-SADI, S., & FREDRIKSSON, R. (2014). Histological Analysis of SLC38A6 (SNAT6) Expression in Mouse Brain Shows Selective Expression in Excitatory Neurons with High Expression in the Synapses. *PLoS ONE*, 9(4), e95438. <https://doi.org/10.1371/journal.pone.0095438>
8. BELKADI, A., BOLZE, A., ITAN, Y., COBAT, A., VINCENT, Q. B., ANTIPENKO, A., SHANG, L., BOISSON, B., CASANOVA, J.-L., & ABEL, L. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences*, 112(17), 5473–5478. <https://doi.org/10.1073/pnas.1418631112>
9. BEND, R., COHEN, L., CARTER, M. T., LYONS, M. J., NIYAZOV, D., MIKATI, M. A., ROJAS, S. K., PERSON, R. E., SI, Y., WENTZENSEN, I. M., TORTI, E., LEE, J. A., BOYCOTT, K. M., BASEL-SALMON, L., FERREIRA, C. R., & GONZAGA-JAUREGUI, C. (2020). Phenotype and mutation expansion of the PTPN23 associated disorder characterized by neurodevelopmental delay and structural brain abnormalities. *European Journal of Human Genetics*, 28(1), 76–87. <https://doi.org/10.1038/s41431-019-0487-1>
10. BENITO-LOZANO, J., LÓPEZ-VILLALBA, B., ARIAS-MERINO, G., POSADA DE LA PAZ, M., & ALONSO-FERREIRA, V. (2022). Diagnostic delay in rare diseases: data from the Spanish rare diseases patient registry. *Orphanet Journal of Rare Diseases*, 17(1), 418. <https://doi.org/10.1186/s13023-022-02530-3>
11. CARVALHO, T., ALMEIDA, F., CALAPEZ, A., LAFARGA, M., BERCIANO, M. T., & CARMO-FONSECA, M. (1999). The Spinal Muscular Atrophy Disease Gene Product, Smn. *The Journal of Cell Biology*, 147(4), 715–728. <https://doi.org/10.1083/jcb.147.4.715>

12. CHANG, H. (2017). Cleave but not leave: Astrotactin proteins in development and disease. *IUBMB Life*, 69(8), 572–577. <https://doi.org/10.1002/iub.1641>
13. CHUNG, B. H. Y., CHAU, J. F. T., & WONG, G. K.-S. (2021). Rare versus common diseases: a false dichotomy in precision medicine. *Npj Genomic Medicine*, 6(1), 19. <https://doi.org/10.1038/s41525-021-00176-x>
14. DAICH VARELA, M., MOTTA, F. L., WEBSTER, A. R., & ARNO, G. (2022). A rare canonical splice-site variant in VPS13B causes attenuated Cohen syndrome. *Ophthalmic Genetics*, 43(1), 110–115. <https://doi.org/10.1080/13816810.2021.1970194>
15. DESVIAT, L. R., PÉREZ, B., & UGARTE, M. (2012). *Minigenes to Confirm Exon Skipping Mutations* (pp. 37–47). https://doi.org/10.1007/978-1-61779-767-5_3
16. DI FONZO, A., MONFRINI, E., & ERRO, R. (2018). Genetics of Movement Disorders and the Practicing Clinician; Who and What to Test for? *Current Neurology and Neuroscience Reports*, 18(7), 37. <https://doi.org/10.1007/s11910-018-0847-1>
17. EFTHYMIOU, S., HAN, W., ILYAS, M., LI, J., YU, Y., SCALA, M., MALINTAN, N. T., ILYAS, M., VAVOURAKI, N., MANKAD, K., MAROOFIAN, R., ROCCA, C., SALPIETRO, V., LAKHANI, S., MALLACK, E. J., PALCULICT, T. B., LI, H., ZHANG, G., ZAFAR, F., ... HOULDEN, H. (2024). Human mutations in SLITRK3 implicated in GABAergic synapse development in mice. *Frontiers in Molecular Neuroscience*, 17. <https://doi.org/10.3389/fnmol.2024.1222935>
18. FERNANDEZ-MARMIESSE, A., GOUVEIA, S., & COUCE, M. L. (2018). NGS Technologies as a Turning Point in Rare Disease Research , Diagnosis and Treatment. *Current Medicinal Chemistry*, 25(3), 404–432. <https://doi.org/10.2174/0929867324666170718101946>
19. GANDASI, N. R., ARAPI, V., MICKAEL, M. E., BELEKAR, P. A., GRANLUND, L., KOTHEGALA, L., FREDRIKSSON, R., & BAGCHI, S. (2021). Glutamine Uptake via SNAT6 and Caveolin Regulates Glutamine–Glutamate Cycle. *International Journal of Molecular Sciences*, 22(3), 1167. <https://doi.org/10.3390/ijms22031167>
20. GONZAGA-JAUREGUI, C. (2021). Challenges and opportunities in rare diseases research. In *Genomics of Rare Diseases* (pp. 263–284). Elsevier. <https://doi.org/10.1016/B978-0-12-820140-4.00013-2>
21. GONZAGA-JAUREGUI, C., & ZEPEDA MENDOZA, C. J. (2021). Genomic sequencing of rare diseases. In *Genomics of Rare Diseases* (pp. 61–95). Elsevier. <https://doi.org/10.1016/B978-0-12-820140-4.00001-6>
22. GRUSS, O. J., MEDURI, R., SCHILLING, M., & FISCHER, U. (2017). UsnRNP biogenesis: mechanisms and regulation. *Chromosoma*, 126(5), 577–593. <https://doi.org/10.1007/s00412-017-0637-6>
23. HAENDEL, M., VASILEVSKY, N., UNNI, D., BOLOGA, C., HARRIS, N., REHM, H., HAMOSH, A., BAYNAM, G., GROZA, T., MCMURRY, J., DAWKINS, H., RATH, A., THAXTON, C., BOCCI, G., JOACHIMIAK, M. P., KÖHLER, S., ROBINSON, P. N., MUNGALL, C., & OPREA, T. I. (2020). How many rare diseases are there? *Nature Reviews Drug Discovery*, 19(2), 77–78. <https://doi.org/10.1038/d41573-019-00180-y>
24. HAGEMANN, I. S. (2015). Overview of Technical Aspects and Chemistries of Next-Generation Sequencing. In *Clinical Genomics* (pp. 3–19). Elsevier. <https://doi.org/10.1016/B978-0-12-404748-8.00001-0>
25. HEBERT, M. D., SZYMCZYK, P. W., SHPARGEL, K. B., & MATERA, A. G. (2001). Coilin forms the bridge between Cajal bodies and SMN, the Spinal Muscular Atrophy protein. *Genes & Development*, 15(20), 2720–2729. <https://doi.org/10.1101/gad.908401>

26. HUSEDZINOVIC, A., NEUMANN, B., REYMAN, J., DRAEGER-MEURER, S., CHARI, A., ERFLE, H., FISCHER, U., & GRUSS, O. J. (2015). The catalytically inactive tyrosine phosphatase HD-PTP/PTPN23 is a novel regulator of SMN complex localization. *Molecular Biology of the Cell*, 26(2), 161–171. <https://doi.org/10.1091/mbc.E14-06-1151>
27. HUSHMANDI, K., EINOLLAHI, B., SAADAT, S. H., LEE, E. H. C., FARANI, M. R., OKINA, E., HUH, Y. S., NABAVI, N., SALIMIMOGHADAM, S., & KUMAR, A. P. (2024). Amino acid transporters within the solute carrier superfamily: Underappreciated proteins and novel opportunities for cancer therapy. *Molecular Metabolism*, 84, 101952. <https://doi.org/10.1016/j.molmet.2024.101952>
28. JANSSEN, B. D. E., VAN DEN BOOGAARD, M. H., LICHTENBELT, K., SEABY, E. G., STALS, K., ELLARD, S., NEWBURY-ECOB, R., DIXIT, A., ROHT, L., PAJUSALU, S., ÖUNAP, K., FIRTH, H. V., BUCKLEY, M., WILSON, M., ROSCIOLI, T., TIDWELL, T., MAO, R., ENNIS, S., HOLWERDA, S. J., ... VAN JAARSVELD, R. H. (2022). De novo putative loss-of-function variants in *TAF4* are associated with a neuro-developmental disorder. *Human Mutation*, 43(12), 1844–1851. <https://doi.org/10.1002/humu.24444>
29. JINNAH, H. A., ALBANESE, A., BHATIA, K. P., CARDOSO, F., DA PRAT, G., DE KONING, T. J., ESPAY, A. J., FUNG, V., GARCIA-RUIZ, P. J., GERSHANIK, O., JANKOVIC, J., KAJI, R., KOTSCHET, K., MARRAS, C., MIYASAKI, J. M., MORGANTE, F., MUNCHAU, A., PAL, P. K., RODRIGUEZ OROZ, M. C., ... GATTO, E. M. (2018). Treatable inherited rare movement disorders. *Movement Disorders*, 33(1), 21–35. <https://doi.org/10.1002/mds.27140>
30. KARIMI, M., IGNASIAK, M. T., CHAN, B., CROFT, A. K., RADOM, L., SCHIESSER, C. H., PATTISON, D. I., & DAVIES, M. J. (2016). Reactivity of disulfide bonds is markedly affected by structure and environment: implications for protein modification and stability. *Scientific Reports*, 6(1), 38572. <https://doi.org/10.1038/srep38572>
31. KUNII, M., DOI, H., HASHIGUCHI, S., MATSUSHI, T., SAKAI, Y., IAI, M., OKUBO, M., NAKAMURA, H., TAKAHASHI, K., KATSUMOTO, A., TADA, M., TAKEUCHI, H., ISHIKAWA, T., MIYAKE, N., SAITSU, H., MATSUMOTO, N., & TANAKA, F. (2020). De novo CACNA1G variants in developmental delay and early-onset epileptic encephalopathies. *Journal of the Neurological Sciences*, 416, 117047. <https://doi.org/10.1016/j.jns.2020.117047>
32. LAURIE, S., PISCIA, D., MATALONGA, L., CORVO, A., GARCIA, C., FERNANDEZ-CALLEJO, M., HERNANDEZ, C., LUENGO, C., NTALIS, A. P., PROTASSIO, J., MARTINEZ, I., PICO, D., THOMPSON, R., TONDA, R., BAYES, M., BULLICH, G., CAMPS, J., PARAMONOV, I., TROTTA, J., ... BELTRAN, S. (2022). The RD-Connect Genome-Phenome Analysis Platform: Accelerating diagnosis, research, and gene discovery for rare diseases. *Human Mutation*. <https://doi.org/10.1002/humu.24353>
33. LEE, C. E., SINGLETON, K. S., WALLIN, M., & FAUNDEZ, V. (2020). Rare Genetic Diseases: Nature's Experiments on Human Development. *IScience*, 23(5), 101123. <https://doi.org/10.1016/j.isci.2020.101123>
34. LINDBERG, F. A., NORDENANKAR, K., & FREDRIKSSON, R. (2022). SLC38A10 Knockout Mice Display a Decreased Body Weight and an Increased Risk-Taking Behavior in the Open Field Test. *Frontiers in Behavioral Neuroscience*, 16. <https://doi.org/10.3389/fnbeh.2022.840987>
35. LIPMAN, A. R., FAN, X., SHEN, Y., & CHUNG, W. K. (2022). Clinical and genetic characterization of CACNA1A-related disease. *Clinical Genetics*, 102(4), 288–295. <https://doi.org/10.1111/cge.14180>
36. LIU, Z., ZHU, L., ROBERTS, R., & TONG, W. (2019). Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We? *Trends in Genetics*, 35(11), 852–867. <https://doi.org/10.1016/j.tig.2019.08.006>

37. LOPES-JÚNIOR, L. C., FERRAZ, V. E. F., LIMA, R. A. G., SCHUAB, S. I. P. C., PESSANHA, R. M., LUZ, G. S., LAIGNIER, M. R., NUNES, K. Z., LOPES, A. B., GRASSI, J., MOREIRA, J. A., JARDIM, F. A., LEITE, F. M. C., FREITAS, P. DE S. S., & BERTOLINI, S. R. (2022). Health Policies for Rare Disease Patients: A Scoping Review. *International Journal of Environmental Research and Public Health*, 19(22), 15174. <https://doi.org/10.3390/ijerph192215174>
38. LUQUE, J., MENDES, I., GÓMEZ, B., MORTE, B., LÓPEZ DE HEREDIA, M., HERRERAS, E., CORROCHANO, V., BUEREN, J., GALLANO, P., ARTUCH, R., FILLAT, C., PÉREZ-JURADO, L. A., MONTOLIU, L., CARRACEDO, Á., MILLÁN, J. M., WEBB, S. M., PALAU, F., & LAPUNZINA, P. (2022). CIBERER: Spanish national network for research on rare diseases: A highly productive collaborative initiative. *Clinical Genetics*, 101(5–6), 481–493. <https://doi.org/10.1111/cge.14113>
39. MARWAHA, S., KNOWLES, J. W., & ASHLEY, E. A. (2022). A guide for the diagnosis of rare and undiagnosed disease: beyond the exome. *Genome Medicine*, 14(1), 23. <https://doi.org/10.1186/s13073-022-01026-w>
40. MONTAUT, S., TRANCHANT, C., DROUOT, N., RUDOLF, G., GUISSART, C., TARABEUX, J., STEMMELEN, T., VELT, A., FOURRAGE, C., NITSCHKÉ, P., GERARD, B., MANDEL, J.-L., KOENIG, M., CHELLY, J., & ANHEIM, M. (2018). Assessment of a Targeted Gene Panel for Identification of Genes Associated With Movement Disorders. *JAMA Neurology*, 75(10), 1234. <https://doi.org/10.1001/jamaneurol.2018.1478>
41. NOVOTNÝ, I., MALINOVÁ, A., STEJSKALOVÁ, E., MATĚJŮ, D., KLIMEŠOVÁ, K., ROITHOVÁ, A., ŠVÉDA, M., KNEJZLÍK, Z., & STANĚK, D. (2015). SART3-Dependent Accumulation of Incomplete Spliceosomal snRNPs in Cajal Bodies. *Cell Reports*, 10(3), 429–440. <https://doi.org/10.1016/j.celrep.2014.12.030>
42. NG, S. B., BUCKINGHAM, K. J., LEE, C., BIGHAM, A. W., TABOR, H. K., DENT, K. M., HUFF, C. D., SHANNON, P. T., JABS, E. W., NICKERSON, D. A., SHENDURE, J., & BAMSHAD, M. J. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*, 42(1), 30–35. <https://doi.org/10.1038/ng.499>
43. NGUENGANG WAKAP, S., LAMBERT, D. M., OLRÝ, A., RODWELL, C., GUEYDAN, C., LANNEAU, V., MURPHY, D., LE CAM, Y., & RATH, A. (2020). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics*, 28(2), 165–173. <https://doi.org/10.1038/s41431-019-0508-0>
44. OLGATI, S., QUADRI, M., & BONIFATI, V. (2016). Genetics of movement disorders in the next-generation sequencing era. *Movement Disorders*, 31(4), 458–470. <https://doi.org/10.1002/mds.26521>
45. QUENTAL, R., SAMPAIO, M., ALONSO, I., QUENTAL, S., LEÃO, M., & SOUSA, R. (2023). A Novel Homozygous Splice Site Variant in AIMP1 Gene Causing Hypomyelinating Leukodystrophy: Case Report and Review of the Literature. *Neuropediatrics*, 54(02), 120–125. <https://doi.org/10.1055/s-0042-1760366>
46. RABBANI, B., TEKIN, M., & MAHDIEH, N. (2014). The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*, 59(1), 5–15. <https://doi.org/10.1038/jhg.2013.114>
47. RAMOS-FUENTES, F., GONZÁLEZ-MENESES, A., ARS, E., & HERNÁNDEZ-JARAS, J. (2020). Genetic Diagnosis of Rare Diseases: Past and Present. *Advances in Therapy*, 37(S2), 29–37. <https://doi.org/10.1007/s12325-019-01176-1>
48. RICHARDS, S., AZIZ, N., BALE, S., BICK, D., DAS, S., GASTIER-FOSTER, J., GRODY, W. W., HEGDE, M., LYON, E., SPECTOR, E., VOELKERDING, K., & REHM, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American

- College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>
49. RIERA-MESTRE, A. (2022). Rare diseases in Spain: a look into the future. *Medicina Clínica (English Edition)*, 158(6), 274–276. <https://doi.org/10.1016/j.medcle.2021.10.008>
 50. RIET, J., RAMOS, L. R. V., LEWIS, R. V., & MARINS, L. F. (2017). Research Article Improving the PCR protocol to amplify a repetitive DNA sequence. *Genetics and Molecular Research*, 16(3). <https://doi.org/10.4238/gmr16039796>
 51. SCHULER, B. A., NELSON, E. T., KOZIURA, M., COGAN, J. D., HAMID, R., & PHILLIPS, J. A. (2022). Lessons learned: next-generation sequencing applied to undiagnosed genetic diseases. *Journal of Clinical Investigation*, 132(7). <https://doi.org/10.1172/JCI154942>
 52. SEABY, E. G., PENGELLY, R. J., & ENNIS, S. (2016). Exome sequencing explained: a practical guide to its clinical application. *Briefings in Functional Genomics*, 15(5), 374–384. <https://doi.org/10.1093/bfgp/elv054>
 53. SEWERIN, S., AURNHAMMER, C., SKUBIC, C., BLAGOTINŠEK COKAN, K., JERUC, J., ROZMAN, D., PFISTER, F., DITTRICH, K., MAYER, B., SCHÖNAUER, R., PETZOLD, F., & HALBRITTER, J. (2024). Mechanisms of pathogenicity and the quest for genetic modifiers of kidney disease in branchiootorenal syndrome. *Clinical Kidney Journal*, 17(1). <https://doi.org/10.1093/ckj/sfad260>
 54. SHEN, C.-H. (2023). Quantification and analysis of proteins. In *Diagnostic Molecular Biology* (pp. 231–257). Elsevier. <https://doi.org/10.1016/B978-0-323-91788-9.00002-8>
 55. SKOPKOVA, M., STUFKOVA, H., RAMBANI, V., STRANECKY, V., BRENNEROVA, K., KOLNIKOVA, M., PIETRZYKOVA, M., KARHANEK, M., NOSKOVA, L., TESAROVA, M., HANSIKOVA, H., & GASPERIKOVA, D. (2023). ATAD3A-related pontocerebellar hypoplasia: new patients and insights into phenotypic variability. *Orphanet Journal of Rare Diseases*, 18(1), 92. <https://doi.org/10.1186/s13023-023-02689-3>
 56. SLEEMAN, J. E., TRINKLE-MULCAHY, L., PRESCOTT, A. R., OGG, S. C., & LAMOND, A. I. (2003). Cajal body proteins SMN and Coilin show differential dynamic behaviour in vivo. *Journal of Cell Science*, 116(10), 2039–2050. <https://doi.org/10.1242/jcs.00400>
 57. SMIGIEL, R., LANDSBERG, G., SCHILLING, M., RYDZANICZ, M., POLLAK, A., WALCZAK, A., STODOLAK, A., STAWINSKI, P., MIERZEWSKA, H., SASIADEK, M. M., GRUSS, O. J., & PLOSKI, R. (2018). Developmental epileptic encephalopathy with hypomyelination and brain atrophy associated with PTPN23 variants affecting the assembly of UsnRNPs. *European Journal of Human Genetics*, 26(10), 1502–1511. <https://doi.org/10.1038/s41431-018-0179-2>
 58. SMITH, C. I. E., BERGMAN, P., & HAGEY, D. W. (2022). Estimating the number of diseases – the concept of rare, ultra-rare, and hyper-rare. *IScience*, 25(8), 104698. <https://doi.org/10.1016/j.isci.2022.104698>
 59. SOUCHE, E., BELTRAN, S., BROSENS, E., BELMONT, J. W., FOSSUM, M., RIESS, O., GILISSEN, C., ARDESHIRDAVANI, A., HOUGE, G., VAN GIJN, M., CLAYTON-SMITH, J., SYNOFZIK, M., DE LEEUW, N., DEANS, Z. C., DINCER, Y., ECK, S. H., VAN DER CRABBEN, S., BALASUBRAMANIAN, M., GRAESSNER, H., ... WEISS, M. M. (2022). Recommendations for whole genome sequencing in diagnostics for rare diseases. *European Journal of Human Genetics*, 30(9), 1017–1021. <https://doi.org/10.1038/s41431-022-01113-x>
 60. STRAITON, J., FREE, T., SAWYER, A., & MARTIN, J. (2019). From Sanger Sequencing to Genome Databases and Beyond. *BioTechniques*, 66(2), 60–63. <https://doi.org/10.2144/btn-2019-0011>
 61. SUNYAEV, S. R. (2012). Inferring causality and functional significance of human coding DNA variants. *Human Molecular Genetics*, 21(R1), R10–R17. <https://doi.org/10.1093/hmg/ddc385>

62. TAN, T. Y., LUNKE, S., CHONG, B., PHELAN, D., FANJUL-FERNANDEZ, M., MARUM, J. E., KUMAR, V. S., STARK, Z., YEUNG, A., BROWN, N. J., STUTTERD, C., DELATYCKI, M. B., SADEDIN, S., MARTYN, M., GORANITIS, I., THORNE, N., GAFF, C. L., & WHITE, S. M. (2019). A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis. *European Journal of Human Genetics*, 27(12), 1791–1799. <https://doi.org/10.1038/s41431-019-0471-9>
63. TESI, B., BOILEAU, C., BOYCOTT, K. M., CANAUD, G., CAULFIELD, M., CHOUKAIR, D., HILL, S., SPIELMANN, M., WEDELL, A., WIRTA, V., NORDGREN, A., & LINDSTRAND, A. (2023). Precision medicine in rare diseases: What is next? *Journal of Internal Medicine*, 294(4), 397–412. <https://doi.org/10.1111/joim.13655>
64. THENGANATT, M. A., & JANKOVIC, J. (2019). Psychogenic (Functional) Movement Disorders. *CONTINUUM: Lifelong Learning in Neurology*, 25(4), 1121–1140. <https://doi.org/10.1212/CON.0000000000000755>
65. TURN, R. E., LINNERT, J., GIGANTE, E. D., WOLFRUM, U., CASPARY, T., & KAHN, R. A. (2021). Roles for ELMOD2 and Rootletin in ciliogenesis. *Molecular Biology of the Cell*, 32(8), 800–822. <https://doi.org/10.1091/mbc.E20-10-0635>
66. UMLAI, U.-K. I., BANGARUSAMY, D. K., ESTIVILL, X., & JITHESH, P. V. (2022). Genome sequencing data analysis for rare disease gene discovery. *Briefings in Bioinformatics*, 23(1). <https://doi.org/10.1093/bib/bbab363>
67. VEERAMACHANENI, V. (2020). Data Analysis in Rare Disease Diagnostics. *Journal of the Indian Institute of Science*, 100(4), 733–751. <https://doi.org/10.1007/s41745-020-00189-y>
68. WANG, R. C., & WANG, Z. (2022). *Synchronization of Cultured Cells to G1, S, G2, and M Phases by Double Thymidine Block* (pp. 61–71). https://doi.org/10.1007/978-1-0716-2736-5_5
69. WRIGHT, C. F., FITZGERALD, T. W., JONES, W. D., CLAYTON, S., MCRAE, J. F., VAN KOGELBERG, M., KING, D. A., AMBRIDGE, K., BARRETT, D. M., BAYZETINOVA, T., BEVAN, A. P., BRAGIN, E., CHATZIMICHALI, E. A., GRIBBLE, S., JONES, P., KRISHNAPPA, N., MASON, L. E., MILLER, R., MORLEY, K. I., ... FIRTH, H. V. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *The Lancet*, 385(9975), 1305–1314. [https://doi.org/10.1016/S0140-6736\(14\)61705-0](https://doi.org/10.1016/S0140-6736(14)61705-0)
70. WRIGHT, C. F., FITZPATRICK, D. R., & FIRTH, H. V. (2018). Paediatric genomics: diagnosing rare disease in children. *Nature Reviews Genetics*, 19(5), 253–268. <https://doi.org/10.1038/nrg.2017.116>
71. XIA, R., PENG, H.-F., ZHANG, X., & ZHANG, H.-S. (2024). Comprehensive review of amino acid transporters as therapeutic targets. *International Journal of Biological Macromolecules*, 260, 129646. <https://doi.org/10.1016/j.ijbiomac.2024.129646>
72. YANG, S.-A., SALAZAR, J. L., LI-KROEGER, D., & YAMAMOTO, S. (2022). *Functional Studies of Genetic Variants Associated with Human Diseases in Notch Signaling-Related Genes Using Drosophila* (pp. 235–276). https://doi.org/10.1007/978-1-0716-2201-8_19

8. ANEXOS

8.1. ANEXO I: FIGURAS Y TABLAS SUPLEMENTARIAS

Tabla Suplementaria 1. Parejas de cebadores diseñadas, longitud de los amplicones (en pares de bases, pb) y temperatura de hibridación (Tm) para las amplificaciones por PCR.

Familia	Gen	Cebadores	Secuencia (5'-3')	Amplicón	Tm
fMD-275	<i>CACNA1G</i> (NM_018896.5)	CACNA1G_E27_D	TGGCGTCTAATGTAGGTGGG	617 pb	60°C
		CACNA1G_E27_R	CACATCAACTAACATCTCTGAGCC		
	<i>TAF4</i> (NM_003185.4)	TAF4_E1_D	GGGTCAGCGGCCAAC	176 pb	60°C*
		TAF4_E1_R	CGCCCTTGTCATAGTTG		
	<i>CROCC</i> (NM_014675.5)	CROCC_E1_D	ACTGAGCTAGTCTTGGGG	184 pb	58°C
		CROCC_E1_R	CCACTCCTGTTAAGTCCAC		
	<i>CACNA1A</i> (NM_023035.3)	CACNA1A_E47_D	CCGCACGTGTCCGAG	333 pb	58°C**
		CACNA1A_E47_R	CCAATCATCGTCACTCTC		
fMD-279	<i>SLC38A6</i> (NM_001172702.2)	SLC38A6_E9_D	CCTACCACAGGTGACCATG	544 pb	60°C
		SLC38A6_E9_R	ATCTGTTTAAGAGCATTGAGGG		
	<i>SLITRK3</i> (NM_001318810.2)	SLITRK3_E1_D	ATGTGCGCACTATTGAGCTG	512 pb	60°C
		SLITRK3_E1_R	TTCTTGGGCTGATGAAACAG		
	<i>ASTN1</i> (NM_004319.3)	ASTN1_E19_F	TCCTCTGACCCATGTAATATGC	657 pb	62°C
		ASTN1_E19_R	CAGGCAAAACTCTTAAATGGAC		
		ASTN1_E20_F	AGCTCTCAGGTGTGATGTGTATCC	521 pb	62°C
		ASTN1_E20_R	ACATTCTCTATAATGCCACCTGGG		

*No se consiguió amplificación a ninguna temperatura de las testadas (58-60-62°C), ni haciendo la PCR con MgCl₂, DMSO ni con ambos, quizá por la complejidad de la región, muy repetitiva y rica en GCs. **Para la amplificación de esta región se requirió 1μL de MgCl₂ (20mM) adicional.

TGCTGTTATGAACAATCCTAAGGTGAAGGCTCATGGCAAGAAGGTGCTGGCTGCCTTCAGTGAGGGTCTGAGTC
ACCTGGACAACCTCAAAGGCACCTTTTGCTAAGCTGAGTGAACCTGCTGCTGACAAGCTGCACGTCCTAGAGT
CGACCCAGCAGTAAAGTAATACATGTAATGCAACCTATACAAATAGCAATAGTAGCATTAGTAGCAATAATA
 ATAGCAATAGTTGTGTGGTCCATAGTAATCATAGAAATATAGGAAAAATTAAGACAAAGAAAAATAGACAGGTT
 AATTGATAGACTAATAGAAAGAGCAGAAGACAGTGGCAATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGA
 TGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTTGATGATCTGTAGTGTCTACAGAAAAATTTGGGGTCAC
 AGTCTATTATGGGGTACGGGATCACCAGAATTCTGGAGC **INSERTO SLC38A6**
 CTAGCAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGGGACCCAGAA
ATTGTAACGCACAGTTTTTAATTGTGGAGGGGAATTTTCTACTGTAATTCACACAACCTGTTTAATAGTACTTG
 GTTTAATAGTACTTGGAGTACTGAAGGTCAAATAACACTGAAGGAAGTGACACAATCACACTCCCATGCAGAA
 TAAAACAATTTATAAACATGTGGCAGGAAGTAGGAAAAGCAATGTATGCCCCCTCCCATCAGCGGACAAATTAGA
 TGTTCAATCAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAG
 ACCTGGAGGAGAGATATGAGGGACAATGGGAGAAGTGAATTATATAAAATATAAGTAGTAAAAATTGAACCAT
 TAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTG
 TTCCTTGGGTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACA
 ATTTATGTCCTGTGATGTCAGCAGCAGCAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTGCAAC
 TCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTC
 CTGGGGATTTGGGGTTGCTCTGGAATACTTTGACCACCTGCTGTGCCTTGGGAATGCTAGTTGGAGTAATAA
 ATCTCTGGAACAGATTTGGAATCACACGACGTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAA
 TACACTCCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAAATGAACAAGAATTATTGGAATTAGATAAATGG
 GCAAGTTTGTGGAATTGGTTAACAATAACAATTTGCTGTGGTATATAAAATTATTCAATATGATAGTAGGAGG
 CTTGGTAGGTTTAAGAATAGTTTTTGTGTACTTTCTGTAGTGAATAGAGTTAGGCAGGGATATTCACCATTAT
 CGTTTCAG**ACCTGGAGATCTCCCGAGGGGACCCGACAGGCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGA**
CAGAGACAGATCCATTTGACCAATTCCTCCTCAGGTGCAGGCTGCCTATCAGAAGGTGGTGGCTGGTGTGGC
CAATGCCCTGGCTCACAAATACCACTGAGATCCAGACATGATAAGATACATTGATGAGTTTGGACA

exón SD6 pSPL3
 pSPL3seq_F
 pSPL3seq_R
 exón SA2 pSPL3

Figura Suplementaria 1. Pareja de cebadores diseñados (pSPL3seq_F/pSPL3_R) y su localización respecto al constructo introducido en el vector pSPL3.

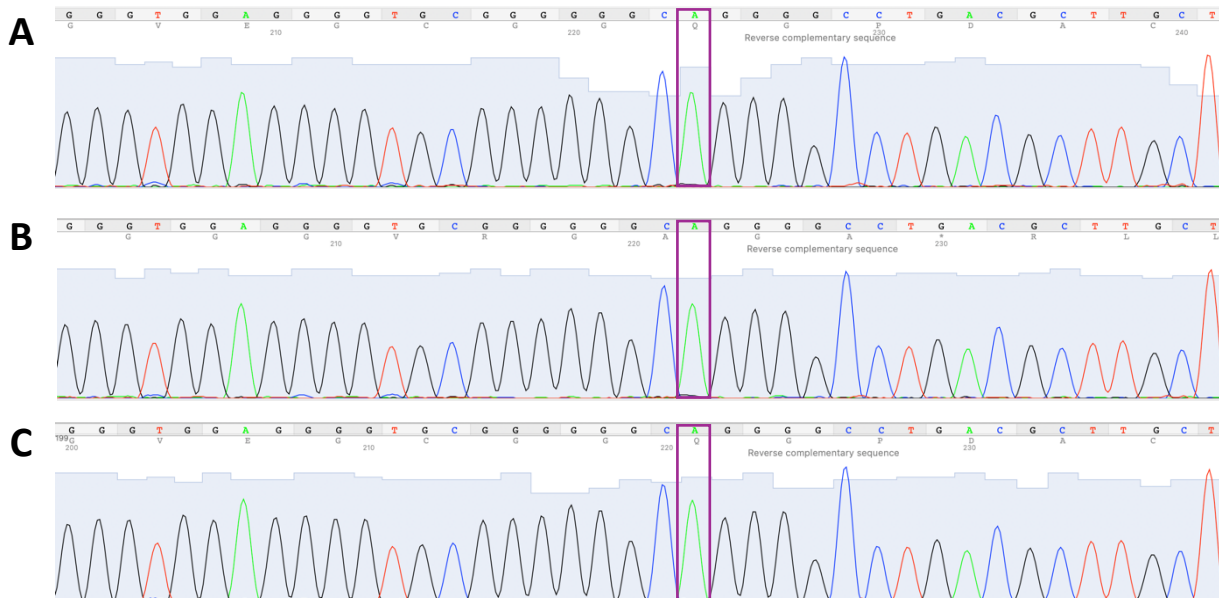


Figura Suplementaria 2. Electroferograma de la validación negativa del cambio c.4760-42A>T en el gen CACNA1G. El recuadro morado señala el lugar donde debería encontrarse la variante. A) Probando (MD-618) B) Madre (MD-619) C) Padre (MD-620)

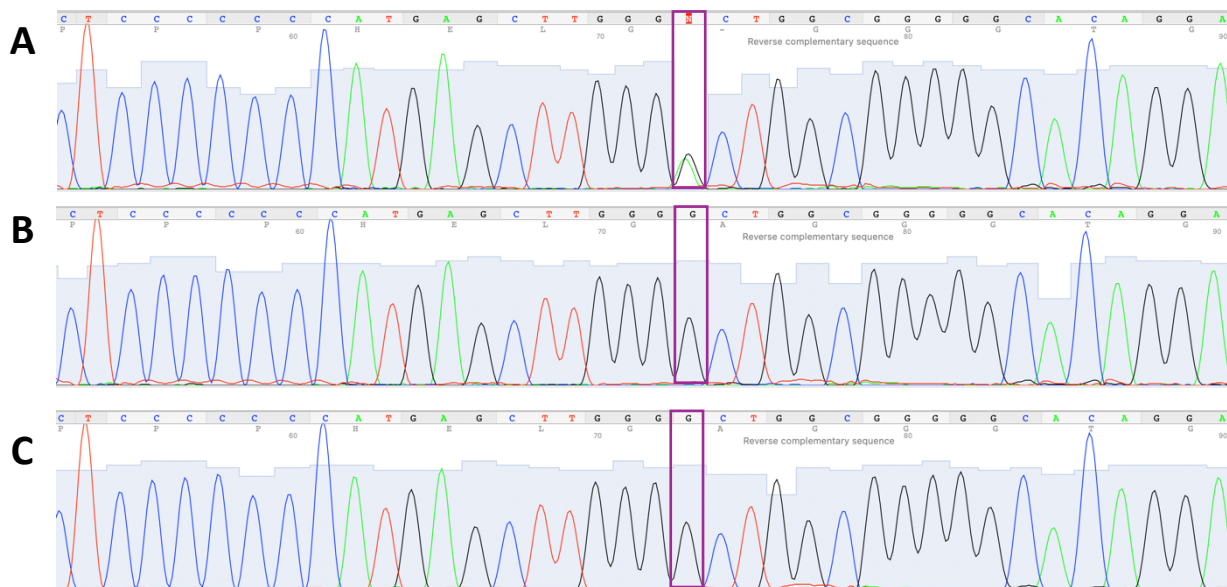


Figura Suplementaria 3. Electroferograma de la validación positiva del cambio c.12G>A en el gen *CROCC*. El recuadro morado señala el lugar donde debería encontrarse la variante. **A)** Probando (MD-618) **B)** Madre (MD-619) **C)** Padre (MD-620)

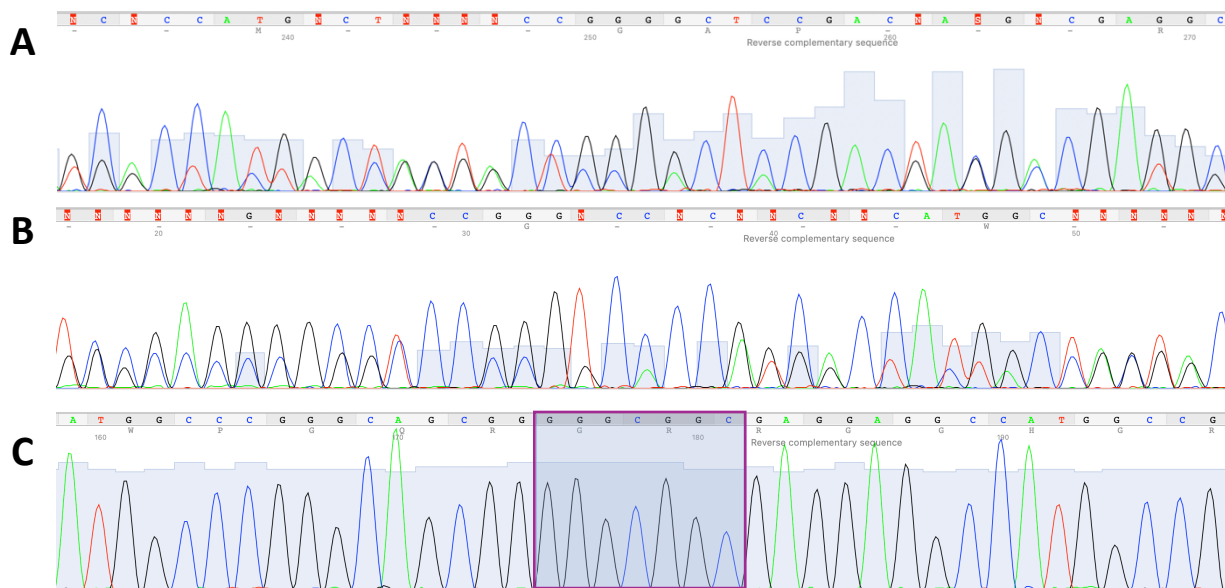


Figura Suplementaria 4. Electroferograma de la validación positiva del cambio c.7266_7271delCAGCGG en el gen *CACNA1A*. El recuadro morado señala el lugar donde debería encontrarse la variante. En este caso, la delección provoca dobles picos perfectos en el electroferograma en el probando y la madre al encontrarse en heterocigosis. **A)** Probando (MD-618) **B)** Madre (MD-619) **C)** Padre (MD-620)

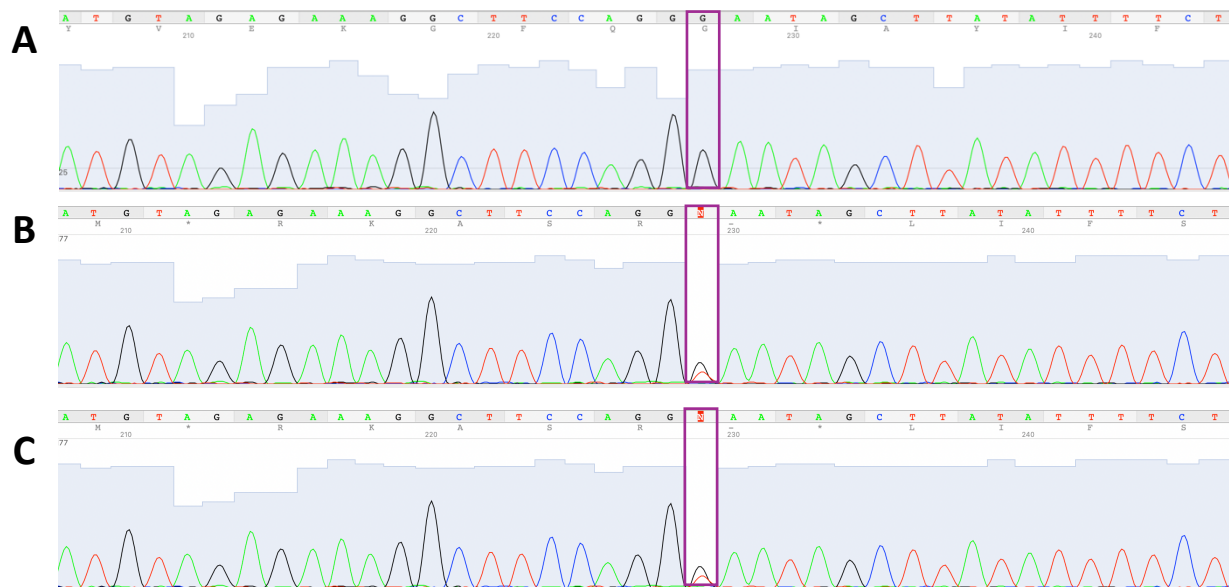


Figura Suplementaria 5. Electroferograma de la validación positiva del cambio c.690+T>G en el gen *SLC38A6*. El recuadro morado señala el lugar donde debería encontrarse la variante. En este caso, los progenitores son heterocigotos (presentan una T y una G), de manera que el probando hereda la G de ambos. **A)** Probando (MD-626) **B)** Madre (MD-627) **C)** Padre (MD-628)

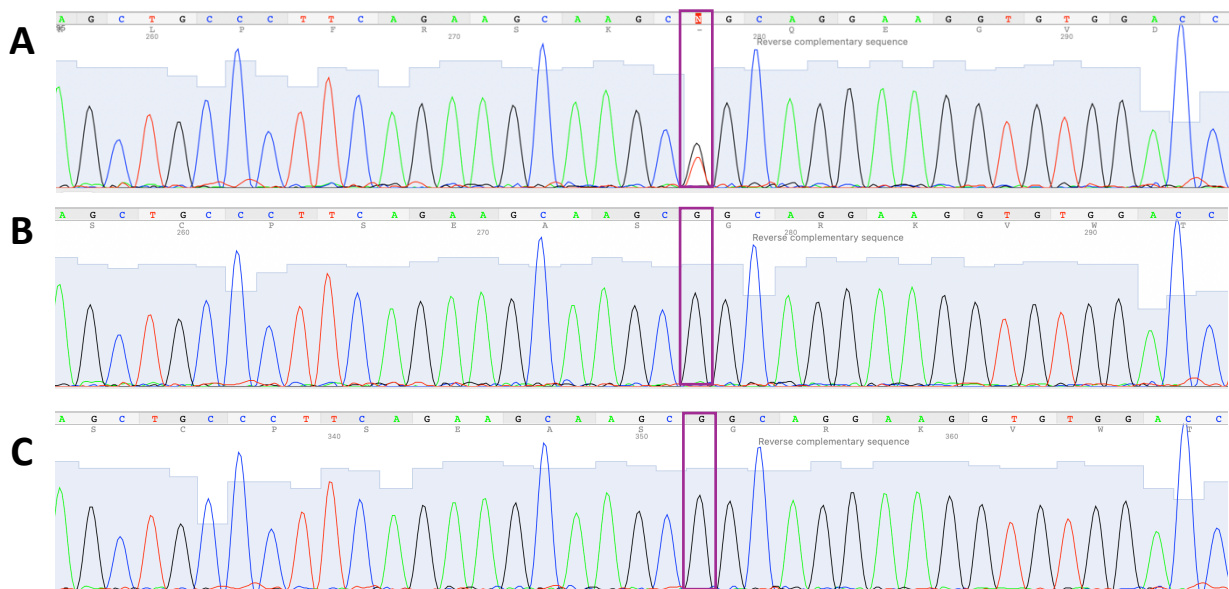


Figura Suplementaria 6. Electroferograma de la validación positiva del cambio c.2075G>T en el gen *SLITRK3*. El recuadro morado señala el lugar donde debería encontrarse la variante. En este caso, el cambio nucleotídico aparece *de novo* en heterocigosis con la variante normal en el probando. **A)** Probando (MD-626) **B)** Madre (MD-627) **C)** Padre (MD-628)

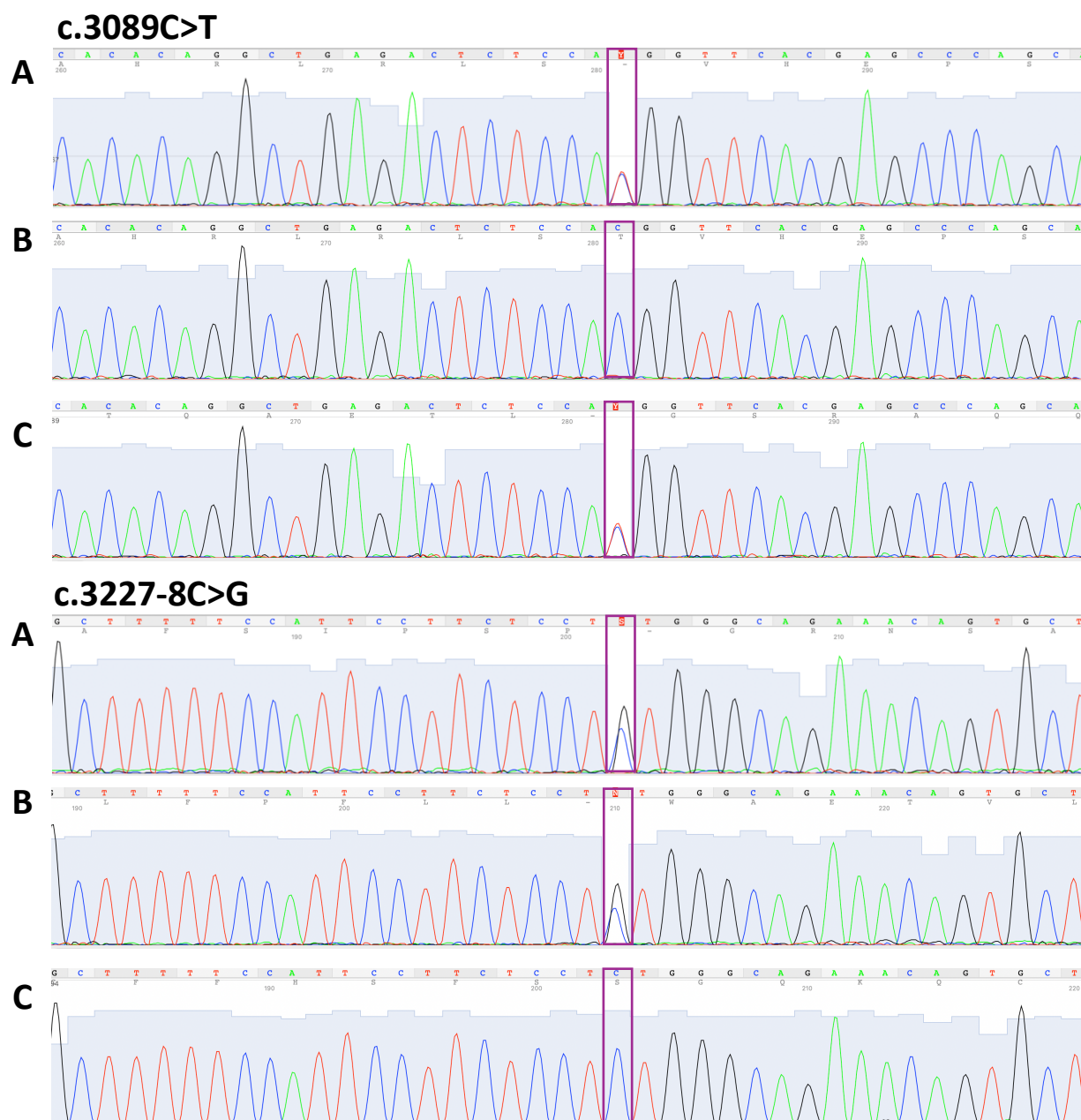


Figura Suplementaria 7. Electroferograma de la validación positiva de los cambios c.3089C>T y c.3227-8C>G en el gen *ASTN1*. El recuadro morado señala el lugar donde debería encontrarse la variante. En este caso, al tratarse de heterocigosis compuesta, el primer cambio (c.3089C>T) procede del padre y el segundo (y c.3227-8C>G) de la madre. **A)** Probando (MD-626) **B)** Madre (MD-627) **C)** Padre (MD-628)

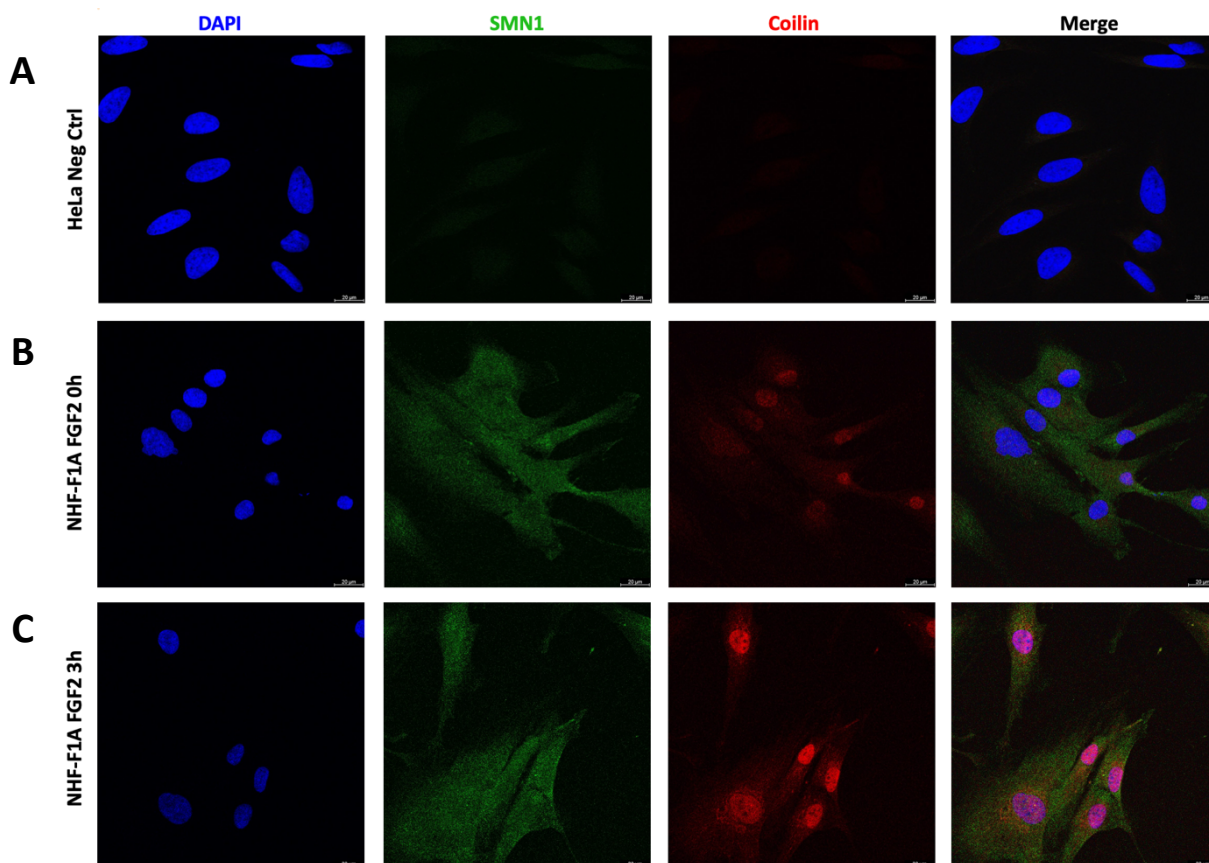


Figura Suplementaria 8. Imágenes representativas otros experimentos de inmunofluorescencia. Se observan células HeLa sin anticuerpo primario (control negativo) **(A)** y fibroblastos de donante sano (NHF-F1A) sin tratar **(B)** y tratados con FGF2 **(C)**, con anticuerpos anti-SMN1 y anti-Coilin, este último empleado para marcar los cuerpos de Cajal (CBs) y detectar la posible acumulación de SMN1 en los mismos. Los núcleos celulares se tiñen con DAPI (azul). Barra de escala: 20μm.

8.2. ANEXO II: CLONACIÓN DE *SLC38A6* EN VECTOR PCDNA3

Tabla Suplementaria 2. Cebadores empleados para la amplificación del ADNc del constructo WT previa a la clonación en el vector pCDNA3. En azul se marcan las secuencias de reconocimiento de las enzimas de restricción, y por delante, en minúscula, cuatro nucleótidos que favorecen el corte de las mismas.

Cebador	Secuencia (5'-3')	Temperatura hibridación	Tamaño amplicón
SD6_HindIII_F	ggaa AAGCTT TCTGAGTCACCTGGACAACC	55°C	329 bp (WT) o 263 bp (sin inserto del gen <i>SLC38A6</i>)
SA2_BamHI_R	gcgc GGATC CATCTCAGTGGTATTTGTGAGC		

Tabla Suplementaria 3. Reactivos para la digestión del inserto y del vector pcDNA3 con HindIII y BamHI.

Reactivo y concentración inicial	Digestión con HindIII (DÍA 1) Volumen a añadir (V _F = 50 µL)		Digestión con BamHI (DÍA 2)* Volumen a añadir (V _F = 50 µL)	
	INSERTO	VECTOR	INSERTO	VECTOR
ADN (vector: 2000 ng/µL)	20 µL	1,5 µL de vector + 18, 5µL H ₂ O = 20 µL	20 µL	20 µL
Tampón (10X) (Thermo Scientific)	Tampón R: 5 µL		Tampón BamHI: 5 µL	
Enzima (10 U/µL)	1 µL		1 µL	
H ₂ O libre de ADNsas y ARNsas	24 µL		24 µL	

*Antes de la digestión se deben purificar el inserto y el vector por columna con el kit Macherey-Nagel™ NucleoSpin Plasmid QuickPure™ Kit (ThermoFisher Scientific).

$$ng \text{ inserto} = 50 \times \frac{\text{tamaño del inserto (pb)}}{\text{tamaño del vector (pb)}} \times \frac{3}{1} = 50 \times \frac{329}{5428} \times \frac{3}{1} = 9,09 \text{ ng}$$

(Ecuación 2)

Tabla Suplementaria 4. Reactivos para la ligación del inserto (constructo WT) y del vector pcDNA3 digeridos.

Reactivo y concentración inicial	Concentración en la reacción	Volumen a añadir (µL) (V _F = 20µL)
pcDNA3 digerido (50 ng/µL)	2,50 ng/µL	1µL
Constructo WT digerido (74,1ng/µL)	0,19 ng/µL	1,23µL del inserto diluido 1/10 en H ₂ O
T4 ligasa (5U/µL)	0,05U/µL	0,2µL
Tampón T4 ligasa (10X)	1X	2 µL
H ₂ O libre de ADNsas y ARNsas	-	15,8µL

Tabla Suplementaria 5. Cebadores empleados para la PCR de colonias de la clonación del constructo WT en el vector pcDNA3.

Cebador	Secuencia (5'-3')	Temperatura hibridación	Tamaño amplicón
pCMV_D	TTGACGCAAATGGGCGGTAGG	56°C	278pb vector religado; 329pb o 263pb vector con <u>inserto</u>
BGH_R	TAGAAGGCACAGTCGAGG		

Tabla Suplementaria 6. Reactivos de la PCR de colonias y programa del termociclador.

Reactivo y concentración inicial	Concentración en la reacción de PCR	Volumen a añadir ($V_f = 25\mu\text{L}$)	Etapas del programa	Temperatura	Tiempo	Ciclos
H ₂ O libre de ADNsas y ARNsas	-	19 μL	Desnaturalización inicial	96°C	7 min	1
Tampón Taq polimerasa 10X (con MgCl ₂)	1X	2,5 μL	Desnaturalización	96°C	1 min	35
dNTPs (20 mM)	0,4 mM	0,5 μL	Annealing o hibridación	56°C	30 s	
Cebador directo, 5' (10 μM)	0,2 μM	0,5 μL	Extensión	72°C	45 s	
Cebador reverso, 3' (10 μM)	0,2 μM	0,5 μL	Extensión final	72°C	7 min	1
Taq ADN polimerasa, Biotools (1 U/ μL)	0,02 U/ μL	0,5 μL	Conservación	4°C	∞	
MgCl ₂ (20mM)	0,4 mM	0,5 μL				
ADN molde (~ 100-200 ng/ μL)	~ 4-8 ng/ μL	1 μL				