

Calidad de los fenotipos humanos en pacientes con discapacidad intelectual con estudios de secuenciación de exoma en un hospital pediátrico peruano

Quality of human phenotypes in patients with intellectual disabilities undergoing exome sequencing tests at a Peruvian pediatric hospital

Pedro Custodio-Sánchez ^{1,a}, Hugo Abarca-Barriga ^{1,2,b}

¹ Instituto de Investigaciones de Ciencias Biomédicas, Universidad Ricardo Palma. Lima, Perú.

² Servicio de Genética, Instituto Nacional de Salud del Niño-Breña. Lima, Perú.

^a Estudiante de medicina. ORCID: <https://orcid.org/0009-0008-6399-8265>

^b Médico genetista, doctor en medicina. ORCID: <https://orcid.org/0000-0002-0276-2557>

An Fac med. 2024;85(3)./ DOI: <https://doi.org/10.15381/anales.v85i3.28859>.

Correspondencia:

Pedro Custodio Sánchez
201720630@urp.edu.pe

Recibido: 22 de abril 2024

Aprobado: 06 de agosto 2024

Publicación en línea: 30 de agosto 2024

Conflictos de interés: Los autores declaran no tener conflictos de interés.

Fuente de financiamiento: Autofinanciado.

Contribuciones de autoría: Ambos autores participaron en la concepción y diseño del estudio, recolección y análisis de datos, redacción del manuscrito, revisión crítica del artículo y aprobaron la versión final del manuscrito. Además, asumen la responsabilidad por lo publicado.

Citar como: Custodio-Sánchez P, Abarca-Barriga H. Calidad de los fenotipos humanos en pacientes con discapacidad intelectual con estudios de secuenciación de exoma en un hospital pediátrico peruano. An Fac med. 2024; 85(3): . DOI: <https://doi.org/10.15381/anales.v85i3.28859>.

Resumen

Introducción. La secuenciación masiva y la ontología del fenotipo humano han revolucionado el diagnóstico genético, permitiendo una comprensión más clara de las enfermedades. Los fenotipos descritos en las historias clínicas pueden ser específicos o generales, cuanto más detallada y específica sea la descripción, mayor será la calidad de la información. **Objetivos.** Determinar la relación entre la calidad de los fenotipos humanos en pacientes con discapacidad intelectual y los hallazgos de secuenciación masiva en el Instituto Nacional de Salud del Niño durante el periodo 2020-2023. **Métodos.** Se analizaron 124 pacientes con discapacidad intelectual con secuenciación de exoma. Se evaluó la calidad de los fenotipos humanos mediante diferentes índices de calidad. **Resultados.** Se identificaron 278 términos fenotípicos humanos. La precisión fenotípica mejoró después del filtrado, destacando la importancia de descripciones fenotípicas específicas (de 103 a 86 términos únicos). El índice de especificidad del conjunto de datos fue más bajo en el grupo con variantes de significado incierto, en comparación con el grupo con variantes patogénicas (0 vs. 0,05). Los fenotipos más frecuentes fueron el retraso psicomotor (12,9% - 15,9%) y la microcefalia (4,5% - 5,2%). **Conclusiones.** Se destaca la importancia de evaluar la calidad de los fenotipos humanos. La presencia de variantes inciertas sugiere una complejidad genotipo-fenotipo, subrayando la necesidad de mayor claridad en la interpretación genética. La consistencia de los coeficientes de IC y el índice de especificidad del conjunto de datos, muestra la competencia del personal médico-genetista local.

Palabras clave: Discapacidad Intelectual; Fenotipo; Secuenciación del Exoma; Hospitales Pediátricos; Perú (fuente: DeCS BIREME).

Abstract

Introduction. Massive sequencing and the Human Phenotype Ontology have revolutionized genetic diagnosis, allowing for a clearer understanding of diseases. The phenotypes described in medical records can be specific or general, the more detailed and specific the description, the higher the quality of the information. **Objectives.** To determine the relationship between the quality of human phenotypes in patients with intellectual disability and the findings from massive sequencing test at the Instituto Nacional de Salud del Niño. **Methods.** We analyzed 124 patients with intellectual disabilities with exome sequencing. The quality of human phenotypes was evaluated using different quality index. **Results.** A total of 278 human phenotype terms were identified. Phenotypic precision improved after filtering, highlighting the importance of specific phenotypic descriptions (from 103 to 86 unique terms). The specificity index of the dataset was lower in the group with variants of uncertain significance compared to the group with pathogenic variants (0 vs. 0.05). The most frequent phenotypes were psychomotor delay (12.9% - 15.9%) and microcephaly (4.5% - 5.2%). **Conclusions.** The importance of evaluating the quality of human phenotypes is highlighted. The presence of uncertain variants suggests a genotype-phenotype complexity, emphasizing the need for greater clarity in genetic interpretation. Consistency in confidence intervals and the specificity index of the dataset underscores show the competence of local medical-genetic personnel.

Keywords: Intellectual Disability; Exome Sequencing; Phenotype; Hospitals, Pediatric; Peru (source: MeSH NLM).

INTRODUCCIÓN

El 1 a 3% de la población tienen discapacidad intelectual (DI), caracterizada por limitaciones significativas tanto en el funcionamiento intelectual, como en el comportamiento adaptativo, y dificultades en áreas conceptuales, sociales y prácticas de la vida. La DI es diagnóstica mediante diferentes pruebas neuropsicológicas en niños entre cinco a 18 años [1-3].

La secuenciación masiva junto a herramientas bioinformáticas han cambiado el diagnóstico de las enfermedades genéticas [4,5]. Es así que el análisis bioinformático ayuda a evaluar el significado de las variantes genéticas, así como posibilita la identificación de nuevos genes vinculados a patologías o cambios en el genoma que determinan la susceptibilidad a enfermedades específicas [5].

La relación entre los datos de secuenciación y la clínica de los pacientes pueden utilizarse para extraer conclusiones significativas, en este contexto, la Ontología del Fenotipo Humano (HPO, *Human Phenotype Ontology*) juega un papel clave [6]. Los fenotipos humanos o HPO (por sus siglas en inglés) son un conjunto estandarizado de términos ordenados jerárquicamente que se utilizan para describir los fenotipos clínicos observados en los pacientes, proporcionando un marco sistemático y consistente para la anotación de características clínicas, lo que permite una interpretación unificada [7].

Es imperativo realizar una buena caracterización fenotípica del paciente, con una clínica adecuada y específica [5]. La información insuficiente en la caracterización de los pacientes puede derivarse por ejemplo por la complejidad fenotípica, experiencia del examinador y los límites de tiempo durante las consultas. Esto puede provocar un problema, sobre todo si existe una característica clínica en común, obviando los otros fenotipos, los cuales si son descritos aumentarían la especificidad. Esta insuficiencia en la caracterización fenotípica puede conducir a diagnósticos imprecisos y terapéuticas subóptimas, por lo que identificar subgrupos dentro de las cohortes es fundamental para mejorar el diagnóstico [8]. Esto se puede alcanzar mejorando la calidad o la profundidad

del fenotipado, el cual es definido como el análisis preciso y exhaustivo de las características clínicas de un individuo, en el cual se observan y describen los componentes individuales del fenotipo, el cual puede ser determinado mediante diferentes pruebas (ej. índice de especificidad del conjunto de datos) [7,9].

Esta investigación tuvo como objetivo evaluar la calidad de los fenotipos humanos en una cohorte de pacientes con discapacidad intelectual que cuentan con secuenciación de exoma. Haciendo uso de diversas herramientas computacionales y estadísticas, buscamos explorar las relaciones entre los resultados genéticos y fenotípicos, para mejorar la comprensión de las otras características clínicas que acompañan a la discapacidad intelectual.

MÉTODOS

Diseño del estudio

Se realizó un estudio observacional y descriptivo, en el Instituto Nacional de Salud del Niño Breña (INSN), ubicado en la ciudad de Lima Metropolitana, Perú.

Población y muestra

La población fueron los 269 pacientes atendidos en el Servicio de Genética y Errores Innatos del Metabolismo del INSN entre los años 2020-2023 con diagnóstico de discapacidad intelectual (DI). La DI fue determinada con una prueba neuropsicológica (Weschler) realizada en el departamento de salud mental. La muestra fue de 124 pacientes con DI que tuvieron la secuenciación de exoma, quienes previamente tuvieron estudios negativos del análisis cromosómico por micromatrices o la determinación del número de tripletes CGG del gen *FMR1*.

Variables

La secuenciación de exoma permitió clasificar a los participantes en cuatro grupos: variantes patogénicas, probablemente patogénica, de significado incierto y sin variantes (probablemente benignas o benignas). Los pacientes con variantes patogénicas o probablemente patogénicas fueron recategorizados como variantes positivas, mientras que los pacientes con ausencia de variantes o con variantes de significado incierto (VUS) fueron con-

siderados como variantes negativas. Las otras variables fueron sexo y edad.

Análisis de la calidad

Los fenotipos fueron extraídos de la base de datos utilizando los códigos HPO (*Human Phenotype Ontology*) [10]. Para el análisis de la calidad se utilizaron diferentes criterios como los términos únicos por cada paciente, promedio, percentil 90, porcentaje de términos HPO con términos secundarios más específicos. Además, se determinó la amplitud y profundidad del fenotipado, utilizando los siguientes conceptos: el uso del contenido de información (IC), frecuencia y distribución de términos, y el índice de especificidad del conjunto de datos (Dsl). Adicionalmente, se ejecutó un análisis de conglomerados y distribución de contenidos de información, mediante la agrupación por métodos de similitud, análisis de componentes principales (PCA), y análisis de clusterización jerárquica (ACJ). Los detalles de las fórmulas se encuentran en el Material Suplementario 1.

Los términos únicos de HPO fueron definidos como el número de términos HPO distintos utilizados para describir a los pacientes de cada grupo. El promedio de los términos de HPO por paciente es el número medio de términos HPO. El percentil 90 del HPO se estableció al clasificar a los pacientes de mayor a menor longitud del perfil de fenotipo, estableciendo la cantidad de información fenotípica disponible para los miembros de cada grupo con perfiles fenotípicos menos detallado. El porcentaje de términos HPO con términos secundarios más específicos se definió como el porcentaje de fenotipos asignados a pacientes en cada grupo para los cuales hay términos más específicos disponibles dentro del HPO para caracterizar al paciente.

Uso del contenido de información

El IC fue utilizado como una medida para cuantificar la especificidad, el valor informativo de dicho término y comprender las características particulares de cada grupo de pacientes [10].

Análisis de frecuencia y distribución de términos de fenotipos humanos

La ontología utilizada en este análisis es la última versión (abril de 2021), que

incluye más de 13 000 términos. La frecuencia de cada término de HPO se definió como el número de pacientes que presentan un fenotipo específico dividido por el número total de pacientes. La especificidad del fenotipado se calculó midiendo la profundidad de estos términos dentro de la estructura jerárquica del HPO (<https://hpo.jax.org/>), con valores que varían entre uno (nodo raíz) a 16 (el más específico). La profundidad puede determinarse considerando la frecuencia de aparición del término en la muestra o utilizando el conjunto de términos únicos en cada subgrupo.

Al comparar la profundidad de los términos con la profundidad promedio de todos los términos en el HPO, podemos determinar si los pacientes de la muestra tienden a ser fenotipados con términos más o menos específicos.

Índice de especificidad del conjunto de datos

El índice de especificidad del conjunto de datos (Dsl) evalúa la especificidad de los términos HPO en nuestra muestra en comparación con su jerarquía. El HPO se dividió en dos secciones: la sección baja que incluye los niveles desde la raíz hasta el nivel con el mayor número de términos; y la sección alta, que abarca desde el nivel más alto (Lmax) hasta el término más profundo en la ontología (Material Suplementario 2).

Para un conjunto de términos en un subgrupo, penalizamos los términos en los niveles más bajos de la jerarquía y recompensamos los términos en los niveles más altos. Dado que los términos HPO cercanos a la raíz son más accesibles y tienen múltiples términos secundarios, aumentamos la penalización/recompensa según la profundidad del término en

comparación con Lmax. Por el contrario, los términos muy profundos son muy específicos y proporcionan información más precisa^[9].

Un valor de 1 señaló que la contribución de ambas secciones es equivalente; pero en la práctica, los valores de Dsl suelen ser inferiores a 1 debido a la dificultad de alcanzar alta calidad en la ontología^[9].

Análisis de conglomerados y distribución de contenidos de información

La agrupación por métodos de similitud se realizó basada en la similitud de sus perfiles, utilizando el coeficiente de Jaccard, que produce puntuaciones normalizadas entre 0 y 1^[11]. Estas puntuaciones se usaron para crear una matriz de similitud y los resultados se presentaron en mapas de calor que muestran la similitud de los perfiles entre los pacientes.

El Análisis de Componentes Principales (PCA) fue usado para procesar los datos fenotípicos y se generaron cincuenta resultados factoriales para cada paciente, mostrando su contribución a la variabilidad total. Para representar la posición relativa de cada paciente, se utilizaron las dos primeras puntuaciones factoriales, F1 y F2. El PCA permitió reducir la complejidad del conjunto de datos y visualizar las relaciones entre los pacientes según sus fenotipos.

El Análisis de Clusterización Jerárquica (ACJ) clasificó y agrupó las observaciones basado en la representación bidimensional obtenida del PCA. El ACJ organizó las observaciones en grupos con características similares utilizando el método de aglomeración de Ward y la distancia euclidiana para calcular la disimilitud. Este método generó tres grupos distintos, identificados con el índice de Hartigan para el truncamiento del dendrograma.

Todos los cálculos matemáticos y gráficos se realizaron con Excel[®] y Xlstat[®]. Además, se utilizó R Studio[®] junto con la biblioteca ggplot2 para elaborar los gráficos de dispersión.

Aspectos éticos

El presente estudio cuenta con la autorización de los comités de ética de la Facultad de Medicina Humana de la Universidad Ricardo Palma (PG-008-2024) y del Instituto Nacional de Salud del Niño – Breña (N°284/2021- CIEI-INSN). La información fue revisada solo con fines de investigación y se mantuvo confidencial. La identidad y otros datos personal de los pacientes no fueron revelados. Se cumplió las normativas de Helsinki y de confidencialidad.

RESULTADOS

Se analizaron 124 pacientes con DI y con resultados de secuenciación del exoma. Fueron identificados 278 HPO. La mediana de la edad fue de 6,3 años y el 57,3% fueron varones (Tabla 1).

Análisis de calidad del fenotipo

En la muestra después del filtrado en el refinamiento del fenotipo que contenía 103 términos HPO únicos, se redujo a 86 términos, ofreciendo un panorama fenotípico más específico. Se observó una tendencia similar en los subgrupos sin variantes, patogénicas o probablemente patogénicas y VUS (Tabla 2).

En el grupo sin variantes se identificaron 46 términos HPO únicos y 80 términos específicos. Entre los pacientes con variantes patogénicas o probablemente patogénicas hubo 47 términos HPO únicos y 70 términos más específicos. En contraste, la cohorte de pacientes con VUS mostró 41 términos HPO únicos y 93

Tabla 1. Características generales relacionadas de los pacientes con discapacidad intelectual.

| Variable | General (n = 124) | Sin variantes (n = 38) | Patogénico o probablemente patogénico (n = 38) | VUS (n = 48) |
|----------------------------|-------------------|------------------------|--|--------------|
| Edad (años) (mediana; RIC) | 6,3 (4,6) | 6,9 (5,15) | 5,2 (5,45) | 6,7 (5,72) |
| Sexo (% femenino) | 42,7 | 47,4 | 52,6 | 31,3 |

RIC= Rango intercuartílico. VUS= Variantes de significado incierto.

Tabla 2. Características de los términos HPO en pacientes con discapacidad intelectual con resultados de secuenciación de exoma en un hospital pediátrico peruano.

| Nombre | Muestra | | | | Muestra con exclusión de pacientes con menos de tres términos HPO | | | |
|---|---------|---------------|---------------------------------------|------|---|---------------|---------------------------------------|------|
| | General | Sin Variantes | Patogénico o probablemente patogénico | VUS | General | Sin Variantes | Patogénico o probablemente patogénico | VUS |
| Tamaño de muestra | 124 | 38 | 38 | 48 | 112 | 36 | 32 | 44 |
| Términos únicos de HPO | 103 | 46 | 47 | 41 | 86 | 40 | 24 | 27 |
| Términos HPO con términos secundarios más específicos | 175 | 80 | 70 | 93 | 167 | 75 | 66 | 90 |
| Términos de HPO por paciente (promedio) | 5,65 | 5,53 | 5,32 | 6,02 | 6,07 | 5,75 | 6,00 | 6,39 |
| Términos HPO por paciente: percentil 90 | 9 | 8,3 | 9,3 | 10 | 9 | 8,5 | 9,9 | 10 |
| Porcentaje de términos HPO con términos secundarios más específicos | 63,0 | 63,5 | 59,8 | 69,4 | 66,0 | 65,2 | 73,3 | 76,9 |

términos más específicos. El promedio de términos HPO fue mayor en el subgrupo de pacientes con VUS (6,02 y 6,39) en comparación de los grupos sin variantes (5,32-5,75) (Tabla 2).

Después del filtrado se observan que los porcentajes de términos HPO fueron más altos, lo que sugiere una caracterización más detallada de los fenotipos (69,4% vs 63%). En el mismo sentido, se observó que en el subgrupo VUS presentaron mayores porcentajes posfiltración (76,9%) con relación a los otros dos grupos (65,2-73,3%) (Tabla 2). El percentil 90 de los términos HPO por paciente mostró que el 90% de los pacientes utilizaron en promedio menos de nueve términos HPO ^(8,3-10).

Las características clínicas más frecuentes en todos los subgrupos fueron

el retraso del desarrollo psicomotor (12,9 - 15,9%) y la microcefalia (4,5- 5,2%). Entre los pacientes con variantes patogénicas o probablemente patogénicas existió una mayor frecuencia de convulsiones (5,2%). Entre los pacientes con variantes de significado incierto (VUS) el trastorno del espectro autista (TEA) y la talla baja fue más frecuente que en los otros grupos (Material Suplementario 3). La ubicación jerárquica dentro de la ontología del fenotipo humano mostró que los fenotipos más frecuentes tienen una profundidad variable entre 5 y 10 (Material Suplementario 4).

Se comparó la distribución de los niveles de todos los términos HPO basado en una cohorte hipotética (Material suplementario 2). Independiente del sub-

grupo, se observó que los términos HPO utilizados tienden a estar más a la izquierda, por lo tanto, fueron más generales y menos específicos, mientras que los términos HPO únicos tienden a ser más específicos y detallados. Sin embargo, en el grupo con VUS, la curva tiende a estar más hacia la izquierda, lo que significa una menor especificidad (Figura 1).

El Dsl mostró en toda la muestra una correlación débil de los términos en los niveles alto y bajo, incluso después del filtrado (Dsl = 0,02). Por el contrario, al examinar los subgrupos sin variantes y variantes patogénicas o probablemente patogénicas, los valores iniciales de Dsl fueron de 0,07 y 0,05, respectivamente, mostrando un aumento a 0,12 y 0,09 cuando se centraron sólo en pacientes

Tabla 3. Índice de especificidad del conjunto de datos (Dsl) para cada subgrupo de pacientes con discapacidad intelectual.

| Frecuencia | Muestra | | | | Muestra con exclusión de pacientes con menos de tres términos HPO | | | |
|-----------------|---------|---------------|---------------------------------------|-------|---|---------------|---------------------------------------|-------|
| | General | Sin variantes | Patogénico o probablemente patogénico | VUS | General | Sin variantes | Patogénico o probablemente patogénico | VUS |
| Cohorte | 0,02 | 0,07 | 0,05 | 0 | 0,02 | 0,12 | 0,09 | 0 |
| Términos únicos | 4,88 | 2,55 | 5,52 | 10,48 | 4,86 | 3,06 | 5,47 | 10,26 |

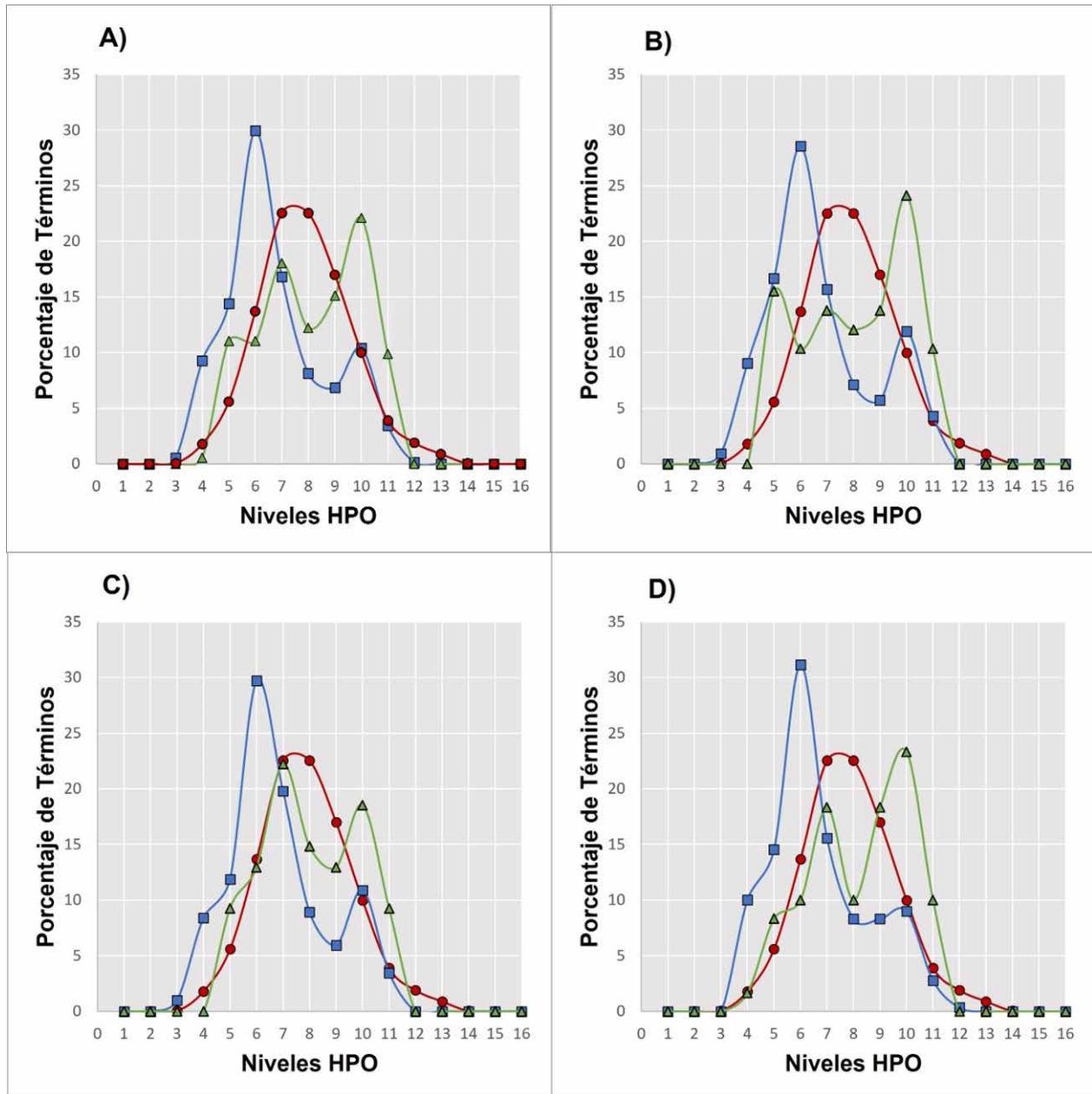


Figura 1. Gráficos de distribución porcentual de términos de HPO para los grupos. Las curvas verdes representan la frecuencia de términos HPO únicos para describir la cohorte, las curvas azules tienen en cuenta la frecuencia de la totalidad de términos de la cohorte y las curvas rojas muestran la frecuencia hipotética de términos incluidos en el HPO en cada nivel.

A. Muestra global, la distribución de todos los términos HPO utilizados en el conjunto de datos (curva azul) es similar a la curva de la Ontología HP (curva roja), con una elevación a partir del nivel 3, un pico en el nivel 5, un descenso en el nivel 8, y luego un repunte que sigue la tendencia de la curva HPO en el nivel 10. Los términos HPO únicos utilizados en el conjunto de datos (curva verde) muestran una distribución medianamente similar a la curva de la Ontología HP (curva roja), con picos en los niveles 5, 7 y 10. **B.** Subgrupo sin variantes, la distribución es similar al gráfico de la muestra global, con la diferencia de que los picos de la curva de los términos HPO únicos (curva verde) en los niveles 5 y 10 son ligeramente más altos. La representación de todos los términos (curva azul) mantiene una similitud marcada al grupo general, con la diferencia de un pico ligeramente más alto en el nivel 10. **C.** Subgrupo con variantes patogénicas o probablemente patogénicas, la distribución de términos HPO únicos (curva verde) muestra un ascenso desde el nivel 4, un pico máximo en el nivel 7, un descenso hasta el nivel 9, y un último pico en el nivel 11, descendiendo hasta el nivel 12. La frecuencia de todos los términos (curva azul) presenta un pico alto en el nivel 6 y se mantiene por debajo de la curva roja de la ontología. Este desplazamiento hacia la derecha sugiere que muchos pacientes han sido fenotipados profundamente. **D.** Subgrupo con variantes de significado incierto, la curva verde, que representa la frecuencia de términos únicos, exhibe un ascenso desde el nivel 4, un pico en el nivel 7, un descenso en el nivel 9, y un último pico en el nivel 11. La frecuencia de todos los términos (curva azul) muestra un pico en el nivel 6, sugiriendo que muchos pacientes del grupo VUS han sido fenotipados de manera más superficial en comparación con el análisis de términos únicos.

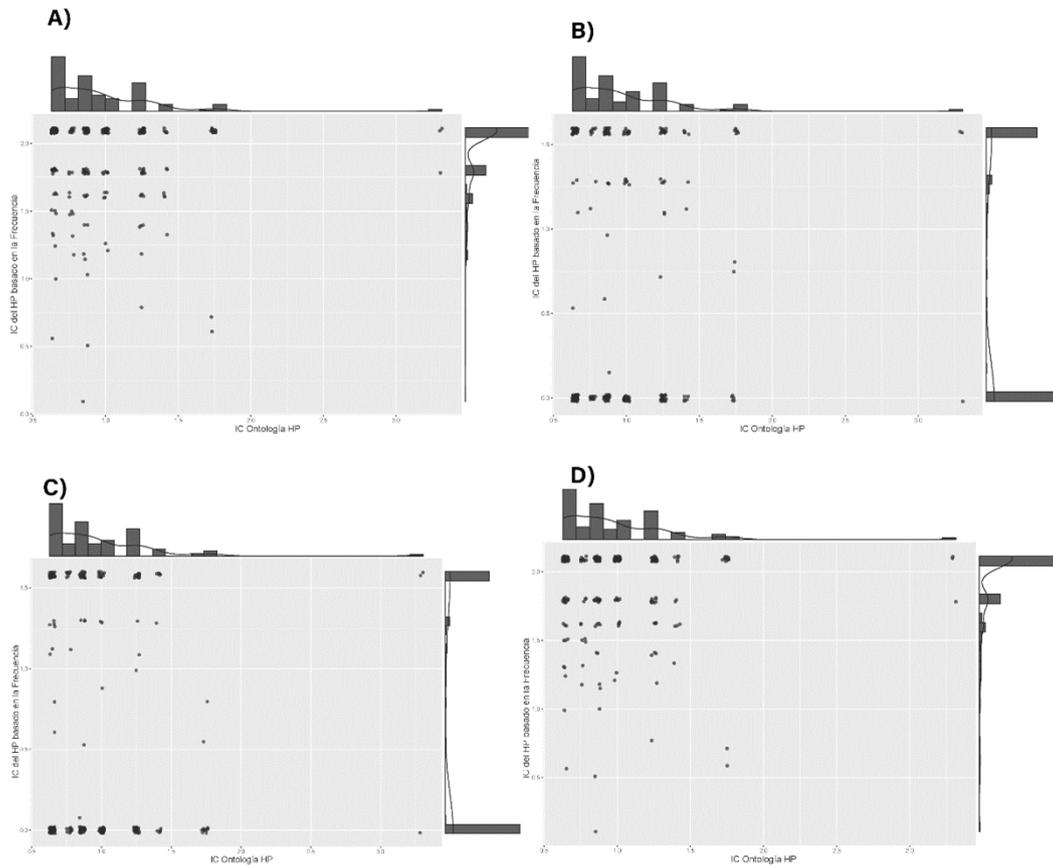


Figura 2. Distribución del contenido de información (IC). **A.** General. **B.** Sin variantes. **C.** Patogénico o probablemente patogénico. **D.** VUS. Los valores de CI basado en frecuencia, se calcularon utilizando los pacientes de los tres subgrupos.

con más de 2 términos. En el subgrupo VUS no se observaron diferencias del Dsl en los niveles alto o bajo, ni tampoco después del filtrado (Dsl = 0). Esto implicaría un fenotipo más complejo y variado dentro de este grupo (Tabla 3).

En la muestra, el Dsl utilizando términos únicos fue de 4,88. Al considerar a pacientes con más de dos términos, la diversidad fenotípica se mantuvo casi estable (4,86). En los subgrupos sin variantes y variantes patogénicas o probablemente patogénicas, el Dsl fue de 2,55 y 5,52, respectivamente. Al filtrar con más de dos términos, el Dsl fue de 3,06 y 5,47, respectivamente. En el subgrupo de pacientes con VUS muestra una riqueza fenotípica sustancial, con 10,48 términos distintos.

En la muestra se observó que los HPO en su mayoría utiliza términos con IC bajos (Figura 2A). En los subgrupos sin variantes

o con variantes patogénicas o probablemente patogénicas muestran una similitud, observándose dos grupos con valores altos y bajos de IC (Figura 2B y 2C).

Comparación de perfiles fenotípicos

El índice de Jaccard mostró vínculos fenotípicos y estructuras específicas dentro de cada subgrupo. En la muestra global se observó una mayor diversidad fenotípica, mientras que en los subgrupos hubo agrupamientos más precisos y asociados a determinadas características clínicas o perfiles genéticos (Material Suplementario 5).

Análisis de componentes principales

El PCA evaluó la distribución y la variabilidad de los fenotipos en una base de datos de 278 términos HPO de los 124 pacientes. Estos resultados factoriales representan la contribución de cada paciente a la variabilidad total del conjunto

de datos. Optamos por seleccionar las dos primeras puntuaciones factoriales, F1 y F2, para representar la posición relativa de cada paciente en un espacio bidimensional definido por los dos primeros componentes principales, los cuales muestran una heterogeneidad (Material Suplementario 6).

Análisis de clusterización jerárquica

El índice de Hartigan para el truncamiento del dendograma, resultó en tres grupos con una distribución parecida (Material Suplementario 7).

Subclasificación en cohortes de variantes positivas y negativas

En el análisis bidimensional entre los tres grupos extraídos para evaluar la distribución de pacientes que presentaron una variante positiva o negativa se observó que es similar (Material Suplementario 8).

En la muestra, el retraso en el desarrollo psicomotor fue el más frecuente. Además, en un mismo paciente se pueden observar varios fenotipos como convulsiones, trastorno del espectro autista (TEA), microcefalia y talla baja (Material Suplementario 9).

DISCUSIÓN

Según nuestros hallazgos, los principales componentes de calidad como los análisis del Dsl y la distribución de términos HPO, muestran que el filtrado aumenta la especificidad fenotípica y ayuda a identificar patrones distintivos; y que el filtrado de fenotipos -pacientes con más de dos términos HPO- mejora la precisión en pacientes con variantes patogénicas o probablemente patogénicas.

La evaluación de la exhaustividad y especificidad de las descripciones fenotípicas reveló una variabilidad significativa. La aplicación de filtros para refinar el fenotipo demostró mejorar la precisión fenotípica. Así, el número de términos HPO únicos disminuyó en un 83%, indicando un panorama fenotípico más enfocado y específico, el cual no fue observado en la cohorte de pacientes con DI de la publicación de Rojano *et al.* [9]. Esta reducción de los términos únicos podría indicar que existe una mayor homogeneidad, especificidad y profundidad del fenotipo. En el grupo que fue eliminado, no significa que necesariamente el fenotipado sea defectuoso, sino que podría deberse a la falta de precisión al momento del examen clínico [9].

Se observa que, en comparación al estudio previo, las medias de HPOs por paciente es mucho más alta en esta investigación (5,65 vs 2,71), aunque el porcentaje de términos secundarios más específicos (PTSME) fue menor, pudiendo deducirse una mayor descripción clínica (63% vs 81,71%) [9]. Además, se puede observar que en nuestra cohorte hay una tendencia a que estos PTSME después del filtrado tienden a aumentar (de 63% a 66%), mientras que en la publicación previa se observa una disminución (de 81,7% a 78,7%) [9]. Estos hallazgos sugieren que en nuestra cohorte el filtrado ha enriquecido la especificidad y profundidad del fenotipado.

El patrón resultante del análisis del IC en los diferentes grupos muestra términos HPO informativos, sin embargo, se identifica una cantidad mínima de estos con un IC alto. Esto indica que la mayoría de los pacientes dentro de este grupo presentan combinaciones con otros términos menos específicos. No obstante, el grupo estudiado presenta una distribución ligeramente más homogénea que un estudio previamente realizado [9].

Al comparar los valores de términos únicos de HPO (percentil 90) con la cohorte de pacientes con DI en la publicación previa, se observa que los pacientes analizados en esta investigación presentaron un fenotipado más profundo y exhaustivo (10 vs 1) [9].

Al evaluar el Dsl, observamos que en la muestra global el fenotipado es general, aunque esto mejora en los subgrupos con variantes patogénicas y probablemente patogénicas, mostrando así que existe una mayor especificidad, lo cual podría haber ayudado en el análisis bioinformático.

Al centrarnos en los subgrupos de pacientes sin variantes y con variantes patogénicas o probablemente patogénicas, notamos un cambio significativo. Los valores iniciales de Dsl de 0,07 y 0,05, respectivamente, aumentaron a 0,12 y 0,09 después del filtro, destacando la importancia del filtrado para mejorar la especificidad fenotípica en estas poblaciones genéticas. Sin embargo, observamos que estos valores están por debajo a lo reportado previamente en pacientes con otras patologías genéticas (0,13- 0,27), esto significa que la calidad de los términos es baja [9], aunque en el grupo con solo discapacidad intelectual el valor que obtuvieron es cero.

Si bien el Dsl brinda una indicación clara de la profundidad del fenotipado, debe interpretarse junto con otras métricas, como el porcentaje de términos HPO con términos secundarios más específicos [9]. Se observó una distribución de los HPO de la muestra global similar a los reportados por Rojano *et al.* Sin embargo, la tendencia de los HPO únicos revela una mayor especificidad con relación al mismo estudio [10].

El análisis detallado de los fenotipos reveló patrones distintivos, destacando el

retraso del desarrollo psicomotor como la característica más prevalente en todas las categorías de variantes genéticas. Esta observación concuerda con investigaciones previas que describen el deterioro cognitivo y alteraciones relacionadas a esta como un fenotipo patológico comúnmente asociado con pacientes que tienen discapacidad intelectual [10,12].

Teniendo en cuenta la variabilidad fenotípica, se examinaron las diferencias en los resultados de secuenciación entre varios grupos de pacientes. El análisis encontró patrones de agrupamiento y diferencias entre estos grupos, lo que indica relaciones específicas entre la variabilidad genética y fenotípica. Estos resultados respaldan estudios previos que indican que, debido a un espectro fenotípico estrecho, la cohorte de discapacidad intelectual tiende a formar grupos cercanos [9].

Es importante mencionar que la mayoría de valores que evalúan la calidad en la cohorte de pacientes con VUS nos muestran en general una mayor calidad de HPO y un menor número de términos; sin embargo, esto no fue suficiente para que el análisis bioinformático pueda encontrar la etiología de la DI.

En cuanto a los resultados de la secuenciación de exoma en pacientes con DI, se observó que los pacientes con variantes patogénicas o probablemente patogénicas se encuentra dentro del rango de la frecuencia estimada en metanálisis previos (30,6% vs 30-42%) o en otros estudios descriptivos (30,6% vs 27,6-48,8%) [13-16]. Aunque, podría observarse que existen estudios que tienen una mayor tasa de detección de variantes, el cual podría asociarse, entre otras causas, a que las poblaciones donde se realizaron estas pruebas genómicas, tuvieron previamente una mayor proporción de controles sanos.

Las limitaciones del presente estudio son que se presenta una muestra pequeña y que los resultados con variantes de significado incierto fueron superior al grupo sin variantes o con variantes probablemente patogénicas. Es probable que los pacientes con VUS podrían tener resultados con una relación genotipo-fenotipo clara. No obstante, este análisis es una de

las primeras cohortes de pacientes con discapacidad intelectual y estudios de secuenciación de exoma, que realizó una evaluación de la calidad de los fenotipos humanos con esta metodología novedosa.

La descripción clínica como resultado de un entrenamiento en los fenotipos humanos es de importancia para el análisis bioinformático, el cual sirve para poder encontrar con una mayor precisión los cambios genómicos etiológicos. La calidad de los HPO descritos en esta investigación, a través de los coeficientes de IC o del DSI, son similares a lo reportado previamente, mostrando que el personal médico-genetista de nuestro medio tiene las herramientas clínicas suficientes. El uso de las tecnologías de secuenciación masiva es solicitado cada vez con mayor frecuencia por los profesionales expertos en genética, por lo que es relevante difundir datos sobre la calidad de los HPO con la finalidad de optimizar el uso de las nuevas tecnologías genómicas.

REFERENCIAS BIBLIOGRÁFICAS

1. Wolfe K, Stueber K, McQuillin A, Jichi F, Patch C, Flinter F, *et al.* Genetic testing in intellectual disability psychiatry: Opinions and practices of UK child and intellectual disability psychiatrists. *Research Intellect Disabil.* 2018;31(2):273-84. DOI:10.1111/jar.12391
2. Brown KA, Parikh S, Patel DR. Understanding basic concepts of developmental diagnosis in children. *Transl Pediatr.* 2020;9(S1):S9-22. DOI:10.21037/tp.2019.11.04
3. American Psychiatric Association, editor. *Diagnostic and statistical manual of mental disorders: DSM-5-TRTM. Fifth edition, text revision.* Washington, DC: American Psychiatric Association Publishing; 2022.
4. Kadri S. Advances in Next-Generation Sequencing Bioinformatics for Clinical Diagnostics. *Advances in Molecular Pathology.* 2018;1(1):149-66. DOI:10.1016/j.yamp.2018.06.010
5. Hurtado C. Medicina de precisión: conceptos, aplicaciones y proyecciones. *Rev. méd. Clín Las Condes.* 2022;33(1):7-16. DOI: doi:10.1016/j.mmlc.2022.01.002
6. Köhler S, Schulz MH, Krawitz P, Bauer S, Dölken S, Ott CE, *et al.* Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies. *Am J Hum Genet.* 2009;85(4):457-64. DOI: 10.1016/j.ajhg.2009.09.003
7. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, *et al.* The Human Phenotype Ontology in 2017. *Nucleic Acids Res.* 2017;45(D1):D865-76. DOI: 10.1093/nar/gkw1039
8. Bodde NMG, Van Der Kruijs SJM, Ijff DM, Lazeron RHC, Vonck KEJ, Boon PAJM, *et al.* Subgroup classification in patients with psychogenic non-epileptic seizures. *Epilepsy Behav.* 2013;26(3):279-89. DOI: 10.1016/j.yebeh.2012.10.012
9. Rojano E, Córdoba-Caballero J, Jabato FM, Gallego D, Serrano M, Pérez B, *et al.* Evaluating, Filtering and Clustering Genetic Disease Cohorts Based on Human Phenotype Ontology Data with Cohort Analyzer. *JPM.* 2021;11(8):730. DOI: 10.3390/jpm11080730
10. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. *Am J Hum Genet.* 2008;83(5):610-5. DOI: 10.1016/j.ajhg.2008.09.017
11. Hwang T, Atluri G, Xie M, Dey S, Hong C, Kumar V, *et al.* Co-clustering phenome-genome for phenotype classification and disease gene discovery. *Nucleic Acids Res.* 2012;40(19):e146-e146. DOI: 10.1093/nar/gks615
12. Salvador-Carulla L, Bertelli M. 'Mental Retardation' or 'Intellectual Disability': Time for a Conceptual Change. *Psychopathology.* 2008;41(1):10-6. DOI: 10.1159/000109950
13. Abarca-Barriga HH, Vásquez-Sotomayor F. Utilidad diagnóstica de la secuenciación de segunda y tercera generación en pacientes con discapacidad intelectual: revisión rápida. *Psiquiatr Biol.* 2023;30(1). DOI: 10.1016/j.psq.2023.100392.
14. Jo YH, Choi SH, Yoo HW, Kwak MJ, Park KH, Kong J, *et al.* Clinical use of whole exome sequencing in children with developmental delay/intellectual disability. *Pediatr Neonatol.* 2024 Jan 23:S1875-9572(24)00004-4. DOI: 10.1016/j.pedneo.2023.05.015.
15. Ballesta-Martínez MJ, Pérez-Fernández V, López-González V, Sánchez-Soler MJ, Serrano-Antón AT, Rodríguez-Peña LI, *et al.* Validation of clinical exome sequencing in the diagnostic procedure of patients with intellectual disability in clinical practice. *Orphanet J. Rare Dis.* 2023;18(1):201. DOI: 10.1186/s13023-023-02809-z
16. Sánchez-Luque KY, Carpena MX, Karam SM, Tovo-Rodríguez 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.* 2022;790:108428. DOI: 10.1016/j.mrrev.2022.108428