

Determinación de variaciones genéticas relacionadas a enfermedades neurodegenerativas mediante secuenciación masiva por exoma clínico de personas mayores de Panamá.

Autores:

Kensy Alvelo¹ †, Carolina E. Rodríguez R.² †, Estefani C. Sanchez Castillo³, Daniela N. Creus⁴, Paulett L. Gonzalez⁴, Juan Mewa⁵, Yaxelis Mendoza⁵, Giselle A. Rangel^{2, 6*}, Alcibiades E Villarreal^{2, 6*} y Gabrielle Britton^{2, 6*}.

Afiliaciones:

1. Facultad de Ciencias Naturales, Exactas y Tecnología. Universidad de Panamá (UP). Ciudad de Panamá, República de Panamá.
2. Instituto de Investigaciones Científicas y Servicios de Alta Tecnología (INDICASAT-AIP), Ciudad del Saber, República de Panamá.
3. Instituto de Investigación, Innovación y Gestión del Conocimiento, Caja de Seguro Social (CSS), Ciudad de la Salud, República de Panamá.
4. Facultad de Medicina. Escuela de Tecnología Médica. Universidad de Panamá (UP), Ciudad de Panamá, República de Panamá.
5. Instituto Conmemorativo Gorgas de Estudios de la Salud (ICGES), Ciudad de Panamá, República de Panamá.
6. Sistema Nacional de Investigación (SNI) in Panama's Secretaría Nacional de Ciencia Tecnología e Innovación (SENACYT), Panama.

† Estos autores contribuyeron de manera equitativa en el artículo.

Autores por correspondencia:

Giselle Rangel, PhD. E-mail: grangel@indicat.org.pa; Teléfono: (+507) 69975525.

Alcibiades E. Villarreal, PhD. E-mail: avillarreal@indicat.org.pa; Teléfono (+507) 6070-7966, Dirección Postal: Edificio 208, Ciudad del Saber, Panamá.

DOI: <https://doi.org/10.66253/conalac.v9.n1.a04>

Abstract

Neurodegenerative diseases (ND) represent a growing worldwide problem due to the increase in population aging that is advancing rapidly and where age is the most recognized risk factor for the development of multiple ND. Population aging is occurring more rapidly in low- and middle-income countries, such as Panama, making it a priority for the country to identify genetic factors associated with these conditions at an early stage, which is necessary information for the development of effective preventive and therapeutic strategies. The objective of this study was to identify genetic variants related to ND in older adults in Panama. A total of 44 DNA samples from participants ≥ 50 years recruited in the community by the Panama Aging Research Initiative-Health Disparities (PARI-HD) were analyzed by massive sequencing using clinical exome. The data obtained were analyzed using a customized virtual panel for the identification of variants in 138 genes related to ND. A total of 294 genetic variants were identified, 73.9% being benign, 20.7% variants of uncertain significance (VUS), 3.4% probably pathogenic and 2.0% pathogenic. Presenting variants such as CR1 and TREM2 related to Alzheimer's disease, PRKN, NEFH and GIGYF2 related to Parkinson's disease, HLA-DRB1 with multiple sclerosis, amyotrophic lateral sclerosis and GRN variant related to frontotemporal dementia. This study establishes the basis for a better understanding of genetic risk factors in our population, and is conducive to the integration of massive sequencing as a tool for the personalization of the clinical management of these pathologies.

Resumen

Las enfermedades neurodegenerativas (EN) son un problema creciente sobre la población mundial debido al incremento del envejecimiento poblacional que avanza rápidamente y en donde la edad es el factor de riesgo más reconocido para el desarrollo de múltiples EN. El envejecimiento poblacional está ocurriendo más rápidamente en países de bajo y mediano ingreso, como lo es Panamá, por lo que resulta prioritario para el país, la identificación temprana de factores genéticos asociados a estas condiciones, información necesaria para el desarrollo de estrategias preventivas y terapéuticas efectivas. Este estudio tuvo como objetivo identificar variantes genéticas relacionadas a EN en adultos mayores de Panamá. Un total de 44 muestras de ADN de participantes ≥ 50 años reclutados en la comunidad por la Iniciativa de Investigación sobre el envejecimiento en Panamá-Disparidades de Salud (PARI-HD) fueron analizadas por secuenciación masiva utilizando exoma clínico. Los datos obtenidos fueron analizados utilizando un panel virtual personalizado para la identificación de variantes en 138 genes relacionados con EN. Se identificaron un total de 294 variantes genéticas, resultando el 73,9% benignas, el 20,7% variantes de

variantes de significado incierto (VUS, por sus siglas en inglés), el 3,4% probablemente patogénicas y el 2,0% patogénicas. Presentando variantes como CR1 y TREM2 relacionadas con la enfermedad de Alzheimer, PRKN, NEFH y GIGYF2 relacionadas con la enfermedad de Parkinson, HLA-DRB1 con la esclerosis múltiple, esclerosis lateral amiotrófica y la variante GRN relacionada con la demencia frontotemporal. Este estudio establece las bases para una mejor comprensión de los factores de riesgo genéticos en nuestra población, y que propicia a la integración de la secuenciación masiva como herramienta para la personalización del manejo clínico de estas patologías.

Introducción

Las enfermedades neurodegenerativas (EN) como la Enfermedad de Alzheimer (EA), Enfermedad de Parkinson (EP), Esclerosis Lateral Amiotrófica (ELA), Demencia Frontotemporal, Demencia por cuerpos de Lewy, Distrofia y Demencia de inicio temprano, entre otras, son un problema creciente a nivel mundial debido a su estrecha relación con el envejecimiento poblacional. Dos de cada tres personas en edad avanzada viven en las regiones de Latinoamérica, el Caribe y Asia, este número se incrementará en el 2050 a 4 de 57 (Alvarez-Mora et al., 2023) representando para Panamá un aumento del 24% de la población. De acuerdo con las estadísticas sanitarias mundiales, durante el 2019 la Organización Mundial de Salud reporta que hubo un incremento a 7 de 10 (44%) la frecuencia de muertes en el mundo causadas por enfermedades no transmisibles, posicionando la Enfermedad de Alzheimer en tercer lugar tanto en América como en Europa representando el 60% a 80% en los casos de demencia, destacando a las mujeres como el grupo más afectado (Alzheimer's Association., 2023) (Organización Mundial de la Salud, 2020).

Por tanto, la identificación temprana de factores genéticos asociados a estas condiciones es crucial para el desarrollo de estrategias preventivas y terapéuticas efectivas.

Se han logrado grandes avances en nuestra capacidad para determinar la etiología genética de muchas enfermedades neurológicas mediante la utilización de plataformas de secuenciación de nueva generación (NGS por su sigla en inglés). Esta tecnología ha proporcionado nuevos avances en la búsqueda de los mecanismos causantes de las enfermedades neurológicas (Bras et al., 2012). La utilización de NGS ofrece la posibilidad de tamizar regiones regulatorias de proteínas codificantes y no codificantes de los genes candidatos a biomarcadores, así, generar conocimiento sobre cómo se pueden determinar diferentes perfiles de genes como factores de riesgo para la aparición de las EN en la población panameña donde podría ayudar a dirigir una atención personalizada al individuo beneficiando el pronóstico de la enfermedad. Este estudio tuvo como propósito la determinación de variaciones genéticas vinculadas a enfermedades neurodegenerativas en una muestra de personas mayores de Panamá mediante técnica NGS por exoma clínico.

MATERIALES Y MÉTODOS

Características y tamaño de la muestra

Se realizó un estudio analítico descriptivo y observacional de una submuestra de 44 participantes reclutados por La Iniciativa Panameña de Investigación sobre el Envejecimiento- Disparidades en Salud (PARI-HD).

PARI-HD es un estudio longitudinal sobre la salud cognitiva y funcional de los adultos mayores en Panamá, donde se reclutan personas de 50 o más años de edad, alfabetizados y sin demencia reclutados en la comunidad mediante muestreo por conveniencia a través de medios de comunicación disponibles, de boca a boca y la divulgación en la comunidad.

Extracción de ADN

El ADN de los participantes fue extraído a partir de muestras sanguíneas, utilizando el kit BIO-TEK D3392-02 E.Z.N.A.® Blood DNA, basado en el sistema de columnas con membrana de gel de sílice. Se determinó la concentración de las muestras de ADN por espectrofotometría con un multilector de absorbancia EPOCH™ (BioTek).

Evaluación de la integridad de la muestra

Se seleccionaron las muestras más adecuadas para la secuenciación NGS considerando como criterio una cuantificación del ADN ≥ 30 ng/ μ l, una ratio de pureza 260/280 ≥ 1.6 y una gráfica dentro de una absorbancia 10 mm hasta de 0.9 en una longitud de onda de 260 nm obtenido con Nanodrop 2000 (Thermo Scientific™). Además, se realizó electroforesis en gel de agarosa al 0.7% en buffer TBE 1X, visualizando el ADN con GelRed® y aplicando 100V durante 40-60 minutos.

Construcción de librerías, secuenciación

Para la construcción de librerías, se utilizó kit de preparación de ADN enriquecido (Illumina). Se utilizaron ≥ 50 ng de ADN por muestra para obtener librerías enriquecidas e indexadas. Estas fueron desnaturalizadas e introducidas en el secuenciador NextSeq 550 (NB552767), ubicado en las instalaciones del Laboratorio de Genómica del INDICASAT.AIP (Instituto de Investigaciones Científicas y Servicios de Alta Tecnología), Ciudad de Panamá. Y los datos generados se almacenaron en la plataforma BaseSpace® Sequence Hub de Illumina.

Análisis de los datos de secuenciación

El análisis primario de los datos se realizó con el software Sequencing Analysis Viewer, evaluando los parámetros de calidad de la corrida. El análisis secundario se llevó a cabo con DRAGEN, una herramienta en la nube de BaseSpace, que permitió ensamblar las lecturas, filtrarlas por calidad y alinearlas con un genoma de referencia a partir de archivos FastQ. Finalmente, en el análisis terciario, se utilizaron los archivos VCF obtenidos para identificar y clasificar variantes mediante la plataforma Franklin (Genoox). En la tabla 1, se presenta el listado de genes incluido en el panel virtual para el análisis de los datos se secuenciación.

Tabla 1. Listado de genes incluidos en panel virtual para determinar variantes genéticas relacionadas a Enfermedades neurodegenerativas en muestra de estudio.

Enfermedad neurodegenerativa	Genes incluidos	Referencias
Enfermedad de Alzheimer	APP, PSEN1, PSEN2, APOE, CLU, PICALM, CR1, BIN1, CD33, MS4A4A, MS4A6E, CD2AP, EPHA1, ABCA7, CASS4, CELF1, FERMT2, INPP5D, MEF2C, NME8, PTK2B, SLC24A4, RIN3, SORL1, ZCWPW1, TREM2, ADAM10, PVRL2, ABI3, PLCG2, CSF1R, TP53INP1, GRN, TOMM40, PNRP, MAPT, AP4E1, AP4M1, EIF2B2, PFN1, CYB561, TOMM40, WWOX/MAF, SCIMP/RABEP1, MAPT, HLA-DRB1, PSMB8, IQCK, ADAMTS4, HESX1, CLNK, OARD1, CNTNAP2, SPI1, KAT8, SCIMP, CYB561.	(Gaiteri et al., 2016; Karran et al., 2011; Lambert et al., 2013)
Enfermedad de Parkinson	VPS35, SNCA, LRRK2, PRKRA, GBA, RAB39B, TMEM230, GCH1, VPS13C, PARK2, PINK1, PARK7, ATP13A2, PLA2G6, FBXO7, SYNJ1, DNAJC6, SCARB2, CHCHD2, POLG, TAF1, EIF4G1, GIGYF2, HTRA2, SLC18A2, OPA1, TMEM175, SCN3A, VAPB, TH, C9orf72, SNCAIP, TMEM175, LRRN3, PRKN, DNAJC13, UCHL1, PODXL, PTRHD1.	(Healy et al., 2008; Klein & Westenberger, 2012; Lesage & Brice, 2009; Paisá N-Ruiz et al., 2004)
Esclerosis Lateral Amiotrófica (ELA)	SOD1, TARDBP, FUS, PFN1, SQSTM1, ANG, DCTN1, FIG4, VAPB, MAPT, TARDBP, C9orf72, KIF5A, PONI, SPG11.	(Chiò et al., 2018; Don W. Cleveland, 2001; Renton et al., 2014; Van Deerlin et al., 2008)
Distonia	TOR1A, THAP1, APTX, ATM, PRRT2, ANO3, TH, ATP1A3, TAF1, SLC39A14, SLC30A10, SPR, SLC6A3.	(Albanese et al., 2019; Fuchs et al., 2013; Jinnah & Factor, 2015)
Epilepsia	SCN1A, PRRT2, SCN8A.	(Harrow & Jobe, 2007; Steinlein, 2001)

Consideraciones éticas

Las muestras que se analizaron pertenecen a participantes reclutados en la cohorte comunitaria de PARI-HD, protocolo denominado "Estudio de cohortes prospectivo para evaluar el valor diagnóstico y pronóstico de biomarcadores en sangre para el deterioro cognitivo leve y la progresión a la enfermedad de Alzheimer".

La cual cuenta con la aprobación del comité Institucional de Ética e investigación de la Caja de Seguro Social, Pol. Manuel Ferrer Valdés (CIEI-CSS-P-083-16). Este protocolo también está registrado en RESEGIS (código 35).

Se ejecutó un análisis descriptivo calculando las frecuencias y porcentajes para las variables categóricas y medias y desviación estándar para las variables continuas. Se realizó un análisis descriptivo de variantes genéticas encontradas en la muestra analizada, incluyendo el reporte de la frecuencia de presentación de variantes repetidas en la muestra, con su intervalo de confianza al 95%.

RESULTADOS

Características demográficas y clínicas de la muestra de estudio.

Las características demográficas y clínicas de la muestra se resumen en la Tabla 2. Los participantes eran en su mayoría mujeres (70.5%) con una edad media de 70.3 años (DE = 7.4). Los participantes completaron 15.3 años de educación formal (DE = 4.2), y la mayoría de la muestra obtuvo un título universitario (52.3%). Las mediciones del índice de masa corporal (IMC) (M = 27,5 kg/m², DE = 3.6) indicaron que la mayoría de participantes tenía sobrepeso (79.5%). Los participantes declararon padecer al menos una enfermedad crónica en 84% de los casos, de las cuales las más comunes eran la hipertensión (52.3%), la osteoporosis (29.5%) y diabetes (22.7%). Además, los participantes declararon tener dificultades para dormir (45.5%), deficiencias visuales (29.5%), tabaquismo pasado o actual (34.1%) y polifarmacia (22.7%). Una proporción menor de participantes cumplía los criterios de depresión geriátrica (15.9%) según el GDS-15

(Escala de Depresión Geriátrica de 15 ítem), y la prueba MMSE (Mini Prueba del Estado Mental), una evaluación neuropsicológica que emplea un puntaje de corte de 25 para determinar la presencia de deterioro cognitivo global, esta reveló que ninguno de los participantes presentaba deterioro, ya que todos obtuvieron puntajes superiores a 25. Sin embargo, muchos de los participantes (43.2%) reportaron quejas subjetivas de memoria.

Tabla 2. Características demográficas y clínicas de la muestra de estudio.

		Total (N = 44) n (%) / M (DS)
Edad (años)		70.3 (7.4)
	Rango	60.2-91.8
Sexo, mujer		31 (70.5)
Status marital		
	Sin pareja	19 (43.2)
	Con pareja	25 (56.8)
Años de educación		15.3 (4.2)
	Rango	6-25
Máximo nivel de educación alcanzado		
	Primaria	7 (15.9)
	Secundaria	14 (31.8)
	Universidad	23 (52.3)
Ingreso mensual (dólares)		
	< 500	5 (11.4)
	501-1200	15 (34.1)
	>1200	24 (54.5)
Índice de masa corporal (IMC)		27.5 (3.6)
	Rango	19.9-36.4
Categorías de IMC		
	Bajo (<18.5)	0 (0)
	Normal (18.5-24.9)	8 (18.2)
	Sobrepeso (>25)	35 (79.5)
Enfermedades crónicas		
	0	7 (15.9)
	1-4	35 (79.4)
	>5	2 (4.5)
Condiciones crónicas (Si)		
	Diabetes	10 (22.7)
	Hipertensión	23 (52.3)
	Asma	6 (13.6)
	Cáncer	5 (11.4)
	Enfermedad cardiovascular	6 (13.6)
	Enfermedad cerebrovascular	3 (6.8)
	Enfermedad pulmonar	5 (11.4)
	Enfermedad renal	2 (4.5)
	Artritis	5 (11.4)
	Osteoporosis	13 (29.5)
Deterioro visual (Si)		13 (29.5)
Deterioro auditivo (Si)		4 (9.1)
Dificultad para dormir (Si)		20 (45.5)
Consumo de tabaco (Si)		15 (34.1)
Presión sistólica (mmHg)		144.3 (24.9)
Polifarmacia (Si)		10 (22.7)
Síntomas depresivos según criterio DGS-15 (depresión, Si)		7 (15.9)
Caidas en el último año (Si)		12 (27.3)
Caidas recurrentes en el último año		
	1	6 (13.6)
	≥2	6 (13.6)
Deterioro cognitivo subjetivo (Si)		19 (43.2)
Deterioro cognitivo objetivo (puntaje MMSE)		27.6 (1.4)
	Rango	25-30

Resultados de la secuenciación y análisis primario

Se generaron 149.4 Gbp de información cumpliendo con todas las métricas de calidad, incluyendo controles de calidad de las lecturas por línea y la celda de flujo. Las métricas de lectura registradas se presentan en la **tabla 3**.

Tabla 3. Métrica de lectura de secuenciación masiva ejecutada.

Lecturas	Ciclos	Rendimiento (Gbp)	Rendimiento del proyecto (Gbp)	Alineado (%)	Tasa de error (%)	Ciclo de intensidad 1	% > Q30
Lectura 1	149	70.4	70.4	0.9	0.5	8816.4	88.0
Lectura 2	10	4.3	4.3	0	0	7612.4	95.8
Lectura 3	10	4.3	4.3	0	0	6555.6	94.9
Lectura 4	149	70.4	70.4	0.9	0.5	8394.8	87.3
Total de lecturas no indexadas	298	140.9	140.9	0.9	0.5	8605.6	87.7
Total	318	149.4	149.9	0.9	0.5	7844.8	88.1

El análisis primario demostró que el 93.0% de las lecturas fueron identificadas por índices y el 6.9% representó el porcentaje de conglomerados que pasaron el filtro y no pudieron ser asignados a un índice. El coeficiente de variación para el número de recuentos en todos los índices del carril 1 fue de 0.6 y los valores mínimos y máximos de representación para cualquier índice fue de 0.6 y 5.0, respectivamente. Las métricas de calidad de la indexación para los carriles 2, 3 y 4, son similares a los presentados con el carril 1 y donde se confirma que todas las muestras indexadas estén correctamente representadas. Las puntuaciones de calidad de la secuenciación miden la probabilidad de que se llame incorrectamente a una base. La gráfica Q-Score (Figura 1) evalúa la distribución del número de bases por puntuación de calidad de la corrida acumulativa que fue de 132.9 G, representando el 87.8%. Solo se incluyen las bases de las lecturas que pasan el filtro de calidad.

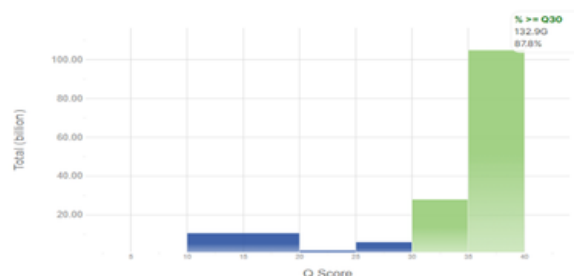


Figura 1. Gráfica de puntuación de calidad. El mapa de calor Q-Score proporciona una visión general de las puntuaciones de calidad en todos los ciclos donde para los primeros 100 ciclos en ambas superficies se alcanza no menos del 80% del puntaje de calidad. Los datos por carril permiten ver métricas de calidad por carril, la densidad de los clústeres ejecutados fue similar, con un promedio total de los 4 carriles en 205.5 ± 6.3 K/mm². Los resultados sobre la densidad de clústeres obtenida se mantienen en el rango recomendado por la casa comercial que es de 170–220 K/mm² utilizando el equipo NetSeq. Se determinó la densidad óptima de clústeres para examinar la diferencia en las métricas de calidad entre carriles. En la figura 2 se muestra que para todos los carriles la densidad de los clústeres ejecutados fue similar, con un promedio total de los 4 carriles en 205.5 ± 6.3 K/mm². Los resultados sobre la densidad de clústeres obtenida se mantienen en el rango recomendado por el fabricante que es de 170–220 K/mm² utilizando el equipo NetSeq.

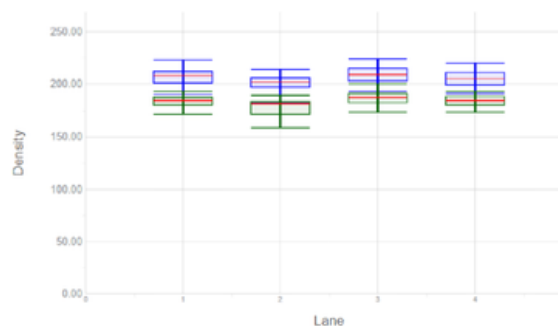


Figura 2. Gráfica de datos por carril. Se presentan los resultados de calidad por carril sobre la densidad de clústeres alcanzada en la corrida de secuenciación.

Resultados de la secuenciación y análisis secundario

El análisis secundario consistió en ensamblar las lecturas obtenidas durante la secuenciación con el genoma humano de referencia Homo sapiens (1000 Genomes hg38-altmaskedv3-graph-enabled) utilizando el software DRAGEN. Mediante este análisis las lecturas se evaluaron y filtraron según su calidad. Hay cuatro grupos diferentes de métricas que se generan en diferentes etapas del análisis: Métricas de enriquecimiento, Mapeo y alineación, VCF, Métricas de duración y Métricas e informes de enriquecimiento. En promedio, el 99.3% de las lecturas obtenidas pasaron el filtro de alineación con el genoma de referencia, por tanto, el 73.4% de las lecturas resultaron enriquecidas con el alineamiento.

Asimismo, a nivel de bases, el 99.3% de los nucleótidos identificados durante la secuenciación se alinearon con el genoma de referencia. La longitud mediana del promedio fragmento secuenciados fue de 206.7 ± 84.4 , con un rango de 33.1-628.0.

Sin embargo, solo se alcanzó un porcentaje de cobertura con una profundidad de 20X del 30.1%. con respecto a las variantes, se identificaron en promedio 13,827 variantes de un solo nucleótido y 1,166 inserciones/delecciones (indels).

Resultados de la secuenciación y análisis terciario

Para el análisis terciario, utilizando la plataforma en línea Franklin, se utilizaron los archivos VCF. Se filtraron los resultados mediante la aplicación de un panel virtual personalizado que incluyó 138 genes. En un total de 44 muestras incluidas en el análisis,

se identificaron 294 variaciones genéticas que han sido vinculadas a enfermedades neurodegenerativas. De estas variables genéticas, 217(73.9%) son benignas, 61(20.7%) variantes de significado incierto (VUS), 10(3.4%) variantes como probablemente patogénicas y 6(2.0%) como patogénica (**Figura 3**).

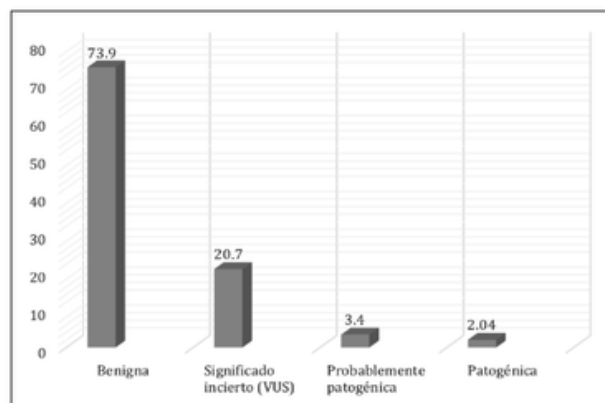


Figura 3. Porcentaje de variantes asociadas a enfermedades neurodegenerativas.

En la tabla 4 se describen las variantes asociadas a cada enfermedad. Para la enfermedad de Alzheimer, se encontraron un total de 4 variantes distribuidas en los genes CR1 y TREM2. De estas, 1 variante fue clasificada como benigna (25%) y correspondió a CR1 (c.6151A>G), con una frecuencia del 9.1%. Las otras 3 variantes (75%) fueron clasificadas como VUS (variantes de significado incierto): CR1 (c.1333A>G), CR1 (c.1253-9_1253-8insG) y TREM2 (c.469C>T), cada una con una frecuencia del 9.1%. Esto refleja que la mayoría de las variantes asociadas a Alzheimer en esta muestra son inciertas respecto a su patogenicidad.

En las variantes relacionadas con ELA, se identificó una única variante en el gen NEFH (c.1973_1978del), clasificada como VUS (100%), con una frecuencia del 9.1%. Esto indica que no se detectaron variantes benignas o patogénicas claras en esta muestra para ELA. En las variantes asociadas a Esclerosis Múltiple (EM), se identificó una única variante en el gen HLA-DRB1 (c.41C>T), clasificada como VUS (100%), con una frecuencia del 13.6%.

Este hallazgo sugiere que no se encontraron variantes benignas ni patogénicas conocidas relacionadas con EM en esta muestra. En las variantes asociadas a Demencia Frontotemporal, se identificó una variante en el gen GRN (c.1973_1978del). Esta variante fue clasificada como VUS (variante de significado incierto), con una frecuencia del 9.1% (4 de 44 muestras). Esto indica que no se detectaron variantes benignas o patogénicas claras asociadas con Demencia Frontotemporal en esta muestra. En las variantes asociadas a Enfermedad de Parkinson (EP) identificadas en las 44 muestras, se encontraron un total de tres variantes distribuidas en los genes PRKN, NEFH y GIGYF2. De estas, dos variantes fueron clasificadas como benignas (66.7%) y se localizaron en los genes PRKN (c.500G>A) y GIGYF2 (c.3626_3646del), cada una con una frecuencia de aparición del 9.1%. La variante restante, NEFH (c.1973_1978del), fue categorizada como una variante de significado incierto (VUS), representando el 33.3% de las variantes asociadas a EP, con una frecuencia del 9.1%. Estos hallazgos reflejan que la mayoría de las variantes relacionadas con Parkinson en esta muestra fueron benignas.

Tabla 4. Variantes genéticas que comparten los participantes de la muestra de estudio

Gen	Variante	Polimorfismo	Categorización de la variante	Enfermedad asociada	Frecuencia de aparición en la muestra	Intervalo de confianza al 95%
HLA-D RB1	C.41C>T	rs9270302	VUS	Esclerosis múltiple	13.6 (6/44)	5.2-27.1
CR1	c.1333A>G	rs3818565	VUS	Enfermedad de Alzheimer (EA)	11.4 (5/44)	4.5-24.4
TREM2	c.469C>T	rs2234255	VUS	EA	9.1 (4/44)	3.0-21.7
CR1	c.1253-9_1253	Rs17047661	VUS	EA	9.1 (4/44)	3.0-21.7
GIGYF2	c.3626_3646del	rs527464858	Benigna	Enfermedad de Parkinson (EP)	9.1(4/44)	2.9-20.3
CR1	c.6151A>G	rs149099494	Benigna	EA	9.1(4/44)	2.9-20.3
NEFH	c.1973_1978del	rs149571560	VUS	EP	9.1(4/44)	2.9-20.3
PRKN	c.500G>A	Rs1801474	Benigna	EP	9.1(4/44)	2.9-20.3

Discusión

El presente estudio utilizó secuenciación masiva (NGS) por exoma clínico para identificar variantes genéticas asociadas con enfermedades neurodegenerativas en muestras de ADN de personas ≥ 60 años de Panamá.

Sobre el rendimiento de la corrida de secuenciación y análisis primario podemos resaltar que se generaron 149.4 Gbp de datos en un total de 318 ciclos. El porcentaje de bases con Q30 fue del 87.8%, un valor que supera el umbral mínimo de calidad para proyectos de secuenciación poblacional y asegura una baja tasa de errores en las lecturas. Estudios previos han demostrado que porcentajes de Q30 por encima del 80% son fundamentales para garantizar resultados reproducibles en análisis genómicos de alto rendimiento. Por ejemplo, se ha reportado valores de Q30 superiores al 85% en análisis realizados con secuenciadores Illumina (Goodwin et al., 2016), destacando su importancia para estudios de poblaciones humanas y animales. Asimismo, se ha documentado que tecnologías como Illumina HiSeq 2000 y MiSeq presentan Q30 consistentes alrededor del 90% en corridas de hasta 300 ciclos, mientras que otros métodos, como Ion Torrent, tienden a mostrar una caída en calidad en lecturas largas (Quail et al., 2012).

El mapa de calor mostró ligeras disminuciones en la calidad de las bases ($>80\%$ Q30) en ciclos posteriores, lo cual es consistente con lo reportado en la literatura debido a errores acumulados en lecturas largas. Según algunos autores (Quail et al., 2008) en plataformas Illumina como HiSeq 2000 y MiSeq, la calidad de las bases tiende a decrecer en los últimos ciclos, con porcentajes de Q30 que caen de un promedio de 85%-90%

al inicio de la corrida a valores de 75%-80% hacia el final, particularmente en lecturas mayores a 250 pb. Además (Fleige & Pfaff, 2006) señalaron que estas disminuciones en calidad pueden impactar la precisión en la identificación de variantes genómicas complejas, como las inserciones y deleciones en regiones de baja complejidad. Sin embargo (Glenn, 2011) argumenta que, a pesar de estas limitaciones, los valores de Q30 >80% son aceptables para la mayoría de los análisis genómicos, siempre y cuando se utilicen herramientas de corrección bioinformática para mitigar errores acumulativos.

La densidad promedio de clústeres alcanzada (205.5 K/mm²) estuvo dentro del rango óptimo (170–220 K/mm²) recomendado por el fabricante del equipo NextSeq 550, garantizando lecturas eficientes y una distribución uniforme de las muestras en la celda de flujo (Bentley et al., 2008). Johnson y colaboradores (Johnson et al., 2019) analizaron densidades de clústeres en proyectos de microbioma humano (n = 300 muestras) utilizando secuenciadores NovaSeq y encontraron que densidades mayores a 230 K/mm² aumentaron los índices de error en regiones repetitivas, especialmente en amplicones largos (>500 pb).

En contraste, valores dentro del rango óptimo (170–220 K/mm²) aseguraron una cobertura uniforme de las regiones génicas y una tasa de éxito superior al 95% en la identificación de variantes. Sobre el análisis secundario podemos destacar que el porcentaje de lecturas alineadas superó el 99% en todas las muestras, reflejando tanto la calidad de las lecturas como la precisión del genoma de referencia empleado, el cual fue GRCh38, en donde la última versión del genoma de

referencia empleado, el cual fue GRCh38, en donde la última versión del genoma de referencia humano fue publicada por el Genome Reference Consortium en 2013, un esfuerzo conjunto entre el National Center for Biotechnology Information (NCBI), el European Bioinformatics Institute (EBI) y el Wellcome Trust Sanger Institute (Genome Reference Consortium, 2013). Este valor está en el rango superior de calidad esperada para estudios de secuenciación de alto rendimiento, donde típicamente se reportan valores >95% (Sims et al., 2014). El enriquecimiento de lecturas dirigidas osciló entre 70% y 84%, lo que indica una alta especificidad del diseño experimental para capturar las regiones objetivo (Gnirke et al., 2009).

Sobre los hallazgos genéticos y las variantes de relevancia identificadas podemos mencionar el gen HLA-DRB1: Variantes como c.41C>T con polimorfismo rs9270302 y c.640_641insTG con polimorfismo rs35837054, asociadas a esclerosis múltiple (EM), fueron recurrentes en el análisis. Este hallazgo es consistente con estudios previos que identifican a HLA-DRB1 como uno de los principales loci genéticos asociados con predisposición a EM (Sawcer et al., 2014). Para el gen NOTCH3 (c.3058G>T) con polimorfismo rs35769976, el cual ha sido relacionado con enfermedades cerebrovasculares; su presencia en este análisis refuerza su relevancia clínica y concuerda con hallazgos reportados anteriormente (Joutel et al., 1996). En el gen CR1 (c.1333A>G) con polimorfismo rs3818565: Clasificada como VUS, está relacionada con la enfermedad de Alzheimer en la literatura. Sin embargo, estudios previos señalan que su impacto puede variar dependiendo del contexto genético y ambiental, subrayando la necesidad de análisis adicionales (Lambert et al., 2013). A pesar de la concordancia general con la literatura, se identificaron ciertos puntos que merecen atención. Por un lado, algunas variantes consideradas benignas en este análisis, como ABCA7,

han sido reportadas como factores de riesgo en Alzheimer. Esto podría deberse a diferencias en la población analizada o a limitaciones en las bases de datos utilizadas. Estudios han explorado cómo la pérdida de función de ABCA7 afecta la inflamación microglial y las propiedades mitocondriales, exacerbando el daño neuronal (Wang et al., 2025). Esto refuerza la importancia del transporte de lípidos y la función microglial en la patogénesis de esta enfermedad (Bellenguez et al., 2022). Adicionalmente, un meta-análisis reciente demostró que variantes específicas de ABCA7 están significativamente asociadas con el riesgo de Alzheimer en diferentes poblaciones, aunque algunos polimorfismos muestran variabilidad en función de la etnicidad y otros factores genéticos y ambientales (Dib et al., 2021).

Las variantes que son consideradas benignas podrían tener diferentes efectos en las poblaciones de América Latina en comparación con otras poblaciones estudiadas en la literatura. Es posible que no sean de riesgo en esta población en particular debido a diferencias genéticas o ambientales. Por tanto, las herramientas utilizadas para clasificar las variantes podrían no incluir suficiente información sobre las variantes específicas en poblaciones latinoamericanas. Esto podría llevar a que algunas variantes de riesgo pasen inadvertidas o se clasifiquen como benignas (De Oliveira et al., 2023). Por otro lado, la cobertura en ciertas regiones (<75%) sugiere ineficiencias en la captura de objetivos específicos, un desafío técnico recurrente en experimentos de secuenciación exómica. Esto significa que estas regiones no fueron suficientemente representadas en el análisis, lo que podría dificultar la identificación precisa de variantes en estas áreas. Una causa probable son las regiones difíciles de secuenciar ya que algunas regiones del genoma,

como aquellas con un alto contenido de bases guanina (G), citosina (C) o regiones repetitivas, son conocidas por ser más difíciles de capturar y amplificar durante la preparación de librerías. Esto es un desafío técnico común en estudios de exoma (Clark et al., 2011)

Este estudio presenta varias limitaciones. Primero, un tamaño de muestra limitado para establecer asociaciones estadísticamente significativas. Segundo, muestra de estudio fue obtenida de participantes reclutados por PARI-HD en Panamá, lo que podría afectar la representatividad de la población y restringir la generalización de los hallazgos. Tercero, el enfoque metodológico aplicado por exoma, excluye la identificación de variantes regulatorias e intrónicas que pueden ser relevantes en ciertas patologías. Por último, la base de datos utilizada para la interpretación de los datos de la secuenciación podría estar subrepresentadas para variantes específicas de poblaciones latinoamericanas.

Conclusión

Este estudio demuestra la aplicabilidad de la secuenciación masiva utilizando exoma clínico para la identificación de variantes genéticas asociadas con enfermedades neurodegenerativas, destacando hallazgos relevantes y congruentes con la literatura. Sin embargo, las limitaciones técnicas y poblacionales resaltan la necesidad de investigaciones futuras que optimicen la sensibilidad de la metodología y amplíen la interpretación de las variantes en contextos poblacionales diversos.

Agradecimientos

Al Instituto de Investigaciones Científicas y Servicios de Alta Tecnología de Panamá INDICASAT AIP.

Conflicto de Interés

Los autores declaran no tener algún conflicto de interés.

Referencias

1. Albanese, A., Di Giovanni, M., & Lalli, S. (2019). Dystonia: Diagnosis and Management. *European Journal of Neurology*, 26(1), 5–17. <https://doi.org/10.1111/ene.13762>
2. Alvarez-Mora, M. I., Rodríguez-Revenga, L., Jodar, M., Potrony, M., Sanchez, A., Badenas, C., Oriola, J., Villanueva-Cañas, J. L., Muñoz, E., Valldeoriola, F., Cámara, A., Compta, Y., Carreño, M., Martí, M. J., Sánchez-Valle, R., & Madrigal, I. (2023). Implementation of Exome Sequencing in Clinical Practice for Neurological Disorders. *Genes*, 14(4). <https://doi.org/10.3390/genes14040813>
3. Alzheimer's Association. (2023, June). What is Alzheimer's Disease? *Alzheimer's Association*. <https://www.alz.org/alzheimers-dementia/what-is-alzheimers?lang=en-us>
4. Bellenguez, C., Küçükali, F., Jansen, I. E., Kleindam, 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. (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>
5. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boultell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., ... Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), 53–59. <https://doi.org/10.1038/nature07517>
6. Bras, J., Guerreiro, R., & Hardy, J. (2012). Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nature Reviews Neuroscience*, 13(7), 453–464. <https://doi.org/10.1038/nrn3271>
7. Chiò, A., Mazzini, L., D'Alfonso, S., Corrado, L., Canosa, A., Moglia, C., Manera, U., Bersano, E., Brunetti, M., Barberis, M., Veldink, J. H., Van Den Berg, L. H., Pearce, N., Sproviero, W., McLaughlin, R., Vajda, A., Hardiman, O., Rooney, J., Mora, G., ... Al-Chalabi, A. (2018). The multistep hypothesis of ALS revisited. *Neurology*, 91(7). <https://doi.org/10.1212/WNL.0000000000005996>
8. Clark, M. J., Chen, R., Lam, H. Y. K., Karczewski, K. J., Chen, R., Euskirchen, G., Butte, A. J., & Snyder, M. (2011). Performance comparison of exome DNA sequencing technologies. *Nature Biotechnology*, 29(10), 908–916. <https://doi.org/10.1038/nbt.1975>
9. De Oliveira, T. C., Secolin, R., & Lopes-Cendes, I. (2023). A review of ancestry and admixture in Latin America and the Caribbean focusing on native American and African descendant populations. *Frontiers in Genetics*, 14. <https://doi.org/10.3389/fgene.2023.1091269>
10. Dib, S., Pahnke, J., & Gosselet, F. (2021). Role of ABCA7 in Human Health and in Alzheimer's Disease. *International Journal of Molecular Sciences*, 22(9), 4603. <https://doi.org/10.3390/ijms22094603>
11. Don W. Cleveland, J. D. R. (2001). From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. *Nature reviews. Neuroscience*, 2(11), 806–819. <https://doi.org/10.1038/35097565>
12. Fleige, S., & Pfaffl, M. W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine*, 27(2–3), 126–139. <https://doi.org/10.1016/j.mam.2005.12.003>
13. Fuchs, T., Saunders-Pullman, R., Masuho, I., Luciano, M. S., Raymond, D., Factor, S., Lang, A. E., Liang, T. W., Troesch, R. M., White, S., Ainehsazan, E., Hervé, D., Sharma, N., Ehrlich, M. E., Martemyanov, K. A., Bressman, S. B., & Ozelius, L. J. (2013). Mutations in GNAL cause primary torsion dystonia. *Nature Genetics*, 45(1), 88–92. <https://doi.org/10.1038/ng.2496>
14. Gaiteri, C., Mostafavi, S., Honey, C. J., De Jager, P. L., & Bennett, D. A. (2016). Genetic variants in Alzheimer disease-molecular and brain network approaches. *Nature Reviews Neuroscience*, 12(7), 413–427. <https://doi.org/10.1038/nrneurol.2016.84>
15. Genome Reference Consortium. (2013). GRCh38: Genome Reference Consortium Human Build 38. *National Center for Biotechnology Information (NCBI)*. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/
16. Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 11(5), 759–769. <https://doi.org/10.1111/j.1755-0998.2011.03024.x>
17. Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., & Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, 27(2), 182–189. <https://doi.org/10.1038/nbt.1523>
18. Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>


- 19.Harrow, M., & Jobe, T. H. (2007). Factors involved in outcome and recovery in Schizophrenia patients not on antipsychotic medications: A 15-year multifollow-up study. *Journal of Nervous and Mental Disease*, 195(5), 406–414.
<https://doi.org/10.1097/01.nmd.0000253783.32338.6e>
- 20.Healy, G. N., Dunstan, D. W., Salmon, J. O., Cerin, E., Shaw, J. E., Zimmet, P. Z., & Owen, N. (2008). Breaks in sedentary time beneficial associations with metabolic risk. *Diabetes care*, 31(4) 661-666.
<https://doi.org/10.2337/dc07>
- 21.Jinnah, H. A., & Factor, S. A. (2015). Diagnosis and treatment of Dystonia. *Neurologic Clinics*, 33(1), 77–100.
<https://doi.org/10.1016/j.ncl.2014.09.002>
- 22.Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., Sodergren, E., & Weinstock, G. M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, 10(1).
<https://doi.org/10.1038/s41467-019-13036-1>
- 23.Joutel, A., Corpechot, C., Ducros, A., & Vahedi, K. (1996). Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature*, 383(6602), 707-710. <https://doi.org/10.1038/383707a0>
- 24.Karran, E., Mercken, M., & Strooper, B. De. (2011). The amyloid cascade hypothesis for Alzheimer's disease: An appraisal for the development of therapeutics. *Nature Reviews Drug Discovery*, 10(9), 698–712.
<https://doi.org/10.1038/nrd3505>
- 25.Klein, C., & Westenberger, A. (2012). Genetics of Parkinson's disease. *Cold Spring Harbor Perspectives in Medicine*, 2(1).
<https://doi.org/10.1101/cshperspect.a008888>
- 26.Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A. L., Bis, J. C., Beecham, G. W., Grenier-Boley, B., Russo, G., Thornton-Wells, T. A., Jones, N., Smith, A. V., Chouraki, V., Thomas, C., Ikram, M. A., Zelenika, D., ... Seshadri, S. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics*, 45(12), 1452–1458.
<https://doi.org/10.1038/ng.2802>
- 27.Lesage, S., & Brice, A. (2009). Parkinson's disease: From monogenic forms to genetic susceptibility factors. *Human Molecular Genetics*, 18(R1).
<https://doi.org/10.1093/hmg/ddp012>
- 28.Organización Mundial de la Salud. (2020). Global Health Estimates 2020: Deaths by cause, age, sex, by country and by region, 2000–2019. <https://www.who.int/data/global-health-estimates>.
- 29.Paisá N-Ruíz, C., Jain, S., Evans, E. W., Gilks, W. P., Simó, J., Van Der Brug, M., Ló Pez De Munain, A., Aparicio, S., Gil, A. M., Khan, N., Johnson, J., Ruiz Martinez, J., & Nicholl, D. (2004). Cloning of the gene containing mutations that cause PARK8 linked Parkinson's Disease. *Neuron* (44) 595–600.
<https://doi.org/10.1016/j.neuron.2004.10.023>
- 30.Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., Swerdlow, H., & Turner, D. J. (2008). A large genome center's improvement to the Illumina sequencing system. *Nature Methods*, 5(12), 1005–1010. <https://doi.org/10.1038/nmeth.1270>
- 31.Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13(1).
<https://doi.org/10.1186/1471-2164-13-341>
- 32.Renton, A. E., Chiò, A., & Traynor, B. J. (2014). State of play in Amyotrophic Lateral Sclerosis genetics. *Nature Neuroscience*, 17(1), 17–23.
<https://doi.org/10.1038/nn.3584>
- 33.Sawcer, S., Franklin, R. J. M., & Ban, M. (2014). Multiple Sclerosis genetics. *The Lancet Neurology*, 13(7), 700–709. [https://doi.org/10.1016/S1474-4422\(14\)70041-9](https://doi.org/10.1016/S1474-4422(14)70041-9)
- 34.Sims, D., Sudbery, I., Illott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), 121–132.
<https://doi.org/10.1038/nrg3642>
- 35.Steinlein, O. K. (2001). Ion channels and Epilepsy. *American Journal of Medical Genetics - Seminars in Medical Genetics*, 106(2), 146–159.
<https://doi.org/10.1002/ajmg.1582>
- 36.Van Deerlin, V. M., Leverenz, J. B., Bekris, L. M., Bird, T. D., Yuan, W., Elman, L. B., Clay, D., Wood, E. M. C., Chen-Plotkin, A. S., Martinez-Lage, M., Steinbart, E., McCluskey, L., Grossman, M., Neumann, M., Wu, I. L., Yang, W. S., Kalb, R., Galasko, D. R., Montine, T. J., ... Yu, C. E. (2008). TARDBP mutations in Amyotrophic Lateral Sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *The Lancet Neurology*, 7(5), 409–416.
[https://doi.org/10.1016/S1474-4422\(08\)70071-1](https://doi.org/10.1016/S1474-4422(08)70071-1)

37.Wang, N., Pan, Y., Starling, S. C., Haskell, D. H., Quintero, A. C., Kawatani, K., Inoue, Y., Shue, F., Ma, X., Aikawa, T., Martens, Y. A., Kurti, A., Parsons, T. M., Perkerson, R. B., Roy, B., Raulin, A. C., Ren, Y., DeTure, M., Dickson, D. W., ... Kanekiyo, T. (2025). Neuronal ABCA7 deficiency aggravates mitochondrial dysfunction and neurodegeneration in Alzheimer's disease. *Alzheimer's and Dementia*, 21(3). <https://doi.org/10.1002/alz.70112>


¿Sabías que...?

Variantes genéticas relacionadas a enfermedades neurodegenerativas

Basado en un estudio de 44 muestras de ADN de adultos mayores en Panamá.



78% de las variantes encontradas son benignas



Sólo el **5.4%** resultaron potencialmente patogénicas



Las variantes encontradas incluyeron genes relacionados con Alzheimer, Parkinson, ELA y otras condiciones

La secuenciación masiva permite entender los riesgos genéticos de nuestra población

Fuente: Determinación de variaciones genéticas relacionadas a enfermedades neurodegenerativas mediante secuenciación masiva por exoma clínico de personas mayores de Panamá.



COLEGIO NACIONAL DE LABORATORISTAS CLÍNICOS DE PANAMÁ

+ (507) 225-3508 + (507) 6264-4893 info@conalac.com.pa

UBICACIÓN TEMPORAL:
BELLA VISTA CALLE 43, PH COLORES DE BELLA VISTA,
PISO 17 OFICINA. A-B, PANAMÁ REPÚBLICA DE PANAMÁ.