

UAM

Universidad Autónoma
de Madrid



Programa de doctorado en Biociencias Moleculares

**IDENTIFICACIÓN MOLECULAR DE ENFERMEDADES
OXPHOS: APROXIMACIONES DIAGNÓSTICAS Y
BIOINFORMÁTICAS MEDIANTE SECUENCIACIÓN MASIVA**

Tesis Doctoral

Aitor Delmiro Magdalena

Madrid, 2019

Departamento de Bioquímica
FACULTAD DE MEDICINA



Universidad Autónoma
de Madrid

**IDENTIFICACIÓN MOLECULAR DE ENFERMEDADES
OXPHOS: APROXIMACIONES DIAGNÓSTICAS Y
BIOINFORMÁTICAS MEDIANTE SECUENCIACIÓN MASIVA**

Memoria presentada por el Licenciado en Biología y Bioquímica
Aitor Delmiro Magdalena
para optar al grado de Doctor en Biociencias Moleculares por la
Universidad Autónoma de Madrid

Director de Tesis:
Dr. Miguel Ángel Martín Casanueva

**Instituto de Investigación Sanitaria i+12
Hospital Universitario 12 de Octubre - Servicio de Bioquímica Clínica
Laboratorio de Enfermedades Mitocondriales y Neuromusculares**

A mis padres y a mi hermana Aida

A Nagore, Belatz y Ekaitz

Agradecimientos

Quiero dedicar estas líneas a las personas con las que he compartido más directamente una importante etapa profesional de mi vida que me marcó para siempre. En primer lugar al Dr. Miguel Ángel Martín, porque sin él ésta tesis no habría visto la luz, gracias por tu incansable esfuerzo por mantener viva la llama investigadora en este Servicio y por intentar siempre acoger a todos los residentes a los que nos “pica el gusanillo investigador”. Al Dr. Joaquín Arenas por ser el pionero que abrió los caminos y luchó (lucha) incansablemente por la investigación biomédica con tesón y dedicación en casi todos los frentes; en este país no es un trabajo especialmente grato, se quedan cortas las palabras... gracias de corazón a los dos.

A la Dra. Inés García-Consuegra, con tu cariñosa “mala leche”... todavía recuerdo el miedo que tenía cuando pipeteaba las primeras veces bajo tu supervisión. Al Dr. Juan Carlos Rubio, compañero desde las primeras batallas, siempre en todos los “fregaos” (y los que nos quedan). A Alberto Blázquez, la primera persona que vi en aquel ruinoso laboratorio de mitocondriales en obras en el Centro de Investigación, siempre una fuente de saber, espero seguir aprendiendo cosas de ti. Al Dr. Javier Fernández, por todos estos años compartidos organizando la genómica y la bioinformática del Hospital.

A mis compañeras del laboratorio, sin ellas en la “gogotera” este tiempo no habría sido tan entretenido: Pilar del Hoyo, lo dice el nombre, desde los inicios un sólido pilar del laboratorio; Sara Jiménez, siempre con una sonrisa y una conversación para compartir; Charo que llegó más tarde y puso firmes a los de suministros; Miguel Navas que estuvo algo más de un año secuenciando genes por Sanger cuando todavía la NGS era una ilusión; Henry Rivera que siempre estaba para echar una mano con cualquier cosa; y Laura Rufián con la que he compartido bastantes horas de poyata y alguna que otra “pifia” en los albores de la NGS. Este tiempo con todos vosotros siempre lo llevaré guardado con mucho cariño.

A los que llegaron más tarde y siguieron trabajando y explorando nuevas vías en el laboratorio: Adrián, Pablo, Rafa, Guillermo, Jorge, Susana, José, Irene. Todo un placer el haber trabajado con vosotros y poder seguir haciéndolo con algunos.

A los IPs que con tenacidad siguen luchando cada día por la investigación biomédica en el Hospital: Dr. Paco Martínez, Dra. María Morán, Dra. Cristina Ugalde y Dr. Alberto García, gracias por continuar en “la brecha” y crear una cantera ejemplar.

A los facultativos que trabajan de manera espléndida “al otro lado” con los pacientes siempre con una mirada hacia el laboratorio: Dra. Cristina Domínguez, Dra. Montse Morales, Dra. Mayte García-Silva, Dra. Elena Martín y Dra. Pilar Quijada.

Y por último recordar a los que ya no están porque se doctoraron y emprendieron vuelo: Rosa, Alberto, Teresa, David, Lorena, guardo muy buenos recuerdos de vuestro paso.

Progress in science depends on new techniques, new discoveries and new ideas, probably in that order

Sydney Brenner

Resumen/Summary

Las enfermedades del sistema de fosforilación oxidativa mitocondrial (EM-OXPHOS) se pueden originar por mutaciones en genes codificados en dos genomas: el DNA mitocondrial (mtDNA) o el DNA nuclear (nDNA). Esto obliga a establecer estrategias de análisis que tengan en consideración las particularidades propias de cada genoma.

Por lo general, estas enfermedades cursan con afectación multisistémica y presentan una amplia variabilidad clínica. Identificar la causa genética resulta fundamental para: a) evitar el peregrinaje de los pacientes entre distintos especialistas y los diferentes niveles del sistema sanitario, con el consiguiente retraso en el diagnóstico de certeza; b) clasificar adecuadamente la EM-OXPHOS concreta, debido a que cada vez se conocen más cuadros clínicos con afectación de diferentes órganos y tejidos asociados a diferentes genes y mutaciones; y c) contribuir a identificar los posibles tratamientos aplicables específicamente a pacientes con una determinada causa genética.

El objetivo de esta tesis doctoral es el desarrollo, mejora y aplicación de metódicas genético-moleculares y bioinformáticas con el propósito de su traslación al ámbito del diagnóstico genético de las EM-OXPHOS. Se han desarrollado y adaptado estrategias de análisis molecular para analizar el mtDNA en profundidad y detectar desde variaciones simples de un nucleótido hasta grandes deleciones, empleándose tanto métodos genéticos tradicionales, como aproximaciones de secuenciación masiva. En este último caso se ha desarrollado un *pipeline* completo de análisis de mtDNA a partir de datos de la plataforma PGM (*Life Technologies*).

El estudio del genoma nuclear fue abordado inicialmente con paneles de genes, mediante la integración y optimización en *scripts* propios de un sistema de filtrado de variantes frecuentes, un informe de cobertura apropiado (que indica cobertura mínima en las regiones de interés) y el programa de anotación *Annovar*. Posteriormente, se realizó una aproximación mediante secuenciación del Exoma (WES) con la oportuna adaptación del citado sistema de anotación, la inclusión del análisis de mtDNA residual y, finalmente, la detección de variaciones en el número de copias (CNVs).

El enfoque fundamentalmente traslacional de la investigación desarrollada tuvo en cuenta el balance coste/beneficio y su potencial viabilidad dentro del Sistema Sanitario. En total, y gracias a las estrategias de análisis empleadas, en el periodo de tiempo estudiado (5 años) se ha logrado el diagnóstico genético de 147 pacientes y se han encontrado 52 variantes distintas con significado incierto.

Finalmente, se propone un diagrama de flujo que aborda el diagnóstico de estas patologías mediante la integración de las diferentes metodologías y tipos de muestras empleadas en la actualidad con el objetivo de orientar el diagnóstico hacia un uso racional de los recursos de acuerdo con la información fenotípica inicial.

Mitochondrial disorders due to defects of oxidative phosphorylation system (OXPHOS disorders) are a group of rare diseases that can be produced by mutations in genes encoded by two genomes: mitochondrial DNA (mtDNA) or nuclear DNA (nDNA). This fact makes necessary to establish analysis strategies which take into account the specific characteristics of each genome.

Most of these diseases show multisystemic phenotypes with a wide clinical variability. Reaching an accurate genetic characterization is crucial: i) to avoid the diagnostic odyssey and consequently the delay in diagnosis, ii) to classify properly the mitochondrial disorder, as there are more and more clinical pictures with involvement of different organs and tissues that are associated with distinct genes and mutations, and iii) to contribute to identify possible treatments which can be particularly tailored for patients with a specific genetic cause.

The purpose of this doctoral thesis is the development, improvement, and application of molecular genetic methods and bioinformatics approaches with the aim of being transferred to the area of clinical genetic diagnosis of OXPHOS disorders.

In this work, strategies of molecular analysis were developed and adapted to analyze mtDNA in depth and to detect from single nucleotide variations to large deletions. During the process, traditional methods and massively parallel sequencing (MPS) technologies were applied. A complete pipeline for mtDNA analysis by MPS was developed using data from the PGM platform (Life Technologies).

The study of the nuclear genome was approached initially by gene panels, through the integration and optimization of own scripts to establish a system for filtering frequent variants, to generate an appropriate coverage report (that indicates minimum coverage in the regions of interest), and the *Annovar* annotation system. Then, exome sequencing (WES) studies were carried out with the adaptation of the previously developed annotation system, including the analysis of residual mtDNA and a pipeline for the detection of copy number variants (CNVs).

This work presents a translational focus, bearing in mind the cost-benefit analysis and its viability within a Health System. Thus, the application of the developed tools for analysis, in the period of time studied (5 years), led to the genetic characterization of 147 patients, and 52 different variants of uncertain significance.

In conclusion, a diagnostic flow chart for OXPHOS disorders is proposed based on initial phenotypic information and the integration of different methodologies and types of biological samples, with the aim of guiding the diagnosis under a rational use of resources.

Índices

Índice general

Abreviaturas	1
Introducción	5
Cadena Respiratoria Mitocondrial (sistema OXPHOS).....	7
Genética Mitocondrial	10
Genes nucleares y sistema OXPHOS	14
Diagnóstico de enfermedades mitocondriales	15
Herramientas de análisis y desarrollo de <i>scripts</i>	17
Objetivos	19
Materiales y Métodos	23
Muestras biológicas	25
Protocolo extracción de DNA músculo	25
Protocolo extracción de DNA sangre	26
Cuantificación de DNA (Espectrofotómetro NanoDrop ND-1000)	26
Cuantificación de dsDNA (Fluorímetro Qubit 2.0, BR Assay kit)	26
Preparación de geles de agarosa	27
Western blots en tejidos <i>post-mortem</i>	28
Electroforesis en una dimensión de tejido (<i>One-dimensional gel electrophoresis</i>).....	28
Identificación de proteínas mediante MALDI-TOF/TOF.....	28
Creación de materiales de referencia	29
Minisecuenciación.....	32
<i>PCR-RFLP</i> con cuantificación microfluídica	37
Estudio de la integridad del mtDNA.....	38
<i>Long-range PCR</i>	38
<i>Southern blot</i>	39
Estudio número de copias de mtDNA (Depleción)	39
Secuenciación masiva del mtDNA en la plataforma PGM (<i>Personal Genome Machine</i>).....	39
Secuenciación masiva del mtDNA en la plataforma MiSeq	40
Preparación de ficheros para análisis de resultados de secuenciación masiva.....	41
Detección de grandes deleciones en mtDNA y análisis de cobertura	44
Estudio de mtDNA residual en datos de secuenciación de exomas	47
Generación ficheros referencia para informe de cobertura de paneles de genes	48

Bases de datos empleadas en anotación de variantes.....	49
Algoritmos de predicción de patogenicidad.....	50
Otros programas/herramientas complementarios y lenguajes empleados	51
Resultados	53
Desarrollo de una estrategia molecular global para la identificación coste-efectiva de anomalías genética en el DNA mitocondrial (mtDNA).	55
1. Materiales de referencia para la evaluación de heteroplasmia.....	55
2. Minisequenciación (mutaciones puntuales mtDNA y haplogrupo).....	57
a. Validación metodología Minisequenciación.....	57
b. Estudio de costes Minisequenciación	63
c. Hallazgos mutaciones puntuales mtDNA mediante minisequenciación	63
3. PCR-RFLP con cuantificación microfluídica (Agilent Bioanalyzer 2100)	65
4. Secuenciación de mtDNA a alta profundidad.....	67
5. Comparación secuenciación mtDNA en PGM y MiSeq	68
6. Desarrollo de <i>scripts</i> para la automatización de procesos de anotación y filtrado de variantes	69
7. Detección de deleciones y análisis de cobertura	73
8. Epidemiología análisis mtDNA alta profundidad.....	80
9. Estudio de mtDNA residual en datos de secuenciación de exomas.....	83
10. Caracterizar variantes genéticas modificadoras del fenotipo en pacientes con mutaciones patogénicas homoplásmicas en síndromes característicos del mtDNA.....	83
Diseñar paneles de genes nucleares por resecuenciación masiva de escala media para diferentes grupos de patología mitocondrial: epidemiología molecular, estudios depleción, deleciones múltiples y estudio funcional tisular de la mutación p.T108M en el gen <i>TK2</i>	85
11. Panel genes de mantenimiento del mtDNA.....	93
12. Panel genes del Sistema de fosforilación oxidativa mitocondrial (OXPHOS).....	99
13. Estudio depleción en niños menores de 4 años (hallazgos).....	107
14. Estudio deleciones múltiples.....	108
15. Estudio funcional tisular de la mutación p.T108M en el gen <i>TK2</i>	109
Identificar nuevas mutaciones en genes asociados a enfermedad OXPHOS mediante resecuenciación de exoma en familias candidatas.	116
16. Gen <i>MTO1</i> (P1)	127
17. Gen <i>NDUFB3</i> (P2).....	129
18. Gen <i>ATAD3A</i> (P3).....	130
19. Gen <i>FDXR</i> (P4).....	132

Establecer un diagrama de flujo para el diagnóstico genético-molecular de las enfermedades OXPHOS en el ámbito del Sistema Nacional de Salud.	134
Discusión	137
Desarrollo de una estrategia molecular global para la identificación coste-efectiva de anomalías genéticas en el DNA mitocondrial (mtDNA).....	139
Caracterizar variantes genéticas modificadoras del fenotipo en pacientes con mutaciones patogénicas homoplásmicas en síndromes característicos del mtDNA.	143
Diseñar paneles de genes nucleares de escala media por resecuenciación masiva para diferentes grupos de patología mitocondrial: epidemiología molecular, estudios depleción, deleciones múltiples y estudio funcional tisular de la mutación p.T108M en el gen <i>TK2</i>	144
Identificar nuevas mutaciones en genes asociados a enfermedad OXPHOS mediante resecuenciación de exoma en familias candidatas.....	146
Establecer un diagrama de flujo para el diagnóstico genético-molecular de las enfermedades OXPHOS en el ámbito del Sistema Nacional de Salud.	149
Conclusiones	151
Bibliografía	155
Publicaciones en las que ha participado el Doctorando	166

Índice de tablas

Tabla 1. Mutaciones seleccionadas y enfermedades asociadas	32
Tabla 2. Cebadores Multiplex PCR.....	33
Tabla 3. Sondas (Reacción minsecuenciación)	34
Tabla 4. Master Mix <i>SequalPrep™ Long PCR</i>	38
Tabla 5. Condiciones PCR <i>SequalPrep™ Long PCR</i> mt16426F y mt16425R	38
Tabla 6. Condiciones PCR <i>SequalPrep™ Long PCR</i> mt3361F y mt3360R	39
Tabla 7. Comandos para la adaptación de bases de MITOMAP, MitImpact y MitoTIP.....	41
Tabla 8. Comandos para creación de filtro de variantes frecuentes.....	44
Tabla 9. Fichero bed “Torrent Server Targeted regions mtDNA”	45
Tabla 10. Fichero bed “Torrent Server Hotspots mtDNA”	46
Tabla 11. <i>Script</i> para cálculo del mínimo de cobertura.....	49
Tabla 12. <i>Script</i> con criterios de selección	49
Tabla 13. Resultados validación mutaciones puntuales.....	61
Tabla 14. Resultados validación Haplogrupos.....	62
Tabla 15. Frecuencia mutaciones minisequenciación	64
Tabla 16. Resultados heteroplasmia Bioanalizador 2100 vs <i>Last-Cycle Hot PCR</i>	66
Tabla 17. Validación de mutaciones en mtDNA patogénicas.....	68
Tabla 18. Diferencias observadas entre Sanger/Mitochip vs PGM (N=16)	69
Tabla 19. Diferencias observadas entre Sanger/Mitochip vs MiSeq (N=18).....	69
Tabla 20. <i>Script</i> de anotación de mtDNA completo	69
Tabla 21. <i>Script</i> de análisis de deleciones en mtDNA.....	73
Tabla 22. <i>Script</i> R de lanzamiento para detección y análisis de deleciones únicas y cobertura	75
Tabla 23. <i>Script</i> R de cálculo y generación de PDF	75
Tabla 24. Haplogrupos presentes en la población estudiada.....	81
Tabla 25. Mutaciones en tRNAs del mtDNA.....	81
Tabla 26. Mutaciones en genes estructurales del mtDNA	82
Tabla 27. Resumen hallazgos secuenciación mtDNA familias LHON	84
Tabla 28. <i>Script</i> de anotación de paneles de genes con Annovar	86
Tabla 29. Tiempos de análisis desde VCF	89
Tabla 30. <i>Script</i> para generar informe de cobertura de paneles	90
Tabla 31. Mutaciones sospechosas halladas con el panel de mantenimiento	96
Tabla 32. Mutaciones descritas ClinVar (07/2019) Panel Mantenimiento (Genes con hallazgos) ..	98

Tabla 33. Mutaciones sospechosas halladas con el panel OXPHOS	103
Tabla 34. Mutaciones descritas ClinVar (08/2019) Panel OXPHOS (Genes con hallazgos).....	106
Tabla 35. Resumen paciente estudio TK2	110
Tabla 36. Script general análisis CNVs.....	116
Tabla 37. Script R para la creación de fichero de referencias.....	118
Tabla 38. Script R para análisis de CNVs	119
Tabla 39. Script para anotación final de CNVs	124
Tabla 40. Pacientes con mutaciones patogénicas (estudio exoma)	127
Tabla 41. Mutación homocigosis <i>MTO1</i> (c.C1390T)	128
Tabla 42. Mutación homocigosis <i>NDUFB3</i> (c.T64C).....	129
Tabla 43. Mutación homocigosis <i>ATAD3A</i> (c.T1217G)	131
Tabla 44. Mutación heterocigosis <i>FDXR</i> (c.C814T)	132

Índice de figuras

Figura 1. Representación esquemática del mtDNA.....	8
Figura 2. Esquema de la cadena respiratoria	9
Figura 3. Sistema de replicación mitocondrial	13
Figura 4. Diferentes rutas del sistema OXPHOS condicionadas por genes nucleares	16
Figura 5. Vector pGEM®-T Easy	29
Figura 6. Esquema sistema selección de colonias	31
Figura 7. Placas agar obtenidas para selección de colonias	31
Figura 8. Esquema reacción Minisequenciación.....	36
Figura 9. Esquema <i>Last cycle Hot PCR</i>	37
Figura 10. Base de datos Mitomap adaptada.....	42
Figura 11. Base de datos adaptada Mitomap.....	43
Figura 12. Base de datos adaptada MitoTIP.....	43
Figura 13. <i>Workflow</i> para análisis de mtDNA residual	47
Figura 14. Formato resultados análisis mtDNA residual	48
Figura 15. Gel agarosa clones posición 3243 del mtDNA.....	55
Figura 16. Secuenciaciones Sanger de clones	56
Figura 17. Clon M10 recuperado tras congelación y sometido a <i>PCR-RFLP</i>	56
Figura 18. Esquema panel haplogrupos	57
Figura 19. Electroferograma PCR multiplex mutaciones puntuales.....	58
Figura 20. Electroferograma PCR multiplex haplogrupos	58
Figura 21. Electroferograma mutaciones puntuales normal.....	59
Figura 22. Electroferograma mutaciones puntuales con la mutación m.8993T>G	60
Figura 23. Electroferogramas de caracterización de haplogrupo	60
Figura 24. Electroferograma de muestra de orina (Haplogrupo H2)	61
Figura 25. Ejemplos de heterocigotos nucleares (minisequenciación)	62
Figura 26. Clones A8344G (100%M, 75%, 50%, 25%, 100%WT)	63
Figura 27. Desglose de costes por tipo de reactivo	63
Figura 28. Frecuencia mutaciones minisequenciación	64
Figura 29. Gel con marcaje radiactivo. PCR-RFLP con bsp120I (m.3243A>G)	65
Figura 30. Estudio correlación heteroplasmia Bioanalyzer 2100 vs <i>Last-Cycle Hot PCR</i>	66
Figura 31. Perfiles cobertura mtDNA en PGM.....	67
Figura 32. Perfil cobertura mtDNA con lecturas desglosadas.....	68

Figura 33. Captura de pantalla de una ejecución del <i>script</i> en la terminal.....	73
Figura 34. Ejemplo fichero Excel final con variantes anotadas.....	73
Figura 35. Informe de cobertura mtDNA: cabecera y perfil de cobertura.....	77
Figura 36. Informe de cobertura mtDNA: histogramas de cobertura (<i>Depth of coverage</i>).....	78
Figura 37. Informe de cobertura mtDNA: cobertura (<i>Depth of coverage</i>) desglosada por gen	79
Figura 38. Ejemplos de perfiles de cobertura de mtDNA con detección de deleciones.....	80
Figura 39. Esquema del pipeline de anotación de paneles de genes	86
Figura 40. Ejemplo de tabla de cobertura con las regiones con problemas de cobertura	92
Figura 41. Rutas implicadas en el mantenimiento del mtDNA	93
Figura 42. Resumen resultados pacientes Panel Mantenimiento.	94
Figura 43. Mutaciones Panel Mantenimiento.....	95
Figura 44. Resumen resultados Panel OXPHOS	101
Figura 45. Mutaciones Panel OXPHOS	102
Figura 46. Hallazgos panel genes mantenimiento en pacientes con depleción mitocondrial.....	107
Figura 47. Resultados Southern Blot + / LR-PCR +	108
Figura 48. Resultados Southern Blot - / LR-PCR +	109
Figura 49. Cambios histológicos en músculo esquelético	111
Figura 50. Análisis del número de copias del mtDNA	112
Figura 51. <i>Long range PCR</i> y <i>Southern Blot</i>	112
Figura 52. <i>Western Blot</i> subunidades OXPHOS.....	113
Figura 53. Densitometría del diafragma (OXPHOS y P62).....	113
Figura 54. Análisis degradación proteica	114
Figura 55. Electroforesis en una dimensión de diafragma	115
Figura 56. <i>MTO1</i> (c.C1390T): Segregación familiar y electroferograma Sec. Sanger	128
Figura 57. <i>MTO1</i> (c.C1390T): Estudio conservación del aminoácido 464.....	129
Figura 58. <i>NDUFB3</i> (c.T64C): Segregación familiar y electroferograma Sec. Sanger.....	130
Figura 59. <i>ATAD3A</i> (c.T1217G): Segregación familiar y electroferograma Sec. Sanger	131
Figura 60. <i>FDXR</i> : Segregación familiar y electroferograma Sec. Sanger	132
Figura 61. Perfiles cobertura posible CNV en gen <i>FDXR</i>	133
Figura 62. Diagrama flujo diagnóstico enfermedades OXPHOS	134

Abreviaturas

ACMG	<i>American American College of Medical Genetics and Genomics</i>
ARMS	<i>Amplification Refractory Mutation System</i>
ATP	Adenosín trifosfato
BAM	Versión en binario del fichero SAM
CNV	<i>Copy number variation</i>
COX	Citocromo c oxidasa o complejo IV
CPEO	Oftalmoplejia progresiva externa crónica
CPK	Creatina fosfoquinasa
DANN	<i>Deleterious annotation of genetic variants using neural networks</i>
DDD	<i>Database of genomic variation and phenotype in humans using Ensembl resources</i>
DGV	<i>Database of genomic Variants</i>
DoC	<i>Depth of coverage</i>
dsDNA	ADN doble hebra
EM-OXPHOS	Enfermedades del sistema de fosforilación oxidativa mitocondrial
ESOC	<i>Epilepsy, Strokes, Optic atrophy and Cognitive decline</i>
EXIT	<i>Exercise Intolerance</i>
FADH	Flavín adenín dinucleótido
FASTQ	Fichero de lecturas con calidad asociada
FBSN	<i>Familial Bilateral Striatal Necrosis</i>
GERP	<i>Genomic Evolutionary Rate Profiling</i>
gnomAD	<i>Genome Aggregation Database</i>
HPLC	<i>High Performance Liquid Chromatography</i>
HRM	<i>High Resolution Melt Profiling</i>
LHON	<i>Leber Hereditary Optic Neuropathy (LEBER)</i>
LR-PCR	<i>Long Range PCR</i>
LS	<i>Leigh Syndrome</i>
LDYT	<i>Leber's hereditary optic neuropathy and Dystonia</i>
MELAS	<i>Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like episodes</i>
MERFF	<i>Myoclonic epilepsy with ragged red fibers</i>
MM	Miopatía mitocondrial
mtDNA	ADN mitocondrial
NADH	Nicotin adenin dinucleótido

NAION	<i>Nonarteritic Anterior Ischemic Optic Neuropathy</i>
NARP	<i>Neurogenic muscle weakness, ataxia and retinitis pigmentosa</i>
nDNA	ADN nuclear
NGS	<i>Next Generation Sequencing</i>
NUMTs	<i>Mitochondrial DNA-like sequences in the nucleus</i>
ORF	<i>Open reading frame</i> . Secuencia codificante sin intrones
OXPHOS	Sistema de fosforilación oxidativa mitocondrial
pb	Pares de bases
PCR	Reacción en cadena de la polimerasa
PGM	<i>Personal Genome Machine</i> (Life Technologies)
rCRS	<i>Revised Cambridge Reference Sequence</i>
RFLP	<i>Restriction Fragment Length Polymorphism</i>
SAM	Fichero de texto con secuencias de bases alineadas frente al genoma de referencia
SANDO	<i>Sensory Ataxic Neuropathy, Dysarthria, and Ophthalmoparesis</i>
SB	<i>Southern Blot</i>
SDH	Succinato deshidrogenasa
SIFT	<i>Sorting Intolerant From tolerant</i> (Algoritmo bioinformático de predicción de patogenicidad)
SNP	<i>Single Nucleotide Polymorphism</i>
TS	<i>Torrent Server</i>
TTGE	<i>Temporal Temperature Gradient Gel Electrophoresis</i>
VCF	<i>Variant caller file</i> . Fichero de variantes
VUS	Variante de significado incierto
WES	<i>Whole Exome Sequencing</i>
WGS	<i>Whole Genome Sequencing</i>

Introducción

Las enfermedades mitocondriales o enfermedades del sistema de fosforilación oxidativa mitocondrial (sistema OXPHOS) pueden estar ocasionadas por mutaciones en genes codificados por el DNA mitocondrial (mtDNA) o genes codificados en el DNA nuclear (nDNA). Estas enfermedades suelen ser de afectación multisistémica y presentan una gran variabilidad clínica. Se encuadran dentro de las enfermedades raras (prevalencia <1:2000) y en su conjunto se estima una mínima prevalencia de enfermedad OXPHOS de 1:5000-10000¹, aunque la frecuencia de portadores de mutaciones en el genoma mitocondrial (mtDNA) es de alrededor de 1 en 200².

El mtDNA humano es una molécula de doble cadena circular cerrada, con una longitud de 16569 pb (Figura 1). No contiene intrones, así que prácticamente toda la secuencia es codificante, excepto el bucle de desplazamiento. Contiene 37 genes que codifican 13 polipéptidos diferentes de la cadena respiratoria mitocondrial. Los otros 24 genes son necesarios para la traducción de proteínas: 22 codifican RNA de transferencia (tRNA) y 2 codifican RNA ribosomales (rRNA).

El resto de las proteínas, codificadas por más de 1100 genes³, que forman parte del sistema OXPHOS y todas las proteínas implicadas en la regulación de la replicación y transcripción de la molécula, así como las que participan en la síntesis proteica intramitocondrial están codificadas por el DNA nuclear (nDNA), y son importadas al interior de la mitocondria mediante translocasas localizadas en la membrana externa e interna^{4,5}.

Cadena Respiratoria Mitocondrial (sistema OXPHOS)

La cadena respiratoria mitocondrial se encuentra localizada en la membrana interna mitocondrial y su función principal es la producción de energía por la oxidación de los equivalentes reductores (NADH y FADH) generados en el ciclo de Krebs a partir del acetil-CoA proveniente de diferentes rutas como la glucólisis o el catabolismo de ácidos grasos. El flujo de electrones se canaliza por distintas reacciones redox cuyo aceptor final es el oxígeno. Este se reduce y se sintetiza agua. La energía liberada en estas reacciones genera un potencial eléctrico y un gradiente de protones a través de la membrana interna mitocondrial debido al bombeo de los mismos de un lado a otro de la membrana en tres lugares específicos de la cadena respiratoria. El gradiente electroquímico resultante es empleado por el complejo V (ATP sintasa) de la cadena respiratoria para sintetizar ATP. La función del sistema OXPHOS es fundamental ya que es el principal proveedor de ATP en el organismo^{6,7}.

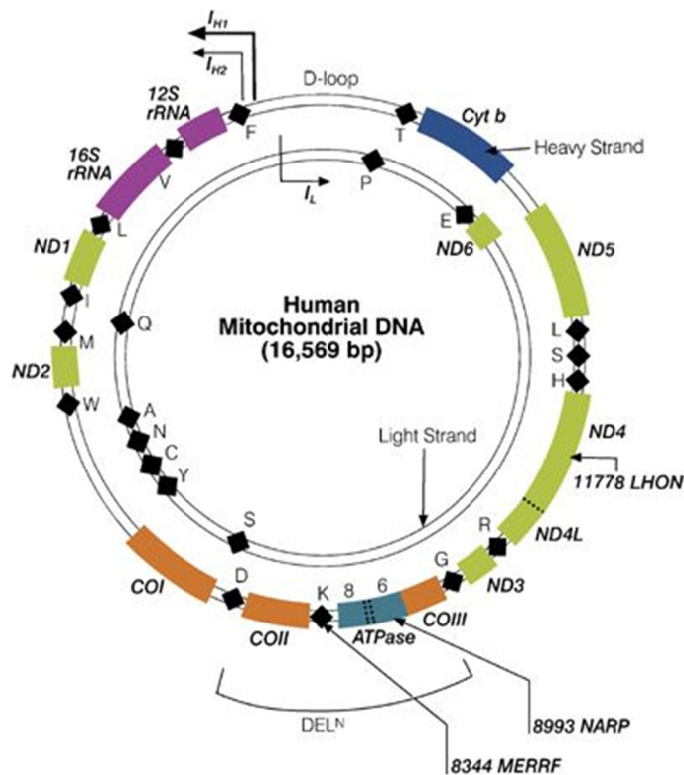


Figura 1. Representación esquemática del mtDNA

ND: genes codificantes de subunidades proteicas del Complejo I; CO: genes codificantes de subunidades del Complejo IV; ATPasas 6 y 8: genes codificantes de estas subunidades del Complejo V; rRNA 12S y 16S: genes codificantes de los RNA ribosomales mitocondriales. Los genes codificantes de los tRNA se encuentran representados con el código estándar. D-loop: bucle-D, región no codificante del mtDNA (iwate-kokyo.info).

La cadena respiratoria está integrada por cinco complejos enzimáticos y dos moléculas que actúan a modo de puntos de unión o lanzadera: el coenzima Q ó ubiquinona (CoQ) y el citocromo *c*⁷.

- **Complejo I (NADH ubiquinona (CoQ) oxidorreductasa)**

Constituido por al menos 46 subunidades, de las cuales siete están codificadas en el mtDNA, y componentes no proteicos como flavina mononucleótido (FMN), núcleos Fe-S y fosfolípidos. El complejo I cataliza la transferencia de electrones desde el NADH (proveniente de la glucólisis, la piruvato deshidrogenasa o el ciclo de Krebs) al *pool* de CoQ, lo que conlleva un salto energético que será utilizado para translocar dos protones al espacio intermembrana.

- **Complejo II (Succinato CoQ oxidorreductasa)**

Es el único complejo que no contiene subunidades codificadas por el mtDNA. Se puede dividir en dos fracciones: la soluble (succinato deshidrogenasa: SDH) y la insoluble que está constituida por las proteínas que participan en el anclaje a la membrana. El complejo II

cataliza la oxidación de succinato a fumarato, liberando FADH_2 y transfiriendo los electrones de este al *pool* de ubiquinona (coenzima Q).

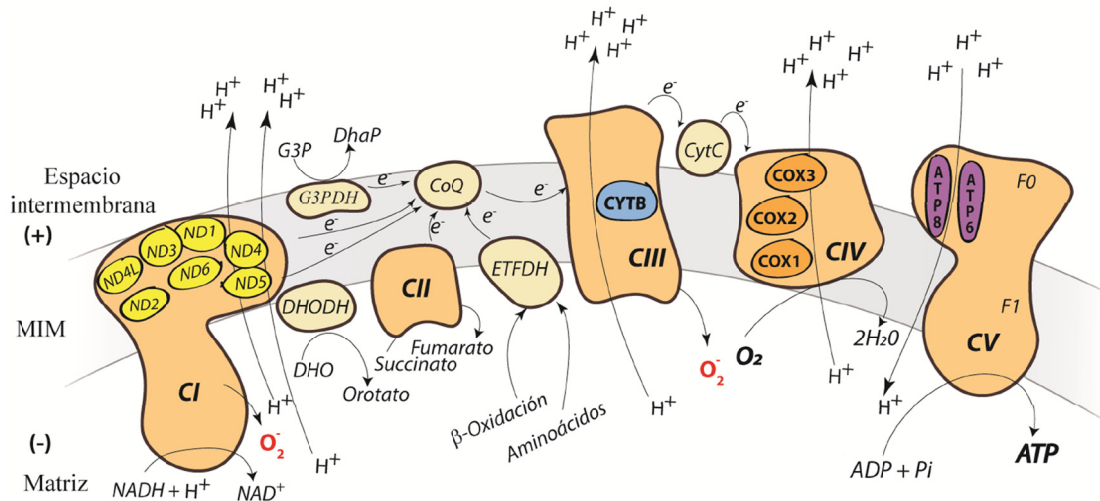


Figura 2. Esquema de la cadena respiratoria

Representación del flujo de electrones a través de la cadena de transporte electrónico y del bombeo de los protones desde la matriz al espacio intermembrana mediante los Complejos I, III y IV. Los protones pasarán de nuevo a la matriz mitocondrial a través del Complejo V produciendo ATP ⁸.

- **Complejo III (Ubiquinona (QH_2) citocromo c oxidorreductasa)**

Está formado por 11 subunidades de las cuales solo una de ellas está codificada por el mtDNA: el citocromo b. Su función es transportar electrones del ubiquinol al citocromo c. Esta transferencia está acoplada al bombeo de protones a través de la membrana interna mitocondrial (cuatro por cada par de electrones).

- **Complejo IV (Citocromo c oxidasa: COX)**

Se compone de 13 subunidades de las cuales las tres de mayor tamaño están codificadas por el mtDNA (I, II, III). Este complejo cataliza la transferencia de cuatro electrones desde el citocromo c al oxígeno molecular, acoplándose el bombeo de dos protones al espacio intermembrana debido a la energía liberada.

- **Complejo V ($\text{F}_0\text{-F}_1$ ATP sintasa)**

De este complejo se han logrado identificar 16 subunidades en humanos, dos de las cuales están codificadas por el mtDNA (ATPasa A6 y A8). Este complejo está constituido por dos elementos: la fracción F_0 , que está unida a la membrana y se encarga de la conducción de protones desde el espacio intermembrana hacia el interior de la matriz mitocondrial. Y la fracción F_1 (denominada fracción catalítica) que emplea la energía obtenida en el transporte

de protones citado para sintetizar ATP a partir de ADP y Pi. Es una enzima funcionalmente reversible, capaz de sintetizar ATP y de hidrolizarlo (función ATPasa) bombeando protones en contra de gradiente.

Las enfermedades OXPHOS son normalmente condiciones progresivas que provocan discapacidad y, en algunos casos, muerte prematura debida con frecuencia a defectos en la conducción cardíaca o a la presencia de convulsiones⁹. Dado que la mitocondria es la principal fuente de producción de energía celular en mamíferos, los tejidos que presentan una mayor afectación son aquellos que tienen mayores necesidades energéticas: sistema nervioso central, nervio óptico, retina, músculo esquelético (particularmente la musculatura extra-ocular), corazón y sistema endocrino. La presentación clínica es muy variable en cuanto a la edad de inicio, los síntomas, la gravedad y el pronóstico.

Las mutaciones puntuales del mtDNA pueden afectar a las regiones codificantes de proteínas y a genes codificantes de los RNA de transferencia (tRNA) que alteran la síntesis intramitocondrial de proteínas. Las deleciones únicas obedecen generalmente a eventos esporádicos y parece que no serían heredadas, aunque se han descrito algunos casos de transmisión familiar¹⁰. La depleción y las deleciones múltiples son efectos secundarios en el mantenimiento del mtDNA debidos a mutaciones en genes nucleares que codifican proteínas mitocondriales. Cada célula contiene varias mitocondrias, y cada mitocondria contiene numerosos genomas. En este contexto, surge la característica genética de la heteroplasmia, con la coexistencia en la misma célula (o mitocondria) de proporciones variables de genomas mutados y normales (*wild-type*). El grado de heteroplasmia influye en la probabilidad y gravedad de la enfermedad. El principal factor que afecta al nivel de heteroplasmia se produce durante la ovogénesis y se conoce como “cuello de botella” mitocondrial¹¹. Este hace que la cantidad total de mitocondrias del organismo se replique solo a partir de un pequeño *pool* de genomas.

Genética Mitocondrial

La genética mitocondrial presenta una serie de características peculiares que la diferencian de la genética nuclear¹².

- Herencia por vía materna

Durante la fertilización prácticamente todo el mtDNA que contiene el cigoto es aportado por el óvulo. Una madre que porte una mutación en su mtDNA puede transmitirla a su descendencia, pero solo sus hijas la transmitirán a la siguiente generación. Sin embargo, la dosis

de mtDNA mutado que posea cada uno de sus ovocitos puede variar ampliamente debido al proceso de restricción y amplificación mencionado anteriormente. Según este modelo, durante una etapa temprana de la embriogénesis de la mujer, en la formación de su línea germinal, el número de moléculas de mtDNA en cada oocito disminuiría drásticamente. Más tarde tendría lugar una amplificación del mtDNA que generaría las aproximadamente 100.000 copias que existen en el oocito maduro. Este fenómeno sería el causante de las diferencias de cantidad relativa de mtDNA mutado existente entre los ovocitos.

Se han publicado diversos casos en los que se ha detectado presencia de mtDNA de origen paterno¹³ y se ha llegado a detectar una mutación en el mtDNA heredada por vía paterna en el músculo esquelético con un porcentaje de heteroplasmia suficiente para manifestar patología¹⁴.

- Poliplasmia

Cada célula puede contener cientos de mitocondrias, y cada mitocondria contiene entre 2 a 10 copias de mtDNA, por lo tanto, podrían existir miles de copias de mtDNA y de los genes que codificados por éste.

- Segregación mitótica

La división de las mitocondrias y la replicación del mtDNA son procesos independientes del ciclo celular. Cuando la célula se divide, las mitocondrias se distribuyen entre las células hijas al azar. Todo ello da lugar al fenómeno de segregación mitótica. Según este, en un paciente portador de mtDNA normal y mutado en sucesivas divisiones pueden generarse tres tipos distintos de poblaciones: células con mitocondrias portadoras de mtDNA normal, entonces hablaremos de homoplasmia normal; células en las que sus mitocondrias contengan material genético mutado, en este caso estaremos ante homoplasmia mutante; y células que porten los dos tipos de mitocondrias, normales y mutantes, en cuyo caso las definiremos como heteroplásmicas. Las variantes nucleotídicas no deletéreas suelen ser homoplásmicas, mientras que las mutaciones patogénicas pueden resultar heteroplásmicas, aunque esto es una generalización con muchas excepciones.

Destacar que la proporción de moléculas de mtDNA mutadas puede variar de un tejido a otro e incluso en uno mismo, y a lo largo del tiempo. Esto implica que los síntomas observados en un paciente pueden variar con el paso del tiempo y que la edad de comienzo, para una misma alteración, sea diferente de unos pacientes a otros.

- Efecto umbral

Cuando un paciente hereda una mutación en el mtDNA la proporción de moléculas mutadas puede variar de un tejido a otro. El efecto umbral representa la proporción mínima de mtDNA mutado necesario para que se produzca la disfunción en un determinado órgano o tejido. Este porcentaje varía de unos tejidos a otros dependiendo de los requerimientos energéticos de los mismos. Tejidos como el cerebro, corazón o músculo esquelético tienen grandes demandas de energía por vía oxidativa y sus umbrales son menores que los de otros tejidos como la piel o la sangre. Por este motivo la presencia de una mutación en una determinada proporción en hígado puede ser clínicamente silente, mientras que la misma proporción en músculo o cerebro puede dar lugar a expresión fenotípica. La proporción de mtDNA mutado en el mismo tejido puede variar con el tiempo y las demandas energéticas. Por ejemplo, la depleción parcial de mtDNA en músculo no produce expresión clínica los primeros meses de vida; pero causa una miopatía cuando la alteración del metabolismo oxidativo no puede hacer frente al incremento de la demanda energética¹⁵. Y se da también el caso extremo de tejidos con alto índice de recambio, como el hemático, en los que una mutación puede llegar a desaparecer debido a una selección clonal negativa.

Replicación del mtDNA

El mtDNA está continuamente replicándose de una forma compleja y lenta en un proceso totalmente independiente a la replicación del DNA nuclear (no está ligado a un momento específico del ciclo celular, fase S)^{16,17}. El mecanismo de replicación fue descrito por Kasamatsu hace más de treinta años¹⁸ y estudiado en detalle por Clayton¹⁹⁻²¹.

Hasta el año 2000 se aceptaba que la replicación del mtDNA es asincrónica con dos orígenes de replicación unidireccionales. Comienza en el origen O_H y sigue la dirección de las agujas del reloj, mientras que la otra cadena comienza su replicación en el origen O_L cuando ya se han sintetizado dos terceras partes de la cadena pesada y sigue el sentido contrario de las agujas del reloj. Pero, en el año 2000, Holt demostró en células de mamíferos la replicación en paralelo de la hebra guía y la hebra retardada²².

Se necesita la unión del factor de transcripción mitocondrial A (TFAM) que permite la separación de las cadenas. En ese punto hibrida un RNA, sintetizado por una RNA polimerasa mitocondrial (mtRNAP), y la cadena ligera desplaza a la cadena H dando lugar al bucle D. Un defecto en el homólogo de esta RNA polimerasa en levaduras se ha asociado a depleción del

genoma mitocondrial, sin producir ninguna alteración a nivel estructural²³. Para que comience la síntesis de DNA debe eliminarse el RNA cebador por la RNaseH1 (endorribonucleasa), generando el cebador necesario para la actuación de la DNA polimerasa y mitocondrial (POLG). Otras proteínas (Figura 3) que participan en el proceso son: Twinkle (helicasa, necesaria para la abertura de la doble hélice de DNA), mtSSB (proteína de unión a la hebra simple de mtDNA para su estabilización), DNA ligase III (para el sellado de los cortes en el mtDNA), MGME1 (exonucleasa tipo RecB), FEN1 (endonucleasa) y DNA2 (helicasa/nucleasa)²⁴

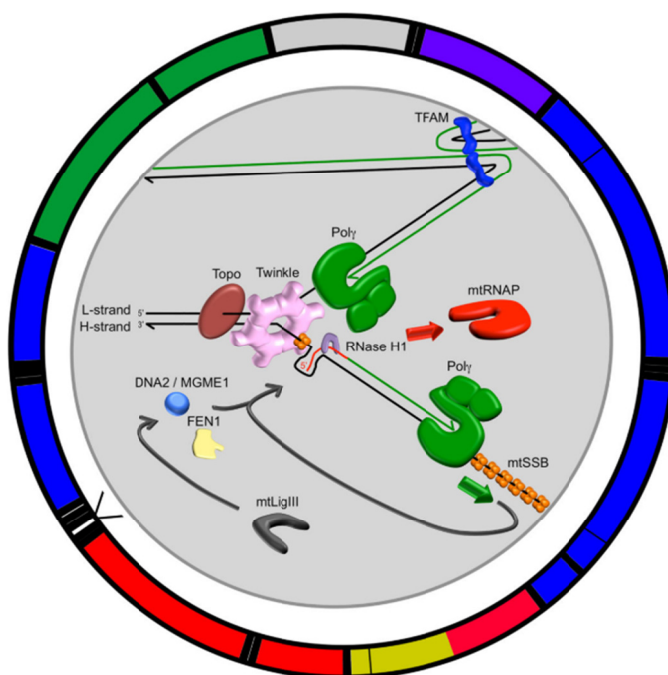


Figura 3. Sistema de replicación mitocondrial

Se muestran los componentes principales del sistema de replicación del mtDNA. Las líneas negras representan la hebra de mtDNA molde y las verdes la nueva hebra sintetizada²⁴

Transcripción del mtDNA

La transcripción tiene lugar a partir de los promotores de ambas cadenas (L, H₁, H₂)²⁵ situados en el bucle D, y evoluciona en sentido contrario y de forma sincrónica. La transcripción de la cadena H comienza en dos puntos (IT_{H1} e IT_{H2}), mientras que la cadena ligera se transcribe a partir de uno sólo (IT_L). Los transcritos a partir de IT_{H2} e IT_L son policistrónicos y corresponden prácticamente a secuencias completas de las dos cadenas. Desde el punto IT_{H2}, el transcrito generado termina después del gen rRNA 16S. El inicio de la transcripción requiere una RNA polimerasa mitocondrial específica, y el factor mtTFA²⁶ que reconoce secuencias específicas de los

promotores, además de esta función también juega un papel de empaquetador cromosómico. La afinidad de este factor es diferente por ambos promotores, siendo mayor para la cadena ligera, lo que supone un aumento del número de transcritos de la misma respecto a la cadena pesada²⁷. La existencia de dos unidades distintas de transcripción en el promotor de la cadena pesada explica en parte la activación preferencial de la síntesis de rRNA frente al mRNA policistrónico total²⁸. En esta regulación parece intervenir también un factor regulador terminador (mTERF)²⁹.

Existen mecanismos reguladores que adaptan la síntesis de las subunidades que forman parte de la cadena respiratoria según la demanda energética de la célula. Se ha observado que niveles elevados de ATP estimulan la transcripción de los rRNA y de la cadena L, mientras que niveles disminuidos de ATP estimulan la transcripción de la cadena H completa³⁰.

Generalmente los RNA maduros corresponden a un solo gen y sólo dos de ellos (ND4 y ND4L, subunidades 6 y 8 de la ATPasa) contienen dos genes cada uno con el marco de lectura solapado.

Traducción del mtDNA

Los mRNAs mitocondriales son traducidos en la matriz mitocondrial mediante sus propios ribosomas. El coeficiente de sedimentación de estos es 55S, con subunidades 28S y 39S, que contienen 32 y 52 proteínas respectivamente, todas ellas codificadas por el nDNA, además de los dos rRNAs codificados por el mtDNA.

El código genético utilizado por la mitocondria tiene una serie de diferencias con respecto al universal. El codon UGA codifica triptófano en lugar de terminación; AUA indica metionina en lugar de isoleucina; AGA y AGG especifican terminación en lugar de arginina. Este fenómeno asegura que solo los mensajes codificados por el mtDNA serán traducidos satisfactoriamente en la matriz.

Otra característica particular del sistema genético mitocondrial es el uso de un modelo inusual de reconocimiento de codones, que permite la lectura del código con solo 22 tRNAs codificados por el mtDNA.

Genes nucleares y sistema OXPHOS

Las enfermedades relacionadas con el sistema OXPHOS afectan a 1 de cada 5000 nacidos vivos y se caracterizan por la presencia de defectos en la cadena respiratoria mitocondrial. Se

trata de patologías clínicamente heterogéneas que pueden debutar en la infancia o en la edad adulta, resultar más o menos severas y son generalmente son multisistémicas. Las características clínicas más comunes son: miopatía, acidosis láctica, convulsiones, ataxia, neuropatía periférica, defectos visuales y auditivos, problemas de motilidad gastrointestinal, fallo hepático y cardiomiopatía. La herencia puede ser materna, autosómica recesiva o dominante, o ligada al cromosoma X, aunque la mayoría de los casos son esporádicos.

La causa genética subyacente de estas enfermedades pueden ser mutaciones en el mtDNA o en el nDNA. Se estima que entre 15%-30% de todos los casos infantiles se deben a mutaciones en el mtDNA, por lo que la mayoría estarían relacionados con mutaciones en el nDNA. Cada año se describen más de 10 nuevos genes nucleares OXPHOS.

Los síndromes relacionados con el mantenimiento del mtDNA están causados por mutaciones en más de 10 genes nucleares (Figura 4) implicados en la replicación del mtDNA, el balance de nucleótidos en la mitocondria y la dinámica mitocondrial. En este trabajo se han hallado mutaciones en varios pacientes en algunos de estos genes y se ha podido desarrollar un estudio en tejidos *post-mortem* de uno de ellos que era portador homocigoto de la mutación p.Thr108Met en el gen *TK2* (pág. 109).

Diagnóstico de enfermedades mitocondriales

El diagnóstico de las enfermedades mitocondriales se inicia en la actualidad a partir de la información aportada por los médicos, patólogos, bioquímicos y genetistas moleculares. A pesar de todas estas fuentes de información, una gran proporción de pacientes con enfermedades OXPHOS no poseen en la actualidad un diagnóstico genético conocido. Hasta la fecha se han descrito mutaciones en más de 350 genes, tanto nucleares como mitocondriales³¹.

El uso de diferentes plataformas de diagnóstico molecular permite identificar con mucha mayor celeridad los defectos genéticos causantes de estos trastornos y posibilita establecer la existencia de perfiles genómicos asociados a la variabilidad fenotípica. Conocer la causa genética es de gran importancia para el cuidado de los pacientes por diversos motivos: i) evita el peregrinaje entre distintos especialistas y niveles del sistema sanitario, con el consiguiente retraso en el diagnóstico de certeza; ii) clasifica adecuadamente la patología mitocondrial concreta, pues cada vez existen más cuadros clínicos (fenotipos) con afectación de diferentes órganos y tejidos, asociados a diferentes genes y mutaciones; iii) es cada vez más frecuente identificar posibles

tratamientos que son aplicables específicamente a un grupo de pacientes con una determinada causa genética, hecho particularmente importante en las enfermedades OXPHOS.

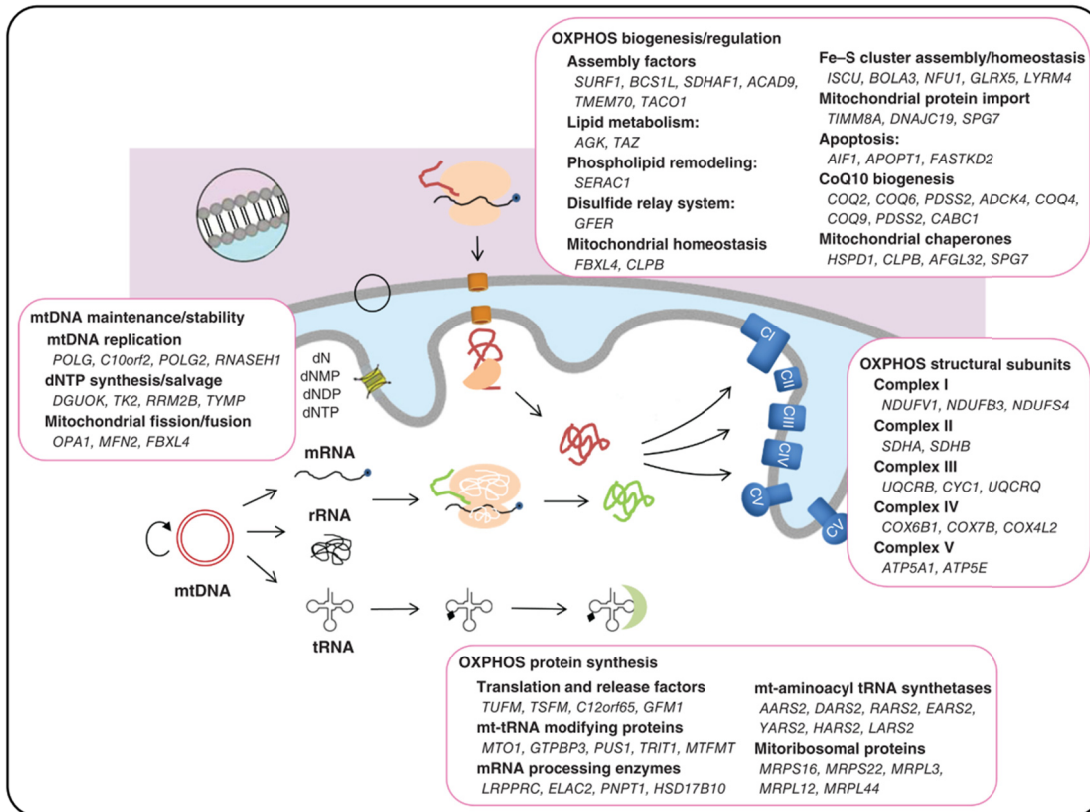


Figura 4. Diferentes rutas del sistema OXPHOS condicionadas por genes nucleares³².

Hasta fechas muy recientes el diagnóstico genético-molecular de enfermedades mitocondriales del sistema OXPHOS se había basado esencialmente en hallazgos bioquímicos, histoquímicos, enzimológicos y algunos estudios genéticos mitocondriales y nucleares muy concretos. Con esta estrategia se ha podido identificar la causa genética de la enfermedad únicamente en alrededor de un 20-30% de los pacientes. En la actualidad, las nuevas metodologías de secuenciación masiva (NGS, *Next Generation Sequencing*) han devenido mucho más asequibles, y permiten obtener de manera rápida^{33,34} resultados de un gran número de genes o regiones genéticas diana. La llegada de las plataformas de sobremesa ha permitido una reducción considerable de los costes y la posibilidad de realizar estudios con un bajo número de muestras, especialmente interesante en enfermedades de baja prevalencia. Se puede estimar que con estos nuevos métodos se podría llegar al diagnóstico de certeza en más del 60% de los pacientes de nuestro entorno.

Desde el punto de vista técnico (química, secuenciadores y sistemas informáticos) la generación rápida de un gran número de datos genéticos se va resolviendo, pero el proceso

presenta actualmente un cuello de botella muy importante: la manipulación de estos resultados para generar información anotada, comprensible y útil que pueda ser empleada para hallar la causa genético-molecular de las patologías de los pacientes³⁵. En el caso de las enfermedades OXPHOS, se añade el hecho del estudio del genoma mitocondrial, con sus particularidades, además del genoma nuclear.

Herramientas de análisis y desarrollo de *scripts*

Las nuevas plataformas de NGS han permitido abordar a gran escala la secuenciación completa del mtDNA. Existen más de 47.000 genomas mitocondriales humanos completos de múltiples haplogrupos publicados en GenBank^a. Para el análisis de esta información han surgido diferentes herramientas bioinformáticas basadas fundamentalmente en Servidores Web de uso público como Haplogrep^{36b}, mtDNA Server^{37c} o MitImpact³⁸.

Para el análisis del mtDNA se ha desarrollado una herramienta de análisis (Tabla 20) adaptada a la secuenciación en PGM con las siguientes premisas: no depender excesivamente de elementos externos, poder ejecutarse íntegramente en la infraestructura informática local y centrarse en la traslación a su uso clínico para su aplicación directa en el diagnóstico molecular de enfermedades mitocondriales. Para ello se han empleado herramientas disponibles en el Servidor *Torrent Server* (TS) que realiza el análisis primario y una parte del secundario hasta la generación del fichero de variantes (en formato VCF o Excel).

Se ha creado un *script* con un lenguaje de programación de alto nivel (Perl) para la anotación de las variantes resultantes del proceso de análisis del TS. Para ello se ha generado una base de SNPs a partir de datos de la base pública MITOMAP (Tabla 7).

Para la anotación de resultados de paneles de genes nucleares se ha integrado ANNOVAR³⁹ en un *script* propio (Tabla 28). Se trata de un programa de línea de comandos basado en Perl que permite anotar con información biológica y funcional resultados de ensayos de secuenciación masiva. Es un software gratuito fuera del ámbito comercial y usado ampliamente en este campo (usado en el 52% de los casos del estudio CLARITY)³³. Es muy versátil y permite seleccionar las bases de datos que se deseen emplear o incluso introducir bases propias.

^a[https://www.ncbi.nlm.nih.gov/nuccore/?term=\(015400\[SLEN\]:016600\[SLEN\]\)+AND+Homo\[Organism\]+AND+mitochondrion\[FILT\]+NOT+\(Homo+sp.+Altai+OR+Denisova+hominin+OR+neanderthalensis+OR+heidelbergensis+OR+consensus+OR+ancient+human+remains\)](https://www.ncbi.nlm.nih.gov/nuccore/?term=(015400[SLEN]:016600[SLEN])+AND+Homo[Organism]+AND+mitochondrion[FILT]+NOT+(Homo+sp.+Altai+OR+Denisova+hominin+OR+neanderthalensis+OR+heidelbergensis+OR+consensus+OR+ancient+human+remains))

^b <http://haplogrep.uibk.ac.at/>

^c <https://mtdna-server.uibk.ac.at/index.html>

Un tipo de anomalías genéticas estructurales que debe tenerse también en cuenta son las CNVs: variaciones en el número de copias, que incluyen normalmente fragmentos entre 1kb y 5Mb⁴⁰. Este tipo de variantes son una fuente más de variabilidad genética y están implicadas en una parte nada desdeñable de patologías⁴¹.

En la actualidad existen más 30 programas bioinformáticos para detectar CNVs. La mayoría están diseñados para trabajar con datos de WGS y existen 12 herramientas optimizadas para datos de WES⁴². Los cuatro métodos más frecuentes que se emplean son: *paired-end mapping*, *split read*, *de novo assembly* y *depth of coverage (DoC)*. En este trabajo se ha empleado *ExomeDepth*⁴³: una herramienta basada en R que emplea el método DoC, que cuenta con una consolidada trayectoria. Se la considera sensible y rápida⁴² y ha sido evaluada en diversas intercomparaciones⁴²⁻⁴⁷. *ExomeDepth* emplea un robusto método beta-binomial para generar una referencia óptima para cada caso y calcula la probabilidad de la existencia de una delección o una duplicación a nivel de exón entre cada muestra y la selección de muestras de referencia.

Objetivos

El objetivo general de esta tesis doctoral consiste en el desarrollo, mejora y aplicación de metodologías genético-moleculares y bioinformáticas con el propósito de su traslación al ámbito del diagnóstico genético de enfermedades mitocondriales del sistema OXPHOS. Dado el carácter genómico dual de estos trastornos, este objetivo se ha subdividido en dos líneas principales: i) la secuenciación en profundidad del genoma mitocondrial completo con el fin de cuantificar fiablemente el grado de heteroplasmia y detectar grandes deleciones, y ii) la secuenciación de genes nucleares mediante el diseño de paneles específicos o la secuenciación completa del exoma.

Objetivos específicos:

- 1) Establecer una estrategia molecular global para la identificación coste-efectiva de anomalías genéticas en el DNA mitocondrial (mtDNA).
- 2) Caracterizar variantes genéticas modificadoras del fenotipo en pacientes con mutaciones patogénicas homoplásmicas en síndromes característicos del mtDNA.
- 3) Diseñar paneles de genes nucleares de escala media por resecuenciación masiva para diferentes grupos de patología mitocondrial: epidemiología molecular, estudios de depleción, deleciones múltiples y estudio funcional tisular de la mutación p.T108M en el gen *TK2*.
- 4) Identificar nuevas mutaciones en genes asociados a enfermedad OXPHOS mediante resecuenciación de exoma en familias candidatas.
- 5) Establecer un diagrama de flujo para el diagnóstico genético-molecular de las enfermedades OXPHOS en el ámbito del Sistema Nacional de Salud.

Una parte importante del trabajo a desarrollar se enmarcará en el terreno de la bioinformática. Las nuevas plataformas de secuenciación son capaces de generar una gran cantidad de información y en estos momentos el cuello de botella principal es la manipulación y análisis de este tipo peculiar de datos.

Se trata por tanto de un proyecto fundamentalmente traslacional que pretende adaptar e incorporar al ámbito clínico nuevas metodologías genéticas y genómicas de diagnóstico coste/efectivas. Con ello se pretende establecer una estrategia de identificación genético-molecular de enfermedades mitocondriales del sistema OXPHOS, con el fin de cubrir la necesidad del diagnóstico genético y la identificación de posibles factores que provocan la heterogeneidad clínica de estas patologías complejas.

Materiales y Métodos

Muestras biológicas

Se utilizaron muestras de DNA extraídas a partir de biopsias musculares y de sangre de pacientes portadores de las mutaciones de interés. Dichos pacientes estaban previamente diagnosticados mediante los estudios genéticos apropiados y confirmados por secuenciación directa. Todas las muestras estaban implicadas en diversos proyectos (PI08/90246, PI06/90088, PI12/01683, PI15/00431, PI18/01374) y contaban con la aprobación del Comité de Ética del Hospital 12 de Octubre.

Muestra primaria

- A) Sangre total:** Se extrajeron 10 mL de sangre depositándose en un tubo con EDTA, para evitar la coagulación. Se procedió a su procesamiento inmediato para la obtención de DNA total o se conservó en congelación a -20°C hasta su extracción.
- B) Músculo esquelético:** Dicho tejido se obtuvo realizando una biopsia muscular abierta de los músculos vasto externo y deltoides. Todas las biopsias de músculo esquelético se congelaron inmediatamente tras su extracción en nitrógeno líquido y se conservaron hasta su análisis en un congelador a -80°C .

Protocolo extracción de DNA músculo

Se parte de una cantidad variable de tejido congelado en nitrógeno líquido (entre 15-40 mg) que se tritura en un mortero evitando la descongelación. Se deposita el pulverizado en un tubo de polipropileno estéril (2 mL) al que se le añaden 180 μL de tampón TSE (Tris-HCl 100 mM pH=7,4, ClNa 10 mM, EDTA 25 mM) y 20 μL de solución acuosa de proteinasa K (20 mg/mL). Se centrifuga brevemente a 14.000 rpm, se mezcla con la pipeta y se incuba 5 min. a 55°C . Se añaden otros 180 μL de tampón TSE, 20 μL de proteinasa K y 40 μL de dodecil sulfato sódico (SDS). Se mezcla de nuevo y se incuba a 55°C otros 60 min. Posteriormente se añaden 20 μL más de proteinasa K y se incuba de nuevo durante 30 min. Se incorporan 50 μL de ClNa 5M, se mezcla, y se añade 1 mL de una mezcla de fenol:cloroformo:alcohol isoamílico (25:24:1), a pH superior a 8. Se agita y se centrifuga a 14.000 rpm brevemente para separar las fases. Se recoge el sobrenadante y se lava dos veces con 1 mL de éter dietílico. Una vez evaporado el éter, se añaden 1,5 mL de etanol al 98%, obteniendo el precipitado de DNA. Éste se recoge con un asa estéril en un tubo y se lleva a evaporación el resto de etanol. Por último, se resuspende en 50-100 μL de tampón Tris-EDTA (TE: Tris-HCl 10 mM, EDTA 1mM, pH=7,4). La muestra de DNA se puede conservar a -20°C .

En algunos casos la extracción de DNA también se realizó con *QIAamp DNA Mini Kit* (Qiagen, Hilden, Germany) siguiendo las recomendaciones del fabricante^d.

Protocolo extracción de DNA sangre

El DNA se extrajo a partir de 10mL de sangre total en EDTA con el kit *Nucleon BACC3* (Amersham-Pharmacia) siguiendo las instrucciones del fabricante. El método se basa en la precipitación de proteínas con perclorato sódico y la separación del DNA con cloroformo y una resina patentada.

Cuantificación de DNA (Espectrofotómetro NanoDrop ND-1000)

De la solución final resuspendida en Tris-EDTA tomamos un pequeño volumen para valorar la concentración, midiendo la absorbancia a 260 y 280 nanómetros (nm):

Se toman 1,5 μL de agua destilada para hacer un blanco y 1,5 μL de la solución problema (DNA del paciente) que serán medidos utilizando un espectrofotómetro NanoDrop ND-1000. Programa: Ácidos Nucléicos.

La relación entre los valores de densidades ópticas de 260/280nm indicará la pureza del DNA (que no debe contener un exceso de material proteico o de solvente orgánico), siendo aconsejable que este valor se encuentre entre 1,7 y 2⁴⁸, y nunca por debajo de 1,5. El cálculo de la concentración en $\mu\text{g}/\text{mL}$ viene dado por la multiplicación del valor de la absorbancia a 260 nm por el factor de dilución y por 50, que es el factor específico para el DNA. Se expresa la concentración en $\text{ng}/\mu\text{L} = \mu\text{g}/\mu\text{L} \times 1000$.

Cuantificación de dsDNA (Fluorímetro Qubit 2.0, BR Assay kit)

Atemperar los estándares 1 y 2 BR (guardados a 4°C) 30 minutos antes de ser utilizados. El resto de reactivos se conservan a T^a ambiente: *Buffer* y *Fluoróforo Qubit Reagent* (evitar contacto con la luz).

Crear la **Working Solution**:

- 1 μL /muestra de *Qubit Reagent*
- 199 μL /muestra de *Buffer*

Preparación de los **estándares** ($V_f = 200\mu\text{L}$) en tubos 0,5 específicos del Qubit.

- 190 μL *Working Solution*
- 10 μL *Standards Kit BR(1, 2)*

^d <https://www.qiagen.com/us/resources/resourcedetail?id=62a200d6-faf4-469b-b50f-2b59cf738962&lang=en>

Preparación de las **muestras** ($V_f=200\mu\text{L}$) en tubos 0,5 específicos del Qubit.

- 180-199 μL *Working Solution*
- 20-1 μL DNA (muestra)

Vortex durante 2-3 segundos cada tubo.

Incubar durante 2 minutos TA.

Encender el Qubit, seleccionar **DNA** y **YES** para utilizar nuevos estándares. Se introduce el Standard 1 y a continuación el 2 para calcular la recta de concentración que usará.

Es importante seleccionar la cantidad de muestra añadida (1-20 μL) y las unidades en la que queremos expresar la concentración DNA ($\text{ng}/\mu\text{L}$).

Preparación de geles de agarosa

Se prepararon geles de agarosa en un molde de ocho pocillos de la siguiente manera:

Añadir 60 μL de TAE 1X en un Matraz Erlenmeyer de 250 mL.

Pesar los gramos necesarios de agar según el % de concentración del gel de trabajo.

Tabla guía:

% concentración del gel de trabajo	Gramos Agarosa
1	60
1,5	90
1,8	108
2	120
3	180

Mezclar la agarosa con el TAE 1X en el Erlenmeyer y agitar suavemente.

Calentar en el microondas hasta hervir (aproximadamente 2 minutos a temperatura máxima).

Observar si está totalmente disuelto (medio transparente). Si no es así volver a repetir el paso anterior.

Dejar enfriar en campana de gases hasta que no salga vapor, añadir 3 μL de Bromuro de Etidio (1 μL Bromuro de Etidio /20 mL de gel) y agitar con suavidad.

Verter la solución suavemente sobre el molde de electroforesis, eliminar las burbujas con una punta de pipeta y colocar el peine. Dejar solidificar (aproximadamente 15 minutos).

Colocar el molde con el gel en la cubeta de electroforesis y rellenar con TAE 1X hasta cubrir totalmente el gel.

Gel agarosa para visualización de mtDNA completo (1 amplicón) SEAKEM®GTG® 0,8%

Preparación del gel: 3,2 g agarosa, 400mL TAE 1X y 20 µL Bromuro de Etidio.

Preparación Marker II Boheringer: 12 µL Marker + 6 µL Buffer de carga.

Preparación Muestras: 3 µL producto L-PCR + 6 µL Buffer de carga.

Los productos de PCR se corren en gel de agarosa 0,8% a 4°C (cámara fría), a un voltaje inicial de 65V unos 15 minutos y posteriormente a una intensidad de 140mA y 45-50V durante 4-5 horas.

Western blots en tejidos *post-mortem*

Se realizaron WB a partir de lisados celulares de tejidos mediante homogenización con buffer RIPA que incluye inhibidores de proteasas y fosfatasa.

El análisis del sistema OXPHOS se ha realizado con anticuerpos específicos frente a SDHA, CV alpha, COX1, ATP8 y GAPDH. Para los estudios de recambio proteico y degradación por autofagia y el sistema ubiquitina-proteasoma, se han analizado por western blot los niveles de P62/SQSTM1, proteínas ubiquitinadas y las subunidades alpha y beta del proteasoma 20S. La α -tubulina y la β -actina se han empleado como controles de carga.

Electroforesis en una dimensión de tejido (*One-dimensional gel electrophoresis*)

Se incubaron 20µg de homogenado de diafragma (paciente y control) con buffer LDS [buffer (0.4 mM EDTA, 8% (w/v) lauryl dodecyl sulfate (SDS), 50 mM Tris-HCl (pH 6.8), 4% glycerol, 0.075% Serva Blue G250 (w/v) y 0.025% Phenol Red (w/v)] durante 10 minutos a temperatura ambiente.

Dichas muestras se cargaron en un gel de poliacrilamida al 12% y se realizó la electroforesis con voltaje constante de 100V con buffer de carrera (buffer (25 mM Tris, 192 mM glycine, 0.1% sodium dodecyl sulfate). Tras la migración el gel se tiñó con *Brilliant Blue G-Colloidal Concentrate* (Sigma) para poder visualizar las bandas de las proteínas separadas⁴⁹.

Identificación de proteínas mediante MALDI-TOF/TOF

Este análisis se llevó a cabo en la Unidad de Proteómica de la Facultad de Farmacia de la Universidad Complutense de Madrid. A partir del gel mencionado en el apartado anterior, determinadas bandas se escindieron manualmente y fueron digeridas con tripsina: se lavaron 3 veces con 25 mM ammonium bicarbonate/50% acetonitrile y se secaron posteriormente en un equipo *Speed-Vac*. Después se añadieron a las piezas de gel seco 10 µL de tripsina a 25 µg/mL (*modified bovine trypsin*, ABSciex, USA) en 25 mM ammonium bicarbonate y se realizó una

incubación toda la noche a 37°C. La extracción de los péptidos resultantes se realizó mediante la adicción en 3 ocasiones de 10% formic acid/50% acetonitrile con liofilización posterior en Speed-Vac. Finalmente el producto se resuspendió en acetonitrile/formic acid solution con α -cyano-4-hydroxycinnamic acid (1:1) y se realizaron alícuotas en placas para el espectrómetro de masas. El espectro de masas en un rango de masas entre 700 y 4500 Da se obtuvo con un sistema MALDI-TOF/TOF (4800 Proteomics Analyzer, Applied Biosystems, Foster City, CA, USA) mediante un ensayo dato dependiente que selecciona los 6 picos más intensos de cada muestra excluyendo aquellos relacionados con la matriz, la autólisis de tripsina o picos de acrilamida. Los picos de autólisis de tripsina se emplearon como calibradores internos permitiendo una sensibilidad de hasta 20 ppm. El análisis de los espectros se realizó mediante en el *Global Protein Server Workstation* (Applied Biosystems, Foster City, CA, USA) que emplea el software Mascot (v2.1.1. Matrix Science Ltd, UK). Las búsquedas se realizaron empleando la base de datos SwissProt.

Creación de materiales de referencia

Se utilizaron bacterias *Escherichia coli* ultracompetentes (JM109, Promega) y como vector el plásmido “pGEM®-T Easy” de Promega (Figura 5).

El DNA plasmídico se purificó mediante el sistema “Wizard® Plus SV Minipreps DNA purification system” (Promega).

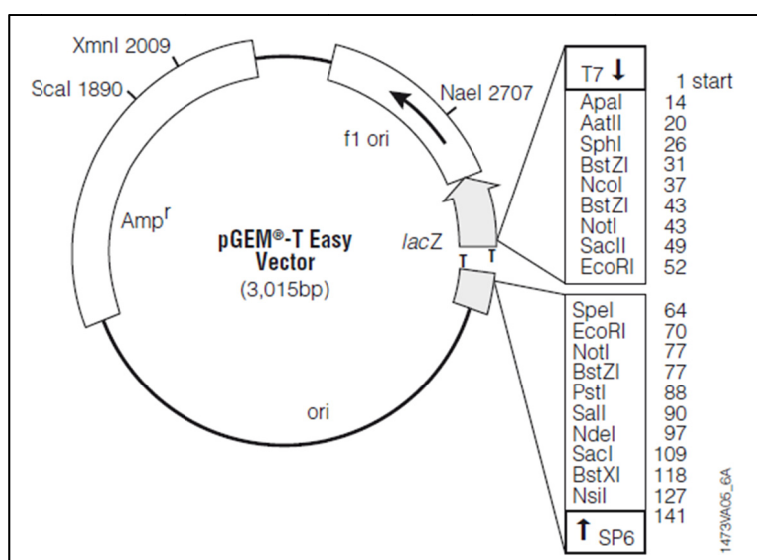


Figura 5. Vector pGEM®-T Easy^e

Medio LB-ampicilina

^e <https://www.promega.es/-/media/files/resources/protocols/technical-manuals/0/pgem-t-and-pgem-t-easy-vector-systems-protocol.pdf?la=en>

1L: 10 g bacto-tryptone, 5 g bacto- yeast extract, 5 g NaCl.

Añadir todo a un Erlenmeyer, agitar con imán, poner gasa + cinta marrón de esterilizar y poner en el autoclave. Cuando esté frío, añadir 1 mL de ampicilina 1:1000 o hasta concentración final de 100 µL/mL.

Medio LB-agar

1L de medio LB. Añadir 15 g de agar, autoclavar, añadir 1mL de ampicilina 1:1000 o hasta concentración final de 100 µl/mL. Añadir 30-35 mL del medio en cada placa petri.

LIGACIÓN

2x Rapid ligation Buffer, T4 DNA ligase	5 µL
pGEM- T (50 ng)	1 µL
PCR product (50 ng/mL)	1 µL
Agua estéril	2 µL
T4 DNA Ligase (3 weiss units/µl)	1 µL

Dejar toda la noche a 4°C o 1 hora a Tª ambiente.

Transformación

Adicionar 2 µl de la reacción de ligación a un eppendorf de 1,5 mL.

Agitar suavemente las células ultracompetentes.

Tomar 50 µl de células ultracompetentes y añadir las al eppendorf de 1,5 mL.

Dejar en hielo durante 20 minutos.

Preparar un baño a 42°C exactamente.

Tras el tiempo en hielo, introducir los eppendorf en el baño durante 45-50 segundos.

Inmediatamente volver a poner en hielo durante 2 minutos.

Adicionar 950 µl de medio SOC. Mezclar.

Incubar durante 1,5 hora a 37°C, con agitación (en el orbital).

Preparar placas de petri añadiendo: IPTG (isopropil-β-D-tio-galactósido), inductor artificial del operón *lac* y X-GAL (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside), sustrato de la β-galactosidasa que, al ser hidrolizado, produce un compuesto (indoxil) que en contacto con el aire presenta un intenso color azul.

Tras el tiempo de incubación, pipetear 100 y 200 µl de cada reacción de ligación en las placas de petri. Incubar durante 16-24 h a 37°C.

Selección de clones

Deben seleccionarse las colonias de color transparente. Éstas son las que han incluido el plásmido portador de la secuencia de interés.

Crecer dichas colonias en medio líquido LB-Ampicilina a 37°C y 225 rpm durante toda la noche (16h).

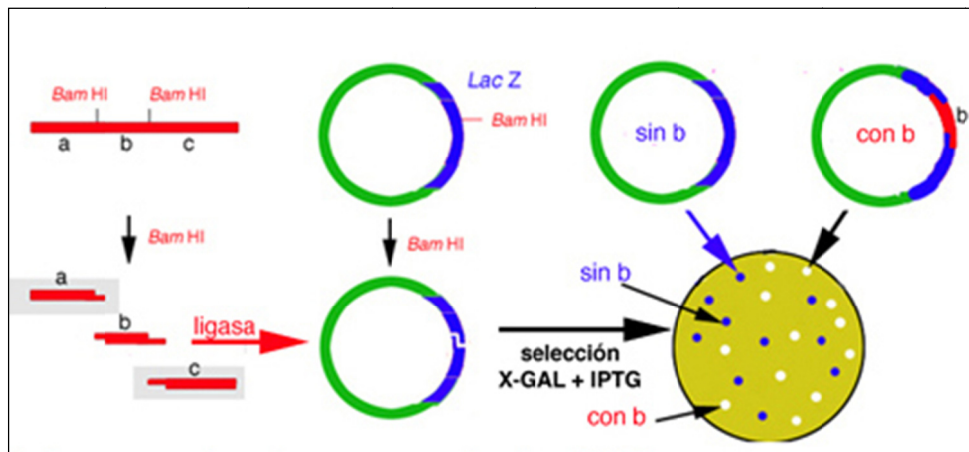


Figura 6. Esquema sistema selección de colonias

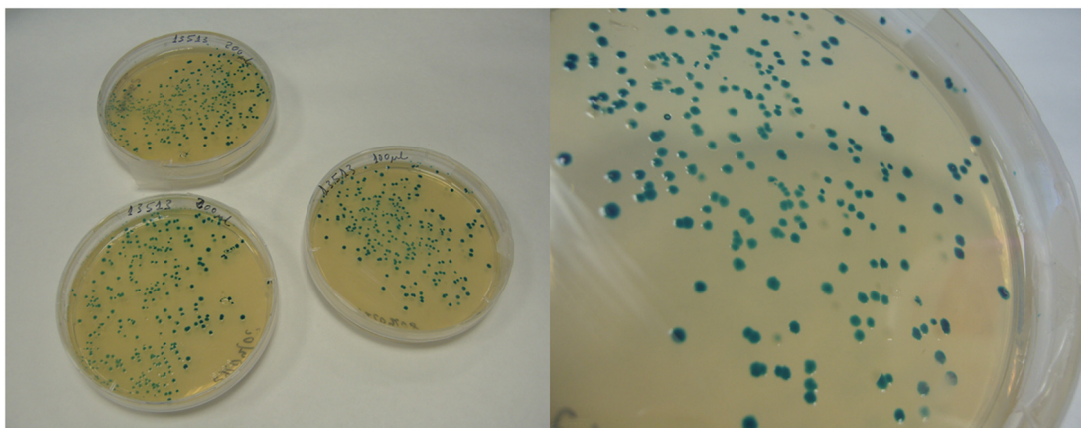


Figura 7. Placas agar obtenidas para selección de colonias

Minisequenciación

Para el panel de mutaciones puntuales se seleccionaron 16 posiciones del mtDNA implicadas en patologías frecuentes y una mutación frecuente en el gen *POLG* (gdNA) (Tabla 1).

Tabla 1. Mutaciones seleccionadas y enfermedades asociadas

Mutación	Gen	Cambio	Enfermedad
m.10158T>C	<i>MT-ND3</i>	p.S34P Ser>Pro	LS (<i>Leigh Syndrome</i>)
m.14459G>A	<i>MT-ND6</i>	p.A72V Ala>Val	LDYT/LS (<i>Leber's hereditary optic neuropathy and Dystonia / Leigh Syndrome</i>)
m.3460G>A	<i>MT-ND1</i>	p.A52T Ala>Thr	LHON (<i>Leber Hereditary Optic Neuropathy</i>)
m.10191T>C	<i>MT-ND3</i>	p.S45P Ser>Pro	LS-like/ESOC (<i>Leigh Syndrome like / Epilepsy, Strokes, Optic atrophy and Cognitive decline</i>)
m.14484T>C	<i>MT-ND6</i>	p.M64V Met>Val	LHON (<i>Leber Hereditary Optic Neuropathy</i>)
m.3243A>G	<i>MT-TL1</i>	tRNA Leu	MELAS, LS (<i>Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like episodes / Leigh Syndrome</i>)
m.11777C>A	<i>MT-ND4</i>	p.R340S Arg>Ser	LS (<i>Leigh Syndrome</i>)
m.8344A>G	<i>MT-TK</i>	tRNA Lys	MERRF (<i>Myoclonic epilepsy with ragged red fibers</i>)
m.13513G>A	<i>MT-ND5</i>	p.D393N Asp>Asn	MELAS/LS (<i>Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like episodes / Leigh Syndrome</i>)
m.14487T>C	<i>MT-ND6</i>	p.M63V Met>Val	LS/Dystonia (<i>Leigh Syndrome</i>)
m.11832G>A	<i>MT-ND4</i>	p.W358X Trp>Ter	EXIT (<i>Exercise Intolerance</i>)
m.13514A>G	<i>MT-ND5</i>	p.D393G Asp>Gly	MELAS/LS (<i>Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like episodes / Leigh Syndrome</i>)
m.11778G>A	<i>MT-ND4</i>	p.R340H Arg>His	LHON (<i>Leber Hereditary Optic Neuropathy</i>)
c.1399G>A	<i>Chr15-POLG</i>	p.A467T Ala>Thr	PEO, SANDO, MERRF, Ataxia-Neuropathy, ataxia-myopathy, Alpers, severe childhood multi-system disorder.
m.9176T>C	<i>MT-ATP6</i>	p.L217P Leu>Pro	LS/FBSN (<i>Leigh Syndrome</i>)
m.9176T>G	<i>MT-ATP6</i>	p.L217R Leu>Arg	LS (<i>Leigh Syndrome</i>)
m.14482C>A	<i>MT-ND6</i>	p.M64I Met>Ile	LHON (<i>Leber Hereditary Optic Neuropathy</i>)
m.14482C>G	<i>MT-ND6</i>	p.M64I Met>Ile	LHON (<i>Leber Hereditary Optic Neuropathy</i>)
m.8993T>G	<i>MT-ATP6</i>	p.L156R Leu>Arg	NARP (<i>Neurogenic muscle weakness, ataxia and retinitis pigmentosa</i>)
m.8993T>C	<i>MT-ATP6</i>	p.L156P Leu>Pro	LS/NARP (<i>Leigh Syndrome / Neurogenic muscle weakness, ataxia and retinitis pigmentosa</i>)

Cebadores de PCR y diseño de sondas (Tabla 2 y Tabla 3)

Para la detección de mutaciones puntuales se emplearon las condiciones anteriormente descritas⁵⁰ con las siguientes modificaciones: i) la inclusión de dos nuevos amplicones para las mutaciones m.8344A>G (*MT-TK tRNA Lys*) y c.1399G>A (*Mitochondrial DNA polymerase gamma, POLG EC 2.7.7.7*)⁵¹ y ii) el diseño de 5 nuevas sondas para la reacción de minisequenciación, la mayoría en la cadena pesada (H) para evitar la posible interferencia con polimorfismos descritos en MITOMAP (<http://www.mitomap.org>)⁵² y iii) ajustes en las concentraciones de cebadores y sondas. Los cebadores han sido sintetizados por Invitrogen (UK) y purificados mediante *cartridge* para eliminar secuencias erróneas. Las sondas fueron purificadas mediante *normal phase chromatography* para eliminar sales residuales.

Para la caracterización de haplogrupos mitocondriales se diseñaron desde el inicio nuevos amplicones. En ambos casos los cebadores han sido diseñados con el software Primer3 version 0.4.0⁵³ para obtener una temperatura de meeting (T_m) en torno a los 60°C y han sido verificados con BLAST (NCBI, USA) y el software SNPcheck v2 (Genetic Medicine, St Mary's Hospital Oxford Road, Manchester) para asegurar su especificidad y evaluar la presencia de polimorfismos en la secuencia de los cebadores. Por otro lado, también se ha verificado la posibilidad de estructuras en bucle y problemas de formación de dímeros de cebadores mediante el programa Autodimer (web-based version)⁵⁴. En el caso de las sondas se verificaron los polimorfismos cercanos para decidir la cadena más óptima y para diseñar sondas degeneradas en las posiciones polimórficas que no son evitables de ningún otro modo.

Tabla 2. Cebadores Multiplex PCR

Primers diseño mutaciones puntuales mtDNA				
Positions	Forward 5'-3'	Reverse 5'-3'	Amplicon Size (bp)	Conc. (μ M)
3243/3460	tataccacaccaccaag	gcggtgatgtagagggtgat	328	0.15 μ M
8344	accactttcaccgctacacg	atggtggccatacggtagt	272	0.20 μ M
8993/9176	aatgccctagcccacttctt	tcattaggagggtgagagg	396	0.20 μ M
10158/10191	tcaacaccctcctagcctta	gggtaaaaggagggaattt	197	0.40 μ M
11777/11778/11832	cacgggcttacatcctcatt	gggggtaaggcgaggtag	158	0.30 μ M
13513/13514	attggcagcctagcattagc	cagggaggtagcgatgagag	132	0.30 μ M
14459/14482/14484/14487	ctccatcgtaaccccacta	ttctgaattttgggggaggt	170	0.30 μ M
PolG c.1399G>A	cagaggcacagggcactta	aaggcctggctacctcttc	110	0.80 μ M
Cebadores diseño haplogrupo mtDNA (Multiplex PCR)				
Positions	Forward 5'-3'	Reverse 5'-3'	Amplicon Size (bp)	Conc. (μ M)
A663G	ccctaacaccagcctaacca	aggctaagcgttttgagctg	426 pb	0.2 μ M
T4216C, G4580A, A4917G	gattccgctacgaccaactc	caactgctgctatgatgga	828 pb	0.4 μ M
C5178A	caaaccagctacgaaaaat	gagaggagggtggatggaat	236 pb	0.2 μ M

C7028C	ggcctgactggcattgtatt	agcgaaggcttctcaaatca	378 pb	0.3 μ M
G9055A	actgcaggccactactcat	gaggagcggttatggagtgga	317 pb	0.2 μ M
A10398G, C10400T, T10873T	ccctaccatgagccctacaa	tactggataagtggcgttg	741 pb	0.2 μ M
A12308G	atgccccatgtctaacaac	tggctcagtgctcagttcgag	311 pb	0.2 μ M
G13708A, A13263G	ccccctagcagaaaatagcc	gcgagggctgtgagtttag	684 pb	0.3 μ M
C14766C	tctcgacggactacaacc	aggagtgagccgaagttca	180 pb	0.1 μ M

Tabla 3. Sondas (Reacción minsecuenciación)

Diseño mutaciones puntuales mtDNA					
Position	Locus	Probe 5'-3' (Extension Primer)	Strand	Probe Size	μ M
10158T>C	MT-ND3	acaactcaacggctacatagaaaa	L	25	0.10
3460G>A	MT-ND1	(GACT) ₂ ctcttggtgaagagttttatRg	H	31	0.60
14459G>A	MT-ND6	(GACT) ₂ GACctcaggatactctcaatagccatc	L	36	0.10
10191T>C	MT-ND3	(GACT) ₃ cgagtgcggttcgaccctata	L	34	0.20
14484T>C	MT-ND6	(GACT) ₃ atcgctgtagtatatccaaagacaacYa	L	40	0.50
3243A>G	MT-TL1	(GACT) ₄ Gaacagggtttgtaagatggcag	L	40	0.20
11777C>A	MT-ND4	(GACT) ₅ Gcaactacgaacgcactcacagt	L	44	0.20
8344A>G	MT-TK	(GACT) ₅ ggcatttcaactgtaaagaggtgtYgg	H	46	0.30
13513G>A	MT-ND5	(GACT) ₇ ttcctcacaggtttactccaaa	L	52	0.10
14487T>C	MT-ND6	(GACT) ₈ gttttttaatttatttaggggga	H	56	0.30
11832G>A	MT-ND4	(GACT) ₇ GACTcaactctactcccactaatagctttt	L	60	0.10
13514A>G	MT-ND5	(GACT) ₁₀ Ttgcggtttcgatgatgtgg	H	60	0.20
11778G>A	MT-ND4	(GACT) ₉ Gagaagtccttgagagaggattatgatg	H	64	0.40
c.1399G>A	Cr15-POLG	(GACT) ₁₀ gagatgaagaagtgcgttgatggatctg	L	67	0.30
9176T>C/G	MT-ATP6	(GACT) ₁₁ Gatccaagcctacgttttcacacttc	L	70	0.30
14482C>A/G	MT-ND6	(GACT) ₁₁ atcgctcgctgtagtatatccaaagacaac	L	73	0.70
8993T>G/C	MT-ATP6	(GACT) ₁₃ GcctactattcaaccaatagcYc	L	76	0.30
Diseño haplogrupo mtDNA (Multiplex PCR)					
Position	Probe 5'-3' (Extension Primer)		Strand	Probe Size	μ M
663A>G	(GACT)ccccataaacaataggtttgtcct		L	30	0.05
13263A>G	(GACT)GGaaatcgtagccttctccactcaagYca		L	34	0.20
5178C>A	(GACT)GAtggatggaattaagggttttagtcatgtta		H	36	0.05
7028T>C	(GACT) ₂ GAagctcctattgataggacatagtggagtg		H	40	0.10
14766C	(GACT) ₄ GTaccaatgacccaatacgcacaaa		L	42	0.15
4580G>A	(GACT) ₃ gatttttacctgagtaggcctagaaataaacat		L	46	0.20
12308A>G	(GACT) ₆ GActatccattggcttaggcccaa		L	50	0.10
9055G>A	(GACT) ₆ GAccactactcatgcacctaattggaaRc		L	54	0.05
13708G>A	(GACT) ₈ GAcctgcgaataggcttccggcYg		H	56	0.65
4917A>G	(GACT) ₇ tctcaatcatataccaaatctctccctcacta		L	60	0.20
4216T>C	(GACT) ₉ ctggagattgtaatgggtatggagacat		H	64	0.25
10400C>T	(GACT) ₇ agtcgaaatcattcgttttgttaaactatataccaaYtc		H	68	0.40
A10398A>G	(GACT) ₁₀ cctatgagtgactacaaaaaggattagactga		L	72	0.65
10873T	(GACT) ₁₁ GTcccacagcctaattattagcatcatccc		L	74	0.50

Amplificación mediante PCR multiplex

El primer paso es una PCR multiplex con 8 parejas de cebadores. La amplificación se realiza a partir de 5ng/ μ L de DNA genómico total con *DNA QIAGEN Kit Multiplex PCR* (Qiagen, Hilden, Germany)^f. La reacción consta de 12.5 μ L de 2x *QIAGEN Multiplex PCR Master Mix*, 2.5 μ L de *10x Primer Mix* (ver concentraciones finales en la Tabla 2), 9.5 μ L de agua RNase-free and 0.5 μ L del DNA molde (se obtiene un volumen final de reacción de 25 μ L). La PCR se realiza en un termociclador Verity[®] (Applied Biosystems) con las siguientes condiciones:

Activación HotStarTaq DNA Polymerase 95°C durante 15 minutos, Touchdown step de 5 ciclos 94°C de 30 segundos, 65°C a 61°C con rampa 1°C por ciclo durante 60 segundos, 72°C durante 30 segundos y después desnaturalización a 94°C durante 30 segundos, annealing 60°C durante 120 segundos, extensión 72°C durante 45 segundos un total de 30 ciclos y una extensión final de 15 minutos a 72°C.

Los productos de la PCR se tratan con ExoSAP-IT[®] (Affymetrix, Inc.) para eliminar cebadores libres y dNTPs. 3 μ L del producto de PCR se incuban durante 1 hora a 37°C con 1.5 μ L de ExoSAP-IT[®]. Finalmente se inactivan las enzimas a 80°C durante 15 minutos

Reacción de Minisequenciación

Este segundo paso consiste en una extensión de bases (*multiplex single-base primer extension*) mediante el kit SNaPshot (Applied Biosystems). La reacción tiene lugar a partir de 10 μ L que contienen 1 μ L del producto de PCR purificado en el primer paso, 3 μ L de *SNaPshot ready reaction mix*, 1 μ L de sondas y 5 μ L de agua destilada.

La reacción se lleva a cabo en un Verity[®] Thermocycler (Applied Biosystems) con condiciones estándar: 96°C durante 10 segundos, 50°C durante 5 segundos, 60°C durante 30 segundos en un total de 25 ciclos. El producto resultante (10 μ L) se trata con 1 μ L de *Shrimp Alkaline Phosphatase* (SAP, Affymetrix, Inc.) durante 1 hora a 37°C y 15 minutos a 75°C para la inactivación de la enzima. La defosforilación de los dNTPs no incorporados, que lleva a cabo esta fosfatasa, altera su migración y previene interferencias en la electroforesis capilar.

^f <https://www.qiagen.com/us/resources/resourcedetail?id=a541a49c-cd06-40ca-b1d2-563d0324ad6c&lang=en>

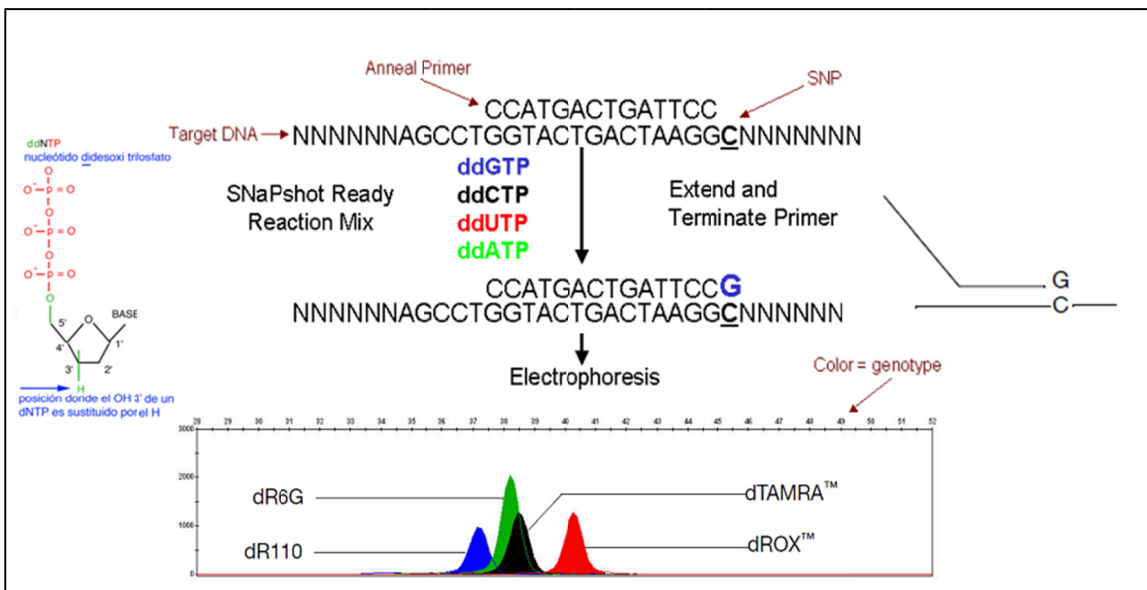


Figura 8. Esquema reacción Minisequenciación⁸

Electroforesis Capilar (Análisis de fragmentos)

Una alícuota (1.5µL) del producto de la reacción SNaPshot se mezcla con 10µL de *HiDi™ formamide* (Applied Biosystems) y 25µL de *GeneScan 120 LIZ ladder* (Applied Biosystems). Esta mezcla se desnaturaliza a 95°C durante 5 minutos y se deja en hielo antes de su carga en el secuenciador usando polímero POP-7 y un capilar de 50cm (*ABI PRISM 3130xl Genetic Analyzer*, Applied Biosystems). Los datos obtenidos se analizan con GeneMapper v4.0 (Applied Biosystems).

Optimización

Se ajustó la concentración final de cada pareja de cebadores, las condiciones de PCR y la concentración del DNA de partida para obtener una concentración similar de cada amplicón (aproximadamente 2-5 ng/µL tras el tratamiento con ExoSAP-IT®).

El volumen del producto empleado en la reacción de SNaPshot y la concentración de las sondas también se ajustaron para obtener unos picos en el electroferograma entre 2000-7000 unidades relativas de fluorescencia (rfus).

Para la asignación e interpretación del haplogrupo mitocondrial se ha creado una tabla Excel que indica de manera automática el haplogrupo tras la introducción del color de los picos del electroferograma.

⁸ https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_042107.pdf

PCR-RFLP con cuantificación microfluídica

Procedimiento PCR-RFLP-Bioanalizador 2100 (Agilent)

- 1.- Se realiza una PCR convencional.
- 2.- Digestión con enzima de restricción (RFLP).
- 3.- Electroforesis capilar y visualización con Bioanalizador 2100 (según protocolo fabricante).

Procedimiento *Last cycle Hot PCR*⁵⁵

- 1.- Se realiza una PCR convencional.
- 2.- Se para en el último ciclo, se añade dCTP* marcado con ³²P (en un volumen final de 30 μ L se añaden 0,3 μ L de dCTP*).
- 3.- Se realiza un último ciclo y una extensión larga (20 min).
- 4.- Digestión con enzima de restricción (RFLP).
- 5.- Visualización en gel de poliacrilamida (10%).

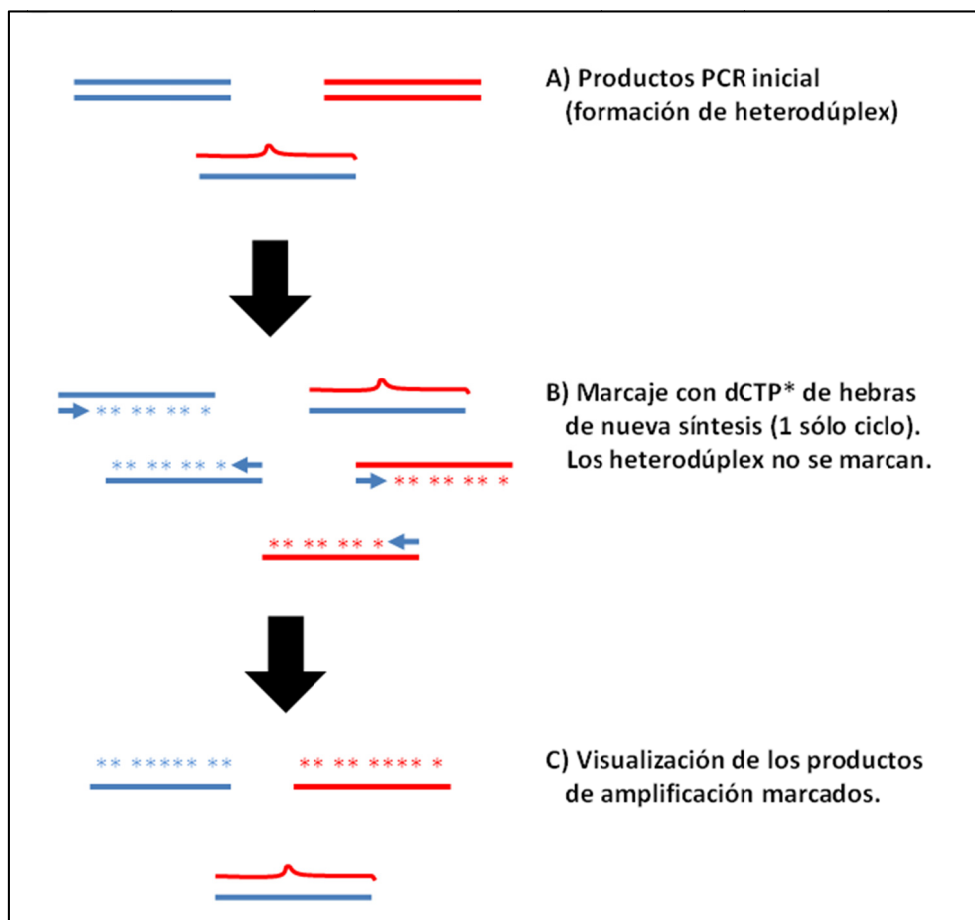


Figura 9. Esquema *Last cycle Hot PCR*

Estudio de la integridad del mtDNA

Long-range PCR

El mtDNA se amplifica en un primer paso a partir de 5-10ng de gDNA en un solo amplicón mediante *SequalPrep™ Long PCR Kit (Life Technologies)* siguiendo las recomendaciones del fabricante^h. También se ha ensayado otra polimerasa (*Takara LA PCR kit*) siguiendo un protocolo previamente descrito⁵⁶ con resultados equivalentes.

Cebadores empleados: mt16426F (5-CCGCACAAGAGTGCTACTCTCTC-3); mt16425R (5-GATATTGATTCACGGAGGATGGTG-3)

Tabla 4. Master Mix *SequalPrep™ Long PCR*

Componentes	Volumen (µL) X1 muestra	Concentración final en PCR
SequalPrep™ Reaction Buffer	2 µL	1x
DMSO	0,4 µL	-
SequalPrep™ 10x Enhancer B	2 µL	1x
SequalPrep™ Long Polimerasa, 5U/µL	0,36 µL	1,8 U
DNAasa- free H2O	4,24 µL	-
Primer mt16426F (2 pmol)	5 µL	0,5 pmol
Primer mt16425R (2 pmol)	5 µL	
TOTAL	19 µL	-
DNA muscle (5ng)		
DNA blood (50- 100 ng)	1 µL	

Tabla 5. Condiciones PCR *SequalPrep™ Long PCR* mt16426F y mt16425R

Activación	Desnat	Anneal	Desnat	Anneal	Extend FINAL	FINAL
	10 ciclos		20 ciclos			
94°C/2 min	94°C/10s	68°C/17min	94°C/10s	68°C/17min Δ20s/ciclo	72°C/5min	4°C/∞

El producto de PCR se cuantifica mediante un fluorímetro Qubit 2.0 para ajustar la concentración a 100ng necesaria en los pasos posteriores de generación de librerías (Pág.26).

Nota: La PCR se puede realizar también con otra pareja de primers 3361F-3360R (2 pmol) en caso de que no amplifique o aparezca delección de 12,8kb⁵⁷. En este caso habría que secuenciar por Sanger la región que comprende dichos primers.

Cebadores empleados: mt3361F (5-TTCCTAATGCTTACCGAACGAA -3); mt3360R (5-TGCCATTGCGATTAGAATGGG-3)

^h https://assets.thermofisher.com/TFS-Assets/LSG/manuals/sequalprep_man.pdf

Tabla 6. Condiciones PCR *SequalPrep™ Long PCR* mt3361F y mt3360R

Activación	Desnat	Anneal	Extend	Desnat	Anneal	Extend	Extend FINAL	FINAL
	10 ciclos			20 ciclos				
94°C/2 min	94°C/10s	56°C /17min	68°C /17min	94°C/10s	56°C/30s	68°C/17min Δ20s/ciclo	72°C/5min	4°C/∞

Southern blot

Ésta técnica⁵⁸ se llevó a cabo en DNA extraído de tejido (principalmente músculo) usando el método no radiactivo de *Boehringer Mannheim*. Para linealizar el DNA se emplea la enzima de restricción PvuII. El gel de agarosa se prepara al 0,8% (SEAKEM GTG®). Las bandas se visualizan mediante marcaje con digoxigenina (*Roche Diagnostics*).

Estudio número de copias de mtDNA (Depleción)

La cuantificación relativa del número de copias del mtDNA frente al nDNA se realizó en DNA total extraído de biopsia muscular mediante real-time PCR usando sondas TaqMan frente al gen rRNA 12S mitocondrial (*MT-RNR1*) y frente al gen nuclear *RNase P* en un sistema *HT7500 Real-Time PCR System* (Applied Biosystems, USA) como se ha descrito anteriormente⁵⁹ y según consta en el procedimiento técnico del laboratorio PT-BIO-BM-MIT-01. Los resultados se informan respecto al mismo tejido de controles de igual edad, según se ha descrito anteriormente⁶⁰.

Secuenciación masiva del mtDNA en la plataforma PGM (*Personal Genome Machine*)

Actualmente existen varias guías con recomendaciones para el desarrollo y análisis de resultados de NGS (*Next-Generation Sequencing*)⁶¹⁻⁶⁴. En nuestro caso se han tenido en cuenta los estándares de laboratorio clínico indicados por ACMG⁶³ y las recomendaciones aplicables en la plataforma PGM.

La secuenciación completa del mtDNA a alta profundidad parte de un DNA que proviene de una PCR larga (*Long-range PCR*) tal y como se describe en el apartado correspondiente anterior (pág. 38). Las librerías (colecciones de fragmentos) se han generado mediante el kit *Ion Xpress plus Fragment Library* (*Life Technologies*) para 200pb y para 400pb. Las purificaciones se han realizado con AMPure beads y la comprobación de los tamaños de los fragmentos y su concentración se han determinado mediante un *High Sensivity Bioanalyzer 2100 chips* (*Agilent*). La PCR en emulsión (*Ion One Touch 2, Life Technologies*) y la secuenciación se realizaron siguiendo las

recomendaciones del fabricante. Cada carrera se realizó con chips v316 cargados con 10 muestras para obtener una cobertura media superior a 1000X.

El análisis de los resultados se llevó a cabo en Torrent Server (TS) usando como alineador TMAP, y los plugins Variantcaller (v4.2.1.0 a v5.10.0.18), VariantcallerformtDNA (v3.0.0) y Coverageanalysis (v.4.2.1.4 a v5.10.0.3). El Variantcaller se ha usado con los parámetros ajustados para línea somática para frecuencias alélicas mayores del 2,5% y con alta sensibilidad. El parámetro *downsample coverage* se ha elevado también a 10000 para evitar la pérdida de información. Por otro lado se creó un fichero .bed con para nombrar los genes presentes en el mtDNA y un fichero de “hotspots” con las 19 mutaciones más frecuentes en el mtDNA para que el sistema pueda mostrar información sobre esas posiciones independientemente de si la mutación está presente o no.

La anotación biológica de estos resultados se realizó adaptando las bases de datos MITOMAPⁱ y MitImpact³⁸ mediante *scripts* propios (Tabla 7).

La detección de haplogrupos se realizó mediante el plugin VariantcallerformtDNA (v3.0.0). Este programa, integrado en Torrent Server, genera un análisis usando librerías de Haplogrep³⁶ y una secuencia consenso generada para búsquedas EMPOP (EDNAP mtDNA Population Database)^j.

Secuenciación masiva del mtDNA en la plataforma MiSeq

La secuenciación en esta plataforma se llevó a cabo en colaboración con la Dra. García Arumi en el Hospital Vall d’Hebrón. Las muestras se prepararon con Nextera XT DNA Sample Preparation kit (Illumina) siguiendo las recomendaciones del fabricante^k para fragmentos de hasta 150pb y lecturas pareadas (pair-end). Para preparar cada librería se necesita 1ng de mtDNA. Los productos de PCR se purifican con Ampure beads XT (Beckman Coulter), se normalizan y se unifican para proceder con MiSeq Reagent kit V2 (300 ciclos, 2x150) y secuenciar en la plataforma MiSeq. Normalmente se unifican unas 40 librerías para obtener una cobertura media de 5000X. El software del secuenciador convierte los ficheros de imágenes brutos en ficheros BAM y se crean finalmente los ficheros VCF con GATK tools^{65,66}. Estos ficheros de variantes se analizan con MiSeq Reporter Software^l.

ⁱ <https://www.mitomap.org//MITOMAP>

^j <https://empop.online/>

^k https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-reference-guide-15031942-03.pdf

^l <https://emea.illumina.com/systems/sequencing-platforms/miseq/products-services/miseq-reporter.html?langsel=/es/>

Preparación de ficheros para análisis de resultados de secuenciación masiva

Para generar las base de anotación de mtDNA se seleccionaron tres bases de datos y se adaptaron (*parsing*) para anotar los resultados provenientes de TS: *mtDNA Variants (Coding y RNA Variants, N=8954)* y *mtDNA Mutations with reports of disease-Associations (Coding & Control Region Mutations, N=281 y rRNA/tRNA Mutations, N=301)*. Además se incluyó la nueva base de predictores de patogenicidad MitImpact³⁸, que se puede descargar libremente en formato texto, para evaluar el efecto sobre los polipéptidos codificados por el mtDNA.

La adaptación de las bases de datos descargadas de MITOMAP realizó mediante el editor de texto Notepad++^m y sobre una terminal GNU-Linux en UBUNTU 18.04 usando los lenguajes de programación Bash (*Bourne again shell*), AWK y otras aplicaciones de GNU⁶⁷. Los comandos básicos empleados han sido: sed, cut, paste y sort. En la Tabla 7 se muestran los comandos empleados y las Figuras 10, 11, y 12 muestran una captura de los resultados obtenidos.

Tabla 7. Comandos para la adaptación de bases de MITOMAP, MitImpact y MitoTIP

PREPARACIÓN	base	SNPs	"PolymorphismsCoding	MITOMAP	MITOWEB"	y
"MutationsCodingControl MITOMAP MITOWEB"						
# Base 1 (Polimorfismos)						
sed 1d mitosnp.txt > dbA; cut -f1,3 dbA > dbA.2; sed -e 's/-/_/g' dbA.2 sed -e 's/\t/_/g' > dbA.3; paste dbA.3 dbA sed 's/ /_/g' > dbA.4; cut -f 1,2,7,8 dbA.4> dbA.5; sort -k1 -d dbA.5 > dbA_final						
# Base 2 (Mutaciones genes) Cambiar los "espacio / espacio" por "/"						
awk -F'\t' -v OFS="\t" '{print \$1,\$2,\$5,\$6,\$10,\$9,\$7,\$8,\$3}' mitocoding.txt > mitocoding ; sed 1d mitocoding > dbB.1; cut -f1,3 dbB.1 > dbB.2; sed -e 's/-/_/g' dbB.2 sed -e 's/\t/_/g' > dbB.3; paste dbB.3 dbB.1 sed 's/ /_/g' > dbB.4; cut -f 1,2,5,6,7,8,9,10 dbB.4> dbB.5; sort -k1 -d dbB.5 > dbB_final						
#Base 3 (Mutaciones RNA) Cambiar los "espacio / espacio" por "/", los "espacio & espacio" por "&" y los "espacios" por "_"						
awk -F'\t' -v OFS="\t" '{print \$1,\$2,\$4,\$5,\$9,\$8,\$6,\$7,\$3}' mitorna.txt > mitorna ; sed 1d mitorna > dbC.1						
# Tiene allele (col3) con número y letra.						
cut -f3 dbC.1 > dbC.2; sed 's/^[ATCG][a-z]*/& /;s/[0-9][0-9]*/& /' dbC.2 > dbC.3; awk '{print \$2,\$1,\$3}' dbC.3 > dbC.4; sed -e 's/ /_/g' dbC.4> dbC.5; paste dbC.5 dbC.1 sed 's/ /_/g' > dbC.6; cut -f 1,2,5,6,7,8,9,10 dbC.6> dbC.7; sort -k1 -d dbC.7 > dbC_final						
# Base 4 (Predicciones MitImpact). A partir de fichero con selección de columnas y sustituyendo "/" por "-"						
cut -f 1,2 mut.txt sed -e 's/\t/_/g' - sed -e 's/-/_/g' > dbD.1; paste dbD.1 mut.txt cut -f1,4-26 sort -k1 -d - > dbD_final						
# Base 5 (MitoTIP). Sustituir espacio por \0						

^m <https://notepad-plus-plus.org/>


```
cut -f 1-3 MitoTIP_raw.txt | sed -e 's/\t/_/g' - > dbE.1; paste dbE.1
MitoTIP_raw.txt | cut -f1,4,5 | sort -k1 -d - > dbE_final
```

Se añade finalmente una columna con las categorías de patogenicidad en función del Score.

Después de realizar las transformaciones anteriores se detectaron duplicidades entre bases que se corrigieron mediante los comandos siguientes (se eliminan de la base de polimorfismos las variantes que aparecen duplicadas en las bases de mutaciones):

```
# Eliminar duplicados entre bases
awk 'FNR==NR{a[$1];next};!($1 in a)' dbB_final dbA_final > dbA_final_filtered.0
awk 'FNR==NR{a[$1];next};!($1 in a)' dbC_final dbA_final_filtered.0 >
dbA_final_filtered
```

1	10086_A_G	10086	N-D 127	Reported	+	-	Hypertensive_end-stage_renal_disease	
2	10156_T_C	10158	S-F 0	Cfzm	+	+	Leigh_Disease	
3	10191_T_C	10191	S-F 0	Cfzm	-	+	Leigh_Disease/Leigh-like_Disease/ESOC	
4	10197_G_A	10197	A-T 4	Cfzm	+	+	Leigh_Disease/Dystonia/Stroke/LDXT	
5	10237_T_C	10237	I-T 42	Reported	+	-	LHON	
6	10398_A_A	10398	I-T 0	"Reported;haplogroup_HN1UUVNWX2_marker"	+	-	"Invasive_Breast_Cancer_risk_factor;AD;PD;BD_lithium_response;Type_2_DM"	
7	10398_A_G	10398	T-A 11313	"Reported;haplogroup_IJK_marker"	+	-	PD_protective_factor/longevity/altared_cell_ph/metabolic_syndrome/breast_cancer_risk	
8	10543_A_G	10543	R-R 0	Reported	-	+	LHON	
9	10591_T_G	10591	F-C 0	Reported	-	+	LHON	
10	10652_T_C	10452	syn 42	Reported	-	+	BD/MDD-associated	
11	10663_T_C	10463	V-A 1	Cfzm	+	-	LHON	
12	10680_G_A	10480	A-T 11	Reported_possibly_synergistic	+	-	LHON	
13	11084_A_G	11084	I-A 146	"Reported;F.M."	+	+	"AD;PD;MELAS"	
14	11232_T_C	11232	L-F 0	Reported	-	+	CFEO	
15	11253_T_C	11253	I-T 170	Reported	+	+	"LHON;PD"	
16	11365_T_C	11365	syn 74	Reported	-	-	found_in_1_HCM_patient	
17	11467_A_G	11467	syn 3581	Reported	+	-	Altered_brain_ph	
18	114_C_T_114	norcoding	131	Reported	+	-	BD-associated	
19	11622_TA_del	11622	frameshift	0	Reported	-	+	CFEO_exercise_intolerance
20	11696_G_A	11496	V-I 153	Reported_possibly_synergistic	+	+	LHON/LDXT/DEAF/hypertension_helper_mut.	
21	11777_C_A	11777	R-S 0	Cfzm	-	+	Leigh_Disease	
22	11778_G_A	11778	R-H 113	Cfzm	+	+	LHON/Progressive_Dystonia	
23	11832_G_A	11832	W-Ter	0	Reported	-	+	EXIT/oncocytoma
24	11874_C_A	11874	T-N 0	Reported	+	-	LHON	
25	11919_C_T	11919	S-F 0	Reported	+	-	Thyroid_Cancer_Cell_Line	
26	11994_C_T	11994	T-I 0	Conflicting_reports	+	-	OAT	
27	12026_A_G	12026	I-V 148	Reported	+	-	DM	
28	12027_T_C	12027	I-T 2	Reported	.	.	SZ-associated	
29	12038_T_C	12038	M-T 64	Conflicting_reports	+	-	DEAF1555_increased_penetrance/LHON	
30	12363_A_G	12361	F-A 163	Reported	-	-	Nonalcoholic_fatty_liver_disease	
31	12372_G_A	12372	syn 3832	Reported	+	-	Altered_brain_ph	
32	12397_A_G	12397	T-A 134	Reported	+	-	PD_early_onset	
33	12425_A_del	12425	N-frameshift	0	Reported	-	+	Mitochondrial_Myopathy_i_Renal_Failure
34	12477_T_C	12477	syn 173	Reported	+	-	possible_HCM_susceptibility	
35	12622_G_A	12422	V-I 3	Significance_unclear	+	+	Leigh_Disease	
36	12634_A_G	12434	I-V 74	Reported	+	-	Thyroid_Cancer_Cell_Line	
37	12706_T_C	12706	F-L 0	Cfzm	-	+	Leigh_Disease	
38	12770_A_G	12770	E-G 0	Reported	-	+	MELAS	
39	12782_T_G	12782	I-S 0	Reported	-	+	LHON	
40	12811_T_C	12811	Y-H 214	Reported	+	-	Possible_LHON_factor	
41	12848_C_T	12848	A-V 0	Reported	-	+	LHON	
42	13042_G_A	13042	A-T 1	Reported	-	+	Optic_neuropathy/retinopathy/LD	
43	13045_A_C	13045	M-L 0	Reported	-	+	MELAS/LHON/Leigh_overlap_syndrome	
44	13051_G_A	13051	G-S 0	Reported	-	+	LHON	
45	13063_G_A	13063	V-I 1	Reported	-	+	Adult-onset_Encephalopathy/Ataxia	
46	13084_A_T	13084	S-C 0	Reported	-	+	MELAS/Leigh_Disease	
47	13094_T_C	13094	V-A 0	Reported	+	+	Ataxia+PEO/MELAS_LD_myoclonus_fatigue	
48	13135_G_A	13135	A-S 247	Reported	+	-	possible_HCM_susceptibility	
49	13271_T_C	13271	L-F 1	Reported	-	+	Exercise_intolerance_(EXII)	
50	13282_T_C	13282	M-A 0	Reported	-	+	PD	

Figura 10. Base de datos Mitomap adaptada

Generación de filtro para variantes comunes propias

Existen diversas variantes que aparecen de manera recurrente en gran parte de los pacientes debido a sesgos de secuenciación o a variantes frecuentes en la población objeto de estudio. Para facilitar y simplificar el análisis de datos se creó un fichero con las variantes que se repiten en más del 25% de los pacientes secuenciados para realizar un filtrado previo. Para ello se parte de los ficheros anotados sin filtro y se realizan diversas transformaciones mediante la terminal de GNU-Linux:

Tabla 8. Comandos para creación de filtro de variantes frecuentes

```
# Buscar variantes comunes en ficheros anotados
Anotar todos los VCFs sin filtro con Annotar
# Preparar ficheros (repetir en cada fichero)
awk '{print col $10,$1,$2,$3,$4,$5}' 01.FINAL > file.01 ; sed 1d file.01 >
filen.01
# Unir ficheros y obtener número de variantes iguales
cat filen.* > union
sort -k1 -f union > ordenado
uniq -d -c ordenado repetidos
sort -k1nr repetidos > final
# Preparar el filtro. Eliminar filas con pocas repeticiones (Ejemplo: de la línea
78 a la 92)
sed '78,92d' final > filter
awk '{print col $2,$3,$4,$5,$6}' filter > filter.1
sed -e 's/\s/./g' filter.1 > filter.final
```

Detección de grandes deleciones en mtDNA y análisis de cobertura

Para la detección de deleciones únicas se desarrolló un *script* en R⁶⁸ que se integró y automatizó con Perl. Los resultados se muestran en un fichero PDF generado con *R Markdown*⁶⁹. Éste entorno de trabajo permite generar informes con gran flexibilidad, tanto estáticos como dinámicos en diversos formatos (HTML, PDF) y permite integrar diversos lenguajes de programación (R, Python, Perl).

El *script* de R consta de dos partes: una que lanza el proceso en cada muestra y otra que realiza los cálculos y genera el fichero PDF de resultados.

Por otro lado, se crearon para el TS dos ficheros en formato .bed con información acerca del mtDNA: para nombrar regiones (genes) y para *Hotspots*, que aportan más información a los datos

emitidos por el TS y permiten informar siempre de las posiciones indicadas como mutaciones puntuales frecuentes (Tabla 9 y Tabla 10).

Tabla 9. Fichero bed “Torrent Server Targeted regions mtDNA”

track type=bedDetail ionVersion=4.0 name="mtDNA_design" description="CoveredBases_mtDNA" color=77,175,74 priority=2							
chrM	0	109	1-109	0	+	.	GENE_ID=D-loop
chrM	109	441	110-441	0	+	.	GENE_ID=MT-OHR
chrM	441	576	442-576	0	+	.	GENE_ID=D-loop
chrM	576	647	577-647	0	+	.	GENE_ID=tRNA-Phe
chrM	647	1601	648-601	0	+	.	GENE_ID=rRNA-12S
chrM	1601	1670	1602-1670	0	+	.	GENE_ID=tRNA-Val
chrM	1670	3229	1671-3229	0	+	.	GENE_ID=rRNA-16S
chrM	3229	3304	3230-3304	0	+	.	GENE_ID=tRNA-Leu
chrM	3304	3306	3305-3306	0	+	.	GENE_ID=non-coding-nucleotides
chrM	3306	4262	3307-4262	0	+	.	GENE_ID=MT-ND1
chrM	4262	4328	4263-4328	0	+	.	GENE_ID=tRNA-Ile
chrM	4328	4331	4329-4331	0	+	.	GENE_ID=tRNA-Ile-Gln
chrM	4331	4400	4332-4400	0	+	.	GENE_ID=tRNA-Gln
chrM	4400	4401	4401	0	+	.	GENE_ID=non-coding-nucleotides
chrM	4401	4469	4402-4469	0	+	.	GENE_ID=tRNA-Met
chrM	4469	5511	4470-5511	0	+	.	GENE_ID=MT-ND2
chrM	5511	5579	5512-5579	0	+	.	GENE_ID=tRNA-Trp
chrM	5579	5586	5580-5586	0	+	.	GENE_ID=non-coding-nucleotides
chrM	5586	5655	5587-5655	0	+	.	GENE_ID=tRNA-Ala
chrM	5655	5656	5656	0	+	.	GENE_ID=non-coding-nucleotides
chrM	5656	5720	5657-5720	0	+	.	GENE_ID=tRNA-Asn
chrM	5720	5729	5721-5729	0	+	.	GENE_ID=tRNA-Asn_MT-ORL
chrM	5729	5760	5730-5760	0	+	.	GENE_ID=MT-ORL
chrM	5760	5798	5761-5798	0	+	.	GENE_ID=tRNA-Cys_MT-ORL
chrM	5798	5825	5799-5825	0	+	.	GENE_ID=tRNA-Cys
chrM	5825	5826	5826	0	+	.	GENE_ID=tRNA-Cys-Tyr
chrM	5826	5891	5827-5891	0	+	.	GENE_ID=tRNA-Tyr
chrM	5891	5903	5892-5903	0	+	.	GENE_ID=non-coding-nucleotides
chrM	5903	7445	5904-7445	0	+	.	GENE_ID=MT-CO1
chrM	7445	7514	7446-7514	0	+	.	GENE_ID=tRNA-Ser
chrM	7514	7517	7515-7514	0	+	.	GENE_ID=non-coding-nucleotides
chrM	7517	7585	7518-7585	0	+	.	GENE_ID=tRNA-Asp
chrM	7585	8269	7586-8269	0	+	.	GENE_ID=MT-CO2
chrM	8269	8294	8270-8294	0	+	.	GENE_ID=non-coding-nucleotides
chrM	8294	8364	8295-8364	0	+	.	GENE_ID=tRNA-Lys
chrM	8364	8365	8365	0	+	.	GENE_ID=non-coding-nucleotides
chrM	8365	8526	8366-8526	0	+	.	GENE_ID=MT-ATP8
chrM	8526	8572	8527-8572	0	+	.	GENE_ID=MT-ATP8-ATP6
chrM	8572	9206	8573-9206	0	+	.	GENE_ID=MT-ATP6
chrM	9206	9207	9207	0	+	.	GENE_ID=MT-ATP6_CO3
chrM	9207	9990	9208-9990	0	+	.	GENE_ID=MT-CO3
chrM	9990	10058	9991-10058	0	+	.	GENE_ID=tRNA-Gly
chrM	10058	10404	10059-10404	0	+	.	GENE_ID=MT-ND3
chrM	10404	10469	10405-10469	0	+	.	GENE_ID=tRNA-Arg
chrM	10469	10759	10470-10759	0	+	.	GENE_ID=MT-ND4L
chrM	10759	10766	10760-10766	0	+	.	GENE_ID=MT-ND4L-ND4
chrM	10766	12137	10767-12137	0	+	.	GENE_ID=MT-ND4
chrM	12137	12206	12138-12206	0	+	.	GENE_ID=tRNA-His
chrM	12206	12265	12207-12265	0	+	.	GENE_ID=tRNA-Ser
chrM	12265	12336	12266-12336	0	+	.	GENE_ID=tRNA-Leu
chrM	12336	14148	12337-14148	0	+	.	GENE_ID=MT-ND5
chrM	14148	14673	14149-14673	0	+	.	GENE_ID=MT-ND6
chrM	14673	14742	14674-14742	0	+	.	GENE_ID=tRNA-Glu

chrM	14742	14746	14743-14746	0	+	.	GENE_ID=non-coding-nucleotides
chrM	14746	15887	14747-15887	0	+	.	GENE_ID=MT-CYB
chrM	15887	15953	15888-15953	0	+	.	GENE_ID=tRNA-Thr
chrM	15953	15955	15954-15955	0	+	.	GENE_ID=non-coding-nucleotides
chrM	15955	16023	15956-16023	0	+	.	GENE_ID=tRNA-Pro
chrM	16023	16569	16024-16569	0	+	.	GENE_ID=D-loop

Tabla 10. Fichero bed “Torrent Server Hotspots mtDNA”

track type=bedDetail name="mtDNA_HotSpots" description="mtDNA_Mutations"							
chrM	1554	1555	DEAF	0	+	REF=A;OBS=G;ANCHOR=G	rRNA-12S
chrM	3242	3243	MELAS-LS	0	+	REF=A;OBS=G;ANCHOR=G	tRNA-leu
chrM	3459	3460	LHON	0	+	REF=G;OBS=A;ANCHOR=C	MT-ND1
chrM	8343	8344	MERRF	0	+	REF=A;OBS=G;ANCHOR=A	tRNA-lys
chrM	8992	8993	LS-NARP	0	+	REF=T;OBS=C;ANCHOR=C	MT-ATP6
chrM	8992	8993	NARP	0	+	REF=T;OBS=G;ANCHOR=C	MT-ATP6
chrM	9175	9176	LS	0	+	REF=T;OBS=G;ANCHOR=C	MT-ATP6
chrM	9175	9176	LS-FBSN	0	+	REF=T;OBS=C;ANCHOR=C	MT-ATP6
chrM	10157	10158	LS	0	+	REF=T;OBS=C;ANCHOR=A	MT-ND3
chrM	10190	10191	LS-ESOC	0	+	REF=T;OBS=C;ANCHOR=A	MT-ND3
chrM	11776	11777	LS	0	+	REF=C;OBS=A;ANCHOR=T	MT-ND4
chrM	11777	11778	LHON	0	+	REF=G;OBS=A;ANCHOR=C	MT-ND4
chrM	11831	11832	EXIT	0	+	REF=G;OBS=A;ANCHOR=T	MT-ND4
chrM	13512	13513	MELAS-LS-LHON.MELAS	0	+	REF=G;OBS=A;ANCHOR=A	MT-ND5
chrM	13513	13514	MELAS-LS	0	+	REF=A;OBS=G;ANCHOR=G	MT-ND5
chrM	14458	14459	LDYT-LS	0	+	REF=G;OBS=A;ANCHOR=C	MT-ND6
chrM	14481	14482	LHON	0	+	REF=C;OBS=A;ANCHOR=C	MT-ND6
chrM	14483	14484	LHON	0	+	REF=T;OBS=C;ANCHOR=A	MT-ND6
chrM	14486	14487	LS-DIST-ATAXIA	0	+	REF=T;OBS=C;ANCHOR=A	MT-ND6

Estudio de mtDNA residual en datos de secuenciación de exomas

Se utilizó en este caso como herramienta de análisis el servidor *Galaxy Server*^{70–72} (<https://main.g2.bx.psu.edu/>). Se importaron en esta plataforma los ficheros FASTQ provenientes de la secuenciación de exomas en plataformas de Illumina. Se aplicó para el análisis un *workflow* general (publicado en <http://usegalaxy.org/heteroplasmy>) con modificaciones para poder identificar cambios con una baja cobertura (Figura 13).

Los pasos iniciales del proceso son el alineamiento con el algoritmo BWA usando como referencia el genoma “Homo_sapiens_nuHg19_mtrCRS” (hg19 con el mtDNA cambiado por rCRS)^{73,74}. En este punto se empleó el software *Qualimap v0.6* para evaluar los resultados de alineador⁷⁵.

Posteriormente se seleccionaron las lecturas con una calidad aceptable (phred scale >30) que mapeaban adecuadamente. Los pasos siguientes (hasta 47) transforman los datos mediante filtros para unirlos en una lista que contiene las variantes alélicas que presentan una frecuencia superior a 0,01 como se ha descrito previamente⁷⁶. Finalmente cada posición se anotó con el nombre que presenta el gen en el mtDNA de referencia (rCRS NC_012920, Figura 14).

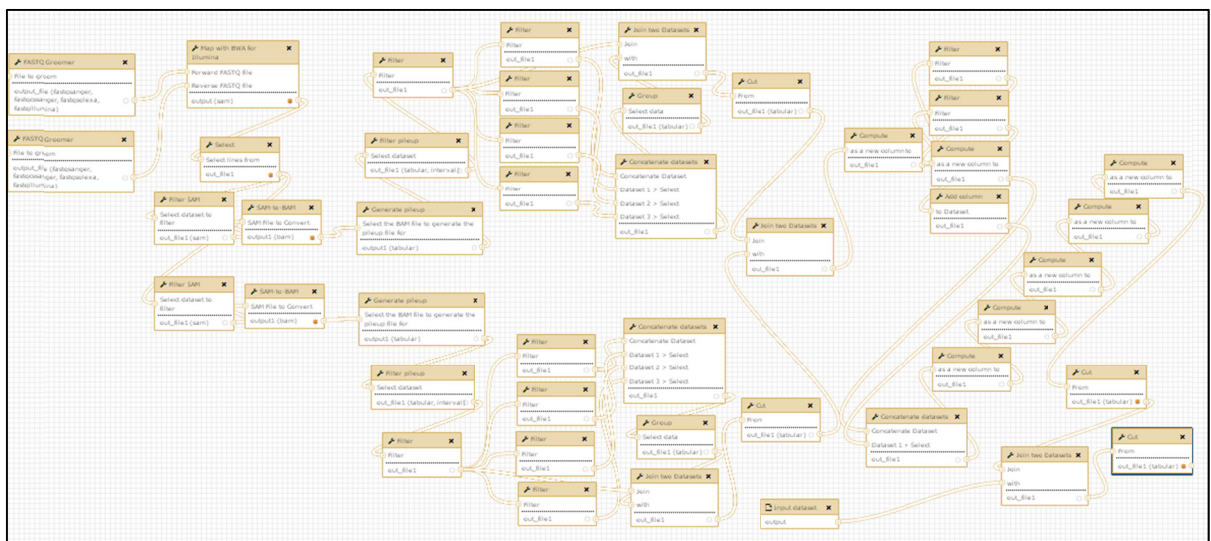


Figura 13. *Workflow* para análisis de mtDNA residual

Chromosome	Position	Reference	"A"	"C"	"G"	"T"	Coverage	Nº variant sites	Gene
chrM	3580	C	0	67	0	0	67	0	MT-ND1
chrM	3581	C	0	64	0	0	64	0	MT-ND1
chrM	3582	C	0	66	0	0	66	0	MT-ND1
chrM	3583	A	54	0	0	0	54	0	MT-ND1
chrM	3584	A	52	0	0	0	52	0	MT-ND1
chrM	3585	C	0	68	0	0	68	0	MT-ND1
chrM	3586	C	0	66	0	0	66	0	MT-ND1
chrM	3587	C	0	69	0	0	69	0	MT-ND1
chrM	3588	C	0	70	0	0	70	0	MT-ND1
chrM	3589	C	0	67	0	0	67	0	MT-ND1
chrM	3590	T	0	0	0	66	66	0	MT-ND1
chrM	3591	G	0	0	65	0	65	0	MT-ND1
chrM	3592	G	0	0	66	0	66	0	MT-ND1
chrM	3593	T	0	0	0	60	60	0	MT-ND1
chrM	3594	C	0	61	0	0	61	0	MT-ND1
chrM	3595	A	63	0	0	0	63	0	MT-ND1

Figura 14. Formato resultados análisis mtDNA residual

[1] Cromosoma, [2] Posición, [3] Base en el genoma de referencia, [4] Número de As, [5] Número de Cs, [6] Número de Gs, [7] Número de Ts, [8] Cobertura en esa posición, [9] Número de variantes (diferencias respecto a la referencia mayores del 1%), [10] Nombre del gen.

Generación ficheros referencia para informe de cobertura de paneles de genes

Se siguieron los pasos descritos en Biostarsⁿ para obtener las pistas de datos adecuadas.

```
# Descarga de inicial de ficheros:
track: UCSC Genes
table: knownCanonical
output format: select fields from primary and related tables
press get output
select fields from hg19.knownCanonical: chrom, chromStart, chromEnd,
transcript
transcript select fields from hg19.kgXref: geneSymbol
press get output

track: UCSC Genes
table: knownGene
output format: BED - browser extensible data
press get output
select option Coding Exons
press get BED

# Modificaciones en las tablas descargadas
awk '{split ($4,a,"_"); {print $1"\t"$2"\t"$3"\t"a[1]"\t"a[3]"\t"$6}}'
UCSC_exons.bed > UCSC_exons_modif.bed
join -1 4 -2 4 <(sort -k4 UCSC_exons_modif.bed) <(sort -k4 UCSC_canonical.bed) |
awk '{print $2"\t"$3"\t"$4"\t"$10"\t"$5"\t"$6}' | bedtools sort -i "-" >
UCSC_exons_modif_canonical.bed
```

ⁿ <https://www.biostars.org/p/93011/>

```
# Curación manual de las tablas obtenidas
También es necesario para acotar correctamente el análisis el fichero bed del
diseño del panel. En este fichero también conviene eliminar el padding
configurado (normalmente 5pb o 10pb). Para quitar 10pb de padding y quedarse sólo
con los exones se usa el siguiente código:
awk '{print $1,$2+10,$3-10,$4,$5,$6}' OFS="\t" your_file.bed > new_file.bed
```

Por último se usaron dos pequeños *scripts* adicionales configurables para calcular el mínimo de cobertura y para los criterios de selección que en este caso son: i) no presentar el 100% de bases cubiertas y ii) que la profundidad mínima (*depth of coverage*) sea menor de 10.

Tabla 11. Script para cálculo del mínimo de cobertura

```
#!/bin/awk -f
!a[$4]++
```

Tabla 12. Script con criterios de selección

```
#!/bin/awk -f
BEGIN {
OFS="\t"
}
{ if ($10<1.0 || $25<10) print $1,$2,$3,$6,$7,$25,$8,$9-$8,$9,$10,$12,$13,$15,$17
}
```

Bases de datos empleadas en anotación de variantes

- hg19 refGene: RefSeqGene, a subset of NCBI's Reference Sequence (RefSeq) project, defines genomic sequences to be used as reference standards for well-characterized genes
<https://www.ncbi.nlm.nih.gov/refseq/rsg/>
- ClinVar: ClinVar is a freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence
<https://www.ncbi.nlm.nih.gov/clinvar/>
- snp138: small variations
<https://www.ncbi.nlm.nih.gov/SNP/>
- UCSC annotations⁷⁷ <https://genome.ucsc.edu/>
 - PhastConsElements46way
 - GenomicSuperDups
- 1000g: The 1000 Genomes Project ran between 2008 and 2015, creating the largest public catalogue of human variation and genotype data.

- <http://www.internationalgenome.org/>
- Esp6500: NHLBI Exome Sequencing Project (ESP). Exome Variant Server
<http://evs.gs.washington.edu/EVS/>
 - ExAC: The data set provided on this website spans 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies.
<http://exac.broadinstitute.org/>
 - Kaviar⁷⁸: is a compilation of SNVs, indels, and complex variants observed in humans, designed to facilitate testing for the novelty and frequency of observed variants.
<http://db.systemsbiology.net/kaviar/>
 - Genome Aggregation Database (gnomAD): the data set provided on this website spans 123,136 exome sequences and 15,496 whole-genome sequences from unrelated individuals sequenced as part of various disease-specific and population genetic studies.
<http://gnomad.broadinstitute.org/>
 - HGMD: The Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff <http://www.hgmd.cf.ac.uk>
 - Database of genomic variation and phenotype in humans using Ensembl resources (Decipher, DDD): <https://decipher.sanger.ac.uk/>
 - Database of genomic Variants (DGV): <http://dgv.tcag.ca/dgv/app/home>

Algoritmos de predicción de patogenicidad

- InterVar⁷⁹ <http://wintervar.wglab.org/search.php>
- SIFT⁸⁰ <http://sift.jcvi.org/>
- PolyPhen2⁸¹ <http://genetics.bwh.harvard.edu/pph2/>
- M-CAP⁸² <http://bejerano.stanford.edu/mcap/index.html>
- S-CAP⁸³ <http://bejerano.stanford.edu/scap/>
- LRT⁸⁴ (*Likelihood ratio test*)
- MutationTaster⁸⁵ <http://www.mutationtaster.org/>
- MutationAssessor⁸⁶ <http://mutationassessor.org/r3/>
- FATHMM⁸⁷ <http://fathmm.biocompute.org.uk/>
- RadialSVM⁸⁸
- LR⁸⁹
- VEST
- GERP++⁹⁰ <http://mendel.stanford.edu/SidowLab/downloads/gerp/>

- PhyloP⁷⁷
- SiPhy⁹¹
- CADD⁹²
- DANN⁹³
- Eigen⁹⁴
- Spidex⁹⁵
- MitImpact³⁸ <http://mitimpact.css-mendel.it/>
- MitoTIP⁹⁶ <https://www.mitomap.org/foswiki/bin/view/MITOMAP/MitoTipInfo>

Otros programas/herramientas complementarios y lenguajes empleados

- R software⁶⁸ <https://www.r-project.org/>
- Perl <https://www.perl.org/>
- Python Language Reference v2.7. <http://www.python.org>
- SAMtools <http://samtools.sourceforge.net/>
- Bedtools <https://bedtools.readthedocs.io/en/latest/>
- GATKtools^{65,66,97} <https://software.broadinstitute.org/gatk/>
- ANNOVAR³⁹ <http://annovar.openbioinformatics.org/>
- Picard <http://broadinstitute.github.io/picard>
- ExomeDepth⁴³ <https://github.com/vplagnol/ExomeDepth>
- AnnotSV⁹⁸ <https://lbgi.fr/AnnotSV/>
- Varsome⁹⁹ <https://varsome.com/>

Resultados

Desarrollo de una estrategia molecular global para la identificación coste-efectiva de anomalías genética en el DNA mitocondrial (mtDNA).

1. Materiales de referencia para la evaluación de heteroplasmia

Una de las actividades fundamentales de los laboratorios de diagnóstico es establecer unos protocolos adecuados de control de calidad. Los ensayos de biología molecular en general carecen materiales de control y programas de evaluación de resultados, aunque éste hecho está cambiando en los últimos años. En nuestro caso se planteó inicialmente la creación de materiales de referencia estables y duraderos que fuesen más allá del uso de muestras reales y se pudiesen emplear en los procesos del análisis que lo requiriesen. En particular, en este caso, el uso fundamental es para asegurar la cuantificación de los niveles de heteroplasmia.

Se realizó el clonaje de segmentos de mtDNA de diversos pacientes en los que se encontraron las mutaciones siguientes: m.3243A>G (*MT-TL1*, *tRNA Leu*), m.3460G>A (*MT-ND1*), m.8344A>G (*MT-TK*, *tRNA Lys*), m.8993T>G (*MT-ATP6*), m.8993T>C (*MT-ATP6*), m.9176T>C (*MT-ATP6*), m.13513A>G (*MT-ND5*), m.13514A>G (*MT-ND5*), m.14484T>C (*MT-ND6*).

Debido al fenómeno de la heteroplasmia debe realizarse al azar una selección de colonias supuestamente portadoras y confirmar la integridad del segmento introducido por secuenciación directa (Figura 16). Se muestran como ejemplo los resultados obtenidos para la mutación m.3243A>G. Se consiguió introducir el DNA de interés en seis de los diez clones probados (Figura 15). De éstos se comprobó por secuenciación directa que tres clones presentan la mutación y los restantes son *wild type*.

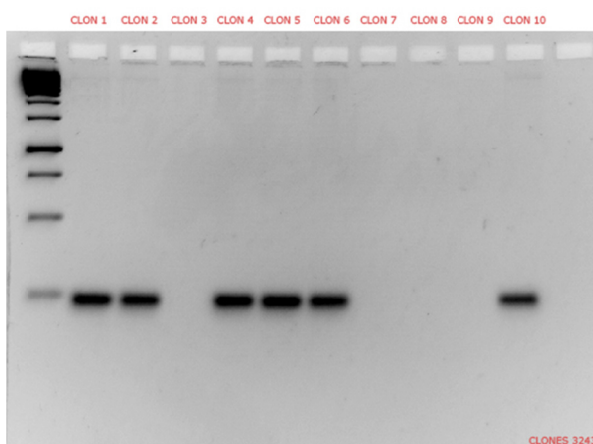


Figura 15. Gel agarosa clones posición 3243 del mtDNA

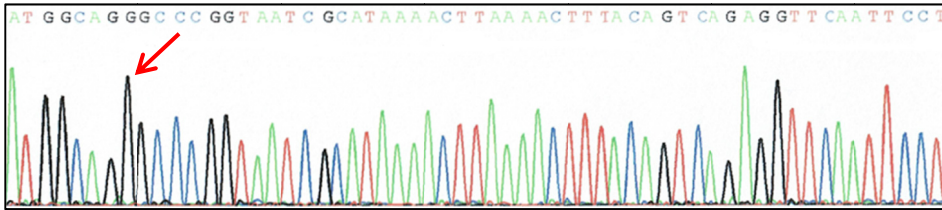
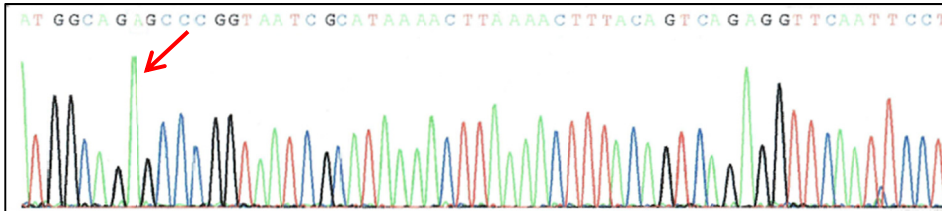
Mutante (MT-TL1 m.3243A>G)**Normal (Wild type)**

Figura 16. Secuenciaciones Sanger de clones (Flecha roja indica posición 3243 de mtDNA)

Para evaluar la viabilidad del almacenamiento de este material de referencia se realizó una recuperación de clones congelados durante dos meses. Se descongeló el Clon WT1 y el Clon M10 (Fig. 13). Dichas bacterias se crecieron en Falcon de 15 mL durante 24 horas con 5 mL de medio LB suplementado con ampicilina (inóculo de 1 uL y 2 uL). La recuperación ha tenido éxito y se consiguieron unos 200 ng/ μ L de producto purificado.

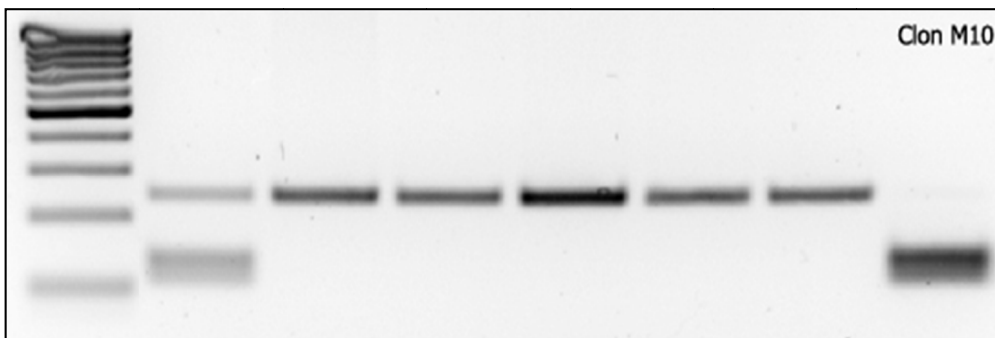


Figura 17. Clon M10 recuperado tras congelación y sometido a PCR-RFLP

2. Minisequenciación (mutaciones puntuales mtDNA y haplogrupo)

El método más común para el análisis de mutaciones en el mtDNA se basa el método de *PCR-RFLP* (*restriction fragment length polymorphism*). El *gold standard* para la cuantificación del grado de heteroplasmia es el método conocido como *Last-cycle hot PCR* consistente en la adición de un nucleótido marcado radiactivamente en la extensión del último ciclo de *PCR*⁵⁵. Éstos métodos son poco coste/efectivos ya que son relativamente caros y consumen mucho tiempo.

Las técnicas de minisequenciación se usan desde hace tiempo en el campo de la genética forense y la investigación clínica^{100,101}. En la actualidad son técnicas bastante robustas y una buena opción para el análisis simultáneo de múltiples SNPs (*Single Nucleotide polymorphisms*).

a. Validación metodología Minisequenciación

En este apartado se optimizó un método de minisequenciación para el cribado de mutaciones relacionadas con las enfermedades mitocondriales más comunes. Por otro lado, se diseñó un panel de mutaciones para la caracterización de los haplogrupos mitocondriales principales en nuestro entorno (Figura 18). Los resultados de la PCR multiplex inicial se muestran en la Figura 19 y la Figura 20 y los electroferogramas resultantes en la Figura 21 y la Figura 22.

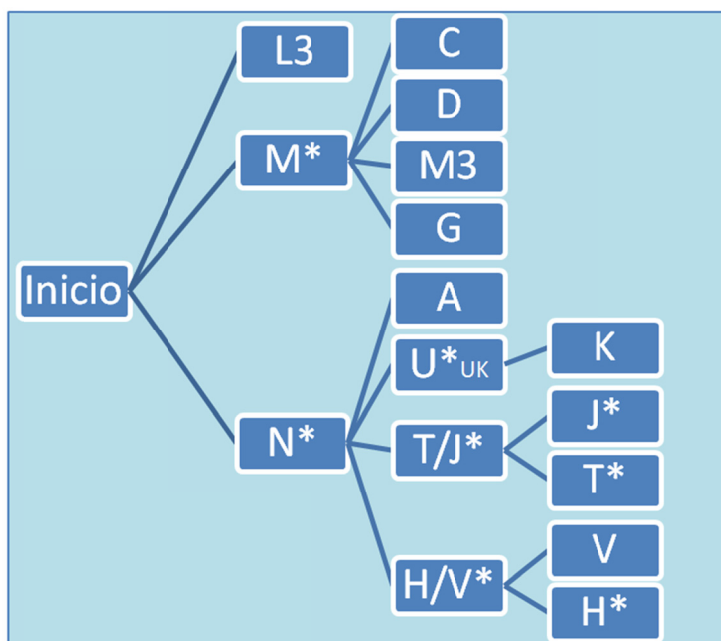


Figura 18. Esquema panel haplogrupos

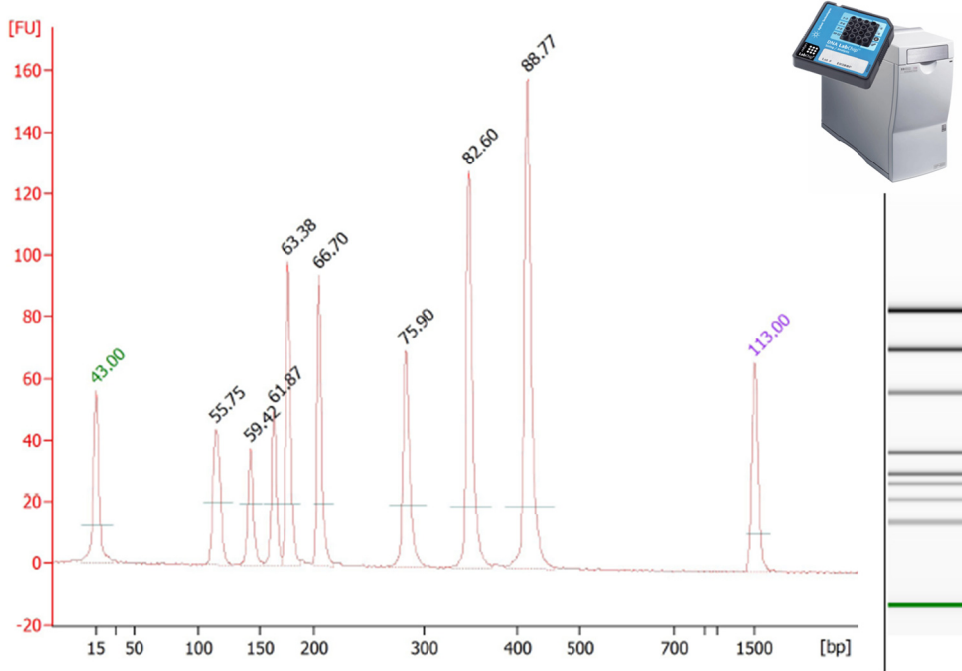


Figura 19. Electroferograma PCR multiplex mutaciones puntuales

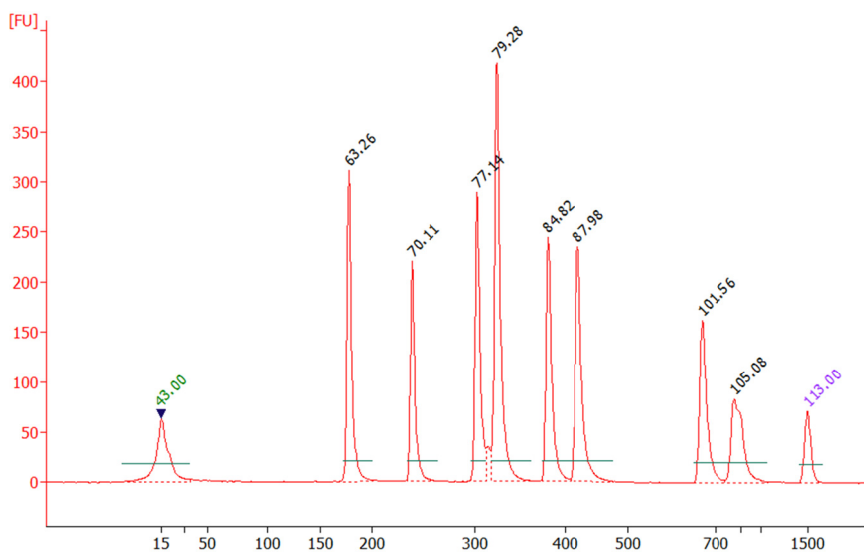


Figura 20. Electroferograma PCR multiplex haplogrupos

Dicho método se validó con pacientes portadores de mutaciones frecuentes en nuestro entorno, en concreto con DNA proveniente de sangre y músculo de 19 pacientes remitidos al Laboratorio de Enfermedades Mitocondriales en los que se confirmó la presencia de mutaciones relacionadas con patologías como MELAS, MERFF, LHON, Síndrome de Leigh, NARP. También se ensayaron 2 pacientes portadores en heterocigosis y homocigosis de la mutación c.1399G>A en el

gen nuclear de la *POLG* (Tabla 13). Todas las mutaciones fueron confirmadas por *PCR-RFLP* y secuenciación directa en un *3130xl Genetic Analyzer* (Applied Biosystems).

El diseño para caracterizar el haplogrupo se validó con 16 muestras (Tabla 14) en las que se había realizado la secuenciación completa del mtDNA por secuenciación enzimática (Sanger). Los electroferogramas resultantes se muestran en la Figura 23. Asimismo, se realizaron pruebas en muestras de orina con éxito (Figura 24), lo que la convierte en una muestra de óptima y no invasiva para caracterizar con rapidez el haplogrupo.

No se ha obtenido ningún falso negativo ni positivo, por lo que los resultados fueron completamente satisfactorios.

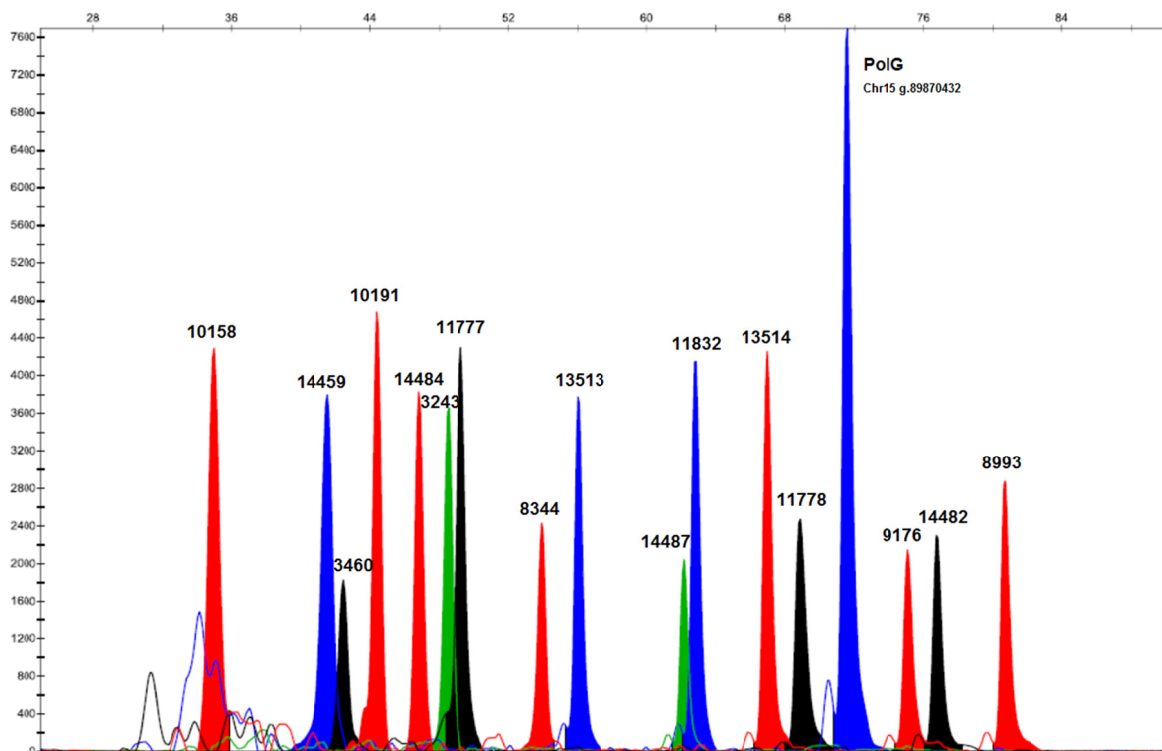


Figura 21. Electroferograma mutaciones puntuales normal

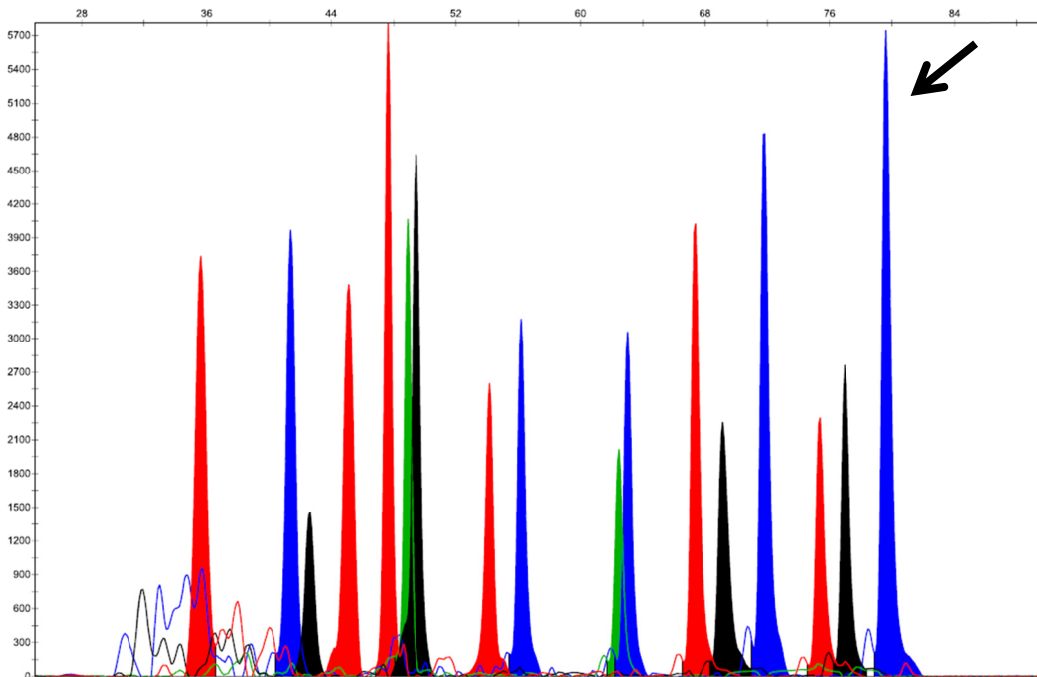


Figura 22. Electroferograma mutaciones puntuales con la mutación m.8993T>G (flecha negra)

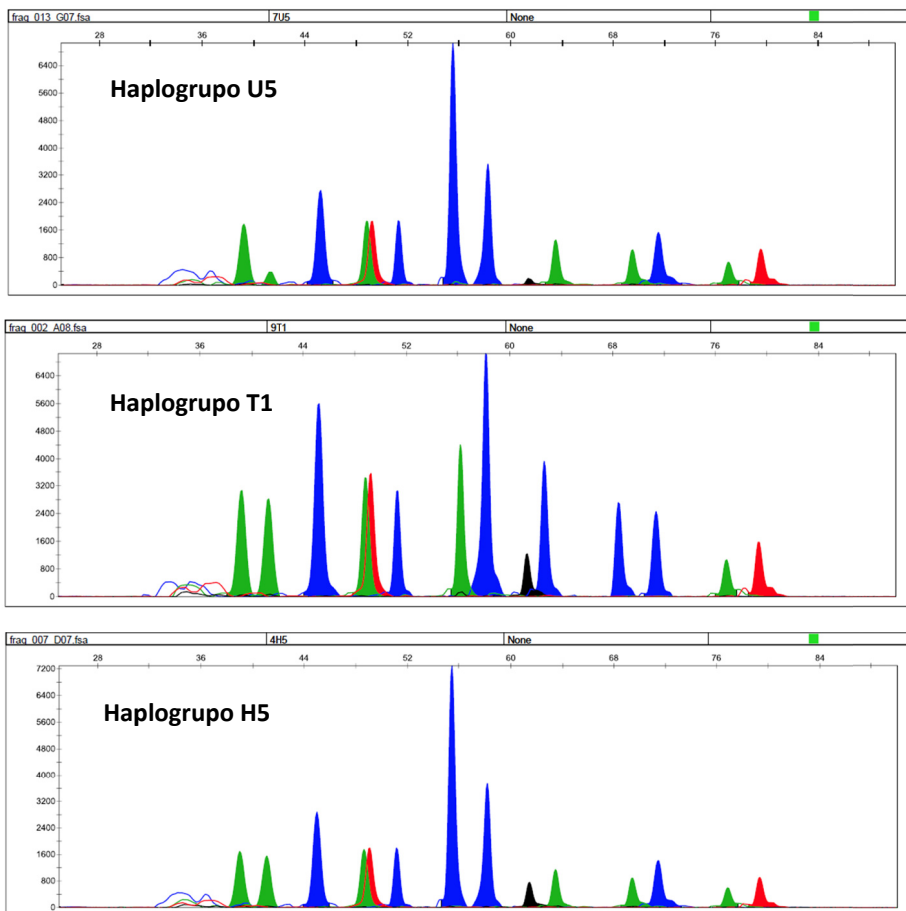


Figura 23. Electroferogramas de caracterización de haplogrupo

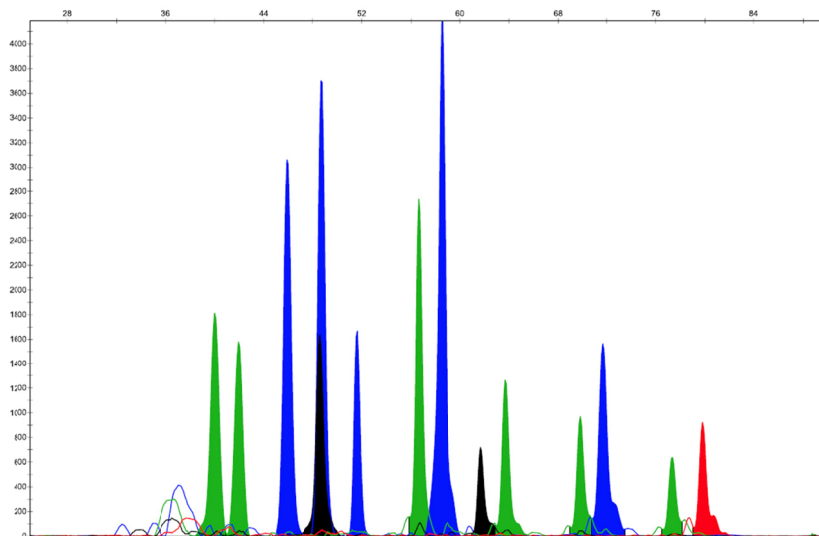


Figura 24. Electroferograma de muestra de orina (Haplogrupo H2)

Tabla 13. Resultados validación mutaciones puntuales

Paciente/Mutación DNA músculo	Resultado previo	Minisequenciación
P1 m.3460G>A	47%	Heteroplasmia
P2 m.3460G>A	85%	Heteroplasmia
P3 m.14484T>C	96%	Homoplasmia
P4 m.14484T>C	58%	Heteroplasmia
P5 m.3243A>G	77%	Heteroplasmia
P6 m.3243A>G	81%	Heteroplasmia
P7 m.8344A>G	53%	Heteroplasmia
P8 m.8344A>G	34%	Heteroplasmia
P9 m.13513G>A	19%	Heteroplasmia
P10 m.13514A>G	67%	Heteroplasmia
P11 m.11778G>A	77%	Heteroplasmia
P12 m.11778G>A	63%	Heteroplasmia
P13 c.1399G>A	Homocigoto	Homocigoto
P14 c.1399G>A	Heterocigoto	Heterocigoto
P15 m. 9176T>G	76%	Heteroplasmia
P16 m.8993T>G	99%	Homoplasmia
P17 m.8993T>C	98%	Homoplasmia
P18 m.8993T>G	89%	Heteroplasmia
P19 m.14487T>C	100%	Homoplasmia

Tabla 14. Resultados validación Haplogrupos

Paciente	Haplogrupo Inicial [‡]	Haplogrupo Minisec
1H	H	H*
2H1	H1	H*
3H2	H2	H*
4H5	H5	H*
6U2	U2	U*
7U5	U5	U*
8U6	U6	U*
9T1	T1	T*
10J1	J1	J*
11V	V	V
12M	M	M*→C
13I	I	N*
14K	K	K
15M	M	M*→G
16J1	J1	J*
17L3	L3	L3

[‡]Obtenido mediante secuenciación completa del mtDNA con método enzimático (Sanger)

Se realizó un estudio de correlación entre los niveles de heteroplasmia obtenidos mediante *PCR-RFLP* y visualización con *Bioanalyzer 2100* (Agilent) y los obtenidos midiendo la altura de los picos del electroferograma de minisequenciación. Se observó una falta de concordancia debida fundamentalmente a que la intensidad de fluorescencia de cada nucleótido no es equivalente tal y como se puede apreciar en la Figura 25 que muestra diferentes tipos de heterocigotos de gDNA.

En el caso del mtDNA podrían establecerse curvas de calibración con plásmidos en función del cambio para ajustar la medida (Figura 26), pero sería muy laborioso y costoso.

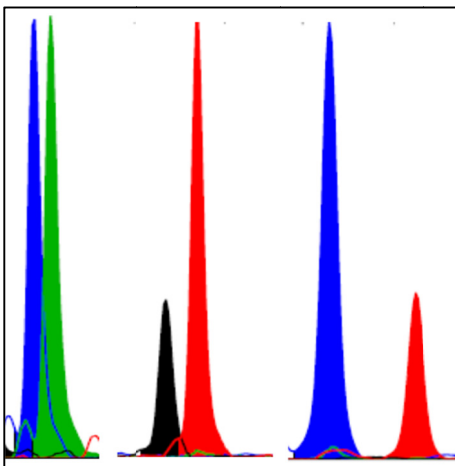


Figura 25. Ejemplos de heterocigotos nucleares (minisequenciación)

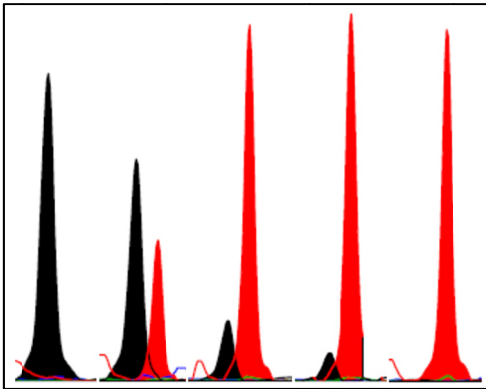


Figura 26. Clones A8344G (100%M, 75%, 50%, 25%, 100%WT)

b. Estudio de costes Minisequenciación

En cuanto a los costes por paciente de la prueba se ha calculado un gasto en reactivos (sin tener en cuenta amortización de equipos) de 6 € (IVA no incluido) procesando 16 pacientes a la vez. El tiempo empleado para procesar estas muestras es de aproximadamente 9 horas: 3 horas de manipulación directa y 6 horas de procesamiento no supervisado en equipos.

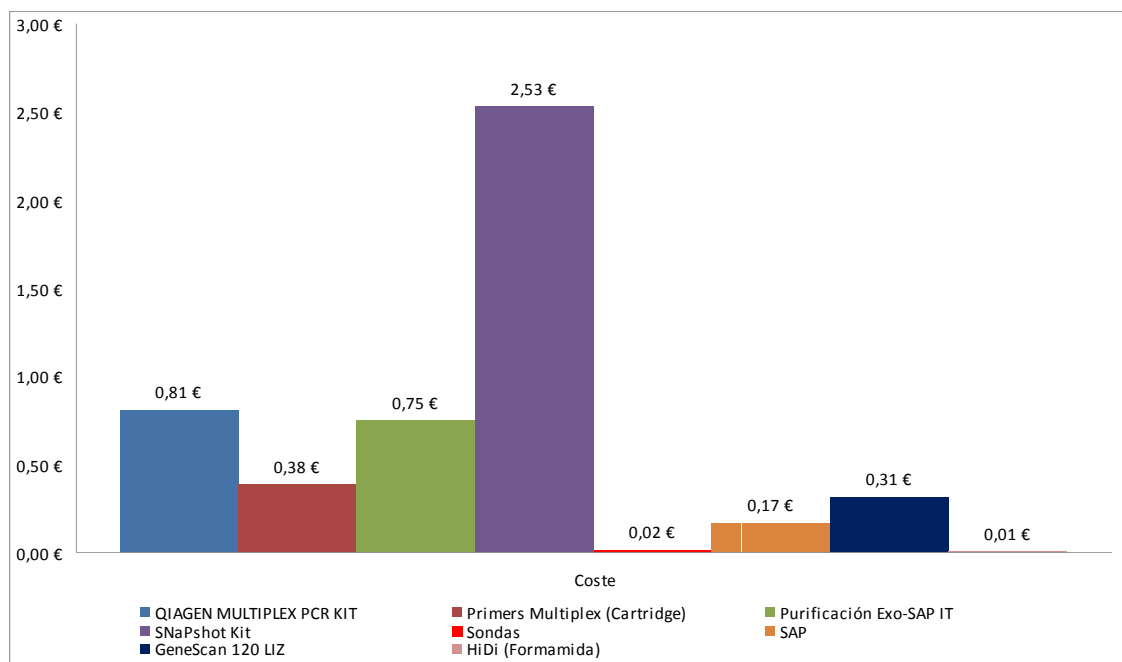


Figura 27. Desglose de costes por tipo de reactivo (excluida electroforesis)

c. Hallazgos mutaciones puntuales mtDNA mediante minisequenciación

En un periodo de 5 años se cribaron 1200 muestras en el laboratorio de Enfermedades Mitocondriales mediante minisequenciación. Han resultado positivos 66 casos (5,7%) con una frecuencia de mutaciones que se indica en la Tabla 15.

Tabla 15. Frecuencia mutaciones minisequenciación

Mutación	Nº casos	Porcentaje del total
m.3243A>G	27	40,3%
m.11778G>A	9	13,4%
m.13513G>A	6	9,0%
m.14484T>C	6	9,0%
m.8993T>G	4	6,0%
c.1399G>A	4	6,0%
m.14487T>C	2	3,0%
m.3460G>A	2	3,0%
m.14459G>A	2	3,0%
m.8344A>G	2	3,0%
m.13514A>G	1	1,5%
m.9176T>C	1	1,5%
m.10158T>C	1	1,5%

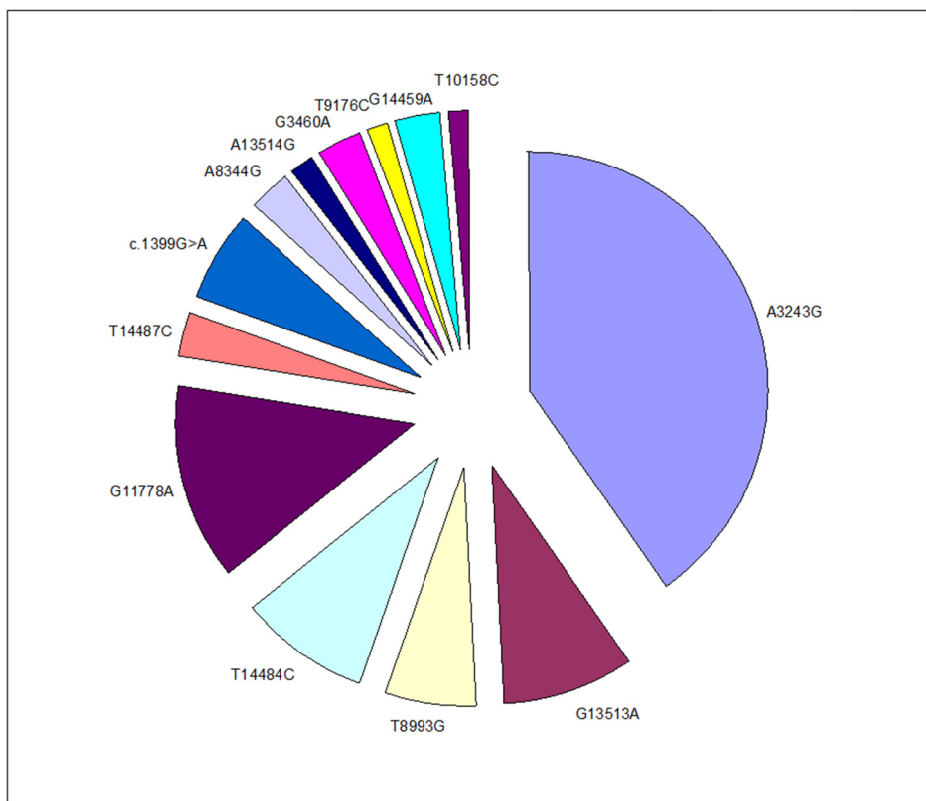


Figura 28. Frecuencia mutaciones minisequenciación

3. PCR-RFLP con cuantificación microfluídica (Agilent Bioanalyzer 2100)

Como ya se ha mencionado el método de referencia para la cuantificación de los niveles de heteroplasma en mtDNA es la llamada *Last cycle Hot PCR*. Éste método evita el marcaje de los heterodúplex que se forman en la *PCR* y permite una correcta y precisa cuantificación del grado de heteroplasma. Pero se trata de un método radiactivo y muy laborioso. Para evitar estos inconvenientes se planteó la posibilidad de aplicar otras metodologías basadas en chips microfluídicos de DNA en el sistema con el objeto de establecer un método sensible, rápido y no radiactivo para la cuantificación de los niveles de heteroplasma en mutaciones puntuales del mtDNA.

Se realizó un estudio de correlación entre el método de referencia radiactivo *Last-Cycle Hot PCR* y el método *PCR-RFLP* y visualización en Bioanalyzer 2100 (Agilent Technologies). Se emplearon 8 mezclas de plásmidos con heteroplasmas (rango 3,12%-85%) simuladas a partir de clones homoplásmicos puros portadores de la mutación m.3243A>G y se verificaron también 2 muestras de DNA de muscular de pacientes heteroplásmicos para la misma mutación en los que se había establecido por anteriormente unas heteroplasmas del 77% y 17% respectivamente.

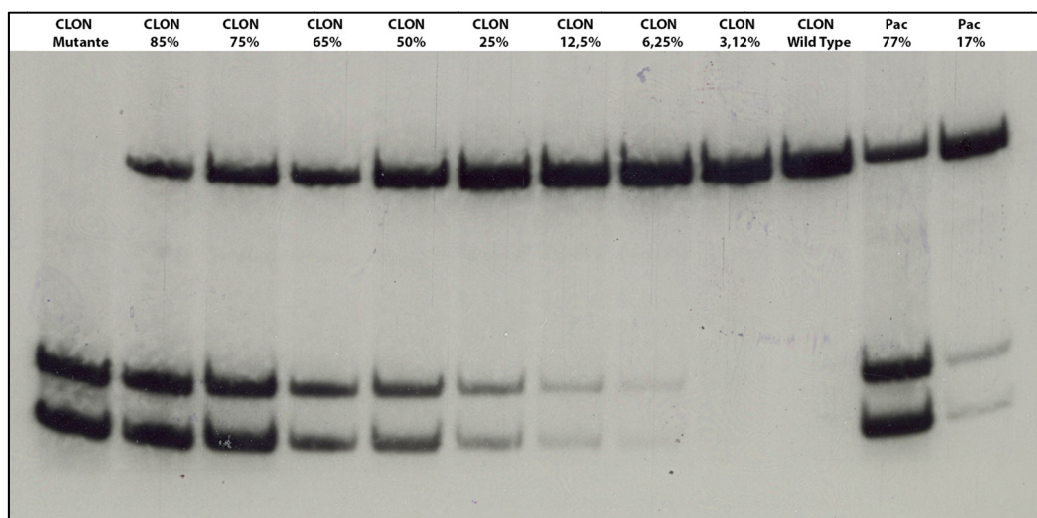
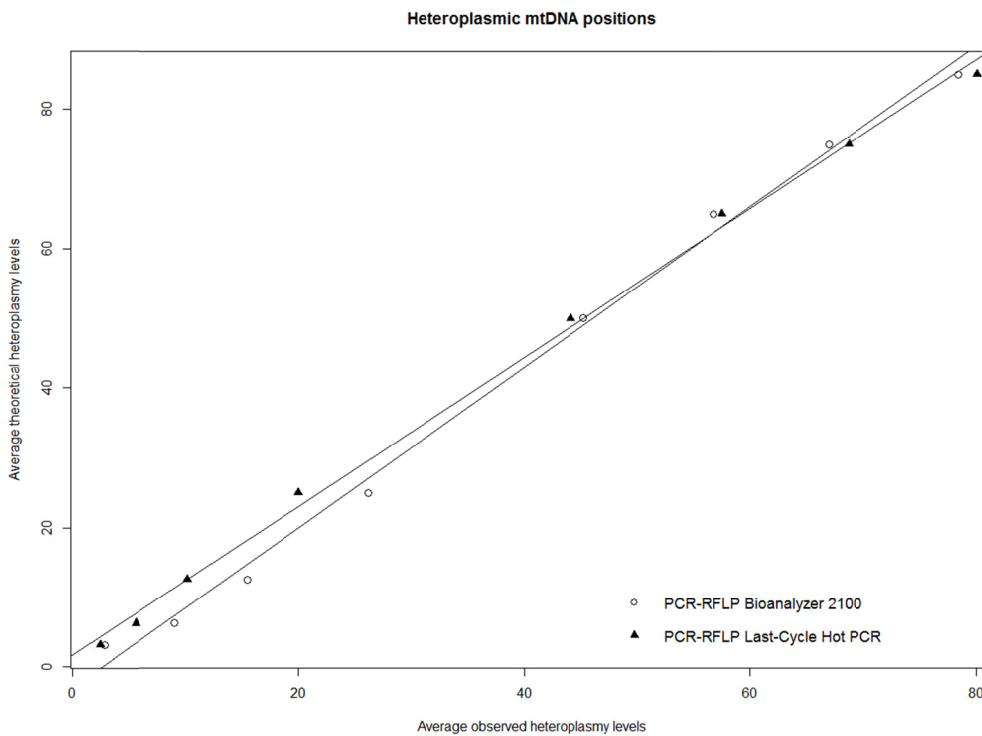


Figura 29. Gel con marcaje radiactivo. PCR-RFLP con *bsp120I* (m.3243A>G)

Los resultados de las cuantificaciones por ambos métodos se muestran en la Tabla 16 y el estudio de correlación en la Figura 30. Como se puede apreciar existe un alto grado de correlación ($r=0.9973$) estadísticamente significativo entre ambos métodos que los hace intercambiables.

Tabla 16. Resultados heteroplasmia Bioanalizador 2100 vs Last-Cycle Hot PCR

Muscle DNA	PCR-RFLP Bionalizador 2100					PCR-RFLP Last-Cycle Hot PCR				
	Mean	Median	SD	CV	N	Mean	Median	SD	CV	N
C 85%	78,4	79,9	2,3221	3,0%	5	80,1	78,6	3,6297	4,5%	5
C 75%	67,0	66,6	3,3327	5,0%	5	68,8	69,1	1,7631	2,6%	5
C 65%	56,7	57,8	3,5801	6,3%	5	57,5	57,0	0,8023	1,4%	5
C 50%	45,2	46,2	3,0646	6,8%	5	44,1	43,3	2,4936	5,7%	4
C 25%	26,2	26,2	2,8662	10,9%	5	20,0	19,2	3,6108	18,1%	4
C 12,5%	15,5	16,9	2,3227	15,0%	5	10,2	9,0	3,1710	31,1%	4
C 6,25%	9,1	9,9	1,5764	17,3%	5	5,7	4,0	3,7859	66,8%	3
C 3,12%	2,9	3,0	0,4583	15,8%	3	2,5	2,5	N/A	N/A	1
PAC 1	74,8	74,7	0,4163	0,6%	3	72,7	71,0	5,8	0,1	3
PAC 2	77,0	77,2	0,4301	0,6%	5	75,4	75,1	1,9	0,0	3
PAC 3	18,5	18,4	1,9962	10,8%	5	15,0	15,0	2,8	0,2	5



Método 1	Método 2	Coefficiente Correlación	p-valor	Nivel de significación
Bioanalizador 2100	Last-Cycle Hot-PCR	0.9972904	4.963408e-08	***
Bioanalizador 2100	Theoretical	0.9977430	2.869452e-08	***
Last-Cycle Hot-PCR	Theoretical	0.9988353	3.946545e-09	***

Figura 30. Estudio correlación heteroplasmia Bioanalyzer 2100 vs Last-Cycle Hot PCR

4. Secuenciación de mtDNA a alta profundidad

En la actualidad las diferentes tecnologías de NGS se están aplicando al estudio del mtDNA completo y a la determinación del grado de heteroplasmia de mutaciones puntuales. Con estas metodologías se pueden alcanzar altos niveles de profundidad (*Depth of coverage*) que permiten la detección de niveles de heteroplasmia muy bajos. En este trabajo se ha validado un protocolo de secuenciación propuesto por *Life Technologies* con algunas modificaciones en la plataforma PGM (*Personal Genome Machine*). Este protocolo se basa en la amplificación inicial por PCR del mtDNA y no está exento de controversia por el posible falseamiento de los niveles de heteroplasmia debido a pseudogenes mitocondriales (*Numts*)¹⁰².

Se obtuvo una cobertura media de 2800X con un perfil irregular (perfil Manhattan, Figura 31)

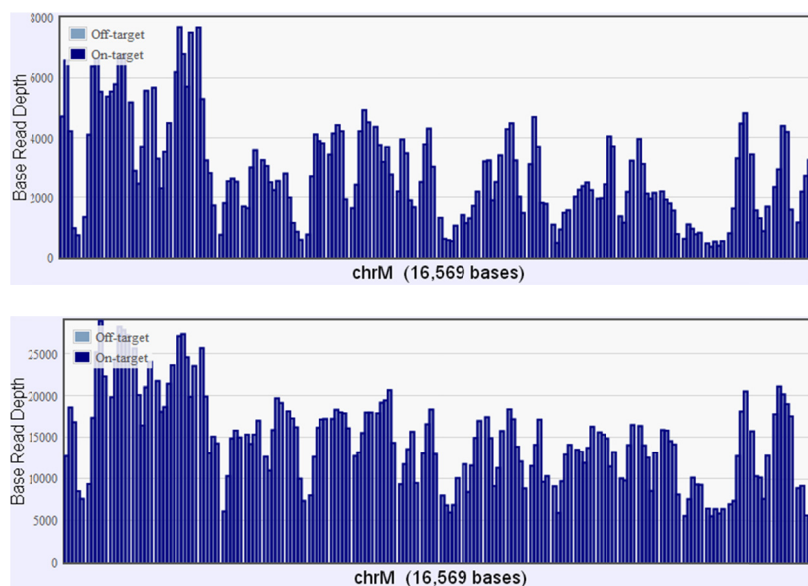


Figura 31. Perfiles cobertura mtDNA en PGM
Arriba química 200pb y abajo química 400pb.

Se ensayó también con el protocolo de secuenciación de 400pb obteniéndose resultados similares a los anteriores (Figura 31 y Figura 32). El perfil heterogéneo sigue presente y los “valles” son prácticamente los mismos. La única mejora sin explicación aparente de la química de 400pb es la mejora del perfil de cobertura en las lecturas en sentido R (*reverse*) tal y como se puede observar en la Figura 32. Perfil cobertura mtDNA con lecturas desglosadas.

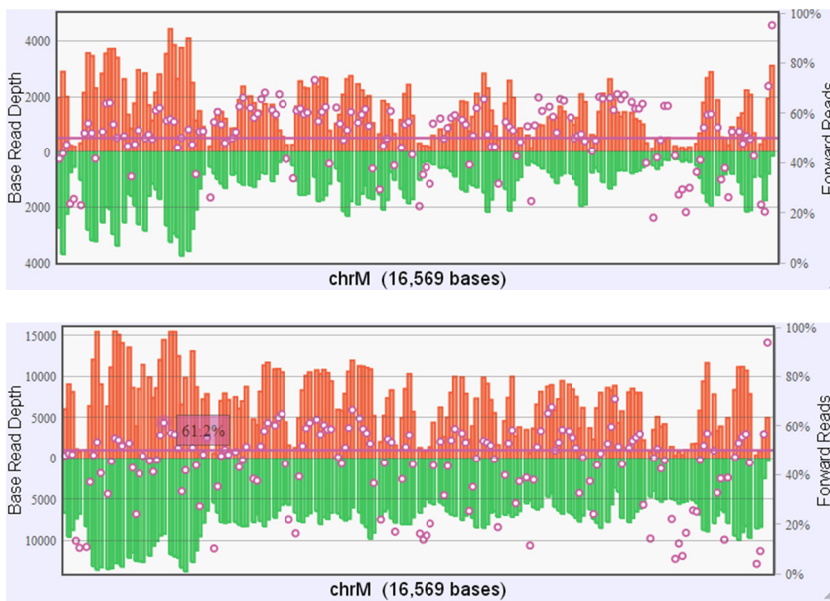


Figura 32. Perfil cobertura mtDNA con lecturas desglosadas.
Arriba química 200pb y abajo química 400pb.

El resumen de la validación se presenta en la siguiente tabla:

Tabla 17. Validación de mutaciones en mtDNA patogénicas

Mutación conocida	Heteroplasmia Bioanalyzer 2100	Heteroplasmia PGM	Cobertura (Profundidad)
m.1555A>G	100%	100%	3507
m.1555A>G	100%	100%	4636
m.3243A>G	80%	82%	2737
m.3243A>G	30%	29%	1010
m.3243A>G	3%	4%	2363
m.3243A>G	17%	9%	3244
m.3243A>G	35%	33%	3083
m.3243A>G	91%	91%	3169
m.3271T>C	78%	78%	3409
m.3460G>A	90%	91%	1540
m.5658T>C	17%	18%	506
m.7191T>C	90%	83%	3237
m.8344A>G	86%	89%	2198
m.8391G>A	38%	33%	819
m.8993T>C	94%	97%	2029
m.9204delAT	95%	96%	1636
m.13094T>C	78%	77%	2614
m.13513G>A	70%	71%	2498
m.15853C>A	100%	100%	3105

5. Comparación secuenciación mtDNA en PGM y MiSeq

Las dos plataformas de segunda generación más usadas en la actualidad son *Ion Torrent Personal Machine* y *MiSeq Desktop Sequencer*. Dichas plataformas difieren sustancialmente en la

forma de generar las librerías y en la química empleada. En colaboración con el Hospital Vall d'Hebron se compararon ambas entre sí para evaluar su rendimiento en cuanto a sensibilidad y especificidad para detectar variantes en el mtDNA (usando como referencia Secuenciación Sanger o Mitochip) y cuantificar el grado de heteroplasmia. Se muestra en las tablas 18 y 19 un resumen de los resultados ya publicados¹⁰³.

Tabla 18. Diferencias observadas entre Sanger/Mitochip vs PGM (N=16)

Location	Total Variants	False positives	False negatives	Sensitivity	Specificity
Whole mtDNA (1-16569)	409	7 (2%)	6 (1%)	98.53%	99.96%
D-loop (16024-576)	118	6 (5%)	6 (5%)	94.92%	99.41%
"Coding region" (577-16023)	291	1 (0.3%)	0	100.00%	99.99%

Tabla 19. Diferencias observadas entre Sanger/Mitochip vs MiSeq (N=18)

Location	Total Variants	False positives	False negatives	Sensitivity	Specificity
Whole mtDNA (1-16569)	439	9 (2%)	0	100.00%	99.94%
D-loop (16024-576)	121	9 (7%)	0	100.00%	99.12%
"Coding region" (577-16023)	318	0	0	100.00%	100.00%

6. Desarrollo de *scripts* para la automatización de procesos de anotación y filtrado de variantes

El *script* mostrado a continuación (Tabla 20) permite la anotación en paralelo a nivel de núcleo de múltiples ficheros en estaciones de trabajo multinúcleo mediante el módulo *Parallel-ForkManager* e incluye la posibilidad de filtrado previo de variantes conocidas que sean polimorfismos habituales en la población objeto de estudio o sesgos de secuenciación conocidos en la plataforma de análisis.

Tabla 20. Script de anotación de mtDNA completo

```
#!/usr/bin/perl -w #
# SCRIPT: mitomap.pl #
# Este script anota con varias bases de MITOMAP y MitImpact las variantes
del mtDNA
# Entrada: Deben indicarse los directorios de entrada (ficheros .xls), de
salida y el fichero de filtrado #
# filter0 = sin filtrado ; filter1= mtDNA completo #
# Ejemplo: perl annovar.pl directorio_input directorio_results filter0 #
# Salida: tabla de variantes anotadas con Mitomap, MitImpact y MitoTIP en
formato xls #
#####
# Activar la declaracion estricta de variables
use strict;
```

```
use Term::ANSIColor;
use Parallel::ForkManager;
my $max_processors=12;
## Paso 1. Adquirir el nombre del directorio
my ($directorio_in,$directorio_out,$FILTER,$n_argumentos);
my $database1 = "dbA_final_filtered"; #Base MITOMAP
my $database2 = "dbB_final"; #Base MITOMAP
my $database3 = "dbC_final"; #Base MITOMAP
my $database4 = "dbD_final"; #Base MitImpact
my $database5 = "dbE_final"; #Base MitoTIP
$n_argumentos = scalar(@ARGV);
($n_argumentos == 3) or
print_error("Error argumentos. Indica directorios de entrada, salida y
filtro");
($directorio_in,$directorio_out,$FILTER) = @ARGV;
print_ok("Lectura de directorios indicados");
## Paso 2. Abrir el directorio
my ($fichero,$fichero_in,$fichero_out,$fichero_wrk);
my ($nombre,$comando,$rtcd);
opendir(DIR,$directorio_in);
print_ok("Acceso al directorio de ficheros para anotar: $directorio_in");
my $super_fork= new Parallel::ForkManager($max_processors);
while ($fichero = readdir(DIR))
{
    ## Control sobre los directorios y ejecución de programas externos
    next unless ($fichero =~ m/\.xls$/);
    next if ($fichero =~ /\^\.\/);
    ##### Start fork #####
my $pid= $super_fork->start and next;
## Paso 3. Convertir xls de entrada en txt
($nombre = $fichero) =~ s/\.xls//;
$fichero_in = "$directorio_in/$fichero";
$fichero_out = "$directorio_in/$nombre";
$comando = "ssconvert $fichero_in $fichero_out.txt";
$rtcd = system($comando);
rt_ctrl($rtcd, "Conversión xls a txt: $nombre");
## Paso 4. Convertir txt en formato correcto para comparar con bases de
datos
$fichero_in = "$fichero_out.txt";
$fichero_wrk = "$directorio_out/$nombre.wrk";
$fichero_out = "$directorio_out/$nombre.convert";
$comando = "cut -f1,2,3,4,5,7,8,10,11,12,13,14,20,22,24 $fichero_in
| sed ld | sed '/Absent/d' | sed -e 's/ /_/g' - > $fichero_wrk ; cut -
f2,3,4 $fichero_wrk | sed -e 's/-/del/g' | sed -e 's/\t/_/g' - | paste -
$fichero_wrk | sort -k1 -d > $fichero_out";
$rtcd = system($comando);
rt_ctrl($rtcd, "Preparación archivo muestra: $nombre");
## Paso 5. Filtrar por variantes conocidas propias
$fichero_in = "$fichero_out";
$fichero_out = "$directorio_out/$nombre.filtered";
$comando = "cat $fichero_in | grep -v -w -f $FILTER >
$fichero_out";
$rtcd = system($comando);
rt_ctrl($rtcd, "Filtrado de variantes comunes: $nombre");
## Paso 6. Anotar con 3 bases de datos
$fichero_in = "$fichero_out";
$fichero_wrk = "$directorio_out/$nombre.wrk2";
$fichero_out = "$directorio_out/$nombre.annotated1";
$comando = "join -a 1 -1 1 $fichero_in $database1 | join -a 1 -1 1
```

```
- $database2 | join -a 1 -1 1 - $database3 | sed 's/ /\t/g' >
$fichero_wrk ; cut -f1-16,18,19,20,21,22,23 $fichero_wrk > $fichero_out ;
sed -i
'1i1\t2\t3\t4\t5\t6\t7\t8\t9\t10\t11\t12\t13\t14\t15\t16\t17\t18\t19\t20\
t21\t22' $fichero_out";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Anotación con Mitomap (MitImpact y MitoTIP)
muestra: $nombre");
## Paso 7. Generar tabulaciones en las columnas vacías y añadir valor
nulo NA
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.annotated2";
    $comando = "ssconvert $fichero_in $fichero_out.xls ; ssconvert
$fichero_out.xls $fichero_out.txt ; bash fill-empty.sh $fichero_out.txt
NA ; sed 's/, /\t/g' $fichero_out.txt | sed 1d - > $fichero_out";
    $rtcd = system($comando);
## Paso 8. Anotar con base MitImpact
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.annotated3";
    $comando = "join -a 1 -1 1 $fichero_in $database4 | sed 's/ /\t/g'
> $fichero_out.1 > $fichero_out";
    $rtcd = system($comando);
## Paso 9. Generar tabulaciones en las columnas vacías y añadir valor
nulo NA
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.annotated4";
    $comando = "ssconvert $fichero_in $fichero_out.xls ; ssconvert
$fichero_out.xls $fichero_out.txt ; bash fill-empty.sh $fichero_out.txt
NA ; sed 's/, /\t/g' $fichero_out.txt > $fichero_out";
    $rtcd = system($comando);
## Paso 10. Anotar con base MitoTIP
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.annotated5";
    $comando = "join -a 1 -1 1 $fichero_in $database5 | sed 's/ /\t/g'
> $fichero_out.2 ; cut -f2-48 $fichero_out.2 > $fichero_out; sed -i
'liChr\tPosition\tRef\tVariant\tZygosity\tHet%\tQual\tType\tSource\tDisea
se\tGene\tRegion\tCoverage\tCov+
\tCov-
\tAACChange\tGB_Freq\tStatus\tHomoplasmy\tHeteroplasmy\tDisease_Mitomap\tA
mino_acid_position\tamino_acid_substitution\tcodon_substitution\tPolyPhen
2_prediction\tPolyPhen2_score\tSIFT_prediction\tSIFT_score\tFatHmM_predic
tion\tFatHmM_score\tPROVEAN_prediction\tPROVEAN_score\tMutAss_prediction\
tMutAss_score\tCarol_prediction\tCarol_score\tCondel_prediction\tCondel_s
core\tPhyloP_46V\tPhastCons_46V\tSiteVar\tUniformity_score\tUniformity_sc
ore_II\tConsensus_Uniformity_score\tMitoTIP_score\tMitoTIP_interpret'
$fichero_out";
    $rtcd = system($comando);
## Paso 11. Generar tabulaciones en las columnas vacías y añadir valor
nulo NA
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.annotated6";
    $comando = "ssconvert $fichero_in $fichero_out.xls ; ssconvert
$fichero_out.xls $fichero_out.txt ; bash fill-empty.sh $fichero_out.txt
NA ; sed 's/, /\t/g' $fichero_out.txt > $fichero_out";
    $rtcd = system($comando);
## Paso 12. Convertir a xls
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.mtDNAanotado";
    $comando = "ssconvert $fichero_in $fichero_out.xls";
    $rtcd = system($comando);
```

```
    rt_ctrl($rtcd, "Conversión a xls: $nombre");
##### end fork #####
$super_fork->finish;
}
$super_fork->wait_all_children;

## Paso 13. Borrado archivos intermedios
    my $dir1 = "$directorio_out";
    unlink glob "$dir1/*.wrk & $dir1/*.wrk2 & $dir1/*.convert &
$dir1/*.filtered & $dir1/*.annotated*";
    my $dir2 = "$directorio_in";
    unlink glob "$dir2/*.xls & $dir2/*.txt";
## Paso 14. Lanzar análisis de coverage
    $comando = "./delecciones.pl $directorio_in $directorio_out";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Análisis de cobertura y delecciones");
# Cerrar el directorio
closedir(DIR);

print_ok(":-) Finalizado. Resultados de cada muestra en directorio:
$directorio_out");

exit(0);
#-----

# Rutinas informativas
print_ok("Inicio correcto");
if (scalar(@ARGV) == 0)
{ print_ok("Parametros"); }
else
{ print_error("Parametros");}
print_ok("Final correcto");
sub rt_ctrl
{
    my $rt = $_[0];
    my $mess = $_[1];
    if ($rt == 0)
    {print_ok($mess);}
    else
    {print_error($mess);}
}
sub print_ok
{
    my $mess = $_[0];
    print STDERR color("bold green"), "$mess\t\t[OK]\n\n";
    print STDERR color(" reset ");
}
sub print_error
{
    my $mess = $_[0];
    print STDERR color("bold red"), "$mess\t\t[ERROR]\n\n";
    print STDERR color(" reset ");
    #die();
    exit(10);
}
```

```

aitor@ubuntu:~/mtanovar$ ./mitomapv2.pl input results filter0
Lectura de directorios indicados [OK]

Acceso al directorio de ficheros: input [OK]

Conversión xls a txt: 007.10165AM2 [OK]

Preparación archivo muestra: 007.10165AM2 [OK]

Filtrado de variantes comunes: 007.10165AM2 [OK]

Anotación con Mitomap muestra: 007.10165AM2 [OK]

Conversión a xls: 007.10165AM2 [OK]

:-) Finalizado. Resultados de cada muestra en directorio: results [OK]
  
```

Figura 33. Captura de pantalla de una ejecución del script en la terminal.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI			
1	chrM	263	A	G	Homozyg	100	34528	SNP	Novel	---	MT-OHR	110-441	2161	1219	942	NA	NA	NA	NA	NA	NA																	
2	chrM	750	A	G	Homozyg	100	47213	SNP	Novel	---	RNA-12S	648-601	2964	1517	1447	RNA	26457	NA	NA	NA	NA	NA																
4	chrM	951	G	A	Homozyg	100	40884	SNP	Novel	---	RNA-12S	648-601	2547	1344	1203	RNA	215	NA	NA	NA	NA	NA																
5	chrM	6647	A	G	Heterozyg	69.4	30452	SNP	Novel	---	MT-CO1	5904-7445	3580	1902	1678	syn.L	67	NA	NA	NA	NA	NA																
6	chrM	7691	T	C	Homozyg	100	17629	SNP	Novel	---	MT-CO2	7585-8269	2348	1245	1103	non-syn.F	8	NA	NA	NA	NA	NA	26 F-L	Tic-Ctc	benign	0.01	neutral	0.75	neutral	1.71	neutral	0.95	neutral	-0.42	neutral			
7	chrM	8860	A	C	Heterozyg	3.3	41016	SNP	Novel	---	MT-ATP6	8573-9206	2614	1436	1178	non-syn.T	1	NA	NA	NA	NA	NA	112 T-P	Aca-Gca	benign	0.28	neutral	0.22	neutral	4.17	deleterious	-5.25	medium	3.08	neutral			
8	chrM	8860	A	C	Heterozyg	96.7	41016	SNP	Novel	---	MT-ATP6	8573-9206	2614	1436	1178	non-syn.T	26536	NA	NA	NA	NA	NA	112 T-A	Aca-Gca	benign	0	neutral	0.61	neutral	4.23	deleterious	-3.97	medium	2.15	neutral			
9	chrM	8371	C	T	Homozyg	100	46902	SNP	Novel	---	MT-CO3	9208-9990	2947	1543	1404	syn.Y	17	NA	NA	NA	NA	NA																
10	chrM	13759	G	A	Homozyg	100	15013	SNP	Novel	---	MT-ND5	12237-141	942	1	941	non-syn.A	804	NA	NA	NA	NA	NA	435 A-T	Gca-Aca	benign	0.01	neutral	0.47	neutral	3.76	neutral	1.66	neutral	-2.07	neutral			
11	chrM	15326	A	G	Homozyg	100	22151	SNP	Novel	---	MT-CYB	14747-158	1385	440	945	non-syn.T	26506	NA	NA	NA	NA	NA	154 T-A	Aca-Gca	benign	0.02	neutral	0.49	neutral	3.23	neutral	0.31	neutral	-1.01	neutral			
12	chrM	15721	T	C	Homozyg	100	61557	SNP	Novel	---	MT-CYB	14747-158	2601	1375	1226	syn.Y	120	NA	NA	NA	NA	NA																
13	chrM	16394	C	T	Heterozyg	89.7	60572	SNP	Novel	---	D-loop	16024-165	4675	1090	3585	NA	NA	NA	NA	NA	NA																	
14	chrM	16357	T	C	Heterozyg	10.7	16032	SNP	Novel	---	D-loop	16024-165	4709	1138	3571	NA	NA	NA	NA	NA	NA																	
15	chrM	16519	T	C	Heterozyg	100	28247	SNP	Novel	---	D-loop	16024-165	1770	1493	277	NA	NA	NA	NA	NA	NA																	
16	chrM	16526	G	A	Heterozyg	4.5	62701	SNP	Novel	---	D-loop	16024-165	1689	1425	264	NA	NA	NA	NA	NA	NA																	
17																																						
18																																						
19																																						
20																																						
21																																						
22																																						

Figura 34. Ejemplo fichero Excel final con variantes anotadas

7. Detección de deleciones y análisis de cobertura

Las deleciones únicas en el mtDNA son habitualmente esporádicas e incluyen tres entidades fenotípicas principales: Síndrome de Pearson, Síndrome de Kearns-Sayre con debut antes de los 20 años (SKS) y Oftalmoplejia progresiva externa aislada (PEO)³².

En la base de datos MITOMAP un 83,5% de las deleciones únicas registradas son grandes deleciones que presentan un tamaño entre 2000-8000pb.

Se desarrolló un script que permite visualizar de una manera rápida la existencia de una gran deleción y realizar un cálculo aproximado de los puntos de corte. Consta de una parte para lanzar el proceso general en cada muestra y otra para realizar los cálculos y generar un fichero PDF con los resultados subdividida a su vez en un proceso de lanzamiento y otro de operaciones.

Tabla 21. Script de análisis de deleciones en mtDNA

```

#!/usr/bin/perl -w #
# SCRIPT: deleciones.pl #
# Entrada: Deben indicarse los directorios de entrada (ficheros .bam) y
de salida
# Ejemplo: ./deleciones.pl directorio_input directorio_output
#####
  
```

```
# Activar la declaracion estricta de variables
use strict;
use Term::ANSIColor;
use Parallel::ForkManager;
my $max_processors=16;
my $mtDNA_genes="mtDNA_genes";
my $work_dir=~"/Desktop/mtannovar";
## Paso 1. Adquirir el nombre del directorio
my ($directorio_in,$directorio_out,$n_argumentos);
$n_argumentos = scalar(@ARGV);
($n_argumentos == 2) or
print_error("Error argumentos. Indica directorios de entrada y salida");
($directorio_in,$directorio_out) = @ARGV;
print_ok("Lectura de directorios");
## Paso 2. Abrir el directorio
my ($fichero,$fichero_in,$fichero_out,$fichero_wrk);
my ($nombre,$comando,$rtcd);
opendir(DIR,$directorio_in);
print_ok("Acceso al directorio de ficheros para cobertura:
$fichero_in");
my $super_fork= new Parallel::ForkManager($max_processors);
while ($fichero = readdir(DIR))
{
    ## Control sobre los directorios y ejecución de programas externos
    next unless ($fichero =~ m/\.bam$/);
    next if ($fichero =~ /\^\.\/);

##### Start fork #####
my $pid= $super_fork->start and next;
## Paso 3. Samtools mpileup
($nombre = $fichero) =~ s/\.bam//;
$fichero_in = "$directorio_in/$fichero";
$fichero_out = "$directorio_out/$nombre.pileup";
$comando = "samtools mpileup -A -B -q 1 -d 1000000 $fichero_in >
$fichero_out";
$rtcd = system($comando);
rt_ctrl($rtcd, "Contaje: $nombre");
## Paso 4. Cut
$fichero_in = "$fichero_out";
$fichero_out = "$directorio_out/$nombre.1";
$comando = "cut -f4 $fichero_in > $fichero_out";
$rtcd = system($comando);
rt_ctrl($rtcd, "Selección de datos: $nombre");
## Paso 5. Paste
$fichero_in = "$fichero_out";
$fichero_out = "$directorio_out/$nombre.coverage";
$comando = "paste $fichero_in $mtDNA_genes > $fichero_out";
$rtcd = system($comando);
rt_ctrl($rtcd, "Fusión nombre de genes: $nombre");

##### end fork #####

$super_fork->finish;
}
$super_fork->wait_all_children;
## Paso 5. Ejecutar script R para predicción deleciones
$comando = "mv $work_dir/results/*.coverage $work_dir | Rscript
mtDNA_covered_report.R";
$rtcd = system($comando);
```

```

rt_ctrl($rtcd, "Cálculos con R");
## Paso 6. Borrado archivos intermedios
my $dir1 = $directorio_out;
unlink glob "$dir1/*.pileup & $dir1/*.1";
my $dir2 = "$directorio_in";
unlink glob "$dir2/*.bam";
my $dir3 = "$work_dir";
unlink glob "$dir3/*.coverage";
rt_ctrl($rtcd, "Limpieza archivos temporales");
# Cerrar el directorio
closedir(DIR);
exit(0);

```

Tabla 22. Script R de lanzamiento para detección y análisis de deleciones únicas y cobertura

```

setwd("~/mtannovar")
files <- list.files(pattern=".coverage")
for (input in files){rmarkdown::render("~/mtannovar/mtDNA_markdown.Rmd",
                                     output_format = "pdf_document",
                                     output_file =
paste("Estudio_Cobertura_", input, ".pdf", sep=''),
                                     output_dir =
'~/mtannovar/results')
}

```

Tabla 23. Script R de cálculo y generación de PDF

```

---
title: "Análisis cobertura mtDNA"
author: "Laboratorio de Enfermedades Mitocondriales Hospital 12 de
Octubre"
output: pdf_document
---
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
```{r Funcion para eliminar outliers
#Fuente: http://stackoverflow.com/questions/4787332/how-to-remove-outliers-from-a-datase
remove_outliers <- function(x, na.rm = TRUE, ...) {
 qnt <- quantile(x, probs=c(.25, .75), na.rm = na.rm, ...)
 H <- 1.5 * IQR(x, na.rm = na.rm)
 y <- x
 y[x < (qnt[1] - H)] <- NA
 y[x > (qnt[2] + H)] <- NA
 y
}
```
```{r Calcular breakpoints y otros graficos, echo=FALSE}
options(warn=-1)
df <- as.data.frame(read.table(input,header=F))
sum <- summary(df[,1])
colnames(df) <- c("coverage", "gene")
print (input, quote=FALSE)
print (sum)
posicion <- c(1:16569)

```



```

df2 <- data.frame(posicion,df)
colnames(df2) <- c("position", "coverage", "gene")
Gráfico Breakpoints
mean10 <- 0.1*mean(df2[,2])
breakp <-subset(df2, df2[,2]<mean10)
head <-head(breakp$position, n=20L) #Selección de los 20 primeros
valores por debajo de la media
head[is.null(head)]<-0 ; head[is.na(head)]<-0
tail <-tail(breakp$position, n=20L)
tail[is.null(tail)]<-0 ; tail[is.na(tail)]<-0
head <-remove_outliers(head) #Eliminar outliers
tail <-remove_outliers(tail)
par(mfrow=c(1,1))

plot(df[,1],type="h",col="grey",xlim=c(1,16569),ylim=c(0,9000),ylab="Depth
of coverage", xlab="mtDNA position", main="Coverage distribution")
if (na.omit(head)[1]==na.omit(tail)[1] ||
na.omit(tail)[1]<=na.omit(head)[1]+30){
 print("Inicio deleción no detectada")
} else {
 abline(v=na.omit(head)[1], col="red") #Genera una línea vertical roja
al comienzo del breakpoint
 text((na.omit(head)[1]), 9000, na.omit(head)[1],
pos=2,cex=0.8,col="purple", font=2)
}
if (na.omit(tail)[1]==na.omit(head)[1] ||
na.omit(head)[1]>=na.omit(tail)[1]-30){
 print("Final deleción no detectada")
} else {
 abline(v=na.omit(tail)[1], col="red") #Genera una línea vertical roja
al final del breakpoint
 text((na.omit(tail)[1]), 9000, na.omit(tail)[1],
pos=2,cex=0.8,col="purple", font=2)
}
Histograma
hist(df[,1],xlab="Depth of
coverage", ylab="Reads",col="grey",main="Histogram Coverage")
Barplot
a<-NROW(subset(df, coverage<500))
b<-NROW(subset(df, coverage>=501&coverage<=2500))
c<-NROW(subset(df, coverage>=2501&coverage<=5000))
d<-NROW(subset(df, coverage>5000))
new<-cbind(a,b,c,d)
colnames(new)<-c("<500", "500-2500","2500-5000", ">5000")
barplot(new, main="Depth of coverage", ylab="Reads", col="grey")
Gráfico coverage por gen del mtDNA (RNAs y subunidades)
genes<-df$gene
tRNAs<-subset(df,grepl("RNA", genes), select=coverage:gene)
tRNAs$gene<-droplevels(tRNAs$gene)
MT<-subset(df,grepl("^M", genes), select=coverage:gene)
MT$gene<-droplevels(MT$gene)
plot(tRNAs$coverage~tRNAs$gene,cex.axis=.55,cex=.5,las=3,cex.lab=.8,ylab=
"Depth of coverage",xlab="",main="Depth of Coverage tRNAs",
ylim=c(1,max(tRNAs$coverage)))
plot(MT$coverage~MT$gene,cex.axis=.55,
cex=.55,las=3,cex.lab=.8,ylab="Depth of coverage",xlab="",main="Depth of
Coverage MT genes", ylim=c(1,max(MT$coverage)))
Gráfico Low coverage
par(mfrow=c(2,1))

```

```

nocov <- subset(df2, coverage<100)
nocov2 <- nocov[c(1)]
colnames(nocov2) <- "mtDNA positions with low coverage (<100)"
plot(nocov2, xlab="mtDNA position", main="mtDNA positions with low
coverage (<100)", xlim=c(1,16569))
```


**mtDNA positions with low coverage (<100)**



```

```{r No cubiertas, echo=FALSE}
print (nocov[,2-3])
```

```


```

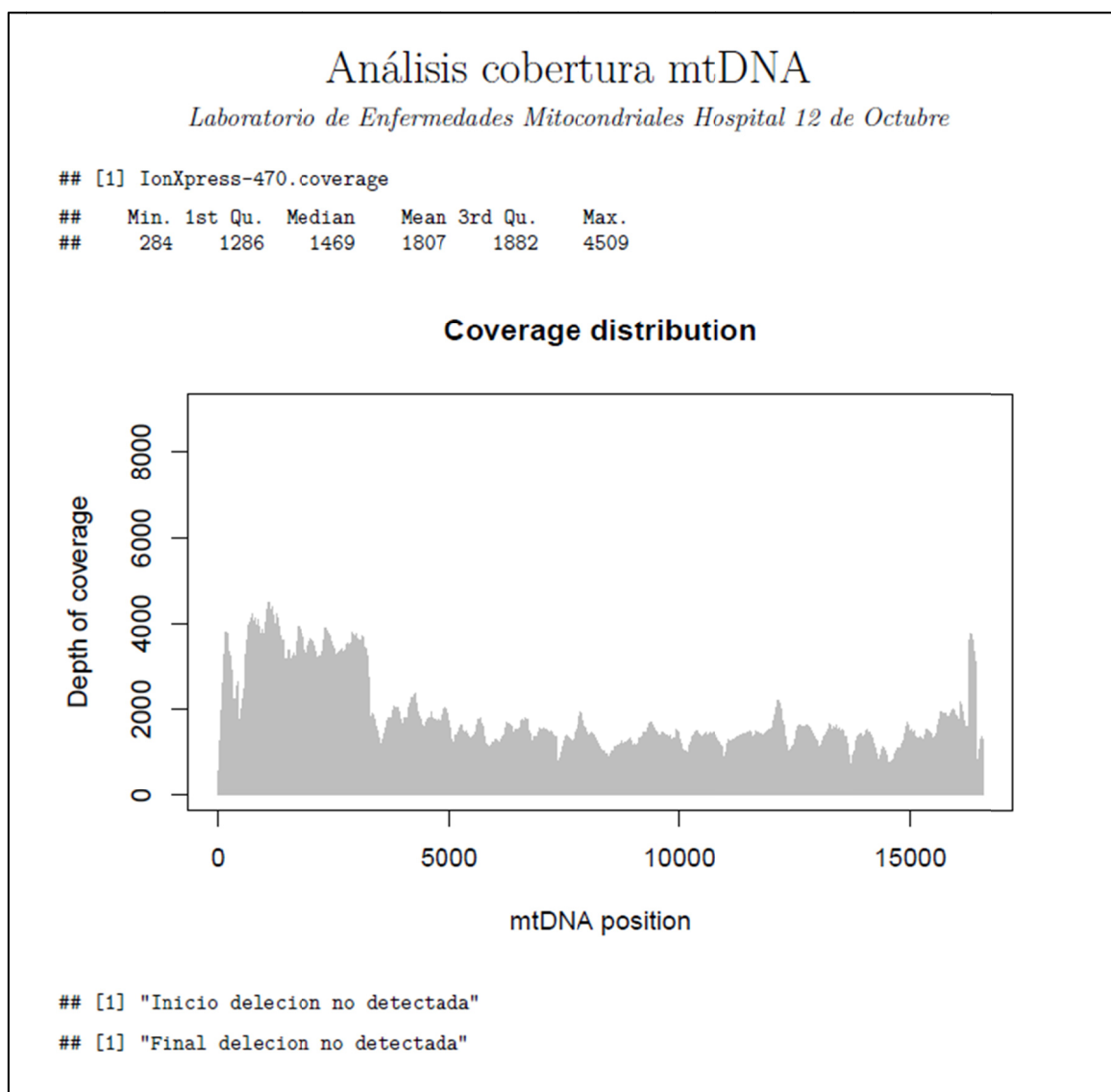


Figura 35. Informe de cobertura mtDNA: cabecera y perfil de cobertura

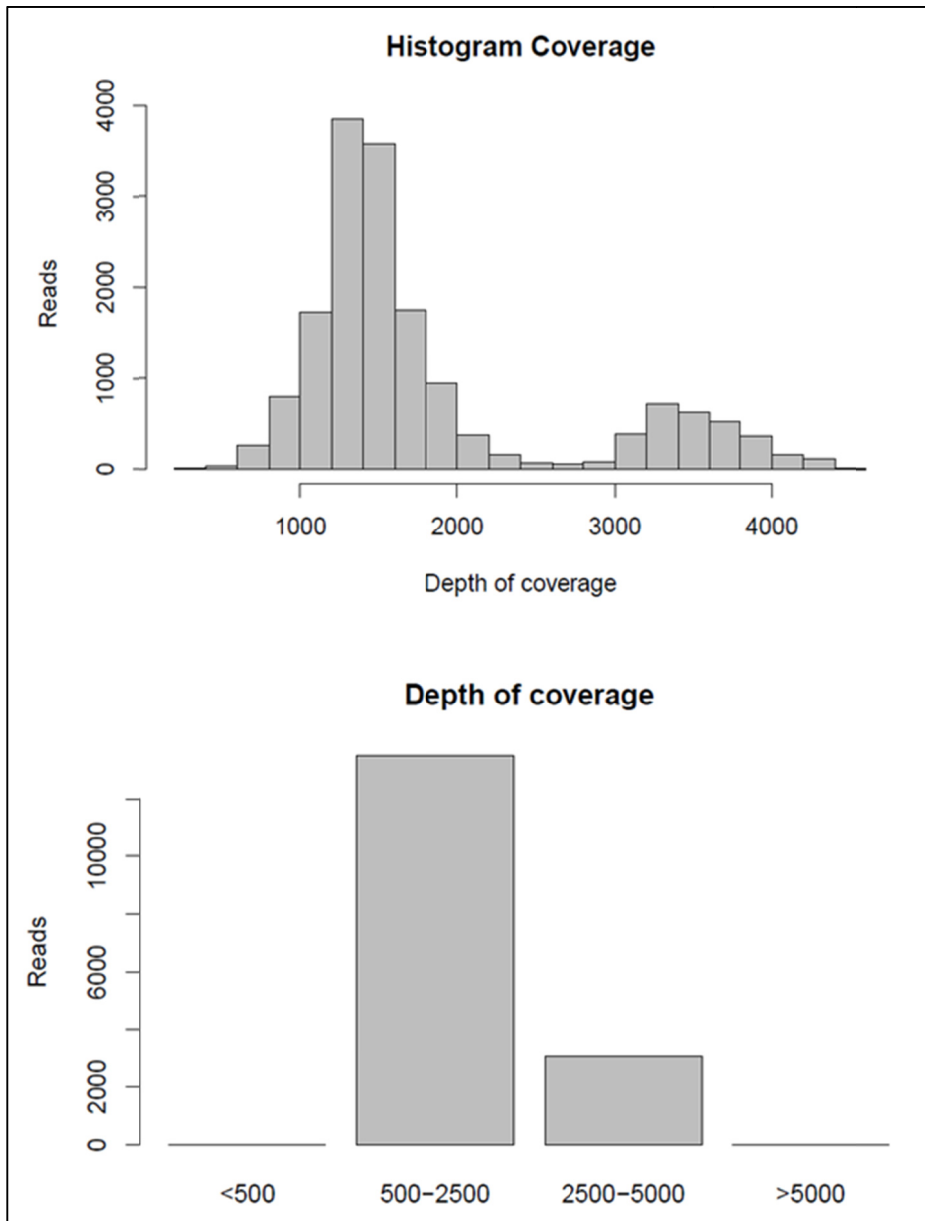


Figura 36. Informe de cobertura mtDNA: histogramas de cobertura (*Depth of coverage*)

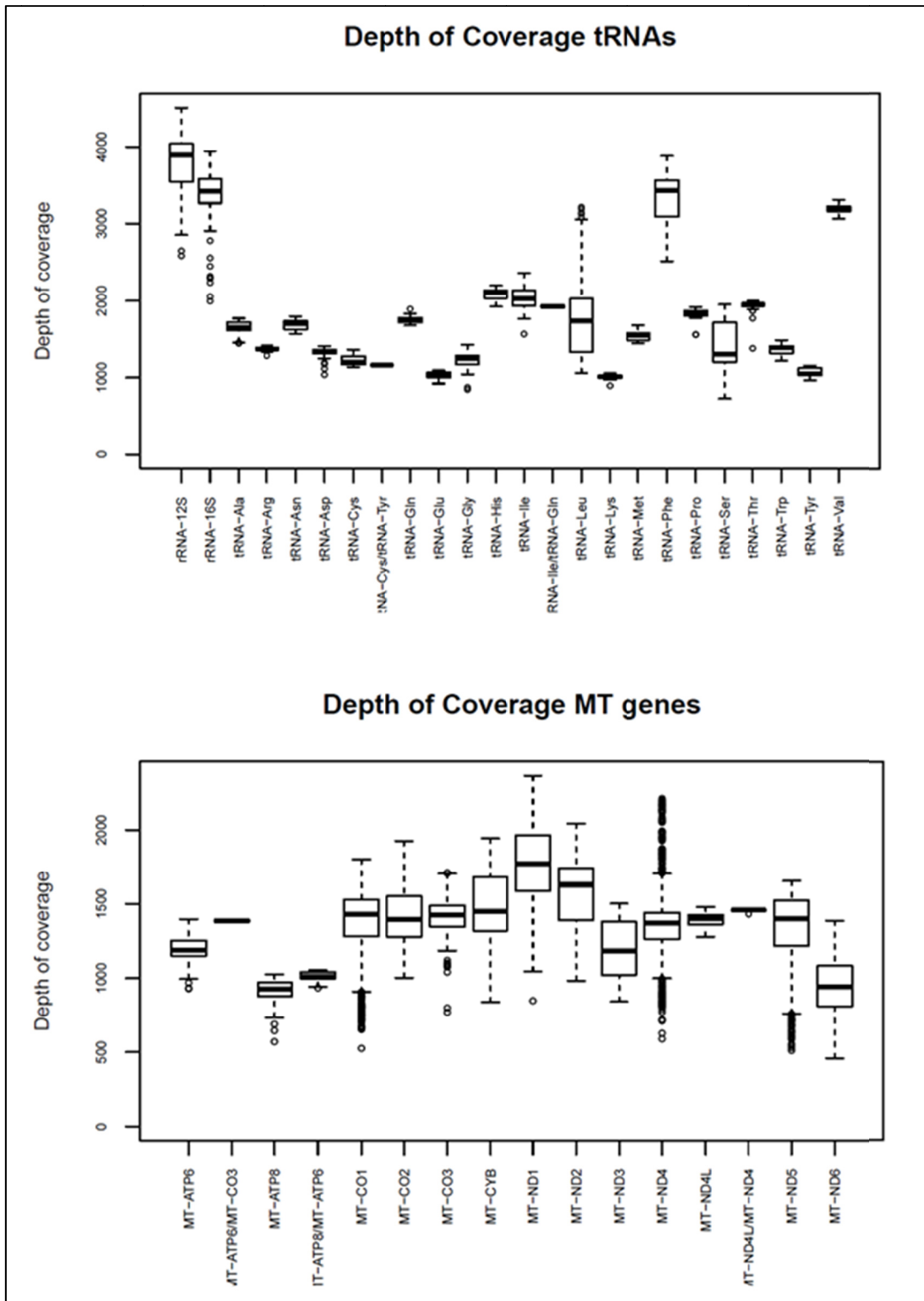


Figura 37. Informe de cobertura mtDNA: cobertura (*Depth of coverage*) desglosada por gen

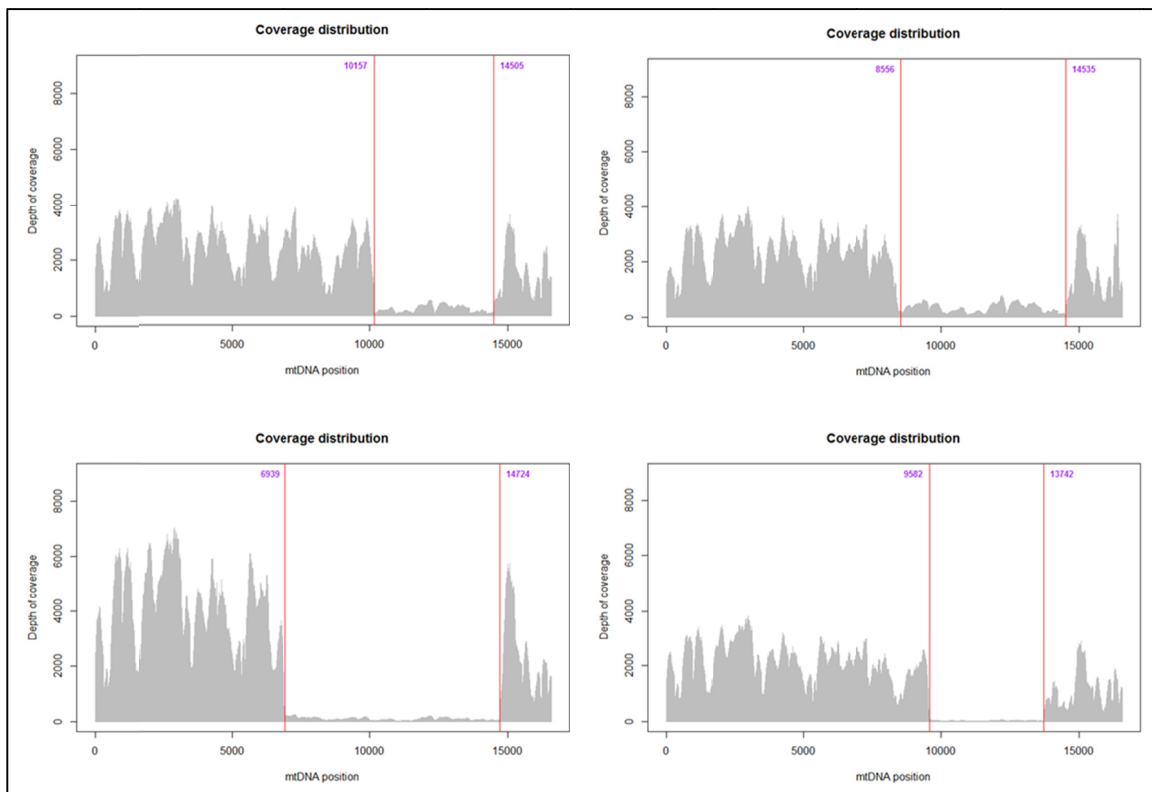


Figura 38. Ejemplos de perfiles de cobertura de mtDNA con detección de deleciones

8. Epidemiología análisis mtDNA alta profundidad

Se realizó la secuenciación del mtDNA a alta profundidad a 367 pacientes. En la Tabla 24 se muestra la distribución de los haplogrupos de la población estudiada con un evidente predominio, como era de esperar, de haplogrupos europeos y con un porcentaje de haplogrupo H coincidente con lo previsto para España.

En cuanto a las variantes encontradas se han identificado mutaciones patogénicas confirmadas o probablemente patogénicas en 39 casos (11%) y variantes de significado incierto en 15 casos (4%). Para caracterizar la patogenicidad de las mutaciones en tRNAs se ha tenido en cuenta el *Pathogenicity Scoring System* revisado en el año 2011¹⁰⁴. En las tablas siguientes (Tabla 25 y Tabla 26) se muestran las mutaciones separadas por su localización en tRNAs o en genes estructurales. Se han hallado en total 16 mutaciones diferentes en tRNAs, 23 mutaciones diferentes en genes estructurales del mtDNA y se han detectado también 10 deleciones únicas diferentes.

Tabla 24. Haplogrupos presentes en la población estudiada

Europeo	Haplogrupo	Pacientes	
92,3%	H	147	41,8%
	HV	13	3,7%
	I	5	1,4%
	J	24	6,8%
	K	21	6,0%
	R	3	0,9%
	T	27	7,7%
	U	54	15,3%
	V	15	4,3%
	W	7	2,0%
X	9	2,6%	
Asiático			
7,7%	B	11	3,1%
	C	4	1,1%
	D	1	0,3%
	M	8	2,3%
	N	3	0,9%
Africano			
3,7%	L0	1	0,3%
	L1	3	0,9%
	L2	2	0,6%
	L3	7	2,0%

Tabla 25. Mutaciones en tRNAs del mtDNA

Variantes patogénicas o probablemente patogénicas							
tRNA	Mutación	% Het.	MitoTIP	Nº GB Seqs	Cons.	Patología reportada	Refs
tRNA-Asn	m.5688T>C	73	<i>possibly pathogenic</i>	0 (0)	97.78%		
tRNA-Asn	m.5702A>G	60	<i>possibly pathogenic</i>	0 (0)	97.78%		
tRNACys	m.5803C>A	72	<i>possibly pathogenic</i>	0 (0)	95.56%		
tRNA-Gly	m.10010T>C	Homop.	<i>confirmed pathogenic</i>	0 (0)	100.00%	PEM	105,106
tRNA-Leu	m.12334G>A	22	<i>possibly pathogenic</i>	0 (0)	86.67%		
tRNA-Lys	m.8356T>C	21	<i>confirmed pathogenic</i>	0 (0)	26.67%	MERRF	107,108
tRNA-Lys	m.8317T>C	74	<i>possibly pathogenic</i>	0 (0)	91.11%		
tRNA-Pro	m.16015T>C	17	<i>possibly pathogenic</i>	0 (0)	100.00%	MM	109
tRNA-Ser	m.7511T>C	Homop.	<i>confirmed pathogenic</i>	1 (0)	91.11%	SNHL	110,111
tRNA-Trp	m.5517T>A	85	<i>possibly pathogenic</i>	0 (0)	86.67%		

Variantes con significado incierto							
tRNA	Mutación	% Het.	MitoTIP	Nº GB Seqs	Cons.	Patología reportada	Refs
tRNA-Arg	m.10440T>C	Homop.	<i>possibly pathogenic</i>	1 (0)	91.11%		112
tRNA-Gly	m.10004dupA	Homop.	NA	0 (0)	33.33%		
tRNA-Leu	m.12320A>G	95	<i>possibly benign</i>	0 (0)	97.78%	MM	105,113
tRNA-Ser	m.7466delC	Homop.	<i>possibly pathogenic</i>	12 (0)	97.78%		114,115
tRNA-Val	m.1643A>G	20	<i>possibly benign</i>	1 (0)	97.78%	LOFMD	116
tRNA-Val	m.1651A>G	Homop.	<i>possibly pathogenic</i>	0 (0)	46.67%		

PEM: Progressive Encephalomyopathy; MERRF: Myoclonic Epilepsy and Ragged Red Fiber disease; SNHL: Sensory Neural Hearing Loss; MM: Mitochondrial Myopathy; LOFMD: Late infantile onset fatal mitochondrial disease % Het.: Porcentaje Heteroplasmia; Homop.: Homoplásmica.

Tabla 26. Mutaciones en genes estructurales del mtDNA

Variantes patogénicas o probablemente patogénicas						
Gen	Mutación	% Het.	Nº GB Seqs	Cons.	Patología reportada	Refs
MT-ND1	m.3460G>A; p.A52T	25	23 (0)	91.11%	LHON	117–119
MT-ND1	m.3688G>A; p.A128T	85	0 (0)	93.33%	LS	120,121
MT-CO1	m.5992G>A; p.G30D	44	0 (0)	100.00%		
MT-ATP8	m.8424T>C; p.L20P	94	0 (0)	88.89%		
MT-ATP8	m.8403T>C; p.I13T	100	1 (0)	84.44%	PN	122
MT-ATP6	m.8579C>T; p.P18L	98	0 (0)	86.67%		
MT-ATP6	m.8959G>A; p.E145K	93	4 (0)	97.78%	DD / ID	123
MT-ATP6	m.9035T>C; p.L170P	100	0 (0)	100.00%	AS	124,125
MT-ATP6	m.9124A>G; p.200A	100	1 (0)	91.11%		
MT-ATP6	m.9176T>C; p.L217P	100	3 (0)	100.00%	FBSN / LS	126–128
MT-ATP6	m.9185T>C; p.L220P	100	3 (0)	97.78%	LS / AS / NARP-Id	129–131
MT-ND4	m.11778G>A; p.R340H	100	120 (0)	100.00%	LHON / PD	132,133
MT-ND5	m.12706T>C; p.F124L	79	0 (0)	100.00%	LS	134,135
MT-ND5	m.13340T>C; p.F335S	85	0 (0)	100.00%		
MT-ND6	m.14484T>C; p.M64V	100	56 (0)	31.11%	LHON	136,137
Variantes con significado incierto						
MT-ND1	m.3914G>A; p.G203E	79	1 (0)	100.00%		
MT-CO1	m.6510G>A; p.A203T	100	3 (0)	100.00%		138
MT-CO3	m.9333C>T; p.L43F	92	2 (0)	88.89%		
MT-CO3	m.9802A>C; p.V199A	42	6 (0)	57.78%		
MT-ND4	m.11360A>G; p.M201V	100	0 (0)	84.44%		
MT-CYB	m.15120C>G; p.A125G	100	0 (0)	86.67%		
MT-CYB	m.15209T>C; p.Y155H	48	4 (0)	100.00%	PW	139
MT-CYB	m.15776A>T; p.S344C	100	0 (0)	15.56%		

LHON: Leber's Hereditary Optic Neuropathy; LS: Leigh Syndrome; NARP: Neurogenic Muscle Weakness, Ataxia, and Retinitis Pigmentosa; NARP-Id: NARP-like disease; AS: Ataxia syndromes; FBSN: Familial Bilateral Striatal Necrosis; PD: Progressive Dystonia; PW: Prader-Willi syndrome; PN: Progressive Neuropathy; DD: Developmental Delay; ID: Intellectual Disability; % Het.: Porcentaje Heteroplasmia.

9. Estudio de mtDNA residual en datos de secuenciación de exomas

El uso de los métodos descritos en el apartado anterior (pág. 47) permitió la identificación de una variante en el mtDNA que ha dado lugar al siguiente artículo: *Whole-Exome Sequencing Identifies a Variant of the Mitochondrial MT-ND1 Gene Associated with Epileptic Encephalopathy: West Syndrome Evolving to Lennox-Gastaut Syndrome*. Human Mutation, 2013¹⁴⁰. La variante detectada es una mutación en heteroplasmia en el gen mitocondrial *ND1* (NC 012920.1:m.3946G>A, p.E214K) que codifica una subunidad del complejo I y que había sido asociada anteriormente a otras patologías.

En general, en las muestras analizadas con WES, se obtiene una cobertura media en casi todos los casos que ronda las 50 lecturas con una desviación estándar de 25. En este caso con la captura realizada con el kit *Agilent SureSelect Human all Exon 44Mb* se “mapearon” frente al mtDNA un 2,9% del total de lecturas, aunque este factor, tal y como se ha descrito, está muy condicionado por el tipo de kit de captura empleado¹⁴¹.

10. Caracterizar variantes genéticas modificadoras del fenotipo en pacientes con mutaciones patogénicas homoplásmicas en síndromes característicos del mtDNA

El objetivo de este apartado fue la identificación de alteraciones cuantitativas y variantes genéticas en el mtDNA que pudiesen contribuir a explicar la variabilidad fenotípica en la Neuropatía Óptica Hereditaria de Leber (LHON) causada por mutaciones homoplásmicas en el mtDNA.

Para ello se utilizaron las metodologías y desarrollos descritos en los apartados anteriores: secuenciación a alta profundidad del mtDNA (pág. 39) y un método de PCR cuantitativa (pág. 39).

Se estudiaron las siguientes familias y varios pacientes no relacionados (23 individuos en total):

- 2 familias con la mutación m.11778G>A (ambas familias con 1 afecto varón y 3 mujeres asintomáticas).
- 1 familia con la mutación m.3460G>A (1 afecto varón, 1 afecta mujer, 5 mujeres asintomáticas y 1 varón asintomático).
- 1 familia con la mutación m.14484T>C (1 afecto varón, 1 afecta mujer, 2 mujeres asintomáticas y 1 varón asintomático).
- Adicionalmente se estudiaron 6 pacientes LHON no relacionados: 2 con la mutación m.11778G>A y 4 con la mutación m.3460G>A.

Para el análisis del número de copias del mtDNA (depleción) se emplearon como controles muestras de DNA anónimas de pacientes entre 30 y 50 años donantes del banco de sangre del Hospital. No se encontraron diferencias significativas entre pacientes sintomáticos y asintomáticos de forma global, ni intrafamiliar.

En el análisis de mtDNA a alta profundidad tampoco se encontraron variantes en baja heteroplasmia potencialmente modificadoras del fenotipo. Se realizó un análisis global e intrafamiliar para detectar las variantes de mtDNA que eran diferentes en los probandos respecto a los familiares asintomáticos. En general, el resultado no fue positivo. En la familia con la mutación m.3460G>A, con dos probandos, únicamente uno de ellos mostró 3 variantes adicionales a bajos niveles de heteroplasmia (<12%) en los genes *MT-ND6*, *tRNA-Ala*, y *MT-ATP8* (esta última no considerable por su baja calidad). En la familia con la mutación m.14484T>C, se encontró una delección en el gen *MT-CO1*, a alto nivel de heteroplasmia, pero no pudo ser confirmada por secuenciación Sanger. En una de las familias portadoras de la mutación m.11778G>A no se priorizaron variantes diferentes. En la segunda familia el probando presentó dos variantes, una de la zona D-loop del mtDNA (una región muy polimórfica), y una inserción no considerable por mala calidad de secuencia (Tabla 27).

También se realizó este mismo análisis, pero buscando variantes en los asintomáticos, es decir como factores de protección; no encontrándose ninguna variante adicional en asintomáticos diferente de los probandos en ninguna de las familias con mutaciones primarias LHON estudiadas.

Estudio LHON		Posición	Ref	Variante	Estado	Het%	Qual	Tipo	Gen	Region	Cobertura	AA Cambio	GB Freq
FAM1 3460	P006	Sin variantes diferentes al resto de familiares asintomáticos											
	P009	14153	T	C	Het	11,9	76,43	SNP	MT-ND6	14149-14673	201	non-syn:N-S	7
		5592	A	G	Het	7,3	127,18	SNP	tRNA-Ala	5587-5655	826	tRNA	21
		8561	C	A	Het	2,5	6,42	SNP	MT-ATP8-ATP6	8527-8572	162		
FAM2 14484	P005	7423	A	-	Het	91,5	17136	DEL	MT-CO1	5904-7445	1595		
FAM3 11778	P001	Sin variantes diferentes al resto de familiares asintomáticos											
FAM4 11778	P003	14504	-	A	Het	2,3	12,09	INS	MT-ND6	14149-14673	1012		
		295	C	T	Het	95,1	19088	SNP	MT-OHR	110-441	1366		

Tabla 27. Resumen hallazgos secuenciación mtDNA familias LHON

No se encontró ninguna asociación con modificadores del fenotipo a nivel de copias mtDNA, ni de variantes puntuales del mtDNA a baja heteroplasmia. Como limitación, resaltar el bajo número de familias con miembros suficientes para poder incluirlas en el estudio.

Diseñar paneles de genes nucleares por resecuenciación masiva de escala media para diferentes grupos de patología mitocondrial: epidemiología molecular, estudios depleción, deleciones múltiples y estudio funcional tisular de la mutación p.T108M en el gen *TK2*.

Las capacidades de la secuenciación masiva están en constante evolución con la aparición de nuevos equipos y el desarrollo de diferentes aplicaciones. En estos momentos se encuentra mayoritariamente extendida la 2ª generación (NGS) post-secuenciación enzimática (Sanger). Esto implica la necesidad de amplificación del DNA para permitir su posterior detección y secuenciación mediante diferentes métodos basados en la detección de fluorescencia (Illumina), cambios de pH (Lifetechnologies) o liberación de pirofosfatos (Roche). Esta generación también está condicionada por la secuenciación de pequeños fragmentos que no suelen pasar de las 900pb como máximo en el mejor de los casos.

Existe una etapa fundamental que es la preparación del DNA para su secuenciación. En la actualidad se pueden diferenciar dos tipos de metodologías básicas para la generación de las colecciones de fragmentos (librerías): por un lado los métodos de amplicones y por otro los métodos de captura. Ambas metodologías presentan ventajas e inconvenientes que conviene valorar adecuadamente antes de realizar un proyecto de secuenciación.

En general, los métodos de amplicones son menos laboriosos, puede emplearse DNA más deteriorado y suelen ser más económicos. Las metodologías de captura tienen como puntos fuertes el garantizar el balance alélico y generalmente mejores coberturas.

Las principales ventajas del método de amplicones de *LifeTechnologies* (Ampliseq) empleado en el presente trabajo son el contar con una buena plataforma de diseño online de cebadores específicos, el uso de muy poco DNA de partida (10ng) y la posibilidad de generar amplicones de diversos tamaños (140pb, 275pb o 275pb). Los inconvenientes serían el no garantizar el balance alélico, la posibilidad de expansión clonal de errores de secuenciación y peores coberturas en ciertas regiones.

En este trabajo se desarrolló un *script* (Tabla 28) para integrar ANNOVAR en un flujo de trabajo automático (Figura 39) que permita anotar de manera múltiple y en paralelo diferentes muestras. Para ello se usó como lenguaje de programación Perl y se empleó el módulo gestor *Parallel::ForkManager* para paralelizar el análisis a nivel de núcleo.

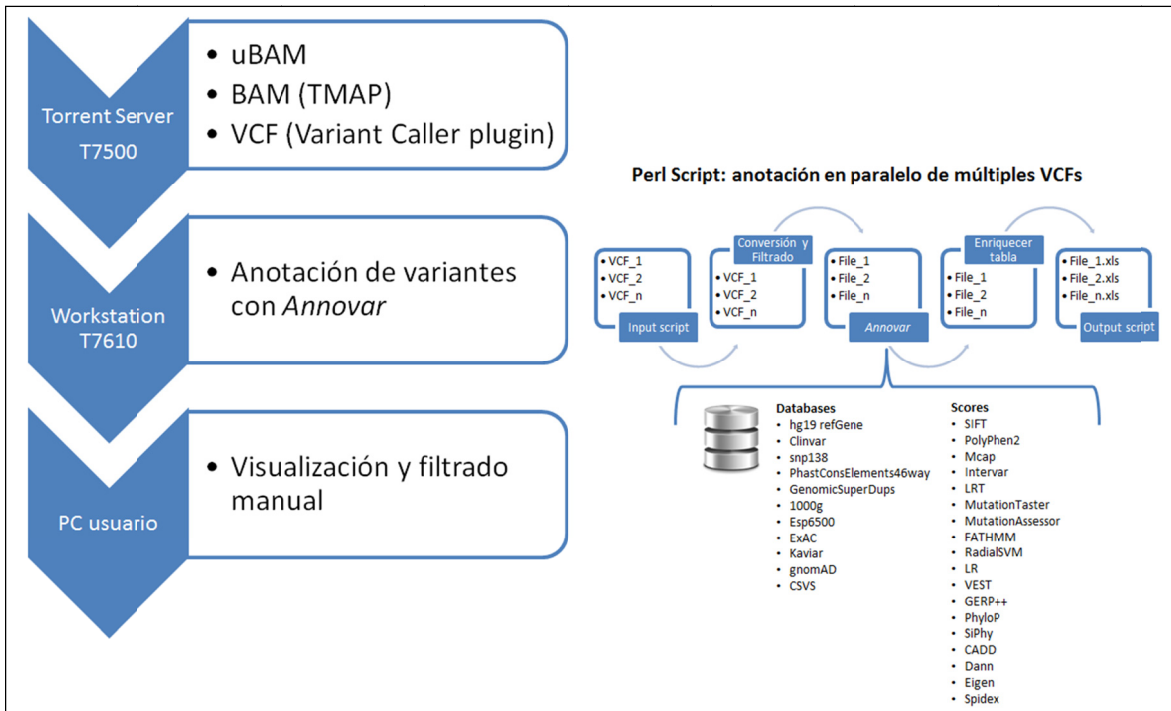


Figura 39. Esquema del pipeline de anotación de paneles de genes

Se crearon también filtros de variantes a partir de resultados propios para eliminar variantes frecuentes en nuestra población y simplificar el análisis.

El *Script* desarrollado se encarga de buscar y abrir los ficheros VCF que se encuentren en el directorio definido y comenzar el análisis en paralelo de tantas muestras como hilos tenga el equipo informático utilizado. Finalmente se generará una tabla en formato Excel con las variantes anotadas con las bases de datos y algoritmos deseados.

El tiempo de análisis desde el fichero VCF para este panel se optimizó para realizarse en aproximadamente 5 minutos (2x Intel Xeon E5-2650v2 (10 cores) y 64 GB RAM).

Tabla 28. Script de anotación de paneles de genes con Annotvar

```
#!/usr/bin/perl -w #
# SCRIPT: annovar.pl #
# Entrada: Deben indicarse los directorios de entrada (ficheros .vcf), de salida y el fichero de filtrado #
# filter0 = sin filtrado ; filter1 = panel mantenimiento ; filter2 = panel haloplex mitotrans ; filter2.1 = haloplex strand bias 1.0 ; filter3 = panel OXPPOS ; filter4 = panel intolerancias
# Ejemplo: perl annovarcadd.pl directorio_vcf directorio_results filter0
```

```
# Salida: tabla de variantes anotadas #
#####

# Activar la declaracion estricta de variables
use strict;
use Term::ANSIColor;
use Parallel::ForkManager;
my $max_processors=20;
my $mitocarta="/home/nfs1/annovar/humandb/mitocarta.txt";
my $bcftools="/home/nfs1/programs/bcftools/bcftools";
my $hg19ref="/home/nfs1/annovar/humandb/hg19torrent.fasta";
## Paso 1. Adquirir el nombre del directorio
my ($directorio_in,$directorio_out,$FILTER,$n_argumentos);
$n_argumentos = scalar(@ARGV);
($n_argumentos == 3) or
print_error("Error argumentos. Indica directorios de entrada, salida y
filtro");
($directorio_in,$directorio_out,$FILTER) = @ARGV;
print_ok("Lectura de directorios");
## Paso 2. Abrir el directorio
my ($fichero,$fichero_in,$fichero_out,$fichero_out2,$fichero_wrk);
my ($nombre,$comando,$rtcd);
opendir(DIR,$directorio_in);
print_ok("Acceso al directorio de ficheros: $directorio_in");
my $super_fork= new Parallel::ForkManager($max_processors);
while ($fichero = readdir(DIR))
{
    ## Control sobre los directorios y ejecución de programas externos
    next unless ($fichero =~ m/\.vcf$/);
    next if ($fichero =~ /^\.\/);
##### Start fork #####
my $pid= $super_fork->start and next;
## Paso 3. Normalización a la izquierda. Paso 1
($nombre = $fichero) =~ s/\.vcf//;
    $fichero_in = "$directorio_in/$fichero";
    $fichero_out = "$directorio_out/$nombre.step1.vcf";
    $comando = "$bcftools norm -m-both $fichero_in -o $fichero_out";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Transformar loci multialélicos en bialélicos:
$nombre");
## Paso 4. Normalización a la izquierda. Paso 2
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.leftnorm.vcf";
    $comando = "$bcftools norm -f /$hg19ref $fichero_in -o $fichero_out";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Normalización izquierda: Referencia ->
hg19torrent.fasta: $nombre");
## Paso 5. Convertir VCF a formato Annovar
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.convert";
    $comando = "perl convert2annovar.pl -format vcf4 $fichero_in >
$fichero_out";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Conversión de VCF: $nombre");
## Paso 6. Filtrar por variantes conocidas propias
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.txt";
    $comando = "cat $fichero_in | grep -v -w -f $FILTER >
$fichero_out";
```

```

    $rtcd = system($comando);
    rt_ctrl($rtcd, "Filtrado de variantes comunes: $nombre");
## Paso 7. Generar tabla anotada
    $fichero_in = $fichero_out;
    $fichero_out = $fichero_out.".hg19_multianno.txt";
    $comando = "perl table_annotar.pl $fichero_in humandb/ -buildver
hg19 -out $fichero_out -remove -v -protocol
refGene,clinvar_20170905,esp6500si_all,esp6500_ea,1000g2015aug_all,gnomad
_exome,gnomad_genome,exac03,kaviar,intervar_20180118,dbnsfp33a,cadd,spide
x,dbscsnv11,snp138NonFlagged,phastConsElements46way,genomicSuperDups,popf
req_max_20150413,bier_frecmax,bd_interna_260617 -operation
g,f,f,f,f,f,f,f,f,f,f,f,f,f,r,r,f,f,f -arg '-splicing
15',,,,,,,,,,,,,,,,,,,,,, -nastring NA";
    $rtcd = system($comando);
## Paso 8. Incluir encabezamiento
    $fichero_in = $fichero_in;
    $comando = "sed -i
'liChr\tStart\tEnd\tRef\tAlt\tZygotity\tQual\tCoverage' $fichero_in";
    $rtcd = system($comando);
## Paso 9. Fusion de tablas
    $fichero_wrk = $fichero_in;
    $fichero_in = $fichero_out;
    $fichero_out = "$directorio_out/$nombre.fusion";
    $comando = "paste $fichero_wrk $fichero_in > $fichero_out";
    $rtcd = system($comando);
## Paso 10. Eliminar columnas repetidas
    $fichero_in = $fichero_out;
    $fichero_out = "$directorio_out/$nombre-annotadohg19.1";
    $comando = "cut -f9,10,11,12,13 --complement $fichero_in | cut -
f51-78 --complement | sed -e 's/ /_/g' - > $fichero_out"; #¡OJO! Se
eliminan columnas con anotaciones de intervar
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Anotación de: $nombre");
## Paso 11. Reordenar columnas
    $fichero_in = $fichero_out;
    $fichero_out = "$directorio_out/$nombre-annotadohg19";
    $comando = "awk -f ordenamiento.awk $fichero_in > $fichero_out";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Reordenando: $nombre");
## Paso 12. Filtrar con Mitocarta
    $fichero_in = $fichero_out;
    $fichero_out2 = "$directorio_out/$nombre-annotadohg19-Mito";
    $comando = "cat $fichero_in | grep -w -f $mitocarta > $fichero_out2
| sed -i
'liChr\tStart\tEnd\tRef\tAlt\tZygotity\tQual\tCoverage\tFunc.refGene\tGen
e.refGene\tGeneDetail.refGene\tExonicFunc.refGene\tAACChange.refGene\tCLIN
SIG\tCLNDBN\tbd_interna_260617\tppopfreq_max\tbier_frecmax\t1000g2015aug_a
ll\tgnomAD_exome_ALL\tgnomAD_genome_ALL\tExAC_ALL\tKaviar_AF\tKaviar_AC\t
esp6500si_all\tInterVar(automated)\tM-CAP_score\tM-
CAP_pred\tCADD_phred\tSIFT_score\tSIFT_pred\tPolyphen2_HDIV_score\tPolyph
en2_HDIV_pred\tPolyphen2_HVAR_score\tPolyphen2_HVAR_pred\tLRT_score\tLRT_
pred\tMutationTaster_score\tMutationTaster_pred\tMutationAssessor_score\t
MutationAssessor_pred\tFATHMM_score\tFATHMM_pred\tPROVEAN_score\tPROVEAN_
pred\tVEST3_score\tMetaSVM_score\tMetaSVM_pred\tMetaLR_score\tMetaLR_pred
\tCADD_raw\tDANN_score\tfathmm-MKL_coding_score\tfathmm-
MKL_coding_pred\tEigen_coding_or_noncoding\tEigen-raw\tEigen-PC-
raw\tGenoCanyon_score\tintegrated_fitCons_score\tintegrated_confidence_va
lue\tGERP++_RS\tphyloP100way_vertibrate\tphyloP20way_mammalian\tphastCons
100way_vertibrate\tphastCons20way_mammalian\tSiPhy_29way_logOdds\tInterpr

```

```

o_domain\tCADD\tCADD_Phred\tdpsi_max_tissue\tdpsi_zscore\tdbscSNV_ADA_SCO
RE\tdbscSNV_RF_SCORE\tSNP138NonFlagged\tphastConsElements46way\tgenomicSu
perDups\tesp6500_ea\tCLNACC\tCLNDSDB\tCLNDSDBID\tgnomAD_exome_AFR\tgnomAD
_exome_AMR\tgnomAD_exome_ASJ\tgnomAD_exome_EAS\tgnomAD_exome_FIN\tgnomAD
_exome_NFE\tgnomAD_exome_OTH\tgnomAD_exome_SAS\tgnomAD_genome_AFR\tgnomAD
_genome_AMR\tgnomAD_genome_ASJ\tgnomAD_genome_EAS\tgnomAD_genome_FIN\tgnom
AD_genome_NFE\tgnomAD_genome_OTH\tExAC_AFR\tExAC_AMR\tExAC_EAS\tExAC_FIN\
tExAC_NFE\tExAC_OTH\tExAC_SAS' $fichero_out2";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Filtrando con Mitocarta: $nombre");
## Paso 13. Convertir a xls
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre";
    $comando = "python tsv2xlsx_annoar.py $fichero_in";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Conversión a xlsx: $fichero_out");
## Paso 14. Convertir a xls
    $fichero_in = "$fichero_out2";
    $fichero_out = "$directorio_out/$nombre";
    $comando = "python tsv2xlsx_annoar.py $fichero_in";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Conversión a xlsx: $fichero_out2");

##### end fork #####

$super_fork->finish;
}
$super_fork->wait_all_children;
## Paso 15. Borrado archivos intermedios
    my $dir = $directorio_out;
    unlink glob "$dir/*.txt & $dir/*.step1.vcf & $dir/*.fusion &
$dir/*.convert & $dir/*-anotadohg19.1 & $dir/*-anotadohg19 & $dir/*-
anotadohg19-Mito";
# Cerrar el directorio
closedir(DIR);
print_ok(":-) Finalizado. Resultados de cada muestra en directorio:
$directorio_out");
exit(0);

```

Los tiempos de análisis del script anterior se han optimizado para obtener los resultados siguientes (Tabla 29) en un sistema con 2 núcleos de 10 cores (Intel Xeon) y 64GB de RAM:

Tabla 29. Tiempos de análisis desde VCF

Panel Mantenimiento		Panel OXPHOS	
t_{real}	4m51.853s	t_{real}	4m55.100s

Generación de un informe de cobertura

En el ámbito clínico es muy relevante conocer exactamente las regiones no secuenciadas o con coberturas muy bajas para poder garantizar la ausencia de mutaciones patogénicas en los genes estudiados. Las herramientas comerciales disponibles no suelen ser demasiado exhaustivas y prácticamente nada configurables. Por ello se ha desarrollado un *script* que permite analizar en

paralelo diversas muestras y generar un informe final en el que se indican los genes y el número de exón que presenta alguna condición comprometida. Por defecto se emplea el criterio de que aparezcan los exones sin cobertura completa (horizontal) y los exones con cobertura (vertical: Depth of coverage) menor de 10.

El *script* usa como programas accesorios *Bedtools* y la herramienta AWK, y se apoya en datos de secuencia obtenidos del UCSC. Se parte de ficheros bam y se obtiene un fichero xls.

Tabla 30. Script para generar informe de cobertura de paneles

```
#!/usr/bin/perl -w #
# SCRIPT: nocovered.pl #
# Entrada: Deben indicarse los directorios de entrada (ficheros .bam), de
# salida y el fichero bed file#
# Ejemplo: perl nocovered.pl directorio_bams directorio_results bedfile
# Salida: tabla en formato xls con las posiciones no cubiertas #
#####

# Activar la declaracion estricta de variables
use strict;
use Term::ANSIColor;
use Parallel::ForkManager;
my $max_processors=20;
## Paso 1. Adquirir el nombre del directorio
my ($directorio_in,$directorio_out,$BEDFILE,$n_argumentos);
$n_argumentos = scalar(@ARGV);
($n_argumentos == 3) or
print_error("Error argumentos. Indica directorios de entrada, salida y
filtro");
($directorio_in,$directorio_out,$BEDFILE) = @ARGV;
print_ok("Lectura de directorios");
## Paso 2. Abrir el directorio
my
($fichero,$fichero_in,$fichero_in2,$fichero_out,$fichero_out1,$fichero_ou
t2,$fichero_out3);
my ($nombre,$comando,$rtcd);
opendir(DIR,$directorio_in);
print_ok("Acceso al directorio de ficheros: $directorio_in");
my $super_fork= new Parallel::ForkManager($max_processors);
while ($fichero = readdir(DIR))
{
    ## Control sobre los directorios y ejecución de programas externos
    next unless ($fichero =~ m/\.bam$/);
    next if ($fichero =~ /\^\.\/);
##### Start fork #####
my $pid= $super_fork->start and next;
## Paso 3. Conversión de BAM mediante Bamtobed
($nombre = $fichero) =~ s/\.bam//;
$fichero_in = "$directorio_in/$fichero";
$fichero_out1 = "$directorio_out/$nombre.1";
$comando = "bedtools bamtobed -i $fichero_in > $fichero_out1";
$rtcd = system($comando);
rt_ctrl($rtcd, "Conversión BAMtoBED: $nombre");
## Paso 4. Contaje en fichero 1 global por región
```

```
$fichero_in = "$fichero_out1";
$fichero_out2 = "$directorio_out/$nombre.2";
$comando = "bedtools coverage -a $fichero_in -b $BEDFILE | sort -V
-k1,1 -k2,2 - | bedtools intersect -wao -a - -b
UCSC_exons_modif_canonical.bed > $fichero_out2";
$rtcd = system($comando);
rt_ctrl($rtcd, "Contaje global: $nombre");
## Paso 5. Contaje en fichero 1 por posición
$fichero_in = "$fichero_out1";
$fichero_out3 = "$directorio_out/$nombre.3";
$comando = "bedtools coverage -d -a $fichero_in -b $BEDFILE | sort
-k8n - | awk -f min.awk - | sort -V -k1,1 -k2,2 - > $fichero_out3";
$rtcd = system($comando);
rt_ctrl($rtcd, "Contaje por posición: $nombre");
## Paso 6. Fusión de tablas
$fichero_in = $fichero_out2;
$fichero_in2 = $fichero_out3;
$fichero_out = "$directorio_out/$nombre.4";
$comando = "paste $fichero_in $fichero_in2 > $fichero_out";
$rtcd = system($comando);
## Paso 7. Reordenar columnas, seleccionar y añadir encabezado
$fichero_in = "$fichero_out";
$fichero_out = "$directorio_out/$nombre-nocubierto";
$comando = "awk -f selection.awk $fichero_in | (echo
'Chr\tStart_region\tEnd_region\tGene\tCov_Mean\tCov_Min\tBases_Cov\tBases
_Nocov\tTotal_Bases\tCov%\tStart_Exon\tEnd_Exon\tExon_number\tExon_Size';
cat -) > $fichero_out";
$rtcd = system($comando);
rt_ctrl($rtcd, "Reordenando: $nombre");
## Paso 7. Convertir a xls
$fichero_in = "$fichero_out";
$fichero_out = "$directorio_out/$nombre";
$comando = "python tsv2xlsx_annovar.py $fichero_in";
$rtcd = system($comando);
rt_ctrl($rtcd, "Conversión a xls: $nombre");

##### end fork #####

$super_fork->finish;
}
$super_fork->wait_all_children;
## Paso 9. Borrado archivos intermedios
my $dir = $directorio_out;
unlink glob "$dir/*.1 & $dir/*.2 & $dir/*.3 & $dir/*.4 & $dir/*-
nocubierto";
# Cerrar el directorio
closedir(DIR);
print_ok(":-) Finalizado. Resultados de cada muestra en directorio:
$directorio_out");
exit(0);
```


Chr	Start_region	End_region	Gene	Cov_Mear	Cov_Min	Bases_Cov	Bases_Nocov	Total_Bases	Cov%	Start_Exon	End_Exon	Exon_number	Exon_Size
chr1	12049225	12049400	MFN2	23	6	175	0	175	1	12049225	12049400	Exon_1	175
chr1	12052611	12052747	MFN2	12	8	136	0	136	1	12052611	12052747	Exon_2	136
chr1	12056212	12056375	MFN2	38	8	163	0	163	1	12056212	12056375	Exon_3	163
chr1	12057353	12057478	MFN2	2	2	125	0	125	1	12057353	12057478	Exon_4	125
chr1	12061457	12061611	MFN2	8	7	154	0	154	1	12061457	12061611	Exon_7	154
chr1	12061825	12061893	MFN2	9	9	68	0	68	1	12061825	12061893	Exon_8	68
chr1	12062038	12062160	MFN2	7	7	122	0	122	1	12062038	12062160	Exon_9	122
chr1	12066594	12066750	MFN2	8	7	156	0	156	1	12066594	12066750	Exon_14	156
chr1	12067109	12067306	MFN2	21	6	197	0	197	1	12067109	12067306	Exon_15	197
chr2	27534766	27534819	MPV17	12	5	53	0	53	1	27534766	27534819	Exon_6	53
chr2	27535081	27535114	MPV17	1	1	33	0	33	1	27535081	27535114	Exon_5	33
chr2	27535860	27535976	MPV17	4	4	116	0	116	1	27535860	27535976	Exon_2	116
chr2	27545314	27545384	MPV17	0	0	0	70	70	0	27545314	27545384	Exon_1	70
chr2	74154037	74154179	DGUOK	20	7	142	0	142	1	74154037	74154179	Exon_1	142
chr2	74177711	74177859	DGUOK	28	6	148	0	148	1	74177711	74177859	Exon_4	148
chr2	84650869	84650896	SUCLG1	12	7	27	0	27	1	84650869	84650896	Exon_9	27
chr2	84652538	84652727	SUCLG1	26	6	189	0	189	1	84652538	84652727	Exon_8	189
chr2	84658631	84658783	SUCLG1	30	9	152	0	152	1	84658631	84658783	Exon_7	152
chr2	84668170	84668228	SUCLG1	6	6	58	0	58	1	84668170	84668228	Exon_5	58
chr2	84676772	84676876	SUCLG1	15	2	104	0	104	1	84676772	84676876	Exon_2	104
chr3	193311166	193311198	OPA1	3	3	32	0	32	1	193311166	193311198	Exon_1	32
chr3	193332511	193332830	OPA1	60	9	319	0	319	1	193332511	193332830	Exon_2	319
chr3	193333462	193333559	OPA1	0	0	0	97	97	0	193333462	193333559	Exon_3	97
chr3	193334966	193335074	OPA1	54	0	100	8	108	0,9259259	193334966	193335074	Exon_4	108
chr3	193349400	193349454	OPA1	20	1	54	0	54	1	193349400	193349454	Exon_8	54
chr3	193353206	193353311	OPA1	38	3	105	0	105	1	193353206	193353311	Exon_9	105
chr3	193354983	193355070	OPA1	24	5	87	0	87	1	193354983	193355070	Exon_10	87
chr3	193361316	193361416	OPA1	33	0	84	16	100	0,84	193361316	193361416	Exon_15	100
chr3	193361763	193361894	OPA1	30	3	131	0	131	1	193361763	193361894	Exon_16	131
chr3	193363516	193363589	OPA1	24	5	73	0	73	1	193363516	193363589	Exon_18	73
chr3	193372650	193372816	OPA1	24	0	146	20	166	0,8795181	193372650	193372816	Exon_22	166
chr3	193374868	193375021	OPA1	20	3	153	0	153	1	193374868	193375021	Exon_23	153
chr3	193377270	193377350	OPA1	28	7	80	0	80	1	193377270	193377350	Exon_25	80
chr3	193409851	193409916	OPA1	9	9	65	0	65	1	193409851	193409916	Exon_30	65
chr4	186064526	186064637	SLC25A4	25	0	109	2	111	0,981982	186064526	186064637	Exon_1	111
chr4	186065917	186066404	SLC25A4	41	5	487	0	487	1	186065917	186066404	Exon_2	487
chr4	186066912	186067053	SLC25A4	1	1	141	0	141	1	186066912	186067053	Exon_3	141
chr4	186067967	186068125	SLC25A4	19	5	158	0	158	1	186067967	186068125	Exon_4	158
chr8	103220360	103220513	RRM2B	28	6	153	0	153	1	103220360	103220513	Exon_9	153
chr8	103231041	103231175	RRM2B	36	1	134	0	134	1	103231041	103231175	Exon_6	134

Figura 40. Ejemplo de tabla de cobertura con las regiones con problemas de cobertura

11. Panel genes de mantenimiento del mtDNA

En colaboración con el Hospital Vall'hebron se diseñó un panel con 13 genes que se encuentran implicados en la estabilidad del mtDNA y en los que se han descrito mutaciones patogénicas. Cuatro genes están implicados en la maquinaria de replicación del mtDNA: *POLG* (*Mitochondrial DNA Polymerase Gamma Catalytic Subunit*, GeneID 5428), *POLG2* (*Mitochondrial DNA Polymerase Accessory Subunit*, GeneID 11232), *TWINK* (*C10orf2*, *Twinkle T7-Like Mitochondrial DNA Helicase*, GeneID 56652); 7 genes se relacionan con el pool de dNTPs: *SLC25A4* (*ADP/ATP Translocase*, GeneID 291), *TK2* (*Thymidine Kinase 2*, GeneID 7084), *DGUOK* (*Deoxyguanosine Kinase*, GeneID 1716), *RRM2B* (*Ribonucleotide Reductase M2 B TP53 Inducible*, GeneID 50484), *MPV17* (*Mitochondrial Inner Membrane Protein*, GeneID 4358), *SUCLA2* (*ATP-Specific Succinyl-CoA Synthetase, Beta Subunit*, GeneID 8803), *SUCLG1* (*Succinyl-CoA Synthetase Subunit Alpha*, GeneID 8802); dos con la dinámica y remodelado mitocondrial: *MFN2* (*Mitofusin 2*, GeneID 9927) y *OPA1* (*Mitochondrial Dynamin Like GTPase*, GeneID 4976) y uno relacionado con el síndrome MNGIE^{142,143}: *TYMP* (*Thymidine Phosphorylase*, GeneID 1890).

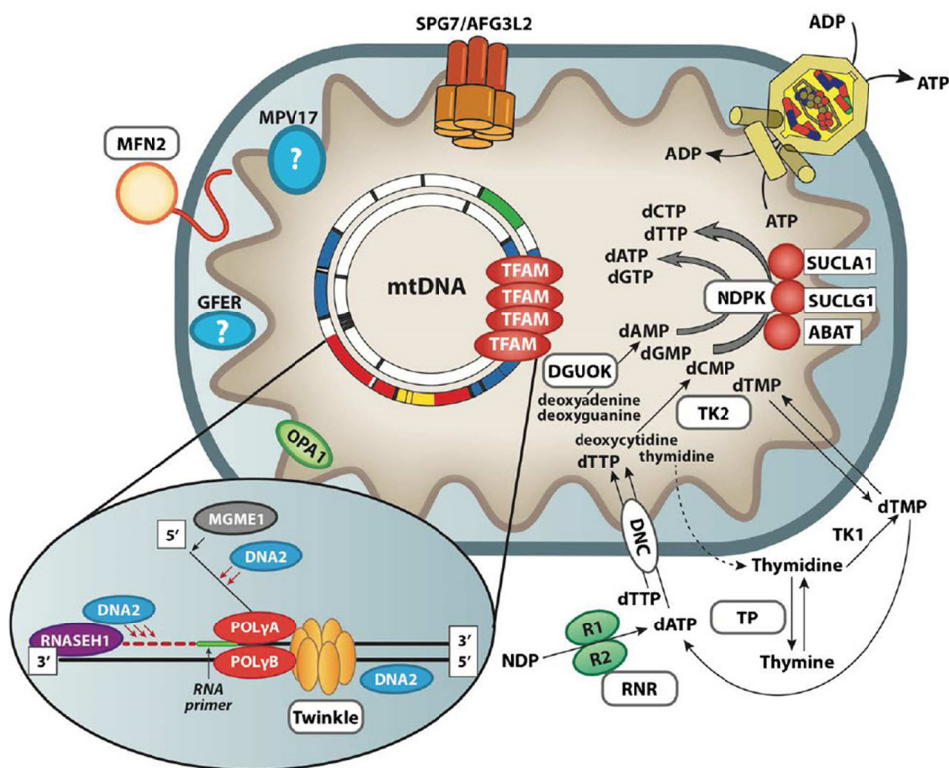


Figura 41. Rutas implicadas en el mantenimiento del mtDNA¹⁴⁴

Los genes de este panel causan lesiones en el mtDNA tales como deleciones múltiples y reducción del número de copias (síndrome de depleción mitocondrial). Los fenotipos más frecuentemente asociados a mutaciones en estos genes son oftalmoplejía progresiva externa, miopatía, hipotonía, encefalopatía, hepatopatía y disfunción gastrointestinal.

Se ha analizado con este panel a 351 pacientes con sospecha o con pruebas de depleción o deleción positivas. Se encontraron mutaciones patogénicas, probablemente patogénicas o de significado incierto (VUS) en el 23% de los probandos (Figura 42).

Por otro lado, los tres genes que acumulan un mayor número de mutaciones y además mayoritariamente patogénicas son *TK2*, *POLG* y *TWINKLE* (Figura 43).

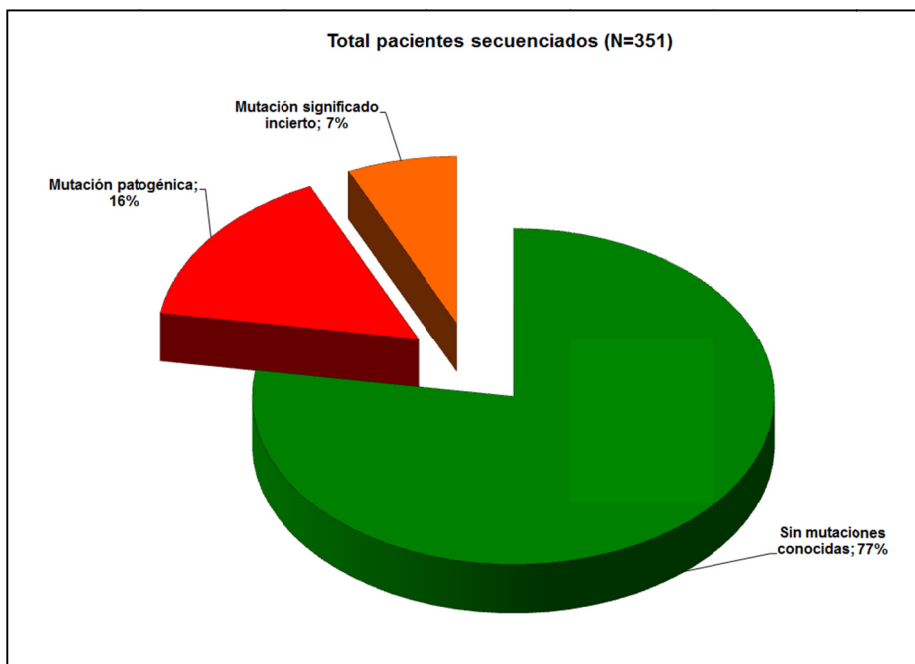


Figura 42. Resumen resultados pacientes Panel Mantenimiento.

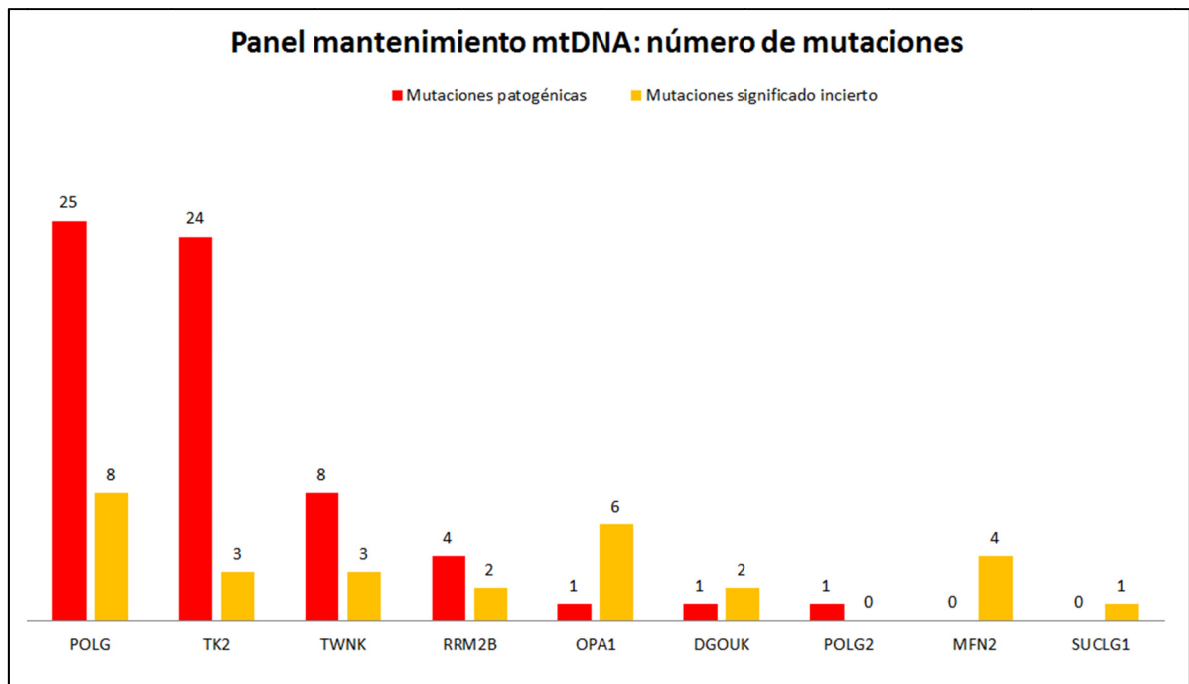


Figura 43. Mutaciones Panel Mantenimiento

En total se detectaron 93 mutaciones candidatas (60 diferentes) desde variantes patogénicas hasta variantes de significado incierto (VUS, Figura 43). Éstas últimas suponen un 31% del total.

En la Tabla 31 se indican las mutaciones halladas que presentan evidencias bibliográficas de patogenicidad o se clasifican según los criterios de la ACMG (*American American College of Medical Genetics and Genomics*)¹⁴⁵ como probablemente patogénicas o de significado incierto. La codificación en categorías proviene de resultados de Intervar y Varsome^{79,99} basados en la mencionada clasificación de la ACMG del siguiente modo: 1=*Benign*, 2=*Likely benign*, 3=*Uncertain significance*, 4=*Likely pathogenic*, 5=*Pathogenic*.

Tabla 31. Mutaciones sospechosas halladas con el panel de mantenimiento

Mutaciones descritas como patogénicas, probablemente patogénicas o con significado incierto			
TK2, Thymidine Kinase 2, Mitochondrial HGNC: 11831; Entrez Gene: 7084; Ensembl: ENSG00000166548	Estado	Ref.	Cat.
NM_004614:exon5:c.C323T:p.T108M	Het./Hom.	146	
NM_004614:exon5:c.C361A:p.H121N	Heterocigosis	147	
NM_004614:exon8:c.604_606del:p.202_202del	Het./Hom.	148	
NM_004614:exon10:c.C760T:p.R254*	Homocigosis	149	
NM_004614:exon2:c.150dupA:p.S51fs	Heterocigosis	148	
NM_004614:exon8:c.G575A:p.R192K	Heterocigosis	150,151	
NM_004614:exon6:c.C388T:p.R130W	Heterocigosis	151	
NM_004614:exon9:c.A623G:p.Y208C	Homocigosis	151	
NM_004614:exon7:c.503delT:p.I168Tfs*2	Heterocigosis		5
NM_004614:exon6:c.G415A:p.A139T	Heterocigosis	152	3
NM_004614:exon10:c.A763G:p.I255V	Heterocigosis		3
POLG, Mitochondrial DNA Polymerase Accessory Subunit HGNC: 9179; Entrez Gene: 5428; Ensembl: ENSG00000140521	Estado	Ref.	Cat.
NM_001126131:exon10:c.C1760T:p.P587L	Heterocigosis	153,154	
NM_001126131:exon3:c.C752T:p.T251I	Het./Hom.	154,155	
NM_002693:exon16:c.C2557T:p.R853W	Heterocigosis	156	
NM_001126131:exon18:c.C2827T:p.R943C	Heterocigosis	157	
NM_001126131:exon10:c.C1760T:p.P587L	Het./Hom.	153	
NM_002693:exon18:c.G2828A:p.R943H	Heterocigosis	155	
NM_001126131:exon22:c.G3550A:p.D1184N	Heterocigosis	158	
NM_001126131:exon16:c.G2542A:p.G848S	Heterocigosis		5
NM_002693:exon2:c.632_633insC:p.L211Ffs*33	Heterocigosis		5
NM_001126131:exon18:c.A2864G:p.Y955C	Heterocigosis		4
NM_001126131:exon7:c.G1399A:p.A467T	Heterocigosis		4
NM_002693:exon4:c.T911G:p.L304R	Heterocigosis		4
NM_001126131:exon16:c.C2573T:p.T858I	Heterocigosis		4
NM_001126131:exon16:c.A2492G:p.Y831C	Heterocigosis		3
NM_001126131:exon18:c.A2834T:p.H945L	Heterocigosis		3

NM_001126131:exon13:c.A2207G:p.N736S	Heterocigosis		3
NM_001126131:exon7:c.A1402G:p.N468D	Heterocigosis		3
NM_002693:exon22:c.G3631A:p.G1211R	Heterocigosis		3
NM_002693:exon10:c.C1832G:p.P611R	Heterocigosis		3
NM_001126131:exon3:c.G803C:p.G268A	Heterocigosis		3
<i>TWINK</i>, Twinkle T7-Like Mitochondrial DNA Helicase HGNC: 1160; Entrez Gene: 56652; Ensembl: ENSG00000107815	Estado	Ref.	Cat.
NM_001163812:exon1:c.G1070C:p.R357P	Heterocigosis	¹⁵⁹	
NM_001163812:exon1:c.G1084C:p.A362P	Heterocigosis	¹⁶⁰	
NM_001163812:exon1:c.G908A:p.R303Q	Heterocigosis	¹⁶¹	
NM_001163812:exon1:c.G1121A:p.R374Q	Heterocigosis	¹⁶²	
NM_021830:exon2:c.T1433G:p.F478C	Heterocigosis		3
NM_001163812:exon1:c.C937T:p.R313W	Heterocigosis		3
NM_021830:exon1:c.T1087A:p.W363R	Heterocigosis		3
NM_021830:exon2:c.T1361G:p.V454G	Heterocigosis		3
<i>RRM2B</i>, Ribonucleotide Reductase M2 B TP53 Inducible HGNC: 17296; Entrez Gene: 50484; Ensembl: ENSG00000048392	Estado	Ref.	Cat.
NM_015713:exon8:c.G817A:p.G273S	Heterocigosis	¹⁶³	
NM_015713:exon4:c.G329A:p.R110H	Homocigosis	¹⁶⁴	
NM_015713:exon9:c.G1045A:p.A349T	Heterocigosis	¹⁶⁵	3
NM_015713:exon2:c.A139G:p.I47V	Heterocigosis		3
NM_015713:exon8:c.A791T:p.E264V	Heterocigosis		3
NM_015713:exon4:c.G343T:p.V115L	Heterocigosis		3
<i>OPA1</i>, Mitochondrial Dynamin Like GTPase HGNC: 8140; Entrez Gene: 4976; Ensembl: ENSG00000198836	Estado	Ref.	Cat.
NM_130837:exon23:c.C2296T:p.R766*	Heterocigosis		5
NM_130837:exon12:c.A1189G:p.K397E	Heterocigosis		3
NM_130837:exon15:c.T1418C:p.I473T	Heterocigosis		3
NM_130837:exon15:c.T1418A:p.I473N	Heterocigosis		3
NM_130837:exon27:c.A2703T:p.L901F	Heterocigosis		3
NM_130837:exon27:c.A2706C:p.K902N	Heterocigosis		3
NM_130837:exon12:c.1150-8A>G	Heterocigosis		3

<i>DGUOK</i>, Mitochondrial Dynamin Like GTPase HGNC: 2858; Entrez Gene: 1716; Ensembl: ENSG00000114956	Estado	Ref.	Cat.
NM_080916:exon5:c.A677G:p.H226R	Homocigosis	¹⁶⁶	
NM_080916:exon2:c.C211G:p.P71A	Heterocigosis		3
NM_080916:exon3:c.G289A:p.D97N	Heterocigosis		3
<i>POLG2</i>, Mitochondrial DNA Pol. Accessory Subunit HGNC: 9180; Entrez Gene: 11232; Ensembl: ENSG00000256525	Estado	Ref.	Cat.
NM_007215:exon5:c.A1105G:p.R369G	Heterocigosis	¹⁶⁷	
<i>MFN2</i>, Mitofusin 2 HGNC: 16877; Entrez Gene: 9927; Ensembl: ENSG00000116688	Estado	Ref.	Cat.
NM_001127660:exon2:c.C58T:p.H20Y	Heterocigosis		3
NM_001127660:exon13:c.G1403A:p.R468H	Heterocigosis		3
<i>SUCLG1</i>, Succinate-CoA Ligase Alpha Subunit HGNC: 11449; Entrez Gene: 8802; Ensembl: ENSG00000163541	Estado	Ref.	Cat.
NM_003849:exon3:c.G236A:p.G79D	Heterocigosis		3

Tabla 32. Mutaciones descritas ClinVar (07/2019) Panel Mantenimiento (Genes con hallazgos)

	<i>POLG</i>	<i>TK2</i>	<i>TWNK</i>	<i>RRM2B</i>	<i>OPA1</i>	<i>DGUOK</i>	<i>POLG2</i>	<i>MFN2</i>	<i>SUCLG1</i>
ClinVar	111	51	37	66	108	26	14	62	16
Cat. 5	21%	46%	37%	63%	46%	35%	31%	23%	26%
ClinVar	50	3	13	9	28	4	2	39	7
Cat. 4	9%	3%	13%	9%	12%	6%	4%	14%	12%
ClinVar	368	56	49	30	100	43	29	169	38
Cat. 3	70%	51%	49%	28%	42%	59%	65%	63%	62%
ClinVar Total	529	110	99	105	236	73	45	270	61
ORF (pb)	3720	798	2829	2172	3048	1404	1458	2274	1041
Nº mut/pb	3,0	6,4	1,3	3,0	3,5	1,9	1,0	2,7	1,5

El gen que presenta con gran diferencia más variantes descritas (de Cat. 3 a Cat.5) en ClinVar es *POLG*, y ha sido también el gen en el que se han detectado más variantes en nuestra serie, aunque si se considera el tamaño de la región codificante el gen que acumula un mayor número de mutaciones en términos relativos es *TK2*. En ClinVar en segundo y tercer lugar se sitúan dos genes relacionados con la dinámica y remodelado mitocondrial: *MFN2* y *OPA1*, mientras que en nuestra serie destaca el gen *TK2* relacionado con el *pool* de nucleótidos y *TWNK*

implicado con la maquinaria de replicación. Es destacable la alta tasa de mutaciones patogénicas o probablemente patogénicas encontradas en nuestra población en el gen *TK2*.

En este diseño *TYMP* ha presentado problemas de cobertura graves (por problemas de unión de los cebadores diseñados) que lo han invalidado prácticamente por completo.

12. Panel genes del Sistema de fosforilación oxidativa mitocondrial (OXPHOS)

Este panel lo conforman 133 genes del sistema de cadena respiratoria y fosforilación oxidativa mitocondrial. Se incluyen todas las subunidades estructurales de la cadena respiratoria y ATP sintasa codificadas por el nDNA y los factores de ensamblaje del sistema OXPHOS conocidos hasta el momento.

Los fenotipos más frecuentemente asociados a mutaciones en estos genes son miopatía, acidosis láctica, convulsiones, ataxia, neuropatía periférica, ceguera, sordera, dismotilidad gastrointestinal, fallo hepático, y disfunción de la médula ósea¹.

Los genes incluidos en este panel son: *ACAD9* (*acyl-CoA dehydrogenase family member 9*), *ATPAF1* (*ATP synthase mitochondrial F1 complex assembly factor 1*), *ATPAF2* (*ATP synthase mitochondrial F1 complex assembly factor 2*), *BCS1L* (*BCS1 homolog, ubiquinol-cytochrome c reductase complex chaperone*), *CMC1* (*C-X9-C motif containing 1*), *CMC2* (*C-X9-C motif containing 2*), *COA1* (*cytochrome c oxidase assembly factor 1 homolog*), *COA3* (*cytochrome c oxidase assembly factor 3*), *COA4* (*cytochrome c oxidase assembly factor 4 homolog*), *COA5* (*cytochrome c oxidase assembly factor 5*), *COA6* (*cytochrome c oxidase assembly factor 6*), *COA7* (*cytochrome c oxidase assembly factor 7 (putative)*), *COX10* (*cytochrome c oxidase assembly factor heme A:farnesyltransferase COX10*), *COX11* (*cytochrome c oxidase copper chaperone COX11*), *COX14* (*cytochrome c oxidase assembly factor COX14*), *COX15* (*cytochrome c oxidase assembly homolog COX15*), *COX16* (*cytochrome c oxidase assembly factor COX16*), *COX17* (*cytochrome c oxidase copper chaperone COX17*), *COX18* (*cytochrome c oxidase assembly factor COX18*), *COX19* (*cytochrome c oxidase assembly factor COX19*), *COX20* (*cytochrome c oxidase assembly factor COX20*), *COX4I1* (*cytochrome c oxidase subunit 4I1*), *COX4I2* (*cytochrome c oxidase subunit 4I2*), *COX5A* (*cytochrome c oxidase subunit 5A*), *COX5B* (*cytochrome c oxidase subunit 5B*), *COX6A1* (*cytochrome c oxidase subunit 6A1*), *COX6A2* (*cytochrome c oxidase subunit 6A2*), *COX6B1* (*cytochrome c oxidase subunit 6B1*), *COX6B2* (*cytochrome c oxidase subunit 6B2*), *COX6C* (*cytochrome c oxidase subunit 6C*), *COX7A1* (*cytochrome c oxidase subunit 7A1*), *COX7A2* (*cytochrome c oxidase subunit 7A2*), *COX7B* (*cytochrome c oxidase subunit 7B*), *COX7B2*

(cytochrome c oxidase subunit 7B2), COX7C (cytochrome c oxidase subunit 7C), COX8A (cytochrome c oxidase subunit 8A), COX8C (cytochrome c oxidase subunit 8C), CYC1 (cytochrome c1), ECSIT (ECSIT signalling integrator), FASTK (Fas activated serine/threonine kinase), FASTKD1 (FAST kinase domains 1), FASTKD2 (FAST kinase domains 2), FASTKD3 (FAST kinase domains 3), FASTKD5 (FAST kinase domains 5), FOXRED1 (FAD dependent oxidoreductase domain containing 1), LRPPRC (leucine rich pentatricopeptide repeat containing), LYRM7 (LYR motif containing 7), NDUFA1 (NADH:ubiquinone oxidoreductase subunit A1), NDUFA10 (NADH:ubiquinone oxidoreductase subunit A10), NDUFA11 (NADH:ubiquinone oxidoreductase subunit A11), NDUFA12 (NADH:ubiquinone oxidoreductase subunit A12), NDUFA13 (NADH:ubiquinone oxidoreductase subunit A13), NDUFA2 (NADH:ubiquinone oxidoreductase subunit A2), NDUFA3 (NADH:ubiquinone oxidoreductase subunit A3), NDUFA4 (NDUFA4, mitochondrial complex associated), NDUFA5 (NADH:ubiquinone oxidoreductase subunit A5), NDUFA6 (NADH:ubiquinone oxidoreductase subunit A6), NDUFA7 (NADH:ubiquinone oxidoreductase subunit A7), NDUFA8 (NADH:ubiquinone oxidoreductase subunit A8), NDUFA9 (NADH:ubiquinone oxidoreductase subunit A9), NDUFAB1 (NADH:ubiquinone oxidoreductase subunit AB1), NDUFAF1 (NADH:ubiquinone oxidoreductase complex assembly factor 1), NDUFAF2 (NADH:ubiquinone oxidoreductase complex assembly factor 2), NDUFAF3 (NADH:ubiquinone oxidoreductase complex assembly factor 3), NDUFAF4 (NADH:ubiquinone oxidoreductase complex assembly factor 4), NDUFAF5 (NADH:ubiquinone oxidoreductase complex assembly factor 5), NDUFAF6 (NADH:ubiquinone oxidoreductase complex assembly factor 6), NDUFAF7 (NADH:ubiquinone oxidoreductase complex assembly factor 7), NDUFB1 (NADH:ubiquinone oxidoreductase subunit B1), NDUFB10 (NADH:ubiquinone oxidoreductase subunit B10), NDUFB11 (NADH:ubiquinone oxidoreductase subunit B11), NDUFB2 (NADH:ubiquinone oxidoreductase subunit B2), NDUFB3 (NADH:ubiquinone oxidoreductase subunit B3), NDUFB4 (NADH:ubiquinone oxidoreductase subunit B4), NDUFB5 (NADH:ubiquinone oxidoreductase subunit B5), NDUFB6 (NADH:ubiquinone oxidoreductase subunit B6), NDUFB7 (NADH:ubiquinone oxidoreductase subunit B7), NDUFB8 (NADH:ubiquinone oxidoreductase subunit B8), NDUFB9 (NADH:ubiquinone oxidoreductase subunit B9), NDUFC1 (NADH:ubiquinone oxidoreductase subunit C1), NDUFC2 (NADH:ubiquinone oxidoreductase subunit C2), NDUFS1 (NADH:ubiquinone oxidoreductase core subunit S1), NDUFS2 (NADH:ubiquinone oxidoreductase core subunit S2), NDUFS3 (NADH:ubiquinone oxidoreductase core subunit S3), NDUFS4 (NADH:ubiquinone oxidoreductase subunit S4), NDUFS5 (NADH:ubiquinone oxidoreductase subunit S5), NDUFS6 (NADH:ubiquinone oxidoreductase subunit S6), NDUFS7 (NADH:ubiquinone oxidoreductase core subunit S7), NDUFS8 (NADH:ubiquinone oxidoreductase core subunit S8), NDUFV1 (NADH:ubiquinone oxidoreductase core subunit V1),

NDUFV2 (NADH:ubiquinone oxidoreductase core subunit V2), NDUFV3 (NADH:ubiquinone oxidoreductase subunit V3), NUBPL (nucleotide binding protein like), OXA1L (OXA1L, mitochondrial inner membrane protein), SCO1 (SCO1, cytochrome c oxidase assembly protein), SCO2 (SCO2, cytochrome c oxidase assembly protein), SDHA (succinate dehydrogenase complex flavoprotein subunit A), SDHAF1 (succinate dehydrogenase complex assembly factor 1), SDHAF2 (succinate dehydrogenase complex assembly factor 2), SDHB (succinate dehydrogenase complex iron sulfur subunit B), SDHC (succinate dehydrogenase complex subunit C), SDHD (succinate dehydrogenase complex subunit D), SURF1 (SURF1, cytochrome c oxidase assembly factor), TACO1 (translational activator of cytochrome c oxidase I), TBRG4 (transforming growth factor beta regulator 4), TIMMDC1 (translocase of inner mitochondrial membrane domain containing 1), TMEM126B (transmembrane protein 126B), TMEM70 (transmembrane protein 70), TTC19 (tetratricopeptide repeat domain 19), UQCC1 (ubiquinol-cytochrome c reductase complex assembly factor 1), UQCC2 (ubiquinol-cytochrome c reductase complex assembly factor 2), UQCR10 (ubiquinol-cytochrome c reductase, complex III subunit X), UQCR11 (ubiquinol-cytochrome c reductase, complex III subunit XI), UQCRB (ubiquinol-cytochrome c reductase binding protein), UQCRC1 (ubiquinol-cytochrome c reductase core protein 1), UQCRC2 (ubiquinol-cytochrome c reductase core protein 2), UQCRFS1 (ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1), UQCRH (ubiquinol-cytochrome c reductase hinge protein), UQCRQ (ubiquinol-cytochrome c reductase complex III subunit VII).

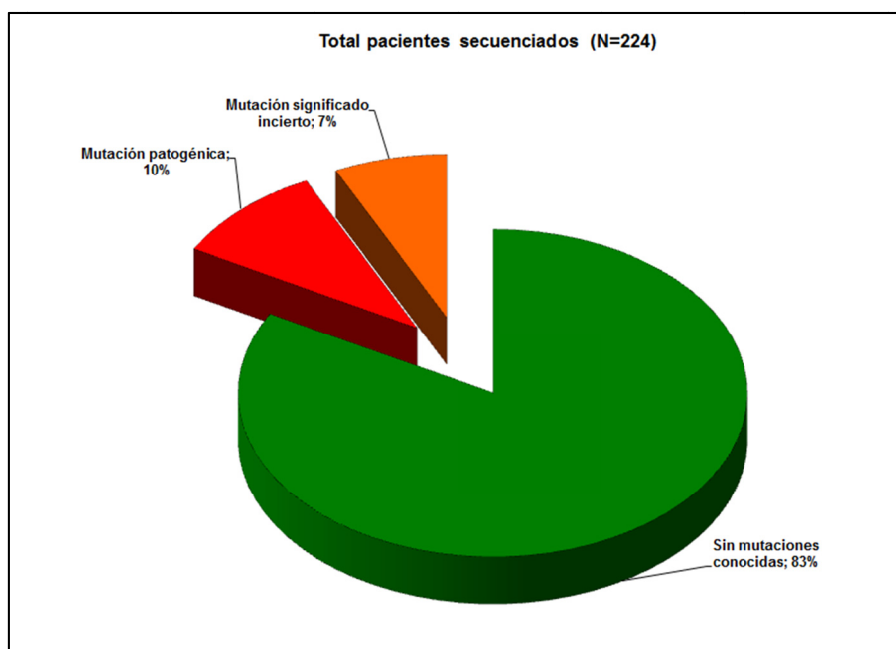


Figura 44. Resumen resultados Panel OXPHOS

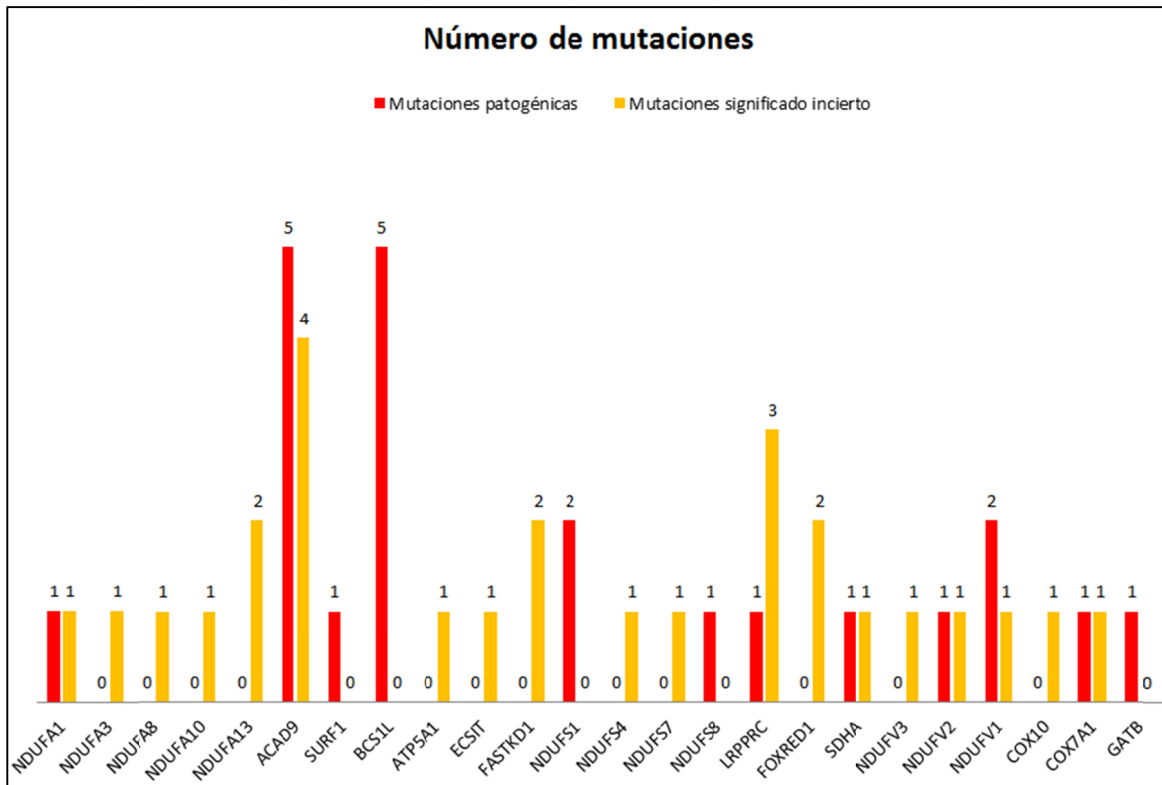


Figura 45. Mutaciones Panel OXPHOS

Se han detectado en total 49 mutaciones candidatas (46 diferentes) desde variantes patogénicas hasta variantes de significado incierto (VUS, Figura 45). Éstas últimas suponen un 55% del total.

Tabla 33. Mutaciones sospechosas halladas con el panel OXPHOS

Mutaciones descritas como patogénicas, probablemente patogénicas o con significado incierto			
ACAD9, Acyl-CoA Dehydrogenase Family Member 9 HGNC: 21497; Entrez Gene: 28976; Ensembl: ENSG00000177646	Estado	Ref.	Cat.
NM_014049:exon7:c.G797A:p.R266Q	Homocigosis	¹⁶⁸	
NM_014049:exon4:c.359delT:p.F120fs (p.Phe120Serfs*9)	Heterocigosis	¹⁶⁹	
NM_014049:exon12:c.1150-1G>A	Heterocigosis		5
NM_014049:exon3:c.C281T:p.P94L	Heterocigosis		4
NM_014049:exon7:c.C803T:p.S268F	Heterocigosis		4
NM_014049:exon2:c.T160G:p.F54V	Heterocigosis		3
NM_014049:exon9:c.G954C:p.L318F	Heterocigosis		3
NM_014049:exon7:c.C796T:p.R266W	Heterocigosis		3
NM_014049:exon6:c.A623G:p.N208S	Heterocigosis		3
SURF1, Cytochrome C Oxidase Assembly Factor HGNC: 11474; Entrez Gene: 6834; Ensembl: ENSG00000148290	Estado	Ref.	Cat.
NM_003172:exon9:c.845_846del:p.S282fs	Homocigosis	¹⁷⁰	
BCS1L, Ubiquinol-Cytochrome C Reductase Complex Chap. HGNC: 1020; Entrez Gene: 617; Ensembl: ENSG00000074582	Estado	Ref.	Cat.
NM_004328:exon3:c.C166T:p.R56*	Heterocigosis		5
NM_004328:exon4:c.C355T:p.R119X*	Heterocigosis		5
NM_004328:exon3:c.A148G:p.T50A	Heterocigosis		4
NM_004328:exon5:c.C550T:p.R184C	Heterocigosis		4
LRPPRC, Leucine Rich Pentatricopeptide Repeat Containing HGNC: 15714; Entrez Gene: 10128; Ensembl: ENSG00000138095	Estado	Ref.	Cat.
NM_133259:exon25:c.2726_2728del:p.K909del	Heterocigosis	¹⁷¹	
NM_133259:exon34:c.T3821C:p.L1274P	Heterocigosis		3
NM_133259:exon12:c.1370-2->GT	Heterocigosis		3
NM_133259:exon1:c.C58T:p.L20F	Heterocigosis		3
FOXRED1, FAD Dependent Oxidoreductase Domain HGNC: 26927; Entrez Gene: 55572; Ensembl: ENSG00000110074	Estado	Ref.	Cat.
NM_017547:exon5:c.T628G:p.Y210D	Heterocigosis		3
NM_017547:exon11:c.C1273T:p.H425Y	Heterocigosis		3

SDHA, Succinate Dehydrogenase Complex Flavoprotein HGNC: 10680; Entrez Gene: 6389; Ensembl: ENSG00000073578	Estado	Ref.	Cat.
SDHA:NM_004168:exon13:c.C1753T:p.R585W	Heterocigosis		4
SDHA:NM_004168:exon11:c.G1535A:p.R512Q	Heterocigosis		3
NDUFV1, NADH:Ubiquinone Oxidoreductase Subunit V1 HGNC: 7716; Entrez Gene: 4723; Ensembl: ENSG00000167792	Estado	Ref.	Cat.
NM_007103:exon3:c.T158C:p.L53P	Heterocigosis	170,171	
NM_007103:exon4:c.C365T:p.P122L	Heterocigosis	170,171	
NM_007103:exon8:c.1162+4A>C	Heterocigosis		4
NDUFV2, NADH:Ubiquinone Oxidoreductase Subunit V2 HGNC: 7717; Entrez Gene: 4729; Ensembl: ENSG00000178127	Estado	Ref.	Cat.
NM_021074:exon5:c.394dupA:p.I132Nfs*15	Heterocigosis		5
NM_021074:exon6:c.G547A:p.A183T	Heterocigosis		3
NDUFV3, NADH:Ubiquinone Oxidoreductase Subunit V3 HGNC: 7719 Entrez Gene: 4731 Ensembl: ENSG00000160194	Estado	Ref.	Cat.
NM_021075:exon2:c.A168C:p.K56N	Homocigosis		3
GATB, Glutamyl-TRNA Amidotransferase Subunit B HGNC: 8849; Entrez Gene: 5188 ; Ensembl: ENSG00000059691	Estado	Ref.	Cat.
NM_004564:exon3:c.T408G:p.F136L	Homocigosis	172	
NDUFA1, NADH:Ubiquinone Oxidoreductase Subunit A1 HGNC: 7683; Entrez Gene: 4694; Ensembl: ENSG00000125356	Estado	Ref.	Cat.
NM_004541:exon1:c.G94C:p.G32R	Heterocigosis	173	
NM_004541:exon1:c.A100G:p.K34E	Heterocigosis		3
NDUFA3, NADH:Ubiquinone Oxidoreductase Subunit A3 HGNC: 7686; Entrez Gene: 4696; Ensembl: ENSG00000170906	Estado	Ref.	Cat.
NM_004542:exon4:c.G173A:p.R58H	Heterocigosis		3
NDUFA8, NADH:Ubiquinone Oxidoreductase Subunit A8 HGNC: 7692; Entrez Gene: 4702; Ensembl: ENSG00000119421	Estado	Ref.	Cat.
NM_014222:exon3:c.G293T:p.R98L	Homocigosis		3
NDUFA10, NADH:Ubiquinone Oxidoreductase Subunit A10 HGNC: 7684; Entrez Gene: 4705; Ensembl: ENSG00000130414	Estado	Ref.	Cat.
NM_004544:exon1:c.G9C:p.L3F	Heterocigosis		3

NDUFA13, NADH:Ubiquinone Oxidoreductase Subunit A13 HGNC: 17194; Entrez Gene: 51079; Ensembl: ENSG00000186010	Estado	Ref.	Cat.
NM_015965:exon2:c.T107C:p.L36P	Heterocigosis		3
NM_015965:exon3:c.194delT:p.L65Pfs	Heterocigosis		3
ATP5F1A, ATP Synthase F1 Subunit Alpha HGNC: 823; Entrez Gene: 498; Ensembl: ENSG00000152234	Estado	Ref.	Cat.
NM_001001937:exon5:c.G418A:p.V140I	Heterocigosis		3
ECSIT, ECSIT Signalling Integrator HGNC: 29548; Entrez Gene: 51295; Ensembl: ENSG00000130159	Estado	Ref.	Cat.
NM_016581:exon8:c.A1055C:p.E352A	Homocigosis		3
FASTKD1, FAST Kinase Domain-Containing Protein 1, Mitochondrial HGNC: 26150; Entrez Gene: 79675; Ensembl: ENSG00000138399	Estado	Ref.	Cat.
NM_024622:exon13:c.C2303G:p.S768*	Heterocigosis		3
NDUFS1, NADH:Ubiquinone Oxidoreductase Core Subunit S1 HGNC: 7707; Entrez Gene: 4719; Ensembl: ENSG00000232228	Estado	Ref.	Cat.
NM_005006:exon18:c.C2029T:p.Q677*	Heterocigosis		5
NM_005006:exon8:c.C691G:p.L231V	Homocigosis		4
NDUFS4, NADH:Ubiquinone Oxidoreductase Subunit S4 HGNC: 7711; Entrez Gene: 4724 ; Ensembl: ENSG00000164258	Estado	Ref.	Cat.
NM_002495:exon3:c.350+5G>A	Homocigosis		3
NDUFS7, NADH:Ubiquinone Oxidoreductase Core Subunit S7 HGNC: 7714; Entrez Gene: 374291; Ensembl: ENSG00000115286	Estado	Ref.	Cat.
NM_024407:exon6:c.C425T:p.P142L	Heterocigosis		3
NDUFS8, NADH:Ubiquinone Oxidoreductase Core Subunit S8 HGNC: 7715; Entrez Gene: 4728; Ensembl: ENSG00000110717	Estado	Ref.	Cat.
NM_002496:exon6:c.G436A:p.D146N	Heterocigosis		4
COX10, Cytochrome C Oxidase Assembly Homolog 10 HGNC: 2260; Entrez Gene: 1352; Ensembl: ENSG00000006695	Estado	Ref.	Cat.
NM_001303:exon5:c.C637T:p.P213S	Heterocigosis		3
COX7A1, Cytochrome C Oxidase Subunit 7A1 HGNC: 2287; Entrez Gene: 1346; Ensembl: ENSG00000161281	Estado	Ref.	Cat.
NM_001864:exon1:c.15+1delG	Heterocigosis		4

Tabla 34. Mutaciones descritas ClinVar (08/2019) Panel OXPPOS (Genes con hallazgos)

	<i>ACAD9</i>	<i>SURF1</i>	<i>BCS1L</i>	<i>LRPPRC</i>	<i>FOXRED1</i>	<i>SDHA</i>	<i>NDUFV1</i>	<i>NDUFV2</i>
ClinVar Cat. 5	15 21%	68 61%	30 31%	22 11%	61 56%	148 23%	18 26%	77 78%
ClinVar Cat. 4	12 16%	23 21%	29 30%	50 26%	10 9%	23 4%	8 11%	3 3%
ClinVar Cat. 3	46 63%	20 18%	37 39%	120 63%	37 34%	469 73%	44 63%	19 19%
ClinVar Total	73	111	96	192	108	640	70	99
ORF (pb)	1866	903	1260	4185	1461	1752	1395	750
Nº mut/pb	0,80	7,53	2,38	0,53	4,18	8,45	1,29	10,27

	<i>NDUFV3</i>	<i>GATB</i>	<i>NDUFA1</i>	<i>NDUFA3</i>	<i>NDUFA8</i>	<i>NDUFA10</i>	<i>NDUFA13</i>	<i>ATP5F1A</i>
ClinVar Cat. 5	66 92%	21 95%	146 96%	18 95%	25 93%	73 55%	11 85%	40 89%
ClinVar Cat. 4	1 1%	1 5%	1 1%	0 0%	1 4%	6 5%	0 0%	1 2%
ClinVar Cat. 3	5 7%	0 0%	5 3%	1 5%	1 4%	53 40%	2 15%	4 9%
ClinVar Total	72	22	152	19	27	132	13	45
ORF (pb)	327	1674	213	255	387	1068	435	771
Nº mut/pb	20,18	1,25	68,54	7,06	6,46	6,84	2,53	5,19

	<i>ECSIT</i>	<i>FASTKD1</i>	<i>NDUFS1</i>	<i>NDUFS4</i>	<i>NDUFS7</i>	<i>NDUFS8</i>	<i>COX10</i>	<i>COX7A1</i>
ClinVar Cat. 5	10 77%	12 71%	26 35%	17 55%	18 50%	9 26%	67 60%	8 73%
ClinVar Cat. 4	1 8%	1 6%	14 19%	2 6%	4 11%	10 29%	6 5%	0 0%
ClinVar Cat. 3	2 15%	4 24%	35 47%	12 39%	14 39%	16 46%	38 34%	3 27%
ClinVar Total	13	17	75	31	36	35	111	11
ORF (pb)	1545	2544	2184	528	642	633	1332	240
Nº mut/pb	0,65	0,47	1,19	3,22	2,80	1,42	5,03	3,33

Del grupo de genes con hallazgos en esta serie presentan un mayor número de mutaciones patogénicas (Cat. 5) descritas en ClinVar dos genes que codifican unidades estructurales de la cadena respiratoria: *SDHA* y *NDUFA1*; aunque si tenemos en cuenta el tamaño de los exones de estos genes, *NDUFA1* presenta una concentración de mutaciones mucho más elevada dado su pequeño tamaño. En cambio, en la serie analizada, quienes que han presentado un mayor número de mutaciones patogénicas o probablemente patogénicas han sido genes implicados en el ensamblaje de complejos cadena respiratoria mitocondrial: *ACAD9*, implicado en el

metabolismo de ácidos grasos y en el ensamblaje del Complejo I y BCS1L, implicado en el ensamblaje el Complejo III.

13. Estudio depleción en niños menores de 4 años (hallazgos)

Se ha realizado el panel de genes de mantenimiento en una cohorte de 32 niños menores de 4 años que presentaban un estudio previo de depleción positivo y no tenían filiada ninguna causa genética conocida.

Se han encontrado mutaciones causales sólo en dos casos en el grupo de 3 pacientes que presentaban los valores de depleción más bajos (<10%):

- Mutación en homocigosis en el exon 10 del gen *TK2* (NM_004614, c.C760T, p.R254*).
- Mutación en heterocigosis en el exón 7 del gen *TK2* (NM_004614, c.503delT, p.I168fs y c.C388T, p.R130W).

En el grupo de 19 pacientes con niveles de depleción por debajo del 30% se han encontrado 3 mutaciones probablemente patogénicas:

- Mutación en heterocigosis en el exon 13 de la *POLG* (NM_001126131, c.A2207G, p.N736S)
- Dos mutaciones que podrían conformar un heterocigoto compuesto en los exones 2 y 4 del gen *DGUOK* (NM_080916, c.C211G, p.P71A y c.A509G, p.Q170R), aunque podría tratarse de polimorfismos ya que aparecen en otros pacientes.

En otros 10 pacientes con niveles de depleción entre el 30-40% no se ha priorizado ninguna variante potencialmente patogénica en los genes estudiados (Figura 46).

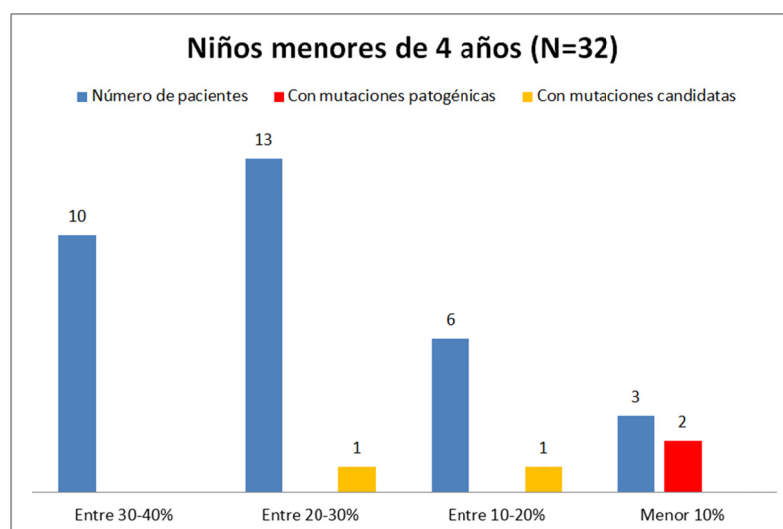


Figura 46. Hallazgos panel genes mantenimiento en pacientes con depleción mitocondrial

14. Estudio deleciones múltiples

Se ha realizado un estudio comparativo entre dos técnicas empleadas usualmente para la detección de deleciones múltiples en el mtDNA: *Southern Blot* y *Long Range PCR (LR-PCR)*.

En la evaluación del grado de concordancia de resultados positivos para deleciones múltiples en una población (N=40) con edad media de 54 años [rango 18-79] se han encontrado variantes patogénicas o probablemente patogénicas en el 48% de los casos, lo que supone una alta tasa de positivos (Figura 47).

Los genes localizados como causales que presentaban variantes patogénicas o probablemente patogénicas fueron: *TK2*, *TWINKLE*, *POLG* y *RRM2B*.

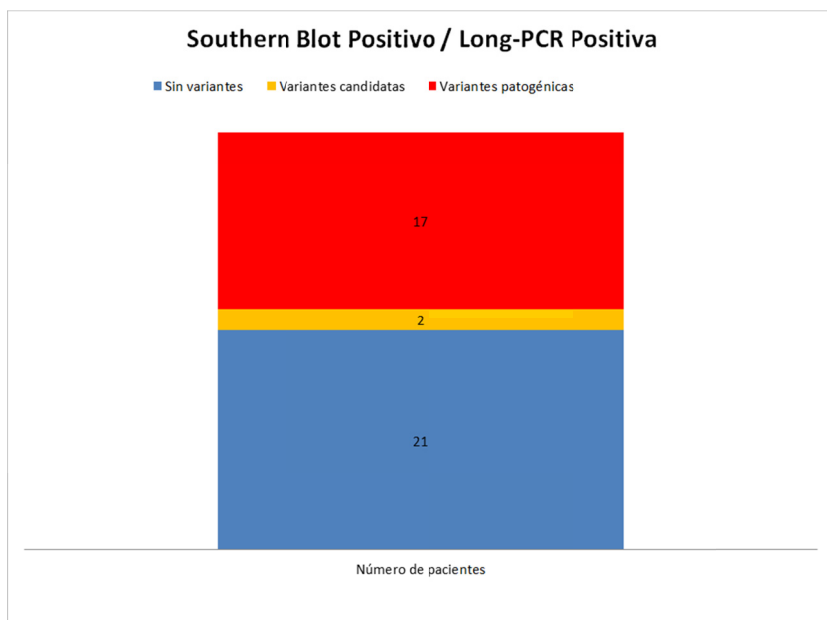


Figura 47. Resultados Southern Blot + / LR-PCR +

En una cohorte (N=31) con resultado de *Southern Blot* negativo, pero LR-PCR positiva se han encontrado variantes patogénicas o probablemente patogénicas en el 20% de los casos. La tasa de positivos es considerablemente más baja que en el primer caso, pero mediante LR-PCR se detectan algunos positivos que han pasado desapercibidos por *Southern Blot* (Figura 48).

En este caso los genes hallados con variantes patogénicas o probablemente patogénicas han sido: *POLG*, *OPA1* y *SUCLG1*.

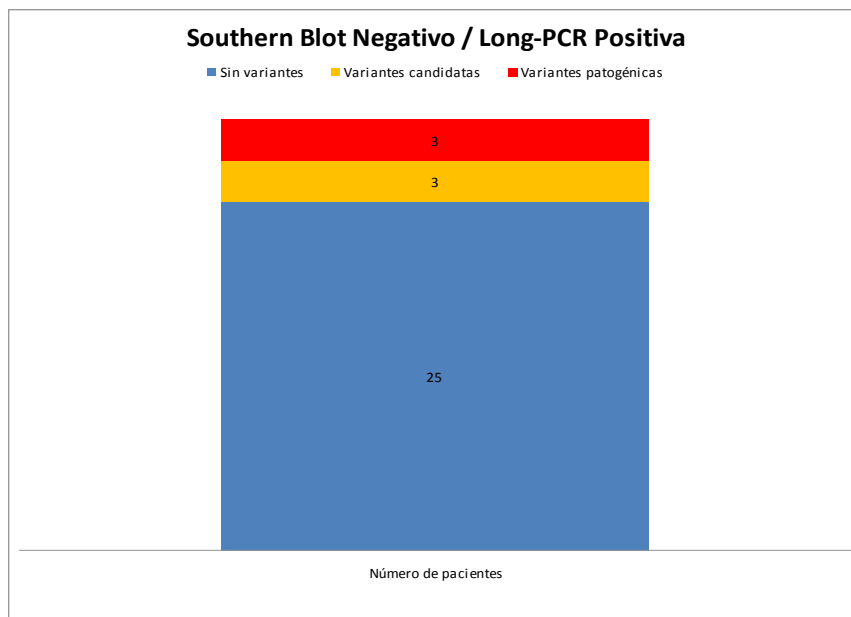


Figura 48. Resultados Southern Blot - / LR-PCR +

15. Estudio funcional tisular de la mutación p.T108M en el gen *TK2*.

Las mutaciones en la Timidina quinasa 2 (*TK2*, RefSeq NP_004605.4) se han asociado frecuentemente con depleción del mtDNA y miopatía fatal de debut infantil o con deleciones múltiples del mtDNA en adultos con miopatía u oftalmoplegía externa progresiva (CPEO). Se han reportado más de 60 pacientes con mutaciones autosómicas recesivas¹⁷⁴. La mutación p.T108M (c.C323T; NM_004614) ha sido descrita en 15 pacientes de 12 familias diferentes: 4 pacientes presentaban una miopatía leve con deleciones múltiples sin depleción en músculo esquelético¹⁵¹; 3 pacientes mostraban una miopatía de lenta progresión, con deleciones múltiples y depleción del mtDNA^{146,175}; 7 pacientes eran casos pediátricos severos que fallecieron¹⁷⁶⁻¹⁸⁰ y un último caso pediátrico seguía con vida a los 12 años¹⁸⁰. Hasta la fecha solo se ha realizado un estudio a nivel molecular de la función mitocondrial en diversos tejidos (músculo esquelético, cerebro e hígado) en 7 niños finlandeses con depleción del mtDNA¹⁸¹, que concluye que deben considerarse las mutaciones en *TK2* también en las encefalomiopatías progresivas infantiles.

En el presente trabajo se han llevado a cabo diferentes estudios en tejidos donados por un paciente y un control. Se ha estudiado la integridad del mtDNA, el sistema OXPHOS, la función glicolítica, la autofagia y el proteasoma. Todos los procedimientos empleados han sido aprobados por el Comité de Ética del Hospital y se obtuvo el consentimiento informado apropiado de los dos sujetos implicados en el estudio. Se obtuvieron diversos tejido post-mortem (músculo, cerebro,

hígado, diafragma, intestino y corazón) del paciente y se emplearon, como control en paralelo, tejidos equivalentes de otro sujeto de la misma edad fallecido a causa de un tumor sólido y sin evidencias de patología neuromuscular, ni metabólica hereditaria.

El paciente procede de padres consanguíneos (primos) y desarrolló un rendimiento físico normal hasta los 18 años. A partir de entonces comenzó a padecer dificultades progresivas para subir escaleras, correr, levantarse de la silla, y una disfagia de curso progresivo, insuficiencia ventilatoria restrictiva progresiva desde los 24 años sin presentar afectación cardíaca, ni de otros órganos. Éste último es un hecho poco frecuente, ya que lo descrito hasta la fecha era que la insuficiencia respiratoria en adultos estaba asociada a un debut temprano en niños⁹¹. Todo ello derivó en la necesidad de ventilación mecánica desde los 30 años, con ingresos frecuentes por infecciones respiratorias y aumento progresivo de las necesidades de ventilación. Con alimentación por sonda de gastrostomía desde los 40 años. Finalmente falleció a los 42 años debido a una insuficiencia respiratoria aguda, desnutrición y bradicardia.

A los 30 años se le realiza una biopsia muscular en la que en los estudios histológicos se observan numerosas fibras rojo rasgadas (*ragged-red fibers*, RRF) y fibras COX negativas (Figura 49). En los estudios bioquímicos destaca un déficit combinado de la actividad de la cadena respiratoria, en concreto de los complejos CI, CIII y CIV. En suero presenta una CPK de 1400UI.

Edad inicio	Exitus	Fenotipo	Histoquímica Músculo	Cadena Respiratoria	Bioquímica	Mutación
18 a	42 a	Debilidad muscular, disfagia y atrofia progresivas Afectación bulbar e insuficiencia respiratoria severa	RRF; COX -; NEC	CI; CIII; CIV	CPK 1400 UI	NM_004614, exon5 c.C323T, p.T108M Homocigosis

Tabla 35. Resumen paciente estudio TK2

a: años; RRF: *ragged-red fibers*; COX -: fibras musculares negativas para tinción de Citocromo C; NEC: células necróticas; CPK: creatina-fosfoquinasa (EC 2.7.3.2).

Los estudios histomorfológicos en el músculo esquelético muestran diferentes alteraciones tales como fibras *ragged-red*, células necróticas, células con incrementos de SDH, así como fibras COX negativas (Figura 49).

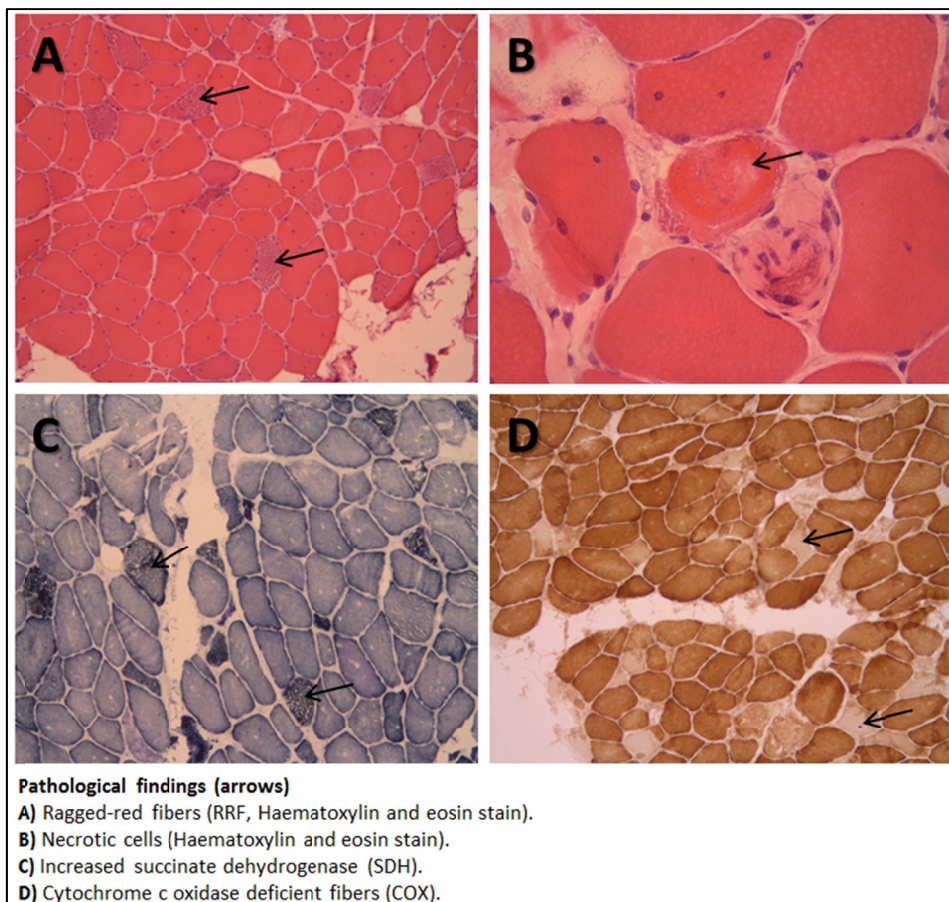


Figura 49. Cambios histológicos en músculo esquelético. Imágenes cedidas por el Dr. Aurelio Hernández Laín, S. Anatomía Patológica, Hospital 12 de Octubre, Madrid.

El análisis del número de copias de mtDNA (Método pág.39) muestra una reducción generalizada en todos los tejidos, en un rango entre 12% y 64% respecto a los tejidos control. Esta reducción del número de copias del mtDNA es muy leve en el caso de corazón y cerebro y especialmente alta en el caso del diafragma (Figura 50).

Los estudios de integridad del mtDNA (*Southern blot* y *Long range PCR*) muestran la presencia de una deleción única tan sólo en el músculo esquelético (Figura 51) y no parece existir ninguna deleción en los demás tejidos analizados.

Estos hallazgos indican que en algunas miopatías con mutaciones en *TK2* el número de copias del mtDNA no sólo se puede encontrar reducido en el músculo esquelético sino que puede estar también disminuido en otros tejidos. En cambio, la presencia de deleciones parece restringirse a tejido muscular. Por otro lado, este caso remarca que la depleción del mtDNA no estaría restringida sólo a los casos pediátricos más severos tal y como se pensaba.

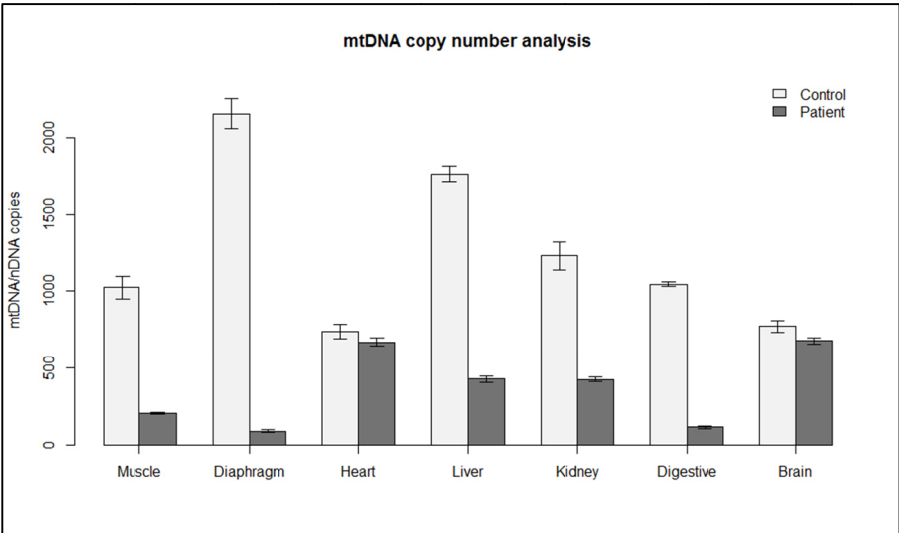


Figura 50. Análisis del número de copias del mtDNA

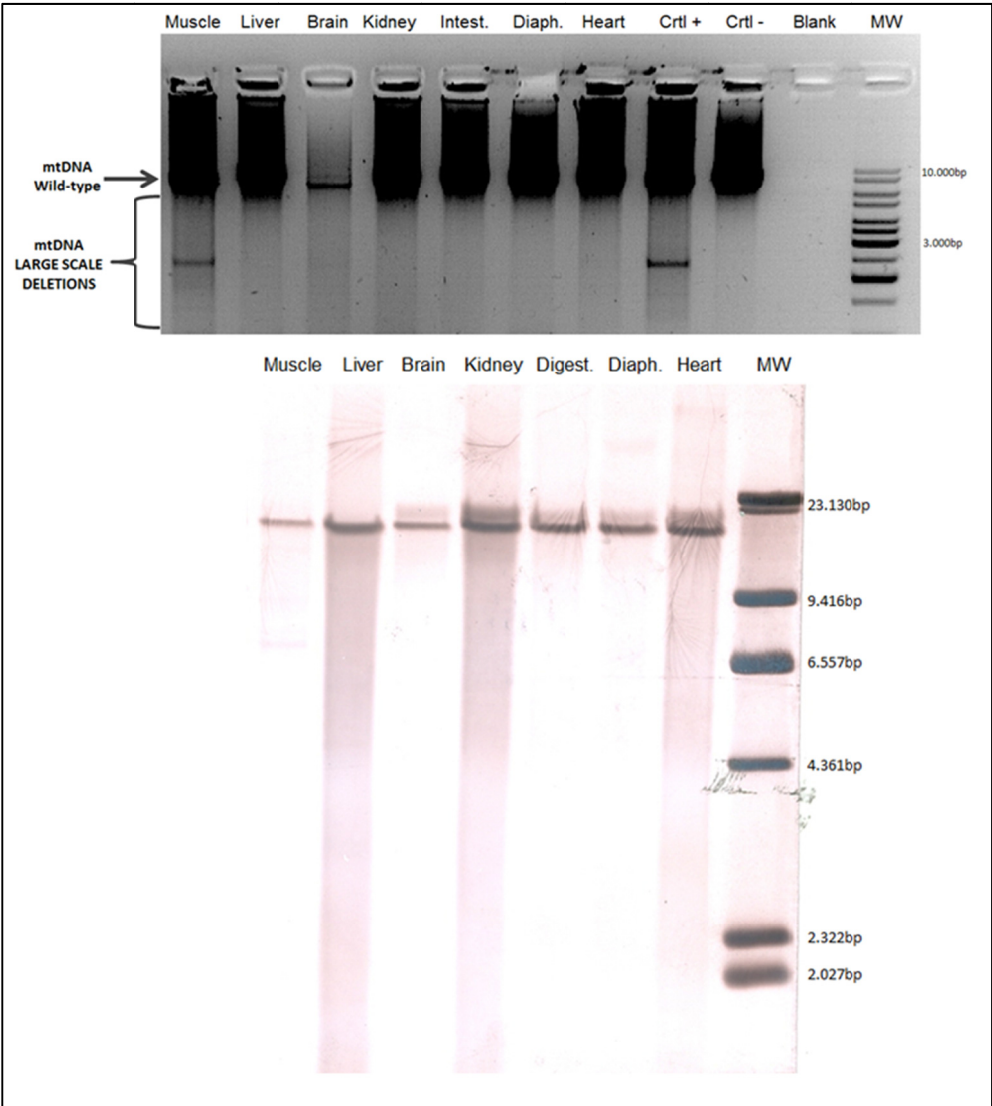


Figura 51. Long range PCR (arriba) y Southern Blot (abajo)

En cuanto a la función OXPHOS, analizada por *Western blot*, se puede apreciar especialmente en el diafragma del paciente una drástica reducción de SDHA, Complejo V alpha subunit, COX1, ATP8 y GAPDH (Figura 52 y Figura 53). En el resto de tejidos no se aprecian cambios o incluso existen incrementos en el paciente respecto al control (Ej. Muscle/Kidney ATP8, Kidney COX1/SDHA).

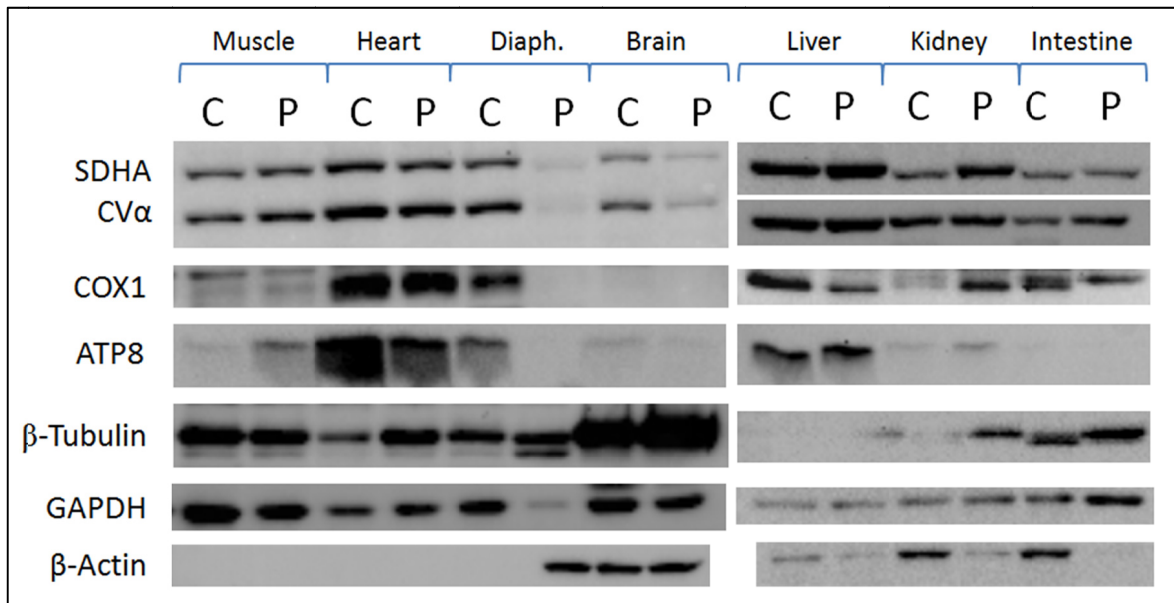


Figura 52. *Western Blot* subunidades OXPHOS

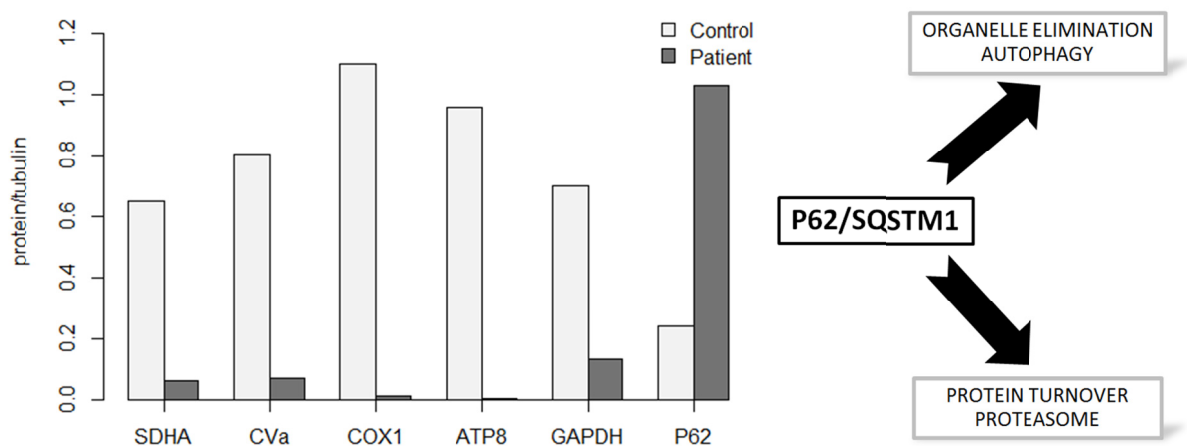


Figura 53. Densitometría del diafragma (OXPHOS y P62)

En el estudio de las vías de degradación de proteínas mediante el receptor de proteínas ubiquinadas p62/SQSTM1 (sequestosome 1)¹⁸² se observa un patrón similar entre paciente y control en todos los tejidos con la excepción del diafragma en el que se encuentra claramente incrementado. Asimismo, el patrón de ubiquinación también se encuentra especialmente

alterado en el diafragma (Figura 54). Por otra parte, el perfil global de proteínas de dicho tejido es marcadamente diferente respecto al control (Figura 55) con la presencia de una gran cantidad de albúmina (Figura 55, banda 1) y la desaparición de importantes proteínas implicadas en la glucólisis como la fructosa-bisfosfato aldolasa (Figura 55, banda 2).

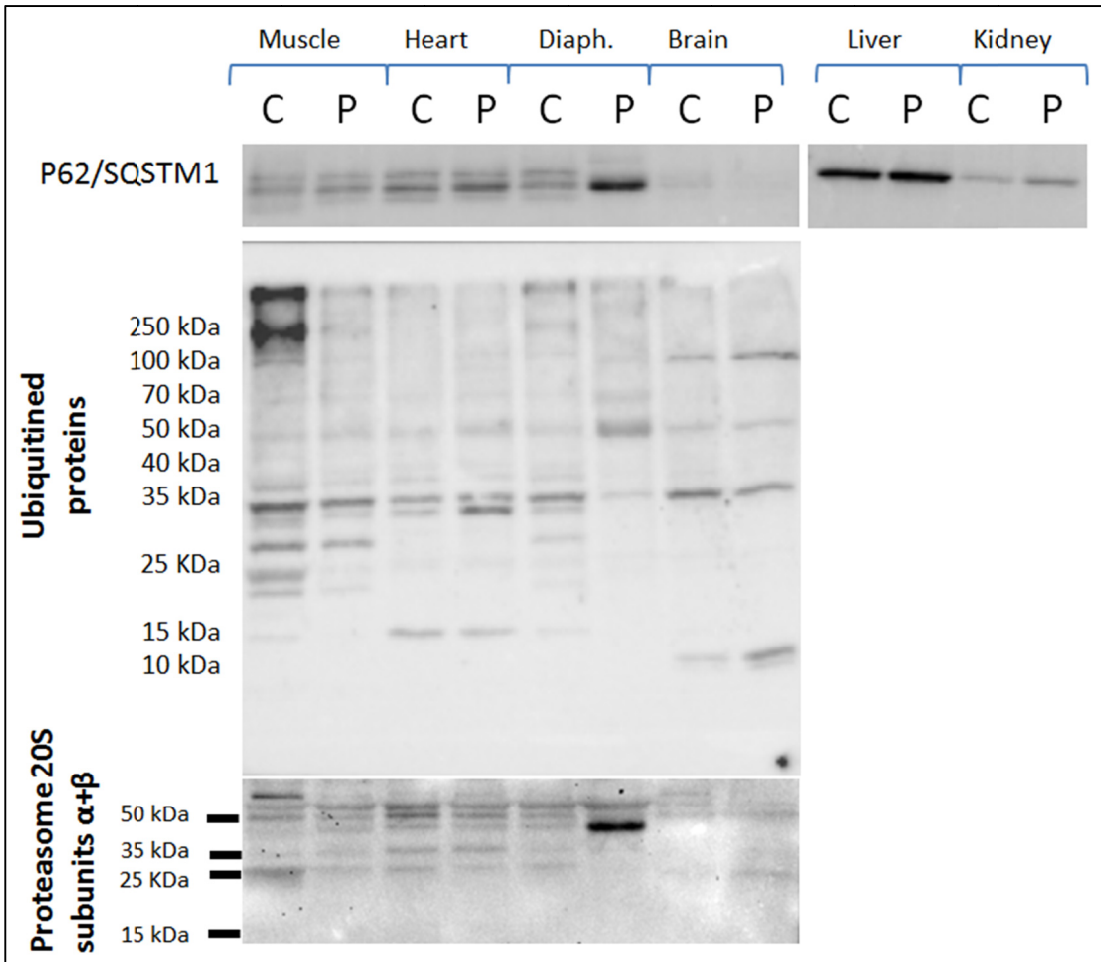


Figura 54. Análisis degradación proteica

El diafragma es el tejido más afectado y en este estadio final muestra una incapacidad casi absoluta para la generación de ATP vía ciclo de Krebs o glucólisis ya que desaparecen por completo proteínas de gran relevancia como la GAPDH (Figura 52) o la fructosa-bisfosfato aldolasa (Figura 55). Además parecen estar incrementadas las vías de degradación de proteínas a través de proteasoma o autofagia y posiblemente exista una desregulación de la transcripción nuclear que está originando proteínas que en condiciones normales no aparecen en el diafragma como la β -actina (Figura 52).

Se ha descrito¹⁸³ que en pacientes con ventilación mecánica prolongada y con una inactividad total del diafragma existen tres vías que juegan un papel relevante en el desarrollo de

una atrofia del mismo: *ubiquitin-proteasome pathway* (UPP), *marked decreases in myosin heavy chain* (MyHC), *and atrophic protein kinase B–forkhead Box O signaling* (AKT-FOXO). En este caso al tratarse de una etapa terminal, con varios años de uso de ventilación mecánica, sólo se observa un incremento en la ubiquitinación de proteínas de alto peso molecular (bandas 50-70kDa), posiblemente porque las más pequeñas han sido ya ampliamente degradadas, y un claro incremento de p62/SQSTM1 (Figura 54) que indicaría la activación de la proteólisis a través de los sistemas principales (lisosoma y proteasoma) que contribuiría de una forma crucial a la atrofia del tejido y a la generación del patrón extremo que se observa únicamente en el diafragma. En el resto de tejidos las rutas de degradación son similares entre control y paciente.

Con estas evidencias se podría especular que, al igual que ocurre con la inactividad del diafragma debida a la ventilación mecánica, se podrían activar también los sistemas proteolíticos como consecuencia de los fallos en la síntesis de proteínas mitocondriales debidos a la mutación en *TK2*, con la consiguiente depleción energética parcial, el origen de manera paulatina de la atrofia del diafragma y el consecuente fenotipo respiratorio severo observado en el paciente.

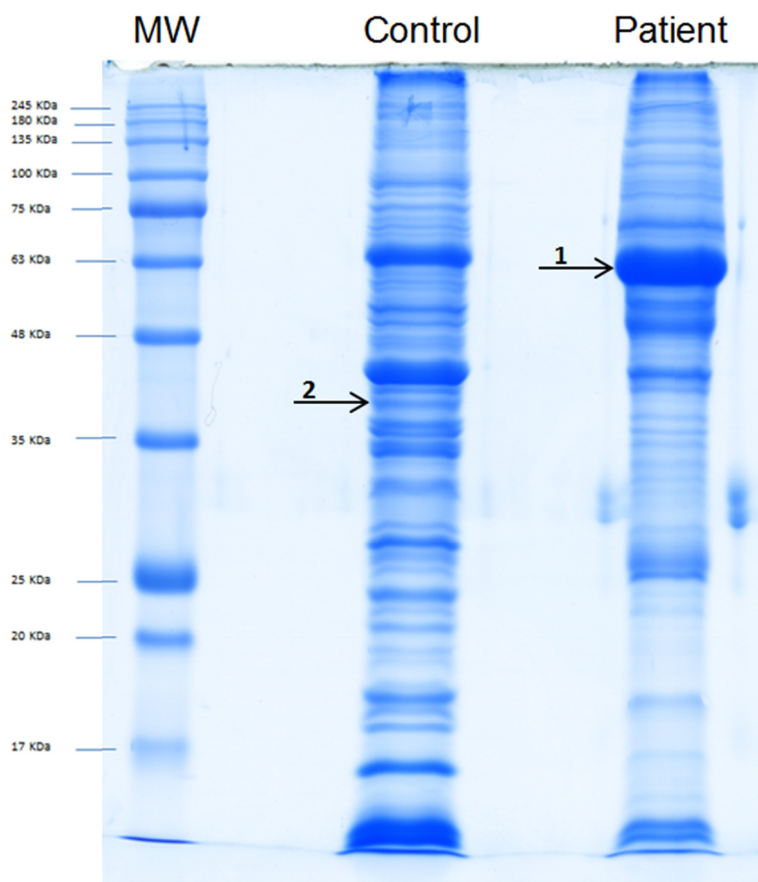


Figura 55. Electroforesis en una dimensión de diafragma (Ver pág. 28)

Identificar nuevas mutaciones en genes asociados a enfermedad OXPPOS mediante resecuenciación de exoma en familias candidatas.

Se ha secuenciado el exoma completo de 8 pacientes pediátricos (rango 3 días - 8 años) que presentaban cuadros clínicos característicos de hipotonía, cardiomiopatía hipertrófica o retraso psicomotor y una serie de criterios bioquímicos concretos como déficits combinados de la cadena respiratoria en músculo, acidosis láctica o ausencia de alteraciones en el mtDNA (mutaciones puntuales, deleciones o depleción).

La secuenciación se llevó a cabo en Centro Nacional de Análisis Genómicos (CNAG) en un equipo HiSeq (Illumina).

Adicionalmente, en otra serie de pacientes, también con trastornos mitocondriales definidos y sin causa genética conocida, se ha realizado la secuenciación del exoma en el Instituto de Investigación Biomédica i+12 en un equipo Nextseq 500 (Illumina).

En este apartado, para la detección y análisis de CNVs se ha realizado una integración y modificación de diversos scripts de R ya publicados que emplean *ExomeDepth* (Tabla 37 y Tabla 38), en un *pipeline* propio (Tabla 36 y Tabla 39) que comienza con la lectura del directorio de ficheros .bam y .bai y finaliza en ficheros .xlsx con los CNVs detectados y anotados con AnnotSV.

La preparación inicial de los ficheros se ha realizado siguiendo las recomendaciones de GATK: tras el mapeo frente al genoma de referencia (bwa), se marcaron duplicados (Picard) y se recalibraron los ficheros bam antes de su procesado por *ExomeDepth*.

Este análisis genera una media de 355 posibles CNVs por exoma (Mediana= 308, Máximo= 872, Mínimo= 141) que deberán ser convenientemente filtrados y analizados.

Tabla 36. Script general análisis CNVs

```
#!/usr/bin/perl -w #
# SCRIPT: CNV_analisis.pl #
# Entrada: Deben indicarse los directorios de entrada (ficheros bam y
bai), de salida y las referencias (fichero .bed y .fasta #
# Se necesitan los scripts: annotsv.pl, ExomeDepth-referencias.R y
ExomeDepth-Analisis.R, el paquete de R ExomeDepth y AnnotSV.
# Salida: Resultados análisis CNVs con Exomedeph y anotados con AnnotSV #
# Ejemplo de ejecución: ./CNV_analisis.pl ~/ExomeDepth/bams/
~/ExomeDepth/Resultados ~/ExomeDepth/refs/UCSC_RefSeqgenes.bed
~/ExomeDepth/refs/ucsc.hg19.fasta
#####

# Activar la declaracion estricta de variables
use strict;
use Term::ANSIColor;
```

```
## Paso 1. Adquirir el nombre del directorio
my ($directorio_in,$directorio_out,$bedfile,$fastafile,$n_argumentos);
$n_argumentos = scalar(@ARGV);
($n_argumentos == 4) or
print_error("Error argumentos. Indica directorios de entrada (.bam y
.bai) y de salida, ruta de bed file y ruta de fasta file");
($directorio_in,$directorio_out,$bedfile,$fastafile) = @ARGV;
print_ok("Lectura de directorios");
## Paso 2. Abrir el directorio y crear nuevos directorios de trabajo
my ($fichero,$fichero_in,$fichero_out);
my ($nombre,$comando,$rtcd);
system ("mkdir -p $directorio_out 2> /dev/null") == 0
    or die "no ha sido posible crear $directorio_out. abortando...\n";
print_ok("Creado directorio de salida $directorio_out");
my $directorio_exomedeph_annotados = "$directorio_out/exomedeph_annotados";
my $directorio_exomedeph_resultados =
"$directorio_out/exomedeph_resultados";
system ("mkdir -p $directorio_exomedeph_annotados 2> /dev/null") == 0
    or die "no ha sido posible crear $directorio_exomedeph_annotados.
abortando...\n";
system ("mkdir -p $directorio_exomedeph_resultados 2> /dev/null") == 0
    or die "no ha sido posible crear $directorio_exomedeph_resultados.
abortando...\n";
print_ok("Directorios creados y listos para ejecutar análisis CNVs");
opendir(DIR,$directorio_in);
print_ok("Acceso al directorio de BAMs: $directorio_in");
while ($fichero = readdir(DIR))
{
    ## Control sobre los directorios y ejecución de programas externos
    next unless ($fichero =~ m/\.bam$/);
    next if ($fichero =~ /\./);
## Paso 3. Crear listado con la ruta de los ficheros .bam (bamlist)
($nombre = $fichero) =~ s/\.bam//;
$fichero_in = "$directorio_in/$fichero";
$fichero_out = "$directorio_out/bamlist.txt";
$comando = "find $directorio_in -type f | grep -v 'bai' >
$fichero_out ; sed -i 's|^||g' $fichero_out";
$rtcd = system($comando);
}
## Paso 4. Ejecutar ExomeDeph-referencias.R
$fichero_in = "$fichero_out";
$comando = "Rscript ExomeDeph-referencias.R -b $fichero_in -f
$bedfile -r $fastafile -o $directorio_out -v > $directorio_out/log-
referencias.txt";
$rtcd = system($comando);
rt_ctrl($rtcd, "Creación de fichero referencia: counts.rda");
## Paso 5. Ejecutar ExomeDeph-analisis.R
$fichero_in = "$directorio_out/counts.rda";
$directorio_out = "$directorio_exomedeph_resultados";
$comando = "Rscript ExomeDeph-Analysis.R -c $fichero_in -f $bedfile
-o $directorio_out ";
$rtcd = system($comando);
rt_ctrl($rtcd, "Análisis de CNVs");
## Paso 6. Ejecutar anotación con AnnotSV
$directorio_in = "$directorio_out";
$directorio_out = "$directorio_exomedeph_annotados";
$comando = "./annotsv.pl $directorio_in $directorio_out";
$rtcd = system($comando);
```

```

    rt_ctrl($rtcd, "Anotación de CNVs");
# Cerrar el directorio
closedir(DIR);
print_ok("Análisis CNVs finalizado. Resultados en directorio:
$directorio_out");
exit(0);

```

Tabla 37. Script R para la creación de fichero de referencias

```

# Creado por Eric Minikel (https://github.com/ericminikel/ngs-scripts)
# Modificado y adaptado por Aitor Delmiro

start_time = Sys.time()
require(ExomeDepth)
suppressPackageStartupMessages(require(optparse))
suppressPackageStartupMessages(require(GenomicAlignments))
options(stringsAsFactors=FALSE) # crucial for handling BAM filenames as
strings
option_list = list(
  make_option(c("-b", "--bamlist"), action="store", default='',
             type='character', help="Path to list of BAMs"),
  make_option(c("-o", "--outdir"), action="store", default='./',
             type='character', help="Output directory [default
%default]"),
  make_option(c("-f", "--bedfile"), action="store", default='',
             type='character', help="Path to bed file"),
  make_option(c("-r", "--fastafile"), action="store", default='',
             type='character', help="Path to fasta file"),
  make_option(c("-v", "--verbose"), action="store_true", default=TRUE,
             help="Print verbose output [default %default]"),
  make_option(c("-q", "--quiet"), action="store_false", dest="verbose",
             help="Do not print verbose output (this is the default)")
)
opt = parse_args(OptionParser(option_list=option_list))
print(opt, file=stderr())
# Bed file
if (file.exists(opt$bedfile)) {
  exons.hg19.propios = read.table(opt$bedfile, header=TRUE)
} else {
  cat("You need to specify a valid bed file using -f.\n", file=stderr())
  cat(paste("The bed file you specified was
'", opt$bedfile, "'.", sep=' '), file=stderr())
  stop()
}
# Fasta file
if (file.exists(opt$fastafile)) {
  fasta.file = (opt$fastafile)
} else {
  cat("You need to specify a valid fasta file using -r.\n", file=stderr())
  cat(paste("The fasta file you specified was
'", opt$fastafile, "'.", sep=' '), file=stderr())
  stop()
}
# Read list of BAMs
if (file.exists(opt$bamlist)) {
  bams = read.table(opt$bamlist, header=FALSE)$V1
} else {
  cat("You need to specify a valid BAM list using -b.\n", file=stderr())

```

```

cat(paste("The filename you specified was
",opt$bamlist,".",sep=''),file=stderr())
stop()
}
if (file.exists(opt$outdir)) {
  setwd(opt$outdir)
} else {
  cat("You need to specify a valid output directory using -
o.\n",file=stderr())
  cat(paste("The directory you specified was
",opt$outdir,".",sep=''),file=stderr())
  stop()
}
if (opt$verbose) {
  cat(paste("Read BAM list from ",opt$bamlist,"\n",sep=''),file=stdout())
}
counts = getBamCounts(bed.frame = exons.hg19.propios, bam.files = bams,
min.mapq = 20,
                      include.chr = FALSE, referenceFasta = fasta.file)
if (opt$verbose) {
  cat(paste("Calculated counts\n",sep=''),file=stdout())
}
save(counts,file="counts.rda")
if (opt$verbose) {
  cat(paste("Generado fichero de referencias counts.rda en
",getwd(),"\n",sep=''),file=stdout())
}
duration = format(Sys.time() - start_time)
if (opt$verbose) {
  cat(paste("Proceso realizado en ",duration,".\n",sep=''),file=stdout())
}

```

Tabla 38. Script R para análisis de CNVs

```

# Creado por Eric Minikel (https://github.com/ericminikel/ngs-scripts)
# Modificado y adaptado por Aitor Delmiro

start_time = Sys.time()
require(ExomeDepth)
suppressPackageStartupMessages(require(GenomicAlignments))
suppressPackageStartupMessages(require(optparse))
options(stringsAsFactors=TRUE) # crucial for handling BAM filenames as
strings
option_list = list(
  make_option(c("-c", "--countsfile"), action="store", default='',
              type='character', help="Full path to counts file"),
  make_option(c("-o", "--outdir"), action="store", default='./',
              type='character', help="Output directory for CNV calls
[default %default]"),
  make_option(c("-f", "--bedfile"), action="store", default='',
              type='character', help="Path to bed file"),
  make_option(c("-g", "--genome"), action="store", default='', #
recommend using long name --genome; otherwise R thinks -g is specifying
the gui.
              type='character', help="PLINK .genome IBD file"),
  make_option(c("-e", "--exclude"), action="store", default='',
              type='character', help="txt file listing samples to exclude
from reference"),

```

```

make_option(c("-t", "--threshold"), action="store", default=.03125,
            type='numeric', help="Maximum IBD to still use as reference
[default %default]"),
make_option(c("-s", "--callsamples"), action="store", default='',
            type='character', help="Path to list of samples to call
CNVs on [default all]"),
make_option(c("-v", "--verbose"), action="store_true", default=TRUE,
            help="Print verbose output [default %default]"),
make_option(c("-q", "--quiet"), action="store_false", dest="verbose",
            help="Do not print verbose output (this is the default)")
)
opt = parse_args(OptionParser(option_list=option_list))
if (opt$verbose) {
  cat(paste("Execution started at
",start_time,"\n",sep=""),file=stdout())
}
# Bed file
if (file.exists(opt$bedfile)) {
  exons.hg19.propios = read.table(opt$bedfile,header=TRUE)
} else {
  cat("You need to specify a valid bed file using -f.\n",file=stderr())
  cat(paste("The bed file you specified was
",opt$bedfile,"'.'",sep=' '),file=stderr())
  stop()
}
if (opt$verbose) {
  cat("Running getExomeDepthCNVs.r with the following
parameters:\n",file=stdout())
  print(opt,file=stdout())
}
if (opt$verbose) {
  cat("Loading ExomeDepth package...\n",file=stdout())
}
# try to load the ExomeDepth package in a safe and informative way
tryCatch ({
  # try
  suppressPackageStartupMessages(require(ExomeDepth))
  if (opt$verbose) {
    cat("Loaded successfully.\n",file=stdout())
  }
}, error = function(cond) {
  cat("Experienced an error trying to load the ExomeDepth
package.\n",file=stderr())
  print(cond, file=stderr())
  stop()
})
if(opt$verbose) {
  cat("Casting count data to a data frame...\n",file=stdout())
}
load(opt$countfile)
countdf = as.data.frame(counts)
if(opt$verbose) {
  cat("Column names in the count data frame:\n",file=stdout())
  print(colnames(countdf),file=stdout())
}
if(opt$verbose) {
  cat("Casting count data to a matrix...\n",file=stdout())
}
countmat = as.matrix(countdf[,7:dim(countdf)[2]]) # remove cols 1-5

```

```
metadata
if(opt$verbose) {
  cat("OK, here's the head of the matrix...\n",file=stdout())
  #print(head(countmat),file=stdout())
}
if(opt$callsamples=='') {
  # default mode: all samples (colnames of count matrix)
  # will have CNVs called on them
  call_indices = 1:dim(countmat)[2]
  if(opt$verbose) {
    cat("Default mode - calling CNVs on all samples:\n",file=stdout())
    print(colnames(countmat)[call_indices],file=stdout())
  }
} else {
  # user has specified a file listing the samples on which to call CNVs
  if(file.exists(opt$callsamples)) {
    cat("Reading list of samples to call CNVs on from:\n",file=stdout())
    cat(paste(opt$callsamples,"\n",sep=""),file=stdout())
    call_sample_names = read.table(opt$callsamples,header=FALSE)$V1
    call_names = make.names(call_sample_names)
    cat("Sample list read successfully.\n",file=stdout())
    # figure out if they specified samples that aren't in the count data
    missing_samples = call_sample_names[!call_sample_names %in%
colnames(countmat)]
    if(length(missing_samples) > 0) {
      cat("You specified some samples that aren't in the counts
file:\n",file=stdout())
      print(missing_samples,file=stdout())
    }
    # get indices of ones they specified that are actually in the counts
file
    call_indices = match(call_sample_names,colnames(countmat))
    call_indices = call_indices[!is.na(call_indices)] # remove NAs
    if(length(call_indices) > 0) {
      cat("Will be calling CNVs on the following
samples:\n",file=stdout())
      print(colnames(countmat)[call_indices],file=stdout())
    } else {
      cat("***Didn't find any specified sample names in the counts
file.\n",file=stdout())
    }
  } else {
    cat("Couldn't find the callsamples list you specified with -s:
\n",file=stderr())
    cat(paste(opt$callsamples,"\n",file=stderr())
stop()
  }
}
stopifnot(exists("call_indices"))
stopifnot(length(call_indices) > 0)
call_names = colnames(countmat)[call_indices]
# Now try to read .genome file
if (file.exists(opt$genome)) {
  if (opt$verbose) {
    cat(paste("Reading .genome IBD file from: ",opt$genome,
"\n\nin order to exclude relateds from reference sets.\n\n",
sep=""),file=stdout())
  }
  genome = read.table(opt$genome,header=TRUE)
```

```

exclude = genome[genome$PI_HAT > opt$threshold, c("IID1", "IID2")]
exclude_proportion = as.numeric(dim(exclude)[1]/dim(genome)[1])
if (opt$verbose) {
  cat(paste("Based on your threshold of ", opt$threshold*100, "% IBD, ",
            formatC(exclude_proportion*100, digits=2), "% of possible\n",
            "combinations of people will be excluded from the
reference\n\n",
            sep=""), file=stdout())
}
# make the sample names from PLINK match the R-ified BAM names
exclude$IID1 = paste(make.names(exclude$IID1), ".bam", sep="")
exclude$IID2 = paste(make.names(exclude$IID2), ".bam", sep="")
# match to the names of samples to call on
genome_iids = unique(c(genome$IID1, genome$IID2))
missing_names = call_names[!(call_names %in% genome_iids)]
if (length(missing_names) > 0) {
  cat("Warning: The following samples that you want to
call\n", file=stderr())
  cat("CNVs on are not in your .genome IBD file:\n", file=stderr())
  print(missing_names, file=stderr())
}
# create a named list where the name is the sample name
# and the value is a vector of other samples to exclude
exclude_list = list()
for (individual in call_names) {
  exclude_list[[individual]] = unique(
    c(exclude$IID1[exclude$IID2==individual],
      exclude$IID2[exclude$IID1==individual]))
}
} else {
  if (opt$genome=='') {
    if (opt$verbose) {
      cat("You didn't specify a .genome IBD file, so we're just
going\n", file=stdout())
      cat("to assume that no individuals are related.\n", file=stdout())
    }
    exclude_list = list()
    for (individual in call_names) {
      exclude_list[[individual]] = character(0)
    }
  } else {
    cat("***Couldn't find the .genome file you
specified:\n", file=stderr())
    cat(opt$genome, file=stderr())
    cat("\n\n", file=stderr())
    stop()
  }
}
if (file.exists(opt$exclude)) {
  if (opt$verbose) {
    cat(paste("Loading list to exclude from ", opt$exclude, "
...\n", sep=""), file=stdout())
  }
  exclude_all = read.table(opt$exclude)$V1
  exclude_all = paste(make.names(exclude_all), ".bam", sep="")
  if (opt$verbose) {
    cat(paste("Loaded.\n", sep=""), file=stdout())
  }
  if (opt$verbose) {

```

```

    cat("The following samples will be excluded from the reference for
all CNV calls:\n", file=stdout())
    print(exclude_all, file=stdout())
  }
  for (individual in call_names) {
    exclude_list[[individual]] = c(exclude_list[[individual]],
exclude_all)
  }
} else if (opt$exclude == '') {
  if(opt$verbose) {
    cat(paste("No exclude all list (-e) detected,
continuing...\n", sep=""), file=stdout())
  }
  exclude_all = character(0) # just make it an empty vector
} else {
  cat("***Exclude all list file not found. You specified this
path:\n", file=stderr())
  cat(paste(opt$exclude, "\n", sep=""), file=stderr())
}
# now we are done reading input, so switch to output directory.
if (file.exists(opt$outdir)) {
  setwd(opt$outdir)
  if(opt$verbose) {
    cat(paste("Will write CNV calls to:
", opt$outdir, "\n", sep=""), file=stdout())
  }
} else {
  cat(paste("You specified this directory:
", opt$outdir, "\n", sep=""), file=stdout())
  cat(paste("And it doesn't exist.\n", sep=""), file=stdout())
  stop()
}
##Inicio Análisis
for (i in call_indices) {
  sample_name = colnames(countmat)[i]
  if(opt$verbose) {
    cat(paste("Now about to call CNVs on
", sample_name, "...", sep=""), file=stdout())
    cat("Excluding the following samples from
reference:\n", file=stdout())
    print(exclude_list[[sample_name]], file=stdout())
  }
  exclude_indices = match(exclude_list[[sample_name]], colnames(countmat))
# individual excludes
exclude_indices =
c(exclude_indices, match(exclude_all, colnames(countmat))) # excludes from
all samples
  exclude_indices = exclude_indices[!is.na(exclude_indices)] # remove NA
  # call ExomeDepth to select best reference set, excluding
  # the sample itself as well as all IBD-excluded samples from the
  # set of potential references.
  reference_list = select.reference.set(test.counts = countmat[,i],
reference.count =
as.matrix(countmat[,c(-i, -exclude_indices)]), # as.matrix in case only 1
col
countdf$start)/1000,
bin.length=(countdf$end-
n.bins.reduced = 10000)
  if(opt$verbose) {

```



```

    cat("Using the following samples as reference:\n",file=stdout())
  }
  # whether verbose or not, write to disk what the reference set was, for
  later use
  refpath = paste(sample_name, ".reflist.txt", sep="")
  cat(paste("# SAMPLE: ", sample_name, "\n", sep=""), file=refpath)
  cat(paste("# REFERENCE:
", paste(reference_list$reference.choice, collapse="
"), "\n", sep=""), file=refpath, append=TRUE)
cat(paste(paste(colnames(countmat), collapse="\t"), "\n", sep=""), file=refpa
th, append=TRUE)
  cat(paste(paste(1*(colnames(countmat) %in%
reference_list$reference.choice), collapse="\t"), "\n", sep=""), file=refpath
, append=TRUE)
  reference_set = apply(
    X = as.matrix(countdf[, reference_list$reference.choice]),
    MAR=1, FUN=sum)
  all_exons = new('ExomeDepth', test=countmat[,i],
    reference=reference_set,
    formula = 'cbind(test,reference) ~ 1')
  all_exons = CallCNVs(x = all_exons, transition.probability=10^-4,
    chromosome=countdf$space, start=countdf$start,
    end=countdf$end, name=countdf$names)
# Guardar matriz de datos para usar en gráficos posteriormente si se
necesita
rdspath = paste(sample_name, ".rds", sep='')
saveRDS(all_exons, file = rdspath)
# write out cnv calls
cnvpath = paste(sample_name, ".csv", sep='')
cat("", file=cnvpath)
write.table(all_exons@CNV.calls, file=cnvpath, append=TRUE,
  sep=',', row.names=TRUE, col.names=TRUE, quote=FALSE)
  if (opt$verbose) {
cat(paste("..done.\n", sep=''), file=stdout())
  }
}
duration = format(Sys.time() - start_time)
if(opt$verbose) {
  cat(paste("Completed execution in
", duration, "\n", sep=''), file=stdout())
}

```

Tabla 39. Script para anotación final de CNVs

```

#!/usr/bin/perl -w
# SCRIPT: annotsv.pl
# Entrada: Deben indicarse los directorios de entrada y de salida
# Ejemplo: perl annotsv.pl directorio_input directorio_results
# Salida: Fichero xlsx anotado con AnnotSV
#####

# Activar la declaracion estricta de variables
use strict;
use Term::ANSIColor;
use Parallel::ForkManager;
my $max_processors=12;
## Paso 1. Adquirir el nombre del directorio
my ($directorio_in,$directorio_out,$n_argumentos);

```

```

$n_argumentos = scalar(@ARGV);
($n_argumentos == 2) or
print_error("Error argumentos. Indica directorios de entrada y salida");
($directorio_in,$directorio_out) = @ARGV;
print_ok("Lectura de directorios");
## Paso 2. Abrir el directorio
my ($fichero,$fichero_in,$fichero_out);
my ($nombre,$comando,$rtcd);
opendir(DIR,$directorio_in);
print_ok("Acceso al directorio de ficheros: $directorio_in");
my $super_fork= new Parallel::ForkManager($max_processors);
while ($fichero = readdir(DIR))
{
    ## Control sobre los directorios y ejecución de programas externos
    next unless ($fichero =~ m/\.csv$/);
    next if ($fichero =~ /^\.\/);
##### Start fork #####
my $pid= $super_fork->start and next;
## Paso 3.
    ($nombre = $fichero) =~ s/\.csv//;
    $fichero_in = "$directorio_in/$fichero";
    $fichero_out = "$directorio_out/$nombre.1";
    $comando = "sed -e 's/,/\t/g' $fichero_in | cut -f2-16 - >
$fichero_out";
    $rtcd = system($comando);
## Paso 4.
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.2";
    $comando = "cut -f5-6 $fichero_in | paste - $fichero_in >
$fichero_out";
    $rtcd = system($comando);
## Paso 5.
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.bed";
    $comando = "cut -f9 $fichero_in | paste - $fichero_in >
$fichero_out";
    $rtcd = system($comando);
## Paso 6.
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.tsv";
    $comando = "~/AnnotSV_2.2/bin/AnnotSV/AnnotSV.tcl -SVinputFile
$fichero_in -SVinputInfo 1 -outputFile $fichero_out -svtBEDcol 6";
    $rtcd = system($comando);
## Paso 7. Eliminar columnas redundantes y nombrar columnas definitivas
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre.annotSV";
    $comando = "cut -f1-9,14-85 $fichero_in | sed 1d - > $fichero_out ;
sed -i
'liAnnotSV_ID\tSV_chrom\tSV_start\tSV_end\tSV_length\tstart.p\tend.p\tSV_
type\tnexons\tBF\treads.expected\treads.observed\treads.ratio\tAnnotSV_ty
pe\tGene_name\tNM\tCDS_length\ttx_length\tlocation\tlocation2\tintersectS
tart\tintersectEnd\tDGV_GAIN_IDs\tDGV_GAIN_n_samples_with_SV\tDGV_GAIN_n_
samples_tested\tDGV_GAIN_Frequency\tDGV_LOSS_IDs\tDGV_LOSS_n_samples_with_
SV\tDGV_LOSS_n_samples_tested\tDGV_LOSS_Frequency\tGD_ID\tGD_AN\tGD_N_HE
T\tGD_N_HOMALT\tGD_AF\tGD_POPMAX_AF\tGD_ID_others\tDDD_SV\tDDD_DUP_n_samp
les_with_SV\tDDD_DUP_Frequency\tDDD_DEL_n_samples_with_SV\tDDD_DEL_Freque
ncy\t1000g_event\t1000g_AF\t1000g_max_AF\tIMH_ID\tIMH_AF\tIMH_ID_others\t
promoters\tbdbVar_event\tbdbVar_variant\tbdbVar_status\tTADcoordinates\tENCO
DEexperiments\tGCcontent_left\tGCcontent_right\tRepeats_coord_left\tRepea

```

```
ts_type_left\tRepeats_coord_right\tRepeats_type_right\tACMG\tHI_CGscore\t
TriS_CGscore\tDDD_status\tDDD_mode\tDDD_consequence\tDDD_disease\tDDD_pmi
ds\tHI_DDDpercent\tsynZ_ExAC\tmisZ_ExAC\tpLI_ExAC\tDelZ_ExAC\tDupZ_ExAC\t
cnvZ_ExAC\tmorbidGenes\tmorbidGenesCandidates\tMim_Number\tPhenotypes\tIn
heritance\tAnnotSV_ranking' $fichero_out";
    $rtcd = system($comando);
## Paso 8. Convertir a xls
    $fichero_in = "$fichero_out";
    $fichero_out = "$directorio_out/$nombre";
    $comando = "python tsv2xlsx_annoar.py $fichero_in";
    $rtcd = system($comando);
    rt_ctrl($rtcd, "Conversión a xlsx: $fichero_out");

##### end fork #####

$super_fork->finish;
}
$super_fork->wait_all_children;
## Paso 9. Borrado archivos intermedios
    my $dir = "$directorio_out";
    unlink glob "$dir/*.1 & $dir/*.2 $dir/*.annotSV $dir/*.bed
$dir/*.tsv";
# Cerrar el directorio
closedir(DIR);
print_ok(":-) Finalizado. Resultados de cada muestra en directorio:
$directorio_out");
exit(0);
```

La priorización de genes candidatos se realizó en base a 6 criterios:

- 1) Herencia autosómica recesiva (variantes homocigosis o heterocigosis compuesta)
- 2) Frecuencia de la variante en bases poblacionales <0.5%
- 3) Genes presentes en la base Mitocarta
- 4) Sospecha de variante con implicaciones estructurales (nonsynonymous SNV, frameshift, splicing, stop gain, etc.)
- 5) Variantes encontradas en un solo paciente con compartidas con el resto de miembros analizados.
- 6) En el caso del análisis de CNVs, el filtrado final se ha realizado de forma manual seleccionando los candidatos según el ranking de AnnotSV, seleccionando las categorías 4 y 5 con el BF (*Bayes factor*) más alto calculado por *ExomeDepth*.

Tras el filtrado anterior se emplearon algoritmos de predicción de patogenicidad (*SIFT*, *Polyphen*, *Mutation Taster*) para seleccionar variantes candidatas para verificación mediante secuenciación Sanger y estudios de segregación familiar.

Se han encontrado mutaciones patogénicas en 4 pacientes (Tabla 40) en 4 genes diferentes: *NDUFB3*, *MTO1*, *ATAD3A* y *FDXR*.

	Edad debut	Fenotipo	Historia familiar	Histoquímica músculo	Déficit cadena respiratoria	Citrato sintasa	mtDNA sec.	mtDNA deleciones	mtDNA depleción
P1	8a	HCM, LA	NO	RRF; COX -	CI; CIII; CIV	↑	NPC	Negativo	Negativo
P2	NAC	HY, LA, RD	NO	NORM	CI; CIII	NORM	NPC	Negativo	Negativo
P3	NAC	HY severa, PE, HCM, LA, HP	CONSG	NA	NORM	NORM	NA	NA	Posible (17%)
P4	ADOL	SN, AOB, Signos MCP	NO	RF; COX-	NORM	↑	NPC	NA	NA

Tabla 40. Pacientes con mutaciones patogénicas (estudio exoma)

a: años; NAC: nacimiento; ADOL: adolescencia; F: femenino; M: masculino; LA: acidosis láctica; HCM: cardiomiopatía hipertrófica; RD: retraso del desarrollo; HP: Hipoplasia cerebelosa; HY: hipotonía; PE: encefalopatía progresiva; SN: sordera neurosensorial; AOB: atrofia óptica bilateral, MCP: mielopatía cordonal posterior; CONSG: consanguinidad (primos hermanos); RRF: *ragged-red fibers*; RS: refuerzos subsarcolémicos; COX -: fibras musculares negativas para tinción Citocromo C; ↑: aumentada; NORM: normal; NPC: no presenta cambios; NA: sin datos

16. Gen *MTO1* (P1)

Los primeros casos de mutaciones en la proteína MTO1 fueron descritos en el año 2012^{184,185}. Actualmente se han publicado 35 casos con mutaciones patogénicas en el gen *MTO1*¹⁸⁶. Este gen codifica la *mitochondrial tRNA translation optimization 1 protein* (OMIM#614667; UniProtKB - Q9Y2Z2) que es una de las dos subunidades de una enzima implicada en la 5-carboximetilaminometilación de la uridina de los tRNAs mitocondriales de leucina, triptófano, glutamina, ácido glutámico y lisina^{187,188}.

En el momento del hallazgo de la mutación en homocigosis c.C1390T (p.R464C; NM012123.4→NP036255) en este estudio, dicha mutación no estaba descrita, pero actualmente se conocen 6 casos que afectan a ese mismo residuo en homocigosis y 2 casos en heterocigosis compuesta^{186,189}. El estudio de segregación familiar demostró que los padres eran heterocigotos para esa mutación (Figura 56) y se confirmó mediante Secuenciación Sanger la presencia en homocigosis en el probando (Figura 56). También se puede apreciar el alto grado de conservación que presenta ese residuo (Figura 57), no encontrarse en la población control sana, y su

clasificación como variante patógena por todos los algoritmos bioinformáticos empleados (Tabla 41).

La característica clínica más frecuente en la presentación inicial de las mutaciones en *MTO1* es la cardiomiopatía hipertrófica, acompañada por acidosis láctica, déficit del complejo IV (normalmente acompañado de CI y/o CIII) de la cadena respiratoria mitocondrial y retraso del desarrollo cognitivo. Nuestro paciente debuta a los dos años y presenta todas las características mencionadas anteriormente, aunque existen casos que no cumplen todos estos requisitos¹⁸⁶ y dificulta el diagnóstico final. Este aspecto hace más relevante aún la realización de un correcto análisis genético que permita identificar deficiencias en *MTO1* para poder abordar un correcto consejo genético y un manejo de la patología más adecuado.

Tabla 41. Mutación homocigosis *MTO1* (c.C1390T)

gnomAD Exome (frecuencia alélica total)	DANN	GERP	SIFT	Mutation Taster
0.00000398 Ningún homocigoto	0.9995	5.48	Damaging (0)	Disease causing (1)

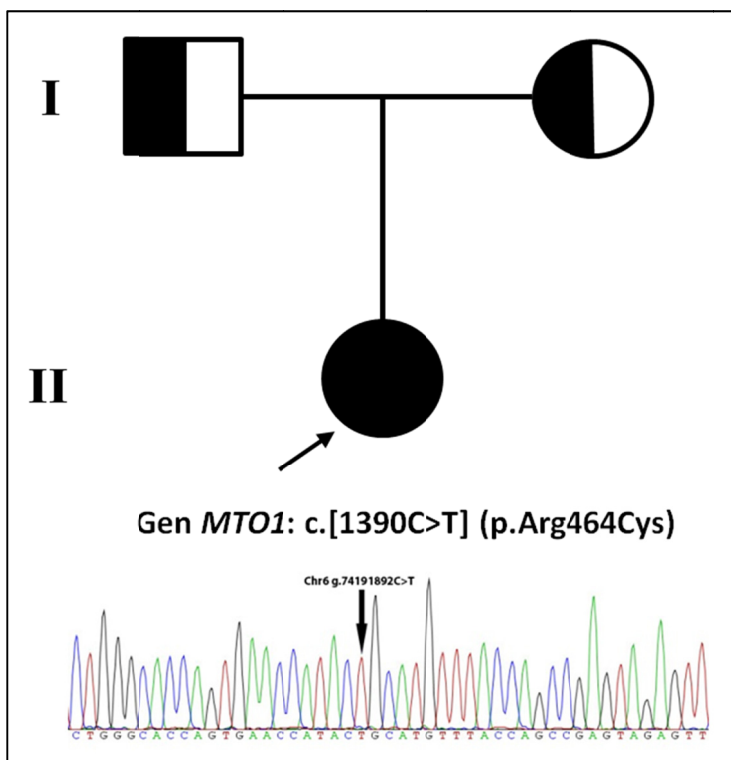


Figura 56. *MTO1* (c.C1390T): Segregación familiar y electroferograma Sec. Sanger

	464
HOMO_SAPIENS	L I D D L T T L G T S E P Y R M F T S R V E F R L S L R P D N A D S R L T L
PONGO_ABELII	L I D D L T T L G T S E P Y R M F T S R V E F R L S L R P D N A D S R L T L
MUS_MUSCULUS	L I D D L T T L G T S E P Y R M F T S R V E F R L S L R P D N A D T R L T F
DANIO_RERIO	L I D D L V C H G V T E P Y R M F T S R A E F R T S L R P D N A D L R L T I
PELODISCUS_SINENSIS	L I D D L T T L G T K E P Y R M F T S R A E F R M S L R P D N A D R R L T L
MONODELPHIS_DOMESTICA	L I D D L T T L G T S E P Y R M F T S R V E F R L S L R P D N A D S R L T L
SCHIZOSACHAROMYCES_CEREVISIAE	L I D D L I N N G V I E P Y R M F T S R S E F R I S V R A D N A D F R L T P

Figura 57. *MTO1* (c.C1390T): Estudio conservación del aminoácido 464

17. Gen *NDUFB3* (P2)

Al igual que en el caso anterior, las primeras mutaciones en este gen fueron descritas en dos casos en el año 2012^{190,191}. El gen *NDUFB3* codifica la *NADH Dehydrogenase (Ubiquinone) 1 Beta Subcomplex, 3* (OMIM#603839; UniProtKB-O43676) que es una subunidad del Complejo I (NADH coenzima Q1 oxidorreductasa) de la cadena respiratoria mitocondrial.

La mutación hallada en homocigosis c.T64C (p.W22R; NM_002491→NP_002482) ha sido publicada en heterocigosis compuesta junto con la mutación c.208G>T (p.G70*) en un caso¹⁹⁰ y en homocigosis en 11 casos más^{191,192}. Dicha mutación se confirmó mediante secuenciación Sanger y el estudio de segregación familiar reveló la presencia de la mutación en heterocigosis tanto en la vía paterna como la materna, heredada por ambas familias de manera independiente (Figura 58).

Todos los casos descritos presentaron déficit del Complejo I en músculo o fibroblastos y cursaron normalmente con acidosis, hipotonía y problemas de desarrollo desde el nacimiento.

Tabla 42. Mutación homocigosis *NDUFB3* (c.T64C)

gnomAD Exome (frecuencia alélica total)	DANN	GERP	SIFT	Mutation Taster
0.000853	0.9962	5.11	Damaging (0)	Disease causing (1)

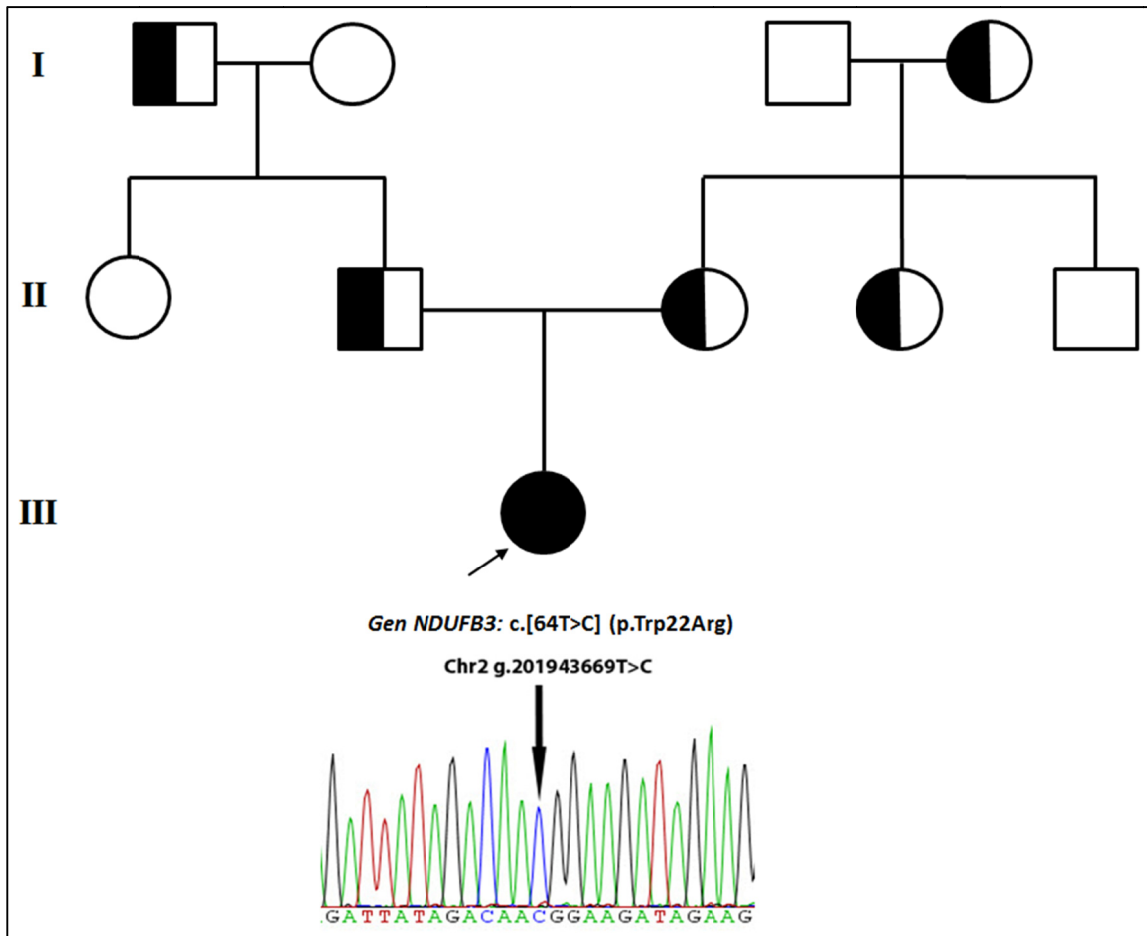


Figura 58. *NDUFB3* (c.T64C): Segregación familiar y electroferograma Sec. Sanger

18. Gen *ATAD3A* (P3)

En este gen han sido descritas diversas mutaciones patológicas con un patrón de herencia tanto dominante como recesivo a partir del año 2016¹⁹³⁻¹⁹⁶. El gen *ATAD3A* (OMIM#612316; UniProtKB-Q9NVI7) codifica la proteína 3A de la familia de las ATPasas con dominio-AAA implicada en diversos procesos celulares entre los que se incluyen: dinámica mitocondrial, participación en el mantenimiento y replicación del mtDNA, crecimiento celular y metabolismo del colesterol. Este gen es un miembro de la familia ATPasa con dominio AAA que incluye otros dos genes parálogos en *Homo sapiens* (*ATAD3B* y *ATAD3C*). Estos tres genes se distribuyen en tándem (*ATAD3C*, *ATAD3B* y *ATAD3A*) cerca del telómero del cromosoma 1.

La mutación c.T1217G (p.L406R; NM_001170536.2) fue hallada homocigosis en el probando y en dos de sus hermanas que presentaban un fenotipo similar al mismo (Figura 59). Otros cuatro hermanos del probando no pudieron ser analizados por no disponer de muestra o por fallecimiento prematuro de los mismos.

Este caso se ha sido aceptado para su publicación en la revista *Molecular Genetics and Metabolism: Novel ATAD3A recessive mutation associated to fatal cerebellar hypoplasia with multiorgan involvement and mitochondrial structural abnormalities*¹⁹⁷.

Tabla 43. Mutación homocigosis *ATAD3A* (c.T1217G)

gnomAD Exome (frecuencia alélica total)	DANN	GERP	SIFT	Mutation Taster
Sin dato de frecuencia	0.9967	4.42	Damaging (0.001)	Disease causing (1)

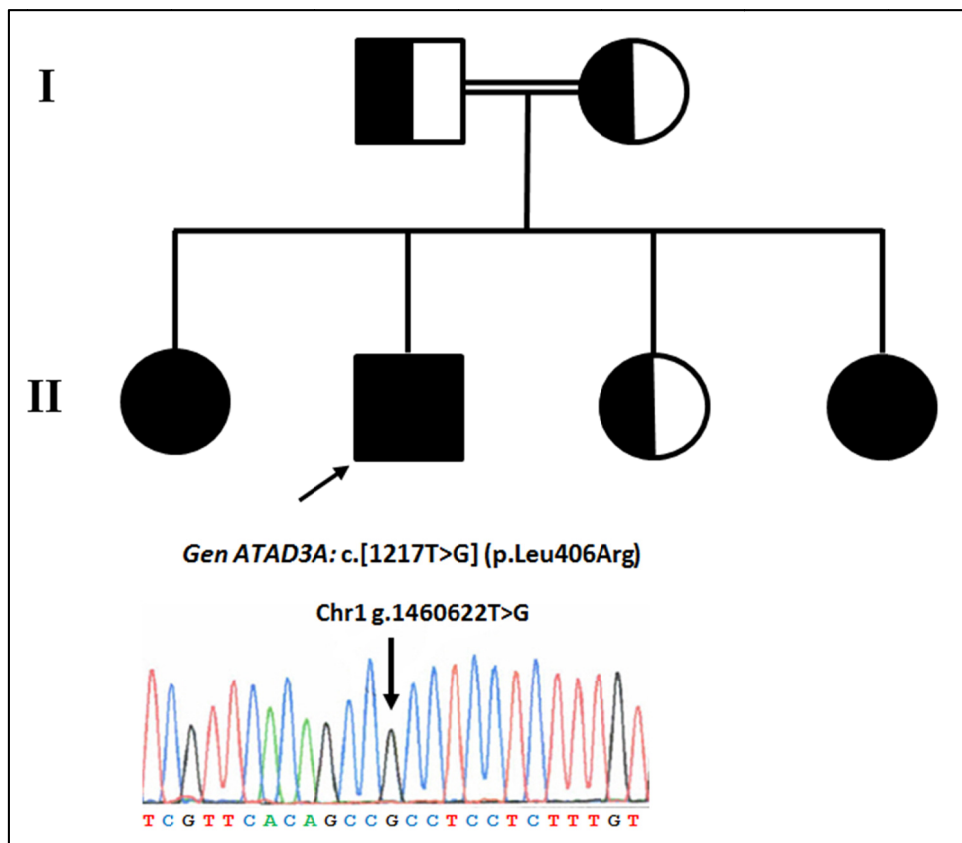


Figura 59. *ATAD3A* (c.T1217G): Segregación familiar y electroferograma Sec. Sanger

19. Gen *FDXR* (P4)

Las primeras mutaciones descritas en este gen y asociadas a neuropatías sensoriales son recientes¹⁹⁸. El gen *FDXR* (OMIM#103270; UniProtKB-P22570) codifica la flavoproteína mitocondrial ferredoxina reductasa, enzima implicada en la formación de varias hormonas esteroideas y en el ensamblaje de los *clusters* Fe-S de la cadena respiratoria mitocondrial.

La mutación c.C814T (p.R272C; NM_001258012) se detectó inicialmente en homocigosis y se catalogó como una variante de significado incierto. En el análisis de segregación, esta variante no está presente ni en la madre, ni en la hermana sana por lo que podría recatalogarse como probablemente patogénica.

Tabla 44. Mutación heterocigosis *FDXR* (c.C814T)

gnomAD Exome (frecuencia alélica total)	DANN	GERP	SIFT	Mutation Taster
0.000183	0.9994	5.62	Damaging (0)	Disease causing (1)

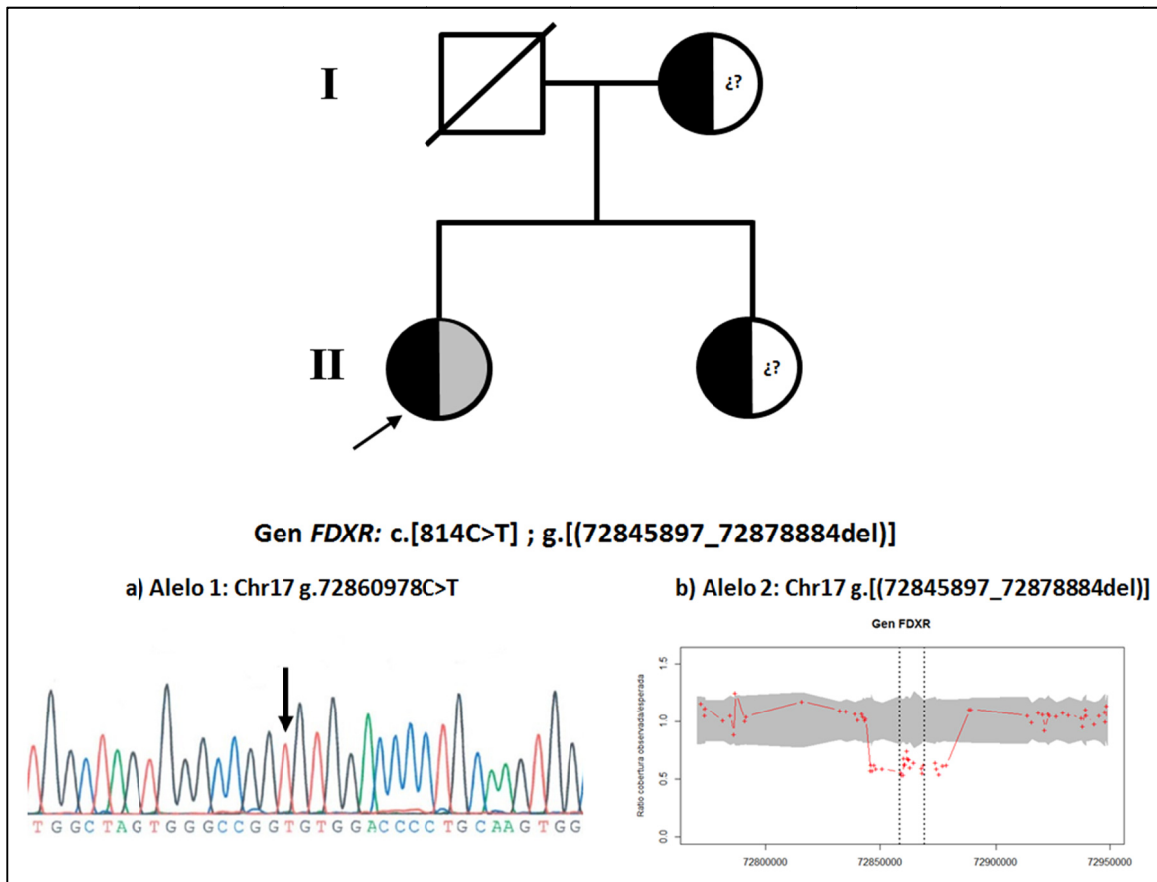


Figura 60. *FDXR*: Segregación familiar y electroferograma Sec. Sanger

a) Electroferograma Secuenciación Sanger. b) Gráfico ratio cobertura observada/cobertura esperada. La zona gris representa el intervalo de confianza al 95% de la ratio calculada y las líneas de puntos acotan el gen *FDXR*.

Estos hallazgos llevan a plantear la hipótesis de una posible delección en el alelo complementario de la paciente. Se realizó un análisis de CNVs sobre los datos del exoma y detectó una delección de casi 33kb en la que se incluye el gen *FDXR*. Como se puede observar en la Figura 60 b) y en la Figura 61, se aprecia una clara disminución de la cobertura en todo el gen *FDXR* que se correspondería con una delección completa en uno de los alelos. Dicha CNV es catalogada como clase 4 por *AnnotSV* y no está presente en las bases de datos DGV, DDD o 1000G. En la base DDD se describen 4 casos que presentaron en heterocigosis pérdidas mucho mayores que las de éste caso que oscilan entre 1,46Mb y 66,75Mb, casi todos los casos son *de novo* (un caso es desconocido) y se clasifican como patogénicos o probablemente patogénicos.

Todo indica que ambas alteraciones en heterocigosis compuesta podrían ser las responsables finales de la patología autosómica recesiva de la paciente.

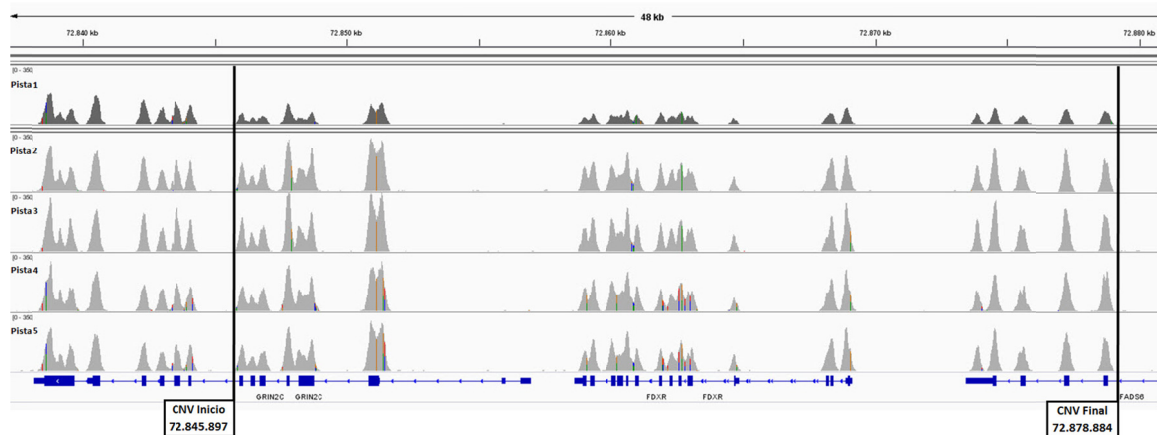


Figura 61. Perfiles cobertura posible CNV en gen *FDXR*

La Pista 1 se corresponde con el paciente y las Pistas 2 a 5 son controles

Establecer un diagrama de flujo para el diagnóstico genético-molecular de las enfermedades OXPPOS en el ámbito del Sistema Nacional de Salud.

El abordaje diagnóstico de las enfermedades raras es siempre un reto dada la heterogeneidad fenotípica que normalmente presentan. En la Figura 62 se muestra el diagrama de flujo propuesto para el diagnóstico de enfermedades del sistema OXPPOS. Se ha tenido especial consideración con los tipos de muestra que se deberían emplear en cada caso y etapa, y con hacer un uso racional de los recursos para generar un balance coste/beneficio adecuado y sostenible. Debe tenerse en cuenta que si se emplea sangre como muestra para secuenciación de mtDNA las deleciones podrían no ser detectadas.

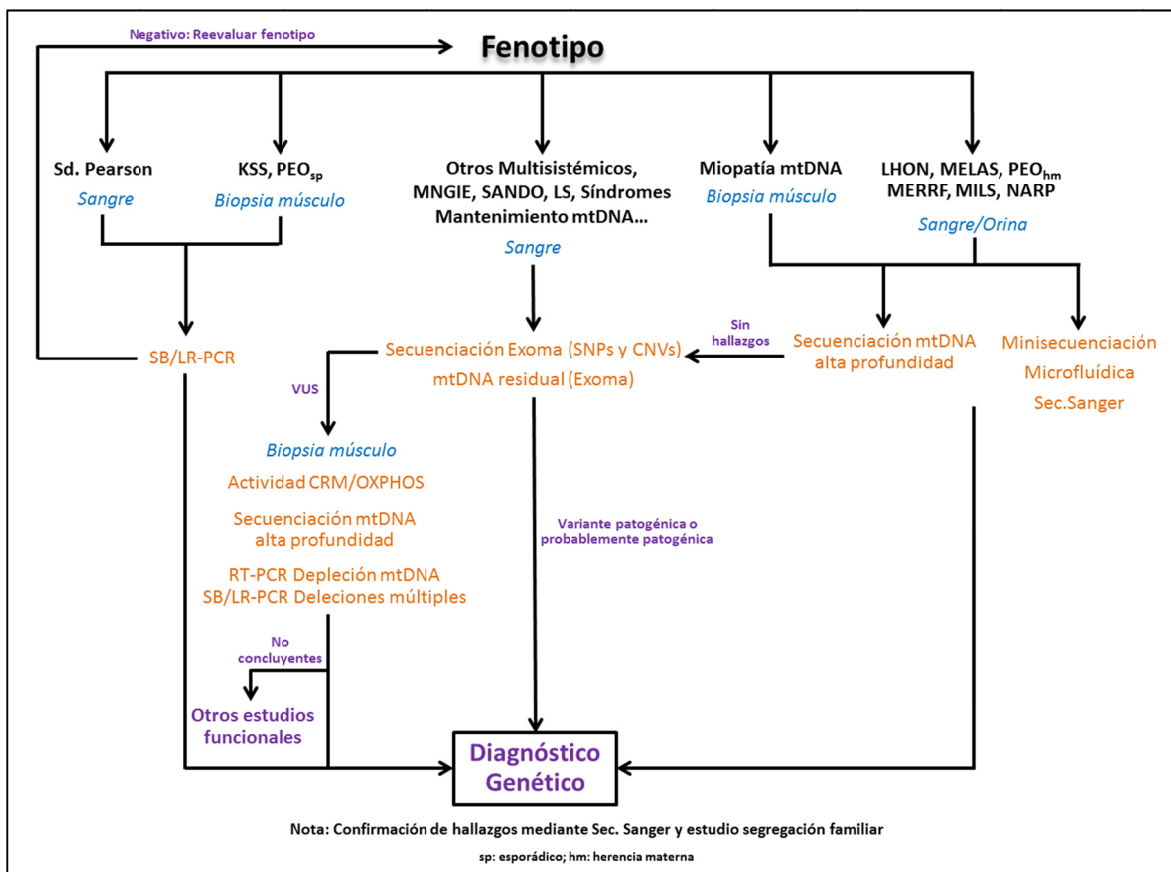


Figura 62. Diagrama flujo diagnóstico enfermedades OXPPOS

Como se puede apreciar entre las técnicas de elección se ha omitido el uso de paneles de genes dirigidos como los empleados hasta la fecha. Consideramos que el estado actual de desarrollo tecnológico y costes permite realizar la secuenciación del exoma con suficientes garantías (sin olvidar que en el fondo no deja de ser un panel amplio en el que siempre existirán

zonas conflictivas mal capturadas). Para el análisis optamos por realizar “ventanas” (paneles virtuales) sobre esos resultados que permitan agilizar el proceso de caracterización de variantes al focalizar el estudio sobre los genes candidatos más compatibles con el fenotipo del paciente, y a la vez obtener una gran flexibilidad para seguir explorando otras regiones no previstas *a priori* en el caso de no llegar a resultados concluyentes.

Discusión

Desarrollo de una estrategia molecular global para la identificación coste-efectiva de anomalías genéticas en el DNA mitocondrial (mtDNA)

Materiales de referencia para la cuantificación de heteroplasmia

Garantizar la validez analítica de cualquier determinación en un laboratorio clínico exige el uso de materiales de control adecuados. En otros ámbitos estos materiales de control son comercializados por diferentes empresas especializadas; pero, en el caso de las pruebas genéticas, es menos frecuente encontrar este tipo de soluciones, y aún menos para pruebas de detección de mutaciones en el mtDNA.

El grado de heteroplasmia es un parámetro relevante en los estudios del mtDNA y, para garantizar una cuantificación óptima y verificar el rendimiento de varias estrategias de cuantificación, se han creado materiales de referencia propios para mutaciones frecuentes. El sistema de clonaje elegido en este caso ha sido *pGEM[®]-T Easy* y ha resultado el más adecuado para los fines que se perseguían. Este vector presenta una alta eficiencia¹⁹⁹, un corto periodo de ligación, un sistema de selección relativamente sencillo y numerosas dianas de restricción para liberar el inserto (Figura 5).

Aunque actualmente existen otros sistemas de clonación como *pSpark^{®200}*, que parece superar a los sistemas *TA cloning*, el método elegido ha resultado eficaz y no precisa de ninguna otra prestación o mejora adicional. A pesar de ser procedimientos de clonaje “rápidos”, se trata de protocolos relativamente laboriosos que requieren el cumplimiento de varios requisitos, con vigilancia especial al mantenimiento de las condiciones de esterilidad para evitar contaminaciones.

La forma de congelación de las células JM109 en glicerol (15%) a -80°C también ha sido adecuada y permite aventurar largos periodos de tiempo de conservación en condiciones óptimas de recuperación: el material genético clonado ha podido ser recuperado en cantidades suficientes, ha sido purificado con facilidad y ha demostrado ser conforme para las aplicaciones ensayadas.

Cuantificación de heteroplasmia en mtDNA: *PCR-RFLP* con cuantificación microfluídica, secuenciación a alta profundidad y Minisequenciación.

Hasta la fecha se han descrito numerosos métodos que pretenden garantizar una correcta cuantificación del grado de heteroplasmia²⁰¹: secuenciación Sanger, *HPLC*, *Snapshot*, *HRM*, *TTGE*, *ARMS*, *NGS*, etc. Y se postula la PCR digital^{202,203} como el método con mayor sensibilidad.

Aquí se ha comprobado que el uso de un método enzimático tradicional como la *PCR-RFLP*, y la visualización con chips microfluídicos en el Bioanalizador 2100 (Agilent) genera un rendimiento similar al método de referencia radioactivo (*Last cycle Hot PCR*). El uso de estos chips para la visualización de los productos de la digestión aporta como ventaja un marcaje equitativo de todos los fragmentos de DNA con independencia de su tamaño. Y, aunque no evita el marcaje de los heterodúplex que se forman en el proceso de *PCR*, gracias a su alta sensibilidad permite visualizarlos y recalcular el porcentaje de heteroplasma adecuadamente. Su facilidad en el manejo y la posibilidad de evitar el empleo de material radiactivo lo convierten en la herramienta idónea para cuantificar la heteroplasma. Aparte de que se obtienen resultados también óptimos en heteroplasmas inferiores al 12% que lo sitúan al nivel de otros métodos de alta sensibilidad como la *RT ARMS-qPCR*²⁰⁴.

Se han verificado las posibilidades de cuantificación de heteroplasma mediante secuenciación del mtDNA a alta profundidad por NGS. En este caso se ha empleado una estrategia que parte de la amplificación del mtDNA, mediante una *PCR* larga con una polimerasa de alta fidelidad, y se han obtenido resultados satisfactorios similares a los del abordaje anterior. En cambio, las pruebas realizadas con Minisequenciación muestran que esta metodología no sería conveniente para asegurar con precisión el grado de heteroplasma.

En la actualidad, los métodos de cuantificación de heteroplasma mediante NGS se están convirtiendo en predominantes y las restantes aproximaciones se emplean para confirmar los resultados de los primeros²⁰³. Los dos métodos descritos anteriormente, y que han demostrado ser adecuados para la cuantificación de heteroplasma, pueden emplearse de manera complementaria de una manera dirigida y de bajo coste con el primero (*PCR-RFLP-Microfluídica*) y en búsquedas de novo o sin diana conocida en el segundo caso (*NGS*). Todo ello teniendo en cuenta el número de muestras que deban analizarse para conseguir una relación coste/efectividad adecuada.

Minisequenciación (*SnapShot*)

Esta metodología de cribado de mutaciones presenta como principales ventajas: la necesidad de muy poca cantidad de DNA de partida (5ng); la obtención de resultados en un periodo de tiempo breve (1-2 días); ser totalmente reconfigurable; ser automatizable y su precio relativamente bajo.

Los principales inconvenientes serían: el tiempo necesario para realizar el diseño y su optimización; la posibilidad de analizar un máximo de unos 40 *SNPs* y los posibles problemas si aparecieran otras variantes frecuentes en las bases contiguas a ambos lados de la posición diana.

En el contexto empleado, el uso de esta aproximación ha sido exitoso, superando la validación de manera completamente satisfactoria. Su bajo coste sitúa a esta metodología como una buena opción de cribado rápido para su traslación a un laboratorio de diagnóstico clínico.

El diseño de haplogrupos propuesto permite realizar una rápida y asequible caracterización de los principales en nuestro entorno, aunque sin llegar al nivel que se consigue con la secuenciación completa del mtDNA a alta profundidad.

Desde el punto de vista epidemiológico, con esta herramienta, en el periodo estudiado, las mutaciones más prevalentes han sido: m.3243A>G (MELAS), m.11778G>A (LHON), m.13513G>A (MELAS/LS), m.14484T>C (LHON), m.8993T>G (NARP) y c.1399G>A. Por lo tanto, las patologías que han sido caracterizadas más frecuentemente por Minisequenciación han sido: MELAS (Encefalopatía mitocondrial, acidosis láctica y episodios tipo ictus) con casi la mitad de los pacientes (49,3%) y LHON (Neuropatía óptica hereditaria de Leber) con cerca de una cuarta parte de los individuos analizados (22,4%).

Secuenciación del mtDNA a alta profundidad en la plataforma PGM

La validación del método ha sido completamente satisfactoria y todas las muestras han sido correctamente caracterizadas con unos niveles de heteroplasmia que se ajustan bien a los obtenidos mediante *PCR-RFLP*-Microfluídica. Cabe destacar la presencia de un perfil de cobertura poco homogéneo (perfil Manhattan) que presenta numerosas regiones “valle” con una fuerte caída del número de lecturas (Figura 31). Este hecho podría tener su explicación en la forma de fragmentar el DNA mediante enzimas. En los ensayos con la química de fragmentos más largos (400pb) no mejora sustancialmente la homogeneidad de la cobertura. Tan solo se aprecia una leve mejoría en las lecturas en sentido R (*reverse*). Todo indica que los sesgos de cobertura observados y el desbalance de lecturas según el sentido podrían tener su origen en la citada estrategia de fragmentación mediante enzimas. Aunque no parece un aspecto crítico para el rendimiento analítico de la plataforma, posiblemente otras enzimas o el uso de fragmentación física podrían mejorar los aspectos citados. En todo caso, debe tenerse en cuenta que esta disparidad de cobertura observada entre las lecturas en sentido F y en sentido R podría ocasionar problemas para la correcta detección de la heteroplasmia en algunas posiciones concretas del mtDNA.

La elección de una estrategia de secuenciación basada en la generación previa de un solo amplicón mediante una PCR larga⁵⁶ genera una cobertura más uniforme que posibilita la detección de grandes deleciones, y permite reducir el impacto que tienen sobre los niveles de heteroplasmia las amplificaciones de DNA nuclear (*NUMTs*). De este modo es posible usar para el alineamiento de las secuencias la *Revised Cambridge Reference Sequence* (rCRS) y no sería imprescindible usar la referencia completa (hg19 o hg38)²⁰⁵. En cualquier caso, la presencia de estas secuencias de mtDNA en el genoma nuclear es siempre un problema para las técnicas de amplificación mediante PCR. Se han descrito hasta la fecha diferentes métodos que pretenden resolver este reto: el estudio más reciente²⁰⁵ resalta que debe tenerse especial precaución al interpretar variantes heteroplásmicas en el *D-loop* y entre las posiciones 5.000 y 10.000 del mtDNA. También debe tenerse precaución cuando se detectan en la *Long-range PCR* los conocidos como *mtDNA sublimons*⁵⁷ y realizar una nueva PCR con otros cebadores y condiciones indicadas en la página 38.

La comparación entre las dos plataformas más habituales (PGM y MiSeq) realizada en colaboración con el Hospital Vall d'Hebron¹⁰³ muestra una sensibilidad y especificidad en la región codificante muy satisfactoria en los dos casos. El rendimiento empeora ligeramente en la región *D-loop*, pero esta zona es muy variable y tiene poca utilidad desde el punto de vista clínico. Por tanto, ambas plataformas resultan apropiadas para su uso en el diagnóstico genético. En los inicios de la secuenciación con PGM en el año 2013 se describía que era necesario optimizar el software de alineamiento para un uso forense adecuado²⁰⁶, posteriormente en el año 2015 ya se ratificaba que PGM presentaba una alta precisión²⁰⁷.

Aplicar una metodología de este tipo en el diagnóstico clínico de rutina de enfermedades raras requiere una plataforma que pueda usarse con un número bajo de muestras a un coste razonable para así poder ofrecer unos tiempos de respuesta adecuados. En el caso de PGM se puede realizar una carrera con 10 muestras de manera coste/efectiva con un chip 316. También se podría analizar una sola muestra en un chip 314 (precio más elevado por muestra) o unas 18 muestras en un chip 318. Esto la convierte en una de las plataformas más versátiles en este aspecto y extremadamente útil en el ámbito de las enfermedades raras. A ello se suma la capacidad para cuantificar el grado de heteroplasmia con exactitud y la obtención del haplogrupo con precisión.

Detección de mtDNA residual en datos de secuenciación de exoma

El análisis desarrollado permite obtener con facilidad las posiciones del mtDNA que difieren con la referencia, ya sean homoplásmicas o heteroplásmicas. El caso publicado¹⁴⁰ pone de manifiesto que es posible detectar mutaciones en el mtDNA con resultados de WES; aunque la cobertura se encuentre reducida tal y como ha sido reportado en varios estudios^{141,208,209}. Datos propios del laboratorio no publicados sugieren una alta correlación en la detección de mutaciones entre la secuenciación de mtDNA a alta profundidad y el estudio de mtDNA residual en WES. Sin olvidar las posibles interferencias que pueden ocasionar los pseudogenes nucleares previsiblemente capturados mediante WES, este análisis de mtDNA residual (sin coste económico añadido) podría emplearse en ciertos casos como una herramienta de cribado previa al uso de la secuenciación del mtDNA a alta profundidad que sí permite también ratificar y asegurar, si procede, el grado de heteroplasmia.

Caracterizar variantes genéticas modificadoras del fenotipo en pacientes con mutaciones patogénicas homoplásmicas en síndromes característicos del mtDNA.

Estudios publicados previamente^{119,210–212} señalan una posible relación entre el desarrollo de LHON y el número de copias de mtDNA. En concreto se observa un incremento estadísticamente significativo en el número de copias de los portadores de mutaciones que no desarrollan LHON. La mayoría de los estudios se centran en familias concretas. El más reciente²¹¹ emplea un número considerable de individuos aislados sin tener en cuenta los haplogrupos mitocondriales, ni el fondo genético, y llega a la misma conclusión: el incremento del número de copias de mtDNA parece ser un factor protector compensatorio del déficit de la cadena respiratoria.

El estudio planteado en este trabajo pretendía añadir más datos esclarecedores en un número más amplio de familias, y además añadir la búsqueda de polimorfismos en el mtDNA que pudiesen estar involucrados; pero, en las cuatro familias estudiadas, no se ha podido ratificar ninguna asociación entre el número de copias de mtDNA y el desarrollo de la enfermedad. Tampoco se ha podido corroborar la presencia de variantes en el mtDNA que pudiesen ejercer un papel desencadenante o, por el contrario, presentasen una función protectora.

Tal como se ha apreciado en otros trabajos¹¹⁹, el número de copias del mtDNA está condicionado por otros factores externos que introducen variabilidad como la edad, el estado

fisiológico hormonal o los hábitos tabáquicos. Y continúa sin estar clara cuál es la base genética (o genético-ambiental) que permite la compensación exitosa de la enfermedad.

Diseñar paneles de genes nucleares de escala media por resecuenciación masiva para diferentes grupos de patología mitocondrial: epidemiología molecular, estudios depleción, deleciones múltiples y estudio funcional tisular de la mutación p.T108M en el gen *TK2*.

El uso de paneles de genes dirigidos es teóricamente una opción asequible que no requiere un equipamiento de altas capacidades y que permite obtener unas buenas tasas de detección de mutaciones causales. En este trabajo, con el panel más pequeño de genes relacionados con el mantenimiento del mtDNA, se ha logrado el diagnóstico genético en un 16% de los casos y un posible diagnóstico en un 7%. En el caso del panel OXPHOS el diagnóstico tuvo lugar en el 10% de los casos y un porcentaje probable también del 7%. Estas aproximaciones mejoran sin duda las tasas de detección existentes en la etapa anterior a la NGS, pero parecen quedarse ligeramente atrás respecto a los resultados que se están obteniendo actualmente con aproximaciones más globales (y costosas) de WES y WGS que presentan una tasa media de diagnóstico molecular del 31%, aunque con un amplio abanico según el estudio (rango 24-68%)²¹³.

En un estudio reciente que analiza la experiencia del análisis con NGS en una amplia serie (5 años) se obtuvo una tasa diagnóstica del 24% y un posible diagnóstico del 9,5%²¹⁴. El porcentaje inferior obtenido con nuestra aproximación quizá tenga que ver con el tipo de patología estudiada y con el tamaño de los paneles empleados. En nuestro caso el panel con mayor número de genes no ha sido el más rentable en términos diagnósticos y pone de manifiesto que el uso de paneles dirigidos seguramente está muy condicionado por la información y caracterización previa de la patología objeto de estudio.

Parece recomendable, siempre que las condiciones económicas y técnicas lo permitan, realizar un abordaje de secuenciación más global (WES) y centrar el análisis en paneles virtuales que se puedan ir expandiendo a demanda en función de otros hallazgos, la aportación de nuevos datos clínicos o la publicación de nueva información sobre el desarrollo de la patología.

Entre los dos paneles, se han localizado 142 mutaciones (Tabla 31 y Tabla 33), de las que 56 (39%) son variantes de significado incierto que requieren la realización de más estudios complementarios y abren la puerta a futuras investigaciones.

Cabe una mención especial al gen *TK2* en el que se puede predecir, según datos de ClinVar (Tabla 32) una alta concentración de mutaciones que efectivamente se confirma en nuestra serie y lo convierte en un gen altamente susceptible a mutaciones deletéreas que debe ser estudiado con detalle.

En cuanto a los estudios de depleción, los resultados obtenidos en la cohorte de niños analizada sugieren que tan solo deberían considerarse como verdaderos positivos los niveles de depleción por debajo del 30-20%. Este valor coincide con el descrito anteriormente por otros autores que consideran la existencia de depleción cuando los niveles de mtDNA se sitúan por debajo del 30% respecto a controles de la misma edad^{215,216}, o depleción moderada si superan ligeramente este punto¹⁴⁴. Por el contrario una revisión reciente indica como punto de corte empírico para el diagnóstico de un síndrome de depleción primario una reducción menos acusada del 60-65%²¹⁷, lo que pone de relieve la diversidad existente para establecer un punto de consenso concreto.

Con la información disponible, actualmente los puntos de decisión del laboratorio, referidos a controles ajustados por edad, para considerar depleción de mtDNA en músculo esquelético son: i) por encima del 40% se considera normal, ii) entre 20%-40% se consideran niveles de mtDNA reducidos y iii) por debajo del 20% se trataría de depleción de mtDNA.

Por último, debe destacarse nuevamente el protagonismo del gen *TK2* en este grupo de niños, ya que aparece mutado en dos de los tres casos que presentaban la reducción del número de copias más acusada.

El análisis de los datos obtenidos en el estudio de deleciones múltiples permite concluir que el uso conjunto de las técnicas de *Southern blot* y *LR-PCR* aumenta de manera efectiva la tasa diagnóstica. Todo indica que la *LR-PCR* presenta una alta sensibilidad y reduce los falsos negativos del *Southern blot*, aunque no deben olvidarse los artefactos que puede cometer la *PCR* y que pueden llevar a un error de interpretación.

En cuanto a la fisiopatología, se considera desde los años 90 que las moléculas de mtDNA delecionadas tendrían cierta ventaja a la hora de replicarse²¹⁸; pero estudios recientes en *C. elegans*²¹⁹ no encuentran evidencia de que el proceso de expansión clonal dependa del tamaño del mtDNA. Sigue abierto el debate sobre cómo tiene lugar el fenómeno de replicación del mtDNA y se barajan dos hipótesis principales²²⁰: i) el propio genoma mitocondrial (de un modo “egoísta”) se apoderaría de ciertas rutas celulares, o ii) se trataría de otra faceta de una relación simbiótica muy adaptable entre célula y mitocondria guiada por mecanismos específicos de tejido y especie.

Tal y como se ha mencionado anteriormente, el gen *TK2* aparece destacado en los estudios realizados en este trabajo y se han podido realizar ciertos estudios adicionales en varios tejidos de un paciente afectado por una mutación (página 109).

El amplio abanico fenotípico observado en las mutaciones de *TK2* no puede explicarse ni por el tipo de mutación, ni por los efectos moleculares sobre el mtDNA (depleción, delección), ni por la actividad *TK2* residual, por lo que se ha postulado¹⁵¹ que debería existir algún mecanismo bioquímico desconocido compensatorio del defecto en *TK2* que permita la supervivencia hasta la edad adulta de algunos individuos.

Se ha realizado recientemente un estudio retrospectivo de historia natural del déficit de *TK2*²²¹ que establece tres tipos fenotipos característicos en función de la edad de debut de la enfermedad: en el primer año de vida (*Infantile-onset myopathy*), entre 1-12 años (*Childhood-onset myopathy*) y a partir de los 12 años (*Late-onset myopathy*); y asocia las mutaciones halladas a cada uno de ellos. La mutación p.T108M (c.C323T; NM_004614) aparece relacionada con los tres fenotipos mencionados.

En el caso del diafragma, que es un tejido afectado en casi todos los casos, parece que la variabilidad de respuesta individual que presenten diferentes vías relacionadas con la proteólisis y la atrofia muscular¹⁸³ tiene un papel fundamental en la progresión y gravedad de la patología, y permite especular que la amplia variabilidad fenotípica observada podría deberse a cómo se ven afectadas por la causa primaria (alteración del metabolismo de los desoxirribonucleótidos mitocondriales) otras rutas metabólicas relativamente lejanas como las mencionadas.

Como nota ilusionante cabe mencionar que se han publicado recientemente resultados muy prometedores²²² del impacto del tratamiento con nucleósidos en pacientes con déficit de *TK2*, que son especialmente favorables cuando se inicia el tratamiento en edades tempranas.

Identificar nuevas mutaciones en genes asociados a enfermedad OXPHOS mediante resecuenciación de exoma en familias candidatas.

Los estudios de secuenciación del exoma en familias o pacientes seleccionados han permitido encontrar mutaciones deletéreas en varios genes:

Gen *MTO1*: aunque inicialmente se consideraba que la cardiomiopatía hipertrófica era la característica patognomónica del déficit de *MTO1*, hoy en día, como ocurre en la mayoría de

enfermedades raras, se sabe que el espectro fenotípico es más amplio e incluye también retraso cognitivo y del desarrollo, hipotonía y patología ocular.

Desde el punto de vista bioquímico es frecuente la acidosis láctica, la elevación en plasma de alanina y la presencia de metabolitos intermediarios del ciclo de Krebs en la orina. Pero la prueba por excelencia es la actividad de los complejos de la cadena respiratoria en músculo. Hasta la fecha todos los pacientes descritos presentan déficits combinados de los complejos I-III-IV o los complejos I-IV¹⁸⁶.

Gen *NDUFB3*: este gen no presenta una casuística muy elevada. En el panel de genes OXPHOS no se ha encontrado ninguna mutación candidata en nuestra serie. En la base de datos HGMD solo se hace mención a tres publicaciones ya mencionadas anteriormente (página 129) y, en ClinVar, tan solo aparecen descritas como patogénicas dos mutaciones puntuales más en región codificante: c.182T>C (p.V61A) y c.200del (p.F67fs).

Al igual que el paciente aquí estudiado, los casos descritos hasta la fecha tienen como características fenotípicas comunes la hipotonía y el retraso del desarrollo. Desde el punto de vista bioquímico se ha observado acidosis láctica y déficit del complejo I de la cadena respiratoria mitocondrial. El estudio más completo hasta hoy está realizado en 10 pacientes irlandeses (8 familias) que presentaron la mutación p.W22R en homocigosis¹⁹². En esta serie los autores destacan la presencia de características faciales peculiares y baja estatura como signos relevantes para sospechar la presencia de esta mutación. A pesar de que varios de los casos descritos han tenido un desenlace fatal a edad temprana, en la última serie irlandesa mencionada anteriormente, también se han observado casos con un desarrollo mucho más favorable, independientemente de los problemas metabólicos agudos graves que padecieron algunos pacientes.

Gen *ATAD3A*: desde el punto de vista clínico, mutaciones en este gen aparecen asociadas con distintos síndromes neurológicos y fenotipos que incluyen retraso del desarrollo psicomotor, discapacidad intelectual, hipotonía, ataxia, espasticidad y neuropatía periférica¹⁹³⁻¹⁹⁵. En los casos con deleciones bialélicas se asocia con hipoplasia cereberal severa y letalidad neonatal¹⁹³.

En cuanto a estudios complementarios de laboratorio se ha observado en algunos casos disfunción mitocondrial¹⁹³, actividades de los complejos OXPHOS en el límite de la normalidad¹⁹⁵ y alteraciones de la dinámica mitocondrial¹⁹⁴.

Gen *FDXR*: la implicación de este gen en el desarrollo de una enfermedad mitocondrial de herencia autosómica recesiva, con atrofia óptica y cierto grado de ataxia como manifestaciones

principales, ha sido recientemente estudiada en cultivos celulares y en un modelo de ratón²²³. También se han observado efectos neurodegenerativos en el cerebro por acumulación de hierro tal y como ocurre en otras patologías como la Ataxia de Frederich o la Enfermedad de Parkinson²²⁴.

En nuestro caso, la detección en homocigosis de una mutación de significado incierto (c.C814T; p.R272C) sin madre portadora condujo a la búsqueda de una hipotética delección en el alelo complementario y se desarrolló un *pipeline* de análisis que permitiese el análisis de CNVs (página 116). En las publicaciones encontradas hasta la fecha existe disparidad en cuanto al rendimiento de las diferentes herramientas para la detección de CNVs. Tal y como advierte en una de ellas²²⁵ se observan bajas concordancias entre las diferentes herramientas que ponen de manifiesto el estado actual de inmadurez que presenta la detección de CNVs en WES. *ExomeDepth* también ha mostrado una gran disparidad de resultados en diferentes trabajos, aunque se ha publicado²²⁶ que es una herramienta especialmente eficaz si se emplean muestras que han sido secuenciadas del mismo modo, con poca variabilidad técnica por parte del laboratorio (misma librería, mismo secuenciador,...).

Teniendo en cuenta las fortalezas y debilidades de cada técnica todo indica que sí están aportando resultados relevantes incluso usándolas con paneles de genes. Para esta última aplicación mencionada se ha propuesto como mejores elecciones *DECoN*^{42,227} y *ExomeDepth*⁴², y la combinación de otras en función de los requisitos de sensibilidad y especificidad que se necesiten. Las utilidades con mejor rendimiento evaluadas en una publicación reciente⁴² serían: *DECoN*, *ExomeDepth*, *ExomeCNV*, *CANOES*, and *CoNVaDING*.

El estudio realizado en este trabajo sobre datos del exoma mediante el uso de *ExomeDepth* y *AnnotSV* muestra una posible CNV en el gen *FDXR* que, en heterocigosis compuesta, parece explicar la patología de la paciente. Quedaría pendiente su confirmación mediante otra tecnología de alta sensibilidad como podría ser un Array de SNPs (*High-resolution single-nucleotide polymorphism genotyping array*^{228,229}).

Este caso es un ejemplo de la complejidad y dificultad que en ocasiones presenta el análisis genético y pone de relieve la necesidad de estudios complementarios más allá de la única secuenciación exoma para poder llegar a una conclusión certera.

Establecer un diagrama de flujo para el diagnóstico genético-molecular de las enfermedades OXPHOS en el ámbito del Sistema Nacional de Salud.

El diagrama propuesto en la página 134 tiene como objetivo adaptarse a las metodologías presentes en la actualidad en el laboratorio de Enfermedades Mitocondriales y Neurometabólicas. Posee un enfoque práctico que podría ser asumible por cualquier laboratorio clínico de análisis molecular que pretenda afinar el coste/beneficio en el diagnóstico de enfermedades mitocondriales. Dada la heterogeneidad clínica de estas patologías se han publicado diferentes propuestas al respecto^{230,231}.

Algunos autores y autoras proponen emplear en todo caso las metodologías de secuenciación antes que la realización de la biopsia muscular²³¹, dejando a esta en un segundo nivel. Se ha calculado que los test genéticos podrían evitar la biopsia en más del 50% de los pacientes. Aunque el uso de estudios genéticos es de gran utilidad en gran parte de los casos, la biopsia muscular puede presentar ventajas en ciertos casos como, por ejemplo, para el diagnóstico de la Oftalmoplejia Progresiva Externa (PEO). Este tejido permite confirmar la presencia de disfunción mitocondrial en estudios histopatológicos, la obtención de mtDNA para los estudios de deleciones y la consiguiente orientación de estudios genéticos en función de los resultados. No cabe duda de que los estudios en tejidos son de gran utilidad cuando no se puede ratificar un diagnóstico o a la hora de interpretar resultados genéticos inciertos y dilucidar la patogenicidad de determinadas mutaciones.

Actualmente las tecnologías de secuenciación masiva aumentan y agilizan de manera considerable el diagnóstico de las enfermedades mitocondriales. Es un hecho que en algunas entidades se ha pasado en los últimos años de una tasa diagnóstica menor del 20% a superar el 60%²³².

Continúa abierto el debate sobre la conveniencia de realizar paneles de genes dirigidos o el uso de una estrategia de secuenciación de Exoma completo (*WES*) o incluso Genoma (*WGS*). Un metaanálisis reciente²¹³, que incluye 37 estudios llevados a cabo entre los años 2011 y 2017 en 20.068 niños, propone el uso de *WES* o *WGS* como análisis de primera línea en niños con sospecha de enfermedad genética. Por el momento el uso de *WES* o *WGS* no presenta diferencias, pero sí parece concluyente que tienen mejor rendimiento si se usan tríos (madre, padre, hijo) y que superan claramente a los *microarrays*.

Se puede aventurar que, de manera progresiva, el uso de *WGS* se irá generalizando y será la estrategia de elección, a la vez que se desarrollen diferentes métodos y formas de análisis más o

menos focalizados en regiones concretas y se enriquezcan las bases de datos con más datos poblacionales.

En el momento actual, se observan amplias diferencias en las tasas de diagnóstico publicadas que, aunque emplean la misma tecnología, probablemente estén reflejando la diferente complejidad y heterogeneidad de los diferentes grupos de enfermedades, la experiencia y conocimiento de las bases fisiopatológicas de las diferentes patologías por parte de los analistas y las posibilidades de realización de pruebas complementarias preanálisis o postanálisis que presenta cada laboratorio.

Desde un punto de vista más global, el modelo organizativo para el diagnóstico de enfermedades raras no diagnosticadas en marcha en Estados Unidos desde el año 2014 es la *Undiagnosed Diseases Network (UDN)*²³³ creada por los *National Institutes of Health*. Dicha red está conformada por un centro coordinador, siete centros clínicos, dos cores de secuenciación, un biorrepositorio centralizado, un core de estudios metabolómicos y un centro de modelos animales. En esta misma línea, en España, el CIBERER (Centro de Investigación Biomédica en Red - Enfermedades Raras) ha puesto en marcha recientemente una iniciativa (*ENoD: Programa CIBERER de Enfermedades raras No Diagnosticadas*)²³⁴ que pretende alcanzar el diagnóstico genético en aquellos pacientes que siguen sin el mismo tras pasar por los cauces asistenciales establecidos en la actualidad.

Para concluir hay que destacar, como apuntan también otros autores y autoras²³⁵, que transcurridos ya más de treinta años desde el comienzo de la llamada “medicina mitocondrial”, se cuenta con la tecnología y las herramientas necesarias para conseguir un diagnóstico genético más rápido y seguro, y dar el salto a la exploración de alternativas terapéuticas eficaces para tratar este tipo de enfermedades tan complejas.

Conclusiones

1. El clonaje de fragmentos de mtDNA en el sistema *pGEM-T* es adecuado para su uso como material de control de calidad estable para pruebas de cuantificación del grado de heteroplasmia.
2. La metodología basada en minisequenciación es rápida y de bajo coste para el genotipado a pequeña escala de mutaciones puntuales (hasta 17 mutaciones diferentes), aunque no es apropiada para la cuantificación precisa del grado de heteroplasmia del mtDNA.
3. El método basado en *PCR-RFLP-Microfluídica* permite una adecuada cuantificación del grado de heteroplasmia, es adaptable a diferentes mutaciones en el mtDNA, y evita el uso de instalaciones radiactivas, por lo que puede sustituir al método de referencia (*Last-Cycle Hot PCR*).
4. El flujo de trabajo propuesto para el análisis de mtDNA a alta profundidad ha sido validado con éxito. Permite cuantificar la heteroplasmia adecuadamente (sin que exista repercusión aparente de los *NUMTs*) y facilita la detección de grandes deleciones y la caracterización del haplogrupo mitocondrial con alto grado de definición.
5. La secuenciación de mtDNA en las dos plataformas más utilizadas en la actualidad (PGM y Miseq) presenta un rendimiento similar y ambas aproximaciones serían aptas para el diagnóstico genético.
6. Los *pipelines* de análisis de genes nucleares desarrollados en este trabajo permiten una extensa anotación de variantes de una manera rápida y eficiente. Mediante los dos paneles de genes diseñados se han hallado 54 mutaciones patogénicas o probablemente patogénicas diferentes y 52 variantes con significado incierto.
7. Mediante secuenciación de exoma se han identificado tres mutaciones deletéreas en homocigosis en los genes *MTO1*, *NDUFB3* y *ATAD3A* y una en heterocigosis compuesta (SNP/CNV) en el gen *FDXR*. Asimismo, se ha verificado que puede realizarse un cribado de mutaciones puntuales en el mtDNA con resultados de *WES*.
8. Se ratifica como punto de corte el valor del 20% para considerar depleción de mtDNA y se confirma la alta sensibilidad que posee la *LR-PCR* para la detección de deleciones en el mtDNA.

-
9. Los estudios moleculares en tejidos de la mutación p.T108M en el gen *TK2* muestran que los efectos son dependientes de tejido y sugieren que el espectro fenotípico que se observa podría depender de la respuesta particular de cada tejido a los efectos ocasionados por la causa primaria (alteración del metabolismo de los desoxirribonucleótidos mitocondriales).
 10. Se propone un diagrama de flujo para el diagnóstico genético-molecular de enfermedades OXPHOS que pretende ser coste/efectivo. Se considera que la secuenciación del exoma y su análisis centrado en paneles virtuales es en estos momentos la mejor estrategia para los casos que presenten características multisistémicas y los fenotipos más difusos.

Bibliografía

1. McFarland, R., Taylor, R. W. & Turnbull, D. M. A neurological perspective on mitochondrial disease. *The Lancet Neurology* **9**, (2010).
2. Elliott, H. R., Samuels, D. C., Eden, J. A., Relton, C. L. & Chinnery, P. F. Pathogenic Mitochondrial DNA Mutations Are Common in the General Population. *Am. J. Hum. Genet.* **83**, 254–260 (2008).
3. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251-7 (2016).
4. Attardi, G. & Schatz, G. Biogenesis of mitochondria. *Annu. Rev. Cell Biol.* **4**, 289–333 (1988).
5. Herrmann, J. M. & Neupert, W. Protein transport into mitochondria. *Curr. Opin. Microbiol.* **3**, 210–4 (2000).
6. Shanske, S. Mitochondrial encephalomyopathies: defects of nuclear DNA. *Brain Pathol.* **2**, 159–62 (1992).
7. Rubio, J. C. *et al.* [Enzyme complex defects of the mitochondrial respiratory chain]. *Rev. Neurol.* **26 Suppl 1**, S15-20 (1998).
8. Palacios Zambrano, S. Identificación y caracterización de nuevos genes implicados en la función OXPHOS y su posible asociación a patologías humanas (Tesis Doctoral). (Universidad Autónoma de Madrid, 2018).
9. Schapira, A. H. V. Mitochondrial disease. *Lancet* **368**, 70–82 (2006).
10. Chinnery, P. F. *et al.* Risk of developing a mitochondrial DNA deletion disorder. *Lancet* **364**, 592–596 (2004).
11. Chinnery, P. F. *et al.* The spectrum of hearing loss due to mitochondrial DNA defects. *Brain* **123 (Pt 1)**, 82–92 (2000).
12. DiMauro, S., Andreu, A. L., Musumeci, O. & Bonilla, E. Diseases of oxidative phosphorylation due to mtDNA mutations. *Semin. Neurol.* **21**, 251–60 (2001).
13. Luo, S. *et al.* Biparental Inheritance of Mitochondrial DNA in Humans. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 13039–13044 (2018).
14. Schwartz, M. & Vissing, J. Paternal inheritance of mitochondrial DNA. *N. Engl. J. Med.* **347**, 576–80 (2002).
15. Moraes, C. T. *et al.* Two novel pathogenic mitochondrial DNA mutations affecting organelle number and protein synthesis. Is the tRNA(Leu(UUR)) gene an etiologic hot spot? *J. Clin. Invest.* **92**, 2906–15 (1993).
16. Bogenhagen, D. & Clayton, D. A. Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. *Cell* **11**, 719–27 (1977).
17. Clayton, D. A. Replication of animal mitochondrial DNA. *Cell* **28**, 693–705 (1982).
18. Kasamatsu, H. & Vinograd, J. Unidirectionality of replication in mouse mitochondrial DNA. *Nat. New Biol.* **241**, 103–5 (1973).
19. Clayton, D. A. Transcription and replication of mitochondrial DNA. *Hum. Reprod.* **15 Suppl 2**, 11–7 (2000).
20. Shadel, G. S. & Clayton, D. A. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* **66**, 409–35 (1997).
21. Lecrenier, N. & Foury, F. New features of mitochondrial DNA replication system in yeast and man. *Gene* **246**, 37–48 (2000).
22. Holt, I. J., Lorimer, H. E. & Jacobs, H. T. Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* **100**, 515–24 (2000).
23. Lisowsky, T. Removal of an intron with unique 3' branch site creates an amino-terminal protein sequence directing the scERV1 gene product to mitochondria. *Yeast* **12**, 1501–10 (1996).
24. Young, M. J. & Copeland, W. C. Human mitochondrial DNA replication machinery and disease. *Curr. Opin. Genet. Dev.* **38**, 52–62 (2016).
25. Montoya, J., Christianson, T., Levens, D., Rabinowitz, M. & Attardi, G. Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 7195–9 (1982).
26. Parisi, M. A. & Clayton, D. A. Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science* **252**, 965–9 (1991).
27. Tracy, R. L. & Stern, D. B. Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Curr. Genet.* **28**, 205–16 (1995).
28. Montoya, J., Gaines, G. L. & Attardi, G. The pattern of transcription of the human mitochondrial

- rRNA genes reveals two overlapping transcription units. *Cell* **34**, 151–9 (1983).
29. Kruse, B., Narasimhan, N. & Attardi, G. Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination. *Cell* **58**, 391–7 (1989).
 30. Gaines, G., Rossi, C. & Attardi, G. Markedly different ATP requirements for rRNA synthesis and mtDNA light strand transcription versus mRNA synthesis in isolated human mitochondria. *J. Biol. Chem.* **262**, 1907–15 (1987).
 31. Rahman, J. & Rahman, S. Mitochondrial medicine in the omics era. *Lancet* **0**, 1–15 (2018).
 32. Blázquez, A., Arenas, J. & Martín, M. A. Molecular Genetics of OXPHOS Disorders. *eLS* 1–10 (2016). doi:10.1002/9780470015902.a0025341
 33. Brownstein, C. a *et al.* An international effort towards developing standards for best practices in analysis, interpretation and reporting of clinical genome sequencing results in the CLARITY Challenge. *Genome Biol.* **15**, R53 (2014).
 34. Taylor, R. W. *et al.* Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. *Jama* **312**, 68–77 (2014).
 35. Oliver, G. R., Hart, S. N. & Klee, E. W. Bioinformatics for Clinical Next Generation Sequencing. *Clin. Chem.* **61**, 124–135 (2015).
 36. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, 386–394 (2009).
 37. Weissensteiner, H. *et al.* mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res.* gkw247 (2016). doi:10.1093/nar/gkw247
 38. Castellana, S., Rónai, J. & Mazza, T. MitImpact: an exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants. *Hum. Mutat.* **36**, E2413-22 (2015).
 39. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
 40. Freeman, J. L. *et al.* Copy number variation: new insights in genome diversity. *Genome Res.* **16**, 949–61 (2006).
 41. Zhang, F., Gu, W., Hurles, M. E. & Lupski, J. R. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* **10**, 451–81 (2009).
 42. Roca, I., González-Castro, L., Fernández, H., Couce, M. L. & Fernández-Marmiesse, A. Free-access copy-number variant detection tools for targeted next-generation sequencing data. *Mutat. Res.* **779**, 114–125
 43. Plagnol, V. *et al.* A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* **28**, 2747–54 (2012).
 44. Guo, Y. *et al.* Comparative study of exome copy number variation estimation tools using array comparative genomic hybridization as control. *Biomed Res. Int.* **2013**, 915636 (2013).
 45. Ellingford, J. M. *et al.* Validation of copy number variation analysis for next-generation sequencing diagnostics. *Eur. J. Hum. Genet.* **25**, 719–724 (2017).
 46. Tan, R. *et al.* An evaluation of copy number variation detection tools from whole-exome sequencing data. *Hum. Mutat.* **35**, 899–907 (2014).
 47. de Ligt, J. *et al.* Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum. Mutat.* **34**, 1439–48 (2013).
 48. Sambrook, J., Fritsch, E.F., Maniatus, T. *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbour Laboratory Press, 1989).
 49. Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**, 255–262 (1988).
 50. Alvarez-Iglesias, V., Barros, F., Carracedo, A. & Salas, A. Minisequencing mitochondrial DNA pathogenic mutations. *BMC Med. Genet.* **9**, 26 (2008).
 51. Cohen, B. H., Chinnery, P. F. & Copeland, W. C. POLG-Related Disorders. *GeneReviews at GeneTests: Medical Genetics Information Resource (database online).* University of Washington, Seattle. 1997-2010 Available at: <http://www.genetests.org>. (Accessed: 1st January 2010)
 52. Lott, M. T. *et al.* mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr. Protoc. Bioinforma.* **44**, 1.23.1-26 (2013).

53. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–86 (2000).
54. Vallone, P. M. & Butler, J. M. AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* **37**, 226–31 (2004).
55. Moraes, C. T., Ricci, E., Bonilla, E., DiMauro, S. & Schon, E. A. The mitochondrial tRNA(Leu(UUR)) mutation in mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes (MELAS): genetic, biochemical, and morphological correlations in skeletal muscle. *Am. J. Hum. Genet.* **50**, 934–49 (1992).
56. Zhang, W., Cui, H. & Wong, L.-J. C. Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin. Chem.* **58**, 1322–31 (2012).
57. Kajander, O. A. *et al.* Human mtDNA sublimons resemble rearranged mitochondrial genomes found in pathological states. *Hum. Mol. Genet.* **9**, 2821–35 (2000).
58. Southern, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–17 (1975).
59. Andreu, A. L., Martinez, R., Marti, R. & García-Arumí, E. Quantification of mitochondrial DNA copy number: pre-analytical factors. *Mitochondrion* **9**, 242–6 (2009).
60. Morten, K. J. *et al.* Liver mtDNA content increases during development: a comparison of methods and the importance of age- and tissue-specific controls for the diagnosis of mtDNA depletion. *Mitochondrion* **7**, 386–95 (2007).
61. Matthijs, G. *et al.* Guidelines for diagnostic next-generation sequencing. *Eur. J. Hum. Genet.* **24**, 2–5 (2016).
62. Hume, S. *et al.* CCMG practice guideline: laboratory guidelines for next-generation sequencing. *J. Med. Genet.* (2019). doi:10.1136/jmedgenet-2019-106152
63. Rehm, H. L. *et al.* ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med.* **15**, 733–47 (2013).
64. Roy, S. *et al.* Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J. Mol. Diagnostics* **20**, 4–27 (2018).
65. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–303 (2010).
66. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1-33 (2013).
67. Tange, O. GNU Parallel : The Command-Line Power Tool. *USENIX Mag.* **36**, 42–47 (2011).
68. Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2017). Available at: <https://www.r-project.org/>.
69. R Markdown. Available at: <https://rmarkdown.rstudio.com/>. (Accessed: 30th August 2018)
70. Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86 (2010).
71. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, T. J. Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* **19**, 1–21 (2010).
72. Giardine, B. *et al.* Galaxy : A platform for interactive large-scale genome analysis. *Genome Res.* **15**, 1451–1455 (2005).
73. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).
74. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
75. García-Alcalde, F. *et al.* Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**, 2678–9 (2012).
76. Goto, H. *et al.* Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biol.* **12**, R59 (2011).
77. Speir, M. L. *et al.* The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res.* **44**, D717-25 (2016).
78. Glusman, G., Caballero, J., Mauldin, D. E., Hood, L. & Roach, J. C. Kaviar : an accessible system for

- testing SNV novelty. *Bioinformatics* **27**, 3216–3217 (2011).
79. Li, Q. & Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am. J. Hum. Genet.* 1–14 (2017). doi:10.1016/j.ajhg.2017.01.004
 80. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–4 (2003).
 81. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–9 (2010).
 82. Jagadeesh, K. A. *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* (2016). doi:10.1038/ng.3703
 83. Jagadeesh, K. A. *et al.* S-CAP extends pathogenicity prediction to genetic variants that affect RNA splicing. *Nat. Genet.* (2019). doi:10.1038/s41588-019-0348-4
 84. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Identif. deleterious Mutat. within three Hum. genomes.* **19**, 1553–1561 (2009).
 85. Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
 86. Reva, B., Antipin, Y. & Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.* **8**, R232 (2007).
 87. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2013).
 88. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
 89. Agresti, A. *Categorical Data Analysis.* (Wiley-Interscience, 2003).
 90. Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–13 (2005).
 91. Li, X., Li, R., Lin, X. & Guan, M.-X. Isolation and characterization of the putative nuclear modifier gene MTO1 involved in the pathogenesis of deafness-associated mitochondrial 12 S rRNA A1555G mutation. *J. Biol. Chem.* **277**, 27256–64 (2002).
 92. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–5 (2014).
 93. Quang, D., Chen, Y. & Xie, X. DANN: A deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761–763 (2015).
 94. Ionita-Laza, I., McCallum, K., Xu, B. & Buxbaum, J. D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48**, 214–20 (2016).
 95. Xiong, H. Y. *et al.* RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
 96. Sonney, S. *et al.* Predicting the pathogenicity of novel variants in mitochondrial tRNA with MitoTIP. *PLoS Comput. Biol.* **13**, e1005867 (2017).
 97. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–8 (2011).
 98. Geoffroy, V. *et al.* AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics* **34**, 3572–3574 (2018).
 99. Kopanos, C. *et al.* VarSome: the human genomic variant search engine. *Bioinformatics* **35**, 1978–1980 (2019).
 100. Filippini, S. *et al.* Multiplex SNaPshot for detection of BRCA1/2 common mutations in Spanish and Spanish related breast/ovarian cancer families. *BMC Med. Genet.* **8**, 40 (2007).
 101. Quintáns, B. *et al.* Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci. Int.* **140**, 251–7 (2004).
 102. Jayaprakash, A. D. *et al.* Stable heteroplasmy at the single-cell level is facilitated by intercellular exchange of mtDNA. *Nucleic Acids Res.* **43**, 2177–2187 (2015).
 103. Gago, L. C. Configuración de una estrategia para identificación genético-molecular de pacientes con enfermedad mitocondrial (Tesis Doctoral). (Universitat Autònoma de Barcelona, 2017).
 104. Yarham, J. W. *et al.* A comparative analysis approach to determining the pathogenicity of mitochondrial tRNA mutations. *Hum. Mutat.* **32**, 1319–25 (2011).
 105. Gorman, G. S. *et al.* Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann. Neurol.* **77**, 753–9 (2015).
 106. Crimi, M. *et al.* Mitochondrial-DNA nucleotides G4298A and T10010C as pathogenic mutations: the

- confirmation in two new cases. *Mitochondrion* **3**, 279–83 (2004).
107. Haas, R. H. & Zolkipli, Z. Mitochondrial disorders affecting the nervous system. *Semin. Neurol.* **34**, 321–40 (2014).
 108. Silvestri, G., Moraes, C. T., Shanske, S., Oh, S. J. & DiMauro, S. A new mtDNA mutation in the tRNA(Lys) gene associated with myoclonic epilepsy and ragged-red fibers (MERRF). *Am. J. Hum. Genet.* **51**, 1213–7 (1992).
 109. Hardy, S. A. *et al.* Pathogenic mtDNA mutations causing mitochondrial myopathy: The need for muscle biopsy. *Neurol. Genet.* **2**, e82 (2016).
 110. Sue, C. M. *et al.* Maternally inherited hearing loss in a large kindred with a novel T7511C mutation in the mitochondrial DNA tRNA(Ser(UCN)) gene. *Neurology* **52**, 1905–8 (1999).
 111. Lechowicz, U. *et al.* Application of next-generation sequencing to identify mitochondrial mutations: Study on m.7511T>C in patients with hearing loss. *Mol. Med. Rep.* **17**, 1782–1790 (2018).
 112. Tang, S. *et al.* Transition to next generation analysis of the whole mitochondrial genome: a summary of molecular defects. *Hum. Mutat.* **34**, 882–93 (2013).
 113. Weber, K. *et al.* A new mtDNA mutation showing accumulation with time and restriction to skeletal muscle. *Am. J. Hum. Genet.* **60**, 373–80 (1997).
 114. Rydzanicz, M. *et al.* Screening of the general Polish population for deafness-associated mutations in mitochondrial 12S rRNA and tRNA Ser(UCN) genes. *Genet. Test. Mol. Biomarkers* **13**, 167–72 (2009).
 115. Debray, F.-G. *et al.* Long-term outcome and clinical spectrum of 73 pediatric patients with mitochondrial diseases. *Pediatrics* **119**, 722–33 (2007).
 116. del Mar O’Callaghan, M. *et al.* New mitochondrial DNA mutations in tRNA associated with three severe encephalomyopathic phenotypes: neonatal, infantile, and childhood onset. *Neurogenetics* **13**, 245–50 (2012).
 117. Johns, D. R., Smith, K. H. & Miller, N. R. Leber’s hereditary optic neuropathy. Clinical manifestations of the 3460 mutation. *Arch. Ophthalmol. (Chicago, Ill. 1960)* **110**, 1577–81 (1992).
 118. Cock, H. R., Cooper, J. M. & Schapira, A. H. Functional consequences of the 3460-bp mitochondrial DNA mutation associated with Leber’s hereditary optic neuropathy. *J. Neurol. Sci.* **165**, 10–7 (1999).
 119. Bianco, A. *et al.* Leber’s hereditary optic neuropathy, intellectual disability and epilepsy presenting with variable penetrance associated to the m.3460G >A mutation and a heteroplasmic expansion of the microsatellite in MTRNR1 gene - case report. *BMC Med. Genet.* **19**, 129 (2018).
 120. Valente, L. *et al.* Identification of novel mutations in five patients with mitochondrial encephalomyopathy. *Biochim. Biophys. Acta* **1787**, 491–501 (2009).
 121. Nesbitt, V. *et al.* A national perspective on prenatal testing for mitochondrial disease. *Eur. J. Hum. Genet.* **22**, 1255–9 (2014).
 122. Auré, K. *et al.* Episodic weakness due to mitochondrial DNA MT-ATP6/8 mutations. *Neurology* **81**, 1810–8 (2013).
 123. Larson, A. A. *et al.* Biochemical signatures mimicking multiple carboxylase deficiency in children with mutations in MT-ATP6. *Mitochondrion* **44**, 58–64 (2019).
 124. Sikorska, M. *et al.* Identification of ataxia-associated mtDNA mutations (m.4052T>C and m.9035T>C) and evaluation of their pathogenicity in transmitochondrial cybrids. *Muscle Nerve* **40**, 381–94 (2009).
 125. Pfeffer, G. *et al.* Adult-onset spinocerebellar ataxia syndromes due to MTATP6 mutations. *J. Neurol. Neurosurg. Psychiatry* **83**, 883–6 (2012).
 126. Thyagarajan, D., Shanske, S., Vazquez-Memije, M., De Vivo, D. & DiMauro, S. A novel mitochondrial ATPase 6 point mutation in familial bilateral striatal necrosis. *Ann. Neurol.* **38**, 468–72 (1995).
 127. Campos, Y. *et al.* Leigh syndrome associated with the T9176C mutation in the ATPase 6 gene of mitochondrial DNA. *Neurology* **49**, 595–7 (1997).
 128. Wei, Y., Cui, L. & Peng, B. Mitochondrial DNA mutations in late-onset Leigh syndrome. *J. Neurol.* **265**, 2388–2395 (2018).
 129. Moslemi, A.-R., Darin, N., Tulinius, M., Oldfors, A. & Holme, E. Two new mutations in the MTATP6 gene associated with Leigh syndrome. *Neuropediatrics* **36**, 314–8 (2005).
 130. Brum, M., Semedo, C., Guerreiro, R. & Pinto Marques, J. Motor Neuron Syndrome as a New Phenotypic Manifestation of Mutation 9185T>C in Gene MTATP6. *Case Rep. Neurol. Med.* **2014**, 701761 (2014).
 131. Pelena, D. *et al.* Complete mtDNA sequencing reveals mutations m.9185T>C and m.13513G>A in

- three patients with Leigh syndrome. *Mitochondrial DNA. Part A, DNA mapping, Seq. Anal.* **29**, 1115–1120 (2018).
132. Howell, N., Xu, M., Halvorson, S., Bodis-Wollner, I. & Sherman, J. A heteroplasmic LHON family: tissue distribution and transmission of the 11778 mutation. *Am. J. Hum. Genet.* **55**, 203–6 (1994).
 133. Bolhuis, P. A. *et al.* Rapid shift in genotype of human mitochondrial DNA in a family with Leber's hereditary optic neuropathy. *Biochem. Biophys. Res. Commun.* **170**, 994–7 (1990).
 134. Taylor, R. W., Morris, A. A. M., Hutchinson, M. & Turnbull, D. M. Leigh disease associated with a novel mitochondrial DNA ND5 mutation. *Eur. J. Hum. Genet.* **10**, 141–4 (2002).
 135. Swalwell, H. *et al.* Respiratory chain complex I deficiency caused by mitochondrial DNA mutations. *Eur. J. Hum. Genet.* **19**, 769–75 (2011).
 136. Johns, D. R., Heher, K. L., Miller, N. R. & Smith, K. H. Leber's hereditary optic neuropathy. Clinical manifestations of the 14484 mutation. *Arch. Ophthalmol. (Chicago, Ill. 1960)* **111**, 495–8 (1993).
 137. Carelli, V. *et al.* Biochemical features of mtDNA 14484 (ND6/M64V) point mutation associated with Leber's hereditary optic neuropathy. *Ann. Neurol.* **45**, 320–8 (1999).
 138. Li, S. *et al.* Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population. *Eur. J. Hum. Genet.* **22**, 1040–5 (2014).
 139. Yiş, U. *et al.* A novel mutation in the mitochondrial DNA cytochrome b gene (MTCYB) in a patient with Prader Willi syndrome. *J. Child Neurol.* **30**, 378–81 (2015).
 140. Delmiro, A. *et al.* Whole-Exome Sequencing Identifies a Variant of the Mitochondrial MT-ND1 Gene Associated with Epileptic Encephalopathy: West Syndrome Evolving to Lennox-Gastaut Syndrome. *Hum. Mutat.* **34**, 1623–7 (2013).
 141. Picardi, E. & Pesole, G. Mitochondrial genomes gleaned from human whole-exome sequencing. *Nat. Methods* **9**, 523–4 (2012).
 142. DiMauro, S., Schon, E. A., Carelli, V. & Hirano, M. The clinical maze of mitochondrial neurology. *Nat. Rev. Neurol.* **9**, 429–44 (2013).
 143. El-Hattab, A. W. & Scaglia, F. Mitochondrial DNA depletion syndromes: review and updates of genetic basis, manifestations, and therapeutic options. *Neurotherapeutics* **10**, 186–98 (2013).
 144. Viscomi, C. & Zeviani, M. MtDNA-maintenance defects: syndromes and genes. *J. Inherit. Metab. Dis.* **40**, 587–599 (2017).
 145. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–24 (2015).
 146. Paradas, C. *et al.* TK2 mutation presenting as indolent myopathy. *Neurology* **80**, 504–6 (2013).
 147. Saada, a *et al.* Mutant mitochondrial thymidine kinase in mitochondrial DNA depletion myopathy. *Nat. Genet.* **29**, 342–4 (2001).
 148. Martí, R. *et al.* Hearing loss in a patient with the myopathic form of mitochondrial DNA depletion syndrome and a novel mutation in the TK2 gene. *Pediatr. Res.* **68**, 151–4 (2010).
 149. Carozzo, R. *et al.* Mutation analysis in 16 patients with mtDNA depletion. *Hum. Mutat.* **21**, 453–4 (2003).
 150. Wang, L. *et al.* Molecular insight into mitochondrial DNA depletion syndrome in two patients with novel mutations in the deoxyguanosine kinase and thymidine kinase 2 genes. *Mol. Genet. Metab.* **84**, 75–82 (2005).
 151. Cámara, Y. *et al.* Severe TK2 enzyme activity deficiency in patients with mild forms of myopathy. *Neurology* **84**, 2286–8 (2015).
 152. Galbiati, S. *et al.* New mutations in TK2 gene associated with mitochondrial DNA depletion. *Pediatr. Neurol.* **34**, 177–85 (2006).
 153. Van Goethem, G. *et al.* Novel POLG mutations in progressive external ophthalmoplegia mimicking mitochondrial neurogastrointestinal encephalomyopathy. *Eur. J. Hum. Genet.* **11**, 547–9 (2003).
 154. Scuderi, C. *et al.* The in cis T251I and P587L POLG1 base changes: description of a new family and literature review. *Neuromuscul. Disord.* **25**, 333–9 (2015).
 155. Lamantea, E. *et al.* Mutations of mitochondrial DNA polymerase gammaA are a frequent cause of autosomal dominant or recessive progressive external ophthalmoplegia. *Ann. Neurol.* **52**, 211–9 (2002).
 156. Rempe, T. *et al.* Early-onset parkinsonism due to compound heterozygous POLG mutations. *Parkinsonism Relat. Disord.* **29**, 135–7 (2016).

157. Wong, L.-J. C. *et al.* Molecular and clinical genetics of mitochondrial diseases due to POLG mutations. *Hum. Mutat.* **29**, E150-72 (2008).
158. González-Vioque, E. *et al.* Association of novel POLG mutations and multiple mitochondrial DNA deletions with variable clinical phenotypes in a Spanish population. *Arch. Neurol.* **63**, 107–111 (2006).
159. Rivera, H. *et al.* Mild ocular myopathy associated with a novel mutation in mitochondrial twinkle helicase. *Neuromuscul. Disord.* **17**, 677–80 (2007).
160. Fratter, C. *et al.* The clinical, histochemical, and molecular spectrum of PEO1 (Twinkle)-linked adPEO. *Neurology* **74**, 1619–26 (2010).
161. Van Hove, J. L. K. *et al.* Finding twinkle in the eyes of a 71-year-old lady: a case report and review of the genotypic and phenotypic spectrum of TWINKLE-related dominant disease. *Am. J. Med. Genet. A* **149A**, 861–7 (2009).
162. Spelbrink, J. N. *et al.* Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nat. Genet.* **28**, 223–31 (2001).
163. Fratter, C. *et al.* RRM2B mutations are frequent in familial PEO with multiple mtDNA deletions. *Neurology* **76**, 2032–4 (2011).
164. Shaibani, A. *et al.* Mitochondrial neurogastrointestinal encephalopathy due to mutations in RRM2B. *Arch. Neurol.* **66**, 1028–32 (2009).
165. Pitceathly, R. D. S. *et al.* Adults with RRM2B-related mitochondrial disease have distinct clinical and molecular characteristics. *Brain* **135**, 3392–403 (2012).
166. Dimmock, D. P. *et al.* Clinical and molecular features of mitochondrial DNA depletion due to mutations in deoxyguanosine kinase. *Hum. Mutat.* **29**, 330–1 (2008).
167. Young, M. J. *et al.* Biochemical analysis of human POLG2 variants associated with mitochondrial disease. *Hum. Mol. Genet.* **20**, 3052–66 (2011).
168. Haack, T. B. *et al.* Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat. Genet.* **42**, 1131–4 (2010).
169. Collet, M. *et al.* High incidence and variable clinical outcome of cardiac hypertrophy due to ACAD9 mutations in childhood. *Eur. J. Hum. Genet.* **24**, 1112–6 (2016).
170. Tiranti, V. *et al.* Mutations of SURF-1 in Leigh disease associated with cytochrome c oxidase deficiency. *Am. J. Hum. Genet.* **63**, 1609–21 (1998).
171. Oláhová, M. *et al.* LRPPRC mutations cause early-onset multisystem mitochondrial disease outside of the French-Canadian population. *Brain* **138**, 3503–19 (2015).
172. Friederich, M. W. *et al.* Pathogenic variants in glutamyl-tRNAGln amidotransferase subunits cause a lethal mitochondrial cardiomyopathy disorder. *Nat. Commun.* **9**, 4065 (2018).
173. Potluri, P. *et al.* A novel NDUFA1 mutation leads to a progressive mitochondrial complex I-specific neurodegenerative disease. *Mol. Genet. Metab.* **96**, 189–95 (2009).
174. Alston, C. L. *et al.* Late-onset respiratory failure due to TK2 mutations causing multiple mtDNA deletions. *Neurology* **81**, 2051–3 (2013).
175. Béhin, A. *et al.* Adult cases of mitochondrial DNA depletion due to TK2 defect: an expanding spectrum. *Neurology* **78**, 644–8 (2012).
176. Oskoui, M. *et al.* Clinical spectrum of mitochondrial DNA depletion due to mutations in the thymidine kinase 2 gene. *Arch. Neurol.* **63**, 1122–6 (2006).
177. Mancuso, M. *et al.* Mitochondrial myopathy of childhood associated with mitochondrial DNA depletion and a homozygous mutation (T77M) in the TK2 gene. *Arch. Neurol.* **60**, 1007–9 (2003).
178. Mancuso, M. *et al.* Mitochondrial DNA depletion: mutations in thymidine kinase gene with myopathy and SMA. *Neurology* **59**, 1197–202 (2002).
179. Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* **369**, 1502–11 (2013).
180. Wang, L. *et al.* Molecular insight into mitochondrial DNA depletion syndrome in two patients with novel mutations in the deoxyguanosine kinase and thymidine kinase 2 genes. *Mol. Genet. Metab.* **84**, 75–82 (2005).
181. Götz, A. *et al.* Thymidine kinase 2 defects can cause multi-tissue mtDNA depletion syndrome. *Brain* **131**, 2841–50 (2008).
182. Kirkin, V., McEwan, D. G., Novak, I. & Dikic, I. A role for ubiquitin in selective autophagy. *Mol. Cell* **34**, 259–69 (2009).

183. Levine, S. *et al.* Increased proteolysis, myosin depletion, and atrophic AKT-FOXO signaling in human diaphragm disuse. *Am. J. Respir. Crit. Care Med.* **183**, 483–90 (2011).
184. Ghezzi, D. *et al.* Mutations of the mitochondrial-tRNA modifier MTO1 cause hypertrophic cardiomyopathy and lactic acidosis. *Am. J. Hum. Genet.* **90**, 1079–87 (2012).
185. Vasta, V., Merritt li, J. L., Saneto, R. P., Hahn, S. H. & Merritt, J. L. Next-generation sequencing for mitochondrial diseases: a wide diagnostic spectrum. *Pediatr. Int.* **54**, 585–601 (2012).
186. O’Byrne, J. J. *et al.* The genotypic and phenotypic spectrum of MTO1 deficiency. *Mol. Genet. Metab.* **123**, 28–42 (2018).
187. Powell, C. A., Nicholls, T. J. & Minczuk, M. Nuclear-encoded factors involved in post-transcriptional processing and modification of mitochondrial tRNAs in human disease. *Front. Genet.* **6**, 79 (2015).
188. Suzuki, T., Nagao, A. & Suzuki, T. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. *Annu. Rev. Genet.* **45**, 299–329 (2011).
189. Martín, M. Á. *et al.* The homozygous R504C mutation in MTO1 gene is responsible for ONCE syndrome. *Clin. Genet.* **91**, 46–53 (2017).
190. Haack, T. B. *et al.* Molecular diagnosis in mitochondrial complex I deficiency using exome sequencing. *J. Med. Genet.* **49**, 277–83 (2012).
191. Calvo, S. E. *et al.* Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci. Transl. Med.* **4**, 118ra10 (2012).
192. Alston, C. L. *et al.* A recurrent mitochondrial p.Trp22Arg NDUF3 variant causes a distinctive facial appearance, short stature and a mild biochemical and clinical phenotype. *J. Med. Genet.* **53**, 634–41 (2016).
193. Harel, T. *et al.* Recurrent De Novo and Biallelic Variation of ATAD3A, Encoding a Mitochondrial Membrane Protein, Results in Distinct Neurological Syndromes. *Am. J. Hum. Genet.* **99**, 831–845 (2016).
194. Cooper, H. M. *et al.* ATPase-deficient mitochondrial inner membrane protein ATAD3A disturbs mitochondrial dynamics in dominant hereditary spastic paraplegia. *Hum. Mol. Genet.* **26**, 1432–1443 (2017).
195. Desai, R. *et al.* ATAD3 gene cluster deletions cause cerebellar dysfunction associated with altered mitochondrial DNA and cholesterol metabolism. *Brain* **140**, 1595–1610 (2017).
196. Peeters-Scholte, C. M. P. C. D. *et al.* Genotype-phenotype correlation in ATAD3A deletions: not just of scientific relevance. *Brain* **140**, e66 (2017).
197. Cristina Ugalde, Susana Peralta, Adrian Gonzalez, Marta Ybarra, Aitor Delmiro, Rafael Pérez-Pérez, Jorge Docampo, Joaquín Arenas, Alberto Blázquez, M. A. M. Novel ATAD3A recessive mutation associated to fatal cerebellar hypoplasia with multiorgan involvement and mitochondrial structural abnormalities. *Mol. Genet. Metab.* **Accepted**, (2019).
198. Paul, A. *et al.* FDXR Mutations Cause Sensorial Neuropathies and Expand the Spectrum of Mitochondrial Fe-S-Synthesis Diseases. *Am. J. Hum. Genet.* **101**, 630–637 (2017).
199. Litterer, L. Comparing Cloning Efficiency of the pGEM[®]-T and pGEM[®]-T Easy Vectors to the TOPO TA Cloning[®] Vectors. (2009). Available at: <http://www.promega.es/resources/pubhub/enotes/comparing-cloning-efficiency-of-pgemt-and-pgemt-easy-vectors-to-topo-ta-cloning-vectors/>. (Accessed: 14th May 2018)
200. Canvax Biotech. Advantages of pSpark[®] over other popular DNA Cloning systems on the market: pGEM[®]-T and TOPO TA cloning[®]. Available at: https://www.chemie-brunschwig.ch/documents/suppliers-information/pdf/downloads/CAN_AdvantagespSpark.pdf. (Accessed: 10th March 2019)
201. Sobenin, I. A. *et al.* Quantitative assessment of heteroplasmy of mitochondrial genome: perspectives in diagnostics and methodological pitfalls. *Biomed Res. Int.* **2014**, 292017 (2014).
202. Belmonte, F. R. *et al.* Digital PCR methods improve detection sensitivity and measurement precision of low abundance mtDNA deletions. *Sci. Rep.* **6**, 25186 (2016).
203. Duan, M., Tu, J. & Lu, Z. Recent Advances in Detecting Mitochondrial DNA Heteroplasmic Variations. *Molecules* **23**, (2018).
204. Bai, R.-K. & Wong, L.-J. C. Detection and quantification of heteroplasmic mutant mitochondrial DNA by real-time amplification refractory mutation system quantitative PCR analysis: a single-step approach. *Clin. Chem.* **50**, 996–1001 (2004).
205. Santibanez-Koref, M. *et al.* Assessing mitochondrial heteroplasmy using next generation

- sequencing: A note of caution. *Mitochondrion* (2018). doi:10.1016/j.mito.2018.08.003
206. Parson, W. *et al.* Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). *Forensic Sci. Int. Genet.* **7**, 543–9 (2013).
207. Seo, S. B. *et al.* Underlying Data for Sequencing the Mitochondrial Genome with the Massively Parallel Sequencing Platform Ion Torrent PGM. *BMC Genomics* **16**, S4 (2015).
208. Dinwiddie, D. L. *et al.* Diagnosis of mitochondrial disorders by concomitant next-generation sequencing of the exome and mitochondrial genome. *Genomics* **102**, 148–156 (2013).
209. Puusepp, S. *et al.* Effectiveness of whole exome sequencing in unsolved patients with a clinical suspicion of a mitochondrial disorder in Estonia. *Mol. Genet. Metab. Reports* **15**, 80–89 (2018).
210. Nishioka, T., Soemantri, A. & Ishida, T. mtDNA/ndDNA ratio in 14484 LHON mitochondrial mutation carriers. *J. Hum. Genet.* **49**, 701–705 (2004).
211. Bianco, A. *et al.* Mitochondrial DNA copy number in affected and unaffected LHON mutation carriers. *BMC Res. Notes* **11**, 911 (2018).
212. Giordano, C. *et al.* Efficient mitochondrial biogenesis drives incomplete penetrance in Leber's hereditary optic neuropathy. *Brain* **137**, 335–53 (2014).
213. Clark, M. M. *et al.* Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ genomic Med.* **3**, 16 (2018).
214. Hartman, P. *et al.* Next generation sequencing for clinical diagnostics: Five year experience of an academic laboratory. *Mol. Genet. Metab. reports* **19**, 100464 (2019).
215. Rahman, S. & Poulton, J. Diagnosis of mitochondrial DNA depletion syndromes. *Arch. Dis. Child.* **94**, 3–5 (2009).
216. Bai, R.-K., Perng, C.-L., Hsu, C.-H. & Wong, L.-J. C. Quantitative PCR analysis of mitochondrial DNA content in patients with mitochondrial disease. *Ann. N. Y. Acad. Sci.* **1011**, 304–9 (2004).
217. Nogueira, C. *et al.* Syndromes associated with mitochondrial DNA depletion. *Ital. J. Pediatr.* **40**, 34 (2014).
218. Wallace, D. C. Mitochondrial genetics: a paradigm for aging and degenerative diseases? *Science* **256**, 628–32 (1992).
219. Gitschlag, B. L. *et al.* Homeostatic Responses Regulate Selfish Mitochondrial Genome Dynamics in *C. elegans*. *Cell Metab.* **24**, 91–103 (2016).
220. Picard, M., Vincent, A. E. & Turnbull, D. M. Expanding Our Understanding of mtDNA Deletions. *Cell Metab.* **24**, 3–4 (2016).
221. Garone, C. *et al.* Retrospective natural history of thymidine kinase 2 deficiency. *J. Med. Genet.* **55**, 515–521 (2018).
222. Domínguez-González, C. *et al.* Deoxynucleoside Therapy for Thymidine Kinase 2-Deficient Myopathy. *Ann. Neurol.* **86**, 293–303 (2019).
223. Peng, Y. *et al.* Biallelic mutations in the ferredoxin reductase gene cause novel mitochondriopathy with optic atrophy. *Hum. Mol. Genet.* **26**, 4937–4950 (2017).
224. Slone, J. *et al.* Biallelic mutations in FDXR cause neurodegeneration associated with inflammation. *J. Hum. Genet.* **63**, 1211–1222 (2018).
225. Yao, R. *et al.* Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. *Mol. Cytogenet.* **10**, 1–7 (2017).
226. Kadalayil, L. *et al.* Exome sequence read depth methods for identifying copy number changes. *Brief. Bioinform.* **16**, 380–392 (2014).
227. Fowler, A. *et al.* Accurate clinical detection of exon copy number variants in a targeted NGS panel using DECoN. *Wellcome open Res.* **1**, 20 (2016).
228. Cooper, G. M., Zerr, T., Kidd, J. M., Eichler, E. E. & Nickerson, D. A. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nat. Genet.* **40**, 1199–203 (2008).
229. Fang, L. & Wang, K. Identification of Copy Number Variants from SNP Arrays Using PennCNV. *Methods Mol. Biol.* **1833**, 1–28 (2018).
230. Paiva Coelho, M., Martins, E. & Vilarinho, L. Diagnosis, management, and follow-up of mitochondrial disorders in childhood: a personalized medicine in the new era of genome sequence. *Eur. J. Pediatr.* **178**, 21–32 (2019).
231. Witters, P. *et al.* Revisiting mitochondrial diagnostic criteria in the new era of genomics. *Genet. Med.* **20**, 444–451 (2018).

232. Gorman, G. S. *et al.* Mitochondrial diseases. *Nat. Rev. Dis. Prim.* **2**, 16080 (2016).
233. Splinter, K. *et al.* Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. *N. Engl. J. Med.* **379**, 2131–2139 (2018).
234. CIBERER-ENOD. Available at: <https://enod.ciberer.es/>. (Accessed: 18th October 2019)
235. Carelli, V. & La Morgia, C. Clinical syndromes associated with mtDNA mutations: where we stand after 30 years. *Essays Biochem.* **62**, 235–254 (2018).

Publicaciones en las que ha participado el Doctorando

Publicaciones derivadas en parte de los resultados de este trabajo de tesis doctoral.

1. Delmiro, A., Rivera, H., García-Silva, MT., García-Consuegra, I., Martín-Hernández, E., Quijada-Fraile, P., Simón de las Heras, R., Moreno-Izquierdo, A. Martín, MA., Arenas, J., Martínez-Azorín, F. (2013). **Whole-Exome Sequencing Identifies a Variant of the Mitochondrial MT-ND1 Gene Associated with Epileptic Encephalopathy: West Syndrome Evolving to Lennox-Gastaut Syndrome.** *Human Mutation*, 34(12), 1623–7. <http://doi.org/10.1002/humu.22445>. PMID: 24105702
2. Hernández-Laín, A., Guerrero, A, Domínguez-González, C., Fernández-Vázquez, I., Maya, D., Delmiro, A., Ruiz, J., Blázquez, A., Morán, M., Martín, MA. (2015). **A novel RRM2B gene variant associated with Telbivudine-induced mitochondrial myopathy.** *Journal of the Neurological Sciences*, 358, 481–483. <http://doi.org/10.1016/j.jns.2015.08.1550>. PMID: 26359855
3. Cámara, Y., Carreño-Gago, L., Martín, MA., Melià, MJ., Blázquez, A., Delmiro, A., Garrabou, G., Constanza, M., Díaz-Manera, J., Gallardo, E., Bornstein, B., López-Gallardo, E., Hernández-Lain, A., San Millán, B., Cancho, E., Rodríguez Vico, JS., Martí, R., García-Arumi, E. (2015). **Severe TK2 enzyme activity deficiency in patients with mild forms of myopathy.** *Neurology*, 84, 1–3. PMID: 25948719
4. M. Fernández-Moreno, M., Hermida-Gómez, T., Dalmao-Fernández, A., Rego-Pérez, I., Vázquez-Mosquera, M., Cortés-Pereira, E., Relaño, S., Oreiro-Villar, N., Fernández-López, C., Gallardo-Pérez, E., Delmiro, A., Martín, MA., Garesse, R., Blanco, F. (2016). **In vitro studies help us to understand the relationship between mitochondrial DNA (mtDNA) haplogroups and OA pathogenesis.** *Osteoarthritis and Cartilage*, 24, S143–S144.
5. Martín, MA., García-Silva, MT., Barcia, G., Delmiro, A., Rodríguez-García, ME., Blázquez, A., Francisco-Álvarez, R., Martín-Hernández, E., Quijada-Fraile, P., Tejada-Palacios, P., Arenas, J., Santos, C., Martínez-Azorín, F. (2016). **The homozygous R504C mutation in MTO1 gene is responsible for ONCE syndrome.** *Clinical Genetics*, 91(1), 46–53. <https://doi.org/10.1111/cge.12815>. PMID: 27256614.
6. Ugalde, C., Peralta, S., González, A., Ybarra, M., Delmiro, A., Pérez-Pérez, R., Docampo, J., Arenas, J., Blázquez, A., Martín, MA. (2019). **Novel ATAD3A recessive mutation associated to fatal cerebellar hypoplasia with multiorgan involvement and mitochondrial structural abnormalities.** *Mol. Genet. Metab.* Accepted for publication.

Otros trabajos en los que ha colaborado el doctorando.

7. Rivera, H., Martín-Hernández, E., Delmiro, A., García-Silva, M. T., Quijada-Fraile, P., Muley, R., ... Martínez-Azorín, F. (2013). **A new mutation in the gene encoding mitochondrial seryl-tRNA synthetase as a cause of HUPRA syndrome.** *BMC Nephrology*, 14(1), 195. <http://doi.org/10.1186/1471-2369-14-195>

8. Morán, M., Delmiro, A., Blázquez, A., Ugalde, C., Arenas, J., & Martín, M. a. (2014). **Bulk autophagy, but not mitophagy, is increased in cellular model of mitochondrial disease.** *Biochimica et Biophysica Acta*, 1842(7), 1059–70. <http://doi.org/10.1016/j.bbadis.2014.03.013>
9. Fiuza-Luces, C., Delmiro, A., Soares-Miranda, L., González-Murillo, Á., Martínez-Palacios, J., Ramírez, M., ... Morán, M. (2014). **Exercise training can induce cardiac autophagy at end-stage chronic conditions: insights from a graft-versus-host-disease mouse model.** *Brain, Behavior, and Immunity*, 39, 56–60. <http://doi.org/10.1016/j.bbi.2013.11.007>
10. Castro-Gago, M., Dacruz-Alvarez, D., Pintos-Martínez, E., Beiras-Iglesias, A., Delmiro, A., Arenas, J., ... Martínez-Azorín, F. (2014). **Exome sequencing identifies a CHKB mutation in Spanish patient with Megaconial Congenital Muscular Dystrophy and mtDNA depletion.** *European Journal of Paediatric Neurology: EJPN: Official Journal of the European Paediatric Neurology Society*, 2–6. <http://doi.org/10.1016/j.ejpn.2014.06.005>
11. Yubero D, Montero R, Martín MA, Montoya J, Ribes A,... Brea-Calvo G. **Secondary coenzyme Q(10) deficiencies in oxidative phosphorylation (OXPHOS) and non-OXPHOS disorders.** *Mitochondrion*. 2016 Jun 30;30:51-58. [Epub ahead of print] PMID: 27374853.
12. Pérez-Pérez, R., Lobo-Jarne, T., Milenkovic, D., Mourier, A., Bratic, A., García-Bartolomé, A., ... Ugalde, C. (2016). **COX7A2L Is a Mitochondrial Complex III Binding Protein that Stabilizes the III2+IV Supercomplex without Affecting Respirasome Formation.** *Cell Reports*. <http://doi.org/10.1016/j.celrep.2016.07.081> PMID: 27545886
13. Boehm, E., Zornoza, M., Jourdain, A., Delmiro Magdalena, A., Garcia-Consuegra, I., Torres Merino, R., ... Simarro, M. (2016). **Role of FAST kinase domains 3 (FASTKD3) in post-transcriptional regulation of mitochondrial gene expression.** *The Journal of Biological Chemistry*, 3(1), jbc.M116.730291. <http://doi.org/10.1074/jbc.M116.730291> PMID: 27789713
14. Fiuza-Luces, C., Díez-Bermejo, J., Fernández-De La Torre, M., Rodríguez-Romo, G., Sanz-Ayán, P., Delmiro, A., ... Morán, M. (2018). **Health Benefits of an Innovative Exercise Program for Mitochondrial Disorders.** *Medicine and Science in Sports and Exercise*, 50(6). <http://doi.org/10.1249/MSS.0000000000001546> PMID:29315169
15. García Del Río, A., Delmiro, A., Martín, M. A., Cantalapiedra, R., Carretero, R., Durántez, C., ... Simarro, M. (2018). **The Mitochondrial Isoform of FASTK Modulates Nonopsonic Phagocytosis of Bacteria by Macrophages via Regulation of Respiratory Complex I.** *Journal of Immunology* (Baltimore, Md. : 1950), 201(10), 2977–2985. <https://doi.org/10.4049/jimmunol.17010>