



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



**Universitat Autònoma
de Barcelona**

Unitat d'Antropologia Biològica

Departament de Biologia Animal, Biologia Vegetal i Ecologia

Doctorat en Biodiversitat

**ANÁLISIS DE LA VARIABILIDAD DEL DNA
MITOCONDRIAL MEDIANTE LA
SECUENCIACIÓN MASIVA: APLICACIÓN
AL ESTUDIO DEL GLIOMA**

María del Mar González García

Tesis Doctoral

2019

Directoras:

Dra. Maria Pilar Aluja Paris

Dra. Cristina Pereira Santos

Tutora:

Dra. Maria Pilar Aluja Paris

Trabajo financiado por el Ministerio de Ciencia e Innovación
Proyecto: CGL2009-08205

Análisis de la variabilidad del DNA mitocondrial mediante la secuenciación masiva: aplicación al estudio del glioma

Memoria presentada por María del Mar González García para optar al título de Doctora en Biodiversidad del Departament de Biologia Animal, Biologia Vegetal i Ecologia de la Universitat Autònoma de Barcelona.

Directoras:

- Dra. Maria Pilar Aluja Paris, Profesora Titular de la Unitat d'Antropologia Biològica del Departament de Biologia Animal, Biologia Vegetal i Ecologia de la Universitat Autònoma de Barcelona.
- Dra. Cristina Maria Pereira Dos Santos, Profesora Agregada de la Unitat d'Antropologia Biològica del Departament de Biologia Animal, Biologia Vegetal i Ecologia de la Universitat Autònoma de Barcelona.

Tutora:

- Dra. Maria Pilar Aluja Paris, Profesora Titular de la Unitat d'Antropologia Biològica del Departament de Biologia Animal, Biologia Vegetal i Ecologia de la Universitat Autònoma de Barcelona.

Dra. Maria Pilar Aluja Paris

Dra. Cristina Pereira Santos

María del Mar González García

“Eso me temo... has perdido la cabeza, estás completamente loco. Pero te diré un secreto: las mejores personas lo están”.

El Sombrero Loco

Alicia en el País de las Maravillas



AGRADECIMIENTOS

En primer lugar, me gustaría agradecer a mis directoras, Cristina y Pilar, por darme la oportunidad de realizar una tesis doctoral y por enseñarme y estar a mi lado durante todos estos años. Por ayudarme a crecer y madurar a nivel profesional y por confiar en mí siempre.

Al Dr. Juan José Acebes que, siendo Jefe de Servicio de la Unidad de Neurocirugía de la Ciudad Sanitaria y Universitaria de Bellvitge, nos facilitó el muestreo de forma desinteresada, siendo receptivo en todo momento a cualquier necesidad para que nuestro proyecto pudiera realizarse. Por valorar la importancia y necesidad del estudio y, sobre todo, por confiar en nosotros.

Al Dr. Carlos Alarcón, el artífice de la recogida de muestras, por estar siempre atento y dispuesto a ayudarnos en todo momento.

A la Dra. Mònica Cos, por toda su ayuda y sus consejos sobre neuroimagen.

A todos los miembros del Servicio de Neurocirugía que, con una paciencia infinita, soportaban nuestra presencia en el quirófano en unas largas y complejas intervenciones.

Quisiera manifestar un agradecimiento especial a todos los pacientes y sus familiares, que consintieron nuestro acceso a las muestras en un momento especialmente difícil para ellos, particularmente por el tipo de recogida, de enfermedad y de pronóstico que suele acompañar a la misma.

A la Dra. Garcia-Barceló y todo su equipo, por darme la oportunidad de hacer una estancia de 3 meses en The University of Hong Kong y por ayudarme a adquirir experiencia en el ámbito de la bioinformática. Por el fantástico trato recibido y por hacer que la estancia en Hong Kong fuera una de las mejores experiencias de mi vida.

A mis padres, mi hermano y mi cuñi, por todo su apoyo desde el inicio, por estar ahí en los momentos en que más cuesta arriba se me hacía el camino y por animarme siempre a seguir adelante en todo. Y a toda mi familia, ¡la mejor!

A Amanda, por todo. Por enseñarme y darme ánimos cuando más lo necesitaba, por aguantarme y por darme consejos. Por esos cafés para charlar de todo. Y por compartir todo tipo de momentos y entenderme. ¡Por ser tan “Dori” como yo!

A “la Mery”, porque juntas empezamos esta larga etapa. Por el gran apoyo durante tantos años y por la compañía que me hacía en los buenos y malos momentos.

A Gems, Nika y Pere. Por los ratos tan divertidos que hemos pasado tanto dentro como fuera de la universidad. Por todas las anécdotas vividas y por los dolores de barriga y lágrimas derramadas de las risas que nos hemos echado juntos. ¡Por todas nuestras frases célebres!

A Aida (mi descubrimiento), Diana (my girl) y Maria (¿qué hubiera sido de Cenicienta sin Gus?). Por el vínculo tan especial que hemos creado y la complicidad que hemos llegado a tener. Por todas nuestras locuras... ¡Y las que quedan!

A “la Txessy” y Núria, por aguantar todas mis payasadas, y por los buenos ratos de charla y risas. Por estar pendientes siempre de mí y ayudarme en todo lo que necesitara. ¡Y por alimentarme tan bien siempre!

A Blanca, ¡por toda su ayuda, colaboración y dedicación para que la tesis quedara genial!

A las Palcamitas, a las que conozco desde hace muchísimos años. Por preocuparse y estar siempre pendientes de mí.

Al mundo del básquet, por todos los buenos momentos. Por ayudarme a despejar mi mente y a desconectar de todo.

Y a ti, Santi, porque entraste en mi vida en la última etapa de mi tesis, pero me diste la fuerza necesaria para afrontar la recta final y para estar anímicamente perfecta.

A todos... ¡GRACIAS!

RESUMEN

La presente tesis doctoral se ha centrado en el reto que significa el estudio de la variabilidad del DNA mitocondrial (DNAm) por secuenciación masiva y el impacto de dicha variabilidad en el glioma cerebral. Debido al papel significativo de la mitocondria en la producción de energía celular, se ha postulado que alteraciones en el DNAm pueden contribuir al desarrollo de determinadas enfermedades, especialmente en aquellos tejidos que requieren más energía, como el cerebral. Así pues, el DNAm se ha asociado con múltiples eventos relacionados con enfermedades neurodegenerativas y procesos carcinogénicos, y por esta razón, el estudio del glioma cerebral podría ser muy interesante en términos de análisis mitocondrial. Con los avances en la secuenciación del DNAm mediante la secuenciación masiva, es posible generar una cantidad de datos mayor que los obtenidos mediante la secuenciación tradicional y aumentar la sensibilidad de detección de mutaciones. Estas mejoras permiten realizar estudios sobre la variabilidad mitocondrial de forma más precisa y exhaustiva, particularmente en aquellos centrados en la detección de la heteroplasmia mitocondrial. Sin embargo, la detección de la heteroplasmia a bajas frecuencias no es un proceso sencillo debido a la gran dificultad para discriminar correctamente la presencia de falsos positivos y negativos, sin que de momento exista un criterio único que determine cuál es la mejor metodología para establecer un límite de detección fiable.

Considerando estos aspectos, en la presente tesis doctoral se han planteado 2 objetivos principales: 1) Evaluar la capacidad de detección y reproducibilidad de la *Next Generation Sequencing* (NGS), utilizando la metodología de Nextera XT® y la plataforma MiSeq (Illumina), en la detección fiable de la variabilidad del genoma mitocondrial tanto en homoplasmia como en heteroplasmia; y 2) Valorar la implicación del genoma mitocondrial en el desarrollo de gliomas cerebrales de diferentes grados de malignidad.

Tanto la metodología como los resultados y discusión se han organizado en 4 capítulos, a fin de poder responder a los diferentes objetivos planteados. En el primer capítulo se han descrito todos los aspectos relacionados con el material y la metodología empleada en el presente trabajo. Se ha detallado el proceso de obtención de las

muestras empleadas en los diferentes estudios presentados y el procedimiento llevado a cabo en el laboratorio hasta la secuenciación del DNAm. Asimismo, se ha detallado todo el procedimiento bioinformático aplicado a los datos obtenidos mediante la NGS y se han descrito los análisis realizados para evaluar la variabilidad mitocondrial detectada. En el segundo capítulo se han presentado los resultados y la discusión en relación al análisis de la calidad de los datos generados a partir de la secuenciación de las muestras de DNAm. Se han analizado todos los parámetros relacionados con la calidad de los archivos Fastq obtenidos de NGS y se han valorado los sesgos generados debido al uso de la metodología Nextera XT® y la plataforma MiSeq (Illumina). En el tercer capítulo se han descrito los resultados y la discusión en relación a la determinación de la fiabilidad y reproducibilidad de los resultados obtenidos en el análisis del DNAm, a partir de la generación de librerías de NGS independientes. Además, se ha establecido una nueva estrategia para determinar los límites de detección de la heteroplasma mitocondrial mediante NGS. Por último, en el cuarto capítulo se han presentado los resultados y la discusión basados en el análisis de la variabilidad mitocondrial de individuos afectados de glioma, considerando una muestra tumoral y una muestra de sangre periférica. Para cada individuo, se ha establecido su haplotipo y haplogrupo a partir de las mutaciones detectadas, tanto en homoplasma como en heteroplasma, y se han clasificado las mutaciones en somáticas o germinales. Para las mutaciones en homoplasma, se ha analizado su distribución a lo largo del genoma mitocondrial y se han determinado las mutaciones sobrerrepresentadas respecto a la población general, las localizadas en posiciones estables y las que tenían cierto impacto a nivel patogénico. Además, se ha examinado la carga mutacional de cada individuo. Para las heteroplasmas, se ha determinado si existen diferencias entre la proporción de individuos heteroplásmicos afectados de glioma con individuos heteroplásmicos que no padecieron esta afectación. Además, también se ha analizado su distribución a lo largo del genoma mitocondrial y se ha determinado las mutaciones localizadas en posiciones estables, así como las que tenían cierto impacto a nivel patogénico.

El presente trabajo ha permitido ampliar los conocimientos respecto a la detección de la heteroplasma mitocondrial mediante la tecnología de NGS y el papel del DNAm en el glioma cerebral. A nivel de la detección de la heteroplasma, se ha presentado una nueva estrategia para establecer sus límites de detección, intentando reducir errores en la interpretación de los resultados obtenidos. Respecto al estudio del glioma, este trabajo ha sido el primero en establecer una asociación de los haplogrupos mitocondriales J y T y el glioma cerebral y ha inferido en la probabilidad de que los tumores puedan progresar a grados de mayor malignidad según el haplogrupo. Asimismo, los resultados obtenidos respecto a la baja cantidad tanto de mutaciones sobrerrepresentadas en el estudio como de mutaciones en posiciones estables que pueden tener cierto impacto a nivel patogénico, y la baja carga mutacional detectada en los individuos afectados de glioma, han sugerido una preferencia por parte del glioma de no acumular mutaciones en el DNAm, con el fin de no generar una gran cantidad de especies reactivas de oxígeno (ROS) en el ambiente y poder progresar a estadios de mayor malignidad tumoral.

ABSTRACT

This PhD thesis was focused on the challenge of studying the variability of mitochondrial DNA (DNAm_t) by massive parallel sequencing and the impact of such variability on the glioma. Due to the significant role of mitochondria in cellular energy production, it has been postulated that mtDNA alterations might contribute to the development of certain diseases, particularly in those tissues that require more energy, such as the brain. Thus, mtDNA has been associated with multiple events related to neurodegenerative diseases and carcinogenic processes, and for this reason, the study of glioma might be very interesting in terms of mitochondrial analysis. The advances in mtDNA sequencing with massive parallel sequencing enable to generate a larger amount of data than with traditional sequencing and to increase the sensitivity of mutation detection. These improvements allow studies based on mitochondrial variability to be carried out more precisely and thoroughly, particularly those focused on mitochondrial heteroplasmy detection. However, heteroplasmy detection at low frequencies is not a straightforward process due to the great difficulty in discriminating the presence of false positives and negatives correctly, and so far, there is not specific criteria that determines the best methodology to establish a reliable detection limit.

Considering these aspects, two main goals were proposed in this doctoral thesis: 1) To evaluate the detection and reproducibility capacity of Next Generation Sequencing (NGS), using the Nextera XT[®] methodology and the MiSeq platform (Illumina), in detecting reliably mitochondrial genome variability in both homoplasmy and heteroplasmy; 2) To assess the involvement of mitochondrial genome in the development of gliomas with different malignancy degrees.

Both the methodology and the results and discussion were organised in 4 chapters in order to respond to the different aims. In the first chapter, the material and the methodology used in this work were described. Sample collection for the different analyses and the laboratory procedure to finally achieve mtDNA sequencing were detailed. Likewise, the bioinformatics procedure applied to the data obtained

through the NGS and the analysis to evaluate the detected mitochondrial variability were described. In the second chapter, the results and discussion regarding the analysis of the quality of the data generated from the sequencing of the mtDNA samples were reported. All the parameters based on the quality of Fastq files obtained from NGS were analysed and the biases generated as a result of using the Nextera XT[®] methodology and the MiSeq platform (Illumina) were assessed. In the third chapter, the results and discussion of the reliability and reproducibility of the results obtained in mtDNA analysis by generating independent NGS libraries were reported. Moreover, a new strategy was established to define heteroplasmy detection limits using NGS. Finally, in the fourth chapter the results and discussion of the analysis of mitochondrial variability in individuals affected by glioma, using tumour and peripheral blood samples, were described. For each individual, both haplotype and haplogroup were established based on homoplasmy and heteroplasmy mutations, and mutations were classified as somatic or germinal. For homoplasmic mutations, their distribution along the mitochondrial genome was analysed, and the mutations that were overrepresented as compared to the general population, those that were located in stable positions and those that had a certain pathogenic impact were detected. In addition, the mutational load of each individual was examined. For heteroplasmic mutations, differences between the proportion of heteroplasmic individuals affected by glioma and heteroplasmic individuals who did not suffer from this condition were determined. In addition, their distribution along the mitochondrial genome was also analysed and mutations located in stable positions were detected, as well as those that might have some impact at the pathogenic level.

This work has allowed us to expand our knowledge regarding to mitochondrial heteroplasmy detection using NGS and the role of mtDNA in glioma. Concerning mitochondrial heteroplasmy detection, a new strategy was presented to establish its detection limits, in an attempt to reduce misinterpretations of results. Regarding the study of glioma, the present work was the first to establish an association between mitochondrial haplogroups J and T and glioma, and it inferred the likelihood of tumours to progress to higher levels of malignancy according to the haplogroup. Likewise, the low amount of both mutations overrepresented in the study and mutations in stable positions that might have some impact at the pathogenic level, and the low mutational load detected in individuals affected by glioma, suggested that glioma tends to not accumulate mtDNA mutations to avoid generating a large amount of reactive oxygen species (ROS) in the environment and to progress to stages of higher levels of malignancy.

ÍNDICE

AGRADECIMIENTOS.....	ix
RESUMEN	xi
ABSTRACT	xiii
1. INTRODUCCIÓN	5
1.1. LA MITOCONDRIA.....	7
1.1.1. Origen y funciones de la mitocondria	7
1.1.2. El DNA mitocondrial	8
Organización y características del DNAMt.....	8
Mutaciones mitocondriales y heteroplasmia	12
Variación del DNAMt	15
1.2. DNA MITOCONDRIAL y CÁNCER	17
1.2.1. Mutaciones mitocondriales en cáncer	22
1.2.2. El glioma	26
Antecedentes del estudio del DNAMt en glioma	28
1.3. METODOLOGÍAS PARA LA SECUENCIACIÓN DEL DNA MITOCONDRIAL	37
1.3.1. Secuenciación de Sanger	37
1.3.2. Next Generation Sequencing	37
La tecnología Illumina	39
1.3.3. Detección de mutaciones en el DNA mitocondrial	42
2. HIPÓTESIS del TRABAJO y OBJETIVOS	45
2.1. PLANTEAMIENTO del TRABAJO e HIPÓTESIS	47
2.2. OBJETIVOS del TRABAJO	49

3. CAPÍTULO I Material y Métodos	51
3.1. SELECCIÓN de las MUESTRAS	55
3.1.1. Formación de quimeras a partir de muestras de sangre de diferentes individuos (Figura 5; Fase a.1)	55
3.1.2. Recogida de muestras tumorales y sanguíneas de pacientes afectos de glioma (Figura 5; Fase a.2)	56
3.1.3. Muestras de cerebro de individuos sin patología (Figura 5; Fase a.2)	57
3.2. EXTRACCIÓN y AMPLIFICACIÓN del DNAMt	57
3.2.1. Quimeras a partir de muestras de sangre de diferentes individuos (Figura 5; Fase a.1)	57
3.2.2. Tumores y sangre de pacientes afectos de glioma y muestras de cerebro de individuos sin patología (Figura 5; Fase a.2)	59
3.3. PURIFICACIÓN, CUANTIFICACIÓN y EQUIMOLARIDAD del DNAMt (Figura 5; Fase b)	60
3.4. PREPARACIÓN de las LIBRERÍAS y posterior SECUENCIACIÓN de NGS (Figura 5; Fase c).....	61
3.5. ANÁLISIS BIOINFORMÁTICO del DNAMt (Figura 4; Fase c)	62
3.5.1. Calidad de los archivos Fastq	62
3.5.2. Análisis de Fastq con mtDNA-Server	63
3.6. ANÁLISIS de VARIANTES DETECTADAS en las QUIMERAS (Figura 5; Fase d.1)	65
3.6.1. Secuenciación de Sanger vs. Secuenciación de NGS	65
3.6.2. Análisis de las mezclas de bases a partir de la posible co-amplificación entre el DNAMt y el NUMT.....	65
3.6.3. Alineamiento <i>BWA-mem</i> vs. Alineamiento <i>Bowtie2</i>	66
3.7. ANÁLISIS de VARIANTES MITOCONDRIALES DETECTADAS en PACIENTES AFECTOS de GLIOMA (Figura 5; Fase d.2).....	67
3.7.1. Variación del DNAMt en individuos afectos de glioma	67
3.7.2. Distribución de haplogrupos mitocondriales en pacientes afectos de glioma ...68 Compilación de la distribución de haplogrupos mitocondriales en pacientes afectos de glioma de poblaciones europeas	68
3.7.3. Análisis de mutaciones en homoplasmia	69
Distribución de las mutaciones en homoplasmia a lo largo del genoma mitocondrial	69
Análisis de la carga mutacional	70

Mutaciones en homoplasmia sobrerrepresentadas en glioma respecto a la población general	71
Mutaciones en homoplasmia detectadas en posiciones estables	71
3.7.4. Análisis de mutaciones en heteroplasmia	73

4. CAPÍTULO II RESULTADOS y DISCUSIÓN 1..... 75

VALORACIÓN de la CALIDAD de los archivos FASTQ

4.1. RESULTADOS	77
4.1.1. Calidad de los archivos Fastq en las quimeras generadas a partir de muestras de sangre de diferentes individuos	77
4.1.2. Calidad de los archivos Fastq en las muestras tumorales y sanguíneas de individuos afectados de glioma y controles de cerebro de individuos sanos	83
Patrón de <i>coverage</i> de las muestras	88
4.2. DISCUSIÓN	89

5. CAPÍTULO III RESULTADOS y DISCUSIÓN 2..... 93

DETECCIÓN de HETEROPLASMIAS MITOCONDRIALES a través de NGS

5.1. RESULTADOS	95
5.1.1. Profundidad de lectura promedio en cada <i>run</i>	95
5.1.2. Secuenciación de Sanger vs. Secuenciación por NGS	96
Alineamiento BWA-mem vs. Alineamiento Bowtie2 en NGS	96
5.1.3. Secuenciación de la región que puede presentar co-amplificación entre el DNAmT y el NUMT	100
Detección de co-amplificación entre el DNAmT y el NUMT	100
Mezclas de bases en posiciones no idénticas entre el DNAmT y el NUMT	101
Mezclas de bases en posiciones idénticas entre el DNAmT y el NUMT	105
Alineamiento BWA-mem vs. Alineamiento Bowtie2 entre el DNAmT y el NUMT	106
Establecimiento de los límites de detección de la heteroplasmia mitocondrial	106
5.2. DISCUSIÓN	108

6. CAPÍTULO IV RESULTADOS y DISCUSIÓN 3113

ANÁLISIS del DNAmT en INDIVIDUOS AFECTOS de GLIOMA

6.1. RESULTADOS	115
------------------------------	------------

6.1.1. Variaciones del DNAMt en individuos afectados de glioma	115
6.1.2. Distribución de los haplogrupos mitocondriales de las muestras del estudio. 124	
Recopilación de la distribución de haplogrupos mitocondriales en pacientes afectos de glioma	127
6.1.3. Análisis de mutaciones en homoplasmia	128
Distribución de las mutaciones en homoplasmia a lo largo de genoma mitocondrial	128
Análisis de la carga mutacional	129
Mutaciones en homoplasmia sobrerrepresentadas en pacientes afectados de glioma respecto a la población general	130
Mutaciones en homoplasmia detectadas en posiciones estables	134
6.1.4. Análisis de mutaciones en heteroplasmia	142
Distribución de las mutaciones en heteroplasmia a lo largo de genoma mitocondrial	143
Mutaciones en heteroplasmia detectadas en posiciones estables	144
6.2. DISCUSIÓN	149
6.2.1. Asociación de haplogrupos mitocondriales y el riesgo de sufrir glioma	149
6.2.2. Distribución de las mutaciones a lo largo del genoma mitocondrial	151
6.2.3. Mutaciones detectadas en pacientes afectados de glioma	153
6.2.4. Carga mutacional	159
6.2.5. Implicaciones del DNAMt en el glioma cerebral	159
7. CONSIDERACIONES FINALES.....	163
8. CONCLUSIONES	169
9. BIBLIOGRAFÍA	175
10. ANEXOS.....	213
10.1. ANEXO I.....	215
10.2. ANEXO II.....	217
10.3. ANEXO III.....	228

1

INTRODUCCIÓN

1. INTRODUCCIÓN

1.1. LA MITOCONDRIA

La mitocondria es un orgánulo citoplasmático intracelular, presente en casi todas las células eucariotas y altamente especializado, que se encarga de producir energía para las células (Chinnery and Hudson 2013). Su estructura consta de una doble membrana: la membrana externa, que separa la mitocondria del citosol, y la interna, estructura que alberga principalmente las enzimas responsables de la respiración celular y que presenta invaginaciones en forma de crestas que sobresalen y delimitan la matriz mitocondrial (Taanman 1999). Gracias a la presencia de estas invaginaciones, la matriz puede aumentar su superficie y permitir una mayor presencia de dichas enzimas (Carew and Huang 2002).

1.1.1. Origen y funciones de la mitocondria

Aunque el origen de la mitocondria permanece todavía incierto, la teoría endosimbiótica presentada por Lynn Margulis en 1967 es la más extendida en la actualidad. Según esta teoría, la mitocondria tiene su origen en un organismo procariota que fue fagocitado por una célula eucariota, estableciendo así una relación simbiogenética entre una célula proto-eucariota glicolítica y una bacteria oxidativa (Carew and Huang 2002; Gray and Doolittle 1982; Schwartz and Dayhoff 1978). De esta forma, la mitocondria se ha convertido en la única estructura no nuclear en mamíferos que posee su propio material genético: el DNA mitocondrial (DNAMt) (Basso et al. 2007).

La mitocondria se caracteriza por su elevada eficiencia para utilizar oxígeno u otros sustratos como glucosa o piruvato y producir energía celular en forma de adenosín trifosfato (ATP) (Mattson et al. 2008). En las células eucariotas, la generación de energía se inicia en la matriz mitocondrial a través del ciclo de Krebs y continúa en la cadena de transporte de electrones, donde se lleva a cabo la fosforilación oxidativa (OXPHOS) (Van Houten et al. 2006). La cadena de transporte de electrones

consta de 5 complejos enzimáticos – complejos I, II, III, IV, V – formados por 74 subunidades codificadas por genes nucleares y 13 subunidades codificadas por genes mitocondriales (Penta et al. 2001).

Aunque la principal función de la mitocondria es la de la producción de energía durante la respiración celular, este orgánulo también desarrolla otras funciones: 1) actúa como mediador del metabolismo para dar soporte a diferentes funciones celulares como la descarboxilación oxidativa del piruvato, el ciclo del ácido tricarboxílico y la oxidación de ácidos grasos; 2) regula la homeostasis de iones como el calcio o el fosfato; 3) sirve de soporte para la síntesis de lípidos, aminoácidos y nucleótidos; 4) puede influir en la circulación de fármacos citotóxicos; y 5) tiene un papel importante en los procesos de proliferación celular y muerte celular programada (apoptosis) (Allegra et al. 2006; Modica-Napolitano and Singh 2004).

1.1.2. El DNA mitocondrial

Organización y características del DNAm

El DNAm es una molécula circular de doble cadena, una ligera rica en pirimidinas y una pesada rica en purinas, que representa entre el 0,1% y el 2% del total de DNA en el caso de las células de mamífero (Druzhyňa et al. 2008). Se localiza en la matriz mitocondrial, lugar donde se hallan también una gran cantidad de enzimas y ribosomas para la síntesis de proteínas. En el caso del genoma mitocondrial humano, la molécula contiene 16.569 pares de bases (pb), siendo uno de los genomas mitocondriales más pequeños encontrados en eucariotas (Penta et al. 2001).

En el DNAm se pueden diferenciar 2 regiones (Figura 1; Tabla 1 del Anexo I):

- *Región codificante*. Codifica un total de 37 genes desprovistos de intrones, distribuidos de forma continua a lo largo del genoma mitocondrial, que corresponden a 22 RNAs de transferencia (RNAt) encargados de sintetizar proteínas mitocondriales en la matriz mitocondrial, 2 subunidades –12S y 16S– de RNA ribosómico (RNAr) y 13 polipéptidos que forman parte de los complejos enzimáticos del sistema OXPHOS:
 - 7 subunidades del Complejo I (NADH deshidrogenasa): *ND1, ND2, ND3, ND4, ND4L, ND5, ND6*.
 - 1 subunidad del Complejo III (citocromo b): *CYB*.
 - 3 subunidades del Complejo IV (citocromo c oxidasa): *CO1, CO2 y CO3*.
 - 2 subunidades del Complejo V (ATP sintasa): *ATP6 y ATP8*.

En la cadena pesada es donde se codifica una mayor cantidad de información –2 RNAr, 14 RNAt y 12 polipéptidos–, mientras que la cadena ligera contiene la información genética para únicamente 1 polipéptido y 8 RNAt (Singh et al. 2014).

Las 87 subunidades polipeptídicas restantes de los complejos I, III, IV y V, junto con la totalidad del Complejo II (succinato deshidrogenasa), y 1500 proteínas que tienen su función en la mitocondria, están codificadas por genes nucleares, sintetizadas en el citosol e importadas al interior de las mitocondrias mediante un sistema de importación de proteínas (Druzhyina et al. 2008).

- *Región no codificante.* Dentro de esta región, el DNAm cuenta con una región control de 1,1kb donde se localiza la región denominada D-loop que forma una estructura que separa ambas cadenas (Basso et al. 2007). El D-loop consta de 3 regiones hipervariables (HVR) –HVRI (posiciones 16024-16365); HVRII (posiciones 73-340); y HVRIII (posiciones 438-574)– y contiene el origen de replicación para la síntesis de la cadena H, así como los promotores para la transcripción mitocondrial (Lutz et al. 1997; Penta et al. 2001; Taanman 1999). Además, en dicha región se hallan 3 bloques de secuencias conservadas (CSB I, II y III), así como la secuencia asociada a terminación (TAS) prematura del ciclo replicativo y la 7S DNA (Falkenberg 2018).

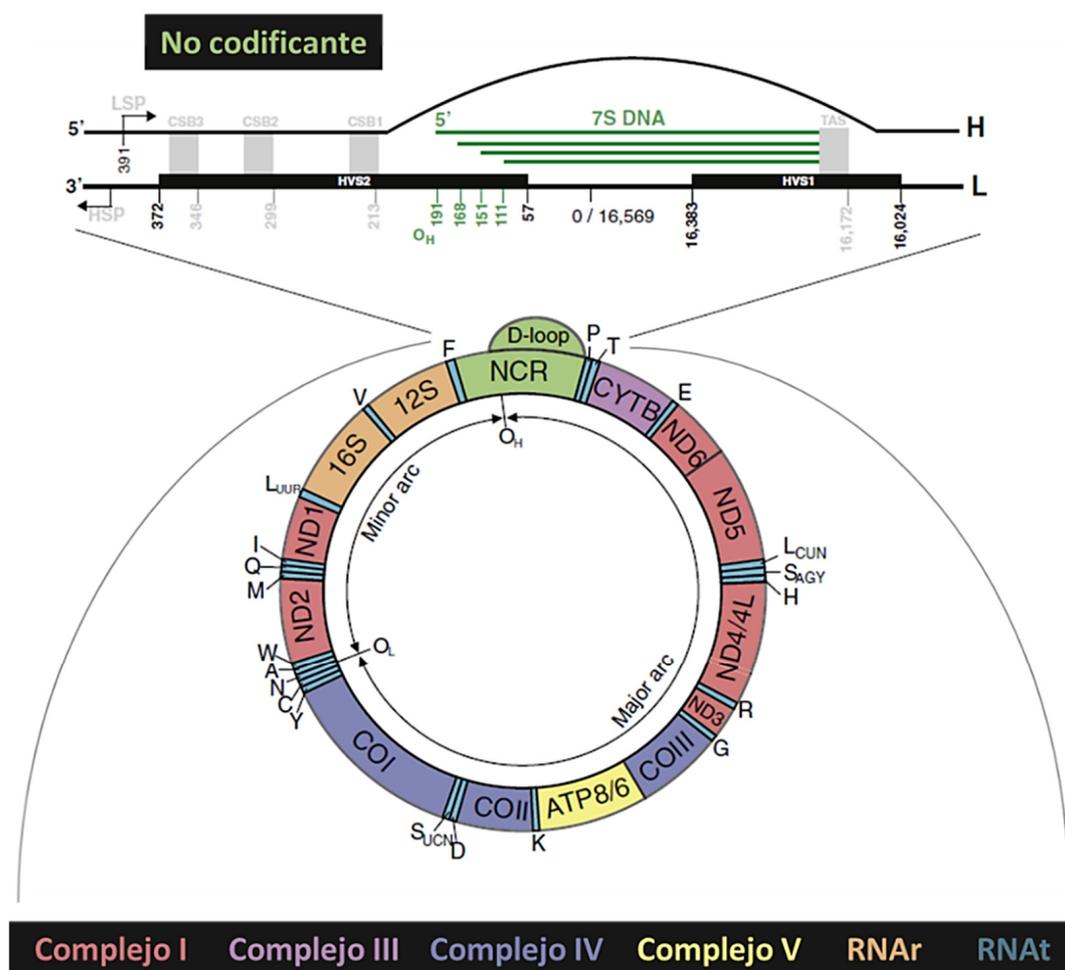


Figura 1. Esquema del genoma mitocondrial. Se puede diferenciar la región codificante y no codificante. Dentro de la región codificante se puede observar la distribución de los genes mitocondriales. Imagen modificada de Nicholls and Minczuk (2014).

Tanto la replicación como la transcripción se producen de forma semiautónoma en la mitocondria (Druzhyina et al. 2008) y ambos procesos están estrechamente controlados y coordinados con la síntesis de un gran número de proteínas mitocondriales que están codificadas por el DNA nuclear (DNAn) (Barbieri and Scorrano 2013), siendo la interacción y coordinación entre el genoma nuclear y mitocondrial necesaria para la síntesis de proteínas y funciones mitocondriales (Penta et al. 2001). Del mismo modo, recientemente, se han identificado proteínas –como MOTS-c y humaninas, entre otras– que están codificadas por el DNAm cuya función aún está en estudio (Lee et al. 2015; Okada et al. 2017). Así, aunque el DNAm representa un porcentaje muy bajo del DNA celular total, los genes codificados en la mitocondria son esenciales para el funcionamiento normal de la célula (Modica-Napolitano and Singh 2004).

Existe evidencia de regulación epigenética del DNAm en mamíferos desde 1984. Sin embargo, su hallazgo permaneció controvertido hasta el 2011 cuando se demostró en varios estudios la translocación de la DNA (citosina-5)-metiltransferasa (DNMT1) desde el núcleo hacia la mitocondria (Yu et al. 2018; Zhang et al. 2010). Debido a que el DNAm se encuentra empaquetado en nucleoides –compuestos por proteínas entre las que se encuentra el factor de transcripción mitocondrial A (TFAM)–, las alteraciones producidas en el contenido de TFAM pueden determinar cambios en el mitocromosoma y exponer el genoma mitocondrial a la acción de las DNMTs, que regularían la expresión de los genes mitocondriales a través de la metilación de residuos de citosina (Akouchekian et al. 2009).

Además, el DNAm tiene ciertas propiedades que son únicas de este genoma y que, por lo tanto, lo convierten en un buen candidato para estudios poblacionales, evolutivos, forenses y biomédicos:

- *Elevado número de copias.* Las células humanas pueden contener cientos o incluso miles de orgánulos mitocondriales (Witas and Zawicki 2004) y cada mitocondria puede albergar entre 2 y 10 copias de DNAm (Basso et al. 2007). El elevado número de copias, juntamente con el hecho de que se localice fuera del núcleo, permite que el DNAm sea relativamente fácil de recuperar. Por estos motivos, ha sido el marcador elegido prioritariamente en estudios de DNA antiguo y en aplicaciones forenses.
- *Herencia materna.* Aunque la transmisión del DNAm por vía paterna en algunos animales es común y recurrente, la herencia del DNAm en la especie humana es únicamente materna (Chinnery and Hudson 2013). Sin embargo, en el 2002 se hizo público un caso donde el DNAm del tejido muscular de un hombre, con una intolerancia severa al ejercicio, tenía contribución tanto materna como paterna (Schwartz and Vissing 2002) y, posteriormente, se encontró

recombinación entre el DNAm_t materno y paterno (Kraytsberg et al. 2004). Ante este hallazgo, varios estudios se centraron en la posible transmisión del DNAm_t por vía paterna, tanto en pacientes con enfermedades mitocondriales (Filosto et al. 2003; Schwartz and Vissing 2002; Taylor et al. 2003) como en individuos aparentemente sanos (Pyle et al. 2015; Santos et al. 2008a; Santos et al. 2005), y en ningún caso se detectó contribución paterna de DNAm_t humano. Sin embargo, recientemente se han descrito otros casos en los que se informa sobre la transmisión biparental (Kurelac et al. 2013; Luo et al. 2018). Existen varios factores, no mutuamente excluyentes, que podrían explicar la ausencia de contribución de DNAm_t de origen paterno: 1) El proceso de ubiquitinización que sufre el DNAm_t de origen paterno en el oocito, ya que durante la espermatogénesis, las mitocondrias del espermatozoide quedan marcadas por una ubiquitina, de forma que cuando el espermatozoide entra dentro del oocito los proteosomas degradan dichas mitocondrias marcadas (Zimmerman and Sutovsky 2009); 2) La baja representación de las copias de DNAm_t de origen paterno debido a que el espermatozoide contiene alrededor de un centenar de copias de DNAm_t en comparación a las aproximadamente 100.000 copias que contiene el oocito, de manera que el efecto de cuello de botella del DNAm_t provocaría la exclusión de los alelos paternos (Chinnery and Hudson 2013).

- *Recombinación.* El hecho de que el DNAm_t se herede únicamente por línea materna en la mayoría de animales supone que no necesita llevar a cabo ningún proceso de recombinación, ya que el papel principal de dicho proceso sería la mezcla de la información genética de diferentes linajes con el fin de obtener una mayor diversidad y la eliminación de posibles variantes deletéreas, aportando así una ventaja selectiva en el proceso de evolución (Chen 2013). Sin embargo, el descubrimiento comentado anteriormente sobre el proceso de recombinación en el caso de transmisión paterna cuestiona el paradigma de que el DNAm_t no recombina. Aunque la contribución paterna de DNAm_t a nivel poblacional parece no ser importante y, por lo tanto, no es de esperar que tenga un impacto evolutivo, también se han postulado otras rutas para el posible fenómeno de recombinación: 1) la posible recombinación entre diferentes moléculas de DNAm_t originado por nuevas mutaciones presentes en un mismo individuo, dando lugar a diferentes mitotipos recombinantes, tanto en individuos sanos como con patología (Zsurka et al. 2005); y 2) la posible recombinación entre el DNAm_t y las inserciones nucleares de origen mitocondrial (NUMTs) (Eyre-Walker and Awadalla 2001; Morris and Lightowlers 2000).
- *Tasa de mutación.* El DNAm_t se caracteriza por tener una tasa de mutación entre 10-100 veces mayor respecto a la tasa de mutación del DNAn (Tan et al. 2008). Esta elevada tasa de mutación podría relacionarse con las especies reactivas de oxígeno (ROS) producidas durante el OXPHOS (Wallace 2005). Además, como se ha comentado, a diferencia del DNAn, el DNAm_t no está asociado

a histonas que podrían protegerlo (Gray 1999; Shigenaga et al. 1994), aunque se ha postulado que los nucleoides formados a partir de la unión de la proteína TFAM al DNAm podrían actuar como elementos de protección (Rebelo et al. 2011). Asimismo, la replicación del DNAm se da de forma más rápida que la replicación del DNA (Tan et al. 2008) y, aunque el DNAm tiene sus propios sistemas de reparación, éstos son menos eficientes que los que se encuentran en el DNA y, por lo tanto, el daño provocado en el DNAm es más difícil de subsanar. Como consecuencia de todos estos factores, la elevada tasa de mutación del DNAm hace que este genoma tenga una tasa de evolución mucho más alta que la del genoma nuclear y sea altamente polimórfico. Por último, a nivel somático, se pueden producir mutaciones a lo largo de la vida, éstas pueden bloquear la RNA polimerasa mitocondrial (POLRMT) e impedir la transcripción del DNAm. Esto puede provocar que las proteínas de la cadena de transporte de electrones no se puedan codificar y provoque una propagación de ROS y una oxidación del DNAm (Basso et al. 2007). Ante esta situación, las mitocondrias que tienen su DNAm muy dañado tienden a fusionarse con otras mitocondrias para juntar las copias de DNAm de ambas mitocondrias con el objetivo de que la cantidad de copias de DNAm no dañadas sea mayor que la de copias de DNAm dañadas (Youle and van der Bliek 2012). De este modo, si una mitocondria ha sufrido mutaciones en su material genético o ha acumulado demasiado ROS, al fusionarse con otra se pueden intentar corregir los problemas. Cuando el proceso de fusión no consigue recuperar el buen funcionamiento de la célula, es necesario eliminar la mitocondria dañada. Para llevarlo a cabo, la mitocondria es fisionada en fragmentos más pequeños y eliminados seguidamente por un proceso de autofagia mitocondrial denominada mitofagocitosis o mitofagia (Kornblihtt et al. 2016; Martorell 2014).

Mutaciones mitocondriales y heteroplasmia

La presencia de múltiples copias de DNAm por célula puede generar que, debido a la ocurrencia de nuevas mutaciones, no todas las copias sean idénticas y que, por lo tanto, puedan coexistir diferentes genotipos mitocondriales dentro de una misma mitocondria, célula, tejido o individuo, fenómeno que se denomina heteroplasmia (Coller et al. 2001; He et al. 2010). Cuando surge una nueva mutación en el DNAm, ésta se puede ir estableciendo dentro de ciertas mitocondrias, células y tejidos. A medida que se va produciendo la segregación replicativa –proceso por el cual las células se van replicando y las mitocondrias se reparten de forma aleatoria dentro de las 2 células hijas–, el porcentaje de DNAm mutado puede ir variando y aumentar hasta llegar a alcanzar el estado de homoplasmia (mutación fijada), es decir, un estado donde todas las células contienen la mutación

mitocondrial (Wallace and Chalkia 2013). Por lo tanto, la heteroplasmia podría considerarse el estadio intermedio en la fijación de una mutación en el genoma mitocondrial (Ramos et al. 2013). Así, las mutaciones mitocondriales observadas en un individuo se pueden clasificar en:

- *Mutaciones germinales*. Son aquellas que se heredan del DNAm_t materno y que, por lo tanto, están presentes en una etapa temprana del desarrollo embrionario, anterior a la diferenciación de los tejidos. Así, cuando este tipo de mutación se presenta en heteroplasmia se puede encontrar, *a priori*, en varios o todos los tejidos (Samuels et al. 2013).
- *Mutaciones somáticas*. Son aquellas que se dan en células o tejidos específicos y que se propagan a partir de la división celular posterior a la mutación (Samuels et al. 2013). Dependiendo del momento del desarrollo en que se producen pueden afectar a varios tejidos o sólo a un tejido concreto.

Dependiendo del tipo de mutación que se produce, se pueden generar heteroplasmas puntuales –cuando ciertas copias de DNAm_t sufren una sustitución nucleotídica– o heteroplasmas de longitud –cuando determinadas copias del DNAm_t sufren una inserción o una delección–.

Aunque en los primeros estudios llevados a cabo sobre la heteroplasmia mitocondrial, ésta se consideraba un fenómeno poco común y se correlacionaba con procesos patológicos y de envejecimiento (Corral-Debrinski et al. 1992; Cortopassi et al. 1992; Holt et al. 1989; Holt et al. 1988), se ha demostrado que la heteroplasmia, debida a la aparición de nuevas mutaciones, no es una condición excepcional (Santos et al. 2008b). Últimamente, algunos autores han considerado la heteroplasmia una condición universal (Payne et al. 2013) y estudios más recientes, basados en datos del proyecto de los 1000 Genomas (Diroma et al. 2014), apuntan a una frecuencia de individuos heteroplásmicos a nivel poblacional superior al 82%, de los que un 90% presentan más de una posición heteroplásmica.

Asimismo, el estudio de la heteroplasmia puntual en la población general indica que un porcentaje muy elevado de heteroplasmas (80%) son de origen germinal. Además, aunque la heteroplasmia de longitud ocurre con una mayor frecuencia en el DNAm_t y sobre todo en las regiones poli-citosina (*poly-C*) de la región control (Ramos et al. 2013; Santos et al. 2008b), estudios más recientes, basados tanto en población general como en el análisis de diferentes patologías como el cáncer, describen

principalmente mutaciones puntuales debido a la dificultad de discriminar heteroplasmas de longitud mediante las nuevas metodologías de secuenciación (Ghaffarpour et al. 2014).

En el caso de las heteroplasmas germinales, el porcentaje de DNAm mutado y no mutado puede diferir entre tejidos y, por lo tanto, dar lugar a diferencias a nivel bioenergético y fenotípico. Además, en estos casos no sólo se debe tener en cuenta la segregación replicativa, sino que también es importante valorar el hecho de que no todos los órganos tienen la misma sensibilidad frente a cambios mitocondriales, de manera que se podrían observar diferentes fenotipos en tejidos específicos (Wallace and Chalkia 2013). En este sentido el sistema nervioso, seguido por el sistema cardíaco, muscular, renal y endocrino, son los más sensibles frente a las alteraciones mitocondriales (Wallace 2005; Wallace and Chalkia 2013).

El impacto que puede generar la aparición de heteroplasmas a lo largo del genoma mitocondrial puede variar según su localización y según el tipo de mutación. Aunque la presencia de heteroplasma puede darse a lo largo de todo el genoma mitocondrial, es de esperar que ésta ocurra en posiciones con una mayor tasa de mutación, siendo la región control la que contiene una tasa más elevada (Zapico 2016). Sin embargo, algunos estudios han demostrado un exceso de mutaciones en heteroplasma en posiciones estables y la mayoría de éstas comprendían transiciones (Li et al. 2010; Ramos et al. 2013). Otros estudios han determinado que, dentro de la región codificante, la proporción de heteroplasmas no sinónimas (NS) es más elevada cuando se encuentran en baja frecuencia (Li et al. 2016). Todos estos resultados sugieren que pueden existir una gran cantidad de mutaciones a nivel heteroplásmico, pero que nunca se llegan a fijar dentro del genoma mitocondrial del individuo ya que podrían existir determinadas fuerzas evolutivas, como la deriva genética o la selección purificadora, que actuarían para poder evitar su fijación en los individuos (Ramos et al. 2013). Ahora bien, también se pueden observar mutaciones deletéreas en heteroplasma, siempre y cuando la frecuencia del alelo minoritario permanezca por debajo de cierto umbral y la mitocondria funcione correctamente. En el momento en que la frecuencia del alelo minoritario supera este umbral, las alteraciones mitocondriales podrían dar lugar a problemas en el funcionamiento mitocondrial y causar patologías (Li et al. 2015a). En este sentido, se han presentado una gran cantidad de enfermedades relacionadas con mutaciones heteroplásmicas –sordera, diabetes, epilepsia, demencia progresiva, hipoventilación, insuficiencia cardíaca, disfunción renal y aparición repentina de ceguera– que no manifiestan ningún síntoma hasta que el nivel de heteroplasma supera el umbral límite (Hancock et al. 2005). Además, Samuels et al. (2013) detectaron que, en ciertos tejidos, especialmente del riñón, del hígado y del músculo

esquelético, se detectaban mutaciones mitocondriales de forma recurrente que eran indetectables en otros tejidos de un mismo individuo y que también fueron encontradas en diferentes individuos no relacionados. Los resultados indicaban que la heteroplasmia se generaba por un proceso no aleatorio que daba lugar a mutaciones particulares únicamente en tejidos específicos.

El hecho de encontrar heteroplasmas en posiciones del genoma mitocondrial que están asociadas a patologías en la población general, debido probablemente al poco efecto que generan, permite que estas heteroplasmas se acumulen con el tiempo y puedan llegar a tener un impacto en enfermedades relacionadas con la edad (Elliott et al. 2008; Ye et al. 2014b). No obstante, aunque la heteroplasmia mitocondrial se haya relacionado con la edad y con enfermedades implicadas principalmente con la neurodegeneración (Wallace 2010), también se han descrito mutaciones en heteroplasmia en células tumorales que pueden llegar a tener un papel destacado en el proceso carcinogénico (Stefano and Kream 2016). Por el contrario, otros estudios han sugerido que algunas posiciones en heteroplasmia podrían estar relacionadas con el aumento de la longevidad (Ramos et al. 2016; Salvioli et al. 2008).

Aunque hasta el momento únicamente se ha hablado de la deriva genética y de la selección purificadora como posibles fuerzas evolutivas que actúan sobre las heteroplasmas, en varios estudios también se ha considerado la presencia de la selección positiva para ciertas heteroplasmas. Por ejemplo, Li et al. (2015a) detectaron en el tejido hepático un elevado número de heteroplasmas que tenían un impacto a nivel aminoacídico y proteico. La actuación de la selección positiva para estas mutaciones somáticas permitía reducir la función mitocondrial en el hígado, órgano donde se lleva a cabo una importante cantidad de procesos metabólicos, permitiendo así reducir la síntesis de productos que pueden ser dañinos para el DNA. Del mismo modo, Chinnery et al. (2000) también sugieren que las mitocondrias que contienen niveles elevados de DNAm mutado pueden proliferar selectivamente como respuesta a posibles defectos en la cadena respiratoria. Por lo tanto, la selección puede actuar tanto de forma positiva como negativa, de manera que es posible detectar células con una mayor o una menor cantidad de DNAm dentro de un mismo órgano o tejido.

Variación del DNAm

La acumulación de polimorfismos mitocondriales durante la historia de las poblaciones humanas ha permitido diferenciar en las poblaciones humanas diferentes linajes maternos, denominados

haplogrupos de DNAm_t (Torroni et al. 1996). Dichos linajes salieron de África en el momento de la migración del hombre anatómicamente moderno y fueron colonizando el resto del mundo. A medida que iban apareciendo nuevas mutaciones en el DNAm_t, se fueron formando nuevos haplogrupos y añadiendo nuevas ramas al árbol filogenético de la variación mitocondrial humana (Wallace and Chalkia 2013). Estas variaciones mitocondriales características de cada uno de los haplogrupos son el resultado del papel ejercido tanto de la deriva genética como la selección adaptativa, con el objetivo de que la mitocondria funcione bien dentro de un determinado ambiente (van Gisbergen et al. 2015). Así, si una mutación es beneficiosa para los individuos que viven en un mismo ambiente, ésta se irá transmitiendo a la descendencia y estableciendo como una nueva variante dentro de cada haplotipo individual –conjunto de variantes mitocondriales de un individuo detectadas en su secuencia de DNAm_t–, y aumentando así su frecuencia (Wallace 2010). El conjunto de haplotipos de individuos relacionados dentro de una misma región geográfica es lo que dará lugar a la aparición de un haplogrupo mitocondrial (Wallace and Chalkia 2013). Por ejemplo, el macrohaplogrupo N incluye 2 variantes –G10398A en el gen *ND3* y G8701A en el gen *ATP6*– que alteran el pH de la membrana mitocondrial y la regulación de Ca²⁺ confiriendo una ventaja para adaptarse a climas fríos; en otro ejemplo, la variante 3394C en el gen *ND1* que se localiza en los haplogrupos M y sugiere una ventaja adaptativa a elevadas altitudes (Wallace 2015).

Todos los linajes mitocondriales humanos aparecen de un ancestro común que surgió en África hace unos 130.000–170.000 años y que se denominó la “Eva mitocondrial” (Cann et al. 1987). A partir de dicho ancestro aparecieron todos los linajes africanos englobados dentro del macrohaplogrupo L. La aparición del haplogrupo L3 hace 65.000–70.000 años dio lugar a 2 macrohaplogrupos más, M y N, los cuales colonizaron el resto del mundo (Wallace 2015) (Figura 2). No obstante, sus subhaplogrupos abarcan zonas geográficas más específicas: los haplogrupos HV, JT, N1, N2, U y X se instalaron en Europa, el suroeste asiático y el norte de África, los haplogrupos R5–R8, M2–M6 y M4’67 en el sur de Asia y los haplogrupos A–G, Z y M7–M9 en Asia Oriental. Asimismo, los haplogrupos A–D y X migraron hasta América mediante 3 eventos demográficos distintos, dando lugar a los haplogrupos A2, B2, C1b, C1c, C1d, C1d1, D1, D4h3a y D4e1c tanto en América del Norte y Sur, los haplogrupos C4c, X2a y X2g en la costa este de Estados Unidos y los haplogrupos A2a, A2b y D3 en Canadá y Groenlandia. Por último, los haplogrupos M14–M15, M27–M29, Q, P, O y S se distribuyeron por Oceanía, concretamente Australia y Melanesia, mientras que el haplogrupo B4a1a1 se situó en las islas del sud-este asiático (Kivisild 2015) (Figura 2).

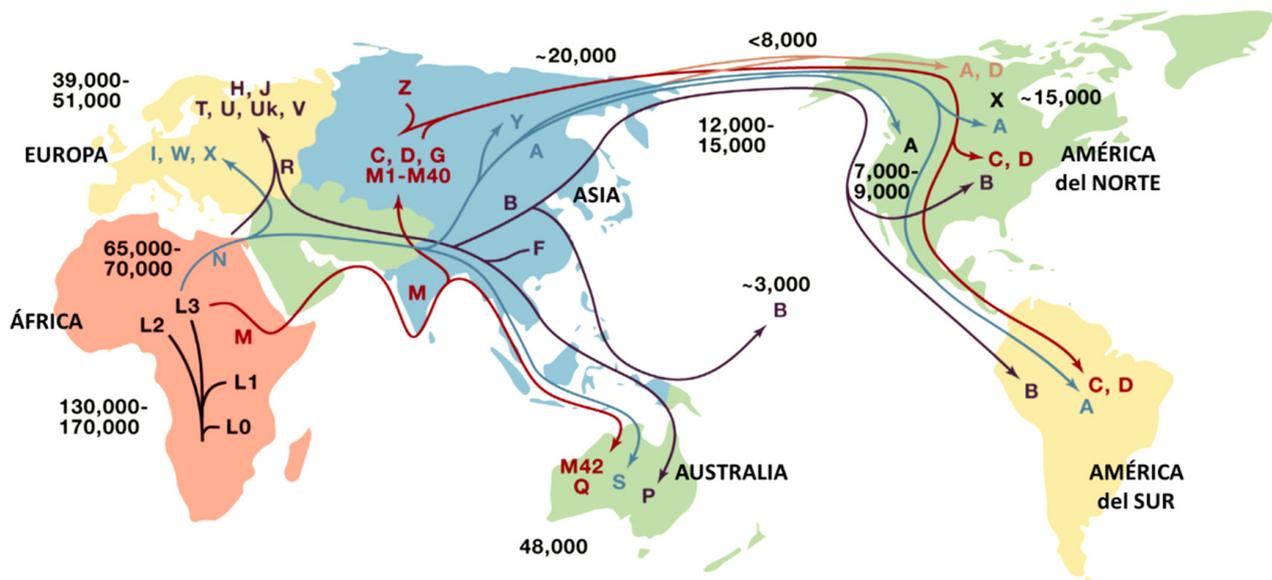


Figura 2. Proceso de migración de las diferentes poblaciones humanas y diversificación de los linajes de DNAm. Imagen modificada de Wallace (2015).

El hecho de que los genes del DNAm jueguen un papel esencial en OXPPOS y que a su vez OXPPOS juegue un papel en la fisiología celular, hace que determinados polimorfismos mitocondriales localizados en regiones funcionales puedan tener también efectos a nivel fenotípico y que, por lo tanto, diferentes haplogrupos mitocondriales se puedan relacionar con determinadas enfermedades (Wallace 2015) como, por ejemplo, el Alzheimer, el Parkinson o la neuropatía hereditaria de Leber (LHON), entre otras (Johns et al. 1992; Khusnutdinova et al. 2008; Santoro et al. 2010). Asimismo, la variante 10398G>A descrita anteriormente como una variante que puede dar cierta ventaja para adaptarse a ambientes fríos, también ha sido descrita como un factor de riesgo de sufrir Parkinson, Alzheimer, desorden bipolar, incluso algunos tipos de cáncer (Canter et al. 2005; Kato et al. 2001; Kazuno et al. 2006; van der Walt et al. 2004; van der Walt et al. 2003).

1.2. DNA MITOCONDRIAL y CÁNCER

Debido a que la mitocondria es la mayor responsable en la generación de energía celular, la presencia de mutaciones somáticas en el DNAm afecta sobre todo a aquellos tejidos que requieren una mayor demanda de energía, es decir, el tejido cerebral, nervioso y muscular (Wallace 2005; Wong 2007a). Así, el DNAm se ha asociado a multitud de fenómenos relacionados con el desarrollo, alteraciones neurológicas tales como Alzheimer o Parkinson, y también a procesos cancerígenos (Mattson et al. 2008). Respecto a estos últimos, recientemente se han producido grandes avances en el estudio y descubrimiento de la genómica del cáncer gracias a la implementación de las nuevas tecnologías de secuenciación masiva (*Next Generation Sequencing*, NGS)– en estudios de gran escala como *Whole Exome Sequencing* (WES) o *Whole Genome Sequencing* (WGS) (Koboldt et al. 2013).

Desde hace décadas, numerosos estudios se han centrado en el estudio de la disfunción mitocondrial en los procesos tumorales. Existen evidencias que implican a la mitocondria en el proceso de apoptosis y, aunque la célula cancerígena utilice, en general, la vía glicolítica como alternativa para la producción de energía, existe una gran dependencia en este orgánulo para su funcionalidad (Liang and Grootveld 2011). Así, cualquier alteración en la función mitocondrial puede contribuir al desarrollo y progresión del cáncer:

- *Alteraciones en el metabolismo energético: el efecto Warburg.* En la década de los 30, Otto Warburg (Warburg 1956) investigó las diferencias entre las células normales y tumorales respecto al metabolismo energético, desarrollando así la teoría del *efecto Warburg*. En un ambiente donde la célula dispone de suficiente oxígeno, la mitocondria es capaz de generar 38 moléculas de ATP mediante OXPHOS. Sin embargo, cuando la célula se encuentra en una situación de hipoxia, es decir, cuando no dispone de suficiente oxígeno, ésta utiliza como vía alternativa la glicólisis anaeróbica para la obtención de ATP. En este caso, la célula es capaz de generar únicamente 2 moléculas de ATP, siendo un proceso metabólico altamente ineficiente. Según la teoría de Warburg, aunque la glicólisis anaeróbica se puede dar en células normales, por ejemplo, cuando se realiza una actividad física excesiva, las células tumorales utilizan esta vía incluso cuando hay suficiente oxígeno para llevar a cabo OXPHOS, proceso denominado glicólisis aeróbica. A medida que el número de células cancerígenas va aumentando, su ratio de multiplicación es mucho más elevada que la de la angiogénesis (formación de vasos sanguíneos). Este desfase hace que no exista el oxígeno suficiente para que todas las células puedan acceder a él y puedan llevar a cabo OXPHOS. No obstante, la cantidad de glucosa presente es suficiente para dichas células, aun requiriendo un consumo de glucosa muy elevado para poder cubrir las necesidades energéticas (Devic 2016). Por lo tanto, según esta teoría, las células cancerígenas utilizarían la glicólisis aeróbica para llevar a cabo su metabolismo energético. Sin embargo, más recientemente ha surgido la teoría de que la célula normal se convierte en cancerígena en el momento en que, para metabolizar la glucosa, la vía glicolítica se convierte en la vía metabólica prioritaria. Esto hace que se plantee la posibilidad de que el efecto Warburg pueda no ser una causa del cambio de metabolismo de las células cancerígenas sino más bien una consecuencia (Devic 2016).

Ahora bien, el concepto de reprogramación metabólica, es decir, la capacidad que tienen las células cancerígenas de cambiar su metabolismo glicolítico a OXPHOS para adaptarse a nuevos ambientes y seguir progresando a nivel tumoral, ha cogido fuerza en los últimos años (Kim 2014) y varios estudios se han centrado en profundizar sobre el metabolismo de las células

cancerígenas. Por ejemplo, Strickland and Stoll (2017) describen, a través de varios estudios, que la glicólisis aeróbica descrita por Warburg no explicaría el 100% de la producción energética en muchos tipos de cáncer (Vander Heiden et al. 2009). Dicha hipótesis fue propuesta por primera vez por Guppy et al. (2002), determinando que la mayor parte de la energía obtenida (80%) en una línea celular humana de adenocarcinoma de mama (MCF-7) se generaba por vía oxidativa. Otros estudios señalan que las células cancerígenas tienen capacidad de reprogramarse y pasar de un metabolismo glicolítico a uno oxidativo en momentos en que se dispone de una baja cantidad de glucosa o se presenta una acidosis láctica en el ambiente (Lin et al. 2017a; Sharma et al. 2019). Le et al. (2014) demostraron la coexistencia tanto de un metabolismo glicolítico como oxidativo en linfocitos B en situaciones de hipoxia. El mismo autor también detectó que, en contra de lo que cabría esperar, en situaciones de hipoxia, algunas células cancerígenas mostraban bajas concentraciones de la subunidad alfa del factor 1 inducible por hipoxia (HIF1 α), un factor de transcripción que aumenta en situaciones donde se dispone de bajas concentraciones de oxígeno (Indran et al. 2011) y que promueve la mitofagia y la angiogénesis (Pavlidis et al. 2012), y un mayor consumo de oxígeno mediante OXPHOS. Estos resultados sugerían, por lo tanto, que determinadas células eran capaces de seguir con un metabolismo oxidativo en condiciones de hipoxia.

- *Alteraciones en la producción de ROS.* Las mutaciones mitocondriales pueden afectar a la generación de energía celular y desencadenar una mayor producción de ROS. La presencia de gran cantidad de ROS puede dar lugar a una situación en la que se produzca más cantidad de radicales libres de lo que los mecanismos antioxidantes de la célula puedan eliminar, originando así una situación de estrés oxidativo y comprometiendo la integridad mitocondrial (Reuter et al. 2010).
- *Alteraciones en la inducción de la apoptosis.* La presencia de mutaciones puede conducir a una disfunción mitocondrial y promover la resistencia a la apoptosis por medio de diferentes mecanismos como la glicólisis aeróbica, la regulación a la baja de factores proapoptóticos – por ejemplo, ROS, iones de calcio y proteínas como Bax y ciclofilina D (CypD)– o la regulación a la alza de factores antiapoptóticos –por ejemplo, la familia de proteínas Bcl-2, el gen *ANT2*, chaperonas y agentes antioxidantes– (Indran et al. 2011). Esto permitiría el crecimiento tumoral y la progresión del cáncer, y sugiere que las mitocondrias pueden servir como vía entre la muerte celular y el crecimiento celular anormal, contribuyendo de este modo al proceso neoplásico (Carew et al. 2003).

Se debe de tener en cuenta que las mutaciones mitocondriales pueden tener un mayor o menor impacto y tener diferente efecto en la mitocondria: mientras que las mutaciones con un mayor impacto pueden inhibir OXPHOS, aumentar la producción de ROS y promover la proliferación de las células tumorales, las mutaciones con un menor impacto pueden permitir la adaptación de los tumores a nuevos ambientes. Por este motivo, cualquier mutación a lo largo del genoma mitocondrial podría tener un efecto adverso debido a que el DNAm_t se expresa y se replica continuamente (Tan et al. 2008).

Paralelamente, en los últimos años se ha evidenciado la posible asociación entre el número de copias del DNAm_t y el riesgo de sufrir cáncer (Bussard and Siracusa 2017). Sin embargo, dicha asociación sigue siendo un tema a debatir debido a la discordancia de resultados obtenidos. La asociación puede variar según el tipo de cáncer –en algunos se ha asociado el aumento en el número de copias mitocondriales con un mayor riesgo de padecerlo, mientras que en otros se ha asociado la disminución con un mayor riesgo de sufrir cáncer– (Tabla 1). Ahora bien, también se han encontrado resultados contradictorios dentro de un mismo tipo de cáncer, teniendo en cuenta el tipo de muestras utilizadas para los análisis. Por ejemplo, se ha detectado tanto un aumento –comparando muestras de sangre de los pacientes con controles– como una disminución –comparando muestras de tejido tumoral con muestras de tejido adyacente– en el número de copias de DNAm_t en cáncer de células renales, mama y pulmón (Dai et al. 2013; Hofmann et al. 2014; Hosgood et al. 2010; Mambo et al. 2005; Meierhofer et al. 2004; Shen et al. 2010a; van Gisbergen et al. 2015; Xing et al. 2008). Incluso en el caso del cáncer de las células renales, se ha detectado tanto un aumento como una disminución en el contenido de DNAm_t en 2 estudios en que ambos compararon muestras de sangre de los pacientes con controles (Hofmann et al. 2014; van Gisbergen et al. 2015; Xing et al. 2008). Esta discrepancia de resultados también se observa en el caso del glioma cerebral, donde se detecta tanto un aumento como una disminución del número de copias en 2 estudios donde se compararon muestras tumorales con controles (Correia et al. 2011; Marucci et al. 2013; van Gisbergen et al. 2015). El incremento en el número de copias de DNAm_t podría producirse como un mecanismo de compensación frente a un defecto de OXPHOS, debido a un aumento en la generación de ROS, (Lee et al. 2000; van Gisbergen et al. 2015), mientras que la disminución podría ser causada por mutaciones en el gen *p53*, en la subunidad gamma de la DNA polimerasa (*POLG*) y en el factor de transcripción mitocondrial A (*TFAM*) (Guo et al. 2011; Lebedeva et al. 2009; Singh et al. 2009; van Gisbergen et al. 2015).

Tabla 1. Estudios basados en el análisis del contenido del DNAMt en diferentes tipos de cáncer.

Tipo de cáncer	Contenido de DNAMt	Metodología (Tejido en casos vs. Controles/ Tejidos no afectados)	Referencia
Células renales			(Hofmann et al. 2014; van Gisbergen et al. 2015)
Colorrectal			(Qu et al. 2011)
Mama	Aumento	Sangre pacientes vs. Controles	(Shen et al. 2010a; van Gisbergen et al. 2015)
Próstata			(Moore et al. 2017)
Pulmón			(Hosgood et al. 2010; van Gisbergen et al. 2015)
Cuello y cabeza			(Jiang et al. 2005; van Gisbergen et al. 2015)
Endometrio			(van Gisbergen et al. 2015; Wang et al. 2005)
Glioblastoma*	Aumento	Tumor pacientes vs. Controles	(Marucci et al. 2013; van Gisbergen et al. 2015)
Ovario			(van Gisbergen et al. 2015; Wang et al. 2006)
Fibrolamelar			(van Gisbergen et al. 2015; Vivekanandan et al. 2010)
Colorrectal			(Chen et al. 2011; van Gisbergen et al. 2015)
Esófago	Aumento	Tejido tumoral pacientes vs. Tejido no tumoral pacientes	(Lin et al. 2010; van Gisbergen et al. 2015)
Próstata			(Mizumachi et al. 2008; van Gisbergen et al. 2015)
Tiroides			(Mambo et al. 2005; van Gisbergen et al. 2015)
Células renales			(Xing et al. 2008)
Vejiga	Depleción	Sangre pacientes vs. Controles	(Williams et al. 2015)
Astrocitoma**			(Correia et al. 2011; van Gisbergen et al. 2015)
Sarcoma de Ewings	Depleción	Tumor pacientes vs. Controles	(van Gisbergen et al. 2015; Yu et al. 2010)
Células renales			(Meierhofer et al. 2004; van Gisbergen et al. 2015)
Colorrectal			(Cui et al. 2013)
Fibrolamelar primario			(van Gisbergen et al. 2015; Vivekanandan et al. 2010)
Hepatocelular	Depleción	Tejido tumoral pacientes vs. Tejido no tumoral pacientes	(van Gisbergen et al. 2015; Vivekanandan et al. 2010)
Gástrico			(Li et al. 2004a; van Gisbergen et al. 2015; Wu et al. 2005)
Mama			(Mambo et al. 2005; van Gisbergen et al. 2015)
Pulmón			(Dai et al. 2013; van Gisbergen et al. 2015)

*El glioblastoma representa el grado de mayor malignidad de cáncer cerebral. **En el estudio se incluyen astrocitomas de diferentes grados de malignidad: astrocitoma de menor malignidad, astrocitoma de mayor malignidad, glioblastoma.

1.2.1. Mutaciones mitocondriales en cáncer

La medida en que la presencia de las alteraciones mitocondriales son la causa del proceso carcinogénico o bien, son la consecuencia de este proceso, todavía se desconoce. Algunos estudios (Ju et al. 2014; Kim 2014) han determinado que las mutaciones mitocondriales en cáncer podrían derivar de errores durante la replicación mitocondrial y que no habría correlación entre las mutaciones mitocondriales y el proceso cancerígeno. Sin embargo, los mismos autores consideran más apropiado realizar estudios de asociación de las mutaciones somáticas con la enfermedad de interés –ya que en una enfermedad no siempre se encuentran involucradas las mismas mutaciones somáticas–, y teniendo en cuenta los cambios generados a nivel de la proteína (Ju et al. 2014; Kim 2014). Además, estas variaciones en el DNAm también pueden actuar individualmente, en combinación con otras variaciones mitocondriales o a través de la interacción con genes nucleares o factores ambientales, e influir en dicho proceso (Bai et al. 2007).

Tal y como se ha comentado anteriormente, cuando una mutación se genera en una de las copias de DNAm, esta molécula mutada coexiste con una mayoría de moléculas que no incorporan dicha mutación. Por lo tanto, es sorprendente que algunos estudios presenten una alta frecuencia de mutaciones mitocondriales somáticas fijadas únicamente en el tejido tumoral (Coller et al. 2001). Respecto a esto, Shen et al. (2010b) proponen que, en las células cancerígenas, las mutaciones que aparecen pueden adquirir cierto nivel de heteroplasmia, y a medida que el tumor va progresando, se produce una selección del fenotipo mutado debido a que las mutaciones mitocondriales son importantes para carcinogénesis. De esta forma, las mutaciones pueden llegar al estado de homoplasmia y ser predominantes en las últimas etapas del desarrollo tumoral. En esta misma línea, varios estudios sugieren que las mutaciones mitocondriales somáticas en células tumorales podrían escapar de una selección purificadora o bien sufrir una selección más relajada o incluso neutra, favoreciendo la acumulación de estas mutaciones y contribuyendo al desarrollo tumoral (Pereira et al. 2012; Stafford and Chen-Quin 2010). Ahora bien, otros estudios determinan que las mutaciones perjudiciales que pueden dar lugar a proteínas truncadas o alteraciones en los RNAt tendrían una desventaja selectiva y no contribuirían a la expansión clonal, sugiriendo que la célula cancerígena tiene una cantidad considerable de mitocondrias intactas que son esenciales para su correcto funcionamiento gracias a la presencia de una presión selectiva en cáncer (Ju et al. 2014; Zong et al. 2016).

La hipótesis de que en las células tumorales se podría dar un deterioro en OXPHOS para proteger las células de la apoptosis (Pereira et al. 2012; Shidara et al. 2005), o que las células tumorales se vuelven menos dependientes de la cadena de transporte de electrones en el momento que empiezan a sintetizar ATP a través de la glicólisis aeróbica (Stafford and Chen-Quin 2010), podrían explicar el papel de la selección en los procesos cancerígenos. Ahora bien, otros estudios también han descrito una fuerte selección contra las mutaciones de DNAm_t perjudiciales en células tumorales, debido a que las mitocondrias han de funcionar correctamente para que se lleve a cabo la carcinogénesis (Khaidakov and Shmookler Reis 2005). No obstante, Coller et al. (2001) y Vega et al. (2004) también consideran la posibilidad de que las mutaciones mitocondriales en homoplasmia aparecen únicamente por un proceso de mutación al azar, sin ofrecer ninguna ventaja fisiológica o requerimiento a nivel tumoral, y no por un proceso de selección. Así, dado los resultados contradictorios que se han publicado en diferentes estudios, el papel de las mutaciones en el DNAm_t sigue siendo todavía un desafío por resolver.

En muchos tipos de cáncer se han descrito un gran número de mutaciones somáticas tanto en homoplasmia como en heteroplasmia, siendo el de mama uno de los más estudiados seguido del de pulmón, el colorrectal y el de próstata junto con el carcinoma oral, y el estudio de la molécula mitocondrial completa el procedimiento más utilizado (Tabla 2). Sin embargo, algunos estudios han criticado la gran cantidad de mutaciones reportadas en la literatura. En este sentido, en un estudio presentado por Salas et al. (2005) se demuestra que muchas de las mutaciones mitocondriales que se relacionan con procesos tumorales en ciertos estudios no están realmente relacionadas con la aparición y el desarrollo del tumor, sino que realmente son polimorfismos neutros de las diferentes poblaciones humanas o *hotspots* mutacionales. Además, en los estudios basados en cáncer, la contaminación tanto a nivel inter como intraindividual puede influir y dar lugar a resultados erróneos. Por ejemplo, en un estudio de Bandelt y Salas (2009) reanalizaron las mutaciones mitocondriales detectadas en varios estudios de carcinomas de células escamosas orales (Prior et al. 2006) y de células bucales (Tan et al. 2008) mediante análisis filogenéticos. La presencia de diferentes haplotipos demostró que dichas mutaciones mitocondriales descritas son, en realidad, falsos positivos. Asimismo, Ramos et al. (2011) muestran que algunas mutaciones detectadas en heteroplasmia en cáncer pueden ser falsas mutaciones y ser en realidad el resultado de la co-amplificación de DNAm_t y DNA nuclear.

Tabla 2. Estudios basados en el análisis del DNAm en diferentes tipos de cáncer a nivel somático. N: número de estudios.

Tipo de cáncer	Región analizada				
	N	DNAm completo	D-loop	Parte Región Codificante	D-loop + Parte Región Codificante
Adenoma pituitario esporádico	1	(Kurelac et al. 2013)			
Cáncer cervical	4	(Kabekkodu et al. 2014)	(Sharma et al. 2005; Warowicka et al. 2013; Zhao et al. 2010)		
Cáncer de colon	1			(Lorenc et al. 2003)	
Cáncer colorrectal	12	(Ericson et al. 2012; Habano et al. 1999; He et al. 2010; Li et al. 2018; Lievre et al. 2005; Linkowska et al. 2015; Polyak et al. 1998; Skonieczna et al. 2018; Wang et al. 2011)	(Akouchekian et al. 2009; Chang et al. 2009)	(Akouchekian et al. 2011)	
Cáncer de cuello y cabeza	5	(Dasgupta et al. 2010; Fliss et al. 2000; Mithani et al. 2007; Zhou et al. 2007)	(Liu et al. 2015)		
Cáncer endometrial	4	(Guerra et al. 2014; Guerra et al. 2011)	(Czarnecka et al. 2009)	(Lorenc et al. 2003)	
Cáncer de esófago	3	(Gochhait et al. 2008; Tan et al. 2006)	(Hibi et al. 2001)		
Cáncer gástrico	3	(Hung et al. 2010)	(Rigoli et al. 2008; Wu et al. 2005)		
Carcinoma hepatocelular	6	(Li et al. 2018; Yin et al. 2019; Yu et al. 2018)	(Lee et al. 2004; Nomoto et al. 2002; Zhang et al. 2010)		
Cáncer de mama	18	(Fendt et al. 2011; Gasparre et al. 2007; Gochhait et al. 2008; McMahon and LaFramboise 2014; Mosquera-Miguel et al. 2008; Parrella et al. 2001; Tan et al. 2002; Tseng et al. 2011)	(Cai et al. 2011; Kuo et al. 2010; Losanoff et al. 2008; Tseng et al. 2006; Yu et al. 2007)	(Ghaffarpour et al. 2014; Martínez-Ramírez et al. 2018)	(Covarrubias et al. 2008; De Vitto et al. 2013; Shen et al. 2011)
Carcinoma oral	8	(Chattopadhyay et al. 2016; Lai et al. 2013; Palodhi et al. 2019)	(Datta et al. 2015; Lin et al. 2015; Liu et al. 2011; Mondal et al. 2013)		(Toure et al. 2019)
Cáncer de ovario	4	(Guerra et al. 2014)	(Van Trappen et al. 2007)	(Lorenc et al. 2003)	(Aikhionbare et al. 2007)
Cáncer de páncreas	5	(Hardie et al. 2017; Hopkins et al. 2018; Jones et al. 2001; Kassaei et al. 2006)	(Navaglia et al. 2006)		
Cáncer de piel	1		(Prior et al. 2009)		
Cáncer de próstata	8	(Arnold et al. 2015; Kalsbeek et al. 2016; Kloss-Brandstatter et al. 2010; Lindberg et al. 2013; McCrow et al. 2016; Parr et al. 2006)		(Petros et al. 2005; Ray et al. 2009)	
Cáncer de pulmón	13	(Dasgupta et al. 2012; Dasgupta et al. 2009; Fang et al. 2013; Fliss et al. 2000; Jakupciak et al. 2008; Jin et al. 2007; Kazdal et al. 2018; Suzuki et al. 2009; Wang et al. 2015)	(Chen et al. 2016; Fang et al. 2015; Yang Ai et al. 2013)	(Lorenc et al. 2003)	
Cáncer rectal	1	(Arstad et al. 2017)			
Cáncer de riñón	3	(Davis et al. 2014; Jakupciak et al. 2008; Nagy et al. 2003)			
Cáncer testicular	1	(Refinetti et al. 2017)			
Cáncer de tiroides	2	(Gasparre et al. 2007; Su et al. 2016)			
Cáncer del tracto urinario	2		(Duberow et al. 2016)	(Tzen et al. 2007)	
Cáncer de vejiga	2	(Fliss et al. 2000; Jakupciak et al. 2008)			
Leucemia	3	(Pagani et al. 2017; Silkjaer et al. 2013)	(Tyagi et al. 2018)		

Por otro lado, existen evidencias sobre una posible contribución de determinadas mutaciones germinales al desarrollo de ciertos tipos de cáncer en diferentes poblaciones (van Gisbergen et al. 2015). En este sentido, existe una gran variabilidad de resultados respecto a las asociaciones entre haplogrupos mitocondriales y cáncer en la bibliografía descrita, y mientras que algunos estudios detectan que ciertos haplogrupos pueden contribuir al aumento del riesgo a sufrir determinados tipos de cáncer, otros pueden tener un efecto protector ante la enfermedad. Del mismo modo, se ha encontrado que un mismo haplogrupo puede tener diferentes efectos dependiendo del tipo de cáncer analizado (Tabla 3).

Tabla 3. Estudios que describen asociaciones entre determinados haplogrupos mitocondriales y diferentes tipos de cáncer. +: Mayor riesgo o más probabilidad de sufrir la enfermedad; -: menor riesgo o efecto protector ante la enfermedad.

Tipo de Cáncer	Población de estudio	Haplogrupo	Efecto	Estudio
Carcinoma vulvar	Polaca	U, K	+	(Klemba et al. 2010)
		H	-	
Colorrectal	Escocesa	H, T	+	(Theodoratou et al. 2010, Li et al. 2015b)
		U	-	
Endometrio	China (sudoeste asiático)	D	+	(Xu et al. 2006)
	Polaca	H	-	(Czarnecka et al. 2009)
Esófago	China	D4, D5	+	(Li et al. 2011)
Hígado	China	M7	-	(Chen et al. 2017)
Mama	Caucásica	K	+	(Bai et al. 2007)
	India	N		(Darvishi et al. 2007)
	Polaca	I		(Czarnecka et al. 2010)
	China	M*, D5		(Fang et al. 2010)
	Caucásica	U		(Bai et al. 2007; Gutierrez Povedano et al. 2015)
Nasofaríngeo	China (sudoeste asiático)	R9, F1	+	(Hu et al. 2014)
Próstata	Blanca	U	+	(Booker et al. 2006)
Renal	Norteamericana			
Pulmón	China	G, M7	+	(Zheng et al. 2012)
		D, F	-	
Sarcoma de Kaposi	Iraní	UK	-	(Jalilvand et al. 2015)
Tiroides	China	D4a	+	(Fang et al. 2010)
	China	A4, B4a, B4g	+	(Su et al. 2016)
	Caucásica (sudeste europeo)	K	-	(Cocos et al. 2018)

*Asociación no encontrada en población con metástasis.

1.2.2. El glioma

La glía son células del sistema nervioso que tienen la función de proteger las neuronas y mantener su funcionamiento ayudando en la neurotransmisión, manteniendo el equilibrio iónico en el espacio extracelular y aislando los axones para acelerar la transmisión de impulso nervioso (Allen and Barres 2009). La población de células gliales se puede dividir en la macroglía, que incluye los astrocitos, los oligodendrocitos y los progenitores gliales NG2, y la microglía (Peng et al. 2014). La macroglía, al igual que las neuronas, proviene de células precursoras derivadas de la capa germinal denominada neuroectodermo, mientras que la microglía, que se compone de células que forman parte del sistema inmunitario y con carácter fagocítico, provienen del mesodermo (Jakel and Dimou 2017).

El glioma es un tipo de neoplasia que se desarrolla en el cerebro o en la médula espinal y engloba todos los tumores que tienen su origen en las células gliales (Schwartzbaum et al. 2006). Existen 3 tipos de células gliales que pueden producir tumores: los astrocitos, los oligodendrocitos y las células ependimales (American Brain Tumor Association 2019c). En 2007, la Organización Mundial de la Salud (OMS) elaboró una clasificación de los tipos de glioma según diferentes aspectos, comentada en la tabla 4 y esquematizado en la figura 3.

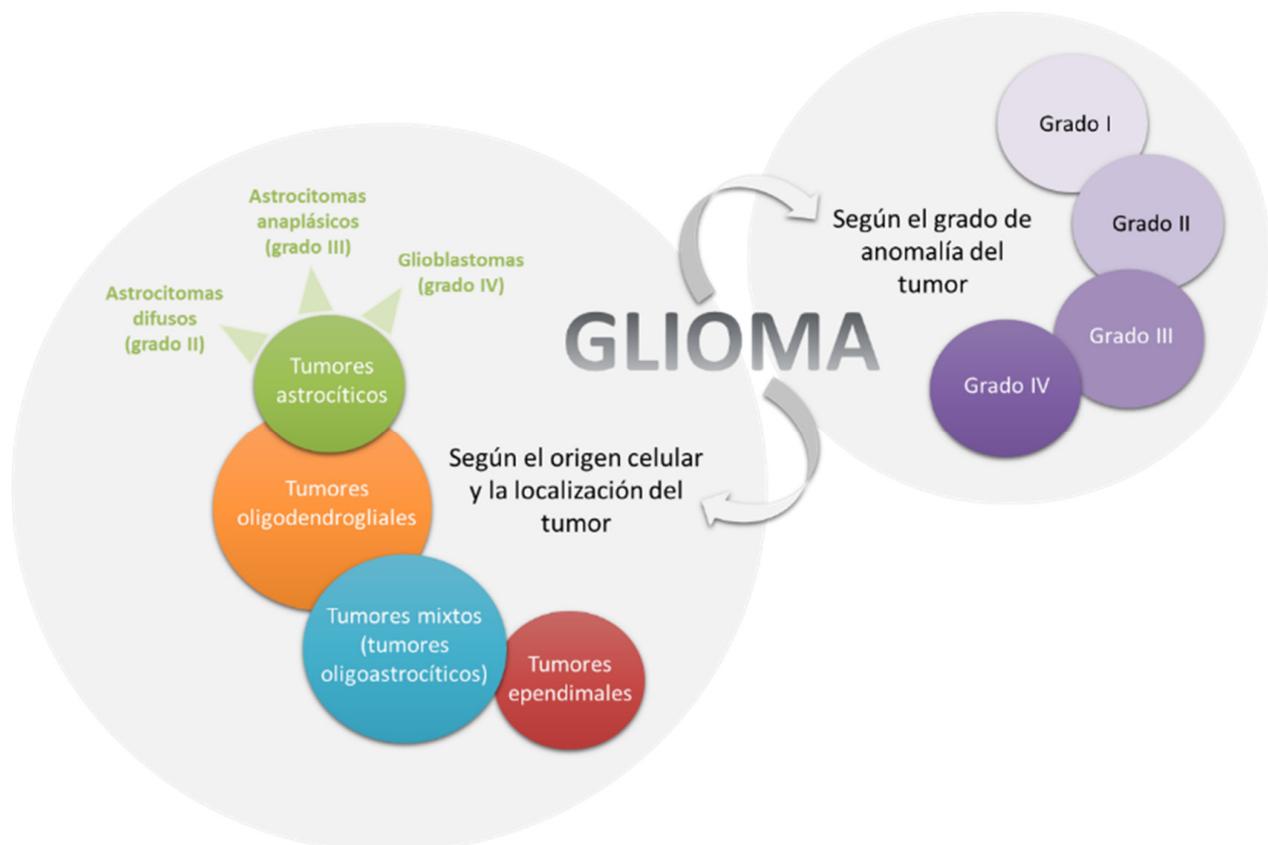


Figura 3. Clasificación de los gliomas según el grado de anomalía del tumor y según el origen celular y la localización del tumor.

Tabla 4. Clasificación de los gliomas según el grado de anomalía del tumor y según el origen celular y la localización del tumor.

Según el grado de anomalía del tumor

Tabla adaptada de *Brain Tumor Center* (Johns Hopkins Medicine 2019) ; (Louis et al. 2007)

Grado I (bajo grado). Las células tumorales se asemejan más a las células normales a nivel citológico. Los tumores crecen lentamente (bajo potencial proliferativo) y no suelen propagarse a tejidos cercanos. Son los tumores con mejor pronóstico.

Grado II. Son tumores de crecimiento relativamente lento y tienen un aspecto ligeramente anormal respecto las células normales a nivel citológico, incluyendo atipia nuclear y actividad mitótica. En algunos casos pueden propagarse a los tejidos cercanos y pueden volver a reproducirse. Además, algunos tumores pueden pasar a grados superiores.

Grado III. Las células tumorales son muy diferentes a las células normales a nivel citológico y tienen una gran actividad mitótica. Los tumores pueden propagarse a tejidos cercanos y tienden a reproducirse con el mismo grado de malignidad o bien, cambiar a un grado superior.

Grado IV (alto grado). Las células tumorales no se parecen a las células normales (a nivel citológico) y se propagan muy rápidamente. El tumor puede crecer en los tejidos cercanos fácilmente y forman vasos sanguíneos que facilitan su rápido crecimiento. Se puede observar, además, áreas de células muertas en el centro del tumor (necrosis). Generalmente, son los tumores con peor pronóstico.

Según el origen celular y la localización del tumor

Tumores astrocíticos. Son tumores que se originan a partir de los astrocitos y que se localizan tanto en el cerebro como en la médula espinal. En adultos, los tumores astrocíticos son los más comunes, representando un porcentaje superior al 75% de todas las neoplasias intracraneales (Wesseling et al. 2015). Dentro de los tumores astrocíticos, se puede encontrar:

Astrocitomas difusos (grado II). Son tumores con capacidad infiltrante, con un aumento en la cantidad de células (celularidad) y anomalía nuclear (atipia), pero sin actividad mitótica, proliferación endotelial o necrosis (Feichtinger et al. 2014; Louis et al. 2007).

Astrocitomas anaplásicos (grado III). Son tumores con mayor celularidad, atipia nuclear más marcada e hiperplasia, es decir, con un aumento de la cantidad de cromatina en el núcleo celular. También tienen actividad mitótica, pero no proliferación endotelial ni necrosis (Feichtinger et al. 2014; Louis et al. 2007).

Glioblastomas (grado IV). Son tumores con una gran celularidad, pleomórficos y con gran actividad mitótica con proliferación microvascular y necrosis (Feichtinger et al. 2014; Louis et al. 2007). Los hemisferios cerebrales son las regiones más comunes donde se suelen localizar los glioblastomas, con un 95% que alcanzan la región supratentorial. Con menos frecuencia se localizan en el tronco cerebral, cerebelo y médula espinal (Nakada et al. 2011). Existen 2 subgrupos: a) glioblastomas primarios que se desarrollan rápidamente y manifiestan un grado de lesión elevado desde el inicio; y b) glioblastomas secundarios que se desarrollan lentamente a través de la progresión a partir del grado II o III (American Brain Tumor Association 2019b). Son el tipo de tumor cerebral primario más común en adultos, representando unas 2/3 partes del total de tumores astrocíticos (Wesseling et al. 2015) y el que muestra mayor malignidad y peor pronóstico. Aunque se puede desarrollar a cualquier edad, el glioblastoma tiene un pico de incidencia entre los 50-70 años de edad.

Tumores oligodendrogliales. Son tumores que se originan a partir de células denominadas oligodendrocitos y que tienen como función cubrir y proteger las neuronas del cerebro y la médula espinal. Este tipo de tumores constituyen, junto a los tumores oligoastrocíticos, entre el 5-20% de todos los gliomas (van den Bent 2007). Principalmente, se ubican en los hemisferios cerebrales en adultos y se tratan de tumores bien diferenciados que se infiltran de forma difusa. Los oligodendrogliomas son tumores que están bien diferenciados y son considerados de grado II. No obstante, los tipos de oligodendrogliomas más anaplásicos (pérdida de la diferenciación celular) y, por lo tanto, más malignos, se clasifican como grado III. Mientras que los oligodendrogliomas se caracterizan por atipia nuclear y mitosis ocasional, los oligodendrogliomas anaplásicos se caracterizan por una elevada actividad mitótica, proliferación microvascular y una posible necrosis (Engelhard et al. 2002). Además, los oligodendrogliomas anaplásicos pueden formarse de *novo*, o bien evolucionar de un oligodendroglioma bien diferenciado (Engelhard et al. 2003).

Tumores mixtos (tumores oligoastrocíticos). Son tumores que presentan células neoplásicas gliales con características morfológicas tanto de astrocitos como oligodendrocitos (Wesseling et al. 2015). Se pueden localizar en los hemisferios cerebrales, aunque los lóbulos frontales y temporales son las zonas más comunes (American Brain Tumor Association 2019d). Tal y como se ha comentado anteriormente, representan entre el 5-20% de todos los gliomas junto con los tumores oligodendrogliales (van den Bent 2007).

Tumores ependimales. Son tumores neuroepiteliales que se originan en las células que revisten el canal central de la médula espinal o los ventrículos. Se trata de un tipo de tumor cerebral poco común, representando únicamente el 1-3% de gliomas en adultos (Chamberlain 2003), y se pueden clasificar en 4 subtipos según el grado de anomalía: subependimoma (grado I); ependimoma mixopapilar (grado II); ependimoma (grado III); y ependimoma anaplásico (grado IV) (American Brain Tumor Association 2019a). El tejido tumoral muestra un crecimiento expansivo dentro de unos límites bien definidos y, en algunos casos, pueden crecer infiltrándose e invadiendo zonas adyacentes (Zhang et al. 2016).

Aunque en el 2007 la OMS realizó una clasificación de los tumores del sistema nervioso central a partir de datos histológicos, en el año 2016 se llevó a cabo una reclasificación de estos tumores (Louis et al. 2016). En esta nueva clasificación, no solamente se tuvieron en cuenta los datos histológicos, sino que se incluyeron datos obtenidos a nivel molecular como el análisis de los genes *ATRX*, *IDH* y *TP53*, o la co-delección cromosómica de 1p/19q, de manera que muchos de estos tumores fueron reestructurados y se incorporaron nuevas entidades. La clasificación más detallada de estos tumores permitirá pronósticos y tratamientos mucho más precisos.

Antecedentes del estudio del DNAm en glioma

Recientemente, diferentes estudios han publicado tanto alteraciones genéticas como aberraciones en las vías de señalización que contribuyen al entendimiento de la formación de los gliomas. Hasta el momento se han descritos varios biomarcadores que están fuertemente asociados al diagnóstico y prognosis de los gliomas: 1) la co-delección de los cromosomas 1p/19q; 2) la metilación del promotor de la O⁶-metilguanina DNA metiltransferasa (MGMT); 3) mutaciones en los genes de la isocitrato deshidrogenasa (IDH) IDH1/2; 4) amplificación de los receptores de los factores de crecimiento epidemial (EGFR); y 5) la fusión o mutación puntual del gen *BRAF* que codifica para la proteína B-Raf (Siegal 2015). No obstante, varios estudios también se han centrado en el estudio del DNAm en diferentes tipos de glioma (Tabla 5). De manera concisa, los principales hallazgos de los estudios anteriores que se describen en la tabla 5 son los siguientes:

- *Antecedentes en el estudio de las alteraciones mitocondriales y el glioma.* Desde hace varias décadas, numerosos estudios se han basado en la detección de alteraciones en el DNAm en gliomas (Tabla 5). Además, los estudios están incluyendo cada vez más trabajos a nivel de predicción de estructuras relacionadas con RNAs mitocondriales, así como análisis de las estructuras tridimensionales de las proteínas sintetizadas a partir de determinados genes del DNAm, aunque la mayoría de estos estudios se basan en el análisis de líneas celulares (Lloyd et al. 2015; Soon et al. 2017; Yeung et al. 2014). Asimismo, también hay estudios que combinan análisis tanto a nivel mutacional como de proteína, aunque éstos se focalizan principalmente en el análisis de glioblastomas, es decir, los tumores más agresivos y con peor pronóstico (grado IV) (Vidone et al. 2015).

A excepción de aquellos basados en líneas celulares de glioma, casi todos los estudios analizan el tumor y otros tejidos como, por ejemplo, sangre o tejido adyacente al tumor, a fin de diferenciar las mutaciones que son específicas de tumor (Feichtinger et al. 2014; Kirches et

al. 2001; Kirches et al. 2002; Larman et al. 2012; Lueth et al. 2009; Lloyd et al. 2015; Mohamed Yusoff et al. 2017; Montanini et al. 2005; Vega et al. 2004; Vidone et al. 2015). En este sentido, la mayoría de los estudios detectan muy pocas mutaciones que puedan llegar a ser patogénicas. Además, a diferencia de la región control, la gran mayoría de mutaciones detectadas en la región codificante no son compartidas por varios tumores, es decir, únicamente se detectan en 1 individuo. En este sentido, varios estudios postulan la idea de que las variantes que se detectan en glioma tienen una baja probabilidad de que ejerzan un papel importante en la aparición o el desarrollo del tumor (Kirches et al. 2001; Montanini et al. 2005; Vega et al. 2004). De hecho, varios describen que parte de las mutaciones que detectan son polimorfismos y variantes comunes en la población (DeHaan et al. 2004; Kirches et al. 2001; Mohamed Yusoff et al. 2017; Montanini et al. 2005; Vega et al. 2004). Sin embargo, otros estudios sí consideran que las mutaciones mitocondriales que se detectan pueden tener cierta importancia en el desarrollo del glioma (Mohamed Yusoff et al. 2017; Yeung et al. 2014). Incluso Yeung et al. (2014) describen 4 mutaciones, 3 de ellas no sinónimas, que se detectan en 22 tumores analizados, siendo el único estudio en que detecta una frecuencia tan elevada de mutaciones compartidas por varios tumores en la región codificante.

Considerando, por lo tanto, los resultados obtenidos en los diferentes estudios basados en glioma y las discrepancias respecto a la implicación de las variantes mitocondriales detectadas en el desarrollo del glioma, un análisis más profundo de variantes mitocondriales detectadas sería un punto muy interesante para poder profundizar sobre su posible implicación en el desarrollo tumoral. Además, sería muy interesante comprobar si existen diferencias a nivel de variabilidad mitocondrial según el grado de malignidad tumoral, ya que hay pocos estudios que incluyen dicha diferenciación en sus muestras (Feichtinger et al. 2014; Kirches et al. 2001; Mohamed Yusoff et al. 2017; Montanini et al. 2005; Soon et al. 2017; Vega et al. 2004).

- *Antecedentes en el estudio de la heteroplasmia mitocondrial y el glioma.* Hasta hace poco más de una década, la mayoría de estudios de glioma se centraban únicamente en el estudio del D-loop mediante la secuenciación de Sanger (Kirches et al. 1999; Montanini et al. 2005; Vega et al. 2004), incluyendo en algunos de estos estudios el análisis complementario de alguna región codificante (DeHaan et al. 2004; Kirches et al. 2001). En la mayoría de éstos se presentaban mutaciones en heteroplasmia, muchas de ellas de longitud en la región control, sin embargo, la baja sensibilidad de la secuenciación de Sanger, no permitía detectar heteroplasmas a bajas frecuencias y, por lo tanto, es posible que en estos estudios se estuviera perdiendo cierta cantidad de información a nivel mutacional. Aunque algunos estudios ya empezaron a analizar gran parte o la totalidad del genoma mitocondrial mediante la secuenciación Sanger

(Feichtinger et al. 2014; Lueth et al. 2010; Lueth et al. 2009), no fue hasta la aparición de la NGS cuando la mayor parte de los estudios iniciaron análisis más amplios o completos del DNAm y pudieron llevar a cabo una detección más precisa de la heteroplasmia mitocondrial (Larman et al. 2012; Lloyd et al. 2015; Vidone et al. 2015; Yeung et al. 2014). Ahora bien, estos estudios se basan únicamente en el análisis de glioblastoma y, además, algunos de ellos presentan limitaciones en sus estudios. Por ejemplo, el estudio de Vidone et al. (2015) analiza el número más alto de muestras, aunque gran parte de la detección de heteroplasmas la realiza por medio de datos extraídos de WXS y WGS, lo que puede generar falsos positivos o negativos debido a la disminución de la cobertura en la secuenciación del DNAm y a la secuenciación de regiones nucleares en las que se encuentran NUMTs. Por otro lado, en el estudio de Lloyd et al. (2015) únicamente analizan los complejos III y IV del DNAm. Así pues, sería muy interesante llevar a cabo un estudio donde se puedan analizar de forma exhaustiva posibles mutaciones en heteroplasmia detectadas a lo largo del genoma mitocondrial completo y comparar los resultados obtenidos entre gliomas que tienen diferentes grados de malignidad tumoral.

- *Antecedentes en el estudio de la heterogeneidad mitocondrial y el glioma.* Tal y como se ha comentado anteriormente respecto a la reprogramación metabólica en cáncer (véase el apartado anterior 1.2), es posible que las células del glioma no siempre utilicen la vía glicolítica para la producción de ATP. Asimismo, también se ha descrito que las células cancerígenas de rápido crecimiento dependen más de la vía glicolítica que no las células que tienen un crecimiento más lento (Jose et al. 2011). Considerando que los gliomas cerebrales pueden crecer de forma ralentizada o de forma rápida dependiendo el grado de malignidad tumoral (Tabla 4), varios estudios precisan que dichos tumores podrían presentar cierta heterogeneidad metabólica para interactuar de distintas maneras con el mismo entorno extracelular, sobrevivir y proliferar hacia estadios de mayor malignidad (Griguer et al. 2005). De hecho, se ha descrito que ciertas áreas del glioma podrían estar bajo condiciones limitadas de glucosa, de manera que dichos tumores se caracterizarían por tener un fenotipo más oxidativo para poder producir energía de forma más eficiente (Gochhait et al. 2008). Existen varios estudios que señalan la gran capacidad de adaptación de las células del glioma a posibles cambios en el entorno. Griguer et al. (2005) demuestran que determinadas líneas celulares en humanos y ratones son capaces de sobrevivir en condiciones con poca disponibilidad de glucosa. Estas líneas celulares tendrían una expresión diferencial de la isoforma A de la lactato deshidrogenasa (LHD-A), principal enzima del metabolismo glicolítico, mientras que las líneas celulares con actividad glicolítica tienen una expresión diferencial de la isoforma LHD-B. Así pues, este estudio hipotetiza que, del mismo modo que existe una heterogeneidad tanto a nivel estructural como en la

vascularización del tumor, es posible que también puedan existir diferentes sustratos según si la región tumoral dispone de mayor o menor cantidad de oxígeno. Dicha hipótesis también fue planteada por Lin et al. (2017b), quienes observaron que las células de glioblastoma de un cultivo primario eran altamente oxidativas, considerando que la glicólisis aeróbica podría coexistir con la oxidación de otros sustratos para la obtención de energía. Otros estudios determinan que existe cierta heterogeneidad entre la región central y periférica del glioma, de manera que las mitocondrias de las regiones más aisladas del tumor, que muestran una funcionalidad mitocondrial superior, tienen una frecuencia respiratoria mayor que el centro del tumor (Lin et al. 2017a; Linkowska et al. 2015).

Teniendo en cuenta estos planteamientos expuestos sobre la posible heterogeneidad metabólica en los gliomas cerebrales, análisis más exhaustivos son necesarios para conocer en profundidad la actividad oxidativa en dichos tumores, siendo posible que la glicólisis aeróbica definida por Warburg no refleje completamente el metabolismo llevado a cabo por estas células para la obtención de energía.

- *Antecedentes en el estudio de asociación entre haplogrupos mitocondriales y glioma.* Aunque anteriormente se haya hecho referencia a diferentes estudios que muestran posibles asociaciones entre haplogrupos mitocondriales y diferentes tipos de cáncer (véase el apartado anterior 1.2.1), en la actualidad no existen estudios que hagan referencia al análisis de posibles asociaciones entre haplogrupos mitocondriales y glioma. De hecho, muy pocos estudios determinan los haplogrupos mitocondriales de las muestras de glioma en estudio (Vega et al. 2004; Vidone et al. 2015). En este sentido, un estudio de este tipo de asociación podría ser interesante para conocer con más profundidad el posible papel de variantes frecuentes y heredadas en el desarrollo del glioma.

Tabla 5. Estudios basados en el análisis del DNAm en glioma. N: tamaño de la muestra. GI: grado I; GII: grado II; GIII: grado III; GIV: grado IV. HRM: *High Resolution Melting*; TCGA: *The Cancer Genome Atlas*; WGS: *Whole Genome Sequencing*; WES: *Whole Exome Sequencing*; ND: no disponible.

Región DNAmt analizada	N	Población	Tipo de muestra	Tipo de secuenciación	Tipo de glioma	Resultados del estudio	Autor
D-loop (región HVRII)	12	ND	Tumor	Sanger y clonación	Astrocitomas	Bajo grado de heteroplasmia (>10%). La mayoría son heteroplasmias de longitud corresponden a la poly-C de la HVRII y muy pocas son heteroplasmias puntuales.	(Kirches et al. 1999)
D-loop En 1 paciente: estudio del gen ND1, gen ND4, varios RNAt y parte del gen CYTB	55	Alemana	Tumor Sangre (únicamente de 17 pacientes con glioblastoma)	Sanger	50 glioblastomas 4 astrocitomas (GII) 1 astrocitoma (GIII)	17 pacientes con mutaciones en homoplasmia: 12 no muestran cambios en la poly-C de la HVRII y 16 presentan la secuencia común C7TC6 en dicha región. 38 pacientes con bajos niveles de heteroplasmia (>10%): 27 presentan variaciones en la secuencia C7TC6 de la poly-C de la HVRII. Glioblastomas: Únicamente se detectan diferencias entre tumor y sangre en 7 pacientes. En 5 glioblastomas se detectan nuevos cambios en la poly-C a nivel somático. Detección de inserciones y deleciones en la poly-C de la HVRII en el 24% de las muestras. Elevado número de mutaciones en homoplasmia, aunque la mayoría se sitúan en posiciones polimórficas y ninguna de ellas es deletérea. Los mecanismos que generan los polimorfismos mitocondriales heredados se potencian en el glioma y podrían dar lugar también a mutaciones somáticas.	(Kirches et al. 2001)
D-Loop	10	Alemana	Tumor Sangre	Sanger	Astrocitomas de bajo grado	No se detecta ningún cambio mutacional en ninguna muestra tumoral. No existen evidencias de un menor porcentaje de mutaciones mitocondriales en tumores de bajo grado en comparación a los de mayor grado.	Kirches et al. 2002)
D-loop	53	Española	Tumor Sangre	Sanger	1 astrocitoma 14 astrocitomas (GII) 7 astrocitomas (GIII) 1 astrocitoma pilocítico 1 ganglioglioma anaplásico 17 glioblastomas (IV) 1 oligoastrocitoma 2 oligoastrocitoma (GII) 1 oligoastrocitoma (GIII) 1 oligodendrogloma 3 oligodendrogloma (GII) 3 oligodendrogloma (GIII) 1 xantoastrocitoma pleomórfico	Determinación de los haplogrupos mitocondriales de las muestras. Las transiciones en las posiciones 146 y 152 se detectan en varias muestras de distintos haplogrupos. Detección de 33 mutaciones somáticas: 33 heteroplasmias de longitud en la poly-C de la HVRII, 8 inserciones/deleciones en la misma región, 5 heteroplasmias, 4 cambios dinucleotídicos y 2 mutaciones puntuales. La mayoría de cambios asociados al tumor son hotspots mutacionales y polimorfismos humanos comunes. Casi todas las variantes detectadas son neutrales respecto a la funcionalidad mitocondrial y es poco probable que influyan en la aparición y el desarrollo del tumor.	(Vega et al. 2004)

Tabla 5. Continuación.

Región DNAm analizada	N	Población	Tipo de muestra	Tipo de secuenciación	Tipo de glioma	Resultados del estudio	Autor
D-loop 13 regiones codificantes	4	ND	Líneas Celulares	Sanger	Glioblastomas	23 alteraciones diferentes detectadas: 19 descritas como polimorfismos, 1 no descrita que no da lugar a ningún cambio a nivel aminoacídico, 2 variantes no descritas situadas en el D-loop y 1 variante no descrita que da lugar a cambio a nivel aminoacídico (T14634C). Las subunidades del complejo I son hotspots mutacionales en los tumores.	(DeHaan et al. 2004)
D-loop	42	Italiana	Tumor Linfocitos	Sanger	39 glioblastomas 2 astrocitomas anaplásicos 1 oligoastrocitoma anaplásico	Detección de 16 mutaciones somáticas: 3 detectadas en la HVRI (A16220; T16336C y T16311C) y 13 en la poly-C de la región HVRI. Detección de 6 nuevos polimorfismos Baja probabilidad de que las mutaciones mitocondriales sean importantes para el diagnóstico y el pronóstico de los gliomas. Tanto la inestabilidad mitocondrial como las mutaciones en el DNAm no confieren una mayor agresividad en los tumores malignos.	(Montanini et al. 2005)
DNAm completo	16	Italiana	Tumor	MitoAll resequencing kit	Gliomas Otros tumores (tiroideos y mama)	Detección de 2 variantes potencialmente patogénicas en heteroplasma (T4016G y T11204C). Mutaciones mitocondriales detectadas frecuentemente en regiones que codifican para proteínas.	(Gasparre et al. 2007)
DNAm completo	19	Alemana	Tumor Sangre	Sanger	Astrocitomas pilocíticos	34 mutaciones somáticas detectadas, 3 en heteroplasma. Únicamente 3 mutaciones son no sinónimas y se detectan en homoplasma (A8704G; C15452A y C6237A). 262 mutaciones germinales detectadas, todas ellas en homoplasma.	(Lueth et al. 2009)
DNAm completo	32	ND	Tumor No tumor (datos de TCGA)	NGS	Glioblastomas Otros tumores (adenocarcinoma de colon y rectal, leucemia mielóide aguda y citadenocarcinoma seroso de ovario)	Detección de 5 mutaciones somáticas, 4 de ellas no sinónimas (T3344C; T4762C; G9655A; G10586A). Detección de 292 mutaciones germinales. El glioblastoma es el tumor donde se detecta un menor número de mutaciones somáticas deletéreas (13%) en comparación al resto de tumores analizados. Probablemente pueden conferir una ventaja selectiva en el proceso oncogénico.	(Larman et al. 2012)

Tabla 5. Continuación.

Región DNAMt analizada	N	Población	Tipo de muestra	Tipo de secuenciación	Tipo de glioma	Resultados del estudio	Autor
Complejos I, IV y RNAt	25	Austriaca	Tumor Tejido adyacente	CEQ 2000 Analysis System	8 astrocitomas difusos (GII) 5 astrocitomas anaplásicos (GIII) 12 glioblastomas (GIV)	En 7 tumores, en los que se detectó una mayor deficiencia de los complejos I y IV, se secuenciaron los genes que codifican para OXPPOS y RNAt. Únicamente se detectaron 2 mutaciones somáticas no sinónimas (G7775A y G13099A). Posible flexibilidad metabólica en los gliomas. En tumores de bajo grado se detecta una elevada actividad OXPPOS. Por el contrario, en tumores de alto grado se detecta una menor actividad OXPPOS. Reducción significativa del contenido de DNAMt en glioblastomas.	Feichtinger et al. (2014)
DNAMt completo	8	Australiana	Líneas Celulares (12) Tumor (22) Tejido cerebral sano (13)	NGS HRM	Glioblastomas	Detección de variantes mitocondriales en las diferentes líneas celulares, todas ellas detectadas en heteroplasmia (rango: 3,1% a 55,5%). Detección de variantes en los 22 glioblastomas mediante HRM. Del total de mutaciones detectadas, 8 se detectan en la región control y 17 en la región codificante. 4 de ellas (G6999A; G8251A; C8552A y T11361C) se detectan en los 22 tumores. A excepción de la mutación G8251A, todas son mutaciones no sinónimas. Las variantes detectadas en el complejo I podrían actuar como iniciadores de glioblastomas. D-loop: región con mayor número de variantes; ND6: gen con mayor susceptibilidad a sufrir mutaciones; ND4: gen con mayor frecuencia de mutaciones.	(Yeung et al. 2014)
DNAMt completo (estudio centrado en el análisis de los complejos III y IV).	42	Inglesa	Líneas Celulares (10) Tumor y Sangre (32)	NGS	Glioblastomas	Amplio espectro de mutaciones asociadas a glioblastomas. 200 mutaciones detectadas, la mayoría con una baja carga mutacional. 25 mutaciones no sinónimas detectadas en los complejos III y IV. 9 de éstas son funcionales y ocurren principalmente en el gen CYTB. De estas 25 mutaciones, únicamente 2 son somáticas, ambas detectadas en heteroplasmia (G6619A y G9655A).	(Lloyd et al. 2015)

Tabla 5. Continuación.

Región DNAMt analizada	N	Población	Tipo de muestra	Tipo de secuenciación	Tipo de glioma	Resultados del estudio	Autor
						WXS y WGS	
						1211 mutaciones detectadas en tumores.	
						1193 mutaciones específicas en tumor (98,5%) tienen un nivel de heteroplasmia muy bajo (<10%) y sin efecto patogénico. Subestimación de posible contaminación.	
						Del resto de heteroplasmas (18), 9 fueron detectadas como probables mutaciones deletéreas y 3 fueron detectadas como potencialmente patogénicas (G1913A; G2652A y G2976A), pero ninguna de ellas muestra cambios en el plegamiento y estabilidad de la molécula. 1 mutación (G4403A) se localiza en el RNAt, pero tampoco afecta a la estabilidad del RNAt.	(Vidone et al. 2015)
						Sanger	
DNAMt completo	149	Italiana	WXS (21), WGS (28) y WXS + WGS (4) (datos de TCGA) Tumores (104)	NGS (45) Sanger (104)	Glioblastomas	951 mutaciones detectadas: 417 mutaciones definitorias de haplogrupo y 534 mutaciones no definitorias de haplogrupo.	
						247 mutaciones detectadas en regiones codificantes y 287 mutaciones detectadas en el D-loop y en regiones no codificantes.	
						40 mutaciones detectadas son potencialmente deletéreas. Únicamente 3 mutaciones fueron detectadas también como somáticas en NGS (datos TCGA) (G11693A en heteroplasmia; G13063A en heteroplasmia y G13135A).	
						Línea celular GII: 6 mutaciones, todas ellas en heteroplasmia. De las 6 mutaciones, 4 se localizan en la región codificante. Únicamente 2 son mutaciones no sinónimas (C6340T y A14181C).	
						Línea celular GIII: 2 mutaciones en homoplasmia, ambas localizadas en la región codificante. Únicamente la mutación T7964C es no sinónima.	
DNAMt completo	3	Caucásica	Líneas Celulares	GeneChip Resequencing Array	Glioma (GII, GIII, y GIV)	Línea celular GIV: 5 mutaciones, 3 de ellas en heteroplasmia. De las 5 mutaciones, 3 se localizan en la región codificante. Ninguna de ellas son mutaciones no sinónimas.	(Soon et al. 2017)
						Mayor estrés oxidativo en tumores de GII y GIII.	
						Mayor preservación de la mitocondria para asegurar su supervivencia y el progreso tumoral en GIV.	

Tabla 5. Continuación.

Región DNAm analizada	N	Población	Tipo de muestra	Tipo de secuenciación	Tipo de glioma	Resultados del estudio	Autor
D-loop	34	Malaya	Tumor Sangre	Sanger	6 astrocitomas pilocíticos (GI) 2 astrocitomas (GII) 2 ependimomas (GII) 3 oligodendrogliomas (GII) 5 astrocitomas anaplásicos (GIII) 16 glioblastomas (GIV) Otros: 15 meningiomas (GI)	<p>En el 51% de los pacientes se detectan 48 mutaciones somáticas en 27 posiciones del D-loop. El 94% de éstas se detectan en homoplasmia y el 19% en la poly-c de la HVRII.</p> <p>210 polimorfismos en 26 posiciones diferentes, siendo las mutaciones A263G, T16189C y C16261T las que se detectan con una mayor frecuencia.</p> <p>Respecto a las características clínico-patológicas, no hay diferencias significativas entre el grupo de pacientes con mutaciones en el D-loop vs. el grupo de pacientes sin mutaciones.</p> <p>Se sugiere que las mutaciones en el D-loop pueden ser biomarcadores potenciales en el tumor cerebral y promover la en la aparición y el desarrollo tumoral en el cerebro.</p>	(Mohamed Yusoff et al. 2017)*

* Los resultados hacen referencia al total de tumores cerebrales analizados (gliomas y meningiomas) (N=49).

1.3. METODOLOGÍAS PARA LA SECUENCIACIÓN DEL DNA MITOCONDRIAL

1.3.1. Secuenciación de Sanger

Hasta hace pocos años, la secuenciación Sanger, tecnología basada en la polimerización del DNA a partir de la utilización de dideoxinucleótidos marcados fluorescentemente (ddNTPs) y la realización de una electroforesis capilar, ha sido una de las metodologías más utilizadas para la detección de mutaciones en el DNAm. Sin embargo, esta tecnología se ve condicionada por las limitaciones de velocidad, rendimiento, resolución y costes asociados (McElhoe et al. 2014), así como la restricción para llevar a cabo análisis de múltiples muestras de forma paralela (Arsenic et al. 2015).

1.3.2. Next Generation Sequencing

Durante la última década, los métodos de secuenciación han conllevado grandes avances, tanto en la química como en la metodología utilizada para la detección de variantes. Dichos avances posibilitan a las nuevas tecnologías de secuenciación NGS tratar de forma paralela un elevado número de muestras, permitiendo el análisis de una gran cantidad de hebras de DNA en una sola secuenciación (Singh et al. 2016). Así, se generan millones de lecturas y se aporta una cantidad de datos mucho mayor que la que se podría obtener mediante los métodos tradicionales de secuenciación. Además, el aumento de la sensibilidad de detección permite realizar estudios mucho más exhaustivos de las variantes mitocondriales, sobre todo en la detección de la heteroplasmia mitocondrial. En este sentido, la NGS permite detectar variantes con una frecuencia muy baja (<5%), una frecuencia que es imperceptible con la secuenciación de Sanger, que es capaz de determinar variantes con un límite de detección entre el 8%-20% (Just et al. 2015; Ramos et al. 2013).

Por estos motivos, en muchas disciplinas la tecnología de NGS ha ido substituyendo a la clásica secuenciación de Sanger, que se está utilizando cada vez más como una técnica complementaria que como una técnica independiente (Just et al. 2015). La gran cantidad de información generada mediante NGS ha provocado, a la vez, la necesidad de desarrollar un gran número de herramientas computacionales para poder manipular los datos generados (Jia et al. 2012). Así, numerosos estudios se han centrado en el uso de NGS para el estudio del DNAm en diferentes ámbitos como la genética médica, evolución molecular, antropología molecular, epidemiología y metagenómica (Just et al. 2015).

Actualmente, existen numerosas plataformas basadas en secuenciación masiva como Genome Analyzer/HiSeq/MiSeq (Illumina Solexa), SOLiD System (Thermo Fisher Scientific), Ion PGM/Ion Proton (Thermo Fisher Scientific), HeliScope Sequencer (Helicos BioSciences) y Pacific Bioscience (Arsenic et al. 2015; Mardis 2008; Voelkerding et al. 2009). Todas estas plataformas utilizan una metodología basada en 3 fases:

- Fase 1: *Preparación de plantillas*. Esta fase consiste en la construcción de una librería de ácidos nucleicos –DNA o DNA complementario (DNAC)– mediante la fragmentación del DNA –de forma física (sonicación) o química (tagmentación)– y la adición de unos adaptadores –conjunto de oligonucleótidos de secuencias conocidas– a ambos extremos de los fragmentos generados. Posteriormente, las librerías se amplifican utilizando cebadores (*primers*) complementarios a los adaptadores para su posterior secuenciación (Berglund et al. 2011).
- Fase 2: *Secuenciación e imagen*. Es la fase donde se obtiene la secuencia de los ácidos nucleicos (lecturas o *reads*) a partir de la amplificación de las librerías. Los fragmentos de DNA se pueden secuenciar en un solo sentido o en 2 sentidos, siendo ésta última la que permitirá una alineación más precisa respecto al genoma de referencia (Berglund et al. 2011).
- Fase 3: *Análisis de datos*. Una vez obtenidos los datos de secuenciación, se lleva a cabo el análisis de los datos brutos obtenidos con el fin de generar un registro de la variación genética obtenida. Dicho análisis consta de diferentes fases:
 - Mapeo de las lecturas. Las lecturas generadas se mapean respecto al genoma de referencia de estudio de forma independiente. Sin embargo, se ha de tener en cuenta que en algunas zonas del genoma mitocondrial las lecturas son más difíciles de alinear –por ejemplo, en las zonas donde se localizan inserciones o deleciones (InDels)– y, por lo tanto, es posible que se obtenga un mal alineamiento en ciertas regiones. Esto puede dar lugar a resultados inexactos que pueden variar según los parámetros y la tecnología utilizada, e influir en la detección y genotipado de las mutaciones. Este problema se puede agudizar en aquellos proyectos que utilizan diferentes tecnologías de secuenciación y diferentes estrategias bioinformáticas (DePristo et al. 2011).
 - Una vez las lecturas se han mapeado, alineado y calibrado, se aplican diferentes parámetros basados en modelos estadísticos con el fin de detectar variaciones genéticas como sustituciones, InDels, inversiones, reordenamientos o variaciones en el número de copias (CNVs) con una mayor sensibilidad y especificidad (DePristo et al. 2011).

Aunque la base metodológica de las nuevas tecnologías de secuenciación es la misma en la mayoría de los casos, las plataformas pueden incluir aspectos únicos en cada fase y generar diferencias entre ellas. Así, cada plataforma puede presentar unos valores de calidad y precisión que no sean equivalentes a los del resto de plataformas y que, por lo tanto, se encuentren diferencias respecto a los datos y resultados obtenidos según la estrategia seguida (Grada and Weinbrecht 2013; Metzker 2010).

La tecnología Illumina

Illumina (Illumina Inc. 2019) es una tecnología de secuenciación de alto rendimiento que se basa en la formación de clones y el uso de terminadores reversibles para poder secuenciar a gran escala de forma rápida y precisa (Illumina Inc. 2010). Siguiendo la metodología comentada anteriormente, la tecnología Illumina lleva a cabo la secuenciación del DNA mediante el siguiente procedimiento:

- *Fragmentación del DNA*. Tal y como se ha comentado anteriormente, el DNA se fragmenta de forma aleatoria y se le añaden unos adaptadores en los extremos para generar las librerías de DNA (Figura 4.A).
- *Amplificación del DNA*. Los fragmentos del DNA se fijan en una superficie sólida (*flowcell*). Cada *flowcell* contiene carriles revestidos con oligonucleótidos que son complementarios a los adaptadores. Los fragmentos de DNA se hibridan en la *flowcell* en la que se distribuyen de forma aleatoria y se amplifican mediante la formación de puentes (Berglund et al. 2011). A medida que se van amplificando los fragmentos de DNA se van generando miles de copias que se colocan de forma lineal y próximas entre sí, y se van generando agrupaciones de DNAm amplificados (clústeres) (Schirmer et al. 2015) (Figura 4.B).
- *Secuenciación del DNA*. La principal característica de la tecnología Illumina es que la secuenciación del DNA se realiza por síntesis, al igual que la secuenciación de Sanger. Esta tecnología utiliza dideoxinucleótidos que actúan como terminadores reversibles marcados fluorescentemente para la polimerización. En cada ciclo se añaden los 4 terminadores reversibles (A, C, G, T), los *primers* y la DNA polimerasa (DNAPol) (Figura 4.C). Debido al terminador reversible bloqueado en el extremo 3' del nucleótido que se incorpora (Metzker 2010), únicamente se realiza la adición de una base en cada ciclo. Para poder realizar la correcta identificación de la base añadida, los fluoróforos se iluminan con un láser rojo para los nucleótidos adenina (A) y citosina (C), y un láser verde para los nucleótidos guanina (G) y timina (T), y se van creando imágenes mediante

diferentes filtros para identificar los 4 nucleótidos. Posteriormente a la identificación, se realiza un ciclo de lavado para poder retirar tanto las etiquetas fluorescentes como el resto de terminadores reversibles no añadidos. Una vez finalizado el lavado, se desbloquea el terminador reversible bloqueado en el extremo 3' del nucleótido incorporado para continuar con la síntesis y se empieza otro ciclo (Schirmer et al. 2015). En el caso de Illumina, esta plataforma incluye además un control del DNA estándar en cada carrera (*run*) para poder discriminar entre problemas generados por el instrumental o reactivos, o problemas generados por la propia muestra (Berglund et al. 2011).

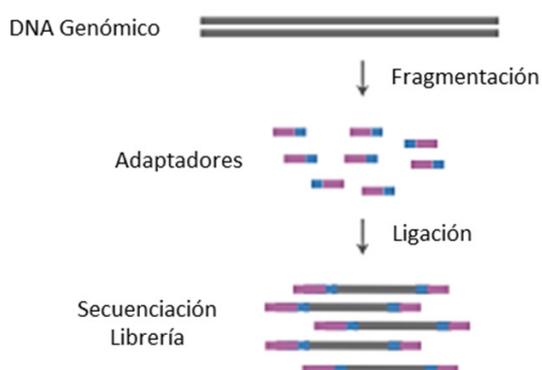
- *Análisis de datos*. Finalmente, se realiza el alineamiento de las lecturas y posteriormente la designación de las variantes (*variant calling*) (Figura 4.D). Durante la última década, varias herramientas y softwares informáticos se han propuesto para el análisis del DNAm, tanto a partir de la secuencia de exomas (WES *Whole Exome Sequencing*) y genomas completos (WGS *Whole Genome Sequencing*), como a partir de la secuenciación del DNAm:

- *MitoSeek*. Una de las primeras herramientas utilizadas fue MitoSeek (Guo et al. 2013). Este software realiza el ensamblaje a partir de las lecturas de WES o WGS que no han estado mapeadas con el genoma nuclear con el fin de evitar el ensamblaje incorrecto de lecturas que provienen de NUMTs. Permite detectar mutaciones a nivel somático cuando se comparan tejidos normales y tumorales y cambios estructurales del DNAm, así como establecer una estimación relativa en el número de copias mitocondriales. Además, ajusta el umbral de detección de la heteroplasmia mitocondrial al nivel apropiado según la profundidad de lectura analizada a partir de la calidad del alineamiento.
- *MToolBox*. Es otra de las herramientas publicadas para analizar las lecturas de WES o WGS (Calabrese et al. 2014). Esta herramienta permite también extraer información para generar un archivo con el formato estándar VCF (*Variant Call Format*) con el análisis de las variantes mitocondriales tales como heteroplasmas e InDels, y la asignación de haplogrupo. Dickins et al. (2014) también desarrollaron un procedimiento bioinformático diseñado para detectar la posibilidad de contaminación a bajos niveles que puede dar lugar a errores de secuenciación y resultados erróneos. Recientemente, Santorsola et al. (2015) añadieron nuevos parámetros a MToolBox para poder discriminar variantes que contribuyen a fenotipos defectuosos.
- *Mit-o-Matic*. Más recientemente, Vellarikkal et al. (2015) publicaron Mit-o-Matic, una herramienta bioinformática para evaluación clínica que analiza datos generados por NGS

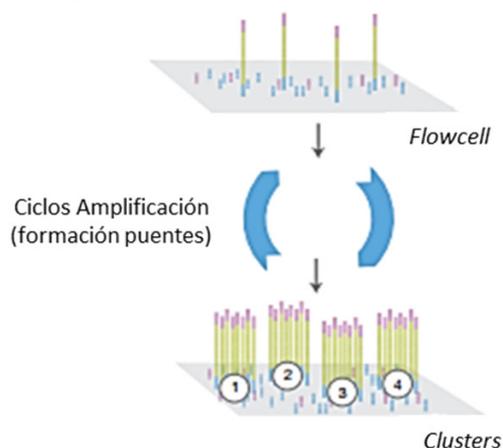
y permite determinar posiciones en heteroplasma, haplogrupos y asociaciones genéticas y fenotípicas a partir de las variantes mitocondriales detectadas.

- *mtDNA-Server*. Durante el 2016 se desarrolló *mtDNA-Server*, un programa que implementa un procedimiento bioinformático centrado en el análisis exclusivo del DNAm, con la finalidad de detectar heteroplasmas a muy baja frecuencia ($\leq 1\%$) y posible contaminación (Weissensteiner et al. 2016a) (véase más adelante en el apartado 3.5.2). Además, Weissensteiner et al. (2016a) llevaron a cabo comparaciones con los programas comentados anteriormente, demostrando generar una información más precisa y evitando reportar casos de falsos positivos de heteroplasmas.

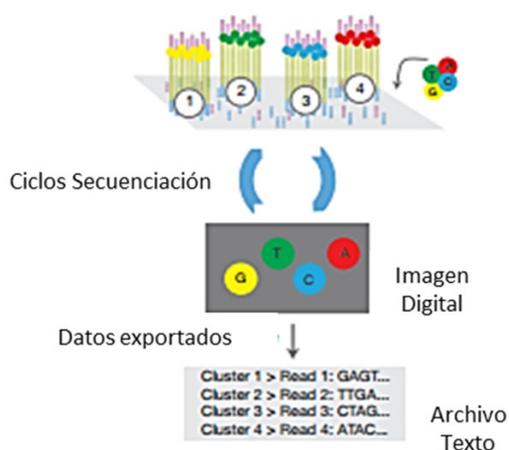
A. Preparación Librerías



B. Amplificación Clusters



C. Secuenciación



D. Alineamiento y Análisis Datos

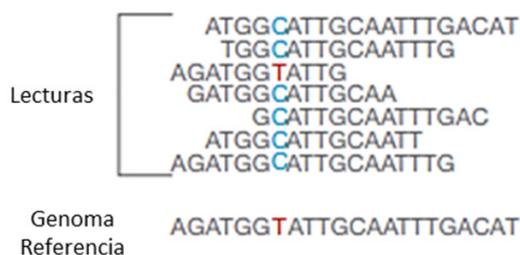


Figura 4. Fases del proceso de secuenciación de la tecnología Illumina. a) Preparación de las librerías; b) Formación de clústeres y posterior amplificación; c) Secuenciación por síntesis; y d) Análisis de los datos obtenidos. Imagen modificada de *An Introduction to Next-Generation Sequencing Technology* (Illumina Inc. 2017).

1.3.3. Detección de mutaciones en el DNA mitocondrial

Durante más de una década numerosos artículos han descrito mutaciones mitocondriales mediante la secuenciación de Sanger. Sin embargo, gracias a la aparición de la NGS, se ha logrado aumentar la sensibilidad de detección de la heteroplasmia mitocondrial, llegando a detectar variantes con una frecuencia imperceptible mediante la secuenciación tradicional (<5%) (Just et al. 2015). No obstante, aunque con la NGS se ha conseguido detectar heteroplasmas a menor frecuencia, los niveles de heteroplasmia descritos en la literatura varían ampliamente. Gran parte de esta variabilidad viene dada por el límite que se establece para la detección de mutaciones mitocondriales. Este límite suele ser una frecuencia del alelo minoritario elegida (MAF) que se debe superar para determinar que la secuencia mitocondrial que se estudia es diferente a la secuencia de referencia utilizada (Mantione 2016). Por ejemplo, se pueden encontrar estudios que detecten heteroplasmas a un nivel del 5% (Tang and Huang 2010), 2% (Skonieczna et al. 2015), 1% (Diroma et al. 2014), e incluso a un nivel inferior del 1% (Kloss-Brandstatter et al. 2015; Li et al. 2015). Estas discrepancias encontradas en los diferentes estudios pueden ser debidas a diferentes factores:

- *Contaminación*. El riesgo de contaminación sobre todo en estudios multiplex donde se lleva a cabo la manipulación de un gran número de muestras muy similares, hace que la detección de variantes mitocondriales sea un proceso complejo debido a la dificultad de saber discriminar falsos positivos de verdaderas variantes genéticas (Dickins et al. 2014).
- *Co-amplificación de DNAn*. La presencia de inserciones nucleares de origen mitocondrial o NUMTs hace que existan secuencias que pueden ser idénticas entre el DNAm y el DNAn y que, por lo tanto, exista la posibilidad de detectar falsos positivos debido a la co-amplificación entre los 2 genomas. La co-amplificación con NUMTs se puede dar por varios motivos: 1) el diseño incorrecto de *primers* en el caso de trabajar con amplicones del DNAm; 2) errores a nivel bioinformático relacionados con el alineamiento de lecturas que no corresponden a la secuencia del DNAm; y 3) la purificación inadecuada de las mitocondrias antes de la secuenciación del DNAm (Mantione 2016). No obstante, estos factores variarán según el tipo de estrategia seguida para la elaboración de las librerías y la posterior secuenciación.

Goios et al. (2008) evidencian que no existe riesgo de contaminación significativo por la amplificación de NUMTs utilizando las técnicas de laboratorio rutinarias y secuenciación de Sanger, sin embargo remarcan que ésta depende de las características de la muestra, así como del tipo de tejido. Se ha de tener en cuenta, además, que el nivel de co-amplificación de DNAn puede ser inferior al que se puede detectar con las técnicas de secuenciación tradicionales y,

por lo tanto, únicamente ser perceptible con NGS. En este sentido, algunos estudios consideran que los trabajos que detectan abundantes variantes a niveles $<2\%$, podrían ser falsos positivos debido a la presencia de NUMTs (Ye et al. 2014a) y que deberían ser excluidos del análisis.

Existen estudios que tienen en cuenta los NUMTs para sus análisis y siguen estrategias para evitar su amplificación (Li et al. 2012). No obstante, Albayrak et al. (2016) determinan que las bases de datos de NUMTs conocidos en las que se establecen las estrategias pueden no ser suficientes para evitar la amplificación de NUMTs, de manera que todavía se podrían estar aceptando determinadas variantes como falsos positivos.

- *Errores de secuenciación.* Las nuevas tecnologías de secuenciación no están ausentes de errores durante el proceso de secuenciación. Todas las tecnologías de NGS tienen un promedio de tasa de error considerablemente más alta que la observada para las secuencias Sanger de alta calidad –de 1/10.000 a 1/100.000– (Kircher and Kelso 2010). Estos errores pueden diferir según la tecnología utilizada en cada una de las plataformas. En el caso de la tecnología Illumina la principal fuente de errores es la de tipo sustitución nucleotídica errónea (*miscalls*), debido a que existe una fuerte correlación respecto a la intensidad y la limitación de los filtros que separan las señales entre A y C –detectados por láser rojo– y G y T –detectados por láser verde–. Además, los nucleótidos A y C son los que generan intensidades de señal más elevadas y los que son más propensos a errores de sustitución en comparación a los nucleótidos G y T (Schirmer et al. 2015). Por otro lado, problemas originados durante la fase (*phasing*) y la prefase (*prephasing*) pueden causar ruido en la señal del clúster. En la tecnología Illumina los clústeres se identifican durante los 4 primeros ciclos de secuenciación. En cada uno de los ciclos se incorpora un único nucleótido, sin embargo, en determinados clústeres se pueden incorporar por error más nucleótidos o bien no incorporar ningún nucleótido en alguno de los ciclos (Berglund et al. 2011). Cuando un fragmento de DNA no logra incorporar un nucleótido en un ciclo y queda rezagado respecto a la secuenciación del resto de fragmentos se denomina fase (*phasing*) (Ledergerber and Dessimoz 2011). Las causas que pueden dar lugar a esta situación pueden estar relacionadas con la cinética enzimática o la eliminación incompleta de los terminadores del extremo 3' o de los fluoróforos (Schirmer et al. 2015). En cambio, si se sintetizan bases múltiples en un sólo ciclo se denomina prefase (*prephasing*) (Ledergerber and Dessimoz 2011). En este caso, las causas pueden estar relacionadas con un lavado inadecuado de la *flowcell* que puede provocar que la síntesis avance demasiado rápido, por el salto de algún ciclo de incorporación de nucleótidos en las secuencias de algún clúster o por la incorporación de nucleótidos sin un terminador en el extremo 3' (Schirmer et al. 2015). En todos los casos el número de secuencias afectadas aumenta con cada ciclo y esto puede afectar y limitar la longitud de las lecturas generados. Así, la

presencia de estos desfases puede provocar una mayor tasa de error hacia el final de las lecturas (Kircher et al. 2009). Por último, también se ha descrito que la posible oxidación del DNA debido a diferentes factores tales como el método de extracción utilizado, el tipo de almacenamiento a largo plazo del DNA en tampones acuosos, el calor, entre otros (Bruskov et al. 2002; Ravanat et al. 2002), pueden dar lugar también a errores de secuenciación (Costello et al. 2013; Zhang et al. 2016).

A nivel general, en la tecnología Illumina se ha observado que la ratio de error en las lecturas en sentido reverso es el doble que en las lecturas en sentido directo (0,0042 y 0,0021, respectivamente) (Schirmer et al. 2016). En concreto, para la plataforma MiSeq (Illumina) la ratio de error es significativamente mayor en las lecturas en sentido reverso (0,0107) en comparación a las lecturas en sentido directo (0,0064) (Schirmer et al. 2016; Schirmer et al. 2015). Además, también se ha comprobado que ciertas posiciones de la lectura tienden a generar más errores respecto a otras (Schirmer et al. 2016).

- *Análisis de datos*. Las diferencias en los niveles de detección podrían ser debidas al manejo erróneo que se realiza de los datos, como por ejemplo a nivel de alineamiento de secuencias, sobre todo cuando hay la presencia de heteroplasmas de longitud (Cerezo et al. 2009).

Todos estos factores hacen que tanto los errores aleatorios como los no aleatorios, que son más aparentes en la secuenciación masiva que en la secuenciación de Sanger, influyan en la capacidad de discriminar una heteroplasma de baja frecuencia auténtica del ruido de fondo presente (Just et al. 2015) y que se requieran potentes herramientas para poder distinguir las verdaderas mutaciones de los errores instrumentales por lo que a la detección y cuantificación de heteroplasmas se refiere, sobre todo aquellas de baja frecuencia (Tang and Huang 2010). En este sentido, se pretende que los futuros estudios proporcionen mejoras en las tasas de mutación y que éstas sean específicas para cada posición. De este modo se podría determinar la probabilidad de hallar una heteroplasma, teniendo en cuenta la tasa de mutación que tiene la posición donde se localiza, sobre todo para la región codificante (Mantione 2016).

2

HIPÓTESIS del TRABAJO y OBJETIVOS

2. HIPÓTESIS del TRABAJO y OBJETIVOS

2.1. PLANTEAMIENTO del TRABAJO e HIPÓTESIS

El genoma mitocondrial ha sido considerado una pieza clave en el metabolismo energético de las células cancerígenas debido a su estrecha relación con el proceso oxidativo. Además, la fuerte demanda de oxígeno y de ATP por parte del tejido cerebral hace de este tejido un claro candidato para el estudio del DNAm y el análisis de su papel dentro del desarrollo tumoral. Así pues, varios aspectos han de tenerse en cuenta.

En primer lugar, dado que el genoma mitocondrial puede presentar mutaciones en heteroplasmia, sería de gran interés conocer su posible efecto dentro del proceso tumoral y profundizar en cómo las fuerzas evolutivas pueden llegar a influir en ellas. A nivel del análisis del glioma, parte de los estudios recogidos en la literatura, que describen mutaciones en heteroplasmia, se basan en el análisis del D-loop a través de la secuenciación de Sanger. La aparición de la secuenciación masiva mediante diferentes tecnologías, como Illumina, propició la publicación de trabajos en los que ya se realizaba el estudio de gran parte o la totalidad de la molécula mitocondrial, permitiendo análisis más exhaustivos. Así pues, el análisis del genoma mitocondrial completo puede ser relevante en el estudio del desarrollo del glioma, tanto a nivel de mutación en homoplasmia como de heteroplasmia, a cualquier nivel. No obstante, los niveles de heteroplasmia detectados por NGS varían ampliamente en la literatura existente debido a que se ha hallado una extensa variabilidad en los límites de detección establecidos para determinar las heteroplasmas. Además, se han descrito errores producidos en la secuenciación masiva que pueden influir en la capacidad para discriminar una verdadera heteroplasmia de baja frecuencia. Por este motivo, se considera interesante disponer de un modelo que establezca unos límites rigurosos de detección de la misma que puedan ser aplicados posteriormente en el estudio de heteroplasmia mitocondrial en glioma o en otros contextos en los que la detección de heteroplasmia es relevante.

En segundo lugar, y centrándonos ya en el estudio del glioma cerebral propiamente dicho, hay que tener en cuenta que algunos estudios consideran que los tumores necesitan la maquinaria mitocondrial intacta para poder progresar (Khaidakov and Shmookler Reis 2005) y que una proporción de mitocondrias deberían permanecer intactas para funcionar correctamente (Ju et al. 2014; Stewart and Chinnery 2015; Zong et al. 2016). Por lo tanto, se plantea la necesidad, por parte del tumor, de mantener un ambiente donde se genere una alta cantidad de agentes antioxidantes. Si bien la presencia de ROS puede inducir la vía apoptótica en una célula normal, la célula tumoral evitaría la apoptosis pasando de un metabolismo oxidativo a uno glicolítico. Ahora bien, es posible que esta teoría no explique completamente el metabolismo del glioma y que dicho tumor pueda mostrar cierta heterogeneidad metabólica para interactuar de distintas maneras con el entorno extracelular (Duan et al. 2018a; Griguer et al. 2005; Lin et al. 2017b; Santandreu et al. 2008; Wanka et al. 2012). La posible capacidad por parte de las células del glioma de obtener energía también por vía OXPHOS, permite hipotetizar que los gliomas no tienden a acumular mutaciones a nivel somático para evitar la acumulación de ROS en el ambiente. Los resultados obtenidos hasta el momento, como ya se ha expuesto en el apartado de introducción, muestran que, a nivel somático, los gliomas acumulan pocas mutaciones consideradas patogénicas. Asimismo, muchas de las mutaciones detectadas en varios tumores se presentan en posiciones consideradas polimórficas o *hotspots* mutacionales de la región control (DeHaan et al. 2004; Kirches et al. 2001; Mohamed Yusoff et al. 2017; Montanini et al. 2005; Vega et al. 2004). No obstante, la mayoría de los estudios ya realizados incluyen un sólo tipo de glioma, centrándose en los más agresivos (DeHaan et al. 2004; Larman et al. 2012; Lloyd et al. 2015; Vidone et al. 2015; Yeung et al. 2014). Así pues, sería de gran interés ampliar el análisis del estudio de la variabilidad mitocondrial para determinar su posible implicación en la aparición y el desarrollo del glioma, valorando posibles diferencias entre los tumores de distintos grados de malignidad.

Por último, dados los antecedentes presentados respecto a evidencias sobre asociaciones de determinados haplogrupos mitocondriales y el riesgo de sufrir algún tipo de cáncer, hace que se considere el hecho de que ciertos haplogrupos mitocondriales puedan ser más susceptibles a desarrollar un glioma. En la bibliografía se describe una gran variabilidad de resultados respecto las asociaciones entre haplogrupos mitocondriales y cáncer: mientras que en algunos casos se detecta que ciertos haplogrupos pueden contribuir al aumento del riesgo a sufrir determinados tipos de cáncer, en otros parece detectarse un efecto protector ante la enfermedad. En este sentido, es conocido que la producción de ROS depende, entre otros factores, del haplogrupo mitocondrial, lo que estaría en la línea anteriormente presentada que hipotetiza que las células tumorales evitan la acumulación de ROS en el ambiente, llevando a pensar que aquellos haplogrupos en que la producción de ROS es

menor incrementarían el riesgo de desarrollar un glioma. Sin embargo, hasta el momento, no existen estudios que hagan referencia a la determinación y asociación de los haplogrupos mitocondriales en glioma. Aportar una nueva visión basada en una posible asociación entre determinados haplogrupos y el riesgo de sufrir un glioma, pudiera ser un punto de partida muy interesante para conocer con más profundidad el posible papel de las variantes mitocondriales frecuentes en el desarrollo tumoral.

Considerando los distintos aspectos que se acaban de exponer, en la presente tesis doctoral se propone, una vez validada la secuenciación del genoma mitocondrial completo por NGS, testar la hipótesis de que la elevada dependencia de ATP por parte del cerebro debido a su elevado consumo de oxígeno hace que determinadas mutaciones mitocondriales tanto fijadas como en heteroplasmia puedan detectarse en los genomas mitocondriales de los gliomas cerebrales. Sin embargo, la heterogeneidad metabólica de las células cancerígenas en el glioma, así como la necesidad de mantener el buen funcionamiento mitocondrial por parte del tumor para progresar y alcanzar estadios de mayor malignidad, contempla la posibilidad de que dicho tumor no tienda a acumular una gran cantidad de variaciones a nivel somático. A nivel germinal, los haplogrupos mitocondriales, con las variaciones polimórficas que los definen, pueden aumentar o disminuir el riesgo de sufrir un glioma en función de su eficiencia metabólica y producción de ROS, generando así posibles asociaciones con la enfermedad.

2.2. OBJETIVOS del TRABAJO

Para contrastar nuestra hipótesis se plantean los siguientes objetivos:

1.- Evaluar la capacidad de detección y reproducibilidad de la NGS, utilizando la metodología de Nextera XT[®] y la plataforma MiSeq (Illumina), en la detección fiable de la variabilidad del genoma mitocondrial tanto en homoplasmia como en heteroplasmia (Capítulos II y III):

1.I. Valorar la calidad de los datos generados por NGS utilizando la metodología de Nextera XT[®] y la plataforma MiSeq (Illumina) (Capítulo II);

1.II. Determinar la fiabilidad y reproducibilidad de los resultados obtenidos en el análisis del DNAm^t a partir de la generación de librerías de NGS independientes, utilizando la metodología de Nextera XT[®], y la secuenciación en diferentes *runs*, mediante la plataforma MiSeq (Illumina) (Capítulo III);

1.III. Proporcionar una estrategia alternativa a las existentes en la literatura que permita establecer unos límites de detección para la heteroplasmia mitocondrial por NGS con la plataforma MiSeq (Illumina) (Capítulo III).

2.- Valorar la implicación del genoma mitocondrial en el desarrollo de gliomas cerebrales de diferentes grados de malignidad (Capítulo IV):

2.I. Evaluar si los haplogrupos mitocondriales desempeñan alguna función en relación al riesgo de presentar glioma y si éstos podrían modular la malignidad tumoral;

2.II. Identificar el papel que juega la variabilidad mitocondrial, tanto en homoplasmia como en heteroplasmia, detectada en el glioma;

2.III. Determinar si la carga mutacional acumulada guarda relación con el desarrollo del glioma o bien, con el grado de malignidad tumoral.

3

CAPÍTULO I MATERIAL y MÉTODOS

3. CAPÍTULO I Material y Métodos

Todo el proceso de caracterización de las muestras estudiadas se resume en el esquema de la figura 5.

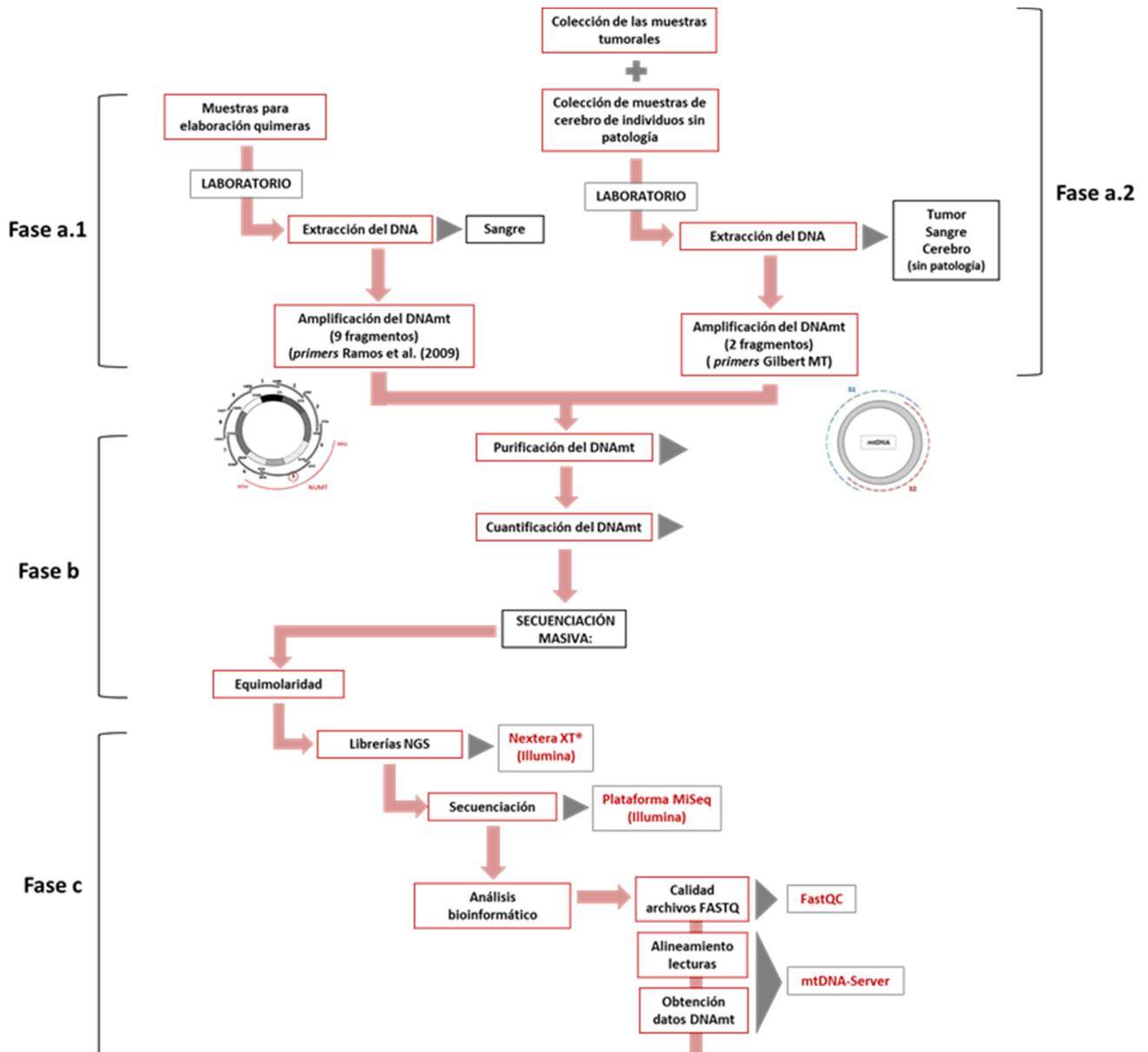


Figura 5. A continuación.

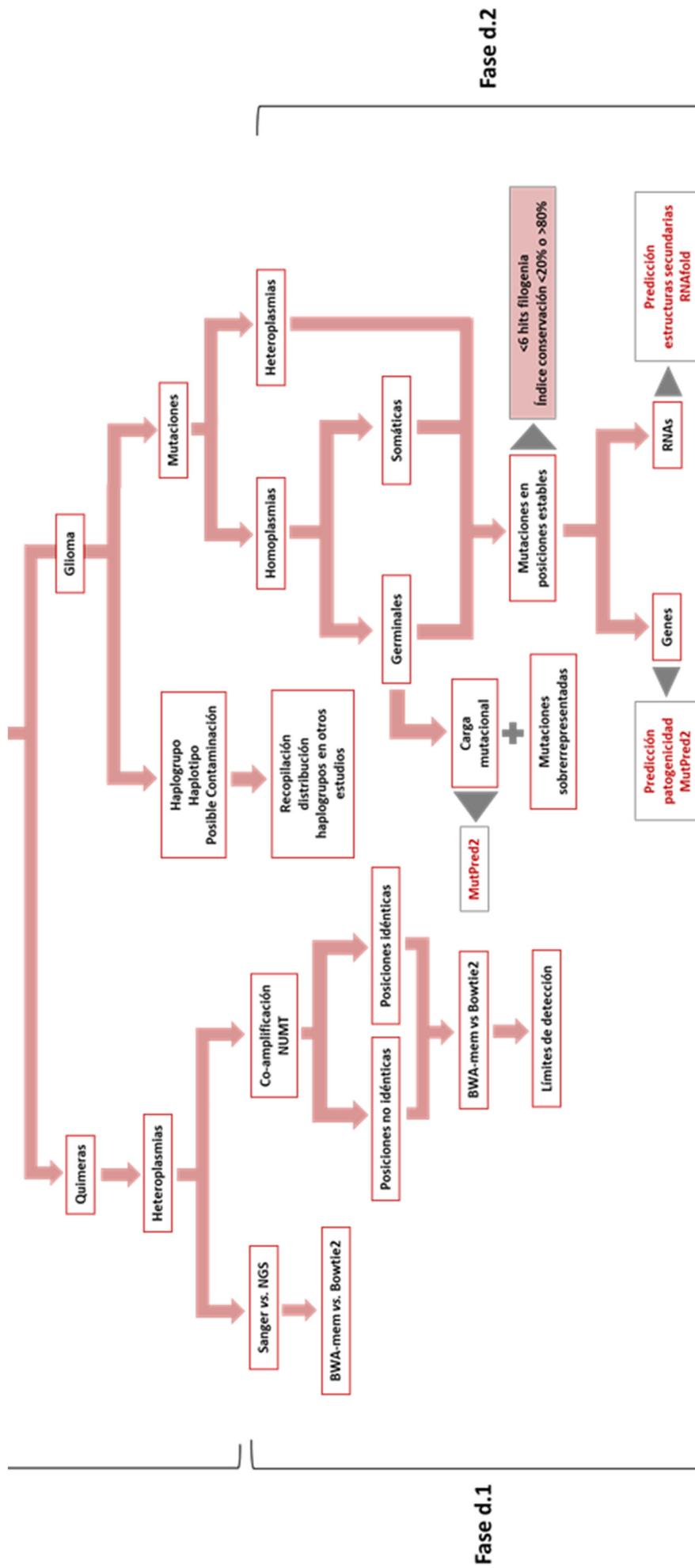


Figura 5. Fases del proceso seguido para el análisis de las mutaciones mitocondriales detectadas en las muestras estudiadas. Consta de 3 fases: a.1) Obtención de muestras de sangre para la creación de las quimeras, extracción del DNA y amplificación por fragmentos del DNAm; a.2) Obtención de las muestras tumorales y sanguíneas de pacientes afectados de glioma y de las muestras de cerebro de individuos sin patología. Extracción del DNA y amplificación del DNAm por fragmentos; b) Purificación, cuantificación del DNAm y equimolaridad de los fragmentos amplificados; c) Elaboración de librerías, secuenciación por NGS y análisis bioinformático para todas las muestras; d.1) Análisis de los datos obtenidos para las quimeras; y d.2) Análisis de los datos obtenidos para las muestras tumorales y sanguíneas de pacientes afectados de glioma.

3.1. SELECCIÓN de las MUESTRAS

3.1.1. Formación de quimeras a partir de muestras de sangre de diferentes individuos (Figura 5; Fase a.1)

Se generaron 2 genomas mitocondriales artificiales o quimeras (Q1 y Q2) con la finalidad de tener representado en dichos genomas varias posiciones en heteroplasmia.

Se utilizaron muestras de sangre de 8 individuos, de origen español y sin relación materna, para los que previamente se secuenció y caracterizó el genoma mitocondrial completo mediante secuenciación de Sanger (Ramos et al. 2013). El estudio y el consentimiento informado por escrito fueron aprobados por el comité de ética de la Junta de Atención Especializada en el Complejo de Salud de Zamora y autorizados por su Director Médico, así como por el Comité de ética de la Universitat Autònoma de Barcelona (UAB). Todos los participantes firmaron el consentimiento informado.

En la Q1 se mezclaron fragmentos mitocondriales de 2 individuos distintos (Tabla 6), mientras que en la Q2 se mezclaron fragmentos mitocondriales de 6 individuos (Tabla 7).

Tabla 6. Fragmentos mitocondriales de diferentes individuos utilizados para la generación de la quimera 1 (Q1). Se describen las mutaciones que se detectaron mediante la secuenciación de Sanger (Ramos et al. 2013).

Fragmentos de DNAm ^t	Código del individuo	Haplogrupo	Mutaciones detectadas mediante la secuenciación de Sanger (Ramos et al. 2013)
Fragmento 1 14898-151	Z90	H1	16209, 16519
Fragmento 2 16488-1677	Z90	H1	16519, 185, 263, <i>Het Poly-C 303-309</i> , 750, 1438
Fragmento 3 1404-3947	Z90	H1	1438, <i>Het 2887Y</i> , 3010, 3107d
Fragmento 4 3734-6739	Z225	U	4639, 4769, <i>Het 5306Y</i>
Fragmento 5 6511-9220	Z90	H1	<i>Length Het 8289-8290CCCCCTCTA</i> , 8860
Fragmento 6 8910-10648	Z90	H1	
Fragmento 7 10360-12226	Z90	H1	11215
Fragmento 8 11977-13830	Z90	H1	12810
Fragmento 9 13477-15349	Z90	H1	15326

Tabla 7. Fragmentos mitocondriales de diferentes individuos utilizados para la generación de la quimera 2 (Q2). Se describen las mutaciones que se detectaron mediante la secuenciación de Sanger (Ramos et al. 2013).

Fragmentos de DNAm ^t	Código del individuo	Haplogrupo	Mutaciones detectadas mediante la secuenciación de Sanger (Ramos et al. 2013)
Fragmento 1 14898-151	Z75	J1b	73, 15326, 15452A, 16069, 16126, 16145, 16162, Het Poly-C 16184-16193 , 16261, 16286
Fragmento 2 16488-1677	Z144	K1a4a1	73, 263, 309.1C, 315.1C, 497, Length Het Poly AC 514-525 , 750, 1189, 1438, 16519
Fragmento 3 1404-3947	Z144	K1a4a1	1438, 1811, 2706, 3480
Fragmento 4 3734-6739	Z31	V	Het 4055Y , 4580, 4769
Fragmento 5 6511-9220	Z24	HV7	7028, Het 7912R , 8380, 8860
Fragmento 6 8910-10648	Snb61	U5b2a1a1	Het 9449Y , 9477
Fragmento 7 10360-12226	Snb61	U5b2a1a1	Het 11016R , 11467, 11719
Fragmento 8 11977-13830	Snb48	K2b1a1	12308, 12372, 13135, Het 13590R
Fragmento 9 13477-15349	Snb61	U5b2a1a1	13617, 13637, 14182, 14766, Het 15244R , 5326

3.1.2. Recogida de muestras tumorales y sanguíneas de pacientes afectados de glioma (Figura 5; Fase a.2)

Un total de 32 pacientes que se diagnosticaron con glioma no recidivo, atendidos en el Servicio de Neurología del Hospital de Bellvitge (Hospitalet de Llobregat, Barcelona, España) han sido incluidos en el estudio. Se obtuvieron datos clínicos y radiológicos por resonancia magnética (RM) de todos los pacientes. Más concretamente, los datos clínicos incluyeron síntomas clínicos y la capacidad de los pacientes para realizar tareas normales a partir de la escala de Karnofsky (KPS), tanto en el periodo de ingreso como al recibir el alta, mientras que los radiológicos por RM incluyeron el tamaño de la lesión (mm), la presencia de lesión multifocal y la localización de la lesión. Dicha información se utilizó para determinar el tipo y la afectación del glioma cerebral en el *Institut de Diagnòstic per la Imatge Duran y Reynals*, asociado al Hospital de Bellvitge (Hospitalet de Llobregat, Barcelona, España). Para la correcta clasificación de los gliomas, en aquellos casos en los que se consideró necesario, se les efectuó un estudio por resonancia magnética de difusión y perfusión cerebrales, así como una espectroscopia (Fayed-Miguel et al. 2006). Además, cada uno de los 32 tumores se clasificó según su

tipología a partir de los análisis de anatomía patológica llevados a cabo y por su grado de malignidad según la Organización Mundial de la Salud (OMS) (Louis et al. 2007).

Una vez los pacientes fueron informados, se procedió a la recogida de las muestras en el momento de la intervención quirúrgica para la extirpación del tumor glial. Previamente, se solicitó el consentimiento informado de los participantes conforme la Declaración de Helsinki y atendiendo a lo dispuesto en la Ley 14/2007 del 3 de julio de Investigación Biomédica. Asimismo, el protocolo que se siguió fue sometido al Comité de Ética del Hospital de Bellvitge y de la UAB para su aprobación. De cada paciente se obtuvieron 1 muestra de tejido tumoral y 1 muestra de tejido sanguíneo.

3.1.3. Muestras de cerebro de individuos sin patología (Figura 5; Fase a.2)

Se obtuvo 1 muestra de tejido cerebral de 10 individuos adultos sanos de edad comprendida entre los 35 y los 70 años y que no padecían aparentemente ninguna afección neurológica en el momento de su muerte. Estas muestras habían sido depositadas previamente en el *Banco de Tejidos del Hospital Clínico y Provincial de Barcelona*. Asimismo, para tener una mayor robustez en los análisis, a los resultados obtenidos a partir de las 10 muestras mencionadas se le añadieron 142 muestras de cerebro de individuos que no padecieron cáncer, previamente analizadas por Li et al. (2015a).

3.2. EXTRACCIÓN y AMPLIFICACIÓN del DNAMt

3.2.1. Quimeras a partir de muestras de sangre de diferentes individuos (Figura 5; Fase a.1)

El DNA de las muestras de sangre utilizadas para la elaboración de las quimeras fue extraído mediante el Kit *JETQUICK Blood DNA Spin* (Genomed) para su posterior amplificación.

Para poder generar los 2 genomas mitocondriales artificiales, el DNAMt se amplificó mediante 9 fragmentos de PCR (*Polymerase Chain Reaction*) solapantes, utilizando los *primers* y condiciones previamente descritos por Ramos et al. (2009) (Tabla 8). En ambas quimeras, la región correspondiente al fragmento 5 (6511-9220) fue amplificada por la pareja de *primers* que se muestra en la Tabla 8, que, tal y como describe Ramos et al. (2011), podrían potencialmente co-amplificar una parte de un NUMT localizado en el cromosoma 1 (GRCh38_Chr.1 629081-634924). Dicho NUMT presenta una identidad del 98% con la correspondiente región mitocondrial (región mitocondrial 3911-9755).

Tabla 8. *Primers* utilizados en la amplificación del DNA mitocondrial en 9 fragmentos solapantes (Ramos et al. 2009). pb: pares de bases; Ta: temperatura de *annealing*. Tm: temperatura de *melting*. Tabla modificada de Ramos et al. (2009).

Fragmento de DNAmt	Tamaño del fragmento (pb)	Primers	Secuencia (5'-3')	Longitud del primer (pb)	Ta	Tm
1	1822	14898F	TAGCCATGCACTACTCACCAGA	22	60	60,3
		151R	GGATGAGGCAGGAATCAAAGAC	22		
2	1758	16488F	CTGTATCCGACATCTGGTTCCT	22	60	60,3
		1677R	GTTTAGCTCAGAGCGGTCAAGT	22		
3	2543	1404F	ACTTAAGGGTCGAAGGTGGATT	22	57	58,4
		3947R	TCGATGTTGAAGCCTGAGACTA	22		
4	3005	3734F	AAGTCACCCTAGCCATCATTCTA	23	61	58,9
		6739R	GATATCATAGCTCAGACCATAACC	23		
5	2709	6511F	CTGCTGGCATCACTATACTACTA	23	58	58,9
		9220R	GATTGGTGGGTCATTATGTGTTG	23		
6	1738	8910F	CTTACCACAAGGCACACCTACA	22	61	60,3
		10648R	GGCACAATATTGGCTAAGAGGG	22		
7	1866	10360F	GTCTGGCCTATGAGTGACTACA	22	61	60,3
		12226R	CAGTCTTGAGCTTTCTCGG	22		
8	1853	11977F	CTCCCTCTACATATTTACCACAAC	24	63	59,3
		13830R	AAGTCCTAGGAAAGTGACAGCGA	23		60,6
9	1872	13477F	GCAGGAATACCTTTCCTCACAG	22	63	60,3
		15349R	GTGCAAGAATAGGAGGTGGAGT	22		

Las condiciones utilizadas para la amplificación de cada uno de los fragmentos fueron: 50pmol de cada *primer*, 200mM de cada dNTP, 2mM de MgCl₂, 10NH₄-based de *buffer*, 1 U de *BIOTAQ DNA polymerase* (Ecogen), y 30 ng de DNA en un volumen final de 50 µl. Las amplificaciones se llevaron a cabo mediante el termociclador S1000™ *Thermal Cycler* (BioRad). El programa para la amplificación consta de un paso de desnaturalización de 5 minutos, seguido de 35 ciclos de PCR –1 minuto a 94°C, 40 segundos a la temperatura de *annealing* (Ta) (Tabla 8) y 2,5 minutos a 72°C)– y un paso de extensión final de 5 minutos a 72°C. La correcta amplificación de los fragmentos se visualizó mediante electroforesis en geles de agarosa al 1,5%.

3.2.2. Tumores y sangre de pacientes afectados de glioma y muestras de cerebro de individuos sin patología (Figura 5; Fase a.2).

Tanto las muestras tumorales como las de tejido cerebral de individuos sin patología fueron previamente limpiadas con SSF para eliminar los restos de sangre y congeladas a -80°C . La extracción del DNA de las muestras tumorales y de sangre obtenidas de cada paciente, así como las muestras obtenidas de tejido cerebral de individuos sin patología, se realizó con el kit QIAamp DNA Investigator Kit (Qiagen), siguiendo las especificaciones indicadas por el fabricante para cada uno de los tejidos.

Una vez extraído el DNA, se procedió a amplificar la totalidad de la molécula de DNAm_t mediante 2 fragmentos solapantes (Set 1 –posición 17 a 9050– y Set2 –posición 8537 a 937–), usando los *primers* utilizados por Mikkelsen et al. (2009) (Tabla 9).

Las condiciones utilizadas para la PCR fueron: 10pmol de cada *primer*, 10mM de cada dNTP, 2mM de MgSO_4 , 5 NH_4 -based de *buffer*, 0,2 U de *Platinum™ Taq DNA polymerase High Fidelity* (Invitrogen), y 30 ng de DNA en un volumen final de 50 μl . Las amplificaciones se llevaron a cabo mediante el termociclador S1000™ *Thermal Cycler* (BioRad). El programa para la amplificación consta de un paso de desnaturalización de 2 minutos, seguido de 35 ciclos de PCR –30 segundos a 94°C , 30 segundos a la temperatura de *annealing* (T_a) (59°C) y 10 minutos a 68°C –, y un paso de extensión final de 7 minutos a 72°C .

En las muestras tumorales con problemas de degradación del DNA la región que comprende el Set 1 (17-9050) se amplificó en 7 fragmentos solapantes más pequeños (Fragmentos 2, 3.A, 3.B, 4.A, 4.B, 5.A y 5.B de la Tabla 9), mientras que la región que comprende el Set 2 (8537-937) se amplificó en 2 fragmentos más pequeños (Fragmentos Set2.A y Set2.B de la Tabla 9). Para estas amplificaciones se utilizaron tanto *primers* diseñados por Ramos et al. (2009) como *primers* diseñados *de novo*, siguiendo la misma estrategia utilizada por Ramos et al. (2009) para evitar la co-amplificación con NUMTs (Tabla 9).

Los resultados de las PCRs se visualizaron mediante electroforesis con geles de agarosa al 1% o al 1,5%.

Tabla 9. *Primers* utilizados para la amplificación del DNAm_t en fragmentos solapantes. *Fragmentos utilizados para la amplificación de DNAm_t en aquellas muestras en las que era posible obtener la totalidad de la molécula en 2 fragmentos. En aquellas muestras en que no se podía obtener el DNAm_t en 2 fragmentos, se amplificaron fragmentos más pequeños.

Fragmentos de DNAm _t	Primer Sentido directo 5'-3'	Primer Sentido reverso 5'-3'	Referencia
SET 1*	posición 17	posición 9050	(Mikkelsen et al. 2009)
	CCCTATTAACCACTCACGGGAGC	CCAATTAGGTGCATGAGTAGGTGG	
SET 2*	posición 8537	posición 937	(Mikkelsen et al. 2009)
	ATCTGTTTCGCTTCATTGCCCC	ACGCCGGCTTCTATTGACTTGGG	
FRAGMENTO 2	posición 16488	posición 1677	(Ramos et al. 2009)
	CTGTATCCGACATCTGGTTCCT	GTTTAGCTCAGAGCGGTCAAGT	
FRAGMENTO 3.A	posición 1404	posición 2801	Presente trabajo
	ACTTAAGGGTCGAAGGTGGATT	TAATGCAGGTTTGGTAGTTTA	
FRAGMENTO 3.B	posición 2646	posición 4421	Presente trabajo
	GGTTCAGCTGTCTCTTACTTT	CTTATTTAGCTGACCTTACTT	
FRAGMENTO 4.A	posición 3734	posición 5017	(Ramos et al. 2009)
	AAGTCACCCTAGCCATCATTCTA	GAGGAGTATGCTAAGATTTTG	
FRAGMENTO 4.B	posición 4914	posición 6739	(Ramos et al. 2009)
	CATATACCAAATCTCTCCCTCAC	GATATCATAGCTCAGACCATAACC	
FRAGMENTO 5.A	posición 6511	posición 8167	(Ramos et al. 2009)
	CTGCTGGCATCACTATACTACTA	CTCCACAGATTCAGAGCATTGA	
FRAGMENTO 5.B	position 7713	position 9647	(Ramos et al. 2009)
	TCCTAACACTCACAACAAAAC	AGCTCAGGTGATTGATACTCC	
Set2.A	posición 8537	posición 12226	Presente trabajo
	ATCTGTTTCGCTTCATTGCCCC	CAGTTCTTGAGCTTTCTCGG	
Set2.B	posición 11977	posición 937	Presente trabajo
	CTCCCTCTACATATTTACCACAAC	ACGCCGGCTTCTATTGACTTGGG	

3.3. PURIFICACIÓN, CUANTIFICACIÓN y EQUIMOLARIDAD del DNAm_t (Figura 5; Fase b)

Una vez los fragmentos de DNAm_t fueron amplificados, se purificaron mediante el kit *NucleoSpin Gel and PCR Clean-up* (Macherey-Nagel). A continuación, se cuantificaron por fluorescencia mediante el fluorómetro Qubit[®] 2.0 y utilizando el kit *Qubit dsDNA HS Assay*. Esta tecnología es altamente selectiva para la detección del DNA de doble cadena, de manera que se descarta la posible cuantificación de cadena sencilla (RNA o *primers*) y se consigue una cuantificación más precisa.

Con la finalidad de tener representada la misma cantidad de DNAm_t de todos los fragmentos amplificados para cada muestra, a partir de la cuantificación del DNAm_t, se procedió a la mezcla de los

fragmentos en cantidades equimolares. Los fragmentos se diluyeron a una concentración de 0,2 ng/ μ l y, posteriormente, se mezclaron según la cantidad de picomoles presentes en cada uno de ellos. La cantidad de picomoles se calculó a partir del tamaño del fragmento y la cantidad de DNAm disponible mediante el programa *Biomath Calculators* de Promega (Promega Corp. 2019).

3.4. PREPARACIÓN de las LIBRERÍAS y posterior SECUENCIACIÓN de NGS (Figura 5; Fase c)

Para todas las muestras se realizaron las librerías a partir de la mezcla equimolar de los fragmentos amplificados. En el caso de Q1 y Q2, se realizaron 2 y 4 librerías independientes, respectivamente. Para las muestras de glioma y de cerebro de individuos sanos se realizó una librería y se repitió en caso de obtener una baja calidad.

Para la preparación de las librerías de NGS se utilizó el kit *Nextera® XT DNA Library Prep* (Illumina), que requiere únicamente 1ng de DNA de partida, siguiendo los pasos que se describen a continuación:

- *Tagmentación*. La tecnología *Nextera®* se basa en el uso de transposomas, complejos formados por una transposasa (Tn5) y un transposón, que se encargan de fragmentar de forma simultánea el DNA inicial, dando lugar a roturas de doble cadena (Figura 6). Durante este proceso, el extremo 3' del transposón se une mediante unos adaptadores al extremo 5' del DNA (Schirmer et al. 2016).
- *PCR de ciclos limitados*. Durante la PCR, el DNA se amplifica y se le añaden unos índices o etiquetas en el extremo 3' necesarios para la posterior formación de los *clusters*. La PCR de ciclos limitados consta de 2 pasos de 3 minutos a 72°C y de 30 segundos a 95°C respectivamente, seguido de 12 ciclos de PCR –10 segundos a 95°C, 30 segundos a 55°C, y 30 segundos a 72°C–, y un último paso de 5 minutos a 72°C (Figura 6).
- *Purificación y normalización de las librerías*. Con el objetivo de eliminar aquellos fragmentos demasiado cortos, se lleva a cabo un proceso de purificación. Posteriormente, las librerías se normalizan para conseguir una representación equilibrada de cada una de ellas respecto al total de librerías.
- *Formación de clusters y posterior secuenciación mediante la plataforma de secuenciación MiSeq (Illumina)*.

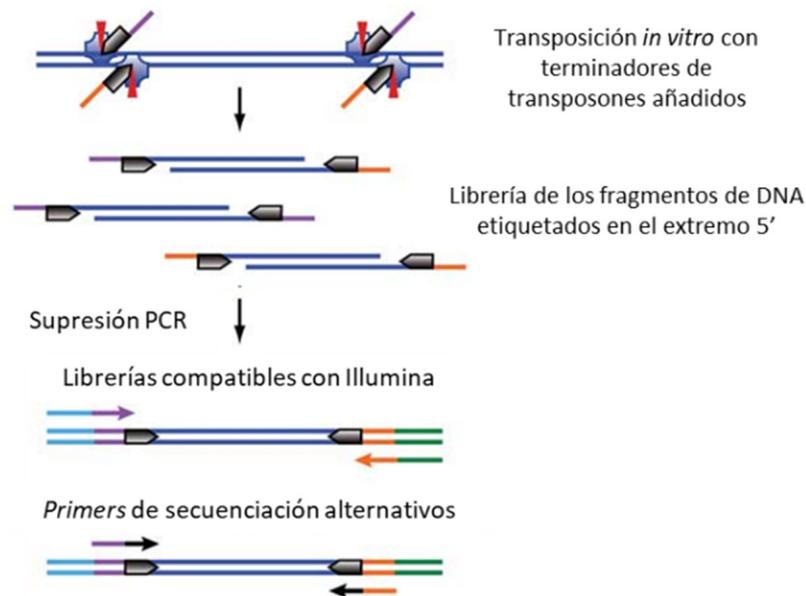


Figura 6. Fase de tagmentación del DNA mediante el uso de transposomas. Los transposomas fragmentan el DNA de forma aleatoria y se añaden unos adaptadores en cada uno de los fragmentos. Posteriormente, se realiza una PCR de ciclos limitados para añadir los índices. Figura modificada de Syed et al. (2009).

La secuenciación de todas las librerías se llevó a cabo mediante 5 *runs* independientes de 24 muestras cada *run* en la plataforma MiSeq (Illumina) generando lecturas de extremos apareados (2x250pb), es decir, lecturas de 250pb tanto en sentido directo como reverso. En este caso, se realiza la secuenciación de un extremo, se produce el giro de la molécula y, a continuación, se realiza la secuenciación del otro extremo.

3.5. ANÁLISIS BIOINFORMÁTICO del DNAMt (Figura 5; Fase c)

3.5.1. Calidad de los archivos Fastq

Previo análisis de los datos obtenidos mediante NGS, se procedió a un análisis de la calidad de los archivos obtenidos directamente de la secuenciación masiva. Dichos datos tienen un formato Fastq (.fastq), una simple extensión del formato FASTA con la capacidad de almacenar un valor de calidad para cada base de la secuencia (Cock et al. 2010). En este caso, al generar lecturas de extremos apareados (2x250pb), se han obtenido 2 archivos Fastq para cada muestra: Fastq-R1 (en referencia al primer ciclo de lecturas, en sentido directo y reverso, de la cadena de DNAMt) y Fastq-R2 (en referencia al segundo ciclo de lecturas, en sentido directo y reverso, de la cadena complementaria de DNAMt).

Para su análisis se ha utilizado la aplicación java FastQC (Andrews 2010) que permite llevar a cabo un control de calidad de los datos brutos que provienen de la secuenciación masiva, proporcionando una visión general de los datos e indicando qué parámetros están sujetos a más problemas.

En el caso de las muestras amplificadas mediante los Sets 1 y 2, se comprobó si el patrón de cobertura (*coverage*) obtenido estaba relacionado con la distribución del contenido de GC a lo largo del genoma mitocondrial. El DNAm se dividió en regiones de 250pb y, para cada una de ellas, se calculó el porcentaje (%) de guanina-citosina (GC) y la profundidad de lectura promedio. A continuación, se calculó el coeficiente de correlación lineal de *Pearson* utilizando el programa SPSS versión 19.0. (IBM Corp. 2010), considerando un valor de significancia de 0,05.

3.5.2. Análisis de Fastq con mtDNA-Server

Para el análisis de los datos en Fastq, se aplicó el proceso bioinformático (*pipeline*) implementado en el programa mtDNA-Server v.1.0.5. (Weissensteiner et al. 2016a). Dicho procedimiento se basa en diferentes fases (Figura 7):

- *Alineamiento de las lecturas.* Las lecturas que se obtienen se alinean mediante la herramienta *BWA-mem* (*Burrow-Wheeler Alignment - maximal exact match*) v.0.7.5 (Li 2013), diseñado para alinear lecturas de un tamaño entre 70 pb y 1 Mb.
- *Control de calidad.* A partir de los archivos Bam obtenidos se analiza la calidad de los alineamientos a partir de diferentes parámetros relacionados con el mapeado de las lecturas, la detección de duplicados o problemas con la secuencia de referencia.
- *Análisis de los archivos Bam y detección de variantes.* Para la detección de variantes, se utiliza el *software HadoopBAM* para dividir los archivos Bam. Para cada fragmento, se descartan aquellas lecturas que tienen una calidad de mapeo <20, una longitud <25 (Zhidkov et al. 2011), y aquellas detectadas como duplicadas. Además, también se excluyen aquellas lecturas con una calidad de alineamiento ≤ 30 y se aplica a todas las lecturas una calidad de alineamiento por base (BAQ) con la herramienta *Genome Analysis Toolkit* (GATK) (McKenna et al. 2010), a fin de poder trabajar con la circularidad del genoma mitocondrial. Posteriormente, para todas las lecturas que pasan los filtros anteriores, se descartan aquellas en las que todas sus bases tienen un *Phred Score* (probabilidad de que una base sea incorrecta) <20. Por último, se realiza un recuento de todas

las bases para cada una de las posiciones –A, C, G, T, N (base desconocida) y d (delección)– y se obtienen las mutaciones presentes en las muestras analizadas.

- *Detección de heteroplasmas*. Para todas las posiciones en las que se observa mezcla de bases, se eliminan aquellas que muestran un *coverage* <10 por cadena. Para el resto de posiciones, aquellas que muestran una frecuencia del alelo minoritario (*minor allele frequency* (MAF)) ≥ 1 por cada cadena independientemente y una cobertura alélica de mínimo de 5 bases por cadena, se les aplica un modelo de máxima verosimilitud (ML) (Ye et al. 2014b). Este modelo tiene en cuenta errores de secuenciación y se aplica para cada cadena independientemente, de forma que todas las posiciones con un logaritmo ML ≥ 5 son identificadas como heteroplásmicas. Además, se seleccionan aquellas posiciones con un sesgo entre cadenas <1 (Guo et al. 2012a; Guo et al. 2012b) y se calcula el intervalo de confianza *Agresti-Coull* para las posiciones heteroplásmicas finales (Calabrese et al. 2014). El nivel de heteroplasma asignado es la media ponderada de heteroplasma de la cadena en sentido directo y reverso. Por último, se indican aquellas posiciones heteroplásmicas que se sitúan en zonas homopoliméricas (LCR) (Zhidkov et al. 2011), así como los NUMTs de naturaleza polimórfica reportados por Dayama et al. (2014) que pueden ser responsables de la mezcla de base observada. De esta forma, se obtienen las heteroplasmas presentes en las muestras analizadas.
- *Detección de haplogrupo y contaminación*. La asignación del haplogrupo se hace de acuerdo a la filogenia de DNAMt disponible en el Phylotree (mtDNA tree Build 16) (van Oven and Kayser 2009) utilizando el programa Haplogrep (Weissensteiner et al. 2016b). Para comprobar si las mezclas de bases que se detectan no son debido a una posible contaminación, se generan 2 perfiles basados en la frecuencia de las variantes alélicas –uno menor (<50%) y uno principal (>50%)–. Si se detecta la presencia de diferentes secuencias mitocondriales, el programa Haplogrep muestra perfiles que dan lugar a diferentes haplogrupos.

Para este estudio, los hotspots (puntos calientes) mitocondriales –posiciones 309-315 y 3107– descritos en la rCRS (*Revised Cambridge Reference Sequence*) (Andrews et al. 1999) y los InDels no fueron analizados debido a la dificultad de obtener una profundidad de lectura y una calidad de alineamiento suficientemente elevadas como para obtener unos resultados de confianza utilizando esta metodología.

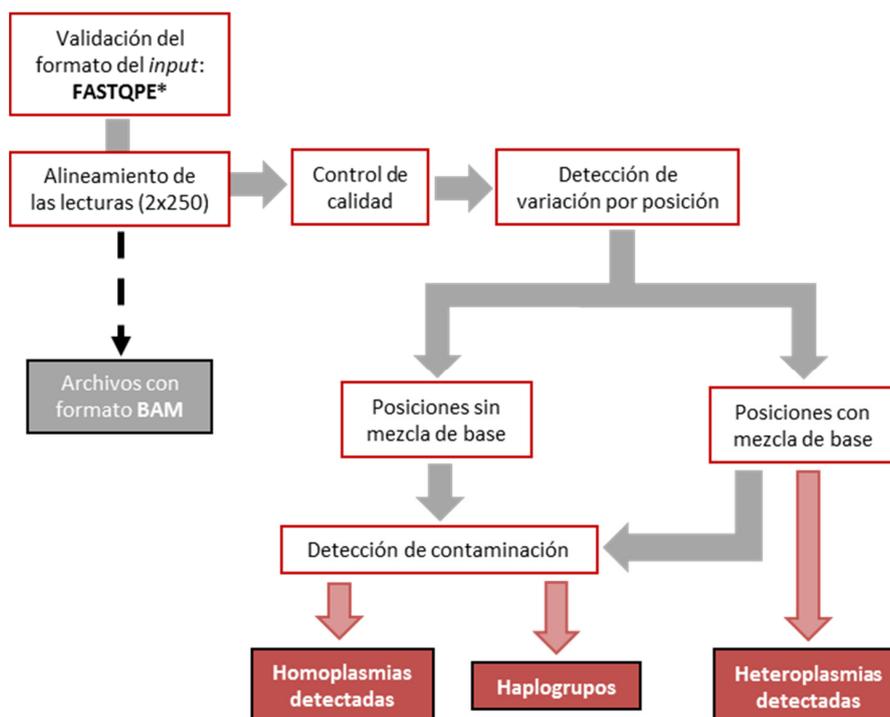


Figura 7. Esquema del proceso bioinformático realizado por el programa mtDNA-Server. Imagen modificada de Weissensteiner et al. (2016a). *Corresponde a Fastq *paired-end* (lecturas de extremos apareados).

3.6. ANÁLISIS de VARIANTES DETECTADAS en las QUIMERAS (Figura 5; Fase d.1)

3.6.1. Secuenciación de Sanger vs. Secuenciación de NGS

A partir de los datos obtenidos de Q1 y Q2 mediante NGS, se analizaron las mutaciones detectadas tanto a nivel de homoplasma como de heteroplasma y se contrapusieron con los resultados obtenidos mediante la secuenciación Sanger, previamente descritos por Ramos et al. (2013). Para estos análisis, las posiciones localizadas en las regiones solapantes entre los fragmentos fueron excluidas.

3.6.2. Análisis de las mezclas de bases a partir de la posible co-amplificación entre el DNAm_t y el NUMT

De la amplificación de la región correspondiente a las posiciones 6511-9220 del DNAm_t (fragmento 5, véase apartado 3.2.1.), se espera co-amplificación de la región nuclear –posiciones 631683 a 634389 del cromosoma 1 (GRCh38)–. Así, las posiciones no idénticas entre el segmento del DNAm_t y su correspondiente en el DNAn (NUMT) darán lugar a mezclas de bases con una proporción más o menos elevada, dependiendo del grado de co-amplificación entre el DNAm_t y el DNAn. Así pues,

la identificación de posiciones tanto idénticas como no idénticas entre el DNAmT y el NUMT se realizó a partir del alineamiento de las secuencias anteriormente obtenidas por Ramos et al (2013) para los individuos que configuran Q1 y Q2 –utilizando primers que evitan la co-amplificación de NUMTs– , con la secuencia del NUMT –posiciones 631910 a 634079 del cromosoma 1 (GRCh38)–, utilizando el programa BioEdit (Hall 1999) (Figura 1 del Anexo II). Las posiciones heteroplásmicas mitocondriales previamente confirmadas por Ramos et al. (2013) no se incluyeron para el presente análisis y se excluyeron las regiones solapantes entre los fragmentos F4-F5 y F5-F6 (véase Tabla 7), lo que representa el análisis de la región comprendida entre las posiciones 6739 y 8910 del DNAmT.

Para determinar si la identificación de mezcla de bases en posiciones no idénticas depende de la profundidad de lectura obtenida, se llevó a cabo un análisis de correlación de *Pearson* entre la profundidad de lectura y el porcentaje de mezcla de base detectado. Asimismo, se calculó el posible sesgo entre cadenas (valores >1) para detectar posibles resultados erróneos en posiciones no idénticas entre el mtDNA y el NUMT (Guo et al. 2012a; Guo et al. 2012b).

En el caso de mezclas de bases localizadas en posiciones idénticas entre el mtDNA y el NUMT, se analizó la posible variabilidad del NUMT utilizando los datos de los 1000 Genomas (Auton et al. 2015), con el propósito de discriminar si estas mezclas de bases pueden ser verdaderas heteroplasmias o son, en realidad, posiciones no idénticas debido a la posible presencia de posiciones polimórficas en los genomas nucleares de los individuos utilizados en la Q1 y Q2.

3.6.3. Alineamiento BWA-mem vs. Alineamiento Bowtie2

A fin de determinar posibles diferencias en los resultados obtenidos en las quimeras según el tipo de alineamiento aplicado, se llevó a cabo un alineamiento alternativo mediante la herramienta *Bowtie2* para todos los fragmentos mitocondriales secuenciados para Q1 y Q2. Tanto *BWA-mem* como *Bowtie2* sirven para alinear lecturas de gran longitud (>50pb) y se basan en un proceso de indexación del genoma de referencia para facilitar el alineamiento de las lecturas, agilizando así el proceso de alineamiento y reduciendo la memoria necesaria (Langmead et al. 2009; Li 2012). No obstante, existen ciertas diferencias en cuanto a los algoritmos aplicados. *Bowtie2* alinea las lecturas respecto al genoma de referencia mediante el algoritmo de coincidencia *Ferragina and Manzini* (Ferragina and Manzini 2000), incluyendo modificaciones para poder tener en cuenta errores de secuenciación y variaciones genéticas, y presenta un algoritmo de calidad para las lecturas alineadas. Además, permite la presencia

de ciertos desajustes y brechas (*gaps*), sobre todo hacia el final de la lectura, incrementado el límite de calidad en estos casos (Langmead and Salzberg 2012; Langmead et al. 2009). *BWA-mem* se basa en un algoritmo donde la coincidencia es exacta (Li 2012) y en el algoritmo modificado *Smith-Waterman* para aquellas lecturas que no han podido ser mapeadas (Li 2013).

3.7. ANÁLISIS de VARIANTES MITOCONDRIALES DETECTADAS en PACIENTES AFECTOS de GLIOMA (Figura 5; Fase d.2)

3.7.1. Variación del DNAMt en individuos afectados de glioma

A partir de los datos obtenidos por NGS se determinó el haplotipo mitocondrial para cada tumor y muestra de sangre de cada uno de los pacientes analizados, considerando tanto las mutaciones detectadas en homoplasmia como las detectadas en heteroplasmia, y se estableció el haplogrupo mitocondrial en cada caso. Además, tal y como se ha comentado anteriormente, se llevó a cabo la detección de posible contaminación utilizando el programa Haplogrep (Weissensteiner et al. 2016a; Weissensteiner et al. 2016b). Asimismo, se clasificaron todas las mutaciones detectadas, tanto en homoplasmia como en heteroplasmia, en somáticas –cuando la mutación se detecta únicamente en el tejido tumoral– o germinales –cuando la mutación se detecta tanto en tejido tumoral como sanguíneo–. Respecto a las heteroplasmas, a nivel somático, se consideraron únicamente aquellas que tenían una frecuencia $>3\%$. A nivel germinal, se tuvieron en cuenta aquellas que se detectaban con una frecuencia $>3\%$ en al menos 1 de los 2 tejidos.

Para determinar si existe algún desequilibrio en la distribución de mutaciones a lo largo del genoma mitocondrial en el presente estudio respecto de la población general, se llevó a cabo un test de proporciones mediante el programa Openstat (versión 2008), considerando un valor de significancia de 0,05. Para ello, se calculó el número de mutaciones detectadas en cada una de las regiones –la región control (HVRI, HVRII, HVRIII) y la región codificante (*RNAr 12S*, *RNAr 16S*, *RNAr* y subunidades de la cadena respiratoria)– tanto en el presente estudio como en la población general y se comparó la proporción encontrada entre ambos. Las mutaciones detectadas en la población general se obtuvieron a partir de los datos de 7605 genomas completos de individuos europeos de la base de datos *Human Mitochondrial Database* (HmtDB) (Clima et al. 2017). Asimismo, se aplicó la corrección de Bonferroni para corregir los sesgos generados por las comparaciones múltiples.

3.7.2. Distribución de haplogrupos mitocondriales en pacientes afectados de glioma

A partir de la determinación de los haplogrupos mitocondriales de los pacientes afectados de glioma analizados en el presente estudio, se calculó la frecuencia de cada uno de los haplogrupos detectados, así como su respectivo intervalo de credibilidad bayesiana (IC 95%) utilizando el software SAMPLING (V. Macaulay, comunicación personal). La frecuencia de haplogrupos obtenida en los pacientes de glioma tanto a nivel global como según el grado de malignidad tumoral se comparó con la frecuencia en la población general española de la costa mediterránea (CCAA: Cataluña, Comunidad Valenciana y Murcia) (Barral-Arca et al. 2016), utilizando el test exacto de Fisher y la Odds Ratio (IC 95%). Para el análisis se ha utilizado el programa SPSS versión 19.0. (IBM Corp. 2010), considerando un valor de significancia de 0,05.

Compilación de la distribución de haplogrupos mitocondriales en pacientes afectados de glioma de poblaciones europeas

Para determinar una posible asociación entre alguno de los haplogrupos analizados y el riesgo de sufrir glioma, se compilaron los datos obtenidos de diferentes cohortes de pacientes europeos con glioma en los que se disponía de información respecto a la asignación de haplogrupo mitocondrial (Gasparre et al. 2007; Vega et al. 2004; Vidone et al. 2015), juntamente con los del presente estudio (Figura 10). Siguiendo la misma metodología antes descrita para la población analizada en la presente tesis, para cada uno de estos estudios se comparó la frecuencia de los haplogrupos detectados en pacientes con su frecuencia en la población general (Barral-Arca et al. 2016; Brisighelli et al. 2012) (Figura 8). Asimismo, para combinar en un mismo análisis todos los estudios disponibles se realizó una regresión logística binaria en la que se tuvo en cuenta la presencia o ausencia de glioma como variable dependiente, y las variables haplogrupo y población, como covariables categóricas. Para el análisis se ha utilizado el programa SPSS versión 19.0. (IBM Corp. 2010), considerando un valor de significancia de 0,05.

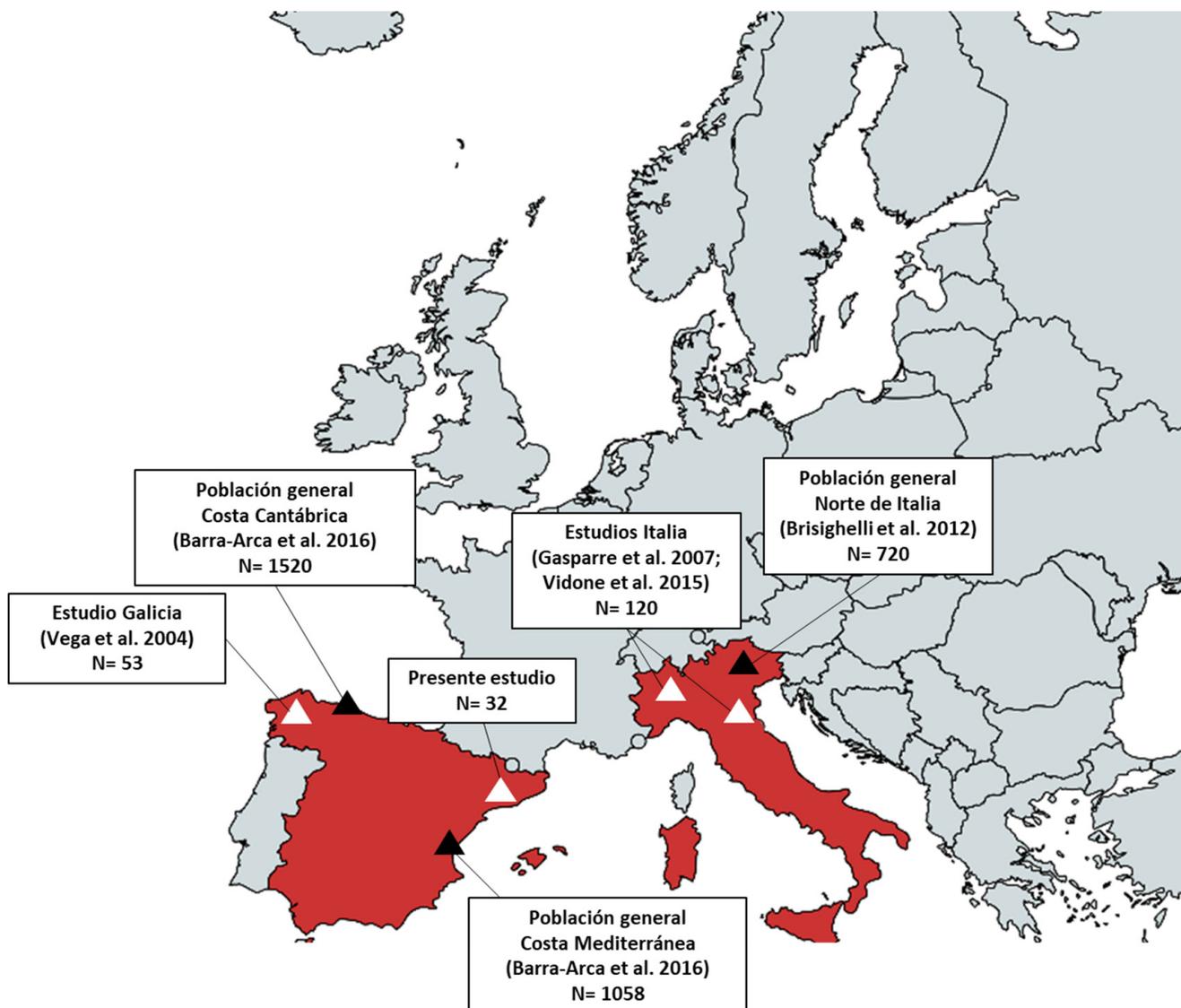


Figura 8. Mapa europeo donde se indica la procedencia de las cohortes europeas compiladas (triángulo blanco), así como el tamaño muestral en cada uno de ellos. Asimismo, se indican las poblaciones generales utilizadas (triángulo negro) para la comparación de las frecuencias de cada uno de los haplogrupos para cada cohorte de pacientes.

3.7.3. Análisis de mutaciones en homoplasmia

Distribución de las mutaciones en homoplasmia a lo largo del genoma mitocondrial

A partir de la detección de las mutaciones en homoplasmia, se determinó su distribución a lo largo del genoma mitocondrial –la región control (HVRI, HVRII, HVRIII) y la región codificante (RNAr 12S, RNAr 16S, RNAt y subunidades de la cadena respiratoria)– y se calculó la ratio de mutaciones por par de bases (pb) en cada una de las regiones mitocondriales (mutaciones/pares de bases de la región mitocondrial). Asimismo, para determinar si las mutaciones en homoplasmia se localizan de forma diferencial en alguna región mitocondrial, se realizó un test de proporciones mediante el programa Openstat (versión 2008), considerando un nivel de significancia de 0,05. Para ello, se compararon

entre sí todas las regiones mitocondriales comentadas anteriormente, considerando el número de mutaciones detectadas y el tamaño (pb) de la región. Se aplicó la corrección de Bonferroni para corregir los sesgos generados por las comparaciones múltiples.

Análisis de la carga mutacional

Para el cálculo de la carga mutacional de cada uno de los individuos, se seleccionaron todas las mutaciones en homoplasmia localizadas en genes mitocondriales que codifican para subunidades de la cadena respiratoria y que daban lugar a cambios a nivel aminoacídico. Para cada una de las mutaciones se calculó el nivel de patogenicidad mediante el programa MutPred2 (rango de 0 a 1) (Pejaver et al. 2017). Así, valores de MutPred2 $>0,5$ indican que la mutación probablemente tenga un efecto deletéreo sobre la función de las proteínas, mientras que valores $<0,3$ indican que la mutación probablemente no tenga un efecto deletéreo (Pienaar et al. 2017).

A continuación, para tener en cuenta el efecto que puede tener el número de variantes observadas en un individuo debido a la proximidad de su secuencia mitocondrial con la rCRS en la filogenia, se realizó un ajuste de la carga mutacional dividiendo la carga mutacional total de cada secuencia por el número de variantes no sinónimas (Pienaar et al. 2017). A partir de dicho cálculo, únicamente se tuvieron en cuenta para los posteriores análisis aquellas cargas mutacionales ajustadas que superaban el valor de 0,5. Tal y como presenta Venter et al. (2017), se ha de tener en cuenta que existen variantes comunes –variantes con una frecuencia $>1\%$ en la población– de bajo impacto que pueden diferir ampliamente entre poblaciones y su inclusión en el cálculo de la carga mutacional podría introducir cierto ruido con probablemente poco impacto fenotípico. De esta forma, se consigue destacar el impacto de las variantes poco frecuentes y reducir el efecto de la estratificación de la población.

En el caso de los RNAs, también se seleccionaron aquellas mutaciones localizadas en regiones que codifican para RNAt para determinar su carga mutacional ajustada. El nivel de patogenicidad en estos casos se calculó a partir de MitoTIP *tRNA Scoring* (Sonney et al. 2017). Valores $<8,44$ indican que las variantes son probablemente benignas, valores entre 8,44 y 12,66 indican que las variantes son posiblemente benignas, valores entre 12,66 y 16,25 indican que las variantes son posiblemente patogénicas y valores $>16,25$ indican que las variantes son probablemente patogénicas.

A fin de determinar si existe alguna diferencia en la carga mutacional teniendo en cuenta el grado tumoral, se llevó a cabo un análisis adicional comparando los tumores de menor malignidad (grado

II y III) con los tumores de mayor malignidad (grado IV) tanto a nivel de mutaciones en genes mitocondriales que codifican para subunidades de la cadena respiratoria como de mutaciones en los RNAs. Se calcularon los estadísticos descriptivos –media, mediana y desviación típica– para cada uno de los grupos y se comparó la carga mutacional ajustada detectada entre los grupos mediante el test de Mann-Whitney para ambas regiones mitocondriales, utilizando el programa SPSS versión 19.0. (IBM Corp. 2010) y considerando un nivel de significancia de 0,05.

Mutaciones en homoplasmia sobrerrepresentadas en glioma respecto a la población general

Para poder determinar si existen mutaciones en homoplasmia que se detectan con una frecuencia superior a la esperada, se comparó su frecuencia respecto a la encontrada en la población general mediante el Test exacto de Fisher, utilizando el programa SPSS versión 19.0. (IBM Corp. 2010) y considerando un nivel de significancia más estricto de 0,01, a fin de minimizar los posibles errores tipo 1. Los datos de la población general se obtuvieron de los 7605 genomas mitocondriales completos de individuos europeos de la base de datos HmtDB (Clima et al. 2017), ya antes mencionada. Para cada frecuencia obtenida se calcularon los intervalos de credibilidad bayesiana (IC 95%) utilizando el software SAMPLING (V. Macaulay, comunicación personal).

Las mutaciones que estaban sobrerrepresentadas en el estudio se clasificaron en definitorias o no definitorias de haplogrupo. Respecto a las mutaciones sobrerrepresentadas, con una frecuencia >2 individuos, no definitorias de haplogrupo y presentes en la región no codificante, se calculó la Odds Ratio (IC 95%) utilizando el programa SPSS versión 19.0. (IBM Corp. 2010) y se llevó a cabo un estudio de predicción de su estructura secundaria y de la estabilidad de los enlaces (entropía) mediante el servicio web RNAfold (Gruber et al. 2008), utilizando los modelos estructurales de diferentes porciones de la región no codificante del DNAm descritos por Pereira et al. (2008).

Mutaciones en homoplasmia detectadas en posiciones estables

Paralelamente, a partir de la detección de las mutaciones en homoplasmia, se determinó la distribución de cada una de ellas según si se localizan en posiciones estables o no estables. Se consideraron posiciones estables aquellas que tenían <6 ocurrencias o hits en la filogenia de DNAm y un índice de conservación <20% o >80%. El número de ocurrencias o hits en la filogenia se calculó mediante el mtDNA tree Build 16 de Phylotree (van Oven and Kayser 2009). Los índices de conservación para la región codificante se calcularon a partir de 12 especies de primates (Figura 9) incluidas en la herramienta MitoMaster

(Brandon et al. 2009). En el caso de la región control, los índices se obtuvieron del GenBank (Benson et al. 2013) y se alinearon las secuencias de 48 especies de primates (Figura 10).

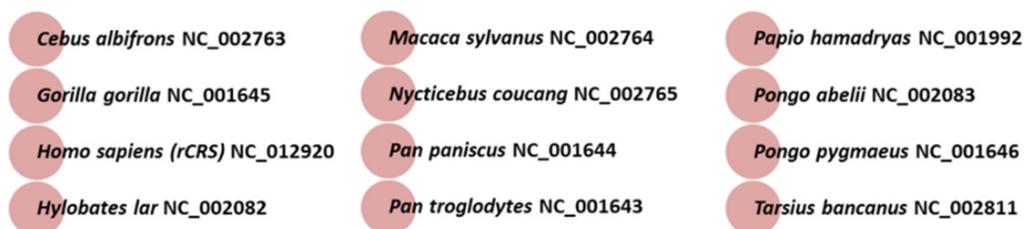


Figura 9. Especies de primates incluidas en la herramienta MitoMaster (Brandon et al. 2009) para el cálculo del índice de conservación de las mutaciones mitocondriales detectadas en la región codificante.



Figura 10. Especies de primates utilizadas para el cálculo del índice de conservación de las mutaciones mitocondriales detectadas en la región control

Para las mutaciones en homoplasia localizadas en posiciones estables, se llevó a cabo una representación gráfica de la distribución y frecuencia de cada una de ellas mediante el programa Geneious Prime® 2019.1.3, y se calculó la ratio de mutaciones (mutaciones/pares de bases (pb) de la región mitocondrial) en cada una de las regiones mitocondriales –la región control (HVRI, HVRII, HVRIII) y la región codificante (RNAr 12S, RNAr 16S, RNAt y subunidades de la cadena respiratoria)–.

Para las mutaciones localizadas en los genes mitocondriales que codifican para subunidades de la cadena respiratoria, se analizaron las mutaciones que daban lugar a cambio a nivel aminoacídico (substituciones no sinónimas) mediante el programa mtDNA-Server (Weissensteiner et al. 2016a). Para las que tienen un impacto a nivel aminoacídico, se aplicó el modelo de predicción de patogenicidad MutPred2 (Pejaver et al. 2017). Respecto a las localizadas en regiones que codifican RNAr –12S y 16S– también se realizó un estudio de predicción de su estructura secundaria y de la entropía mediante el servicio web RNAfold (Gruber et al. 2008). Por último, para aquellas mutaciones localizadas en regiones que codifican RNAt, se llevó a cabo una predicción de patogenicidad mediante la herramienta MitoTIP *tRNA Scoring* (Sonney et al. 2017) de la base de datos del genoma mitocondrial humano MitoMap (Lott et al. 2013). Para aquellas que fueron consideradas posible o probablemente patogénicas, se realizó un estudio a nivel de estructura secundaria. Se obtuvieron los modelos de estructura de los RNAt de una compilación de RNAt mitocondriales de mamíferos (Mamit-tRNA) (Pütz et al. 2007) y, posteriormente, se predijeron sus estructuras mediante el servicio web RNAfold (Gruber et al. 2008).

3.7.4. Análisis de mutaciones en heteroplasma

Para determinar si existen diferencias significativas entre el número de individuos heteroplásmicos afectados de glioma respecto a individuos que no padecen cáncer, se compilaron los datos obtenidos en el presente estudio para cerebros de individuos sanos (N=10) con los publicados previamente por Li et al. (2015a) de muestras de cerebro de individuos que no padecieron cáncer (N total= 152). Para la comparación se utilizó una regresión logística binaria con el programa SPSS versión 19.0. (IBM Corp. 2010). Se tuvo en cuenta la presencia o ausencia de glioma como variable dependiente y la presencia o ausencia de heteroplasma como variable independiente. Además, para evitar que los resultados pudieran estar influenciados por la edad de los individuos, se consideró esta variable como covariable. Adicionalmente, se calculó la frecuencia y los intervalos de credibilidad bayesiana (IC 95%) utilizando el software SAMPLING tanto para las heteroplasmas somáticas como germinales detectadas en el presente estudio y se comparó con las frecuencias detectadas en la población control. En este caso, se manejaron únicamente los datos presentados por Li et al. (2015a) de muestras de cerebro de individuos que no padecieron cáncer (N= 142) y se utilizó la información descrita por Ramos et al. (2016) sobre la diferenciación de estas heteroplasmas en somáticas y germinales.

El análisis de la distribución de las mutaciones en heteroplasma a lo largo de genoma mitocondrial y el análisis de mutaciones en heteroplasma detectadas en posiciones estables se realizó siguiendo la misma estrategia antes descrita para las mutaciones en homoplasma.

4

CAPÍTULO II RESULTADOS y DISCUSION 1

4. CAPÍTULO II RESULTADOS y DISCUSIÓN 1

VALORACIÓN de la CALIDAD de los archivos FASTQ

4.1. RESULTADOS

4.1.1. Calidad de los archivos Fastq en las quimeras generadas a partir de muestras de sangre de diferentes individuos

Para cada quimera secuenciada –Q1_run1; Q1_run2 y Q2_run1; Q2_run2; Q2_run3; Q2_run4– se generaron 2 archivos Fastq: Fastq-R1 (en referencia al primer ciclo de lecturas, en sentido directo y reverso, de la cadena de DNAm) y Fastq-R2 (en referencia al segundo ciclo de lecturas, en sentido directo y reverso, de la cadena complementaria de DNAm). Con el programa FastQC (Andrews 2010), se han evaluado varios parámetros para cada uno de los archivos (Tabla 10).

Respecto a todos los módulos analizados con el programa FastQC, ambas quimeras muestran buenos resultados para varios módulos en todos los *runs* (Tabla 10). Sin embargo, en algunos casos, varios módulos muestran una disminución en la calidad:

Calidad de la secuencia por base. Los resultados de las quimeras muestran que las lecturas tienen una buena calidad por base en todos los casos, a excepción de las lecturas R2 de Q1_run1, Q2_run3 y Q2_run4, que muestran una disminución notable de la calidad hacia el final de la lectura (Tabla 10; Figura 11).

Tabla 10. Visión general de los resultados obtenidos de cada uno de los módulos analizados en el programa FastQC, para cada una de las quimeras en cada uno de los *runs*. El color verde indica que la calidad en ese módulo es correcta; el color naranja advierte que la calidad de ese módulo no es del todo correcta; el color rojo indica que la calidad de ese módulo falla. R1: primer archivo Fastq generado; R2: segundo archivo Fastq generado.

Informe FastQC	Q1_run1		Q1_run2		Q2_run1		Q2_run2		Q2_run3		Q2_run4	
	R1	R2										
Estadística básica ¹	Green	Green										
Calidad de la secuencia por base	Green	Red	Green	Green	Green	Green	Green	Green	Green	Red	Green	Red
Calidad por cuadrante de la <i>flowcell</i>	Green	Green										
Valores promedio de calidad por secuencia	Green	Green										
Proporción de base por posición	Red	Red	Orange	Orange	Orange	Orange	Orange	Red	Red	Red	Orange	Red
Contenido de GC por secuencia en comparación con una distribución normal teórica	Orange	Orange	Green	Green	Orange	Orange	Orange	Orange	Red	Orange	Red	Red
Contenido de N (veces que no se ha realizado una lectura) por base	Green	Green										
Distribución de la longitud de las secuencias	Orange	Orange										
Niveles de duplicados por secuencia	Red	Red	Red	Red	Red	Red	Red	Red	Red	Orange	Red	Red
Secuencias sobrerrepresentadas ²	Orange	Green	Green	Green	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange
Contenido de adaptadores ³	Green	Green										
Contenido K-mer ⁴	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Red	Orange	Orange	Orange	Orange

¹Información básica como el nombre y tipo de archivo, la plataforma de secuenciación utilizada, el número total de secuencias, la longitud de las secuencias y el porcentaje (%) GC. ²Se muestran aquellas secuencias que se considera que están sobrerrepresentadas en el genoma. ³Indica si existe un porcentaje de secuencias que corresponden a las secuencias de los adaptadores y que, por lo tanto, son secuencias que se pueden acumular en los extremos de las lecturas. ⁴Contabiliza la cantidad de secuencias *K-mers*, es decir, todas las posibles subsecuencias de longitud *K* que pueden contener las lecturas. Con las secuencias *K-mers* se tienen en cuenta aquellas secuencias que no son exactamente iguales y que, por lo tanto, no se contabilizan ni como duplicados ni secuencias sobrerrepresentadas.

Proporción de base por posición. Los resultados de las quimeras muestran que el 50% de los archivos Fastq presentan diferencias >10% entre A-T o G-C (indicado en color naranja en la Tabla 10), mientras que el otro 50% presentan diferencias >20% (indicado en color rojo en la Tabla 10), sobre todo en el inicio y hacia el final de las lecturas (Figura 12). Un resultado óptimo no mostraría diferencias entre la proporción de A-T y G-C a lo largo de las lecturas.

Contenido de GC por secuencia. Todos los resultados de las quimeras muestran una desviación respecto la teórica distribución normal del contenido de GC calculada a partir de los datos observados, a excepción de las lecturas en ambos sentidos de Q1_run2: 7 de los casos muestran una desviación >15% (indicado en color naranja en la Tabla 10) y 3 de los casos una desviación >30% (indicado en color rojo en la Tabla 10) (Figura 13).

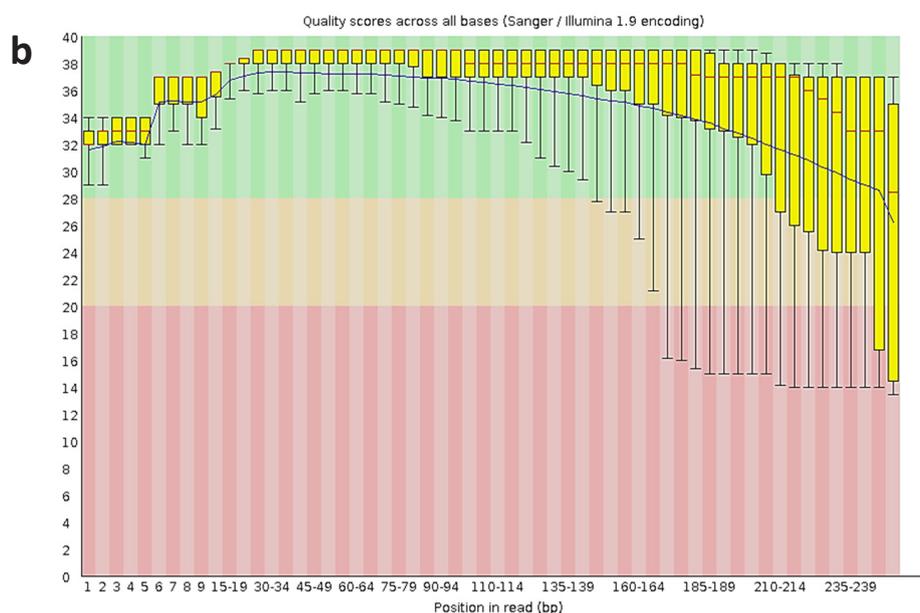
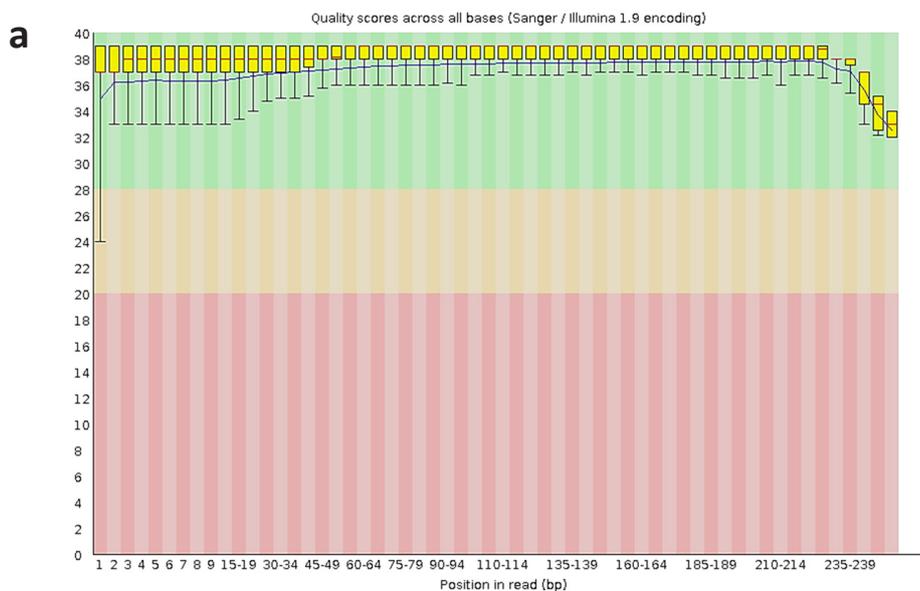


Figura 11. Calidad por base a lo largo de las lecturas. El eje de abscisas muestra la longitud de las lecturas (250 pb) y el eje de ordenadas el puntaje de calidad otorgado. Se diferencian 3 rangos: rango de buena calidad (verde), de calidad intermedia (naranja) y baja calidad (rojo). Las líneas rojas representan el valor medio y la línea azul indica la calidad promedio. a) Ejemplo de una buena calidad por base en las lecturas R1 de Q1_run2. b) Ejemplo de la disminución de la calidad por base en las lecturas R2 de Q2_run3. En este caso, se observa que la calidad va disminuyendo hacia el final de las lecturas R2 en Q2_run3.

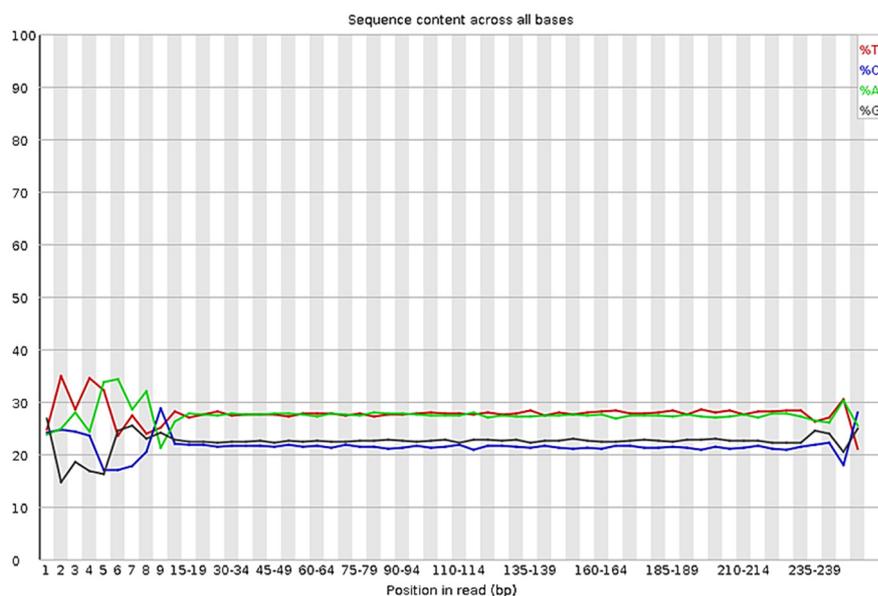


Figura 12. Ejemplo de la proporción de bases obtenida a lo largo de las lecturas R2 de Q2_run2. Se observa que las diferencias entre la proporción de A-T y G-C son >20%, sobre todo en los primeros 20pb de las lecturas, y hacia el final de la lectura.

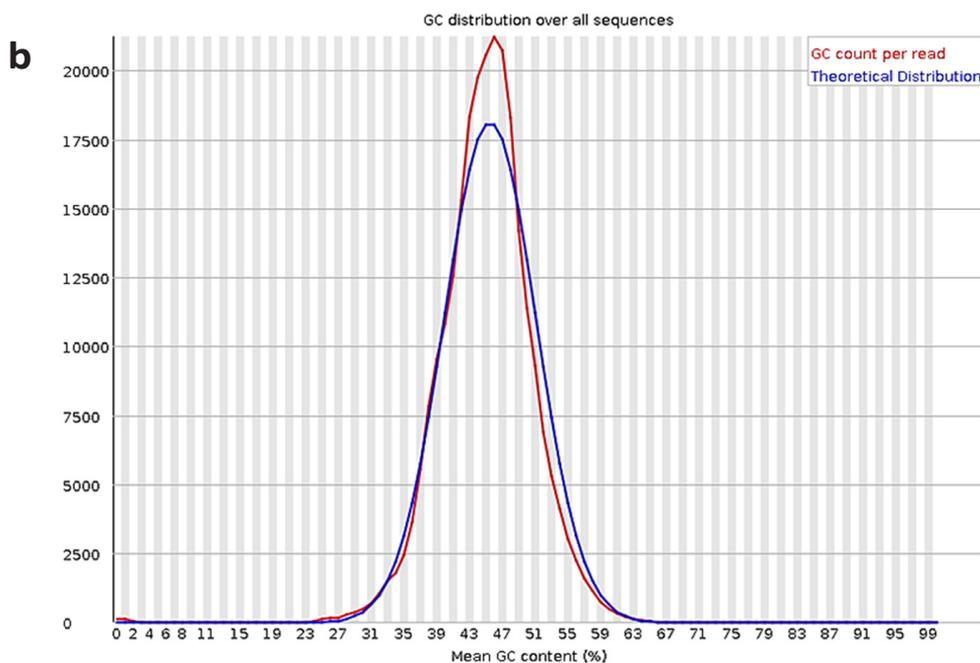
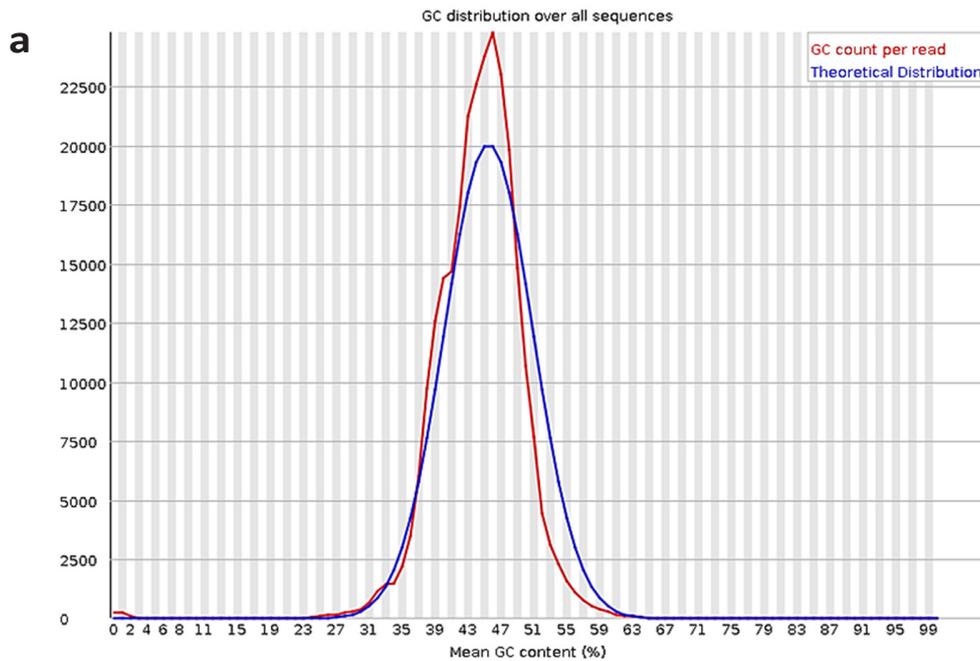


Figura 13. En ambos gráficos, la línea roja corresponde al contenido de GC de las lecturas y la línea azul corresponde a la teórica distribución normal del contenido de GC del genoma de estudio. a) Ejemplo de las lecturas obtenidas en R1 de Q1_run1. El 15% de las lecturas presentan una desviación respecto a la distribución normal del contenido de GC. b) Ejemplo de las lecturas obtenidas en R2 de Q1_run2. En este caso, las lecturas presentan una correcta distribución en el contenido de GC.

Distribución de la longitud de las secuencias. En este estudio las quimeras se secuenciaron mediante lecturas de 2x250pb, por lo tanto, es de esperar que todas las secuencias generadas tengan una longitud de 250pb. Los resultados muestran que todas las lecturas tienen una pequeña proporción de secuencias con una longitud menor (Tabla 10; Figura 14).

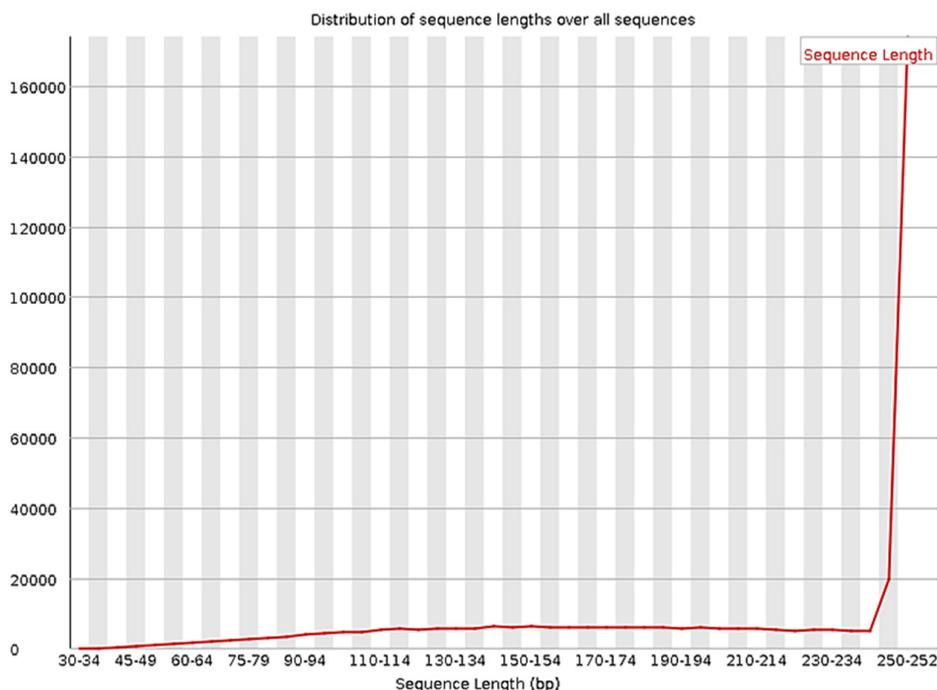


Figura 14. Ejemplo de la longitud de las lecturas obtenidas en R1 de Q2_run1. Se observa que la mayoría de secuencias tienen la longitud esperada (250pb), sin embargo, también se puede observar una pequeña proporción de secuencias con longitudes <250pb (entre 60 y 240pb).

Niveles de secuencias duplicadas por secuencia. Los resultados de las quimeras muestran en todos los casos que la presencia de secuencias no únicas es >50% del total de secuencias (indicado en color rojo en la Tabla 10), a excepción de las lecturas R2 de Q2_run3, donde la presencia de secuencias no únicas está entre el 20–50% (Figura 15).

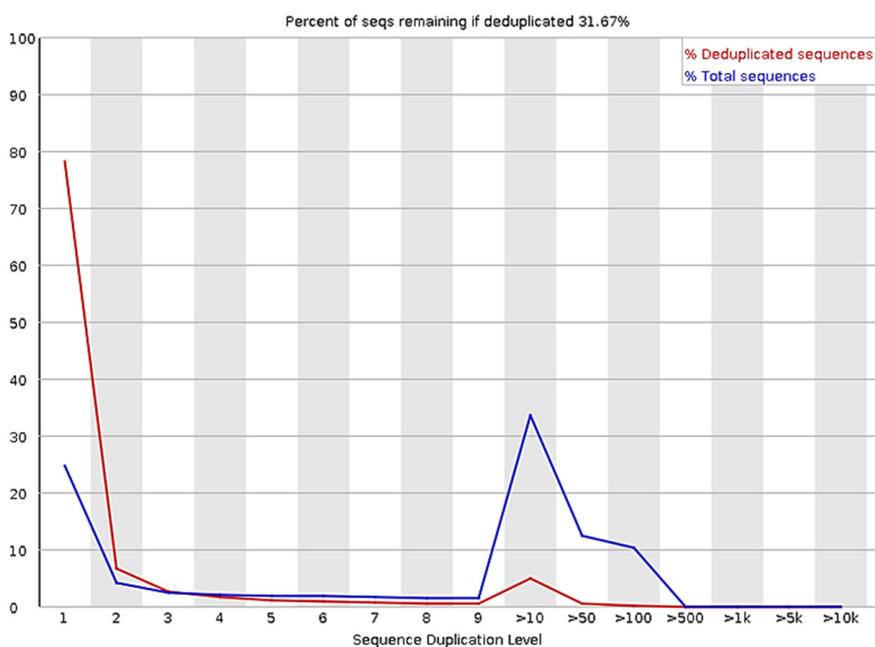


Figura 15. Ejemplo del porcentaje de secuencias duplicadas en las lecturas R1 de Q2_run2. En este caso, el porcentaje de secuencias no únicas es >50%.

Secuencias sobrerrepresentadas. En la mayoría de resultados se encuentran secuencias que representan más de 0,1% del genoma total, pero menos del 1% (indicado en color naranja en la Tabla 10) (Tabla 11).

Tabla 11. Ejemplo de secuencias sobrerrepresentadas en las lecturas R1 de Q2_run2. Las 4 secuencias sobrerrepresentadas se encuentran en un porcentaje entre el 0,1% y 1% del genoma total. En todos los casos, no es posible detectar la fuente que da lugar a esta sobrerrepresentación.

Secuencia	Recuento	Porcentaje	Posible fuente
TATGTGCCTGCGTTCAGGCGTTCTGGCTGTTGCCCTCATCGGGTGATGAT	631	0,161	Sin resultado
AGGATTATGATGCGACTGTGAGTGCGTTCGTAGTTTGAGTTTGCTAGGCA	448	0,115	Sin resultado
TCCTTGACTATCCCTATGAGGCATAATTATAACAAGCTCCATCTGCCTA	443	0,113	Sin resultado
TCTCACTGCCAAGAAGACTATCAAACCTCTGAGCCAACAACCTAATATGAC	424	0,108	Sin resultado
GTTGTAGTCCGTGCGAGAATAATGATGTATGCTTTGTTTCTGTTGAGTGT	407	0,104	Sin resultado

Contenido K-mer. Lo resultados de las quimeras muestran en todos los casos secuencias K-mers, es decir, secuencias de k nucleótidos en la secuencia del DNAm (en este caso, secuencias 7-mers). El contenido K-mer está en desequilibrio con un p-valor binomial $<0,01$, mostrando desviaciones significativas a partir del coverage uniforme del genoma mitocondrial (indicado en color naranja en la Tabla 10), a excepción las lecturas R2 de Q2_run2, donde se encuentra alguna secuencia K-mer en desequilibrio con un p-valor binomial $<10^{-5}$ (indicado en color rojo en la Tabla 10) (Figura 16; Tabla 12).

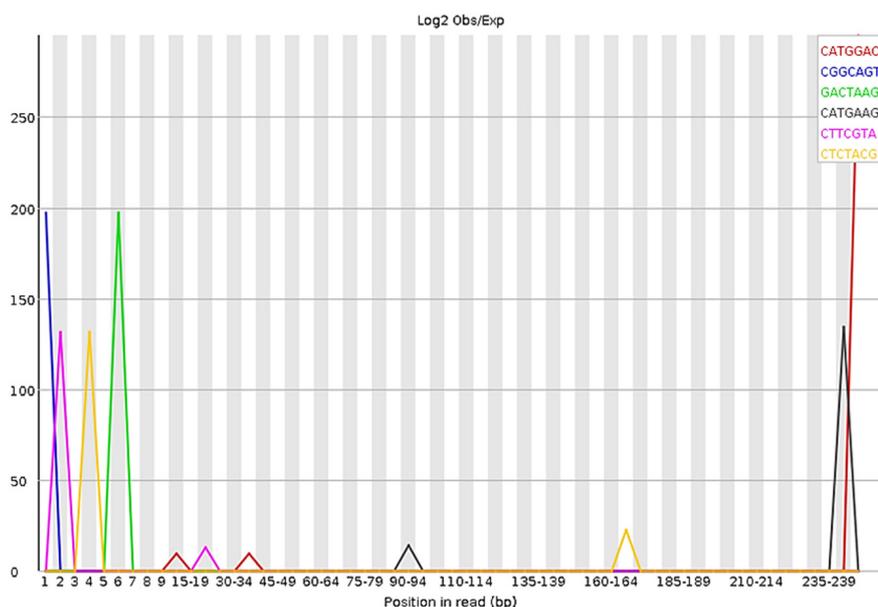


Figura 16. Ejemplo de secuencias 6 K-mers más sesgadas en las lecturas R1 de Q2_run1. Principalmente, se observa la presencia de secuencias k-mers al inicio y al final de las lecturas.

Tabla 12. Representación de todas las secuencias que se detectan en el gráfico. En este caso, algunos k-mers se encuentran en desequilibrio con un p-valor binomial $<0,01$.

Secuencia	Recuento	Valor-P	Máximos Observados/Esperados	Máximos Posiciones Observados/Esperados
CATGGAC	20	8,48 ⁻⁴	294,727	245
CGGCAGT	10	0,0028	197,714	1
GACTAAG	10	0,0028	197,714	6
CATGAAG	15	0,0089	134,611	240-244
CTTCGTA	15	0,0094	131,81	2
CTCTACG	15	0,0094	131,81	4
TATTCGT	200	1,85 ⁻⁴	58,945	245
GCACGCG	160	0,0057	55,261	245
TCGAATG	160	0,0057	55,261	245
CTACAAC	730	0	48,448	245
CCTGGGG	415	3,1304 ⁻⁶	42,611	245
CGGTTCA	195	0,0012	40,157	235-239
CAATTCG	190	0,0012	40,069	230-234
ACGAATA	100	0,0013	39,543	1
TCGCACG	345	0	36,953	235-239
AGTAGAC	405	1,3547 ⁻⁴	36,386	245
CCTATAC	1215	0	36,386	245
CGGTCGT	165	1,0113 ⁻⁵	35,948	7
CTTATAC	335	0,0023	35,191	245
CTAGGGC	770	5,1914 ⁻⁹	34,449	245

4.1.2. Calidad de los archivos Fastq en las muestras tumorales y sanguíneas de individuos afectados de glioma y controles de cerebro de individuos sanos

Siguiendo la misma estrategia llevada a cabo para las quimeras, se ha analizado la calidad de los datos obtenidos a partir de NGS tanto para las muestras tumorales y sanguíneas de individuos afectados de glioma como para los controles de cerebro de individuos sanos. Así, se han evaluado también varios parámetros en los 2 archivos Fastq generados para cada muestra (Tablas 13 y 14).

Tabla 13. Visión general de los resultados obtenidos de cada uno de los módulos analizados en el programa FastQC para cada una de las muestras secuenciadas en los diferentes *runs*. Del mismo modo que para las quimeras, el color verde indica que la calidad en ese módulo es correcta; el color naranja advierte que la calidad de ese módulo no es del todo correcta; el color rojo indica que la calidad de ese módulo falla. R1: primer archivo Fastq generado; R2: segundo archivo Fastq generado. *La definición de determinados módulos fueron definidos anteriormente (véase apartado 4.1.1, Tabla 10).

Muestra	Tejido	Run	Número de ciclo	Estadística básica*	Calidad de la secuencia por base	Calidad por cuadrante de la <i>flowcell</i>	Valores promedio de calidad por secuencia	Proporción de base por posición	Contenido de GC por secuencia en comparación con una distribución normal teórica	Contenido de N (veces que no se ha realizado una lectura) por base	Distribución de la longitud de las secuencias	Niveles de duplicados por secuencia*	Secuencias sobrerrepresentadas*	Contenido de adaptadores	Contenido K-mer*
3	Tumor	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
	Sangre	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
5	Tumor	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
	Sangre	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
6	Tumor	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
	Sangre	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
7	Tumor	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
	Sangre	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
10	Tumor	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
	Sangre	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
13	Tumor	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
	Sangre	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
16	Tumor	5	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
	Sangre	5	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
17	Tumor	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
	Sangre	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
18	Tumor	5	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
	Sangre	1	R1	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde
			R2	Verde	Verde	Verde	Verde	Rojo	Naranja	Verde	Verde	Rojo	Naranja	Verde	Verde

Tabla 13. Continuación.

Muestra	Tejido	Run	Número de ciclo	Estadística básica*	Calidad de la secuencia por base	Calidad por cuadrante de la flowcell	Valores promedio de calidad por secuencia	Proporción de base por posición	Contenido de GC por secuencia en comparación con una distribución normal teórica	Contenido de N (veces que no se ha realizado una lectura) por base	Distribución de la longitud de las secuencias	Niveles de duplicados por secuencia*	Secuencias sobrerrepresentadas*	Contenido de adaptadores	Contenido K-mer*
19	Tumor	1	R1												
			R2												
	Sangre	1	R1												
			R2												
20	Tumor	1	R1												
			R2												
	Sangre	1	R1												
			R2												
21	Tumor	3	R1												
			R2												
	Sangre	3	R1												
			R2												
26	Tumor	1	R1												
			R2												
	Sangre	1	R1												
			R2												
27	Tumor	5	R1												
			R2												
	Sangre	2	R1												
			R2												
30	Tumor	3	R1												
			R2												
	Sangre	3	R1												
			R2												
31	Tumor	5	R1												
			R2												
	Sangre	4	R1												
			R2												
31BIS	Tumor	4	R1												
			R2												
	Sangre	4	R1												
			R2												
33	Tumor	5	R1												
			R2												
	Sangre	4	R1												
			R2												
34	Tumor	5	R1												
			R2												
	Sangre	3	R1												
			R2												

Tabla 13. Continuación.

Muestra	Tejido	Run	Número de ciclo	Estadística básica*	Calidad de la secuencia por base	Calidad por cuadrante de la flowcell	Valores promedio de calidad por secuencia	Proporción de base por posición	Contenido de GC por secuencia en comparación con una distribución normal teórica	Contenido de N (veces que no se ha realizado una lectura) por base	Distribución de la longitud de las secuencias	Niveles de duplicados por secuencia*	Secuencias sobrerrepresentadas*	Contenido de adaptadores	Contenido K-mer*
35	Tumor	4	R1												
			R2												
	Sangre	4	R1												
			R2												
36	Tumor	5	R1												
			R2												
	Sangre	3	R1												
			R2												
41	Tumor	2	R1												
			R2												
	Sangre	5	R1												
			R2												
43	Tumor	2	R1												
			R2												
	Sangre	2	R1												
			R2												
44	Tumor	2	R1												
			R2												
	Sangre	5	R1												
			R2												
45	Tumor	2	R1												
			R2												
	Sangre	5	R1												
			R2												
50	Tumor	4	R1												
			R2												
	Sangre	5	R1												
			R2												
51	Tumor	5	R1												
			R2												
	Sangre	4	R1												
			R2												
53	Tumor	5	R1												
			R2												
	Sangre	4	R1												
			R2												
54	Tumor	5	R1												
			R2												
	Sangre	4	R1												
			R2												

Tabla 13. Continuación.

Muestra	Tejido	Run	Número de ciclo	Estadística básica*	Calidad de la secuencia por base	Calidad por cuadrante de la flowcell	Valores promedio de calidad por secuencia	Proporción de base por posición	Contenido de GC por secuencia en comparación con una distribución normal teórica	Contenido de N (veces que no se ha realizado una lectura) por base	Distribución de la longitud de las secuencias	Niveles de duplicados por secuencia*	Secuencias sobrerrepresentadas*	Contenido de adaptadores	Contenido K-mer*
55	Tumor	5	R1												
			R2												
	Sangre	5	R1												
			R2												
56	Tumor	5	R1												
			R2												
	Sangre	4	R1												
			R2												
59	Tumor	5	R1												
			R2												
	Sangre	4	R1												
			R2												

Tabla 14. Visión general de los resultados obtenidos de cada uno de los módulos analizados en el programa FastQC para los controles secuenciados en diferentes runs. *La definición de determinados módulos fueron definidos anteriormente (véase apartado 4.1.1, Tabla 10).

Muestra	Run	Número de ciclo	Estadística básica*	Calidad de la secuencia por base	Calidad por cuadrante de la flowcell	Valores promedio de calidad por secuencia	Proporción de base por posición	Contenido de GC por secuencia en comparación con una distribución normal teórica	Contenido de N (veces que no se ha realizado una lectura) por base	Distribución de la longitud de las secuencias	Niveles de duplicados por secuencia*	Secuencias sobrerrepresentadas*	Contenido de adaptadores	Contenido K-mer*
Control 2	2	R1												
		R2												
Control 4	2	R1												
		R2												
Control 8	2	R1												
		R2												
Control 10	2	R1												
		R2												
Control 11	2	R1												
		R2												
Control 12	3	R1												
		R2												
Control 13	3	R1												
		R2												
Control 15	5	R1												
		R2												
Control 20	3	R1												
		R2												
Control 21	3	R1												
		R2												

Patrón de coverage de las muestras

Todas las muestras, a excepción de aquellas que fueron amplificadas mediante más fragmentos (muestras 21 y 30 tumorales), siguen el mismo patrón respecto al *coverage* obtenido, con determinadas zonas donde se obtiene una mayor profundidad de lectura y otras donde se obtiene una menor profundidad (Figura 17). Respecto a la posible relación entre el *coverage* y el contenido de GC, no se observa una relación significativa entre ambos parámetros (correlación de *Pearson*: $-0,058$; $p=0,642$) (Figura 18).

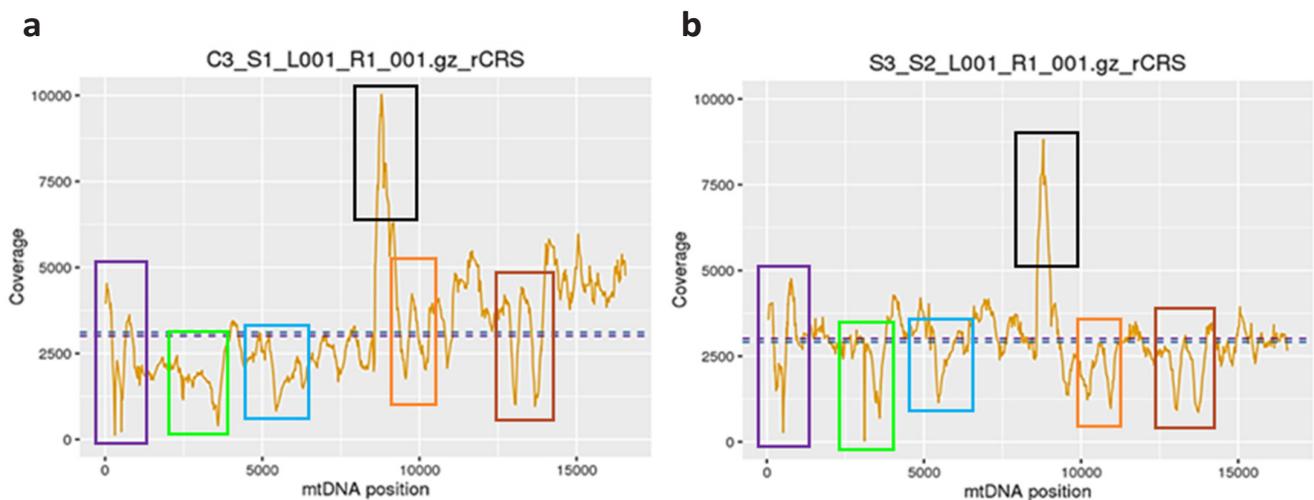


Figura 17. a) Ejemplo del *coverage* obtenido a lo largo del genoma mitocondrial de la muestra 3 de tejido tumoral. b) Ejemplo del *coverage* obtenido a lo largo del genoma mitocondrial de la muestra 3 de tejido sanguíneo.

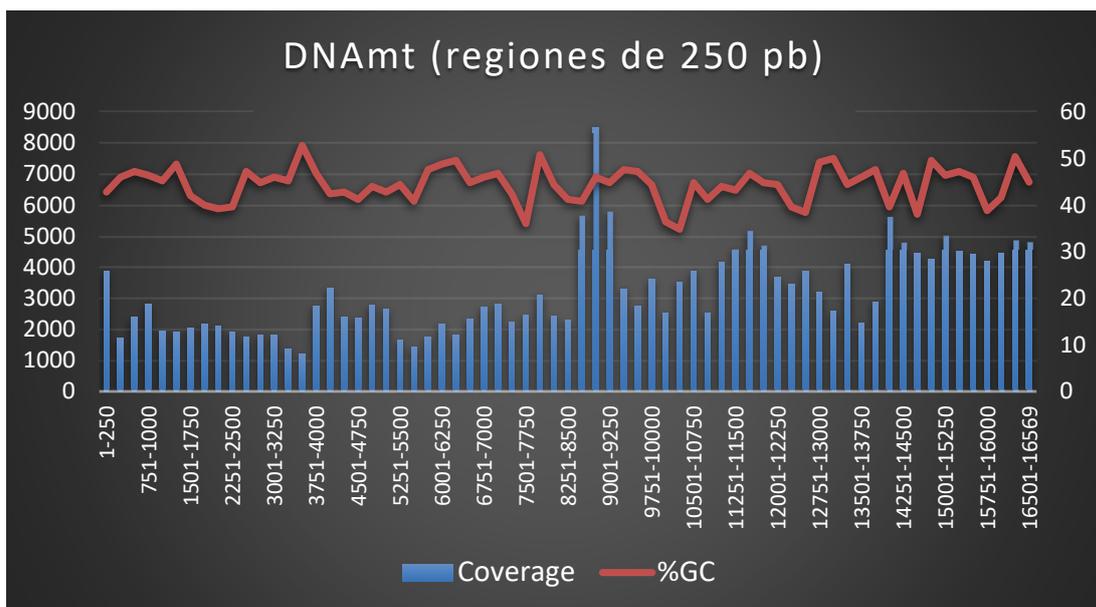


Figura 18. Relación entre la profundidad de lectura obtenida y el contenido de GC (%) para cada fragmento mitocondrial de 250 pb de la muestra 3 tumoral. A la izquierda, los valores del eje de ordenadas indican la profundidad de lectura. A la derecha, los valores del eje de ordenadas indican el porcentaje de GC.

4.2. DISCUSIÓN

Para valorar todos los parámetros de calidad de los archivos Fastq, en primer lugar, se ha de tener en cuenta que dichos parámetros no están dirigidos a ningún tipo de genoma, metodología para la elaboración de las librerías, o plataforma de secuenciación específica. Por lo tanto, los sesgos que se introduzcan por alguno de estos motivos serán detectados como advertencia o error en la calidad a nivel global o a nivel de *run*, y no a nivel de muestra. Así pues, es importante conocer bien la metodología (Nextera XT®) y la tecnología (MiSeq, Illumina) que se está utilizando para poder interpretar, entender y valorar los resultados obtenidos. En este sentido, varios factores deben tenerse en cuenta.

En primer lugar, la estrategia seguida para la secuenciación incluye la realización de librerías en que se realiza una amplificación del DNA, hecho que puede influir en la aparición de determinados sesgos. Por un lado, varios estudios (Bansal 2017; Jones et al. 2015; Kozarewa et al. 2009) describen que se puede producir una elevada presencia de secuencias duplicadas en aquellas librerías de Illumina en las que su protocolo incluya una amplificación del DNA, a fin de enriquecer los fragmentos de DNA con adaptadores en los extremos, tal y como se detecta en los presentes resultados. En algunos casos, parte de estos duplicados pueden corresponder a las secuencias de los adaptadores que se han añadido (Kozarewa et al. 2009). No obstante, los problemas que puede originar la presencia de duplicados desaparecerán en el momento que éstos sean eliminados durante el procedimiento bioinformático implementado (Weissensteiner et al. 2016a). Por otro lado, varios estudios han comprobado que también se puede generar un sesgo relacionado con el contenido de GC en las librerías de Illumina con PCR (Aird et al. 2011; Benjamini and Speed 2012). Como consecuencia, ambos sesgos pueden ser identificados como secuencias sobrerrepresentadas dentro del análisis de la calidad de los Fastq, ya que la amplificación del DNA puede ser desigual en algunas zonas y generar secuencias que estén representadas de forma excesiva dentro del genoma (Kozarewa et al. 2009). Así pues, estos 3 sesgos que se reproducen en todas las muestras analizadas en la presente tesis se podrían justificar, *a priori*, por la realización de librerías que incluyen una amplificación del DNA. Por lo tanto, es importante considerar los aspectos de la metodología empleada para llevar a cabo la secuenciación, ya que se pueden generar ciertos sesgos por defecto respecto a la calidad de las lecturas generadas. Además, en el presente trabajo se ha de considerar que para poder seleccionar únicamente el DNAm de las muestras analizadas es necesario llevar a cabo una PCR previa adicional, de manera que se pueden ir arrastrando sesgos de amplificación incluso antes de la PCR que incluye el protocolo de las librerías.

En segundo lugar, el uso de transposomas en el kit Nextera XT[®] utilizado para la elaboración de las librerías podría tener un papel relevante. Aunque la tagmentación es un proceso aleatorio, pueden existir fragmentos de DNA con determinadas secuencias que son preferentemente tagmentadas (Schirmer et al. 2016). Esto puede generar una distribución desigual de bases que provoque, a la vez, una sobrerrepresentación de secuencias *k-mers*, sobre todo al inicio de las lecturas (en los primeros 20pb aproximadamente), tal y como se observa en los resultados obtenidos. Además, se ha descrito que los transposomas utilizados en Nextera XT[®] utilizan transposasas Tn5 modificadas que aumentan su capacidad para unirse al DNA (Marine et al. 2011) y tienden a cortarlo por regiones ricas en A/T (TTATA) y flanqueadas por regiones GC, hecho que puede contribuir también a la distribución desigual de bases al inicio de las lecturas (Ason and Reznikoff 2004; Schirmer et al. 2016).

Finalmente, tal y como se ha comentado anteriormente, la plataforma Illumina realiza la secuenciación por síntesis, es decir, que no se secuencian totalmente una molécula antes de secuenciar otra, sino que todas las moléculas se van secuenciando simultáneamente, de manera que en cada ciclo se incorpora una base de cada una de las secuencias. Es por eso que errores de fase y prefase (véase apartado 1.3.3) pueden provocar que se detecten diferentes bases en un mismo ciclo de secuenciación y, como consecuencia, mezclas de señales que provocan una progresiva pérdida de calidad sobre todo hacia el final de las lecturas (Schirmer et al. 2016). Por otro lado, la plataforma Illumina genera clústeres, es decir, un conjunto de moléculas idénticas para que juntas puedan emitir suficiente señal para ser detectada. Cuando la densidad de los clústeres es demasiado alta puede dificultar las lecturas R2 y hacer que la calidad de ésta disminuya significativamente. Esto podría explicar el hecho de que la calidad de las lecturas R2 sea inferior a la calidad de las lecturas R1 en la mayoría de casos. Asimismo, aunque se ha comentado que la mayoría de secuencias *K-mers* se dan al inicio de las lecturas, es posible que se puedan dar en una menor proporción hacia el final debido al sesgo que se genera a medida que avanza el proceso de secuenciación (Schirmer et al. 2015). Por último, en relación a la longitud de las lecturas, aunque en el presente estudio se espera que todas las lecturas tengan una longitud de 250pb, en todas las muestras secuenciadas se presenta una pequeña proporción de lecturas de un tamaño inferior al esperado. No obstante, en determinadas plataformas generan, por defecto, una proporción de lecturas de un tamaño menor, de manera que dicho resultado no tendría trascendencia respecto a este parámetro.

Aun habiendo obtenido estos resultados a nivel de calidad, un alineamiento posterior de las lecturas permitirá valorar mejor los resultados obtenidos en la secuenciación. Obtener unos buenos resultados de calidad no significa obtenerlos también en los alineamientos de las lecturas. Del mismo modo, unos resultados no tan buenos respecto a la calidad pueden generar igualmente datos útiles para posteriores

análisis. No obstante, el análisis de la calidad sí que puede dar una idea sobre la profundidad de lectura que se podrá obtener una vez se apliquen los filtros de calidad establecidos para el análisis bioinformático.

Respecto al patrón de *coverage* a lo largo del genoma obtenido en el presente estudio, se ha hallado un resultado similar al presentado por McElhoe et al. (2014) (Figura 19), desconociéndose la razón específica de la distribución desigual que se observa. En el estudio de McElhoe et al. (2014) el DNAm_t también fue amplificado en 2 fragmentos –2480–10858 y 10653–2688– y se utilizó el mismo kit para la construcción de las librerías (Nextera XT[®]) y la misma plataforma de secuenciación (MiSeq, Illumina).

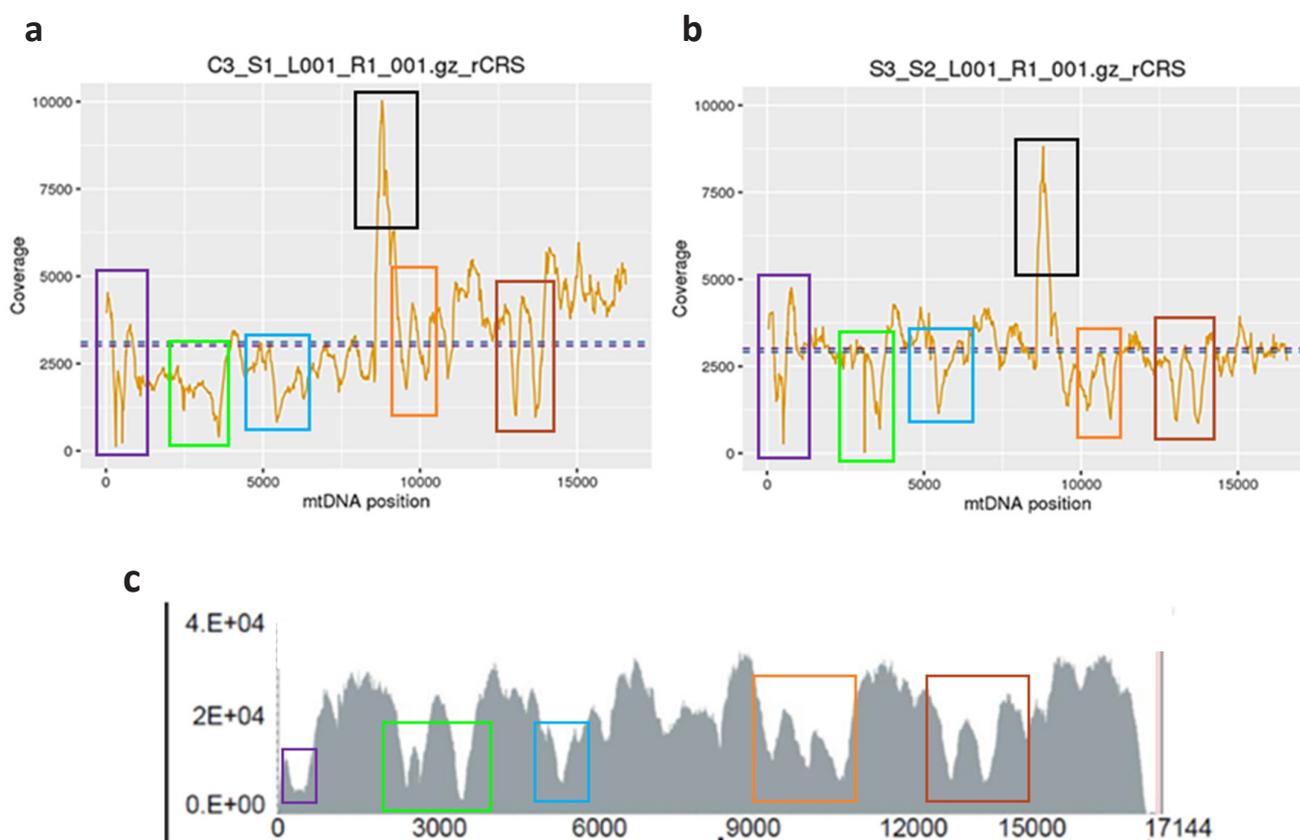


Figura 19. a) *Coverage* a lo largo del genoma mitocondrial de la muestra 3 de tejido tumoral. b) *Coverage* a lo largo del genoma mitocondrial de la muestra 3 de tejido sanguíneo. c) Patrón de *coverage* a lo largo del genoma mitocondrial encontrado en una muestra del estudio reportado por McElhoe et al. (2014). Las disminuciones en el *coverage* a lo largo del genoma mitocondrial coinciden en las 3 muestras (marcadas con cuadros de diferentes colores en los 3 casos). Además, en ambos tejidos de la muestra 3 se detecta un pico de *coverage* (indicado en color negro), que coincide con la zona solapante del Set 1 y Set 2.

Tal y como se ha comentado, las diferencias en el contenido de GC a lo largo del genoma de la mayoría de especies han estado relacionadas con la aparición de diferentes sesgos generados por razones tecnológicas, metodológicas y biológicas (por ejemplo, el tipo de genoma analizado, los procesos de amplificación previos, el tipo de preparación de la librería, el tipo de plataforma de secuenciación,

entre otras) y que pueden afectar a los análisis llevados a cabo posteriormente (Romiguier and Roux 2017). Estos sesgos pueden darse tanto en regiones con un elevado o un bajo contenido de GC y en zonas homopoliméricas (Ross et al. 2013). Así pues, mientras que algunos estudios presentan una correlación positiva entre el número de lecturas y el contenido de GC en genomas de diferentes organismos, otros describen una correlación negativa. Por ejemplo, se ha detectado una correlación positiva en la secuenciación del genoma de *Helicobacter acinonychis* con la plataforma Solexa (Dohm et al. 2008), pero también una correlación negativa en el genoma mitocondrial de 2 especies de vertebrados en la plataforma Illumina (Bentley et al. 2008; Ekblom et al. 2014). En otros estudios se ha determinado un bajo *coverage* tanto en aquellas zonas pobres en GC como en zonas ricas en GC (Chen et al. 2013). Dentro del genoma humano se ha observado una correlación negativa entre el *coverage* y áreas donde el contenido de GC es extremo (<4% y >76%) (McElhoe et al. 2014). Ahora bien, teniendo en cuenta el resultado obtenido en el presente estudio, el contenido de GC parece no estar relacionado con las diferencias en el patrón de *coverage* obtenidos en las muestras analizadas. Así, son necesarios análisis más profundos para poder conocer los factores que influyen en el patrón de *coverage* obtenido en la secuenciación del genoma mitocondrial.

5

CAPÍTULO III
RESULTADOS y DISCUSION 2

5. CAPÍTULO III RESULTADOS y DISCUSIÓN 2

DETECCIÓN de HETEROPLASMIAS MITOCONDRIALES a través de NGS

5.1. RESULTADOS

Las dos quimeras sintetizadas (Q1 y Q2) portadoras de mutaciones en homoplasma, pero también en heteroplasma, se han utilizado como patrón control para estandarizar algunos parámetros que permitirán validar la variabilidad obtenida, fundamentalmente en heteroplasma, mediante NGS en los gliomas cerebrales.

5.1.1. Profundidad de lectura promedio en cada *run*

En cada quimera se han obtenido diferentes valores de profundidad de lectura promedio en cada uno de los *runs* (Figura 20).

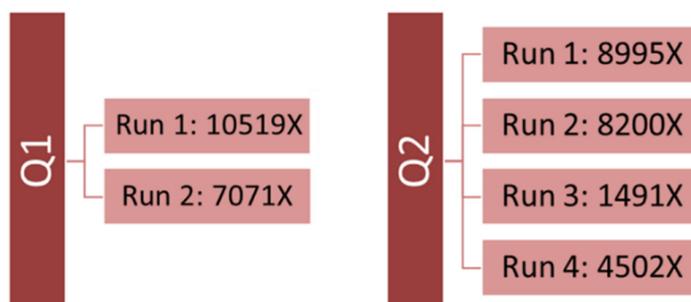


Figura 20. Profundidad promedio obtenida en la Q1 y Q2 en cada uno de los *runs*.

5.1.2. Secuenciación de Sanger vs. Secuenciación por NGS

A partir de las diferencias obtenidas en la profundidad de lectura obtenida dentro del mismo *run* y entre *runs*, varios resultados han sido valorados.

Para los fragmentos de DNAm 1, 2, 3, 4, 6, 7, 8 y 9 (anteriormente descritos en el apartado de metodología), todas las mutaciones previamente detectadas en homoplasmia (Ramos et al. 2013) mediante la secuenciación de Sanger, se identificaron correctamente en todos los *runs* de NGS de ambas quimeras (Tablas 15-16). Sin embargo, a través de NGS, en la Q1, la posición 185 se detecta en heteroplasmia en el *run* 1 (1,07%), mientras que en el *run* 2 aparece como una mutación fijada (Tabla 15). Del mismo modo, en la Q2, la posición 16286 también se determina como heteroplasmia en el *run* 1, *run* 2 y *run* 4 (1,85%, 1,42%, 1,25%, respectivamente) (Tabla 16).

Respecto a las mutaciones en heteroplasmia, todas aquellas que fueron detectadas a través de la secuenciación de Sanger, también lo fueron mediante NGS (Tablas 17-18). No obstante, la frecuencia de heteroplasmia detectada con NGS es notablemente superior respecto la secuenciación de Sanger, a excepción de la heteroplasmia detectada en la posición 5306 de la Q1, donde el porcentaje se mantiene independientemente de la tecnología de secuenciación utilizada (~46%) (Tabla 17). Además, en la Q2, la posición 16000 se detecta en heteroplasmia con una frecuencia entre el 9,41% y 11,5% mediante NGS, por lo que debería haberse detectado, *a priori*, también mediante Sanger (Tabla 18). En relación a las heteroplasmas que únicamente se detectan mediante NGS, se observa que todas presentan una frecuencia muy baja de la variante minoritaria (entre 0,79% y 2,75%), pero ninguna es reproducible en todos los *runs* en ambas quimeras (Tablas 17-18).

Alineamiento BWA-mem vs. Alineamiento Bowtie2 en NGS

Las posiciones en homoplasmia se detectan todas con una profundidad de lectura muy parecida tanto con el alineamiento *BWA-mem* como con el alineamiento *Bowtie2*; asimismo, las posiciones en heteroplasmia también se detectan casi todas con una profundidad de lectura y una frecuencia similar. No obstante, existen ciertas discrepancias entre los 2 tipos de alineamiento respecto a la detección de heteroplasmas cuando éstas presentan una frecuencia $\leq 1\%$ (Tablas 19-20). Además, con el alineamiento *BWA-mem* se detecta la posición 3635 como heteroplasmia con una frecuencia de 2,75% en el *run* 4 de la Q2, pero con una profundidad de lectura baja (472X) (Tabla 20).

Tabla 15. Mutaciones esperadas en homoplasmia, teniendo en cuenta el resultado de la secuenciación de Sanger, detectadas en la Q1 en cada uno de los runs. DIR/REV: lecturas en sentido directo y en sentido reverso

Posición	Mutación	DIR/REV Run 1 (ratio)	Profundidad de lectura Run 1	DIR/REV Run 2 (ratio)	Profundidad de lectura Run 2
185*	G>A	909/863 (1,05)	1772	1481/1496 (0,99)	2977
263	A>G	440/529 (0,83)	969	854/812 (1,05)	1666
750	A>G	5272/1612 (3,27)	6884	2890/4300 (0,67)	7190
1438	A>G	2673/7280 (0,37)	9953	5648/4093 (1,38)	9741
3010	G>A	7501/6748 (1,11)	14249	4627/5181 (0,89)	9808
4639	T>C	9310/9369 (0,99)	18679	3302/3239 (1,02)	6541
4769	A>G	9644/8037 (1,2)	17681	2904/3270 (0,89)	6174
11215	C>T	6177/3207 (1,93)	9384	4077/4999 (0,82)	9076
12810	A>G	2239/4535 (0,49)	6774	4157/3230 (1,29)	7387
15326	A>G	6035/6137 (0,98)	12172	4895/4780 (1,02)	9675
16209	T>C	4532/7333 (0,62)	11865	4247/3617 (1,17)	7864
16519	T>C	2427/4827 (0,5)	7254	3509/2980 (1,18)	6489

*Posición esperada en homoplasmia, teniendo en cuenta el resultado de la secuenciación de Sanger, detectada como heteroplasmia en el run 1 (1,07%).

Tabla 16. Mutaciones esperadas en homoplasmia, teniendo en cuenta el resultado de la secuenciación de Sanger, detectadas en la Q2 en cada uno de los runs. DIR/REV: lecturas en sentido directo y en sentido reverso

Posición	Mutación	DIR/REV Run 1 (ratio)	Profundidad de lectura Run 1	DIR/REV Run 2 (ratio)	Profundidad de lectura Run 2	DIR/REV Run 3 (ratio)	Profundidad de lectura Run 3	DIR/REV Run 4 (ratio)	Profundidad de lectura Run 4
263	A>G	1004/1070 (0,94)	2074	905/1023 (0,88)	1928	163/165 (0,99)	328	521/610 (0,85)	1131
497	C>T	1858/2067 (0,9)	3925	1098/1464 (0,75)	2562	242/129 (1,88)	371	505/344 (1,47)	849
750	A>G	4111/6229 (0,66)	10340	2888/4506 (0,64)	7394	1376/459 (3)	1835	3219/1771 (1,82)	4990
1189	T>C	6120/4483 (1,37)	10603	5031/3982 (1,26)	9013	744/1344 (0,55)	2088	2615/3198 (0,82)	5813
1438	A>G	6429/4023 (1,6)	10452	5193/3440 (1,51)	8633	503/1817 (0,28)	2320	2256/4143 (0,54)	6399
1811	A>G	1815/1932 (0,94)	3747	1640/1785 (0,92)	3425	516/254 (2,03)	770	1067/919 (1,16)	1986
2706	A>G	2541/1523 (1,67)	4064	1616/778 (2,08)	2394	196/632 (0,31)	828	541/1256 (0,43)	1797
3480	A>G	890/1028 (0,87)	1918	412/762 (0,54)	1174	81/74 (1,09)	155	223/195 (1,14)	418
4580	G>A	1136/1371 (0,83)	2507	1216/1325 (0,92)	2541	290/293 (0,99)	583	659/708 (0,93)	1367
4769	A>G	1184/1402 (0,84)	2586	1177/1428 (0,82)	2605	286/262 (1,09)	548	730/549 (1,33)	1279
9477	G>A	4733/4347 (1,09)	9080	3146/3167 (0,99)	6313	627/523 (1,2)	1150	1434/1538 (0,93)	2972
11467	A>G	20377/19738 (1,03)	40115	14894/14977 (0,99)	29871	3515/2849 (1,23)	6364	9340/8913 (1,05)	18253
11719	G>A	17858/15395 (1,16)	33253	16249/15088 (1,08)	31337	1932/3659 (0,53)	5591	7427/9179 (0,81)	16606
12308	A>G	552/666 (0,83)	1218	557/617 (0,9)	1174	165/82 (2,01)	247	434/382 (1,14)	816

Tabla 16. Continuación.

Posición	Mutación	DIR/REV Run 1 (ratio)	Profundidad de lectura Run 1	DIR/REV Run 2 (ratio)	Profundidad de lectura Run 2	DIR/REV Run 3 (ratio)	Profundidad de lectura Run 3	DIR/REV Run 4 (ratio)	Profundidad de lectura Run 4
12372	G>A	601/812 (0,74)	1413	603/715 (0,84)	1318	159/113 (1,41)	272	473/381 (1,24)	854
13135	G>A	419/642 (0,65)	1061	292/453 (0,64)	745	107/45 (2,38)	152	260/145 (1,79)	405
14182	T>C	12416/12123 (1,02)	24539	10377/10642 (0,98)	21019	1537/1542 (1)	3079	5611/5379 (1,04)	10990
14766	C>T	7607/8618 (0,88)	16225	6163/7915 (0,78)	14078	1866/979 (1,91)	2845	4858/3341 (1,45)	8199
15452	C>A	3648/3211 (1,14)	6859	2871/2519 (1,14)	5390	472/502 (0,94)	974	1349/1557 (0,87)	2906
16069	C>T	3109/2609 (1,19)	5718	2894/2505 (1,16)	5399	524/688 (0,76)	1212	1878/2019 (0,93)	3897
16126	T>C	3229/2681 (1,2)	5910	2742/2318 (1,18)	5060	427/716 (0,6)	1143	1618/1972 (0,82)	3590
16145	G>A	2722/2930 (0,93)	5652	2001/2408 (0,83)	4409	449/575 (0,78)	1024	1641/1574 (1,04)	3215
16162	A>G	2768/2838 (0,98)	5606	2027/2276 (0,89)	4303	370/508 (0,73)	878	1469/1420 (1,03)	2889
16261	C>T	3979/2343 (1,7)	6322	2781/1948 (1,43)	4729	236/571 (0,41)	807	986/1685 (0,59)	2671
16286*	C>T	3770/2511 (1,5)	6281	2726/2205 (1,24)	4931	253/530 (0,48)	783	1075/1566 (0,69)	2641

*Posición esperada en homoplasmia, teniendo en cuenta el resultado de la secuenciación de Sanger, detectada como heteroplasmia en el run 1, run 2 y run 4 (1,85%, 1,42%, 1,25%, respectivamente).

Tabla 17. Posiciones en que se detecta heteroplasmia por NGS en la Q1 en cada uno de los runs. rCRS: secuencia de referencia del genoma mitocondrial. MAYOR.MINOR: base mayoritaria y base minoritaria; % Het: frecuencia de heteroplasmia (% de la base minoritaria); DIR/REV: diferencia entre las lecturas en sentido directo y en sentido reverso.

Posición	rCRS	MAYOR. MINOR.	% Het Sanger	% Het Run 1	DIR/REV Run 1 (ratio)	Profundidad de lectura Run 1	% Het Run 2	DIR/REV Run 2 (ratio)	Profundidad de lectura Run 2
185*	G	A/G		1,07	909/863 (1,05)	1772		1481/1496 (0,99)	2977
2887**	T	T/C	8,05	19,95	6602/7886 (0,84)	14488	21,08	5295/4783 (1,11)	10078
3482*	A	A/G			701/479 (1,46)	1180	0,84	554/1597 (0,35)	2151
3578*	T	T/C			430/505 (0,85)	935	0,96	752/803 (0,94)	1555
3864*	T	T/C		2,59	7578/2400 (3,16)	9978	2,25	3190/4089 (0,78)	7279
5306**	C	C/T	45,58	46,6	4073/9235 (0,44)	13308	46,76	4980/3761 (1,32)	8741

*Posiciones esperadas en homoplasmia según la secuenciación de Sanger. **Posiciones esperadas en heteroplasmia según la secuenciación de Sanger.

Tabla 18. Posiciones en que se detecta heteroplasmia por NGS en la Q2 en cada uno de los runs. rCRS: secuencia de referencia del genoma mitocondrial. MAYOR.MINOR: base mayoritaria y base minoritaria; % Het: frecuencia de heteroplasmia (% de la base minoritaria); DIR/REV: diferencia entre las lecturas en sentido directo y en sentido reverso; Prof: profundidad de lectura.

Posición	rCRS	MAYOR. MINOR.	% Het Sanger	% Het Run 1	DIR/REV Run 1 (ratio)	Prof. Run 1	% Het Run 2	DIR/REV Run 2 (ratio)	Prof. Run 2	% Het Run 3	DIR/REV Run 3 (ratio)	Prof. Run 3	% Het Run 4	DIR/REV Run 4 (ratio)	Prof. Run 4
272*	A	A/G		0,79	985/1033 (0,95)	2018		872/957 (0,91)	1529		154/155 (0,99)	309		474/596 (0,80)	1070
333*	T	T/C		0,76	1545/809 (1,91)	2354		1187/632 (1,88)	1819		67/167 (0,4)	234		149/545 (0,27)	694
2422*	T	T/C			1052/2467 (0,43)	3519	0,79	837/1951 (0,43)	2788		728/151 (4,82)	879		1383/333 (4,15)	1716
3635*	G	G/A		1,26	805/939 (0,86)	1744		661/778 (0,85)	1439		79/73 (1,08)	152	2,75	265/207 (1,28)	472
4055**	T	T/C	9,62	16,23	1538/1683 (0,91)	3221	16,43	1539/1733 (0,89)	3272	18,87	302/175 (1,73)	477	15,02	817/674 (1,21)	1491
4695*	T	T/C			1237/1381 (0,90)	2618		1295/1333 (0,97)	2628		284/303 (0,94)	587	1,15	636/672 (0,95)	1308
6243*	G	G/A		1,11	1190/979 (1,22)	2169		824/812 (1,01)	1636		157/227 (0,69)	384		359/431 (0,83)	790
6367*	T	T/C			924/1046 (0,88)	1970	0,94	586/687 (0,85)	1273		127/133 (0,95)	260		357/265 (1,35)	622
9449**	C	T/C	16,26	25,76	6969/5730 (1,22)	12699	26,44	4860/4145 (1,17)	9005	25,05	758/859 (0,88)	1617	26,29	1895/2715 (0,70)	4610
10927*	T	T/C			4718/4693 (1,01)	9411		4318/4383 (0,99)	8701	1,46	726/574 (1,26)	1300		1936/1919 (1,01)	3855
11016**	G	G/A	17,18	32,33	8143/13013 (0,63)	21156	31,18	8295/12362 (0,67)	20657	31,69	2093/955 (2,19)	3048	31,01	5345/2852 (1,87)	8197
11048*	T	T/C			10206/15759 (0,65)	25965		9665/14767 (0,65)	24432	0,82	2556/989 (2,58)	3545		6761/3159 (2,14)	9920
12261*	T	T/C		1,12	531/633 (0,84)	1164		623/664 (0,94)	1287		175/77 (2,27)	252		497/354 (1,40)	851
12749*	T	T/C			948/790 (1,2)	1738	0,92	896/834 (1,07)	1730		120/169 (0,71)	289		472/528 (0,89)	1000
16000*	G	G/A		11,45	3046/2903 (1,05)	5949	11,5	3256/2902 (1,12)	6158	9,41	607/743 (0,82)	1350	11,29	2127/2215 (0,96)	4342
16286*	C	T/C		1,85	3770/2511 (1,5)	6281	1,42	2726/2205 (1,24)	4931		253/530 (0,48)	783	1,25	1075/1566 (0,69)	2641

*Posiciones esperadas en homoplasmia según la secuenciación de Sanger. **Posiciones esperadas en heteroplasmia según la secuenciación de Sanger.

Tabla 19. Diferencias en la detección de heteroplasmas entre los 2 tipos de alineamiento en la Q1. ID: referencia; rCRS: secuencia de referencia del genoma mitocondrial; % Het: frecuencia de heteroplasma (base minoritaria); MAYOR.MINOR: base mayoritaria y base minoritaria.

Q1											
BWA-mem						BOWTIE2					
ID	Posición	rCRS	% Het	MAYOR. MINOR.	Profundidad de lectura	ID	Posición	rCRS	% Het	MAYOR. MINOR.	Profundidad de lectura
						Q1_run1	3517	T	0,89	T/C	2141
						Q1_run2	3443	T	0,85	T/C	3902
Q1_run2	3482	A	0,84	A/G	2151						

Tabla 20. Diferencias en la detección de heteroplasmas entre los 2 tipos de alineamiento en la Q2. ID: referencia; rCRS: secuencia de referencia del genoma mitocondrial; % Het: frecuencia de heteroplasma (base minoritaria); MAYOR.MINOR: base mayoritaria y base minoritaria.

Q2											
BWA-mem						BOWTIE2					
ID	Posición	rCRS	% Het	MAYOR. MINOR.	Profundidad de lectura	ID	Posición	rCRS	% Het	MAYOR. MINOR.	Profundidad de lectura
Q2_run1	333	T	0,76	T/C	2354						
						Q2_run1	344	T	0,86	T/C	2690
						Q2_run1	3323	T	0,79	T/C	3796
Q2_run2	2422	T	0,79	T/C	2787						
Q2_run2	12749	T	0,92	T/C	1730						
						Q2_run2	13336	T	0,89	T/C	1458
						Q2_run3	14325	T	0,84	T/C	2741
						Q2_run3	14459	G	1,04	G/A	2303
						Q2_run3	14530	T	0,87	T/A	2310
						Q2_run3	14534	C	0,78	C/A	2816
						Q2_run3	14569	G	0,93	G/A	2681
						Q2_run3	14819	T	0,71	T/C	2392
						Q2_run3	14853	T	0,89	T/C	2368
Q2_run4*	3635	G	2,75	G/A	472						

*Heteroplasma detectada en el run 1 en ambos tipos de alineamiento.

5.1.3. Secuenciación de la región que puede presentar co-amplificación entre el DNAmT y el NUMT

Detección de co-amplificación entre el DNAmT y el NUMT

En base al alineamiento entre la secuencia del mtDNA de los individuos Q1 y Q2 obtenida por la secuenciación de Sanger (Ramos et al. 2013) y la secuencia del NUMT, todas aquellas posiciones que se espera que estén presentes como mezclas de bases porque son posiciones no idénticas entre el DNAmT y el NUMT son detectadas, como mínimo, en alguno de los runs en ambas quimeras (Tabla 21). En

este sentido, se evidencia la existencia de co-amplificación del NUMT en ambas muestras analizadas. Atendiendo al porcentaje de co-amplificación, se observa que en la Q1 es superior respecto a la Q2: mientras que en la Q1 el porcentaje promedio se sitúa en un 3,95% en el *run 1* y 3,64% en el *run 2* (valor promedio de los *runs*= 3,79%), el porcentaje promedio en la Q2 se sitúa en un 2,14%, 1,87%, 1,63% y 1,70% en el *run 1*, *run 2*, *run 3* y *run 4*, respectivamente (valor promedio de los *runs*= 1,83%). Por ejemplo, la posición 8140 presenta un porcentaje de co-amplificación entre el 1,24% y 2,44% en la Q2, mientras que en la Q1 este porcentaje aumenta hasta situarse entre un 4,6% y 5,05% (Tabla 21).

Mezclas de bases en posiciones no idénticas entre el DNAmT y el NUMT

De las 28 mezclas de bases esperadas para la Q1, todas se han detectado en algún *run*, y 26 (92,86%) fueron detectadas por NGS en todos los *runs*. En el caso de la Q2, todas las 29 mezclas de bases esperadas se han detectado en algún *run*, pero sólo 17 (58,62%) se detectaron en todos los *runs* (Tabla 21).

Teniendo en cuenta la profundidad de lectura obtenida en cada *run* para ambas quimeras (Tabla 21) y el número de posiciones no idénticas no detectadas, los resultados muestran una correlación negativa significativa entre la profundidad de lectura y la aparición de falsos negativos en posiciones no idénticas: a medida que la profundidad de lectura disminuye, la cantidad de falsos negativos aumentan (Correlación de *Pearson* 0,468; $p < 0,01$). En este sentido, el *run 3* de la Q2 presenta la profundidad de lectura más baja (profundidad promedio de 1491,59X) y, por lo tanto, el menor número de posiciones no idénticas detectadas (13/30) (Tabla 21).

Asimismo, algunos falsos negativos en posiciones no idénticas se detectaron incluso obteniendo una profundidad elevada, lo que sugiere que otros factores pueden conducir a una subestimación de las mezclas de bases. Por un lado, las posiciones 8455 y 8461 en la Q1 se detectaron en el *run 2*, pero no en el *run 1* con una profundidad mayor (Tabla 21). Para estas posiciones se analizaron detenidamente los filtros que se aplicaron para la detección de las heteroplasmas (véase apartado 3.5.2). Los datos brutos revelan la presencia de estas posiciones como mezclas de bases, aunque el sesgo producido entre cadenas (valor >1) hace que la frecuencia de cada base no sea consistente entre la cadena en sentido directo y la cadena en sentido reverso, provocando que ambas mezclas de bases se descarten en este paso (Guo et al. 2012a; Guo et al. 2012b) (Tabla 22). Además, observando los resultados de la calidad de las lecturas en los archivos Fastq, se advierte que la calidad disminuye en las lecturas R2 (Figura 21). Alternativamente, dicho sesgo (valor >1) se detectó también utilizando el alineamiento *Bowtie2* (Tabla 23).

Tabla 21. Mezclas de bases detectadas en la región analizada del DNAm entre 6739-8910 pb en la Q1 y la Q2. Se espera que todas las posiciones sean no idénticas entre la secuencia del mtDNA y del NUMT, excepto aquellas señaladas con * en las que se espera que sean posiciones idénticas. La base más frecuente se indica en mayúscula, mientras que la base minoritaria se indica en minúscula. ND: mezclas de bases no detectadas.

Muestra	Posición	mtDNA individuo/NUMT		Run 1 (Profundidad promedio 10519,62X)		Run 2 (Profundidad promedio 7071,62X)		Muestra	Posición	mtDNA individuo/NUMT		Run 1 (Profundidad promedio 8995,70X)		Run 2 (Profundidad promedio 8200,77X)		Run 3 (Profundidad promedio 1491,59X)		Run 4 (Profundidad promedio 4502,56X)								
		% mezcla bases	Profundidad de lectura	% mezcla bases	Profundidad de lectura	% mezcla bases	Profundidad de lectura			% mezcla bases	Profundidad de lectura	% mezcla bases	Profundidad de lectura	% mezcla bases	Profundidad de lectura	% mezcla bases	Profundidad de lectura	% mezcla bases	Profundidad de lectura	% mezcla bases	Profundidad de lectura	% mezcla bases				
Q1	6935	C/t	2,56	16550	2,64	7188	6935	C/t	1,3	7134	1,6	6134	1,43	1185	1,66	3665	6938	C/t	1,1	7151	1,44	6174	1,75	1198	1,56	3662
	7146	A/g	3,16	15946	3	6723	7028	T/c	2,15	6750	2,04	5492	1,77	1188	2	3307	7146	A/g	2,08	6650	2,18	5089	1,29	1320	1,67	3407
	7232	C/t	2,53	17525	2,07	6858	7232	C/t	1,28	6788	1,55	5341	ND	1407	ND	3462	7256	C/t	1,47	6056	1,74	5060	ND	1309	1,02	3220
	7256	C/t	2,88	16686	2,55	6157	7256	C/t	1,47	6056	1,74	5060	ND	1309	1,02	3220	7270*	T/c	3,69	6047	2,85	5013	2,77	1264	3,54	3164
	7316	G/a	3,24	14567	2,69	5432	7316	G/a	1,41	5386	1,1	4536	ND	1203	1,23	2765	7521	G/a	1,03	4672	1,26	4359	ND	1002	0,96	2591
	7521	G/a	2,81	12855	2,41	4943	7521	G/a	1,89	6285	1,57	4972	1,33	1206	1,22	2865	7650	C/t	1,89	6285	1,57	4972	1,33	1206	1,22	2865
	7650	C/t	3,13	14225	2,9	6414	7705	T/c	4	14223	3,79	6909	1,59	1211	1,93	2951	7810	C/t	1,97	7191	1,59	5273	1,24	1211	1,93	2951
	7663*	C/t	1,29	14103	ND	6326	7810	C/t	4,58	13947	4,8	7089	1,78	1111	1,56	2878	7868	C/t	3	7239	1,78	5392	1,44	1111	1,56	2878
	7705	T/c	4	14223	3,79	6909	7868	C/t	4,16	12630	3,77	6638	2,03	1084	1,56	2954	7891	C/t	3,25	7835	2,03	5873	1,66	1084	1,56	2954
	7891	C/t	4,16	12630	3,77	6638																				

Tabla 21. Continuación.

										Q1										Q2																																																																																																																																																																																																																																																																																											
7912	G/a	4,77	12244	4,9	7059	7891	C/t	2,7	7361	1,65	5503	2,08	1008	1,3	2762	7922*	T/c	1,4	8045	1,58	5942	1,57	1082	1,39	2802	8021	A/g	5,07	11309	4,6	6693	8021	A/g	3,61	8195	2,52	5645	1,94	1030	2,07	2848	8065	C/t	5,05	9351	4,6	5689	8065	G/a	2,86	8189	1,91	5433	ND	1040	1,76	2847	8140	G/a	5,44	8386	4,48	5005	8140	C/t	2,44	6917	1,95	4831	1,24	1051	1,66	2774	8167	T/c	6,12	7483	5,09	4305	8152	G/a	2,18	6552	2,15	4649	ND	955	1,81	2600	8203	C/t	6,79	6908	5,85	3829	8155*	G/a	ND	6851	0,79	4782	ND	967	ND	2642	8254*	C/t	3,18	5724	3,07	6063	8157*	T/c	4,33	6622	3,56	4546	4,52	996	3,44	2644	8392	G/a	5,51	6927	3,62	3338	8203	C/t	2,22	5856	1,64	4443	1,56	1088	1,29	2554	8455	C/t	ND	9747	3	4566	8380	C/t	2,45	4812	2,25	6322	1,8	1043	1,21	3052	8461	C/t	ND	10264	3,08	4939	8392	G/a	2,7	4474	2,34	3799	ND	767	1,51	1920	8503	T/c	3,17	11403	3,2	5346	8455	C/t	2,05	4984	2,29	4200	ND	838	1,81	2215	8545	G/a	3,1	8490	3,85	4158	8461	C/t	1,99	5380	2,3	4395	1,71	875	2,16	2318	8655	C/t	3,09	13862	3,56	6400	8503	T/c	2,21	5830	1,79	4534	1,95	870	1,84	2442	8677	A/c	3,23	13285	3,15	6200	8545	G/a	2,36	4242	ND	3194	ND	596	2,19	1825	8701	A/g	3	13976	2,92	6757	8655	C/t	1,86	6511	1,97	5534	ND	1059	1,89	2905	8718	A/g	3,26	14135	3,44	6926	8677	A/c	1,87	6408	1,88	5544	ND	1039	2,2	2824	8878*	C/t	1,16	13238	1,05	6383	8701	A/g	1,88	7115	1,76	5975	ND	1041	2,21	2983	Promedio**					3,95	12660,85	3,64	5938,86	Promedio**					2,14	6403,46	1,87	5132,85	1,63	1084,75	1,70	2788,37
Promedio**					3,95	12660,85	3,64	5938,86	Promedio**					2,14	6403,46	1,87	5132,85	1,63	1084,75	1,70	2788,37																																																																																																																																																																																																																																																																																										

**Promedio calculado sin tener en cuenta las posiciones no detectadas ni tampoco las posiciones idénticas entre las secuencias del DNAmT y el NUMT.

Tabla 22. Cálculo del valor del sesgo entre cadenas para las posiciones 8455 y 8461 en ambos *runs* (valores >1) con el alineamiento *BWA-mem* en la Q1. En color rojo se observa las diferencias en el porcentaje de cada base para la cadena en sentido directo y en sentido reverso. ID: referencia; %: porcentaje de mezcla de bases (base minoritaria); Prof. DIR: profundidad de lectura en sentido directo; Prof. REV: profundidad de lectura en sentido reverso; A/a: frecuencia de adeninas; C/c: frecuencia de citosinas; G/g: frecuencia de guaninas; T/t: frecuencia de timinas; d: frecuencia de deleciones; SC: valor del sesgo entre cadenas.

BWA-mem															
ID	POS.	%	Prof. DIR.	Prof. REV.	Cadena sentido directo					Cadena sentido reverso					SC
					A	C	G	T	d	a	c	g	t	d	
Q1 Run 1	8455		8183	1564	0,0010	0,9708	0,0000	0,0281	0,0001	0,0006	0,9277	0,0000	0,0716	0,0001	1,24*
Q1 Run 2	8455	3	1579	2987	0,0000	0,9626	0,0000	0,0367	0,0006	0,0000	0,9736	0,0000	0,0264	0,0006	0,34
Q1 Run 1	8461		8329	1935	0,0001	0,9718	0,0000	0,0280	0,0001	0,0005	0,9292	0,0000	0,0703	0,0001	1,18*
Q1 Run 2	8461	3,08	1807	3132	0,0000	0,9629	0,0000	0,0371	0,0000	0,0003	0,9722	0,0003	0,0271	0,0000	0,32

*El valor de sesgo entre cadenas es >1.

Tabla 23. Cálculo del valor del sesgo entre cadenas para las posiciones 8455 y 8461 en ambos *runs* (valores >1) con el alineamiento *Bowtie2* en la Q1. En color rojo se observa las diferencias en el porcentaje de cada base para la cadena en sentido directo y en sentido reverso. ID: referencia; %: porcentaje de mezcla de bases (base minoritaria); Prof. DIR: profundidad de lectura en sentido directo; Prof. REV: profundidad de lectura en sentido reverso; %A/%a: porcentaje de adeninas; %C/%c: porcentaje de citosinas; %G/%g: porcentaje de guaninas; %T/%t: porcentaje de timinas; %d: porcentaje de deleciones; SC: valor del sesgo entre cadenas.

BOWTIE2															
ID	POS.	%	Prof. DIR.	Prof. REV.	Cadena sentido directo					Cadena sentido reverso					SC
					A	C	G	T	d	a	c	g	t	d	
Q1 Run 1	8455		8175	1659	0,0012	0,9705	0,0000	0,0283	0,0000	0,0036	0,9265	0,0000	0,0675	0,0000	1,13*
Q1 Run 2	8455	2,97	1593	3025	0,0000	0,9636	0,0000	0,0358	0,0006	0,0000	0,9732	0,0003	0,0264	0,0006	0,32
Q1 Run 1	8461		8315	2181	0,0004	0,9725	0,0000	0,0271	0,0001	0,0018	0,9335	0,0000	0,0628	0,0001	1,04*
Q1 Run 2	8461	3	1865	3165	0,0000	0,9630	0,0000	0,0359	0,0011	0,0003	0,9728	0,0003	0,0265	0,0011	0,31

*El valor de sesgo entre cadenas es >1

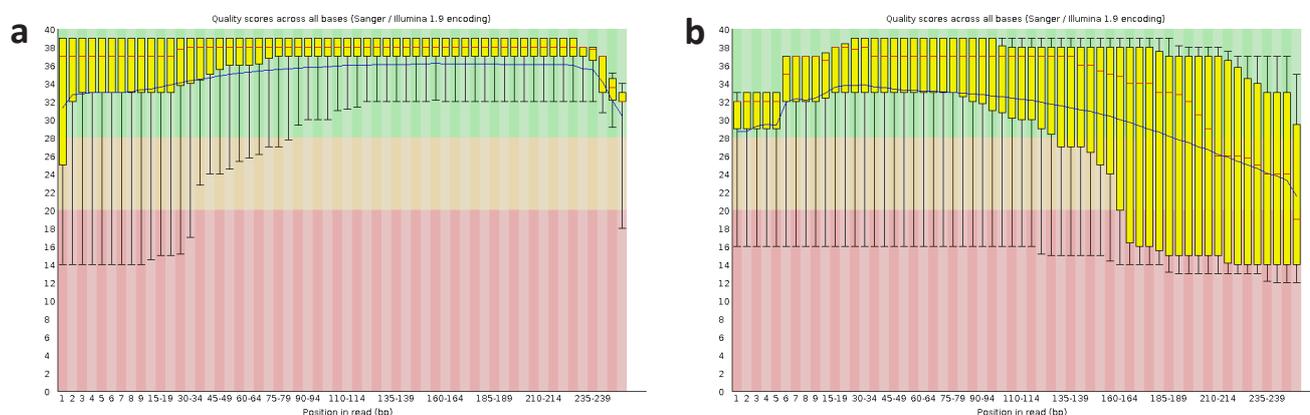


Figura 21. a) Calidad de las lecturas R1 de Q1_run1. Se observa que la calidad es buena a lo largo de los 250pb. b) Calidad de las lecturas R2 de Q1_run1. Se observa que la calidad disminuye notablemente a partir de los 150 pb.

Por otro lado, en la Q2 la posición 8545 se detecta como mezcla de bases únicamente en el *run 1* y *run 4*, pero no en el *run 2* y *run 3* (Tabla 20). El hecho de que no se detecte en el *run 3* puede ser debido a la baja profundidad de lectura obtenida (596X). Sin embargo, la profundidad de lectura obtenida en el *run 2* para esta posición (3194X) es superior a la obtenida en el *run 4* (1825X) y, por lo tanto, este factor tampoco debería ser la razón por la que no se detecte mezcla de bases en el *run 2*. Si se analizan todos los datos obtenidos de esta posición en el *run 2*, se observa que la base minoritaria detectada con las lecturas en sentido directo difiere de la base minoritaria detectada con las lecturas en sentido reverso, por lo que esta mezcla de bases también se elimina a partir de los datos brutos (Tabla 24).

Tabla 24. Posición 8545 de la Q2 en el *run 2*. En color rojo se observa que la base minoritaria detectada en sentido directo difiere de la base detectada en sentido reverso. ID: referencia; % HET: porcentaje de heteroplasmia; rCRS: secuencia de referencia del genoma mitocondrial; BASE MAYOR. DIR: base mayoritaria detectada con la cadena en sentido directo; BASE MINOR. DIR: base minoritaria detectada con la cadena en sentido directo; BASE MAYOR. REV: base mayoritaria detectada con la cadena en sentido reverso; BASE MINOR. REV: base minoritaria detectada con la cadena en sentido reverso.

ID	Posición	rCRS	BASE MAYOR. DIR.	BASE MINOR. DIR.	BASE MAYOR. REV	BASE MINOR. REV
Q2 Run2	8545	G	G	T	G	A

Mezclas de bases en posiciones idénticas entre el DNAmT y el NUMT

Un total de 3 posiciones (7663, 8254 y 8878) en la Q1 y 4 posiciones (7270, 7922, 8155 y 8157) en la Q2 se detectaron como mezclas de bases en posiciones idénticas entre las secuencias del DNAmT y el NUMT (Tabla 21). Teniendo en cuenta que el NUMT que se co-amplifica puede ser polimórfico y presentar variabilidad inter e intrapoblacional, se ha valorado la la variabilidad en la base de datos de los 1000 Genomas (Auton et al. 2015). Únicamente en la posición 632834 (GCRh38_Chr.1) del NUMT, que corresponde a la posición 7663 del DNAmT de la Q1, presenta variación. Este resultado indicaría que el individuo con el que se ha amplificado el fragmento 5 podría presentar variabilidad en dicha posición del NUMT, de manera que la mezcla de bases detectada podría ser una posición no idéntica entre el DNAmT del individuo y el NUMT (Figura 22). Respecto al resto de posiciones, parecen ser posiciones idénticas y, por lo tanto, podrían ser verdaderas heteroplasmas, aunque no se debe descartar la posibilidad de que estas posiciones también pueden ser falsos positivos.

Por último, es importante destacar que 5 de las 7 mezclas de bases (posiciones 8254 y 8878 en Q1, y las posiciones 7270, 7922 y 8157 en Q2) se detectaron en todos los *runs*, por lo que es poco probable que puedan ser falsos positivos. Las 2 mezclas de bases restantes (posiciones 7663 y 8155 en Q1 y Q2, respectivamente), que se detectaron a un nivel <1,5%, no fueron reproducibles en todos los *runs*, por lo que pueden ser falsos positivos.

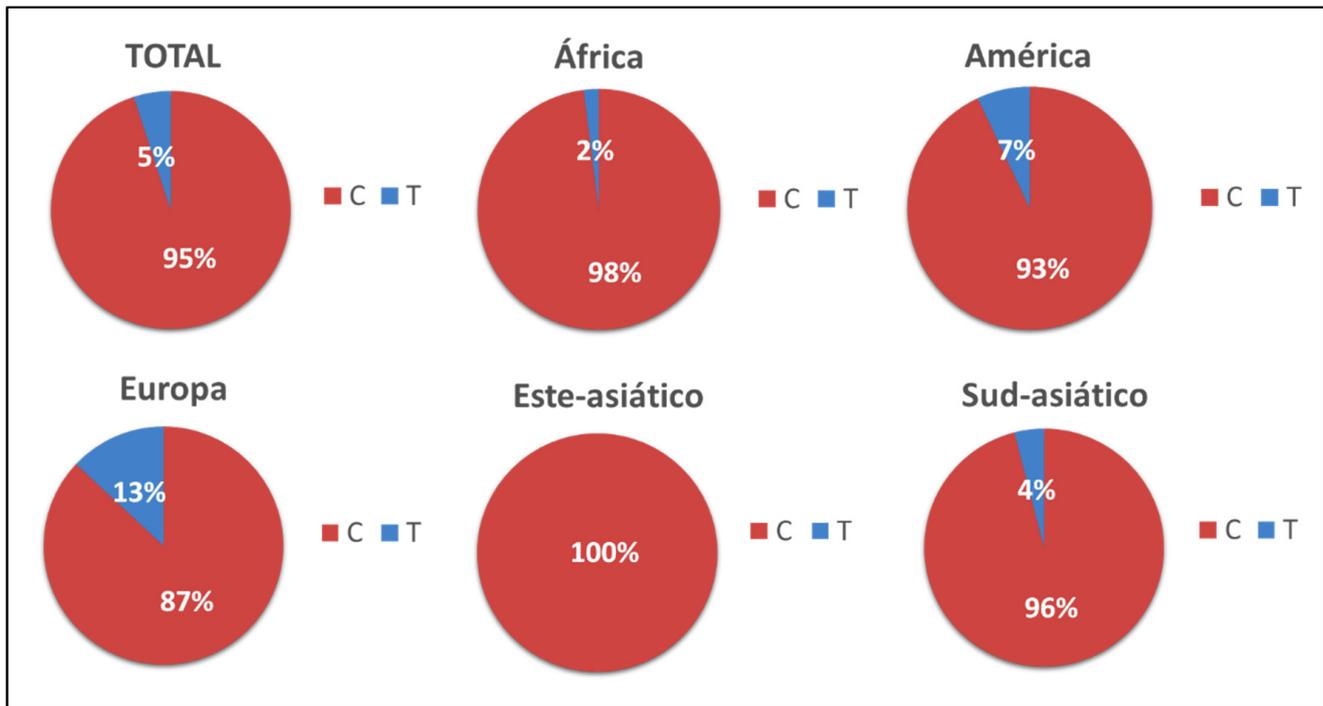


Figura 22. Frecuencias alélicas de la variante rs373437560 del NUMT (GCRh38_Ch1 posición 632834, que corresponde a la posición 7663 del DNAmT de la Q1) en grandes grupos poblacionales. Datos obtenidos de la base de datos de los 1000 Genomas (Auton et al. 2015).

Alineamiento BWA-mem vs. Alineamiento Bowtie2 entre el DNAmT y el NUMT

Si se aplica el alineamiento *Bowtie2* se obtiene el mismo resultado que utilizando el alineamiento *BWA-mem*, tanto a nivel de frecuencia como de profundidad de lectura. Sin embargo, mediante *Bowtie2* también se detecta la posición 8427Y como heteroplasma en el *run* 1 de la Q2 (porcentaje de heteroplasma= 0,68%; profundidad de lectura= 5025X). Según el alineamiento, esta posición es idéntica entre la secuencia mitocondrial y el NUMT y, por lo tanto, puede considerarse tanto una posible posición en heteroplasma como un falso positivo.

Establecimiento de los límites de detección de la heteroplasma mitocondrial

A partir de las profundidades de lectura obtenidas es posible establecer varios límites de detección de heteroplasma, considerando las frecuencias de las bases minoritarias detectadas. En primer lugar, todas las mezclas de bases con un nivel >3% se detectan en todos los casos, incluso en el *run* 3 de la Q2, donde la profundidad de lectura en este *run* se sitúa ~1000X. Por lo tanto, cuando la frecuencia de la heteroplasma es >3%, la presencia de falsos positivos y falsos negativos parece estar prácticamente

ausente con una profundidad de lectura de $\sim 1000X$. En segundo lugar, es posible detectar mezclas de bases a un nivel entre 1,5% y 3% cuando la profundidad de lectura es más elevada ($\sim 3000X$). Este límite se puede establecer observando los *runs* de la Q2 (Tabla 25): mientras que algunas mezclas de bases se detectan en los *runs* 1 y 2 (la profundidad promedio en estas posiciones es de 6079X en el *run* 1 y 4892X en el *run* 2), estas mismas no se detectan en el *run* 3 (la profundidad promedio en estas posiciones es de 1052X). No obstante, cuando la profundidad de lectura aumenta en el *run* 4 (la profundidad promedio en estas posiciones es de 2733X), estas mezclas de bases pueden detectarse. Por lo tanto, con una profundidad de lectura de $\sim 3000X$ ya no se detectan, *a priori*, ni falsos positivos ni falsos negativos. Por último, existen muchas discrepancias respecto a las mezclas de bases con una frecuencia muy baja ($<1,5\%$). Algunas mezclas de bases a esta frecuencia se pueden detectar en algunos casos, pero también se detectan falsos negativos, sobre todo en el *run* 3 de la Q2 donde se obtiene una profundidad de lectura más baja. Por esta razón, cuando la frecuencia de heteroplasmia es $<1,5\%$, la discriminación entre falsos positivos y verdaderas heteroplasmas es muy compleja incluso teniendo una profundidad de lectura elevada y no es posible, además, descartar la presencia de falsos negativos.

Tabla 25. Análisis de determinadas mezclas de bases en la Q2 para poder establecer el límite de detección de la heteroplasmia mitocondrial de $1,5\% < X < 3\%$.

Muestra	Posición	mtDNA individuo/ NUMT	Run 1		Run 2		Run 3		Run 4	
			% mezcla bases	Profundidad de lectura						
Q2	7232	C/t	1,28	6788	1,55	5341	-	1407	-	3462
	7256	C/t	1,47	6056	1,74	5060	-	1309	1,02	3220
	7316	G/a	1,41	5386	1,1	4536	-	1203	1,23	2765
	7521	G/a	1,03	4672	1,26	4359	-	1002	0,96	2591
	8065	G/a	2,86	8189	1,91	5433	-	1040	1,76	2847
	8152	G/a	2,18	6552	2,15	4649	-	955	1,81	2600
	8167	T/c	2,41	5811	2,46	4279	-	962	1,86	2467
	8392	G/a	2,7	4474	2,34	3799	-	767	1,51	1920
	8455	C/t	2,05	4984	2,29	4200	-	838	1,81	2215
	8655	C/t	1,86	6511	1,97	5534	-	1059	1,89	2905
	8677	A/c	1,87	6408	1,88	5544	-	1039	2,2	2824
8701	A/g	1,88	7115	1,76	5975	-	1041	2,21	2983	

5.2. DISCUSIÓN

Los resultados de la secuenciación de las quimeras muestran la elevada dificultad de poder detectar mezclas de bases a bajas frecuencias, considerando los posibles sesgos obtenidos en relación a las diferencias observadas en la profundidad de lectura obtenida dentro del mismo *run* y entre *runs*. Todas las mezclas de bases que han sido detectadas a una baja frecuencia son difíciles de reproducir, no sólo a nivel de *runs*, sino también según el tipo de alineamiento seleccionado. En este sentido, cuando se detecta una mezcla de bases a baja frecuencia (<1,5%) es complicado determinar si se trata de un falso positivo debido a posibles errores introducidos durante el proceso de amplificación, de la elaboración de las librerías o de la secuenciación. Del mismo modo, es posible también que no se lleguen a detectar verdaderas mezclas de bases en determinadas posiciones, generándose falsos negativos.

Por otro lado, existen notables diferencias en el porcentaje de heteroplasma establecido en Sanger y en NGS, siendo superior en este último en la mayoría de casos. Esto hace que incluso la heteroplasma detectada en la posición 16000 en NGS, a un porcentaje ~10%, no se llegue a detectar en Sanger. En este sentido, el poder determinar, con una mayor frecuencia, heteroplasmas en NGS con un menor porcentaje de mezcla, puede explicar el por qué esta posición no se detectó en Sanger, ya que el límite de detección de la misma es inferior (Ramos et al. 2013).

En relación a los resultados obtenidos en la secuenciación del fragmento en que se puede producir co-amplificación de DNAmT y DNA nuclear (posición 6511-9220 del DNAmT), la presencia de mezcla de bases en aquellas posiciones no idénticas entre el DNAmT del individuo y el NUMT confirma la posibilidad de co-amplificación entre el DNAmT y el DNAn a través de los NUMTs y su detección por NGS, incluso siendo muy apreciable en algunas posiciones donde la co-amplificación alcanza hasta el 6%. Este resultado está de acuerdo con otros estudios que determinan que la presencia de secuencias idénticas o casi idénticas entre los NUMT y el DNAmT puede introducir un sesgo y generar falsos positivos a bajas frecuencias a partir de NGS (Albayrak et al. 2016). Además, durante la amplificación de DNAmT, la cantidad de NUMT amplificada es mayor en la Q1 que en la Q2 debido posiblemente a razones estocásticas: la diferencia observada puede estar relacionada con el estado de las muestras, de manera que, durante el proceso de amplificación, la cantidad de NUMT amplificado ha sido superior en la muestra de la Q1 (individuo 1_Z90) comparativamente a la muestra de la Q2 (individuo 4_Z24). Estos resultados refuerzan la importancia de llevar a cabo un diseño exhaustivo de *primers* de PCR para amplificar selectivamente el DNAmT, como ya sugirieron otros autores anteriormente (Ramos et

al. 2009; Ramos et al. 2011). Sin este análisis exhaustivo se podría estar aceptando ciertas mezclas de bases como verdaderas heteroplasmias que son, en realidad, producto de la co-amplificación.

Teniendo en cuenta los resultados obtenidos en el presente estudio, la estrategia de co-amplificación del NUMT se puede utilizar para establecer varios límites de detección de heteroplasmia. Las mezclas de bases observadas a bajas frecuencias (<1,5%) no siempre se pueden detectar en todos los *runs*, independientemente de la profundidad obtenida. Este resultado coincide con el que se ha encontrado en el resto de fragmentos y refuerza la teoría de que la discriminación de heteroplasmia a bajas frecuencias es un procedimiento muy complejo (Duan et al. 2018b; White et al. 2017). Sin embargo, es importante tener en cuenta que, aunque la sensibilidad de detección es más alta en el presente estudio –algunas mezclas de bases se detectan con una frecuencia <1,5% en algunos *runs*–, es posible ignorar algunas mezclas de bases verdaderas (falsos negativos) o aceptar errores de secuenciación (falsos positivos) a este nivel. Esto es un factor muy importante a tener en cuenta en algunos tipos de estudios, como aquellos basados en el análisis de la carga mutacional a nivel de heteroplasmia o estudios de caso-control, donde es posible obtener resultados sesgados dando lugar a malas interpretaciones en los resultados de la investigación. Por lo tanto, aunque se pueden detectar algunas heteroplasmias, consideramos que todas las mutaciones presentes en una frecuencia <1,5% deben ser excluidas de los análisis para evitar interpretaciones erróneas y generar resultados sesgados.

No obstante, cuando las mezclas de bases presentan valores ligeramente más altos –entre el 1,5% y el 3%–, éstas pueden ser reproducidas en todos los *runs* y en ambos tipos de alineamiento, siempre y cuando la profundidad de lectura sea más elevada (~3000X). Por lo tanto, en este rango la profundidad juega un papel muy importante en la detección de heteroplasmia mitocondrial (Jayaprakash et al. 2015). Finalmente, con una profundidad de ~1000X hasta ~3000X, se considera que una frecuencia mínima del 3% es necesaria para evitar, *a priori*, la detección de falsos positivos y negativos.

Los casos donde se produjeron diferentes errores de secuenciación, independientemente de la profundidad de lectura obtenida, conducen a una subestimación de determinadas mezclas de bases. En el caso de las posiciones 8455 y 8461 de la Q1, la baja calidad de las lecturas en sentido reverso hace que muchas de ellas no pasen los posteriores filtros de calidad, generando una disminución en la profundidad de las lecturas en este sentido y produciendo un sesgo entre cadenas (valores >1). Además, la presencia de este sesgo se detecta en ambos tipos de alineamiento, un resultado que se asemeja al publicado por Guo et al. (2012b), donde concluyen que el valor de sesgo entre cadenas parece no darse por un problema en el tipo de alineamiento. No obstante, teniendo en cuenta los valores obtenidos

(valores cercanos a 1), es importante tener en cuenta que, aunque el valor de sesgo puede ayudar a discriminar entre verdaderas heteroplasmas y falsos positivos, la aplicación de un valor de sesgo demasiado restrictivo puede generar falsos negativos y excluir verdaderas heteroplasmas (Guo et al. 2012a). Del mismo modo, la detección de diferentes bases entre las cadenas en sentido directo y reverso en una misma posición, como en el caso de la posición 8545 de la Q2, puede producir también un sesgo y generar resultados contradictorios, lo que evitaría detectar verdaderas mezclas de bases.

Varios trabajos se han enfocado en establecer su propio límite de detección de la heteroplasma (Tabla 26). Comparando los resultados obtenidos en el presente estudio con los descritos por otros autores se pueden remarcar varios aspectos. En primer lugar, nuestros resultados permiten establecer límites de detección de heteroplasma más bajos que los de otros estudios (Li et al. 2010; Tang and Huang 2010) que presentan sus propios límites con una baja reproducibilidad y detectando tanto falsos positivos como negativos. En segundo lugar, los resultados de He et al. (2010) estarían de acuerdo con los presentados en el presente estudio. Aunque detectan heteroplasmas con una sensibilidad >1%, establecen un límite de detección de heteroplasma de 1,6% con una profundidad de lectura muy elevada, lo que refuerza la teoría de que una profundidad de lectura mayor es necesaria para detectar heteroplasmas entre 1,5% y 3%, y que heteroplasmas inferiores a 1,6% no se pueden detectar con fiabilidad. De forma similar, algunos estudios han podido detectar variantes de bajo nivel (~2%) con una baja profundidad de lectura y obteniendo una buena reproducibilidad (Li and Stoneking 2012), pero no descartan que algunas de las heteroplasmas detectadas en su estudio puedan ser falsas debido a problemas de contaminación o lecturas quiméricas. En tercer lugar, aunque algunos estudios han podido establecer un límite de detección muy bajo ($\leq 1\%$) (Kloss-Brandstatter et al. 2015; McElhoe et al. 2014) la profundidad de lectura necesaria para establecerlo es muy superior en comparación con la descrita en el presente estudio. Sin embargo, es muy remarcable el hecho de que se puedan observar falsos negativos incluso con una profundidad de lectura de 35000X (Kloss-Brandstatter et al. 2015). Finalmente, otros estudios muestran más discrepancias. En comparación con nuestros resultados, Goto et al. (2011) fueron capaces de disminuir el nivel de detección de variantes a un nivel de ~2% con una profundidad de lectura de ~1000X, mientras que Li et al. (2015a) establecieron un límite de detección de heteroplasma de 0,5% con una profundidad de lectura de ~3500X, logrando unos resultados fiables y una buena reproducibilidad en ambos estudios. Es importante destacar que las diferencias en los análisis de los datos pueden dar lugar a discrepancias en los resultados. En este sentido, el flujo de trabajo aplicado en el estudio actual de mtDNA-Server está diseñado principalmente con el propósito de detectar y validar heteroplasmas a muy baja frecuencia ($\geq 1\%$). Además, aunque se usa la misma plataforma de NGS (Illumina), los modelos de secuenciación empleados pueden ser

diferentes (Genome Analyzer, Genome Analyzer II, HiSeq 2000, HiSeq 2500, MiSeq), y de este modo, muy pocos estudios aplican el modelo MiSeq para secuenciar (McElhoe et al. 2014).

Tabla 26. Trabajos basados en el establecimiento del límite de detección de la heteroplasma, dependiendo principalmente de la profundidad obtenida, para lograr la máxima sensibilidad y reproducibilidad.

Referencia	Límite de detección heteroplasma	Profundidad de lectura	Sensibilidad/Reproducibilidad
(He et al. 2010)	1,6%	16700X	Elevada sensibilidad.
(Tang and Huang 2010)	≥5%	1785X	Detección de falsos positivos al 1%. Especificidad baja a este nivel.
(Li et al. 2010)	10%	76X	Detección de falsos positivos y negativos.
(Goto et al. 2011)	2%	1170X	No detección de falsos positivos ni negativos con una ratio de error del 0,001 y una frecuencia del alelo minoritario del 1%.
(Li and Stoneking 2012)	2,3%	500X	Los 3 métodos aplicados muestran una buena especificidad (sin falsos positivos y negativos). Tasa de detección de falsos negativos <1%.
(McElhoe et al. 2014)	1%	5000X	Las variantes se observan de fácilmente y de forma reproducible.
	0,5%	8000X	
(Kloss-Brandstatter et al. 2015)	1%	35000X	Falsos negativos detectados*
(Li et al. 2015a)	0,5%	3458X	Los hallazgos originales son reproducibles con una ratio de falsos positivos <0,01.
Presente estudio (2018)	3%	1000X	Falsos positivos y negativos no se detectan aparentemente.
	1,5%-3%	3000X	

*Detectado por nuestro grupo en el Material Suplementario del artículo.

Considerando la gran cantidad de metodologías y estrategias utilizadas, así como los diferentes criterios y resultados descritos hasta el momento en la literatura, es muy difícil evaluar qué estrategia es más precisa para ajustarse a unos resultados más fiables, así como establecer un valor límite único para detectar la heteroplasma de DNAm. Teniendo en cuenta que la detección de heteroplasma a niveles muy bajos sigue siendo un desafío a resolver, la nueva metodología presentada en este estudio basada en la generación de mezclas de bases por co-amplificación de DNAm y DNAn proporciona un nuevo enfoque para establecer los límites de detección de la heteroplasma. A partir de todos los resultados mencionados anteriormente, esta nueva estrategia basada en la implementación de la tecnología Nextera XT® y la plataforma MiSeq asegura la detección de heteroplasmas con una frecuencia del 3% con una alta reproducibilidad y una profundidad relativamente baja de lectura

(~1000X). Además, los presentes resultados confirman el importante papel que ejerce la profundidad de lectura obtenida durante el proceso de NGS para obtener resultados precisos relacionados con la detección de heteroplasmia a niveles muy bajos.

6

CAPÍTULO IV
RESULTADOS y DISCUSION 3

6. CAPÍTULO IV RESULTADOS y DISCUSIÓN 3

ANÁLISIS del DNAm en INDIVIDUOS AFECTOS de GLIOMA

6.1. RESULTADOS

De los 32 pacientes analizados afectados de glioma (6 mujeres y 26 hombres), 8 presentan tumores de grado II, 7 tumores de grado III y 17 tumores de grado IV. No se ha observado ningún paciente con glioma de grado I. Tanto las características clínicas como demográficas de las muestras se describen en la Tabla 27. Respecto a la edad media de detección del glioma según el grado de malignidad del tumor, se observa que es prácticamente idéntica en los tumores de grado II y III (43,25 años en los tumores de grado II y 43 años en los tumores de grado III) y superior en los tumores de grado IV (56,71 años).

6.1.1. Variaciones del DNAm en individuos afectados de glioma

En la Tabla 28, para cada paciente analizado se presenta el haplotipo mitocondrial del tejido tumoral y sanguíneo. A partir de las variantes detectadas se estableció el haplogrupo mitocondrial de cada uno de los individuos que se muestra también en la tabla 28.

Tabla 27. Características demográficas y clínicas de las 32 muestras de estudio. H: hombre; M: mujer; KPS: Escala Karnofsky, forma estándar de medir la capacidad de los pacientes con cáncer para realizar tareas normales; ND: No disponible.

Muestra	Sexo	Edad	Anatomía patológica del Tumor	Grado del Tumor	Datos Clínicos		Datos radiológicos			
					Presentación Clínica*	KPS al ingreso	KPS al alta	Tamaño Lesión (mm)	Lesión Multifocal	Localización
3	H	49	Glioblastoma	IV	Focalidad, hemiparesia derecha Afasia	70	70	41x31x44	No	Frontal izquierdo
5	H	71	Glioblastoma	IV	Focalidad, afasia de expresión Desorientación	90	90	ND	No	Frontal izquierdo
6	H	64	Glioblastoma con componente oligodendroglial	IV	Focalidad, hemiparesia derecha Afasia Cefalea	80	80	40x40x50	No	Temporal derecho
7	H	37	Astrocitoma difuso	II	Crisis comiciales parciales	100	100	ND	No	Insular derecho
10	H	29	Astrocitoma alto grado	III	Focalidad, hemiparesia derecha	80	80	43x44x46	No	Ganglio basal izquierdo
13	H	53	Glioblastoma	IV	Focalidad, hemiparesia izquierda Astenia	90	90	45	No	Temporal derecho y cingulo
16	H	43	Glioblastoma con componente oligodendroglial	IV	Crisis comiciales parciales	90	90	ND	No	Parietal derecho
17	M	56	Glioblastoma	IV	Crisis comiciales parciales	90	90	18x17x24	No	Parietal izquierdo
18	M	34	Oligoastrocitoma de predominio astrocitario con foco de oligoastrocitoma atípico	III	Crisis comiciales parciales	90	90	ND	No	Insular derecho
19	H	43	Glioblastoma	IV	Crisis convulsiva TCG	90	80	ND	Sí	Temporal izquierdo
20	H	53	Oligoastrocitoma	II	Crisis convulsiva TCG	70	60	40	No	Frontal izquierdo
21	H	59	Glioblastoma	IV	Focalidad: hemiparesia izquierda Deterioro cognitivo	80	90	50x40	No	Parietal derecho
26	H	58	Glioblastoma con componente oligodendroglial	IV	Cefalea Deterioro cognitivo	90	90	ND	No	Temporal derecho
27	M	46	Astrocitoma difuso fibrilar	II	Crisis convulsiva TCG	90	90	ND	No	Frontal izquierdo
30	H	31	Oligoastrocitoma (pérdida alélica D1p y 19q)	II	Crisis comiciales parciales	100	100	53x33x35	No	Insular derecho
31	H	73	Glioblastoma	IV	Focalidad: problemas en la mano derecha Deterioro cognitivo	90	50	24x20x22	No	Parietal izquierdo

Tabla 27. Continuación.

Muestra	Sexo	Edad	Anatomía patológica del Tumor	Grado del Tumor	Datos Clínicos		Datos radiológicos			
					Presentación Clínica*	KPS al ingreso	KPS al alta	Tamaño Lesión (mm)	Lesión Multifocal	Localización
31BIS	H	53	Glioblastoma con componente oligodendroglial (pérdida alélica D1p)	IV	Crisis convulsiva TCG Déficit cognitivo	70	70	61x44x53	Sí	Frontal izquierdo
33	M	70	Glioblastoma	IV	Cefalea Deterioro cognitivo	90	90	55x35x47	No	Frontal derecho
34	H	52	Astrocitoma gemistocítico (ausencia de pérdida alélica D1p y 19q)	II	Crisis comiciales parciales Generalización secundaria	90	90	ND	No	Temporal-medial derecho
35	H	53	Astrocitoma difuso	II	Crisis comiciales parciales Generalización secundaria Cefalea	100	90	ND	No	Temporal izquierdo
36	H	31	Astrocitoma anaplásico	III	Crisis comiciales parciales Generalización secundaria	100	50	ND	No	Insular izquierdo
41	H	34	Astrocitoma anaplásico	III	Crisis convulsiva TCG	100	70	37x36x39	No	Frontal izquierdo
43	H	50	Astrocitoma anaplásico	III	Crisis comiciales parciales	100	100	ND	No	Insular derecho
44	H	59	Glioblastoma	IV	Crisis comiciales parciales Hematoma tumoral	90	90	ND	Sí	Temporal derecho
45	H	35	Astrocitoma difuso con componente gemistocítico	II	Cefalea Focalidad	80	80	ND	No	ND
50	H	56	Glioblastoma	IV	Focalidad	60	40	ND	ND	ND
51	H	63	Astrocitoma anaplásico	III	ND	90	90	ND	ND	ND
53	H	57	Glioblastoma	IV	Crisis comiciales Focalidad	70	70	ND	ND	ND
54	M	65	Glioblastoma	IV	Cefalea Focalidad	80	80	ND	ND	ND
55	M	35	Glioblastoma de células gigantes	IV	Focalidad	80	80	ND	ND	ND
56	H	60	Astrocitoma anaplásico	III	Crisis comiciales	90	90	ND	ND	ND
59	H	39	Oligoastrocitoma (no pérdida de 1p ni 19q, positivo para p53)	II	Crisis comiciales	80	80	ND	ND	ND

*Afasia: alteración del lenguaje a nivel general (sin especificación del tipo de trastorno); Astenia: estado de fatiga, cansancio; Crisis comiciales: crisis epilépticas; Crisis comiciales parciales: crisis epilépticas en las que no se ve alterado el nivel de consciencia; Crisis convulsiva TCG: crisis epiléptica Tónico-Clónica Generalizada. Se basa en una primera fase tónica (rigidez corporal) seguida de una fase clónica (rápidas convulsiones en las extremidades); Focalidad: anomalías en la función nerviosa; Focalidad, afasia de expresión: trastorno del lenguaje relacionado con la expresión del lenguaje (el paciente entiende, pero no habla bien); Focalidad, hemiparesia derecha o izquierda: falta de fuerza para el movimiento de las extremidades izquierdas o derechas; Generalización secundaria: crisis epiléptica que inicia parcial (afecta a una parte del cuerpo) y no altera el nivel de consciencia y que posteriormente se generaliza (afecta a todo el cuerpo y se pierde la consciencia).

Tabla 28. Haplotipo y haplogrupo establecido a partir de las muestras de tumor y sangre de cada uno de los pacientes analizados. Las mutaciones no definitivas de haplogrupo se indican en negrita, mientras que las mutaciones que sólo se detectan en un tejido se indican mediante subrayado. Las mutaciones en heteroplasmia se indican con diferentes colores: en azul aparecen aquellas heteroplasmas en que la variante minoritaria presenta un porcentaje >5%; en naranja aparecen aquellas heteroplasmas en que la variante minoritaria presenta un porcentaje entre 3-5%; en rojo aparecen aquellas heteroplasmas en que la variante minoritaria presenta un porcentaje que no supera el 3%, pero que se detecta en el otro tejido con un porcentaje >3%.

Muestra	Haplogrupo	Haplotipo tumoral	Haplotipo sanguíneo
3	J1c2e2	73G, 185A, 188G, 228A, 263G, 295T, 462T, 489C, 709A, 750G, 1438G, 2706G, 3010A, 4216C, 4232C, 4769G, 7028T, 8404C, 8860G, 10398G, 11251G, 11719A, 12612G, 13708A, 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16278T, 16366T, 16311C , 16519C	73G, 185A, 188G, 228A, 263G, 295T, 462T, 489C, 709A, 750G, 1438G, 2706G, 3010A, 4216C, 4232C, 4769G, 7028T, 8404C, 8860G, 10398G, 11251G, 11719A, 12612G, 13708A, 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16278T, 16366T, 16311C , 16519C
5	J1c5c1	73G, 185A (4,72%) , 188G (29%) , 228A, 263G, 295T, 462T, 489C, 750G, 1040Y (3,37%) , 1438G, 2706G, 3010A, 3531A, 3721G, 4216C, 4769G, 5198G, 7028T, 8734A, 8860G, 8994A, 10398G, 11251G, 11719A, 12612G, 13708A, 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16300G, 16519Y (8,31%)	73G, 185A (3,79%) , 188G (37,34%) , 228A, 263G, 295T, 462T, 489C, 750G, 1040Y (15,48%) , 1438G, 2706G, 3010A, 3531A, 3721G, 4216C, 4769G, 5198G, 7028T, 8734A, 8860G, 8994A, 10398G, 11251G, 11719A, 12612G, 13708A, 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16300G
6	T2b2b	73G, 263G, 709A, 750G, 930A (6,28%) , 1438G, 1888A, 2706G, 4216C, 4769G, 4917G, 5147A, 7028T, 8697A, 8860G, 10463C, 11242G, 11251G, 11719A, 11812G, 13368A, 13404C , 14233G, 14766T, 14905A, 15326G, 15452A, 15607G, 15928A, 16126C, 16294T, 16296T, 16519C	73G, 263G, 709A, 750G, 930A (2,74%) , 1438G, 1888A, 2706G, 4216C, 4769G, 4917G, 5147A, 7028T, 8697A, 8860G, 10463C, 11242G, 11251G, 11719A, 11812G, 13368A, 14233G, 13404C , 14766T, 14905A, 15326G, 15452A, 15607G, 15928A, 16126C, 16294T, 16296T, 16519C
7	H89	263G, 750G, 1438G, 1787R (5,07%) , 4769G, 6932G, 8068T, 8860G, 10057C , 12696C, 15326G, 16465T , 16519C	263G, 750G, 1438G, 4769G, 6932G, 8068T, 8860G, 10057C , 12696C, 15326G, 16465T , 16519C
10	J2a1a1	73G, 150T, 152C, 195C, 215G, 263G, 295T, 319C, 489C, 513A, 750G, 1438G, 2392C , 2706G, 4216C, 4769G, 7028T, 7476T, 7789A, 8860G, 10398G, 10499G, 11251G, 11377A, 11719A, 12612G, 13708A, 13722G, 14133G, 14766T, 15257A, 15326G, 15452A, 16069T, 16126C, 16145A, 16231C, 16261T	73G, 150T, 152C, 195C, 215G, 263G, 295T, 319C, 489C, 513A, 750G, 1438G, 2392C , 2706G, 4216C, 4769G, 7028T, 7476T, 7789A, 8860G, 10398G, 10499G, 11251G, 11377A, 11719A, 12612G, 13708A, 13722G, 14133G, 14766T, 15257A, 15326G, 15452A, 16069T, 16126C, 16145A, 16231C, 16261T
13	J1c2c	73G, 146C, 185A, 188G, 228A, 263G, 295T, 462T, 489C, 750G, 1438G, 2706G, 3010A, 4216C, 4769G, 7028T, 8860G, 10398G, 10685A, 11251G, 11719A, 12612G, 13281C, 13708A, 13933G, 14106C , 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16519C	73G, 146C, 185A, 188G, 228A, 263G, 295T, 462T, 489C, 653R (5,36%) , 750G, 1438G, 2706G, 3010A, 4216C, 4769G, 7028T, 8860G, 10398G, 10685A, 11251G, 11719A, 12612G, 13281C, 13708A, 13933G, 14106C , 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16519C

Tabla 28. Continuación.

Muestra	Haplogrupo	Haplotipo tumoral	Haplotipo sanguíneo
16	U4b1a	73G, 195C, 263G, 499A, 750G, 1438G, 1811G, 2706G, 3672G, 4224T , 4646C, 4769G, 5999C, 6047G, 7028T, 7705C, 8860G, 11332T, 11339C, 11467G, 11719A, 12308G, 12372A, 13731G , 14091G , 14524G , 14620T, 14766T, 15326G, 15693C, 16186T (16,13%) , 16189C , 16356C, 16519C	73G, 195C, 263G, 499A, 750G, 1438G, 1811G, 2706G, 3672G, 4224T , 4646C, 4769G, 5999C, 6047G, 7028T, 7705C, 8860G, 11332T, 11339C, 11467G, 11719A, 12308G, 12372A, 13731G , 14091G , 14524G , 14620T, 14766T, 15326G, 15693C, 16186T (13,3%) , 16189C , 16356C, 16519C
17	H1j	263G, 750G, 1438G, 3010A, 4733C, 4769G, 8860G, 9361Y (1,16%) , 13911G , 15326G, 16519C	263G, 750G, 1438G, 3010A, 4733C, 4769G, 8860G, 9361Y (3,97%) , 13911G , 15326G, 16519C
18	H1ap1	152C, 263G, 292C, 750G, 1438G, 1709A, 3010A, 4807C (40,32%) , 4769G, 5780A, 8410T, 8860G, 13135R (4,4%) , 14374C , 15326G, 16093C , 16183C , 16189C, 16201T (8,10%) , 16293G , 16519C	152C, 263G, 292C, 750G, 1438G, 1709A, 3010A, 4769G, 5780A, 8410T, 8860G, 14374C , 15326G, 16093C , 16183C , 16189C, 16201T (20,32%) , 16293G , 16519C
19	U5a2c	73G, 150T , 263G, 750G, 1438G, 2706G, 3173A , 3197C, 4769G, 7028T, 8860G, 9145A , 9477A, 10619T, 11467G, 11719A, 12308G, 12372A, 13617C, 13763Y (7,36%) , 14766T, 14793G, 15316G , 15326G, 16192T, 16256T, 16270T, 16526A	73G, 150T , 263G, 750G, 1438G, 2706G, 3173A , 3197C, 4769G, 7028T, 8860G, 9145A , 9477A, 10619T, 11467G, 11719A, 12308G, 12372A, 13617C, 13763Y (16,04%) , 14766T, 14793G, 15316G , 15326G, 16192T, 16256T, 16270T, 16526A
20	H1r1	146C, 263G, 750G, 1438G, 3010A, 4769G, 8860G, 9356T, 11711R (4,48%) , 15326G, 16278T, 16519C	146C, 263G, 750G, 1438G, 3010A, 4769G, 8860G, 9356T, 15326G, 16278T, 16519C
21	K1a1	73G, 114T, 152Y (39,66%) , 263G, 497T, 750G, 1189C, 1438G, 1518T , 1811G, 2706G, 3480G, 4769G, 7028T, 7699T , 8860G, 9055A, 9698C, 10398G, 10550G, 11084G , 11299C, 11467G, 11719A, 11914A, 12046G , 12308G, 12372A, 12597G , 13578T , 14167T, 14766T, 14798C, 15326G, 15340G , 16224C, 16311C, 16519C	73G, 114T, 263G, 497T, 750G, 1189C, 1438G, 1518T , 1811G, 2706G, 3480G, 4769G, 7028T, 7699T , 8860G, 9055A, 9698C, 10398G, 10550G, 11084G , 11299C, 11467G, 11719A, 11914A, 12046G , 12308G, 12372A, 12597G , 14167T, 14766T, 14798C, 15326G, 16224C, 16311C, 16519C
26	T2b2b	73G, 263G, 709A, 750G, 930A, 1438G, 1809C , 1888A, 2706G, 3338C , 4011T (5,78%) , 4216C, 4769G, 4879G , 4917G, 5147A, 7028T, 8697A, 8860G, 10463C, 11242G, 11251G, 11719A, 11812G, 13368A, 14233G, 14428T , 14766T, 14905A, 15326G, 15452A, 15607G, 15928A, 16126C, 16294T, 16296T, 16311C , 16519C	73G, 263G, 709A, 750G, 930A, 1438G, 1809C , 1888A, 2706G, 3338C , 4011T (18,39%) , 4216C, 4769G, 4879G , 4917G, 5147A, 7028T, 8697A, 8860G, 10463C, 11242G, 11251G, 11719A, 11812G, 13368A, 14233G, 14428T , 14766T, 14905A, 15326G, 15452A, 15607G, 15928A, 16126C, 16294T, 16296T, 16311C , 16519C

Tabla 28. Continuación.

Muestra	Haplogrupo	Haplotipo tumoral	Haplotipo sanguíneo
27	U6a3a1	73G, 263G, 750G, 1438G, 1767R (2,44%) , 2706G, 3348G, 4769G, 4820A, 7028T, 7805A, 8860G, 9491G , 11467G, 11719A, 12308G, 12372A, 14179G, 14364A, 14766T, 15326G, 15790T, 16172C, 16183C , 16189C, 16219G, 16278T, 16291T (34,55%)	73G, 263G, 750G, 1438G, 1438G, 1767R (13,49%) , 2706G, 3348G, 4769G, 4820A, 7028T, 7805A, 8860G, 9491G , 11467G, 11719A, 12308G, 12372A, 14179G, 14364A, 14766T, 15326G, 15790T, 16172C, 16183C , 16189C, 16219G, 16278T
30	H58a	215G (27,78%) , 263G, 750G, 1438G, 1693M (9,78%) , 2251R (3,69%) , 4214R (49,19%) , 4769G, 6340T , 8860G, 11908G , 14793R (7,83%) , 15326G, 15466A, 15721C, 16519C	215R (10,64%) , 263G, 750G, 1438G, 4769G, 6340T , 8860G, 11908G , 14793R (30,64%) , 15326G, 15466A, 15721C, 16519C
31	K1a4a1	73G, 146C , 263G, 497T, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4769G, 5054A , 6260A, 7028T, 8860G, 8911C , 9055A, 9698C, 10398G, 10550G, 11299C, 11467G, 11485C, 11719A, 11840T, 12308G, 12372A, 13740C, 14167T, 14766T, 14798C, 15326G, 16093C (3,87%) , 16224C, 16311C, 16519C	73G, 146C , 263G, 497T, 750G, 1189C, 1438G, 1811G, 2706G, 3480G, 4157W (5,34%) , 4769G, 5054A , 6260A, 7028T, 8860G, 8911C , 9055A, 9698C, 10398G, 10550G, 11299C, 11467G, 11485C, 11719A, 11840T, 12308G, 12372A, 13740C, 14167T, 14766T, 14798C, 15326G, 16035R (3,42%) , 16093C (4,38%) , 16224C, 16311C, 16519C
31BIS	L2a1c	73G, 143A, 146C, 152C, 195C, 263G, 739T, 750G, 769A, 1018A, 1438G, 2416C, 2706G, 2789T, 3010A, 3594T, 4104G, 4769G, 6663G, 7028T, 7175C, 7256T, 7274T, 7521A, 7771G, 8206A, 8701G, 8860G, 9221G, 9230C , 9540C, 10115C, 10398G, 10873C, 11719A, 11914A, 11944C, 12351C , 12693G, 12705T, 13590A, 13650T, 13803G, 14566G, 14766T, 15301A, 15326G, 15784C, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C	73G, 143A, 146C, 152C, 195C, 263G, 739T, 750G, 769A, 1018A, 1438G, 2416C, 2706G, 2789T, 3010A, 3594T, 4104G, 4157W (3,65%) , 4769G, 6663G, 7028T, 7175C, 7256T, 7274T, 7521A, 7771G, 8206A, 8701G, 8860G, 9221G, 9230C , 9540C, 10115C, 10398G, 10873C, 11719A, 11914A, 11944C, 12351C , 12693G, 12705T, 13590A, 13650T, 13803G, 14566G, 14766T, 15301A, 15326G, 15784C, 16223T, 16278T, 16294T, 16309G, 16390A, 16519C
33	H1v	146C (49,30%) , 263G, 750G, 1438G, 3010A, 4769G, 8185C , 8860G, 10314T, 15326G, 16519C	263G, 750G, 1438G, 3010A, 4157W (3,21%) , 4769G, 8185C , 8860G, 10314T, 15326G, 16519C
34	H5b1	146C, 195C, 263G, 449C , 456T, 499A , 750G, 1438G, 4769G, 5471A, 8860G, 14497G, 15326G, 16304C	146C, 195C, 263G, 449C , 456T, 499A , 750G, 1438G, 4769G, 5471A, 8860G, 14497G, 15326G, 16304C
35	J1c2r	73G, 185A, 188G, 228A, 263G, 295T, 462T, 489C, 750G, 1438G, 2706G, 3010A, 4157W (5,15%) , 4216C, 4769G, 7028T, 8860G, 10398G, 11002G, 11251G, 11719A, 12612G, 13032G, 13708A, 14325C, 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16186T, 16325C , 16519C	73G, 185A, 188G, 228A, 263G, 295T, 462T, 489C, 750G, 1438G, 2706G, 3010A, 4157W (0,78%) , 4216C, 4769G, 7028T, 8860G, 10398G, 11002G, 11251G, 11719A, 12612G, 13032G, 13708A, 14325C, 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16186T, 16325C , 16519C

Tabla 28. Continuación.

Muestra	Haplogrupo	Haplotipo tumoral	Haplotipo sanguíneo
36	H45b	263G, 750G, 1438G, 4164G, 4769G, 8843C, 8860G, 12130C, 14443T, 15326G, 16519C	263G, 750G, 1438G, 4164G, 4769G, 8843C, 8860G, 12130C, 14443T, 15326G, 16519C
41	C	10C, 73G, 146C, 195C, 263G, 489C, 750G, 1243C, 1438G, 1693M (26,01%) , 1888A, 2449R (5,46%) , 2706G, 3552A, 4715G, 4769G, 5054A, 6531T (28,67%) , 7028T, 7196A, 7368Y (7,92%) , 8188G, 8584A, 8701G, 8860G, 9540C, 9545G, 10398G, 10400T, 10873C, 11719A, 11914A, 12372A, 12705T, 13263G, 14318C, 14766T, 14783C, 15043A, 15301A, 15326G, 15487T, 15930A, 16051G , 16223T, 16298C, 16325C , 16327T	10C, 73G, 146C, 195C, 263G, 489C, 750G, 1243C, 1438G, 1888A, 2706G, 3552A, 4715G, 4769G, 5054A, 6531T, 7028T, 7196A, 8188G, 8584A, 8701G, 8860G, 9540C, 9545G, 10398G, 10400T, 10873C, 11719A, 11914A, 12372A, 12705T, 13263G, 14318C, 14766T, 14783C, 15043A, 15301A, 15326G, 15487T, 15930A, 16051G , 16223T, 16298C, 16325C , 16327T
43	T2c1d1	73G, 146C, 152C, 195C, 263G, 279C, 531Y (9,19%) , 709A, 750G, 1438G, 1888A, 2706G, 3958R (3,24%) , 4216C, 4769G, 4917G, 5187T, 6261A, 7028T, 7873T, 8697A, 8860G, 10463C, 10586A , 10822T, 11251G, 11719A, 11812G, 11914A, 13368A, 14233G, 14766T, 14882R (10,84%) , 14905A, 15326G, 15452A, 15607G, 15928A, 16126C, 16220G , 16292T, 16294T, 16519C	73G, 146C, 152C, 195C, 263G, 279C, 709A, 750G, 1438G, 1888A, 2706G, 4216C, 4769G, 4917G, 5187T, 6261A, 7028T, 7873T, 8697A, 8860G, 10463C, 10586A , 10822T, 11251G, 11719A, 11812G, 11914A, 13368A, 14233G, 14766T, 14882G (46,80%) , 14905A, 15326G, 15452A, 15607G, 15928A, 16126C, 16220G , 16292T, 16294T, 16519C
44	V	72C, 263G, 750G, 1438G, 2416C , 2706G, 4580A, 4769G, 7028T, 8860G, 12172G, 14502C , 15326G, 15424T , 15904T, 16092C , 16298C	72C, 263G, 750G, 1438G, 2416C , 2706G, 4580A, 4769G, 7028T, 8860G, 12172G, 14502C , 15326G, 15424T , 15904T, 16092C (3,33%) , 16293R (5,27%) , 16298C
45	J1c1b1	73G, 185A, 188G , 228A, 263G, 295T, 462T, 482C, 489C, 750G, 1438G, 2706G, 3010A, 3394C, 4216C, 4769G, 5773A, 7028T, 7184G, 8760C , 8860G, 10398G, 11251G, 11719A, 12192A , 12612G, 13708A, 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16201T, 16362C	73G, 185A, 188G , 228A, 263G, 295T, 462T, 482C, 489C, 750G, 1438G, 2706G, 3010A, 3394C, 4216C, 4769G, 5773A, 7028T, 7184G, 8760C , 8860G, 10398G, 11251G, 11719A, 12192A , 12612G, 13708A, 14766T, 14798C, 15326G, 15452A, 16069T, 16126C, 16201T, 16362C
50	J2a1a1a2	73G, 150T, 152C, 195C, 215G, 263G, 295T, 319C, 489C, 513A, 750G, 1438G, 1850C, 2706G, 3447G, 4157W (0,83%) , 4216C, 4769G, 7028T, 7476T, 7789A, 8860G, 10398G, 10499G, 11251G, 11377A, 11719A, 12612G, 13708A, 13722G, 14133G, 14766T, 15257A, 15326G, 15452A, 16069T, 16126C, 16145A, 16231C, 16261T	73G, 150T, 152C, 195C, 215G, 263G, 295T, 319C, 489C, 513A, 750G, 1438G, 1850C, 2706G, 3447G, 4157W (4,85%) , 4216C, 4769G, 7028T, 7476T, 7789A, 8860G, 10398G, 10499G, 11251G, 11377A, 11719A, 12612G, 13708A, 13722G, 14133G, 14766T, 15257A, 15326G, 15452A, 16069T, 16126C, 16145A, 16231C, 16261T

Tabla 28. Continuación.

Muestra	Haplogrupo	Haplotipo tumoral	Haplotipo sanguíneo
51	T2b	73G, 263G, 709A, 750G, 930A (4,02%) , 1438G, 1688C , 1888A, 2706G, 4216C, 4769G, 4917G, 5147A, 7028T, 8697A, 8860G, 10463C, 11251G, 11719A, 11812G, 11950G , 13368A, 14233G, 14766T, 14791T (29,89%) , 14905A, 15326G, 15452A, 15607G, 15871G , 15928A, 16093C (9,96%) , 16126C, 16294T, 16296T, 16304C, 16519C	73G, 263G, 709A, 750G, 930A, 1438G, 1688C , 1888A, 2706G, 4157W (5,16%) , 4216C, 4769G, 4917G, 5147A, 7028T, 8697A, 8860G, 10463C, 11251G, 11719A, 11812G, 11950G , 13368A, 14233G, 14766T, 14791Y (25,79%) , 14905A, 15326G, 15452A, 15607G, 15871G , 15928A, 16093C , 16126C, 16294T, 16296T, 16304C, 16519C
53	J1b1a1	73G, 152C , 242T, 263G, 295T, 462T, 489C, 750G, 1438G, 2158C, 2706G, 3010A, 4216C, 4769G, 5460A, 7028T, 8269A, 8557A, 8860G, 9031T (36,32%) , 10398G, 11251G, 11719A, 12007A, 12612G, 12746C (14,47%) , 13708A, 13879C, 14766T, 15326G, 15452A, 16069T, 16126C, 16145A, 16172C, 16222T, 16261T, 16316G	73G, 152C , 242T, 263G, 295T, 462T, 489C, 750G, 1438G, 2158C, 2706G, 3010A, 4216C, 4769G, 5460A, 7028T, 8269A, 8557A, 8860G, 10398G, 11251G, 11719A, 12007A, 12612G, 13708A, 13879C, 14766T, 15326G, 15452A, 16069T, 16126C, 16145A, 16172C, 16222T, 16261T, 16316G
54	U5b1b1d	73G, 150T, 263G, 723G , 750G, 1438G, 2706G, 3197C, 4769G, 5656G, 5964C , 7028T, 7385G, 7768G, 8736W (47,16%) , 8860G, 9477A, 10927C, 11467G, 11719A, 12308G, 12372A, 12618A, 13617C, 14182C, 14766T, 15326G, 15884A, 16183C , 16189C, 16270T	73G, 150T, 263G, 723G, 750G, 1438G, 2706G, 3197C, 4769G, 5656G, 5964C , 7028T, 7385G, 7768G, 8736W (30,89%) , 8860G, 9477A, 10927C, 11467G, 11719A, 12308G, 12372A, 12618A, 13617C, 14182C, 14766T, 15326G, 15884A, 16183C , 16189C, 16270T
55	W6a	73G, 143A , 189G, 194T, 195C, 204C, 207A, 263G, 709A, 750G, 1243C, 1438G, 2706G, 3505G, 4093G, 4769G, 5046A, 5460A, 7028T, 7131R (45,96%) , 8251A, 8610C, 8614C, 8860G, 8994A, 11674T, 11719A, 11947G, 12414C, 12705T, 14766T, 15326G, 15884C, 16192T, 16223T, 16292T, 16325C, 16519C	73G, 143A , 189G, 194T (17,43%) , 195C, 204C, 207A, 263G, 709A, 750G, 1243C, 1438G, 2706G, 3505G, 4093G, 4769G, 5046A, 5460A, 7028T, 7131R (4,15%) , 8251A, 8610C, 8614C, 8860G, 8994A, 11674T, 11719A, 11947G, 12414C, 12705T, 14766T, 15326G, 15884C, 16192T, 16223T, 16292T (3,62%) , 16325C, 16519C
56	H15	55C (6,57%) , 57C, 189R (3,14%) , 263G, 750G, 1438G, 3010A , 4769G, 8860G, 15326G, 16189C , 16519C	55C, 57C, 263G, 750G, 1438G, 3010A , 4157W (3,51%) , 4769G, 4897R (6,4%) , 8860G, 15326G, 16189C , 16519C
59	H1r1	146C, 263G, 750G, 1167R (3,41%) , 1438G, 3010A, 4769G, 8554R (9,39%) , 8860G, 9356T, 11585W (5,75%) , 13015K (3,10%) , 15326G, 16278T, 16519C	146C, 263G, 750G, 1438G, 3010A, 4769G, 8860G, 9356T, 15326G, 16278T, 16519C

En las muestras tumorales de los 32 pacientes analizados, fueron detectadas 923 mutaciones en 323 posiciones mitocondriales diferentes. De estas 923 mutaciones, 877 son mutaciones en homoplasma localizadas en 281 posiciones diferentes, mientras que las 46 restantes corresponden a mutaciones en heteroplasma localizadas en 42 posiciones diferentes.

En relación a las mutaciones en homoplasma observadas en los tumores, el 100% son de origen germinal, es decir, detectadas tanto en el tejido tumoral como sanguíneo. En relación a las mutaciones en heteroplasma, 27 son heteroplasmas de origen somático (58,70%), lo que significa que se detectan únicamente en el tejido tumoral, mientras que las 19 restantes son de origen germinal (41,30%) (Figura 23).

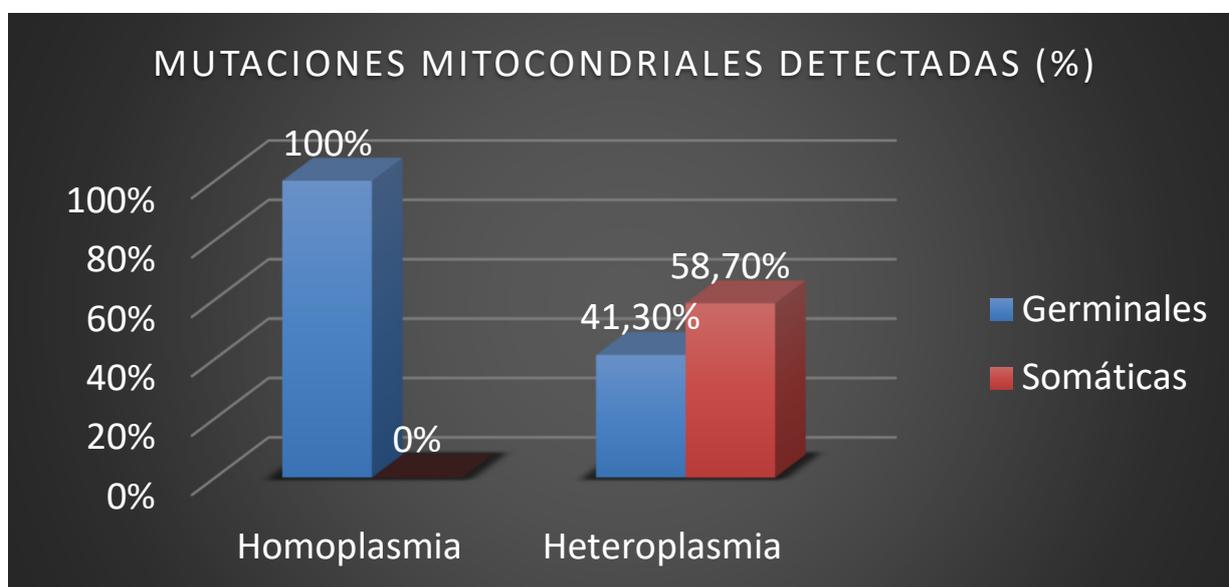


Figura 23. Porcentaje de las mutaciones mitocondriales detectadas en las muestras tumorales de los 32 pacientes del estudio, según el tipo de mutación (homoplasma-heteroplasma) y el origen de la mutación (germinal-somática).

Comparando las mutaciones detectadas teniendo en cuenta su localización en el genoma mitocondrial con las mutaciones detectadas en la población general, se detecta una sobrerrepresentación de mutaciones en los RNAr, concretamente en el RNAr 12S ($Z= 11,525$; $p<0,0001$) y en el complejo V ($Z= 4,519$; $p<0,0001$), así como una subrepresentación de mutaciones en la HVRI ($Z= 3,443$; $p<0,0003$), y en el complejo IV ($Z= 3,205$; $p<0,0007$) respecto a la población general (Tabla 29).

Tabla 29. Mutaciones detectadas en el estudio y en la población general –datos de genomas completos obtenidos de HmtDB (Clima et al. 2017)–, teniendo en cuenta la región mitocondrial.

Región mitocondrial	Presente estudio N= 32 individuos		Población general N= 7605 individuos		Test estadístico Z	Probabilidad
	Nº mutaciones N=923	Frecuencia (%)	Nº mutaciones N=130440	Frecuencia (%)		
Región control	266	28,82	42167	32,33	2,271	0,0116
HVRI	93	10,08	18289	14,02	3,443	0,0003*
HVRII	121	13,11	14944	11,46	1,57	0,0582
HVRIII	25	2,71	2667	2,04	1,419	0,078
Región Codificante	657	71,18	88273	67,67	2,271	0,0116
RNAs	169	18,31	19774	15,16	2,658	0,0039
RNAr	145	15,71	14096	10,81	4,774	<0,0001*
12S	84	9,1	3635	2,79	11,525	<0,0001*
16S	61	6,61	10461	8,02	1,574	0,0578
RNAt	24	2,6	5678	4,35	2,604	0,0046
Subunidades de la cadena respiratoria	487	52,76	68499	52,51	0,151	0,44
Complejo I	263	28,49	39280	30,11	1,069	0,1426
Complejo III	102	11,05	11292	8,66	2,575	0,005
Complejo IV	66	7,15	13535	10,38	3,205	0,0007*
Complejo V	56	6,07	4392	3,37	4,519	<0,0001*

*Regiones mitocondriales donde se detecta una sobrerrepresentación o una subrepresentación de mutaciones en el presente estudio respecto a la población general, considerando una corrección de Bonferroni de $p < 0,0033$.

6.1.2. Distribución de los haplogrupos mitocondriales de las muestras del estudio

En las tablas 30 y 31 se observa la distribución a nivel de haplogrupo de los pacientes afectados de glioma (Tabla 30) y según el grado del tumor en el caso de los haplogrupos detectados con una frecuencia superior a 1: H, J, T y U (Tabla 31). Asimismo, a título comparativo, se presentan las frecuencias observadas de los haplogrupos en la población general de la Costa Mediterránea española (CCAA: Cataluña, Comunidad Valenciana y Murcia) (Barral-Arca et al. 2016).

Comparando la distribución de los haplogrupos encontrados en el presente estudio respecto a los hallados en la población general de la Costa Mediterránea, se observa una sobrerrepresentación significativa del haplogrupo J en los pacientes analizados (Test exacto de Fisher: $p = 0,0021$), ya que un 25% de los pacientes pertenece al haplogrupo J frente a un 7% de los individuos de la población general. En este sentido, nuestros resultados indican que los individuos pertenecientes al haplogrupo J tienen 4,43 veces más riesgo de sufrir glioma que el resto de individuos con otro haplogrupo (OR: 4,43; 95% IC 1,92-10,21) (Tabla 31). En relación al resto de haplogrupos, no se observan diferencias significativas entre los pacientes y la población general.

Considerando la distribución de haplogrupos según el grado de los tumores, la frecuencia de haplogrupo H presenta una tendencia inversa respecto a la malignidad: un 62,5% de los tumores de grado II, un 42,86% de los de grado III y un 11,76% de los de grado IV pertenecen a individuos con haplogrupo H. (Tabla 31). Además, se observa una sobrerrepresentación de tumores de grado II respecto a los de grado IV (Test exacto de Fisher: $p=0,017$), ya que un 62,5% de los pacientes con haplogrupo H tienen un tumor de grado II en comparación con el 11,76% de los pacientes con haplogrupo H con tumor de grado IV. Así, los pacientes con haplogrupo H tienen 12,5 veces más riesgo de sufrir un tumor de grado II que un tumor de grado IV (OR: 12,5; 95% IC 1,6-97,65).

Por último, si se comparan las frecuencias de los diferentes grados tumorales del haplogrupo H respecto a la población general, se detecta una subrepresentación significativa de tumores de grado IV en pacientes con haplogrupo H, ya que un 11,76% de los pacientes con haplogrupo H tienen un tumor de grado IV frente a un 37% de los individuos de la población general (Test exacto de Fisher: $p=0,04$) (Tabla 31). Los resultados indican, por lo tanto, que los pacientes de haplogrupo H tienen 4,4 veces menos riesgo de sufrir un tumor de grado IV respecto a la población general (OR: 4,4; 95% IC 1-19,33). Del mismo modo, también se detecta una sobrerrepresentación significativa de tumores de grado IV en pacientes con haplogrupo J, ya que un 29,41% de los pacientes con haplogrupo J tienen un tumor de grado IV frente a un 7% de los individuos de la población general (Test exacto de Fisher: $p=0,0058$) (Tabla 31).

Tabla 30. Comparación de la frecuencia de haplogrupos entre el presente estudio y la población general de la Costa Mediterránea española. N= tamaño muestral.

Haplogrupos	Glioma (Presente estudio) N= 32		Población general (Barral-Arca y col. 2016) N= 1058	
	N	Frecuencia (%) (95% IC)	N	Frecuencia (%) (95% IC)
C	1	3,13 (0,7-15,8)	10	1 (0,5-1,7)
H	10	31,25 (18,0-48,7)	391	37 (34,1-39,9)
J	8	25 (13,3-42,3)	74	7 (5,6-8,7)
L	1	3,13 (0,7-15,8)	21	2 (1,3-3)
T	4	12,5 (5,1-28,2)	95	9 (7,4-10,9)
U	6	18,75 (9-35,5)	232	22 (19,5-24,5)
V	1	3,13 (0,7-15,8)	21	2 (1,3-3)
W	1	3,13 (0,7-15,8)	10	1 (0,5-1,7)

Tabla 31. Comparación de la frecuencia de haplogrupos de poblaciones europeas (%) entre el presente estudio y la población general de la Costa Mediterránea española a nivel general y según el grado de malignidad tumoral. N= tamaño muestral.

Haplogrupos	Glioma (Presente estudio)												Población general (Barral-Arca et al. 2016)	
	Total N= 32			Grado de malignidad tumoral II N= 8			Grado de malignidad tumoral III N= 7			Grado de malignidad tumoral IV N= 17				Total N= 1058
	N	Frecuencia (%)	Test exacto de Fisher	N	Frecuencia (%)	Test exacto de Fisher	N	Frecuencia (%)	Test exacto de Fisher	N	Frecuencia (%)	Test exacto de Fisher		N
H	10	31,25	0,5802 (0,36-1,65)	5	62,5	0,1558 (0,68-11,96)	3	42,86	0,714 (0,28-5,75)	2	11,76	0,04* (1-19,33)	391	37
J	8	25	0,0021* (1,92-10,21)	2	25	0,1062 (0,88-22,35)	1	14,29	0,4011 (0,26-18,65)	5	29,41	0,0058* (1,9-16,15)	74	7
T	4	12,5	0,525 (0,5-4,22)	0	0		2	28,57	0,1275 (0,78-21,18)	2	11,76	0,6609 (0,3-6)	95	9
U	6	18,75	0,829 (0,33-2,02)	1	12,5	1 (0,24-16,06)	0	0		5	29,41	0,5533 (0,52-4,25)	232	22

*El test exacto de Fisher muestra diferencias significativas ($p < 0,05$).

Recopilación de la distribución de haplogrupos mitocondriales en pacientes afectados de glioma

En la tabla 32 se presentan las frecuencias de haplogrupos en los pacientes afectados de glioma de las diferentes poblaciones europeas y en la población general respectiva. Se observa una frecuencia superior de individuos con haplogrupo J y T en los estudios de glioma en comparación a las poblaciones generales respectivas, así como una frecuencia inferior de individuos con haplogrupo U. Los resultados estadísticos muestran que los individuos con haplogrupo J y T tienen una mayor disposición de sufrir glioma ($p < 0,001$ y $p = 0,007$, respectivamente), siendo de hasta 2,41 veces superior en el caso del haplogrupo J (OR: 2,407; 95% IC 1,538-3,764) y 1,82 veces superior en el caso del haplogrupo T (OR: 1,817; 95% IC 1,176-2,806) (Tabla 33).

Tabla 32. Comparación de las frecuencias (%) de haplogrupos en los pacientes afectados de glioma de las diferentes poblaciones europeas y en la población general respectiva.

Haplogrupos	CATALUÑA				GALICIA				NORTE de ITALIA			
	Glioma (Presente estudio) N=32		Población general de la Costa Mediterránea (Barral-Arca et al. 2016) N=1058		Glioma (Vega et al. 2004) N=53		Población general de la Costa Cantábrica (Barral-Arca et al. 2016) N=1520		Glioma (Gasparre et al. 2007; Vidone et al. 2015) N=120		Población general del Norte de Italia (Brisighelli et al. 2012) N=702	
	N	Frecuencia (%)	N	Frecuencia (%)	N	Frecuencia (%)	N	Frecuencia (%)	N	Frecuencia (%)	N	Frecuencia (%)
H	10	31,25	391	36,96	27	50,94	790	51,97	59	49,17	386	54,99
J	8	25	74	6,99	8	15,09	121	7,96	10	8,33	28	3,99
T	4	12,5	95	8,98	5	9,43	121	7,96	18	15	49	6,98
U	6	18,75	232	21,93	8	15,09	288	18,95	15	12,5	140	19,94

Tabla 33. Resultados de la regresión logística para cada uno de los haplogrupos. Para cada uno de los análisis se tuvo en cuenta el haplogrupo, la presencia o ausencia de glioma, la población (Cataluña, Galicia, Norte de Italia) y la frecuencia detectada.

	Haplogrupo H	Haplogrupo J	Haplogrupo T	Haplogrupo U
Efecto	Sin efecto	Predisposición	Predisposición	Sin efecto
P valor	0,147	<0,001*	0,007*	0,059
Odds ratio (IC 95%)	1,238 (0,928-1,653)	2,407 (1,538-3,764)	1,817 (1,176-2,806)	1,480 (0,985-2,225)

*Los resultados ponen en evidencia que los haplogrupos J y T tienen un efecto ante la enfermedad ($p < 0,05$).

6.1.3. Análisis de mutaciones en homoplasma

Distribución de las mutaciones en homoplasma a lo largo de genoma mitocondrial

Se han detectado un total de 877 mutaciones en homoplasma. En la tabla 34 se representa su distribución en las distintas regiones del genoma mitocondrial, así como la ratio de mutaciones detectada en cada región.

En cuanto a la comparación de la ratio de mutaciones por par de bases en las diferentes regiones del genoma mitocondrial, se detecta una mayor ratio de mutaciones en la región control en comparación con la región codificante y en HVRII respecto a HVRI. También se observa una mayor ratio de mutaciones en la subunidad 12S de los RNAr respecto a los RNAs, así como una menor ratio en los RNAt respecto a los RNAs. Por último, se observa tanto una mayor ratio de mutaciones en los complejos III (CYB) y V (ATP) como una menor ratio de mutaciones en el complejo IV (CO) respecto al total de subunidades de la cadena respiratoria (Tablas 34 y 35).

Tabla 34. Distribución de las mutaciones en homoplasma detectadas en el estudio. Cálculo de la ratio de mutaciones por par de bases (mutaciones/pb región mitocondrial). pb: pares de bases.

Región mitocondrial	Nº mutaciones en homoplasma N=877	Tamaño de la región (pb)	Mutaciones/pb
Región control	252	1180	0,214
HVRI	88	360	0,244
HVRII	115	316	0,364
HVRIII	24	137	0,175
Región Codificante	624	15389	0,041
Nucleótidos no codificantes	1	58	0,017
RNAs	159	4021	0,04
RNAr	135	2513	0,054
12S	80	954	0,084
16S	55	1559	0,035
RNAt	24	1508	0,016
Subunidades de la cadena respiratoria	465	11395	0,041
Complejo I	251	6356	0,039
Complejo III	99	1141	0,087
Complejo IV	62	3010	0,021
Complejo V	53	888	0,06

Tabla 35. Comparación entre regiones mitocondriales de la ratio de mutaciones encontradas en cada una de ellas. La ratio de mutaciones en cada región se obtiene de la relación entre el número de mutaciones detectadas y el tamaño (pb) de la región. Únicamente se muestran aquellas comparaciones en que el resultado es significativo, considerando una corrección Bonferroni de $p < 0,0042$.

Comparación entre regiones mitocondriales		Test estadístico Z	Probabilidad
Región Control*	Región codificante	25,595	$\leq 0,0001$
HVRII*	HVRI	3,633	0,0001
HVRII*	HVRIII	4,001	$\leq 0,0001$
RNAr 12S*	RNAs	5,754	$\leq 0,0001$
RNAs*	RNAr	4,374	$\leq 0,0001$
Complejo III*	Complejos I, III, IV y V	7,14	$\leq 0,0001$
Complejos I, III, IV y V*	Complejo IV	5,253	$\leq 0,0001$
Complejo V*	Complejos I, III, IV y V	2,696	0,0035

*La proporción de mutaciones es significativamente mayor en estas regiones mitocondriales.

Análisis de la carga mutacional

Los resultados obtenidos indican que la mayoría de las variantes encontradas en pacientes con glioma son variantes comunes (variantes con una frecuencia en la población general $>1\%$) con valores de patogenicidad bajos (valores de MutPred2 $< 0,5$ y MitoTIP $< 12,66$: valor actualizado en julio 2019) (Tablas 1 y 2 del Anexo III). Respecto al cálculo de la carga mutacional ajustada de las variantes localizadas en genes que codifican para proteínas de la cadena respiratoria, tan solo 4 individuos (12,5%) presentan una carga mutacional que supera el valor de 0,5. Estos individuos presentan variabilidad respecto al grado de malignidad tumoral: 2 de ellos (individuos 5 y 19) presentan tumores más agresivos (grado IV), mientras que los otros 2 (individuos 36 y 45) presentan tumores menos agresivos (grados II y III) (Tabla 36; Tabla 3 del Anexo III). La carga mutacional de estos individuos viene dada por mutaciones únicas localizadas en el gen *ND1* en 2 casos (en la posición 3721 en el individuo 5 y en la posición 3394 en el individuo 45) y en el gen *ATP6* en los otros 2 casos (posición 9145 en el individuo 19 y posición 8843 en el individuo 36). Además, estas variantes que podrían tener un efecto probablemente deletéreo tienen una frecuencia $<1\%$ en la población general (Tabla 36). Respecto a la carga mutacional ajustada de las variantes localizadas en regiones que codifican para RNAr, ningún individuo presenta una carga mutacional que supere el valor de 12,66, ni tampoco se detecta ninguna variante en concreto que supere dicho valor para que pueda considerarse, como mínimo, posiblemente patogénica (Tabla 4 del Anexo III).

Tabla 36. Individuos con una carga mutacional ajustada >0,5. La carga mutacional ajustada se obtiene dividiendo la carga mutacional total (en este caso, la suma de todas las variantes que tienen una carga mutacional >0,5) entre el número de variantes.

Individuo	Edad	Grado de malignidad tumoral	Haplogrupo	Variantes con valor de carga mutacional ajustada >0,5			Frecuencias (%) de las mutaciones en la población general N=7605
				Carga Mutacional	Variantes	Carga mutacional ajustada	
5	71	IV	J1c5c1	0,62	3721 (ND1)	0,62	0,08
19	43	IV	U5a2c	0,563	9145 (ATP6)	0,563	0,13
36	31	III	H45b	0,592	8843 (ATP6)	0,592	0,42
45	35	II	J1c1b1	0,651	3394 (ND1)	0,651	0,97

Aplicando el mismo principio, pero comparando los tumores de menor y mayor grado de malignidad entre sí (grados II y III vs. grado IV), tampoco se detecta una mayor carga mutacional ajustada en individuos con un grado tumoral más severo, ni a nivel de variantes localizadas en genes que codifican proteínas de los complejos de la cadena respiratoria ni en RNAt (Tabla 37; Tablas 3 y 4 del Anexo III).

Tabla 37. Cálculo de estadísticos descriptivos (media, mediana y desviación típica) de la carga mutacional detectada según el grado de malignidad tumoral (grado II y III vs. grado IV), tanto en genes que codifican proteínas de los complejos de la cadena respiratoria y los RNAt. Comparación de la carga mutacional ajustada detectada entre los grados de malignidad tumoral en ambas regiones mitocondriales mediante el test Mann-Whitney.

Región mtDNA	Grado de malignidad tumoral	Estadísticos descriptivos			Test Mann-Whitney
		Media	Mediana	Desviación típica	
Genes que codifican proteínas de los complejos de la cadena respiratoria	II y III	0,188	0,193	0,044	Carga= 109,0 Z= -0,704 p= 0,481
	IV	0,178	0,179	0,04	
RNAt	II y III	3,271	0,016	4,529	Carga= 111,5 Z= -0,634 p= 0,526
	IV	4,278	0,706	5,305	

*El test Mann-Whitney no muestra diferencias significativas ($p > 0,05$).

Mutaciones en homoplasmia sobrerrepresentadas en pacientes afectados de glioma respecto a la población general

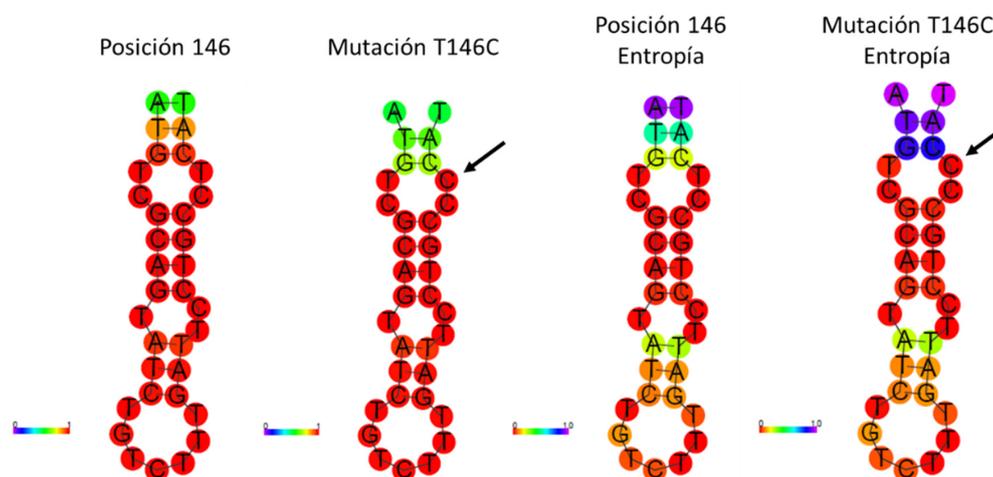
Del total de mutaciones detectadas en homoplasmia (N= 877), se detectan 122 mutaciones (13,91%) en 33 posiciones mitocondriales que están sobrerrepresentadas en el presente estudio respecto a la población general (Test exacto de Fisher: $p < 0,01$) (Tabla 5 del Anexo III). De éstas 122 mutaciones, 87 (71,31%) localizadas en 10 posiciones son definitorias de haplogrupo, la mayoría del haplogrupo J (68 de 87; 78,16%) (Tabla 38). Las 35 mutaciones restantes (28,69%) localizadas en 23 posiciones no son definitorias de haplogrupo (Tabla 39). Se observa que la mayoría se detectan únicamente en

1 o 2 individuos en el presente estudio y únicamente las mutaciones T146C [locus: región control (HVRII) / sitio de unión a la membrana (ATT) origen de replicación (OH) / 7S DNA] y la T16325C [locus: región control (HVRI) / ATT / 7S DNA] se detectan en 8 y 3 individuos, respectivamente. Nuestros hallazgos indican que con la mutación T146C hay un riesgo de 3,31 veces más de sufrir glioma (Test exacto de Fisher: $p=0,0071$; OR: 3,31; 95% IC 1,48-7,39), mientras que con la mutación T16325C lo hay de 12,79 veces más (Test exacto de Fisher: $p=0,0023$; OR: 12,79; 95% IC 3,8-43,12). La predicción de las estructuras secundarias muestra que la mutación T146C da lugar a la pérdida de un enlace y cambios a nivel de entropía, mientras que la mutación T16325C provoca un cambio importante a nivel conformacional y de entropía (Figura 24).

a) Mutación T146C: cambios a nivel estructural y de entropía.

- Energía libre en la región HVRII (116-149): $-8,14$ kcal/mol.

- Energía libre en la región HVRII (116-149) con la mutación T146C: $-6,92$ kcal/mol.b)



b) Mutación T16325C: cambios a nivel estructural y de entropía.

- Energía libre en la región HVRI (16310-16332): $-1,71$ kcal/mol.

- Energía libre en la región HVRI (16310-16332) con la mutación T16325C: $-0,70$ kcal/mol.

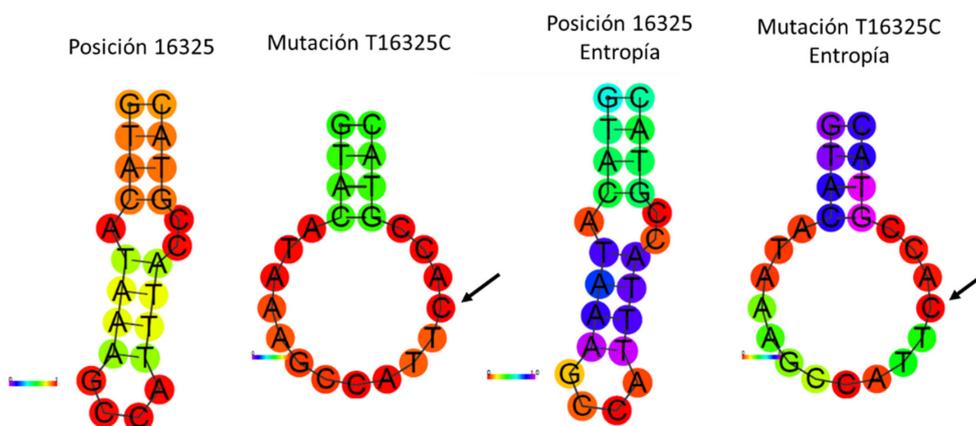


Figura 24. a) Cambios a nivel estructural y de entropía en la región control (HVRII) que comprende las posiciones 116 a 149 cuando se produce la mutación en la posición 146; b) Cambios a nivel estructural y de entropía en la región control (HVRI) que comprende las posiciones 16310 a 16332 cuando se produce la mutación en la posición 16325.

Tabla 38. Mutaciones en homoplasma sobrerrepresentadas que definen haplogrupo (Test exacto de Fisher <0,01) (N=87). Los datos de la población general se han obtenido de la base de datos HmtDB (Clima et al. 2017). IC: intervalos de confianza. ATT: sitio de unión a la membrana; OH: origen de replicación; TAS2: secuencia asociada a la terminación 2; TFY: sitio de unión del factor de transcripción 1 (TF1).

Posición	rCRS	Mutación	Región mitocondrial	Haplogrupo Individuos	Hits Filogenia	Presente estudio		Base de datos población general europea	
						Individuos N=32	Frecuencia (%) IC (95%)	Individuos N= 7605	Frecuencia (%) IC (95%)
188	A	A>G	HVRII / ATT / OH / 7S DNA	J1c2c, J1c2r, J1c2e2, J1c1b1	6	4	12,5 (5,1-28,2)	189	2,49 (2,2-2,9)
295	C	C>T	HVRII / ATT / OH / TFY	J1c2e2, J1c5c1, J2a1a1, J1c2c, J1c2r, J1c1b1, J2a1a2, J1b1a1	3	8	25 (13,3-42,3)	692	9,1 (8,5-9,8)
489	T	T>C	HVRIII / ATT	J1c2e2, J1c5c1, J2a1a1, J1c2c, J1c2r, J1c1b1, J2a1a2, J1b1a1, C	3	9	28,13 (15,6-45,5)	721	9,48 (8,8-10,2)
4216	T	T>C	ND1	J1c2e2, J1c5c1, J2a1a1, J1c2c, J1c2r, J1c1b1, J2a1a2, J1b1a1, T2b, T2b2b, T2b2b, T2c1d1	13	12	37,5 (22,9-54,9)	1347	17,71 (16,9-18,6)
11242	C	C>G	ND4	T2b2b, T2b2b	1	2	6,25 (1,9-20,2)	8	0,11 (0,1-0,2)
11251	A	A>G	ND4	J1c2e2, J1c5c1, J2a1a1, J1c2c, J1c2r, J1c1b1, J2a1a2, J1b1a1, T2b, T2b2b, T2b2b, T2c1d1	1	12	37,5 (22,9-54,9)	1326	17,44 (16,6-18,3)
12612	A	A>G	ND5	J1c2e2, J1c5c1, J2a1a1, J1c2c, J1c2r, J1c1b1, J2a1a2, J1b1a1	7	8	25 (13,3-42,3)	688	9,05 (8,4-9,7)
15452	C	C>A	CYB	J1c2e2, J1c5c1, J2a1a1, J1c2c, J1c2r, J1c1b1, J2a1a2, J1b1a1, T2b, T2b2b, T2b2b, T2c1d1	2	12	37,5 (22,9-54,9)	1325	17,42 (16,6-18,3)
16069	C	C>T	HVRI / ATT	J1c2e2, J1c5c1, J2a1a1, J1c2c, J1c2r, J1c1b1, J2a1a2, J1b1a1	6	8	25 (13,3-42,3)	693	9,11 (8,5-9,8)
16126	T	T>C	HVRI / ATT / TAS2 / 7S DNA	J1c2e2, J1c5c1, J2a1a1, J1c2c, J1c2r, J1c1b1, J2a1a2, J1b1a1, T2b, T2b2b, T2b2b, T2c1d1	19	12	37,5 (22,9-54,9)	1380	18,14 (17,3-19)

Tabla 39. Mutaciones en homoplasmia sobrerrepresentadas que no definen haplogrupo (Test exacto de Fisher <0,01) (N=35). Los datos de la población general se han obtenido de la base de datos HmtDB (Clima et al. 2017). IC: intervalos de confianza. ATT: sitio de unión a la membrana.

Posición	rCRS	Mutación	Región mitocondrial	Haplogrupo Individuos		Hits Filogenia	Presente estudio		Base de datos población general europea	
				Individuos N=32	Frecuencia (%) IC (95%)		Individuos N=7605	Frecuencia (%) IC (95%)		
146	T	T>C	HVRII / ATT / OH / 7S DNA	K1a4a1, C, L2a1c, H1r1, H1r1, H5b1, T2c1d1, J1c2c	120(2)	8	25 (13,3-42,3)	696	9,15 (8,5-9,8)	
1518	C	C>T	RNAr 12S	K1a1	0	1	3,13 (0,7-15,8)	0	0	
1688	A	A>C	RNAr 16S	T2b	0	1	3,13 (0,7-15,8)	0	0	
2416	T	T>C	RNAr 16S	L2a1c, V	5	2	6,25 (1,9-20,2)	28	0,37 (0,3-0,5)	
3173	G	G>A	RNAr 16S	U5a2c	0	1	3,13 (0,7-15,8)	0	0	
4879	C	C>G	ND2	T2b2b	0	1	3,13 (0,7-15,8)	0	0	
5054	G	G>A	ND2	K1a4a1, C	10	2	6,25 (1,9-20,2)	15	0,2 (0,1-0,3)	
6932	A	A>G	CO1	H89	3	1	3,13 (0,7-15,8)	0	0	
7699	C	C>T	CO2	K1a1	0	1	3,13 (0,7-15,8)	1	0,01 (0-0,1)	
8068	C	C>T	CO2	H89	1	1	3,13 (0,7-15,8)	0	0	
8760	T	T>C	ATP6	J1c1b1	0	1	3,13 (0,7-15,8)	0	0	
8911	T	T>C	ATP6	K1a4a1	2	1	3,13 (0,7-15,8)	1	0,01 (0-0,1)	
9230	T	T>C	CO3	L2a1c	2	1	3,13 (0,7-15,8)	1	0,01 (0-0,1)	
9356	C	C>T	CO3	H1r1	2	2	6,25 (1,9-20,2)	9	0,12 (0,1-0,2)	
10057	T	T>C	RNAr glicina	H89	0	1	3,13 (0,7-15,8)	1	0,01 (0-0,1)	
10314	C	C>T	ND3	H1v	2	1	3,13 (0,7-15,8)	1	0,01 (0-0,1)	
12046	A	A>G	ND4	K1a1	0	1	3,13 (0,7-15,8)	0	0	
12597	A	A>G	ND5	K1a1	0	1	3,13 (0,7-15,8)	1	0,01 (0-0,1)	
14443	C	C>T	ND6	H45b	0	1	3,13 (0,7-15,8)	1	0,01 (0-0,1)	
14524	A	A>G	ND6	U4b1a	1	1	3,13 (0,7-15,8)	0	0	
15424	C	C>T	CYB	V	0	1	3,13 (0,7-15,8)	1	0,01 (0-0,1)	
15871	A	A>G	CYB	T2b	0	1	3,13 (0,7-15,8)	0	0	
16325	T	T>C	HVR I / ATT / 7S DNA	W6a, J1c2r, C	24(1 del)	3	9,38 (3,4-24,3)	61	0,8 (0,6-1)	

Mutaciones en homoplasma detectadas en posiciones estables

En la figura 25 se esquematiza la distribución de las mutaciones en homoplasma en las distintas regiones del genoma mitocondrial, teniendo en cuenta si se localizan en posiciones estables, es decir,

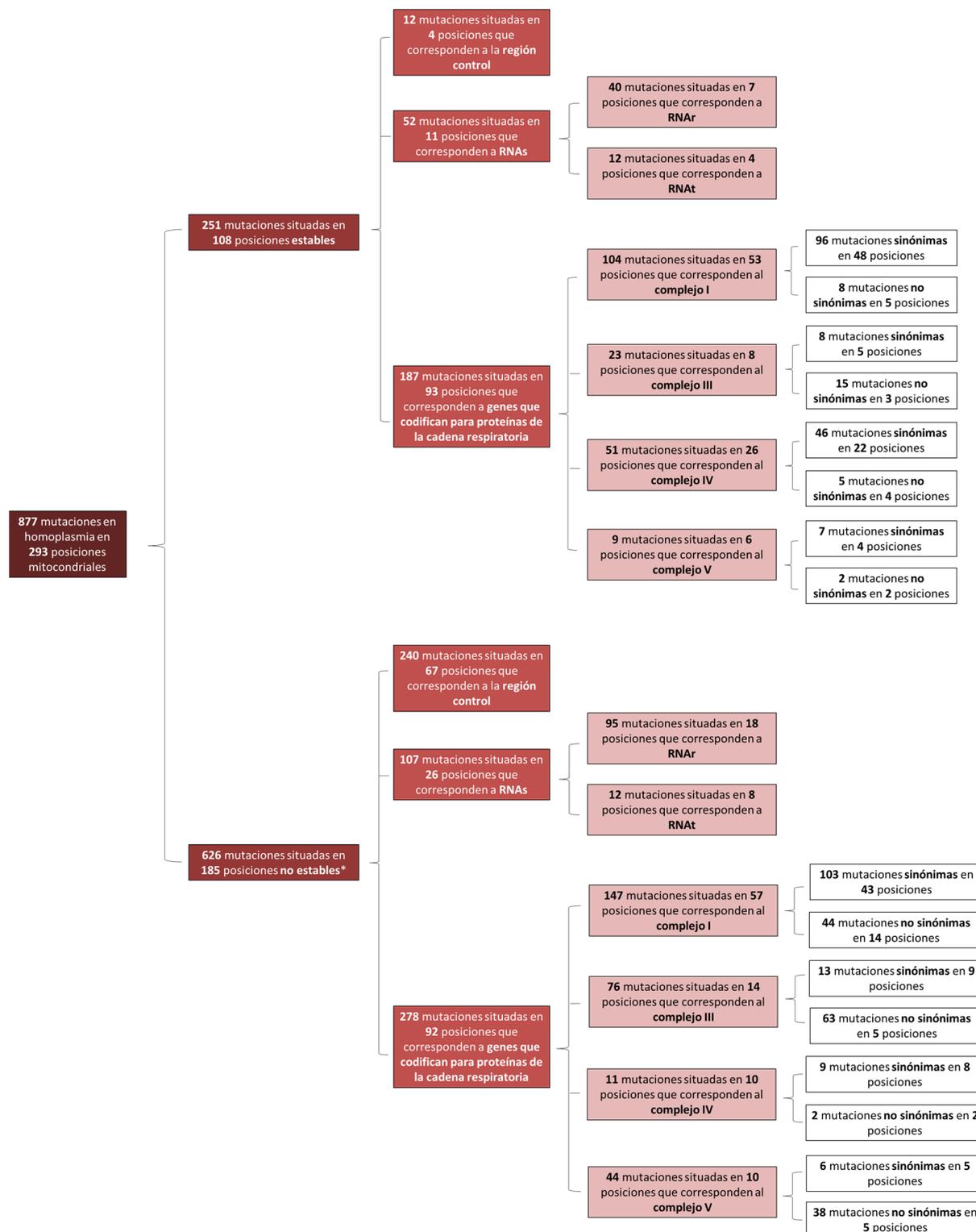


Figura 25. Esquema de las mutaciones en homoplasma detectadas a lo largo del genoma mitocondrial (región control, RNAs y genes que codifican para proteínas de la cadena respiratoria), teniendo en cuenta si se detectan en posiciones estables (posiciones que tienen <6 hits en la filogenia y que tienen un índice de conservación <20% o >80%) o no estables. *De las 626, la mutación A5656G se sitúa en una posición no codificante fuera de la región control. Dicha mutación no se tiene en cuenta en la posterior clasificación.

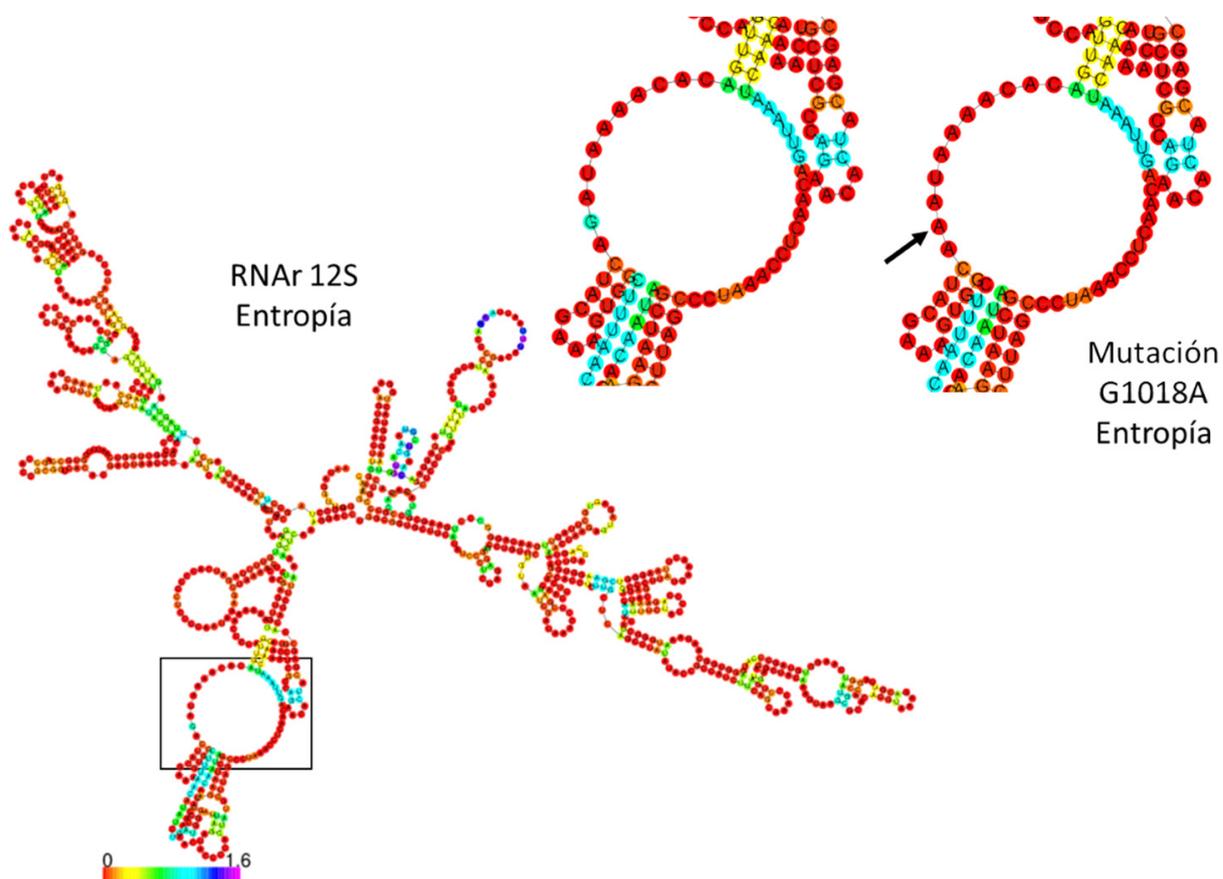
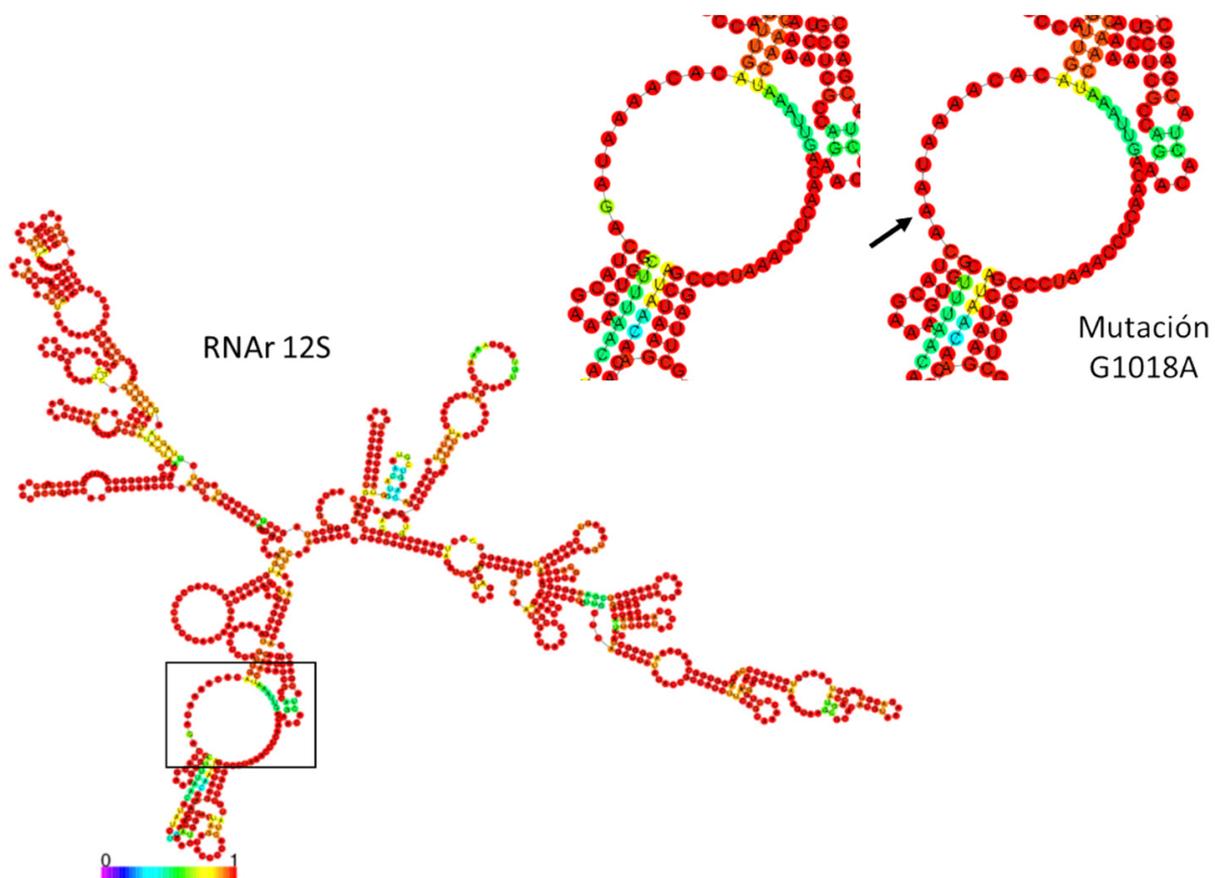
De las 187 mutaciones detectadas en 93 posiciones estables que codifican para proteínas de la cadena respiratoria, 30 (16,04%) son mutaciones no sinónimas, es decir, que dan lugar a cambio a nivel aminoacídico (Figura 25). De estas mutaciones se observó que únicamente 2 [A3721G (*ND1*) y G9145A (*ATP6*)] muestran valores de patogenicidad MutPred2 >0,5, indicando que probablemente tengan un efecto deletéreo.

Para las 52 mutaciones detectadas en 11 posiciones estables que corresponden a RNAs se llevó a cabo la predicción de las estructuras secundarias. Únicamente 4 mutaciones [G1018A (*RNAr 12S*), T1850C (*RNAr 16S*), G5780A (*RNAt* cisteína) y T10057C (*RNAt* glicina)] presentan variación a nivel de entropía (Figura 27) y 3 mutaciones [C739T (*RNAr 12S*) y T1243C (*RNAr 12S*) en 2 individuos] variación a nivel tanto estructural como de entropía (Figura 28).

a) Mutación G1018A (RNAr 12S): cambios a nivel de entropía.

- Energía libre RNAr 12S: -213,14 kcal/mol.

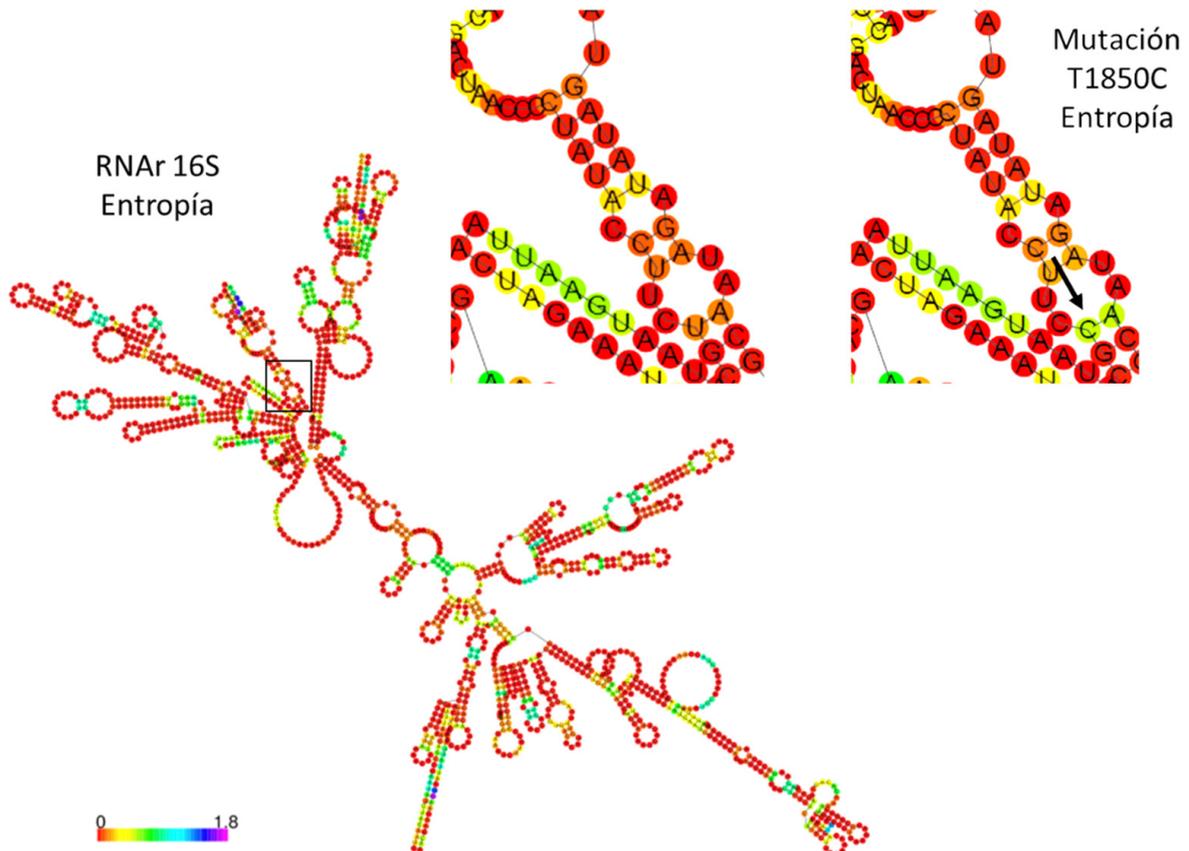
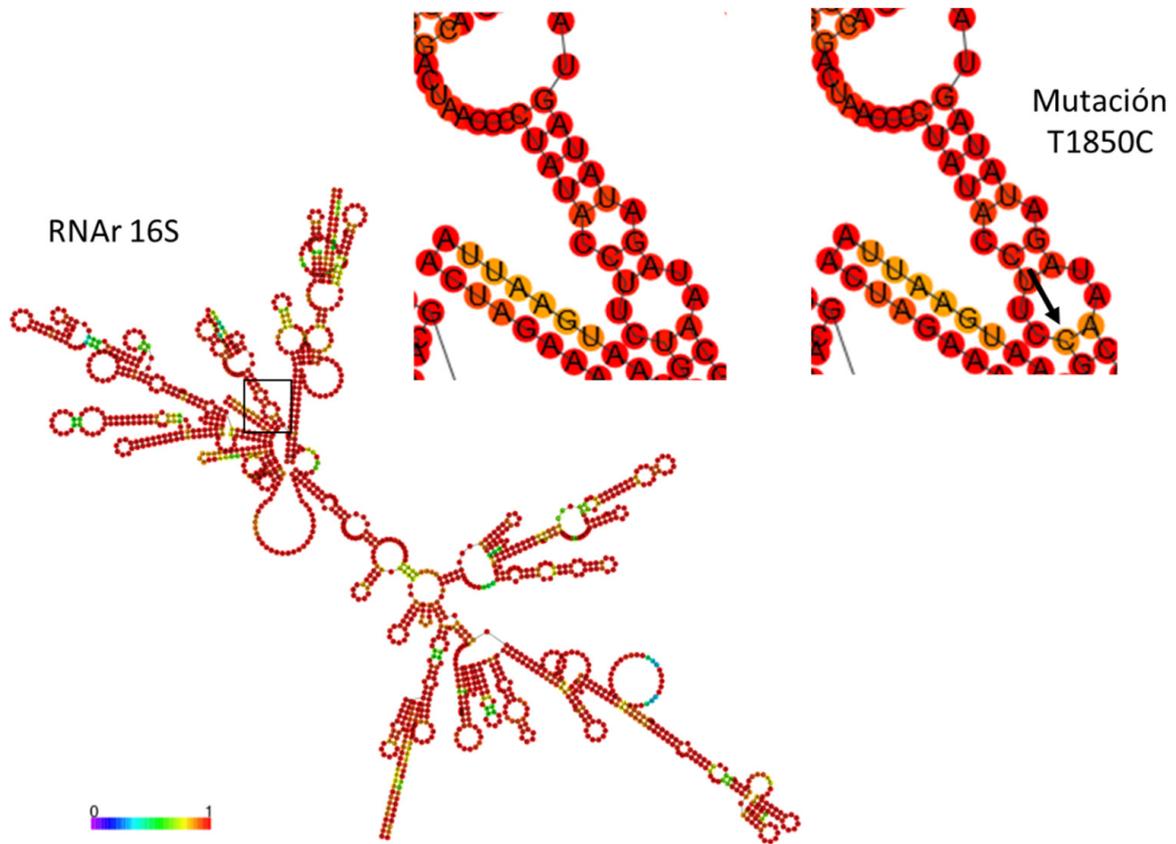
- Energía libre RNAr 12S con la mutación G1018A: -212,91 kcal/mol.



b) Mutación T1850C (RNAr 16S): cambios a nivel de entropía.

- Energía libre RNAr 16S: -336,35 kcal/mol.

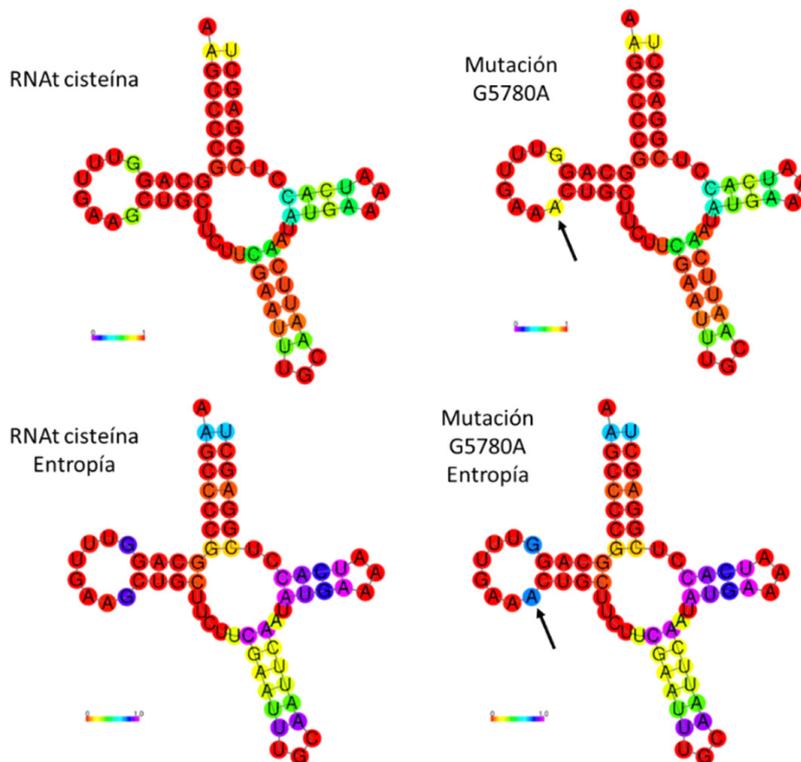
- Energía libre RNAr 16S con la mutación T1850C: -335,32 kcal/mol.



c) Mutación G5780A (RNAt cisteína): cambios a nivel de entropía.

- Energía libre RNAt cisteína: $-16,98$ kcal/mol.

- Energía libre RNAt cisteína con la mutación G5780A: $-17,20$ kcal/mol.



d) Mutación T10057C (RNAt glicina): cambios a nivel de entropía.

- Energía libre RNAt glicina: $-9,12$ kcal/mol.

- Energía libre RNAt glicina con la mutación T10057C: $-8,55$ kcal/mol.

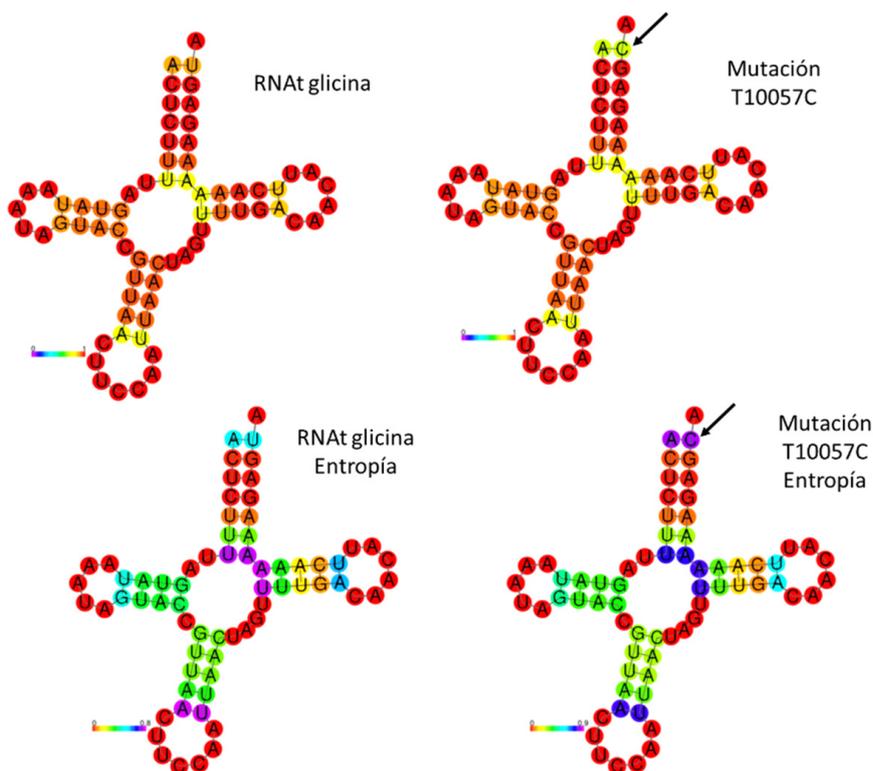
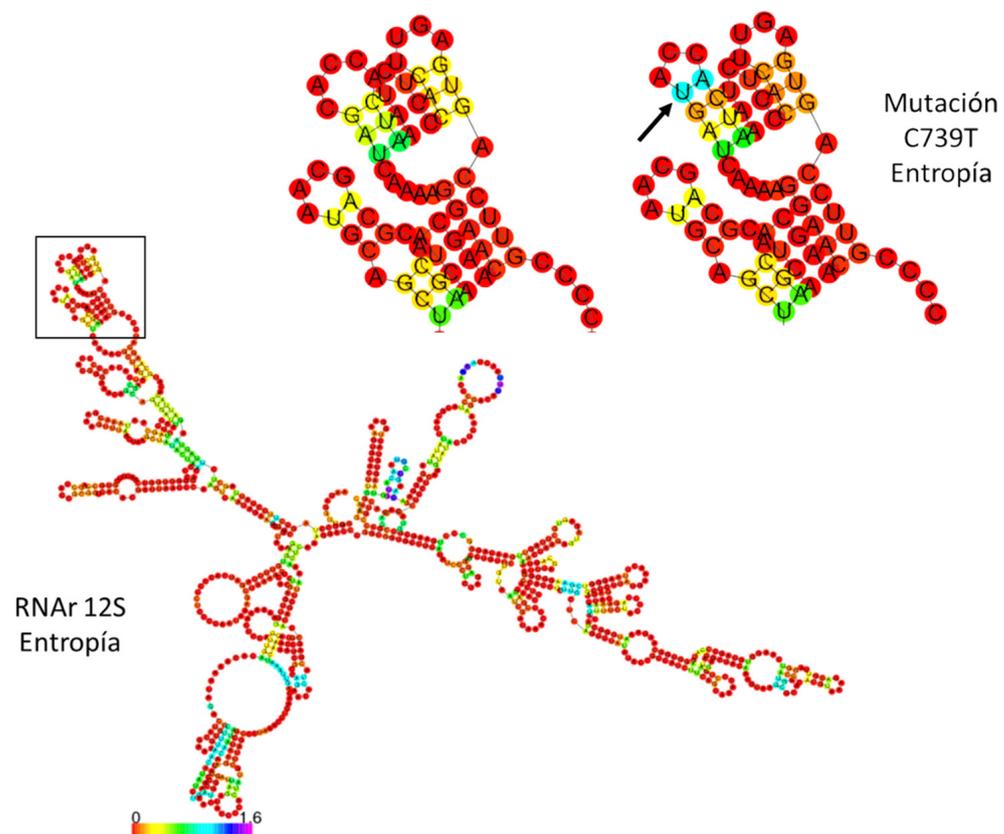
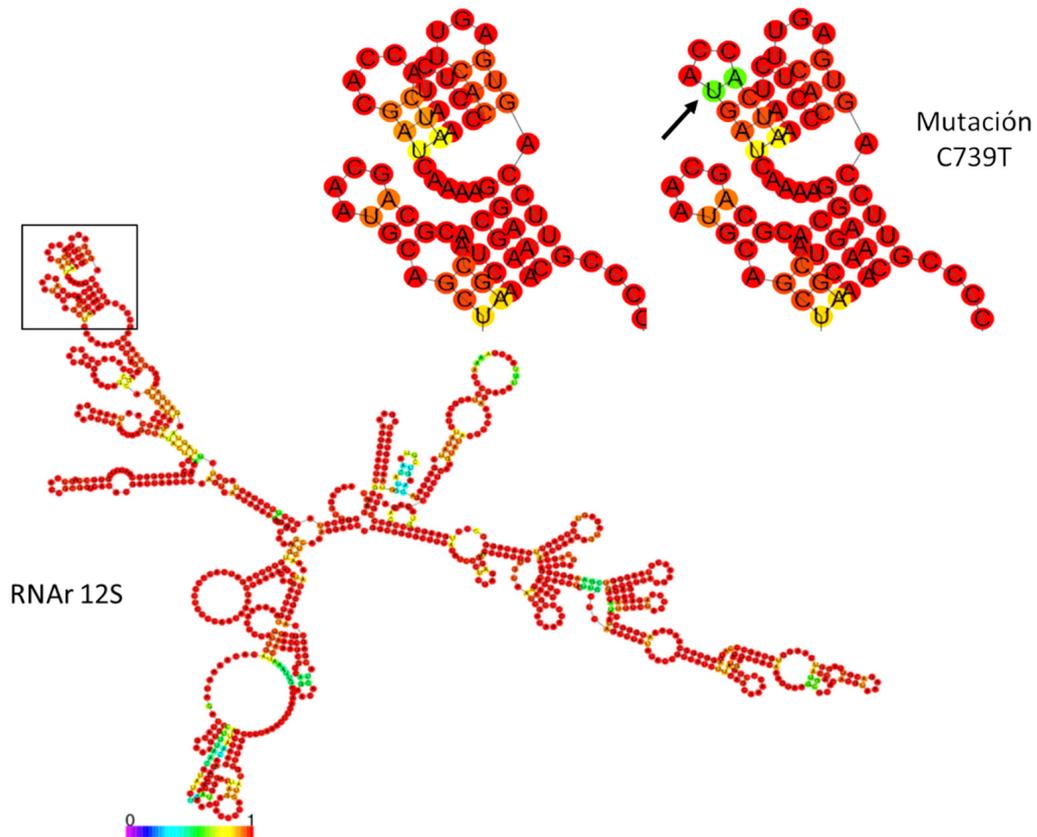


Figura 27. Mutaciones en homoplasma localizadas en posiciones estables que dan cambios a nivel de entropía. a) Mutación G1018A (RNAr 12S); b) T1850C (RNAr 16S); c) G5780A (RNAt cisteína); d) T10057C (RNAt glicina).

a) Mutación C739T (RNAr 12S): cambios a nivel estructural y de entropía.

- Energía libre RNAr 12S: -213,14 kcal/mol.

- Energía libre RNAr 12S con la mutación C739T: -213,77 kcal/mol.



- b) Mutación T1243C (RNAr 12S): cambios a nivel estructural y de entropía.
- Energía libre RNAr 12S: $-213,14$ kcal/mol.
- Energía libre RNAr 12S con la mutación T1243C: $-212,36$ kcal/mol.

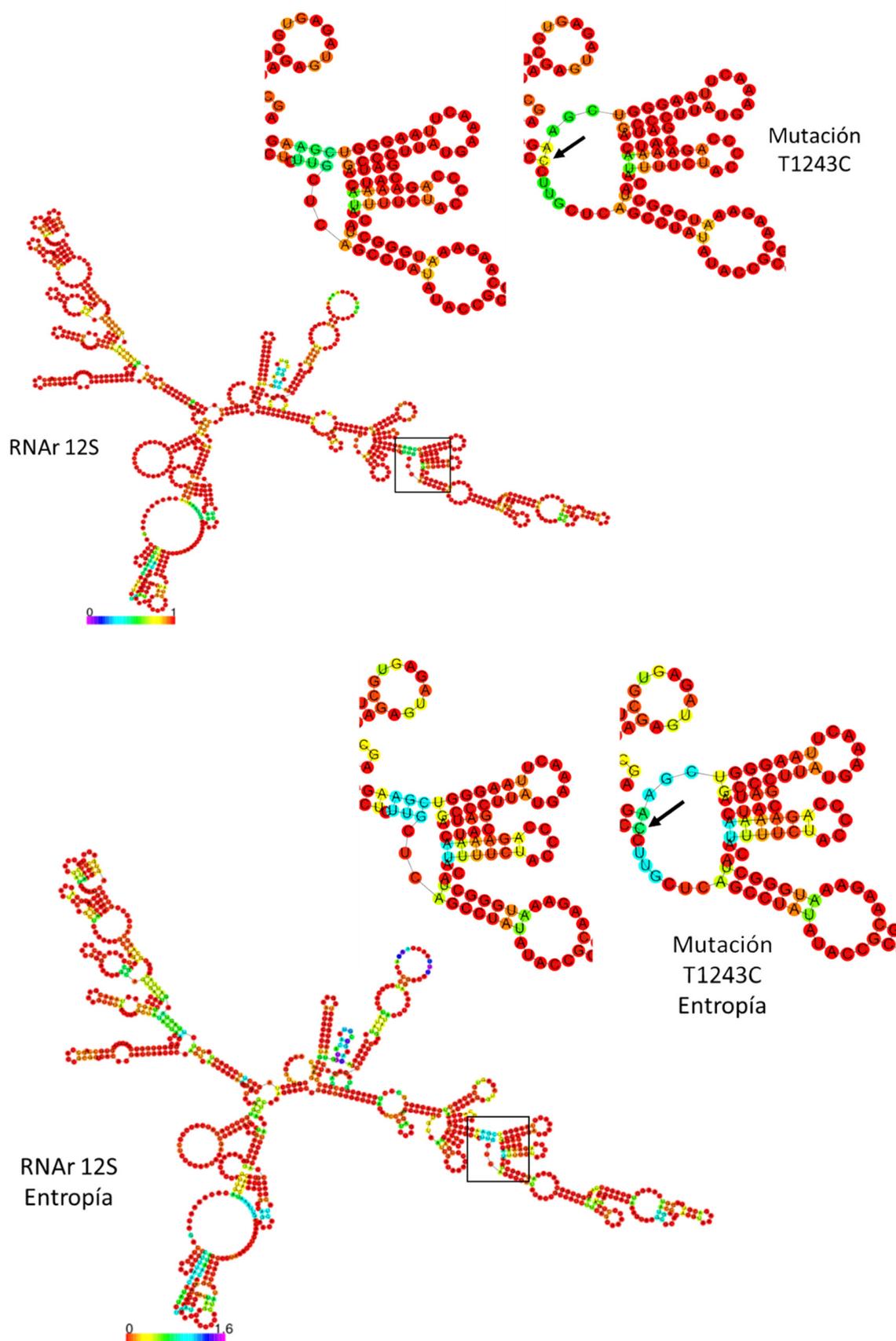


Figura 28. Mutaciones en homoplasmia localizadas en posiciones estables que dan cambios tanto a nivel estructural como de entropía. a) Mutación C739T (RNAr 12S); b) T1243C (RNAr 12S).

6.1.4. Análisis de mutaciones en heteroplasmia

En la figura 29 se observa la distribución de las heteroplasmas detectadas en el tejido tumoral en los pacientes estudiados, tanto a nivel germinal como somático. Así, 24 pacientes (75%) presentan alguna mutación en heteroplasma, mientras que los 8 restantes (25%) no presentan ninguna mutación en heteroplasma.

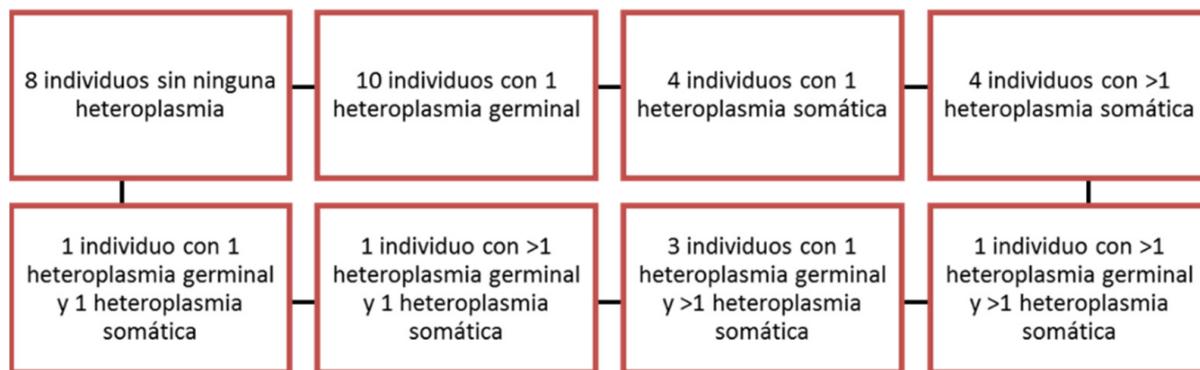


Figura 29. Distribución de las heteroplasmas detectadas (germinales y somáticas) en las muestras de los 32 pacientes analizados.

Atendiendo al porcentaje de individuos con heteroplasma, los resultados muestran un aumento significativo de individuos heteroplásmicos afectados de glioma (75%) respecto a los que no padecen cáncer (60,53%) ($p < 0,05$; OR: 2,836; 95% IC 2,48-3,24). Asimismo, se observa que la gran mayoría de heteroplasmas detectadas en la población que no padece cáncer son de origen germinal (93,42%; IC 95% 90,5-95,5), mientras que en los individuos afectados de glioma se observa que la frecuencia de heteroplasmas somáticas es superior a la de las germinales (58,7%; IC 95% 44,3-71,7) (Figura 30).

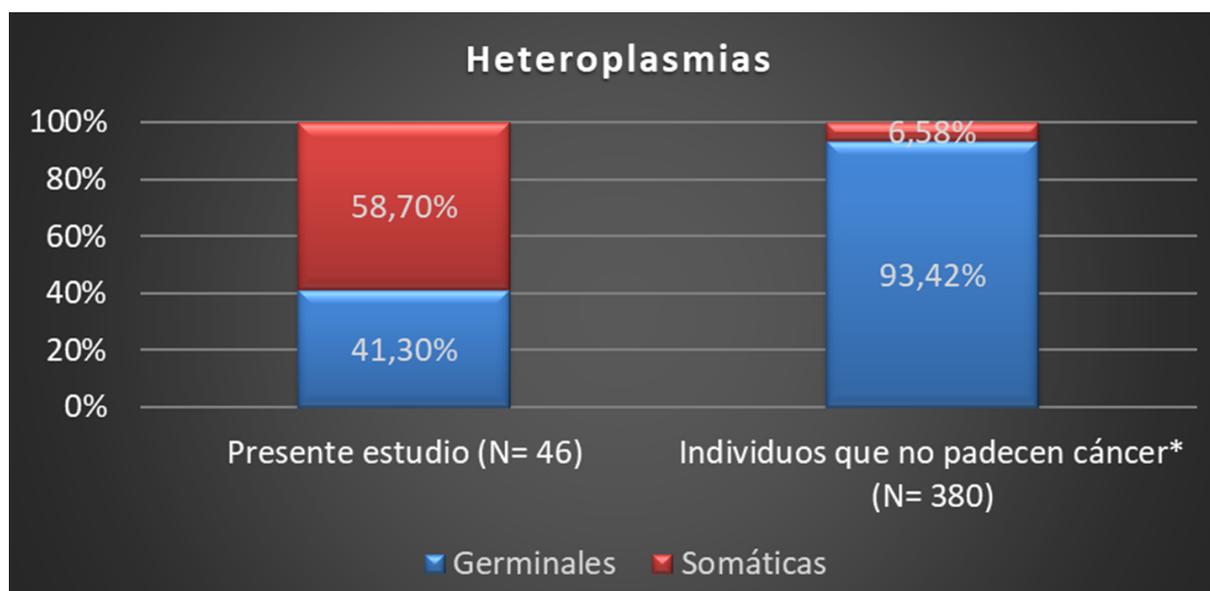


Figura 30. Porcentaje de heteroplasmas germinales y somáticas detectadas en el presente estudio y en individuos que no padecen cáncer. *Muestras de tejido cerebral de individuos que no padecen cáncer. Datos reportados por Li et al. 2015 (niveles de heteroplasma $\geq 3\%$) y clasificados por Ramos et al. (2016).

Por último, diferenciando los tumores más agresivos –grado de malignidad IV (N=17)– y menos agresivos –grado de malignidad II y III (N=15)–, el porcentaje de mutaciones germinales difiere en ambos grupos, siendo en los tumores menos agresivos del 36,84% y en los más agresivos del 63,16%. Por el contrario, en relación a las heteroplasmas somáticas, el porcentaje detectado en los tumores menos agresivos es 81,48% y en los más agresivos 18,52%.

Distribución de las mutaciones en heteroplasma a lo largo de genoma mitocondrial

Se han detectado un total de 46 mutaciones en heteroplasma. En la tabla 40 se representa su distribución en las distintas regiones del genoma mitocondrial, así como la ratio de mutaciones por par de bases detectada en cada región. Únicamente se detecta una mayor ratio de mutaciones en la región control respecto a la región codificante ($Z= 6,156$; $p<0,0001$; Corrección Bonferroni $p<0,0042$).

Tabla 40. Distribución de las mutaciones en heteroplasma detectadas en el estudio. Cálculo de la ratio de mutaciones por par de bases (mutaciones/pb región mitocondrial). pb: pares de bases.

Región mitocondrial	Nº mutaciones en heteroplasma N=46	Tamaño de la región (pb)	Mutaciones/pb
Región control	14	1180	0,012
HVRI	5	360	0,014
HVRII	6	316	0,019
HVRIII	1	137	0,007
Región Codificante	32	15389	0,002
RNAs	10	4021	0,002
RNAr	10	2513	0,004
12S	4	954	0,004
16S	6	1559	0,004
Subunidades de la cadena respiratoria	22	11395	0,002
Complejo I	12	6356	0,002
Complejo III	3	1141	0,003
Complejo IV	4	3010	0,001
Complejo V	3	888	0,003

Mutaciones en heteroplasma detectadas en posiciones estables

En la figura 31 se esquematiza la distribución de las mutaciones en heteroplasma en las distintas regiones del genoma mitocondrial, teniendo en cuenta si se localizan en posiciones estables o no

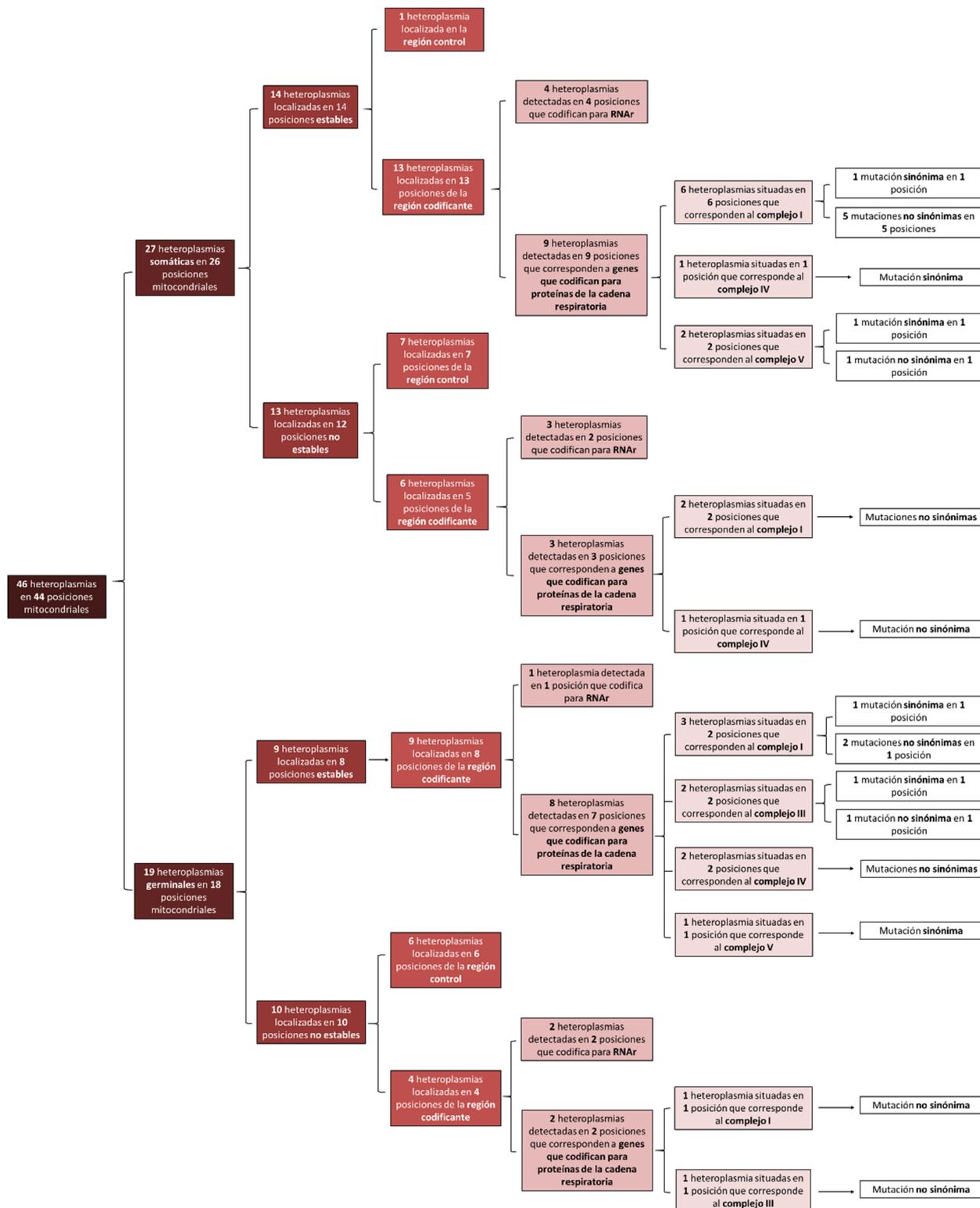


Figura 31. Esquema de las mutaciones detectadas en heteroplasma a lo largo del genoma mitocondrial (región control, RNAs y genes que codifican para proteínas de la cadena respiratoria, tanto a nivel somática como germinal. Del mismo modo que en el caso de las mutaciones en homoplasma, se consideran posiciones estables cuando tienen <6 hits en la filogenia y con un índice de conservación <20% o >80%.

estables. De las 46 heteroplasmas, 27 (58,7%) heteroplasmas son somáticas –se han detectado únicamente en el tejido tumoral–, y 19 (41,3%) son germinales –se han detectado tanto en el tejido tumoral como sanguíneo–. En el caso de las heteroplasmas somáticas, 14 (51,85%) se localizan en 14 posiciones estables: 1 (7,14%) en la región control y las 13 restantes (92,86%) en 13 posiciones de la región codificante. De éstas últimas, 4 heteroplasmas (30,77%) se localizan en 4 posiciones que codifican para RNAs y 9 heteroplasmas (69,23%) en 9 posiciones que codifican para proteínas de la cadena respiratoria. En relación a las heteroplasmas germinales, 9 (47,37%) se localizan en 8 posiciones estables, todas ellas de la región codificante. De éstas, 1 heteroplasmia (11,11%) se localiza en una posición que codifica para RNA y las 8 restantes (88,89%) en 7 posiciones que codifican para proteínas de la cadena respiratoria (Figura 31; Tabla 7 del Anexo III).

Respecto a las 23 mutaciones en heteroplasmas localizadas en posiciones estables, en la figura 32 se describe gráficamente su localización dentro del genoma mitocondrial y su frecuencia.

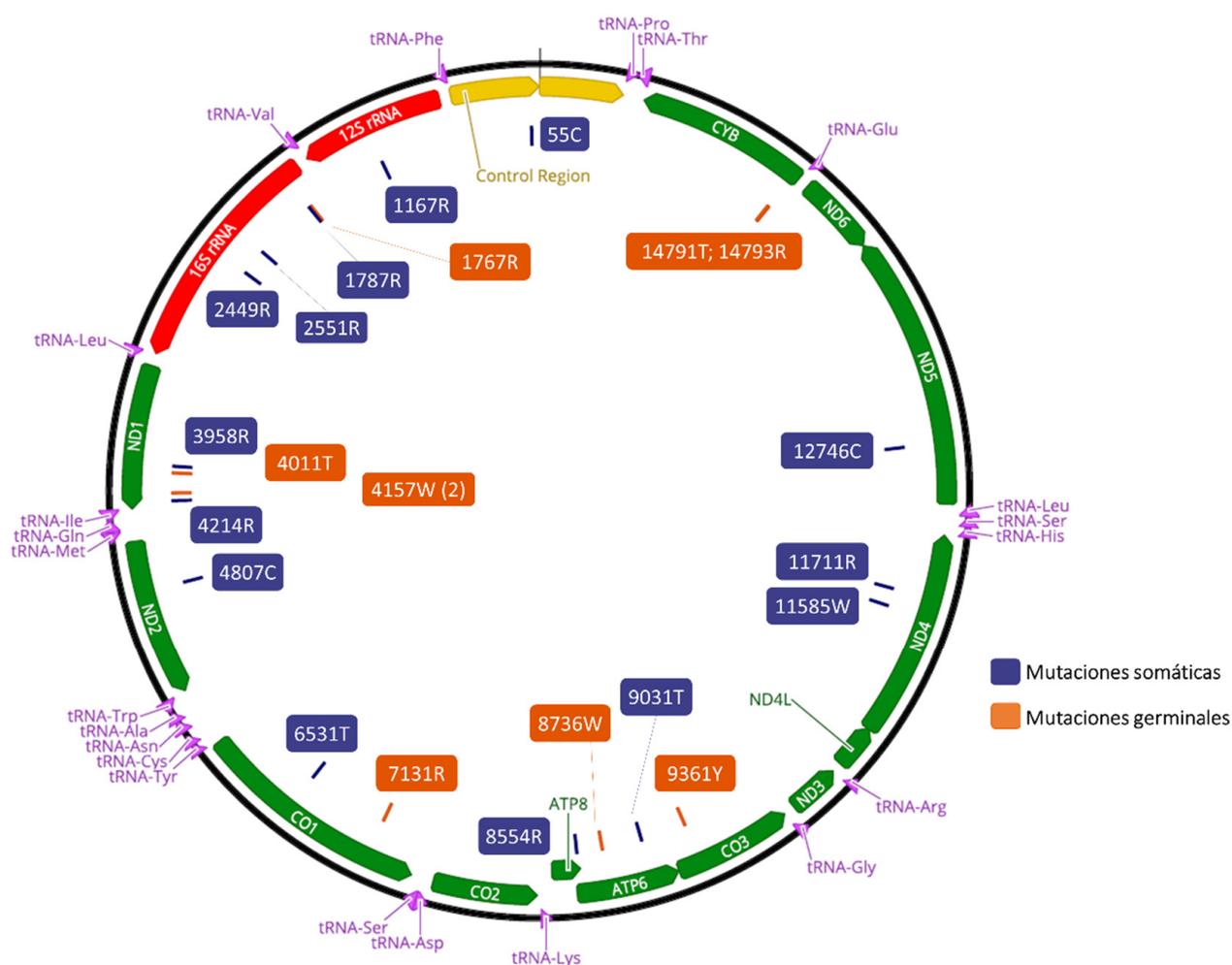


Figura 32. Gráfico que muestra la localización de las mutaciones en heteroplasmia detectadas en posiciones estables (<6 hits en la filogenia y que tienen un índice de conservación <20% o >80%) (N=23). Únicamente la heteroplasmia 4157W se detecta con una frecuencia >1 individuo.

De las 9 heteroplasmas somáticas detectadas en 9 posiciones estables que codifican para proteínas de la cadena respiratoria, 6 (66,67%) son mutaciones no sinónimas (Figura 31). Para estas heteroplasmas, los valores de Mutpred2 (valores >0,5) determinan que 4 de ellas [3958R (3,24%) (ND1); 4807C (40,32%) (ND2); 11585W (5,75%) (ND4); 12746C (14,47%) (ND5)] pueden tener un efecto deletéreo. Asimismo, para las 4 heteroplasmas localizadas en posiciones que codifican para RNAr, únicamente la mutación 2449R (5,46%) (RNAr 16S) presenta cambios a nivel tanto estructural como de entropía en la predicción de sus estructuras secundarias (Figura 33).

Del mismo modo, de las 8 heteroplasmas germinales detectadas en 7 posiciones estables que codifican para proteínas de la cadena respiratoria, 5 (62,5%) son mutaciones no sinónimas. Para estas heteroplasmas, los valores de Mutpred2 (valores >0,5) determinan que únicamente la heteroplasmia 9361Y (1,16% en tumor y 3,97% en sangre) (CO3) puede tener un efecto deletéreo. Respecto a la heteroplasmia 1767R (2,44% en tumor y 13,5% en sangre) (RNAr 16S) localizada en una posición que codifica para RNAr, la predicción de su estructura secundaria determina una variación tanto a nivel estructural como de entropía en la predicción de sus estructuras secundarias (Figura 34).

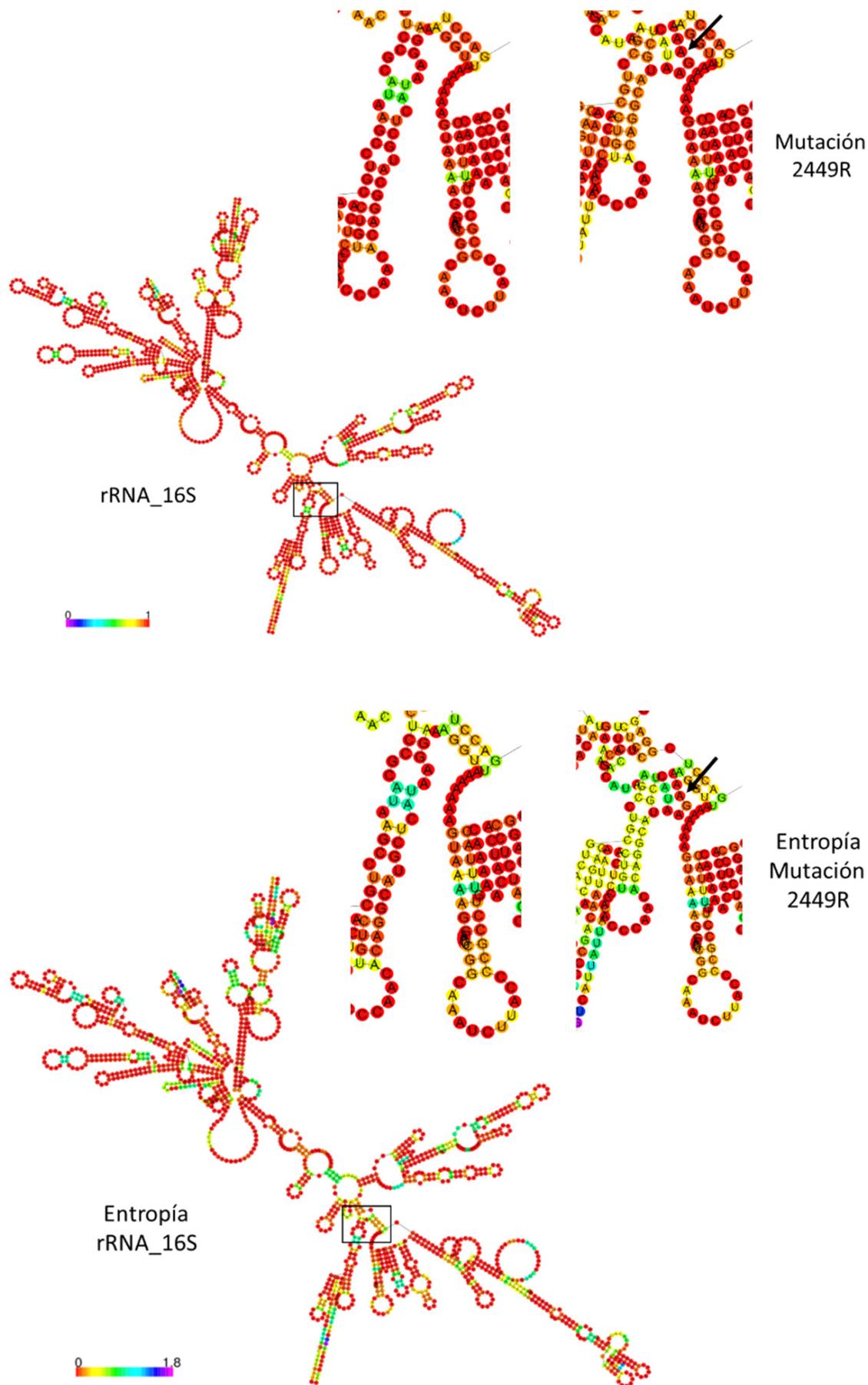


Figura 33. Cambios a nivel estructural y de entropía del RNAr 16S con la heteroplasmia somática 2449R (5,46%).
 Energía libre RNAr 16S: -336,35 kcal/mol; Energía libre RNA 16S con la mutación 2449R: -332,06 kcal/mol.

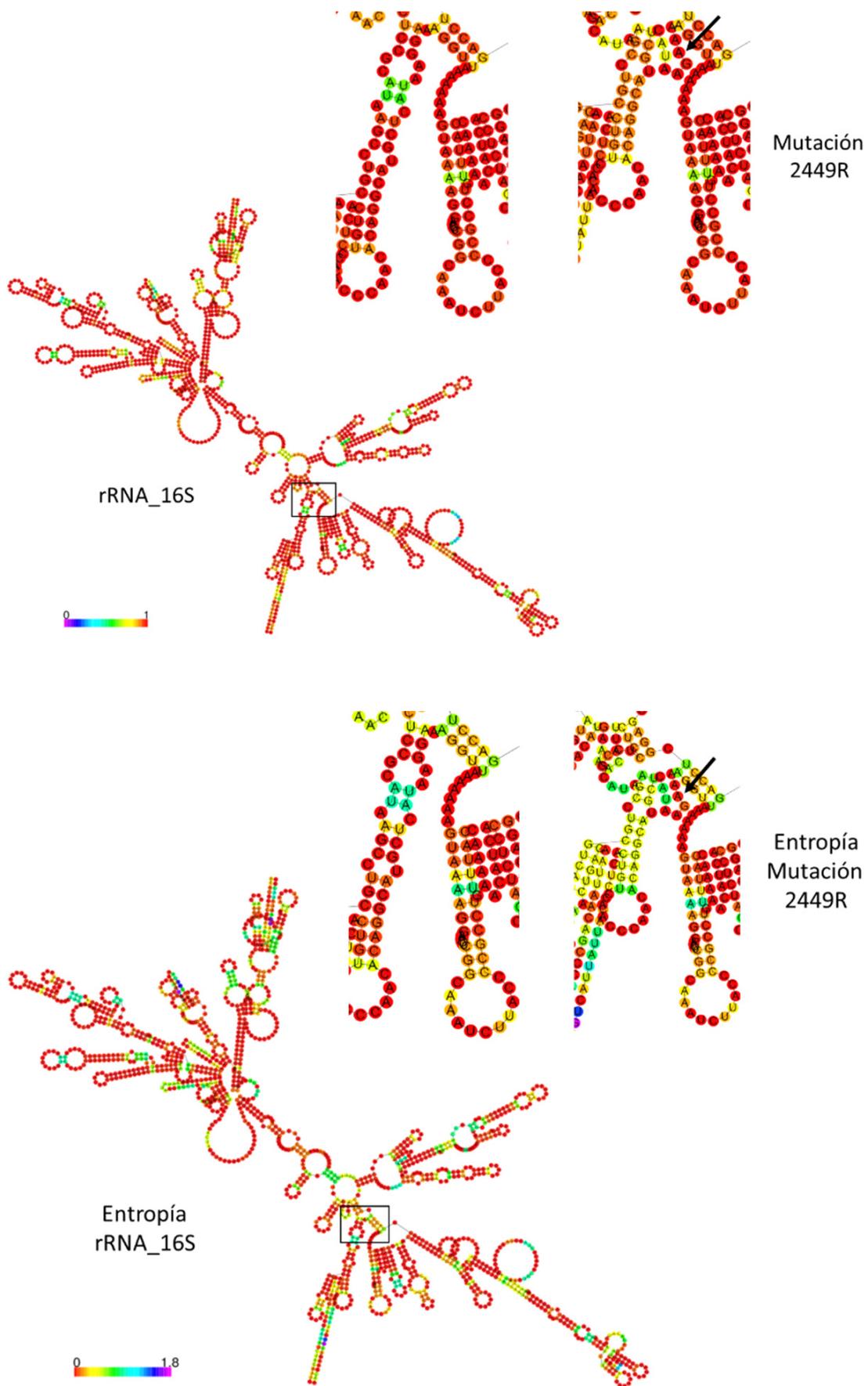


Figura 34. Cambios a nivel estructural y de entropía del RNAr 16S con la heteroplasma germinal 1767R (2,44% en tumor y 13,5% en sangre). *Energía libre RNAr 16S: -336,35 kcal/mol; Energía libre RNA 16S con la mutación 1767R: -331,81 kcal/mol.*

6.2. DISCUSIÓN

6.2.1. Asociación de haplogrupos mitocondriales y el riesgo de sufrir glioma

La función mitocondrial presenta cambios interindividuales debido a las variantes presentes en el DNAm. Entre estas variantes, determinados polimorfismos adaptativos que se han acumulado por el efecto de la selección positiva a lo largo del tiempo, debido a que conferían cierta ventaja evolutiva para la adaptación ambiental de las poblaciones humanas, pueden, a su vez, predisponer al desarrollo ciertas patologías (Wallace 2013). En este sentido, y tal y como se ha comentado anteriormente, existen varios estudios que relacionan haplogrupos mitocondriales con la predisposición a sufrir determinados tipos de cáncer (véase apartado 1.2.1, Tabla 3). No obstante, hasta la fecha, no se ha publicado ningún estudio de asociación entre haplogrupos mitocondriales y el riesgo de sufrir glioma, siendo por tanto el presente estudio el primero de este tipo.

En la presente tesis se describe una asociación entre el haplogrupo J y el aumento del riesgo de sufrir glioma, observándose un aumento significativo de la frecuencia de individuos pertenecientes a este haplogrupo respecto a la población general. Los estudios anteriormente publicados (Gasparre et al. 2007; Vega et al. 2004; Vidone et al. 2015), en los que se asigna a los individuos afectados de glioma su haplogrupo mitocondrial correspondiente, no se han centrado en analizar posibles asociaciones. Sin embargo, un reanálisis de estos trabajos ha permitido verificar que se detecta la misma tendencia encontrada en el presente estudio, es decir, un aumento en la frecuencia de individuos afectados de glioma pertenecientes al haplogrupo J respecto de la población general, aunque en ninguno de ellos esta tendencia es significativa. Ahora bien, combinando nuestro estudio con otros ya realizados a nivel europeo, se ha podido determinar que el haplogrupo J podría predisponer a sufrir glioma, de manera que los individuos pertenecientes a este haplogrupo tendrían un riesgo superior a los de otros haplogrupos de padecer esta enfermedad (OR: 2,407; 95% IC 1,538-3,764). Teniendo en cuenta estos resultados, cabe la posibilidad de que dicho haplogrupo favorezca el desarrollo del glioma. Varios estudios han asociado la frecuencia más alta del haplogrupo J caucásico en las regiones del norte de Europa con la adaptación a ambientes fríos (Ruiz-Pesini et al. 2004). Este haplogrupo se caracteriza por la acumulación preferencial de mutaciones no sinónimas que producen cambios a nivel aminoacídico en los complejos I (NADH) y III (CYB) de la cadena de transporte de electrones: T4216C (Y304H) (ND1), A10398G (T114A) (ND3), G13708GA (A458T) (ND5) y C15452A (L236I) (CYB) (Mishmar et al. 2003; Pierron et al. 2011). Esto permitiría un desacoplamiento de la cadena respiratoria y permitiría producir más calor a cambio de generar menos ATP. El hecho de producir menos ATP puede disminuir la

producción de ROS generado por el incremento de oxidación de la cadena de transporte de electrones, reduciendo así el daño oxidativo (menor cantidad de ROS) y, por lo tanto, el riesgo de provocar apoptosis (Ruiz-Pesini et al. 2004). Este efecto permitiría que el haplogrupo J tuviera un efecto protector frente a determinadas enfermedades neurodegenerativas como el Parkinson (van der Walt et al. 2003), ya que se conoce que el estrés oxidativo con el tiempo puede provocar la degeneración de neuronas y la acumulación de proteínas fibrilares (Betarbet et al. 2000; Sherer et al. 2002; van der Walt et al. 2003) y estaría relacionado con una mayor longevidad respecto a la población (Rea et al. 2013). Sin embargo, dicha propiedad también se podría relacionar con el proceso tumoral, dado que, a medida que el proceso tumoral avanza, un incremento desproporcionado de ROS intracelular podría dañar a la célula y conducir a un proceso de apoptosis (Liou and Storz 2010). Ante esta situación, la célula tumoral reduce la generación de ROS gracias a una rápida respuesta antioxidante, a fin de mantener la integridad del genoma mitocondrial, evitando el proceso de apoptosis y permitiendo así que el proceso tumoral pueda avanzar (Seoane et al. 2011; Soon et al. 2017). En este sentido, la generación de una menor cantidad de ROS por parte de este haplogrupo permitiría que los propios agentes antioxidantes de las células tumorales eliminen con mayor facilidad dicha cantidad de ROS, evitando así entrar en apoptosis (Liou and Storz 2010). Este hecho, a su vez, explicaría que el tumor pudiera progresar a grados de mayor malignidad, tal y como se ha hallado por primera vez en el presente estudio, en el que se observa una sobrerrepresentación significativa de individuos pertenecientes al haplogrupo J con glioma con grado de malignidad IV. Así pues dicho haplogrupo parece contar con determinadas variantes mitocondriales que podrían propiciar un ambiente favorable para el desarrollo del glioma.

Bajo esta misma línea, los resultados obtenidos respecto al haplogrupo T muestran que, aunque el aumento de la frecuencia de individuos con haplogrupo T en el presente estudio no es significativo, sí se detecta la misma tendencia observada en otros trabajos (Gasparre et al. 2007; Vega et al. 2004; Vidone et al. 2015). Tal y como sucede con el haplogrupo J, los resultados a nivel europeo combinando nuestro estudio con todos los ya realizados demuestran que también el haplogrupo T proporcionaría cierto riesgo a sufrir glioma (OR: 1,817; 95% IC 1,176-2,806). Varios estudios han asociado el haplogrupo T con el riesgo de sufrir varias patologías tales como la astenozoospermia (Ruiz-Pesini et al. 2000), la arteriopatía coronaria y retinopatía diabética (Kofler et al. 2009), la obesidad (Ebner et al. 2015) o el cáncer colorrectal (Li et al. 2015b). Sin embargo, del mismo modo que sucede con el haplogrupo J, no se ha hallado ningún estudio que relacione el haplogrupo T con el riesgo de sufrir este tipo de cáncer.

Se conoce que el haplogrupo T deriva del haplogrupo JT, de manera que comparte determinadas mutaciones con el haplogrupo J [T4216C (ND1), A11251G (ND4) y C15452A (CYB)]. Tal y como se ha

comentado para el haplogrupo J, 2 de estas mutaciones dan lugar a cambios a nivel aminoacídico [T4216C (Y304H) (*ND1*) y C15452A (L236I) (*CYB*)], de forma que el haplogrupo T también tendría una menor eficiencia energética debido a la presencia de estos polimorfismos que provocan el desacoplamiento parcial de la cadena respiratoria, produciendo así una menor cantidad de ROS y una menor susceptibilidad al estrés oxidativo (Aldamiz-Echevarria et al. 2019; Arning et al. 2010; Mueller et al. 2012). Además, el haplogrupo T posee la mutación no sinónima A4917G (N150D), que se caracteriza por ser un polimorfismo localizado en una posición altamente conservada del gen *ND2* y que confiere una ventaja adaptativa ambiental a los portadores de dicho haplogrupo (Ruiz-Pesini et al. 2004; Wallace 2005). Esta mutación podría, por lo tanto, contribuir al desacoplamiento de OXPHOS y a la disminución en la generación de ROS.

Respecto al grado de malignidad tumoral, como se ha mencionado anteriormente, el haplogrupo J podría generar un ambiente favorable para que el tumor pueda ir desarrollándose y alcanzar un mayor grado de malignidad. Por otro lado, en relación a los resultados obtenidos respecto al haplogrupo H, que presenta una frecuencia en pacientes de glioma similar a la de la población general, se observa una tendencia negativa en relación a la malignidad tumoral, existiendo una subrepresentación significativa de individuos con grado de malignidad tumoral IV respecto de la población general y una sobrerrepresentación de tumores de grado II respecto a los de grado IV. Se considera que las mutaciones mitocondriales que definen el haplogrupo H pueden provocar la generación de una mayor cantidad de ROS (Martinez-Redondo et al. 2010), al contrario que los individuos con haplogrupo J y T. La gran cantidad de ROS generada por dicho haplogrupo podría obstaculizar al tumor el poder eliminar por sí mismo todo el ROS generado y, por lo tanto, tener mayores problemas para poder progresar. Así pues, se puede considerar que los individuos con haplogrupo H podrían tener cierta ventaja para evitar la progresión del tumor cerebral a grados de mayor malignidad. En este sentido, Liou and Storz (2010) consideran que un aumento de ROS intracelular puede inducir la detención del ciclo celular de las células cancerígenas, la senescencia y la apoptosis a partir del agotamiento de los agentes antioxidantes de las células tumorales o la generación de ROS por parte de las células sanas. Esto, por lo tanto, se correlacionaría con las propiedades de los individuos del haplogrupo H y los resultados obtenidos en el presente estudio.

6.2.2. Distribución de las mutaciones a lo largo del genoma mitocondrial

Los resultados del presente estudio muestran diferencias en el patrón de distribución de las mutaciones entre los pacientes que sufren glioma y la población general, detectando una sobrerrepresentación de

mutaciones en las regiones *RNAr 12S* y el Complejo V, y subrepresentación en las regiones HVRI y el Complejo IV. Ante estos resultados, cabe resaltar varios aspectos. Por un lado, es destacable el hecho de que no haya una sobrerrepresentación de mutaciones en los Complejos I y III en las muestras de glioma, puesto que dichos complejos son los que generan una mayor producción de ROS, debido a una mayor fuga de electrones durante el proceso de OXPHOS (Bottje 2018). Por otro lado, la sobrerrepresentación de mutaciones en los RNAs 12S podría estar relacionado con alteraciones en la función de los MOTS-c, péptidos que son codificados por este gen ribosomal y que regulan el metabolismo energético, estimulando la absorción y utilización de la glucosa e inhibiendo la respiración oxidativa (Yang et al. 2019).

Basándonos únicamente en las mutaciones en homoplasmi de nuestro estudio, también se ha detectado que las mutaciones se distribuyen de forma diferencial en determinadas regiones mitocondriales. Por un lado, la región control es la que tiene una ratio de mutaciones mayor respecto a la región codificante, considerándose una de las regiones más susceptibles de sufrir mutaciones en cáncer (Iommarini et al. 2013; Lloyd et al. 2015). Además, aunque se considera que la tasa de sustitución de la HVRI es hasta 2 veces superior a la observada en la HVRII debido principalmente a la elevada frecuencia de transiciones pirimidina-pirimidina, en la presente tesis se detecta una ratio de mutaciones superior en la HVRII. Este resultado se asemeja al descrito en otros estudios basados en cáncer donde también detectan una frecuencia de mutaciones superior en la HVRII respecto a la HVRI (Shu et al. 2019). No obstante, se ha de tener en cuenta que las mutaciones A73G (HVRII) y A263G (HVRII) se detectan en 21 y 32 individuos, respectivamente, de manera que el 46,09% de las mutaciones detectadas en esta región se explicaría únicamente con estos 2 polimorfismos asociados a haplogrupo. Esto mismo es lo que sucede si se comparan los complejos de la cadena respiratoria. Aunque el complejo III (CYB) y el complejo V (ATP) son las regiones donde se detecta una ratio de mutaciones superior al resto de complejos, en ambas regiones se detectan variantes con una frecuencia muy elevada dentro del estudio. Por un lado, en el complejo III, el 65,66% de las mutaciones detectadas se debe a 3 polimorfismos no sinónimos asociados a haplogrupo: C14766T (T7I) en 21 individuos, A15326G (T194A) en 32 individuos y C15452A (L236I) en 12 individuos. Asimismo, en los presentes resultados se muestra que 32 de las 49 mutaciones (65,31%) detectadas en el complejo V, concretamente en el gen *ATP6*, corresponden al polimorfismo A8860G (T112A) que está presente en todos los haplogrupos detectados en los 32 individuos del estudio. Resultados similares se han encontrado en diversos tipos de cáncer como el colorrectal, el de mama o de ovario donde aparece este mismo polimorfismo en un porcentaje muy elevado de individuos (75-100%) (Aikhionbare et al. 2004; Aikhionbare et al. 2007; Ghaffarpour et al. 2014; Mehrabi et al. 2010), así como en enfermedades como Alzheimer, trastorno bipolar o LHON

(Fauser et al. 2002; Kirk et al. 1999; Mostafaie et al. 2004). Además, el gen *ATP6* es el más variable dentro del genoma mitocondrial humano (Mishmar et al. 2003) y en el que se encuentra un exceso de polimorfismos no sinónimos tanto asociados como no asociados a haplogrupo (Elson et al. 2004).

Por el contrario, el complejo IV (CO) tiene una ratio de mutaciones menor respecto al resto de complejos. Existen varios estudios basados en las mutaciones de tumores a nivel somático donde describen también una ratio de mutaciones inferiores en esta región. Por ejemplo, Pereira et al. (2012) describen una frecuencia significativamente menor de mutaciones en el gen *CO3* en tumores oncocíticos, y determina una cierta protección del gen *CO1* a sufrir mutaciones, mientras que Stafford and Quin (2010) consideran que los tumores podrían tener cierta dependencia a la actividad del complejo IV o, concretamente, a la función del gen *CO1* para poder progresar debido a que se ha detectado selección negativa más fuerte para este gen.

Respecto a los RNAs, se detecta una ratio de mutaciones mayor en la región del *RNAr 12S*. Del mismo modo que se ha comentado anteriormente, se puede observar que las mutaciones A750G y A1438G se detectan en el 100% de los individuos y son definitorias de haplogrupo en todos los casos, de manera que el 80% de las mutaciones vienen dadas por estos 2 polimorfismos localizados en los 32 individuos. Por el contrario, los *RNAt* tienen una ratio de mutaciones menor, obteniendo un resultado similar al descrito en un estudio a nivel tumoral de mutaciones somáticas, donde describe también una ratio más baja y sugiriendo una posible restricción por selección negativa en las regiones de *RNAt* (Li et al. 2017).

6.2.3. Mutaciones detectadas en pacientes afectos de glioma

Los presentes resultados muestran que la mayoría de las mutaciones que se detectan son de origen germinal (95,02%) y muy pocas mutaciones son de origen somático (4,98%). Respecto a estas últimas, del total de mutaciones localizadas en genes que codifican proteínas de la cadena respiratoria, el 75% son mutaciones que producen un cambio de aminoácido. Este resultado hace que se considere el hecho de que estas mutaciones somáticas pueden generar alteraciones funcionales en la mitocondria que proporcionarían cierta ventaja para el desarrollo tumoral. Sin embargo, todas las mutaciones somáticas detectadas en el estudio se detectan en heteroplasmia. Teniendo en cuenta que el defecto bioenergético se agrava a medida que la población de *DNAmt* mutado respecto al *DNAmt* de tipo salvaje aumenta (Wallace and Chalkia 2013), se considera un valor umbral mínimo de 60% para que se pueda considerar que la mutación pueda tener cierto impacto funcional (Tuppen et al. 2010). En los resultados obtenidos, únicamente 2 mutaciones [4807C (F113S) (*ND2*) y 12746C (L137P) (*ND5*)] pueden

tener efecto patogénico y superan el valor umbral mínimo establecido para considerar que ambas pueden contribuir al desarrollo de la enfermedad (60% y 85,53%, respectivamente). Ahora bien, dicho umbral puede variar según la mutación y el tejido, de manera que los tejidos que dependen altamente del metabolismo OXPHOS podrían presentar afectación a niveles de heteroplasmia menores que los tejidos menos dependientes (Tuppen et al. 2010). Es ampliamente conocido que el tejido cerebral es uno de los tejidos que requiere una mayor demanda energética –el cerebro representa sólo un 2% de la masa corporal y consume hasta el 20% del oxígeno (Wallace 2013)–. Además, es uno de los más susceptibles a los efectos del ROS debido a su alta actividad metabólica y la relativa baja capacidad de regeneración celular (Rinaldi et al. 2016). Así, aunque en el presente trabajo el nivel de heteroplasmia en las muestras tumorales sea bajo en la mayoría de casos –15 de las 27 mutaciones (55,56%) no superan el 10% de heteroplasmia– podría ser que alguna de estas mutaciones tenga relevancia en la enfermedad, aunque el nivel de DNAm_t mutado no supere el umbral mínimo establecido para tener consecuencias a nivel fenotípico. Por otro lado, cabe considerar que el porcentaje de DNAm_t mutado detectado puede variar según la zona del tejido seleccionado para el análisis, y que, por lo tanto, los niveles de heteroplasmia reportados sean aproximados.

Existen numerosos estudios que describen mutaciones somáticas en cáncer que se han llegado a fijar en el genoma mitocondrial (véase 1.2.1, Tabla 2), indicando que ciertas mutaciones con impacto a nivel aminoacídico pueden conferir cierta ventaja para el desarrollo tumoral. Estas mutaciones, que aparecerían inicialmente en un estado de heteroplasmia, pueden llegar a fijarse dentro del genoma mitocondrial a partir de una posible actuación de la selección positiva (Li et al. 2018; McMahon and LaFramboise 2014) o bien, de la relajación de la selección negativa (Stafford and Chen-Quin 2010). No obstante, otros autores consideran que no hay necesidad de que haya una ventaja fisiológica o que exista un papel relevante de las mutaciones mitocondriales en la aparición o el desarrollo del tumor, y que no habría necesidad de la actuación de la selección para estas mutaciones, sino que se fijarían por efecto de la deriva genética (Coller et al. 2001; Chinnery et al. 2002; Jones et al. 2001; Vega et al. 2004). Ahora bien, es llamativo que, en el presente estudio, el 100% de las mutaciones somáticas están en heteroplasmia. Este resultado podría ser debido a diferentes motivos. Por un lado, aunque se ha descrito que gran parte de las mutaciones somáticas en homoplasmia se detectan en los últimos estadios del tumor (Park et al. 2009), el tiempo transcurrido para que estas mutaciones que pueden ayudar a la proliferación tumoral se fijen, podría no haber sido suficiente en este caso. De hecho, algunas de estas heteroplasmas somáticas sí se han detectado en homoplasmia en otros estudios basados en cáncer, como la mutación T16093C en cáncer de mama y próstata (Brandon et al. 2006; Chen et al. 2002; Parrella et al. 2001) y la mutación T16519C en cáncer de gástrico, de ovario,

Tabla 41. Resultado de las mutaciones que pueden tener relevancia en el presente estudio.

Posición	Mutación	Localización en el genoma mitocondrial	Frecuencia	Estado de la mutación	Origen de la mutación	% Base minoritaria	Sobrerrepresentación en el presente estudio respecto a la población general	Estabilidad de la posición	Efecto (MutPred2) (MitoTIP)	Estructuras secundarias
146	T>C	Región control (HVRII) / ATT / OH / 7S DNA	8	Homoplasmia	Germinal		Sí	No estable		Cambios a nivel conformacional y de entropía.
739	C>T	RNAr 12S	1	Homoplasmia	Germinal		No	Estable		Cambios a nivel conformacional y de entropía.
1018	G>A	RNAr 12S	1	Homoplasmia	Germinal		No	Estable		Cambios a nivel de entropía.
1243	T>C	RNAr 12S	2	Homoplasmia	Germinal		No	Estable		Cambios a nivel conformacional y de entropía.
1767	G/A	RNAr 16S	1	Heteroplasmia	Germinal	2,44% Tumor 13,5% Sangre				Cambios a nivel conformacional y de entropía.
1850	T>C	RNAr 16S	1	Homoplasmia	Germinal		No	Estable		Cambios a nivel de entropía.
2449	G/A	RNAr 16S	1	Heteroplasmia	Somática	5,46%				Cambios a nivel conformacional y de entropía.
3721	A>G	ND1	1	Homoplasmia	Germinal		No	Estable	Efecto deletéreo (T139A)	
3958	G/A	ND1	1	Heteroplasmia	Somática	3,24%			Efecto deletéreo (G218S)	
4807	C/T	ND2	1	Heteroplasmia	Somática	40,32%			Efecto deletéreo (F113S)	
5780	G>A	RNAt cisteína	1	Homoplasmia	Germinal		No	Estable	Posiblemente polimórfica	Cambios a nivel de entropía.
9145	G>A	ATP6	1	Homoplasmia	Germinal		No	Estable	Efecto deletéreo (A207T)	
9361	T/C	CO3	1	Heteroplasmia	Germinal	1,16% Tumor 3,97% Sangre			Efecto deletéreo (L52P)	
10057	T>C	RNAt glicina	1	Homoplasmia	Germinal		Sí	Estable	Posiblemente patogénica	Cambios a nivel de entropía.
11585	T/A	ND4	1	Heteroplasmia	Somática	5,75%			Efecto deletéreo (C276S)	
12746	C/T	ND5	1	Heteroplasmia	Somática	14,47%			Efecto deletéreo (L137P)	
16325	T>C	Región control (HVRI) / ATT / 7S DNA	3	Homoplasmia	Germinal		Sí	No estable		Cambios a nivel conformacional y de entropía.

de próstata y de pulmón (Bragoszewski et al. 2008; Brandon et al. 2006; Chen et al. 2002; Fliss et al. 2000). Incluso algunas mutaciones como T146C, T152C, A189G, G13135A, T16519C también se han detectado como fijadas en otros estudios de glioma (DeHaan et al. 2004; Kirches et al. 2001; Vega et al. 2004; Vidone et al. 2015). Por otro lado, no se puede obviar el hecho de que la ausencia de mutaciones somáticas fijadas en el tumor también podría ser debido a la gran dificultad de seleccionar en nuestras muestras exclusivamente tejido tumoral, de manera que es posible que, en determinadas zonas, sobre todo en la periferia tumoral, se encuentren tanto células transformadas como no transformadas y que se presente así cierta heterogeneidad celular. Del mismo modo, la posibilidad de que haya una vascularización del tumor para poder seguir desarrollándose (angiogénesis tumoral) también dificultaría una perfecta discriminación y selección de células tumorales. Por último, a esto se puede añadir la posible intervención de procesos celulares, como la fusión mitocondrial, para intentar corregir problemas generados en determinadas mitocondrias (Martorell 2014).

Dentro de la variabilidad mitocondrial detectada en las muestras de glioma del presente estudio, combinando los distintos análisis realizados, se destacan ciertas mutaciones que pueden tener relevancia respecto a su papel dentro del desarrollo tumoral (Tabla 41). De éstas, la gran mayoría se detecta en 1 individuo. De hecho, la mayor parte de la sobrerrepresentación de las mutaciones que se detectan en el estudio viene dada porque: i) la mayoría de las que son definitorias de haplogrupo son un reflejo de la sobrerrepresentación del haplogrupo J (78,16%), o ii) la mayoría de las mutaciones que no son definitorias de haplogrupo se detectan únicamente en 1 o 2 individuos (91,3%), pero la frecuencia en la que se detectan en la población general es muy baja o incluso 0% en algunos casos.

De todas las mutaciones descritas, únicamente 3 están sobrerrepresentadas en el presente estudio respecto a la población general, de manera que se podría considerar que pueden tener cierta relación con la enfermedad. Por un lado, la mutación T146C se ha detectado en 8 pacientes del presente estudio. Dicha mutación ha sido descrita en numerosos estudios, incluso en estudios basados en glioma. Vega et al. (2004) detectan esta mutación en varios tipos de glioma y otros tumores cerebrales tanto de grado de malignidad II como III y IV (astrocitomas de grado II y III, oligoastrocitomas de grado II y III, glioblastomas y meningiomas), no obstante, los autores consideran que este polimorfismo, que está presente en varios haplogrupos como H, K, J, M y U6 y que tiene una tasa de mutación 30 veces superior que la región control, considerándose un *hotspot* mutacional, no conferiría ninguna ventaja selectiva y que, por lo tanto, no contribuiría a la carcinogénesis. Ahora bien, otros estudios describen que este polimorfismo se detecta con una mayor frecuencia en pacientes con cáncer que en los controles. Por ejemplo, Guo et al. (2016) detectan que el cambio T146C constituiría un factor de

riesgo para el cáncer de colon. A nivel somático, dicha mutación se ha descrito en varias patologías tanto en homoplasmia en cáncer de ovario (Liu et al. 2001) como en heteroplasmia en determinados tejidos, como fibroblastos y tejido cerebral en individuos de mayor edad (Attardi 2002; Coskun et al. 2004; Michikawa et al. 1999), y patologías como cáncer de próstata (Jeronimo et al. 2001), Alzheimer (Coskun et al. 2004) u oftalmoplejía (PEO) debida a alteraciones en la DNA polimerasa gamma (POLG) (Del Bo et al. 2003). Asimismo, cabe recalcar que esta mutación también se ha detectado en el presente estudio como heteroplasmia a nivel somático en un individuo, a un nivel del 50% aproximadamente. Si nos centramos en la localización de la posición dentro del genoma mitocondrial, esta mutación se sitúa en el origen de replicación de la cadena pesada (OH). Se ha postulado la necesidad de una mayor actividad replicativa del DNAm_t para la aparición y expansión de mutaciones somáticas y, por lo tanto, determinadas mutaciones localizadas en el origen de replicación podrían estar implicadas en procesos como el cancerígeno (Zhang et al. 2003). Sin embargo, esta mutación también se ha relacionado con la longevidad (Rose et al. 2007) y se ha descrito que mutaciones situadas en el origen de replicación pueden provocar un incremento de la replicación del DNAm_t para compensar el daño oxidativo que se puede ir acumulando con la edad (Zhang et al. 2003). Así pues, el hecho de que una misma mutación pueda estar involucrada en procesos tan distintos como la longevidad y el cáncer revela que estos procesos comparten muchas características bioquímicas y genéticas, sobre todo relacionadas con el metabolismo mitocondrial (Wallace 2005). Por otro lado, la mutación T16325C, detectada en 3 pacientes afectos de glioma, no ha sido descrita previamente como asociada a otras enfermedades. Sin embargo, dicha mutación provoca un reordenamiento conformacional y una reducción de la estabilidad en la estructura secundaria de la posición localizada en un punto de unión a la membrana y en la 7S DNA, un fragmento corto de DNA que se transcribe a partir de la cadena L parental mientras la cadena H se desplaza formando el D-loop (Jemt et al. 2015). La 7S DNA permite abrir la conformación del DNAm_t, facilitando el acceso a los sitios de unión de proteínas que están involucradas tanto la replicación como la transcripción del DNAm_t (Nicholls and Minczuk 2014). Tal y como se ha visto en los resultados, esta mutación genera cambios importantes en la estructura secundaria, tanto a nivel conformacional como de entropía. Además, se localiza entre elementos control (mt3 y mt5) que están involucrados en la transcripción mitocondrial (Falah et al. 2017; Suzuki et al. 1991), de manera que la mutación T16325C podría alterar dicho proceso. Por último, la sobrerrepresentación de la mutación T10057C (RNAt glicina) viene dada porque, aunque la mutación únicamente se detecta en 1 individuo en el estudio, la frecuencia en la que se detectan en la población general es muy baja (0,01%).

Las mutaciones A3721G (T139A) y G9145A (A207T) son las únicas mutaciones germinales en homoplasmia, localizadas en posiciones estables de genes que codifican para proteínas de la cadena

respiratoria (*ND1* y *ATP6*, respectivamente), que tienen un notable efecto deletéreo. Estas 2 mutaciones, detectadas además en tumores más agresivos (grado IV), confieren una carga mutacional elevada (>0,5) para los individuos portadores de estas mutaciones (individuos 5 y 19, respectivamente). Este hecho sugiere que ambas variantes podrían tener cierta relevancia dentro de la enfermedad, sin embargo, se ha comprobado que no existe una sobrerrepresentación de ellas en el estudio, y que este resultado se deba probablemente a un efecto aleatorio. Además, no existen muchos estudios que hagan referencia a la presencia de estas mutaciones en algún proceso patogénico, y únicamente la mutación G9145A se ha descrito en otro estudio en un único paciente con enfermedad hepática colestática con atresia biliar (Koh et al. 2018). Asimismo, algunas de las mutaciones germinales en homoplasmia, localizadas en posiciones estables de RNAs que dan cambios a nivel de entropía y/o a nivel conformacional, han sido descritas anteriormente: la mutación T1243C (*RNAr 12S*) se ha descrito a nivel somático en un estudio de líneas celulares de cáncer de páncreas (Jones et al. 2001), la mutación G1018A se ha detectado en determinadas patologías como la pérdida auditiva y la no compactación del ventrículo izquierdo (Li et al. 2004b; Tang et al. 2010) y en un individuo con pérdida auditiva y cardiomiopatía hipertrófica (Chamkha et al. 2011), y la mutación T1850C se ha detectado en un paciente con LHON (Fauser et al. 2002). A su vez, la mutación G5780A se ha considerado patogénica en un estudio donde la detectan a nivel germinal en heteroplasmia (92% en tejido muscular y 56% en tejido sanguíneo) en un paciente con pérdida auditiva neurosensorial (SNHL) (Lehtonen et al. 2003). A nivel de heteroplasmia, únicamente la heteroplasmia somática 3958R (*ND1*), detectada a un nivel únicamente del 3,24%, ha sido descrita en homoplasmia en un paciente con LHON leve (Ji et al. 2016).

Por último, un 35,29% de las mutaciones que pueden tener relevancia en el estudio están en posiciones que codifican para *RNAr*. Tal y como se ha comentado anteriormente, las mutaciones localizadas en el *RNAr 12S* pueden estar involucradas en alteraciones en los MOTS-s, péptidos que desempeñan un papel en el metabolismo celular. Respecto a las mutaciones en posiciones localizadas en el *RNAr 16S*, se ha de considerar que otros péptidos como las humaninas (HN) y pequeños péptidos similares a las humaninas (SHLPs) se codifican en este *RNAr*. Las HN se han relacionado con la neuroprotección y con efectos anti-oxidantes y anti-apoptóticos, mientras que los SHLPs se han relacionado con la citoprotección y el metabolismo (Yang et al. 2019). Así pues, mutaciones localizadas en los *RNAr* que pueden tener cierto impacto pueden alterar la actividad de estos péptidos dentro del glioma.

6.2.4. Carga mutacional

Si bien los estudios de asociación han sido considerados estudios clásicos para determinar el posible papel del DNAm en enfermedades complejas, en ocasiones estas asociaciones no son detectadas en otras cohortes o incluso se descubren nuevas asociaciones con distintos haplogrupos, de manera que estos estudios pueden ser controvertidos debido a la baja reproducibilidad de los resultados (Salas and Elson 2015; Venter et al. 2017). En este sentido, el análisis basado en el cálculo de la carga mutacional permite centrar los estudios en la determinación de si los efectos acumulativos de varias mutaciones mitocondriales pueden ser factor de riesgo a nivel individual, generando así nuevos modelos de asociación a la vez que se evitan posibles falsos positivos en estudios de asociación a nivel de haplogrupo o de posición (Pienaar et al. 2017; Salas and Elson 2015). En el presente estudio, los resultados obtenidos muestran que la carga mutacional de los individuos afectados de glioma tiende a ser baja y compuesta por variantes comunes en la población general. Sólo 4 individuos tienen una carga mutacional relevante, todos ellos con diferentes rangos de edad (de 31 a 71 años), grado de malignidad tumoral (II, III y IV en 2 individuos) y haplogrupo (H, J en 2 casos y U). Aunque se podría esperar que el número de variantes poco frecuentes en la población general, con elevada carga de patogenicidad, fuera común en los pacientes afectados de glioma, o que su carga mutacional fuera significativa, dichos factores no parecen tener un papel fundamental en la enfermedad, ni a nivel general ni considerando los diferentes grados del tumor. No obstante, aunque la mayoría de individuos cuentan con una carga mutacional poco relevante, no se puede obviar el hecho de que una posible acumulación de mutaciones con leve carga también podría contribuir, en cierto modo, a la etiología de la enfermedad. Esta teoría la propone Vidone et al. (2015) en su estudio basado en glioblastomas. En dicho estudio se detecta una baja presencia de mutaciones patogénicas en los genes que codifican para proteínas de la cadena respiratoria, pero consideran que la acumulación de mutaciones podría explicar el desarrollo tumoral y la flexibilidad metabólica que tienen estos tumores tan agresivos. Asimismo, otro aspecto que se debe de valorar es que la carga mutacional seguirá incluyendo marcadores de haplogrupo y que, por lo tanto, no está libre del contexto del haplogrupo mitocondrial (Pienaar et al. 2017).

6.2.5. Implicaciones del DNAm en el glioma cerebral

Teniendo en cuenta todo lo expuesto anteriormente, es importante recalcar determinados aspectos que pueden respaldar la hipótesis propuesta en el presente estudio. Por un lado, los haplogrupos mitocondriales en que sus polimorfismos provocan un desacoplamiento de la cadena respiratoria como medida de adaptación al entorno y, por lo tanto, generan una menor cantidad de ROS y evitan

situaciones de estrés oxidativo para la célula, son los que tienen más riesgo de sufrir este tipo de tumor. Por otro lado, las consecuencias fenotípicas de las mutaciones detectadas tanto a nivel somático como germinal dependerán de muchos factores tales como la carga mutacional, la demanda energética del tejido o el tipo de mutación (Wong 2007b). En este sentido, la baja cantidad tanto de mutaciones sobrerrepresentadas en el estudio –la mayoría se detectan en un único individuo– como de mutaciones en posiciones estables que puedan llegar a ser patogénicas, y la poca acumulación de mutaciones que dé lugar a una elevada carga mutacional individual, sugiere una preferencia por parte del tumor de no acumular mutaciones en el DNAm. Esta idea también queda reflejada en el aumento significativo de hasta 2,836 veces más de individuos heteroplásmicos afectados de glioma respecto a los que no padecen cáncer, indicando que, durante el proceso cancerígeno, se observaría un incremento en la generación de mutaciones, pero que posiblemente se intente evitar que se fijen en el genoma mitocondrial por la acción de una posible selección purificadora, otras fuerzas evolutivas o procesos celulares como la mitofagia.

Todo esto conlleva a la determinación de que en el desarrollo del glioma cerebral se preserva el buen funcionamiento de la maquinaria mitocondrial y que muy pocas mutaciones podrían tener algún tipo de implicación en el proceso cancerígeno. La razón podría estar relacionada con la versatilidad metabólica, tal y como se ha postulado anteriormente (apartado 1.2.2. de la introducción). Aunque las alteraciones relacionadas con el proceso tumoral pueden redirigir el metabolismo de las células cancerígenas hacia la vía glicolítica, no se puede excluir la posible capacidad de las células del glioma de utilizar también la vía OXPHOS, debido a que han de sobrevivir en diferentes condiciones metabólicas porque pueden tener zonas de suministro sanguíneo heterogéneo (Feichtinger et al. 2014; Wanka et al. 2012). De hecho, es posible que en algunas zonas del tumor aumente la disponibilidad de oxígeno gracias a los vasos sanguíneos próximos al tumor o a los generados por la propia angiogénesis del tumor (Le et al. 2014). Incluso también se ha considerado que las mitocondrias de las regiones periféricas del tumor muestran una funcionalidad mitocondrial y una frecuencia respiratoria mayor que el centro del tumor (Duan et al. 2018a; Santandreu et al. 2008). Si bien el efecto Warburg considera que las células tumorales utilizan la vía glicolítica en vez de la oxidativa, cabe la posibilidad de que, a partir de la angiogénesis tumoral, se disponga de una mayor cantidad de oxígeno en vez de glucosa en zonas que conservan una adecuada irrigación sanguínea, generando la ya comentada reprogramación metabólica. Además, considerando que el cerebro es uno de los órganos que más oxígeno consume para obtener energía, también cabe la posibilidad de que se genere un ambiente donde se disponga de una mayor cantidad de oxígeno en comparación con la glucosa, o que la vía glicolítica no sea suficiente para proporcionar toda la energía necesaria para el tumor cerebral. Ahora bien, para que el tumor pueda

seguir progresando, sería necesario que no se generara una gran cantidad de ROS que conlleve a una situación de estrés oxidativo. En este sentido, si parte de las células tumorales son capaces de seguir utilizando OXPHOS, se puede suponer la necesidad de que no se acumulen mutaciones que puedan ser perjudiciales y que generen ROS en el ambiente. Bajo esta suposición, el proceso de mitofagia en las células cancerígenas podría eliminar aquellas mitocondrias que están muy dañadas y que generan un aumento de ROS inducido por las mutaciones generadas en su DNAm (Chourasia et al. 2015; Vara-Perez et al. 2019). No obstante, algunos autores consideran que el concepto de mitofagia dentro del glioma, concretamente del glioblastoma, es aún un concepto controvertido debido a la elevada heterogeneidad de procesos que intervienen en dicho tumor (Ulasov et al. 2018).

Asimismo, el grado de malignidad tumoral también puede generar diferencias en los gliomas cerebrales. Por un lado, los glioblastomas, es decir, los tumores más agresivos, se caracterizan por tener un rápido crecimiento en comparación a los tumores menos agresivos. Por otro lado, algunos autores consideran que pueden existir diferencias en el metabolismo tumoral dependiendo de la velocidad de su crecimiento (Jose et al. 2011). Relacionando este aspecto con los presentes resultados, los tumores de mayor grado de malignidad podrían utilizar más la vía glicolítica, debido a que la velocidad a la que crecen es superior en comparación a la cantidad de oxígeno de la que disponen, y los de menor grado de malignidad más la vía oxidativa. Se ha descrito que los glioblastomas tienen una menor cantidad de copias de DNAm y una menor actividad OXPHOS y, en consecuencia, una menor cantidad de ROS generado en comparación a los gliomas de menor grado de malignidad, circunstancia que podría ser debida a una remodelación y reducción de la densidad tumoral (Feichtinger et al. 2014; Soon et al. 2017). Esta menor cantidad de ROS generada por parte de los glioblastomas hace que en dicho tumor se prevenga la presencia y acumulación de mutaciones mitocondriales. En este sentido, se podría esperar que los tumores de grado II y III puedan llegar a acumular más mutaciones somáticas si se comparan con los tumores de grado IV. Nuestros resultados concuerdan con lo expuesto, ya que el porcentaje de mutaciones somáticas en tumores de menor grado de malignidad (grados II y III) es muy superior en comparación a los de mayor grado de malignidad (grado IV) (81,48% y 18,52%, respectivamente). Este resultado también concuerda con el hecho de que casi el 30% de los glioblastomas pertenecen a individuos de haplogrupo J, que se caracteriza por producir menor cantidad de energía y, por lo tanto, una menor generación de ROS.

Asimismo, la baja presencia de mutaciones patogénicas en el presente estudio también respaldaría la idea de que la no acumulación de mutaciones es imprescindible para que no se genere una cantidad de ROS elevada en el ambiente. En este sentido, otros estudios como el de Larman et al. (2012) determinan

que sólo 6 de los 32 glioblastomas que analizan (18,75%) contienen alguna mutación somática, siendo detectadas todas ellas, además, en heteroplasmia. Aunque los porcentajes de heteroplasmia detectados por Larman et al. (2012) son superiores respecto a los hallados en el presente estudio –varía entre el 30,2% y 85,1%–, los resultados podrían ser comparables en ambos. Además, tal y como comentan Vidone et al. (2015), las mutaciones potencialmente perjudiciales se detectan en heteroplasmia en la mayoría de casos de cáncer (Brandon et al. 2006; Iommarini et al. 2013; Larman et al. 2012), y en los estudios centrados en glioma no detectan o detectan una baja cantidad de mutaciones que podrían ser de interés en la enfermedad (Gasparre et al. 2007; Kirches et al. 2001; Larman et al. 2012; Yeung et al. 2014).

La presente tesis doctoral demuestra la importancia de la preservación de la maquinaria mitocondrial para el desarrollo tumoral, aportando una visión ampliada sobre la biogénesis del glioma cerebral. Además, el presente estudio es el primero en determinar asociaciones de los haplogrupos mitocondriales J y T que con un mayor riesgo de sufrir glioma. Estos haplogrupos son precisamente los que presentan una producción de ATP y de ROS más bajos. Por lo tanto, estudios adicionales, y con un mayor tamaño muestral, serían interesantes para poder corroborar los hallazgos encontrados en el presente estudio y comprender mejor el papel del DNAm en la aparición o desarrollo de la enfermedad.

7

CONSIDERACIONES FINALES

7. CONSIDERACIONES FINALES

Los numerosos estudios publicados hasta el momento sobre el DNAm han relacionado a este genoma con multitud de procesos celulares, así como con la aparición y el desarrollo de numerosas alteraciones y enfermedades. En este sentido, la implicación del DNAm en el cáncer, concretamente en el glioma, ha sido debatido y analizado por varios autores desde las últimas dos décadas debido a la implicación de este genoma en la respiración celular y en la generación de ROS dentro de un órgano, cuya cantidad de oxígeno necesaria es elevada para su correcto funcionamiento. Gracias a la aparición de la secuenciación masiva, el análisis más profundo del DNAm en este tipo de cáncer ha permitido obtener una mayor información, sobre todo a nivel de heteroplasmia mitocondrial. Sin embargo, la ausencia de un criterio único para determinar qué nivel de heteroplasmia es el adecuado para evitar aceptar tanto falsos positivos como negativos, dificulta la comparación entre diferentes estudios y el poder profundizar en el comportamiento mitocondrial. Así pues, uno de los objetivos de la presente tesis ha sido establecer un propio límite de detección de la heteroplasmia mitocondrial utilizando una metodología propia y aplicando una determinada tecnología de secuenciación a partir de la implementación de la tecnología Nextera XT[®] y plataforma MiSeq de Illumina. Previamente al establecimiento de un propio límite de detección, la valoración de la calidad de los archivos Fastq generados a partir de la secuenciación ha permitido determinar posibles sesgos introducidos por defecto debido a la metodología y tecnología utilizada tanto a nivel de amplificación, de la elaboración de las librerías o de la plataforma de secuenciación seleccionada. En este sentido, es importante saber discriminar qué sesgos pueden aparecer, por defecto, debido a estos factores, o qué sesgos pueden ser debidos a problemas de laboratorio. Además, los sesgos generados nos darán la información necesaria para saber qué filtros bioinformáticos será necesario aplicar posteriormente, a fin de evitar la generación resultados confusos. Así pues, con el límite establecido –mezclas de bases a partir de un 3%–, la presencia de falsos positivos y negativos parece estar ausente con una profundidad de lectura ~1000X. Asimismo, la detección de mezclas de bases con una frecuencia menor –entre 1,5% y 3%– requiere la obtención de una mayor profundidad de lectura (~3000X), confirmando el importante

papel que ejerce la profundidad de secuenciación en la detección de heteroplasmas. Por último, con la nueva estrategia presentada, no se pueden discriminar correctamente mezclas de bases con una frecuencia del alelo minoritario $\leq 1,5\%$, de manera que se puede estar aceptando variantes generadas por posibles errores durante el proceso de amplificación o de secuenciación (falsos positivos) o ignorando verdaderas mezclas de bases (falsos negativos). El establecimiento de unos propios límites de detección a partir de la generación de mezclas de bases ha permitido tomar como referencia el límite de detección del 3% para nuestro posterior estudio, en relación a la variabilidad heteroplásmica detectada en diferentes muestras de gliomas cerebrales. La detección de heteroplasmas a bajas frecuencias pueden llegar a ser de gran importancia porque permitiría en algunos casos el diagnóstico preventivo y el tratamiento de enfermedades que aún no han mostrado ninguna manifestación fenotípica (Duan et al. 2018b). Por lo tanto, es importante volver a remarcar la necesidad de mejorar la detección de la heteroplasma mitocondrial a todos los niveles para poder tener resultados mucho más precisos en el actual estudio.

Varios resultados obtenidos en el presente estudio mediante NGS son destacables referentes al glioma. Por un lado, aunque existen numerosos estudios de asociación entre haplogrupos mitocondriales y el riesgo de sufrir determinados tipos de cáncer, este trabajo sería el primero en establecer una asociación de los haplogrupos mitocondriales J y T y el glioma cerebral, ampliando, además, la información sobre la probabilidad de que los tumores puedan progresar a grados de mayor malignidad o no según el haplogrupo. Por otro lado, los resultados obtenidos respecto a la baja cantidad tanto de mutaciones sobrerrepresentadas en el estudio como de mutaciones en posiciones estables que pueden tener cierto impacto a nivel patogénico, y la baja carga mutacional detectada en los individuos afectados de glioma, sugieren una preferencia por parte del glioma de no acumular mutaciones en el DNAm con el fin de no generar una gran cantidad de ROS en el ambiente y poder progresar a estadios de mayor malignidad tumoral. Este comportamiento también se refleja en el aumento significativo de individuos heteroplásmicos afectados de glioma en comparación a individuos heteroplásmicos que no padecen cáncer y en el hecho de que el 100% de las mutaciones somáticas se manifiestan en heteroplasma, sugiriendo la posible existencia de fuerzas evolutivas o proceso celular que evite la fijación de mutaciones en el genoma mitocondrial. Respecto a este último punto, no se puede obviar el hecho de que la presencia de cierta heterogeneidad celular en los tumores, por ejemplo, por la formación de vasos sanguíneos por parte del tumor, pueda generar la presencia tanto de células transformadas como no transformadas y generar mezclas de diferentes genomas mitocondriales, llegando a generar posibles resultados imprecisos respecto al estado de las mutaciones detectadas.

En relación al nivel de malignidad tumoral, se ha observado un menor porcentaje de mutaciones somáticas en los tumores de mayor grado de malignidad (81,48%) respecto a los de menor grado de malignidad (18,52%), resultado que podría relacionarse con la idea de que los glioblastomas tienen una menor cantidad de copias de DNAmT y una menor actividad OXPHOS y, en consecuencia, una menor cantidad de ROS generado en comparación a los gliomas de menor grado de malignidad, circunstancia que podría ser debida a la remodelación y reducción de la densidad tumoral (Feichtinger et al. 2014; Soon et al. 2017). En este sentido, el análisis adicional del contenido mitocondrial en los gliomas sería de gran interés para profundizar en el comportamiento de los gliomas más agresivos. Todos estos resultados se pueden relacionar con el hecho de que parte de las células tumorales en glioma pueden llevar a cabo una reprogramación metabólica, y que la actividad OXPHOS por parte de estas células conlleve la necesidad de preservar la maquinaria mitocondrial y generar un ambiente idóneo para el desarrollo tumoral.

Por lo tanto, el presente trabajo ha permitido ampliar los conocimientos respecto a la detección de la heteroplasmia mitocondrial mediante la tecnología de NGS y el papel del DNAmT en el glioma cerebral, y ha generado nuevas hipótesis para seguir avanzando en el estudio y ampliar los conocimientos descritos hasta el momento sobre la etiología de dicha enfermedad. Se ha planteado la importancia de la preservación de la maquinaria mitocondrial para el desarrollo tumoral, aportando una visión ampliada sobre la biogénesis del glioma cerebral. Asimismo, los resultados obtenidos han generado un mayor interés en seguir trabajando bajo esta línea de investigación aumentando el número de muestras de tumores con diferentes grados de malignidad y realizando análisis adicionales como, por ejemplo, el análisis del contenido mitocondrial o la actividad OXPHOS.

No obstante, el presente trabajo también presenta varias limitaciones que se han de tener en cuenta. La elevada dificultad presentada en la amplificación del DNA mitocondrial del tumor, debido posiblemente a la elevada degradación del DNA, ha conllevado a la necesidad de reducir el número de muestras, así como considerar cualquier parte del tumor como un único tejido. Para aquellas muestras en las que no se obtuvo un buen resultado a nivel bioinformático, se tuvo que volver a extraer el DNA y repetir todo el proceso, ya que los errores detectados se podían haber introducido en cualquier punto de la metodología llevada a cabo. Aun así, algunas muestras seguían sin generar buenos resultados y tuvieron que ser excluidas finalmente del estudio. Estos factores han hecho que se tuviera que reducir el tamaño muestral considerablemente y excluir la posibilidad de obtener una diferenciación intratumoral. Por otro lado, aún y obteniendo material amplificado, la cantidad de DNA mitocondrial no fue suficiente en la mayoría de casos para utilizar cualquier tipo de kit para la construcción de las

librerías. Así, nos tuvimos que limitar al uso del kit Nextera[®], que requería una cantidad menor de partida (1ng de DNA). Asimismo, la dificultad de poder obtener muestras control de tejido cerebral sano supone un hándicap para poder comparar resultados entre muestras tumorales y no tumorales. No obstante, este factor se ha podido solventar mediante la obtención de datos extraídos del trabajo de Li et al. (2015a). Por último, respecto a las mutaciones detectadas en las muestras de glioma, los análisis *in-silico* realizados no permiten conocer el impacto que pueden llegar a tener en la enfermedad en profundidad y, por lo tanto, la realización de estudios funcionales de la variación mitocondrial detectada puede ser un objetivo muy interesante para futuros estudios. Del mismo modo, la relación de los resultados obtenidos con ciertas características del metabolismo celular del glioma está basada en evidencias meramente teóricas. No obstante, estas posibles evidencias han generado un gran interés por llevar a cabo estudios basados en el análisis del número de copias de DNAm^t y de la actividad OXPHOS en las células del glioma en un futuro.

8

CONCLUSIONES

8. CONCLUSIONES

1. El análisis de los archivos Fastq, obtenidos a partir de la NGS, demuestra que la obtención de una buena calidad depende de muchos factores tales como la naturaleza del genoma analizado, la realización de una PCR previa, la tecnología implementada para la elaboración de las librerías o la plataforma de secuenciación utilizada. La obtención de una menor calidad en algunos parámetros analizados no siempre implica que generen unos datos poco útiles para posteriores análisis y pueden dar una idea sobre la profundidad de lectura que se podrá obtener una vez se apliquen los filtros de calidad establecidos para el análisis bioinformático.
2. La secuenciación de las quimeras mediante NGS muestran la elevada dificultad de poder detectar mezclas de bases a bajas frecuencias. Todas las mezclas de bases que han sido detectadas a una baja frecuencia son difíciles de reproducir, no sólo a nivel de *runs*, sino también según el tipo de alineamiento seleccionado. Asimismo, se ha demostrado que la profundidad de lectura obtenida ejerce un papel clave para evitar interpretaciones erróneas en la detección de la heteroplasmia mitocondrial a bajas frecuencias. En este sentido, es muy importante generar el máximo de lecturas que superen los filtros de calidad establecidos, a fin de obtener una profundidad de lectura suficientemente alta para poder detectar heteroplasmas de forma fiable.
3. Se ha establecido una estrategia alternativa para establecer unos límites de detección de la heteroplasmia mitocondrial usando la tecnología Nextera XT® y la plataforma MiSeq (Illumina). Siguiendo esta estrategia, se requiere una frecuencia mínima del alelo minoritario del 3% y una profundidad de lectura mínima de 1000X para garantizar que tanto los falsos positivos como negativos están prácticamente ausentes. Sin embargo, cuando la frecuencia del alelo minoritario presenta valores ligeramente más bajos (entre el

1,5% y el 3%), es necesario una profundidad de lectura mayor (~3000X) para no detectar, *a priori*, ni falsos positivos ni negativos. Por último, las mezclas de bases observadas a bajas frecuencias (<1,5%) no siempre se pueden detectar en todos los *runs*, independientemente de la profundidad obtenida.

4. A nivel de la heteroplasmia mitocondrial, cuando el análisis incluye la amplificación previa del DNAMt, es necesario llevar a cabo un diseño exhaustivo de los *primers* para evitar resultados equivocados debido a la co-amplificación de NUMTs.
5. Se ha determinado que los individuos con haplogrupos mitocondriales J y T tienen un riesgo significativamente mayor de sufrir glioma. Estos haplogrupos poseen determinados polimorfismos que provocan un desacoplamiento de la cadena respiratoria como medida de adaptación al entorno y, por lo tanto, generan una menor cantidad de ROS y evitan situaciones de estrés oxidativo para la célula. Esta adaptación puede generar, a su vez, un ambiente favorable para el desarrollo tumoral. El presente estudio es el primero en establecer una asociación de haplogrupos mitocondriales con este tipo de cáncer.
6. En relación al haplogrupo mitocondrial J, se ha observado una mayor frecuencia de individuos con glioblastoma respecto a individuos con gliomas menos agresivos y respecto a la población general. La generación de una menor cantidad de ROS por parte este haplogrupo permite que los propios agentes antioxidantes de las células tumorales sean capaces de eliminarlo de forma eficaz, generando así un ambiente favorable para que los tumores puedan alcanzar mayor grado de malignidad. Así pues, el haplogrupo J podría considerarse un indicador de mal pronóstico en cuanto a la agresividad tumoral.
7. En relación al haplogrupo mitocondrial H, se considera que las variantes que definen este haplogrupo pueden generar una mayor cantidad de ROS en el ambiente y dar lugar a una situación de estrés oxidativo. Esta situación puede provocar problemas al tumor para poder eliminar por sí mismo todo el ROS generado y, por lo tanto, tener más problemas para poder progresar. Así pues, se puede considerar que los individuos con haplogrupo H podrían tener cierta ventaja para evitar la progresión del tumor cerebral a grados de mayor malignidad.

8. El 95,02% de las mutaciones detectadas son de origen germinal. La baja cantidad de mutaciones sobrerrepresentadas en el estudio respecto a la población general y de mutaciones que podrían tener algún tipo de implicación en el proceso cancerígeno, sugiere una preferencia por parte del tumor de no acumular mutaciones en el DNAm. Además, la carga mutacional de los individuos afectados de glioma tiende a ser baja y compuesta por variantes comunes en la población general, tanto a nivel general como analizando los diferentes grados de malignidad tumoral. Estos resultados plantean la idea de que, en el glioma cerebral, se preserva el buen funcionamiento de la maquinaria mitocondrial.
9. A nivel somático, el 100% de las mutaciones se han detectado en heteroplasmia. Sin embargo, no se puede obviar el hecho de que la ausencia de mutaciones somáticas fijadas en el tumor podría ser debido a la heterogeneidad celular generada en el momento de seleccionar exclusivamente tejido tumoral. Del mismo modo, el proceso de angiogénesis llevado a cabo por el tumor también dificultaría una perfecta discriminación y selección de células tumorales.
10. El porcentaje de mutaciones somáticas en los tumores de menor grado de malignidad (grados II y III) es muy superior en comparación a los de mayor grado de malignidad (grado IV). Este resultado se puede relacionar con el hecho de que los glioblastomas tienen una menor cantidad de copias de DNAm y una menor actividad OXPHOS y, en consecuencia, una menor cantidad de ROS generado en comparación a los gliomas de menor grado de malignidad.
11. El aumento significativo de individuos heteroplásmicos afectados de glioma respecto a los que no padecen cáncer indicaría que, durante el proceso cancerígeno, se observaría un incremento en la generación de mutaciones, pero que posiblemente se intente evitar que se fijen en el genoma mitocondrial por la acción de una posible selección purificadora u otros procesos evolutivos y celulares, como la mitofagia.
12. Los resultados obtenidos plantean la posibilidad de que las células del glioma cerebral sean capaces de utilizar también la vía OXPHOS para obtener energía en situaciones donde se disponga una mayor cantidad de oxígeno o una menor cantidad de glucosa. Esta situación respaldaría la idea de que la no acumulación de mutaciones es imprescindible para que no se genere una cantidad de ROS suficientemente alta y se preserve la integridad de la maquinaria mitocondrial.

9

BIBLIOGRAFÍA

9. BIBLIOGRAFÍA

- Aikhionbare FO, Khan M, Carey D, Okoli J, Go R (2004) Is cumulative frequency of mitochondrial DNA variants a biomarker for colorectal tumor progression? *Mol Cancer* 3: 30.
- Aikhionbare FO, Mehrabi S, Kumaresan K, Zavareh M, Olatinwo M, Odunsi K, Partridge E (2007) Mitochondrial DNA sequence variants in epithelial ovarian tumor subtypes and stages. *J Carcinog* 6: 1.
- Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* 12: R18.
- Akouchekian M, Houshmand M, Akbari MH, Kamalidehghan B, Dehghan M (2011) Analysis of mitochondrial ND1 gene in human colorectal cancer. *J Res Med Sci* 16: 50-5.
- Akouchekian M, Houshmand M, Hemati S, Ansari-pour M, Shafa M (2009) High rate of mutation in mitochondrial DNA displacement loop region in human colorectal cancer. *Dis Colon Rectum* 52: 526-30.
- Albayrak L, Khanipov K, Pimenova M, Golovko G, Rojas M, Pavlidis I, Chumakov S, Aguilar G, Chavez A, Widger WR, Fofanov Y (2016) The ability of human nuclear DNA to cause false positive low-abundance heteroplasmy calls varies across the mitochondrial genome. *BMC Genomics* 17: 1017.
- Aldamiz-Echevarria T, Resino S, Bellon JM, Jimenez-Sousa MA, Miralles P, Medrano LM, Carrero A, Diez C, Perez-Latorre L, Fanciulli C, Garcia-Broncano P, Berenguer J (2019) European mitochondrial haplogroups predict liver-related outcomes in patients coinfecting with HIV and HCV: a retrospective study. *J Transl Med* 17: 244. doi: 10.1186/s12967-019-1997-x
- Allegra E, Garozzo A, Lombardo N, De Clemente M, Carey TE (2006) Mutations and polymorphisms in mitochondrial DNA in head and neck cancer cell lines. *Acta Otorhinolaryngol Ital* 26: 185-90.
- Allen N, Barres B (2009) Glia — more than just brain glue. *Nature* 457: 675-677.
- American Brain Tumor Association (2019a) Ependymoma. https://www.abta.org/tumor_types/ependymoma/. Accessed 2015

- American Brain Tumor Association (2019b) Glioblastoma (GBM). https://www.abta.org/tumor_types/glioblastoma-gbm/. Accessed 2015
- American Brain Tumor Association (2019c) Glioma. https://www.abta.org/tumor_types/glioma/. Accessed 2015
- American Brain Tumor Association (2019d) Oligoastrocytoma. https://www.abta.org/tumor_types/oligoastrocytoma/. Accessed 2015
- Andrews RM, Kubacka I, Chinnery PF, Lightowers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147. doi: 10.1038/13779
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Arning L, Haghikia A, Taherzadeh-Fard E, Saft C, Andrich J, Pula B, Hoxtermann S, Wiczorek S, Akkad DA, Perrech M, Gold R, Epplen JT, Chan A (2010) Mitochondrial haplogroup H correlates with ATP levels and age at onset in Huntington disease. *J Mol Med (Berl)* 88: 431-6. doi: 10.1007/s00109-010-0589-2
- Arnold RS, Fedewa SA, Goodman M, Osunkoya AO, Kissick HT, Morrissey C, True LD, Petros JA (2015) Bone metastasis in prostate cancer: Recurring mitochondrial DNA mutation reveals selective pressure exerted by the bone microenvironment. *Bone* 78: 81-6. doi:10.1016/j.bone.2015.04.046
- Arsenic R, Treue D, Lehmann A, Hummel M, Dietel M, Denkert C, Budczies J (2015) Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer. *BMC Clin Pathol* 15: 20.
- Arstad C, Refinetti P, Kristensen AT, Giercksky KE, Ekstrom PO (2017) Is detection of intraperitoneal exfoliated tumor cells after surgical resection of rectal cancer a prognostic factor of survival? *BMC Cancer* 17: 406. doi: 10.1186/s12885-017-3365-7
- Ason B, Reznikoff WS (2004) DNA sequence bias during Tn5 transposition. *J Mol Biol* 335: 1213-25.
- Attardi G (2002) Role of mitochondrial DNA in human aging. *Mitochondrion* 2: 27-37.
- Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. *Nature* 526: 68-74.
- Bai RK, Leal SM, Covarrubias D, Liu A, Wong LJ (2007) Mitochondrial genetic background modifies breast cancer risk. *Cancer Res* 67: 4687-94.
- Bandelt HJ, Salas A (2009) Contamination and sample mix-up can best explain some patterns of mtDNA instabilities in buccal cells and oral squamous cell carcinoma. *BMC Cancer* 9: 113.
- Bansal V (2017) A computational method for estimating the PCR duplication rate in DNA and RNA-seq experiments. *BMC Bioinformatics* 18: 43. doi: 10.1186/s12859-017-1471-9

- Barbieri E, Scorrano L (2013) When numbers matters: mitochondrial DNA and gliomagenesis. *Cell Death Differ* 20: 1601-2.
- Barral-Arca R, Pischedda S, Gomez-Carballa A, Pastoriza A, Mosquera-Miguel A, Lopez-Soto M, Martinon-Torres F, Alvarez-Iglesias V, Salas A (2016) Meta-Analysis of Mitochondrial DNA Variation in the Iberian Peninsula. *PLoS One* 11: e0159735.
- Basso D, Navaglia F, Fogar P, Zambon CF, Greco E, Schiavon S, Fasolo M, Stranges A, Falda A, Padoan A, Fadi E, Pedrazzoli S, Plebani M (2007) DNA repair pathways and mitochondrial DNA mutations in gastrointestinal carcinogenesis. *Clin Chim Acta* 381: 50-5.
- Benjamini Y, Speed TP (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res* 40: e72.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. *Nucleic Acids Res* 41: D36-42. doi: 10.1093/nar/gks1195
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara ECM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53-9.
- Berglund EC, Kiialainen A, Syvanen AC (2011) Next-generation sequencing technologies and applications for human genetic history and forensics. *Investig Genet* 2: 23.
- Betarbet R, Sherer TB, MacKenzie G, Garcia-Osuna M, Panov AV, Greenamyre JT (2000) Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nat Neurosci* 3: 1301-6. doi: 10.1038/81834
- Booker LM, Habermacher GM, Jessie BC, Sun QC, Baumann AK, Amin M, Lim SD, Fernandez-Golarz C, Lyles RH, Brown MD, Marshall FF, Petros JA (2006) North American white mitochondrial haplogroups in prostate and renal cancer. *J Urol* 175: 468-72; discussion 472-3.

- Bottje WG (2018) Oxidative metabolism and efficiency: the delicate balancing act of mitochondria. *Poult Sci.* doi: 10.3382/ps/pey405
- Bragoszewski P, Kupryjanczyk J, Bartnik E, Rachinger A, Ostrowski J (2008) Limited clinical relevance of mitochondrial DNA mutation and gene expression analyses in ovarian cancer. *BMC Cancer* 8: 292.
- Brandon M, Baldi P, Wallace DC (2006) Mitochondrial mutations in cancer. *Oncogene* 25: 4647–62.
- Brandon MC, Ruiz-Pesini E, Mishmar D, Procaccio V, Lott MT, Nguyen KC, Spolim S, Patil U, Baldi P, Wallace DC (2009) MITOMASTER: a bioinformatics tool for the analysis of mitochondrial DNA sequences. *Hum Mutat* 30: 1–6.
- Brisighelli F, Alvarez-Iglesias V, Fondevila M, Blanco-Verea A, Carracedo A, Pascali VL, Capelli C, Salas A (2012) Uniparental markers of contemporary Italian population reveals details on its pre-Roman heritage. *PLoS One* 7: e50794.
- Bruskov VI, Malakhova LV, Masalimov ZK, Chernikov AV (2002) Heat-induced formation of reactive oxygen species and 8-oxoguanine, a biomarker of damage to DNA. *Nucleic Acids Res* 30: 1354–63.
- Bussard KM, Siracusa LD (2017) Understanding Mitochondrial Polymorphisms in Cancer. *Cancer Res* 77: 6051–6059.
- Cai FF, Kohler C, Zhang B, Chen WJ, Barekati Z, Garritsen HS, Lenner P, Toniolo P, Zhang JJ, Zhong XY (2011) Mutations of mitochondrial DNA as potential biomarkers in breast cancer. *Anticancer Res* 31: 4267–71.
- Calabrese C, Simone D, Diroma MA, Santorsola M, Gutta C, Gasparre G, Picardi E, Pesole G, Attimonelli M (2014) MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. *Bioinformatics* 30: 3115–7.
- Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325: 31–35.
- Canter JA, Kallianpur AR, Parl FF, Millikan RC (2005) Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res* 65: 8028–33.
- Carew JS, Huang P (2002) Mitochondrial defects in cancer. *Mol Cancer* 1: 9.
- Carew JS, Zhou Y, Albitar M, Carew JD, Keating MJ, Huang P (2003) Mitochondrial DNA mutations in primary leukemia cells after chemotherapy: clinical significance and therapeutic implications. *Leukemia* 17: 1437–47.
- Cerezo M, Bandelt HJ, Martin-Guerrero I, Ardanaz M, Vega A, Carracedo A, Garcia-Orad A, Salas A (2009) High mitochondrial DNA stability in B-cell chronic lymphocytic leukemia. *PLoS One* 4: e7902.

- Clima R, Preste R, Calabrese C, Diroma MA, Santorsola M, Scioscia G, Simone D, Shen L, Gasparre G, Attimonelli M (2017) HmtDB 2016: data update, a better performing query system and human mitochondrial DNA haplogroup predictor. *Nucleic Acids Res* 45: D698–D706.
- Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM (2010) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res* 38: 1767–71.
- Cocos R, Schipor S, Badiu C, Raicu F (2018) Mitochondrial DNA haplogroup K as a contributor to protection against thyroid cancer in a population from southeast Europe. *Mitochondrion* 39: 43–50.
- Coller HA, Khrapko K, Bodyak ND, Nekhaeva E, Herrero-Jimenez P, Thilly WG (2001) High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. *Nat Genet* 28: 147–50.
- Corral-Debrinski M, Horton T, Lott MT, Shoffner JM, Beal MF, Wallace DC (1992) Mitochondrial DNA deletions in human brain: regional variability and increase with advanced age. *Nat Genet* 2: 324–9.
- Correia RL, Oba-Shinjo SM, Uno M, Huang N, Marie SK (2011) Mitochondrial DNA depletion and its correlation with TFAM, TFB1M, TFB2M and POLG in human diffusely infiltrating astrocytomas. *Mitochondrion* 11: 48–53.
- Cortopassi GA, Shibata D, Soong NW, Arnheim N (1992) A pattern of accumulation of a somatic deletion of mitochondrial DNA in aging human tissues. *Proc Natl Acad Sci U S A* 89: 7370–4.
- Coskun PE, Beal MF, Wallace DC (2004) Alzheimer's brains harbor somatic mtDNA control-region mutations that suppress mitochondrial transcription and replication. *Proc Natl Acad Sci U S A* 101: 10726–31.
- Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, Fostel JL, Friedrich DC, Perrin D, Dionne D, Kim S, Gabriel SB, Lander ES, Fisher S, Getz G (2013) Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res* 41: e67.
- Covarrubias D, Bai RK, Wong LJ, Leal SM (2008) Mitochondrial DNA variant interactions modify breast cancer risk. *J Hum Genet* 53: 924–8.
- Cui H, Huang P, Wang Z, Zhang Y, Zhang Z, Xu W, Wang X, Han Y, Guo X (2013) Association of decreased mitochondrial DNA content with the progression of colorectal cancer. *BMC Cancer* 13: 110.
- Czarnecka AM, Klemba A, Semczuk A, Plak K, Marzec B, Krawczyk T, Kofler B, Golik P, Bartnik E (2009) Common mitochondrial polymorphisms as risk factor for endometrial cancer. *Int Arch Med* 2: 33.

- Czarnecka AM, Krawczyk T, Plak K, Klemba A, Zdrozny M, Arnold RS, Kofler B, Golik P, Szybinska A, Lubinski J, Mossakowska M, Bartnik E, Petros JA (2010) Mitochondrial genotype and breast cancer predisposition. *Oncol Rep* 24: 1521-34.
- Chamberlain MC (2003) Ependymomas. *Curr Neurol Neurosci Rep* 3: 193-9.
- Chamkha I, Mkaouar-Rebai E, Aloulou H, Chabchoub I, Kifagi C, Fendri-Kriaa N, Kammoun T, Hachicha M, Fakhfakh F (2011) A novel m.3395A>G missense mutation in the mitochondrial ND1 gene associated with the new tRNA(Ile) m.4316A>G mutation in a patient with hypertrophic cardiomyopathy and profound hearing loss. *Biochem Biophys Res Commun* 404: 504-10. doi: 10.1016/j.bbrc.2010.12.012
- Chang SC, Lin PC, Yang SH, Wang HS, Liang WY, Lin JK (2009) Mitochondrial D-loop mutation is a common event in colorectal cancers with p53 mutations. *Int J Colorectal Dis* 24: 623-8.
- Chattopadhyay E, De Sarkar N, Singh R, Ray A, Roy R, Paul RR, Pal M, Ghose S, Ghosh S, Kabiraj D, Banerjee R, Roy B (2016) Genome-wide mitochondrial DNA sequence variations and lower expression of OXPHOS genes predict mitochondrial dysfunction in oral cancer tissue. *Tumour Biol* 37: 11861-11871. doi: 10.1007/s13277-016-5026-x
- Chen C, Ba Y, Li D, Du X, Lia X, Yang H, An J, Xing J, Yang H, Dong G, Guo X (2017) Genetic variations of mitochondrial genome modify risk and prognosis of hepatocellular carcinoma patients. *Clin Res Hepatol Gastroenterol* 41: 378-385.
- Chen JZ, Gokden N, Greene GF, Mukunyadzi P, Kadlubar FF (2002) Extensive somatic mitochondrial mutations in primary prostate cancer using laser capture microdissection. *Cancer Res* 62: 6470-4.
- Chen T, He J, Shen L, Fang H, Nie H, Jin T, Wei X, Xin Y, Jiang Y, Li H, Chen G, Lu J, Bai Y (2011) The mitochondrial DNA 4,977-bp deletion and its implication in copy number alteration in colorectal cancer. *BMC Med Genet* 12: 8.
- Chen XJ (2013) Mechanism of homologous recombination and implications for aging-related deletions in mitochondrial DNA. *Microbiol Mol Biol Rev* 77: 476-96.
- Chen XZ, Fang Y, Shi YH, Cui JH, Li LY, Xu YC, Ling B (2016) Mitochondrial D310 instability in Chinese lung cancer patients. *Mitochondrial DNA A DNA Mapp Seq Anal* 27: 1177-80. doi: 10.3109/19401736.2014.936426
- Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC (2013) Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One* 8: e62856.
- Chinnery PF, Hudson G (2013) Mitochondrial genetics. *Br Med Bull* 106: 135-59.
- Chinnery PF, Samuels DC, Elson J, Turnbull DM (2002) Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism? *Lancet* 360: 1323-5.

- Chinnery PF, Thorburn DR, Samuels DC, White SL, Dahl HM, Turnbull DM, Lightowlers RN, Howell N (2000) The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? *Trends Genet* 16: 500-5.
- Chourasia AH, Boland ML, Macleod KF (2015) Mitophagy and cancer. *Cancer Metab* 3: 4. doi: 10.1186/s40170-015-0130-8
- Dai JG, Zhang ZY, Liu QX, Min JX (2013) Mitochondrial genome microsatellite instability and copy number alteration in lung carcinomas. *Asian Pac J Cancer Prev* 14: 2393-9.
- Darvishi K, Sharma S, Bhat AK, Rai E, Bamezai RN (2007) Mitochondrial DNA G10398A polymorphism imparts maternal Haplogroup N a risk for breast and esophageal cancer. *Cancer Lett* 249: 249-55.
- Dasgupta S, Koch R, Westra WH, Califano JA, Ha PK, Sidransky D, Koch WM (2010) Mitochondrial DNA mutation in normal margins and tumors of recurrent head and neck squamous cell carcinoma patients. *Cancer Prev Res (Phila)* 3: 1205-11.
- Dasgupta S, Soudry E, Mukhopadhyay N, Shao C, Yee J, Lam S, Lam W, Zhang W, Gazdar AF, Fisher PB, Sidransky D (2012) Mitochondrial DNA mutations in respiratory complex-I in never-smoker lung cancer patients contribute to lung cancer progression and associated with EGFR gene mutation. *J Cell Physiol* 227: 2451-60.
- Dasgupta S, Yung RC, Westra WH, Rini DA, Brandes J, Sidransky D (2009) Following mitochondrial footprints through a long mucosal path to lung cancer. *PLoS One* 4: e6533.
- Datta S, Chattopadhyay E, Ray JG, Majumder M, Roy PD, Roy B (2015) D-loop somatic mutations and approximately 5 kb "common" deletion in mitochondrial DNA: important molecular markers to distinguish oral precancer and cancer. *Tumour Biol* 36: 3025-33. doi: 10.1007/s13277-014-2937-2
- Davis CF, Ricketts CJ, Wang M, Yang L, Cherniack AD, Shen H, Buhay C, Kang H, Kim SC, Fahey CC, Hacker KE, Bhanot G, Gordenin DA, Chu A, Gunaratne PH, Biehl M, Seth S, Kaiparettu BA, Bristow CA, Donehower LA, Wallen EM, Smith AB, Tickoo SK, Tamboli P, Reuter V, Schmidt LS, Hsieh JJ, Choueiri TK, Hakimi AA, The Cancer Genome Atlas Research N, Chin L, Meyerson M, Kucherlapati R, Park WY, Robertson AG, Laird PW, Henske EP, Kwiatkowski DJ, Park PJ, Morgan M, Shuch B, Muzny D, Wheeler DA, Linehan WM, Gibbs RA, Rathmell WK, Creighton CJ (2014) The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* 26: 319-330. doi: 10.1016/j.ccr.2014.07.014
- Dayama G, Emery SB, Kidd JM, Mills RE (2014) The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res* 42: 12640-9.

- De Vitto H, Mendonca BS, Elseth KM, Vesper BJ, Portari EA, Gallo CV, Paradise WA, Rumjanek FD, Radosevich JA (2013) Part II. Mitochondrial mutational status of high nitric oxide adapted cell line BT-20 (BT-20-HNO) as it relates to human primary breast tumors. *Tumour Biol* 34: 337-47.
- DeHaan C, Habibi-Nazhad B, Yan E, Salloum N, Parliament M, Allalunis-Turner J (2004) Mutation in mitochondrial complex I ND6 subunit is associated with defective response to hypoxia in human glioma cells. *Mol Cancer* 3: 19.
- Del Bo R, Bordoni A, Sciacco M, Di Fonzo A, Galbiati S, Crimi M, Bresolin N, Comi GP (2003) Remarkable infidelity of polymerase gammaA associated with mutations in POLG1 exonuclease domain. *Neurology* 61: 903-8. doi: 10.1212/01.wnl.0000092303.13864.be
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytzky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43: 491-8.
- Devic S (2016) Warburg Effect - a Consequence or the Cause of Carcinogenesis? *J Cancer* 7: 817-22.
- Dickins B, Rebolledo-Jaramillo B, Su MS, Paul IM, Blankenberg D, Stoler N, Makova KD, Nekrutenko A (2014) Controlling for contamination in re-sequencing studies with a reproducible web-based phylogenetic approach. *Biotechniques* 56: 134-6, 138-41.
- Diroma MA, Calabrese C, Simone D, Santorsola M, Calabrese FM, Gasparre G, Attimonelli M (2014) Extraction and annotation of human mitochondrial genomes from 1000 Genomes Whole Exome Sequencing data. *BMC Genomics* 15 Suppl 3: S2.
- Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* 36: e105.
- Druzhyna NM, Wilson GL, LeDoux SP (2008) Mitochondrial DNA repair in aging and disease. *Mech Ageing Dev* 129: 383-90.
- Duan K, Liu ZJ, Hu SQ, Huo HY, Xu ZR, Ruan JF, Sun Y, Dai LP, Yan CB, Xiong W, Cui QH, Yu HJ, Yu M, Qin Y (2018a) Lactic acid induces lactate transport and glycolysis/OXPHOS interconversion in glioblastoma. *Biochem Biophys Res Commun* 503: 888-894. doi: 10.1016/j.bbrc.2018.06.092
- Duan M, Tu J, Lu Z (2018b) Recent Advances in Detecting Mitochondrial DNA Heteroplasmic Variations. *Molecules* 23.
- Duberow DP, Brait M, Hoque MO, Theodorescu D, Sidransky D, Dasgupta S, Mathies RA (2016) High-performance detection of somatic D-loop mutation in urothelial cell carcinoma patients by polymorphism ratio sequencing. *J Mol Med (Berl)* 94: 1015-24. doi: 10.1007/s00109-016-1407-2

- Ebner S, Mangge H, Langhof H, Halle M, Siegrist M, Aigner E, Paulmichl K, Paulweber B, Datz C, Sperl W, Kofler B, Weghuber D (2015) Mitochondrial Haplogroup T Is Associated with Obesity in Austrian Juveniles and Adults. *PLoS One* 10: e0135622.
- Eklblom R, Smeds L, Ellegren H (2014) Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. *BMC Genomics* 15: 467.
- Elson JL, Turnbull DM, Howell N (2004) Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am J Hum Genet* 74: 229-38. doi: 10.1086/381505
- Elliott HR, Samuels DC, Eden JA, Relton CL, Chinnery PF (2008) Pathogenic mitochondrial DNA mutations are common in the general population. *Am J Hum Genet* 83: 254-60.
- Engelhard HH, Stelea A, Cochran EJ (2002) Oligodendroglioma: pathology and molecular biology. *Surg Neurol* 58: 111-7; discussion 117.
- Engelhard HH, Stelea A, Mundt A (2003) Oligodendroglioma and anaplastic oligodendroglioma: clinical features, treatment, and prognosis. *Surg Neurol* 60: 443-56.
- Ericson NG, Kulawiec M, Vermulst M, Sheahan K, O'Sullivan J, Salk JJ, Bielas JH (2012) Decreased mitochondrial DNA mutagenesis in human colorectal cancer. *PLoS Genet* 8: e1002689.
- Eyre-Walker A, Awadalla P (2001) Does human mtDNA recombine? *J Mol Evol* 53: 430-5.
- Falah M, Farhadi M, Kamrava SK, Mahmoudian S, Daneshi A, Balali M, Asghari A, Houshmand M (2017) Association of genetic variations in the mitochondrial DNA control region with presbycusis. *Clin Interv Aging* 12: 459-465. doi: 10.2147/CIA.S123278
- Falkenberg M (2018) Mitochondrial DNA replication in mammalian cells: overview of the pathway. *Essays Biochem* 62: 287-296.
- Fang H, Shen L, Chen T, He J, Ding Z, Wei J, Qu J, Chen G, Lu J, Bai Y (2010) Cancer type-specific modulation of mitochondrial haplogroups in breast, colorectal and thyroid cancer. *BMC Cancer* 10: 421.
- Fang Y, Huang J, Zhang J, Wang J, Qiao F, Chen HM, Hong ZP (2013) Detecting the somatic mutations spectrum of Chinese lung cancer by analyzing the whole mitochondrial DNA genomes. *Mitochondrial DNA*.
- Fang Y, Yang HY, Shi YH, Cui JH, Li LY, Xu YC, Shao JL (2015) Mitochondrial DNA haplogroups and somatic mutations are associated with lung cancer in patients from Southwest China. *Genet Mol Res* 14: 5031-43. doi: 10.4238/2015.May.12.6
- Fauser S, Luberichs J, Besch D, Leo-Kottler B (2002) Sequence analysis of the complete mitochondrial genome in patients with Leber's hereditary optic neuropathy lacking the three most common pathogenic DNA mutations. *Biochem Biophys Res Commun* 295: 342-7.

- Fayed-Miguel N, Morales-Ramos H, Modrego-Pardo PJ (2006) [Magnetic resonance imaging with spectroscopy, perfusion and cerebral diffusion in the diagnosis of brain tumours]. *Rev Neurol* 42: 735-42.
- Feichtinger RG, Weis S, Mayr JA, Zimmermann F, Geilberger R, Sperl W, Kofler B (2014) Alterations of oxidative phosphorylation complexes in astrocytomas. *Glia* 62: 514-25.
- Fendt L, Niederstatter H, Huber G, Zelger B, Dunser M, Seifarth C, Rock A, Schafer G, Klocker H, Parson W (2011) Accumulation of mutations over the entire mitochondrial genome of breast cancer cells obtained by tissue microdissection. *Breast Cancer Res Treat* 128: 327-36.
- Ferragina P, Manzini G (2000) Opportunistic Data Structures with Applications. *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*: 390.
- Filosto M, Mancuso M, Vives-Bauza C, Vila MR, Shanske S, Hirano M, Andreu AL, DiMauro S (2003) Lack of paternal inheritance of muscle mitochondrial DNA in sporadic mitochondrial myopathies. *Ann Neurol* 54: 524-6.
- Fliss MS, Usadel H, Caballero OL, Wu L, Buta MR, Eleff SM, Jen J, Sidransky D (2000) Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. *Science* 287: 2017-9.
- Gasparre G, Porcelli AM, Bonora E, Pennisi LF, Toller M, Iommarini L, Ghelli A, Moretti M, Betts CM, Martinelli GN, Ceroni AR, Curcio F, Carelli V, Rugolo M, Tallini G, Romeo G (2007) Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors. *Proc Natl Acad Sci U S A* 104: 9001-6.
- Ghaffarpour M, Mahdian R, Fereidooni F, Kamalidehghan B, Moazami N, Houshmand M (2014) The mitochondrial ATPase6 gene is more susceptible to mutation than the ATPase8 gene in breast cancer patients. *Cancer Cell Int* 14: 21. doi: 10.1186/1475-2867-14-21
- Gochhait S, Bhatt A, Sharma S, Singh YP, Gupta P, Bamezai RN (2008) Concomitant presence of mutations in mitochondrial genome and p53 in cancer development - a study in north Indian sporadic breast and esophageal cancer patients. *Int J Cancer* 123: 2580-6.
- Goios A, Prieto L, Amorim A, Pereira L (2008) Specificity of mtDNA-directed PCR-influence of Nuclear MTDNA insertion (NUMT) contamination in routine samples and techniques. *Int J Legal Med* 122: 341-5.
- Goto H, Dickins B, Afgan E, Paul IM, Taylor J, Makova KD, Nekrutenko A (2011) Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biol* 12: R59.
- Grada A, Weinbrecht K (2013) Next-generation sequencing: methodology and application. *J Invest Dermatol* 133: e11.
- Gray MW (1999) Evolution of organellar genomes. *Curr Opin Genet Dev* 9: 678-87.

- Gray MW, Doolittle WF (1982) Has the endosymbiont hypothesis been proven? *Microbiol Rev* 46: 1-42.
- Griguer CE, Oliva CR, Gillespie GY (2005) Glucose metabolism heterogeneity in human and mouse malignant glioma cell lines. *J Neurooncol* 74: 123-33.
- Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL (2008) The Vienna RNA websuite. *Nucleic Acids Res* 36: W70-4. doi: 10.1093/nar/gkn188
- Guerra F, Girolimetti G, Perrone AM, Procaccini M, Kurelac I, Ceccarelli C, De Biase D, Caprara G, Zamagni C, De Iaco P, Santini D, Gasparre G (2014) Mitochondrial DNA genotyping efficiently reveals clonality of synchronous endometrial and ovarian cancers. *Mod Pathol* 27: 1412-20. doi: 10.1038/modpathol.2014.39
- Guerra F, Kurelac I, Cormio A, Zuntini R, Amato LB, Ceccarelli C, Santini D, Cormio G, Fracasso F, Selvaggi L, Resta L, Attimonelli M, Gadaleta MN, Gasparre G (2011) Placing mitochondrial DNA mutations within the progression model of type I endometrial carcinoma. *Hum Mol Genet* 20: 2394-405. doi: 10.1093/hmg/ddr146
- Guo J, Zheng L, Liu W, Wang X, Wang Z, Wang Z, French AJ, Kang D, Chen L, Thibodeau SN, Liu W (2011) Frequent truncating mutation of TFAM induces mitochondrial DNA depletion and apoptotic resistance in microsatellite-unstable colorectal cancer. *Cancer Res* 71: 2978-87.
- Guo Y, Cai Q, Samuels DC, Ye F, Long J, Li CI, Winther JF, Tawn EJ, Stovall M, Lahteenmaki P, Malila N, Levy S, Shaffer C, Shyr Y, Shu XO, Boice JD, Jr. (2012a) The use of next generation sequencing technology to study the effect of radiation therapy on mitochondrial DNA mutation. *Mutat Res* 744: 154-60.
- Guo Y, Li J, Li CI, Long J, Samuels DC, Shyr Y (2012b) The effect of strand bias in Illumina short-read sequencing data. *BMC Genomics* 13: 666.
- Guo Y, Li J, Li CI, Shyr Y, Samuels DC (2013) MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics* 29: 1210-1.
- Guo Z, Zhao S, Fan H, Du Y, Zhao Y, Wang G (2016) Identification of sequence polymorphisms in the D-Loop region of mitochondrial DNA as a risk factor for colon cancer. *Mitochondrial DNA A DNA Mapp Seq Anal* 27: 4244-4245. doi: 10.3109/19401736.2014.1003920
- Guppy M, Leedman P, Zu X, Russell V (2002) Contribution by different fuels and metabolic pathways to the total ATP turnover of proliferating MCF-7 breast cancer cells. *Biochem J* 364: 309-15. doi: 10.1042/bj3640309
- Gutierrez Povedano C, Salgado J, Gil C, Robles M, Patino-Garcia A, Garcia-Foncillas J (2015) Analysis of BRCA1 and mtDNA haplotypes and mtDNA polymorphism in familial breast cancer. *Mitochondrial DNA* 26: 227-31.

- Habano W, Sugai T, Yoshida T, Nakamura S (1999) Mitochondrial gene mutation, but not large-scale deletion, is a feature of colorectal carcinomas with mitochondrial microsatellite instability. *Int J Cancer* 83: 625-9.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.
- Hancock DK, Tully LA, Levin BC (2005) A Standard Reference Material to determine the sensitivity of techniques for detecting low-frequency mutations, SNPs, and heteroplasmies in mitochondrial DNA. *Genomics* 86: 446-61.
- Hardie RA, van Dam E, Cowley M, Han TL, Balaban S, Pajic M, Pinese M, Iconomou M, Shearer RF, McKenna J, Miller D, Waddell N, Pearson JV, Grimmond SM, Australian Pancreatic Cancer Genome I, Sazanov L, Biankin AV, Villas-Boas S, Hoy AJ, Turner N, Saunders DN (2017) Mitochondrial mutations and metabolic adaptation in pancreatic cancer. *Cancer Metab* 5: 2. doi: 10.1186/s40170-017-0164-1
- He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA, Jr., Kinzler KW, Vogelstein B, Papadopoulos N (2010) Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* 464: 610-4.
- Hibi K, Nakayama H, Yamazaki T, Takase T, Taguchi M, Kasai Y, Ito K, Akiyama S, Nakao A (2001) Mitochondrial DNA alteration in esophageal cancer. *Int J Cancer* 92: 319-21.
- Hofmann JN, Hosgood HD, 3rd, Liu CS, Chow WH, Shuch B, Cheng WL, Lin TT, Moore LE, Lan Q, Rothman N, Purdue MP (2014) A nested case-control study of leukocyte mitochondrial DNA copy number and renal cell carcinoma in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *Carcinogenesis* 35: 1028-31.
- Holt IJ, Harding AE, Cooper JM, Schapira AH, Toscano A, Clark JB, Morgan-Hughes JA (1989) Mitochondrial myopathies: clinical and biochemical features of 30 patients with major deletions of muscle mitochondrial DNA. *Ann Neurol* 26: 699-708.
- Holt IJ, Harding AE, Morgan-Hughes JA (1988) Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* 331: 717-9.
- Hopkins JF, Denroche RE, Aguiar JA, Notta F, Connor AA, Wilson JM, Stein LD, Gallinger S, Boutros PC (2018) Mutations in Mitochondrial DNA From Pancreatic Ductal Adenocarcinomas Associate With Survival Times of Patients and Accumulate as Tumors Progress. *Gastroenterology* 154: 1620-1624 e5. doi: 10.1053/j.gastro.2018.01.029
- Hosgood HD, 3rd, Liu CS, Rothman N, Weinstein SJ, Bonner MR, Shen M, Lim U, Virtamo J, Cheng WL, Albanes D, Lan Q (2010) Mitochondrial DNA copy number and lung cancer risk in a prospective cohort study. *Carcinogenesis* 31: 847-9.

- Hu SP, Du JP, Li DR, Yao YG (2014) Mitochondrial DNA haplogroup confers genetic susceptibility to nasopharyngeal carcinoma in Chaoshanese from Guangdong, China. *PLoS One* 9: e87795.
- Hung WY, Wu CW, Yin PH, Chang CJ, Li AF, Chi CW, Wei YH, Lee HC (2010) Somatic mutations in mitochondrial genome and their potential roles in the progression of human gastric cancer. *Biochim Biophys Acta* 1800: 264-70.
- IBM Corp. (2010) IBM SPSS Statistics for Windows, Version 19.0. Armonk, NY: IBM Corp.
- Illumina Inc. (2010) Illumina Sequencing Technology. https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf. Accessed 2013
- Illumina Inc. (2017) An introduction to Next-Generation Sequencing Technology. Accessed 2013
- Illumina Inc. (2019) Illumina sequencing platforms. <https://www.illumina.com/>.
- Indran IR, Tufo G, Pervaiz S, Brenner C (2011) Recent advances in apoptosis, mitochondria and drug resistance in cancer cells. *Biochim Biophys Acta* 1807: 735-45. doi: 10.1016/j.bbabbio.2011.03.010
- Iommarini L, Calvaruso MA, Kurelac I, Gasparre G, Porcelli AM (2013) Complex I impairment in mitochondrial diseases and cancer: parallel roads leading to different outcomes. *Int J Biochem Cell Biol* 45: 47-63.
- Jakel S, Dimou L (2017) Glial Cells and Their Function in the Adult Brain: A Journey through the History of Their Ablation. *Front Cell Neurosci* 11: 24.
- Jakupciak JP, Maragh S, Markowitz ME, Greenberg AK, Hoque MO, Maitra A, Barker PE, Wagner PD, Rom WN, Srivastava S, Sidransky D, O'Connell CD (2008) Performance of mitochondrial DNA mutations detecting early stage cancer. *BMC Cancer* 8: 285.
- Jalilvand S, Shoja Z, Marashi SM, Shahmahmoodi S, Safaie-Naraghi Z, Nourijelyani K, Nesheli AB, Mokhtari-Azad T (2015) Mitochondrial haplogroups and control region polymorphisms in Kaposi's sarcoma patients. *J Med Virol* 87: 1608-15.
- Jayaprakash AD, Benson EK, Gone S, Liang R, Shim J, Lambertini L, Toloue MM, Wigler M, Aaronson SA, Sachidanandam R (2015) Stable heteroplasmy at the single-cell level is facilitated by intercellular exchange of mtDNA. *Nucleic Acids Res* 43: 2177-87.
- Jemt E, Persson O, Shi Y, Mehmedovic M, Uhler JP, Davila Lopez M, Freyer C, Gustafsson CM, Samuelsson T, Falkenberg M (2015) Regulation of DNA replication at the end of the mitochondrial D-loop involves the helicase TWINKLE and a conserved sequence element. *Nucleic Acids Res* 43: 9262-75. doi: 10.1093/nar/gkv804
- Jeronimo C, Nomoto S, Caballero OL, Usadel H, Henrique R, Varzim G, Oliveira J, Lopes C, Fliss MS, Sidransky D (2001) Mitochondrial mutations in early stage prostate cancer and bodily fluids. *Oncogene* 20: 5195-8. doi: 10.1038/sj.onc.1204646

- Ji Y, Liang M, Zhang J, Zhu L, Zhang Z, Fu R, Liu X, Zhang M, Fu Q, Zhao F, Tong Y, Sun Y, Jiang P, Guan MX (2016) Mitochondrial ND1 Variants in 1281 Chinese Subjects With Leber's Hereditary Optic Neuropathy. *Invest Ophthalmol Vis Sci* 57: 2377–89.
- Jia P, Li F, Xia J, Chen H, Ji H, Pao W, Zhao Z (2012) Consensus rules in variant detection from next-generation sequencing data. *PLoS One* 7: e38470.
- Jiang WW, Masayeva B, Zahurak M, Carvalho AL, Rosenbaum E, Mambo E, Zhou S, Minhas K, Benoit N, Westra WH, Alberg A, Sidransky D, Koch W, Califano J (2005) Increased mitochondrial DNA content in saliva associated with head and neck cancer. *Clin Cancer Res* 11: 2486–91.
- Jin X, Zhang J, Gao Y, Ding K, Wang N, Zhou D, Jen J, Cheng S (2007) Relationship between mitochondrial DNA mutations and clinical characteristics in human lung cancer. *Mitochondrion* 7: 347–53.
- Johns DR, Neufeld MJ, Park RD (1992) An ND-6 mitochondrial DNA mutation associated with Leber hereditary optic neuropathy. *Biochem Biophys Res Commun* 187: 1551–7.
- Johns Hopkins Medicine (2019) Brain Tumor Grades: Biopsy and Prognosis. https://www.hopkinsmedicine.org/neurology_neurosurgery/centers_clinics/brain_tumor/diagnosis/brain-tumor-grade.html. Accessed 2015
- Jones JB, Song JJ, Hempen PM, Parmigiani G, Hruban RH, Kern SE (2001) Detection of mitochondrial DNA mutations in pancreatic cancer offers a “mass”-ive advantage over detection of nuclear DNA mutations. *Cancer Res* 61: 1299–304.
- Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, Fabani MM, Seguritan V, Green J, Pride DT, Yooseph S, Biggs W, Nelson KE, Venter JC (2015) Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci U S A* 112: 14024–9. doi: 10.1073/pnas.1519288112
- Jose C, Bellance N, Rossignol R (2011) Choosing between glycolysis and oxidative phosphorylation: a tumor's dilemma? *Biochim Biophys Acta* 1807: 552–61.
- Ju YS, Alexandrov LB, Gerstung M, Martincorena I, Nik-Zainal S, Ramakrishna M, Davies HR, Papaemmanuil E, Gundem G, Shlien A, Bolli N, Behjati S, Tarpey PS, Nangalia J, Massie CE, Butler AP, Teague JW, Vassiliou GS, Green AR, Du MQ, Unnikrishnan A, Pimanda JE, Teh BT, Munshi N, Greaves M, Vyas P, El-Naggar AK, Santarius T, Collins VP, Grundy R, Taylor JA, Hayes DN, Malkin D, Foster CS, Warren AY, Whitaker HC, Brewer D, Eeles R, Cooper C, Neal D, Visakorpi T, Isaacs WB, Bova GS, Flanagan AM, Futreal PA, Lynch AG, Chinnery PF, McDermott U, Stratton MR, Campbell PJ (2014) Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *Elife* 3.
- Just RS, Irwin JA, Parson W (2015) Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Sci Int Genet* 18: 131–9.

- Kabekkodu SP, Bhat S, Mascarenhas R, Mallya S, Bhat M, Pandey D, Kushtagi P, Thangaraj K, Gopinath PM, Satyamoorthy K (2014) Mitochondrial DNA variation analysis in cervical cancer. *Mitochondrion* 16: 73-82.
- Kalsbeek AM, Chan EF, Grogan J, Petersen DC, Jaratlerdsiri W, Gupta R, Lyons RJ, Haynes AM, Horvath LG, Kench JG, Stricker PD, Hayes VM (2016) Mutational load of the mitochondrial genome predicts pathological features and biochemical recurrence in prostate cancer. *Aging (Albany NY)* 8: 2702-2712. doi: 10.18632/aging.101044
- Kassauei K, Habbe N, Mullendore ME, Karikari CA, Maitra A, Feldmann G (2006) Mitochondrial DNA mutations in pancreatic cancer. *Int J Gastrointest Cancer* 37: 57-64.
- Kato T, Kunugi H, Nanko S, Kato N (2001) Mitochondrial DNA polymorphisms in bipolar disorder. *J Affect Disord* 62: 151-64.
- Kazdal D, Harms A, Endris V, Penzel R, Oliveira C, Kriegsmann M, Longuespee R, Winter H, Schneider MA, Muley T, Pfarr N, Weichert W, Stenzinger A, Warth A (2018) Subclonal evolution of pulmonary adenocarcinomas delineated by spatially distributed somatic mitochondrial mutations. *Lung Cancer* 126: 80-88. doi: 10.1016/j.lungcan.2018.10.024
- Kazuno AA, Munakata K, Nagai T, Shimozono S, Tanaka M, Yoneda M, Kato N, Miyawaki A, Kato T (2006) Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics. *PLoS Genet* 2: e128.
- Khaidakov M, Shmookler Reis RJ (2005) Possibility of selection against mtDNA mutations in tumors. *Mol Cancer* 4: 36.
- Khusnutdinova E, Gilyazova I, Ruiz-Pesini E, Derbeneva O, Khusainova R, Khidiyatova I, Magzhanov R, Wallace DC (2008) A mitochondrial etiology of neurodegenerative diseases: evidence from Parkinson's disease. *Ann N Y Acad Sci* 1147: 1-20.
- Kim A (2014) Mitochondrial DNA somatic mutation in cancer. *Toxicol Res* 30: 235-42.
- Kircher M, Kelso J (2010) High-throughput DNA sequencing-- concepts and limitations. *Bioessays* 32: 524-36.
- Kircher M, Stenzel U, Kelso J (2009) Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol* 10: R83.
- Kirches E, Krause G, Warich-Kirches M, Weis S, Schneider T, Meyer-Puttlitz B, Mawrin C, Dietzmann K (2001) High frequency of mitochondrial DNA mutations in glioblastoma multiforme identified by direct sequence comparison to blood samples. *Int J Cancer* 93: 534-8.
- Kirches E, Krause G, Weis S, Mawrin C, Dietzmann K (2002) Comparison between mitochondrial DNA sequences in low grade astrocytomas and corresponding blood samples. *Mol Pathol* 55: 204-6.

- Kirches E, Michael M, Woy C, Schneider T, Warich-Kirches M, Schneider-Stock R, Winkler K, Wittig H, Dietzmann K (1999) Loss of heteroplasmy in the displacement loop of brain mitochondrial DNA in astrocytic tumors. *Genes Chromosomes Cancer* 26: 80-3.
- Kirk R, Furlong RA, Amos W, Cooper G, Rubinsztein JS, Walsh C, Paykel ES, Rubinsztein DC (1999) Mitochondrial genetic analyses suggest selection against maternal lineages in bipolar affective disorder. *Am J Hum Genet* 65: 508-18.
- Kivisild T (2015) Maternal ancestry and population history from whole mitochondrial genomes. *Investig Genet* 6: 3. doi: 10.1186/s13323-015-0022-2
- Klemba A, Kowalewska M, Kukwa W, Tonska K, Szybinska A, Mossakowska M, Scinska A, Golik P, Koper K, Radziszewski J, Kukwa A, Czarnecka AM, Bartnik E (2010) Mitochondrial genotype in vulvar carcinoma - cuckoo in the nest. *J Biomed Sci* 17: 73.
- Kloss-Brandstatter A, Schafer G, Erhart G, Huttenhofer A, Coassin S, Seifarth C, Summerer M, Bektic J, Klocker H, Kronenberg F (2010) Somatic mutations throughout the entire mitochondrial genome are associated with elevated PSA levels in prostate cancer patients. *Am J Hum Genet* 87: 802-12.
- Kloss-Brandstatter A, Weissensteiner H, Erhart G, Schafer G, Forer L, Schonherr S, Pacher D, Seifarth C, Stockl A, Fendt L, Sottas I, Klocker H, Huck CW, Rasse M, Kronenberg F, Kloss FR (2015) Validation of Next-Generation Sequencing of Entire Mitochondrial Genomes and the Diversity of Mitochondrial DNA Mutations in Oral Squamous Cell Carcinoma. *PLoS One* 10: e0135643.
- Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. *Cell* 155: 27-38.
- Kofler B, Mueller EE, Eder W, Stanger O, Maier R, Weger M, Haas A, Winker R, Schmut O, Paulweber B, Iglseder B, Renner W, Wiesbauer M, Aigner I, Santic D, Zimmermann FA, Mayr JA, Sperl W (2009) Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study. *BMC Med Genet* 10: 35.
- Koh H, Park GS, Shin SM, Park CE, Kim S, Han SJ, Pham HQ, Shin JH, Lee DW (2018) Mitochondrial Mutations in Cholestatic Liver Disease with Biliary Atresia. *Sci Rep* 8: 905. doi: 10.1038/s41598-017-18958-8
- Kornblihtt L, Cavaliere V, Blanco GA (2016) Mitofagia: una forma selectiva de autofagia que participa en la homeostasis de las células hematopoyéticas. *Hematología* 20: 216-23.
- Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ (2009) Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* 6: 291-5. doi: 10.1038/nmeth.1311

- Kraytsberg Y, Schwartz M, Brown TA, Ebraldise K, Kunz WS, Clayton DA, Vissing J, Khrapko K (2004) Recombination of human mitochondrial DNA. *Science* 304: 981.
- Kuo SJ, Chen M, Ma GC, Chen ST, Chang SP, Lin WY, Chen YC, Lee TH, Lin TT, Liu CS (2010) Number of somatic mutations in the mitochondrial D-loop region indicates poor prognosis in breast cancer, independent of TP53 mutation. *Cancer Genet Cytogenet* 201: 94–101.
- Kurelac I, MacKay A, Lambros MB, Di Cesare E, Cenacchi G, Ceccarelli C, Morra I, Melcarne A, Morandi L, Calabrese FM, Attimonelli M, Tallini G, Reis-Filho JS, Gasparre G (2013) Somatic complex I disruptive mitochondrial DNA mutations are modifiers of tumorigenesis that correlate with low genomic instability in pituitary adenomas. *Hum Mol Genet* 22: 226–38.
- Lai CH, Huang SF, Liao CT, Chen IH, Wang HM, Hsieh LL (2013) Clinical significance in oral cavity squamous cell carcinoma of pathogenic somatic mitochondrial mutations. *PLoS One* 8: e65578.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–9.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25. doi: 10.1186/gb-2009-10-3-r25
- Larman TC, DePalma SR, Hadjipanayis AG, Protopopov A, Zhang J, Gabriel SB, Chin L, Seidman CE, Kucherlapati R, Seidman JG (2012) Spectrum of somatic mitochondrial mutations in five cancers. *Proc Natl Acad Sci U S A* 109: 14087–91.
- Le A, Stine ZE, Nguyen C, Afzal J, Sun P, Hamaker M, Siegel NM, Gouw AM, Kang BH, Yu SH, Cochran RL, Sailor KA, Song H, Dang CV (2014) Tumorigenicity of hypoxic respiring cancer cells revealed by a hypoxia-cell cycle dual reporter. *Proc Natl Acad Sci U S A* 111: 12486–91. doi: 10.1073/pnas.1402012111
- Lebedeva MA, Eaton JS, Shadel GS (2009) Loss of p53 causes mitochondrial DNA depletion and altered mitochondrial reactive oxygen species homeostasis. *Biochim Biophys Acta* 1787: 328–34.
- Ledergerber C, Dessimoz C (2011) Base-calling for next-generation sequencing platforms. *Brief Bioinform* 12: 489–97.
- Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, Kim SJ, Mehta H, Hevener AL, de Cabo R, Cohen P (2015) The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab* 21: 443–54.
- Lee HC, Li SH, Lin JC, Wu CC, Yeh DC, Wei YH (2004) Somatic mutations in the D-loop and decrease in the copy number of mitochondrial DNA in human hepatocellular carcinoma. *Mutat Res* 547: 71–8.
- Lee HC, Yin PH, Lu CY, Chi CW, Wei YH (2000) Increase of mitochondria and mitochondrial DNA in response to oxidative stress in human cells. *Biochem J* 348 Pt 2: 425–32.

- Lehtonen MS, Moilanen JS, Majamaa K (2003) Increased variation in mtDNA in patients with familial sensorineural hearing impairment. *Hum Genet* 113: 220-7.
- Li D, Du X, Guo X, Zhan L, Li X, Yin C, Chen C, Li M, Li B, Yang H, Xing J (2017) Site-specific selection reveals selective constraints and functionality of tumor somatic mtDNA mutations. *J Exp Clin Cancer Res* 36: 168.
- Li F, Wang X, Han C, Lin J (2004a) Decreased mtDNA Copy Number of Gastric Cancer: a New Tumor Marker? *Chinese Journal of Clinical Oncology* 1.
- Li H (2012) Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* 28: 1838-44. doi: 10.1093/bioinformatics/bts280
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.
- Li M, Rothwell R, Vermaat M, Wachsmuth M, Schroder R, Laros JF, van Oven M, de Bakker PI, Bovenberg JA, van Duijn CM, van Ommen GJ, Slagboom PE, Swertz MA, Wijmenga C, Kayser M, Boomsma DI, Zollner S, de Knijff P, Stoneking M (2016) Transmission of human mtDNA heteroplasmy in the Genome of the Netherlands families: support for a variable-size bottleneck. *Genome Res* 26: 417-26.
- Li M, Schonberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M (2010) Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet* 87: 237-49.
- Li M, Schroder R, Ni S, Madea B, Stoneking M (2015a) Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc Natl Acad Sci U S A* 112: 2491-6.
- Li M, Schroeder R, Ko A, Stoneking M (2012) Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Res* 40: e137.
- Li M, Stoneking M (2012) A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biol* 13: R34.
- Li R, Greinwald JH, Jr., Yang L, Choo DI, Wenstrup RJ, Guan MX (2004b) Molecular analysis of the mitochondrial 12S rRNA and tRNA^{Ser}(UCN) genes in paediatric subjects with non-syndromic hearing loss. *J Med Genet* 41: 615-20. doi: 10.1136/jmg.2004.020230
- Li X, Guo X, Li D, Du X, Yin C, Chen C, Fang W, Bian Z, Zhang J, Li B, Yang H, Xing J (2018) Multi-regional sequencing reveals intratumor heterogeneity and positive selection of somatic mtDNA mutations in hepatocellular carcinoma and colorectal cancer. *Int J Cancer* 143: 1143-1152. doi: 10.1002/ijc.31395

- Li XY, Guo YB, Su M, Cheng L, Lu ZH, Tian DP (2011) Association of mitochondrial haplogroup D and risk of esophageal cancer in Taihang Mountain and Chaoshan areas in China. *Mitochondrion* 11: 27–32.
- Li Y, Beckman KB, Caberto C, Kazma R, Lum-Jones A, Haiman CA, Le Marchand L, Stram DO, Saxena R, Cheng I (2015b) Association of Genes, Pathways, and Haplogroups of the Mitochondrial Genome with the Risk of Colorectal Cancer: The Multiethnic Cohort. *PLoS One* 10: e0136796. doi: 10.1371/journal.pone.0136796.
- Liang BC, Grootveld M (2011) The importance of mitochondria in the tumorigenic phenotype: gliomas as the paradigm (review). *Int J Mol Med* 27: 159–71.
- Lievre A, Chapusot C, Bouvier AM, Zinzindohoue F, Piard F, Roignot P, Arnould L, Beaune P, Faivre J, Laurent-Puig P (2005) Clinical value of mitochondrial mutations in colorectal cancer. *J Clin Oncol* 23: 3517–25.
- Lin CS, Chang SC, Wang LS, Chou TY, Hsu WH, Wu YC, Wei YH (2010) The role of mitochondrial DNA alterations in esophageal squamous cell carcinomas. *J Thorac Cardiovasc Surg* 139: 189–197 e4.
- Lin CY, Tsai LC, Hsieh HM, Huang CH, Yu YJ, Tseng B, Linacre A, Lee JC (2017a) Investigation of length heteroplasmy in mitochondrial DNA control region by massively parallel sequencing. *Forensic Sci Int Genet* 30: 127–133. doi: 10.1016/j.fsigen.2017.07.003
- Lin H, Patel S, Affleck VS, Wilson I, Turnbull DM, Joshi AR, Maxwell R, Stoll EA (2017b) Fatty acid oxidation is required for the respiration and proliferation of malignant glioma cells. *Neuro Oncol* 19: 43–54.
- Lin JC, Wang CC, Jiang RS, Wang WY, Liu SA (2015) Impact of somatic mutations in the D-loop of mitochondrial DNA on the survival of oral squamous cell carcinoma patients. *PLoS One* 10: e0124322. doi: 10.1371/journal.pone.0124322
- Lindberg J, Mills IG, Klevebring D, Liu W, Neiman M, Xu J, Wikstrom P, Wiklund P, Wiklund F, Egevad L, Gronberg H (2013) The mitochondrial and autosomal mutation landscapes of prostate cancer. *Eur Urol* 63: 702–8.
- Linkowska K, Jawien A, Marszalek A, Malyarchuk BA, Tonska K, Bartnik E, Skonieczna K, Grzybowski T (2015) Mitochondrial DNA Polymerase gamma Mutations and Their Implications in mtDNA Alterations in Colorectal Cancer. *Ann Hum Genet* 79: 320–328. doi: 10.1111/ahg.12111
- Liou GY, Storz P (2010) Reactive oxygen species in cancer. *Free Radic Res* 44: 479–96.
- Liu SA, Jiang RS, Chen FJ, Wang WY, Lin JC (2011) Somatic mutations in the D-loop of mitochondrial DNA in oral squamous cell carcinoma. *Eur Arch Otorhinolaryngol* 269: 1665–70.

- Liu SA, Jiang RS, Wang WY, Lin JC (2015) Somatic mutations in the D-loop of mitochondrial DNA in head and neck squamous cell carcinoma. *Head Neck* 37: 878–83. doi: 10.1002/hed.23680
- Liu VW, Shi HH, Cheung AN, Chiu PM, Leung TW, Nagley P, Wong LC, Ngan HY (2001) High incidence of somatic mitochondrial DNA mutations in human ovarian carcinomas. *Cancer Res* 61: 5998–6001.
- Lorenc A, Bryk J, Golik P, Kupryjanczyk J, Ostrowski J, Pronicki M, Semczuk A, Szolkowska M, Bartnik E (2003) Homoplasmic MELAS A3243G mtDNA mutation in a colon cancer sample. *Mitochondrion* 3: 119–24.
- Losanoff JE, Zhu W, Qin W, Mannello F, Sauter ER (2008) Can mitochondrial DNA mutations in circulating white blood cells and serum be used to detect breast cancer? *Breast* 17: 540–2.
- Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, Procaccio V, Wallace DC (2013) mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr Protoc Bioinformatics* 44: 1 23 1–26.
- Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, Kleihues P (2007) The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 114: 97–109.
- Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW (2016) The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* 131: 803–20.
- Lueth M, von Deimling A, Pietsch T, Wong LJ, Kurtz A, Henze G, Driever PH (2010) Medulloblastoma harbor somatic mitochondrial DNA mutations in the D-loop region. *J Pediatr Hematol Oncol* 32: 156–9.
- Lueth M, Wronski L, Giese A, Kirschner-Schwabe R, Pietsch T, von Deimling A, Henze G, Kurtz A, Driever PH (2009) Somatic mitochondrial mutations in pilocytic astrocytoma. *Cancer Genet Cytogenet* 192: 30–5.
- Luo S, Valencia CA, Zhang J, Lee NC, Slone J, Gui B, Wang X, Li Z, Dell S, Brown J, Chen SM, Chien YH, Hwu WL, Fan PC, Wong LJ, Atwal PS, Huang T (2018) Biparental Inheritance of Mitochondrial DNA in Humans. *Proc Natl Acad Sci U S A* 115: 13039–13044. doi: 10.1073/pnas.1810946115
- Lutz S, Weisser HJ, Heizmann J, Pollak S (1997) A third hypervariable region in the human mitochondrial D-loop. *Hum Genet* 101: 384.
- Lloyd RE, Keatley K, Littlewood DT, Meunier B, Holt WV, An Q, Higgins SC, Polyzoidis S, Stephenson KF, Ashkan K, Fillmore HL, Pilkington GJ, McGeehan JE (2015) Identification and functional prediction of mitochondrial complex III and IV mutations associated with glioblastoma. *Neuro Oncol* 17: 942–52.

- Mambo E, Chatterjee A, Xing M, Tallini G, Haugen BR, Yeung SC, Sukumar S, Sidransky D (2005) Tumor-specific changes in mtDNA content in human cancer. *Int J Cancer* 116: 920-4.
- Mantione KJ (2016) Mitochondrial Heteroplasmy: Detection and Significance. *American Journal of Bioinformatics Research* 6: 14-18.
- Mardis ER (2008) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9: 387-402.
- Marine R, Polson SW, Ravel J, Hatfull G, Russell D, Sullivan M, Syed F, Dumas M, Wommack KE (2011) Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Appl Environ Microbiol* 77: 8071-9.
- Martinez-Ramirez M, Coral-Vazquez RM, Tenorio A, Mendez JP, Benitez-Granados J, Maffuz Asis A, Rodriguez Cuevas S, Dominguez Reyes C, Erazo-Valle A, Canto P (2018) Complete sequence of the ATP6 and ND3 mitochondrial genes in breast cancer tissue of postmenopausal women with different body mass indexes. *Ann Diagn Pathol* 32: 23-27. doi: 10.1016/j.anndiagpath.2017.09.001
- Martinez-Redondo D, Marcuello A, Casajus JA, Ara I, Dahmani Y, Montoya J, Ruiz-Pesini E, Lopez-Perez MJ, Diez-Sanchez C (2010) Human mitochondrial haplogroup H: the highest VO₂max consumer--is it a paradox? *Mitochondrion* 10: 102-7.
- Martorell A (2014) Breve introducción a la dinámica mitocondrial. *Encuentros en la biología* 7: 179-81.
- Marucci G, Maresca A, Caporali L, Farnedi A, Betts CM, Morandi L, de Biase D, Cerasoli S, Foschini MP, Bonora E, Vidone M, Romeo G, Perli E, Giordano C, d'Amati G, Gasparre G, Baruzzi A, Carelli V, Eusebi V (2013) Oncocytic glioblastoma: a glioblastoma showing oncocytic changes and increased mitochondrial DNA copy number. *Hum Pathol* 44: 1867-76.
- Mattson MP, Gleichmann M, Cheng A (2008) Mitochondria in neuroplasticity and neurological disorders. *Neuron* 60: 748-66.
- McCrow JP, Petersen DC, Louw M, Chan EK, Harmeyer K, Vecchiarelli S, Lyons RJ, Bornman MS, Hayes VM (2016) Spectrum of mitochondrial genomic variation and associated clinical presentation of prostate cancer in South African men. *Prostate* 76: 349-58. doi: 10.1002/pros.23126
- McElhoe JA, Holland MM, Makova KD, Su MS, Paul IM, Baker CH, Faith SA, Young B (2014) Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. *Forensic Sci Int Genet* 13: 20-9.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297-303.

- McMahon S, LaFramboise T (2014) Mutational patterns in the breast cancer mitochondrial genome, with clinical correlates. *Carcinogenesis* 35: 1046-54.
- Mehrabi S, Akwe JA, Adams G, Jr., Grizzle W, Yao X, Aikhionbare FO (2010) Analysis of mtDNA sequence variants in colorectal adenomatous polyps. *Diagn Pathol* 5: 66.
- Meierhofer D, Mayr JA, Foetschl U, Berger A, Fink K, Schmeller N, Hacker GW, Hauser-Kronberger C, Kofler B, Sperl W (2004) Decrease of mitochondrial DNA content and energy metabolism in renal cell carcinoma. *Carcinogenesis* 25: 1005-10.
- Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11: 31-46.
- Michikawa Y, Mazzucchelli F, Bresolin N, Scarlato G, Attardi G (1999) Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. *Science* 286: 774-9. doi: 10.1126/science.286.5440.774
- Mikkelsen M, Rockenbauer E, Wächter A, Fendt L, Zimmermann B, Parson W, Nielsen SA, Gilbert T, Willerslev E, Morling N (2009) Application of full mitochondrial genome sequencing using 454 GS FLX pyrosequencing. *Forensic Science International: Genetics Supplement Series*. doi: doi:10.1016/j.fsigss.2009.09.028
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* 100: 171-6.
- Mithani SK, Taube JM, Zhou S, Smith IM, Koch WM, Westra WH, Califano JA (2007) Mitochondrial mutations are a late event in the progression of head and neck squamous cell cancer. *Clin Cancer Res* 13: 4331-5.
- Mizumachi T, Muskhelishvili L, Naito A, Furusawa J, Fan CY, Siegel ER, Kadlubar FF, Kumar U, Higuchi M (2008) Increased distributional variance of mitochondrial DNA content associated with prostate cancer cells as compared with normal prostate cells. *Prostate* 68: 408-17.
- Modica-Napolitano JS, Singh KK (2004) Mitochondrial dysfunction in cancer. *Mitochondrion* 4: 755-62.
- Mohamed Yusoff AA, Mohd Nasir KN, Haris K, Mohd Khair SZN, Abdul Ghani ARI, Idris Z, Abdullah JM (2017) Detection of somatic mutations in the mitochondrial DNA control region D-loop in brain tumors: The first report in Malaysian patients. *Oncol Lett* 14: 5179-5188. doi: 10.3892/ol.2017.6851
- Mondal R, Ghosh SK, Talukdar FR, Laskar RS (2013) Association of mitochondrial D-loop mutations with GSTM1 and GSTT1 polymorphisms in oral carcinoma: a case control study from northeast India. *Oral Oncol* 49: 345-53.
- Montanini L, Regna-Gladin C, Eoli M, Albarosa R, Carrara F, Zeviani M, Bruzzone MG, Broggi G, Boiardi A, Finocchiaro G (2005) Instability of mitochondrial DNA and MRI and clinical correlations in malignant gliomas. *J Neurooncol* 74: 87-9.

- Moore A, Lan Q, Hofmann JN, Liu CS, Cheng WL, Lin TT, Berndt SI (2017) A prospective study of mitochondrial DNA copy number and the risk of prostate cancer. *Cancer Causes Control* 28: 529–538.
- Morris AA, Lightowers RN (2000) Mitochondrial DNA recombination. *Lancet* 356: 941.
- Mosquera-Miguel A, Alvarez-Iglesias V, Carracedo A, Salas A, Vega A, Carracedo A, Milne R, de Leon AC, Benitez J, Carracedo A, Salas A (2008) Is mitochondrial DNA variation associated with sporadic breast cancer risk? *Cancer Res* 68: 623–5; author reply 624.
- Mostafaie N, Rossmanith W, Hombauer H, Dechat T, Raffelsberger T, Bauer K, Worofka B, Kittl E, Hofmann J, Hejtman M, Kirchmeyr W, Schreiber W, Weissgram S, Jungwirth S, Fischer P, Bittner R, Huber K (2004) Mitochondrial genotype and risk for Alzheimer’s disease: cross-sectional data from the Vienna-Transdanube-Aging “VITA” study. *J Neural Transm (Vienna)* 111: 1155–65.
- Mueller EE, Brunner SM, Mayr JA, Stanger O, Sperl W, Kofler B (2012) Functional differences between mitochondrial haplogroup T and haplogroup H in HEK293 cybrid cells. *PLoS One* 7: e52367.
- Nagy A, Wilhelm M, Kovacs G (2003) Mutations of mtDNA in renal cell tumours arising in end-stage renal disease. *J Pathol* 199: 237–42.
- Nakada M, Kita D, Watanabe T, Hayashi Y, Teng L, Pyko IV, Hamada J (2011) Aberrant signaling pathways in glioma. *Cancers (Basel)* 3: 3242–78.
- Navaglia F, Basso D, Fogar P, Sperti C, Greco E, Zambon CF, Stranges A, Falda A, Pizzi S, Parenti A, Pedrazzoli S, Plebani M (2006) Mitochondrial DNA D-loop in pancreatic cancer: somatic mutations are epiphenomena while the germline 16519 T variant worsens metabolism and outcome. *Am J Clin Pathol* 126: 593–601.
- Nicholls TJ, Minczuk M (2014) In D-loop: 40 years of mitochondrial 7S DNA. *Exp Gerontol* 56: 175–81.
- Nomoto S, Yamashita K, Koshikawa K, Nakao A, Sidransky D (2002) Mitochondrial D-loop mutations as clonal markers in multicentric hepatocellular carcinoma and plasma. *Clin Cancer Res* 8: 481–7.
- Okada AK, Teranishi K, Lobo F, Isas JM, Xiao J, Yen K, Cohen P, Langen R (2017) The Mitochondrial-Derived Peptides, HumaninS14G and Small Humanin-like Peptide 2, Exhibit Chaperone-like Activity. *Sci Rep* 7: 7802.
- Pagani IS, Kok CH, Saunders VA, Van der Hoek MB, Heatley SL, Schwarzer AP, Hahn CN, Hughes TP, White DL, Ross DM (2017) A Method for Next-Generation Sequencing of Paired Diagnostic and Remission Samples to Detect Mitochondrial DNA Mutations Associated with Leukemia. *J Mol Diagn* 19: 711–721. doi: 10.1016/j.jmoldx.2017.05.009

- Palodhi A, Ghosh S, Biswas NK, Basu A, Majumder PP, Maitra A (2019) Profiling of genomic alterations of mitochondrial DNA in gingivobuccal oral squamous cell carcinoma: Implications for disease progress. *Mitochondrion* 46: 361–369. doi: 10.1016/j.mito.2018.09.006
- Park JS, Sharma LK, Li H, Xiang R, Holstein D, Wu J, Lechleiter J, Naylor SL, Deng JJ, Lu J, Bai Y (2009) A heteroplasmic, not homoplasmic, mitochondrial DNA mutation promotes tumorigenesis via alteration in reactive oxygen species generation and apoptosis. *Hum Mol Genet* 18: 1578–89.
- Parr RL, Dakubo GD, Crandall KA, Maki J, Reguly B, Aguirre A, Wittcock R, Robinson K, Alexander JS, Birch-Machin MA, Abdel-Malak M, Froberg MK, Diamandis EP, Thayer RE (2006) Somatic mitochondrial DNA mutations in prostate cancer and normal appearing adjacent glands in comparison to age-matched prostate samples without malignant histology. *J Mol Diagn* 8: 312–9.
- Parrella P, Xiao Y, Fliss M, Sanchez-Céspedes M, Mazzarelli P, Rinaldi M, Nicol T, Gabrielson E, Cuomo C, Cohen D, Pandit S, Spencer M, Rabitti C, Fazio VM, Sidransky D (2001) Detection of mitochondrial DNA mutations in primary breast cancer and fine-needle aspirates. *Cancer Res* 61: 7623–6.
- Pavlidis S, Vera I, Gandara R, Sneddon S, Pestell RG, Mercier I, Martinez-Outschoorn UE, Whitaker-Menezes D, Howell A, Sotgia F, Lisanti MP (2012) Warburg meets autophagy: cancer-associated fibroblasts accelerate tumor growth and metastasis via oxidative stress, mitophagy, and aerobic glycolysis. *Antioxid Redox Signal* 16: 1264–84. doi: 10.1089/ars.2011.4243
- Payne BA, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, Taylor RW, Samuels DC, Santibanez-Koref M, Chinnery PF (2013) Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet* 22: 384–90.
- Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H, Mort M, Cooper DN, Sebat J, Iakoucheva LM, Mooney SD, Radivojac P (2017) MutPred2: inferring the molecular and phenotypic impact of amino acid variants. bioRxiv 134981 doi: <https://doi.org/10.1101/134981>
- Peng L, Parpura V, Verkhratsky A (2014) EDITORIAL Neuroglia as a Central Element of Neurological Diseases: An Underappreciated Target for Therapeutic Intervention. *Curr Neuropharmacol* 12: 303–7.
- Penta JS, Johnson FM, Wachsman JT, Copeland WC (2001) Mitochondrial DNA in human malignancy. *Mutat Res* 488: 119–33.
- Pereira F, Soares P, Carneiro J, Pereira L, Richards MB, Samuels DC, Amorim A (2008) Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region. *Mol Biol Evol* 25: 2759–70. doi: 10.1093/molbev/msn225

- Pereira L, Soares P, Maximo V, Samuels DC (2012) Somatic mitochondrial DNA mutations in cancer escape purifying selection and high pathogenicity mutations lead to the oncocytic phenotype: pathogenicity analysis of reported somatic mtDNA mutations in tumors. *BMC Cancer* 12: 53.
- Petros JA, Baumann AK, Ruiz-Pesini E, Amin MB, Sun CQ, Hall J, Lim S, Issa MM, Flanders WD, Hosseini SH, Marshall FF, Wallace DC (2005) mtDNA mutations increase tumorigenicity in prostate cancer. *Proc Natl Acad Sci U S A* 102: 719-24.
- Pienaar IS, Howell N, Elson JL (2017) MutPred mutational load analysis shows mildly deleterious mitochondrial DNA variants are not more prevalent in Alzheimer's patients, but may be under-represented in healthy older individuals. *Mitochondrion* 34: 141-146.
- Pierron D, Chang I, Arachiche A, Heiske M, Thomas O, Borlin M, Pennarun E, Murail P, Thoraval D, Rocher C, Letellier T (2011) Mutation rate switch inside Eurasian mitochondrial haplogroups: impact of selection and consequences for dating settlement in Europe. *PLoS One* 6: e21543.
- Polyak K, Li Y, Zhu H, Lengauer C, Willson JK, Markowitz SD, Trush MA, Kinzler KW, Vogelstein B (1998) Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat Genet* 20: 291-3.
- Prior SL, Griffiths AP, Baxter JM, Baxter PW, Hodder SC, Silvester KC, Lewis PD (2006) Mitochondrial DNA mutations in oral squamous cell carcinoma. *Carcinogenesis* 27: 945-50.
- Prior SL, Griffiths AP, Lewis PD (2009) A study of mitochondrial DNA D-loop mutations and p53 status in nonmelanoma skin cancer. *Br J Dermatol* 161: 1067-71.
- Promega Corp. (2019) Biomath Calculators <https://www.promega.es/resources/tools/biomath/>. Accessed 2013
- Pütz J, Dupuis B, Sissler M, Florentz C (2007) Mamit-tRNA, a database of mammalian mitochondrial tRNA primary and secondary structures. *RNA* 13: 1184-1190.
- Pyle A, Hudson G, Wilson IJ, Coxhead J, Smertenko T, Herbert M, Santibanez-Koref M, Chinnery PF (2015) Extreme-Depth Re-sequencing of Mitochondrial DNA Finds No Evidence of Paternal Transmission in Humans. *PLoS Genet* 11: e1005040.
- Qu F, Liu X, Zhou F, Yang H, Bao G, He X, Xing J (2011) Association Between Mitochondrial DNA Content in Leukocytes and Colorectal Cancer Risk. *Cancer* DOI: 10.1002/cncr.25906.
- Ramos A, Aluja MP, Lima M, Santos C (2016) Somatic vs Germinal Mutations in Mitochondrial DNA: Is There Any Relation with Human Health and Aging? In: Lima M, Ramos A, Santos C (eds) *Anthropology: Current and Future Developments*, vol 2, pp 49-63
- Ramos A, Santos C, Alvarez L, Nogues R, Aluja MP (2009) Human mitochondrial DNA complete amplification and sequencing: a new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification. *Electrophoresis* 30: 1587-93.

- Ramos A, Santos C, Barbena E, Mateiu L, Alvarez L, Nogues R, Aluja MP (2011) Validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification: a revision based on the New Human Genome Reference Sequence (GRCh37). *Electrophoresis* 32: 782-3.
- Ramos A, Santos C, Mateiu L, Gonzalez Mdel M, Alvarez L, Azevedo L, Amorim A, Aluja MP (2013) Frequency and pattern of heteroplasmy in the complete human mitochondrial genome. *PLoS One* 8: e74636.
- Ravanat JL, Douki T, Duez P, Gremaud E, Herbert K, Hofer T, Lasserre L, Saint-Pierre C, Favier A, Cadet J (2002) Cellular background level of 8-oxo-7,8-dihydro-2'-deoxyguanosine: an isotope based method to evaluate artefactual oxidation of DNA during its extraction and subsequent work-up. *Carcinogenesis* 23: 1911-8.
- Ray AM, Zuhlke KA, Levin AM, Douglas JA, Cooney KA, Petros JA (2009) Sequence variation in the mitochondrial gene cytochrome c oxidase subunit I and prostate cancer in African American men. *Prostate* 69: 956-60.
- Rea IM, McNerlan SE, Archbold GP, Middleton D, Curran MD, Young IS, Ross OA (2013) Mitochondrial J haplogroup is associated with lower blood pressure and anti-oxidant status: findings in octo/nonagenarians from the BELFAST Study. *Age (Dordr)* 35: 1445-56.
- Rebelo AP, Dillon LM, Moraes CT (2011) Mitochondrial DNA transcription regulation and nucleoid organization. *J Inherit Metab Dis* 34: 941-51.
- Refinetti P, Arstad C, Thilly WG, Morgenthaler S, Ekstrom PO (2017) Mapping mitochondrial heteroplasmy in a Leydig tumor by laser capture micro-dissection and cycling temperature capillary electrophoresis. *BMC Clin Pathol* 17: 6. doi: 10.1186/s12907-017-0042-3
- Reuter S, Gupta SC, Chaturvedi MM, Aggarwal BB (2010) Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic Biol Med* 49: 1603-16.
- Rigoli L, Di Bella C, Verginelli F, Falchetti M, Bersiga A, Rocco A, Nardone G, Mariani-Costantini R, Caruso RA (2008) Histological heterogeneity and somatic mtDNA mutations in gastric intraepithelial neoplasia. *Mod Pathol* 21: 733-41.
- Rinaldi M, Caffo M, Minutoli L, Marini H, Abbritti RV, Squadrito F, Trichilo V, Valenti A, Barresi V, Altavilla D, Passalacqua M, Caruso G (2016) ROS and Brain Gliomas: An Overview of Potential and Innovative Therapeutic Strategies. *Int J Mol Sci* 17.
- Romiguier J, Roux C (2017) Analytical Biases Associated with GC-Content in Molecular Evolution. *Front Genet* 8: 16.
- Rose G, Passarino G, Scornaienchi V, Romeo G, Dato S, Bellizzi D, Mari V, Feraco E, Maletta R, Bruni A, Franceschi C, De Benedictis G (2007) The mitochondrial DNA control region shows genetically

- correlated levels of heteroplasmy in leukocytes of centenarians and their offspring. *BMC Genomics* 8: 293. doi: 10.1186/1471-2164-8-293
- Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB (2013) Characterizing and measuring bias in sequence data. *Genome Biol* 14: R51.
- Ruiz-Pesini E, Lapena AC, Diez-Sanchez C, Perez-Martos A, Montoya J, Alvarez E, Diaz M, Urries A, Montoro L, Lopez-Perez MJ, Enriquez JA (2000) Human mtDNA haplogroups associated with high or reduced spermatozoa motility. *Am J Hum Genet* 67: 682-96.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303: 223-6.
- Salas A, Elson JL (2015) Mitochondrial DNA as a risk factor for false positives in case-control association studies. *J Genet Genomics* 42: 169-72.
- Salas A, Yao YG, Macaulay V, Vega A, Carracedo A, Bandelt HJ (2005) A critical reassessment of the role of mitochondria in tumorigenesis. *PLoS Med* 2: e296.
- Salvioli S, Capri M, Santoro A, Raule N, Sevini F, Lukas S, Lanzarini C, Monti D, Passarino G, Rose G, De Benedictis G, Franceschi C (2008) The impact of mitochondrial DNA on human lifespan: a view from studies on centenarians. *Biotechnol J* 3: 740-9.
- Samuels DC, Li C, Li B, Song Z, Torstenson E, Boyd Clay H, Rokas A, Thornton-Wells TA, Moore JH, Hughes TM, Hoffman RD, Haines JL, Murdock DG, Mortlock DP, Williams SM (2013) Recurrent tissue-specific mtDNA mutations are common in humans. *PLoS Genet* 9: e1003929.
- Santandreu FM, Brell M, Gene AH, Guevara R, Oliver J, Couce ME, Roca P (2008) Differences in mitochondrial function and antioxidant systems between regions of human glioma. *Cell Physiol Biochem* 22: 757-68. doi: 10.1159/000185559
- Santoro A, Balbi V, Balducci E, Pirazzini C, Rosini F, Tavano F, Achilli A, Siviero P, Minicuci N, Bellavista E, Mishto M, Salvioli S, Marchegiani F, Cardelli M, Olivieri F, Nacmias B, Chiamanti AM, Benussi L, Ghidoni R, Rose G, Gabelli C, Binetti G, Sorbi S, Crepaldi G, Passarino G, Torrioni A, Franceschi C (2010) Evidence for sub-haplogroup h5 of mitochondrial DNA as a risk factor for late onset Alzheimer's disease. *PLoS One* 5: e12037.
- Santorsola M, Calabrese C, Girolimetti G, Diroma MA, Gasparre G, Attimonelli M (2015) A multi-parametric workflow for the prioritization of mitochondrial DNA variants of clinical interest. *Hum Genet* 135: 121-36.
- Santos C, Montiel R, Arruda A, Alvarez L, Aluja MP, Lima M (2008a) Mutation patterns of mtDNA: empirical inferences for the coding region. *BMC Evol Biol* 8: 167.

- Santos C, Montiel R, Sierra B, Bettencourt C, Fernandez E, Alvarez L, Lima M, Abade A, Aluja MP (2005) Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal). *Mol Biol Evol* 22: 1490-505.
- Santos C, Sierra B, Alvarez L, Ramos A, Fernandez E, Nogues R, Aluja MP (2008b) Frequency and pattern of heteroplasmy in the control region of human mitochondrial DNA. *J Mol Evol* 67: 191-200.
- Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C (2016) Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics* 17: 125.
- Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* 43: e37.
- Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. *N Engl J Med* 347: 576-80.
- Schwartz RM, Dayhoff MO (1978) Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* 199: 395-403.
- Schwartzbaum JA, Fisher JL, Aldape KD, Wrensch M (2006) Epidemiology and molecular pathology of glioma. *Nat Clin Pract Neurol* 2: 494-503; quiz 1 p following 516.
- Seoane M, Mosquera-Miguel A, Gonzalez T, Fraga M, Salas A, Costoya JA (2011) The mitochondrial genome is a "genetic sanctuary" during the oncogenic process. *PLoS One* 6: e23327.
- Sharma H, Singh A, Sharma C, Jain SK, Singh N (2005) Mutations in the mitochondrial DNA D-loop region are frequent in cervical cancer. *Cancer Cell Int* 5: 34.
- Sharma N, Pasala MS, Prakash A (2019) Mitochondrial DNA: Epigenetics and environment. *Environ Mol Mutagen*. doi: 10.1002/em.22319
- Shen J, Platek M, Mahasneh A, Ambrosone CB, Zhao H (2010a) Mitochondrial copy number and risk of breast cancer: a pilot study. *Mitochondrion* 10: 62-8.
- Shen L, Fang H, Chen T, He J, Zhang M, Wei X, Xin Y, Jiang Y, Ding Z, Ji J, Lu J, Bai Y (2010b) Evaluating mitochondrial DNA in cancer occurrence and development. *Ann N Y Acad Sci* 1201: 26-33.
- Shen L, Wei J, Chen T, He J, Qu J, He X, Jiang L, Qu Y, Fang H, Chen G, Lu J, Bai Y (2011) Evaluating mitochondrial DNA in patients with breast cancer and benign breast disease. *J Cancer Res Clin Oncol* 137: 669-75. doi: 10.1007/s00432-010-0912-x
- Sherer TB, Betarbet R, Stout AK, Lund S, Baptista M, Panov AV, Cookson MR, Greenamyre JT (2002) An in vitro model of Parkinson's disease: linking mitochondrial impairment to altered alpha-synuclein metabolism and oxidative damage. *J Neurosci* 22: 7006-15. doi: 20026721
- Shidara Y, Yamagata K, Kanamori T, Nakano K, Kwong JQ, Manfredi G, Oda H, Ohta S (2005) Positive contribution of pathogenic mutations in the mitochondrial genome to the promotion of cancer by prevention from apoptosis. *Cancer Res* 65: 1655-63.

- Shigenaga MK, Hagen TM, Ames BN (1994) Oxidative damage and mitochondrial decay in aging. *Proc Natl Acad Sci U S A* 91: 10771-8.
- Shu HY, Li HC, Xie WQ, Ni B, Zhou HY (2019) Mitochondrial DNA variations in tongue squamous cell carcinoma. *Biomed Rep* 10: 23-28.
- Siegel T (2015) Clinical impact of molecular biomarkers in gliomas. *J Clin Neurosci* 22: 437-44.
- Silkjaer T, Norgaard JM, Aggerholm A, Ebbesen LH, Kjeldsen E, Hokland P, Nyvold CG (2013) Characterization and prognostic significance of mitochondrial DNA variations in acute myeloid leukemia. *Eur J Haematol* 90: 385-96.
- Singh AK, Pandey P, Tewari M, Pandey HP, Shukla HS (2014) Human mitochondrial genome flaws and risk of cancer. *Mitochondrial DNA* 25: 329-34.
- Singh KK, Ayyasamy V, Owens KM, Koul MS, Vujcic M (2009) Mutations in mitochondrial DNA polymerase-gamma promote breast tumorigenesis. *J Hum Genet* 54: 516-24.
- Singh RR, Rajyalakshmi L, Routbort MJ, Patel KP, Medeiros JL (2016) Implementation of next generation sequencing in clinical molecular diagnostic laboratories: advantages, challenges and potential. *Expert Rev Precis Med Drug Dev* 1: 109-120.
- Skonieczna K, Malyarchuk B, Jawien A, Marszalek A, Banaszkiwicz Z, Jarmocik P, Borcz M, Bala P, Grzybowski T (2015) Heteroplasmic substitutions in the entire mitochondrial genomes of human colon cells detected by ultra-deep 454 sequencing. *Forensic Sci Int Genet* 15: 16-20.
- Skonieczna K, Malyarchuk B, Jawien A, Marszalek A, Banaszkiwicz Z, Jarmocik P, Grzybowski T (2018) Mitogenomic differences between the normal and tumor cells of colorectal cancer patients. *Hum Mutat* 39: 691-701. doi: 10.1002/humu.23402
- Sonney S, Leipzig J, Lott MT, Zhang S, Procaccio V, Wallace DC, Sondheimer N (2017) Predicting the pathogenicity of novel variants in mitochondrial tRNA with MitoTIP. *PLoS Comput Biol* 13: e1005867. doi: 10.1371/journal.pcbi.1005867
- Soon BH, Abdul Murad NA, Then SM, Abu Bakar A, Fadzil F, Thanabalan J, Mohd Haspani MS, Toh CJ, Mohd Tamil A, Harun R, Wan Ngah WZ, Jamal R (2017) Mitochondrial DNA Mutations in Grade II and III Glioma Cell Lines Are Associated with Significant Mitochondrial Dysfunction and Higher Oxidative Stress. *Front Physiol* 8: 231.
- Stafford P, Chen-Quin EB (2010) The pattern of natural selection in somatic cancer mutations of human mtDNA. *J Hum Genet* 55: 605-12.
- Stefano GB, Kream RM (2016) Mitochondrial DNA heteroplasmy in human health and disease. *Biomed Rep* 4: 259-262.
- Stewart JB, Chinnery PF (2015) The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet* 16: 530-42.

- Strickland M, Stoll EA (2017) Metabolic Reprogramming in Glioma. *Front Cell Dev Biol* 5: 43.
- Su X, Wang W, Ruan G, Liang M, Zheng J, Chen Y, Wu H, Fahey TJ, Guan M, Teng L (2016) A Comprehensive Characterization of Mitochondrial Genome in Papillary Thyroid Cancer. *Int J Mol Sci* 17.
- Suzuki H, Hosokawa Y, Nishikimi M, Ozawa T (1991) Existence of common homologous elements in the transcriptional regulatory regions of human nuclear genes and mitochondrial gene for the oxidative phosphorylation system. *J Biol Chem* 266: 2333-8.
- Suzuki S, Takei M, Ehara T, Nishio S, Inaba H, Hashizume K (2009) Atypical early onset of diabetes, deafness and lung cancer in a male patient with mitochondrial mutations in peripheral mononuclear cells. *BMJ Case Rep* 2009.
- Syed F, Grunenwald H, Caruccio N (2009) Next-generation sequencing library preparation: simultaneous fragmentation and tagging using in vitro transposition. *Nat Methods* 6: 783-856.
- Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1410: 103-23.
- Tan D, Goerlitz DS, Dumitrescu RG, Han D, Seillier-Moisewitsch F, Spornak SM, Orden RA, Chen J, Goldman R, Shields PG (2008) Associations between cigarette smoking and mitochondrial DNA abnormalities in buccal cells. *Carcinogenesis* 29: 1170-7.
- Tan DJ, Bai RK, Wong LJ (2002) Comprehensive scanning of somatic mitochondrial DNA mutations in breast cancer. *Cancer Res* 62: 972-6.
- Tan DJ, Chang J, Liu LL, Bai RK, Wang YF, Yeh KT, Wong LJ (2006) Significance of somatic mutations and content alteration of mitochondrial DNA in esophageal cancer. *BMC Cancer* 6: 93.
- Tang S, Batra A, Zhang Y, Ebenroth ES, Huang T (2010) Left ventricular noncompaction is associated with mutations in the mitochondrial genome. *Mitochondrion* 10: 350-7. doi: 10.1016/j.mito.2010.02.003
- Tang S, Huang T (2010) Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. *Biotechniques* 48: 287-96.
- Taylor RW, McDonnell MT, Blakely EL, Chinnery PF, Taylor GA, Howell N, Zeviani M, Briem E, Carrara F, Turnbull DM (2003) Genotypes from patients indicate no paternal mitochondrial DNA contribution. *Ann Neurol* 54: 521-4.
- Theodoratou E, Din FV, Farrington SM, Cetnarskyj R, Barnetson RA, Porteous ME, Dunlop MG, Campbell H, Tenesa A (2010) Association between common mtDNA variants and all-cause or colorectal cancer mortality. *Carcinogenesis* 31: 296-301.
- Torrioni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144: 1835-50.

- Toure S, Mbaye F, Gueye MD, Fall M, Dem A, Sembene M (2019) Somatic Mitochondrial Mutations in Oral Cavity Cancers among Senegalese Patients. *Asian Pac J Cancer Prev* 20: 2203–2208. doi: 10.31557/APJCP.2019.20.7.2203
- Tseng LM, Yin PH, Chi CW, Hsu CY, Wu CW, Lee LM, Wei YH, Lee HC (2006) Mitochondrial DNA mutations and mitochondrial DNA depletion in breast cancer. *Genes Chromosomes Cancer* 45: 629–38.
- Tseng LM, Yin PH, Yang CW, Tsai YF, Hsu CY, Chi CW, Lee HC (2011) Somatic mutations of the mitochondrial genome in human breast cancers. *Genes Chromosomes Cancer* 50: 800–11.
- Tuppen HA, Blakely EL, Turnbull DM, Taylor RW (2010) Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta* 1797: 113–28.
- Tyagi A, Pramanik R, Vishnubhatla S, Ali S, Bakhshi R, Chopra A, Singh A, Bakhshi S (2018) Pattern of mitochondrial D-loop variations and their relation with mitochondrial encoded genes in pediatric acute myeloid leukemia. *Mutat Res* 810: 13–18. doi: 10.1016/j.mrfmmm.2018.05.002
- Tzen CY, Mau BL, Wu TY (2007) ND4 mutation in transitional cell carcinoma: Does mitochondrial mutation occur before tumorigenesis? *Mitochondrion* 7: 273–8.
- Ulasov IV, Lenz G, Lesniak MS (2018) Autophagy in glioma cells: An identity crisis with a clinical perspective. *Cancer Lett* 428: 139–146. doi: 10.1016/j.canlet.2018.04.034
- van den Bent M (2007) Anaplastic Oligodendroglioma and Oligoastrocytoma. *Neurol Clin* 25: 1089–1109.
- van der Walt JM, Dementieva YA, Martin ER, Scott WK, Nicodemus KK, Kroner CC, Welsh-Bohmer KA, Saunders AM, Roses AD, Small GW, Schmechel DE, Murali Doraiswamy P, Gilbert JR, Haines JL, Vance JM, Pericak-Vance MA (2004) Analysis of European mitochondrial haplogroups with Alzheimer disease risk. *Neurosci Lett* 365: 28–32.
- van der Walt JM, Nicodemus KK, Martin ER, Scott WK, Nance MA, Watts RL, Hubble JP, Haines JL, Koller WC, Lyons K, Pahwa R, Stern MB, Colcher A, Hiner BC, Jankovic J, Ondo WG, Allen FHJ, Goetz CG, Small GW, Mastaglia F, Stajich JM, McLaurin AC, Middleton LT, Scott BL, Schmechel DE, Pericak-Vance MA, Vance JM (2003) Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am J Hum Genet* 72: 804–811.
- van Gisbergen MW, Voets AM, Starmans MH, de Coo IF, Yadak R, Hoffmann RF, Boutros PC, Smeets HJ, Dubois L, Lambin P (2015) How do changes in the mtDNA and mitochondrial dysfunction influence cancer and cancer therapy? Challenges, opportunities and models. *Mutat Res Rev Mutat Res* 764: 16–30.
- Van Houten B, Woshner V, Santos JH (2006) Role of mitochondrial DNA in toxic responses to oxidative stress. *DNA Repair (Amst)* 5: 145–52.

- van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30: E386-94.
- Van Trappen PO, Cullup T, Troke R, Swann D, Shepherd JH, Jacobs IJ, Gayther SA, Mein CA (2007) Somatic mitochondrial DNA mutations in primary and metastatic ovarian cancer. *Gynecol Oncol* 104: 129-33.
- Vander Heiden MG, Cantley LC, Thompson CB (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324: 1029-33. doi: 10.1126/science.1160809
- Vara-Perez M, Felipe-Abrio B, Agostinis P (2019) Mitophagy in Cancer: A Tale of Adaptation. *Cells* 8. doi: 10.3390/cells8050493
- Vega A, Salas A, Gamborino E, Sobrido MJ, Macaulay V, Carracedo A (2004) mtDNA mutations in tumors of the central nervous system reflect the neutral evolution of mtDNA in populations. *Oncogene* 23: 1314-20.
- Vellarikkal SK, Dhiman H, Joshi K, Hasija Y, Sivasubbu S, Scaria V (2015) mit-o-matic: a comprehensive computational pipeline for clinical evaluation of mitochondrial variations from next-generation sequencing datasets. *Hum Mutat* 36: 419-24.
- Venter M, Malan L, van Dyk E, Elson JL, van der Westhuizen FH (2017) Using MutPred derived mtDNA load scores to evaluate mtDNA variation in hypertension and diabetes in a two-population cohort: The SABPA study. *J Genet Genomics* 44: 139-149.
- Vidone M, Clima R, Santorsola M, Calabrese C, Girolimetti G, Kurelac I, Amato LB, Iommarini L, Trevisan E, Leone M, Soffietti R, Morra I, Faccani G, Attimonelli M, Porcelli AM, Gasparre G (2015) A comprehensive characterization of mitochondrial DNA mutations in glioblastoma multiforme. *Int J Biochem Cell Biol* 63: 46-54.
- Vivekanandan P, Daniel H, Yeh MM, Torbenson M (2010) Mitochondrial mutations in hepatocellular carcinomas and fibrolamellar carcinomas. *Mod Pathol* 23: 790-8.
- Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 55: 641-58.
- Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* 39: 359-407.
- Wallace DC (2010) Mitochondrial DNA mutations in disease and aging. *Environ Mol Mutagen* 51: 440-50.
- Wallace DC (2013) A mitochondrial bioenergetic etiology of disease. *J Clin Invest* 123: 1405-12. doi: 10.1172/JCI61398
- Wallace DC (2015) Mitochondrial DNA variation in human radiation and disease. *Cell* 163: 33-8.
- Wallace DC, Chalkia D (2013) Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb Perspect Biol* 5: a021220.

- Wang CY, Li H, Hao XD, Liu J, Wang JX, Wang WZ, Kong QP, Zhang YP (2011) Uncovering the profile of somatic mtDNA mutations in Chinese colorectal cancer patients. *PLoS One* 6: e21613.
- Wang Y, Liu VW, Xue WC, Cheung AN, Ngan HY (2006) Association of decreased mitochondrial DNA content with ovarian cancer progression. *Br J Cancer* 95: 1087–91.
- Wang Y, Liu VW, Xue WC, Tsang PC, Cheung AN, Ngan HY (2005) The increase of mitochondrial DNA content in endometrial adenocarcinoma cells: a quantitative study using laser-captured microdissected tissues. *Gynecol Oncol* 98: 104–10.
- Wang Z, Choi S, Lee J, Huang YT, Chen F, Zhao Y, Lin X, Neuberger D, Kim J, Christiani DC (2015) Mitochondrial Variations in Non-Small Cell Lung Cancer (NSCLC) Survival. *Cancer Inform* 14: 1–9. doi: 10.4137/CIN.S13976
- Wanka C, Steinbach JP, Rieger J (2012) Tp53-induced glycolysis and apoptosis regulator (TIGAR) protects glioma cells from starvation-induced cell death by up-regulating respiration and improving cellular redox homeostasis. *J Biol Chem* 287: 33436–46. doi: 10.1074/jbc.M112.384578
- Warburg O (1956) On the origin of cancer cells. *Science* 123: 309–14.
- Warowicka A, Kwasniewska A, Gozdicka-Jozefiak A (2013) Alterations in mtDNA: a qualitative and quantitative study associated with cervical cancer development. *Gynecol Oncol* 129: 193–8.
- Weissensteiner H, Forer L, Fuchsberger C, Schopf B, Kloss-Brandstatter A, Specht G, Kronenberg F, Schonherr S (2016a) mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res* 44: W64–9.
- Weissensteiner H, Pacher D, Kloss-Brandstatter A, Forer L, Specht G, Bandelt HJ, Kronenberg F, Salas A, Schonherr S (2016b) HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* 44: W58–63.
- Wesseling P, van den Bent M, Perry A (2015) Oligodendroglioma: pathology, molecular mechanisms and markers. *Acta Neuropathol* 129: 809–27.
- White EJ, Ross T, Lopez E, Nikiforov A, Gault C, Batorsky R, Darcy C, Campagna D, Fleming M, Thompson J (2017) Chasing a moving target: Detection of mitochondrial heteroplasmy for clinical diagnostics. <https://doi.org/10.1101/222109>
- Williams SB, Ye Y, Huang M, Chang DW, Kamat AM, Pu X, Dinney CP, Wu X (2015) Mitochondrial DNA Content as Risk Factor for Bladder Cancer and Its Association with Mitochondrial DNA Polymorphisms. *Cancer Prev Res (Phila)* 8: 607–13.
- Witas HW, Zawicki P (2004) Mitochondrial DNA and human evolution: A review. *Prz Antropol-Anthropol Rev* 67: 97–110.
- Wong LJ (2007a) Diagnostic challenges of mitochondrial DNA disorders. *Mitochondrion* 7: 45–52.

- Wong LJ (2007b) Pathogenic mitochondrial DNA mutations in protein-coding genes. *Muscle Nerve* 36: 279-93.
- Wu CW, Yin PH, Hung WY, Li AF, Li SH, Chi CW, Wei YH, Lee HC (2005) Mitochondrial DNA mutations and mitochondrial DNA depletion in gastric cancer. *Genes Chromosomes Cancer* 44: 19-28.
- Xing J, Chen M, Wood CG, Lin J, Spitz MR, Ma J, Amos CI, Shields PG, Benowitz NL, Gu J, de Andrade M, Swan GE, Wu X (2008) Mitochondrial DNA content: its genetic heritability and association with renal cell carcinoma. *J Natl Cancer Inst* 100: 1104-12.
- Xu L, Hu Y, Chen B, Tang W, Han X, Yu H, Xiao C (2006) Mitochondrial polymorphisms as risk factors for endometrial cancer in southwest China. *Int J Gynecol Cancer* 16: 1661-7.
- Yang Ai SS, Hsu K, Herbert C, Cheng Z, Hunt J, Lewis CR, Thomas PS (2013) Mitochondrial DNA mutations in exhaled breath condensate of patients with lung cancer. *Respir Med* 107: 911-8.
- Yang Y, Gao H, Zhou H, Liu Q, Qi Z, Zhang Y, Zhang J (2019) The role of mitochondria-derived peptides in cardiovascular disease: Recent updates. *Biomed Pharmacother* 117: 109075. doi: 10.1016/j.biopha.2019.109075
- Ye F, Samuels DC, Clark T, Guo Y (2014a) High-throughput sequencing in mitochondrial DNA research. *Mitochondrion* 17: 157-63.
- Ye K, Lu J, Ma F, Keinan A, Gu Z (2014b) Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc Natl Acad Sci U S A* 111: 10654-9.
- Yeung KY, Dickinson A, Donoghue JF, Polekhina G, White SJ, Grammatopoulos DK, McKenzie M, Johns TG, St John JC (2014) The identification of mitochondrial DNA variants in glioblastoma multiforme. *Acta Neuropathol Commun* 2: 1.
- Yin C, Li DY, Guo X, Cao HY, Chen YB, Zhou F, Ge NJ, Liu Y, Guo SS, Zhao Z, Yang HS, Xing JL (2019) NGS-based profiling reveals a critical contributing role of somatic D-loop mtDNA mutations in HBV-related hepatocarcinogenesis. *Ann Oncol*. doi: 10.1093/annonc/mdz105
- Youle RJ, van der Bliek AM (2012) Mitochondrial fission, fusion, and stress. *Science* 337: 1062-5.
- Yu C, Wang X, Huang L, Tong Y, Chen L, Wu H, Xia Q, Kong X (2018) Deciphering the Spectrum of Mitochondrial DNA Mutations in Hepatocellular Carcinoma Using High-Throughput Sequencing. *Gene Expr* 18: 125-134. doi: 10.3727/105221618X15185539348147
- Yu M, Wan Y, Zou Q (2010) Decreased copy number of mitochondrial DNA in Ewing's sarcoma. *Clin Chim Acta* 411: 679-83.
- Yu M, Zhou Y, Shi Y, Ning L, Yang Y, Wei X, Zhang N, Hao X, Niu R (2007) Reduced mitochondrial DNA copy number is correlated with tumor progression and prognosis in Chinese breast cancer patients. *IUBMB Life* 59: 450-7.

- Zapico SC (2016) *Mechanisms Linking Aging, Diseases and Biological Age Estimation*. CRC Press, Boca Raton, FL
- Zhang J, Asin-Cayuela J, Fish J, Michikawa Y, Bonafe M, Olivieri F, Passarino G, De Benedictis G, Franceschi C, Attardi G (2003) Strikingly higher frequency in centenarians and twins of mtDNA mutation causing remodeling of replication origin in leukocytes. *Proc Natl Acad Sci U S A* 100: 1116–21. doi: 10.1073/pnas.242719399
- Zhang R, Zhang F, Wang C, Wang S, Shiao YH, Guo Z (2010) Identification of sequence polymorphism in the D-Loop region of mitochondrial DNA as a risk factor for hepatocellular carcinoma with distinct etiology. *J Exp Clin Cancer Res* 29: 130.
- Zhang TH, Wu NC, Sun R (2016) A benchmark study on error-correction by read-pairing and tag-clustering in amplicon-based deep sequencing. *BMC Genomics* 17: 108.
- Zhao Y, Wang J, Chen X, Wu Q (2010) Significance of mutation in the D-loop region in cervical cancer. *West Indian Med J* 59: 291–4.
- Zheng S, Qian P, Li F, Qian G, Wang C, Wu G, Li Q, Chen Y, Li J, Li H, He B, Ji F (2012) Association of mitochondrial DNA variations with lung cancer risk in a Han Chinese population from southwestern China. *PLoS One* 7: e31322.
- Zhidkov I, Nagar T, Mishmar D, Rubin E (2011) MitoBamAnnotator: A web-based tool for detecting and annotating heteroplasmy in human mitochondrial DNA sequences. *Mitochondrion* 11: 924–8.
- Zhou S, Kachhap S, Sun W, Wu G, Chuang A, Poeta L, Grumbine L, Mithani SK, Chatterjee A, Koch W, Westra WH, Maitra A, Glazer C, Carducci M, Sidransky D, McFate T, Verma A, Califano JA (2007) Frequency and phenotypic implications of mitochondrial DNA mutations in human squamous cell cancers of the head and neck. *Proc Natl Acad Sci U S A* 104: 7540–5.
- Zimmerman S, Sutovsky P (2009) The sperm proteasome during sperm capacitation and fertilization. *J Reprod Immunol* 83: 19–25.
- Zong WX, Rabinowitz JD, White E (2016) Mitochondria and Cancer. *Mol Cell* 61: 667–76.
- Zsurka G, Kraysberg Y, Kudina T, Kornblum C, Elger CE, Khrapko K, Kunz WS (2005) Recombination of mitochondrial DNA in skeletal muscle of individuals with multiple mitochondrial DNA heteroplasmy. *Nat Genet* 37: 873–7.

10 ANEXOS

10. ANEXOS

10.1. ANEXO I

Tabla 1. Localización y tamaño de las diferentes regiones dentro del genoma mitocondrial humano (NC_012920.1). pb: pares de bases. HVRI: región hipervariable I; HVRII: región hipervariable II; HVRIII: región hipervariable III; MT-TA: RNAt alanina; MT-ATP6: ATP sintasa 6; MT-ATP8: ATP sintasa 8; MT-TC: RNAt cisteína; MT-CYB: citocromo b; MT-CO1: citocromo c oxidasa 1; MT-CO2: citocromo oxidasa 2; MT-CO3: citocromo oxidasa 3; MT-TD: RNAt ácido aspártico; MT-TE: RNAt ácido glutámico; MT-TF: RNAt fenilalanina; MT-TG: RNAt glicina; MT-TH: RNAt histidina; MT-TI: RNAt isoleucina; MT-TK: RNAt lisina; MT-TL1: RNAt leucina 1; MT-TL2: RNAt leucina2; MT-TM: RNAt metionina; MT-TN: RNAt asparagina; MT-ND1: NADH deshidrogenasa 1; MT-ND2: NADH deshidrogenasa 2; MT-ND3: NADH deshidrogenasa 3; MT-ND4: NADH deshidrogenasa 4; MT-ND4L: NADH deshidrogenasa 4L; MT-ND5: NADH deshidrogenasa 5; MT-ND6: NADH deshidrogenasa 6; MT-TQ: RNAt glutamina; MT-TR: RNAt arginina; MT-TS1: RNAt serina 1; MT-TS2: RNAt serina 2; MT-TV: RNAt valina; MT-TW: RNAt triptófano; MT-TY: RNAt tirosina; 12S: subunidad 12S del RNAr; 16S: subunidad 16S del RNAr. pb: pares de bases.

Región mitocondrial		Subregión mitocondrial	Inicio	Final	Total (pb)
Región control		HVRI	16024	16383	360
		HVRII	57	372	316
		HVRIII	438	574	137
Región Codificante	RNAr	12S	648	1601	954
		16S	1671	3229	1559
	RNAt	MT-TF	577	647	71
		MT-TV	1602	1670	69
		MT-TL1	3230	3304	75
		MT-TI	4263	4331	69
		MT-TQ	4329	4400	72
		MT-TM	4402	4469	68
		MT-TW	5512	5579	68
		MT-TA	5587	5655	69
		MT-TN	5657	5729	73
		MT-TC	5761	5826	66
		MT-TY	5826	5891	66
		MT-TS1	7446	7514	69
		MT-TD	7518	7585	68
		MT-TK	8295	8364	70
		MT-TG	9991	10058	68

Tabla 1. Continuación.

Región mitocondrial		Subregión mitocondrial	Inicio	Final	Total (pb)	
Región Codificante	RNAs	RNA _t	MT-TR	10405	10469	65
			MT-TH	12138	12206	69
			MT-TS2	12207	12265	59
			MT-TL2	12266	12336	71
			MT-TE	14674	14742	69
			MT-TT	15888	15953	66
			MT-TP	15956	16023	68
	Subunidades de la cadena respiratoria	Complejo I	MT-ND1	3307	4262	956
			MT-ND2	4470	5511	1042
			MT-ND3	10059	10404	346
			MT-ND4L	10470	10766	297
			MT-ND4	10760	12137	1378
			MT-ND5	12337	14148	1812
		Complejo III	MT-CYB	14747	15887	1141
		Complejo IV	MT-CO1	5904	7445	1542
			MT-CO2	7586	8269	684
			MT-CO3	9207	9990	784
		Complejo V	MT-ATP6	8527	9207	681
			MT-ATP8	8366	8572	207

10.2. ANEXO II

a

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6520      6530      6540      6550      6560
Sec_mit_Q1  CTGCTGGCAT CACTATACTA CTAACAGACC GCAACCTCAA CACCACCTTC
NUMT_Chr1   CCGCTGGCAT CACTATACTA CTAACAGACC GTAACCTCAA CACCACCTTC

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6570      6580      6590      6600      6610
Sec_mit_Q1  TTCGACCCG CCGGAGGAGG AGACCCCAT CTATACCAAC ACCTATTCTG
NUMT_Chr1   TTCGACCCA CCGGAGGAGG AGACCCCAT CTATACCAAC ACCTATTCTG

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6620      6630      6640      6650      6660
Sec_mit_Q1  ATTTTTCGGT CACCCTGAAG TTTATATTCT TATCCTACCA GGCTTCGGAA
NUMT_Chr1   ATTTTTCGGT CACCCTGAAG TTTATATTCT CATCCTACCA GGCTTCGGAA

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6670      6680      6690      6700      6710
Sec_mit_Q1  TAATCTCCCA TATTGTAACT TACTACTCCG G-AAAAAAG AACCATTTGG
NUMT_Chr1   TAATCTCCCA TATTGTAACT TACTACTCCG GGAAAAAAG AACCATTTGG

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6720      6730      6740      6750      6760
Sec_mit_Q1  ATACATAGGT ATGGTCTGAG CTATGATATC AATTGGCTTC CTAGGGTTTA
NUMT_Chr1   ATACATAGGT ATGGTCTGAG CTATGATATC AATTGGCTTC CTAGGGTTTA

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6770      6780      6790      6800      6810
Sec_mit_Q1  TCGTGTGAGC ACACCATATA TTTACAGTAG GAATAGACGT AGACACACGA
NUMT_Chr1   TCGTGTGAGC ACACCATATA TTTACAGTAG GAATAGACGT AGACACACGA

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6820      6830      6840      6850      6860
Sec_mit_Q1  GCATATTTCA CCTCCGCTAC CATAATCATC GCTATCCCCA CCGGCGTCAA
NUMT_Chr1   GCATATTTCA CCTCCGCTAC CATAATCATC GCTATCCCCA CCGGCGTCAA

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6870      6880      6890      6900      6910
Sec_mit_Q1  AGTATTTAGC TGA CTCCGCA CACTCCACGG AAGCAATATG AAATGATCTG
NUMT_Chr1   AGTATTTAGC TGA CTCCGCA CACTCCACGG AAGCAATATG AAATGATCTG

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6920      6930      6940      6950      6960
Sec_mit_Q1  CTGCAGTGCT CTGAGCCCTA GGATTGATCT TTCTTTTCAC CGTAGGTGGC
NUMT_Chr1   CTGCAGTGCT CTGAGCCCTA GGATTGATTT TTCTTTTCAC CGTAGGTGGC

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
          6970      6980      6990      7000      7010
Sec_mit_Q1  CTGACTGGCA TTGTATTAGC AAAC TCATCA CTAGACATCG TACTACACGA
NUMT_Chr1   CTGACTGGCA TTGTATTAGC AAAC TCATCA CTAGACATCG TACTACACGA

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7020      7030      7040      7050      7060
Sec_mit_Q1  CACGTACTAC GTTGTAGCCC ACTTCCACTA TGTCCTATCA ATAGGAGCTG
NUMT_Chrl  CACGTACTAC GTTGTAGCCC ACTTCCACTA TGTCCTATCA ATAGGAGCTG

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7070      7080      7090      7100      7110
Sec_mit_Q1  TATTTGCCAT CATAGGAGGC TTCATTCACT GATTTCCCCT ATTCTCAGGC
NUMT_Chrl  TATTTGCCAT CATAGGAGGC TTCATTCACT GATTTCCCCT ATTCTCAGGC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7120      7130      7140      7150      7160
Sec_mit_Q1  TACACCCTAG ACCAAACCTA CGCCAAAATC CATTTCAACTA TCATATTCAT
NUMT_Chrl  TACACCCTAG ACCAAACCTA CGCCAAAATC CATTTCGGCTA TCATATTCAT

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7170      7180      7190      7200      7210
Sec_mit_Q1  CGGCGTAAAT CTAACTTTTCT TCCCACAACA CTTTCTCGGC CTATCCGGAA
NUMT_Chrl  CGGCGTAAAT CTAACTTTTCT TCCCACAACA CTTTCTCGGC CTATCCGGAA

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7220      7230      7240      7250      7260
Sec_mit_Q1  TGCCCCGACG TTACTCGGAC TACCCGATG CATACACCAC ATGAAACATC
NUMT_Chrl  TGCCCCGACG TTACTCGGAC TACCCGATG CATACACCAC ATGAAATATC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7270      7280      7290      7300      7310
Sec_mit_Q1  CTATCATCTG TAGGCTCATT CATTTTCTCTA ACAGCAGTAA TATTAATAAT
NUMT_Chrl  CTATCATCTG TAGGCTCATT CATTTTCTCTA ACAGCAGTAA TATTAATAAT

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7320      7330      7340      7350      7360
Sec_mit_Q1  TTTCATGAATT TGAGAAGCCT TCGCTTCGAA GCGAAAAGTC CTAATAGTAG
NUMT_Chrl  TTTCATAAAATT TGAGAAGCCT TCGCTTCGAA GCGAAAAGTC CTAATAGTAG

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7370      7380      7390      7400      7410
Sec_mit_Q1  AAGAACCCTC CATAAACCTG GAGTGACTAT ATGGATGCCC CCCACCCTAC
NUMT_Chrl  AAGAACCCTC CATAAACCTG GAGTGACTAT ATGGATGCCC CCCACCCTAC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7420      7430      7440      7450      7460
Sec_mit_Q1  CACACATTCG AAGAACCCTG ATACATAAAA TCTAGACAAA AAAGGAAGGA
NUMT_Chrl  CACACATTCG AAGAACCCTG ATACATAAAA TCTAGACAAA AAAGGAAGGA

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7470      7480      7490      7500      7510
Sec_mit_Q1  ATCGAACCCC CCAAAGCTGG TTTCAAGCCA ACCCCATGGC CTCCATGACT
NUMT_Chrl  ATCGAACCCC CCAAAGCTGG TTTCAAGCCA ACCCCATGGC CTCCATGACT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7520      7530      7540      7550      7560
Sec_mit_Q1 TTTTCAAAAA GTATTAGAA AAACCATTTC ATAAC TTTGT CAAAGTTAAA
NUMT_Chr1  TTTTCAAAAA GATTATTAGAA AAACCATTTC ATAAC TTTGT CAAAGTTAAA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7570      7580      7590      7600      7610
Sec_mit_Q1 TTATAGGCTA AATCCTATAT ATCTTAATGG CACATGCAGC GCAAGTAGGT
NUMT_Chr1  TTATAGGCTA AATCCTATAT ATCTTAATGG CACATGCAGC GCAAGTAGGT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7620      7630      7640      7650      7660
Sec_mit_Q1 CTACAAGACG CTA CTTCCCC TATCATAGAA GAGCTTATCA CTTTCATGA
NUMT_Chr1  CTACAAGACG CTA CTTCCCC TATCATAGAA GAGCTTATCA TCTTCATGA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7670      7680      7690      7700      7710
Sec_mit_Q1 TCACGCCCTC ATAATCATTT TCCTTATCTG CTTCTAGTC CTGTATGCC
NUMT_Chr1  TCACGCCCTC ATAATCATTT TCCTTATCTG CTTCTAGTC CTGTATGCC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7720      7730      7740      7750      7760
Sec_mit_Q1 TTTTCCTAAC ACTCACAACA AAAC TAACTA ATACTAACAT CTCAGACGCT
NUMT_Chr1  TTTTCCTAAC ACTCACAACA AAAC TAACTA ATACTAACAT CTCAGACGCT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7770      7780      7790      7800      7810
Sec_mit_Q1 CAGGAAATAG AAACCGTCTG AACTATCCTG CCCGCCATCA TCCTAGTCCT
NUMT_Chr1  CAGGAAATAG AAACCGTCTG AACTATCCTG CCCGCCATCA TCCTAGTCCT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7820      7830      7840      7850      7860
Sec_mit_Q1 CATCGCCCTC CCATCCCTAC GCATCCTTTA CATAACAGAC GAGGTCAACG
NUMT_Chr1  CATCGCCCTC CCATCCCTAC GCATCCTTTA CATAACAGAC GAGGTCAACG

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7870      7880      7890      7900      7910
Sec_mit_Q1 ATCCCTCCCT TACCATCAAAA TCAATTGGCC ACCAATGGTA CTGAACCTAC
NUMT_Chr1  ATCCCTCCCT TACCATCAAAA TCAATTGGCC ATCAATGGTA CTGAACCTAC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7920      7930      7940      7950      7960
Sec_mit_Q1 GAGTACACCG ACTACGGCGG ACTAATCTTC AACTCCTACA TACTTCCCC
NUMT_Chr1  GAGTACACCG ACTACGGCGG ACTAATCTTC AACTCCTACA TACTTCCCC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|
      7970      7980      7990      8000      8010
Sec_mit_Q1 ATTATTCCTA GAACCAGGCG ACCTGCGACT CCTTGACGTT GACAATCGAG
NUMT_Chr1  ATTATTCCTA GAACCAGGCG ACCTGCGACT CCTTGACGTT GACAATCGAG

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8020      8030      8040      8050      8060
Sec_mit_Q1 TAGTACTCCC GATTGAAGCC CCCATTTCGTA TAATAATTAC ATCACAAGAC
NUMT_Chrl  TAGTACTCCC GATTGAAGCC CCCATTTCGTA TAATAATTAC ATCACAAGAC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8070      8080      8090      8100      8110
Sec_mit_Q1 GTCTTGCACT CATGAGCTGT CCCACACATTA GGCTTAAAAA CAGATGCAAT
NUMT_Chrl  GTCTTACACT CATGAGCTGT CCCACACATTA GGCTTAAAAA CAGATGCAAT

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8120      8130      8140      8150      8160
Sec_mit_Q1 TCCCGGACGT CTAAACCAA CCACTTTCAC CGCTACACGA CCGGGGTAT
NUMT_Chrl  TCCCGGACGT CTAAACCAA CCACTTTCAC TGCTACACGA CCAGGGGTAT

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8170      8180      8190      8200      8210
Sec_mit_Q1 ACTACGGTCA ATGCTCTGAA ATCTGTGGAG CAAACCACAG TTTCATGCC
NUMT_Chrl  ACTACGGCCA ATGCTCTGAA ATCTGTGGAG CAAACC--AG TTTTATGCC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8220      8230      8240      8250      8260
Sec_mit_Q1 ATCGTCCTAG AATTAATTCC CCTAAAAATC TTTGAAATAG GGCCCGTATT
NUMT_Chrl  ATCGTCCTAG AATTAATTCC CCTAAAAATC TTTGAAATAG GGCCCGTATT

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8270      8280      8290      8300      8310
Sec_mit_Q1 TACCCTATAG CACCCCTCT ACCCCCTCTA GAGCCCACTG TAAAGCTAAC
NUMT_Chrl  TACCCTATAG CACCCCTCT ACCCCCTCTA GAGCCCACTG TAAAGCTAAC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8320      8330      8340      8350      8360
Sec_mit_Q1 TTAGCATTAA CCTTTTAAGT TAAAGATTAA GAGAACCAAC ACCTCTTTAC
NUMT_Chrl  TTAGCATTAA CCTTTTAAGT TAAAGATTAA GAGAACCAAC ACCTCTTTAC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8370      8380      8390      8400      8410
Sec_mit_Q1 AGTGAAATGC CCCAACTAAA TACTACCGTA TGGCCCAACA TAATTACCC
NUMT_Chrl  AGTGAAATGC CCCAACTAAA TACTACCGTA TGACCCCAACA TAATTACCC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8420      8430      8440      8450      8460
Sec_mit_Q1 CATACTCCTT AACTATTCC TCATCACCCA ACTAAAAATA TTAACACAA
NUMT_Chrl  CATACTCCTT AACTATTCC TCATCACCCA ACTAAAAATA TTAACATCAA

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8470      8480      8490      8500      8510
Sec_mit_Q1 ACTACCACCT ACCTCCCTCA CCAAAGCCCA TAAAAATAAA AAATTATAAC
NUMT_Chrl  ATTACCACCT ACCTCCCTCA CCAAAGCCCA TAAAAATAAA AAACATAAC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8520      8530      8540      8550      8560
Sec_mit_Q1 AAACCCTGAG AACCAAAATG AACGAAAATC TGTTCCCTTC ATTCATTGCC
NUMT_Chr1  AAACCCTGAG AACCAAAATG AACGAAAATC TGTTCCACTTC ATTCATTGCC

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8570      8580      8590      8600      8610
Sec_mit_Q1 CCCACAATCC TAGGCCTACC CGCCGCAGTA CTGATCATTC TATTTCCCCC
NUMT_Chr1  CCCACAATCC TAGGCCTACC CGCCGCAGTA CTGATCATTC TATTTCCCCC

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8620      8630      8640      8650      8660
Sec_mit_Q1 TCTATTGATC CCCACCTCCA AATATCTCAT CAACAACCGA CTAATCACCA
NUMT_Chr1  TCTATTGATC CCCACCTCCA AATATCTCAT CAACAACCGA CTAATTACCA

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8670      8680      8690      8700      8710
mtDNA_S1  CCCAACAATG ACTAATCAA CTAACCTCAA AACAAATGAT AACCATACAC
NUMT_Chr1  CCCAACAATG ACTAATCAA CTAACCTCAA AACAAATGAT ACCCATACAC

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8720      8730      8740      8750      8760
mtDNA_S1  AACACTAAG GACGAACCTG ATCTCTTATA CTAGTATCCT TAATCATTTT
NUMT_Chr1  AACACTAAG GACGAACCTG ATCTCTTATA CTAGTATCCT TAATCATTTT

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8770      8780      8790      8800      8810
Sec_mit_Q1 TATTGCCACA ACTAACCTCC TCGGACTCCT GCCTCACTCA TTTACACCAA
NUMT_Chr1  TATTGCCACA ACTAACCTCC TCGGACTCCT GCCTCACTCA TTTACACCAA

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8820      8830      8840      8850      8860
Sec_mit_Q1 CCACCCAAC TATCTATAAAC CTAGCCATGG CCATCCCCTT ATGAGCGGGC
NUMT_Chr1  CCACCCAAC TATCTATAAAC CTAGCCATGG CCATCCCCTT ATGAGCGGGC

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8870      8880      8890      8900      8910
Sec_mit_Q1 GCAGTGATTA TAGGCTTTTCG CTCTAAGATT AAAAAATGCC TAGCCCCTT
NUMT_Chr1  GCAGTGATTA TAGGCTTTTCG CTCTAAGATT AAAAAATGCC TAGCCCCTT

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8920      8930      8940      8950      8960
Sec_mit_Q1 CTTACCACAA GGCACACCTA CACCCTTAT CCCCATACTA GTTATTATCG
NUMT_Chr1  CTTACCACAA GGCACACCTA CACCCTTAT CCCCATACTA GTTATTATCG

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8970      8980      8990      9000      9010
Sec_mit_Q1 AAACCATCAG CCTACTCATT CAACCAATAG CCCTGGCCGT ACGCCTAACC
NUMT_Chr1  AAACCATCAG CCTACTCATT CAACCAATAG CCCTGGCCGT ACGCCTAACC

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          9020          9030          9040          9050          9060
Sec_mit_Q1 GCTAACATTA CTGCAGGCCA CCTACTCATG CACCTAATTG GAAGCGCCAC
NUMT_Chrl  GCTAACATTA CTGCAGGCCA CCTACTCATG CACCTAATTG GAAGCGCCAC

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          9070          9080          9090          9100          9110
Sec_mit_Q1 CCTAGCAATA TCAACCATTA ACCTTCCCTC TACTACTTATC ATCTTCACAA
NUMT_Chrl  CCTAGCAATA TCAACTATTA ACCTTCCCTC TACTACTTATC ATCTTCACAA

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          9120          9130          9140          9150          9160
Sec_mit_Q1 TTCTAATTCT ACTGACTATC CTAGAAATCG CTGTTCGCCTT AATCCAAGCC
NUMT_Chrl  TTCTAATTCT ACTGACTATC CTAGAAATCG CTGTTCGCCTT AATCCAAGCC

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          9170          9180          9190          9200          9210
Sec_mit_Q1 TACGTTTTCA CACTTCTAGT AAGCCTCTAC CTGCACGACA ACACATAATG
NUMT_Chrl  TACGTTTTTA CACTTCTAGT AAGCCTCTAC CTGCACGACA ACACATAATG

```

```

.....|.....| .
          9220
Sec_mit_Q1 ACCCACCAAT C
NUMT_Chrl  ACCCACCAAT C

```

b

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          6520          6530          6540          6550          6560
Sec_mit_Q2 CTGCTGGCAT CACTATACTA CTAACAGACC GCAACCTCAA CACCACCTTC
NUMT_Chrl  CCGCTGGCAT CACTATACTA CTAACAGACC GTAACCTCAA CACCACCTTC

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          6570          6580          6590          6600          6610
Sec_mit_Q2 TTCGACCCCG CCGGAGGAGG AGACCCCAT TCTATACCAAC ACCTATTCTG
NUMT_Chrl  TTCGACCCAG CCGGAGGAGG AGACCCCAT TCTATACCAAC ACCTATTCTG

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          6620          6630          6640          6650          6660
Sec_mit_Q2 ATTTTTTCGGT CACCCTGAAG TTTATATTCT TATCTACCA GGCTTCGGAA
NUMT_Chrl  ATTTTTTCGGT CACCCTGAAG TTTATATTCT TCATCTACCA GGCTTCGGAA

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          6670          6680          6690          6700          6710
Sec_mit_Q2 TAATCTCCCA TATTGTAACT TACTACTCCG G-AAAAAAAG AACCATTTGG
NUMT_Chrl  TAATCTCCCA TATTGTAACT TACTACTCCG GGAAAAAAAG AACCATTTGG

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
          6720          6730          6740          6750          6760
Sec_mit_Q2 ATACATAGGT ATGGTCTGAG CTATGATATC AATTGGCTTC CTAGGGTTTA
NUMT_Chrl  ATACATAGGT ATGGTCTGAG CTATGATATC AATTGGCTTC CTAGGGTTTA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      6770      6780      6790      6800      6810
Sec_mit_Q2  TCGTGTGAGC ACACCATATA TTTACAGTAG GAATAGACGT AGACACACGA
NUMT_Chr1  TCGTGTGAGC ACACCATATA TTTACAGTAG GAATAGACGT AGACACACGA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      6820      6830      6840      6850      6860
Sec_mit_Q2  GCATATTTCA CCTCCGCTAC CATAATCATC GCTATCCCCA CCGGCGTCAA
NUMT_Chr1  GCATATTTCA CCTCCGCTAC CATAATCATC GCTATCCCCA CCGGCGTCAA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      6870      6880      6890      6900      6910
Sec_mit_Q2  AGTATTTAGC TGACTIONGCA CACTCCACGG AAGCAATATG AAATGATCTG
NUMT_Chr1  AGTATTTAGC TGACTIONGCA CACTCCACGG AAGCAATATG AAATGATCTG

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      6920      6930      6940      6950      6960
Sec_mit_Q2  CTGCAGTGCT CTGAGCCCTA GGATTGATCT TTCTTTTCAC CGTAGGTGGC
NUMT_Chr1  CTGCAGTGCT CTGAGCCCTA GGATTGATCT TTCTTTTCAC CGTAGGTGGC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      6970      6980      6990      7000      7010
Sec_mit_Q2  CTGACTGGCA TTGTATTAGC AAACCTCATCA CTAGACATCG TACTACACGA
NUMT_Chr1  CTGACTGGCA TTGTATTAGC AAACCTCATCA CTAGACATCG TACTACACGA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7020      7030      7040      7050      7060
Sec_mit_Q2  CACGTACTAC GTTGTAGCTC ACTTCCACTA TGTCCTATCA ATAGGAGCTG
NUMT_Chr1  CACGTACTAC GTTGTAGCTC ACTTCCACTA TGTCCTATCA ATAGGAGCTG

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7070      7080      7090      7100      7110
Sec_mit_Q2  TATTTGCCAT CATAGGAGGC TTCATTCACT GATTTCCCTT ATTCTCAGGC
NUMT_Chr1  TATTTGCCAT CATAGGAGGC TTCATTCACT GATTTCCCTT ATTCTCAGGC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7120      7130      7140      7150      7160
Sec_mit_Q2  TACACCCTAG ACCAAACCTA CGCCAAAATC CATTTCAGCTA TCATATTCAT
NUMT_Chr1  TACACCCTAG ACCAAACCTA CGCCAAAATC CATTTCAGCTA TCATATTCAT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7170      7180      7190      7200      7210
Sec_mit_Q2  CGGCGTAAAT CTAACCTTCT TCCACAACA CTTTCTCGGC CTATCCGGAA
NUMT_Chr1  CGGCGTAAAT CTAACCTTCT TCCACAACA CTTTCTCGGC CTATCCGGAA

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7220      7230      7240      7250      7260
Sec_mit_Q2  TGCCCCGACG TTACTIONGAC TACCCCGATG CATAACCCAC ATGAAACATC
NUMT_Chr1  TGCCCCGACG TTACTIONGAC TACCCCGATG CATAACCCAC ATGAAACATC

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7270      7280      7290      7300      7310
Sec_mit_Q2  CTATCATCTG TAGGCTCATT CATTCTCTCTA ACAGCAGTAA TATTAATAAT
NUMT_Chr1  CTATCATCTG TAGGCTCATT CATTCTCTCTA ACAGCAGTAA TATTAATAAT

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7320      7330      7340      7350      7360
Sec_mit_Q2  TTTCATGATT TGAGAAGCCT TCGCTTCGAA GCGAAAAGTC CTAATAGTAG
NUMT_Chr1  TTTCATAATT TGAGAAGCCT TCGCTTCGAA GCGAAAAGTC CTAATAGTAG

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7370      7380      7390      7400      7410
Sec_mit_Q2  AAGAACCCTC CATAAACCTG GAGTGACTAT ATGGATGCCC CCCACCCTAC
NUMT_Chr1  AAGAACCCTC CATAAACCTG GAGTGACTAT ATGGATGCCC CCCACCCTAC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7420      7430      7440      7450      7460
Sec_mit_Q2  CACACATTCG AAGAACCCGT ATACATAAAA TCTAGACAAA AAAGGAAGGA
NUMT_Chr1  CACACATTCG AAGAACCCGT ATACATAAAA TCTAGACAAA AAAGGAAGGA

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7470      7480      7490      7500      7510
Sec_mit_Q2  ATCGAACCCC CCAAAGCTGG TTTCAAGCCA ACCCCATGGC CTCCATGACT
NUMT_Chr1  ATCGAACCCC CCAAAGCTGG TTTCAAGCCA ACCCCATGGC CTCCATGACT

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7520      7530      7540      7550      7560
Sec_mit_Q2  TTTTCAAAAA GGTATTAGAA AAACCATTTC ATAACCTTGT CAAAGTTAAA
NUMT_Chr1  TTTTCAAAAA GATATTAGAA AAACCATTTC ATAACCTTGT CAAAGTTAAA

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7570      7580      7590      7600      7610
Sec_mit_Q2  TTATAGGCTA AATCCTATAT ATCTTAATGG CACATGCAGC GCAAGTAGGT
NUMT_Chr1  TTATAGGCTA AATCCTATAT ATCTTAATGG CACATGCAGC GCAAGTAGGT

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7620      7630      7640      7650      7660
Sec_mit_Q2  CTACAAGACG CTAATCCTAT TATCATAGAA GAGCTTATCA CCTTTCATGA
NUMT_Chr1  CTACAAGACG CTAATCCTAT TATCATAGAA GAGCTTATCA TCTTTCATGA

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7670      7680      7690      7700      7710
Sec_mit_Q2  TCACGCCCTC ATAATCATT TCCTTATCTG CTCCTAGTC CTGTATGCCC
NUMT_Chr1  TCACGCCCTC ATAATCATT TCCTTATCTG CTCCTAGTC CTGTATGCCC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      7720      7730      7740      7750      7760
Sec_mit_Q2  TTTTCCTAAC ACTCACAACA AAACCTAATA ATACTAACAT CTCAGACGCT
NUMT_Chr1  TTTTCCTAAC ACTCACAACA AAACCTAATA ATACTAACAT CTCAGACGCT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7770      7780      7790      7800      7810
Sec_mit_Q2 CAGGAAATAG AAACCGTCTG AACTATCCTG CCCGCCATCA TCCTAGTCCT
NUMT_Chr1 CAGGAAATAG AAACCGTCTG AACTATCCTG CCCGCCATCA TCCTAGTCCT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7820      7830      7840      7850      7860
Sec_mit_Q2 CATCGCCCTC CCATCCCTAC GCATCCTTTA CATAACAGAC GAGGTCAACG
NUMT_Chr1 CATCGCCCTC CCATCCCTAC GCATCCTTTA CATAACAGAC GAGGTCAACG

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7870      7880      7890      7900      7910
Sec_mit_Q2 ATCCCTCCCT TACCATCAAA TCAATTGGCC ACCAATGGTA CTGAACCTAC
NUMT_Chr1 ATCCCTCCCT TACCATCAAA TCAATTGGCC ATCAATGGTA CTGAACCTAC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7920      7930      7940      7950      7960
Sec_mit_Q2 GAGTACACCG ACTACGGCGG ACTAATCTTC AACTCCTACA TACTTCCCCC
NUMT_Chr1 GAAATACACCG ACTACGGCGG ACTAATCTTC AACTCCTACA TACTTCCCCC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      7970      7980      7990      8000      8010
Sec_mit_Q2 ATTATTCCTA GAACCAGGCG ACCTGCGACT CCTTGACGTT GACAATCGAG
NUMT_Chr1 ATTATTCCTA GAACCAGGCG ACCTGCGACT CCTTGACGTT GACAATCGAG

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8020      8030      8040      8050      8060
Sec_mit_Q2 TAGTACTCCC GATTGAAGCC CCCATTCGTA TAATAATTAC ATCACAAGAC
NUMT_Chr1 TAGTACTCCC GTTGAAGCC CCCATTCGTA TAATAATTAC ATCACAAGAC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8070      8080      8090      8100      8110
Sec_mit_Q2 GTCTTGCACT CATGAGCTGT CCCACATTA GGCTTAAAAA CAGATGCAAT
NUMT_Chr1 GTCTTGCACT CATGAGCTGT CCCACATTA GGCTTAAAAA CAGATGCAAT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8120      8130      8140      8150      8160
Sec_mit_Q2 TCCCGGACGT CTAAACCAAA CCACTTTCAC CGCTACACGA CCGGGGTAT
NUMT_Chr1 TCCCGGACGT CTAAACCAAA CCACTTTCAC TGCTACACGA CCGGGGTAT

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8170      8180      8190      8200      8210
Sec_mit_Q2 ACTACGGTICA ATGCTCTGAA ATCTGTGGAG CAAACCACAG TTTCATGCCC
NUMT_Chr1 ACTACGGCCA ATGCTCTGAA ATCTGTGGAG CAAACC--AG TTTTATGCCC

```

```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
      8220      8230      8240      8250      8260
Sec_mit_Q2 ATCGTCCTAG AATTAATTCC CCTAAAAATC TTTGAAATAG GGCCCCTATT
NUMT_Chr1 ATCGTCCTAG AATTAATTCC CCTAAAAATC TTTGAAATAG GGCCCCTATT

```

```

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8270      8280      8290      8300      8310
Sec_mit_Q2  TACCCTATAG CACCCCCTCT ACCCCCCTCTA GAGCCCCTG TAAAGCTAAC
NUMT_Chr1   TACCCTATAG CACCCCCTCT ACCCCCCTCTA GAGCCCCTG TAAAGCTAAC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8320      8330      8340      8350      8360
Sec_mit_Q2  TTAGCATTAA CCTTTTAAGT TAAAGATTAA GAGAACCAAC ACCTCTTTAC
NUMT_Chr1   TTAGCATTAA CCTTTTAAGT TAAAGATTAA GAGAACCAAC ACCTCTTTAC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8370      8380      8390      8400      8410
Sec_mit_Q2  AGTGAAATGC CCCAACTAAA CACTACCGTA TGGCCACCA TAATTACCCC
NUMT_Chr1   AGTGAAATGC CCCAACTAAA TACTACCGTA TGAACCCACCA TAATTACCCC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8420      8430      8440      8450      8460
Sec_mit_Q2  CATACTCCTT ACACTATTCC TCATCACCCA ACTAAAAATA TTAAACACAA
NUMT_Chr1   CATACTCCTT ACACTATTCC TCATCACCCA ACTAAAAATA TTAAATACAA

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8470      8480      8490      8500      8510
Sec_mit_Q2  ACTACCACCT ACCTCCCTCA CCAAAGCCCA TAAAAATAAA AAATTATAAC
NUMT_Chr1   ATTACCACCT ACCTCCCTCA CCAAAGCCCA TAAAAATAAA AAACATAAC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8520      8530      8540      8550      8560
Sec_mit_Q2  AAACCCTGAG AACCAAATG AACGAAAATC TGTTGCTTC ATTCATTGCC
NUMT_Chr1   AAACCCTGAG AACCAAATG AACGAAAATC TGTTCACTTC ATTCATTGCC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8570      8580      8590      8600      8610
Sec_mit_Q2  CCCACAATCC TAGGCCTACC CGCCGCAGTA CTGATCATTC TATTTCCCCC
NUMT_Chr1   CCCACAATCC TAGGCCTACC CGCCGCAGTA CTGATCATTC TATTTCCCCC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8620      8630      8640      8650      8660
Sec_mit_Q2  TCTATTGATC CCCACCTCCA AATATCTCAT CAACAACCGA CTAATCACCA
NUMT_Chr1   TCTATTGATC CCCACCTCCA AATATCTCAT CAACAACCGA CTAATTACCA

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8670      8680      8690      8700      8710
Sec_mit_Q2  CCCAACAATG ACTAATCAA CTAACCTCAA AACAAATGAT AACCATACAC
NUMT_Chr1   CCCAACAATG ACTAATCAA CTAACCTCAA AACAAATGAT AGCCATACAC

.....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      8720      8730      8740      8750      8760
Sec_mit_Q2  AACACTAAG GACGAACCTG ATCTCTTATA CTAGTATCCT TAATCATTTT
NUMT_Chr1   AACACTAAG GACGAACCTG ATCTCTTATA CTAGTATCCT TAATCATTTT

```

```

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
            8770      8780      8790      8800      8810
Sec_mit_Q2 TATTGCCACA ACTAACCTCC TCGGACTCCT GCCTCACTCA TTTACACCAA
NUMT_Chr1 TATTGCCACA ACTAACCTCC TCGGACTCCT GCCTCACTCA TTTACACCAA

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
            8820      8830      8840      8850      8860
Sec_mit_Q2 CCACCCAAC TCTATAAAC CTAGCCATGG CCATCCCCTT ATGAGCGGGC
NUMT_Chr1 CCACCCAAC TCTATAAAC CTAGCCATGG CCATCCCCTT ATGAGCGGGC

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
            8870      8880      8890      8900      8910
Sec_mit_Q2 GCAGTGATTA TAGGCTTTCG CTCTAAGATT AAAAATGCC TAGCCCACTT
NUMT_Chr1 GCAGTGATTA TAGGCTTTCG CTCTAAGATT AAAAATGCC TAGCCCACTT

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
            8920      8930      8940      8950      8960
Sec_mit_Q2 CTTACCACAA GGCACACCTA CACCCCTTAT CCCAATACTA GTTATTATCG
NUMT_Chr1 CTTACCACAA GGCACACCTA CACCCCTTAT CCCTAATACTA GTTATTATCG

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
            8970      8980      8990      9000      9010
Sec_mit_Q2 AAACCATCAG CCTACTCATT CAACCAATAG CCCTGGCCGT ACGCCTAACC
NUMT_Chr1 AAACCATCAG CCTACTCATT CAACCAATAG CCCTGGCCGT ACGCCTAACC

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
            9020      9030      9040      9050      9060
Sec_mit_Q2 GCTAACATTA CTGCAGGCCA CCTACTCATG CACCTAATTG GAAGCGCCAC
NUMT_Chr1 GCTAACATTA CTGCAGGCCA CCTACTCATG CACCTAATTG GAAGCGCCAC

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
            9070      9080      9090      9100      9110
Sec_mit_Q2 CCTAGCAATA TCAACCATTA ACCTTCCCTC TACTACTTATC ATCTTCACAA
NUMT_Chr1 ACTAGCAATA TCAACTATTA ACCTTCCCTC TACTACTTATC ATCTTCACAA

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
            9120      9130      9140      9150      9160
Sec_mit_Q2 TTCTAATTCT ACTGACTATC CTAGAAATCG CTGTGCCTT AATCCAAGCC
NUMT_Chr1 TTCTAATTCT ACTGACTATC CTAGAAATCG CTGTGCCTT AATCCAAGCC

      ....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
            9170      9180      9190      9200      9210
Sec_mit_Q2 TACGTTTTCA CACTTCTAGT AAGCCTCTAC CTGCACGACA ACACATAATG
NUMT_Chr1 TACGTTTTTA CACTTCTAGT AAGCCTCTAC CTGCACGACA ACACATAATG

      ....|.....| .
            9220
Sec_mit_Q2 ACCCACCAAT C
NUMT_Chr1 ACCCACCAAT C

```

Figura 1. Alineamientos mediante programa *BioEdit* (Hall 1999) *Sequence Alignment Editor*. a) Alineamiento entre el DNAmT del individuo (Q1) y el NUMT. b) Alineamiento entre el DNAmT del individuo (Q2) y el NUMT. En ambos casos, las posiciones no idénticas están marcadas en amarillo.

10.3. ANEXO III

Tabla 1. Cálculo total de patogenicidad a partir de las variantes mitocondriales detectadas en genes que codifican para complejos de la cadena respiratoria que dan lugar a cambios a nivel aminoacídico para cada una de las muestras. *Cálculo de la frecuencia de cada una de las variantes en la población general (N= 7605). Ala: alanina; Asn: asparagina; Asp: ácido aspártico; Arg: arginina; Gly: glicina; His: histidina; Ile: isoleucina; Leu: leucina; Met: metionina; Phe: fenilalanina; Thr: treonina; Tyr: tirosina; Val: valina.

Muestra	Posición	Base alternativa	Frecuencia	Región mitocondrial	Hits Filogenia	Cambio aminoacídico	MutPred2	Frecuencia en la población (%)*
3	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
3	4232	C	1	ND1	6	I309T (Ile-Thr)	0,136	0,11
3	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
3	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
3	13708	A	8	ND5	37	A458T (Ala-Thr)	0,067	11,14
3	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
3	14798	C	7	CYB	4	F18L (Phe-Leu)	0,127	14,74
3	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
3	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
							1,162	
5	3721	G	1	ND1	1	T139A (Thr-Ala)	0,62	0,08
5	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
5	8734	A	1	ATP6	1	L70I (Leu-Ile)	0,202	0,09
5	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
5	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
5	13708	A	8	ND5	37	A458T (Ala-Thr)	0,067	11,14
5	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
5	14798	C	7	CYB	4	F18L (Phe-Leu)	0,127	14,74
5	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
5	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
							1,848	
6	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
6	4917	G	4	ND2	4	N150D (Asn-Asp)	0,402	8,67
6	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
6	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
6	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
6	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
							1,031	
7	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
7	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
							0,385	
10	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
10	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
10	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
10	13708	A	8	ND5	37	A458T (Ala-Thr)	0,067	11,14
10	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03

Tabla 1. Continuación.

Muestra	Posición	Base alternativa	Frecuencia	Región mitocondrial	Hits Filogenia	Cambio aminoacídico	MutPred2	Frecuencia en la población (%)*
10	15257	A	2	CYB	3	D171N (Asp-Asn)	0,281	2,34
10	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
10	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
1,18								
13	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
13	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
13	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
13	13708	A	8	ND5	37	A458T (Ala-Thr)	0,067	11,14
13	13933	G	1	ND5	3	T533A (Thr-Ala)	0,115	0,37
13	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
13	14798	C	7	CYB	4	F18L (Phe-Leu)	0,127	14,74
13	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
13	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
1,141								
16	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
16	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
16	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
16	15693	C	1	CYB	1	M316T (Met-Thr)	0,163	2,14
0,612								
17	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
17	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
0,385								
18	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
18	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
0,385								
19	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
19	9145	A	1	ATP6	1	A207T (Ala-Thr)	0,563	0,13
19	9477	A	2	CO3	4	V91I (Val-Ile)	0,328	8,32
19	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
19	14793	G	1	CYB	1	H16R (His-Arg)	0,224	4,21
19	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
1,564								
20	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
20	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
0,385								
21	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
21	9055	A	2	ATP6	12	A177T (Ala-Thr)	0,297	8,9
21	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
21	11084	G	1	ND4	7	T109A (Thr-Ala)	0,461	0,21
21	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
21	14798	C	7	CYB	4	F18L (Phe-Leu)	0,127	14,74

Tabla 1. Continuación.

Muestra	Posición	Base alternativa	Frecuencia	Región mitocondrial	Hits Filogenia	Cambio aminoacídico	MutPred2	Frecuencia en la población (%)*
21	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
								1,537
26	3338	C	1	ND1	3	V11A (Val-Ala)	0,185	0,12
26	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
26	4879	G	1	ND2	0	A137G (Ala-Gly)	0,387	0
26	4917	G	4	ND2	4	N150D (Asn-Asp)	0,402	8,67
26	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
26	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
26	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
26	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
								1,603
27	7805	A	1	CO2	9	V74I (Val-Ile)	0,051	1,46
27	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
27	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
27	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
								0,5
30	6340	T	1	CO1	0	T146I (Thr-Ile)	0,134	0,13
30	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
30	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
								0,519
31	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
31	9055	A	2	ATP6	12	A177T (Ala-Thr)	0,297	8,9
31	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
31	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
31	14798	C	7	CYB	4	F18L (Phe-Leu)	0,127	14,74
31	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
								1,076
31bis	6663	G	1	CO1	1	I254V (Ile-Val)	0,167	0,07
31bis	8701	G	2	ATP6	9	T59A (Thr-Ala)	0,085	2,49
31bis	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
31bis	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
31bis	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
31bis	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
								0,904
33	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
33	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
								0,385
34	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
34	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
								0,385

Tabla 1. Continuación.

Muestra	Posición	Base alternativa	Frecuencia	Región mitocondrial	Hits Filogenia	Cambio aminoacídico	MutPred2	Frecuencia en la población (%)*
35	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
35	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
35	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
35	13708	A	8	ND5	37	A458T (Ala-Thr)	0,067	11,14
35	14325	C	1	ND6	3	N117D (Asn-Asp)	0,273	0,11
35	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
35	14798	C	7	CYB	4	F18L (Phe-Leu)	0,127	14,74
35	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
35	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
							1,299	
36	8843	C	1	ATP6	9	I106T (Ile-Thr)	0,592	0,42
36	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
36	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
							0,977	
41	8584	A	1	ATP6	5	A20T (Ala-Thr)	0,243	0,41
41	8701	G	2	ATP6	9	T59A (Thr-Ala)	0,085	2,49
41	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
41	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
41	14318	C	1	ND6	1	N119S (Asn-Ser)	0,366	0,16
41	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
41	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
							1,346	
43	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
43	4917	G	4	ND2	4	N150D (Asn-Asp)	0,402	8,67
43	6261	A	1	CO1	10	A120T (Ala-Thr)	0,324	0,6
43	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
43	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
43	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
43	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
							1,355	
44	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
44	14502	C	1	ND6	6	I58V (Ile-Val)	0,364	0,18
44	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
							0,749	
45	3394	C	1	ND1	7	Y30H (Tyr-His)	0,651	0,97
45	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
45	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
45	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
45	13708	A	8	ND5	37	A458T (Ala-Thr)	0,067	11,14
45	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
45	14798	C	7	CYB	4	F18L (Phe-Leu)	0,127	14,74

Tabla 1. Continuación.

Muestra	Posición	Base alternativa	Frecuencia	Región mitocondrial	Hits Filogenia	Cambio aminoacídico	MutPred2	Frecuencia en la población (%)*
45	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
45	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
1,677								
50	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
50	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
50	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
50	13708	A	8	ND5	37	A458T (Ala-Thr)	0,067	11,14
50	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
50	15257	A	2	CYB	3	D171N (Asp-Asn)	0,281	2,34
50	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
50	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
1,18								
51	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
51	4917	G	4	ND2	4	N150D (Asn-Asp)	0,402	8,67
51	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
51	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
51	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
51	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
1,031								
53	4216	C	12	ND1	13	Y304H (Tyr-His)	0,094	17,71
53	5460	A	2	ND2	38(1)	A331T (Ala-Thr)	0,112	5,1
53	8557	A	1	ATP6	7	A11T (Ala-Thr)	0,047	0,95
53	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
53	10398	G	12	ND3	21	T114A (Thr-Ala)	0,203	20,51
53	13708	A	8	ND5	37	A458T (Ala-Thr)	0,067	11,14
53	13879	C	1	ND5	9	S515P (Ser-Pro)	0,199	0,87
53	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
53	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
53	15452	A	12	CYB	2	L236I (Leu-Ile)	0,086	17,42
1,257								
54	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
54	9477	A	2	CO3	4	V91I (Val-Ile)	0,328	8,32
54	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
54	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
54	15884	A	1	CYB	20	A380T (Ala-Thr)	0,136	1,42
0,913								
55	3505	G	1	ND1	5	T67A (Thr-Ala)	0,133	2,16
55	4093	G	1	ND1	5	T263A (Thr-Ala)	0,223	0,16
55	5046	A	1	ND2	5	V193I (Val-Ile)	0,099	2,72
55	5460	A	2	ND2	38(1)	A331T (Ala-Thr)	0,112	5,1
55	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95

Tabla 1. Continuación.

Muestra	Posición	Base alternativa	Frecuencia	Región mitocondrial	Hits Filogenia	Cambio aminoacídico	MutPred2	Frecuencia en la población (%)*
55	14766	T	21	CYB	4	T7I (Thr-Ile)	0,064	50,03
55	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
55	15884	C	1	CYB	20	A380P (Ala-Thr)	0,292	2,18
							1,308	
56	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
56	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
							0,385	
59	8860	G	32	ATP6	4	T112A (Thr-Ala)	0,288	98,95
59	15326	G	32	CYB	8	T194A (Thr-Ala)	0,097	98,93
							0,385	

Tabla 2. Cálculo total de patogenicidad a partir de las variantes mitocondriales detectadas en RNAt en cada una de las muestras. *Cálculo de la frecuencia de cada una de las variantes en la población general (N= 7605).

Muestra	Posición	Base alternativa	Frecuencia	MapLocus	Hits Filogenia	Valor patogenicidad MitoTIP	Frecuencia en la población (%)*
3	NO		NO	NO	NO	0	0
						0	
5	NO		NO	NO	NO	0	0
						0	
6	10463	C	4	TR (RNAt arginina)	5	2,703	8,76
6	15928	A	4	TT (RNAt treonina)	5	7,54	8,52
						10,243	
7	10057	C	1	TG (RNAt glicina)	0	11,292	0,01
						11,292	
10	7476	T	2	TS1 (RNAt serina1)	1	0,016	1,91
						0,016	
13	NO		NO	NO	NO	0	0
						0	
16	12308	G	6	TL2 (RNAt leucina2)	3	11,851	23,37
						11,851	
17	NO		NO	NO	NO	0	0
						0	
18	5780	A	1	TC (RNAt cisteína)	1	10,728	0,07
						10,728	
19	12308	G	6	TL2 (RNAt leucina2)	3	11,851	23,37
						11,851	
20	NO		NO	NO	NO	0	0
						0	
21	12308	G	6	TL2 (RNAt leucina2)	3	11,851	23,37
						11,851	

Tabla 2. Continuación.

Muestra	Posición	Base alternativa	Frecuencia	MapLocus	Hits Filogenia	Valor patogenicidad MitoTIP	Frecuencia en la población (%)*
26	10463	C	4	TR (RNAt arginina)	5	2,703	8,76
26	15928	A	4	TT (RNAt treonina)	5	7,54	8,52
						10,243	
27	12308	G	6	TL2 (RNAt leucina2)	3	11,851	23,37
						11,851	
30	NO		NO	NO	NO	0	0
						0	
31	12308	G	6	TL2 (RNAt leucina2)	3	11,851	23,37
						11,851	
33	NO		NO	NO	NO	0	0
						0	
31bis	7521	A	1	TD (RNAt ácido aspártico)	10	0,706	1,13
						0,706	
34	NO		NO	NO	NO	0	0
						0	
35	NO		NO	NO	NO	0	0
						0	
36	NO		NO	NO	NO	0	0
						0	
41	15930	A	1	TT (RNAt treonina)	14	2,389	0,27
						2,389	
43	10463	C	4	TR (RNAt arginina)	5	2,703	8,76
43	15928	A	4	TT (RNAt treonina)	5	7,54	8,52
						10,243	
44	12172	G	1	TH (RNAt histidina)	10	3,327	0,25
44	15904	T	1	TT (RNAt treonina)	1	1,691	3,68
						5,018	
45	5773	A	1	TC (RNAt cisteína)	13	2,435	0,5
45	12192	A	1	TH (RNAt histidina)	4	2,659	0,14
						5,094	
50	7476	T	2	TS1 (RNAt serina1)	1	0,016	1,91
						0,016	
51	10463	C	4	TR (RNAt arginina)	5	2,703	8,76
51	15928	A	4	TT (RNAt treonina)	5	7,54	8,52
						10,243	
53	NO		NO	NO	NO	0	0
						0	
54	12308	G	6	TL2 (RNAt leucina2)	3	11,851	23,37
						11,851	
55	NO		NO	NO	NO	0	0
						0	
56	NO		NO	NO	NO	0	0
						0	
59	NO		NO	NO	NO	0	0
						0	

Tabla 3. Cálculo de la carga mutacional y la carga mutacional ajustada a partir de variantes mitocondriales en genes que codifican para complejos de la cadena respiratoria que dan lugar a cambios a nivel aminoacídico para cada una de las muestras. La carga mutacional se ha calculado a partir de diferentes valores de patogenicidad: variantes totales, variantes con valor de patogenicidad >0,3, variantes con valor de patogenicidad >0,5 y variantes con valor de patogenicidad >0,6.

Muestra	Edad	Grado	Haplogrupo	Variantes totales			Variantes con valor de patogenicidad >0,3			Variantes con valor de patogenicidad >0,5			Variantes con valor de patogenicidad >0,6			
				Carga Mutacional	Número de Variantes	Carga mutacional ajustada	Carga Mutacional	Número de Variantes	Carga mutacional ajustada	Carga Mutacional	Número de Variantes	Carga mutacional ajustada	Carga Mutacional	Número de Variantes	Carga mutacional ajustada	
3	49	IV	J1c2e2	1,162	9	0,129	0	0	0	0	0	0	0	0	0	0
5	71	IV	J1c5c1	1,848	10	0,185	0,62	1	0,62	0,62	1	0,62	0,62	1	0,62	0,62
6	64	IV	T2b2b	1,031	6	0,172	0,402	1	0,402	0	0	0	0	0	0	0
7	37	II	H89	0,385	2	0,193	0	0	0	0	0	0	0	0	0	0
10	29	III	J2a1a1	1,18	8	0,148	0	0	0	0	0	0	0	0	0	0
13	53	IV	J1c2c	1,141	9	0,127	0	0	0	0	0	0	0	0	0	0
16	43	IV	U4b1a	0,612	4	0,153	0	0	0	0	0	0	0	0	0	0
17	56	IV	H1j	0,385	2	0,193	0	0	0	0	0	0	0	0	0	0
18	34	III	H1ap1	0,385	2	0,193	0	0	0	0	0	0	0	0	0	0
19	43	IV	U5a2c	1,564	6	0,261	0,891	2	0,4455	0,563	1	0,563	0	0	0	0
20	53	II	H1r1	0,385	2	0,193	0	0	0	0	0	0	0	0	0	0
21	59	IV	K1a1	1,537	7	0,220	0,461	1	0,461	0	0	0	0	0	0	0
26	58	IV	T2b2b	1,603	8	0,200	0,789	2	0,3945	0	0	0	0	0	0	0
27	46	II	U6a3a1	0,5	4	0,125	0	0	0	0	0	0	0	0	0	0
30	31	II	H58a	0,519	3	0,173	0	0	0	0	0	0	0	0	0	0
31	73	IV	K1a4a1	1,076	6	0,179	0	0	0	0	0	0	0	0	0	0
31bis	53	IV	L2a1c	0,904	6	0,151	0	0	0	0	0	0	0	0	0	0
33	70	IV	H1v	0,385	2	0,193	0	0	0	0	0	0	0	0	0	0
34	52	II	H5b1	0,385	2	0,193	0	0	0	0	0	0	0	0	0	0
35	53	II	J1c2r	1,299	9	0,144	0	0	0	0	0	0	0	0	0	0

Tabla 4. Cálculo de la carga mutacional y la carga mutacional ajustada a partir de variantes mitocondriales localizadas en regiones que codifican para RNAt. La carga mutacional se ha calculado a partir de diferentes valores de patogenicidad: variantes totales y variantes con valor de patogenicidad >12,66.

Muestra	Edad	Grado	Haplogrupo	Variantes totales			Variantes con valores de patogenicidad >12,66		
				Carga Mutacional	Número de Variantes	Carga mutacional ajustada	Carga Mutacional	Número de Variantes	Carga mutacional ajustada
3	49	IV	J1c2e2	0	0	0	0	0	0
5	71	IV	J1c5c1	0	0	0	0	0	0
6	64	IV	T2b2b	10,243	2	5,122	0	0	0
7	37	II	H89	11,292	1	11,292	0	0	0
10	29	III	J2a1a1	0,016	1	0,016	0	0	0
13	53	IV	J1c2c	0	0	0	0	0	0
16	43	IV	U4b1a	11,851	1	11,851	0	0	0
17	56	IV	H1j	0	0	0	0	0	0
18	34	III	H1ap1	10,728	1	10,728	0	0	0
19	43	IV	U5a2c	11,851	1	11,851	0	0	0
20	53	II	H1r1	0	0	0	0	0	0
21	59	IV	K1a1	11,851	1	11,851	0	0	0
26	58	IV	T2b2b	10,243	2	5,122	0	0	0
27	46	II	U6a3a1	11,851	1	11,851	0	0	0
30	31	II	H58a	0	0	0	0	0	0
31	73	IV	K1a4a1	11,851	1	11,851	0	0	0
31bis	53	IV	L2a1c	0,706	1	0,706	0	0	0
33	70	IV	H1v	0	0	0	0	0	0
34	52	II	H5b1	0	0	0	0	0	0
35	53	II	J1c2r	0	0	0	0	0	0
36	31	III	H45b	0	0	0	0	0	0
41	34	III	C	2,389	1	2,389	0	0	0

Tabla 4. Continuación.

Muestra	Edad	Grado	Haplogrupo	Variantes totales			Variantes con valores de patogenicidad >12,66		
				Carga Mutacional	Número de Variantes	Carga mutacional ajustada	Carga Mutacional	Número de Variantes	Carga mutacional ajustada
43	50	III	T2c1d1	10,243	2	5,122	0	0	0
44	59	IV	V	5,018	2	2,509	0	0	0
45	35	II	J1c1b1	5,094	2	2,547	0	0	0
50	56	IV	J2a1a1a2	0,016	1	0,016	0	0	0
51	63	III	T2b	10,243	2	5,122	0	0	0
53	57	IV	J1b1a1	0	0	0	0	0	0
54	65	IV	U5b1b1d	11,851	1	11,851	0	0	0
55	35	IV	W6a	0	0	0	0	0	0
56	60	III	H15	0	0	0	0	0	0
59	39	II	H1r1	0	0	0	0	0	0

Tabla 5. Comparación de las frecuencias de las mutaciones en homoplasma detectadas en el presente estudio vs. la población general europea -datos de genomas completos obtenidos de de HmtDB (Clima et al. 2017)- (N=877). ATT: sitio de unión a la membrana; CBS1: bloque de secuencia conservada 1; HPR: primer de replicación; MT5: elemento control; OH: origen de replicación; TAS2: secuencia asociada a la terminación 2; TFY: sitio de unión del factor de transcripción 1 (TF1); TFX: sitio de unión de TF1.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Presente estudio		Base de datos de población europea		Test exacto de Fisher
						Individuos N=32	Frecuencia (%)	Individuos N=7605	Frecuencia (%)	
10	T	T>C	Región control / ATT / 7S DNA	1	1	1	3,13	5	0,07	0,0249
57	T	T>C	HVRII / ATT / 7S DNA / OH57	1	4	1	3,13	32	0,42	0,1296
72	T	T>C	HVRII / ATT / 7S DNA	1	6	1	3,13	292	3,84	1
73	A	A>G	HVRII / ATT / 7S DNA	21	19	21	65,63	4111	54,06	0,216
114	C	C>T	HVRII / ATT / OH / 7S DNA	1	8(3)	1	3,13	63	0,83	0,2365
143	G	G>A	HVRII / ATT / OH / 7S DNA	2	28(2)	2	6,25	49	0,65	0,0191
146	T	T>C	HVRII / ATT / OH / 7S DNA	8	120(2)	8	25	696	9,15	0,0071*
150	C	C>T	HVRII / ATT / OH / 7S DNA	4	74(2)	4	12,50	755	9,93	0,5535
152	T	T>C	HVRII / ATT / OH / 7S DNA	6	206(2)	6	18,75	1524	20,04	1
185	G	G>A	HVRII / ATT / OH / 7S DNA	4	26(2)	4	12,50	475	6,24	0,1378
188	A	A>G	HVRII / ATT / OH / 7S DNA	4	6	4	12,50	189	2,49	0,0082*
189	A	A>G	HVRII / ATT / OH / 7S DNA	1	39	1	3,13	285	3,75	1
195	T	T>C	HVRII / ATT / OH	8	116(4)	8	25	1341	17,63	0,2521
204	T	T>C	HVRII / ATT / OH	1	42	1	3,13	440	5,78	1
207	G	G>A	HVRII / ATT / OH	1	43(3)	1	3,13	311	4,09	1
215	A	A>G	HVRII / ATT / CSB1 / OH	2	6(1)	2	6,25	87	1,15	0,0531
228	G	G>A	HVRII / ATT / CSB1 / OH	5	13(2)	5	15,63	429	5,64	0,0328
242	C	C>T	HVRII / ATT / TFX / OH	1	1	1	3,13	55	0,72	0,2102
263	A	A>G	HVRII / ATT / OH	32	13	32	100	7526	98,96	1
279	T	T>C	HVRII: ATT / TFY / OH	1	4	1	3,13	32	0,42	0,1296
292	T	T>C	HVRII: ATT / TFY / OH	1	1 (1 ins)	1	3,13	2	0,03	0,0125
295	C	C>T	HVRII: ATT / TFY / OH	8	3	8	25	692	9,1	0,0069*
319	T	T>C	HVRII: ATT / HPR / OH	2	1(1)	2	6,25	76	1	0,0419
449	T	T>C	HVRIII: ATT	1	0	1	3,13	2	0,03	0,0125
456	C	C>T	HVRIII: ATT	1	8(1) (1 del /1 ins)	1	3,13	223	2,93	0,615
462	C	C>T	HVRIII: ATT	6	2	6	18,75	508	6,68	0,0182
482	T	T>C	HVRIII: ATT	1	6	1	3,13	61	0,8	0,23
489	T	T>C	HVRIII / ATT	9	3	9	28,13	721	9,48	0,0024*
497	C	C>T	HVRIII: ATT	2	1	2	6,25	334	4,39	0,6506
499	G	G>A	HVRIII: ATT	2	7	2	6,25	165	2,17	0,1543
513	G	G>A	HVRIII: ATT	2	21(1)	2	6,25	125	1,64	0,0985
709	G	G>A	RNAr 12S	6	57(1)	6	18,75	1111	14,61	0,456
723	A	A>G	RNAr 12S	1	9	1	3,13	21	0,28	0,0884
739	C	C>T	RNAr 12S	1	2	1	3,13	3	0,04	0,0167
750	A	A>G	RNAr 12S	32	4	32	100	7491	98,5	1
769	G	G>A	RNAr 12S	1	5	1	3,13	69	0,91	0,2557
930	G	G>A	RNAr 12S	1	8	1	3,13	273	3,59	1
1018	G	G>A	RNAr 12S	1	2	1	3,13	68	0,89	0,2525
1189	T	T>C	RNAr 12S	2	3	2	6,25	528	6,94	1

Tabla 5. Continuación.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Presente estudio		Base de datos de población europea		Test exacto de Fisher
						Individuos N=32	Frecuencia (%)	Individuos N=7605	Frecuencia (%)	
1243	T	T>C	RNAr 12S	2	5	2	6,25	165	2,17	0,1543
1438	A	A>G	RNAr 12S	32	11	32	100	7302	96,02	0,6375
1518	C	C>T	RNAr 12S	1	0	1	3,13	0	0	0,0042*
1688	A	A>C	RNAr 16S	1	0	1	3,13	0	0	0,0042*
1709	G	G>A	RNAr 16S	1	9	1	3,13	24	0,32	0,0998
1809	T	T>C	RNAr 16S	1	3	1	3,13	5	0,07	0,0249
1811	A	A>G	RNAr 16S	3	6	3	9,38	1007	13,24	0,7924
1850	T	T>C	RNAr 16S	1	1	1	3,13	55	0,72	0,2102
1888	G	G>A	RNAr 16S	5	13	5	15,63	653	8,59	0,1921
2158	T	T>C	RNAr 16S	1	3	1	3,13	57	0,75	0,2169
2392	T	T>C	RNAr 16S	1	0	1	3,13	3	0,04	0,0167
2416	T	T>C	RNAr 16S	2	5	2	6,25	28	0,37	0,0069*
2706	A	A>G	RNAr 16S	22	10	22	69,00	4199	55,21	0,1541
2789	C	C>T	RNAr 16S	1	2	1	3,13	23	0,3	0,096
3010	G	G>A	RNAr 16S	13	14	13	40,63	1862	24,48	0,0403
3173	G	G>A	RNAr 16S	1	0	1	3,13	0	0	0,0042*
3197	T	T>C	RNAr 16S	2	4	2	6,25	644	8,47	1
3338	T	T>C	ND1	1	3	1	3,13	9	0,12	0,0411
3348	A	A>G	ND1	1	1	1	3,13	127	1,67	0,4184
3394	T	T>C	ND1	1	7	1	3,13	74	0,97	0,2713
3447	A	A>G	ND1	1	4	1	3,13	75	0,99	0,2744
3480	A	A>G	ND1	2	3	2	6,25	653	8,59	1
3505	A	A>G	ND1	1	5	1	3,13	164	2,16	0,5036
3531	G	G>A	ND1	1	10	1	3,13	30	0,39	0,1223
3552	T	T>A	ND1	1	6	1	3,13	10	0,13	0,0452
3594	C	C>T	ND1	1	1	1	3,13	66	0,87	0,2461
3672	A	A>G	ND1	1	1	1	3,13	15	0,2	0,065
3721	A	A>G	ND1	1	1	1	3,13	6	0,08	0,029
4093	A	A>G	ND1	1	5	1	3,13	12	0,16	0,0532
4104	A	A>G	ND1	1	1	1	3,13	65	0,85	0,2429
4164	A	A>G	ND1	1	4	1	3,13	7	0,09	0,033
4216	T	T>C	ND1	12	13	12	37,50	1347	17,71	0,0084*
4224	C	C>T	ND1	1	1	1	3,13	2	0,03	0,0125
4232	T	T>C	ND1	1	6	1	3,13	8	0,11	0,0371
4580	G	G>A	ND2	1	2	1	3,13	264	3,47	1
4646	T	T>C	ND2	1	5	1	3,13	178	2,34	0,5326
4715	A	A>G	ND2	1	3	1	3,13	27	0,36	0,1111
4733	T	T>C	ND2	1	2	1	3,13	82	1,08	0,2956
4769	A	A>G	ND2	32	6	32	100	7356	96,73	0,6259
4820	G	G>A	ND2	1	6	1	3,13	20	0,26	0,0845
4879	C	C>G	ND2	1	0	1	3,13	0	0	0,0042*
4917	A	A>G	ND2	4	4	4	12,50	659	8,67	0,3549
5046	G	G>A	ND2	1	5	1	3,13	207	2,72	0,5875
5054	G	G>A	ND2	2	10	2	6,25	15	0,2	0,0022*

Tabla 5. Continuación.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Presente estudio		Base de datos de población europea		Test exacto de Fisher
						Individuos N=32	Frecuencia (%)	Individuos N=7605	Frecuencia (%)	
5147	G	G>A	ND2	3	22	3	9,38	293	3,85	0,1256
5187	C	C>T	ND2	1	1	1	3,13	21	0,28	0,0884
5198	A	A>G	ND2	1	1	1	3,13	44	0,58	0,1726
5460	G	G>A	ND2	2	38(1)	2	6,25	388	5,1	0,6779
5471	G	G>A	ND2	1	8	1	3,13	79	1,04	0,2866
5656	A	A>G	NC4	1	5	1	3,13	157	2,06	0,4885
5773	G	G>A	TC (RNAt cisteína)	1	13	1	3,13	38	0,5	0,1514
5780	G	G>A	TC (RNAt cisteína)	1	1	1	3,13	5	0,07	0,0249
5964	T	T>C	CO1	1	4	1	3,13	9	0,12	0,0411
5999	T	T>C	CO1	1	3	1	3,13	164	2,16	0,5036
6047	A	A>G	CO1	1	2	1	3,13	162	2,13	0,4993
6260	G	G>A	CO1	1	10	1	3,13	100	1,31	0,3475
6261	G	G>A	CO1	1	10	1	3,13	46	0,6	0,1796
6340	C	C>T	CO1	1	0	1	3,13	10	0,13	0,0452
6663	A	A>G	CO1	1	1	1	3,13	5	0,07	0,0249
6932	A	A>G	CO1	1	3	1	3,13	0	0	0,0042*
7028	C	C>T	CO1	22	2	22	68,75	4200	55,23	0,1541
7175	T	T>C	CO1	1	1	1	3,13	18	0,24	0,0768
7184	A	A>G	CO1	1	1	1	3,13	20	0,26	0,0845
7196	C	C>A	CO1	1	3	1	3,13	26	0,34	0,1074
7256	C	C>T	CO1	1	3	1	3,13	66	0,87	0,2461
7274	C	C>T	CO1	1	1	1	3,13	19	0,25	0,0806
7385	A	A>G	CO1	1	1	1	3,13	57	0,75	0,2169
7476	C	C>T	TS1 (RNAt serina1)	2	1	2	6,25	145	1,91	0,1256
7521	G	G>A	TD (RNAt ácido aspártico)	1	10	1	3,13	86	1,13	0,3075
7699	C	C>T	CO2	1	0	1	3,13	1	0,01	0,0084*
7705	T	T>C	CO2	1	1	1	3,13	30	0,39	0,1223
7768	A	A>G	CO2	1	1	1	3,13	318	4,18	1
7771	A	A>G	CO2	1	2	1	3,13	20	0,26	0,0845
7789	G	G>A	CO2	2	6	2	6,25	82	1,08	0,0479
7805	G	G>A	CO2	1	9	1	3,13	111	1,46	0,3773
7873	C	C>T	CO2	1	1	1	3,13	21	0,28	0,0884
8068	C	C>T	CO2	1	1	1	3,13	0	0	0,0042*
8185	T	T>C	CO2	1	1	1	3,13	2	0,03	0,0125
8188	A	A>G	CO2	1	8	1	3,13	4	0,05	0,0208
8206	G	G>A	CO2	1	4	1	3,13	28	0,37	0,1148
8251	G	G>A	CO2	1	25	1	3,13	396	5,21	1
8269	G	G>A	CO2	1	7	1	3,13	151	1,99	0,4752
8404	T	T>C	ATP8	1	3	1	3,13	4	0,05	0,0208
8410	C	C>T	ATP8	1	3	1	3,13	9	0,12	0,0411
8557	G	G>A	ATP8	1	7	1	3,13	72	0,95	0,2651
8584	G	G>A	ATP6	1	5	1	3,13	31	0,41	0,126

Tabla 5. Continuación.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Presente estudio		Base de datos de población europea		Test exacto de Fisher
						Individuos N=32	Frecuencia (%)	Individuos N=7605	Frecuencia (%)	
8610	T	T>C	ATP6	1	2	1	3,13	10	0,13	0,0452
8614	T	T>C	ATP6	1	6	1	3,13	15	0,2	0,065
8697	G	G>A	ATP6	4	3	4	12,50	656	8,63	0,3528
8701	A	A>G	ATP6	2	9	2	6,25	189	2,49	0,1885
8734	C	C>A	ATP6	1	1	1	3,13	7	0,09	0,033
8760	T	T>C	ATP6	1	0	1	3,13	0	0	0,0042*
8843	T	T>C	ATP6	1	9	1	3,13	32	0,42	0,1296
8860	A	A>G	ATP6	32	4	32	100	7525	98,95	1
8911	T	T>C	ATP6	1	2	1	3,13	1	0,01	0,0084*
8994	G	G>A	ATP6	2	11	2	6,25	202	2,66	0,2101
9055	G	G>A	ATP6	2	12	2	6,25	677	8,9	1
9145	G	G>A	ATP6	1	1	1	3,13	10	0,13	0,0452
9221	A	A>G	CO3	1	1	1	3,13	27	0,36	0,1111
9230	T	T>C	CO3	1	2	1	3,13	1	0,01	0,0084*
9356	C	C>T	CO3	2	2	2	6,25	9	0,12	0,0009*
9477	G	G>A	CO3	2	4	2	6,25	633	8,32	1
9491	A	A>G	CO3	1	0	1	3,13	2	0,03	0,0125
9540	T	T>C	CO3	2	1	2	6,25	172	2,26	0,1646
9545	A	A>G	CO3	1	8	1	3,13	55	0,72	0,2102
9698	T	T>C	CO3	2	2	2	6,25	691	9,09	1
10057	T	T>C	TG (RNAt glicina)	1	0	1	3,13	1	0,01	0,0084*
10115	T	T>C	ND3	1	2	1	3,13	30	0,39	0,1223
10314	C	C>T	ND3	1	2	1	3,13	1	0,01	0,0084*
10398	A	A>G	ND3	12	21	12	37,50	1560	20,51	0,0263
10400	C	C>T	ND3	1	1	1	3,13	77	1,01	0,2805
10463	T	T>C	TR (RNAt arginina)	4	5	4	12,50	666	8,76	0,5227
10499	A	A>G	ND4L	2	2	2	6,25	90	1,18	0,0564
10550	A	A>G	ND4L	2	1	2	6,25	644	8,47	1
10586	G	G>A	ND4L	1	4	1	3,13	11	0,14	0,0492
10619	C	C>T	ND4L	1	2	1	3,13	14	0,18	0,0611
10685	G	G>A	ND4L	1	14(1)	1	3,13	42	0,55	0,1656
10822	C	C>T	ND4	1	1	1	3,13	35	0,46	0,1406
10873	T	T>C	ND4	2	1	2	6,25	169	2,22	0,1602
10927	T	T>C	ND4	1	2	1	3,13	61	0,8	0,23
11002	A	A>G	ND4	1	8	1	3,13	24	0,32	0,0998
11084	A	A>G	ND4	1	7	1	3,13	16	0,21	0,069
11242	C	C>G	ND4	2	1	2	6,25	8	0,11	0,0007*
11251	A	A>G	ND4	12	1	12	37,5	1326	17,44	0,0078*
11299	T	T>C	ND4	2	6	2	6,25	644	8,47	1
11332	C	C>T	ND4	1	1	1	3,13	160	2,1	0,495
11339	T	T>C	ND4	1	4	1	3,13	26	0,34	0,1074
11377	G	G>A	ND4	2	7	2	6,25	165	2,17	0,1543
11467	A	A>G	ND4	6	2	6	18,75	1782	23,43	0,677

Tabla 5. Continuación.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Presente estudio		Base de datos de población europea		Test exacto de Fisher
						Individuos N=32	Frecuencia (%)	Individuos N=7605	Frecuencia (%)	
11485	T	T>C	ND4	1	3	1	3,13	135	1,78	0,438
11674	C	C>T	ND4	1	2	1	3,13	167	2,2	0,51
11719	G	G>A	ND4	21	4	21	65,63	3854	50,68	0,1107
11812	A	A>G	ND4	4	2	4	12,50	493	6,48	0,1517
11840	C	C>T	ND4	1	1	1	3,13	86	1,13	0,3075
11908	A	A>G	ND4	1	3	1	3,13	4	0,05	0,0208
11914	G	G>A	ND4	4	43	4	12,50	254	3,34	0,0218
11944	T	T>C	ND4	1	8	1	3,13	44	0,58	0,1726
11947	A	A>G	ND4	1	3	1	3,13	166	2,18	0,5078
11950	A	A>G	ND4	1	1	1	3,13	4	0,05	0,0208
12007	G	G>A	ND4	1	14	1	3,13	69	0,91	0,2557
12046	A	A>G	ND4	1	0	1	3,13	0	0	0,0042*
12130	T	T>C	ND4	1	2	1	3,13	9	0,12	0,0411
12172	A	A>G	TH (RNAt histidina)	1	10	1	3,13	19	0,25	0,0806
12192	G	G>A	TH (RNAt histidina)	1	4	1	3,13	11	0,14	0,0492
12308	A	A>G	TL2 (RNAt leucina2)	6	3	6	18,75	1777	23,37	0,677
12351	T	T>C	ND5	1	1	1	3,13	5	0,07	0,0249
12372	G	G>A	ND5	7	7	7	21,88	1793	23,58	1
12414	T	T>C	ND5	1	8	1	3,13	170	2,24	0,5162
12597	A	A>G	ND5	1	0	1	3,13	1	0,01	0,0084*
12612	A	A>G	ND5	8	7	8	25	688	9,05	0,0067*
12618	G	G>A	ND5	1	9	1	3,13	107	1,41	0,3666
12693	A	A>G	ND5	1	1	1	3,13	19	0,25	0,0806
12696	T	T>C	ND5	1	7	1	3,13	9	0,12	0,0411
12705	C	C>T	ND5	3	1	3	9,38	681	8,95	0,7614
13032	A	A>G	ND5	1	4	1	3,13	7	0,09	0,033
13263	A	A>G	ND5	1	3	1	3,13	28	0,37	0,1148
13281	T	T>C	ND5	1	5	1	3,13	24	0,32	0,0998
13368	G	G>A	ND5	4	12	4	12,50	657	8,64	0,3535
13404	T	T>C	ND5	1	4	1	3,13	21	0,28	0,0884
13578	C	C>T	ND5	1	0	1	3,13	2	0,03	0,0125
13590	G	G>A	ND5	1	11	1	3,13	52	0,68	0,2001
13617	T	T>C	ND5	2	2	2	6,25	629	8,27	1
13650	C	C>T	ND5	1	2	1	3,13	72	0,95	0,2651
13708	G	G>A	ND5	8	37	8	25	847	11,14	0,0218
13722	A	A>G	ND5	2	7	2	6,25	78	1,03	0,0439
13731	A	A>G	ND5	1	1	1	3,13	3	0,04	0,0167
13740	T	T>C	ND5	1	2	1	3,13	87	1,14	0,3104
13803	A	A>G	ND5	1	1	1	3,13	21	0,28	0,0884
13879	T	T>C	ND5	1	9	1	3,13	66	0,87	0,2461
13911	A	A>G	ND5	1	1	1	3,13	2	0,03	0,0125
13933	A	A>G	ND5	1	3	1	3,13	28	0,37	0,1148

Tabla 5. Continuación.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Presente estudio		Base de datos de población europea		Test exacto de Fisher
						Individuos N=32	Frecuencia (%)	Individuos N=7605	Frecuencia (%)	
14091	A	A>G	ND5	1	1	1	3,13	3	0,04	0,0167
14106	T	T>C	ND5	1	1	1	3,13	2	0,03	0,0125
14133	A	A>G	ND5	2	6	2	6,25	85	1,12	0,051
14167	C	C>T	ND6	2	1	2	6,25	654	8,6	1
14179	A	A>G	ND6	1	2	1	3,13	89	1,17	0,3162
14182	T	T>C	ND6	1	10	1	3,13	327	4,3	1
14233	A	A>G	ND6	4	5	4	12,50	511	6,72	0,1662
14318	T	T>C	ND6	1	1	1	3,13	12	0,16	0,0532
14325	T	T>C	ND6	1	3	1	3,13	8	0,11	0,0371
14364	G	G>A	ND6	1	9	1	3,13	39	0,51	0,155
14374	T	T>C	ND6	1	1	1	3,13	3	0,04	0,0167
14428	C	C>T	ND6	1	0	1	3,13	2	0,03	0,0125
14443	C	C>T	ND6	1	0	1	3,13	1	0,01	0,0084*
14497	A	A>G	ND6	1	1	1	3,13	8	0,11	0,0371
14502	T	T>C	ND6	1	6	1	3,13	14	0,18	0,0611
14524	A	A>G	ND6	1	1	1	3,13	0	0	0,0042*
14566	A	A>G	ND6	1	2	1	3,13	19	0,25	0,0806
14620	C	C>T	ND6	1	2	1	3,13	163	2,14	0,5015
14766	C	C>T	CYB	21	4	21	65,63	3805	50,03	0,1095
14783	T	T>C	CYB	1	1	1	3,13	78	1,03	0,2835
14793	A	A>G	CYB	1	1	1	3,13	320	4,21	1
14798	T	T>C	CYB	7	4	7	21,88	1121	14,74	0,3122
14905	G	G>A	CYB	4	8	4	12,50	649	8,53	0,3478
15043	G	G>A	CYB	1	7	1	3,13	344	4,52	1
15257	G	G>A	CYB	2	3	2	6,25	178	2,34	0,1736
15301	G	G>A	CYB	2	6	2	6,25	151	1,99	0,1341
15316	A	A>G	CYB	1	0	1	3,13	5	0,07	0,0249
15326	A	A>G	CYB	32	8	32	100	7524	98,93	1
15340	A	A>G	CYB	1	1	1	3,13	8	0,11	0,0371
15424	C	C>T	CYB	1	0	1	3,13	1	0,01	0,0084*
15452	C	C>A	CYB	12	2	12	37,5	1325	17,42	0,0078*
15466	G	G>A	CYB	1	6	1	3,13	15	0,2	0,065
15487	A	A>T	CYB	1	2	1	3,13	26	0,34	0,1074
15607	A	A>G	CYB	4	3	4	12,50	648	8,52	0,3471
15693	T	T>C	CYB	1	1	1	3,13	163	2,14	0,5015
15721	T	T>C	CYB	1	7	1	3,13	22	0,29	0,0922
15784	T	T>C	CYB	1	11	1	3,13	82	1,08	0,2956
15790	C	C>T	CYB	1	1	1	3,13	21	0,28	0,0884
15871	A	A>G	CYB	1	0	1	3,13	0	0	0,0042*
15884	G	G>A	CYB	1	20	1	3,13	108	1,42	0,3693
15884	G	G>C	CYB	1	20	1	3,13	166	2,18	0,5078
15904	C	C>T	TT (RNAt treonina)	1	1	1	3,13	280	3,68	1
15928	G	G>A	TT (RNAt treonina)	4	5	4	12,50	648	8,52	0,3471

Tabla 5. Continuación.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Presente estudio		Base de datos de población europea		Test exacto de Fisher
						Individuos N=32	Frecuencia (%)	Individuos N=7605	Frecuencia (%)	
15930	G	G>A	TT (RNAt treonina)	1	14	1	3,13	21	0,27	0,0884
16051	A	A>G	HVRI: ATT	1	22	1	3,13	171	2,25	0,5183
16069	C	C>T	HVRI: ATT	8	6	8	25	693	9,11	0,007*
16093	T	T>C	HVRI: ATT / TAS2	1	50(10)	1	3,13	389	5,12	1
16126	T	T>C	HVRI: ATT / TAS2 / 7S DNA	12	19	12	37,50	1380	18,14	0,0095*
16145	G	G>A	HVRI: ATT / 7S DNA	3	32	3	9,38	229	3,01	0,072
16172	T	T>C	HVRI: ATT / TAS2 / 7S DNA	2	50	2	6,25	357	4,7	0,6615
16183	A	A>C	HVRI: ATT / 7S DNA	3	NA	3	9,38	377	4,96	0,2116
16186	C	C>T	HVRI: ATT / 7S DNA	1	3	1	3,13	144	1,9	0,4592
16189	T	T>C	HVRI: ATT / 7S DNA	5	111(5)	5	15,63	1130	14,86	0,8058
16192	C	C>T	HVRI: ATT / 7S DNA	2	37	2	6,25	443	5,82	0,71
16201	C	C>T	HVRI: ATT / mt5 / 7S DNA	1	1	1	3,13	16	0,21	0,069
16219	A	A>G	HVRI: ATT / 7S DNA	1	5	1	3,13	144	1,9	0,4592
16220	A	A>G	HVRI: ATT / 7S DNA	1	4	1	3,13	6	0,08	0,029
16222	C	C>T	HVRI: ATT / 7S DNA	1	6(1)	1	3,13	87	1,14	0,3104
16223	C	C>T	HVRI: ATT / 7S DNA	3	30(1)	3	9,38	710	9,34	1
16224	T	T>C	HVRI: ATT / 7S DNA	2	11	2	6,25	648	8,52	1
16231	T	T>C	HVRI: ATT / 7S DNA	2	5	2	6,25	86	1,13	0,0521
16256	C	C>T	HVRI: ATT / 7S DNA	1	36(1)	1	3,13	395	5,2	1
16261	C	C>T	HVRI: ATT / 7S DNA	3	30	3	9,38	274	3,6	0,1084
16270	C	C>T	HVRI: ATT / 7S DNA	2	17	2	6,25	663	8,72	1
16278	C	C>T	HVRI: ATT / 7S DNA	5	50(2)	5	15,63	360	4,73	0,0168
16292	C	C>T	HVRI: ATT / 7S DNA	1	24	1	3,13	218	2,87	0,6066
16293	A	A>G	HVRI: ATT / 7S DNA	1	20(2)	1	3,13	143	1,88	0,4569
16294	C	C>T	HVRI: ATT / 7S DNA	5	29	5	15,63	753	9,9	0,2418
16296	C	C>T	HVRI: ATT / 7S DNA	3	2(1)	3	9,38	336	4,42	0,1678
16298	T	T>C	HVRI: ATT / 7S DNA	2	17	2	6,25	355	4,67	0,6605
16300	A	A>G	HVRI: ATT / 7S DNA	1	13	1	3,13	19	0,25	0,0806
16304	T	T>C	HVRI: ATT / 7S DNA	2	27(1)	2	6,25	574	7,55	1
16309	A	A>G	HVRI: ATT / 7S DNA	1	13(2)	1	3,13	42	0,55	0,1656
16311	T	T>C	HVRI: ATT / 7S DNA	4	133(4)	4	12,50	1313	17,27	0,6403
16316	A	A>G	HVRI: ATT / 7S DNA	1	8	1	3,13	49	0,65	0,1899
16325	T	T>C	HVRI: ATT / 7S DNA	3	24(1 del)	3	9,38	61	0,8	0,0023*
16327	C	C>T	HVRI: ATT / 7S DNA	1	10	1	3,13	23	0,3	0,096
16356	T	T>C	HVRI: ATT / 7S DNA	1	15	1	3,13	265	3,49	1
16362	T	T>C	HVRI: ATT / 7S DNA	1	85(1)	1	3,13	492	6,47	0,7202
16366	C	C>T	HVRI: ATT / 7S DNA	1	2	1	3,13	24	0,32	0,0998
16390	G	G>A	Región Control: ATT / 7S DNA	1	32	1	3,13	121	1,59	0,4033
16465	C	C>T	Región Control: ATT / 7S DNA	1	8	1	3,13	27	0,35	0,1111
16519	T	T>C	Región Control: ATT / 7S DNA	21	NA	21	65,63	5025	66,08	1
16526	G	G>A	Región Control: ATT / 7S DNA	1	10	1	3,13	113	1,49	0,3826

*El test exacto de Fisher muestra diferencias significativas ($p < 0,01$).

Tabla 6. Mutaciones en homoplasmia localizadas en posiciones estables (N=251). ATT: sitio de unión a la membrana; OH: origen de replicación; TFY: sitio de unión de TF1.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Índice de Conservación (%)
57	T	T>C	HVRII / ATT / 7S DNA / OH57	1	4	93,8
292	T	T>C	HVRII: ATT / TFY / OH	1	1 (1 ins)	18,8
489	T	T>C	HVRIII: ATT	9	3	18,8
739	C	C>T	RNAr 12S	1	2	16,67
750	A	A>G	RNAr 12S	32	4	100
1018	G	G>A	RNAr 12S	1	2	8,33
1243	T	T>C	RNAr 12S	2	5	83,33
1850	T	T>C	RNAr 16S	1	1	83,33
3173	G	G>A	RNAr 16S	1	0	8,33
3197	T	T>C	RNAr 16S	2	4	0
3348	A	A>G	ND1	1	1	91,67
3447	A	A>G	ND1	1	4	100
3480	A	A>G	ND1	2	3	100
3505	A	A>G	ND1	1	5	83,33
3672	A	A>G	ND1	1	1	100
3721	A	A>G	ND1	1	1	91,67
4093	A	A>G	ND1	1	5	83,33
4104	A	A>G	ND1	1	1	100
4164	A	A>G	ND1	1	4	100
4224	C	C>T	ND1	1	1	91,67
4580	G	G>A	ND2	1	2	100
4646	T	T>C	ND2	1	5	100
4715	A	A>G	ND2	1	3	100
4733	T	T>C	ND2	1	2	83,33
4879	C	C>G	ND2	1	0	100
4917	A	A>G	ND2	4	4	83,33
5187	C	C>T	ND2	1	1	83,33
5780	G	G>A	TC (RNAc cisteína)	1	1	91,67
5964	T	T>C	CO1	1	4	83,33
5999	T	T>C	CO1	1	3	100
6047	A	A>G	CO1	1	2	91,67
6340	C	C>T	CO1	1	0	83,33
6663	A	A>G	CO1	1	1	91,67
6932	A	A>G	CO1	1	3	100
7028	C	C>T	CO1	22	2	100
7175	T	T>C	CO1	1	1	100
7184	A	A>G	CO1	1	1	100
7196	C	C>A	CO1	1	3	100
7256	C	C>T	CO1	1	3	100
7274	C	C>T	CO1	1	1	100
7385	A	A>G	CO1	1	1	100
7699	C	C>T	CO2	1	0	91,67

Tabla 6. Continuación.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Índice de Conservación (%)
7705	T	T>C	CO2	1	1	100
7771	A	A>G	CO2	1	2	100
7873	C	C>T	CO2	1	1	100
8068	C	C>T	CO2	1	1	100
8185	T	T>C	CO2	1	1	100
8206	G	G>A	CO2	1	4	100
8404	T	T>C	ATP8	1	3	91,67
8697	G	G>A	ATP6	4	3	91,67
8734	C	C>A	ATP6	1	1	100
8760	T	T>C	ATP6	1	0	100
8911	T	T>C	ATP6	1	2	100
9145	G	G>A	ATP6	1	1	100
9230	T	T>C	CO3	1	2	100
9356	C	C>T	CO3	2	2	91,67
9477	G	G>A	CO3	2	4	83,33
9491	A	A>G	CO3	1	0	91,67
9540	T	T>C	CO3	2	1	91,67
9698	T	T>C	CO3	2	2	T:83,33
10057	T	T>C	TG (RNAt glicina)	1	0	83,33
10314	C	C>T	ND3	1	2	83,33
10463	T	T>C	TR (RNAt arginina)	4	5	91,67
10499	A	A>G	ND4L	2	2	100
10550	A	A>G	ND4L	2	1	100
10586	G	G>A	ND4L	1	4	100
10619	C	C>T	ND4L	1	2	100
11242	C	C>G	ND4	2	1	100
11251	A	A>G	ND4	12	1	100
11467	A	A>G	ND4	6	2	100
11485	T	T>C	ND4	1	3	100
11674	C	C>T	ND4	1	2	100
11719	G	G>A	ND4	21	4	91,67
11812	A	A>G	ND4	4	2	100
11947	A	A>G	ND4	1	3	91,67
11950	A	A>G	ND4	1	1	100
12308	A	A>G	TL2 (RNAt leucina 2)	6	3	100
12597	A	A>G	ND5	1	0	91,67
12693	A	A>G	ND5	1	1	100
13032	A	A>G	ND5	1	4	100
13263	A	A>G	ND5	1	3	100
13281	T	T>C	ND5	1	5	83,33
13404	T	T>C	ND5	1	4	100
13617	T	T>C	ND5	2	2	91,67
13650	C	C>T	ND5	1	2	83,33

Tabla 6. Continuación.

Posición	rCRS	Mutación	Región mitocondrial	Frecuencia	Hits Filogenia	Índice de Conservación (%)
13731	A	A>G	ND5	1	1	100
13740	T	T>C	ND5	1	2	100
13911	A	A>G	ND5	1	1	83,33
14091	A	A>G	ND5	1	1	100
14106	T	T>C	ND5	1	1	91,67
14167	C	C>T	ND6	2	1	100
14233	A	A>G	ND6	4	5	83,33
14374	T	T>C	ND6	1	1	91,67
14428	C	C>T	ND6	1	0	100
14443	C	C>T	ND6	1	0	100
14497	A	A>G	ND6	1	1	100
14524	A	A>G	ND6	1	1	83,33
14566	A	A>G	ND6	1	2	100
14620	C	C>T	ND6	1	2	100
14793	A	A>G	CYB	1	1	91,67
15257	G	G>A	CYB	2	3	83,33
15316	A	A>G	CYB	1	0	16,67
15340	A	A>G	CYB	1	1	91,67
15452	C	C>A	CYB	12	2	91,67
15487	A	A>T	CYB	1	2	100
15607	A	A>G	CYB	4	3	100
15790	C	C>T	CYB	1	1	16,67
16219	A	A>G	HVRI: ATT / 7S DNA	1	5	18,8

Tabla 7. Mutaciones en heteroplasmia localizadas en posiciones estables (N=23). (T): tumor; (S): sangre. ATT: sitio de unión a la membrana.

Posición	rCRS	Mutación (base mayoritaria/ base minoritaria)	Porcentaje de heteroplasmia (% base minoritaria)		Región mitocondrial	Origen	Frecuencia	Hits Filogenia	Índice de Conservación (%)
55C	T	C/T	6,57		Región control: ATT / 7S DNA	Somático	1	3	T:93,8
1167R	A	A/G	3,41		RNAr 12S	Somático	1	0	A:100
1767R	G	G/A	2,44 (T)	13,49 (S)	RNAr 16S	Germinal	1	0	G:100
1787R	G	G/A	5,07		RNAr 16S	Somático	1	0	G:100
2251R	A	A/G	3,69		RNAr 16S	Somático	1	1	A:91,67
2449R	G	G/A	5,46		RNAr 16S	Somático	1	0	G:100
3958R	G	G/A	3,24		ND1	Somático	1	0	G:100
4011T	C	T/C	5,78 (T)	18,39 (S)	ND1	Germinal	1	4	C:100
4157W	A	A/T	5,15 (T)	0,78 (S)	ND1	Germinal	2	2	A:100
			0,83 (T)	4,85 (S)					
4214R	G	G/A	49,19		ND1	Somático	1	0	G:100
4807C	T	C/T	40,32		ND2	Somático	1	0	T:100
6531T	C	T/C	28,67		CO1	Somático	1	1	C:100
7131R	G	G/A	45,96 (T)	4,15 (S)	CO1	Germinal	1	0	G:100
8554R	A	A/G	9,39		ATP6	Somático	1	1	A:100
8736W	T	T/A	47,16 (T)	30,89 (S)	ATP6	Germinal	1	0	T:100
9031T	C	T/C	36,32		ATP6	Somático	1	1	C:100
9361Y	T	T/C	1,16 (T)	3,97 (S)	CO3	Germinal	1	0	T:91,67
11585W	T	T/A	5,75		ND4	Somático	1	0	T:91,67*
11711R	G	G/A	4,48		ND4	Somático	1	0	G:100
12746C	T	C/T	14,47		ND5	Somático	1	0	T:100
14791T	C	T/C	29,89 (T)	25,79 (S)*	CYB	Germinal	1	0	C:100
14793R	A	A/G	7,83 (T)	30,64 (S)	CYB	Germinal	1	1	A:91,67

*La base minoritaria difiere respecto al tejido tumoral.

