



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Caracterización genómica de un grupo de individuos con Deterioro Cognitivo Leve (DCL) en población del Atlántico, Colombia.

Johan Hernando Largo González

Universidad Nacional de Colombia

Facultad de Ingeniería

Bogotá, Colombia

2025

Caracterización genómica de un grupo de individuos con Deterioro Cognitivo Leve (DCL) en población del Atlántico, Colombia.

Johan Hernando Largo González

Trabajo de grado presentado como requisito parcial para optar al título de:

Magíster en Bioinformática

Director:

Ph D. Andrés Mauricio Pinzón Velasco

Codirector:

Ph D. Carlos Eduardo Arboleda Bustos

Línea de Investigación:

Biología de Sistemas

Grupo de Investigación:

Grupo de Bioinformática y Biología de Sistemas - GIBBS

Universidad Nacional de Colombia

Facultad de Ingeniería

Bogotá, Colombia

2025

(Dedicatoria)

*A la ciencia, que es para mi como la espada
del augurio, y me deja ver más allá de lo
evidente.*

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto). Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Johan Hernando Largo González

C.C. 1.073.518.389

jlargog@unal.edu.co

Agradecimientos

A mi madre, por ser ella, por su fortaleza y por abrirnos el camino a mí y a mis hermanos hacia la universidad. A mis hermanos, por ser también mis amigos incondicionales en todo momento.

A mi gato Tom, por alegrarme las tardes con su compañía, y a La Chata y La Chiku, las gatas más cariñosas, que siempre me sacaron de la rutina y me acompañaron con su ternura.

A "la vieja Betsy", mi fiel laptop HP 14-ck0001la con su procesador Dual-Core Intel® Celeron® N4000 @ 1.10GHz, que ha resistido desde mi pregrado hasta hoy, siendo testigo de dos tesis completas: una de pregrado y esta de posgrado.

A la Universidad Nacional de Colombia, que, a pesar de todo, finalmente me acogió como una de sus estudiantes y al Grupo de Investigación en Bioinformática y Biología de Sistemas GIBBS, por abrirme sus puertas.

Al profesor Andrés, por adoptarme académicamente desde mi pregrado y seguir confiando en mí hasta hoy, guiándome con compromiso y entusiasmo en la presentación de este trabajo de grado.

Y al profesor Carlos, por su paciencia, dedicación y valiosa enseñanza a lo largo de este proceso investigativo y de formación profesional.

A todos, gracias.

Resumen

Título en español: Caracterización genómica de un grupo de individuos con Deterioro Cognitivo Leve (DCL) en población del Atlántico, Colombia.

El deterioro cognitivo leve (DCL) es una etapa temprana de la enfermedad de Alzheimer (EA), caracterizada por una disminución en la cognición sin llegar a ser demencia. Factores genéticos influyen en su desarrollo, y se han identificado variantes asociadas a su progresión, destacando el alelo APOE- ϵ 4 como un marcador clave en la transición de DCL a EA.

La prevalencia de DCL varía entre 5,0% y 36,7% a nivel mundial, mientras que en Colombia afecta al 17,5% de la población, con mayor incidencia en la región del Atlántico. A pesar de las herramientas diagnósticas actuales, es necesario un enfoque multiómico para una comprensión más profunda de la enfermedad.

Los avances en secuenciación de nueva generación (NGS) permiten identificar factores genéticos clave en enfermedades neurodegenerativas, facilitando el desarrollo de estrategias preventivas y terapéuticas. En este contexto, surge la oportunidad de realizar un estudio de asociación y caracterización genómica en pacientes con DCL y controles de la población del Atlántico, Colombia, con el objetivo de profundizar en los mecanismos genéticos involucrados en esta región, que presenta una alta prevalencia de enfermedades neurodegenerativas.

Palabras clave: Deterioro Cognitivo Leve, Atlántico, GWAS, Genómica.

Abstract

Título en inglés: Genomic Characterization of a Cohort with Mild Cognitive Impairment (MCI) in the Population of Atlántico, Colombia

Mild Cognitive Impairment (MCI) is an early stage of Alzheimer's disease (AD), characterized by a decline in cognition without reaching the threshold for dementia. Genetic factors influence its development, and several variants associated with its progression have been identified, with the APOE- ϵ 4 allele standing out as a key marker in the transition from MCI to AD.

The prevalence of MCI ranges from 5.0% to 36.7% worldwide, while in Colombia, it affects 17.5% of the population, with the highest incidence in the Atlántico region. Despite the availability of current diagnostic tools, a multi-omics approach is necessary for a deeper understanding of the disease.

Advances in next-generation sequencing (NGS) enable the identification of key genetic factors in neurodegenerative diseases, facilitating the development of preventive and therapeutic strategies. In this context, the opportunity arises to conduct an association study and genomic characterization of MCI patients and controls from the Atlántico region, Colombia, aiming to gain deeper insights into the genetic mechanisms involved in this population, which exhibits a high prevalence of neurodegenerative diseases.

Keywords: Mild Cognitive Impairment , Atlántico, GWAS, Genomics.

Este Trabajo Final de maestría fue calificado en junio de 2025 por la siguiente evaluadora:

Liliana Lopez Kleine
Profesora Departamento de Estadística
Facultad de Ciencias
Universidad Nacional de Colombia, Sede Bogotá

Contenido

	Pág.
Introducción.....	1
1. Antecedentes.....	5 ^[Obj]
2. identificación del Problema.....	7
2.1 Planteamiento del problema.....	7
2.2 Justificación.....	8
3. Objetivo general y objetivos específicos.....	10
3.1 Objetivo General.....	10
3.2 Objetivos específicos.....	10
4. Marco Teórico.....	11
4.1 El envejecimiento cognitivo (EC).....	11
4.2 El Deterioro Cognitivo (DC).....	11
4.2.1 El DC como un posible estado de transición entre el EC y la demencia.....	11
4.3 El Deterioro Cognitivo Leve (DCL).....	12
4.3.1 En La Región.....	12
4.3.2 DCL En Cifras.....	12
4.4 GWAS y EN, aproximaciones y diseño experimental.....	13
4.4.1 Asociaciones y variabilidad en DCL.....	13
4.4.2 Disponibilidad de información, bases de datos y consorcios genéticos.....	14
5. Metodología.....	15
5.1 Diagnostico del grupo de pacientes.....	15
5.2 Toma de muestra.....	22
5.3 Extracción de ADN.....	22
5.4 Secuenciación.....	22
5.4.1 Control de calidad.....	22
5.5 Procesamiento de los datos.....	23
Mapeo de lecturas al genoma de referencia.....	23
Preparación de los ensamblajes.....	23
Llamado de variantes.....	24
Filtrado y preparación para el llamado de variantes.....	24
Exportación de variantes y visualización.....	25
5.6 Subtítulos nivel 2.....	25
5.6.1 Subtítulos nivel 3.....	25
6. Conclusiones y recomendaciones.....	27

6.1 Conclusiones.....27
6.2 Recomendaciones.....27

Lista de figuras

	Pág.
Figura 1: Calidad de secuencia por base.....	23
Figura 2: Puntuaciones de calidad por secuencia.....	24
Figura 3: Distribución de asociaciones genéticas obtenidas por GWAS en pacientes con DCL del Atlántico.....	27
Figura 4: Variantes genéticas (SNPs) con asociación sugerente en pacientes con DCL. 31	
Figura 5: Proporción de variantes missense y sinónimas detectadas en el análisis funcional con VEP.....	32
Figura 6: Número de variantes asociadas al gen cercano con VEP.....	34
Figura 7: Enriquecimiento de términos en g:Profiler.....	35
Figura 8: Distribución de diferencias de frecuencias Alélicas.....	37
Figura 9: Comparación de frecuencias alélicas.....	38
Figura 10: PCA con Clustering K-means.....	40
Figura 11: PCA diferenciando casos y controles.....	41
Figura 12: Gráfico de Ancestría en la población de DCL.....	43

Lista de tablas

Pág.

Tabla 1: Variantes genéticas (SNPs) con asociación sugerente en pacientes con DCL.. 30

Lista de Símbolos y Abreviaturas

Símbolos con letras latinas

Símbolo	Término	Unidad SI	Definición
p	Valor de significancia estadística	-	Probabilidad de obtener un resultado igual o más extremo que el observado bajo la hipótesis nula.
OR	Razón de momios (odds ratio)	-	Medida de asociación que indica la probabilidad de un evento en dos grupos.
CHISQ	Estadístico Chi-cuadrado	-	Valor que mide la diferencia entre resultados observados y esperados.
k -mer	Subcadena de longitud k en una secuencia	-	Fragmento de longitud k en una secuencia, útil en análisis bioinformáticos.

Símbolos con letras griegas

Símbolo	Término	Unidad SI	Definición
ϵ	Épsilon (en APOE- ϵ 4)	-	Subtipo del gen APOE, asociado al riesgo genético de Alzheimer.

Superíndices

Superíndices Término

rsID	Identificador de polimorfismo de nucleótido único
IGFBP2	Insulin-like Growth Factor Binding Protein 2
10 ⁻⁵	Notación científica para valores pequeños, usados en p-valores sugestivos
10 ⁻⁸	Notación científica para significancia genómica en GWAS

Abreviaturas

Abreviatura Término

DCL	Deterioro Cognitivo Leve
EA	Enfermedad de Alzheimer
EP	Enfermedad de Parkinson
DFT	Demencia Frontotemporal
ADD	Alzheimer's Disease and related Dementias
LOAD	Late-Onset Alzheimer's Disease
GWAS	Estudio de Asociación del Genoma Completo
SNP	Polimorfismo de un solo nucleótido
APOE	Apolipoproteína E
WGS	Secuenciación del Genoma Completo
WES	Secuenciación del Exoma Completo
VEP	Predictor de Efecto de Variantes
GATK	Kit de Análisis del Genoma
FASTQ	Formato de archivo de lecturas y calidad de secuenciación
VCF	Formato de archivo para variantes genéticas
BAM	Archivo binario de alineamiento
SAM	Archivo de alineamiento en texto
BQSR	Recalibración de calidad de bases
VQSR	Recalibración de calidad de variantes
PCA	Análisis de Componentes Principales
MDS	Escalamiento Multidimensional
DLG	Genómica basada en aprendizaje profundo
MAF	Frecuencia del Alelo Menor
ND	Enfermedad Neurodegenerativa
UPS	Sistema Ubiquitina-Proteasoma
ALP	Vía Autofagia-Lisosoma
DUBs	Enzimas desubiquitinantes
UTR	Región no traducida del ARNm
lncRNA	ARN largo no codificante

NMD	Degradación mediada por codón de parada
ADNI	Iniciativa de Neuroimagen de la EA
TRKB	Receptor tirosina quinasa tipo B
BDNF	Factor neurotrófico derivado del cerebro
GABA	Ácido γ -aminobutírico
TDAH	Trastorno por Déficit de Atención e Hiperactividad

Introducción

El deterioro cognitivo leve (DCL) es una condición que se distingue del envejecimiento normal, ya que involucra alteraciones cognitivas mayores a las esperadas para la edad, pero sin alcanzar la gravedad propia de una demencia (Petersen, 2004). Actualmente, se reconocen diferentes subtipos de DCL, como el de dominio único, múltiple amnésico y no amnésico, lo cual indica que no todos los casos son progresivos ni representan una fase inicial de demencia, especialmente en el caso del DCL no amnésico (Anderson, 2019; Angela M. Sanford, 2017; Roberts & Knopman, 2013).

En este contexto, el DCL se ha descrito como una etapa temprana de algunas enfermedades neurodegenerativas, incluyendo la enfermedad de Alzheimer (EA), la enfermedad de Parkinson (EP) y las demencias frontotemporales (DFT) (Aarsland et al., 2021; de Mendonça et al., 2004; Kelley & Petersen, 2007). Persiste el debate sobre si el DCL es una entidad clínica independiente o una etapa temprana de enfermedades neurodegenerativas, ya que su clasificación como “demencia cuestionable” en el sistema CDR (Hughes et al. 1982) ha llevado a algunos autores a no considerarla como una patología aparte, sino más bien como una etapa inicial o limítrofe de trastornos como la EA (Anderson, 2019). No obstante, aunque su uso como categoría diagnóstica es controversial, se ha considerado útil como apoyo en la detección temprana y prevención de la demencia (Moreira & Bond, 2008).

Ahora bien, desde el inicio del Proyecto del Genoma Humano, se ha reconocido el papel clave de los factores genéticos en la patogénesis de enfermedades complejas como las demencias (X. Zhang, 2020). Sin embargo, la investigación genética en el DCL ha sido limitada y ambigua, en parte porque suele considerarse un estado prodrómico de la enfermedad de la EA y no una entidad clínica autónoma. En este contexto, uno de los genes más estudiados es *APOE*, la presencia del alelo *APO-ε4* se ha asociado con un

mayor riesgo de desarrollar DCL con una reducción acelerada del volumen hipocampal y la amígdala (Roberts & Knopman, 2013). Pero en contraste, el alelo *APO-ε2* ha mostrado un posible efecto protector al vincularse con una menor atrofia en la corteza temporal transversal (Hostage et al., 2013). Otros estudios sobre variantes genéticas enriquecidas en poblaciones amerindias o africanas, han encontrado una interacción nominalmente significativa entre la variante rs8112679, enriquecida en África, y el alelo *APOE-ε4* en el contexto de DCL (Granot-HersHKovitz et al., 2023). Además, otras variantes como *CLU*-RS11136000-G y *MS4A6A*-RS610932-C también se han asociado con mayor riesgo de deterioro cognitivo y progresión a DCL/LOAD, pero contrario a la variante *PICALM*-rs3851179-G que ha mostrado un efecto protector (Carrasquillo et al., 2015).

Sin embargo, la utilidad del *APOE-ε4* como marcador diagnóstico aislado sigue siendo cuestionable. Aunque se puede asociar tanto a pacientes con DCL como con Alzheimer de inicio tardío (LOAD), este alelo resulta más informativo cuando se combina con criterios clínicos o hallazgos post-mortem (Mayeux et al., 1998; Reitz & Mayeux, 2010).

En la última década, técnicas como el aprendizaje profundo han mejorado el análisis de variantes genéticas en DCL. Wen et al. (2022) identificaron 17 asociaciones entre variantes genéticas independientes (ISV) y subtipos neuroanatómicos en 1.487 participantes del proyecto ADNI, ocho de ellas significativas. Además, se han reportado variantes en los cromosomas 8, 15 y 1 (rs7840202, rs11637611, rs12752888) vinculadas con progresión acelerada de DCL a EA, explicando hasta un 16,6 % de los casos (X. Hu et al., 2011). Otras variantes, como rs17090219, rs192470679 y rs10903488, se relacionan con mayor deterioro cognitivo (Lee et al., 2017), mientras que rs6877893 y rs6448456 se asociaron con menor riesgo de DCL. Finalmente, la variante rs723672 ha mostrado una mayor susceptibilidad, aunque con resultados no consistentes (J. Zhang et al., 2023). Finalmente, se han identificado genes potencialmente relevantes para el estudio del DCL, como el caso de *SORL1*, cuya baja expresión se ha asociado con mayor deterioro cognitivo y posible progresión a Alzheimer (Sager et al., 2007).

En Colombia, diversos estudios han abordado el deterioro cognitivo desde una perspectiva epidemiológica y de prevalencia. En Manizales, Díaz Cabezas et al. (2013) reportaron una prevalencia del 39,4% en adultos mayores de 65 años, asociada con mayor edad, baja escolaridad, polifarmacia, hipertensión y dislipidemia,

recomendando pruebas de tamizaje para su detección. De manera similar, Alvarado et al. (2014) encontraron que el 40% de 803 adultos mayores evaluados con la prueba del reloj presentaban signos de deterioro cognitivo, especialmente entre los 70 y 79 años, y en personas con hipertensión, diabetes o bajo nivel educativo. Por su parte, el análisis de la encuesta SABE Colombia 2015, con 23.694 adultos mayores, evidenció una prevalencia del 19,7%, mayor en mujeres, y asociada con factores socioeconómicos como el trabajo activo (factor de riesgo) y la educación primaria (factor protector), con diferencias según el sexo (Sarmiento Buitrago et al., 2024).

Específicamente en cuanto a DCL, Henao Arboleda et al. (2008) reportaron una prevalencia del 9,7% de DCL amnésico en mayores de 50 años del Valle de Aburrá, con mayor frecuencia en hombres y menor en personas con mayor escolaridad, sugiriendo su posible rol como etapa inicial del Alzheimer. Cabe destacar que este tipo de patologías representan un desafío creciente para el sistema de salud, especialmente en el contexto del envejecimiento poblacional y en regiones como el Atlántico, que registra la mayor prevalencia de patologías mentales (12,9%) en el país (Ministerio de Salud y Protección Social, 2017).

A pesar de estos avances, la investigación genética en DCL sigue siendo limitada. Los estudios de asociación pueden ofrecer claves sobre la susceptibilidad a enfermedades complejas como el DCL, pero en Colombia los esfuerzos se han centrado principalmente en la prevención primaria y terciaria (Ministerio de Salud y Protección Social, 2017). Se requiere avanzar hacia estrategias de prevención secundaria que permitan el diagnóstico y tratamiento oportuno. Por ello, este estudio busca reducir la brecha de conocimiento entre genética y deterioro cognitivo leve en una población subrepresentada como la colombiana, analizando la relación entre características genómicas en pacientes con DCL y controles sanos del departamento del Atlántico.

1. Antecedentes

En el ámbito de la genómica, con las tecnologías de nueva generación se ha podido avanzar en la caracterización de EN a nivel mundial. Por un lado tenemos la secuenciación del genoma completo – WGS (en inglés Whole Genome Sequencing) que permite identificar en gran medida la variación genética de un individuo, este enfoque se ha visto como promisorio debido a las existentes y nuevas tecnologías en desarrollo que han permitido genotipificar organismos complejos (Ng & Kirkness, 2010). Además, los estudios de asociación del genoma completo – GWAS (en inglés Genome-wide association study) han contribuido en gran medida a comprender enfermedades complejas a la hora de evaluar la variación genética asociada que se puede presentar con cada enfermedad/fenotipo, proporcionando nuevas ideas y métodos para explorar la patogénesis genética de las enfermedades (X. Zhang, 2020).

Los estudios GWAS involucran procedimientos fundamentales para su desarrollo, en una primera fase la colecta de los datos implica la selección de la población de estudio como su posterior genotipado. En este punto los estudios pueden discrepar al realizar el genotipado que puede estar dirigido a variantes específicas y preseleccionadas utilizando microarrays, o realizar estudios de secuenciación de todo el exoma – WES (en inglés Whole Exome Sequencing) o un WGS (en inglés Whole Genome Sequencing) que tiene como objetivo capturar toda variación genética (Uffelmann et al., 2021).

Por otro lado, se han documentado asociaciones de DCL a otro tipo de enfermedades como la enfermedad cardiovascular (ECV) con la identificación de SNP en variantes genéticas en ARHGAP26 que se asociaron con EA y ECV (K. Wang et al., 2022). También se ha encontrado asociación de la enfermedad de las arterias coronarias (EAC) como factor de riesgo en EA, y EAC relacionado con DCL en la progresión a EA (Elman et al., 2019). Asimismo, los factores asociados al sueño pueden jugar un papel fundamental en la progresión de una demencia en donde en la comunidad latina se cree que las medidas genéticas subyacentes al insomnio, la somnolencia diurna y la duración del sueño están asociadas con el riesgo de DCL (Y. Zhang et al., 2023). Además se ha

estudiado el papel de la capacidad auditiva como predictor de DCL (Fu et al., 2023), junto con factores ambientales asociados con genes truncados en ABCA7 (Xi et al., 2020).

1.1 Variabilidad Genética

Se han hecho principalmente estudios relacionando polimorfismos de un solo nucleótido (SNP) de EA como marcadores de riesgo para el DCL. Por un lado, se ha observado que EA esta asociada a una perdida de volumen cerebral en donde existen SNP's en el gen GRIN2B implicado en el aprendizaje y la memoria (Stein et al., 2010) o variantes tempranas en RBFOX1 que puede ser un potencial marcador para DCL (Raghavan et al., 2020) junto con asociaciones a genes como APOE, CYP19A1 y NCAM2 (Han et al., 2010). Por ejemplo se ha identificado un efecto en el volumen cortical entorrinal en un intron del gen ZNF292 y otro SNP en el gen ARPP-21 con efectos en el grosor cortical, este tipo de variaciones sugieren una relación con la progresión a la EA utilizando datos de imágenes para la identificación de SNP (Furney et al., 2011; Gomar, 2011; Kong et al., 2015). Hoy en día hay un creciente desarrollo de nuevos métodos para detectar variantes discriminantes características de pacientes que disminuyen la heterogeneidad de la enfermedad, por ejemplo, la técnica Deep-Learning e genómica de aprendizaje profundo (DLG) permitió identificar nuevos marcadores genéticos en la progresión de EA incluidos rs6311 y rs6313 en HTR2A, rs1354269 en NAV2 y rs690705 en RFC3 (L. Li et al., 2021).

1.2 Otras aproximaciones

Sumado a lo anterior, no solo se tiene asociaciones entre la EA y DCL, sino también asociaciones con la Enfermedad de Parkinson (EP) en donde se han identificado SNP en el gen PICALM, sin embargo enfatizan que dicho gen pueden jugar un papel protector en la enfermedad de Parkinson y posibles alteraciones se pueden relacionar con demencias (Periñán et al., 2021). Se debe tener en cuenta que la EA esta asociada a un deterioro cognitivo y EP a un deterioro motor asociado y que esta puede ser discriminada según la batería neuropsicológica que se aplique (Ophey et al., 2021).

2. Identificación del Problema

2.1 Planteamiento del problema

El DCL conlleva una serie de problemáticas asociadas, siendo su heterogeneidad uno de los aspectos más destacados. Actualmente, existe una discusión vigente sobre la clasificación de subtipos dentro del DCL. Algunas de las clasificaciones propuestas incluyen el DCL de dominio único, el DCL de dominio múltiple amnésico y el DCL no amnésico (Anderson, 2019), esto representa un factor adicional que sugiere que no todos los casos de DCL son precursores de la demencia y no todos son progresivos como puede ser en DCL de tipo no amnésico (Angela M. Sanford, 2017). Dicha heterogeneidad permite discutir sobre las aproximaciones que se deben abordar a la hora de realizar estudios de variantes genéticas, si realizar el típico caso-control o intentar dividir en diferentes endofenotipos el estudio (Espinosa et al., 2018).

Además de los daños neuronales asociados al DCL, los pacientes experimentan diversas dificultades cognitivas que se relacionan con una disminución en su calidad de vida diaria, un aumento en los niveles de depresión, ansiedad y estrés percibido, así como un menor bienestar mental en general (Stites et al., 2018); esto refleja una participación social deficiente y provoca que los pacientes vean afectado su desempeño en las actividades cotidianas y en sus relaciones personales más íntimas (Anderson, 2019; Hussenoeder et al., 2020).

Sumado a lo anterior, la dificultad al diagnosticar el DCL y el dilema de los diversos endofenotipos que pueden generar un falso positivo y que a la hora de enfocar la terapia en el individuo puede generar problemas adversos al realizar un diagnóstico impropio (del Carmen Díaz-Mardomingo et al., 2017). El DCL al ser

un estado preclínico de enfermedades de mayor complejidad, se hace necesario desarrollar estrategias que posiblemente nos permitan realizar un diagnóstico temprano con el fin de reducir la tasa de crecimiento de las EN en Colombia.

Para Colombia este tipo de enfermedades podría representar una mayor carga en el sistema de salud; teniendo en cuenta el envejecimiento poblacional y resaltando específicamente al departamento de Atlántico con la mayor prevalencia de patologías mentales con un 12.9% en su población (Ministerio de Salud y Protección Social, 2017). Se plantea ¿Existe relación entre las características genómicas de pacientes con DCL y controles en la población del Atlántico?

2.2 Justificación

El deterioro cognitivo leve (DCL) es una EN que afecta principalmente personas de la tercera edad. Se cree que el DCL es un estado preclínico de la Enfermedad del Alzheimer (EA) y su estudio se enfoca principalmente en conocer cómo el DCL podría funcionar como una herramienta de apoyo para la detección temprana de enfermedades neurodegenerativas más complejas (Angela M. Sanford, 2017; Morozova et al., 2022) como las enfermedades de Alzheimer, Parkinson o Huntington (Aarsland et al., 2021; Duff et al., 2010; Kelley & Petersen, 2007).

Existen diversos factores de riesgo asociados a DCL, pero principalmente la edad es un punto a tener en cuenta en el diagnóstico; sin dejar de lado comorbilidades como la obesidad. Asimismo, se cree que la genética contribuye a marcadores que pueden ser susceptibles para el desarrollo de DCL (Campbell et al., 2013; J. Q. Li et al., 2016).

Según estimaciones del Departamento Administrativo Nacional de Estadística (DANE) para el año 2070 se cree que Colombia tendrá 183 adultos mayores por cada 100 personas menores de 15 años, esta relación es mayor con respecto a otros de la región como México, Argentina o Perú; Además, la pandemia COVID-

19 agravó las estimaciones de la población mayor de 60 años que crece con mayor velocidad; se cree que para el año 2070 el porcentaje de personas mayores de 80 años pasará de 8.1% al 9.7% (Mesa Interinstitucional de Población, 2023). Esto prevé un aumento de trastornos cognitivos como el DCL y enfermedades neurodegenerativas como la EA, EP y Huntington en la población Colombiana en los próximos años (Ministerio de Salud y Protección Social, 2017).

Las EN se destacan entre el plan decenal de salud pública 2012-2021 con una creciente incidencia en la población Colombiana que no solo afecta directamente la calidad de vida del paciente sino en forma indirecta a sus familiares cercanos (Ministerio de Salud y Protección Social, 2013).

Dado que los elementos genéticos desempeñan un rol crucial en la aparición de enfermedades complejas (X. Zhang, 2020). La implementación de nuevas metodologías como los estudios de asociación de genoma completo, posibilitará la identificación de variantes genéticas de riesgo o protección que nos permitirá tener una mejor comprensión de la bases genética del DCL en la población del Atlántico. Estos nuevos conocimientos ayudarán a la planificación de aproximaciones innovadoras para el diagnóstico y detección temprana del DCL y otras enfermedades neurodegenerativas en nuestro país.

3. Objetivo general y objetivos específicos

3.1 Objetivo General

Analizar la relación entre las características genómicas identificadas entre pacientes con DCL y controles sanos en una muestra de la población del Atlántico, Colombia.

3.2 Objetivos específicos

- Obtener la secuencia de los genomas completos de pacientes con DCL y controles de la población del departamento del Atlántico.
- Describir las características genómicas más relevantes de los genomas ensamblados utilizando un análisis *in silico*.
- Analizar las principales variantes genéticas asociadas a enfermedades neurodegenerativas detectadas en Europa en la población del Atlántico.

4.Marco Teórico

4.1 El envejecimiento cognitivo (EC)

El envejecimiento cognitivo (EC) normalmente se presenta en las personas cuando alcanzan una edad avanzada, manifestándose a través de la disminución de la memoria, la fluidez verbal, la capacidad de razonamiento matemático y la eficiencia en el análisis, además de la reducción en la velocidad de procesamiento (Whitley et al., 2016); esta acompañado de cambios morfológicos como la pérdida de volumen cerebral y el adelgazamiento de la corteza frontal, ambos importantes para el desarrollo de procesos cerebrales regulares (Borrás Blasco & Viña Ribes, 2016).

4.2 El Deterioro Cognitivo (DC)

El deterioro cognitivo (DC) se define como una alteración cognitiva que excede las expectativas esperadas para la edad y otros factores demográficos que influirían en su desarrollo (Benavides-Caro, 2017). El DC implica la reducción de las capacidades cognitivas y está relacionado con cambios en la estructura morfológica, factores fisiológicos y ambientales que pueden variar de forma única en cada individuo (Borrás Blasco & Viña Ribes, 2016). De esta manera, el DC desempeña un papel específico en la transición de un EC normal hacia la posible aparición de demencia.

4.2.1 El DC como un posible estado de transición entre el EC y la demencia.

Debido al envejecimiento de la población actual, se ha observado un aumento a nivel mundial en la incidencia de enfermedades neurodegenerativas, que incluyen demencias (Gjøra et al., 2021). Estas enfermedades, en la actualidad, carecen de tratamientos curativos o que modifiquen su curso (García-Cifuentes et al., 2019). Entre estas

enfermedades, la Enfermedad del Alzheimer (EA) es la variante más común, representando entre el 50% y el 60% de todos los casos (Blennow et al., 2006), Para el año 2015, se estimaba una prevalencia de 46,8 millones de personas afectadas por EA, con una proyección alarmante de 131,5 millones de afectados para el año 2050, lo que implica un nuevo caso cada 3,2 segundos (Prince et al., 2015); En Colombia, se estima que la incidencia de demencia puede alcanzar una tasa de 13,1 casos por cada 1.000 habitantes (Pradilla A. et al., 2003).

4.3 El Deterioro Cognitivo Leve (DCL)

El DCL se considera una enfermedad neurodegenerativa que conlleva el desarrollo de un déficit cognitivo mayor de lo normal y que a su vez no cumple con los criterios de una demencia. Los criterios para su diagnóstico incluyen el declive o deterioro cognitivo, dependiendo del manual de diagnóstico utilizado (DSM-IV, CIE-10); declaración propia o de familiares de déficit de memoria u deterioro en otros dominios cognitivos y la alteración en las actividades de la vida diaria (Gutiérrez Rodríguez & Guzmán Gutiérrez, 2017; Petersen, 2004, 2016).

4.3.1 En La Región

Para latinoamérica, en la población mexicana se han identificado 10 sitios CpG que se metilaron de manera diferencial entre los controles e individuos con DCL, estas variaciones se encontraron en cuatro regiones metiladas diferencialmente en *TMEM232*, *SLC17A8*, *ALOX12* y *SEPT8*, dichos genes están relacionados con la inflamación sistémica instigada metabólicamente y la interrupción del transporte vesicular sináptico (Pathak et al., 2019).

4.3.2 DCL En Cifras

Además, la prevalencia de DCL varía considerablemente en la población mundial, oscilando entre el 5,0% y el 36,7% (Bai et al., 2022; Sachdev et al., 2015), Sin embargo, en Colombia, esta condición representa una preocupación para las entidades estatales, ya que se ha observado un aumento en la población colombiana, alcanzando un porcentaje del 17,5% (Ministerio de Salud y Protección Social & Colciencias, 2015). La aplicación de pruebas neuropsicológicas es la herramienta estándar para la detección temprana de los cambios cognitivos. Entre estas podemos mencionar la prueba Mini-Mental State Examination (MMSE), el Addenbrooke's Cognitive Examination - revisado (ACE-R) o el Montréal Cognitive Test Assessment (MoCA) utilizadas principalmente para

la identificación del deterioro cognitivo en estadios tempranos asociado a demencia (Gil et al., 2015; Ospina García, 2015; Pedraza L et al., 2014). En Colombia se han realizado contribuciones al entendimiento del DCL siendo estudiado principalmente desde perfiles neuropsicológicos que permitieron la caracterización de esta enfermedad (Ríos-Gallardo et al., 2017) así como la identificación de síntomas neuropsiquiátricos asociados a DCL de tipo amnésico (Henao-Arboleda et al., 2010).

4.4 GWAS y EN, aproximaciones y diseño experimental.

4.4.1 Asociaciones y variabilidad en DCL

En DCL se han identificado loci en *BR5* y Tay-Sachs, estos se asociaron con una disminución significativa de la cognición, así como con la conversión de sujetos con DCL a un diagnóstico de EA (X. Hu et al., 2011). Además, se resaltan los genes *APOE* y *TOMM40* como fuertes marcadores relacionados con la pérdida de densidad de grosor cortical de la materia gris, o el gen *NXPH1* clave en la integridad sináptica distintivo en la EA (L. Shen et al., 2010). Con respecto a *APOE*, se cree que el alelo *APOE-ε4* tiene un pequeño efecto en pacientes de DCL que evolucionan a EA (Lacour et al., 2017).

Existen otros SNP asociados a DCL a EA en genes como *BDH1*, *ST6GAL1*, *RAB20*, *PDS5B*, *ADARB2* y *SPSB1* (Lee et al., 2017). No obstante, también se han estudiado los niveles de plasma de Beta-secretasa 1 en la predicción de EA en pacientes con DCL, dicha investigación presenta variaciones en el gen *BACE1* que se encuentra aumentada en pacientes con DCL (Y. Shen et al., 2018). Finalmente, otras variantes como en el gen *PDE4D*, asociado directamente con DCL, es un factor de riesgo para las enfermedades degenerativas neuronales (Xiang et al., 2020) u otros genes relacionados con la neurogénesis como el gen *ADORA2A* en donde está involucrado en el crecimiento del hipocampo (Horgusluoglu-Moloch et al., 2017).

4.4.2 Disponibilidad de información, bases de datos y consorcios genéticos

Finalmente, se tiene información de consorcios especializados en documentar todos los metadatos asociados a las EN como la Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI) que incluyen información sobre genotipos y biomarcadores, incluidas medidas de volumen derivadas para estructuras relevantes como el hipocampo y la amígdala (Brabec et al., 2021). Este consorcio a encontrado principales genes como *APOE*, *BIN1*, *CLU*, *CR1* y *PICALM* relacionados con la EA (Saykin et al., 2015). Además se tiene bien documentado el proteoma del cerebro humano, como una herramienta útil a la hora de enfocar nuestros esfuerzos en la proteómica de la enfermedad (Robins et al., 2021).

5. Metodología

Para llevar a cabo el objetivo general del trabajo de grado se determinó un estudio observacional de tipo caso control. Para tal fin se establecieron puntos críticos para el desarrollo del mismo.

5.1 Diagnóstico del grupo de pacientes

El proceso de diagnóstico fue llevado a cabo por profesionales del Grupo de Investigación de Neurociencias del Caribe de la Universidad Simón Bolívar. El diagnóstico de Deterioro Cognitivo Leve (DCL) siguió un protocolo de cinco pasos: (1) socialización del estudio al paciente, (2) verificación del cumplimiento de los criterios de inclusión/exclusión, (3) lectura y firma del consentimiento informado, (4) evaluación neurológica inicial, y (5) evaluación neuropsicológica.

Posteriormente, se dio paso a la aplicación de baterías neuropsicológicas, comenzando con la Escala de Inteligencia de Wechsler para Adultos (WAIS IV) (Wechsler, 2012) para evaluar la capacidad cognitiva. Si el paciente cumplía los criterios de inclusión, se continuaba con pruebas de rastreo como el Mini-Mental State Examination (MMSE) (Folstein et al., 2014) y la Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2014), las cuales evaluaban áreas clave como la memoria, atención, lenguaje y habilidades visoespaciales.

El proceso, basado en los criterios de Petersen (1999, 2014) y Winblad et al. (2014), incluyó tanto la evaluación neurológica como la neurocognitiva, considerando distintos subtipos de DCL (amnésico, multidominio y no amnésico). Las pruebas midieron funciones relacionadas con la memoria, funcionalidad diaria y otras capacidades cognitivas, asegurando así una evaluación integral del paciente. Para este trabajo de grado, se tomaron en cuenta 45 muestras de pacientes.

5.2 Toma de muestra

El procedimiento realizado por el grupo de investigación de la Universidad Simón Bolívar se basó en la recolección de muestras de sangre mediante venopunción a los pacientes y controles. Las muestras, con un volumen de entre 3 y 5 mL, fueron depositadas en tubos con EDTA y conservadas a -20°C hasta su procesamiento para la secuenciación.

5.3 Extracción de ADN

Para la extracción de ADN, se utilizó el kit comercial QIAwave DNA Blood & Tissue Kit, siguiendo las recomendaciones del fabricante. El ADN fue purificado a partir de 100 μl de sangre completa de los pacientes, a partir de lo cual se obtuvo un rendimiento de 3–6 μg de ADN.

El protocolo incluyó una fase de lisado con proteína K, un posterior cargue del lisado en la columna DNeasy Mini Spin del kit, la centrifugación de las muestras y la unión del ADN a la membrana. Finalmente, los contaminantes restantes y los inhibidores de enzimas fueron eliminados durante el proceso de lavado, y el ADN fue eluido en agua o tampón.

Este procedimiento fue llevado a cabo por el Grupo de Investigación Neurociencias del Caribe de la Universidad Simón Bolívar.

5.4 Secuenciación

La secuenciación de los 45 pacientes se realizó a través de una empresa privada especializada que emplea la tecnología DNA Nanoball por MGI Tech (Porreca, 2010; Y. Xu et al., 2019), asegurando una alta precisión y calidad en los datos genómicos obtenidos (T. Hu et al., 2024; Mak et al., 2017).

5.5 Control de calidad

Los distintos pasos de la secuenciación de nueva generación (NGS, por sus siglas en inglés) pueden estar influenciados por errores instrumentales o técnicos que reducen la calidad del proceso de secuenciación (Rollano & Mollinedo, 2017). Estos errores pueden ocasionar fallos en los análisis posteriores, por lo que se recomienda realizar una visualización preliminar del estado de las secuencias y llevar a cabo un control de calidad si es necesario (Smith & de Sena Brandine, 2021). Comúnmente, se examina la calidad por base en las lecturas, así como la calidad de los k-mers o la presencia de contaminación por adaptadores u otros elementos añadidos durante la preparación de las librerías (Adam & Nathan, 2020).

Por esta razón, se realizó un análisis inicial utilizando herramientas como FastQC (Andrews, 2010) y MultiQC (Ewels et al., 2016) con el fin de identificar posibles errores en el proceso de secuenciación.

5.6 Procesamiento de los datos

Para el procesamiento de los datos, se utilizó el flujo de trabajo Genome Analysis Toolkit (GATK) del Instituto Broad, Estados Unidos (DePristo et al., 2011; McKenna et al., 2010). Asimismo, se tuvo en cuenta la documentación proporcionada por la Universidad de Melbourne, Australia (The University of Melbourne, n.d.).

5.6.1 Preparación de Archivos para Análisis Genómicos

Se utilizó una colección de archivos estándar para análisis en humanos, proporcionada por el equipo del Genome Analysis Toolkit (GATK Team) a través del GATK Resource Bundle, alojado en un depósito de almacenamiento en Google Cloud.

Para automatizar el proceso de obtención de estos archivos, se diseñó el script *download_resources.sh*, el cual facilitó la descarga de los datos mediante el uso de la herramienta *wget*. Este script, junto con otros mencionados en este documento, se encuentra disponible en el repositorio correspondiente del proyecto en GitHub (<https://github.com/jlargog/Pipeline-Variantes-Geneticas>).

En el script se definió una lista específica de archivos en formato VCF y sus respectivos índices, fundamentales para la anotación y el análisis de variantes genéticas, incluyendo información de genomas humanos, SNPs e indels. Estos archivos fueron descargados automáticamente desde el servidor del Broad Institute para su posterior uso en el análisis.

Entre los archivos descargados se encontraron los siguientes:

- 1000G_omni2.5.hg38.vcf.gz
- 1000G_phase1.snps.high_confidence.hg38.vcf.gzAxiom_Exome_Plus.genotypes.all_populations.poly.hg38.vcf.gz
- Homo_sapiens_assembly38.dbsnp138.vcf
- Homo_sapiens_assembly38.known_indels.vcf.gz
- Mills_and_1000G_gold_standard.indels.hg38.vcf.gz
- hapmap_3.3.hg38.vcf.gz

5.6.2 Mapeo de lecturas al genoma de referencia

Con el fin de generar ensamblajes de alta calidad y facilitar la identificación precisa de variantes genéticas, se realizó el alineamiento de las lecturas contra el genoma de referencia humano GRCh38.p13 (Genome Reference Consortium, 2019). Para ello, se utilizó el archivo Homo_sapiens_assembly38.fasta, correspondiente al conjunto de referencia proporcionado por GATK (Broad Institute), el cual representa una versión modificada del ensamblaje GRCh38 optimizada para análisis genómicos. Esta versión incorpora mejoras como el enmascaramiento de regiones repetitivas, la inclusión de contigs alternativos y secuencias decoy, características que aumentan la fidelidad del mapeo y reducen ambigüedades en regiones complejas del genoma.

Ahora bien, fue esencial contar con un genoma de referencia apropiado que facilitara una anotación precisa. En este contexto, el flujo de trabajo recomendado por GATK sugiere el uso del alineador BWA para llevar a cabo el mapeo de las lecturas (H. Li & Durbin, 2009). En consecuencia, se descargó el archivo Homo_sapiens_assembly38.fasta, el cual fue indexado conforme a los requerimientos del algoritmo de BWA y utilizado posteriormente en el proceso de alineamiento.

5.6.3 Preparación de los ensamblajes

Los archivos de salida generados por BWA fueron obtenidos en formato SAM, lo que implicó la necesidad de realizar una conversión a formato binario BAM para su adecuada manipulación. Esta conversión se llevó a cabo utilizando el paquete de herramientas SamTools (Danecek et al., 2021).

Adicionalmente, se realizó un proceso de marcado de duplicados, correspondiente a lecturas provenientes de un único fragmento de ADN. Cabe resaltar que estas duplicaciones también podían originarse a partir de un único

grupo de amplificación, el cual pueden ser erróneamente detectados como múltiples grupos por el sensor óptico del secuenciador (Broad Institute, 2019).

Finalmente, se efectuó una recalibración del puntaje de calidad base (BQSR). Las herramientas de GATK permitieron detectar errores sistemáticos derivados de fallas técnicas en el proceso de secuenciación o en la calidad de los reactivos. El procedimiento de BQSR ajustó la calidad base mediante un algoritmo de aprendizaje automático, con el objetivo de mejorar la precisión de las llamadas de variantes en etapas posteriores del análisis.

5.6.4 Llamado de variantes

El siguiente paso consistió en realizar la llamada de variantes. Para ello, se utilizó GATK4 (van der Auwera & O'Connor, 2020), el cual se enfocó en generar archivos en formato de llamada genómica de variantes (GVCF, por sus siglas en inglés). Este enfoque fue especialmente relevante debido a su capacidad para manejar datos a gran escala, facilitando así un genotipado conjunto eficiente.

En primer lugar, se empleó la herramienta HaplotypeCaller (Poplin et al., 2017), la cual permitió identificar simultáneamente variantes germinales de nucleótido único (SNVs) y pequeñas inserciones o deleciones (Indels), mediante un ensamblaje local basado en haplotipos. El resultado de esta etapa fue un archivo GVCF que proporcionó información detallada sobre las coordenadas genómicas secuenciadas.

Dado que se trabajó con múltiples genomas, se recomendó utilizar la herramienta CombineGVCFs, con el objetivo de fusionar varios archivos GVCF individuales en uno solo. Esta estrategia simplificó significativamente el análisis conjunto de los datos de variantes.

5.6.5 Filtrado y preparación para el llamado de variantes

Se realizó un filtrado de variantes debido a que el archivo VCF inicial no representaba completamente variantes reales, por lo que fue necesario identificar y eliminar posibles falsos positivos. Para ello, se empleó la Recalibración del Puntaje de Calidad de Variantes (VQSR), un procedimiento que permitió generar un archivo VCF filtrado, conteniendo únicamente variantes de alta calidad.

Una vez obtenido el archivo filtrado, se llevó a cabo un proceso de anotación con identificadores RS utilizando la herramienta del bcftools *annotate* (H. Li, 2011) y la base de datos dbSNP (versión dbsnp138). Este paso generó un archivo VCF comprimido e indexado, listo para ser utilizado en el proceso de asociación genética.

5.6.6 GWAS

Posteriormente, se realizó un análisis de asociación genética utilizando PLINK (Purcell et al., 2007). Para ello, el archivo VCF anotado fue convertido a un formato compatible con PLINK (generación de los archivos .bed, .bim y .fam), incluyendo la opción `--allow-extra-chr` para evitar errores relacionados con cromosomas no estándar. A continuación, se generó el archivo de fenotipos .fam, clasificando las muestras en casos y controles según el formato requerido por PLINK.

Como paso previo al análisis GWAS, se aplicaron filtros estrictos para mejorar la calidad de los datos: se eliminaron variantes con más del 5% de datos faltantes (`--geno 0.05`) y se evaluaron los individuos con más del 10% de genotipos no determinados (`--mind 0.1`), aunque en este caso ninguna muestra fue excluida.

El análisis de asociación genética se ejecutó con el siguiente comando:

```
plink --bfile filtered_data_maf_45 --assoc -out  
analysis_results_assoc_45 --allow-extra-chr --allow-no-sex
```

Este comando aplica una prueba estadística caso-control basada en el test de Chi-cuadrado para cada variante genética. Como todos los individuos fueron registrados con sexo ambiguo, PLINK generó un archivo adicional (`analysis_results_assoc_45.nosex`) para documentar esta información. El proceso finalizó correctamente, y los resultados se almacenaron en el archivo `analysis_results_assoc_45.assoc`, que contiene los valores estadísticos detallados por variante.

Posteriormente, se llevó a cabo un análisis preliminar y visualización de los resultados en RStudio. El script empleado filtró inicialmente los datos, eliminando filas con valores de p faltantes (NA) y conservando solo aquellos entre 0 y 1. Se excluyeron los cromosomas sexuales (X, Y y MT), así como cualquier identificador cromosómico no numérico, centrándose únicamente en los cromosomas autosómicos (1 al 22).

El análisis permitió identificar SNPs altamente significativos ($p \leq 5 \times 10^{-8}$), con fuerte evidencia de asociación, así como SNPs con valores sugestivos de asociación ($1 \times 10^{-5} > p > 5 \times 10^{-8}$), potencialmente relevantes para estudios futuros. Además, se extrajeron los identificadores rs vinculados con el DCL para facilitar análisis posteriores.

Finalmente, se generaron visualizaciones exploratorias en R usando los paquetes `ggplot2` y `qqman`, enfocadas en la distribución genómica de los valores de p. Se incluyó principalmente un Manhattan plot (M. H. Wang et al., 2019), diseñado para identificar regiones con señales de asociación relevantes y facilitar la priorización de variantes candidatas.

5.6.7 Anotación funcional con VEP

Se utilizó el listado de SNPs significativos $P \leq 5 \times 10^{-8}$ y sugerentes ($1 \times 10^{-5} > P > 5 \times 10^{-8}$), para llevar a cabo un análisis de anotación funcional con Variant Effect Predictor (VEP) (McLaren et al., 2016). Este análisis permitió recuperar información estructural y funcional de cada variante. Para ello, se utilizó la versión dbSNP138 como base de referencia y se ejecutó VEP sobre los identificadores RS previamente extraídos. Posteriormente, se descargaron los resultados en formato csv con las anotaciones correspondientes para las siguientes etapas de análisis.

5.6.8 Análisis de enriquecimiento génico

Con base en los resultados del análisis de anotación funcional, se extrajo una lista de genes asociados a los SNPs anotados. Esta lista se introdujo en la plataforma STRING (Szklarczyk et al., 2023), con el objetivo de obtener información funcional y estructural de las proteínas codificadas por los genes implicados. Adicionalmente, se empleó la plataforma g:Profiler (Raudvere et al., 2019) para realizar un análisis complementario de enriquecimiento funcional, enfocado en categorías como funciones moleculares, procesos biológicos y motivos reguladores.

5.6.9 Comparación de frecuencias alélicas con metaanálisis de referencia

Con el objetivo de identificar coincidencias o discrepancias en las variantes de riesgo para Alzheimer, se compararon las frecuencias alélicas obtenidas con los resultados del metaanálisis GWAS de (Bellenguez et al., 2022) Bellenguez et al. (2022). Para este análisis, se tomó el archivo VCF sin filtrado adicional y se ejecutó nuevamente un análisis de asociación. Posteriormente, se utilizó un archivo de texto (*rsids_list.txt*) que contenía los identificadores RS reportados por Bellenguez et al. para filtrar los SNPs presentes en los resultados del análisis local. Este filtrado se realizó con un comando *awk*, generando un archivo con las variantes compartidas.

Finalmente, se procesaron los datos en RStudio para evaluar la distribución de las diferencias de frecuencia alélica entre ambas poblaciones.

5.6.10 Análisis de ancestría

Se utilizó el proyecto del Proyecto 1000 Genomas, eliminando SNPs e individuos con alta proporción de datos faltantes y filtrando por frecuencia alélica. Luego, identifica SNPs comunes entre los datos propios (cohorte DCL) y los de referencia, y se sincronizaron las posiciones genómicas y corrigiendo posibles

discrepancias en la orientación de los alelos (strand flips) (Belsare et al., 2019; Marees et al., 2018). Posteriormente, se eliminaron SNPs conflictivos y se fusionan ambas bases de datos. Finalmente, se realiza un *pruning* para remover SNPs en desequilibrio de ligamiento y se lleva a cabo un análisis de escalamiento multidimensional (MDS) (L. Liu et al., 2013), que permite visualizar la relación genética entre individuos y detectar agrupamientos por ancestría.

6.Resultados y Discusión

6.1 Ensamblaje de Genomas.

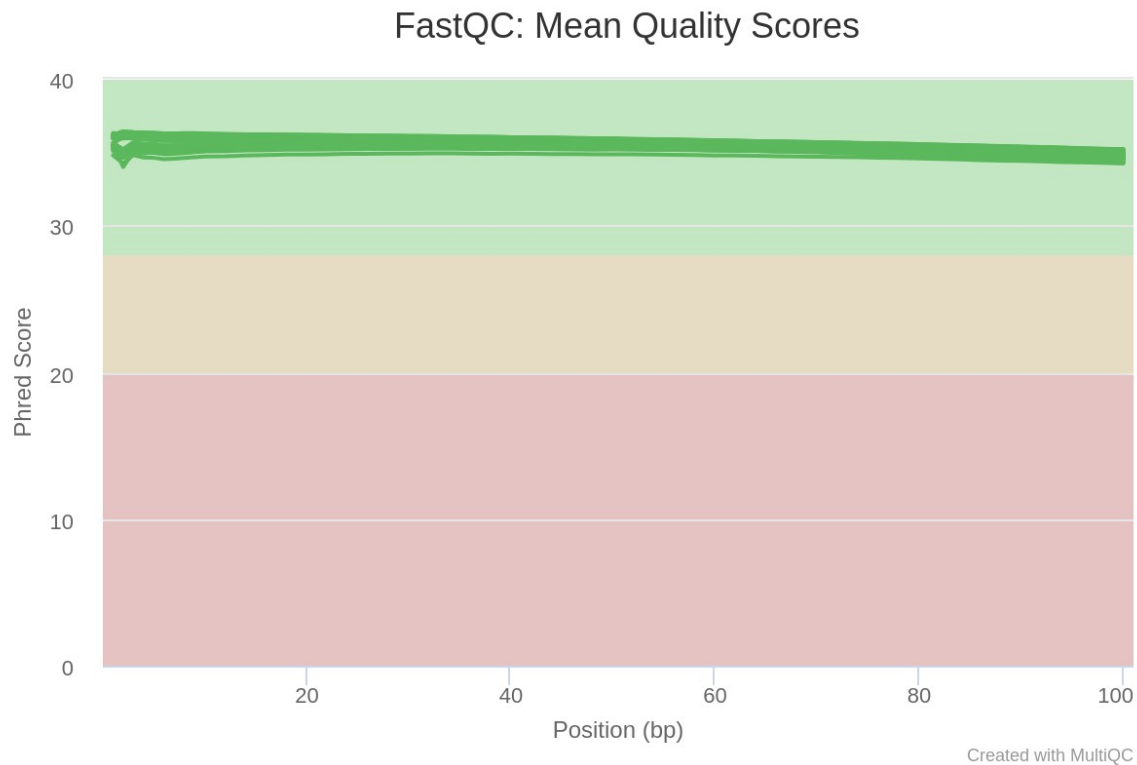
Se obtuvieron los respectivos archivos *forward* y *reverse* de la secuenciación de la empresa comercial, obteniendo 90 archivos FASTQ, dos para cada paciente analizado.

6.1.1 Control de calidad

Para realizar el control de calidad, se diseñó un script en Python denominado *fastqc.py*, que permite procesar múltiples archivos simultáneamente y generar reportes de calidad individuales mediante la herramienta FastQC. En este caso particular, se tomó una muestra aleatoria de 40 archivos FASTQ sobre los cuales se ejecutó dicho script. Posteriormente, se utilizó la herramienta MultiQC para obtener una vista global e integrada de los resultados de calidad.

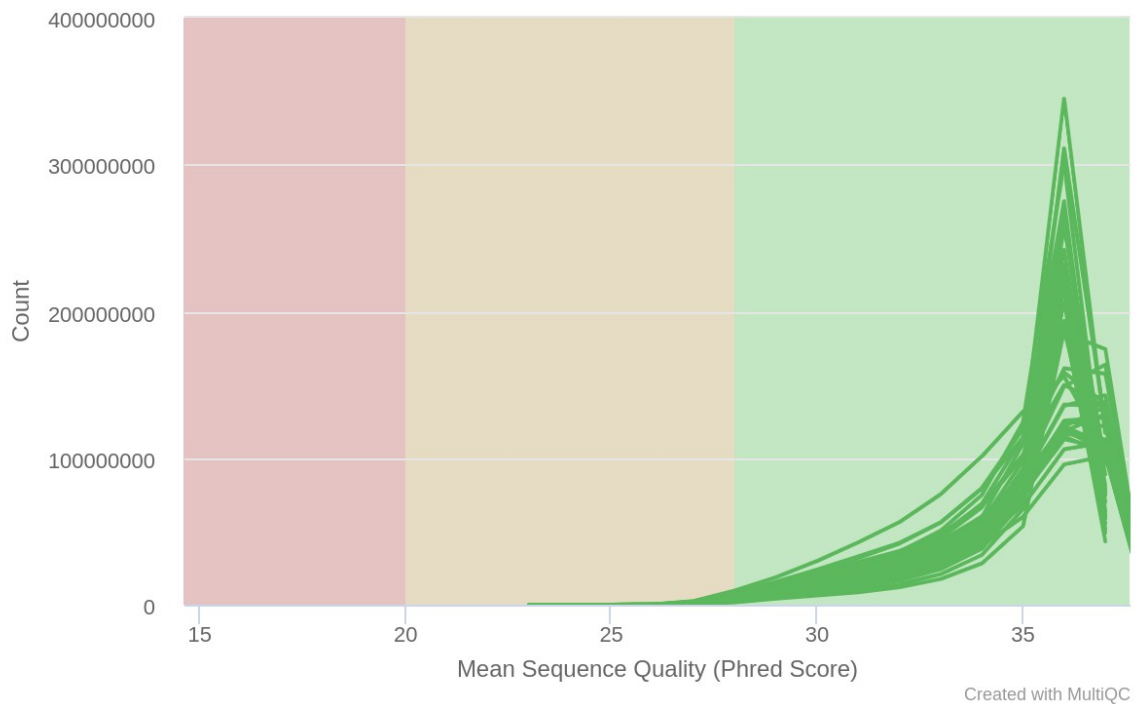
Los archivos FASTQ analizados tenían un promedio de 534.41 lecturas por archivo, con una longitud de 100 pb, un contenido promedio de GC del 40.75% y una tasa promedio de duplicación del 15.96%. La media de calidad por posición de cada base mostró puntuaciones Phred buenas, ubicando las secuencias crudas en la región verde, indicativa de buena calidad (ver Figura 1).

Figura 1: Calidad de secuencia por base



Para la media de calidad por lectura, también se evidenció que las secuencias presentaban una calidad con valores Phred Score superiores a 25 (ver Figura 2).

Figura 2: Puntuaciones de calidad por secuencia.



En cuanto al contenido de N bases, se presentó una presencia menor al 1%. De igual forma, los niveles de duplicación se mantuvieron en umbrales que el programa de control de calidad calificó como apropiados. Finalmente, las 40 muestras presentaban menos del 1% de secuencias sobre-representadas.

Teniendo en cuenta que las lecturas tenían una calidad óptima, no se realizó ningún proceso de corte de secuencias (trimming) y se inició el procesamiento de los datos.

6.1.2 Ensamblaje de Genomas

Se llevó a cabo el alineamiento de secuencias al genoma de referencia mediante el script *alignment.py*, desarrollado en Python, el cual automatiza la generación de archivos BAM. Este script utiliza la herramienta BWA (Burrows-Wheeler Aligner) para realizar el alineamiento de secuencias FASTQ mediante el comando `bwa mem`, utilizando 40 hilos de procesamiento en paralelo. Posteriormente, los archivos resultantes en formato SAM se convierten al formato BAM con Samtools, optimizando nuevamente el rendimiento al emplear 40 hilos adicionales. Este flujo de trabajo mejoró significativamente la eficiencia del alineamiento y la conversión de datos, facilitando así los análisis bioinformáticos posteriores. Todos los scripts mencionados pueden consultarse en el repositorio de GitHub correspondiente.

6.1.3 Preparación de los archivos BAM

El proceso de alineamiento y preprocesamiento de secuencias se realizó utilizando la plataforma MGI, cuyo formato de encabezado en los archivos FASTQ presenta diferencias con otros estándares. La asignación de Read Groups se llevó a cabo mediante el script *read_group.py*, asegurando la correcta identificación de las lecturas en los archivos BAM. Posteriormente, los archivos BAM fueron ordenados con el script *sort.py*, organizándolos según sus coordenadas genómicas. Para minimizar el sesgo en el análisis, se aplicó el marcado de duplicados con *mark_duplicates.py*, generando métricas sobre las lecturas redundantes. Además, se consolidaron estos datos en un archivo CSV mediante *evaluated_duplicate.py* para facilitar su análisis. Finalmente, se llevó a cabo la recalibración de calidad de bases en dos pasos: primero, con BaseRecalibrator de GATK, utilizando *recalibrate_part1.py*, y luego aplicando BQSR con *recalibrate_part2.py*, obteniendo archivos BAM calibrados y listos para el llamado de variantes. Todos estos pasos fueron automatizados mediante scripts en Python, optimizando el procesamiento y análisis de datos genómicos.

6.1.4 Llamado de variantes

Se generó un archivo VCF de alta calidad mediante un flujo de trabajo automatizado, diseñado para ser utilizado en estudios genómicos, incluyendo GWAS, análisis de ancestría y comparación de SNPs entre casos y controles. Para ello, se implementaron múltiples procesos optimizados, entre ellos el llamado de variantes con GATK HaplotypeCaller, herramienta que genera archivos VCF comprimidos a partir de lecturas alineadas en formato BAM. Posteriormente, estos archivos individuales fueron consolidados en un único archivo mediante GATK CombineGVCFs, facilitando así el análisis conjunto de variantes en cohortes.

El archivo combinado se sometió a un proceso de genotipado con GATK GenotypeGVCFs, seguido por una recalibración de la calidad de las variantes utilizando bases de datos de referencia (hapmap, 1000G, dbSNP), considerando métricas clave como QD, MQRankSum, FS, SOR y DP. Luego, se aplicaron filtros rigurosos mediante GATK VariantFiltration para eliminar variantes de baja calidad, basándose en criterios como profundidad de lectura, calidad del mapeo y sesgos en la posición de las lecturas.

Dado que el archivo VCF inicialmente generado no contenía identificadores de variantes (RS), se llevó a cabo un proceso de anotación para incorporarlos. Para ello, se utilizó la base de datos dbSNP (versión dbsnp138) como referencia, y la asignación de los identificadores se realizó mediante la herramienta bcftools annotate. Este paso permitió asociar los IDs correspondientes a cada variante presente en el archivo updated_45_snps.vcf.gz.

Finalmente, se obtuvo el archivo final comprimido e indexado. El conjunto de datos generado contiene 45 muestras con un total de 14,830,256 registros, todos ellos correspondientes a SNPs (sin MNPs ni indels). Dentro de estos registros, se identificaron 190,154 sitios multialélicos, todos ellos SNPs. El índice general de transiciones/transversiones (Ts/Tv) fue de 2.08, aumentando a 2.10 al considerar únicamente el primer alelo alternativo. Adicionalmente, se identificaron 5,059,568 SNPs tipo singleton (con un único alelo alternativo en la población), de los cuales 3,429,526 corresponden a transiciones y 1,630,042 a transversiones.

6.2 Caracterización de Rasgos Genómicos Relevantes

6.2.1 GWAS

Se aplicó un filtro para excluir variantes con una frecuencia alélica menor al 5% (MAF < 0.05), lo que redujo el número total de variantes de 14.830.256 a

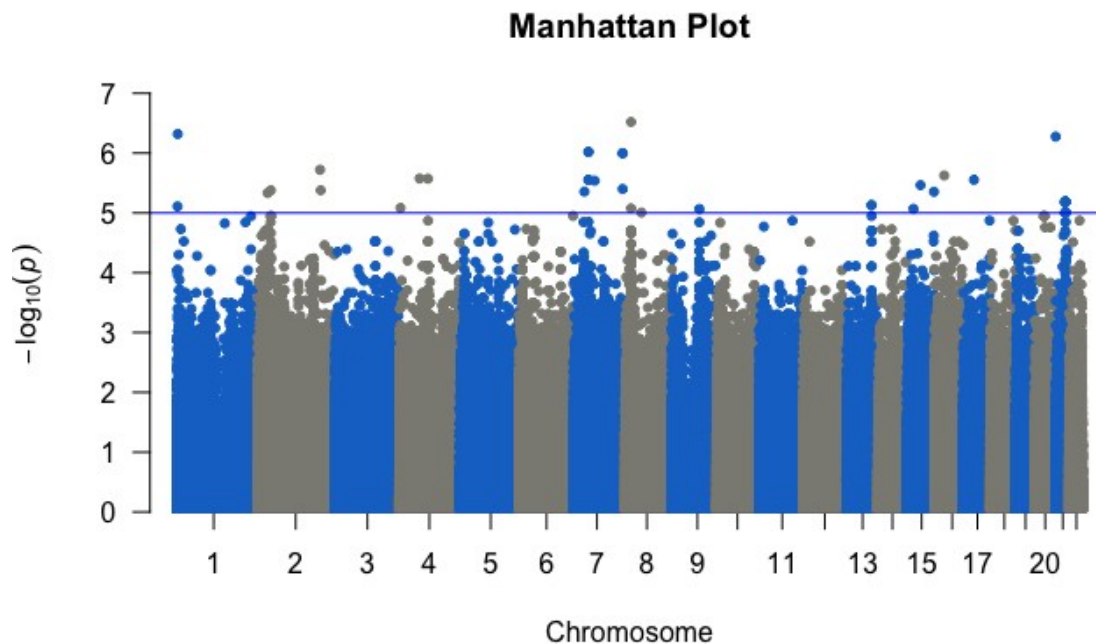
6.377.687, asegurando así un conjunto de datos más robusto y confiable para los análisis posteriores.

Se ejecuto el comando utilizando la herramienta PLINK, durante el procesamiento, el sistema detectó un total de 515.623 MB de RAM, de los cuales se reservaron 257.811 MB para la ejecución del análisis. Se cargaron exitosamente las 6.377.687 variantes desde el archivo .bim y los 45 individuos desde el archivo .fam. Los fenotipos fueron correctamente asignados, con una distribución de 26 casos y 19 controles, confirmando que el estudio se enmarca en un diseño de tipo caso-control.

En cuanto a la calidad general de los datos, la tasa de genotipado fue de 0.999661, lo que indica una cobertura prácticamente completa. Tras aplicar todos los filtros de calidad, se conservaron la totalidad de las variantes y de los individuos, validando la integridad del conjunto de datos utilizado en el análisis final.

Como una primera visualización, se generó un Manhattan plot en RStudio (ver Figura 3). En esta gráfica no se observan SNPs que superen el umbral de significancia genómica estricta ($p \leq 5 \times 10^{-8}$, es decir $-\log_{10}(p) \geq 7.3$); sin embargo, se identifican varios SNPs que cruzan el umbral sugestivo de asociación ($1 \times 10^{-5} > p > 5 \times 10^{-8}$), principalmente en los cromosomas 1, 6, 7, 9, 15 y 20, lo que sugiere posibles regiones candidatas para estudios posteriores.

Figura 3: Distribución de asociaciones genéticas obtenidas por GWAS en pacientes con DCL del Atlántico.



Nota. Manhattan plot generado a partir de un análisis de asociación genética tipo GWAS. En el eje x se representan los cromosomas del 1 al 22, organizados de forma alternada en colores azul y gris para facilitar la visualización. En el eje y se muestra el valor de asociación de cada SNP expresado como $-\log_{10}(p)$, lo que permite visualizar más fácilmente los SNPs con mayor significancia estadística. Cada punto representa un polimorfismo de nucleótido único (SNP), y su altura indica la fuerza de la asociación con el fenotipo estudiado. La línea horizontal azul marca el umbral de significancia sugestiva, generalmente fijado en $-\log_{10}(p) = 5$ (correspondiente a $p = 1 \times 10^{-5}$), lo que señala posibles regiones de interés genético.

Se identificaron un total de cuarenta y nueve (49) variantes sugestivas asociadas al deterioro cognitivo leve (DCL) (ver Tabla 3). Con el objetivo de explorar su relevancia funcional, se realizó una búsqueda detallada en bases de datos como NCBI dbSNP, GeneCard y UCSC Genome Browser para identificar genes cercanos a estas variantes. Dentro de los SNPs con odds ratios (OR) elevados, que sugieren mayor riesgo, destacan rs215631 en el cromosoma 7 (OR = 9.118, $p = 4.4 \times 10^{-6}$), ubicado cerca del gen *PDE1C*, y rs67090483 (OR = 19.44, $p = 2.89 \times 10^{-6}$), asociado a un transcrito no codificante (*ENSG00000287985*) relacionado con *NONHSAG047778.2* y *HSALNG0058455*, según GeneCaRNA. Asimismo, rs13048344 y varios SNPs cercanos en el cromosoma 21 (OR ≈ 9.263 , $p = 6.55 \times 10^{-6}$) se encuentran en la vecindad del gen *RUNX1*, ampliamente conocido por su rol en regulación transcripcional. Estos SNPs presentan

frecuencias alélicas significativamente mayores en los casos que en los controles, lo que refuerza su potencial implicación en la susceptibilidad al DCL.

En contraste, se identificaron SNPs con OR bajos que podrían ejercer un efecto protector. Por ejemplo, rs10159041 en el cromosoma 1 (OR = 0.0551, $p = 4.796 \times 10^{-7}$) se localiza próximo al gen *ATAD3A*, implicado en funciones mitocondriales, mientras que rs35559182 en el cromosoma 2 (OR = 0.03007) y rs11203917 en el cromosoma 8 (OR = 0.08089) se asocian a regiones cercanas a elementos no codificantes, como *RP11-156K13.1* sinónimo de *MTUS1-DT* un lncRNA.

Además, se observan agrupaciones de SNPs en regiones específicas, particularmente en el cromosoma 7 (como rs61740895, rs3735495, rs9791417, rs10234808) y en el cromosoma 21 (junto a rs13048344), lo que sugiere la existencia de regiones genómicas de interés para estudios funcionales y de validación adicionales.

Ahora bien, se identificaron varias variantes presentes en el GWAS Catalog con implicaciones potenciales en el DCL. Destacan los SNPs ubicados en la región 7q36.1: rs9640171, rs78225463 y rs60234513, todos mapeados a los genes *KCNH2* asociado a enfermedades del sistema nervioso central (Zhou et al., 2023) y *AOC1* asociado a pacientes con migraña, insomnio, déficit de atención con hiperactividad (TDAH) o fibromialgia (Fortes Marin et al., 2025). Aunque no presentan valores de OR disponibles, sus valores p (rs9640171: $p = 3.972 \times 10^{-6}$; rs78225463 y rs60234513: $p = 1.006 \times 10^{-6}$) los sitúan entre las variantes más sugestivas. Estas variantes se encuentran en regiones intergénicas o regulatorias, lo que sugiere un posible papel en la regulación de la expresión génica en loci con funciones celulares relevantes. Por otro lado, en la región 21q22.12, los SNPs rs2409620 (OR = 9.263; $p = 6.551 \times 10^{-6}$) y rs8133814 (OR = 8.365; $p = 9.934 \times 10^{-6}$), cercanos al gen *RUNX1*, presentan una elevada frecuencia en los casos frente a los controles. *RUNX1*, conocido principalmente por su rol en hematopoyesis, también se ha implicado en procesos de regulación génica que podrían estar involucrados en patologías neurodegenerativas (Fukui et al., 2018).

En cuanto a las variantes que no presentaban un identificador rs asociado, se realizó una búsqueda en UCSC Genome Browser para la posición chr15:58769435, encontrando que se localiza en las cercanías del transcrito *RP11-30K9.6*. Posteriormente, al consultar dicha posición en dbSNP, se determinó que corresponde a la variante rs1367939775, vinculada al gen *MINDY2*, miembro de una familia de nuevas desubiquitininas (DUBs) descritas recientemente (Abdul Rehman et al., 2021). Las enfermedades neurodegenerativas se relacionan con la disfunción de vías de degradación como

el sistema ubiquitina-proteasoma (UPS) y la autofagia-lisosoma (ALP), procesos en los que la ubiquitinación cumple un papel esencial (B. Liu et al., 2022). Aunque *MINDY2* se ha estudiado sobre todo en cáncer pancreático, se ha asociado con progresión tumoral e invasión neurológica, lo que sugiere un posible papel funcional relevante (P. Liu et al., 2023). Por lo anterior, se requiere más adelante investigar acerca de este tipo de familias desubiquitinadas relacionadas con las EN.

Finalmente, para la variante ubicada en chr21:5052044, aunque no se obtuvo información relevante en UCSC Genome Browser, se realizó una búsqueda en dbSNP, identificándose como la variante rs1245335245. Sin embargo, este SNP no se encuentra caracterizado y la información disponible sobre él es limitada.

Tabla 1: Variantes genéticas (SNPs) con asociación sugerente en pacientes con DCL.

CHR	SNP	BP	A1	F_A	F_U	A2	CHISQ	P	OR
1	rs13303330	933993	C	0.1154	0.5526	G	19.99	7,79E-03	0,1056
1	rs10159041	1518541	C	0.05769	0.5263	T	25.34	4,80E-04	0,0551
2	rs600015	31595231	G	0.25	0.7368	C	20.98	4,64E-03	0,119
2	rs35559182	41684300	A	0.01923	0.3947	T	21.18	4,18E-03	0,03007
2	rs12104475	194244627	C	0	0.3684	T	22.69	1,91E-03	0
2	rs1406217	196430887	G	0.01923	0.3947	A	21.18	4,18E-03	0,03007
4	rs112351502	3766269	C	0.1731	0.6316	A	19.88	8,26E-03	0,1221
4	rs115488222	63201518	T	0.2885	0.7895	C	22.04	2,66E-03	0,1081
4	rs368487632	89108585	*	0	0.3611	G	22.03	2,68E-03	0
7	rs215631	32325362	C	0.6731	0.1842	A	21.08	4.4e-06	9,118
7	rs61740895	45083124	A	0.1346	0.6316	G	24.01	9,56E-04	0,09074
7	rs3735495	45084866	T	0.1731	0.6579	A	21.95	2,79E-03	0,1088
7	rs9791417	45086176	A	0.1731	0.6579	G	21.95	2,79E-03	0,1088
7	rs10234808	45087994	T	0.1346	0.6316	C	24.01	9,56E-04	0,09074
7	rs67090483	64422033	G	0.5192	0.05263	A	21.89	2,89E-03	19,44
7	rs9640171	150903757	A	0.4231	0	G	21.28	3,97E-03	NA
7	rs78225463	150905601	G	0.4615	0	A	23.92	1,01E-03	NA
7	rs60234513	150907318	C	0.4615	0	T	23.92	1,01E-03	NA
8	rs4999541	17819952	G	0.1346	0.5789	A	19.85	8,39E-03	0,1131
8	rs11203917	17842156	G	0.1346	0.6579	C	26.24	3,02E-04	0,08089
8	rs28648923	49791805	G	0.6731	0.1944	*	19.53	9,88E-03	8,529
9	rs1187325	84670680	G	0.2885	0.7632	C	19.8	8,60E-03	0,1258
13	rs9561806	95195864	T	0.03846	0.4211	C	20.09	7,40E-03	0,055
13	rs12429872	95206031	T	0.03846	0.4211	C	20.09	7,40E-03	0,055
15	rs111540891	36856774	T	0.1538	0.6053	C	19.81	8,54E-03	0,1186
15	.	58769435	*	0.4423	0	A	21.56	3,43E-03	NA

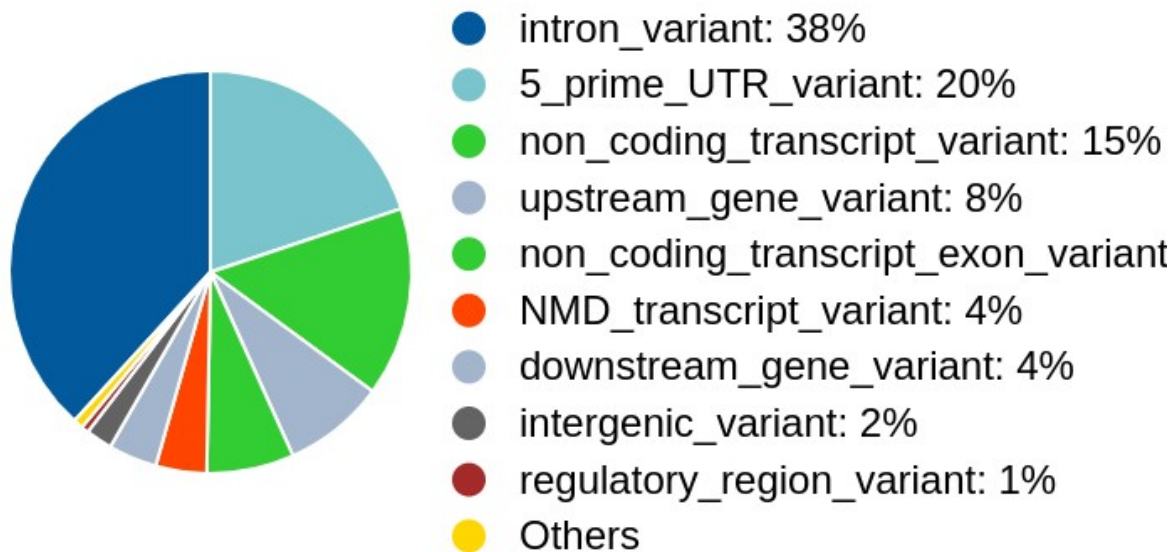
15	rs4328423	100535743	T	0.1	0.5556	C	21.07	4,44E-03	0,08889
16	rs369580269	30403289	G	0.09615	0.5526	T	22.27	2,37E-03	0,08612
17	rs2934233	32639265	G	0.1731	0.6579	A	21.95	2,79E-03	0,1088
21	.	5052044	A	0.01923	0.4474	T	25.15	5,30E-04	0,02422
21	rs13048344	35318094	C	0.6346	0.1579	A	20.32	6,55E-03	9,263
21	rs7277833	35318686	C	0.6346	0.1579	T	20.32	6,55E-03	9,263
21	rs2409620	35318891	C	0.6346	0.1579	T	20.32	6,55E-03	9,263
21	rs732569	35319155	A	0.6346	0.1579	G	20.32	6,55E-03	9,263
21	rs35640344	35319895	T	0.6346	0.1579	G	20.32	6,55E-03	9,263
21	rs35553297	35319905	C	0.6346	0.1579	G	20.32	6,55E-03	9,263
21	rs2834905	35320056	T	0.6346	0.1579	G	20.32	6,55E-03	9,263
21	rs8133814	35320152	C	0.6538	0.1842	G	19.52	9,93E-03	8,365
21	rs13052959	35320745	G	0.6538	0.1842	A	19.52	9,93E-03	8,365
21	rs2242717	35321492	G	0.6538	0.1842	A	19.52	9,93E-03	8,365
21	rs28460413	35321790	T	0.6538	0.1842	C	19.52	9,93E-03	8,365
21	rs13050284	35321894	A	0.6538	0.1842	G	19.52	9,93E-03	8,365
21	rs13050093	35321897	T	0.6538	0.1842	C	19.52	9,93E-03	8,365
21	rs2834907	35321948	C	0.6538	0.1842	T	19.52	9,93E-03	8,365
21	rs2834908	35322309	A	0.6538	0.1842	G	19.52	9,93E-03	8,365
21	rs1040313	35339820	C	0.6346	0.1579	T	20.32	6,55E-03	9,263
21	rs13051961	35361254	C	0.6346	0.1579	T	20.32	6,55E-03	9,263
21	r13051252	35362636	T	0.6346	0.1579	A	20.32	6,55E-03	9,263
21	rs62218300	35363416	C	0.6346	0.1579	T	20.32	6,55E-03	9,263

6.2.2 Análisis de Anotación de Genes Sugestivos

Se introdujo la lista de RS (Reference SNPs) en VEP (Variant Effect Predictor) con el objetivo de anotar e identificar sus funciones asociadas. Posteriormente, se descargaron los resultados del análisis realizado con VEP para su interpretación.

El análisis de variantes muestra que la mayoría de las alteraciones identificadas se encuentran en regiones no codificantes, con un 38% en intrones, lo que sugiere un posible impacto en el *splicing* o en la regulación de la expresión génica. Además, un 20% de las variantes están en la región 5' UTR, lo que puede influir en la estabilidad y traducción del ARNm, mientras que un 15% afectan transcritos no codificantes, pudiendo desempeñar funciones regulatorias. También se detectaron variantes en regiones upstream (8%) y downstream (4%) de los genes, las cuales podrían modificar sus niveles de expresión. Un 7% de las variantes afectan exones de transcritos no codificantes y un 4% están asociadas a transcritos sujetos a degradación por el mecanismo Nonsense-Mediated Decay (NMD). Las variantes intergénicas (2%) son menos frecuentes y su impacto funcional es incierto. Finalmente, solo un 1% de las variantes corresponden a cambios de sentido (*missense variants*), lo que sugiere una baja proporción de alteraciones que podrían modificar la estructura y función de las proteínas. Ver figura 4.

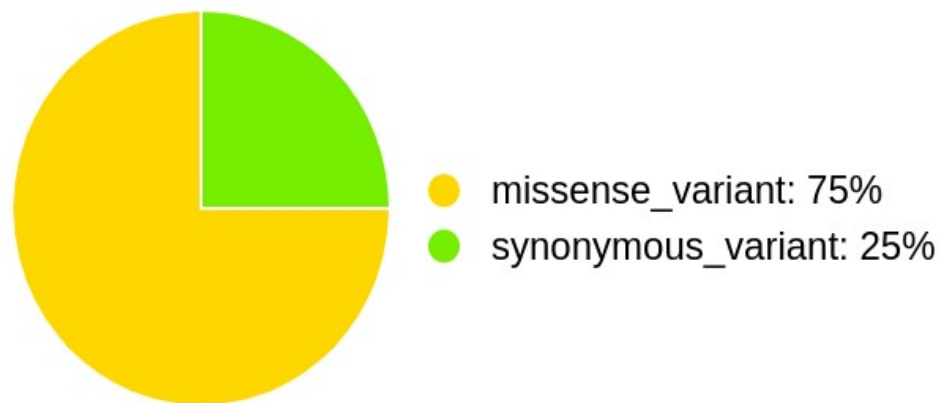
Figura 4: Variantes genéticas (SNPs) con asociación sugerente en pacientes con DCL.



En cuanto a las variantes codificantes, el análisis revela que el 75 % corresponde a variantes de cambio de sentido (missense variants), las cuales modifican la secuencia de aminoácidos de la proteína y, por ende, podrían afectar su estructura y función. En contraste, el 25 % restante está representado por variantes sinónimas (synonymous variants), que no alteran el aminoácido codificado, pero podrían influir en la eficiencia del splicing, la estabilidad del ARNm o la tasa de traducción (ver Figura 5).

Sin embargo, la variante missense rs61740895 fue identificada en el gen *NACAD* y presenta un impacto moderado sobre la estructura y función de la proteína codificada. Por otro lado, la variante sinónima rs3735495, también localizada en el gen *NACAD*, no altera la secuencia aminoacídica; sin embargo, podría afectar la eficiencia del splicing, la estabilidad del ARNm o la tasa de traducción.

Figura 5: Proporción de variantes missense y sinónimas detectadas en el análisis funcional con VEP



Ahora bien, se tomó la lista de genes obtenida a partir del análisis con VEP, identificándose un total de 15 genes asociados a los SNPs analizados (ver Figura 6). Esta lista fue introducida en dos plataformas para el análisis de enriquecimiento génico, una de ellas fue STRING, la cual permitió obtener una descripción detallada de cada uno de los genes implicados.

El análisis reveló un conjunto de genes con alteraciones potencialmente relevantes. Entre ellos, *RUNX1* fue el más frecuentemente afectado, con 18 variantes identificadas. Este gen codifica un factor de transcripción esencial para la regulación de la expresión génica y la diferenciación celular. Se ha sugerido que *RUNX1* podría actuar como un precursor en la neurogénesis en ratones, participando en etapas tempranas de la neurogénesis adulta (Fukui et al., 2018).

Otro gen destacado es *NACAD*, con cuatro variantes. Este gen está involucrado en la regulación del direccionamiento de proteínas hacia el retículo endoplasmático, y ha despertado interés debido a su posible implicación en una enfermedad llamada malformaciones cavernosas cerebrales (CCM) (Gallione et al., 2022).

Asimismo, se identificaron dos variantes en el gen *ABCC4*, que codifica un transportador de aniones orgánicos clave en la detoxificación celular. En el sistema nervioso central, los transportadores ABC (ATP-Binding Cassette) protegen el cerebro y la médula espinal al eliminar compuestos tóxicos, incluidos algunos fármacos (Dean et al., 2022).

Un hallazgo especialmente relevante fue la identificación de una variante en *NTRK2*, gen que codifica un receptor clave para el desarrollo del sistema nervioso. Variantes en este gen se han asociado con encefalopatía del desarrollo

y epiléptica (DEE), una condición rara caracterizada por epilepsia temprana y retraso neurológico (Yoganathan et al., 2021). Este receptor participa en múltiples procesos neuronales, incluyendo supervivencia, proliferación, migración, diferenciación y plasticidad sináptica, a través de su interacción con factores neurotróficos como *BDNF*, *NTF3* y *NTF4*.

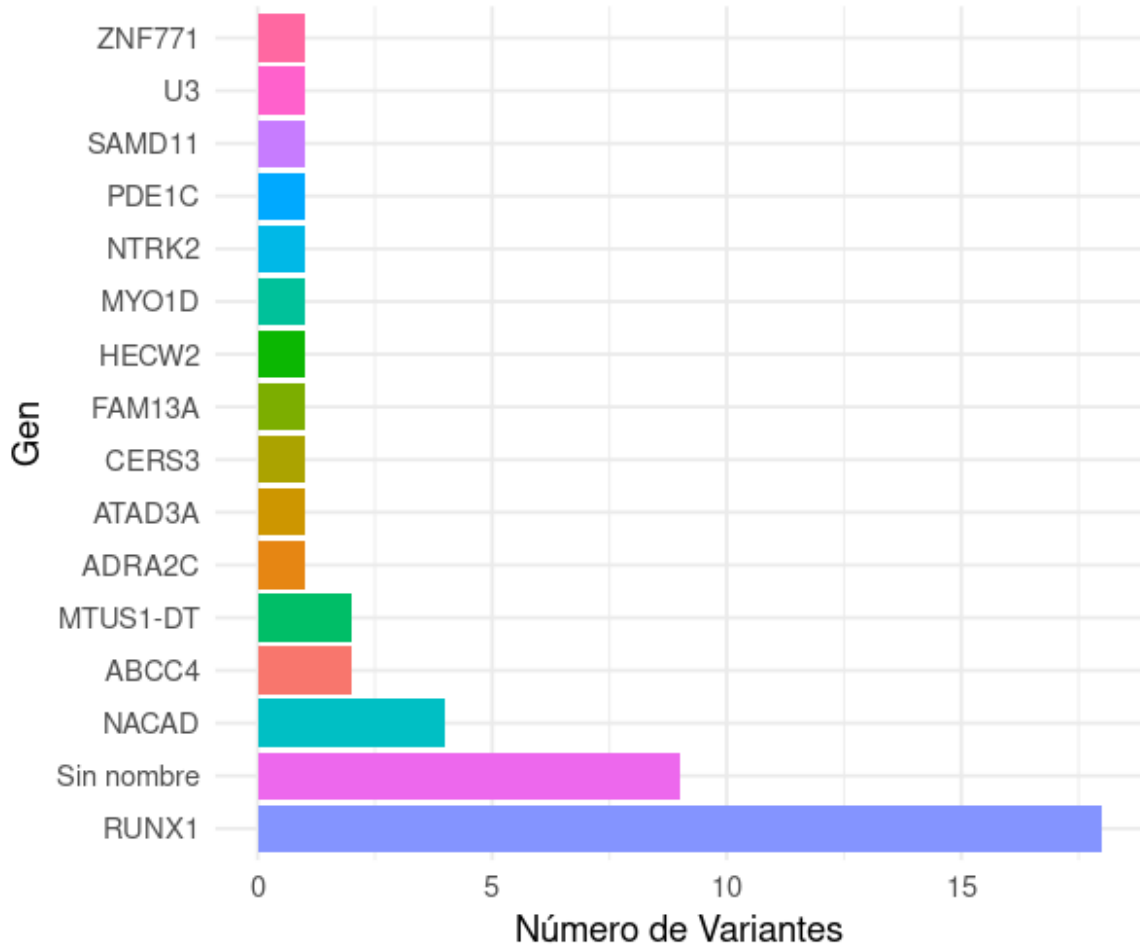
También se encontraron variantes en genes con funciones clave a nivel neuronal, como *ZNF771*, un factor de transcripción involucrado en la regulación epigenética. Este gen regula grupos de genes co-metilados en el cerebro, participando en procesos de diferenciación y plasticidad neuronal mediante su interacción con regiones sensibles a modificaciones como 5mC y 5hmC (Maulik et al., 2018). Asimismo, el gen *ATAD3A*, clave en la organización mitocondrial y el metabolismo celular, ha presentado variantes asociadas a Huntington. Su oligomerización excesiva podría causar fragmentación mitocondrial, disrupción del ADN mitocondrial y neurodegeneración. (Zhao et al., 2019).

Por su parte, el gen *CERS3* se asocia a la síntesis de ceramidas de cadena ultra larga, esenciales para la homeostasis lipídica. Las ceramidas no solo son componentes estructurales de las membranas celulares y mitocondriales, sino que también participan como mediadores bioactivos en procesos clave como proliferación, diferenciación, inflamación e inducción de la apoptosis (Delcheva et al., 2024). También el gen *ADRA2C*, que codifica un receptor adrenérgico $\alpha 2$, cumple un rol importante en la modulación del sistema nervioso simpático. Se ha relacionado con diversos trastornos, incluyendo enfermedad de Parkinson, trastornos de pánico y Huntington (Riess et al., 1994). Recientemente, se ha identificado como un gen clave (hub gene) dentro de redes asociadas a la neurotransmisión monoaminérgica, y se considera que podría tener un papel funcional importante en el trastorno por déficit de atención e hiperactividad (ADHD) (Yadav et al., 2021).

Finalmente, se identificaron variantes en *SAMD11*, gen que podría desempeñar un papel crucial en el desarrollo de los fotorreceptores. Estudios recientes han demostrado que *SAMD11* y *SAMD7* son componentes del complejo epigenético PRC1 en fotorreceptores, donde *SAMD7* actúa como regulador principal y *SAMD11* como un factor accesorio que contribuye a establecer la identidad específica de las células bastón (Kubo et al., 2021).

En conjunto, la variabilidad genética observada en estos genes sugiere que podrían estar implicados en procesos clave como la señalización celular, la regulación epigenética y transcripcional, el metabolismo energético y el neurodesarrollo.

Figura 6: Número de variantes asociadas al gen cercano con VEP.



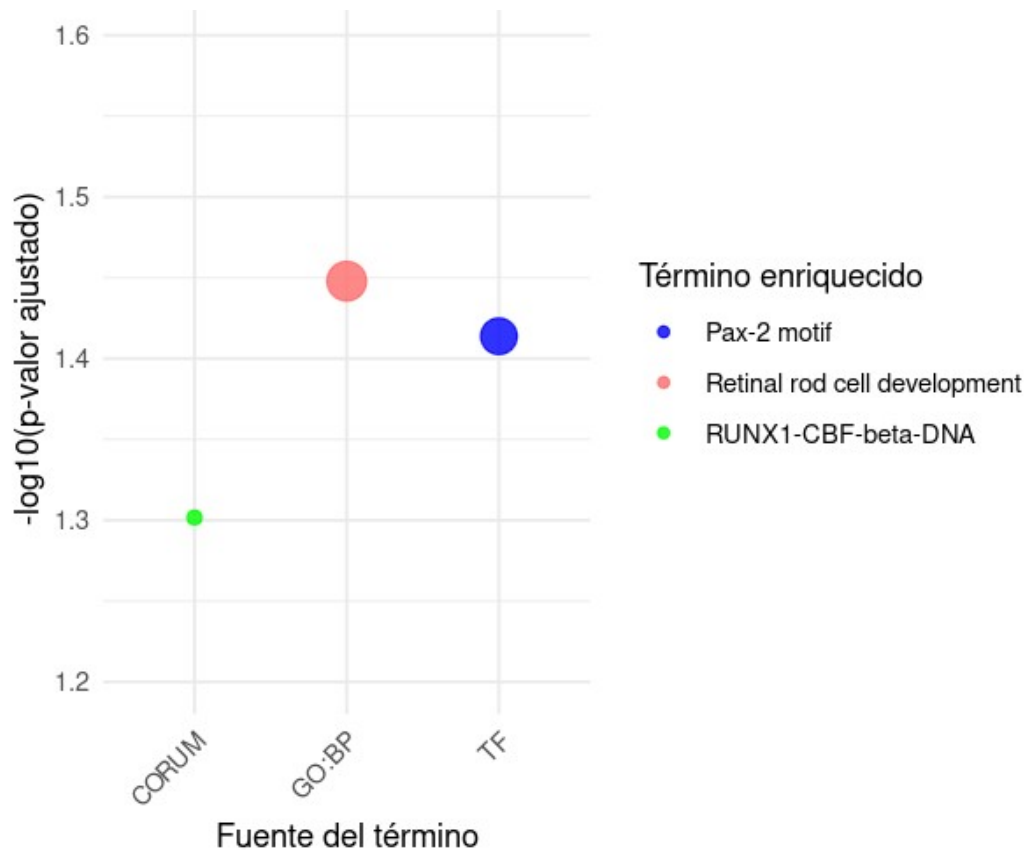
Con el fin de visualizar de manera más clara las posibles interacciones funcionales, se realizó un análisis de enriquecimiento funcional utilizando la plataforma g:Profiler, el cual arrojó únicamente tres términos con enriquecimiento no significativo (ver Figura 7).

Entre estos términos, destaca particularmente el factor de transcripción Pax-2 (motivo: NCGTCACGCNYSRNYGCNYYN, clase de coincidencia: 1), que, a pesar del reducido número de genes analizados, mostró intersección con los genes: *RUNX1*, *ABCC4*, *ADRA2C*, *ATAD3A*, *CERS3*, *HECW2*, *MYO1D*, *NTRK2*, *PDE1C*, *SAMD11* y *ZNF771*, alcanzando un valor de $-\log_{10}(p\text{-ajustado}) = 1.41$.

Específicamente, Pax2 regula el destino celular de los precursores GABAérgicos durante el desarrollo del cerebelo y la médula espinal. Su mutación o delección puede provocar alteraciones en la regulación transcripcional, cambios en los niveles de GABA en el sistema nervioso central, desequilibrio sináptico entre excitación e inhibición (E-I) y anomalías en los circuitos neuronales, eventos que

en conjunto podrían contribuir al desarrollo de trastornos del neurodesarrollo (Lv et al., 2021).

Figura 7: Enriquecimiento de términos en *g:Profiler*.



6.3 Genética Comparada de Enfermedades Neurodegenerativas

El estudio de Bellenguez et al. (2022) corresponde a un meta-análisis GWAS que integró múltiples cohortes para identificar variantes genéticas asociadas a la enfermedad de Alzheimer y otras demencias relacionadas (ADD). La cohorte principal fue la European Alzheimer & Dementia Biobank (EADB), con 20,464 casos clínicos y 22,244 controles de 15 países europeos. El análisis permitió identificar 75 loci de riesgo, de los cuales 42 eran nuevos y 9 correspondían a loci previamente reportados en estudios relacionados con los datos utilizados.

Con el fin de identificar variaciones poblacionales en las frecuencias alélicas de los SNPs asociados a la enfermedad de Alzheimer reportados por Bellenguez et al. (2022), en comparación con las observadas en la población del Atlántico, se llevó a cabo un procesamiento de datos en dos fases. En primer lugar, se utilizó el archivo genotípico sin filtros adicionales y se realizó una asociación por SNPs

mediante el comando ASSOC de PLINK, con el objetivo de obtener las frecuencias alélicas observadas:

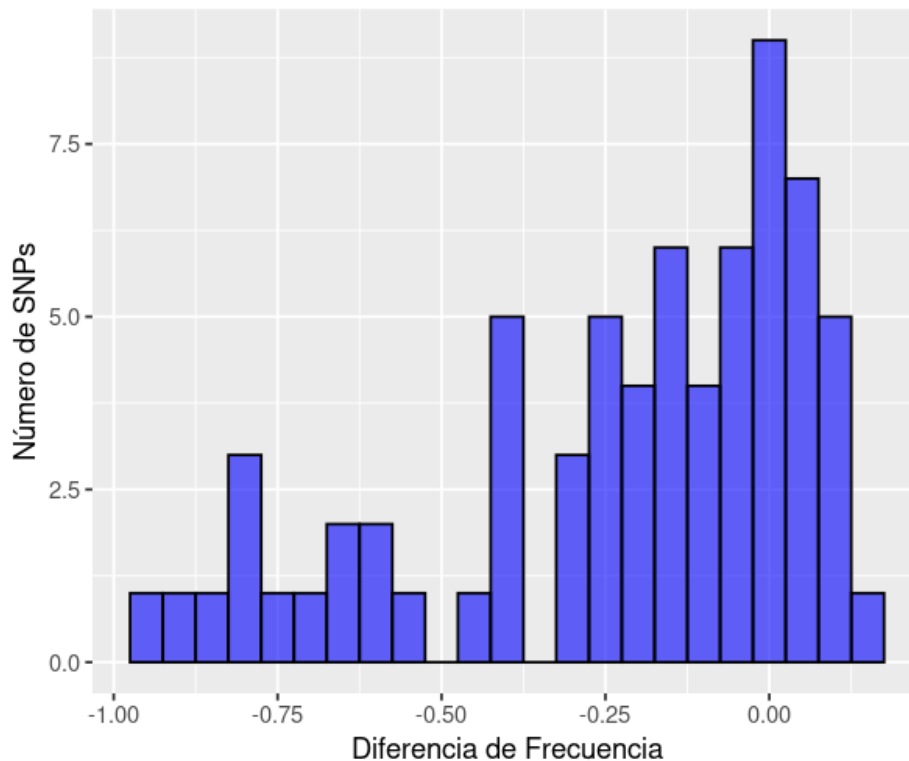
```
plink --bfile update_45 --assoc --out update_45_fa --allow-extra-chr --allow-no-sex
```

Posteriormente, se empleó una lista de identificadores **rsIDs** proveniente del estudio de Bellenguez (*rsids_list.txt*), la cual fue utilizada para filtrar los SNPs correspondientes en nuestros resultados:

```
awk 'NR==FNR {snps[$1]; next} ($2 in snps)' rsids_list.txt  
../../../../3_vcf_45/1_vcf_files/update_45_fa.assoc >  
filtered_assoc_45_corrected.txt
```

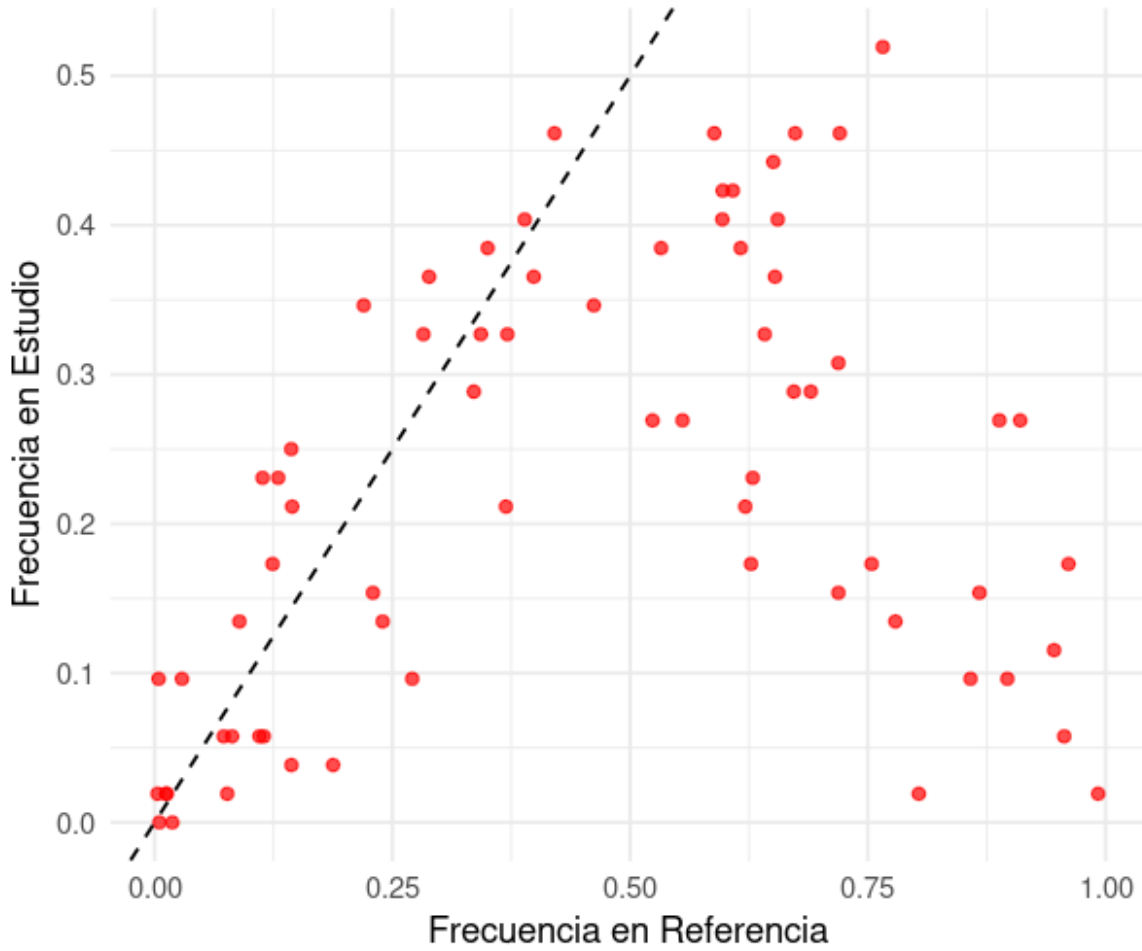
Finalmente, en RStudio se analizó la distribución de las diferencias de frecuencia alélica entre ambas poblaciones (referencia y muestra local). La Gráfica 8 muestra dicha distribución. Si la mayoría de las diferencias se sitúan cerca de 0, esto indica una alta similitud entre poblaciones. Sin embargo, se observó una distribución leptocúrtica con asimetría negativa (desviación hacia la derecha), lo cual sugiere la presencia de algunos SNPs con frecuencias significativamente diferentes respecto a la cohorte de referencia, lo que podría reflejar diferencias poblacionales relevantes.

Figura 8: Distribución de diferencias de frecuencias Alélicas.



Estas diferencias en las frecuencias alélicas también pueden observarse en el Gráfico de Dispersión de Frecuencias, correspondiente a la Figura 9. En este gráfico, si los puntos se alinean con la diagonal ($y = x$), indica que las frecuencias son similares entre ambas poblaciones. No obstante, se observa la presencia de varios puntos dispersos alejados de la diagonal, lo que evidencia la existencia de SNPs con diferencias marcadas en frecuencia alélica. Esta dispersión sugiere la posible presencia de señales de diferenciación poblacional, que podrían reflejar variabilidad genética entre la población del estudio y la cohorte de referencia utilizada por Bellenguez et al. (2022).

Figura 9: Comparación de frecuencias alélicas.



6.3.1 Análisis de ancestría

En el análisis de frecuencias alélicas se evidenciaron discrepancias importantes con respecto a los valores de referencia, lo que motivó la realización de un análisis de componentes principales (PCA, por sus siglas en inglés) como primer paso para explorar la estructura genética de la muestra.

Para el PCA se utilizaron los archivos de PLINK, con el objetivo de evaluar posibles agrupaciones o patrones de variabilidad genética entre los individuos. Para ello, se emplearon los archivos genotípicos `update_45.bed`, `update_45.bim` y `update_45.fam`, a partir de los cuales se calcularon los 10 primeros componentes principales. Este análisis generó los archivos `pca_update_45.eigenval` (valores propios) y `pca_update_45.eigenvec` (coordenadas de los individuos en el espacio de los componentes principales).

Posteriormente, se cuantificó la varianza explicada por cada componente mediante la normalización de los valores propios respecto a la varianza total, lo que permitió identificar el porcentaje de variabilidad genética capturada por cada uno de los componentes principales.

```
plink --bfile update_45 --pca 10 --out pca_update_45
```

```
awk '{sum+=$1} END {for (i=1; i<=NR; i++) print "PC" i, $i/sum*100"%"}'  
pca_update_45.eigenval
```

```
awk '{sum+=$1} END {print sum}' pca_update_45.eigenval
```

```
total=$(awk '{sum+=$1} END {print sum}' pca_update_45.eigenval)
```

```
awk -v total="$total" '{printf "PC%d %.2f%%\n", NR, ($1/total)*100}'  
pca_update_45.eigenval
```

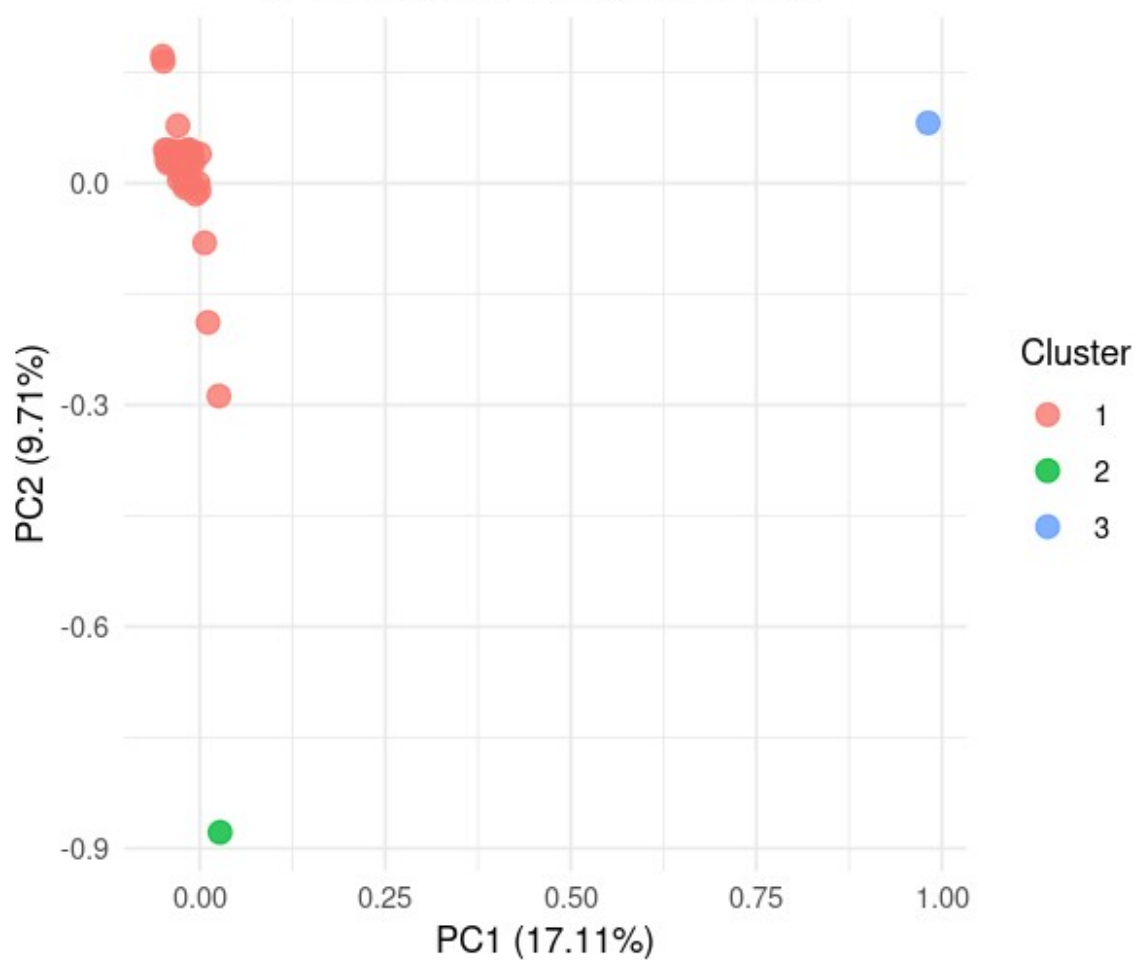
Luego, en RStudio, se ejecutó un script que realiza un análisis de componentes principales (PCA) complementado con un algoritmo de clustering K-means, con el objetivo de identificar posibles agrupaciones dentro de la muestra genética analizada.

En primer lugar, se cargó el archivo `pca_update_45.eigenvec`, que contiene los valores correspondientes a los 10 primeros componentes principales obtenidos previamente con PLINK, y se renombraron las columnas para facilitar su manipulación. Posteriormente, se implementó un modelo de clustering K-means estableciendo tres grupos ($K = 3$), utilizando las dos primeras componentes principales (PC1 y PC2), capturando la mayor proporción de la variabilidad genética de la muestra.

La asignación de cada individuo a uno de los tres clusters se incorporó como una nueva columna al dataframe y se generó una visualización de dispersión

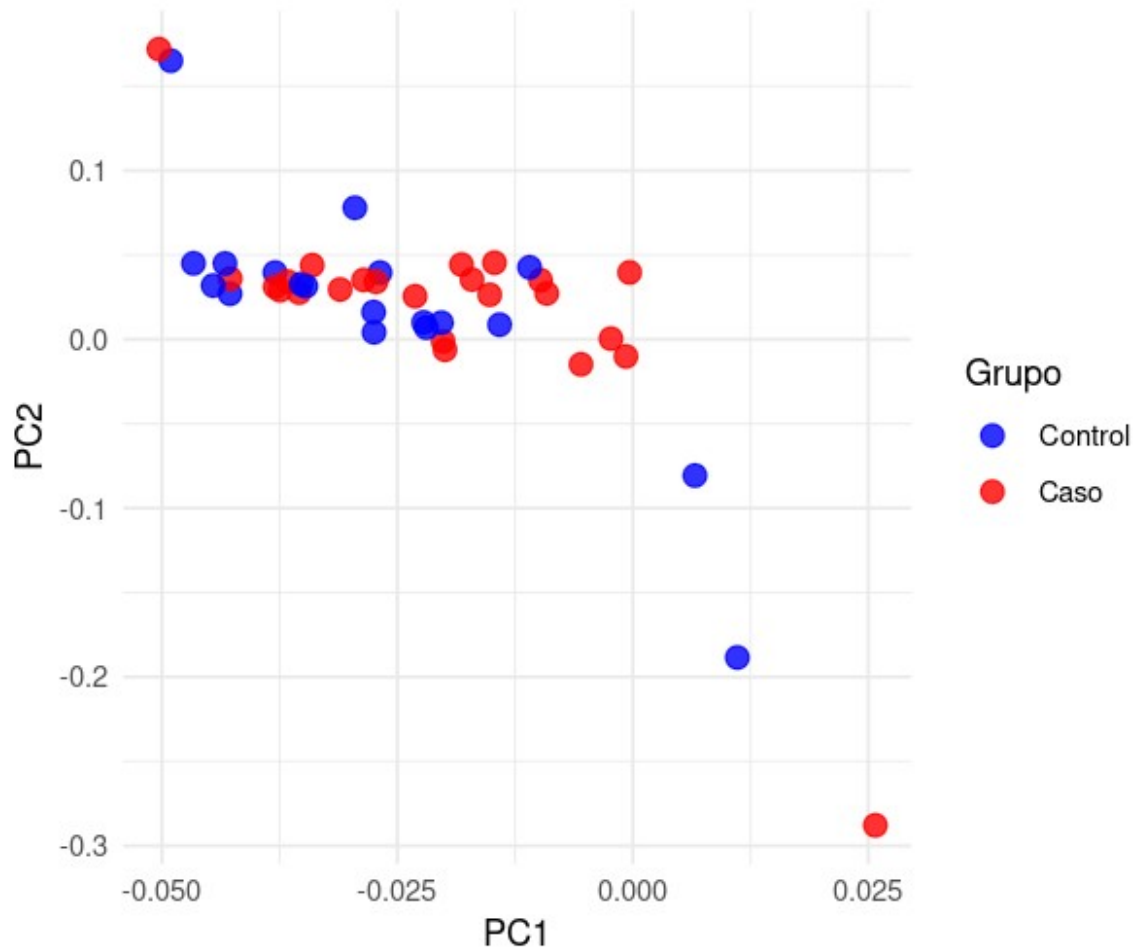
empleando el paquete ggplot2 (Wickham, 2016) , en la cual los individuos son representados en función de PC1 y PC2, y coloreados según el grupo asignado por el algoritmo de clustering.

Figura 10: PCA con Clustering K-means.



En la Figura 10, se observa que la mayoría de los individuos (Cluster 1, en rojo) están agrupados cerca del origen, lo que sugiere una alta similitud genética dentro de esta población. Sin embargo, hay dos individuos claramente diferenciados: uno en el Cluster 2 (verde) correspondiente al caso 21 y otro en el Cluster 3 (azul) correspondiente al caso 82 , los cuales presentan una separación considerable en los ejes PC1 y PC2, lo que podría indicar diferencias significativas en la estructura genética o posibles outliers.

Figura 11: PCA diferenciando casos y controles.



Con el fin de visualizar la distribución de casos y controles después de aplicar el agrupamiento no supervisado mediante el algoritmo K-means con $K = 3$ clusters, se utilizaron como entrada las dos primeras componentes principales (PC1 y PC2), que capturan la mayor parte de la variabilidad genética. Los grupos resultantes se incorporaron al dataframe principal, permitiendo asignar a cada individuo su respectivo clúster.

Posteriormente, se excluyeron dos individuos específicos identificados como posibles outliers (IID 21 y 82), con el objetivo de mejorar la interpretación gráfica de los patrones de agrupamiento. A partir de estos datos filtrados, se generó un gráfico de dispersión utilizando el paquete ggplot2, en el cual los individuos fueron representados según sus coordenadas en PC1 y PC2, y coloreados de azul para los controles y rojo para los casos.

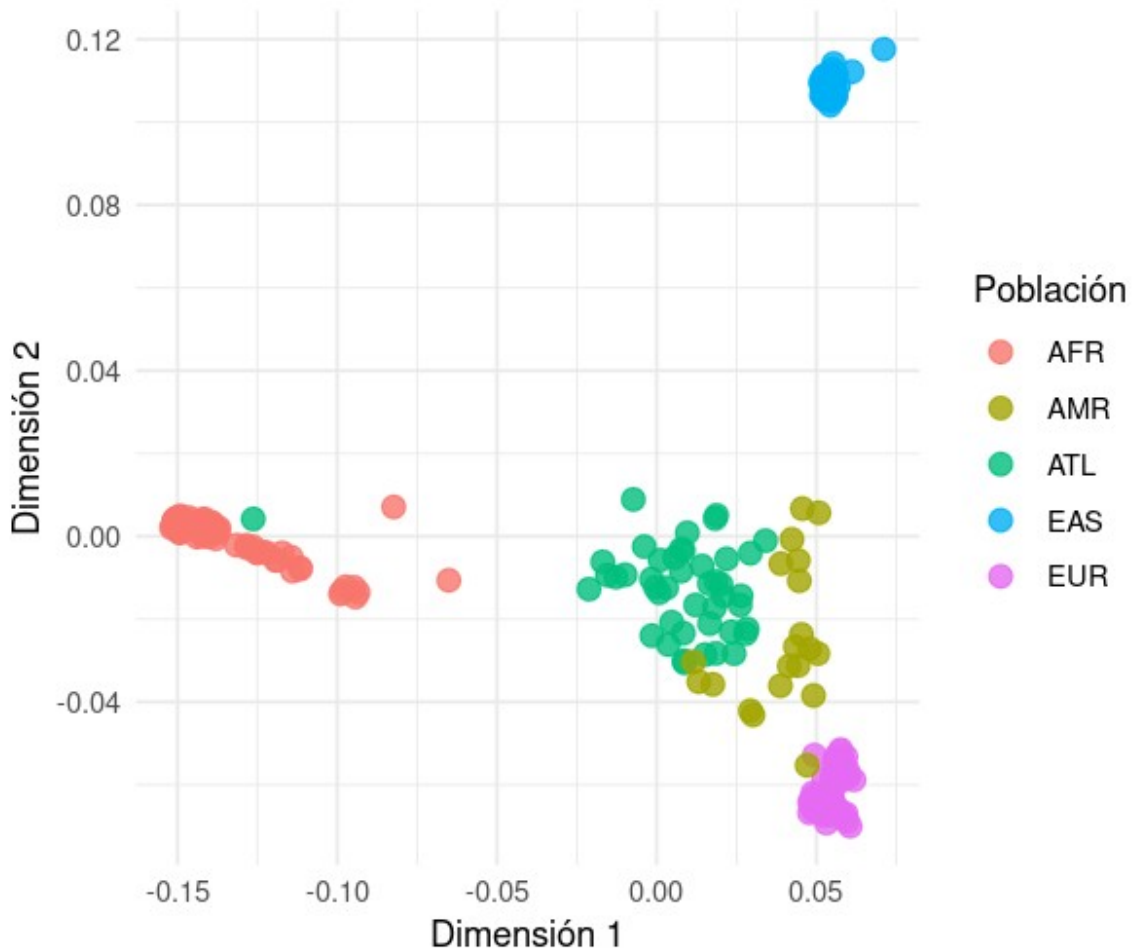
En la Figura 11 se puede observar que, incluso tras la exclusión de los outliers, se mantiene un patrón claro de agrupamiento entre las muestras, lo cual refuerza la presencia de una estructura genética diferenciada entre casos y controles.

Dado que se evidenciaron agrupamientos y la presencia de posibles outliers en la muestra, se procedió a realizar un análisis de ancestría utilizando como referencia los datos del Proyecto de los 1000 Genomas. Este análisis comenzó con un proceso de control de calidad (QC) sobre los datos de referencia, en el cual se eliminaron SNPs e individuos con alta proporción de datos faltantes, y se aplicaron filtros por frecuencia alélica para asegurar la robustez de las comparaciones.

Posteriormente, se identificaron los SNPs comunes entre los datos propios (cohorte DCL) y los del 1000 Genomas. A continuación, se sincronizaron las posiciones genómicas y se corrigieron posibles discrepancias en la orientación de los alelos (strand flips), lo cual es esencial para evitar errores en la fase de fusión de datos. Tras la eliminación de SNPs conflictivos, se procedió a la fusión de ambas bases de datos.

Como paso final antes del análisis de ancestría, se aplicó un pruning para eliminar SNPs en desequilibrio de ligamiento, con el fin de reducir la redundancia y el ruido en los datos. Finalmente, se realizó un análisis de escalamiento multidimensional (MDS), que permitió visualizar la relación genética entre los individuos y detectar patrones de agrupamiento según su ancestría. Este enfoque es crucial para controlar la estratificación poblacional en estudios de asociación genética, garantizando así la validez estadística de los resultados obtenidos.

Figura 12: Gráfico de Ancestría en la población de DCL.



En la Figura 12, correspondiente al análisis de ancestría realizado con los 45 individuos de la muestra, se observa que la población del Atlántico (*representada en color turquesa*) muestra una mayor proximidad genética con los marcadores de ancestría americana, así como una relación significativa con la ascendencia europea. No obstante, se identifica un individuo cuya posición en el espacio de componentes sugiere una mayor cercanía con la población africana, lo que indica una posible contribución genética de dicha ancestría en ese caso particular.

6.4 Documentación en repositorio Github

Finalmente, los scripts mencionados y la carpeta en Drive con las tablas correspondientes pueden consultarse en el repositorio respectivo en GitHub. Allí se documentan las diferentes etapas del proyecto, así como la disponibilidad de los scripts utilizados a lo largo del análisis.

7. Conclusiones y recomendaciones

7.1 Conclusiones

En cuanto al primer objetivo, se obtuvo la secuencia de los genomas completos de pacientes con DCL y controles de la población del departamento del Atlántico. La tecnología de MGI demostró un excelente desempeño en la generación de librerías y archivos FASTQ, alcanzando Phred Scores superiores a 25 (ver Figura 2), lo que evidencia una alta calidad en las lecturas.

En cuanto a la calidad general del genotipado, se observó una tasa de éxito del 99.9661%, lo que indica una cobertura prácticamente completa. Este resultado concuerda con estudios previos que reportan que las plataformas de MGI presentan una mayor tasa de concordancia en el genotipado de SNPs y en la llamada de variantes, en comparación con otras plataformas comerciales (Kim et al., 2021).

En lo que concierne al segundo objetivo, hemos de destacar que este estudio es uno de los primeros en realizar análisis de genoma completo en pacientes con DCL en una población subrepresentada como la colombiana, lo que permite identificar variantes locales potencialmente novedosas. No obstante, enfrenta limitaciones como la corrección por múltiples pruebas ($p < 5 \times 10^{-8}$), que reduce la detección de variantes de pequeño efecto y limita la explicación de la heredabilidad. Además, muchas señales se localizan en regiones no codificantes, dificultando su interpretación funcional y requiriendo integración con datos ómicos. Aunque los GWAS han sido clave en estos avances, el modelo omnigénico sugiere que múltiples genes pueden influir indirectamente en los rasgos complejos (Tam et al., 2019).

El análisis permitió identificar 49 variantes genéticas sugestivas asociadas al DCL, algunas con odds ratios elevados como rs215631 (*PDE1C*) y rs67090483, lo que sugiere una posible implicación en el riesgo de la enfermedad. Particularmente, variantes cercanas a *RUNX1* en el cromosoma 21, como rs13048344 y rs2409620, refuerzan su rol potencial en la regulación transcripcional implicada en procesos neurodegenerativos (Fukui et al., 2018). Por otro lado, SNPs como rs10159041 (*ATAD3A*) y rs35559182 (*MTUS1-DT*) mostraron posibles efectos protectores. Además, se detectaron variantes en regiones intergénicas reguladoras, como las ubicadas en 7q36.1 (*KCNH2* y *AOC1*), previamente reportadas en el GWAS Catalog (Fortes Marin et al., 2025; Zhou et al., 2023). Un hallazgo novedoso fue la variante rs1367939775 vinculada a *MINDY2*, gen poco explorado pero relacionado con vías de degradación proteica implicadas en enfermedades neurodegenerativas (Abdul Rehman et al., 2021; B. Liu et al., 2022).

Ahora bien, los análisis de anotación con la plataforma VEP se enfocaron en ciertos genes de relevancia como la identificación de 18 SNPs intrónicos en el gen *RUNX1* con valores de p entre 6.5×10^{-6} y 9.9×10^{-6} , y odds ratios elevados (8.3–9.3), lo que sugiere un impacto funcional relevante en el DCL. Particularmente *RUNX1* participa en hematopoyesis e inflamación, y su desregulación, junto a *RUNXOR*, *MAP2*, *NGF*, *BDNF* e *IL-10*, ha sido asociada a esclerosis múltiple, indicando un posible rol en neurodegeneración e inmunorregulación (Haridy et al., 2023).

Además, se identificaron dos variantes intrónicas en *ABCC4* (rs9561806 y rs12429872) con valores de p = 7.404×10^{-6} y OR = 0.055, lo que sugiere un posible efecto protector en DCL. *ABCC4* es el único transportador de su familia con función protectora demostrada frente a compuestos tóxicos cerebrales como los análogos de camptotecina (Lin et al., 2013). Junto con *ABCC1* y *ABCC2*, podría tener roles específicos en contextos patológicos (Baltira et al., 2024), y se

ha asociado al riesgo de Alzheimer tardío (LOAD), como en el caso de *ABCA7*-rs3764650-C (Carrasquillo et al., 2015).

De igual forma, la variante rs1187325 en la región 5' UTR del gen *NTRK2* ($p = 8.601 \times 10^{-6}$; OR = 0.1258), con posible efecto protector. *NTRK2* codifica el receptor TRKB, activado por BDNF, proteína clave para la supervivencia neuronal, la plasticidad sináptica, la LTP y la protección frente a la apoptosis (Pradhan et al., 2019). Estudios de variantes en *NTRK2* mostraron asociaciones significativas en haplotipos de 2 ($p = 0.012$) y 3 loci ($p = 0.009$), involucrando SNPs en regiones asociadas a splicing alternativo y regulación transcripcional. Estas combinaciones podrían afectar la señalización neurotrófica mediada por el receptor TRKB (Chen et al., 2008). Además, se ha reportado mutaciones en el gen que pueden interrumpir esta señalización neurotrófica esencial, lo cual se ha asociado con la aparición de trastornos neurológicos graves desde etapas tempranas del desarrollo (Yoganathan et al., 2021).

La variante rs369580269 en la región upstream de *ZNF771* ($p = 2.369 \times 10^{-6}$; OR = 0.0861) sugiere un posible efecto protector. *ZNF771*, como otros genes ZNF, regula la expresión génica y funciones neuronales. Se ha vinculado esta familia con enfermedades neurodegenerativas, incluyendo *GLIS3*, *ZFX* y *ZC3H14*, asociados a la acumulación y modificación de proteína tau y procesamiento de ARN (Bu et al., 2021; Calderari et al., 2018; S. Xu et al., 2016). Al participar en módulos epigenéticos vinculados a la diferenciación neuronal y sinapsis, podría ser un gen clave en enfermedades como Alzheimer, Parkinson o esquizofrenia. (Bu et al., 2021). Además, la interacción significativa de la variante rs8112679 en el exón 4 del gen *ZNF222*, enriquecida en poblaciones africanas ($p = 0.01$), refuerza el papel potencial de los genes ZNF en la susceptibilidad a enfermedades neurodegenerativas (Granot-HersHKovitz et al., 2023).

Ahora bien, se identificó la variante intrónica rs10159041 en el gen *ATAD3A*, con una asociación sugestiva ($p = 4.796 \times 10^{-7}$; OR = 0.0551), lo que sugiere un posible efecto protector frente al DCL. Estudios han demostrado que la oligomerización de *ATAD3A* en modelos de Alzheimer induce la acumulación de

colesterol en las membranas asociadas al retículo endoplásmico mitocondrial (MAMs) al inhibir la expresión de *CYP46A1*, una enzima clave en el metabolismo del colesterol cerebral, contribuyendo así a la pérdida sináptica y al deterioro cognitivo característicos de la enfermedad (Zhao et al., 2022) . De igual forma se identificó una variante en el gen *CERS3*, rs4328423 ($p = 4.437 \times 10^{-6}$, OR = 0.08889). específicamente *CERS3* codifica una enzima clave en la síntesis de esfingolípidos, cuya desregulación se ha asociado con enfermedades neurodegenerativas como el Parkinson y el Alzheimer. (McInnis et al., 2024). Por último, la variante rs112351502, ubicada en la región upstream del gen *ADRA2C*, mostró una asociación sugestiva ($p = 8.264 \times 10^{-6}$) con un odds ratio de 0.1221. Estudios han reportado que la expresión del ARNm de *ADRA2C* en el hipocampo está significativamente reducida en sujetos con enfermedad de Alzheimer y demencia con cuerpos de Lewy, lo que indica una alteración específica de este receptor en dichas patologías (Szot et al., 2006).

El análisis de g:Profiler destacó la anotación de *PAX2*, cuya implicación en la neurogénesis embrionaria y adulta. Específicamente las mutaciones en *PAX2* pueden alterar redes regulatorias esenciales para la diferenciación y mantenimiento neuronal, contribuyendo al deterioro de circuitos cerebrales. Así, *PAX2* emerge como un nodo clave cuya disfunción puede vincularse tanto con enfermedades del desarrollo como con patologías neurodegenerativas (Lv et al., 2021). Recientemente en modelos murinos la eliminación específica de Pax2 en el sistema nervioso de ratones provoca conductas repetitivas restringidas y déficits en la flexibilidad cognitiva, asociadas a alteraciones en los circuitos neuronales del hipocampo y una reducción de IGFBP2 (Insulin-like Growth Factor Binding Protein 2) (Y. Wang et al., 2024).

En conclusión, los resultados resaltan regiones genómicas clave para futuros estudios funcionales y subrayan la importancia de investigar variantes poco caracterizadas en poblaciones subrepresentadas como la colombiana. No obstante, deben interpretarse con cautela, ya que los GWAS con muestras pequeñas, como en este caso, tienden a detectar solo variantes comunes o de gran efecto, limitando la comprensión integral de la arquitectura genética de la

enfermedad (Tam et al., 2019). También es importante considerar la complejidad del modelo biológico. En plantas como *Arabidopsis thaliana*, algunos rasgos están controlados por pocos loci de gran efecto, lo que permite detectarlos con muestras pequeñas. En cambio, en humanos, los rasgos complejos suelen depender de muchos loci de efecto pequeño, por lo que se requieren decenas de miles de individuos para alcanzar significancia genómica, como ocurre en estudios de depresión mayor, donde se necesitan más de 50,000 casos y controles. (Korte & Farlow, 2013).

Finalmente, en relación con el objetivo de analizar variantes asociadas a enfermedades neurodegenerativas detectadas en Europa en la población del Atlántico, se compararon las frecuencias alélicas de los loci reportados en la cohorte europea con nuestra muestra local. Este análisis permitió identificar diferencias que podrían tener implicaciones clínicas y funcionales, ya que algunas variantes comunes en una población pueden ser raras en otra, reflejando posibles diferencias en la estructura genética como se vio en nuestro análisis de ancestría (ver Figura 12).

De manera similar, estudios en poblaciones latinoamericanas, como el realizado en 35 individuos mexicanos con Alzheimer, han evidenciado variaciones alélicas significativas frente a bases de datos de referencia como GnomAD, lo que destaca la necesidad de caracterizar genéticamente a poblaciones subrepresentadas (Genis-Mendoza et al., 2020). Estas diferencias alélicas pueden señalar variantes que merecen un análisis funcional más profundo debido a su posible especificidad poblacional. No obstante, es importante reiterar que el tamaño de muestra representa una limitación clave en los resultados de este estudio.

7.2 Recomendaciones

Ampliar el tamaño muestral en estudios futuros y mejorar la capacidad de detección de variantes de pequeño efecto asociadas al DCL, se recomienda expandir la cohorte analizada. Estudios GWAS con mayor número de individuos permiten identificar con mayor robustez asociaciones significativas a nivel genómico (Tam et al., 2019; Korte & Farlow, 2013).

Realizar estudios funcionales de las variantes sugestivas, como las cercanas a *RUNX1*, *NTRK2*, *ATAD3A*, *ZNF771* y *ABCC4*, mostraron efectos potenciales sobre la regulación génica, la neuroprotección o la neurodegeneración. Se sugiere realizar estudios funcionales (in vitro o in vivo) para validar su implicación biológica.

Integrar enfoques multi-ómicos dado que muchas variantes se encuentran en regiones no codificantes o regulatorias, es esencial integrar transcriptómica, epigenómica y proteómica para una interpretación funcional más completa y para validar el efecto de las variantes sobre rutas moleculares específicas.

Algunas variantes identificadas mostraron odds ratios bajos, lo que sugiere un potencial efecto protector frente al DCL (por ejemplo, en *ABCC4*, *CERS3*, *ZNF771*, *ADRA2C*). Estas podrían ser prioritarias para el desarrollo de biomarcadores o futuras estrategias terapéuticas.

Bibliografía

- Aarsland, D., Batzu, L., Halliday, G. M., Geurtsen, G. J., Ballard, C., Ray Chaudhuri, K., & Weintraub, D. (2021). Parkinson disease-associated cognitive impairment. *Nature Reviews Disease Primers*, 7(1). <https://doi.org/10.1038/s41572-021-00280-3>
- Abdul Rehman, S. A., Armstrong, L. A., Lange, S. M., Kristariyanto, Y. A., Gräwert, T. W., Knebel, A., Svergun, D. I., & Kulathu, Y. (2021). Mechanism of activation and regulation of deubiquitinase activity in MINDY1 and MINDY2. *Molecular Cell*, 81(20), 4176-4190.e6. <https://doi.org/10.1016/j.molcel.2021.08.024>
- Adam, F., & Nathan, W. (2020). Best Practices for De Novo Transcriptome Assembly with Trinity. *Harvard FAS Informatics*. <https://informatics.fas.harvard.edu/best-practices-for-de-novo-transcriptome-assembly-with-trinity.html>
- Alvarado, C., Gómez, J. F., Etayo, E., Giraldo, C. E., Pineda, A., & Toro, E. (2014). Estudio EDECO (Estudio poblacional de deterioro cognitivo en población colombiana). *Acta Médica Colombiana*, 264–271. <https://doi.org/10.36104/amc.2014.196>
- Anderson, N. D. (2019). State of the science on mild cognitive impairment (MCI). In *CNS Spectrums* (Vol. 24, Issue 1). <https://doi.org/10.1017/S1092852918001347>
- Andrews, S. (2010). FastQC. *Babraham Bioinformatics*. <http://www.bioinformatics.babraham.ac.uk/projects/>
- Angela M. Sanford. (2017). Mild Cognitive Impairment. *Clinics in Geriatric Medicine*, 33(3), 325–337. <http://dx.doi.org/10.1016/j.cger.2017.02.005>
- Bai, W., Chen, P., Cai, H., Zhang, Q., Su, Z., Cheung, T., Jackson, T., Sha, S., & Xiang, Y. T. (2022). Worldwide prevalence of mild cognitive impairment among community dwellers aged 50 years and older: a meta-analysis and systematic review of epidemiology studies. *Age and Ageing*, 51(8). <https://doi.org/10.1093/ageing/afac173>
- Baltira, C., Aronica, E., Elmquist, W. F., Langer, O., Löscher, W., Sarkaria, J. N., Wesseling, P., de Gooijer, M. C., & van Tellingen, O. (2024). The impact of ATP-binding cassette transporters in the diseased brain: Context matters. *Cell Reports Medicine*, 5(6), 101609. <https://doi.org/10.1016/j.xcrm.2024.101609>
- Bellenguez, C., Küçükali, F., Jansen, I. E., Kleinedam, L., Moreno-Grau, S., Amin, N., Naj, A. C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., Holmans, P. A.,

- Boland, A., Damotte, V., van der Lee, S. J., Costa, M. R., Kuulasmaa, T., Yang, Q., de Rojas, I., Bis, J. C., ... Lambert, J.-C. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature Genetics*, *54*(4), 412–436. <https://doi.org/10.1038/s41588-022-01024-z>
- Belsare, S., Levy-Sakin, M., Mostovoy, Y., Durinck, S., Chaudhuri, S., Xiao, M., Peterson, A. S., Kwok, P.-Y., Seshagiri, S., & Wall, J. D. (2019). Evaluating the quality of the 1000 genomes project data. *BMC Genomics*, *20*(1), 620. <https://doi.org/10.1186/s12864-019-5957-x>
- Benavides-Caro, C. . (2017). Deterioro cognitivo en el adulto mayor. *Revista Mexicana de Anestesiología*, *40*(2).
- Blennow, K., de Leon, M. J., & Zetterberg, H. (2006). Alzheimer's disease. *The Lancet*, *368*(9533), 387–403. [https://doi.org/10.1016/S0140-6736\(06\)69113-7](https://doi.org/10.1016/S0140-6736(06)69113-7)
- Borrás Blasco, C., & Viña Ribes, J. (2016). Neurofisiología y envejecimiento. Concepto y bases fisiopatológicas del deterioro cognitivo. *Revista Espanola de Geriatria y Gerontologia*, *51*. [https://doi.org/10.1016/S0211-139X\(16\)30136-6](https://doi.org/10.1016/S0211-139X(16)30136-6)
- Brabec, J. L., Lara, M. K., Tyler, A. L., & Mahoney, J. M. (2021). System-Level Analysis of Alzheimer's Disease Prioritizes Candidate Genes for Neurodegeneration. *Frontiers in Genetics*, *12*. <https://doi.org/10.3389/fgene.2021.625246>
- Broad Institute. (2019). Picard toolkit. In *Broad Institute, GitHub repository*.
- Bu, S., Lv, Y., Liu, Y., Qiao, S., & Wang, H. (2021). Zinc Finger Proteins in Neuro-Related Diseases Progression. *Frontiers in Neuroscience*, *15*. <https://doi.org/10.3389/fnins.2021.760567>
- Calderari, S., Ria, M., Gérard, C., Nogueira, T. C., Villate, O., Collins, S. C., Neil, H., Gervasi, N., Hue, C., Suarez-Zamorano, N., Prado, C., Cnop, M., Bihoreau, M.-T., Kaisaki, P. J., Cazier, J.-B., Julier, C., Lathrop, M., Werner, M., Eizirik, D. L., & Gauguier, D. (2018). Molecular genetics of the transcription factor GLIS3 identifies its dual function in beta cells and neurons. *Genomics*, *110*(2), 98–111. <https://doi.org/10.1016/j.ygeno.2017.09.001>
- Campbell, N. L., Unverzagt, F., LaMantia, M. A., Khan, B. A., & Boustani, M. A. (2013). Risk factors for the progression of mild cognitive impairment to dementia. In *Clinics in Geriatric Medicine* (Vol. 29, Issue 4). <https://doi.org/10.1016/j.cger.2013.07.009>
- Carrasquillo, M. M., Crook, J. E., Pedraza, O., Thomas, C. S., Pankratz, V. S., Allen, M., Nguyen, T., Malphrus, K. G., Ma, L., Bisceglia, G. D., Roberts, R. O., Lucas, J. A., Smith, G. E., Ivnik, R. J., Machulda, M. M., Graff-Radford, N. R., Petersen, R. C., Younkin, S. G., & Ertekin-Taner, N. (2015). Late-onset Alzheimer's risk variants in memory decline, incident mild cognitive impairment, and Alzheimer's disease. *Neurobiology of Aging*, *36*(1), 60–67. <https://doi.org/10.1016/j.neurobiolaging.2014.07.042>

- Chen, Z., Simmons, M. S., Perry, R. T., Wiener, H. W., Harrell, L. E., & Go, R. C. P. (2008). Genetic Association of Neurotrophic Tyrosine Kinase Receptor Type 2 (NTRK2) With Alzheimer's Disease. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *147B*(3), 363–369. <https://doi.org/10.1002/ajmg.b.30607>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). <https://doi.org/10.1093/gigascience/giab008>
- de Mendonça, A., Ribeiro, F., Guerreiro, M., & Garcia, C. (2004). Frontotemporal mild cognitive impairment. *Journal of Alzheimer's Disease*, *6*(1), 1–9. <https://doi.org/10.3233/JAD-2004-6101>
- Dean, M., Moitra, K., & Allikmets, R. (2022). The human ATP-binding cassette (ABC) transporter superfamily. *Human Mutation*, *43*(9), 1162–1182. <https://doi.org/10.1002/humu.24418>
- del Carmen Díaz-Mardomingo, M., García-Herranz, S., Rodríguez-Fernández, R., Venero, C., & Peraíta, H. (2017). Problems in classifying mild cognitive impairment (MCI): One or multiple syndromes? *Brain Sciences*, *7*(9). <https://doi.org/10.3390/brainsci7090111>
- Delcheva, G., Stefanova, K., & Stankova, T. (2024). Ceramides—Emerging Biomarkers of Lipotoxicity in Obesity, Diabetes, Cardiovascular Diseases, and Inflammation. *Diseases*, *12*(9), 195. <https://doi.org/10.3390/diseases12090195>
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. <https://doi.org/10.1038/ng.806>
- Díaz Cabezas, R., Marulanda Mejía, F., & Martínez Arias, M. H. (2013). Prevalencia de deterioro cognitivo y demencia en mayores de 65 años en una población urbana colombiana. *Acta Neurológica Colombiana*, *29*(3). <https://actaneurologica.com/index.php/anc/article/view/1376>
- Duff, K., Paulsen, J., Mills, J., Beglinger, L. J., Moser, D. J., Smith, M. M., Langbehn, D., Stout, J., Queller, S., & Harrington, D. L. (2010). Mild cognitive impairment in prediagnosed Huntington disease. *Neurology*, *75*(6). <https://doi.org/10.1212/WNL.0b013e3181eccfa2>
- Elman, J. A., Panizzon, M. S., Logue, M. W., Gillespie, N. A., Neale, M. C., Reynolds, C. A., Gustavson, D. E., Rana, B. K., Andreassen, O. A., Dale, A. M., Franz, C. E., Lyons, M. J., & Kremen, W. S. (2019). Genetic risk for coronary heart disease alters the influence of Alzheimer's genetic risk on mild cognitive impairment. *Neurobiology of Aging*, *84*, 237.e5-237.e12. <https://doi.org/10.1016/j.neurobiolaging.2019.06.001>

- Espinosa, A., Hernández-Olasagarre, B., Moreno-Grau, S., Kleineidam, L., Heilmann-Heimbach, S., Hernández, I., Wolfsgruber, S., Wagner, H., Rosende-Roca, M., Mauleón, A., Vargas, L., Lafuente, A., Rodríguez-Gómez, O., Abdelnour, C., Gil, S., Marquié, M., Santos-Santos, M. A., Sanabria, Á., Ortega, G., ... Ruiz, A. (2018). Exploring Genetic Associations of Alzheimer's Disease Loci With Mild Cognitive Impairment Neurocognitive Endophenotypes. *Frontiers in Aging Neuroscience, 10*. <https://doi.org/10.3389/fnagi.2018.00340>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32*(19). <https://doi.org/10.1093/bioinformatics/btw354>
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (2014). Mini-Mental State Examination. In *PsycTESTS Dataset*. <https://doi.org/10.1037/t07757-000>
- Fortes Marin, E., Carrera Marcolin, L., Martí Melero, L., Tintoré Gazulla, M., & Beltran Porres, M. (2025). The Prevalence of Single Nucleotide Polymorphisms of the AOC1 Gene Associated with Diamine Oxidase (DAO) Enzyme Deficiency in Healthy Newborns: A Prospective Population-Based Cohort Study. *Genes, 16*(2). <https://doi.org/10.3390/genes16020141>
- Fu, X., Eikelboom, R. H., Tian, R., Liu, B., Wang, S., & Jayakody, D. M. P. (2023). The Relationship of Age-Related Hearing Loss with Cognitive Decline and Dementia in a Sinitic Language-Speaking Adult Population: A Systematic Review and Meta-Analysis. *Innovation in Aging, 7*(1). <https://doi.org/10.1093/geroni/igac078>
- Fukui, H., Rünker, A., Fabel, K., Buchholz, F., & Kempermann, G. (2018). Transcription factor Runx1 is pro-neurogenic in adult hippocampal precursor cells. *PLOS ONE, 13*(1), e0190789. <https://doi.org/10.1371/journal.pone.0190789>
- Furney, S. J., Simmons, A., Breen, G., Pedroso, I., Lunnon, K., Proitsi, P., Hodges, A., Powell, J., Wahlund, L.-O., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Spenger, C., Lathrop, M., Shen, L., Kim, S., Saykin, A. J., ... Lovestone, S. (2011). Genome-wide association with MRI atrophy measures as a quantitative trait locus for Alzheimer's disease. *Molecular Psychiatry, 16*(11), 1130–1138. <https://doi.org/10.1038/mp.2010.123>
- Gallione, C. J., Detter, M. R., Sheline, A., Christmas, H. M., Lee, C., & Marchuk, D. A. (2022). Genetic genealogy uncovers a founder deletion mutation in the cerebral cavernous malformations 2 gene. *Human Genetics, 141*(11), 1761–1769. <https://doi.org/10.1007/s00439-022-02458-5>
- Garcia-Cifuentes, E., Jaramillo-Jimenez, A., Aguillon, D., Gómez-Vega, M., Velez-Hernandez, J. E., Cano Gutiérrez, C., & Lopera, F. (2019). Prevenir la demencia: un reto para la salud pública en Colombia. *Acta Neurológica Colombiana, 35*(4), 208–210. <https://doi.org/10.22379/24224022269>
- Genis-Mendoza, A., Martínez-Magaña, J., Téllez Martínez, J. A., Jiménez-Guenchi, J., Roche Bergua, A., Castañeda, C., Tovilla-Zarate, C. A., & Nicolini, H. (2020).

- Identification of high impact variants in TREM2 and ABCA7 in Mexican individuals diagnosed with Alzheimer's disease. *Revista Mexicana de Psiquiatría y Salud Mental*, 1(8), 224–229. <https://goo.su/L1u26>
- Genome Reference Consortium. (2019). GRCh38.p13 Genome Reference Assembly. *National Center for Biotechnology Information*. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39/
- Gil, L., Ruiz De Sánchez, C., Gil, F., Romero, S. J., & Pretelt Burgos, F. (2015). Validation of the Montreal Cognitive Assessment (MoCA) in Spanish as a screening tool for mild cognitive impairment and mild dementia in patients over 65 years old in Bogotá, Colombia. *International Journal of Geriatric Psychiatry*, 30(6). <https://doi.org/10.1002/gps.4199>
- Gjøra, L., Strand, B. H., Bergh, S., Borza, T., Brækhus, A., Engedal, K., Johannessen, A., Kvello-Alme, M., Krokstad, S., Livingston, G., Matthews, F. E., Myrstad, C., Skjellegrind, H., Thingstad, P., Aakhus, E., Aam, S., & Selbæk, G. (2021). Current and future prevalence estimates of mild cognitive impairment, dementia, and its subtypes in a population-based sample of people 70 years and older in Norway: The HUNT study. *Journal of Alzheimer's Disease*, 79(3). <https://doi.org/10.3233/JAD-201275>
- Gomar, J. J. (2011). Utility of Combinations of Biomarkers, Cognitive Markers, and Risk Factors to Predict Conversion From Mild Cognitive Impairment to Alzheimer Disease in Patients in the Alzheimer's Disease Neuroimaging Initiative. *Archives of General Psychiatry*, 68(9), 961. <https://doi.org/10.1001/archgenpsychiatry.2011.96>
- Granot-HersHKovitz, E., Xia, R., Yang, Y., Spitzer, B., Tarraf, W., Vásquez, P. M., Lipton, R. B., Daviglus, M., Argos, M., Cai, J., Kaplan, R., Fornage, M., DeCarli, C., Gonzalez, H. M., & Sofer, T. (2023). Interaction analysis of ancestry-enriched variants with APOE-ε4 on MCI in the Study of Latinos-Investigation of Neurocognitive Aging. *Scientific Reports*, 13(1), 5114. <https://doi.org/10.1038/s41598-023-32028-2>
- Gutiérrez Rodríguez, J., & Guzmán Gutiérrez, G. (2017). Definición y prevalencia del deterioro cognitivo leve. *Revista Española de Geriatría y Gerontología*, 52. [https://doi.org/10.1016/s0211-139x\(18\)30072-6](https://doi.org/10.1016/s0211-139x(18)30072-6)
- Han, M.-R., Schellenberg, G. D., & Wang, L.-S. (2010). Genome-wide association reveals genetic effects on human Aβ 42 and τ protein levels in cerebrospinal fluids: a case control study. *BMC Neurology*, 10(1), 90. <https://doi.org/10.1186/1471-2377-10-90>
- Haridy, S. F. A., Shahin, N. N., Shabayek, M. I., Selim, M. M., Abdelhafez, M. A., & Motawi, T. K. (2023). Diagnostic and prognostic value of the RUNXOR/RUNX1 axis in multiple sclerosis. *Neurobiology of Disease*, 178, 106032. <https://doi.org/10.1016/j.nbd.2023.106032>
- Henao-Arboleda, E., Moreno-Carrillo, C., Ramos, V., Aguirre-Acevedo, D. C., Pineda, D., & Lopera, F. (2010). Caracterización de síntomas neuropsiquiátricos en pacientes

con DCL de tipo amnésico en una población colombiana. *Revista Chilena de Neuropsicología*, 5(2), 153–159.

Henao Arboleda, E., Aguirre Acevedo, D. C., Muñoz, C., Pineda Salazar, D. A., & Lopera Restrepo, F. (2008). Prevalencia de deterioro cognitivo leve de tipo amnésico en una población colombiana. *Revista de Neurología*, 46(12), 709.

<https://doi.org/10.33588/rn.4612.2007569>

Horgusluoglu-Moloch, E., Nho, K., Risacher, S. L., Kim, S., Foroud, T., Shaw, L. M., Trojanowski, J. Q., Aisen, P. S., Petersen, R. C., Jack, C. R., Lovestone, S., Simmons, A., Weiner, M. W., & Saykin, A. J. (2017). Targeted neurogenesis pathway-based gene analysis identifies ADORA2A associated with hippocampal volume in mild cognitive impairment and Alzheimer's disease. *Neurobiology of Aging*, 60, 92–103. <https://doi.org/10.1016/j.neurobiolaging.2017.08.010>

Hostage, C. A., Roy Choudhury, K., Doraiswamy, P. M., & Petrella, J. R. (2013). Dissecting the Gene Dose-Effects of the APOE ϵ 4 and ϵ 2 Alleles on Hippocampal Volumes in Aging and Alzheimer's Disease. *PLoS ONE*, 8(2), e54483.

<https://doi.org/10.1371/journal.pone.0054483>

Hu, T., Chen, J., Lin, X., He, W., Liang, H., Wang, M., Li, W., Wu, Z., Han, M., Jin, X., Kristiansen, K., Xiao, L., & Zou, Y. (2024). Comparison of the DNBSEQ platform and Illumina HiSeq 2000 for bacterial genome assembly. *Scientific Reports*, 14(1), 1292.

<https://doi.org/10.1038/s41598-024-51725-0>

Hu, X., Pickering, E. H., Hall, S. K., Naik, S., Liu, Y. C., Soares, H., Katz, E., Paciga, S. A., Liu, W., Aisen, P. S., Bales, K. R., Samad, T. A., & John, S. L. (2011). Genome-wide association study identifies multiple novel loci associated with disease progression in subjects with mild cognitive impairment. *Translational Psychiatry*, 1(11), e54–e54. <https://doi.org/10.1038/tp.2011.50>

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140(6).

<https://doi.org/10.1192/bjp.140.6.566>

Hussenoeder, F. S., Conrad, I., Roehr, S., Fuchs, A., Pentzek, M., Bickel, H., Moesch, E., Weyerer, S., Werle, J., Wiese, B., Mamone, S., Brettschneider, C., Hesel, K., Kleineidam, L., Kaduszkiewicz, H., Eisele, M., Maier, W., Wagner, M., Scherer, M., ... Riedel-Heller, S. G. (2020). Mild cognitive impairment and quality of life in the oldest old: a closer look. *Quality of Life Research*, 29(6).

<https://doi.org/10.1007/s11136-020-02425-5>

Kelley, B. J., & Petersen, R. C. (2007). Alzheimer's Disease and Mild Cognitive Impairment. In *Neurologic Clinics* (Vol. 25, Issue 3, pp. 577–609).

<https://doi.org/10.1016/j.ncl.2007.03.008>

Kim, H.-M., Jeon, S., Chung, O., Jun, J. H., Kim, H.-S., Blazyte, A., Lee, H.-Y., Yu, Y., Cho, Y. S., Bolser, D. M., & Bhak, J. (2021). Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina

- sequencing benchmark for whole-genome sequencing. *GigaScience*, *10*(3).
<https://doi.org/10.1093/gigascience/giab014>
- Kong, X., Liu, Z., Huang, L., Wang, X., Yang, Z., Zhou, G., Zhen, Z., & Liu, J. (2015). Mapping Individual Brain Networks Using Statistical Similarity in Regional Morphology from MRI. *PLOS ONE*, *10*(11), e0141840.
<https://doi.org/10.1371/journal.pone.0141840>
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*(1), 29. <https://doi.org/10.1186/1746-4811-9-29>
- Kubo, S., Yamamoto, H., Kajimura, N., Omori, Y., Maeda, Y., Chaya, T., & Furukawa, T. (2021). Functional analysis of Samd11, a retinal photoreceptor PRC1 component, in establishing rod photoreceptor identity. *Scientific Reports*, *11*(1), 4180.
<https://doi.org/10.1038/s41598-021-83781-1>
- Lacour, A., Espinosa, A., Louwersheimer, E., Heilmann, S., Hernández, I., Wolfsgruber, S., Fernández, V., Wagner, H., Rosende-Roca, M., Mauleón, A., Moreno-Grau, S., Vargas, L., Pijnenburg, Y. A. L., Koene, T., Rodríguez-Gómez, O., Ortega, G., Ruiz, S., Holstege, H., Sotolongo-Grau, O., ... Ruiz, A. (2017). Genome-wide significant risk factors for Alzheimer's disease: role in progression to dementia due to Alzheimer's disease among subjects with mild cognitive impairment. *Molecular Psychiatry*, *22*(1), 153–160. <https://doi.org/10.1038/mp.2016.18>
- Lee, E., Giovanello, K. S., Saykin, A. J., Xie, F., Kong, D., Wang, Y., Yang, L., Ibrahim, J. G., Doraiswamy, P. M., & Zhu, H. (2017). Single-nucleotide polymorphisms are associated with cognitive decline at Alzheimer's disease conversion within mild cognitive impairment patients. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *8*(1), 86–95. <https://doi.org/10.1016/j.dadm.2017.04.004>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, *27*(21), 2987–2993.
<https://doi.org/10.1093/bioinformatics/btr509>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.
<https://doi.org/10.1093/bioinformatics/btp324>
- Li, J. Q., Tan, L., Wang, H. F., Tan, M. S., Tan, L., Xu, W., Zhao, Q. F., Wang, J., Jiang, T., & Yu, J. T. (2016). Risk factors for predicting progression from mild cognitive impairment to Alzheimer's disease: A systematic review and meta-analysis of cohort studies. *Journal of Neurology, Neurosurgery and Psychiatry*, *87*(5).
<https://doi.org/10.1136/jnnp-2014-310095>
- Li, L., Yang, Y., Zhang, Q., Wang, J., Jiang, J., & Neuroimaging Initiative, A. D. (2021). Use of Deep-Learning Genomics to Discriminate Healthy Individuals from Those with Alzheimer's Disease or Mild Cognitive Impairment. *Behavioural Neurology*, *2021*, 1–15. <https://doi.org/10.1155/2021/3359103>

- Lin, F., Marchetti, S., Pluim, D., Iusuf, D., Mazzanti, R., Schellens, J. H. M., Beijnen, J. H., & van Tellingen, O. (2013). Abcc4 Together with Abcb1 and Abcg2 Form a Robust Cooperative Drug Efflux System That Restricts the Brain Entry of Camptothecin Analogues. *Clinical Cancer Research*, *19*(8), 2084–2095. <https://doi.org/10.1158/1078-0432.CCR-12-3105>
- Liu, B., Ruan, J., Chen, M., Li, Z., Manjengwa, G., Schlüter, D., Song, W., & Wang, X. (2022). Deubiquitinating enzymes (DUBs): decipher underlying basis of neurodegenerative diseases. *Molecular Psychiatry*, *27*(1), 259–268. <https://doi.org/10.1038/s41380-021-01233-8>
- Liu, L., Zhang, D., Liu, H., & Arendt, C. (2013). Robust methods for population stratification in genome wide association studies. *BMC Bioinformatics*, *14*(1), 132. <https://doi.org/10.1186/1471-2105-14-132>
- Liu, P., Liu, S., Zhu, C., Li, Y., Li, Y., Fei, X., Hou, J., Wang, X., & Pan, Y. (2023). The deubiquitinating enzyme MINDY2 promotes pancreatic cancer proliferation and metastasis by stabilizing ACTN4 expression and activating the PI3K/AKT/mTOR signaling pathway. *Frontiers in Oncology*, *13*, 1169833. <https://doi.org/10.3389/fonc.2023.1169833>
- Lv, N., Wang, Y., Zhao, M., Dong, L., & Wei, H. (2021). The Role of PAX2 in Neurodevelopment and Disease. *Neuropsychiatric Disease and Treatment, Volume 17*, 3559–3567. <https://doi.org/10.2147/NDT.S332747>
- Mak, S. S. T., Gopalakrishnan, S., Carøe, C., Geng, C., Liu, S., Sinding, M.-H. S., Kuderna, L. F. K., Zhang, W., Fu, S., Vieira, F. G., Germonpré, M., Bocherens, H., Fedorov, S., Petersen, B., Sicheritz-Pontén, T., Marques-Bonet, T., Zhang, G., Jiang, H., & Gilbert, M. T. P. (2017). Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *GigaScience*, *6*(8). <https://doi.org/10.1093/gigascience/gix049>
- Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, *27*(2), e1608. <https://doi.org/10.1002/mpr.1608>
- Maulik, U., Sen, S., Mallik, S., & Bandyopadhyay, S. (2018). Detecting TF-miRNA-gene network based modules for 5hmC and 5mC brain samples: a intra- and inter-species case-study between human and rhesus. *BMC Genetics*, *19*(1), 9. <https://doi.org/10.1186/s12863-017-0574-7>
- Mayeux, R., Saunders, A. M., Shea, S., Mirra, S., Evans, D., Roses, A. D., Hyman, B. T., Crain, B., Tang, M.-X., & Phelps, C. H. (1998). Utility of the Apolipoprotein E Genotype in the Diagnosis of Alzheimer's Disease. *New England Journal of Medicine*, *338*(8), 506–511. <https://doi.org/10.1056/NEJM199802193380804>
- McInnis, J. J., Sood, D., Guo, L., Dufault, M. R., Garcia, M., Passaro, R., Gao, G., Zhang, B., & Dodge, J. C. (2024). Unravelling neuronal and glial differences in ceramide

- composition, synthesis, and sensitivity to toxicity. *Communications Biology*, 7(1), 1597. <https://doi.org/10.1038/s42003-024-07231-0>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1), 122. <https://doi.org/10.1186/s13059-016-0974-4>
- Mesa Interinstitucional de Población. (2023). 3.3. ENVEJECIMIENTO Y DERECHOS DE LAS PERSONAS MAYORES. In Fondo de Población de las Naciones Unidas(UNFPA) (Ed.), *Análisis de Situación de Población (ASP Colombia 2023) Presentación y Capítulo dinámica demográfica* (p. 28). Departamento Nacional de Planeación. https://colombia.unfpa.org/sites/default/files/pub-pdf/3.3_envejecimiento.pdf
- Ministerio de Salud y Protección Social. (2013). *Plan Decenal de Salud Pública 2012-2021*. Ministerio de Salud y Protección Social. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/PSP/PDSP.pdf>
- Ministerio de Salud y Protección Social. (2017). *Boletín de salud mental Demencia*. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/ENT/Boletin-demencia-salud-mental.pdf>
- Ministerio de Salud y Protección Social, & Colciencias. (2015). *Política Colombiana de Envejecimiento Humano y Vejez 2015-2024* (p. 54). Ministerio de Salud y Protección Social. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/PS/Política-colombiana-envejecimiento-humano-vejez-2015-2024.pdf>
- Moreira, T., & Bond, J. (2008). Does the prevention of brain ageing constitute anti-ageing medicine? Outline of a new space of representation for Alzheimer's Disease. *Journal of Aging Studies*, 22(4), 356–365. <https://doi.org/10.1016/j.jaging.2008.05.008>
- Morozova, A., Zorkina, Y., Abramova, O., Pavlova, O., Pavlov, K., Soloveva, K., Volkova, M., Alekseeva, P., Andryshchenko, A., Kostyuk, G., Gurina, O., & Chekhonin, V. (2022). Neurobiological Highlights of Cognitive Impairment in Psychiatric Disorders. *International Journal of Molecular Sciences*, 23(3). <https://doi.org/10.3390/ijms23031217>
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2014). Montreal Cognitive Assessment. In *PsycTESTS Dataset*. <https://doi.org/10.1037/t27279-000>

- Ng, P. C., & Kirkness, E. F. (2010). Whole genome sequencing. In *Methods in Molecular Biology* (Vol. 628). https://doi.org/10.1007/978-1-60327-367-1_12
- Ophey, A., Wolfsgruber, S., Roeske, S., Polcher, A., Spottke, A., Frölich, L., Hüll, M., Jessen, F., Kornhuber, J., Maier, W., Peters, O., Ramirez, A., Wiltfang, J., Liepelt-Scarfone, I., Becker, S., Berg, D., Schulz, J. B., Reetz, K., Wojtala, J., ... Kalbe, E. (2021). Cognitive profiles of patients with mild cognitive impairment due to Alzheimer's versus Parkinson's disease defined using a base rate approach: Implications for neuropsychological assessments. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *13*(1). <https://doi.org/10.1002/dad2.12223>
- Ospina García, N. (2015). *Adaptación y validación en Colombia del addenbrooke's cognitive examination-revisado (ACE-R) en pacientes con deterioro cognoscitivo leve y demencia*. [Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/handle/unal/52292>
- Pathak, G. A., Silzer, T. K., Sun, J., Zhou, Z., Daniel, A. A., Johnson, L., O'Bryant, S., Phillips, N. R., & Barber, R. C. (2019). Genome-Wide Methylation of Mild Cognitive Impairment in Mexican Americans Highlights Genes Involved in Synaptic Transport, Alzheimer's Disease-Precursor Phenotypes, and Metabolic Morbidities. *Journal of Alzheimer's Disease*, *72*(3), 733–749. <https://doi.org/10.3233/JAD-190634>
- Pedraza L, O. L., Sánchez, E., Plata, S. J., Montalvo, C., Galvis, P., Chiquillo, A., & Arévalo-Rodríguez, I. (2014). Puntuaciones del MoCA y el MMSE en pacientes con deterioro cognitivo leve y demencia en una clínica de memoria en Bogotá. *Acta Neurológica Colombiana*, *30*(1).
- Periñán, M. T., Macías-García, D., Labrador-Espinosa, M. Á., Jesús, S., Buiza-Rueda, D., Adarmes-Gómez, A. D., Muñoz-Delgado, L., Gómez-Garre, P., & Mir, P. (2021). Association of PICALM with Cognitive Impairment in Parkinson's Disease. *Movement Disorders*, *36*(1), 118–123. <https://doi.org/10.1002/mds.28283>
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3). <https://doi.org/10.1111/j.1365-2796.2004.01388.x>
- Petersen, R. C. (2016). Mild Cognitive Impairment. *CONTINUUM: Lifelong Learning in Neurology*, *22*(2, Dementia), 404–418. <https://doi.org/10.1212/CON.0000000000000313>
- Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2017). *Scaling accurate genetic variant discovery to tens of thousands of samples*. <https://doi.org/10.1101/201178>
- Porreca, G. J. (2010). Genome sequencing on nanoballs. *Nature Biotechnology*, *28*(1), 43–44. <https://doi.org/10.1038/nbt0110-43>

- Pradhan, J., Noakes, P. G., & Bellingham, M. C. (2019). The Role of Altered BDNF/TrkB Signaling in Amyotrophic Lateral Sclerosis. *Frontiers in Cellular Neuroscience*, *13*. <https://doi.org/10.3389/fncel.2019.00368>
- Pradilla A., G., Vesga A., B. E., & León-Sarmiento, F. E. (2003). Estudio neuroepidemiológico nacional (EPINEURO) colombiano. *Revista Panamericana de Salud Pública*, *14*(2). <https://doi.org/10.1590/s1020-49892003000700005>
- Prince, M., Wimo, A., Guerchet, M., Gemma-Claire, A., Wu, Y.-T., & Prina, M. (2015). World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends. *Alzheimer's Disease International*. <https://doi.org/10.1111/j.0963-7214.2004.00293.x>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. <https://doi.org/10.1086/519795>
- Raghavan, N. S., Dumitrescu, L., Mormino, E., Mahoney, E. R., Lee, A. J., Gao, Y., Bilgel, M., Goldstein, D., Harrison, T., Engelman, C. D., Saykin, A. J., Whelan, C. D., Liu, J. Z., Jagust, W., Albert, M., Johnson, S. C., Yang, H.-S., Johnson, K., Aisen, P., ... Mayeux, R. (2020). Association Between Common Variants in RBF1, an RNA-Binding Protein, and Brain Amyloidosis in Early and Preclinical Alzheimer Disease. *JAMA Neurology*, *77*(10), 1288. <https://doi.org/10.1001/jamaneurol.2020.1760>
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198. <https://doi.org/10.1093/nar/gkz369>
- Reitz, C., & Mayeux, R. (2010). Use of Genetic Variation as Biomarkers for Mild Cognitive Impairment and Progression of Mild Cognitive Impairment to Dementia. *Journal of Alzheimer's Disease*, *19*(1), 229–251. <https://doi.org/10.3233/JAD-2010-1255>
- Riess, O., Thies, U., Siedlaczek, I., Potisek, S., Graham, R., Theilmann, J., Grimm, T., Epplen, J. T., & Hayden, M. R. (1994). Precise Mapping of the Brain α 2-Adrenergic Receptor Gene within Chromosome 4p16. *Genomics*, *19*(2), 298–302. <https://doi.org/10.1006/geno.1994.1061>
- Ríos-Gallardo, Á. M., Muñoz-Bernal, L. F., Aldana-Camacho, L. V., Santamaría-Íñiguez, M. F., & Villanueva-Bonilla, C. (2017). Perfil neuropsicológico de un grupo de adultos mayores diagnosticados con deterioro cognitivo leve. *Revista Mexicana de Neurociencia*, *18*(5).
- Roberts, R., & Knopman, D. S. (2013). Classification and Epidemiology of MCI. *Clinics in Geriatric Medicine*, *29*(4), 753–772. <https://doi.org/10.1016/j.cger.2013.07.003>

- Robins, C., Liu, Y., Fan, W., Duong, D. M., Meigs, J., Harerimana, N. V., Gerasimov, E. S., Dammer, E. B., Cutler, D. J., Beach, T. G., Reiman, E. M., De Jager, P. L., Bennett, D. A., Lah, J. J., Wingo, A. P., Levey, A. I., Seyfried, N. T., & Wingo, T. S. (2021). Genetic control of the human brain proteome. *The American Journal of Human Genetics*, *108*(3), 400–410. <https://doi.org/10.1016/j.ajhg.2021.01.012>
- Rollano, O. M., & Mollinedo, P. (2017). Análisis Bioinformático De Arn-Seq Con Una Perspectiva Para Bolivia. *Revista Boliviana de Química*, *34*(2). <https://bit.ly/3vk7vX5>
- Sachdev, P. S., Lipnicki, D. M., Kochan, N. A., Crawford, J. D., Thalamuthu, A., Andrews, G., Brayne, C., Matthews, F. E., Stephan, B. C. M., Lipton, R. B., Katz, M. J., Ritchie, K., Carrière, I., Ancelin, M. L., Lam, L. C. W., Wong, C. H. Y., Fung, A. W. T., Guaita, A., Vaccaro, R., ... Lobo, E. (2015). The prevalence of mild cognitive impairment in diverse geographical and ethnocultural regions: The COSMIC Collaboration. *PLoS ONE*, *10*(11). <https://doi.org/10.1371/journal.pone.0142388>
- Sager, K. L., Wu, J., Leurgans, S. E., Rees, H. D., Gearing, M., Mufson, E. J., Levey, A. I., & Lah, J. J. (2007). Neuronal LR11/sorLA expression is reduced in mild cognitive impairment. *Annals of Neurology*, *62*(6), 640–647. <https://doi.org/10.1002/ana.21190>
- Sarmiento Buitrago, A. F., Cerón Perdomo, D., & Mayorga Bogota, M. A. (2024). Asociación entre el deterioro cognitivo y factores socioeconómicos y sociodemográficos en adultos mayores colombianos. *Revista Colombiana de Psiquiatría*, *53*(2), 134–141. <https://doi.org/10.1016/j.rcp.2022.02.005>
- Saykin, A. J., Shen, L., Yao, X., Kim, S., Nho, K., Risacher, S. L., Ramanan, V. K., Foroud, T. M., Faber, K. M., Sarwar, N., Munsie, L. M., Hu, X., Soares, H. D., Potkin, S. G., Thompson, P. M., Kauwe, J. S. K., Kaddurah-Daouk, R., Green, R. C., Toga, A. W., & Weiner, M. W. (2015). Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimer's & Dementia*, *11*(7), 792–814. <https://doi.org/10.1016/j.jalz.2015.05.009>
- Shen, L., Kim, S., Risacher, S. L., Nho, K., Swaminathan, S., West, J. D., Foroud, T., Pankratz, N., Moore, J. H., Sloan, C. D., Weiner, M. W., & Saykin, A. J. (2010). Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *NeuroImage*, *53*(3), 1051–1063. <https://doi.org/10.1016/j.neuroimage.2010.01.042>
- Shen, Y., Wang, H., Sun, Q., Yao, H., Keegan, A. P., Mullan, M., Wilson, J., Lista, S., Leyhe, T., Laske, C., Rujescu, D., Levey, A., Wallin, A., Blennow, K., Li, R., & Hampel, H. (2018). Increased Plasma Beta-Secretase 1 May Predict Conversion to Alzheimer's Disease Dementia in Individuals With Mild Cognitive Impairment. *Biological Psychiatry*, *83*(5). <https://doi.org/10.1016/j.biopsych.2017.02.007>
- Smith, A. D., & de Sena Brandine, G. (2021). Falco: High-speed FastQC emulation for quality control of sequencing data. *F1000Research*, *8*. <https://doi.org/10.12688/f1000research.21142.2>

- Stein, J. L., Hua, X., Morra, J. H., Lee, S., Hibar, D. P., Ho, A. J., Leow, A. D., Toga, A. W., Sul, J. H., Kang, H. M., Eskin, E., Saykin, A. J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M. J., Craig, D. W., Gerber, J. D., Allen, A. N., ... Thompson, P. M. (2010). Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. *NeuroImage*, *51*(2), 542–554. <https://doi.org/10.1016/j.neuroimage.2010.02.068>
- Stites, S. D., Harkins, K., Rubright, J. D., & Karlawish, J. (2018). Relationships between cognitive complaints and quality of life in older adults with mild cognitive impairment, mild Alzheimer disease dementia, and normal cognition. *Alzheimer Disease and Associated Disorders*, *32*(4). <https://doi.org/10.1097/WAD.0000000000000262>
- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*(D1), D638–D646. <https://doi.org/10.1093/nar/gkac1000>
- Szot, P., White, S. S., Greenup, J. L., Leverenz, J. B., Peskind, E. R., & Raskind, M. A. (2006). Compensatory Changes in the Noradrenergic Nervous System in the Locus Ceruleus and Hippocampus of Postmortem Subjects with Alzheimer's Disease and Dementia with Lewy Bodies. *The Journal of Neuroscience*, *26*(2), 467–478. <https://doi.org/10.1523/JNEUROSCI.4265-05.2006>
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. In *Nature Reviews Genetics* (Vol. 20, Issue 8). <https://doi.org/10.1038/s41576-019-0127-1>
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 59. <https://doi.org/10.1038/s43586-021-00056-9>
- van der Auwera, G., & O'Connor, B. D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Incorporated. <https://books.google.com.co/books?id=wwiCswEACAAJ>
- Wang, K., Lu, Y., Morrow, D. F., Xiao, D., & Xu, C. (2022). Associations of ARHGAP26 Polymorphisms with Alzheimer's Disease and Cardiovascular Disease. *Journal of Molecular Neuroscience*, *72*(5). <https://doi.org/10.1007/s12031-022-01972-5>
- Wang, M. H., Cordell, H. J., & Van Steen, K. (2019). Statistical methods for genome-wide association studies. *Seminars in Cancer Biology*, *55*, 53–60. <https://doi.org/10.1016/j.semcancer.2018.04.008>
- Wang, Y., Wang, Y., Tang, J., Li, R., Jia, Y., Yang, H., & Wei, H. (2024). Impaired neural circuitry of hippocampus in Pax2 nervous system-specific knockout mice leads to restricted repetitive behaviors. *CNS Neuroscience & Therapeutics*, *30*(4). <https://doi.org/10.1111/cns.14482>

- Wechsler, D. (2012). Test de inteligencia de Wechsler para adultos-IV (WAIS-IV). *Explicación Del Test*.
- Wen, J., Cui, Y., Yang, Z., Bao, J., Chen, J., Erus, G., Abdulkadir, A., Mamourian, E., Singh, A., Yang, S., Fan, Y., Saykin, A. J., Thompson, P. M., Jun, G. R., Ritchie, M. D., Shen, L., Wolk, D. A., Shou, H., Nasrallah, I. M., & Davatzikos, C. (2022). Genetic heterogeneity of four MCI/AD neuroanatomical dimensions discovered via deep learning. *Alzheimer's & Dementia*, *18*(S6). <https://doi.org/10.1002/alz.065223>
- Whitley, E., Deary, I. J., Ritchie, S. J., Batty, G. D., Kumari, M., & Benzeval, M. (2016). Variations in cognitive abilities across the life course: Cross-sectional evidence from Understanding Society: The UK Household Longitudinal Study. *Intelligence*, *59*. <https://doi.org/10.1016/j.intell.2016.07.001>
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. In *Springer-Verlag New York*. <https://ggplot2.tidyverse.org>
- Xi, J., Ding, D., Zhao, Q., Liang, X., Zheng, L., Guo, Q., Hong, Z., Fu, H., Xu, J., & Xiao, Q. (2020). Joint Effect of ABCA7 rs4147929 and Body Mass Index on Progression from Mild Cognitive Impairment to Alzheimer's Disease: The Shanghai Aging Study. *Current Alzheimer Research*, *17*(2), 185–195. <https://doi.org/10.2174/1567205017666200317095608>
- Xiang, J., Wang, X., Gao, Y., Li, T., Cao, R., Yan, T., Ma, Y., Niu, Y., Xue, J., & Wang, B. (2020). Phosphodiesterase 4D Gene Modifies the Functional Network of Patients With Mild Cognitive Impairment and Alzheimer's Disease. *Frontiers in Genetics*, *11*. <https://doi.org/10.3389/fgene.2020.00890>
- Xu, S., Duan, P., Li, J., Senkowski, T., Guo, F., Chen, H., Romero, A., Cui, Y., Liu, J., & Jiang, S.-W. (2016). Zinc Finger and X-Linked Factor (ZFX) Binds to Human SET Transcript 2 Promoter and Transactivates SET Expression. *International Journal of Molecular Sciences*, *17*(10). <https://doi.org/10.3390/ijms17101737>
- Xu, Y., Lin, Z., Tang, C., Tang, Y., Cai, Y., Zhong, H., Wang, X., Zhang, W., Xu, C., Wang, J., Wang, J., Yang, H., Yang, L., & Gao, Q. (2019). A new massively parallel nanoball sequencing platform for whole exome research. *BMC Bioinformatics*, *20*(1), 153. <https://doi.org/10.1186/s12859-019-2751-3>
- Yadav, S. K., Bhat, A. A., Hashem, S., Nisar, S., Kamal, M., Syed, N., Temanni, M.-R., Gupta, R. K., Kamran, S., Azeem, M. W., Srivastava, A. K., Bagga, P., Chawla, S., Reddy, R., Frenneaux, M. P., Fakhro, K., & Haris, M. (2021). Genetic variations influence brain changes in patients with attention-deficit hyperactivity disorder. *Translational Psychiatry*, *11*(1), 349. <https://doi.org/10.1038/s41398-021-01473-w>
- Yoganathan, S., Arunachal, G., Gowda, V. K., Vinayan, K. P., Thomas, M., Whitney, R., & Jain, P. (2021). NTRK2-related developmental and epileptic encephalopathy: Report of 5 new cases. *Seizure*, *92*, 52–55. <https://doi.org/10.1016/j.seizure.2021.08.008>

- Zhang, J., Wang, X., Duan, H., Chen, C., Lu, Z., Zhang, D., & Li, S. (2023). The Association of Calcium Signaling Pathway Gene Variants, Bone Mineral Density and Mild Cognitive Impairment in Elderly People. *Genes*, *14*(4), 828. <https://doi.org/10.3390/genes14040828>
- Zhang, X. (2020). Review of genome-wide association study. In *Kexue Tongbao/Chinese Science Bulletin* (Vol. 65, Issue 8, pp. 671–683). Chinese Academy of Sciences. <https://doi.org/10.1360/TB-2019-0063>
- Zhang, Y., Elgart, M., Granot-Herskovitz, E., Wang, H., Tarraf, W., Ramos, A. R., Stickel, A. M., Zeng, D., Garcia, T. P., Testai, F. D., Wassertheil-Smoller, S., Isasi, C. R., Daviglius, M. L., Kaplan, R., Fornage, M., DeCarli, C., Redline, S., González, H. M., & Sofer, T. (2023). Genetic associations between sleep traits and cognitive ageing outcomes in the Hispanic Community Health Study/Study of Latinos. *EBioMedicine*, *87*, 104393. <https://doi.org/10.1016/j.ebiom.2022.104393>
- Zhao, Y., Hu, D., Wang, R., Sun, X., Ropelewski, P., Hubler, Z., Lundberg, K., Wang, Q., Adams, D. J., Xu, R., & Qi, X. (2022). ATAD3A oligomerization promotes neuropathology and cognitive deficits in Alzheimer's disease models. *Nature Communications*, *13*(1), 1121. <https://doi.org/10.1038/s41467-022-28769-9>
- Zhao, Y., Sun, X., Hu, D., Prosdocimo, D. A., Hoppel, C., Jain, M. K., Ramachandran, R., & Qi, X. (2019). ATAD3A oligomerization causes neurodegeneration by coupling mitochondrial fragmentation and bioenergetics defects. *Nature Communications*, *10*(1), 1371. <https://doi.org/10.1038/s41467-019-09291-x>
- Zhou, Y., Hao, N., Sander, J. W., Lin, X., Xiong, W., & Zhou, D. (2023). KCNH2 variants in a family with epilepsy and long QT syndrome: A case report and literature review. *Epileptic Disorders*, *25*(4), 492–499. <https://doi.org/10.1002/epd2.20046>