



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Química Biológica

Utilización de técnicas de secuenciación masiva para el diagnóstico molecular de enfermedades poco frecuentes.

Tesis presentada para optar al título de Doctora de la Universidad de Buenos Aires en el área de Química Biológica.

Guadalupe Buda

Director de tesis: Dr. Marcelo Adrián Martí

Consejera de estudios: Dra. Rosanna Ramhorst

Lugar de trabajo: Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.

Buenos Aires, 2023

Utilización de técnicas de secuenciación masiva para el diagnóstico molecular de enfermedades poco frecuentes.

Resumen

En Argentina, las enfermedades poco frecuentes de origen genético, afectan a alrededor de 3,5 millones de habitantes. La lista se ha expandido como consecuencia de la aplicación de tecnologías de secuenciación masiva y, actualmente, comprende un número creciente de trastornos monogénicos, con un patrón de herencia mendeliana. En este contexto, el análisis del genoma completo y especialmente de su parte codificante, el exoma, se ha convertido en una herramienta de gran utilidad en el diagnóstico de las mismas. El presente proyecto de tesis buscó beneficiar a aquellos pacientes imposibilitados de acceder y costear estas tecnologías, otorgándoles una potencial oportunidad de arribar a un diagnóstico. De esta manera, nos propusimos como objetivo global implementar tecnologías de secuenciación masiva para lograr el diagnóstico molecular preciso de entidades poco frecuentes en la clínica y, en particular y de forma específica, desarrollar un protocolo general que facilitara el análisis e identificación de variantes genéticas aplicado tanto a casos particulares como al estudio de exomas de pacientes incorporados en tres ediciones del Proyecto PAMPA, con diagnóstico presuntivo de Enfermedades Autoinflamatorias. A partir del análisis bioinformático logramos identificar, para cada caso particular, las variantes responsables del fenotipo observado, y aportar información de relevancia que permitió profundizar el conocimiento sobre la función de los genes *AFG3L2* e *ISG15* a nivel molecular y describir fenotipos novedosos. Este trabajo de tesis incluyó tres ediciones del Proyecto PAMPA, incorporó a más de 160 pacientes, y reflejó una tasa de éxito -casos donde se arribó a un potencial diagnóstico molecular- entre un 16-34%, lo que resulta consistente con estudios realizados a nivel internacional.

Palabras claves: bioinformática- secuenciación masiva- exoma- genética humana- enfermedades autoinflamatorias- enfermedades neurodegenerativas.

Next-generation sequencing for molecular diagnosis of rare diseases

Abstract

In Argentina, rare diseases affect about 3.5 million habitants. The list of rare diseases has expanded as a consequence of the application of next-generation sequencing technologies and now includes a growing number of monogenic disorders with a Mendelian inheritance pattern. In this context, the analysis of the whole genome, and especially its coding region, the exome, has become a very useful tool in the diagnosis of these diseases. The present thesis sought to benefit those patients unable to access and afford these technologies, giving them a potential opportunity to reach a diagnosis. Thus, our main objective was to implement massive sequencing technologies to achieve the precise molecular diagnosis of rare diseases in clinical genomics and specifically, to develop a general protocol to facilitate the analysis and identification of variants applied both to cases and to the study of patients with suspicion of Autoinflammatory Diseases, who participated in the PAMPA Projects. From the bioinformatic perspective, we were able to identify, for each case, the variants responsible for the phenotype and to provide relevant information that allowed us to deepen the knowledge about the function of *AFG3L2* and *ISG15* genes at a molecular level and to describe novel phenotypes. In the case of the PAMPA Project, it had three editions during my thesis work, included more than 160 patients and reflected a success rate -cases where a potential molecular diagnosis was reached- of between 16-34%, which is consistent with international studies.

Keywords: bioinformatics- NGS- exome- human genetics- autoinflammatory diseases- neurodegenerative diseases.

Índice

Índice	3
Introducción	5
Desde la investigación científica hasta el diagnóstico clínico mediante técnicas de secuenciación masiva	6
NGS como herramienta diagnóstica	8
Bioinformática de “nueva generación”	9
Genómica poblacional (El concepto de variante)	11
Frecuencias Alélicas	12
Frecuencias alélicas y la magnitud del efecto	13
Bases moleculares de las enfermedades genéticas	16
Distintos tipos de variantes y sus efectos moleculares	16
Clasificación de variantes	19
Hallazgo de variantes de significado incierto (VUS)	20
Relación fenotipo-genotipo en las enfermedades genéticas	21
Enfermedades poco frecuentes	23
Enfermedades autoinflamatorias	24
Diagnóstico de una enfermedad genética en la era NGS	32
Elección del mejor enfoque: paneles, WES o WGS	32
Dificultad de implementación local de estas tecnologías	36
Objetivos	37
Materiales y Métodos	38
Selección de Pacientes	38
Consentimiento informado	38
Protocolo para la Secuenciación Exómica	38
Hallazgos incidentales	39
Precisión y tasa de error de un experimento de secuenciación por NGS	39
Análisis bioinformático	39
Etapas generales	39
Predictores bioinformáticos	48
<i>Polyphen-2</i>	48
<i>SIFT</i>	48
<i>Mutation Taster</i>	49
Bases de datos	49
ENSEMBL	50

Uniprot	50
Online Mendelian Inheritance in Man (OMIM)	51
dbSNP	52
Clinical Variants (ClinVar)	52
Infovers	54
Registro Eurofever	54
Protocolo de priorización de variantes	55
Análisis de variantes estructurales	58
Validación de Resultados	59
Resultados	61
Caso I:	61
Introducción	61
Análisis bioinformático	67
Validación de la segregación familiar	71
Correlación genotipo- fenotipo	72
Discusión	80
Caso II:	82
Introducción	82
Análisis bioinformático	86
Discusión	96
Caso III:	98
Introducción	98
Materiales y Métodos	102
Resultados	105
Discusión	121
Discusión	123
Desafíos en el diagnóstico molecular de enfermedades mendelianas por NGS	126
Perspectivas a futuro	129
Anexo 1	131
Congresos	131
Publicaciones	131
Bibliografía	132

Introducción

De acuerdo con el catálogo OMIM (*Online Mendelian Inheritance in Man* [1]), en la actualidad se conocen alrededor de 8 mil enfermedades genéticas -la mayoría de ellas de tipo monogénicas- que, en conjunto, afectan aproximadamente al 4-8% de la población mundial en función de lo establecido por la Organización Mundial de la Salud (OMS) [2,3]. El 80% de las enfermedades poco frecuentes posee un origen genético definido, con la implicación de uno o varios genes. Otras, son causadas por infecciones (bacterianas o virales), alergias, o se deben a causas degenerativas, proliferativas o teratógenas (productos químicos, radiación, etc.) y para una minoría, aún se desconoce la etiología. Si bien pueden presentarse a cualquier edad, alrededor del 70% de los casos se manifiesta en la edad pediátrica. En Argentina, de acuerdo con la Federación Argentina de Enfermedades Poco Frecuentes (FADEPOF), estas patologías se definen como aquellas cuya prevalencia en la población es igual o inferior a 1 persona cada 2.000 habitantes, según lo establece la Ley N° 26.6891 de "Cuidado integral de la salud de las personas con EPOF y sus familias", promulgada en junio 2011 y reglamentada en el año 2015 por el Decreto 794/15. En su totalidad, en nuestro país afectan a alrededor de 3.6 millones de habitantes [4]. La mayoría de estas afecciones son crónicas y degenerativas, pudiendo llegar a producir algún tipo de discapacidad y en algunos casos, sin un adecuado diagnóstico y tratamiento, ponen en riesgo la vida de los pacientes. Es por ello que obtener un diagnóstico molecular preciso (es decir, conocer la/s variante/s responsable/s) y temprano es fundamental para un adecuado manejo terapéutico y asesoramiento genético, por ejemplo, para la futura planificación familiar.

Las enfermedades monogénicas se heredan de forma mendeliana, siguiendo diversos patrones hereditarios conocidos: autosómico dominante, autosómico recesivo, ligados al X, ligados al cromosoma Y, disomía uniparental y trastornos de *imprinting* [5]. La mayoría están causadas por variantes patogénicas, localizadas preferentemente en las regiones codificantes -exones- (o intrónicas adyacentes) de los genes, produciendo alteraciones en el ARN o la proteína resultante y ocasionando, de este modo, el fenotipo patológico.

El diagnóstico genético (o molecular) de las mismas se basa en la identificación de las variantes patogénicas en el/los gen/es asociados con la enfermedad en estudio. Es importante destacar que, usualmente, debido a la gran heterogeneidad genética existente, el arribo a un diagnóstico molecular preciso lleva años, impidiendo que en ese lapso los pacientes reciban medidas terapéuticas y de rehabilitación específicas para su enfermedad, o que sus familiares puedan entrar en programas preventivos dirigidos a la detección temprana mediante estudios genéticos [2,6].

En el presente trabajo de tesis, abordaremos dos casos de pacientes con enfermedades poco frecuentes, en los que resultó indispensable la aplicación de tecnologías

de nueva generación para llegar a un diagnóstico preciso y el desarrollo de tres ediciones del Proyecto PAMPA [7], que permitió obtener resultados y conclusiones interesantes a partir de la aplicación de tecnologías NGS a grupos seleccionados de pacientes con enfermedades autoinflamatorias (EAI).

En general, la estrategia actual de diagnóstico de las EAI consiste en iniciar el estudio de el/los gen/es más frecuentemente asociado/s al diagnóstico presuntivo mediante secuenciación por el método de Sanger: si se detecta la variante patogénica, se confirma el diagnóstico molecular; si esto no ocurre, se continúa con la secuenciación de otros genes asociados hasta poder hallar -o no- la/s variante/s causante/s de la enfermedad, conllevando altos costos en tiempo y dinero.

El avance de las técnicas de secuenciación masivas, también conocidas como de próxima generación (NGS del inglés *Next Generation Sequencing*) amplía, día a día, nuestra capacidad de realizar este tipo de diagnósticos, pero para que esto sea posible es necesario desarrollar y evaluar -en base a casos reales- los protocolos de biología molecular (asociadas a la preparación de muestras), de secuenciación, y de bioinformática (asociados al procesamiento y análisis de los datos) necesarios para su implementación en la clínica. En este contexto, el análisis del exoma, se ha convertido en una herramienta de gran utilidad en el diagnóstico de enfermedades poco frecuentes con componente hereditario.

En la presente tesis se plantea, de forma global, la implementación de tecnologías de secuenciación masiva para lograr el diagnóstico molecular preciso de entidades poco frecuentes en el ámbito de la genómica clínica. En particular y de forma específica, buscamos desarrollar un protocolo para el análisis e identificación de variantes genéticas, aplicado tanto a casos particulares como a un proyecto nacional, que engloba a un gran número de pacientes con sospecha de Enfermedades Autoinflamatorias.

Desde la investigación científica hasta el diagnóstico clínico mediante técnicas de secuenciación masiva

A partir del conocimiento y tecnologías derivadas del Proyecto Genoma Humano - anunciado el 14 de abril del 2003 por el Instituto Nacional de Investigación del Genoma Humano (NHGRI), el Departamento de Energía (DOE) y sus socios del Consorcio Internacional para la Secuenciación del Genoma Humano- se produjo un punto de quiebre en las ciencias biológicas y, particularmente, en la genética molecular humana [8,9]. A pesar de los continuos avances y grandes logros de la genética que permitieron, por ejemplo, el descubrimiento y secuenciación del gen *BRCA1*, o del gen causante de la fibrosis quística *CFTR* mediante el método convencional de Sanger; sus limitaciones a nivel de costos,

tiempos, rendimiento y resolución quedaron expuestas y restringieron su aplicación a escala masiva, y a una amplia variedad de patologías.

Antes de la llegada de la tecnología de secuenciación de próxima generación (NGS), la genética se centraba en la caracterización de genes únicos asociados con el desarrollo de determinadas patologías monogénicas bien definidas (p. ej. Fibrosis Quística, Enfermedad de Huntington y determinados cánceres hereditarios). El surgimiento de las tecnologías NGS en los inicios del siglo 21 lograron disminuir dramáticamente los costos y aumentar la eficiencia de la secuenciación, propiciando de este modo un cambio de paradigma en la forma en la que se practicaba la medicina, produciendo un salto cualitativo desde el abordaje de terapia tradicional a un enfoque personalizado basado en la constitución genética del paciente, en lo que actualmente se conoce como medicina personalizada [10,11].

En contraste con la secuenciación tradicional de Sanger, en la que un pequeño número de reacciones de secuenciación, de longitud de lectura relativamente largo (700 – 1000 bp) producen una modesta cantidad de datos; las tecnologías de NGS se basan en la ejecución simultánea de millones de reacciones de secuenciación en paralelo, en volúmenes muy pequeños, de longitud de lectura relativamente corta (típicamente 100-150 pb) y en la generación de gigabases (GB) de datos por experimento. Los pasos comunes en la mayoría de los enfoques NGS consisten en (i) fragmentación al azar del ADN a secuenciar, (ii) adición de adaptadores específicos a ambos extremos de los fragmentos (iii) amplificación clonal mediante el anclaje del fragmento de ADN, a través de sus adaptadores, a una superficie sólida como microesferas o placa de secuenciación (e.g. emulsión PCR o *bridge* PCR) (vi) secuenciación (por síntesis o ligación) de los fragmentos y detección simultánea de las bases de forma masiva y paralela (v) adquisición de datos crudos -*raw data*- (por ejemplo, captura de imágenes o detección de iones) y (vi) conversión de los datos crudos en bases de nucleótidos (o sea las secuencias y/o lecturas) [12-16].

Las lecturas (o “*reads*”) de aproximadamente 100 pb obtenidos de una muestra a partir de NGS, pueden ensamblarse mediante procesos informáticos, es decir, unirse para formar una secuencia más larga hasta lograr la cobertura del genoma o de una región determinada (*ensamblado de novo*). En el caso de genomas conocidos -como el humano-, los fragmentos se mapean y alinean contra un genoma de referencia, con el objetivo de determinar las diferencias o variantes intraespecíficas (ver Materiales y Métodos, Figura 5). Usualmente, se secuencian varias veces cada una de las tres mil millones de bases del genoma humano (o las regiones de interés), proporcionando una alta profundidad de lectura (dependiendo de la tecnología seleccionada, se “lee” en un promedio de 100 a 400 veces cada una de las bases) con el fin de obtener datos más precisos.

Teniendo en cuenta que el costo de un experimento de NGS es directamente proporcional a la cantidad de bases que se secuencian, estas tecnologías pueden utilizarse

para secuenciar genomas completos, o puede restringirse a la secuenciación de regiones específicas de interés [17], como las codificantes de los aproximadamente 21.000 genes humanos [18] -en lo que tradicionalmente se conoce como secuenciación de exoma completo- o un pequeño número de genes individuales [19] tradicionalmente llamado panel de genes. Con fines diagnósticos, suele recurrirse con mayor frecuencia a la secuenciación del exoma completo del paciente [20]. De esta forma, resulta mucho más eficaz secuenciar las regiones que sólo representan el 1-2% del genoma, pero que contienen aproximadamente el 85% de las variantes conocidas causantes de enfermedades [21,22]. Por otro lado, la secuenciación selectiva es otro enfoque para obtener la secuencia de unos pocos genes o un área genómica de interés y puede ser asistida con las tecnologías de enriquecimiento de Roche/NimbleGen, RainDance Technologies, entre otras [16,23,24]. Es importante resaltar que, si bien la secuenciación de Sanger sigue siendo considerada el estándar de calidad de secuenciación, resulta ineficiente al momento de secuenciar exomas o genomas completos para detectar variantes genéticas en proporciones menores al heterocigota común [25]. **En conclusión, las tecnologías de NGS resultan ideales para el hallazgo y el diagnóstico de enfermedades genéticas [26].**

Finalmente, se destaca el gran desarrollo y auge que tienen actualmente las tecnologías de secuenciación de tercera generación, que evitan la amplificación de los fragmentos de ADN mediante PCR, reduciendo el tiempo de trabajo y el precio global, mediante la secuenciación de una molécula única (*single-molecule sequencing*) [27]. Entre ellas, se incluyen las tecnologías Pacific Biosciences (PacBio) basadas en métodos de detección por fluorescencia -sin necesidad de amplificación previa- y los nanoporos de Oxford Nanopore [28,29], que permiten secuenciar moléculas únicas sin depender de reactivos de secuenciación, ya que la secuencia es leída a medida que pasa por un poro biológico sin requerir sintetizar una nueva hebra.

NGS como herramienta diagnóstica

Las tecnologías de NGS han tenido un profundo impacto en nuestra comprensión de la genética y biología del genoma [30,31]. Como se mencionó anteriormente, el uso y la implementación de las mismas permitió llevar a cabo secuenciaciones masivas en paralelo, lo que produjo el descenso de los costos, acelerando las investigaciones e incluyendo secuenciaciones *de novo*, de genomas completos, exomas, targets genómicos, perfiles transcriptómicos, análisis metagenómicos, estudios de interacción proteína-DNA (ChIP-seq) y de metilación [30,32]. Durante la última década, se han realizado enormes progresos en términos de velocidad, longitud de lecturas y rendimiento, propiciando el desarrollo de nuevas aplicaciones en ciencias biológicas y genética molecular, así como en áreas de investigación

traslacional: diagnósticos clínicos, agrogenómica y ciencias forenses, entre otras [33,34]. En este contexto, y representando el punto de partida de esta tesis, durante la última década, han sido publicados una serie de estudios que implementan las tecnologías NGS para identificar variantes genéticas subyacentes a algunos casos particularmente difíciles, incluyendo varias enfermedades poco frecuentes. Podemos destacar que los estudios diagnósticos basados en NGS llevaron al descubrimiento de más de 300 genes causantes de enfermedades, anteriormente no reconocidos, y siguen siendo utilizados para investigar diferentes condiciones médicas incluyendo las metabólicas, endocrinas [35-39], oncológicas [40-42], cardiológicas [43], testeos prenatales [44-47], enfermedades infecciosas [48,49], gastrointestinales [50], y las pruebas de histocompatibilidad para trasplante [51,52].

A pesar de que el enfoque de la secuenciación exómica puede perderse ciertas categorías de variantes que son difíciles de detectar, como las variaciones en el número o de copias (CNVs del inglés *Copy Number Variation*), o que son difíciles de reconocer a partir del conocimiento actual (variantes *noncoding* o regulatorias en regiones intergénicas), diversos grupos de investigadores reportaron un porcentaje de éxito del 15 al 35% [53] en la identificación de variantes causales de patologías poco frecuentes, aportando información útil para el consejo genético de los pacientes y sus familias [54].

Existen varias características que convierten a las tecnologías de NGS en una herramienta idónea para ser usada en el diagnóstico clínico de enfermedades genéticas: son simples, rentables -cerca de cinco órdenes de magnitud más económicas comparadas con la secuenciación convencional por Sanger-, menos laboriosas y generan una gran cantidad de datos en tiempos más cortos. De esta manera, permiten analizar un mayor número de genes, siempre dentro de las limitaciones típicas de los diagnósticos médicos. Esto resulta particularmente importante en el caso de varias enfermedades de herencia mendeliana que presentan heterogeneidad de locus (es decir, un mismo fenotipo clínico puede ser causado por variantes en varios -incluso centenas de- genes distintos).

Bioinformática de “nueva generación”

La transición desde la secuenciación tradicional a plataformas con una mayor capacidad de producción de datos, como las de NGS, ha forzado también el desarrollo de nuevos algoritmos bioinformáticos y métodos de análisis necesarios para la correcta manipulación, almacenamiento e interpretación de los mismos. De hecho, varios autores señalan que la utilidad clínica de las tecnologías de NGS se ve más obstaculizada por nuestra limitación a la hora de interpretar los datos (lo que requiere de desarrollo y conocimiento bioinformático), que por nuestra capacidad para generarlos [55,56].

El aumento sostenido en la capacidad de generación de datos ha sobrepasado con creces los recursos bioinformáticos tanto humanos como de equipamiento disponibles hasta la fecha, por lo que el análisis bioinformático se está convirtiendo en un bien escaso, y altamente apreciado. Por otro lado, debido a la tasa de error inherente a la tecnología (0.1-1%, o algo mayor en equipos de tercera generación), a la corta longitud de las lecturas (de 100-150 pb) y, a la mayor profundidad con la que se realiza la secuenciación, el análisis de los datos de NGS, y en especial datos de humanos, es computacionalmente costoso, intensivo y complejo. En este contexto, sin lugar a duda, surge la necesidad de desarrollar herramientas bioinformáticas que permitan aplicar con éxito la secuenciación masiva al diagnóstico genético en la práctica clínica [57].

A diferencia del proceso de generación de datos -el cual es protocolizado y moderadamente estable para cada plataforma-; el análisis de los datos no sigue un modelo "*gold-standard*". A su vez, la combinación de diferentes piezas de *software*, así como su integración con diferentes bases de datos, proporcionan resultados muy diversos. **En el estudio de genomas o exomas humanos, la identificación y caracterización del conjunto de variantes presentes en una muestra es uno de los procesos clave y posiblemente más sensibles, que determina la correcta aplicación de la secuenciación masiva en el diagnóstico genético** [58].

Las lecturas que devuelve el secuenciador son procesadas utilizando un *pipeline* bioinformático que consiste, -a rasgos generales-, en el **control de calidad, mapeo-alineamiento, llamado y anotación de variantes**. Una vez obtenidas las lecturas y **controlada su calidad** -habiendo descartado todas aquellas que no superaron el umbral mínimo de calidad-, es necesario ordenarlas y conocer su ubicación en el genoma humano de referencia, lo que se conoce como **Mapeo**. Una vez que las lecturas fueron mapeadas, estas se **Alinean** a fin de corregir detalles, para finalmente determinar las diferencias (variantes) entre la muestra secuenciada (usualmente del paciente) y el genoma de referencia, en un proceso denominado **Llamado de variantes**. Finalmente, el procedimiento de adición de información a las variantes encontradas recibe el nombre de **Anotación**, en el que toma relevancia la recopilación de información proveniente de bases de datos biológicas, y la posterior asociación de la misma a las variantes encontradas.

El desarrollo y establecimiento de protocolos bioinformáticos de trabajo eficientes, precisos y validados que permitan garantizar estándares de calidad es una tarea larga y compleja, debido tanto a la necesidad de un exhaustivo conocimiento del funcionamiento de las plataformas de NGS, como de la biología del objeto de estudio (por ejemplo, detalles de la enfermedad particular y su fisiopatología molecular). En esta tesis, nos propusimos como objetivo, avanzar en una solución a este problema, para progresar hacia un diagnóstico más preciso y un pronóstico más temprano de las enfermedades poco frecuentes (EPoF) [59-62].

Genómica poblacional (El concepto de variante)

El genoma de un individuo puede representarse como una cadena de letras; donde la mayoría de las posiciones se conservan en todos los miembros de la misma especie mientras que, en algunos sitios, una minoría presenta una letra diferente. La mínima diferencia de una sola letra se conoce como una variante de nucleótido único o SNV (*Single Nucleotide Variant*, por sus siglas en inglés). Las SNVs que caen en regiones codificantes y que poseen mayor probabilidad de influir en una enfermedad, son aquellos que alteran la secuencia aminoacídica de la proteína codificada y se encuentran presentes usualmente a una tasa más baja, presentando frecuencias alélicas menores que las sustituciones sinónimas (en general, silenciosas); evidenciando la existencia de una selección que actúa en contra de los alelos perjudiciales o deletéreos a lo largo de la evolución humana.

Desde la publicación del genoma humano en 2003, los genetistas han estado interesados también en determinar la variabilidad de los genomas en la población. El Consorcio Internacional HapMap (“*International HapMap Consortium*”, en inglés) fue el primero en abordar este tema, y se esforzó por catalogar las variantes más comunes en nuestro genoma. Principalmente, reportaron y catalogaron “todas” aquellas variantes con una frecuencia de alelos minoritarios (MAF) de, al menos, 5% en uno o más grupos étnicos. Para el 2008, el catálogo del proyecto HapMap contenía 3,5 millones de variantes (usualmente llamadas polimorfismos). Sin embargo, para ampliar la información disponible de variantes asociadas a patologías e incluir aquellas de baja frecuencia poblacional, fue necesaria la utilización de tecnologías de secuenciación masiva, lo que permitió identificar variantes a partir de los estudios de asociación, mejorar el conocimiento disponible para los investigadores interesados en la biología evolutiva, y sentar las bases para predecir la predisposición a enfermedades. Más recientemente, el consorcio 1000 Genomas (*1000 Genomes*) se materializó para abordar estas necesidades, al proporcionar principalmente una base de datos de referencia de secuencias y la caracterización profunda de las variaciones del genoma humano como base para el estudio de la relación fenotipo-genotipo, a partir de la secuenciación del genoma completo (del inglés, *Whole Genome Sequencing*) a un conjunto diverso de individuos de múltiples poblaciones [63]. El proyecto finalizó en el 2015, tras la reconstrucción de los genomas de unos 2.500 individuos de 26 poblaciones (o etnias) diferentes -mediante una combinación de secuenciación del genoma completo de baja cobertura, secuenciación del exoma, genotipificación de microarreglos y la caracterización de más de 88 millones de variantes (84.7 millones de polimorfismos de un solo nucleótido (SNP), 3.6 millones de inserciones/delecciones cortas (Indels) y 60.000 variantes estructurales).

Unos años después, surgieron diferentes agrupaciones de investigadores que analizaron, reorganizaron, agregaron y armonizaron los datos obtenidos a partir de

numerosos proyectos de secuenciación exómica a gran escala, partiendo de datos de miles de genomas y exomas completos, haciéndolos disponibles a la comunidad científica. De esta forma, grandes conjuntos de datos de referencia, como los del Consorcio del Proyecto 1000 Genomes [63], el Proyecto de Secuenciación del Exoma (ESP) [64], el Consorcio de Agregación del Exoma (ExAC) [65] y la Base de datos de agregación del genoma (GnomAD) [66] se han hecho públicos en beneficio de la comunidad biomédica. Estos conjuntos de datos aportan información de enorme relevancia con numerosas aplicaciones, tanto para la investigación clínica como para la básica. En particular, dado que la frecuencia de una variante es uno de los mejores indicadores de su potencial carácter deletéreo, los genetistas clínicos utilizan los mismos para distinguir las variantes patogénicas de los polimorfismos benignos. Además, se basan en los datos de frecuencia de variantes y alelos para inferir la función de los genes y variantes (por ejemplo, si un gen es esencial), así como para el análisis de genética de poblaciones. Estos proyectos a gran escala publican los datos en bruto en forma de archivos VCF (formato resultante del llamado de variantes), que requieren conocimientos de bioinformática para analizarlos y crearlos. Sin embargo, están disponibles navegadores del genoma que han permitido visualizar grandes conjuntos de datos a escala genómica (por ejemplo, USCS, ExAC y GnomAD), llevar a cabo análisis a nivel de genes y evaluar datos de las variantes de un solo nucleótido, inserciones y delecciones y variantes del número de copias. Por supuesto, la información detallada sobre cada variante, incluidas las anotaciones y las métricas de calidad, es de suma importancia. Adicionalmente, una visualización igualmente relevante es la de la ausencia de variación, es decir, si una variante ausente implica una falta de variación observada, una cobertura baja o nula en la región genómica, o una variación filtrada debido a la mala calidad [67]. La existencia de estos navegadores posibilita el aporte de información de relevancia que constantemente facilita el proceso de filtrado y priorización de variantes.

Frecuencias Alélicas

Se puede definir a los alelos como cada una de las formas alternativas de una posición o segmento de secuencia de ADN en un lugar particular de un cromosoma, que se puede manifestar en modificaciones concretas de la función del gen (también pueden considerarse marcadores moleculares). Relacionando conceptos, es claro que una variante es necesariamente un alelo alternativo, siendo la secuencia del genoma de referencia el “alelo de referencia”. En cierto sentido, variante y alelo son sinónimos. Lo importante es que el concepto de alelo tiene una larga tradición de uso en genética clásica y en genómica de poblaciones. Cuando el mismo alelo es observado en diversos individuos de una población,

se puede calcular su frecuencia. El alelo de mayor frecuencia suele ser la referencia, y por ello nos referimos al alelo de menor aparición, como al alternativo.

La población humana tiene en término medio una diversidad genética relativamente limitada, lo que refleja su corta edad evolutiva e histórico tamaño pequeño. Existen muchas variantes genéticas “raras” o infrecuentes en la población humana, pero la mayor parte de la heterogeneidad poblacional es atribuible a los alelos comunes (es decir, aquellos que están presentes con una frecuencia mayor al 1% en la población general, usualmente denominados polimorfismos). Las variantes poco frecuentes (es decir, aquellas que están presentes con una frecuencia menor al 1-0.5% en la población general) constituyen, en muchos casos, las causas primarias de las enfermedades genéticas de herencia mendeliana y generalmente son de origen “reciente” y altamente penetrantes. Cuantificar los efectos en el *fitness* - representación cuantitativa del éxito reproductivo individual- subyacentes a la aparición de nuevas variantes en el genoma humano resulta clave para resolver debates importantes en la genética médica y evolutiva [68]. Diferentes factores influyen en las frecuencias alélicas de una población. Uno de ellos es la selección natural: si una variante es deletérea o perjudicial, la selección natural evitará que se transmita de generación en generación o, en el caso de ser beneficiosa, probablemente aumente su frecuencia. También influye la migración de personas, trasladando su acervo genético de un lugar a otro y la deriva génica. Sin lugar a duda, la contribución de las variantes raras a la arquitectura genética de las enfermedades humanas y los rasgos complejos es una cuestión de interés fundamental, que puede arrojar luz sobre la acción de la selección negativa [69,70].

Frecuencias alélicas y la magnitud del efecto

En líneas generales, se pueden considerar dos propiedades de las variantes en relación con su impacto en la salud [71]. La primera, se basa en las frecuencias alélicas descritas en el párrafo anterior, que se muestran en el eje x de la Figura 1. El rango abarca desde variantes comunes (definidas como mayor o igual al 5% de la frecuencia del alelo minoritario (MAF¹)), de baja frecuencia (<5% MAF), raras (<0,5% MAF), o muy poco frecuentes (<0,1% MAF). La segunda, es la magnitud del efecto que producen (*effect size*, eje Y de la Figura 1), que suele ser cuantificado como un *odds ratio* (OR; [72]). El OR es la probabilidad de que ocurra un suceso dividido por la probabilidad de que no ocurra, comparando personas con el alelo de riesgo frente a personas sin él y constituye una medida de la asociación entre la aparición de, en este caso, una variante genética y el resultado que

1 MAF (por sus siglas en *inglés*, *Minor Allele Frequency*), se define como la frecuencia del *alelo* menos frecuente en un determinado *locus*, dentro de una población. Es un parámetro utilizado para estudiar la *variación genética*, ya que proporciona información que permite diferenciar entre variantes *frecuentes* y *raras* en la población, resultando un dato fundamental en estudios de *genética poblacional*.

ésta produce -por ejemplo, el desarrollo de una enfermedad-. Un OR igual a 1 implica que la presencia de la variante no afecta las probabilidades de desarrollar la enfermedad, mientras que un OR mayor a 1 conlleva una mayor probabilidad de asociación entre una variante y la aparición de una determinada patología. Un OR de 2 implica que el desarrollo de la patología es 2 veces más probable en presencia de la variante que en su ausencia. Por el contrario, un OR menor a 1 se relaciona con una menor probabilidad de asociación entre una variante y una dada patología.

Como se mencionó anteriormente, el principal objetivo de la genómica humana es identificar variantes causantes de patologías (o de riesgo), es por esto que podemos centrarnos en la región entre las líneas punteadas de la Figura 1, alelos de muy baja frecuencia -o raros- con una alta magnitud de efecto (*effect size*), superior a 10, tienden a causar enfermedades monogénicas de herencia mendeliana (como se puede observar en el margen izquierdo de la figura mencionada). Alelos de baja frecuencia tienden a tener efectos menos perjudiciales (centro de la figura) mientras que algunos alelos frecuentes -o comunes, identificados mediante estudios de asociación genómica (GWAS)- tienen un efecto (u OR) bajo, cercano a 1, pero aún así contribuyen al desarrollo de enfermedades comunes (como se puede observar en el margen inferior derecho de la figura). Hay pocos ejemplos de variantes frecuentes que tienen una gran contribución a patologías comunes (margen superior derecho), mientras que variantes muy infrecuentes con bajo efecto son muy difíciles de identificar. El espectro de la variación genética es mayor de lo que se preveía en un principio y sugiere que la arquitectura genómica subyacente de muchas enfermedades y rasgos podría ser más compleja. La disponibilidad de chips de genotipado de nueva generación y de tecnologías de secuenciación ha suscitado la esperanza de que los futuros estudios de variantes poco comunes y raras puedan aumentar la tasa de descubrimiento y explicar una fracción adicional de la heredabilidad, y eventualmente acercarse a un rendimiento discriminatorio clínicamente útil para los modelos de riesgo genético [73].

En la presente tesis, nos enfocamos en variantes de baja (o muy baja) frecuencia que son, como se mencionó anteriormente, aquellas con mayor probabilidad de ser las causantes de enfermedades mendelianas.

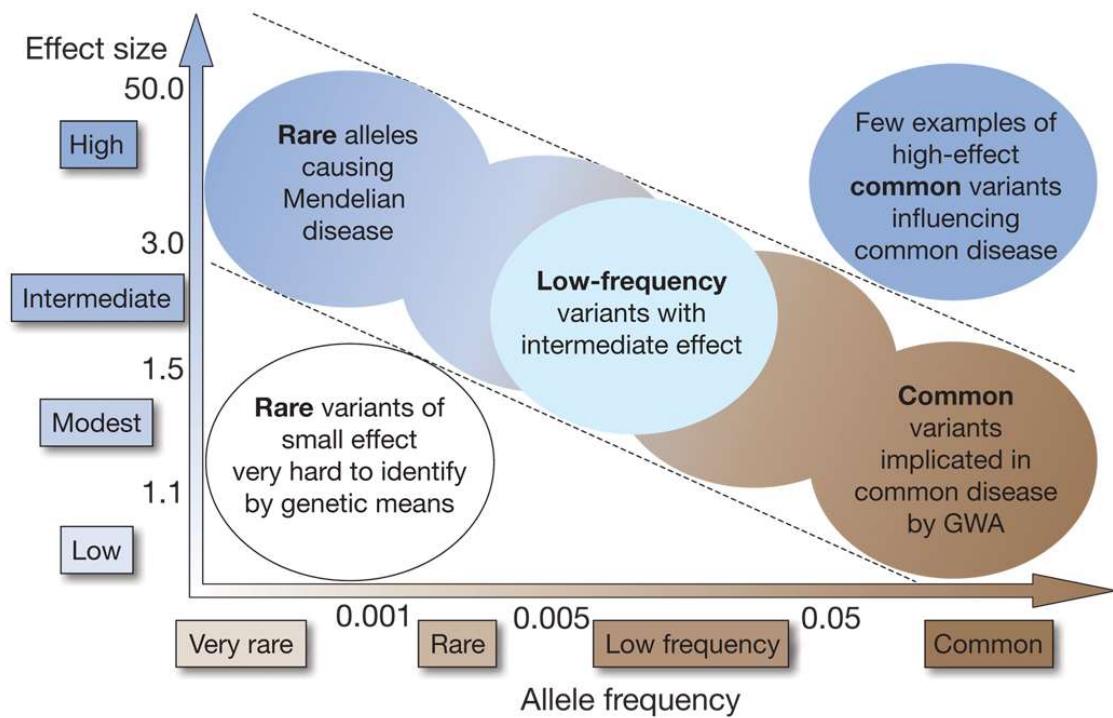


Figura 1. Frecuencia alélica y magnitud del efecto. Extraído de Manolio et al., 2009 [71].

Fuerza o magnitud de un efecto genético (odds ratio, eje Y) en función de las frecuencias alélicas (eje X). El mayor énfasis está puesto en la región contenida entre las líneas punteadas, donde la magnitud del efecto (effect size) puede ser High: alta, Intermediate: intermedia, Modest: moderada, Low: baja y las frecuencias alélicas Very rare: muy poco frecuentes, Rare: poco frecuentes, Low frequency: baja frecuencia, Common: polimorfismos, alta frecuencia.

Bases moleculares de las enfermedades genéticas

Los organismos presentan una tendencia inherente a “sufrir” variantes que tradicionalmente se categorizan en cromosómicas o genéticas, dependiendo del nivel al que ocurran. Las variantes cromosómicas afectan a cromosomas enteros, a segmentos de los mismos o incluso a dotaciones cromosómicas completas. Estos cambios no van acompañados necesariamente de variantes génicas y los efectos de estas se deben, con mayor frecuencia, a una nueva disposición de los cromosomas y de los genes que éstos contienen [74]. Por otro lado, las variantes genéticas usualmente corresponden a variantes puntuales o pequeñas delecciones e inserciones. En promedio, cada persona posee entre **3 y 5 millones de variantes** en su genoma, la mayoría de las cuales se encuentran localizadas en regiones intergénicas, repetitivas y no codificantes y, por lo general, no tienen ningún efecto directo en la salud del individuo. Sin embargo, las variantes localizadas dentro de un gen o una región regulatoria suelen estar asociadas a un mayor impacto, pudiendo afectar la función y/o expresión del gen y del ARN/proteína resultante.

El término “enfermedad molecular”, introducido seis décadas atrás, refiere a aquellas enfermedades en donde la principal causa es la alteración -heredada o adquirida- de uno o varios genes, viéndose afectada su función y/o expresión [75].

A pesar de que la mayoría de las enfermedades monogénicas conocidas son producto de variantes en la región codificante del ADN; en las últimas décadas se descubrieron algunas excepciones: enfermedades causadas por variantes en los genes de los ARN no codificante (ncRNA), en los microARNs (miRNAs) -que regulan targets específicos- y en genes mitocondriales que codifican para ARN de transferencia (tRNAs). Resulta esencial entender las enfermedades genéticas a nivel molecular y bioquímico, porque este conocimiento es la base para el desarrollo de una terapia racional.

En esta tesis el centro de atención estará puesto sobre enfermedades monogénicas causadas principalmente por defectos o alteraciones en la región codificante de los genes de interés para cada patología estudiada.

Distintos tipos de variantes y sus efectos moleculares

Tal como lo establece el dogma de la biología molecular, para obtener una proteína biológicamente activa, la información contenida en el ADN debe ser transcripta a un mensajero (mRNA), que es procesado hasta su madurez, para luego traducirse en un polipéptido, que atravesará distintos procesos biológicos hasta llegar a convertirse en una proteína funcional en el lugar adecuado.

Cualquiera de las variantes mencionadas a continuación puede interrumpir alguna de estas etapas, dando lugar a cambios en la cantidad, calidad o naturaleza de la proteína resultante.

Las variantes se clasifican, en primera instancia, de acuerdo a su naturaleza en relación con su localización en regiones funcionales del genoma, y de acuerdo al efecto molecular que producen.

Las sustituciones se pueden dividir en:

- Sinónimas o silenciosas, son aquellas que involucran un cambio de nucleótido en un codón, sin conllevar cambios en el aminoácido codificado. Por esta razón, no suelen tener un efecto **aparente** en la proteína resultante.
- No sinónimas (*missense*), son aquellas en las que un cambio de nucleótido conduce a una sustitución aminoacídica, pudiendo (o no) provocar cambios estructurales y/o funcionales en la proteína resultante y su consiguiente alteración o pérdida de función.

Adicionalmente, la **pérdida de función de la proteína mutada** suele ser un efecto asociado a los siguientes tipos de variantes:

- Desplazamiento del marco de lectura (*frameshift*).
- Ganancia de un codón *stop* prematuro (*stop gained*)
- Pérdida de un codón *stop* (*stop-loss*) ó delección del codón de inicio de la traducción (*start-loss*).

La severidad de una enfermedad debida a la pérdida de función de una proteína en particular, se relaciona usualmente con la cantidad de la función perdida; por lo que, en aquellos casos donde se conserva un pequeño porcentaje residual de la función, se ven reducidas las consecuencias fenotípicas deletéreas [76].

Otro efecto de una variante missense, puede ser el de la **ganancia de función** -la adquisición de una propiedad nueva-, o la **expresión del gen en un momento o lugar erróneo** -expresión heterocrónica o ectópica, respectivamente-. Las variantes de ganancia de función se clasifican en dos tipos diferentes: las que incrementan la producción de la proteína normal (aumento del dosage génico que resulta, en general, de la duplicación de una parte del cromosoma), o las que mejoran (o cambian) la función normal de la misma (suelen tener un efecto negativo en el rol fisiológico de la proteína). En algunas enfermedades, un cambio en la secuencia aminoacídica de una proteína le confiere una nueva propiedad, sin afectar necesariamente su función normal. Estos tipos de variantes son infrecuentes ya que la mayoría de las sustituciones aminoacídicas resultan ser neutrales o afectan negativamente la función o estabilidad de la proteína en cuestión.

La última clase importante de variantes incluye a aquellas que llevan a una expresión inapropiada de un gen en un momento, o lugar determinado. Las mismas ocurren

generalmente en la región regulatoria de los genes, siendo un ejemplo ampliamente estudiado en diferentes tipos de cáncer el caso de la expresión anormal de un oncogen, que resulta de la mutación de un protooncogen, y promueve la proliferación celular.

Adicionalmente, la expresión de un gen también puede verse afectada por la presencia de variantes intrónicas (*intron variant*), pudiendo afectar el proceso de *splicing alternativo*. Las variantes que se encuentran en regiones río arriba/abajo del gen, en regiones no traducibles, se denominan variantes 5'UTR o 3'UTR, respectivamente. Las mismas pueden modificar la interacción de las regiones UTR con elementos reguladores, proteínas y microARN. Las consecuencias funcionales pueden afectar la modulación de la transcripción del ARNm, su estructura secundaria, estabilidad, localización, traducción o el acceso de reguladores como los microARNs y las proteínas de unión al ARN (RBPs). Se sabe que las alteraciones de los mecanismos reguladores podrían modificar vías moleculares y procesos celulares, pudiendo dar lugar también a procesos patológicos.

Además de conocer la ubicación de una variante y su efecto a nivel molecular, **nuestro principal objetivo es determinar qué efecto produce sobre el fenotipo del paciente**, es decir, que consecuencias fisiopatológicas tiene o podría tener. Predecir el efecto (sobre todo a nivel organismo) de una variante resulta complejo debido principalmente a la heterogeneidad de los mecanismos moleculares que subyacen a las enfermedades. Es importante destacar que el impacto de una determinada variante depende principalmente de su contexto genómico y celular. En resumen, se puede analizar el efecto de una variante desde dos perspectivas: en primer lugar, analizando la naturaleza de la variante (posición y efecto molecular), y posteriormente realizando un análisis extensivo, investigando su entorno y rol biológico.

De forma independiente, se puede evaluar la frecuencia alélica, la conservación evolutiva, su potencial efecto en la estructura de la proteína, propiedades fisicoquímicas de los aminoácidos involucrados en el cambio, y las características funcionales de la misma a partir de estudios funcionales *in vitro* o *in vivo*.

Desde el punto de vista del organismo y en el contexto de una posible enfermedad se puede analizar también el modo de herencia para corroborar si el genotipo de la variante se corresponde con el esperado, -recomendando también la evaluación de la segregación de la variante en los familiares sanos y afectados-, buscar asociaciones previas en la literatura con el diagnóstico presuntivo y estudiar las vías metabólicas y las relaciones génicas/proteicas involucradas en la patología en estudio.

La/s variante/s candidata/s son aquellas que reúnen la evidencia suficiente para ser consideradas como tales y que aportan la información necesaria para orientar al médico a cargo hacia un diagnóstico molecular preciso, permitiéndole explicar -total o parcialmente- el fenotipo del paciente.

Clasificación de variantes

En la última década, las tecnologías de secuenciación masiva han permitido ampliar de manera significativa la realización de pruebas genéticas que abarcan desde genes individuales, paneles genéticos, exomas, genomas hasta transcriptomas, facilitando el surgimiento de nuevos desafíos en la interpretación de los datos obtenidos. Por esta razón, la determinación precisa de la patogenicidad de una variante (hecho necesario para un diagnóstico certero) es una de las tareas más complejas y relevantes. En el 2013, el Colegio Americano de Genética Médica y Genómica (ACMG) convocó a un grupo de trabajo compuesto por representantes del ACMG, la Asociación de Patología Molecular (AMP), y el Colegio de Patólogos Americanos para revisar las pautas vigentes para la interpretación de variantes y publicaron lo que se conoce usualmente como “las recomendaciones de la ACMG/AMP” [77], un conjunto sistemático de reglas para valorar y combinar la evidencia de cada variante, que permite clasificarla de acuerdo a su patogenicidad. Las recomendaciones se basan en el análisis y asignación de 28 tipos de evidencia asociados a, entre otras características, frecuencia alélica, análisis funcionales, predicciones *in-silico*, análisis de segregación familiar, reportes previos en bases de datos curadas, relación genotipo-fenotipo, que son valorados con un código alfanumérico, y luego son combinados para clasificar a cada variante en una de las 5 categorías siguientes: patogénica (*Pathogenic*, P), probablemente patogénica (*Likely Pathogenic*, LP), de significado incierto (*Variant of Uncertain Significance*, VUS), probablemente benigna (*Likely Benign*, LB), ó Benigna (Benign, B). Comprender en profundidad estos criterios y saber asignar cada uno de los 28 códigos en base a la evidencia que presenta cada variante, requiere el conocimiento y el manejo de varios campos disciplinarios (medicina, genética, bioquímica y bioinformática). Recientemente, surgieron las guías de ClinGen [78] (del inglés, *Clinical Genome Resource*), un proyecto financiado por los Institutos Nacionales de Salud (NIH) y dedicado a crear un recurso central que defina la relevancia clínica de genes y variantes para su uso en medicina de precisión e investigación. Su implementación permitió refinar los criterios previamente establecidos por el ACMG y ponderarlos.

Resulta importante destacar que, durante los últimos años, para ayudar a la interpretación y clasificación de las variantes, se han desarrollado numerosos programas y bases de datos asociadas, que proveen una valoración (semi)-automática de cada uno de los niveles de evidencia, y luego, en función de la misma, realizan la asignación del criterio de patogenicidad correspondiente y otorgan como resultado la clasificación tentativa de la variante.

Hallazgo de variantes de significado incierto (VUS)

En la genómica clínica, cuando lo que se busca es un diagnóstico molecular preciso que explique la sintomatología y/o causa de la enfermedad de un paciente, muchas veces nos enfrentamos con la realidad de que el resultado del análisis genómico, es la presencia de una (o más) variantes de significado incierto.

La mayoría de las variantes pueden clasificarse fácilmente como benignas o probablemente benignas según la frecuencia poblacional de las mismas, y/o evidencia contundente de benignidad en bases de datos curadas. Sin embargo, cuando se encuentran variantes novedosas (sin datos de frecuencia), usualmente se realiza un trabajo de búsqueda de evidencia de su potencial patogenicidad, utilizando los criterios descritos anteriormente. En general, la mayoría de las búsquedas no proporcionan datos o información suficiente para determinar la relevancia clínica de una variante nueva y, por lo tanto, poco frecuente. Como usualmente la búsqueda del diagnóstico molecular busca la elección precisa de un abordaje terapéutico, y/o al menos una confirmación (o corrección) del diagnóstico clínico presuntivo - que ponga fin a lo que muchas veces es una odisea de visitas médicas y estudios- la obtención como resultado de una VUS deja, tanto a médicos como a pacientes y sus familiares, con un sinsabor. Es por ello que gran parte del esfuerzo de los investigadores en genómica consiste en determinar si las VUS serán benignas o patogénicas.

De acuerdo con los criterios de ACMG/AMP, una variante es clasificada como VUS cuando:

- a) La combinación de valoraciones de la evidencia de la misma no alcanzan para clasificarla como Posiblemente Patogénica o posiblemente Benigna; ó
- b) La variante presenta evidencia contradictoria.

De los 4 grupos de niveles de evidencia que puede poseer cada variante, la frecuencia poblacional siempre suele aportar información importante, ya sea porque la variante no existe en las bases de datos poblaciones, o porque posee muy baja frecuencia poblacional. Es en el análisis del efecto molecular, donde el tipo de variante (por ejemplo, *missense*), y/o la posición en el gen/proteína, y/o la ausencia de ensayos funcionales, y/o la inconsistencia de las predicciones bioinformáticas, pueden fallar en aportar suficiente evidencia para clasificar a la variante como probablemente patogénica. La clave en estos casos suele estar asociada al análisis de la información provista por las bases de datos clínicas en relación con el fenotipo del paciente. Si el fenotipo del paciente (y/o la historia familiar) son altamente específicos y coinciden con aquellos observados en pacientes previamente reportados con variantes en el gen donde se halla la variante, es importante evaluar reportar la variante y especificar que, si bien la misma formalmente se clasifica como VUS, la evidencia circunstancial sugiere su posible patogenicidad y, por lo tanto, su capacidad de precisar el diagnóstico clínico.

En este contexto si bien el profesional no contará con un diagnóstico preciso, tendrá una hipótesis de trabajo que, en función de, por ejemplo, el estudio posterior de la segregación de la variante en el grupo familiar y/o la realización propia (o por parte de terceros) de estudios funcionales, y/o la aparición de la variante como candidata en pacientes con la misma sintomatología, pueda convertir a la variante en (probablemente) patogénica y/o permitan descartarla por benigna. Este proceso de reevaluación posterior es indicado/sugerido en el informe y debe realizarse un seguimiento adecuado del mismo. **En todos los casos analizados y comentados en la presente tesis, se sugirió -y efectuó en algunos- la validación de la segregación de la/s variante/s reportada/s en los integrantes disponibles del grupo familiar.**

Relación fenotipo-genotipo en las enfermedades genéticas

Encontrar la relación entre el fenotipo y el genotipo fue y será el principal objetivo de la genética. En el contexto de la genómica clínica, sobre todo cuando se analizan variantes en muchos genes, la clave está en estudiar cuál de todos los genes que potencialmente se ven alterados por la presencia de las variantes encontradas, dará lugar a un fenotipo que se solape con el observado en el paciente [79].

La complejidad de la relación entre el fenotipo observado y el genotipo posee varios orígenes. Por ejemplo, en algunas enfermedades genéticas, una variante concreta no siempre produce un fenotipo aberrante en todos los individuos que la portan. Este concepto se denomina penetrancia de la mutación o variante. En otros casos, los individuos portadores de la misma variante muestran una gama de fenotipos que varían en su gravedad. La expresividad es el grado en que la expresión de un rasgo difiere entre individuos. A diferencia de la penetrancia, la expresividad describe la variabilidad individual, no la variabilidad estadística entre una población de genotipos. En muchos casos, tanto la penetrancia fenotípica como la expresividad de una mutación varían debido a las diferentes combinaciones de alelos modificadores que están presentes en un *background* genético - en comparación con otro- [80]. Estos dos factores sumados a la heterogeneidad fenotípica - asociación de más de un fenotipo con variantes en un único gen- contribuyen a la heterogeneidad total reflejada en una amplia gama de fenotipos clínicos observados para una condición médica determinada [81].

Adicionalmente, la secuenciación del exoma completo ha revelado un alto nivel de heterogeneidad alélica -presencia de variantes diferentes en un locus génico específico que producen la misma expresión fenotípica de una enfermedad o afección, o una expresión fenotípica similar- y heterogeneidad genética -presencia de variantes en distintos genes que

causan un mismo fenotipo clínico o afección- asociada a las enfermedades mendelianas simples [82].

El objetivo principal del Proyecto del Genoma Humano (HGP), y de los posteriores proyectos de secuenciación de genomas y exomas humanos, fue mejorar la comprensión de las enfermedades humanas, permitiendo identificar y caracterizar las consecuencias de variantes germinales en los más de 20.000 genes humanos. Las variantes asociadas a patologías relacionan directamente a las variantes presentes en el ADN con la función o dosificación de la proteína alterada y, finalmente, con los fenotipos observados; transformando así la comprensión de la biología básica del desarrollo y la homeostasis involucrada en la regulación de la transición salud-enfermedad. De hecho, mucho de lo que se conoce acerca de la relación entre la función de los genes y los fenotipos observados se basa en el estudio de variantes poco frecuentes que causan fenotipos específicos como los patológicos y que poseen herencia de tipo mendeliana [82].

Es importante tener en cuenta que, partiendo de un diagnóstico clínico tentativo, se deben evaluar todos los genes potencialmente causales del fenotipo observado, y que el número de genes involucrados en una patología particular varía dependiendo de la cantidad de proteínas diferentes involucradas en el proceso fisiológico afectado.

En el contexto de la relación genotipo-fenotipo, y como ya se mencionó, es importante destacar que a la hora de decidir si una variante está realmente relacionada con el diagnóstico clínico aportado por el profesional, resulta clave estudiar si existen variantes previamente reportadas que afecten al mismo gen y produzcan fenotipos clínicos semejantes al observado. Es por ello que cobra una enorme importancia la existencia de una base de datos biológica exhaustiva, autorizada y curada que contenga información sobre todos los genes humanos conocidos y las enfermedades genéticas asociadas a los mismos, incluyendo bibliografía citada, que vincule variantes patogénicas en un determinado gen con el desarrollo de una enfermedad particular. Fue así como el Dr. Victor A. McKusick de la Universidad Johns Hopkins creó, a partir de los libros -publicados desde 1966- sobre Herencia Mendeliana en el hombre (MIM, por sus siglas en inglés *Mendelian Inheritance in Man*) el catálogo versión online de uso público OMIM [83] con el objetivo de incentivar la investigación y apoyar la educación en genética humana.

Finalmente, es importante resaltar que el descubrimiento de la base genética de un fenotipo mendeliano establece un vínculo causal entre el genotipo y el fenotipo, haciendo posible la detección de portadores sanos, el screening poblacional y el diagnóstico directo. Dichos descubrimientos contribuyen también al conocimiento de la función de los genes, la regulación genética, y los mecanismos biológicos que pueden llegar a utilizarse para el desarrollo de nuevas terapias.

Enfermedades poco frecuentes

De acuerdo con la Federación Argentina de Enfermedades Poco Frecuentes (FADEPOF) [4], las enfermedades poco frecuentes (EPOF) -o también llamadas "enfermedades raras"- son condiciones de salud con patrones específicos de signos, síntomas y hallazgos clínicos definidos, que afectan a un escaso número de personas en una población determinada y pueden variar según el país y/o región.

Basados en los datos epidemiológicos disponibles en Orphanet [84], se ha determinado que en el mundo hay descritas clínicamente al menos 8.000 enfermedades poco frecuentes definidas por prevalencia puntual (sin contemplar cánceres, enfermedades infecciosas o intoxicaciones bacterianas o virales poco frecuentes). Por esta razón, se reconoce que el número total es considerablemente mayor, pudiendo llegar a más de 10.000 EPOF [85]. Estas enfermedades afectan entre el 4 y el 8% de la población mundial [86] e incluso, publicaciones recientes refieren que podrían afectar hasta al 10% de la población mundial, estimando que más de 300 millones de personas en el mundo viven con alguna enfermedad poco frecuente. Suelen constituir enfermedades crónicas, complejas, progresivas, incapacitantes y, en ciertos casos, potencialmente mortales. Siete de cada diez son de origen genético, y de estas, alrededor del 70% se manifiesta al nacer o durante la niñez, resultando en la muerte de 3 de cada 10 niños antes de cumplir los 5 años [86,87].

Las personas que viven con una EPOF requieren de una atención biopsicosocial, que contemple tanto la asistencia clínica especializada -en atención primaria y/o de alta complejidad- como de servicios sociales y apoyo psicológico destinado al propio paciente como a su grupo familiar, bajo una atención integral y coordinada. La dificultad para localizar a un médico especialista, sumado a la dispersión geográfica hace que las personas con EPOF y sus familiares deban trasladarse a grandes distancias y pasar largas estadías fuera del lugar de residencia para la realización de estudios, controles y seguimiento de su enfermedad, ocasionando pérdida de escolarización y/o productividad laboral [88].

Llegar a un diagnóstico certero es el mayor desafío: en promedio, puede llevar entre 5 y 10 años y requerir hasta 8 visitas a distintos especialistas antes de obtenerlo. Además, se estima que más de 4 de cada 10 personas reciben al menos un diagnóstico errado durante el proceso [89]. Debido a la baja prevalencia de estas enfermedades, la experiencia médica y conocimientos generales son escasos, la oferta asistencial inadecuada y la investigación limitada. A pesar de su gran número global, los pacientes con enfermedades poco frecuentes suelen ser huérfanos de los sistemas sanitarios, a los que a menudo se les niega el diagnóstico, el tratamiento y los beneficios de la investigación. **Es por esta razón que buscamos incluir en esta tesis a pacientes con fuerte sospecha de EPOF y bajas**

posibilidades de costear estudios genéticos, de forma de facilitarles el acceso a un potencial diagnóstico molecular certero.

Enfermedades autoinflamatorias

Un ejemplo de enfermedades poco frecuentes lo constituyen las enfermedades autoinflamatorias (EAI), caracterizadas principalmente por crisis inflamatorias generalizadas recurrentes o persistentes, que frecuentemente incluyen fiebre, erupciones cutáneas, compromiso de serosas o artritis y elevación de reactantes de fase aguda (RFA), pero se debe tener en cuenta una gran variabilidad que puede contemplar vasculopatías, enteropatías y alteraciones neurológicas. Se definen por la continua sobreproducción y liberación de mediadores proinflamatorios, especialmente interleucina-1-beta (IL-1 β), y sobreactivación de células del sistema inmune innato, como consecuencia de la desregulación del control del proceso inflamatorio, generalmente sin asociarse con una respuesta celular T con antígenos definidos, ni a altos títulos de anticuerpos [90,91]. Pueden simular también patologías autoinmunes, infecciosas y neoplásicas, por lo que es necesario descartar cualquiera de estas etiologías. Debido a su origen genético, suelen iniciarse en la niñez. Durante los intervalos asintomáticos, se alcanza el bienestar clínico y la normalización de los parámetros inflamatorios, aunque la amiloidosis representa una grave complicación a largo plazo [92]. El diagnóstico es clínico y se basa en las características fenotípicas. El diagnóstico genético es muy importante, pero debe ser realizado e interpretado correctamente. El tratamiento con agentes biológicos que bloquean citocinas proinflamatorias, particularmente IL-1 β , ha demostrado ser efectivo en muchos pacientes [92,93].

Biología molecular de las enfermedades autoinflamatorias

La larga batalla entre los seres humanos y diversos insultos físicos, químicos y biológicos que causan daño tisular o lesiones celulares (por ejemplo, productos de daño tisular, metabolitos y/o infecciones) ha llevado a la evolución de diversas respuestas adaptativas. Estas respuestas se desencadenan por el reconocimiento de patrones moleculares asociados a daños (DAMPs) y/o patrones moleculares asociados a patógenos (PAMPs), generalmente por parte de las células del sistema inmune innato [94]. Los DAMPs y PAMPs son reconocidos por los receptores de reconocimiento de patrones (PRRs) expresados por las células inmunes innatas; este reconocimiento desencadena la inflamación. Los PRR incluyen los receptores tipo Toll (TLR), los receptores tipo Nod (NLR), los receptores de lectina tipo C (CLR) y los receptores tipo RIG-I (RLR) que reconocen los DAMP y los PAMP para iniciar las respuestas inmunitarias [95,96]. Estas proteínas también se denominan receptores inmunitarios innatos. Las variantes de ganancia de función de los receptores inmunitarios innatos o las variantes de pérdida de función de sus inhibidores están

relacionadas con las enfermedades autoinflamatorias [97]. Las mismas están fuertemente asociadas a la desregulación de los interactomas de los PRR, entre los que se encuentran los inflamasomas, los signalosomas activadores de NF-κB, los signalosomas inductores de interferón de tipo I y los inmunoproteasomas; dando lugar las alteraciones de la regulación de estos interactomas a las Inflamasomopatías, Relopatías, Interferonopatías y Síndromes autoinflamatorios asociados a los proteasomas, respectivamente [98]. Esto explica la patogénesis de las enfermedades autoinflamatorias que implican brotes inflamatorios recurrentes en ausencia de autoanticuerpos o linfocitos T antígeno-específicos [99]. El conocimiento de los mecanismos moleculares que subyacen a las funciones de estos receptores inmunitarios innatos es útil para el tratamiento y la gestión de los individuos con enfermedades autoinflamatorias (Figura 2).

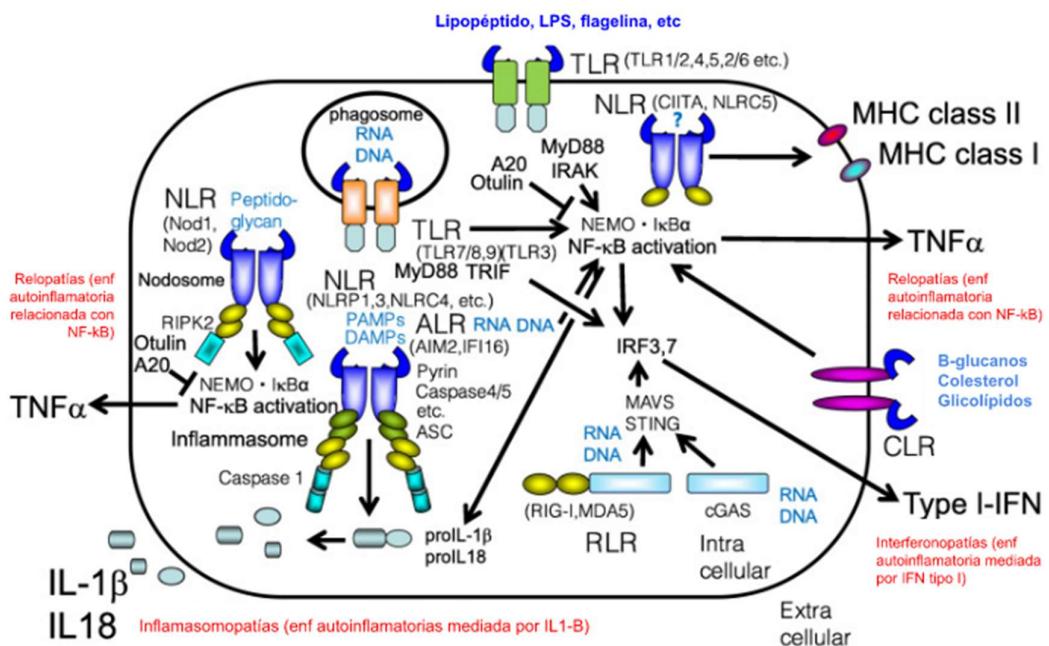


Figura 2. Receptores de reconocimiento de patrones en las células inmunes innatas. Adaptado de Masumoto et al. 2021 [97]

Los diagnósticos diferenciales de las enfermedades autoinflamatorias son complejos y dependen en gran medida del contexto clínico. No existe una única lista de signos y síntomas o estudios que excluyan efectivamente todas las posibles causas de inflamación recurrente de causa incierta. Los estudios deben adaptarse a cada persona y estar dirigidos por una historia clínica exhaustiva y detallada. En general, los estudios iniciales tienden a centrarse en excluir inmunodeficiencia, infección, y otras enfermedades genéticas poco frecuentes en los niños, mientras que en los adultos los trastornos adquiridos, como la

autoinmunidad, vasculitis, sarcoidosis y malignidad, adquieren mayor importancia [100]. En ambos casos, contar por lo tanto con un diagnóstico molecular preciso, es una herramienta y contribución indispensable.

Estrategias de bloqueo del inflamasoma y las citoquinas en las enfermedades autoinflamatorias

Muchas enfermedades autoinflamatorias -en particular los defectos monogénicos- se asocian con alteraciones de los inflamasomas. De acuerdo con la estructura del receptor de tipo NOD (NLR) implicado, se han descrito 4 tipos de inflamasomas hasta la fecha:

- El NLRP1, que puede ser activado por una toxina derivada del ántrax [101]
- El NLRC4, activado por la flagelina bacteriana [102]
- El AIM2, un inflamasoma de la familia PYHIN, que se activa mediante la unión a ADN de doble cadena viral y bacteriana de patógenos intracelulares [103]
- El NLRP3, el más estudiado, que puede ser activado por una gran variedad de PAMPs y DAMPs, por especies reactivas de oxígeno (ROS, del inglés *Reactive Oxygen Species*), proteasas liberadas en el citosol, daño lisosomal causado por amiloide-β, sílice, catepsina B, cristales de colesterol o salida de potasio de la célula [104].

Los inflamasomas son estructuras multimoleculares complejas que responden a las señales de "peligro" mediante la activación de citoquinas [105]. Además de los PAMPs y DAMPs, el inflamasoma puede interactuar con y ser estimulada por proteínas como la pirina (codificada por el gen *MEFV*), la proteína que interactúa con la prolina-serina-treonina fosfatasa 1 (codificada por el gen *PSTPIP1*), la mevalonato quinasa (codificada por el gen *MVK*) y la proteína NLRP7. Todas ellas también pueden verse alteradas en su estructura y función por la presencia de variantes patogénicas, que dan lugar a enfermedades autoinflamatorias.

Tras la activación del inflamasoma, la procaspasa 1 va a ser activada la caspasa 1, que a su vez escinde la pro-interleucina-1β y la convierte en su forma activa, la IL-1β [106]. La IL-1β es, además del TNF-α, uno de los principales mediadores de la fiebre y la inflamación en humanos y se ha denominado "citoquina de alarma prototípica", ya que coordina la respuesta más temprana del sistema inmune en respuesta al daño de origen exógeno o endógeno, reconociendo su papel fundamental en la inmunidad innata [107,108]. Una vez liberada, la IL-1β conduce a la dimerización de la proteína accesoria del receptor de IL-1 (IL-1RAcP) con el receptor de IL-1 tipo 1 (IL-1R1) [109]. Esto provoca una cascada de señales descendentes que finalmente dan lugar a la activación de NFκB y la producción y liberación de otras citoquinas inflamatorias [110]. Como consecuencia de la activación del inflamasoma, se producen una gran variedad de citoquinas liberadas por las células del sistema inmune

innato (monocitos, macrófagos y células dendríticas). Entre ellas, se encuentran las familias IL-1 (IL-1, IL-18, IL-33), TNF (TNF- α , LT- α), IL-6 (IL-6, IL-11), IL-17 (IL-17A, IL-25) y los IFN de tipo 1 (IFN- α , IFN- β) [111]. La inhibición de la cascada inflamatoria reviste un gran interés, tanto desde el punto de vista científico como, tal vez, desde el punto de vista clínico. Por ello, se han identificado una serie de mediadores y sustancias capaces de modular o incluso bloquear el proceso inflamatorio en distintos puntos [112].

En cuanto a la patogénesis inflamatoria implicada en la enfermedad, se han clasificado los trastornos autoinflamatorios monogenéticos como mediados por IL-1 (CAPS y DIRA), parcialmente mediados por la IL-1 (FMF, HIDS, PAPA) y mediados por otras vías (TRAPS, Síndrome de Blau, Síndrome de Majeed, Querubismo y Deficiencia del receptor de IL-10) [113]. En el contexto de las EAI, éstas son las más frecuentes, son enfermedades monogénicas clásicas, su modelo de herencia y genes afectados son conocidos, y existen tratamientos disponibles. En la Tabla 1 se resume dicha información.

En relación con los tratamientos, debido al papel destacado de la IL-1 β en la inflamación humana y a los efectos favorables del bloqueo de la IL-1 β , estas estrategias han sido las más estudiadas y utilizadas en el CAPS, la FMF resistente a la colchicina [114], el TRAPS, el HIDS y el DIRA. La inhibición de la IL-1 puede conseguirse a diferentes niveles: algunas sustancias están dirigidas directamente a la molécula de IL-1 β (Canakinumab [115], Rilonacept [116], Gevokizumab [117]), mientras que otras son antagonistas del receptor de la IL-1 (Anakinra) y han sido eficaces no sólo en trastornos mediados por la IL-1, sino también en trastornos metabólicos como la gota o la diabetes, y en enfermedades autoinflamatorias genéticamente complejas como la enfermedad de Behcet, y la Artritis Idiopática Juvenil Sistémica o la Enfermedad de Still en el adulto [118, 119]. Otros bloqueantes de citocinas utilizados en la práctica clínica son los inhibidores del TNF (Etanercept [117, 120], Infliximab [121] y Adalimumab [122]) y el inhibidor de la IL-6 (Tocilizumab [123]).

Por su parte, la colchicina se ha utilizado durante siglos para el tratamiento de trastornos inflamatorios [122–124], resultando eficaz en la gota [125], pero también en la enfermedad de Behcet [126] y en la FMF, donde es capaz de prevenir la amiloidosis [127]. El fármaco afecta a muchos tipos de células y se acumula preferentemente en los neutrófilos. Aunque su modo de acción aún no está claro, tiene propiedades desestabilizadoras de los microtúbulos que pueden ser parte de sus efectos. Efectos adicionales como la alteración de la expresión de moléculas de adhesión, quimiotaxis y la generación de ROS también influyen en la inflamación [128]. En general, la colchicina tiene buena tolerancia. Sin embargo, se producen efectos secundarios gastrointestinales, hematológicos y neuromusculares cuando la dosis administrada es demasiado alta [129].

Otros biológicos como el Abatacept (inhibe la activación de las células T, utilizado en la artritis reumatoide [130]), Belimumab (anticuerpo monoclonal humano que inhibe el factor

activador de las células B, aprobado por la FDA para el lupus sistémico [131]) o Rituximab (anticuerpo monoclonal quimérico anti-CD20, utilizado en artritis reumatoide, púrpura idiopática o nefritis lúpica [132]) son eficaces en las enfermedades autoinmunes, pero no existen informes sobre su aplicación en trastornos autoinflamatorios.

Hasta la fecha, el bloqueo de la IL-1 β es el tratamiento más eficaz en la mayoría de las enfermedades autoinflamatorias monogenéticas [133]. Los efectos más favorables se observan en el tratamiento de los síndromes periódicos como Criopirinopatías, aunque también eficaz en otras enfermedades como DIRA, TRAPS, PFAPA, FMF resistente a la colchicina, entre otros. La inhibición de la IL-1 también desempeña un papel en enfermedades autoinflamatorias multifactoriales y comunes como la diabetes, la gota y la arterosclerosis.

Es claro que el conocimiento de las bases moleculares que llevan al desarrollo de las EAI, derivado del conocimiento de los genes afectados por variantes patogénicas, ha contribuido enormemente al desarrollo de una amplia variedad de potenciales terapias y que en la actualidad su tratamiento ofrece un panorama complejo pero esperanzador. **Es por ello que en esta tesis buscamos contribuir al avance del conocimiento de las bases moleculares de las EAI mediante el análisis de datos genómicos de pacientes.**

Tabla 1. Clasificación de las enfermedades autoinflamatorias sistémicas hereditarias, genes responsables, modelo de herencia y enfoque terapéutico disponible.

Síndromes hereditarios de fiebre periódica					
Nombre	OMIM	Gen- Proteína afectada	Modelo de herencia	Principales características	Enfoque terapéutico
Fiebre mediterránea familiar (FMF)	249100	MEFV- Pirina/marenostrina	AR (AD en pocos casos)	<ul style="list-style-type: none"> ● Debut: <20 años. ● Fiebre de duración 1-3 días, 10 a 12 veces/año. ● Manifestaciones digestivas, peritonitis estéril. ● Poliartralgias monoartritis. ● Mialgias ocasionales. ● Eritema de tipo erisipela. ● Adenopatías ocasionales. ● Amiloidosis variable. 	Colchicina, Canakinumab (para no respondedores a colchicina)
Síndrome periódico asociado al receptor del TNF (TRAPS)	260920	TNFRSF1A - Receptor 1 del TNF	AD	<ul style="list-style-type: none"> ● Debut: <10 años. ● Fiebre de duración >7 días, 3-6 veces por año. ● Manifestaciones digestivas muy frecuentes, dolor abdominal. ● Poliartralgias Monoartritis. ● Mialgias muy frecuentes migratorias. ● Muy frecuentes Migratorias. ● Muy frecuentes: edema periorbital y conjuntivitis. ● Adenopatías ocasionales. ● Amiloidosis variable (2-25%). 	- Agentes antiinflamatorios no-esteroides para el tratamiento sintomático. - Corticosteroides para episodios inflamatorios, pero a menudo se requieren dosis crecientes. - Etanercept, un inhibidor de TNF, ha demostrado ser eficaz (disminuye con el tiempo). - Anakinra y Canakinumab también han sido utilizados proporcionando un control óptimo de las manifestaciones inflamatorias a largo plazo.

Síndrome de hiper-IgD y fiebre periódica (HIDS)	142680	MVK- Mevalonato Kinasa	AR	<ul style="list-style-type: none"> ● Debut: <12 meses ● Fiebres de duración 4-6 días, 9-10 veces/año ● Manifestaciones digestivas, dolor abdominal y diarrea. ● Poliartralgias artritis ocasionales. ● Mialgias ocasionales. ● Exantema maculopapular Adenopatías muy frecuentes. ● Amiloidosis muy frecuente. 	<ul style="list-style-type: none"> - Altas dosis de Prednisona. - Anakinra o, con menor eficacia, Etanercept. - Canakinumab,
---	--------	------------------------	----	--	---

Síndromes periódicos asociados a criopirina (CAPS)

FCAS	120100	<i>NLRP3 (CIAS1) - Criopirina</i>	AD	Episodios recurrentes de fiebre autolimitada, urticaria, dolor articular, conjuntivitis, náuseas, sed extrema, sudoración y dolor de cabeza.	<p>Sintomático, antiinflamatorios no esteroideos, corticoides, y analgésicos.</p> <p>El tratamiento consiste en inhibidores de IL-1β: Anakinra, Rilonacept y Canakinumab.</p>
Síndrome de Muckle-Wells	191900			Episodios recurrentes de fiebre y urticaria (rash), dolor articular, inflamación ocular, pérdida auditiva gradual, riesgo de amiloidosis con daño renal, fatiga, dolor de cabeza, náuseas, sed extrema y escalofríos.	
Síndrome de CINCA-NOMID	607115			Meningitis crónica progresiva, pérdida de visión y audición, problemas con el aprendizaje, artritis, retraso del crecimiento óseo, amiloidosis con daño renal, rash permanente, inflamación ocular, esplenomegalia, náuseas, vómitos, fatiga y dolor de cabeza.	

Granulomatosis sistémicas pediátricas						
Síndrome de Blau y Sarcoidosis de inicio precoz	186580	<i>NOD2- Proteína que contiene el dominio de oligomerización de unión a nucleótidos 2</i>	AD	Afectación de piel, ojos y articulaciones. Los síntomas se manifiestan en general en niños menores a 4 años con dermatitis, artritis granulomatosa y uveítis.	El síndrome de Blau no tiene cura, pero altas dosis de corticosteroides pueden ayudar a reducir la inflamación. Los medicamentos inmunosupresores también pueden ayudar y el Adalimumab puede reducir la inflamación en combinación con corticosteroides, inmunosupresores o ambos.	
Otras enfermedades autoinflamatorias sistémicas hereditarias						
Síndrome PAPA	604416	<i>PSTPIP1- Proteína de interacción prolina-serina-treonina fosfatasa 1</i>	AD	La artritis se inicia en la primera década de la vida y es progresivamente destructiva. Episodios de traumatismo menor pueden desencadenarla. A veces aparecen úlceras con mala cicatrización y bordes despegados, a menudo en localizaciones de lesión (p. ej., sitios de vacunación). Por lo general, el acné es noduloquístico y, de no ser tratado, causa cicatrización. En la pubertad, la artritis tiende a desaparecer y predominan los síntomas cutáneos.	La artritis y las lesiones en la piel responden a los glucocorticoides. El tratamiento con Etanercept o Anakinra puede ser útil. El acné se trata con tetraciclina o isotretinoína por vía oral.	
CRMO, Síndrome Majeed	609628	<i>LPIN2- Lipin 2</i>	AR	Puede afectar a uno o varios huesos y el dolor óseo es el síntoma más frecuente (no se debe a infección). Suele haber sensibilidad y dolor en la zona afectada. Algunos pacientes pueden desarrollar artritis y fatiga.	El tratamiento suele comenzar con AINEs (ibuprofeno, naproxeno, meloxicam), pero algunos pacientes necesitan medicamentos más potentes, como metotrexato, biológicos (Etanercept, Adalimumab, Infliximab, Anakinra) y bifosfonatos (pamidronato, ácido zoledrónico).	

CAPS: Síndrome periódico asociado a la criopirina, FCAS: Síndrome autoinflamatorio familiar inducido por el frío, CINCA: Síndrome articular, cutáneo y neurológico crónico infantil o NOMID: Enfermedad inflamatoria multisistémica de inicio neonatal, PAPA: Artritis piógena, pioderma gangrenoso y acné, CRMO: Osteomielitis crónica multifocal recurrente, AINEs: Medicamentos antiinflamatorios no esteroides, AD: autosómico dominante, AR: autosómico recesivo.

Diagnóstico de una enfermedad genética en la era NGS

El diagnóstico de enfermedades genéticas implica tradicionalmente un examen clínico integral que consta de, al menos, tres elementos principales [134]:

1. Examen físico;
2. Antecedentes familiares detallados;
3. Resultados de pruebas clínicas, estudios especializados y de laboratorio (en el caso que corresponda y estén disponibles).

Si bien un médico general o de cabecera no siempre tiene la capacidad de determinar el diagnóstico definitivo de una enfermedad genética, su función es esencial en recopilar los detalles de los antecedentes familiares, evaluar la posibilidad del desarrollo de una enfermedad genética tras un diagnóstico presuntivo, ordenar la realización de estudios complementarios y, si es adecuado, remitir al paciente a un especialista en genética médica. En este contexto, las tecnologías de secuenciación de próxima generación (NGS) se introdujeron e insertaron sólidamente en laboratorios y clínicas que ofrecían la posibilidad de realizar análisis genéticos, luego de convertirse en un método confiable, preciso y rentable aplicado tanto al diagnóstico molecular como al descubrimiento de nuevos genes asociados con patologías hereditarias de etiología desconocida o complejas, como las cardiomiyopatías, arritmias cardíacas, trastornos del tejido conectivo o neuronales y autismo, donde una gran cantidad de genes están involucrados en un gran espectro fenotípico. Dado que las plataformas de NGS permiten realizar la secuenciación masiva y paralela de diferentes regiones “target” u objetivo, se utilizan ampliamente tanto en el diagnóstico de enfermedades infecciosas, inmunes, trastornos poco frecuentes como en el diagnóstico prenatal no invasivo y, más recientemente, en la toma de decisiones terapéuticas para el tratamiento del cáncer.

Es aquí donde los desafíos que los médicos están comenzando a enfrentar hoy en día contemplan la elección de qué tipo de estrategia es la más adecuada para aplicar a cada caso clínico. Actualmente, hay tres opciones disponibles: paneles de genes dirigidos, secuenciación del exoma (WES), o secuenciación del genoma completo (WGS).

Elección del mejor enfoque: paneles, WES o WGS

Los profesionales de la salud suelen plantearse si deben comenzar su análisis con una prueba dirigida al diagnóstico clínico presuntivo, o si, por el contrario, resulta preferible iniciar la búsqueda de un diagnóstico solicitando directamente la secuenciación del exoma o genoma. El costo de una prueba de secuenciación de múltiples genes dirigida a una determinada enfermedad generalmente varía de 150 a 1000 USD, dependiendo de la

tecnología seleccionada, la profundidad de secuenciación y la empresa con la que se desea trabajar. Por otro lado, las pruebas clínicas de secuenciación del exoma varían entre 350 a 1000 USD, mientras que el costo de un genoma podría contemplarse entre 1000 a 2000 USD.

La secuenciación dirigida de los genes de interés seleccionados ofrece una buena cobertura, con una profundidad media de entre 300X-500X, dependiendo de las plataformas y el número de muestras analizadas, para asegurarse una cobertura pareja y completa de la totalidad del panel analizado, que justifique el mayor precio y el objetivo acotado de genes (regiones) a analizar. En el caso de existir, aquellas regiones específicas no detectadas, se pueden secuenciar por la técnica de Sanger, con el fin de cubrir todas las regiones de interés y/o validar los resultados obtenidos [135, 136]. Por ejemplo, las pruebas con PCR de largo alcance se pueden realizar en paralelo para compensar las fallas de NGS en detectar aquellos genes con pseudogenes, como es el caso del gen *STRC14* [137]. Además, los laboratorios a menudo complementan los paneles de genes con métodos de detección de variantes en el número de copias CNVs (del inglés, Copy Number Variation: variación del número de copias) para identificar grandes delecciones o inserciones y otros, que los servicios de secuenciación del exoma no detectan actualmente. Este enfoque de secuenciación dirigida se ha adoptado para determinados trastornos genéticos, que poseen un número chico o moderado de genes asociados a los mismos, como es el caso de sordera no sindrómica, síndrome nefrótico, cardiomielopatías, enfermedades lisosomales o en el diagnóstico de otras condiciones mendelianas. Otra clara ventaja del uso del panel dirigido es la reducción del número de hallazgos incidentales y/o de variantes de significado incierto. Su principal limitación es la restricción de búsqueda de variantes a un número acotado de genes, pudiendo perderse información de relevancia. La aplicación de paneles de secuenciación masiva al diagnóstico genético requiere del estudio de los genes asociados a la patología, del diseño bioinformático del panel y de su validación, que permita evaluar los parámetros de calidad del mismo, como la reproducibilidad, cobertura media, sensibilidad, especificidad, detección de delecciones e *indels* ("inserción o delección"), previa a su aplicación en el diagnóstico de enfermedades genéticas. Dado que el campo genético se encuentra en constante cambio, la aparición de evidencia que relacione nuevos genes con un fenotipo clínico particular provocaría la necesidad de rediseñar y revalidar el panel [137,138].

El Colegio Americano de Genética Médica y Genómica sugiere que los enfoques de secuenciación del genoma (WGS) o exoma completo (WES) deben reservarse para aquellos casos en los que las pruebas dirigidas al diagnóstico presuntivo (paneles) sean negativas, o tengan pocas probabilidades de devolver un resultado positivo de forma adecuada y costo efectiva [139]. En este contexto, el beneficio de la secuenciación del exoma radica en proveer una gran cantidad de información genética correspondiente a un elevado número de genes (cerca de 18 mil), incluso si en la práctica, la cobertura completa de todos los exones

codificantes no es factible. La cobertura típica de los exones es aproximadamente del 90 al 95% [140]. Sin embargo, al analizar pequeños conjuntos de genes que están implicados en una determinada enfermedad, la cobertura puede ser mucho menor, y esto es algo que suele y debe ser evaluado previo a la decisión de realizar la secuenciación. La aplicación del WES puede ser útil, por ejemplo, en casos que presentan fenotipos complejos y heterogéneos y que sugieren la existencia de múltiples enfermedades o son poco claros y/o atípicos, o cuando una prueba genética anterior no ha tenido éxito. Los estudios de WES, al incluir no sólo los exones de aquellos genes asociados al fenotipo en estudio, también permiten la identificación de nuevos genes asociados a enfermedades con base molecular conocida, y/o la identificación de aquellos genes causales de una enfermedad genética para la cual los mismos son desconocidos. Este enfoque permitió la identificación de nuevas variantes en genes responsables de numerosos síndromes, como el de Freeman-Sheldon, el Síndrome de Miller y el Síndrome de Schinzel-Giedion. Otra ventaja del WES es que el análisis de variantes puede limitarse a los genes de interés, y posteriormente ampliarse si es necesario, minimizando la posibilidad de detección de hallazgos incidentales o secundarios. En base a estas ventajas, en muchos casos, se implementa el uso de WES como el primer análisis genético a realizar. Su costo moderado (comparado con WGS) hace usualmente posible el análisis de tríos (secuenciación del caso índice y sus progenitores) de una forma costo efectiva, facilitando la interpretación de los datos y aumentando el rendimiento diagnóstico considerablemente (20-30%) y aún más si se acompaña por alguna tecnología de secuenciación de CNVs [141,142]. En cuanto a las contras de la utilización de WES, es importante tener en cuenta que aproximadamente el 10% de las bases secuenciadas obtienen una profundidad de lectura similar o menor a 20 (necesaria para un diagnóstico certero), y que aproximadamente solo el 85% de los genes asociados a enfermedades humanas presentes en la base de datos de OMIM recibe la cobertura adecuada. La cobertura deficiente puede deberse a varios factores: que no se hayan alineado las sondas específicas para ciertos genes, ya sea porque los genes no se eligieron para su inclusión durante el desarrollo del ensayo o porque las secuencias repetitivas impidieron su inclusión, y/o relacionado a las sondas con bajo rendimiento debido a la riqueza de GC o la baja calidad de mapeo. Además, al igual que ocurre en los paneles de genes, las herramientas que han sido desarrolladas para la detección de CNVs mediante WES siguen sin ser tan precisas como las utilizadas para WGS [143,144]. Por estos motivos, entre otros, es posible que incluso para un diagnóstico presuntivo asociado claramente a una enfermedad genética, el estudio de WES no permita identificar la variante responsable del fenotipo. Es por lo que, en ciertos casos, surge la necesidad de recurrir a la secuenciación del genoma completo.

La secuenciación del genoma completo, conocida como WGS, se ha utilizado en gran medida como una herramienta de investigación, pero actualmente se está introduciendo en

el diagnóstico clínico. Es una estrategia que utiliza tecnología innovadora y avanzada, constituyendo una herramienta clave para llevar a cabo nuevos descubrimientos, obteniendo todos los datos genéticos de un individuo en un solo experimento y accediendo a la detección tanto de SNPs, SNVs y CNVs. Actualmente, los costos de secuenciación y la capacidad de cómputo para procesar y almacenar los datos crudos constituyen unas de las principales desventajas de la utilización de WGS, así también como el hallazgo de millones de variantes de significado incierto, principalmente debido a la falta de información (frecuencia, evidencia de correlación genotipo-fenotipo, predicción de patogenicidad, etc.) de variantes presentes en regiones no codificantes, regulatorias o repetitivas. Por otro lado, la secuenciación del genoma completo tiene la ventaja de brindar una cobertura continua, permitir la identificación de variantes por fuera de los exones, mejorar la identificación de CNVs y lograr una mayor eficiencia diagnóstica. La búsqueda de variantes, como en WES, podría dirigirse inicialmente a las regiones de interés, ya que la interpretación de variantes en regiones intrónicas, intergénicas o regulatorias es mucho más compleja. Tales variantes podrían cobrar relevancia en un futuro cercano, incrementando el valor de los datos obtenidos mediante WGS [145].

En resumen, cuando se comparan todas las opciones (Paneles, WES, WGS), parece claro que la secuenciación del genoma completo (WGS) constituye teóricamente el mejor enfoque, ya que produce una mayor cantidad de datos genómicos de un individuo. Ofrece importantes ventajas: cobertura más uniforme que WES y permite además la detección de variantes intrónicas y de CNVs, incluyendo pérdidas de un solo exón. Sin embargo, obtener una cobertura suficiente incrementa los costos y el tiempo de análisis limita su implementación en el diagnóstico genético de rutina, ya que se detectan miles de variantes de difícil interpretación. Por el contrario, el WES es considerablemente más económico que el WGS y sigue siendo considerado como una estrategia eficiente [146]. El análisis de paneles representa una alternativa menos costosa, pero que solo tendrá éxito si el gen que causa la enfermedad está incluido en el panel. La ventaja es, además del corto tiempo de resultado, que permite una mayor cobertura a un coste menor que los enfoques de exoma o genoma y reduce la posibilidad de hallazgos inespecíficos o accidentales. Esto puede solucionarse en WES y WGS dirigiendo el análisis a unos pocos genes seleccionados y conocidos y, si no se encuentra nada clínicamente significativo, ampliando la búsqueda de variantes a otros genes relevantes y realizando modificaciones en la aplicación de filtros durante el análisis bioinformático [145]. **Por los motivos expuestos, en la presente tesis, hemos adoptado principalmente un enfoque basado en WES con un análisis dirigido a paneles asociados al diagnóstico clínico presuntivo.**

Dificultad de implementación local de estas tecnologías

Como muchos de los desarrollos tecnológicos, la Medicina de Precisión Genómica tiene su epicentro en países europeos, Estados Unidos, Canadá, Japón y China, que financian y fomentan el desarrollo de programas masivos que pretenden secuenciar cientos de miles de genomas humanos. En los países periféricos, como la Argentina y el resto de Latinoamérica, aún no existen programas de esta envergadura, principalmente debido a la escasez de recursos humanos formados en el área, a la falta de capital para el financiamiento de proyectos de tal extensión, y a la tendencia a adquirir este tipo de productos y servicios tecnológicos en el exterior. Esto resulta en un acceso restringido y sobrevaluado al servicio. Dicha situación sólo puede ser modificada con la formación local de profesionales de la salud (y otras disciplinas) en estas técnicas y la demostración de casos de éxito realizados localmente. Paradójicamente, en Argentina y varios países de Latinoamérica (Brasil, Chile, Uruguay, etc.), se encuentra disponible la capacidad tecnológica para secuenciar genomas humanos, y la misma se halla ampliamente subutilizada. La implementación de estas potentes tecnologías en laboratorios clínicos certificados o en Universidades presenta un gran desafío debido a su complejidad, al gran volumen de datos generados, la dificultad del almacenamiento y a las implicaciones médico-legales de los datos generados [147]. Particularmente, los costos de los reactivos y el mantenimiento de los equipos constituyen el principal obstáculo ya que, aunque en los últimos años han comenzado a descender, se encuentran en moneda extranjera, significando un alto y fluctuante costo para nuestro país. Además, un laboratorio molecular clínico debe cumplir con los diversos requisitos y pautas de calidad, lo que ralentiza el proceso. Por otro lado, es importante destacar que la genómica clínica es un área del conocimiento netamente interdisciplinaria que requiere del trabajo conjunto y sinérgico de médicos, bioquímicos, biólogos, estadísticos e informáticos en pos de un objetivo común, y que se encuentra en un incipiente desarrollo en la región.

Objetivos

El objetivo principal de esta tesis fue poner a punto y desarrollar un protocolo general de aplicación de filtros sucesivos que permitiera facilitar la priorización y el análisis de variantes genéticas derivadas de secuenciación de exomas con la finalidad de llegar al diagnóstico genético-molecular de pacientes con fuerte sospecha de Enfermedades Poco Frecuentes, ya sea en el marco de casos individuales, como en el contexto de proyectos de secuenciación clínica de mediana envergadura.

Especificamente nos propusimos:

- Desarrollar el caso de una familia con numerosos miembros afectados que evidenciaron un fenotipo poco frecuente y atípico, y resultaron ser portadores de dos variantes en el gen *AFG3L2*.
- Exponer y profundizar el caso de una paciente con un fenotipo autoinflamatorio poco frecuente y atípico, portadora de dos variantes en el gen *ISG15*.
- Analizar de forma global e integrada los resultados de la secuenciación de 165 exomas de pacientes con diagnóstico presuntivo de Enfermedades Autoinflamatorias, en el marco del proyecto PAMPA, clasificando la información obtenida de acuerdo a la relevancia de los resultados.

Materiales y Métodos

Selección de Pacientes

La selección de pacientes se menciona en detalle en los capítulos I, II y III, y se adaptó a los objetivos específicos.

Consentimiento informado

Todos los casos analizados formaron parte de: i) protocolos de investigación aprobados por los comités de ética de las instituciones que atendieron a los pacientes, con los consentimientos informados correspondientes; o alternativamente ii) casos enmarcados en un proceso de innovación clínica donde, de acuerdo con los principios expresados en la Declaración de Helsinki de la Asociación Médica Mundial (AMM) [148], el médico responsable del caso otorga y explica el consentimiento informado al paciente y/o a sus tutores, quienes consienten de manera informada sobre la realización del estudio.

Protocolo para la Secuenciación Exómica

Una vez seleccionados los casos, e informados el médico y paciente, se procedió a la firma de los consentimientos informados y declaraciones juradas. Posteriormente, se derivó a cada paciente a un laboratorio de análisis clínicos o al área de bioquímica del hospital postulante para realizar la extracción de una muestra de sangre periférica. La purificación de ADN de linfocitos circulantes la realizó, en la mayoría de los casos, un laboratorio tercerizado, mediante la utilización de kits comerciales QIAamp DNA Mini Kit (QIAGEN), logrando una concentración final mínima de 50 ng/ul y una pureza mayor a 1,8 en relación de absorbancia 260nm/280nm. En algunos casos, las muestras de ADN fueron directamente aportadas por las instituciones de los pacientes postulados. La captura exómica se realizó con el kit de “*Agilent SureSelect Human All Exon V5 (Caso I, Caso II y Edición PAMPA 2018)*”, “*Agilent SureSelect Human All Exon V7 (Ediciones PAMPA 2019-2021)*” y las muestras fueron secuenciadas mediante la tecnología “*Illumina HiSeq 4000*” con una longitud de lectura de 100 pares de bases y una profundidad promedio de 100X. Todas las muestras incorporadas al Proyecto PAMPA fueron anonimizadas en todos los pasos, desde la extracción de sangre hasta la entrega de resultados finales al médico responsable.

Hallazgos incidentales

En todos los casos, el análisis se limitó a aquellos genes previamente consensuados con el profesional como directamente relacionados con el diagnóstico presuntivo, para disminuir la posibilidad de hallazgos incidentales. En ninguno de los casos analizados para esta tesis se presentó/reportó un hallazgo incidental.

Precisión y tasa de error de un experimento de secuenciación por NGS

Cuando se realiza un experimento de NGS, como ya se mencionó anteriormente, lo que se termina secuenciando son pequeños fragmentos de aproximadamente 100 pb que se denominan lecturas. Cada lectura es un fragmento al azar de la secuencia completa y la tasa de error de éstas ronda el 0.1-1% por base leída. Es por esto que, en la práctica, se deben “leer” muchas lecturas superpuestas, lo que permite luego reconstruir informáticamente la secuencia de ADN original. En este punto aparecen dos parámetros claves que determinan la resolución y tasa de error de un experimento de secuenciación por NGS: i) la profundidad (también conocida como cobertura vertical) y ii) la cobertura horizontal. El primero refiere al número promedio de lecturas que cubren a cada base dentro de la muestra de ADN -en otras palabras, cuantas veces es leída (o secuenciada) cada base-, mientras que la segunda se relaciona con la extensión de la secuencia “leída” en relación a la longitud total del ADN que uno desea secuenciar. A modo de ejemplo, un genoma completo secuenciado al 90% de cobertura horizontal y 60X de profundidad, significa que 9 de cada 10 bases del genoma fueron leídas al menos una vez y que, en promedio, cada base en el genoma fue “leída” 60 veces. Cuando no se especifica lo contrario, la palabra “cobertura” suele utilizarse para hacer referencia a la profundidad o cobertura vertical.

Análisis bioinformático

Etapas generales

Se pre-procesaron los resultados que entregó el secuenciador- las lecturas y datos asociados- utilizando un *pipeline* basado en las buenas prácticas de GATK del Broad Institute [149], que consistió en los siguientes pasos i) en control de calidad de las lecturas -con descarte de las que no superaran el umbral mínimo de calidad-, ii) mapeo-alineamiento, iii) llamado de variantes. Luego se procedió a la anotación de las variantes (iv) y, finalmente, al análisis y priorización de las mismas en el contexto de la clínica de cada caso.

En primer lugar, se procedió a marcar las lecturas que pudieran ser producto de duplicaciones de PCR por medio del software PICARD, se detectaron regiones con varias

lecturas de baja calidad y se realizó un realineamiento local con el fin de detectar posibles delecciones o inserciones pequeñas. El siguiente paso consistió en detectar las diferencias entre el consenso de las lecturas obtenidas y el genoma de referencia, en un proceso denominado “llamado de variantes” (del inglés “variant calling”) mediante la herramienta “*Haplotype Caller*” de GATK. A partir de la información de calidad de las variantes se realizó un filtrado, con el protocolo “*Hard Filtering*”, según las recomendaciones detalladas por GATK, con el fin de detectar y filtrar aquellas variantes llamadas de baja calidad (falsos positivos). Como último paso, se procedió a anotar el VCF (vincular las variantes con datos biológicos) con información externa de bases de datos utilizando el paquete de software SnpEff/SnpSift [150] y Annovar [151]. Para ello se utilizaron las siguientes fuentes: dbSNP, ExAC, 1000 Genomas, GnomAD para frecuencia poblacional [152], ClinVar [153] para la relevancia clínica y Polyphen [154], SIFT [155] y Mutation Taster [156], entre otros, para las predicciones de patogenicidad. La Figura 3 resume el procesamiento bioinformático que tuvo lugar, indicando los pasos seguidos y los programas utilizados que luego serán explicados en detenimiento [149].

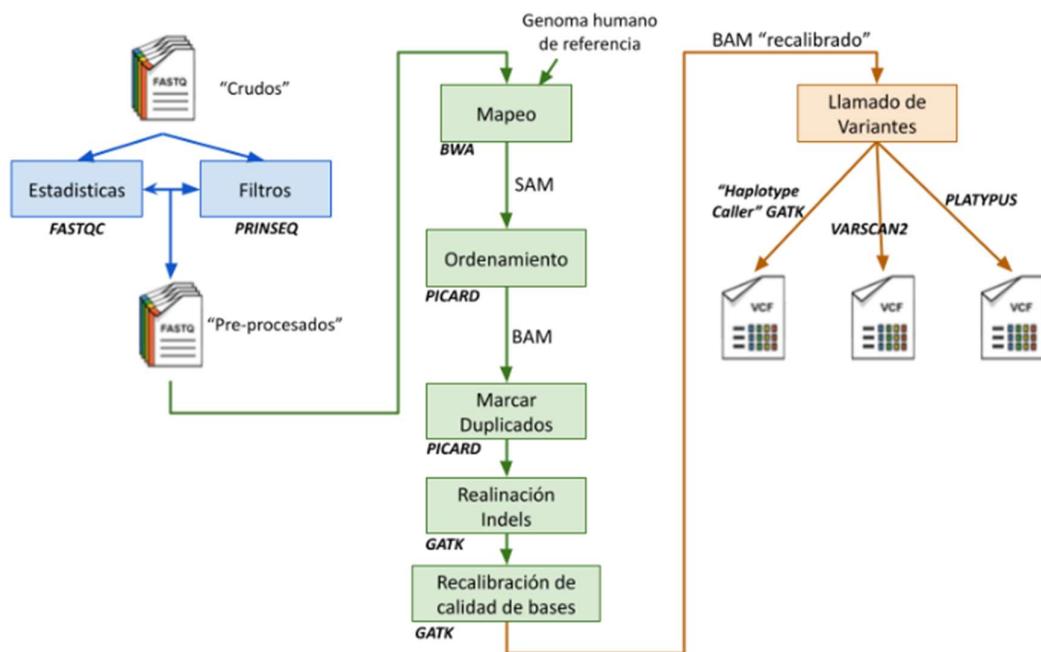


Figura 3: Esquema de pasos para la obtención de variantes. Desde los archivos del secuenciador “FASTQ” hasta los archivos de variantes “VCF”. Se detallan en **ítalic** y **negrita** los distintos programas que se aplicaron en cada paso. Como *llamador de variantes* se utilizó el “*Haplotype Caller*” de GATK.

i) Control de calidad de las lecturas

- Análisis y preprocessamiento de los archivos “.fastq”

Como producto de la secuenciación se obtuvo, para cada muestra, un conjunto de lecturas (*reads*) en dos archivos con formato “.fastq”, similar a los archivos “.fasta” usualmente utilizados para secuencias de nucleótidos, pero con la incorporación de una línea al final con información de la calidad de lectura para cada base. Un archivo FASTQ contiene múltiples líneas de texto, las cuales se agrupan cada 4 filas (o líneas) para representar cada lectura (Figura 4).

Previamente, a los pasos de alineamiento y mapeo necesarios para reconstruir la secuencia, debió controlarse la calidad de las lecturas. Para ello, se analizó la longitud promedio de las mismas, el contenido de C-G -que debe ajustarse a la probabilidad esperada- la calidad de lectura por base, etc. Un parámetro importante es la cantidad de lecturas repetidas, derivadas del proceso de duplicación por PCR y no por fragmentación al azar de la muestra de ADN, que pueden influir erróneamente en el llamado de variantes. Sin embargo, en caso de existir, son fácilmente excluibles en pasos posteriores.

```
@sequence_id
GAAAATGGCTTCGAAATTGGTTTTAGCTCCCCACTTATGACACACTTCTGTAAATGA
+
CCCFFFEHGGHCG:GA:DHHE>EEHHHCDBD?B@FFFHHHIJIJEHCFHGDEHGIJGIGII
```

Figura 4: Ejemplo de un archivo “.fastq”. La primera línea, precedida por un símbolo “@” (arroba), contiene el identificador del read (generalmente consiste en el nombre del equipo, el id de la flowcell utilizada, las coordenadas en la misma, si corresponde al par 1 o 2 -si es pair-end- y una descripción). La segunda contiene la secuencia de nucleótidos del read, de longitud variable dependiendo del experimento de secuenciación (representadas por las letras A, C, G y T). La tercera línea es un separador, que contiene un símbolo “+” (suma), y la última línea indica la calidad de lectura en el cual cada símbolo del alfabeto ASCII corresponde a un nivel de calidad en escala Phred, determinada por el secuenciador en base a los resultados del proceso de secuenciación.

Para obtener un análisis general de las lecturas se utilizó el programa FastQC (Babraham Institute. Versión 0.11.4) con el fin de obtener estadísticas en relación a la cantidad de lecturas totales, el tamaño promedio, el contenido de C-G, la cantidad de lecturas repetidas y el contenido de adaptadores que no se hayan filtrado en los pasos previos, entre otros parámetros. Como paso de filtrado, se decidió únicamente eliminar las lecturas que tengan un mínimo de calidad de 10 promedio en valores de Phred score (lo que equivale a

eliminar lecturas con menos del 90% de precisión en las bases, ó que se encuentra 1 error cada 10 bases secuenciadas), en base al programa PRINSEQ versión 0.20.4 [157].

ii) Mapeo y alineamiento contra el genoma humano de referencia

Una vez obtenidas las lecturas y controlada su calidad, fue necesario ordenarlas y conocer su ubicación en el genoma humano de referencia, lo que se conoce como proceso de Mapeo. Es importante distinguir la diferencia entre dos conceptos importantes, similares y complementarios, pero técnicamente diferentes: **alineamiento y mapeo**. En un **alineamiento**, se comparan dos secuencias, estableciendo una correspondencia mutua de bases. Por otro lado, al **mapear** una secuencia, se establece su posición correspondiente dada una secuencia de referencia (en este caso el genoma humano).

En los exomas, se vuelve enormemente útil mapear primero las lecturas frente al genoma de referencia humano, aportado por el Genome Reference Consortium (GRC). De esta forma, se obtuvo un “ensamblado” con todas las lecturas superpuestas ubicadas en su lugar correspondiente en el genoma de referencia (Figura 5). Desde la publicación del primer “borrador” del genoma humano en 2001, se han liberado numerosas versiones que mejoran la información ya presentada, al completar baches en la secuencia o reemplazar bases por aquellas más representativas de todos los genomas analizados. Por esta razón, resulta esencial decidir y aclarar con qué versión del genoma humano de referencia se trabaja, y de este modo obtener resultados comparables y transferibles. En esta tesis se usó en todos los casos el genoma de referencia GRCh37/hg19.

Al ubicar cada lectura en su posición correspondiente, se compara cada base del genoma de referencia con aquellas presentes en las lecturas que se superponen en esa región. Cualquier diferencia (significativa) identificada es reportada como una variante (homocigota o heterocigota).

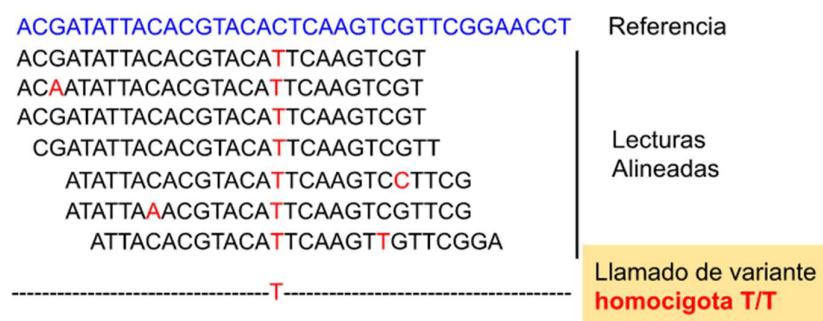


Figura 5: Representación de un mapeo de lecturas en relación con la secuencia de referencia. Cada lectura puede presentar pequeñas diferencias con respecto a la referencia, pero globalmente el alineamiento está bien hecho. Esas diferencias pueden representar errores de la secuenciación o variantes reales presentes en la muestra. Los programas “llamadores de variantes”

analizan las diferencias a lo largo de todas las lecturas secuenciadas y seleccionan aquellas que representan una variación significativa respecto a la referencia, identificándolas como variantes.

El resultado del proceso de alineamiento/mapeo -para el que se suelen utilizar algoritmos de mapeo basados en la transformación de Burrow-Wheeler- es un archivo de formato “.sam” [158]. Se creó con el objetivo de constituir un formato genérico de almacenamiento de alineamiento frente a secuencias de referencia, con la suficiente flexibilidad para manejar lecturas derivadas de secuenciación *single* ó *pair-end*, compacto en tamaño y de eficiente acceso por diferentes fuentes. Consiste en un archivo de texto tabulado que presenta un encabezado con información general del alineamiento y luego renglones correspondientes a cada lectura alineada. Cada línea está dividida en campos tales como nombre del fragmento, nombre de la referencia, posición de mapeo, calidad de mapeo, código CIGAR (código de cómo está alineado a la referencia), la longitud del fragmento y su secuencia, entre otros (Figura 6).

El SAM generado suele estar desordenado (el alineador agrega una línea por cada lectura analizada según aparecen las mismas en el fastQ original), por lo que, para que el mismo pueda ser correctamente “leído” por futuros programas que trabajan de forma secuencial, es necesario un proceso de “ordenado” por coordenadas. Además, para que el archivo ocupe menos espacio de disco (un archivo SAM de un exoma humano puede llegar a pesar 5 GB), se lo puede comprimir a su versión binaria e indexada denominada BAM (“.bam”), el cual ocupa menor espacio, pero no es accesible por un procesador de texto (no es “human-readable” o legible por humanos).

```
@HD VN:1.5 SO:coordinate
@SQ SN:ref LN:51304566
r001 99 ref 7 30 8M2I4M1D3M = 37 39 TTAGATAAAGGATACTG *
r002 0 ref 9 30 3S6M1P1I4M * 0 0 AAAAGATAAGGATA *
r003 0 ref 9 30 5S6M * 0 0 GCCTAAGCTAA *
SA:Z:ref,29,-,6H5M,17,0;
r004 0 ref 16 30 6M14N5M * 0 0 ATAGCTTCAGC *
r003 2064 ref 29 17 6H5M * 0 0 TAGGC *
SA:Z:ref,9,+,5S6M,30,1;
r001 147 ref 37 30 9M = 7 -39 CAGCGGCAT * NM:i:1
```

Figura 6: Ejemplo de archivo en formato SAM. Comienza con un encabezado indicado con “@”, que informan la versión del formato (VN), el tipo de ordenamiento de los alineamientos (SO, por coordenadas), y a continuación el listado de alineamientos. Por ejemplo, la primer línea especifica que el nombre del fragmento es “r0001”, un flag de “99” (suma de números que indican la condición del alineamiento), el nombre de la secuencia de referencia “ref”, su posición de mapeo “7” (en relación a la referencia), su calidad de mapeo de “30”, su código CIGAR “8M2I4M1D3M” - 8 match seguidos, 2 inserciones, 4 match, 1 delección y 3 match -, la referencia del siguiente read en el fragmento es la misma (=) y su posición es “37”, la longitud observada del template es “39”, la secuencia del fragmento es “TTAGATAAAGGATACTG”, y por último la calidad de secuenciación (representada igual que en el formato “.fastq”) en este caso no se almacenó “*”.

En nuestro caso, las lecturas de los “.fastq” procesados fueron alineados al genoma humano de referencia versión GRCh37.75 (GRCh37.p13, EnsEMBL release 75 - February 2014), utilizando el algoritmo BWA-MEM de BWA (Burrows-Wheeler Aligner) de modo de obtener un mapeo en formato “.sam” (SAM/BAM format specification working group, 2013). Este archivo fue ordenado y transformado a formato binario del SAM “.bam” mediante la función SortSam del programa Picard (<http://broadinstitute.github.io/picard/> Nov 2017), y se obtuvieron estadísticas en cuanto al alineamiento con la opción “flagstats” de SamTools [159]. También se calculó la profundidad de cobertura mediante el comando “depth” de SamTools.

- *Procesamiento del archivo “.bam”*

Como se mencionó previamente, los múltiples procesos de PCR realizados desde los pasos de tagmentación hasta la secuenciación propiamente dicha, se pueden generar duplicados que contribuyen a la profundidad de lectura de una forma ficticia. Las lecturas duplicadas resultan no informativas y no deberían ser contadas como evidencia a favor o en contra de una variante en particular. Estos duplicados por lo general no son eliminados, pero son identificados y marcados con un “FLAG” en los archivos SAM o BAM para que los programas llamadores de variantes (ver “Llamado de variantes”) no los tomen en cuenta a la hora de establecer los genotipos para cada variante. Mediante el programa Picard se procedió a marcar los duplicados de PCR del archivo “.bam” mediante la opción “MarkDuplicates”.

Posteriormente al mapeo, se realiza además una **realineación local** alrededor de los indels encontrados, principalmente porque los algoritmos utilizados en los pasos iniciales de mapeo tienden a producir varios tipos de artefactos en las regiones circundantes. En particular, las lecturas que alinean con los bordes de los indels usualmente son mapeadas con “mismatches” en las bases que podrían levantar falsa evidencia para el descubrimiento de SNPs (*single nucleotide polymorphisms* ó polimorfismos de nucleótido único). El proceso de realineamiento identifica la ubicación más consistente de las lecturas relativas al indel para deshacerse de estos artificios. Este procedimiento ocurre en dos pasos, primero, se utiliza la opción “RealignerTargetCreator” de Genome Analysis Toolkit (GATK, versión 3.5 - McKenna et al. 2010), en el cual el programa identifica los intervalos que deben ser realineados y en un segundo paso, mediante la opción “IndelRealigner” del mismo paquete de herramientas, se determina la secuencia consenso óptima y se realinean las lecturas.

Por último, los algoritmos de llamado de variantes, se apoyan en los puntajes de calidad asignados a cada base durante el proceso de secuenciación, los cuales, equivalen a una estimación por base de los errores emitidos por los secuenciadores. Estos suelen ser subestimaciones del error real en los datos debido al aporte de varias fuentes de error sistemático. Se realiza entonces una recalibración de los puntajes de calidad (BQSR, por

Base Quality Score Recalibration) en los que se aplica un proceso “*machine learning*” para modelar estos errores empíricamente y ajustarlos a los puntajes de calidad. En consecuencia, esto permite obtener valores de calidad más precisos, aumentando la precisión de los llamadores de variantes. Este paso también se realiza en dos etapas, primero generando un modelo de la covariación de los datos provistos con un set de variantes conocidas, utilizando la opción “BaseRecalibrator” y luego, se ajusta la calidad de las bases en la muestra basándose en el modelo creado mediante la opción “PrintReads”, ambos comandos son herramientas de GATK.

iii) Llamado de variantes

Una vez mapeadas y alineadas las lecturas contra la referencia, se puede proseguir a llamar las “variantes” (en inglés, *Variant Calling*), es decir, determinar los sitios que en la muestra difieren con la referencia. La naturaleza de estas variantes puede ser de cambios de un único nucleótido (SNP), inserciones o delecciones (*Insertions, Deletions*; INDELs) o variantes estructurales (*Structural Variants*; SV) como las variantes del número de copias (*Copy Number Variations*; CNVs). Los programas llamadores de variantes analizan las regiones donde se presentan variaciones con respecto al genoma de referencia y seleccionan aquellas que cumplan con determinados criterios que las hagan elegibles como “variantes verdaderas”, tales como: calidad de la base secuenciada, calidad de mapeo y el número de lecturas independientes que den evidencia a favor de su presencia. Esta información se suele volcar a un archivo de formato tipo VCF (por sus siglas *Variant Call Format*) [160,161], introducido por el consorcio responsable del proyecto 1000 Genomas [63]. El mismo, consiste en un archivo de texto tabulado en el cual cada línea corresponde a una variante y, al eliminarse toda la información redundante con el genoma de referencia, sólo se obtiene información sobre las particularidades del ADN analizado, lo que lo hace más compacto y fácil de leer (Figura 7).

Para cada variante, en un archivo VCF se informa: el número cromosoma en el que se encuentra (“CHROM”), su posición dentro del cromosoma (“POS”), un identificador (“ID”), el nucleótido presente en el genoma de referencia ('REF'), el nucleótido presente en la muestra secuenciada (“ALT”), la calidad de la secuenciación (“QUAL”), un filtro personalizable para la selección de variantes (“FILTER”), información adicional que se le quiera agregar a la variante (“INFO”) y, por último, un conjunto de campos que nos brindan información de las lecturas y el genotipo de la variante (“FORMAT” que incluye un campo por cada muestra representada en el VCF).

```

##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS    ID      REF ALT     QUAL FILTER INFO FORMAT    NA 00001      NA00002
20    14370  rs6054257 G   A     29  PASS  NS=2;DP=14;AF=0.5;DB;H2      GT:GQ:DP:HQ
0|0:48:1:51,51 1|0:48:8:51,51
20    17330  .       T   A     3    q10  NS=2;DP=11;AF=0.017      GT:GQ:DP:HQ
0|0:49:3:58,50 0|1:3:5:65,3
20    1110696 rs6040355 A   G,T   67  PASS  NS=2;DP=10;AF=0.333,0.667;AA=T,DB
GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2
20    1230237 .       T   .     47  PASS  NS=2;DP=13;AA=T      GT:GQ:DP:HQ
0|0:54:7:56,60 0|0:48:4:51,51
20    1234567 microsat1 GTCT  G,GTACT 50  PASS  NS=2;DP=9;AA=G      GT:GQ:DP
0/1:35:4    0/2:17:2

```

Figura 7: Fragmento de archivo VCF. El archivo VCF contiene un encabezado (header) que nos brinda información acerca del contenido de los campos *INFO* y *FORMAT*. Cada línea del encabezado indica el nombre de cada uno de estos contenidos y que tipo de valor se espera encontrar (número, lista, palabras, etc.).

Para el llamado de variantes se utilizó la herramienta “HapotypeCaller” de GATK (DePristo *et al.*, 2011; Van der Auwera *et al.*, 2013), programa capaz de llamar SNPs e Indels de manera simultánea a través de un ensamblado *de novo* local, lo que implica que cuando el programa encuentra una región que muestra señales de variación, deja a un lado la información existente del mapeo y reensambla completamente las lecturas en la región. Este procedimiento le confiere una mayor precisión a la hora de descubrir variantes en regiones difíciles como, por ejemplo, cuando contienen diferentes tipos de variantes cercanas entre sí. Además, al ser capaz de manejar información perteneciente a organismos no diploides, se vuelve de utilidad para encontrar variantes con frecuencias menores a las típicas heterocigotas (es decir, posibles mosaicismos). **El resultado de este proceso es la obtención de un archivo de tipo VCF para cada una de las muestras secuenciadas, que incluye una lista de variantes.**

iv) Anotación de variantes

Posteriormente, la anotación de variantes o anotación del VCF se basa en la utilización de bases de datos con distinto tipo de información biológica. En un principio la misma suele incluir información genómica-molecular (gen, exón, variante en secuencia codificante y en la proteína, etc.), lo que se conoce como **anotación estructural**, a la que luego se agrega información funcional, como frecuencia poblacional, predicciones bioinformáticas de patogenicidad/conservación, información fenotípica o clínica vinculada a la variante; etc. en el proceso que se conoce como **anotación funcional** [162].

Para cada uno de los genes del exoma, se trajeron de las bases de datos de Clinvar [163], ExAC (*Exome Aggregation Consortium*, Versión 0.3), dbSNP [165], InterVar [164] y GnomAD [152] anotaciones de variantes de tipo SNP y pequeños Indels (Ver apartado “llamado de variantes” para información sobre el formato VCF) y se combinaron en un único VCF con el algoritmo vcf-merge de VCFtools 1.13 [161].

El archivo VCF contiene un gran volumen de variantes, unas 50-100 mil para un exoma. Analizar cada variante individualmente resultaría extremadamente laborioso, por lo que es necesaria la aplicación de filtros para la priorización de las mismas. Usualmente, es deseable conocer la posición de las variantes en el genoma; si caen dentro de la región exónica un gen, si produce algún cambio en el marco de lectura, si introduce codones *stop* prematuros, etc.

La información recopilada proveniente de bases de datos biológicas está asociada a las variantes en el campo “INFO”, especificando los detalles de cada anotación en el encabezado.

Para anotar las variantes, se utilizó el programa de anotación estructural y predicción de efectos “SnpEff” versión 4.2 [150], que agrega al campo “INFO” del vcf, información pertinente a dónde se encuentra la variante (cerca o dentro de un gen, en la zona regulatoria, en un exón o intrón), el efecto que produce (cambios sinónimos o no sinónimos, pérdida de codón de iniciación o aparición de codón *stop* prematuros, pérdida de sitios dadores o aceptores de *splicing*, etc) y, le asigna un valor de efecto de patogenicidad *LOW*, *MEDIUM* o *HIGH* para zonas codificantes, o *MODIFIER* para variantes en zonas más inciertas.

De la misma manera, se utilizaron los siguientes programas de predicción automática de patogenicidad: Polyphen [154], SIFT [155] y Mutation Taster [156], entre otros.

Las predicciones automáticas de este tipo son esenciales para interpretar grandes conjuntos de datos que incluyen variantes genéticas nuevas o poco frecuentes, y permiten dirigir el análisis posterior a las variantes candidatas más prometedoras, ahorrando tiempo y recursos.

Predictores bioinformáticos

Polyphen-2

PolyPhen-2 [154] es una herramienta automática para la predicción del posible impacto de una sustitución aminoacídica en la estabilidad, estructura y función de una proteína humana utilizando consideraciones evolutivas estructurales y comparativas. La predicción se basa en una serie de características estructurales, filogenéticas y de secuencia que caracterizan la sustitución. A continuación, estima la probabilidad de que la mutación sin sentido sea perjudicial en función de una combinación de todas estas propiedades. Para una sustitución de aminoácidos dada, PolyPhen-2 extrae varias características de secuencia y estructura del sitio de sustitución, y las utiliza para “alimentar” a un clasificador probabilístico. Las características de PolyPhen-2 incluyen un alineamiento múltiple de secuencias de proteínas de alta calidad y un método de predicción que emplea una clasificación automática y es capaz de analizar grandes volúmenes de datos producidos por proyectos de secuenciación de próxima generación, gracias al soporte integrado para entornos informáticos de alto rendimiento. Es uno de los métodos más utilizados para la predicción del efecto de variantes.

SIFT

SIFT (por sus siglas en inglés *Sorting Intolerant from Tolerant*, [155]) es un programa que se basa únicamente en la secuencia aminoacídica de una proteína y utiliza resultados de identidad de secuencia para predecir el efecto de una sustitución no sinónima a la función de la proteína, es decir, si ésta es capaz de alterar el fenotipo. SIFT se aplica a bases de datos de variantes humanas y puede distinguir variantes patogénicas de polimorfismos neutros. Supone que los aminoácidos importantes se conservan en las familias de proteínas, por lo que cambios en las posiciones conservadas evolutivamente tienden a predecirse como nocivos. Por ejemplo, si una determinada posición en un alineamiento de una familia de proteínas sólo contiene el aminoácido isoleucina, se supone que la isoleucina es necesaria para la función de las mismas, por lo que se predice que un cambio de isoleucina por cualquier otro aminoácido resultará perjudicial. Si, por el contrario, una posición dada en un alineamiento múltiple contiene -en distintas proporciones- isoleucina, valina y leucina, (aminoácidos hidrofóbicos) entonces SIFT asume que, en efecto, esta posición solo puede contener aminoácidos con carácter hidrofóbico. En esta posición, los cambios a otros aminoácidos hidrofóbicos suelen predecirse como tolerados, pero se prevé que cambios a otros tipos de residuos (tales como cargados o polares) afecten la función de la proteína.

Para predecir si una sustitución de aminoácido en una proteína afectará la función de la proteína, SIFT considera entonces la posición en la que ocurrió el cambio y el tipo de sustitución aminoacídica. En cuanto a las proteínas, dada una secuencia aminoacídica, SIFT elige proteínas relacionadas -con alta identidad de secuencia- y obtiene un alineamiento múltiple. Basado en los aminoácidos que aparecen en cada posición del alineamiento, SIFT calcula la probabilidad de que un cambio de aminoácido en una posición sea tolerado, relativizado al aminoácido tolerado con mayor frecuencia. Si este valor normalizado es menor que un valor de corte, se predice que la sustitución será perjudicial.

Mutation Taster

Mutation Taster [156] es una herramienta gratuita, basada en la evaluación rápida del potencial patogénico de alteraciones en la secuencia de ADN. Integra información de diferentes bases de datos biomédicas y utiliza herramientas de análisis preestablecidas. Los análisis comprenden la conservación evolutiva, modificaciones en el sitio de empalme de los exones, pérdida de características de proteínas, y cambios que pueden afectar la cantidad de ARNm. Los resultados de las pruebas luego son evaluados por un clasificador que predice el potencial patogénico de la variante. Dependiendo de la naturaleza de la alteración, *Mutation Taster* elige entre tres modelos de predicción diferentes, que están dirigidos a alteraciones "silenciosas" o intrónicas, a alteraciones que afectan a un solo aminoácido o, a alteraciones que causan cambios complejos en la secuencia de aminoácidos.

Bases de datos

Una base de datos biológica constituye un repertorio de información, resultado de experimentos científicos, literatura publicada, tecnología de alto rendimiento y análisis computacional. Las bases de datos biológicas incluyen, en general, información sobre la función, estructura y localización (tanto celular como cromosómica) de genes, efectos clínicos de variantes, frecuencias poblacionales, clasificación de variantes, relación genotipo-fenotipo, así como similitudes de secuencias y estructuras biológicas.

Este tipo de bases de datos constituyen una herramienta esencial para almacenar, estructurar, organizar, actualizar y manipular datos biológicos. La variedad de éstos datos, así como también su rápido crecimiento, hacen a las bases de datos una herramienta clave.

Existe una gran variedad de bases de datos públicas, disponibles gratuitamente que pueden ser consultadas libremente y suelen constituir una referencia obligada en el área de la genética médica [166]. Como se explicó anteriormente, para la anotación funcional de las variantes se utiliza información de múltiples bases de datos, ya sean las que contienen

información asociadas a genes/proteínas (ENSEMBL, Uniprot), como aquellas que recopilan información clínica de cada variante conocida, o previamente reportada (OMIM, ClinVar, dbSNP). A continuación, se mencionan algunas de las más frecuentemente usadas y que fueron utilizadas en esta tesis.

ENSEMBL

El proyecto ENSEMBL [167] se inició en 1999, unos años antes de que se completara el genoma humano. Este proyecto es un trabajo de colaboración entre el European Bioinformatic Institute (EMBL-EBI) [168] y el Wellcome Trust Sanger Institute [169], con el objetivo de anotar automáticamente genomas de vertebrados y otras especies eucariotas, integrar esta anotación con otros datos biológicos, centralizar la información disponible y poner todo esto a disposición del público a través de la web. Los conjuntos de genes de ENSEMBL están basados en la evidencia de las bases de datos de secuencias incluyendo Uniprot [170] y RefSeq [171]. Cuando un transcripto de ENSEMBL es coincidente con un transcripto en RefSeq, los dos transcriptos se vinculan. Mientras que los modelos de genes ENSEMBL se anotan directamente en el genoma de referencia, RefSeq anota en las secuencias de ARNm. En este trabajo de tesis se utilizó ENSEMBL para asignar y analizar los transcriptos.

Uniprot

Uniprot (proveniente de *Universal Protein*) [170] es un repositorio central de datos gratuito sobre secuencias proteicas y anotaciones funcionales relacionadas. Las bases de datos que la componen son UniprotKB (Uniprot Knowledgebase), UniRef (Uniprot Reference Cluster) y Uniprot Archive (UniParc).

Esta interfaz nos permite buscar información funcional sobre proteínas con abundantes anotaciones. A su vez, UniprotKB está dividida en dos partes, según sus entradas hayan sido manualmente curadas (Swiss-Prot) o subidas de manera automática (TrEMBL).

La información de la proteína -expresada por el gen afectado- contenida en Uniprot es sumamente importante porque nos da herramientas para intentar comprender el efecto que puede tener una determinada variante sobre la función proteica y constituye un valor agregado a la hora de priorizar las variantes, y de asignarles significancia clínica. La información a la que podemos acceder en cada entrada de Uniprot está dividida en diferentes secciones. La primera proporciona información útil sobre la función de la proteína, principalmente conocimientos biológicos (actividad catalítica de la enzima, propiedades biofísicas y fisicoquímicas, vías metabólicas asociadas, localización de sitios activos).

También se informan los sitios de unión para cualquier grupo químico (coenzima, grupo prostético, etc.), se describe la interacción entre un aminoácido y otra entidad química, entre otras. De este modo es posible conocer si la variante analizada afecta a un aminoácido involucrado en los sitios importantes de la proteína, potencialmente perturbando su función. Otra sección interesante proporciona información sobre variantes conocidas previamente reportadas, y las enfermedades y fenotipos asociados a las mismas, que se extrae de la literatura científica y de la base de datos OMIM (descrita más abajo).

Asimismo, Uniprot provee información sobre la expresión proteica, especificidad tisular, expresión de isoformas, interacción con otras proteínas (o complejos de proteínas), estructura terciaria y secundaria. Por último, se encuentra una sección de secuencias donde se muestra la secuencia canónica de la proteína y todas las isoformas descritas en la literatura, incluyendo sus longitudes y pesos moleculares.

Online Mendelian Inheritance in Man (OMIM)

Como se mencionó anteriormente, OMIM [1] es un amplio compendio de genes humanos y sus fenotipos asociados. Cataloga todas las enfermedades que posean un componente genético y genera relaciones bibliográficas con otras bases de datos genómicas. A cada entrada OMIM se le asigna un número único de seis dígitos cuyo primer dígito indica si su herencia es autosómica, ligada al X, ligada al Y o mitocondrial. Además, las entradas se clasifican de acuerdo si contienen información sobre genes, fenotipos o ambos. Esto se denota por el símbolo que precede a un número “MIM”. Un asterisco (*) antes de un número de entrada indica un gen de secuencia conocida, el símbolo numeral (#) indica que es una entrada descriptiva de un fenotipo, un signo más (+) contiene tanto la descripción de un gen de secuencia conocida como el fenotipo asociado y el símbolo de porcentaje (%) describe un fenotipo mendeliano confirmado para el cual no se conoce la base molecular subyacente.

Cada entrada de OMIM tiene un resumen de texto completo de el fenotipo y/o gen determinado, y numerosos enlaces a otras bases de datos genéticas, referencias de PubMed, bases de datos de variantes genéticas y específicas de locus, MapViewer, GeneTests, ClinVar y grupos de apoyo para pacientes, entre otros. Esta base de datos es sumamente importante el trabajo de genómica clínica ya que es la que utilizamos principalmente para evaluar cuán buena es la coincidencia entre el fenotipo observado en el paciente, y el fenotipo esperado (según la descripción de OMIM) para un paciente con una variante dada que estamos evaluando como potencial candidata a ser diagnóstica.

dbSNP

La base de datos de polimorfismos de un solo nucleótido es un archivo público gratuito que contiene información sobre la variación genética dentro y entre diferentes especies, desarrollada por el Centro Nacional de Información Biotecnológica en colaboración con el Instituto Nacional de Investigación del Genoma Humano. dbSNP (dbSNP: *the NCBI database of genetic variation*) contiene variaciones humanas de un solo nucleótido, microsatélites e inserciones y delecciones pequeñas, junto con *links* a publicaciones, frecuencia poblacional, consecuencia molecular e información de mapeo genómico, tanto para polimorfismos como para variantes clínicas. Desde su creación en septiembre de 1998, la base de datos dbSNP ha servido como depósito central y público de variantes genéticas. Una vez identificadas y catalogadas dichas variaciones en la base de datos, otros laboratorios pueden utilizar la información de la secuencia en torno al polimorfismo y las condiciones experimentales específicas para otras aplicaciones de investigación. Al igual que todos los recursos del NCBI, los datos de dbSNP están disponibles de forma gratuita y en una variedad de formas. Si bien en los albores de la presente tesis, dbSNP era ampliamente utilizada por ejemplo para conocer la frecuencia poblacional de una variante en estudio, en la actualidad para el estudio de genomas humanos, su uso ha sido reemplazado por bases de datos poblacionales, como ExAC y GnomAD.

Clinical Variants (ClinVar)

Esta plataforma fue desarrollada para facilitar el acceso a la interpretación de la relación entre variantes y fenotipos médicaamente relevantes [163]. La base de datos está estrechamente relacionada con dbSNP y dbVar, que aportan información sobre la ubicación de las variantes en el genoma humano, y se basa en las descripciones fenotípicas incluidas en MedGen (<http://www.ncbi.nlm.nih.gov/medgen>). El contenido de ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) se divide en cinco categorías principales: remitente, variante, fenotipo, interpretación y evidencia. La combinación única entre remitente, variación y fenotipo determina una unidad de registro y se le asigna una entrada con el formato SCV000000000.0 (SCV). El remitente puede actualizar la información enviada en cualquier momento, en cuyo caso se le asigna una nueva versión. Para facilitar la evaluación de la relevancia médica de cada variante reportada, ClinVar agrega información de otras bases de datos de NCBI e informa si hay interpretaciones clínicas contradictorias. Los datos están disponibles en múltiples formatos que incluyen presentaciones con detalles del fenotipo, interpretación de significado funcional y clínico, metodología utilizada para capturar las variantes y evidencia de respaldo. En este sentido, ClinVar representa los distintos valores

de significancia clínica con un código numérico (ASN.1), de acuerdo con las recomendaciones del Colegio Americano de Genética Médica y Genómica [77].

Tabla 2. Numeración correspondiente a cada una de las categorías incluidas en ClinVar.

Categoría		
0	Desconocido o de Significancia Incierta	Unknown, Uncertain significance
1	Sin respaldo (incluye a los casos donde la información respaldatoria no está disponible o el remitente es desconocido)	Untested
2	Benignas	Benign
3	Probablemente benignas	Likely benign
4	Probablemente patogénicas	Likely pathogenic
5	Patogénicas	Pathogenic
6	Responden a drogas	Drug-response
7	Histocompatibilidad	Histocompatibility
255	Otros	Other

Es importante notar que cada variante puede haber sido reportada con más de una categoría, y si los remitentes no están de acuerdo con la interpretación del significado clínico, el registro es catalogado como con reportes en conflicto. Dicho de otro modo, si un remitente envía una variante y la clasifica como benigna, y otro remitente la clasifica como patogénica, ClinVar la terminará reportando como de “registro en conflicto”. Por otro lado, las entradas se clasifican también según la forma de obtención de datos (por ejemplo, pruebas clínicas, evaluación e investigación de la literatura) y el nivel o estado de revisión (presentación única, panel de expertos, etc.). De esta forma, ClinVar brinda a todos los usuarios acceso a un conjunto más amplio de interpretaciones clínicas y a un sitio integral que permite obtener datos actuales e históricos de las variantes. En genómica clínica y en esta tesis, utilizaremos ClinVar principalmente como un catálogo de referencia de variantes conocidas y con una clasificación ACMG determinada. De este modo, para aquellas variantes encontradas en un paciente particular, que ya se encuentren catalogadas en ClinVar, tendremos una clasificación independiente que sustente (o no) la derivada de nuestro análisis.

Infevers

Infevers [172] (*Internet Fevers*; <https://infevers.umai-montpellier.fr/web/>) es un sitio web dedicado al registro de variantes responsables de enfermedades autoinflamatorias hereditarias que se creó en 2001 y desde entonces no ha dejado de evolucionar. Recopila información actualizada sobre variantes responsables de trastornos inflamatorios hereditarios: es decir, Fiebre Mediterránea Familiar (FMF), TRAPS (Síndrome Asociado al Receptor 1A del TNF), HIDS (Síndrome de HiperIgD), MWS (Síndrome de *Muckle-Wells*)/FCU (Urticaria Fría Familiar)/CINCA (Síndrome Cutáneo y Articular Neurológico Infantil Crónico), entre otros. Los colaboradores, médicos o investigadores envían sus variantes novedosas a través de un formulario. En función de la enfermedad de la que se trate, se solicita automáticamente a un miembro del consejo editorial que examine y valide las nuevas propuestas a través de una interfaz web segura especial. Si se acepta, la nueva mutación está disponible en el sitio web de Infevers y el descubridor, al que se informa por correo electrónico, es acreditado con su nombre y la fecha del descubrimiento. El portal INFEVERS proporciona a investigadores y médicos un lugar de acceso común a la información sobre enfermedades autoinflamatorias, lo que permite obtener una visión general rápida de los defectos genéticos descubiertos hasta el momento. Además, es interactivo y ampliable en función de los últimos genes descubiertos. Actualmente, incluye 48 genes y contiene más de 2600 variantes de secuencia. Para cada gen, el portal ofrece: 1) una lista tabular de las variantes que pueden ordenarse según varios parámetros; 2) un gráfico de genes que ofrece una representación esquemática de las variantes a lo largo del gen; 3) un análisis estadístico de los datos según el fenotipo, el tipo de alteración y la localización de la mutación en el gen; 4) las secuencias de ADNc y ADNg de cada gen, que muestran los cambios de nucleótidos a lo largo de la secuencia, con un código basado en colores que resalta los dominios del gen, el codón *start* y *stop* de terminación; y 5) un menú de "descarga" que pone a disposición de los usuarios todas las tablas y figuras. La base de datos se curó para cumplir con las directrices de nomenclatura del Comité de Nomenclatura Genética de HUGO (HGNC) y del HGVS.

Registro Eurofever

La base de datos Infevers otorga información sobre variantes encontradas en genes asociados con enfermedades autoinflamatorias, aunque sin proporcionar asociaciones genotipo-fenotipo, sino sólo el fenotipo clínico del primer o los primeros pacientes descritos para cada variante. Las enfermedades autoinflamatorias constituyen una familia en expansión de afecciones diferentes. Sin embargo, existe un grado considerable de solapamiento en su etiología, manifestaciones clínicas y tratamiento. El Grupo de Trabajo de Enfermedades

Autoinflamatorias de PReS, en colaboración con PRINTO (por sus siglas en inglés, *Paediatric Rheumatology INternational Trials Organisation*), ha creado el registro internacional Eurofever [84] (<https://www.printo.it/eurofever/registry>), de enfermedades autoinflamatorias con el objetivo de desarrollar asociaciones genotipo-fenotipo observadas en pacientes con fiebres recurrentes hereditarias. El principal objetivo consistió en recopilar información sobre la presentación clínica, edad media de inicio, evolución de la enfermedad (recurrente o crónica), duración media de los episodios de fiebre, manifestaciones asociadas y atípicas, complicaciones médicas y respuesta al tratamiento de pacientes afectados por las principales enfermedades autoinflamatorias. Por esta razón, esta herramienta es complementaria a la base de datos Infevers.

En 2019 se propuso un nuevo conjunto de criterios de clasificación basados en la evidencia Eurofever/PRINTO (EPCC), en el contexto del proyecto INSAID fundado por E-rare, para todas las fiebres recurrentes hereditarias. Por primera vez, una combinación de variables tanto genéticas como clínicas permitió establecer criterios de clasificación específicos. En particular y como ejemplo, permitió identificar dos grupos distintos de pacientes con Fiebre Mediterránea Familiar (FMF), que difirieron en características clínicas, enfoque terapéutico y respuesta al tratamiento. Los pacientes EPCC+ evidenciaron características típicas de la FMF, mientras que los pacientes EPCC- presentaron un fenotipo más variable, con un menor porcentaje de respuesta a la colchicina [173].

Protocolo de priorización de variantes

En línea con nuestro principal objetivo, buscamos establecer un protocolo general para facilitar el proceso de priorización y análisis de variantes, basándonos en la información previamente asociada a las mismas durante el proceso de anotación. El mismo fue diseñado durante el desarrollo de la tesis en base a la experiencia previa y a las características esperables de variantes asociadas a enfermedades mendelianas poco frecuentes. El protocolo pretendió sistematizar el análisis de los exomas aplicando secuencialmente grupos de filtros (descritos en la Figura 8), que permitieran comenzar la búsqueda de variantes relevantes siendo lo más estrictos posible (y obteniendo un número reducido de variantes que pudieran ser fácilmente estudiadas en profundidad), y luego ampliando la búsqueda en la medida en que no se obtuvieran variantes candidatas con clara relevancia clínica.

La aplicación de filtros sucesivos fue adaptada a cada caso particular, teniendo en cuenta: los genes candidato seleccionados, la posible presencia de variantes previamente reportadas como patogénicas o probablemente patogénicas en bases de datos, el potencial impacto de las variantes a nivel molecular, el modelo de herencia sospechado -en consecuencia, la cigosidad esperada de las variantes-, y su correspondiente frecuencia (por

ejemplo, se espera que la frecuencia sea mayor para variantes que en homocigosis están asociadas a enfermedades recesivas, ya que en heterocigosis no ocasionan la enfermedad y los individuos heterocigotas suelen ser portadores sanos, mientras que se espera que la frecuencia de variantes asociadas a enfermedades dominantes sea menor y extremadamente baja -o nula- en el caso de variantes de novo). Para la aplicación iterativa de filtros, se utilizó la herramienta informática B_Platform, desarrollada en conjunto por el grupo de investigación del director de tesis y Bitgenia, en la cual se sube el VCF anotado (la lista de alrededor de 100 mil variantes identificadas en un exoma y toda la información asociada a cada una de ellas). Este software funciona dentro de un servidor accesible por la red que facilita consultas y aplicación de filtros iterativos definidos por el usuario sobre los datos de variantes exómicas. Adicionalmente, permite resaltar y/o consultar en bases de datos externas la información dispuesta en la anotación para cada variante. El diseño de los filtros, el orden iterativo y sucesivo de aplicación de los mismos, y la valoración de las características analizadas para cada variante, que es usuario dependiente, son el corazón del protocolo general de priorización y análisis de variantes.

A partir de la Historia Clínica de cada caso particular, y con el aporte de los médicos postulantes, se seleccionaron genes candidatos a albergar variantes patogénicas que pudieran explicar cada patología. En el caso del Proyecto PAMPA, el panel de genes fue preestablecido y actualizado anualmente, en conjunto con el *Advisory Board*. Los paneles de genes analizados se mencionan en los capítulos correspondientes.

Cada variante fue categorizada, en primer lugar, en función de su impacto a nivel molecular. Se le asignó un impacto **alto** a aquellas variantes que resultaran en ganancia o pérdida de codones de inicio o finalización de la traducción y/o cambios en el marco de lectura. El impacto se consideró **moderado** cuando las modificaciones a nivel proteico involucraron cambios no sinónimos de un único aminoácido y/o pequeñas inserciones o delecciones que mantuvieron el marco de lectura. El resto se consideró de **bajo** impacto. Para cada una de las variantes seleccionadas se analizó luego su frecuencia poblacional (si la hubiera), la evidencia previa de asociaciones clínicas, (para lo cual se utilizó la calificación de ClinVar considerando que valores de ClinSig 4 o 5, se corresponden con las categorías “*Probably pathogenic*” y “*Pathogenic*” respectivamente), el efecto fenotípico previsto, el modelo de herencia para variantes ya reportadas en ese gen según OMIM y, si fuera posible, el efecto sobre la estructura y función proteica (predicciones de algoritmos bioinformáticos y modelado de estructuras proteicas).

En primer lugar, nuestro protocolo de priorización sugirió comenzar el análisis con la búsqueda de variantes con reportes previos de patogenicidad en bases de datos de asociaciones clínicas (por ejemplo, ClinVar). En segundo lugar, de no obtenerse resultados relevantes con la aplicación del primer conjunto de filtros, o con el objetivo de descartar la

presencia de otra/s variante/s potencialmente interesante/s, se recomendó continuar con la búsqueda de variantes novedosas (sin respaldo de asociación a enfermedades en bases de datos clínicas) presentes en los genes candidatos, que tuvieran alto impacto y baja frecuencia poblacional, de acuerdo con el modelo de herencia (para esto se diseñó el filtro compuesto correspondiente y se lo implementó en el software de análisis). En el caso de no detectarse variantes de interés con la aplicación de estos filtros, se suele terminar el análisis con la búsqueda de variantes novedosas de impacto alto o moderado por fuera del panel de genes propuesto, pero en genes relacionados indirectamente con la patología. Los mismos pueden obtenerse fácilmente a partir de paneles comerciales preexistentes asociados, por ejemplo, a enfermedades similares o diagnósticos diferenciales, pueden ser facilitados o sugeridos por los profesionales médicos a cargo o pueden encontrarse utilizando programas específicos que asocian signos y síntomas (fenotipo) con potenciales genes de interés (con un score de probabilidad). Ejemplos de estos programas lo constituyen VarElect [186] o Phenomizer [185].

Los resultados obtenidos a partir de la aplicación de grupos de filtros se analizan de forma detallada y aquellas variantes que presentan características de interés son las priorizadas, y consideradas como variantes candidatas. Estas variantes son luego investigadas individualmente utilizando bases de datos y herramientas de evaluación funcional (que incluyen los proporcionados por NCBI, ENSEMBL, Varsome [174], UniProt, LitVar [175], Infevers [176], Eurofever Project [84] y LOVD [177], entre otros) e incluidas en un reporte final que se le proporciona al médico a cargo.

Todos los casos mencionados en la presente tesis fueron sometidos al protocolo de priorización de variantes mencionado previamente, con la finalidad de facilitar este proceso descartando variantes que no cumplieran con los requisitos para ser consideradas como “patogénicas”, “posiblemente patogénicas” o de “significado incierto” según los criterios del *American College of Medical Genetics and Genomics* (ACMG), y de enriquecer los resultados con aquellas variantes candidatas a ser las causales de la sintomatología presentada.

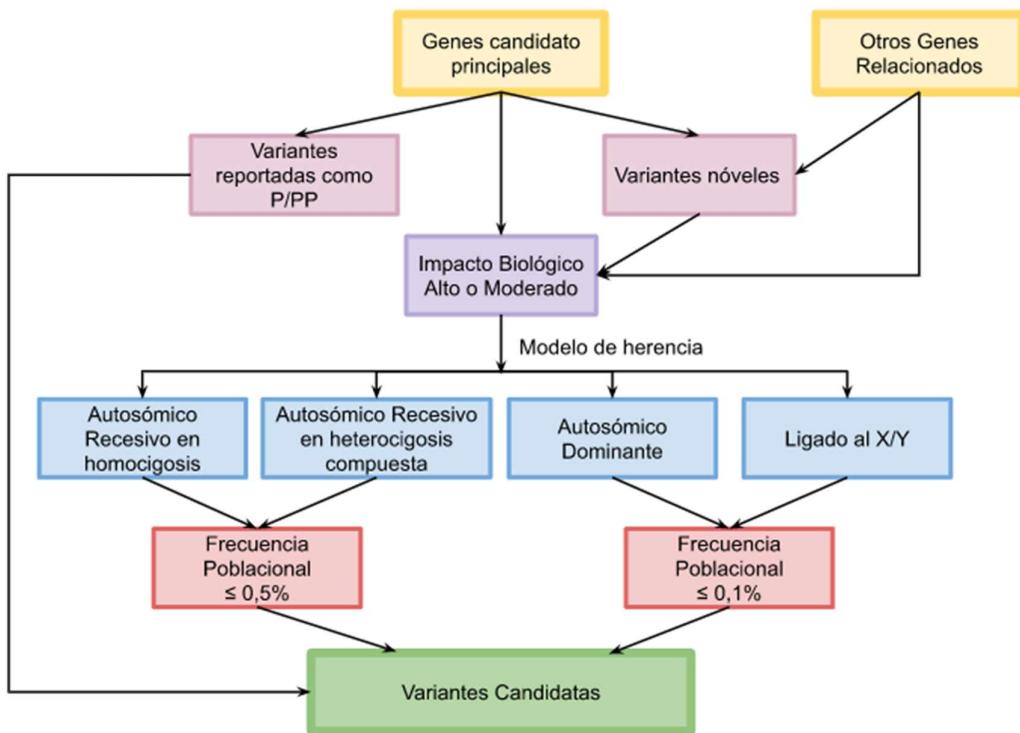


Figura 8: Esquema de priorización de variantes. En base a la historia clínica, se seleccionaron genes candidatos (Caso I y II). En primer lugar se buscaron variantes reportadas previamente como patogénicas (P) o probablemente patogénicas (PP) en ClinVar dentro de los mismos. Luego, se analizaron variantes de alto impacto y baja frecuencia poblacional, según el modelo de herencia/cigosidad esperado/a ó nóveles en los genes candidatos. De no encontrarse resultados de relevancia, se continuó con la búsqueda de variantes de impacto alto o moderado en genes indirectamente relacionados a la patología (que pudieran explicar parcialmente los signos y síntomas evidenciados por el paciente en estudio).

- Análisis de variantes estructurales

Aunque los llamadores de variantes son eficaces a la hora de descubrir variantes de pequeña longitud como SNP o pequeños Indels, resultan ineficaces en detectar variantes de mayor alcance, como en el caso de grandes delecciones, duplicaciones, inversiones y translocaciones. En esta tesis y, debido a la técnica de captura y profundidad de secuenciación utilizada, no fue recomendable analizar la presencia de variantes estructurales y únicamente se llevó a cabo la búsqueda de CNVs a través de la utilización de algoritmos especializados (DECoN [178, 179] y OncoCNV [180]) en pocos casos en los que los médicos postulantes o integrantes del *Advisory Board* así lo solicitaron. **No se encontraron variantes de número de copias en ninguno de esos casos.**

- Validación de Resultados

Aunque un estudio reciente generó discusión en la comunidad científica ya que sugirió que los resultados de la aplicación de técnicas de secuenciación masiva deberían ser considerados imperfectos, pero altamente fiables por sí solos [181], cuando se encuentran variantes potencialmente patogénicas, los protocolos de trabajo actuales sugieren verificar la presencia de la/s misma/s en el paciente (o en el trío, de ser posible), mediante la aplicación de otra tecnología de secuenciación, idealmente Sanger (*gold standard*). Es importante tener en cuenta las limitaciones (baja calidad de las secuencias en los primeros 15-40 pb y la incapacidad de distinguir SNPs en segmentos largos > 1000 pb), el tipo de variante y la proporción de alelos mutados (especialmente en la detección de mutaciones somáticas), de modo de abordar de forma óptima la validación [182]. El análisis en los padres permite establecer, además, de manera precisa, el modelo de herencia de la variante. Asimismo, la secuenciación por Sanger continúa siendo también la técnica de elección para el estudio de variantes conocidas en rutinas de diagnóstico prenatal.

La evaluación de la segregación familiar de las variantes encontradas en los casos I y II se llevó a cabo y los resultados se detallan en los apartados correspondientes, mientras que la validación por Sanger se terciarizó y llevó a cabo en algunos casos de PAMPA en los cuales pudo obtenerse financiamiento.

Para la familia del Caso I, los *primers* fueron diseñados como muestra la Tabla 3, de forma de capturar la región flanqueante a las variantes detectadas en los exones 4 y 16 y fueron encargados a Macrogen. También se indica la temperatura a la que se realizó la PCR partiendo del ADN genómico de cada integrante de la familia y la longitud en pb del producto esperado.

Los pasos seguidos se resumen a continuación:

1. Ampliación mediante PCR de las regiones flanqueantes a las variantes investigadas en el gen *AFG3L2*.
2. Preparación de las reacciones de secuenciación de ADN y separación de bandas por electroforesis capilar.
3. Envío de tacos conteniendo los productos de PCR de interés para su secuenciación en Macrogen.
4. Análisis bioinformático de las secuencias obtenidas por comparación con las secuencias de referencia utilizando FinchTv.

MÉTODO: PCR / SECUENCIACIÓN EN ABI310 (SANGER)

Tabla 3. Primers Fw y Rv diseñados y especificaciones de PCR. Se utilizaron los siguientes pares de primers para validar la presencia o ausencia de las variantes del exón 4 y 16, detectadas en el probando II.5 por secuenciación exómica, en el resto de los familiares.

Exón	Temperatura PCR	Primer Fw (5' a 3')	Primer Rv (5' a 3')	Longitud del producto de PCR
4	58°C	AGC CTC CCT GAT TGG TAA GC	CTC TGG GCC ACC ATC ATA ACC	340 pb
16	60°C	AGG TGT TAT GCC TGG GCT TC	ACA TCA GAA CGA ACG GAC CC	520 pb

Para el Caso II, la validación de la segregación de las variantes fue realizada en el Laboratorio de Genética humana y Enfermedades infecciosas, INSERM, París, Francia. Para los pocos casos de PAMPA validados, la secuenciación por Sanger fue terciarizada al Laboratorio VZ.

Resultados

Caso I:

Paciente con sospecha clínica de Neurodegeneración asociada a Pantotenato Quinasa (PKAN)

Introducción

Se presentó el caso particular de un paciente (II.5) de 19 años al momento de la consulta (año 2018), con una alteración de la marcha de progresión lenta y dificultades de aprendizaje, siendo el quinto de 8 hermanos, hijo de padres no consanguíneos. La exploración física de los hermanos II.1, II.4, II.5 y II.6 reveló ataxia cerebelosa, signos piramidales, nistagmo y disminución de la sensación vibratoria en miembros inferiores. El paciente II.1, varón de 27 años, presentó los signos más graves, además de disartria, marcha espástica, cambios de comportamiento y estrabismo (Figura 9).

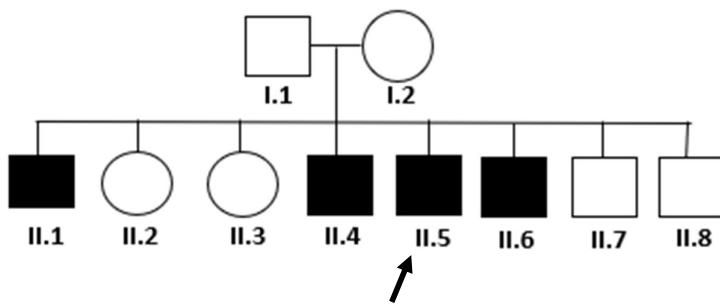


Figura 9: Familiograma. Se grafica la historia familiar de la que se dispone para favorecer la determinación del modelo de herencia autosómico recesivo.

Dada la evolución clínica y los estudios realizados hasta ese momento, el profesional a cargo propuso como diagnóstico clínico presuntivo Neurodegeneración asociada a Pantotenato Quinasa (PKAN, por sus siglas en inglés). PKAN constituye una enfermedad poco frecuente de herencia autosómica recesiva que suele presentarse en las primeras décadas de vida y se caracteriza por depósito de hierro en los ganglios basales, causando disfunción extrapiramidal progresiva. Las manifestaciones clínicas suelen ser diversas y se caracterizan por: distonía progresiva, alteración del habla, retinopatía pigmentaria y déficit cognitivo. Sin embargo, alrededor del 25% de los pacientes presentan un curso atípico, caracterizado por un inicio más tardío de los síntomas (> 10 años), trastornos psiquiátricos y progresión gradual de la enfermedad [183]. El signo clásico de “ojo de tigre” (EOT, por sus

siglas en inglés) se evidencia en las imágenes ponderadas en T2, en la resonancia magnética cerebral y hace referencia a una alta intensidad de señal central en la cara anteromedial del globo pálido y a una baja intensidad de señal circundante debido al exceso de hierro. Aunque el signo de EOT se considera casi patognomónico de la neurodegeneración asociada a pantotenato quinasa, también se ha descrito en otros trastornos e incluso en un adulto sano [184].

La resonancia magnética (RM) cerebral de los 4 hermanos afectados reveló el signo de ojo de tigre en las imágenes ponderadas en T2, **pero no se detectó la acumulación excesiva de hierro esperada en T2*** ni en la **imagen de susceptibilidad (SWI)** -una nueva técnica que utiliza las diferencias de susceptibilidad magnética del tejido para generar un contraste único- (Figura 10). Los resultados del examen físico y la resonancia magnética cerebral del resto de los hermanos y de los padres resultaron normales.

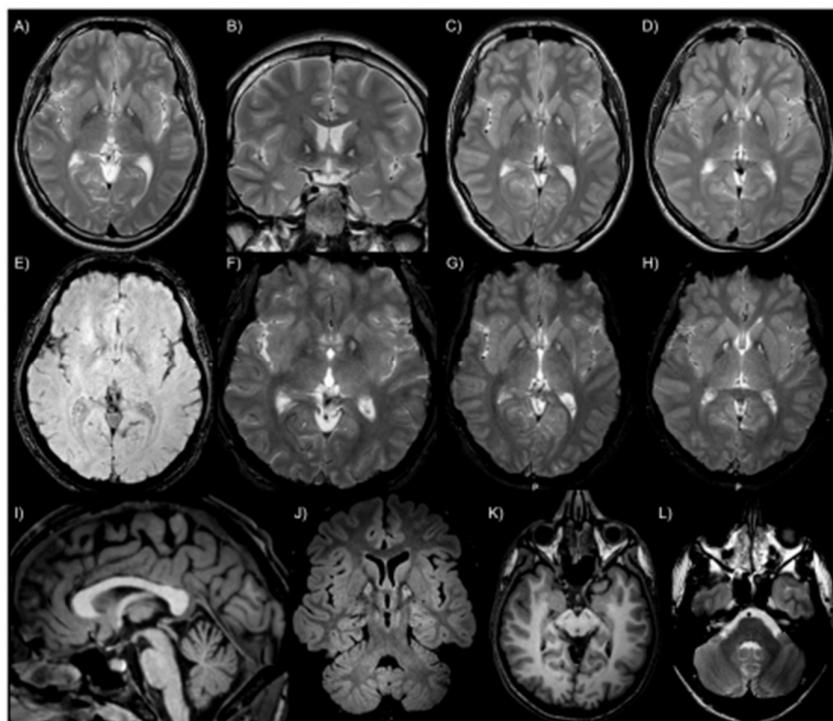


Figura 10: Resonancias magnéticas cerebrales. Señal hiperintensa central de ambos GP en imágenes axiales ponderadas en T2 (A = paciente II.1, B = paciente II.4, C = paciente II.5, D = paciente II.6) sin señal hipointensa circundante sobre susceptibilidad (E = paciente II.1) o imagen ponderada en T2* (F = paciente II.4, G = paciente II.5, H = paciente II.6). RM cerebral del paciente II.5: C) señal hiperintensa del borde posterior del núcleo lenticular en la imagen axial potenciada en T2; I) atrofia del vermis en la imagen sagital ponderada en T1; J) señal hiperintensa de GP y SN en el plano coronal oblicuo; K) señal hipointensa de SN en imagen axial ponderada en T1; L) leve señal hiperintensa de ambos núcleos dentados en la imagen axial ponderada en T2.

Referencias: GP= globo pálido, SN= sustancia negra.

En la Tabla 4 se detalla la clínica evidenciada y los hallazgos en la resonancia magnética nuclear de los cuatro pacientes.

Tabla 4. Signos, síntomas y hallazgos en la RMN de los pacientes estudiados. Referencias: TG= triglicéridos, GP= globo pálido, SN= sustancia negra, LN = núcleo lenticular. DN = núcleo dentado, NCS / EMG = estudio de conducción nerviosa / electromiografía. EAS: edad de aparición de síntomas, EA: edad actual.

Paciente/ Género	EAS/EA	Signos piramidales	Signos Cerebelares	Otras manifestaciones clínicas	Hallazgos en la RMN
II.1/ masculino	8 años/ 30 años	Clonus de tobillo bilateral, escotoma en altas frecuencias bilateral, clonus bilateral por debajo de la rodilla, marcha espástica.	Ataxia apendicular, TG alterados, nistagmo evocado por la mirada, disartria.	Estrabismo, dificultad de aprendizaje, cambios de comportamiento *, ↓ Sensación vibratoria en miembros inferiores, Romberg (+)	GP: EOT-like ^{w/o} /↓ señal en T2*/SWI SN: ↓ señal T1, ↑ señal T2-FLAIR, ^{w/o} /↓ señal T2*/SWI Cerebelo: atrofia leve del vermis superior y hemisferios cerebelosos adyacentes.
II.4/ masculino	Desconocido/ 25 años	Ataxia de miembros superiores derechos, escotoma en altas frecuencias del oído derecho, clonus de rodilla bilateral	LUL ataxia, TG alterados, nistagmo evocado por la mirada.	↓ Sensación vibratoria en miembros inferiores bilateral, Romberg (+), NCS / EMG: polineuropatía axonal de miembros superiores e inferiores.	GP: EOT-like ^{w/o} /↓ señal en T2*/SWI SN: ↓ señal T1, ↑ señal T2-FLAIR, ^{w/o} /↓ señal T2*/SWI Cerebelo: atrofia leve del vermis superior y hemisferios cerebelosos adyacentes.
II.5/ masculino	8 años/ 21 años	Clonus de rodilla bilateral.	Ataxia de miembros inferiores, TG alterados, nistagmo evocado por la mirada.	Dificultad de aprendizaje *, ↓ Sensación vibratoria en miembros inferiores.	GP: EOT-like ^{w/o} /↓ señal en T2*/SWI LN: ↑ señal en T2/FLAIR borde posterior SN: ↓ señal T1, ↑ señal T2-FLAIR, ^{w/o} /↓ señal T2*/SWI Cerebelo: atrofia leve del vermis superior y hemisferios cerebelosos adyacentes. DN: leve señal hiperintensa T2-FLAIR
II.6/ masculino	Desconocido/ 19 años	Clonus de rodilla y tobillo derecho.	Ataxia apendicular, TG alterados, nistagmo evocado por la mirada.	↓ Sensación vibratoria en miembros inferiores, pes cavus bilateral, Romberg (+)	GP: EOT-like ^{w/o} /↓ señal en T2*/SWI SN: ↓ señal T1, ↑ señal T2-FLAIR, ^{w/o} /↓ señal T2*/SWI. Cerebelo: atrofia leve del vermis superior y de los hemisferios cerebelosos adyacentes.

El diagnóstico de PKAN se basa en una evaluación clínica exhaustiva y en la realización de varias pruebas especializadas. El examen neurológico está dirigido a evaluar distonía, rigidez, movimientos anormales, espasticidad y el desarrollo del habla y lenguaje. También es apropiado el examen oftalmológico para evaluar una posible retinopatía. El diagnóstico se confirma con la realización de un examen genético, con el que se esperan detectar variantes en el gen *PANK2*. Aproximadamente el 95% de los pacientes son portadores de dos variantes en heterocigosis identificables en este gen y sólo el 5% poseen una única mutación en homocigosis. En algunos casos, la secuenciación dirigida del gen *PANK2* es solicitada por el médico. Sin embargo, también existen opciones de paneles multigénicos o secuenciación del exoma completo para abarcar el análisis de genes asociados con otras enfermedades neurodegenerativas por acumulación cerebral de hierro (NBIA), ayudando a descartar -o confirmar- diagnósticos diferenciales. Las NBIA se definen como un grupo de trastornos extrapiramidales progresivos con evidencia radiográfica de acumulación focal de hierro en el cerebro, generalmente en los ganglios basales. Los distintos tipos de NBIA se incluyen en la Tabla 5.

Teniendo en cuenta la gran variedad de síndromes que podrían explicar total o parcialmente los signos y síntomas evidenciados por el probando II.5, el médico solicitó la secuenciación de su exoma completo, el cual fue financiado por el Hospital del Cruce.

Tabla 5. Tipos de enfermedades asociadas con Neurodegeneración con acumulación cerebral de hierro (NBIA). Referencias: MDH = Modelo de herencia; AD = Autosómico dominante; AR = Autosómico recesivo; XL = Ligada al X.

Gen	Nombre de la condición	MDH	Progresión
<i>ATP13A2</i>	Síndrome de Kufor-Rakeb	AR	NBIA de aparición tardía, lentamente progresiva con inicio después de la primera década de vida.
<i>C19orf12</i>	Neurodegeneración asociada a las proteínas de la membrana mitocondrial (MPAN)	AR	Inicio temprano (durante la 1 ^a década) o aparición tardía (después de la 1 ^a década), NBIA lentamente progresiva.
<i>CP</i>	Aceruloplasminemia	AR	NBIA de aparición tardía, lentamente progresiva con inicio después de la primera década de vida.
<i>CoASY</i>	Neurodegeneración asociada a la proteína COASY (CoPAN)	AR	NBIA de inicio temprano, lentamente progresivo con inicio durante la primera década de vida.

DCAF17	Síndrome de Woodhouse-Sakati	AR	NBIA de inicio temprano, lentamente progresivo con inicio durante la primera década de vida.
FA2H	Neurodegeneración asociada a la hidroxilasa de ácidos grasos (FAHN)	AR	NBIA de inicio temprano, lentamente progresivo con inicio durante la primera década de vida.
FTL	Neuroferritinopatía	AD	NBIA de aparición tardía, lentamente progresiva con inicio después de la primera década de vida.
PANK2	Neurodegeneración asociada a pantotenato quinasa (PKAN) atípica	AR	NBIA de aparición tardía, lentamente progresiva con inicio después de la primera década de vida.
	Neurodegeneración asociada a pantotenato quinasa (PKAN) clásica		NBIA de inicio temprano rápidamente progresivo con inicio durante la primera década de vida.
PLA2G6	Distrofia neuroaxonal atípica	AR	NBIA de inicio temprano, lentamente progresivo con inicio durante la primera década de vida.
	Distrofia neuroaxonal infantil		NBIA de inicio temprano, rápidamente progresivo con inicio durante la primera década de vida.
	Distonia-parkinsonismo asociado a <i>PLA2G6</i>		NBIA de aparición tardía, lentamente progresiva con inicio después de la primera década de vida.
WDR45	Neurodegeneración asociada a proteínas de hélice beta (BPAN)	XL	NBIA de inicio temprano, lentamente progresivo con inicio durante la primera década de vida.
Desconocido	NBIA Idiopática	--	NBIA de aparición tardía, lentamente progresiva con inicio después de la primera década de vida.

Como el objetivo final consistió en encontrar variantes con alta probabilidad de ser patogénicas, resultó imprescindible buscar genes candidatos presentes en paneles comerciales preexistentes e interiorizarnos con casos similares mediante la lectura de bibliografía específica, a fin de generar una base de datos propia que nos permitiera comparar los datos obtenidos del análisis del paciente con los reportados previamente. Puede asumirse

que las variantes previamente publicadas representan un set de datos independientes, así como también permiten identificar la presencia de “hot-spots” mutacionales, exones o dominios proteicos mayormente afectados, y los tipos de efectos fenotípicos más comúnmente observados en los pacientes. De esta forma, dada la historia clínica, la hipótesis del profesional y teniendo en cuenta el familiograma y la fuerte sospecha de enfermedad recesiva, pudimos establecer la siguiente lista de genes a analizar (Tabla 6).

Tabla 6. Paneles de genes seleccionados. Se construyeron en función de la sospecha clínica y los signos y síntomas del paciente y sus familiares afectados.

Panel	Genes contemplados
Panel para NBIA	<i>ATP13A2, C19orf12, CP, COASY, DCAF17, FA2H, FTL, PANK2, PLA2G6, WDR45.</i>
Panel asociado a paraparesia espástica	<i>ABCD1, AFG3L2, ALDH18A1, ALS2, AP4B1, AP4E1, AP4M1, AP4S1, AP5Z1, ARG1, ATAD3A, ATL1, ATP2B4, B4GALNT1, BICD2, BSCL2, BTD, C19orf12, CAPN1, COASY, CTNNB1, CYP27A1, CYP2U1, CYP7B1, DDHD1, DDHD2, DSTYK, ERLIN2, FA2H, FARS2, FXN, GALC, GBA2, GBE1, GCH1, GJC2, HACE1, HSPD1, IBA57, KDM5C, KIF1A, KIF1C, KIF5A, KLC2, L1CAM, L2HGDH, MARS2, NIPA1, NT5C2, PAH, PLP1, PNPLA6, REEP1, RTN2, SACS, SETX, SLC16A2, SLC25A15, SLC33A1, SPAST, SPG11, SPG21, SPG7, SPR, TECPR2, TH, VAMP1, VPS37A, WASHC5, ZFYVE26, ZFYVE27, MTRFR, DARS1, RARS1, SPG20 (SPART).</i>
Genes priorizados de acuerdo a los signos y síntomas del paciente²	<i>AARS2, ABCB7, ABCD1, ABHD12, ACOX1, ADGRG1, AFG3L2, AHI1, ALS2, ANO10, AP5Z1, APP, APTX, ATCAY, ATM, ATN1, ATP13A2, ATP1A3, ATP6AP2, ATP8A2, ATXN1, ATXN2, ATXN3, ATXN7, ATXN8, ATXN8OS, AUH, BSCL2, C19orf12, C9orf72, CA8, CACNA1A, CC2D2A, CEP290, CLN6, COASY, CP, CTC1, CTDP1, CYP27A1, CYP7B1, DARS2, DCAF17, DDC, DNAJC6, DNM1L, DNMT1, EIF2B5, ELOVL4, ERLIN2, FA2H, FAM126A, FBXO7, FGF14, FTL, FUS, FXN, GAD1, GAN, GBA1, GBA2, GBE1, GCH1, GJB1, GJB2, GJC2, GRID2, GRM1, GRN, HSD17B4, HTRA1, HTT, IFT140, IFT172, INPP5E, ITM2B, ITPR1, JPH3, KCNA1, KCNC3, KIF1A, KIF1C, L2HGDH, LMNB1, LRRK2, MAG, MAPT, MARS2, MECP2, MFN2, MRE11, MTFMT, MTPAP, NAGA, NALCN, NDUFS4, NDUFV1, NOP56, NPC1, NPC2, OPA3, OPTN, PANK2, PARK7, PAX6, PDGFB, PDGFRB, PDYN, PIK3R5, PINK1, PLA2G6, PLP1, PMPCA, PNKP, PNPLA6, POLG, POLR3A, POLR3B, PPARGC1A, PRKCG, PRNP, PSEN1, PSEN2, RNF216, RPGRIP1L, RRM2B, RUBCN, SACS, SCN1A, SCN2A, SCN8A, SDHA, SDHD, SETX, SIL1, SLC17A5, SLC19A3, SLC1A3, SLC20A2, SLC25A15, SLC2A1, SLC6A3, SNCA, SOD1, SPAST, SPG11, SPG21, SPG7, SPR, SPTBN2, SQSTM1, STUB1, SURF1, SYNE1, SYT14, TARDBP, TBP, TCTN2, TGM6, TH, TMEM67, TP53, TPP1, TREM2, TTBK2, TTPA, TWNK, UCHL1, VCP, VLDLR, VPS13A, VPS13C, WDR45, WDR81, WWOX, XPR1, ZFYVE26, ZNF592.</i>

² Se utilizaron dos herramientas bioinformáticas (Phenomizer, VarElect) para encontrar genes directamente asociados a los siguientes términos: *Signos piramidales + Bradiquinesia + Ataxia cerebelosa + Disminución de sensibilidad en miembros inferiores + Cambios de personalidad + Discapacidad intelectual.*

Previamente a la extracción de la muestra de sangre del paciente, se determinó la cobertura del kit de captura exómico utilizado para los genes candidatos. Los resultados mostraron que el kit Agilent V5 presentó un buen porcentaje de cobertura teórica de los exones codificantes de los genes a estudiar, minimizando la posibilidad de perder información de relevancia que pudiera estar presente en regiones no capturadas.

Análisis bioinformático

Priorización de variantes

Una vez recibidos los datos de la secuenciación exómica del paciente II.5, los mismos fueron procesados de acuerdo con el *pipeline* bioinformático descrito en Materiales y Métodos y comenzamos con el análisis de las variantes detectadas utilizando la herramienta B_platform (<https://www.bitgenia.com/b-platform/>). Se obtuvieron inicialmente un total de 114.817 variantes. La búsqueda comenzó en el grupo de genes asociados a NBIA y no arrojó hallazgos significativos, ya que todas las variantes encontradas resultaron benignas/probablemente benignas. Este primer paso permitió descartar la presencia de variantes relevantes en las regiones codificantes del gen *PANK2*, directamente asociada con la sospecha diagnóstica del profesional responsable del caso. En segundo lugar, se procedió a utilizar el panel de paraparesia espástica, a partir de la sugerencia del médico postulante, teniendo en cuenta el fenotipo evidenciado por el paciente. Nuevamente, la búsqueda no arrojó resultados interesantes y la mayoría de las variantes detectadas en esos genes resultaron ser benignas/probablemente benignas, o de significado incierto con una cigosidad diferente a la esperada para la/s condición/es estudiadas. Finalmente, luego de debatir los resultados obtenidos, decidimos seguir adelante buscando en Phenomizer [185] genes que pudieran estar directamente asociados a los signos/síntomas distintivos del paciente, entre ellos: signos piramidales, bradiquinesia, ataxia cerebelosa, disminución de sensibilidad en miembros inferiores, cambios de personalidad y discapacidad intelectual. Se aplicó un segundo filtro sobre los genes obtenidos utilizando el programa VarElect [186] y seleccionamos únicamente aquellos con un alto score de asociación con la lista de síntomas mencionada. Una vez definido este tercer panel de genes a estudiar, aplicamos filtros adicionales de impacto alto/moderado y baja frecuencia poblacional (<0,001). De esta manera, detectamos dos variantes nóveles (ausentes de bases de datos poblacionales) en el gen *AFG3L2* (OMIM: 604581), que podrían explicar a *priori*, con un modelo de herencia heterocigota compuesto, el cuadro del paciente. Las variantes halladas fueron *AFG3L2(NM_006796.3):c.383G>A - p.(Trp128Ter)* y *AFG3L2(NM_006796.3):c.2083A>G - p.(Arg695Gly)* (Tabla 7/Figura 11).

Tabla 7. Variantes detectadas (Genoma de referencia: GRCh37) en el paciente II.5 y su clasificación de acuerdo con los criterios del ACMG.

Crom.	Pos.	Gen	Variante	Exón	Transcripto	Clasificación ACMG
18	12367291	AFG3L2	C>T c.383G>A p.(Trp128Ter)	4	ENST00000537174	Patogénica PVS1, PP3, PM2
18	12337432	AFG3L2	T>C c.2083A>G p.(Arg695Gly)	16	ENST00000537174	Probablemente patogénica PM1, PM2, PM3, PP2, PP3

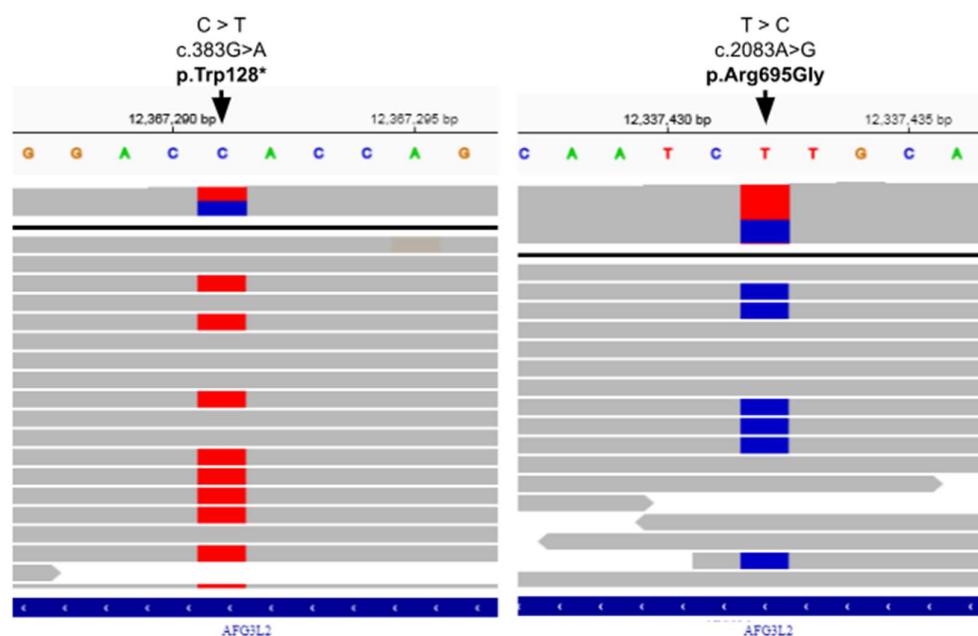


Figura 11: Visualización de las variantes detectadas en el archivo BAM. Ambas variantes se observaron en heterocigosis.

La primera variante se trató de un cambio de nucleótido único G>A en la región codificante (NM_006796.3) del exón 4, que produjo la ganancia de un codón *stop* prematuro en el residuo 128 de la proteína resultante, constituyendo una variante de pérdida de función, mecanismo que fue previamente asociado al desarrollo patológico (PVS1). La misma posee muy baja frecuencia poblacional (PM2, ausente de la base de datos de GnomAD), cuenta con predicción patogénica por numerosos algoritmos bioinformáticos BayesDel_addAF, DANN, EIGEN, FATHMM-MKL y MutationTaster (PP3) y no se encuentra reportada previamente en

bases de datos y/o publicaciones científicas. Por otro lado, la segunda variante hallada, correspondió a un cambio de nucleótido único A>G en de la región codificante (NM_006796.3) del exón 16, que produjo la sustitución del aminoácido Arginina (R) 695 por Glicina (G) en un *hot-spot* mutacional (PM1) de la proteína resultante. La misma posee muy baja frecuencia poblacional (PM2, ausente de la base de datos de GnomAD), cuenta con predicción de patogenicidad por los algoritmos bioinformáticos BayesDel_addAF, DANN, DEOGEN2, FATHMM-MKL, LIST-S2, M-CAP, MutationAssessor, MutationTaster y SIFT (PP3) y no se encuentra previamente reportada en bases de datos y/o publicaciones científicas. Adicionalmente, es importante destacar que el residuo R695 se encuentra al comienzo de una alfa hélice del dominio de peptidasa (Figura 12). Como se observa en la Figura 13, AFG3L2 es una proteína altamente conservada, la Arginina está presente en todos los vertebrados e incluso en bacterias. Como modelo, encontramos la estructura cristalina de la proteína FtsH homóloga de *T. thermophilus* (PDB id: 2DHR) - proteasa dependiente de ATP asociada a la membrana plasmática-, que muestra una Lisina en esa posición. La proteína AFG3L1 también es muy similar, está altamente conservada, y alberga Glutamina (Q) o Ácido glutámico (E) en su lugar. Curiosamente, otras variantes patogénicas en el mismo dominio y exón (exón 16), especialmente las más cercanas a la posición 695, dan como resultado cambios significativos en la carga y polaridad del residuo. Un ejemplo concreto lo constituyen las variantes E691K, A694E, E700K y R702Q (Figura 15). Nuestra hipótesis es que la variante R695G podría desestabilizar la hélice (un hecho que también está respaldado por la predicción del impacto del cambio en la estructura secundaria, ya que la Glicina es un residuo que desfavorece la formación de hélices) y/o que los residuos cargados en esta región resultan claves para el establecimiento de asociaciones o interacciones proteína/proteína adecuadas.

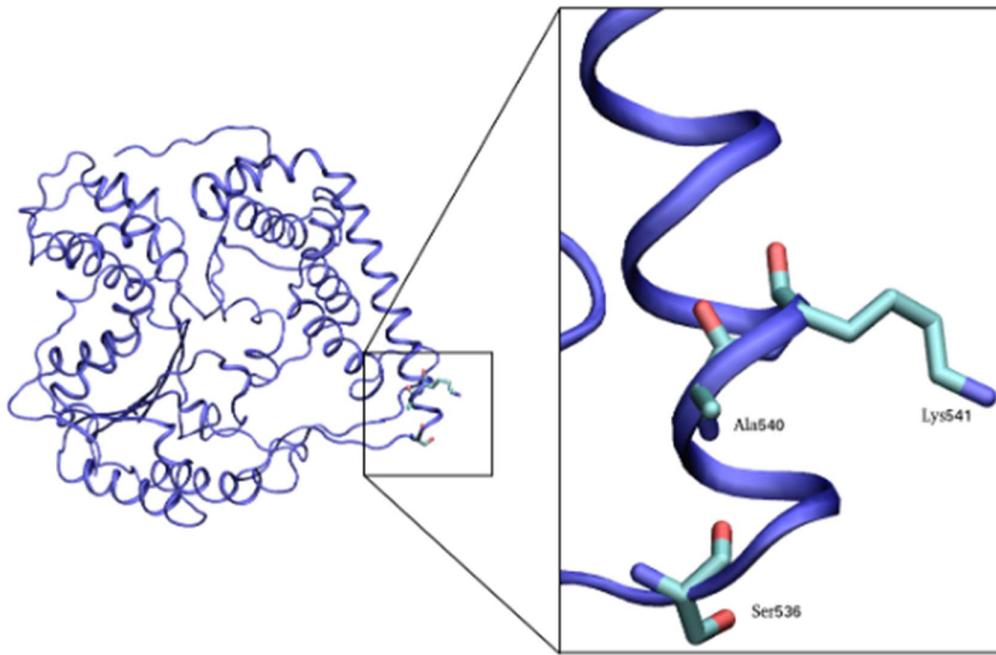


Figura 12: Estructura cristalina de la proteína *Thermus thermophilus* FTSH (PDB id: 2DHR). El residuo Lys541 es homólogo a Arg695 en la proteína AFG3L2 del *Homo sapiens*. El residuo Ala540 es homólogo del Ala694 de AFG3L2, cuya mutación se ha informado previamente como patogénica [187]. Ser536 es un residuo altamente conservado en toda la superfamilia peptidasa_M41.

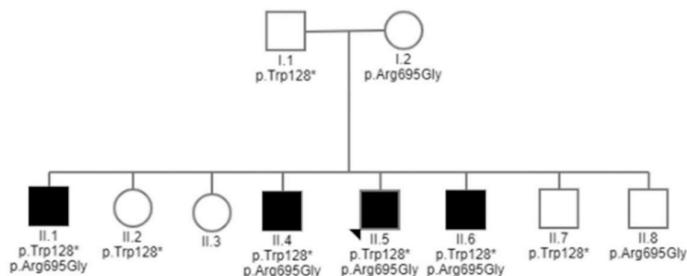
	695	
AFG3-l2(<i>Homo sapiens</i>)		
AFG3-l2(<i>Mus musculus</i>)		
AFG3-l2(<i>Fulmarus glacialis</i>)		
AFG3-l2(<i>Danio rerio</i>)		
Hfl(<i>Basidiobolus meristosporus</i>)		
FtsH(<i>Escherichia coli</i>)		
FtsH(<i>Thermus thermophilus</i>)		
AFG3-l1(<i>Mus musculus</i>)		
AFG3-l1(<i>Rattus norvegicus</i>)		
AFG3-like(<i>Caenorhabditis elegans</i>)		
AFG3l1(<i>Xenopus tropicalis</i>)		
SI:ch1073-174d20.2(<i>Danio rerio</i>)		
VTQAYAQIVQFGMNEKVGQISFDLPRQG-----DMVLEKPYSEATARLIDDEVRLIND	707	
VTQAYAQIVQFGMNEKVGQISFDLPRQG-----DMVLEKPYSEATARLIDDEVRLISD	706	
VTQAYAQIVQFGMNEKVGQISFDLPRQG-----DMVLEKPYSEATARLIDDEVRLINI	573	
VTQAYAQIVQFGMNEKVGQISFDLPRQG-----ELVLEKPYSEATARLIDTEVRNLIST	702	
VTQAYAQIVQFGMNEKVGQISFDLPRQG-----EQQFKPYSEETGRIIIDNEARKMIGN	678	
ATNLARNMVMTQWGFSEKLGPELLYAAEEEGEVFLGRSVAKAKHMSDETARIIDQEVKALIER	555	
ATELARRMITTEWGMHPEFGPVAYAVREDTYLGGYDV--ROYSEETAKRIDEAVRRLIEE	553	
VTQAYAQIVQFGMSEKLGVQSFDFPRQG-----ETMVEKPYSEATAQLIDEEVRCLVRS	699	
VTQAYAQIVQFGMSEKLGVQSFDFPRQG-----ETMVEKPYSEATAQLIDEEVRCLVRS	699	
VTQAYAQIVQFGMSEKLGVQSFDFPRQG-----EMAFDKPYSEATAQLIDQEVRDLVMN	692	
VTQAYAQIVQFGMSEKLGVQSFDFPRQG-----EMLAEKPYSEATAELIDQEARNLINS	695	
VTQAYAQIVQFGMSEKLGVQSFDFPRQG-----ETVLEKPYSEATAELIDEEVRDLVDR	659	

Figura 13: Alineamiento múltiple de proteínas homólogas a AFG3L2 en diferentes especies. El rectángulo rojo marca la posición del residuo 695 en la proteína humana.

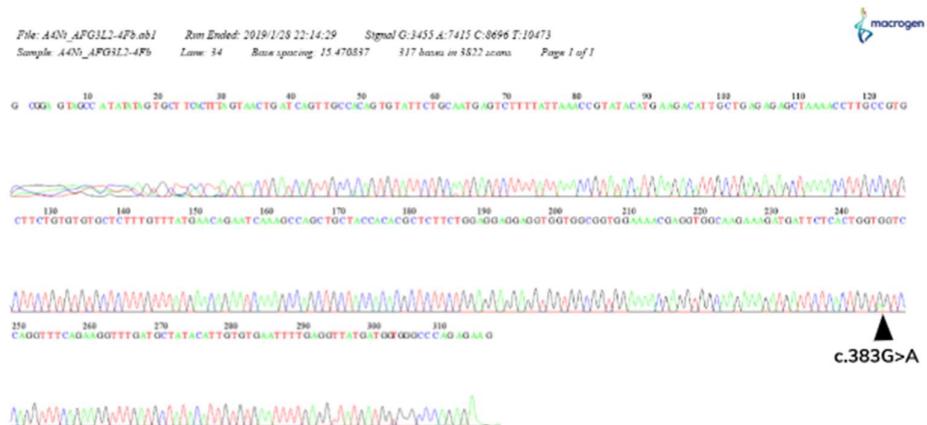
Validación de la segregación familiar

Se llevó a cabo la validación por técnica de Sanger de la presencia/ausencia de ambas variantes en el extenso grupo familiar, siguiendo los pasos mencionados en Materiales y Métodos, obteniéndose la segregación evidenciada en la Figura 14. De esta forma, se confirmó que todos los hermanos afectados de la familia resultaron portadores de ambas variantes (PP1, PM3), mientras que cada progenitor fue portador de una única variante en heterocigosis. Esto confirmó que las variantes se encontraban en *trans* en los individuos afectados, es decir, en alelos diferentes.

A)



B)



C)

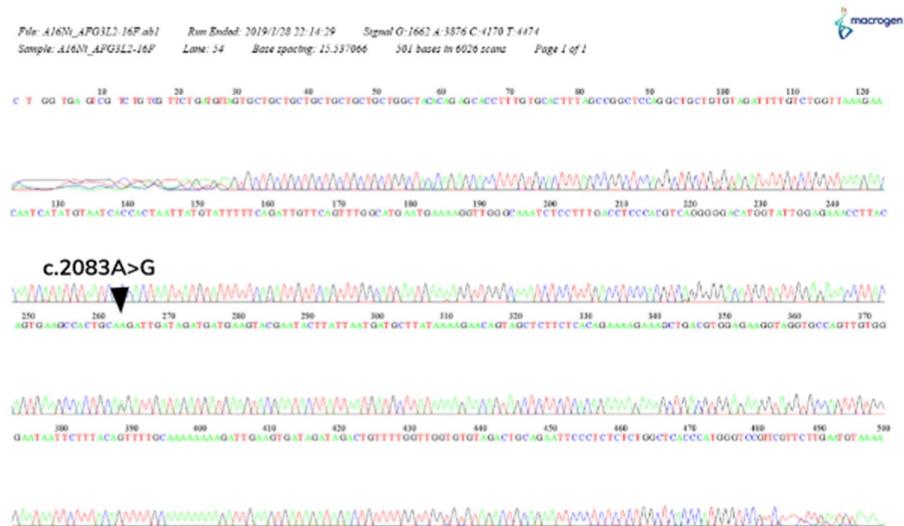


Figura 14: Segregación familiar de las variantes detectadas. A) Se evidencia la cosegregación de ambas variantes en los hermanos afectados (cuadrados rellenos). B) Cromatograma resultante del método de PCR / secuenciación en ABI310 (SANGER) que valida la presencia de la variante en el Exón 4 en el paciente II.5 C) Cromatograma resultante del método de PCR / secuenciación en ABI310 (SANGER) que valida la presencia de la variante en el Exón 16 en el paciente II.5.

Este hallazgo resultó crucial a la hora de reevaluar la relevancia diagnóstica de estas variantes, permitiéndole al médico postulante analizar en mayor profundidad los hallazgos informados y establecer una mejor correlación del genotipo hallado con el fenotipo observado. Para facilitar esta tarea, decidimos realizar una búsqueda bibliográfica exhaustiva de variantes previamente reportadas en el gen *AFG3L2*.

Correlación genotipo- fenotipo

El gen de tipo 3 de la familia ATPasa (*AFG3L2*) se encuentra localizado en la región cromosómica 18p11.21 y codifica para una subunidad de la proteasa m-AAA mitocondrial (ATPasa asociada a varias actividades celulares), localizada en la membrana interna mitocondrial. La metaloproteasa AAA participa en la calidad del control proteolítico y en actividades similares a las de las chaperonas, degradando proteínas mal plegadas y promoviendo el ensamblado de los complejos participantes en la cadena respiratoria [188]. *AFG3L2* forma complejos homo o heterooligoméricos con *SPG7*, otra proteasa m-AAA cuyas variantes conducen a una forma relativamente frecuente de paraplejia espástica familiar. El complejo *SPG7-AFG3L2* está involucrado en varias vías cruciales para la correcta función mitocondrial, incluyendo el control de calidad de las proteínas mitocondriales y la

homeostasis. La pérdida de función de AFG3L2 puede conducir a alteraciones en la síntesis de proteínas mitocondriales, respiración celular, mantenimiento de redes mitocondriales, transporte axonal, así como disfunciones en el flujo de Ca^{2+} y muerte celular inducida por Ca^{2+} [189].

Se ha reportado que la mayoría de las variantes de cambio de aminoácido (*missense*) detectadas en este gen fueron clasificadas como patogénicas [190–192], lo que sugiere que es un gen poco tolerante a variantes de este tipo (PP2).

Variantes en el gen *AFG3L2* son responsables de las formas SCA28 y SPAX5 de ataxia espinocerebelosa [187, 193]. Las ataxias espinocerebelosas (SCA, del inglés, *SpinoCerebellar Ataxia*) representan un grupo clínica y genéticamente heterogéneo de trastornos neurológicos hereditarios con fenotipos superpuestos y muy variables, caracterizados por incoordinación progresiva, disartria y alteración de movimientos oculares.

La Ataxia Espinocerebelosa de tipo 28 (SCA28, OMIM: 610246) es un subtipo muy poco común de Ataxia Cerebelosa autosómica dominante de tipo 1 (ADCA tipo I). Se caracteriza por inicio juvenil y ataxia cerebelosa de progresión lenta debida a la degeneración de las células de Purkinje. Su prevalencia es desconocida, pero la SCA28 representa aproximadamente el 1,5% de todos los casos europeos de ataxia cerebelar de herencia autosómica dominante [194]. Algunos pacientes muestran deterioro cognitivo y, en etapas más avanzadas de la enfermedad, se han descrito también oftalmoparesia, ptosis y signos piramidales como síntomas adicionales. La SCA28 ha sido localizada en el locus 18p11.22-q11.2 y su pronóstico es relativamente bueno, los pacientes pueden caminar sin asistencia hasta la séptima década de vida. Adicionalmente, los estados epilépticos y las convulsiones intratables que pudieran poner en riesgo sus vidas o la probabilidad de una disfagia grave son poco frecuentes.

Hasta la fecha, se han identificado 13 variantes de cambio de aminoácido (*missense*) heterocigotas en familias de ascendencia francesa, italiana y alemana (Figura 15). En la mayoría de los casos, las variantes se detectaron en los exones 15 y 16, con una sola excepción (c.1295A>C, p.N432T) detectada en el exón 10 [187]. Los estudios funcionales de la proteasa m-AAA codificada con alguna variante en los exones 15 o 16 revelaron un dominio peptidasa M41 en el extremo C-terminal anormal que comprometía la estabilidad y la actividad proteolítica del complejo m-AAA, posiblemente a través de un proceso dominante negativo. Por el contrario, la mutación localizada en el exón 10, modificó la estructura de entrada del canal, en la que los sustratos se extraen para degradarlos [187]. Además, se han informado dos delecciones en el gen *AFG3L2* en pacientes con SCA28, una del gen completo [195] y otra que incluyó desde el exón 14 hasta el 16 [196], así como una variante de cambio en el marco de lectura en el exón 15, que eliminó los últimos 144 aminoácidos [197]. **Por lo tanto, la haploinsuficiencia de *AFG3L2* constituye una de las posibles causas de SCA28.**

Adicionalmente, se informó una nueva mutación c.346A>G, p.G116R en el dominio N-terminal de la proteína AFG3L2 en un paciente que presentó mioclonías espontáneas asociadas a signos piramidales. Hasta el momento de la publicación, el paciente no presentó ataxia cerebelosa, anomalías oculomotoras ni déficits sensoriales [198].

Por otro lado, la Ataxia Espástica-5 (SPAX5, OMIM: 614487) constituye un trastorno neurodegenerativo autosómico recesivo que se caracteriza por el inicio en la infancia de paraparesia espástica lentamente progresiva de las extremidades inferiores y ataxia cerebelosa (con disartria, dificultad para tragar, degeneración motora), junto con neuropatía sensitivo-motora (que incluye debilidad muscular y amiotrofia distal en las extremidades inferiores) y epilepsia mioclónica progresiva. También se pueden asociar signos oculares (ptosis, apraxia oculomotora), dismetría, disdiadococinesia, movimientos distónicos y mioclonías [199]. La primera variante asociada a SPAX5 (c.1847G>A, Y616C) se detectó en homocigosis en 2 hermanos con ataxia espástica de inicio temprano. La variante Y616C interfiere, pero no inhibe completamente, la actividad de la proteasa m-AAA homo-oligomérica [193]. Posteriormente, se describió una mutación homocigota (c.1875G>A, M625I) en 2 pacientes italianos no emparentados con un fenotipo leve y un historial clínico de ataxia diferente, ninguno de los cuales presentó espasticidad o neuropatía temprana [200]. Ambas variantes mencionadas se localizaron en el exón 15. Adicionalmente, se reportaron hermanos sauditas que evidenciaron apraxia oculomotora, características extrapiramidales y epilepsia mioclónica [201].

Finalmente, Caporali y col. (2020) publicaron un *case report* donde 2 pacientes con atrofia óptica (OPA12, OMIM: 618977) resultaron portadores de variantes bialélicas en el dominio ATPasa (dominio AAA). Ambos presentaron pérdida de visión, déficit en la visión de los colores, fotofobia, discos ópticos pálidos, disminución del grosor de la capa de fibras nerviosas de la retina, ataxia, espasticidad, mioclonías, corea, distonía y signos cerebelosos [202]. La distribución de variantes patogénicas asociadas con atrofia óptica aislada, que se localizan predominantemente en el dominio ATPasa, es distinta de las asociadas con SCA28 y SPAX5, que se localizan principalmente en el dominio proteolítico [203].

Hasta la fecha de nuestra publicación, sólo se habían informado 9 pacientes con variantes autosómicas recesivas en el gen *AFG3L2*, sumándose luego los dos pacientes informados por Caporali et al. 2020 [202]. A diferencia de los pacientes estudiados en este trabajo de tesis, todos resultaron portadores de variantes homocigotas y presentaron una edad de inicio más temprana, una progresión clínica más rápida y epilepsia mioclónica progresiva, mientras que ninguno de ellos evidenció el signo del ojo de tigre en las imágenes de resonancia magnética.

En levaduras, variantes homocigotas en el gen *AFG3L2* alteraron la respiración mitocondrial, la actividad de la citocromo c oxidasa y la actividad proteolítica del complejo de

metaloproteasa m-AAA. Es probable que estas deficiencias condujeran al compromiso de la cadena respiratoria y a una mayor generación de especies reactivas de oxígeno, lo que podría conducir a la degeneración de las células de Purkinje y a la disfunción cerebelosa [204].

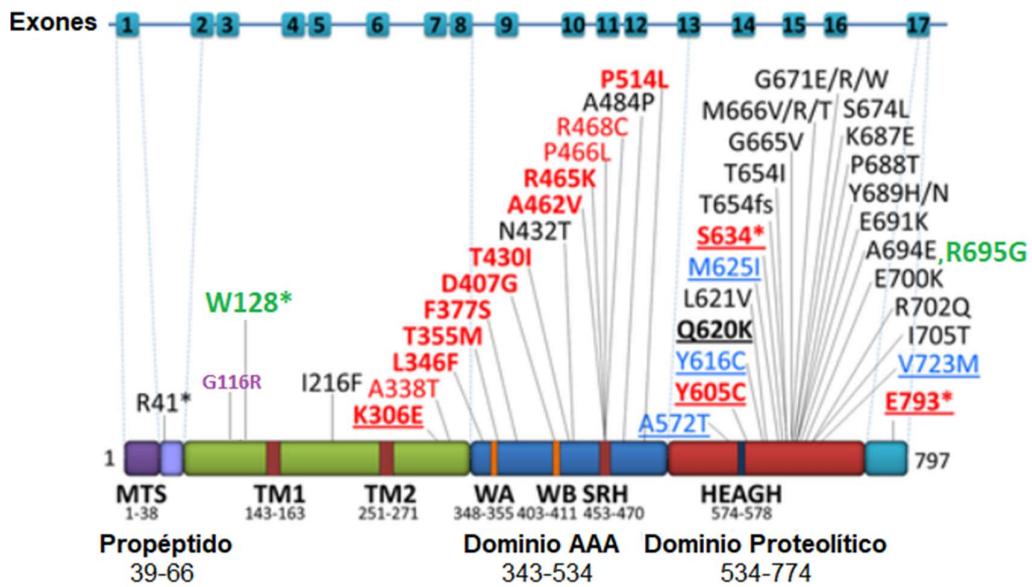


Figura 15. Distribución de variantes conocidas en la proteína AFG3L2. La estructura del gen AFG3L2 (17 exones) y la proteína (dominios y posiciones de aminoácidos) se describen en la parte superior e inferior de la figura, respectivamente, con todas las variantes patogénicas informadas hasta la fecha en AFG3L2 (NM_006796.3). Las variantes patogénicas asociadas con fenotipos de Atrofia Óptica (rojo), Ataxia Espinocerebelosa 28 (SCA28; negro) y Síndrome de Ataxia Espástica 5 (SPAX5; azul) [203] se indican con diferentes colores. La variante G116R (violeta) se encontró en un paciente con mioclonía y signos piramidales. Las variantes informadas en este estudio se indicaron en verde y negrita. Las variantes bialélicas están subrayadas.

En la Tabla 8 se resumen las características y hallazgos clínicos de los pacientes reportados hasta la fecha, portadores de variantes patogénicas en homocigosis o heterocigosis compuesta en el gen AFG3L2.

Tabla 8. Casos de pacientes reportados en la literatura científica. Se contemplan aquellos pacientes con variantes bialélicas en el gen *AFG3L2* previamente reportados en la literatura (en orden cronológico), incluyendo un resumen de los pacientes presentados en esta tesis. P=paciente, G=género, EI=edad de inicio; C=consanguinidad; A=ancestría; EMP=epilepsia mioclónica progresiva; GTC=generalizadas tonicoclónicas; MI=miembros inferiores; ↓=disminuido; NAA=N-acetil aspartato; mtDNA= DNA mitocondrial; EEG=electroencefalografía; ↑=elevado; MRS=espectroscopia de resonancia magnética; Hz=hertz; EOT=signo “ojo de tigre”; NCS / EMG= estudio de conducción nerviosa / electromiografía; MS=miembros superiores.

Autores	P/G/EI	C/A	Hallazgos clínicos	Hallazgos RMN	Mutación en <i>AFG3L2</i>	Otros hallazgos
<i>Pierson TM et al., 2011 [193]</i>	IV-1/masculino/2 años	Sí/Colombianos	Marcha espástica. EMP, GTC y convulsiones mioclónicas a los 8 años. Disartria, disfagia y marcha con asistencia. Paraparesia con atrofia distal LL. Último examen: ataxia cerebelosa, ptosis, apraxia oculomotora, distonía, mioclonías inducidas por estímulos, con cognición normal.	Atrofia moderada del cerebelo / pontino, leve adelgazamiento del cuerpo caloso, ↓ valores de NAA en pontino y cerebelo. Madre con leve atrofia de cerebelo.	c.1847G>A c.1847G>A	Paciente IV-1: neuropatía sensitivo-motora axonal de MI. Biopsia del nervio sural: “bulbo de cebolla”. Biopsia muscular (microscopía electrónica): mitocondrias fuera de lugar con gotitas de lípidos, ↓ número de copias de ADNmt. EEG: EMP. Estudios funcionales en levaduras y fibroblastos de pacientes.
	IV-2/masculino/2 años		Marcha espástica, nunca caminó independientemente. EMP, GTC y convulsiones mioclónicas a los 8 años. Disartria, disfagia. Paraparesia con atrofia distal MI. Falleció de neumonía a los 13 años.	No disponible		

Muona M et al., 2015 [205]	PME62-1/ masculino/ 10 años	No/ Italianos	Mioclonías severas y convulsiones tonicoclónicas, ataxia, ligero deterioro cognitivo. En silla de ruedas por mioclonías y ataxia a los 22 años.	No disponible	c.1875G>A c.1875G>A	-1.75-Mb mutación homocigota con polimorfismos idénticos flanqueando la región.
	PME32-1/ femenino/ 10 años	No/ Italianos	Mioclonías y ataxia progresiva. Convulsiones tonicoclónicas a los 22 años. Defecto cognitivo leve. Silla de ruedas a partir de los 48 años.	No disponible	c.1875G>A c.1875G>A	
Eskandrania et al., 2017 [201]	IV.2/femenino/ 8 meses	Si/ Sauditas	Hitos de regresión, espasmos infantiles, epilepsia refractaria, microcefalia, espasticidad. Vivo, 5 años.	Señal hiperintensa en putamen bilateral y núcleos caudados con ↑lactato en MRS.	c.1714G>A c.1714G>A	Afectación retiniana bilateral por electrorretinograma. ↑ lactato sanguíneo. El EEG evolucionó de un patrón hipsarrítmico a una actividad moderada bien organizada y moderadamente bien desarrollada de 4 a 5 Hz, bilateralmente sincrónico y simétrico, con delta rítmico de alta amplitud con frecuentes picos superpuestos máximos en el área occipital. Paciente IV.7: fibroblastos cutáneos con ↓ actividad del complejo I / III y IV de la cadena respiratoria mitocondrial.
	IV.3/masculino/ 6 meses		Hitos de regresión, espasmos infantiles, epilepsia refractaria, microcefalia, espasticidad. Murió a los 2 años y medio.	Señal hiperintensa en putamen.		
	IV.5/femenino/ 8 meses		Hitos de regresión, convulsiones mioclónicas, epilepsia refractaria, microcefalia, espasticidad. Vivo, 5 años	Señal hiperintensa en putamen.		
	IV.7/masculino/ 7 meses		Hitos de regresión, espasmos infantiles, epilepsia refractaria, microcefalia, espasticidad. Vivo, 4 años y medio	Señal hiperintensa en putamen.		
	IV.8/femenino/ 8 meses		Hitos de regresión, espasmos infantiles, epilepsia refractaria, microcefalia, espasticidad. Murió a los 8 años.	Señal hiperintensa en putamen bilateral y núcleos caudados.		

Calandra CR et al., 2020 [206]	II.1/masculino/ 8 años	No/Europeos- Americanos	Dificultad de aprendizaje; cambios de comportamiento lentamente progresivos, marcha espástica y atáxica; signos piramidales; ataxia cerebelosa apendicular; disartria; nistagmo evocado por la mirada; disminución de la sensación vibratoria en MI; Signo de Romberg (+). Vivo, 28 años.	Signo similar a EOT; Atrofia leve del vermis cerebeloso; señal hipointensa en imagen ponderada en T1 de la sustancia negra.	c.383G>A c.2083A>G	Paciente II.4 NCS / EMG: polineuropatía axonal sensitivo-motora de MI y MS.
	II.4/masculino/ desconocido		Ataxia cerebelosa troncal y apendicular leve, signos piramidales de MI, disminución de la sensación vibratoria de MI; nistagmo evocado por la mirada; Signo de Romberg (+). Vivo, 23 años.			
	II.5/masculino/ 8 años		Dificultad de aprendizaje, ataxia cerebelosa de MI, marcha levemente atáxica, signos piramidales en MI, disminución de la sensación vibratoria en MI, nistagmo evocado por la mirada. Vivo, 19 años.	Ídem II.1 y II.4 incluyendo señal hiperintensa del borde posterior del núcleo lenticular y del núcleo dentado en imagen ponderada en T2.		
	II.6/masculino/ desconocido		Ataxia cerebelosa apendicular y troncal leve, signos piramidales de MI, nistagmo evocado por la mirada, pie cavo bilateral, signo de Romberg (+). Vivo, 17 años.	Signo similar a EOT; Atrofia leve del vermis cerebeloso, señal hipointensa en imagen ponderada en T1 de la sustancia negra.		

<i>Caporali L et al., 2020</i> [202]	Familia 6 II- I/femenino/ 36 años	<p>Presentó atrofia óptica desde los 2 años de edad, que dio lugar a ceguera bilateral a los 15 años.</p> <p>Padeció mioclonías infantiles de las extremidades desde los 8 años y un trastorno progresivo de la marcha espástica-atáxica desde los 15 años.</p> <p>A los 35 años presentó ceguera, ataxia cerebelosa de leve a moderada, paraespasticidad, mioclonías de las extremidades y corea de las manos.</p>	<p>No se observaron otras anomalías neurológicas y el examen oftalmológico y la resonancia magnética cerebral fueron negativos.</p>	c.1385C>T c.1858C>A	----
	Familia 11, II- I/masculino/ 18 años	<p>Desde los 4 años con dificultad para caminar y ataxia de la marcha.</p> <p>Luego desarrolló movimientos hipercinéticos involuntarios de las extremidades superiores y una postura distónica. El déficit visual se hizo evidente a los 9 años y se detectó atrofia óptica. Ahora presenta ataxia, dismetría, disgracia y distonía.</p>	<p>A los 6 años, la resonancia magnética reveló hiperintensidad en T2 en putamen, globos pálidos y sustancia negra. El seguimiento por resonancia magnética mostró afectación del núcleo caudado (11 años) y del subtálamo (14 años).</p>	c.1901_1902delCT c.916A>G	----

Discusión

En el presente caso, se encontraron dos variantes (p.Trp128Ter y p.Arg695Gly) en el gen *AFG3L2* a partir de la secuenciación exómica del paciente estudiado, que contaron con el potencial de explicar el fenotipo observado y aportar información de relevancia a nivel molecular. Resulta importante mencionar que el análisis de dicho nos resultó sumamente complejo, por lo que creemos importante mencionar los desafíos con los que nos encontramos durante el estudio de esta familia, teniendo en cuenta los siguientes aspectos:

- La sospecha diagnóstica inicial del médico, que era orientativa de un grupo de desórdenes asociados a NBIA. Al no encontrar variantes de relevancia en este grupo de 10 genes, particularmente en *PANK2*, se tuvieron múltiples reuniones con el médico, que propiciaron un mejor entendimiento de la clínica y facilitaron las búsquedas sucesivas de genes relacionados con paraparesia espástica (75 genes) o potencialmente relacionados con la clínica del paciente usando herramientas disponibles en línea para su incorporación al estudio, adicionando 178 genes al análisis, entre los cuales se encontraba *AFG3L2*.
- La disponibilidad reducida de información de relevancia asociada a las principales variantes candidatas encontradas como resultado de la aplicación de los filtros en la plataforma usada para el análisis del caso en el año 2018. Esto resultó inevitablemente en la necesidad de buscar información adicional en diversas bases de datos de forma manual, lo que conllevo un mayor tiempo de análisis y le asoció un mayor nivel de dificultad a este caso. Adicionalmente, el difícil acceso a herramientas que facilitaran la clasificación de variantes al momento del análisis del caso resultó ser otro obstáculo (desconocimiento o inaccesibilidad a clasificadores semi-automáticos como Intervar o Varsome). Actualmente, la plataforma cuenta con mayor información que, a simple vista, permite descartar o priorizar variantes sin necesidad de recurrir a un análisis exhaustivo adicional (por ejemplo, clasificación semiautomática según criterios de la ACMG, asignación de frecuencia poblacional según GenomAD, 1000 genomas y de frecuencia poblacional local, más de 10 predictores bioinformáticos de patogenicidad, links a bases de datos de interés como ClinVar, PMED, UniProt y estadísticas comparativas con casos previos).
- La compleja correlación genotipo-fenotipo, teniendo en cuenta lo previamente reportado para variantes bialélicas en el gen *AFG3L2* hasta principios del año 2018, donde solo se habían descrito 9 pacientes con ataxia espástica 5 de herencia recesiva. A diferencia de lo observado para el caso analizado, todos

ellos tenían una edad de inicio más temprana, una progresión clínica más rápida y/o epilepsia mioclónica progresiva. Más recientemente, cerca de la fecha de la publicación del *Case Report*, se reportó un paciente con variantes en heterocigosis compuesta en *AFG3L2*, que presentaba varias características comunes con el fenotipo de ataxia espinocerebelosa tipo 28, usualmente de herencia autosómica dominante. La misma asocia un fenotipo más leve e incluye ataxia de la marcha y de las extremidades de aparición tardía y progresión muy lenta que provoca problemas de coordinación y equilibrio, disartria, ptosis, nistagmo y oftalmoparesia. Sin embargo, las manifestaciones clínicas de la madre heterocigota no fueron concluyentes y el padre del paciente no estuvo disponible para la evaluación clínica.

Inicialmente, el profesional a cargo consideró una forma atípica de PKAN como el diagnóstico más probable debido a la presencia del signo ojo de tigre (EOT-like) -lesión temprana en el globo pálido que se torna en un área de gliosis-. También existía la posibilidad de delecciones más grandes en *PANK2* no detectadas por la secuenciación del exoma. Sin embargo, la ausencia de acumulación excesiva de hierro en el globo pálido tras al menos 15 años de enfermedad argumenta fuertemente en contra de la posibilidad de NBIA. Se sospechó que la ataxia espinocerebelar -trastornos progresivos de herencia autosómica dominante en los que el cerebelo degenera lentamente, a menudo acompañados de cambios degenerativos en el tronco encefálico y otras partes del sistema nervioso central- y la paraplejia espástica hereditaria como posibles diagnósticos, pero en ninguno de los dos se consideró una fenocopia de imagen del signo EOT. Esto nos llevó a considerar el mejor enfoque diagnóstico y a ampliar el espectro fenotípico asociado con variantes bialélicas en el gen *AFG3L2* con la descripción de los cuatro hermanos reportados en este caso, que presentaron ataxia espástica con signo del ojo del tigre, aportando información novedosa que permitió la publicación de un *Case Report* en la revista *Parkinsonism & Related Disorders* en el año 2020 [206].

El trabajo y la dedicación puestas en este caso no solo le brindaron a una gran familia la posibilidad de recibir un potencial diagnóstico molecular y la posible causa subyacente al fenotipo de los integrantes afectados, sino también un nuevo punto de partida que incluyó asesoramiento genético y la evaluación de posibles estrategias terapéuticas.

Caso II:

Paciente con sospecha de Enfermedad Autoinflamatoria

Introducción

Este caso incluye a una paciente con lesiones cutáneas ulcerativas y enfermedad pulmonar. La paciente nació en 2013 de una familia argentina no consanguínea, que se completa con un hermano menor sano y un hermanastro paterno sano (Figura 16). Recibió la vacuna BCG (Bacillus Calmette–Guérin) al nacer. A los 6 meses de edad presentó lesiones ulceradas pápulo-ulcerativas en cuello y axilas. En ese momento ingresó en el hospital y recibió tratamiento antibiótico sin respuesta clínica, seguido de resolución espontánea 2 meses después. Entre los 6 y los 15 meses de edad, y debido a las úlceras cutáneas recurrentes, inicialmente se sospechó que presentaba BCG-itis - una reacción poco frecuente, secundaria a la vacunación, que se manifiesta localmente como lesiones polimorfas granulomatosas-. Al año de edad presentó novedosas lesiones similares en región inguinal y genitales externos, de localización bilateral y asimétrica y sin respuesta a tratamientos locales ni sistémicos. Las lesiones mostraron cicatrización lenta (3 meses) y la biopsia informó ulceración de dermis y epidermis, con necrosis de hipodermis subyacente y sin reacción inflamatoria. Las reacciones ácido-Schiff periódicas, bacilos ácido-alcohol resistentes y Ziehl-Neelsen fueron negativas. Los cultivos para bacterias piógenas, micobacterias y hongos también fueron negativos. La inmunofluorescencia de muestras de piel fue negativa para inmunoglobulinas A, G, M y C3.

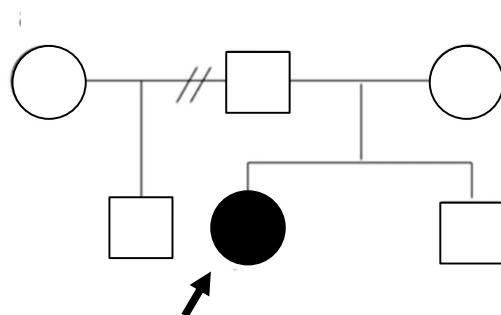


Figura 16: Familiograma. Se grafica la historia familiar, que permite sospechar de un modelo de herencia recesivo.

A partir de los 2 años y 8 meses de edad, la paciente comenzó a tener manifestaciones respiratorias que incluían sibilancias recurrentes y neumonías (siete episodios aislados y no relacionados en el lapso de 4 años). Tres de ellos requirieron hospitalización, antibióticos y oxigenoterapia (por un período de 5 días en promedio); mientras que uno presentó imágenes

indicativas de una neumonía necrosante con pruebas microbiológicas negativas. Se mantuvo asintomática entre episodios. La tomografía computarizada (TC) de pulmón y tórax (Figura 17a) mostró algunas lesiones inespecíficas y residuales, mientras que la TC cerebral evidenció calcificaciones de los ganglios basales (Figura 17b). Por las manifestaciones pulmonares se sospechó nuevamente infección por micobacterias. Sin embargo, los médicos descartaron BCG-itis, ya que: (i) no se aislaron especies de micobacterias (u otros microorganismos) de biopsias de piel o de múltiples lavados broncoalveolares (BAL); (ii) las lesiones cutáneas cicatrizaron sin el uso de ningún tratamiento antimicobacteriano, y (iii) las manifestaciones pulmonares, aunque parecían progresivas, se presentaron en múltiples episodios y afectaron diferentes secciones del parénquima pulmonar.

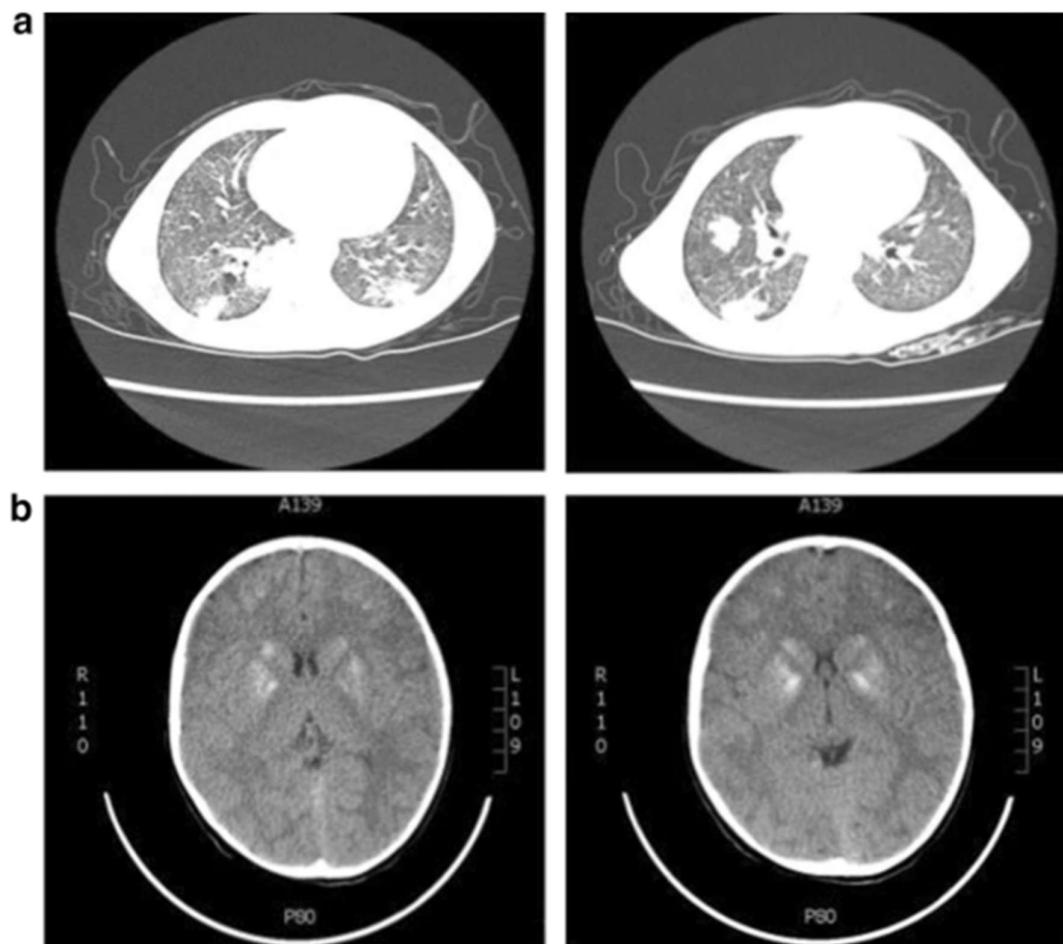


Figura 17. a) **Tomografía computarizada de tórax.** Áreas focales bilaterales de consolidaciones parenquimatosas, algunas de ellas con broncograma aéreo y de aspecto nodular, distribuidas en las porciones periféricas de las zonas inferiores. También se observa engrosamiento septal generalizado bilateral y cambios en el árbol periférico en ambos pulmones. Se encontró atrapamiento de aire lobular al espirar. b) **Tomografía computarizada del cerebro.** En las imágenes se muestran calcificaciones a nivel de los núcleos caudados, núcleos lenticulares y región subcortical de ambos lóbulos frontales.

Las pruebas de laboratorio mostraron reactantes de fase aguda elevados, principalmente la velocidad de sedimentación globular (VSG). El número y porcentaje de células T, B y NK mostró estar en un rango adecuado para la edad, así como IgG, C3 y C4. Además, se encontró una respuesta inadecuada de anticuerpos a algunos antígenos proteicos y polisacáridos. El ensayo de dihidrorodamina en neutrófilos fue normal. El estudio autoinmune reveló anticoagulante lúpico positivo transitorio y auto-anticuerpo inespecífico de ASMA. Debido al cuadro multisistémico de la paciente, que afectó predominantemente a la piel (úlceras crónicas y recurrentes) y al tracto respiratorio inferior (enfermedad pulmonar asociada a hipoxia y lesiones persistentes a pesar del tratamiento adecuado), se sospechó inicialmente que la paciente tenía una enfermedad autoinflamatoria. Durante el año 2020, la paciente presentó talla y peso adecuados para su edad. Las lesiones en cicatrización todavía estuvieron presentes en regiones retroauriculares, cervicales, axilares e inguinales bilaterales y en labios mayores, con lipoatrofia asociada. Presentó una dentición retrasada y peculiar con ausencia de caninos y sin otras anomalías. Además, evidenció enfermedad pulmonar crónica con insuficiencia respiratoria hipoxémica y fue tratada con oxigenoterapia en su domicilio (0,5 L/min en sueño y 1 L/min durante el día) [207].

De acuerdo con la fuerte sospecha diagnóstica de enfermedad autoinflamatoria indefinida de origen genético y en base a la historia clínica proporcionada, se contempló la utilización de paneles asociados a Enfermedades Autoinflamatorias e Inmunodeficiencias Primarias (Tabla 9). Como en todos los casos y previamente a la extracción de muestra de la paciente, se determinó la cobertura del kit de captura exómico para los genes candidatos. Los resultados mostraron que el kit Agilent V5 presentaba un buen porcentaje de cobertura teórica de los genes a estudiar, minimizando la posible pérdida de información relevante que pudiera estar presente en regiones no capturadas (por ejemplo: intrones, regiones promotoras, 5'UTR, 3'UTR, zonas repetitivas, etc.).

Tabla 9. Paneles de genes seleccionados. Se muestran los paneles priorizados en función de la sospecha clínica y el fenotipo evidenciado.

Panel	Genes contemplados
Panel para Enfermedades Autoinflamatorias	<i>ACP5, ADAM17, ADAR, ADGRE2, AGBL3, AP1S3, CARD14, CARD15, CECR1, CIAS1, COPA, DDX58, DNASE2, IFIH1, IL10, IL10RA, IL10RB, IL1RN, IL36RN, LPIN2, MEFV, MVK, NALP12, NALP3, NCSTN, NLRC4, NLRP1, NLRP12, NLRP3, NLRP7, NOD2, OSMR, OTULIN, PLCG2, POLA1, POMP, PSMA3, PSMB4, PSMB8, PSMB9, PSMG2, PSTPIP1, RBC1, RNASEH2A, RNASEH2B, RNASEH2C, RNF31, SAMHD1, SH3BP2, SHARPIN, SLC29A3, TMEM173, TNFAIP3, TNFRSF11A, TNFRSF1A, TREX1, TRNT1, USP18, WDR1.</i>
Panel de Inmunodeficiencias primarias	<i>ACD, ACP5, ADA, ADA2, ADAM17, ADAR, AICDA, AIRE, AK2, AP3B1, ATM, B2M, BCL10, BLNK, BLOC1S6, BTK, CARD11, CARD14, CARD9, CASP10, CASP8, CD247, CD27, CD3E, CD3G, CD40LG, CD79A, CD79B, CD8A, CDK1, CEBPE, CHD7, CIITA, CLPB, COPA, CR2, CSF3R, CTC1, CTLA4, CTPS1, CTSC, CXCR4, CYBA, CYBB, DCLRE1A, DCLRE1B, DNMT3B, DOCK2, DOCK8, ELANE, EPG5, FADD, FAS, FASLG, FERMT3, FOXN1, FOXP3, FPR1, G6PC3, GATA2, GFI1, HAX1, ICOS, IFIH1, IFNGR1, IFNGR2, IGLL1, IKBKB, IL10RA, IL10RB, IL12B, IL12RB1, IL17F, IL17RA, IL17RC, IL1RN, IL21, IL21R, IL2RA, IL2RG, IL36RN, IL7R, IRAK4, IRF7, IRF8, ISG15, ITCH, ITGB2, ITK, JAGN1, JAK3, LAMTOR2, LCK, LIG4, LPIN2, LRBA, LYST, MAGT1, MALT1, MAP3K14, MEFV, MOGS, MVK, MYD88, NBN, NCF2, NCF4, NFAT5, NFKB2, NFKBIA, NHEJ1, NHP2, NLRC4, NLRP12, NLRP3, NOD2, NOP10, ORAI1, PARN, PGM3, PIK3CD, PIK3R1, PLCG2, PMS2, PNP, POLE, PRF1, PRKCD, PSMB8, PSTPIP1, PTPRC, RAB27A, RAC2, RAG1, RAG2, RBC1, RFX5, RFXANK, RFXAP, RHOH, RNASEH2A, RNASEH2B, RNASEH2C, RORC, RTEL1, SAMHD1, SEMA3E, SH2D1A, SH3BP2, SLC29A3, SLC35C1, SLC37A4, SLC7A7, SMARCAL1, SP110, SPINK5, STAT1, STAT2, STAT3, STAT5B, STIM1, STK4, STX11, STXBP2, TAP1, TAP2, TAPBP, TAZ, TBK1, TCN2, TERC, TERT, TICAM1, TINF2, TLR3, TMC6, TMC8, TMEM173, TNFRSF13B, TNFRSF13C, TNFRSF1A, TNFRSF4, TNFSF12, TPP2, TRAF3, TRAF3IP2, TREX1, TRNT1, TTC7A, TYK2, UNC13D, UNC93B1, UNG, VPS13B, VPS45, WAS, WIPF1, XIAP, ZAP70, ZBTB24.</i>

Análisis bioinformático

Priorización de variantes

Realizamos la secuenciación del exoma de la paciente, procesamos los datos y comenzamos el análisis con la herramienta B_platform. Se obtuvieron un total de 100.341 variantes. El primer análisis se centró en los genes del panel de EAI, buscando variantes con potencial impacto molecular, de acuerdo con la sospecha diagnóstica. La búsqueda no arrojó hallazgos significativos, y no se encontraron variantes previamente reportadas como patogénicas, probablemente patogénicas o VUS (de significado incierto), dentro de los genes directamente relacionados con patologías autoinflamatorias. Decidimos entonces, con el aval del equipo médico, buscar variantes de impacto alto/moderado y baja frecuencia (<0,001) en genes asociados con Inmunodeficiencias Primarias, de acuerdo con los signos y síntomas descritos, de manera tal de ampliar la sospecha diagnóstica. De esta forma, se logró la detección de dos variantes de alto impacto en heterocigosis en el gen *ISG15* (OMIM: 147571) (Tabla 10, Figura 18), del inglés “*interferon-stimulated gene 15*”.

Tabla 10. Variantes detectadas (Genoma de referencia: GRCh37 (hg19)) en la paciente y su clasificación de acuerdo con los criterios ACMG.

Crom.	Pos.	Gen	Variante	Exón	Transcripto	Clasificación ACMG
1	949644	<i>ISG15</i>	AC>A c.285delC p.(Tyr96ThrfsTer5)	2	ENST00000379389	Patogénica PVS1, PS3, PM2, PM3
1	949659	<i>ISG15</i>	c.299_312del p.(Leu100Argfs)	2	ENST00000379389	Patogénica PVS1, PS3, PM2, PM3

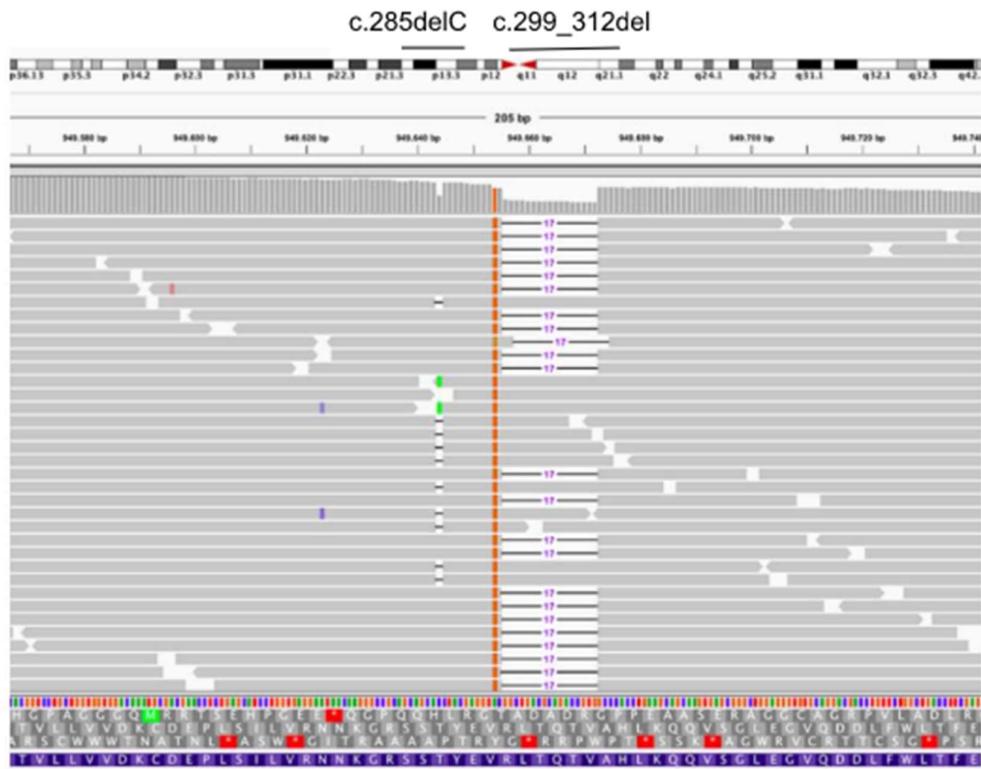


Figura 18: Visualización de las variantes detectadas en el archivo BAM. Ambas se observaron en heterocigosis.

La primera variante (NM_005101.4):c.285del, también conocida como c.284del, corresponde a una delección de un solo nucleótido que produce un cambio en el marco de lectura en el exón 2 (p.Tyr96Thrfs*5) con la consecuente ganancia de un codón stop prematuro río abajo del residuo Treonina 95.

La segunda variante, (NM_005101.4):c.299_312del, también conocida como c.297_313del, corresponde a una delección de 14 nucleótidos, que modifica el tamaño de la proteína resultante y también produce el desplazamiento del marco de lectura río abajo del residuo Leucina 100.

Ambas variantes se localizaron dentro del dominio proteico *Ubiquitina-like 2* (Figura 19), presentaron muy baja frecuencia poblacional (ausentes en la base de datos poblacionales como GnomAD) y no se reportaron previamente en bases de datos clínicas ni publicaciones científicas.

Según los criterios del ACMG, ambas variantes se clasificaron como patogénicas, de acuerdo a la combinación de los siguientes criterios:

- PVS1 (patogénico muy fuerte), ya que ambas son variantes nulas.
- PS3 (patogénico fuerte 3), cuando células HEK293T fueron transfectadas con cada una de las variantes identificadas en este caso, se encontró que ninguna condujo a la producción de proteína ISG15 detectable, dando como resultado la pérdida de expresión de acuerdo con el sistema de sobreexpresión *in vitro* diseñado por Dusan y col., en el que se compararon ambas variantes con la forma *wild type* de ISG15. Por su parte, la variante c.299_312del afectó los niveles de mRNA de ISG15, produciéndose su degradación por el mecanismo *non-stop decay*.
- PM2 (patogénico moderado 2), por sus bajas frecuencias poblacionales (sin reportes o muy baja frecuencia de alelos en bases de datos como ExAC o GnomAD);
- PM3 (patogénico moderado 3), ya que se confirmó por secuenciación de Sanger que la paciente heredó ambas variantes en *trans*, cada una a partir de un progenitor (Figura 22).

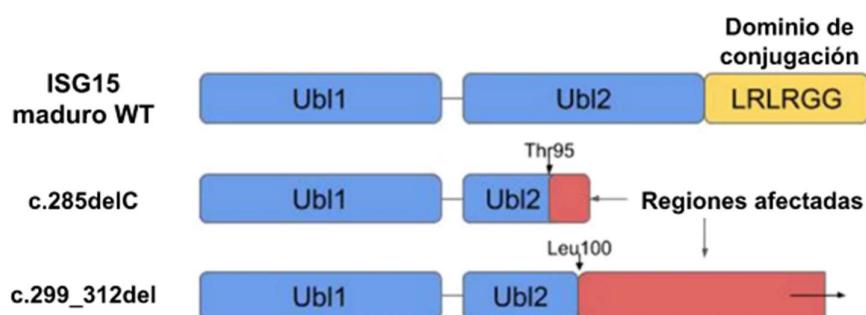


Figura 19: ISG15 consta de dos dominios *UBL* (azul) que están unidos por un residuo de prolina. ISG15 se sintetiza como un precursor de 17 kDa que se procesa proteolíticamente en una forma madura de 15 kDa. Este procesamiento expone un motivo LRLRGG carboxi-terminal, requerido para ISGylation (amarillo). La variante c.285delC produce un desplazamiento de marco de lectura después del residuo Treonina 95, dando como resultado la ganancia de un codón stop prematuro río abajo. La variante c.299_312del produce un desplazamiento del marco de lectura después del residuo Leucina 100.

ISG15 es un miembro de la familia de la ubiquitina, que incluye también a los modificadores similares a la ubiquitina (Ubls). La ubiquitina y los Ubls intervienen en la regulación de diversas actividades celulares, como la estabilidad de las proteínas, el tráfico intracelular, el control del ciclo celular y la modulación inmunitaria. Mientras que algunas Ubls están constitutivamente presentes en la célula, otras son inducidas por diferentes estímulos [208]. ISG15 es una de las proteínas similares a la ubiquitina más fuerte y rápidamente inducidas, que existe en tres estados distintos: libre dentro de la célula, secretada al espacio

extracelular o conjugada con proteínas target. ISG15 y los miembros de la cascada enzimática (E1, E2 y E3), que median la conjugación de la ISG15 (ISGilación) con un residuo de lisina de las proteínas target, son fuertemente inducidos por los interferones de tipo I. Se ha avanzado en la definición de algunos de los mecanismos por los que la ISGilación de proteínas tanto virales como del huésped inhibe la replicación viral y las estrategias de evasión viral que han evolucionado para eludir la acción de ISG15, pero los últimos avances en este campo han puesto de manifiesto la complejidad de esta vía [209]. Además, ISG15 existe como una proteína no conjugada que funciona como una citoquina y puede interactuar con varias proteínas intracelulares mediante interacciones no covalentes, regulando la replicación viral y las respuestas inmunitarias del huésped [210]. La forma de ISG15 secretada puede inducir la proliferación de células NK y actuar como factor quimiotáctico para los neutrófilos y como citoquina inductora de IFN-gamma, desempeñando un papel esencial en la inmunidad antimicrobiana. Por otro lado, los principales mecanismos de acción conocidos mediante los cuales ISG15 ejerce su efecto protector frente a las infecciones virales son: la inhibición de la salida del virus, la conjugación con proteínas virales y la modificación de proteínas celulares [211].

La proteína ISG15 está compuesta por dos repeticiones en tandem de dominios ubiquitina, el dominio N-terminal es necesario para la transferencia, mediada por la ligasa E3, de ISG15 desde la enzima E2 UbCH8 a sus sustratos, aunque es prescindible en los pasos de activación y transtiolación [212]. Los dos dominios de ubiquitina de ISG15 juegan papeles críticos en la ISGilación, que es único entre los modificadores de ubiquitina y similares a la ubiquitina.

Ambas variantes *frameshifts* interrumpen claramente el segundo dominio, provocando la pérdida del extremo C-terminal que contiene una secuencia de aminoácidos muy conservada Leu Arg Leu Arg Gly Gly (LRLRGG), correspondiente al sitio de ISGilación, y evitando la producción de proteína ISG15 detectable [213] (Figuras 19 y 20). En comparación con la ubiquitina, que se encuentra altamente conservada entre distintas especies (cerca del 100% de identidad), la proteína ISG15 tiene una conservación entre especies relativamente baja. Su ausencia en numerosas especies eucariotas sugiere que no es un gen esencial, posibilitando su diversificación a lo largo de la evolución. Las variantes patogénicas en el gen *ISG15* son una de las causas recientemente definidas de un grupo de Interferonopatías de tipo I [214].

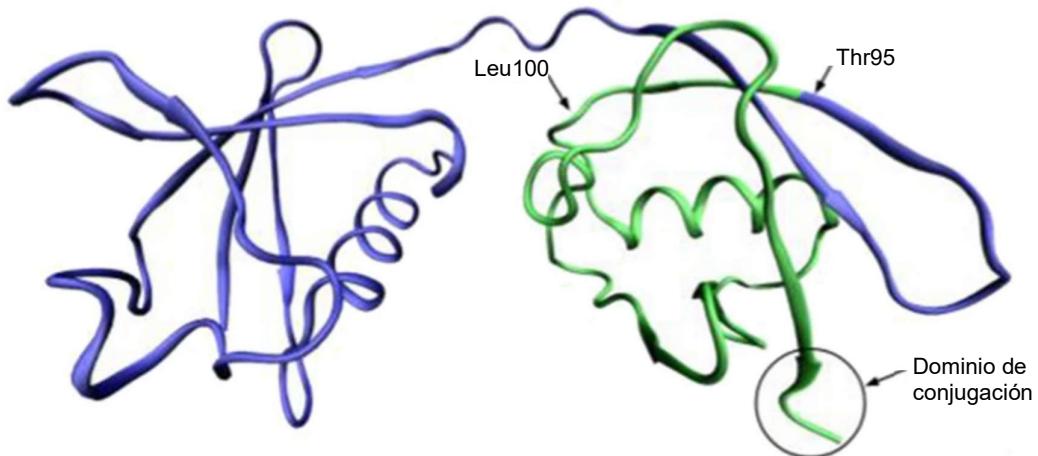


Figura 20. Estructura cristalográfica de ISG15 humano. El área afectada por las variantes se resaltó en verde. Thr95 marca el comienzo del cambio del marco de lectura asociado a la variante 285_del. Leu100 marca el comienzo del cambio del marco de lectura asociado a la variante 299_312del. El extremo C-terminal de la estructura corresponde a los primeros 3 residuos del dominio de conjugación (LRL).

Variantes bialélicas de pérdida de función en el gen *ISG15* se describieron inicialmente en pacientes con Inmunodeficiencia 38 (OMIM: 616126), un trastorno autosómico recesivo que confiere Susceptibilidad mendeliana sindrómica a las enfermedades micobacterianas (MSMD), y que afecta a aproximadamente 1 de cada 50.000 individuos expuestos a micobacterias ambientales o a cepas de la vacuna BCG [215]. Sorprendentemente, y a pesar de las funciones antivirales de ISG15 descritas en ratones, los pacientes descritos hasta la fecha no presentan fenotipo de susceptibilidad viral [215]. Sin embargo, la deficiencia hereditaria de ISG15 se ha manifestado con dos presentaciones distintas hasta la fecha: susceptibilidad a la enfermedad micobacteriana y calcificaciones intracraneales por la producción hipomórfica de interferón-II (IFN-II) o respuesta excesiva a IFN-I, respectivamente (Figura 21).

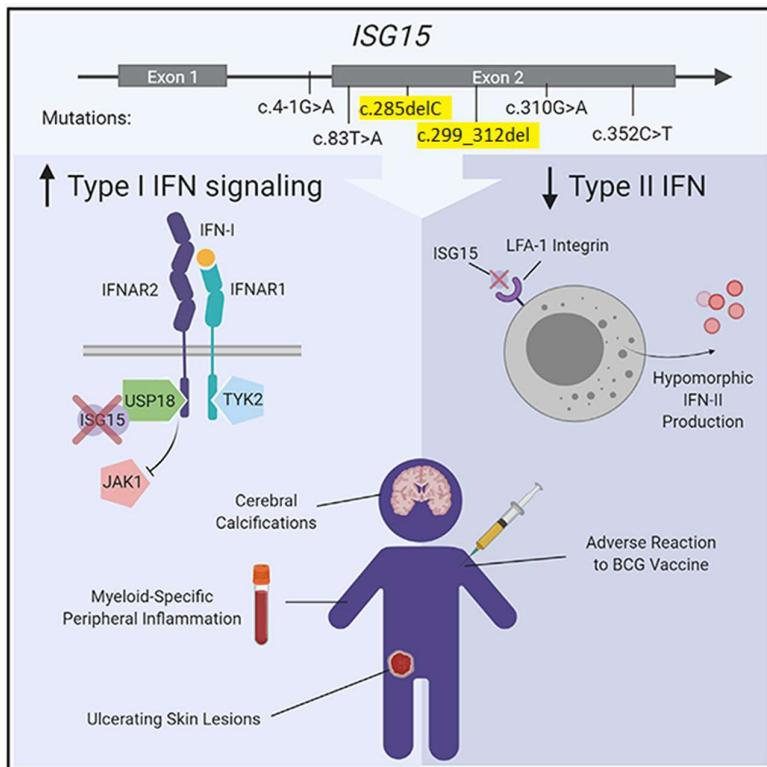


Figura 21. Figura extraída del abstract gráfico del trabajo de Martin-Fernandez et al. [216]. Panel superior: variantes reportadas. Panel inferior: efectos moleculares de las variantes en la señalización del Interferón de tipo I y II y fenotípicos generales.

Las pruebas genéticas realizadas en pacientes con pérdida de expresión de *ISG15* han revelado la importancia de esta proteína en la regulación de la respuesta al interferón de tipo I. Desde el punto de vista bioquímico, la falta de *ISG15* intracelular da lugar a niveles inestables de *USP18* [214,217], un potente regulador negativo del receptor de IFN-I, favoreciendo la señalización continua de JAK-STAT *downstream* [218] y a un aumento de los niveles de expresión de genes estimulados por interferón (ISG) en la sangre.

El circuito *ISG15-IFN-γ* opera principalmente entre los granulocitos y las células NK [219]. La expresión elevada de ISG y la calcificación cerebral son características consistentes en pacientes con variantes patogénicas bialélicas en *ISG15*, que presentan a su vez, un fuerte solapamiento neuroradiológico con el Síndrome de Aicardi-Goutières (AGS) [220]. Los médicos mencionaron cierta similitud con pacientes con Vasculopatía asociada a STING de inicio en la infancia (conocida como síndrome SAVI, por sus siglas en inglés), un fenotipo inflamatorio sistémico de inicio temprano caracterizado por una vasculopatía cutánea grave y una importante enfermedad pulmonar intersticial. El síndrome SAVI suele ser causado por una mayor sensibilidad o activación independiente del ligando (constitutiva) de un componente receptor no nucleico (por ejemplo, una molécula adaptadora) de la vía de

señalización inducida por el IFN, como en el caso de las variantes patogénicas *de novo* y heredadas en el gen *TMEM173*, que codifica para la proteína STING (estimuladora de los genes del interferón) [221].

Gracias al hallazgo de ambas variantes, se estableció una colaboración con científicos del Imagine Institute, INSERM, Universidad de París, Francia y del Icahn School of Medicine at Mount Sinai, NY, USA. Ambos grupos vienen estudiando la proteína ISG15 hace años y encontraron que celularmente, se observó una sorprendente especificidad en la respuesta de IFN-I, que anteriormente se suponía que era universal. En sangre periférica, las células mieloides mostraron las firmas de IFN-I más sólidas. En la piel afectada, la señalización de IFN-I se observa en los queratinocitos de la epidermis, el endotelio y los monocitos y macrófagos de la dermis. Estos hallazgos permitieron identificar cuáles son las células específicas que causan la inflamación circulatoria y dermatológica y, de esta manera, amplían el espectro clínico de la deficiencia de ISG15 a presentaciones dermatológicas como un tercer fenotipo codominante, junto con manifestaciones infecciosas y neurológicas [216].

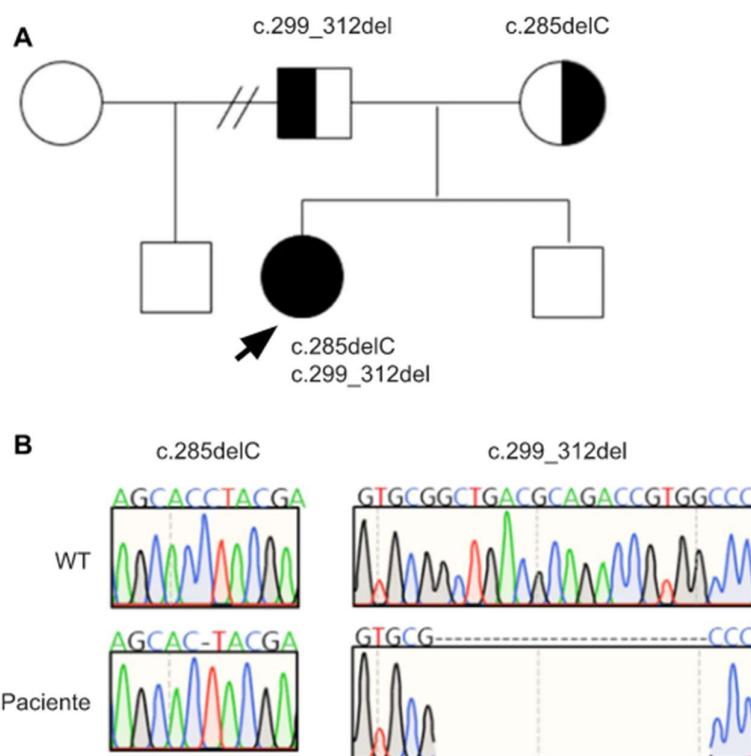


Figura 22. A. Segregación familiar de las variantes por Sanger. Se evidencia la cosegregación de ambas variantes en la paciente (indicada con una flecha). Cada progenitor es portador de una variante en heterocigosis. **B. Cromatogramas de la secuencia parcial de ISG15 en la paciente.** Se muestran las secuencias de ADN de los alelos wild type (WT, tipo salvaje) e ISG15 mutados.

En la Tabla 11 se resumen las características y hallazgos clínicos de los pacientes portadores de variantes bialélicas en *ISG15* previamente reportados en la literatura, incluyendo a la paciente presentada en esta tesis.

Tabla 11. Casos de pacientes reportados previamente en la literatura científica (Science 2012 [215], Nature 2015 [214], Cell Report 2020 [216]).

MSMD=Susceptibilidad mendeliana a micobacterias; IBCG= calcificación de los ganglios basales. Los pacientes de China e Irán son hermanos. ** Estas variantes se han reportado en el Cell Report como c.284del y c.297_313del.

País de origen	Paciente	Presentación clínica			Hallazgos genéticos			Otros Comentarios
		Vacunación BCG	Lesiones cutáneas	Calcificación	Herencia	Variantes detectadas	Caracterización	
China	P I femenino	No	No	De los ganglios basales	AR	c.163C >T, p.Gln55*	Pérdida de expresión	Epilepsia.
	P II femenino							-
	P III femenino							-
Turquía	Paciente femenino	Sí	Indeterminado	Intracraneal	AR	c.379G>T, p.Glu127*	Pérdida de expresión	Linfadenopatía luego de vacunación con BCG (MSMD) Lesión herpética en la zona escapular.
Irán	P I masculino		No	Intracraneal	AR	c.336_337insG, p.Leu114fs	Pérdida de expresión	Linfadenitis cercana al sitio de vacunación (MSMD).
	P II masculino		No	Intracraneal	AR			Úlceras y linfadenopatías postulantes luego de vacunación con BCG.

Argentina	Paciente de sexo femenino reportada en esta tesis	Sí	Sí	De los ganglios basales	AR	c.285delC, p.Tyr96fs y c.299_312del, p.Leu100Argfs (NM_005101.4, GRCh37(hg19)**	Pérdida de expresión Pérdida de expresión	Sin BCGitis después de la vacunación con BCG al nacer, Manifestaciones respiratorias recurrentes y neumonías, ulceraciones cutáneas de pierna y vulva.
US	Paciente de sexo femenino	No	Sí	Intracraneal	AR	c.310G>A, p.Val104Met y c.352C>T, p.Gln118*	Pérdida de función Pérdida de expresión	Lesiones cutáneas en el área inguinal.
España	Paciente de sexo femenino	No	Sí	Cerebelares y de los ganglios basales	AR	c.285del, p.Tyr96fs y c.83T>A, p.Leu28Gln	Pérdida de expresión Pérdida de función	Úlcera cutánea de pierna, vulva, ingle y región perianal.
Arabia Saudita	PI masculino	Sí	Sí	Intracraneal	AR	c.4-1G>A p. Gly2AlafsTer4*	Pérdida de expresión	Linfadenopatías inguinales luego de vacunación con BCG (MSMD) Lesiones en el cuello y regiones inguinales y alopecia en la región occipitoparietal izquierda del cuero cabelludo.
	PII masculino	Sí	Sí	Intracraneal		Alteraciones de splicing	Pérdida de expresión	Lesiones cutáneas en la axila derecha.

Discusión

En este caso, la ampliación del análisis del exoma de la paciente incluyendo genes asociados a Inmunodeficiencias Primarias, favoreció la detección de dos variantes (p.Tyr96ThrfsTer5 y p.Leu100Argfs) de alto impacto en el gen *ISG15*. Las mismas se incluyeron en el reporte y la médica dermatóloga postulante del Hospital Militar fue la encargada de informar a la familia. Unos meses después, recibimos novedades del sector de Inmunología del Hospital Garrahan, contándonos que la paciente se encontraba en seguimiento por pediatras e inmunólogos e internada. Adicionalmente, nos informaron que estaban en comunicación con el Instituto de París y el Icahn School of Medicine at Mount Sinai, quienes se encontraban estudiando pacientes con fenotipos similares y variantes en *ISG15*. Como se mencionó anteriormente, la deficiencia de *ISG15* consiste en dos formas clínicas superpuestas: un fenotipo de inmunodeficiencia y una interferonopatía de tipo I [222]. Se describió por primera vez en dos familias de Turquía e Irán como una susceptibilidad mendeliana a la enfermedad micobacteriana (MSMD) [215], en los cuales las variantes detectadas alteraron la secreción leucocitaria de *ISG15*. Al inducir el IFNy, *ISG15* desempeña un papel esencial en la optimización de la respuesta inmunitaria antimicobacteriana de los linfocitos Th1 y los macrófagos [223]. También regula negativamente la vía de señalización IFNR mediante la estabilización de USP18 [224] y se ha descrito una regulación positiva del IFN tipo I, caracterizada principalmente por un fenotipo neurológico con calcificaciones de los ganglios basales y epilepsia [225]. Recientemente, nuestros colaboradores notificaron variantes de pérdida de función de *ISG15* y signos de activación de IFN en queratinocitos, células endoteliales cutáneas y células mieloides de sangre periférica en cinco pacientes afectados por lesiones ulcerosas cutáneas en las zonas perianal, inguinal, cuero cabelludo, cuello y axilar. Estos hallazgos pueden sugerir otro fenotipo con un patrón único de localización cutánea de la inflamación vasculopática [216].

Como resultado del hallazgo de un fenotipo novedoso, se pudo establecer un trabajo multidisciplinario con las instituciones mencionadas y al mismo tiempo, incluir a la paciente en un *Case report* que permitió reportar la ampliación del fenotipo asociado a la presencia de variantes bialélicas en ese gen. El mismo se caracterizó por incluir enfermedad pulmonar recurrente y manifestaciones inflamatorias cutáneas, con la particularidad de no evidenciar infecciones micobacterianas.

Los estudios y experimentos llevados a cabo por los colaboradores del Instituto de París y el Icahn School of Medicine at Mount Sinai permitieron destacar el papel de *ISG15* como factor inmunomodulador, cuyas variantes de pérdida de función pueden dar lugar a presentaciones clínicas heterogéneas. Gracias a su trabajo, se pudo establecer que ambas

variantes detectadas en esta paciente provocaron la pérdida completa de la producción y, por lo tanto, función de la proteína resultante.

Nuestro trabajo resalta la capacidad de la secuenciación exómica de permitir la detección de variantes patogénicas en pacientes con disregulación inmune con gran precisión, facilitando el arribo a un diagnóstico certero, incluso en casos aislados (no familiares). En esta ocasión, como se mencionó anteriormente, la detección de estas variantes permitió la ampliación del fenotipo reportado para pacientes con variantes bialélicas en *ISG15*. En lo que respecta al mecanismo molecular que conduce a la enfermedad, la hipótesis de los profesionales es que podría haber un factor disparador como la infección, el estrés o eventos de vacunación que podría desencadenar las manifestaciones inflamatorias de la enfermedad multisistémica observada, que inicialmente apareció con ulceraciones en la piel y continuó con afectación pulmonar. Los médicos plantearon iniciar un tratamiento con inhibidor de JAK/STAT (JAKinib), inhibidor de la vía de señalización intracelular *Janus kinase/signal transducer and activator of transcription*, pero lamentablemente perdieron el seguimiento de la paciente y no pudieron evaluar la eficacia del tratamiento pensado con el objetivo de mejorar sus cuadros pulmonares.

Caso III:

Proyecto PAMPA

Introducción

El concepto de enfermedad autoinflamatoria sistémica (EAI) fue propuesto en 1999 por el doctor Kastner para agrupar algunas enfermedades poco frecuentes con manifestaciones clínicas similares (episodios febriles e inflamatorios recurrentes), y se estableció que la ausencia de autoanticuerpos a títulos elevados o de células T específicas de antígenos propios, eran características diferenciales respecto de las enfermedades autoinmunes [226]. Desde entonces, el número de enfermedades autoinflamatorias hereditarias ha ido aumentando progresivamente, debido principalmente a un creciente conocimiento del funcionamiento del sistema inmune, y a los avances en genética humana.

En la Tabla 12 se incluyen algunos ejemplos de EAI poligénicas o multifactoriales y se las distingue de las formas monogénicas, divididas en tres grupos: Infammosomopatías, Interferonopatías y Relopatías. Adicionalmente, se las clasifica de una manera gráfica teniendo en cuenta la vía afectada, sus manifestaciones clínicas más destacadas, y el *target* terapéutico [227–231].

Tabla 12. Enfermedades autoinflamatorias monogénicas y poligénicas: clasificación, vía afectada, manifestaciones clínicas características y target terapéutico [227–231].

Monogénicas				Poligénicas
Inflamasomopatías	Interferonopatías	Relopatías	Otras	
<ul style="list-style-type: none"> - FMF - HIDS/MKD - TRAPS - CAPS - PAPA 	<ul style="list-style-type: none"> - Síndrome de Aicardi-Goutières - SAVI - CANDLE - Síndrome COPA 	<ul style="list-style-type: none"> - HA20 - Deficiencia de OTULIN - Síndrome de Blau 		<ul style="list-style-type: none"> ● Enfermedad de Behçet ● PFAPA ● CRMO/CNO ● Síndrome SAPHO (forma esporádica) ● Enfermedad de Crohn, colitis ulcerosa ● AIJs/ Enfermedad de Still ● Síndrome de Schnitzler ● Psoriasis/artritis psoriásica ● Gota ● Otras
Vía afectada				
IL-1	IFN tipo I	NF- κ B IL-1, IL-6, TNF		
Clínica y tratamiento				
Fiebre y rash <i>Distinto espectro de severidad</i>	Calcificación CNS, vasculopatía y fibrosis acral y pulmonar	Granuloma, Uveítis, Úlcera		
Sobreactivación del inflamasoma	Respuesta aumentada a IFN	Ubiquitilación anormal		
Tratamiento: Inhibidores de IL-1	Tratamiento: Inhibidores JAK	Tratamiento: Anti-TNF		

FMF, Fiebre Mediterránea Familiar; MKD/HIDS, Deficiencia de mevalonato quinasa (Síndrome de hiper-IgD); TRAPS, Síndrome Periódico Asociado al Receptor del Factor de Necrosis Tumoral; CAPS, Síndrome de Fiebre Periódica Asociada a Criopirina; PAPA, Síndrome de artritis piógena estéril, pioderma gangrenoso y acné; SAVI, Vasculopatía asociada a STING con inicio en la infancia; Síndrome de dermatosis neutrofílica crónica atípica con lipodistrofia y temperatura elevada; HA20, Haploinsuficiencia A20; DADA2, Deficiencia de la adenosina desaminasa-2; DIRA, Deficiencia del antagonista del receptor de la Interleucina 1; DITRA, Deficiencia del antagonista del receptor de la Interleucina-36; PFAPA, Fiebre periódica asociada a adenitis, faringitis y estomatitis aftosa; CNO/CRMO, Osteomielitis crónica no bacteriana/osteomielitis crónica multifocal recurrente; AIJs, Artritis idiopática juvenil sistémica; SAPHO, Sinovitis, Acné, Pustulosis, Hiperostosis y Osteítis; NF- κ B, Factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas; TNF, Factor de necrosis tumoral, IFN: Interferón, IL: interleucina.

Como se muestra en la Tabla 12 y según su origen, las Enfermedades autoinflamatorias se pueden clasificar en [232]:

- Enfermedades autoinflamatorias monogénicas: son aquellas causadas por variantes en genes que codifican proteínas implicadas en vías inflamatorias, cuya alteración conduce a una respuesta inflamatoria sistémica exagerada por la producción descontrolada de citocinas proinflamatorias, principalmente la IL-1 β , en el caso de las inflamasomopatías (FMF, TRAPs, MKD y CAPS). La mayoría son heredadas y sigue un patrón de herencia mendeliano dominante o recesivo, aunque algunas enfermedades se han descrito asociadas a variantes de *novo*. Generalmente, se presentan durante la infancia o adolescencia, pero en algunas ocasiones alcanzan la edad adulta debido al retraso diagnóstico. Se caracterizan por episodios de inflamación sistémica órgano-específica, con un amplio espectro de las características inflamatorias.
- Enfermedades autoinflamatorias poligénicas o multifactoriales: son enfermedades que presentan una base autoinflamatoria pero etiología genética desconocida [233]. Algunas de ellas están mediadas por la IL-1, ya que responden a su bloqueo específico. Entre éstas se encuentran el Síndrome de PFAPA [234], Osteomielitis crónica no bacteriana/osteomielitis crónica multifocal recurrente (CRMO/CNO), la Enfermedad de Crohn, la Artritis Idiopática Juvenil Sistémica (AIJ sistémica), la enfermedad de Still del adulto, el Síndrome de Schnitzler [233], el Síndrome SAPHO (acrónimo de sinovitis, acné, pustulosis, hiperostosis y osteítis) y la gota, entre otras.
- Enfermedades autoinflamatorias indiferenciadas (o indefinidas): corresponden a cuadros o síndromes clínicos que cursan con fiebre recurrente o persistente, normalmente junto a otras manifestaciones inflamatorias, en los que se han descartado las causas infecciosas, neoplásicas y autoinmunes, y el estudio genético es negativo o no concluyente (p.ej. presencia de variantes benignas o de penetrancia incompleta, no asociadas con la expresión clínica). Además, suelen responder a fármacos utilizados en las enfermedades monogénicas, como la colchicina, dosis altas de glucocorticoides y/o agentes anti-IL-1. Actualmente, se reconoce que un fenotipo autoinflamatorio clásico o atípico, también puede ser ocasionado por variantes determinantes de mosaicismos somáticos, gonadales o gonosomales [234,235]. Los pacientes sin variantes identificables continúan siendo un enigma y el mosaicismo somático podría proveer una explicación a algunos de estos dilemas clínicos, generando nuevos desafíos en la comprensión y el tratamiento de las enfermedades de transmisión genética [236].

Actualmente, el diagnóstico de las EAI es clínico y se basa en las características fenotípicas evidenciadas en los pacientes. Las EAI más frecuentes se asocian a fiebre

periódica, erupciones cutáneas, compromiso de serosas y elevación de reactantes de fase aguda (RFA). Sin embargo, existe una gran variabilidad clínica y los autoanticuerpos, aunque usualmente ausentes, pueden observarse en ciertos síndromes [237].

La variabilidad clínica de las EAI, y el elevado número de genes asociados a las mismas convierte a este grupo de desórdenes en un caso ideal para su abordaje por técnicas de secuenciación masiva. Como se mencionó anteriormente, el diagnóstico genético es sumamente importante, pero debe ser realizado e interpretado correctamente. Por lo tanto, es necesario contar con un equipo profesional con acceso a tecnologías adecuadas y con la experiencia suficiente en la interpretación y validación de mutaciones de significado incierto y de nuevas variantes, que se detectan con mucha frecuencia. El correcto diagnóstico, permite en muchos casos un abordaje terapéutico con agentes biológicos que bloquean citocinas proinflamatorias, particularmente IL-1 β [238], anti-TNF [239] y bloqueantes de IFN [240,241], que han demostrado ser efectivos en muchos pacientes.

A comienzos del año 2018 se comenzó a gestar el **Programa Argentino de Medicina de Precisión en enfermedades Autoinflamatorias**, -mejor conocido como PAMPA (<https://apps.bitgenia.com/pampa/>)-, con el apoyo de Novartis y el acompañamiento de un consejo asesor (*Advisory Board*), compuesto por médicos y bioquímicos especializados en Diagnóstico molecular, Inmunología y Reumatología. En esta tesis se incluyeron resultados obtenidos durante tres Ediciones del Proyecto, que incorporó a más de 160 pacientes con sospecha de EAI, con el principal objetivo de brindarles la posibilidad de acceder a un diagnóstico preciso de forma gratuita. De esta forma, se promovió la oportunidad de recibir un tratamiento adecuado en el contexto de una verdadera medicina personalizada.

Materiales y Métodos

La difusión del proyecto tuvo un enorme impacto en todo el territorio argentino y permitió la inclusión de pacientes provenientes de instituciones de salud tanto públicas como privadas a lo largo y ancho del país. La postulación de casos fue realizada a través de un formulario *online* por diversos profesionales de la salud, especializados en identificar pacientes con sospecha de Enfermedad Autoinflamatoria. La postulación de los casos incluyó además la firma por parte de los pacientes del consentimiento informado correspondiente, y la aprobación del mismo por parte del comité de ética de la institución postulante (ver sección correspondiente). Como criterio de selección se dio prioridad a aquellos casos en los cuales el profesional médico tuviera una fuerte sospecha clínica de EAI. Los escenarios clínicos más frecuentes de presentación de estas enfermedades, en función de lo establecido por Federici y Gattorno [242], fueron adaptados a las ediciones de PAMPA y contemplaron la inclusión de pacientes con:

- Fiebre recurrente o periódica;
- Fiebre crónica y eczema;
- Formas atípicas de artritis recurrente o crónica;
- Lesiones pustulosas e inflamación sistémica;
- Lesiones osteolíticas inflamatorias estériles;
- Lesiones cutáneas afebriles.

En todos estos escenarios, el ejercicio clínico de diagnóstico diferencial con enfermedades infecciosas, inflamatorias autoinmunes, y oncológicas resultó fundamental para enfocar el diagnóstico de estas entidades. Se descartaron aquellos casos con una fuerte sospecha de enfermedad multifactorial o que tuvieran una alta probabilidad de ser causados por modificaciones genéticas no detectables por medio de la secuenciación exómica. Todos los casos postulados fueron analizados en profundidad por los integrantes del Advisory Board y se seleccionaron aquellos que cumplieron con los criterios de inclusión. La totalidad de los casos analizados formaron parte de: i) protocolos investigación aprobados por los comités de ética de las instituciones que atienden los pacientes, con los consentimientos informados correspondientes; o alternativamente ii) casos enmarcados en un proceso de innovación clínica [243] donde el médico responsable del caso otorga y explica el consentimiento informado al paciente. El análisis se limitó a aquellos genes previamente consensuados con el profesional como directamente relacionados con el diagnóstico presuntivo, para disminuir la posibilidad de hallazgos incidentales.

Una vez aprobado el caso, e informados el médico y paciente, se solicitó la firma del consentimiento informado y la declaración jurada. Luego se procedió con la extracción de una muestra de sangre periférica y a la purificación de ADN de linfocitos circulantes en cada caso, generalmente mediante la utilización de kits comerciales QIAamp DNA Mini Kit (QIAGEN) - tercerizado a laboratorio asociado VZ o a la institución/laboratorio del Hospital al cual pertenecía el paciente-.

La captura exómica se realizó con el kit de “Agilent SureSelect Human All Exon V5 (Edición 2018)”, “Agilent SureSelect Human All Exon V7 (Edición 2019-2020)” y las muestras fueron secuenciadas mediante la tecnología “Illumina HiSeq 4000” con una longitud de lectura de 100 pares de bases y una profundidad promedio de 100X. Todas las muestras fueron anonimizadas en todos los pasos del análisis, desde la extracción de sangre hasta la entrega de resultados finales al médico responsable.

Para la priorización de variantes se utilizó un panel de genes asociados a EAI, que fueron determinados por el Advisory Board y actualizados anualmente. El panel principal incluyó, entre otros, los genes contenidos en la Tabla 13.

Tabla 13. Panel de genes asociados con enfermedades autoinflamatorias.

Panel	Genes contemplados
Enfermedades Autoinflamatorias	<i>ACP5, ACTB, ADAM17, ADAR, ADGRE2, AGBL3, AP1S3, ARPC1B, CARD14, CECR1, COPA, DDX58, DNASE2, EGFR, IFIH1, IL10RA, IL10RB, IL1RN, IL36RN, ISG15, LACC1, LPIN2, MEFV, MVK, NCSTN, NLRC4, NLRP1, NLRP12, NLRP3, NLRP7, NOD2, OSMR, OTULIN, PLCG2, POLA1, POMP, PSMA3, PSMB4, PSMB8, PSMB9, PSMG2, PSTPIP1, RBCK1, RELA, RNASEH2A, RNASEH2B, RNASEH2C, RNF31, SAMHD1, SH3BP2, SHARPIN, SLC29A3, STIM1, TMEM173, TNFAIP3, TNFRSF11A, TNFRSF1A, TREX1, TRNT1, USP18, WDR1.</i>

En caso de no encontrar variantes de relevancia clínica en el panel principal, se procedió a analizar un panel ampliado en aquellos casos en los que el médico postulante así lo autorizó, contemplando el estudio adicional de genes de interés para cada caso en particular. En general, la búsqueda adicional se concentró en paneles de Inmunodeficiencias primarias, Enfermedades autoinmunes o Enfermedad inflamatoria intestinal, de acuerdo con la historia clínica del paciente en cuestión y en función de las sugerencias y recomendaciones realizadas por el Advisory Board y la médica genetista del equipo. Todos los casos fueron sometidos al protocolo de priorización de variantes diseñado previamente. Además, se buscó información adicional en Infevers [172,176,244,245] para cada una de las variantes encontradas dentro de los genes incluidos en esa base de datos. Finalmente, para las variantes reportadas en pacientes con FMF, MKD, TRAPS y CAPS, también se recurrió a las bases de datos Eurofevers [246], que recientemente incluyó información sobre la correlación genotipo-fenotipo de estas entidades. Procesados y analizados los datos, se generó un reporte de resultados que fue presentado al profesional médico correspondiente, quién fue responsable de entregarle una copia a la familia, orientarlos en la correcta interpretación de los datos y brindarles el asesoramiento genético pertinente.

Resultados

En el marco de las tres primeras ediciones del proyecto PAMPA se postularon un total de 239 casos, de los cuales 165 fueron aprobados para proceder con la secuenciación exómica. En concordancia con las prevalencias reportadas en Orphanet [247] para las enfermedades autoinflamatorias más frecuentes, se obtuvo un mayor número de postulaciones con sospecha de FMF (1-5/10.000), seguidos por casos con sospecha de TRAPS (Alemania: 1/1.785.000) y, posteriormente de CAPS (Francia: 1/360.000, USA: 1-2/1.000.000). Adicionalmente, se incluyeron pacientes con sospecha de enfermedades muy poco frecuentes, entre las que podemos resaltar los Síndromes de DIRA/DITRA, PAPA, Blau, Espondilo Condrodisplasia y Síndrome H (< 1/1.000.000). En la Figura 23 se puede observar el número de casos totales postulados, y aceptados de acuerdo con el diagnóstico presuntivo. Es importante remarcar también el carácter federal del proyecto, resaltando que se recibieron casos provenientes de numerosas provincias argentinas.

Para analizar la eficiencia diagnóstica, cada caso fue clasificado en una de las siguientes categorías: “positivo”, “hipótesis de trabajo” y “negativo”, de acuerdo al grado de asociación entre el diagnóstico presuntivo (o los síntomas) del paciente y el fenotipo patológico reportado para defectos en el gen donde se encontraron la(s) variante(s) potencialmente diagnóstica(s), y teniendo en cuenta también el tipo y nivel de evidencia disponible para cada una de las variantes reportadas (Ver Tabla 14).

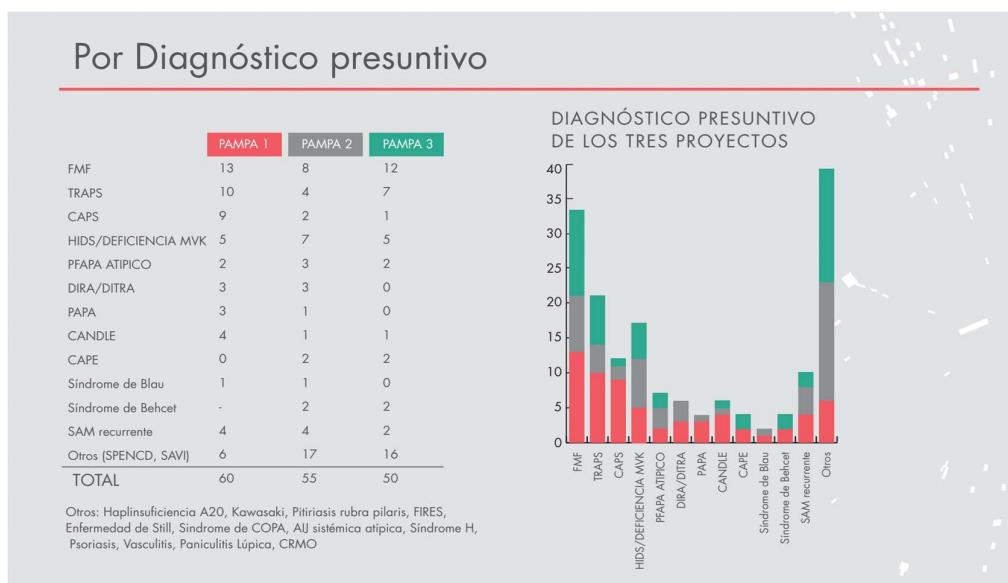


Figura 23. Número de casos aceptados agrupados en función del diagnóstico tentativo.
Se observa un mayor número de postulaciones de las EAI más frecuentes, en concordancia con lo esperado. Asimismo, llama la atención el gran número de postulaciones recibidas con sospechas diagnósticas muy poco frecuentes.

Se consideraron como "positivos" aquellos casos en los que se informó una (o más) variantes directamente relacionadas con la sospecha clínica y/o con el fenotipo evidenciado por el paciente, pudiendo considerarse el resultado como diagnóstico. De acuerdo con este criterio, el porcentaje de casos diagnosticados fue del 16%.

Aquellos casos en los que se informaron alguna/s potencial/es variante/s que podría/n considerarse como "hipótesis de trabajo", y explicar parcialmente el fenotipo observado fueron agrupados en la segunda categoría, abarcando éstas al 34% de los casos.

Finalmente, se consideraron como resultados "negativos" aquellos casos donde no se encontró ninguna variante relevante para ser informada, tanto en el/los gen/es directamente relacionados con la sospecha clínica, como en otros genes que podrían asociarse con los signos y síntomas de los pacientes estudiados (50%).

Los resultados de la Tabla 15 muestran 16 variantes interesantes encontradas, potencialmente asociadas a EAI, en 12 de los genes candidato estudiados (*MEFV*, *NLRP3*, *MVK*, *NOD2*, *RIPK1*, *ACP5*, *SLC29A3*, *ISG15*, *TRAP1*, *COPA*, *NLRP12* y *ADA2*). Estos hallazgos resaltan la relevancia de este tipo de estudios sistemáticos para ampliar el conocimiento de la genética molecular subyacente a las EAI, y demuestra un gran potencial para lograr una mayor eficiencia diagnóstica futura.

Tabla 14. Número de casos “positivos”, “hipótesis de trabajo” y “negativos” en función del criterio mencionado.

Resultado	Observaciones	Número de casos
Positivo	<i>Se consideran como resultados “positivos” aquellos casos en los cuales se informaron una o más variantes directamente relacionadas con el diagnóstico clínico presuntivo.</i>	<u>Pampa 1:</u> 14 <u>Pampa 2:</u> 3 <u>Pampa 3:</u> 9 Total: 26
Hipótesis de trabajo	<i>Se consideran como “Hipótesis de trabajo” aquellos donde se informaron alguna/s potencial/es variante/s que podría/n explicar parcialmente el fenotipo observado.</i>	<u>Pampa 1:</u> 15 <u>Pampa 2:</u> 16 <u>Pampa 3:</u> 25 Total: 56
Negativo	<i>Se consideran como resultados “negativos” aquellos casos donde no se encontró ninguna variante relevante para ser informada tanto en el/los gen/es directamente relacionados con la sospecha clínica, como en otros genes que podrían estar asociados con los signos y síntomas evidenciados.</i>	<u>Pampa 1:</u> 31 <u>Pampa 2:</u> 36 <u>Pampa 3:</u> 16 Total: 83
Total	<i>Número total de casos incorporados, secuenciados, procesados y analizados para todas las ediciones del Proyecto.</i>	165

Tabla 15. Variantes detectadas en genes seleccionados. Se distinguen el número de variantes previamente reportadas como patogénicas o probablemente patogénicas en la literatura/bases de datos de aquellas “noveles” que presentaron evidencia de asociación con Enfermedades Autoinflamatorias. **D:** Número de variantes totales **detectadas**. **PR:** Número de variantes **previamente reportadas** en la literatura como Patogénicas o Probablemente patogénicas.

Gen	D	PR	Variantes Noveles potencialmente asociadas con EAI		
MEFV	10	3	1	NM_000243.2:c.2230G>T - p.(Ala744Ser) Clasificación ACMG: VUS (PM1, PM2, PP2, BP4)	
NLRP3	7	5	1	NM_001079821.2:c.980G>A - p.(Arg327Gln) Clasificación ACMG: VUS (PM2, BP4)	
NOD2	5	-	1	NM_022162.2:c.1558C>T - p.(His520Tyr) Clasificación ACMG: Probablemente patogénica (PS2, PM1, PM2, PP3)	
MVK	3	4	1	NM_000431.4:c.151C>T - p.(Leu51Phe) Clasificación ACMG: VUS (PM2, PP2, PP3)	
SLC29A3	2	-	2	NM_018344.5:c.728T>G - p.(Leu243Arg) Clasificación ACMG: VUS (PM2, PM1, PP3) NM_018344.5:c.1077_1084delCCTATGTG - p.(Asp359fs) Clasificación ACMG: Patogénica (PVS1, PM2, PP3)	
ACP5	2	1	1	NM_001111036.3:c.632T>C - p.(Ile211Thr) Clasificación ACMG: Probablemente Patogénica (PM1, PM2, PP2, PP3)	

				NM_005101.4:c.285del- p.(Tyr96ThrfsTer5) Clasificación ACMG: Patogénica (PVS1, PS3, PM2, PM3)
<i>ISG15</i>	2	-	2	NM_005101.4:c.299_312del- p.(Leu100Argfs) Clasificación ACMG: Patogénica (PVS1, PS3, PM2, PM3, PP3)
<i>RIPK1</i>	1	1	1	NM_003804.6:c.1814T>G- p.(Leu605Arg) Clasificación ACMG: Probablemente Patogénica (PM1, PM2, PP2, PP3)
<i>COPA</i>	1	-	1	NM_001098398.2:c.718T>G - p.(Trp240Gly) Clasificación ACMG: VUS (PM5, PM2, PP3, BP1)
<i>NLRP12</i>	4	-	2	NM_001277126.2:c.1463_1471del - p.(Gly488_Asp490del) Clasificación ACMG: VUS (PM2, PM4, PP3) NM_001277126.1:c.3154C>T- p.(Arg1052Ter) Clasificación ACMG: Patogénica (PVS1, PM2, PP3)
<i>TRAP1</i>	4	-	2	NM_016292.2:c.319T>G - p.(Ser107Ala) Clasificación ACMG: VUS (PM2, PP3, BP1) NM_016292.2:c.1711G>A - p.(Ala571Thr) Clasificación ACMG: VUS/Probablemente benigna (PM2, BP1, BP4)
<i>ADA2</i> (CECR1)	4	2	1	NM_001282225.2:c.145C>T- p.(Arg49Trp) Clasificación ACMG: Probablemente Patogénica (PM5, PM1, PM2, PP2, PP3)
<i>AP1S3</i>	1	1	-	

<i>RBCK1</i>	1	-	-	
<i>TNFAIP3</i>	2	-	-	
<i>TNFRSF1A</i>	2	-	-	
<i>PSTPIP1</i>	2	-	-	
<i>TNFRSF11A</i>	1	-	-	

De los casos mencionados en la Tabla 14, se seleccionaron 6 que resultaron de particular interés por las características particulares de la relación genotipo-fenotipo. La descripción clínica detallada y relevancia de los hallazgos moleculares reportados en 6 casos seleccionados por su representatividad se encuentran incluidos en los BOX 1-6.

BOX 1

Paciente de sexo femenino de 35 años a la fecha, con fiebre autolimitada y episódica de inicio temprano, dolor articular y abdominal, artritis, rash, urticaria, aftas, enfermedad intestinal indeterminada, alteración general del crecimiento, adenitis, hepatomegalia y esplenomegalia, reactantes de fase aguda persistentemente elevados y leucocitosis episódica. Se descartaron ALPS, Enfermedad celíaca, AIJ sistémica, enfermedades infecciosas y linfoproliferativas. El cariotipo evidenció mosaicismo para Síndrome de Turner. Como diagnóstico tentativo, los médicos postulantes contemplaron Deficiencia de Mevalonato Kinasa y, como alternativo, Síndrome autoinflamatorio asociado a la criopirina (CAPS).

A partir de la secuenciación exómica de la paciente se buscaron variantes con probada patogenicidad dentro del panel de genes asociados con Enfermedades Autoinflamatorias, principalmente en *MVK* (asociado a Deficiencia de Mevalonato Kinasa) y *NLRP3* (asociado a CAPS).

Se encontraron dos variantes de tipo *missense* en heterocigosis en el gen *MVK* (OMIM: 251170), que codifica para la mevalonato quinasa, una enzima que participa en la síntesis de isoprenoides y colesterol.

La primera variante NM_000431.3:c.803T>C - p.(Iso268Thr) produjo un cambio en la región codificante (ENST00000228510) del exón 9, provocando la sustitución del residuo Isoleucina 268 por Treonina dentro del dominio conservado C-terminal GHMP kinasa. La misma posee predicción patogénica por numerosos algoritmos bioinformáticos, se encuentra reportada en Infevers, UniProt y ClinVar como patogénica, asociada con Síndrome de hiperinmunoglobulina D (HIDS; OMIM: 260920) y Aciduria Mevalónica (OMIM: 610377). Sander et. al. realizó estudios *in vitro* en bacterias E. coli,

comparando el porcentaje de actividad residual de la proteína mevalonato kinasa en bacterias portadoras de la variante I268T en función de la actividad máxima registrada en bacterias *wild type* para el gen *MVK*, concluyendo que la misma produce un efecto deletéreo y reduce la actividad de la proteína resultante. Por estas razones, esta variante se clasifica como patogénica.

La segunda variante, NM_000431.3:c.644G>A - p.(Arg215Gln), corresponde a un cambio en la región codificante (ENST00000228510) del exón 7, produciendo la sustitución del residuo Arginina 215 por Glutamina dentro del tercer dominio S5. La misma posee predicción patogénica por numerosos predictores bioinformáticos y se encuentra reportada en Infevers y UniProt como patogénica, en base a estudios funcionales, asociada a HIDS (OMIM: 260920). Por estas razones, esta variante se clasifica como probablemente patogénica.

Variantes en homocigosis o heterocigosis compuesta en este gen son responsables del Síndrome de Hiperinmunoglobulina D (HIDS; OMIM: 260920) y de un fenotipo más grave conocido como Aciduria Mevalónica (OMIM: 610377), ambos se tratan de trastornos autoinflamatorios de herencia autosómica recesiva, caracterizados principalmente por episodios recurrentes de fiebre asociada a linfadenopatía, artralgia, enfermedad inflamatoria intestinal y rash [248]. La aciduria mevalónica se caracteriza además por dismorfología, retraso psicomotor, ataxia cerebelosa progresiva y crisis febres recurrentes, que suelen manifestarse en la primera infancia [249]. La deficiencia de *MVK* (MKD) conduce al agotamiento del geranilgeranil pirofosfato, lo que resulta en la inactivación de RhoA [250]. Dado que la inactivación de RhoA activa el inflamasoma, la MKD conduce a una inflamasomopatía [251].

Este es un ejemplo de un caso exitoso, donde se hallaron variantes que se alinean con el diagnóstico tentativo y permitieron descartar el alternativo. Se ha demostrado que el tratamiento con antagonistas del receptor de interleucina 1 (por ejemplo, Anakinra) dio lugar a la remisión a largo plazo de las características digestivas y a la mejora de los marcadores inflamatorios en varios pacientes con HIDS, resultando también en una alternativa terapéutica disponible [252]. Adicionalmente, el tratamiento con Canakinumab, un anticuerpo monoclonal anti-IL-1 β , resulta eficaz para la MKD, lo que sugiere que la IL-1 β es un mediador común de estas enfermedades [253].

BOX 2

Otro ejemplo de diagnóstico exitoso y certero lo comprende el caso de un paciente pediátrico que comienza a ser estudiado por baja talla, atribuida a displasia esquelética, espondilodisplasia (platispondilia) metafisaria (ensanchamiento metafisario incipiente), fiebre prolongada asociada a episodio infeccioso >20 días (agudo) y picos esporádicos intermitentes, sin frecuencia establecida. Adicionalmente, presentó hepatomegalia, RFA elevados y leucocitosis recurrente. Se sospechó de Espondiloencondrodisplasia (SPENCD) basándose en la presencia de displasia esquelética, baja talla y disregulación inmune: hepatitis con disminución persistente de niveles de complemento, picos febriles intermitentes, marcadores inflamatorios moderadamente elevados e inmunodeficiencia humoral con impacto celular.

A partir de la secuenciación exómica del paciente se buscaron variantes con probada patogenicidad dentro del panel de genes asociados con Enfermedades Autoinflamatorias, principalmente en *ACP5* (asociado a SPENCD).

Se encontraron dos variantes de tipo *missense* en heterocigosis en el gen *ACP5* (OMIM: 171640), que codifica para la enzima Tartrato-resistente ácido fosfatasa tipo 5 (TRAP), principal regulador de la actividad de la osteopontina. La osteopontina fosforilada, una proteína implicada en la reabsorción ósea y en la regulación inmunitaria, se acumula en el suero, la orina y las células en cultivo de individuos con deficiencia de TRAP. Las células dendríticas derivadas de los pacientes presentan un perfil de citoquinas alterado y son más potentes que células control a la hora de estimular la proliferación de células T alogénicas en cultivos mixtos de linfocitos [254].

La primera variante detectada, *ACP5*(NM_001111036.3):c.632T>C - p.(Ile211Thr), corresponde a un cambio A>G en la región codificante (ENST00000218758) del exón número 6 y produce la sustitución del residuo Isoleucina 211 por Treonina en la proteína resultante. La variante posee muy baja frecuencia poblacional en la base de datos de GnomAD (ausencia de homocigotas), cuenta con predicción patogénica por numerosos algoritmos bioinformáticos y se encuentra previamente reportada en ClinVar como variante de significado incierto, resultando su clasificación final de significado incierto.

La segunda variante, *ACP5*(NM_001111036.3):c.791T>A - p.(Met264Lys) corresponde a un cambio A>T en la región codificante (ENST00000218758) del exón número 7, produciendo la sustitución del residuo Metionina 264 por una Lisina en la proteína resultante. La variante posee muy baja frecuencia poblacional en la base de datos de GnomAD (ausencia de homocigotas) y se encuentra previamente reportada como patogénica en homocigosis en Uniprot, asociada a la espondiloencondrodisplasia con disregulación inmune (SPENCDI - OMIM: 607944). En ClinVar posee una clasificación de significado incierto, contando con una entrada de significado incierto y una patogénica. Sin embargo, la clasificación final de la variante es probablemente patogénica.

Variantes bialélicas en el gen *ACP5* se asocian a Espondiloencondrodisplasia con disregulación inmune, una displasia inmuno ósea que combina las lesiones óseas metafisarias y vertebrales típicas de la espondiloencondrodisplasia (SPENCD) con disfunción inmunitaria y afectación neurológica. Se presenta con afectación esquelética característica, baja estatura,

micromelia rizomélica, aumento de la lordosis lumbar, tórax en tonel, dismorfias faciales y torpeza motora [254]. La afectación del sistema nervioso central incluye espasticidad, retraso mental y calcificaciones cerebrales, y la desregulación inmunitaria varía desde la autoinmunidad hasta la inmunodeficiencia. Se ha observado una gran variabilidad en las manifestaciones esqueléticas, neurológicas y autoinmunitarias en diferentes combinaciones, incluso entre integrantes de una misma familia [255].

El fenotipo de la paciente asocia displasia esquelética, estatura baja, platispondilia y hepatomegalia con episodios de fiebre podría explicarse por las variantes identificadas en el gen *ACP5*, de confirmarse la presencia de las mismas en trans (no se llevó a cabo confirmación por Sanger).

BOX 3

Paciente de sexo masculino de 49 años a la fecha, con fiebre autolimitada y recurrente, dolor articular, artritis, rash, maculopapular, urticaria, hipoacusia neurosensorial, conjuntivitis, adenopatías, reactantes de fase aguda elevados y leucocitosis recurrente. Posee diagnóstico de artritis reumatoide seronegativa. El cuerpo médico contempló como diagnóstico tentativo el Síndrome de Muckle-Wells.

A partir de la secuenciación exómica de la paciente, se buscaron variantes con probada patogenicidad dentro del panel de genes asociados con Enfermedades Autoinflamatorias, principalmente en el gen *NLRP3* (asociado con el Síndrome de Muckle-Wells).

Se detectó una variante en heterocigosis en el gen *NLRP3* (OMIM: 606416), NM_001079821.2:c.784C>T - p.(Arg262Ter), que corresponde a la ganancia de un codón *stop* prematuro en algunos transcriptos y al cambio de aminoácido de la Arginina en la posición 262 por un Triptófano en otros, con muy baja frecuencia poblacional (sin reporte en GnomAD) y predicción patogénica por numerosos algoritmos bioinformáticos. Esta variante produce un cambio en la región codificante (ENST00000336119) del residuo Arginina 262 por un Triptófano o por un codón terminal en el gen *NLRP3*, en el cual la mayoría de las variantes *nonsense* están reportadas como un mecanismo conocido del desarrollo de la enfermedad. La misma se ubica en el exón 3, dentro del dominio proteico de unión a nucleótidos NACHT y se encuentra reportada en Infevers, Eurofevers y ClinVar como patogénica, asociada al Síndrome periódico asociado a la criopirina (CAPS, OMIM: 191900), una enfermedad inflamatoria caracterizada por erupción cutánea episódica, artralgias y fiebre asociadas a sordera neurosensorial de inicio tardío y amiloidosis renal. Por lo mencionado anteriormente, la variante se clasifica como patogénica.

A través del modelado tridimensional de la proteína codificada por el gen *NLRP3*, se ha demostrado que variantes en el exón 3 pueden afectar la unión a nucleótidos y la oligomerización de las proteínas que forman parte del inflamasoma, produciendo la disfunción de este complejo multiproteico. *NLRP3* controla la secreción de citoquinas inflamatorias IL-1 β /18 y la piroptosis mediante el ensamblaje del inflamasoma [256]. Variantes patogénicas en *NLRP3* (como la reportada en este caso) conducen a una activación prolongada del inflamasoma y están relacionadas con un grupo de síndromes de fiebres periódicas poco frecuentes de herencia autosómica dominante, conocidos como criopirinopatías, entre los que se encuentran CINCA/NOMID, Síndrome de Muckle-Wells y Síndrome inflamatorio familiar por frío de tipo 1. De esta manera, el hallazgo concuerda con la orientación diagnóstica tentativa y explica la sintomatología clínica presentada por el paciente.

Se realizó secuenciación por Sanger (terciarizada) usando un par de *primers* específicos para validar la presencia/ausencia de esta variante en el hijo sintomático del paciente. La misma fue detectada también en heterocigosis y reportada.

De esta forma, la secuenciación exómica no sólo reveló la presencia de una variante patogénica en el probando, sino que también favoreció la identificación de la misma en su hijo sintomático por secuenciación dirigida por Sanger, permitiéndole al médico responsable arribar a un diagnóstico molecular certero, evaluar potenciales terapias o enfoques terapéuticos para los familiares afectados y derivarlos para asesoramiento genético.

BOX 4

Paciente pediátrico de sexo femenino, sin historia familiar de relevancia, con artritis exuberante de carpos y tobillos (bilateral) y tenosinovitis marcada, dolor articular, rash maculopapular generalizado, sarcoidosis confirmada por biopsia cutánea, BAAR negativo y PAS negativo para elementos micóticos. Los médicos a cargo propusieron como diagnóstico tentativo el Síndrome de Blau.

A partir de la secuenciación exómica de la paciente, se buscaron variantes con probada patogenicidad dentro del panel de genes asociados con Enfermedades Autoinflamatorias, principalmente en el gen *NOD2* (asociado con el Síndrome de Blau).

Como se esperaba, se detectó una variante en heterocigosis en el gen *NOD2* (OMIM:605956), NM_022162.2:c.1558C>T - p.(His520Tyr), que produce un cambio de nucleótido único C>T en el exón 4, generando la sustitución del residuo Histidina 520 por Tirosina en la proteína resultante. La misma posee muy baja frecuencia poblacional (ausencia de homocigotas en GnomAD), cuenta con predicción patogénica por varios algoritmos bioinformáticos y se encuentra reportada en Infevers como probablemente patogénica, habiéndose detectado en un paciente sintomático con diagnóstico de Síndrome de Blau (OMIM: 186580), un trastorno de herencia autosómica dominante que se caracteriza por incluir, entre los síntomas principales, artritis granulomatosa, uveítis y rash [257]. El síndrome de Blau se asocia, usualmente, a polimorfismos de nucleótido único de ganancia de función en el dominio NACHT del receptor inmune innato NOD2. NOD2 es un miembro de la familia de receptores de reconocimiento de patrones NLR citosólicos y constituye una proteína activada por el componente dipeptido murmil [258]. Una vez activa, interactúa con el receptor adaptador RIP2, para iniciar vías de señalización proinflamatorias que involucran el factor nuclear kappa B (NF-κB) y quinasas de estrés. De esta manera, desempeña un papel importante en la respuesta a la infección bacteriana, incluida la activación de la autofagia [258,259].

Adicionalmente, se analizó la segregación de la variante en el trío (paciente y ambos padres), demostrando la presencia de esta en heterocigosis únicamente en la paciente (*de novo*). Este hallazgo aportó la evidencia necesaria que permitió reclasificar esta variante de tipo VUS a probablemente patogénica, de acuerdo con las recomendaciones y guías del ACMG. Este caso representa un ejemplo interesante en el cual la sospecha diagnóstica del profesional a cargo se termina confirmando mediante la detección de una variante probablemente patogénica a partir de la secuenciación del exoma de la paciente, en conjunto con el análisis de la segregación familiar de la misma por Sanger.

BOX 5

Paciente pediátrico de sexo femenino, sin historia familiar de relevancia, que entre los síntomas principales presentó fiebre episódica, dolor óseo, dismorfias faciales y osteoarticulares, rash, lipodistrofia, alteración general del crecimiento, esplenomegalia, reactantes de fase aguda elevados y persistentes. Como diagnóstico clínico tentativo se contempló CANDLE y, como alternativo, Síndrome H.

A partir de la secuenciación exómica de la paciente se buscaron variantes con probada patogenicidad dentro del panel de genes asociados con Enfermedades Autoinflamatorias, principalmente en los genes *PSMB4*, *PSMA3*, *PSMB9*, *PSMB8* (asociados con CANDLE) y *SLC29A3* (asociado con Síndrome H).

Se detectaron dos variantes en heterocigosis en el gen *SLC29A3* (OMIM: 612373), que codifica para un transportador que regula el influjo y eflujo de nucleósidos, localizado en las membranas de lisosomas, mitocondrias y otras organelas intracelulares.

La primera variante NM_018344.5:c.728T>G - p.(Leu243Arg) produce un cambio de nucleótido único en la región codificante del exón 5, generando la sustitución del residuo Leucina 243 por Arginina (un cambio significativo desde una perspectiva bioquímica), dentro del dominio transportador de nucleósidos [260]. La misma posee muy baja frecuencia poblacional, cuenta con predicción patogénica por numerosos algoritmos bioinformáticos y no se encuentra previamente descrita en bases de datos ni publicaciones científicas.

La segunda variante, NM_018344.5:c.1077_1084delCCTATGTG - p.(Asp359fs), corresponde a una delección de 8 nucleótidos, que produce el desplazamiento del marco de lectura en la región codificante del exón 6 y la ganancia de un codón *stop* prematuro. La misma posee muy baja frecuencia poblacional y no se encuentra reportada en bases de datos clínicas ni publicaciones científicas.

La primera variante se clasificó como de significado incierto (por falta de evidencia adicional y no por poseer propiedades que se asocian con una variante benigna) y la segunda, como probablemente patogénica.

Variantes en homocigosis o heterocigosis compuesta en el gen *SLC29A3* están asociadas con el Síndrome de histiocitosis con linfadenopatía (OMIM: 602782), un grupo de condiciones que incluye al Síndrome H y produce signos y síntomas multisistémicos compartidos. Una característica común es la histiocitosis, crecimiento y acumulación de histiocitos en los tejidos, especialmente en los ganglios linfáticos del cuello, causando inflamación (linfadenopatía). Otros síntomas pueden incluir parches cutáneos hiperpigmentados con hipertricosis, diabetes, anomalías cardíacas, hipogonadismo, hepatoesplenomegalia, bajo peso y estatura y pérdida de la audición [261].

Variantes patogénicas en el gen *SLC29A3* que reducen o eliminan la actividad de la proteína codificada, conducen a la acumulación de nucleósidos en los lisosomas y otras estructuras celulares, resultando perjudicial para las mismas y afectando la producción de energía celular.

La enfermedad autoinflamatoria causada por la presencia de ambas variantes en *SLC29A3* generó en la paciente un estado de inflamación crónica con compromiso multisistémico progresivo y

severo:

- Hematológico: Anemia,
- Dermatológico: hipertricosis, hiperpigmentación, cambios esclerodermiformes,
- Oftalmológico: opacidades corneales.
- Óseo: osteopenia, camptodactilia, contracturas,
- Endocrinológico: hipocrecimiento severo, hasta el momento la niña sólo hizo hiperglicemias pero se debe monitorizar por su riesgo aumentado de diabetes mellitus,
- Ginecológico: hipogonadismo hipogonadotrófico,
- Cardíaco: taquiarritmia ventricular mejorada pero aún con riesgo aumentado de muerte súbita de origen cardíaco y necesidad de controles permanentes con cardiología y arritmología,

Dado que uno de los importantes efectos de las variantes en *SLC29A3* es aumentar la expresión de IL-6, los inhibidores de IL-6 (por ejemplo, Tocilizumab) han demostrado ser beneficiosos en estos pacientes. Por tal motivo inició tratamiento con Tocilizumab 8 mg/Kg/dosis cada 15 días según recomendaciones para Síndrome H.

En este caso, la validación de la segregación familiar de las variantes no pudo llevarse a cabo, ya que ninguno de los progenitores se encontró disponible para el análisis. Sin embargo, los hallazgos informados aportaron información de relevancia consistente con la sospecha clínica alternativa, posibilitando el arribo a un diagnóstico molecular preciso y ampliando el espectro fenotípico del poco frecuente Síndrome H.

BOX 6

Paciente de sexo femenino de 17 años a la fecha. Entre los síntomas principales presenta fiebres episódicas, dolor articular, dolor abdominal, artritis, osteomielitis estéril, inflamación intestinal indeterminada, uveítis, adenopatías, esplenomegalia, reactantes de fase aguda elevados y leucocitosis recurrente. El diagnóstico clínico tentativo de los médicos postulantes fue el de Síndrome de artritis piógena estéril, pioderma gangrenoso y acné (PAPA).

A partir de la secuenciación exómica de la paciente se buscaron variantes con probada patogenicidad dentro del panel de genes asociados con Enfermedades Autoinflamatorias, principalmente en el gen *PSTPIP1* (asociado al Síndrome de PAPA).

Para nuestra sorpresa y la de los médicos a cargo, se encontraron dos variantes patogénicas en el gen *MEFV* (OMIM: 608107).

La primera variante NM_000243.2:c.2040G>A - p.(Met680Ile) constituye un cambio de nucleótido único G>A en la región codificante (ENST00000219596) del exón 10, produciendo la sustitución del residuo Metionina 680 por Isoleucina, mientras que la segunda variante NM_000243.2:c.2080A>G - p.(Met694Val) genera un cambio de nucleótido único A>G en la región codificante (ENST00000219596) del exón 10, provocando la sustitución del residuo Metionina 694 por Valina en la proteína resultante.

Ambas poseen baja frecuencia poblacional, pero se las suele encontrar en poblaciones armenias, y fueron reportadas en Infevers y ClinVar como patogénicas, asociadas con Fiebre Mediterránea Familiar (FMF, OMIM: 608107), un trastorno autoinflamatorio de herencia autosómica dominante o autosómica recesiva que presenta, entre los síntomas principales, dolor abdominal, fiebre, artritis y dolor de pecho [262]. Se ha reportado que variantes de alta penetrancia en el exón 10, que se ubican dentro del dominio PRYSPRY -particularmente p.M680I y p.M694V- producen fenotipos graves en niños, pudiendo presentar eritema similar a la erisipela, úlceras cutáneas, esplenomegalia, vasculitis, dolor muscular recurrente aislado y osteomielitis multifocal recurrente crónica (CRMO) [263].

El gen *MEFV* tiene 10 exones y codifica para la pirina (o marenostrina), una proteína que se expresa principalmente en neutrófilos, eosinófilos y monocitos activados por citocinas, modula el procesamiento de IL-1 β , la activación de NF- κ B y regula la actividad de varias proteínas *target* involucradas en inmunidad innata y en procesos de autofagia (principalmente de componentes del inflamasoma), apoptosis e inflamación [264]. Asimismo, la pirina recluta y se une a la prolina-serina-treonina fosfatasa 1 (codificada por el gen *PSTPIP1*) y ambas interactúan con proteínas de actina en el proceso de migración celular.

Dado que las variantes asociadas a la enfermedad en *PSTPIP1* mejoran la unión de la pirina, se cree que el síndrome de PAPA y la FMF comparten una etiopatogenia común.

Varias formas de artritis se asocian con fiebre mediterránea familiar, incluida la mono/oligoartritis aguda en extremidades inferiores, espondilitis anquilosante, sacroileítis, artritis de la articulación de la cadera y artritis crónica juvenil. Las complicaciones graves que amenazan la vida, como la obstrucción intestinal, la amiloidosis renal y los síntomas infrecuentes resultan, en algunas

ocasiones, la única forma de manifestación de la fiebre mediterránea familiar [265]. Particularmente, la variante p.M694V -tanto en homocigosis como en heterocigosis- está asociada con amiloidosis [266].

Numerosas publicaciones han señalado que la FMF no puede ni debe considerarse una enfermedad con síntomas universales [267]. La asociación de variantes en el gen *MEFV* con trastornos que comparten uno o varios síntomas con Fiebre Mediterránea Familiar sugiere que las mismas podrían contribuir, en diferentes individuos, a manifestaciones fenotípicas complejas e indicativas de otro síndrome autoinflamatorio [266].

Al igual que en la FMF, los pacientes con síndrome de PAPA desarrollan derrames estériles ricos en neutrófilos en las articulaciones. Las manifestaciones cutáneas del síndrome de PAPA incluyen acné quístico, abscesos estériles en los sitios de inyecciones y lesiones ulcerativas purulentas en expansión conocidas como pioderma gangrenoso. Aunque estos hallazgos son más dramáticos de lo que se suelen observar en la FMF, las manifestaciones cutáneas de ambos se consideran dermatosis neutrofílicas [268].

Las similitudes clínicas entre el síndrome de FMF y PAPA pueden estar relacionadas con el hecho de que estos trastornos son causados por variantes respectivas en pirina y PSTPIP1, proteínas con una distribución celular y subcelular similar que interactúan entre sí [269].

Este caso constituyó un claro ejemplo en el cual la sospecha diagnóstica basada en la clínica de la paciente no coincidió con el diagnóstico molecular. El mismo le brindó al cuerpo médico y a sus familiares un nuevo horizonte y los orientó hacia un diagnóstico certero de Fiebre Mediterránea Familiar atípico. En este caso, se sugirió la validación de la presencia de las variantes por Sanger en la paciente y sus padres (para validar segregación en trans). Sin embargo, a la fecha, desconocemos si se llevó a cabo.

Desde una perspectiva molecular es de destacar que la mayoría de las variantes reportadas corresponden a variantes *missense* (una totalidad de 12, siendo las restantes 3 *frameshift* y 1 *nonsense*) y que por lo tanto son las más difíciles de interpretar y de asignar criterios de patogenicidad. Para profundizar este análisis y en el marco de una colaboración interna, en la Figura 24 se muestra un score de patogenicidad que utiliza información de las distintas bases de datos como GnomAD, Clinvar y el análisis de la conservación a lo largo de las proteínas para los genes *NLRP3*, *ISG15*, *NOD2* y *MEFV*. Con esta información se estableció un score que varía entre 0-1, el cual permite determinar zonas más o menos propensas a poseer variantes patogénicas o benignas. Como se puede observar para los 4 genes, todos presentan regiones donde las variantes tienden a ser patogénicas y regiones donde las mismas tienden a ser benignas. Interesantemente, las regiones benignas suelen ubicarse en los extremos, y específicamente en zonas con ausencia de dominios. La utilización de este análisis en relación con su contribución para la clasificación de variantes se hace evidente al analizar las variantes previamente reportadas (PR) y las potencialmente

noveles encontradas en estos 4 genes en el presente trabajo. Para *NLRP3*, las variantes se ubican en zonas neutras (amarillas) o con tendencia a ser patogénicas (Naranja/Rojo), encontrando la variante p.Arg327Gln en una zona naranja, lo que sostiene su valor diagnóstico. Lo mismo ocurre con las variantes noveles encontradas en *ISG15*, *NOD2* y *MEFV*, que se localizan en zonas con alta probabilidad de patogenicidad.

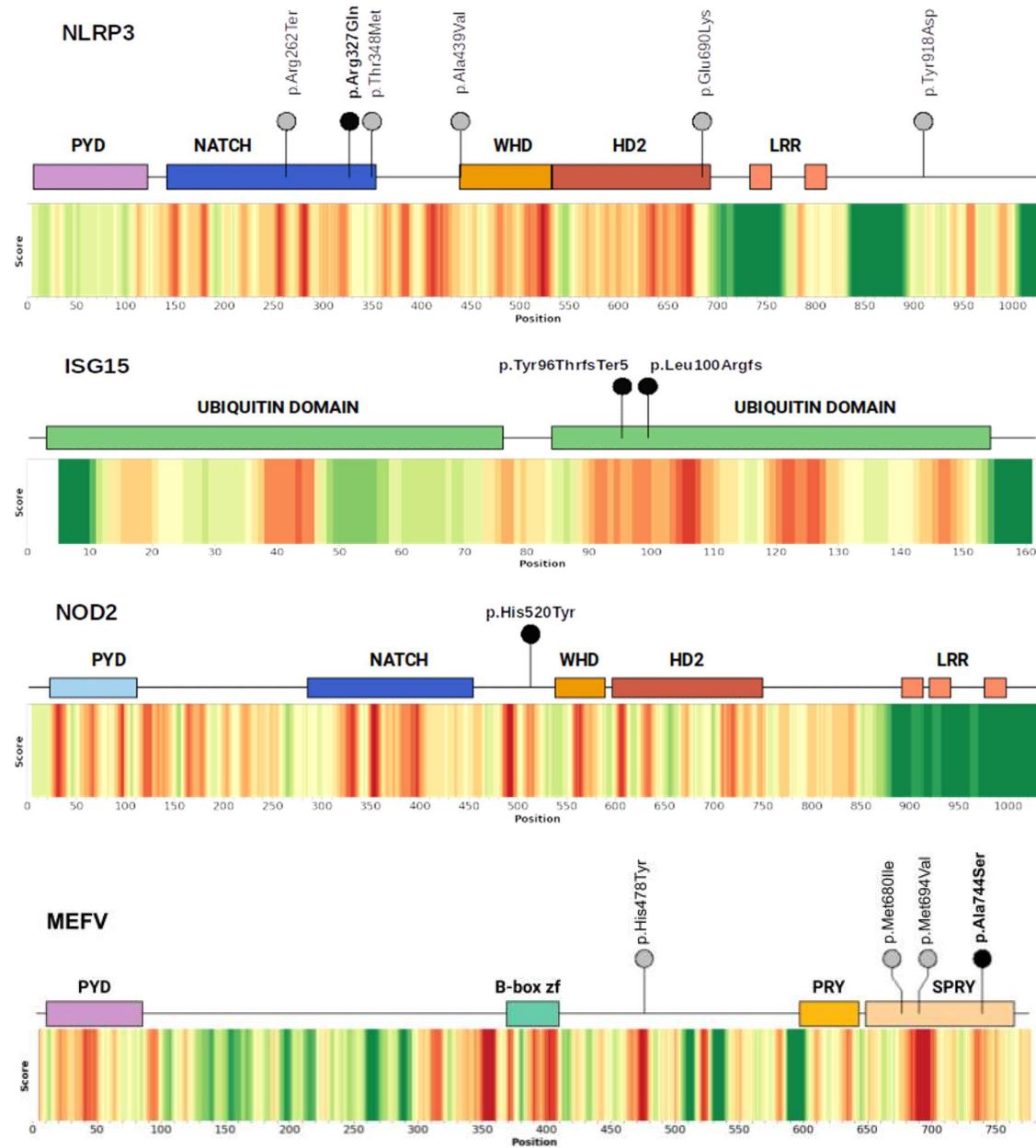


Figura 24: Representación gráfica del score de patogenicidad a lo largo de la secuencia codificante de los genes *NLRP3*, *ISG15*, *NOD2* y *MEFV*. Naranja/Rojo indica mayor probabilidad de variante patogénica, Verde indica mayor probabilidad de variante benigna. Para cada uno de los genes se indican las variantes noveles (círculos negros) y previamente reportadas (círculos grises) encontradas en los pacientes estudiados.

Discusión

Las enfermedades autoinflamatorias comprenden una amplia gama de síndromes causados por una desregulación de la respuesta inmune innata. Son difíciles de diagnosticar debido a su heterogeneidad fenotípica y expresividad variable. Por esta razón, el origen genético de la enfermedad sigue siendo indeterminado para una proporción importante de pacientes y/o puede ser pasado por alto con facilidad. Normalmente, los pacientes acuden a múltiples especialistas antes de obtener una sospecha diagnóstica, para lo que resulta clave la evaluación del fenotipo observado y el análisis de los estudios realizados. Es importante señalar que se trata de diagnósticos de "alto valor", ya que la mayoría de estos síndromes pueden controlarse de forma muy eficaz, mejorando drásticamente la calidad de vida y proporcionando protección contra el desarrollo de complicaciones irreversibles como la amiloidosis AA [100]. Este grupo de enfermedades constituye un ejemplo de éxito de la medicina traslacional inversa y han contribuido a desarrollar enormemente nuestra comprensión del sistema inmune innato.

El Proyecto PAMPA nació en el 2018 y sus tres primeras ediciones permitieron estudiar los exomas de 165 pacientes con sospecha de Enfermedades Autoinflamatorias, con el objetivo de identificar variantes que pudieran explicar los fenotipos observados y probar las ventajas y limitaciones del exoma "panelizado" aplicado al diagnóstico molecular. A pesar de su principal limitación de no incluir CNVs, regiones intrónicas, regulatorias, intergénicas, secuencias de pseudogenes y repetitivas, la secuenciación exómica resultó un enfoque apropiado para detectar probables variantes causales de enfermedades autoinflamatorias.

Si bien los resultados de los Proyectos PAMPA no incluyeron una cantidad de casos suficiente como para realizar afirmaciones estadísticamente significativas sobre la capacidad de la secuenciación exómica para diagnosticar de manera precisa las causas moleculares de las EAI, permitieron avanzar en el descubrimiento de nuevos genes y/o la detección de nuevas variantes potencialmente patogénicas en genes con conocida asociación con estas enfermedades. Adicionalmente, se pusieron a prueba las capacidades locales con sus virtudes y dificultades, para la implementación de estas tecnologías en contextos de investigación científica e innovación clínica, y compararlos con experiencias similares realizadas en países centrales. Los resultados de este capítulo muestran que la tasa de éxito -casos donde se arribó a un diagnóstico molecular- se estimó entre un 16-34% [53], evidenciando que tanto los procesos, como la eficiencia y calidad diagnóstica son equivalentes a aquellos observados a nivel internacional [54,270].

El programa posibilitó el hallazgo de algunas variantes noveles, que además de permitir establecer un potencial diagnóstico, constituyeron interesantes hipótesis de trabajo. Mediante la realización de estudios funcionales posteriores -como se ha realizado para las

variantes detectadas en *ISG15*- se buscó profundizar nuestra comprensión de los mecanismos asociados a la etiopatología molecular de las EAI y al funcionamiento del sistema inmune a nivel molecular-celular. Además, el análisis de las variantes novedosas de tipo *missense* (tanto las reportadas como las descartadas), en el contexto de aquellas conocidas y asignadas como Patogénicas o Benignas, permitió obtener una mayor comprensión del potencial efecto de las mismas, contribuyendo a evaluar una gran cantidad de variantes de significado incierto detectadas regularmente en estos genes.

Finalmente, y desde una perspectiva sociológica, la experiencia de estos proyectos permitió darle peso y reforzar la siguiente filosofía de trabajo interdisciplinario: si bien la división de tareas es un proceso lógico, práctico y adecuado para muchos trabajos, tales como el relevamiento de la historia clínica por parte del médico, la toma y procesamiento de la muestra por el bioquímico de laboratorio, el procesamiento de datos por parte del bioinformático, la priorización de variantes por parte de bioquímicos/biólogos moleculares/médicos genetistas y finalmente la interpretación de las mismas y devolución al paciente por parte del médico; la eficiencia y las chances de éxito se maximizan cuando todos los profesionales trabajan en conjunto. Particularmente, esto se evidencia tanto en la etapa de postulación y selección de pacientes como en la priorización e interpretación de las variantes, la cual requiere integrar conocimientos asociados al experimento de NGS (cobertura, profundidad, calidad de la variante, entre otras), al impacto a nivel molecular de las mismas y, por supuesto, a su potencial relación con la clínica. Adicionalmente, queremos resaltar que la probabilidad de éxito se vio incrementada en aquellos casos en los que se evidenció la identificación de un diagnóstico presuntivo preciso, y el profundo conocimiento de los genes asociados a enfermedades autoinflamatorias por parte del equipo médico postulante.

A lo largo de estos años de experiencia, destacamos el compromiso del equipo de Novartis y Bitgenia, del *Advisory Board* y de cada médico postulante; pero por sobre todas las cosas el carácter federal del Programa, que brindó la posibilidad a pacientes de diversos puntos del país y de instituciones públicas o privadas de formar parte de un programa de medicina de precisión gratuito.

El Proyecto PAMPA tuvo un enorme impacto, cambiando el enfoque terapéutico de muchos de los pacientes incorporados, algunos de los cuales fueron mencionados en el Capítulo II y los boxes 1-6. El mismo sigue vigente hoy en día, culminando a fines de marzo con la quinta Edición y comenzando a planear la recepción de postulaciones de la sexta versión, demostrando su enorme éxito y difusión entre reumatólogos, inmunólogos y clínicos del país y sentando las bases para la implementación del Proyecto Cónedor en Colombia.

Discusión

La identificación de variantes patogénicas en la secuencia de ADN de pacientes con enfermedades hereditarias poco frecuentes es una de los principales aportes de la genética humana a la clínica. Las técnicas de secuenciación de próxima generación permiten la secuenciación de cientos de genes candidatos, genomas o exomas completos, conduciendo inevitablemente a la detección de un gran número de variantes, etapa en la que la evaluación de la patogenicidad de cada una de ellas resulta crucial para arribar a un diagnóstico molecular certero. Particularmente, la secuenciación exómica resulta un enfoque útil para identificar variantes que producen enfermedades monogénicas (Bamshad et al., 2011), ya que la mayoría de las patologías mendelianas son causadas por variantes presentes en la región codificante (exones) de un único gen.

La genómica, mediante la aplicación de tecnologías de NGS, ha cambiado el paradigma de la medicina en los siguientes aspectos fundamentales: la prevención, el diagnóstico, y el tratamiento de un amplio espectro de enfermedades, incluidas las condiciones genéticas, patologías crónicas y enfermedades infecciosas, y se prevé que en un futuro cercano su creciente aplicación en la clínica generará resultados favorables para lograr el diagnóstico molecular en un número mayor de pacientes y a un menor costo, potenciando la revolución de la medicina personalizada y de precisión.

Como el objetivo del diagnóstico clínico consiste en proporcionar a los médicos información fiable -en tiempos razonables y a un costo aceptable-, la secuenciación de próxima generación constituye -sin duda- uno de los avances tecnológicos más significativos de los últimos 30 años que permite la confirmación de diagnósticos clínicos presuntivos a través de: la caracterización molecular de los genes involucrados, la correlación fenotipo-genotipo, la identificación de los portadores de la mutación en la familia, el aporte de asesoramiento genético, y el establecimiento de terapias personalizadas o tratamientos particulares.

En cuanto a la prevención, el genoma del individuo puede contribuir, al menos parcialmente en el potencial -o riesgo- de desarrollar enfermedades complejas de alta prevalencia en el adulto, como por ejemplo las Enfermedades cardiovasculares, Diabetes tipo 2, Obesidad, Alzheimer, Parkinson o Cáncer, entre otras. El análisis de riesgo diferencial le permite al paciente tomar acciones preventivas, que incluyen cambios de hábitos poco saludables, abandono del tabaquismo, ajustes en la dieta y aumento de la frecuencia de ejercicio físico, etc.

En el diagnóstico, principalmente de las Enfermedades poco frecuentes, es donde las tecnologías NGS tuvieron un mayor impacto, al permitirnos buscar y encontrar de manera precisa la/s variante/s responsable/s del desarrollo de la enfermedad, entregando al paciente

y su familia un diagnóstico molecular certero en un porcentaje moderado de los casos. En algunos, el diagnóstico molecular le permite al médico encontrar e implementar un tratamiento adecuado (por ejemplo, en el caso de pacientes con enfermedades autoinflamatorias clásicas, donde se suelen usar anticuerpos monoclonales, esteroides o colchicina). Algunos tratamientos farmacológicos también pueden adaptarse a los pacientes en función de sus características genéticas, con el objetivo de mejorar la eficacia, optimizar su uso y reducir sus efectos secundarios. Este es el paradigma de la farmacogenómica, que estudia las bases moleculares de las diferencias individuales en la respuesta a los fármacos y pueden atribuirse a variantes, alteraciones epigenéticas o polimorfismos genéticos.

En este contexto, la bioinformática, disciplina que utiliza bases de datos y algoritmos para analizar proteínas, genes y genomas, cobra enorme relevancia y se vuelve fundamental. A medida que la exploramos como herramienta para el estudio de patologías hereditarias, constantemente nos enfrentamos con la complejidad inherente a todos los sistemas biológicos. Incluso cuando se descubre el gen que, debido a la presencia de variantes patogénicas, produce una enfermedad, nuestro desafío es intentar conectar el genotipo con el fenotipo, integrando y sintetizando información del contexto biológico digitalizada en una amplia variedad de bases de datos biológicas, con las que debemos trabajar e integrar informáticamente.

Resulta importante resaltar, para el tipo de estudios presentados en esta tesis, la necesidad y relevancia de las distintas etapas de procesamiento y análisis bioinformático que permitieron el arribo exitoso a un diagnóstico molecular en la mayoría de los casos. En tal sentido, la obtención de lecturas a partir del experimento de secuenciación, el mapeo contra el genoma de referencia, el llamado de variantes, la anotación funcional -que permite el análisis contextual de las variantes encontradas en relación con la información disponible en las bases de datos consultadas-, la predicción de patogenicidad por parte de algoritmos bioinformáticos e incluso el aprovechamiento de herramientas de clasificación automáticas, fueron elementos fundamentales en el análisis de los datos. Nuestra hipótesis de trabajo es que una de las cuestiones claves del éxito de la implementación de tecnologías NGS al diagnóstico, no radica únicamente en contar con una herramienta o plataforma sólida que permita la integración de datos genómicos y su análisis, sino también resaltar la importancia del desarrollo de un protocolo de priorización de variantes. De esta manera, el análisis de las variantes requirió de una serie de pasos que permitiera y facilitara obtener, a partir de todas las lecturas crudas obtenidas del secuenciador, unas pocas variantes candidatas interesantes para analizar con mayor profundidad. Esto fue posible gracias a la implementación de un esquema o flujo de priorización de variantes.

Con este trabajo, buscamos cubrir principalmente la necesidad de los pacientes y sus familias de llegar a un diagnóstico certero, que les permitiera establecer un nuevo punto de

partida y les facilitará el acceso a un tratamiento personalizado (en el caso de estar disponible). Los resultados obtenidos permitieron arribar a conclusiones interesantes y, en algunos casos, posibilitaron definir y delinear fenotipos novedosos.

El resultado exitoso y la descripción, tanto de casos individuales en profundidad, como de las generalidades derivadas de un análisis de múltiples casos para un grupo de enfermedades acotadas, permite la reflexión sobre varios de los aspectos importantes asociados al proceso y resultados obtenidos.

El primer caso, lo comprende un paciente que se presentó a la consulta con una alteración de la marcha de progresión lenta y dificultades de aprendizaje. Resultó ser el quinto de ocho hermanos, todos hijos de una pareja estable no consanguínea. Cinco integrantes presentaron ataxia cerebelosa, signos piramidales, nistagmo y disminución de la sensación vibratoria en miembros inferiores. Inicialmente, el neurólogo consideró una forma atípica de neurodegeneración asociada a pantotenato-quinasa (PKAN), una enfermedad caracterizada por la degeneración progresiva del sistema nervioso con acumulación de hierro en el cerebro, como el diagnóstico más probable, debido a la presencia del signo “ojo de tigre” en la resonancia magnética nuclear (en secuencias ponderadas en T2 consiste en un foco hiperintenso en la porción anteromedial del globo pálido, rodeado de un área hipointensa que cubre las porciones posteriores de este núcleo [271,272]). Sin embargo, la ausencia de acumulación de hierro en el cerebro y la detección de dos variantes heterocigotas en el gen *AFG3L2*, posibilitaron el cambio de enfoque, el arribo a un diagnóstico certero y la posterior publicación de un *Case Report* [206].

En segundo lugar, decidimos dedicar un capítulo particular al caso de una paciente con fiebre aguda y autolimitada, pioderma gangrenoso, úlceras recidivantes en regiones de pliegues, neumonía intersticial, alteración general del crecimiento, adenitis, leucocitosis episódica y vasculitis leucocitoclástica con cambios hipóxicos e isquémicos secundarios, con alta sospecha de Enfermedad autoinflamatoria monogénica. La detección de dos variantes en heterocigosis en el gen *ISG15*, en un cuadro heterocigota compuesto, posibilitó la detección y publicación de un fenotipo atípico [207]. En este contexto, la incorporación de esta paciente al Proyecto PAMPA y la implementación de tecnologías de secuenciación masiva permitió brindar un diagnóstico certero a la familia y la publicación de trabajos conjuntos con el instituto INSERM de París y la Universidad de Rockefeller en Estados Unidos [216].

El Proyecto PAMPA, en cada una de sus ediciones, comprendió el estudio genómico de pacientes con Enfermedades autoinflamatorias (EAI). Estas engloban un conjunto de condiciones heterogéneas caracterizadas por la presencia de inflamación sistémica crónica o recurrente, definida por la sobreactivación de mediadores y células del sistema inmune innato, como consecuencia de una disregulación del control del proceso inflamatorio. Muchos de los síndromes autoinflamatorios exhiben características inflamatorias sistémicas y/u orgánicas

específicas tales como fiebre periódica recurrente y episódica, serositis, artritis y/o inflamación cutánea, sobreproducción de IL-1 β y activación de células inmunes innatas, particularmente monocitos. La variabilidad clínica (heterogeneidad fenotípica) de las EAI, y el elevado número de genes asociados al desarrollo de las mismas convierte a este grupo de desórdenes en un caso ideal para su abordaje por técnicas de secuenciación masiva.

En todos los casos, resultó extremadamente importante contar con el apoyo de los médicos postulantes. Una adecuada valoración del paciente que deviniera en la entrega de una historia clínica completa y entendible, incluyendo una descripción detallada de los signos, síntomas y sospecha diagnóstica constituyó un punto de partida fundamental para el análisis de las miles de variantes obtenidas del experimento de secuenciación. Esto nos permitió adquirir una rápida comprensión inicial del caso y facilitó la selección de los principales genes a estudiar, con el consecuente impacto en el número de variantes a priorizar. Como resultado, particularmente en el Caso I, la importante contribución del médico a cargo maximizó las chances de encontrar variantes que pudieran explicar el fenotipo del paciente y de todos sus familiares afectados y posibilitó la ampliación de los fenotipos asociados a variantes bialélicas en el gen *AFG3L2*. Tanto en el Caso I como en el II se realizaron extensas búsquedas bibliográficas y comparaciones entre fenotipos previamente reportados en la literatura y los presentados por los pacientes. Por su parte, el enfoque de PAMPA con un panel predefinido acotado para un grupo específico de enfermedades permitió poner a prueba la potencia de un enfoque tipo panel que aumentó de manera significativa la eficiencia del análisis.

Desafíos en el diagnóstico molecular de enfermedades mendelianas por NGS

Los principales desafíos en el diagnóstico molecular de enfermedades monogénicas los constituyen:

- La **clasificación final de las variantes de significado incierto (VUS)**, ya sea como benignas o patogénicas. La mayoría de las variantes detectadas en un estudio genómico (ya sea de panel, exoma, o genoma) son variantes benignas, o VUS. Una variante suele ser clasificada como VUS cuando el conocimiento disponible sobre la misma no es suficiente para decidir si pudiera (o no) ser responsable de la patología. Este es un enorme obstáculo al momento de decidir su relevancia en la clínica. Si bien, la constante aplicación de NGS en esta área genera un aumento continuo en la cantidad de variantes sobre las que se conoce e informa un efecto deletéreo, y que por lo tanto pueden ser clasificadas como patogénicas, el número de variantes en la práctica es casi infinito, y en el corto o mediano plazo es imposible que todas hayan sido analizadas. Es por esta razón que se deben desarrollar nuevas y mejores

metodologías que faciliten el proceso de clasificación y estudio de variantes. Por un lado, existe un creciente desarrollo de métodos bioinformáticos para la valoración de variantes, que son cada vez más precisos y complejos, y han incorporado recientemente elementos de Inteligencia Artificial, mejorando de manera significativa su performance. Sin embargo, la mayoría de los métodos no son capaces de proveer información relevante para aquellas variantes que no han sido previamente reportadas en la literatura, o que no se encuentran dentro de un dominio proteico conservado o poseen muy baja frecuencia poblacional, que es donde más se los necesita. A esto se le suma la inexistencia de bases de datos públicas sistematizadas de ensayos funcionales de variantes en genes/proteínas, o de variantes que alteran el proceso de *splicing*.

- Otro punto relevante, es la correcta **priorización de genes**. Contar con una lista de genes candidatos en función del fenotipo patológico observado en el paciente resulta fundamental para acotar la cantidad de variantes y asegurarnos de que la búsqueda se realice en regiones codificantes de interés. Actualmente, si bien existen diversas herramientas (findZebra, phenomizer, VarElect, etc.) que permiten, dada una lista de signos y síntomas, obtener una lista de genes candidatos, las mismas son poco precisas, y suelen devolver largas listas de genes. Por lo tanto, el avance en el desarrollo de modelos cuantitativos que faciliten la priorización y valoración de los genes en función de la historia clínica ingresada resultaría de gran ayuda y ahorraría mucho del tiempo invertido en la búsqueda del panel de genes candidatos a estudiar en cada caso.
- **Identificación de variaciones en el número de copias (CNV, copy number variants) a partir de NGS.** Los CNV constituyen un tipo de variantes estructurales, como delecciones, duplicaciones, inserciones, inversiones y translocaciones, que implican un cambio en el número de copias respecto al genoma de referencia. Aunque son relativamente comunes en la estructura de nuestros genomas y contribuyen en gran medida a la variabilidad interindividual, también ha sido probada su implicación en un gran número de enfermedades como Autismo, Parkinson, Alzheimer, Enfermedades mitocondriales o Cáncer, entre otras. La patogenicidad de un CNV depende del tipo, tamaño y ubicación del mismo [273]. Dependiendo de si existiera un CNV de ganancia o pérdida de material genético, el mismo provocaría triplosensibilidad o haploinsuficiencia, respectivamente. El tamaño es importante, porque cuanto más grande sea el CNV, más genes se verán

afectados y el efecto causado será más deletéreo. La ubicación también es muy importante: los CNV intergénicos, es decir, los que se producen entre genes, no suelen ser patogénicos, mientras que los que abarcan regiones codificantes sí lo son. La técnica estándar para detectar este tipo de variantes es MLPA (*Multiplex Ligation-dependent Probe Amplification*) o la utilización de hibridación genómica completa por microarreglos (CGH microarray). Sin embargo, cuando tenemos datos de secuenciación NGS, algunos CNV también pueden ser detectados de forma fiable dependiendo de su localización y tamaño [274]. Para ello, se recurre a herramientas como DECoN [179], ONCOCNV [180] o VarScan2 [275] que permiten identificarlas de forma comparativa. Normalmente se utilizan como mínimo 20 muestras de una misma corrida, es decir, que las muestras utilizadas en el cálculo idealmente deben tener las mismas condiciones experimentales. Resultaría interesante incorporar el análisis de CNVs como un estudio complementario al de NGS, enfocado en los genes candidatos de interés, especialmente en aquellos casos en los que no se encontraron variantes de nucleótido único o delecciones/inserciones pequeñas y el reporte de resultados resultó Negativo. Esto se aplicó a ciertos casos en los cuales la sospecha diagnóstica era fuertemente orientativa de una enfermedad en particular - asociada a un grupo de genes reducidos- y el resultado del exoma resultó negativo. En todos esos casos, no se hallaron CNVs reportables.

Además de las complejidades y dificultades mencionadas, también resulta relevante discutir que el análisis de los datos procedentes de la **secuenciación exómica, particularmente el análisis de CNVs presenta limitaciones** que deben ser tenidas en cuenta, ya que pueden afectar a la precisión de los resultados [276,277]. Entre ellas podemos destacar:

- Solo permite definir desequilibrios de información genómica, mientras que los reordenamientos que no impliquen una pérdida o ganancia neta, como las inversiones o translocaciones balanceadas, pasarán desapercibidas.
- No se dispone de la secuencia de todo el genoma, por lo que se ve afectada la detección de alteraciones que incluyen regiones no capturadas (intrones y regiones intergénicas).
- La preparación de la biblioteca y el proceso de captura pueden dar lugar a regiones con distintas profundidades y cobertura de secuenciación.
- Las secuencias repetitivas dificultan el mapeo único de las lecturas procedentes de la secuenciación.

- La caracterización de CNVs sigue siendo complicada, por lo que es necesario conocer y tener presente sus limitaciones. Además, resulta clave llevar a cabo una correcta validación de aquellas variantes encontradas, especialmente en el entorno clínico. El desarrollo de bases de datos de acceso público, tanto en individuos controles como en pacientes, contribuirá en el futuro a dilucidar el verdadero significado de todas las CNVs analizadas.

Perspectivas a futuro

De cara al futuro, y en base a lo observado en las últimas décadas, esperamos que los avances tecnológicos den paso a métodos de secuenciación más rápidos y precisos. Se busca que las futuras plataformas de secuenciación requieran menos cantidad de ADN y reactivos, y sean portátiles para su uso en diagnósticos médicos, agrícolas, ecológicos y otras investigaciones de campo. Los avances en bioinformática, robótica, manipulación de líquidos y preparación de ácidos nucleicos contribuirán a facilitarlo.

Sin embargo, para la mayoría de las aplicaciones cotidianas, muchos investigadores siguen prefiriendo el alto rendimiento, la alta precisión y la fiabilidad de las técnicas de lectura corta bien establecidas, como las que ofrece Illumina. El bajo costo por base y la diversidad de configuraciones de preparación de bibliotecas hacen que la secuenciación de lectura corta sea ventajosa para una amplia gama de aplicaciones clínicas y de investigación.

El avance vertiginoso de las tecnologías de secuenciación masiva se dirige hacia el diagnóstico con especificidad cercana al 100% al mejorar los algoritmos bioinformáticos, en donde no sea necesaria la confirmación de variantes patogénicas mediante secuenciación por Sanger, y esto permita el estudio de cualquier trastorno molecular en un único proceso, incluyendo el análisis de variantes estructurales y secuencias repetitivas. Se espera que el estudio del genoma se aplique a la detección de variantes patogénicas en regiones no estudiadas por los métodos actuales y permita profundizar el conocimiento sobre la influencia de las regiones reguladoras, teniendo el potencial de transformarse en una prueba genética para el diagnóstico y seguimiento de pacientes. Adicionalmente, se esperan mejoras o avances en los algoritmos o bases de datos que aporten información de relevancia a la hora de clasificar las variantes VUS y permitan encontrar una relación fenotipo-genotipo clara.

De la misma manera, la evolución de bases de datos poblacionales (que adicionen información de otras etnias/poblaciones), el avance en el estudio de mecanismos moleculares, funcionales y elementos regulatorios involucrados en enfermedades y el desarrollo del transcriptoma y metiloma, aportarán a la medicina personalizada desde el diagnóstico hasta el tratamiento, introduciendo nuevas oportunidades y retos. No obstante, en función de los

constantes desarrollos e innovaciones de las tecnologías NGS, es inminente la necesidad de modificar la formación de los médicos, educar a los pacientes y mejorar la comunicación entre asesores genéticos, investigadores, científicos clínicos, informáticos y médicos. Asimismo, resulta indispensable continuar profundizando y avanzando en la interpretación de las variantes, el almacenamiento de los datos bioinformáticos y los procedimientos de consentimiento informado para las pruebas genéticas. En este contexto, es necesaria la formación de equipos transdisciplinarios que puedan integrar el conocimiento de distintos campos, para el eficiente desarrollo de un proyecto que incluya tecnologías de nueva generación en la práctica clínica. Esto tendrá el potencial de implementar colaboraciones que acerquen a profesionales (médicos-científicos) de las distintas disciplinas y ámbitos profesionales (hospital-instituto/universidad) a fin de generar un lenguaje común que permita el flujo eficiente de conocimiento.

Anexo 1

Publicaciones en revistas especializadas y presentaciones a congresos relacionados con el presente trabajo de tesis.

Congresos

- XVII Congreso Latinoamericano de Genética, ALAG. Póster: MUTACIONES NOVELES EN GEN *ISG15* DEFINIENDO EL ESPECTRO FENOTÍPICO DE INMUNODEFICIENCIA ASOCIADA A MICOBACTERIAS EN HUMANOS". Mendoza-Argentina.

Publicaciones

- CUATRO AÑOS DE PAMPA: DIAGNÓSTICO MOLECULAR DE ENFERMEDADES AUTOINFLAMATORIAS EN ARGENTINA, publicado en la revista Química Viva en Agosto 2022.
- Keitelman IA, Shiromizu CM, Zgajnar NR, Danielián S, Jancic CC, Martí MA, Fuentes F, Yancoski J, Vera Aguilar D, Rosso DA, Goris V, Buda G, Katsicas MM, Galigniana MD, Galletti JG, Sabbione F, Trevani AS. The interplay between serine proteases and caspase-1 regulates the autophagy-mediated secretion of Interleukin-1 beta in human neutrophils. *Front Immunol.* 2022 Aug 25;13:832306. doi: 10.3389/fimmu.2022.832306. PMID: 36091026; PMCID: PMC9458071.
- Buda, G., Valdez, R.M., Biagioli, G. et al. Inflammatory cutaneous lesions and pulmonary manifestations in a new patient with autosomal recessive ISG15 deficiency case report. *Allergy Asthma Clin Immunol* 16, 77 (2020). <https://doi.org/10.1186/s13223-020-00473-7>
- Martin-Fernandez M, Bravo García-Morato M, Gruber C, Murias Loza S, Malik M NH, Alsohime F, Alakeel A, Valdez R, Buta S, Buda G, Marti MA, Larralde M, Boisson B, Feito Rodriguez M, Qiu X, Chrabieh M, Al Ayed M, Al Muhsen S, Desai JV, Ferre EMN, Rosenzweig SD, Amador-Borrero B, Bravo-Gallego LY, Olmer R, Merkert S, Bret M, Sood AK, Al-Rabiah A, Temsah MH, Halwani R, Hernandez M, Pessler F, Casanova JL, Bustamante J, Lionakis MS, Bogunovic D. Systemic Type I IFN Inflammation in Human ISG15 Deficiency Leads to Necrotizing Skin Lesions. *Cell Rep.* 2020 May 12;31(6):107633. doi: 10.1016/j.celrep.2020.107633. PMID: 32402279; PMCID: PMC7331931.
- Calandra CR, Buda G, Vishnopska SA, Oliveri J, Olivieri FA, Pérez Millán MI, Biagioli G, Miquelini LA, Pellec AL, Marti MA. Spastic ataxia with eye-of-the-tiger-like sign in 4 siblings due to novel compound heterozygous AFG3L2 mutation. *Parkinsonism Relat Disord.* 2020 Apr;73:52-54. doi: 10.1016/j.parkreldis.2020.03.020. Epub 2020 Mar 24. PMID: 32248051.

Bibliografía

1. Amberger JS, Bocchini CA, Schietecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 2015;43: D789–98.
2. Lee K. The World Health Organization (WHO). Routledge; 2008.
3. Prakash V, Moore M, Yáñez-Muñoz RJ. Current Progress in Therapeutic Gene Editing for Monogenic Diseases. *Mol Ther.* 2016;24: 465–474.
4. FADEPOF. [cited 20 Feb 2022]. Available: <https://fadepof.org.ar/epof.php>
5. Gonzaga-Jauregui C, Lupski JR. Genomics of Rare Diseases: Understanding Disease Genetics Using Genomic Approaches. Academic Press; 2021.
6. Genetics, Disease Prevention and Treatment FAQ. In: Genome.gov [Internet]. [cited 20 Feb 2022]. Available: <https://www.genome.gov/FAQ/Genetics-Disease-Prevention-and-Treatment>
7. Bitgenia Proyecto PAMPA Programa Argentino de Medicina de Precisión en Enfermedades Autoinflamatorias. [cited 20 Feb 2022]. Available: <https://www.bitgenia.com/pampa/>
8. Manolio TA. Incorporating Whole-Genome Sequencing Into Primary Care: Falling Barriers and Next Steps. *Annals of Internal Medicine.* 2017. p. 204. doi:10.7326/m17-1518
9. Meng Q, Yu J. Next Generation DNA Sequencing Technologies and Applications. *Next Generation Sequencing and Whole Genome Selection in Aquaculture.* 2010. pp. 35–56. doi:10.1002/9780470958964.ch3
10. Smedley D, Robinson PN. Phenotype-driven strategies for exome prioritization of human Mendelian disease genes. *Genome Medicine.* 2015. doi:10.1186/s13073-015-0199-2
11. Diagnostic Clinical Genome and Exome Sequencing. *New England Journal of Medicine.* 2014. pp. 1169–1170. doi:10.1056/nejmc1408914
12. Zhang X, Min LI, Zhang X-J. Exome sequencing and its application. *Hereditas (Beijing).* 2011. pp. 847–856. doi:10.3724/sp.j.1005.2011.00847
13. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* 2008;9: 387–402.
14. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010;11: 31–46.
15. Thompson JF, Milos PM. The properties and applications of single-molecule DNA sequencing. *Genome Biol.* 2011;12: 217.
16. Voelkerding KV, Dames S, Durtschi JD. Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. *J Mol Diagn.* 2010;12: 539–551.
17. Frank M, Prenzler A, Eils R, Graf von der Schulenburg J-M. Genome sequencing: a systematic review of health economic evidence. *Health Econ Rev.* 2013;3: 29.
18. Willyard C. New human gene tally reignites debate. *Nature.* 2018;558: 354–355.
19. Dattani MT. The Candidate Gene Approach to the Diagnosis of Monogenic Disorders. *Hormone Research in Paediatrics.* 2009. pp. 14–21. doi:10.1159/000192431

20. van Nimwegen KJM, van Soest RA, Veltman JA, Nelen MR, van der Wilt GJ, Vissers LELM, et al. Is the \$1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. *Clin Chem.* 2016;62: 1458–1464.
21. Ku C-S, Cooper DN, Polychronakos C, Naidoo N, Wu M, Soong R. Exome sequencing: dual role as a discovery and diagnostic tool. *Ann Neurol.* 2012;71: 5–14.
22. Stranneheim H, Wedell A. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. *Journal of Internal Medicine.* 2016. pp. 3–15. doi:10.1111/jim.12399
23. Hoppman-Chaney N, Peterson LM, Klee EW, Middha S, Courteau LK, Ferber MJ. Evaluation of oligonucleotide sequence capture arrays and comparison of next-generation sequencing platforms for use in molecular diagnostics. *Clin Chem.* 2010;56: 1297–1306.
24. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem.* 2009;55: 641–658.
25. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med.* 2013;369: 1502–1511.
26. Precone V, Del Monaco V, Esposito MV, De Palma FDE, Ruocco A, Salvatore F, et al. Cracking the Code of Human Diseases Using Next-Generation Sequencing: Applications, Challenges, and Perspectives. *Biomed Res Int.* 2015;2015: 161648.
27. Xiao T, Zhou W. The third generation sequencing: the advanced approach to genetic diseases. *Transl Pediatr.* 2020;9: 163–173.
28. Zhang R. Oxford Nanopore sequencing and library construction v1. *protocols.io.* doi:10.17504/protocols.io.btcwnixe
29. Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE. Landscape of Next-Generation Sequencing Technologies. *Analytical Chemistry.* 2011. pp. 4327–4341. doi:10.1021/ac2010857
30. Ahn S-M. Phenotype-Genotype Association Study: From GWAS to NGS. *Endocrinology and Metabolism.* 2011. p. 187. doi:10.3803/enm.2011.26.3.187
31. Sablok G, Kumar S, Ueno S, Kuo J, Varotto C. Advances in the Understanding of Biological Sciences Using Next Generation Sequencing (NGS) Approaches. Springer; 2015.
32. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol.* 2008;26: 1135–1145.
33. van Dijk EL, Auger H, Jaszczyzyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet.* 2014;30: 418–426.
34. Wong L-JC. Next Generation Sequencing Based Clinical Molecular Diagnosis of Human Genetic Disorders. Springer; 2017.
35. Meur G, Simon A, Harun N, Virally M, Dechaume A, Bonnefond A, et al. Insulin gene mutations resulting in early-onset diabetes: marked differences in clinical presentation, metabolic status, and pathogenic effect through endoplasmic reticulum retention. *Diabetes.* 2010;59: 653–661.
36. Alexander Valencia C, Ali Pervaiz M, Husami A, Qian Y, Zhang K. Next Generation Sequencing Technologies in Medical Genetics. Springer Science & Business Media; 2013.
37. Pellini M, Miller V, Palmer G. L2.2 Utilizing Next Generation Sequencing (NGS) in Clinical Oncology. *Annals of Oncology.* 2012. p. v6. doi:10.1016/s0923-7534(20)31263-1
38. Vuyisich M. Next generation sequencing (NGS)technologies and applications. 2012. doi:10.2172/1050518
39. Yubero D, Artuch R. NGS for Metabolic Disease Diagnosis. *EJIFCC.* 2018;29: 227–229.

40. Guan Y-F, Li G-R, Wang R-J, Yi Y-T, Yang L, Jiang D, et al. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chin J Cancer*. 2012;31: 463–470.
41. Gagan J, Van Allen EM. Next-generation sequencing to guide cancer therapy. *Genome Med*. 2015;7: 80.
42. Roychowdhury S, Van Allen EM. *Precision Cancer Medicine: Challenges and Opportunities*. Springer Nature; 2020.
43. Blue GM, Winlaw DS. Next Generation Sequencing in Congenital Heart Disease: Gene Discovery and Clinical Application. *Journal of Next Generation Sequencing & Applications*. 2015. doi:10.4172/2469-9853.1000113
44. Odibo AO, Krantz DA. Prenatal Screening and Diagnosis, An Issue of the Clinics in Laboratory Medicine, E-Book. Elsevier Health Sciences; 2016.
45. Muzzey D. Understanding the Basics of NGS in the Context of NIPT. *Noninvasive Prenatal Testing (NIPT)*. 2018. pp. 7–24. doi:10.1016/b978-0-12-814189-2.00002-5
46. Pös O, Budis J, Kubiritsova Z, Kucharik M, Duris F, Radvanszky J, et al. Identification of Structural Variation from NGS-Based Non-Invasive Prenatal Testing. *International Journal of Molecular Sciences*. 2019. p. 4403. doi:10.3390/ijms20184403
47. Prenatal DNA Testing. *Prenatal Medicine*. 2016. pp. 299–308. doi:10.3109/9781420016369-20
48. Ma L, Jakobiec FA, Dryja TP. A Review of Next-Generation Sequencing (NGS): Applications to the Diagnosis of Ocular Infectious Diseases. *Seminars in Ophthalmology*. 2019. pp. 223–231. doi:10.1080/08820538.2019.1620800
49. Ampofo K, Pavia A, Blaschke AJ, Schlaberg R. Detection of Respiratory Pathogens in Parapneumonic Effusions by Hypothesis-free, Next-Generation Sequencing (NGS). *Open Forum Infectious Diseases*. 2017. pp. S17–S17. doi:10.1093/ofid/ofx162.044
50. Harb WJ. Gastrointestinal Polyposis Syndromes. *Inherited Cancer Syndromes*. 2011. pp. 105–125. doi:10.1007/978-1-4419-6821-0_5
51. De Santis D, Dinauer D, Duke J, Erlich HA, Holcomb CL, Lind C, et al. 16(th) IHIW : review of HLA typing by NGS. *Int J Immunogenet*. 2013;40: 72–76.
52. Mack SJ, Milius RP, Gifford BD, Sauter J, Hofmann J, Osoegawa K, et al. Minimum information for reporting next generation sequence genotyping (MIRING): Guidelines for reporting HLA and KIR genotyping via next generation sequencing. *Hum Immunol*. 2015;76: 954–962.
53. Zhang X. Exome sequencing greatly expedites the progressive research of Mendelian diseases. *Front Med*. 2014;8: 42–57.
54. Sawyer SL, Hartley T, Dyment DA, Beaulieu CL, Schwartzenruber 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: 275–284.
55. Grody WW, Nakamura RM, Kiechle FL, Strom C. *Molecular Diagnostics: Techniques and Applications for the Clinical Laboratory*. Academic Press; 2009.
56. Genomic and Personalized Medicine. Academic Press; 2012.
57. Kanzi AM, San JE, Chimukangara B, Wilkinson E, Fish M, Ramsuran V, et al. Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. *Front Genet*. 2020;11: 544162.
58. Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics. *J Pathol Inform*. 2012;3: 40.

59. Abe H. Expectations for innovation of personalized medicine. *Personalized Medicine Universe*. 2013. p. 1. doi:10.1016/j.pmu.2013.05.003
60. Wong L-JC. *Next Generation Sequencing: Translation to Clinical Diagnostics*. Springer Science & Business Media; 2013.
61. Glass DA, Nuara AAA. Whole Exome Sequencing. *Dermatitis*. 2013. pp. 33–34. doi:10.1097/der.0b013e3182759e43
62. Sabater-Molina M, García-Molina E, Tovar I, Ruiz-Espejo F, Gimeno JR, Valdés M. Cost-effectiveness of genetic studies in inherited heart diseases. *Cardiogenetics*. 2013. doi:10.4081/cardigenetics.2013.e5
63. A global reference for human genetic variation. *Nature*. 2015;526: 68–74.
64. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012;337: 64–69.
65. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536: 285–291.
66. The Genome Aggregation Database (gnomAD). [cited 20 Feb 2022]. Available: <https://www.nature.com/immersive/d42859-020-00002-x/index.html>
67. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 2017;45: D840–D845.
68. Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet*. 2008;4: e1000083.
69. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet*. 2012;13: 135–145.
70. Schoech AP, Jordan DM, Loh P-R, Gazal S, O'Connor LJ, Balick DJ, et al. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat Commun*. 2019;10: 790.
71. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461: 747–753.
72. Szumilas M. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry*. 2010;19: 227–229.
73. Park J-H, Gail MH, Weinberg CR, Carroll RJ, Chung CC, Wang Z, et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proc Natl Acad Sci U S A*. 2011;108: 18026–18031.
74. Boehm CD, Kazazian HH Jr. The molecular basis of genetic disease. *Curr Opin Biotechnol*. 1990;1: 180–187.
75. Strasser BJ, Fantini B. Molecular diseases and diseased molecules: ontological and epistemological dimensions. *Hist Philos Life Sci*. 1998;20: 189–214.
76. Reeb J, Hecht M, Mahlich Y, Bromberg Y, Rost B. Predicted Molecular Effects of Sequence Variants Link to System Level of Disease. *PLoS Comput Biol*. 2016;12: e1005047.
77. 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. *Genet Med*. 2015;17: 405–424.

78. Dalton KP, Rehm HL, Wright MW, Mandell ME, Krysiak K, Babb L, et al. Accessing clinical-grade genomic classification data through the ClinGen Data Platform. *Pac Symp Biocomput.* 2023;28: 531–535.
79. Altarescu G. Clinically Relevant Examples of Genotype–Phenotype Correlation. *Fabry Disease.* 2010. pp. 105–109. doi:10.1007/978-90-481-9033-1_5
80. Zlotogora J. Penetrance and expressivity in the molecular age. *Genetics in Medicine.* 2003. pp. 347–352. doi:10.1097/01.gim.0000086478.87623.69
81. McClellan J, King M-C. Genetic Heterogeneity in Human Disease. *Cell.* 2010. pp. 210–217. doi:10.1016/j.cell.2010.03.032
82. Lu JT, Campeau PM, Lee BH. Genotype-phenotype correlation. *Obstet Gynecol Surv.* 2014;69: 728–730.
83. OMIM - NCBI. [cited 18 Aug 2021]. Available: <https://www.ncbi.nlm.nih.gov/omim>
84. Papa R, Doglio M, Lachmann HJ, Ozen S, Frenkel J, Simon A, et al. A web-based collection of genotype-phenotype associations in hereditary recurrent fevers from the Eurofever registry. *Orphanet J Rare Dis.* 2017;12: 167.
85. Tisdale A, Cutillo CM, Nathan R, Russo P, Laraway B, Haendel M, et al. The IDeaS initiative: pilot study to assess the impact of rare diseases on patients and healthcare systems. *Orphanet J Rare Dis.* 2021;16: 429.
86. Haendel M, Vasilevsky N, Unni D, Bologa C, Harris N, Rehm H, et al. How many rare diseases are there? *Nat Rev Drug Discov.* 2020;19: 77–78.
87. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. *Nature Reviews Genetics.* 2018. pp. 253–268. doi:10.1038/nrg.2017.116
88. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet.* 2020;28: 165–173.
89. Paz JFL. Aproximación a las dimensiones y factores asociados al contexto de enfermedades raras o poco frecuentes. *Revista Iberoamericana de Bioética.* 2021. pp. 01–13. doi:10.14422/rib.i15.y2021.006
90. Bousfiha A, Jeddane L, Picard C, Ailal F, Bobby Gaspar H, Al-Herz W, et al. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. *J Clin Immunol.* 2018;38: 129–143.
91. Ciccarelli F, Martinis M, Ginaldi L. An Update on Autoinflammatory Diseases. *Current Medicinal Chemistry.* 2013. pp. 261–269. doi:10.2174/09298673113206660303
92. Meiorin S, Espada G, Rosè C. Enfermedades autoinflamatorias en pediatría. *Arch Argent Pediatr* 2013;111(3):237-243 <https://dx.doi.org/10.5546/aap.2013.237>
93. Llobet Agulló MP, Moure González JD. Diagnóstico y manejo de las enfermedades autoinflamatorias en Pediatría. *Protoc diagn ter pediatr.* 2019; 2:453-69.
94. Kumar V, Abbas AK, Aster JC. Robbins and Cotran Pathologic Basis of Disease. W B Saunders Company; 2014.
95. Janeway CA Jr. Pillars article: approaching the asymptote? Evolution and revolution in immunology. *Cold spring harb symp quant biol.* 1989. 54: 1-13. *J Immunol.* 2013;191: 4475–4487.
96. Tang T, Gong T, Jiang W, Zhou R. GPCRs in NLRP3 Inflammasome Activation, Regulation, and Therapeutics. *Trends Pharmacol Sci.* 2018;39: 798–811.

97. Masumoto J, Zhou W, Morikawa S, Hosokawa S, Taguchi H, Yamamoto T, et al. Molecular biology of autoinflammatory diseases. *Inflammation and Regeneration*. 2021; doi:10.1186/s41232-021-00181-8
98. McDermott A, Jacks J, Kessler M, Emanuel PD, Gao L. Proteasome-associated autoinflammatory syndromes: advances in pathogeneses, clinical presentations, diagnosis, and management. *Int J Dermatol*. 2015;54: 121–129.
99. McDermott MF, Aksentijevich I, Galon J, McDermott EM, Ogunkolade BW, Centola M, et al. Germline mutations in the extracellular domains of the 55 kDa TNF receptor, TNFR1, define a family of dominantly inherited autoinflammatory syndromes. *Cell*. 1999;97: 133–144.
100. Oldham J, Lachmann HJ. The systemic autoinflammatory disorders for dermatologists. Part 1: overview. *Clin Exp Dermatol*. 2020;45: 962–966.
101. Kovarova M, Hesker PR, Jania L, Nguyen M, Snouwaert JN, Xiang Z, et al. NLRP1-dependent pyroptosis leads to acute lung injury and morbidity in mice. *J Immunol*. 2012;189: 2006–2016.
102. Zhao Y, Yang J, Shi J, Gong Y-N, Lu Q, Xu H, et al. The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature*. 2011;477: 596–600.
103. Hornung V, Ablasser A, Charrel-Dennis M, Bauernfeind F, Horvath G, Caffrey DR, et al. AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature*. 2009;458: 514–518.
104. Halle A, Hornung V, Petzold GC, Stewart CR, Monks BG, Reinheckel T, et al. The NALP3 inflammasome is involved in the innate immune response to amyloid-beta. *Nat Immunol*. 2008;9: 857–865.
105. Jiang Y, Wang M, Huang K, Zhang Z, Shao N, Zhang Y, et al. Oxidized low-density lipoprotein induces secretion of interleukin-1 β by macrophages via reactive oxygen species-dependent NLRP3 inflammasome activation. *Biochem Biophys Res Commun*. 2012;425: 121–126.
106. Martinon F, Burns K, Tschopp J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. *Mol Cell*. 2002;10: 417–426.
107. Hua K-F, Chernikov O, Chen A, Moon J-S. NLRP3 Inflammasome: Regulatory Mechanisms, Role in Health and Disease and Therapeutic Potential. *Frontiers Media SA*; 2021.
108. Bauernfeind FG, Horvath G, Stutz A, Alnemri ES, MacDonald K, Speert D, et al. Cutting edge: NF-kappaB activating pattern recognition and cytokine receptors license NLRP3 inflammasome activation by regulating NLRP3 expression. *J Immunol*. 2009;183: 787–791.
109. Hannum CH, Wilcox CJ, Arend WP, Joslin FG, Dripps DJ, Heimdal PL, et al. Interleukin-1 receptor antagonist activity of a human interleukin-1 inhibitor. *Nature*. 1990;343: 336–340.
110. Ozkurede VU, Franchi L. Immunology in clinic review series; focus on autoinflammatory diseases: role of inflammasomes in autoinflammatory syndromes. *Clin Exp Immunol*. 2012;167: 382–390.
111. Doherty TA, Brydges SD, Hoffman HM. Autoinflammation: translating mechanism to therapy. *J Leukoc Biol*. 2011;90: 37–47.
112. Bodmer MW, Henderson B. *Therapeutic Modulation of Cytokines*. CRC Press; 1996.
113. Henderson C, Goldbach-Mansky R. Monogenic IL-1 mediated autoinflammatory and immunodeficiency syndromes: finding the right balance in response to danger signals. *Clin Immunol*. 2010;135: 210–222.
114. Ozen S, Bilginer Y, Aktay Ayaz N, Calguneri M. Anti-interleukin 1 treatment for patients with familial Mediterranean fever resistant to colchicine. *J Rheumatol*. 2011;38: 516–518.

115. Chakraborty A, Tannenbaum S, Rordorf C, Lowe PJ, Floch D, Gram H, et al. Pharmacokinetic and pharmacodynamic properties of canakinumab, a human anti-interleukin-1 β monoclonal antibody. *Clin Pharmacokinet*. 2012;51: e1–18.
116. Kapur S, Bonk ME. Rilonacept (arcalyst), an interleukin-1 trap for the treatment of cryopyrin-associated periodic syndromes. *P T*. 2009;34: 138–141.
117. Moreau A, Le Vée M, Jouan E, Denizot C, Parmentier Y, Fardel O. Effect of Gevokizumab on Interleukin-1 β -Mediated Cytochrome P450 3A4 and Drug Transporter Repression in Cultured Human Hepatocytes. *Eur J Drug Metab Pharmacokinet*. 2017;42: 871–878.
118. Terkeltaub R, Sundy JS, Schumacher HR, Murphy F, Bookbinder S, Biedermann S, et al. The interleukin 1 inhibitor rilonacept in treatment of chronic gouty arthritis: results of a placebo-controlled, monosequence crossover, non-randomised, single-blind pilot study. *Ann Rheum Dis*. 2009;68: 1613–1617.
119. Schlesinger N, De Meulemeester M, Pikhla A, Yücel AE, Richard D, Murphy V, et al. Canakinumab relieves symptoms of acute flares and improves health-related quality of life in patients with difficult-to-treat Gouty Arthritis by suppressing inflammation: results of a randomized, dose-ranging study. *Arthritis Res Ther*. 2011;13: R53.
120. Kerensky TA, Gottlieb AB, Yaniv S, Au S-C. Etanercept: efficacy and safety for approved indications. *Expert Opin Drug Saf*. 2012;11: 121–139.
121. Danese S. Mechanisms of action of infliximab in inflammatory bowel disease: an anti-inflammatory multitasker. *Dig Liver Dis*. 2008;40 Suppl 2: S225–8.
122. Gelfand J. Faculty Opinions recommendation of Clinical response to adalimumab: relationship to anti-adalimumab antibodies and serum adalimumab concentrations in rheumatoid arthritis. Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature. 2007. doi:10.3410/f.1065979.518943
123. Ogata A, Tanaka T. Tocilizumab for the treatment of rheumatoid arthritis and other systemic autoimmune diseases: current perspectives and future directions. *Int J Rheumatol*. 2012;2012: 946048.
124. Molad Y. Update on colchicine and its mechanism of action. *Curr Rheumatol Rep*. 2002;4: 252–256.
125. Gurcel L, Abrami L, Girardin S, Tschopp J, van der Goot FG. Caspase-1 activation of lipid metabolic pathways in response to bacterial pore-forming toxins promotes cell survival. *Cell*. 2006;126: 1135–1145.
126. Yurdakul S, Mat C, Tüzün Y, Ozyazgan Y, Hamuryudan V, Uysal O, et al. A double-blind trial of colchicine in Behcet's syndrome. *Arthritis Rheum*. 2001;44: 2686–2692.
127. Sari I, Yuksel A, Kozaci D, Selcuk S, Gokce G, Yildiz Y, et al. The effect of regular colchicine treatment on biomarkers related with vascular injury in newly diagnosed patients with familial Mediterranean fever. *Inflammation*. 2012;35: 1191–1197.
128. Franchi L, Kanneganti T-D, Dubyak GR, Núñez G. Differential requirement of P2X7 receptor and intracellular K $^{+}$ for caspase-1 activation induced by intracellular and extracellular bacteria. *J Biol Chem*. 2007;282: 18810–18818.
129. Leung YY, Yao Hui LL, Kraus VB. Colchicine--Update on mechanisms of action and therapeutic uses. *Semin Arthritis Rheum*. 2015;45: 341–350.
130. Hirose W, Harigai M, Amano K, Hidaka T, Itoh K, Aoki K, et al. Real-world effectiveness and safety of tofacitinib and abatacept in patients with rheumatoid arthritis. *Rheumatol Adv Pract*. 2022;6: rkac090.

131. Lai Y, Li B, Huang J, Du J, Yue M, Shen X, et al. Different pregnancy outcomes in patients with systemic lupus erythematosus treated with belimumab. *Lupus*. 2022; 9612033221141805.
132. Lake-Bakaar G. *Rituximab: Pharmacology, Clinical Uses and Health Effects*. Nova Science Pub Incorporated; 2012.
133. Inflamasome and cytokine blocking strategies in autoinflammatory disorders. *Clin Immunol*. 2013;147: 242–275.
134. Genetic Alliance, New York-Mid-Atlantic Consortium for Genetic and Newborn Screening Services. *Understanding Genetics: A New York, Mid-Atlantic Guide for Patients and Health Professionals*. Lulu.com; 2009.
135. Schully SD, Lam TK, Dotson WD, Chang CQ, Aronson N, Birkeland ML, et al. Evidence synthesis and guideline development in genomic medicine: current status and future prospects. *Genet Med*. 2015;17: 63–67.
136. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet*. 2013;14: 295–300.
137. Francey LJ, Conlin LK, Kadesch HE, Clark D, Berrodin D, Sun Y, et al. Genome-wide SNP genotyping identifies the Stereocilin (STRC) gene as a major contributor to pediatric bilateral sensorineural hearing impairment. *Am J Med Genet A*. 2012;158A: 298–308.
138. Duzkale H, Shen J, McLaughlin H, Alfares A, Kelly MA, Pugh TJ, et al. A systematic approach to assessing the clinical significance of genetic variants. *Clin Genet*. 2013;84: 453–463.
139. ACMG Board of Directors. Points to consider in the clinical application of genomic sequencing. *Genet Med*. 2012;14: 759–761.
140. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*. 2011. pp. 745–755. doi:10.1038/nrg3031
141. 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: 485.
142. Clark MM, Stark Z, Farnaes L, Tan TY, White SM, Dimmock D, et al. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *npj Genomic Medicine*. 2018;3: 1–10.
143. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. *Cell*. 2013;155: 27–38.
144. Standage DS. Scalable and reproducible genome analysis in the age of next-generation genome sequencing. doi:10.31274/etd-180810-4829
145. Di Resta C, Galbiati S, Carrera P, Ferrari M. Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities. *EJIFCC*. 2018;29: 4–14.
146. Pediátrica REE. *Revista Española Endocrinología Pediátrica - Nuevas tecnologías aplicadas en la detección de alteraciones genéticas*. [cited 1 Mar 2022]. Available: https://www.endocrinologiapediatrica.org/modules.php?name=articulos&idarticulo=462&idlangart=ES&preproduccion=&in_window=1
147. Challenges of diagnostic genomics in Latin America. *Curr Opin Genet Dev*. 2021;66: 101–109.
148. Declaración de Helsinki de la AMM - Principios éticos para las investigaciones médicas en seres humanos. [cited 9 Apr 2023]. Available: http://www.anmat.gov.ar/comunicados/HELSINSKI_2013.pdf

149. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20: 1297–1303.
150. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* . 2012;6: 80–92.
151. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38: e164.
152. Klymkowsky M. Faculty Opinions recommendation of Analysis of protein-coding genetic variation in 60,706 humans. Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature. 2019. doi:10.3410/f.726646180.793565293
153. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016;44: D862–8.
154. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;Chapter 7: Unit7.20.
155. Advances in Protein Chemistry and Structural Biology. Academic Press; 2014.
156. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods.* 2010;7: 575–576.
157. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011;27: 863–864.
158. Robinson P, Hansen P. SAM/BAM Format. Computational Exome and Genome Analysis. 2017. pp. 111–128. doi:10.1201/9781315154770-9
159. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25: 1754–1760.
160. Lubin IM, Aziz N, Babb LJ, Ballinger D, Bisht H, Church DM, et al. Principles and Recommendations for Standardizing the Use of the Next-Generation Sequencing Variant File in Clinical Settings. *J Mol Diagn.* 2017;19: 417–426.
161. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27: 2156–2158.
162. ROL DE NGS EN PATOLOGIA MIELOIDE. Available: http://www.sah.org.ar/revista/numeros/vol23/n2_educacional/2-ROL%20DE%20NGS%20EN%20PATOLOGIA%20MIELOIDE-jauk-n23-ext.pdf. [cited 20 Mar 2022].
163. 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.* 2018;46: D1062–D1067.
164. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *The American Journal of Human Genetics.* 2017. pp. 267–280. doi:10.1016/j.ajhg.2017.01.004
165. Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research.* 2001. pp. 308–311. doi:10.1093/nar/29.1.308
166. Database tools in genetic diseases research. *Genomics.* 2013;101: 75–85.
167. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. *Nucleic Acids Res.* 2018;46: D754–D761.

168. "The European Bioinformatics Institute. EMBL-EBI" www.ebi.ac.uk/. Accessed 15 Dec. 2021.
169. Wellcome Sanger Institute. [cited 20 Mar 2022]. Available: <https://www.sanger.ac.uk/>
170. Consortium TU, The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*. 2019. pp. D506–D515. doi:10.1093/nar/gky1049
171. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44: D733–45.
172. Sarrauste de Menthière C, Terrière S, Pugnère D, Ruiz M, Demaile J, Touitou I. INFEVERS: the Registry for FMF and hereditary inflammatory disorders mutations. *Nucleic Acids Res.* 2003;31: 282–285.
173. Bustaffa M, Koné-Paut I, Ozen S, Amaryan G, Papadopoulou-Alataki E, Gallizzi R, et al. The impact of the Eurofever criteria and the new InFevers MEFV classification in real life: Results from a large international FMF cohort. *Semin Arthritis Rheum.* 2022;52: 151957.
174. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. *Bioinformatics*. 2019;35: 1978–1980.
175. Allot A, Peng Y, Wei C-H, Lee K, Phan L, Lu Z. LitVar: a semantic search engine for linking genomic variant data in PubMed and PMC. *Nucleic Acids Research*. 2018. pp. W530–W536. doi:10.1093/nar/gky355
176. Milhavet F, Cuisset L, Hoffman HM, Slim R, El-Shanti H, Aksentijevich I, et al. The infevers autoinflammatory mutation online registry: update with new genes and functions. *Hum Mutat.* 2008;29: 803–808.
177. Fokkema IFAC, Fokkema IFA, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, et al. LOVD v.2.0: the next generation in gene variant databases. *Human Mutation*. 2011. pp. 557–563. doi:10.1002/humu.21438
178. Moreno-Cabrera JM, Del Valle J, Castellanos E, Feliubadaló L, Pineda M, Brunet J, et al. Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *Eur J Hum Genet.* 2020;28: 1645–1655.
179. Fowler A, Mahamdallie S, Ruark E, Seal S, Ramsay E, Clarke M, et al. Accurate clinical detection of exon copy number variants in a targeted NGS panel using DECoN. *Wellcome Open Res.* 2016;1: 20.
180. Boeva V, Popova T, Lienard M, Toffoli S, Kamal M, Le Tourneau C, et al. Multi-factor data normalization enables the detection of copy number aberrations in amplicon sequencing data. *Bioinformatics*. 2014;30: 3443–3450.
181. Beck TF, Mullikin JC, NISC Comparative Sequencing Program, Biesecker LG. Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clin Chem.* 2016;62: 647–654.
182. Mu W, Lu H-M, Chen J, Li S, Elliott AM. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *J Mol Diagn.* 2016;18: 923–932.
183. Gregory A, Hayflick SJ. Pantothenate Kinase-Associated Neurodegeneration. In: Adam MP, Arlinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, et al., editors. *GeneReviews*. Seattle (WA): University of Washington, Seattle; 2002.
184. Lee J-H, Gregory A, Hogarth P, Rogers C, Hayflick SJ. Looking Deep into the Eye-of-the-Tiger in Pantothenate Kinase-Associated Neurodegeneration. *AJNR Am J Neuroradiol.* 2018;39: 583–588.

185. Köhler S, Schulz MH, Krawitz P, Bauer S, Dölkens S, Ott CE, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet.* 2009;85: 457–464.
186. Stelzer G, Plaschkes I, Oz-Levi D, Alkelai A, Olander T, Zimmerman S, et al. VarElect: the phenotype-based variation prioritizer of the GeneCards Suite. *BMC Genomics.* 2016;17 Suppl 2: 444.
187. Di Bella D, Lazzaro F, Brusco A, Plumari M, Battaglia G, Pastore A, et al. Mutations in the mitochondrial protease gene AFG3L2 cause dominant hereditary ataxia SCA28. *Nat Genet.* 2010;42: 313–321.
188. Atorino L, Silvestri L, Koppen M, Cassina L, Ballabio A, Marconi R, et al. Loss of m-AAA protease in mitochondria causes complex I deficiency and increased sensitivity to oxidative stress in hereditary spastic paraplegia. *J Cell Biol.* 2003;163: 777–787.
189. Kondadi AK, Wang S, Montagner S, Kladt N, Korwitz A, Martinelli P, Herholz D, Baker MJ, Schauss AC, Langer T, Rugarli EI. Loss of the m-AAA protease subunit AFG₃L₂ causes mitochondrial transport defects and tau hyperphosphorylation. *EMBO J.* 2014 May 2;33(9):1011–26. doi: 10.1002/embj.201387009. Epub 2014 Mar 28. PMID: 24681487; PMCID: PMC4193934.
190. Rugarli EI, Langer T. Translating m-AAA protease function in mitochondria to hereditary spastic paraplegia. *Trends in Molecular Medicine.* 2006. pp. 262–269. doi:10.1016/j.molmed.2006.04.002
191. DeLuca GC, Ebers GC, Esiri MM. The extent of axonal loss in the long tracts in hereditary spastic paraplegia. *Neuropathology and Applied Neurobiology.* 2004. pp. 576–584. doi:10.1111/j.1365-2990.2004.00587.x
192. Nolden M, Ehses S, Koppen M, Bernacchia A, Rugarli EI, Langer T. The m-AAA Protease Defective in Hereditary Spastic Paraplegia Controls Ribosome Assembly in Mitochondria. *Cell.* 2005. pp. 277–289. doi:10.1016/j.cell.2005.08.003
193. Pierson TM, Adams D, Bonn F, Martinelli P, Cherukuri PF, Teer JK, et al. Whole-Exome Sequencing Identifies Homozygous AFG3L2 Mutations in a Spastic Ataxia-Neuropathy Syndrome Linked to Mitochondrial m-AAA Proteases. *PLoS Genetics.* 2011. p. e1002325. doi:10.1371/journal.pgen.1002325
194. Cagnoli C, Stevanin G, Brussino A, Barberis M, Mancini C, Margolis RL, et al. Missense mutations in the AFG3L2 proteolytic domain account for ~1.5% of European autosomal dominant cerebellar ataxias. *Hum Mutat.* 2010;31: 1117–1124.
195. Myers KA, Warman Chardon J, Huang L, Boycott KM. Deletion of AFG3L2 associated with spinocerebellar ataxia type 28 in the context of multiple genomic anomalies. *Am J Med Genet A.* 2014;164A: 3209–3212.
196. Smets K, Deconinck T, Baets J, Sieben A, Martin J-J, Smouts I, et al. Partial deletion of AFG3L2 causing spinocerebellar ataxia type 28. *Neurology.* 2014;82: 2092–2100.
197. Musova Z, Kaiserova M, Kriegova E, Fillerova R, Vasovcak P, Santava A, et al. A Novel Frameshift Mutation in the AFG3L2 Gene in a Patient with Spinocerebellar Ataxia. *The Cerebellum.* 2014. pp. 331–337. doi:10.1007/s12311-013-0538-z
198. Mancini C, Orsi L, Guo Y, Li J, Chen Y, Wang F, et al. An atypical form of AOA2 with myoclonus associated with mutations in SETX and AFG3L2. *BMC Med Genet.* 2015;16: 16.
199. Bouchard J-P, Brais B, Dupré N, Rouleau GA. Chapter 9 Hereditary Ataxias and Spastic Parapareses in Northeastern Canada. *Spinocerebellar Degenerations: The Ataxias and Spastic Paraplegias.* 2007. pp. 222–243. doi:10.1016/s1877-184x(09)70083-8
200. Colucci F, Neri M, Fortunato F, Ferlini A, Carrozzo R, Torraco A, et al. AFG3L2 Biallelic Mutation: Clinical Heterogeneity in Two Italian Patients. *Cerebellum.* 2022. doi:10.1007/s12311-022-01497-y

201. Eskandani A, AlHashem A, Ali E-S, AlShahwan S, Tlili K, Hundallah K, et al. Recessive AFG3L2 Mutation Causes Progressive Microcephaly, Early Onset Seizures, Spasticity, and Basal Ganglia Involvement. *Pediatric Neurology*. 2017. pp. 24–28. doi:10.1016/j.pediatrneurol.2017.03.019
202. Caporali L, Magri S, Legati A, Del Dotto V, Tagliavini F, Balistreri F, et al. ATPase Domain AFG3L2 Mutations Alter OPA1 Processing and Cause Optic Neuropathy. *Ann Neurol*. 2020;88: 18–32.
203. Charif M, Roubertie A, Salime S, Mamouni S, Goizet C, Hamel CP, et al. A novel mutation of AFG3L2 might cause dominant optic atrophy in patients with mild intellectual disability. *Front Genet*. 2015;6: 311.
204. Maricich SM, Zoghbi HY. Dominantly Inherited Spinocerebellar Syndromes. *Neuromuscular Disorders of Infancy, Childhood, and Adolescence*. 2015. pp. 1003–1032. doi:10.1016/b978-0-12-417044-5.00050-0
205. Muona M, Berkovic SF, Dibbens LM, Oliver KL, Maljevic S, Bayly MA, et al. A recurrent de novo mutation in KCNC1 causes progressive myoclonus epilepsy. *Nat Genet*. 2015;47: 39–46.
206. Calandra CR, Buda G, Vishnopska SA, Oliveri J, Olivier FA, Pérez Millán MI, et al. Spastic ataxia with eye-of-the-tiger-like sign in 4 siblings due to novel compound heterozygous AFG3L2 mutation. *Parkinsonism Relat Disord*. 2020;73: 52–54.
207. Buda G, Valdez RM, Biagioli G, Olivier FA, Affranchino N, Bouso C, et al. Inflammatory cutaneous lesions and pulmonary manifestations in a new patient with autosomal recessive ISG15 deficiency case report. *Allergy, Asthma & Clinical Immunology*. 2020. doi:10.1186/s13223-020-00473-7
208. Perng Y-C, Lenschow DJ. ISG15 in antiviral immunity and beyond. *Nat Rev Microbiol*. 2018;16: 423–439.
209. Loeb KR, Haas AL. The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins. *J Biol Chem*. 1992;267: 7806–7813.
210. Bogunovic D, Boisson-Dupuis S, Casanova J-L. ISG15: leading a double life as a secreted molecule. *Exp Mol Med*. 2013;45: e18.
211. Campbell JA, Lenschow DJ. Emerging roles for immunomodulatory functions of free ISG15. *J Interferon Cytokine Res*. 2013;33: 728–738.
212. Chang Y-G, Yan X-Z, Xie Y-Y, Gao X-C, Song A-X, Zhang D-E, et al. Different roles for two ubiquitin-like domains of ISG15 in protein modification. *J Biol Chem*. 2008;283: 13370–13377.
213. Zhang D, Zhang D-E. Interferon-stimulated gene 15 and the protein ISGylation system. *J Interferon Cytokine Res*. 2011;31: 119–130.
214. Zhang X, Bogunovic D, Payelle-Brogard B, Francois-Newton V, Speer SD, Yuan C, et al. Human intracellular ISG15 prevents interferon- α/β over-amplification and auto-inflammation. *Nature*. 2015;517: 89–93.
215. Bogunovic D, Byun M, Durfee LA, Abhyankar A, Sanal O, Mansouri D, et al. Mycobacterial disease and impaired IFN- γ immunity in humans with inherited ISG15 deficiency. *Science*. 2012;337: 1684–1688.
216. Martin-Fernandez M, Bravo García-Morato M, Gruber C, Murias Loza S, Malik MNH, Alsohime F, et al. Systemic Type I IFN Inflammation in Human ISG15 Deficiency Leads to Necrotizing Skin Lesions. *Cell Rep*. 2020;31: 107633.
217. Francois-Newton V, Livingstone M, Payelle-Brogard B, Uzé G, Pellegrini S. USP18 establishes the transcriptional and anti-proliferative interferon α/β differential. *Biochemical Journal*. 2012. pp. 509–516. doi:10.1042/bj20120541

218. Stark GR, Darnell JE. The JAK-STAT Pathway at Twenty. *Immunity*. 2012. pp. 503–514. doi:10.1016/j.jimmuni.2012.03.013
219. Trinchieri G. Interleukin-12 and the regulation of innate resistance and adaptive immunity. *Nat Rev Immunol*. 2003;3: 133–146.
220. Crow YJ, Manel N. Aicardi–Goutières syndrome and the type I interferonopathies. *Nature Reviews Immunology*. 2015. pp. 429–440. doi:10.1038/nri3850
221. Jeremiah N, Neven B, Gentili M, Callebaut I, Maschalidi S, Stolzenberg M-C, et al. Inherited STING-activating mutation underlies a familial inflammatory syndrome with lupus-like manifestations. *J Clin Invest*. 2014;124: 5516–5520.
222. Volpi S, Picco P, Caorsi R, Candotti F, Gattorno M. Type I interferonopathies in pediatric rheumatology. *Pediatric Rheumatology*. 2016. doi:10.1186/s12969-016-0094-4
223. d'Angelo DM, Di Filippo P, Breda L, Chiarelli F. Type I Interferonopathies in Children: An Overview. *Front Pediatr*. 2021;9: 631329.
224. Moreira-Teixeira L, Mayer-Barber K, Sher A, O'Garra A. Type I interferons in tuberculosis: Foe and occasionally friend. *J Exp Med*. 2018;215: 1273–1285.
225. Rodero MP, Crow YJ. Type I interferon-mediated monogenic autoinflammation: The type I interferonopathies, a conceptual overview. *J Exp Med*. 2016;213: 2527–2538.
226. Kastner DL, Aksentijevich I, Goldbach-Mansky R. Autoinflammatory disease reloaded: a clinical perspective. *Cell*. 2010;140: 784–790.
227. Demir S, Sag E, Dedeoglu F, Ozen S. Vasculitis in Systemic Autoinflammatory Diseases. *Front Pediatr*. 2018;6: 377.
228. Davidson S, Steiner A, Harapas CR, Masters SL. An Update on Autoinflammatory Diseases: Interferonopathies. *Curr Rheumatol Rep*. 2018;20: 38.
229. Harapas CR, Steiner A, Davidson S, Masters SL. An Update on Autoinflammatory Diseases: Inflammasomopathies. *Curr Rheumatol Rep*. 2018;20: 40.
230. Steiner A, Harapas CR, Masters SL, Davidson S. An Update on Autoinflammatory Diseases: Relopathies. *Curr Rheumatol Rep*. 2018;20: 39.
231. Manthiram K, Zhou Q, Aksentijevich I, Kastner DL. Corrigendum: The monogenic autoinflammatory diseases define new pathways in human innate immunity and inflammation. *Nat Immunol*. 2017;18: 1271.
232. Cuadernos de Autoinmunidad. [cited 17 Sep 2021]. Available: <https://sademi.com/wp-content/uploads/2016/01/Cuadernos-Autoinmunidad-A%C3%B1o-13-num-2.pdf>
233. Al-Mayouf S, Alsonbul A. Monogenic and multifactorial autoinflammatory diseases: Clinical and laboratory characterization in a pediatric Saudi population. *Pediatric Rheumatology*. 2015. doi:10.1186/1546-0096-13-s1-p160
234. Vanoni F, Theodoropoulou K, Hofer M. PFAPA syndrome: a review on treatment and outcome. *Pediatr Rheumatol Online J*. 2016;14: 38.
235. Labrousse M, Kevorkian-Verguet C, Boursier G, Rowczenio D, Maurier F, Lazaro E, et al. Mosaicism in autoinflammatory diseases: Cryopyrin-associated periodic syndromes (CAPS) and beyond. A systematic review. *Crit Rev Clin Lab Sci*. 2018;55: 432–442.
236. Hoffman HM, Broderick L. Editorial: It Just Takes One: Somatic Mosaicism in Autoinflammatory Disease. *Arthritis & rheumatology (Hoboken, N.J.)*. 2017. pp. 253–256.
237. Ryan JG, Goldbach-Mansky R. The spectrum of autoinflammatory diseases: recent bench to bedside observations. *Curr Opin Rheumatol*. 2008;20: 66–75.

238. Jesus AA, Goldbach-Mansky R. IL-1 Blockade in Autoinflammatory Syndromes. *Annual Review of Medicine*. 2014. pp. 223–244. doi:10.1146/annurev-med-061512-150641
239. Caorsi R, Federici S, Gattorno M. Biologic drugs in autoinflammatory syndromes. *Autoimmun Rev*. 2012;12: 81–86.
240. Reinhardt RL, Liang H-E, Bao K, Price AE, Mohrs M, Kelly BL, et al. A novel model for IFN- γ -mediated autoinflammatory syndromes. *J Immunol*. 2015;194: 2358–2368.
241. Efthimiou P. *Auto-Inflammatory Syndromes: Pathophysiology, Diagnosis, and Management*. Springer; 2019.
242. Federici S, Gattorno M. A practical approach to the diagnosis of autoinflammatory diseases in childhood. *Best Practice & Research Clinical Rheumatology*. 2014. pp. 263–276. doi:10.1016/j.berh.2014.05.005
243. Holzer FS, Mastroleo ID. Ethical Aspects of Precision Medicine: An Introduction to the Ethics and Concept of Clinical Innovation. *Precision Medicine*. 2018. pp. 1–19. doi:10.1016/b978-0-12-805364-5.00001-9
244. Van Gijn ME, Ceccherini I, Shinar Y, Carbo EC, Slofstra M, Arostegui JI, et al. New workflow for classification of genetic variants' pathogenicity applied to hereditary recurrent fevers by the International Study Group for Systemic Autoinflammatory Diseases (INSAID). *Journal of Medical Genetics*. 2018. pp. 530–537. doi:10.1136/jmedgenet-2017-105216
245. Touitou I, Lesage S, McDermott M, Cuisset L, Hoffman H, Dode C, et al. Infevers: an evolving mutation database for auto-inflammatory syndromes. *Hum Mutat*. 2004;24: 194–198.
246. Eurofever Project - The Eurofever registry. [cited 8 Aug 2021]. Available: <https://www.printo.it/eurofever/registry>
247. Orphanet: an online rare disease and orphan drug database. Copyright, INSERM 1997. Disponible en <http://www.orpha.net>. Accessed : August 21, 2017.
248. D'Osualdo A, Picco P, Caroli F, Gattorno M, Giacchino R, Fortini P, et al. MVK mutations and associated clinical features in Italian patients affected with autoinflammatory disorders and recurrent fever. *Eur J Hum Genet*. 2005;13: 314–320.
249. Prietsch V, Mayatepek E, Krastel H, Haas D, Zundel D, Waterham HR, et al. Mevalonate kinase deficiency: enlarging the clinical and biochemical spectrum. *Pediatrics*. 2003;111: 258–261.
250. Mandey SHL, Kuijk LM, Frenkel J, Waterham HR. A role for geranylgeranylation in interleukin-1 β secretion. *Arthritis & Rheumatism*. 2006. pp. 3690–3695. doi:10.1002/art.22194
251. Normand S, Massonnet B, Delwail A, Favot L, Cuisset L, Grateau G, et al. Specific increase in caspase-1 activity and secretion of IL-1 family cytokines: a putative link between mevalonate kinase deficiency and inflammation. *European Cytokine Network*. 2009. pp. 101–107. doi:10.1684/ecn.2009.0163
252. Bader-Meunier B, Martins AL, Charbit-Henrion F, Meinzer U, Belot A, Cuisset L, et al. Mevalonate Kinase Deficiency: A Cause of Severe Very-Early-Onset Inflammatory Bowel Disease. *Inflamm Bowel Dis*. 2021;27: 1853–1857.
253. Lachmann HJ, Lauwers B, Miettunen P, Kallinich T, Jansson A, Rosner I, et al. Canakinumab improves patient-reported outcomes in children and adults with autoinflammatory recurrent fever syndromes: results from the CLUSTER trial. *Clin Exp Rheumatol*. 2021;39 Suppl 132: 51–58.
254. Lausch E, Janecke A, Bros M, Trojandt S, Alanay Y, De Laet C, et al. Genetic deficiency of tartrate-resistant acid phosphatase associated with skeletal dysplasia, cerebral calcifications and autoimmunity. *Nat Genet*. 2011;43: 132–137.

255. Niewold T. Faculty Opinions recommendation of Tartrate-resistant acid phosphatase deficiency causes a bone dysplasia with autoimmunity and a type I interferon expression signature. Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature. 2011. doi:10.3410/f.8249957.8661070
256. Niu T, De Rosny C, Chautard S, Rey A, Patoli D, Groslambert M, et al. NLRP3 phosphorylation in its LRR domain critically regulates inflammasome assembly. *Nat Commun.* 2021;12: 5862.
257. Parkhouse R, Boyle JP, Monie TP. Blau syndrome polymorphisms in NOD2 identify nucleotide hydrolysis and helical domain 1 as signalling regulators. *FEBS Lett.* 2014;588: 3382–3389.
258. Girardin SE, Boneca IG, Viala J, Chamaillard M, Labigne A, Thomas G, et al. Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem.* 2003;278: 8869–8872.
259. Philpott DJ, Sorbara MT, Robertson SJ, Croitoru K, Girardin SE. NOD proteins: regulators of inflammation in health and disease. *Nature Reviews Immunology.* 2014. pp. 9–23. doi:10.1038/nri3565
260. Limviphuvadh V, Tan CS, Konishi F, Jenjaroenpun P, Xiang JS, Kremenska Y, et al. Discovering novel SNPs that are correlated with patient outcome in a Singaporean cancer patient cohort treated with gemcitabine-based chemotherapy. *BMC Cancer.* 2018;18: 1–16.
261. Molho-Pessach V, Ramot Y, Camille F, Doviner V, Babay S, Luis SJ, et al. H syndrome: the first 79 patients. *J Am Acad Dermatol.* 2014;70: 80–88.
262. Cekin N, Akyurek ME, Pinarbası E, Ozen F. MEFV mutations and their relation to major clinical symptoms of Familial Mediterranean Fever. *Gene.* 2017;626: 9–13.
263. Babaoglu H, Varan O, Kucuk H, Turan A, Tufan A. Osteitis as a manifestation of familial Mediterranean fever. *Rheumatology.* 2017;56: 2035–2036.
264. Chae JJ, Komarow HD, Cheng J, Wood G, Raben N, Paul Liu P, et al. Targeted Disruption of Pyrin, the FMF Protein, Causes Heightened Sensitivity to Endotoxin and a Defect in Macrophage Apoptosis. *Molecular Cell.* 2003. pp. 591–604. doi:10.1016/s1097-2765(03)00056-x
265. Daia A, Kini V, Taha RZ, El-Shanti H, Fathalla B. Chronic recurrent multifocal osteomyelitis in a patient with Familial Mediterranean Fever. *Pediatric Rheumatology.* 2015. doi:10.1186/1546-0096-13-s1-p104
266. Moradian MM, Sarkisian T, Amaryan G, Hayrapetyan H, Yeghiazaryan A, Davidian N, et al. Patient management and the association of less common familial Mediterranean fever symptoms with other disorders. *Genet Med.* 2014;16: 258–263.
267. Maggio MC, Corsello G. FMF is not always “fever”: from clinical presentation to “treat to target.” *Italian Journal of Pediatrics.* 2020. doi:10.1186/s13052-019-0766-z
268. Wallach D. Neutrophilic Dermatoses: An Overview. *Neutrophilic Dermatoses.* 2018. pp. 5–9. doi:10.1007/978-3-319-72649-6_2
269. Shoham NG, Centola M, Mansfield E, Hull KM, Wood G, Wise CA, et al. Pyrin binds the PSTPIP1/CD2BP1 protein, defining familial Mediterranean fever and PAPA syndrome as disorders in the same pathway. *Proc Natl Acad Sci U S A.* 2003;100: 13501–13506.
270. Kosukcu C, Taskiran EZ, Batu ED, Sag E, Bilginer Y, Alikasifoglu M, et al. Whole exome sequencing in unclassified autoinflammatory diseases: more monogenic diseases in the pipeline? *Rheumatology.* 2021;60: 607–616.
271. Kumar N, Boes CJ, Babovic-Vuksanovic D, Boeve BF. The “Eye-of-the-Tiger” Sign Is Not Pathognomonic of the PANK2 Mutation. *Archives of Neurology.* 2006. p. 292. doi:10.1001/archneur.63.2.292

272. Chang C, Lin C. Eye-of-the-Tiger sign is not Pathognomonic of Pantothenate Kinase-Associated Neurodegeneration in Adult Cases. *Brain and Behavior*. 2011. pp. 55–56. doi:10.1002/brb3.8
273. Sismani C, Koufaris C, Voskarides K. Copy Number Variation in Human Health, Disease and Evolution. *Genomic Elements in Health, Disease and Evolution*. 2015. pp. 129–154. doi:10.1007/978-1-4939-3070-8_6
274. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC Bioinformatics*. 2013;14 Suppl 11: S1.
275. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22: 568–576.
276. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Research*. 2012. pp. 1525–1532. doi:10.1101/gr.138115.112
277. Yamamoto T, Shimojima K, Ondo Y, Imai K, Chong PF, Kira R, et al. Challenges in detecting genomic copy number aberrations using next-generation sequencing data and the eXome Hidden Markov Model: a clinical exome-first diagnostic approach. *Hum Genome Var*. 2016;3: 16025.