

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

CARACTERIZACIÓN MOLECULAR Y FENOTÍPICA DE LA ATAXIA CON
APRAXIA OCULOMOTORA EN COSTA RICA MEDIANTE ENFOQUES
MOLECULARES Y BIOINFORMÁTICOS

Tesis sometida a la consideración de la Comisión del Programa de Estudios de
Posgrado en Biología para optar al grado y título de Maestría Académica en Biología
con énfasis en Genética y Biología Molecular

HILDA GUADALUPE TORRES ULATE

Ciudad Universitaria Rodrigo Facio, Costa Rica

2022

Agradecimientos

Para mis padres por su apoyo, consejos, ayuda en los momentos difíciles, y por brindarme los recursos necesarios para estudiar. Me han dado todo lo que soy como persona, mis valores, mis principios, mi carácter, mi empeño, mi coraje para conseguir mis objetivos y no desmayar en los problemas que se presentaban, enseñándome a encarar las adversidades sin perder nunca la dignidad ni desfallecer en el intento. A mis hermanos por su compañía y sabias palabras en los momentos necesarios.

Agradezco también a mi comité asesor. A Ph.D. Alejandro Leal Esquivel, Ph.D. Gabriela Chavarría Soley y a Ph.D. José Guevara Coto por toda la colaboración y el apoyo. Gracias por su enseñanza durante este proceso.

A, Martha Gómez, Saby Cruz, Lidia Benítez, Ricardo Villalobos, Marcelo Castro, María de los Ángeles Ulate, Monserrat Lobo y Carolina Céspedes por toda la ayuda, compañía, el aliento brindado y por sus valiosos aportes para el desarrollo de esta investigación.

“Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Biología de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Biología con énfasis en Genética y Biología Molecular

Dr. Andrey Sequeira Cordero
**Representante de la Decana
Sistema de Estudios de Posgrado**

Dr. Alejandro Leal Esquivel
Profesor Guía

Dra. Gabriela Chavarría Soley
Lectora

Dr. José Andrés Guevara Coto
Lector

Dr. Eric Fuchs Castillo
Director del Programa de Posgrado en Biología

Hilda Guadalupe Torres Ulate
Sustentante

Contenido

Agradecimientos	ii
Hoja de aprobación	iii
Contenido.....	iv
Resumen.....	v
Lista de Cuadros	vii
Introducción.....	1
Metodología.....	8
Objetivo General.....	15
Objetivos Específicos	15
Artículo: Analysis of ataxia in Costa Rica through an integrative approach.....	16
Conclusiones.....	38
Referencias Bibliográficas.....	40

Resumen

Las ataxias son un grupo de trastornos clínicamente heterogéneos, que pueden ser hereditables o no dependiendo de la etiología. Las manifestaciones clínicas de las ataxias hereditarias son una incoordinación progresiva del movimiento y del habla (disartria), y una marcha inestable, descoordinada y de base amplia. Además, los pacientes pueden desarrollar oftalmoplejía (limitaciones del movimiento ocular), espasticidad, neuropatía y dificultades cognitivas. Este trabajo se enfoca en la ataxia con apraxia oculomotora (AOA), que es una enfermedad de herencia autosómica recesiva. Hasta la fecha se han descrito cuatro tipos de AOA (AOA1-AOA4). Los diferentes tipos de ataxia comparten varias de sus características clínicas, los cuales se diagnostican mediante la historia familiar, el examen físico, la neuroimagen y las pruebas moleculares. Debido a que la información sobre las mutaciones causantes de AOA en Costa Rica es escasa, se planea identificar molecularmente las mutaciones causantes de AOA en pacientes costarricenses mediante los métodos de secuenciación de Sanger y secuenciación de nueva generación, por lo que este trabajo contribuiría a caracterizar las mutaciones prevalentes en los pacientes que padecen ataxia con apraxia oculomotora en el país, así como el fenotipo asociado. Además, sentaría las bases para desarrollar protocolos de trabajo para el diagnóstico molecular de la enfermedad.

El desafío de determinar si las manifestaciones fenotípicas están asociadas con grupos específicos de proteínas que comparten una relación subyacente, representa una oportunidad para estudiar cómo surgen las diferentes manifestaciones fenotípicas. Pero además, los enfoques bioinformáticos recientes en el estudio de las proteínas y las relaciones entre aminoácidos podrían permitir el análisis de agrupamiento e identificación de grupos de residuos involucrados en los diferentes tipos de AOA. Esto con el objetivo de asociar los

fenotipos clínicos reportados a grupos de residuos proteicos que se encuentran conservados o relacionados con otros residuos dentro de la misma estructura de la proteína.

Lista de Cuadros

Cuadro 1. Resultados de pruebas serológicas y sus variaciones según el tipo de AOA ...	3
Cuadro 2. Mutaciones descritas e incluidas en OMIM hasta la fecha, para cada uno de los tipos de AOA	6



UNIVERSIDAD DE
COSTA RICA

SEP Sistema de
Estudios de Posgrado

Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, Hilda Guadalupe Torres Ulate, con cédula de identidad 114780692, en mi condición de autor del TFG titulado CARACTERIZACIÓN MOLECULAR Y FENOTÍPICA DE LA ATAXIA CON APRAXIA OCULOMOTORA EN COSTA RICA MEDIANTE ENFOQUES MOLECULARES Y BIOINFORMÁTICOS

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI NO *

*En caso de la negativa favor indicar el tiempo de restricción: 2 año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

Introducción

Las ataxias son un grupo de trastornos dentro de las enfermedades cerebelares, de las cuales es posible encontrar registros que datan de mediados desde el siglo XIX y principios del siglo XX cuando se hicieron los primeros esfuerzos por tratar de clasificarlas (Holmes, 1908). Estos trastornos clínica y genéticamente heterogéneos, se caracterizan por una descoordinación lentamente progresiva de la marcha. A menudo son asociadas con una mala coordinación en el movimiento de las manos, movimientos oculares, habla, y con frecuencia se produce la atrofia del cerebelo. Las ataxias pueden ser:

- a. No hereditarias: cuando aparecen en la adultez (Klockgether, 2010) son causadas por factores tales como alcoholismo crónico, deficiencias vitamínicas, enfermedad vascular, atrofia muscular espinal, etc. (Jayadev & Bird, 2013).
- b. Hereditarias: cuando son causadas por una mutación causal en un gen. Hasta la fecha, se sabe que hay más de 30 formas autosómicas dominantes y más de 60 formas que son autosómicas recesivas (Ruano, Melo, Silva, & Coutinho, 2014). Muchas de las ataxias hereditarias tienen presentaciones superpuestas y existe un alto grado de heterogeneidad genética (Inlora et al., 2017).

Las manifestaciones clínicas de las ataxias hereditarias son una incoordinación progresiva del movimiento y del habla (disartria), y una marcha inestable, descoordinada y de base amplia. Además, los pacientes pueden desarrollar limitaciones del movimiento ocular (oftalmoplejía), espasticidad, neuropatía y dificultades cognitivas (Jayadev & Bird, 2013). Para establecer el diagnóstico de ataxia hereditaria se requiere:

- Detectar signos clínicos típicos en el examen neurológico (Klockgether, 2010).

- Documentación de la posible naturaleza hereditaria de la enfermedad mediante el desarrollo de un historial familiar (Jayadev & Bird, 2013). En caso de no poder identificar una causa genética es necesario proceder con la exclusión de causas no genéticas (Klockgether, 2010).

Dentro de las ataxias de herencia autosómica dominante, Schöls et al; 2004 definen a las ataxias espinocerebelosas (SCA) como un grupo clínica y genéticamente heterogéneo que causa la degeneración progresiva del cerebelo y sus conexiones aferentes y eferentes. Estos trastornos están causados por una mutación de expansión de repetición del triplete CAG en regiones codificantes de los genes. (Pulst et al., 1996). Las SCA más comunes son SCA1, SCA2, SCA3 y SCA6 (Schmitz-Hübsch et al., 2008). Esto puede variar en ciertas regiones debido a un efecto fundador (Manto, 2005).

Los pacientes suelen presentar un síndrome cerebeloso lentamente progresivo con diversas combinaciones de trastornos oculomotores, disartria, dismetría/temblor cinético y/o marcha atáxica. Al ser genéticamente heterogéneas puede presentarse un traslape de los fenotipos entre los diferentes subtipos (Manto, 2005).

Durante los últimos quince años se han descrito el segundo grupo de ataxias cerebelosas con la mayor prevalencia a nivel mundial dentro de las ataxias con herencia autosómica recesiva (Ruano et al., 2014).

Estas ataxias son causadas por mutaciones en genes responsables de reparación del ADN, terminación prematura de la proteína, deficiencias de maduración de la proteína, o ambos. Dentro del mismo se incluyen la ataxia con apraxia oculomotora (AOA), ataxia

telangiectasia (AT) y el más inusual, el trastorno similar a ataxia telangiectasia (ATLD) (Mariani et al., 2017).

Con respecto a la AOA, hasta la fecha se han descrito cuatro tipos: AOA1, AOA2, AOA3 (Tassan et al., 2012) y AOA4/CMT2B (Leal et al., 2018). Los diferentes tipos de AOA comparten síntomas entre ellos (Inlora et al., 2017).

La AOA1 tiene una edad de inicio que va desde la infancia hasta la preadolescencia (2-12 años) y es causada por mutaciones en el gen *APTX* (Coutinho & Barbot, 2002) (Cuadro 1), que se localiza en 9p21.1; tiene 7 exones y un tamaño de 28.217 pb. (Kent et al., 2002). *APTX* codifica para la proteína aprataxina, que se encarga de catalizar la liberación nucleofílica de los grupos adenilato unidos covalentemente a los extremos 5' fosfato en rupturas de una sola hebra, dando como resultado la producción de los extremos 5' fosfato que se pueden unir de manera eficiente durante la reparación del ADN (MIM 2089200; Prasad et al., 2009).

Cuadro 1. Mutaciones descritas e incluidas en OMIM hasta la fecha, para cada uno de los tipos de AOA

AOA1 <i>APTX</i>	AOA2 <i>SETX</i>	AOA3 <i>PIK3R5</i>	AOA4 <i>PNKP</i>
<ul style="list-style-type: none"> ● p.Lys247 ● p.Pro206Leu ● 689insT ● 318delT ● p.Val89Gly ● p.His27Arg ● Trp279Ter ● p.Trp279Arg ● Del. 7 exones ● Leu223Pro ● c.689dupT ● p.Glu232Gly- fsTer38 	<ul style="list-style-type: none"> ● c.839-2A>G ● p.Thr2154Met ● p.Arg168Trp ● p.Leu1976Arg ● p.Glu65Lys 	<ul style="list-style-type: none"> ● p.Pro629Ser 	<ul style="list-style-type: none"> ● c0.1189-15_1191del18 ● p.Gly375Trp ● Gly442AlafsTer27 ● Thr408del ● Gln517LeufsTer24

Fuente: Gatti et al., 2019; Laurencin et al., 2015; MIM. 2089200; Duquette et al., 2005; Mignarri, Tessa, Federico, Santorelli, & Dotti, 2015; Szpisjak, Obal, Engelhardt, Vecsei, Klivenyi, 2016; Tassan et al., 2012; Bras, et al., 2015 and Schiess et al., 2017.

La aprataxina es un miembro de la superfamilia de la triada de histidinas (HIT) (van Minkelen et al., 2015), que son una antigua superfamilia de nucleótido hidrolasas y transferasas que actúan sobre el α -fosfato de ribonucleótidos o en sustratos que contienen nucleótidos en vías de señalización importantes para el crecimiento celular, apoptosis y metabolismo de ADN, ARN y carbohidratos (Brenner, 2002). La aprataxina se expresa en los siguientes tejidos: cerebro, tiroides, paratiroides, glándulas adrenales, pulmones, páncreas, hígado, vesícula biliar, vejiga, corazón, riñón, hígado, pulmón, linfa, mioblastos, miotubos, páncreas, placenta, músculo esquelético, cordón espinal y tálamo (Uhlen et al., 2015).

La AOA2 se describió por primera vez en el año 2000 (M. C. Moreira et al., 2004). Es ocasionada por mutaciones en el gen *SETX* (Cuadro 3) y tiene una edad media de inicio en la pre adolescencia (12.7 años) (Schiess, et al., 2017). Codifica para una proteína que por su homología con la proteína Sen1p de los hongos, se denominó Senataxina. La senataxina tiene actividad de ARN helicasa codificada por un dominio helicasa de ADN/ARN en el extremo C-terminal, lo que sugiere que puede estar involucrada en el procesamiento de ADN y ARN (Prasad et al., 2009). Se expresa en: músculo esquelético, corteza cerebral, apéndice, testículos, músculo cardíaco, cerebelo, músculo liso, pulmón, nasofaringe, intestino delgado, hipocampo, piel, tiroides, endometrio, médula ósea, bronquios, riñón, recto, trompas de Falopio, estómago, epidídimo y colon. (Uhlen et al., 2015)

Tassan y colaboradores (2012) identificaron la AOA3 en una familia consanguínea de Arabia Saudita, cuyos miembros afectados presentaban características clínicas similares a las de los individuos con AOA2, pero con una edad media de inicio en la adolescencia (15.6 años). Este tipo de ataxia es causada por mutaciones en el gen *PIK3R5* (Tassan et al., 2012) (Cuadro 3), que se localiza en 17p13.1, tiene 18 exones y un tamaño de 30.869

pb (Kent et al., 2002). Este gen codifica por la subunidad reguladora 5 del complejo de clase fosfatidilinositol 3-quinasa gamma (PIK3 γ), que es una subunidad reguladora de 101 kD del complejo de clase I PIK3 γ . PIK3R5 es un enzima dimérica (Uhlen et al., 2015). Las fosfatidilinositol 3-quinasa (PI3K) son miembros de una familia única y muy conservada de quinasa intracelulares que fosforilan el grupo 3'-hidroxilo del fosfatidilinositol y fosfoinosítidos. Esta reacción tiene como resultado, la activación de múltiples vías de señalización intracelular que regulan funciones diversas e importantes, como el metabolismo celular, la supervivencia y la polaridad, y el tráfico de vesículas (Engelman, Luo, & Cantley, 2006). Esta proteína se expresa en duodeno, estómago, apéndice, pulmón, nasofaringe, vesícula biliar, tiroides, glándula adrenal, intestino delgado, colon, recto, riñón, epidídimo, vesícula seminal, cérvix, bronquios, placenta, glándulas salivales, trompas de Falopio y médula ósea (Prasad *et al.*, 2009).

El último tipo descrito de AOA presenta una edad de inicio promedio de la enfermedad en la infancia (4.3 años) (Schiess et al., 2017) y en algunos casos se caracteriza por ataxia cerebelosa, apraxia oculomotora, polineuropatía y atrofia cerebelosa en la resonancia magnética, dicho fenotipo se clasificó como ataxia con apraxia oculomotora tipo 4 (AOA4) (Gatti et al., 2019). En otros casos, el fenotipo clínico presentado se caracterizó por la presencia de una polineuropatía motora y sensorial parecido a una forma de la enfermedad de Charcot-Marie-Tooth e identificado como CMT2B2 (Leal et al., 2018; Pedroso et al., 2015). La causa se debe a mutaciones en el gen *PNKP* (Cuadro 3), el cual se ubica en 19q13.33, cuenta con 16 exones y tiene un tamaño de 6.337 pb. *PNKP* codifica por una fosfatasa polinucleótido quinasa 3', que en respuesta al daño por radiación ionizante o al daño oxidativo del ADN, cataliza la fosforilación 5' de ácidos nucleicos y también tiene una actividad fosfatasa 3' asociada, que predice una función importante en la reparación del ADN (Uhlen *et al.*, 2015). Se ha reportado la expresión

de esta proteína en cerebro, colon, corazón, riñón, hígado, pulmón, ovarios, páncreas, placenta, próstata, músculo esquelético, intestino delgado, bazo y testículos (Prasad *et al.*, 2009).

Debido a que los diferentes tipos de ataxia comparten varias de sus características clínicas (Mariani *et al.*, 2017), estos se diagnostican mediante la historia familiar, el examen físico, marcadores serológicos (Cuadro 2), la neuroimagen y las pruebas moleculares, las cuales permiten detectar mutaciones nuevas o previamente descritas, como las que se muestran en el Cuadro 3.

Cuadro 2. Resultados de pruebas serológicas y sus variaciones según el tipo de AOA

AOA 1	AOA 2	AOA 3	AOA 4
<ul style="list-style-type: none"> ● Albúmina disminuida ● Colesterol total aumentado ● Alfa fetoproteína normal ● Deficiencia de coenzima Q10 a nivel muscular puede ayudar, pero no es definitiva. 	<ul style="list-style-type: none"> ● Conteo sanguíneo, creatinquinasa y colesterol normales ● Alfa fetoproteína elevada 	<ul style="list-style-type: none"> ● Alfa fetoproteína elevada 	<ul style="list-style-type: none"> ● Albúmina elevada ● Colesterol elevado ● Alfafetoproteína normal.

Fuente: van Minkelen *et al.*, 2015; Becherel, *et al.*, 2015; Tassan *et al.*, 2012 and Schiess, Zee, Siddiqui, Szolics, & El-Hattab, 2017

En cuanto a los resultados de pruebas serológicas, se encuentran diferencias en los biomarcadores listados en el Cuadro 3, sin embargo, estos no pueden utilizarse como un factor discriminativo ya que como ha sido reportado por Coutinho y Barbot (2015) así como por Mignarri, *et al.* (2015), los resultados de los marcadores serológicos pueden traslaparse en los diferentes tipos de AOA, e incluso se han reportado casos de individuos con un mismo tipo de AOA y resultados variables en los biomarcadores serológicos.

Hasta ahora no se han estudiado las AOAs y su causa genética en Costa Rica. No obstante, Leal *et al.* (2018) reportaron las mutaciones p.Gln517ter, en condición homocigota y junto con Thr408del en el gen *PNKP* en pacientes heterocigotos compuestos, como responsable de Charcot Marie Tooth (CMT) con afectación cerebelar.

Esta mutación había sido asociada previamente con la AOA4 (Bras et al., 2015). Por esta razón, es muy probable que en la población costarricense se encuentren familias con AOA4, que carguen la mutación Thr408del también en condición homocigota.

El posible efecto de las mutaciones en el desarrollo de los fenotipos de enfermedad es uno de los focos de interés en la investigación en genética humana. El desafío de determinar si las manifestaciones fenotípicas están asociadas con grupos específicos de residuos que comparten una relación subyacente representa una oportunidad para estudiar cómo surgen las diferentes manifestaciones fenotípicas. Sin embargo, los enfoques bioinformáticos recientes en el estudio de las proteínas y sus relaciones subyacentes han abierto una puerta para que en este proyecto se proponga llevar a cabo un análisis con las familias proteicas involucradas en los diferentes tipos de AOA. Esto mediante un enfoque computacional que buscará agrupar residuos proteicos e identificar aquellos sitios donde exista coevolución, esto se refiere a los cambios coordinados que pueden producirse en biomoléculas, por lo general para mantener o refinar las interacciones funcionales entre esos pares (De Juan, Pazos, & Valencia, 2013) y de esta manera establecer una asociación con los fenotipos clínicos reportados así como identificar nuevos residuos de interés que compartan patrones de coevolución con los residuos que sean asociados a fenotipos de la AOA.

Metodología

ANÁLISIS MOLECULAR

Comité de ética

Esta investigación es parte del proyecto CMT que es dirigido por el Dr.rer.nat. Alejandro Leal Esquivel, el cual se encuentra inscrito ante la Vicerrectoría de Investigación y cuenta con aprobación del comité ético científico de la UCR.

Selección de pacientes

El proceso de reclutamiento de pacientes comenzó con la generación de una base de datos con información de contacto (correo electrónico, número de teléfono, hospital(es) en el que labora) de cada uno de los neurólogos registrados a nivel nacional en la Asociación Costarricense de Ciencias Neurológicas, además se cotejaron estos datos con los de los directorios de diferentes hospitales del ámbito público y privado en el país. Estos datos se utilizaron para establecer líneas de comunicación con todos los neurólogos, se les informó sobre el proyecto de investigación, y se les proporcionó el teléfono de la Escuela de Biología de la UCR, para que, ante un caso sospechoso de AOA, ellos le indicaran al paciente que si lo deseaban podían comunicarse con el equipo investigador.

A cada voluntario se le asignó un código para asegurar la anonimidad de las muestras y cada expediente fue revisados para determinar quiénes contaban con características fenotípicas que se correspondían a las esperadas en ataxia. Una vez completado el estudio de casos, se realizaron dos viajes, uno a Pérez Zeledón y otro a la provincia de Alajuela para tomar muestras de saliva de los familiares de los pacientes seleccionados con el kit prepIT.L2P y se procedió a la fase de análisis de dichas muestras.

Adicional a este reclutamiento se solicitó a la Vicerrectoría de Investigación el acceso a muestras que se encuentran en el Instituto de Investigaciones en Salud (INISA) y que forman parte de otro estudio de neuropatías periféricas realizado en dicha Institución.

Obtención y procesamiento de muestras

El procesamiento comenzó con la extracción de ADN de las 48 muestras obtenidas durante el reclutamiento, para las muestras de sangre se utilizó el kit QIAamp DNA Blood Mini Kit (250) y para las muestras de saliva se utilizó el protocolo de purificación de ADN del kit prepIT.L2P. Una vez extraído el ADN de todas las muestras se procedió a la debida cuantificación de ADN de las mismas con el equipo Thermo Scientific NanoDrop 2000 Uv-vis Spectrophotometer. Todas las muestras contaron con una concentración de ADN adecuada para avanzar a la fase de pruebas de Reacción en Cadena de la Polimerasa para *PNKP*.

El siguiente paso fue la purificación de los productos de las amplificaciones siguiendo el protocolo disponible en el manual de procesos del laboratorio 240 de la Escuela de Biología de la UCR que se basa en un método orgánico a base de sales. Una vez realizada dicha purificación se procedió con el análisis de secuenciación bidireccional de Sanger mediante el método *dye terminator*, para el que se utilizó el secuenciador 3130 Genetic Analyzer (Applied Biosystems) que se encuentra en el laboratorio 240 y con los iniciadores *PNKP-del-F* GGGTTTGTGTTGTCGATGG, *PNKP-del-R* TCTGCCGATCTGTTTGTGAC para la mutación Thr408del y *PNKP-50364522-F* ATGTCTAAAGTGCTCATGCCAGG y *PNKP-50364522-R* GGTACTGTTGGGGATAGCAGG para la mutación p.Gln517X..

Detección de variantes

Todas las regiones de exones de todos los genes humanos (~22.000) fueron capturadas por xGen Exome Research Panel v2 (Integrated DNA Technologies, Coralville, Iowa, USA). Las regiones capturadas del genoma se secuenciaron con Novaseq 6000 (Illumina, San Diego, CA, EE.UU.). El análisis de los datos brutos de la secuenciación del genoma, incluida la alineación con el genoma humano de referencia GRCh37/hg19, la llamada de variantes y la anotación, se llevó a cabo con herramientas bioinformáticas de código abierto y software interno perteneciente a 3billion. El software de interpretación automática de variantes, EVIDENCE, se utilizó para priorizar las variantes basándose en las directrices del ACMG (Richards et al., 2015) y en el fenotipo de cada paciente. Este sistema tiene tres pasos principales; filtración de variantes, clasificación y puntuación de similitud para fenotipo del paciente (Seo et al., 2020). En primer lugar, gnomAD (Karczewski et al., 2020) como base de datos del genoma de la población y la base de datos de 3 mil millones de genomas se utilizaron para estimar la frecuencia alélica.

Filtrado y validación

Las variantes comunes con una frecuencia alélica menor de >5% se filtraron de acuerdo con la BA1 de la directriz del ACMG (Richards et al., 2015). En segundo lugar, se extrajeron datos de evidencia sobre la patogenicidad de variantes de una serie de literaturas científicas y bases de datos de enfermedades, incluyendo ClinVar (Landrum et al., 2018) y UniProt (Bateman et al., 2021). La patogenicidad de cada variante en sus enfermedades asociadas se evaluó de acuerdo con las recomendaciones de la directriz del ACMG (Richards et al., 2015). En tercer lugar, los fenotipos clínicos de los pacientes se transformaron en los correspondientes términos estandarizados de la ontología del fenotipo humano (Köhler et al., 2021) y se accedió a ellos para medir la similitud (Greene, Richardson, & Turro, 2016; Köhler et al., 2009) con cada una de las ~7.000 enfermedades

genéticas raras (Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). La puntuación de similitud entre el fenotipo de cada paciente y los síntomas asociados a esa enfermedad, causados por variantes priorizadas usando como base aquellas asociadas a desórdenes neurológicos y según las directrices del ACMG, osciló entre 0 y 10. Las variantes de un solo nucleótido se confirmaron mediante secuenciación bidireccional de Sanger.

Reacción en cadena de la polimerasa (PCR) de largo alcance

En el caso del individuo AT-004, se realizó una PCR de largo alcance para la GAA-TR expandida en el gen FXN utilizando el kit Quantabio AccuStart Long Range SuperMix bajo las condiciones especificadas en el protocolo proporcionado dentro del mismo.

Análisis clínicos

Cinco individuos afectados por una neuropatía periférica no diagnosticada (AT-004, AT-010, AT-026, AT-042 y 1071) se sometieron a una evaluación clínica por parte de neurólogo. Se realizó un examen clínico estándar, que incluía una revisión de la historia clínica completa de cada paciente y exámenes físicos para medir los reflejos, el tono muscular, los movimientos oculares, la afectación sensorial y otros signos que ayudaran a establecer una relación fenotipo-genotipo.

ANÁLISIS BIOINFORMÁTICO

Se realizó el análisis de información mutua para la identificación de residuos coevolucionados y correlacionados con otros residuos dentro de las mismas secuencias proteicas en las diferentes superfamilias a las que pertenecen las proteínas asociadas con AOA, siguiendo las especificaciones de Xia et al (2017).

Obtención de datos

Usando el sitio de Centro Nacional para la Información en Biotecnología (Altschul et al., 1997), se obtuvieron las secuencias de las superfamilias proteicas tomando como referencia las secuencias o números de acceso [Q7Z2E3](#) (APTX), [Q7Z333](#) (SETX), [Q8WYR1](#) (PIK3R5) y [Q96T60](#) (PNKP) del *Homo sapiens*, ya que el ser humano es el caso de estudio y el sistema biológico a analizar. Utilizando PSI-BLAST (Altschul et al., 1997), con los parámetros predeterminados, se llevaron a cabo 3 iteraciones. El resultado fue un output compuesto de aproximadamente de 500 secuencias proteicas de cada superfamilia. La redundancia de las secuencias se redujo mediante CD-HIT (programa de agrupamiento por similitud) (Li, Jaroszewski, & Godzik, 2001) utilizando un corte de identidad de secuencia del 95% para mayor astringencia. El resultado obtenido fue un set de trabajo que contenía al menos 150 secuencias de cada superfamilia proteica.

Curado de datos y alineamientos de secuencias

Para disminuir las posibilidades de una sobrerrepresentación de proteínas similares con diferentes números de entrada se utilizó la herramienta MEGA7 (Kumar, Stecher, Tamura, & Dudley, 2016) y se realizó un curado manual de los datos para asegurar la generación de un conjunto de datos de secuencia de alta calidad para la alineación.

Usando el software ClustalW (Thompson, Gibson, & Higgins, 2002) se alinearon cada uno de los conjuntos de datos de trabajo. Para curar aún más los archivos de los alineamientos se realizaron ajustes manuales de acuerdo a los datos obtenidos.

Como último paso de esta fase se llevó a cabo un alineamiento de refinamiento por medio del software MUSCLE (Edgar, 2004).

Análisis de información mutual

Mediante el uso de MISTIC (Servidor de Información Mutua para Inferir Coevolución) se determinó la correlación entre dos posiciones de residuos en los archivos de las múltiples secuencias. La identificación de los pares de residuos que tienen correlaciones evolutivas significativas (valores por encima del umbral de 6.5), fueron definidos por MISTIC.

Los residuos se visualizaron por medio de un gráfico Circos y una red en Cytoscape (incorporados dentro de MISTIC), donde aquellos aminoácidos significativos se reportaron de acuerdo a una escala de colores que representa su conservación evolutiva y su correlación, esto de acuerdo a lo descrito por (Simonetti, Teppa, Chernomoretz, Nielsen, & Marino Buslje, 2013).

Análisis de predicción de estabilidad proteica

Para predecir los posibles efectos causados en la estabilidad de los aminoácidos por los cambios de secuencia, utilizamos el iStable 2.0. Esta herramienta integra 11 predictores diferentes en un único sistema.

Los modelos de iStable 2.0 se basan en clasificadores binarios (0/1) para la estabilidad o la inestabilidad y en un regresor para los valores continuos que utiliza características basadas en la secuencia para calcular la predicción de delta delta G ($\Delta\Delta G$), así como la estabilidad y la inestabilidad después de un cambio de aminoácido (Chen et al., 2020).

Predicción de funcionalidad proteica

Para predecir el potencial patogénico de las alteraciones de la secuencia de ADN encontradas en el paciente AT-042, se utilizó la herramienta de software MutationTaster (Schwarz, et al., 2014). Se analizó la variante presentada en la posición dada en todas las

transcripciones factibles de Ensembl (Howe et al., 2021), utilizando modelos *Random Forest* para las predicciones. El voto del árbol indica cuántos árboles de decisión del bosque aleatorio sugieren una alteración deletérea frente a cuántos sugieren una alteración benigna (Schwarz et al., 2014).

Objetivo General

Determinar la causa genética de las AOAs en pacientes costarricenses por medio de secuenciación, así como relacionar el fenotipo de los pacientes con las características de conservación y correlación de residuos en las superfamilias de las proteínas asociadas con AOA mediante el uso de herramientas bioinformáticas, esto con el fin de contribuir con una clasificación clínica más certera de las AOAs en Costa Rica.

Objetivos Específicos

- Identificar mutaciones en genes involucrados en AOA en pacientes costarricenses mediante los métodos de secuenciación de Sanger y secuenciación de nueva generación, para generar datos sobre las mutaciones causantes de la AOAs en la población del país.
- Establecer la relación fenotipo-genotipo en las familias con AOA en Costa Rica mediante análisis clínicos, con el fin de poder formular recomendaciones posteriores para un diagnóstico eficiente de la causa genética de la enfermedad.
- Identificar residuos coevolutivos conservados en cada una de las proteínas involucradas en los diferentes tipos de AOA, así como las interacciones entre aminoácidos y residuos coevolutivos en las diferentes familias de proteínas en estudio, mediante el análisis de información mutual. Esto con el fin de establecer una asociación con los fenotipos clínicos, y de presentarse el caso, identificar nuevos residuos de interés que compartan patrones de coevolución con los residuos que sean asociados a fenotipos de las AOAs.

Este trabajo sentaría las bases para llevar a cabo el diagnóstico molecular de las neuropatías y ataxias asociadas a este gen, así como en otros genes involucrados en AOA y generaría datos fenotípicos relacionados con AOA en Costa Rica.

Artículo:**ANALYSIS OF ATAXIA IN COSTA RICA THROUGH AN INTEGRATIVE APPROACH**

Hilda Torres-Ulate¹, Sixto Bogantes², Eugene Lee³, Go Hun Seo³, José Guevara-Coto⁴,
Gabriela Chavarría⁵, Alejandro Leal⁵

¹ School of Biology, University of Costa Rica. 11501 San José, Costa Rica.

² Faculty of Medicine, University of Costa Rica. 11501 San José, Costa Rica.

³ 3billion, Inc. 06193, Seoul, Republic of Korea

⁴ School of Computer Science and Informatics, University of Costa Rica. 11501 San José, Costa Rica.

⁵ Section of Genetics and Biotechnology, School of Biology, University of Costa Rica. 11501 San José, Costa Rica.

Abstract

The ataxias are a group of clinically heterogeneous disorders, which may or may not be heritable depending on the etiology. The clinical manifestations of the hereditary ataxias are progressive incoordination of movement and speech (dysarthria), and an unsteady, uncoordinated, broad-based gait. In addition, patients may develop ophthalmoplegia (eye movement limitations), spasticity, neuropathy, and cognitive difficulties. This paper focuses on ataxia with oculomotor apraxia (AOA), which is an autosomal recessively inherited disease. Four types of AOA (AOA1-AOA4) have been described to date. Since information on AOA-causing mutations in Costa Rica is scarce, here we characterize AOA-causing mutations in Costa Rican patients using Sanger sequencing and next-generation sequencing methods. The challenge of determining whether phenotypic manifestations are associated with specific groups of residues that share an underlying relationship has not yet been validated. To address this, we constructed a multiple sequence alignment and analyzed the resulting files with mutual information. In each protein family, Mutual Information (MI) identified pairs of residues with functional importance due to their location within important domains and regions. This work would contribute to characterizing the prevalent mutations in patients suffering from ataxia with oculomotor apraxia in the country, as well as the associated phenotype, and would lay the groundwork for developing working protocols for the molecular diagnosis of the disease in Costa Rica.

Keywords

Ataxia, apraxia, dysarthria, peripheral neuropathy, molecular genetics, sequencing, multiple sequence alignment, mutual information, coevolution.

Introduction

Ataxias are a group of disorders within cerebellar diseases. The first efforts to classify them were made by Gordon Holmes (1908) back in the mid-nineteenth and early twentieth century. These can be clinically and genetically heterogeneous, with slowly

progressive gait incoordination as its main trait. Ataxias can be non-hereditary, when they appear in adulthood (Klockgether, 2010), and are caused by external factors such as chronic alcoholism, vitamin deficiencies, vascular disease, primary or metastatic tumors and paraneoplastic diseases associated with occult carcinoma of the ovary, breast or lung, idiopathic degenerative disease and multiple system atrophy (spinal muscular atrophy) (Jayadev & Bird, 2013). Ataxias can also be hereditary when associated with a causal pathogenic variation in certain genes. These can cause, among other things, cerebellar atrophy, gait and speech problems of varying severity. In addition, individuals may develop eye movement limitations (ophthalmoplegia), spasticity, neuropathy, and cognitive difficulties (Jayadev & Bird, 2013).

For inherited disorders, all three modes of inheritance can be observed (Perlman, 2022). Regarding the autosomal dominant ataxias, there are more than 30 forms known to this date, with a prevalence of 1.2 to 1.9 per hundred thousand inhabitants in countries such as Spain and Brazil, respectively (Ruano et al., 2014). Showing this pattern of inheritance, spinocerebellar ataxias (SCA) are a group of progressive ataxia disorders caused by degeneration of the cerebellum and its afferent and efferent connections (Schöls et al; 2004), linked to a CAG triplet repeat expansion mutation in coding regions of genes (Manto, 2005; Pulst et al., 1996). The prevalence of SCA is estimated to be approximately 1-5 per 100,000 inhabitants (Ruano, et al., 2014). The most common are SCA1, SCA2, SCA3 y SCA6 (Schmitz-Hübsch et al., 2008) with SCA3 being the most common worldwide (Bird, 2019). This may vary in certain regions due to a founder effect, as exemplified by SCA2 in Cuba and SCA10 in Mexico (Manto, 2005).

There are large variations among SCA subtypes (Schöls, Ludger; Bauer, Peter; Schmidt, Thorsten; Schulte Thorsten; Riess, 2004), symptoms of SCA1, SCA2, SCA3, SCA7, SCA8, SCA12, SCA13, SCA17, or SCA25 may begin in the first decade, whereas ataxia may appear after the age of 65 years in SCA6. Regarding SCA17 the penetrance is reduced (Zühlke et al., 2003) while SCA8 shows a complex inheritance pattern with extremes of incomplete penetrance, often with only one or two affected individuals in a family (Manto, 2005).

Symptoms commonly developed by individuals with these conditions are a slowly progressive cerebellar syndrome with various combinations of oculomotor disorders, dysarthria, dysmetria/kinetic tremor and/or ataxic gait. They may also be affected with pigmentary retinopathy, extrapyramidal movement disorders (parkinsonism, dyskinesias, dystonia, chorea), pyramidal signs, cortical symptoms (seizures, cognitive impairment/behavioral symptoms) and peripheral neuropathy (Table 1). An overlap of phenotypes between the different subtypes can occur due to SCAs being genetically heterogeneous. (Manto, 2005).

Regarding ataxias of autosomal recessive inheritance, these account for approximately 3 cases per 100,000 inhabitants, with more than 60 forms being Friedreich's ataxia, ataxia-telangiectasia and oculomotor apraxia being the most common of them (Ruano et al., 2014). Recessive ataxias have a reported prevalence in Spain of 7.2 per 100,000 inhabitants (Ruano et al., 2014). Many of the hereditary ataxias have overlapping presentations and there is a high degree of genetic heterogeneity (Inlora et al., 2017).

Table 1. Clinical presentation of the most common SCAs

Signs and symptoms	SCA 1	SCA2	SCA3	SCA 4	SCA7	SCA8	SCA10	SCA12	SCA1 3	SCA17	SCA18	SCA2 5
Oculomotor disorder	X	X	X		X							
Dysmetria/kinetic tremor						X						
Pigmentary retinopathy					X							
Extrapyramidal movement disorders	X	X	X					X		X		
Pyramidal signs	X	X	X	X	X	X		X	X			
Peripheral neuropathy	X	X	X	X		X		X			X	X
Cognitive disability	X	X	X						X	X		
Seizures							X			X		

Source: Schöls, Ludger; Bauer, Peter; Schmidt, Thorsten; Schulte Thorsten; Riess, 2004.

Within the autosomal recessively inherited disorders, during the last fifteen years the ataxia with oculomotor apraxia (AOA) has been described, this is a group involving axonal sensorimotor neuropathy, and extrapyramidal features (Bras et al., 2015). These were identified as the second most prevalent collection of autosomal recessive ataxias worldwide (Ruano et al., 2014) with an incidence varying depending on the type of AOA. They are caused by mutations in genes responsible for DNA repair, premature protein termination, protein maturation deficiencies, or both. Four types have been described to date: AOA1, AOA2, AOA3 (Tassan et al., 2012) and AOA4/CMT2B2 (Leal et al., 2018). The different types of AOA share many symptoms with each other (Inlora et al., 2017) (Table 2).

AOA1 has an age of onset ranging from infancy to preadolescence (2-12 years) and is caused by mutations in the Aprataxin gen (*APTX*) (Coutinho & Barbot, 2002), which is located on 9p21.1; it has 7 exons and a size of 28,217 bp. (Kent et al., 2002). *APTX* encodes for the protein aprataxin (*APTX*), which is a nuclear protein, present in both the nucleoplasm and nucleolus. It is responsible for catalyzing the nucleophilic release of adenylate groups covalently attached to the 5' phosphate ends at single-strand breaks, resulting in the production of 5' phosphate ends that can be efficiently joined. It is associated with the other DNA repair proteins, playing a role in single-stranded DNA repair through its nucleotide-binding activity and its diadenosine polyphosphate hydrolase activity (MIM 2089200; Prasad et al., 2009).

Németh et al. (2000) first identified a novel locus for a primary autosomal recessive cerebellar ataxia in the *SETX* gene know as AOA2, with an average age of onset in pre-adolescence (12.7 years). (Schiess et al., 2017). *SETX* is located at 9q34.13, has 26 exons and is 93,630 bp in size (Bateman et al., 2021). It encodes for a protein that, because of its homology to the fungal Sen1p protein, was named Senataxin. It has RNA helicase activity encoded by a DNA/RNA helicase domain at the C-terminal end, suggesting that

it may be involved in DNA and RNA processing (Prasad et al., 2009). In association with Rrp45, directs the RNA exosome complex to sites of transcription-induced DNA damage (Richard, Feng, & Manley, 2013b). Contains at its C-terminal end a classic seven-motif domain found in superfamily 1 of the helicase (M. C. Moreira et al., 2004). This domain has strong homology with the human *RENT1* and *IGHMBP2* genes, which are two genes encoding proteins known to have functions in RNA processing (Becherel et al., 2015).

Table 2. Signs presented by the different types of AOA

Sign	AOA type			
	AOA1	AOA2	AOA3	AOA4/CMT2B2
Evolution	Severe	Benign	Benign	Severe
Oculomotor apraxia	X	X	X	X
Dystonia	X	X	Not mentioned	X
Axonal neuropathy	X	X	X	X
Cognitive disability	X	Not mentioned	Not mentioned	X
Cerebellar atrophy	X	X	X	X
Chorea	X	X	Not mentioned	Not mentioned
Dysmetria	Not mentioned	Not mentioned	X	Not mentioned
Dysarthria	X	X	X	X
Saccadic eye movements	X	X	X	X

Source: Coutinho & Barbot, 2002; Tassan et al., 2012; Bras et al., 2015; Szpisjak, Obal, Engelhardt, Vecsei, & Klivenyi, 2016; Inlora et al., 2017; Mariani et al., 2017; Schiess et al., 2017

Tassan et al. (2012) identified AOA3 in a consanguineous family in Saudi Arabia, whose affected members had clinical features similar to those of individuals with AOA2, but with a mean age of onset in adolescence (15.6 years). This type of ataxia is caused by mutations in *PIK3R5*, which is located at 17p13.1. It has 18 exons and a size of 30,869 bp (Kent et al., 2002). This gene encodes regulatory subunit 5 of the phosphatidylinositol 3-kinase gamma class complex (PIK3 γ), which is a 101 kD regulatory subunit of the class I PIK3 γ complex. PIK3R5 is a dimeric enzyme, consisting of a 110 kD catalytic gamma subunit and a 55.87 or 101 kD regulatory subunit that interacts with class 1B phosphoinositide-3-kinase (PI3K) (Uhlen et al., 2015). It acts through high-affinity interaction with G-beta-gamma proteins to recruit the catalytic subunit from the cytosol to the plasma membrane (Uhlen et al., 2015). Phosphatidylinositol 3-kinase (PI3K) is a member of a unique and highly conserved family of intracellular kinases that phosphorylate the 3'-hydroxyl group of phosphatidylinositol and phosphoinositides. This reaction results in the activation of multiple intracellular signaling pathways that regulate diverse and important functions, such as cellular metabolism, survival and polarity, and vesicle trafficking (Engelman et al., 2006).

Within the rare group of cerebellar ataxias to which AOA belongs, patients with AOA associated to mutations in *PNKP* have an average age of disease onset in childhood (4.3 years) (Schiess et al., 2017) and in some cases is characterized by cerebellar ataxia, oculomotor apraxia, polyneuropathy, and cerebellar atrophy at MRI. Such phenotype was

classified as ataxia with oculomotor apraxia type 4 (AOA4) (Gatti et al., 2019). In other cases, the presenting clinical phenotype was characterized by the presence of a motor and sensory axonal polyneuropathy, resembling a form of Charcot–Marie–Tooth disease and identified as CMT2B2 (Leal et al., 2018; Pedroso et al., 2015). In these patients, slurred speech and cerebellar atrophy at brain MRI were also described (Leal et al., 2018).

PNKP is located at 19q13.33, has 16 exons and is 6,337 bp in size and encodes for a polynucleotide kinase 3' phosphatase, which in response to ionizing radiation or oxidative damage, catalyzes the 5' phosphorylation of nucleic acids and also has an associated 3' phosphatase activity, which predicts an important role in DNA repair after ionizing radiation or oxidative damage (Gatti et al., 2019; Leal et al., 2018; Uhlen et al., 2015).

In view of the foregoing on the different phenotypes that can be identified in the same disorder, to establish the diagnosis of hereditary ataxia, the following are recommended: detecting typical clinical signs on neurological examination, including poorly coordinated hand movements, dysarthria, eye movement abnormalities, identification of distinctive features on magnetic resonance imaging or computed tomography scans (Klockgether, 2010; Perlman, 2022), documentation of the possible hereditary nature of the disease by developing a family history searching for cases of relatives with ataxia, identifying a mutation causing ataxia or recognizing a clinical phenotype characteristic of a genetic form of ataxia (Jayadev & Bird, 2013; Perlman, 2022). If a genetic cause cannot be identified, it is necessary to proceed with the exclusion of non-genetic causes of ataxia by means of CT scans, magnetic resonance imaging and detection of multiple antibodies (Klockgether, 2010).

Since the challenge of determining whether phenotypic manifestations are associated with specific groups of residues that share an underlying relationship, the possible effect of mutations in the development of disease phenotypes is a focus of interest and represents an opportunity to study how different phenotypic manifestations arise.

As a bioinformatics approach in the study of proteins and their underlying relationships, mutual information can be used to estimate the extent of the mutual coevolutionary relationship between two positions in a protein family and it can be applied to predict positional correlations in a multiple sequence alignment to make possible the analysis of those positions structurally or functionally important in a given fold or protein family (Dunn, Wahl, Gloor; 2008), such as the ones associated with Ataxia with Oculomotor Apraxia (AOA). An integrative approach based on analyzing full-length sequence alignments from proteins and the subsequent identification of protein sectors using statistical analysis has been proposed in previous studies (Guevara Coto, Schwartz, & Wang, 2014; Halabi et al, 2009; McLaughlin et al, 2012; Morcos et al., 2011) aiming to cluster amino acid residues based on an underlying interaction, such as co-evolution, to understand how mutations in these related sites can lead to different disease phenotypes.

Given that AOAs have not been studied in Costa Rica this project aimed to molecularly characterize the variants associated with ataxia in the population of the country, as well as to study the relationship between these variants and the protein conformation they produce, with the phenotype of patients.

Methodology

Ethics Committee

This research is part of the Project “Determination of genes related to peripheral neuropathies and a group of ataxias” (No. 111-B8-37), registered with the Vice Rector's Office for Research and has been approved by the Ethics Committee.

Recruitment process

Neurologists registered in the main hospitals of the country were contacted and informed about the project so that, in the event of a suspected case of AOA, they could indicate to the patient that if they wished, they could contact the research team.

In addition to this recruitment, the Ethics Committee was asked for permission to access to samples from the Institute of Health Research (INISA) of the University of Costa Rica, which were part of another study on ataxias carried out in that institution.

Neurologists working in Costa Rica, invited patients to establishing contact with our research group. After receiving a detailed explanation about the project, if the individuals agreed to participate in the study, they proceeded to sign an informed consent form and a code was assigned to ensure the anonymity of the samples. Individuals with ataxia or probable ataxia phenotypic were taken into account. Saliva or blood samples were collected.

Variants detection

We obtained DNA samples of 48 individuals who showed phenotypic characteristics that corresponded to those expected in ataxia and did not have a definitive clinical diagnosis provided by a neurologist. Of these, 28 samples met the quality criteria established to allow them to continue to be processed, so all exon regions of all human genes were captured by xGen Exome Research Panel v2 (Integrated DNA Technologies, Coralville, Iowa, USA). The captured regions of the genome were sequenced with Novaseq 6000 (Illumina, San Diego, CA, USA). Analysis of the raw genome sequencing data, including alignment with the GRCh37/hg19 human reference genome, variant calling and annotation, was performed with open-source bioinformatics tools and 3Billion in-house software. The automatic variant interpretation software, EVIDENCE has three main steps; variant filtering, ranking, and similarity scoring for patient phenotype (Seo et al., 2020). First, gnomAD (Karczewski et al., 2020) as a population genome database and the 3Billion genome database were used to estimate allele frequency.

Filtering and Validation

Common variants with an allele frequency less than >5% were filtered according to BA1 of the ACMG guideline (Richards et al., 2015). Second, evidence data on variant pathogenicity were extracted from a range of scientific literatures and disease databases, including ClinVar (Landrum et al., 2018) and UniProt (Bateman et al., 2021). The pathogenicity of each variant in its associated diseases was assessed according to the recommendations of the ACMG guideline (Richards et al., 2015). Third, patients' clinical phenotypes were transformed into corresponding standardized human phenotype ontology terms (Köhler et al., 2021) and accessed to measure similarity (Greene et al., 2016; Köhler et al., 2009) to each of the ~7,000 rare genetic diseases rare genetic diseases

(Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). The similarity score between each patient's phenotype and the symptoms associated with that disease caused by the prioritized variants using as a starting point those associated with neurological diseases and according to ACMG guidelines ranged from 0 to 10. The single nucleotide variants and all indels were confirmed by bidirectional Sanger sequencing.

Long range polymerase chain reaction (PCR)

In the case of individual AT-004, a long PCR for expanded GAA-TR in the *FXN* gene was performed using Quantabio AccuStart Long Range SuperMix Kit under the conditions specified in the protocol provided in the package.

Clinical analysis

Because some samples belonged to a previous study, these individuals could not be located for clinical evaluation, thus only five individuals affected with an undiagnosed peripheral neuropathy (AT-004, AT-010, AT-026, AT-042 and 1071) underwent clinical analysis. A standard clinical examination was performed, which included a review of the complete clinical history of each patient and physical examinations to measure reflexes, muscle tone, eye movements, sensory involvement and other signs that would help to establish a phenotype-genotype relationship.

Data Acquisition

For the proteins associated to AOA, the sequences of the protein superfamilies were obtained from the National Center for Biotechnology Information (Altschul et al., 1997) using as reference the sequences or accession numbers [Q7Z2E3](#) (APTX), [Q7Z333](#) (SETX), [Q8WYR1](#) (PIK3R5) and [Q96T60](#) (PNKP) from *Homo sapiens*, since the human being is the case study and the biological system to be analyzed. Using PSI-BLAST (Altschul et al., 1997), with the predetermined parameters, 3 iterations were carried out. The result was an output composed of approximately 500 protein sequences from each superfamily. The redundancy of the sequences was reduced by CD-HIT (similarity clustering program). (Li et al., 2001) using a 95% sequence identity cutoff for astringency. The result obtained was a working set containing at least 150 sequences from each protein superfamily.

Data curation and sequence alignments

To reduce the chances of overrepresentation of similar proteins with different entry numbers, the MEGA7 tool was used. (Kumar et al., 2016) and manual data curation was performed to ensure the generation of a high quality sequence dataset for alignment. Using ClustalW (Thompson et al., 2002) each of the working data sets were aligned. To further curate the alignment files, manual adjustments were made according to the data obtained. As the last step of this phase, a refinement alignment was carried out using MUSCLE (Edgar, 2004).

Mutual information

Using Mutual Information Server to Infer Coevolution (MISTIC), the correlation between two residue positions in the multiple sequence files was determined. The identification of pairs of residues that have significant evolutionary correlations (values above the threshold of 6.5) were defined by MISTIC. Residues were visualized by means of a Circos plot and a network in Cytoscape (incorporated within MISTIC), where those significant amino acids were reported according to a color scale representing their evolutionary conservation and correlation, this according to those described by (Simonetti et al., 2013).

Protein stability prediction analysis

To predict the possible effects caused in amino acid stability caused by sequence changes, we used the iStable 2.0. The tool integrates 11 different predictors in a single system. The iStable 2.0 models are based on binary classifiers (0/1) for stability or instability and a regressor for continuous values that use sequence based features to calculate the to predict delta delta G (ddG) as well as stability and instability after an amino acid change (Chen et al., 2020). (Fig 1).

Functionality prediction analysis

To predict the pathogenic potential of DNA sequence alterations found in patient AT-042, the software tool MutationTaster (Schwarz, et al., 2014) was used. It analyzed the submitted variant at the given position in all feasible Ensembl transcripts (Howe et al., 2021), using Random Forest models for predictions. Tree vote indicates how many decision trees of the Random Forest are suggestive of deleteriousness vs. how many are suggestive of a benign alteration (Schwarz et al., 2014).

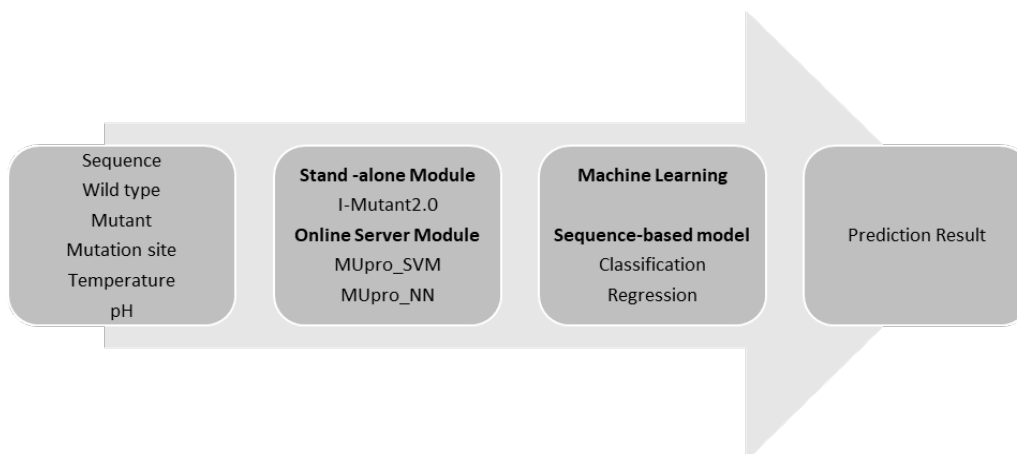


Figure 1. System architecture with the key points of the process for protein stability prediction.

Results

Phenotypic characterization

All the examined patients have as a common factor the instability (ataxia), although they can be differentiated into two groups (Table 3):

The first group includes individuals AT-004 and AT-026 where the ataxia is the most disabling clinical sign, with mild or absent signs of motor neuropathy, although they have phenotypically distinct forms of ataxia. AT-004 presents a sensory ataxia (neurophysiology consistent with this modality) associated with pyramidal signs, without cerebellar involvement in neuroimaging and with non-motor manifestations of her disease suggesting that it is a systemic condition where the cause of the ataxia is not a primary cerebellar disease. On the other hand, individual AT-026 presents a form of late-onset ataxia with pancerebellar symptoms and very important atrophic changes in his cerebellum evidenced in the neuroimaging suggesting a primary cerebellar disease as the cause of ataxia.

The second group includes the other three individuals who are the most disabled ones, and where the functional limitation that has led to the loss of walking is associated with a clear peripheral motor nerve disease. All of them have lost walking autonomy and have trophic changes in the lower extremities (distal muscle atrophy). These individuals present extrapyramidal signs of variable severity (in AT-042 and AT-010 are evident while in 1071 are mild). The signs of cerebellar dysfunction are mild. In this group AT-042 has a family history of three uncles suffering from a neurodegenerative disease with neuropathy and ataxia, one of them died at the age of 24. Otherwise, AT-010 was first diagnosed with cerebral palsy since he presented perinatal insult (hypoglycemia), his severe disability can be explained by this reason. Individual 1071 has normal cognitive function with a university education but is currently retired due to the severity of peripheral motor involvement, she also has subtle signs of brainstem dysfunction.

Table 3. Clinical evaluation result for the evaluated patients.

	Group 1			Group 2	
	(AT-004)	(AT- 026)	AT-010	(1071)	(AT- 042)
Gender	Female	Male	Male	Female	Male
Age	22	70	26	44	21
Decade of age at onset	First	Third	First	Third	First
First sign	Syncope	Polyneuropathy	Polyneuropathy	Polyneuropathy	Choreathetosis
More prominent sign	Ataxia	Polyneuropathy	Polyneuropathy	Polyneuropathy	Polyneuropathy
OMA	-	+	++	+/-	-
Slurred speech	+	+	+++	+	+
Mobility	Mobile, severe gait ataxia	Mobile, mild gait ataxia	Wheelchair	Wheelchair	Wheelchair
Dystonia	-	-	+	-	++
Cognitive impairment	-	-	Mild	-	Mild
UE Muscle strength (I/D)	5/5	4/3	4/1	3/0	5/5
LE Muscle strength (P/I/D)	5/5/5	5/3/0	1/0/0	5/0/0	4/0/0
Reflexes (UE/Knee/Ankle)	2/0/0	2/3/3	2/0/0	1/0/0	2/0/0
UE Sensory involvement (T/Pa/V/Po)	1/1/1/1	2/1/0/0	0/0/0/0	2/1/0/0	2/2/1/1
LE Sensory involvement (T/Pa/V/Po)	1/1/1/0	1/0/0/0	0/0/0/0	0/0/0/0	1/1/1/1
Atrophy	None	Intrinsic muscles of hands and feet, calves	Anterior and posterior compartments of forearm, intrinsic muscles of hands and feet, calves	Intrinsic muscles of hands and feet, calves	Calves
Others	Marcus Gunn sign (relative afferent pupillary defect)	REM sleep behavior disorder	Lymphedema	Lymphedema, right hemifacial spasms	Three uncles affected with clinically similar disease
Deformities	Left ventricular hypertrophy	Claw hand, pes cavus and hammertoes	Claw and dropped hand, scoliosis	Claw hand	Head hyperextension, scoliosis
Pyramidal signs	Babinski	-	-	-	-
Obesity	++	-	-	-	-
MRI findings	Normal	CA, no WM abnormalities	CA, no WM abnormalities	CA, no WM abnormalities	CA, periventricular gliosis
Electrophysiological data	Axonal sensory neuropathy	Normal	-	-	Severe sensorimotor neuropathy

OMA, oculomotor apraxia. UE, upper extremity; LE, lower extremity; P, proximal (knee extensor or flexor); I, intermediate (hand extensor or flexor (UE), foot extensor or flexor (LE)); D, distal (intrinsic hand muscles (UE), intrinsic foot muscles (LE)), Motor scale: 5, normal; 4, mild weakness; 3, ability to lift against gravity; 2, not able to lift against gravity; but movement visible, 1, no movement, but tendon contraction visible; 0, complete paralysis, Reflexes/sense of vibration or position: 2, normal; 1, reduced; 0, absent; T, touch; Pa, pain; V, vibration; Po, position, Sensory involvement: 2, normal; 1, mildly reduced: distally to wrist level (UE) or malleoli level (LE), 0, severely reduced: distally to elbow level (UE) or knee level (LE). CA, cerebellar atrophy. WM, white matter.

Molecular characterization

Five variants associated with peripheral neuropathies causing disorders were identified in 9 patients (Table 4).

Table 4. Identified variants associated with peripheral neuropathies

Sample	Gen	Variant	Presentation	Disorder
AT-026	<i>KCNC3</i>	19-50826951-C-T NM_004977.3:c.1259G>A (NP_004968.2:p.Arg420His)	Heterozygous	Spinocerebellar ataxia 13
AT-038	<i>PMPCA</i>	9-139313299-G-A NM_015160.2:c.1129G>A (NP_055975.1:p.Ala377Thr)	Homozygous	Spinocerebellar ataxia, autosomal recessive 2 (SCAR2)
AT-004	<i>FXN</i>	Two expanded alleles of 908 and 1116 triplets in intron 1	Homozygous	Friedreich ataxia
AT-021	<i>PNKP</i>	19-50364522-G-A NM_007254.3:c.1549C>T (NP_009185.2:p.Gln517Ter)	Homozygous	AOA4/CMT2B
1207	<i>PNKP</i>	19-50364522-G-A NM_007254.3:c.1549C>T (NP_009185.2:p.Gln517Ter)	Homozygous	AOA4/CMT2B
AT-010	<i>PNKP</i>	19-50365103-CGTG-C NM_007254.3:c.1221_1223del (NP_009185.2:p.Thr408del)	Homozygous	AOA4
AT-040	<i>PNKP</i>	19-50365103-CGTG-C NM_007254.3:c.1221_1223del (NP_009185.2:p.Thr408del)	Homozygous	AOA4/CMT2B
AT-046	<i>PNKP</i>	19-50365103-CGTG-C NM_007254.3:c.1221_1223del (NP_009185.2:p.Thr408del)	Homozygous	AOA4/CMT2B
1071	<i>PNKP</i>	19-50365103-CGTG-C NM_007254.3:c.1221_1223del (NP_009185.2:p.Thr408del)	Compound Heterozygous	CMT2B
AT-042	<i>PNKP</i>	19-50364522-G-A NM_007254.3:c.1549C>T (NP_009185.2:p.Gln517Ter) 19-50365103-CGTG-C NM_007254.3:c.1221_1223del (NP_009185.2:p.Thr408del) 19-50365031-G-GGC NM_007254.4:c.1294_1295dup (NP_009185.2:p.Arg433ProfsTer35)	Compound Heterozygous	AOA4

The variant linked to Spinocerebellar Ataxia 13 (SCA13) in both a French and a Filipino families (M. F. Waters et al., 2005) was identified in one patient (AT-026): a transition in exon 2 c.G1259A in the potassium channel, voltage-gated, shaw-related subfamily, member 3 gen (*KCNC3*) resulting in a substitution of an amino acids that is 100% conserved among members of the human KCNC family (M. F. Waters et al., 2006). *KCNC3* is located within the linkage interval on chromosome 19q13.33, the voltage-gated potassium channel plays an important role in the rapid repolarization of fast-firing brain neurons. The channel displays rapid activation and inactivation kinetics. It is involved in the regulation of the frequency, shape and duration of action potentials in

Purkinje cells and is required for normal survival of cerebellar neurons, therefore it is necessary for normal motor function (Bateman et al., 2021).

Another variant identified in a patient (AT-038) is p.Ala377Thr, it was associated to Spinocerebellar ataxia, autosomal recessive 2 (SCAR2) in three families of Christian Lebanese Maronite origin (Jobling et al., 2015). It is caused by a missense mutation in exon 10, c.G1129A in the mitochondrial processing peptidase-alpha gene (*PMPCA*), leading to defective substrate recognition and binding leading to inadequate enzyme activity within the mitochondria (Jobling et al., 2015). *PMPCA* cytogenetic location is 9q34.3 (Bateman et al., 2021). It encodes the alpha subunit of the mitochondrial processing peptidase (MPP), which cleaves the targeting peptide of nuclear-encoded mitochondrial precursor proteins upon their import into mitochondria (Choquet et al., 2016)

Lastly, for seven patients, three variants were identified in *PNKP*, that is also located on chromosome 19q13.33 and contains a kinase and a phosphates domain that are both involved in DNA binding (Bateman et al., 2021) through its two catalytic activities which ensure that DNA termini are compatible with extension and ligation by either removing 3'-phosphates from, or by phosphorylating 5'-hydroxyl groups on, the ribose sugar of the DNA backbone (Bateman et al., 2021). Two of those variants were already associated with AOA4 (Gatti et al., 2019).

A transition c.C1549T in exon 17 of the polynucleotide kinase 3'-phosphatase (*PNKP*) gene. This variant causes a nonsense mutation (p.Gln517ter) predicted to truncate the last five amino acids (Leal et al., 2018). The second variant identified consists of a three-base deletion in exon 14, c.1221_1223del, this variant results in the deletion of residue Thr408 (Thr408del) (MIM 616267). This variant was found in a homozygous presentation in three patients.

Two patients were found to be compound heterozygous for variants in the *PNKP* gene. The first one presented the two variants already described in Costa Rica, p.Gln517ter and Thr408del. For the second patient, the variants found were Thr408del and Arg433Ter, a variant that has not yet been described and is predicted to cause a premature termination in the kinase region of the protein. It represents a nonsense mutation in exon 14, c.1294_1295dup. (Arg433ProfsTer35).

Mutual information

Mutual information (MI) identified functionally important positions known to cause AOA in three of the superfamilies. For APTX three positions were linked, for PIK3R5 one position was recognized and for PNKP two position were identified. In the case of SETX, six positions linked to genetically related disorders with AOA2 were found (Moreira & Koenig, 2018), four positions known to cause spinocerebellar ataxia with axonal neuropathy (SCAN2) and one position associated with Amyotrophic lateral sclerosis 4 (ALS4) respectively were identified as well (Bateman et al., 2021)(See table 5).

Table 1. Functional important positions known to cause AOA in each superfamily.

APTX	SETX	PIK3R5	PNKP/CMT2B
206P, 247L and 279W	305W, 413P, 496P and 2213P (SCAN2) 1554C (ALS4)	629P	375G, 399L, 409C, 424T, 429A, 439R, 442G, 462R, 465E, 515Y and 517Q

Additionally, potential candidate positions that could give rise to disease phenotypes if mutated were found in all the super families (Table 2).

Table 2. Potential candidates that could cause AOA if mutated for each protein superfamily.

APTX	SETX	PIK3R5	PNKP/CMT2B
254P, 260H, 262H, 319C and 322C	4C, 6W, 7C, 38C, 40C, 43C, 55P, 62W, 113P, 121P, 133C, 138C, 154P, 163P, 170W, 195C, 208P, 225P, 238W, 242C, 285P, 287W, 288P, 292C, 305W, 311P, 342P, 352C, 370W, 375C, 376P, 379C, 380P, 410W, 451C, 474C, 476H, 479W, 485W, 492C, 524H, 535Y, 546G, 550G, 555C, 568G, 573G, 574W, 587C, 612C, 788C, 976F, 977P, 1093W 1266P, 1270P, 1271P, 1278P, 1280P, 1293P, 1318G, 1345S, 1497F, 1503P, 1509C, 1549C, 1554C, 1565C, 1568H, 1594P, 1599F, 1622P, 1708W, 1720G, 1721P, 1722P, 1734P, 1737F, 1749P, 1763W, 1804Y, 1805P, 1877C, 1909P, 1914F, 1915C, 1916T, 1953P, 1959C, 1962H, 1963G, 1964P, 2038C, 2039, 2047G, 2067H, 2103G, 2152C, 2153C, 2159G, 2160G, 2174P, 2177C, 2187C, 2194P, 2199C, 2208P, 2212P, 2250C, 2261H, 2262P, 2265C, 2267F, 2268P, 2288C, 2292W, 2293P, 2294F, 2296P, 2384C, 2389C, 2434W, 2476P, 2481P, 2491P and 2578P	8C, 33W, 42W, 80P, 115W, 116P, 118P, 120C, 137P, 284P, 293W, 504P, 545R, 546P, 560P, 616P, 617W 634C, 663C, 726P, 744W, 750W, 760C, 818C, 833C, 838C, 865C, 869C and 871P	11W, 18G, 24L, 25P, 27D, 28G, 33L, 34G, 35R, 36G, 37P, 41V, 43D, 46C, 48R, 68G, 70N, 71P, 79L, 82G, 90G, 97N, 98G, 100H, 101P, 147W, 156F, 160G, 163P, 169G, 170F, 171D, 173D, 174G, 182G, 185F, 186P, 189P, 191D, 192W, 196Y, 197P, 200P, 210G, 211Y, 216F, 221S, 225G, 232F, 246P, 256G, 261P, 264G, 266W, 270Q, 286F, 288G, 289D, 292G, 294P, 297W, 308C, 321F, 324P, 331W, 336F, 339P, 355P, 374P, 402W, 405C, 409C, 426P, 433R, 434Y, 437C, 443V, 444F, 447C, 459H, 471H, 489P, 494G, 505L and 518F

Protein Stability

For the three variants found in PNKP, prediction scenarios were run for all possible amino acid changes. Regression models were used to determine the consequences of variations for the protein stability (Table 7) using the changes in ddG values and translating them into a binary results.

Table 7. Results of the predictions made for each simulated residue variation.

Protein Stability	Variant		
	Gln517	Arg433	T408del
Decreased	G, A, V, N, I, K, M, R, D, H, E, F, W, Y, S, T, C and P	G, A, V, N, I, K, M, R, D, H, L, E, F, W, Y, S, T, C and P	G, A, V, N, I, K, M, R, D, H, L C, P and Q
Increased	L	None	E, F, W, Y, S and T

Prediction of protein functionality

Since the variant NM_007254.3: c.1294_1295dup found in patient AT-042 has not been previously described in literature, a prediction of protein functionality was performed. It was anticipated that around 55 amino acids will be lost corresponding to the protein's kinase region. Therefore, the variant was predicted to be deleterious with a Tree vote: 190|10. The number before the vertical bar (|) always represents deleterious predictions and the number after the vertical bar benign predictions (Schwarz et al., 2014).

Discussion

In this work, we clinically described five Costa Rican patients presenting signs of three different neuropathies. Through molecular analysis we found five variants associated with ataxia in the samples analyzed.

Disease associated mutations in *KCNC3* and *PMPCA* were identified in two patients, AT-026 and AT-038. Variant p.Arg420His in *KCNC3* is linked to Spinocerebellar ataxia type 13 (SCA13), a rare sub-type of spinocerebellar ataxias (Michael F. Waters et al., 2006). This protein is required for normal survival of cerebellar neurons because it plays a role in the regulation of the frequency, shape and duration of action potentials in Purkinje cells (Bateman et al., 2021). While the known pathogenic variants in *KCNC3* are associated with different phenotypes, data to date are too limited to make any genotype-phenotype correlations (M.F. Waters, 2020). Nonetheless p.Arg420His is predicted to cause loss of channel activity; decreased in protein abundance and protein stability; reduced localization to the plasma membrane and impaired N-glycosylation (Zhao, Zhu, & Thornhill, 2013).

Variant p.Ala377Thr, found in another patient (AT-042), is associated to Spinocerebellar ataxia, autosomal recessive 2, the first pure cerebellar ataxia to be described in which the inheritance is autosomal recessive (Rosenberg & Khemani, 2015). Jobling et al., 2015 found that this particular variant leads to more abnormalities of mitochondrial function and impacts the maturation process of frataxin (FXN), the protein which is depleted in Friedreich ataxia (FRDA).

For patient AT-004, it was possible to perform additional testing in *FXN*, thus lead to a diagnosis of Friedreich ataxia, the most common of the inherited ataxias (Delatycki, Williamson, & Forrest, 2000) and caused by a GAA repeat, generally in intron one. *FXN*

is a nuclear-encoded mitochondrial iron chaperone involved in iron-sulfur biogenesis and heme biosynthesis (MIM 229300). Different studies have shown that the size of the GAA repeat length in each allele is important in predicting the age of onset and some features of FRDA and that there is a far greater contribution from the smaller than larger allele to disease parameters and complications (Delatycki et al., 2000).

Regarding *PNKP*, more than 40 individuals carrying *PNKP* gene mutations have been reported around the world so far. We identified both homozygous and compound heterozygous mutations in seven of the 28 patients whose samples were analyzed. This results agree with the previously variants linked to AOA4/CMT2B reported in Costa Rica by Leal et al., 2018, p.Gln517Ter and p.Thr408del, leading us to suggest the possibility of a founder effect to explain why only these variants, with the exception of AT-042, were found.

Additionally to both p.Gln517Ter and p.Thr408del, a variant that has not been previously described was found, p.Arg433ProfsTer35. The pathogenicity of this sequence alterations could not be verified through molecular approaches since additional family members were not available for segregation analysis.

Nevertheless, this mutation could be causative for the disease phenotype due to the suspected production of truncated, partially or fully functional protein, thus contributing to the protein deficiency that is the distinguishing feature of many recessive genetic disorders (Nickless, Bailis, & You, 2017). With this in mind, different predictive tests were carried out using bioinformatics methods that provide information on the degree of conservation, stability and protein functionality.

Multiple sequence alignments were analyzed to identify coevolving residues using a method that involves Mutual Information (MI), this was based on its ability to identify underlying relationships such as coevolution between pairs of residues (Moreno-Brid & Ruiz-Nápoles, 2009) (De Juan et al., 2013). In the four super families of proteins associated to AOA, functionally important residues already known to cause each of the existing types were inferred among the highly conserved positions, as well as potential candidate sites potential candidates that, due to their degree of conservation, could cause phenotypes similar to those currently described if they were to mutate. One of the candidate sites identified through MI is 433R, the same mutated residue that was identified in patient AT-042, this finding can serve as a real proof of concept observed in a patient.

Although not all identified sites that are highly conserved are disease-causing, consistency between predictions and observed reality is detected and worth of following up in other types of studies with a larger cohort. A possible explanation to understand why some of the predicted conserved residues may not have been associated with disease yet is that they may be so pathogenic that the change becomes incompatible with life.

The candidate sites identified in the super families represent positions of interest that would require validation through clinical reports or experimental data. Our results suggest that MI may be a novel approach for the identification of new disease candidate sites, while providing valuable insight into how underlying amino acid relationships may shape their role in the occurrence of different phenotypic manifestations.

Genetic variations, such as hydrogen bonding networks, conformational dynamics, protein activity and protein interaction networks, particularly at the level of functional assemblies, can have dramatic effects on protein stability and constitute one of the main molecular mechanisms underlying several mutation-induced diseases (Sanavia et al., 2020). Although the protein stability prediction tool used in this study performed the tests presented by calculating the effects of amino acid substitution and not with early termination, the result was a loss of stability in most scenarios for all three variants found in *PNKP*.

Since it has been demonstrated that protein stability is a major factor contributing to monogenic disease (Yue, Li, & Moulton, 2005), these results can be used as a basis for inferring that an early termination mutation would have a similar effect to the ones reviewed, if the mutation has any effect on the stability at the $\Delta\Delta G$ level, either increased or decreased, resulting in a truncated protein, that would likely cause the protein to be inelastic or denatured and the checkpoints responsible for detecting these defects would proceed to discard it.

However, to assess the detrimental effect of the variants found, protein stability is necessary but not sufficient to infer protein function, since proteins are not necessarily optimized to maximize their stability (Chen et al., 2020). Therefore, it was considered crucial to understanding how the unknown will functionally impact important sites. Functionality prediction analysis for p.Arg433ProfsTer35 indicated that the resulting amino acid sequence change will cause the loss of more than 10% of the protein structure (kinase domain) leading to increased DNA damage with subsequent cell death due to the loss of the catalytic function that enables the protein to transfer phosphate molecules usually from ATP to other substrates (Bateman et al., 2021).

Taking into account that the resulting prediction points to such severity of the new variant, a possible explanation for why the AT-042 phenotype is less severe than expected goes hand in hand with the proposal by Bermúdez-Guzmán, et al., 2020 of mutational survivorship bias. This suggests a higher tolerance in the kinase domain, which groups most of the deleterious variants described to date, as opposed to the phosphatase domain. It is supported by previous studies proposing that the phosphatase domain is functionally more important and necessary for DNA repair and therefore there are fewer variants described as they may cause individuals to be non-viable. This theory could provide an explanation as to why in the present case of an individual with compound heterozygous condition (AT-042) in which the variants reported in the kinase domain are deleterious and even one allele is completely unusable, the enzymatic capacity of the protein is only diminished, and this allows some level of functionality to be expressed.

The p.Thr408del is a known variant, present in homozygous form for AT-010 and as a compound heterozygous for 1071 and AT-042, the T408 has a location within the kinase region so its deletion probably has consequences for protein conformation and function (Bras et al., 2015). The second variant found in 1071, p.Gln517Ter, causes the loss of those amino acids of the enzyme, that play a role in the stabilization of the protein, anchoring the kinase domain to the phosphatase domain (Leal et al., 2018).

Residues located from positions 402–521 of *PNKP* are coevolved and considerably conserved among the respective superfamily, this suggest that the C-terminal tail, as well as its interactions are critical for the protein's function which in response to ionizing

radiation or oxidative damage catalyzes 5' phosphorylation of nucleic acids and also has an associated 3' phosphatase activity, predicting an important role in DNA repair after ionizing radiation or oxidative damage, multiple pathways involved in DNA-damage repair, including single-strand breaks (SSBs) and double-strand breaks (DSB) (Bras et al., 2015; Uhlen et al., 2015).

In this way, the molecular involvement of PNKP in the different pathways mentioned above, might contribute to the disease phenotypes associated with the deletion (p.Thr408del), the premature termination (p.Gln517Ter), as well as the potential candidate (Arg433ProfsTer) since this defects can result in protein alterations leading to damaged DNA, mainly due to oxidative stress in the nervous system, with ensuing interference with transcription, reduced enzyme activity and, finally, cell death.

In the case of the 19 patients in whom no mutations were found, through whole exome sequencing, our hypothesis is related to the sequencing method used, since there are regions and genetic variants that cannot be technically covered by whole exome sequencing method such as structural chromosomal aberrations, trinucleotide repeat expansion, epigenetic factors, variants in genes with corresponding pseudogenes or other highly homologous sequences, and noncoding regions including untranslated regions, introns, and intergenic regions. For this patients, further testing will be required to reach a diagnosis.

In conclusion, this work lays the groundwork for molecular diagnosis of neuropathies and ataxias associated with this gene, as well as in other genes involved in peripheral neuropathies, and would generate phenotypic data related to ataxias in Costa Rica. To the best of our knowledge, this is the first time that homozygous p.Thr408del patients are reported in Costa Rica, and these findings along to those by Leal et al, (2018) suggest that mutations in *PNKP* are the most frequent cause of AOA in the country.

Our investigation also provides an approach to prioritize mutations discovered in large-scale sequencing projects which serves to validate these bioinformatics tools and recognize their predictive method in indicating sites of interest.

Acknowledgments

We would like to thank the patients and families for their participation in this study. We would also like to thank M.Sc. Melissa Vásquez Cerdas for her collaborations in regards to the samples located in the INISA, MBBS, PhD. Sanjay I. Bidichandani, CHF Claire Gordon Duncan Chair in Genetics, University of Oklahoma College of Medicine for performing the test for FRDA diagnosis, Lidia Benitez and Marcelo Castro-Alpizar for their valuable input while performing tests and writing this paper.

References

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>

- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., ... Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- Becherel, O. J., Sun, J., Yeo, A. J., Nayler, S., Fogel, B. L., Gao, F., ... Lavin, M. F. (2015). A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. *Human Molecular Genetics*, *24*(20), 5759–5774. <https://doi.org/10.1093/hmg/ddv296>
- Bermúdez-Guzmán, L., Jimenez-Huezo, G., Arguedas, A., & Leal, A. (2020). Mutational survivorship bias: The case of PNKP. *PLoS ONE*, *15*(12 December), 1–25. <https://doi.org/10.1371/journal.pone.0237682>
- Bird, T. D. (2019). Hereditary Ataxia Overview. Retrieved from GeneReviews™ website: <https://www.ncbi.nlm.nih.gov/books/NBK1138/>
- Bras, J., Alonso, I., Barbot, C., Costa, M. M., Darwent, L., Orme, T., ... Guerreiro, R. (2015). Mutations in PNKP cause recessive ataxia with oculomotor apraxia type 4. *American Journal of Human Genetics*, *96*(3), 474–479. <https://doi.org/10.1016/j.ajhg.2015.01.005>
- Brenner, C. (2002). Hint, Fhit, and GalT: Function, structure, evolution, and mechanism of three branches of the histidine triad superfamily of nucleotide hydrolases and transferases. *Biochemistry*, *41*(29), 9003–9014. <https://doi.org/10.1021/bi025942q>
- Brenner, C., Biecanowski, P., Pace, H. C., & Huebner, K. (1999). The histidine triad superfamily of nucleotide-binding proteins. *Journal of Cellular Physiology*, *181*(2), 179–187. [https://doi.org/10.1002/\(SICI\)1097-4652\(199911\)181:2<179::AID-JCP1>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-4652(199911)181:2<179::AID-JCP1>3.0.CO;2-8)
- Chen, C. W., Lin, M. H., Liao, C. C., Chang, H. P., & Chu, Y. W. (2020). iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules. *Computational and Structural Biotechnology Journal*, *18*, 622–630. <https://doi.org/10.1016/j.csbj.2020.02.021>
- Choquet, K., Zurita-Rendón, O., La Piana, R., Yang, S., Dicaire, M. J., Boycott, K. M., ... Tétéreault, M. (2016). Autosomal recessive cerebellar ataxia caused by a homozygous mutation in PMPCA. *Brain*, *139*(3), e19. <https://doi.org/10.1093/brain/awv362>
- Coutinho, P., & Barbot, C. (2002). Ataxia with Oculomotor Apraxia Type 1. *GeneReviews™*, 1–12. <https://doi.org/NBK1456> [bookaccession]
- De Juan, D., Pazos, F., & Valencia, A. (2013). Emerging methods in protein co-evolution. *Nature Reviews Genetics*, *14*(4), 249–261. <https://doi.org/10.1038/nrg3414>
- Delatycki, M. B., Williamson, R., & Forrest, S. M. (2000). Friedreich ataxia: An overview. *Journal of Medical Genetics*, *37*(1), 1–8. <https://doi.org/10.1136/jmg.37.1.1>
- Engelman, J. A., Luo, J., & Cantley, L. C. (2006). The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nature Reviews Genetics*, *7*(8), 606–619. <https://doi.org/10.1038/nrg1879>
- Gatti, M., Magri, S., Nanetti, L., Sarto, E., Di Bella, D., Salsano, E., ... Taroni, F. (2019). From congenital microcephaly to adult onset cerebellar ataxia: Distinct and overlapping phenotypes in patients with PNKP gene mutations. *American Journal of Medical Genetics, Part A*, *179*(11), 2277–2283. <https://doi.org/10.1002/ajmg.a.61339>
- Greene, D., Richardson, S., & Turro, E. (2016). Phenotype Similarity Regression for Identifying the Genetic Determinants of Rare Diseases. *American Journal of Human Genetics*, *98*(3), 490–499. <https://doi.org/10.1016/j.ajhg.2016.01.008>
- Guevara Coto, J., Schwartz, C. E., & Wang, L. (2014). Protein sector analysis for the

- clustering of disease-associated mutations. *BMC Genomics*, 15(Suppl 11), S4. <https://doi.org/10.1186/1471-2164-15-S11-S4>
- Halabi, N., Rivoire, O., Leibler, S., & Ranganathan, R. (2009). Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell*, 138(4), 774–786. <https://doi.org/10.1016/j.cell.2009.07.038>
- Holmes, G. (1908). An attempt to classify cerebellar disease , with a note on Marie’s Hereditary Cerebellar Ataxia. *Brain*, 30(4), 545–567. <https://doi.org/https://doi.org/10.1093/brain/30.4.545>
- Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Ridwan Amode, M., ... Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, 49(D1), D884–D891. <https://doi.org/10.1093/nar/gkaa942>
- Inlora, J., Sailani, M. R., Khodadadi, H., Teymurinezhad, A., Takahashi, S., Bernstein, J. A., ... Snyder, M. P. (2017). Identification of a novel mutation in the APTX gene associated with ataxia-oculomotor apraxia. *Cold Spring Harbor Molecular Case Studies*, 3(6), 1–10. <https://doi.org/10.1101/mcs.a002014>
- Jayadev, S., & Bird, T. D. (2013). Hereditary ataxias: Overview. *Genetics in Medicine*, 15(9), 673–683. <https://doi.org/10.1038/gim.2013.28>
- Jobling, R. K., Assoum, M., Gakh, O., Blaser, S., Raiman, J. A., Mignot, C., ... Yoon, G. (2015). PMPCA mutations cause abnormal mitochondrial protein processing in patients with non-progressive cerebellar ataxia. *Brain*, 138(6), 1505–1517. <https://doi.org/10.1093/brain/awv057>
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., ... MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434–443. <https://doi.org/10.1038/s41586-020-2308-7>
- Kent, W., Sugnet, C., Furey, T., Roskin, K., Pringle, T., Zahler, A., & Haussler, D. (2002). The human genome browser at UCSC. Retrieved March 10, 2018, from https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr9%3A32972606-33001628&hgslid=679854365_a5kEusvN9FluLmsqGveaiqayHGGD
- Klockgether, T. (2010). Sporadic ataxia with adult onset: classification and diagnostic criteria. *The Lancet Neurology*, 9(1), 94–104. [https://doi.org/10.1016/S1474-4422\(09\)70305-9](https://doi.org/10.1016/S1474-4422(09)70305-9)
- Köhler, S., Gargano, M., Matentzoglou, N., Carmody, L. C., Lewis-Smith, D., Vasilevsky, N. A., ... Robinson, P. N. (2021). The human phenotype ontology in 2021. *Nucleic Acids Research*, 49(D1), D1207–D1217. <https://doi.org/10.1093/nar/gkaa1043>
- Köhler, S., Schulz, M. H., Krawitz, P., Bauer, S., Dölken, S., Ott, C. E., ... Robinson, P. N. (2009). Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies. *American Journal of Human Genetics*, 85(4), 457–464. <https://doi.org/10.1016/j.ajhg.2009.09.003>
- Kumar, S., Stecher, G., Tamura, K., & Dudley, J. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., ... Magglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46, D1062–D1067. <https://doi.org/DOI:10.1093/nar/gkx1153>
- Leal, A., Bogantes-Ledezma, S., Ekici, A. B., Uebe, S., Thiel, C. T., Sticht, H., ... Reis, A. (2018). The polynucleotide kinase 3’-phosphatase gene (PNKP) in Charcot-

- Marie-Tooth disease (CMT2B2) previously related to MED25. *Neurogenetics*. <https://doi.org/https://doi.org/10.1007/s1048-018-0555-7>
- Li, W., Jaroszewski, L., & Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. In *BIOINFORMATICS APPLICATIONS NOTE* (Vol. 17). Retrieved from <http://bioinformatics>.
- Manto, M. U. (2005). The wide spectrum of spinocerebellar ataxias (SCAs). *Cerebellum*, 4(1), 2–6. <https://doi.org/10.1080/14734220510007914>
- Mariani, L. L., Rivaud-Péchoux, S., Charles, P., Ewenczyk, C., Meneret, A., Monga, B. B., ... Anheim, M. (2017). Comparing ataxias with oculomotor apraxia: A multimodal study of AOA1, AOA2 and at focusing on video-oculography and alpha-fetoprotein. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-15127-9>
- McLaughlin, R. N., Poelwijk, F. J., Raman, A., Gosal, W. S., & Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. *Nature*, 491(7422), 138–142. <https://doi.org/10.1038/nature11500>
- Medicine, M.-N. I. of G., & Johns Hopkins University (Baltimore, M. (n.d.). Online Mendelian Inheritance in Man, OMIM®. Retrieved March 10, 2018, from <https://omim.org/>
- Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., ... Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences of the United States of America*, 108(49). <https://doi.org/10.1073/pnas.1111471108>
- Moreira, M. C., Klur, S., Watanabe, M., Németh, A. H., Le Ber, I., Moniz, J. C., ... Koenig, M. (2004). Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature Genetics*, 36(3), 225–227. <https://doi.org/10.1038/ng1303>
- Moreira, M., & Koenig, M. (n.d.). Ataxia with Oculomotor Apraxia Type 2 (E. Adam MP, Ardinger HH, Pagon RA, et al., Ed.). Retrieved from GeneReviews® [Internet] website: <https://www.ncbi.nlm.nih.gov/books/NBK1154/>
- Moreno-Brid, J. C., & Ruiz-Nápoles, P. (2009). La educación superior y el desarrollo económico en América Latina. *Estudios y Perspectivas CEPAL*, 106, 46.
- Nemeth, A. H., Bochukova, E., Dunne, E., Huson, S. M., Elston, J., Hannan, M. A., ... Taylor, A. M. R. (2000). Autosomal recessive cerebellar ataxia with oculomotor apraxia (ataxia-telangiectasia-like syndrome) is linked to chromosome 9q34. *American Journal of Human Genetics*, 67(5), 1320–1326. [https://doi.org/10.1016/S0002-9297\(07\)62962-0](https://doi.org/10.1016/S0002-9297(07)62962-0)
- Nickless, A., Bailis, J. M., & You, Z. (2017). Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell and Bioscience*, 7(1), 1–12. <https://doi.org/10.1186/s13578-017-0153-7>
- Pedroso, J. L., Rocha, C. R. R., Macedo-Souza, L. I., De Mario, V., Marques, W., Barsottini, O. G. P., ... Kok, F. (2015). Clinical/scientific notes. *Neurology: Genetics*, 1(4), 1–4. <https://doi.org/10.1212/NXG.0000000000000030>
- Perlman, S. (2022). Hereditary Ataxia Overview 1 . Clinical Characteristics of Primary Hereditary Ataxia. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK1138/>
- Prasad, T. S. K., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., ... Kishore, A. (2009). Human Protein Reference Database. Retrieved March 12, 2018, from http://hprd.org/index_html
- Pulst, S. M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X. N., Lopes-Cendes, I., ... Sahba, S. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type. *Nature Genetics*, 14(3), 269–276.

- <https://doi.org/10.1038/ng1196-269>
- Richard, P., Feng, S., & Manley, J. L. (2013a). A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes and Development*, *27*(20), 2227–2232. <https://doi.org/10.1101/gad.224923.113>
- Richard, P., Feng, S., & Manley, J. L. (2013b). A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes and Development*, *27*(20), 2227–2232. <https://doi.org/10.1101/gad.224923.113>
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–424. <https://doi.org/10.1038/gim.2015.30>
- Rosenberg, R. N., & Khemani, P. (2015). The Inherited Ataxias. In *Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease: Fifth Edition* (Fifth Edit, pp. 811–832). <https://doi.org/10.1016/B978-0-12-410529-4.00071-1>
- Ruano, L., Melo, C., Silva, M. C., & Coutinho, P. (2014). The global epidemiology of hereditary ataxia and spastic paraplegia: A systematic review of prevalence studies. *Neuroepidemiology*, *42*(3), 174–183. <https://doi.org/10.1159/000358801>
- Sanavia, T., Birolo, G., Montanucci, L., Turina, P., Capriotti, E., & Fariselli, P. (2020). Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. *Computational and Structural Biotechnology Journal*, *18*, 1968–1979. <https://doi.org/10.1016/j.csbj.2020.07.011>
- Schiess, N., Zee, D. S., Siddiqui, K. A., Szolics, M., & El-Hattab, A. W. (2017). Novel PNKP mutation in siblings with ataxia-oculomotor apraxia type 4. *Journal of Neurogenetics*, *31*(1–2), 23–25. <https://doi.org/10.1080/01677063.2017.1322079>
- Schmitz-Hübsch, T., Coudert, M., Bauer, P., Giunti, P., Globas, C., Baliko, L., ... Klockgether, T. (2008). Spinocerebellar ataxia types 1, 2, 3, and 6: Disease severity and nonataxia symptoms. *Neurology*, *71*(13), 982–989. <https://doi.org/10.1212/01.wnl.0000325057.33666.72>
- Schöls, Ludger; Bauer, Peter; Schmidt, Thorsten; Schulte Thorsten; Riess, O. (2004). Autosomal dominant cerebellar ataxias: clinical features, genetics and pathogenesis. *Neurology*, *3*(2), 8. <https://doi.org/10.4324/9780429456916-3>
- Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014). Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nature Methods*, *11*(4), 361–362. <https://doi.org/10.1038/nmeth.2890>
- Seo, G. H., Kim, T., Choi, I. H., Park, J. young, Lee, J., Kim, S., ... Lee, B. H. (2020). Diagnostic yield and clinical utility of whole exome sequencing using an automated variant prioritization system, EVIDENCE. *Clinical Genetics*, *98*(6), 562–570. <https://doi.org/10.1111/cge.13848>
- Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M., & Marino Buslje, C. (2013). MISTIC: Mutual information server to infer coevolution. *Nucleic Acids Research*, *41*(Web Server issue), 8–14. <https://doi.org/10.1093/nar/gkt427>
- Tassan, N. Al, Khalil, D., Shinwari, J., Sharif, L. Al, Bavi, P., Abduljaleel, Z., ... Bohlega, S. (2012). A Missense mutation in PIK3R5 gene in a family with ataxia and oculomotor apraxia. *Human Mutation*, *33*(2), 351–354. <https://doi.org/10.1002/humu.21650>
- Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2002). Multiple Sequence Alignment Using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, 1–22.

- <https://doi.org/10.1002/0471250953.bi0203s00>
- Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... Ponten, F. (2015). Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419–1260419. <https://doi.org/10.1126/science.1260419>
- van Minkelen, R., Guitart, M., Escofet, C., Yoon, G., Elfferich, P., Bolman, G. M., ... van den Ouweland, A. M. W. (2015). Complete APTX deletion in a patient with ataxia with oculomotor apraxia type 1. *BMC Medical Genetics*, *16*(1), 1–4. <https://doi.org/10.1186/s12881-015-0213-y>
- Waters, M. F., Fee, D., Figueroa, K. P., Nolte, D., Müller, U., Advincula, J., ... Pulst, S. M. (2005). An autosomal dominant ataxia maps to 19q13: Allelic heterogeneity of SCA13 or novel locus? *Neurology*, *65*(7), 1111–1113. <https://doi.org/10.1212/01.wnl.0000177490.05162.41>
- Waters, Michael F. (2020). Spinocerebellar Ataxia Type 13 - GeneReviews® - NCBI Bookshelf. Retrieved June 19, 2022, from <https://www.ncbi.nlm.nih.gov/books/NBK1225/>
- Waters, Michael F., Minassian, N. A., Stevanin, G., Figueroa, K. P., Bannister, J. P. A., Nolte, D., ... Pulst, S. M. (2006). Mutations in voltage-gated potassium channel KCNC3 cause degenerative and developmental central nervous system phenotypes. *Nature Genetics*, *38*(4), 447–451. <https://doi.org/10.1038/ng1758>
- Yue, P., Li, Z., & Moulton, J. (2005). Loss of protein structure stability as a major causative factor in monogenic disease. *Journal of Molecular Biology*, *353*(2), 459–473. <https://doi.org/10.1016/j.jmb.2005.08.020>
- Zhao, J., Zhu, J., & Thornhill, W. B. (2013). Spinocerebellar ataxia-13 Kv3.3 potassium channels: Arginine-to-histidine mutations affect both functional and protein expression on the cell surface. *Biochemical Journal*, *454*(2), 259–265. <https://doi.org/10.1042/BJ20130034>
- Zühlke, C., Gehlken, U., Hellenbroich, Y., Schwinger, E., & Bürk, K. (2003). Phenotypical variability of expanded alleles in the TATA-binding protein gene: Reduced penetrance in SCA17? *Journal of Neurology*, *250*(2), 161–163. <https://doi.org/10.1007/s00415-003-0958-7>

Conclusiones

En el mundo se han identificado cuatro genes causantes de los diferentes tipos de AOA identificados hasta la fecha, en este estudio se identificaron en *PNKP* las variantes p.Gln517Ter y p.Thr408del que están asociadas a AOA4 en condición tanto homocigota como heterocigota compuesta. Estas variantes habían sido caracterizadas en estudios previos como causantes de AOA4/CMT2B en la población de Costa Rica.

Además de las variantes mencionadas previamente, se identificó una variante no descrita en la literatura (Arg433Ter) y que los análisis bioinformáticos de predicción de estabilidad y funcionalidad proteica señalan como una posible causante de la enfermedad. Sin embargo, más estudios son necesarios para confirmar la hipótesis que relaciona el fenotipo observado con esta nueva variante.

Las mutaciones identificadas en *PNKP* -deleciones y responsables de terminaciones tempranas- pueden dar lugar a alteraciones proteicas que conducen a ineficiencia de reparación del ADN dañado. De hecho, debido a estas mutaciones, se desestabiliza la estructura proteica por debilitamiento del anclaje del dominio quinasa al dominio fosfatasa, se reduce de la actividad enzimática (fosforilación 5' de ácidos nucleicos y actividad fosfatasa 3' que son catalizadas en respuesta al estrés oxidativo en el sistema nervioso) y, finalmente, se provoca la muerte celular.

En todas las superfamilias de proteínas analizadas con información mutual se infirieron entre las posiciones conservadas en diferentes grados, residuos funcionalmente importantes ya conocidos para cada uno de los tipos descritos de AOA. Adicionalmente se identificaron posibles sitios candidatos que, por su grado de conservación, podrían causar fenotipos similares a los actualmente descritos si mutaran. Un ejemplo de esto es el residuo 4333R, el mismo residuo mutado que fue identificado en el paciente AT-042,

lo que puede servir como una prueba real de concepto observada en un paciente. Este trabajo aporta evidencia que apoya lo reportado por otros autores sobre la utilidad de las herramientas bioinformáticas y sus posibles aplicaciones en proyectos de secuenciación a gran escala al visibilizar cómo las relaciones de aminoácidos subyacentes pueden configurar su papel en la aparición de diferentes manifestaciones fenotípicas.

Hay que aclarar que, aunque no todos los sitios identificados que están altamente conservados, han sido reportados como causantes de la enfermedad, se detecta una consistencia entre las predicciones y la realidad observada que merece la pena seguir en otros estudios con una cohorte mayor.

Por último, además de contribuir a la identificación de las variantes presentes en Costa Rica y el establecimiento de un diagnóstico específico, los resultados de este trabajo brindan un estímulo psicológico para los individuos afectados y sus familias por el hecho de poner nombre a una enfermedad hasta el momento desconocida para ellos. También, permiten una discusión informada sobre el pronóstico y los tratamientos disponibles, además de abrir la puerta para el desarrollo de nuevas terapias.

Referencias Bibliográficas

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., ... Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1), D480–D489. <https://doi.org/10.1093/nar/gkaa1100>

Becherel, O. J., Sun, J., Yeo, A. J., Nayler, S., Fogel, B. L., Gao, F., ... Lavin, M. F. (2015). A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. *Human Molecular Genetics*, 24(20), 5759–5774. <https://doi.org/10.1093/hmg/ddv296>

Bird, T. D. (2019). Hereditary Ataxia Overview. Retrieved from GeneReviews™ website: <https://www.ncbi.nlm.nih.gov/books/NBK1138/>

Bras, J., Alonso, I., Barbot, C., Costa, M. M., Darwent, L., Orme, T., ... Guerreiro, R. (2015). Mutations in PNKP cause recessive ataxia with oculomotor apraxia type 4. *American Journal of Human Genetics*, 96(3), 474–479. <https://doi.org/10.1016/j.ajhg.2015.01.005>

Brenner, C. (2002). Hint, Fhit, and GalT: Function, structure, evolution, and mechanism of three branches of the histidine triad superfamily of nucleotide hydrolases and transferases. *Biochemistry*, 41(29), 9003–9014. <https://doi.org/10.1021/bi025942q>

Brenner, C., Biecanowski, P., Pace, H. C., & Huebner, K. (1999). The histidine triad superfamily of nucleotide-binding proteins. *Journal of Cellular Physiology*, 181(2), 179–

187. [https://doi.org/10.1002/\(SICI\)1097-4652\(199911\)181:2<179::AID-JCP1>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-4652(199911)181:2<179::AID-JCP1>3.0.CO;2-8)

Chen, C. W., Lin, M. H., Liao, C. C., Chang, H. P., & Chu, Y. W. (2020). iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules. *Computational and Structural Biotechnology Journal*, 18, 622–630. <https://doi.org/10.1016/j.csbj.2020.02.021>

Coutinho, P., & Barbot, C. (2002). Ataxia with Oculomotor Apraxia Type 1. *GeneReviewsTM*, 1–12. <https://doi.org/NBK1456> [bookaccession]

De Juan, D., Pazos, F., & Valencia, A. (2013). Emerging methods in protein co-evolution. *Nature Reviews Genetics*, 14(4), 249–261. <https://doi.org/10.1038/nrg3414>

Duquette, A., Roddier, K., McNabb-Baltar, J., Gosselin, I., St-Denis, A., Dicaite, M. J., ... Brais, B. (2005). Mutations in senataxin responsible for Quebec cluster of ataxia with neuropathy. *Annals of Neurology*, 57(3), 408–414. <https://doi.org/10.1002/ana.20408>

Engelman, J. A., Luo, J., & Cantley, L. C. (2006). The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nature Reviews Genetics*, 7(8), 606–619. <https://doi.org/10.1038/nrg1879>

Gatti, M., Magri, S., Nanetti, L., Sarto, E., Di Bella, D., Salsano, E., ... Taroni, F. (2019). From congenital microcephaly to adult onset cerebellar ataxia: Distinct and overlapping phenotypes in patients with PNKP gene mutations. *American Journal of Medical Genetics, Part A*, 179(11), 2277–2283. <https://doi.org/10.1002/ajmg.a.61339>

Greene, D., Richardson, S., & Turro, E. (2016). Phenotype Similarity Regression for Identifying the Genetic Determinants of Rare Diseases. *American Journal of Human Genetics*, 98(3), 490–499. <https://doi.org/10.1016/j.ajhg.2016.01.008>

Holmes, G. (1908). An attempt to classify cerebellar disease , with a note on Marie's Hereditary Cerebellar Ataxia. *Brain*, 30(4), 545–567.
<https://doi.org/https://doi.org/10.1093/brain/30.4.545>

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Ridwan Amode, M., ... Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, 49(D1), D884–D891.
<https://doi.org/10.1093/nar/gkaa942>

Inlora, J., Sailani, M. R., Khodadadi, H., Teymurinezhad, A., Takahashi, S., Bernstein, J. A., ... Snyder, M. P. (2017). Identification of a novel mutation in the APTX gene associated with ataxia-oculomotor apraxia. *Cold Spring Harbor Molecular Case Studies*, 3(6), 1–10. <https://doi.org/10.1101/mcs.a002014>

Jayadev, S., & Bird, T. D. (2013). Hereditary ataxias: Overview. *Genetics in Medicine*, 15(9), 673–683. <https://doi.org/10.1038/gim.2013.28>

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., ... MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434–443. <https://doi.org/10.1038/s41586-020-2308-7>

Kent, W., Sugnet, C., Furey, T., Roskin, K., Pringle, T., Zahler, A., & Haussler, D. (2002). The human genome browser at UCSC. Retrieved March 10, 2018, from https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr9%3A32972606-33001628&hgside=679854365_a5kEusvN9FluLmsqGveaiqayHGGD

- Klockgether, T. (2010). Sporadic ataxia with adult onset: classification and diagnostic criteria. *The Lancet Neurology*, 9(1), 94–104. [https://doi.org/10.1016/S1474-4422\(09\)70305-9](https://doi.org/10.1016/S1474-4422(09)70305-9)
- Köhler, S., Gargano, M., Matentzoglou, N., Carmody, L. C., Lewis-Smith, D., Vasilevsky, N. A., ... Robinson, P. N. (2021). The human phenotype ontology in 2021. *Nucleic Acids Research*, 49(D1), D1207–D1217. <https://doi.org/10.1093/nar/gkaa1043>
- Köhler, S., Schulz, M. H., Krawitz, P., Bauer, S., Dölken, S., Ott, C. E., ... Robinson, P. N. (2009). Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies. *American Journal of Human Genetics*, 85(4), 457–464. <https://doi.org/10.1016/j.ajhg.2009.09.003>
- Kumar, S., Stecher, G., Tamura, K., & Dudley, J. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., ... Magglott, D. R. (2018). ClinVAR: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46, D1062–D1067. <https://doi.org/DOI:10.1093/nar/gkx1153>
- Laurencin, C., Anheim, M., Larrieu, L., Tilikete, C., Koenig, M., & Thobois, S. (2015). Pitfalls in ataxia with ocular motor apraxia type 1: pseudodominant inheritance and very late onset. *Journal of Neurology*, 262(5), 1366–1368. <https://doi.org/10.1007/s00415-015-7717-4>
- Leal, A., Bogantes-Ledezma, S., Ekici, A. B., Uebe, S., Thiel, C. T., Sticht, H., ... Reis, A. (2018). The polynucleotide kinase 3'-phosphatase gene (PNKP) in Charcot-Marie-

Tooth disease (CMT2B2) previously related to MED25. *Neurogenetics*.
<https://doi.org/https://doi.org/10.1007/s1048-018-0555-7>

Li, W., Jaroszewski, L., & Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. In *BIOINFORMATICS APPLICATIONS NOTE* (Vol. 17). Retrieved from <http://bioinformatics>.

Manto, M. U. (2005). The wide spectrum of spinocerebellar ataxias (SCAs). *Cerebellum*, 4(1), 2–6. <https://doi.org/10.1080/14734220510007914>

Mariani, L. L., Rivaud-Péchoux, S., Charles, P., Ewencyk, C., Meneret, A., Monga, B. B., ... Anheim, M. (2017). Comparing ataxias with oculomotor apraxia: A multimodal study of AOA1, AOA2 and at focusing on video-oculography and alpha-fetoprotein. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-15127-9>

Medicine;, M.-N. I. of G., & Johns Hopkins University (Baltimore, M. (n.d.). Online Mendelian Inheritance in Man, OMIM®. Retrieved March 10, 2018, from <https://omim.org/>

Mignarri, A., Tessa, A., Federico, A., Santorelli, F. M., & Dotti, M. T. (2015). Ataxia with oculomotor apraxia type 2: not always an easy diagnosis. *Neurological Sciences*, 36(8), 1505–1507. <https://doi.org/10.1007/s10072-015-2119-z>

Moreira, M. C., Klur, S., Watanabe, M., Németh, A. H., Le Ber, I., Moniz, J. C., ... Koenig, M. (2004). Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature Genetics*, 36(3), 225–227. <https://doi.org/10.1038/ng1303>

Pedroso, J. L., Rocha, C. R. R., Macedo-Souza, L. I., De Mario, V., Marques, W., Barsottini, O. G. P., ... Kok, F. (2015). Clinical/scientific notes. *Neurology: Genetics*, 1(4), 1–4. <https://doi.org/10.1212/NXG.0000000000000030>

Prasad, T. S. K., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., ... Kishore, A. (2009). Human Protein Reference Database. Retrieved March 12, 2018, from http://hprd.org/index_html

Pulst, S. M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X. N., Lopes-Cendes, I., ... Sahba, S. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type. *Nature Genetics*, 14(3), 269–276. <https://doi.org/10.1038/ng1196-269>

Richard, P., Feng, S., & Manley, J. L. (2013). A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes and Development*, 27(20), 2227–2232. <https://doi.org/10.1101/gad.224923.113>

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>

Ruano, L., Melo, C., Silva, M. C., & Coutinho, P. (2014). The global epidemiology of hereditary ataxia and spastic paraplegia: A systematic review of prevalence studies. *Neuroepidemiology*, 42(3), 174–183. <https://doi.org/10.1159/000358801>

Schiess, N., Zee, D. S., Siddiqui, K. A., Szolics, M., & El-Hattab, A. W. (2017). Novel PNKP mutation in siblings with ataxia-oculomotor apraxia type 4. *Journal of Neurogenetics*, 31(1–2), 23–25. <https://doi.org/10.1080/01677063.2017.1322079>

Schmitz-Hübsch, T., Coudert, M., Bauer, P., Giunti, P., Globas, C., Baliko, L., ... Klockgether, T. (2008). Spinocerebellar ataxia types 1, 2, 3, and 6: Disease severity and

nonataxia symptoms. *Neurology*, 71(13), 982–989.
<https://doi.org/10.1212/01.wnl.0000325057.33666.72>

Schöls, Ludger; Bauer, Peter; Schmidt, Thorsten; Schulte Thorsten; Riess, O. (2004). Autosomal dominant cerebellar ataxias: clinical features, genetics and pathogenesis. *Neurology*, 3(2), 8. <https://doi.org/10.4324/9780429456916-3>

Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014). Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nature Methods*, 11(4), 361–362. <https://doi.org/10.1038/nmeth.2890>

Seo, G. H., Kim, T., Choi, I. H., Park, J. young, Lee, J., Kim, S., ... Lee, B. H. (2020). Diagnostic yield and clinical utility of whole exome sequencing using an automated variant prioritization system, EVIDENCE. *Clinical Genetics*, 98(6), 562–570. <https://doi.org/10.1111/cge.13848>

Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M., & Marino Buslje, C. (2013). MISTIC: Mutual information server to infer coevolution. *Nucleic Acids Research*, 41(Web Server issue), 8–14. <https://doi.org/10.1093/nar/gkt427>

Szpisjak, L., Obal, I., Engelhardt, J. I., Vecsei, L., & Klivenyi, P. (2016). A novel SETX gene mutation producing ataxia with oculomotor apraxia type 2. *Acta Neurologica Belgica*, 116(3), 405–407. <https://doi.org/10.1007/s13760-015-0569-y>

Tassan, N. Al, Khalil, D., Shinwari, J., Sharif, L. Al, Bavi, P., Abduljaleel, Z., ... Bohlega, S. (2012). A Missense mutation in PIK3R5 gene in a family with ataxia and oculomotor apraxia. *Human Mutation*, 33(2), 351–354. <https://doi.org/10.1002/humu.21650>

Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2002). Multiple Sequence Alignment Using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, 1–22. <https://doi.org/10.1002/0471250953.bi0203s00>

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... Ponten, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220), 1260419–1260419. <https://doi.org/10.1126/science.1260419>

van Minkelen, R., Guitart, M., Escofet, C., Yoon, G., Elfferich, P., Bolman, G. M., ... van den Ouweland, A. M. W. (2015). Complete APTX deletion in a patient with ataxia with oculomotor apraxia type 1. *BMC Medical Genetics*, 16(1), 1–4. <https://doi.org/10.1186/s12881-015-0213-y>

Zühlke, C., Gehlken, U., Hellenbroich, Y., Schwinger, E., & Bürk, K. (2003). Phenotypical variability of expanded alleles in the TATA-binding protein gene: Reduced penetrance in SCA17? *Journal of Neurology*, 250(2), 161–163. <https://doi.org/10.1007/s00415-003-0958-7>