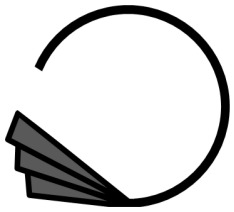
Instituto de Cibernética, Matemática y Física
Departamento de Física Teórica

TESIS DE MAESTRÍA

# ENVEJECIMIENTO, ALZHEIMER Y GLIOBLASTOMA EN EL ESPACIO DE EXPRESIÓN GENÉTICA

**Autor:** Joan Andrés Nieves Cuadrado

**Tutor:** Dr. Augusto González García, *ICIMAF*



La Habana, 2024

*Lo que sabemos es una gota de agua;*
*lo que ignoramos es el océano.*
*Isaac Newton*

# Resumen

Los datos disponibles de la materia blanca de cerebro permiten localizar los atractores normal (homeostático), Glioblastoma y Alzheimer en el espacio de expresión genética e identificar caminos relacionados con transiciones como la carcinogénesis o la aparición del Alzheimer. También se aprecia una trayectoria predefinida para el envejecimiento, lo cual es consistente con la hipótesis del envejecimiento programado. Adicionalmente, suposiciones razonables sobre la fortaleza relativa de los atractores permite dibujar un panorama esquemático del *fitness*: diagrama de Wright. Estos sencillos diagramas reproducen relaciones conocidas entre el envejecimiento, el glioblastoma y el Alzheimer, y plantea cuestiones interesantes como la posible conexión entre el envejecimiento programado y el glioblastoma en este tejido. Prevemos que múltiples diagramas similares en otros tejidos podrían ser útiles en el entendimiento de la biología de enfermedades o trastornos aparentemente no relacionados, y para descubrir pistas inesperadas para su tratamiento.

# Abstract

Available data for white matter of the brain allows to locate the normal (homeostatic), glioblastoma and Alzheimer's disease attractors in gene expression space and to identify paths related to transitions like carcinogenesis or Alzheimer's disease onset. A predefined path for aging is also apparent, which is consistent with the hypothesis of programmatic aging. In addition, reasonable assumptions about the relative strengths of attractors allow to draw a schematic landscape of fitness: a Wright's diagram. These simple diagrams reproduce known relations between aging, glioblastoma and Alzheimer's disease, and rise interesting questions like the possible connection between programmatic aging and glioblastoma in this tissue. We anticipate that similar multiple diagrams in other tissues could be useful in the understanding of the biology of apparently unrelated diseases or disorders, and in the discovery of unexpected clues for their treatment.

# Índice general

**Resumen** **ii**

**Abstract** **iii**

**Índice general** **iv**

**Introducción** **1**

**1 Materiales y métodos** **5**
1.1 Reducción de la dimensionalidad . . . . . 5
1.2 Datos de expresión genética . . . . . 10

**2 Diagrama de tres atractores** **12**
2.1 El diagrama de N + GB + EA . . . . . 12
2.2 Panorama del *fitness* . . . . . 16
2.3 Limitaciones . . . . . 18

**3 Resultados** **19**
3.1 Principales resultados . . . . . 19
3.2 Perspectiva cuantitativa . . . . . 25

**Conclusiones** **27**

**Recomendaciones** **28**

**Apéndices** **29**

**A Modelo de ratón para el envejecimiento** **29**

**B Eje PC1 (envejecimiento)** **30**

**C Eje PC2 (GB vs. EA)** **33**

**D Transición de O a la EA** **36**

**Bibliografía** **37**

# Introducción

El *fitness* celular refiere a la capacidad de una célula de sobrevivir y proliferar en un ambiente determinado. Abarca varios factores como la capacidad de la célula para adaptarse al ambiente, resistir al estrés y mantener la homeostasis. Este concepto es particularmente importante en la biología del cáncer, donde los niveles de *fitness* celular pueden determinar su supervivencia y dominancia dentro de un tejido.

Un paradigma conocido en la biología molecular expresa que los máximos locales de *fitness*, en el espacio de expresión genética, están relacionados con estados biológicos accesibles. Un diagrama de Wright es una representación gráfica reducida del panorama genético, donde los picos y valles representan diferentes genotipos y sus niveles relativos de *fitness* [1], ver Fig. 1.

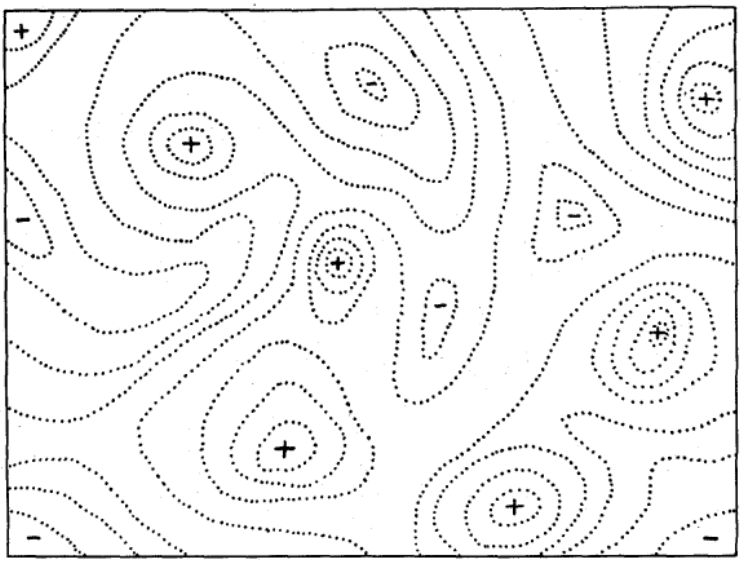


Figura 1: Representación esquemática de un diagrama de Wright. Las zonas de mayor *fitness* (+) están asociadas con estados biológicos accesibles. Las de menor *fitness* (-) son solo zonas de transito. Las lineas discontinuas delimitan los distintos niveles de *fitness*. Tomada de la referencia [1].

Esta representación ha sido aplicada a la descripción del destino celular a lo largo de una linea de diferenciación [2]. Sin embargo, hasta donde sabemos, no hay gráficos basados en datos reales para un tejido dado que represente, al menos de forma parcial, un diagrama de Wright con más de dos máximos. En el presente trabajo, mostramos un diagrama para la materia blanca del cerebro en el cual el estado normal (N) se representa junto con el atractor de glioblastoma (GB) y el máximo relacionado con la enfermedad del Alzheimer (EA).

La EA se caracteriza por la pérdida progresiva de células neuronales, mientras que el GB es un tipo de cáncer cerebral que implica la proliferación descontrolada de células. Algunas investigaciones muestran que pacientes con la EA podrían tener un menor riesgo de desarrollar GB. Por ejemplo, en la referencia [3] se muestra que ambas enfermedades presentan un aumento en el estrés oxidativo, en la EA, esto hace que las células neuronales sean más vulnerables a la muerte, mientras que en el GB, las células cancerosas se vuelven más resistentes. Además, se plantea que la degeneración de células que secretan acetilcolina en la EA podría tener un efecto protector contra el cáncer, ya que esta sustancia puede estimular el crecimiento de células cancerosas.

Los estudios epidemiológicos señalan que aquellas personas que padecen una de estas enfermedades tiene un riesgo menor de desarrollas la otra. Concretamente, los pacientes con cáncer tienen un riesgo alrededor de 31 y 35 % menor de desarrollar la EA en comparación con aquellos sin antecedente de cáncer. Mientras que las personas que padecen la EA tienen un riesgo entre 41 y 61 % menor de desarrollar cáncer [4–6].

Esta idea de la EA y el GB como alternativas opuestas también está apoyada por un gran número experimentos de biología molecular. Por ejemplo, en la referencia [7] los autores encuentran que las vías de señalización ERK/MAPK están aumentadas en GB y disminuidas en la EA. Estas vías conectan señales extracelulares, como factores de crecimiento y citocinas, a respuestas intracelulares que regulan la proliferación celular, diferenciación y supervivencia. También muestran que las vías de señalización de la angiopoyetina están aumentadas en la EA y disminuidas en GB, estas relacionadas con la regulación de la angiogénesis y la estabilidad vascular.

El estudio realizados por C. Lanni y colaboradores en la referencia [8] destaca varios actores moleculares, especialmente PIN1 y p53, que están involucrados en interacciones moleculares complejas asociadas con la correlación inversa entre estas enfermedades. El aumento en la expresión de PIN1 esta relacionado con un retraso en la edad de inicio de la EA, mientras

niveles bajos de expresión se asocian con un menor riesgo de desarrollar varios tipos de cáncer.

Muchas de estas investigaciones no solo muestran una oposición entre la EA y el GB, sino que también señalan el envejecimiento como un factor de riesgo común para ambas enfermedades [5–9]. Esta compleja relación entre ambas enfermedades cerebrales queda representada en nuestro diagrama de Wright. Además, se puede apreciar un camino o corredor hacia el envejecimiento normal, en línea con la teoría del envejecimiento programado [10, 11].

A nivel genético, hay genes que varían de la misma manera en los procesos de envejecimiento, progresión al cáncer y la EA, mientras que también hay genes que indican una situación disyuntiva entre la EA y el GB. Un ejemplo de este último es el gen codificador de proteínas MMP9, que juega un papel importante en la invasión tumoral [12, 13], pero también conocido como neuroprotector, controlando las interacciones entre axones y fibras beta-amiloide [14]. Las desviaciones del valor de expresión génica de su referencia en el tejido normal pueden indicar una progresión potencial a la EA (subexpresión) o al GB (sobreexpresión).

Esta visión inusual puede ayudar a comprender las relaciones entre la EA y el GB, e identificar marcadores génicos útiles para ambos procesos. Como un bono adicional, la representación permite encontrar preguntas muy interesantes que se discutirán a continuación.

El trabajo presentado como tesis de diploma consta de una introducción, tres capítulos, las conclusiones y las recomendaciones. El contenido se distribuye de la siguiente forma:

- **Introducción:** Se presenta análisis del panorama genético y epidemiológico relacionado con la EA, el GB y el envejecimiento. Se describe la relación inversa entre la EA y el GB basada en evidencia epidemiológica y molecular. Además, se hace referencia a la idea del envejecimiento programado como factor de riesgo para ambas enfermedades.
- **Capítulo 1:** Se centra en los métodos y materiales utilizados para analizar los datos de expresión genética. Se detalla la aplicación de técnicas de reducción de dimensionalidad, para simplificar y visualizar conjuntos de datos complejos. Además, se describen las dos principales fuentes de datos. Se explica también el procesamiento de los datos y su conversión a una expresión diferencial logarítmica para facilitar el análisis comparativo.
- **Capítulo 2:** Se centra en la construcción y análisis del diagrama de tres atractores que emergen en el espacio de expresión genética. Mediante técnicas de reducción dimensional se visualizan las posiciones relativas de estos atractores y se observa cómo las muestras de tejido se desplazan de un estado a otro. Además, se introduce un panorama del fitness

celular a través de un diagrama de Wright que ilustra las barreras y la fuerza relativa de cada atractor, enfatizando las diferencias en la transición de un estado a otro.

- **Capítulo 3:** Presenta los resultados obtenidos del análisis de los datos de expresión genética. En este capítulo se destacan, por un lado, los hallazgos cualitativos, en los que se identifican claramente los atractores asociados al estado normal, al glioblastoma y a la enfermedad de Alzheimer, evidenciando las trayectorias y transiciones entre estos estados; y, por otro lado, se ofrece una perspectiva cuantitativa, respaldada por representaciones gráficas y análisis estadísticos, que valida la relación inversa entre el glioblastoma y el Alzheimer y resalta el papel del envejecimiento en estas transiciones.
- **Conclusiones:** Se sintetiza como el análisis de la expresión génica en la materia blanca del cerebro permite identificar tres atractores y ponen de manifiesto una relación inversa entre el desarrollo del glioblastoma y la aparición del Alzheimer.
- **Recomendaciones:** Se destaca la necesidad de realizar estudios futuros que afinen y amplíen el modelo propuesto, proponiendo la integración de técnicas analíticas más robustas y la incorporación de datos de plataforma unificada para superar las limitaciones actuales de heterogeneidad.

# Capítulo 1

# Materiales y métodos

## 1.1 Reducción de la dimensionalidad

La reducción de la dimensionalidad es un paso crucial en el análisis de datos, especialmente cuando se trata de grandes conjuntos con muchas variables. Este facilita la visualización y comprensión de los datos, lo que permite una rápida identificación de patrones y tendencias importantes. Algunas técnicas comunes de reducción dimensional son PCA (*Principal Component Analisis*), t-SNE (*t-Distributed Stochastic Neighbor Embedding*) y UMAP (*Uniform Manifold Approximation and Projection*). Estas técnicas son herramientas poderosas que ayudan a simplificar la complejidad de los datos sin perder información crítica, lo que permite un análisis más eficiente y efectivo.

El t-SNE es un método no linear y estocástico. Su funcionamiento se puede separar en dos etapas. En la primera, se seleccionan los vecinos de cada punto. Para ello se utiliza una distribución gaussiana alrededor de él, donde los más cercanos tienen un probabilidad mayor de ser seleccionados que los lejanos. Este paso permite al modelo preservar las estructuras locales. Durante la segunda etapa, se asignan posiciones iniciales aleatorias en un espacio de menor dimensión (generalmente 2 o 3 dimensiones). Luego, se define una distribución de probabilidad similar para los puntos en el nuevo espacio y se minimiza la divergencia entre las dos distribuciones. Esta etapa ayuda a mantener la fidelidad de la representación en el espacio reducido. De esta forma, el algoritmo logra una transformación de los datos hacia una dimensión reducida que preserva la similitud entre los vecinos cercanos [15, 16].

UMAP es una técnica muy similar a t-SNE. Una de las diferencias principales es que, durante la selección de los vecinos, se asume que los datos forman una variedad de menor dimensión que el espacio original. Esto le permite ser más eficiente en términos de tiempo de cómputo y más efectivo en la conservación de relaciones a gran escala. Otra característica de este método es que, para hacer la representación reducida, minimiza la entropía cruzada en lugar de la divergencia entre las distribuciones. Estas diferencias permiten a este algoritmo preservar mejor tanto la estructura local como la global de los datos, además de hacerlo capaz de trabajar con datos que no se ajustan necesariamente a una distribución normal. [17, 18]

Por otro lado, el PCA es una técnica lineal y determinista que transforma variables correlacionadas en un conjunto reducido de variables no correlacionadas, conocidas como componentes principales. Al igual que en los casos anteriores, su funcionamiento se puede dividir en dos etapas. En la primera etapa, se centran los datos en su media aritmética y se calcula la matriz de covarianza. Esta matriz captura la variabilidad conjunta entre múltiples variables aleatorias y permite comprender las relaciones entre ellas. Luego, durante la segunda etapa, se obtienen los autovalores, ordenados de mayor a menor, y sus correspondientes autovectores. Los autovectores representan las direcciones de máxima varianza en los datos, mientras que los autovalores indican la cantidad de varianza que se encuentra en cada una de estas direcciones. Al proyectar los datos originales sobre los primeros autovectores, se obtiene una representación de los datos en un sistema ortogonal que maximiza la conservación de la varianza, utilizando el menor número posible de componentes [19].

t-SNE y UMAP se han vuelto muy populares últimamente debido a su eficiencia y la gran capacidad de para visualizar datos de alta dimensión en un espacio de 2 o 3 dimensiones. Pero su naturaleza no linear y estocástica hace que sea complejo una interpretación cuantitativa de los resultados. Por otro lado, PCA es una técnica lineal y determinista que junto a su relativa sencillez permite realizar varias interpretaciones de sus resultados.

Sin embargo, la aplicación directa del PCA puede presentar algunas complicaciones. Una de las principales dificultades es la construcción de la matriz de covarianza, ya que el número de elementos que contiene es igual al cuadrado de la dimensión original de los datos. Esto hace que sea imposible almacenarla en la memoria RAM de la mayoría de los equipos de cómputo personales. Por ejemplo, si el número de componentes de los datos es $5 \times 10^4$, la matriz de covarianza tendría $2{,}5 \times 10^9$ elementos. Suponiendo que cada elemento ocupe 8 bytes, el tamaño total de la matriz sería aproximadamente 19 GB. Si se hace uso de que

esta es una matriz simétrica, se podría guardar solo los elementos de la triangular superior (o inferior), permitiendo reducir el almacenamiento necesario casi a la mitad. Sin embargo, aún así se requeriría mucho espacio y este escala rápidamente con el aumento de la dimensión de los datos. Por lo tanto, en general, es necesario recurrir al almacenamiento en disco, que tiene una velocidad de lectura y escritura menor que la memoria RAM.

Otro problema grave que enfrenta el algoritmo estándar del PCA es el cálculo de los autovalores y autovectores. La mayoría de las implementaciones de los métodos directos usados para este cálculo no permiten que se apliquen a grandes matrices debido a las limitaciones de la memoria RAM. Una característica del PCA que resulta de gran ayuda en esta parte es que, en general, no es necesario calcular todos los autovalores, solo los más grandes y sus correspondientes autovectores.

Un algoritmo relativamente sencillo de implementar y que permite calcular solo los autovalores más grandes y sus correspondientes autovectores es el método de Lanczos. En el caso del análisis de datos de expresión genética, donde solo un subconjunto diferente de genes se expresa en cada tejido, la matriz de covarianza podría tener un número elevado de valores nulos. Esto es beneficioso para el método de Lanczos, ya que funciona mejor con matrices dispersas.

El método de Lanczos puede ser de gran utilidad en algunos problemas donde las implementaciones estándar pueden verse limitadas. Sin embargo, en la práctica se utilizan otras técnicas para realizar el PCA de forma indirecta. Una de las alternativas es la descomposición en valores singulares (SVD, por sus siglas en inglés). A continuación se describirá brevemente como funciona la SVD, algunas de sus propiedades y como realizar el PCA a partir de esta.

## Descomposición en valores singulares

La SVD provee una descomposición numéricamente estable de matrices que puede ser usado en una gran variedad de propósitos. Como resultado de aplicar este algoritmo se obtiene una descomposición matricial única que existe para toda matriz de valores complejos $\mathbf{X} \in \mathbb{C}^{n \times m}$:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \tag{1.1}$$

donde $\mathbf{U} \in \mathbb{C}^{n \times n}$ y $\mathbf{V} \in \mathbb{C}^{m \times m}$ son matrices unitarias con columnas ortonormales, y $\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$ una matriz con valores reales no negativos en la diagonal y ceros fuera de la diagonal. Aquí $^*$

denota la transpuesta conjugada.

Cuando $n \geq m$, la matriz $\mathbf{\Sigma}$ tiene como máximo $m$ valores distintos de cero en la diagonal y puede ser escrita como $\mathbf{\Sigma} = \begin{bmatrix} \hat{\mathbf{\Sigma}} \\ 0 \end{bmatrix}$. Por lo tanto, es posible representar $\mathbf{X}$ de forma exacta usando la versión reducida de SVD:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* = \begin{bmatrix} \hat{\mathbf{U}} & \mathbf{U}^\perp \end{bmatrix} \begin{bmatrix} \hat{\mathbf{\Sigma}} \\ 0 \end{bmatrix} \mathbf{V}^*, \tag{1.2}$$

las columnas de $\mathbf{U}^\perp$ abarcan un espacio vectorial que es complementario y ortogonal a $\hat{\mathbf{U}}$. Las columnas de $\mathbf{U}$ son llamadas vectores singulares izquierdos de $\mathbf{X}$ y forman una base del espacio de los vectores columnas de $\mathbf{X}$. Las columnas de $\mathbf{V}$ son los vectores singulares derechos y forman una base para los vectores filas de $\mathbf{X}$. Los elementos diagonales de $\hat{\mathbf{\Sigma}} \in \mathbb{C}^{m \times m}$ , los llamados valores singulares, están ordenados de mayor a menor. El rango de $\mathbf{X}$ es igual a la cantidad de valores singulares distintos de cero.

Para ver la relación de esta técnica con el PCA, partimos de la matriz de covarianza. Esta se construye a partir de la siguiente expresión:

$$\sigma_{ij} = \frac{1}{N-1} \sum_{l=1}^{N} \left( x_i^{(l)} - \mu_i \right) \left( x_j^{(l)} - \mu_j \right), \tag{1.3}$$

donde $i$ y $j$ son la característica $i$-ésima y $j$-ésima del conjunto de datos estudiado, en nuestro caso son el gen $i$ y $j$, respectivamente. La variable $l$ se recorre por todas las muestras. $\sigma_{ij}$ es el elemento $(i,\ j)$ de la matriz de covarianza, es decir, es la covarianza entre el gen $i$ y el $j$ para los elementos no diagonales y la varianza del gen $i$ para los elementos de la diagonal principal. El número total de muestras es $N$, $x_i^{(l)} \in \mathbb{R}$ es el valor de la componente $i$ en la muestra $l$, y, por último, $\mu_i$ es la media de los valores de la componente $i$ sobre todas las muestras.

Si los datos ya han sido previamente centrados, es decir, se cumple que $\mu_i = 0$ para todo $i$, entonces la ecuación (1.3) puede reducirse a:

$$\sigma_{ij} = \frac{1}{N-1} \sum_{l=1}^{N} x_i^{(l)} x_j^{(l)}, \tag{1.4}$$

Si definimos una matriz $\mathbf{X} \in \mathbb{R}^{n \times m}$, tal que el elemento $(i,\ j)$ es igual a $x_i^{(j)}/(N-1)$, la ecuación (1.4) queda representada en forma matricial como:

$$\mathbf{C} = \mathbf{X}^T\mathbf{X}. \tag{1.5}$$

Si usamos la SVD de $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ y la propiedad de ortonormalidad de $\mathbf{U}$ y $\mathbf{V}$, obtenemos:

$$\mathbf{C} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{1.6}$$

$$= \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T, \tag{1.7}$$

es decir, $\mathbf{\Sigma}$ y $\mathbf{V}$ son la solución del siguiente problema de autovalores:

$$\mathbf{C}\mathbf{V} = \mathbf{V}\mathbf{\Sigma}^2. \tag{1.8}$$

En otras palabras, cada valor singular de $\mathbf{X}$ distinto de cero es la raíz cuadrada positiva de los autovalores de su matriz de covarianza, y las columnas de $\mathbf{V}$ son los autovectores. En términos del PCA, las columnas de $\mathbf{V}$ son las componentes principales de $\mathbf{X}$ y los elementos de la diagonal de $\mathbf{\Sigma}^2$ representan la varianza de los datos en cada una de las componente.

Esta es la manera usual en la que los algoritmo actuales realizan el PCA, por ejemplo, la librería *scikit-learn* de Python [20]. En general, nosotros preferimos usar directamente SVD para hacer el PCA, ya que brinda mayor flexibilidad y control. Por ejemplo, si en lugar usar una matriz de media cero, los datos se centran en el valor medio de un subconjunto, las componentes principales van a indicar la dirección en la cual hay más dispersión respecto a este subjconjunto.

Una ventaja significativa de este método para realizar el PCA, es que se obtiene las componentes principales directamente, sin tener que calcular la matriz de covarianza. Esto permite ahorrar tiempo y recursos computacionales, algo que toma mayor importancia en la medida que aumenta la cantidad de datos y el números de variables a procesar. Un procedimiento similar se siguió en las referencias [21, 22].

## 1.2 Datos de expresión genética

Utilizamos datos de expresión genética obtenidos de dos experimentos distintos. El primero de ellos contiene muestras patológicamente normales (o "sanas") y con glioblastoma, y fueron tomadas del Atlas del Genoma del Cáncer (TCGA, https://www.cancer.gov/tcga) [23, 24]. Estas son tomadas durante procedimientos quirúrgicos. Los tumores se pueden localizar en diferente zonas cerebrales pero, como es común en el glioblastoma, estos son tumores tomados de la sustancia blanca del cerebro [25].

Hay 5 muestras normales de pacientes con edades en el rango entre 49 y 74 años, mientras que el intervalo de edades de las 169 muestras de glioblastoma es entre 21 y 89 años. Las pacientes femeninas representan aproximadamente dos tercios de la cohorte.

El segundo grupo de datos proviene del estudio longitudinal del Instituto Allen sobre el envejecimiento y la demencia (https://aging.brain-map.org/) [26]. Las muestras son tomadas *post mortem*. El grupo de control está conformado por 47 muestras, mientras que en el otro hay 28 muestras. El intervalo de edad para todas estas muestras es entre 77 y 101 años. El diagnóstico de la enfermedad del Alzheimer está respaldado por pruebas cognitivas y otras pruebas clínicas. Alrededor del 40 % de la cohorte son mujeres.

En ambos experimentos los valores de expresión genética se encuentran en unidades FPKM (*fragments per kilobase of transcript per million fragments mapped*), que es una unidad común en este tipo de estudios. En términos simples, dicha unidad de medida significa: la tasa de fragmentos por base multiplicada por un número muy grande ($10^9$). El cálculo de FPKM para el gen $i$ se realiza por medio de la siguiente fórmula [27]:

$$\begin{aligned} FPKM_i &= \frac{q_i}{(l_i/10^3)(\sum_j q_j/10^6)}, \\ &= \frac{q_i}{l_i \sum_j q_j} * 10^9, \end{aligned} \tag{1.9}$$

donde $q_i$ es la cantidad de fragmentos contados, $l_i$ es la longitud del gen, y $\sum_j q_j$ corresponde al número total de fragmentos.

El uso de esta unidad de medida puede resultar complicado debido a que el valor de expresión genética de muchos genes es cero o muy cercano a cero, mientras que un conjunto muy reducido puede presentar valores entre $10^2$ y $10^4$. Por esta razón, en muchos estudios

de análisis de expresión genética se utiliza una variable conocida como expresión diferencial logarítmica ($e_{fold}$). Para un gen determinado $i$, esta se calcula de la siguiente manera:

$$e^i_{fold} = \log_2 \left( \frac{e^i}{e^i_{ref}} \right), \tag{1.10}$$

donde $e^i$ es la expresión del gen $i$ y $e^i_{ref}$ es un valor de referencia para dicho gen, ambos en unidades FPKM. Esta nueva variable permite concentrar mejor los puntos, y su distribución tiende a asemejarse a una distribución gaussiana.

Usualmente, empleamos como valor de referencia del gen $i$ la media geométrica de su expresión en el conjunto de muestras normales (o "sanas"), tras haber sido desplazadas por una constante pequeña, generalmente 0,1 o 0,01. De esta forma, los valores negativos o positivos de la expresión diferencial logarítmica indican la subexpresión o sobreexpresión de un gen, respectivamente.

# Capítulo 2

# Diagrama de tres atractores

## 2.1 El diagrama de N + GB + EA

Nuestro punto de partida es el diagrama de los datos de expresión genética del análisis de componentes principales para la materia blanca del cerebro, mostrado en la Fig. 2.1. Como se puede notar en la figura, las dos primeras componentes principales capturan más del 80 % de la varianza del sistema. Por lo tanto, es una representación bidimensional adecuada de la distribución real de los puntos en el espacio de expresión genética.

En la figura se pueden apreciar 4 grupos de muestras. Las muestras marcadas como N y GB corresponden, a especímenes patológicamente normales y tumorales en los datos del TCGA para el glioblastoma [23]. Los centros de las nubes de muestras de N y GB en el espacio de expresión genética definen, respectivamente, los atractores Normal (homeostático) y Glioblastoma de Kauffman [21, 28]. De hecho, la acumulación de puntos en una determinada región de este espacio indica que esta es un atractor de la red de regulación Genética que gobierna la dinámica del sistema.

Por otro lado, los grupos etiquetados como EA y O corresponden a las muestras de la materia blanca del cerebro de la enfermedad del Alzheimer y del grupo de control (*old*) en el estudio del Instituto Allen [26].

La Fig. 2.2 es una reconstrucción de la figura 3a de la referencia [22]. En esta se muestran los resultados del PCA para los datos de expresión genética de la materia blanca del cerebro del Instituto Allen. La primera componente principal (PC1), la cual contiene el 24,7 % de la

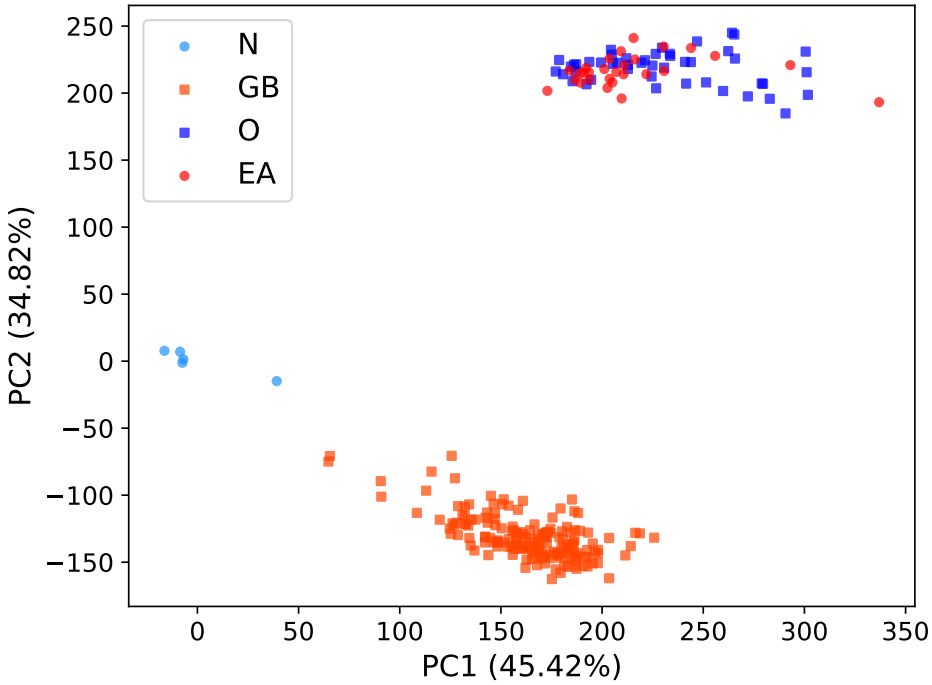



Figura 2.1: Análisis de componentes principales para los datos de expresión genética.

varianza total, discrimina entre las muestra de O y EA. La posición del centro de la nube de O en este eje es $\langle x_1 \rangle = 0$, y para la EA es $\langle x_1 \rangle = 40{,}97$. Sin embargo, los radios de las nubes de las muestras de O y EA son más grandes que la distancia entre los centros, que son 80,69 y 72,64 respectivamente.

Es bien conocido el papel de la edad en la EA, especialmente en ancianos [29]. Por lo tanto, podemos usar la edad como una variable de tiempo para seguir la transición. A pesar del número relativamente pequeño de muestras, se realizó un análisis de regresión lineal de la posición media de $\langle x_1 \rangle$ en función de la edad en las muestras de O, Fig. 2.3. Esta figura es una reproducción de la 4a de la referencia [22] y en la cual se ilustra que $\langle x_1 \rangle = -287{,}12 + 3{,}24 \cdot edad$. En las muestras de la EA, sin embargo, no se encontró correlación entre $\langle x_1 \rangle$ y la edad observada. Por lo tanto, la posición de la zona EA es aproximadamente fija, y la nube de muestras de O evidencia una deriva hacia el mínimo de la EA a medida que aumenta la edad.

Una mejor ilustración de este hecho viene representada en la Fig. 2.4, donde se compara la densidad de probabilidad de las muestras de O y de la EA. Esta reproduce la imagen S2 del material suplementario de la referencia [22]. Para construir dicha figura, se definieron

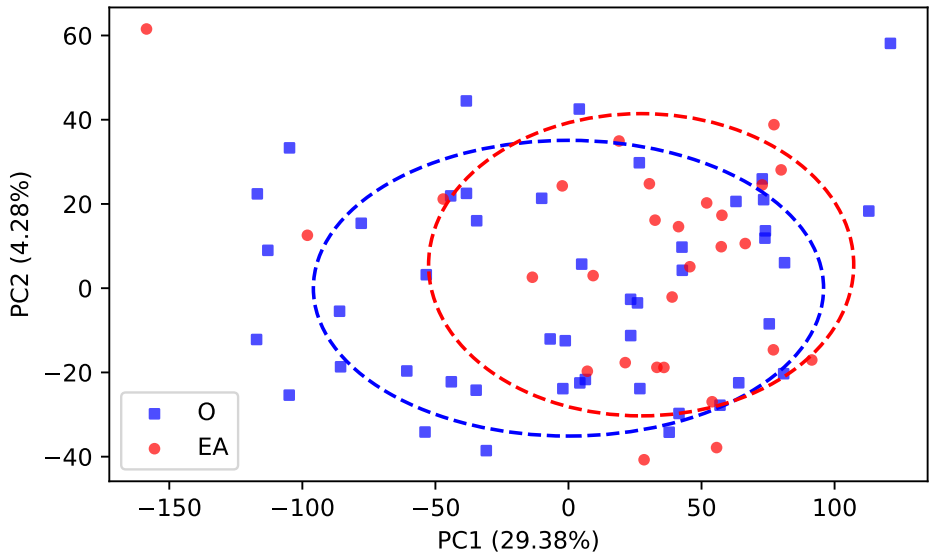



Figura 2.2: Análisis de componentes principales de los datos de expresión genética del Instituto Allen de la materia blanca del cerebro. Se representan las muestras de O y la EA. Las elipses discontinuas se dibujan de acuerdo con las desviaciones estándar de cada conjunto. Tomada de la referencia [22].

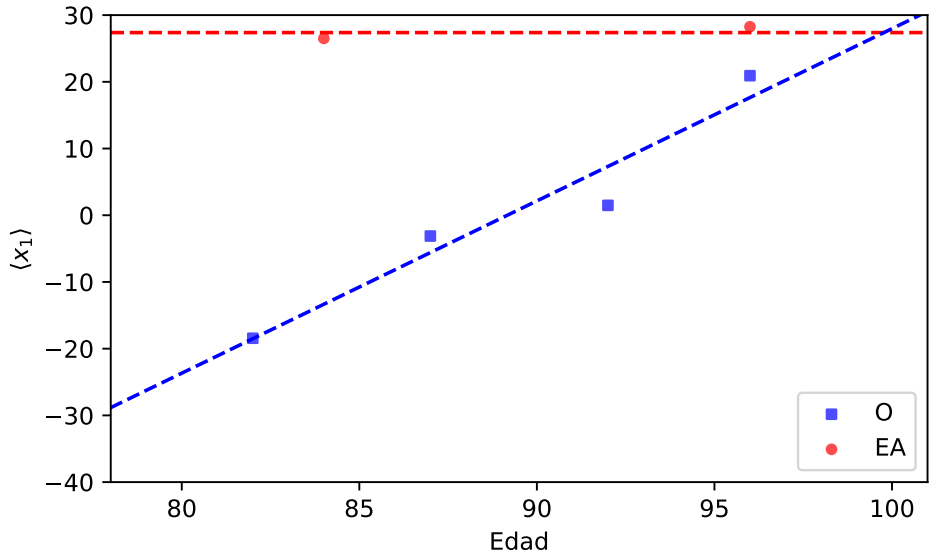



Figura 2.3: Posición media de la muestra a lo largo del eje PC1 en función de la edad. A medida que aumenta la edad, las muestras de O experimentan una deriva hacia la región de la EA, cuyo centro es aproximadamente independiente de la edad. Tomada de la referencia [22].

cuatro intervalos de edades, que contienen aproximadamente la misma cantidad de muestras de O: [77, 84], [84, 90], [90, 95], [95, 100+]. La probabilidad total de las muestras de la EA es mostrada en los cuatro paneles. El solapamiento creciente entre las densidades de probabilidad evidencia una transición gradual de las muestras del grupo O hacia valores característicos del grupo EA.

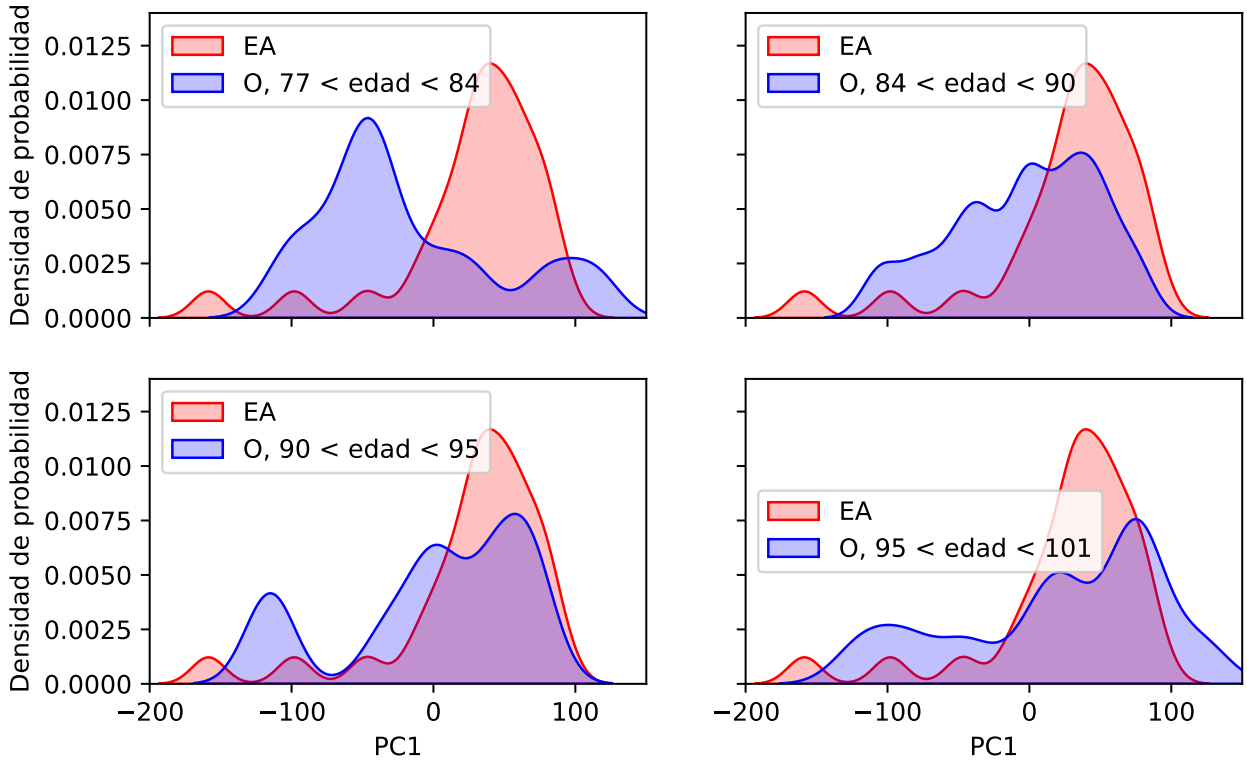



Figura 2.4: Densidad de probabilidad de las muestras de O y de la EA a lo largo del eje de PC1. Cada panel es para un intervalo de edad para las muestras de O. La probabilidad de la EA, la cual es aproximadamente independiente de la edad, es mostrada en los cuatro paneles. Tomada de la referencia [22].

Esta propiedad sugiere que el centro de la nube de muestras de la EA define un atractor en el espacio de expresión genética. Las muestras de O parecen ser atrapadas por el atractor de la EA en el proceso del envejecimiento.

Así, en nuestra aproximación, obtenemos un panorama en el espacio de expresión genética de tres atractores: N, GB y EA, y un conjunto de muestras de O que se desplaza hacia la EA. Las posiciones relativas y las principales transiciones entre los atractores se resumen en

la Fig. 2.5. Asumimos que estas transiciones están determinadas por la biología subyacente a los procesos en los tejidos. La transición de N a la EA se denomina "EA anticipada" para enfatizar que también existe una vía hacia la EA a través del envejecimiento: "EA tardía". La figura también indica una vía para el GB y para el envejecimiento.

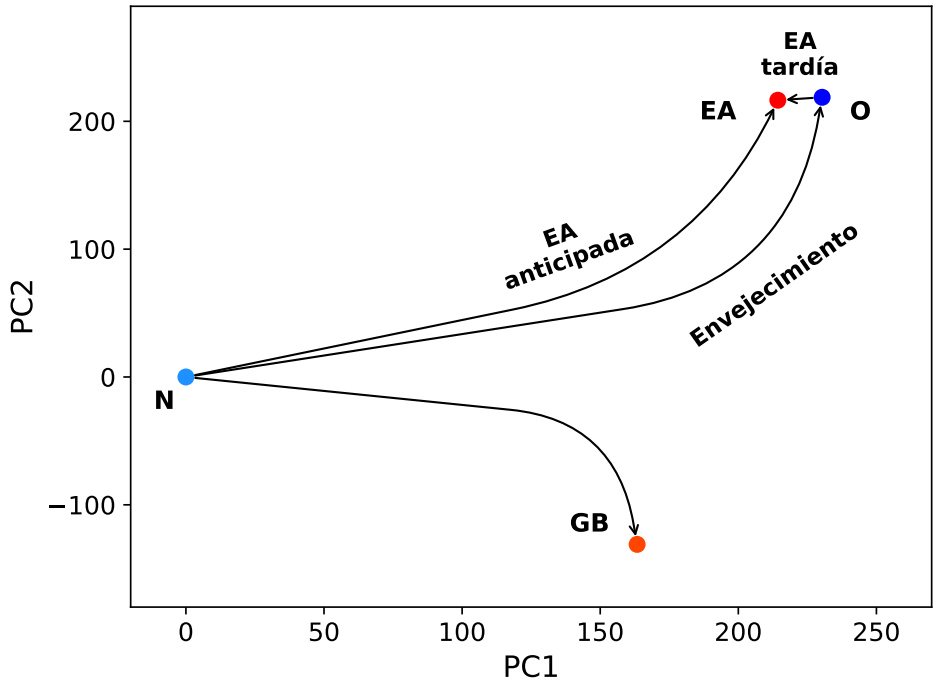



Figura 2.5: Posiciones relativas y principales transiciones entre los atractores.

## 2.2 Panorama del fitness

Existe información cualitativa que puede introducirse en nuestra descripción. Esta se relaciona con una variable de *fitness*, de modo que dibujamos una especie de diagrama de Wright [1]. En la Fig. 2.6 se muestra un diagrama esquemático que contiene un gráfico de contorno hipotético del *fitness*. Los atractores N y GB son máximos de *fitness* y deberían estar separados por una barrera de bajo *fitness* [22]. El GB debería ser el máximo más alto de los tres actores representados [22,30]. Por otro lado, la transición de O a la EA es casi continua, con un número relativamente pequeño de genes expresados diferencialmente [22]. Esto significa que existe una barrera muy pequeña, o incluso una ruta sin barrera, que conecta a O y a la EA. Esperamos

una barrera de bajo *fitness* que impida las transiciones directas de O a la EA, y un máximo de la EA pequeño, ya que este atractor se encuentra en la región de bajo *fitness*, lejos de N.

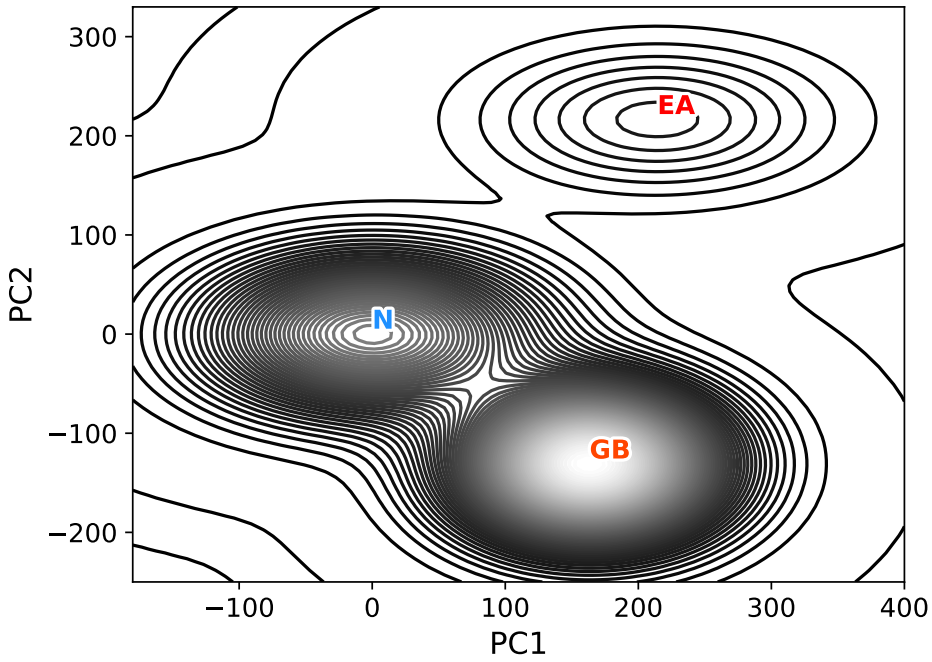



Figura 2.6: Diagrama de Wright que muestra un gráfico de contorno hipotético del *fitness*. El máximo absoluto corresponde con el estado de GB. El atractor de la EA se representa como un ligero máximo local.

Todos estos hechos se representan en la Fig. 2.6. El esquema se construye a partir de una suma de gaussianas centradas en los atractores, con desviaciones estándar proporcionales a los valores reales observados en la Fig. 2.1 y con alturas que respetan cualitativamente la fuerza relativa de los atractores.

Resaltemos el significado de un diagrama de Wright en el tejido cerebral. En otros tejidos, la evolución somática se relaciona principalmente con la replicación de células madre. Sin embargo, en su estado normal, el cerebro es un tejido de replicación muy lenta [31]. Los cambios en pequeñas regiones cerebrales, es decir, los desplazamientos en el diagrama, son básicamente daños acumulados, es decir, envejecimiento [32]. Sin embargo, una vez que se produce la transición al estado GB, se produce un enorme aumento de la tasa de replicación de las células tumorales. Observemos, además, que los cambios relacionados con el envejecimiento son muy evidentes en la sustancia blanca [33].

## 2.3 Limitaciones

En este trabajo se utilizaron datos de expresión genética, en formato FPKM, de las referencias [23, 26]. Estos fueron obtenidos usando diferentes plataformas. Nosotros tomamos aproximadamente 30 000 genes que están perfectamente identificados en ambas plataformas y realizamos un sencillo análisis de componentes principales [19], como se definió en la sección 1.1. Para definir los valores de la expresión diferencial logarítmica y calcular la matriz de covarianza utilizada para el PCA, se utilizó como referencia común la media geométrica en el conjunto de muestras N.

Debido al uso de estos datos, proveniente de dos experimentos distintos, para realizar un solo calculo de PCA surgen problemas tanto técnicos como conceptuales. Por ejemplo, la referencia N corresponde precisamente al estado normal del cerebro, sino que son un conjunto de muestras patológicamente normales que fueron tomadas de individuos con GB. Además, dos de los pacientes tienen más de 70 años. Desde el punto de vista computacional, por otro lado, se podrían utilizar correcciones por lotes [34, 35], que corrigen parcialmente los sesgos asociados a cada grupo de muestras, pero también pueden introducir problemas incontrolados.

En lugar de introducir procedimientos muy avanzados, preferimos extraer los datos directamente de las fuentes y utilizar la técnica de PCA más sencilla. No creemos que ninguna corrección altere sustancialmente el análisis cualitativo derivado del diagrama de tres atractores que se muestras en la Fig. 2.1.

La situación ideal podría ser repetir el experimento dentro de un único marco tecnológico, incluyendo datos de personas jóvenes sanas, que se utilizarían para establecer la referencia de los cálculos de la expresión genética diferencial, incluyendo datos de pacientes con GB y la EA, y datos de pacientes sanos en diferentes rangos de edad. Este es un experimento complejo, pero podría ser particularmente factible en un modelo de ratones [36], por ejemplo. Consideramos nuestro diagrama de la Fig. 2.1 como una aproximación cualitativa de este experimento ideal.

# Capítulo 3

# Resultados

## 3.1 Principales resultados

Sobre la base de nuestros diagramas, podemos formular las siguientes observaciones o afirmaciones, que son los principales resultados del trabajo.

1. **Existe una dirección en el espacio de expresión genética, que a grandes rasgos se puede identificar con el eje PC1, asociada al envejecimiento y a un aumento del riesgo de padecer GB y la EA.**

De hecho, el desplazamiento en esta dirección implica escalar parcialmente las barreras de bajo *fitness* que separan N de los estados GB y EA, y por lo tanto aumentar el riesgo tanto para GB como para la EA.

Vale la pena observar los principales genes involucrados en este proceso. Para ello, observamos el vector unitario a lo largo del eje PC1. Los genes se clasifican según su contribución al vector unitario.

En la Tabla B.1 de los apéndices se encuentra una lista con los 100 primeros genes del *ranking*. En la Fig. 3.1 se representan los 30 primeros genes ordenados por su valor absoluto. Las amplitudes positivas definen genes cuya expresión aumenta con el desplazamiento a lo largo de la dirección positiva de PC1, mientras que las amplitudes negativas se refieren a genes silenciados. Estos genes deberían desempeñar simultáneamente un papel crucial en el envejecimiento, el GB y la EA.

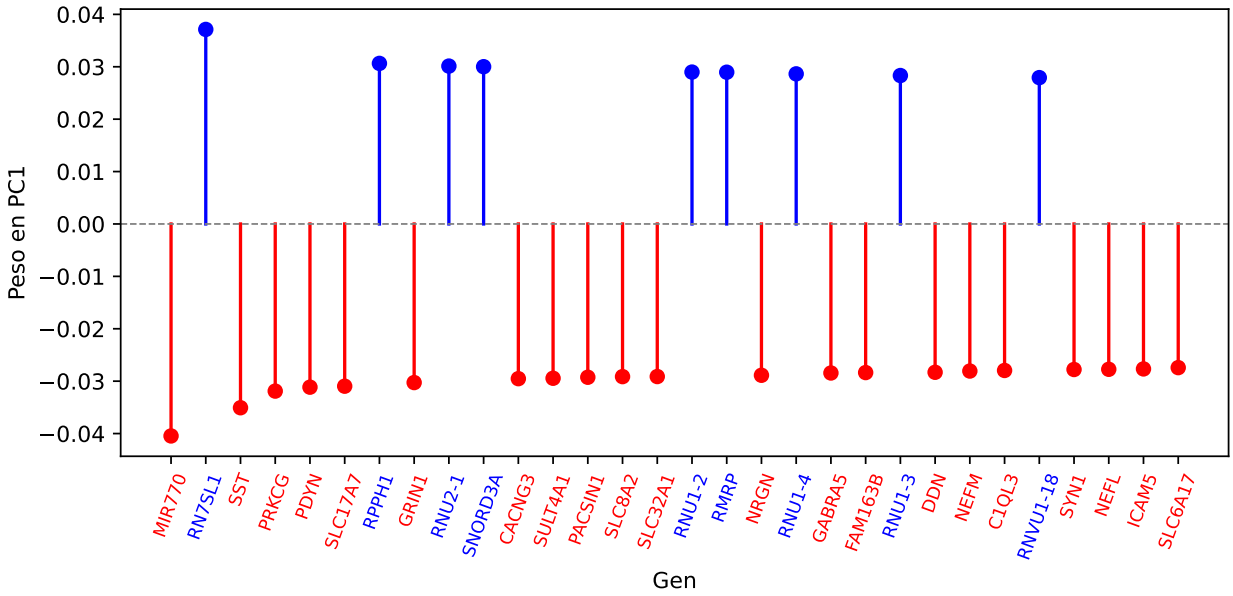



Figura 3.1: Genes con mayor peso a lo largo del vector PC1.

Por supuesto, debido al valor puramente cualitativo de nuestro análisis, los genes, y especialmente el *ranking*, deben considerarse con cautela. Sin embargo, cabe destacar que 20 de los genes silenciados están relacionados con los procesos principales de transmisión a través de sinapsis químicas. En la Tabla B.2 del Apéndice B se enumeran los procesos principales del *Reactome* (https://reactome.org/) asociadas a estos genes [37]. Este conjunto incluye 56 genes anotados.

La disminución de la función sináptica es una característica conocida del cerebro envejecido, según la revisión [38]. La segunda característica principal, según esta referencia, es un aumento de la función inmunitaria, que no es particularmente evidente en nuestro conjunto de genes. En cambio, observamos genes relacionados con la neurotoxicidad de las toxinas de clostridium [39], con la disminución de la actividad mitocondrial [40], micro ARN compartidos entre la EA y el GB [41], etc.

2. **Hay una dirección en el espacio de expresión genética, que puede identificarse aproximadamente con el eje PC2, que muestra que la EA y GB son alternativas excluyentes.**

De hecho, la EA y el GB aparecen en semiplanos opuestos. La evidencia clínica [3–6] y los

estudios de biología molecular [7, 8] respaldan esta disyunción. En consecuencia, el eje PC2 involucra genes con desregulación inversa en la EA y el GB.

En la Tabla C.1 de los apéndices, se listan los 100 genes principales definidos por el vector unitario a lo largo del eje PC2. En la Fig. 3.2 se muestran los 30 primeros genes ordenados por su valor absoluto. Los pesos positivos corresponden a genes cuya expresión aumenta en la transición de N a la EA. Por otro lado, las amplitudes negativas corresponden a genes cuya expresión aumenta en la transición de N a GB.

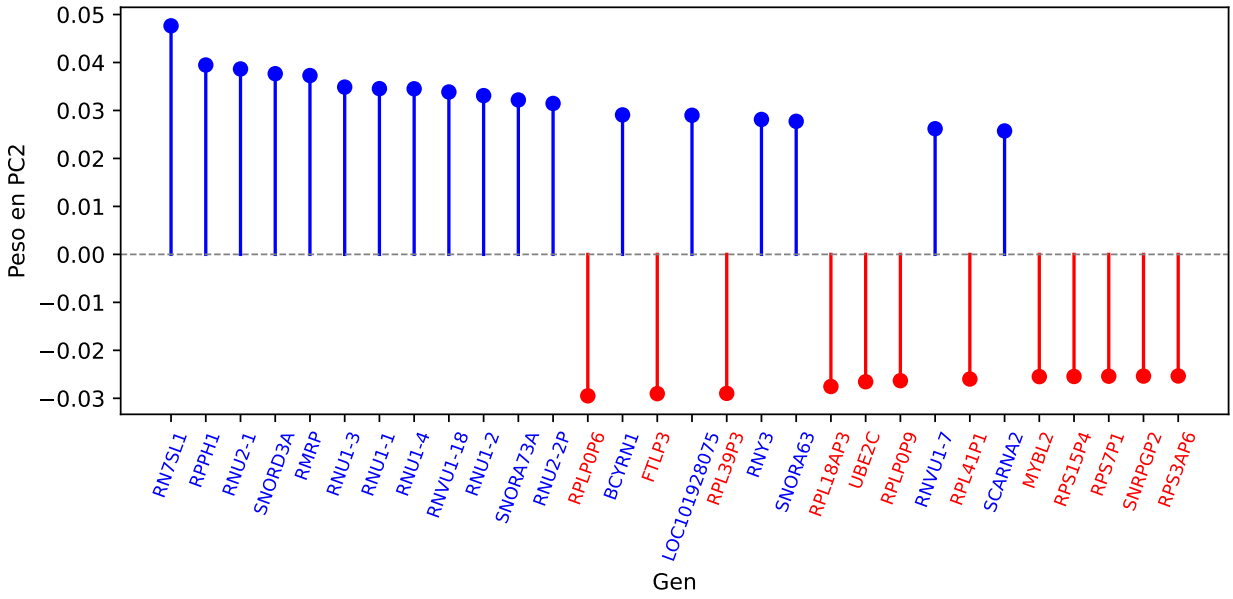



Figura 3.2: Genes con mayor peso a lo largo del vector PC2.

Los procesos de *Reactome* relacionadas con estos genes se muestran en la Tabla C.2 del Apéndice C. Se relacionan principalmente con el control del ciclo celular, la replicación del ADN, la apoptosis, la modificación de la matriz extracelular, etc., es decir, con las características distintivas del cáncer [42–44].

Anteriormente, mencionamos MMP9 como un ejemplo de genes que desempeñan funciones opuestas en el GB y la EA. El gen codificador de la proteína UBE2C es otro gen conocido con esta característica [45, 46]. Por otro lado, el gen BCYRN1, también entre los primeros 100 genes del *ranking*, parece estar subexpresado en GB [47] y sobreexpresado en la EA [48]. La Fig. 3.3 muestra gráficos de violín para la expresión diferencial de los genes MMP9 y BCYRN1

en muestras de N, GB y la EA.

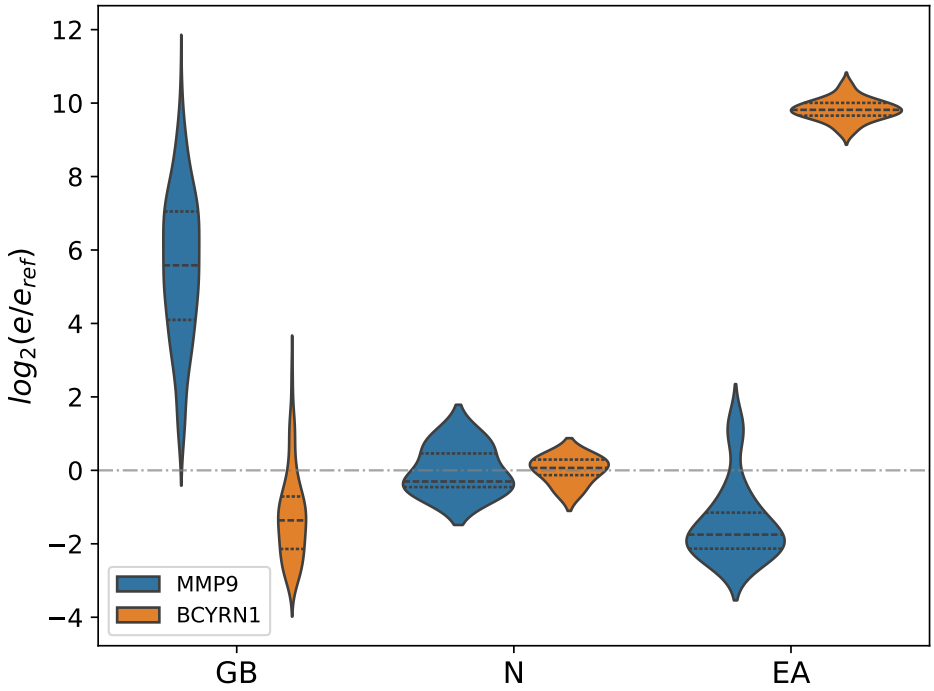



Figura 3.3: Gráficos de violín para la expresión diferencial logarítmica de los genes MMP9 y BCYRN1 en los estados N, GB y EA.

Noté también que en la Tabla C.1 hay numerosos genes codificadores de proteínas ribosomales, de núcleo pequeño, de microARN y de otros tipos, regulados inversamente en ambos procesos. Solo 18 de este conjunto de genes están anotados en los procesos *Reactome*. Este es un problema habitual en el análisis de procesos, donde las funciones biológicas de muchos genes no están anotadas.

3. **Existe un corredor de envejecimiento, es decir un camino preferencial para el envejecimiento en el espacio de expresión genética.**

En nuestros datos, existen muestras en la región N y muestras correspondientes a cerebros de edad normal, ubicadas en una región definida cerca del atractor de la EA. En otras palabras, el proceso de envejecimiento parece definir una trayectoria o corredor de decrecimiento continuo del *fitness*, del cual los datos O muestran el último segmento. Sin embargo, faltan muestras en la región intermedia.

En lugar de incluir muestras adicionales en nuestra figura, lo cual introduciría efectos de lote adicionales, utilizamos resultados recientes en un modelo de ratón [36] que muestran indudablemente un corredor continuo para el envejecimiento. Presentamos en la Fig. A.1 del Apéndice A una representación gráfica de sus datos para el cuerpo calloso, una región rica en materia blanca. En el panel izquierdo, se grafican las dos primeras componentes principales para los centros de los subgrupos de muestras. Se consideran edades de ratón entre 3 y 28 meses; este último equivale aproximadamente a 80 años en una escala humana. Un corredor para el envejecimiento es evidente. El panel derecho, por otro lado, muestra distancias reales incluyendo todos los componentes. Por lo tanto, las proyecciones en el plano (PC1, PC2) son una representación fiel de la distribución real de puntos.

En nuestro esquema (Fig. 2.5), se delinea un corredor de envejecimiento. La Fig. 2.6 sugiere que este corredor es una dirección con una mínima disminución del *fitness*.

Una dirección o corredor preferencial para el envejecimiento es consistente con la hipótesis del envejecimiento programado [10, 11], es decir, la idea de que el envejecimiento está programado en nuestros genes.

4. **El corredor de envejecimiento predeterminado podría estar relacionado con la presión de evitar el fuerte atractor GB.**

Una pregunta muy interesante se refiere a la selección de la dirección preferida para el envejecimiento. Nuestro esquema simplificado (Fig. 2.6) ofrece una respuesta inesperada: en la materia blanca, esta dirección podría estar relacionada con la presión para evitar el atractor GB más fuerte.

De hecho, para cada pequeña porción de tejido, se puede modelar el envejecimiento como una especie de movimiento aleatorio que comienza en la región N. Obsérvese que en la referencia [49] se utilizó un modelo de saltos aleatorios en el espacio de expresión genética para describir la evolución somática de diferentes tejidos hacia el cáncer. Primero, asumimos que la dirección de los saltos es aleatoria en el plano que se muestra en la Fig. 2.6.

Solo existen cuatro posibilidades para el destino de las trayectorias aleatorias en el plano que comienzan en la región N. Primero, la trayectoria permanece en la región N. Segundo, llega a una región de *fitness* muy baja y las células mueren. Este es el destino de muchas células en el cerebro envejecido. Tercero, es capturada por el atractor GB. Y, finalmente, la cuarta posibilidad es la captura por el atractor de la EA.

Debido al elevado valor de *fitness* del atractor GB, debería existir una probabilidad relativamente alta de que las trayectorias sean capturadas por el GB, lo que conduce a la iniciación de un tumor. Esto implica un enorme aumento del *fitness*, la propagación del tumor en el cerebro y una esperanza de vida para el individuo de tan solo unos dos años tras la iniciación [50]. Esto podría afectar a los individuos en edad reproductiva. Por lo tanto, evitar el atractor GB podría ser objeto de la presión selectiva.

Como comprobación indirecta, podemos comparar las incidencias de GB y la EA. Estas deberían ser proporcionales a las tasas de captura de trayectorias aleatorias por los atractores de GB y EA. Como se mencionó, en un modelo donde la dirección de los saltos es aleatoria, la incidencia de GB debería ser mucho mayor que la de la EA. Sin embargo, la incidencia global de glioblastoma es inferior a 10 por 100 000 personas [51], en contraste con el 5 % de la EA en personas de 65 a 74 años y el 13 % en personas de 75 a 84 años [52]. La incidencia observada sugiere que se evita el movimiento hacia el centro de GB.

5. **La aparición tardía de la EA podría ser el resultado de la captura por parte del atractor de la EA de microestados cerebrales envejecidos.**

La imagen es, por lo tanto, la siguiente. El proceso de envejecimiento se relaciona inicialmente con un desplazamiento a lo largo del corredor de envejecimiento, con la correspondiente disminución del *fitness*. En los últimos pasos, los estados O son capturados por el atractor débil de la EA.

Como se mencionó anteriormente, la afirmación sobre el corredor de envejecimiento está respaldada por el experimento en un modelo de ratón, mientras que la captura por parte del centro de la EA está sugerida por los cálculos de la referencia [22], particularmente por los resultados que se muestran en la Fig. 2.3, que es un reconstrucción de la Fig. 4 de esa referencia.

En la Tabla D.1 se muestran los 10 genes principales en la transición de O a la EA. Se trata de genes incluidos en la Tabla B.1, pero que varían en dirección opuesta, es decir, en la dirección negativa del eje PC1. Este hecho se representa en el diagrama esquemático de la Fig. 2.5.

## 3.2 Perspectiva cuantitativa

Debido al carácter cualitativo de nuestro estudio, el análisis de genes y procesos relevantes no se aborda adecuadamente en este artículo. Sin embargo, analicemos cualitativamente siete marcadores conocidos para la EA y el GB según nuestro esquema. Los gráficos de violín para estos genes se muestran en Fig. 3.4. La figura muestra que los genes MAPT (proteína tau [53]) y APP (beta amiloide [54]) están subexpresados tanto en el GB como en la EA y, por lo tanto, según nuestro esquema, son genes tipo PC1, principalmente relacionados con el envejecimiento. En cierto modo, esto concuerda con los hallazgos del estudio del Instituto Allen sobre los patrones de proteína tau y placas amiloides en el cerebro envejecido. Por supuesto, ciertas mutaciones de estos genes podrían conducir a un envejecimiento cerebral acelerado y a la aparición temprana de la EA. El gen APOE [55], por el contrario, está desregulado inversamente en el GB y la EA. Es un gen del tipo PC2.

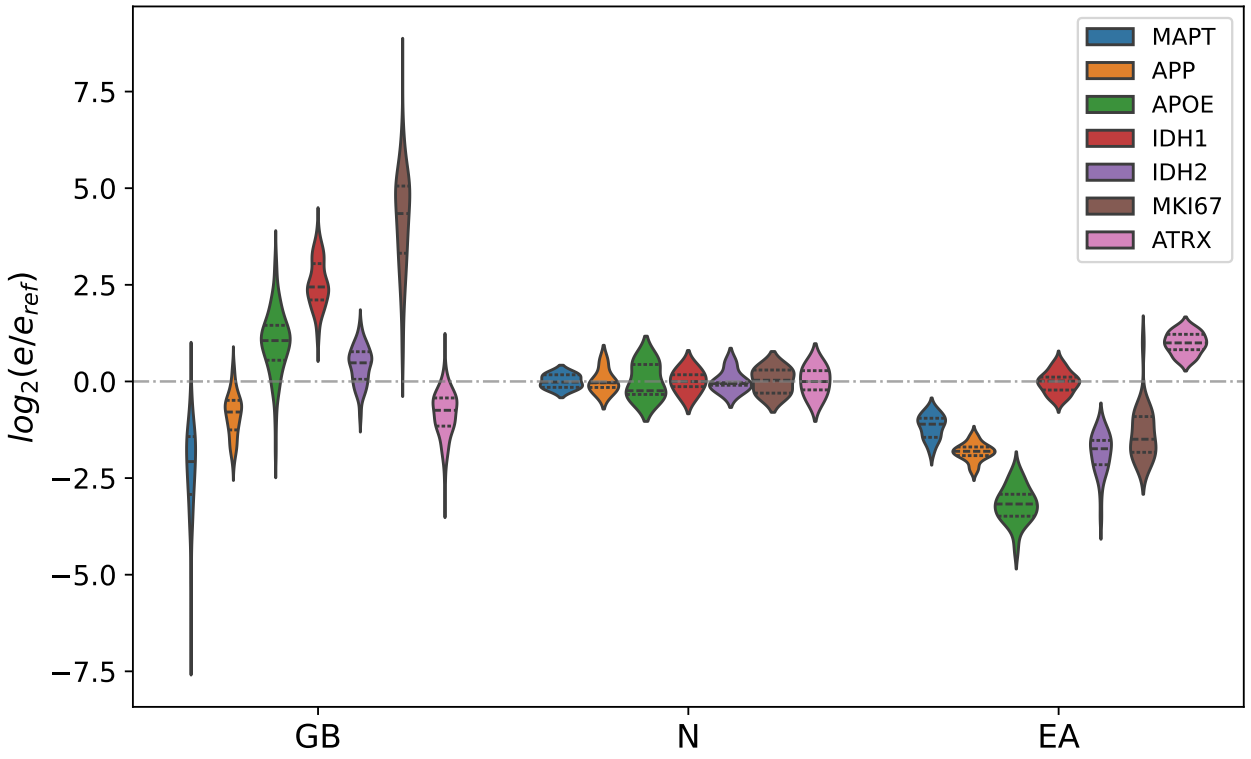



Figura 3.4: Gráficos de violín para marcadores conocidos en la EA y el GB.

Por otro lado, el marcador IDH1 [56] está sobreexpresado en la GB, pero es irrelevante en

la EA. Los tres marcadores de GB, IDH2 [56], MKI67 [57] y ATRX [58], son genes similares a PC2. Se han estudiado principalmente en relación con la GB, pero el hecho de que sean genes PC2 indica que podrían desempeñar un papel importante también en la EA.

Consideremos, como último ejemplo, la reciente demostración de la relevancia de TREM2 en la EA [59]. Se ha demostrado que la activación de TREM2 en la EA mejora el metabolismo microglial. Según nuestro análisis, TREM2 es un gen PC2, subexpresado en la EA y sobreexpresado en el GB. Por lo tanto, esperamos que la inhibición de TREM2 en células de glioma pueda tener un efecto importante en el GB. Este hecho, según nuestro sencillo esquema, se encuentra actualmente en estudio [60].

# Conclusiones

Nuestros sencillos esquemas cualitativos identifican direcciones en el espacio de expresión genética asociadas a diferentes procesos biológicos: envejecimiento, carcinogénesis, aparición de la enfermedad de Alzheimer. Cada una de estas direcciones se caracteriza por un "metagén" o perfil de expresión genética, del cual se pueden extraer los principales genes que contribuyen al proceso.

Algunos de nuestros resultados confirman conocimientos previos, pero otros requieren mayor corroboración. Por ejemplo, la idea de que el envejecimiento programático podría estar relacionado con evitar el fuerte atractor GB, o la aparición tardía de la EA como la captura por el atractor de EA de muestras de edad normal. Esperamos que estos resultados motiven la investigación experimental en estas direcciones. El experimento en un modelo de ratones es particularmente factible, como lo demuestra la referencia [36].

Cabe destacar que incluso datos o métodos computacionales más refinados no podrían modificar sustancialmente nuestros esquemas cualitativos con solo tres atractores. Sus posiciones relativas podrían variar, pero las afirmaciones formuladas se mantendrán.

# Recomendaciones

El diagrama de la Fig. 2.1 debe completarse con datos correspondientes a otros tipos de demencia o trastornos cerebrales. En particular, se espera un área de la enfermedad de Parkinson cercana al atractor de la EA y opuesta al GB [61]. El panorama completo puede revelar una topología aún más precisa del espacio de expresión genética y un diagrama de Wright más completo.

Anticipamos que diagramas similares en otros tejidos, además de proporcionar una perspectiva integral, podrían ser útiles en la comprensión de la biología de enfermedades o trastornos aparentemente no relacionados, y en el descubrimiento de pistas inesperadas para su tratamiento.

# Apéndice A

# Modelo de ratón para el envejecimiento

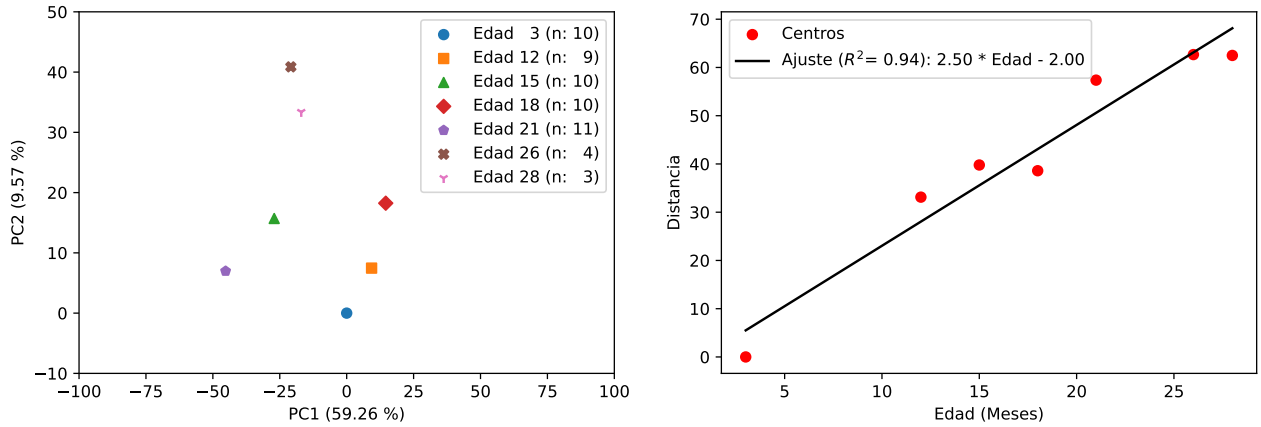



Figura A.1: Representación gráfica de los datos de la referencia [36] en un modelo de ratón para el cuerpo calloso, una región rica en materia blanca.

Panel izquierdo: Análisis de componentes principales de los datos. Se muestran los centros de los subgrupos de muestras. Se consideran edades entre 3 y 28 meses. Se observa una dirección de envejecimiento.

Panel derecho: Distancias completas (incluyendo todos los componentes) hasta el punto inicial (3 meses). Esta figura muestra que la proyección al plano (PC1, PC2) es una representación fiel.

# Apéndice B

# Eje PC1 (envejecimiento)

Cuadro B.1: Los 100 principales genes que contribuyen al vector unitario a lo largo de PC1.

| Gen | PC1 | Gen | PC1 |
|---|---|---|---|
| RN7SL1 | 0.04 | WBSCR17 | -0.03 |
| RPPH1 | 0.03 | SV2B | -0.03 |
| RNU2-1 | 0.03 | SLC30A3 | -0.03 |
| SNORD3A | 0.03 | ATP1A3 | -0.03 |
| RNU1-2 | 0.03 | CPNE6 | -0.03 |
| RMRP | 0.03 | CAMK2A | -0.03 |
| RNU1-4 | 0.03 | SLC6A7 | -0.03 |
| RNU1-3 | 0.03 | GABRA1 | -0.03 |
| RNVU1-18 | 0.03 | KCNJ4 | -0.03 |
| RNU1-1 | 0.03 | CDK5R2 | -0.03 |
| SNORA73A | 0.02 | CALY | -0.03 |
| RNVU1-7 | 0.02 | GNG3 | -0.03 |
| RNU2-2P | 0.02 | CALB2 | -0.03 |
| WARS2-IT1 | 0.02 | CHD5 | -0.03 |
| VGF | -0.02 | VSNL1 | -0.03 |
| SLC12A5 | -0.02 | CRYM | -0.03 |
| HIPK4 | -0.02 | CCKBR | -0.03 |
| SSTR1 | -0.02 | NEUROD6 | -0.03 |
| CPLX1 | -0.02 | HRH3 | -0.03 |
| SYT5 | -0.02 | SYT1 | -0.03 |
| MIR657 | -0.02 | SNCB | -0.03 |
| INA | -0.02 | CHRM1 | -0.03 |

Cuadro B.1: Los 100 principales genes que contribuyen al vector unitario a lo largo de PC1.

| Gen | PC1 | Gen | PC1 |
|---|---|---|---|
| CCK | -0.02 | MAL2 | -0.03 |
| GABRD | -0.02 | SNORD113-3 | -0.03 |
| MIR1249 | -0.02 | PNMA5 | -0.03 |
| WIF1 | -0.02 | CPLX2 | -0.03 |
| NECAB2 | -0.02 | TMEM130 | -0.03 |
| FBXL16 | -0.02 | LOC105373377 | -0.03 |
| GDA | -0.02 | SYN2 | -0.03 |
| KIAA1045 | -0.02 | SLC6A17 | -0.03 |
| SYNPR | -0.02 | ICAM5 | -0.03 |
| HMP19 | -0.02 | NEFL | -0.03 |
| CHGA | -0.02 | SYN1 | -0.03 |
| SNCG | -0.02 | C1QL3 | -0.03 |
| SVOP | -0.02 | NEFM | -0.03 |
| LRTM2 | -0.02 | DDN | -0.03 |
| WNT10B | -0.02 | FAM163B | -0.03 |
| MTND4P12 | -0.02 | GABRA5 | -0.03 |
| PHYHIP | -0.02 | NRGN | -0.03 |
| RAB3A | -0.02 | SLC32A1 | -0.03 |
| RASAL1 | -0.02 | SLC8A2 | -0.03 |
| HPCA | -0.03 | PACSIN1 | -0.03 |
| SERTM1 | -0.03 | SULT4A1 | -0.03 |
| GABRG2 | -0.03 | CACNG3 | -0.03 |
| CAMKV | -0.03 | GRIN1 | -0.03 |
| CREG2 | -0.03 | SLC17A7 | -0.03 |
| SNAP25 | -0.03 | PDYN | -0.03 |
| GAD2 | -0.03 | PRKCG | -0.03 |
| SYT13 | -0.03 | SST | -0.04 |
| KLK7 | -0.03 | MIR770 | -0.04 |

Cuadro B.2: Los principales procesos de *Reactome* relacionados con los 100 primeros genes en el *ranking* a lo largo de la dirección PC1

| Proceso | Entidades | | | | Reacciones | |
|---|---|---|---|---|---|---|
| | encontrado | tasa | *p-value* | FDR | encontrado | tasa |
| Transmisión a través de sinapsis químicas | 20/344 | 0.022 | 3.72e-13 | 9.16e-11 | 63/167 | 0.012 |
| Sistema neuronal | 20/490 | 0.032 | 1.97e-10 | 2.42e-08 | 72/221 | 0.015 |

# Apéndice C

# Eje PC2 (GB vs. EA)

Cuadro C.1: Los 100 principales genes que contribuyen al vector unitario a lo largo de PC2.

| Gen | PC1 | Gen | PC1 |
|---|---|---|---|
| RN7SL1 | 0.05 | RPL37AP1 | -0.02 |
| RPPH1 | 0.04 | AURKB | -0.02 |
| RNU2-1 | 0.04 | FAUP1 | -0.02 |
| SNORD3A | 0.04 | MMP9 | -0.02 |
| RMRP | 0.04 | FTLP2 | -0.02 |
| RNU1-3 | 0.03 | RPL7P9 | -0.02 |
| RNU1-1 | 0.03 | GPX1P1 | -0.02 |
| RNU1-4 | 0.03 | CHI3L1 | -0.02 |
| RNVU1-18 | 0.03 | MIR621 | -0.02 |
| RNU1-2 | 0.03 | PI3 | -0.02 |
| SNORA73A | 0.03 | RPL6P27 | -0.02 |
| RNU2-2P | 0.03 | IGFBP2 | -0.02 |
| BCYRN1 | 0.03 | BIRC5 | -0.02 |
| LOC101928075 | 0.03 | TPI1P1 | -0.02 |
| RNY3 | 0.03 | FTH1P8 | -0.02 |
| SNORA63 | 0.03 | RPS27AP16 | -0.02 |
| RNVU1-7 | 0.03 | RPS27AP5 | -0.02 |
| SCARNA2 | 0.03 | LOC105369550 | -0.02 |
| SNORA48 | 0.02 | PLA2G2A | -0.02 |
| CYCSP30 | 0.02 | NDUFA4P1 | -0.02 |
| SCARNA10 | 0.02 | FTH1P2 | -0.02 |
| RN7SL2 | 0.02 | TMSB4XP8 | -0.02 |

Cuadro C.1: Los 100 principales genes que contribuyen al vector unitario a lo largo de PC2.

| Gen | PC1 | Gen | PC1 |
|---|---|---|---|
| LOC729348 | 0.02 | RPL26P19 | -0.02 |
| SCARNA5 | 0.02 | COL1A1 | -0.02 |
| SCARNA7 | 0.02 | COX6A1P2 | -0.02 |
| SNORA81 | 0.02 | COL3A1 | -0.02 |
| SNORD97 | 0.02 | RPS27P4 | -0.02 |
| DNAJC19P8 | 0.02 | SAA1 | -0.02 |
| FCF1P1 | 0.02 | RPL14P1 | -0.02 |
| SNORA57 | 0.02 | RPS27P3 | -0.02 |
| SNORD10 | 0.02 | SNRPGP10 | -0.02 |
| SNORA49 | 0.02 | MT2P1 | -0.02 |
| SNORA54 | 0.02 | RPS3AP26 | -0.02 |
| RAB9AP2 | 0.02 | PBK | -0.02 |
| RAB9AP5 | 0.02 | TOP2A | -0.02 |
| RNY1 | 0.02 | CPXM1 | -0.02 |
| TPT1P9 | -0.02 | MIR3682 | -0.02 |
| RPL31P4 | -0.02 | RPS18P12 | -0.03 |
| RPL10P9 | -0.02 | RPS3AP6 | -0.03 |
| GAPDHP65 | -0.02 | SNRPGP2 | -0.03 |
| RPL12P4 | -0.02 | RPS7P1 | -0.03 |
| RPS20P14 | -0.02 | RPS15P4 | -0.03 |
| FAM64A | -0.02 | MYBL2 | -0.03 |
| RPS2P46 | -0.02 | RPL41P1 | -0.03 |
| RPS3AP5 | -0.02 | RPLP0P9 | -0.03 |
| RPL7P1 | -0.02 | UBE2C | -0.03 |
| YBX1P1 | -0.02 | RPL18AP3 | -0.03 |
| FTH1P7 | -0.02 | RPL39P3 | -0.03 |
| RPL35P5 | -0.02 | FTLP3 | -0.03 |
| RPL13AP25 | -0.02 | RPLP0P6 | -0.03 |

Cuadro C.2: Los principales procesos de *Reactome* relacionados con los 100 primeros genes en el *ranking* a lo largo de la dirección PC2

| Proceso | Entidades | | | | Reacciones | |
|---|---|---|---|---|---|---|
| | encontrado | tasa | *p-value* | FDR | encontrado | tasa |
| G0 y G1 temprano | 4/38 | 0.002 | 1.58e-04 | 0.032 | 4/27 | 0.002 |
| TFAP2A actúa como un represor transcripcional durante la diferenciación celular inducida por ácido retinoico | 20/344 | 0.022 | 3.72e-13 | 9.16e-11 | 63/167 | 0.012 |
| Señalización de interleucina-4 e interleucina-13 | 6/211 | 0.014 | 0.004 | 0.254 | 2/47 | 0.003 |
| SUMOilación de proteínas de replicación del ADN | 3/50 | 0.003 | 0.005 | 0.272 | 4/8 | 5.52e-04 |
| Transcripción de dianas E2F bajo control negativo por p107 (RBL1) y p130 (RBL2) en complejo con HDAC1 | 2/20 | 0.001 | 0.009 | 0.336 | 2/8 | 5.52e-04 |
| Ensamblaje de fibrillas de colágeno y otras estructuras multiméricas | 3/67 | 0.004 | 0.012 | 0.336 | 16/26 | 0.002 |
| Degradación del colágeno | 3/69 | 0.004 | 0.013 | 0.336 | 16/34 | 0.002 |
| Transcripción de dianas E2F bajo control negativo por el complejo DREAM | 2/25 | 0.002 | 0.013 | 0.336 | 2/12 | 8.29e-04 |
| TP53 regula la transcripción de varios genes de muerte celular adicionales cuyas funciones específicas en la apoptosis dependiente de p53 siguen siendo inciertas | 2/28 | 0.002 | 0.017 | 0.355 | 2/19 | 0.001 |

# Apéndice D

# Transición de O a la EA

Cuadro D.1: Los 10 primeros genes en el *ranking* de PCA para la transición de O a la EA.

| Gen | Nombre | PC1 |
|---|---|---|
| SNAP25 | *Synaptosome Associated Protein 25* | 1.00 |
| VSNL1 | *Visinin Like 1* | 0.90 |
| STMN2 | *Stathmin 2* | 0.90 |
| ENC1 | *Ectodermal-Neural Cortex 1* | 0.89 |
| NEFL | *Neurofilament Light Chain* | 0.89 |
| SYT1 | *Synaptotagmin 1* | 0.88 |
| RGS4 | *Regulator Of G Protein Signaling 4* | 0.86 |
| CHN1 | *Chimerin 1* | 0.84 |
| GABRA1 | *Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha1* | 0.77 |
| GABRG2 | *Gamma-Aminobutyric Acid Type A Receptor Subunit Gamma2* | 0.77 |

Los resultados se basan en cálculos de la referencia [22]. Los pesos en el vector unitario a lo largo de PC1 se normalizan al valor más alto.

# Bibliografía


[1] S. Wright, "The roles of mutation, inbreeding, crossbreeding and selection in evolution," in *Proceedings of the sixth international congress of Genetics*, vol. 1, pp. 356–366, 1932.

[2] M. J. Casey, P. S. Stumpf, and B. D. MacArthur, "Theory of cell fate," *WIREs Systems Biology and Medicine*, vol. 12, p. e1471, Dec. 2019.

[3] S.-M. Ou, Y.-J. Lee, Y.-W. Hu, C.-J. Liu, T.-J. Chen, J.-L. Fuh, and S.-J. Wang, "Does Alzheimer's Disease Protect against Cancers? A Nationwide Population-Based Study," *Neuroepidemiology*, vol. 40, pp. 42–49, Oct. 2012.

[4] C. M. Roe, A. L. Fitzpatrick, C. Xiong, W. Sieh, L. Kuller, J. P. Miller, M. M. Williams, R. Kopan, M. I. Behrens, and J. C. Morris, "Cancer linked to Alzheimer disease but not vascular dementia," *Neurology*, vol. 74, pp. 106–112, Jan. 2010.

[5] J. A. Driver, A. Beiser, R. Au, B. E. Kreger, G. L. Splansky, T. Kurth, D. P. Kiel, K. P. Lu, S. Seshadri, and P. A. Wolf, "Inverse association between cancer and Alzheimer's disease: results from the Framingham Heart Study," *BMJ*, vol. 344, pp. e1442–e1442, Mar. 2012.

[6] M. Musicco, F. Adorni, S. Di Santo, F. Prinelli, C. Pettenati, C. Caltagirone, K. Palmer, and A. Russo, "Inverse occurrence of cancer and Alzheimer disease: A population-based incidence study," *Neurology*, vol. 81, pp. 322–328, July 2013.

[7] T. Liu, D. Ren, X. Zhu, Z. Yin, G. Jin, Z. Zhao, D. Robinson, X. Li, K. Wong, K. Cui, H. Zhao, and S. T. C. Wong, "Transcriptional signaling pathways inversely regulated in Alzheimer's disease and glioblastoma multiform," *Scientific Reports*, vol. 3, Dec. 2013.

[8] C. Lanni, M. Masi, M. Racchi, and S. Govoni, "Cancer and Alzheimer's disease inverse relationship: an age-associated diverging derailment of shared pathways," *Molecular Psychiatry*, vol. 26, pp. 280–295, May 2020.

[9] J. A. Driver and K. Ping Lu, "Pin1: A New Genetic Link between Alzheimers Disease, Cancer and Aging," *Current Aging Sciencee*, vol. 3, pp. 158–165, Dec. 2010.

[10] J. P. Magalhães, "Programmatic features of aging originating in development: aging mechanisms beyond molecular damage?," *The FASEB Journal*, vol. 26, pp. 4821–4826, Sept. 2012.

[11] D. Gems, "The hyperfunction theory: An emerging paradigm for the biology of aging," *Ageing Research Reviews*, vol. 74, p. 101557, Feb. 2022.

[12] G. Choe, J. K. Park, L. Jouben-Steele, T. J. Kremen, L. M. Liau, H. V. Vinters, T. F. Cloughesy, and P. S. Mischel, "Active Matrix Metalloproteinase 9 Expression Is Associated with Primary Glioblastoma Subtype1," *Clinical Cancer Research*, vol. 8, pp. 2894–2901, 09 2002.

[13] Q. Xue, L. Cao, X.-Y. Chen, J. Zhao, L. Gao, S.-Z. Li, and Z. Fei, "High expression of MMP9 in glioma affects cell proliferation and is associated with patient survival rates," *Oncology Letters*, vol. 13, pp. 1325–1330, Jan. 2017.

[14] A. Kaminari, N. Giannakas, A. Tzinia, and E. C. Tsilibary, "Overexpression of matrix metalloproteinase-9 (MMP-9) rescues insulin-mediated impairment in the 5XFAD model of Alzheimer's disease," *Scientific Reports*, vol. 7, Apr. 2017.

[15] L. van der Maaten and G. Hinton, "Viualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 11 2008.

[16] S. Jung, T. Dagobert, J.-M. Morel, and G. Facciolo, "A Review of t-SNE," *Image Processing On Line*, vol. 14, pp. 250–270, Oct. 2024. `https://doi.org/10.5201/ipol.2024.528`.

[17] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," 2018.

[18] E. Becht, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, "Evaluation of UMAP as an alternative to t-SNE for single-cell data," Apr. 2018.

[19] J. Lever, M. Krzywinski, and N. Altman, "Principal component analysis," *Nature Methods*, vol. 14, pp. 641–642, 7 2017.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[21] A. Gonzalez, D. A. Leon, Y. Perera, and R. Perez, "On the gene expression landscape of cancer," *PLOS ONE*, vol. 18, p. e0277786, Feb. 2023.

[22] A. Gonzalez, J. Nieves, D. A. Leon, M. L. Bringas Vega, and P. V. Sosa, "Gene expression rearrangements denoting changes in the biological state," *Scientific Reports*, vol. 11, Apr. 2021.

[23] C. W. Brennan *et al.*, "The Somatic Genomic Landscape of Glioblastoma," *Cell*, vol. 155, pp. 462–477, Oct. 2013.

[24] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "Review The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge," *Contemp Oncol (Pozn)*, vol. 1A, pp. 68–77, 2015.

[25] B. Ellingson, A. Lai, R. Harris, J. Selfridge, W. Yong, K. Das, W. Pope, P. Nghiemphu, H. Vinters, and L. a. Liau, "Probabilistic radiographic atlas of glioblastoma phenotypes," *American Journal of neuroradiology*, vol. 34, no. 3, pp. 533–540, 2013.

[26] J. A. Miller *et al.*, "Neuropathological and transcriptomic characteristics of the aged brain," *eLife*, vol. 6, Nov. 2017.

[27] Y. Zhao, M.-C. Li, M. M. Konaté, L. Chen, B. Das, C. Karlovich, P. M. Williams, Y. A. Evrard, J. H. Doroshow, and L. M. McShane, "TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository," *Journal of Translational Medicine*, vol. 19, jun 2021.

[28] S. Huang, I. Ernberg, and S. Kauffman, "Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective," *Seminars in Cell & Developmental Biology*, vol. 20, pp. 869–876, Sept. 2009.

[29] "2019 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 15, pp. 321–387, Mar. 2019.

[30] A. Gonzalez, F. Quintela, D. A. Leon, M. L. Bringas-Vega, and P. A. Valdes-Sosa, "Estimating the number of available states for normal and tumor tissues in gene expression space," *Biophysical Reports*, vol. 2, no. 2, p. 100053, 2022.

[31] K. L. Spalding, R. D. Bhardwaj, B. A. Buchholz, H. Druid, and J. Frisén, "Retrospective birth dating of cells in humans," *Cell*, vol. 122, no. 1, pp. 133–143, 2005.

[32] B. Schumacher, J. Pothof, J. Vijg, and J. H. Hoeijmakers, "The central role of DNA damage in the ageing process," *Nature*, vol. 592, no. 7856, pp. 695–703, 2021.

[33] C. R. Guttmann, F. A. Jolesz, R. Kikinis, R. J. Killiany, M. B. Moss, T. Sandor, and M. S. Albert, "White matter changes with normal aging," *Neurology*, vol. 50, no. 4, pp. 972–978, 1998.

[34] L. Haghverdi, A. T. L. Lun, M. D. Morgan, and J. C. Marioni, "Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors," *Nature biotechnology*, vol. 36, no. 5, pp. 421–427, 2018.

[35] Y. Zhang, G. Parmigiani, and W. E. Johnson, "ComBat-seq: batch effect adjustment for RNA-seq count data," *NAR Genomics and Bioinformatics*, vol. 2, no. 3, p. lqaa078, 2020.

[36] O. Hahn, A. G. Foltz, M. Atkins, B. Kedir, P. Moran-Losada, I. H. Guldner, C. Munson, F. Kern, R. Pálovics, N. Lu, *et al.*, "Atlas of the aging mouse brain reveals white matter as vulnerable foci," *Cell*, vol. 186, no. 19, pp. 4117–4133.e22, 2023.

[37] M. Gillespie, B. Jassal, R. Stephan, *et al.*, "The reactome pathway knowledgebase 2022," *Nucleic Acids Research*, vol. 50, pp. D687–D692, Nov. 2021.

[38] S. Ham and S.-J. V. Lee, "Advances in transcriptome analysis of human brain aging," *Experimental & Molecular Medicine*, vol. 52, pp. 1787–1797, Nov. 2020.

[39] M. Biazzo, M. Allegra, and G. Deidda, "Clostridioides difficile and neurological disorders: New perspectives," *Frontiers in Neuroscience*, vol. 16, Sept. 2022.

[40] N. Sun, R. J. Youle, and T. Finkel, "The Mitochondrial Basis of Aging," *Molecular Cell*, vol. 61, pp. 654–666, Mar. 2016.

[41] L. Thomas, T. Florio, and C. Perez-Castro, "Extracellular Vesicles Loaded miRNAs as Potential Modulators Shared Between Glioblastoma, and Parkinson's and Alzheimer's Diseases," *Frontiers in Cellular Neuroscience*, vol. 14, Nov. 2020.

[42] D. Hanahan and R. A. Weinberg, "The Hallmarks of Cancer," *Cell*, vol. 100, pp. 57–70, Jan. 2000.

[43] D. Hanahan and R. Weinberg, "Hallmarks of Cancer: The Next Generation," *Cell*, vol. 144, pp. 646–674, Mar. 2011.

[44] D. Hanahan, "Hallmarks of Cancer: New Dimensions," *Cancer Discovery*, vol. 12, pp. 31–46, Jan. 2022.

[45] R. Ma, X. Kang, G. Zhang, F. Fang, Y. Du, and H. Lv, "High expression of UBE2C is associated with the aggressive progression and poor outcome of malignant glioma," *Oncology Letters*, vol. 11, pp. 2300–2304, Feb. 2016.

[46] S. K. Jaladanki, A. Elmas, G. S. Malave, and K.-l. Huang, "Genetic dependency of Alzheimer's disease-associated genes across cells and tissue types," *Scientific Reports*, vol. 11, June 2021.

[47] M. Mu, W. Niu, X. Zhang, S. Hu, and C. Niu, "Correction: LncRNA BCYRN1 inhibits glioma tumorigenesis by competitively binding with miR-619-5p to regulate CUEDC2 expression and the PTEN/AKT/p21 pathway," *Oncogene*, vol. 40, pp. 5972–5973, Aug. 2021.

[48] Y. Zhang, Y. Zhao, X. Ao, W. Yu, L. Zhang, Y. Wang, and W. Chang, "The Role of Non-coding RNAs in Alzheimer's Disease: From Regulated Mechanism to Therapeutic Targets and Diagnostic Biomarkers," *Frontiers in Aging Neuroscience*, vol. 13, July 2021.

[49] R. Herrero, D. A. Leon, and A. Gonzalez, "A one-dimensional parameter-free model for carcinogenesis in gene expression space," *Scientific Reports*, vol. 12, Mar. 2022.

[50] M. T. C. Poon, C. L. M. Sudlow, J. D. Figueroa, and P. M. Brennan, "Longer-term ($\geq$ 2 years) survival in patients with glioblastoma in population-based studies pre- and post-2005: a systematic review and meta-analysis," *Scientific Reports*, vol. 10, July 2020.

[51] H. Ohgaki and P. Kleihues, "Epidemiology and etiology of gliomas," *Acta Neuropathologica*, vol. 109, pp. 93–108, Jan. 2005.

[52] *Alzheimer's & Dementia*, vol. 19, pp. 1598–1695, Mar. 2023.

[53] K. H. Strang, T. E. Golde, and B. I. Giasson, "MAPT mutations, tauopathy, and mechanisms of neurodegeneration," *Laboratory Investigation*, vol. 99, pp. 912–928, July 2019.

[54] J. TCW and A. M. Goate, "Genetics of $\beta$-Amyloid Precursor Protein in Alzheimer's Disease," *Cold Spring Harbor Perspectives in Medicine*, vol. 7, p. a024539, Dec. 2016.

[55] A.-C. Raulin, S. V. Doss, Z. A. Trottier, T. C. Ikezu, G. Bu, and C.-C. Liu, "ApoE in Alzheimer's disease: pathophysiology and therapeutic strategies," *Molecular Neurodegeneration*, vol. 17, Nov. 2022.

[56] A. L. Cohen, S. L. Holmen, and H. Colman, "IDH1 and IDH2 Mutations in Gliomas," *Current Neurology and Neuroscience Reports*, vol. 13, Mar. 2013.

[57] W.-J. Chen, D.-S. He, R.-X. Tang, F.-H. Ren, and G. Chen, "Ki-67 is a valuable prognostic factor in gliomas: Evidence from a systematic review and meta-analysis," *Asian Pacific Journal of Cancer Prevention*, vol. 16, pp. 411–420, Feb. 2015.

[58] S. Haase, M. B. Garcia-Fabiani, S. Carney, D. Altshuler, F. J. Núñez, F. M. Méndez, F. Núñez, P. R. Lowenstein, and M. G. Castro, "Mutant ATRX: uncovering a new therapeutic target for glioma," *Expert Opinion on Therapeutic Targets*, vol. 22, pp. 599–613, June 2018.

[59] B. van Lengerich, L. Zhan, D. Xia, D. Chan, *et al.*, "A TREM2-activating antibody with a blood–brain barrier transport vehicle enhances microglial metabolism in Alzheimer's disease models," *Nature Neuroscience*, Jan. 2023.

[60] R. Sun, R. Han, C. McCornack, S. Khan, G. T. Tabor, Y. Chen, J. Hou, H. Jiang, K. M. Schoch, D. D. Mao, R. Cleary, A. Yang, Q. Liu, J. Luo, A. Petti, T. M. Miller, J. D. Ulrich, D. M. Holtzman, and A. H. Kim, "TREM2 inhibition triggers antitumor cell activity of myeloid cells in glioblastoma," *Science Advances*, vol. 9, May 2023.

[61] P. Mencke, Z. Hanss, I. Boussaad, P.-E. Sugier, A. Elbaz, and R. Krüger, "Bidirectional relation between parkinson's disease and glioblastoma multiforme," *Frontiers in Neurology*, vol. 11, Aug. 2020.