

Universidad Miguel Hernández de Elche

Facultad de Ciencias Experimentales

Grado en Biotecnología

Curso 2024-2025



Idoneidad de las aproximaciones ómicas en estudios oncológicos a partir de tejido fijado y parafinado.

Trabajo de fin de grado

Autora: Alejandra Ocaña Vidal

Tutor: Luis Miguel Valor Becerra

Departamento: Bioquímica y Biología Molecular

Área: Investigación en Genética, Epigenética y Transcripción en
Neuropatologías

Índice

ÍNDICE.....	1
RESUMEN	3
ABSTRACT	3
INTRODUCCIÓN	4
1. GLIOMAS	4
1.1 <i>Astrocitomas</i>	4
1.2 <i>Oligodendrogliomas</i>	5
1.3 <i>Glioblastoma</i>	5
2. EL GLIOBLASTOMA EN MAYOR DETALLE	5
2.1 <i>Prevalencia y factores de riesgo</i>	5
2.2 <i>Síntomas y diagnóstico</i>	6
2.3 <i>Clasificación histológica</i>	7
2.4 <i>Clasificación molecular</i>	7
2.4.1. <i>Mutaciones somáticas</i>	7
2.4.2. <i>Alteraciones transcripcionales</i>	8
3. EPIGENÉTICA	9
3.1 <i>Definición y modificaciones epigenéticas</i>	9
3.2 <i>Metilación del ADN</i>	11
3.3 <i>Alteraciones epigenéticas en glioblastoma</i>	11
4. FIJACIÓN POR FORMALINA Y PARAFINADO	12
ANTECEDENTES Y OBJETIVOS.....	13
1. ANTECEDENTES.....	13
2. OBJETIVO	14
MATERIALES Y MÉTODOS	15
1. SETS DE DATOS	15
2. RECURSOS INFORMÁTICOS <i>ONLINE</i>	15
3. PROGRAMAS	16
3.1 <i>Excel</i>	16
3.2 <i>RStudio</i>	16
4. VALIDACIÓN EXPERIMENTAL	17
4.1 <i>Muestras</i>	17
4.2 <i>Método</i>	17
RESULTADOS.....	18

1.	IDENTIFICACIÓN DE LOS GENES MÁS DIFERENCIALMENTE EXPRESADOS ENTRE TUMORES FRESCOS Y FFPE	18
1.1	<i>Filtrado de los datos transcriptómicos</i>	18
1.2	<i>Clasificación de genes</i>	18
2.	LOS GENES MEJOR EXPRESADOS EN TUMORES FRESCOS TIENDEN A SER MÁS CORTOS Y CON MAYOR CONTENIDO EN G/C QUE LOS GENES EXPRESADOS EN TUMORES FFPE	19
3.	LOS GENES MEJOR EXPRESADOS EN TEJIDO FRESCO TIENDEN A TENER UN MAYOR VALOR DE MFE (MINIMUM FREE ENERGY) QUE LOS GENES EN TEJIDO FFPE.	21
4.	LOS GENES MÁS REPRESENTADOS EN TEJIDO FRESCO Y PARAFINADO ESTÁN ASOCIADOS A DIFERENTES FUNCIONES BIOLÓGICAS.....	24
5.	VALIDACIÓN DE LAS DIFERENCIAS DE ABUNDANCIA DE TRANSCRITOS ENTRE TEJIDO FRESCO Y PARAFINADO MEDIANTE QPCR.	26
6.	LOS PERFILES DE METILACIÓN ENTRE AMBOS TIPOS DE TEJIDOS PRESENTAN DIFERENCIAS EN SU POSICIÓN GENÓMICA.....	28
	DISCUSIÓN.....	31
	CONCLUSIONES Y PROYECCIÓN FUTURA.....	35
	BIBLIOGRAFÍA.....	36



Resumen

El glioblastoma (GBM) es uno de los tumores cerebrales más agresivos y de peor pronóstico en la actualidad, lo que resalta la importancia clínica del uso de biomarcadores fiables para su diagnóstico y seguimiento. En los últimos años, se ha prestado una mayor atención a los efectos que el proceso de fijación y embebido en parafina (FFPE) puede tener sobre el material genético, y cómo estos pueden comprometer la fiabilidad de ciertos biomarcadores moleculares.

En este estudio se ha llevado a cabo un análisis bioinformático de más de 30.000 genes procedentes de tejido fresco y más de 40.000 de tejido parafinado, con el objetivo de identificar posibles diferencias estructurales y funcionales en el ARN que puedan influir en su susceptibilidad al proceso de parafinado. Algunos de los resultados obtenidos fueron validados experimentalmente mediante PCR a tiempo real (qPCR). Además, se analizaron los perfiles de metilación en sitios CpG de ambos tipos de tejido, con el fin de evaluar si existen variaciones epigenéticas que puedan contribuir a las diferencias observadas.

Palabras clave: Glioblastoma, FFPE, parafinado, metilación

Abstract

Glioblastoma (GBM) is one of the most aggressive and with the worst prognosis cerebral tumor nowadays, highlighting the clinical importance of reliable biomarkers for its diagnosis and monitoring. In recent years, the effects of the fixation and paraffin-embedding process (FFPE) on the genetic material and how this process may compromise the reliability of certain biomarkers have been considered.

In this study, a bioinformatic analysis of more than 30.000 genes from fresh tissues and more than 40.000 genes from paraffin-embedded tissues was carried out to identify if there are structural and functional differences in RNA that may influence their susceptibility to the paraffin-embedded process. Some of these results were experimentally validated by real-time PCR (qPCR). Moreover, metylation profiles of CpG sites from both tissue types were analyzed in order to evaluate whether there are epigenetic variations that may contribute to the observed differences.

Key words: Glioblastoma, FFPE, paraffin-embedded, metilation

Introducción

1. Gliomas

Los gliomas son el tipo de tumor más frecuente del sistema nervioso central (SNC). Se clasifican en distintos subtipos generalmente asociados a diferentes tasas de supervivencias. Originalmente esta clasificación de los gliomas se realizaba a partir de la histología del tumor, pero en los últimos años se ha visto que los patrones moleculares son altamente relevantes en la clínica, permitiendo realizar una clasificación más exacta a la hora de determinar el pronóstico de los pacientes. Así, con la 5ª y última edición de la clasificación de los tumores del SNC de la Organización Mundial de la Salud (OMS o WHO por sus siglas en inglés, *World Health Organization*), aprobada en 2021, se han asignado los gliomas adultos en tres tipos principales de tumores: **Astrocitomas**, que presentan mutaciones en Isocitrato Deshidrogenasa 1 (IDH1, aminoácido R132) o 2 (IDH2, aminoácido R172) ; **Oligodendrogliomas** que además de ser IDH mutantes también presentan co-delección en el brazo corto del cromosoma 1 y en el brazo largo del cromosoma 19 (-1p/-19q); y los **Glioblastomas (GBM)**, que no presentan mutaciones en IDH1/2 (IDH nativos). [13]

1.1 Astrocitomas

Los astrocitomas son tumores que se dan en el cerebro de adultos. Normalmente se diagnostican en jóvenes adultos (menores de 40) y es raro en mayores de 60, a diferencia del GBM, que se da mayoritariamente en ancianos. Además, presenta una incidencia 1.5 veces mayor en hombres que en mujeres.[13, 14]. En cuanto a la esperanza de vida de los pacientes varía con el grado del tumor, siendo de más de 10 años para aquellos con astrocitomas de grado 2, pero bajando a 5 en aquellos de grado 3 e incluso menos en los diagnosticados de grado 4, que son muy poco frecuentes. [13].

Con la última revisión de la OMS, los GBM y astrocitomas son considerados entidades separadas en función de la mutación en IDH [14], y actualmente los astrocitomas incluyen a los previamente conocidos como glioblastomas secundarios, presumiblemente malignizados a partir de un astrocitoma [26]. También presentan normalmente la inactivación de los genes *ATRX* y *TP53*, el primero produciendo una proteína esencial para el correcto desarrollo, ya que

se relaciona con la remodelación de la cromatina; y el segundo expresando la proteína p53, considerado el guardián del genoma y el principal supresor de tumores. [13]

1.2 Oligodendrogliomas

Los oligodendrogliomas son tumores intracraneales que representan entre el 5% y el 25% de todos los gliomas y el 2-5% de los tumores cerebrales primarios. Se manifiestan principalmente en adultos de mediana edad (entre 40 y 60 años), con ligera mayor prevalencia en hombres y son bastante raros en niños. Se clasifican como de grado 2 o 3, por lo que generalmente son menos agresivos que los otros gliomas difusos de tipo adulto [15], con una media superior a los 10 años. Primariamente estos tumores se diagnosticaban únicamente con pruebas histológicas, mientras que ahora se le añaden las pruebas moleculares para confirmar que también presentan mutación en IDH1/2 y codeleción completa de 1p19q. No obstante, estos marcadores pueden no estar disponibles o arrojar resultados ambiguos. En dichas circunstancias, el diagnóstico se fundamenta exclusivamente en las características histológicas, clasificándose entonces como no especificado (NOS, *not otherwise specified*). [15]

1.3 Glioblastoma

El GBM es un tipo de cáncer que se desarrolla principalmente en el cerebro, con pocos casos en la médula espinal. Desde el punto de vista histológico, se clasifica como grado 4, el más agresivo, debido a su rápido e invasivo crecimiento. Una de sus principales características es su capacidad de infiltrarse en las periferias del parénquima cerebral. A continuación, nos centraremos en este cáncer cerebral por su relevancia clínica.

2. El glioblastoma en mayor detalle

2.1 Prevalencia y factores de riesgo

Los GBM constituyen la mitad de los gliomas y son el tipo de tumor primario más común del SNC, aunque su incidencia anual es relativamente baja, entorno a los 3,1 casos por cada 100.000 habitantes, en comparación con otros cánceres como el de mama (171,2 de 100.000) o el de próstata (201,4 de 100.000). Esta incidencia parece variar también con la raza,

ya que por ejemplo desciende a los 0,6 casos por cada 100.000 en Asia, y con la edad, pasando de 0,15 casos por cada 100.000 habitantes en niños a 15 por cada 100.000 en mayores de 75 años. [3]

Debido a su carácter agresivo, la supervivencia media de los pacientes es de aproximadamente 12 meses con los tratamientos actuales, aumentando a 15 con algunos ensayos clínicos. El porcentaje de supervivencia en 5 años varía dependiendo de la bibliografía consultada, aunque normalmente tiene valores bajos (entre el 5-7%), y está inversamente asociado a la edad del paciente. [2, 17]

Los factores de riesgo que llevan al desarrollo del GBM a día de hoy son en su mayoría desconocidos, ya que el único factor ambiental conocido es la radiación ionizante y menos del 1% de los GBM se pueden asociar a un carácter hereditario, entre los que se incluyen la neurofibromatosis tipo 1 y 2 o el síndrome de Turcot. [1,3]

2.2 Síntomas y diagnóstico

Los síntomas de este tumor dependen principalmente de su localización y del área del cerebro a la que afecte. Además, según adquiere tamaño va aumentando la presión intracraneal, llevando a síntomas como problemas de memoria y habla, pérdidas de conciencia, convulsiones, vómitos, cambios de ánimo y comportamiento o dificultades de visión. [1]

Ante estos síntomas persistentes, el paciente se deriva a un análisis radiológico, como la **resonancia magnética (MRI)**, la cual permite localizar el tumor, diferenciarlo de otras patologías y hallar las áreas de edemas e isquemias relativas al GBM, y la **tomografía computarizada (CT)**, que suele ser la primera prueba que se realiza a los pacientes con síntomas graves y que permite encontrar las zonas que podrían contener tumores junto con aquellas en las que se haya dado recientemente sangrado o calcificación. La primera aproximación es la más frecuente, mientras que la segunda se realiza bajo situaciones comprometidas, como cuando hay sospecha de hemorragia o no es posible hacer MRI. [1,3]

Una vez localizado el tumor, se puede llevar a cabo una biopsia, lo que permite extraer un fragmento del tumor, con el objetivo de analizarlo histológica y molecularmente, o una cirugía para eliminar gran parte o la totalidad del mismo. Estas pruebas permiten realizar el

examen histológico, que se confirma con el análisis de la mutación en *IDH1/IDH2*, además de otras alteraciones moleculares de interés clínico como veremos a continuación. Con esta información, es posible afinar el pronóstico y, en casos muy concretos, personalizar el tratamiento ya que este se puede focalizar en las dianas sensibles del tumor. [1, 3]

2.3 Clasificación histológica.

Los biomarcadores moleculares son fundamentales para el diagnóstico, con el objetivo de disminuir la dependencia de los factores morfológicos en la clasificación de los tumores. Esto se debe a que la clasificación histopatológica es altamente subjetiva, mostrando variabilidad intra- e Inter-observador. [13]

El GBM es un cáncer que suele ser muy heterogéneo celularmente y cuyos patrones histológicos se caracterizan principalmente necrosis, actividad mitótica y proliferación microvascular [3,6]. También es común la presencia de células astrocíticas con marcada atipia nuclear o, en otras palabras, células gliales que presentan una alteración en la forma y tamaño de su núcleo. Los GBM pueden clasificarse según el predominio de un tipo celular específico, como el **glioblastoma de células gigantes** (GC-GBM), el **gliosarcoma** (GS) y el **glioblastoma epitelioide** (E-GBM). El GC-GBM contiene numerosas células gigantes multinucleadas, mientras que el E-GBM se caracteriza por células epitelioides y algunas células rabdoides con alta actividad mitótica. [13, 16]

2.4 Clasificación molecular.

2.4.1. Mutaciones somáticas

Como hemos mencionado anteriormente, la principal característica del GBM es que no presenta mutaciones en *IDH1/2*. La diferencia más importante entre estos genes es que codifican para proteínas con diferente localización subcelular, encontrándose IDH1 en el citoplasma y el peroxisoma, mientras que IDH2 es mitocondrial. Las proteínas IDH son homodímeros que catalizan la descarboxilación oxidativa del isocitrato (IC) a α -ceto-glutarato (α -CG) y CO_2 al reducirse NADP^+ a NADPH. El α -CG se utiliza para la generación de glutatión reducido, el principal antioxidante celular, que favorece la supervivencia celular y la antiapoptosis. Las mutaciones en estos genes son de pérdida de función, dando lugar a la

generación de 2-hidroxiglutarato (2-HG), considerado un oncometabolito, y causando una disminución en los niveles de α -CG y un incremento en los niveles de radicales libres. [27]

Otros patrones moleculares característicos incluyen mutaciones en los siguientes genes:

- *TP53*, que codifica para **p53**, el “guardián del genoma”.
- Genes reguladores del receptor tirosina quinasa (**RTK**), del sarcoma de rata (**RAS**) o fosfatidilinositol 3 quinasa (**PI3K**).
- La proteína de señalización del retinoblastoma (**RB**), un supresor tumoral .
- Promotor del gen **TERT**, involucrado en el mantenimiento de telómeros.
- **EGFR**, el Receptor del Factor de Crecimiento Epidérmico o *Epidermic Grow Factor Receptor* en inglés, que controla el crecimiento y proliferación celular.
- Los genes **CDKN2A/2B** que controlan el ciclo celular y la apoptosis.

Además, suelen presentar también aberraciones cromosómicas, siendo las más destacadas la ganancia de una copia del cromosoma 7 y/o la pérdida de una copia del cromosoma 10 (-7/+10). [13, 20]

2.4.2. Alteraciones transcripcionales

Para una comprensión más profunda del GBM y su microambiente tumoral, se ha incorporado el análisis transcriptómico. Esta aproximación permite identificar genes y rutas de señalización con patrones de expresión alterados en el tumor, lo que puede contribuir tanto al diagnóstico como al pronóstico del paciente. Entre los hallazgos más relevantes, se han descrito tres firmas transcripcionales principales que permiten clasificar el GBM en los subtipos **proneural** (PN), **clásico** (CL) y **mesenquimal** (MES). Asimismo, en un mismo GBM pueden coexistir múltiples subtipos, pudiendo incluso variar la proporción de cada uno de ellos a lo largo del tiempo. Esta heterogeneidad hace que los GBMs sean más reacios a responder a las diferentes terapias, dificultando su tratamiento. Además, se ha observado que la firma transcripcional con peor pronóstico clínico es la del subtipo MES. A su vez, la secuenciación de ARN de célula única (sc-RNAseq) y los estudios de trayectoria de linajes han mostrado que las células del GBM poseen plasticidad y capacidad para pasar de un estado a

otro, lo que sugiere que tanto el microambiente del tumor, la genética y los programas de neurodesarrollo pueden influir en las firmas transcripcionales. Los 4 estados principales son: *tipo células progenitoras neurales*, *tipo células progenitoras de oligodendrocitos*, *tipo mesenquimal* y *tipo astrocitario*. Por tanto, los subtipos de GBM no son entidades estáticas, ya que reflejan diferentes proporciones de distintos estados celulares, que pueden cambiar a lo largo del desarrollo del tumor. [28, 29]. Aunando scRNA-seq y transcriptómica espacial se han identificado distintas subpoblaciones dentro del GBM, incluidas las **células proliferativas malignas**, las **células similares a las células troncales**, las **células similares a las células mesenquimales** y las **células relacionadas con el sistema inmunitario**, que contribuyen a la progresión tumoral, la resistencia al tratamiento y las interacciones en el microambiente del tumor, mostrando diferencias tanto en la distribución espacial como en los perfiles de expresión génica y las características funcionales de estas subpoblaciones, entre las que destacan diferencias en las actividades metabólicas entre el núcleo tumoral y la periferia del mismo. [30]

Estas alteraciones transcripcionales pueden suceder por alteraciones epigenéticas que se explica a continuación en mayor detalle.

3. Epigenética

3.1 Definición y modificaciones epigenéticas

La epigenética hace referencia al estudio de los cambios que se producen en el ADN que no afectan a la secuencia original y que, por tanto, pueden ser reversibles. Estas marcas epigenéticas cambian la forma en la que se expresan los genes, y hay casos en los que dichas modificaciones pueden transmitirse de células madre a hijas. Al conjunto de todos los cambios epigenéticos de un genoma se denomina **epigenoma**. [9, 10] La epigenética no solo permite la regulación transitoria de la expresión génica, sino que también permite el “encendido/apagado” de gran parte del genoma. Este control da lugar a la diferenciación y mantenimiento de la identidad celular a lo largo de la vida a partir de un único genoma presente en todas las células nucleadas del organismo, pero permitiendo la expresión únicamente de los genes que le corresponden a cada tipo y tejido celular. Además, la

epigenética también regula eventos moleculares como la impronta génica, la inactivación de un cromosoma X en mujeres, el silenciamiento de algunos elementos transponibles y, en general, contribuye a la estabilidad del genoma. [11]

Existen varios tipos de marcas epigenéticas, que podemos dividir en los siguientes grupos:

- **Modificaciones covalentes del ADN:** 5- metilcitosina (5-mC), 5- hidroximetilcitosina (5-hmC), 5-formilcitosina (5-fC), 5-carboxilcitosina (5-caC), N6-metiladenina (6-mA) [11]
- **Modificaciones de histonas:** Acetilación y metilación de las lisinas de las colas de histonas (Kac, Kme1, Kme2, Kme3 si están mono, di o trimetiladas, respectivamente), Fosforilación, Ubiquitinación, Sumoilación, similar a la ubiquitinación pero mediada por la proteína SUMO, etc.
- **Remodelación de histonas dependientes de ATP:** familias ISWI (*Imitation Switch*), CHD (helicasa con cromodominio de unión a ADN), SWI/SNF (*Switch/Sucrose Non Fermentable*), INO80
- **Intercambio de variantes de histonas:** H3.3, H2A.Z, H2A.X...
- **ARN no codificantes:** Xist, piARN, ANRIL, HOTAIR...

Aprovechando la naturaleza reversible de estas modificaciones, se han desarrollado varios fármacos epigenéticos, como inhibidores de ADN metiltransferasas o DNMTs o inhibidores de la deacetilasa de histonas. Estos fármacos van dirigidos a enzimas conocidas como “**escritoras**” (introducen marcas epigenéticas), “**lectoras**” (reconocen dichas marcas) o “**borradoras**” (eliminan las marcas), tanto en histonas como en el propio ADN. No obstante, dado que muchas de estas enzimas son ubicuas y constitutivas, es frecuente que se produzcan efectos adversos, afectando también a células normales y alterando sus patrones epigenéticos. Esta falta de especificidad terapéutica se ha observado con mayor impacto en niños y adolescentes, ya que los tejidos en desarrollo son especialmente sensibles. En contraste, en adultos, la toxicidad suele ser menor, ya que estos fármacos tienden a actuar preferentemente sobre células en proliferación, mientras que las células diferenciadas presentan una menor susceptibilidad. [11]

3.2 Metilación del ADN

La metilación o adición de grupos metilo al carbono 5 de la citosina (5-mC) está asociado con una desactivación o “apagado” del gen, por lo que no se expresará. Esta modificación está catalizada por las **DNMTs**. Por su parte, la desmetilación es en gran parte desconocida, pero se sabe que uno de los posibles mecanismos se produce a través de las enzimas TET (del inglés *Ten-Eleven Translocation*) y TDG (*Thymine DNA Glycosylase*), involucradas en la oxidación de 5-mC para producir hidroximetilcitosina (5-hmC), formilcitosina (5-fC) y carboxilcitosina (5-caC) y la posterior escisión de estos últimos.

En el ser humano dicha metilación se produce en la citosina de los dinucleótidos CpG, los cuales suelen concentrarse en regiones específicas denominadas **islas CpG**. Estas regiones ricas en CpG se suelen asociar con la región reguladora 5' (promotores) de muchos genes. Se ha visto que la hipometilación global a nivel de ADN contribuye a la carcinogénesis al disminuir la estabilidad cromosómica, reactivar algunos elementos transponibles o alterar la identidad celular y la impronta génica. Sin embargo, también se ha visto que un aumento de metilación de manera particular en islas CpG de los promotores de genes supresores de tumores como p15 o p16 pueden inactivar dichos genes. [9, 10, 11, 31]

Estos estados anómalos en el metiloma de los tumores pueden ser usados como dianas terapéuticas. Por ejemplo, se plantea el uso de fármacos desmetiladores para combatir la hipermetilación de las islas de CpGs y así reactivar algunos genes de interés. Un ejemplo es la **decitabina**, una molécula similar a la citosina pero que no se puede metilar, lo que da como resultado la expresión de varios genes antes silenciados. Sin embargo, como mencionamos previamente, también puede causar inestabilidad cromosómica y carece de una gran selectividad, afectando a células sanas. Por ello, se han desarrollado otros fármacos de 2ª generación, como la **guadecitabina**, con mejor estabilidad molecular y captación por las células cancerosas, o **GSK3685032**, un inhibidor reversible y selectivo de DNMTs con una mejor tolerabilidad. [11]

3.3. Alteraciones epigenéticas en glioblastoma

El epigenoma es dinámico y puede variar en función de factores ambientales, señales fisiológicas y necesidades celulares transitorias. Sin embargo, cuando estos cambios

epigenéticos no están adecuadamente regulados, pueden contribuir al desarrollo de diversas patologías. Las alteraciones epigenéticas aberrantes se han asociado con una amplia gama de enfermedades, que incluyen desde trastornos autoinmunes hasta distintos tipos de cáncer. En el caso del glioblastoma se han descrito múltiples alteraciones, pero cabe destacar por su implicación clínica la hipermetilación del promotor de **MGMT** que lleva al silenciamiento del gen que codifica para la O-6-metilguanina-ADN metiltransferasa, una enzima reparadora del ADN. De esta manera, el tumor tiene más posibilidades de ser sensible a la quimioterapia con temozolomida (TMZ), un agente alquilante que introduce grupos metilo en el ADN promoviendo roturas que pueden ser reparadas por *MGMT* [2, 6, 13]. También se ha visto que una baja expresión de *DNMT1* se asocia a GBM resistentes a TMZ. [31]

Por su parte, la modificación de las histonas se ha visto que cumple un papel esencial en la regulación de la proliferación y propagación celular así como en la resistencia a ciertos regímenes terapéuticos en el GBM. Un ejemplo sería la proteína potenciadora del homólogo zeste 2 o “*Enhancer of Zeste Homolog 2*” (**EZH2**), una histona metiltransferasa implicada en incremento de expresión de c-MYC, estrechamente implicado en la tumorigénesis del GBM y que disminuye las tasas de supervivencia. Otras enzimas involucradas en los procesos de metilación y acetilación de histonas pueden ver alteradas su expresión, contribuyendo a la proliferación, desarrollo y resistencia a terapia de dicho tumor. [31]

4. Fijación por formalina y parafinado

Las muestras fijadas con formalina y embebidas en parafina o FFPE (*Formalin-Fixed Paraffin-Embedded*) son aquellas obtenidas mediante biopsia o cirugía y que se conservan en este estado para poder acceder a ellas nuevamente en un futuro, bien con un fin de investigación o para un segundo diagnóstico. Estas muestras se suelen conservar en departamentos de Anatomía Patológica y Servicios de Biobancos y son una herramienta importante de estudio, especialmente en los campos de cáncer, inmunología y desarrollo de fármacos. Las muestras de FFPE bien conservadas permiten, incluso después de varios años, no solo el análisis de sus características morfológicas, sino también estudios genómicos y transcriptómicos, ya que son una fuente de ADN y ARN.

Brevemente, una vez obtenida la muestra se fija lo antes posible con el objetivo de que conserve al máximo sus características naturales. Para ello, las muestras se conservan en formaldehído, el cual penetra en el tejido y detiene los procesos celulares, conservando sus componentes estructurales y moleculares. Luego, la muestra se deshidrata con disoluciones con un grado creciente de etanol, lo que ayuda a una mayor penetración de la parafina, y se lava para eliminar la grasa del tejido. Por último, el tejido es embebido en parafina caliente, de unos 60°C. Cuando se solidifica, la parafina aportará un soporte estructural a la muestra, lo que facilitará su disección. [8]

Aunque este tipo de material es valioso, se caracteriza por comprometer la integridad de los ácidos nucleicos: por ejemplo, el proceso de fijación puede inducir sustituciones de bases nitrogenadas (principalmente C por T / G por A) y otras modificaciones químicas como la adición de grupos metilo que pueden promover dimerización de adeninas. [24, 25, 33, 34]

Aunque se ha descrito que de manera general existe una gran concordancia a nivel de perfiles transcriptómicos y epigenómicos entre muestras frescas y fijadas-parafinadas, es deseable conocer si existen cambios específicos derivados del proceso de conservación en parafina que habría que tener en cuenta en estudios de biomarcadores en glioblastoma que utilicen este material. [25, 33, 35, 36]

Antecedentes y objetivos

1. Antecedentes

El grupo de investigación liderado por el tutor ha llevado a cabo diversas investigaciones relacionadas con marcadores transcripcionales y epigenéticos en tumores de GBM y otros cánceres del SNC como astrocitomas u oligodendrogliomas. Entre los hallazgos más significativos se encuentran los siguientes:

- La detección de variantes de la histona H3, una de las proteínas estructurales de la cromatina, podría mejorar la clasificación de los gliomas en alto y bajo grado, ya que una mayor expresión de H3.3, la variante mayoritaria de histona H3 en cerebro adulto,

se relaciona con gliomas de bajo grado, mientras que la inducción de H3.1/H3.2 se produce en tumores más agresivos. [21]

- Las alteraciones genéticas y de expresión en los genes reguladores de la epigenética se producen principalmente en aquellos encargados de la acetilación de histonas, tanto en los que añaden (acetiltransferasas de lisinas o *KAT*) como los que eliminan dicha modificación (histonas desacetilasas o *HDAC*). Además, parece existir una firma transcripcional asociada a diferentes niveles de acetilación global de histona H3 que puede modularse mediante tratamiento farmacológico (inhibidores de *HDAC*) y que se asocia a buen pronóstico. [22]
- Más recientemente, se ha creado la base más importante de datos transcriptómicos y de patrones de metilación del ADN a partir de muestras en formato FFPE, aportando una cohorte independiente para validación de posibles biomarcadores. A diferencia de otras bases de datos públicas, esta tiene en cuenta las posibles degradaciones y alteraciones derivadas del proceso de fijación y parafinado, además de incluir información de tumores pobremente estudiados [23]

Los datos ómicos generados durante estos estudios son el punto de partida de este TFG.

2. Objetivo

Con este estudio se busca identificar qué cambios en la expresión de genes y patrones de metilación del ADN se ven más afectados en las muestras de FFPE con respecto a las muestras en fresco, es decir, si existen características intrínsecas del ADN o el ARN que sean más susceptibles al daño inducido por fijación y embebimiento y promuevan diferencias que habría que tener en cuenta a la hora de proponer nuevos biomarcadores de utilidad en la práctica clínica en el diagnóstico del GBM.

Para ello, compararemos los resultados de ambos tipos de muestras tanto de manera *in silico* como de manera experimental en el laboratorio. Analizaremos datos basados en la secuenciación masiva (NGS, *next generation sequencing*) y en hibridación de chips (plataforma MethylationEPIC v2.0), para a continuación validar los posibles resultados experimentalmente con muestras independientes.

Materiales y métodos

1. Sets de datos

Se ha trabajado con los siguientes sets de datos de GBM IDH nativo, originados de investigaciones previas del laboratorio de acogida y de libre acceso:

- GSE272042, RNA-seq mediante captura de exomas de 110 GBMs procedentes de tejido FFPE.
- GSE185861, RNA-seq de RNA total con depleción de RNA ribosómico de 20 GBMs procedentes de tejido fresco.
- RNA-seq de tejido fresco extraídas de la base de datos TCGA (<https://portal.gdc.cancer.gov/>)
- GSE274910, hibridación de 112 GBMs de tejido FFPE en beadchips Infinium MethylationEPIC v2.0.
- GSE147391, hibridación de 18 GBMs de tejido fresco en beadchips Infinium MethylationEPIC v2.0.

2. Recursos informáticos *online*

- Venny (<https://bioinfogp.cnb.csic.es/tools/venny/>): para crear **diagramas de Venn** que permiten comparar hasta **cuatro listas de elementos**.
- HGNC (HUGO Gene Nomenclature Committee, <https://www.genenames.org/>): se ha utilizado la herramienta **Multi-symbol Checker** para trabajar con la misma versión más actualizada de nomenclatura *Gene Symbol* en todas las listas de genes .
- Biomart (<https://www.ensembl.org/biomart/martview/>): es una herramienta web integrada en el portal **Ensembl**, diseñada para facilitar la extracción masiva de datos genómicos sin requerir conocimientos de programación, y que se utilizó para obtener la longitud y el contenido de GC de los genes seleccionados.
- DAVID Bioinformatics (<https://davidbioinformatics.nih.gov/>): este portal, ofrece un conjunto integral de herramientas bioinformáticas diseñadas para la anotación funcional y el análisis de enriquecimiento de listas génicas para facilitar la interpretación biológica de estudios de alto rendimiento, como *microarrays* o secuenciación masiva. En este

proyecto, usamos *Functional Annotation Tool*, que permite identificar procesos biológicos enriquecidos en una lista de genes.

- RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>): permite predecir la estructura secundaria de secuencias de ARN o ADN monocatenario a partir de su secuencia primaria, basado en el modelo de energía libre mínima (*Minimum Free Energy*, MFE) que asume que la molécula adoptará la conformación más termodinámicamente estable.
- Nucleic Acid Converter (https://skaminsky115.github.io/nac/DNA-mRNA-Protein_Converter.html): para la transcripción de secuencias de ADNc a ARN mensajero.

3. Programas

3.1 Excel

Para la visualización de los datos en tablas dinámicas.

3.2 RStudio

La mayoría del análisis, tratamiento y filtrado de los datos se hizo utilizando el programa de RStudio, versión 2024.12.0 + 467, entorno de desarrollo integrado (IDE) diseñado específicamente para el lenguaje de programación R, ampliamente utilizado en estadística, análisis de datos y ciencia de datos. Proporciona una interfaz gráfica intuitiva que facilita la escritura de código, la ejecución de scripts, la visualización de resultados y la generación de informes reproducibles.

Además del código base que incluye el análisis estadístico, se utilizaron los siguientes paquetes para realizar tareas específicas de forma más eficiente y especializada, y son fundamentales para el trabajo en estadística, ciencia de datos, bioinformática, visualización, aprendizaje automático, entre otros campos. Las principales librerías utilizadas fueron:

- I. **dplyr**: proporciona una gramática intuitiva para realizar operaciones como filtrar, seleccionar, ordenar, agrupar y resumir datos. Su sintaxis basada en funciones encadenadas con el operador %>% facilita la escritura de código legible y eficiente.
- II. **ggplot2**: permite la creación de gráficos estadísticos basada en la gramática de los gráficos (*Grammar of Graphics*), facilitando la visualización de resultados .

- III. **readxl**: permite la importación de datos desde archivos de Excel (.xls y .xlsx) directamente a R sin necesidad de depender de software externo como Java y **Openxlsx** que permite tanto leer como escribir archivos Excel, ideal para generar hojas automatizadas en Excel desde R sin requerir software adicional.
- IV. **limma**: facilita el análisis reproducible de datos biológicos de alto rendimiento y que utilizamos para el análisis de metilación diferencial mediante la creación de modelos lineales basados en métodos bayesianos. PCR cuantitativa (qPCR). [37]

4. Validación experimental

4.1 Muestras

Para la validación de resultados de expresión génica, llevamos a cabo ensayos de qPCR en muestras independientes de GBM (21 en FFPE y 15 en fresco). Su uso cuenta con autorización del Comité de Ética de Investigación con medicamentos (CEIm) del Hospital General Universitario Puerta del Mar (ref. PI2022-128) de acuerdo a la legislación regional, nacional y europea según la Carta de Derechos Fundamentales de la Unión Europea, la Convención Europea sobre Derechos Humanos y la Declaración de Helsinki del 2013, Ley 14/2007 de Investigación biomédica, Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales y Reglamento General de Protección de Datos y Reglamento (UE) 2016/679 del Parlamento europeo y del Consejo de 27 de abril de 2016 de Protección de Datos (RGPD).

4.2. Método

Se analizó la expresión de un total de 8 genes mediante cebadores (*primers*) específicos para *IRF2BPL*, *CD83*, *SLC17A7*, *ATXN7*, *HIST2H3D/H3C13*, *HBA1/2* y usando *TBP* y *GAPDH* como genes normalizados o *housekeeping*. Estos genes se seleccionaron en base a la disponibilidad de *primers* en el laboratorio.

Para llevar a cabo la qPCR se utilizó la *master mix* PowerUp SYBR Green (Applied Biosystems, Thermo Fisher, Madrid, España) y el termociclador QuantStudio 12K Flex (Thermo Fisher, Madrid, España), añadiendo en cada pocillo 10µL (5 µL *master mix* 2x + 1-2 µL ADNc + 2-3 µL H₂O + 1µL *primers* 5µM). El programa de qPCR fue el siguiente:

1º) 50 °C durante 2 minutos

2º) 95 °C durante 10 minutos

3º) 40 ciclos a 95 °C por 15s

4º) 60 °C durante 1 minuto

Los resultados se analizaron utilizando el método $2^{-\Delta\Delta C_t}$, usando cada normalizador por separado y promediando los valores de cambio (*fold change*).

Resultados

1. Identificación de los genes más diferencialmente expresados entre tumores frescos y FFPE

1.1 Filtrado de los datos transcriptómicos

Se partió de 2 sets de datos, uno procedente de tejido fresco y otro de tejido FFPE, con 30451 y 40233 genes, respectivamente. Se comprobó si la nomenclatura de los genes era correcta utilizando Multi-Symbol Checker de HGNC, con el fin de actualizar los nombres de genes para las dos listas de genes. Una vez actualizada la nomenclatura, se introdujeron ambas listas de genes en la página Venny, la cual devolvió un total de 26368 genes comunes. Se introdujo esta lista en Rstudio para poder realizar los cálculos y cambios necesarios. Adicionalmente, se observó que existían genes con una mediana de 0 para uno o ambos tejidos, por lo que también fueron eliminados.

1.2 Clasificación de genes

Dadas las diferencias en profundidad de lectura y preparación de las librerías de secuenciación de los dos sets de datos [21, 23] no fue posible aplicar métodos de expresión diferencial como “DESeq2”, por lo que se optó por elaborar un *ranking* con los genes comunes a los dos sets de datos, en el que se asignó a cada uno de los genes un valor ordinal para cada set de datos “Fresco” y “FFPE”, en función de su nivel de expresión en cada tipo de muestra, siendo 1 la máxima y 26368 la mínima. Una vez establecidos los *rankings* para ambas condiciones, se calculó la diferencia de posiciones entre ellos mediante la fórmula [Δ = Posición en Fresco – Posición en FFPE]. Este valor es una primera estimación del grado de variación en la expresión génica entre ambos tipos de tejido.

Posteriormente, se obtuvo el valor absoluto de dicha diferencia ($|\Delta|$) para cada gen y se examinaron los genes más expresados en cada condición, en este caso los 1000 genes con mayor $|\Delta|$: 514 genes con mayor expresión en “Fresco” [Tabla 1] y 486 genes con mayor expresión en “FFPE” [Tabla 2]. Dada la similitud en el número de genes, denominaremos a esta selección como Top500 de forma indistinta para las dos condiciones.

Posición	Gen	rankF	rankFFP	delta
Top 1	RPS18	37	12842	-12805
Top 2	HLA-C	89	12842	-12753
Top 3	RPS9	177	12842	-12665
Top 4	MT-TP	32	12242	-12210
Top 5	MTND2P28	195	12242	-12047
Top 6	FGD5-AS1	536	12242	-11706
Top 7	AIF1	1220	12842	-11622
Top 8	MAPT	1307	12842	-11535
Top 9	SNN	323	11718	-11395
Top 10	CUTA	370	11718	-11348
Top 11	IFI27L2	911	12242	-11331
Top 12	NELFE	988	12242	-11254
Top 13	TMEM47	1642	12842	-11200
Top 14	GABBR1	1044	12242	-11198
Top 15	GNG12	2080	12842	-10762
Top 16	NRGN	967	11718	-10751
Top 17	SH3GL2	1598	12242	-10644
Top 18	NAP1L5	1665	12242	-10577
Top 19	DDR1	179	10468	-10289
Top 20	CAMK2N2	2558	12842	-10284
Top 21	ASCL1	1479	11718	-10239
Top 22	NPTXR	1542	11718	-10176
Top 23	CTXN1	723	10849	-10126
Top 24	SPTSSA	1660	11718	-10058
Top 25	C4orf3	531	10468	-9937

Tabla 1. Representación de los 25 genes con mayor diferencia de expresión a favor del tejido “Fresco”.

Posición	Gen	rankF	rankFFP	delta
Top 1	H1-4	13066	243	12823
Top 2	NPIP3	12817	15	12802
Top 3	H2BC18	13089	292	12797
Top 4	SNORD15B	12771	17	12754
Top 5	GUSBP9	12995	283	12712
Top 6	MATR3	13076	396	12680
Top 7	MAGED4B	12934	261	12673
Top 8	H4C3	13142	549	12593
Top 9	SNORA53	12724	150	12574
Top 10	RPS10	12535	137	12398
Top 11	POLR2J3	12556	234	12322
Top 12	SNORA12	12563	266	12297
Top 13	GUSBP13	13089	803	12286
Top 14	EIF4A1	12535	283	12252
Top 15	SNORA22	12307	101	12206
Top 16	ZBTB20	12934	753	12181
Top 17	RASA4	12943	765	12178
Top 18	SERPINA3	12410	316	12094
Top 19	GABARAP	12242	232	12010
Top 20	NDUFA13	12160	167	11993
Top 21	PILRB	12731	815	11916
Top 22	SNORD17	11898	7	11891
Top 23	SNORD67	12707	892	11815
Top 24	PFDN5	12027	259	11768
Top 25	SNORA7B	12976	1210	11766

Tabla 2. Representación de los 25 genes con mayor diferencia de expresión a favor del tejido “FFPE”.

2. Los genes mejor expresados en tumores frescos tienden a ser más cortos y con mayor contenido en G/C que los genes expresados en tumores FFPE

Se obtuvieron tanto el contenido en G/C como la longitud de los genes para los Top500 de Fresco y FFPE para determinar si estas dos variables podrían estar relacionadas con las diferencias encontradas. Se añadió también una tercera lista que denominamos “Null”, en el que aparecían los 1000 genes con un Δ más cercano a 0, como set de genes más estables entre las dos condiciones y que podría servir de referencia para ambas listas. Las comparaciones arrojaron diferencias significativas para ambas variables, longitud y %GC, en los tres grupos [Fresco vs Null, FFPE vs Null y Fresco vs FFPE], como se puede observar en las tablas 3 y 4.

Archive	Statistic	p_value	Confidence_Interval_Lowe	Confidence_Interval_Uppe	Mean_Group	Mean_Group	stderr
FrescovsFFPE	-5.1624338	2.9435E-07	-3.500372485	-1.572183211	45.55212371	48.08840156	0.491295
FFPEvsNull	-3.1254131	0.00182673	-2.105592516	-0.4813448	45.55212371	46.84559237	0.4138553
FrescovsNull	2.85576192	0.00438365	0.388795015	2.096823365	48.08840156	46.84559237	0.4351936

Tabla 3. Resultados por pares del análisis estadístico “t de student” para el contenido en GC de los datasets (%GC).

Archive	Statistic	p_value	Confidence_Interval_Lowe	Confidence_Interval_Uppe	Mean_Group	Mean_Group	stderr
FrescovsFFPE	13.3723639	1.1579E-35	2954.655077	3972.05318	6416.738144	2953.384016	258.9934
FFPEvsNull	9.08842808	1.5929E-18	1850.683143	2871.094351	6416.738144	4055.849398	259.7687
FrescovsNull	-10.358297	2.8868E-24	-1311.25323	-893.6775336	2953.384016	4055.849398	106.4331

Tabla 4. Resultados por pares del análisis estadístico “t de student” para la longitud de los transcritos de los datasets.

Datasets	Statistic	p_value	Confidence_Interval_Lowe	Confidence_Interval_Uppe	Mean_Group	Mean_Group	Stderr
Fresco2 vs FFPE	8.91994003	1.8976E-18	3.192246825	4.992682222	49.64458824	45.55212371	0.4573256
Fresco2 vs Null	7.030086	3.2175E-12	2.017969257	3.580022474	49.64458824	46.84559237	0.3981453
FFPE vs Null	-3.1254131	0.00182673	-2.105592516	-0.4813448	45.55212371	46.84559237	0.4122206

Tabla 5. Resultados por pares del análisis estadístico “t de student” para el contenido en GC de los datasets (%GC).

Datasets	Statistic	p_value	Confidence_Interval_Lowe	Confidence_Interval_Uppe	Mean_Group	Mean_Group	Stderr
Fresco2 vs FFPE	-10.858919	1.4036E-25	-3575.638763	-2480.705173	3388.566176	6416.738144	275.7398
Fresco2 vs Null	-4.4971797	7.5787E-06	-958.4034644	-376.1629779	3388.566176	4055.849398	148.3782
FFPE vs Null	9.08842808	1.5929E-18	1850.683143	2871.094351	6416.738144	4055.849398	256.4109

Tabla 6. Resultados por pares del análisis estadístico “t de student” para la longitud de los transcritos de los datasets.

Además, se observó que los genes de la lista “Fresco” tenían un mayor contenido en G/C y son más cortos. Para confirmar estos resultados, se hizo una nueva comparación con otra cohorte de tejido en fresco, al que se le denominó “Fresco 2”, derivada de la base de datos TCGA. Los resultados fueron similares a los anteriores [Tabla 5-6], incluyendo los valores promedio para las dos listas de “Frescos”. Para una mejor visualización se han representado los resultados en gráficos de tipo cajas y bigotes [Figura 1].

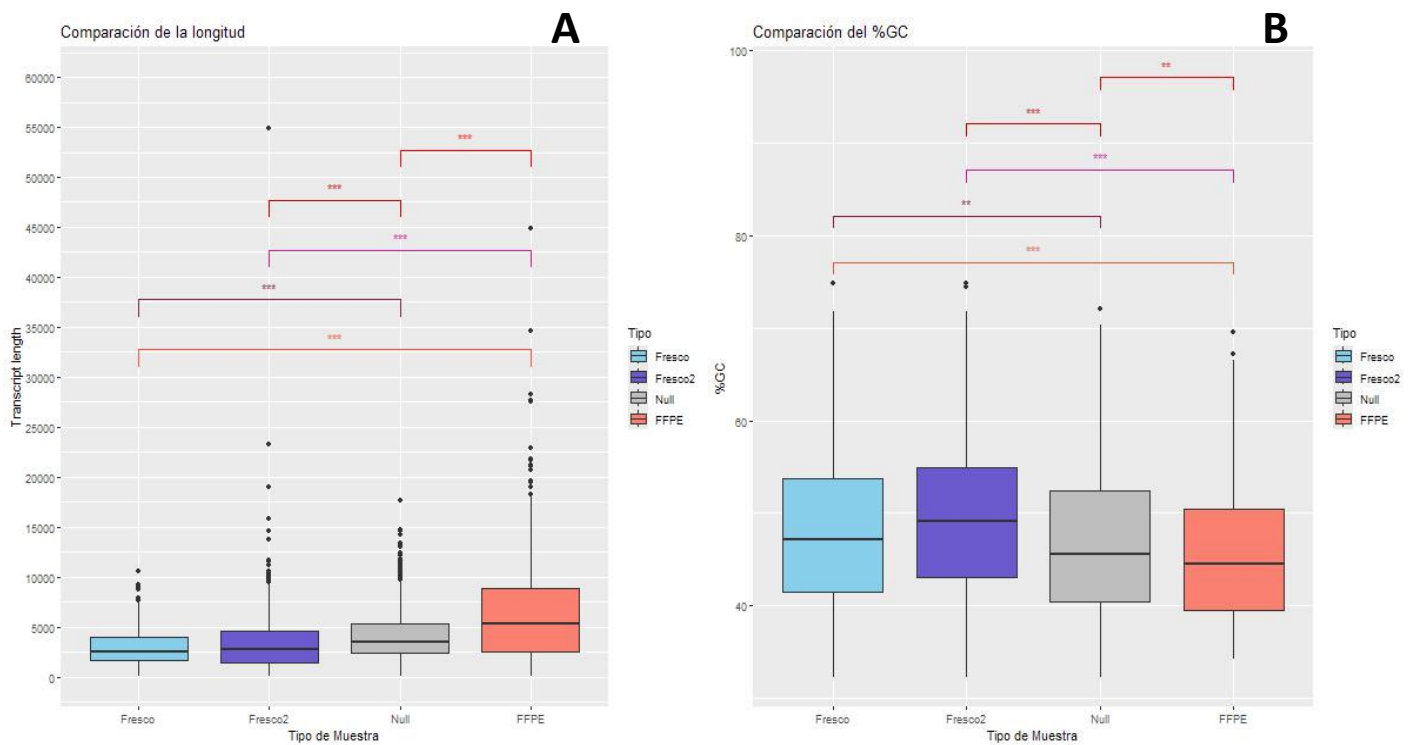


Figura 1. Representación mediante diagramas de caja y bigotes (“boxplots”) de las cuatro categorías analizadas en el estudio estadístico. **(A)** Comparación longitud **(B)** Comparación del contenido GC.

3. Los genes mejor expresados en tejido fresco tienden a tener un mayor valor de MFE (Minimum Free Energy) que los genes en tejido FFPE.

El MFE es un cálculo de estabilidad del ARNm, en el cual cuanto menor mínimo de energía libre tenga un transcrito, más estable será. Esta variable también podría explicar las diferencias encontradas entre muestras en fresco y FFPE. Se seleccionaron genes pareados con la misma longitud de cada lista (“Fresco” y “FFPE”) para poder realizar una comparación más precisa de los MFE sin la variable de longitud ya que los transcritos son generalmente más estables a mayor longitud; finalmente seleccionamos 20 parejas y 2 tríos de genes que cubrían todo el espectro de longitudes, desde 1216 hasta 9540 nt. A continuación, se calculó el MFE de estos 46 genes usando la herramienta RNAfold. De las 22 comparaciones, 13 de ellas tenían un menor valor de MFE en “Fresco”, implicando que, al ser más negativos, son más estables. En la gráfica de dispersión [Figura 2a], se aprecia mejor que a bajas longitudes de ARNm los valores de ambos tipos de genes son similares y estos no comienzan a diferenciarse claramente hasta que se alcanzan los 3750 nt de longitud, aproximadamente. Esto se ve mejor reflejado al añadir las líneas de regresión, que se solapan a longitudes bajas, pero según

aumenta la extensión de los transcritos estas se van separando [Figura 2a]. Este comportamiento diferencial según la longitud del ARNm hace que la comparación global no salga significativa, aunque en la gráfica de cajas y bigotes se puede observar cierta tendencia de los genes de “Fresco” a tener valores más negativos de MFE [Figura 2b].

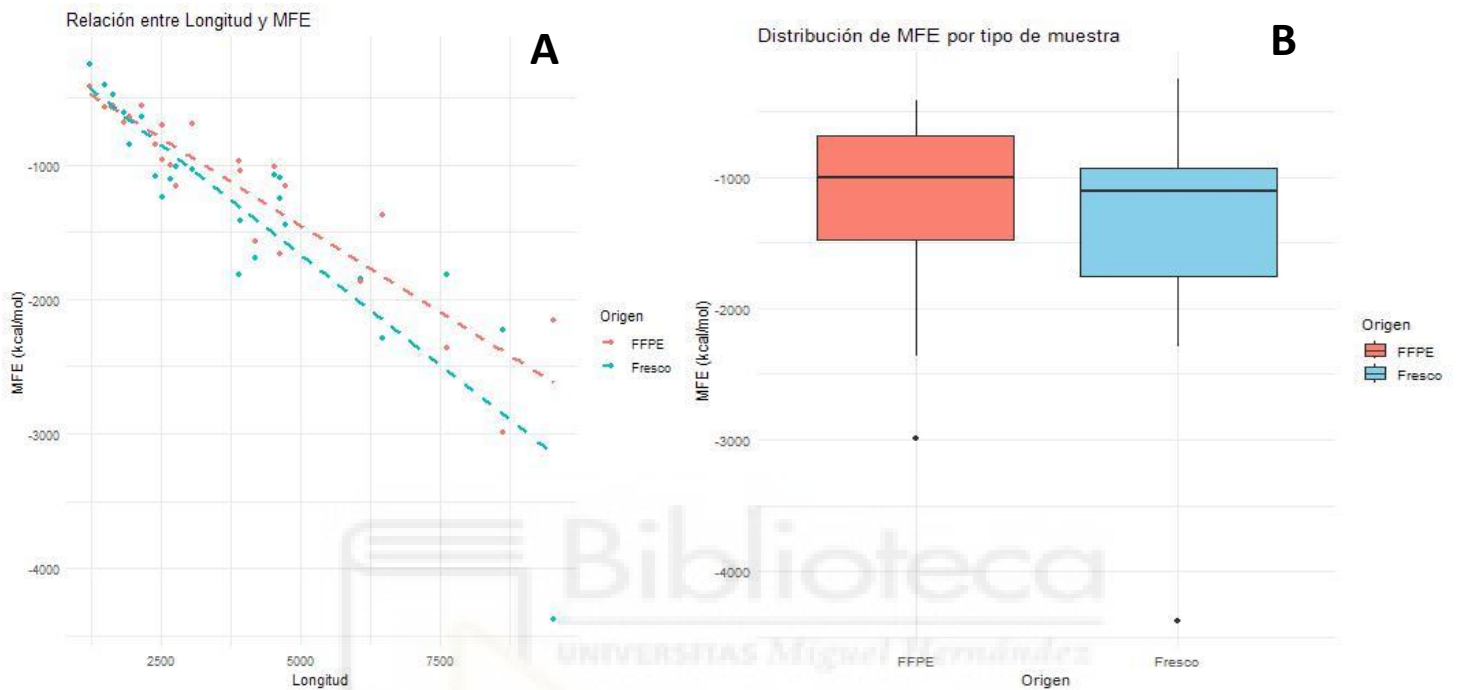


Figura 2. A) Gráficos de dispersión de los MFE de los genes de “FFPE” (rojo) y “Fresco” (azul) en función de la longitud de sus transcritos. **B)** Representación mediante “cajas y bigotes”, en el que se compara los valores de MFE de los genes de ambos tipos de muestras.

Para demostrar si la hipótesis de que los transcritos en “Fresco” son más estables a mayores longitudes que los de “FFPE”, se hizo un segundo estudio con otros 10 genes de cada set de datos con longitudes > 5000 nt. Al representar los valores obtenidos nuevamente en la gráfica de dispersión [Figura 3a], se puede apreciar cierta tendencia de los genes procedentes de tejido a fresco a tener menores MFE, aunque no obtuvimos ninguna diferencia significativa de manera global (p-valor = 0.65, Figura 3b). lo que sugiere que necesitamos un mayor número de genes en el análisis.

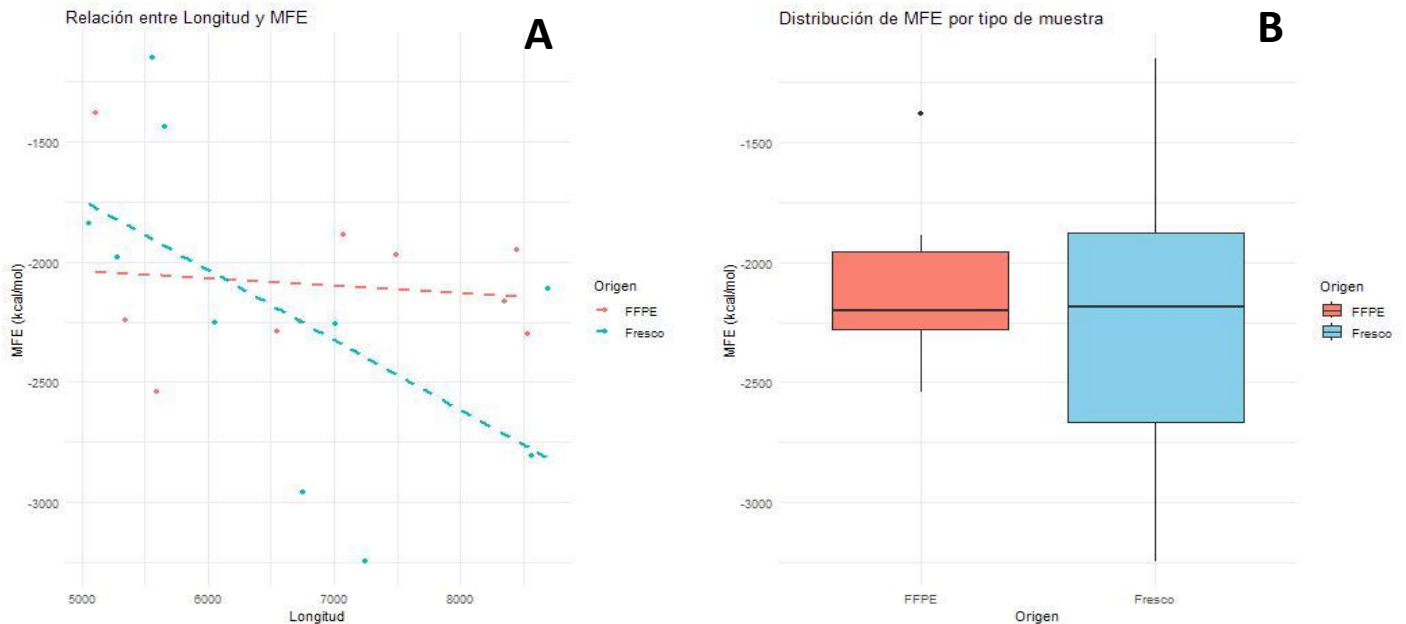


Figura 3. A) Gráficos de dispersión de los MFE de los genes de “FFPE” (rojo) y “Fresco” (azul) en función de la longitud de sus transcritos (>5000). **B)** Representación mediante “cajas y bigotes”, en el que se compara los valores de MFE de los genes de ambos tipos de muestras.

Para ilustrar mejor esta tendencia, la Figura 4 muestra imágenes de las representaciones tridimensionales de dos parejas de mRNAs de igual longitud pero con MFE muy diferentes. Se puede apreciar como aquellos ARNm con valores de MFE más negativos y, por tanto más estables, tienden a tener estructuras más lineales, con menos ramificaciones, mientras que aquellos con valores más altos presentan organizaciones más complejas [Figura 4].

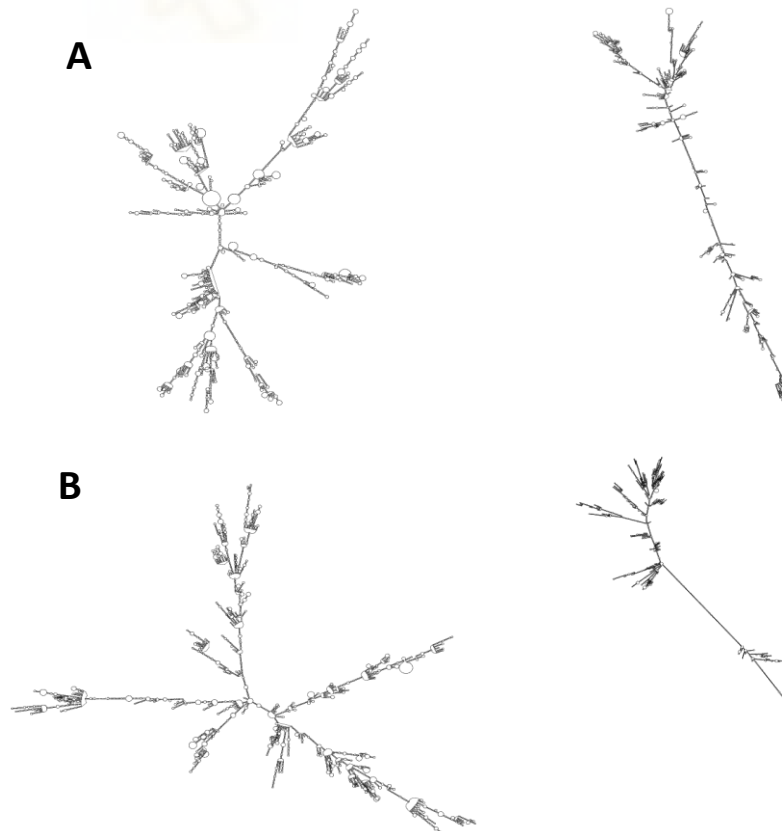


Figura 4. Representaciones de las estructuras secundarias de los ARNm. **A)** ARNm DNAJB14 (izquierda) procedente de “FFPE” y ARNm NPIP5 (derecha) procedente de “Fresco”, ambos con una longitud de 6463nt **B)** ARNm UBA6 (izquierda) procedente de “FFPE” y ARNm FEM1A (derecha) procedente de “Fresco”, ambos con una longitud de 9540nt.

4. Los genes más representados en tejido fresco y parafinado están asociados a diferentes funciones biológicas.

Se realizó un análisis de las funciones más representadas en las listas de genes *Top500*, utilizando la herramienta de anotación funcional de DAVID bioinformatics basada en la terminología de *Gene Ontology* (GO) [Tabla 7]. Se obtuvo que los genes del *Top500* de “Fresco” estaban principalmente involucrados en 5 categorías de procesos biológicos: (i) *Procesos mitocondriales*; (ii) *Procesos sinápticos*; (iii) *Neurodesarrollo*; (iv) *Metabolismo* y (v) *Regulación de procesos celulares*, en los que destacan el procesamiento y presentación de antígenos y la transcripción mediante la ARN polimerasa II. Las categorías con más representantes son **Procesos sinápticos** y **Regulación de procesos celulares** con 6 términos GO, seguido de **Neurodesarrollo** con 4 tipos de genes. Sin embargo, los términos GO con mayor “Fold Enrichment”, que indica el grado de sobre-representación de nuestros genes en el término GO correspondiente, se relaciona con el complejo II de la cadena respiratoria mitocondrial, seguido por genes implicados en la inhibición de la desacetilación de tubulina.

Term	PValue	Fold Enrichment
GO:0048167~regulation of synaptic plasticity	0.000589	5.516
GO:0016079~synaptic vesicle exocytosis	0.001365	10.029
GO:0046928~regulation of neurotransmitter secretion	0.002237	8.825
GO:0031630~regulation of synaptic vesicle fusion to presynaptic active zone membrane	0.009935	18.912
GO:0050805~negative regulation of synaptic transmission	0.016529	14.709
GO:0097091~synaptic vesicle clustering	0.028979	11.032
GO:1904428~negative regulation of tubulin deacetylation	0.007203	22.064
GO:0030308~negative regulation of cell growth	0.019330	2.942
GO:0050848~regulation of calcium-mediated signaling	0.044100	8.825
GO:0032482~Rab protein signal transduction	0.049662	8.274
GO:0006366~transcription by RNA polymerase II	0.021323	2.206
GO:0019882~antigen processing and presentation	0.021397	4.694
GO:0022010~central nervous system myelination	0.004266	11.767
GO:0061564~axon development	0.009443	5.963
GO:0007423~sensory organ development	0.018255	7.060
GO:1990416~cellular response to brain-derived neurotrophic factor stimulus	0.020355	13.238
GO:0006044~N-acetylglucosamine metabolic process	0.005163	11.032
GO:0006790~sulfur compound metabolic process	0.005163	11.032
GO:0018146~keratan sulfate biosynthetic process	0.049662	8.274
GO:0032543~mitochondrial translation	0.000277	4.694
GO:0034553~mitochondrial respiratory chain complex II assembly	0.004875	26.476

Procesos Mitocondriales
Procesos Sinápticos
Neurodesarrollo
Metabolismo
Regulación de Procesos Celulares

Tabla 7. Representación de los procesos biológicos más enriquecidos entre los genes del *Top500* correspondientes al tejido “Fresco”. Los procesos se agrupan en seis categorías funcionales principales, cada una representada con un color distinto para facilitar su visualización.

Por su parte, los genes del *Top500* de “FFPE” [Tabla 8] se dividen en 11 categorías: (i) *Regulación de la diferenciación y proliferación celular*; (ii) *Regulación de la transcripción y procesamiento del ARN*; (iii) *Remodelación y organización de la cromatina*; (iv) *Reparación y mantenimiento del ADN*; (v) *Desarrollo y diferenciación celular*; (vi) *Señalización celular*; (vii) *Respuesta celular y mantenimiento*; (viii) *Otros procesos biológicos*; (ix) *Recombinación y división celular*; (x) *Degradación de proteínas* y (xi) *Procesos relacionados con los telómeros*. En este caso las categorías con más representantes son ***Regulación de la transcripción y procesamiento del ARN*** y ***Desarrollo y diferenciación celular*** con 9 términos GO, seguidos de cerca por ***Reparación y mantenimiento del ADN*** con 8 términos, mientras que los términos GO con mayor enriquecimiento (“Fold enrichment”) son los relacionados con la telomerasa.

Por tanto, obtenemos enriquecimientos de genes relacionados con procesos nerviosos, neurodesarrollo, integridad y correcto funcionamiento tanto del ARN como el ADN, entre otras funciones, asociados a diferentes longitudes, contenido GC y estabilidad de mRNA.



Term	PValue	Fold Enrichment
GO:0007131~reciprocal meiotic recombination	0.005206	7.037
GO:0044818~mitotic G2/M transition checkpoint	0.013243	7.933
GO:0097695~establishment of protein-containing complex localization to telomere	0.045214	43.631
GO:1904884~positive regulation of telomerase catalytic core complex assembly	0.045214	43.631
GO:0035360~positive regulation of peroxisome proliferator activated receptor signaling pathway	0.020793	13.089
GO:0006511~ubiquitin-dependent protein catabolic process	0.045124	2.034
GO:0032436~positive regulation of proteasomal ubiquitin-dependent protein catabolic process	0.043434	3.117
GO:0018105~peptidyl-serine phosphorylation	0.010011	3.818
GO:0051056~regulation of small GTPase mediated signal transduction	0.001567	3.698
GO:0007264~small GTPase-mediated signal transduction	0.006680	3.245
GO:0010804~negative regulation of tumor necrosis factor-mediated signaling pathway	0.007272	6.416
GO:0007224~smoothened signaling pathway	0.015525	3.471
GO:0035023~regulation of Rho protein signal transduction	0.033254	5.630
GO:0007265~Ras protein signal transduction	0.043434	3.117
GO:0006355~regulation of DNA-templated transcription	0.032836	1.459
GO:0006396~RNA processing	0.000003	2.517
GO:0045893~positive regulation of DNA-templated transcription	0.002270	1.826
GO:0006366~transcription by RNA polymerase II	0.003676	2.545
GO:0045944~positive regulation of transcription by RNA polymerase II	0.004655	1.541
GO:0000122~negative regulation of transcription by RNA polymerase II	0.033330	1.457
GO:0097694~establishment of RNA localization to telomere	0.045214	43.631
GO:0006357~regulation of transcription by RNA polymerase II	0.047266	1.293
GO:0010629~negative regulation of gene expression	0.006053	2.223
GO:0042776~proton motive force-driven mitochondrial ATP synthesis	0.016437	4.028
GO:0046854~phosphatidylinositol phosphate biosynthetic process	0.030848	4.195
GO:1903566~positive regulation of protein localization to cilium	0.007364	21.816
GO:0071456~cellular response to hypoxia	0.005433	3.073
GO:0071277~cellular response to calcium ion	0.017177	3.394
GO:0160063~multi-pass transmembrane protein insertion into ER membrane	0.020793	13.089
GO:1900034~regulation of cellular response to heat	0.045012	8.726
GO:0072660~maintenance of protein location in plasma membrane	0.045214	43.631
GO:0010839~negative regulation of keratinocyte proliferation	0.016830	7.272
GO:0045668~negative regulation of osteoblast differentiation	0.021975	3.740
GO:0043616~keratinocyte proliferation	0.023131	6.464
GO:0051149~positive regulation of muscle cell differentiation	0.039597	9.350
GO:0045931~positive regulation of mitotic cell cycle	0.042120	5.133
GO:0045599~negative regulation of fat cell differentiation	0.045868	3.698
GO:0007399~nervous system development	0.003260	2.064
GO:0042692~muscle cell differentiation	0.007507	9.696
GO:0031175~neuron projection development	0.009258	3.068
GO:0003170~heart valve development	0.020793	13.089
GO:0007420~brain development	0.022401	2.296
GO:0003140~determination of left/right asymmetry in lateral mesoderm	0.025034	11.899
GO:0009791~post-embryonic development	0.025816	3.586
GO:0061061~muscle structure development	0.029593	10.908
GO:0003281~ventricular septum development	0.042120	5.133
GO:0006281~DNA repair	0.002917	2.407
GO:0032508~DNA duplex unwinding	0.008360	3.966
GO:0036297~interstrand cross-link repair	0.010777	5.741
GO:0000723~telomere maintenance	0.011916	4.363
GO:0000722~telomere maintenance via recombination	0.025034	11.899
GO:0006974~DNA damage response	0.028000	1.958
GO:0000724~double-strand break repair via homologous recombination	0.033524	2.909
GO:0006287~base-excision repair, gap-filling	0.045012	8.726
GO:0006338~chromatin remodeling	0.000001	2.189
GO:0140588~chromatin looping	0.000475	3.396
GO:0006325~chromatin organization	0.009116	2.373
GO:0006334~nucleosome assembly	0.049281	2.407

Regulación de la Diferenciación y Proliferación Celular
Regulación de la Transcripción y Procesamiento de RNA
Remodelación y Organización de la Cromatina
Reparación y Mantenimiento del DNA
Desarrollo y Diferenciación Celular
Señalización Celular
Respuesta Celular y Mantenimiento
Otros Procesos Biológicos
Recombinación y división celular
Degradación de proteínas
Procesos relacionados con los telómeros

Tabla 8. Representación de los procesos biológicos más enriquecidos entre los genes del Top500 correspondientes al tejido “FFPE”. Los procesos se agrupan en once categorías funcionales principales, cada una representada con un color distinto para facilitar su visualización

5. Validación de las diferencias de abundancia de transcritos entre tejido fresco y parafinado mediante qPCR.

Dado que el análisis de los datos de RNA-seq reveló una tendencia diferencial en la expresión de ciertos genes entre tejido fresco y tejido parafinado, se propuso validar estos

resultados mediante un método independiente, la cuantificación por PCR a tiempo real (qPCR). A diferencia del RNA-seq en el que se utilizaron dos estrategias diferentes de secuenciación masiva para “Fresco” y “FFPE”, en este caso empleamos muestras tratadas con DNasa y retrotranscritas a ADN complementario (ADNc) siguiendo los mismos procedimientos (TURBO DNA-free kit y RevertAid First-DNA Synthesis kit, ambos de Thermo Fisher) [21]. Para ello, se seleccionaron tres genes con sobreexpresión en muestras “Fresco” y tres genes con sobreexpresión en “FFPE”, los cuales fueron analizados en un total de 36 muestras, incluyendo ambos tipos de tejido. Además, se incluyeron también dos genes de control como normalizadores. No obstante, descartamos el gen *HIST2H3D* (*H3C13*) del análisis final debido a la baja calidad de sus resultados por su baja expresión en todas las muestras. Tras las reacciones de qPCR se realizaron los cálculos para obtener la magnitud de cambio (FC, *Fold Change*) a partir del CT (*cycle threshold*) de cada muestra. El valor de FC indica cuánto cambia la expresión de un gen entre ambas condiciones.

De esta manera, confirmamos que los genes *IRF2BPL*, *CD83* y *SLC17A7* no muestran prácticamente expresión en tejido parafinado, mientras que *HBA1/2* tiende a expresarse mucho más en “FFPE” que en “Fresco”. Por su parte, *ATXN7* muestra una tendencia para expresarse más en tejido fresco en los ensayos de qPCR, al contrario del comportamiento en los análisis de RNA-seq [Figura 5]. A pesar de esta discrepancia, parece que de manera general validamos que las diferencias de expresión entre tejido fresco y parafinado se deben a propiedades intrínsecas de los ARNm y no de las diferencias en la preparación de librerías y alineamiento de las lecturas en el genoma de referencia.

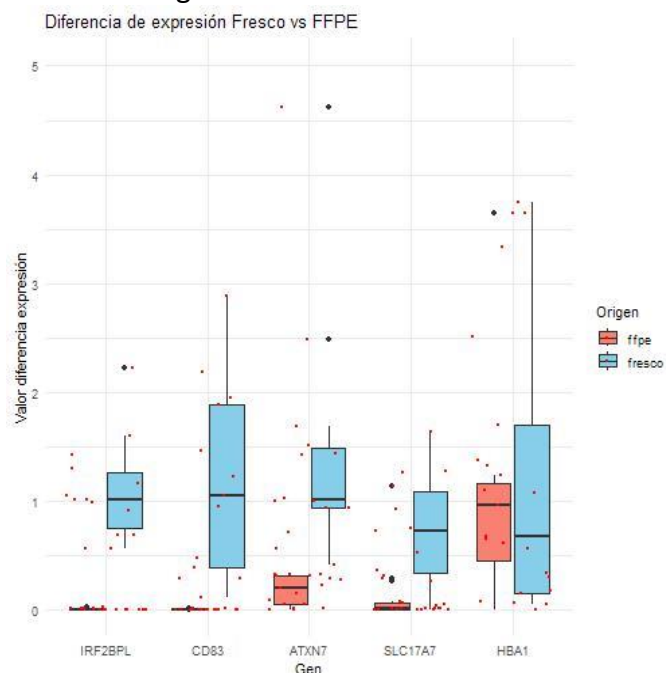


Figura 5. Representación en cajas y bigotes de la expresión diferencial entre tejidos para cada gen. Los puntos rojos indican los valores obtenidos

6. Los perfiles de metilación entre ambos tipos de tejidos presentan diferencias en su posición genómica.

A continuación, se compararon los perfiles de metilación de ambos tejidos con el fin de observar si había alguna tendencia significativa en los mismos que pudiera influir durante el proceso de parafinado. Cada dinucleótido CpG representado en el chip posee un valor beta entre 0 y 1, siendo 0 completamente no metilado en todas las células y 1 completamente metilado en todas las células. Un primer análisis mostró que no existía ningún gen con valor de 0 o 1, aunque sí varios cercanos a los mismos, lo que significa que en mayor o menor proporción coexisten células metiladas y no metiladas en los CpG examinados. Además, se llevó a cabo un análisis estadístico de t de student entre ambos grupos; sin embargo, ambos sets de datos poseen la misma media y no se obtuvieron diferencias significativas [Figura 6].

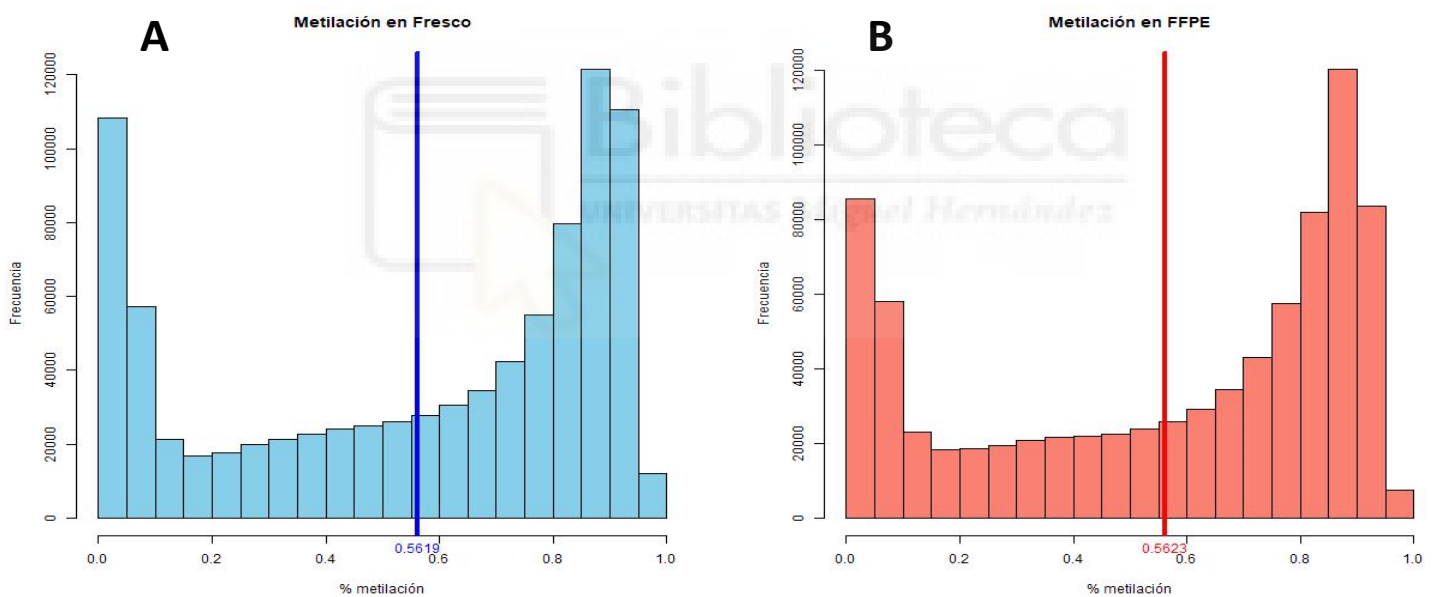


Figura 6. Histogramas con la distribución de los perfiles de metilación y la media de cada grupo (línea vertical). El eje x representa el porcentaje de metilación y el eje y indica la frecuencia absoluta de observaciones **(A)** Perfil de metilación en Fresco, con media de 0.5619; **(B)** Perfil de metilación en FFPE, con media de 0.5623.

Se calculó el Δ entre los valores de “Fresco” y “FFPE” para cada CpG [Δ = Valores Fresco – Valores FFPE], de la misma manera que se hizo con los datos de RNA-seq. En este caso, y dado el rango tan estrecho de valores (entre 0 y 1) se obtuvieron diferencias muy pequeñas. Por ello, se seleccionaron aquellos con mayor valor absoluto de delta ($|\Delta| > 0,3$). Con estos

filtros se obtuvieron finalmente 22 CpG con valor negativo (mayor metilación en “FFPE”) [Tabla 9] y 52 genes con valor positivo (mayor metilación en “Fresco”) [Tabla 10].

Posición	Gen	Fresco	FFPE	Delta
Top 1	cg08394412	0.1789	0.5890	-0.4101
Top 2	cg16896868	0.3443	0.7300	-0.3857
Top 3	cg01184784	0.0658	0.4310	-0.3652
Top 4	cg25336332	0.2251	0.5898	-0.3646
Top 5	cg05310764	0.1295	0.4831	-0.3536
Top 6	cg08986950	0.0782	0.4308	-0.3526
Top 7	cg09556826	0.4016	0.7449	-0.3433
Top 8	cg03090734	0.3747	0.7167	-0.3420
Top 9	cg14301932	0.0701	0.4112	-0.3411
Top 10	cg04927931	0.2840	0.6081	-0.3241
Top 11	cg25698741	0.3053	0.6280	-0.3226
Top 12	cg13265740	0.3325	0.6541	-0.3216
Top 13	cg01156747	0.4766	0.7981	-0.3216
Top 14	cg21063722	0.1752	0.4929	-0.3177
Top 15	cg25476565	0.3109	0.6285	-0.3176
Top 16	cg15082166	0.5423	0.8566	-0.3143
Top 17	cg21082028	0.2993	0.6130	-0.3136
Top 18	cg16327926	0.2533	0.5665	-0.3132
Top 19	cg02743713	0.1430	0.4534	-0.3104
Top 20	cg00838919	0.1394	0.4485	-0.3091
Top 21	cg22876812	0.2485	0.5490	-0.3005
Top 22	cg07228824	0.2017	0.5019	-0.3002

Tabla 9. Representación de los 22 genes con mayor diferencia de metilación en “FFPE”.

Posición	Gen	Fresco	FFPE	Delta
Top 1	cg27292870	0.9256	0.0530	0.8726
Top 2	cg14315558	0.5054	0.0934	0.4120
Top 3	cg18932798	0.5503	0.1435	0.4068
Top 4	cg00154357	0.5840	0.1776	0.4064
Top 5	cg26322231	0.5199	0.1256	0.3943
Top 6	cg25096149	0.7486	0.3660	0.3826
Top 7	cg09352518	0.5137	0.1416	0.3721
Top 8	cg05299094	0.5338	0.1631	0.3707
Top 9	cg13421331	0.7059	0.3453	0.3606
Top 10	cg07959030	0.5905	0.2301	0.3604
Top 11	cg08323960	0.7079	0.3501	0.3577
Top 12	cg05911086	0.7155	0.3627	0.3528
Top 13	cg06324373	0.5430	0.1912	0.3518
Top 14	cg10836173	0.4280	0.0800	0.3480
Top 15	cg01759384	0.4431	0.0955	0.3476
Top 16	cg04683289	0.7488	0.4019	0.3470
Top 17	cg06555487	0.7003	0.3565	0.3438
Top 18	cg25899817	0.5769	0.2333	0.3436
Top 19	cg16822939	0.4219	0.0784	0.3435
Top 20	cg09426679	0.6142	0.2759	0.3383
Top 21	cg21920270	0.5327	0.1948	0.3378
Top 22	cg14478058	0.6740	0.3379	0.3361
Top 23	cg22408421	0.6919	0.3559	0.3360
Top 24	cg08876558	0.7774	0.4422	0.3352
Top 25	cg19915738	0.3694	0.0391	0.3303

Tabla 10. Representación del top 25 genes con mayor diferencia de metilación en “Fresco”.

A continuación, se realizó un análisis con Limma para obtener un análisis estadístico fiable, gracias a que la tecnología empleada para generar estos datos fue la misma. Este análisis permitió identificar un total de 12.846 CpGs significativamente más metilados en “Fresco”, frente a solo 1.103 en “FFPE”, lo que indica que aproximadamente el 92 % de los resultados significativos provienen de metilaciones detectadas en tejido fresco pero no en FFPE. Se compararon estos resultados con los obtenidos mediante el análisis basado en Δ , observándose que el 51 % de los CpGs identificados por este último método (38 de 74) también estaban presentes entre los resultados significativos de Limma.

Se analizó, además, la distribución de los CpGs significativos por cromosoma y en relación con su localización respecto a las islas CpG. Para el análisis cromosómico, se contabilizó el número de CpGs significativos en cada cromosoma y su porcentaje respecto al total, tanto en “Fresco” como en “FFPE”. Estos datos se representaron mediante una gráfica de barras [Figura 7], en la que se observa que ambas condiciones presentan una distribución

cromosómica globalmente similar, aunque se observaron diferencias puntuales, como un relativo mayor número de CpGs en tejido fresco en los cromosomas 3 y 4, y mayor número de CpG en tejido FFPE en los cromosomas 17 y 19.

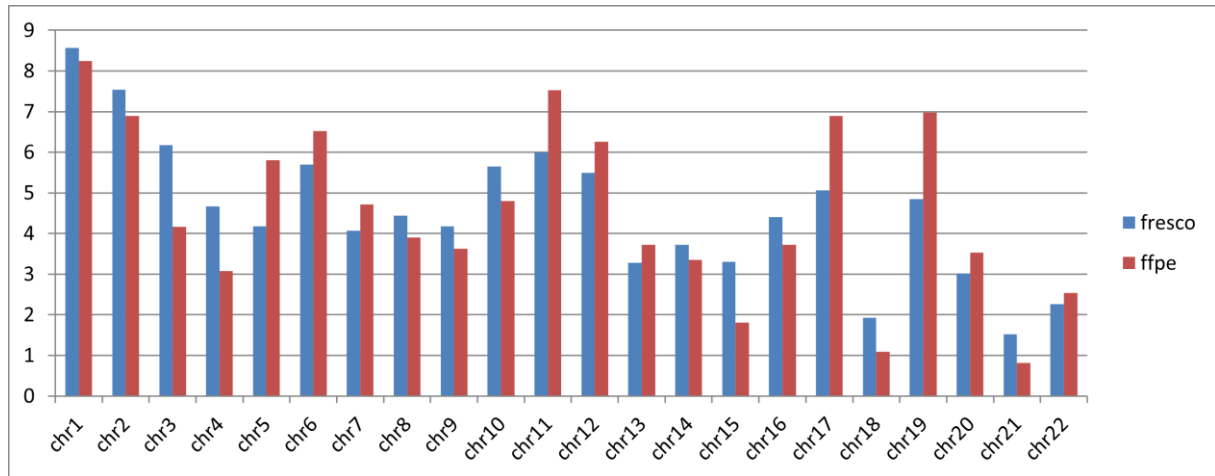


Figura 7. Gráfica de barras comparando la distribución de los CpGs por cromosoma y por condición.

También se calculó la proporción de CpGs significativos en función de su localización relativa a las islas CpG. Las categorías consideradas fueron:

- **Island:** Dentro de la isla de CpGs
- **Shore:** Adyacentes a islas CpGs, hasta 2kb de la isla CpG
- **Shelf:** De 2 a 4kb de la isla CpG
- **OpenSea:** Regiones alejadas de las islas CpGs (a más de 4kb)

Los resultados se representaron mediante una gráfica de barras [Figura 8], que compara la proporción de cada categoría entre las cohortes “Fresco” y “FFPE”. Además, se incluyeron diagramas circulares (“pie charts”) individuales [Figura 9] para visualizar la distribución relativa de cada categoría dentro de cada condición por separado. En ellas se observa como en tejido fresco la mayoría de elementos metilados no se encuentra en islas CpG, sino en otras regiones del genoma. Estas diferencias podrían tener repercusiones en la regulación de la expresión génica.

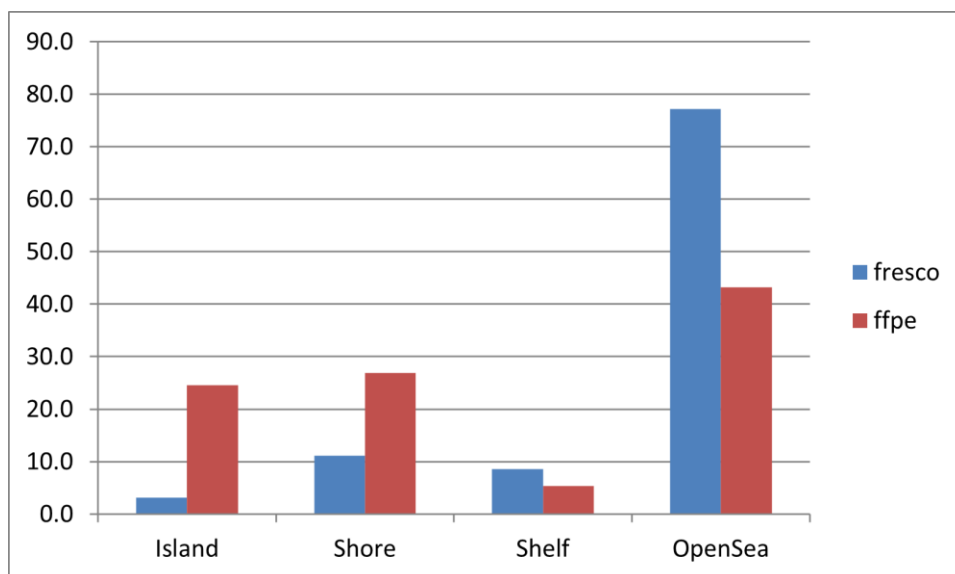


Figura 8. Gráfica de barras comparando las variables de la posición relativa a las islas de CpGs por condición.

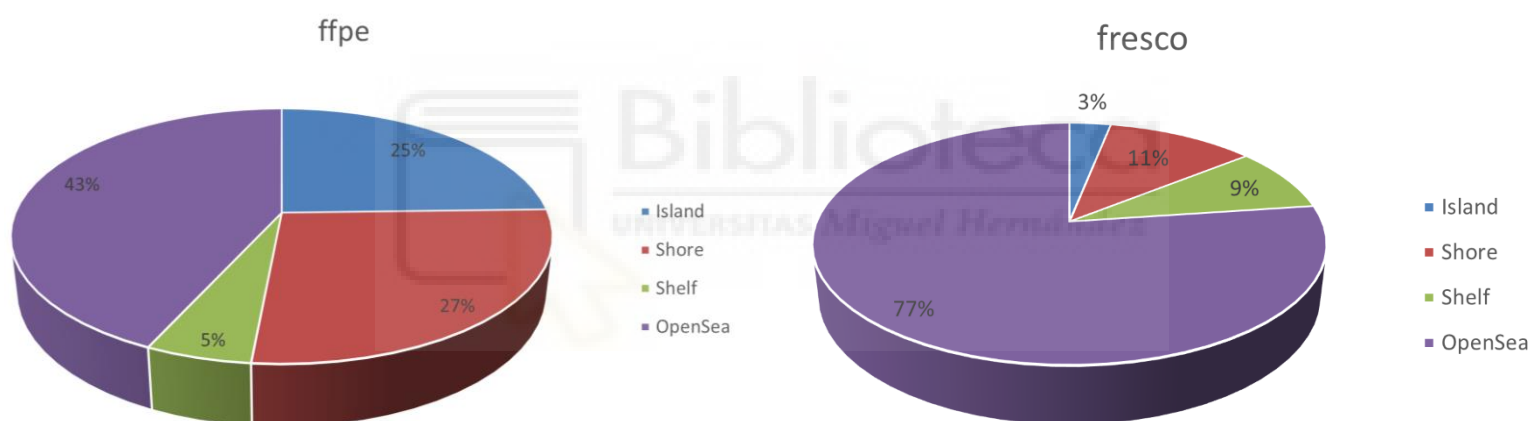


Figura 9. Gráficas “pie chart” con la distribución de cada variables de la posición relativa a las islas de CpGs en cada condición.

Discusión

Debido al carácter agresivo de los gliomas, en concreto de los GBM, es necesario entender en profundidad de qué manera afecta la fijación y parafinado en las muestras tumorales, como método de conservación a largo plazo más extendido en hospitales. Por ello, es relevante desde el punto de vista clínico y de la investigación oncológica estudiar las características intrínsecas del ADN y ARNm que pueden llevar a cambios tras el proceso de

conservación y que puede tener un impacto importante en el uso de biomarcadores para estas patologías.

En primer lugar, utilizamos dos sets de datos de RNA-seq de acceso libre, originados por el propio laboratorio en momentos y cohortes diferentes. Aunque lo deseable hubiera sido procesar en paralelo los mismos tumores en fresco y parafinado usando los mismos protocolos, la comparación de estos sets se puede considerar una primera aproximación para investigar estas características intrínsecas en los ARNm afectados por la fijación-parafinado. Esto nos ha llevado a realizar una comparación de *ránking* de expresión en lugar de crear una matriz de datos común, ya que la normalización de diferentes experimentos plantea riesgos de introducir artefactos cuando las condiciones experimentales son tan diferentes (diferentes estrategias de preparación de librerías, profundidad de lectura, etc).

Tras obtener una lista de genes con las diferencias más notables, comparamos tres características: longitud, contenido GC y estabilidad de los ARNm. Durante la evaluación de los transcritos con mayor expresión diferencial en “FFPE” y en “Fresco”, observamos que aquellos que aparecían en “FFPE” poseían una extensión significativamente mayor y un contenido más bajo de GC en comparación con aquellos procedentes de tejido fresco. Como confirmación de estos resultados, se introdujo un tercer grupo de genes, para los que no se encontraron diferencias entre ambas condiciones y que mostraron un comportamiento intermedio. Estos hallazgos son consistentes con lo descrito en la literatura, donde se ha señalado que los transcritos más cortos, con mayor contenido en GC o de baja abundancia son más susceptibles a degradación o alteración durante el proceso de fijación y embebido en parafina [24, 25]. Por tanto, nuestros resultados respaldan esta evidencia previa. En estos trabajos se plantea como posibles causas:

- a) La mayor representatividad de los extremos 3' en muestras frescas, más crítica en ARNm largos, siendo más consistente la cobertura a lo largo de todo el gen en muestras FFPE. Sin embargo, no se ha utilizado una estrategia de captura de oligo-dT en nuestro set de datos “Fresco” que justifique esta desviación.
- b) El mapeo incorrecto en intrones que aumenta el % de GC, un artefacto típico de librerías de ARN total que no se da en nuestro caso al construirse mediante captura de exomas.

También analizamos la energía libre mínima (MFE) y, aunque los resultados no alcanzaron significación estadística, se observó una tendencia a que los ARNm procedentes de tejido fresco presenten valores de MFE más negativos, lo que sugiere una mayor estabilidad estructural en comparación con los transcritos predominantes en muestras FFPE. Sin embargo, la literatura disponible sobre la relación entre MFE y la degradación del ARN en tejidos parafinados es limitada. En el estudio consultado, los autores indican que los transcritos más degradados en FFPE tienden a presentar un mayor MFE y una menor longitud [32], de acuerdo con nuestros resultados en los que los transcritos más representados en la cohorte “FFPE” muestran, en promedio, valores de MFE más elevados que los de la cohorte “Fresco”. Esta aparente discrepancia podría atribuirse a diferencias metodológicas, al tamaño reducido de la muestra o a la selección específica de genes analizados. Por tanto, se considera necesario ampliar el estudio con un mayor número de transcritos y condiciones experimentales más controladas, con el fin de determinar si existe una tendencia sistemática a la degradación del ARN en función de su MFE, y si dicha tendencia se asocia de forma consistente a valores más altos o más bajos de esta métrica.

Por tanto, los mecanismos exactos por los que la longitud del transcrito, contenido en GC y energía libre mínima explican las diferencias observadas aún no se comprenden completamente. Una posible hipótesis es que, dado que el proceso de parafinado y posterior desparafinado puede inducir fragmentación del ARN, los transcritos de menor longitud podrían degradarse completamente, impidiendo su detección. En cambio, los transcritos más largos podrían conservar fragmentos suficientes para ser detectados durante el análisis transcriptómico, lo que explicaría su mayor representación en muestras FFPE. Por otro lado, las estructuras estables son menos atacables por las condiciones de degradación de la fijación-parafinado.

Asimismo, otra variable encontrada en la literatura es la abundancia del ARNm [24,25], la cuál no hemos analizado por falta de tiempo, pero que podría respaldar parte de estas diferencias en la detección de transcritos.

Para validar los resultados de la secuenciación NGS, se llevó a cabo una comprobación experimental mediante qPCR con algunos genes representativos de cada set de datos. En concreto, los genes *IRF2BPL*, *CD83* y *SLC17A7* mostraron una alta expresión en tejido fresco y

una expresión prácticamente nula en tejido parafinado, en concordancia con los análisis computacionales. Por otro lado, el gen *HBA1/2* presentó una sobreexpresión en muestras “FFPE” respecto a las de “Fresco”. El único gen cuya expresión no se ajustó a lo previsto fue *ATXN7*, que no mostró diferencias significativas entre ambos tipos de tejido. Esta discrepancia podría deberse a que el cambio de expresión observado informáticamente no era tan marcado como en el caso de *HBA1/2* (de 4,98 a 53 frente a 3,7 a 300, respectivamente). Este comportamiento podría sugerir que los genes con mayor expresión en tejido parafinado podrían estar más expuestos a variabilidad técnica o a efectos impredecibles derivados del proceso de fijación, especialmente cuando las diferencias de expresión no son suficientemente amplias. En cambio, los genes que muestran una mayor expresión en tejido fresco tienden a reflejar con mayor fidelidad los resultados bioinformáticos de secuenciación gracias a una mayor integridad.

Más novedoso fue el análisis de los perfiles de metilación en ambas cohortes, con el objetivo de identificar posibles patrones diferenciales entre tejidos frescos y parafinados. En este caso, contamos con la ventaja de haber utilizado la misma tecnología (mismo formato de beadchips y protocolo de marcado e hibridación) y sí se pudo realizar un análisis de metilación diferencial, que resultó mucho más informativo que la comparación de *rankings* (mayor número de resultados, acompañados de estadístico). Durante el análisis llevado a cabo con “limma”, se observó que los resultados significativos provenían principalmente de tejido fresco (92%). Esto podría indicar que en el tejido parafinado hay una pérdida de metilación en comparación con el tejido fresco. Esta diferencia podría deberse a una mayor degradación producida por el tratamiento con bisulfito, que es clave en el análisis del metiloma, tal y como se ha documentado previamente [38]. Este efecto parece además agravarse cuanto más antigua es la muestra parafinada.

Por otra parte, se han observado diferencias en la distribución de la metilación según su posición respecto a las islas CpG entre ambas condiciones. En “Fresco” la mayor parte de estas modificaciones se encuentran en regiones alejadas de las islas CpG, mientras que en “FFPE” esta distribución es más homogénea y encontramos mayor proporción en islas CpG y regiones adyacentes. Se desconoce cuál es la causa de dichas diferencias, aunque podrían ser debidas a que la metilación en regiones cercanas o en la propia isla CpG aporta una mayor

protección, haciendo que se degraden o modifiquen menos en comparación a otras regiones del genoma.

Por último, se analizó también la distribución de CpGs por cromosoma. Debido a la falta de tiempo no se ha podido analizar en profundidad las diferencias en la distribución y cuál es su origen, aunque la distribución y proporción de eucromatina (más activa transcripcionalmente) y heterocromatina (inactiva en la transcripción) pueden ser un factor interesante a analizar de cara al futuro, ya que estas cambian según el tipo de célula y sus condiciones y podrían modular la accesibilidad a la degradación del ADN.

Conclusiones y proyección futura

Las principales conclusiones que hemos obtenido son:

- Los transcritos más sobre-representados en tejido fresco suelen ser más cortos y con un mayor contenido en GC, además de presentar una tendencia a poseer un MFE menor, haciéndolos más estables.
- Se detectan más eventos de metilación en dinucleótidos CpG en muestras frescas con respecto a tejido parafinado.

Como proyección futura, sería conveniente profundizar y ampliar el estudio de factores como MFE (asociación empírica entre la estabilidad estructural del ARN, reflejada en valores de MFE) y su susceptibilidad a la degradación durante el proceso de fijación en parafina) y los perfiles de metilación, los cuales no han podido analizarse en profundidad debido a limitaciones de tiempo, ya que no ha sido posible validar algunos candidatos mediante pirosecuenciación. Por otro lado, el análisis detallado de los perfiles de metilación podría permitir identificar factores epigenéticos que contribuyan a las diferencias observadas en la distribución de CpGs entre tejidos frescos y parafinados. En particular, sería de interés explorar si estas diferencias están relacionadas con el estado de la cromatina (eucromatina o heterocromatina) u otras características estructurales del ADN, lo que podría aportar información relevante sobre la accesibilidad y estabilidad del material genético en función del tipo de tejido y su procesamiento.

Bibliografía

1. Oncoclínicas. Glioblastoma. Disponible en: <https://gruponcoclinicas.com/en/tudo-sobre-o-cancer/tipos-de-cancer/glioblastoma#:~:text=Glioblastoma%20is%20the%20type%20of,and%20difficulty%20learning%20are%20symptoms>.
2. Schaff LR, Mellinghoff IK. Glioblastoma and Other Primary Brain Malignancies in Adults: A Review. JAMA. 2023 Feb 21;329(7):574-587. doi: 10.1001/jama.2023.0023. PMID: 36809318; PMCID: PMC11445779.
3. Hans-Georg Wirsching, Evanthia Galanis, Michael Weller, Chapter 23 - Glioblastoma, Editor(s): Mitchel S. Berger, Michael Weller, Handbook of Clinical Neurology, Elsevier, Volume 134, 2016, Pages 381-397, ISSN 0072-9752, ISBN 9780128029978, <https://doi.org/10.1016/B978-0-12-802997-8.00023-2>. (<https://www.sciencedirect.com/science/article/pii/B9780128029978000232>)
4. Melero Gea J. Modulación epigenética en el tratamiento personalizado del glioblastoma [Trabajo de Fin de Grado]. Elche: Universidad Miguel Hernández; 2024. Disponible en: <https://hdl.handle.net/11000/32762>.
5. Haynes HR, Killick-Cole CL, Hares KM, Redondo J, Kemp KC, Moutasim KA, Faulkner C, Wilkins A, Kurian KM. Evaluation of the quality of RNA extracted from archival FFPE glioblastoma and epilepsy surgical samples for gene expression assays. J Clin Pathol. 2018 Aug;71(8):695-701. doi: 10.1136/jclinpath-2017-204969. Epub 2018 Feb 20. PMID: 29463577.
6. Guo X, Gu L, Li Y, Zheng Z, Chen W, Wang Y, Wang Y, Xing H, Shi Y, Liu D, Yang T, Xia Y, Li J, Wu J, Zhang K, Liang T, Wang H, Liu Q, Jin S, Qu T, Guo S, Li H, Wang Y, Ma W. Histological and molecular glioblastoma, IDH-wildtype: a real-world landscape using the 2021 WHO classification of central nervous system tumors. Front Oncol. 2023 Jul 6;13:1200815. doi: 10.3389/fonc.2023.1200815. PMID: 37483487; PMCID: PMC10358772.
7. Aldape K, Zadeh G, Mansouri S, Reifenberger G, von Deimling A. Glioblastoma: pathology, molecular mechanisms and markers. Acta Neuropathol. 2015 Jun;129(6):829-48. doi: 10.1007/s00401-015-1432-1. Epub 2015 May 6. PMID: 25943888.
8. Lexogen. What are FFPE samples and how are they prepared? 2025. Disponible en: <https://www.lexogen.com/what-are-ffpe-samples-and-how-are-they-prepared/>
9. National Human Genome Research Institute. Epigenética [Internet]. Bethesda (MD): National Human Genome Research Institute; 2023 [citado 3 de abril de 2025]. Disponible en: <https://www.genome.gov/es/genetics-glossary/Epigenetica>

10. MedlinePlus. ¿Qué es la epigenética? [Internet]. Bethesda (MD): MedlinePlus; 2023 [citado 3 de abril de 2025]. Disponible en: <https://medlineplus.gov/spanish/genetica/entender/comofuncionangenese/epigenetica/>
11. Esteller Badosa M. Epigenetic and epitranscriptomic pharmacological compound. *An Real Acad Farm.* 2024;90(1):7–19.
12. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, Hawkins C, Ng HK, Pfister SM, Reifenberger G, Soffietti R, von Deimling A, Ellison DW. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol.* 2021 Aug 2;23(8):1231-1251. doi: 10.1093/neuonc/noab106. PMID: 34185076; PMCID: PMC8328013.
13. Antonelli M, Poliani PL. Adult type diffuse gliomas in the new 2021 WHO Classification. *Pathologica.* 2022 Dec;114(6):397-409. doi: 10.32074/1591-951X-823. PMID: 36534419; PMCID: PMC9763975.
14. Gaillard F, Mazini B, Sharma R, et al. Astrocytoma, IDH-mutant. Reference article, Radiopaedia.org (Accessed on 07 May 2025) <https://doi.org/10.53347/rID-14598>
15. Gaillard F, Walizai T, Schultz K, et al. Oligodendroglioma. Reference article, Radiopaedia.org (Accessed on 07 May 2025) <https://doi.org/10.53347/rID-1772>
16. Gaillard F, Le L, Abu Kamesh M, et al. Glioblastoma, IDH-wildtype. Reference article, Radiopaedia.org (Accessed on 07 May 2025) <https://doi.org/10.53347/rID-4910>
17. Miller KD, Ostrom QT, Kruchko C, Patil N, Tihan T, Cioffi G, Fuchs HE, Waite KA, Jemal A, Siegel RL, Barnholtz-Sloan JS. Brain and other central nervous system tumor statistics, 2021. *CA Cancer J Clin.* 2021 Sep;71(5):381-406. doi: 10.3322/caac.21693. Epub 2021 Aug 24. PMID: 34427324.
18. National Cancer Institute. Diccionario de cáncer [Internet]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/gen-idh1>
19. National Cancer Institute. Diccionario de cáncer [Internet]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/gen-idh2>
20. Neilsen BK, Sleightholm R, McComb R, Ramkissoon SH, Ross JS, Corona RJ, Miller VA, Cooke M, Aizenberg MR. Comprehensive genetic alteration profiling in primary and recurrent glioblastoma. *J Neurooncol.* 2019 Mar;142(1):111-118. doi: 10.1007/s11060-018-03070-2. Epub 2018 Dec 9. PMID: 30535594.
21. Hervás-Corpión I, Gallardo-Orihuela A, Catalina-Fernández I, Iglesias-Lozano I, Soto-Torres O, Geribaldi-Doldán N, Domínguez-García S, Luna-García N, Romero-García R, Mora-López F, Iriarte-Gahete M, Morales JC, Campos-Caro A, Castro C, Gil-Salú JL, Valor LM. Potential Diagnostic Value of the Differential Expression of Histone H3 Variants between Low- and High-Grade Gliomas.

- Cancers (Basel). 2021 Oct 20;13(21):5261. doi: 10.3390/cancers13215261. PMID: 34771425; PMCID: PMC8582563.
22. Hervás-Corpión I, Navarro-Calvo J, Martín-Climent P, Iriarte-Gahete M, Geribaldi-Doldán N, Castro C, Valor LM. Defining a Correlative Transcriptional Signature Associated with Bulk Histone H3 Acetylation Levels in Adult Glioblastomas. *Cells*. 2023 Jan 19;12(3):374. doi: 10.3390/cells12030374. PMID: 36766715; PMCID: PMC9913072.
 23. García-Heredia A, Guerra-Núñez L, Martín-Climent P, Rojas E, López-Domínguez R, Alcántara-Domínguez C, Alenda C, Valor LM. Transcriptomics and epigenomics datasets of primary brain cancers in formalin-fixed paraffin embedded format. *Sci Data*. 2025 Feb 15;12(1):273. doi: 10.1038/s41597-025-04597-6. PMID: 39955294; PMCID: PMC11830079.
 24. Graw S, Meier R, Minn K, Bloomer C, Godwin AK, Fridley B, et al. Robust gene expression and mutation analyses of RNA-sequencing of formalin-fixed diagnostic tumor samples. *Sci Rep*. 2015;5:12335. doi:10.1038/srep12335.
 25. Liu Y, Bhagwate A, Winham SJ, Stephens MT, Harker BW, McDonough SJ, et al. Quality control recommendations for RNASeq using FFPE samples based on pre-sequencing lab metrics and post-sequencing bioinformatics metrics. *BMC Med Genomics*. 2022;15(1):195. doi:10.1186/s12920-022-01355-0.
 26. Seifert M, Schackert G, Temme A, Schröck E, Deutsch A, Klink B. Molecular characterization of astrocytoma progression towards secondary glioblastomas utilizing patient-matched tumor pairs. *Cancers (Basel)*. 2020;12(6):1696. doi:10.3390/cancers12061696
 27. Ricaurte O, Neita K, Valero D, Ortega-Rojas J, Arboleda-Bustos CE, Zubieta C, et al. Estudio de mutaciones en los genes IDH1 e IDH2 en una muestra de gliomas de población colombiana. *Biomedica [Internet]*. 2018 May 1 [citado 2025 Jun 29];38(Sup.1):86–92. Disponible en: <https://revistabiomedica.org/index.php/biomedica/article/view/3708>
 28. Xiong A, Zhang J, Chen Y, Zhang Y, Yang F. Integrated single-cell transcriptomic analyses reveal that GPNMB-high macrophages promote PN-MES transition and impede T cell activation in GBM. *EBioMedicine*. 2022 Sep;83:104239. doi:10.1016/j.ebiom.2022.104239. PMID: 36054938; PMCID: PMC9437813.
 29. Gangoso E, Southgate B, Bradley L, Rus S, Galvez-Cancino F, McGivern N, et al. Glioblastomas acquire myeloid-affiliated transcriptional programs via epigenetic immunoediting to elicit immune evasion. *Cell*. 2021;184(9):2454–70.e26. doi:10.1016/j.cell.2021.03.023. PMID: 33857425; PMCID: PMC8099351.

30. Fu Y, Yi Y, Shao Y, Jiang J, Deng Q. Single-cell and spatial transcriptomic insights into glioma cellular heterogeneity and metabolic adaptations. *Front Immunol.* 2025;16:1561388. doi:10.3389/fimmu.2025.1561388.
31. Bou Zerdan M, Atoui A, Hijazi A, Basbous L, Abou Zeidane R, Alame SM, Assi HI. Latest updates on cellular and molecular biomarkers of gliomas. *Front Oncol.* 2022;12:1030366. doi:10.3389/fonc.2022.1030366.
32. Lin Y, Dong ZH, Ye TY, Yang JM, Xie M, Luo JC, et al. Optimization of FFPE preparation and identification of gene attributes associated with RNA degradation. *NAR Genom Bioinform.* 2024;6(1):lqae008. doi:10.1093/nargab/lqae008.
33. Esteve-Codina A, et al. A comparison of RNA-Seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. *PLoS One.* 2017;12:e0170632. doi:10.1371/journal.pone.0170632.
34. Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res.* 1999;27:4436–43. doi:10.1093/nar/27.22.4436.
35. Shohdy KS, et al. Functional comparison of exome capture-based methods for transcriptomic profiling of formalin-fixed paraffin-embedded tumors. *NPJ Genom Med.* 2021;6:66. doi:10.1038/s41525-021-00231-7.
36. Noguera-Castells A, Garcia-Prieto CA, Alvarez-Errico D, Esteller M. Validation of the new EPIC DNA methylation microarray (900K EPIC v2) for high-throughput profiling of the human DNA methylome. *Epigenetics.* 2023;18:2185742. doi:10.1080/15592294.2023.2185742.
37. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47. doi:10.1093/nar/gkv007.
38. Herrera Jehison A., Segura Yenifer Y., Montero Wendy J., Medina Vilma, Moreno-Acosta Pablo, Romero Alfredo E. et al. Evaluación de la metilación del DNA en tejidos de cáncer colorrectal fijados en formalina y embebidos en parafina (FFPE). *rev.colomb.cancerol.* [Internet]. 2021 June [cited 2025 July 14];25(2):110-114. Available from: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-90152021000200110&lng=en. Epub May 06, 2022. <https://doi.org/10.35509/01239015.704>