



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

**Rendimiento diagnóstico de la  
secuenciación de exomas completos en  
trío e individual en una cohorte de  
pacientes colombianos con sospecha de  
enfermedad genética de origen  
monogénico**

**Paula Estefanía Rodríguez Alvarino**

Universidad Nacional de Colombia  
Facultad de Medicina, Instituto de Genética Humana  
Bogotá, Colombia

2023

II Rendimiento diagnóstico de la secuenciación de exomas completos en trio e individual en una cohorte de pacientes colombianos con sospecha de enfermedad genética de origen monogénico

---

# **Rendimiento diagnóstico de la secuenciación de exomas completos en trio e individual en una cohorte de pacientes colombianos con sospecha de enfermedad genética de origen monogénico**

**Paula Estefanía Rodríguez Alvarino**

Tesis presentada como requisito parcial para optar al título de:

**Magister en Genética Humana**

Directora:

Marcela Gálvez Bermúdez, MD., MSc.

Codirectora:

Sandra Yaneth Ospina Lagos, MD., MSc.

Línea de Investigación:

Genética Clínica

Universidad Nacional de Colombia

Facultad de Medicina, Instituto de Genética Humana

Bogotá, Colombia

2023

IV Rendimiento diagnóstico de la secuenciación de exomas completos en trio e individual en una cohorte de pacientes colombianos con sospecha de enfermedad genética de origen monogénico

---

*Dedicado a:*

*Mi padre, por su cariño y sabiduría, por desarrollar mi amor por la ciencia y la investigación, y a mi madre por su paciencia, amor y motivación por no dejarme rendir y enseñarme el valor de la responsabilidad.*

VI Rendimiento diagnóstico de la secuenciación de exomas completos en trio e individual en una cohorte de pacientes colombianos con sospecha de enfermedad genética de origen monogénico

---

## Declaración de obra original


Yo declaro lo siguiente:

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

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

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.



---

Estefanía Rodríguez Alvarino

Fecha 03/02/2023

## **Agradecimientos**

A la Dra. Marcela Gálvez, directora de este proyecto de tesis, por ser mi mentora y desarrollar mi gusto por la genética desde que realicé mi pregrado en medicina, por cada una de sus enseñanzas y orientación, por su paciencia y dedicación para desarrollar este proyecto y por toda la confianza depositada en mi para lograr mi título de Magister en Genética.

A la Dra. Sandra Ospina, codirectora de la tesis por transmitir con tanto cariño sus conocimientos y hacerme pensar en la virtud de encontrar sentido en la investigación para beneficiar a mis pacientes.

A la Dra. Sandra Bello por su tiempo y paciencia orientando el análisis de los casos del presente proyecto.

A Diego Otero, por su dedicación y disposición para enseñarme bioinformática para brindarme bases para el análisis de los datos de este proyecto.

A Gencell Pharma, por permitirme desarrollar la tesis con su equipo y facilitar las herramientas para que todo este proyecto se hiciera posible.

Al Dr. Juan José Yunis, Dr. Mauricio Rey, Dr. William Usaquén por todas sus enseñanzas y conocimientos transmitidos, por ser maestros y ser siempre críticos para generar una mejor versión de sus estudiantes.

A mis amigos de maestría Francisco, Fernanda, David, Jair y Luisa por cada seminario, laboratorio y clase que sin su compañía no habrían hecho esta maestría tan feliz y llena de risas.



## Resumen

### **Rendimiento diagnóstico de la secuenciación de exomas completos en trio e individual en una cohorte de pacientes colombianos con sospecha de enfermedad genética de origen monogénico**

Durante los últimos años, la secuenciación del exoma completo ha tenido una gran aceptación para la evaluación de trastornos genéticos. Gracias al acceso a la secuenciación de siguiente o nueva generación (NGS), a las herramientas bioinformáticas y a los expertos profesionales, estas pruebas moleculares representan una herramienta diagnóstica de primera línea para muchas enfermedades genéticas.

El objetivo de este estudio fue determinar el rendimiento diagnóstico de la secuenciación de exoma en trio e individual en un grupo de pacientes colombianos con sospecha clínica de enfermedad genética monogénica, así como el rendimiento diagnóstico por grupos de fenotipos, entre junio de 2020 y mayo de 2021.

Se realizó un estudio descriptivo transversal, retrospectivo, aprobado por el comité de ética en investigación. Se revisaron las historias clínicas para obtener los datos clínicos y fenotípicos, y se realizó secuenciación de exoma completo en trio. El llamado de variantes se realizó utilizando las mejores prácticas de GATK, con la posterior anotación y filtrado utilizando términos de *Human Phenotype Ontology* (HPO) y las variantes se clasificaron utilizando las directrices del Colegio Americano de Genética Médica y Genómica (ACMG, por sus siglas en inglés) de 2015. Se realizaron análisis tanto de exoma individual como análisis de exoma en trio en cada caso. Los casos se consideraron positivos cuando se identificaron variantes patogénicas (P) o probablemente patogénicas (PP), con un mecanismo de herencia que explicaba el fenotipo del paciente. Además, los casos se estratificaron por grupos según el fenotipo clínico, para evaluar el rendimiento específico de la prueba.

Se incluyeron un total de 100 casos. La edad media fue de 8.6 años, y siete casos eran hijos de padres consanguíneos. El rendimiento diagnóstico fue del 29% para el análisis de exomas únicos y del 32% para el análisis en trio. Entre los casos positivos, el 19% presentaba una enfermedad autosómica dominante, el 8% una enfermedad ligada al cromosoma X, el 4% una enfermedad autosómica recesiva y el 1% un trastorno de herencia digénica. Dos casos tenían hallazgos incidentales en los genes *PMS2* y *MSH6*, y otros dos casos tenían condiciones monogénicas concomitantes.

El fenotipo más prevalente fue el retraso del neurodesarrollo en el 47% de los individuos, con un rendimiento del 38.3%. El mayor rendimiento se encontró en la parálisis (71.3%), seguido de la hipoacusia (55.5%), la microcefalia (50%), la hipertensión pulmonar (50%) y las alteraciones hematológicas (50%). Los movimientos anormales tuvieron la tasa de diagnóstico más baja (10%).

Este estudio mostró un rendimiento 3% mayor para el análisis del exoma en trio en comparación con el individual. Esto se explica principalmente por la identificación de variantes *de novo* y mutaciones heterocigotas compuestas confirmadas, que no pueden establecerse en los exomas únicos. El exoma trio mostró el mejor rendimiento en los fenotipos neurológicos, especialmente en individuos con parálisis.

**Palabras clave: Diagnóstico, secuenciación del exoma completo, NGS, pruebas genéticas, genética, fenotipo.**

## Abstract

### **Diagnostic yield of trio and individual Whole Exome Sequencing in a cohort of Colombian patients with a suspected monogenic disease**

During last years, Whole Exome Sequencing (WES) has had great acceptance for the evaluation of genetic disorders. Due to the access to next-generation sequencing (NGS), bioinformatics tools, and professionals' experts, these molecular tests represent a first-tier diagnostic tool for many genetic diseases.

The aim was to compare the diagnostic yield of singleton and trio WES in a Colombian cohort with clinical suspicion of a genetic disease, as well as the yield per phenotype groups, between June 2020 to May 2021.

A cross-sectional descriptive, retrospective study, approved by an ethics board, was carried out. Medical records were reviewed to obtain phenotypic data and trio WES was performed. Variant calling was done using *GATK* best practices, with subsequent annotation and filtering using Human Phenotype Ontology (HPO) terms and variants were classified using the American College of Medical Genetics and Genomics (ACMG) 2015 guidelines. Analysis as singleton exome and trio exome were performed for every case. Cases were considered positive when pathogenic (P) or likely pathogenic (PP) variants were identified, with a mechanism of inheritance explaining the patient's phenotype. Additionally, cases were stratified by groups according to the main clinical phenotype, to evaluate the specific yield of the test.

A total of 100 cases were included. The average age was of 8.6 years, and seven cases were born of consanguineous parents. The diagnostic yield was 29% for singleton and 32% for trio WES. Among positive cases, 19% had an autosomal dominant disease, 8% an X-linked, 4% autosomal recessive, and 1% a digenic inheritance disorder. Two cases had

incidental findings in *PMS2* and *MSH6* genes, and two other cases had concomitant monogenic conditions.

The most prevalent phenotype was neurodevelopmental delay in 47% individuals, with a yield of 38.3%. The highest yield was found for paralysis (71.3%), followed by hypoacusis (55.5%), microcephaly (50%), pulmonary hypertension (50%) and hematologic alterations (50%). Abnormal movements had the lowest diagnostic rate (10%).

This study showed a 3% better yield for trio exome analysis in comparison to singleton. This is mainly explained by the identification of *de novo* variants and confirmed compound heterozygous mutations, which cannot be established in singleton exomes. WES showed the best performance in neurological phenotypes, especially paralysis.

**Keywords: Diagnostics, exome sequencing, NGS, genetic testing, genomics, phenotype.**

# Contenido

	Pág.
<b>Resumen</b> .....	<b>IX</b>
<b>Lista de figuras</b> .....	<b>XV</b>
<b>Lista de tablas</b> .....	<b>XVI</b>
<b>Lista de Símbolos y abreviaturas</b> .....	<b>XVII</b>
<b>Planteamiento del problema</b> .....	<b>1</b>
<b>Justificación</b> .....	<b>5</b>
<b>1 Objetivos</b> .....	<b>7</b>
1.1 Objetivo General .....	7
1.1.1 Objetivos específicos.....	7
<b>2 Marco teórico</b> .....	<b>9</b>
2.1 Un recorrido a través de la historia.....	9
2.2 Enfermedades monogénicas .....	18
2.3 El genoma y sus regiones codificantes: .....	20
2.4 Técnicas de secuenciación de nueva generación:.....	21
2.5 Técnica de secuenciación MGI DNBseq .....	27
2.6 Análisis computacional de datos biológicos .....	28
2.7 Lineamientos para la clasificación de variantes .....	34
2.8 Definición de un estudio de exoma.....	45
2.9 Tipos de estudios de exomas .....	46
2.10 Revisión de la literatura del rendimiento de exomas en la práctica clínica .....	47
<b>3 Metodología</b> .....	<b>51</b>
3.1 Tipo de estudio.....	51
3.2 Diagrama de la investigación.....	51
3.3 Población.....	51
3.4 Criterios de selección.....	52
3.4.1 Criterios de inclusión .....	52
3.4.2 Criterios de exclusión .....	52
3.5 Muestra.....	52
3.6 Recolección de la información.....	52
3.7 Técnicas de análisis para la evaluación de variantes y su interpretación clínica.....	53
3.8 Análisis estadístico.....	58
3.9 Control de errores y sesgos.....	58

XIV Rendimiento diagnóstico de la secuenciación de exomas completos en trio e individual en una cohorte de pacientes colombianos con sospecha de enfermedad genética de origen monogénico

---

3.9.1	Sesgos de selección.....	58
3.9.2	Sesgos de información.....	59
3.9.3	Sesgos de confusión.....	59
<b>4</b>	<b>Consideraciones éticas .....</b>	<b>61</b>
4.1	Declaración de existencia de conflicto de intereses.....	62
4.2	Propiedad intelectual.....	62
<b>5</b>	<b>Resultados .....</b>	<b>63</b>
5.1	Definición de la muestra de estudio.....	63
5.2	Descripción de las características demográficas, clínicas y fenotípicas de la población estudiada. ....	63
5.3	Rendimiento diagnóstico del exoma completo individual y en trio.....	67
5.4	Descripción de casos diagnósticos y variantes identificadas. ....	68
5.4.1	Identificación de variantes nuevas.....	72
5.4.2	Hallazgos Incidentales.....	77
5.5	Rendimiento diagnóstico por características sociodemográficas y fenotípicas..	80
<b>6</b>	<b>Discusión.....</b>	<b>87</b>
6.1	Limitaciones .....	96
<b>7</b>	<b>Conclusiones .....</b>	<b>97</b>
	<b>Bibliografía .....</b>	<b>99</b>

## Lista de figuras

	<b>Pág.</b>
Figura 2-1 Técnicas de secuenciación de segunda generación. ....	25
Figura 2-2. Resumen de análisis clínico y bioinformático de datos biológicos. ....	33
Figura 3-1 Diagrama de la investigación.....	51
Figura 3-2 Proceso de evaluación, clasificación e interpretación de variantes.....	55
Figura 5-1 Mecanismos de herencia de condiciones encontradas en análisis de exomas .....	72
Figura. 5-2. Rendimiento de edad por quinquenios.....	81
Figura 5-3. Rendimiento fenotípico por característica fenotípica. ....	84

## Lista de tablas

	<b>Pág.</b>
<i>Tabla 2-1 Comparación de kits de captura para exoma</i> .....	23
Tabla 2-2 Bases de datos poblacionales .....	35
Tabla 2-3 Bases de datos clínicas.....	35
Tabla 2-4. Bases de datos para secuencias de referencia.....	35
Tabla 2-5 Programas de predicción in silico .....	36
Tabla 2-6 Criterios para la clasificación de variantes patogénicas.....	37
Tabla 5-1. Características demográficas.....	64
Tabla 5-2. Resumen de la edad de la muestra estudiada .....	65
Tabla 5-3. Caracterización fenotípica de la población .....	66
Tabla 5-4. Rendimiento diagnóstico según tipo de análisis de exoma. ....	68
Tabla 5-5. Descripción de casos diagnósticos en exoma trio.....	68
Tabla 5-6 Variantes nuevas.....	73
Tabla 5-7. Casos hallazgos incidentales.....	78
Tabla 5-9. Rendimiento diagnóstico de acuerdo con las características sociodemográficas y clínicas .....	81
Tabla 5-10. Rendimiento diagnóstico por departamentos de Colombia. ....	83
Tabla 5-11 Rendimiento diagnóstico en los subgrupos fenotípicos de retraso psicomotor. ....	85



# Lista de Símbolos y abreviaturas

## Símbolos

<b>Símbolo</b>	<b>Término</b>
A	Adenina
C	Citosina
G	Guanina
Gb	Gigabase
kb	Kilobase
Mb	Megabase
pb	Pares de bases
ng	Nanogramo
T	Timina
ug	Microgramo
>	Cambio de una base nitrogenada por otra
Ala	Alanina
Arg	Arginina
Asn	Asparagina
Asp	Ácido aspártico
Cys	Cisteína
Gln	Glutamina
Glu	Ácido glutâmico
Gly	Glicina
His	Histidina
Ile	Isoleucina
Leu	Leucina
Lys	Lisina
Met	Metionina
Phe	Fenilalanina
Pro	Prolina
Ser	Serina
Thr	Treonina
Trp	Triptofano
Tyr	Tirosina
Val	Valina

## Abreviaturas

<b>Abreviatura</b>	<b>Término</b>
<i>ACMG</i>	American College of Medical Genetics
<i>ADN</i>	Ácido desoxirribonucleico
<i>ARN</i>	Ácido ribonucleico
<i>BAM</i>	Binary Alignment Map
<i>ClinVar</i>	Clinically Relevant Variation
<i>CNV</i>	Variante en el número de copias
<i>cRNA</i>	ARN complementario
<i>DECIPHER</i>	Databases of genomic variation and Phenotype in Humans using Ensembl Resources
<i>DNB</i>	DNA-nanoballs, Nanobolas de ADN
<i>GATK</i>	Genome Analysis Toolkit
<i>GRCh37</i>	Genome Reference Consortium Human Build 37
<i>HPO</i>	Human Phenotype Ontology
<i>InDel</i>	Inserciones/deleciones
<i>MAF</i>	Frecuencia del alelo menor
<i>NCBI</i>	Centro Nacional de Información Biotecnológica
<i>NGS</i>	Next Generation Sequencing, Secuenciación de Nueva Generación
<i>NIH</i>	Instituto Nacional de Salud de Estados Unidos
<i>OMIM</i>	Online Mendelian Inheritance in Man
<i>OMS</i>	Organización Mundial de la Salud
<i>P</i>	Variante patogénica
<i>PCR</i>	Reacción en cadena polimerasa
<i>PP</i>	Variante probablemente patogénica
<i>SAM</i>	Sequence Alignment Map
<i>SIVIGILA</i>	Sistema de Vigilancia en Salud Pública
<i>SNP</i>	Polimorfismo de nucleótido simple
<i>SNV</i>	Variante de nucleótido simple
<i>vcf</i>	Variant Call Format
<i>VSI</i>	Variante de significado incierto
<i>WES</i>	Whole Exome Sequencing

## Planteamiento del problema

La llegada de las técnicas de secuenciación de siguiente o nueva generación NGS (Next Generation Sequencing) han acelerado la investigación en medicina, permitiendo identificar una gran cantidad de enfermedades de origen genético y haciendo que en la actualidad se dispongan de más de 2000 pruebas de diagnóstico y aproximadamente 350 productos basados en biotecnología(1). Hasta los años 70s las únicas técnicas disponibles para acceder a la visualización del ADN eran las de citogenética, y con estas solo se lograba documentar alteraciones cromosómicas como aneuploidías, poliploidías, traslocaciones y grandes deleciones o duplicaciones superiores a 5 megabases, lo cual generaba un conocimiento muy limitado de las enfermedades que parecían transmitirse de una generación a otra, debido a que no era posible establecer cambios de un único nucleótido. Fue hasta 1977 que Frederick Sanger desarrolló el método de secuenciación que lleva su apellido, Sanger, con el cual logró la primera secuenciación del primer genoma completo de un bacteriófago y se identificaron por primera vez pequeños cambios en la secuencia polinucleotídica del ADN con el potencial de explicar enfermedades familiares(2). La secuenciación Sanger le dio el premio nobel a Frederick Sanger junto con Walter Gilbert en 1980 y fue la base para realizar el Proyecto Genoma Humano que inició en 1990, cuyos resultados iniciales fueron publicados en el 2001, con el primer borrador del 90% de la secuencia de ADN humano, y en 2003, una segunda publicación con el 92% del genoma(2). Es de destacar que este proyecto, que tuvo un costo de 2.700 millones de dólares, terminó cuatro años antes de lo esperado gracias a la llegada de técnicas de secuenciación de nueva generación que permitieron terminar el proyecto antes de los quince años estimados inicialmente(3), sin embargo, fue hasta el 2022 que se publicó el 100% del genoma a través del consorcio T2T (Telómero a Telómero), debido al uso de técnicas de secuenciación de lecturas largas lo cual permitió conocer la secuencia de regiones altamente repetitivas como los satélites de los cromosomas acrocéntricos(4).

La identificación de la secuencia consenso de ADN en el ser humano permitió descubrir variantes que causan enfermedades monogénicas y en la actualidad, según registros de OMIM (Online Mendelian Inheritance in Man) se conoce de 6.575 enfermedades con alteraciones genéticas documentadas(5). Si bien, como regla general, la mayoría de las enfermedades genéticas se consideran enfermedades huérfanas por su baja prevalencia y su pobre interés por parte de las compañías farmacéuticas para ser estudiadas, pues solo existe tratamiento disponible para menos del 3% de estas enfermedades(6), están empezando a cobrar una mayor relevancia, debido a que la mortalidad por enfermedades infecciosas se ha reducido, y las enfermedades crónicas al igual que las enfermedades monogénicas han adquirido un papel protagónico como causa de morbimortalidad.

Lo anterior, junto con la llegada de técnicas de secuenciación de nueva generación (NGS), ha permitido la identificación de nuevos genes causales de enfermedad, ha dilucidado su fisiopatología y ha permitido el desarrollo de tratamientos específicos e incluso terapia génica.

En efecto, estos antecedentes han conducido a un aumento considerable por parte de los médicos especialistas en la solicitud de estudios de secuenciación de tipo exoma, que consiste en el análisis las regiones codificantes del material genético de un individuo, con el fin de lograr documentar la etiología genética de su condición y brindarle mejores opciones terapéuticas y de rehabilitación; sin embargo, es importante reconocer que este tipo de estudios arrojan una gran cantidad de datos y variantes cuya interpretación biológica persiste desconocida, logrando un rendimiento diagnóstico que oscila entre el 20-50%, con un rendimiento adicional del 10% si se reanalizan en un periodo de 18-36 meses(7,8), donde aquellas enfermedades con manifestaciones desde el nacimiento logran mayores rendimientos diagnósticos respecto a aquellas que aparecen tardíamente en la infancia o en la adultez(9). Los exomas han tenido un interés particular ya que se conoce que las mutaciones en regiones exónicas o codificantes ocurren en el 85% de las enfermedades monogénicas(10) y solo se analiza entre el 1.5% y el 2% del total del genoma.

En el caso de Colombia, no existe información respecto al rendimiento diagnóstico de los exomas en patologías genéticas, sin embargo, se cree que al conocer la secuencia de las

regiones codificantes y las regiones de la unión intrón-exón se logrará identificar con precisión el origen genético de la mayoría de las enfermedades y se le da un alto valor a esta prueba esperando lograr rendimientos diagnósticos muy altos, pero lo que se ha visto a nivel mundial, es que el rendimiento diagnóstico parece ser más bajo que el esperado, donde los rendimientos reales no alcanzan el 50%. También surgen algunas limitantes, entre ellas, no hay suficiente información en la literatura para interpretar las variantes y su asociación con la aparición de fenotipos patológicos; tampoco se conoce la frecuencia poblacional de las variantes en individuos colombianos debido a que no se cuenta con una base de datos única que reúna la información de Colombia obtenida por los laboratorios e instituciones genéticas a nivel nacional; también es probable que haya desconocimiento de la metodología de análisis del exoma entre los profesionales, y además se ha encontrado que los médicos solicitantes no son ampliamente descriptivos de los fenotipos y esto limita los filtros bioinformáticos y la interpretación de las variantes, generando un aumento en las variantes de significado incierto(11). Por otro lado, en algunos estudios se ha observado una discordancia en la interpretación de variantes entre laboratorios del 12.8% por las diferencias en la clasificación de las variantes a pesar de la utilización de las guías del Colegio Americano de Genética Médica (ACMG) (12).

A la vez, cuando se solicita un estudio de exoma, surgen algunos interrogantes, como por ejemplo, determinar si ¿es superior el estudio de exoma en trio (probando con sus padres) respecto al estudio exclusivo en el probando? Y si ¿existen grupos fenotípicos donde sea posible lograr un mejor rendimiento diagnóstico en los exomas? Resolver estos interrogantes, será de interés para la comunidad clínico-científica, pues se mejoraría la aproximación clínica de los pacientes para solicitarles un estudio de exoma completo y aumentaría su rendimiento diagnóstico en población colombiana. Además, se brindaría la posibilidad de realizar diagnósticos tempranos y oportunos que impacten en la calidad de vida de los pacientes con enfermedades raras de origen genético y sus familias.



## Justificación

Existen cerca de 7000 enfermedades raras(13) que afectan entre el 3.5 al 5.9% de la población mundial (263 a 446 millones de personas afectadas en el mundo), de las cuales el 71.9% de ellas son de origen mendeliano(14). El 75% tienen manifestaciones en la infancia, y son responsables del 35% de la mortalidad en el primer año de vida en países desarrollados(9). El tiempo estimado para el diagnóstico molecular de una enfermedad genética se encuentra entre 1 y 10 años(15), y este retraso impide que los pacientes reciban medidas terapéuticas y de rehabilitación específicas y oportunas, así como que sus familiares entren en programas preventivos y reciban asesoramiento genético oportuno. La llegada de la secuenciación masiva en paralelo está cambiando los modelos diagnósticos en genética, permitiendo aproximarnos cada vez más al origen genético de las enfermedades, los mecanismos moleculares que conducen a su fisiopatología y sus desenlaces a nivel molecular; sin embargo, los médicos se enfrentan a la decisión de elegir el método más preciso y costo-efectivo, lo que ha conducido a un incremento en la solicitud de estudios de secuenciación de tipo exoma, como una de las primeras alternativas de aproximación diagnóstica para una amplia variedad de patologías(15).

No obstante, los estudios de exoma presentan algunos retos como la interpretación de los datos biológicos y bioinformáticos obtenidos, pues al obtener información sobre el 1.5% del genoma se encuentran aproximadamente unas 20.000 variantes genéticas(9), donde gran parte de ellas aún permanecen sin interpretación clínica generando una gran limitante en cuanto a la capacidad de establecer si una variante es responsable de determinado fenotipo, y esto conlleva a la falta de un diagnóstico conclusivo, incertidumbre en los médicos tratantes, los pacientes y sus familias, y múltiples dificultades para realizar un asesoramiento genético adecuado(16).

Las dificultades en la interpretación de los datos biológicos y bioinformáticos sumados a la falta de información de variantes en la población colombiana generan la necesidad de desarrollar guías para el manejo de la enorme cantidad de información genética generada.

En Colombia la solicitud de un estudio de exoma se encuentra cubierta en el plan de beneficios en salud desde el año 2018, sin embargo, involucra necesariamente la valoración por un médico especialista e incluso por un subespecialista cuando no hay un genetista disponible, esto genera retraso en la atención y en la posibilidad de brindar un diagnóstico oportuno. Además, se deben considerar algunos obstáculos que surgen durante el abordaje clínico como la regulación que ejercen las empresas aseguradoras cuando se realiza solicitud de estudios de alto costo, dificultando aún más el abordaje diagnóstico del paciente.

Adicionalmente, surgen algunos desafíos tecnológicos en Colombia, como los altos costos de las pruebas, la poca disponibilidad de laboratorios de diagnóstico molecular, la formación académica en el análisis e interpretación de estudios como el exoma que son técnicas relativamente nuevas y la falta de bases de datos de población colombiana.

Este será el primer estudio en Colombia que haga una aproximación al rendimiento diagnóstico del estudio de exoma y además compare si es beneficioso hacer los análisis en trio respecto a los análisis en solo el probando. Adicionalmente, permitirá aportar en el conocimiento de nuestra población en cuanto al rendimiento diagnóstico por grupos fenotípicos y compararlos con lo reportado en la literatura.



# 1 Objetivos

## 1.1 Objetivo General

Determinar el rendimiento diagnóstico de la secuenciación de exoma trio e individual en un grupo de pacientes colombianos con sospecha clínica de enfermedad genética y su rendimiento por grupos fenotípicos.

### 1.1.1 Objetivos específicos

- Describir las características demográficas y clínicas de la población de estudio.
- Determinar el rendimiento diagnóstico del exoma individual para la identificación de variantes patogénicas y probablemente patogénicas en el grupo de pacientes estudiados.
- Determinar el rendimiento diagnóstico del exoma trio para la identificación de variantes patogénicas y probablemente patogénicas en el grupo de pacientes estudiados.
- Comparar los rendimientos diagnósticos del análisis en trio con el análisis individual de los exomas.
- Identificar el rendimiento diagnóstico de la secuenciación de exoma por grupos fenotípicos.
- Proponer indicaciones para la solicitud de exomas en el abordaje de pacientes con sospecha de patologías de origen genético monogénicas.

8 Rendimiento diagnóstico de la secuenciación de exomas completos en trio e individual en una cohorte de pacientes colombianos con sospecha de enfermedad genética de origen monogénico

---

## **2 Marco teórico**

### **2.1 Un recorrido a través de la historia**

El estudio del ADN ha sido uno de los principales temas de investigación durante los últimos setenta años. Algunos hitos de la historia han sido determinantes para el rápido avance que ha tenido la genética humana como disciplina, también para entender la evolución, el poblamiento de las culturas, las enfermedades, el desarrollo de técnicas diagnósticas y recientemente algunos tratamientos bajo ingeniería genética. La historia hasta la llegada de la genómica comprende: En 1953, Watson y Crick describieron el modelo de doble cadena helicoidal, complementaria y antiparalela del ADN, lo que los hizo merecedores del premio Nobel en 1962, a partir de los datos cristalográficos desarrollados por Rosalind Franklin y Maurice Wilkins(1). En 1977, Frederick Sanger desarrolló el método de secuenciación Sanger usando cuatro reacciones en tubos separados con desoxinucleótidos radiomarcados diferentes para cada tubo, dideoxinucleótidos, ADN polimerasa y cebadores(2), es decir cadenas cortas de ADN monocatenario que son complementarias a la hebra molde y permiten iniciar la actividad de la ADN polimerasa para secuenciar. En 1986 se realizó la automatización de la técnica, Prober describió la técnica con cuatro dideoxinucleótidos de terminación fluoromarcados con succinil fluoresceína y un espectro de emisión diferente para desarrollar la reacción en un mismo ensayo, y Applied Biosystems en este mismo año lanzó el primer secuenciador de ADN automatizado (ABI370A) basado en esta técnica(2). En 1983 con la aparición de la reacción de cadena polimerasa PCR desarrollada por el Dr. Mullis(17), y la llegada de la electroforesis capilar, la técnica mejoró incluso hasta ser la base crucial para el Proyecto de secuenciación del Genoma Humano. En 1996 se lanzó el ABI Prism 310, el primer secuenciador de ADN asociado a electroforesis capilar(2) logrando un procesamiento de 1000 bases en menos de una hora.

Posterior a la secuenciación Sanger, surgió la secuenciación de nueva generación o NGS, que consiste en técnicas desarrolladas para analizar el ADN de forma masiva y en paralelo, es decir grandes cantidades de ADN secuenciadas al mismo tiempo, reduciendo los tiempos de trabajo de la técnica y a su vez los costos. Las primeras técnicas de secuenciación de nueva generación tienen tres pasos fundamentales: preparación de librerías, que incluyen fragmentación y ligación de adaptadores, amplificación de la librería por PCR y secuenciación con diferentes metodologías(18) como pirosecuenciación, secuenciación por síntesis, por ligación, conducción iónica, entre otras.

En el año 2003 Roche introdujo al mercado el primer secuenciador de siguiente generación (NGS) que utilizaba la técnica de pirosecuenciación, cuyo principio se basa en la medición de una señal quimioluminiscente generada por una reacción enzimática tras la liberación de pirofosfato durante la polimerización del ADN(19). En 2006 Solexa (hoy Illumina) lanzó el secuenciador GenomeAnalyzer basado en la técnica de secuenciación por síntesis que utiliza nucleótidos fluoromarcados complementarios a la hebra molde durante la reacción de polimerización por complementariedad(19). En 2007 Applied Biosystems lanzó el secuenciador SOLiD (Sequencing by Oligo Ligation and Detection), que utiliza una metodología como lo describe su nombre en la cual se aprovecha la especificidad de las ADN ligasas para ligar oligonucleótidos fluorescentes a primers de una manera dependiente de la secuencia(2). En 2010 Pacific BioSciences revolucionó las técnicas de secuenciación lanzando la secuenciación en tiempo real de una sola molécula(18), en 2014 Oxford Nanopore Technologies comercializó el MiniION que utiliza la secuenciación de hebra única basada en sensores iónicos en nanoporos detectando cambios de voltaje en cada posición de la hebra de ADN(18) y en 2016 en China, el Beijing Genomics Institute (BGI) lanzó una plataforma NGS llamada BGISEQ-500 donde se realiza un proceso de amplificación basado en replicación en círculo de los fragmentos de ADN, generando nanobolas de ADN o DNB (DNA-nanoballs) las cuales son ancladas a un array y posteriormente secuenciadas por ligación a través de un método sonda-ancla.(20)(21).

En la era de la genética y genómica, todos estos avances en las tecnologías de secuenciación han favorecido la investigación en biología molecular genética, y durante

las tres últimas décadas han surgido algunos consorcios para comprender la estructura, la función y la variabilidad del genoma humano; dentro de los más importantes encontramos:

### Proyecto Genoma Humano

Este proyecto inició en el año de 1990 con la colaboración del Instituto Nacional de Salud (NIH) y el Departamento de Energía de Estados Unidos (DOE) en colaboración con varios países, y financiado por entidades públicas y privadas cuyo objetivo era determinar la secuencia nucleotídica del genoma humano, mapear los genes que lo componen, generar mapas de enlace para explicar enfermedades genéticas y discutir las implicaciones éticas del conocimiento del genoma(22). Este proyecto fue dirigido por James Watson y posteriormente por Francis Collins y su desarrollo fue planeado a 15 años. En el 2001, después de 11 años, se obtuvo el primer borrador de la secuencia nucleotídica haploide con el 90% del genoma humano secuenciado, y que fue publicada en las revistas *Nature* y *The Lancet*(3), con la identificación de alrededor de 20.500 genes. Para el 2003 se publicó cerca del 92% del genoma humano y más tarde en 2007 *Venter y colaboradores* publicaron la secuencia diploide de un individuo. El proyecto del genoma humano tardó menos tiempo del estimado, debido a que los cinco últimos años se realizaron con tecnologías de secuenciación de nueva generación desplazando la secuenciación Sanger, disminuyendo los tiempos de secuenciación y mejorando el rendimiento(1).

Cabe aclarar que fue hasta el año 2022 que el consorcio T2T (Telómero a Telómero) publicó la totalidad de la secuencia del genoma humano, logrando completar el 8% de la secuencia faltante, principalmente porque se solucionaron limitaciones técnicas, gracias al uso de tecnologías de secuenciación de lecturas largas de Oxford Nanopore y PacBioHifi que alcanzan lecturas de un millón y veinte mil pares de bases respectivamente, lo cual permitió mejorar la secuenciación en regiones de heterocromatina altamente repetitivas, entre ellas regiones centroméricas, pericentroméricas, los satelitales de los brazos cortos de los cinco cromosomas acrocéntricos y regiones subteloméricas principalmente, reduciendo las brechas generadas por las técnicas de lecturas cortas. En la actualidad se sabe que el genoma humano comprende 3.055 millones de pares de bases y 19.969 genes, y esta última aproximación permitió conocer 200 millones de pares de bases que contienen 1956 genes y se prevé que 99 codifican proteínas(4).

### Proyecto HapMap:

Estableció los mapas de haplotipos en diferentes poblaciones con ascendencia asiática, europea y africana a través del análisis de polimorfismos de nucleótido simple (SNP) y tuvo una duración de siete años entre el 2002 y 2009(23)(24). Un haplotipo es un grupo de variantes de ADN que se heredan en conjunto o en bloque por su proximidad, y que se recombinan poco durante la meiosis para heredarse. El proyecto HapMap, tuvo como objetivo desarrollar un mapa de haplotipos para conocer la variabilidad del genoma humano (polimorfismos) y en consecuencia poder usarlos como marcadores para realizar estudios de asociación, encontrar genes relacionados con enfermedad, documentar dianas para la respuesta a fármacos, entender la recombinación y la variación estructural(25).

Se identificaron alrededor de 10 a 11 millones de SNP, cuya frecuencia es mayor al 5% en la población estudiada, y hay un SNP cada 300 pares de bases, sin embargo el Proyecto HapMap permitió establecer los bloques haplotípicos donde se heredan grupos de 30 a 40 SNPs de forma conjunta entre poblaciones, y también los TagSNP que son las regiones que le dan la variación a cada bloque entre poblaciones y permiten reducir la cantidad de SNPs de interés para realizar estudios de asociación, para un total de 500.000 TagSNP(25,26).

### Proyecto 1000 Genomas

Inició en el 2008 en Reino Unido, y es una colaboración internacional del Wellcome Trust Sanger Institute (Reino Unido), el Instituto Max Planck de Genética Molecular (Alemania), el Beijing Genomics Institute (China) y el National Human Genome Research Institute (Estados Unidos), cuyo objetivo fue inicialmente secuenciar 1.000 genomas humanos de diversas poblaciones sanas de los cinco continentes, para caracterizar la variabilidad genética de los individuos que ocurre con una frecuencia superior al 1%, utilizando plataformas de NGS y cuyos resultados fueran de acceso gratuito para futuras investigaciones(27).

Este proyecto estuvo compuesto por tres fases, una fase piloto de 180 muestras del proyecto HapMap secuenciadas con una profundidad de 2 a 4x, una fase 1 con 1092

individuos, una fase 2 de manejo de datos y una fase 3 con 2504 individuos. El informe final se publicó en el año 2015, en total se incluyeron 26 poblaciones, se generaron 95 terabases de información, se identificaron 88.3 millones de variantes incluyendo variantes de nucleótido simple, InDel, variantes en el número de copias y grandes variantes estructurales(28).

En la actualidad, la plataforma del proyecto está disponible con más de 3.000 muestras incluidas, y se utiliza para diferentes enfoques como el filtrado de variantes en estudios diagnósticos de NGS según su frecuencia poblacional, análisis de poblaciones, estudios de asociación, entre otros(28).

#### Proyecto del Consorcio Enciclopedia de los elementos del ADN (ENCODE)

Se diseñó en 2007 para explorar y entender todos los elementos funcionales codificados en el genoma humano, que incluyen regiones transcritas de ARN, de codificación de proteínas, sitios de unión a factores de transcripción, estructura de la cromatina y sitios de metilación, a través del estudio de 147 tipos celulares diferentes(29).

Se encontró que el 80.4% del genoma participa en al menos un evento bioquímico asociado con el ARN o la cromatina en al menos un tipo de célula. El 62% del genoma representa diferentes tipos de ARN, también el 56.1% representa regiones para modificaciones de histonas, el 15.2% contiene cromatina abierta y el 4.6% corresponde a motivos de unión de factor de transcripción(29).

Todos estos consorcios y proyectos han permitido comprender la variabilidad de las poblaciones, la evolución y entender las enfermedades a nivel molecular, además de reconocer el papel protagónico que tiene la bioinformática para el análisis e interpretación de estos datos genómicos en la era de la genética molecular.

La bioinformática es una disciplina que se encarga del almacenamiento, análisis e interpretación de datos biológicos(30). Sin embargo, su llegada a la medicina ha hecho que se hable de bioinformática biomédica, que es una aproximación que permite la unión entre la biología molecular y la clínica genómica, para entender enfermedades, aplicar

diagnósticos y ofrecer tratamientos de medicina personalizada(30). El Instituto Nacional de Salud (NIH) de Estados Unidos ha sido el promotor de la era de medicina genómica que propone y facilita a la comunidad científica el uso de información genómica como parte de la atención clínica para revolucionar este campo(30)(31).

El acceso a la información implica la generación de bases de datos de referencia que permiten el análisis comparativo de los datos del caso en estudio con los datos de referencia contenidos en las bases de datos, para filtrar las variantes, clasificar aquellas ubicadas en genes que explican el desarrollo de enfermedad y definir su representación en la población general. Esto implica el acceso gratuito(32) que ofrecen los grandes consorcios del mundo al genoma humano como, ENCODE, 1000 Genomas, entre otros(31). También la disponibilidad de bases de datos para evaluar el impacto de las variantes en la función de las proteínas, sus consecuencias celulares y clínicas ofrecidas en su gran mayoría por el Instituto Europeo de Bioinformática (EMBL-EBI) o el Centro Nacional de Información Biotecnológica (NCBI) donde se encuentran herramientas como BLAST, Ensembl, GenBank entre muchas otras(2), y además se encuentran disponibles bases de datos para la interpretación clínica que vinculan los datos moleculares con los clínicos, que requirieron el desarrollo de términos médicos para integrar la búsqueda y mejorar la precisión de la aproximación clínica como la ontología del fenotipo humano (términos HPO) y que posteriormente facilitan el uso de herramientas como ClinVar, OMIM, DECIPHER y otras(32).

La disponibilidad de bases de datos abiertas de referencia ha sido de gran utilidad en la clínica genómica para el análisis de estudios diagnósticos a partir de NGS como paneles, exomas y genomas en la clasificación de variantes. Los laboratorios de genética que realizan diagnóstico utilizan diversas herramientas durante el análisis de NGS desde el ensamblaje de las secuencias o el alineamiento de las lecturas con el genoma de referencia, también para el llamado, la anotación, filtrado y asignación de patogenicidad de las variantes incluyendo bases para revisar las frecuencias alélicas, realizar predicciones de la pérdida de función in-silico, establecer conservación, identificar dominios proteicos, consultar bancos de genes asociados a enfermedades o fenotipos, revelar mecanismos de herencia y consultar variantes de importancia clínica previamente reportadas.



Dentro de las herramientas de mayor relevancia para en análisis y clasificación de variantes en estudios diagnósticos se encuentran:

BLAST: *Basic Local Alignment Search Tool*, esta herramienta permite cargar y realizar el alineamiento de secuencias nucleotídicas o de aminoácidos para compararlas con secuencias de la base de datos de diversas especies para encontrar regiones de similitud. No se usa en diagnóstico, pero permite evaluar relaciones evolutivas y funcionales para establecer familias de genes por homología(33).

Ensembl: Es un buscador de genomas de vertebrados que permite anotar genes, calcula alineaciones, recopila información de asociación con enfermedad(34).

NIH Nucleotide Database: Es un repositorio que almacena secuencias nucleotídicas de diferentes bases de datos, entre ellas, GeneBank, RefSeq, entre otras(35).

GenBank: *Genetic Sequence Data Bank*: Consiste en una base de datos que almacena una colección anotada de todas las secuencias de ADN públicamente disponibles. Hace parte de una colaboración internacional entre el banco de datos de ADN de Japón (DDBJ), el Laboratorio Europeo de Biología Molecular (EMBL) y GeneBank de NCBI, las cuales actualizan sus datos a diario para mantenerlos abiertos a la comunidad científica(35).

UniProt: *Universal Protein Resource* Es un consorcio que contiene una colección de secuencias de proteínas donde se pueden consultar diferentes características de las secuencias proteicas como dominios funcionales y estructurales, isoformas, variantes reportadas, entre otros. UniProt es una colaboración entre el Instituto Europeo de Bioinformática (EMBL-EBI), el Instituto Suizo de Bioinformática (SIB), y el Recurso de Información de Proteínas (PIR) (36).

RefSeq: *Reference Sequence*, es una colección de secuencias de proteínas, transcritos (ARN) y ADN genómico seleccionado y no redundante producido por NCBI. RefSeq es una referencia estable para la anotación del genoma, la identificación y caracterización de genes, el análisis de mutaciones y polimorfismos, los estudios de expresión y los análisis comparativos, y es posible ingresar a dicha colección a través de *nucleotide* y *protein* en NCBI (35).

GnomAD: Anteriormente era el Exome Aggregation Consortium (ExAC), es una base de datos que contiene secuencias de exomas y genomas de poblaciones de todo el mundo,

en la actualidad no contiene cohortes para enfermedades pediátricas ni secuencias de familiares en primer y segundo grado para funcionar como referencia de frecuencias alélicas, sin embargo, puede tener casos no eliminados(37). Adicionalmente, cabe mencionar que los datos contenidos en GnomAD tiene secuencias obtenidas a partir de diferentes kits de captura y técnicas de secuenciación por lo que la cobertura es variable entre individuos, sin embargo, la base de datos lo tiene en cuenta al calcular la frecuencia de las variantes(37).

dbSNP: Database of Short Genetic Variations incluye un banco de variantes anotadas y no redundantes de nucleótido simple SNV, InDels de tamaño pequeño y microsatélites, brinda información de la frecuencia poblacional y datos genotípicos, tanto para variantes de tipo polimorfismo como aquellas clínicamente relevantes(35), contiene un número identificador *rs* (*reference SNP ID*) para las variantes reportadas o contenidas que permite identificar la variante a pesar del ensamblaje genómico. Al igual que GnomAD cuenta con muestras de individuos no relacionados, y combina información de ExaAC/GnomAD, 1000 Genomas, HapMap e información de laboratorios públicos y organizaciones, en la actualidad contiene alrededor de dos mil millones de SNP enviados por remitentes(38).

ClinVar: *Clinically Relevant Variation* Es un repositorio público del NIH que contiene información de las variantes reportadas en humanos y su clasificación en el tiempo para relacionarlas con el estado de salud humana(39), es decir provee información de la clasificación de las variantes según su patogenicidad. Cuenta con un sistema de clasificación para las variantes que envían y someten diferentes remitentes y evalúa su respaldo en diversas fuentes. Más adelante, al tratar la anotación clínica de las variantes se describirá el sistema de puntuación de esta base de datos.

OMIM: *Online Mendelian Inheritance in Man*, es un compendio creado en los años sesenta por el Dr. Victor A. McKusick donde se reúnen todos los trastornos mendelianos conocidos e información de más de 15.000 genes del genoma humano, centrándose en la documentación de las relaciones genotipo-fenotipo(40)

HGMD: *Human Gene Mutation Database* es un compendio de variantes genéticas conocidas responsables de enfermedad mendeliana en humanos, también incluye polimorfismos de relevancia clínica asociados a enfermedades ampliamente descritos en

la literatura. No incluye variantes somáticas ni en genoma mitocondrial, en general tampoco incluye variantes silentes a menos de que estén descritas como mecanismo de enfermedad(41). Para el 2020 contaba con más de 289.000 variantes descritas en más de 11.000 genes manualmente curadas(41).

CGD *Clinical Genomic Database*: Es una base de datos de genómica clínica del NIH que incluye todas las condiciones genéticas conocidas, incluye información sobre el gen, los mecanismos de herencia, la edad de presentación (pediátrica o adulta), una descripción clínica y las posibles intervenciones. Es exclusiva para condiciones monogénicas y no cuenta con descripciones de genes contiguos, tampoco incluye asociaciones genéticas o factores de susceptibilidad(42).

DECIPHER *Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources*: Es una herramienta diseñada para la interpretación de variantes genómicas tanto de nucleótido simple como en el número de copias, permite la interacción con otras bases de datos que permiten realizar un análisis detallado de las variantes, utiliza la secuencia de referencia a través de Ensembl para definir los genes involucrados en una variante estructural como CNV o para predecir la posición de una variante puntual en la proteína en sus diferentes dominios(43).

HPO *Human Phenotype Ontology*: Es un proyecto desarrollado por el Hospital Universitario Charité de Berlín, el Laboratorio Nacional Lawrence Berkeley, la Universidad de Cambridge y la Universidad de Oregón que surgió en el 2008 con el fin de desarrollar un esquema para vincular enfermedades humanas con modelos animales mediante la anotación de fenotipo basada en ontología. Esta base de datos contiene la ontología de anomalías fenotípicas médicamente relevantes encontradas en enfermedades humanas, anotaciones de fenotipos de enfermedades y los algoritmos que operan en estos (44), donde incluye signos y síntomas, parámetros bioquímicos e imagenológicos. Para el 2020 contenía 15.247 términos clínicos con un *HP ID*, que se conectan desde fenotipos generales (ontologías generales) hasta fenotipos muy específicos (subontologías) como modificadores de edad de inicio y la frecuencia de las características y permite vincularlos con genes relacionados con el fenotipo clínico. Al ser una alternativa para unificar términos clínicos se utiliza para desarrollar aplicaciones bioinformáticas en diagnóstico y en investigación(44). En diagnóstico es particularmente útil para filtrar genes candidatos a enfermedad en un enfoque de fenotipo a genotipo o en búsqueda de diagnósticos

diferenciales y en investigación se utiliza principalmente para encontrar genes nuevos asociados a enfermedad(45).

## 2.2 Enfermedades monogénicas

Una enfermedad monogénica es aquella que ocurre por alteraciones en un único gen y se transmite usualmente por mecanismos de herencia Mendeliana. De acuerdo con la base de datos de OMIM, al día 10 de febrero de 2023 existen 6.575 enfermedades con un origen genético documentado(5) y se ha encontrado que son más frecuentes en países en vía de desarrollo, particularmente en áreas rurales por el aumento de uniones consanguíneas, relacionadas con enfermedades de herencia mendeliana autosómica recesiva. Según la Organización Mundial de la Salud (OMS), la mortalidad infantil en países en desarrollo se debe mayormente a enfermedades infecciosas y esto enmascara el puesto que ocupan las enfermedades congénitas hereditarias en la mortalidad infantil y la preocupación de los gobiernos por su detección temprana(46).

Gran parte de las enfermedades monogénicas pertenece al grupo de las enfermedades raras o huérfanas, cuya definición varía entre países y continentes. En Estados Unidos una enfermedad rara se define como aquella que afecta a menos de 200.000 personas(47), diferente de la Unión Europea donde su definición es aquella enfermedad que afecta a menos de 1 en 2.000 personas, en Rusia menos de 1 en 10.000, en Turquía menos de 1 en 100.000, en el Reino Unido utilizan otros criterios como aquellas que ocurren en menos de 500 ciudadanos al año y requieren atención médica especializada(6). En Colombia según la ley 1438 de 2011 las enfermedades huérfanas son aquellas “crónicamente debilitantes, graves, que amenazan la vida y con una prevalencia menor de 1 por cada 5.000 personas, comprenden, las enfermedades raras, las ultrahuérfanas y olvidadas”(48).

El costo de las enfermedades huérfanas para los sistemas de salud es muy elevado y genera un gran impacto económico. Por ejemplo, en Estados Unidos el costo anual por tratamientos con fármacos para enfermedades huérfanas es de 137.782 dólares por paciente(47), sin tener en cuenta los costos que implican el manejo integral de estas enfermedades, como acceso a especialistas, rehabilitación, hospitalizaciones, seguimiento periódico, entre otros. En Colombia, el Decreto 1954 de 2012 dicta las disposiciones para

implementar el sistema de información con enfermedades huérfanas. En este decreto se establece el reporte obligatorio de estas enfermedades al Sistema de Vigilancia en Salud Pública, SIVIGILA(49) y que la Administradora de los Recursos del Sistema General de Seguridad Social en salud (ADRES) dispondrá del registro para asignar los recursos empleados en el tratamiento de los pacientes. En 2017 *Mateus y colaboradores* publicaron el primer registro nacional de enfermedades raras, identificando 13.215 pacientes con 653 enfermedades raras. Los autores describen que el 69% de los pacientes tenían entidades de origen genético(50), sin embargo, también se encontró un subregistro en las zonas de los centros urbanos del país. De acuerdo con la Resolución 0023 de 2023 del Ministerio de Salud y Protección social donde se actualiza el listado de enfermedades huérfanas en Colombia, actualmente hay 2.247 enfermedades huérfanas registradas para realizar seguimiento(51) y según el Informe de Enfermedades Huérfanas – Raras, Colombia del primer semestre del 2022 del Instituto Nacional de Salud (INS), se notificaron al SIVIGILA 6.657 casos de enfermedades huérfanas en Colombia, de los cuales el 61.2% (4.075) vive en Antioquia, Bogotá y Cali (52).

En presentación de evento del INS del informe de Enfermedades Huérfanas – Raras en Colombia 2022 hasta el periodo epidemiológico XII disponible en la página web del INS, se habían notificado 13.227 casos de enfermedades huérfanas, donde las enfermedades que comprometen el sistema nervioso central son las más frecuentes en un 20.5%, seguidas por malformaciones congénitas, deformidades y anomalías cromosómicas en un 16.0%, y en tercer lugar se encontraron las enfermedades que comprometen el sistema osteomuscular y el tejido conectivo en el 14.5% (53). Actualmente no existe información pública disponible respecto al impacto económico del tratamiento y la atención de los pacientes con enfermedades raras que genera en el sistema de salud colombiano.

El diagnóstico de una enfermedad monogénica no es sencillo, pues existen múltiples mecanismos que originan estas condiciones como, por ejemplo: la sustitución de un nucleótido único, deleciones y duplicaciones de diversos tamaños, mutaciones dinámicas, alteraciones en la metilación y la disomía uniparental que afecta regiones de impronta genómica(54), las cuales pueden encontrarse en la totalidad de las células o en mosaicismo. Adicionalmente, la expresión fenotípica de las enfermedades monogénicas puede verse afectada por la expresividad variable que implica que las manifestaciones clínicas y la severidad de una enfermedad genética son diversas incluso en individuos de la misma familia con la misma variante, y por la penetrancia incompleta, donde menos del

100% de los individuos que portan una mutación patogénica expresan la enfermedad. De otro lado, puede existir heterogeneidad de locus en la que una misma enfermedad puede ser causada por diferentes genes donde por ejemplo se ha descrito que la discapacidad intelectual puede ser causada por alteraciones monogénicas en más de 700 genes(9); y también puede presentarse heterogeneidad alélica, en la cual variantes en un mismo gen producen fenotipos diferentes, como por ejemplo el gen *FGFR3* que se ha asociado con acondroplasia, hipocondroplasia, displasia tanatofórica y síndrome Cruzon entre otros(55).

Lograr documentar el origen de una enfermedad monogénica en un paciente, suele ser un proceso largo y difícil, que por toda la variabilidad previamente descrita, hace que los fenotipos se sobrelapen, que cada caso sea único y que sea un reto, que de superarlo, le permite a los especialistas conducir el manejo de la enfermedad, ampliar el conocimiento con respecto al pronóstico, la calidad de vida, el asesoramiento genético y riesgo reproductivo, e incluso le permite a los pacientes y sus familias acceder a los beneficios que brinda el Estado, incluyendo mejor acceso al sistema de salud como la facilidad en la programación de citas médicas con especialistas, la eliminación de copagos para los pacientes pertenecientes al régimen contributivo, entre otros(48)(49).

## 2.3 El genoma y sus regiones codificantes:

En la actualidad se conoce que el genoma humano corresponde a una secuencia de  $3.05 \times 10^9$  nucleótidos, que contienen 63.494 genes de los cuales 19.969 codifican para proteínas(4). Solo los exones (la región codificante del genoma), representan entre el 1.5% y el 2% del genoma humano y corresponden a aproximadamente 48 a 60Mb, que son alrededor de 562.164 exones aproximadamente(56).

Además del genoma nuclear, el genoma humano también se compone del genoma mitocondrial, cuya secuencia fue publicada antes que la del genoma nuclear en 1981(57). Este genoma circular es de 16.6kb de longitud, el 97% de su contenido es codificante, y contiene 37 genes, de los cuales 13 codifican para proteínas estructurales de la mitocondria, 2 para ARN ribosomal y 22 para ARN de transferencia(57)(58). Sin embargo, muchas proteínas necesarias para el funcionamiento de la mitocondria se encuentran en

el ADN nuclear. Se han descrito múltiples enfermedades de origen mitocondrial, que ocurren por sustituciones puntuales en la secuencia de ADN mitocondrial o deleciones, por lo que las técnicas de secuenciación de exoma también permiten el estudio de patologías de posible origen mitocondrial(58).

## 2.4 Técnicas de secuenciación de nueva generación:

La secuenciación de nueva generación (NGS) es una técnica que permite identificar cada una de las bases nucleotídicas de una cadena de ácidos nucleicos (ADN o ARN) posición a posición. A medida que aparecen técnicas que aumentan la velocidad de procesamiento, disminuyen el tiempo de trabajo y disminuyen la tasa de error, las técnicas o tecnologías de secuenciación se han clasificado en generaciones:

**Secuenciación de primera generación:** Esta técnica hace referencia a la secuenciación por método de Sanger, que consiste en utilizar la cadena molde que se desea amplificar, primers forward y reverse, es decir cebadores o iniciadores en sentido 5' a 3' (forward) que se unen a la cadena molde para elongarla y cebadores en sentido 3' a 5' (reverse) que se unen a la cadena complementaria, una ADN polimerasa, deoxinucleótidos (dNTPs) y dideoxinucleótidos con radioactividad (ddNTPs) que tienen un hidrógeno en el carbono 3 de la ribosa que al no tener un grupo OH funcionará como un nucleótido de terminación. Con estos reactivos se realizan cuatro reacciones de polimerización con los cuatro tipos de ddNTPs (ddATP para Adenina, ddGTP para Guanina, ddCTP para Citocina y ddTTP Timina) y se van agregando tanto los dNTPs (dATP, dGTP, dCTP y dTTP) por complementariedad. Un ddNTP se polimerizará aleatoriamente finalizando la amplificación, y formando así fragmentos de ADN de diferentes tamaños. Estos fragmentos se analizan a través de un gel de poliacrilamida que separa los fragmentos de acuerdo con su peso molecular (los fragmentos más pequeños migran más rápido) y permite identificar el último nucleótido añadido en ese fragmento (ddNTP) reorganizando así la secuencia según la posición de cada nucleótido(59). Años después, la técnica se automatizó utilizando ddNTPs fluoromarcados que permiten realizar la secuenciación en una única reacción y adicionalmente se utilizó electroforesis capilar para su análisis, leyendo los ddNTPs fluoromarcados de los fragmentos según su longitud de onda generando un electroferograma; la electroforesis capilar puede tener hasta 96 capilares en un solo secuenciador (59). Esta técnica es capaz de secuenciar un fragmento de aproximadamente 1kb en una corrida.

Su principal limitante es que secuencia un bajo número de bases por corrida consumiendo una gran cantidad de tiempo para secuencias de gran tamaño, además de ser elevado el costo por base(18). A pesar de esto, se considera como el *gold standard* de la secuenciación y hasta hace algunos años, se recomendaba la confirmación Sanger de los hallazgos que se obtenían utilizando las nuevas generaciones de secuenciación(18).

**Secuenciación de segunda generación o de lecturas cortas:** A partir de esta generación se conoce como Secuenciación de Nueva Generación, NGS, y recibe su nombre debido a que las técnicas desarrolladas tienen la capacidad de realizar secuenciación masiva en paralelo en una misma reacción, es decir que utiliza la Reacción en Cadena de la Polimerasa (PCR) para amplificar el ADN y generar millones de fragmentos para su lectura al mismo tiempo, además de usar técnicas diferentes para identificar las bases nitrogenadas en cada posición, lo que disminuye el tiempo de corrida(59)(18).

Los principios generales de esta técnica son:

1. Preparación de librerías
  2. Amplificación de la librería por PCR
  3. Secuenciación con diferentes metodologías(18).
- 
1. La preparación de librerías tiene dos pasos fundamentales: la fragmentación del ADN que se quiere secuenciar y la ligación de los fragmentos adaptadores que forman parte de los kits de enriquecimiento que capturan las regiones génicas de interés (existen kits para genes específicos, por grupo de enfermedades, para exomas completos, para genes HLA, entre otros) que son distribuidos por diferentes casas comerciales.
    - 1.1 La fragmentación del ADN consiste en la ruptura del ADN en segmentos cortos y puede realizarse por técnicas físicas como sonicación o nebulización, o a través de técnicas enzimáticas como transposasas o enzimas de restricción que generan fragmentos de longitud variable según la técnica utilizada entre 60-350pb(18).
    - 1.2 La ligación de adaptadores consiste en una serie de reacciones enzimáticas que permiten la reparación de los extremos, adenilación y adición de adaptadores en los extremos 3' y 5' de cada fragmento de ADN, los cuales están debidamente etiquetados en una especie de código de barras, lo cual permitirá la unión de los fragmentos



previamente obtenidos, a las perlas o a la superficie sólida (flowcell) de las regiones enriquecidas para la posterior amplificación, además de servir como guía para la indexación de los datos generados a partir de la lectura de los fragmentos previo al alineamiento durante el análisis bioinformático(60). Este paso para la preparación de librerías puede realizarse en simultáneo en una única reacción con transposasas y se conoce como tagmentación(60).

Tabla 2-1 Comparación de kits de captura para exoma.

Kit de secuenciación de exoma	Regiones secuenciadas	Número de sondas	Tipo de sonda (adaptador)	DNA genómico o requerido por muestra	Técnica de unión de adaptadores	Longitud de la sonda	Diseño de la sonda	Precio por captura (dólares)	Genoma de referencia para su construcción	Tiempo de hibridación en horas
Agilent SureSelect Human All Exon V8	35,24 Mb	~758,086	Primer de cRNA biotinilado	50-200 ng	Ligación	120	Lecturas pair end que no se superponen y se usan para PCR en puente.	No información	GRCh37 (hg19)	16
IDT xGEN Exome Panel	39 Mb	429,826	Primer de ADN biotinilado	500 ng	Ligación	No descrito	Primers no superpuestos	\$250	GRCh37 (hg19)	4
Illumina Nextera Rapid Capture Expanded Exome	62 Mb	>340,000	Primer de ADN biotinilado	50 ng	Transposasa	95	Primers no superpuestos, adyacentes entre sí.	\$250	GRCh37 (hg19)	24-48
Roche Nimblegen SeqCap EZ Exome v3.0	64 Mb	>2,100,000	Primer de ADN biotinilado	1 ug	Ligación	60 - 90	Primers superpuestos	\$600	GRCh37 (hg19)	72
MGI Easy Exome Capture V5 Probe Set	69Mb	No información	Primer de cRNA biotinilado	250-1000 ng	Ligación	90	Lecturas pair end	\$160	GRCh37 (hg19)	18-24

\* traducido y adaptado de (61–64).

- Este paso consiste en la amplificación clonal de las librerías; se puede realizar haciendo PCR por emulsión con perlas, PCR en puente o replicación en círculo generando nanobolas de ADN (DNBs). La PCR por emulsión con perlas consiste en la unión de los fragmentos con adaptadores a unas micelas que contienen unas perlas cubiertas con adaptadores que se unen por complementariedad a los adaptadores de los fragmentos, también contienen primers en sus extremos, dNTPs y ADN polimerasa. La PCR ocurre al interior de las micelas al someterlas a diferentes cambios de temperatura(21). La PCR por emulsión es usada para las plataformas de

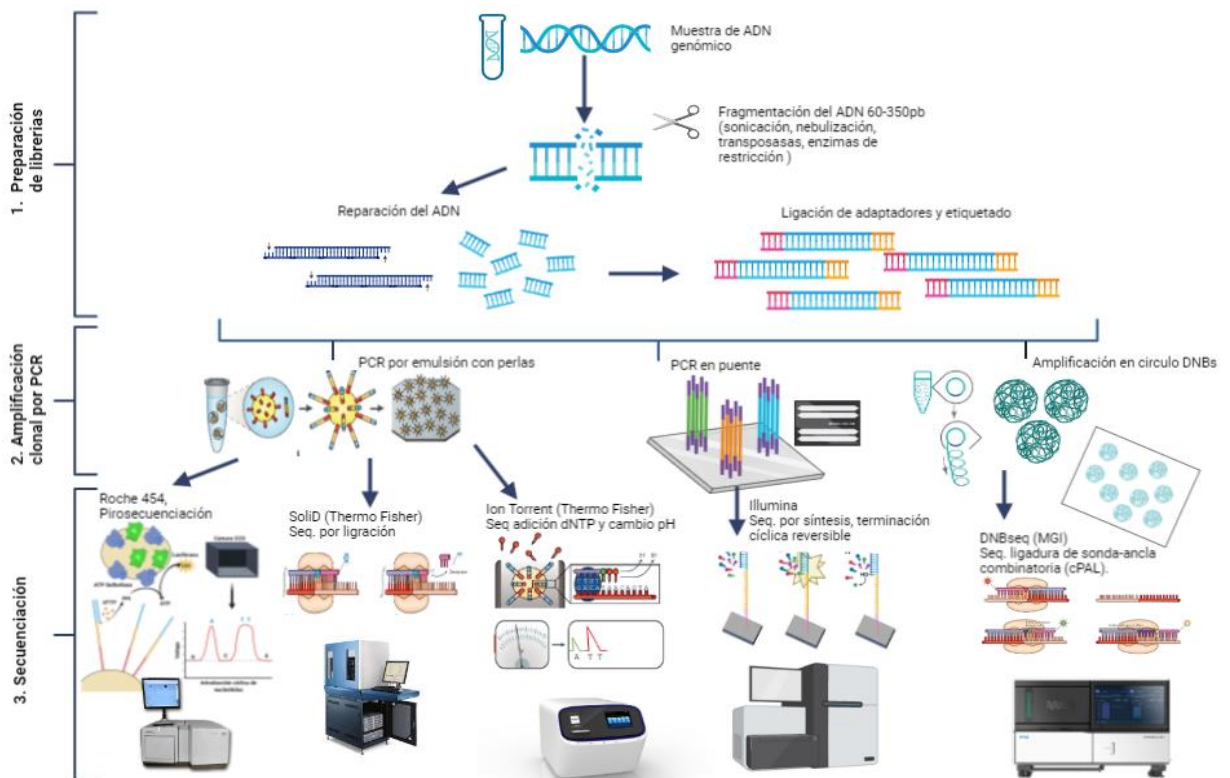
pirosecuenciación de Roche, SOLiD (Thermo Fisher), GeneReader (Qiagen) y IONTorrent(Thermo Fisher)(18). Por otro lado, la PCR en puente es una técnica utilizada por Illumina Solexa y consiste en la unión de los fragmentos unidos a los adaptadores a otras secuencias adaptadoras fijadas a un soporte sólido o flowcell que también contiene un primer que actuará de molde para la polimerasa, se somete a termociclación y cuando se polimeriza una hebra se une a otro pozo con adaptadores para hacer una nueva amplificación pero en forma reversa, generando clústeres de los fragmentos de interés, que se observan como puentes durante la amplificación(61). La replicación en círculo (rolling circle replication) es utilizada por MGI DNBseq, en la cual los fragmentos de ADN de la librería se circularizan y por PCR se generan múltiples copias clonales de los fragmentos generando nanobolas de ADN que posteriormente serán cargadas a una celda o flowcell que contiene pozos (DNA spots) de aproximadamente 220nm de diámetro para posteriormente ser secuenciados(21).

3. El último paso es la secuenciación donde a través de diferentes plataformas y técnicas se lee cada base nucleotídica y se interpreta en un software con la letra que identifica la base nitrogenada. Las técnicas se dividen en dos grupos según la química involucrada: Secuenciación por síntesis y secuenciación por ligación(18). La secuenciación por síntesis incluye la técnica de pirosecuenciación (Roche 454), donde cada vez que se añade un nucleótido complementario a la secuencia se libera pirrofosfato inorgánico que se convierte en ATP por acción de la enzima sulfurilasa, y una enzima de luciferasa aprovecha el ATP para convertirlo en una señal quimioluminiscente que varía según cada nucleótido y forma un pirograma(19)(18). La plataforma que por excelencia realiza secuenciación por síntesis es Illumina Solexa que actualmente es líder en el mercado de NGS, y consiste en el empleo de terminadores reversibles unidos a los dNTPs que se incorporan por complementariedad a los fragmentos plantilla de ADN y que por una señal lumínica permiten identificar el nucleótido de la secuencia y generar una imagen digital del nucleótido leído para posteriormente separarse(61). La tasa de error de esta metodología es muy baja y varía del 0.1 al 1%(18). Se muestra un resumen de estas metodologías de secuenciación en la figura 1(18). De acuerdo con la velocidad de

procesividad, Illumina ha creado diferentes secuenciadores conocidos como Miseq, Nextseq, Hiseq, Novaseq, entre otros.

Otra técnica de secuenciación por síntesis es la utilizada por Ion Torrent de ThermoFisher, que a diferencia de Illumina que utiliza terminación reversible cíclica, se basan en una sola señal para la adición de un único nucleótido dNTP en una cadena de elongación. Ion Torrent detecta los iones H<sup>+</sup> que se liberan cuando cada dNTP es incorporado, el cambio en el pH es detectado por un semiconductor de óxido de metal y un transistor sensible a iones(21). BGI cuenta con un sistema de secuenciación utilizando un enfoque de ligación de sonda-ancla combinatoria, Applied Biosystems ABI-SOLID también realiza secuenciación por ligación y está basada en sondas de oligonucleótidos de longitudes variables, etiquetadas por fluorescencia sensible al nucleótido que identifica por excitación con láser y son ligadas utilizando una ligasa de ADN(18).

Figura 2-1 Técnicas de secuenciación de segunda generación.



\*Diseñado con Biorender.com \* Traducido y adaptado de (18,21)

Dentro de las limitantes de estas técnicas de NGS están que no es precisa para detectar mutaciones dinámicas de tipo expansión de tripletas, tiene una menor sensibilidad para

detectar variantes en regiones muertas del NGS, que son aquellas regiones ricas en GC donde es difícil lograr secuenciación de alta calidad, y en general, también es menos sensible para identificar traslocaciones e inversiones, en particular por su dificultad para establecer puntos de corte en variantes estructurales ya que es también dependiente del kit de captura y la región secuenciada(61).

**Secuenciación de tercera generación o de lecturas largas:** Estas tecnologías permiten la secuenciación de una única hebra de ADN en tiempo real, eliminando el paso de amplificación y el sesgo debido al error que puede generarse en esta etapa con las tecnologías de segunda generación. Esta tecnología es empleada por Pacific Biosciences. Esta plataforma tiene unos nanopozos que tienen un orificio llamado zero-mode waveguide (ZMW), por donde pasa un rayo láser. En cada nanopozo se inmoviliza un fragmento de ADN con una polimerasa que realizará la amplificación utilizando nucleótidos fluoromarcados generando una longitud de onda diferente según la base nitrogenada de cada posición que se producirá por la excitación con láser y luego será liberado el nucleótido(18)(59), este proceso se puede repetir varias veces para generar una secuencia consenso.

La ventaja de las técnicas de secuenciación de tercera generación respecto a las de segunda, es que reducen el tiempo en el flujo de trabajo por la eliminación de los pasos de la PCR, también utiliza fragmentos más largos de lectura lo cual facilita el alineamiento, facilita definir variantes en *cis* o *trans* cuando son cercanas y reduce los segmentos sin cobertura, sin embargo, tienen la desventaja de ser técnicas muy costosas, y de aumentar las tasas de error que varía entre un 1-15%(21).

Oxford Nanopore Technologies desarrolló una tecnología ION, en la cual toda una hebra de ADN pasa a través de un nanoporo a la que se le ligan adaptadores para capturar el fragmento hacia el nanoporo y ligar la polimerasa, la hebra se desplaza a través del nanoporo durante la amplificación y detecta cambios mínimos de voltaje por la carga iónica permitiendo la secuenciación. Esta técnica tiene como objetivo realizar la secuenciación *in situ* es decir realizar la secuenciación directa sobre una única célula con ADN de cualquier longitud(21).

Tiene la ventaja de requerir una baja cantidad de ADN para la secuenciación, ahorra el paso de amplificación por PCR, también utiliza equipos mucho más pequeños respecto a los de lecturas cortas, incluso del tamaño de una USB haciéndolo más cómodo para el usuario y muestra los datos de lectura en tiempo real. Su mayor limitante es la tasa de error que varía del 1-30% particularmente de tipo InDel; debido a que realiza una única secuenciación y no secuenciación repetitiva, obteniendo una profundidad de 1x, sin embargo, si se secuencian varias células se obtiene una profundidad de secuenciación mayor y se reduce significativamente esta tasa de error(18)(59).

## 2.5 Técnica de secuenciación MGI DNBseq

La técnica empleada para la secuenciación de los exomas del presente estudio será una técnica de segunda generación conocida como DNBseq, la cual pertenece a la compañía MGI. En primer lugar, para la preparación de librerías, se requiere una cantidad mínima de 250ng de ADN, la fragmentación se realiza enzimáticamente y el tamaño de los fragmentos es controlado por los tiempos de incubación y la temperatura(62). Posteriormente el ADN genómico es seleccionado a través de perlas magnéticas de selección para elegir fragmentos entre 100 a 1000pb que en promedio son de 350pb para secuenciación pair-end 150. Después se lleva a cabo una purificación de los fragmentos, se realiza la reparación de los extremos de las cadenas de ADN a través de ligasas, y adenilación o A-taling, que consiste adicionar un nucleótido no modelado al extremo 3' a los fragmentos de ADN bicatenarios y romos(62). Luego se realiza la ligación de los primeros adaptadores en los extremos 3' y 5' de cada fragmento de ADN de cadena doble lineal, los cuales están debidamente etiquetados y se circularizan. En consiguiente, el molde de ADN circular es clivado cadena abajo de la secuencia de adaptadores con una endonucleasa de tipo II, tres rondas adicionales de ligación de adaptadores, circularización y clivaje generan un molde de ADN de cadena doble circular con cuatro tipos de adaptadores en el círculo(21).

Posteriormente, sigue un proceso de replicación en círculo rodante (Rolling circle replication), donde se amplifican los moldes de ADN circular al escindir los adaptadores generando una gran masa de concatémeros (secuencias de ADN que contiene múltiples copias de la misma secuencia en serie) llamados nanobolas de ADN, que se conocen como DNBs (DNA nanoballs)(21). Esta técnica tiene la ventaja de hacer múltiples copias exclusivamente de la hebra original. Luego, previo a iniciar el proceso de secuenciación, se cargan las nanobolas de ADN en una celda de flujo y se fijan en sus pozos, conocidos

como DNB spots, en un array donde ocurrirá una reacción de secuenciación por ligación con la respectiva lectura de las señales por fluorescencia(21). Esta secuenciación tiene un enfoque de ligación de sonda-ancla combinatorio, que contiene una sonda (secuencia con una base conocida y las demás degeneradas) marcada con fluoróforo y un ancla (fragmento de anclaje) complementaria a una de las cuatro secuencias adaptadoras. El complejo sonda-ancla se van hibridando en la DNB en su posición y se generan imágenes a través de un láser de excitación para que identifique la primera base en el lado 3' o 5' del ancla. A continuación, el complejo sonda-ancla se elimina y el proceso comienza de nuevo con la misma ancla pero diferente sonda con la base conocida en la posición n+1(21). Esto se repite hasta cinco bases desde el extremo 3' del ancla y cinco bases desde el extremo 5' del ancla. Luego, hay una nueva ronda de hibridación utilizando anclas desplazadas 5 posiciones. Este proceso ocurre desde las 4 secuencias de adaptadores múltiples veces para obtener fragmentos de lectura pair-end de 100-150pb(21). Esta tecnología tiene la ventaja de que al detectar los fluoróforos de los adaptadores, elimina automáticamente la señal emitida por los adaptadores y solo lee los datos de la secuencia de interés. Esto ahorra un paso en la limpieza de los datos crudos y mejora la calidad de las lecturas(21)(63)(64).

## 2.6 Análisis computacional de datos biológicos

El software del secuenciador hace la interpretación de cada posición de los fragmentos de ADN o ARN y arroja un archivo de lectura con una gran cantidad de información por los millones de copias producidas, el cual requerirá una limpieza y análisis de calidad para ser ensamblado y generar una secuencia consenso para su posterior interpretación(65).

El primer archivo que se obtiene con los datos crudos se conoce como archivo FASTQ el cual integra un archivo FASTA con los datos de secuenciación y un archivo Q que mira la calidad al hacer el llamado de cada base, conocido como *base calling*(65). El tamaño del archivo FASTQ de un genoma a profundidad media de 30x es de aproximadamente 75 a 100Gb y el de un exoma a profundidad media de 100x varía entre de 8 y 14 Gb(66). Posteriormente, las lecturas se almacenan en un formato SAM (Sequence Alignment Map) que describe donde se mapean las lecturas de secuenciación en el genoma de referencia y alcanza los 150Gb para genoma y 15Gb para exoma (65).

Luego el archivo SAM se comprime en un archivo BAM (Binary Alignment Map) que es la versión en lenguaje binario del mapa de secuenciación alineado y en el caso del genoma humano de 3000 millones de pares de bases, genera un archivo de aproximadamente 40Gb de información y de 4 a 5Gb para un exoma, sin embargo, este tamaño puede variar de acuerdo con la profundidad de la secuenciación(66). El tamaño del archivo BAM estará determinado por la profundidad (el número promedio de veces que se lee cada base) y la cobertura (longitud de lectura en un experimento de secuenciación) (66).

Llamado de variantes: hace referencia al proceso para identificar las bases nucleotídicas que son diferentes respecto a una secuencia de referencia del genoma humano. Para realizar el llamado es importante tener los datos del caso índice limpios, de buena calidad y con una amplia cobertura, así como elegir una secuencia de referencia de buena calidad. El llamado de variantes es más confiable cuando se tiene una mayor profundidad de lectura, y su precisión depende de la tasa de error de la plataforma de secuenciación, la presencia de regiones genómicas repetitivas, y el algoritmo utilizado para ensamblar las lecturas en la secuencia de referencia(66). Para la detección de variantes de nucleótido simple (SNV) se recomienda una profundidad de 100x en secuenciación de exoma completo y de al menos 60x para identificar InDel (66). El llamado de variantes es almacenado en un archivo estandarizado conocido como VCF (Variant Call Format) que tiene un tamaño de aproximadamente 150Mb para genoma y 30Mb para exoma(65), el cual en diagnóstico será el archivo utilizado por los clínicos para definir las variantes de importancia según el fenotipo de cada individuo en estudio y para hacer la anotación de las variantes. Algunos de los softwares utilizados para generar los archivos VCF son SAMtools, GATK, FreeBayes, Deepvariant y Atlas2(66)(67). Uno de los más preferidos es GATK por su capacidad para identificar pequeñas inserciones/delecciones (InDel), variantes de nucleótido simple (SNVs), Variantes en el número de copias (CNVs) y algunas variantes estructurales (SV), usando un modelo Bayesiano y con la capacidad de aprender los filtros de los datos(66).

Anotación de variantes: Es el proceso por el cual se observan las consecuencias biológicas o funcionales de una variante en la proteína a través de modelos de predicción. Algunos programas de anotación son ANNOVAR para anotar SNVs, InDels y CNVs que explora las consecuencias biológicas de las variantes e indica la puntuación de PolyPhen, SIFT, CADD, Mutation taser, SIFT, entre otros(66).

Filtrado de variantes: Es la elección de las variantes que serán incluidas para el análisis, de cada caso, utilizando bases de datos clínicas de tipo ClinVar, OMIM, DECIPHER, HGMD(68), entre otras. Este proceso también incluye la identificación de la MAF (frecuencia de alelo menor) para cada variante ya que una variante poco frecuente en la población general, pero frecuente dentro de un grupo de pacientes con un mismo fenotipo, aumenta su valor de patogenicidad. Por lo general, se descartan aquellas variantes que no tienen repercusiones funcionales en modelos biológicos o según los modelos de predicción, también aquellas altamente frecuentes en la población como para causar una enfermedad monogénica rara, y aquellas cuyo mecanismo de herencia no coincide con la presentación de la variante para causar enfermedad. Este paso en el análisis bioinformático puede variar entre laboratorios, de acuerdo con la disponibilidad de bases de datos, y en algunos programas de predicción *in silico* la variante puede tener repercusión biológica mientras que en otros programas no(66)(65).

En general, a nivel diagnóstico, cada laboratorio desarrolla un algoritmo bioinformático para filtrar las variantes de interés clínico, este proceso se conoce en inglés como *pipeline* bioinformático y a pesar de que cada laboratorio es independiente para desarrollarlo, existen algunos filtros comunes que son aplicados:

- Términos HPO asociados al fenotipo clínico del probando, con el fin de obtener variantes en aquellos genes que coinciden con el cuadro clínico y han sido ampliamente reportados en la literatura.
- Fenotipo OMIM: similar al filtro anterior, esta depuración se realiza con el fin de obtener genes que se asocian con una enfermedad mendeliana y coinciden con un fenotipo previamente descrito. Este filtro es necesario particularmente cuando el análisis de los datos de secuenciación tiene fines diagnósticos y no de investigación.
- Frecuencia alélica: En general cuando una variante tiene una frecuencia alélica superior al 5% se considera que es una variante benigna y es poco probable que explique un fenotipo patológico en una enfermedad de herencia mendeliana ya que habría muchos afectados en la población general(69). Por lo tanto, los laboratorios aplican una frecuencia alélica menos (MAF) por lo menos inferior al 5% ( $<0.05$ ) tomando como referencia las frecuencias reportadas en bases de datos



poblacionales como ExAC, GnomAD, 1000Genomas entre otras. Existen variantes de interés clínico con una alta frecuencia poblacional que no deben ser excluidas del análisis.

- Frecuencia de profundidad respecto al alelo de referencia: Cuando se detecta un alelo alternativo al alelo de referencia, se espera que la frecuencia en la que se secuencia el alelo alternativo sea cercana al 50% de las lecturas, cuando está presente en heterocigosis, y cercana al 100% en caso de encontrarse en homocigosis. Cuando los valores son muy inferiores al 50%, es posible que se haya presentado la detección de un alelo alternativo por un error que surge en la secuenciación o porque dicha región génica no estaba suficientemente enriquecida para detectar la variante.
- Región de interés: Según la región de interés en el caso de los exomas por ejemplo es posible filtrar para regiones exónicas y las regiones próximas al sitio donador y aceptor de splicing.
- Tipo de mutación: En general las variantes sinónimas que no generan ningún efecto deletéreo en las proteínas a menos de que afecten el sitio canónico del splicing, pueden ser descartadas, sin embargo, aquellas con una alta evidencia de patogenicidad pueden ser rescatadas.
- Bases de datos clínicas: ClinVar (Clinically Relevant Variation) es un repositorio público y constantemente actualizado que contiene la interpretación de variantes clínicamente relevantes del NIH de Estados Unidos, el cual se nutre a través del reporte de variantes que realiza la comunidad científica como laboratorios de diagnóstico, clínicos, investigadores, organizaciones o paneles de expertos (39). ClinVar admite diferentes niveles de complejidad de reporte que varían desde variantes estudiadas por modelos *in silico*, hasta variantes con estudios funcionales en modelos animales o documentadas en cohortes de pacientes; las cuales son analizadas y curadas por expertos de ClinGen programa asociado del NIH que valida y clasifica las variantes para ser publicadas posteriormente en ClinVar(70–72). Adicionalmente, ClinVar cuenta con un sistema de puntuación por estrellas para conocer el grado de revisión y aprobación que tiene la clasificación de la variante en la base de datos: 4 estrellas indican que la variante forma parte de guías de práctica, 3 estrellas que fue revisada por un panel de expertos, 2 estrellas hacen referencia a que dos o más remitentes proporcionaron la misma interpretación y no

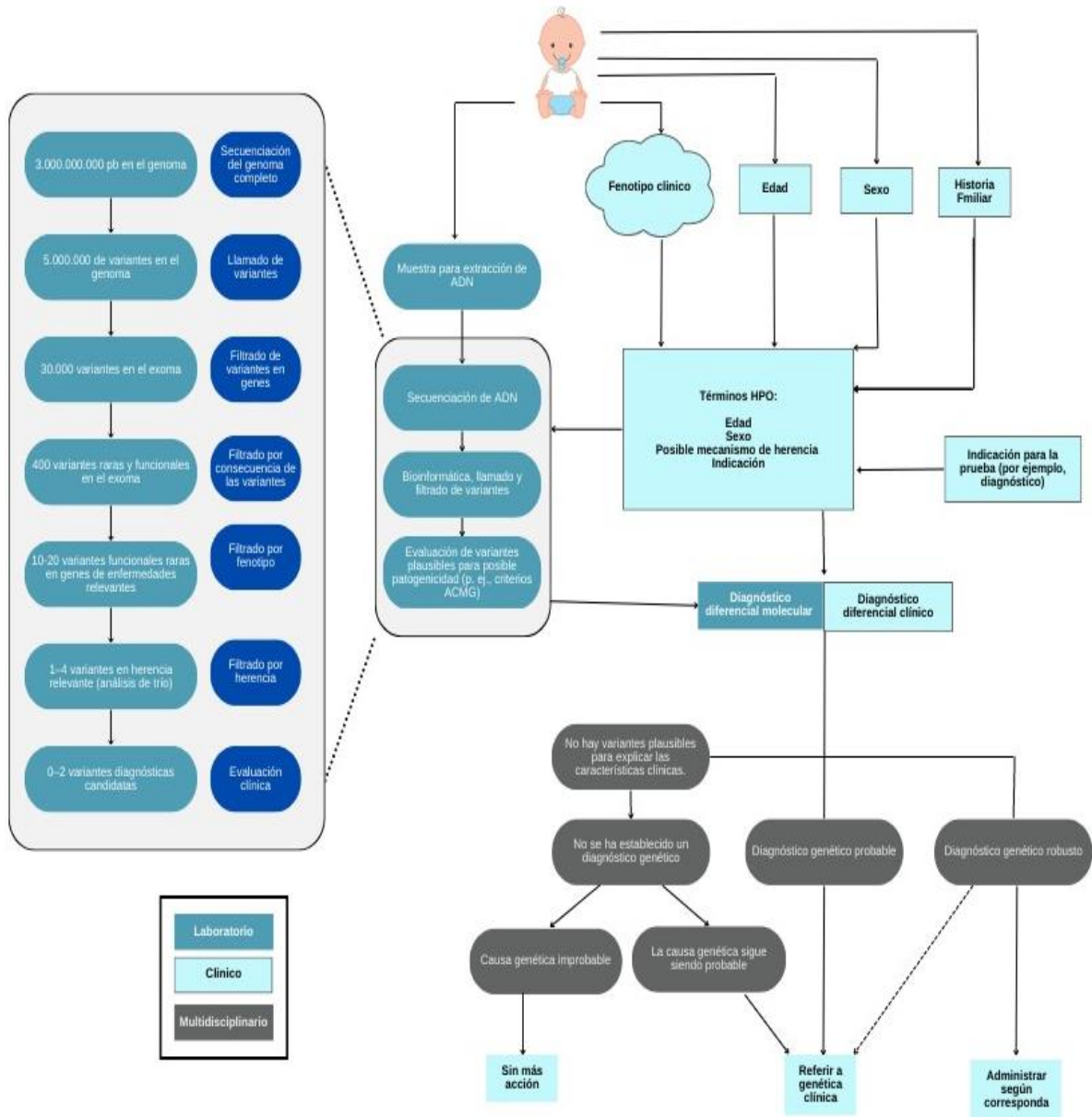
hay conflictos en la misma, 1 estrella indica que hay múltiples remitentes pero con interpretaciones contradictorias o que la información de la variante ha sido proporcionada por un único remitente, finalmente 0 estrellas indican que se envió una variante pero sin presentar criterios de afirmación o clasificación(73). Estas características, permiten que esta base de datos sea un filtro inicial particularmente útil en diagnóstico para eliminar aquellas variantes benignas ampliamente descritas, y no deben ser usadas para descartar de una variante que no ha sido reportada, pues la ausencia de una variante en Clinvar no resta su posible patogenicidad(70).

Posteriormente, cada laboratorio es libre de realizar filtros utilizando predictoras *in silico* donde es posible usar predictores basados en evolución, que utilizan información filogenética para definir probabilidad de patogenicidad, también puede usar predictores para evaluar el grado de tolerancia de variantes missense, sinónimas, del sitio de splicing y no codificantes en las proteínas.

Asignación de patogenicidad: En el ámbito diagnóstico, una vez elegidas las variantes candidatas, se debe definir si alguna de ellas es capaz de aproximarse al diagnóstico del caso índice, proceso que es conocido como asignación de patogenicidad, el cual es usualmente realizado por un clínico que se encargará de realizar una revisión exhaustiva de la literatura para definir si el fenotipo del paciente en estudio se corresponde con la variante en un gen que afecta su función. Para ello se utilizan bases de datos clínicas para realizar la curación donde se reportan las variantes patogénicas, que incluyen (entre otras) ClinVar, HGMD, LOVD, pero su uso debe ser muy cauteloso respecto a sus actualizaciones, secuencias de referencia y casos reportados. Otras opciones de análisis han intentado estandarizar la búsqueda utilizando vocabulario clínico u ontologías estructuradas como HPO (Human Phenotype Ontology) para filtrar de acuerdo con las características clínicas más relevantes (30). Finalmente, para la asignación de la patogenicidad, se recomienda el uso de guías para la clasificación de variantes, para el presente estudio se utilizaron las guías americanas del Colegio Americano de Genética Médica y Genómica y la Asociación de Patología Molecular (ACMG-AMP) que establece los lineamientos para la clasificación de variantes como se explica más adelante(55)(9).

Ver figura 2-2. Existen otras guías disponibles, pero el consenso del ACMG es ampliamente utilizado en Europa y América.

Figura 2-2. Resumen de análisis clínico y bioinformático de datos biológicos.



\* Traducido y adaptado de Wright y colaboradores 2018 (9)

## 2.7 Lineamientos para la clasificación de variantes

El Colegio Americano de Genética Médica y Genómica (ACMG) junto con la Asociación de Patología Molecular (AMP), en el año 2015 publicó los estándares y guías para la interpretación de variantes de secuenciación con el fin de unificar el análisis y la interpretación de los resultados de exámenes de tipo panel, exoma, genoma y transcriptoma de pacientes con sospecha de enfermedades de origen genético. En esta guía, se realizan recomendaciones respecto a la unificación de la nomenclatura para el reporte de las variantes, así como, las herramientas para tener en cuenta para su clasificación en categorías, bases de datos clínicas, poblacionales, bases bioinformáticas de predicción y datos funcionales y de segregación(69) De acuerdo con estas guías, no es recomendable utilizar el término mutación y polimorfismo y debe ser reemplazado por el termino variante. Las variantes a su vez, según su significancia clínica para enfermedades mendelianas se dividen en 5 categorías: patogénicas, probablemente patogénicas, de significado incierto, probablemente benignas y benignas (69).

Variante patogénica (P): Es una variante con suficiente evidencia para producir enfermedad, con una certeza superior al 99% (69,74).

Variante probablemente patogénica (PP): Es un cambio con una evidencia fuerte a favor de patogenicidad y se utiliza para indicar una certeza superior al 90%.

Variante de significado Incierto (VSI): Son variantes cuya evidencia es insuficiente o contradictoria para definir si son benignas o patogénicas. Se recomienda su reporte si el gen en el cual se documentó puede estar asociado al fenotipo y la probabilidad de patogenicidad es del 10 al 90%

Variante probablemente benigna (PB): Es un cambio con una evidencia fuerte a favor de benignidad y se utiliza para indicar una certeza superior al 90% o una probabilidad de patogenicidad entre el 1 y menor a 10%.

Variante benigna (B): Es aquella con suficiente evidencia en contra de su patogenicidad. Es decir, con una certeza de patogenicidad inferior al 1%.

Para clasificar las variantes, las guías recomiendan utilizar las siguientes bases de datos: (69) (Tabla 2-2, 2-3, 2-4, 2-5)

Tabla 2-2 Bases de datos poblacionales

Bases de datos poblacionales	
Exome Aggregation Consortium	<a href="http://exac.broadinstitute.org/">http://exac.broadinstitute.org/</a>
Exome Variant Server	<a href="http://evs.gs.washington.edu/EVS">http://evs.gs.washington.edu/EVS</a>
1000 Genomes Project	<a href="http://browser.1000genomes.org">http://browser.1000genomes.org</a>
dbSNP	<a href="http://www.ncbi.nlm.nih.gov/snp">http://www.ncbi.nlm.nih.gov/snp</a>
dbVar	<a href="http://www.ncbi.nlm.nih.gov/dbvar">http://www.ncbi.nlm.nih.gov/dbvar</a>

\*Traducida de Standards and guidelines for the interpretation of sequence variants ACMG 2015 (69)

Tabla 2-3 Bases de datos clínicas

Bases de datos clínicas	
ClinVar	<a href="http://www.ncbi.nlm.nih.gov/clinvar">http://www.ncbi.nlm.nih.gov/clinvar</a>
OMIM	<a href="http://www.omim.org">http://www.omim.org</a>
Human Gene Mutation Database	<a href="http://www.hgmd.org">http://www.hgmd.org</a>
Human Genome Variation Society	<a href="http://www.hgvs.org/dblist/dblist.html">http://www.hgvs.org/dblist/dblist.html</a>
Leiden Open Variation Database	<a href="http://www.lovd.nl">http://www.lovd.nl</a>
DECIPHER	<a href="http://decipher.sanger.ac.uk">http://decipher.sanger.ac.uk</a>

\*Traducida de Standards and guidelines for the interpretation of sequence variants ACMG 2015 (69)

Tabla 2-4. Bases de datos para secuencias de referencia

Bases de datos para secuencias de referencia	
NCBI Genome	<a href="http://www.ncbi.nlm.nih.gov/genome">http://www.ncbi.nlm.nih.gov/genome</a>
RefSeqGene	<a href="http://www.ncbi.nlm.nih.gov/refseq/rsg">http://www.ncbi.nlm.nih.gov/refseq/rsg</a>
Locus Reference Genomic (LRG)	<a href="http://www.lrg-sequence.org">http://www.lrg-sequence.org</a>
MitoMap	<a href="http://www.mitomap.org/MITOMAP/HumanMitoSeq">http://www.mitomap.org/MITOMAP/HumanMitoSeq</a>

\*Traducida de Standards and guidelines for the interpretation of sequence variants ACMG 2015 (69)

Tabla 2-5 Programas de predicción in silico

<b>Programas de predicción <i>in silico</i></b>	
<b>Predicción de variantes <i>missense</i></b>	
ConSurf	<a href="http://consurftest.tau.ac.il">http://consurftest.tau.ac.il</a>
FATHMM	<a href="http://fathmm.biocompute.org.uk">http://fathmm.biocompute.org.uk</a>
MutationAssessor	<a href="http://mutationassessor.org">http://mutationassessor.org</a>
PANTHER	<a href="http://www.pantherdb.org/tools/csnpscoreForm.jsp">http://www.pantherdb.org/tools/csnpscoreForm.jsp</a>
PhD-SNP	<a href="http://snps.biofold.org/phd-snp/phd-snp.html">http://snps.biofold.org/phd-snp/phd-snp.html</a>
SIFT	<a href="http://sift.jcvi.org">http://sift.jcvi.org</a>
SNPs&GO	<a href="http://snps-and-go.biocomp.unibo.it/snps-and-go">http://snps-and-go.biocomp.unibo.it/snps-and-go</a>
Align GVGD	<a href="http://agvgd.iarc.fr/agvgd_input.php">http://agvgd.iarc.fr/agvgd_input.php</a>
MAPP	<a href="http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html">http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html</a>
MutationTaster	<a href="http://www.mutationtaster.org">http://www.mutationtaster.org</a>
MutPred	<a href="http://mutpred.mutdb.org">http://mutpred.mutdb.org</a>
PolyPhen-2	<a href="http://genetics.bwh.harvard.edu/pph2">http://genetics.bwh.harvard.edu/pph2</a>
PROVEAN	<a href="http://provean.jcvi.org/index.php">http://provean.jcvi.org/index.php</a>
nsSNPAnalyzer	<a href="http://snpanalyzer.uthsc.edu">http://snpanalyzer.uthsc.edu</a>
Condel	<a href="http://bg.upf.edu/fannsdb/">http://bg.upf.edu/fannsdb/</a>
CADD	<a href="http://cadd.gs.washington.edu">http://cadd.gs.washington.edu</a>
<b>Predicción variantes en el sitio de splicing</b>	
GeneSplicer	<a href="http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml">http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml</a>
Human Splicing Finder	<a href="http://www.umd.be/HSF/">http://www.umd.be/HSF/</a>
MaxEntScan	<a href="http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html">http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html</a>
NetGene2	<a href="http://www.cbs.dtu.dk/services/NetGene2">http://www.cbs.dtu.dk/services/NetGene2</a>
NNSplice	<a href="http://www.fruitfly.org/seq_tools/splice.html">http://www.fruitfly.org/seq_tools/splice.html</a>
FSPLICE	<a href="http://www.softberry.com/berry.phtml?topic=fssplice&amp;group=programs&amp;subgroup=gfind">http://www.softberry.com/berry.phtml?topic=fssplice&amp;group=programs&amp;subgroup=gfind</a>
<b>Predicción de conservación de nucleótidos</b>	
GERP	<a href="http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html">http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html</a>
PhastCons	<a href="http://compugen.bscb.cornell.edu/phast/">http://compugen.bscb.cornell.edu/phast/</a>
PhyloP	<a href="http://compugen.bscb.cornell.edu/phast/">http://compugen.bscb.cornell.edu/phast/</a>

\*Traducida de Standards and guidelines for the interpretation of sequence variants ACMG 2015 (69)

Utilizando las bases de datos anteriormente descritas, la guía sugiere 28 criterios para clasificar las variantes patogénicas como se muestra a continuación:

Tabla 2-6 Criterios para la clasificación de variantes patogénicas

Criterios para clasificar variantes patogénicas		
Evidencia de patogenicidad	Categoría	Notas y advertencias
Muy fuerte	PVS1 variante nula (nonsense, framehift, $\pm 1$ o 2 sitios de splicing canónico, codón de iniciación, deleción única o multiexón) en un gen donde la pérdida de función (LOF) es un mecanismo conocido de enfermedad.	Advertencias: <ul style="list-style-type: none"> <li>• Tenga cuidado con los genes donde LOF no es un mecanismo conocido de enfermedad.</li> <li>• Tenga cuidado al interpretar las variantes de LOF en el extremo 3' de un gen.</li> <li>• Tenga cuidado con las variantes de splicing que se predice que provocarán la omisión del exón, pero dejarán el resto de la proteína intacta</li> <li>• Tenga precaución en la presencia de múltiples transcritos.</li> </ul>
Fuerte	PS1 Cambio por el mismo aminoácido de una variante patogénica establecida previamente, independientemente del cambio de nucleótidos	Advertencia: Tenga cuidado con los cambios que afectan el splicing en lugar de a nivel de aminoácidos / proteínas
	PS2 De novo (confirmada la maternidad y la paternidad) en un paciente con la enfermedad y sin antecedentes familiares de esta.	
	PS3 Estudios funcionales in vitro o in vivo bien establecidos que respaldan un efecto dañino en el gen o producto del gen	Nota: Estudios funcionales que han sido validados y demostrados como reproducibles y robustos en clínica. El laboratorio diagnóstico se considera el más establecido.
	PS4 La prevalencia de la variante en los individuos afectados aumenta significativamente en comparación con la prevalencia en controles	Nota 1: El riesgo relativo u OR, en los estudios de casos y controles, es $> 5.0$ , y el intervalo de confianza no incluye 1.0. Nota 2: en casos de variantes muy raras donde los estudios de casos y controles no alcanzan la significancia estadística, la observación previa de la variante en múltiples pacientes no relacionados con el mismo fenotipo y su ausencia en controles, pueden usarse como nivel de evidencia moderado.

Moderada	PM1 Ubicada en un punto caliente mutacional y / o dominio funcional crítico y bien establecido (por ejemplo, sitio activo de una enzima) sin variación benigna.	
	PM2 ausente en los controles (o con una frecuencia extremadamente baja si es recesiva) en el Proyecto de Exome Sequencing Project, Proyecto 1000 Genomas, o Consorcio de Exome Aggregation.	Advertencia: los datos de población para inserciones / deleciones pueden tener un mal llamado por la secuenciación de siguiente generación.
	PM3 Para trastornos recesivos, detectados en trans con una variante patogénica	Nota: Esto requiere pruebas de los padres (o descendientes) para determinar la fase.
	PM4 La longitud de la proteína cambia como resultado de deleciones/inserciones in-frame en una región no repetitiva o variantes con pérdida de parada stop-loss.	
	PM5 Variante missense nueva en un residuo de aminoácido donde se determina que un cambio de sentido erróneo es patogénico y se ha visto previamente.	Advertencia: Tenga cuidado con los cambios que afectan el splicing en lugar de a nivel de aminoácidos / proteínas.
	PM6 Asumido de novo, pero sin confirmación de paternidad y maternidad	
Apoya	PP1 Cosegregación con enfermedad en múltiples miembros de la familia afectados en un gen que definitivamente se sabe que causa enfermedad	
	PP2 Variante missense en un gen que tiene una baja tasa de variación missense benigna y en el que las variantes missense son un mecanismo común de enfermedad	



	<p>PP3 Múltiples líneas de evidencia computacional respaldan un efecto nocivo sobre el gen o producto génico (conservación, impacto evolutivo, splicing, etc.)</p>	<p>Advertencia: Debido a que muchos algoritmos in silico usan la misma entrada para sus predicciones, cada algoritmo no debe contarse como un criterio independiente. PP3 se puede usar solo una vez en la evaluación de una variante.</p>
	<p>PP4 El fenotipo del paciente o la historia familiar es altamente específica para una enfermedad con una etiología genética única.</p>	
	<p>PP5 Una fuente confiable reporta recientemente que la variante es patogénica, pero la evidencia no está disponible para el laboratorio para realizar una evaluación independiente.</p>	

\*Traducida de Standards and guidelines for the interpretation of sequence variants ACMG 2015 (69)

Tabla 2-7 Criterios para la clasificación de variantes benignas

Criterios para clasificar variantes benignas		
Evidencia de impacto benigno	Categoría	Notas y advertencias
Suficiente	<p>BA1 La frecuencia de alelos es &gt;5 % en Exome Sequencing Project, 1000 Genomes Project o Exome Aggregation Consortium</p>	
Fuerte	<p>BS1 La frecuencia del alelo es mayor que la esperada para la enfermedad.</p>	
	<p>BS2 Observado en un individuo adulto sano para un trastorno recesivo (homocigoto), dominante (heterocigoto) o ligado al cromosoma X (hemicigoto), con penetrancia completa esperada a una edad temprana</p>	
	<p>BS3 Los estudios funcionales in vitro o in vivo bien establecidos no muestran ningún efecto perjudicial sobre el empalme o la función de las proteínas</p>	

	BS4 Falta de segregación en miembros afectados de una familia	Advertencia: la presencia de fenocopias para fenotipos comunes (es decir, cáncer, epilepsia) puede imitar la falta de segregación entre las personas afectadas. Además, las familias pueden tener más de una variante patogénica que contribuye a un trastorno autosómico dominante, lo que confunde aún más la aparente falta de segregación.
Soporte	BP1 Variante sin sentido en un gen para el cual se sabe que principalmente las variantes truncadas causan enfermedad	
	BP2 Observado en trans con una variante patogénica para un gen/trastorno dominante totalmente penetrante u observado en cis con una variante patogénica en cualquier patrón de herencia	
	BP3 deleciones/inserciones in-frame en una región repetitiva sin una función conocida	
	BP4 Múltiples líneas de evidencia computacional sugieren que no hay impacto en el gen o el producto del gen (conservación, evolutivo, impacto de empalme, etc.)	Advertencia: Debido a que muchos algoritmos in silico usan la misma entrada o una muy similar para sus predicciones, cada algoritmo no puede contarse como un criterio independiente. BP4 se puede usar solo una vez en cualquier evaluación de una variante.
	BP5 Variante encontrada en un caso con una base molecular alternativa para la enfermedad	

	BP6 Una fuente acreditada informa recientemente que la variante es benigna, pero el laboratorio no dispone de pruebas para realizar una evaluación independiente	
	BP7 Una variante sinónima (silente) para la cual los algoritmos de predicción de empalme no predicen ningún impacto en la secuencia de consenso de empalme ni la creación de un nuevo sitio de empalme Y el nucleótido no está muy conservado	

\*Traducida de Standards and guidelines for the interpretation of sequence variants ACMG 2015 (69)

Teniendo en cuenta estos 28 criterios, se deben aplicar las siguientes reglas para la clasificación de las variantes: (69)

### **Clasificación de variantes patogénicas**

1. 1 criterio muy fuerte (PVS1) y
  - a)  $\geq 1$  Fuerte (PS1 – PS4) o
  - b)  $\geq 2$  Moderado (PM1 – PM6) o
  - c) 1 Moderado (PM1 – PM6) y 1 de apoyo (PP1 – PP5) o
  - d)  $\geq 2$  de apoyo (PP1 – PP5)
2.  $\geq 2$  Fuerte (PS1 – PS4) o
3. 1 fuerte (PS1 – PS4) y
  - a)  $\geq 3$  Moderado (PM1 – PM6) o
  - b) 2 Moderado (PM1 – PM6) y  $\geq 2$  Apoyo (PP1 – PP5) o
  - c) 1 Moderado (PM1 – PM6) Y  $\geq 4$  Apoyo (PP1 – PP5)

### **Clasificación de variantes probablemente patogénicas:**

1. 1 muy fuerte (PVS1) y 1 moderado (PM1–PM6) o
2. 1 fuerte (PS1 – PS4) Y 1–2 moderado (PM1 – PM6) o
3. 1 fuerte (PS1 – PS4) Y  $\geq 2$  de apoyo (PP1 – PP5) o
4.  $\geq 3$  Moderado (PM1 – PM6) o
5. 2 moderado (PM1 – PM6) y  $\geq 2$  de apoyo (PP1 – PP5) o

- 42 Rendimiento diagnóstico de la secuenciación de exomas completos en trio e individual en una cohorte de pacientes colombianos con sospecha de enfermedad genética de origen monogénico
- 

6. 1 moderado (PM1 – PM6) y  $\geq 4$  de apoyo (PP1 – PP5)

#### **Clasificación de variantes benignas**

1. 1 Suficiente (BA1) o
2.  $\geq 2$  Fuerte (BS1–BS4)

#### **Clasificación de variantes probablemente benignas**

1. 1 Fuerte (BS1–BS4) y 1 de soporte (BP1–BP7) o
2.  $\geq 2$  Soporte (BP1–BP7)

#### **Clasificación de variantes de significado incierto**

1. Las variantes deben clasificarse como VSI si no cumplen otros criterios o si los criterios de benignidad y patogenicidad son contradictorios.

En la práctica han surgido dificultades con la aplicación de algunos criterios durante la clasificación de variantes, lo cual ha conducido a la publicación de algunas actualizaciones para aumentar o reducir el peso que se le da a los criterios ACMG y también para aplicar reglas dentro de algunos criterios.

El Grupo de Trabajo de Interpretación de Variantes de secuencia de ClinGen en noviembre de 2017 publicó las modificaciones en el peso de los criterios y sus cambios en la nomenclatura, allí se indica que los criterios se pueden aumentar o disminuir de peso utilizando el juicio profesional, para ello se recomienda utilizar el código del criterio original, seguido de un guion bajo y su nuevo nivel de fuerza asignado(75).

Algunas de las actualizaciones más importantes son las recomendaciones para interpretar el criterio PVS1 para variantes de pérdida de función y modificar su peso según las características de la variante(76) Se recomienda en primer lugar determinar si se trata de una variante nula nonsense, frameshift, de pérdida de exón, en el sitio canónico del splicing o de pérdida del codón de inicio, también se indica no aplicar PVS1 y PM4 concomitantemente. Adicionalmente, se debe considerar el impacto de un codón de parada prematuro para la degradación del ARN mensajero mediada por mutaciones terminadoras, conocida por sus siglas en inglés NMD (Nonsense Mediated Decay) y tener en cuenta que en general no se predice que ocurra si el codón de parada prematuro se

encuentra en el último exón y hasta 50pb del 3'UTR del penúltimo exón, esta consideración puede bajar el peso del criterio. También indica que si no se predice que ocurra NMD se debe determinar si la región truncada es crítica para la función de la proteína, también se debe tener en cuenta la tolerancia del exón a variantes de pérdida de función según lo reportado en la literatura, se recomienda revisar la posibilidad de generación de splicing alternativos o saltos de exón y además contemplar el tamaño de la región faltante en la proteína pues, una pérdida menor al 10% es menos probable que genere pérdida de función(76). Estas recomendaciones incluyen un árbol de decisión para la asignación de peso al evaluar este criterio.

De otro lado, ClinGen también publicó recomendaciones para la aplicación de la regla BA1, pues se han documentado variantes con carácter de patogenicidad que tienen una  $MAF > 0.05$ , en ella se incluye una lista de nueve variantes que no deben ser consideradas benignas a pesar de cumplir este criterio. Dichas variantes se encuentran en los genes: *ACAD9*, *GJB2*, *HFE*, *MEFV*, *PIBF1*, *ACADS* y *BTD*(77).

Se emitieron también recomendaciones para descender el criterio PM2 de moderado a criterio de soporte, y recomiendan también crear un nuevo criterio para que aquellas variantes clasificadas como PVS1 y además cumplan con la regla del PM2 se clasifiquen como probablemente patogénicas(78).

Adicionalmente el Grupo de Trabajo de Interpretación de Variantes de secuencia de ClinGen emitió una revisión para el uso del criterio PM3 y asignar el peso adecuado. Se recomienda tener en cuenta la clasificación de la otra variante documentada en trans y asignarle un puntaje si es patogénica de 0.5 y de 0.25 si es probablemente patogénica, también asigna puntaje si la fase está confirmada o no, y además recomienda tener en cuenta la frecuencia alélica para la asignación del peso(79).

Para las variantes *de novo* con criterios PS2 y PM6, actualmente recomiendan tener en cuenta tres parámetros, la paternidad confirmada del caso vs un estado de paternidad presumida, la consistencia del fenotipo y el número de observaciones *de novo*. A estas consideraciones se les asigna un puntaje y de acuerdo con la suma de estos se debe asignar el peso al criterio(80).

Se emitieron recomendaciones para el criterio de evidencia funcional PP3 y PB3, y se propone practicar una serie de cuatro pasos para definir la validez clínica de este criterio:

1. Definir el mecanismo de la enfermedad, 2. Evaluar la aplicabilidad de ensayos utilizados en el campo, 3. Evaluar la validez de las consideraciones específicas de los ensayos; 4. Aplicar la evidencia para la interpretación individual de la variante. Indican que se requiere un mínimo de once controles de variantes patogénicas y benignas en total para alcanzar evidencia de nivel moderado en ausencia de un análisis estadístico riguroso(81).

En la página web de ClinGen se encuentra publicada una carta al editor de la revista *Genetics in Medicine* donde se propone discontinuar el uso los criterios de soporte de fuentes confiables PP5 y BP6, sin embargo, aclara que la interpretación de las variantes es responsabilidad de cada director de laboratorio(82).

El reporte de las variantes debe realizarse utilizando la nomenclatura designada por el Human Genome Variation Society (HGVS), (<http://www.hgvs.org/mutnomen>), también incluir la secuencia de referencia del National Center for Biotechnology Information RefSeq database (<http://www.ncbi.nlm.nih.gov/RefSeq/>), la ubicación de la variante en el genoma “g”, ADN codificante “c” y en la proteína “p”. Además, se debe reportar la cigocidad, y la categoría de patogenicidad a la que pertenece. (69)

La realización de estudios de secuenciación de exoma o genoma genera entre 20.000 y 5´000.000 de variantes respectivamente, lo cual al no ser dirigidos al fenotipo aumentan la probabilidad de encontrar variantes incidentales o no asociadas al fenotipo del paciente en estudio(9). Esto además conduce a un aumento en la cantidad de variantes de significado incierto generando incertidumbre en los pacientes, pues una variante de significado incierto no debe ser considerada como accionable en el ámbito clínico precisamente porque no es posible asignarle patogenicidad(83). Las guías recomiendan incluir las VSI en los informes de laboratorio cuando existen pruebas de asociación del gen a la enfermedad y no se puede establecer la patogenicidad(69). Las variantes pueden ser reclasificadas en el tiempo a medida que aumenta la evidencia de estas(84), en general se conoce que la mayoría de VSI se reclasifican hacia benignas y en raras ocasiones se reclasifican hacia patogénicas porque hay evidencia del aumento en su importancia clínica(83). Un estudio realizado por Mersch y colaboradores documento la prevalencia de reclasificación de variantes en una cohorte de 1´450.000 personas con estudio genético para cáncer, encontrando que se pudo reclasificar el 6,4% de las variantes (2868/44.777), donde el

91.2% (1867/2.048) de las VSI se reclasificaron a benignas o probablemente benignas y el 7% (178/2048) fueron reclasificadas a patogénicas y probablemente patogénicas(85).

## 2.8 Definición de un estudio de exoma

Un estudio de exoma hace referencia a la evaluación de la secuenciación masiva de los exones del genoma, es decir, las regiones codificantes de los genes, que darán origen al ARN mensajero maduro, y que, al ser traducido en los ribosomas, darán origen a las proteínas, este estudio, también incluye las regiones de la unión intrón exón donde se lleva a cabo el splicing. En el caso del genoma humano, la secuenciación del exoma completo corresponde al 1.5-2% del genoma (54) y permite la identificación de variantes asociadas a enfermedad. Suele ser un examen solicitado a pacientes con un panel genético previamente negativo u otros estudios de genética molecular sin identificación de la causa de la enfermedad y también se ha considerado como primera línea en el estudio de condiciones neurogenéticas principalmente (68,86). Un estudio de exoma identifica en promedio 20.000 variantes genéticas, y es deber del laboratorio o grupo que lo analiza, identificar aquellas que tienen relevancia clínica en cada caso. Se conoce que alrededor del 85% de las enfermedades genéticas de herencia mendeliana se originan en mutaciones que ocurren en regiones exónicas (55), y es preciso utilizar programas bioinformáticos para lograr identificar las variantes que causan enfermedad.

El estudio de exoma logra una cobertura cercana al 95-98% de todas las regiones codificantes con los kits de captura actuales(87), tiene una alta sensibilidad para documentar variantes de nucleótido simple SNVs, que pueden ser missense, nonsense, en regiones canónicas del sitio de splicing, e InDels, y además a través de algoritmos bioinformáticos es posible buscar variantes en el número de copias CNVs, sin embargo, no es suficiente un solo exoma para ello, pues requiere el uso de varias muestras para poder compararlas y realizar el llamado de variantes y no es posible en todos los casos establecer puntos de corte de las variantes(68). Algunas ventajas del exoma son por ejemplo su utilidad para el diagnóstico de enfermedades con un mismo fenotipo, pero con múltiples genes causantes (heterogeneidad de locus) como el retraso en el neurodesarrollo, o encefalopatía epiléptica, porque al tener muchos genes que causan el fenotipo, puede ser más costo efectivo estudiar todo el exoma que un grupo limitado de

genes (88). Otra ventaja que ofrece el estudio de exoma, a nivel de investigación, por ejemplo, es proponer genes candidatos como causantes de enfermedad que en caso de repetirse en un grupo de pacientes con el mismo fenotipo lo clasificaría como patogénico(88). También, se pueden identificar nuevas variantes patogénicas en un gen conocido de enfermedad(10).

El objetivo final de un estudio de exoma a nivel diagnóstico, para el caso de pacientes con sospecha de enfermedad monogénica, será documentar la región génica precisa donde ocurrió una alteración y que conlleva al desarrollo de todas las características clínicas del caso en estudio, es decir documentar el genotipo que genera el fenotipo patológico y que coincida con un mecanismo de enfermedad(9).

## 2.9 Tipos de estudios de exomas

En el mercado existen diferentes tipos de estudios de exoma que pueden ser solicitados según los requerimientos de cada paciente y brindan diferentes ventajas unos respecto a otros. Dentro de este grupo se encuentran:

- Exoma completo: También conocido como WES (por sus siglas en inglés, whole exome sequencing) abarca el estudio de los 20.000 genes en un individuo único (probando) utilizando kits de captura especiales para cubrir todas las regiones codificantes del genoma y las regiones de unión intrón-exón.
- Exoma clínico: También conocido como mendelioma: Utiliza kits de captura para 5.000-7.000 genes que se conoce que se encuentran asociados a enfermedades humanas conocidas de origen hereditario, reportadas en bases de datos como OMIM, Orphanet o HGMD(89).
- Exoma dirigido: Es un exoma clínico que cuenta con librerías enriquecidas para los genes que se sospecha que podrían ser los causantes de la enfermedad en el paciente en estudio. Puede formar parte de los paneles de genes para condiciones que tienen una amplia heterogeneidad genética(54).
- Exoma individual: Es el análisis de la secuenciación de exoma de un probando.
- Exoma trio: Es el análisis del exoma del caso índice (probando) junto al de sus progenitores.



El estudio de exomas en trio ha tenido una gran acogida para el estudio de enfermedades monogénicas, pues facilita la interpretación de variantes benignas raras que se segregan en la familia, o definir si son *de novo*(9), debido a que con el análisis en trio, es posible descartar con mayor confianza aquellas variantes raras localizadas en genes asociados a enfermedad con un mecanismo de herencia autosómico dominante y con penetrancia completa, que se identifican en alguno de los padres y en el probando y que no parecen generar ningún fenotipo en el padre portador. También en el caso de entidades recesivas, el exoma trio permite definir el estado de portador en los padres en lugar de realizar secuenciación Sanger para evaluar segregación de las variantes, esto para algunos especialistas, se refleja como una reducción en el tiempo de interpretación de variantes(9).

A pesar de las ventajas que parece ofrecer el análisis en trio de los exomas, no existen estudios con tamaños de muestra lo suficientemente grandes para establecer si realmente el análisis en trio mejora el rendimiento diagnóstico de los estudios de exoma. En un estudio realizado por *Tan y colaboradores*, tomaron un grupo de 30 pacientes con sospecha de enfermedad monogénica, en los cuales 11 de ellos (36.7%) fueron diagnósticos y el estudio en trio solo mejoró para la identificación de un caso más (12/30, es decir, 40%) sin una mejoría significativa en el rendimiento de la técnica, el estudio en trio duplicó los costos, pero obvió la necesidad del estudio de segregación por Sanger y el número de variantes para curación (7)

## **2.10 Revisión de la literatura del rendimiento de exomas en la práctica clínica**

Son pocos los estudios que comparan los rendimientos diagnósticos del análisis de exoma único o en trio y especialmente, que establecen cuál es el grupo de enfermedades para el cual hay un mayor rendimiento. En la actualidad no existen guías que definan los criterios para solicitar un estudio de exoma, sin embargo, existen algunas recomendaciones y, en general los motivos de solicitud son(9):

1. Trastorno del neurodesarrollo: por ejemplo, retraso del desarrollo y / o dificultades en el aprendizaje, encefalopatía epiléptica o parálisis cerebral severa.

2. Anomalías congénitas: múltiples anomalías congénitas (dos o más anomalías mayores) o una sola anomalía mayor junto con un trastorno del desarrollo neurológico, crecimiento aberrante, características dismórficas o comportamiento inusual.
3. Parámetros de crecimiento anormales (talla, peso, circunferencia occipitofrontal): dos o más parámetros >3 DE por encima o por debajo de la media o un parámetro único >4 DE por encima o por debajo de la media (excepto para la obesidad donde el umbral para obesidad aislada es >4.5 DE junto con una fuerte sospecha de una etiología genética).
4. Características dismórficas.
5. Fenotipo conductual inusual junto con una o más de las características anteriores con sospecha de una base genética.
6. Trastorno de impacto considerable para el cual se cree una base genética con lo siguiente: varios afectados en la familia; otro miembro de la familia afectado con un fenotipo raro, consistente y distintivo; o un solo caso con un fenotipo particularmente severo.
7. No ha sido posible el diagnóstico con otras pruebas genéticas(65)

El rendimiento general de la secuenciación de exoma completo está estimado entre un 25 a un 52%(68), y hasta el 2016 según *Bertier y colaboradores* y en 2017 *Schwarze y colaboradores*, en dos revisiones sistemáticas documentaron que no se podía determinar si el exoma se indicaba como estudio de primera línea para sospecha de entidades genéticas(90)(91).

Respecto al rendimiento diagnóstico general de la secuenciación de exoma, *Yang y colaboradores 2014* reportaron un diagnóstico molecular de 25.2% (504/2000 pacientes) en pacientes con sospecha de enfermedad mendeliana(92). El consorcio FORGE de Canada reportó por *Sawyer y colaboradores 2016* un rendimiento diagnóstico general de los exomas del 29% (105 de 362) para pacientes con sospecha de enfermedades genéticas(93). Otros estudios como el de *Dillon y colaboradores* en 2018, encontraron que el exoma mejora el rendimiento diagnóstico en un 23% respecto a un panel genético dirigido en pacientes con enfermedades con sospecha genética de fenotipo heterogéneo(88).

Según los grupos de enfermedad, el rendimiento en alteraciones neurogenéticas ha sido evaluado por diferentes autores logrando los mejores resultados de rendimiento diagnóstico dentro de los diferentes grupos de enfermedad; *Dong y colaboradores 2019* evaluaron en una cohorte de pacientes chinos entre 1 y 16 años con alteraciones del neurodesarrollo, encontrando un rendimiento diagnóstico del exoma de 41,38% y concluyen que puede ser un test de primera línea para el diagnóstico de este tipo de enfermedades(86).

En Latinoamérica, *Córdoba y colaboradores 2019* en Argentina, quisieron explorar la utilidad de la secuenciación de exoma completo en pacientes con sospecha de una condición neurogenética y evaluar la rentabilidad de su implementación como primera línea. Este estudio incluyó 40 pacientes, de diferentes edades con sospecha de enfermedad neurogenética. Encontraron que el exoma tiene un rendimiento de 40% (16/40) para el diagnóstico de estas entidades, el tiempo promedio desde el inicio de los síntomas y el diagnóstico con exoma fue de 11 años, al hacer el estudio retrospectivo de los múltiples exámenes y valoraciones el costo promedio de la trayectoria diagnóstica es de 1.646 USD y se estima que se podrían ahorrar el 60% de los recursos solicitando un exoma de forma temprana. Los autores concluyen que la secuenciación de exoma podría ser un estudio de primera línea de entidades neurogenéticas, ahorrando tiempo y costos(8)

*Reuter y colaboradores 2020* evaluaron el rendimiento de la secuenciación de genoma completo con análisis en trio en población pediátrica con anomalías cardíacas identificando un rendimiento del 12.6% (14/111 familias)(94). En población adulta *Siedelman y colaboradores 2017* encontraron un rendimiento de exoma de 26.5% en un grupo de 200 pacientes con enfermedad cardiovascular donde el mayor rendimiento se obtuvo para pacientes con QT prolongado(95).

En hematología, *Saes y colaboradores 2018* encontraron un rendimiento diagnóstico de secuenciación de exoma completo del 17% en un grupo de pacientes con diátesis hemorrágica(96). Para la evaluación de alteraciones plaquetarias *Romasko y colaboradores 2017* documentaron un rendimiento de 23.8% de la secuenciación de exoma(97). Para el estudio de condiciones nefrológicas *Groopman y colaboradores 2019* lograron un diagnóstico en el 10% de una cohorte de 3.315 pacientes con enfermedad renal crónica de instauración temprana(98), pero para el estudio de pacientes con

síndrome nefrótico resistente a esteroides el exoma tuvo un rendimiento del 25% de acuerdo con un estudio realizado por *Warejko y colaboradores 2017* (99).

*Polla y colaboradores 2015* reportaron un rendimiento diagnóstico de secuenciación de exoma dirigido de 22% para displasias esqueléticas(100). *Kingsmore y colaboradores 2019* reportaron un rendimiento diagnóstico del 24% utilizando genoma o exoma en un grupo de pacientes críticamente enfermos admitidos en unidad de cuidado intensivo, sin embargo, se sospecha que el bajo rendimiento fue debido a una mala selección de los pacientes no valorados por genetista(101).

El análisis de exomas también ha sido estudiado en diagnóstico perinatal; *Vora y colaboradores 2020* evaluaron 102 exomas en trio de casos con ecografía con al menos una malformación fetal e identificaron un rendimiento del 20.6% (21 casos positivos), sin embargo, consideran que la información disponible en bases de datos de fenotipos prenatales es escasa(102).

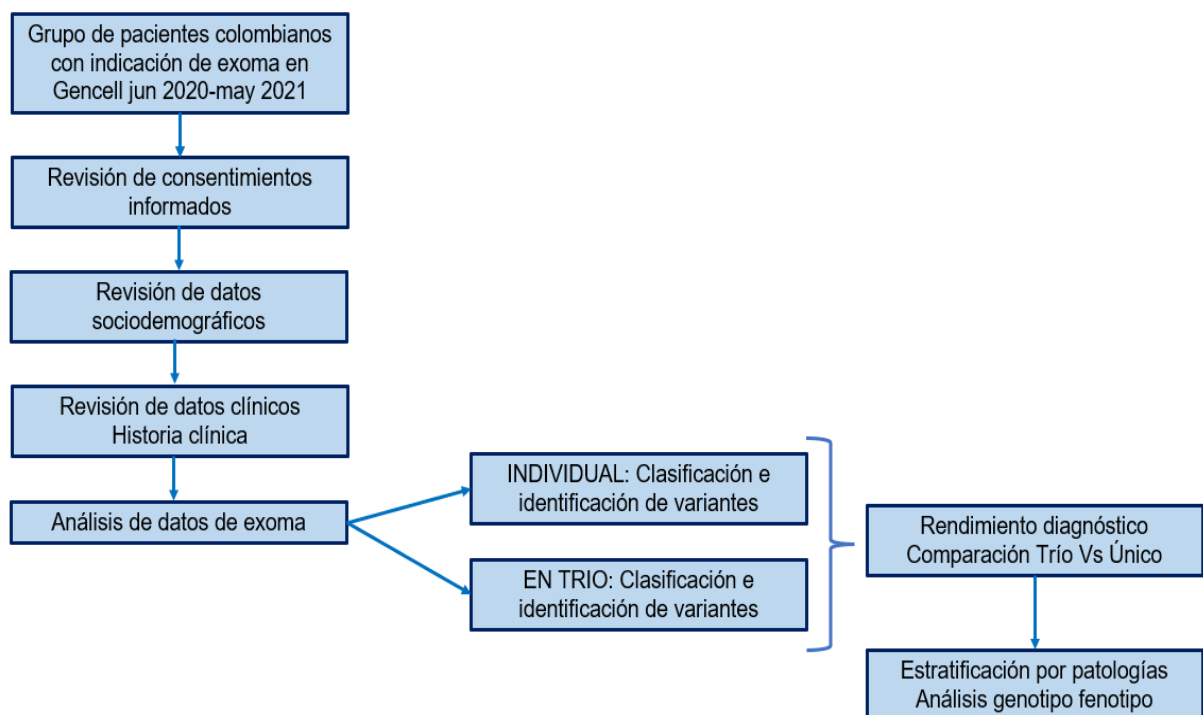
## 3 Metodología

### 3.1 Tipo de estudio

Estudio de corte transversal observacional con componente analítico.

### 3.2 Diagrama de la investigación

Figura 3-1 Diagrama de la investigación



### 3.3 Población

Grupo de pacientes colombianos a quienes se les realizó estudio de exoma, en el laboratorio Gencell de Bogotá, en un periodo de tiempo entre junio de 2020 hasta mayo de 2021, que cumplían los criterios de selección.

## **3.4 Criterios de selección**

### **3.4.1 Criterios de inclusión**

- Pacientes colombianos (nacidos en Colombia según documento de identificación) a quienes se les haya realizado estudio de exoma por indicación médica por sospecha de enfermedad genética en el laboratorio Gencell de Bogotá entre junio de 2020 y mayo de 2021.
- Exomas realizados inhouse utilizando el MGIEasy Exome Capture V5 siguiendo el protocolo dispuesto por la casa comercial para preparación de librerías(62) y con el secuenciador DNBSEQ-G400.
- Personas que hayan firmado consentimiento informado para la utilización de sus datos de forma anónima en estudios de investigación científica y para el caso de menores de edad, sus representantes legales.

### **3.4.2 Criterios de exclusión**

- Exomas de familiares en primer grado ya incluidos en el estudio.
- Exomas de pacientes sin disponibilidad de historia clínica relacionada con el motivo de la solicitud.
- Pacientes donde no se encuentre disponible la información del trio o donde no sea posible asumir la paternidad o maternidad de acuerdo con los datos genómicos.

## **3.5 Muestra**

Se utilizó una muestra a conveniencia no probabilística, en la que se incluyeron 100 casos de pacientes colombianos que cumplían los criterios de selección previamente descritos, en el periodo de tiempo establecido.

## **3.6 Recolección de la información**

La información se obtuvo a través de dos fuentes principales, historia clínica y datos de secuenciación: La copia de la historia clínica, que es una fuente secundaria de la cual se

tomaron los datos demográficos fenotípicos y clínicos de cada uno de los casos descritos por el especialista que realizó la solicitud de exoma.

Archivos fastq y vcf, estos últimos obtenidos después de la limpieza y análisis de calidad de los datos crudos generados por el secuenciador masivo de ADN **DNBSEQ-G400**, y kit de preparación de librerías **MGIEasy Exome Capture V5 probe set**, de los exomas de los pacientes incluidos en el estudio, este kit tiene una cobertura de la región on-target, de 69Mb para las regiones codificantes del genoma, las regiones de la unión intrón-exón, y el genoma mitocondrial, así como las regiones asociadas a enfermedad. Utiliza las bases de datos CCDS, GENCODE, RefSeq, miRBase, MGI Clinical Database (62) para el diseño de sondas de enriquecimiento, requiere 250ng de ADN para su adecuada hibridización(103). El llamado de variantes para la obtención del archivo vcf de todos los casos, se realizó de acuerdo con los flujos de trabajo del Broad Institute utilizando las mejores prácticas de GATK (Genome Analysis Toolkit) v3.3, a partir de la versión del genoma humano GRCh37 (hg19), con la posterior anotación y filtrado utilizando el Software bioinformático, GenomeOne.

### **3.7 Técnicas de análisis para la evaluación de variantes y su interpretación clínica**

A continuación, se presenta en forma de diagrama el proceso de análisis de datos para la evaluación, clasificación e interpretación de variantes en el presente estudio, este análisis se realizó tanto para el análisis de los exomas individuales como en trio Figura 3-2:

El análisis inició a partir de los archivos vcf cargados en el software GenomeOne que arroja en promedio entre 20.000 y 23.000 variantes. En primer lugar, se realizó un filtro por términos HPO de las manifestaciones clínicas de cada paciente referidas en la historia clínica, en este filtro los genes obtenidos en cada término HPO son aditivos y no excluyentes. En este punto, se obtiene un listado donde se debe evaluar en conjunto si cada variante está asociada con el fenotipo presentado en el paciente y para ello se revisa la base de datos OMIM y se filtra si el gen de cada variante corresponde con la clínica del paciente. Es importante aclarar que el abordaje utilizado en este estudio fue principalmente de fenotipo a genotipo, pues al tratarse de un estudio de rendimiento diagnóstico, no se estaban identificando posibles genes candidatos asociados a enfermedad.

Una vez seleccionadas aquellas variantes asociadas al fenotipo se definió si la fracción del alelo alternativo identificado en cada variante correspondía con un estado de cigosidad. Cuando las fracciones son cercanas al 50% corresponde a un estado heterocigoto y cuando son cercanas al 100% corresponde a un estado homocigoto. En este paso, aquellas variantes menores al 25% fueron descartadas por criterios de calidad y posiblemente corresponder a un artefacto.

Posteriormente, para las variantes preseleccionadas en los pasos anteriores se analizaba si la cigosidad de la variante coincidía con el **mecanismo de herencia** de la enfermedad causada por ese gen, lo anterior para dilucidar si se trataba de una enfermedad dominante con una variante heterocigota que explique la condición del paciente, o si se trataba de una enfermedad recesiva con una variante en homocigosis o dos variantes en heterocigosis compuesta que puedan explicar la condición. Para las enfermedades ligadas al cromosoma X se revisaba en los hombres si estaba en hemicigosis, cercana a 100% y en las mujeres si se encontraba en heterocigosis cercana al 50%.

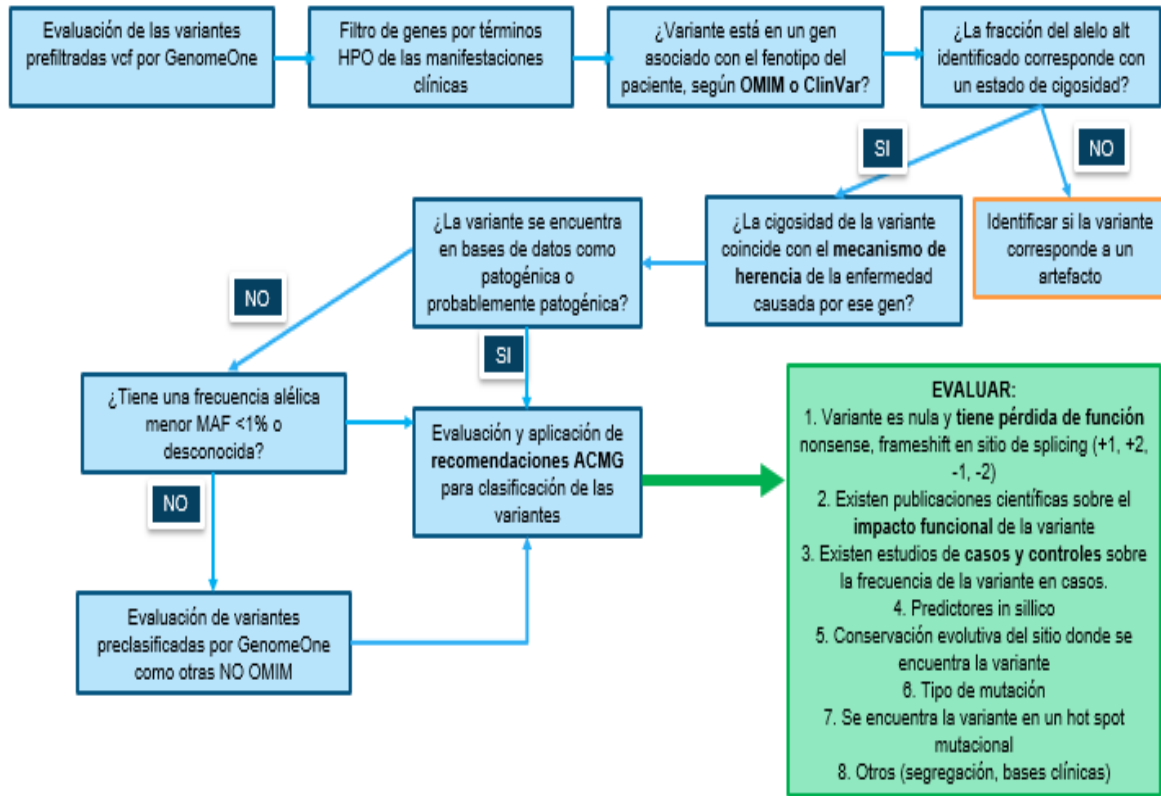
Luego se realizaba la pregunta: ¿La variante se encuentra en bases de datos como patogénica o probablemente patogénica? Para responderla se revisaron bases de datos como Clinvar, The Human Mutation Database, MalaCards, HGMD, LOVD, ClinVar Miner, InterVar, DGV, Varsome. Si se trataba de una variante ampliamente descrita, se procedió con el resto del análisis aplicando los criterios ACMG.

Cuando no se identificaba una variante ampliamente descrita, se evaluó la frecuencia poblacional revisando si la frecuencia alélica menor (MAF) de la variante es menor al 1% o desconocida en bases de datos poblacionales. En caso de tratarse de una variante rara o desconocida se continuaba con la aplicación de los criterios ACMG. Por otro lado, cuando se trataba de una variante muy frecuente y no estaba descrita en la literatura como de importancia clínica se decidió realizar su eliminación.

Dado que el software preclasificaba algunas variantes como patogénicas o probablemente patogénicas a pesar de no estar asociada a ningún gen OMIM, se realizaba la respectiva revisión de la literatura para identificar si el gen había sido descrito por su asociación a enfermedad para posteriormente continuar con la clasificación.



Figura 3-2 Proceso de evaluación, clasificación e interpretación de variantes



Para realizar el filtrado de las variantes y asignación de la patogenicidad se utilizaron:

- Términos de Human Phenotype Ontology (HPO)
- Bases de datos poblacionales: 1000 Genomes, ExAc gnomAD, Ensembl.
- Bases de datos clínicas: OMIM, ClinVar, The Human Mutation Database, MalaCards, HGMD, LOVD, ClinVar Miner, InterVar, DGV, Varsome.
- Predictores in silico: Mutation Taster, PolyPhen-2, SIFT, SNPAnalyzer, Condel, GeneSplicer, CAAD Score, Revel, SpliceAI.
- Otras bases de datos: NCBI, Pubmed, UCSC Genome Browser, UniProt,

El último paso fue la evaluación y aplicación de las recomendaciones ACMG 2015 para la clasificación de las variantes, que incluyó la verificación y anotación manual de las variantes preseleccionadas, y posterior clasificación como patogénicas (P), probablemente patogénicas (PP), de significado incierto (VSI), benignas o probablemente benignas.

También se aplicaron las nuevas actualizaciones propuestas por ClinGen ya descritas. Figura 3-4.

Se evaluó si las variantes eran nulas o asociadas a un gen donde la pérdida de función es un mecanismo de pérdida de función, para la aplicación de este criterio se utilizaba el árbol de decisión de Clingen(76) y herramientas como Auto PVS1, Decipher, UCSC Genome Browser y los scores de Clingen. Para evaluar si una variante con diferente nucleótido producía el mismo cambio de aminoácido y estaba descrito como patogénico (criterio PS1), se usaron herramientas como Clinvar, GnomAD, UCSC, UniProt. Estas herramientas también fueron aplicadas para asignar el criterio PM5.

El criterio PS2 y PM6 solo podía utilizarse en la etapa de análisis de los exomas en trio, al confirmar la paternidad y verificar la segregación de las variantes. Los criterios PS3 y BS3 para documentar la presencia de estudios funcionales del efecto deletéreo en la variante y el criterio PS4 para buscar estudios de casos y controles se aplicó a través de revisión de publicaciones en la literatura Pubmed, Mastermind, Franklin, Varsome, Clinvar entre otros.

Se utilizaron herramientas como UCSC y Uniprot para evaluar el criterio PM1 de variante localizada en un hotspot mutacional o en un dominio de la proteína; para el criterio PM2, BA1 y BS2 se usaron bases poblacionales 1000Genomas y GnomAD principalmente. El criterio PM3 para establecer si las variantes estaban en cis o trans con una variante patogénica en un gen con mecanismo de herencia autosómico recesivo solo podía usarse en los análisis en trio a través de la segregación de la variante. Los criterios PM4 y BP3 para evaluar los cambios en la longitud de la proteína se aplicaban a través de UCSC, UniProt.

En la aplicación de PP2 se utilizaron bases de datos como Varsome y GnomAD teniendo en cuenta valores de ZS >3.08. Para examinar los predictores in silico se utilizó principalmente Revel con un puntaje > 0.644 para la asignación de PP3 y <0.29 para BP4, pues esta herramienta utiliza en su algoritmo varios predictores in silico, también se usó SpliceAI para aquellas variantes en el sitio de splicing. Para PP5 y BP6 donde se tiene en cuenta si la variante fue evaluada previamente por una fuente confiable se usó Clinvar, Franklin, Varsome, HGMD.

En caso de obtener un resultado positivo se buscó información en la literatura, para definir si existe relación genotipo-fenotipo.

Figura 3-4 Proceso de evaluación, clasificación e interpretación de variantes

	Benigno			Patogénico		
	Fuerte	Soporte	Soporte	Moderado	Fuerte	Muy Fuerte
Bases de datos poblacionales	MAF es demasiado alta para el trastorno BA1/BS1 O la observación en controles inconsistentes con la penetrancia de la enfermedad BS2			Ausente en bases de datos poblacionales PM2	La prevalencia en afectados esta significativamente aumenta comparada con controles PS4	
Datos computacionales y predictivos		Múltiples líneas de evidencia computacional sugieren que no hay impacto en el gen o producto génico BP4  Variante missense en un gen para el cual se sabe donde solo variantes truncadas causan la enfermedad BP1  Variante sinónima con la cual los algoritmos de predicción de empalme no predicen ningún impacto BP7	Múltiples líneas de evidencia computacional respaldan un efecto deletéreo en el gen/producto génico PP3	Variante missense nueva en un aminoácido donde existe otro cambio missense patogénico PM5  Variante de cambio de longitud de proteína PM4	Mismo cambio de aminoácido que una variante patogénica establecida PS1	Variante nula en un gen donde LOF es un mecanismo conocido de enfermedad PVS1
Datos funcionales	Estudios funcionales bien establecidos no muestran efectos deletéreos BS3		Variante missense en un gen donde se ha reportado una baja frecuencia de variantes missenses benignas y además donde las variantes missenses son causa frecuente de enfermedad PP2	Hotspot mutacional o dominio funcional bien estudiado sin variación benigna PM1	Estudios funcionales bien establecidos muestran un efecto nocivo en PS3	
Datos de segregación	No segregación con enfermedad BS4		Segregación conjunta con enfermedad en varios miembros de la familia afectada PP1	Incremento en los datos de segregación		
Datos de novo				De novo (sin paternidad y maternidad confirmada) PM6	De novo (paternidad y maternidad confirmada) PS2	
Datos Alélicos		Observado en <i>trans</i> con una variante dominante PB2  Observado en <i>cis</i> con una variante patogénica BP2		Para trastornos recesivos, detectados en <i>trans</i> con una variante patogénica PM3		
Otras bases de datos		Fuente acreditada sin datos compartidos = BP6 benigno	Fuente acreditada = PP5 patogénica			
Otros datos		Encontrado en caso con una causa alternativa BP5	Fenotipo del paciente o FH altamente específico para gen PP4			

\*Traducida de Standards and guidelines for the interpretation of sequence variants ACMG 2015 (69)

Se exploró la presencia de hallazgos incidentales evaluando los genes incluidos en la lista más reciente propuesta por la ACMG SF V3.1 del 2022 que incluye un grupo de 78 genes asociados a cánceres, enfermedades cardiovasculares, enfermedades metabólicas y otras condiciones con guías de manejo clínico establecidas que incluye los siguientes

genes(104): *ACTA2, ACTC1, ACVRL1, APC, APOB, ATP7B, BAG3, BMPR1A, BRCA1, BRCA2, BTD, CACNA1S, CASQ2, COL3A1, DES, DSC2, DSG2, DSP, ENG, FBN1, FLNC, GAA, GLA, HFE, HNF1A, KCNH2, KCNQ1, LDLR, LMNA, MAX, MEN1, MLH1, MSH2, MSH6, MUTYH, MYBPC3, MYH11, MYH7, MYL2, MYL3, NF2, OTC, PALB2, PCSK9, PKP2, PMS2, PRKAG2, PTEN, RB1, RBM20, RET, RPE65, RYR1, RYR2, SCN5A, SDHAF2, SDHB, SDHC, SDHD, SMAD3, SMAD4, STK11, TGFBR1, TGFBR2, TMEM127, TMEM43, TNNC1, TNNT1, TNNT2, TP53, TPM1, TRDN, TSC1, TSC2, TTN, TTR, VHL, WT1*, donde se exploraron solo las variantes patogénicas o probablemente patogénicas.

### **3.8 Análisis estadístico**

En primer lugar, se realizó estadística descriptiva para las variables demográficas y clínicas de los individuos incluidos en el estudio. Posteriormente, se realizó el cálculo de las frecuencias de las variables clínicas, de las variantes genéticas y de los grupos fenotípicos.

Luego, se calculó el rendimiento diagnóstico de los exomas en análisis único, en análisis en trio y por grupo de fenotipos, es decir la proporción de resultados positivos en pacientes respecto al total de pacientes analizados.

El análisis estadístico se realizó utilizando el software estadístico SPSS versión 26(105).

### **3.9 Control de errores y sesgos**

#### **3.9.1 Sesgos de selección**

La selección de los casos depende de los convenios del laboratorio, esto hace referencia a que es posible que la mayoría de los casos que accede al estudio de exoma pertenece más al régimen contributivo colombiano que al régimen subsidiado debido a los convenios de contratación del laboratorio y por la mayor facilidad que tienen los médicos para solicitar un estudio de exoma a un paciente del régimen contributivo.

### 3.9.2 Sesgos de información

- Imposibilidad de la valoración presencial de los pacientes incluidos en la población de estudio y de evaluar su fenotipo clínico, debido a que los datos clínicos fueron obtenidos a partir de una fuente secundaria, y no fue posible verificar las características clínicas y el fenotipo de los casos índice.
- Sesgo de memoria en la historia clínica, donde el informante no registra la totalidad de las características clínicas relevantes de cada paciente y/o la historia familiar.
- Sesgo de detección según la cobertura del exoma realizado en regiones particulares que podrían ser de interés para un caso determinado, y la calidad del llamado de las variantes.

### 3.9.3 Sesgos de confusión

El único sesgo de confusión identificado es la aparición de variantes de significado incierto (VSI), debido a que son aquellas que a la luz del conocimiento actual se desconoce si son las responsables del fenotipo causante de la enfermedad porque se carece de literatura disponible que soporte su interpretación como una variante patogénica y debe ser reclasificada periódicamente, con el fin de otorgarle un criterio de benignidad o patogenicidad en el futuro. Para controlar este sesgo no se tuvo en cuenta como aporte en el rendimiento diagnóstico.



## 4 Consideraciones éticas

De acuerdo con la Resolución 8430 de 1993 el presente estudio, se clasificó como un estudio sin riesgo, debido a que implicó la recolección de datos de fuentes retrospectivas, el uso de herramientas bioinformáticas de análisis genético que no conllevaba a ningún perjuicio para los sujetos de investigación o sus datos.

Para el desarrollo de este proyecto se trabajó con datos de pacientes colombianos con indicación médica de exoma, los datos fueron disociados y anonimizados para proteger la confidencialidad de los individuos de estudio. Además, para el desarrollo de este protocolo, no se requirió contactar a los pacientes, debido a que es deber del laboratorio informar los resultados obtenidos por solicitud del médico tratante, además previo a la realización del exoma por solicitud médica, los individuos o sus representantes legales (en caso de ser menores de edad o con discapacidad cognitiva) firmaron un consentimiento informado en el cual se dio autorización para el uso de los datos para protocolos de investigación de forma anónima.

Este estudio fue aprobado por el Comité de Ética en Investigación de la Facultad de Medicina de la Universidad Nacional de Colombia bajo el Acta de evaluación No. 008-051 del 22 de abril de 2021.

La investigadora principal y maestrante participaron en la ejecución de este protocolo a partir de la revisión de consentimientos informados para definir los pacientes que fueron incluidos en el estudio, también participaron en la recolección de la información de las historias clínicas, el procesamiento de gran parte de estudios de secuenciación, la totalidad del análisis bioinformático a partir de los archivos .vcf, la organización de las variables de estudio, el análisis y la interpretación de los datos.

Para el tratamiento de los datos biológicos del presente estudio, se firmó un acuerdo de confidencialidad y no divulgación, entre la tesista Estefanía Rodríguez Alvarino y el representante legal de Gencell S.A.S, sobre la propiedad y confidencialidad de los datos biológicos y bioinformáticos que surjan en el presente estudio y que para ser publicados se respetará la anonimidad de los mismos.

Este estudio se rigió bajo las directrices establecidas en la declaración de Helsinki del año 2013 como principios éticos para las investigaciones médicas en seres humanos. Se desarrolló en el marco de la Ley 23 de 1981, Código de ética Médica en Colombia, y la declaración de Taipéi de 2016 sobre las consideraciones éticas de las bases de datos de salud y biobancos.

#### **4.1 Declaración de existencia de conflicto de intereses**

La investigadora y la codirectora de este proyecto de investigación declaran no tener conflicto de intereses para la realización de este trabajo. Sin embargo, se declara que la Dra. Marcela Gálvez, actualmente ejerce en la dirección médica-científica de Gencell y que el análisis y resultados obtenidos de esta investigación no se utilizaran como beneficio comercial del laboratorio.

#### **4.2 Propiedad intelectual**

Este proyecto de investigación se rige por lo estipulado en el acuerdo No. 35 de 2013 de la Universidad Nacional de Colombia por el cual se expide el reglamento de propiedad intelectual, que define que la propiedad intelectual “es un derecho complejo de dominio especial sobre las creaciones del talento humano que se concede a los autores o inventores”, y que, por su carácter de proyecto de tesis de maestría, se concede a los autores de este proyecto, respetando la confidencialidad sobre los hallazgos del mismo.



## **5 Resultados**

### **5.1 Definición de la muestra de estudio**

Inicialmente, se tomaron 121 casos a los que se les realizó estudio de exoma en el periodo de tiempo comprendido entre junio de 2020 y mayo del 2021, al revisar los criterios de elegibilidad, 14 casos no contaban con historia clínica del médico solicitante o no habían aceptado la realización de estudios de investigación de forma anonimizada. Para los 107 casos restantes, 4 de ellos fueron excluidos del estudio debido a que tuvieron que ser analizados en dúo (madre y probando), porque según los datos genómicos, la paternidad no podía ser asumida, 2 casos más fueron descartados porque eran hermanos de un caso ya incluido en el estudio, y otro caso fue excluido porque tenía diagnóstico de trisomía 21 y ya había sido previamente documentada por cariotipo, y por tanto no aportaría al rendimiento diagnóstico del presente estudio. Lo anterior, resultó en un tamaño de muestra final de 100 individuos, los cuales fueron analizados de manera individual y en trio (en conjunto con el exoma de sus dos progenitores), para un total de 300 exomas.

### **5.2 Descripción de las características demográficas, clínicas y fenotípicas de la población estudiada.**

La muestra de pacientes estudiada, con indicación de realización de exoma entre junio de 2020 y mayo de 2021 estuvo compuesta en un 51% por mujeres. Los 100 individuos estudiados eran procedentes de 16 departamentos diferentes de Colombia, donde Bogotá fue el lugar con mayor número de individuos remitidos en un 41%, seguido por Antioquia 30% y Risaralda 5%.

El 7% de los pacientes tenían antecedente de consanguinidad y el 18% tenían historia familiar de un afectado en primer grado con fenotipo similar. Al realizar la revisión de las historias clínicas, se encontró que el 34% tenían estudios genéticos previos de tipo cariotipo, hibridación genómica comparada (CGH) por microarreglos, MLPA, análisis de genes únicos o paneles. De otro lado, se encontró que la especialidad que más remite fue genética en un 96%, seguido por neuropediatría en un 2%. (Tabla 5-1).

Tabla 5-1. Características demográficas

<b>SEXO</b>		
	Frecuencia	Porcentaje %
F	51	51.0
M	49	49.0
Total	100	100.0
<b>DEPARTAMENTO</b>		
	Frecuencia	Porcentaje %
Bogotá	41	41.0
Antioquia	30	30.0
Risaralda	5	5.0
Boyacá	3	3.0
Sucre	3	3.0
Caldas	3	3.0
Cundinamarca	3	3.0
Tolima	3	3.0
Córdoba	2	2.0
Valle del Cauca	1	1.0
Quindio	1	1.0
Bolivar	1	1.0
Santander	1	1.0
Huila	1	1.0
Meta	1	1.0
Nariño	1	1.0
<b>CONSANGUINIDAD</b>		
	Frecuencia	Porcentaje %
NO	93	93.0
SI	7	7.0

<b>HISTORIA FAMILIAR</b>		
	Frecuencia	Porcentaje %
NO	82	82.0
SI	18	18.0

<b>OTROS ESTUDIOS GENÉTICOS PREVIOS</b>		
	Frecuencia	Porcentaje %
NO	66	66.0
SI	34	34.0

<b>ESPECIALIDAD QUE REMITE</b>		
	Frecuencia	Porcentaje
Genética	96	96.0
Hematooncología	1	1.0
Inmunología	1	1.0
Neuropediatría	2	2.0

La media (DE) de la edad de los 100 individuos estudiados fue de 8.76 (8.06). El paciente más joven tenía 0 años y el mayor 41 años. El 50% de los pacientes tenía edad entre 3 y 6 años, valores que corresponden al percentil 25 y 75 respectivamente. La mediana o percentil 50 fue de 6 años. Con respecto a la edad, el valor de  $p < 0.05$  indica que los datos observados son atípicos bajo un modelo estadístico que asume que la hipótesis nula de normalidad de los datos es cierta. Por lo tanto, se rechazó la hipótesis nula de normalidad. (Tabla 5-2).

Tabla 5-2. Resumen de la edad de la muestra estudiada.

<b>EDAD</b>			
Media (DE)	Mediana (RIQ)	Máximo - Mínimo	Valor p
8.76 (8.06)	6 (3-6)	0-41	0.00001

Al evaluar los fenotipos, se encontró que la característica fenotípica más frecuente en la muestra estudiada fue el retardo en el desarrollo psicomotor en el 47% de los individuos, seguido por retardo en el desarrollo del lenguaje (34%), dimorfismo facial (32%), epilepsia (30%) y anomalías menores (26%). En la tabla 5-3 se presenta un resumen descriptivo del número de individuos con 41 características fenotípicas identificadas. Cabe aclarar que dos o más características fenotípicas podían estar presentes en un individuo.

El compromiso del sistema nervioso a nivel funcional o estructural es el grupo de características fenotípicas mayormente representadas para realizar estudio de exoma.

Tabla 5-3. Caracterización fenotípica de la población

Descripción fenotípica	
Característica fenotípica	Individuos (n)
Retardo en el desarrollo psicomotor	47
Retardo en el desarrollo del lenguaje	34
Dismorfismo facial	32
Epilepsia	30
Anomalías menores	26
Baja talla	22
Alteraciones estructurales SNC	20
Discapacidad intelectual	18
Bajo peso	18
Trastorno del comportamiento	17
Compromiso óseo	17
Infecciones a repetición	14
Compromiso cardíaco	13
Compromiso inmunológico	13
Alteraciones endocrinas	13
Hipotonía	13
Compromiso gastrointestinal	13
Microcefalia	12
C. Visual	12
Compromiso articular	12
Trastorno de la deglución	12
Compromiso en piel	12
Autismo	11
Movimientos anormales / estereotipias	10
C. Auditivo	9
Alteraciones en metabolitos	9
Dolor / parestesias	8
Parálisis	7
Neuropatía	7
Apnea del sueño	7

Compromiso renal	7
Alteraciones de los reflejos	7
Compromiso hematológico	6
Ataxia / marcha inestable	6
Macrocefalia	6
Compromiso pulmonar	5
Disautonomía	5
Miopatía	4
Hipertensión pulmonar	4
Compromiso muscular	4
Alteración de la diferenciación sexual	3

### 5.3 Rendimiento diagnóstico del exoma completo individual y en trio

Para el análisis de rendimiento diagnósticos, se consideraron positivos aquellos casos donde se identificaron variantes patogénicas (P) o probablemente patogénicas (PP), de acuerdo con las recomendaciones de la ACMG, en genes que al estar comprometidos explicaba el fenotipo descrito del paciente, y el mecanismo de herencia estaba representado por la cigosidad de las variantes. Todos los demás, fueron considerados casos negativos. La identificación de variantes de significado incierto VSI no se tuvo en cuenta en el rendimiento diagnóstico porque por definición no existe evidencia suficiente para considerarlas como causantes de enfermedad. El rendimiento diagnóstico para el análisis de exomas de forma individual fue del 29% (29).

Este rendimiento fue superado en un 3% por el análisis en trio, para un rendimiento total de la prueba de 32%, es decir 32 casos considerados positivos. Los tres casos cuyo resultado fue positivo en el análisis trio, pero no en el individual, corresponden a dos casos de heterocigosis compuesta en los cuales fue posible identificar la segregación de las variantes y asignar el criterio de PM3 en el análisis en trio al confirmar el estado en *trans* de las variantes, mientras que en el análisis individual fueron clasificadas como VSI por falta de este criterio. El tercer caso corresponde a un paciente en quien se asignó una clasificación de VSI en el análisis individual, pero al confirmar que se trata de un hallazgo *de novo*, fue posible agregar el criterio PS2\_Moderado a la puntuación y por tanto reclasificarla como probablemente patogénica. *Tabla 5-4.*

Tabla 5-4. Rendimiento diagnóstico según tipo de análisis de exoma.

Rendimiento diagnóstico estudio de exoma completo	
Exoma Individual	29(29,0%)
Exoma Trio	32(32,0%)

A continuación, se presenta el listado de casos diagnósticos en los exomas trio con la descripción de las variantes documentadas. Tabla 5-5.

## 5.4 Descripción de casos diagnósticos y variantes identificadas.

De los 100 casos analizados, 90 tuvieron al menos una variante de significado incierto a ser reportada por su probable asociación con el fenotipo y 10 casos no tuvieron hallazgos genéticos clínicamente relevantes. En total se identificaron 113 variantes de significado incierto, 18 variantes patogénicas y 19 variantes probablemente patogénicas.

Es de anotar que el análisis se realizó con un enfoque fenotipo-genotipo utilizando términos HPO para mejorar la curación de variantes; en promedio se utilizaron 9.8 ≈ 10 términos por cada caso, con un mínimo de un término y un máximo de 26.

Tabla 5-5. Descripción de casos diagnósticos en exoma trio

ID	Gen Afectado Seq de Referencia	Variante	Tipo de variante	Enfermedad	OMIM	Herencia	Cigosidad	De novo / herencia	Criterios ACMG	ACMG	Caso diagnóstico en individual
1	CACNAS1 NM_000069	c.3716G>A (p.Arg1239His)	missense	Parálisis periódica hipokalémica tipo 1	170400	AD	Heterocigosis	De novo	PM1 PM2 PM5 PM6 PP3	P	Si
	FIG4 NM_014845	c.122T>C (p.Ile41Thr)	missense	Esclerosis lateral amiotrófica 11	612577	AD	Heterocigosis	Herencia (madre)	PS3 PM2 PP2 PP3	PP	
3	TFAP2B NM_003221	c.218C>G (p.Pro73Arg)	missense	Síndrome Char	169100	AD	Heterocigosis	Herencia (madre)	PS3 PM2 PP2 PP4	PP	Si

ID	Gen Afectado Seq de Referencia	Variante	Tipo de variante	Enfermedad	OMIM	Herencia	Cigosis	De novo / heredad	Criterios ACMG	ACMG	Caso diagnóstico en individual
9	COL4A2 NM_001846	c.2587+1G>A p.?	sitio canónico del splicing	Enfermedad cerebral de vasos pequeños	614483	AD	Heterocigosis	De novo	PVS1 PM6 PM2 PP3	P	Si
10	SPAST NM_014946	c.898A>T (p.Lys300*)	nonsense	Paraplejía espástica AD	182601	AD	Heterocigosis	De novo	PVS1 PM2 PM6 PP3	PP	Si
	DDX3X NM_001356	c.562G>A (p.Gly188Arg)	missense	Trastorno del desarrollo intelectual, ligado al cromosoma X, síndrome Snijders Blok	300958	LXD	Heterocigosis	De novo	PM1 PM2 PM6 PP2 PP3	PP	
12	FLNA NM_001456	c.1331_1338del (p.Cys444Serfs*33)	frameshift	Síndrome Melnick-Needles	309350	LXD	Heterocigosis	Hereditaria (madre)	PVS1 PM2 PP3	P	Si
23	FGFR3 NM_000142	c.1620C>G (p.Asn540Lys)	missense	Síndrome Cruzon, Acondroplasia, Hipocondroplasia, Síndrome Muenke	100800 , 612247 , 602849	AD	Heterocigosis	De novo	PS1 PM1 PM2 PM6 PP2	PP	Si
24	KIF1A NM_004321	c.296C>T (p.Thr99Met)	missense	Síndrome NESCAV, paraplejía espástica AD 30	614255 , 610357	AD	Heterocigosis	De novo	PS3 PM1 PM2 PM6 PP3	P	Si
30	ERBB3 NM_001982	c.2425C>T (p.Gln809*)	nonsense	Susceptibilidad a eritroleucemia familiar	133180	AD	Heterocigosis	Hereditaria (padre)	PVS1 PM2 PP3	PP	Si
31	SELENON NM_020451	c.943G>C (p.Gly315Arg)	missense	Miopatía congénita con desproporción de tipos de fibras	255310	AD	Heterocigosis	Hereditaria (padre)	PS3 PM2 PP3	PP	Si
39	PACS1 NM_018026	c.607C>T (p.Arg203Trp)	missense	Síndrome Schuurs-Hoeijmakers	615909	AD	Heterocigosis	De novo	PS3 PM2 PM6 PP2 PP3	PP	Si
41	SCN4A NM_000334	c.4483A>G (p.Ile1495Val)	missense	Parálisis periódica hiperkalemica	170500	AD	Heterocigosis	Hereditaria (madre)	PM1 PM2 PM5 PP2 PP3	PP	Si
44	ARID1B NM_001374828	c.4856-2A>G p.?	sitio canónico del splicing	Síndrome Coffin-Siris 1	135900	AD	Heterocigosis	De novo	PVS1 PM2 PM6	P	Si
46	NPHP4	c.3044G>A	missense		606966					PP	Si

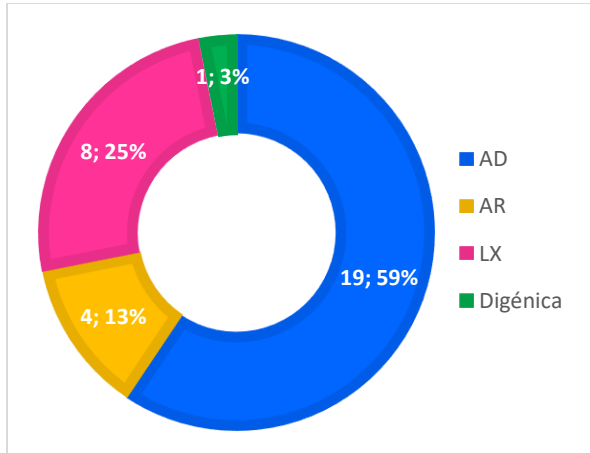




ID	Gen Afectado Seq de Referencia	Variante	Tipo de variante	Enfermedad	OMIM	Herencia	Cigosis	De novo / heredad	Criterios ACMG	ACMG	Caso diagnóstico en individual
77	<i>FAR1</i> NM_032228	c.1439G>A (p.Arg480His)	missense	Cataratas, paraparesia espástica y retraso del habla	619338	AD	Heterocigosis	Heredad (madre)	PS1 PM2 PP3 PP5	PP	Si
79	<i>GJB2</i> NM_004004	c.101T>C (p.Met34Thr)	missense	Síndrome de Vohwinkel o queratodermia	124500	AD	Heterocigosis	Heredad (padre)	PS3 PM2 PP3 PP5	P	Si
84	<i>PACS1</i> NM_018026	c.607C>T (p.Arg203Trp)	missense	Síndrome Schuurs-Hoeijmakers	615009	AD	Heterocigosis	De novo	PS3 PM2 PM6 PP2 PP3	P	Si
85	<i>SIN3A</i> NM_015477	c.816T>G (p.Tyr272*)	nonsense	Síndrome Witteveen-Kolk	613406	AD	Heterocigosis	De novo	PVS1 PM2 PM6 PP3	P	Si
88	<i>MECP2</i> NM_004992	c.502C>T (p.Arg168*)	nonsense	Síndrome Rett	312750	LXD	Heterocigosis	De novo	PVS1 PS3 PM2 PM6 PP3 PP5	P	Si
90	<i>HFE</i> NM_000410	c.845G>A (p.Cys282Tyr)	missense	Porfiria variegata, susceptibilidad	176200	AD	Heterocigosis	Heredad (padre)	PS1 PM1 PP2 PP3	PP	Si
93	<i>B3GAT3</i> NM_012200	c.175C>T (p.Arg59*) c.985T>C (p.Ser329Pro)	nonsense missense	Múltiples dislocaciones articulares, baja estatura, dismorfismo craneofacial, con o sin defectos cardíacos congénitos	245600	AR	Heterocigosis compuesta	Heredad (madre) Heredad (padre)	PVS1 PM2 PP3 PP5 PM2 PM3 PP2 PP3	P PP	No
113	<i>HNRNP2</i> NM_019597	c.616C>T (p.Arg206Trp)	missense	Trastorno del desarrollo intelectual, sindrómico ligado al cromosoma X, tipo Bain	300986	LXD	Heterocigosis	De novo	PS1 PM2 PM5 PM6 PP2 PP3	P	Si
114	<i>MECP2</i> NM_004992	c.502C>T (p.Arg168*)	nonsense	Síndrome Rett	312750	LXD	Heterocigosis	De novo	PVS1 PS3 PM2 PM6 PP3 PP5	P	Si
117	<i>PQBP1</i> NM_005710	c.120C>G (p.Tyr40*)	nonsense	Síndrome Renpenning	309500	LXR	Hemicigosis	Heredad (madre)	PVS1 PM2	PP	Si

De los 32 pacientes con hallazgos genéticos positivos, 19 tenían enfermedades con herencia autosómica dominante, 4 con herencia autosómica recesiva (2 heterocigotos compuestos), 8 con herencia ligada al cromosoma X (solo un caso con herencia recesiva ligada al X), y 1 caso con herencia digénica. Figura 5-1.

Figura 5-1 Mecanismos de herencia de condiciones encontradas en análisis de exomas



Se identificó un total de 37 variantes patogénicas o probablemente patogénicas que explicaban la condición del probando, de las cuales 32 (86.48%) eran puntuales y 5 (13.51%) de tipo InDel; dentro del grupo de variantes puntuales, se identificaron 21 missense, 8 nonsense y 3 en el sitio canónico del splicing; y en el grupo de variantes tipo InDel 4 frameshift, y 1 inframe. En los dos pacientes con variantes en genes incidentales, se identificaron una variante de tipo nonsense y otra frameshift.

Del grupo de pacientes estudiados, los pacientes identificados con los números 88 y 114 tuvieron diagnóstico de síndrome Rett por la misma variante, sin embargo, los pacientes no están relacionados, proceden de regiones geográficas diferentes, y en ambos casos la variante se presentó *de novo*. De igual forma, dos pacientes identificados con los números 39 y 84 se consideraron positivos al tener la misma variante en el gen *PACS1*, *de novo* clasificada como probablemente patogénica.

#### 5.4.1 Identificación de variantes nuevas

De las 39 variantes descritas como patogénicas y probablemente patogénicas en los 32 casos clasificados como diagnósticos o positivos, se identificó un total de 8 variantes

nuevas que no habían sido previamente reportadas en la literatura, 1 comprometió el sitio canónico del splicing, 4 eran nonsense, 3 missense y 2 frameshift. Tabla 5-6.

Tabla 5-6 Variantes nuevas

Variantes nuevas						
ID	Gen Afectado	Variante región codificante	Variante en proteína	Tipo de variante	Clasificación ACMG	Secuencia de referencia
9	<i>COL4A2</i>	c.2587+1G>A	p.?	sitio canónico del splicing	P	NM_001846
10	<i>SPAST</i>	c.898A>T	p.Lys300*	nonsense	P	NM_014946
10	<i>DDX3X</i>	c.562G>A	p.Gly188Arg	missense	PP	NM_001356
12	<i>FLNA</i>	c.1331_1338del	p.Cys444Serfs*33	frameshift	P	NM_001456
30	<i>ERBB3</i>	c.2425C>T	p.Gln809*	nonsense	PP	NM_001982
46	<i>NPHP3</i>	c.2571_2574del	p.Ser857Argfs*4	frameshift	PP	NM_153240
85	<i>SIN3A</i>	c.816T>G	p.Tyr272*	nonsense	P	NM_015477
117	<i>PQBP1</i>	c.120C>G	p.Tyr40*	nonsense	PP	NM_005710

Estas variantes no se encontraron descritas en las bases poblacionales 1000 Genomas, ni GnomAD, tampoco contaban con *rs* (reference SNP ID) en la base de datos dbSNP, no estaban descritas en UniProt. Además, tampoco estaban descritas en bases de datos clínicas públicamente disponibles como ClinVar, LOVD, HGMD y no se encontraron casos descritos en publicaciones científicas en buscadores de Pubmed y Google Scholar, motivo por el cual adquirirían inicialmente un criterio PM2\_soporte(78).

A las 8 variantes nulas en los genes *COL4A2*, *SPAST*, *FLNA*, *ERBB3*, *NPHP3*, *SIN3A*, *PQBP1* se les asignó el criterio PVS1 aplicando el árbol de decisión de ClinGen que sumado al criterio PM2 por su rareza permitía clasificarlas como probablemente patogénicas(76,78).

A la variante en *DDX3X*, se le asignó el criterio PM6 por ser de novo, PM2\_soporte por su rareza, también se le dio el criterio PM1 por encontrarse localizada en un dominio de unión de helicasa de ATP y estar en un hot spot con múltiples variantes patogénicas descritas cercanas y además se le asignó el criterio PP3 pues en diferentes bases de datos se confirmó su efecto deletéreo y encontrarse altamente conservada.

A continuación, se presenta la descripción de las variantes:

**COL4A2 c.2587+1G>A p.? en heterocigosis *de novo***

La variante es un cambio en el sitio canónico del splicing *de novo*, que se encuentra localizada en el intrón 30/47, en un gen que no es haploinsuficiente pLI: 2.11e-06, el salto de exón o el uso del sitio crítico del empalme conserva el marco de lectura, no se documentó como un hot spot mutacional ni se encuentra en un dominio funcional. Las variantes de pérdida de función (LoF) en este exón son poco frecuentes en la población general y no se encontró en bases de datos 1000 Genomas, GnomAD, el exón está presente en transcritos biológicamente relevante(s) y la variante elimina <10 % de la proteína(106–108), No se encuentra descrita en dbSNP, ni las bases de datos clínicas ClinGen, LOVD, HGMD, ni en la bibliografía científica consultada a la fecha, PhyloP 100 y GERP prevee que es una variante conservada. Lo anterior permitió clasificarla como patogénica. Este gen codifica para la cadena 2 del colágeno tipo IV, y es componente estructural de las membranas basales, su región C-terminal o canstatina es un inhibidor de la angiogénesis y el crecimiento tumoral(109). Variantes en este gen se han asociado con enfermedad cerebral de vasos pequeños (MIM 614483) caracterizada por un suministro vascular alterado al cerebro que conduce a degeneración cerebral, hemiplejía, convulsiones y discapacidad intelectual con gravedad variable(110). Esta variante se observó en un paciente de un año con encefalopatía epiléptica, retraso en el neurodesarrollo, microcefalia, leucomalacia crónica periventricular bilateral, hipotonía, trastorno de la deglución, no conexión con el medio, irritabilidad, hiporreflexia y cuadriparesia.

**SPAST c.898A>T p.Lys300\* en heterocigosis *de novo***

Esta variante heterocigota *de novo* de tipo nonsense se encuentra localizada en el exón 6/17 del gen *SPAST*, con evidencia de haploinsuficiencia pLI: 0,997, esta variante predice un codón de parada prematuro que causa degradación del ARN mensajero mediada por mutaciones terminadoras, NMD (Nonsense Mediated Decay) y el exón está presente en un transcrito biológicamente relevante. Al ser nueva no se encontró descrita en la literatura, ni en bases de datos poblacionales 1000 Genomas, GnomAD, además PhyloP 100 y GERP prevé que es una variante conservada. Se clasificó como probablemente patogénica(106–108). Se conoce que el gen *SPAST* codifica para una de la familia de las ATPasas

asociadas a procesos celulares, participa en tráfico celular, motilidad intracelular y biogénesis de organelos y se ha asociado con paraplejía espástica autosómica dominante(109–111). El fenotipo del paciente es descrito más adelante.

#### ***DDX3X c.562G>A p.Gly188Arg en heterocigosis de novo***

Se trata de una variante missense en la cual se reemplaza un glicina por una arginina en la posición 188, de novo, SIFT, Mutation Taster y Polyphen2(106,112,113) predicen su efecto deletéreo en la proteína, además es una variante altamente conservada de acuerdo con PhyloP 100 y GERP, no se encuentra documentada en dbSNP, ni las bases de datos clínicas ClinGen, LOVD, HGMD, es una variante rara no encontrada en bases poblacionales(106,108), se localizada en un dominio de unión de helicasa de ATP de acuerdo con UniProt y esta en un hot spot mutacional con múltiples variantes patogénicas descritas cercanas. Este gen codifica para una helicasa de ARN dependiente de ATP, la cual participa en múltiples procesos celulares a nivel nuclear y citoplasmático, en ensamblaje de pre-RNA y su exportación fuera del núcleo(109,110). Variantes patogénicas en este gen se han asociado con trastorno del desarrollo intelectual(110,111), como ocurre en el paciente de este estudio, una mujer con autismo, microcefalia, agenesia de cuerpo caloso, déficit cognitivo, atrofia de iris bilateral y estereotipias.

#### ***FLNA c.1331\_1338del p.Cys444Serfs\*33***

Es una variante de corrimiento del marco de lectura, se encuentra localizada en el exón 9/47, en un gen que tiene una evidencia suficiente de haploinsuficiencia de acuerdo con la curación manual de ClinGen tiene un HI score de 3 sugestiva de ser una variante patogénica(108,114). Predice un codón de parada prematuro para la degradación del ARN mensajero mediada por mutaciones terminadoras, NMD (Nonsense Mediated Decay), el exón está presente en transcritos biológicamente relevantes(106), se encuentra ausente en bases de datos poblacionales y no se encontró descrita en la literatura. Se encontró en un paciente con hipertensión pulmonar primaria severa, dependiente de oxígeno a alto flujo, comunicación interauricular tipo ostium secundum, luxación congénita del radio bilateral, apnea del sueño central, migraña con aura y fenotipo con micrognatia, epicanto, cubitus valgo, sandal gap, e hiperlaxitud. El gen *FLNA* codifica para la filamina A, una proteína de unión a actina que participa en la remodelación del citoesqueleto para generar cambios de forma y migración celular(107,109,110). El cuadro clínico coincide con el Síndrome Melnick-Needles asociado con este gen(109,110).

### ***ERBB3* c.2425C>T p.Gln809\* en heterocigosis**

Es una variante de tipo nonsense que se encuentra localizada en el exón 20/28 del gen *ERBB3*, en un gen que no es de haploinsuficiencia pLI: 4.36e-06, predice un codón de parada prematuro para la degradación del ARN mensajero mediada por mutaciones terminadoras, NMD (Nonsense Mediated Decay), el exón está presente en un transcrito biológicamente relevante(106–108), no está en dbSNP, ni en bases de datos clínicas ni poblacionales. Este gen codifica un miembro de la familia tirosin cinasa del receptor del factor de crecimiento epidérmico (EGFR) y se asocia con susceptibilidad a eritroleucemia familiar(109–111), y fue documentada en un paciente con síndrome mieloproliferativo, leucopenia, neutropenia, trombocitopenia, hiperplasia mieloide, hiperplasia megacariocítica, mielofibrosis leve, cardiopatía, estenosis pulmonar, ductus arterioso persistente, trastorno del comportamiento, frente amplia, epicanto, telecanto, narinas hipoplásicas, orejas de implantación baja, teletelia e infecciones frecuentes.

### ***NPHP3* c.2571\_2574del p.Ser857Argfs\*4**

Variante de corrimiento en el marco de lectura en el exón 19/27 del gen *NPHP3*, en un gen no haploinsuficiente 1.41e-09, en el cual se predice un codón de parada prematuro para la degradación del ARN mensajero mediada por mutaciones terminadoras, NMD (Nonsense Mediated Decay), además el exón está presente en un transcrito biológicamente relevante(106,107), en una región biológicamente conservada, no fue encontrada en 1000 Genomas ni en GnomAD, tampoco en dbSNP ni en bases de datos clínicas ni en la literatura(108). Este gen codifica para una proteína con dominio de tubulina-tirosina ligasa que interactúa con la nefrocistina, la cual es necesaria para el desarrollo ciliar normal y participa en el desarrollo tubular renal(109,110). Las variantes en este gen se asocian con displasia renal-hepática y pancreática y con nefronoptosis tipo 3 autosómica recesiva (109,110) y se ha documentado en herencia digénica con otros genes *NPHP*(115), y fue documentada en un paciente con Nefronoptosis, enfermedad renal crónica secundaria, fibrosis hepática congénita, HTA, trasplante renal e hiperparatiroidismo.

### ***SIN3A* c.816T>G p.Tyr272\* heterocigota *de novo***

Es una variante de tipo nonsense encontrada *de novo* en el gen *SIN3A* con un score de haploinsuficiencia pli=1 que indica haploinsuficiencia, en el cual se predice un codón de

parada prematuro para la degradación del ARN mensajero mediada por mutaciones terminadoras, se encuentra ubicada en el exón 6/21 en un transcrito biológicamente relevante(106–108). Se desconoce su frecuencia pues no ha sido descrita en bases de datos poblacionales y tampoco fue encontrada en bases de datos clínicas ClinGen, LOVD, HGMD, ni en la bibliografía científica consultada a la fecha. Se predice que se localiza en una región conservada, y codifica para una proteína reguladora de la transcripción que se ha asociado con Síndrome de Witteveen-Kolk de herencia autosómica dominante(109–111). El caso presentado en este estudio se trataba de un paciente con Antecedente de hidrotórax congénito, bajo peso al nacer, síndrome broncoobstructivo recurrente, microcefalia, hipotelorismo, retrognatia, hiperlaxitud, retraso del desarrollo motor e hipoacusia neurosensorial bilateral.

#### ***PQBP1 c.120C>G p.Tyr40\****

Es una variante de tipo nonsense que genera un codón de parada en la posición 40 de la proteína, se encuentra localizada en el exón 2/6, en un gen que tiene una evidencia suficiente de haploinsuficiencia de acuerdo con la curación manual de ClinGen tiene un HI score de 3(106,116), en el cual se predice un codón de parada prematuro para la degradación del ARN mensajero mediada por mutaciones terminadoras (NMD), y se considera que este exón se localiza en un transcrito biológicamente relevante(106–108). No se encontró en bases de datos 1000 Genomas, GnomAD, no está descrita en bases de datos clínicas ni se encontró en la literatura científica. Esta variante fue documentada en Hemicigosis en un paciente con cuadro clínico de holoprosencefalia, hipoplasia cerebral, agenesia de cuerpo calloso, criptorquidia bilateral, fístula perineal, malformación anorrectal, peso y talla bajos para la edad, cráneo asimétrico, plagiocefalia, pabellones auriculares displásicos, fisuras palpebrales oblicuas hacia abajo camptodactilia de segundo y tercer dedo, pliegue palmar único bilateral, uñas hiperconvexas y estudio de microarreglos normal. PQBP1 codifica para una proteína de unión a poliglutamina nuclear que está implicada en la activación de la transcripción y se ha descrito en pacientes con Síndrome Renpenning 1(109–111), fenotipo que coincide con el paciente descrito.

### **5.4.2 Hallazgos Incidentales**

Teniendo en cuenta que se trataba de estudios de exoma completo, se exploraron hallazgos incidentales o secundarios, definidos por la ACMG como aquellos genes

asociados a condiciones genéticas específicas que cuentan con intervenciones establecidas destinadas a prevenir o reducir significativamente la morbilidad y la mortalidad de los pacientes portadores de variantes en estos genes. El listado de genes más reciente publicado por la ACMG V3.1 de 2022 contiene noventa genes de interés clínico para reportar variantes patogénicas conocidas(104). En este estudio, se encontraron hallazgos incidentales en 4 individuos (2 padres y 2 probandos), los cuales corresponden a un caso con una variante patogénica en el gen *MSH6* encontrada tanto en el probando como en el padre (caso 32), y un segundo paciente con una variante probablemente patogénica en el gen *PMS2*, identificada tanto en el probando como en el padre (caso 3), los dos genes asociados con síndrome Lynch. Tabla 5-6.

Tabla 5-7. Casos hallazgos incidentales.

Descripción casos con hallazgos incidentales											
ID	Gen Afectado	Variante	Tipo de variante	Fenotipo	OMIM	Herencia	Cigosidad	De novo / heredada	Criterios ACMG	ACMG	Caso diagnóstico
3	<i>PMS2</i> NM_000535	c.400C>T (p.Arg134*)	nonsense	Síndrome Lynch	614337	AD	Heterocigosis	Heredada (padre)	PVS1 PM2 PP3	P	Hallazgo incidental
32	<i>MSH6</i> NM_000179	c.742delC (p.Arg248G lufs*31)	frameshift	Síndrome Lynch	614337	AD	Heterocigosis	Heredada (padre)	PVS1 PM2 PP3	P	Hallazgo incidental

Es de anotar que, en el caso identificado con el número 1 se documentó una variante patogénica en el gen *CACNAS1*, que se encuentra en el listado de los genes de hallazgos incidentales, sin embargo, esta variante coincidía con el fenotipo del paciente y por tanto no fue considerado como un hallazgo incidental, sino que se asignó como diagnóstico.

Un hallazgo llamativo fue encontrar que 2 casos tenían hallazgos moleculares compatibles con dos condiciones monogénicas concomitantes, que podrían asociarse al fenotipo descrito, es decir que posiblemente se tratan de dobles heterocigotos, sin embargo, esta asignación solo puede ser realizada por el médico tratante de los pacientes.

El primero de ellos corresponde al caso 1, un individuo masculino de 22 años con historia de parálisis episódicas desde los dos años de edad, episodios agudos que inician en miembro superior derecho y progresan al resto del cuerpo, con múltiples crisis a lo largo de la vida, pérdida progresiva de la fuerza y caídas frecuentes, al examen físico con ptosis



palpebral, hipoplasia mediofacial, facies hipomímicas, hipotrofia generalizada, hipotrofia tenar e hipotenar, reflejos osteotendinosos ++/++++ en las cuatro extremidades, marcha con balanceo, incapacidad para realizar marcha en puntas y talones, no hay fenómeno miotónico, hipercekaemia, EMG patrón miopático, biopsia sugestiva de distrofia miotónica considerado como hallazgos clínicos y de laboratorio no concluyentes. Al realizar el estudio de exoma se identificó una variante patogénica *de novo* en el gen *CACNAS1* en estado heterocigoto, asociado con parálisis periódica hipokalémica tipo 1 (MIM 170400) autosómica dominante que podría explicar el cuadro clínico del paciente. Sin embargo, también se documentó una variante probablemente patogénica en el gen *FIG4* en heterocigosis, heredada de la madre, que se ha asociado con esclerosis lateral amiotrófica autosómica dominante (MIM 612577), que puede solaparse con algunos de los signos y síntomas del paciente, pero también es un gen con penetrancia incompleta y expresividad variable que podría estar asociado o no al fenotipo del paciente. Además, la edad promedio de inicio de las manifestaciones clínicas de la enfermedad, según registros de OMIM es a los 56 años.

El segundo posible doble heterocigoto, corresponde al caso 10, una paciente femenina de 11 años con antecedente de retraso global del neurodesarrollo, cuadro clínico de trastorno del espectro autista, microcefalia, agenesia de cuerpo calloso, déficit cognitivo severo, estereotipias, no fijación de la mirada, atrofia del iris bilateral, baja talla y asimetría de extremidades inferiores. En estudio de exoma trio se identificó una variante patogénica en heterocigosis *de novo* en el gen *DDX3X*, el cual se ha asociado con Trastorno del desarrollo intelectual, ligado al cromosoma X, síndrome Snijders Blok (MIM 300958) que se presenta predominantemente en mujeres y cursa con discapacidad intelectual variable, alteraciones del movimiento y/o espasticidad, y anomalías cerebrales como microcefalia, ventriculomegalia, o malformaciones corticales, es de instauración en la infancia y suele presentarse por variantes *de novo*.

Por otro lado, en el mismo caso 10 se documentó una segunda variante patogénica *de novo* en heterocigosis en el gen *SPAST*, relacionado con paraplejía espástica autosómica dominante, que se caracteriza por debilidad y espasticidad progresiva de miembros inferiores, su inicio puede ser en la infancia o en la adultez, además puede cursar con déficit cognitivo, hiperreflexia y marcha espástica. Si bien el fenotipo parece estar predominantemente expresado por las manifestaciones del síndrome Snijders Blok, es

posible que parte de sus características fenotípicas también estén solapadas con la variante en el gen *SPAST*.

El caso 3, en quien se registró una variante probablemente patogénica heredada de la madre en el gen *TFAP2B* asociado con síndrome Char, que describe el fenotipo del paciente, se detectó también una variante en el gen *PMS2* heredada del padre, que se asocia a síndrome Lynch de penetrancia reducida.

Se observó que, de los 100 casos analizados, el 90% tuvieron al menos una variante de significado incierto identificada, que, a pesar de no ser diagnóstica, podría explicar el fenotipo de los pacientes, y deberán ser reclasificadas en el futuro. En contraparte, 10 casos fueron reportados como negativos donde no se identificó ninguna variante de tipo SNP, ni CNV, ni variantes en el genoma mitocondrial, que pudieran explicar total o parcialmente el fenotipo.

## **5.5 Rendimiento diagnóstico por características sociodemográficas y fenotípicas**

El rendimiento diagnóstico del estudio de exoma fue superior en mujeres que en hombres, con un rendimiento diagnóstico del 37.2% (19/51) en mujeres y de 26.5% (13/49) en hombres, de 7 casos con antecedente de consanguinidad confirmada 4 fueron diagnosticados tanto en el análisis en trio como en el individual, es decir se hizo diagnóstico en el 57.1% de los casos con consanguinidad conocida. Sin embargo, dos casos si estaban asociados con condiciones recesivas con alelos homocigotos y dos casos tenían condiciones dominantes en heterocigosis.

El antecedente de historia familiar conocida, donde se tuvo en cuenta haber reportado en la historia clínica un familiar en primer o segundo grado con cuadro clínico similar al del probando, fue identificado en 18 casos, de los cuales el 27.7% fueron diagnósticos en el análisis de exoma, mientras que, en el grupo de 82 pacientes sin antecedente familiar conocido, se hizo diagnóstico en una mayor proporción de pacientes, 32.9%.

Los casos con estudios genéticos previos tuvieron diagnóstico en el 32.3% el cual fue ligeramente superior a no tener ningún estudio genético previo 31.8%. Respecto a la

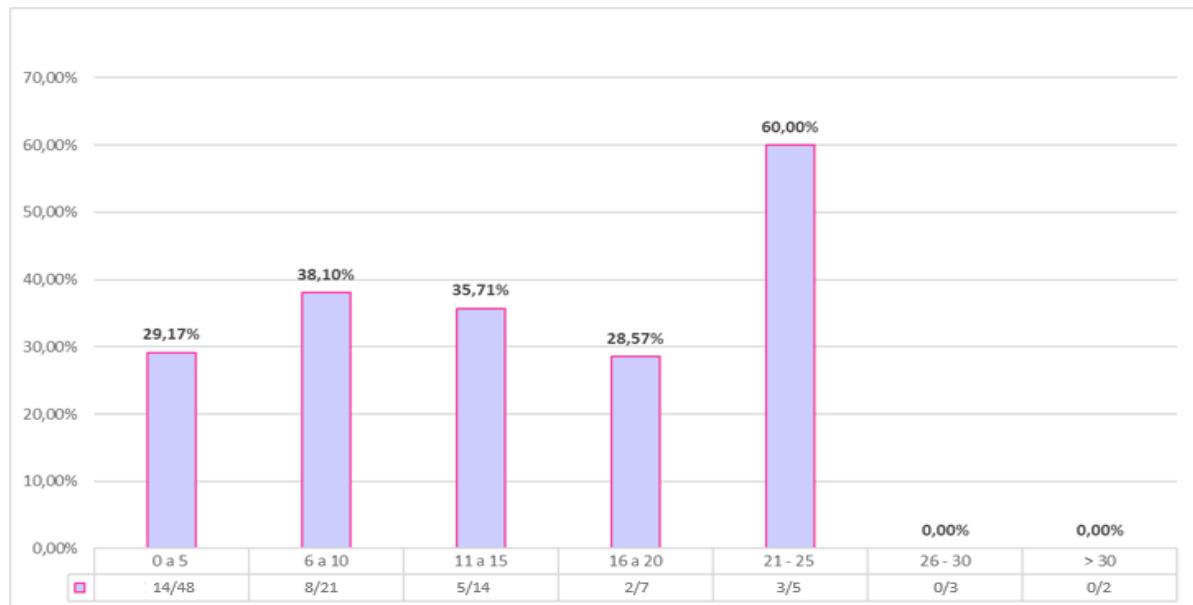
especialidad que solicita el estudio de exoma, se encontró que Genética hizo el mayor número de ordenamientos y la positividad fue del 31.2%, las especialidades de hemato-oncología, inmunología y neuropediatría hicieron cuatro ordenamientos de los cuales el 50% fueron diagnósticos (2/4). (Tabla 5-9).

Tabla 5-9. Rendimiento diagnóstico de acuerdo con las características sociodemográficas y clínicas

<b>Sexo</b>		
	Individuos n (%)	Diagnosticados n (%)
Femenino	51 (51.0)	19 (37.2%)
Masculino	49 (49.0)	13 (26.5%)
Total	100 (100%)	32(32.0%)
<b>Consanguinidad</b>		
No	93 (93.0)	28 (30.1%)
Si	7(7.0)	4 (57.1%)
<b>Historia familiar</b>		
No	82 (82.0)	27 (32.9%)
Si	18 (18.0)	5(27.7%)
<b>Estudios genéticos previos</b>		
No	66 (66.0)	21 (31.8%)
Si	34 (34.0)	11 (32.3%)
<b>Especialidad que solicita el estudio</b>		
Genética	96 (96.0)	30 (31.2%)
Hemato-oncología	1 (1.0)	1 (100%)
Inmunología	1 (1.0)	1 (100%)
Neuropediatría	2 (2.0)	0 (0.0%)

Se agruparon los pacientes por quinquenios encontrando que el grupo de edad más representativo en la muestra estudiada fue el grupo de 0 a 5 años, en quienes se obtuvo un rendimiento diagnóstico de 29.17%. Sin embargo, el mayor rendimiento diagnóstico se obtuvo en el grupo de 21 a 25 años con un 60% de los casos. Figura 5-2. El rendimiento diagnóstico de la población de 10 años o menos fue del 31.88%.

Figura. 5-2. Rendimiento de edad por quinquenios

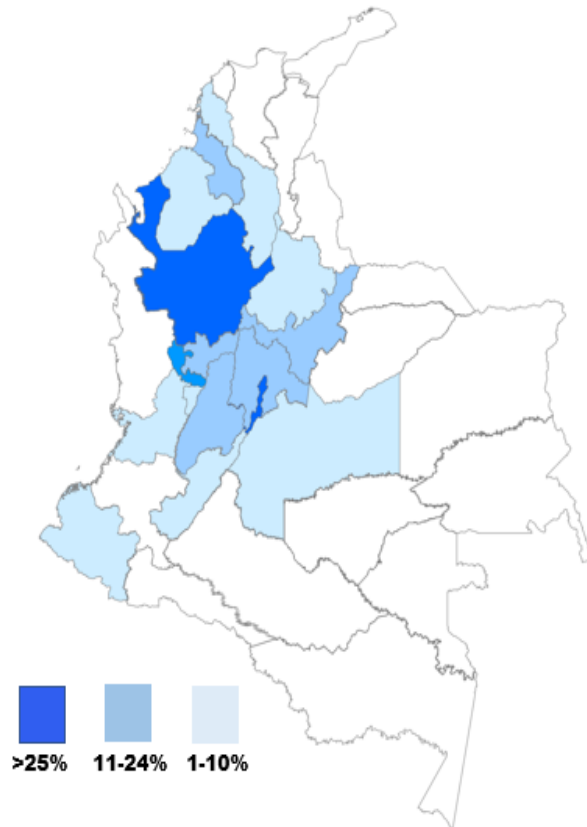


Como se observa en el mapa de calor a continuación, la muestra estudiada estaba distribuida en 16 departamentos de Colombia principalmente en la región andina, parte de la región caribe y el sur de la región pacífica, donde Bogotá, fue la región con 10 casos y un 31.25% dentro del grupo de diagnosticados. El segundo lugar fue ocupado por Antioquia con una proporción del 28.12% de los diagnosticados.

Boyacá tuvo 3 de 3 casos diagnosticados, que al considerar que se trata de una región endogámica podría estar representado por las condiciones autosómicas recesivas, sin embargo, los tres pacientes diagnosticados tienen condiciones ligadas al cromosoma X. El siguiente departamento con mayor rendimiento diagnóstico fue Tolima con 6.25%. Tabla 5-10.

Tabla 5-10. Rendimiento diagnóstico por departamentos de Colombia.

Departamento de Colombia		
	Individuos n	Diagnosticados n (%)
Bogotá	41	10(31,25%)
Antioquia	30	9(28,12%)
Risaralda	5	1(3,12%)
Boyacá	3	3(9,37%)
Sucre	3	1(3,12%)
Caldas	3	1(3,12%)
Cundinamarca	3	1(3,12%)
Tolima	3	2(6,25%)
Córdoba	2	1(3,12%)
Valle del Cauca	1	0(0%)
Quindío	1	0(0%)
Bolívar	1	0(0%)
Santander	1	0(0%)
Huila	1	1(3,12%)
Meta	1	1(3,12%)
Nariño	1	1(3,12%)

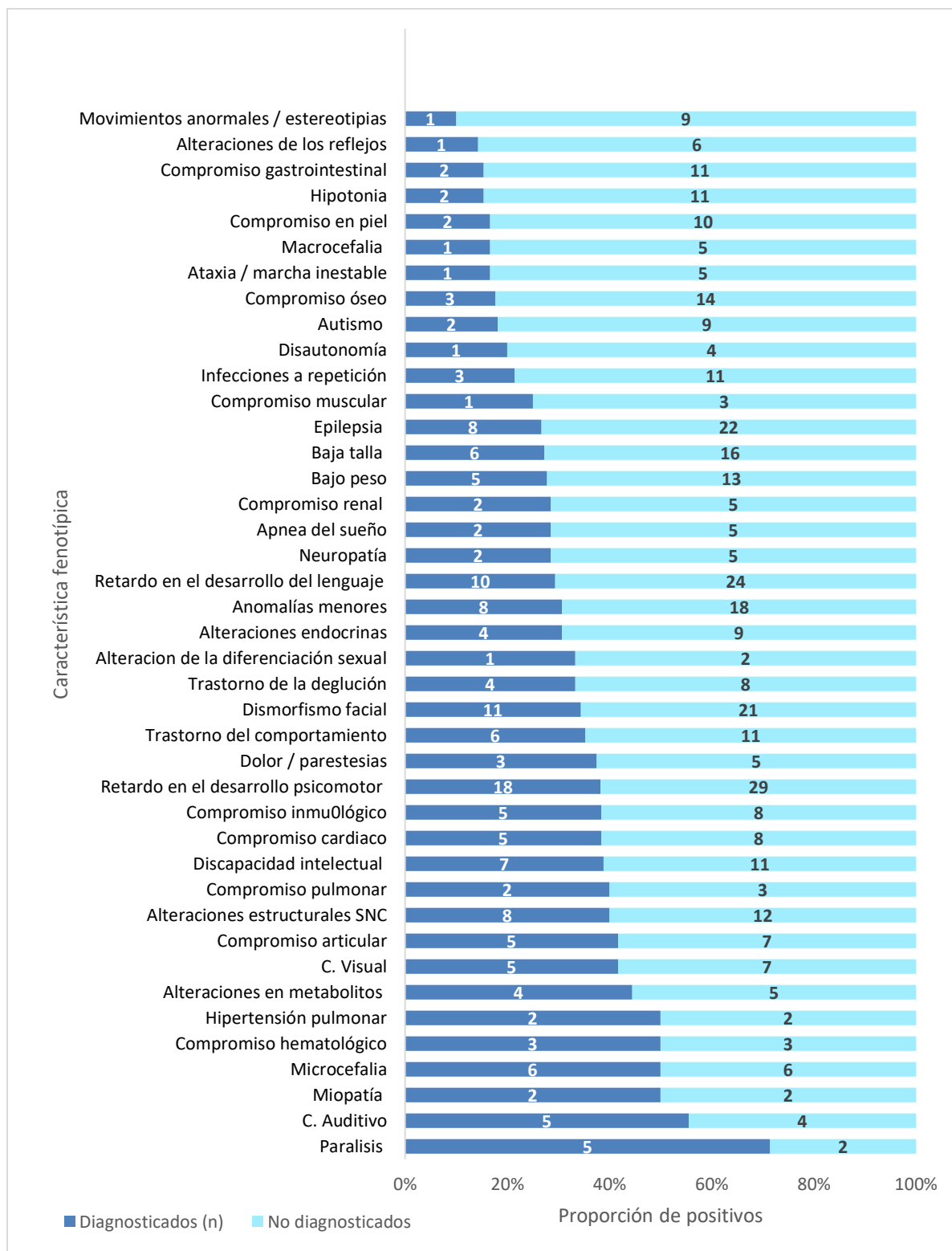


Al revisar las 41 características fenotípicas extraídas de las historias clínicas de los pacientes incluidos en el estudio, se encontró que la parálisis fue el rasgo fenotípico con un mayor rendimiento diagnóstico (71.43%), donde 5 de los 7 individuos con esta manifestación clínicas fueron diagnósticos en el exoma trio.

El 55.56% de los individuos con compromiso auditivo fueron diagnosticados a través del análisis de exoma ocupando el segundo lugar como característica fenotípica con mayor rendimiento diagnóstico.

El antecedente de miopatía, microcefalia, compromiso hematológico e hipertensión pulmonar tuvieron un rendimiento diagnóstico del 50%. Y otras características como alteraciones en los metabolitos, compromiso visual, articular, pulmonar y alteraciones en estructurales en el sistema nervioso central tuvieron un rendimiento diagnóstico superior al 40%. Gráfica 5-2.

Figura 5-3. Rendimiento fenotípico por característica fenotípica.



Teniendo en cuenta que el retraso en el desarrollo psicomotor era la característica fenotípica más frecuente en 47 pacientes, donde 18 casos fueron diagnósticos, se crearon 6 subgrupos fenotípicos: Retraso psicomotor + microcefalia/malformaciones mayores, Retraso psicomotor + alteración estructural del SNC, Retraso psicomotor + Epilepsia, Retraso psicomotor + dismorfismo facial, Retraso psicomotor + retraso del lenguaje, R. psicomotor + Trastorno del comportamiento/autismo. Tabla 5-11.

Tabla 5-11 Rendimiento diagnóstico en los subgrupos fenotípicos de retraso psicomotor.

Subgrupos	Individuos (n)	Diagnóstico (n)	Rendimiento diagnóstico en grupo de 47 individuos con R. psicomotor	Rendimiento diagnóstico en grupo de 18 individuos diagnósticos con R. psicomotor
R. psicomotor + parálisis	4	4	100.00%	22.22%
R. psicomotor + microcefalia/malformaciones mayores	11	7	63.64%	38.89%
R. psicomotor + alteración estructural del SNC	15	8	53.33%	44.44%
R. psicomotor + Epilepsia	18	7	38.89%	38.89%
R. psicomotor + dismorfismo facial	21	8	38.10%	44.44%
R. psicomotor + retraso del lenguaje	29	11	37.93%	61.11%
R. psicomotor + T. del comportamiento/autismo	15	5	33.33%	27.78%

Finalmente, al evaluar el rendimiento diagnóstico en los 6 subgrupos fenotípicos, teniendo en cuenta el total de individuos con retraso psicomotor (n=47), se identificó que el grupo de pacientes con retraso psicomotor y microcefalia o malformaciones mayores tuvo el mayor rendimiento diagnóstico con 63.64%. De otro lado el retraso psicomotor sumado con alteración estructural del sistema nervioso central tuvo un rendimiento del 53.33%; y el tercer lugar lo ocupó el grupo de retraso psicomotor con epilepsia que tuvo un rendimiento diagnóstico de 38.89%.





## 6 Discusión

El presente trabajo de investigación es el primer estudio en Colombia que determina el rendimiento diagnóstico de la secuenciación de exoma en la práctica clínica en población colombiana, y compara el análisis de exomas en trio, que duplica el costo y triplican el almacenamiento de datos, respecto al exoma individual en su capacidad diagnóstica. Además, presenta el rendimiento diagnóstico por grupos fenotípicos y según la principal sintomatología de los pacientes estudiados que fue retraso psicomotor.

Los estudios de secuenciación son hoy en día una de las principales herramientas utilizadas por los clínicos en Colombia para el diagnóstico de pacientes con sospecha de enfermedades de origen genético, debido a la evidencia en su utilidad diagnóstica, su disponibilidad para realizarlo en el país, y también por la ventaja que supone que las pruebas de secuenciación se encuentren incluidas en el plan de beneficios en salud específicamente en la resolución 5857 del 26 de diciembre del 2018, donde el Ministerio de Salud y Protección Social aprueba que las personas afiliadas al sistema pueden acceder a estudios moleculares solo con la orden médica(117) .

La elección de un estudio de exoma depende principalmente de la complejidad fenotípica de los pacientes, y se prefiere su uso cuando existe un fenotipo variado, heterogéneo, con compromiso de múltiples sistemas y se considera una entidad que supone una amplia heterogeneidad alélica; en otros casos se prefiere usarlo cuando estudios moleculares de menor complejidad no han arrojado un diagnóstico.

En este estudio, el rendimiento diagnóstico para el exoma individual fue del 29% y tuvo un rendimiento adicional del 3% al realizar el análisis en trio, particularmente por su capacidad para establecer la segregación familiar de dos casos de enfermedades recesivas con variantes en heterocigosis compuesta y por su capacidad para clasificar variantes de novo

donde es posible revisar la paternidad, y también por la mejoría que genera en la asignación de patogenicidad.

Algunos estudios han mostrado la superioridad del exoma trio respecto al exoma único, *Kingsmore y colaboradores* realizaron un estudio de genoma o exoma rápido y ultrarrápido en 213 niños críticamente enfermos con condiciones de etiología desconocida, encontrando que el rendimiento diagnóstico realizando el análisis solo en el probando fue de 24% (52/213), y el análisis en trio para 147 de estos casos solo mostró un aumento del 0.7% (1/147) en el rendimiento diagnóstico para ese grupo (101)

*Alfares y colaboradores* documentaron un rendimiento diagnóstico para los análisis individuales del 39% (57/145) y del 40% (85/214) para aquellos en trio, en un estudio retrospectivo que incluyó pacientes con sospecha de condiciones monogénicas que se sometieron a WES o WGS (Secuenciación de genoma completo) en Arabia Saudita (118), sin embargo, al separar solo la cohorte de pacientes con WES del grupo con WGS individual versus trio, se encontró que el rendimiento fue superior en los casos individuales 47% (52/110) que en los tríos 46% (79/172). Lo anterior les permitió concluir que no hubo diferencia en la tasa de aciertos y una posible explicación puede ser que la mayoría de las enfermedades documentadas eran autosómicas recesivas y las variantes se encontraban en homocigosis(118).

A diferencia de este proyecto de tesis, en los estudios previamente mencionados no se realiza exoma individual y exoma trio a todos los individuos, sino que por el contrario, se selecciona un grupo de pacientes diferentes para realizar exoma trio y otro grupo para exoma individual, principalmente por la disponibilidad de los padres para realizar el examen, lo cual desde el inicio introduce un sesgo de selección y dificulta establecer con claridad la superioridad del exoma trio sobre el exoma individual.

Solo se encontró un estudio similar en la literatura, realizado por *Yang y colaboradores* donde también buscan controlar ese sesgo de selección y tomaron una muestra de 30 pacientes con sospecha de enfermedad de origen genético previamente valorados y seleccionados por un genetista clínico en Melbourne Australia, sin estudios genéticos previos y con menos de dos años de estudio en su recorrido diagnóstico, a quienes les realizaron tanto análisis individual como en trio en paralelo, encontrando que 11/30 (36.7%) fueron diagnosticados en los dos tipos de análisis y un caso más aumento el rendimiento

en el rendimiento al realizarlo en trio 12/30(40%), donde el caso diagnóstico adicional fue una variante de novo en el gen HNRNPU logrando un aumento en el rendimiento del 3.3% (119). Estos hallazgos son similares en porcentaje a lo documentado en el presente estudio y también muestran la ventaja en rendimiento del exoma individual vs trio.

El aumento en el rendimiento del 3%, podría estar dado por algunas ventajas del exoma trio como son, la reducción en el tiempo de análisis, la facilidad para identificar variantes de novo y asignarles patogenicidad, la posibilidad de identificar heterocigotos compuestos, la reducción cantidad de variantes preseleccionadas para ser filtradas y analizadas y la posibilidad de hacer diagnóstico de hallazgos incidentales en más de un miembro de la familia. Algunos autores han documentado que estas ventajas generan un menor costo de análisis si se considera principalmente la reducción en tiempo y la disminución de estudios confirmatorios (119).

La diferencia tan pequeña en el rendimiento diagnóstico del exoma individual y en trio, podría sugerir que, sería preferible elegir en primera instancia la realización de un exoma individual al momento de enfrentarse a un paciente con sospecha de enfermedad monogénica, no obstante, es necesario realizar estudios futuros de costoefectividad para definir si es recomendable iniciar por un exoma individual y posteriormente ascender a un trio o preponderar el análisis en trio por las ventajas que supone (119).

En este estudio participaron 7 pacientes hijos de parejas consanguíneas y se encontró que 4/7 (57.1%) fueron diagnosticados, no obstante, este porcentaje se mantuvo tanto para el análisis individual como en trio, y de los 4 casos diagnósticos, 2 resultaron ser homocigotos, otro caso fue *de novo* y un caso heredó una variante heterocigota de su padre, lo cual sugiere que el exoma trio no supone superioridad para estudiar hijos de parejas consanguíneas al igual que lo documentó *Alfares y colaboradores* (118)

En contraparte a los hallazgos de este proyecto donde no parece haber una ventaja del exoma trio en parejas consanguíneas, *Karul y colaboradores*, documentaron una ventaja en la capacidad diagnóstica del exoma trio en 190 familias consanguíneas de Turquía con enfermedades neurogenéticas, con una efectividad diagnóstica en el 72% de las familias 137/190, de las cuales 141/172 casos se encontraban en homocigosis (120). Esto podría suponer una superioridad de los exomas trio en poblaciones altamente consanguíneas

como Turquía que reporta una prevalencia de parejas consanguíneas del 24%, pero no en Colombia que tiene una prevalencia de consanguinidad registrada de 1.31% (121).

De otro lado, este estudio tuvo un rendimiento diagnóstico general del 32%, lo cual se aproxima y es ligeramente superior a rendimientos diagnósticos reportados en otros estudios como *Retterer y colaboradores* quienes reportan un rendimiento general del 28.8% para el WES individual en un grupo de 3040 probandos, y 31% en los casos que incluían el análisis en tres miembros de la familia (122). De igual forma, como se observó en el estudio de *Kingsmore y colaboradores* quienes encontraron un rendimiento total del exoma de 24.7% (101), *Yang y colaboradores* reportaron un diagnóstico molecular de 25.2% (504/2000) en pacientes con sospecha de enfermedad mendeliana(92). El consorcio FORGE de Canadá reportó por *Sawyer y colaboradores* un rendimiento diagnóstico general de los exomas del 29% (105 de 362) para pacientes con sospecha de enfermedades genéticas(93).

En contraparte, otros estudios reportan rendimientos diagnósticos mayores a los vistos en esta investigación, *Kim y colaboradores*, en un estudio de 108 niños con retraso en el neurodesarrollo en Korea del Sur, documentaron un rendimiento diagnóstico en el estudio de exoma clínico del 38% (41/108) y un aumento del 3.7% al realizar el análisis en trio para aquellos que no habían sido diagnosticados (45/108), para un diagnóstico total de 41.7% (123). *Stark y colaboradores* en 2016, documentan un rendimiento diagnóstico del 57.5% (46/80) en WES para estudiar niños con sospecha de enfermedades monogénicas(124). *Seo y colaboradores en 2020* documentaron un rendimiento del 42.7% (141/330) en WES en pacientes de familias no consanguíneas con sospecha de condiciones genéticas (125).

Eventualmente, la razón principal para obtener mayores rendimientos diagnósticos respecto a esta investigación fue la selección más estricta de los pacientes, donde la evaluación clínica detallada por un genetista, la edad pediátrica y la selección de fenotipos específicos como múltiples anomalías congénitas y características dismórficas, o condiciones con alta sospecha de una causa monogénica como displasias esqueléticas o alteraciones neurometabólicas, pueden mejorar o disminuir el rendimiento diagnóstico de los WES(101,124). Incluso, la posibilidad de contar con descripciones de fenotipos detallados permite usar sistemas de priorización de variantes automatizado que asignan

un score a la similitud con los síntomas del paciente que permite una aproximación más rápida a las variantes de interés para ser analizadas(125).

El uso de términos HPO, fue una herramienta útil para alcanzar estos rendimientos diagnósticos de forma estandarizada y clasificar correctamente los fenotipos, en este trabajo se documentó una cantidad promedio de 10 términos HPO por caso, con una mínima de 1 término y una máxima de 26. La utilidad de esta aproximación diagnóstica a través de términos HPO es ampliamente reconocida en diversos estudios de NGS (119,120,126), y también se reconoce la necesidad de realizar historiales clínicos exhaustivos con exámenes físicos altamente descriptivos con la mayor información clínica posible, para abordar un estudio WES.

En la actualidad, se ha documentado que el análisis de CNVs aumenta el rendimiento diagnóstico de los estudios de NGS, sin embargo, este aumento es muy variable, pero supone una ventaja para optimizar el abordaje de los pacientes. En este trabajo no se documentaron variantes en el número de copias CNVs, lo cual pudo ocurrir porque una gran parte de los pacientes ya contaban con estudios previos de microarreglos o MLPA, o debido al tamaño muestral limitado, ya que las CNV no son una causa común en enfermedades de herencia mendeliana. Lo anterior, puede ser contrastado con *Gao y colaboradores* que en probandos sin estudios genéticos previos, obtuvieron un rendimiento diagnóstico de 14.8% (8/54) en el análisis de CNVs a partir de exomas trio en un grupo de 54 pacientes con desordenes del neurodesarrollo, lo cual fue significativamente superior ( $p= 0.016$ ) comparado con el rendimiento si se hubiera realizado el análisis exclusivo de SNVs 29.6% (16/54) para un rendimiento diagnóstico total de 44.4% (127). *Dong y colaboradores* reportaron un rendimiento de 12.02% (131/1090) para CNVs a través de secuenciación de exoma en una cohorte de pacientes chinos con enfermedades del neurodesarrollo (86), y *Reuter y colaboradores* reportaron un rendimiento de 1.8% (2/111) para CNVs en un grupo de pacientes con enfermedades cardiacas a quienes se les realizó estudio de genoma completo (94).

Pese a que 34 individuos de esta investigación contaban con un estudio genético previo, el 32.3% (11/34) fueron diagnosticados a través de la secuenciación de exoma completo, lo cual podría sugerir un beneficio del estudio de exoma como primera línea para reducir la odisea diagnóstica en los pacientes con condiciones de etiología desconocida y además

fue posible aumentar en aproximadamente un tercio el diagnóstico en los pacientes a través del exoma. Este beneficio fue estudiado por Córdoba y colaboradores en un grupo de 40 pacientes argentinos con condiciones neurogenéticas, donde se obtuvo un rendimiento diagnóstico del 40% (16/40) para la secuenciación de exoma con un tiempo promedio en la trayectoria diagnóstica de 11 años y un costo de 1.646 USD, que podrían reducirse en un 60% en los recursos si se solicita el exoma en primera línea para sospecha de enfermedades neurogenéticas (8). Será de futuros estudios explorar el beneficio de la implementación de exoma en primera línea de diagnóstico en Colombia en cuanto a la reducción del tiempo diagnóstico y la reducción de los costos en servicios de salud a mediano y largo plazo.

En cuanto a las variables sociodemográficas, se encontró que hubo un rendimiento diagnóstico superior en el grupo de mujeres 37.2% (19/51), respecto al 26.5% (13/49) en el grupo de hombres. Este hallazgo, probablemente es debido a la cantidad de casos de herencia ligada al X dominante que se encontró en este estudio fue elevado y corresponde al 7%.

Respecto a la edad, el grupo etario con mayor rendimiento diagnóstico fue el quinquenio de 21 a 25 años donde fueron diagnosticados el 60% (3/5) de los casos, diferente a lo observado en otros estudios donde es mayor el diagnóstico en probandos menores de 20 años (86,123,127). Sin embargo, este grupo estaba sub-representado (solo 5 casos) y al revisar los grupos de edad con mayor población en la muestra de estudio, el mayor rendimiento se obtuvo en el grupo de 6 a 10 años con un 38.1% (8/21), lo cual es comparable con los estudios mencionados previamente.

Es relevante el grupo etario de 0 a 5 años donde el rendimiento fue inferior 29.17% (14/48), pues se conoce que alrededor del 71.9% de las enfermedades raras son de origen mendeliano (14), y de estas el 69.9% son de inicio exclusivamente pediátrico de acuerdo con registros de Orphanet (14), entonces se esperaría haber obtenido el mayor rendimiento en este grupo, como se ha sido documentado en otras investigaciones como *Stark y colaboradores* con rendimiento del 57.5% (46/80) en WES en niños hasta de 2 años de edad. No obstante, se debe considerar que las condiciones monogénicas son responsables del 35% de la mortalidad en el primer año de vida (9) y por ende puede existir un subdiagnóstico en este grupo etario, además el fenotipo suele ser más severo y

clínicamente reconocido a medida que avanza la edad, lo cual refleja la necesidad de mejorar las bases de datos en etapas prenatales, neonatales e infantiles (102).

Este trabajo involucró individuos de 16 departamentos de Colombia que, al visualizar su distribución geográfica se encontró que la mayoría pertenecen al centro del país principalmente la región Andina y parte de la región Caribe, lo cual se corresponde con la distribución de los especialistas en el país(128).

Un total de 32 pacientes fueron diagnosticados con condiciones monogénicas y se observó que el patrón de herencia más frecuente fue autosómico dominante 19%, seguido por 8% ligado al X, 4% herencia recesiva y 1% herencia digénica. Si bien la herencia autosómica dominante es el mecanismo de herencia más frecuente por prevalencia, es de resaltar la alta frecuencia de condiciones ligadas al cromosoma X dominante 7%, pues según registros de Orphanet la frecuencia de estas condiciones oscila en el 1% (14).

El 2% de los individuos tuvo un diagnóstico dual, los dos con condiciones neurogenéticas, se corresponde con el de otros estudios ya descritos el cual oscila entre el 1 y el 3% (86,127).

Por otro lado, se estableció una frecuencia del 2% de hallazgos incidentales en genes accionables para síndrome Lynch en los genes *PMS2* y *MSH6*, este resultado es ligeramente superior a lo descrito en la literatura, donde se estima que alrededor del 1% de los estudios por NGS presentan una variante en uno de los genes accionables sugeridos por el ACMG (129). En contraparte, se debe tener en cuenta que la cantidad de genes sugeridos por ACMG ha aumentado hasta su última actualización en el 2022 (104), lo cual conlleva a un aumento en la frecuencia de variantes reportadas en los estudios de NGS.

Un caso de herencia digénica fue identificado, en el cual se encontraron dos variantes en los genes *NPHP3* y *NPHP4*, asociados con nefronoptosis, en un individuo cuyo hermano presentaba sintomatología similar. En la actualidad, la prevalencia de la herencia digénica es desconocida, sin embargo, se ha descrito ampliamente en enfermedades de etiología renal como ocurrió en este estudio (130).

Al realizar la clasificación en grupos fenotípicos, se identificaron 41 características, donde la característica con mayor rendimiento fenotípico fue la parálisis cerebral con un 71.43% (5/7). Entiéndase la parálisis cerebral como un grupo de desórdenes permanentes del desarrollo del movimiento y la postura, que causa limitación en la actividad motora, la cual es atribuida a alteraciones no progresivas que ocurren en el desarrollo fetal o en el cerebro infantil, las alteraciones motoras de la parálisis cerebral suelen estar acompañadas de compromiso de la sensorpercepción, cognición, comunicación y comportamiento, por epilepsia y problemas musculoesqueléticos secundarios (131). Se encontró un único trabajo en la literatura desarrollado por *Moreno-De-Luca y colaboradores* quienes documentaron un rendimiento diagnóstico de la secuenciación de exoma del 32.7% en una cohorte de 1345 pacientes pediátricos con parálisis cerebral, y un rendimiento de 10.5% en una cohorte de 181 pacientes adultos con el mismo cuadro clínico (132), el cual muestra rendimientos inferiores a los de esta investigación, pero con un mejor tamaño muestral.

En orden de frecuencia diagnóstica, después de la parálisis, el mayor rendimiento diagnóstico lo ocupó el compromiso auditivo 55.56%, seguido por miopatía, microcefalia, compromiso hematológico e hipertensión pulmonar tuvieron un rendimiento diagnóstico del 50%. Todas estas características, se caracterizan por tener una amplia heterogeneidad genética, lo cual supone mayor posibilidad de obtener un mayor rendimiento diagnóstico al momento de realizar un análisis de exoma. En una cohorte de 100.000 estudios de NGS, se documentó que el compromiso auditivo tuvo un rendimiento diagnóstico superior al 40%, similar a los presentado en este estudio, pero discrepó en el rendimiento para trastornos hematológicos que tuvieron un rendimiento inferior al 20% (126)

La característica más frecuentemente documentada en la población de estudio fue el retraso en el desarrollo psicomotor 47% y obtuvo un rendimiento del 38.3% muy similar a lo encontrado en estudios que tienen rendimientos diagnósticos que oscilan entre el 36% y el 44.4% para trastornos del neurodesarrollo, y que incluso consideran el uso de exoma dentro de la primera línea en el diagnóstico de estas condiciones (8,86,123,127).

Al realizar la subclasificación de los individuos con retraso psicomotor, en subgrupos con otras características clínicas de interés, el mayor rendimiento se obtuvo con la asociación con microcefalia y otras malformaciones mayores 63.64%, comparable con *Dong y*



*colaboradores*, donde también fue la combinación fenotípica con mayor rendimiento diagnóstico con 50.32% (86).

De esta investigación surgen algunas recomendaciones como son considerar el uso de exoma en el abordaje de pacientes con sospecha de condiciones monogénicas por su buen rendimiento diagnóstico y su aplicación de forma temprana en aquellas condiciones que comprometen el sistema nervioso central, en especial para el retraso en el desarrollo psicomotor, que en general tienen un mejor rendimiento si se asocia con microcefalia/malformaciones mayores o compromiso estructural del SNC, sin embargo, estas consideraciones deben ser abordadas en estudios con un tamaño de muestra más grande.

Desde el año 2010 la hibridación genómica por microarreglos es el estudio de primera línea para el diagnóstico molecular de desórdenes del neurodesarrollo y malformaciones congénitas por su rendimiento diagnóstico que oscila entre el 15 y el 20% (133), sin embargo, como se ha expuesto ampliamente el estudio de exoma alcanza rendimientos superiores al 35% para esta condición, y en algunos estudios de costoefectividad se ha documentado que previo al WES en promedio el gasto de un paciente son de 7225 euros, y posterior al WES los estudios de seguimiento se reducen significativamente y los costos bajan en un 80%, por ello debería considerarse como un estudio de primera línea en el abordaje de paciente con compromiso en el neurodesarrollo (134). Por el aumento que se hay ido demostrando en el rendimiento diagnóstico en este grupo fenotípico específico de retraso en el neurodesarrollo, desde el año 2019 el Colegio Americano de Genética Médica ha recomendado usar el estudio de exoma como primera línea en el algoritmo diagnóstico.

Si bien en Colombia no se cuenta con estudios con un tamaño de muestra grande, este estudio es comparable a lo observado en la literatura y podrá ser usado para consideraciones próximas en el desarrollo de guías y recomendaciones para el abordaje de pacientes con retraso en el neurodesarrollo en Colombia.

Finalmente, se deberá explorar en el futuro la reclasificación de las variantes de los casos estudiados para aumentar el rendimiento de la prueba especialmente en los casos con variantes de significado incierto y además continuar en el mejoramiento del abordaje de los estudios de NGS en la era de la medicina genómica.

## 6.1 Limitaciones

La principal limitación de este proyecto radica en la utilización de una muestra no probabilística tomada a conveniencia en un grupo limitado de casos, lo cual impide la generalización y extrapolación de los datos para la totalidad de la población.

La distribución de los pacientes estaba subrepresentada en algunos grupos etarios y por la naturaleza descriptiva de este estudio no fue posible reevaluar a los pacientes y sus padres para mejorar su descripción fenotípica y por tanto el filtrado de variantes.

Debido al tamaño muestral y la poca representatividad de algunos grupos fenotípicos no fue posible cumplir a cabalidad con el último objetivo planteado, para establecer recomendaciones respecto a las indicaciones para solicitar un estudio de exoma.

## 7 Conclusiones

- Se desarrolló el primer estudio en Colombia que evaluó el rendimiento diagnóstico de la secuenciación de exoma en trio e individual.
- El rendimiento del exoma trio fue un 3% superior al análisis del exoma individual.
- El uso del exoma individual tiene un rendimiento alto y comparable con el exoma trio en el grupo de pacientes estudiados, lo que podría sugerir un beneficio en su uso de forma temprana, sin evaluar otras variables de costoefectividad.
- Este estudio mostró la utilidad del estudio de exoma en fenotipos neurológicos comparable con otros estudios reportados en la literatura, especialmente, en el grupo fenotípico con parálisis que tuvo el mayor rendimiento diagnóstico.
- El uso de secuenciación de exoma mostró una buena utilidad en pacientes con retraso en el desarrollo psicomotor asociado a otras características fenotípicas como microcefalia/malformaciones mayores o compromiso estructural del SNC.
- Este trabajo evidencia la necesidad de realizar estudios adicionales aumentando el tamaño de muestra y las características sociodemográficas para esclarecer indicaciones precisas para solicitar estudios de exoma en el ámbito clínico colombiano.



## Bibliografía

1. Durmaz AA, Karaca E, Demkow U, Toruner G, Schoumans J, Cogulu O. Evolution of genetic techniques: Past, present, and beyond. *Biomed Res Int*. 2015.
2. Giani AM, Gallo GR, Gianfranceschi L, Formenti G. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput Struct Biotechnol J [Internet]*. 2020;18:9–19. Disponible en: <https://doi.org/10.1016/j.csbj.2019.11.002>
3. Hood L, Rowen L. The human genome project: Big science transforms biology and medicine. *Genome Med*. 2013;5(9):1.
4. Nurk S, Koren S, Rhie A, Rautiainen M, Bizakadze A V, Mikheenko A, et al. The complete sequence of a human genome. 2023.
5. Entry Statistics - OMIM [Internet]. [citado el 10 de febrero de 2023]. Disponible en: <https://www.omim.org/statistics/entry>
6. Czech M, Baran-Kooiker A, Atikeler K, Demirtshyan M, Gaitova K, Holownia-Voloskova M, et al. A Review of Rare Disease Policies and Orphan Drug Reimbursement Systems in 12 Eurasian Countries. *Front Public Health*. 2020;7(January):1–17.
7. Tan TY, Lunke S, Chong B, Phelan D, Fanjul-Fernandez M, Marum JE, et al. A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis. *European Journal of Human Genetics [Internet]*. 2019; Disponible en: <http://dx.doi.org/10.1038/s41431-019-0471-9>
8. Córdoba M, Rodríguez-Quiroga SA, Vega PA, Salinas V, Pérez-Maturo J, Amartino H, et al. Whole exome sequencing in neurogenetic odysseys: An effective, cost- and time-saving diagnostic approach. *PLoS One*. 2018;13(2):1–13.
9. Wright CF, FitzPatrick DR, Firth H v. Paediatric genomics: Diagnosing rare disease in children. *Nat Rev Genet [Internet]*. 2018;19(5):253–68. Disponible en: <http://dx.doi.org/10.1038/nrg.2017.116>

10. Petersen B, Sabina, Fredrich B, Hoepfner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. 2017;1–13.
11. Kim YE, Ki CS, Jang MA. Erratum: Challenges and Considerations in Sequence Variant Interpretation for Mendelian Disorders. *Ann Lab Med*. 2019;39(6):606.
12. Harrison SM, Dolinsky JS, Knight Johnson AE, Pesaran T, Azzariti DR, Bale S, et al. Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genetics in Medicine*. 2017;19(10):1096–104.
13. Derayeh S, Kazemi A, Rabiei R, Hosseini A, Moghaddasi H. National information system for rare diseases with an approach to data architecture : A systematic review. 2018;7(3):156–63.
14. Nguengang S, Deborah W, Annie ML, Charlotte O, Charlotte R, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases : analysis of the Orphanet database. *European Journal of Human Genetics* [Internet]. 2020;165–73. Disponible en: <http://dx.doi.org/10.1038/s41431-019-0508-0>
15. Boycott KM, Hartley T, Biesecker LG, Gibbs RA, Innes AM, Riess O, et al. Commentary A Diagnosis for All Rare Genetic Diseases : The Horizon and the Next Frontiers. *Cell*. 2019;177(1):32–7.
16. Hoffman-Andrews L. The known unknown: The challenges of genetic variants of uncertain significance in clinical practice. *J Law Biosci*. 2017;4(3):648–57.
17. Mullis KB. The unusual origin of the polymerase chain reaction. *Sci Am*. 1990;262(4):56–65.
18. Gupta N, Verma VK. Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality. 2019;313–41.
19. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*. 2008;24(3):133–41.
20. Zhu FY, Chen MX, Ye NH, Qiao WM, Gao B, Law WK, et al. Comparative performance of the BGISEQ-500 and Illumina HiSeq4000 sequencing platforms for transcriptome analysis in plants. *Plant Methods*. 2018;14(1):1–14.
21. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. *Nat Rev Genet* [Internet]. 2016;17(6):333–51. Disponible en: <http://dx.doi.org/10.1038/nrg.2016.49>
22. What is the Human Genome Project? [Internet]. [citado el 15 de julio de 2020]. Disponible en: <https://www.genome.gov/human-genome-project/What>

23. International HapMap Consortium. International HapMap Consortium. The International HapMap Project. *Nature*. 2003;426(6968):789–96.
24. Valdespino-Gómez VM, Margarita P. Organización estructural y funcional del genoma humano: variación en el número de copias predisponentes de enfermedades degenerativas. *Gaceta Mexicana de Oncología*. 2013;12(6):426–32.
25. Belmont JW, Boudreau A, Leal SM, Hardenbol P, Pasternak S, Wheeler DA, et al. A haplotype map of the human genome. *Nature*. el 27 de octubre de 2005;437(7063):1299–320.
26. HapMap [Internet]. [citado el 6 de febrero de 2023]. Disponible en: <https://www.ncbi.nlm.nih.gov/probe/docs/projhapmap/>
27. Via M, Gignoux C, Burchard EG. The 1000 Genomes Project: New opportunities for research and social challenges. Vol. 2, *Genome Medicine*. 2010.
28. Zheng-Bradley X, Flicek P. Applications of the 1000 Genomes Project resources. *Brief Funct Genomics*. el 1 de mayo de 2017;16(3):163–70.
29. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. el 6 de septiembre de 2012;489(7414):57–74.
30. Vamathevan J, Birney E. A Review of Recent Advances in Translational Bioinformatics: Bridges from Biology to Medicine. *Yearb Med Inform*. 2017;26(1):178–87.
31. Roth SC. What is genomic medicine? *Journal of the Medical Library Association*. 2019;107(3):442–8.
32. Corpas M, Kovalevskaya N V., McMurray A, Nielsen FGG. A FAIR guide for data providers to maximise sharing of human genomic data. *PLoS Comput Biol*. 2018;14(3):1–10.
33. BLAST: Basic Local Alignment Search Tool [Internet]. [citado el 7 de febrero de 2023]. Disponible en: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
34. Ensembl navegador del genoma 108 [Internet]. [citado el 7 de febrero de 2023]. Disponible en: <https://www.ensembl.org/index.html>
35. All Resources - Site Guide - NCBI [Internet]. [citado el 10 de febrero de 2023]. Disponible en: <https://www.ncbi.nlm.nih.gov/guide/all/>
36. About UniProt | UniProt help | UniProt [Internet]. [citado el 10 de febrero de 2023]. Disponible en: <https://www.uniprot.org/help/about>

37. About gnomAD | gnomAD [Internet]. [citado el 10 de febrero de 2023]. Disponible en: <https://gnomad.broadinstitute.org/about>
38. None [Internet]. [citado el 10 de febrero de 2023]. Disponible en: [https://www.ncbi.nlm.nih.gov/snp/docs/RefSNP\\_about/](https://www.ncbi.nlm.nih.gov/snp/docs/RefSNP_about/)
39. What is ClinVar? [Internet]. [citado el 10 de febrero de 2023]. Disponible en: <https://www.ncbi.nlm.nih.gov/clinvar/intro/>
40. About OMIM [Internet]. [citado el 5 de noviembre de 2020]. Disponible en: <https://omim.org/about>
41. Stenson PD, Mort M, Ball E v., Chapman M, Evans K, Azevedo L, et al. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet.* el 1 de octubre de 2020;139(10):1197–207.
42. Clinical Genomic Database [Internet]. [citado el 10 de febrero de 2023]. Disponible en: <https://research.nhgri.nih.gov/CGD/>
43. About - DECIPHER v11.17 [Internet]. [citado el 10 de febrero de 2023]. Disponible en: <https://www.deciphergenomics.org/about/overview>
44. Human Phenotype Ontology [Internet]. [citado el 10 de febrero de 2023]. Disponible en: <https://hpo.jax.org/app/about>
45. Köhler S, Gargano M, Matentzoglou N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The human phenotype ontology in 2021. *Nucleic Acids Res.* el 8 de enero de 2021;49(D1):D1207–17.
46. WHO | Screening the genes [Internet]. [citado el 15 de julio de 2020]. Disponible en: <https://www.who.int/bulletin/volumes/90/8/12-030812/en/>
47. Khosla N, Valdez R. A compilation of national plans, policies and government actions for rare diseases in 23 countries. *Intractable Rare Dis Res.* 2018;7(4):213–22.
48. Jaimes CA, Espinoza DAE. Actores del Sistema General de Seguridad Social en salud. 2011;(4).
49. Ministerio de Salud y Protección Social. Decreto 1954 de 2012. 2012;1–4.
50. Mateus HE, Pérez AM, Mesa ML, Escobar G, Gálvez JM, Montaña JI, et al. A first description of the Colombian national registry for rare diseases. *BMC Res Notes.* 2017;10(1):9–11.



51. Resolución 0023 de 2023/ Por medio de la cual se actualiza el listado de enfermedades huérfanas – raras. Ministerio de Salud y Protección social. 4 de Enero de 2023..
52. Edwin F, Alvarado P. Informe de evento: Enfermedades Huérfanas-Raras Créditos Martha Lucía Ospina Martínez Directora General. Instituto Nacional de Salud. 2022.
53. Enfermedades huérfanas-raras Casos notificados al Sivigila. [citado el 1 de febrero de 2023]; Disponible en: [www.ins.gov.co](http://www.ins.gov.co)
54. Wallace SE, Bean LJH. Educational Materials — Genetic Testing : Current Approaches This discussion addresses clinical tests available through CLIA-certified laboratories in the United States . Many inherited disorders and phenotypes are genetically heterogeneous – that is , pa. 2019;1–13.
55. Berg AO, Berg JS, Brown CW, Burke W, Calonge BN, Chung WK, et al. An evidence framework for genetic testing. *An Evidence Framework for Genetic Testing*. 2017. 1–136 p.
56. Piovesan A, Antonaros F, Vitale L, Strippoli P, Pelleri MC, Caracausi M. Human protein-coding genes and gene feature statistics in 2019. *BMC Res Notes*. el 4 de junio de 2019;12(1).
57. Garcia I, Jones E, Ramos M, Innis-Whitehouse W, Gilkerson R. The little big genome: The organization of mitochondrial DNA. *Frontiers in Bioscience - Landmark*. 2017;22(4):710–21.
58. Li Huanzheng, Slone Jesse, Fei Lin. Mitochondrial DNA Variants and Comon Diseases. *Cells*. 2019.
59. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2016;107(1):1–8.
60. Hess JF, Kohl TA, Kotrová M, Rönsch K, Paprotka T, Mohr V, et al. Library preparation for next generation sequencing: A review of automation strategies. *Biotechnol Adv*. 2020;41(March):107537.
61. Illumina Inc. Illumina sequencing introduction. *Illumina sequencing introduction*. 2017;(October):1–8.
62. Exome Capture V5 Probe Set User Manual MGIEasy II Revision History [Internet]. Disponible en: <https://en.mgi-tech.com/download/files.html#Contents>
63. DNBSEQ-G400: Comprehensive and flexible genome sequencer-MGI [Internet]. [citado el 18 de julio de 2020]. Disponible en: [https://en.mgitech.cn/products/instruments\\_info/2/](https://en.mgitech.cn/products/instruments_info/2/)

64. History-BGI Group Official Website [Internet]. [citado el 18 de julio de 2020]. Disponible en: <https://en.genomics.cn/en-history.html>
65. Lelieveld SH, Veltman JA, Gilissen C. Novel bioinformatic developments for exome sequencing. *Hum Genet.* 2016;135(6):603–14.
66. He KY, Ge D, He MM. Big data analytics for genomic medicine. *Int J Mol Sci.* 2017;18(2):1–18.
67. Bartha Á, Györfy B. Comprehensive outline of whole exome sequencing data analysis tools available in clinical oncology. *Cancers (Basel).* 2019;11(11):1–20.
68. Adams DR, Eng CM. Next-generation sequencing to diagnose suspected genetic disorders. *New England Journal of Medicine.* 2018;379(14):1353–62.
69. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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. *Genetics in Medicine.* 2015;17(5):405–24.
70. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* el 1 de enero de 2018;46(D1):D1062–7.
71. Landrum MJ, Kattman BL. ClinVar at five years: Delivering on the promise. *Hum Mutat.* el 1 de noviembre de 2018;39(11):1623–30.
72. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: Improvements to accessing data. *Nucleic Acids Res.* el 1 de enero de 2020;48(D1):D835–44.
73. The ClinVar record display [Internet]. [citado el 5 de noviembre de 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/clinvar/docs/details/>
74. Tavtigian S v., Harrison SM, Boucher KM, Biesecker LG. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat.* el 1 de octubre de 2020;41(10):1734–7.
75. ClinGen Sequence Variant Interpretation Work Group recommendations for ACMG/AMP guideline criteria code modifications Version 1.0 Working Group Page: [https://clinicalgenome.org/working-groups/sequence-variant-interpretation/svi\\_criteria\\_nomenclature\\_recommendation\\_v1](https://clinicalgenome.org/working-groups/sequence-variant-interpretation/svi_criteria_nomenclature_recommendation_v1). Disponible en: <https://clinicalgenome.org/working-groups/sequence-variant-interpretation/>

76. Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat.* el 1 de noviembre de 2018;39(11):1517–24.
77. Ghosh R, Harrison SM, Rehm HL, Plon SE, Biesecker LG. Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Hum Mutat.* el 1 de noviembre de 2018;39(11):1525–30.
78. ClinGen Sequence Variant Interpretation Recommendation for PM2 - Version 1.0 Working Group Disponible en: [https://clinicalgenome.org/workinggroups/sequence-variant-interpretation/pm2\\_-\\_svi\\_recommendation\\_\\_approved\\_sept\\_2020](https://clinicalgenome.org/workinggroups/sequence-variant-interpretation/pm2_-_svi_recommendation__approved_sept_2020).
79. ClinGen Sequence Variant Interpretation Recommendation for in trans Criterion (PM3)-Version 1.0 Working Group Page: [https://clinicalgenome.org/working-groups/sequence-variant-interpretation/SVI\\_Recommendation\\_for\\_in\\_trans\\_Criterion\\_\(PM3\)-Version\\_1.0](https://clinicalgenome.org/working-groups/sequence-variant-interpretation/SVI_Recommendation_for_in_trans_Criterion_(PM3)-Version_1.0) [Internet]. 2019. Disponible en: <https://clinicalgenome.org/working-groups/sequence-variant-interpretation/>
80. ClinGen Sequence Variant Interpretation Recommendation for de novo Criteria (PS2/PM6)-Version 1.1 Working Group Page: <https://clinicalgenome.org/working-groups/sequence-variant-interpretation> [Internet]. 2018. Disponible en: <https://clinicalgenome.org/working-groups/sequence-variant-interpretation/>
81. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Vol. 12, *Genome Medicine*. BioMed Central Ltd; 2019.
82. Biesecker LG, Harrison SM. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. Vol. 20, *Genetics in Medicine*. Nature Publishing Group; 2018. p. 1687–8.
83. Slavin TP, Manjarrez S, Pritchard CC, Gray S, Weitzel JN. The effects of genomic germline variant reclassification on clinical cancer care [Internet]. Vol. 10, *Oncotarget*. 2019. Disponible en: [www.oncotarget.com](http://www.oncotarget.com)
84. Ellard S, Baple EL, Callaway A, Berry I, Forrester N, Turnbull C, et al. ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020 Recommendations ratified by ACGS Quality Subcommittee on 4 th. 2020; Disponible en: <https://doi.org/10.1101/531210>
85. Mersch J, Brown N, Pirzadeh-Miller S, Mundt E, Cox HC, Brown K, et al. Prevalence of variant reclassification following hereditary cancer genetic testing. *JAMA - Journal of the American Medical Association.* el 25 de septiembre de 2018;320(12):1266–74.

86. Dong X, Liu B, Yang L, Wang H, Wu B, Liu R, et al. Clinical exome sequencing as the first-tier test for diagnosing developmental disorders covering both CNV and SNV: A Chinese cohort. *J Med Genet.* 2020;1–9.
87. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Hum Mutat.* 2015;36(8):815–22.
88. Dillon OJ, Lunke S, Stark Z, Yeung A, Thorne N, Gaff C, et al. Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. *European Journal of Human Genetics.* 2018;26(5):644–51.
89. Pengelly RJ, Ward D, Hunt D, Mattocks C, Ennis S. Comparison of Mendeliome exome capture kits for use in clinical diagnostics. *Sci Rep.* 2020;10(1):1–7.
90. Bertier G, Héту M, Joly Y. Unsolved challenges of clinical whole-exome sequencing: A systematic literature review of end-users' views Donna Dickenson, Sandra Soo-Jin Lee, and Michael Morrison. *BMC Med Genomics.* 2016;9(1):1–12.
91. Schwarze K, Buchanan J, Taylor JC, Wordsworth S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genetics in Medicine.* 2018;20(10):1122–30.
92. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing: Editorial comment. *Obstet Gynecol Surv.* 2015;70(3):164–7.
93. Sawyer SL, Hartley T, Dymant DA, Beaulieu CL, Schwartzentruber J, Smith A, et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: Time to address gaps in care. *Clin Genet.* 2016;89(3):275–84.
94. Reuter MS, Chaturvedi RR, Liston E, Manshaei R, Aul RB, Bowdin S, et al. The Cardiac Genome Clinic: implementing genome sequencing in pediatric heart disease. *Genetics in Medicine [Internet].* 2020;22(6):1015–24. Disponible en: <http://dx.doi.org/10.1038/s41436-020-0757-x>
95. Seidelmann SB, Smith E, Subrahmanyam L, Dykas D, Ziki MDA, Azari B, et al. Application of Whole Exome Sequencing in the Clinical Diagnosis and Management of Inherited Cardiovascular Diseases in Adults. *Circ Cardiovasc Genet.* 2017;10(1).
96. Saes JL, Simons A, de Munnik SA, Nijziel MR, Blijlevens NMA, Jongmans MC, et al. Whole exome sequencing in the diagnostic workup of patients with a bleeding diathesis. *Haemophilia.* 2019;25(1):127–35.

97. Romasko EJ, Devkota B, Biswas S, Jayaraman V, Jairam S, Scarano MI, et al. HHS Public Access. 2019;93(1):8–16.
98. Groopman EE, Marasa M, Cameron-Christie S, Petrovski S, Aggarwal VS, Milo-Rasouly H, et al. Diagnostic Utility of exome sequencing for kidney disease. *New England Journal of Medicine*. 2019;380(2):142–51.
99. Warejko JK, Tan W, Daga A, Schapiro D, Lawson JA, Shril S, et al. Whole exome sequencing of patients with steroid-resistant nephrotic syndrome. *Clinical Journal of the American Society of Nephrology*. 2018;13(1):53–62.
100. Polla DL, Cardoso MTO, Silva MCB, Cardoso ICC, Medina CTN, Araujo R, et al. Use of targeted exome sequencing for molecular diagnosis of skeletal disorders. *PLoS One*. 2015;10(9):1–17.
101. Kingsmore SF, Cakici JA, Clark MM, Gaughran M, Feddock M, Batalov S, et al. A Randomized, Controlled Trial of the Analytic and Diagnostic Performance of Singleton and Trio, Rapid Genome and Exome Sequencing in Ill Infants. *Am J Hum Genet*. 2019;105(4):719–33.
102. Vora NL, Gilmore K, Brandt A, Gustafson C, Strande N, Ramkissoon L, et al. An approach to integrating exome sequencing for fetal structural anomalies into clinical practice. *Genetics in Medicine [Internet]*. 2020;22(5):954–61. Disponible en: <http://dx.doi.org/10.1038/s41436-020-0750-4>
103. MGIEasy Exome Capture V5 Probe Set-Library Prep-MGI [Internet]. [citado el 11 de noviembre de 2020]. Disponible en: [https://en.mgi-tech.com/Products/reagents\\_info/id/10](https://en.mgi-tech.com/Products/reagents_info/id/10)
104. Miller DT, Lee K, Abul-Husn NS, Amendola LM, Brothers K, Chung WK, et al. ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). Vol. 24, *Genetics in Medicine*. Elsevier B.V.; 2022. p. 1407–14.
105. IBM Corp. Publicado en 2019. IBM SPSS Statistics para Windows, versión 26.0. Armonk, Nueva York: IBM Corp.
106. UCSC Genome Browser Gateway [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <https://genome.ucsc.edu/cgi-bin/hgGateway>
107. UniProt [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <https://www.uniprot.org/>
108. gnomAD [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <https://gnomad.broadinstitute.org/>

109. GeneCards - Human Genes | Gene Database | Gene Search [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <https://www.genecards.org/>
110. Home - OMIM [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <https://www.omim.org/>
111. MalaCards - human disease database [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <https://www.malacards.org/>
112. PolyPhen-2: prediction of functional effects of human nsSNPs [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <http://genetics.bwh.harvard.edu/pph2/>
113. MutationTaster [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <https://www.mutationtaster.org/>
114. FLNA curation results [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <https://search.clinicalgenome.org/kb/genes/HGNC:3754>
115. Hoefele J, Wolf MTF, O'Toole JF, Otto EA, Schultheiss U, Dêschenes G, et al. Evidence of oligogenic inheritance in nephronophthisis. *Journal of the American Society of Nephrology*. octubre de 2007;18(10):2789–95.
116. PQBP1 curation results [Internet]. [citado el 16 de febrero de 2023]. Disponible en: <https://search.clinicalgenome.org/kb/genes/HGNC:9330>
117. Resolución 5857 de 2018/ Por medio de la cual se actualiza integralmente el Plan de Beneficios en Salud con cargo a la Unidad de Pago por Capitacion (UPC). Ministerio de Salud y Protección social. 26 de Diciembre de 2018. Republica de Colombia.
118. Alfares A, Alsubaie L, Aloraini T, Alaskar A, Althagafi A, Alahmad A, et al. What is the right sequencing approach? Solo VS extended family analysis in consanguineous populations. *BMC Med Genomics*. el 17 de julio de 2020;13(1).
119. Tan TY, Lunke S, Chong B, Phelan D, Fanjul-Fernandez M, Marum JE, et al. A head-to-head evaluation of the diagnostic efficacy and costs of trio versus singleton exome sequencing analysis. *European Journal of Human Genetics*. 2019;
120. Hiz Kurul S, Oktay Y, Töpf A, Szabó NZ, Güngör S, Yaramis A, et al. High diagnostic rate of trio exome sequencing in consanguineous families with neurogenetic diseases. *Brain*. el 1 de abril de 2022;145(4):1507–18.
121. Liascovich R, Nica Rittler M, Castilla EE, Maternoinfantil R, Sardá B, Aires A. Consanguinity in South America: Demographic Aspects. Vol. 51, *Hum Hered*. 2001.

122. Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, et al. Clinical application of whole-exome sequencing across clinical indications. *Genetics in Medicine*. el 1 de julio de 2016;18(7):696–704.
123. Kim SH, Kim B, Lee JS, Kim HD, Choi JR, Lee ST, et al. Proband-Only Clinical Exome Sequencing for Neurodevelopmental Disabilities. *Pediatr Neurol*. el 1 de octubre de 2019;99:47–54.
124. Stark Z, Tan TY, Chong B, Brett GR, Yap P, Walsh M, et al. A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genetics in Medicine*. el 1 de noviembre de 2016;18(11):1090–6.
125. Seo GH, Kim T, Choi IH, Park J young, Lee J, Kim S, et al. Diagnostic yield and clinical utility of whole exome sequencing using an automated variant prioritization system, EVIDENCE. *Clin Genet*. el 1 de diciembre de 2020;98(6):562–70.
126. Smedley D, Smith K, Martin A, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. *New England Journal of Medicine* Noviembre de 2021.
127. Gao C, Wang X, Mei S, Li D, Duan J, Zhang P, et al. Diagnostic yields of trio-wes accompanied by CnVSeq for rare neurodevelopmental disorders. *Front Genet*. 2019;10(MAY).
128. Fernando Ochoa Zuluaga Universidad del Rosario L. Informe de pasantía y trabajo de grado de la maestría en salud pública: análisis sobre la cantidad de especialistas en el sistema de salud colombiano y la incidencia de los cupos de residencia medica sobre estos. 2018.
129. Olfson E, Cottrell CE, Davidson NO, Gurnett CA, Heusel JW, Stitzel NO, et al. Identification of medically actionable secondary findings in the 1000 genomes. *PLoS One*. el 2 de septiembre de 2015;10(9).
130. Deltas C. Digenic inheritance and genetic modifiers. Vol. 93, *Clinical Genetics*. Blackwell Publishing Ltd; 2018. p. 429–38.
131. Sadowska M, Sarecka-Hujar B, Kopyta I. Cerebral palsy: Current opinions on definition, epidemiology, risk factors, classification and treatment options. Vol. 16, *Neuropsychiatric Disease and Treatment*. Dove Medical Press Ltd; 2020. p. 1505–18.
132. Moreno-De-Luca A, Millan F, Pesacreta DR, Elloumi HZ, Oetjens MT, Teigen C, et al. Molecular Diagnostic Yield of Exome Sequencing in Patients with Cerebral

- Palsy. JAMA - Journal of the American Medical Association. el 2 de febrero de 2021;325(5):467–75.
133. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. Am J Hum Genet. el 14 de mayo de 2010;86(5):749–64.
134. Vrijenhoek T, Middelburg EM, Monroe GR, van Gassen KLI, Geenen JW, Hövels AM, et al. Whole-exome sequencing in intellectual disability; cost before and after a diagnosis. European Journal of Human Genetics. el 1 de noviembre de 2018;26(11):1566–71.