

INTRODUCCIÓN AL FILTRADO, ANÁLISIS Y CURACIÓN DE VARIANTES GENÉTICAS EN PACIENTES CON DISCAPACIDAD INTELECTUAL

JIMENA BARRAZA GARCÍA, CRISTINA CANO MORATILLA, ALBERTO GONZÁLEZ DE LA VEGA

EurofinsMegalab. Madrid

Resumen Actualmente la secuenciación del exoma completo (WES; Whole-exome sequencing) mediante la técnica NGS (Next-generation sequencing) es uno de los estudios genéticos más solicitados dentro del abordaje de pacientes con Discapacidad Intelectual con o sin otras anomalías. Al igual que con otros procedimientos y estudios clínicos, es conveniente que los médicos prescriptores tengan una comprensión clara de los alcances y limitaciones del uso de WES, del proceso de análisis de las variantes genéticas identificadas, así como de aspectos a evaluar acerca de la calidad y estructura de los informes de los estudios de NGS, con el objetivo de que puedan interpretar mejor los resultados de un estudio y plantear de la mejor manera la correlación de los mismos con la clínica observada.

Palabras clave: discapacidad intelectual, secuenciación de exoma completo, curación de variantes

Abstract *Introduction to filtering, analysis and curation of genetic variants in patients with Intellectual Disability*

Currently, Whole exome sequencing (WES) using NGS (Next-generation sequencing) technology is one of the most requested genetic studies within the approach of patients with intellectual disability with or without other anomalies. As with other procedures and clinical studies, it is convenient for prescribing physicians to have a clear understanding of the scope and limitations of the use of WES, the analysis process of the genetic variants identified, as well as aspects to be evaluated regarding quality and structure of the reports of the NGS studies, with the aim that they can better interpret the results of a study, evaluate its quality, and propose in the best way the correlation of the same with the observed phenotype.

Key words: intellectual disability, whole exome sequencing, variant curation

La Discapacidad Intelectual (DI) es un trastorno que comienza durante el período de desarrollo que incluye déficits de funcionamiento tanto intelectuales como adaptativos en los dominios conceptual, social y práctico^{1, 2}. Se estima que del 1 al 3% de la población presenta algún grado de discapacidad intelectual¹⁻⁴. Por ello, es una de las causas más comunes de referencia a la consulta de Neurología y Pediatría.

La DI puede deberse a factores genéticos (monogénicos/cromosómicos), ambientales (infección intrauterina, enfermedad prenatal o postnatal, etc.), o una combinación de ambos^{3, 5}. En un gran porcentaje de casos la etiología es desconocida. Es posible identificar una causa genética para la presencia de DI en alrededor de la mitad de los casos⁴, y los estudios genéticos habitualmente solicitados comprenden el cariotipo, el *array* cromosómico (CMA;

Chromosomal microarray analysis), y los estudios basados en NGS (panel multigenes, secuenciación de exoma, y secuenciación de genoma).

La NGS de manera general se basa en la fragmentación del ADN obtenido del paciente, la captura de fragmentos de interés (si se trata de estudios dirigidos), la separación de esos fragmentos capturados, y su amplificación individual. Los fragmentos amplificados son secuenciados simultáneamente, y posteriormente ensamblados para su comparativa con un genoma de referencia³.

El estudio puede realizarse como un panel dirigido a DI (donde las sondas utilizadas para capturar dentro del DNA del paciente sólo cubren las regiones codificantes de genes previamente relacionados con DI), un exoma (que puede ser clínico, donde las sondas cubrirán solo las regiones codificantes de los genes previamente relacionados con patologías humanas; o completo, donde se cubrirán todos los genes conocidos, hayan sido relacionados o no con enfermedad), o un genoma (donde se secuencia todo el material genético, sea codificante o no).

Estos abordajes difieren en cuanto a región secuenciada, cobertura de la región, tipos de variantes estructurales susceptibles de análisis, coste, tiempo y dificultad de evaluación^{2, 3, 5, 6}.

Actualmente, el abordaje con NGS más utilizado en DI es el análisis de exoma completo (WES), dada la etiología altamente heterogénea de la DI, y el rápido descubrimiento de nuevos genes relacionados con este fenotipo.

En 2021, el Colegio Americano de Genética Médica y Genómica (ACMG) estableció la recomendación del uso de secuenciación de exoma y genoma como prueba de primer o segundo nivel para pacientes con retraso en el desarrollo y/o DI con inicio antes de los 18 años de edad⁷. Recientemente, algunos grupos multidisciplinarios proponen al WES como la prueba de primer nivel para la evaluación de la DI, tomando como base su tasa de diagnóstico consistentemente más alta que CMA^{4, 5}.

Se estima que la tasa diagnóstica de WES para DI con o sin otras anomalías oscila entre el 25 y el 40%, correspondiendo los porcentajes más altos a los abordajes en trío (es decir, incorporando muestra de ambos padres para el análisis)^{4, 7}.

A medida que el análisis de WES se convierte en un estudio frecuentemente solicitado por médicos no genetistas, se vuelve más relevante que tengan una comprensión clara del proceso de análisis de las variantes genéticas identificadas, de los alcances y limitaciones del uso de WES, así como de los aspectos a evaluar acerca de la calidad del estudio.

Proceso de análisis de un caso de DI mediante WES

Previamente al estudio, el médico prescriptor debe proporcionar al paciente información básica acerca del estudio solicitado, y los posibles resultados del mismo, incluyendo la posibilidad de hallazgos incidentales o secundarios, la potencial necesidad de realizar estudios genéticos o clínicos adicionales en el probando o en sus familiares a consecuencia del resultado obtenido, o las posibles implicaciones para otros miembros de la familia. Todas estas cuestiones deben plasmarse en un consentimiento informado, que debe enviarse junto con la muestra o muestras del caso a analizar⁷.

Una historia clínica lo más detallada posible es el punto de partida para la evaluación de un caso por parte de un laboratorio de Genética: ha sido ampliamente demostrado que el fenotipado detallado incrementa la tasa diagnóstica en los estudios de WES, y asimismo, si la información clínica facilitada no es correcta o está incompleta se podría producir una interpretación errónea de los resultados. La selección de los datos clínicos relevantes puede realizarse utilizando términos de HPO (Human Phenotype Ontology) especialmente los casos en

que las herramientas de análisis de variantes lo soliciten para incluirlo en sus algoritmos de filtrado^{2, 6, 8}.

En un estudio de exoma completo, puede iniciarse incorporando paneles virtuales más o menos extensos para luego considerar el panel virtual completo de genes clínicos (exoma clínico) y posteriormente el resto de datos². La complejidad irá en aumento si se debe realizar análisis de variantes de número de copias (CNVs) además de variantes puntuales.

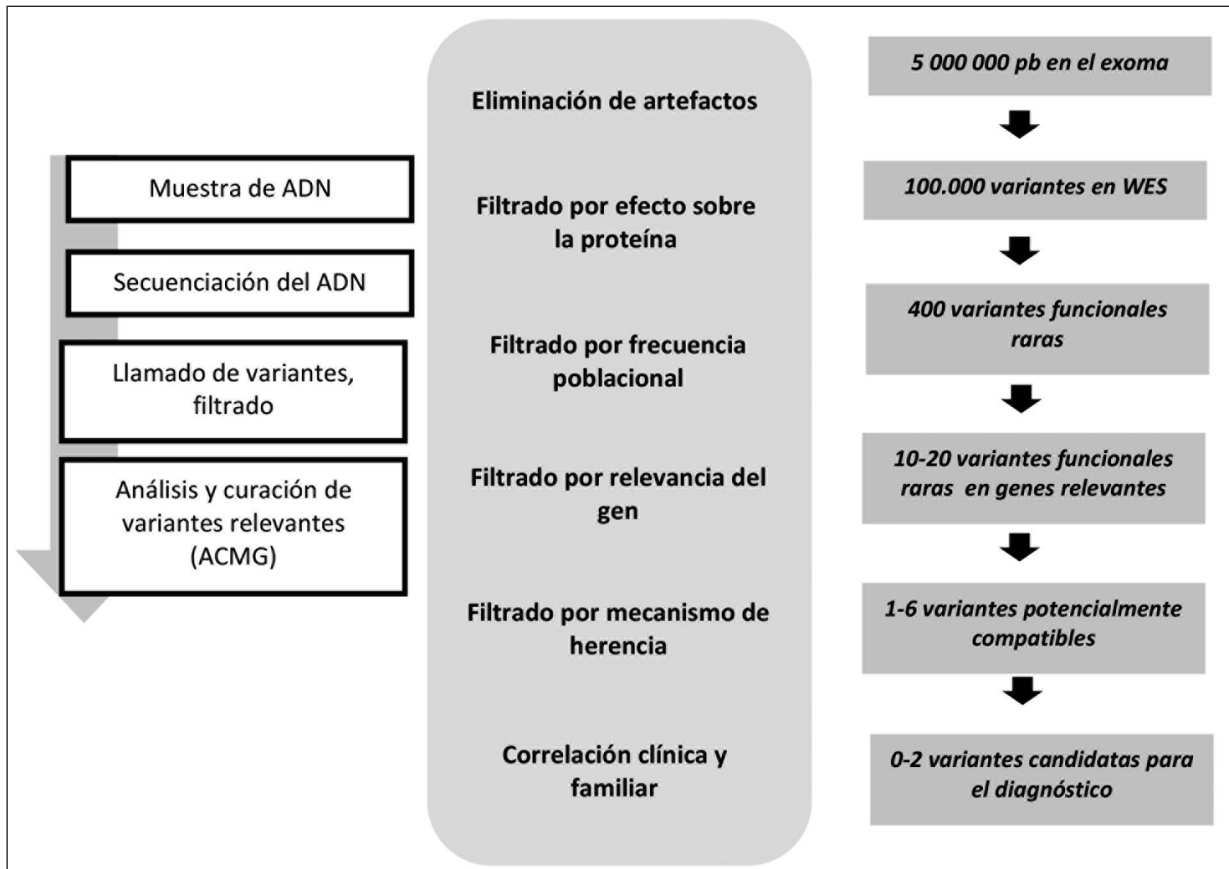
El análisis de las variantes detectadas mediante WES implica varios pasos de filtrado. Por poner un contexto, suelen identificarse de 60.000 a 100.000 SNVs por caso WES^{3, 9, 10}. Las variantes obtenidas suelen ser filtradas o priorizadas de acuerdo a parámetros como calidad de cobertura, frecuencia en población general, efecto sobre la proteína (sinónimas vs. no sinónimas, intrónicas), modo de herencia, transcrito afectado, y relación del gen sobre el fenotipo referido^{9, 10}. Los programas de análisis de datos WES suelen utilizar información disponible en bases de datos tales como OMIM, gnomAD, y predictores de patogenicidad *in silico* (Figura 1).

El número de variantes candidatas después estos filtrados es variable. Para las variantes restantes, debe reunirse toda la evidencia disponible acerca de la relación que podrían tener con lo observado en el paciente. Suelen utilizarse bases de datos especializadas tales como ClinGen, PanelApp, GenCC, GeneReviews, ClinVar, DECIPHER, HGMD, LOVD, AutoPVS1, LitVar, Mastermind, entre otras¹¹. Luego de reunir toda la información, las variantes serán clasificadas de acuerdo con las pautas de 2015 emitidas por el ACMG y la Asociación de Patología molecular (AMP)¹².

Las recomendaciones de la ACMG/AMP evalúan 28 tipos diferentes de evidencias asociadas a características inherentes a cada variante. A cada característica se le asigna un código alfanumérico:

- *Pathogenic Very Strong* (PVS): 1 código (PVS1). Se aplica a variantes de tipo truncamiento de proteína en genes donde la pérdida de función es un mecanismo de enfermedad.
- *Pathogenic Strong* (PS): 4 códigos (PS1 a PS4). Se considera evidencia fuerte de patogenicidad aspectos como que exista una variante ya clasificada como patogénica en el mismo aminoácido (PS1), presencia de novoCON paternidad y maternidad confirmadas (PS2), estudios funcionales que apoyen su patogenicidad (PS3), o prevalencia significativamente incrementada de la variante en afectados con respecto a controles (PS4).
- *Pathogenic Moderate* (PM): 6 códigos (PM1 a PM6). Se considera evidencia moderada de patogenicidad aspectos como localización de la variante en una región funcionalmente relevante (PM1), baja frecuencia o ausencia en controles (PM2), presencia de una segunda variante patogénica en el otro alelo para enfermedad recesiva (PM3), delección/inserción en marco en una región

Fig. 1.– Proceso de razonamiento seguido para la priorización de las variantes



Modificado de Ref. 17

no repetida o pérdida de la señal de stop (PM4), nuevo cambio en un aminoácido donde un cambio diferente ya se considera patogénico (PM5), o presencia *de novo* SIN paternidad y maternidad confirmadas (PM6).

- Pathogenic Supporting (PP): 4 códigos (PP1 a PP4). Se considera evidencia de soporte de patogenicidad aspectos como la co-segregación de la variante en afectados de una familia (PP1), *missense* en un gen en el que hay pocas *missense* en controles y donde este tipo de variantes son causa de enfermedad (PP2), predicciones de patogenicidad *in silico* (PP3), especificidad del fenotipo (PP4), o clasificación como patogénica por fuentes confiables (PP5).

- Benign Stand Alone (BA): 1 código (BA1). Este código aplica a variantes con una frecuencia poblacional superior al 5%.

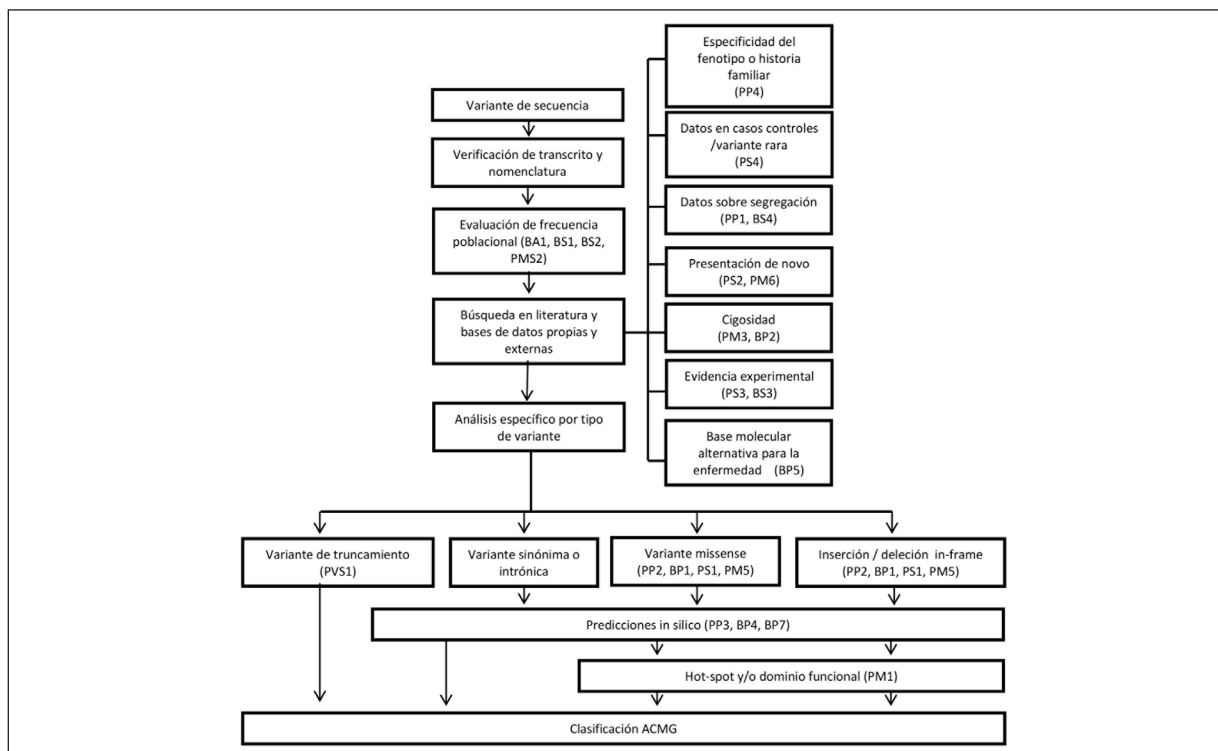
- Benign Strong (BS): 4 códigos (BS1 a BS4). Se considera evidencia fuerte de benignidad aspectos como frecuencia alélica mayor a la esperada para enfermedad (BS1), presencia en un individuo sano homocigoto (herencia recesiva), heterocigoto (herencia dominante), o hemocigoto (herencia ligada al X en enfermedad con

penetrancia completa esperada a edad temprana (BS2), estudios funcionales que no apoyen su patogenicidad (BS3), o fallo de segregación en una familia para fenotipos poco comunes (BS4).

- Benign Supporting (BP): 7 códigos (BP1 a BP7). Se considera evidencia de soporte de benignidad aspectos como que se trate de una variante *missense* en un gen donde la causa de enfermedad es el truncamiento (BP1), presencia en de una variante patogénica dominante completamente penetrante en el alelo opuesto, o en el mismo alelo para cualquier tipo de herencia (BP2), delección/ inserción en una región repetitiva sin función conocida (BP3), predicciones de benignidad *in silico* (BP4), caso con una base molecular alternativa establecida (BP5), clasificación como benigna por fuentes confiables (BP6), o variante sinónima en un nucleótido poco conservado que no parece afectar el *splicing* (BP7).

Una vez asignados los códigos correspondientes a una variante, estos se combinan para dar como resultado la clasificación final de la misma en 5 categorías: patogénica, probablemente patogénica, de significado incierto, probablemente benigna, y benigna (Figura 2).

Fig. 2.– Evaluación de variantes genéticas de acuerdo a criterios de la ACMG



<p>1 <i>Very strong</i> (PVS1) Y</p> <p>a) ≥ 1 <i>Strong</i> (PS1-PS4) O</p> <p>b) ≥ 2 <i>Moderate</i> (PM1-PM6) O</p> <p>c) 1 <i>Moderate</i> (PM1-PM6) y 1 <i>Supporting</i> (PP1-PP5) O</p> <p>d) ≥ 2 <i>Supporting</i> (PP1-PP5)</p>	<p>1 <i>Very strong</i> (PVS1) y 1 <i>Moderate</i> (PM1-PM6)</p>	<p>No se cumplen criterios para las otras clasificaciones</p>	<p>1 <i>Strong</i> (BS1-BS4) y 1 <i>Supporting</i> (BP1-BP7)</p>	<p>1 <i>Stand-alone</i> (BA1)</p>
<p>≥ 2 <i>Strong</i> (PS1-PS4)</p> <p>1 <i>Strong</i> (PS1-PS4) Y</p> <p>a) ≥ 3 <i>Moderate</i> (PM1-PM6)</p> <p>b) ≥ 2 <i>Moderate</i> (PM1-PM6) Y ≥ 2 <i>Supporting</i> (PP1-PP5) O</p> <p>c) 1 <i>Moderate</i> (PM1-PM6) Y ≥ 4 <i>Supporting</i> (PP1-PP5)</p>	<p>1 <i>Strong</i> (PS1-PS4) y 1-2 <i>Moderate</i> (PM1-PM6)</p> <p>1 <i>Strong</i> (PS1-PS4) y ≥ 2 <i>Supporting</i> (PP1-PP5)</p>	<p>Los criterios de benignidad y patogenicidad son contradictorios</p>	<p>≥ 2 <i>Supporting</i> (BP1-BP7)</p>	<p>≥ 2 <i>Strong</i> (BS1-BS4)</p>
	<p>≥ 3 <i>Moderate</i> (PM1-PM6)</p>			
	<p>2 <i>Moderate</i> (PM1-PM6) y ≥ 2 <i>Supporting</i> (PP1-PP5)</p>			
	<p>1 <i>Moderate</i> (PM1-PM6) y ≥ 4 <i>Supporting</i> (PP1-PP5)</p>			

Aunque estas guías definen claramente las categorías principales y el razonamiento de cada criterio, diversos estudios han señalado que en ocasiones los laboratorios pueden realizar interpretaciones discordantes^{6, 12, 13}.

Debe comentarse que estas guías están diseñadas para utilizarse sobre variantes en genes con evidencia sustancial que respalde su papel en DI, y no sobre genes con evidencia limitada o cuestionada: por lo tanto, las variantes identificadas en estos genes deben clasificarse como variantes de significado incierto. Como ejemplo, un estudio reciente que evaluó 156 genes propuestos como relacionados con DI y autismo determinó que 22 (14%) de los genes evaluados tenían evidencia limitadas o discutida en cuanto a su relación con estos fenotipos¹⁴.

Potenciales resultados del estudio realizado

Las variantes consideradas relevantes deben ser consignadas y detalladas en el reporte del estudio. Los posibles resultados del estudio son positivo, no concluyente, y negativo sin variantes.

Es importante contextualizar para el paciente una prueba que no ha identificado variantes relevantes. Los resultados negativos son relativamente comunes en las pruebas genómicas y plantean preocupaciones importantes. Debe comentarse que se trata de un estudio que utiliza una tecnología con limitaciones específicas, que analiza una serie concreta de genes y/o variantes, en un momento dado del conocimiento de la genética. Las limitaciones de WES incluyen dificultad para detectar regiones repetidas o con alta homología (que pueden dar errores de mapeo), e incapacidad para detectar variantes intrónicas profundas. También debido a la variabilidad en la profundidad de la cobertura que resulta del paso de captura, la detección de variantes en el número de copias puede ser problemática^{3,7,15}. Adicionalmente, algunos factores que pueden llevar a la omisión de variantes relevantes son la captura inadecuada de la región, la elección del transcrito incorrecto para un gen, o la evaluación incorrecta de la forma de herencia en un trío (por ejemplo, al codificar como no afectado a un padre levemente afectado).

Un resultado no concluyente describirá una o más variantes de significado incierto (VUS), definidas como variantes para las que no es posible establecer una relación definitiva con el cuadro clínico del paciente, debido a que la información disponible es conflictiva o no concluyente, o porque la información es escasa o está ausente⁴. Ante una VUS, el médico debe realizar correlación clínica, que suele basarse principalmente en una evaluación de las características fenotípicas y los patrones de herencia correspondientes a las enfermedades relacionadas con el gen en particular. En casos seleccionados puede considerarse realizar estudio de co-segregación familiar, ya que puede ser de ayuda para definir la potencial patogenicidad

de una VUS; también puede considerarse la reevaluación de la variante de manera periódica por un profesional.

Anualmente se notifican alrededor de 250-300 nuevos genes causantes de enfermedad y 9200 variantes asociadas a enfermedades^{2,16}. Debido a esto, diversos autores han indicado la conveniencia de reanalizar los datos de WES en casos negativos o no concluyentes cada 1 a 3 años, acompañado de una actualización del fenotipo clínico observado^{6,16}. El rendimiento diagnóstico del reanálisis oscila entre 10 y 15% de acuerdo con diversos estudios^{2,8,15-17}.

Un resultado positivo es en el que se han identificado la etiología molecular del fenotipo referido. El diagnóstico molecular de certeza permite confirmar un diagnóstico clínico, en ocasiones incluso cambiar un diagnóstico previo, y ser informativo acerca del pronóstico, opciones de manejo, vigilancia o prevención, así como plantear el modo de herencia, identificar el riesgo para otros miembros de la familia, y guiar la investigación sobre nuevas terapias^{3,4, 18}.

Otros aspectos a considerar por el médico prescriptor

Dentro del análisis debe considerarse la presencia de hallazgos incidentales o secundarios, que pueden observarse hasta en 0.5%-3.5% de los casos de WES¹⁵. Los hallazgos incidentales tienen implicaciones potenciales para la salud y significado clínico, pero no están relacionados con los síntomas de la enfermedad por los que se solicitó la prueba. Los hallazgos secundarios son buscados activamente en genes seleccionados por su alta penetrancia para enfermedades monogénicas, para las que existe capacidad de actuar clínicamente de forma efectiva⁷.

Es importante que el médico prescriptor de la prueba revise la estructura del informe, la información técnica o de metodología, las estrategias de trazabilidad de las muestras, los genes o paneles incorporados, aun en el caso de exoma, la calidad de la secuenciación, y las bases de datos o referencias utilizadas, así como si es necesario confirmar la variante mediante Sanger, o si ya lo ha hecho el laboratorio. Conviene tomar nota de las certificaciones con las que cuenta el laboratorio, si se cuenta con valores propios de sensibilidad y especificidad para SNVs y CNVs, criterios de selección de genes y variantes, si se detallan los criterios de clasificación y curación de variantes, y si se plantea la posibilidad de reanálisis en el futuro.

Conclusiones

El mayor desafío actual es el filtrado y la interpretación de las variantes detectadas mediante NGS. El papel del médico prescriptor debe incluir la calibración cuidadosa de

la prueba, y una comprensión detallada de los procesos de análisis así como de cuáles pueden ser las fuentes de errores u omisiones en un estudio. El proporcionar información clínica completa y detallada, así como una comunicación directa entre el laboratorio y el médico tratante, son aspectos que se traducen en una clara mejora en el rendimiento de los estudios de secuenciación, ya que los hallazgos y las variantes pueden ser planteadas, evaluadas, discutidas, y seleccionadas de la mejor manera.

Al final, es el médico quien debe determinar si el diagnóstico molecular propuesto por el laboratorio constituye un diagnóstico genético sólido (ya sea en su totalidad o en parte) y si la confianza en esta afirmación es suficiente para guiar el manejo futuro del paciente, así como si el resultado permite realizar análisis familiar, o diagnóstico preimplantacional o prenatal.

Conflicto de intereses: Ninguno para declarar.

Bibliografía

1. Bass N, Skuse D. Genetic testing in children and adolescents with intellectual disability. *Curr Opin Psychiatry* 2018; 31:490-5.
2. Bruel AL, Vitobello A, TranMau-Them F, et al. Next-generation sequencing approaches and challenges in the diagnosis of developmental anomalies and intellectual disability. *Clin Genet* 2020; 98: 433-44.
3. De Luca C, Race V, Keldermans L, Bauters M, Van Esch H. Challenges in molecular diagnosis of X-linked Intellectual disability. *Br Med Bull* 2020; 15; 133: 36-48.
4. Sánchez-Luquez KY, Carpena MX, Karam SM, Tovo-Rodriguez L. The contribution of whole-exome sequencing to intellectual disability diagnosis and knowledge of underlying molecular mechanisms: A systematic review and meta-analysis. *Mutat Res Rev Mutat Res* 2022; 27; 790:108428.
5. Richardson E. The Utility of Whole Exome Sequencing in Patients with Intellectual Disability and Developmental Delay as a First-Tier Diagnostic Testing Strategy. University of South Carolina (Master's thesis). 2020. En: <https://scholarcommons.sc.edu/etd/5717>; consultado noviembre 2020.
6. Johnson B, Ouyang K, Frank L, et al. Systematic use of phenotype evidence in clinical genetic testing reduces the frequency of variants of uncertain significance. *Am J Med Genet A* 2022; 188: 2642-51.
7. Manickam K, McClain MR, Demmer LA, et al. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* 2021; 23: 2029-37.
8. Robertson AJ, Tan NB, Spurdle AB, Metke-Jimenez A, Sullivan C, Waddell N. Re-analysis of genomic data: An overview of the mechanisms and complexities of clinical adoption. *Genet Med* 2022; 24: 798-810.
9. Vorsteveld EE, Hoischen A, van der Made CI. Next-Generation Sequencing in the Field of Primary Immunodeficiencies: Current Yield, Challenges, and Future Perspectives. *Clin Rev Allergy Immunol* 2021; 61: 212-25.
10. Zhang J, Yao Y, He H, Shen J. Clinical Interpretation of Sequence Variants. *Curr Protoc Hum Genet* 2020;106(1): e98.
11. Chunn LM, Nefcy DC, Scouten RW, et al. Mastermind: A Comprehensive Genomic Association Search Engine for Empirical Evidence Curation and Genetic Variant Interpretation. *Front Genet* 2020;13; 11: 577152.
12. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; 17: 405-24.
13. Harrison SM, Biesecker LG, Rehm HL. Overview of Specifications to the ACMG/AMP Variant Interpretation Guidelines. *Curr Protoc Hum Genet* 2019;103(1): e93.
14. Riggs ER, Bingaman TI, Barry CA, et al. Clinical validity assessment of genes frequently tested on intellectual disability/autism sequencing panels. *Genet Med* 2022; 24: 1899-908.
15. Han JY, Lee IG. Genetic tests by next-generation sequencing in children with developmental delay and/or intellectual disability. *Clin Exp Pediatr* 2020; 63: 195-202.
16. Al-Nabhani M, Al-Rashdi S, Al-Murshedi F, et al. Reanalysis of exome sequencing data of intellectual disability samples: Yields and benefits. *Clin Genet* 2018; 94: 495-501.
17. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet* 2018; 19: 253-68.
18. Ellard S, Baple EL, Callaway A, et al. ACGS Best Practice Guidelines for Variant Classification in Rare Disease (2020). En: <https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf>; consultado noviembre 2022.