



FACULTAD DE INGENIERÍA Y CIENCIAS AGROPECUARIAS

IDENTIFICACIÓN DE BIOMARCADORES SANGUÍNEOS EN LA
ENFERMEDAD DE ALZHEIMER MEDIANTE LA APLICACIÓN DE
HERRAMIENTAS BIOINFORMÁTICAS EN EL PROCESAMIENTO DE
MICRO-ARREGLOS Y ANÁLISIS DE REDES DE CO-EXPRESIÓN

AUTOR

Freddy Cristhian Nieto Seraquive

AÑO

2017



FACULTAD DE INGENIERÍA Y CIENCIAS AGROPECUARIAS

IDENTIFICACIÓN DE BIOMARCADORES SANGUÍNEOS EN LA
ENFERMEDAD DE ALZHEIMER MEDIANTE LA APLICACIÓN DE
HERRAMIENTAS BIOINFORMÁTICAS EN EL PROCESAMIENTO DE MICRO-
ARREGLOS Y ANÁLISIS DE REDES DE CO-EXPRESIÓN.

“Trabajo de Titulación presentado en conformidad con los requisitos
establecidos para optar por el título de Ingeniero en Biotecnología”

Profesor guía

Ph.D. Eduardo Tejera Puente

Autor:

Freddy Cristhian Nieto Seraquive

Año

2017

DECLARACIÓN DEL PROFESOR GUÍA

“Declaro haber dirigido este trabajo a través de reuniones periódicas con el estudiante, orientando sus conocimientos y competencias para un eficiente desarrollo del tema escogido y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación”.

Eduardo Tejera Puente

Doctor en Ciencias Farmacéuticas, especialidad Bioquímica

CI: 1756501977

DECLARACIÓN DEL DIRECTOR CIENTÍFICO

“Declaro haber dirigido científicamente al estudiante para la realización de su trabajo experimental de titulación en base al método científico, conduciéndole con coherencia en el conjunto de experimentos realizados, y orientando sus conocimientos para lograr los objetivos propuestos”.

Eduardo Tejera Puente

Doctor en Ciencias Farmacéuticas, especialidad Bioquímica

CI: 1756501977

DECLARACIÓN DEL PROFESOR CORRECTOR

“Declaro haber revisado este trabajo, dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación”.

Fernando Xavier Rivas Romero

Master en Biotecnología Molecular y Celular de Plantas

CI: 1718092701

DECLARACIÓN DE AUTORÍA DEL ESTUDIANTE

“Declaro que este trabajo es original, de mi autoría, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales que protegen los derechos de autor vigentes.”

Freddy Cristhian Nieto Seraquive

CI: 1600602021

AGRADECIMIENTOS

Agradezco a la Dirección General de Investigación de la Universidad de las Américas por su compromiso y respaldo hacia todos los tesisistas de la carrera.

Agradezco de manera especial al Doctor Eduardo Tejera, mi profesor guía, quien de manera desinteresada supo brindarme su apoyo e impartirme todos los conocimientos que requería para realizar mi tesis.

Agradezco a la Doctora Vivian Morera, mi directora de carrera, quien estuvo respaldando todo mi proceso de titulación.

DEDICATORIA

A mis padres y hermanos, quienes han sido mi fortaleza y apoyo incondicional en los momentos más duros; quienes nunca han dejado de creer en mí. Gracias por su amor y sacrificio, ustedes son mi todo, los amo.

A Dios, quien me ha dado las mejores batallas y que con su infinita bondad, en un día de suerte, me colocó en el lugar que estoy ahora.

A mí mismo, que supe tener la determinación de seguir adelante, de superarme, de crecer, de soñar. Así fue que un día decidí, fui y vencí.

A mis amigos, Dany O., Nenis P., Vicho B., Mily O., Shir B., Jany A., Pame T., Gaby G. y Byron F., quienes formaron parte de una etapa extraordinaria de mi vida, en donde reímos, lloramos, pero sobre todo aprendimos. Gracias por tanto, siempre los llevo presente en mi mente y corazón.

A mis amigas, Domi O., Giss L., Naty E. y Ale S., con quienes compartimos buenos momentos en esta última etapa universitaria.

RESUMEN

La enfermedad de Alzheimer es un trastorno neurológico que afecta a la población de edad avanzada, caracterizándose por cuadros de demencia que aumenta con el pasar del tiempo. Esta patología afecta a varios millones de personas en todo el mundo y los tratamientos, aún hoy en día, son deficientes. La identificación de biomarcadores para el diagnóstico temprano de la enfermedad de Alzheimer cada vez se intensifica más. Hoy en día existen biomarcadores bien establecidos para su diagnóstico, como el análisis de beta amiloide, proteína *tau* total en líquido cefalorraquídeo o los procedimientos con resonancia magnética. La especificidad y sensibilidad que se consiguen con estos biomarcadores son muy buenas. Sin embargo, los procedimientos para la recolección de muestra y posterior análisis son complejos y en algunos casos costosos. En este estudio, se identificaron biomarcadores sanguíneos de la enfermedad de Alzheimer mediante la aplicación de herramientas bioinformáticas en el procesamiento de micro-arreglos. Los micro-arreglos obtenidos de la bases de datos públicas *Gene Expression Omnibus* fueron analizados mediante diversos algoritmos y métodos bioinformáticos. Se analizó la diferencia en la expresión génica en casos y controles, para posteriormente construir redes de co-expresión con genes que sólo se diferencien en la enfermedad de Alzheimer. Los análisis de enriquecimiento de procesos biológicos fueron necesarios para comprobar la relación entre la expresión de ciertos genes y el apareamiento de la enfermedad de Alzheimer. Con el propósito de brindar nuevas dianas de estudio, se propuso una lista de genes que son candidatos potenciales para ser biomarcadores en la enfermedad de Alzheimer. La combinación de métodos de selección de rasgos y redes de co-expresión permitió proponer los genes *PRKCB*, *CARD8*, *PIK3CG*, *NOD2*, *MXD1*, *ZFX* y *PCYOX1L* como biomarcadores o dianas terapéuticas potenciales para la enfermedad de Alzheimer.

ABSTRACT

Alzheimer's disease is a neurological disorder that affects elderly population. This pathology affects several million people around the world while the today treatments are still deficient. Researches in identification of biomarkers for the early diagnosis of Alzheimer's disease are intensified. Nowadays we count with well-established biomarkers for diagnosis, such as beta-amyloid analysis, total *tau* protein in cerebrospinal fluid or magnetic resonance imaging (MRI) procedures. The specificity and sensitivity achieved with these biomarkers are good. However, their procedures for sample collection and subsequent analysis have an increased prize and difficulty. In this study, we identified blood biomarkers in Alzheimer's disease by applying bioinformatics tools for processing and analysis of microarrays. Microarrays obtained from the public Gene Expression Omnibus databases were analyzed using several algorithms and bioinformatics methods. Differentiation of gene expression in cases and controls was studied followed by co-expression networks analysis. Enrichment analysis of biological processes and metabolic pathways were necessary to verify the relationship between the expression of certain genes and Alzheimer's disease. In order to provide a new study target, a list of genes that are potential candidates for being biomarkers in Alzheimer's disease was proposed. Combination of features selection methods and co-expression networks allowed us to propose genes *PRKCB*, *CARD8*, *PIK3CG*, *NOD2*, *MXD1*, *ZFX* and *PCYOX1L* as a potential biomarkers or therapeutic targets for Alzheimer's disease.

ÍNDICE

1.CAPÍTULO I. Introducción.....	1
1.1 Antecedentes	1
1.2 Planteamiento del problema	3
1.3 Objetivos.....	5
1.3.1 Objetivo General.....	5
1.3.2 Objetivos específicos.....	5
1.4 Justificación	5
2.CAPÍTULO II. Marco teórico	7
2.1 Biomarcador	7
2.2 Biomarcadores en la enfermedad de Alzheimer.....	8
2.2.1 Biomarcadores en líquido cefalorraquídeo (LCR)	8
2.2.2 Biomarcadores basados en neuroimagen	9
2.2.3 Biomarcadores basados en sangre.....	9
2.3 Micro-arreglos de expresión.....	10
2.3.1 Tipos de micro-arreglos de expresión.....	12
2.3.2 Proceso de cuantificación de la expresión de ARNm en micro- arreglos de expresión	12
2.3.3 Pre-procesamiento de datos de micro-arreglos de expresión	14
2.4 Métodos bioinformáticos	15
2.4.1 Redes de co-expresión.....	16
2.4.2 Selección de rasgos para el análisis de micro-arreglos.....	18
2.4.3 Diferenciación de genes	19
3.CAPITULO III. Diseño del plan experimental.....	20
4.CAPITULO IV. Procedimientos	23
4.1 Población y muestra.....	23
4.2 Materiales y métodos para la obtención de los datos	24
4.2.1 Procesamiento de micro-arreglos	24

4.2.2 Diferenciación génica	24
4.2.3 Análisis de enriquecimiento de procesos biológicos.....	25
4.2.4 Construcción de redes de co-expresión.....	25
4.2.5 Análisis de módulos.....	26
4.2.6 Distancia génica	26
4.4.7 Selección de Rasgos.....	27
5.CAPÍTULO V. Resultados y discusión.....	28
5.1 Integración de plataformas	28
5.2 Análisis de la influencia de la edad y el sexo	31
5.3 Diferenciación génica.....	32
5.4 Análisis de enriquecimiento de procesos biológicos.	34
5.5 Construcción y análisis de redes de co-expresión.....	37
5.6 Análisis de enriquecimiento de procesos biológicos por módulos	41
5.7 Selección de Rasgos	43
6. CONCLUSIONES Y RECOMENDACIONES	48
6.1 Conclusiones.....	48
6.2 Recomendaciones	49
REFERENCIAS	50
ANEXOS.....	65

CAPÍTULO I. Introducción

1.1 Antecedentes

El uso de técnicas de evaluación de la expresión génica, en la actualidad, permite un amplio estudio de enfermedades complejas (Blennow, 2010). El objetivo es comprender las relaciones entre las enfermedades, sus mecanismos y las posibles alternativas terapéuticas para enfrentarlas (Clark et al., 2012). Existen varias técnicas para evaluar la expresión génica y su regulación. Algunas de estas técnicas están bien establecidas, mientras que otras son relativamente nuevas como la *multiplexing* (Drachman, 2006). Entre las técnicas más conocidas se encuentran la hibridación fluorescente in situ (de sus siglas en inglés, *Fluorescent In Situ Hybridization*, FISH), análisis en serie de la expresión génica (de sus siglas en inglés, *Serial Analysis of Gene Expression*, SAGE), reacción en cadena de la polimerasa de transcripción inversa (de sus siglas en inglés, *Reverse Transcription Polymerase Chain Reaction*, RT-PCR), RNA seq (de sus siglas en inglés, *RNA sequencing*), micro-arreglos, entre otras (Forsberg et al., 2008).

En principio, los patrones de perfiles de expresión génica mediante técnicas como los micro-arreglos, son un método poderoso para aislar las diferencias moleculares entre tejido sano y enfermo (García, 2007). Los micro-arreglos, teniendo como base chips de vidrio, plástico o silicio, permiten el análisis de la expresión simultánea de miles de genes y la comparación de perfiles genéticos en diferentes tejidos o condiciones fisiológicas específicas (Teixeira, 2010). Un micro-arreglo contiene sondas que pueden ser secuencias de oligonucleótidos específicas, ADN genómico o fragmentos de ADNc (Hampel, 2007). La tecnología depende de la hibridación entre la muestra de ARNm y las sondas de ADN, que se imprimen en la base (Aguilar, 2013).

La enfermedad de Alzheimer es la forma más común de demencia, afectando a casi la mitad de la población mayor de 85 años (Drachman, 2006). Es una patología neurodegenerativa con una pérdida total de las funciones mentales en sus últimos estados (Henriksen, 2014). La característica que más se asocia con la patología, es la aparición de lesiones con depósitos proteicos localizados a nivel del hipocampo y áreas parietotemporales de la corteza cerebral (Henriksen, 2014). Dichas lesiones constan de placas neuríticas con depósitos extracelulares de β -amiloide y ovillos interneuronales formados por neurofibrillas de proteína *tau* (Hye et al., 2006). A pesar de los intensos esfuerzos, no hay tratamientos modificadores finales para la enfermedad de Alzheimer (Juncal et al., 2000). Entre las posibles razones para la ausencia de medicamentos eficientes, encontramos el problema de atacar una sola diana o el planteamiento de un mecanismo incorrecto (Klunk y Mathis, 2008). Por último, existe una falta de biomarcadores para registrar a los pacientes "correctos" en los ensayos, es decir, pacientes en estado temprano (Kong; Mou y Hu, 2011).

Por otro lado, existe un alto porcentaje de afectados en los cuales, los cambios patológicos relacionados con el Alzheimer comienzan en el cerebro, por lo menos una década antes de la aparición de signos y síntomas clínicos (Li et al., 2012). Esto hace que sea difícil identificar a los pacientes con Alzheimer en estadios tempranos, en donde es más probable que los tratamientos sean eficaces (Li et al., 2012). En vista de ello, investigaciones en todo el mundo buscan descubrir y desarrollar métodos precisos, fiables y rentables para la detección temprana de Alzheimer (Marcello et al., 2009). Es así que, mediante el análisis con micro-arreglos, se está dedicando gran esfuerzo en la identificación de componentes solubles en sangre y líquido cefalorraquídeo (LCR) que podrían servir como biomarcadores de Alzheimer útiles y fiables (Marcello et al., 2009).

En el LCR, los biomarcadores establecidos incluyen el beta amiloide 42 (A β 42), *tau* fosforilado (*p-tau*) y *tau* total (Mattsson; Blennow y Zetterberg, 2009). Los bajos niveles de A β 42 en el LCR de individuos con deterioro cognitivo leve (DCL), se consideran indicativos de la presencia de Alzheimer en curso temprano, pero con probabilidad de progresión rápida (Morris et al., 2009). Una limitación para el uso del LCR en general es el medio por el cual se lo obtiene; a través de una punción espinal lumbar, que se considera invasiva y no exenta de riesgo. Así mismo, se han realizado avances en la detección temprana de Alzheimer, mediante resonancia magnética (RM) y la tomografía por emisión de positrones (PET) (Romero, 2013). En estas técnicas se utilizan marcadores como Florbetapir (18F), Pittsburgh B (PiB) y fluorodeoxiglucosa (Sherman et al., 2007). El alto costo de estos procedimientos y las inconsistencias en la interpretación prohíben su uso como evaluadores de enfermedades iniciales (Romero, 2013).

En contraste, la obtención de sangre es menos invasiva. Proteínas plasmáticas, lípidos, proteínas y microRNAs encerrados dentro de los exosomas son altamente aprovechables (Schneider; Hampel y Buerger, 2009). Estos componentes han demostrado ser biomarcadores prometedores para la detección temprana del Alzheimer (Morris et al., 2009). Sin embargo, la vasta extensión y complejidad de la enfermedad, que afecta a numerosas moléculas, células y sistemas, impide los intentos de determinar qué alteraciones están específicamente asociadas con la patología temprana (Yang; Rieves y Ganley, 2012).

1.2 Planteamiento del problema

La utilización de micro-arreglos que permitan la identificación de biomarcadores en sangre de pacientes con Alzheimer, podría proporcionar nuevas estrategias para abordar la complejidad de la enfermedad (Blalock et al., 2003). La aplicación de la técnica genera gran cantidad de datos, pero el análisis e interpretación de los mismos representaría una labor lenta, costosa e inefectiva

(Miller; Galecki y Shmookler, 2001). Así mismo la interpretación, a menudo se ve obstaculizada por la baja potencia estadística (Barnes y Yaffe, 2016). Pocos estudios de micro-arreglos utilizan tamaños de muestra suficientes para proporcionar la potencia estadística necesaria para evitar errores (Cai; Zhao y Ratka, 2011). Entre los errores más comunes están; frecuencia alta de falsos positivos, falsos negativos e incertidumbre de pertinencia en los criterios de valoraciones funcionales (Ghidoni; Paterlini y Benussi, 2013).

Por tanto, el problema se centra en extraer la valiosa información biológica de esta inmensa cantidad de datos, por lo que es imprescindible la utilización de herramientas y técnicas bioinformáticas (Kong; Mou y Hu, 2011). Entre las herramientas actuales se encuentra el aprendizaje automático (*Machine learning*) y la minería de datos (*Data mining*) (Grossman et al., 2001). Con respecto al análisis de micro-arreglos, es más común la utilización de herramientas de agrupación (*clustering* y *biclustering*) incluidas dentro del aprendizaje automático (Hu et al., 2012). Estas metodologías permitirían encontrar grupos funcionales de genes que muestren comportamientos análogos (Pickering, 2012).

Del mismo modo, las redes de co-expresión desempeñan un papel central en la identificación y análisis de patrones específicos, produciendo nuevos conocimientos significativos (Josephs et al., 2006). Redes de co-expresión a partir de biomarcadores sanguíneos en Alzheimer son insuficientes (Koyama et al., 2013). Por esta razón, aún es deficiente la comprensión de los procesos biológicos complejos que están involucrados en el Alzheimer (Koyama et al., 2013).

1.3 Objetivos

1.3.1 Objetivo General

- ✓ Aplicar herramientas bioinformáticas para la identificación de biomarcadores sanguíneos en la enfermedad de Alzheimer.

1.3.2 Objetivos específicos

- ✓ Analizar micro-arreglos para evaluar la diferencia de la expresión génica mediante métodos bioinformáticos.
- ✓ Construir redes de co-expresión utilizando herramientas bioinformáticas.
- ✓ Aplicar herramientas de selección de rasgos y análisis de redes para la identificación de biomarcadores en Alzheimer.

1.4 Justificación

Los perfiles de transcripción de sangre mediante micro-arreglos, reflejan cambios neuropatológicos y grandes alteraciones somáticas (Lucie; Dwaine y Charles, 2012). Se cree que la sangre, junto con todo su perfil transcripcional, puede proporcionar una herramienta poderosa y mínimamente invasiva para la identificación de nuevos biomarcadores en Alzheimer (O'Bryant et al., 2010). Los esfuerzos hacia la identificación de biomarcadores de cambios patológicos tempranos, así como de protección neuronal, se deben intensificar. De este modo, el desarrollo y la validación de los mismos facilitará en gran medida la identificación de un tratamiento eficaz, junto con estrategias preventivas para esta enfermedad.

El estudio para la identificación de biomarcadores en Alzheimer es de interés mundial, puesto que ofrece la posibilidad de implementar una alternativa a la incapacidad de un diagnóstico precoz para el Alzheimer (Thambisetty et al., 2010). Por otra parte, la implementación de herramientas bioinformáticas, juega un rol esencial para el análisis e interpretación de datos obtenidos a partir de tecnologías de exploración masiva, en este caso los micro-arreglos (Ziemann et

al., 2011). Aguilar y colaboradores (2013), utilizaron algoritmos de aprendizaje automático mediante arboles de decisión, en donde encontraron funciones booleanas a partir de agrupación de datos. La implementación del algoritmo fue mediante el paquete bioinformático WEKA (Universidad de Waikato). Sin embargo, el estudio fue hecho a partir de datos basados en imágenes de resonancia magnética (MRI). Esto plantea la posibilidad de aplicar los mismos algoritmos de aprendizaje automático a partir de biomarcadores sanguíneos que, hasta la fecha no han sido realizados (Zipser, 2016).

Para investigar más a fondo los genes y vías biológicas que intervienen en la aparición de Alzheimer, en el presente proyecto se plantea la aplicación de varias herramientas bioinformáticas. Además se contempla la utilización de consorcios de datos de expresión disponibles de forma libre (*Gene Expression Omnibus, GEO*) que se traduce en un ahorro de tiempo significativo eliminando ensayos previos en laboratorio. El programa de la Unión Europea *AddNeuroMed* es una iniciativa multicéntrica, diseñada para recopilar y validar los datos de biomarcadores para la enfermedad de Alzheimer. Estos datos se encuentran clasificados dentro de la base de datos *GEO*.

CAPÍTULO II. Marco teórico

2.1 Biomarcador

Un biomarcador puede definirse como una característica susceptible de medirse y evaluarse objetivamente siendo un indicador de procesos biológicos normales, procesos patológicos o respuestas farmacológicas a una intervención terapéutica (Britschgi, 2011). Los biomarcadores pueden ser conceptualizados en distintos términos como rasgo, estado y tasa (Blennow, 2010). Un biomarcador de rasgo indica la susceptibilidad a una enfermedad, un biomarcador de estado se dirige al diagnóstico de una enfermedad y un biomarcador de tasa ofrece el seguimiento de la progresión de una enfermedad o un tratamiento (Cai; Zhao y Ratka, 2011). Asimismo, existen biomarcadores clínicos que buscan medir por ejemplo, el rendimiento en una prueba cognitiva. Del mismo modo, biomarcadores basados en la formación de imágenes, tales como la cuantificación volumétrica de una región específica del cerebro con la resonancia magnética (RM), biomarcadores epigenéticos (metilación del ADN), transcriptómicos (expresión de mRNA o miRNA), marcadores genéticos como MAPT H1 / genotipo H1 como un predictor independiente de riesgo en la enfermedad de Parkinson (Chalbot et al., 2011). También se conocen biomarcadores bioquímicos con varios candidatos potenciales para su identificación; éstos incluyen el líquido cefalorraquídeo, componentes de la sangre, orina y piel, entre otros (Clark et al., 2012).

Los biomarcadores son importantes por dos características fundamentales:

- ✓ **Sensibilidad:** cuando nos referimos al grado de individuos con una enfermedad específica que el ensayo predice correctamente (Cummings, 2011).
- ✓ **Especificidad:** se refiere a la fracción de individuos sin la enfermedad que el ensayo descarta correctamente (Cummings, 2011).

Actualmente, los avances en disciplinas nuevas como la lipidómica, proteómica y metabolómica, están abriendo nuevas posibilidades para la obtención de biomarcadores efectivos (Shumaker, 2011).

2.2 Biomarcadores en la enfermedad de Alzheimer

La enfermedad de Alzheimer (EA) es progresiva y no tiene cura. Los tratamientos se ven obstaculizados significativamente por la falta de biomarcadores de fácil acceso, que puedan detectar la presencia de la enfermedad y predecir el riesgo de enfermedad de forma fiable (Shumaker, 2011). A pesar de los intensos esfuerzos, no hay tratamientos modificadores aprobados para la enfermedad de Alzheimer (Ferrer, 2012). Como resultado, los esfuerzos hacia la identificación de biomarcadores de cambios patológicos tempranos, así como biomarcadores indicativos de protección neuronal, se han intensificado (Ferrer, 2012). La identificación de los biomarcadores periféricos que conduzcan a una prueba diagnóstica temprana y eficaz para la EA permitiría la detección pre-sintomática y sería valiosa para el monitoreo de la eficacia de las intervenciones durante los ensayos clínicos (Hampel, 2008). Actualmente, en todo el mundo se manejan tres tipos de biomarcadores para evaluar la progresión de la EA, los primeros basados en líquido cefalorraquídeo (LRC), luego están los biomarcadores basados en neuroimagen y por último, los biomarcadores sanguíneos.

2.2.1 Biomarcadores en líquido cefalorraquídeo (LRC)

Biomarcadores establecidos de EA a partir del líquido cefalorraquídeo (LRC) son muy precisos, pero existen obstáculos para la aplicación clínica (Herrmann; Chau; Kircanski y Lanctot, 2011). El péptido β -Amiloide 42 ($A\beta_{42}$), proteína *tau* total y niveles de proteína *tau* hiperfosforilada son biomarcadores ampliamente caracterizados en la enfermedad de Alzheimer, permitiendo identificar pacientes con EA de individuos cognitivamente normales de edad avanzada (Hampel, 2008). Además, estos biomarcadores también han demostrado el potencial de pronóstico, ya que fueron capaces de separar los sujetos con deterioro cognitivo leve (DCL) que progresaron a EA de aquellos que no lo

hicieron (Herrmann; Chau; Kircanski y Lanctot, 2011). Sin embargo, para la recolección de LCR es necesaria una punción lumbar, considerada una práctica invasiva con una percepción negativa del público (Hampel, 2012). Es una de las limitaciones en la utilización de estos biomarcadores como un sondeo de primera línea (Hampel, 2012).

2.2.2 Biomarcadores basados en neuroimagen

Los enfoques de neuroimagen, como las imágenes de resonancia magnética estructural (RM) de regiones específicas del cerebro, tomografía por emisión de positrones (PET) mediante marcadores como Florbetapir (18F), Pittsburgh B (PiB) y fluorodeoxiglucosa, y resonancia magnética funcional (RMF) se han estudiado ampliamente. Biomarcadores de neuroimagen proporcionarían un valor pronóstico para la conversión de DCL a EA, ya que los pacientes con DCL que son A β 42 positivo son altamente propensos a desencadenar EA, mientras que los son A β 42 negativo no lo son. Una importante limitación con los compuestos Pittsburgh B y Florbetapir (18F) es su corta vida media (menos de 20 minutos) y costo, lo que limita una aplicación más amplia.

2.2.3 Biomarcadores basados en sangre

Un biomarcador basado en la sangre se considera ventajoso dada la accesibilidad, mínima invasividad y costo de extracción (Hye et al., 2006). La sangre posee componentes como suero, células sanguíneas, proteínas, lípidos, y diversos productos metabólicos susceptibles de evaluación. Ray y colaboradores (2007), identificaron un panel de 18 biomarcadores de un grupo de 120 proteínas de señalización en sangre. Asimismo, existen componentes de inmunidad innata y adaptativa que están presentes en la sangre, considerada como el mejor fluido corporal para la evaluación de procesos sistémicos (Hye et al., 2006). Es así que, diversos proyectos de investigación clínica, evalúan componentes inmunológicos como biomarcadores para determinar el estado de salud del paciente, o los resultados de intervenciones farmacológicas (Ray et al., (2007).

Por otro lado, el uso potencial de la sangre, basado en el perfil de expresión génica para el diagnóstico de trastornos cerebrales, también ha sido objeto de estudio (Jack, 2010). Investigaciones demuestran un grado significativo de covariabilidad en la expresión génica entre el tejido cerebral y las células sanguíneas periféricas (Klunk y Mathis, 2008). No obstante, una conexión directa entre el cerebro y la sangre periférica está ausente, debido a una barrera hematoencefálica casi intacta (Klunk y Mathis, 2008). Es por esto que, los biomarcadores basados en la sangre de individuos con EA, serían el primer paso en una proyección de varias etapas para el proceso de reemplazo de biomarcadores basados en LCR o imagen (Mattsson; Blennow y Zetterberg, 2009). El posible desequilibrio en el nivel de expresión de genes específicos en sangre, hace imprescindible su análisis, así un gen desconocido podría convertirse en un biomarcador confiable. Los micro-arreglos de expresión, proporcionan nuevas herramientas para analizar un complejo conjunto de genes con una marcada diferencia en su expresión bajo condiciones específicas. (Schneider; Hampel y Buerger, 2009).

2.3 Micro-arreglos de expresión

Los micro-arreglos de ADN son dispositivos desarrollados en el ámbito biotecnológico que pueden medir el nivel de expresión de miles de genes en paralelo, dentro de una muestra de ARNm en particular (Morris et al., 2009). Estos perfiles de expresión pueden usarse para comparar el nivel de transcripción de genes en condiciones clínicas. Los objetivos pueden ser los de identificar biomarcadores de diagnóstico o pronóstico; controlar la respuesta a una terapia; y/o entender los mecanismos implicados en el origen de una enfermedad (Schneider; Hampel y Buerger, 2009). Por tanto, los micro-arreglos de ADN se consideran importantes herramientas para el descubrimiento en la medicina clínica (Soares, 2012).

En un micro-arreglo cada sonda (secuencias conocidas de oligonucleótidos) se repite miles de veces en un espacio microscópico, y tiene una secuencia

específica que complementa a un fragmento de secuencia de un gen (Tanne, 2012). Cada gen suele estar representado en un micro-arreglo por un conjunto de sondas de secuencias distintas, distribuidas aleatoriamente en la superficie del micro-arreglo (Soares, 2012). Es necesario la utilización de diferentes sondas para obtener una estimación correcta del nivel de expresión de un gen, cubriendo así una amplia extensión del transcrito analizado (Figura 2) (Thambisetty y Lovestone, 2010). El formato utilizado en el micro-arreglo influye en el número de sondas, que puede variar entre tres o cuatro o hasta diez por gen (Thambisetty y Lovestone, 2010).

Los cambios en la expresión de determinados genes pueden estar vinculados a la acción de estímulos particulares, a una determinada etapa fisiológica, metabólica o vía de señalización, y en última instancia la función del producto génico (Thambisetty et al., 2010).

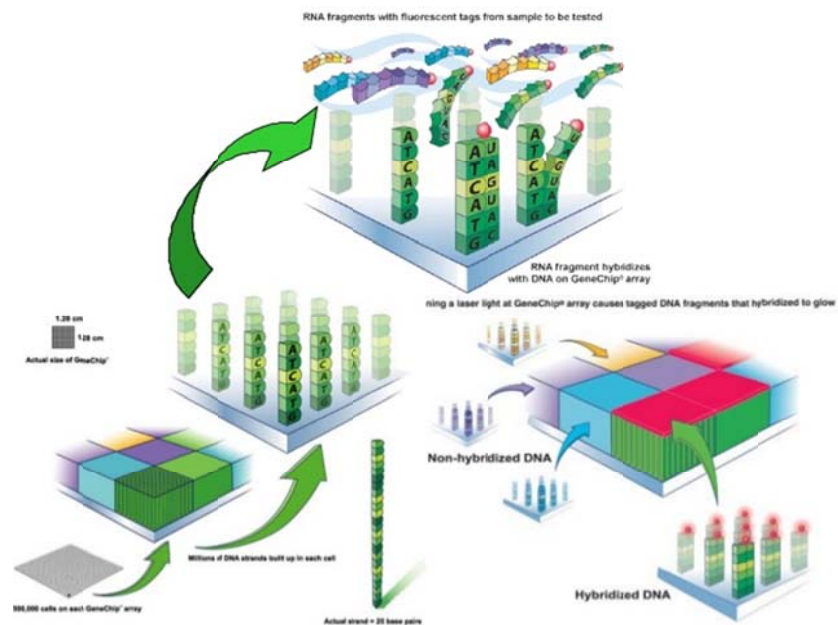


Figura 1. Esquema con la distribución de sondas en una matriz y cadenas marcadas con fluorocromo.

Adaptado de Affymetrix (s.f).

2.3.1 Tipos de micro-arreglos de expresión

Los micro-arreglos se pueden clasificar de acuerdo a tres criterios: la longitud de las sondas; método de fabricación y el número de muestras que se pueden perfilar de forma simultánea en una matriz (Thambisetty et al., 2010). Entre ellos están:

2.3.1.1 Micro-arreglos de oligonucleótidos o de canal único: en este tipo de micro-arreglos las sondas son moléculas de oligonucleótidos sintetizados directamente sobre la superficie o base. Los oligonucleótidos son polímeros cortos de ADN de cadena sencilla de 25, a 60 nucleótidos. En estos micro-arreglos cada gen está representado por varios oligonucleótidos de secuencia distinta y analizan solo un tipo de muestra (Cai; Zhao y Ratka, 2011).

2.3.1.2 Micro-arreglos de cDNA o dos canales: en los micro-arreglos de dos canales las sondas son moléculas de ADN complementario de cadena sencilla que son sintetizados en forma *in vitro* y se colocan sobre la superficie. Cada sonda representa normalmente un gen y pueden analizar dos o más muestras (Cai; Zhao y Ratka, 2011).

2.3.2 Proceso de cuantificación de la expresión de ARNm en micro-arreglos de expresión

1. **Inmovilización de las sondas:** las sondas u oligonucleótidos se inmovilizan en el soporte o matriz en posiciones ordenadas. Existen actualmente varias técnicas de inmovilización, tecnologías *in situ* como la fotolitografía, entre otras (Hampel, 2012).
2. **Extracción de ARNm de la muestra:** el ARNm se extrae de tejidos o células mediante transcripción inversa, debido a la inestabilidad del ARNm. Posteriormente se marca con un colorante fluorescente o biotina. En el caso de micro-arreglos de un solo canal usualmente se marca con biotina mientras que para micro-arreglos de dos canales se marca con colorantes como cianina (Cys-3 y Cys-5) (Hampel, 2012).

3. **Hibridación:** la hibridación se efectúa por complementariedad de cadenas y los lavados se realizan en condiciones de alta rigurosidad para minimizar la probabilidad de hibridación cruzada entre genes similares. Para los micro-arreglos de dos canales se efectúa una hibridación competitiva, por lo que la intensidad y color de la micro-imagen dependerá del grado de hibridación de las sondas con las cadenas marcadas (Figura 3) (Hampel, 2012).
4. **Cuantificación:** en esta etapa se determina la cantidad de muestra hibridada mediante un escáner laser que detecta moléculas marcadas fluorescentemente. La cantidad de fluorescencia medida en cada ubicación específica es directamente proporcional a la cantidad de ARNm con la secuencia complementaria presente en la muestra analizada. La fluorescencia es capturada en una imagen formada por una malla de puntos brillantes para su cuantificación (Hampel, 2012).

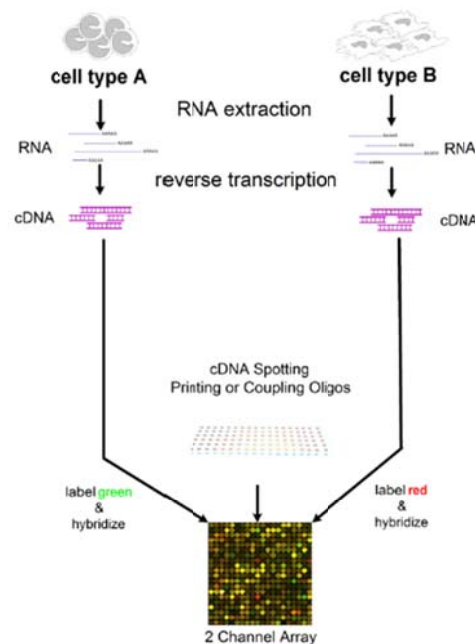


Figura 2. Proceso de hibridación y cuantificación de secuencias de ARNm en micro-arreglos de expresión.

Adaptado de (Cano, 2010).

2.3.3 Pre-procesamiento de datos de micro-arreglos de expresión

Las imágenes resultantes de la hibridación de los micro-arreglos se utilizan para generar un conjunto de datos (Hye et al., 2006). Estos datos necesitan ser "pre-procesados" antes del análisis e interpretación de los resultados (Hye et al., 2006). El pre-procesamiento no es más que transformar los datos sin procesar en datos biológicamente útiles (Klunk y Mathis, 2008). Por lo tanto, es un paso que mejora y extrae las características significativas de los datos y los prepara para la aplicación de métodos de análisis (Klunk y Mathis, 2008).

La normalización es un tipo particular de pre-procesamiento, en donde se busca eliminar ciertas diferencias sistémicas (Mattsson; Blennow y Zetterberg, 2009). Básicamente, se corrigen diferencias en las distribuciones de los valores de intensidad de las distintas matrices y plataformas (Mattsson; Blennow y Zetterberg, 2009). Dependiendo del tipo de plataforma y tecnología del micro-arreglo, existen distintos métodos para normalizar los datos: *Linear lowess*, *Robust linear lowess*, *Rank invariant*, *Quantile normalization* (Soares, 2012).

En general podemos describir tres tipos de normalización:

Normalización intra-muestra: se realiza debido a que la distribución de los datos es altamente asimétrica. Por lo tanto, es necesario lograr una mayor simetría en la distribución mediante la aplicación del logaritmo en base dos para distribuir la frecuencia más uniformemente. Dicha transformación es recomendada en determinadas ocasiones por el proveedor del micro-arreglo (Thambisetty y Lovestone, 2010).

Normalización intra-arreglo: la normalización entre matrices es necesario debido a que los promedios y varianzas son distintos en las muestras procesadas. La normalización dentro de la matriz se toma para cada micro-arreglo por separado y puede involucrar a todos los genes, o a sólo partes de

ellos. Este tipo de normalización tiene como objetivo, entre otras cosas eliminar la variabilidad de los colorantes fluorescentes (Thambisetty y Lovestone, 2010).

Normalización inter-plataforma: la normalización entre plataformas es necesaria debido a que los promedios y varianzas son distintos en las muestras procesadas. Por otro lado, usualmente involucra estudios diversos lo que aumenta la variabilidad de las mediciones. En las matrices los factores que aumentan la varianza son generados por el equipamiento con el que se trabaja, procedimientos en la medición y el operador (Cummings, 2011).

Con la normalización se obtiene una matriz numérica normalizada, que constituye por sí sola una fase donde se aplica un cierto nivel bioinformático. Posteriormente, la aplicación de herramientas bioinformáticas, permite determinar genes con una expresión génica más alta y agruparlos mediante métodos de clasificación como el *clustering*. La determinación de patrones de comportamiento común y agrupación, son procedimientos que están inmersos en la aplicación de paquetes incluidos dentro de lenguajes de programación (Cavalieri y De Filippo, 2005).

2.4 Métodos bioinformáticos

La bioinformática posee múltiples herramientas para el análisis de grandes cantidades de datos biológicos (Tamames et al., 2002). Es así que puede apoyarse en la programación para realizar instrucciones de asociación, cálculo, clasificación y análisis estadístico (Obayashi et al., 2007). Este tipo de instrucciones permiten de forma automática obtener datos valiosos en el ámbito de la biomedicina. Una de sus aplicaciones es el análisis de expresión génica. *Bioconductor* es un proyecto de software de desarrollo abierto escrito en lenguaje de programación R para el análisis y comprensión de los datos genómicos y de expresión génica (Cavalieri y De Filippo, 2005).

Uno de los paquetes más usados en el análisis de expresión es *Limma* (Smyth, 2005). Esta instrucción en R, usada para el análisis estadístico de datos genómicos, ha demostrado ser una opción frecuentemente utilizada para el análisis de los datos de los experimentos con micro-arreglos, PCR, entre otras (Cruz et al., 2016). El paquete está diseñado de tal manera que, después de un pre-procesamiento y la normalización inicial, es posible llevar a cabo el análisis de un sin fin de tecnologías (Smyth, 2005).

2.4.1 Redes de co-expresión

Una red de co-expresión implica la evaluación de la relación entre la expresión de un gen y el resto de los genes analizados. Esto quiere decir que, si un gen sufre un cambio en su expresión (por una determinada condición, ej. patología), este cambio puede generar una modificación en la expresión de otros genes. El estudio de estas relaciones es el estudio de las redes de co-expresión (Tamames et al., 2002).

Las redes de co-expresión ofrecen información sobre el genoma a gran escala y también tienen el potencial para poner en relieve los mecanismos moleculares específicos en la enfermedad, sobre todo si la base biofísica de la co-expresión se integra en las herramientas matemáticas del análisis de redes (ej. módulos y centralidad) (Tamames et al., 2002). Por ejemplo, es común la evaluación del número de vecinos en la red de co-expresión para identificar los genes altamente conectados ("genes *hub*") que pueden estar correlacionados con la enfermedad o ser mediadores putativos de la patología (Cavalieri y De Filippo, 2005). Si bien este enfoque (centralidades o genes *hub*) ha dado lugar a muchas ideas, tiende a centrar la atención en unos pocos genes, y hace caso omiso de las muchas formas en las que la co-expresión se pueden utilizar para generar y transformar los sistemas de conocimientos en predicciones comprobables (Zhang y Horvath, 2005).

Los estudios en múltiples especies, tejidos y plataformas han demostrado que los genes co-expresados tienden a estar funcionalmente relacionados (Obayashi et al., 2007). De forma análoga, los conjuntos de genes que están densamente interconectados dentro de la red global, son comúnmente conocidos como "módulos" (Langfelder y Horvath, 2008) y, consecuentemente estos módulos, usualmente responden a procesos biológicos específicos. Si una fracción significativa de los genes en un módulo se corresponde con un determinado proceso biológico (definido por su ontología genética), entonces consideramos la regla de "culpabilidad por asociación", es decir, el resto de genes en el módulo se supone que también están relacionados a ese mismo proceso (Tamames et al., 2002). De esta forma, el análisis típico en las redes de co-expresión busca asociar los módulos con enfermedades u otros rasgos fenotípicos registrados en el mismo conjunto de datos (Obayashi et al., 2007).

En la red de co-expresión, los genes seleccionados son representados por nodos y las conexiones entre ellos, que representan la fuerza de la conexión ($\text{Alfa}_{i,j}$), son representadas esencialmente por una matriz de adyacencia ponderada con elementos:

$$\text{Alfa}_{i,j} = | \text{cor}(x_i, x_j) |^\beta \quad (\text{Ecuación 1})$$

Donde x_i y x_j son los perfiles de expresión de los genes i y j . Este método considera un intervalo continuo (0,1) en lugar de una matriz de adyacencia discreta (Zhang y Horvath, 2005).

Las correlaciones entre los genes pueden alterarse en la enfermedad, así como la estructura reguladora, sin que esto afecte la expresión diferencial (Zhang et al., 2013). Mientras que la co-expresión diferencial es en sí misma una nueva herramienta, se aplica generalmente para encontrar módulos enteros de genes con conectividad diferente (correlaciones) entre casos y control (Zhang et al., 2013). Ampliando el trabajo previo en la co-expresión diferencial, Hudson y

colaboradores (2009) y Rhinn y colaboradores (2012), muestran cómo la co-expresión diferencial puede ser utilizada para priorizar objetivos moleculares relacionados con la enfermedad.

2.4.2 Selección de rasgos para el análisis de micro-arreglos

La característica principal de los nuevos conjuntos de datos creados por micro-arreglos es su gran dimensionalidad (es decir muchas variables, ej. genes), y en muchas ocasiones un tamaño de muestra pequeño. La idea central en la selección de rasgos es extraer aquellos rasgos (variables) que representen, de forma más completa la información contenida en todo el espacio, disminuyendo de esta forma la dimensionalidad (Liu; Ranka y Kahveci, 2008).

Existen diversos métodos de selección de rasgos en el análisis de micro-arreglos, por ejemplo: métodos univariados y multivariados, métodos evolutivos y otros que utilizan clasificadores lineales y no lineales, como máquinas de vector soporte (de sus siglas en inglés, *Support Vector Machines*, SVMs) y regresión logística (Liu; Ranka y Kahveci, 2008). Debido a la gran complejidad y diversidad de métodos, resulta difícil explicar todos ellos aquí. Por eso nos limitaremos a explicar aquel que será utilizado y que se denomina como “mínima redundancia máxima relevancia” (mRMR) (Hanchuan; Fuhui y Ding, 2005).

Este método considera para su ejecución no sólo la expresión de los genes para la separación de las clases (ej. de casos y controles), sino que adicionalmente analiza las relaciones entre los propios genes (ej. correlaciones), reduciendo de esta forma la posibilidad de seleccionar genes que contribuyan con la misma información (Hanchuan; Fuhui y Ding, 2005). Para ello, este método emplea la información mutua como medida de correlación entre los genes, al mismo tiempo que va creando grupos de los mismos y evaluando el poder de clasificación. Una de las ventajas de este

método es que es rápido y básicamente no lineal (Hanchuan; Fuhui y Ding, 2005).

2.4.3 Diferenciación de genes

La evaluación de la expresión diferencial génica, entre muchos grupos y condiciones diferentes al mismo tiempo, se puede realizar mediante diversas formas. No obstante, una de las más empleadas en el ambiente de *R* es el paquete *limma*, que se encuentra en *Bioconductor* y en especial la función *lmFit* (Smyth, 2005) (Gentleman et al., 2004) (R Core Team, 2013). Esta función acepta una matriz de modelo que describe el diseño experimental, y produce una salida que almacena la información del modelo ajustado para cada gen. Cada uno de estos modelos es posteriormente ajustado por la función *eBayes*, que considera los estadísticos de prueba empíricos de Bayes para cada gen, incluyendo los valores de p y el logaritmo de la expresión diferencial (Johnson; Li y Rabinovic, 2007).

Es importante destacar que, como este análisis es realizado en miles de genes simultáneamente, es importante disminuir el error al rechazar hipótesis nulas en diferencias de expresión génicas, cuando estas son realmente nulas. Para ello, se ajusta el valor de p aplicando los métodos de factor de falso descubrimiento (de sus siglas en inglés, *False Discovery Rate*, *FDR*) (Benjamini y Hochberg, 1995).

CAPITULO III. Diseño del plan experimental

En el presente trabajo se realizó la identificación de biomarcadores sanguíneos en la enfermedad de Alzheimer, usando herramientas bioinformáticas en el procesamiento de micro-arreglos, tales como el análisis de redes de co-expresión y la selección de rasgos. El estudio no constituyó la aplicación de procedimientos experimentales con muestras de individuos dentro de un laboratorio. El programa de la Unión Europea *AddNeuroMed* se encarga del financiamiento, recopilación y distribución de datos de biomarcadores para la enfermedad de Alzheimer. Estos son posteriormente presentados públicamente en varias plataformas de almacenamiento de datos, una de ellas *Gene Expression Omnibus (GEO)*, Edgar; Domrachev y Lash, 2002). Es así que, se utilizaron datos de dos ensayos de micro-arreglos, almacenados y disponibles libremente en la base de datos *GEO*, siguiendo lineamientos de Información Mínima sobre un Experimento de Micro-arreglos (de sus siglas en inglés, *Minimum Information About a Microarray Experiment, MIAME*). Los lineamientos *MIAME* son directrices para presentar la información mínima que debe incluirse cuando se detalla un experimento de micro-arreglos. Dentro del proyecto se manejaron variables dependientes caso/control e independientes como la edad y sexo.

Para alcanzar los objetivos planteados, se elaboró una base de datos de micro-arreglos curada con casos y controles, a partir de la base de datos GSE63060 (Sood et al., 2015) con la plataforma GPL6947 basada en chip de perlas de expresión *Illumina HumanHT-12 V3.0*; y GSE63061 (Sood et al., 2015) con plataforma la GPL10558 basada en chip de perlas de expresión *Illumina HumanHT-12 V4.0*. La distribución y los determinantes de estados de casos y controles de los dos ensayos fueron obtenidos de *ArrayExpress* (Kolesnikov et al., 2015).

Luego, mediante el lenguaje de programación *R* (R Core Team, 2013), junto con *Bioconductor* (Gentleman et al., 2004), se normalizaron los valores de expresión media entre plataformas (*cross-normalization*). La normalización dentro de cada una de las muestras y entre muestras no fue necesaria, debido a que cada ensayo fue normalizado previamente por los investigadores que publicaron dichas bases de datos en *GEO*. Se realizó la integración de las bases de datos, junto con la clasificación epidemiológica de la muestra y número de genes a estudiar.

Posteriormente se realizó un análisis estadístico convencional para evaluar la significancia de las variables edad y sexo. A continuación, se realizó la diferenciación génica entre casos y controles mediante el paquete *limma*. El análisis de enriquecimiento de procesos biológicos fue realizado en los genes seleccionados para establecer sus relaciones con la enfermedad de Alzheimer. Posteriormente, se realizó la construcción de redes de co-expresión (representación gráfica con dendogramas), análisis de módulos, determinación de conectividad y distancia génica.

Por último, la información de las redes de co-expresión se combinó con estrategias de análisis de redes y de selección de rasgos para la determinación de los biomarcadores más promisorios. El resumen del diseño experimental se lo aprecia en la figura 3.

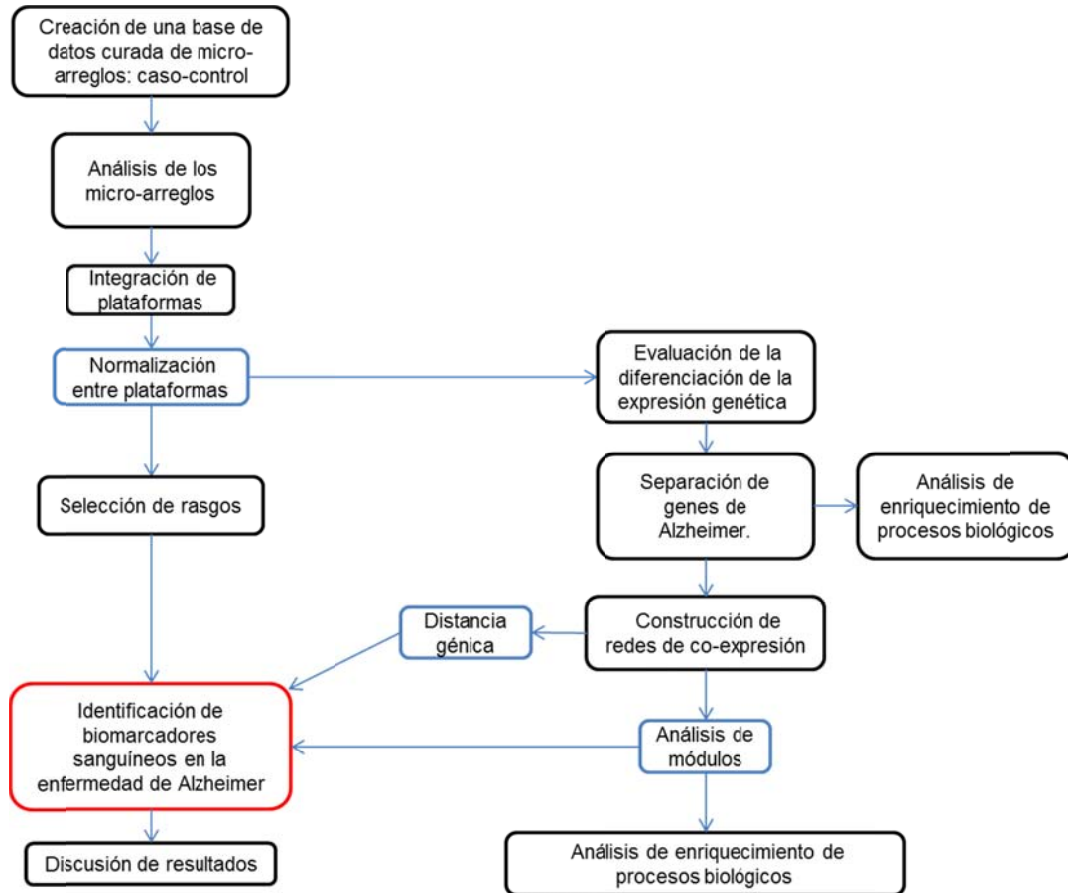


Figura 3. Diagrama de flujo con las fases para el desarrollo del estudio: identificación de biomarcadores sanguíneos en la enfermedad de Alzheimer mediante la aplicación de herramientas bioinformáticas en el procesamiento de micro-arreglos y análisis de redes de co-expresión.

CAPITULO IV. Procedimientos

4.1 Población y muestra

Se seleccionaron dos ensayos de los repositorios *GEO* (Edgar; Domrachev y Lash, 2002) y que también se encuentran en la base de datos *ArrayExpress* (Kolesnikov et al., 2015) que tenían perfiles de expresión génica con relación al Alzheimer en sangre. Los grupos de muestras de casos y controles de Alzheimer se obtienen gracias al financiamiento de la Unión Europea *AddNeuroMed*. Los estudios se apoyan en sangre humana como fuente de ARN. El primer estudio GSE63060 contiene 329 muestras mientras que el segundo GSE63061 contiene un total de 388 muestras (Tabla 1). Los casos corresponden a pacientes con enfermedad de Alzheimer y pacientes con deterioro cognitivo leve. Cada caso tiene descripciones propias de edad, etnia y sexo, disponibles en el repositorio *ArrayExpress* anteriormente mencionado.

Tabla 1.

Total de muestras de los ensayos utilizados y total de genes en común obtenidos a partir del procesamiento en R.

	Número de muestras	Número de genes en común
GSE63061	388	14476
GSE63060	329	
Total	717	

Se asignó a cada variable dentro del estudio un código para su identificación y ejecución de instrucciones a lo largo del procesamiento. Para los casos de Alzheimer se asignó el código (A); para el deterioro cognitivo leve (M) y para los controles (C). Con respecto al sexo, se asignó el código (f) para mujeres y (m) para hombres. La distribución y clasificación epidemiológica de toda la muestra se encuentran en cada fase (Tabla 2).

Tabla 2.

Clasificación epidemiológica con la mediana de la edad y sexo de los individuos según su estado dentro de la muestra.

		Estado						Total	
		Alzheimer (A)		Deterioro cognitivo leve (M)		Control (C)			
Edad		78(58-95)		77(57-100)		74(52-91)			
		n	% del total	n	% del total	n	% del total	n	% del total
Sexo	Mujeres	184	42,40%	144	33,18%	106	24,42%	434	60,53%
	Hombres	101	35,69%	95	33,74%	87	30,74%	283	39,47%
Muestra		n	% del total	n	% del total	n	% del total	n	% del total
		285	39,75%	193	26,92%	239	33,33%	717	100%

4.2 Materiales y métodos para la obtención de los datos

El estudio requirió como equipo de trabajo una portátil con sistema operativo capaz de soportar paquetes bioinformáticos.

4.2.1 Procesamiento de micro-arreglos

Mediante el uso del paquete *GEOquery* (Davis y Meltzer, 2007), en *R*, es posible la comunicación directa con los repositorios de datos utilizados en *GEO*. El mapeo entre las sondas y los genes se realizó de forma manual, empleando las respectivas anotaciones de las plataformas asociadas a cada micro-arreglo. En aquellos casos donde dos o más sondas corresponden a un mismo gen, se empleó la media de los valores para la intensidad final. Para esto se usan las funciones: *collapseRows* (Miller et al., 2011) e *intersect* (Langfelder y Horvath, 2008) implementadas en el paquete *WGCNA*. La normalización inter plataformas se llevó a cabo con el fin de volver a escalar la intensidad y eliminar los efectos de diferencia de intensidad entre estudios. Esto se realizó mediante la función *Combat* en el paquete *SVA* (Johnson; Li y Rabinovic, 2007).

4.2.2 Diferenciación génica

La evaluación de genes, con una diferencia de expresión estadísticamente significativa entre C y grupos (A y M) se obtuvo mediante el paquete *limma*

(Smyth, G. K. 2005). El intervalo de confianza para definir la significación estadística fue $p < 0.05$. Luego, el ajuste de tasa de falso descubrimiento (*FDR*) al valor de p obtenido fue necesario para eliminar los falsos negativos (Benjamini y Hochberg, 1995). Se aplicó el $-\log_{10}$ al ajuste del valor p concerniente a la diferencia de expresión génica, para llevar los datos a una escala positiva y una mejor apreciación visual.

4.2.3 Análisis de enriquecimiento de procesos biológicos.

El análisis de procesos biológicos (de sus siglas en inglés, *Biological Process*, *BP*) se realizó con el recurso bioinformático *DAVID* v6.7 (de sus siglas en inglés, *Database for Annotation, Visualization and Integrated Discovery*, *DAVID*) (Huang; Sherman, y Lempicki, 2009).

4.2.4 Construcción de redes de co-expresión

Para la construcción de las redes de co-expresión, fue necesario plantearse criterios de discriminación para los genes que participan en cada caso. Se consideró sólo los genes que se expresen en la enfermedad de Alzheimer con respecto a los controles, pero que no sean comunes con los que se expresan en el deterioro cognitivo leve. El criterio se aplicó en base al conocimiento de que, no todo deterioro cognitivo leve en un individuo desencadena en la enfermedad de Alzheimer (Petersen et al., 1999) (Flicker; Ferris y Reisberg, 1991). Debido a la gran cantidad de genes, las condiciones para su selección se incrementó. Es así que se ajustó la diferencia en la expresión génica hasta un valor de $p < 0,01$.

Para la construcción de la redes de co-expresión, se utilizó el paquete *WGCNA*. En este estudio, se seleccionó $\beta = 6$, siguiendo el criterio de topología libre de escala, utilizando la función *pickSoftThreshold* (Langfelder et al., 2008) en *WGCNA*.

4.2.5 Análisis de módulos

Los módulos se detectaron utilizando el algoritmo dinámico de árbol cortado con la función *cutreeDynamic* (Langfelder; Zhang y Horvath, 2008) en el paquete *WGCNA*. La conectividad del nodo (k) y la conectividad intramodular del nodo (k_{intra}) se calcularon para cada módulo como se describe en Tejera; Bernardes y Rebelo (2013) y Cruz y colaboradores (2016). Se establecieron colores para cada módulo en inglés para las instrucciones en *R*, y sus equivalencias en español para la presentación de resultados.

Cada uno de los módulos tiene un símbolo (MA) dentro del dendograma, representándose con un color en particular (turquesa, marrón, rojo, azul, amarillo, gris y verde) y agrupando cierta cantidad de genes. De acuerdo al método utilizado, al módulo gris no se le atribuye una función especial o un proceso específico; este agrupa genes que no tienen una función en particular, reflejando el remanente de genes que no se co-expresan de forma suficiente. Este principio se aplica para todos los dendogramas obtenidos.

4.2.6 Distancia génica

Otra forma de comparar la conectividad de los genes, es mediante un análisis de la distancia topológica entre genes en las condiciones estudiadas. Para esto usamos la formulación presentada en Tejera; Bernardes y Rebelo en 2013. La distancia se define como:

$$KDist_i = \frac{1}{k_i^{CoA} + k_i^{CoC}} \sqrt{\sum_{j=1}^C (\alpha_{i,j}^{CoA} - \alpha_{i,j}^{CoC})^2} \quad (\text{Ecuación 2})$$

Donde k^{CoA} y k^{CoC} corresponden con las conectividades totales de un gen i en la red de Alzheimer y Control respectivamente. Por otro lado α^{CoA} y α^{CoC} son los valores de correlación en la matriz de adyacencia entre el gen i y todos los genes vecinos.

4.4.7 Selección de Rasgos

Se realizaron dos procedimientos para la selección de rasgos. El primero (Modelo 1) se definió como clases sólo los casos de Alzheimer (clase=1) y Control (clase=0), mientras que en el segundo (Modelo 2) se consideró igualmente dos clases pero de la siguiente forma: Alzheimer (clase=1) y Control + M (clase=0).

En la selección de rasgos se usaron los 14774 genes, y se predefinió el algoritmo mRMR según Hanchuan; Fuhui y Ding en 2005, para elegir de forma ordenada los 250 genes con mayor relevancia y mínima redundancia. Dicha estrategia ha sido recientemente reportada en estudios de priorización de genes en Parkinson por Cruz y colaboradores (2016). La selección de 250 genes fue, en cierta forma arbitraria, es más para problemas en los cuales fuesen a ser empleados algoritmos de aprendizaje automático, sería conveniente una cantidad mucho menor de variables. No obstante, como se va a realizar un cruzamiento con la información que ya se posea de los módulos y valores de expresión, es necesario un número relativamente alto de genes. De cualquier forma, siendo una lista ordenada, no solo se debió considerar la cantidad de genes en común, sino también la posición (*ranking*) en el cual un gen es identificado por su grado de relevancia y mínima redundancia.

CAPÍTULO V. Resultados y discusión

5.1 Integración de plataformas

En la figura 4, se representa la diferencia de expresión media de cada micro-arreglo (muestra) en ambos estudios.

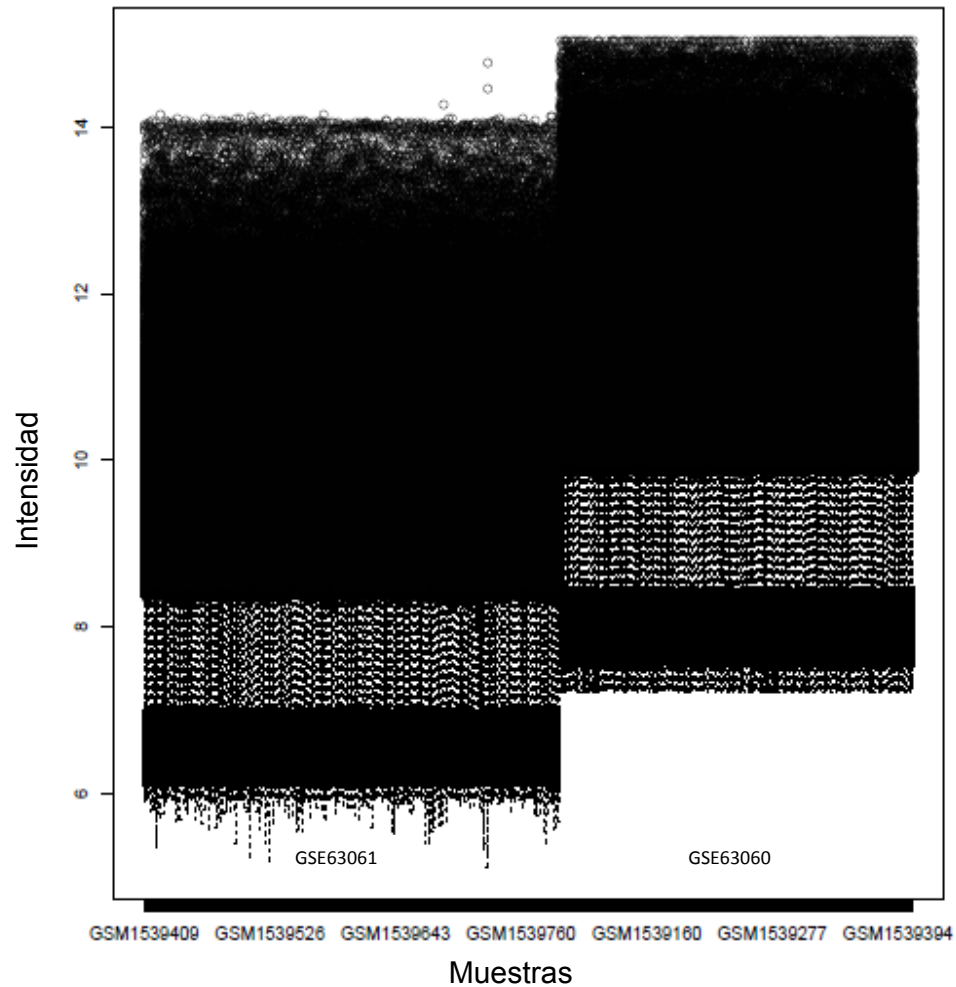


Figura 4. Valores de expresión media para cada micro-arreglo en los estudios GSE63061 y GSE63060 previo a la normalización entre plataformas.

La distribución de las intensidades claramente demuestra que hay grandes diferencias en los dos estudios. Esto se debe a múltiples factores incluyendo diferentes formas de corrección de fondo o formas diferentes de normalización

en cada estudio. En la figura 5, se representa el mismo fenómeno pero usando otra forma de estimación, mediante una proyección de componentes principales (usando la función *cmdscale* de *R*).

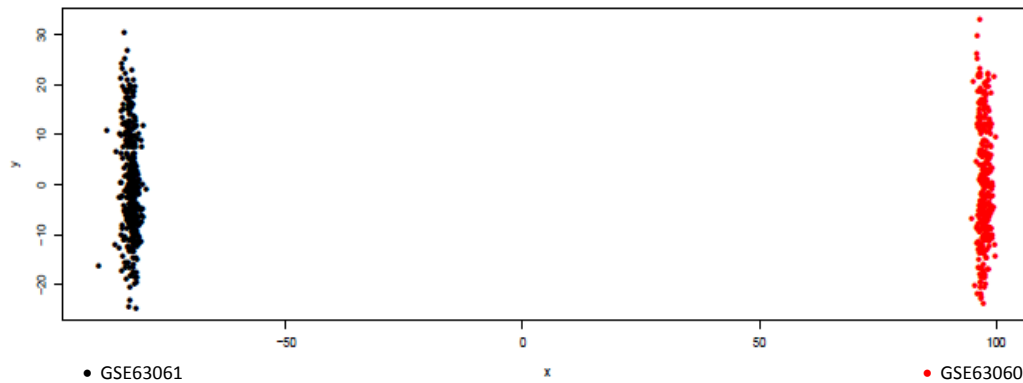


Figura 5. Proyección de componentes principales para cada micro-arreglo en los estudios GSE63061 y GSE63060 previo a la normalización entre plataformas.

El hecho de poder separar visualmente los dos estudios indica, de forma evidente, el sesgo presente en los datos por pertenecer a estudios diferentes. La normalización posterior tiene como objetivo eliminar este tipo de sesgo, para poder establecer una comparación confiable de las expresiones génicas en los dos estudios.

En las figuras 6 y 7, se representa la homogeneidad que existe en el valor de expresión media en cada micro-arreglo entre las dos plataformas posterior a su normalización.

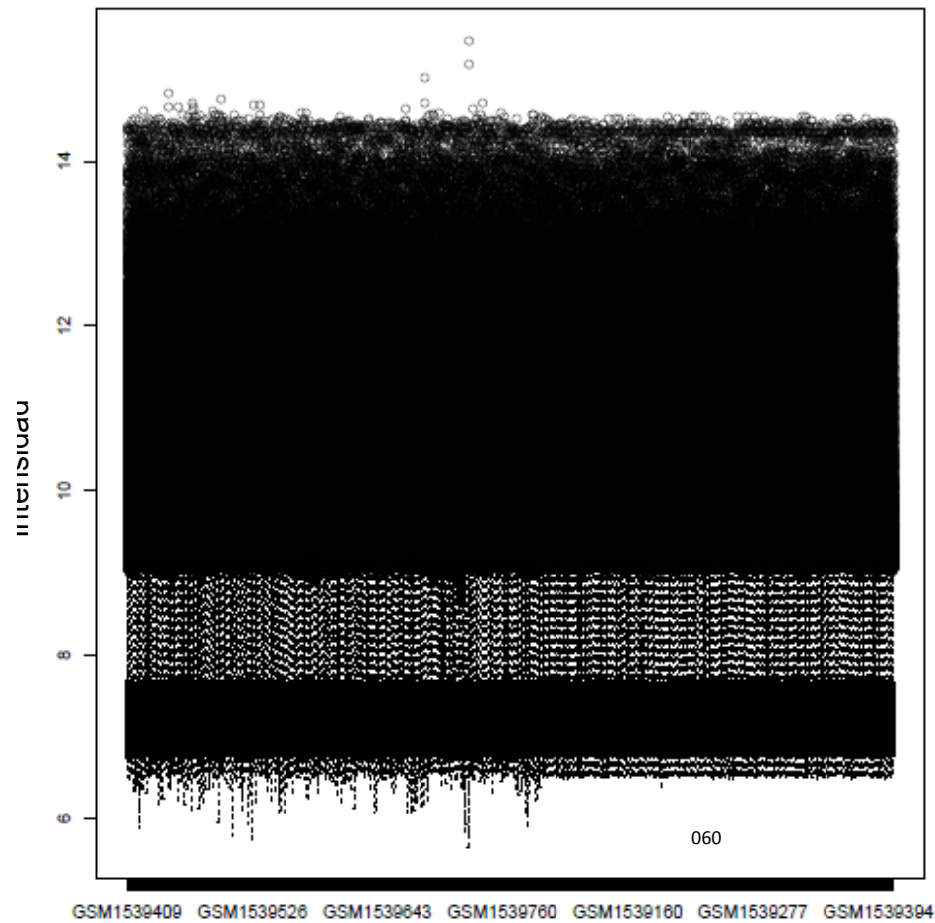


Figura 6. Valores de expresión media para cada micro-arreglo en los estudios GSE63061 y GSE63060 posterior a la normalización entre plataformas.

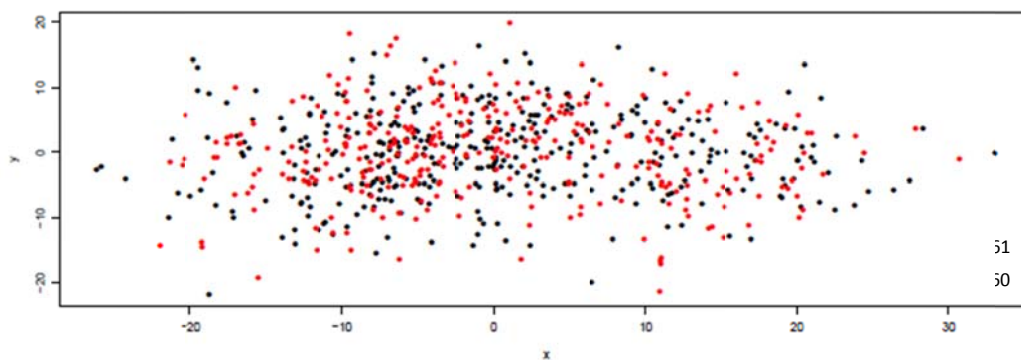


Figura 7. Proyección de componentes principales para cada micro-arreglo en los estudios GSE63061 y GSE63060 posterior a la normalización entre plataformas.

Posteriormente a la integración y mapeo de los genes, se tuvo un total de 14476 genes que deberán ser analizados.

5.2 Análisis de la influencia de la edad y el sexo

Se realizó un análisis de varianza (ANOVA) para determinar si existían diferencias significativas entre la edad y sexo con respecto a la enfermedad de Alzheimer. Debido a que se trabajó con tres variables en el estudio (A, M y C), se realizó un ajuste de Bonferroni, mientras que para la variable categórica (sexo), se utilizó un test de chi-2. En la tabla 3, se presentan los niveles de significancia para las variables edad y sexo con respecto a la enfermedad de Alzheimer en la muestra estudiada.

Tabla 3.

Análisis de varianza (ANOVA) y chi-2 para las variables edad y sexo.

Variable	F	Valor p
Edad	12,15	p<0,01*
Sexo	1,934	n.s.

Nota, *: Estadísticamente significativo; n.s.: no significativo

Los resultados obtenidos para la edad muestran una diferencia estadísticamente significativa en su relación con el Alzheimer y el deterioro cognitivo leve. Por otro lado, el sexo no muestra diferencias estadísticas significativas.

En este sentido, estudios previos han demostrado cómo la edad influye en el apareamiento y progreso de la enfermedad de Alzheimer. Los resultados en los estudios sugieren que la taupatía asociada con la enfermedad de Alzheimer, aparece aproximadamente en el 95% de personas en edades mayores a los 69 años. Además, se reporta que en algunos casos puede comenzar antes de lo esperado, en edades posteriores a la pubertad inclusive antes de dicha etapa (Braak, 2011). Del mismo modo, Ferrer y colaboradores (2012), describe en sus estudios una distinción entre el envejecimiento normal

y la enfermedad de Alzheimer por la presencia de placas seniles y ovillos neurofibrilares conforme se incrementa la edad, con alrededor del 80% de los individuos mayores de 65 años. Por lo tanto, podemos señalar que los resultados obtenidos en nuestro estudio eran esperados y están respaldados con estudios previos.

5.3 Diferenciación génica

Los 14476 genes proporcionados por las 717 muestras de individuos con la enfermedad de Alzheimer (A), deterioro cognitivo leve (M) y controles (C) se procesó con *limma*. El objetivo aquí fue identificar los genes diferenciados significativamente entre Alzheimer-Control (A-C), Alzheimer-Deterioro cognitivo leve (A-M) y deterioro cognitivo leve-controles (M-C). En la Tabla 4, se muestran las cantidades de genes identificados bajo dos criterios de significación estadística.

Tabla 4.

Números de genes identificados en la diferenciación génica.

	Números de genes identificados con un valor $p^* < 0.05$	Números de genes identificados con un valor $p < 0.01$
A-C	3774	2758
A-M	446	31
M-C	5349	4034

*El valor de p corresponde con el ajustado por medio del factor de falso descubrimiento (FDR)

Se sabe que no todo deterioro cognitivo leve en un individuo desencadena en la enfermedad de Alzheimer, por lo tanto, buscando una mayor especificidad, solo nos interesó la identificación de genes que sean capaces de diferenciar los casos de Alzheimer. Para esto fue necesario entonces eliminar aquellos genes que también aparecen como significativos para el M. En la figura 8, se muestran los diagramas de Venn con los genes en cada caso utilizando dos valores de cortes de significación: 0.05 y 0.01 respectivamente.

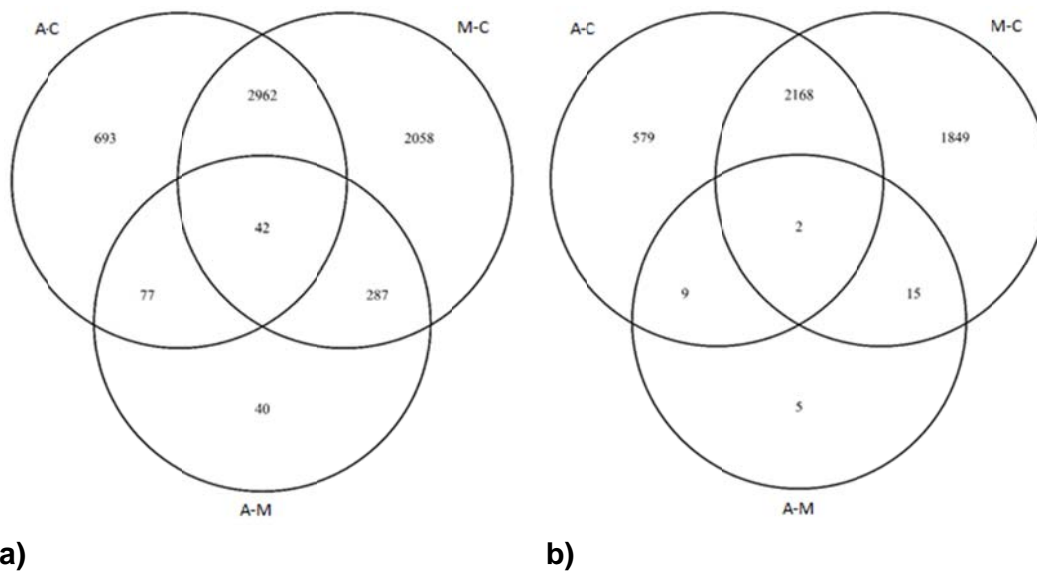


Figura 8. Diagramas de Venn

a) Diagrama de Venn con genes comunes con un valor de corte de significación 0.05.

b) Diagrama de Venn con genes obtenidos con un valor de corte de significación 0.01.

Observando el diagrama de Venn para el valor de corte más estricto, observamos que de los 2758 genes obtenidos con un valor de $p < 0.01$, solo 579 genes están únicamente diferenciados en Alzheimer. Estos genes fueron los que se utilizaron para la construcción de la redes de co-expresión. Los resultados de expresión y valores p obtenidos para estos genes en la comparación Alzheimer-control se muestran en el Anexo 1.

En la figura 9, se muestra de manera clara y representativa a aquellos genes que muestran una mayor diferencia de expresión entre A-C, pero que no representan de forma exclusiva la enfermedad de Alzheimer. El uso de los genes con mayor diferenciación (*RNA28S5*, *CETN2*, *MRPL51*, *RPL36AL*, *RPS25*, *NDUFA1*, *NDUFS5*, *TOMM7*, *RPL26*) como biomarcadores, sería un error, ya que podría incrementar los falsos positivos confundiendo el Alzheimer

y el deterioro cognitivo leve. Contrariamente con esta afirmación, un estudio reportó una co-localización de factores de metilación de estabilización con rRNA (5.8S rRNA), y la proteína ribosomal codificada por el gen *RPL26*. La disminución de estas proteínas en muestras de cerebro de individuos con Alzheimer, sugiere que los mecanismos epigenéticos podrían desempeñar un papel importante para modular la expresión (Mastroeni et al., 2010).

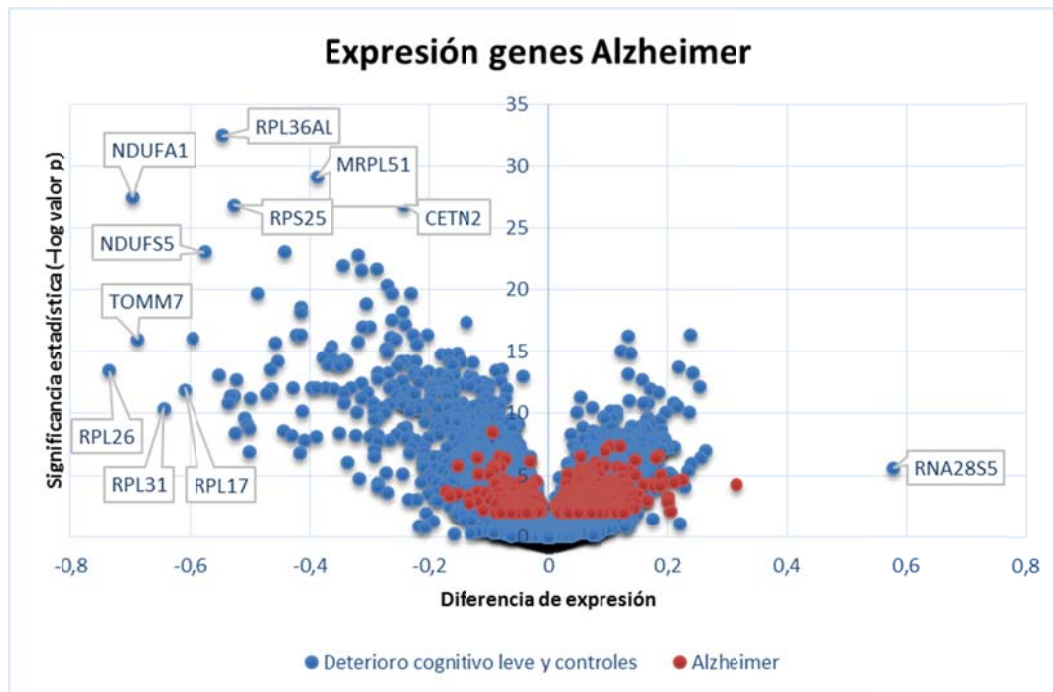


Figura 9. Diferencia de expresión de gene entre A-C junto con genes exclusivos para A.

5.4 Análisis de enriquecimiento de procesos biológicos.

Los análisis de enriquecimiento de procesos biológicos permitieron obtener aquellos procesos que se relacionan significativamente con los 579 genes previamente identificados. La lista que se presenta muestra los 10 procesos biológicos con alta significancia estadística (Tabla 5).

Tabla 5.

Análisis de enriquecimiento de procesos biológicos para genes de Alzheimer (n=579).

Categoría	Termino	Valor <i>p</i>
GO:0045087	Respuesta inmune innata	7,04E-05
GO:0043280	Regulación positiva de la actividad endopeptidasa de tipo cisteína implicada en el proceso apoptótico	1,38E-04
GO:0051607	Respuesta de defensa a virus	1,81E-04
GO:0043124	Regulación negativa de la quinasa I-kappaB y señalización del factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas (NF-kappaB)	7,22E-04
GO:0050718	Regulación positiva de la secreción de la beta interleucina-1	0,0014
GO:0050728	Regulación negativa de la respuesta inflamatoria	0,0026
GO:0090200	Regulación positiva de liberación de citocromo c de la mitocondria	0,0054
GO:0006952	Respuesta de defensa	0,0068
GO:0050707	Regulación de secreción de citoquinas	0,0092
GO:0032760	Regulación positiva de producción del factor de necrosis tumoral	0,0093

Estudios previos realizados en procesos biológicos relacionados con el Alzheimer nos permitieron obtener una visión más amplia. Pennisi y colaboradores (2016), reportaron la presencia de inflamomas (complejos multiprotéicos del sistema inmune innato) activadores de caspasas proinflamatorias, que producen la activación de las citoquinas proinflamatorias interleucina (IL) -1, IL-18 e IL-33, promoviendo el desarrollo de una neuroinflamación y patologías como el Alzheimer. Así mismo, existen receptores codificados por los genes *TLR2* y *TLR4*, que participan en la respuesta inflamatoria mediada por la microglia, durante la formación de placas de beta amiloide en la enfermedad de Alzheimer (Zhang et al., 2012). Adicionalmente, un estudio reportó una referencia al deterioro en la reparación del ADN por escisión de nucleótidos, mediante la vía (TCR) que podría

desempeñar un papel crítico en trastornos neurodegenerativos crónicos, tales como la enfermedad de Alzheimer, mismo proceso que consta como significativo en el análisis de enriquecimiento expuesto (Sepe et al., 2013).

Receptores para productos finales de glicación avanzada fueron reportados por Tanne y colaboradores (2012), en donde explicaron como dos isoformas proteicas desempeñan papeles importantes en la patogénesis del Alzheimer, ya sea a través de la interacción del péptido beta mRAGE con beta amiloide o la inhibición de la vía de señalización activada por beta mRAGE. En otro ámbito, Shih y colaboradores (2015), reportaron la activación de NF-kappa B, especialmente su activación constitutiva en pacientes de estados inflamatorios crónicos y más aún en la enfermedad de Alzheimer. Procesos biológicos en compartimientos celulares también han sido reportados. Santos y colaboradores (2010), expusieron como el mal funcionamiento mitocondrial es una de las principales causas del inicio de la neurodegeneración en el cerebro de personas con Alzheimer.

En el presente estudio también se reportó la regulación positiva en la producción del factor de necrosis tumoral. Un estudio realizado por Gabbita y colaboradores (2012), mostraron cómo factores de necrosis tumoral alfa son mediadores inflamatorios de la neurodegeneración en pacientes con la enfermedad de Alzheimer. La razón se encuentra en los niveles elevados de citoquinas en el líquido cefalorraquídeo, hipocampo y corteza cerebral.

Por tanto, podemos señalar que los procesos biológicos analizados están relativamente descritos en estudios anteriores. Sin embargo, existen procesos biológicos obtenidos que aún no han sido reportados en relación al Alzheimer. En este apartado, no se presentan las tablas con todos los procesos biológicos significativos, sino aquellas con mayor relevancia para nuestra discusión. El

listado de todos los términos significativos puede ser consultados en los Anexos.

5.5 Construcción y análisis de redes de co-expresión

Para la red de co-expresión se emplearon los 579 genes. En la figura 10, se muestra el dendrograma para la enfermedad de Alzheimer. Cada uno de los módulos tiene un símbolo (MA) dentro del dendrograma, representándose con un color en particular (turquesa, marrón, azul y gris) y agrupando cierta cantidad de genes.

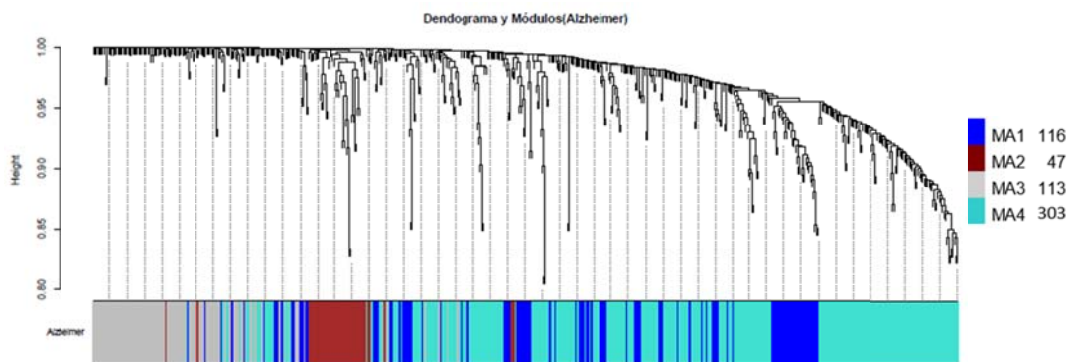


Figura 10. Dendrograma y representaciones de módulos obtenidos en las redes de co-expresión en la enfermedad de Alzheimer (A).

En la figura 11, se muestra cómo las diferencias más altas en expresión se encuentran en los módulos azul, marrón y turquesa. Es así que, en dichos módulos se encuentran genes fuera del comportamiento medio (en especial en el azul y turquesa) de cada módulo con una mayor regulación negativa.

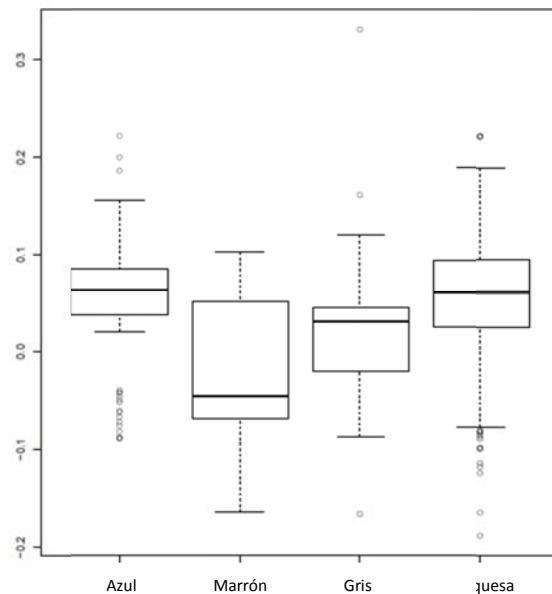


Figura 11. Caja de Tukey de la diferencia de expresión de genes presentes en la enfermedad de Alzheimer junto con valores atípicos.

Con respecto al módulo marrón, la media se encuentra por debajo de cero, es decir el módulo contiene la mayoría de genes que en el Alzheimer fueron sufrieron una regulación negativa. No obstante, la diferencia de expresión no necesariamente implica una mayor o menor importancia de un grupo de genes en específico. Para esto es necesario considerar aspectos como la conectividad (Cruz et al., 2016) (Tejera; Bernardes y Rebelo, 2013).

La conectividad (k) de un gen en un módulo representa la importancia de ese gen en ese módulo. Por lo tanto, mientras más conectado se encuentra un gen, más importante podría ser. En la tabla 6, se muestran los módulos de la red de co-expresión de Alzheimer con los valores de conectividad.

Tabla 6.

Diferentes valores de conectividad en los módulos.

Módulo	N	k Total	k Intramodulo	k Fuera del módulo	k Diferencia
MA1	116	2,25	1,04	1,21	-0,17
MA2	47	0,40	0,23	0,17	0,07
MA3	113	0,04	0,01	0,03	-0,02
MA4	303	3,24	2,78	0,46	2,31

Los módulos azul (MA1) y turquesa (MA4) de la red de co-expresión del Alzheimer son los que muestran una mayor conectividad tanto total como intra-modular. Esto sugiere que estos dos módulos contendrían la mayor cantidad de genes relevantes para la patología, y al mismo tiempo agruparían procesos biológicos específicos.

Paralelamente, para cada uno de los 579 genes, se calculó la distancia topológica. Esta distancia otorga una medida de las diferencias de conectividad de cada uno de los genes en la red control y la red de Alzheimer. De esta manera se evaluó la relevancia de cada gen de una forma independiente. De hecho, como podemos ver en la figura 12, los genes con mayor distancia no necesariamente presentaron mayores o menores valores de diferenciación en cuanto a expresión, pero sí con genes que mostraron una mayor diferencia de conectividades intra-modular con relación a la conectividad total (figura 13).

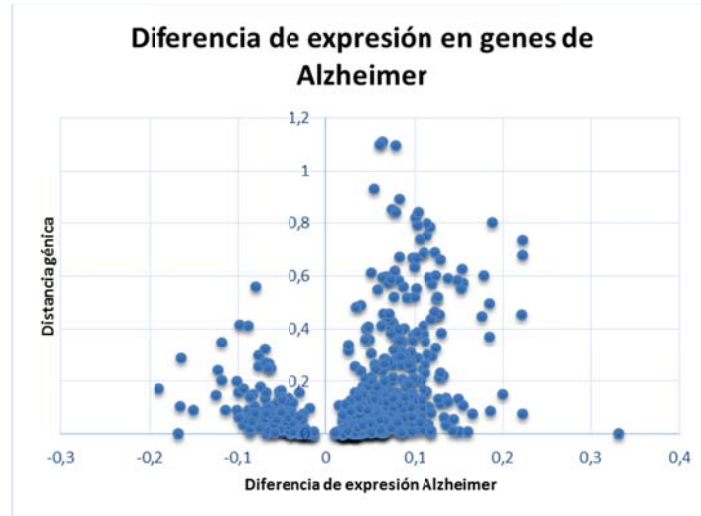


Figura 12. Diferencia de expresión de genes de Alzheimer

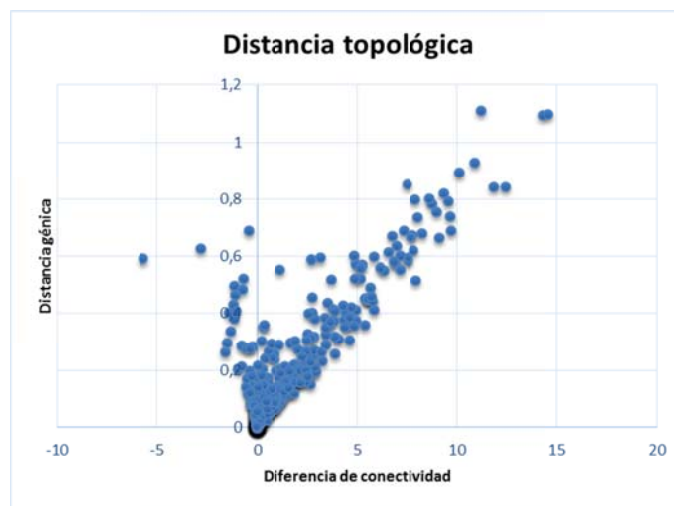


Figura 13. Distancia topológica de genes de Alzheimer y control.

La conectividad media en la red de co-expresión de Alzheimer es mucho mayor ($k=1,68$) que la conectividad media en la red control ($k=0,50$). Esto implica que, en la patología muchos genes aumentan sus relaciones con genes vecinos (en términos funcionales). De esta forma, genes muy conectados en la red patológica, necesariamente implicarán una mayor diferenciación con la red control (como se muestra en la Figura 13). Estos “*hubs*” que a su vez se

encontraron en los módulos MA1 y MA4 serían de gran valor para la descripción biológica (o patogénica) de la enfermedad.

5.6 Análisis de enriquecimiento de procesos biológicos por módulos

Dentro del análisis por módulos se encontraron algunos procesos relacionados con el Alzheimer como se muestran en la tabla 7, 8 y 9.

Tabla 7.

Análisis de enriquecimiento de procesos biológicos para el módulo MA1.

Categoría	Termino	Valor <i>p</i>
GO:0006954	Respuesta inflamatoria	2,53E-04
GO:0035556	Transducción de la señal intracelular	0,0017
GO:0042981	Regulación del proceso apoptótico	0,0045
GO:0006952	Respuesta de defensa	0,0109
GO:0006955	Respuesta inmune	0,0155
GO:0002250	Respuesta inmune adaptativa	0,0159
GO:0002374	Secreción de citoquinas implicada en la respuesta inmune	0,0254
GO:0051607	Respuesta de la defensa al virus	0,0297
GO:0043280	Regulación positiva de la actividad endopeptidasa de tipo cisteína implicada en el proceso apoptótico	0,0416
GO:0032755	Regulación positiva de la producción de interleucina-6	0,0436
GO:0001666	Respuesta a la hipoxia	0,0454
GO:0006468	Fosforilación de proteínas	0,0460
GO:0032570	Respuesta a la progesterona	0,0477
GO:0051591	Respuesta al adenosín monofosfato cíclico	0,0498

Con respecto a procesos apoptóticos, Wang y colaboradores (2012), reportaron un mecanismo por el cual el péptido amiloide induce la apoptosis en astrocitos, contribuyendo significativamente a la neurodegeneración en personas con Alzheimer. Por otro lado, en los procesos de transducción de la señal intracelular, Bot y colaboradores (2011), reportaron la proteína NRXN3 β . Este tipo de señal demostró estar implicada en la regulación de la actividad y la plasticidad de las sinapsis y las vías de transducción.

Tabla 8.

Análisis de enriquecimiento de procesos biológicos para el módulo MA2.

Categoría	Termino	Valor p
GO:0006195	Proceso catabólico de nucleótidos de purina	0,0275
GO:0016569	Modificación covalente de la cromatina	0,0353
GO:0006672	Proceso metabólico de la ceramida	0,0483

Tabla 9.

Análisis de enriquecimiento de procesos biológicos para el módulo MA4.

Categoría	Termino	Valor p
GO:0045087	Respuesta inmune innata	1,51E-07
GO:0051607	Respuesta de defensa a virus	8,74E-04
GO:0050728	Regulación negativa de la respuesta inflamatoria	0,0016
GO:0071222	Respuesta celular a lipopolisacáridos	0,0032
GO:0045944	Regulación positiva de la transcripción del promotor de ARN polimerasa II	0,0035
GO:0043124	Regulación negativa de la quinasa I-kappaB y señalización del factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas (NF-kappaB)	0,0085
GO:0038066	Cascada de proteínas quinasas activadas por mitógenos (P38mapk)	0,0094
GO:0006584	Proceso metabólico de la catecolamina	0,0094
GO:0007165	Transducción de señales	0,0112
GO:0045742	Regulación positiva de la vía de señalización del receptor del factor de crecimiento epidérmico	0,0134
GO:0090200	Regulación positiva de liberación de citocromo c de la mitocondria	0,0229
GO:0050776	Regulación de la respuesta inmune	0,0292
GO:0070935	Estabilización del mRNA mediada por 3'-UTR	0,0367
GO:0009725	Respuesta a hormonas	0,0380
GO:0045444	Diferenciación de células grasas	0,0447

Por último, se ha reportado en estudios, que la ceramida podría participar en la muerte celular neuronal que conduce a la enfermedad de Alzheimer, pero su papel principal sigue siendo desconocido. En el estudio encontraron niveles elevados de Cer16, Cer18, Cer20, y Cer24 en los cerebros de individuos con la enfermedad de Alzheimer (Filippov et al., 2012).

Como se puede observar, básicamente el módulo MA1, corresponde a los procesos inmunes e inflamatorios; el módulo MA4 recoge más específicamente los procesos inmunológicos; mientras que MA2 corresponde a procesos del metabolismo proteico y de estructura nuclear.

5.7 Selección de Rasgos

De los 579 genes identificados previamente, 24 coincidieron con el Modelo 1 y 74 con el Modelo 2. Así, el segundo modelo fue más específico para Alzheimer que el primero. Estos resultados apoyaron la discusión previa de que los 579 son más específicos para Alzheimer. Por esta razón, solo se analizaron los resultados del Modelo 2. Los 20 primeros genes con mayor impacto en el mRMR son: *RPL36AL*, *TNRC6C*, *HLA-DOA*, *DDIT4*, *LRRRC14*, *TCIRG1*, *PPP1R12C*, *ATP6V1E1*, *DCUN1D1*, *NSMCE1*, *ZBTB8OS*, *MRPL51*, *OCRL*, *PIEZO2*, *ZFX*, *SHFM1*, *PCYOX1L*, *TMEM120A*, *PSMD4*, *CDC42SE1*.

En base a estos 20 genes obtenidos, se realizó la búsqueda de estudios que estén en correspondencia con los resultados y sean el respaldo indicado para que los genes se conviertan en biomarcadores útiles en futuras investigaciones. Meng y colaboradores (2016), reportaron el gen *ZFX* como un posible regulador del envejecimiento. El análisis sugirió que los reguladores de la transcripción son los genes más afectados en la enfermedad de Alzheimer, tanto a nivel de metilación del ADN y expresión génica, así como en el proceso de envejecimiento.

Por otra parte, Scheubert y colaboradores (2012), reportaron el gen *PCYOX1L* junto con *LOC283755* y *PTPN3* como genes que necesitan un estudio especial y exhaustivo. Según Cruz y colaboradores (2016), la poca correspondencia de los genes resultantes de la selección de rasgos es la lógica diferente que utiliza, encontrándose genes más distantes desde el punto de vista metabólico. En la tabla se observa que, tanto la distancia topológica como los genes obtenidos por la selección de rasgos, apuntan una vez más a que los módulos MA1 y MA4 presentan la mayor importancia. Es decir, ambos módulos contienen la mayor cantidad de los genes identificados por mRMR. No obstante, debemos observar que estos dos módulos también contienen la mayor cantidad de genes (Tabla 10), por esta razón es necesario hacer un ajuste.

Tabla 10.

Módulos con genes obtenidos mediante selección de rasgos.

Módulos	Distancia Topológica	Cantidad de Genes Identificados por mRMR	Ranking medio (mRMR)	Frecuencia ajustada (mRMR)	Genes
MA1	0,15	24	121.5	1,62	<i>LRRC14</i> <i>NSMCE1</i> <i>CELF2</i> <i>POLR3C</i> <i>PRKCB</i> <i>CARD8</i> <i>C9orf66</i> <i>MTMR4</i> <i>CTSS</i> <i>ARHGAP24</i> <i>VCAN</i> <i>CNOT11</i> <i>KDM3B</i> <i>WDR37</i> <i>PIK3CG</i> <i>AVIL</i> <i>IL10RB</i> <i>NOD2</i> <i>HCG27</i> <i>DUSP1</i> <i>SLC22A4</i> <i>IFNGR2</i> <i>MXD1</i> <i>CD93</i>
MA2	0,08	8	34	1,50	<i>TNRC6C</i> <i>DCUN1D1</i> <i>PSMD4</i> <i>TNFAIP3</i> <i>ERCC5</i>

					SPG11 DPCD PIPSL RSRP1
MA3	0,02	9	58	0,55	DDIT4 PPP1R12C HLA-A C5orf24 FBN2 AHSA1 FAM127B EIF1AY
MA4	0,22	33	106	0,85	TMEM120A CDC42SE1 MARCH6 KLHDC8B DGKD CA4 AURKAIP1 VPREB3 TRAF3IP3 PGS1 ATP6V0B TMX4 H2AFY TBKBP1 SDF4 C2orf68 TPST1 ZNF467 PRKDC PLBD1 CD79B BRI3 HSPA6 S100A11 OPTN CEBPD ALOX5AP CSNK1G2 TRIM25 GATS PLXDC2 SNRPN VNN3

Para esto consideramos la ponderación de la frecuencia de genes encontrados con la cantidad de genes que el mismo módulo posee, de forma semejante a como lo realizó Cruz y colaboradores (2016). Es decir, el valor de “frecuencia ajustada” se define como:

$$Fa = \frac{\left(\frac{n_{mRMR}}{250}\right)}{\left(\frac{n_{MA}}{n_{MA \text{ total}}}\right)} \quad (\text{Ecuación 3})$$

Donde, n_{mRMR} es número de genes comunes con mRMR, n_{MA} corresponde al número de genes en el módulo y $n_{MAtotal}$ es el número de genes en todos los módulos. Este nuevo valor confirma el valor del módulo MA1 y MA2 (azul y marrón) pero decrece el valor del módulo MA4. Este resultado, en conjunto con el análisis modular y el análisis de distancia topológica, sugiere que el módulo MA1 es el de mayor importancia, y que los 24 genes encontrados en ese módulo en común con la selección de rasgos, pueden ser potenciales biomarcadores. Este resultado también sugiere que la respuesta inflamatoria asociada con procesos inmunológicos, tiene una mayor contribución en la enfermedad de Alzheimer que los procesos puramente inmunológicos, más representados en el módulo MA4.

Estos 24 genes son: *LRR14*, *NSMCE1*, *CELF2*, *POLR3C*, *PRKCB*, *CARD8*, *C9orf66*, *MTMR4*, *CTSS*, *ARHGAP24*, *VCAN*, *CNOT11*, *KDM3B*, *WDR37*, *PIK3CG*, *AVIL*, *IL10RB*, *NOD2*, *HCG27*, *DUSP1*, *SLC22A4*, *IFNGR2*, *MXD1* y *CD93*. De los cuales algunos se han reportado en investigaciones previas. En el caso de Gerschutz y colaboradores (2014), expusieron que *PRKCB* y *MAPK1* incrementaron su expresión en el tronco cerebral y el cerebelo en la enfermedad de Alzheimer. Por otro lado, propusieron que las alteraciones en la expresión de estos dos genes, se producen en etapas tempranas de la patogénesis. Además, estipulan que las múltiples vías de transducción son afectadas, dando lugar a la enfermedad de Alzheimer en lugar de un envejecimiento fisiológico normal.

En cambio, Un y colaboradores (2008), mostraron cómo las mujeres de genotipo AA del gen *CARD8* (proteína truncada), tenían un riesgo 2,39 veces mayor de desarrollar la enfermedad de Alzheimer que los sujetos con el genotipo TT del gen *CARD8* (proteína de longitud completa). Sin embargo, en el grupo de hombres era ausente dicho efecto. En otro estudio, Plata y colaboradores (2012), estudiaron el caso del gen *PIK3CG*, gen que está dentro

de un amplio rango genes que incluyen un número previamente vinculado en estudios de expresión génica para la formación de placas beta amiloide en la enfermedad de Alzheimer.

Recientes hallazgos de Chen y colaboradores (2008), demostraron que las funciones del gen *NOD2* en conjunto con *TLR2*, ejercen una regulación positiva en la expresión de receptor del péptido formilo 2 (mFPR2), un receptor acoplado a la proteína G que media la actividad quimiotáctica de patógenos y los péptidos derivados del huésped, incluyendo péptidos como el amiloide beta 42, el factor causante en la enfermedad de Alzheimer.

Las investigaciones de Tseveleki y colaboradores (2010), han ayudado a comprobar la regulación del gen *MXD1* en modelos murinos (ratones). Su regulación negativa sugirió que puede estar implicado en diferentes tipos de neurodegeneración, proponiendo así una nueva diana terapéutica para la inhibición de trastornos del sistema nervioso central.

6. CONCLUSIONES Y RECOMENDACIONES

6.1 Conclusiones

Los genes *PRKCB*, *CARD8*, *PIK3CG*, *NOD2*, *MXD1*, *ZFX* y *PCYOX1L* son siete biomarcadores o dianas terapéuticas potenciales para la enfermedad de Alzheimer identificados por selección de rasgos y/o análisis de co-expresión.

Los genes *LRRC14*, *NSMCE1*, *CELF2*, *POLR3C*, *C9orf66*, *MTMR4*, *CTSS*, *ARHGAP24*, *VCAN*, *CNOT11*, *KDM3B*, *WDR37*, *AVIL*, *IL10RB*, *HCG27*, *DUSP1*, *SLC22A4*, *IFNGR2* y *CD93* priorizados mediante selección de rasgos y análisis de co-expresión, son candidatos completamente desconocidos en relación a la enfermedad de Alzheimer.

La combinación de métodos de selección de rasgos y redes de co-expresión permitieron proponer nuevos genes candidatos relacionados con la enfermedad de Alzheimer que deben ser validados experimentalmente.

La construcción de las redes de co-expresión basadas en casos de enfermedad de Alzheimer, muestran cómo los genes se expresan en forma diferente y cómo su co-expresión puede aumentar el riesgo en el apareamiento de la patología. A esto se adiciona que nuestro análisis sugiere que los procesos inflamatorios e inmunes podrían jugar un papel esencial en la patología.

6.2 Recomendaciones

Llevar a cabo la validación experimental de los genes *RPL36AL*, *TNRC6C*, *HLA-DOA*, *DDIT4*, *LRRC14*, *TCIRG1*, *PPP1R12C*, *ATP6V1E1*, *DCUN1D1*, *NSMCE1*, *ZBTB8OS*, *MRPL51*, *OCRL*, *PIEZO2*, *SHFM1*, *TMEM120A*, *PSMD4*, *CDC42SE1*, *CELF2*, *POLR3C*, *C9orf66*, *MTMR4*, *CTSS*, *ARHGAP24*, *VCAN*, *CNOT11*, *KDM3B*, *WDR37*, *AVIL*, *IL10RB*, *HCG27*, *DUSP1*, *SLC22A4*, *IFNGR2*, *CD93*, *TNFAIP3*, *ERCC5*, *SPG11*, *DPCD*, *PIPSL* y *RSRP1* para comprobar su relación con la enfermedad de Alzheimer .

Realizar técnicas de aprendizaje automático para priorizar aún más la información genética especialmente de aquellos genes identificados mediante la selección de rasgos y evaluar la selectividad y especificidad de cualquier algoritmo de clasificación utilizando estos genes.

REFERENCIAS

- Aguilar, C., Westman, E., Muehlboeck, J.-S., Mecocci, P., Vellas, B., Tsolaki, M., Wahlund, L.O. (2013). Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Research: Neuroimaging*, 212(2). Recuperado el 15 de Octubre de 2016 de <http://dx.doi.org/10.1016/j.pscychresns.2012.11.005>
- Barnes, D. E., & Yaffe, K. (2016). The projected effect of risk factor reduction on Alzheimer's disease prevalence. *The Lancet Neurology*, 10(9). Recuperado el 16 de Octubre de 2016 de [http://dx.doi.org/10.1016/S1474-4422\(11\)70072-2](http://dx.doi.org/10.1016/S1474-4422(11)70072-2) .
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Source Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1). Recuperado el 16 de Octubre de 2016 de <http://www.jstor.org/stable/2346101>
- Blalock, E. M., Chen, K.-C., Sharrow, K., Herman, J. P., Porter, N. M., Foster, T. C., & Landfield, P. W. (2003). Gene Microarrays in Hippocampal Aging: Statistical Profiling Identifies Novel Processes Correlated with Cognitive Impairment. *The Journal of Neuroscience*, 23(9). Recuperado el 18 de Octubre de 2016 de <http://www.jneurosci.org/content/23/9/3807.abstract>.
- Blennow, K. (2010). Biomarkers in Alzheimer's disease drug development. *Nature medicine* (Vol. 16). Recuperado el 19 de Noviembre de 2016 de <https://doi.org/10.1038/nm.2221>
- Blennow, K., Hampel, H., Weiner, M., & Zetterberg, H. (2010). Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nat Rev Neurol*, 6(3). Recuperado el 20 de Octubre de 2016 de <http://dx.doi.org/10.1038/nrneurol.2010.4>.

- Bot, N., Schweizer, C., Ben Halima, S., & Fraering, P. C. (2011). Processing of the synaptic cell adhesion molecule neurexin-3beta by Alzheimer disease alpha- and gamma-secretases. *The Journal of Biological Chemistry*, 286(4). Recuperado el 20 de Octubre de 2016 de <https://doi.org/10.1074/jbc.M110.142521>
- Braak, H., Thal, D. R., Ghebremedhin, E., Del Tredici, K., Nelson, P., Braak, H., Lee, V. (2011). Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. *Journal of Neuro pathology and Experimental Neurology*, 70(11). Recuperado el 20 de Octubre de 2016 de <https://doi.org/10.1097/NEN.0b013e318232a379>
- Britschgi, M., Rufibach, K., Huang, S. L. B., Clark, C. M., Kaye, J. A., Li, G., Wyss-Coray, T. (2011). Modeling of pathological traits in Alzheimer's disease based on systemic extracellular signaling proteome. *Molecular & Cellular Proteomics: MCP*, 10(10). Recuperado el 29 de Octubre de 2016 de <https://doi.org/10.1074/mcp.M111.008862>
- Cai, Z., Zhao, B., & Ratka, A. (2011). Oxidative Stress and β -Amyloid Protein in Alzheimer's Disease. *NeuroMolecular Medicine*, 13(4). Recuperado el 18 de Octubre de 2016 de <http://doi.org/10.1007/s12017-011-8155-9>.
- Cano Guitiérrez, C. (2010). Extracción de conocimiento de microarrays y literatura biomédica para el estudio de la regulación genética. Granada: Editorial de la Universidad de Granada.
- Cavalieri, D., & De Filippo, C. (2005). Bioinformatic methods for integrating whole-genome expression results into cellular networks. *Drug Discovery Today*, 10(10). Recuperado el 18 de Octubre de 2016 de [https://doi.org/10.1016/S1359-6446\(05\)03433-1](https://doi.org/10.1016/S1359-6446(05)03433-1)
- Chalbot, S., Zetterberg, H., Blennow, K., Fladby, T., Andreasen, N., Grundke-Iqbal, I., & Iqbal, K. (2011). Blood-cerebrospinal fluid barrier permeability in Alzheimer's disease. *Journal of Alzheimer's Disease*, 25(3).

Recuperado el 20 de Noviembre de 2016 de <https://doi.org/10.3233/JAD-2011-101959>.

- Chen, K., Zhang, L., Huang, J., Gong, W., Dunlop, N. M., & Wang, J. M. (2008). Cooperation between NOD2 and Toll-like receptor 2 ligands in the up-regulation of mouse mFPR2, a G-protein-coupled A 42 peptide receptor, in microglial cells. *Journal of Leukocyte Biology*, 83(6). Recuperado el 20 de Noviembre de 2016 de <https://doi.org/10.1189/jlb.0907607>
- Clark, C. M., Pontecorvo, M. J., Beach, T. G., Bedell, B. J., Coleman, R. E., Doraiswamy, P. M., Skovronsky, D. M. (2012). Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic B amyloid plaques: A prospective cohort study. *The Lancet Neurology*, 11(8). Recuperado el 20 de Noviembre de 2016 de [https://doi.org/10.1016/S1474-4422\(12\)70142-4](https://doi.org/10.1016/S1474-4422(12)70142-4)
- Cruz-Monteaudo, M., Borges, F., Paz-y-Miño, C., Cordeiro, M. N. D. S., Rebelo, I., Perez-Castillo, Y., Tejera, E. (2016). Efficient and biologically relevant consensus strategy for Parkinson's disease gene prioritization. *BMC Medical Genomics*, 9(1). Recuperado el 20 de Noviembre de 2016 de <https://doi.org/10.1186/s12920-016-0173-x>
- Cummings, J. L. (2011). Biomarkers in Alzheimer's disease drug development. *Alzheimer's & Dementia*, 7(3). Recuperado el 22 de Noviembre de 2016 de <https://doi.org/10.1016/j.jalz.2010.06.004>
- Davis, S., & Meltzer, P. S. (2007). *GEOquery*: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, 23(14). Recuperado el 22 de Noviembre de 2016 de <https://doi.org/10.1093/bioinformatics/btm254>
- Ding, C y Peng, H. (2003). "Minimum Redundancy Feature Selection from Microarray Gene Expression Data". 2nd IEEE Computer Society Bioinformatics Conference (CSB 2003), Stanford, CA, USA.

- Drachman, D. A. (2006). Aging of the brain, entropy, and Alzheimer disease. *Neurology*, 67(8). Recuperado el 22 de Octubre de 2016 de <http://doi.org/10.1212/01.wnl.0000240127.89601.83>.
- Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1). Recuperado el 22 de Noviembre de 2016 de <http://www.ncbi.nlm.nih.gov/pubmed/11752295>
- Ferrer, I. (2012). Defining Alzheimer as a common age-related neurodegenerative process not inevitably leading to dementia. *Progress in Neurobiology*, 97(1). Recuperado el 24 de Noviembre de 2016 de <https://doi.org/10.1016/j.pneurobio.2012.03.005>
- Filippov, V., Song, M. A., Zhang, K., Vinters, H. V, Tung, S., Kirsch, W. M., Duerksen-Hughes, P. J. (2012). Increased ceramide in brains with Alzheimer's and other neurodegenerative diseases. *Journal of Alzheimer's Disease: JAD*, 29(3). Recuperado el 24 de Noviembre de 2016 <https://doi.org/10.3233/JAD-2011-111202>
- Flicker, C., Ferris, S. H., & Reisberg, B. (1991). Mild cognitive impairment in the elderly: predictors of dementia. *Neurology*, 41(7). Recuperado el 26 de Noviembre de 2016 <http://www.ncbi.nlm.nih.gov/pubmed/2067629>
- Fontalba, A., Guti rrez, O., Llorca, J., Mateo, I., Berciano, J., Fern ndez-Luna, J. L., & Combarros, O. (2008). Deficiency of CARD8 Is Associated with Increased Alzheimer's Disease Risk in Women. *Dementia and Geriatric Cognitive Disorders*, 26(3). Recuperado el 26 de Noviembre de 2016 <https://doi.org/10.1159/000160956>.
- Forsberg A, Engler H, Almkvist O, Blomquist G, Hagman G, Wall A, et al. (2008). PET imaging of amyloid deposition in patients with mild cognitive impairment. *Neurobiol Aging*, 29.

- Gabbita, S. P., Srivastava, M. K., Eslami, P., Johnson, M. F., Kobritz, N. K., Tweedie, D., Harris-White, M. E. (2012). Early intervention with a small molecule inhibitor for tumor necrosis factor- α prevents cognitive deficits in a triple transgenic mouse model of Alzheimer's disease. *Journal of Neuroinflammation*, 9. Recuperado el 27 de Noviembre de 2016 de <https://doi.org/10.1186/1742-2094-9-99>
- García, M; Baltá, A; Bello, R; Caballero, Y; Álvarez, D; (2007). Un nuevo algoritmo de selección de rasgos basado en la teoría de los conjuntos aproximados. *Revista Facultad de Ingeniería Universidad de Antioquia*. Recuperado el 27 de Noviembre de 2016 de <http://www.redalyc.org/articulo.oa?id=43004111>.
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, 5(10). Recuperado el 28 de Noviembre de 2016 de <http://genomebiology.com/2004/5/10/R80>
- Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A., & Dudoit, S. (Eds.). (2005). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. New York, NY: Springer New York. Recuperado el 30 de Noviembre de 2016 de <https://doi.org/10.1007/0-387-29362-0>
- Gerschütz, A., Heinsen, H., Grünblatt, E., Wagner, A. K., Bartl, J., Meissner, C., Monoranu, C. M. (2014). Neuron-specific alterations in signal transduction pathways associated with Alzheimer's disease. *Journal of Alzheimer's Disease: JAD*, 40(1). Recuperado el 24 de Noviembre de 2016 de <https://doi.org/10.3233/JAD-131280>
- Ghidoni, R., Paterlini, A., & Benussi, L. (2013). Translational proteomics in Alzheimer's disease and related disorders. *Clinical Biochemistry*, 46(6). Recuperado el 20 de Octubre de 2016 de <http://dx.doi.org/10.1016/j.clinbiochem.2012.10.012>.

- Grossman, M., Miller, B., Dickson, D., & JQ, T. (2001). Clinical and pathological diagnosis of frontotemporal dementia: Report of the work group on frontotemporal dementia and disease. *Archives of Neurology*, 58(11). Recuperado el 20 de Octubre de 2016 de <http://dx.doi.org/10.1001/archneur.58.11.1803>.
- Hampel, H., Bürger, K., Teipel, S. J., Bokde, A. L. W., Zetterberg, H., & Blennow, K. (2008). Core candidate neurochemical and imaging biomarkers of Alzheimer's disease, Madrid, Spain. *Alzheimer's and Dementia*. Recuperado el 1 de Diciembre de 2016 de <https://doi.org/10.1016/j.jalz.2007.08.006>
- Hampel, H., Lista, S., & Khachaturian, Z. S. (2012). Development of biomarkers to chart all Alzheimer's disease stages: The royal road to cutting the therapeutic Gordian Knot. *Alzheimer's and Dementia*. Recuperado el 1 de Diciembre de 2016 de <https://doi.org/10.1016/j.jalz.2012.05.2116>
- Hanchuan Peng, H., Fuhui Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8). Recuperado el 1 de Diciembre de 2016 de <https://doi.org/10.1109/TPAMI.2005.159>
- Henriksen, K., O'Bryant, S. E., Hampel, H., Trojanowski, J. Q., Montine, T. J., Jeromin, A., for the Blood-Based Biomarker Interest Group. (2014). The future of blood-based biomarkers for Alzheimer's disease. *Alzheimer's & Dementia : The nal of the Alzheimer's Association*, 10(1). Recuperado el 21 de Noviembre de 2016 de <http://doi.org/10.1016/j.jalz.2013.01.013>.
- Herrmann, N., Chau, S. A., Kircanski, I., & Lanctôt, K. L. (2011). Current and Emerging Drug Treatment Options for Alzheimer's Disease. *Drugs*, 71(15). Recuperado el 1 de Diciembre de 2016 de <https://doi.org/10.2165/11595870-000000000-00000>

- Hu, W. T., Holtzman, D. M., Fagan, A. M., Shaw, L. M., Perrin, R., Arnold, S. E., Soares, H. (2012). Plasma multianalyte profiling in mild cognitive impairment and Alzheimer disease. *Neurology*, 79(9). Recuperado el 21 de Noviembre de 2016 de <http://doi.org/10.1212/WNL.0b013e318266fa70>
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1). Recuperado el 1 de Diciembre de 2016 de <https://doi.org/10.1093/nar/gkn923>
- Hye, A., Lynham, S., Thambisetty, M., Causevic, M., Campbell, J., Byers, H. L., Lovestone, S. (2006). Proteome-based plasma biomarkers for Alzheimer disease. *Brain*, 129(11). Recuperado el 21 de Octubre de 2016 de <http://brain.oxfordjournals.org/content/129/11/3042.abstract>
- Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9(1). Recuperado el 2 de Diciembre de 2016 de [https://doi.org/10.1016/S1474-4422\(09\)70299-6](https://doi.org/10.1016/S1474-4422(09)70299-6)
- Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)*, 8(1). Recuperado el 1 de Diciembre de 2016 de <https://doi.org/10.1093/biostatistics/kxj037>
- Josephs, K. A., Petersen, R. C., Knopman, D. S., Boeve, B. F., Whitwell, J. L., Duffy, J. R., Dickson, D. W. (2006). Clinicopathologic analysis of frontotemporal and corticobasal degenerations and PSP. *Neurology*, 66(1). Recuperado el 28 de Octubre de 2016 de <http://doi.org/10.1212/01.wnl.0000191307.69661.c3>.

- Klunk, W. E., & Mathis, C. A. (2008). The future of amyloid-beta imaging: a tale of radionuclides and tracer proliferation. *Current Opinion in Neurology*, 21(6). Recuperado el 29 de Noviembre de 2016 de <http://doi.org/10.1006/bbrc.1999.1956>.
- Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., Brazma, A. (2015). ArrayExpress update--simplifying data submissions. *Nucleic Acids Research*, 43(D1). Recuperado el 29 de Noviembre de 2016 de <https://doi.org/10.1093/nar/gku1057>
- Kong, W., Mou, X., Hu, X., 2011. Exploring matrix factorization techniques for Significant genes identification of Alzheimer's disease microarray gene expression data. *BMC Bioinform.* 12, S7.
- Koyama, A., O'Brien, J., Weuve, J., Blacker, D., Metti, A. L., & Yaffe, K. (2013). The Role of Peripheral Inflammatory Markers in Dementia and Alzheimer's Disease: A Meta-Analysis. *The nals of Gerontology Series A: Biological Sciences and Medical Sciences*, 68(4). Recuperado el 22 de Octubre de 2016 de <http://doi.org/10.1093/gerona/gls187>.
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9:559 Recuperado el 29 de Noviembre de 2016 de <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559>
- Langfelder, P., Horvath, S., Fisher, R., Zhou, X., Kao, M., Wong, W., Yandell, B. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1). Recuperado el 29 de Noviembre de 2016 de <https://doi.org/10.1186/1471-2105-9-559>
- Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, 24(5). Recuperado el 30 de Noviembre de 2016 de <https://doi.org/10.1093/bioinformatics/btm563>

- Leek, J.T., Johnson W.E., Parker H.S., Fertig E.J., Jaffe A.E. and Storey J.D. (2016). *sva: Surrogate Variable Analysis*. R package version 3.22.0.
- Li J, Llano DA, Ellis T, LeBlond D, Bhathena A, Jhee SS, et al. (2012). Effect of human cerebrospinal fluid sampling frequency on amyloid-beta levels. *Alzheimers Dement*.
- Liu, J., Ranka, S., & Kahveci, T. (2008). Classification and feature selection algorithms for multi-class CGH data. *Bioinformatics (Oxford, England)*, 24(13). Recuperado el 22 de Octubre de 2016 de <https://doi.org/10.1093/bioinformatics/btn145>
- Lucie, Y., Dwaine, R., & Charles, G. (2012). Brain Amyloid Imaging FDA Approval of Florbetapir F18 Injection. *New England Journal of Medicine*, 367(10). Recuperado el 2 de Noviembre de 2016 de <http://doi.org/10.1056/NEJMp1208061>.
- Marcello, A., Wirths, O., Schneider-Axmann, T., Degerman-Gunnarsson, M., Lannfelt, L., & Bayer, T. A. (2009). Circulating immune complexes of A β and IgM in plasma of patients with Alzheimer's disease. *Journal of Neural Transmission*, 116(7). Recuperado el 2 de Noviembre de 2016 de <http://doi.org/10.1007/s00702-009-0224-y>.
- Mastroeni, D., Grover, A., Delvaux, E., Whiteside, C., Coleman, P. D., & Rogers, J. (2010). Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiology of Aging*, 31(12). Recuperado el 2 de Noviembre de 2016 de <https://doi.org/10.1016/j.neurobiolaging.2008.12.005>
- Mattsson, N., Blennow, K., & Zetterberg, H. (2009). CSF Biomarkers. *Annals of the New York Academy of Sciences*, 1180(1). Recuperado el 2 de Noviembre de 2016 de <http://doi.org/10.1111/j.1749-6632.2009.04944.x>.
- Meng, G., Zhong, X., & Mei, H. (2016). A Systematic Investigation into Aging Related Genes in Brain and Their Relationship with Alzheimer's Disease.

PloS One, 11(3). Recuperado el 10 de Noviembre de 2016 de <https://doi.org/10.1371/journal.pone.0150624>

Miller, J. A., Cai, C., Langfelder, P., Geschwind, D. H., Kurian, S. M., Salomon, D. R., & Horvath, S. (2011). Strategies for aggregating gene expression data: The collapseRows R function. *BMC Bioinformatics*, 12(1). Recuperado el 8 de Noviembre de 2016 de <https://doi.org/10.1186/1471-2105-12-322>

Miller, R. A., Galecki, A., & Shmookler-Reis, R. J. (2001). Interpretation, Design, and Analysis of Gene Array Expression Experiments. *The nals of Gerontology Series A: Biological Sciences and Medical Sciences* , 56(2). Recuperado el 8 de Noviembre de 2016 de <http://doi.org/10.1093/gerona/56.2.B52>.

Morris JC, Roe CM, Grant EA, Head D, Storandt M, Goate AM, et al. Pittsburgh compound B imaging and prediction of progression from cognitive normality to symptomatic Alzheimer disease. *Arch Neurol* 2009.

O'Bryant, S. E., Waring, S. C., Hobson, V., Hall, J. R., Moore, C. B., Bottiglieri, T., Diaz-Arrastia, R. (2010). Decreased C-Reactive Protein Levels in Alzheimer Disease. *nal of Geriatric Psychiatry and Neurology* , 23(1). Recuperado el 14 de Noviembre de 2016 de <http://doi.org/10.1177/0891988709351832>.

Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., Ohta, H. (2007). ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. *Nucleic Acids Research*, 35 (Database). Recuperado el 21 de Noviembre de 2016 de <https://doi.org/10.1093/nar/gkl783>

Pennisi, M., Crupi, R., Di Paola, R., Ontario, M. L., Bella, R., Calabrese, E. J., Calabrese, V. (2016). Inflammasomes, hormesis, and antioxidants in neuroinflammation: Role of NLRP3 in Alzheimer disease. *Journal of*

Neuroscience Research. Recuperado el 19 de Noviembre de 2016 de <https://doi.org/10.1002/jnr.23986>

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., Kokmen, E. (1999). Mild Cognitive Impairment. *Archives of Neurology*, 56(3). Recuperado el 25 de Noviembre de 2016 de <https://doi.org/10.1001/archneur.56.3.303>

Pickering, E. (2012). Plasma biomarkers associated with the apolipoprotein e genotype and alzheimer disease. *Archives of Neurology*, 69(10). Recuperado el 8 de Noviembre de 2016 de <http://dx.doi.org/10.1001/archneurol.2012.1070>

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Recuperado el 29 de Noviembre de 2016 de <http://www.R-project.org/>.

Ray, S., Britschgi, M., Herbert, C., Takeda-Uchimura, Y., Boxer, A., Blennow, K., Wyss-Coray, T. (2007). Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nature Medicine*, 13(11). Recuperado el 3 de Diciembre de de 2016 de <https://doi.org/10.1038/nm1653>

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." *Nucleic Acids Research*, 43(7).

Romero, F. (2013). Un Caso Práctico en Biología Molecular de Sistemas: Análisis de Redes de Co-expresión Génica. Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Sevilla. Recuperado el 14 de Noviembre de 2016 de https://www.cs.us.es/~fran/curso_unia/red.html.

Scheubert, L., Luštrek, M., Schmidt, R., Repsilber, D., & Fuellen, G. (2012). Tissue-based Alzheimer gene expression markers-comparison of multiple machine learning approaches and investigation of redundancy in

small biomarker sets. *BMC Bioinformatics*, 13, 266. Recuperado el 14 de Noviembre de 2016 de <https://doi.org/10.1186/1471-2105-13-266>

Schneider, P., Hampel, H., & Buerger, K. (2009). Biological marker candidates of Alzheimer's disease in blood, plasma, and serum. *CNS Neuroscience & Therapeutics*, 15(4). Recuperado el 2 de Diciembre de 2016 de <https://doi.org/10.1111/j.1755-5949.2009.00104>.

Sepe, S., Payan-Gomez, C., Milanese, C., Hoeijmakers, J. H., & Mastroberardino, P. G. (2013). Nucleotide excision repair in chronic neurodegenerative diseases. *DNA Repair*, 12(8). Recuperado el 2 de Diciembre de 2016 de <https://doi.org/10.1016/j.dnarep.2013.04.009>

Sherman BT, Huang DW, Tan Q, Guo Y, Bour S, Liu D, Stephens R, Baseler MW, Lane HC, Lempicki RA (2007). "DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis". *BMC Bioinformatics*. 8: 426. Recuperado el 14 de Noviembre de 2016 de [doi:10.1186/1471-2105-8-426](https://doi.org/10.1186/1471-2105-8-426).

Shih, R.-H., Wang, C.-Y., & Yang, C.-M. (2015). NF-kappaB Signaling Pathways in Neurological Inflammation: A Mini Review. *Frontiers in Molecular Neuroscience*. Recuperado el 16 de Noviembre de 2016 de <https://doi.org/10.3389/fnmol.2015.00077>

Shumaker, S. D., LeBlanc, S. K., Delaney, P., Marquie-Beck, J., Ueland, S., Ellis, R. J. (2011). Incidence of Post-Dural Puncture Headache in Research Volunteers. *Headache: The Journal of Head and Face Pain*, 51(10). Recuperado el 30 de Noviembre de 2016 de <https://doi.org/10.1111/j.1526-4610.2011.01959>

Silver, M., Janousova, E., Hua, X., Thompson, P. M., Montana, G., & Alzheimer's Disease Neuroimaging Initiative. (2012). Identification of gene pathways implicated in Alzheimer's disease using longitudinal imaging phenotypes with sparse regression. *NeuroImage*, 63(3).

Recuperado el 30 de Noviembre de 2016 de <https://doi.org/10.1016/j.neuroimage.2012.08.002>

Smyth, G. K. (2005). Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pages.

Smyth, G.K. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor. (2005). Bioinformatics and Computational Biology Solutions Using R and Bioconductor. New York, NY: Springer New York.

Soares, H. D., Potter, W. Z., Pickering, E., Kuhn, M., Immermann, F. W., Shera, D. M., Shaw, L. M. (2012). Plasma biomarkers associated with the apolipoprotein E genotype and Alzheimer disease. *Archives of Neurology*, 69(10). Recuperado el 2 de Diciembre de 2016 de <https://doi.org/10.1001/archneurol.2012.1070>

Sood, S., Gallagher, I. J., Lunnon, K., Rullman, E., Keohane, A., Crossland, H., Timmons, J. A. (2015). A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome Biology*, 16(1). Recuperado el 2 de Diciembre de 2016 de <https://doi.org/10.1186/s13059-015-0750-x>

Tamames, J., Clark, D., Herrero, J., Dopazo, J., Blaschke, C., Fernández, J. M., Valencia, A. (2002). Bioinformatics methods for the analysis of expression arrays: data clustering and information extraction. *Journal of Biotechnology*. Recuperado el 2 de Diciembre de 2016 de <http://www.ncbi.nlm.nih.gov/pubmed/12141992>

Tanne, J. H. (2012). US scientists discuss early detection and treatment of Alzheimer's disease. *Bmj*, 344. Recuperado el 1 de Diciembre de 2016 de <https://doi.org/10.1136/bmj.e1068>

Tejera, E., Bernardes, J., & Rebelo, I. (2013). Co-expression network analysis and genetic algorithms for gene prioritization in preeclampsia. *BMC*

Medical Genomics, 6(1). Recuperado el 3 de Diciembre de <https://doi.org/10.1186/1755-8794-6-51>

Thambisetty, M., & Lovestone, S. (2010). Blood-based biomarkers of Alzheimer's disease: challenging but feasible. *Biomarkers in Medicine*, 4(1). Recuperado el 3 de Diciembre de 2016 de <https://doi.org/10.2217/bmm.09.84>

Thambisetty, M., Simmons, A., Velayudhan, L., & et al. (2010). Association of plasma clusterin concentration with severity, pathology, and progression in alzheimer disease. *Archives of General Psychiatry*, 67(7). Recuperado el 25 de Noviembre de 2016 de <http://dx.doi.org/10.1001/archgenpsychiatry.2010.78>.

Thambisetty, M., Tripaldi, R., Riddoch-Contreras, J., Hye, A., An, Y., Campbell, J., Resnick, S. M. (2010). Proteome-based plasma markers of brain B amyloid deposition in non-demented older individuals. *Journal of Alzheimer's Disease*, 22(4). Recuperado el 26 de Noviembre de 2016 de <https://doi.org/10.3233/JAD-2010-101350>

Tseveleki, V., Rubio, R., Vamvakas, S.-S., White, J., Taoufik, E., Petit, E., Probert, L. (2010). Comparative gene expression analysis in mouse models for multiple sclerosis, Alzheimer's disease and stroke for identifying commonly regulated and disease-specific gene changes. *Genomics*, 96(2). Recuperado el 26 de Noviembre de 2016 de <https://doi.org/10.1016/j.ygeno.2010.04.004>

Yang, L., Rieves, D., & Ganley, C. (2012). Brain Amyloid Imaging — FDA Approval of Florbetapir F18 Injection. *New England Journal of Medicine*, 367(10). Recuperado el 29 de Noviembre de 2016 de <http://doi.org/10.1056/NEJMp1208061>.

Zhang, B, Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. Article17.

Ziemann, U., Wahl, M., Hattingen, E., & Tumani, H. (2011). Development of biomarkers for multiple sclerosis as a neurodegenerative disorder. *Progress in Neurobiology*, 95(4). Recuperado el 29 de Noviembre de 2016 de <http://dx.doi.org/10.1016/j.pneurobio.2011.04.007>.

Zipser, B. D., Johanson, C. E., Gonzalez, L., Berzin, T. M., Tavares, R., Hulette, C. M., Stopa, E. G. (2016). Microvascular injury and blood-brain barrier leakage in Alzheimer's disease. *Neurobiology of Aging*, 28(7). Recuperado el 18 de Noviembre de 2016 de <http://doi.org/10.1016/j.neurobiolaging.2016.05.016>.

ANEXOS

Anexo 1. Tablas con los valores de expresión y significación estadística obtenidos mediante Limma para los 579 genes específicos de Alzheimer-Control

GENEID	logFC	Valor p	Valor p (FDR)	GENEID	logFC	Valor p	Valor p (FDR)
11335	-0,093	5,95E-11	3,05E-09	6843	0,043	2,28E-04	1,65E-03
53917	0,104	1,17E-09	4,32E-08	1912	0,056	2,30E-04	1,66E-03
533	0,121	1,17E-09	4,32E-08	4689	0,102	2,44E-04	1,76E-03
10636	0,096	3,11E-09	1,02E-07	29781	0,020	2,45E-04	1,77E-03
1052	0,184	7,71E-09	2,26E-07	59342	0,082	2,46E-04	1,77E-03
10598	-0,081	1,16E-08	3,17E-07	8651	0,044	2,48E-04	1,78E-03
29123	0,054	1,26E-08	3,41E-07	22934	-0,132	2,48E-04	1,78E-03
253018	0,181	1,44E-08	3,84E-07	57763	0,061	2,49E-04	1,79E-03
8664	-0,119	1,49E-08	3,94E-07	4201	-0,043	2,51E-04	1,80E-03
10623	-0,072	1,90E-08	4,93E-07	51266	-0,087	2,55E-04	1,82E-03
22900	0,145	2,15E-08	5,53E-07	84961	0,049	2,55E-04	1,82E-03
51186	-0,031	2,70E-08	6,74E-07	8778	0,092	2,56E-04	1,82E-03
51380	0,077	4,19E-08	9,90E-07	51222	0,027	2,57E-04	1,83E-03
3276	-0,084	5,93E-08	1,34E-06	7421	0,024	2,57E-04	1,83E-03
23299	0,095	7,09E-08	1,57E-06	10284	-0,041	2,61E-04	1,85E-03
9422	-0,078	7,11E-08	1,57E-06	286006	0,061	2,68E-04	1,90E-03
6001	-0,152	7,23E-08	1,60E-06	57053	0,032	2,72E-04	1,92E-03
80315	0,081	8,25E-08	1,79E-06	54998	0,072	2,76E-04	1,95E-03
29082	-0,089	8,95E-08	1,93E-06	9586	0,092	2,78E-04	1,96E-03
197370	-0,092	9,02E-08	1,94E-06	5594	0,050	2,87E-04	2,01E-03
5093	0,117	9,30E-08	1,98E-06	112616	0,055	2,87E-04	2,01E-03
25798	0,134	1,56E-07	3,16E-06	55002	0,065	2,87E-04	2,02E-03
55122	0,125	1,58E-07	3,20E-06	2730	0,033	2,94E-04	2,06E-03
151903	-0,090	1,66E-07	3,34E-06	3659	0,116	2,98E-04	2,08E-03
8621	0,053	1,72E-07	3,43E-06	4287	-0,031	3,00E-04	2,09E-03
49856	0,066	1,79E-07	3,55E-06	9252	0,067	3,02E-04	2,11E-03
3423	0,073	2,05E-07	4,02E-06	3755	-0,028	3,08E-04	2,14E-03
10659	0,061	2,32E-07	4,45E-06	81844	0,079	3,14E-04	2,17E-03
285172	0,085	3,21E-07	5,90E-06	5510	-0,047	3,15E-04	2,18E-03
116369	0,063	3,88E-07	6,95E-06	83596	0,030	3,16E-04	2,18E-03
5710	-0,099	4,04E-07	7,19E-06	6844	0,082	3,21E-04	2,21E-03
171017	0,052	4,16E-07	7,32E-06	971	-0,074	3,21E-04	2,21E-03
8527	0,034	4,24E-07	7,46E-06	9334	0,121	3,22E-04	2,21E-03
3931	0,031	4,94E-07	8,52E-06	10610	0,080	3,24E-04	2,23E-03
83538	-0,116	4,98E-07	8,56E-06	1486	0,083	3,28E-04	2,25E-03
138151	0,072	5,09E-07	8,75E-06	8659	0,012	3,29E-04	2,25E-03
1520	0,146	5,39E-07	9,19E-06	123041	0,041	3,29E-04	2,25E-03
2357	0,189	5,86E-07	9,86E-06	1850	0,054	3,33E-04	2,28E-03
131408	0,063	6,80E-07	1,13E-05	3842	0,028	3,35E-04	2,29E-03
9489	0,125	7,29E-07	1,19E-05	5094	0,062	3,38E-04	2,31E-03
83440	0,069	7,57E-07	1,24E-05	6515	0,120	3,43E-04	2,34E-03
8675	0,063	8,30E-07	1,34E-05	9185	0,064	3,43E-04	2,34E-03
134553	0,051	9,54E-07	1,51E-05	55653	-0,048	3,45E-04	2,35E-03
79746	0,126	9,73E-07	1,54E-05	7057	0,080	3,48E-04	2,37E-03
54442	0,051	1,02E-06	1,60E-05	64221	0,034	3,48E-04	2,37E-03
22884	0,058	1,25E-06	1,90E-05	219771	0,052	3,52E-04	2,39E-03
84245	0,066	1,29E-06	1,95E-05	3034	0,073	3,55E-04	2,41E-03
83463	0,075	1,35E-06	2,01E-05	51409	0,028	3,55E-04	2,41E-03
2130	0,049	1,41E-06	2,10E-05	286254	0,032	3,68E-04	2,49E-03
867	0,106	1,47E-06	2,18E-05	2114	0,079	3,77E-04	2,54E-03
64965	-0,074	1,54E-06	2,28E-05	55689	0,042	3,83E-04	2,57E-03
23514	0,057	1,56E-06	2,29E-05	6240	-0,074	3,87E-04	2,60E-03
83716	0,224	1,61E-06	2,36E-05	4150	0,020	3,93E-04	2,63E-03

5294	0,071	1,66E-06	2,42E-05	6604	0,039	3,93E-04	2,63E-03
10677	0,051	1,75E-06	2,53E-05	6184	0,059	3,94E-04	2,63E-03
1490	-0,021	1,93E-06	2,76E-05	7852	0,082	3,94E-04	2,63E-03
10813	-0,044	1,98E-06	2,82E-05	153577	0,037	4,07E-04	2,71E-03
10299	0,070	2,25E-06	3,17E-05	25855	-0,056	4,14E-04	2,75E-03
154467	-0,089	2,36E-06	3,29E-05	29062	0,037	4,14E-04	2,75E-03
94097	0,066	2,49E-06	3,45E-05	25832	0,032	4,15E-04	2,76E-03
51499	-0,068	2,56E-06	3,53E-05	55646	-0,062	4,16E-04	2,76E-03
51311	0,095	2,60E-06	3,58E-05	241	0,119	4,17E-04	2,77E-03
84270	0,083	2,69E-06	3,68E-05	4084	0,121	4,18E-04	2,77E-03
1462	0,210	2,74E-06	3,74E-05	197259	0,103	4,19E-04	2,78E-03
11314	0,103	2,91E-06	3,94E-05	90522	0,062	4,21E-04	2,79E-03
266971	-0,086	2,93E-06	3,97E-05	151636	0,050	4,21E-04	2,79E-03
4051	0,128	3,01E-06	4,05E-05	11252	0,084	4,24E-04	2,80E-03
27180	0,058	3,01E-06	4,05E-05	79934	0,040	4,25E-04	2,81E-03
9147	-0,062	3,19E-06	4,26E-05	54872	0,034	4,38E-04	2,88E-03
11027	0,170	3,19E-06	4,26E-05	23258	0,085	4,42E-04	2,90E-03
1843	0,193	3,34E-06	4,44E-05	280655	-0,029	4,49E-04	2,94E-03
84885	0,066	3,60E-06	4,73E-05	26270	0,082	4,51E-04	2,95E-03
56882	0,102	3,62E-06	4,75E-05	10970	0,109	4,66E-04	3,04E-03
27034	0,030	3,65E-06	4,78E-05	79027	0,020	4,66E-04	3,04E-03
200942	0,131	3,69E-06	4,82E-05	4191	0,070	4,67E-04	3,04E-03
27439	0,132	3,77E-06	4,91E-05	29108	0,065	4,68E-04	3,05E-03
6817	0,049	3,80E-06	4,94E-05	821	0,063	4,71E-04	3,07E-03
79135	-0,041	3,81E-06	4,96E-05	1084	0,085	4,73E-04	3,08E-03
8870	0,104	4,00E-06	5,15E-05	2535	0,059	4,76E-04	3,09E-03
3105	0,314	4,07E-06	5,24E-05	51493	-0,055	4,76E-04	3,09E-03
3937	0,066	4,40E-06	5,60E-05	7045	0,112	4,82E-04	3,13E-03
3134	0,112	4,59E-06	5,80E-05	4035	0,026	4,83E-04	3,14E-03
22918	0,160	4,65E-06	5,87E-05	9394	0,019	4,90E-04	3,17E-03
51550	-0,028	4,83E-06	6,06E-05	7263	0,099	4,96E-04	3,20E-03
64207	0,176	4,92E-06	6,17E-05	80342	-0,075	5,01E-04	3,23E-03
7097	0,095	5,07E-06	6,32E-05	23173	-0,059	5,06E-04	3,26E-03
102157402	-0,029	5,10E-06	6,34E-05	9129	0,060	5,08E-04	3,27E-03
57639	-0,059	5,29E-06	6,52E-05	55628	0,014	5,08E-04	3,27E-03
51693	-0,054	5,78E-06	7,05E-05	201475	0,043	5,11E-04	3,28E-03
91662	0,124	5,87E-06	7,13E-05	51564	0,019	5,14E-04	3,29E-03
100128537	0,069	5,91E-06	7,17E-05	308	0,101	5,18E-04	3,32E-03
1455	0,187	6,05E-06	7,33E-05	7357	0,059	5,19E-04	3,32E-03
219931	0,079	6,36E-06	7,67E-05	2783	0,067	5,24E-04	3,35E-03
10577	0,111	6,49E-06	7,81E-05	3081	-0,025	5,24E-04	3,35E-03
84458	0,062	7,05E-06	8,41E-05	6942	0,020	5,25E-04	3,35E-03
10077	0,067	8,12E-06	9,56E-05	1185	0,021	5,27E-04	3,36E-03
51072	0,051	8,14E-06	9,58E-05	10658	0,024	5,33E-04	3,39E-03
83862	0,092	8,22E-06	9,64E-05	7942	0,039	5,43E-04	3,44E-03
9846	0,078	8,58E-06	1,00E-04	158248	0,044	5,45E-04	3,45E-03
155061	0,117	8,87E-06	1,03E-04	5591	0,025	5,45E-04	3,45E-03
8500	0,048	8,90E-06	1,03E-04	128646	0,043	5,46E-04	3,46E-03
641	-0,035	9,11E-06	1,05E-04	79887	0,107	5,46E-04	3,46E-03
55571	-0,074	9,49E-06	1,09E-04	54853	0,045	5,49E-04	3,47E-03
9555	0,058	9,76E-06	1,12E-04	6310	0,045	5,50E-04	3,48E-03
143	0,090	9,99E-06	1,14E-04	440574	-0,018	5,55E-04	3,50E-03
8972	0,063	1,08E-05	1,22E-04	8334	0,120	5,57E-04	3,51E-03
84446	0,117	1,26E-05	1,40E-04	65005	-0,034	5,61E-04	3,54E-03
126074	0,061	1,32E-05	1,47E-04	201799	0,129	5,68E-04	3,57E-03
152007	0,095	1,33E-05	1,47E-04	2218	0,065	5,73E-04	3,60E-03

1378	0,049	1,37E-05	1,52E-04	64781	0,035	5,76E-04	3,62E-03
7100	0,137	1,39E-05	1,53E-04	84830	-0,047	5,80E-04	3,64E-03
6890	0,135	1,48E-05	1,62E-04	9463	0,029	5,82E-04	3,65E-03
2073	-0,077	1,48E-05	1,62E-04	54902	-0,068	5,82E-04	3,65E-03
9110	0,070	1,59E-05	1,72E-04	4097	0,019	5,83E-04	3,65E-03
51150	-0,053	1,62E-05	1,75E-04	387	0,102	5,86E-04	3,67E-03
54899	0,046	1,63E-05	1,76E-04	4856	0,091	5,92E-04	3,69E-03
169841	0,032	1,64E-05	1,77E-04	54682	0,084	5,95E-04	3,71E-03
3920	0,105	1,65E-05	1,77E-04	57222	0,099	6,00E-04	3,74E-03
6789	0,078	1,67E-05	1,79E-04	6923	-0,055	6,03E-04	3,75E-03
6840	0,072	1,68E-05	1,80E-04	57580	0,071	6,06E-04	3,76E-03
29802	-0,172	1,73E-05	1,85E-04	84941	0,097	6,07E-04	3,77E-03
25911	-0,071	1,74E-05	1,85E-04	7391	0,067	6,09E-04	3,78E-03
23383	0,076	1,82E-05	1,92E-04	5236	0,066	6,14E-04	3,81E-03
7305	0,102	1,89E-05	1,99E-04	6772	0,115	6,16E-04	3,82E-03
6638	-0,063	1,95E-05	2,04E-04	399967	0,031	6,28E-04	3,88E-03
4676	-0,083	1,97E-05	2,06E-04	26253	0,070	6,37E-04	3,93E-03
54541	0,134	1,99E-05	2,07E-04	11068	0,044	6,39E-04	3,94E-03
10133	-0,118	1,99E-05	2,07E-04	5699	0,070	6,41E-04	3,95E-03
4043	0,057	2,00E-05	2,08E-04	23191	0,032	6,45E-04	3,97E-03
8926	-0,057	2,03E-05	2,11E-04	3735	-0,042	6,55E-04	4,02E-03
818	0,086	2,04E-05	2,12E-04	153443	-0,044	6,61E-04	4,05E-03
79042	0,113	2,08E-05	2,15E-04	80019	0,074	6,62E-04	4,05E-03
2852	0,044	2,11E-05	2,17E-04	54920	0,063	6,68E-04	4,08E-03
54165	0,097	2,12E-05	2,19E-04	81565	0,072	6,91E-04	4,21E-03
57136	0,116	2,13E-05	2,19E-04	6046	0,062	7,09E-04	4,31E-03
23492	0,080	2,22E-05	2,28E-04	317	0,056	7,09E-04	4,31E-03
352954	0,090	2,24E-05	2,29E-04	79837	0,038	7,15E-04	4,34E-03
171389	0,029	2,31E-05	2,36E-04	762	0,144	7,17E-04	4,35E-03
9736	0,051	2,32E-05	2,36E-04	23151	0,049	7,18E-04	4,35E-03
8178	0,098	2,37E-05	2,40E-04	5610	0,134	7,23E-04	4,38E-03
55778	0,051	2,37E-05	2,41E-04	4814	0,110	7,26E-04	4,39E-03
80071	-0,036	2,45E-05	2,46E-04	9766	0,081	7,36E-04	4,44E-03
64748	0,121	2,58E-05	2,58E-04	58191	0,057	7,37E-04	4,45E-03
2091	-0,105	2,65E-05	2,64E-04	23248	0,039	7,38E-04	4,45E-03
55374	0,038	2,69E-05	2,67E-04	314	0,012	7,43E-04	4,48E-03
7707	0,086	2,74E-05	2,71E-04	27036	0,031	7,49E-04	4,50E-03
3490	0,092	2,82E-05	2,78E-04	55144	-0,039	7,53E-04	4,52E-03
26071	0,027	2,84E-05	2,80E-04	1537	0,060	7,57E-04	4,54E-03
9094	0,048	2,84E-05	2,80E-04	387893	0,029	7,58E-04	4,54E-03
23064	0,058	2,84E-05	2,80E-04	25979	0,037	7,66E-04	4,58E-03
168544	0,152	2,85E-05	2,80E-04	79957	0,020	7,77E-04	4,65E-03
91748	0,040	2,94E-05	2,87E-04	3574	-0,015	7,88E-04	4,70E-03
581	-0,038	2,96E-05	2,88E-04	93622	-0,051	7,91E-04	4,72E-03
359948	0,043	2,96E-05	2,88E-04	728524	0,014	7,93E-04	4,73E-03
974	-0,152	3,01E-05	2,92E-04	387103	-0,026	8,04E-04	4,78E-03
51738	0,066	3,05E-05	2,95E-04	5654	0,028	8,12E-04	4,83E-03
56666	0,150	3,06E-05	2,96E-04	5089	0,068	8,14E-04	4,84E-03
57609	0,074	3,10E-05	2,98E-04	84080	0,039	8,18E-04	4,86E-03
23401	0,138	3,11E-05	3,00E-04	432	0,083	8,27E-04	4,91E-03
51130	0,031	3,20E-05	3,07E-04	9658	0,035	8,30E-04	4,92E-03
10404	0,081	3,27E-05	3,13E-04	109	0,061	8,30E-04	4,93E-03
6945	0,030	3,27E-05	3,13E-04	5159	0,044	8,49E-04	5,03E-03
483	0,054	3,43E-05	3,26E-04	7125	-0,021	8,67E-04	5,12E-03
54863	0,062	3,47E-05	3,29E-04	387700	0,055	8,73E-04	5,15E-03
22864	0,033	3,60E-05	3,39E-04	9043	0,047	8,79E-04	5,19E-03

57690	-0,045	3,61E-05	3,41E-04	388799	0,020	8,92E-04	5,25E-03
7706	0,121	3,64E-05	3,43E-04	9308	-0,037	9,01E-04	5,30E-03
1774	0,029	3,66E-05	3,44E-04	132864	0,030	9,10E-04	5,35E-03
81619	0,060	4,00E-05	3,72E-04	6799	0,032	9,14E-04	5,37E-03
912	0,107	4,10E-05	3,81E-04	644	-0,058	9,15E-04	5,37E-03
84679	-0,036	4,35E-05	4,01E-04	55252	0,072	9,17E-04	5,38E-03
339105	0,034	4,43E-05	4,07E-04	10215	-0,076	9,32E-04	5,46E-03
5922	0,061	4,64E-05	4,24E-04	55741	0,050	9,33E-04	5,47E-03
6583	0,109	4,68E-05	4,28E-04	1018	0,018	9,47E-04	5,53E-03
1432	0,049	4,98E-05	4,53E-04	578	0,042	9,54E-04	5,57E-03
285613	0,033	5,01E-05	4,56E-04	402055	-0,109	9,61E-04	5,61E-03
9410	-0,062	5,05E-05	4,58E-04	29765	0,021	9,64E-04	5,62E-03
57035	0,094	5,28E-05	4,78E-04	151516	0,083	9,66E-04	5,62E-03
57820	-0,024	5,32E-05	4,81E-04	1521	0,118	9,67E-04	5,63E-03
10226	0,120	5,40E-05	4,87E-04	6720	0,029	9,67E-04	5,63E-03
284207	0,028	5,42E-05	4,88E-04	4217	0,071	9,70E-04	5,64E-03
85444	-0,037	5,51E-05	4,94E-04	3568	-0,034	9,71E-04	5,64E-03
5226	0,132	5,55E-05	4,98E-04	56670	0,018	9,92E-04	5,75E-03
51759	-0,167	5,57E-05	4,98E-04	85495	0,038	1,02E-03	5,90E-03
123803	-0,065	5,59E-05	5,00E-04	55350	0,093	1,03E-03	5,97E-03
9189	0,079	5,63E-05	5,02E-04	9530	0,026	1,04E-03	5,99E-03
9086	-0,133	5,68E-05	5,05E-04	80781	0,044	1,04E-03	5,99E-03
3096	0,041	5,75E-05	5,12E-04	8543	0,051	1,05E-03	6,02E-03
145165	-0,058	5,76E-05	5,12E-04	83478	0,028	1,05E-03	6,04E-03
27334	-0,045	5,79E-05	5,14E-04	5250	0,038	1,05E-03	6,05E-03
4299	0,037	6,01E-05	5,30E-04	152006	0,028	1,06E-03	6,10E-03
55827	-0,064	6,09E-05	5,37E-04	3455	0,032	1,07E-03	6,14E-03
435	0,024	6,23E-05	5,46E-04	57533	0,074	1,07E-03	6,14E-03
10955	0,039	6,37E-05	5,56E-04	10019	0,068	1,07E-03	6,15E-03
3303	0,201	6,44E-05	5,61E-04	64342	0,059	1,08E-03	6,17E-03
3399	-0,094	6,58E-05	5,71E-04	119	-0,025	1,08E-03	6,21E-03
54776	0,033	6,82E-05	5,90E-04	9183	-0,041	1,09E-03	6,24E-03
116541	-0,081	6,91E-05	5,98E-04	90378	0,032	1,09E-03	6,25E-03
4521	-0,054	7,04E-05	6,07E-04	2353	0,129	1,11E-03	6,31E-03
4850	-0,018	7,10E-05	6,11E-04	4691	-0,052	1,11E-03	6,31E-03
5329	0,103	7,22E-05	6,20E-04	9768	0,025	1,13E-03	6,42E-03
9684	0,054	7,26E-05	6,23E-04	51079	-0,064	1,16E-03	6,60E-03
55577	0,101	7,44E-05	6,36E-04	26511	0,072	1,17E-03	6,64E-03
2153	0,103	7,50E-05	6,40E-04	23070	0,024	1,17E-03	6,65E-03
56255	0,089	7,56E-05	6,44E-04	51279	0,069	1,18E-03	6,68E-03
23259	-0,027	7,57E-05	6,45E-04	90317	-0,022	1,18E-03	6,70E-03
51780	0,067	7,63E-05	6,49E-04	5930	0,018	1,19E-03	6,73E-03
10376	0,091	7,78E-05	6,62E-04	22938	-0,038	1,19E-03	6,73E-03
388969	0,068	7,91E-05	6,71E-04	56907	-0,033	1,19E-03	6,74E-03
56683	-0,053	7,96E-05	6,74E-04	684959	0,058	1,21E-03	6,80E-03
126017	-0,027	8,03E-05	6,77E-04	81030	0,104	1,21E-03	6,82E-03
27173	0,059	8,20E-05	6,89E-04	6282	0,111	1,21E-03	6,84E-03
5170	0,029	8,24E-05	6,92E-04	56851	0,049	1,22E-03	6,88E-03
8668	-0,072	8,44E-05	7,07E-04	388555	0,023	1,24E-03	6,96E-03
115992	0,059	8,51E-05	7,12E-04	678	0,089	1,24E-03	6,97E-03
8460	0,126	8,61E-05	7,20E-04	51267	0,030	1,24E-03	6,97E-03
10910	-0,039	8,69E-05	7,26E-04	27125	0,042	1,24E-03	6,98E-03
196527	0,062	9,22E-05	7,65E-04	196	0,044	1,25E-03	7,01E-03
84898	0,096	9,29E-05	7,70E-04	8115	-0,043	1,25E-03	7,02E-03
9755	0,037	9,35E-05	7,74E-04	6248	-0,032	1,26E-03	7,03E-03
286075	0,031	9,42E-05	7,79E-04	9910	-0,036	1,28E-03	7,14E-03

706	0,087	9,47E-05	7,83E-04	64332	0,047	1,28E-03	7,16E-03
196441	0,052	9,57E-05	7,90E-04	4170	0,084	1,30E-03	7,22E-03
5702	-0,059	9,85E-05	8,10E-04	5978	-0,014	1,30E-03	7,23E-03
217	0,098	9,90E-05	8,14E-04	64180	0,079	1,30E-03	7,25E-03
23231	-0,096	1,00E-04	8,21E-04	9907	0,030	1,30E-03	7,25E-03
9903	0,074	1,01E-04	8,25E-04	29916	0,032	1,30E-03	7,26E-03
23095	0,051	1,01E-04	8,29E-04	10981	0,075	1,31E-03	7,28E-03
146880	0,022	1,01E-04	8,31E-04	51734	0,102	1,31E-03	7,28E-03
1831	0,080	1,02E-04	8,37E-04	2204	0,053	1,32E-03	7,32E-03
154141	0,023	1,03E-04	8,39E-04	348793	-0,031	1,32E-03	7,33E-03
22978	0,102	1,03E-04	8,40E-04	353116	0,052	1,32E-03	7,34E-03
7326	0,040	1,04E-04	8,45E-04	84432	-0,020	1,35E-03	7,48E-03
23528	0,121	1,05E-04	8,54E-04	54994	-0,048	1,36E-03	7,53E-03
9274	-0,025	1,07E-04	8,66E-04	79366	-0,023	1,37E-03	7,54E-03
6281	0,102	1,08E-04	8,75E-04	221496	0,042	1,39E-03	7,64E-03
9895	0,034	1,10E-04	8,86E-04	4763	0,027	1,40E-03	7,71E-03
3805	0,090	1,10E-04	8,88E-04	51754	0,025	1,41E-03	7,78E-03
3588	0,078	1,12E-04	9,06E-04	79022	-0,040	1,43E-03	7,87E-03
8349	0,115	1,15E-04	9,26E-04	115286	-0,017	1,45E-03	7,94E-03
3460	0,107	1,16E-04	9,30E-04	25844	0,079	1,45E-03	7,97E-03
3310	0,114	1,18E-04	9,42E-04	4318	0,204	1,46E-03	7,99E-03
132241	0,025	1,19E-04	9,53E-04	200030	0,031	1,46E-03	8,00E-03
777	0,039	1,23E-04	9,78E-04	121355	-0,039	1,49E-03	8,12E-03
7265	-0,054	1,25E-04	9,94E-04	57655	0,072	1,50E-03	8,19E-03
2358	0,161	1,29E-04	1,02E-03	2002	0,036	1,52E-03	8,27E-03
5730	0,166	1,31E-04	1,03E-03	5768	0,038	1,52E-03	8,29E-03
4615	0,085	1,38E-04	1,08E-03	7284	0,058	1,52E-03	8,29E-03
93129	0,026	1,40E-04	1,10E-03	55696	0,053	1,53E-03	8,30E-03
10288	0,052	1,42E-04	1,11E-03	64127	0,107	1,53E-03	8,31E-03
57649	0,027	1,42E-04	1,11E-03	81928	0,029	1,53E-03	8,32E-03
11065	0,032	1,48E-04	1,15E-03	10541	-0,070	1,55E-03	8,39E-03
220323	0,082	1,48E-04	1,15E-03	6041	0,055	1,55E-03	8,43E-03
7128	0,064	1,50E-04	1,16E-03	1659	0,040	1,55E-03	8,43E-03
80008	-0,040	1,53E-04	1,19E-03	5937	0,038	1,56E-03	8,45E-03
80208	0,083	1,56E-04	1,21E-03	2161	0,036	1,59E-03	8,61E-03
1662	-0,046	1,57E-04	1,21E-03	55600	-0,084	1,60E-03	8,65E-03
7538	0,110	1,57E-04	1,21E-03	348235	-0,031	1,60E-03	8,65E-03
64802	0,032	1,59E-04	1,22E-03	65979	0,045	1,61E-03	8,66E-03
3112	-0,117	1,59E-04	1,22E-03	5899	0,093	1,61E-03	8,70E-03
79188	0,092	1,60E-04	1,23E-03	23254	0,022	1,62E-03	8,72E-03
11262	-0,047	1,63E-04	1,25E-03	115727	0,042	1,62E-03	8,74E-03
8530	0,147	1,66E-04	1,27E-03	283	0,057	1,63E-03	8,76E-03
116540	-0,067	1,67E-04	1,27E-03	57178	0,086	1,63E-03	8,77E-03
79932	0,080	1,68E-04	1,28E-03	8841	0,037	1,63E-03	8,79E-03
4485	0,040	1,71E-04	1,30E-03	1612	0,048	1,64E-03	8,81E-03
22990	0,074	1,72E-04	1,31E-03	55281	0,115	1,65E-03	8,87E-03
114876	0,032	1,75E-04	1,33E-03	81873	-0,054	1,66E-03	8,90E-03
51202	-0,047	1,76E-04	1,33E-03	4860	-0,076	1,67E-03	8,93E-03
10859	0,052	1,79E-04	1,35E-03	54107	-0,060	1,67E-03	8,94E-03
55959	0,090	1,81E-04	1,36E-03	23560	-0,064	1,69E-03	9,00E-03
54386	-0,063	1,86E-04	1,39E-03	23387	0,070	1,70E-03	9,08E-03
117289	0,075	1,86E-04	1,39E-03	2201	0,049	1,71E-03	9,09E-03
8993	0,200	1,87E-04	1,40E-03	10516	-0,042	1,72E-03	9,15E-03
10460	0,097	1,87E-04	1,40E-03	2971	-0,088	1,72E-03	9,15E-03
22849	0,027	1,88E-04	1,41E-03	56254	-0,057	1,75E-03	9,25E-03
7277	0,132	1,97E-04	1,47E-03	114548	0,048	1,76E-03	9,32E-03

4841	0,043	2,06E-04	1,53E-03	55106	0,030	1,77E-03	9,36E-03
8772	0,065	2,07E-04	1,53E-03	51614	-0,047	1,79E-03	9,43E-03
963	0,066	2,08E-04	1,54E-03	171423	-0,037	1,79E-03	9,43E-03
4141	0,050	2,10E-04	1,55E-03	28511	0,019	1,79E-03	9,46E-03
4686	-0,044	2,15E-04	1,58E-03	22933	0,027	1,80E-03	9,48E-03
55784	0,088	2,16E-04	1,59E-03	10409	0,126	1,82E-03	9,59E-03
1376	0,054	2,17E-04	1,59E-03	6348	0,030	1,84E-03	9,68E-03
10552	0,056	2,18E-04	1,60E-03	157983	0,034	1,85E-03	9,72E-03
5894	0,090	2,22E-04	1,63E-03	8934	-0,062	1,85E-03	9,75E-03
5008	0,057	2,24E-04	1,63E-03	5579	0,049	1,86E-03	9,77E-03
84295	-0,012	2,24E-04	1,63E-03	94103	-0,080	1,87E-03	9,81E-03
8693	0,056	2,26E-04	1,65E-03				
10073	-0,045	2,27E-04	1,65E-03				
9798	0,060	2,28E-04	1,65E-03				

Anexo 2. Tabla de análisis de enriquecimiento de procesos biológicos de Alzheimer (n=579).

Categoría	Termino	Valor p
GO:0045087	Respuesta inmune innata	7,04E-05
GO:0043280	Regulación positiva de la actividad endopeptidasa de tipo cisteína implicada en el proceso apoptótico	1,38E-04
GO:0051607	Respuesta de defensa a virus	1,81E-04
GO:0043124	Regulación negativa de I-kappab kinase / NF-kappab señalización	7,22E-04
GO:0050718	Regulación positiva de secreción de la beta interleucina-1	0,0014
GO:0050728	Regulación negativa de la respuesta inflamatoria	0,0026
GO:0090200	Regulación positiva de la liberación de citocromo c a partir de la mitocondria	0,0054
GO:0006952	Respuesta de defensa	0,0068
GO:0050707	Regulación de la secreción de citoquinas	0,0092
GO:0032760	Regulación positiva de la producción de factor de necrosis tumoral	0,0093
GO:2000766	Regulación negativa de la traducción citoplasmática	0,0098
GO:0007165	Transducción de señales	0,0102
GO:1900153	Regulación positiva del proceso catabólico de mRNA transcrito por la radiación nuclear, decadencia dependiente de la muerte	0,0119
GO:0050776	Regulación de la respuesta inmune	0,0128
GO:0016045	Detección de bacterias	0,0150
GO:1902043	Regulación positiva de la vía de señalización extrínseca apoptótica y vía receptores del dominio de muerte celular	0,0150
GO:0006954	Respuesta inflamatoria	0,0152
GO:0032757	Regulación positiva de la producción de interleucina-8	0,0163
GO:0045944	Regulación positiva de la transcripción del promotor de ARN polimerasa II	0,0167
GO:0043065	Regulación positiva del proceso apoptótico	0,0175
GO:0071222	Respuesta celular a lipopolisacáridos	0,0183
GO:0032755	Regulación positiva de la producción de interleucina-6	0,0220
GO:0030168	Activación plaquetaria	0,0220
GO:0071385	Respuesta celular al estímulo glucocorticoide	0,0225
GO:2000353	Regulación positiva del proceso apoptótico de las células endoteliales	0,0225
GO:0002430	Vía de señalización mediada por receptor de complemento	0,0232
GO:0045892	Regulación negativa de la transcripción, modelo de ADN	0,0256
GO:0032570	Respuesta a la progesterona	0,0267

GO:0035556	Transducción de la señal intracelular	0,0277
GO:0009636	Respuesta a sustancias tóxicas	0,0279
GO:0006955	Respuesta inmune	0,0305
GO:0034605	Respuesta celular al calor	0,0309
GO:0051092	Regulación positiva de la actividad del factor de transcripción de NF-kappa	0,0309
GO:2001243	Regulación negativa de la vía de señalización apoptótica intrínseca	0,0316
GO:0038066	Cascada de proteínas quinasas activadas por mitógenos (P38mapk)	0,0316
GO:0039694	Replicación del genoma viral del ARN	0,0316
GO:0006584	Proceso metabólico de la catecolamina	0,0316
GO:0006919	Activación de la actividad endopeptidasa de tipo cisteína implicada en el proceso apoptótico	0,0339
GO:0031663	Lipopolisacárido mediado por la vía de señalización	0,0384
GO:0045444	Diferenciación de las células grasas	0,0407
GO:0002675	Regulación positiva de la respuesta inflamatoria aguda	0,0410
GO:0010818	Quimiotaxis de células T	0,0410
GO:0097011	Respuesta celular al estímulo de colonias de macrófagos de granulocitos	0,0410
GO:1901216	Regulación positiva de la muerte neuronal	0,0423
GO:0045862	Regulación positiva de la proteólisis	0,0423
GO:0006935	Quimiotaxis	0,0465
GO:0097190	Ruta de señalización apoptótica	0,0465

