

Nombre del autor: Ariel Ignacio Valencia Capetillo.

Título del TFG: Interacción Epi-Genética entre los genes *PM20D1* y *CST3* como factor de riesgo ante la enfermedad de Alzheimer.

Nombre de los tutores: Jose Vicente Sánchez Mut y Asia Fernández Carvajal.

Departamento: Instituto de Neurociencias (IN).

Área al que está adscrita la realización del trabajo: Genética.

Titulación: Grado en Biotecnología.

Facultad: Facultad de Ciencias Experimentales.

Curso académico: 2026.



Agradecimientos

En primer lugar, quisiera agradecer a mi tutor Jose Vicente Sanchez Mut, por haberme abierto las puertas de su laboratorio y haberme guiado en toda la investigación.

Igualmente quisiera agradecer a mi tutora Asia Fernández Carvajal, que aun sin vernos en persona, siempre estuvo atenta a mis avances y ayudándome para que saliese bien.

Gracias a mi padre Cristian Eduardo Valencia Morales por haberme ayudado tanto para que pudiese tener la oportunidad de estar aquí y realizar esta investigación.

Gracias a mi madre Estrella Margarita Capetillo Sepulveda por de igual forma ayudarme a poder estar aquí y preocuparse de que estuviera bien mientras investigaba.

Gracias a mi familia en general, que todos siempre me apoyaron y me dieron ánimos de venir a hacer mi investigación.

Gracias a mi director de carrera John Quiñones Díaz, que siempre estuvo atento y me ayudó en todos los temas administrativos para que pudiese estar investigando.

Gracias a mis compañeros de laboratorio, que siempre fueron tan amables conmigo y espero poder seguir investigando un poco más con ellos.

Gracias a mis compañeros de movilidad, quienes me apoyaron y se preocuparon de que estuviera siempre bien mientras investigaba.

Y, por último, gracias tanto a la Universidad de La Frontera como a la Universidad Miguel Hernández por haberme dado los espacios para esta investigación.

Resumen

La enfermedad de Alzheimer es una patología multifactorial asociada a distintas variantes genéticas. El gen *PM20D1* es un locus de rasgo cuantitativo de metilación y expresión asociado al riesgo de desarrollar la enfermedad, cuyo lípido derivado (N-oleoil-leucina) regula positivamente la expresión de la Cistatina C (*CST3*), otro gen implicado en el desarrollo de la enfermedad. Ambos genes han mostrado asociaciones significativas con la enfermedad en estudios genéticos de gen candidato, sin embargo, estas asociaciones individuales no superan el umbral estadístico en estudios de asociación de genoma completo (GWAS, por sus siglas en inglés). Nuestra hipótesis postula que existe una interacción epistática entre ambos genes que enmascara el efecto individual y que su consideración conjunta muestra un efecto sinérgico sobre la carga amiloide, característica en esta enfermedad. Para esto, se analizaron datos de metilación ($n = 2085$), y genéticos ($n = 2462$), donde se identificó una epistasis significativa sobre el biomarcador de Alzheimer, ratio $A\beta_{42/40}$ ($p = 0.0012$), tras ser ajustada por covariables. Estos resultados apoyan la hipótesis de una única ruta metabólica compartida, viéndose limitados en estudios GWAS al evaluar efectos individuales. En conclusión, estos resultados aportan nuevas consideraciones en interacciones no lineales claves para rescatar la heredabilidad oculta, mostrando nuevas perspectivas para la estratificación en medicina de precisión y diseño de paneles diagnósticos de cribado tempranos.

Abstract

Alzheimer's disease is a multifactorial pathology associated with various genetic variants. The *PM20D1* gene is a quantitative methylation trait locus associated with the risk of developing the disease. Its derived lipid (N-oleoyl-leucine) positively regulates the expression of cystatin C (*CST3*), another gene implicated in the development of the disease. Both genes have shown significant associations with the disease in candidate gene studies; however, these individual associations do not exceed the statistical threshold in genome-wide association studies (GWAS). Our hypothesis postulates that there is an epistatic interaction between both genes that masks their individual effects, and that their combined effect shows a synergistic effect on amyloid burden, a characteristic feature of this disease. To this end, methylation ($n = 2085$) and genetic ($n = 2462$) data were analyzed, revealing a significant epistasis on the Alzheimer's biomarker, the $A\beta_{42/40}$ ratio ($p = 0.0012$), after adjustment for covariates. These results support the hypothesis of a single shared metabolic pathway, which is limited in GWAS studies when evaluating individual effects. In conclusion, these results offer new insights into key nonlinear interactions for uncovering hidden heritability, providing new perspectives for stratification in precision medicine and the design of early screening diagnostic panels.

Palabras claves: Alzheimer, PM20D1, Cistatina C, Genética, Epistasis.

INDICE

Introducción

- La enfermedad de Alzheimer
- La Genética del Alzheimer
- El gen *PM20D1*
- La N-oleoil-leucina (OLE)
- El gen *CST3*
- Bases de datos y bioinformática

Antecedentes y objetivos

- Antecedentes e hipótesis
- Objetivos

Materiales y métodos

- Búsqueda de datos
- Epigenética de *PM20D1* y *CST3*
- Metilación de *PM20D1* según la edad
- Regresión cuantílica

Resultados

- Frecuencia alélica de *PM20D1* en tejido cerebral
- Frecuencia alélica de *PM20D1* en sangre total
- Distribución genética de *CST3*
- Epigenética cronológica de *PM20D1*
- Interacción genética entre *PM20D1* y *CST3*

Discusión

- Metilación de *PM20D1*
- Estabilidad cronológica de *PM20D1*
- Epigenética de *CST3*
- Interacción entre *PM20D1* y *CST3*
- Limitaciones de la investigación

Conclusiones y proyecciones futuras

- Conclusiones
- Proyecciones futuras

INTRODUCCIÓN

La enfermedad de Alzheimer

La demencia es caracterizada por un declive significativo de las habilidades cognitivas para realizar las actividades diarias de una persona ([Westphal Filho et al., 2025](#)). La enfermedad de Alzheimer (EA) es el tipo más común de demencia, asociada a un alto deterioro cognitivo ([Liu et al., 2024](#)).

La incidencia de esta enfermedad, su carga social y su carga económica aumenta cada día, planteando importantes problemas médicos y sociales ([Ren et al., 2022](#)). Se cree que para 2050 la prevalencia de la demencia será el doble en Europa y el triple en todo el mundo, lo cual es 3 veces peor si se considera la definición biológica en lugar de la clínica ([Scheltens et al., 2021](#))

Esta enfermedad se caracteriza por la deposición extracelular de péptidos beta-amiloides (A β) mal plegados y agregados, y por la formación ovillos interneuronales compuestos por tau hiperfosforilada ([Heneka et al., 2025](#)). Además de esto, patológicamente hay una pérdida en la sinapsis y neuronas, distrofia neurítica, alteraciones vasculares y respuestas inflamatorias medidas por microglía y astrocitos ([Blumenfeld et al., 2024](#)).

Los síntomas cognitivos de esta enfermedad más comunes incluyen problemas en la memoria a corto plazo, disfunción ejecutiva y visoespacial, y praxis, aunque se han identificado variantes de la enfermedad en la que la memoria se conserva relativamente bien ([Apostolova, 2016](#)). Actualmente, los biomarcadores para esta enfermedad están categorizados en biomarcadores core 1 y core 2, para determinar si se está en un etapa temprana o tardía, respectivamente ([Pichet Binette et al., 2025](#)).

La genética del Alzheimer

El riesgo de desarrollar Alzheimer está altamente influenciado por la genética, en la que más de 75 loci han sido asociado a la enfermedad ([Xue et al., 2024](#)). La primera evidencia de una etiología genética de esta enfermedad se obtuvo debido a observaciones de familias con varias generaciones afectadas por Alzheimer temprano ([Goate et al., 1991](#)).

Al igual que en otras enfermedades, el descubrimiento de las causas genéticas de esta enfermedad podría identificar los mecanismos subyacentes y dar lugar a posibles tratamientos ([Reitz et al., 2023](#)).

Han sido caracterizadas principalmente dos formas la Alzheimer, la temprana y la tardía. La temprana, se da en individuos menores a 65 años, y se caracteriza por mutaciones en los genes PSEN1, PSEN2, y APP ([Bellenguez et al., 2020](#)). La tardía, se da en individuos mayores a 65 años, siendo considerada su mayor factor de riesgo, la edad ([Andrade-Guerrero et al., 2023](#)).

En 1993 fue reportado que, tanto en la forma temprana como tardía de la enfermedad, el alelo épsilon 4 del gen de la apolipoproteína E (APOE4) está altamente asociado con el riesgo de desarrollar Alzheimer ([Strittmatter et al., 1993](#)).

Recientemente, estudios de asociaciones de genoma completo (GWAS) han permitido identificar una suma de genes asociados con la enfermedad. Sin embargo, la relación de estos loci con el riesgo de desarrollar la enfermedad de Alzheimer aún no es bien descrita ([Andrade-Guerrero et al., 2023](#)).

El gen *PM20D1*

La peptidasa con dominio M20 número 1 (*PM20D1*) una enzima secretada, que cataliza tanto la síntesis como la hidrólisis de aminoácidos N-acilados (NAA) a partir de ácidos grasos y aminoácidos libres (Long et al., 2016). Aunque inicialmente se identificó como un factor liberado por adipocitos marrones y beige, en humanos, es altamente expresada en páncreas y piel.

Así mismo, se ha demostrado que su expresión ocurre en todas las regiones anatómicas y linajes celulares del cerebro ([Garro-Núñez et al., 2022](#)), en donde es co-secretada junto complejos lipoproteicos, que vehiculizan e impulsan la biosíntesis y actividad de los NAAs ([Kim et al., 2020](#)), siendo relacionada a varias enfermedades neurodegenerativas y a la regulación negativa de la muerte neuronal ([Villa & Stocco, 2022](#); [Y. Yang et al., 2024](#)).

Dentro de estas enfermedades, en el Alzheimer se ha descrito que una serie de SNPs (rs1172198, rs708727, rs823082, rs823088, rs1361754, y rs960603) que regulan directamente sus niveles de metilación y tasa de transcripción, consolidando a dichas variantes genéticas en locus de rasgos cuantitativos de metilación y expresión (mQTLs y eQTLs, respectivamente) ([Sanchez-Mut et al., 2018, 2020](#)).

La N-oleoil-leucina (OLE)

A nivel metabólico, los aminoácidos N-acilados (NAAs) sintetizados por *PM20D1*, participan en la regulación del gasto energético del organismo, en la reducción de la adiposidad y en la optimización de la homeostasis de la glucosa ([Alves Chagas et al., 2025](#); [Kitaba et al., 2025](#); [Long et al., 2018](#)).

Dentro de los NAAs, la N-oleoil-leucina (OLE) en circulación se ha visto una fuertemente correlacionada con los niveles de *PM20D1* en humanos ([R. Yang et al., 2022](#)). De igual forma, OLE ha demostrado una capacidad notable para ayudar en la pérdida peso, mejorar la intolerancia a la glucosa y aumentar el gasto metabólico basal en modelos murinos ([Robbins et al., 2026](#); [Sivaramakrishna et al., 2025](#); [Tanzo et al., 2023](#)).

Además de esto, estudios recientes han relacionado a OLE con una mejora de la enfermedad de Alzheimer, induciendo el reclutamiento y la asociación de las células de la microglía hacia las placas de beta-amiloide ($A\beta$), promoviendo su fagocitosis y aclaramiento ([Pozzi-Ruiz et al., 2026](#)). Interesantemente, el tratamiento con OLE aumenta los niveles de expresión de la cistatina C (*CST3*) la cual ha demostrado ser suficiente para activar la fagocitosis ab en un estudio reciente independiente (X. Li et al., 2026).

El gen *CST3*

El gen *CST3*, codifica para la cistatina C, un inhibidor de cisteína de tipo 2 que se caracteriza por encontrarse tanto en el espacio intracelular como en el extracelular ([Zi & Xu, 2018](#)). Aunque *CST3* ha sido estudiado por más de 40 años ([Newman, 2002](#)), y se

emplee como marcador de la función renal (Taglieri et al., 2009; Q.-P. Wang et al., 2009), su papel en el sistema nervioso ha adquirido relevancia en la biología actual.

La función biológica de la Cistatina C radica en la protección celular frente la actividad proteolítica de las peptidasas de cisteína endógenas, participando activamente en los mecanismos de respuesta y reparación ante lesiones cerebrales. Si bien algunos de sus aspectos siguen en estudio, su síntesis y secreción en células gliales en respuesta a la deposición amiloide sustenta su implicación en la patogénesis de la enfermedad de Alzheimer ([Mathews & Levy, 2016](#)).

Diversas investigaciones han demostrado una estrecha relación entre los niveles de *CST3* y la progresión de la enfermedad principalmente bajo los valores de los péptidos beta amiloide ($A\beta$), teniendo la capacidad de unirse físicamente a estos (Li et al., 2026) . En concordancia con esto, se ha documentado que las concentraciones en el líquido cefalorraquídeo (LCR) de pacientes diagnosticados con Alzheimer se encuentran significativamente disminuidas en comparaciones con sujetos sanos ([Bhuiya et al., 2026](#); [Kaur & Levy, 2012](#)).

Estudios recientes han identificado *CST3* como el gen que más incrementa su expresión tras el tratamiento con OLE (Pozzi-Ruiz et al., 2026) y, dado que OLE es sintetizado por *PM20D1*, y que tanto *PM20D1* (Sanchez-Mut et al., 2018) como *CST3* (Wang et al., 2021) se han relacionado con la enfermedad de Alzheimer, ambos genes podrían formar parte de una ruta molecular compartida, y mostrar una relación epistática con la enfermedad.

Bases de datos y bioinformática

La bioinformática y la biología computacional tienen un alto potencial para desvelar asociaciones biológicas y patrones ocultos en redes complejas de genes y proteínas ([Niu et al., 2025](#)). En el caso de la enfermedad de Alzheimer, existen miles de datos ómicos guardados en distintos repositorios independientes que exigen estrategias capaces de armonizar todos estos datos desde metilación del DNA y variabilidad genética hasta biomarcadores fisiológicos procedentes de diversas cohortes internacionales.

Para los análisis epigenéticos, se recorrió al internacional *Gene Expression Omnibus* (GEO), el cual alberga perfiles epigenéticos públicos ideales para análisis masivos ([Clough & Barrett, 2016](#)). Por otra parte, la caracterización genética y biomarcadores, se sustentó por el consorcio ADNI (*Alzheimer's Disease Neuroimaging initiative*) ([Veitch et al., 2024](#)), y la cohorte de la Fundación Pasqual Maragall gestionada por el *Barcelona Brain Research Center* (BBRC).

La armonización de estas bases de datos no solo potencia la estadística global, si no también permite cruzar datos de metilación, datos genéticos y biomarcadores, ofreciendo la posibilidad de realizar análisis robustos para ver la interacción y relación entre los genes *PM20D1* y *CST3* con la enfermedad de Alzheimer.

ANTECEDENTES Y OBJETIVOS

Antecedentes e hipótesis

PM20D1 produce OLE, que incrementa los niveles de *CST3* ([Pozzi-Ruiz et al., 2026](#)), y aunque ambos genes han sido asociados con mecanismos de protección neuronal ([Kaeser et al., 2007](#)) su interacción genética aún permanece incierta, por lo que nuestra hipótesis es que, “La variabilidad en la metilación y la interacción genética entre las variantes de *PM20D1* y *CST3* modula el riesgo de desarrollar la enfermedad de Alzheimer a través de un efecto sinérgico que altera los niveles de beta-amiloide, más allá de su contribución individual.”

Objetivos

Para evaluar la hipótesis planteada, el objetivo general de esta investigación consistió en “Analizar la metilación e interacción genética entre las variantes de riesgo de *PM20D1* y *CST3* en la enfermedad de Alzheimer mediante análisis bioinformáticos.”

Para cumplir con el objetivo general, se determinó el objetivo específico primero de “Caracterizar la dinámica epigenética de *PM20D1* y *CST3*, evaluando las frecuencias de sus alelos de riesgo y la deriva epigenética asociada al envejecimiento en muestras

Alzheimer y controles sanos.” y el objetivo específico segundo de “Identificar y cuantificar la existencia de interacciones epistáticas entre las variantes de riesgo de *PM20D1* y *CST3*, en la enfermedad de Alzheimer, mediante el biomarcador $A\beta_{42}/A\beta_{40}$.”

MATERIALES Y MÉTODOS

Búsqueda de datos

El presente estudio se basó en análisis de datos de metilación y genéticos humanos obtenidos en bases de datos de acceso público y restringido. Para los análisis de metilación, se realizó una búsqueda de datos epigenéticos centrada en la base de datos *Gene Expression Omnibus* (GEO) del *National Center for Biotechnology Information* (NCBI). La estrategia de búsqueda empleó las palabras claves “*Alzheimer’s disease*”, “*Alzheimer*”, “*Methylation*”, “*Human*”, “*Epigenomic*” y “*Cognitive decline*”.

Para mitigar el sesgo técnico, los conjuntos de datos se sometieron a criterios de inclusión en donde se buscaron estudios en humanos con un mínimo de muestras únicas de $n \geq 10$; perfiles procesados por plataformas de microarrays de metilación Illumina HumanMethylation27K, HumanMethylation450K o MethylationEPIC 850K; muestras provenientes de tejido cerebral o tejido periférico (sangre total); y disponibilidad de metadatos clínicos en las muestras diagnosticados con la enfermedad de Alzheimer o controles sanos.

Tras aplicar estos filtros, se seleccionaron un total 14 estudios independientes depositados GEO, que cumplieran con los criterios de inclusión. El conjunto de datos armonizados correspondió a un total de 2085 muestras únicas, distribuidas en 963 casos con la enfermedad de Alzheimer (46%), y 1122 controles sanos (54%). Estas muestras han incluido la especificidad del tejido, la plataforma de microarrays y la ascendencia poblaciones, como se detalla en la Tabla 1.

Tabla 1. Resumen de especificaciones de los estudios independientes utilizados en los análisis de metilación.

Nº	GSE	Sonda	Tejido	Plataforma	n	Población
1	GSE45775	cg14893161	Hipocampo	27k	10	Europa

2	GSE59685	cg14893161	Corteza prefrontal	450k	84	Europa
3	GSE76105	cg14893161	Lóbulo temporal	450k	68	USA
4	GSE80970	cg14893161	Corteza prefrontal	450k	142	Europa
5	GSE57361	cg14893161	Corteza prefrontal	450k	15	Europa
6	GSE153712	cg14893161	Sangre	850k	632	Oceania
7	GSE226298	cg14893161	Sangre	850k	41	Europa
8	GSE125895	cg14893161	Corteza prefrontal	450k	68	USA
9	GSE156984	cg14893161	Lóbulo temporal	850k	127	USA
10	GSE105109	cg14893161	Corteza Entorrinal	450k	92	Europa
11	GSE134379	cg14893161	Lóbulo temporal	450k	402	USA
12	GSE144858	cg14893161	Sangre	450k	189	Europa
13	GSE66351	cg14893161	Corteza prefrontal	450k	63	Europa
14	GSE109627	cg14893161	Lóbulo temporal	450k	153	Europa

Para los análisis de interacción genética se accedió a la base de datos la Fundación Pasqual Maragall (FPM), la cual nos proporcionó un tamaño muestral total de 2462 muestras únicas de ascendencia europea con genotipados para *PM201* y *CST3*, así como niveles de los biomarcadores péptido amiloide β_{42} ($A\beta_{42}$), amiloide β_{40} ($A\beta_{40}$), proteína tau fosforilada en el residuo 181 (pTau₁₈₁) y cadena ligera de neurofilamento (NFL).

Para confirmar la interacción genética descubierta en la base de datos PMF, se accedió a la base de datos *Alzheimer's Disease Neuroimaging Initiative* (ADNI) a través del portal *Laboratory of Neuro Imaging* (LONI), en donde las muestras fueron filtradas a través de las diferentes fases del consorcio (ADNI 1, ADNI GO, ADNI 2, y ADNI 3) obteniendo un total de 1117 muestras únicas para la confirmación. Estas muestras fueron

seleccionadas tras poseer los datos de los genotipos y cuantificaciones de biomarcadores requeridos para la confirmación de la interacción.

Epigenética de *PM20D1* y *CST3*

Los datos de metilación tanto de *PM20D1* como de *CST3*, extraídos desde GEO, fueron procesados en el entorno estadístico R (versión 3.6.0) eliminando duplicados y muestras con perfiles clínicos incompletos. Con el fin de mitigar la heterogeneidad biológica, las muestras de *PM20D1* se estratificaron y se analizaron de manera independiente según tejido cerebral ($n = 1223$) o sangre total ($n = 862$).

Para identificar los niveles de metilación de *PM20D1* se usó la sonda cg14893161 y para *CST3* se usaron las sondas cg07041999, cg16589663 y cg25161868. Al tratar a estos loci como loci rasgos cuantitativos (mQTLs), se implementó una inferencia genotípica mediante el algoritmo de particionamiento no supervisado *k-means* ($k = 3$) sobre los niveles de metilación de las sondas correspondientes.

Para mitigar el efecto lote (batch effect), el algoritmo *k-means* se ejecutó independientemente dentro de cada conjunto de datos antes de su armonización, permitiendo clasificar las muestras en homocigoto protector (GG, baja metilación), heterocigoto (AG, metilación intermedia) y homocigoto de riesgo (AA, alta metilación). Para aquellas muestras de pacientes sin registros de sus edades, se calcularon sus edades biológicas según su reloj epigenético de Hannum's ([Hannum et al., 2013](#)).

A partir de estas inferencias genotípicas, se analizaron las distribuciones mediante matrices de contingencia de dos por tres (2×3) y pruebas de chi cuadrado ([Pearson, 1992](#)) de independencia (χ^2). Como control de calidad, las muestras control sanas fueron evaluadas con las proporciones mendelianas teóricas mediante pruebas de bondad de ajuste para el equilibrio de Hardy Weinberg (HWE).

Con el objetivo de homogeneizar la métrica del tamaño, los genotipos diploides se transformaron en sus alelos equivalentes ($2n$ alelos totales). Bajo este nuevo modelo, se compilaban tablas de contingencia de dos por dos (2×2) para calcular la significancia

de la asociación alélica por chi cuadrado y cuantificar la fuerza del riesgo a través del Odds Ratio (OR) con sus respectivos intervalos de confianza del 95% (IC 95%).

Metilación de *PM20D1* según la edad

Para evaluar si en los niveles de metilación de *PM20D1* se encontraban asociados al envejecimiento, se realizó un modelo de regresión sobre las muestras de GEO. Con el propósito de disminuir el ruido estadístico de los extremos homólogos (AA y GG), se analizaron exclusivamente las muestras heterocigotas (AG, $n = 495$) según controles sanos ($n = 192$) o pacientes con la enfermedad ($n = 303$), usando un rango cronológico entre los 60 y los 90 años, concentrando el estudio en la ventana con mayor incidencia de la patología.

Los niveles de metilación (valores beta de la sonda cg14893161) en función de la edad, se ejecutó mediante Modelos Aditivos Generalizados (GAM) a través del paquete estadístico *mgcv* (versión 1.8) en el entorno R, con el fin de identificar relaciones no lineales, para evitar asumir comportamientos estáticos.

Finalmente, para analizar si las trayectorias se desviaban significativamente entre ambos grupos se usó un modelo de interacción global parametrizado (*Edad* $\times$ *Diagnóstico*). La significancia de las curvas y su homogeneidad inter-grupo se evaluó a través de pruebas de Wald generalizadas, fijando un umbral de significancia de un 0.05. Todos los gráficos y representaciones fueron estructuradas y exportadas mediante la librería *ggplot2*.

Regresión cuantílica

Se filtraron las muestras de la base de datos de la Fundación Pasqual Maragall (FPM) sin genotipo de *PM20D1* o *CST3*, sin datos de biomarcadores o con datos duplicados, seleccionando solo la primera visita de cada paciente. Los niveles de los marcadores asociados a *PM20D1*, fueron medidos a través de los valores del polimorfismo de nucleótido único (SNP) rs708727.

Para ver la relevancia funcional de este polimorfismo, se consultó la base de expresión tisular del consorcio GTEx, mostrando una fuerte asociación como eQTL. En tejido cerebral, el alelo de riesgo se asoció significativamente con una menor expresión del gen ($p = 7.2 \times 10^{-31}$, $NES = -0.96$) al igual que en sangre total ($p = 9.1 \times 10^{-105}$, $NES = -0.79$), confirmando un efecto dependiente del alelo sobre la transcripción de *PM20D1* y remarcando un mayor efecto en cerebro que en sangre.

De igual forma, se evaluaron las distintas variantes utilizadas para *CST3* en GTEx. El polimorfismo rs1064039 se identificó como un eQTL significativo para *CST3*, donde el alelo de riesgo se asoció con un descenso en los niveles del transcrito tanto en sangre ($p = 8.8 \times 10^{-25}$, $NES = -0.21$) como en cerebro ($p = 4.2 \times 10^{-5}$, $NES = -0.17$), aunque con efectos atribuibles al alelo muy inferiores a los observados para *PM20D1* (e.g., NES n cerebro para *PM20D1* y *CST3* de -0.96 y -0.17, respectivamente).

Según la tecnología de genotipificación usada en las bases de datos, fueron utilizados los polimorfismos rs6036478, rs3827143 y rs35610040 como marcadores equivalentes a este locus. Estas variantes se encuentran con un alto desequilibrio de ligamiento, específicamente rs3827143, que exhibió métricas idénticas a la variante de referencia.

Por su parte, rs6036478 mostró estadísticas parecidas tanto en sangre ($p = 9.2 \times 10^{-25}$, $NES = -0.20$) como en cerebro ($p = 6.7 \times 10^{-5}$, $NES = -0.16$), mientras que el SNP rs35610040 mostró un efecto aún más pronunciado sobre *CST3* en sangre ($p = 3.0 \times 10^{-25}$, $NES = -0.44$) y una consistencia con los demás SNPs en cerebro ($p = 4.0 \times 10^{-5}$, $NES = -0.16$), garantizando así la homogeneidad de esta inferencia independiente del marcador genético.

Se realizó una regresión cuantílica de la mediana como se ha descrito previamente ([Valletta et al., 2024](#)) para evaluar la interacción genética entre *PM20D1* y *CST3*, sistemáticamente sobre los niveles cada uno de los biomarcadores, $\frac{A\beta_{42}}{A\beta_{40}}$, $A\beta_{42}$, $A\beta_{40}$, $\frac{A\beta_{42}}{pTau_{181}}$, pTau₁₈₁ y NfL.

Estos análisis fueron ajustados por covariables para aislar interferencias biológicas o sociodemográficas. Las variables usadas fueron la edad cronológica, sexo biológico,

años de educación, índice de masa corporal (IMC), y el estado genotípico de riesgo de APOE como se han descrito en la literatura ([Kwon et al., 2025](#); [Li et al., 2021](#)).

El cálculo de la significancia estadística de la interacción para determinar el efecto sinérgico entre los alelos de riesgo se realizó mediante una prueba de Wald ([Wald, 1943](#)). La interacción genética vista en FPM fueron confirmados en la base de datos ADNI. Todos los procedimientos estadísticos y gráficos fueron generados por R (versión 3.6.0).

RESULTADOS

Frecuencia alélica de *PM20D1* en tejido cerebral

El análisis descriptivo del conjunto de datos de metilación para el tejido cerebral ($n = 1233$) permitió ordenar las muestras según la arquitectura genotípica del locus *PM20D1* (rs708727) tras la asignación no supervisada por el algoritmo *k means*. Nótese que se trata de un QTL fuerte en cerebro, con un efecto normalizado atribuible al alelo de riesgo de -0.96 ($p = 7.2 \times 10^{-31}$). Genéticamente, la población global fue conformada por 403 individuos homocigotos protectores (GG; 32.95%), 604 individuos heterocigotos (AG; 49.39%) y 216 sujetos clasificados en el perfil homocigoto de riesgo (AA; 17.66%).

Las estimaciones de las frecuencias alélicas globales determinaron que el Alelo protector G presenta una frecuencia de 0.5765, mientras que el alelo de riesgo A consolidó una frecuencia del 0.4235, en línea con las frecuencias reportadas para la población europea en donde el alelo alternativo A presenta una frecuencia alélica menor (MAF) de aproximadamente 0.40 ([The 1000 Genomes Project Consortium et al., 2015](#)).

Para asegurar la metodología de la inferencia genotípica, se testeó la adherencia a las leyes de segregación mendeliana en los controles sanos ($n = 538$). La prueba de bondad de ajuste confirmó que las frecuencias observadas se mantenían en equilibrio genético, sin desviaciones estadísticas respecto al equilibrio de Hardy-Weinberg ($X^2_{HWE} = 1.3691$, $gl = 1$, $p_{HWE} = 0.24196$).

Al deconstruir los genotipos a número de alelos equivalentes ($2n = 2446$ alelos totales) para evaluar la asociación alélica frente al diagnóstico clínico, la prueba de chi cuadrado de independencia detectó una diferencia significativa entre los grupos ($\chi^2 = 7.0632$, $gl =$

1, $p = 0.0079$). En esta divergencia, marcada por una asimetría proporcional (Tabla 2) el alelo G del grupo control representa un 60.69% ($n = 653$) mientras que, en los pacientes con la enfermedad, se reduce a un 55.26% ($n = 757$). En contraposición el Alelo de riesgo A, incrementa su presencia del 39.31% ($n = 423$) en el grupo control al 44.74% ($n = 613$) en los pacientes con la enfermedad (Figura 1A). La presencia del locus A aumenta de forma significativa el riesgo de desarrollar la enfermedad de Alzheimer, resultando en un Odds Ratio (OR) de 1.25 (IC 95%, 1.06 – 1.47).

Tabla 2. Distribución de la frecuencia alélica para el locus *PM20D1* en tejido cerebral ($2n = 2446$).

Diagnóstico	Alelo de Riesgo (A)	Alelo protector (G)	Total Alelos ($2n$)
Alzheimer	613 (44.74%)	757 (55.26%)	1370
Control Sano	423 (39.31%)	653 (60.69%)	1076
Cohorte Global	1036 (42.35%)	1410 (57.65%)	2446

Nota: Los conteos alélicos se derivaron de la deconstrucción a partir de genotipos virtuales inferidos mediante particionamiento no supervisado de la sonda cg14893161.

Frecuencia alélica de *PM20D1* en sangre total

Al igual que en tejido cerebral, se procesaron las muestras de sangre total ($n = 862$, equivalente a $2n = 1724$ alelos totales). Los controles sanos siguieron frecuencias en conformidad a lo esperado, manteniendo el equilibrio de Hardy Weinberg ($\chi^2_{HWE} = 1.7277$, $p = 0.1887$).

A diferencia de lo observado en tejido cerebral, la distribución del alelo de riesgo A (45.38% en controles vs 40.65% en pacientes con Alzheimer) y el alelo protector G (54.62% en controles vs 59.35% en pacientes con Alzheimer) no replicó el perfil de susceptibilidad como se había observado en las muestras cerebrales.

La prueba de chi cuadrado de independencia confirmó la ausencia de diferencias estadísticamente significativas inter-grupo ($\chi^2 = 3.2323$, $gl = 1$, $p =$

0.0722, Figura 1B), sin significancia tampoco en el tamaño del efecto ($OR = 0.82$, $IC\ 95\%$, $0.67 - 1.01$), indicando una posible especificidad tisular y/o menor potencia estadística a la hora de asociar el alelo en sangre por su menor NES. Nótese también que, a diferencia de las muestras de cerebro, cuyo diagnóstico se basa en el análisis histopatológico postmortem, en sangre se utilizan biomarcadores asociados a la enfermedad, con un cierto margen de error.

Distribución genética de *CST3*

A pesar del efecto limitado de las variantes genéticas de riesgo sobre la regulación del gen *CST3* (i.e., $p = 4.2 \times 10^{-5}$, $NES = -0.17$), procedimos a evaluar su posible asociación con la enfermedad de Alzheimer, utilizando 1213 muestras de GEO. En este análisis el grupo Alzheimer ($n = 615$), tuvo un 41.79% de homocigotos CC ($n = 257$), un 42.76% para el heterocigoto CT ($n = 263$) y un 15.45% para el homocigoto de riesgo TT ($n = 95$).

Por su parte, el grupo control sano ($n = 497$) tuvo una frecuencia del 39.64% para CC ($n = 197$), 41.05% para CT ($n = 204$) y 19.32% para TT ($n = 96$). El grupo control mostró una desviación significativa respecto a las leyes de segregación mendeliana teórica ($\chi^2_{HWE} = 09.8142$, $p = 0.0017$), lo que indica un sesgo en el muestreo o, más probablemente, la incapacidad del análisis por k-means de determinar el fondo genético cuando se trata de un gen con un efecto QTL débil (i.e., $p = 4.0 \times 10^{-5}$, $NES = -0.16$).

Epigenética cronológica de *PM20D1*

Con el propósito de evaluar la deriva epigenética asociada al envejecimiento, se analizaron de forma aislada las muestras heterocigotas (AG, $n_{total} = 495$) ya que, al presentar un modelo de metilación intermedia, permiten observar variaciones en sus niveles sin el efecto suelo-techo de los homocigotos, que muestran patrones de ausencia de metilación o metilación completa.

El modelado no lineal mediante Modelos Aditivos Generalizados (GAM) mostró niveles de metilación de *PM20D1* estables a lo largo de la edad cronológica analizada (60 a 90 años).

No se encontró una contribución significativa ni en la cohorte de controles sanos ($n = 192$, $p = 0.0795$) ni en pacientes con la enfermedad de Alzheimer ($n = 303$, $p = 0.9991$).

Igualmente, el modelo de interacción global no detectó muestras significativas entre las curvas de los grupos ($p_{diferencia} = 0.1036$, figura 1 C), demostrando la robustez de este mQTL a través del tiempo y confirmando la estabilidad de este biomarcador tanto en condiciones fisiológicas como patológicas.

Interacción genética entre *PM20D1* y *CST3*

Mediante una regresión lineal, se identificó una interacción epistática significativa sobre la carga amiloide. En las muestras de la Fundación Pasqual Maragall ($n = 2452$) la interacción (SNP $\times$ SNP) entre de *PM20D1* Y *CST3* mostró una asociación estadísticamente significativa sobre el ratio $\frac{A\beta_{42}}{A\beta_{40}}$ ($p = 0.040$).

Esta epistasis significativa solamente fue detectada en el ratio amiloide, no se identificaron interacciones genéticas significativa sobre los biomarcadores de daño axonal (NfL) o Tau patológica $\left(\frac{A\beta_{42}}{p\text{Tau}_{181}}\right)$.

Debido a la presencia de valores extremos en la distribución del amiloide, se ajustó un modelo de regresión cuantílica centrado en la mediana. Al aplicar este nuevo modelo, la potencia estadística de la interacción epistática incrementó alcanzando una alta significancia estadística sobre el biomarcador $A\beta_{42}/A\beta_{40}$ ($p = 0.0012$; Figura 2A).

Se usó la base de datos ADNI como cohorte de validación externa, para replicar el descubrimiento, revelando una tendencia estadística marginal para la interacción genética ($p = 0.0983$; Figura 2B), mostrando una consistencia direccional con el observado originalmente en los datos de la Fundación Pasqual Maragall.

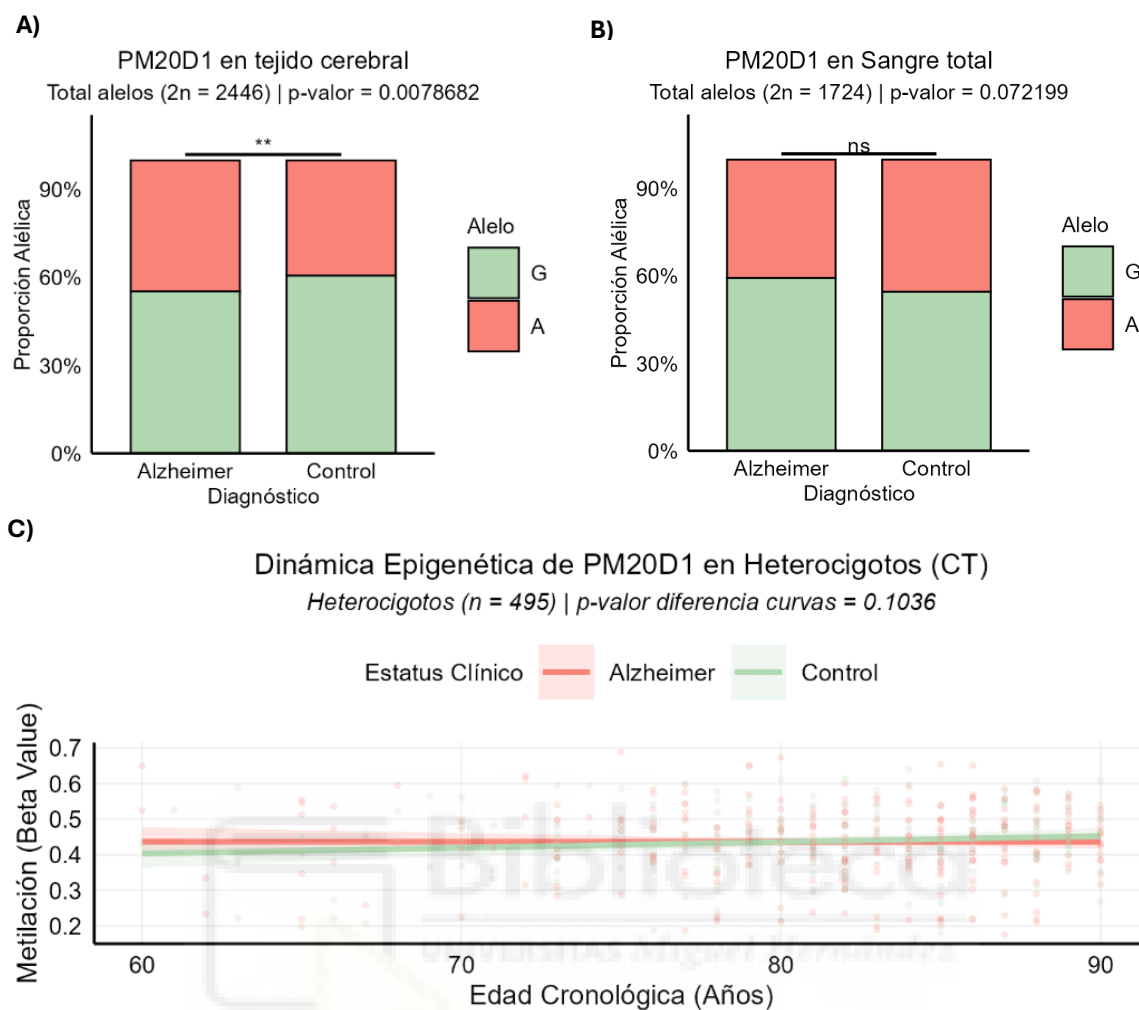


Figura 1. Frecuencia alélica de *PM20D1* Y *CST3*. (A) Frecuencia alélica de *PM20D1* en tejido cerebral de la cohorte de GEO ($n = 2446$) según pacientes con Alzheimer o controles sanos, indicando diferencias estadísticamente significativas mediante la prueba de chi cuadrado de pearson ($p = 0.0079$). (B) Frecuencia alélica de *PM20D1* en sangre total según diagnóstico ($n = 1724$). La evaluación mediante chi cuadrado confirma la homogeneidad estructural de las variantes entre los grupos ($p = 0.0722$). (C) Modelo aditivo de la metilación (valores beta) en función de la edad cronológica (de 60 a 90 años) exclusivamente con muestras heterocigotas (AG, $n = 495$). Las líneas continuas centrales representan la tendencia predicha mediante los modelos aditivos generalizados (GAM) para el grupo control ($n = 192$, verde) y Alzheimer ($n = 303$, rojo), flanqueadas por sombras indicando sus intervalos de confianza del 95%. No se evidenciaron desviaciones significativas en controles ($p = 0.0795$) ni en muestras alzheimer ($p = 0.9991$), viéndose un perfil inter grupo estadísticamente homogéneo ($p_{diferencia} = 0.1036$).

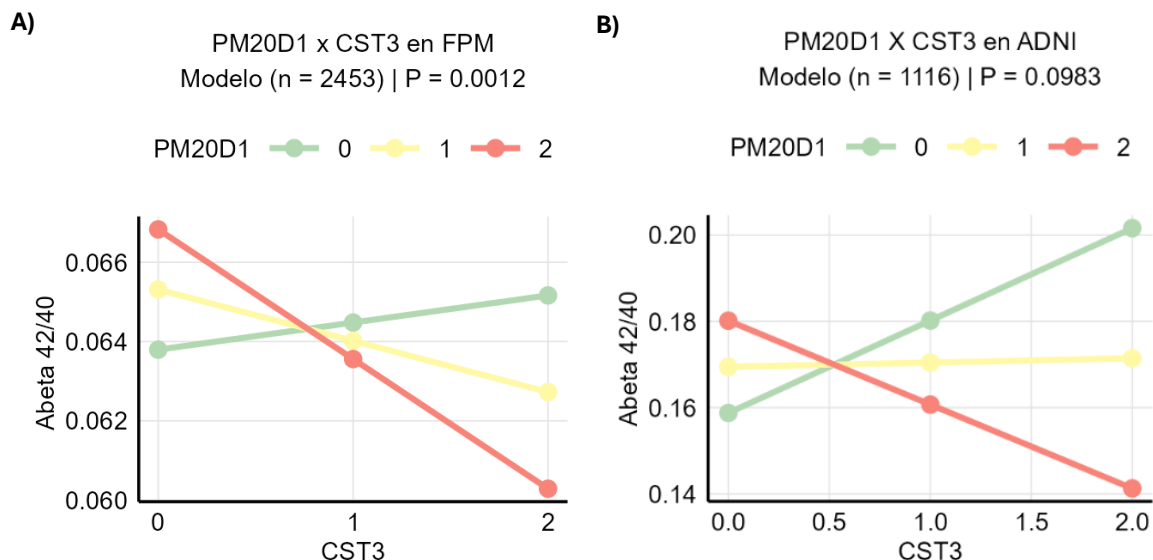


Figura 2. Interacción genética entre *PM20D1* y *CST3* sobre los niveles de $A\beta_{42/40}$. **A)** Regresión cuantílica no lineal estadísticamente significativa ($p = 0.0012$) sobre los niveles del ratio $A\beta_{42/40}$ en la cohorte de la Fundación Pasqual Maragall (FPM, $n = 2452$). **B)** Validación externa del modelo en la cohorte ADNI ($n = 1116$), replicando una tendencia estadística ($p = 0.0983$) consistente direccionalmente con el efecto de riesgo observado en FPM.

DISCUSIÓN

Metilación de *PM20D1*

Los resultados de este trabajo confirman a *PM20D1* como un locus de rasgo cuantitativo de metilación (mQTL) clave en la enfermedad de Alzheimer, al igual como se describió en [Sanchez-Mut et al. \(2018\)](#).

Este fenómeno se ve respaldado por el hecho que *PM20D1* se comporta como un QTL genéticamente robusto en el tejido cerebral, con un efecto normalizado (NES) atribuible al alelo de riesgo de -0.96, indicando que un gran porcentaje de la variabilidad está determinada por el genotipo, y mostrando el alelo A una asociación con la enfermedad, con un Odds Ratio (OR) de 1.25 (IC 95%, 1.06 – 1.47).

A diferencia de la significancia encontrada en tejido cerebral, en sangre total no se encontraron diferencias significativas en las frecuencias alélicas entre controles y

pacientes con la enfermedad de Alzheimer ($p = 0.0722$), sugiriendo una especificidad tisular en las marcas epigenéticas y/o limitaciones en el QTL en el tejido.

Mientras que en tejido cerebral el tamaño del efecto normalizado de este mQTL es dominante ($NES = -0.96$), en tejido sanguíneo el NES decae a un 0.70, implicando que, en la metilación periférica existe un mayor ruido estadístico por factores no controlados, lo que aumenta el margen de error del k means para clasificar los genotipos en comparación al tejido cerebral.

Por otro lado, mientras que el tejido cerebral cuenta con un diagnóstico definitivo por análisis histopatológicos *postmortem*, el diagnóstico de las muestras de sangre total se fundamenta en biomarcadores, los cuales poseen un rango de incertidumbre superior, lo que, en conjunto, podría explicar las diferencias de asociación con la enfermedad entre ambos tejidos.

Estabilidad cronológica de *PM20D1*

El modelado mediante Modelos Aditivos Generalizados (GAM) en los heterocigotos (*AG*, $n = 495$) permitió eliminar el ruido estadístico propio de los extremos homocigotos, pudiendo observar una trayectoria plana tanto en los controles ($p = 0.0795$) como en los pacientes con Alzheimer ($p = 0.9991$) a través de las edades, demostrando la estabilidad de este mQTL en el envejecimiento en tejido cerebral.

Biológicamente, la ausencia de una pendiente significativa en función a los años cronológicos descarta la susceptibilidad de este locus a la deriva epigenética (*epigenetic drift*). La estabilidad estadística ($p_{diferencia} = 0.1036$) ratifica que el equilibrio se mantiene incluso en pacientes con la enfermedad de Alzheimer, lo que sugiere que la metilación de *PM20D1* opera de forma estática a través de su estructura genética a lo largo de las últimas décadas de la vida (de 60 a 90 años).

Estas observaciones difieren con los publicados por Wang y colaboradores (PMID: 33298155), en los que observan una alteración de la metilación de *PM20D1* con la edad, que según nuestros resultados serían debidos a una distribución diferencial de los alelos de riesgo en los diferentes rangos de edad, variable no contemplada en el estudio de (Q. Wang et al., 2020).

Al observarse que los niveles de metilación de en los heterocigotos no experimentan diferencias dependientes de la edad, *PM20D1* se muestra como un biomarcador estable, consolidando su utilidad en la identificación de perfiles vulnerables a la enfermedad de Alzheimer.

Epigenética de *CST3*

A pesar de que variantes genéticas y perfiles transcripcionales de *CST3* han sido ya relacionados a la enfermedad de Alzheimer (Finckh, 2000) actualmente no hay estudios que vinculen su metilación. Los análisis de la metilación de *CST3* en estas 1213 muestras procedentes de GEO, demostraron distribuciones aparentemente estables tanto en controles como pacientes Alzheimer. Sin embargo, los controles revelaron una desviación significativa respecto al equilibrio de Hardy Weinberg ($\chi^2_{HWE} = 9.8142, p = 0.0017$).

La quiebra del equilibrio de Hardy Weinberg (HWE) en el grupo control es indicativa de un sesgo en el muestreo o en la inferencia del fondo genético a partir de la metilación. Nótese que *CST3* se comporta como un QTL débil a nivel de expresión ($NES \approx 0.30$), pero que no hay estudios que soporten su comportamiento como QTL a nivel de metilación. Estos resultados y limitaciones no permiten la continuidad del análisis de asociación posterior como se realizó con *PM20D1*, ya que no se dispone de evidencia suficiente para descartar que se trate de un artefacto matemático desprovisto de significad biológico.

Interacción entre *PM20D1* y *CST3*

Estudios recientes han demostrado que el gen protector contra la enfermedad de Alzheimer, *PM20D1* (Sanchez-Mut et al., 2018), genera una serie de compuestos, en concreto el N-oleoil leucina (OLE) (Long et al., 2016), que actúa sobre las células de la microglía del cerebro, aumentando la expresión de *CST3* (Pozzi-Ruiz et al., 2026), la cual, en último término, facilita la eliminación de los péptidos de beta amiloide ($A\beta$) (X. Li et al., 2026).

Estos estudios describen una vía de señalización protectora contra la enfermedad, que justifica el análisis de ambos loci de manera conjunta, ya que la alteración de cualquiera de ellos podría disminuir la eficiencia de la misma, mostrando efectos epistáticos a partir de un cierto umbral.

La identificación de *PM20D1* y *CST3* como eQTLs otorga una justificación adicional a su análisis conjunto funcional, ya que permite realizar hipótesis de direccionalidad basadas en la cantidad disponible de cada una de estas proteínas en relación con la enfermedad. Nótese que, a diferencia del estudio de metilación, el estudio genético directo permite soslayar el menor impacto del QTL sobre la expresión de *CST3*.

La identificación de una interacción genética epistática estadísticamente significativa sobre el ratio $A\beta_{42/40}$ mediante una regresión cuantílica ($p = 0.0012$) constituye el hallazgo central del estudio y ayuda a entender de mejor manera la genética de la enfermedad de Alzheimer. Estos resultados han mostrado que, la presencia del alelo de riesgo de *CST3* potencia de forma sinérgica el efecto patológico derivado de *PM20D1*.

Esta interacción alcanza una alta significancia en la cohorte de la Fundación Pasqual Maragall (FPM) con una tendencia en la cohorte ADNI. Las muestras de FPM se caracterizan por presentar un deterioro cognitivo leve (MCI), etapas tempranas de la enfermedad en donde la comulación de placas $A\beta$ tiene una correlación lineal con el declive cognitivo.

Por su parte, las muestras de la cohorte ADNI incluyen altamente a pacientes con diagnósticos establecidos de la enfermedad de Alzheimer, etapa avanzada en la cual la deposición amiloide es saturada y pierde linealidad, dando paso a la proteopatía de Tau hiperfosforilada y a la neurodegeneración (Jack et al., 2010).

Dado que *CST3* tiene una función neuroprotectora específica en el aclaramiento de $A\beta$, careciendo de una estrecha relación con Tau, resulta altamente coherente que el efecto de la interacción epistática con *PM20D1* sea significativamente más potente en FPM, donde sus individuos pertenecen a etapas más tempranas de la enfermedad.

Sumado a esto, FPM posee un diseño de centro único, altamente homogéneo y marcadores uniformes, mientras que ADNI introduce ruido debido a su naturaleza multicéntrica y al uso de sondas heterogéneas, diluyendo el efecto de la interacción.

En conjunto, nuestro modelo implica que la convergencia entre la reducción de un estímulo (*PM20D1*) y un defecto en el efector (*CST3*) disminuiría la capacidad del cerebro para limpiar los agregados proteicos, lo que resultaría en una acumulación de placas seniles y, por consiguiente, el marcador del ratio amiloide $A\beta_{42/40}$ detectado en las muestras cerebrales.

Para ver esta interacción genética, el uso de una regresión cuantílica nos permitió ver de mejor manera esta interacción en comparación de una regresión lineal de la media. Al trabajar con la mediana ($p = 0.0012$) en lugar de la media ($p = 0.04$), se pudo disminuir el error estándar y el impacto de los valores extremos que podrían haber afectado en las cohortes clínicas de la enfermedad.

Clínicamente, este efecto epistático sugiere que pacientes con ambos alelos de riesgo, podrían experimentar un declive cognitivo y progresión patológica acelerada. Haciendo fundamental el desarrollo de algoritmos y terapias personalizadas complejas que contemplen las interacciones entre diferentes variables genéticas con una arquitectura genómica no necesariamente lineal.

Limitaciones de la investigación

A pesar de la significancia de los resultados, es importante reconocer las limitaciones de los métodos empleados. En los análisis epigenéticos, las muestras de los pacientes provienen de 14 estudios completamente independientes. Sutiles variaciones en las metodologías de los diferentes laboratorios pueden venir con ligeros errores que pueden influir en la obtención o conclusión de los resultados.

La inferencia genotípica mediante el algoritmo de particionamiento k means, aun demostrando estabilidad biológica rigurosa mediante el equilibrio de Hardy-Weinberg, posee un margen error inherente. Esta aproximación estadística no reemplaza la precisión genotípica característica de los secuenciadores de nueva generación o microarrays de ADN, por lo que cabe la posibilidad de errar en algunas de las inferencias genotípicas.

Una de las mayores limitantes en los análisis genéticos fue en su confirmación, debido a las diferentes sondas empleadas en las distintas cohortes de ADNI. Al haber utilizado diferentes sondas para secuenciar el DNA de las muestras, no se pudo utilizar una única sonda universal que estuviera en cada una de las cohortes para analizar la presencia de *CST3*. A pesar de usar variantes con un robusto desequilibrio de ligamiento ($r^2 = 0.9608$), estas producen un ligero ruido estadístico que podría afectar a la hora de replicar los resultados.

CONCLUSIONES Y PROYECCIÓN FUTURA

Conclusiones

Estos resultados nos permiten ver como la combinación de diferentes bases de datos (GEO, ADNI y FPM) son de gran ayuda para análisis globales genéticos y epigenéticos. El procesamiento en R, la estratificación y particionamiento por k means, nos permitieron inferir los genotipos eficientemente, siguiendo el equilibrio de Hardy-Weinberg.

Los análisis estadísticos sobre las muestras extraídas de GEO ($n = 2085$) mostraron una diferencia significativa en las muestras cerebrales de pacientes con la enfermedad de Alzheimer, caracterizada por el cambio en las frecuencias genotípicas de *PM20D1* en comparación con controles sanos ($p = 0.0074$).

En cuanto al cambio de la metilación según la edad (de 60 a 90 años), el subgrupo heterocigoto (*AG*, $n = 495$) no presento cambios significativos, tanto en controles ($p = 0.0795$) como en pacientes Alzheimer ($p = 0.9991$), junto con un perfil inter-grupo estadísticamente homogéneo ($p_{diferencia} = 0.1036$), confirmando a este mQTL como un biomarcador estático a través del tiempo.

Los análisis aplicados principalmente a la cohorte de la FPM ($n = 2462$) mostraron una epistasis estadística relacionada al aumento de la carga amiloide. Mientras que el modelo lineal mostró una estadística convencional ($p = 0.040$), la regresión de la mediana nos permitió ver una alta interacción ($p = 0.0012$) sobre el biomarcador $A\beta_{42/40}$, mostrando una tendencia en el dataset de confirmación externa ADNI ($p = 0.0983$, $n = 1117$).

Estos hallazgos permiten aceptar formalmente la hipótesis planteada en esta investigación. Se puede observar como la variabilidad genotípica de *PM20D1* y *CST3* modulan el riesgo de la enfermedad de Alzheimer a través de una epistasis, alterando los niveles de la carga amiloide, superando a la suma de sus contribuciones individuales.

Al comparar estos resultados con la literatura reciente, estos concuerdan con la hipótesis que *PM20D1* y *CST3* comparten una misma ruta metabólica relacionada a la protección neuronal, en donde la combinación de estos alelos aumenta la carga amiloide, provocando una progresión patológica acelerada.

Estos descubrimientos demuestran que la consideración conjunta de estos alelos de riesgo, pueden ayudar a identificar la significancia clínica de estos genes, que suele diluirse en estudios GWAS debido a las limitaciones de los modelos lineales aditivos independientes.

Mediante la herramienta de inferencia para modelos de regresión lineal de la Universidad de California (basada en la aproximación de Hulley, Cummings, Browner, Grady y Newman, 2013) se determinó que si el efecto epistático detectado ($n = 2,462$, $p = 0.0012$) fuera nuevamente evaluado, la magnitud de asociación marginal ($r = 0.05$) requeriría una muestra total de 15,819 individuos para alcanzar la potencia estadística del 80% ($\beta = 0.20$) bajo el umbral de significancia en estudios GWAS ($\alpha = 5 \times 10^{-8}$).

Como vemos con este número de muestras, el modelo de regresión cuantílica, optimiza drásticamente los análisis, justificando la incorporación de estos genes dentro del diseño de paneles de cribado y pruebas genéticas de diagnóstico temprano para Alzheimer, antes de la manifestación clínica neurodegenerativa.

Proyecciones futuras

Una vez vista la significancia estadística entre *PM20D1* en el dataset de GEO, se propone realizar una validación externa en los datos de la Fundación Pasqual Maragall (FPM) y ADNI, con el fin de aumentar la potencia estadística, permitiendo una mayor precisión en el impacto de estos alelos sobre la enfermedad.

Incorporar nuevas bases de datos como AIBL (Australian imaging, biomarkers and lifestyle study), permitiría aumentar la potencia muestral y hacer un modelo de regresión cuantílica global que combine los datos genotípicos y los biomarcadores de todas las bases de datos.

Finalmente, la integración de análisis transcriptómicos permitiría evaluar si la interacción epistática encontrada en los análisis genéticos se traduce a una regulación en los niveles de transcripción de *PM20D1* y *CST3*, lo cual además ayudaría a aumentar la potencia estadística del estudio, al fusionar los genotipos inferidos de la metilación, los datos de transcripción y los datos genotipados de los cohortes genéticos, validando de forma definitiva el impacto de esta interacción en la enfermedad.

BIBLIOGRAFÍA

- Alves Chagas, B. C., Jia, H., Brixius, B., & Brixius-Anderko, S. (2025). The curious family of cytochrome P450 4F fatty acid ω -hydroxylases: Recent developments in their function and relevance for human health. *Open Biology*, 15(8), 250115. <https://doi.org/10.1098/rsob.250115>
- Andrade-Guerrero, J., Santiago-Balmaseda, A., Jeronimo-Aguilar, P., Vargas-Rodríguez, I., Cadena-Suárez, A. R., Sánchez-Garibay, C., Pozo-Molina, G., Méndez-Catalá, C. F., Cardenas-Aguayo, M.-C., Diaz-Cintra, S., Pacheco-Herrero, M., Luna-Muñoz, J., & Soto-Rojas, L. O. (2023). Alzheimer's Disease: An Updated Overview of Its Genetics. *International Journal of Molecular Sciences*, 24(4), 3754. <https://doi.org/10.3390/ijms24043754>
- Apostolova, L. G. (2016). Alzheimer Disease. *Continuum*, 22(2), 419–434. <https://doi.org/10.1212/CON.0000000000000307>
- Bellenguez, C., Grenier-Boley, B., & Lambert, J.-C. (2020). Genetics of Alzheimer's disease: Where we are, and where we are going. *Current Opinion in Neurobiology*, 61, 40–48. <https://doi.org/10.1016/j.conb.2019.11.024>
- Bhuiya, J., Shibly, A. Z., Sheikh, A. M., Tabassum, S., Abdullah, F. B., Fahmida, J., Tamegai, H., Aritake, S., Yano, S., & Nagai, A. (2026). Deposition of aggregated cystatin C-amyloid- β complexes and reduced cathepsin B activity modulate

- cerebral amyloid angiopathy pathogenesis. *Journal of Neuropathology & Experimental Neurology*, 85(4), 372–388. <https://doi.org/10.1093/jnen/nlaf138>
- Blumenfeld, J., Yip, O., Kim, M. J., & Huang, Y. (2024). Cell type-specific roles of APOE4 in Alzheimer disease. *Nature Reviews Neuroscience*, 25(2), 91–110. <https://doi.org/10.1038/s41583-023-00776-9>
- Clough, E., & Barrett, T. (2016). The Gene Expression Omnibus Database. In E. Mathé & S. Davis (Eds.), *Statistical Genomics* (Vol. 1418, pp. 93–110). Springer New York. https://doi.org/10.1007/978-1-4939-3578-9_5
- Finckh, U. (2000). Genetic Association of a Cystatin C Gene Polymorphism With Late-Onset Alzheimer Disease. *Archives of Neurology*, 57(11), 1579. <https://doi.org/10.1001/archneur.57.11.1579>
- Garro-Núñez, D., Mora-Cubillo, P., Fonseca-Bone, S., Picado-Martínez, M. J., Fonseca-Brenes, M., Raventós-Vorst, H., & Chavarría-Soley, G. (2022). The many roles of the Alzheimer-associated gene PM20D1. *Journal of Translational Genetics and Genomics*, 6, 361–374. <https://doi.org/10.20517/jtgg.2022.10>
- Goate, A., Chartier-Harlin, M.-C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L., Mant, R., Newton, P., Rooke, K., Roques, P., Talbot, C., Pericak-Vance, M., Roses, A., Williamson, R., Rossor, M., ... Hardy, J. (1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*, 349(6311), 704–706. <https://doi.org/10.1038/349704a0>
- Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., & Zhang, K. (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular Cell*, 49(2), 359–367. <https://doi.org/10.1016/j.molcel.2012.10.016>
- Heneka, M. T., Van Der Flier, W. M., Jessen, F., Hoozemanns, J., Thal, D. R., Boche, D., Brosseron, F., Teunissen, C., Zetterberg, H., Jacobs, A. H., Edison, P., Ramirez, A., Cruchaga, C., Lambert, J.-C., Laza, A. R., Sanchez-Mut, J. V., Fischer, A., Castro-Gomez, S., Stein, T. D., ... Riechers, S.-P. (2025). Neuroinflammation in Alzheimer disease. *Nature Reviews Immunology*, 25(5), 321–352. <https://doi.org/10.1038/s41577-024-01104-7>

- Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9(1), 119–128. [https://doi.org/10.1016/S1474-4422\(09\)70299-6](https://doi.org/10.1016/S1474-4422(09)70299-6)
- Kaesler, S. A., Herzig, M. C., Coomaraswamy, J., Kilger, E., Selenica, M.-L., Winkler, D. T., Staufenbiel, M., Levy, E., Grubb, A., & Jucker, M. (2007). Cystatin C modulates cerebral β -amyloidosis. *Nature Genetics*, 39(12), 1437–1439. <https://doi.org/10.1038/ng.2007.23>
- Kaur, G., & Levy, E. (2012). Cystatin C in Alzheimer's disease. *Frontiers in Molecular Neuroscience*, 5. <https://doi.org/10.3389/fnmol.2012.00079>
- Kim, J. T., Jedrychowski, M. P., Wei, W., Fernandez, D., Fischer, C. R., Banik, S. M., Spiegelman, B. M., & Long, J. Z. (2020). A Plasma Protein Network Regulates PM20D1 and N-Acyl Amino Acid Bioactivity. *Cell Chemical Biology*, 27(9), 1130–1139.e4. <https://doi.org/10.1016/j.chembiol.2020.04.009>
- Kitaba, N. T., Østergaard, T. M., Lønnebotn, M., Accordini, S., Real, F. G., Malinovschi, A., Oudin, A., Benediktsdottir, B., González, F. J. C., Gómez, L. P., Holm, M., Jögi, N. O., Dharmage, S. C., Skulstad, S. M., Schlünssen, V., Svanes, C., & Holloway, J. W. (2025). Father's adolescent body silhouette is associated with offspring asthma, lung function and BMI through DNA methylation. *Communications Biology*, 8(1), 796. <https://doi.org/10.1038/s42003-025-08121-9>
- Kwon, H. S., Hwang, M., Koh, S.-H., Choi, S. H., Lee, J.-H., Kim, H.-J., Park, S.-H., Park, H.-H., Jeong, J. H., Han, M. H., & Kim, J. Y. (2025). Comparison of plasma p-tau217 and p-tau181 in predicting amyloid positivity and prognosis among Korean memory clinic patients. *Scientific Reports*, 15(1), 7791. <https://doi.org/10.1038/s41598-025-90232-8>
- Li, T., Wang, B., Gao, Y., Wang, X., for the Alzheimer's Disease Neuroimaging Initiative, Yan, T., Xiang, J., Niu, Y., Liu, T., Chen, D., Fang, B., Xie, Y., Funahashi, S., & Yan, T. (2021). APOE ϵ 4 and cognitive reserve effects on the functional network in the Alzheimer's disease spectrum. *Brain Imaging and Behavior*, 15(2), 758–771. <https://doi.org/10.1007/s11682-020-00283-w>
- Li, X., Tang, X., Zeng, J., Duan, L., Hou, Z., Li, L., Guo, Y., Chai, C., Liu, J., Wang, Y., Zhu, L.-Q., Li, H., Zhang, T., Wang, Y., He, A., & Lu, Y. (2026). Peripheral cancer

- attenuates amyloid pathology in Alzheimer's disease via cystatin-c activation of TREM2. *Cell*, 189(3), 853-871.e31. <https://doi.org/10.1016/j.cell.2025.12.020>
- Liu, E., Zhang, Y., & Wang, J.-Z. (2024). Updates in Alzheimer's disease: From basic research to diagnosis and therapies. *Translational Neurodegeneration*, 13(1), 45. <https://doi.org/10.1186/s40035-024-00432-x>
- Long, J. Z., Roche, A. M., Berdan, C. A., Louie, S. M., Roberts, A. J., Svensson, K. J., Dou, F. Y., Bateman, L. A., Mina, A. I., Deng, Z., Jedrychowski, M. P., Lin, H., Kamenecka, T. M., Asara, J. M., Griffin, P. R., Banks, A. S., Nomura, D. K., & Spiegelman, B. M. (2018). Ablation of PM20D1 reveals *N*-acyl amino acid control of metabolism and nociception. *Proceedings of the National Academy of Sciences*, 115(29). <https://doi.org/10.1073/pnas.1803389115>
- Long, J. Z., Svensson, K. J., Bateman, L. A., Lin, H., Kamenecka, T., Lokurkar, I. A., Lou, J., Rao, R. R., Chang, M. R., Jedrychowski, M. P., Paulo, J. A., Gygi, S. P., Griffin, P. R., Nomura, D. K., & Spiegelman, B. M. (2016). The Secreted Enzyme PM20D1 Regulates Lipidated Amino Acid Uncouplers of Mitochondria. *Cell*, 166(2), 424–435. <https://doi.org/10.1016/j.cell.2016.05.071>
- Mathews, P. M., & Levy, E. (2016). Cystatin C in aging and in Alzheimer's disease. *Ageing Research Reviews*, 32, 38–50. <https://doi.org/10.1016/j.arr.2016.06.003>
- Newman, D. J. (2002). Cystatin C. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, 39(2), 89–104. <https://doi.org/10.1258/0004563021901847>
- Niu, Y., Zhang, Y., Zha, Q., Shi, J., & Weng, Q. (2025). Bioinformatics to analyze the differentially expressed genes in different degrees of Alzheimer's disease and their roles in progress of the disease. *Journal of Applied Genetics*, 66(1), 73–85. <https://doi.org/10.1007/s13353-024-00827-6>
- Pearson, K. (1992). On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such that it Can be Reasonably Supposed to have Arisen from Random Sampling. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in Statistics* (pp. 11–28). Springer New York. https://doi.org/10.1007/978-1-4612-4380-9_2
- Pichet Binette, A., Smith, R., Salvadó, G., Tideman, P., Glans, I., Van Westen, D., Groot, C., Ossenkoppele, R., Stomrud, E., Parchi, P., Zetterberg, H., Blennow, K.,

- Mattsson-Carlsson, N., Janelidze, S., Palmqvist, S., Hansson, O., Alzheimer's Disease Neuroimaging Initiative, Adegboye, O., Adam Hussen, K., ... Zimmerman, C. (2025). Evaluation of the Revised Criteria for Biological and Clinical Staging of Alzheimer Disease. *JAMA Neurology*, 82(7), 666.
<https://doi.org/10.1001/jamaneurol.2025.1100>
- Pozzi-Ruiz, V., Giner De Gracia, A., Glauser, L., Romani, M., Gunter-Rahman, F., González-Ramón, A., Haj-Yahya, M., Kolla, R., Burns, A. M., Lashuel, H. A., Auwerx, J., Gräff, J., & Sanchez-Mut, J. V. (2026). The PM20D1-OLE pathway induces microglia rewiring to ameliorate Alzheimer disease. *Cell Death & Disease*, 17(1), 561. <https://doi.org/10.1038/s41419-026-08791-1>
- Reitz, C., Pericak-Vance, M. A., Foroud, T., & Mayeux, R. (2023). A global view of the genetic basis of Alzheimer disease. *Nature Reviews Neurology*, 19(5), 261–277.
<https://doi.org/10.1038/s41582-023-00789-z>
- Ren, R., Qi, J., Lin, S., Liu, X., Yin, P., Wang, Z., Tang, R., Wang, J., Huang, Q., Li, J., Xie, X., Hu, Y., Cui, S., Zhu, Y., Yu, X., Wang, P., Zhu, Y., Wang, Y., Huang, Y., ... Wang, G. (2022). The China Alzheimer Report 2022. *General Psychiatry*, 35(1), e100751.
<https://doi.org/10.1136/gpsych-2022-100751>
- Robbins, J. M., Benson, M., Verkerke, A. R. P., Tiwari, G., Deng, S., Rao, P., Tahir, U. A., Avila-Pacheco, J., Shi, X., Guan, Y., Tendoh, F.-G., Barber, J. L., Miller, P. E., Perry, A. S., Hall, M. E., Wood, A. C., Taylor, K. D., Post, W. S., Rich, S. S., ... Gerszten, R. E. (2026). N-Palmitoyl Glutamine Is a Candidate Mediator of Cardiorespiratory Fitness. *Circulation*, 153(1), 47–58.
<https://doi.org/10.1161/CIRCULATIONAHA.125.074187>
- Sanchez-Mut, J. V., Glauser, L., Monk, D., & Gräff, J. (2020). Comprehensive analysis of PM20D1 QTL in Alzheimer's disease. *Clinical Epigenetics*, 12(1), 20.
<https://doi.org/10.1186/s13148-020-0814-y>
- Sanchez-Mut, J. V., Heyn, H., Silva, B. A., Dixsaut, L., Garcia-Esparcia, P., Vidal, E., Sayols, S., Glauser, L., Monteagudo-Sánchez, A., Perez-Tur, J., Ferrer, I., Monk, D., Schneider, B., Esteller, M., & Gräff, J. (2018). PM20D1 is a quantitative trait locus associated with Alzheimer's disease. *Nature Medicine*, 24(5), 598–603.
<https://doi.org/10.1038/s41591-018-0013-y>

- Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Ch  telat, G., Teunissen, C. E., Cummings, J., & Van Der Flier, W. M. (2021). Alzheimer's disease. *The Lancet*, 397(10284), 1577–1590. [https://doi.org/10.1016/S0140-6736\(20\)32205-4](https://doi.org/10.1016/S0140-6736(20)32205-4)
- Sivaramakrishna, D., Manasa, K., Kishore, R., & Swamy, M. J. (2025). Structural and thermal investigations on odd and even-chain N-acyl-L-alanine alkyl esters: Insights into unusual thermal transition parameters and thermal stability. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 727, 138177. <https://doi.org/10.1016/j.colsurfa.2025.138177>
- Strittmatter, W. J., Weisgraber, K. H., Huang, D. Y., Dong, L. M., Salvesen, G. S., Pericak-Vance, M., Schmechel, D., Saunders, A. M., Goldgaber, D., & Roses, A. D. (1993). Binding of human apolipoprotein E to synthetic amyloid beta peptide: Isoform-specific effects and implications for late-onset Alzheimer disease. *Proceedings of the National Academy of Sciences*, 90(17), 8098–8102. <https://doi.org/10.1073/pnas.90.17.8098>
- Taglieri, N., Koenig, W., & Kaski, J. C. (2009). Cystatin C and Cardiovascular Risk. *Clinical Chemistry*, 55(11), 1932–1943. <https://doi.org/10.1373/clinchem.2009.128397>
- Tanzo, J. T., Li, V. L., Wiggenhorn, A. L., Moya-Garzon, M. D., Wei, W., Lyu, X., Dong, W., Tahir, U. A., Chen, Z.-Z., Cruz, D. E., Deng, S., Shi, X., Zheng, S., Guo, Y., Sims, M., Abu-Remaileh, M., Wilson, J. G., Gerszten, R. E., Long, J. Z., & Benson, M. D. (2023). CYP4F2 is a human-specific determinant of circulating N-acyl amino acid levels. *Journal of Biological Chemistry*, 299(6), 104764. <https://doi.org/10.1016/j.jbc.2023.104764>
- The 1000 Genomes Project Consortium, Corresponding authors, Auton, A., Abecasis, G. R., Steering committee, Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., ... Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Valletta, M., Vetrano, D. L., Rizzuto, D., Winblad, B., Canevelli, M., Andersson, S., Dale, M., Fredolini, C., Fratiglioni, L., & Grande, G. (2024). Blood biomarkers of Alzheimer's disease in the community: Variation by chronic diseases and

- inflammatory status. *Alzheimer's & Dementia*, 20(6), 4115–4125.
<https://doi.org/10.1002/alz.13860>
- Veitch, D. P., Weiner, M. W., Miller, M., Aisen, P. S., Ashford, M. A., Beckett, L. A., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Landau, S. M., Morris, J. C., Nho, K. T., Nosheny, R., Okonkwo, O., Perrin, R. J., Petersen, R. C., Rivera Mindt, M., Saykin, A., ... for the Alzheimer's Disease Neuroimaging Initiative. (2024). The Alzheimer's Disease Neuroimaging Initiative in the era of Alzheimer's disease treatment: A review of ADNI studies from 2021 to 2022. *Alzheimer's & Dementia*, 20(1), 652–694. <https://doi.org/10.1002/alz.13449>
- Villa, C., & Stocco, A. (2022). Epigenetic Peripheral Biomarkers for Early Diagnosis of Alzheimer's Disease. *Genes*, 13(8), 1308.
<https://doi.org/10.3390/genes13081308>
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3), 426–482. <https://doi.org/10.1090/S0002-9947-1943-0012401-3>
- Wang, Q., Chen, Y., Readhead, B., Chen, K., Su, Y., Reiman, E. M., & Dudley, J. T. (2020). Longitudinal data in peripheral blood confirm that PM20D1 is a quantitative trait locus (QTL) for Alzheimer's disease and implicate its dynamic role in disease progression. *Clinical Epigenetics*, 12(1), 189. <https://doi.org/10.1186/s13148-020-00984-5>
- Wang, Q.-P., Gu, J.-W., Zhan, X.-H., Li, H., & Luo, X.-H. (2009). Assessment of glomerular filtration rate by serum cystatin C in patients undergoing coronary artery bypass grafting. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, 46(6), 495–500. <https://doi.org/10.1258/acb.2009.009065>
- Wang, Z., Zhang, Q., Lin, J.-R., Jabalameli, M. R., Mitra, J., Nguyen, N., & Zhang, Z. D. (2021). Deep post-GWAS analysis identifies potential risk genes and risk variants for Alzheimer's disease, providing new insights into its disease mechanisms. *Scientific Reports*, 11(1), 20511. <https://doi.org/10.1038/s41598-021-99352-3>
- Westphal Filho, F. L., Moss Lopes, P. R., Menegaz De Almeida, A., Sano, V. K. T., Tamashiro, F. M., Gonçalves, O. R., De Moraes, F. C. A., Kreuz, M., Kelly, F. A., & Silveira Feitoza, P. V. (2025). Statin use and dementia risk: A systematic review

- and updated meta-analysis. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 11(1), e70039. <https://doi.org/10.1002/trc2.70039>
- Xue, D., Blue, E. E., Conomos, M. P., & Fohner, A. E. (2024). The power of representation: Statistical analysis of diversity in US Alzheimer's disease genetics data. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 10(1), e12462. <https://doi.org/10.1002/trc2.12462>
- Yang, R., Hu, Y., Lee, C. H., Liu, Y., Diaz-Canestro, C., Fong, C. H. Y., Lin, H., Cheng, K. K. Y., Pravelil, A. P., Song, E., Lam, K. S. L., & Xu, A. (2022). PM20D1 is a circulating biomarker closely associated with obesity, insulin resistance and metabolic syndrome. *European Journal of Endocrinology*, 186(2), 151–161. <https://doi.org/10.1530/EJE-21-0847>
- Yang, Y., Chen, S., Zhang, L., Zhang, G., Liu, Y., Li, Y., Zou, L., Meng, L., Tian, Y., Dai, L., Xiong, M., Pan, L., Xiong, J., Chen, L., Hou, H., Yu, Z., & Zhang, Z. (2024). The PM20D1-NADA pathway protects against Parkinson's disease. *Cell Death & Differentiation*, 31(11), 1545–1560. <https://doi.org/10.1038/s41418-024-01356-9>
- Zi, M., & Xu, Y. (2018). Involvement of cystatin C in immunity and apoptosis. *Immunology Letters*, 196, 80–90. <https://doi.org/10.1016/j.imlet.2018.01.006>