



UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Medicina

Departamento de Bioquímica

**Identificación y caracterización molecular de mutaciones en pacientes con  
enfermedad de Pompe.**

**Análisis del transcriptoma de sangre periférica**

Cintia Amiñoso Carbonero

Tesis Doctoral

Madrid, 2019

UNIVERSIDAD AUTÓNOMA DE MADRID  
Facultad de Medicina  
Departamento de Bioquímica

## Identificación y caracterización molecular de mutaciones en pacientes con enfermedad de Pompe.

### Análisis del transcriptoma de sangre periférica

Memoria presentada por la Licenciada en Bioquímica

**Cintia Amiñoso Carbonero**

Para optar al grado de Doctor en Bioquímica, por la Universidad Autónoma de Madrid.

Director de tesis:

**Dr. Jesús Solera García**

Este trabajo ha sido realizado en la Unidad de Oncogenética Molecular en el Instituto de Genética Médica y Molecular (INGEMM) en el Hospital Universitario La Paz, bajo la supervisión de Jesús Solera García.



**D. Jesús Solera García**, Profesor asociado del Departamento de Bioquímica de la Universidad Autónoma de Madrid,

**CERTIFICA:**

Que Dña. Cintia Amiñoso Carbonero, licenciada en Bioquímica, ha realizado bajo mi dirección, en la Unidad de Oncogenética Molecular en el Instituto de Genética Médica y Molecular (INGEMM) del Hospital Universitario La Paz, el trabajo titulado:

**Identificación y caracterización molecular de mutaciones en pacientes con enfermedad de Pompe. Análisis del transcriptoma de sangre periférica.**

Una vez supervisado el trabajo considero que reúne todos los requisitos necesarios en cuanto a originalidad y calidad científica requeridos para ser presentados como tesis doctoral, con el objeto de optar al grado de Doctor por la Universidad Autónoma de Madrid.

Y para que conste a todos los efectos firmo el presente certificado en,

Madrid, 22 de octubre de 2019

Fdo.: Jesús Solera García  
Director de la tesis

## AGRADECIMIENTOS

Escribir esta tesis se ha convertido en una tarea muy difícil. Gracias a que he tenido gente alrededor para ayudarme a que todo fuera un poco más sencillo y llevadero. A todos ellos quiero dedicarles una palabras.

Mi primer agradecimiento tiene que ser para mi director de tesis, el Dr. Jesús Solera, que es “el culpable” de que yo esté metida en este lío. Muchísimas gracias Jesús por acogerme en tu laboratorio en el 2005, por animarme a estudiar la carrera porque me distes el empujoncito que necesitaba, porque siempre has creído en mí, incluso más de lo que yo he creído en mi misma. Muchas gracias por tu apoyo todos estos años y por tu tiempo. Sin tí, esto no habría sido ni parecido.

Quiero acordarme del laboratorio en el que tuve mi primer contacto con la investigación. Dar las gracias a Juan y María José por darme la oportunidad de hacer ciencia en su labo y a mis compañeros, Lola, Pilar, Eva y Jesús..... Fue el primer laboratorio en el que estuve y no puedo tener mejor recuerdo de aquella época. Aprendí muchísimo de biología molecular, me enseñasteis a trabajar minuciosamente y la importancia del compañerismo. Gracias chicos por ser como sois y por seguir ahí casi 20 años después.

Gracias a mis chicas del grupo “Antozoicas”. A Julita por esas comidas de fin de año en las que nos ponemos al día de todo lo ocurrido. A ti, Martus, por tu apoyo y por aguantarme estos años, gracias por ser mi amiga y por estar siempre ahí, en los buenos y en los malos momentos.

Gracias a mis compañeros del Master de Bioinformática por hacer que la experiencia fuera un poco más llevadera. En especial, a mi grupo “Microsocial” por toda vuestra ayuda y amistad, durante y después del Máster, por esas entretenidas charlas y esas copiosas cenas.

Gracias a la gente de mi labo del INGEMM. A Pilar y a Sara, que me han regalado su tiempo cuando más lo necesitaba. Estos años, y sobretodo el último en el que ya se notaba la presión, ellas han estado ayudándome en todo lo que han podido y he necesitado. Muchas gracias chicas por vuestra ayuda y amistad. A los que ya no están en nuestra poyata, Ana, Héctor, Laura y María,

que también han formado parte de esto, y que han hecho más llevaderas tantas horas delante de los tubos. Al personal del INGEMM, que con su trabajo han permitido que el mío fuera más fácil.

Quiero dar las gracias a Genzyme-Sanofi como institución y, especialmente, a todas las personas relacionadas con las enfermedades lisosomales tanto en el departamento médico como en el área comercial. Su ayuda logística y económica, han sido determinantes para que este trabajo haya visto la luz.

Muchas gracias a todos los pacientes que han cedido sus muestras para que el conocimiento en esta enfermedad avance, a los médicos que han colaborado y a mis voluntarios sanos que forman una parte imprescindible de este trabajo.

A Luis del Peso, Ramón Díaz Uriarte y a Gonzalo Gómez por su inestimable ayuda con la parte bioinformática.

Por último, y no menos importante, quiero agradecer a mi familia su apoyo incondicional durante todos estos años. A mis padres, que siempre me han inculcado el valor del esfuerzo y del trabajo bien hecho, unos valores que han hecho que sea como soy. Que aunque no siempre han estado de acuerdo con todas mis decisiones, las han acatado y respetado, estando pendientes por si había que entrar al rescate. Gracias por apoyarme en todas mis “aventuras”. A mis hermanas, mis cuñados, mi sobrinillo y a mi tía, por su cariño, ayuda, apoyo, su cercanía, sus buenos consejos....., gracias por estar siempre ahí, os quiero mucho.

Gracias, en definitiva, a todas las personas que han formado parte de esto, de una u otra forma, y que han permitido que esto culmine. Espero no haberme olvidado de nadie.

¡¡¡Muchísimas gracias a todos!!!!



ÍNDICE

ÍNDICE DE FIGURAS .....	1
ÍNDICE DE TABLAS .....	3
ABREVIATURAS .....	5
RESUMEN .....	7
SUMMARY .....	9
INTRODUCCIÓN.....	13
<b>1.1 El lisosoma. ....</b>	<b>13</b>
1.1.1. Características estructurales .....	13
1.1.2. Funciones del lisosoma .....	14
1.1.3. Enfermedades relacionadas con el lisosoma .....	16
<b>1.2. Glucogenosis tipo II. Enfermedad de Pompe .....</b>	<b>17</b>
1.2.1. Aspectos históricos.....	17
1.2.2. Aspectos clínicos de la enfermedad.....	17
1.2.3. Bases moleculares de la enfermedad de Pompe .....	19
1.2.3.1. Estructura del gen GAA .....	19
1.2.3.2. Perfil de expresión tisular del gen GAA.....	20
1.2.3.3. Procesamiento del mRNA.....	21
1.2.3.4. Síntesis de la proteína y procesamiento proteolítico.....	23
1.2.3.5. Mutaciones descritas en el gen GAA.....	24
1.2.4. Diagnóstico de la enfermedad. ....	28
1.2.5. Tratamiento.....	30
<b>1.3. Análisis del transcriptoma mediante RNA-Seq. ....</b>	<b>31</b>
<b>2. OBJETIVOS.....</b>	<b>35</b>
<b>3. MATERIAL Y MÉTODOS .....</b>	<b>39</b>
<b>3.1. Población de estudio .....</b>	<b>39</b>
<b>3.2. Aislamiento y cuantificación de ácidos nucleicos .....</b>	<b>39</b>
<b>3.3. Amplificación mediante PCR del gen GAA. ....</b>	<b>40</b>
<b>3.4. Purificación de los fragmentos amplificados y secuenciación mediante el método Sanger.....</b>	<b>42</b>

<b>3.5. Amplificación de sondas dependiente de ligación múltiple (MLPA) de diseño propio.</b>	<b>42</b>
<b>3.6. Array CGH (Comparative Genomic Hybridization)</b>	<b>46</b>
<b>3.7. Análisis de paternidad</b>	<b>47</b>
<b>3.8. Interpretación biológica de las variantes identificadas en el gen GAA</b>	<b>47</b>
3.8.1. Análisis <i>in silico</i> de variantes missense	47
3.8.2. Análisis funcional de procesamiento de mRNA	50
<b>3.9. Análisis transcriptoma mediante RNA-seq</b>	<b>53</b>
3.9.1. Pacientes analizados	53
3.9.2. RNA-seq. Procedimiento de análisis	53
3.9.2.1. Flujo de análisis.	53
3.9.2.2. Procesamiento de los datos crudos	54
3.9.2.3. Alineamiento con el genoma de referencia. Cuantificación de las lecturas por gen.	54
3.9.2.4. Análisis de expresión diferencial.	54
3.9.2.5. Análisis de enriquecimiento funcional.	56
<b>4. RESULTADOS.</b>	<b>61</b>
<b>4.1. Análisis genético de pacientes con sospecha de enfermedad de Pompe</b>	<b>61</b>
4.1.1. Variantes identificadas en el estudio genético de los pacientes	61
4.1.2. Pacientes con alteraciones genéticas descritas	61
4.1.3. Pacientes con alteraciones genéticas no descritas	65
4.1.4. Caracterización molecular de nuevas variantes identificadas por MLPA	67
4.1.4.1. Caracterización de la variante g.80082249-80121055delins14bp presente en el paciente P11	67
4.1.4.2. Caracterización de la variante c.2799+244_2859+63del presente en el paciente P481	68
4.1.5. Análisis bioinformático de las variantes no descritas y de las descritas como de significado desconocido.	68
4.1.6. Análisis funcional de las variantes con potencial implicación en el procesamiento del mRNA...	71
4.1.6.1. Paciente P70.	71
4.1.6.2. Paciente P210.	73
4.1.6.3. Paciente P290.	74
4.1.6.4. Paciente P300.	75
4.1.6.5. Paciente P318.	76
4.1.6.6. Paciente P345.	78
4.1.6.7. Paciente P346.	79
4.1.6.8. Resto pacientes.	82



## ÍNDICE

---

4.1.7. Cómputo global de resultados de los análisis moleculares.....	82
<b>4.2. Análisis de transcriptoma mediante RNA-Seq.....</b>	<b>83</b>
4.2.1. Análisis de expresión diferencial.....	84
4.2.2. Identificación de rutas transcripcionalmente alteradas .....	88
4.2.2.1. Análisis de enriquecimiento funcional mediante la herramienta <i>GSEA</i> .....	88
4.2.2.2. Análisis de enriquecimiento funcional mediante la herramienta Pathview .....	91
<b>5. DISCUSIÓN .....</b>	<b>99</b>
<b>5.1. Diagnóstico y caracterización molecular de pacientes con sospecha de Enfermedad de Pompe. .</b>	<b>99</b>
5.1.1. Pacientes con fenotipo infantil.....	99
5.1.2. Pacientes con fenotipo tardío .....	102
<b>5.2. Análisis de transcriptoma mediante RNA-seq .....</b>	<b>109</b>
<b>6. CONCLUSIONES .....</b>	<b>115</b>
<b>7. BIBLIOGRAFÍA.....</b>	<b>119</b>
<b>8. PUBLICACIONES.....</b>	<b>135</b>
<b>8.1. ARTÍCULO 1: Identification of the first deletion–insertion involving the complete structure of GAA gene and part of CCDC40 gene mediated by an Alu element. ....</b>	<b>135</b>
<b>8.2. ARTÍCULO 2: Reevaluating the pathogenicity of the mutation c. 1194+5G&gt;A in GAA gene by functional analysis of RNA in a 61-year-old woman diagnosed with Pompe disease by muscle biopsy. ....</b>	<b>135</b>

ÍNDICE DE FIGURAS

<b>Figura 1.</b> Funciones del lisosoma. ....	15
<b>Figura 2.</b> Microfotografías de extensiones de sangre de pacientes con enfermedad de Pompe y control sano.....	18
<b>Figura 3.</b> Estructura genómica del gen <i>GAA</i> . Secuencias de los límites intrón/exón. ....	20
<b>Figura 4.</b> Niveles de expresión <i>GAA</i> en distintos tejidos. Tránscritos expresados en músculo esquelético y en sangre periférica. ....	21
<b>Figura 5:</b> Secuencias consenso implicadas en el procesamiento del mRNA.....	22
<b>Figura 6.</b> Secuencia de eventos que se producen en el procesamiento del mRNA.....	22
<b>Figura 7.</b> Variantes de procesamiento del exón 2 observadas en <i>GAA</i> normal.....	23
<b>Figura 8.</b> Estructura de la proteína <i>GAA</i> .....	24
<b>Figura 9.</b> Tipos de alteraciones moleculares descritas como patogénicas o probablemente patogénica en la base de datos del Pompe Center. ....	25
<b>Figura 10.</b> Localización de variantes puntuales patológicas en la estructura de la proteína <i>GAA</i> . ....	26
<b>Figura 11.</b> Algoritmo diagnóstico para la Enfermedad de Pompe .....	29
<b>Figura 12.</b> Esquema explicativo de la técnica de <i>MLPA</i> . ....	43
<b>Figura 13.</b> Perfil de picos del <i>MLPA</i> de diseño propio obtenido por electroforesis capilar.....	46
<b>Figura 14.</b> Esquema explicativo de la técnica de <i>array CGH</i> . ....	47
<b>Figura 15.</b> Esquema de trabajo para el análisis de las variantes implicadas en procesamiento. ....	51
<b>Figura 16.</b> Algoritmo utilizado para el análisis del <i>RNA-seq</i> .....	53
<b>Figura 17.</b> Esquema general método <i>GSEA</i> .....	56
<b>Figura 18.</b> Localización de las mutaciones detectadas, distribuidas a lo largo de toda la estructura del gen <i>GAA</i> . ....	66
<b>Figura 19.</b> Caracterización de los puntos de ruptura de la deleción detectada en el paciente P11.....	67
<b>Figura 20.</b> Caracterización de los puntos de ruptura de la deleción detectada en el paciente P481.....	68
<b>Figura 21.</b> Análisis del procesamiento de un control normal y del paciente P70.....	72
<b>Figura 22.</b> Análisis del procesamiento de un control normal y del paciente P210.....	73
<b>Figura 23.</b> Análisis del procesamiento de un control normal y del paciente P290.....	74
<b>Figura 24.</b> Análisis del procesamiento de un control normal y del paciente P300.....	76
<b>Figura 25.</b> Análisis del procesamiento de un control normal y del paciente P318.....	77
<b>Figura 26.</b> Análisis del procesamiento de un control normal y del paciente P345.....	79
<b>Figura 27.</b> Análisis del procesamiento de un control normal y del paciente P346.....	81
<b>Figura 28.</b> Análisis del procesamiento de un control normal y del paciente P349.....	82
<b>Figura 29.</b> Cómputo global de los resultados obtenidos. ....	83

## ÍNDICE FIGURAS

---

<b>Figura 30.</b> <i>MDS plot</i> de los datos normalizados.....	84
<b>Figura 31.</b> <i>Heatmap</i> genes diferencialmente expresados. ....	85
<b>Figura 32.</b> <i>Boxplot</i> realizados con los genes con diferencias de expresión más significativas. ....	86
<b>Figura 33.</b> Gráfica del valor de enriquecimiento de las rutas con los valores de <i>NES</i> más altos ( <i>Hallmark</i> ). .....	89
<b>Figura 34.</b> Ruta hsa04722_ <i>Neurotrophin signaling pathway</i> . ....	93
<b>Figura 35.</b> Ruta hsa04668_ <i>TNF signaling pathway</i> .....	94
<b>Figura 36.</b> Ruta hsa04141_ <i>Protein processing in endoplasmic reticulum</i> .....	95

ÍNDICE DE TABLAS

<b>Tabla 1.</b> Oligonucleótidos utilizados para la caracterización molecular de los pacientes. ....	40
<b>Tabla 2.</b> Condiciones de amplificación por PCR de todos los fragmentos del gen <i>GAA</i> amplificados para la caracterización molecular de los pacientes. ....	41
<b>Tabla 3.</b> Condiciones de la PCR de secuenciación de todos los fragmentos del gen <i>GAA</i> para la caracterización molecular de los pacientes. ....	42
<b>Tabla 4.</b> Oligonucleótidos utilizados para la realización del <i>MLPA</i> , mezcla 1. ....	44
<b>Tabla 5.</b> Oligonucleótidos utilizados para la realización del <i>MLPA</i> , mezcla 2. ....	45
<b>Tabla 6.</b> Valores por defecto de las herramientas que integran la plataforma <i>Alamut</i> <sup>®</sup> visual. ....	50
<b>Tabla 7.</b> Condiciones de la qPCR para realizar el ensayo cuantitativo. ....	52
<b>Tabla 8.</b> Cebadores del gen constitutivo <i>β2-Microglobulina</i> usado para la normalización de la qPCR. ....	52
<b>Tabla 9.</b> Pacientes con genotipo confirmatorio de enfermedad de Pompe. ....	62
<b>Tabla 10.</b> Pacientes con genotipo formado por variante patogénica y VSD o dos VSDs. ....	63
<b>Tabla 11.</b> Pacientes en los que se identificó una variante descrita como no patogénica. ....	63
<b>Tabla 12.</b> Pacientes portadores de enfermedad de Pompe. ....	64
<b>Tabla 13.</b> Pacientes portadores de una variante descrita como de significado desconocido. ....	65
<b>Tabla 14.</b> Pacientes en los que se identificó una mutación no descrita previamente. ....	66
<b>Tabla 15.</b> Pacientes portadores de una variante no descrita. ....	66
<b>Tabla 16.</b> Análisis bioinformático de las variantes no descritas y de las descritas como de significado desconocido. ....	70
<b>Tabla 17.</b> Pacientes incluidos en el análisis funcional para analizar la implicación de las variantes de interés en el procesamiento. ....	71
<b>Tabla 18.</b> Muestras analizadas mediante <i>RNA-seq</i> . ....	83
<b>Tabla 19.</b> Descripción de la función de los genes más significativamente expresados positivamente. ....	87
<b>Tabla 20.</b> Descripción de la función de los genes más significativamente expresados negativamente. ....	88
<b>Tabla 21.</b> Resumen del resultado de enriquecimiento funcional positivo obtenido mediante <i>GSEA</i> . ....	88
<b>Tabla 22.</b> Resumen del resultado de enriquecimiento funcional negativo obtenido mediante <i>GSEA</i> . ....	89
<b>Tabla 23.</b> Rutas enriquecidas positiva y negativamente con un <i>FDR</i> <0,01, obtenidas con <i>GSEA</i> con la agrupación <i>Hallmark</i> . ....	89
<b>Tabla 24.</b> Rutas enriquecidas positiva y negativamente con <i>FDR</i> <0,01, obtenidas con <i>GSEA</i> con las agrupaciones de <i>KEGG</i> . ....	90
<b>Tabla 25.</b> Rutas enriquecidas positiva y negativamente con un <i>FDR</i> <0,05, obtenidas con <i>GSEA</i> con la agrupación <i>Biocarta</i> . ....	90

**Tabla 26.** Primeras 10 rutas enriquecidas positivamente, obtenidas con *GSEA* con las agrupaciones de *TFT*.  
..... 91

**Tabla 27.** Rutas de *KEGG* diferencialmente expresadas con  $FDR < 0.1$  obtenidas con la librería de R *Pathview*  
..... 92

### ABREVIATURAS

3'ss	Sitio 3' acceptor (del inglés 3' <i>Splice site</i> ).
5'ss	Sitio 5' donador (del inglés 5' <i>Splice site</i> ).
ACMG	Colegio Americano de Medicina Genética y Genómica (del inglés <i>American College of Medical Genetics and Genomics</i> ).
CGH	Hibridación genómica comparada (del inglés <i>Comparative Genomic Hybridization</i> ).
Chr	Cromosoma.
CLEAR	Del inglés <i>Coordinated Lysosomal Expression and Regulation Element</i> .
CPK	Creatinfosfoquinasa.
CRIM	Del inglés <i>Cross Reactive Immunologic Material</i> .
DBS	Sangre en gota seca (del inglés <i>Dried blood spot</i> ).
dNTPs	Desoxinucleótidos trifosfato.
EDL	Enfermedad de depósito lisosomal.
EE	Endosoma temprano (del inglés <i>Early Endosome</i> ).
ERT	Terapia de reemplazamiento enzimático (del inglés <i>Enzyme replacement therapy</i> ).
ES	Valor de enriquecimiento (del inglés <i>enrichment score</i> ).
ESE	Del inglés <i>Exonic splicing enhancers</i> .
ESS	Del inglés <i>Exonic splicing silencers</i> .
FDR	Tasa de Descubrimiento Falso (del inglés <i>False Discovery Rate</i> ).
Glc4	Tetrasacárido de glucosa.
GS	Del inglés <i>GeneSplicer</i> .
GSDII	Glucogenosis Tipo II (del inglés <i>Glycogen storage disease</i> ).
GSEA	Del inglés <i>Gene Set Enrichment Analysis</i> .
GTEX	Del inglés <i>Genotype-Tissue Expression</i> .
HSF	Del inglés <i>Human Splice Finder</i> .
ISE	Del inglés <i>Intronic splicing enhancers</i> .
ISS	Del inglés <i>Intronic splicing silencers</i> .
Kpb	Kilo pares de bases.
LAMP	Proteína asociada a la membrana del lisosoma (del inglés <i>Lysosome-associated membrane glycoproteins</i> ).
LE	Endosoma tardío (del inglés <i>Late Endosome</i> ).

## ABREVIATURAS

---

M6P	Manosa-6-fosfato.
MES	Del inglés <i>MaxEntScan</i> .
MLPA	Del inglés <i>Multiplex Ligation-dependent Probe Amplification</i> .
MPR	Receptor manosa-6-P.
NES	Valor de enriquecimiento normalizado (del inglés <i>normalized enrichment score</i> ).
NMD	Del inglés <i>Nonsense Mediated Decay</i> .
NNS	Del inglés <i>NNSplice</i> .
PAS	Ácido peryódico de Schiff (del inglés <i>Periodic acid-Schiff</i> ).
PCR	Reacción en Cadena de la Polimerasa.
qPCR	Reacción en Cadena de la Polimerasa cuantitativa.
RE	Retículo endoplásmico.
rhGAA	Glucosidasa ácida lisosomal humana recombinante.
RNA-seq	Secuenciación de RNA.
snRNP	Del inglés Small nuclear RNP.
SNVs	Variantes de un solo nucleótido (del inglés <i>Single nucleotide variants</i> ).
SSF	Del inglés <i>Splice Site Finder like</i> .
SP	Sangre periférica.
SV	Variante de procesamiento (del inglés <i>splicing variant</i> ).
TBE	Tampón Tris Borato y EDTA.
TFEB	Factor de transcripción EB (del inglés <i>Transcription factor EB</i> ).
UTR	Región no traducida (del inglés <i>Untranslated Region</i> ).
VSD	Variante de Significado Desconocido.

Algunos términos en inglés ampliamente utilizados en Biología Molecular y sin clara traducción en español se muestran en cursiva.

### RESUMEN

La Glucogenosis tipo II, también conocida como enfermedad de Pompe, es una enfermedad genética con un patrón de herencia autosómico recesivo, producida por el déficit de la enzima alfa glucosidasa acida lisosomal, responsable de la ruptura del glucógeno en esa organela subcelular. Es una patología con una gran variabilidad clínica, que depende principalmente del grado de actividad enzimática residual, lo que dificulta de manera apreciable el diagnóstico y condiciona su evolución.

En este trabajo hemos estudiado una cohorte de 259 pacientes con sospecha de padecer la enfermedad. Para ello, hemos puesto a punto un abordaje mediante secuenciación Sanger de las regiones exónicas e intrónicas flanqueantes del gen *GAA*. Para identificar algunos tipos de reordenamientos complejos hemos enriquecido el estudio con un *MLPA* de diseño propio. Esta aproximación nos ha permitido alcanzar un diagnóstico genético confirmatorio en 76 pacientes, detectando 20 variantes nuevas, no descritas anteriormente, de las que 2 son deleciones detectadas gracias al *MLPA*, una de ellas afecta a la estructura completa del gen *GAA* y parte del gen contiguo *CCDC40*, siendo la mayor descrita hasta el momento. Además, la mutación c.-32-13T>G ha sido identificada en el 84,3% de los pacientes con la forma tardía de la enfermedad.

Hemos realizado la caracterización funcional de 16 variantes, para explorar su potencial implicación en el procesamiento del mRNA. El método utilizado pone de manifiesto tanto alteraciones cualitativas como cuantitativas que afectan al rendimiento y la calidad del mRNA resultante de un alelo mutado. Esto nos ha permitido dilucidar el papel de 7 de estas variantes en un procesamiento aberrante del RNA mensajero. Entre ellas, hemos detectado la implicación en el procesamiento de la variante *missense* c.1704C>G (p.H568Q), de la *nonsense* c.343C>T (p.Q115\*), de la variante sinónima c.546G>C (p.T182T), así como de las variantes que afectan a sitios canónicos de *splicing* c.1075+1G>A, c.1637-2delA y c.1889-1G>A. También hemos podido establecer la implicación de la variante c.1194+5G>A en el procesamiento aberrante del mRNA y, de ese modo, determinar su patogenicidad. Antes de este trabajo esta variante era considerada como no patogénica.

El análisis de transcriptoma en células de sangre periférica ha permitido determinar la desregulación de rutas relacionadas con la respuesta inmune, como la vía de señalización de TNFa a través de NFkB, así como la apoptosis, entre otras. Asimismo, se ha detectado una desregulación del sistema ubiquitina-proteasoma, implicado principalmente en la degradación de proteínas intracelulares.

Las diferencias de expresión observadas con grupos seleccionados de genes han permitido una agrupación no supervisada, capaz de discriminar entre pacientes y controles. Esta puede ser la base para obtener una firma génica y ayudar en la búsqueda de nuevos biomarcadores.



### SUMMARY

Type II glycogenosis, also known as Pompe disease, is an autosomal recessive disorder. It is produced by the deficit of the lysosomal enzyme acid-alpha-glycosidase, responsible for the breakdown of glycogen in that subcellular organelle. It is a pathology with great clinical variability, which mainly depends on the degree of residual enzymatic activity. It makes diagnosis difficult and determines the evolution.

In this work we have studied a cohort of 259 patients with suspected disease. To this end, we have set up an approach through Sanger sequencing of the exonic and flanking intronic regions the *GAA* gene. To assess some types of complex rearrangements we have enriched the study using a homemade MLPA. This approach allowed us to reach a confirmatory genetic diagnosis in 76 patients. In addition, we have discovered 20 new variants, two of which are deletions detected by the new *MLPA*. One of these deletions spans the complete structure of *GAA* gene and part of the adjacent *CCDC40*, being the largest described so far. In addition, the c.-32-13T>G mutation has been identified in 84.3% of patients with the late form of the disease.

We have performed the functional characterization of 16 variants, to explore its potential involvement in mRNA processing. The procedure reveals both qualitative and quantitative alterations affecting the performance and quality of mRNA resulting from a mutated allele. This has allowed us to elucidate the role of 7 of these variants in aberrant messenger RNA processing. In particular, we have detected the implication in RNA processing of the missense c.1704C>G (p.H568Q), the nonsense variant c.343C>T (p.Q115\*), the synonymous variant c.546G>C (p.T182T). At the same time, several variants that affect canonical splicing sites such as c.1075+1G>A, c.1637-2delA and c.1889-1G>A, were analyzed with the methodology. We have also been able to establish the involvement of the c.1194+5G>A variant in the aberrant processing of mRNA and thereby determine its pathogenicity. Prior to this work this variant was considered non-pathogenic.

The transcriptome analysis in peripheral blood cells has allowed to determine the deregulation of routes related to the immune response, such as the TNF $\alpha$  signaling pathway through NF $\kappa$ B, as well as apoptosis, among others. Likewise, a deregulation of the ubiquitin-proteasome system has been detected, mainly involved in the degradation of intracellular proteins.

The differences in expression observed with selected groups of genes have allowed an unsupervised clustering, capable of discriminating between patients and controls. This may be the basis for obtaining a gene signature and assisting in the search for new biomarkers.

## INTRODUCCIÓN

### INTRODUCCIÓN

La Glucogenosis tipo II (GSDII), también conocida como enfermedad de Pompe, es una patología de base genética que se transmite con un patrón de herencia autosómico recesivo. Está producida por mutaciones en el gen *GAA*, lo que ocasiona la deficiencia de la enzima  $\alpha$ -glucosidasa (maltasa) ácida lisosomal (EC 3.2.1.20), responsable de la degradación del glucógeno en el lisosoma. El déficit enzimático provoca la acumulación intralisosomal progresiva de este metabolito, un proceso que daña las células de un gran número de tejidos, afectando principalmente al músculo esquelético y cardíaco, al sistema nervioso central, al hígado, riñón y linfocitos.

#### 1.1 El lisosoma.

##### 1.1.1. Características estructurales

Los lisosomas fueron originalmente descritos en 1955 por el Dr. Christian de Duve, en estudios realizados en células hepáticas de rata (Duve et al., 1955). Son organelas intracelulares encontradas en el citosol de todas las células eucariotas. Su tamaño difiere de un tipo celular a otro, pero en la mayoría de las células es inferior a 1  $\mu\text{m}$  (Lüllmann-Rauch, 2005). Están limitados por una membrana con estructura de bicapa lipídica, en la que se han identificado más de 100 proteínas integradas que agrupan a transportadores, canales iónicos, o el complejo vATPasa, que mantiene el pH intralisosomal cercano a 4,5 (Mindell, 2012; Schwake et al., 2013). El alto contenido de residuos glicosilados en los dominios intraluminales de las proteínas de membrana parece protegerla de la acción de los enzimas hidrolíticos del interior del lisosoma (Saftig et al., 2010).

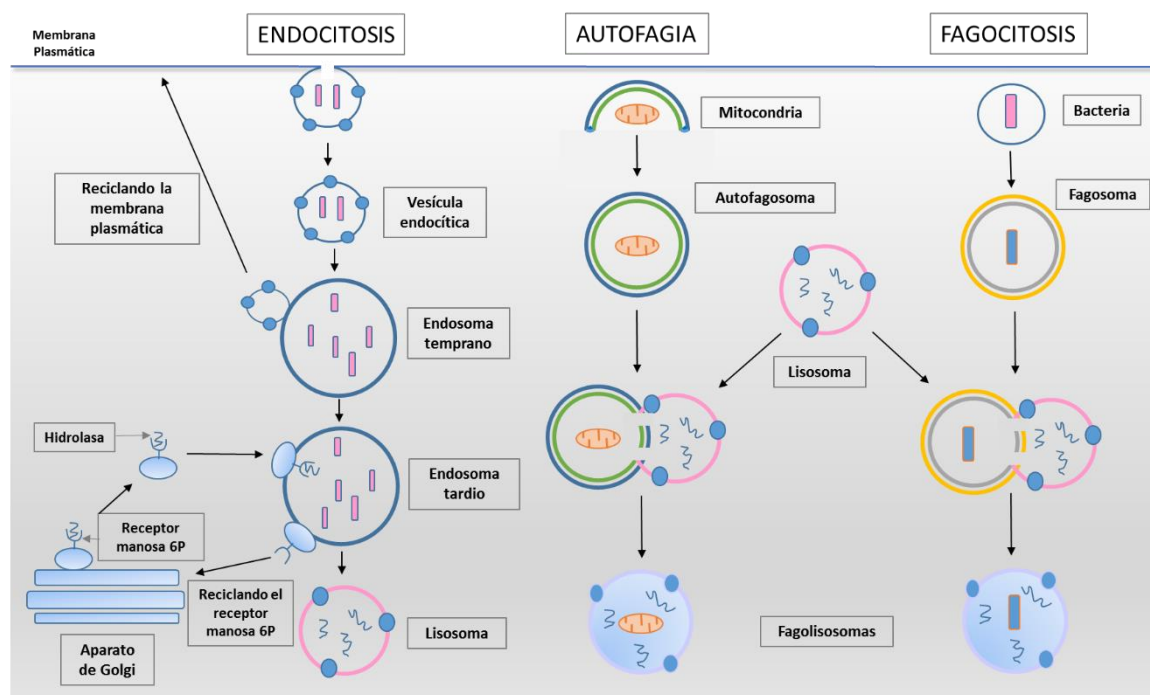
Hasta el momento, se han identificado 60 hidrolasas, incluyendo nucleasas, proteasas, glucosidasas, sulfatasas, fosfatasas y lipasas, entre otras, lo que permite la degradación de macromoléculas de naturaleza muy diferente (Xu and Ren, 2015). Las hidrolasas lisosomales se sintetizan en el retículo endoplásmico rugoso y se transportan a través del aparato de Golgi, hasta la red trans-Golgi, desde donde alcanzan el compartimento lisosomal. La explicación de cómo se reconocen y seleccionan con precisión las enzimas lisosomales entre las otras proteínas en la red trans-Golgi, se basa en que son etiquetadas con un marcador único: el grupo manosa-6-fosfato (M6P), que se agrega a residuos de N-oligosacáridos unidos a las hidrolasas lisosomales (Coutinho et al., 2012). Los residuos M6P de las enzimas lisosomales interactúan con receptores de M6P (MPR), y el complejo enzima-MPR resultante entra en los endosomas tempranos. A medida que el pH disminuye durante la maduración del endosoma, los MPR liberan los enzimas que, a través de una fusión del endosoma tardío con el lisosoma, llegan a esta organela (Braulke and Bonifacino, 2009).

### 1.1.2. Funciones del lisosoma

Los lisosomas son elementos dinámicos, que constituyen el compartimento primordial en el que se producen los procesos catabólicos en las células eucariotas (**Figura 1**) (Saftig and Klumperman, 2009). En ellos son degradados componentes de la matriz extracelular, que son internalizados y alcanzan el lisosoma mediante diferentes procesos, entre ellos la endocitosis. El material entra en la célula por vesículas mediadas por clatrina que derivan de la membrana plasmática y se fusionan con el endosoma temprano (EEs). Los componentes de la membrana se reciclan y los EEs maduran, dando lugar a los endosomas tardíos (LEs). El contenido de los LE es finalmente transferido al lisosoma, donde son degradados por acción de las hidrolasas (Luzio et al., 2009). Por otro lado, esta ruta es utilizada por la célula como un sistema de reposición de los componentes lisosomales, de modo que existe un solapamiento en el contenido molecular de los endosomas tardíos y los lisosomas, incluyendo las proteínas LAMP de membrana y las hidrolasas (Braulke and Bonifacio, 2009). Sin embargo, endosomas tardíos y lisosomas pueden ser distinguidos por la ausencia de receptores de manosa-6-fosfato en estos últimos (Brown et al., 1986).

La autofagia es un proceso por el que elementos intracelulares, como organelas dañadas, proteínas mal plegadas, etc., alcanzan el lisosoma por diferentes rutas que configuran un proceso global de autodigestión (Eskelinen and Saftig, 2009). En el proceso se forman inicialmente autofagosomas (Lamb et al., 2013), que se fusionan más tarde con los lisosomas, en cuyo interior vuelcan su contenido (Reggiori and Ungermann, 2017). La autofagia permite a la célula deshacerse de elementos que ha de desechar. Además, con los productos resultantes suministra nutrientes para mantener funciones vitales, lo que contribuye a la homeostasis celular. En conjunto, se ha podido determinar que la autofagia es un proceso estimulado por el ayuno (Russell et al., 2014) que promueve el mantenimiento y la renovación de elementos celulares, repercutiendo en la salud y la longevidad (Eskelinen and Saftig, 2009; Levine and Kroemer, 2008).

Una tercera ruta de acción catabólica en la que participa el lisosoma es la fagocitosis, una función reservada a células especializadas, como los macrófagos, que son capaces de englobar grandes partículas como bacterias y restos celulares que necesitan ser eliminados. En una secuencia de acontecimientos similar a las rutas anteriores, estas macromoléculas son englobadas en vacuolas fagocíticas (fagosomas) que posteriormente se fusionan con el lisosoma y pueden así ser degradadas (Stuart and Ezekowitz, 2005). Los elementos que forman parte de endosomas, autofagosomas o fagosomas, van a modificar la composición final de los lisosomas con los que se fusionan en los procesos catabólicos.



**Figura 1. Funciones del lisosoma.**

Las moléculas entran en la célula en vesículas endocíticas y se fusionan con el endosoma temprano (EE). Los componentes de la membrana son reciclados y el EE madura a endosoma tardío (LE). Las hidrolasas ácidas con el receptor de M6P salen del Golgi para fusionarse con el LE, lo que da lugar a la formación del lisosoma. En la autofagia, las organelas internas que han de ser renovadas son englobadas por fragmentos de la membrana del retículo endoplásmico, lo que da lugar a los autofagosomas. En la fagocitosis, partículas grandes (como bacterias) son englobadas en el interior de vacuolas fagocíticas o fagosomas. Tanto el autofagosoma como los fagosomas se fusionan con el lisosoma dando lugar al fagolisosoma, en cuyo interior se producirá la digestión.

Más allá de su papel en el catabolismo celular, en los últimos años se han ido descubriendo nuevas funciones de los lisosomas, por lo que actualmente son considerados elementos centrales en el control del metabolismo celular, y que funcionan como una plataforma para la detección de nutrientes y la transducción de señales metabólicas (Lim and Zoncu, 2016; Lawrence and Zoncu, 2019).

Las tareas que realizan los lisosomas presentan una coordinación transcripcional regulada por el factor de transcripción EB (TFEB). El estudio de las regiones promotoras de 96 genes lisosomales permitió identificar una secuencia palindrómica de 10 bases en 68 de ellos. Ese motivo, llamado CLEAR (*Coordinated Lysosomal Expression and Regulation Element*), es reconocido por TFEB, al que se une (Sardiello et al., 2009). En situaciones en las que se altera la función lisosomal, TFEB es transportado del citoplasma al núcleo celular, uniéndose a CLEAR y activando los genes diana. La sobreexpresión de TFEB induce la biogénesis lisosomal y aumenta la degradación de moléculas complejas a través de la autofagia, es decir, ambas acciones están reguladas por TFEB (Settembre et al., 2011).

Otro elemento de regulación transcripcional lo ejerce el complejo mTORC1, situado en la membrana lisosomal. Las actividades de mTOR y el lisosoma están interconectadas, de modo que en un

estado en el que las células están bien nutridas, mTORC1 activo fosforila TFEB, lo que impide que alcance el núcleo celular, y no active la biogénesis lisosomal o la autofagia. Por el contrario, en situaciones de falta de nutrientes, la inactivación de mTORC1 permite que TFEB sea transportado al núcleo y active la expresión de los genes lisosomales, lo que estimula la autofagia e induce la biogénesis lisosomal (Martina et al., 2012; Roczniak et al., 2012).

### 1.1.3. Enfermedades relacionadas con el lisosoma

El concepto de enfermedad de depósito lisosomal (EDL) comenzó a utilizarse a partir de 1963, cuando se estableció la relación entre la deficiencia de alfa-glucosidasa lisosomal y la acumulación de glucógeno en el interior de los lisosomas de los pacientes con enfermedad de Pompe (Hers, 1963).

Las EDL son un grupo de unas 70 alteraciones monogénicas que se transmiten en la mayoría de los casos con patrón de herencia autosómico recesivo. Aunque individualmente son raras, en conjunto afectan a 1 de cada 5000 nacidos vivos (Platt et al., 2018). La mayoría de estas alteraciones provocan el mal funcionamiento de solo una proteína, que puede ser una de las hidrolasas lisosomales o de las proteínas de su membrana. Estas alteraciones suelen provocar la acumulación de material no degradado en el interior del lisosoma.

Se han propuesto varias maneras de clasificar estas enfermedades, pero la más habitual es la que las agrupa en función de la naturaleza de los sustratos acumulados: Enfermedades de almacenamiento de glucógeno (glucogenosis), lipidosis, esfingolipidosis, mucopolisacaridosis, etc. (Boustany, 2013).

El proceso primario de acumulación de sustratos desencadena alteraciones secundarias que condicionan el mal funcionamiento de la organela. La alteración de rutas metabólicas secundarias repercuten sobre el grado de afectación funcional global, no solo de los lisosomas, sino de la célula en su conjunto (Futerman and Van Meer, 2004). Como ejemplo, el papel primordial que realiza el lisosoma en las últimas etapas de la autofagia hace que este proceso se vea afectado de manera particular en las EDL, aunque el grado de afectación varía según la naturaleza de las macromoléculas que se acumulan en el lisosoma y, por tanto, varía en cada EDL (Seranova et al., 2017).

Desde el punto de vista clínico, la edad de inicio de la enfermedad, la severidad de los síntomas, la evolución, etc., de cada una de estas patologías, vendrá condicionado por la naturaleza de la alteración genética que provoca la enfermedad, el grado de afectación que esta alteración ocasione sobre la proteína, la naturaleza y magnitud del sustrato acumulado y, en último término, del tipo celular afectado (Platt et al., 2012).

### 1.2. Glucogenosis tipo II. Enfermedad de Pompe

#### 1.2.1. Aspectos históricos

El epónimo utilizado para identificar esta enfermedad tiene su origen en el patólogo holandés Joannes Cassianus Pompe (1901-1945), quien realizó la primera descripción clínica, tras llevar a cabo la autopsia de una niña de 7 meses fallecida por cardiopatía hipertrófica idiopática, junto con hipotonía muscular generalizada (Pompe, 1932).

Como se ha mencionado en el apartado anterior, la primera conexión entre el déficit de alfa glucosidasa lisosomal, la acumulación de glucógeno y la enfermedad de Pompe se llevó a cabo en el año 1963 (Hers, 1963). En ese trabajo se estableció la influencia del pH sobre la actividad de la alfa glucosidasa, determinándose que la enzima mostraba actividad máxima a pH 4. Cinco pacientes con la forma infantil de la enfermedad fueron estudiados. Se pudo determinar la acumulación tisular de glucógeno en corazón, músculo e hígado. Todos los pacientes mostraban ausencia total de la alfa glucosidasa en biopsia de tejido cardiaco o muscular.

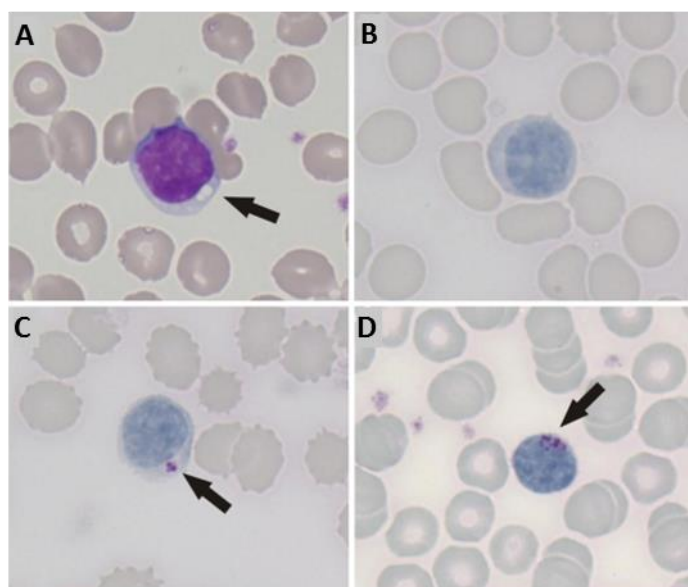
El aislamiento del gen y la descripción de sus características estructurales se realizaron en 1991 (Martiniuk et al., 1991). Los autores prepararon una genoteca genómica, utilizando Lambda como vector. Insertaron en ella fragmentos de entre 8 y 25 Kpb, generados por digestión parcial de DNA genómico humano con la endonucleasa de restricción EcoRI. El rastreo de la genoteca se llevó a cabo con sondas de cDNA (Martiniuk et al., 1986).

#### 1.2.2. Aspectos clínicos de la enfermedad

La Glucogenosis tipo II (GSDII), o enfermedad de Pompe (MIM 232300), presenta un espectro clínico muy amplio, de modo que tanto la severidad como la edad de aparición pueden variar de manera muy significativa. El grado de actividad residual de alfa glucosidasa determina en gran medida el grado de afectación, ya que condiciona que exista mayor o menor acumulación de glucógeno en los distintos tejidos, lo que juega un papel fundamental en la fisiopatología de la enfermedad. La correlación genotipo-fenotipo en esta enfermedad no siempre resulta sencilla de determinar, ya que se postula que puedan existir factores que condicionen el cuadro clínico final. En este sentido, se ha descrito la existencia de pacientes que, a pesar de compartir el mismo genotipo, o genotipos muy similares, manifiestan la enfermedad con grados de severidad diferentes, pudiendo aparecer los primeros síntomas en edades también muy dispares (Kroos et al., 2007; Van der Ploeg and Reuser, 2008).

La forma clásica, o infantil, es la forma más severa. El comienzo del cuadro clínico se produce en los primeros 4 meses de vida, con hipotonía generalizada, dificultad respiratoria, miocardiopatía

hipertrófica y hepatomegalia generada, esta última por estasis venosa, entre otros síntomas. La cuantificación de creatinfosfoquinasa (CPK) muestra niveles elevados en sangre periférica. En biopsias musculares se observan acumulaciones vacuolares Ácido Peryódico-Schiff (PAS) positivas, ocasionadas por el almacenamiento de glucógeno en el interior de los lisosomas (Bembi et al., 2008). Hallazgos similares se pueden obtener en muestras de sangre periférica, en las que se pueden observar vacuolas con las mismas características en el interior de los linfocitos (Hagemans et al., 2010) (**Figura 2**).



**Figura 2. Microfotografías de extensiones de sangre de pacientes con enfermedad de Pompe y control sano.**

**A:** Tinción de Giemsa de un paciente Pompe adulto donde se muestra un linfocito con 2 vacuolas distintas (flecha).  
**B,C,D:** Tinción Ácido Peryódico-Schiff (PAS). **B:** Extensión de sangre de un control sano; **C:** Extensión de sangre de paciente Pompe adulto donde se señala un linfocito con inclusiones PAS + (;  
**D:** Extensión de sangre de paciente Pompe infantil donde se señala un linfocito con múltiples inclusiones PAS + (Imagen tomada de Hagemans et al., 2010)

Las alteraciones moleculares que causan la forma infantil de la enfermedad corresponden con mutaciones muy deletéreas que provocan la ausencia total, o casi total, de actividad de GAA lisosomal (<http://www.pompevariantdatabase.nl>). Por un lado, están aquellas mutaciones que alteran la síntesis de la enzima, por lo que provocan un fenotipo *CRIM* negativo (*Cross Reactive Immunologic Material*); por otro lado, aquellas mutaciones que alteran la funcionalidad del enzima, pero no impiden su síntesis, lo que genera el fenotipo *CRIM* positivo (Reuser et al., 1978). En conjunto, se considera que alrededor del 20% de los pacientes con esta forma de la enfermedad son *CRIM* negativos (Bali et al., 2012). La evolución de la forma infantil de la enfermedad lleva a la muerte del paciente antes del segundo año de vida (Kishnani et al., 2006).

El fenotipo tardío, también conocido como forma juvenil o del adulto, es más leve y presenta una gran variabilidad en aspectos como la edad de aparición, el grado de severidad, pronóstico, etc. La afectación principal es la disfunción del músculo esquelético que, dependiendo del grado de severidad, puede generar afectación pulmonar con insuficiencia respiratoria progresiva, y comprometer la vida del paciente. Esta forma de enfermedad no implica compromiso cardíaco (Hagemans et al., 2005). La actividad enzimática determinada en sangre periférica se encuentra entre el 2 y el 40% de la actividad



normal (Bembi et al., 2008). Asimismo, se suelen presentar niveles elevados de CPK en sangre periférica. Las alteraciones moleculares que provocan las formas tardías son menos lesivas que las que provocan la forma infantil.

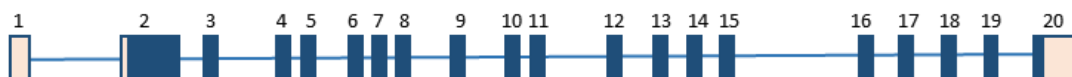
### 1.2.3. Bases moleculares de la enfermedad de Pompe

#### 1.2.3.1. Estructura del gen *GAA*

La enzima  $\alpha$ -glucosidasa ácida lisosomal es el producto de la expresión del gen *GAA*, localizado en el brazo largo del cromosoma 17 (17q25.2-q25.3). En este trabajo de tesis se utiliza como genoma de referencia la versión GRCh38 (<https://www.ncbi.nlm.nih.gov/genome>). La localización genómica de *GAA*, es: chr17:80101704-80119879. El tamaño del gen es de 18.160 pares de bases y contiene 20 exones, 19 de ellos codificantes. (**Figura 3A**)

La región promotora reúne características de un gen constitutivo. El exón 1 representa la mayoría de la región 5' no traducida del gen (5'UTR). El ATG iniciador se encuentra en el exón 2, a 32 pares de bases del comienzo del propio exón. Todos los extremos intrón-exón siguen la regla *AG/GT*, excepto el sitio de procesamiento donador en el límite entre exón e intrón 19, que tiene *GC* en lugar de *GT* en su secuencia (**Figura 3B**). El codón de parada, TAG, está localizado en los primeros 60 nucleótidos del exón 20, que tiene 610 nucleótidos de longitud total (Hoefsloot et al., 1990). ([https://www.ensembl.org/Homo\\_sapiens/Transcript/Exons?db=core;g=ENSG00000171298;r=17:80101556-80119879;t=ENST00000302262](https://www.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000171298;r=17:80101556-80119879;t=ENST00000302262)).

**A** Estructura genómica del gen *GAA*



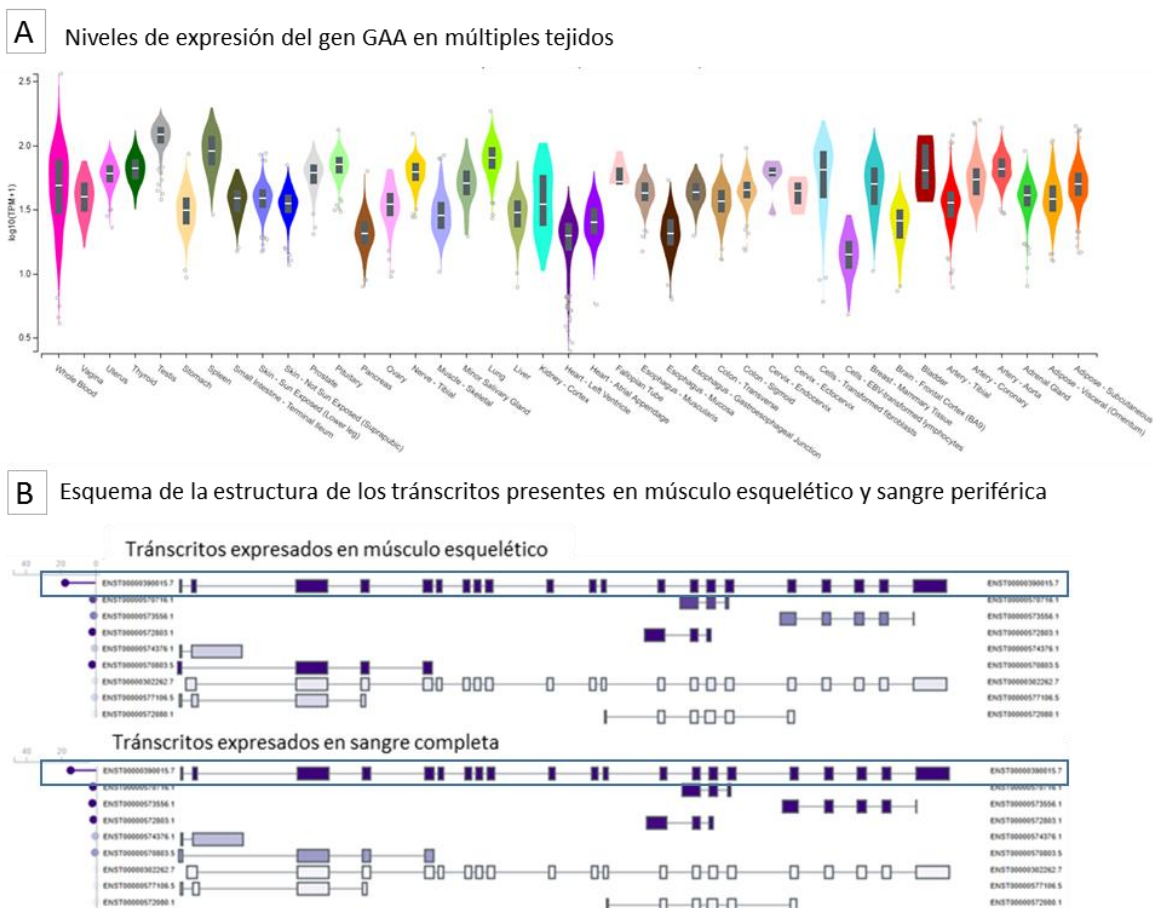
**B** Secuencias de los límites intrón/exón

EXÓN	Longitud Exón	5'donador	INTRÓN	Longitud intrón (pb)	3' <i>acceptor</i>
1	335	CGGgtgaga	1	2664	tctccgcagGCC
2	578	ACGgtgggc	2	616	tctcttctagATC
3	146	GCTgtgagt	3	1662	tgtccgcagGCT
4	166	ACGgtacag	4	77	catgtcccagCCC
5	97	TGGgtaagc	5	393	tcccttccagATG
6	120	TGGgtaggg	6	79	tggcctgcagGAT
7	119	CTGgtgagt	7	89	gtggctgcagGAC
8	132	GTGgtgtgt	8	1116	ctcttcccagGAT
9	111	AAGgtaggg	9	671	cgttgtccagGTA
10	114	ATTgtaagt	10	100	tctcttgcagGAC
11	85	CTGgtcagc	11	957	cctcttccagGGG
12	118	CAGgtgagg	12	477	accaccccagGGC
13	134	CAGgtgagc	13	164	gccctcccagAAA
14	152	CTGgtaggg	14	190	tgccttgcagCCC
15	149	GGAgtagt	15	3601	cccttgcagGTT
16	142	ACGgtgagt	16	490	ctccctccagGTG
17	150	CAGgtacct	17	443	cccttccagGGC
18	165	AATgtgagt	18	295	ctcggcccagAAC
19	153	AAGgcaaga	19	466	tcttttccagGTC
20	609				

**Figura 3. A: Estructura genómica del gen *GAA*.** El gen *GAA* consta de 20 exones 19 de ellos codificantes (cuadros azules). Los cuadros salmón corresponden a las UTRs. **B: Secuencias de los límites intrón/exón.** Secuencias detalladas de las zonas 5'donador y 3'aceptor de todos los exones del gen *GAA* que están implicadas en el procesamiento del mRNA

1.2.3.2. Perfil de expresión tisular del gen *GAA*

Recientemente, el análisis de los datos de expresión de ARN de 1641 muestras de 43 tejidos obtenidas de 175 individuos, como parte del proyecto “*Genotype-Tissue Expression*” (*GTEx*), ha permitido observar que el gen *GAA* se expresa de forma ubicua en todos los tejidos del organismo, con claras diferencias en los niveles de expresión entre ellos (*GTEx Consortium, 2015*) (**Figura 4A**). Se han identificado diferentes transcritos como resultado de la expresión del gen, pero el mRNA de referencia (NM\_000152.3) tiene una longitud de 3.751 pb y da lugar a la proteína funcional *GAA* de 952 aminoácidos ([https://www.ncbi.nlm.nih.gov/nucore/NM\\_000152.3](https://www.ncbi.nlm.nih.gov/nucore/NM_000152.3)). Las características de los transcritos presentes en músculo esquelético y sangre periférica no presentan diferencias entre sí (**Figura 4B**), por lo que cabe esperar que cada una de las características funcionales que se deriven de la secuencia genómica del gen *GAA*, queden reflejadas de manera similar en los productos de la expresión del gen en ambos tejidos.



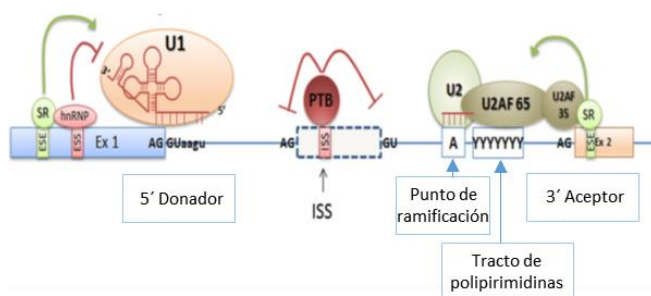
**Figura 4. A: Niveles de expresión GAA en distintos tejidos. B: Transcritos expresados en músculo esquelético y en sangre periférica.**

Los transcritos que aparecen mayoritariamente expresados en cada tejido están marcados en azul más intenso, el transcrito principal está recuadrado (Imágenes adaptadas GTEx Consortium).

### 1.2.3.3. Procesamiento del mRNA

El procesamiento del mRNA es un paso crítico en el que se eliminan los intrones para así obtener un mRNA maduro, capaz de ser traducido a proteína por la maquinaria celular. Es un proceso complejo que depende de secuencias consenso que definen la unión exón-intrón, así como secuencias regulatorias que serán reconocidas por la maquinaria implicada en este proceso. Dentro de estas secuencias consenso podemos incluir el 5' donador (5' ss) y el 3' aceptor (3' ss), altamente conservadas y normalmente definidas por un *GU* y *AG*, además del punto de ramificación y el tracto de polipirimidinas que, junto con el donador y el aceptor, son reconocidos por la maquinaria implicada en el procesamiento. Existen otros elementos reguladores del proceso, localizados fuera de las regiones canónicas de procesamiento, que actúan promoviendo el *splicing*, son los (*ESE-exonic splicing enhancers* y *ISE-intronic splicing enhancers*) o silenciándolo (*ESS-exonic splicing silencers* y *ISS-intronic splicing silencers*) (Abramowicz and Gos, 2018) (Figura 5).

## INTRODUCCIÓN

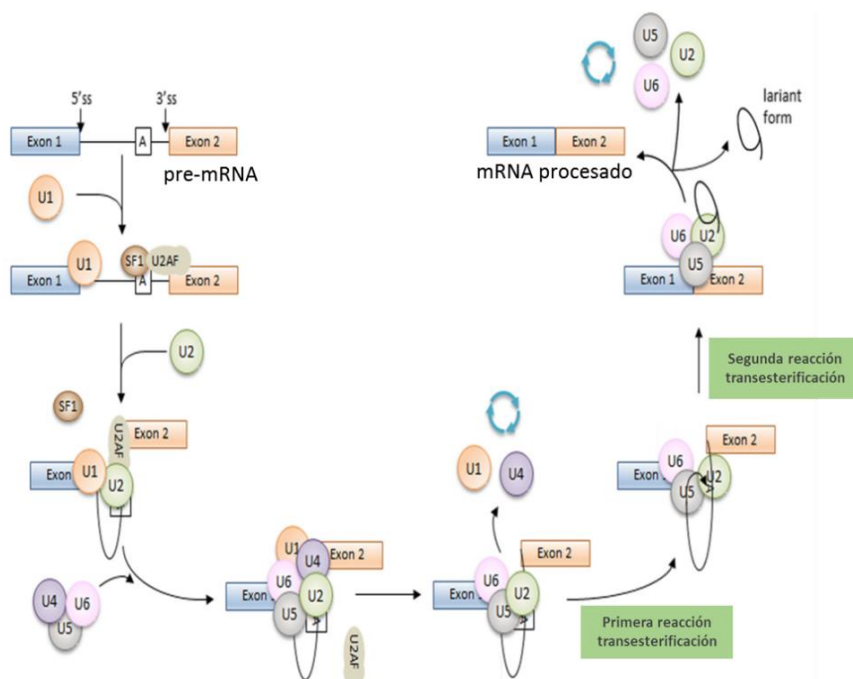


Abramowicz A. et. al\_J Appl Genetics (2018) 59:253–268

**Figura 5: Secuencias consenso implicadas en el procesamiento del mRNA.**

Existen unas secuencias altamente conservadas en las zonas flanqueantes del intrón. En la región 5'SS aparece casi invariablemente un GU y en la región 3'SS un AG que serán reconocidas por la ribonucleoproteína (snRNP) U1 y la U2AF35 respectivamente. El punto de ramificación, que es una secuencia más variable pero que siempre presenta una adenosina (A) necesaria para la unión de la snRNP U2, y el tracto de polipirimidinas, al que se unirá la snRNP U2AF65. Existen otras secuencias reguladoras que promueven el procesamiento, como las secuencias ESE y ISE, o que lo silencian como ESS y los ISS (Imagen adaptada de Abramowicz and Gos, 2018).

La eliminación de los intrones se lleva a cabo por el *spliceosoma*, que implica un conjunto de 5 ribonucleoproteínas (snRNP U1, U2, U4, U5 y U6) a las que se unirán alrededor de 300 proteínas diferentes. Inicialmente, la snRNP U1 reconoce el sitio 5'donador, la subunidad 35 de la U2 (U2AF35) reconoce el sitio 3' aceptor, la subunidad 65 de la U2 (U2AF65) se une al tracto de polipirimidinas y la U2 se une al punto de ramificación. La unión de la U2 es una señal para el reclutamiento y unión del resto de las ribonucleoproteínas, la U4-U5-U6. Esto permite un cambio de conformación que lleva a la liberación de la U1 y U4, permitiendo así la interacción de la U6-U2 lo que inicia las dos reacciones de transesterificación necesarias para eliminar el intrón y ligar los exones (Abramowicz and Gos, 2018) (Figura 6).

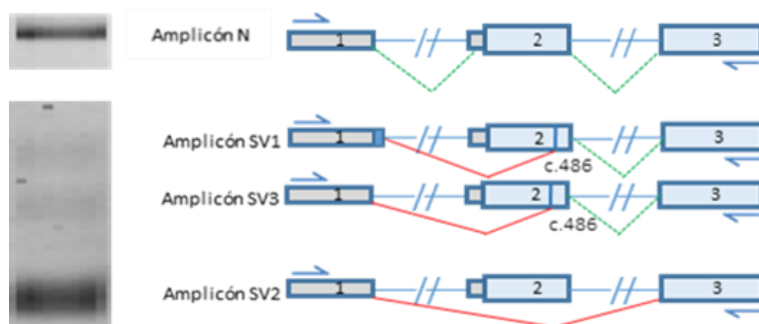


Abramowicz A. et. al\_J Appl Genetics (2018) 59:253–268

**Figura 6. Secuencia de eventos que se producen en el procesamiento del mRNA**

La U1 se une al 5'SS, la U2AF reconoce el sitio 3'SS y el tracto de polipirimidinas, y la U2 se une al punto de ramificación. La unión de la U2 es una señal para el reclutamiento del complejo U4-U5-U6. Esto permite la liberación de la U1 y la U4, iniciándose las dos reacciones de transesterificación necesarias para eliminar el intrón y ligar los exones. (Imagen adaptada de Abramowicz and Gos, 2018).

El procesamiento del mRNA del gen *GAA* muestra algunas particularidades en relación al exón 2, que se han puesto de manifiesto en varios trabajos de investigación (Hule et al., 1994; Boerkoel et al., 1995; Dardis et al., 2013; Bergsma et al., 2015) y que se han observado también en muestras de controles utilizados en este trabajo de tesis. El exón 2 en individuos sanos presenta varias alternativas de procesamiento, que dan lugar a 4 variantes diferentes. En la **Figura 7** se muestra la intensidad relativa de las diferentes poblaciones generadas en el procesamiento normal del exón 2. La variante N sería la única con toda la secuencia codificante capaz de producir la enzima *GAA*.



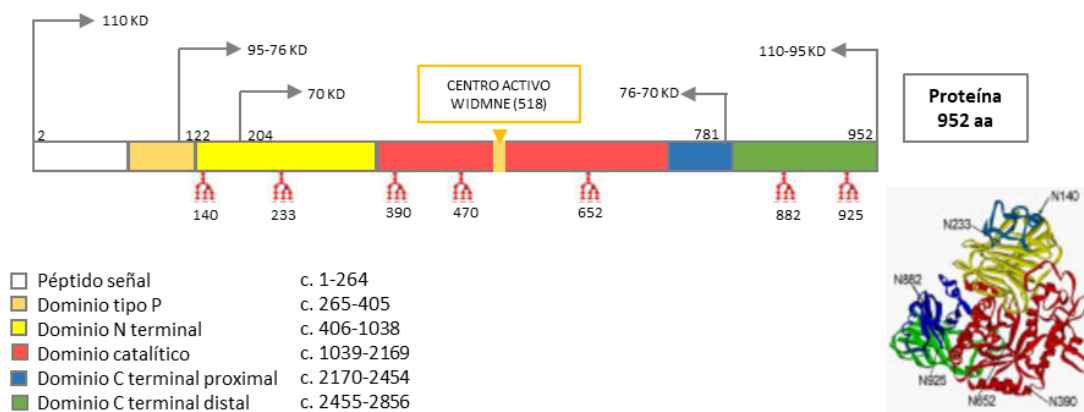
**Figura 7. Variantes de procesamiento del exón 2 observadas en *GAA* normal.**

Electroforesis en gel de agarosa de los fragmentos obtenidos en la amplificación por RT-PCR del exón 2. El esquema que acompaña cada banda, muestra el mecanismo molecular por el que se produce. La variante representada por el amplicón N supone el procesamiento normal e incluye el exón 2 completo, utilizando para ello las secuencias canónicas 5'ss y 3'ss en los intrones 1 y 2. En la variante SV1 se utiliza un sitio criptico donador situado a 36 nucleótidos del extremo 3' del exón 1. Además, no se reconoce el 3'ss del intrón 1 y, en su lugar, se utiliza un sitio criptico 3'ss, localizado dentro del propio exón 2, en c.486, lo que supone que la mayor parte del exón 2 es eliminado, junto con los intrones 1 y 2. En la variante SV3 se utiliza el 5'ss correcto del intrón 1, pero no se utiliza el 3'ss del intrón 1, sino que también es utilizado el sitio criptico 3'ss, c.486, por lo que el exón 2 es eliminado parcialmente. En ambas variantes, solo permanece en el mRNA maduro un fragmento del exón 2 desde c.486 hasta el extremo 3' del mismo. En la variante SV2, la maquinaria de *splicing* no utiliza el 3'ss del intrón 1, ni el 5'ss del intrón 2, lo que supone la eliminación en bloque de intrón 1, exón 2 e intrón 2, quedando un mRNA maduro en el que se ha producido la pérdida total del exón 2.

#### 1.2.3.4. Síntesis de la proteína y procesamiento proteolítico

La enzima alfa-glucosidasa lisosomal (*GAA*) (EC 3.2.1.20) cataliza la hidrólisis de los enlaces alfa-1,4 y alfa-1,6 en sus sustratos naturales glucógeno, maltosa e isomaltosa, generando alfa-glucosa como producto. Su pH óptimo está entre 4 y 5. Esta proteína es sintetizada como un precursor de 952 aminoácidos, con un péptido señal que facilita su transporte al interior del retículo endoplásmico (RE) (Hoefsloot et al., 1988). La estructura primaria de la proteína tiene 7 residuos potenciales de glicosilación, que están todos ocupados (Hermans et al., 1991). Estudios realizados sobre *GAA* purificada de placenta humana, han permitido determinar que tras la escisión del péptido señal de 27 residuos y la glicosilación en el RE, la proteína *GAA* de 110 kDa (residuos 28-952) se transfiere al Golgi, donde sufre escisiones proteolíticas adicionales y se modifican los glicanos incorporados previamente. Entre estos cambios, se

encuentra la adición de residuos de manosa -6-fosfato (Park et al., 2018). Finalmente, la proteína queda constituida por un péptido de 70 kDa (Moreland et al., 2005) (**Figura 8**).



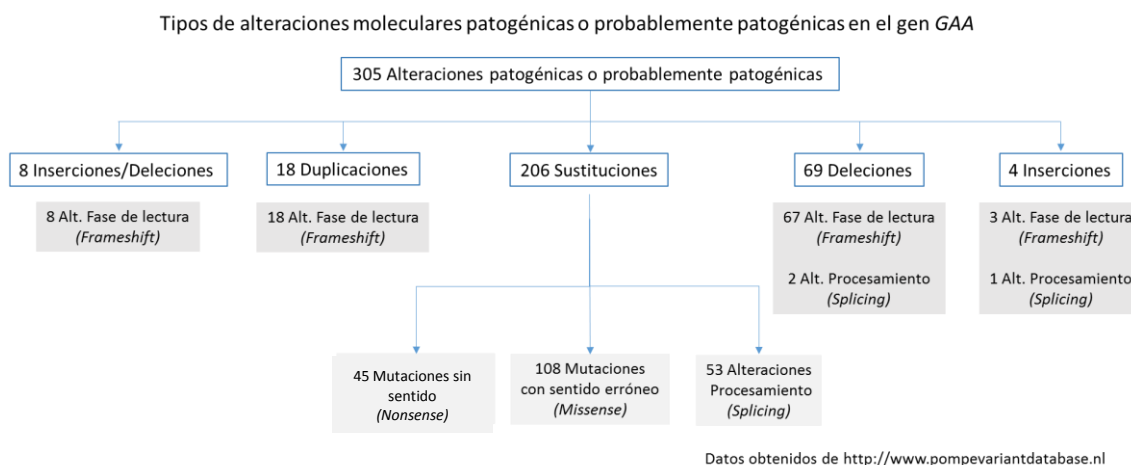
**Figura 8. Estructura de la proteína GAA**

La proteína GAA consta de 952 aminoácidos y 6 dominios. Esta sufre un proceso de glicosilación en 7 residuos (🍷) dando lugar a una proteína de 110kDa. Posteriormente, sufre varios procesamientos proteolíticos (marcados con las flechas) para dar lugar a distintas formas intermedias de la proteína (95kDa, 76kDa) y finalmente, a la forma madura (70kDa) (Imagen adaptada de Kroos et al., 2012).

La determinación de la estructura tridimensional de la enzima GAA humana mediante cristalografía de Rayos-X ha permitido identificar 5 dominios estructurales, conocer la existencia de cinco puentes disulfuro formados entre residuos cisteína, y datos adicionales que permitirán determinar mejor las consecuencias de las alteraciones que afectan al gen GAA (**Figura 8**). Asimismo, se han podido definir particularidades como que los trece residuos D282, W376, D404, L405, I441, W481, W516, D518, M519, R600, D616, W618, F649, L650 y H674, conforman el centro activo (Deming et al., 2017). Estudios previos realizados con mutagénesis dirigida habían determinado el papel crítico de W516 y D518 en el centro catalítico del enzima (Hermans et al., 1991).

### 1.2.3.5. Mutaciones descritas en el gen GAA.

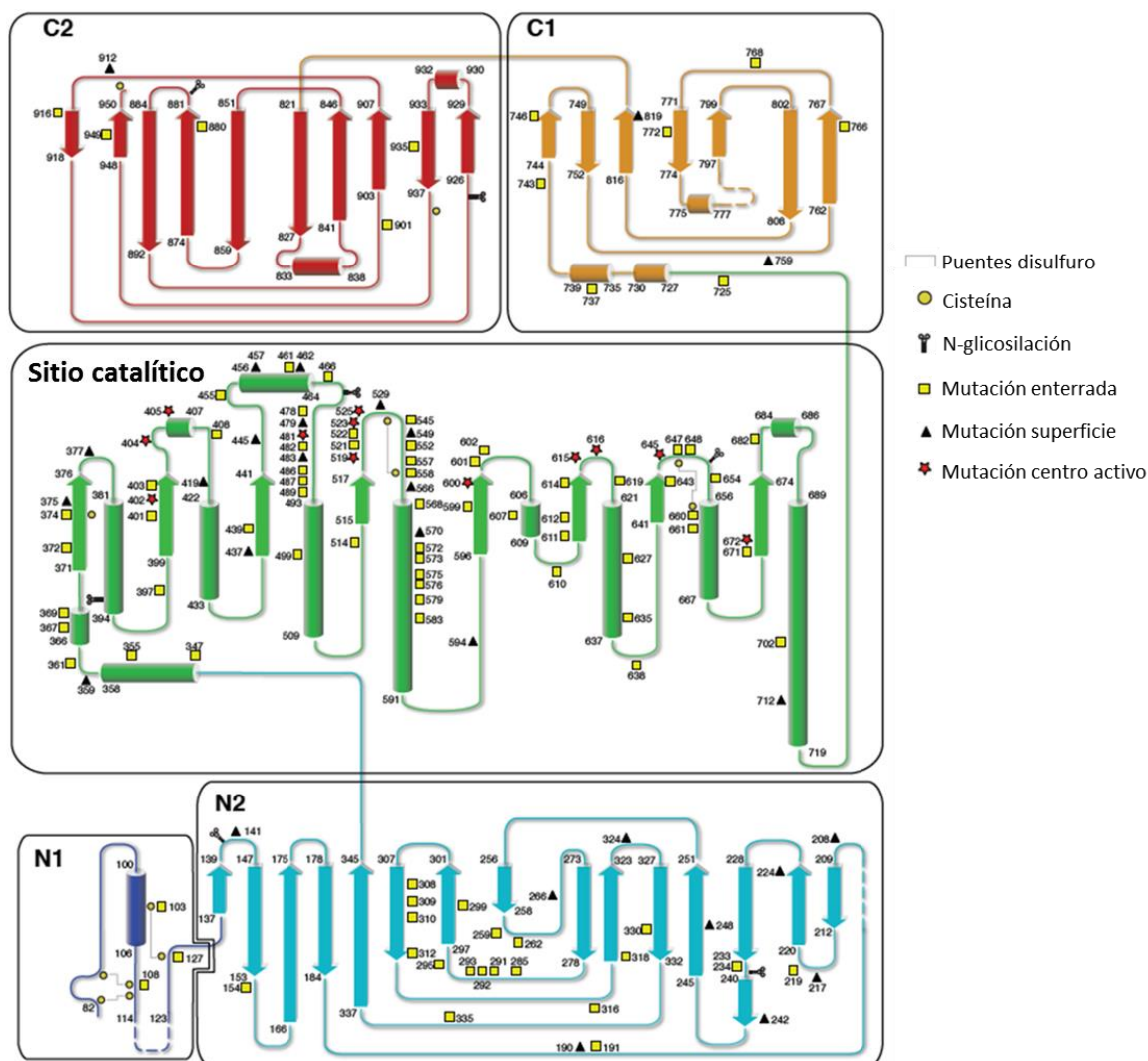
Existe una base de datos desarrollada por el Centro Médico Erasmus, en la Universidad de Rotterdam (<http://www.pompevariantdatabase.nl>), en la que están recogidas las variantes descritas hasta el momento en la enfermedad. Hay identificados 561 eventos moleculares únicos; entre variantes patogénicas y probablemente patogénicas, suman 305; variantes de significado desconocido (VSD), son 214; variantes benignas y probablemente benignas, se han identificado 42. El impacto funcional sobre el RNA o la proteína de las variantes patogénicas o probablemente patogénicas queda representado de manera concisa en la **Figura 9**, donde se reflejan los datos recopilados en la mencionada base de datos.



**Figura 9. Tipos de alteraciones moleculares descritas como patogénicas o probablemente patogénica en la base de datos del Pompe Center.**

El espectro de variantes identificadas en el gen *GAA* es muy heterogéneo. Se encuentran distribuidas a lo largo de toda la estructura del gen y, en consecuencia, se localizan en todos los dominios de la proteína (Deming et al., 2017) (Figura 10).

La mayoría de las variantes se han identificado en familias únicas o en un número muy pequeño de ellas. No obstante, la mutación c.-32-13T>G, localizada en el intrón 1, se observa con mucha frecuencia entre pacientes de origen caucásico con un fenotipo tardío (adulto) de la enfermedad. Inicialmente, fue descrita en un grupo de 41 pacientes procedentes de diferentes centros de Estados Unidos, estando presente en el genotipo del 68,2% de los que presentaban el fenotipo adulto, aunque existía entre ellos una enorme discrepancia en la edad de debut de la enfermedad, que estaba entre 17 y 73 años (Hule et al., 1994). Esa discrepancia ha sido posteriormente confirmada en otro grupo de 56 afectados con la mutación c.-32-13T>G en su genotipo (Kroos et al., 2007). En un estudio de 40 pacientes de origen italiano, se ha identificado como parte del genotipo en el 85% de los mismos (Montalvo et al., 2006). Más recientemente, otro grupo italiano ha encontrado esta mutación en 29 de 30 pacientes con la forma tardía de la enfermedad, lo que supone el 96,6% (Montagnese et al., 2015). Una frecuencia muy diferente fue encontrada en un grupo de 22 pacientes de origen español. En este caso, el grupo de estudio estaba formado por 16 pacientes con la forma tardía y 6 con la forma infantil. El 34,4% de los que presentaban la forma tardía, tenían la mutación c.-32-13T>G formando parte del genotipo enfermo. En este mismo estudio se indica que el 14% de todos los pacientes estudiados tenían la mutación c.1076-1G>C (Gort et al., 2007).



**Figura 10. Localización de variantes puntuales patológicas en la estructura de la proteína GAA.**

Las mutaciones encontradas en la estructura de la proteína GAA están localizadas en la superficie, en zonas profundas o en el sitio activo. N1 y N2 corresponden con los dominios tipo P y N terminal respectivamente. C1 y C2 corresponden con los dominios C proximal y C distal respectivamente (Imagen adaptada Deming et al., 2017)

En un trabajo realizado con 6 pacientes en los que el alelo c.-32-13T>G se presenta en homocigosidad, se pudo observar que presentaban formas clínicas de severidad variada y un debut clínico de la enfermedad que iba de los 12 a los 55 años (Musumeci et al., 2015).

La caracterización funcional de la mutación c.-32-13T>G, ha permitido determinar que propicia la exclusión del exón 2 durante el procesamiento del mRNA del gen *GAA*, provocando una modificación de la eficiencia global de su *splicing*. Como consecuencia, se rompe el balance entre las diferentes isoformas presentes en el procesamiento normal del exón 2 (Dardis et al., 2013).

Muy recientemente, se ha identificado la variante c.510C>T (p.D170D) que parece modificar la variabilidad fenotípica en los pacientes con la mutación c.-32-13T>G, de modo que incrementa el riesgo



de aparición clínica de la enfermedad en etapas más precoces, y con un grado de afectación algo superior en pacientes con la c.-32-13T>G (Bergsma et al., 2019).

La delección completa del exón 18 (c.2481+102\_2646+31del) es una lesión frecuente que provoca una grave alteración de la estructura del gen y una alteración de la fase de lectura. Esta mutación puede ser observada tanto en pacientes con la forma infantil como con la forma tardía, ya que lo que condicionará el fenotipo final del paciente será la mutación que la acompañe en *trans*. Es especialmente frecuente en población de origen holandés (Vanderkraan et al., 1994; Kroos et al., 1995).

Algunas mutaciones se muestran con mayor frecuencia en determinados grupos étnicos. Una de las mejor caracterizadas es la mutación c.2560C>T (p.R854\*), identificada como la mutación más frecuente en pacientes americanos de origen africano, según un estudio en el que 10 de 23 pacientes (43,4%) de origen africano tenían esa mutación en el genotipo enfermo (Becker et al., 1998).

Un tipo de alteración molecular que parece estar siendo infradiagnosticada es aquella que afecta al procesamiento (*splicing*) del mRNA. Esta anomalía puede estar, en cierto modo, justificada por la insuficiente información disponible aún acerca del papel que ejercen elementos que incrementan (*enhancers*) o suprimen (*silencers*) el *splicing* de los exones. Estos elementos se identifican como *ESE* (*exonic splicing enhancers*) o *ESS* (*exonic splicing silencers*), como se ha mencionado anteriormente. Un estudio realizado sobre 27.681 mutaciones exónicas causantes de enfermedad, concretamente, sustituciones que provocan un cambio de mensaje del tipo *missense* o *nonsense*, permitió determinar que aproximadamente el 25% de ellas alteraban la secuencia de alguno de los 238 hexámeros identificados como *ESE* o *ESS*, con la potencial consecuencia sobre la funcionalidad de los mismos (Sterne-Weiler et al., 2011).

En 2016, se llevó a cabo un estudio para evaluar la posible repercusión sobre el procesamiento del exón 10 del gen MLH1, de las 22 variantes que afectan a un único nucleótido (*SNVs*) identificadas en ese exón hasta ese momento. Quince de las variantes eran *missense*, 3 *nonsense* y 4 variantes sinónimas. Solo 9 estaban clasificadas como claramente patogénicas, y una claramente no patogénica. El estudio consistió en la generación de minigenes en los que se incluyó una variante en cada minigen, de modo que se pudiera evaluar aisladamente el efecto de cada una de ellas sobre el procesamiento del exón. El resultado del estudio concluyó con que el 77% de las *SNVs* alteraban el procesamiento del exón 10 (Soukarieh et al., 2016).

La dificultad para poder anticipar el efecto de muchas de las *SNVs* presentes en los exones o en zonas intrónicas, lleva a desarrollar abordajes que permiten identificar posibles anomalías en el procesamiento del mRNA, que pueden pasar desapercibidos con las estrategias de análisis molecular

basadas únicamente en el análisis de secuencias sobre gDNA. En este sentido, se desarrolló un abordaje sobre muestras de RNA, aplicado al propio gen *GAA*, que permite determinar el efecto sobre el *splicing* de cualquier alteración presente en el gDNA (Bergsma et al., 2015). Al considerar esta aproximación una herramienta con capacidad para poder identificar las alteraciones de procesamiento que pueden pasar inadvertidas por medio de la secuenciación de DNA genómico, hemos incluido esta técnica en el estudio de los pacientes de este trabajo de tesis. En la actualidad, el abordaje debería incorporarse a cualquier estrategia de diagnóstico molecular en esta patología.

La secuenciación del gen *GAA* en un gran número de muestras por parte de numerosos grupos de investigación, ha permitido identificar hasta el momento 214 variantes consideradas de significado desconocido (VSD) (<http://www.pompevariantdatabase.nl>). En contraste con las variantes genéticas cuya vinculación con la enfermedad ha podido ser establecida, las VSD no tienen una relación bien definida con la patología. Estudios sucesivos en los que se puede establecer la cosegregación de las VSD en *GAA* con el fenotipo enfermo, y su presencia en genotipos en los que coincide en *trans* con mutaciones patogénicas, pueden contribuir a que estas variantes puedan ser consideradas como probablemente patogénicas. Asimismo, el desarrollo de nuevas aplicaciones bioinformáticas puede facilitar la reclasificación más adecuada de cada una de ellas. En algún caso, puede ser necesario realizar estudios funcionales capaces de dilucidar su posible implicación en la fisiopatología de la enfermedad.

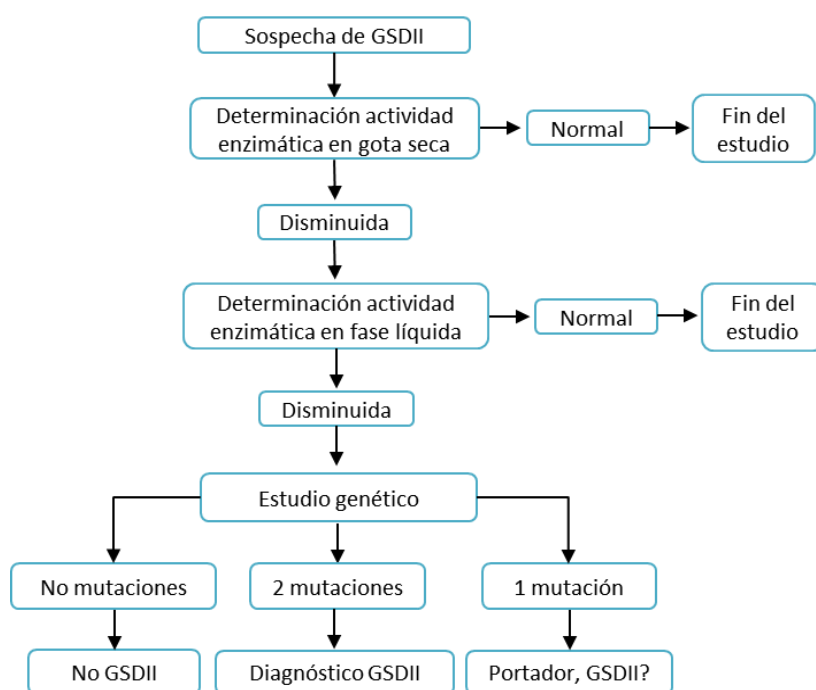
### 1.2.4. Diagnóstico de la enfermedad.

En este apartado hay que diferenciar entre la forma infantil, más severa, y la forma tardía de la enfermedad. En la forma infantil, la aparición de los síntomas en los primeros 3-4 meses de vida, junto con la hipotonía generalizada, insuficiencia respiratoria y la miocardiopatía severa, de evolución rápida y mortal, hacen imprescindible contar con un algoritmo diagnóstico preciso, que facilite un diagnóstico rápido y permita la inmediata instauración del tratamiento adecuado (Banugaria et al., 2013).

El diagnóstico de la forma tardía parece estar condicionado por varias circunstancias que pueden resultar determinantes: La enfermedad tiene una incidencia muy baja, estimada en 1/40.000 para población americana (Martiniuk et al., 1998), una cifra similar fue estimada posteriormente para población holandesa (Ausems et al., 1999), aunque un estudio de rastreo sistemático en recién nacidos realizado en Taiwan sobre 191.786 muestras, permitió identificar 16 pacientes con la enfermedad, lo que implica una incidencia de 1/11.987 (Liao et al., 2014). Por otro lado, la clínica que presentan los pacientes en las formas tardías es ambigua y poco específica, pudiendo ser fácilmente confundida con patologías musculares que presentan síntomas y signos clínicos similares. La edad de aparición es también muy variable, incluso entre pacientes que presentan genotipos similares, como se ha comentado en apartados anteriores. Para intentar mejorar el proceso diagnóstico de estos pacientes, se han elaborado una serie

de guías clínicas en las que se definen las características clínicas, bioquímicas y genéticas de la enfermedad, lo que permite optimizar los procedimientos diagnósticos (Kishnani et al., 2006; Barba-Romero et al., 2012). A pesar de este esfuerzo por introducir elementos de normalización en los protocolos diagnósticos, los pacientes sufren con frecuencia un retraso en el diagnóstico de la enfermedad que puede llegar a los 12 años, sobre todo en aquellos casos en los que la clínica es menos definida (Kishnani et al., 2013). El diagnóstico se basa en la adecuada interpretación de los rasgos clínicos, pruebas bioquímicas y en la demostración de una deficiencia en la actividad de GAA. La determinación del enzima puede realizarse en muestras de sangre en gota seca (DBS), muestra de sangre líquida, fibroblastos o biopsia muscular (Winchester et al., 2008). Una vez constatada la deficiencia enzimática, se indica el estudio genético confirmatorio del diagnóstico, siendo particularmente adecuado en aquellos casos en los que el perfil clínico o bioquímico no permite alcanzar un diagnóstico concluyente (Wokke et al., 1995). En nuestro laboratorio, este estudio molecular ha sido ampliado gracias al desarrollo de un MLPA (Multiplex Ligation-dependent Probe Amplification), con el fin de poder detectar cambios en el número de copias que no pueden ser detectados con el estudio de secuenciación clásica (Amiñoso et al., 2013) (Figura 11).

La mejora en las técnicas de secuenciación masiva promueven el uso de paneles con genes implicados en enfermedades neuromusculares con síntomas compartidos con la enfermedad de Pompe, lo que parece ser una opción válida para reducir los actuales retrasos en el diagnóstico (Lévesque et al., 2016).



**Figura 11. Algoritmo diagnóstico para la Enfermedad de Pompe**

Se muestran los distintos pasos que se realizan ante una sospecha de enfermedad de Pompe. Inicialmente, se realiza una cuantificación de la actividad enzimática de GAA en gota seca. Si es normal se descarta la enfermedad, pero si está disminuida, se realiza una nueva medición mediante el mismo ensayo enzimático en fase líquida. Si se confirma la deficiencia, se realiza el estudio genético con el fin de establecer las mutaciones causantes de la enfermedad.

Para ayudar al diagnóstico y, ocasionalmente, monitorizar la respuesta al tratamiento, ha sido propuesto el tetrasacárido de glucosa (Glc4) como un biomarcador específico (Kishnani et al., 2006). No obstante, Glc4 se ha visto elevado en otras patologías de depósito lisosomal como la *GSD* tipo III (*GSD* por deficiencia de enzima desramificante del glucógeno), *GSD* tipo VI (*GSD* por deficiencia de glucógeno fosforilasa hepática) y en otras enfermedades musculares, como la distrofia muscular de Duchenne (An et al., 2000). Por todo ello, se considera que no existe un biomarcador objetivo y fiable que permita la evaluación o la predicción de la respuesta individual al tratamiento de estos pacientes.

### 1.2.5. Tratamiento.

Actualmente, el único tratamiento aprobado para esta enfermedad es la terapia de reemplazamiento enzimático (*ERT*), que supone la administración intravenosa de glucosidasa ácida lisosomal humana recombinante (*rhGAA*), obtenida de un cultivo de células de ovario de hámster chino (Kishnani et al., 2006). Esta terapia se ha mostrado eficaz en los paciente infantiles con la forma clásica de la enfermedad, prolongando la supervivencia y mejorando la afectación cardiaca, sobre todo cuando se administra muy tempranamente (Koeberl et al., 2007). En los pacientes con la forma adulta, también se ha observado una mejoría de los resultados obtenidos en las pruebas en las que se evalúan diferentes funciones y capacidades, como resistencia al ejercicio o capacidad funcional respiratoria, especialmente en aquellos que han recibido al menos 18 meses de tratamiento (Van der Ploeg et al., 2010). No obstante, este tratamiento conlleva la posible aparición de reacciones adversas asociadas con la infusión de la enzima, incluyendo reacciones anafilácticas severas, urticaria, hiperhidrosis, vómitos o el aumento de la presión arterial (Van der Ploeg et al., 2010). Además, también se ha observado una ineficiente biodistribución de la enzima recombinante, sobre todo a nivel de músculo esquelético, que ha hecho que se estén desarrollando nuevas investigaciones destinadas a descubrir nuevos tratamientos eficaces para esta patología (Lieberman et al., 2012). Estas nuevas terapias van desde el tratamiento mediante RNAs de interferencia dirigidos a reducir la expresión de las principales enzimas encargadas de la producción de glucógeno, con el fin de llevar a cabo una terapia de reducción de sustrato (Douillard-Guilloux et al., 2008). El uso de chaperonas farmacológicas, capaces de estabilizar la *rhGAA* incrementando así su vida media (Porto et al., 2009). Promover la sobreexpresión del factor de transcripción EB (*TFEB*), capaz de regular la biogénesis de lisosomas y la autofagia, induciendo así la generación de nuevos lisosomas e incrementando el número de autofagosomas (Spampanato et al., 2013) y el uso de la terapia génica con virus adeno-asociados para la expresión en músculo de *GAA* (Smith et al., 2013; Corti et al., 2017), y de *TFEB* (Gatto et al., 2017).

### 1.3. Análisis del transcriptoma mediante RNA-Seq.

El transcriptoma se define como el conjunto completo de moléculas de ácido ribonucleico (RNA) expresadas en un tejido o tipo celular. Es decir, el transcriptoma es el producto inicial de la expresión del genoma. Representa un enlace clave entre el DNA y el fenotipo celular y sus características funcionales. Incluye las diferentes especies de RNA presentes en la célula: mRNA, rRNA, tRNA y el resto de RNAs no codificantes. Puede ofrecer información valiosa sobre los procesos biológicos en los que el tejido o tipo celular está involucrado, así como su respuesta a diferentes estímulos o situaciones que pueden alterar su funcionalidad.

La tecnología para el estudio del transcriptoma cuenta con varios métodos: tecnología de microarrays (Eisen et al., 1998), análisis en serie de la expresión génica (SAGE)(Boon et al., 2002) y, más recientemente, la secuenciación de ARN (RNA-Seq) (Hrdlickova et al., 2017).

La enfermedad de Pompe afecta principalmente al tejido muscular, pero otros tejidos muestran también signos de afectación, como es el caso de la sangre periférica donde la acumulación de vacuolas de glucógeno en leucocitos ha sido utilizado como un recurso diagnóstico de la enfermedad (Hagemans et al., 2010).

Hasta el momento, los 2 únicos estudios de transcriptoma realizados con pacientes de enfermedad de Pompe se llevaron a cabo en 2012 mediante *arrays*, en ambos casos con muestras de músculo de pacientes con la forma infantil de la enfermedad y como muestras control se utilizaron biopsias musculares de niños sanos con edades similares a las de los pacientes (Palermo et al., 2012). El primer estudio se centró en el análisis del transcriptoma de muestras de cuádriceps de 11 pacientes, comparando el perfil de expresión antes de comenzar el tratamiento y tras ser tratados con la enzima recombinante Myozyme® durante 52 semanas. Como grupo control se utilizaron biopsias de 7 niños sanos. Esta aproximación permitió identificar perfiles de expresión génica diferentes en pacientes que respondían favorablemente al tratamiento, de aquellos que no lo hacían. En el segundo estudio, publicado en el mismo artículo, se analizaron muestras de bíceps de 9 pacientes no tratados y 9 controles sanos. Los resultados de los estudios permitieron identificar alteraciones en los perfiles de expresión de numerosas rutas, especialmente rutas relacionadas con la respuesta inmune y con el desarrollo del músculo, y más de 900 genes mostraban perfiles de expresión alterados.

El análisis del transcriptoma en pacientes con enfermedad de Pompe, puede ser una vía para conocer la manera en la que la enfermedad está modificando la funcionalidad celular de cada tejido afectado, así como las diferentes vías de respuesta ante la alteración de la función lisosomal. Por este motivo, consideramos el análisis de transcriptoma un elemento de interés en este trabajo de tesis, ya que

nos puede permitir identificar también posibles cambios en los perfiles de expresión de los genes con potencial aplicación como biomarcadores.

## OBJETIVOS

### 2. OBJETIVOS

La enfermedad de Pompe reúne una serie de características por las que con frecuencia los pacientes sufren retrasos prolongados hasta disponer de un diagnóstico confirmado de la enfermedad. El papel del estudio genético es trascendental para poder establecer la condición clínica de cada paciente, con todo lo que ello conlleva en aspectos relacionados con el tratamiento, el pronóstico y el adecuado consejo genético. Además, esta enfermedad presenta una gran variabilidad fenotípica que con frecuencia resulta difícil de correlacionar con el perfil mutacional. Por todo ello, se planteó en nuestro laboratorio la realización de este trabajo de tesis, que surge a partir de una actividad diagnóstica.

Asimismo, se plantea el estudio del transcriptoma en sangre periférica, con el que se puede determinar la presencia de perfiles de expresión génica propios de los pacientes y, con ello, ahondar en aspectos de la fisiopatología e intentar identificar biomarcadores de utilidad clínica para el futuro.

El **objetivo principal** de este trabajo es la mejora del procedimiento diagnóstico genético en esta enfermedad, así como el análisis de transcriptoma para conocer el perfil de expresión génica en sangre periférica asociado a la misma.

Para ello nos planteamos los siguientes **objetivos secundarios**:

1. Puesta a punto de una estrategia de análisis genético que permita la caracterización molecular de pacientes con sospecha de enfermedad de Pompe, incluyendo un *MLPA* de diseño propio.
2. Caracterización funcional de mutaciones para explorar su potencial implicación en el procesamiento del mRNA.
3. Analizar el transcriptoma en sangre periférica, con el fin de detectar perfiles de expresión asociados a la enfermedad, así como rutas diferencialmente expresadas entre pacientes y controles.



## MATERIAL Y MÉTODOS

### 3. MATERIAL Y MÉTODOS

#### 3.1. Población de estudio

En este trabajo de tesis se estudiaron 259 muestras de pacientes procedentes de numerosos centros sanitarios de todo el territorio nacional. Los criterios de selección para la solicitud del estudio genético de cada paciente fueron aplicados por los médicos responsables de su atención sanitaria. Desafortunadamente, en la inmensa mayoría de los casos esos criterios no quedaron reflejados en el formulario de solicitud de análisis genético. Además, se incluyeron muestras de 31 individuos sanos que formaron parte del grupo control en algunos experimentos concretos.

Todos los individuos incluidos en el estudio han dado su visto bueno, rubricando un consentimiento informado aprobado por el comité de ética del Hospital Universitario La Paz, que cumple con los requisitos de la Ley de Investigación Biomédica 14/2007 y con la Declaración de Helsinki de la Asociación Médica Mundial de 1964. En los casos de los pacientes menores de edad fueron sus padres o tutores legales los que dieron su consentimiento para ser incluidos en el estudio.

#### 3.2. Aislamiento y cuantificación de ácidos nucleicos

La extracción de DNA genómico (gDNA) se realizó a partir de muestras de sangre periférica (SP), obtenidas por venopunción y recogidas en tubos con EDTA dipotásico (K<sub>2</sub>EDTA) como anticoagulante. En el caso de las muestras procedentes del hospital La Paz, se utilizaron tubos *BD Vacutainer*<sup>®</sup> (Becton Dickinson).

El procedimiento de extracción se llevó a cabo manualmente con el kit *Gentra*<sup>®</sup> *Puregene*<sup>®</sup> *Blood Kit* (Qiagen), y de forma automática utilizando la tecnología Chemagen, con el equipo *Chemagic MSM I* (Perkin Elmer). En ambos casos se partió de un volumen de 3 ml de sangre total, que fue procesada siguiendo las instrucciones de los proveedores de los kits. El rendimiento final de gDNA obtenido por muestra fue de alrededor de los 75 µg, dependiendo del rendimiento del experimento y de la celularidad de cada muestra.

El ARN total fue extraído a partir de una alícuota de 2 ml de sangre periférica. Se utilizó para ello el *miRNeasy mini kit* (Qiagen). En todo momento se siguieron los procedimientos propuestos por el fabricante del kit. El rendimiento total del RNA extraído fue de alrededor de 10 µg por muestra.

La cuantificación y determinación de la pureza de los ácidos nucleicos aislados se realizó utilizando los espectrofotómetros *Nanodrop*<sup>™</sup> *2000* (Thermo Scientific) y *Infinite*<sup>®</sup> *M200* (Tecan). La relación A260/280 fue muy estable en todas las muestras, y se consideró que un ADN de pureza óptima tenía un

valor entre 1.8-2.0. En aquellos casos en los no se alcanzaron esos niveles de pureza, las muestras fueron repurificadas.

Para valorar la integridad y la calidad de los RNAs se cargaron 300ng de RNA total en un gel de agarosa al 1% (agarosa D2 Pronadisa), con el fin de visualizar la integridad de las bandas correspondientes a las subunidades ribosomales 28S y 18S. Para los ensayos de RNA-seq se utilizó además el *Bioanalyzer RNA 6000 Nano Kit* (Agilent Technologies).

### 3.3. Amplificación mediante PCR del gen GAA.

Las reacciones se llevaron a cabo mediante la técnica de reacción en cadena de la polimerasa (PCR), siguiendo el esquema de análisis molecular de diseño propio para el gen *GAA*, de manera que todos los exones y regiones intrónicas flanqueantes estuvieran incluidas en el conjunto de fragmentos amplificados del gen. Las amplificaciones se llevaron a cabo utilizando oligonucleótidos de diseño propio, cuya secuencia se muestra en la **Tabla 1**. Se utilizó la herramienta bioinformática *BLAT* (<https://genome.ucsc.edu/cgi-bin/hgBlat?command=start>) para confirmar que las secuencias propuestas hibridaban solo en la secuencia de interés. Asimismo, para determinar la temperatura de fusión se empleó la herramienta *In-silico PCR* (<https://genome.ucsc.edu/cgi-bin/hgPcr>). Una vez diseñados, los oligonucleótidos fueron sintetizados por los laboratorios Roche o Sigma.

Nombre	Secuencia (5' - 3')	Localización	Tamaño amplificado (pb)
GAA_EX2F1	AGTCTAGACAGCAGGGCAA	Exón 2	466
GAA_EX2R1	GCGCAATCGAAGCGGCTGTT		
GAA_EX2F2	AGAGCAGTGCCACACAGTG	Exón 2	405
GAA_EX2R2	ACCCACCCCTTGTGAGGTGC		
GAA_EX3F	TGTCCTTGGCGTGC GGGTTGTT	Exón 3	263
GAA_EX3R	GTCGCCCTCCCATCATGCTG		
GAA_EX4-5F	GGTGCTCTCTGGGTGCTCTCAGG	Exón 4	490
GAA_EX4-5R	CATGCGGACCTCCAGTCTCCAGG	Exón 5	
GAA_EX6F	CGCCTGTGATTGGCCATCTGTGG	Exón 6	280
GAA_EX6R	CAGGCCAACGCCACTTCATGAG		
GAA_EX7-8F	TTGTGGGTAGGGCCTGCTCCCTG	Exón 7	518
GAA_EX7-8R	CCACACAGGCACGAGGATGACGC	Exón 8	
GAA_EX9F	CTCACTGTCTCAGTTTTCCC	Exón 9	278
GAA_EX9R	AAATCCCACCGTCTTCTGAGC		
GAA_EX10-11F	CTCAGTGGGGCTTCCATGCAG	Exón 10	443
GAA_EX10-11R	GTGCTAAGTCTCCAGGCCAGA	Exón 11	
GAA_EX12F	AGGGAGGGCACCTTGGAGCCTG	Exón 12	250
GAA_EX12R	GCAGAGGCCCAACCTTGTAGG		
GAA_EX13F	TGCCCTGCTGGTGACAGGGTTCC	Exón 13	273
GAA_EX13R	CCGGCAAGCCTCCCATAGAGGCC		

**Tabla 1. Oligonucleótidos utilizados para la caracterización molecular de los pacientes.**

En esta tabla se detallan las identificaciones y secuencias de todos los oligonucleótidos utilizados en la amplificación completa del gen *GAA*, así como el tamaño de cada fragmento amplificado.

## MATERIAL Y MÉTODOS

GAA_EX14F	ACTTGGCCTGAGCTGGCTCTGC	Exón 14	277
GAA_EX14R	TTCCAGGGGAGAGTCTTGGGT		
GAA_EX15F	TGAGAAGTGCAGCTCTCCCG	Exón 15	312
GAA_EX15R	GACAGGGCTGCCTGGGAGTTACG		
GAA_EX16F	ATTCAGCCTCTTCTGTGCCTCC	Exón 16	332
GAA_EX16R	TTCCGCCACCTGGTCACACC		
GAA_EX17F	GGGAGATGGAGAGCGTGGTTCC	Exón 17	313
GAA_EX17R	CTCCCCACCATCTCCCTGTGC		
GAA_EX18F	TGCTGTACCAGCCTAGCATTCCC	Exón 18	357
GAA_EX18R	AGTGGCAGGTAGCCATCGTG		
GAA_EX19F	CCAGCTGTCTGCTGACACCT	Exón 19	289
GAA_EX19R	CAGGACGCCCAGCACCTTCTGC		
GAA_EX20F1	TCTGAGCGCTGGGGTCTCACTGC	Exón 20	432
GAA_EX20R1	CTAATACTACCCGCATGCCG		
GAA_EX20F2	CGCATCGCTTGTTCACCTC	Exón 20	458
GAA_EX20R2	TGGATACATCCTCCTGCCCTG		

**Tabla 2 (continuación). Oligonucleótidos utilizados para la caracterización molecular de los pacientes.**

En esta tabla se detallan las identificaciones y secuencias de todos los oligonucleótidos utilizados en la amplificación completa del gen *GAA*, así como el tamaño de cada fragmento amplificado.

Para la amplificación de cada fragmento se utilizaron 20 ng de gDNA junto con el reactivo *PCR Master Mix* (Promega™), que contiene todos los componentes necesarios para la reacción de PCR. En cada experimento de amplificación se incluyó un tubo control sin gDNA, usado para detectar una posible contaminación de los reactivos.

Para poner a punto el perfil óptimo de temperaturas de cada amplificación se hicieron una serie de experimentos con gradientes entre 54 y 62°C, con una amplificación cada 2°C. Todas las reacciones de amplificación por PCR se llevaron a cabo en termocicladores *Mastercycler*® (Eppendorf) y *2720 Thermal Cycler* (Applied Biosystems™). Finalmente, se establecieron como condiciones óptimas de amplificación las descritas en la **Tabla 2**.

	Temperatura	Tiempo	Nº de ciclos
<b>Desnaturalización inicial</b>	98 °C	5 min	1
<b>Amplificación</b>	96 °C	30 s	32
	60 °C	30 s	
	72 °C	30 s	
<b>Extensión final</b>	72 °C	5 min	1

**Tabla 3. Condiciones de amplificación por PCR de todos los fragmentos del gen *GAA* amplificados para la caracterización molecular de los pacientes.**

Los amplificadores fueron verificados en geles de agarosa al 2%, utilizándose para ello TBE 1X (Norgen). Cada gel contenía bromuro de etidio al 5% (Sigma). Se utilizaron los marcadores de peso *DNA Molecular Weight Marker VI* (Roche) y *1Kb Plus DNA Ladder* (Invitrogen™). Los geles se fotografiaron en *MiniBIS Pro*® (Bio-Imaging Systems).

### 3.4. Purificación de los fragmentos amplificados y secuenciación mediante el método Sanger.

Una vez realizada la PCR inicial, 40-50 ng de cada uno de los fragmentos amplificados fueron sometidos a un proceso de limpieza enzimática, para eliminar restos de oligonucleótidos y desoxinucleótidos trifosfato (dNTPs) no incorporados. Para ello, se utilizaron los reactivos *ExoProStar* (Illustra) o *Ex'S-Pure* (Nimagen), siguiendo las instrucciones del fabricante. Posteriormente, se realizó una secuenciación bidireccional, utilizando los oligonucleótidos detallados en la **Tabla 1** y los reactivos de secuenciación *BigDye™ Terminator v3.1* (Applied Biosystems) y *BrightDye™ Terminator v3.1* (Nimagen). La PCR de secuenciación se llevó a cabo con las condiciones que se detallan en la **Tabla 3**.

	Temperatura	Tiempo	Nº de ciclos
Desnaturalización inicial	98 °C	5 min	1
Amplificación	96 °C	20 s	25
	55 °C	20 s	
	60 °C	4 min	
Extensión final	72 °C	5 min	1

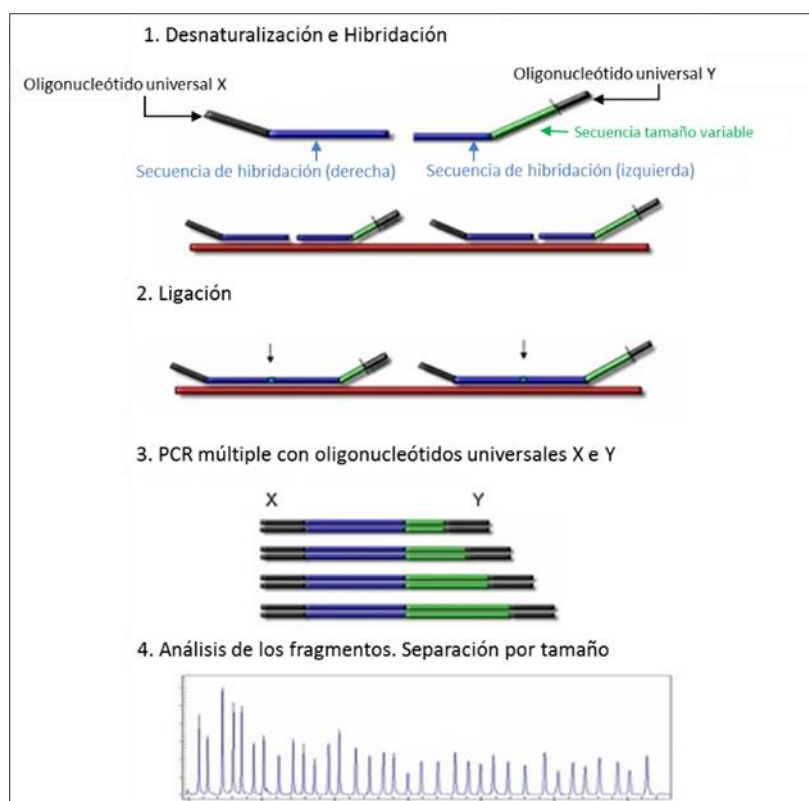
**Tabla 4. Condiciones de la PCR de secuenciación de todos los fragmentos del gen GAA para la caracterización molecular de los pacientes.**

Los productos de secuenciación fueron purificados empleando los reactivos *Agencourt CleanSEQ* (Beckman Coulter), basados en la tecnología de inmovilización reversible en fase sólida y que sirven para la eliminación de nucleótidos fluorescentes no incorporados. Posteriormente, fueron sometidos a electroforesis capilar en el equipo *ABI Prism® 3730* (Applied Biosystems). Los electroferogramas se analizaron utilizando el programa *Sequencher™ v4.9* (Gene Codes) con la secuencia de referencia NM\_000152, obtenida de la base de datos de Secuencias de Referencia (<https://www.ncbi.nlm.nih.gov/refseq/>).

Todas las variantes identificadas en este trabajo de tesis han sido nombradas siguiendo las recomendaciones de nomenclatura de la Human Genome Variation Society (<http://www.hgvs.org/mutnomen/>).

### 3.5. Amplificación de sondas dependiente de ligación múltiple (MLPA) de diseño propio.

La técnica de *MLPA (Multiplex Ligation-dependent Probe Amplification)* (Schouten et al., 2002) es un método que permite una cuantificación relativa de múltiples regiones de interés en una única PCR. Para llevar a cabo la reacción es necesario el diseño de dos sondas por cada fragmento a amplificar, con secuencias complementarias a la región de interés y que hibriden en nucleótidos contiguos. Cada sonda tiene una región de secuencia variable que permite controlar el tamaño de los amplificados, para que puedan ser separados después mediante electroforesis capilar. Además, presentan en cada extremo la secuencia de un oligonucleótido universal, lo que resulta imprescindible para realizar la PCR múltiple. Brevemente, la técnica consiste en cuatro pasos (**Figura 12**):



**Figura 12. Esquema explicativo de la técnica de MLPA.**

La técnica consiste en un primer paso de desnaturalización seguido de una hibridación. El paso de ligación permite unir aquellas sondas que han hibridado en nucleótidos contiguos y que serán amplificadas mediante una PCR múltiple. Los fragmentos obtenidos son separados mediante electroforesis capilar (Imagen adaptada de MRC-Holland).

Se utilizan 200ng de gDNA que se somete a un primer paso de desnaturalización con el fin de separar la doble cadena de DNA. Posteriormente, se añaden las sondas y se permite la hibridación de las mismas mediante una incubación a 60°C durante 16-18h. Pasado ese tiempo, se realiza una reacción de ligación mediante una ligasa termoestable, después se lleva a cabo una reacción de PCR múltiple mediante los oligonucleótidos universales. Cada uno de los fragmentos amplificados será separado por electroforesis capilar. La altura de los picos obtenidos en el paciente son comparados con los obtenidos en los controles, permitiendo así detectar ganancias o pérdidas de dosis génicas en las regiones estudiadas.

Con el fin de explorar la presencia de grandes reordenamientos germinales que pudieran afectar la estructura del gen *GAA*, se diseñó un *MLPA* con sondas específicas para los exones 2, 4, 6, 10, 11, 13, 14, 16, 18, 20 y *UTRs* 5' y 3' del gen, ya que no existía un diseño previo de *MLPA* para este gen (Amiñoso et al., 2013). Todas las sondas tenían una estructura común, que seguía las instrucciones definidas al ser publicada la técnica (Schouten et al., 2002). Además, se introdujeron en el diseño varias sondas control que hibridaban en otras regiones del genoma, para así poder normalizar los resultados. En este caso, se utilizaron cinco sondas control que hibridaban con regiones concretas de los cromosomas 17 y 18 (sondas *RAI1*, *NOL4*, *MAPT*, *CDH2* y *CHST9*). El diseño de las sondas se realizó manualmente, pero se confirmó su hibridación específica en el genoma mediante la herramienta bioinformática *BLAT* (<https://genome.ucsc.edu/cgi-bin/hgBlat?command=start>).

Los oligonucleótidos usados se detallan en la **Tabla 4** y en la **Tabla 5**.

Nombre	Secuencia (5' - 3')	Exón	Localización (Hg38)	Tamaño (pb)
GAA EX6L	<u>GGGTTCCCTAAGGGTTGGAGACAGGTGGGATCCTGGATGTC</u>	Ex6	17: 80108330-80108351	93
GAA EX6R	PH-TACATCTTCTGGGCCAGAGCCCAAGAGTCTAGATTGGATCTTGCTGGCAC		17: 80108352-80108380	
GAA EX10L	<u>GGGTTCCCTAAGGGTTGGAGGAGGACATGGTGGCTGAGTT</u>	Ex10	17: 80110786-80110806	98
GAA EX10R	PH-CCATGACCAGGTGCCCTTCGACGGCATGTGGATTgTCTAGATTGGATCTTGCTGGCAC		17: 80110807-80110840	
GAA EX4L	<u>GGGTTCCCTAAGGGTTGGAGCTGTTCTTTCGGACCAGTTCCTCAGCTG</u>	Ex4	17: 80107579-80107608	103
GAA EX4R	PH-TCCACCTCGTCCCTCGCAGTATATCACAGTCTAGATTGGATCTTGCTGGCAC		17: 80107609-80107639	
GAA EX18L	<u>GGGTTCCCTAAGGGTTGGAcattgagATGGAGAGCCTGGAAGTCTGG</u>	Ex18	17: 80118292-80118316	108
GAA EX18R	PH-AGCGAGGGGCTACACACAGGTATCTTcagacagTCTAGATTGGATCTTGCTGGCAC		17: 80118317-80118345	
RAI1-L	<u>GGGTTCCCTAAGGGTTGGAttgaaCGCAAGGAAAAGTGGCCTTTGCGTGACACAG</u>	In1	17: 17683573-17683603	112
RAI1-R	PH-AAGTGACACATCATTCTGATCTCAGGGCcttTCTAGATTGGATCTTGCTGGCAC	gen <i>RAI1</i>	17: 17683604-17683632	
GAA EX2L	<u>GGGTTCCCTAAGGGTTGGAgcatcatttagtCCCCAGCTACAAGCTGGAGAAC</u>	Ex2	17: 80104985-80105506	114
GAA EX2R	PH-CTGAGCTCCTGAAATGGGCTACgtctgtgataTCTAGATTGGATCTTGCTGGCAC		17: 80105507-80105030	
GAA 3' UTR-L	<u>GGGTTCCCTAAGGGTTGGAgacgatcaacctGCTTCCAACAGTGGGCCTAAGACT</u>	Región 3' UTR	17: 80120842-80120867	121
GAA 3' UTR-R	PH-TTAGGCAGAGGCCAGAAATCTGCCagttacagttatagTCTAGATTGGATCTTGCTGGCAC		17: 80120868-80120893	
NOL4-L	<u>GGGTTCCCTAAGGGTTGGAttcactggtATCCACGGACAACCCAAATTTAAATTCT</u>	Ex1	18: 34223056-34223083	124
NOL4-R	PH-GGGTCAAATCGAAGGGCTCCAGCTGGCCAGaaattgctggtTCTAGATTGGATCTTGCTGGCAC	gen <i>NOL4</i>	18: 34223084-34223115	
MAPT-L	<u>GGGTTCCCTAAGGGTTGGAgaaactCCTGGGACTTTAGGGCTAACAGTTCTCTTGTAAAGG</u>	3' UTR	17: 46025154-46025190	128
MAPT-R	PH-ACTTGTGCCTCTTGGGAGACGTCCACCCGTTTctggattgaatTCTAGATTGGATCTTGCTGGCAC	gen <i>MAPT</i>	17: 46025191-46025223	
CDH2-L	<u>GGGTTCCCTAAGGGTTGGAttgtacattctgggtccgAGGATGTTTGCCTTACTGTTGCTGC</u>	Ex10	18: 27990277-27990302	131
CDH2-R	PH-AGAAAATCAAGTGCATTAGCCAAGGGggccttaagccgggaTCTAGATTGGATCTTGCTGGCAC	gen <i>CDH2</i>	18: 27990303-27990329	
CHST9-L	<u>GGGTTCCCTAAGGGTTGGAggaaatgtccAGAGAACAAGAAAGACTGAAGTTTTACGGGAAAACAAAT</u>	Ex2	18: 27142805-27142844	137
CHST9-R	PH-CATGTGGTCTCAGATTCTGAAATAAGGAGAAATGCataagctgaTCTAGATTGGATCTTGCTGGCAC	gen <i>CHST9</i>	18: 27142845-27142881	

**Tabla 5. Oligonucleótidos utilizados para la realización del MLPA, mezcla 1.**

Identificación de cada uno de los cebadores usados en la mezcla 1, especificándose su secuencia, las coordenadas de la región genómica donde hibridan (chr: coordenada) y el tamaño del amplificado obtenido. Leyenda: ACGT: secuencia de los oligonucleótidos universales; **acgt**: Bases introducidas para ajustar tamaño; ACGT: secuencia del oligonucleótidos específica del exón a analizar; PH-: Primer fosforilado; Ex: Exón; In: Intrón.

Nombre	Secuencia (5' - 3')	Exón	Localización (Hg38)	Tamaño (pb)
GAA EX13L	<u>GGGTTCCCTAAGGGTTGGAGGTGAAGGCTCGGGGGACACGC</u>	Ex13	17: 80112584-80112605	92
GAA EX13R	PH-CCATTGTGATCTCCCGCTCGACCTTTGCTAGATTGGATCTTGCTGGCAC		17: 80112606-80112633	
GAA EX16L	<u>GGGTTCCCTAAGGGTTGGAGAAGGCCGAAGTGACTGGCTACTTCCC</u>	Ex16	17: 80117055-80117081	94
GAA EX16R	PH-CTTGGGCACATGGTACGACCTGCAGTCTAGATTGGATCTTGCTGGCAC		17: 80117082-80117106	
GAA EX11L	<u>GGGTTCCCTAAGGGTTGGAGTACCTACCAGCAGCGCTCTCTTGCCAG</u>	Intrón 10	17: 80110912-80110940	100
GAA EX11R	PH-GACATGAACGAGCCTTCCAACCTCATCAGTCTAGATTGGATCTTGCTGGCAC	Ex11	17: 80110941-80110969	
RAI1-L	<u>GGGTTCCCTAAGGGTTGGAttgaaacGCAAGGAAAAGTGGCCTTTTGCCTGACACAG</u>	Intrón 1 gen <i>RAI1</i>	17: 17683573-17683603	112
RAI1-R	PH-AAGTGGACACATCATTCTGATCTCAGGGCcttTCTAGATTGGATCTTGCTGGCAC		17: 17683604-17683632	
GAA EX14L	<u>GGGTTCCCTAAGGGTTGGAgattcagtgatCCTTCTACCCCTTCATGCGGAAC</u>	Ex14	17: 80112984-80113006	115
GAA EX14R	PH-CACAACAGCCTGCTCAGTCTGGTAGATcagtcattaggTCTAGATTGGATCTTGCTGGCAC	Ex14_In14	17: 80113007-80113031	
GAA EX20L	<u>GGGTTCCCTAAGGGTTGGAgattaccaatgataCTGTGTCTCGCTGTTGATGGGAG</u>	Ex20	17: 80119283-80119305	118
GAA EX20R	PH-AGCAGTTTCTCGTCAGCTGGTGTAGcaatagagttacTCTAGATTGGATCTTGCTGGCAC		17: 80119306-80119331	
NOL4-L	<u>GGGTTCCCTAAGGGTTGGAttcactggtATCCACGGACAACGCCAAATTTAAATCT</u>	Ex1	18: 34223056-34223083	124
NOL4-R	PH-GGTCAAATCGAAGGGCTTCCAGCTGGCCAGaattcgtggtTCTAGATTGGATCTTGCTGGCAC	gen <i>NOL4</i>	18: 34223084-34223115	
MAPT-L	<u>GGGTTCCCTAAGGGTTGGAgaaactCCTGGGACTTTAGGGCTAACCAAGTCTCTTTGTAAGG</u>	3'UTR	17: 46025154-46025190	128
MAPT-R	PH-ACTTGTGCCTCTTGGGAGACGTCCACCCGTTTctggattgaatTCTAGATTGGATCTTGCTGGCAC	gen <i>MAPT</i>	17: 46025191-46025223	
CDH2-L	<u>GGGTTCCCTAAGGGTTGGAattgtacattctgggtccgAGGATGTTTGCCTTACTGTTGCTGC</u>	Ex10	18: 27990277-27990302	131
CDH2-R	PH-AGAAAATCAAGTGCATTAGCCAAGGGggccttaagccgggaTCTAGATTGGATCTTGCTGGCAC	gen <i>CDH2</i>	18: 27990303-27990329	
GAA 5' UTR-L	<u>GGGTTCCCTAAGGGTTGGAgcactacatctatataCGCTCTTCTCCCTGGACTTCCAAGTA</u>	Región 5'UTR	17: 80100476-80100502	134
GAA 5' UTR-R	PH-CCAATGACCTTGGGGCTCTTTGCAGTTTAAAGacatcgatcagtcagTCTAGATTGGATCTTGCTGGCAC		17: 80100503-80100535	
CHST9-L	<u>GGGTTCCCTAAGGGTTGGAggaaatgtccAGAGAACAAGAAAAGACTGAAGTTTACGGGAAAACAAT</u>	Ex2	18: 27142805-27142844	137
CHST9-R	PH-CATGTGGTCTCAGATTCTGAAATAAGGAGAATGCaatgtaTCTAGATTGGATCTTGCTGGCAC	gen <i>CHST9</i>	18: 27142845-27142881	

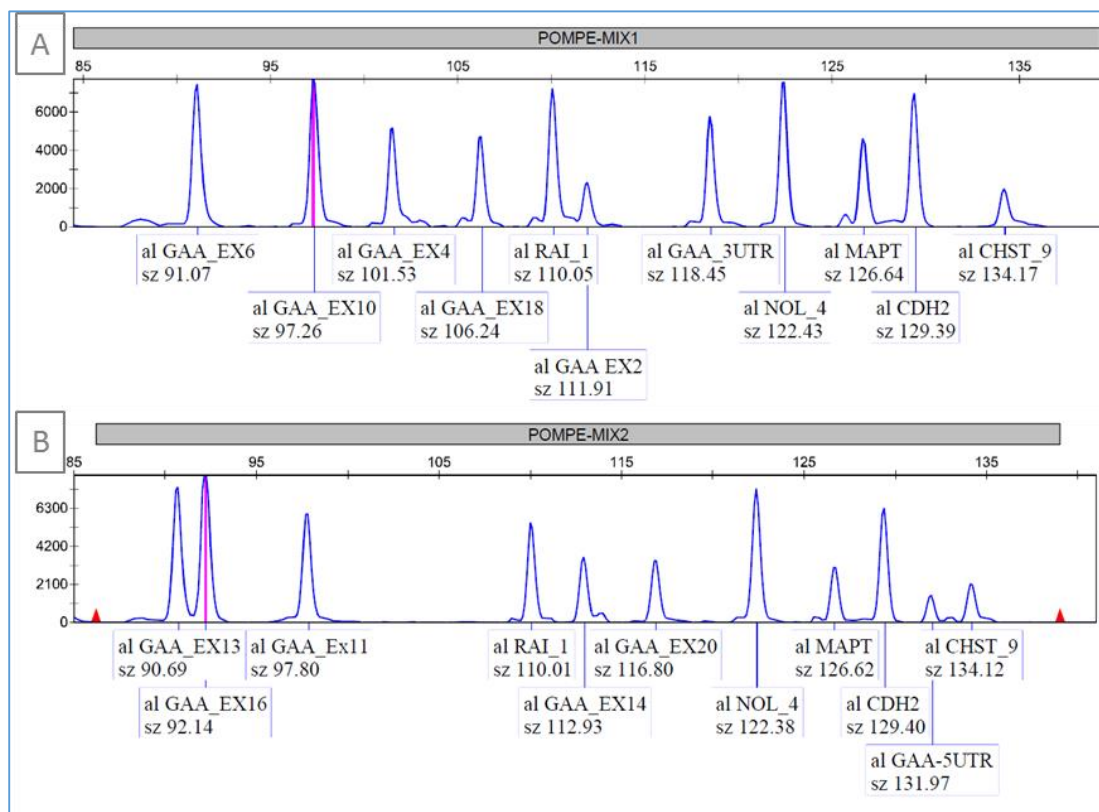
**Tabla 6. Oligonucleótidos utilizados para la realización del MLPA, mezcla 2**

Identificación de cada uno de los cebadores usados en la mezcla 2, especificándose su secuencia, las coordenadas de la región genómica donde hibridan (chr: coordenada) y el tamaño del amplificado obtenido. Leyenda: ACGT: secuencia de los oligonucleótidos universales; **acgt**: Bases introducidas para ajustar tamaño; ACGT: secuencia del oligonucleótidos específica del exón a analizar; PH-: Primer fosforilado; Ex: Exón; In: Intrón.

Las reacciones de desnaturalización e hibridación, la posterior reacción de ligación y, finalmente, las reacciones de PCR se realizaron con reactivos de *MRC-Holland*, siguiendo las instrucciones del fabricante para cada una de las etapas.

Los distintos fragmentos obtenidos fueron separados en un secuenciador *ABI Prism® 3730* (Applied Biosystems). En nuestro caso, para poder obtener un perfil de picos con tamaños suficientemente diferenciados se prepararon dos mezclas de sondas (**Figura 13**). En ambas mezclas se incluyeron seis sondas que hibridaban en el gen *GAA* y las cinco sondas control. Las sondas utilizadas en la mezcla 1 se detallan en la **Tabla 4**, y las de la mezcla 2, en la **Tabla 5**. Posteriormente, las características de cada uno de los cromatogramas fueron analizadas utilizando el programa *Peak Scanner v.1.0* (Applied Biosystems). La normalización de los datos se llevó a cabo considerando la altura de cada pico y su relación con el resto de los picos de los fragmentos control, siguiendo el procedimiento descrito (Schouten et al., 2002).





**Figura 13. Perfil de picos del MLPA de diseño propio obtenido por electroforesis capilar.**

A) Perfil de picos obtenidos con la mezcla 1. B) Perfil de picos obtenidos con la mezcla 2. Debajo de cada pico aparece el nombre de la sonda (al) y el tamaño del fragmento amplificado (sz).

### 3.6. Array CGH (Comparative Genomic Hybridization)

Para validar y caracterizar la deleción detectada mediante *MLPA* en el paciente P11, se utilizó el *MetabolArray®* (Agilent Technologies), que fue desarrollado en el Instituto de Genética Médica y Molecular (INGEMM) del Hospital Universitario La Paz, en colaboración con el Centro de Enfermedades Moleculares del Dpto. de Biología Molecular de la UAM (Desviat et al., 2011). El array se diseñó con sondas específicas para 205 genes implicados en enfermedades metabólicas con base genética. La densidad media de cobertura en las regiones donde se localizan los genes es de 40-300 pb, e incluye una cobertura del genoma entero con una separación media de aproximadamente 132kb. La región donde se encuentra el gen *GAA* se enriqueció con una sonda cada 80pb.

Brevemente, para el estudio con el *array CGH* se utilizó gDNA del paciente y el gDNA de un control. Se llevó a cabo una desnaturalización inicial y el marcaje de cada gDNA con distintos fluorocromos. Se realizó la hibridación con las sondas que se encuentran inmovilizadas en un chip y se cuantificó la intensidad de fluorescencia (**Figura 14**).



**Figura 14. Esquema explicativo de la técnica de array CGH.**

La técnica consiste en un primer paso de desnaturalización del gDNA del paciente y de una muestra control. Seguidamente, se realiza un marcaje de ambos gDNAs con distintos fluorocromos y se permite la hibridación de estos con las sondas inmobilizadas. Si la intensidad de fluorescencia es equitativa no existe ni ganancia ni pérdida. Si hay una menor intensidad de señal del fluorocromo con el que se marcó el paciente es que existe una deleción y si esa intensidad es mayor, es que existe una duplicación.

El análisis del *array CGH* se llevó a cabo con el *software Agilent Genomic WorkBench 6.5*, y se utilizó como criterio de validación que la alteración de la señal afectase, al menos, a tres sondas consecutivas.

### 3.7. Análisis de paternidad

Cada análisis de paternidad se realizó con el kit *AmpFLSTR™ Identifiler™ PCR Amplification* (ThermoFisher Scientific), siguiendo las directrices marcadas por el fabricante en cada una de las etapas de su aplicación. Este kit permite la amplificación de 15 microsatélites (*STR-Short Tandem Repeat*) así como la amelogenina, previo a su separación en el secuenciador *ABI Prism® 3730* (Applied Biosystems). Posteriormente, los cromatogramas fueron analizados utilizando el programa *Peak Scanner v.1.0* (Applied Biosystems). El grado de confirmación de paternidad que otorga este procedimiento es del 99,9999992%.

### 3.8. Interpretación biológica de las variantes identificadas en el gen *GAA*

#### 3.8.1. Análisis *in silico* de variantes missense

Para la clasificación de las variantes se utilizaron las recomendaciones propuestas por el Colegio Americano de Medicina Genética y Genómica (*American College of Medical Genetics and Genomics-ACMG*) y la Asociación de Patología Molecular (*Association for Molecular Pathology*) (Richards et al.,

2015). Éstas fueron clasificadas como patogénica (P), probablemente patogénica (LP), variante de significado desconocido (VSD), probablemente benigna (LB) y benigna (B).

Para determinar la posible patogenicidad de las variantes encontradas se consultaron las siguientes bases de datos:

- *Pompe Center* (<http://www.pompevariantdatabase.nl>).
- *ClinVar* (<http://www.ncbi.nlm.nih.gov/clinvar/>).
- *ExAC* (<http://exac.broadinstitute.org/>).

Se consideraron por su valor clínico aquellas variantes que se clasificaban como patogénicas en la literatura o en las bases de datos consultadas. Asimismo, se clasificaron como patogénicas aquellas que no habiendo sido descritas con anterioridad, variantes “*de novo*”, que producían un truncamiento prematuro de la síntesis del mRNA, una alteración de la fase de lectura o una alteración del procesamiento del mRNA.

Aquellos cambios que no estaban descritos en las bases de datos, o lo estaban como de significado desconocido, fueron analizados bioinformáticamente, con el fin de poder establecer su posible patogenicidad. Se utilizaron las siguientes herramientas:

- *Polyphen2* (<http://genetics.bwh.harvard.edu/pph2/>).
- *Mutation Taster* (<http://www.mutationtaster.org/>).
- *SIFT* (<http://sift.bii.a-star.edu.sg/>).
- *Revel* (<https://sites.google.com/site/revelgenomics/>).
- *InterVar* (<http://wintervar.wglab.org/>).
- *SNAP2* (<https://rostlab.org/services/snap2web/>).
- *DynaMut* (<http://biosig.unimelb.edu.au/dynamut/>).
- Plataforma *Alamut® Visual v.2.7.2* (<https://www.interactive-biosoftware.com/alamut-visual/>).

Las herramientas *Polyphen2* (Adzhubei et al., 2010) y *SIFT* (Kumar et al., 2009) realizan predicciones mediante un análisis de la estructura fisicoquímica del cambio, teniendo además en cuenta el grado de conservación del aminoácido en distintas especies, otorgándoles con ello una puntuación. En el caso de *Polyphen2* esta puntuación va de 0 a 1, siendo 0 cuando predice que la variante es posiblemente benigna, y 1 cuando es posiblemente patogénica. En el caso de *SIFT*, la puntuación también oscila en el rango de 0-1, pero en este caso las variantes con puntuación  $\leq 0,05$  estarían consideradas como patogénicas y las que obtengan un valor  $>0,05$  como toleradas. *Mutation Taster* (Schwarz et al., 2014) analiza la estructura y la conservación de los aminoácidos implicados. Además, examina si hay descritas mutaciones en el mismo aminoácido o si está descrito el cambio en bases de datos poblacionales como *ExAC* y en bases de datos de mutaciones como *HGMD*. Con esos datos, establece para la variante un valor

de probabilidad de ser patogénica o de ser benigna, sin tener un rango de valores como en el caso de las anteriores herramientas.

La herramienta *InterVar* (Li and Wang, 2017) integra los criterios de la ACMG y los resultados de distintas herramientas, para facilitar la adecuada clasificación de las variantes genéticas.

La herramienta *REVEL* (Loannidis et al., 2016) se utilizó para intentar predecir la posible patogenicidad de las variantes *missense*. Para ello, en REVEL han integrado 13 herramientas ampliamente usadas para este fin, como son: *MutPred*, *FATHMM*, *VEST*, *PolyPhen*, *SIFT*, *PROVEAN*, *MutationAssessor*, *MutationTaster*, *LRT*, *GERP*, *SiPhy*, *phyloP*, and *phastCons*. Con ello se obtiene una puntuación que va en un rango de 0 (probablemente neutra) a -1 (probablemente patogénica). Se considera que una variante podría ser probablemente patogénica, cuando la puntuación que se obtiene es superior a 0,5, con una sensibilidad estimada de 0,754 y una especificidad de 0,891.

La herramienta *SNAP2* (Hecht et al., 2015) realiza sus predicciones teniendo en cuenta la conservación evolutiva de los aminoácidos, en función de un alineamiento múltiple generado automáticamente. Considera características estructurales, como la predicción de estructura secundaria. Con ello se obtiene una puntuación que se mueve en un rango entre -100 y 100. *SNAP2* muestra en azul los cambios que se predicen como neutros y en rojo aquellos en los que se puede establecer un posible efecto deletéreo.

Para realizar un análisis más orientado a valorar el papel de los cambios sobre la estructura de la proteína se utilizó la herramienta *DynaMut* (Rodrigues et al., 2018). Esta herramienta integra el paquete de R *Bio3D* y aplica el modelo *ENCoM* (*Elastic Network Contact Model*), para poder efectuar un análisis de la flexibilidad estructural de las proteínas. Esto permite calcular el impacto que las mutaciones tienen sobre la dinámica y la estabilidad de una proteína concreta, como resultado de los cambios de entropía vibracional. El programa compara la energía libre de Gibbs ( $\Delta G$ ) de la proteína normal con la de la proteína mutante, para obtener la puntuación que luego se refleja en el resultado ( $\Delta\Delta G$ ). Establece como neutro un valor de cero. Un valor por encima de cero significa que la variante otorga una mayor estabilidad a la estructura de la proteína, mientras que un valor por debajo de cero supone una desestabilización de la estructura proteica.

Además, se realizó un análisis bioinformático de la repercusión de las variantes en el procesamiento del pre-mRNA, para lo que se utilizó la plataforma *Alamut® Visual v.2.7.2*. Para llevar a cabo este análisis, la plataforma integra los algoritmos: *Splice Site Finder-like (SSF)*, *MaxEntScan (MES)*,

*NNSplice (NNS)*, *GeneSplicer (GS)* y *Human Splice Finder (HSF)*, cuyas características por defecto se muestran en la **Tabla 6**.

Herramienta de predicción (Alamut® Visual)	Umbral 5' donador	Rango	Rango normalizado (100%)	Umbral 3' aceptor	Rango	Rango normalizado (100%)
Splice Site Finder-like (SSF)	70	0-100	70	70	0-100	70
Max Ent Scan (MES)	0	0-12	0	0	0-16	0
NNSPLICE (NNS)	0,4	0-1	40	0,4	0-1	40
Gene Splicer (GS)	0	0-15	0	0	0-15	0
Human Splicing Finder (HSF)	65	0-100	65	65	0-100	65

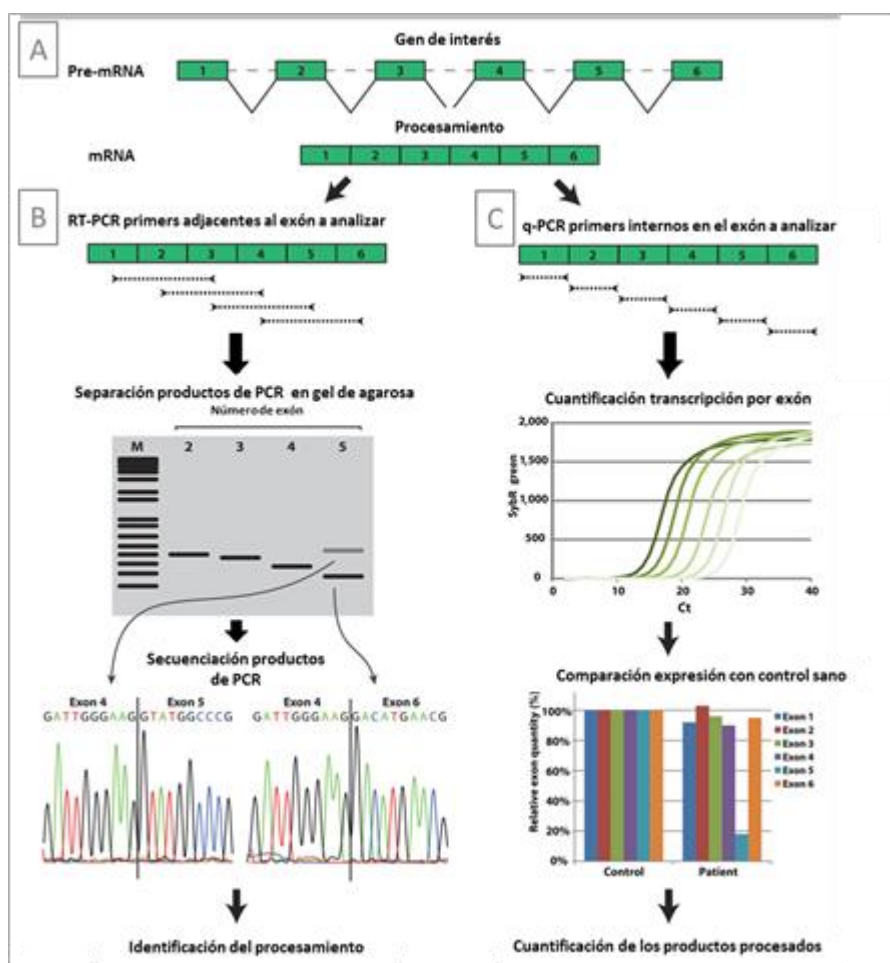
**Tabla 7. Valores por defecto de cada una de las herramientas que integran la plataforma Alamut® visual**

Valores por defecto que se establecen para cada uno de los algoritmos usados tanto para el extremo 5' como para el 3', también se muestra el rango de cada una de ellas y el rango normalizado en base al valor máximo, establecido en el 100%

El informe donde se detallan los resultados de cada variante estudiada muestra la puntuación que cada algoritmo le otorga a la secuencia normal, reflejando así su potencial papel sobre el procesamiento del mRNA. Además, refleja la puntuación que cada herramienta de predicción otorga por la presencia de la variante y, con ello, como podría estar alterando el *splicing* del mRNA. Como cada herramienta utiliza un rango de puntuación diferente, y con el objetivo de facilitar la interpretación de los resultados, el valor máximo de cada aplicación se ha normalizado al 100% y, en base a eso, se han escalado los resultados tanto del sitio normal como del mutado, con el fin de poder realizar una comparación entre ellos.

### 3.8.2. Análisis funcional de procesamiento de mRNA

El estudio funcional de procesamiento de mRNA se realizó siguiendo el procedimiento publicado por Bergsma et al., 2015 (Bergsma et al., 2015). Para llevar a cabo el análisis, 2µg de RNA total fueron tratados con 2U de *RQ1 RNase-Free DNase* (Promega™) durante 1 hora a 37°C. Posteriormente, se llevó a cabo la reacción de retrotranscripción (RT-PCR), utilizando los reactivos *ImProm-II™ Reverse Transcription System* (Promega™). Para comprobar que la RT-PCR había funcionado correctamente, se realizó una amplificación de un gen constitutivo, en este caso la *β-2 Microglobulina*, utilizando para ello las condiciones establecidas en la **Tabla 2** y los cebadores que se muestran en la **Tabla 8**. Con este abordaje se realizaron dos análisis independientes por cada exón (**Figura 15**), utilizando para ello los oligonucleótidos descritos en el citado artículo.



**Figura 15. Esquema de trabajo para el análisis de las variantes implicadas en procesamiento.**

A) Esquema del procesamiento de un mRNA. B) Análisis cualitativo mediante PCR de los posibles efectos de las variantes de procesamiento usando oligonucleótidos que hibridan en exones flanqueantes al exón de análisis. C) Análisis cuantitativos mediante qPCR de los efectos de las variantes de procesamiento usando oligonucleótidos que hibridan en el mismo exón que se quiere cuantificar. (Imagen adaptada Bergsma et al., 2015)

Se realizó un primer análisis en el que se trató de identificar posibles alteraciones en los productos de procesamiento (**Figura 15B**). Para ello se realizaron amplificaciones específicas mediante PCR, utilizando oligonucleótidos que hibridan en los exones flanqueantes al exón que se quiere analizar. Las condiciones de temperatura fueron las mismas que las utilizadas en el análisis mutacional inicial (**Tabla 2**). Finalizada la amplificación, 15 µl de cada reacción de PCR se analizaron en un gel de agarosa al 3% que contenía bromuro de etidio. Se utilizaron los marcadores de peso molecular *DNA Molecular Weight Marker VI* (Roche) y *1Kb Plus DNA Ladder* (Invitrogen). Los gels se fotografiaron en *MiniBIS Pro* (Bio-imaging Systems). En aquellos casos en los que aparecieron bandas de distinto tamaño al esperado, éstas fueron recortadas del gel y purificadas utilizando los reactivos *QIAquick Gel Extraction Purification kit* (Qiagen). A *v3.1* (Applied Biosystems) y *BrightDye™ Terminator v3.1* (Nimagen). Los electroferogramas fueron analizados mediante el software *Sequencher™ v4.9* (Gene Codes).

Posteriormente, se realizó una qPCR a tiempo real con la que se ponía de manifiesto el nivel de expresión de cada uno de los exones (**Figura 15C**). En este caso, los oligonucleótidos hibridan en el interior del exón que se quiere explorar. Se evalúa con ello la expresión relativa del gen *GAA*. Para la puesta a punto de la técnica se realizó un ensayo de eficiencia de los oligonucleótidos, con el fin de establecer la concentración óptima de los mismos para alcanzar una eficiencia de entre 90-110%. El cDNA utilizado como material de partida para la qPCR, se amplificó utilizando el kit *PerfeCTa SYBR® Green FastMix ROX* (QuantaBio). La reacción de cuantificación se llevó a cabo en el equipo *7300 Real-Time PCR System* (Applied Biosystems), siguiendo las condiciones marcadas en la **Tabla 7**.

	Temperatura	Tiempo	Nº de ciclos
Desnaturalización inicial	95 °C	3 min	1
Amplificación	95 °C	15 s	40
	60 °C	45 s	
Curva de disociación	95 °C	15 s	1
	60 °C	20 s	
	95 °C	15 s	
	60 °C	15 s	

**Tabla 8. Condiciones de la qPCR para realizar el ensayo cuantitativo**  
Ensayo cuantitativo con mRNA de todos los exones de *GAA*, así como para el gen constitutivo *B2M* usado para la normalización.

Todas las muestras fueron medidas por triplicado y normalizadas con la  $\beta$ 2-Microglobulina (*B2M*), utilizando para ello los cebadores de la **Tabla 8** con las condiciones establecidas en la **Tabla 7**.

Nombre	Secuencia (5' - 3')	Localización	Tamaño cDNA (pb)
BETA-2MQF	TGCTGTCTCCATGTTTGATGTAT	3' UTR gen <i>B2M</i>	86
BETA-2MQR	TCTCTGCTCCCCACCTCTAAGT		

**Tabla 9. Cebadores del gen constitutivo  $\beta$ 2-Microglobulina usado para la normalización de la qPCR**

El primer paso para llevar a cabo el análisis es calcular el  $\Delta Ct$ , que es la diferencia entre el ciclo umbral ( $Ct$ ) obtenido en el análisis del gen de interés, en este caso *GAA*, y el gen de referencia, el de la  $\beta$ -2M, para cada uno de los grupos de estudio, es decir, pacientes y controles. El  $\Delta\Delta Ct$  corresponde a la diferencia del  $\Delta Ct$  obtenido en el paciente y el obtenido en la muestra control.

$$\Delta Ct (\text{paciente o control}) = Ct (\text{Gen de interés } GAA) - Ct (\text{gen de referencia } B2M).$$

$$\Delta\Delta Ct = \Delta Ct (\text{paciente}) - \Delta Ct (\text{control})$$

La cuantificación final, tras la normalización fue realizada mediante el método  $2^{-\Delta\Delta Ct}$  (Livak and Schmittgen, 2001).

### 3.9. Análisis transcriptoma mediante RNA-seq

#### 3.9.1. Pacientes analizados

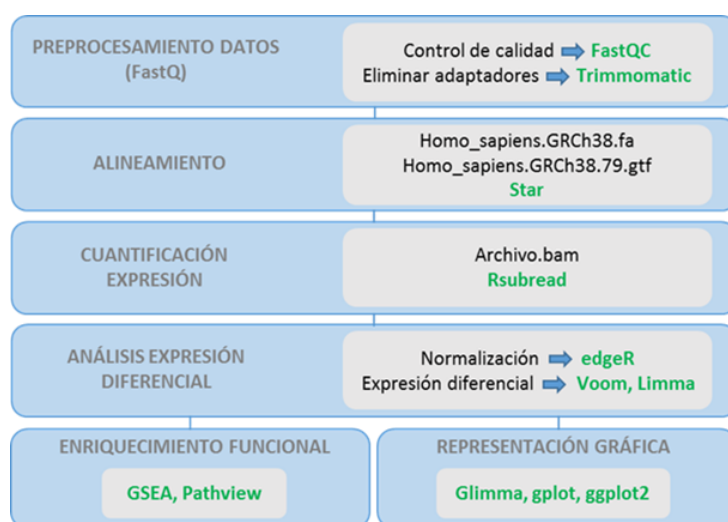
En el estudio de transcriptómica se analizaron un total de 13 muestras de pacientes que presentaban la forma adulta de la enfermedad y de los que se nos facilitó muestra de sangre para disponer de RNA. Asimismo, se analizaron 17 muestras de controles sanos con estudio genético normal para el gen *GAA*. El RNA se extrajo con el *miRNeasy mini kit* (Qiagen) a partir de 2ml de SP y siguiendo las recomendaciones del fabricante. Una vez extraído, el RNA se cuantificó utilizando el espectrofotómetro *Nanodrop™ 2000* (Thermo Scientific). Posteriormente, se prepararon diluciones de 400ng de RNA en un volumen final de 100µl de agua estéril. Estas diluciones fueron analizadas mediante el equipo *Agilent RNA 6000 Nano Kit* (Agilent Technologies), con el fin de chequear la integridad y la calidad del RNA.

#### 3.9.2. RNA-seq. Procedimiento de análisis

Los experimentos de *RNA-seq* fueron realizados en la Unidad de Genómica del Centro Nacional de Investigación Cardiovascular (CNIC). Para preparar las librerías se utilizaron los reactivos *NEBNext® Ultra™ DNA Library Prep Kit for Illumina®* (New England BioLabs), haciéndose una selección de poliA+ para posteriormente secuenciarlo en modo extremo único (*single end*) en el equipo *HiSeq 2500 Sequencing System* (Illumina®).

##### 3.9.2.1. Flujo de análisis.

El flujo de análisis utilizado se describe en la **Figura 16**:



**Figura 16. Algoritmo utilizado para el análisis del RNA-seq.**

Los archivos *FastQ* obtenidos mediante *RNA-seq* se sometieron a un control de calidad y se eliminaron los adaptadores. El alineamiento con el genoma de referencia permitió obtener los archivos *.bam*. Posteriormente, estos se utilizaron para la cuantificación de lecturas necesario para realizar el análisis de expresión diferencial, con ello se obtuvo la lista de genes diferencialmente expresados (DEGs). Se usaron distintas librerías de R para hacer las representaciones gráficas. Para el enriquecimiento funcional se utilizó la herramienta web *GSEA*, así como la librería de R *Pathview*. Los nombres de los programas usados son los que están escritos en letras verdes.



### 3.9.2.2. Procesamiento de los datos crudos

Los *FastQ* de todas las muestras fueron analizados mediante el programa *FastQC* (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), con el fin de poder detectar muestras que hubiera que desechar por baja calidad. Posteriormente, estos archivos se filtraron utilizando el programa *Trimmomatic* (<http://www.usadellab.org/>), para eliminar los adaptadores y lecturas de mala calidad o con un tamaño menor de 36 nucleótidos. Una vez realizado este paso de filtrado se volvió a repetir el análisis con el *FastQC*, con el fin de monitorizar la calidad de los resultados.

### 3.9.2.3. Alineamiento con el genoma de referencia. Cuantificación de las lecturas por gen.

Para llevar a cabo el alineamiento con el genoma de referencia se utilizó el programa *Star*. Se eligió este alineador porque es rápido, eficiente y necesita menos requerimientos computacionales que otros alineadores (Dobin et al., 2013). Inicialmente, se debe generar un genoma indexado, para lo que se utilizaron los archivos descargados de *Ensembl*: *Homo\_sapiens.GRCh38.dna.primary\_assembly.fa* y *Homo\_sapiens.GRCh38.79.gtf*. Este genoma indexado es necesario para llevar a cabo los alineamientos de las secuencias. Una vez alineadas, se obtuvieron archivos *.bam* de todas las muestras, así como archivos finales de calidad de cada alineamiento. Además, se generaron archivos *.bai* a partir de los archivos *.bam*, usando para ello el programa *samtools* (<http://samtools.sourceforge.net/>), necesarios para la visualización de los alineamientos mediante el software *IGV* (<https://software.broadinstitute.org/software/igv/>).

Para realizar el análisis posterior de expresión génica era imprescindible el paso de cuantificación de las lecturas asociadas a cada gen y en cada muestra. En este trabajo se usó la librería *Rsubread* (<http://bioconductor.org/packages/release/bioc/html/Rsubread.html>) (Liao et al., 2019). Con esta librería se pudo asignar cada lectura al gen correspondiente. Aquellas lecturas que solapaban con más de un exón se asignaron al exón en el que el solapamiento de bases era mayor. Esto finalmente genera una tabla en la que se refleja la cuantificación de lecturas de cada uno de los genes por muestra, que se utilizará en el posterior análisis de expresión diferencial.

### 3.9.2.4. Análisis de expresión diferencial.

Para llevar a cabo el análisis de la expresión diferencial se utilizaron la librería *edgeR* (<http://bioconductor.org/packages/release/bioc/html/edgeR.html>) (Robinson et al., 2010; Mc Carthy et al., 2012) y la librería *Limma* (<https://bioconductor.org/packages/release/bioc/html/limma.html>) (Ritchie et al., 2015). El primer paso consistió en filtrar y eliminar todos aquellos genes que tenían pocas lecturas. Para ello, se estableció un punto de corte en el que solo se mantuvieron aquellos genes que eran

expresados en aproximadamente la mitad de las muestras de cada uno de los grupos de estudio. Así, se estableció como punto de corte que se expresaran en al menos 8 muestras del grupo control, o en 6 muestras del grupo de pacientes. Después del filtrado fue necesario normalizar los datos. Para ello, *edgeR* utiliza el método *TMM (Trimmed Mean of M-values)*, en el que se asume que en un análisis masivo de datos la mayoría de los genes estudiados no mostrarán diferencias de expresión entre muestras (Robinson and Oshlack, 2010). Una vez filtrados y normalizados los datos, se aplicó la función *voom()*, con el objetivo de minimizar la heterocedasticidad, como paso previo a la aplicación del modelo de regresión lineal. Para realizar el análisis de la expresión diferencial se utilizó la librería *Limma*, basada en el uso de los modelos lineales para el análisis de datos de RNA-seq (Ritchie et al., 2015). Se han utilizado las funciones *lmFit()* y *eBayes()* que aplica una *t-student* moderada (*t*), usando un modelo bayesiano que intenta compensar el bajo número de muestras. Esto permite obtener una selección de genes diferencialmente expresados (Dündar et al., 2015). Posteriormente, con el fin de obtener el p-valor ajustado para test múltiples, se utilizó la función *topTable()*, que permite elegir el método de ajuste. En este estudio se utilizó el método *BH (Benjamini y Hochberg)*, para obtener la Tasa de Descubrimientos Falsos (*False Discovery Rate*), o (*FDR*), y estableciendo como valor de corte un  $FDR \leq 0,05$  (Benjamini and Hochberg, 1995).

Una vez procesados los datos, se realizó un “paso de anotación”, para lo que se utilizaron las librerías *org.Hs.eg.db()*, *TxDb.Hsapiens.UCSC.hg38.knownGene()*, *GenomicRanges()* y la librería *GenomicFeatures()*. Este paso es necesario para transformar los nombres obtenidos con el análisis de RNA-seq, que correspondían al “*Ensembl gene*”, a nombre del gen oficial, referencia *NCBI*, necesario para realizar el enriquecimiento funcional con *GSEA*, necesario para el análisis con la librería *Pathview* (<https://bioconductor.org/packages/release/bioc/html/pathview.html>) (Luo and Brouwer, 2013). Además, se generó un archivo con todos los genes expresados, ordenados por el estadístico *t*, necesarios para realizar los análisis de enriquecimiento funcional con el programa *GSEA* (<http://software.broadinstitute.org/gsea/index.jsp>), utilizando el modulo “*GSEA preranked*”.

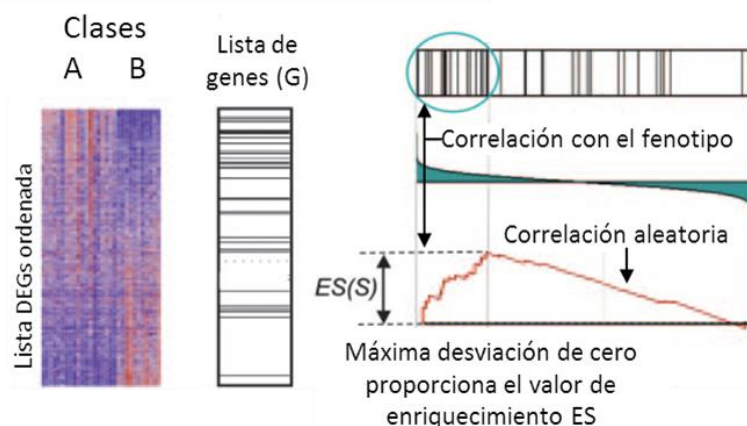
Para el análisis y visualización de los datos, se llevaron a cabo distintas representaciones:

- Análisis de escalamiento multidimensional (*MDS-Multidimensional Scaling*): Este gráfico es similar al *PCA (Principal Component Analysis)*, y sirve para representar en dos dimensiones las distancias que muestran cada par de muestras. Esa distancia se calcula como la raíz cuadrada de la media aritmética de los cuadrados de los valores de los 500 genes con mayor cambio de expresión. Para ello se utilizó la función *gMDSPlot()* de la librería *glimma* (<https://bioconductor.org/packages/release/bioc/html/Glimma.html>).
- “Gráficos de calor” o *heatmaps*: Este tipo de gráficos permite realizar agrupamientos no supervisados en función de los niveles de expresión detectados en las distintas muestras. Para ello se utilizó la función *heatmap.2()* de la librería *gplots* (<https://cran.r-project.org/web/packages/gplots/index.html>).

- Diagrama de cajas o *Boxplot*: Este tipo de gráficos permite visualizar fácilmente la distribución de los datos, permitiendo comparar los valores de expresión entre pacientes y controles de un gen en concreto, obteniéndose información de su mediana, de los cuartiles y de la presencia o no de datos fuera de rango. Para realizar estos gráficos se utilizó la función `ggplot()`, de la librería `ggplot2` (<https://cran.r-project.org/web/packages/ggplot2/index.html>).

### 3.9.2.5. Análisis de enriquecimiento funcional.

El enriquecimiento funcional de los datos obtenidos en el RNA-seq se llevó a cabo mediante el método utilizado en *GSEA* (*Gene Set Enrichment Analysis*) (Subramanian et al., 2005). Este método se utiliza para evaluar las implicaciones funcionales de agrupaciones de genes, como podría ser su participación en una misma ruta biológica. Para realizar esta valoración se analiza una lista de genes (G), obtenida previamente con *Limma*, y ordenada en base al estadístico *t*. La herramienta *GSEA* permite evaluar si el perfil de genes diferencialmente expresados puede tener vinculación con alguna de las rutas definidas en la propia aplicación. Para evaluar la significación estadística, el programa realiza un test no paramétrico de Kolmogorov-Smirnov, considerando un  $FDR \leq 0,25$  como significativo para identificar rutas biológicamente enriquecidas. El principal valor que resulta del análisis con *GSEA* es el denominado valor de enriquecimiento (ES), que se obtiene sumando o restando una cantidad por cada gen de la lista G, en función de si pertenece o no a la lista de genes agrupados por una ruta o función (**Figura 17**).



**Figura 17. Esquema general método GSEA**

Se utiliza una lista de genes diferencialmente expresados (DEGs) ordenada en función del estadístico *t*. Cada uno de los genes de esta lista, se puntúan función de si pertenece o no a una misma ruta otorgando un valor por cada gen. La máxima puntuación, y por lo tanto la máxima desviación de cero que proporciona esa suma, corresponde al valor de enriquecimiento y está correlacionado con el fenotipo. El resto de los genes de la ruta y que no están representados diferencialmente en nuestra lista están relacionados con el fenotipo de forma aleatoria, al azar.

*GSEA* dispone de 8 colecciones principales de genes y varias subcolecciones adicionales en función de diferentes criterios de clasificación, como por ejemplo: *Hallmarks*, *Reactome*, *BioCarta*, *KEGG*, etc.

Para llevar a cabo el análisis se utilizaron distintos repositorios:

- *Hallmark*, que es una agrupación de 50 conjuntos de genes que representan estados o procesos biológicos bien definidos. Cada uno de esos conjuntos ha sido generado computacionalmente basadas en la exploración de otras colecciones de genes, utilizando aquellos solapantes capaces de definir de manera fiel un proceso biológico.
- *KEGG* (*Kyoto Encyclopedia of Genes and Genomes*), formado por un conjunto de 186 colecciones de genes que forman parte de rutas implicadas en procesos biológicos, como el metabolismo, procesos celulares, enfermedades humanas o relacionados con fármacos.
- *BioCarta*, un conjunto de 289 colecciones de genes implicados en rutas relacionadas con metabolismo y señalización.
- *TFT: transcription factor targets*, constituido por 610 conjuntos de genes. Los genes de cada conjunto comparten un motivo regulador en *cis*, que podría funcionar como un sitio potencial de unión de un factor de transcripción.

Dentro de los valores por defecto, el programa va a analizar aquellos conjuntos de genes en los que haya un mínimo de 15 genes de nuestra lista y un máximo de 500.

También se utilizó la librería de R *Pathview* (<https://bioconductor.org/packages/release/bioc/html/pathview.html>), que permite visualizar los datos de los genes diferencialmente expresados obtenidos con *Limma*, localizándolos en las rutas *KEGG* en las que detecta que hay un enriquecimiento funcional con un  $FDR < 0.1$  (Luo and Brouwer, 2013).

## RESULTADOS

### 4. RESULTADOS.

#### 4.1. Análisis genético de pacientes con sospecha de enfermedad de Pompe

En el laboratorio se reciben muestras procedentes de todo el país para el diagnóstico genético de la enfermedad de Pompe. En este estudio de tesis se realizó el análisis genético de 259 muestras de pacientes con sospecha de esta enfermedad. Los criterios clínicos o bioquímicos que indicaban el estudio genético con frecuencia no se reflejaron en las solicitudes de estudio.

El análisis genético principal consistió en la amplificación por PCR, a partir de gDNA, de las regiones exónicas e intrónicas flanqueantes del gen *GAA*, y posterior secuenciación mediante el método Sanger de cada una de ellas. Además, para completar el estudio molecular se puso a punto un *MLPA* de diseño propio, con el fin de detectar grandes deleciones que no podían ser descubiertas mediante la secuenciación. La sensibilidad de este protocolo de análisis genético para identificar variantes potencialmente patogénicas en el gen *GAA* supera el 98%, considerando los eventos moleculares únicos recopilados en la base de datos del Pompe Center (<http://www.pompevariantdatabase.nl>).

##### 4.1.1. Variantes identificadas en el estudio genético de los pacientes

Las variantes identificadas fueron filtradas, de manera que aquellas que tenían una frecuencia alélica en población control superior al 1%, según las bases de datos *ExAc* o *dbSNP*, fueron descartadas, al ser consideradas polimórficas. El resto de variantes se tuvo en cuenta por su posible papel en la etiología de la enfermedad. Se identificaron cambios que tenían un claro efecto deletéreo sobre la funcionalidad del gen o la proteína, como eran los que suponían alteraciones de la fase de lectura del mRNA, la generación prematura de codones de parada, o la alteración de las secuencias canónicas de procesamiento del mRNA. Otras variantes identificadas implicaban cambios en nucleótidos de secuencia codificante y suponían cambios en la estructura primaria de la proteína. En estos últimos casos, su implicación patogénica no estaba asegurada, pero se utilizaron análisis *in silico*, consultas a las bases de datos de mutaciones, predicción de estructuras y estudios funcionales, para intentar determinar su posible efecto deletéreo.

##### 4.1.2. Pacientes con alteraciones genéticas descritas

El procedimiento de análisis permitió determinar que 47 pacientes eran portadores de genotipos con dos mutaciones consideradas patogénicas en la literatura o las bases de datos consultadas (**Tabla 9**). Los genotipos identificados permitieron confirmar el diagnóstico genético de la enfermedad en este grupo de pacientes. En 8 pacientes los genotipos estaban formados por variantes de efecto muy deletéreo sobre la funcionalidad del gen, lo que resultaba compatible con el fenotipo de forma clásica o infantil de la

## RESULTADOS

enfermedad que presentaban. Los 39 pacientes restantes presentaban una forma tardía. Es destacable que 36 de los 39 tenían la mutación c.-32-13T>G en su genotipo patológico, lo que supone el 92,3% de este grupo. Por otro lado, de los 94 alelos de esos 47 pacientes, 46 implicaban alteraciones del procesamiento del mRNA (48,9%), incluida la mutación c.-32-13T>G; 27 mutaciones suponían cambios de sentido erróneo (*missense*) (28,7%); 16 mutaciones entrañaban cambios sin sentido (*nonsense*) (17%) y 5 conllevaban cambios de fase de lectura (5,3%).

ID. pacientes	Alelo 1	Alelo 2	Nº total pacientes
P20	I1 c.-32-13T>G	Ex5 c.925G>A (p.G309R)	1
P22	I1 c.-32-13T>G	Ex16 c.2255_2257del (p.I752del)	1
P34, P125	I1 c.-32-13T>G	Ex9 c.1432G>A (p.G478R)	2
P37	Ex4 c.854C>G (p.P285R)	Ex13 c.1798C>T (p.R600C)	1
P54, P76, P191, P438	I1 c.-32-13T>G	Ex3 c.573C>A (p.Y191*)	4
P60	I1 c.-32-13T>G	Ex2 c.236-246del (p.P79Rfs*13)	1
P61	I1 c.-32-13T>G	Ex18 c.2600_2604delinsA(p.V867Efs*19)	1
P69, P92, P266	I1 c.-32-13T>G	EX5 c.875A>G (p.Y292C)	3
P80	I6 c.1076-1G>C	Ex13 c.1856G>A (p.S619N)	1
P83 (I)	Ex14 c.1927G>A (p.G643R)	Ex14 c.1927G>A (p.G643R)	1
P87, P395	I1 c.-32-13T>G	Ex16 c.2237G>A (p.W746*)	2
P114	I1 c.-32-13T>G	Ex7 c.1115A>T (p.H372L)	1
P119, P454, P478 (I)	I10 c.1551+1G>A	I10 c.1551+1G>A	3
P120	Ex15 c.2173C>T (p.R725W)	Ex15 c.2173C>T (p.R725W)	1
P121 (I)	EX5 c.875A>G (p.Y292C)	Ex14 c.1933G>T (p.D645Y)	1
P137, P391 (I)	Ex18 c.2560C>T (p.R854*)	Ex18 c.2560C>T (p.R854*)	2
P189, P420, P428	I1 c.-32-13T>G	EX2 c.343C>T (p.Q115*)	3
P206	I1 c.-32-13T>G	EX3 c.670C>T (p.R224W)	1
P231	I1 c.-32-13T>G	I10 c.1551+1G>A	1
P298	I1 c.-32-13T>G	Ex14 c.1933G>T (p.D645Y)	1
P247	I1 c.-32-13T>G	Ex5 c.877G>A (p.G293R)	1
P250	I1 c.-32-13 T>G	E16 c.2214 G>A (p.W738*)	1
P257 (I)	Ex13 c.1798C>T (p.R600C)	Ex18 c.2560C>T (p.R854*)	1
P265	I1 c.-32-13T>G	Ex12 c.1724A>G (p.Y575C)	1
P319, P403, P417	I1 c.-32-13T>G	Ex12 c.1655T>C (p.L552P)	3
P329	I1 c.-32-13T>G	Ex13 c.1802C>T (p.S601L)	1
P356, P440	I1 c.-32-13T>G	I1 c.-32-13T>G	2
P364	I1 c.-32-13T>G	Ex2 c.118C>T (p.R40*)	1
P427	I1 c.-32-13T>G	Ex15 c.2104C>T (p.R702C)	1
P405	I1 c.-32-13 T>G	Ex11 c.1564C>A (p.P522T)	1
P74	I1 c.-32-13T>G	c.2189+459_3405del (p.E730_C952del)	1
P385	I1 c.-32-13T>G	c.2481+102_2646+31del	1
<b>TOTAL</b>			<b>47</b>

**Tabla 10. Pacientes con genotipo confirmatorio de enfermedad de Pompe**

Pacientes en los que se les ha detectado dos mutaciones ya descritas. Leyenda: Ex: exón I: Intrón (I): Pacientes con fenotipo clásico

Además, se identificó un grupo de 10 pacientes con genotipos en los que se combinaban una mutación patogénica ya descrita en un alelo, con una variante de significado desconocido (VSD) en el otro, o bien VSDs en ambos alelos (**Tabla 10**). En total, se identificaron 7 VSDs diferentes en estos pacientes. Sin embargo, los resultados obtenidos en nuestro estudio permitieron que algunas de esas VSDs fueran reclasificadas, se utilizaron para ello los criterios de clasificación del *American College of Medical Genetics*

## RESULTADOS

*and Genomics (ACMG)*. Es decir, se pudieron reclasificar 4 de las 7 VSDs, ya que la segregación en *trans* con mutaciones patogénicas en nuestros pacientes hacía que esas VSDs fueran reclasificadas a probablemente patogénicas. De ese modo, se determinó en 8 de esos 10 pacientes el diagnóstico de enfermedad de Pompe. En 6 de los 8 pacientes se identificó la mutación c.-32-13T>G en el genotipo enfermo. Únicamente los dos pacientes portadores de VSDs en ambos alelos quedaban fuera de ser diagnosticados. Se aplicaron diferentes análisis *in silico*, como se expondrá en el apartado 4.1.5, con el fin de averiguar la posible patogenicidad de esas variantes.

ID. pacientes	Alelo 1	Alelo 2	Nº total pacientes
P63	I1 c.-32-13T>G	Ex4 c.844G>C (p.D282H)	1
P70	Ex12: c.1655T>C (p.L552P)	Ex12 c.1704C>G (p.H568Q)	1
P75	I1 c.-32-13T>G	Ex10 c.1445C>T (p.P482L)	1
P105, P122, P413, P433	I1 c.-32-13T>G	Ex9 c.1408_1410delAAC (p. N470del)	4
P299	Ex10 c.1504A>G (p.M502V)	Ex10 c.1504A>G (p.M502V)	1
P318	EX2 c.343C>T (p.Q115*)	Ex10 c.1445C>T (p.P482L)	1
P401	Ex4 c.851C>G (p.A284G)	Ex18 c.2510G>A (p.R837H)	1
		<b>TOTAL</b>	<b>10</b>

**Tabla 11. Pacientes con genotipo formado por variante patogénica y VSD o dos VSDs.**

Pacientes en los que se detectó una mutación descrita como patogénica junto con una mutación descrita como de significado desconocido o no descrita previamente (en negrita). Leyenda: Ex: exón, I: Intrón

Dos pacientes mostraron una situación que estimamos particular (**Tabla 11**). Se trata de los pacientes P346 y P218. El primero, había sido diagnosticado de la forma adulta de la enfermedad mediante biopsia muscular. Era homocigoto para la variante c.1194+5G>A, considerada como no patogénica (Kroos et al., 2012). El paciente P218, era portador en heterocigosidad de esta misma variante en uno de sus alelos y de la variante patogénica c.1781G>A (p.R594H) en el otro. Teniendo en cuenta que la variante c.1194+5G>A era considerada no patogénica, se debía interpretar que el paciente P346 no presentaba alteración alguna en el gen y, el P218, era portador de la enfermedad. Sin embargo, el diagnóstico realizado por biopsia muscular de P346 nos indujo a dudar de la validez con la que la variante c.1194+5G>A estaba siendo considerada, por lo que realizamos un estudio funcional sobre su posible papel en la alteración del procesamiento del mRNA, cuyos resultados se presentan más adelante en el apartado 4.1.6.7 (Amiñoso et al., 2019).

ID. pacientes	Alelo 1	Alelo 2	Nº total pacientes
P218	<b>I7 c.1194+5G&gt;A</b>	Ex13 c.1781G>A (p.R594H)	1
P346	<b>I7 c.1194+5G&gt;A</b>	<b>I7 c.1194+5G&gt;A</b>	1
		<b>TOTAL</b>	<b>2</b>

**Tabla 12. Pacientes en los que se identificó una variante descrita como no patogénica**

Pacientes en los que se detectó la variante c.1194+5G>A descrita en la base de datos del Pompe Center como no patogénica. En el caso del P218 aparece junto con una mutación descrita como patogénica. Leyenda: Ex: exón, I: Intrón



## RESULTADOS

Asimismo, en un grupo de 36 pacientes se identificó una variante patogénica en uno de los alelos y no se identificó ninguna alteración en el otro, lo que suponía la condición de portador de la enfermedad (**Tabla 12**). En este grupo de pacientes, que nos fueron enviados como probandos con sospecha de la enfermedad, se identificó un genotipo que, en sí mismo, no justificaba su cuadro clínico. Al analizar estos resultados se tuvo en cuenta que el estudio molecular no analiza el 100% de la estructura del gen GAA, por lo que en alguno de los pacientes no se descarta la posibilidad de que pudiera existir una variante patogénica no identificada.

El diagnóstico de portador que se determinó en estos pacientes debe ser un elemento importante en cualquier proceso de asesoramiento genético que reciban en el futuro.

Una vez más se puso de manifiesto la prevalencia de la mutación c-32-13T>G en este grupo de pacientes, ya que 25 de ellos eran portadores de este cambio. El resto de los cambios identificados suponen 4 variantes *missense*, 2 que alteran el procesamiento del mRNA y una mutación *nonsense*, que genera un codón de parada prematuro.

ID. pacientes	Alelo 1	Nº total pacientes
P107, P117, P143, P144, P146, P178, P181, P207, P219, P242, P274, P279, P285, P327, P328, P341, P370, P374, P375, P402, P450, P458, P464, P465, P480	I1 c.-32-13T>G	25
P145	Ex15 c.2173C>T (p.R725W)	1
P187, P288	Ex14 c.1933G>T (p.D645Y)	2
P209, P407	Ex16 c.2238G>C (p.W746C)	2
P211	Ex12 c.1655T>C (p.L552P)	1
P243	Ex18 c.2646+2T>A	1
P347, P348, P351	I6 c.1076-1G>C	3
P418	Ex3 c.573C>A (p.Y191*)	1
<b>TOTAL</b>		<b>36</b>

**Tabla 13. Pacientes portadores de enfermedad de Pompe.**

Pacientes en los que se ha detectado una mutación descrita como patogénica. Leyenda: Ex: exón I: Intrón

Otro grupo de 12 pacientes resultaron portadores en uno de los alelos de una VSD, sin presentar ninguna alteración en el otro alelo. Casi todas las VSDs correspondían con cambios en zona codificante, que implicaban sustitución de un aminoácido por otro, excepto en el P345 que presentaba una variante sinónima, concretamente c.546G>C (p.T182T). Esta situación no permitía establecer de manera definitiva el estado de portador de la enfermedad en este grupo de pacientes (**Tabla 13**). No obstante, como consecuencia de la reclasificación de VSDs realizada aplicando los criterios de la ACMG al grupo de 10 pacientes presentados en la **Tabla 10**, el paciente P167, portador en heterocigosidad de la variante c.1445C>T (p.P482L), se consideraría portador de la enfermedad.

## RESULTADOS

ID. pacientes	Alelo 1	Nº total pacientes
P108	Ex18 c.2510G>A (p.R837H)	1
P167	Ex10 c.1445C>T (p.P482L)	1
P208, P350, P412	Ex6 c.1048G>A (p.V350M)	3
P256	Ex14 c.1996G>T (p.A666S)	1
P267	Ex8 c.1310G>A (p.R437H)	1
P345	Ex2 c.546G>C (p.T182T)	1
P355	Ex11 c.1607A>G (p.N536S)	1
P398	Ex7 c.1153C>T (p.R385C)	1
P422	Ex2 c.265C>T (p.R89C)	1
<b>TOTAL</b>		<b>15</b>

**Tabla 14. Pacientes portadores de una variante descrita como de significado desconocido.**

Pacientes en los que se detectó solo una variante descrita como de significado desconocido o no descrita, en las bases de datos consultadas. Leyenda: Ex: exón

### 4.1.3. Pacientes con alteraciones genéticas no descritas

Además de los resultados hasta ahora enumerados, se identificó un grupo de 23 pacientes en los que se observaron variantes no descritas antes de la realización de este trabajo (**Tablas 14 y 15**). En 18 de esos pacientes (**Tabla 14**), se encontró un genotipo compuesto por una variante patogénica en un alelo y una variante no descrita previamente en el otro. Además, en el paciente P129 el genotipo estaba formado por una variante no identificada, la p.S560C, y por la p.E192D, descrita como de significado desconocido. El paciente P479, presentaba en homocigosidad la variante p.D443V, no descrita previamente. Por último, el paciente P349 presentaba la variante p.E471\*, no descrita antes, también en homocigosidad (**Tabla 14**). En 15 de los 21 pacientes incluidos en la **Tabla 14** se pudo determinar el diagnóstico genético de la enfermedad de manera firme, dado que eran portadores de un genotipo compuesto por una variante patogénica conocida y por una variante nueva de naturaleza claramente patogénica. La naturaleza de estas variantes incluía 2 grandes deleciones, detectadas por *MLPA* (**Tabla 14-verde**), 4 variantes de procesamiento que afectan a los sitios canónicos de *splicing* (**Tabla 14-naranja**) y 6 mutaciones *nonsense* o de cambio de fase de lectura (**Tabla 14-rojo**). Por último, 6 pacientes (**Tabla 14-azul**) presentaban una variante *missense* nueva, de patogenicidad incierta. Entre los 15 pacientes, 4 presentaban la forma infantil de la enfermedad, y sus genotipos mostraban variantes muy lesivas para la funcionalidad del gen en ambos alelos.

## RESULTADOS

ID. pacientes	Alelo 1	Alelo 2	Nº total pacientes
P11 (I)	Ex5: c.925G>A (p.G309R)	<b>g.80082249-80121055delins14bp</b>	1
P481	I1 c.-32-13T>G	<b>c.2799+244_2859+63del</b>	1
P103, P158, P210, P473	I1 c.-32-13T>G	<b>I13 c.1889-1G&gt;A</b>	4
P290	I1 c.-32-13T>G	<b>I6 c.1075+1G&gt;A</b>	1
P300	I1 c.-32-13T>G	<b>I11 c.1637-2delA</b>	1
P451 (I)	Ex2 c.236-246del (p.P79Rfs*13)	<b>I5-Ex6 c.956-5_957del (p.?)</b>	1
P124	I1 c.-32-13T>G	<b>Ex13 c.1853G&gt;A (p.W618*)</b>	1
P151	I1 c.-32-13T>G	<b>Ex12 c.1657C&gt;T (p.Q553*)</b>	1
P180	Ex15 c.2173C>T (p.R725W)	<b>Ex16 c.2316G&gt;A (p.W772*)</b>	1
P230 (I)	<b>Ex9 c.1416dupC (p.G473Rfs*33)</b>	Ex19 c.2662G>T (p.E888*)	1
P273	I1 c.-32-13T>G	<b>Ex17 c.2334_2335dupGC (p.P779Rfs*3)</b>	1
P349 (I)	<b>Ex9 c.1411G&gt;T (p.E471*)</b>	<b>Ex9 c.1411G&gt;T (p.E471*)</b>	1
P50	Ex6 c.1064T>C (p.L355P)	<b>Ex8 c.1209C&gt;G (p.N403K)</b>	1
P106	I1 c.-32-13T>G	<b>Ex10 c.1447G&gt;T (p.G483W)</b>	1
P129	Ex3 c.576G>C (p.E192D) #	<b>Ex12 c.1679C&gt;G (p.S560C)</b>	1
P352	I1 c.-32-13T>G	<b>Ex15 c.2141A&gt;C (p.H714P)</b>	1
P388	<b>Ex4 c.713C&gt;T (p.P238L)</b>	Ex6 c.1064T>C (p.L355P)	1
P479	<b>Ex9 c.1328A&gt;T (p.D443V)</b>	<b>Ex9 c.1328A&gt;T (p.D443V)</b>	1
<b>TOTAL</b>			<b>21</b>

**Tabla 15. Pacientes en los que se identificó una mutación no descrita previamente.**

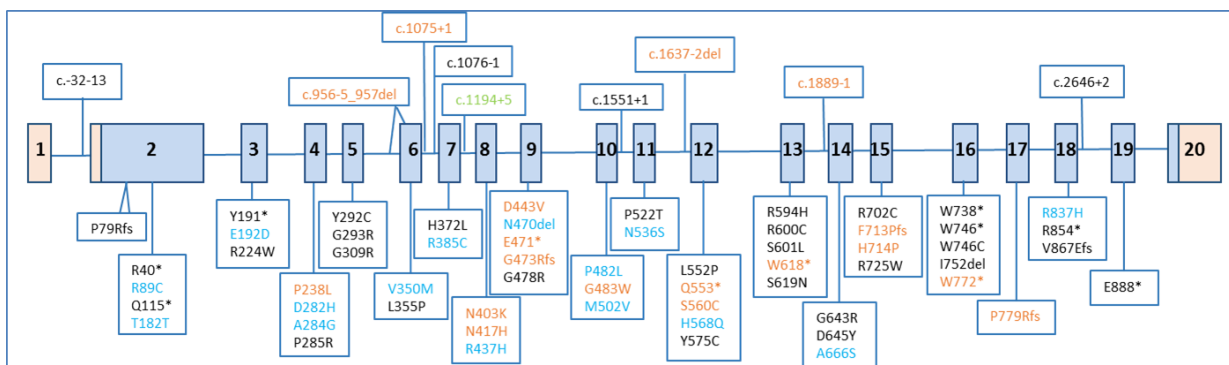
Pacientes en los que se detectó una mutación descrita en la base de datos como patogénica, junto con otra variante no descrita (en negrita). En verde, los cambios detectados mediante MLPA, en naranja los que afectan a los sitios canónicos de *splicing* y en rojo las variantes que implicaban la introducción de un codón de Stop prematuro. En azul, las variantes que implican un cambio de aminoácido y de las que no se puede establecer la patogenicidad. Leyenda: Ex: exón; I: Intrón; # VSD; (I): Pacientes con fenotipo clásico.

Por último, 2 pacientes eran portadores en heterocigosidad de variantes no descritas, pero no se observó alteración alguna en el otro alelo (Tabla 15). En este caso, sólo la variante del paciente P199 era de naturaleza patogénica, al introducir un cambio de fase de lectura. La otra variante era una *missense* y patogenicidad incierta.

ID. pacientes	Alelo 1	Nº total pacientes
P199	Ex15 c.2135_2136delTG (p.F713Pfs*24)	1
P301	Ex8 c.1249A>C (p.N417H)	1
<b>TOTAL</b>		<b>2</b>

**Tabla 16. Pacientes portadores de una variante no descrita.**  
Leyenda: Ex: exón

En la **Figura 18** se muestra la distribución a lo largo del gen de todas las variantes detectadas.



**Figura 18. Localización de las mutaciones detectadas, distribuidas a lo largo de toda la estructura del gen GAA.**

Código de colores: Negro: mutaciones descritas como patogénicas. Naranja: variantes descritas por primera vez en este trabajo. Azul: variantes descritas como de significado desconocido. Verde: variante descrita como no patogénica.

4.1.4. Caracterización molecular de nuevas variantes identificadas por MLPA

4.1.4.1. Caracterización de la variante g.80082249-80121055delins14bp presente en el paciente P11

El análisis del paciente P11 permitió identificar en el exón 5 que era portador en aparente homocigosidad de la mutación c.925G>A (p.G309R), una variante considerada como patogénica. Un resultado que no resulto coherente con los genotipos obtenidos después al estudiar las muestras de los padres. La madre era portadora en heterocigosidad de ese cambio identificado en el caso índice, pero el padre no era portador de dicho cambio. Esto nos llevó a realizar un análisis de segregación de marcadores de paternidad con el kit *Identifiler™*, que nos confirmaron que la muestra que se estaba analizando era realmente del padre del caso índice. Esta situación fue la que nos hizo plantearnos la puesta a punto de un *MLPA* específico para el gen *GAA* que nos permitiera identificar posibles reordenamientos de gran tamaño que no fueran visibles mediante secuenciación *Sanger*.

El estudio del paciente P11 con el nuevo *MLPA* permitió detectar un descenso en el área de los picos de todas las sondas localizadas a lo largo de la estructura del gen, lo que indicaba una pérdida de dosis génica de toda su estructura. Para completar la caracterización se hicieron estudios adicionales mediante un *array-CGH*, que confirmaron la delección. PCRs con oligonucleótidos flanqueantes al punto de ruptura y posterior secuenciación de los fragmentos amplificados, permitieron caracterizar los extremos de la delección (**Figura 19**). Se determinó que implicaba una secuencia *Alu*, y que el tamaño total de la región delecionada correspondía a 38,8Kb. Además, se observó una inserción de 14pb entre los puntos de ruptura. Esta delección suponía la pérdida de todo el gen *GAA* y, aproximadamente, un tercio del gen contiguo *CCDC40*. Siendo la delección más grande detectada en esta enfermedad hasta la fecha (Amiñoso et al., 2013).

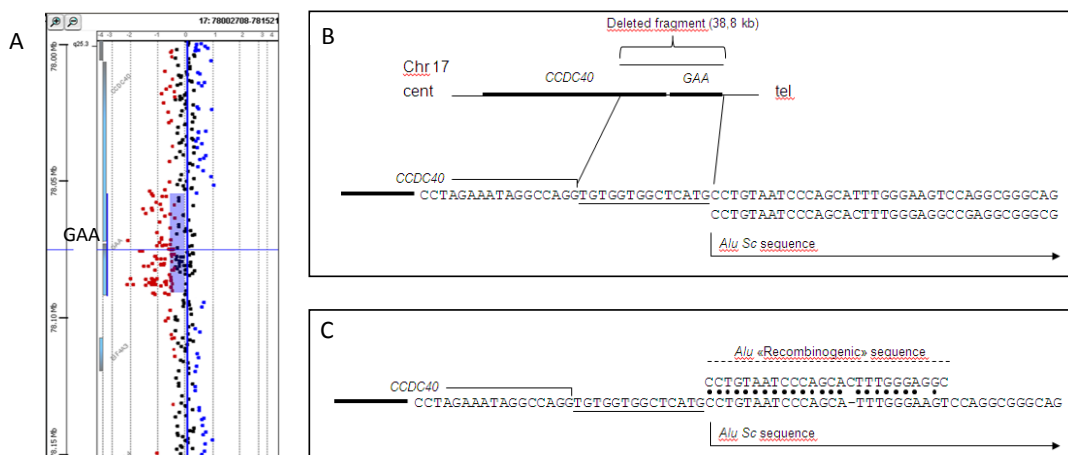
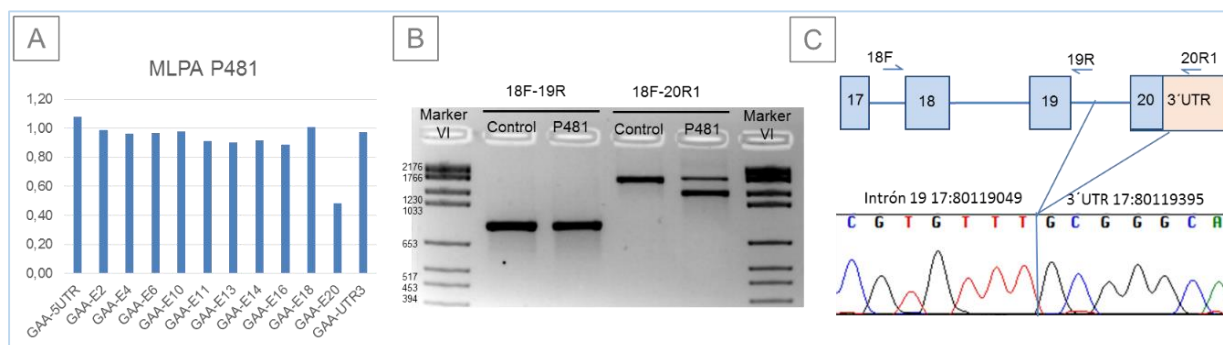


Figura 19. Caracterización de los puntos de ruptura de la delección detectada en el paciente P11.

**A:** Imagen del array CGH con el que se puede apreciar la pérdida de heterocigosidad con respecto al control y que aparece marcada por los puntos rojos. **B-C:** corresponde a la caracterización de los puntos de ruptura de la delección que implicaría secuencia *Alu*. Se muestra la región delecionada (38.8kb) y la secuencia de 14pb insertada (secuencia subrayada).

#### 4.1.4.2. Caracterización de la variante c.2799+244\_2859+63del presente en el paciente P481

En el paciente P481 se detectó una disminución en la señal de la sonda que hibrida en el exón 20, manteniéndose normal la sonda que hibrida en el exón 18 y el resto de las sondas a lo largo del gen, resultado que sugería un deleción en el entorno del exón 20. Se realizó una PCR con oligonucleótidos que podían flanquear la deleción y se pudo determinar mediante secuenciación posterior que la deleción implicaba parte del intrón 19 y toda la parte codificante del exón 20 (**Figura 20**).



**Figura 20. Caracterización de los puntos de ruptura de la deleción detectada en el paciente P481.**

**A:** Imagen del resultado del MLPA en el que se aprecia un descenso de la sonda que hibrida en el ex20. **B:** Imagen del gel de agarosa con cebadores flanqueantes a la zona de la deleción. **C:** Secuenciación realizada para caracterizar los puntos de ruptura y representación gráfica del fragmento que implica la deleción.

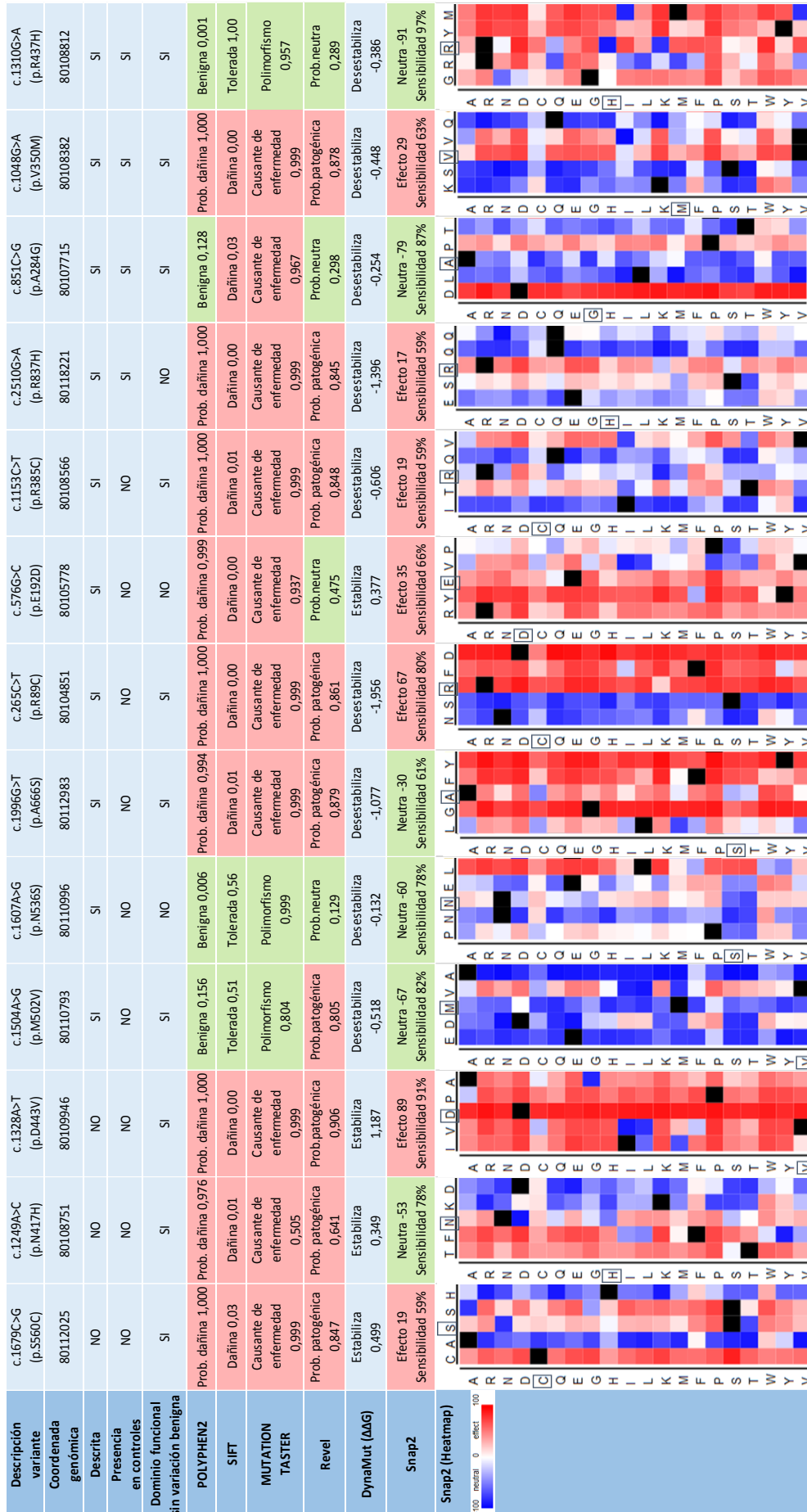
#### 4.1.5. Análisis bioinformático de las variantes no descritas y de las descritas como de significado desconocido.

Los programas informáticos de predicción *in-silico* son elementos habituales en el análisis de variantes cuya patogenicidad no ha sido aún determinada. Trece variantes *missense* no descritas previamente, o consideradas como VSDs, fueron analizadas “*in silico*” con herramientas de predicción de patogenicidad (**Tabla 16**). Se han utilizado distintas herramientas bioinformáticas de predicción, como son *Polyphen2*, *SIFT*, *Mutation Taster* y *Revel*. Con estas herramientas de predicción, ocho variantes fueron clasificadas como probablemente patogénicas o causantes de enfermedad, en función de los criterios específicos aplicados por cada una de ellas. Solo en el caso de las variantes p.N536S y la p.R437H, todas las herramientas coincidían en el carácter benigno de las mismas. Existen algunas variantes en las que el resultado no era unánime, la p.M502V, en la que solo *Revel* le otorga una posible patogenicidad, la p.E192D en la que todas las herramientas le otorgan un posible efecto patogénico excepto *Revel*, y la variante p.A284G, que tanto *Polyphen2* como *Revel* le otorgan un papel neutro frente al resultado que tanto *SIFT* como *Mutation Taster* reportan.

Se utilizó la herramienta *Dynamut* con el fin de determinar el impacto de cada mutación sobre la dinámica y la estabilidad de la estructura proteica. Esta herramienta determinó que la mayoría de los

cambios producirían una desestabilización de la estructura de la proteína. Solo en el caso de las variantes p.S560C, p.N417H, p.D443V y p.E192D anticipó que los cambios introducidos producían una estabilización de la estructura, provocando un aumento de la rigidez de la molécula.

Para medir el grado de conservación de los aminoácidos se aplicó la herramienta *Snap2*. Los aminoácidos S560, N417, D443, A666, R89, E192, R837, V350 y el R4379, presentaban un alto grado de conservación. Las variantes p.S560C, p.D443V, p.R89C, p.E192D, p.R837H y p.V350M implicaban un cambio que previsiblemente podía tener un efecto deletéreo, pero en el caso de los cambios p.N417H, p.A666S y p.R437H serían tolerados. El resto de los aminoácidos estudiados, M502, N536, R385 y A284 tendrían un bajo grado de conservación, por lo tanto, los cambios serían en principio mejor tolerados. Sin embargo, el cambio p.R385C no sería bien tolerado y *Snap2* predice que tendría un efecto deletéreo sobre la proteína.



**Tabla 17. Análisis bioinformático de las variantes no descritas y de las descritas como de significado desconocido.**

En la tabla se muestran las variantes analizadas, si están o no descritas, si se han detectado en controles para lo que se ha usado la base de datos poblacional ExAC y si la variante ha sido detectada o no junto con otra variante patogénica. Se muestra el resultado del análisis con las herramientas de predicción de patogénicidad Polyphen 2, SIFT y Mutation Taster, mostrándose en rojo las que se predicen como patogénica y en verde las que se predicen como benignas. La última columna, se refiere al resultado obtenido con InterVar y muestra la clasificación final que tendrían las variantes en base a todos los datos introducidos. Ex: Exón, VSD: Variante de Significado Desconocido

#### 4.1.6. Análisis funcional de las variantes con potencial implicación en el procesamiento del mRNA.

Las estrategias de análisis genético están principalmente dirigidas a la identificación de variantes potencialmente patogénicas en muestras de gDNA. En ocasiones, las variantes identificadas provocan alteraciones en mecanismos moleculares que no son evidentes por la propia naturaleza de la variante. Entre estos mecanismos que pueden resultar alterados y pasar desapercibidos se encuentra el procesamiento del mRNA.

El análisis bioinformático realizado en este trabajo con la aplicación *Alamut* sobre las variantes cuya patogenicidad no estaba aún bien definida, determinó que durante el proceso de expresión del gen *GAA* algunas de ellas podrían alterar el normal procesamiento de los exones o intrones en los que estaban localizadas. No obstante, ante la enorme dificultad para predecir los efectos de las variantes sobre el procesamiento del mRNA, se consideró oportuno la aplicación de un análisis funcional a todas las recogidas en la **Tabla 17**, de modo que permitiera la identificación y caracterización de las que pudiesen afectar esa etapa en la expresión del gen *GAA*. En esta parte del trabajo resultó limitante la disponibilidad de muestra de sangre periférica de algunos pacientes para la extracción de RNA. El procedimiento se realizó en dos partes. Primero, un análisis cualitativo para identificar productos aberrantes del mRNA en todos los exones del gen y, segundo, una qPCR que permitía cuantificar cada exón individualmente.

P70	Ex12: c.1655T>C (p.L552P)	<b>Ex12 c.1704C&gt;G (p.H568Q)</b>
P106	I1 c.-32-13T>G	<b>Ex10 c.1447G&gt;T (p.G483W)</b>
P108	<b>Ex18 c.2510G&gt;A (p.R837H)</b>	Normal
P145	<b>Ex15 c.2173C&gt;T (p.R725W)</b>	Normal
P210	I1 c.-32-13T>G	<b>I13 c.1889-1G&gt;A</b>
P279	I1 c.-32-13T>G	<b>I6: c.1076-15G&gt;C</b>
P290	I1 c.-32-13 T>G	<b>I6 c.1075+1G&gt;A</b>
P299	<b>Ex10 c.1504A&gt;G (p.M502V)</b>	<b>Ex10 c.1504A&gt;G (p.M502V)</b>
P300	I1 c.-32-13T>G	<b>I11 c.1637-2delA</b>
P301	<b>Ex8 c.1249A&gt;C (p.N417H)</b>	Normal
P318	EX2 c.343C>T (p.Q115*)	<b>Ex10 c.1445C&gt;T (p.P482L)</b>
P345	<b>Ex2 c.546G&gt;C (p.T182T)</b>	Normal
P346	<b>I7 c.1194+5G&gt;A</b>	<b>I7 c.1194+5G&gt;A</b>
P349	<b>Ex9 c.1411G&gt;T (p.E471*)</b>	<b>Ex9 c.1411G&gt;T (p.E471*)</b>
P352	I1 c.-32-13T>G	<b>Ex15 c.2141A&gt;C (p.H714P)</b>
P479	<b>Ex9 c.1328A&gt;T (p.D443V)</b>	<b>Ex9 c.1328A&gt;T (p.D443V)</b>

Tabla 18. Pacientes incluidos en el análisis funcional para analizar la implicación de las variantes de interés (en negrita) en el procesamiento.

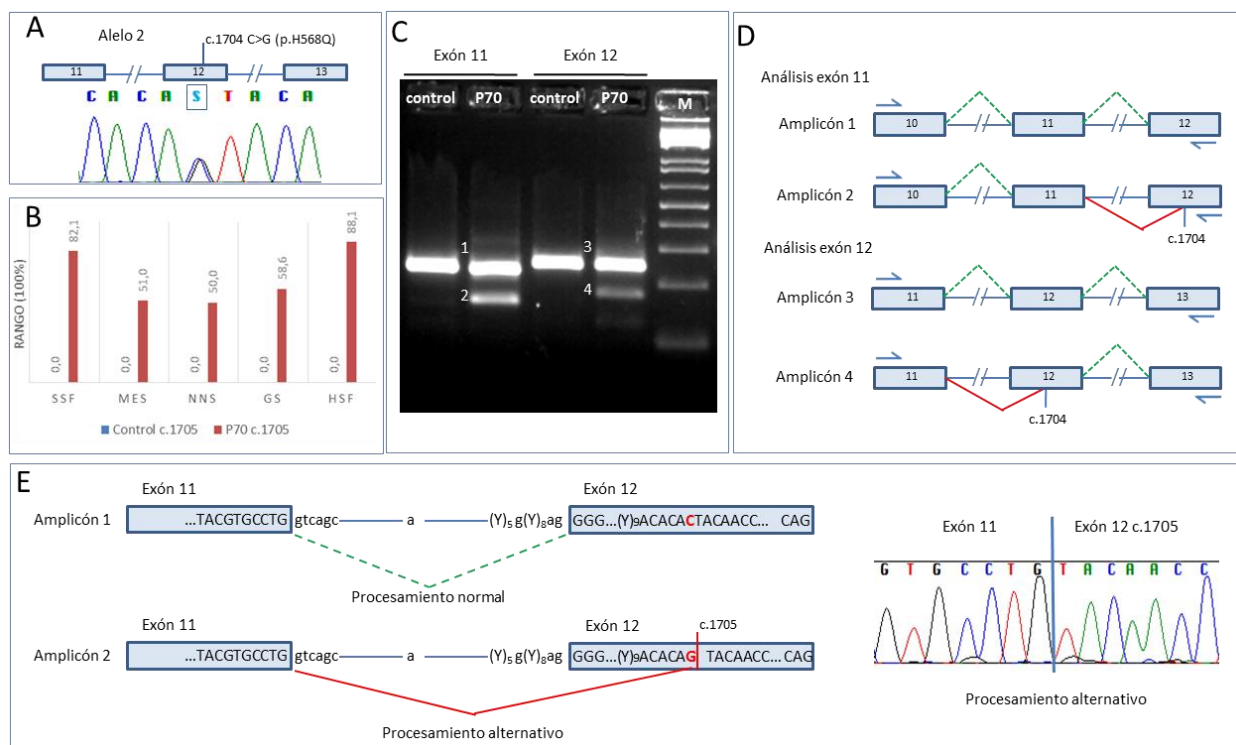
##### 4.1.6.1. Paciente P70.

En el paciente P70 se identificaron las variantes c.1655T>C (p.L552P) y c.1704C>G (p.H568Q), ambas localizadas en el exón 12. Esta última es el objeto del análisis funcional, ya que ha sido descrita en el Pompe Center como de significado desconocido (**Figura 21A**). Todas las herramientas bioinformáticas predecían la aparición de un posible sitio AG, 3' aceptor, en el c.1705, que no existía cuando la secuencia



## RESULTADOS

en el c.1704 era la normal (**Figura 21B**), lo que se evidencia por la diferencia de puntuación entre ambas secuencias. Se hizo el análisis cualitativo y se detectó la aparición de bandas de tamaño anormal, tanto en el análisis del exón 11 (banda 2) como en el del exón 12 (banda 4) (**Figura 21C**), frente a las obtenidas en la muestra control, bandas 1 y 3. Estas bandas corresponden con un procesamiento alternativo que, debido a la localización de los oligonucleótidos utilizados, se evidencia en el análisis de ambos exones (**Figura 21D**). Se secuenciaron las bandas aberrantes y se pudo corroborar que el cambio de c.1704C>G generaba un nuevo 3'aceptor, lo que provocaba un procesamiento alternativo del intrón 11. De esta manera, el extremo 5' donador se procesaba normalmente, pero en el extremo 3' se utilizaría el nuevo AG generado en el alelo mutado, situado en c.1705, dentro del exón 12 (**Figura 21E**), creando así una población de mensajero en el que se pierden 68pb, que implicaría un cambio en la fase de lectura y la introducción de un codón de parada precoz 67 aminoácidos después, ya en el exón 14. El análisis cuantitativo mediante qPCR, no mostró cambios significativos con respecto al control (datos no mostrados).



**Figura 21. Análisis del procesamiento de un control normal y del paciente P70**

**A:** Secuencia genómica del exón 12 realizada en el paciente P70 donde está recuadrada la variante analizada. **B:** Resultado del análisis obtenido con la aplicación *AluMut*. Puntuación con la secuencia normal (azul) y con la secuencia mutada (rojo). **C:** Gel de agarosa del análisis de mRNA de los exones 11 y 12 con oligonucleótidos flanqueantes, realizado tanto en el paciente como en una muestra control. **D:** Representación gráfica del análisis realizado, que explicaría el resultado obtenido en el gel de agarosa. El amplicón 1 y 3 corresponderían con el procesamiento normal y el amplicón 2 y 4 corresponderían con la representación del procesamiento aberrante que se realiza en el paciente. Los exones están representados por las cajas azules y el número de cada exón está en el interior de cada caja. Las líneas azules representan los intrones, las líneas verdes corresponden al procesamiento normal y las líneas rojas marcan el procesamiento aberrante. La localización de los oligonucleótidos utilizados para realizar la amplificación se representan mediante (↔). **E:** Representación detallada donde se muestra el procesamiento normal y el alternativo. En las cajas azules se muestra la secuencia normal marcando en rojo la variante, (Y)<sub>n</sub> se usa para identificar un grupo de pirimidinas y su número. En la parte derecha, está la secuencia obtenida donde se puede también apreciar el resultado del procesamiento aberrante.

4.1.6.2. Paciente P210.

En el paciente P210 se identificó en el alelo 1, la variante c.-32-13T>G, localizada en el intrón 1 y, en el alelo 2, la c.1889-1G>A, variante genética localizada en el dinucleótido consenso del aceptor de *splicing* del intrón 13 (**Figura 22A**), que no estaba descrita en las bases de datos consultadas. Todas las herramientas bioinformáticas utilizadas predecían que el cambio c.1889-1G>A provocaría un debilitamiento del sitio 3' aceptor del intrón 13 (**Figura 22B**). Se realizó el análisis cualitativo con cebadores que hibridaban en los exones flanqueantes. En las condiciones en las que se hicieron las electroforesis en gel de agarosa se pudo apreciar una banda de tamaño anormal (**Figura 22C**). Cuando se secuenció la banda indicada como amplicón 2, se pudo determinar que el cambio c.1889-1G>A hacía que la maquinaria de procesamiento utilizase una secuencia alternativa. Como se puede ver en la representación detallada (**Figura 22D**), en el c.1898-1899 se localiza un AG precedido de una secuencia con un grupo de cinco pirimidinas. Al perderse la secuencia AG del 3' aceptor normal, la maquinaria de procesamiento identifica este otro punto como un posible 3' aceptor, lo que produce un *splicing* alternativo. La consecuencia de esta mutación es que se pierden 11 pb del exón 14, y se produce una alteración de la fase de lectura con la aparición de un codón de parada prematuro. Con el análisis cuantitativo (**Figura 22E**) se pudo detectar una expresión inferior al 40-50% en todos los exones analizados con respecto a la expresión detectada en el control.

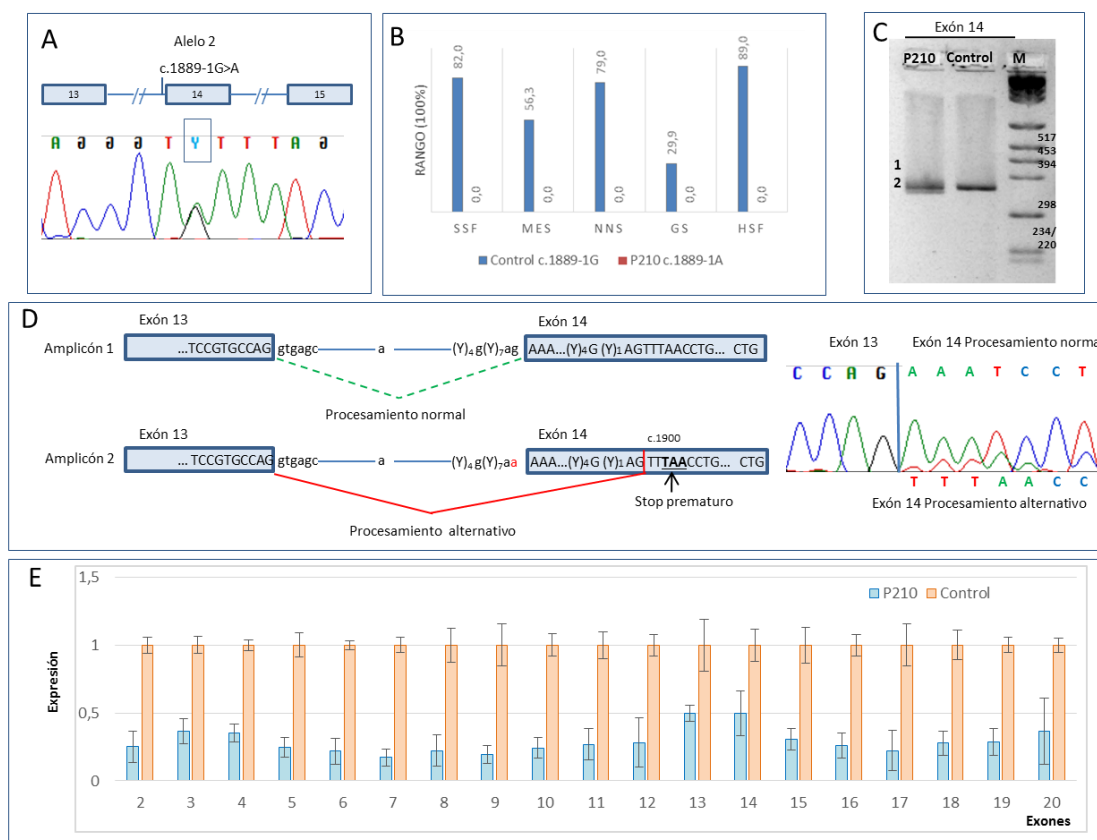


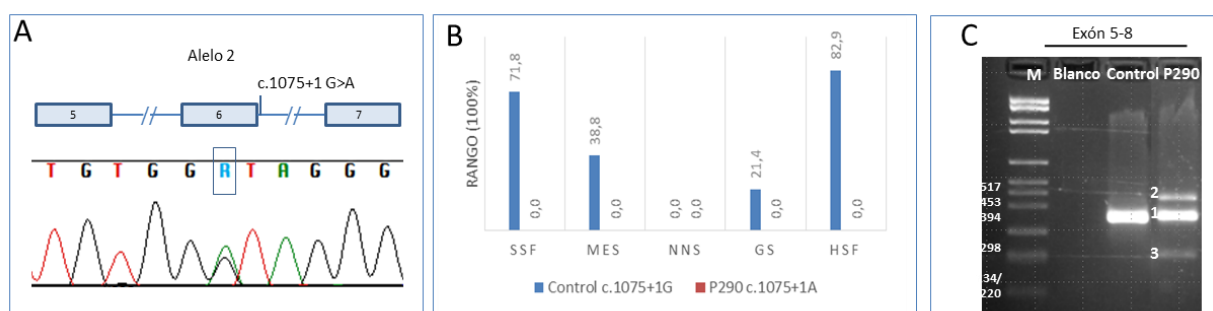
Figura 22. Análisis del procesamiento de un control normal y del paciente P210.

A: Secuencia genómica del exón 14 realizada en el P210 donde está recuadrada la variante analizada. B: Resultado del análisis obtenido con la aplicación *Alamut*. Puntuación con la secuencia normal (azul) y con la secuencia mutada (rojo).

**Figura 22 (continuación): Análisis del procesamiento de un control normal y del paciente P210.** C: Gel de agarosa del análisis de mRNA del exón 14 realizado con oligonucleótidos localizados en los exones 13 y 15, realizado tanto en el paciente como en una muestra control. D: Representación detallada de ambos amplicones con el procesamiento normal y el alternativo. En las cajas azules se muestra la secuencia normal marcando en rojo la variante, (Y)<sub>n</sub> se usa para identificar un grupo de pirimidinas y su número. En negrita y subrayado se marca el codón stop prematuro. En la parte derecha, está la secuencia obtenida donde se puede también apreciar el resultado del procesamiento normal (superior) y del procesamiento aberrante (inferior). E: Representación del análisis de la cuantificación del mRNA mediante qPCR. Se usó la *B2M* para normalizar. Los valores del control se llevaron al 100%. Y las barras de error indican la desviación estándar (n=3)

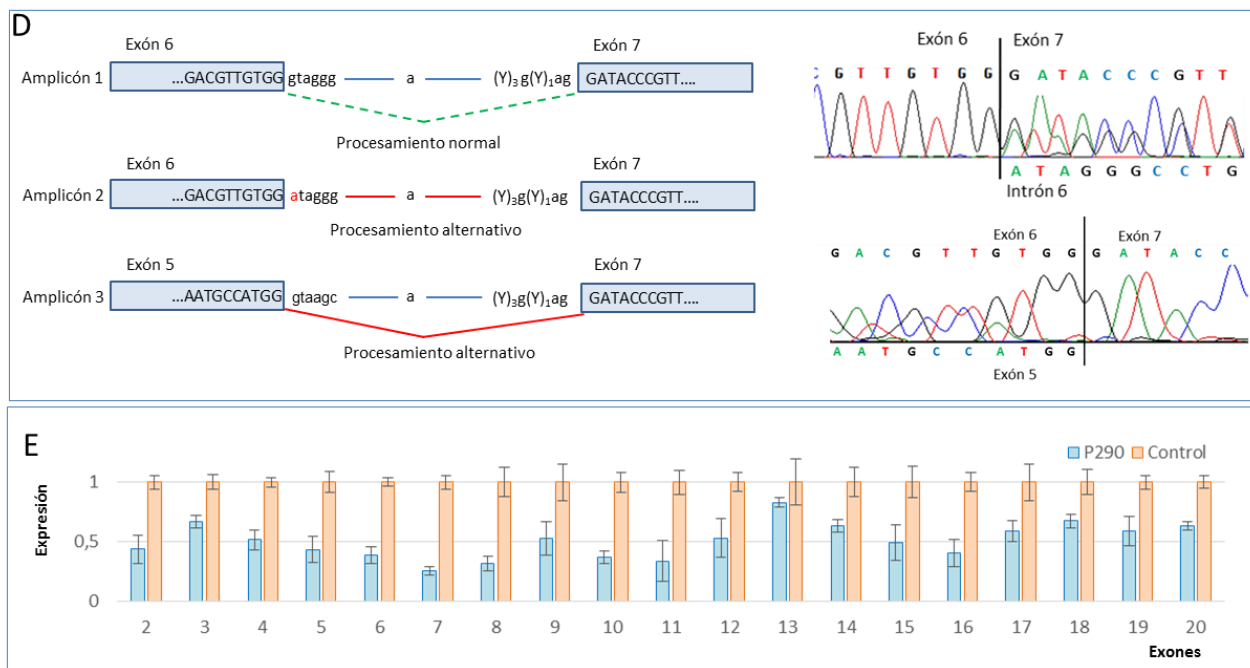
#### 4.1.6.3. Paciente P290.

En el paciente P290 se identificó el genotipo formado por las variantes c.-32-13T>G y c.1075+1G>A. Esta última, localizada en el dinucleótido del 5' donador del intrón 6 (**Figura 23A**), no había sido descrita previamente. Todas las herramientas bioinformáticas utilizadas predecían que el cambio de G>A provocaría un debilitamiento del 5' donador (**Figura 23B**) del intrón 6. Se realizó el análisis cualitativo y se pudo apreciar la aparición de tres bandas. En la **Figura 23C** se muestra el resultado de las amplificaciones con cebadores que hibridaban en los exones 5 y 8, por la mejor definición del patrón de bandas. Se pudo observar una banda de tamaño normal (amplicón 1) y dos de tamaño anormal (amplicones 2 y 3), que no aparecían en la muestra control, lo que sugería un procesamiento aberrante (**Figura 23C**). El amplicón 2 se purificó a partir del gel de agarosa, se secuenció y se observó la aparición de una doble secuencia que correspondía, por un lado, con la permanencia del intrón 6 y, por otro lado, con la secuencia del procesamiento normal del intrón, que era fruto de una contaminación del alelo normal en la banda purificada (**Figura 23D**). La purificación y posterior secuenciación del amplicón 3 puso de manifiesto un segundo mecanismo de procesamiento aberrante en el que se pierde la secuencia completa del exón 6, junto con los intrones 5 y 6. (**Figura 23D**). En el análisis cuantitativo se pudo detectar una expresión claramente disminuida en todos los exones analizados con respecto a la expresión obtenida en la muestra control (**Figura 23E**).



**Figura 23. Análisis del procesamiento de un control normal y del paciente P290.**

**A:** Secuencia genómica del exón 6 donde está recuadrada la variante analizada. **B:** Resultado del análisis obtenido con la aplicación *Alamut*. Puntuación con la secuencia normal (azul), con la secuencia mutada (rojo). **C:** Gel de agarosa del análisis de mRNA con oligonucleótidos localizados en los exones 5 y 8, realizado tanto en la muestra del paciente P290 como en una muestra control. El amplicón 1 corresponde con el procesamiento normal y el amplicón 2 y 3 corresponderían con el procesamiento aberrante que se realiza en el paciente.

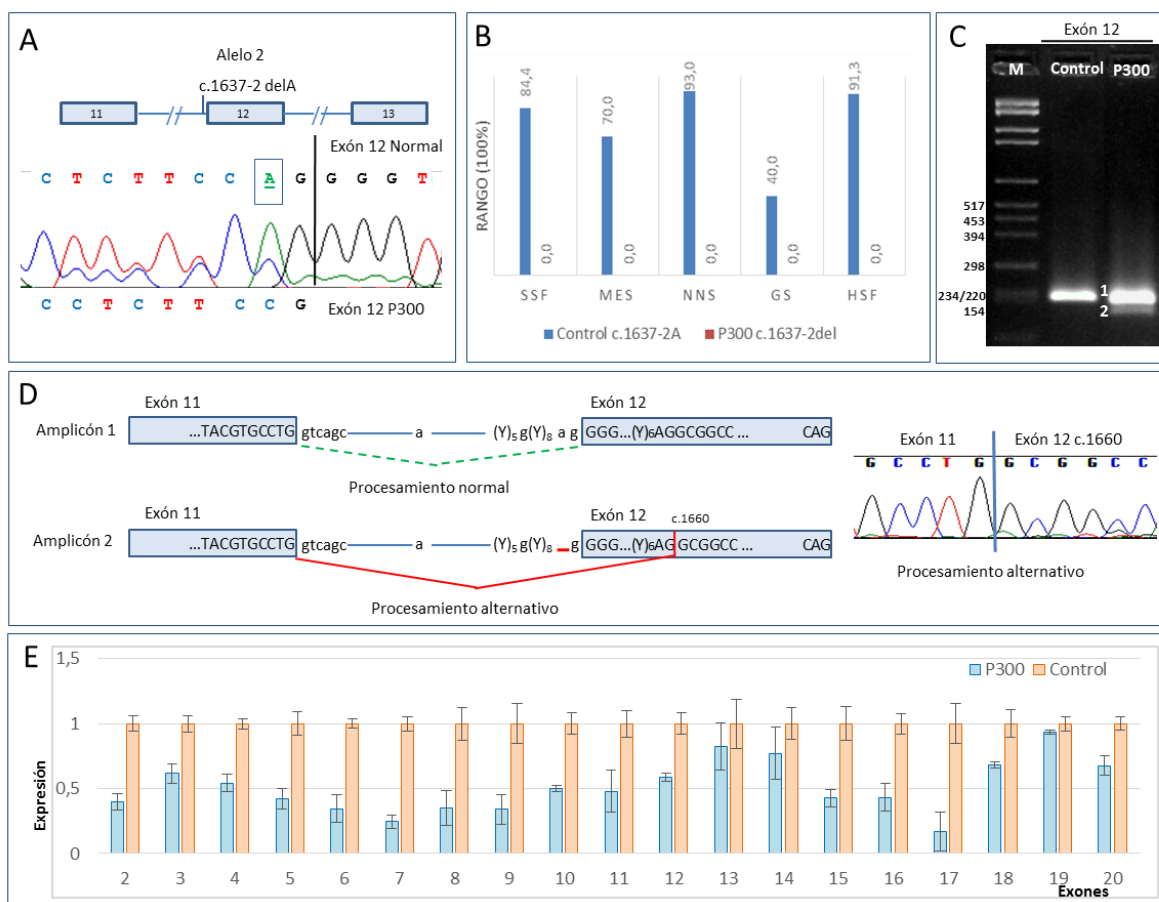


**Figura 23 (continuación): Análisis del procesamiento de un control normal y del paciente P290. D:** Representación detallada del procesamiento de los amplicones 1 y 2 con el procesamiento normal y el alternativo. En las cajas azules se muestra la secuencia normal, marcando en rojo la variante, (Y)<sub>n</sub> se usa para identificar un grupo de pirimidinas y su número. En la parte derecha, está la secuencia obtenida con el resultado del procesamiento normal (superior) y del procesamiento aberrante (inferior). **E:** Representación del análisis de la cuantificación del mRNA mediante qPCR. Se usó la *B2M* para normalizar. Los valores del control se llevaron al 100%. Y las barras de error indican la desviación estándar (n=3)

#### 4.1.6.4. Paciente P300.

En el paciente P300 se identificaron las variantes c.-32-13T>G y c.1637-2delA, localizada esta última en el dinucleótido AG aceptor en el intrón 11 (**Figura 24A**), que no había sido descrita previamente. Las herramientas bioinformáticas predecían la pérdida del 3' aceptor del intrón 11 (**Figura 24B**). Se realizó el análisis cualitativo de todos los exones y en la amplificación del exón 12 se pudo apreciar la banda obtenida del procesamiento normal (amplicón 1), presente tanto en el paciente como en la muestra control, y la aparición de una banda adicional más pequeña y de menor intensidad (amplicón 2) (**Figura 24C**). Se llevó a cabo la secuenciación de ese amplicón y se pudo determinar que la pérdida del sitio normal de *splicing* hacía que la maquinaria de procesamiento utilizase un AG alternativo. Como se puede ver en la representación detallada (**Figura 24D**), en el c.1658-1659 se localiza un AG precedido de una secuencia con un grupo de seis pirimidinas, lo que hacía que al perderse el aceptor normal, la maquinaria reconociera este 3' críptico, llevándose a cabo un *splicing* alternativo que modificaba la pauta de lectura, lo que introducía un codón de parada prematuro 83 aminoácidos después, ya en el exón 14. Con el análisis cuantitativo (**Figura 24E**), se pudo detectar una expresión parcialmente disminuida y desigual en todos los exones analizados.

## RESULTADOS



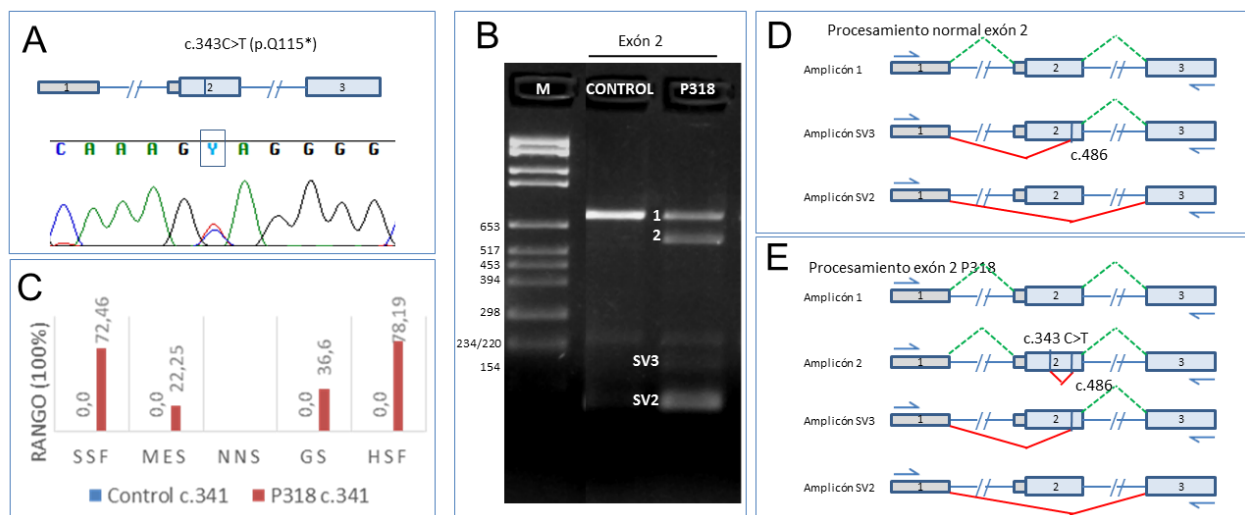
**Figura 24. Análisis del procesamiento de un control normal y del paciente P300.**

**A:** Secuencia genómica del exón 12, donde está recuadrada la variante analizada. **B:** Resultado del análisis obtenido con la aplicación *Alamut*. Puntuación con la secuencia normal (azul) y con la secuencia mutada (rojo). **C:** Gel de agarosa del análisis de mRNA con oligonucleótidos localizados en los exones 11 y 13, realizado tanto en la muestra del paciente P300 como en una muestra control. El amplicón 1 corresponde con el procesamiento normal y el amplicón 2 corresponde con el procesamiento aberrante que se produce en el paciente. **D:** Representación detallada del procesamiento de los amplicones 1 y 2 donde se muestra el procesamiento normal y el alternativo. En las cajas azules se muestra la secuencia normal, marcando en rojo la variante, (Y)<sub>n</sub> se usa para identificar un grupo de pirimidinas y su número. En la parte derecha está la secuencia obtenida, donde se puede apreciar el resultado del procesamiento aberrante. **E:** Representación del análisis de la cuantificación del mRNA mediante qPCR. Se usó la *B2M* para normalizar. Los valores del control se llevaron al 100%, las barras de error indican la desviación estándar (n=3).

### 4.1.6.5. Paciente P318.

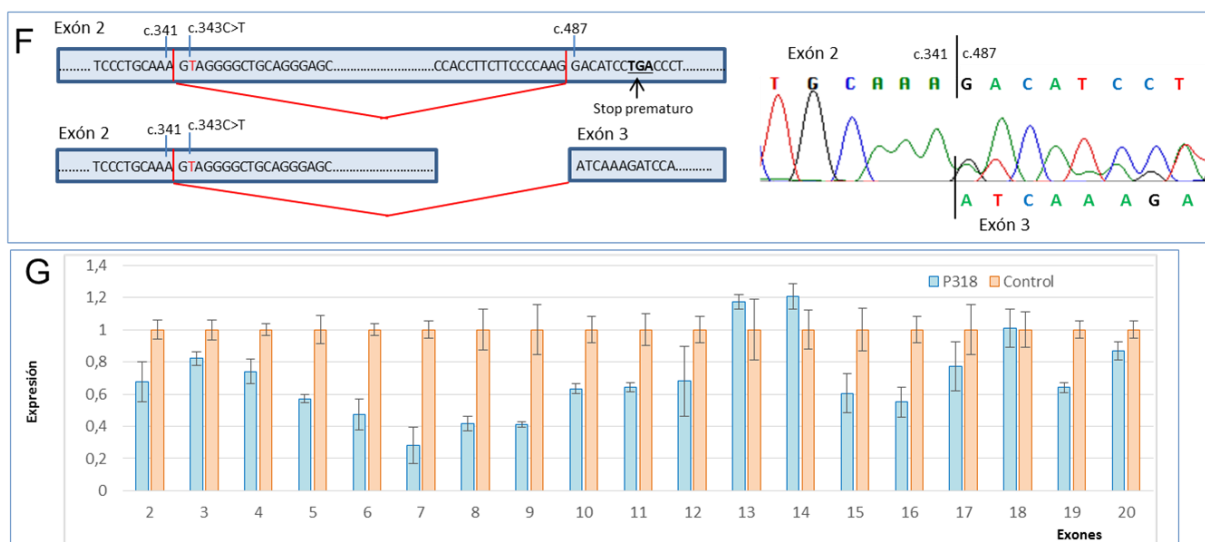
En el paciente P318 se identificó en el alelo 1 la variante c.343C>T (p.Q115\*), localizada en el exón 2, y en el alelo 2 la c.1445C>T (p.P482L), localizada en el exón 10, que había sido descrita previamente en la base de datos del Pompe Center y clasificada como VSD. Las herramientas bioinformáticas de predicción no le otorgaban a la variante c.1445C>T (p.P482L) un papel modificador en el procesamiento normal, pero se hizo el análisis funcional para comprobar este resultado. El análisis cualitativo, mostró un procesamiento normal del exón 10 (datos no mostrados). Como se ha mencionado anteriormente, este paciente presentaba el cambio c.343C>T (p.Q115\*) en el exón 2 (**Figura 25A**). Cuando se analizó este exón, se detectó en el paciente una banda (amplicón 2) de un tamaño menor al normal, que no se detectaba en

el control (**Figura 25B**). Antes del estudio funcional del gen no habíamos realizado un análisis con las herramientas incluidas en *Alamut*<sup>®</sup> de esta variante situada en exón 2, pero a la vista de los resultados obtenidos en el gel de agarosa se realizó ese análisis y todas las herramientas, excepto *NNS* que no arrojó resultado positivo, mostraban la aparición de un 5' donador en el c.341, que no mostraban en la secuencia control (**Figura 25C**). Debido a la complejidad del procesamiento del exón 2 normal (**Figura 25D-E**) y, para poder llevar a cabo la secuenciación del *splicing* aberrante de la forma más limpia posible, se realizó una amplificación con el oligonucleótido GAA\_2F2 (**Tabla 1**) y el utilizado para explorar el exón 2 (5'GAAGGGCTCCTCGGAGAA 3'), que produciría una amplificación más específica de la banda normal (amplicón 1) y del amplicón 2, dejando al margen las variante SV2 y SV3. La secuenciación posterior con los mismos cebadores permitió confirmar el procesamiento aberrante generado por el 5' GT anormal producto de la mutación. Como se puede ver en la representación detallada (**Figura 25F**), hay un *splicing* aberrante con el AG localizado en el c.485-486. Esto implica la pérdida del fragmento entre los nucleótidos c.342 y c.486, que corresponde con una pérdida de 145pb y la introducción de un codón de parada cuatro aminoácidos después del punto de ruptura. En el análisis cuantitativo (**Figura 25G**), se puede observar una disminución desigual de la expresión de los distintos exones comparado con la expresión obtenida en el control normal.



**Figura 25. Análisis del procesamiento de un control normal y del paciente P318.**

**A:** Secuencia genómica del exón 2, donde está recuadrada la variante analizada. **B:** Gel de agarosa del análisis de mRNA con oligonucleótidos localizados en los exones 1 y 3, realizado tanto en la muestra del paciente como en una muestra control. El amplicón 1, SV3 y SV2 corresponden con el procesamiento normal del exón 2 y el amplicón 2 corresponde con el procesamiento aberrante que se produce en el paciente. **C:** Representación de la predicción del papel del c.341 en el procesamiento del exón 2 que las herramientas bioinformáticas le otorgan, si la secuencia es la normal descrita para el gen (control) o si existe la mutación c.343C>T descrita en el paciente. **D:** Representación del procesamiento normal del exón 2. **E:** Representación de los efectos de la variante sobre el procesamiento normal del exón 2. Los exones están representados por las cajas azules y el número de cada exón está en el interior de cada caja. Las líneas azules representan los intrones, las líneas verdes corresponden al procesamiento normal y las líneas rojas marcan el procesamiento aberrante. La localización de los oligonucleótidos utilizados para realizar la amplificación se representan mediante (↔).

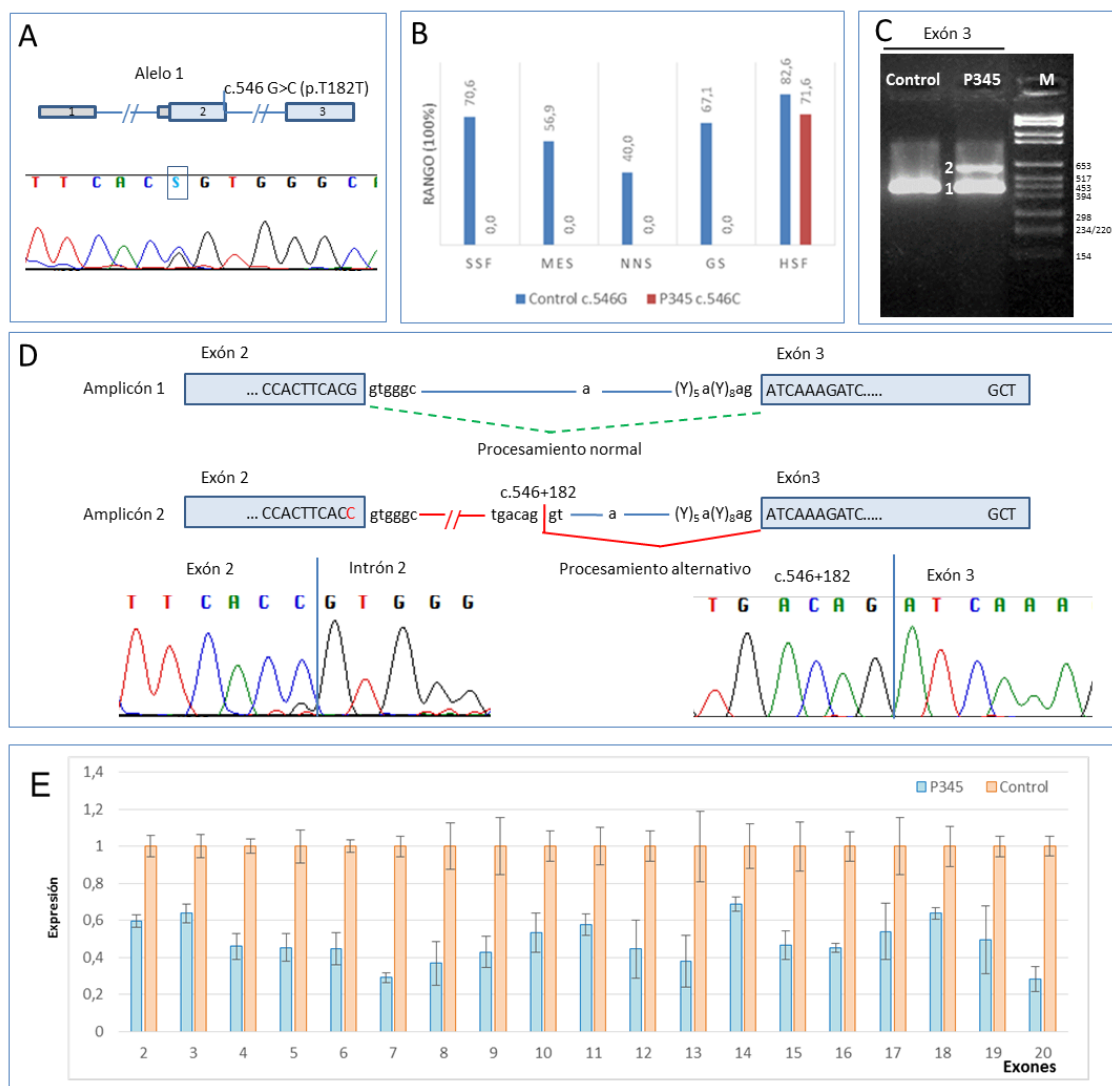


**Figura 25 (continuación): Análisis del procesamiento de un control normal y del paciente P318.** F: Representación detallada del procesamiento observado en la secuenciación del fragmento amplificado. En las cajas azules se muestra la secuencia normal, marcando en rojo la variante. En negrita y subrayado se marca el codón de parada prematuro. En la parte derecha, se presenta la secuencia obtenida donde se aprecia el resultado del procesamiento aberrante. G: Representación del análisis de la cuantificación del mRNA mediante qPCR. Se usó la *B2M* para normalizar. Los valores del control se llevaron al 100%. Y las barras de error indican la desviación estándar (n=3).

#### 4.1.6.6. Paciente P345.

En el paciente P345 se identificó únicamente la variante sinónima c.546G>C (p.T182T) (**Figura 26A**), que afecta al último nucleótido del exón 2. Este nucleótido forma parte de la secuencia de reconocimiento de la ribonucleoproteína U1, dentro del procesamiento del extremo 5' del intrón 2. Esta variante ha sido clasificada en el Pompe Center como variante de significado desconocido y en *ClinVar* como patogénica. En el genotipo de este paciente no se observó ningún cambio de interés en el otro alelo del gen *GAA*. La mayoría de las herramientas bioinformáticas predecían la pérdida del 5' donador del intrón 2 (**Figura 26B**). En el análisis cualitativo se detectó en el gel de agarosa una banda que estaba presente tanto en el paciente como en la muestra control (amplicón 1), además se identificó solo en el paciente una banda adicional de mayor tamaño (amplicón 2), que sugería un posible procesamiento aberrante (**Figura 26C**). La secuenciación de ese amplicón 2 puso de manifiesto que se utilizaba un sitio críptico alternativo, situado 182 bases dentro del propio intrón 2 (**Figura 26D**). Esto provocaría la incorporación de 61 aminoácidos adicionales, alterando la pauta de lectura y generando un codón de parada prematuro en el exón 6. Con el análisis cuantitativo (**Figura 26E**), se pudo detectar una expresión disminuida en todos los exones analizados, con respecto a la expresión detectada en el control.

## RESULTADOS



**Figura 26. Análisis del procesamiento de un control normal y del paciente P345.**

**A:** Secuencia genómica del exón 2 donde está recuadrada la variante analizada. **B:** Resultado del análisis obtenido con la aplicación *Alamut*. Puntuación con la secuencia normal (azul). Puntuación otorgada con la secuencia mutada (rojo). **C:** Gel de agarosa del análisis de mRNA con oligonucleótidos localizados en los exones 1 y 3, realizado tanto en la muestra del paciente P345 como en una muestra control. El amplicón 1 corresponde con el procesamiento normal y el amplicón 2 corresponde con el procesamiento aberrante que se realiza en el paciente. **D:** Representación detallada del procesamiento de los amplicones 1 y 2 donde se muestra el procesamiento normal y el alternativo. En las cajas azules se muestra la secuencia normal marcando en rojo la variante, (Y)<sub>n</sub> se usa para identificar un grupo de pirimidinas y su número. En la parte inferior está la secuencia obtenida donde se puede también apreciar el resultado del procesamiento aberrante. En la parte izquierda se aprecia el mantenimiento del intrón 2 y en la secuencia de la parte derecha se aprecia el punto del procesamiento intrón 2-exón 3. **E:** Representación del análisis de la cuantificación del mRNA mediante qPCR. Se usó la *B2M* para normalizar. Los valores del control se llevaron al 100%. Y las barras de error indican la desviación estándar (n=3).

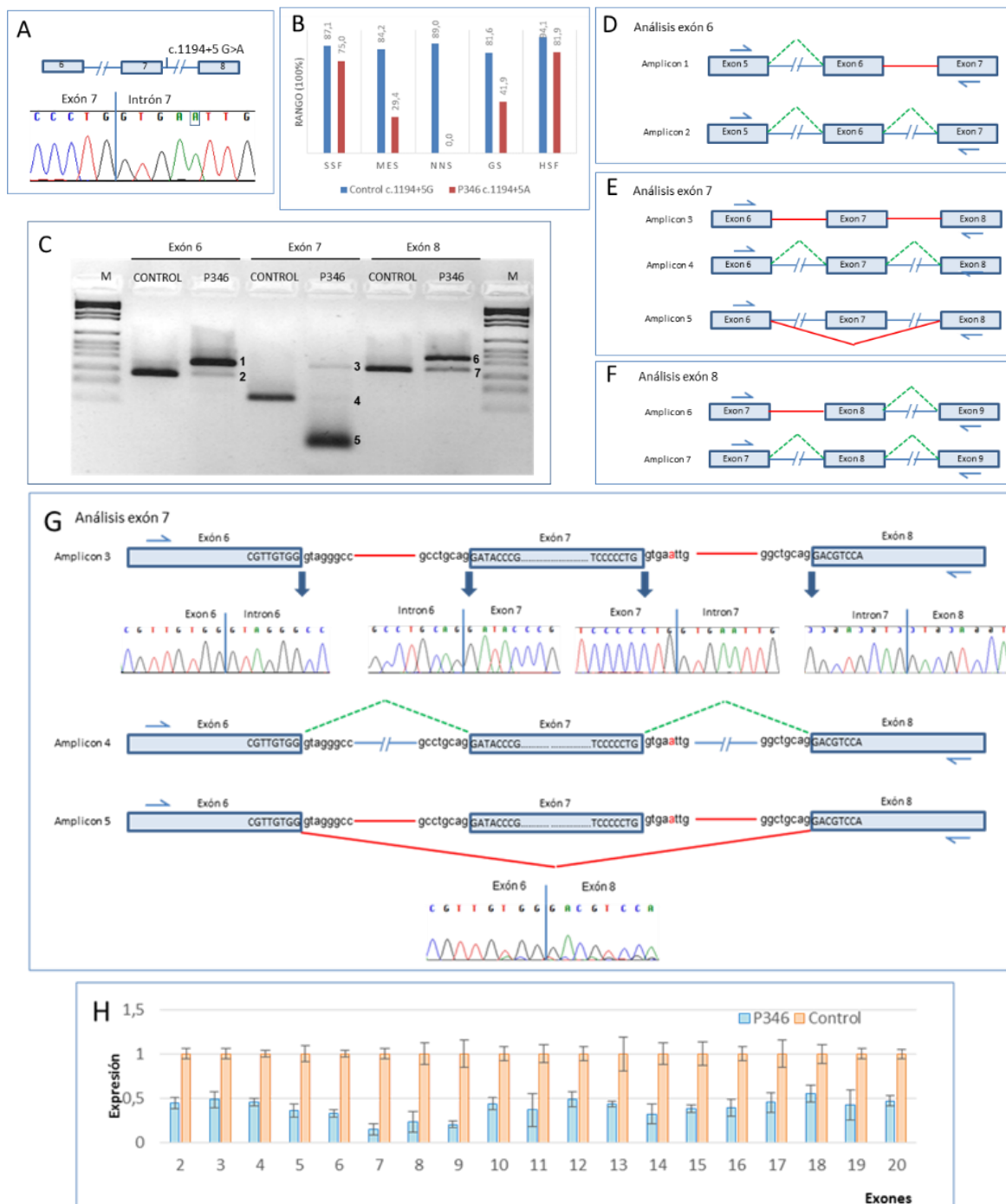
### 4.1.6.7. Paciente P346.

El paciente P346 había sido diagnosticado de enfermedad de Pompe mediante biopsia muscular. El estudio genético permitió identificar en homocigosidad la variante c.1194+5G>A, localizada en el intrón 7 (**Figura 27A**), que había sido descrita previamente (Kroos et al., 2012). En ese artículo, se hizo un análisis “*in silico*” de la variante, utilizando las herramientas bioinformáticas de la plataforma *Alamut® Visual*, y



fue clasificada como no patogénica. En este trabajo de tesis se realizó un nuevo análisis utilizando nuevas versiones de las mismas herramientas bioinformáticas (versión V2.7.2) y se pudo determinar que todas daban una pérdida del sitio donador, siendo la herramienta *NNS* la que mostró una mayor sensibilidad (**Figura 27B**). Se hizo el análisis funcional cualitativo de todo el gen. Tras las amplificaciones, se visualizaron los productos de PCR en un gel de agarosa, pudiéndose observar bandas de tamaño anormal en los análisis de los exones 6, 7 y 8 (bandas 1, 3, 5 y 6), con respecto a las obtenidas como consecuencia del procesamiento normal (bandas 2, 4 y 7) (**Figura 27C**). Todas ellas fueron posteriormente secuenciadas. El análisis del exón 6 mostró 2 bandas, una más intensa (banda 1), que correspondería con la permanencia completa del intrón 6, y una banda de menor tamaño y con menor intensidad (banda 2), que correspondería al procesamiento normal (**Figura 27C-27D**). El estudio del exón 7 mostró tres productos distintos. La banda 3, que correspondería con la permanencia completa de los intrones 6 y 7, la banda 4 que correspondería con el procesamiento normal, y la banda 5, que es la de mayor intensidad, que correspondería a la pérdida del exón 7, haciéndose por lo tanto un procesamiento exón 6-8 (**Figura 27C-27E**). El estudio del exón 8, también reflejó la aparición de 2 bandas de distinto tamaño. La banda 6 que correspondería a la permanencia del intrón 7, y la banda 7 que correspondería al procesamiento normal (**Figura 27C-27F**). Los productos observados en el análisis del exón 7 (**Figura 27E**) reflejarían el resultado que la variante c.1194+5G>A produce sobre el procesamiento, ya que los análisis del exón 6 y 8 se realizaron con oligonucleótidos que anillaban en el exón 7, por lo que solo se estarían amplificando aquellos amplicones en los que el exón 7 está presente (amplicón 3 y 4). En la **Figura 27G**, se puede ver la representación detallada de cada uno de los amplicones obtenidos en el análisis del exón 7, con la secuenciación obtenida de cada una de las zonas de procesamiento. El procesamiento mayoritario, en el que se elimina el exón 7 por completo, junto a los intrones 6 y 7, lo que provoca es un cambio de fase de lectura que implica la introducción de un codón de parada precoz en el exón 10. En el análisis cuantitativo con qPCR (**Figura 27H**), se pudo observar un descenso de la expresión de todos los exones con respecto al control, siendo más acusado en los exones 6, 8 y 9 que estaba en torno al 20-40%, y el exón 7 que mostraba una reducción adicional con una expresión aproximada del 15% con respecto al control normal.

## RESULTADOS

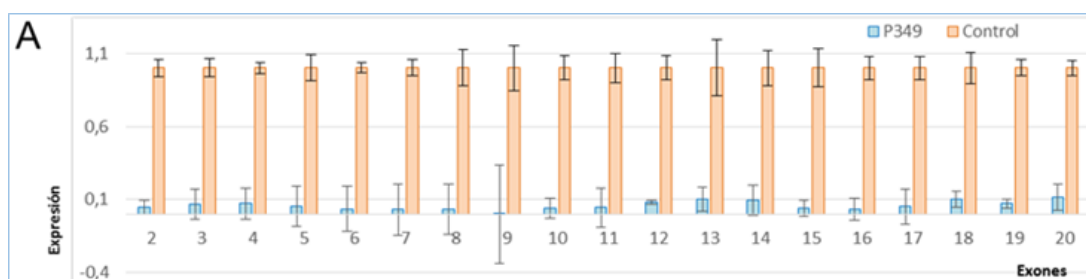


**Figura 27. Análisis del procesamiento de un control normal y del paciente P346.**

**A:** Secuencia genómica del exón 7 donde está recuadrada la variante analizada. **B:** Resultado del análisis obtenido con la aplicación *AluMut*. Puntuación con la secuencia normal (azul). Puntuación otorgada con la secuencia mutada (rojo). **C:** Gel de agarosa del análisis de mRNA con oligonucleótidos localizados en los exones 5, 6, 7, 8 y 9 como se muestra en las figuras D-E-F, realizado tanto en la muestra del paciente P346 como en una muestra control. Los amplicones 2, 4 y 7 corresponderían con el procesamiento normal y los amplicones 1, 3, 5 y 6 corresponderían con el procesamiento aberrante que se realiza en el paciente. **D-E-F:** Representación de los efectos de la variante sobre el procesamiento. Los exones están representados por las cajas azules y el número de cada exón está en el interior de cada caja. Las líneas azules representan los intrones, las líneas verdes corresponden al procesamiento normal y las líneas rojas marcan el procesamiento aberrante. La localización de los oligonucleótidos utilizados para realizar la amplificación se representan mediante (↔). **G:** Representación detallada del procesamiento observado en el análisis del exón 7. El amplicón 4 mostraría el procesamiento normal y los amplicones 3 y 5 mostrarían el procesamiento alternativo. En las cajas azules se muestra la secuencia normal marcando en rojo la variante. En la parte inferior, está la secuencia obtenida donde se puede también apreciar el resultado del procesamiento aberrante. **H:** Representación del análisis de la cuantificación del mRNA mediante qPCR. Se usó la *B2M* para normalizar. Los valores del control se llevaron al 100%. Y las barras de error indican la desviación estándar (n=3).

#### 4.1.6.8. Resto pacientes.

En el paciente P349 se identificó en homocigosidad la variante c.1411G>T (p.E471\*), localizada en el exón 9, no descrita previamente en las bases de datos consultadas. En el análisis cualitativo no se observaron bandas adicionales a las obtenidas en la muestra control (datos no mostrados). Sin embargo, se pudo detectar un descenso muy intenso en el nivel de expresión del gen, en comparación con el control en el análisis de qPCR (**Figura 28**). En este caso, se consideró que el descenso podría ser ocasionado por una activación del fenómeno de *Nonsense Mediated decay (NMD)*, mecanismo que conlleva una degradación del mRNA mediado por la existencia de codones de parada precoces.



**Figura 28. Análisis del procesamiento de un control normal y del paciente P349.**

**A:** Representación del análisis de la cuantificación del mRNA mediante qPCR. Se usó la *B2M* para normalizar. Los valores del control se llevaron al 100%. Y las barras de error indican la desviación estándar (n=3).

En el resto de las muestras analizadas no se observaron resultados anormales en ninguno de los estudios cualitativos o cuantitativos realizados.

#### 4.1.7. Cómputo global de resultados de los análisis moleculares

En este trabajo de tesis se llevaron a cabo 259 estudios genéticos de pacientes con sospecha de enfermedad de Pompe, lo que permitió confirmar el diagnóstico genético en 76 de ellos. De ese grupo de pacientes, 12 tenían la forma infantil de la enfermedad y mostraban genotipos compuestos por variantes genéticas con efecto muy deletéreo sobre la funcionalidad del gen o la proteína. Los 64 restantes presentaban una forma tardía, y en 54 de ellos se encontró la variante c.-32-13T>G formando parte del genotipo enfermo, lo que supone el 84,3%.

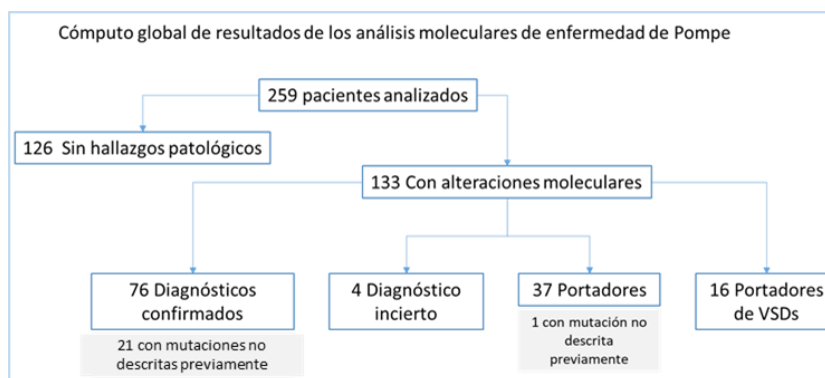
En 21 de los 76 pacientes con genotipo diagnóstico de la enfermedad se encontraron variantes no descritas previamente, entre ellas, la delección más grande identificada hasta el momento en esta patología.

En 4 pacientes se hallaron variantes no descritas o descritas como VSD, en homocigosidad o doble heterocigosidad, lo que no permitió alcanzar un diagnóstico de certeza.

## RESULTADOS

Como resultado de este estudio se diagnosticaron además 37 pacientes portadores de la enfermedad, de los que 1 era portador de una variante patogénica no descrita previamente. Además, se encontraron 16 pacientes con un genotipo en el que solo se identificó una VSD en un alelo, mostrando una secuencia normal en el otro.

Por último, en 126 pacientes no se detectó ningún cambio o reordenamiento que pueda considerarse patológico. (Figura 29).



**Figura 29. Cómputo global de los resultados obtenidos.**

Resumen de los resultados obtenidos en este trabajo de tesis con el número de pacientes en los que se han detectado alteraciones moleculares y en los que no.

### 4.2. Análisis de transcriptoma mediante RNA-Seq.

Para llevar a cabo el análisis de transcriptoma se utilizaron 13 muestras de pacientes con la forma tardía de la enfermedad y genotipo enfermo confirmado. Como población control se analizaron 17 muestras de individuos sanos de edades similares a la de los pacientes. En todos estos pacientes se confirmó un genotipo normal en *GAA* (Tabla 18). Para identificar las muestras (ID. Paciente), se utilizó el siguiente código: P (código de identificación de paciente de Pompe) + Número (corresponde con el orden de entrada a la unidad) + C (control) o P (paciente).

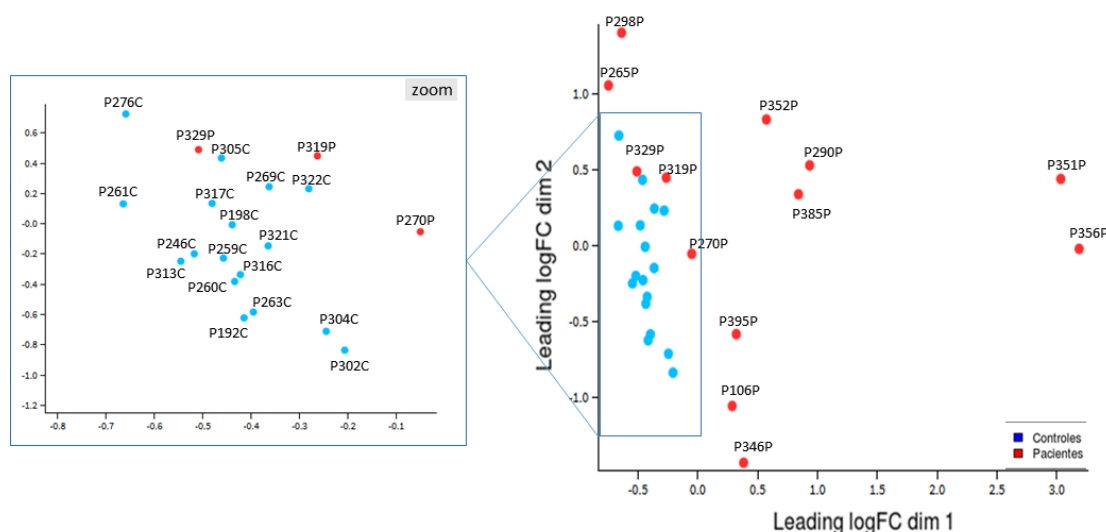
Número	ID. Grupo Paciente	Alelo 1	Alelo 2	Número	ID. Grupo Control
1	P106P	I1 c.-32-13T>G	Ex10 c.1447G>T (p.G483W)	1	P192C
2	P265P	I1 c.-32-13T>G	Ex12 c.1724A>G (p.Y575C)	2	P198C
3	P270P	I1 c.-32-13T>G	Ex10 c.1447G>T (p.G483W)	3	P246C
4	P290P	I1 c.-32-13T>G	I6 c.1075+1G>A	4	P259C
5	P298P	I1 c.-32-13T>G	Ex14 c.1933G>T (p.D645Y)	5	P260C
6	P319P	I1 c.-32-13T>G	Ex12 c.1655T>C (p.L552P)	6	P261C
7	P329P	I1 c.-32-13T>G	Ex13 c.1802C>T (p.S601L)	7	P263C
8	P346P	I7 c.1194+5G>A	I7 c.1194+5G>A	8	P269C
9	P351P	I6 c.1076-1G>C	Ex15 c.2065G>A (p.E689K)	9	P276C
10	P352P	I1 c.-32-13T>G	Ex15 c.2141A>C (p.H714P)	10	P302C
11	P356P	I1 c.-32-13T>G	I1 c.-32-13T>G	11	P304C
12	P385P	I1 c.-32-13T>G	c.2481+102_2646+31del	12	P305C
13	P395P	I1 c.-32-13T>G	Ex16 c.2237G>A (p.W746*)	13	P313C
				14	P316C
				15	P317C
				16	P321C
				17	P322C

**Tabla 19. Muestras analizadas mediante RNA-seq.**

Identificación de las muestras de pacientes incluidos en el estudio con su genotipo asociado, así como la de los controles.

4.2.1. Análisis de expresión diferencial.

El análisis de transcriptoma nos permitió cuantificar más de 65000 transcritos diferentes. Tras el análisis que implica el filtrado y eliminación de los transcritos no expresados en ningún grupo, quedaron un total de 20401, que cumplían los requisitos de estar expresados en al menos 8 controles o en 6 pacientes. Una vez normalizados mediante el método *TMM* se realizó un análisis de escalamiento multidimensional (**Figura 30**), en el que se pudo observar una estratificación de las muestras en base a su grupo de estudio, mostrando mayoritariamente diferencias claras entre pacientes y controles.

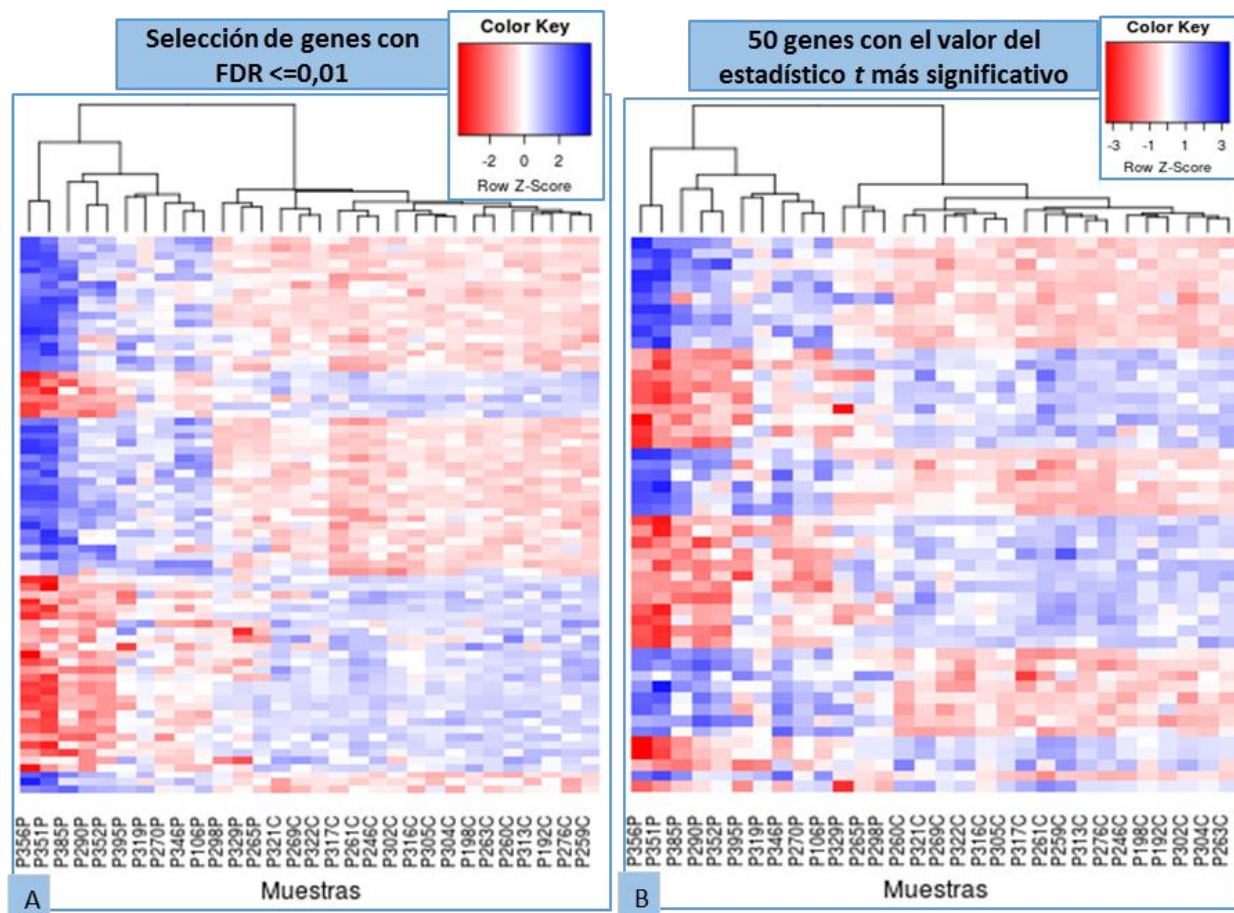


**Figura 30. MDS plot de los datos normalizados.**

Relación espacial entre las muestras incluidas en el estudio. En rojo se muestran los pacientes y en azul los controles. Todas las muestras aparecen identificadas con su ID de paciente.

Después del análisis de expresión diferencial llevado a cabo con *Limma* se pudo discriminar entre genes significativamente expresados de los que no lo estaban, estableciendo un  $FDR \leq 0,05$ . Con este criterio, se obtuvieron un total de 682 genes sobreexpresados (*up*) y 673 genes infraexpresados (*down*), en pacientes con respecto a los controles. Con el fin de transformar los nombres de los genes de la nomenclatura *Ensembl* a la de referencia (NCBI) se realizó un paso de anotación con las librerías *org.Hs.eg.db()* y *TxDb.Hsapiens.UCSC.hg38.knownGene()*. Esto provocó la pérdida de genes que no estaban bien caracterizados, suponiendo una reducción total del número de genes diferencialmente expresados (DE) de 1355 a 980, de los cuales 479 estaban *up*-regulados y 501 *down*-regulados. El modo en que se formula el cambio de nivel en la expresión génica es mediante el logaritmo en base 2 del cambio de proporción ( $\log_2$  *Fold Change* o  $\logFC$ ). Se estableció un punto de corte más restrictivo, con un  $FDR \leq 0,01$  y se obtuvo una lista de 260 genes que se encontraban en un rango de  $\logFC$  entre -1.87 y 3.71, de ellos, se seleccionaron aquellos con los  $\logFC$  más significativos, es decir,  $\logFC$  entre -1,87 y -1, y entre +1 y +3,71. A partir de estos criterios se obtuvieron 74 genes con los que se realizó una agrupación no

supervisada que representamos con un *heatmap* (Figura 31A). Además, se seleccionaron 25 genes *up*-regulados y 25 *down*-regulados, con un valor de *t* más significativo con los que se realizó un agrupamiento no supervisado (Figura 31B).



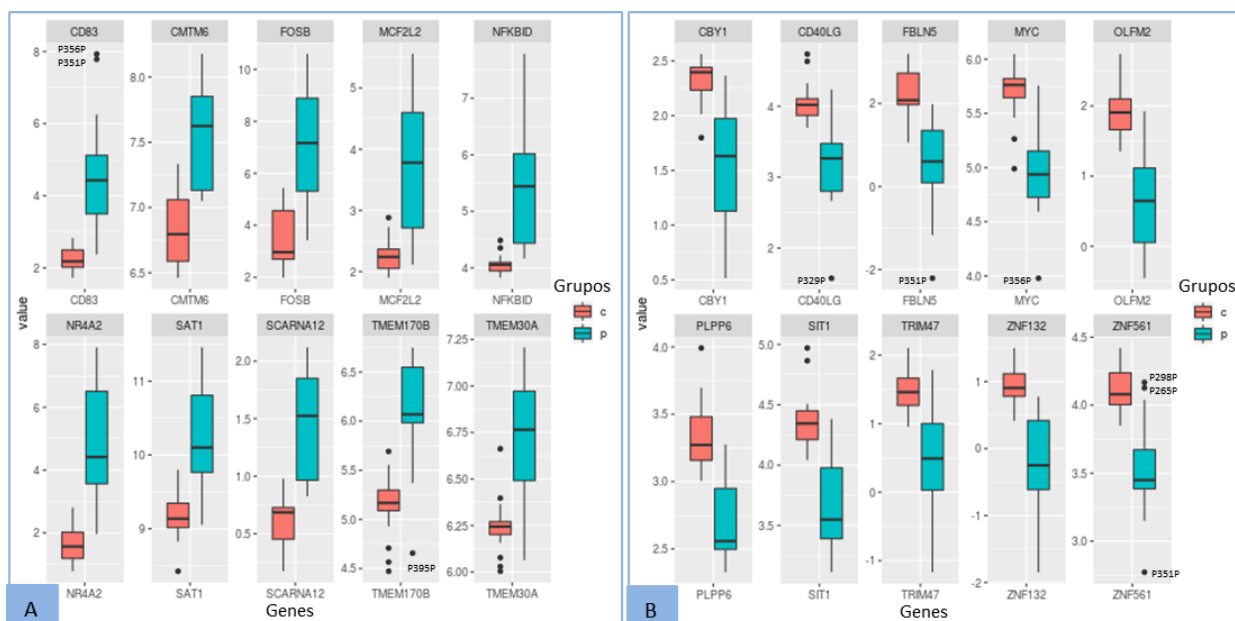
**Figura 31. Heatmap genes diferencialmente expresados.**

A: selección de los genes diferencialmente expresados con un  $FDR \leq 0,01$  B: 50 genes *up* y *down*-regulados con el valor del estadístico *t* más significativo.

Como se observa en la **Figura 31**, ambas agrupaciones no supervisadas mostraban resultados muy similares. En ambos casos, se encontraban clasificados todos los pacientes en un mismo grupo, excepto P298P, P329P y P265P, que tenían un perfil de expresión situado entre el grupo de pacientes y el grupo control. No obstante, destaca la capacidad de discriminación que ambos procedimientos tenían para separar pacientes de controles, como consecuencia de las diferencias obtenidas en los perfiles de expresión de ambas poblaciones.

Posteriormente, se seleccionaron los 10 genes *up*-regulados y los 10 genes *down*-regulados que mostraban los valores de *t* más significativos. Con ellos se realizaron varios *boxplots*, con el fin de mostrar gráficamente la expresión diferencial obtenida en los dos grupos estudiados (**Figura 32**).

## RESULTADOS



**Figura 32. Boxplot realizados con los genes con diferencias de expresión más significativas.**

**A:** Selección de los 10 genes con la *t* con valores más positivos **B:** Selección de los 10 genes con la *t* con valores más negativos. Grupo C (Control) y P (Paciente). Cada *boxplot* corresponde a un gen que está indicado en la parte inferior y superior de cada gráfico.

Todos ellos presentan un p-valor ajustado  $<0,01$ . Las diferencias en las medianas observadas por gen, comparando pacientes frente a controles, oscilan entre los valores 0,49 y -0,55 que corresponden a los genes *TMEM30A* y *ZNF561*, respectivamente, y los observados en los genes *FOSB* y *NR4A2*, que son de 3,71 y 3,10, respectivamente. Esto implica un amplio rango de valores en cuanto a diferencias de expresión. En todos los genes existe una mayor dispersión en los datos obtenidos en el grupo de los pacientes frente al grupo control, lo que vendría a corroborar los resultados previamente obtenidos tanto en el *MDS-plot* como en los *heatmaps*, y que supone el reflejo de la gran variabilidad fenotípica observada en este tipo de patología. En 13 de los 20 genes se observaron datos atípicos, que estarían fuera de los valores máximos y mínimos representados. En 4 de ellos, estos valores atípicos correspondían con los paciente P351P y P356P. En 3 de los gráficos, los valores atípicos correspondían a los pacientes P329P, P265P, P298P y P395P. En 9 de estas representaciones, los valores atípicos se observaron en los controles.

Se utilizó la información disponible en bases de datos como *Uniprot* y *String* con el fin de conocer la función de los 20 genes previamente seleccionados. Esta información aparece detallada en la **Tabla 19** y **Tabla 20**:

- Genes diferencialmente expresados positivamente (Tabla 19)

CD83	Proteína	CD83 antigen
	Descripción	Parece que desempeña un papel importante en la presentación del antígeno o en las interacciones celulares que siguen a la activación de los linfocitos
CMTM6	Proteína	CKLF-like MARVEL transmembrane domain-containing protein 6.
	Descripción	Proteína localizada en el endosoma, implicada en el reciclado y la expresión de PD-L1/CD274, una proteína transmembrana que parece que tiene un papel importante en la tolerancia inmune frente a lo propio y frente a la inmunidad antitumoral.
FOSB	Proteína	Protein fosB
	Descripción	FosB es un factor de transcripción que dimeriza con proteínas de la familia Jun para formar el complejo AP-1. Está implicada en la regulación de la proliferación celular, diferenciación y transformación.
MCF2L2	Proteína	Probable guanine nucleotide exchange factor MCF2L2
	Descripción	Probable función como factor intercambiador de guanina, relacionado con la familia de proteínas Rho.
NFKBID	Proteína	NF-kappa-B inhibitor delta
	Descripción	Regula la expresión de la Interleucina-2, Interleucina-6 y otras citocinas mediante la regulación de la actividad de NF-kappa-B. Tiene un papel importante en la regulación de la respuesta inflamatoria.
NR4A2	Proteína	Nuclear receptor subfamily 4 group A member 2
	Descripción	Es un regulador transcripcional implicado en la diferenciación y el mantenimiento de las neuronas dopaminérgicas durante el desarrollo.
SAT1	Proteína	Diamine acetyltransferase 1
	Descripción	Enzima implicada en catalizar la acetilación de poliaminas, mostrando una mayor especificidad por la espermidina, la espermina y la putrescina. Esta enzima está altamente regulada lo que permite un fino control de la concentración intracelular de poliaminas, que son moléculas con carga positiva como el DNA, lo que permite estabilizarlo.
SCARNA12	Proteína	No proteína
	Descripción	Es un RNA de pequeño tamaño (snoRNA) que se localiza en el núcleo celular y está implicado en las modificaciones post-transcripcionales de otros RNAs.
TMEM170B	Proteína	Transmembrane protein 170B
	Descripción	Regula negativamente la cascada de señalización Wnt en células de cáncer de mama inhibiendo su crecimiento
TMEM30A	Proteína	Transmembrane protein 30A
	Descripción	Componente accesorio del complejo flipasa P4-ATPasa, que cataliza la hidrólisis de ATP acoplado al transporte de aminofosfolípidos desde la cara externa a la interna de varias membranas celulares lo que permite mantener una distribución asimétrica de fosfolípidos.

**Tabla 20. Descripción de la función de los genes más significativamente expresados positivamente.**

Se muestra una breve descripción de los 10 genes más significativamente expresados positivamente en función del estadístico *t*. Información obtenida de la base de datos *Uniprot* y *String*.

- Genes diferencialmente expresados negativamente (Tabla 20)

CBY1	Proteína	Protein chibby homolog 1
	Descripción	Inhibe la vía Wnt uniéndose a CTNNB1 / $\beta$ -catenina e inhibiendo la activación transcripcional mediada por $\beta$ -catenina a través de la competencia con factores de transcripción TCF / LEF. La formación del complejo $\beta$ -catenina-TCF/LEF regula positivamente la expresión de genes que codifican proteínas como ciclina-D1, c-Myc y metaloproteasas de la matriz celular
CD40LG	Proteína	CD40 ligand
	Descripción	Citocina que se une a CD40 / TNFRSF5. Co-estimula la proliferación de células T y la producción de citocinas. Mejora la producción de IL-4 e IL-10. Induce la activación de NF-kappa-B y quinasas MAPK8 y PAK2 en células T.
FBLN5	Proteína	Fibulin-5
	Descripción	Esencial para la formación de fibras elásticas, participa en el ensamblaje del polímero de elastina continua (ELN) y promueve la interacción de las microfibrillas y el ELN. Estabiliza y organiza fibras elásticas en la piel, pulmón y en el sistema vascular.
MYC	Proteína	Myc proto-oncogene protein
	Descripción	Factor de transcripción que se une al ADN de una manera no específica, pero también reconoce específicamente la secuencia central 5'-CAC[GA]TG-3'. Activa la transcripción de genes relacionados con el crecimiento. Se une al promotor VEGFA, promoviendo la producción de VEGFA y con ello la angiogénesis.

**Tabla 21. Descripción de la función de los genes más significativamente expresados negativamente.**

Se muestra una breve descripción de los 10 genes más significativamente expresados negativamente en función del estadístico *t*. Información obtenida de la base de datos *Uniprot* y *String*.



## RESULTADOS

OLFM2	Proteína	Noelin-2
	Descripción	Juega un papel en la diferenciación mediada por TGF-beta de las células del músculo liso desde células madre mesenquimales. Participa en la transformación del factor de crecimiento TGF-beta, inducida por la diferenciación del músculo liso. TGF-beta induce la expresión y la translocación de OLFM2 al núcleo donde se une a SRF, provocando su disociación del represor transcripcional HEY2 / HERP1 y facilitando la unión de SRF a los genes diana. Es un regulador del cambio fenotípico de las células vasculares del músculo liso (SMC), que actúa promoviendo RUNX2 e inhibiendo la unión de MYOCD a SRF. El cambio fenotípico de SMC es el proceso a través del cual las SMC vasculares experimentan una transición entre un fenotipo contráctil inactivo y un fenotipo sintético proliferativo en respuesta a estímulos patológicos.
PLPP6	Proteína	Phospholipid phosphatase 6
	Descripción	Fosfatasa que desfosforila el difosfato de preescualeno (PSDP) en monofosfato de preescualeno (PSMP), lo que sugiere que puede estar implicado indirectamente en la inmunidad innata.
SIT1	Proteína	Signaling threshold-regulating transmembrane adapter 1
	Descripción	Regula negativamente la señalización mediada por receptor de antígeno de células T (TCR). Participa en la selección positiva de células T
TRIM47	Proteína	E3 ubiquitin-protein ligase TRIM47
	Descripción	Implicada en la ubiquitinación y la degradación vía proteosoma de CYLD, que a su vez, juega un papel importante en la regulación de la supervivencia, proliferación y diferenciación vía NF-kappa-B
ZNF132	Proteína	Zinc finger protein 132
	Descripción	Factor de transcripción implicado en la regulación transcripcional.
ZNF561	Proteína	Zinc finger protein 561
	Descripción	Factor de transcripción implicado en la regulación transcripcional.

**Tabla 22 (continuación). Descripción de la función de los genes más significativamente expresados negativamente.**

Se muestra una breve descripción de los 10 genes más significativamente expresados negativamente en función del estadístico *t*. Información obtenida de la base de datos *Uniprot* y *String*.

### 4.2.2. Identificación de rutas transcripcionalmente alteradas

#### 4.2.2.1. Análisis de enriquecimiento funcional mediante la herramienta GSEA

Con los genes obtenidos después de la normalización y ordenados en función del estadístico *t*, se llevó a cabo el análisis de enriquecimiento funcional mediante la herramienta *GSEA*. Para ello se utilizaron las agrupaciones de genes: *Hallmark*, *KEGG*, *Biocarta* y Factores de transcripción, cuyas características han sido descritas previamente.

La **Tabla 21** muestra los conjuntos de genes de todas las agrupaciones estudiadas que aparecen enriquecidos positivamente, es decir que se encuentran elevados en el fenotipo enfermo frente al sano, y en la **Tabla 22** se muestran los conjuntos enriquecidos negativamente, es decir, disminuidos en el fenotipo enfermo frente al sano.

Collecciones	Conjunto genes enriquecidos positivamente	FDR <0,25	FDR <0,05	FDR <0,01
Hallmark	38 / 49	29	25	14
Kegg	97 / 157	57	25	6
Biocarta	94 / 117	65	2	0
Factores de transcripción	520 / 576	411	251	112

**Tabla 23. Resumen del resultado de enriquecimiento funcional positivo obtenido mediante GSEA.**

Conjunto de rutas enriquecidas positivamente (+) de cada una de las agrupaciones estudiadas. Se muestra el número de las agrupaciones diferencialmente enriquecidas con un *FDR* <0,25, < 0,05 y <0,01.

## RESULTADOS

Collecciones	Conjunto genes enriquecidos negativamente	FDR <0,25	FDR <0,05	FDR <0,01
Hallmark	11 / 49	1	1	1
Kegg	60 / 157	26	13	6
Biocarta	23 / 117	3	1	1
Factores de transcripción	56 / 576	3	0	0

**Tabla 24. Resumen del resultado de enriquecimiento funcional negativo obtenido mediante GSEA.**

Conjunto de rutas enriquecidos negativamente (-) de cada una de las agrupaciones estudiadas. Se muestra el número de las agrupaciones diferencialmente enriquecidas con un  $FDR < 0,25$ ,  $< 0,05$  y  $< 0,01$ .

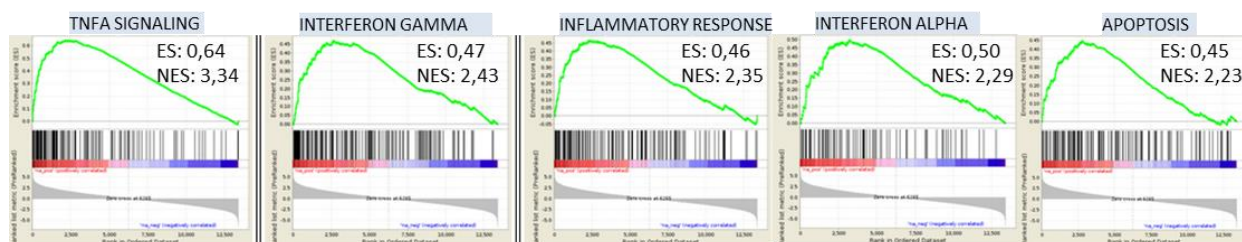
Con el análisis de las rutas enriquecidas funcionalmente mediante la herramienta *GSEA* y la agrupación de genes *Hallmark* se obtuvieron 29 enriquecidas positivamente (+) con un  $FDR < 0,25$ , de las cuales 25 tenían un  $FDR < 0,05$  y 14 tenían un  $FDR < 0,01$ . Solo una ruta se mostró enriquecida negativamente (-) con un  $FDR < 0,25$  y mostraba un  $FDR < 0,01$ .

Las rutas con  $FDR < 0,01$  se muestran en la **Tabla 23**. Dentro de las obtenidas, cabe destacar que de las 5 primeras con los valores de enriquecimiento normalizado (*NES*) más altos, 4 estarían implicadas en la respuesta inmune y la quinta, correspondería a una ruta implicada en apoptosis. En la **Figura 33** se muestran los gráficos del valor de enriquecimiento (*ES*) obtenidos en estas 5 rutas.

RUTAS_HALLMARK	TAMAÑO RUTA	GENES ENRIQUECIDOS	ES	NES	FDR q-val
TNFA_SIGNALING_VIA_NFKB	165	99	0,64	3,34	0,000
INTERFERON_GAMMA_RESPONSE	175	70	0,47	2,43	0,000
INFLAMMATORY_RESPONSE	143	57	0,46	2,35	0,000
INTERFERON_ALPHA_RESPONSE	84	43	0,50	2,29	0,000
APOPTOSIS	137	59	0,45	2,23	0,000
UV_RESPONSE_UP	120	44	0,41	2,01	0,001
HYPOXIA	148	68	0,38	1,96	0,001
CHOLESTEROL_HOMEOSTASIS	64	33	0,44	1,92	0,001
TGF_BETA_SIGNALING	46	26	0,47	1,91	0,001
P53_PATHWAY	171	62	0,34	1,79	0,002
PI3K_AKT_MTOR_SIGNALING	89	39	0,38	1,78	0,002
G2M_CHECKPOINT	178	77	0,33	1,71	0,004
COMPLEMENT	155	59	0,34	1,67	0,007
MTORC1_SIGNALING	191	49	0,31	1,66	0,008
MYC_TARGETS_V2	58	28	-0,46	-1,90	0,001

**Tabla 25. Rutas enriquecidas positiva y negativamente con un  $FDR < 0,01$ , obtenidas con GSEA con la agrupación Hallmark.**

En azul se marcan las rutas enriquecidas positivamente y en rojo se marca la ruta enriquecida negativamente. En la tabla se muestra el número total de genes de cada ruta y el número de genes que están enriquecidos, el valor de enriquecimiento sin normalizar (*ES*) y el normalizado (*NES*), así como el *FDR* obtenido.



**Figura 33. Grafica del valor de enriquecimiento de las rutas con los valores de *NES* más altos (Hallmark).**

Representación gráfica del valor de enriquecimiento (*ES*) obtenido en las 5 rutas con mayor *ES* normalizado (*NES*).

## RESULTADOS

Con la agrupación de genes de rutas *KEGG*, se obtuvieron 57 rutas significativamente enriquecidas positivamente con un  $FDR < 0,25$ , de las 25 tenían un  $FDR < 0,05$  y 6 tenían un  $FDR < 0,01$ . En cuanto a las enriquecidas negativamente, se obtuvieron 26 con un  $FDR < 0,25$ , de las cuales 13 mostraban un  $FDR < 0,05$  y 6 tenían un  $FDR < 0,01$ . Las 6 rutas con un  $FDR$  más significativo, enriquecidas tanto positiva como negativamente se muestran en la **Tabla 24**:

RUTAS_KEGG	TAMAÑO RUTA	GENES ENRIQUECIDOS	ES	NES	FDR q-val
PROTEASOME	38	23	0,59	2,28	0,000
NEUROTROPHIN_SIGNALING_PATHWAY	105	56	0,45	2,16	0,001
PATHOGENIC_ESCHERICHIA_COLI_INFECTION	49	21	0,51	2,09	0,003
B_CELL_RECEPTOR_SIGNALING_PATHWAY	67	36	0,48	2,09	0,003
GAP_JUNCTION	61	36	0,48	2,05	0,003
LONG_TERM_POTENTIATION	48	24	0,47	1,95	0,008
VALINE_LEUCINE_AND_Isoleucine_DEGRADATION	40	22	-0,55	-2,11	0,001
BUTANOATE_METABOLISM	23	13	-0,62	-2,04	0,002
BETA_ALANINE_METABOLISM	17	13	-0,66	-2,03	0,002
AMINOACYL_TRNA_BIOSYNTHESIS	38	19	-0,53	-2,01	0,002
PEROXISOME	66	24	-0,46	-1,95	0,004
MISMATCH_REPAIR	20	15	-0,59	-1,88	0,006

**Tabla 26. Rutas enriquecidas positiva y negativamente con  $FDR < 0,01$ , obtenidas con GSEA con las agrupaciones de KEGG.** En azul se marcan las rutas enriquecidas positivamente y en rojo se marcan las rutas enriquecidas negativamente. En la tabla se muestra el número total de genes de cada ruta y el número de genes que están enriquecidos. Además se muestra el valor de enriquecimiento sin normalizar (*ES*) y normalizado (*NES*), así como el *FDR* obtenido.

Estas rutas están relacionadas con la diferenciación y supervivencia celular, respuesta inmune adaptativa como reacción a patógenos, así como rutas relacionadas con el metabolismo de aminoácidos, lípidos, etc.

Con la agrupación de genes de rutas *Biocarta*, se obtuvieron 65 significativamente enriquecidas positivamente con un  $FDR < 0,25$ , solo 2 rutas tenían un  $FDR < 0,05$  y ninguna con un  $FDR < 0,01$ . En cuanto a las enriquecidas negativamente, se obtuvieron 3 con un  $FDR < 0,25$ , de las cuales 1 tenía un  $FDR < 0,01$  (**Tabla 25**)

RUTAS_BIOCARTA	TAMAÑO RUTA	GENES ENRIQUECIDOS	ES	NES	FDR q-val
PROTEASOME_PATHWAY	27	15	0.58	2.07	0.028
EIF4_PATHWAY	23	11	0.56	1.90	0.049
NO2IL12_PATHWAY	15	7	-0.72	-2.13	0,001

**Tabla 27. Rutas enriquecidas positiva y negativamente con un  $FDR < 0,05$ , obtenidas con GSEA con la agrupación Biocarta.** En azul se marcan las rutas enriquecidas positivamente y en rojo se marca la ruta enriquecida negativamente. En la tabla se muestra el número total de genes de cada ruta y el número de genes que están enriquecidos. Además se muestra el valor de enriquecimiento sin normalizar (*ES*) y normalizado (*NES*), así como el *FDR* obtenido

## RESULTADOS

Por último se exploró la agrupación de factores de transcripción (*TFT*). Se obtuvieron 411 rutas significativamente enriquecidas positivamente con un *FDR* <0,25, de las que 251 tenían un *FDR*<0,05 y 112 mostraban un *FDR*<0,01. En cuanto a las rutas enriquecidas negativamente, se obtuvieron 3 con un *FDR* <0,25 y todas mostraban un *FDR* >0,05. En la **Tabla 26** se muestran las primeras 10 rutas enriquecidas positivamente.

RUTAS_Factores de transcripción	TAMAÑO RUTA	GENES ENRIQUECIDOS	ES	NES	FDR q-val
ATF_01	195	86	0,44	2,29	0,0000
CREB_02	205	76	0,39	2,09	0,0009
E4F1_Q6	230	106	0,39	2,12	0,0011
ATF3_Q6	162	73	0,40	2,07	0,0011
CREB_Q2_01	173	66	0,40	2,07	0,0012
CEBP_C	128	50	0,42	2,08	0,0012
TAXCREB_01	99	37	0,42	2,02	0,0012
FOXO1_02	171	73	0,40	2,06	0,0013
ATF1_Q6	151	59	0,42	2,13	0,0013
TFIIA_Q6	163	61	0,39	2,03	0,0013

**Tabla 28. Primeras 10 rutas enriquecidas positivamente, obtenidas con GSEA con las agrupaciones de TFT.**

En la tabla se muestra las primeras 10 rutas enriquecidas positivamente, especificando el número total de genes de cada ruta y el número de genes que están enriquecidos. Además se muestra el valor de enriquecimiento sin normalizar (*ES*) y normalizado (*NES*), así como el *FDR* obtenido

En estas 10 primeras rutas, aparecen muy representados genes como *FOSB* y *JUN*. Estos están implicados en el aumento de la expresión génica como parte de una respuesta temprana a estímulos de estrés. Juegan un papel clave en un gran número de procesos que implican proliferación, supervivencia y diferenciación, entre otros.

Asimismo, están representados en muchas de estas rutas genes como *TRIB1* y *ZFAND5*, que estarían implicados en la degradación de proteínas. *TRIB1* llevaría a cabo esta función por su interacción con la proteína ligasa de ubiquitina COP1. *ZFAND5*, desempeña un papel en la degradación de proteínas durante la atrofia muscular, regula la activación de *NF-kappa-B* y está relacionado con la activación de la apoptosis. Además, aparecen muy representados otros genes que están descritos en la **Tabla 19**, como *NR4A2*, *CD83*, *SAT1* entre otros.

### 4.2.2.2. Análisis de enriquecimiento funcional mediante la herramienta Pathview

Con los genes diferencialmente expresados obtenidos después del análisis de *Limma*, se utilizó la librería de R *Pathview*, capaz de integrar los datos de la expresión diferencial en las rutas de *KEGG*. Se obtuvieron 11 rutas diferencialmente enriquecidas con un *FDR*<0.1 (**Tabla 27**).

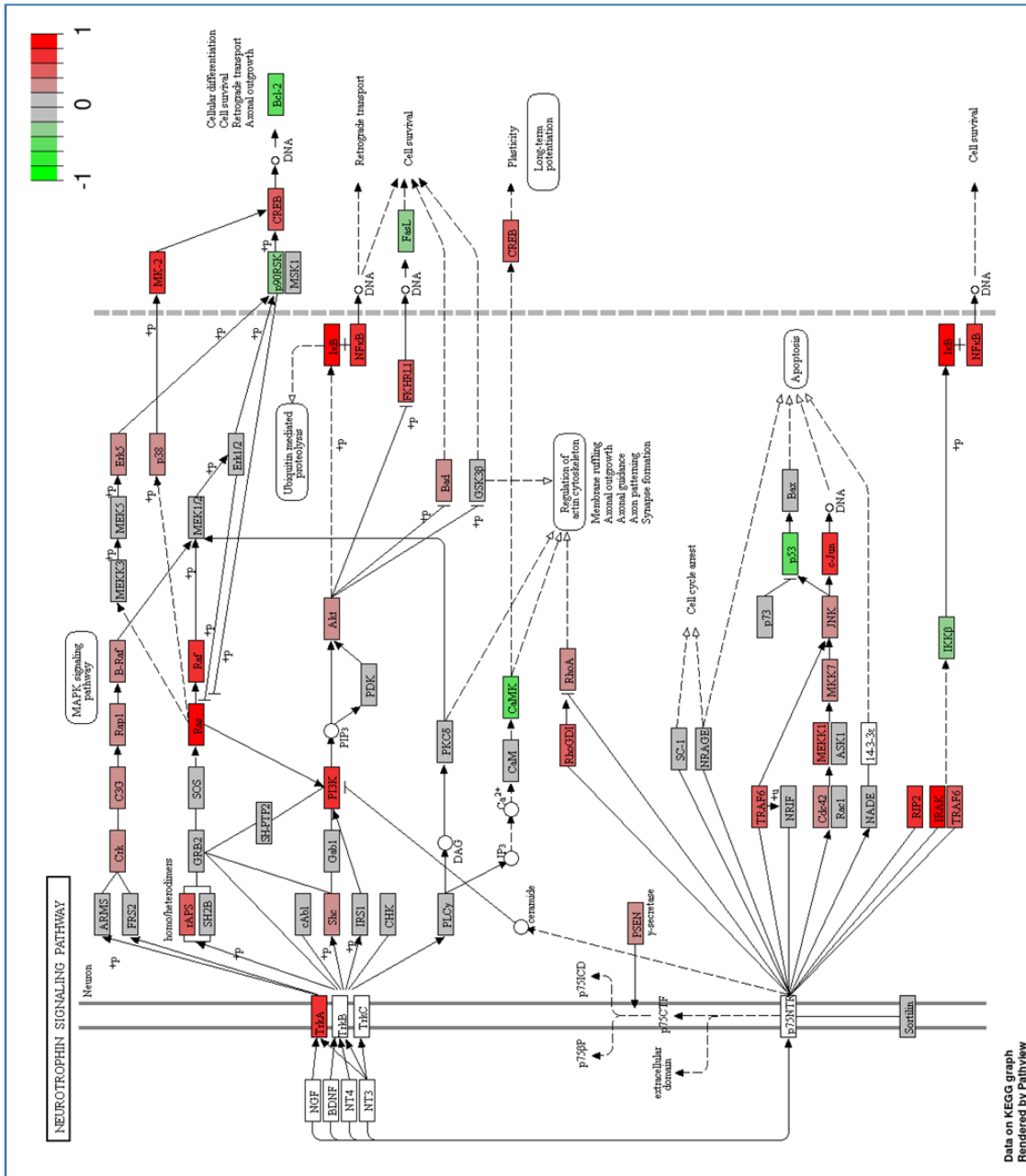
## RESULTADOS

ID. KEGG PATHWAY	FDR	set.size
hsa04722_Neurotrophin signaling pathway	0.02377120	99
hsa04668_TNF signaling pathway	0.02377120	89
hsa04141_Protein processing in endoplasmic reticulum	0.02377120	146
hsa04662_B cell receptor signaling pathway	0.02377120	65
hsa04010_MAPK signaling pathway	0.04099438	214
hsa04380_Osteoclast differentiation	0.08212209	99
hsa04371_Apelin signaling pathway	0.08212209	97
hsa04012_ErbB signaling pathway	0.08212209	67
hsa04540_Gap junction	0.08212209	60
hsa04210_Apoptosis	0.08860698	119
hsa04657_IL-17 signaling pathway	0.09721288	63

Tabla 29. Rutas de KEGG diferencialmente expresadas con  $FDR < 0,1$  obtenidas con la librería de R *Pathview*

Los resultados obtenidos con esta librería vendrían a corroborar los obtenidos con *GSEA*, mostrando en muchos casos las mismas rutas significativamente enriquecidas. Las 3 primeras son:

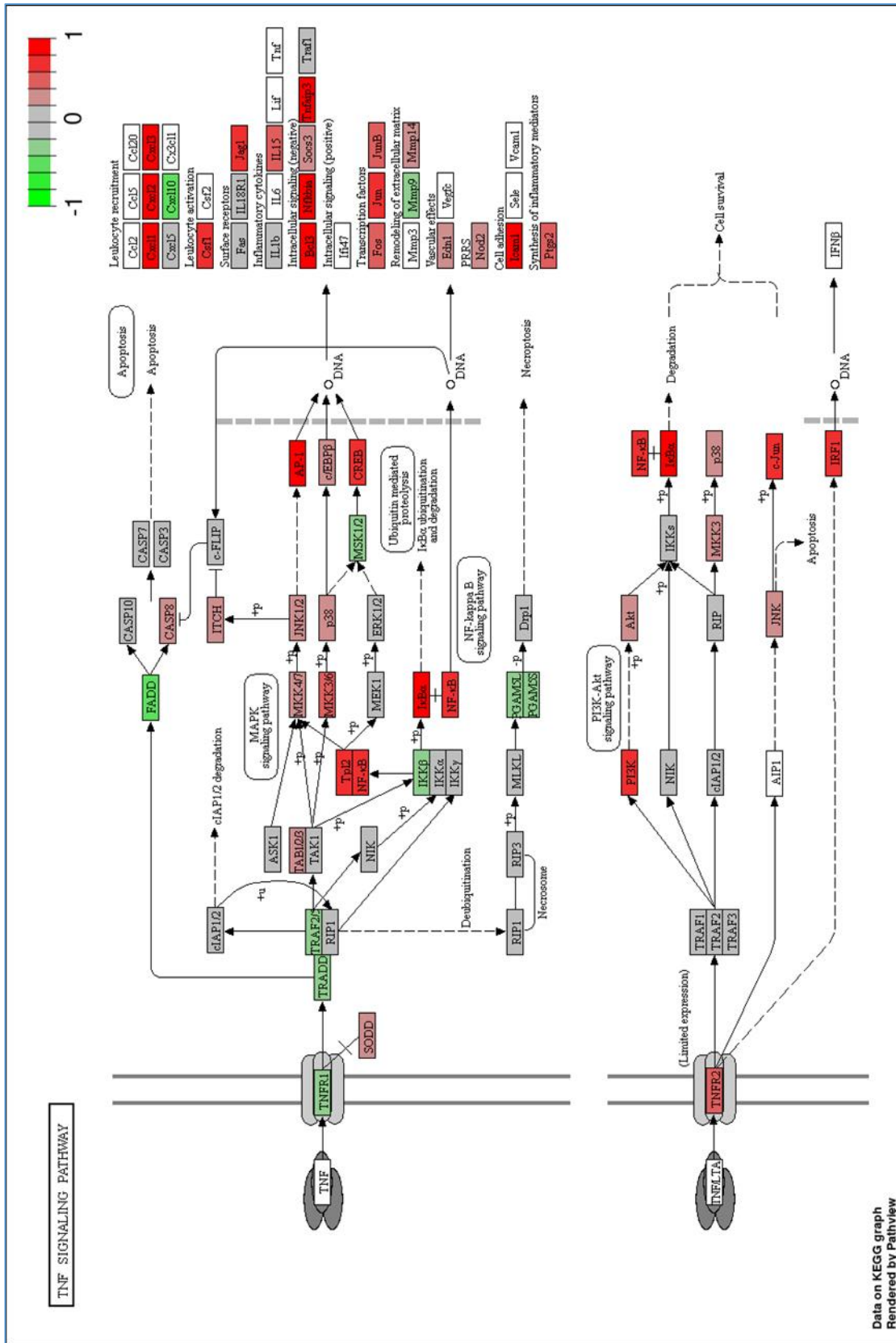
- Hsa04722 (**Figura 34**) estaría implicada en el crecimiento, la diferenciación y supervivencia de las neuronas. Esta ruta está relacionada con la de MAPK, apoptosis, proteólisis mediada por ubiquitina, diferenciación y supervivencia celular.
- Hsa04668 (**Figura 35**) está implicada en un conjunto de procesos que incluyen la apoptosis, supervivencia celular, inflamación e inmunidad.
- Hsa04141 (**Figura 36**) está involucrada en el procesamiento de las proteínas en el retículo endoplásmico, lo que implica su correcto plegamiento y la activación del proteosoma si este plegamiento es incorrecto.



Data on KEGG graph  
Rendered by Pathview

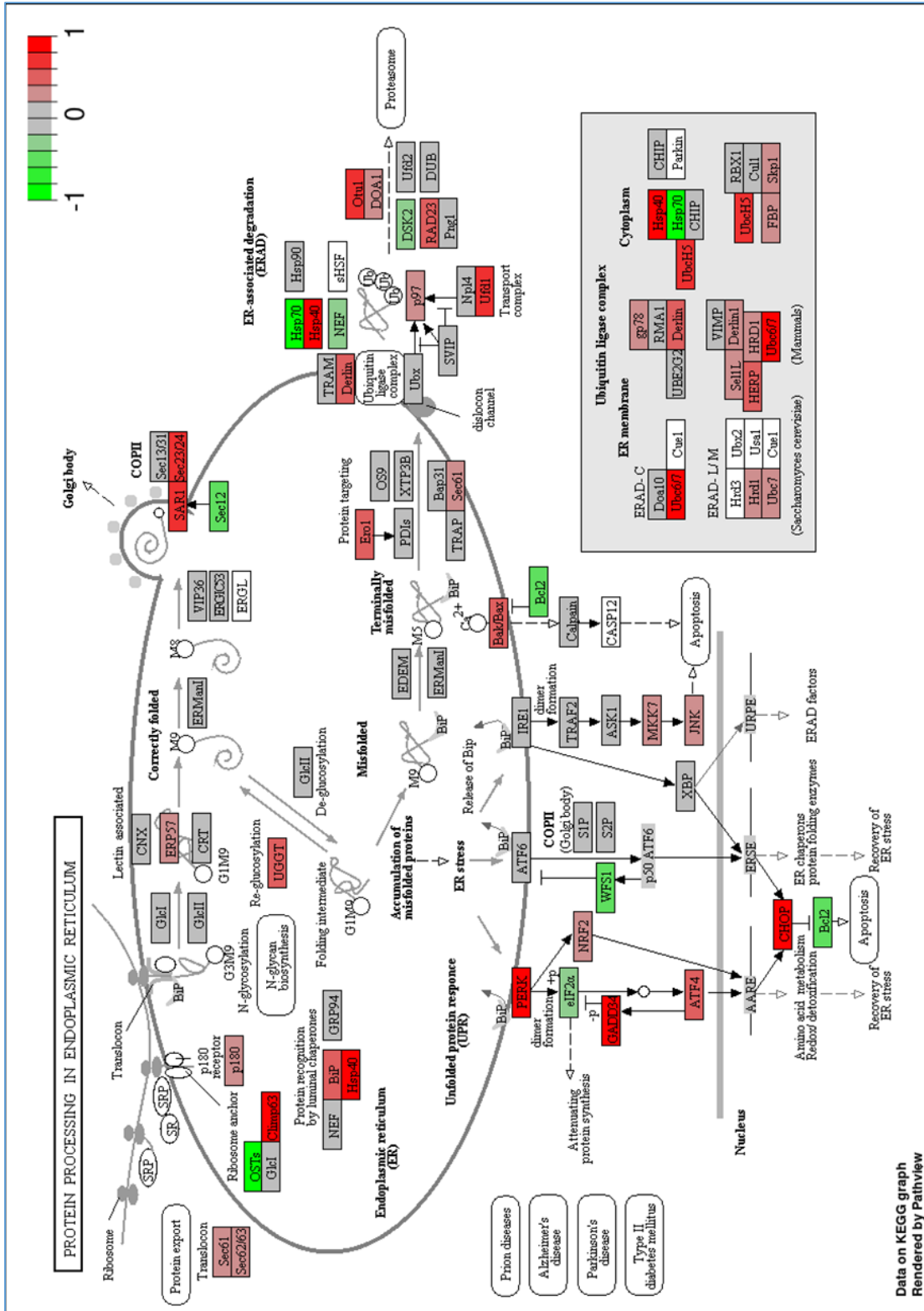
**Figura 34. Ruta hsa04722\_Neurotrophin signaling pathway.**

Análisis realizado con Pathview con los datos obtenidos después del análisis de expresión diferencial llevado a cabo con *Limma*. En rojo aparecen los genes con aumento de expresión, en gris los que no muestran cambios significativos, en verde aquellos genes que muestran un descenso de la expresión diferencial y en blanco aparecen aquellos genes que no forman parte del estudio.



Data on KEGG graph  
 Rendered by Pathview

**Figura 35. Ruta hsa04668\_TNF signaling pathway.** Análisis realizado con Pathview con los datos obtenidos después del análisis de expresión diferencial llevado a cabo con Limma. En rojo aparecen los genes con aumento de expresión, en gris los que no muestran cambios significativos, en verde aquellos que muestran un descenso de la expresión diferencial y en blanco aparecen aquellos genes que no forman parte del estudio.



**Figura 36. Ruta hsa04141\_Protein processing in endoplasmic reticulum.** Análisis realizado con *Pathview* con los datos obtenidos después del análisis de expresión diferencial llevado a cabo con *Limma*. En rojo aparecen los genes con aumento de expresión, en gris los que no muestran cambios significativos, en verde aquellos genes que muestran un descenso de la expresión diferencial y en blanco aparecen aquellos genes que no forman parte del estudio



## DISCUSIÓN

### 5. DISCUSIÓN

#### 5.1. Diagnóstico y caracterización molecular de pacientes con sospecha de Enfermedad de Pompe.

El análisis genético tiene un papel confirmatorio esencial tras el diagnóstico bioquímico de la enfermedad de Pompe. En una situación ideal, el diagnóstico se confirma en un paciente con actividad enzimática reducida en muestra de sangre periférica o en otro tejido, con la identificación posterior de 2 mutaciones patogénicas conocidas en *trans*. En la forma infantil de la enfermedad, la naturaleza de las mutaciones causantes de ese fenotipo es muy deletérea, por lo que el perfil genético suele confirmar el diagnóstico enzimático en el que el nivel de GAA es inferior al 1% de la actividad normal. En las formas tardías de la enfermedad, el déficit enzimático confirmado suele ser el reflejo de un perfil genético con la presencia de mutaciones con menor capacidad lesiva sobre la funcionalidad del gen, y del que la mutación c.-32-13T>G suele formar parte.

En este trabajo de tesis se ha realizado la caracterización molecular de una cohorte de 259 pacientes con sospecha de enfermedad de Pompe. Inicialmente el abordaje, que se puso a punto en 2008, consistía en la amplificación por PCR y posterior secuenciación por técnicas de *Sanger* de las regiones exónicas e intrónicas flanqueantes. Un esquema similar al que había sido utilizado previamente por otros autores (McCready et al., 2007). Sin embargo, la necesidad de caracterizar adecuadamente la alteración presente en el paciente P11, un varón de 7 meses de edad con la forma infantil de la enfermedad, hizo que en nuestro laboratorio se pusiera a punto un procedimiento de *MLPA* específico para esta patología. De este modo, se identificó una delección en este paciente, no descrita previamente, que afectaba a toda la estructura del gen *GAA* y parte del gen contiguo *CCDC40*, y que es la delección más grande detectada en el gen *GAA* hasta la fecha (Amiñoso et al., 2013). Por lo que recomendamos el uso del *MLPA* dentro del abordaje diagnóstico de rutina utilizado en esta patología. Mediante este análisis molecular hemos podido confirmar el diagnóstico 76 pacientes.

##### 5.1.1. Pacientes con fenotipo infantil

Hemos estudiado 12 pacientes con el fenotipo infantil que presentan una combinación de mutaciones muy deletéreas. Dentro de este grupo de pacientes que tienen una clínica más severa, podemos diferenciar entre los que muestran una cantidad residual de enzima (*CRIM+*), que normalmente tienen una mejor respuesta al tratamiento, frente a los que presentan una cantidad de enzima indetectable (*CRIM-*), y que suelen desarrollar una respuesta inmune frente a la *rhGAA*, lo que conlleva una peor respuesta al tratamiento (Berrier et al., 2015). De estos 12 pacientes, 8 presentan mutaciones descritas. En base al tipo de mutación y los criterios utilizados en el Pompe Center respecto a la posible presencia de enzima detectable (*CRIM*), 4 de estos pacientes pueden considerarse *CRIM+*. El paciente

P121, tiene un genotipo compuesto por dos mutaciones *missense* (p.Y292C, p.D645Y). En ambos casos se trata de sustituciones de un aminoácido por otro que generan un estatus *CRIM+*. El paciente en el que hemos encontrado esta combinación de mutaciones presenta un fenotipo muy severo de la enfermedad, lo que nos indica la profunda deficiencia funcional que generan ambas mutaciones. En la base de datos del Pompe Center, la mutación p.Y292C ha sido identificada una vez formando parte de un genotipo, junto con un alelo nulo, del que resulta también una forma severa de la enfermedad. La variante p.D645Y no ha sido descrita en pacientes con la forma severa de la enfermedad, por lo que hasta ahora no disponíamos de datos acerca de su efecto deletéreo sobre la actividad enzimática. La identificación en este trabajo como parte del genotipo de un paciente con una forma grave, permite deducir un efecto muy dañino sobre la actividad del enzima.

Los pacientes P119, P454 y P478, tienen la c.1551+1G>A en homocigidad. Una mutación de procesamiento clasificada como *CRIM+* en Pompe Center, que afecta el 5'ss del intrón 10. Estudios funcionales sobre el procesamiento ponen de manifiesto la pérdida completa del exón 10, pero que no ocasiona un cambio de fase de lectura (Bergsma et al., 2015), por lo que es posible que la proteína resultante pueda ser detectada en sangre periférica.

Dos pacientes (P137 y P391) presentan la mutación p.R854\* en homocigidad lo que implica la presencia prematura de una señal de stop y, con ello, la interrupción de la síntesis de la estructura primaria de la proteína, lo que conlleva una ausencia total de enzima, y resulta un estatus *CRIM-*.

El paciente P257, presenta un genotipo compuesto por la mutación p.R600C, clasificada como *CRIM+*, y la p.R854\* clasificada como *CRIM-*. En este caso, el alelo con la mutación *missense* (p.R600C) condiciona el estado *CRIM+*. La mutación (p.R600C) implica uno de los 13 aminoácidos que conforman el centro activo del enzima (Deming et al., 2017). Desde el punto de vista de la severidad de la enfermedad, ambas mutaciones han sido previamente relacionadas con el fenotipo más severo (Tsuji et al., 2000).

Por último, el paciente P83, presenta en homocigidad la mutación p.G643R, de la que no se han publicado datos acerca de la cantidad de enzima detectable en sangre periférica, por lo que desconocemos si genera un estado *CRIM+* o -, aunque al tratarse de una *missense*, podríamos especular con que fuese *CRIM+*. Esta mutación ha sido previamente descrita en homocigidad en un paciente con la forma severa (Pittis et al., 2008).

Los otros 4 pacientes con fenotipo infantil presentan en su genotipo mutaciones descritas por primera vez en este trabajo. El paciente P11 presentaba en heterocigidad una delección que afecta a toda la estructura del gen *GAA* y parte del gen contiguo *CCDC40* (Amiñoso et al., 2013). Se ha podido caracterizar que en uno de los extremos de la delección, se encuentra una secuencia *Alu* previamente

descrita como un potencial punto caliente de recombinación, ya que se ha descrito su papel como causante de la deleción c.2189+459\_3405del (Huie et al., 1999). Detectamos además una secuencia insertada de 14pb entre los puntos de ruptura que, junto con un fragmento de 15pb de la secuencia del *CCDC40*, forman una secuencia *Alu Jr*. Esta secuencia se ha podido localizar en el intrón 2 del gen *CDH4*, y creemos que podría haberse producido una recombinación entre ambos fragmentos como mecanismo patogénico que explicase esta compleja deleción-inserción. Este paciente tenía la mutación p.G309R, situada en el exón 5, en el otro alelo. Una variante que ha sido descrita en homocigosidad en un paciente con la forma infantil, *CRIM+* (Elder et al., 2013).

Los pacientes P451, P230 y el P349 tienen también un fenotipo infantil y presentan mutaciones con un carácter muy deletéreo.

El paciente P451 presenta un genotipo formado por las mutaciones c.236\_246del (p.P79Rfs\*13) y c.956-5\_957delTCCAGAT, esta última no descrita anteriormente. La variante c.956-5\_957del afecta el 3'ss del intrón 5, y los dos primeros nucleótidos del exón 6. La falta de RNA disponible ha impedido que se pudiera llevar a cabo un estudio funcional de esta variante. No obstante, dada la secuencia implicada, y al estar presente en un paciente con un fenotipo infantil, se puede anticipar la gravedad del cambio. Todas las herramientas de análisis bioinformático proponen el uso de un sitio aceptor alternativo formado por un AG (c.968-969), situado en el exón 6, lo que provocaría un cambio de fase de lectura y la introducción de un codón de parada prematuro 6 aminoácidos después.

El paciente P230 tiene un genotipo formado por las variantes c.1416dupC (p.G473Rfs\*33) y la c.2662G>T (p.E888\*). Ambas mutaciones tienen un carácter muy deletéreo, de modo que se interrumpe de forma prematura la síntesis de la cadena polipeptídica con ambos alelos. La mutación p.E888\* se ha identificado en el 23,1% de población china afectada de enfermedad de Pompe (Chen et al., 2017), etnia a la que pertenece este paciente. La mutación p.G473Rfs\*33 no ha sido previamente identificada e implica un cambio de fase de lectura a partir del dominio catalítico, lo que genera la pérdida de toda la región del carboxilo terminal.

El paciente P349 presenta en homocigosidad la mutación c.1411 G>T (p.E471\*), situada en el exón 9. El paciente es de etnia árabe y aunque no se nos ha facilitado el antecedente de consanguinidad, el resultado de nuestro estudio sugiere esa posibilidad. Con el estudio funcional realizado sobre mRNA hemos podido cuantificar que la expresión del gen es prácticamente nula, lo que sería compatible con el lanzamiento del mecanismo *NMD*.

### 5.1.2. Pacientes con fenotipo tardío

Sesenta y cuatro pacientes de este estudio presentan un fenotipo tardío y 54 de ellos tienen la variante c.-32-13T>G en su genotipo, lo que supone un 84,3%. Esta variante implica un procesamiento aberrante del intrón 1, apareciendo distintas poblaciones de mensajero en las que se produce una pérdida total o parcial del exón 2, lo que altera el balance entre las diferentes isoformas presentes en el procesamiento su normal (Dardis et al., 2013). Se observa con mucha frecuencia entre pacientes de origen caucásico, con un fenotipo tardío y con una frecuencia que varía mucho dependiendo del estudio consultado, oscilando entre el 34,4%, observado en un estudio realizado sobre población española (Gort et al., 2007), hasta el 96,6%, obtenido por un grupo italiano (Montagnese et al., 2015). En la última revisión de la base de datos de mutaciones del Pompe Center se ha identificado esta variante en 253 pacientes, lo que supone una frecuencia global del 86% de los pacientes recopilados con fenotipo tardío (294). Dado el alto número de pacientes recopilados, posiblemente esa frecuencia se aproxime más a un valor real que otros estudios con muestras de pacientes más modestas (Niño et al., 2019). En este trabajo de tesis se ha detectado una frecuencia de la variante c.-32-13T>G del 84,3%, muy similar a la que refleja el Pompe Center. Nuestros resultados muestran una frecuencia de esta variante en población española muy diferente a la del trabajo de Gort et al. (Gort et al., 2007). En nuestra cohorte, de 64 pacientes con fenotipo adulto, solo 2 pacientes eran homocigotos para la variante c.-32-13T>G (3,1%). En el artículo de Musumeci et al. (Musumeci et al., 2015), realizan un cálculo para estimar la prevalencia de homocigotos en población holandesa, alemana e italiana, en base a la prevalencia estimada de la enfermedad en estas poblaciones y el porcentaje de pacientes que presentan esta mutación. Con estos cálculos, establecen que aproximadamente el 16% de los pacientes con fenotipo adulto deberían ser homocigotos para esta variante. Sin embargo la frecuencia reportada en estas poblaciones es más baja y dista del valor detectado en nuestro estudio. Esto parece que podría explicarse con una penetrancia reducida o incompleta de este alelo, resultado de una combinación de diferentes factores genéticos y ambientales (Musumeci et al., 2015).

En 17 pacientes con la forma tardía de la enfermedad hemos identificado en sus genotipos una variante descrita previamente en la literatura junto con una variante no descrita (**Tabla 14**). En conjunto, se trata de 14 variantes nuevas, de las cuales 8 pueden considerarse como patogénicas debido a su naturaleza deletérea. Además, 12 de ellas se encuentran en *trans* con una mutación descrita en el otro alelo. Una de estas variantes corresponde a una delección detectada gracias al *MLPA*. Esta delección, detectada en el paciente P481, implica a parte del intrón 19 y toda la parte codificante del exón 20, incluido el codón de parada. La consecuencia de esta delección supone la pérdida del 3' ss del intrón 19, junto con los 20 aminoácidos del carboxilo terminal resultantes de la traducción del fragmento codificante

del exón 20. Este paciente tiene la mutación c.-32-13T>G en el otro alelo, lo que condiciona el fenotipo de la enfermedad.

Un grupo de 6 pacientes presentan un genotipo en el que la mutación c.-32-13T>G se combina en *trans* con mutaciones de procesamiento no descritas previamente, que afectan las secuencias consenso 5'ss o 3'ss. En 4 de ellos, la mutación es c.1889-1G>A, que afecta el 3' acceptor del intrón 13. Las herramientas bioinformáticas integradas en *Alamut* predicen la alteración del 3' acceptor como consecuencia de la mutación. El análisis funcional realizado en la muestra del paciente P210, ha dado como resultado que la maquinaria de *splicing* utiliza un AG alternativo situado en el interior del exón 14, lo que supone la pérdida de 11 pares de bases de este exón, la alteración de la fase de lectura y la aparición de un codón stop prematuro inmediato. El análisis cuantitativo del mRNA ha puesto de manifiesto un notable descenso de su nivel expresión, lo que sugiere la posible activación del sistema *NMD* para eliminar el mRNA aberrante.

El paciente P300 tiene un genotipo formado por c.-32-13T>G y c.1637-2delA, localizada esta última en el dinucleótido AG acceptor del intrón 11. En este mismo nucleótido se ha identificado en homocigosidad la mutación c.1637-2A>G en una paciente mediante diagnóstico prenatal, lo que permitió un tratamiento prematuro, antes incluso del comienzo de los síntomas (Herzog et al., 2012). Los predictores integrados en *Alamut* indican que ese sitio acceptor se pierde. El análisis funcional ha puesto de manifiesto la utilización de un 3' acceptor críptico alternativo localizado en el exón 12, concretamente el AG situado en c.1658-1659 y precedido de 6 pirimidinas. El resultado es la pérdida de parte del exón 12, la alteración de la fase de lectura y la aparición de un stop prematuro 83 aminoácidos más tarde. La cuantificación del mRNA ha resultado con niveles muy desiguales entre los exones, por lo que el proceso de cuantificación puede reflejar interferencias analíticas.

El paciente P290 tiene un genotipo compuesto por c.-32-13T>G y c.1075+1G>A. Esta mutación está localizada en el 5' donador del intrón 6. Los predictores indican la pérdida del sitio de procesamiento normal. El estudio funcional ha puesto de manifiesto la aparición de 2 poblaciones anormales de mRNA. Por un lado, una banda resultado de la pérdida del exón 6, junto con intrón 5 e intrón 6 y, por otro lado, otra banda generada por la permanencia del intrón 6. La cuantificación del mRNA refleja un descenso aproximado del 50%. En este caso, el descenso es menos acusado que el observado en el análisis del paciente P210 con la mutación en el 3' acceptor del intrón 13.

En 3 pacientes hemos detectado sendos cambios *nonsense* (p.W618\*, p.Q553\* y p.W772\*), no descritas previamente. Las 2 primeras localizadas en el dominio catalítico y la última en el carboxilo terminal. En los pacientes P124 y P151 se encuentran en *trans* con la mutación c.-32-13T>G. En el paciente P180 la mutación *nonsense* aparece formando parte del genotipo en *trans* con la c.2173C>T (p.R725W).

Este paciente comienza con los síntomas a la edad de 20 años, presentando debilidad muscular en miembros inferiores e insuficiencia respiratoria. En este grupo de afectados las mutaciones *nonsense* alteran la síntesis de la cadena polipeptídica, que puede ser eliminada por el proteosoma (Ashida et al., 2006) o bien el mRNA mediante la activación del *NMD* (Kurosaki and Maquat, 2016). En cualquiera de los dos casos, suponen la ausencia de enzima producida por el alelo con ese tipo de mutación. El fenotipo de los 3 pacientes es tardío, como consecuencia de la actividad enzimática producida por el alelo con la mutación c.-32 -13 T>G o la p.R725W. En el caso de esta última, la base de datos del Pompe Center tiene recogidos varios casos asociados a forma juvenil. No se ha realizado estudio funcional de estas variantes al no disponer de muestra apropiada.

El paciente P273 tiene un genotipo formado por la duplicación c.2334\_2335dupGC (p.P779Rfs\*3), no descrita con anterioridad, junto con la c.-32-13T>G. La duplicación genera un cambio de fase de lectura y la aparición de un codón stop muy inmediato, por lo que el efecto deletéreo es muy severo sobre la expresión del gen. Al igual que en otras mutaciones incluidas en este estudio, podría activarse el *NMD* para eliminar el mRNA aberrante o bien se podría eliminar la cadena peptídica sintetizada a través del proteosoma. No se ha realizado estudio funcional alguno sobre esta mutación al no disponer de muestra de sangre para ello.

Por último, se han identificado 6 variantes *missense* nuevas, de las que 4 forman parte de genotipos con mutaciones ya descritas patogénicas en *trans*. Este hecho, implica que esas 4 variantes vayan a ser clasificadas como probablemente patogénicas al aplicar las recomendaciones del *ACMG*, respecto de la interpretación con fines diagnósticos. En el paciente P50 observamos la mutación patogénica c.1064T>C (p.L355P) con la c.1209C>G (p.N403K). Esta última, fue reportada por primera vez por nuestro grupo, en el año 2008, a la base de datos del Pompe Center. El paciente comenzó con síntomas en los primeros años de vida. La mutación p.L355P se localiza en el dominio catalítico, donde parece generar un giro de la proteína en una hélice alfa. Esta variante ha sido asociada con la forma más severa de la enfermedad en varias ocasiones ([www.pompevariantdatabase.nl](http://www.pompevariantdatabase.nl)). La p.N403K ha sido recientemente vinculada a un paciente con la forma juvenil (Wens et al., 2015). Esta mutación podría afectar una lámina beta dentro del dominio catalítico y muy próxima a aminoácidos que forman parte del sitio activo del enzima (Deming et al., 2017).

El paciente P388 presenta la variante c.1064T>C (p.L355P) junto con la c.713C>T (p.P238L), esta última no descrita previamente. Este paciente presenta una forma juvenil de la enfermedad, encontrándose en estos momentos en tratamiento con la enzima recombinante (*rhGAA*). La mutación p.L355P, ha sido descrita previamente en el paciente anterior. La variante p.P238L se localiza en el

dominio amino terminal de la proteína. Dada la severidad de la mutación presente en el otro alelo se puede deducir que esta variante genera una proteína que mantiene una actividad enzimática residual.

El paciente P106 presenta un genotipo constituido por la mutación c.-32-13T>G y la c.1447G>T (p.G483W), situada en el dominio catalítico del enzima. Esta variante se describe por primera vez en este trabajo. Se han identificado varias mutaciones que afectan al mismo aminoácido (p.G483R y p.G483V) ([www.pompevariantdatabase.nl](http://www.pompevariantdatabase.nl)). La variante p.G483W, afecta a un aminoácido que, aunque no corresponde a uno de los críticos que conforman el centro activo, hemos podido determinar mediante la herramienta *Dynamut* (Rodrigues et al., 2018), que forma parte del bolsillo de unión al glucógeno, por lo que la mutación puede modificar su afinidad frente al sustrato.

El paciente P352 muestra un genotipo con la mutación c.-32-13T>G y la c.2141A>C (p.H714P), no descrita con anterioridad. La variante p.H714P se localiza entre el dominio catalítico y el carboxilo terminal, dentro de una hélice alfa, lo que podría generar un giro en su estructura (Deming et al., 2017).

La nueva variante c.1328A>T (p.D443V), detectada en homocigosidad en el paciente P479, se encuentra localizada en un aminoácido altamente conservado. Los algoritmos de predicción utilizados en este trabajo le otorgan un posible papel patogénico. Estos resultados no modifican la clasificación de esta variante que es considerada como de significado incierto en un contexto diagnóstico. Son necesarios estudios posteriores para poder modificar su clasificación. No obstante, el alto grado de conservación encontrado en este aminoácido, nos hace considerar su posible patogenicidad.

El paciente P129 presenta una variante no descrita c.1679C>G (p.S560C) y en el otro alelo c.576G>C (p.E192D), descrita como VSD. Ambos cambios se localizan en aminoácidos altamente conservados y todos los algoritmos les otorgan un posible papel patogénico, excepto la herramienta *Revel* que clasifica la p.E192D como neutra. Al igual que la variante anterior, no se puede considerar que este genotipo pueda ser considerado patogénico en un ámbito diagnóstico.

En 10 pacientes hemos detectado 7 variantes descritas como VSDs en las bases de datos (**Tabla 10**). De ellas, 4 (p.D282H, p.H568Q, p.P482L y p.N470del), presentes en 8 de esos pacientes, han sido detectadas junto con otra mutación patogénica en *trans*, lo que permite modificar su clasificación a probablemente patogénicas, en aplicación de los criterios de la *ACMG*.

La p.D282H afecta a un aminoácido que forma parte del centro activo, lo que podría modificar la afinidad por el sustrato.

La variante c.1704C>G (p.H568Q), localizada en el exón 12, ha sido detectada en el paciente P70 junto con otra mutación descrita como patogénica, la c.1655T>C (p.L552P), lo que en principio modifica



el estatus de la primera a probablemente patogénica, siguiendo las recomendaciones de la ACMG (Richards et al., 2015). La variante p.H568Q fue identificada por primera vez como parte de este trabajo en el año 2008, e informada como tal en un contexto diagnóstico. Los responsables clínicos de la asistencia del paciente publicaron las características clínicas, junto con la identificación de la mutación, sin considerar a nuestro grupo en la autoría del trabajo (Alejaldre et al., 2012). El estudio funcional de procesamiento del mRNA realizado por nosotros ha permitido determinar la implicación de esta variante en la alteración del procesamiento del exón 12, generando un sitio 3' acceptor alternativo en este exón. La consecuencia es la pérdida de parte del exón 12 en el mRNA, lo que provoca un cambio de fase de lectura y un codón de stop 67 aminoácidos después.

La variante c.1445C>T (p.P482L) ha sido detectada en dos pacientes. En el P75 junto con la mutación c.-32-13T>G y en el paciente P318 junto con la c.343C>T (p.Q115\*). La variante p.P482L, afecta a un aminoácido que, aunque no corresponde a uno de los aminoácidos críticos que conforman el centro activo, hemos podido determinar utilizando la herramienta *Dynamut* (Rodrigues et al., 2018), que forma parte del bolsillo de unión al glucógeno, por lo que esta mutación puede modificar la afinidad de GAA frente al sustrato. Además, en el artículo de Deming et al (Deming et al., 2017), tienen en cuenta que esta prolina está implicada en un giro dentro de una lámina beta, por lo que una leucina, previsiblemente va a modificar la estructura terciaria del polipéptido. Hemos realizado un análisis funcional de procesamiento del mRNA con muestra del paciente P318, para ver si c.1445C>T (p.P482L) tiene alguna implicación en el *splicing* del exón 10, donde se localiza. Con este análisis hemos podido determinar que la variante p.P482L no genera un procesamiento aberrante. Como el análisis funcional incluye toda la estructura del gen hemos observado que en el exón 2, donde está localizada la p.Q115\*, presente en el otro alelo, sí que muestra un procesamiento aberrante de este exón. Mediante este estudio hemos podido detectar que el cambio de C>T en el c.343 genera un 5'donador críptico que, en caso de ser utilizado por la maquinaria de *splicing*, provoca la eliminación de parte del exón 2. La complejidad de procesamiento de este exón, incluso en alelos normales, ante la presencia de esta mutación crea alguna alternativa adicional. Por un lado, el 5' críptico junto con el 3' críptico del nucleótido c.486, puede producir la eliminación de un fragmento del exón 2 que se habría convertido en un pseudointrón de 145 nucleótidos. Por otro lado, el 5' críptico generado por la mutación puede ser utilizado junto con el 3' acceptor normal del intrón 2, provocando la deleción de una fracción de 225 nucleótidos del 3' del exón 2. En ambos casos se produce un cambio de fase de lectura, y la introducción de un codón stop prematuro.

La variante c.1408\_1410delAAC (p.N470del), detectada en 4 pacientes, elimina uno de los 7 aminoácidos que son glicosilados en la proteína en el RE, previo a su transporte al Golgi, durante el proceso de maduración. Hay varios artículos en los que realizan una modificación de los sitios de glicosilación en distintas enzimas lisosomales. En ellos determinan que los sitios de glicosilación están

implicados en el plegamiento, el tráfico del RE al Golgi y, en último término, en la actividad de la enzima, dependiendo del sitio de glicosilación que sea afectado (Wujek et al., 2004; Hiraoka et al., 2013). Aunque son necesarios estudios adicionales para valorar el efecto exacto que la mutación p.N470del tiene sobre la proteína, parece clara su probable implicación en la fisiopatología de la enfermedad.

Las otras 3 variantes consideradas como VSDs corresponden al cambio c.1504A>G (p.M502V), detectada en homocigosidad en el paciente P299, y las variantes c.851C>G (p.A284G) y c.2510G>A (p.R837H), detectadas en el paciente P401. Se ha realizado un análisis bioinformático mediante herramientas *in silico* con el fin de explorar su posible papel en la fisiopatología de la enfermedad. Todas las herramientas utilizadas, otorgan un posible papel patogénico a la variante p.R837H, ya que se encuentra localizado en un aminoácido muy conservado. No se ha obtenido un resultado concluyente en las otras 2 variantes.

En dos pacientes detectamos una variante intrónica, la c.1194+5G>A, que ha sido descrita como no patogénica en base a un análisis *in silico* (Kroos et al., 2012). (**Tabla 11**). Esta variante, detectada en homocigosidad en la paciente P346, está localizada en la secuencia consenso de unión de la ribonucleoproteína U1 (AUACUUACCUG), y además afecta a una de las tres bases a las que se une U6, previo al inicio de la primera reacción de transesterificación (Patel and Steitz, 2003). La secuencia 5´donadora a la que se une la U1, tiene distintos grados de conservación, pero estudios de análisis comparativo han puesto de manifiesto el papel fundamental en el procesamiento de los nucleótidos -1 y +5 en la zona 5´de procesamiento (Carmel et al., 2004; Shepard et al., 2011), además de la secuencia GT. Estos datos refuerzan los resultados obtenidos en nuestro estudio, en el que hemos podido caracterizar que la variante c.1194+5G>A provoca una interrupción del uso del sitio 5´donador del exón 7, modificando el procesamiento normal y cuya principal consecuencia es la generación de un transcrito en el que se pierde el exón 7, lo que interrumpe la fase de lectura e introduce un codón de parada precoz, con el consiguiente descenso en la cuantificación de la expresión del mRNA que oscila entre un 20-40%, con una reducción adicional en la expresión del exón 7 que sería de solo el 15%. Esta variante ha sido previamente analizada con herramientas bioinformáticas para valorar el papel de este cambio en el procesamiento. De las 5 herramientas utilizadas, solo la herramienta NNS ha anticipado fehacientemente la implicación de la c.1194+5G>A en un procesamiento aberrante del mensajero, lo que pone de manifiesto la importancia de los análisis funcionales en el esclarecimiento de la posible implicación en el *splicing* de variantes sospechosas. Estos resultados demuestran la patogenicidad de la variante, por lo que debería ser considerada como tal en las bases de datos de mutaciones (Amiñoso et al., 2019).

Hemos analizado una última variante, la c.546G>C (p.T182T), localizada en el paciente P345 y en el que no encontramos otro cambio, siendo considerado como portador de la enfermedad. La variante,

ha sido descrita en el artículo de Gort et al. (Gort et al., 2007), pero solo especulan con la posible implicación en el *splicing*, ya que se encuentra situado en el último nucleótido del exón 2. También se ha descrito la variante c.546G>A (Hermans et al., 2004) y la c.546G>T (Maimaiti et al., 2009; Tsuburaya et al., 2012). Solo en el caso de Tsuburaya et al., detectan un *splicing* aberrante, fruto del procesamiento de la variante c.546G>T en homocigosidad, que implica la pérdida del exón 2 en el mensajero maduro. En nuestro caso, la variante c.546G>C, genera un procesamiento aberrante, ya que al modificarse la secuencia de unión de la ribonucleoproteína U1, la maquinaria utiliza un sitio 5´donador críptico localizado en el intrón 2, lo que provoca la aparición de un RNA mensajero en el que permanecen 182pb del intrón, modificando la pauta de lectura e introduciendo un codón stop precoz en el exón 6.

La consulta a la base de datos del Pompe Center (Niño et al., 2019) permite observar que las variantes descritas en este gen se encuentran distribuidas a lo largo de toda la estructura. Un dato que coincide con los resultados de nuestro estudio, ya que no hemos detectado una distribución de mutaciones enriquecida, a excepción de la c.-32 -13T>G (**Figura 18**). No obstante se han descrito mutaciones con una mayor representación en distintos grupos poblacionales. Uno de los mejores ejemplos serían las mutaciones c.525delT y la c.2481+102\_2646+31del, que son detectadas con una frecuencia alélica del 35% y 31% respectivamente en población holandesa (Kroos et al., 1995). La p.E888\* y la p.D645Y detectadas con una frecuencia alélica de 23,1% y 20,8% respectivamente en población china, siendo la p.E888\* la más frecuente en la población del norte y la p.D645Y la más frecuentemente detectada en el Sur de China (Chen et al., 2017). Otra mutación muy bien caracterizada es la p.R854\*, con una frecuencia alélica estimada del 43,4% en población americana de origen africano (Becker et al., 1998). Para población española, existe un estudio de Gort et al., en el que describen la mutación c.1076-1G>C como la segunda más frecuente después de la c.-32-13T>G (Gort et al., 2007), con un 14% de frecuencia alélica. En nuestro estudio, solo 1 paciente de los 76 diagnosticados presenta esta mutación, lo que estaría muy por debajo del 14% estimado. En este trabajo, no hemos detectado una mutación más prevalente, pero si hemos revelado 7 mutaciones, la p.Y191\*, p.Y292C, p.Q115\*, p.N470del, p.L552P, c.1551+1G>A y la c.1889-1G>A presentes cada una de ellas en 4 pacientes, lo que representaría un 5,26% cada una, lo que estaría por debajo del 14% que estiman Gort et al., para la mutación c.1076-1G>C y muy por debajo de las frecuencias alélicas detectadas en las otras mutaciones previamente descritas.

De los 259 pacientes estudiados, 53 son portadores de un único cambio (20,47%). De estos, 36 presentan un cambio patogénico descrito en la literatura (**Tabla 12**), 15 presentan un cambio descrito como VSD (**Tabla 13**) y 2 presentan un cambio no descrito previamente (**Tabla 15**). En este caso, la variante c.-32-13T>G ha sido detectada en 25 de los 53 pacientes, lo que corresponde con un 47,17%.

Por último, en 126 (48,65%) de las 259 muestras estudiadas, no se ha detectado ningún cambio o reordenamiento que pueda considerarse patológico en el gen *GAA*. Este dato, junto con el de portadores, pone de manifiesto la enorme dificultad que existe para hacer una adecuada selección de los pacientes en esta patología. Esta es una enfermedad minoritaria que no tiene un perfil fenotípico específico, sobre todo en el caso de las formas tardías de la enfermedad en las que la clínica es menos definida.

Los resultados de este amplio estudio sobre el gen *GAA* ponen de manifiesto la complejidad de los mecanismos moleculares presentes en la enfermedad de Pompe, y que van más allá de lo que se puede anticipar en algunos casos por la naturaleza de las variantes identificadas a nivel genómico. Se han podido detectar variantes que inesperadamente tienen una clara repercusión sobre el procesamiento del gen, y grandes reordenamientos que no pueden ser analizados mediante un abordaje clásico, que incluye la tradicional secuenciación de exones y regiones intrónicas flanqueantes. En conjunto, se ha evidenciado la necesidad de introducir estudios moleculares más completos para poder identificar de manera más precisa los mecanismos que condicionan la expresión fenotípica de la enfermedad. La introducción de estudios de secuenciación masiva que incluyan toda la estructura genómica de *GAA*, junto con estudios funcionales realizados con muestras de RNA de sangre periférica pueden permitir la identificación y posterior caracterización de esos complejos mecanismos moleculares.

### 5.2. Análisis de transcriptoma mediante RNA-seq

El análisis del transcriptoma en sangre periférica de la muestra de pacientes con la forma tardía de la enfermedad de Pompe llevado a cabo en este estudio, puede ayudar a conocer cuál es la respuesta transcripcional que se produce ante los eventos moleculares que alteran la funcionalidad del gen *GAA*. Hasta donde hemos podido averiguar, se trata del primer estudio de este tipo realizado en muestras de RNA de sangre periférica en pacientes con esta enfermedad. En esta patología la sangre periférica no es el tejido principalmente afectado, pero la enfermedad sí tiene expresividad en estos tipos celulares (Hagemans et al., 2010). Gracias a la facilidad de obtención de muestras de sangre frente a muestras de tejido muscular se puede inferir que cualquier conocimiento derivado de este estudio, como puede ser la identificación de biomarcadores, podría tener una más fácil aplicación en la clínica.

El análisis de expresión diferencial ha puesto de manifiesto un total de 980 genes que presentan diferencias significativas ( $FDR < 0,05$ ) entre las muestras de pacientes y controles, de ellos, 479 muestran un nivel de expresión significativamente más elevado en los pacientes, mientras que 501 muestran un nivel de expresión significativamente disminuido en ellos.

Con los datos obtenidos en el estudio transcripcional se ha realizado un análisis de escalamiento multidimensional mediante un *MDS-Plot*, que ha mostrado mayoritariamente una separación entre

pacientes y controles, lo que indicaría que la enfermedad induce diferencias de expresión génica en sangre periférica que permiten una segregación de las poblaciones que forman parte de la muestra analizada en este estudio.

También hemos llevado a cabo un agrupamiento no supervisado, utilizando para ello dos criterios distintos, y hemos obtenido resultados muy similares en ambos casos (**Figura 31**). Analizamos el perfil de expresión obtenido con los 74 genes con mayor nivel de significancia ( $FDR < 0,01$ ) y estableciendo como segundo criterio que su  $LogFC$  sea mayor de +1 o menor de -1 (**Figura 31A**). Por otro lado, realizamos un segundo agrupamiento con los 50 genes con el estadístico  $t$  más significativo, tanto positiva como negativamente (**Figura 31B**). En ambos casos, se puede ver que existe una buena separación entre pacientes y controles. A pesar de la capacidad para alcanzar una buena agrupación de las 2 poblaciones incluidas en nuestra muestra de estudio por parte de los grupos de genes mencionados anteriormente, como se muestra en la **Figura 31**, habría que valorar la posibilidad de poder constituir una firma de genes con estas agrupaciones o con parte de ellas. Para generar estas firmas moleculares de genes capaces de identificar a los pacientes con enfermedad de Pompe, se pueden utilizar distintos métodos como la agrupación jerárquica con remuestreo (Wilkerson and Hayes, 2010), métodos de inteligencia artificial (Bartsch et al., 2016) o con redes neuronales (Khan et al., 2001).

Un aspecto que nos parece interesante explorar es si existen rutas metabólicas, de señalización celular, etc., diferencialmente expresadas entre pacientes y controles. Para ello, hemos analizado los datos obtenidos mediante el estudio de expresión diferencial, utilizando las herramientas *GSEA* y la librería de R *Pathview*. Entre las rutas que muestran un mayor número de genes con expresión diferencial elevada en los pacientes están aquellas relacionadas con la respuesta inmune, concretamente son las rutas de señalización de TNF-alfa, vía NFkB; la vía de respuesta de interferón gamma; la ruta definida como de respuesta inflamatoria; la vía de respuesta de interferón alfa y también la ruta de señalización TGF-beta. Un trabajo previo de análisis del transcriptoma realizado con muestras de músculos bíceps y cuádriceps procedentes de niños afectados con la forma infantil, más severa, de la enfermedad, realizado con *array* de expresión, puso de manifiesto un patrón de expresión similar en relación con estas rutas de señalización (Palermo et al., 2012). En nuestro estudio, el número de genes que forman parte de estas vías de señalización con niveles de expresión alterados es muy elevado, y no puede ser comparado con los resultados del estudio previo que se menciona al haber sido realizado con una metodología diferente. No obstante, la concordancia en la alteración de las rutas indicadas sugiere que, a pesar de tratarse de tejidos diferentes y pacientes con formas clínicas también distintas, los componentes de respuestas inflamatoria e inmune parecen estar alterados de un modo muy relevante. La ruta de señalización de TNF-alfa vía NFkB está implicada en procesos de proteólisis y pérdida de masa muscular, induciendo el

desarrollo de atrofia (Jackman and Kandarian, 2004), por otro lado, la alteración de la ruta de TGF-beta estimula la fibrosis en el tejido muscular (Li et al., 2004).

La elevación de los niveles de expresión de genes que forman parte de las rutas mencionadas, junto con las de interferón alfa, gamma y de respuesta inflamatoria, en las células de sangre periférica de las muestras de pacientes incluidos en este estudio, podrían integrar una respuesta inflamatoria global alterada, en la que tanto el tejido muscular como las células de sangre periférica pueden estar involucrados. Serían necesarios estudios adicionales que permitieran comprobar esa posibilidad.

En nuestro estudio se han identificado varias rutas adicionales en la que se han caracterizado un número significativo de genes diferencialmente expresados entre los dos grupos de muestras que conforman nuestro estudio. Así, varios genes involucrados en apoptosis han mostrados niveles de expresión elevados en las muestras de los pacientes. En este sentido es de destacar *SAT1*, un gen que es diana de la acción de p53, y con la que se induce un proceso de muerte celular, con un papel primordial de la peroxidación de lípidos de membrana en un proceso dependiente de hierro, llamado ferroptosis (Stockwell et al., 2017). Se ha podido determinar que en procesos de estrés oxidativo, como consecuencia de la presencia de grandes cantidades de peróxido de hidrógeno en el interior de la célula, se produce una permeabilización de la membrana lisosomal lo que provoca la entrada del peróxido de hidrógeno en el lisosoma, donde ocasiona la generación de radicales altamente reactivos ricos en hierro induciéndose así un fenómeno de apoptosis (Kurz et al., 2008).

Además, las enfermedades de depósito lisosomal se caracterizan por una acumulación progresiva de macromoléculas no digeridas dentro de la célula debido a la disfunción lisosomal. Los lisosomas juegan un papel fundamental en la vía autofágica al fusionarse con el autofagosomas y digerir su contenido (Lieberman et al., 2012). Esto, no solo permite proporcionar energía y aminoácidos para mantener la función celular en condiciones de inanición, sino que tiene un papel importante en la limpieza celular, ya que es capaz de eliminar proteínas mal plegadas, agregados proteicos y orgánulos desgastados como mitocondrias, lo que contribuye a la renovación fisiológica de las células (Raben et al., 2012). En varios trabajos se ha descrito la existencia de una autofagia anormal en biopsias de músculo en pacientes con formas tardías de la enfermedad (Fernández et al., 1999; Lewandowska et al., 2008). Este componente autofágico que es tan prominente en las células musculares de pacientes con la forma adulta, es prácticamente insignificante en enfermos con la forma clásica (Raben et al., 2010). Esta afectación de la autofagia detectada en el tejido muscular, no ha sido detectado en el trabajo de Palermo et al. (Palermo et al., 2012) pero tampoco en nuestro análisis de transcriptoma hemos podido detectar rutas significativamente enriquecidas relacionadas con la autofagia y su función anormal. Por otro lado, sí que hemos podido detectar distintas rutas relacionadas con el proteosoma en el análisis con *GSEA*, tanto

utilizando la agrupación de genes de *KEGG* como con la agrupación de *Biocarta*. En ambos casos se encontraban enriquecidas positivamente, y por lo tanto elevados en el fenotipo enfermo frente al sano. De las 10 rutas enriquecidas positivamente con las agrupaciones de factores de transcripción, en 8 de ellas aparece representado el gen *TRIB1* y en 5 de ellas aparece el gen *ZFAND5*, ambos relacionados con la degradación de proteínas vía ubiquitinación. Además dentro de las rutas diferencialmente expresadas obtenidas con *Pathview*, cabe destacar la ruta hsa04141, en la que se describe el procesamiento dentro del retículo endoplásmico de las proteínas mal plegadas y su relación con el proteosoma y la ubiquitinación (Raben et al., 2008).

El sistema ubiquitina-proteosoma y la autofagia son los dos sistemas principales de degradación y reciclaje de proteínas intracelulares. Aunque sus actividades no son interdependientes, estudios recientes muestran que existen conexiones entre ambos sistemas. Parece que la inhibición de uno de los sistemas provoca el aumento del otro con el fin de mantener la homeostasis (Kocaturk and Gozuacik, 2018). Se ha visto que en neuronas la digestión mediante autofagia se produce no solo por la fusión de los autofagosomas con los lisosomas, sino que hay otro vía que consiste en la fusión de los autofagosomas con el endosoma temprano o tardío lo que daría lugar al anfisoma (Ihara et al., 2012). Habría que explorar la posibilidad de que puedan existir diferencias en cuanto a la relación autofagia-proteosoma que puedan ser dependientes de tejido, y si eso puede explicar las diferencias detectadas en cuanto a la desregulación de la autofagia descrito en músculo esquelético frente a los resultados de nuestro estudio, en el que parece que habría una mayor implicación del sistema ubiquitina-proteosoma en la degradación celular.

## CONCLUSIONES



### 6. CONCLUSIONES

1. La implementación del estudio molecular en la enfermedad de Pompe, mediante la incorporación del *MLPA* a la rutina diagnóstica, ha permitido la mejora de la sensibilidad global del estudio, pudiendo identificar 4 deleciones que no hubieran sido detectadas mediante el abordaje clásico de secuenciación Sanger.
2. En este trabajo hemos diagnosticado 76 pacientes en los que se han podido caracterizar las dos variantes patogénicas que conforman el genotipo enfermo.
3. La mutación c.-32-13T>G ha sido encontrada en el 84,3% de los pacientes con la forma tardía de la enfermedad. Nuestra serie no refleja la existencia de otra mutación prevalente, sino que las mutaciones detectadas se encuentran distribuidas a lo largo de toda la estructura del gen.
4. Se han identificado 20 variantes nuevas no descritas previamente, de las cuales 13 son claramente patogénicas debido a su naturaleza.
5. El análisis funcional de procesamiento de mRNA de 16 variantes ha puesto de manifiesto la implicación de 7 de ellas en un *splicing* aberrante.
6. De los 259 pacientes estudiados, 53 pacientes eran portadores de una única mutación y 126 no mostraban cambio alguno que sea considerado patológico. Esto pone de manifiesto la enorme dificultad que existe para hacer una adecuada selección de los pacientes en esta patología y la necesidad de modificar los criterios de selección.
7. A la vista de los resultados de este trabajo, consideramos que el estudio genético debería incluir un análisis genómico de toda la estructura del gen, con la caracterización de posibles reordenamientos. Además, sería recomendable complementarlo con un análisis funcional de RNA, para detectar posibles procesamientos aberrantes.
8. En el análisis de transcriptoma hemos identificado 980 genes diferencialmente expresados, de los cuales 479 estaban *up*-regulados y 501 *down*-regulados en los pacientes frente a los controles.
9. En el análisis de enriquecimiento funcional hemos descrito la desregulación de rutas implicadas en inflamación y respuesta inmune, así como la alteración de rutas vinculadas con apoptosis y con el sistema ubiquitina-proteosoma.

## BIBLIOGRAFÍA

## 7. BIBLIOGRAFÍA

## A

- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. & Sunyaev, S.R. 2010, "A method and server for predicting damaging missense mutations", *Nature methods*, vol. 7, no. 4, pp. 248-249.
- Alejaldre, A., Díaz-Manera, J., Ravaglia, S., Tibaldi, E.C., D'Amore, F., Morís, G., Muelas, N., Vílchez, J.J., García-Medina, A. & Usón, M. 2012, "Trunk muscle involvement in late-onset Pompe disease: study of thirty patients", *Neuromuscular Disorders*, vol. 22, pp. S148-S154.
- Amiñoso, C., Gordillo-Marañón, M., Hernández, J. & Solera, J. 2019, "Reevaluating the pathogenicity of the mutation c. 1194+5G>A in GAA gene by functional analysis of RNA in a 61-year-old woman diagnosed with Pompe disease by muscle biopsy", *Neuromuscular Disorders*, vol. 29, no. 3, pp. 187-191.
- Amiñoso, C., Vallespin, E., Fernández, L., Arrabal, L.F., Desviat, L.R., Pérez, B., Santos, F. & Solera, J. 2013, "Identification of the first deletion–insertion involving the complete structure of GAA gene and part of CCDC40 gene mediated by an Alu element", *Gene*, vol. 519, no. 1, pp. 169-172.
- An, Y., Young, S.P., Hillman, S.L., Van Hove, J.L., Chen, Y. & Millington, D.S. 2000, "Liquid chromatographic assay for a glucose tetrasaccharide, a putative biomarker for the diagnosis of Pompe disease", *Analytical Biochemistry*, vol. 287, no. 1, pp. 136-143.
- Anna, A. & Monika, G. 2018, "Splicing mutations in human genetic disorders: examples, detection, and confirmation", *Journal of Applied Genetics*, vol. 59, no. 3, pp. 253-268.
- Ashida, H., Maeda, Y. & Kinoshita, T. 2006, "DPM1, the catalytic subunit of dolichol-phosphate mannose synthase, is tethered to and stabilized on the endoplasmic reticulum membrane by DPM3", *The Journal of biological chemistry*, vol. 281, no. 2, pp. 896-904.
- Ausems, M., Verbiest, J., Hermans, M., Kroos, M., Beemer, F., Wokke, J., Sandkuijl, L., Reuser, A. & Van der Ploeg, A. 1999, "Frequency of glycogen storage disease type II in The Netherlands: implications for diagnosis and genetic counselling", *European Journal of Human Genetics*, vol. 7, no. 6, pp. 713-716.

## B

- Bali, D.S., Goldstein, J.L., Banugaria, S., Dai, J., Mackey, J., Rehder, C. & Kishnani, P.S. 2012, "Predicting cross-reactive immunological material (CRIM) status in Pompe disease using GAA mutations: Lessons learned from 10 years of clinical laboratory testing experience", *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* Wiley Online Library, pp. 40-49
- Banugaria, S.G., Prater, S.N., Patel, T.T., DeArme, S.M., Milleson, C., Sheets, K.B., Bali, D.S., Rehder, C.W., Raiman, J.A. & Wang, R.A. 2013, "Algorithm for the early diagnosis and treatment of patients with cross reactive immunologic material-negative classic infantile pompe disease: a step towards improving the efficacy of ERT", *PLoS One*, vol. 8, no. 6, pp. e67052.

- Barba-Romero, M.A., Barrot, E., Bautista-Lorite, J., Gutiérrez-Rivas, E., Illa, I., Jiménez, L.M., Ley-Martos, M., López de Munain, A., Pardo, J. & Pascual-Pascual, S.I. 2012, "Guía clínica de la enfermedad de Pompe de inicio tardío", *Rev Neurol*, vol 54, no 8, pp. 497-507.
- Bartsch, G., Mitra, A.P., Mitra, S.A., Almal, A.A., Steven, K.E., Skinner, D.G., Fry, D.W., Lenehan, P.F., Worzel, W.P. & Cote, R.J. 2016, "Use of artificial intelligence and machine learning algorithms with gene expression profiling to predict recurrent non muscle invasive urothelial carcinoma of the bladder", *The Journal of urology*, vol. 195, no. 2, pp. 493-498.
- Becker, J.A., Vlach, J., Raben, N., Nagaraju, K., Adams, E.M., Hermans, M.M., Reuser, A.J., Brooks, S.S., Tifft, C.J., Hirschhorn, R., Huie, M.L., Nicolino, M. & Plotz, P.H. 1998, "The African origin of the common mutation in African American patients with glycogen-storage disease type II", *American Journal of Human Genetics*, vol. 62, no. 4, pp. 991-994.
- Bembi, B., Cerini, E., Danesino, C., Donati, M.A., Gasperini, S., Morandi, L., Musumeci, O., Parenti, G., Ravaglia, S., Seidita, F., Toscano, A. & Vianello, A. 2008, "Diagnosis of glycogenosis type II", *Neurology*, vol. 71, no. 23 Suppl 2, pp. S4-11.
- Benjamini, Y. & Hochberg, Y. 1995, "Controlling the false discovery rate: a practical and powerful approach to multiple testing", *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289-300.
- Bergsma, A.J., In't Groen, S.L., van den Dorpel, Jan JA, van den Hout, Hannerieke JMP, van der Beek, Nadine AME, Schoser, B., Toscano, A., Musumeci, O., Bembi, B. & Dardis, A. 2019, "A genetic modifier of symptom onset in Pompe disease", *EBioMedicine*, vol. 43, pp. 553-561.
- Bergsma, A.J., Kroos, M., Hoogeveen-Westerveld, M., Halley, D., Ploeg, A.T. & Pijnappel, W. 2015, "Identification and Characterization of Aberrant GAA Pre-mRNA Splicing in Pompe Disease Using a Generic Approach", *Human mutation*, vol. 36, no. 1, pp. 57-68.
- Berrier, K.L., Kazi, Z.B., Prater, S.N., Bali, D.S., Goldstein, J., Stefanescu, M.C., Rehder, C.W., Botha, E.G., Ellaway, C. & Bhattacharya, K. 2015, "CRIM-negative infantile Pompe disease: characterization of immune responses in patients treated with ERT monotherapy", *Genetics in Medicine*, vol. 17, no. 11, pp. 912-918.
- Boerkoel, C.F., Exelbert, R., Nicastrì, C., Nichols, R.C., Miller, F.W., Plotz, P.H. & Raben, N. 1995, "Leaky splicing mutation in the acid maltase gene is associated with delayed onset of glycogenosis type II", *American Journal of Human Genetics*, vol. 56, no. 4, pp. 887-897.
- Boon, K., Osorio, E.C., Greenhut, S.F., Schaefer, C.F., Shoemaker, J., Polyak, K., Morin, P.J., Buetow, K.H., Strausberg, R.L., De Souza, S.J. & Riggins, G.J. 2002, "An anatomy of normal and malignant gene expression", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 17, pp. 11287-11292.
- Boustany, R.N. 2013, "Lysosomal storage diseases—the horizon expands", *Nature Reviews Neurology*, vol. 9, no. 10, pp. 583-598.
- Braulke, T. & Bonifacino, J.S. 2009, "Sorting of lysosomal proteins", *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, vol. 1793, no. 4, pp. 605-614.

Brown, W.J., Goodhouse, J. & Farquhar, M.G. 1986, "Mannose-6-phosphate receptors for lysosomal enzymes cycle between the Golgi complex and endosomes", *The Journal of cell biology*, vol. 103, no. 4, pp. 1235-1247.

---

**C**

---

Carmel, I., Tal, S., Vig, I. & Ast, G. 2004, "Comparative analysis detects dependencies among the 5' splice-site positions", *RNA (New York, N.Y.)*, vol. 10, no. 5, pp. 828-840.

Chen, X., Liu, T., Huang, M., Wu, J., Zhu, J., Guo, Y., Xu, X., Li, F., Wang, J. & Fu, L. 2017, "Clinical and Molecular Characterization of Infantile-Onset Pompe Disease in Mainland Chinese Patients: Identification of Two Common Mutations", *Genetic testing and molecular biomarkers*, vol. 21, no. 6, pp. 391-396.

Corti, M., Liberati, C., Smith, B.K., Lawson, L.A., Tuna, I.S., Conlon, T.J., Erger, K.E., Islam, S., Herzog, R. & Fuller, D.D. 2017, "Safety of Intradiaphragmatic Delivery of Adeno-Associated Virus-Mediated Alpha-Glucosidase (rAAV1-CMV-hGAA) Gene Therapy in children affected by Pompe Disease", *Human Gene Therapy*, vol. 28, no. 4, pp 208-218.

Coutinho, M.F., Prata, M.J. & Alves, S. 2012, "Mannose-6-phosphate pathway: a review on its role in lysosomal function and dysfunction", *Molecular genetics and metabolism*, vol. 105, no. 4, pp. 542-550.

---

**D**

---

Dardis, A., Zanin, I., Zampieri, S., Stuani, C., Pianta, A., Romanello, M., Baralle, F.E., Bembi, B. & Buratti, E. 2013, "Functional characterization of the common c.-32-13T>G mutation of GAA gene: identification of potential therapeutic agents", *Nucleic acids research*, vol. 42, no. 2, pp. 1291-1302.

De Duve, C., Pressman, B.C., Gianetto, R., Wattiaux, R. & Appelmans, F. 1955, "Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue", *The Biochemical journal*, vol. 60, no. 4, pp. 604-617.

Deming, D., Lee, K., McSherry, T., Wei, R.R., Edmunds, T. & Garman, S.C. 2017, "The molecular basis for Pompe disease revealed by the structure of human acid  $\alpha$ -glucosidase", *bioRxiv*, pp. 212837.

Desviat, L., Vallespin, E., Nevado, J., Crespo, C., Santos, F., Rodriguez-Pombo, P., Merinero, B., Perez-Cerda, C., Lapunzina, P. & Ugarte, M. 2011, "Metabolarray: targeted array-CGH for the detection of deletions/duplications in 205 genes involved in inherited metabolic disease", *journal of inherited metabolic diseases*, Springer van godewijkstraat 30, 3311 gz dordrecht, netherlands, , pp. S237.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T.R. 2013, "STAR: ultrafast universal RNA-seq aligner", *Bioinformatics*, vol. 29, no. 1, pp. 15-21.

Douillard-Guilloux, G., Raben, N., Takikita, S., Batista, L., Caillaud, C. & Richard, E. 2008, "Modulation of glycogen synthesis by RNA interference: towards a new therapeutic approach for glycogenosis type II", *Human molecular genetics*, vol. 17, no. 24, pp. 3876-3886.

Dündar, F., Skrabanek, L. & Zumbo, P. 2015, "Introduction to differential gene expression analysis using RNA-seq", *Applied Bioinformatics Core/Weill Cornell Medical College*, pp. 1-67.

### E

---

- Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. 1998, "Cluster analysis and display of genome-wide expression patterns", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 25, pp. 14863-14868.
- Elder, M.E., Nayak, S., Collins, S.W., Lawson, L.A., Kelley, J.S., Herzog, R.W., Modica, R.F., Lew, J., Lawrence, R.M. & Byrne, B.J. 2013, "B-Cell depletion and immunomodulation before initiation of enzyme replacement therapy blocks the immune response to acid alpha-glucosidase in infantile-onset Pompe disease", *The Journal of pediatrics*, vol. 163, no. 3, pp. 847-854.
- Eskelinen, E. & Saftig, P. 2009, "Autophagy: a lysosomal degradation pathway with a central role in health and disease", *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, vol. 1793, no. 4, pp. 664-673.
- Fernández, R., Fernández, J.M., Cervera, C., Teijeira, S., Teijeiro, A., Domínguez, C. & Navarro, C. 1999, "Adult glycogenosis II with paracrystalline mitochondrial inclusions and Hirano bodies in skeletal muscle", *Neuromuscular Disorders*, vol. 9, no. 3, pp. 136-143.

### F

---

- Futerman, A.H. & Van Meer, G. 2004, "The cell biology of lysosomal storage disorders", *Nature reviews Molecular cell biology*, vol. 5, no. 7, pp. 554-565.

### G

---

- Gatto, F., Rossi, B., Tarallo, A., Polishchuk, E., Polishchuk, R., Carrella, A., Nusco, E., Alvino, F.G., Iacobellis, F. & Leonibus, E. 2017, "AAV-mediated transcription factor EB (TFEB) gene delivery ameliorates muscle pathology and function in the murine model of Pompe Disease", *Scientific reports*, vol. 7, no. 1, pp. 15089.
- Gort, L., Coll, M.J. & Chabás, A. 2007, "Glycogen storage disease type II in Spanish patients: high frequency of c. 1076-1G> C mutation", *Molecular genetics and metabolism*, vol. 92, no. 1-2, pp. 183-187.
- GTEx Consortium 2015, "Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans", *Science (New York, N.Y.)*, vol. 348, no. 6235, pp. 648-660.

### H

---

- Hagemans, M.L., Stigter, R.L., van Capelle, C.I., van der Beek, Nadine AME, Winkel, L.P., van Vliet, L., Hop, W.C., Reuser, A.J., Beishuizen, A. & van der Ploeg, Ans T 2010, "PAS-positive lymphocyte vacuoles can be used as diagnostic screening test for Pompe disease", *Journal of inherited metabolic disease*, vol. 33, no. 2, pp. 133-139.
- Hagemans, M.L., Winkel, L.P., Hop, W.C., Reuser, A.J., Van Doorn, P.A. & Van der Ploeg, A.T. 2005, "Disease severity in children and adults with Pompe disease related to age and disease duration", *Neurology*, vol. 64, no. 12, pp. 2139-2141.

- Hecht, M., Bromberg, Y. & Rost, B. 2015, "Better prediction of functional effects for sequence variants", *BMC genomics*, vol. 16, no. 8, pp. S1.
- Hermans, M.M., Leenen, D.v., Kroos, M.A., Beesley, C.E., Van der Ploeg, Ans T, Sakuraba, H., Wevers, R., Kleijer, W., Michelakakis, H. & Kirk, E.P. 2004, "Twenty-two novel mutations in the lysosomal  $\alpha$ -glucosidase gene (GAA) underscore the genotype–phenotype correlation in glycogen storage disease type II", *Human mutation*, vol. 23, no. 1, pp. 47-56.
- Hermans, M.M., Kroos, M.A., van Beeumen, J., Oostra, B.A. & Reuser, A.J. 1991, "Human lysosomal alpha-glucosidase. Characterization of the catalytic site", *The Journal of biological chemistry*, vol. 266, no. 21, pp. 13507-13512.
- Hers, H.G. 1963, "Alpha-Glucosidase deficiency in generalized glycogen storage disease (Pompe's disease)", *The Biochemical journal*, vol. 86, pp. 11-16.
- Herzog, A., Hartung, R., Reuser, A.J., Hermanns, P., Runz, H., Karabul, N., Gökce, S., Pohlenz, J., Kampmann, C. & Lampe, C. 2012, "A cross-sectional single-centre study on the spectrum of Pompe disease, German patients: molecular analysis of the GAA gene, manifestation and genotype-phenotype correlations", *Orphanet journal of rare diseases*, vol. 7, no. 1, pp. 35.
- Hiraoka, M., Okamoto, K., Ohguro, H. & Abe, A. 2013, "Role of N-glycosylation of human lysosomal phospholipase A2 for the formation of catalytically active enzyme", *Journal of lipid research*, vol. 54, no. 11, pp. 3098-3105.
- Hoefsloot, L.H., Hoogeveen-Westerveld, M., Kroos, M., Van Beeumen, J., Reuser, A.J. & Oostra, B. 1988, "Primary structure and processing of lysosomal alpha-glucosidase; homology with the intestinal sucrase-isomaltase complex." *The EMBO journal*, vol. 7, no. 6, pp. 1697-1704.
- Hoefsloot, L.H., Hoogeveen-Westerveld, M., Reuser, A.J. & Oostra, B.A. 1990, "Characterization of the human lysosomal alpha-glucosidase gene", *The Biochemical journal*, vol. 272, no. 2, pp. 493-497.
- Hrdlickova, R., Toloue, M. & Tian, B. 2017, "RNA-Seq methods for transcriptome analysis", *Wiley Interdisciplinary Reviews: RNA*, vol. 8, no. 1, pp. e1364.
- Huie, M., Shanske, A., Kasper, J., Marion, R.W. & Hirschhorn, R. 1999, "A large Alu-mediated deletion, identified by PCR, as the molecular basis for glycogen storage disease type II (GSDII)", *Human genetics*, vol. 104, no. 1, pp. 94-98.
- Hule, M., Chen, A., Tsujino, S., Shanske, S., DiMauro, S., Engel, A.G. & Hirschhorn, R. 1994, "Aberrant splicing in adult onset glycogen storage disease type II (GSDII): molecular identification of an IVS1 (–13T→G) mutation in a majority of patients and a novel IVS10 (1GT→CT) mutation", *Human molecular genetics*, vol. 3, no. 12, pp. 2231-2236.
- 
- Ihara, Y., Morishima-Kawashima, M. & Nixon, R. 2012, "The ubiquitin-proteasome system and the autophagic-lysosomal system in Alzheimer disease", *Cold Spring Harbor perspectives in medicine*, vol. 2, no. 8, a006361.

Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E. & Karyadi, D. 2016, "REVEL: an ensemble method for predicting the pathogenicity of rare missense variants", *The American Journal of Human Genetics*, vol. 99, no. 4, pp. 877-885.

## J

Jackman, R.W. & Kandarian, S.C. 2004, "The molecular basis of skeletal muscle atrophy", *American Journal of Physiology-Cell Physiology*, vol. 287, no. 4, pp. C834-C843.

## K

Khan, J., Wei, J.S., Ringner, M., Saal, L.H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C.R. & Peterson, C. 2001, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks", *Nature medicine*, vol. 7, no. 6, pp. 673-679.

Kishnani, P.S., Amartino, H.M., Lindberg, C., Miller, T.M., Wilson, A. & Keutzer, J. 2013, "Timing of diagnosis of patients with Pompe disease: data from the Pompe registry", *American Journal of Medical Genetics Part A*, vol. 161, no. 10, pp. 2431-2443.

Kishnani, P.S., Hwu, W., Mandel, H., Nicolino, M., Yong, F., Corzo, D. & Infantile-Onset Pompe Disease Natural History Study Group 2006, "A retrospective, multinational, multicenter study on the natural history of infantile-onset Pompe disease", *The Journal of pediatrics*, vol. 148, no. 5, pp. 671-676.

Kishnani, P.S., Steiner, R.D., Bali, D., Berger, K., Byrne, B.J., Case, L.E., Crowley, J.F., Downs, S., Howell, R.R. & Kravitz, R.M. 2006, "Pompe disease diagnosis and management guideline", *Genetics in Medicine*, vol. 8, no. 5, pp. 267-288.

Kishnani, P.S., Nicolino, M., Voit, T., Rogers, R.C., Tsai, A.C., Waterson, J., Herman, G.E., Amalfitano, A., Thurberg, B.L. & Richards, S. 2006, "Chinese hamster ovary cell-derived recombinant human acid  $\alpha$ -glucosidase in infantile-onset Pompe disease", *The Journal of pediatrics*, vol. 149, no. 1, pp. 89-97.

Kocaturk, N.M. & Gozuacik, D. 2018, "Crosstalk between Autophagy and the Ubiquitin-Proteasome System", *Frontiers in Cell and Developmental Biology*, vol. 6, pp. 128.

Koeberl, D., Kishnani, P. & Chen, Y. 2007, "Glycogen storage disease types I and II: treatment updates", *Journal of inherited metabolic disease*, vol. 30, no. 2, pp. 159-164.

Kroos, M., Hoogeveen-Westerveld, M., Michelakakis, H., Pomponio, R., Van der Ploeg, A., Halley, D., Reuser, A., Augoustides-Savvopoulou, P., Ausems, M. & Llona, J.B. 2012, "Update of the Pompe disease mutation database with 60 novel GAA sequence variants and additional studies on the functional effect of 34 previously reported variants", *Human mutation*, vol. 33, no. 8, pp. 1161-1165.

Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A. & Reuser, A.J. 2012, "The genotype-phenotype correlation in Pompe disease", *American Journal of Medical Genetics Part C: Seminars in Medical Genetics Wiley Online Library*, vol. 160C, no. 1, pp. 59-68.

Kroos, M.A., Pomponio, R.J., Hagemans, M.L., Keulemans, J.L., Phipps, M., DeRiso, M., Palmer, R.E., Ausems, M.G., Van der Beek, N.A., Van Diggelen, O.P., Halley, D.J., Van der Ploeg, A.T. & Reuser, A.J. 2007, "Broad spectrum of Pompe disease in patients with the same c.-32-13T->G haplotype", *Neurology*, vol. 68, no. 2, pp. 110-115.



Kroos, M.A., Van der Kraan, M., Van Diggelen, O.P., Kleijer, W.J., Reuser, A.J., Van den Boogaard, M.J., Ausems, M.G., Ploos van Amstel, H.K., Poenaru, L. & Nicolino, M. 1995, "Glycogen storage disease type II: frequency of three common mutant alleles and their associated clinical phenotypes studied in 121 patients", *Journal of medical genetics*, vol. 32, no. 10, pp. 836-837.

Kumar, P., Henikoff, S. & Ng, P.C. 2009, "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm", *Nature protocols*, vol. 4, no. 7, pp. 1073-1081.

Kurosaki, T. & Maquat, L.E. 2016, "Nonsense-mediated mRNA decay in humans at a glance", *Journal of cell science*, vol. 129, no. 3, pp. 461-467.

Kurz, T., Terman, A., Gustafsson, B. & Brunk, U.T. 2008, "Lysosomes and oxidative stress in aging and apoptosis", *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1780, no. 11, pp. 1291-1303.

---

**L**

---

Lamb, C.A., Yoshimori, T. & Tooze, S.A. 2013, "The autophagosome: origins unknown, biogenesis complex", *Nature reviews Molecular cell biology*, vol. 14, no. 12, pp. 759-774.

Lawrence, R.E. & Zoncu, R. 2019, "The lysosome as a cellular centre for signalling, metabolism and quality control", *Nat. Cell Biol.*, vol. 21, pp. 133-142.

Lévesque, S., Auray-Blais, C., Gravel, E., Boutin, M., Dempsey-Nunez, L., Jacques, P., Chenier, S., Larue, S., Rioux, M. & Al-Hertani, W. 2016, "Diagnosis of late-onset Pompe disease and other muscle disorders by next-generation sequencing", *Orphanet journal of rare diseases*, vol. 11, no. 1, pp. 8.

Levine, B. & Kroemer, G. 2008, "Autophagy in the pathogenesis of disease", *Cell*, vol. 132, no. 1, pp. 27-42.

Lewandowska, E., Wierzba-Bobrowicz, T., Rola, R., Modzelewska, J., Stepien, T., Lugowska, A., Pasennik, E. & Ryglewicz, D. 2008, "Pathology of skeletal muscle cells in adult-onset glycogenosis type II (Pompe disease): ultrastructural study", *Folia neuropathologica*, vol. 46, no. 2, pp. 123-133.

Li, Q. & Wang, K. 2017, "InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines", *The American Journal of Human Genetics*, vol. 100, no. 2, pp. 267-280.

Li, Y., Foster, W., Deasy, B.M., Chan, Y., Prisk, V., Tang, Y., Cummins, J. & Huard, J. 2004, "Transforming growth factor- $\beta$ 1 induces the differentiation of myogenic cells into fibrotic cells in injured skeletal muscle: a key event in muscle fibrogenesis", *The American journal of pathology*, vol. 164, no. 3, pp. 1007-1019.

Liao, H., Chiang, C., Niu, D., Wang, C., Kao, S., Tsai, F., Huang, Y., Liu, H., Huang, C. & Gao, H. 2014, "Detecting multiple lysosomal storage diseases by tandem mass spectrometry—a national newborn screening program in Taiwan", *Clinica Chimica Acta*, vol. 431, pp. 80-86.

Liao, Y., Smyth, G.K. & Shi, W. 2019, "The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads", *Nucleic acids research*, vol. 47, no. 8, pp. e47-e47.

- Lieberman, A.P., Puertollano, R., Raben, N., Slaugenhaupt, S., Walkley, S.U. & Ballabio, A. 2012, "Autophagy in lysosomal storage disorders", *Autophagy*, vol. 8, no. 5, pp. 719-730.
- Lim, C.Y. & Zoncu, R. 2016, "The lysosome as a command-and-control center for cellular metabolism", *The Journal of cell biology*, vol. 214, no. 6, pp. 653-664.
- Livak, K.J. & Schmittgen, T.D. 2001, "Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method", *methods*, vol. 25, no. 4, pp. 402-408.
- Lüllmann-Rauch, R. 2005, "History and morphology of the lysosome" in *Lysosomes* Springer, pp. 1-16.
- Luo, W. & Brouwer, C. 2013, "Pathview: an R/Bioconductor package for pathway-based data integration and visualization", *Bioinformatics*, vol. 29, no. 14, pp. 1830-1831.
- Luzio, J.P., Parkinson, M.D., Gray, S.R. & Bright, N.A. 2009, "The delivery of endocytosed cargo to lysosomes", *Biochemical Society transactions*, vol. 37, no. 5, pp. 1019-1021.

---

**M**

---

- Maimaiti, M., Takahashi, S., Okajima, K., Suzuki, N., Ohinata, J., Araki, A., Tanaka, H., Mukai, T. & Fujieda, K. 2009, "Silent exonic mutation in the acid- $\alpha$ -glucosidase gene that causes glycogen storage disease type II by affecting mRNA splicing", *Journal of human genetics*, vol. 54, no. 8, pp. 493-496.
- Martina, J.A., Chen, Y., Gucek, M. & Puertollano, R. 2012, "MTORC1 functions as a transcriptional regulator of autophagy by preventing nuclear transport of TFEB", *Autophagy*, vol. 8, no. 6, pp. 903-914.
- Martiniuk, F., Bodkin, M., Tzall, S. & Hirschhorn, R. 1991, "Isolation and partial characterization of the structural gene for human acid alpha glucosidase", *DNA and cell biology*, vol. 10, no. 4, pp. 283-292.
- Martiniuk, F., Chen, A., Mack, A., Arvanitopoulos, E., Chen, Y., Rom, W.N., Codd, W.J., Hanna, B., Alcabes, P. & Raben, N. 1998, "Carrier frequency for glycogen storage disease type II in New York and estimates of affected individuals born with the disease", *American Journal of Medical Genetics*, vol. 79, no. 1, pp. 69-72.
- Martiniuk, F., Mehler, M., Pellicer, A., Tzall, S., La Badie, G., Hobart, C., Ellenbogen, A. & Hirschhorn, R. 1986, "Isolation of a cDNA for human acid alpha-glucosidase and detection of genetic heterogeneity for mRNA in three alpha-glucosidase-deficient patients", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 24, pp. 9641-9644.
- McCarthy, D.J., Chen, Y. & Smyth, G.K. 2012, "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation", *Nucleic acids research*, vol. 40, no. 10, pp. 4288-4297.
- McCready, M., Carson, N., Chakraborty, P., Clarke, J., Callahan, J., Skomorowski, M., Chan, A., Bamforth, F., Casey, R. & Rupar, C. 2007, "Development of a clinical assay for detection of GAA mutations and characterization of the GAA mutation spectrum in a Canadian cohort of individuals with glycogen storage disease, type II", *Molecular genetics and metabolism*, vol. 92, no. 4, pp. 325-335.
- Mindell, J.A. 2012, "Lysosomal acidification mechanisms", *Annual Review of Physiology*, vol. 74, pp. 69-86.

- Montagnese, F., Barca, E., Musumeci, O., Mondello, S., Migliorato, A., Ciranni, A., Rodolico, C., De Filippi, P., Danesino, C. & Toscano, A. 2015, "Clinical and molecular aspects of 30 patients with late-onset Pompe disease (LOPD): unusual features and response to treatment", *Journal of neurology*, vol. 262, no. 4, pp. 968-978.
- Montalvo, A., Bembi, B., Donnarumma, M., Filocamo, M., Parenti, G., Rossi, M., Merlini, L., Buratti, E., De Filippi, P. & Dardis, A. 2006, "Mutation profile of the GAA gene in 40 Italian patients with late onset glycogen storage disease type II", *Human mutation*, vol. 27, no. 10, pp. 999-1006.
- Moreland, R.J., Jin, X., Zhang, X.K., Decker, R.W., Albee, K.L., Lee, K.L., Cauthron, R.D., Brewer, K., Edmunds, T. & Canfield, W.M. 2005, "Lysosomal acid alpha-glucosidase consists of four different peptides processed from a single chain precursor", *The Journal of biological chemistry*, vol. 280, no. 8, pp. 6780-6791.
- Musumeci, O., Thieme, A., Claeys, K.G., Wenninger, S., Kley, R.A., Kuhn, M., Lukacs, Z., Deschauer, M., Gaeta, M. & Toscano, A. 2015, "Homozygosity for the common GAA gene splice site mutation c.-32-13T> G in Pompe disease is associated with the classical adult phenotypical spectrum", *Neuromuscular Disorders*, vol. 25, no. 9, pp. 719-724.

## N

---

- Niño, M.Y., in't Groen, S.L., Bergsma, A.J., van der Beek, Nadine AME, Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, Ans T & Pijnappel, W.P. 2019, "Extension of the Pompe mutation database by linking disease-associated variants to clinical severity", *Human mutation*, pp. 1-14.

## P

---

- Palermo, A., Palmer, R., So, K., Oba-Shinjo, S., Zhang, M., Richards, B., Madhiwalla, S., Finn, P., Hasegawa, A. & Ciociola, K. 2012, "Transcriptional response to GAA deficiency (Pompe disease) in infantile-onset patients", *Molecular genetics and metabolism*, vol. 106, no. 3, pp. 287-300.
- Park, H., Kim, J., Lee, Y.K., Kim, W., You, S.K., Do, J., Jang, Y., Oh, D., Kim, J.I. & Kim, H.H. 2018, "Four unreported types of glycans containing mannose-6-phosphate are heterogeneously attached at three sites (including newly found Asn 233) to recombinant human acid alpha-glucosidase that is the only approved treatment for Pompe disease", *Biochemical and biophysical research communications*, vol. 495, no. 4, pp. 2418-2424.
- Patel, A.A. & Steitz, J.A. 2003, "Splicing double: insights from the second spliceosome", *Nature reviews Molecular cell biology*, vol. 4, no. 12, pp. 960-970.
- Pittis, M., Donnarumma, M., Montalvo, A., Dominissini, S., Kroos, M., Rosano, C., Stroppiano, M., Bianco, M., Donati, M. & Parenti, G. 2008, "Molecular and functional characterization of eight novel GAA mutations in Italian infants with Pompe disease", *Human mutation*, vol. 29, no. 6, pp. E27-E36.
- Platt, F.M., d'Azzo, A., Davidson, B.L., Neufeld, E.F. & Tiffit, C.J. 2018, "Lysosomal storage diseases", *Nature Reviews Disease Primers*, vol. 4, no. 1, pp. 27.
- Platt, F.M., Boland, B. & van der Spoel, A.C. 2012, "The cell biology of disease: lysosomal storage disorders: the cellular impact of lysosomal dysfunction", *The Journal of cell biology*, vol. 199, no. 5, pp. 723-734.

Pompe, J. 1932, "Over idiopathische hypertrophie van het hart", *Ned Tijdschr Geneeskd*, vol. 76, pp. 304-311.

Porto, C., Cardone, M., Fontana, F., Rossi, B., Tuzzi, M.R., Tarallo, A., Barone, M.V., Andria, G. & Parenti, G. 2009, "The pharmacological chaperone N-butyldeoxynojirimycin enhances enzyme replacement therapy in Pompe disease fibroblasts", *Molecular Therapy*, vol. 17, no. 6, pp. 964-971.

## R

Raben, N., Hill, V., Shea, L., Takikita, S., Baum, R., Mizushima, N., Ralston, E. & Plotz, P. 2008, "Suppression of autophagy in skeletal muscle uncovers the accumulation of ubiquitinated proteins and their potential role in muscle damage in Pompe disease", *Human molecular genetics*, vol. 17, no. 24, pp. 3897-3908.

Raben, N., Ralston, E., Chien, Y., Baum, R., Schreiner, C., Hwu, W., Zaal, K.J. & Plotz, P.H. 2010, "Differences in the predominance of lysosomal and autophagic pathologies between infants and adults with Pompe disease: implications for therapy", *Molecular genetics and metabolism*, vol. 101, no. 4, pp. 324-331.

Raben, N., Wong, A., Ralston, E. & Myerowitz, R. 2012, "Autophagy and mitochondria in Pompe disease: nothing is so new as what has long been forgotten", *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* Wiley Online Library, vol. 160C, pp. 13-21.

Reggiori, F. & Ungermann, C. 2017, "Autophagosome maturation and fusion", *Journal of Molecular Biology*, vol. 429, no. 4, pp. 486-496.

Reuser, A.J., Koster, J.F., Hoogeveen, A. & Galjaard, H. 1978, "Biochemical, immunological, and cell genetic studies in glycogenosis type II", *American Journal of Human Genetics*, vol. 30, no. 2, pp. 132-143.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E. & Spector, E. 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*, vol. 17, no. 5, pp. 405-424.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. & Smyth, G.K. 2015, "limma powers differential expression analyses for RNA-sequencing and microarray studies", *Nucleic acids research*, vol. 43, no. 7, pp. e47-e47.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010, "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data", *Bioinformatics*, vol. 26, no. 1, pp. 139-140.

Robinson, M.D. & Oshlack, A. 2010, "A scaling normalization method for differential expression analysis of RNA-seq data", *Genome biology*, vol. 11, no. 3, pp. R25.

Roczniak-Ferguson, A., Petit, C.S., Froehlich, F., Qian, S., Ky, J., Angarola, B., Walther, T.C. & Ferguson, S.M. 2012, "The transcription factor TFEB links mTORC1 signaling to transcriptional control of lysosome homeostasis", *Science signaling*, vol. 5, no. 228, pp. ra42.

Rodrigues, C.H., Pires, D.E. & Ascher, D.B. 2018, "DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability", *Nucleic acids research*, vol. 46, no. W1, pp. W350-W355.

Russell, R.C., Yuan, H. & Guan, K. 2014, "Autophagy regulation by nutrient signaling", *Cell research*, vol. 24, no. 1, pp. 42-57.

## S

Saftig, P. & Klumperman, J. 2009, "Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function", *Nature reviews Molecular cell biology*, vol. 10, no. 9, pp. 623-635.

Saftig, P., Schroder, B. & Blanz, J. 2010, "Lysosomal membrane proteins: life between acid and neutral conditions", *Biochemical Society transactions*, vol. 38, no. 6, pp. 1420-1423.

Sardiello, M., Palmieri, M., di Ronza, A., Medina, D.L., Valenza, M., Gennarino, V.A., Di Malta, C., Donaudy, F., Embrione, V., Polishchuk, R.S., Banfi, S., Parenti, G., Cattaneo, E. & Ballabio, A. 2009, "A gene network regulating lysosomal biogenesis and function", *Science (New York, N.Y.)*, vol. 325, no. 5939, pp. 473-477.

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwiijnenburg, D., Diepvens, F. & Pals, G. 2002, "Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification", *Nucleic acids research*, vol. 30, no. 12, pp. e57-e57.

Schwake, M., Schröder, B. & Saftig, P. 2013, "Lysosomal membrane proteins and their central role in physiology", *Traffic*, vol. 14, no. 7, pp. 739-748.

Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. 2014, "MutationTaster2: mutation prediction for the deep-sequencing age", *Nature methods*, vol. 11, no. 4, pp. 361-362.

Seranova, E., Connolly, K.J., Zatyka, M., Rosenstock, T.R., Barrett, T., Tuxworth, R.I. & Sarkar, S. 2017, "Dysregulation of autophagy as a common mechanism in lysosomal storage diseases", *Essays in biochemistry*, vol. 61, no. 6, pp. 733-749.

Settembre, C., Di Malta, C., Polito, V.A., Garcia Arencibia, M., Vetrini, F., Erdin, S., Erdin, S.U., Huynh, T., Medina, D., Colella, P., Sardiello, M., Rubinsztein, D.C. & Ballabio, A. 2011, "TFEB links autophagy to lysosomal biogenesis", *Science (New York, N.Y.)*, vol. 332, no. 6036, pp. 1429-1433.

Shepard, P.J., Choi, E., Busch, A. & Hertel, K.J. 2011, "Efficient internal exon recognition depends on near equal contributions from the 3' and 5' splice sites", *Nucleic acids research*, vol. 39, no. 20, pp. 8928-8937.

Smith, B.K., Collins, S.W., Conlon, T.J., Mah, C.S., Lawson, L.A., Martin, A.D., Fuller, D.D., Cleaver, B.D., Clément, N. & Phillips, D. 2013, "Phase I/II trial of adeno-associated virus-mediated alpha-glucosidase gene therapy to the diaphragm for chronic respiratory failure in Pompe disease: Initial safety and ventilatory outcomes", *Human Gene Therapy*, vol. 24, no. 6, pp. 630-640.

Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frébourg, T., Tosi, M. & Martins, A. 2016, "Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools", *PLoS genetics*, vol. 12, no. 1, pp. e1005756.

Spampanato, C., Feeney, E., Li, L., Cardone, M., Lim, J.A., Annunziata, F., Zare, H., Polishchuk, R., Puertollano, R., Parenti, G., Ballabio, A. & Raben, N. 2013, "Transcription factor EB (TFEB) is a new therapeutic target for Pompe disease", *EMBO molecular medicine*, vol. 5, no. 5, pp. 691-706.

## BIBLIOGRAFÍA

---

Sterne-Weiler, T., Howard, J., Mort, M., Cooper, D.N. & Sanford, J.R. 2011, "Loss of exon identity is a common mechanism of human inherited disease", *Genome research*, vol. 21, no. 10, pp. 1563-1571.

Stockwell, B.R., Angeli, J.P.F., Bayir, H., Bush, A.I., Conrad, M., Dixon, S.J., Fulda, S., Gascón, S., Hatzios, S.K. & Kagan, V.E. 2017, "Ferroptosis: a regulated cell death nexus linking metabolism, redox biology, and disease", *Cell*, vol. 171, no. 2, pp. 273-285.

Stuart, L.M. & Ezekowitz, R.A.B. 2005, "Phagocytosis: elegant complexity", *Immunity*, vol. 22, no. 5, pp. 539-550.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. & Mesirov, J.P. 2005, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545-15550.

## T

---

Tsuburaya, R.S., Monma, K., Oya, Y., Nakayama, T., Fukuda, T., Sugie, H., Hayashi, Y.K., Nonaka, I. & Nishino, I. 2012, "Acid phosphatase-positive globular inclusions is a good diagnostic marker for two patients with adult-onset Pompe disease lacking disease specific pathology", *Neuromuscular Disorders*, vol. 22, no. 5, pp. 389-393.

Tsujino, S., Huie, M., Kanazawa, N., Sugie, H., Goto, Y., Kawai, M., Nonaka, I., Hirschhorn, R. & Sakuragawa, N. 2000, "Frequent mutations in Japanese patients with acid maltase deficiency", *Neuromuscular Disorders*, vol. 10, no. 8, pp. 599-603.

## V

---

Van der Ploeg, Ans T, Clemens, P.R., Corzo, D., Escolar, D.M., Florence, J., Groeneveld, G.J., Herson, S., Kishnani, P.S., Laforet, P. & Lake, S.L. 2010, "A randomized study of alglucosidase alfa in late-onset Pompe's disease", *New England Journal of Medicine*, vol. 362, no. 15, pp. 1396-1406.

van der Ploeg, Ans T & Reuser, A.J. 2008, "Pompe's disease", *The lancet*, vol. 372, no. 9646, pp. 1342-1353.

Vanderkraan, M., Kroos, M., Joosse, M., Bijvoet, A., Verbeet, M.P., Kleijer, W. & Reuser, A. 1994, "Deletion of exon 18 is a frequent mutation in glycogen storage disease type II", *Biochemical and biophysical research communications*, vol. 203, no. 3, pp. 1535-1541.

## W

---

Wens, S; Van Doeveren, T; Lequin, M; Van Gelder, C;Verdijk, R; Van der Hout, H; Van Doorn, P; Van der Ploeg, A; De Co, R. 2015, "Muscle MRI in Classic Infantile Pompe Disease", *J Rare Dis Diagn Ther*, Vol. 1, No. 1:10, pp.1-4.

Wilkerson, M.D. & Hayes, D.N. 2010, "ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking", *Bioinformatics*, vol. 26, no. 12, pp. 1572-1573.

## BIBLIOGRAFÍA

---

- Winchester, B., Bali, D., Bodamer, O., Caillaud, C., Christensen, E., Cooper, A., Cupler, E., Deschauer, M., Fumić, K. & Jackson, M. 2008, "Methods for a prompt and reliable laboratory diagnosis of Pompe disease: report from an international consensus meeting", *Molecular genetics and metabolism*, vol. 93, no. 3, pp. 275-281.
- Wokke, J.H., Ausems, M.G., van den Boogaard, Marie-José H, Ippel, E.F., van Diggelen, O., Kroos, M.A., Boer, M., Jennekens, F.G., Reuser, A.J. & van Amstel, Hans Kristian Ploos 1995, "Genotype-phenotype correlation in adult-onset acid maltase deficiency", *Annals of Neurology*, vol. 38, no. 3, pp. 450-454.
- Wujek, P., Kida, E., Walus, M., Wisniewski, K.E. & Golabek, A.A. 2004, "N-glycosylation is crucial for folding, trafficking, and stability of human tripeptidyl-peptidase I", *The Journal of biological chemistry*, vol. 279, no. 13, pp. 12827-12839.

## X

---

- Xu, H. & Ren, D. 2015, "Lysosomal physiology", *Annual Review of Physiology*, vol. 77, pp. 57-80.

## PUBLICACIONES



## 8. PUBLICACIONES

### 8.1. ARTÍCULO 1: Identification of the first deletion–insertion involving the complete structure of GAA gene and part of CCDC40 gene mediated by an Alu element.

Amiñoso, C., Vallespin, E., Fernández, L., Arrabal, L.F., Desviat, L.R., Pérez, B., Santos, F. & Solera, J. 2013, "Identification of the first deletion–insertion involving the complete structure of GAA gene and part of CCDC40 gene mediated by an Alu element", *Gene*, vol. 519, no. 1, pp. 169-172.

### 8.2. ARTÍCULO 2: Reevaluating the pathogenicity of the mutation c. 1194+5G>A in GAA gene by functional analysis of RNA in a 61-year-old woman diagnosed with Pompe disease by muscle biopsy.

Amiñoso, C., Gordillo-Marañón, M., Hernández, J. & Solera, J. 2019, "Reevaluating the pathogenicity of the mutation c. 1194+5G>A in GAA gene by functional analysis of RNA in a 61-year-old woman diagnosed with Pompe disease by muscle biopsy", *Neuromuscular Disorders*, vol. 29, no. 3, pp. 187-191.