

Caracterización *in silico* de exones y secuencias específicas de unión para factores de la transcripción en los intrones del gen *PARKIN* RBR E3 Ubiquitina Proteína Ligasa 3 (*PRKN*) humana

Erika Paola Ardila Sandoval

Trabajo de Grado para Optar al Título de Bióloga

Director

Francisco José Martínez Pérez

Doctor en Ciencias Biológicas

Codirector

Carlos Jaime Barrios Hernández

Doctor en Ciencias

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Programa de Biología

Bucaramanga

2023

Dedicatoria

Hace algún tiempo leí un libro, el cual deseaba leer con todo mi corazón porque hablaba de la superación a la pérdida de un ser amado; hace poco tuve la capacidad y la oportunidad de terminarlo y con su historia pude identificarme de muchas maneras y es que muchas cosas no son para siempre. Segundo, que la historia de uno de los personajes curiosamente se parece tanto a ti mamá: su sabiduría, su consejo y sin quererlo El otro personaje principal se parece a mí: la manera en cómo la admira, la escucha y trata de recordar sus palabras y honrar su memoria y eso quiero con esta dedicatoria. Tú eres mi “Dora” particular, siempre tierna y agradecida, y si has llegado hasta estas páginas es porque habré logrado mi objetivo; hacerte eterna a ti a tu amor hacía quienes te rodeaban, y si he aguantado y e soportado todo por llegar hasta aquí, lo he hecho con amor, emoción, y con la misma lágrima que no termina y que me ha acompañado por 13 años pero que se mantiene sostenida en mi mejilla cuándo te recuerdo, y es que a esa lágrima acudo cuándo no puedo comprender el mundo y ver sus injusticias, y tú me has enseñado que el amor sí que puede ser para siempre y que no hace falta ver las cosas para creer en ellas.

Por todo eso y mucho más sé que sólo tú y yo sabíamos, gracias, mamá.

Te hecho tanto de menos, como te amo.

Agradecimientos

Al Dr. Francisco José Martínez Pérez y al Dr. Carlos Jaime Barrios Hernández por permitirme hacer parte de su grupo de trabajo, en especial al profesor Francisco quien ha sido mi mentor y apoyo en este proceso, a él mil gracias por la confianza depositada en mí.

A la Universidad Industrial de Santander.

A la Facultad de Ciencias de la Universidad Industrial de Santander. A la Escuela de Biología de la Universidad Industrial de Santander. Al Laboratorio de Genómica de Celomados (GCL-UIS).

Al Centro de Supercomputación Cálculo Científico SC3-UIS.

Al Grupo de Investigación Computo Avanzado y a Gran Escala (CAGE).

A mi padre que a pesar de las dificultades es mi compañero de vida y mayor apoyo, y a mi madre quien puso en mí la motivación de siempre luchar por mis sueños.

A todos ustedes lectores.

Agradezco el financiamiento de esta pasantía al proyecto titulado: “Contribución al protocolo de diagnóstico del Coronavirus COVID-19 y del Virus de Influenza A H1N1 por RT-PCR en tiempo real de un paso autorizado por la Organización Mundial de la Salud con la inclusión de genes marcadores requeridos para la infección viral y su validación por secuenciación genómica de nueva generación con la tecnología de concentraciones de nucleótidos”, (Contrato 369-2020 y código No. 1102101576900) de la Mincienciación Correspondiente a la Convocatoria: 1015- Invitación a presentar propuestas de proyectos relacionados con la pandemia de COVID- 19. Además, a la Vicerrectoría de Investigación y Extensión de la Universidad Industrial de Santander por el financiamiento (Proyecto 76900).

Tabla de Contenido

	Pág.
Introducción	12
1. Objetivos	18
1.1 Objetivo de la pasantía.....	18
1.2 Objetivo General.....	18
1.3 Objetivos Específicos.....	18
2. Competencias	19
2.1 Competencias Cognitivas.....	19
2.2 Competencias Actitudinales.....	19
3. Metodología	20
3.1 Identificación del gen y las variantes de ARN mensajero de <i>PRKN Homo sapiens</i> en la base de datos Gene del NCBI	20
3.1.1 Delimitación de intrones del gen <i>PRKN</i>	20
3.2 Validación <i>in silico</i> de exones presentes en los intrones del gen <i>PRKN</i>	21
3.3 Caracterización de secuencias específicas de unión para factores de la transcripción (SEUFT) en intrones	22
3.3.1 Caracterización funcional de exones y marcos abiertos de lectura en el gen <i>PRKN</i>	22
4. Resultados	23
4.1 Identificación de las secuencias del gen <i>PRKN</i> y ARNm en la base de datos Gene.	23
4.1.1 Delimitación de intrones y exones del gen <i>PRKN</i>	25
4.2 Validación <i>in silico</i> de exones en los intrones del gen <i>PRKN</i>	30

4.3 Características funcionales de los exones y marcos abiertos de lectura en los intrones del gen <i>PRKN</i>	35
4.3.1 Caracterización de secuencias específicas de unión para factores de la transcripción (SEUFT)	36
5. Discusión.....	37
6. Conclusiones	42
Referencias Bibliográficas	43

Lista de Tablas

	Pág.
Tabla 1 <i>Variantes de ARNm del gen PRKN en la base de datos GENE</i>	24

Lista de Figuras

	Pág.
Figura 1 <i>Imagen de la base de datos de GenBank correspondiente al gen PRKN</i>	23
Figura 2 <i>Alineamiento entre el primer exón y las variantes del ARNm de PRKN</i>	27
Figura 3 <i>Alineamiento manual entre el primer exón y las variantes del ARNm de PRKN</i>	28
Figura 4 <i>Gráfica del tamaño de los exones e intrones del gen PRKN</i>	30
Figura 5 <i>Primer Marco abierto de lectura del intrón 1 del gen PRKN</i>	31
Figura 6 <i>ORF del intrón 7 del gen PRKN con secuencias repetidas en tándem</i>	32
Figura 7 <i>Cantidad de ORFs sentidos y antisentidos del gen PRKN</i>	33
Figura 8 <i>Distribución de los exones concomitantemente ORFs en los intrones 1 y 7</i>	34
Figura 9 <i>Caracterización de los dominios de los ORFs de los intrones del gen PRKN</i>	35

Lista de Apéndices

Apéndice A. Formatos FASTA del Gen *PRKN* y sus variantes ARN mensajero.

Apéndice B. Tabla suplementaria 1. Posiciones de intrones y exones del gen *PRKN*.

Apéndice C. Alineamientos entre el gen *PRKN* y sus variantes de ARNm.

Apéndice C.1 Gen-ARNm-*PRKN* por defecto.

Apéndice C.2 Var1-3 vs X1-3 ARNm-*PRKN* por defecto.

Apéndice C.3 Var1-3 vs X1-2 ARNm-*PRKN* por defecto.

Apéndice C.4 Var1-3 vs X2 ARNm-*PRKN* por defecto.

Apéndice C.5 Var1-3 ARNm *PRKN* por defecto.

Apéndice C.6 Alin GEN-Var ARNm-*PRKN* por defecto.

Apéndice C.7 Alin Gen-Var ARNm-*PRKN* 9-2-2.

Apéndice D. Mapa de la predicción de exones y ORFs en los intrones del gen *PRKN*. Los nombres de cada archivo indican el resultado de cada intrón con sus softwares AUGUST y GenScan.

Apéndice E. Tabla suplementaria 2. Dominios de los exones de los intrones del gen *PRKN*. Al indicar con el cursor cada dominio se envía a la página con los datos de la base PROSITE.

Apéndice F. Tabla suplementaria 3. Sitios río arriba de los exones de los intrones del gen *PRKN*.

Nota: Los apéndices están adjuntos y puede visualizarlos en la base de datos de la biblioteca UIS

Glosario

EF: Enfermedad de Parkinson.

EXÓN: Secuencia nucleotídica de un gen que codifica para una proteína.

In silico: Análisis experimental de un proceso biológico con elementos computacionales.

INTRÓN: Secuencia nucleotídica de un gen que separa a los exones y que puede o no tener elementos de regulación de la transcripción y traducción.

NCBI: *National Center for Biotechnology Information.*

ORF: Marco abierto de lectura.

pb: pb.

PRKN: Gen que codifica para la enzima E3 Ubiquitina Ligasa

SEUFT: Secuencias nucleotídicas de un gen específicas para la unión de factores para la transcripción.

VARIANTE: Secuencia alterna a la secuencia consenso, producto del procesamiento alternativo del ARNm.

V_n - X_n: Variante del ARN mensajero del gen *PRKN*. La letra n se sustituye por el número de la respectiva variante. La diferencia entre las nomenclaturas de cada variante.

Resumen

Título: Caracterización *in silico* de exones y secuencias específicas de unión para factores de la transcripción en los intrones del gen *PARKIN* RBR E3 Ubiquitina Proteína Ligasa 3 (*PRKN*) humana*

Autor: Erika Paola Ardila Sandoval**

Palabras Clave: Parkinson, Variante, Gen *PRKN*, Intrones, Exones Mitochondrias, Mitofagia, *In silico*.

Descripción: La Enfermedad de Parkinson (EP) es una alteración neurológica degenerativa que afecta la actividad motora de más de 6,1 millones de personas y se estima que para el 2040 habrá alrededor de 17 millones de personas con la EP. Uno de los genes relacionados con la EP es *PRKN* que contribuye en la mitofagia. El gen tiene 1.380.350 pb, sus mutaciones se asocian con la EP y algunas son la causa recesiva más común. Además, se han reportado 6 variantes de ARNm. Por ejemplo, la variante 1 (V1) tiene 4.178 pb. Por lo tanto, más de 1.376.172 pb corresponderán a intrones. No obstante, se desconoce si algunos de ellos tienen elementos de regulación de la transcripción y de la traducción. Para ello, primero se alinearon las variantes de ARNm con el gen. Las posiciones homólogas y su verificación respecto al GenBank mostraron que el primer intrón tuvo más de 300.000 pb, un grupo de 5 intrones de 150.000 a 200.000 pb y otro grupo, menor a 50.000 pb. En estos, los softwares GenScan y Web AUGUSTUS mostraron elementos promotores de la transcripción y 129 exones, con una o más señales de poliadenilación. El programa en línea PROSITE, mostró que más de la mitad de éstos corresponden a sitios de fosforilación, una cuarta parte para miristilación y el resto para otros elementos. Nuestros resultados indican que en los intrones del gen *PRKN* podrían codificar para nuevas proteínas que posiblemente participarán en la EP.

*Trabajo de Grado

**Facultad de Ciencias. Escuela de Biología. Director: Francisco José Martínez Pérez. Doctor en Ciencias Biológicas. Codirector: Carlos Jaime Barrios Hernández. Doctor en Ciencias.

Abstract

Title: *In silico* characterization of exons and specific binding sequences for transcription factors in the introns of the *PARKIN* RBR E3 Ubiquitin Protein Ligase (*PRKN*) gene*

Author(s): Erika Paola Ardila Sandoval**

Key Words: Parkinson's, Variant, *PRKN* Gene, Introns, Exons Mitochondria, Mitophagy, *In Silico*.

Description: Parkinson's disease (PD) is a degenerative neurological disorder that affects the motor activity of more than 6.1 million people, and it is estimated that by 2040 there will be around 17 million people with PD. One of the genes related to PD is *PRKN* which contributes to mitophagy. The gene has 1,380,350 base pairs (bp), its mutations are associated with PD and some are the most common recessive cause. In addition, 6 mRNA variants have been reported. For example, variant 1 (V1) is 4,178 bp. Therefore, more than 1,376,172 bp would correspond to introns. However, it is not known whether some of them have transcription and translation regulatory elements. For this purpose, mRNA variants were first aligned to the gene. The homologous positions and their verification against GenBank showed that the first intron had more than 300 thousand bp, a group of 5 introns from 150 to 200 thousand bp and another group, less than 50 thousand bp. In these, GenScan and Web AUGUSTUS software showed transcription promoter elements and 129 exons, with one or more polyadenylation signals. The PROSITE online program showed that more than half of these correspond to phosphorylation sites, a quarter to myristylation and the rest to other elements. Our results indicate that introns in the *PRKN* gene may encode for new proteins that are likely to participate in PD.

*Degree work

**Science Faculty. School of Biology

Adviser: PhD. Francisco José Martínez Pérez. Adviser: PhD. Carlos Jaime Barrios Hernández.

Introducción

La Enfermedad de Parkinson (EP) recibe su nombre por su descubridor el Dr. James Parkinson (Bautista, 2021). Posterior a la descripción de sus síntomas, se determinó que es una alteración neurológica degenerativa que afecta la actividad motora y se ha propuesto que está presente en más de 6,1 millones de personas en todo el mundo (Feigin et al., 2017), además de considerarse la segunda enfermedad neurodegenerativa del cerebro después de la enfermedad de Alzheimer (Al-kuraishy et al., 2023).

En Colombia, la EP tiene una prevalencia anual de 115 casos por cada 100.000 habitantes (Nichols et al., 2019). Aunado a que el ministerio de salud colombiano indicó que la población de adultos mayores con EP aumentó del 6,9% al 13,8% entre los años 1985 y 2020; además de tener un posible incremento hasta el 16% para el 2030 (Bautista, 2021).

La EP afecta tanto a mujeres como a hombres y de acuerdo con el Ministerio de Salud de Colombia, el Registro Individual de Prestación de Servicios en Salud (RIPS) indica que se diagnosticaron más de 148.000 personas entre 2016 y 2020. De ellas, un 44% corresponde a mujeres y el resto a hombres (Bautista, 2021). En este sentido, el programa integral de la investigación *Global Burden of Disease*, estima que para el 2040 habrá alrededor de 17 millones de personas con la EP. Es de resaltar que para el Departamento de Santander únicamente se determinó un estudio neuroepidemiológico de la EP en el municipio de Piedecuesta (Pradilla Ardila et al., 2002).

La EP se presenta principalmente en la vejez con la ralentización generalizada de los movimientos y síntomas de temblor o rigidez en reposo (Alexoudi et al., 2018; Mirpour et al., 2018). Para su diagnóstico se emplea principalmente la escala unificada de calificación de la

enfermedad de Parkinson de la Sociedad de Trastornos del Movimiento (MDS-UPDRS) (Liu et al., 2002), y en muestras sanguíneas, se utilizan marcadores como la α -sinucleína, IL6 y ADNmt (Borsche et al., 2020; Kluge et al., 2022).

Se ha encontrado que en la EP se ven afectadas las neuronas dopaminérgicas de la sustancia nigra y a su vez, una relación directa con el daño mitocondrial (Vizziello et al., 2021). Los genes *PINK1* y *PRKN* están involucrados en la homeostasis mitocondrial, ya que se encargan de marcar las mitocondrias que serán degradadas vía lisosoma en una autofagia selectiva, a lo que se denomina mitofagia, lo cual ocurre cuando estas ya no funcionan correctamente (Bayne y Trempe, 2019; Fakih et al., 2023; Han et al., 2022; Pickrell y Youle, 2015). Además, Pereira y colaboradores, indican que *PRKN* participa en varios procesos celulares como lo es: la formación de vesículas derivadas de la mitocondria, metabolismo celular, la homeostasis de calcio, el mantenimiento del genoma mitocondrial, la biogénesis mitocondria l y el inicio de la apoptosis (2023). Mutaciones presentes en estos genes promueven la neurodegeneración, al igual que la transmisión recesiva de la EP (Sato et al., 2018).

La estructura y organización del gen *PRKN* es de gran importancia para entender el desarrollo de la enfermedad de Parkinson ya que se ha reportado que los eventos de duplicación y las deleciones de exones, están directamente relacionadas con la enfermedad (Ahmad et al., 2023; Seike et al., 2021). Por ejemplo, las deleciones en los exones 2 y 4, han sido determinadas por autopsias de pacientes con la enfermedad (Jensen et al., 2022; Sakuwa et al., 2021). Mientras que otros reportes indican que la deleción del exón 4 promueve el desplazamiento del marco de lectura entre los exones 3 y 5, así como el desplazamiento codón de paro de la traducción del exón 3, lo que produce una alteración en el dominio funcional de “Dedos tipo Anillo”, el cual es un tipo especializado del dominio Dedos de Zinc (Tan et al., 2005).

Otro tipo de mutación en el gen *PRKN* es la duplicación del exón 7 y las deleciones en el exón 3 y en otras regiones del exón 4, las cuales reducen la cantidad de cuerpos de Lewy (Elfferich et al., 2011), que son empleados en la detección de la EP (Johansen et al., 2018). Por otro lado, se reportó un caso de inversión en el gen, específicamente en intrones circundantes al exón 5, lo que produjo su reemplazo por 98 pb; lo que preservó el ORF, pero se perdió un número representativo de pb (Mor-Shaked et al., 2021).

Por lo anterior, las mutaciones en el gen *PRKN* también son ampliamente estudiadas, ya que se han reportado como la mayor causa de aparición de EP, sin embargo, algunos autores consideran que la presencia de estas puede no tener un impacto en la probabilidad de desarrollar la enfermedad (Zhu et al., 2022). A pesar de lo anterior, es posible relacionarlas directamente con la EP, como sucede con una mutación en el exón 2 (G403>C), que genera el cambio de una glicina a una arginina, alterando la estructura de la proteína (Fang et al., 2019).

Puesto que la mayoría del estudio de la relación mutacional del gen *PRKN* y la EP el conocimiento del patrón mutacional de los intrones es en menor cantidad; por ejemplo, se demostró que una deleción de 80 pb en el intrón 7, reduce casi el doble la fragilidad del gen *PRKN*, lo que demuestra que esta región está involucrada en la supervivencia del gen (Munk et al., 2021).

Aunque los estudios se han enfocado principalmente en la relación de estas variaciones con el desarrollo y aparición de la enfermedad, estos conocimientos también se han aplicado en la elaboración de tratamientos. Se ha demostrado que la deleción del exón 3 o 4, se puede reparar eliminando el exón que se encuentre en el gen (3 o 4), dependiendo de la variante, lo que promueve la transcripción del gen y el buen funcionamiento de la proteína (Asai et al., 2010; Li et al., 2020). Así mismo, el conocimiento de estas formas ha sido utilizado para generar células madre portadoras de mutaciones y alteraciones del gen, como por ejemplo deleción del exón 7

(Castelo Rueda et al., 2022), otra en el exón 3 (Zanon et al., 2019), una delección del exón 2 (Shuvalova et al., 2020), una mutación compuesta en el exón 6; para estudiar el desarrollo de la enfermedad.

Por otro lado, se han identificado cambios puntuales en el gen, relacionados con la EP, los cuales se correlacionan con otros marcadores moleculares de fácil acceso (como el suero sanguíneo). Esto se encontró en un estudio donde analizaron diferentes polimorfismos del gen, y los relacionaron con concentraciones de hierro, cobre y zinc (Sanyal et al., 2020). El estudio de los elementos inmersos en los intrones se ha determinado en otros genes que contienen elementos de regulación de la transcripción y de la traducción como es gen de la Distrofina o los exones, cuya expresión es por procesamiento alternativo o por algún estímulo o en el ciclo de vida. Sin embargo, actualmente gracias a la secuenciación del genoma humano y de otras especies se ha logrado tener una mejor comprensión de las enfermedades ligadas a la herencia ya que actualmente no es necesario la clonación y secuenciación tradicional, lo que limitaba el desarrollo de la ciencia básica y aplicada a la medicina. Ejemplo de lo anterior se ha empleado al síndrome de Axenfeld-Rieger (ARS) que es una enfermedad ocular autosómica dominante la cual está relacionada con los genes *PITX2* (*pituitary homeobox protein 2*) y *FOXC1* (*forkhead box C1*) (Walter et al., 2020).

Aun cuando las tecnologías de secuenciación de ARNm en forma de transcriptomas o exomas y de genomas continúan siendo un reto científico ya que aún es necesario la identificación de nuevos elementos en los genes que aún no han sido establecidos. Una de las causas de estos eventos es la capacidad de secuenciación y se requiere tener una clara identificación de las regiones de inserción y delección de los genes para el desarrollo de los sistemas de diagnóstico en los pacientes (Barseghyan et al., 2017).

Desde el descubrimiento de los intrones en el año 1977, primero en genomas virales y posteriormente en genomas eucariontes, comenzó una nueva era en el estudio de la regulación de la expresión génica, ya que fue indispensable determinar cuáles, eran los procesos moleculares y celulares que se requieren para conformar un ARNm que sea traducido en los ribosomas, es decir, que las regiones codificantes denominadas exones se unan correctamente para así poder obtener una proteína.

Los estudios posteriores al descubrimiento de los intrones mostraron que existen secuencias denominadas sitios donantes y aceptores de procesamiento y un punto de ramificación (BP) ubicado cerca del 3'. En donde, estas regiones están al inicio y final de cada intrón además de los exones que lo flanquean (Kumari et al., 2022). Este proceso se realiza por medio de un conjunto de ribo-nucleoproteínas, proteínas auxiliares y otras regulatorias denominadas complejo de corte y empalme las cuales actúan de una forma ordenada y sistemática para la eliminación del intrón (Black et al., 2023; Kretova et al., 2023).

Puesto que se ha determinado que las secuencias de procesamiento de los intrones y exones tienen elementos conservados en la escala filogenética, el uso de proceso bioinformáticos permite identificar dichas regiones de manera predictiva y con ello posteriormente validar su expresión biológica (Shiraishi et al., 2022). Este poder bioinformático se ha aplicado también en secuencias que constituyen a los promotores, es decir, la región de un gen a la cual se unen diversos factores de la transcripción para que la RNA polimerasa II comience la transcripción del gen y de forma simultánea la eliminación de los intrones para la generación del ARNm maduro (Jian et al., 2014).

Con esta visión, y dada la importancia que tiene el gen *PRKN* de humano en la EP y en la mitogafia, se genera la oportunidad del estudio de los elementos transcripcionales y traduccionales, que aún no han sido determinados en el genoma humano. La generación de este conocimiento

permitirá el establecer estrategias experimentales para la identificación de exones y ORFs en caso de existir, y con ello nuevos sistemas de diagnóstico para pacientes que sufren la EP o los daños generados por agentes virales como fueron los afectados en la pandemia de SARS-CoV-2 (Faquih et al., 2023).

1. Objetivos

1.1 Objetivo de la pasantía

Conocer por métodos *in silico* las secuencias específicas de unión para factores de la transcripción (SEUFT) y posibles exones en los intrones del gen *PRKN* en humano.

1.2 Objetivo General

Determinar *in silico* la presencia de secuencias específicas de unión para factores de la transcripción (SEUFT) y posibles exones en los intrones gen *PRKN* humano para establecer sus posibles variantes de ARN mensajeros.

1.3 Objetivos Específicos

Caracterizar por procedimientos bioinformáticos la presencia de secuencias específicas de unión para factores de la transcripción (SEUFT) en los intrones gen *PRKN* humano.

Identificar por procedimientos *in silico* posibles exones dentro de los intrones del gen *PRKN* humano.

Establecer los posibles marcos abiertos de lectura de los intrones del gen *PRKN* humano.

2. Competencias

2.1 Competencias Cognitivas

Obtiene secuencias de genes reportados en bases de datos públicas para el gen *PRKN*.

Identifica la presencia de secuencias específicas de unión para factores de la transcripción (SEUFT) para los intrones del gen *PRKN* humano. Caracteriza intrones y exones del gen *PRKN* para humanos.

Establece posibles marcos abiertos de lectura de los exones codificados por el gen *PRKN* humano.

2.2 Competencias Actitudinales

Desarrolla habilidades para búsquedas bioinformáticas en bases de datos públicas relacionadas a la biología molecular.

Genera pensamiento crítico y analítico para resultados bioinformáticos. Produce valores y principios éticos para la investigación científica y aplicada.

Genera informes y reportes científicos para la correcta transferencia del conocimiento a la sociedad.

3. Metodología

3.1 Identificación del gen y las variantes de ARN mensajero de *PRKN Homo sapiens* en la base de datos Gene del NCBI

La secuencia del gen *PRKN* de *H. sapiens* se obtuvo en la base de datos Gene del Centro Nacional de Información Biotecnológica (NCBI) (Geer et al., 2010) con dirección electrónica <https://www.ncbi.nlm.nih.gov/gene/5071>. Se emplearon como criterio de búsqueda las “palabras claves”: *PRKN* y homo.

La información obtenida mediante el GenBank muestra las anotaciones actualizadas del gen, referente a: su posición respectiva en el cromosoma 6, los tejidos en los que se ha determinado su expresión y los códigos de las variantes de ARNm de *PRKN*. Además, independientemente del tipo de ácido nucleico de *PRKN*, todas las secuencias se descargaron en formato FASTA para su posterior alineamiento entre el gen y cada ARNm con su respectivo ORF.

3.1.1 Delimitación de intrones del gen *PRKN*

Para establecer las regiones del gen *PRKN* que corresponden a exones e intrones entre las variantes de ARNm de *PRKN* identificadas por el procedimiento anterior, se realizaron alineamientos múltiples entre la secuencia del gen *PRKN* con las variantes V1, V2 y V3 de ARNm, mediante la herramienta Kalign (Madeira et al., 2019), con los parámetros del Modelo de Pérdida de ADN (DNA-LM) (Martínez-Pérez et al., 2002, 2007). El resultado, se guardó en un archivo de texto plano para su edición.

Para esta pasantía las regiones del gen en el alineamiento que no presentaron identidad con uno o más regiones de los ARNm fueron consideradas intrones y las que sí, correspondieron a los exones, es decir, las regiones idénticas tanto en el gen y ARNm de las variantes. A cada exón se

colocó su traducción conceptual la cual se generó con el software Translate de la plataforma ExPASy (<https://web.expasy.org/translate/>) y la delimitación se realizó de manera manual (Gasteiger, 2003). Cuando se presentó discordancia biológica en el alineamiento, en otras palabras, que no se cumpliera la concordancia entre las secuencias del ARNm con el gen, se cambiaron los parámetros de la matriz de alineamiento hasta obtener la mejor afinidad biológica, es decir, la misma posición entre los nucleótidos del ARNm con el gen.

Para corroborar las posiciones obtenidas de los alineamientos y las nuevas variantes reportadas posteriores al alineamiento, se realizó la delimitación manual de las posiciones de exones indicadas en el GenBank para el gen *PRKN* (Sayers et al., 2019). El resultado se tabuló en una hoja de cálculo para facilidad del manejo de la información.

3.2 Validación *in silico* de exones presentes en los intrones del gen *PRKN*

La verificación de los sitios de procesamiento de ARNm, sitios potenciales de exones en sus regiones 5' → 3', se determinaron con los softwares AUGUSTUS (Hoff y Stanke, 2013; Stanke et al., 2006); (<https://bioinf.uni-greifswald.de/augustus/submission.php>) y GenScan (Burge y Karlin, 1997); (<http://hollywood.mit.edu/GENSCAN.html>); mientras que los posibles sitios de poliadenilación con el software PolyApred (Ahmed et al., 2009); (<https://webs.iiitd.edu.in/raghava/polyapred/index.html>). Además, la determinación de secuencias en tándem y microsatélites de alguna región se empleó el programa *Tandem repeats finder* (Benson, 1999) (https://tandem.bu.edu/trf/advanced_submit); cuando se consideró pertinente. Además, la representación gráfica de los ORFs representativos de los intrones se realizó con el software Geneious (Kearse et al., 2012).

3.3 Caracterización de secuencias específicas de unión para factores de la transcripción (SEUFT) en intrones

Los elementos obtenidos en los pasos anteriores se ubicaron de forma manual dentro de cada intrón en la secuencia consenso. Se determinaron los SEUFT, mediante el establecimiento de los elementos canónicos de promotores eucariontes, elementos de sitios funcionales y regulatorios, a partir de los softwares GenScan y Softberry (<http://www.softberry.com/berry.phtml?topic=tsswygroup=programsysubgroup=promoter>).

3.3.1 Caracterización funcional de exones y marcos abiertos de lectura en el gen PRKN

Con los ORFs obtenidos mediante el paso anterior, se hizo la traducción de nucleótidos a aminoácidos, mediante la herramienta Translate de la plataforma ExPASy (Gasteiger, 2003). Además, para determinar los elementos funcionales de cada proteína, los ORFs generados de cada intrón se compararon con las bases de datos de dominios de proteínas PROSITE (Sigrist et al., 2013) (<https://prosite.expasy.org/>).

4. Resultados

4.1 Identificación de las secuencias del gen *PRKN* y ARNm en la base de datos Gene.

En la base de datos *Gene* se determinó que el gen *PRKN* tiene el código de identificación 5071, con el nombre oficial *PRKN* RBR E3 ubiquitin protein ligase y con el símbolo oficial empelado en esta pasantía. Se reporta en dicha página que el gen *PRKN* está en el cromosoma 6. Dentro de la información que suministra también se indican los ARNm caracterizados del gen *PRKN* (Figura 1, Apéndice A).

Figura 1

Imagen de la base de datos de GenBank correspondiente al gen PRKN

Name/Gene ID	Description	Location	Aliases	MIM
<input checked="" type="checkbox"/> PRKN ID: 5071	parkin RBR E3 ubiquitin protein ligase [Homo sapiens (human)]	Chromosome 6, NC_000006.12 (161347417..162727766, complement)	AR-JP, LPRS2, PARK2, PDJ	602544
<input type="checkbox"/> TP53 ID: 7157	tumor protein p53 [Homo sapiens (human)]	Chromosome 17, NC_000017.11 (7668421..7687490, complement)	BCC7, BMFS5, LFS1, P53, TRP53	191170
<input type="checkbox"/> HIF1A ID: 3091	hypoxia inducible factor 1 subunit alpha [Homo sapiens (human)]	Chromosome 14, NC_000014.9 (61695513..61748258)	HIF-1-alpha, HIF-1A, HIF-1alpha, HIF1, HIF1-ALPHA, MOP1, PASD8, BHLHe78	603348
<input type="checkbox"/> BRCA1 ID: 672	BRCA1 DNA repair associated [Homo sapiens (human)]	Chromosome 17, NC_000017.11 (43044295..43125364, complement)	BRCAI, BRCC1, BROVCA1, FANCS, IRIS, PNCA4, PPP1R53, PSCP, RNF53	113705

1: NG_008289.2 RefSeqGene
Range: 5037..1385386
Download: [GenBank](#), [FASTA](#), [Sequence Viewer \(Graphics\)](#)

mRNA and Protein(s)

1: NM_004553.3 → NP_004553.2 E3 ubiquitin-protein ligase parkin isoform 1
See identical proteins and their annotated locations for NP_004553.2

Status: REVIEWED

Description: Transcript Variant. Transcript variant 1 represents the predominant and full-length form of this gene.
Source sequence(s): [AK292590.AL132982.DB023187](#)
Consensus CDS: [CCDS5281.1](#)
UniProtKB/Swiss-Prot: [Q60260](#), [Q8WV07](#)
UniProtKB/TrEMBL: [X5DR79](#)
Related: [ENSP00000355865.1](#), [ENST00000366898.6](#)

Conserved Domains (4) **summary**

cd01798	parkin_N, amino-terminal ubiquitin-like of parkin protein
Location:3 → 72	
smart00213	UBQ, Ubiquitin homologues
Location:1 → 72	
smart00647	IBR, In Between Ring fingers
Location:313 → 375	
pfam01485	IBR, IBR domain, a half RING-finger domain
Location:413 → 453	

Nota. La imagen indica la ubicación y acceso a la información del gen *PRKN* de *H. sapiens*.

Además, en la base de datos se indica que el gen *PRKN* en el genoma humano, está en monocopia y contiene 1.380.350 pb. En tanto para los ARNm que codifica, se identificaron 3 variantes, las cuales se nombraron como: variante 1 (V1), que contenía 12 exones, la variante 2 (V2), que presentó un exón menos respecto a la variante anterior y la variante 3 (V3) con 9 exones. Todas ellas mostraron un rango de 3731 a 5129 pb (Tabla 1).

También se determinaron 3 variantes adicionales, las cuales estaban denominadas como variante X1 y X2 que contenían 12 exones y una variante X3 con un exón menos que las antes indicadas (Tabla 1).

Tabla 1

Variantes de ARNm del gen PRKN en la base de datos GENE

Variante	N° Exones	bp	N° acceso
1	12	4178	NM_004562.3
2	11	4094	NM_013987.3
3	9	3731	NM_013988.3
X1	12	5129	XM_017010908.2
X2	12	4175	XM_011535863.2
X3	11	4014	XM_024446449.2

Nota. La variante 1, V1 es la más completa reportada con 12 exones, la variante 2 V2 cuenta con 11 exones, la variante 3 V3 cuenta con 9 exones. Las variantes X1 y X2 contiene los 12 exones y la variante X3 cuenta con 11 exones.

Un aspecto para resaltar con la fecha inicial a la obtención de los datos reportados en el GenBank se determinó en la posición de los exones de cada variante, el cual cambió en marzo de 2023 como se evidenciará más adelante en este informe. Para esta pasantía, se consideraron únicamente las posiciones de las variantes iniciales y no las posteriores al análisis, ya que como se demostrará en los siguientes resultados, los datos iniciales del GenBank no presentaban una base

molecular consistente a una realidad biológica y que, a la fecha de este informe, lo que actualmente reporta esta base de datos sí concuerda con los resultados iniciales (Apéndice B).

4.1.1 Delimitación de intrones y exones del gen PRKN

El alineamiento tipo ClustalW del gen *PRKN* y sus seis secuencias de ARNm con los parámetros propuestos por el diseñador del software, no generaron un alineamiento acorde a una realidad biológica (Apéndice C.1). Mientras que el mismo procedimiento sin el gen y con las 6 variantes del ARNm, se determinó que la variante X3 presentaba un orden nucleotídico distinto a las otras variantes (Apéndice C.2).

De igual forma, al realizar el alineamiento entre los ARNm de las 3 variantes (V1, V2 y V3) y los ARNm X1 y X2 de *PRKN* se determinó que el ARNm X1 presentaba una mayor diferencia en las primeras 1080 pb del extremo 5' (Apéndice C.3). Finalmente, el alineamiento entre el ARNm de la variante X2 de *PRKN* respecto a las 3 variantes mostró que esta última era similar a la variante V1 con la única diferencia de dos nucleótidos dentro del ORF (Apéndice C.4). Por lo tanto, la variante X2 se consideró idéntica a la V1 y la determinación de la posición de los exones en el gen *PRKN* se realizó únicamente con las variantes de ARNm V1, V2 y V3.

El alineamiento entre ellas mostró la pérdida de regiones en los ARNm de las variantes V2 y V3 lo que confirmó los tamaños reportados mostrados en el Apéndice B y sugirió que la pérdida respecto al ARNm de la variante V1, correspondía a los exones reportados (Apéndice C.5).

El alineamiento entre el gen *PRKN* y las 3 variantes de ARNm, mostraron la identidad de las regiones del gen y sus correspondientes segmentos de ARNm de cada una de las variantes; lo que confirmó que las regiones del gen que no contenían homología respecto al ARNm podían ser considerados como intrones. Mientras que, la ausencia de un segmento de alguna de las variantes V2 y V3 correspondían al procesamiento alternativo en el ARNm de cada una (Apéndice C.6).

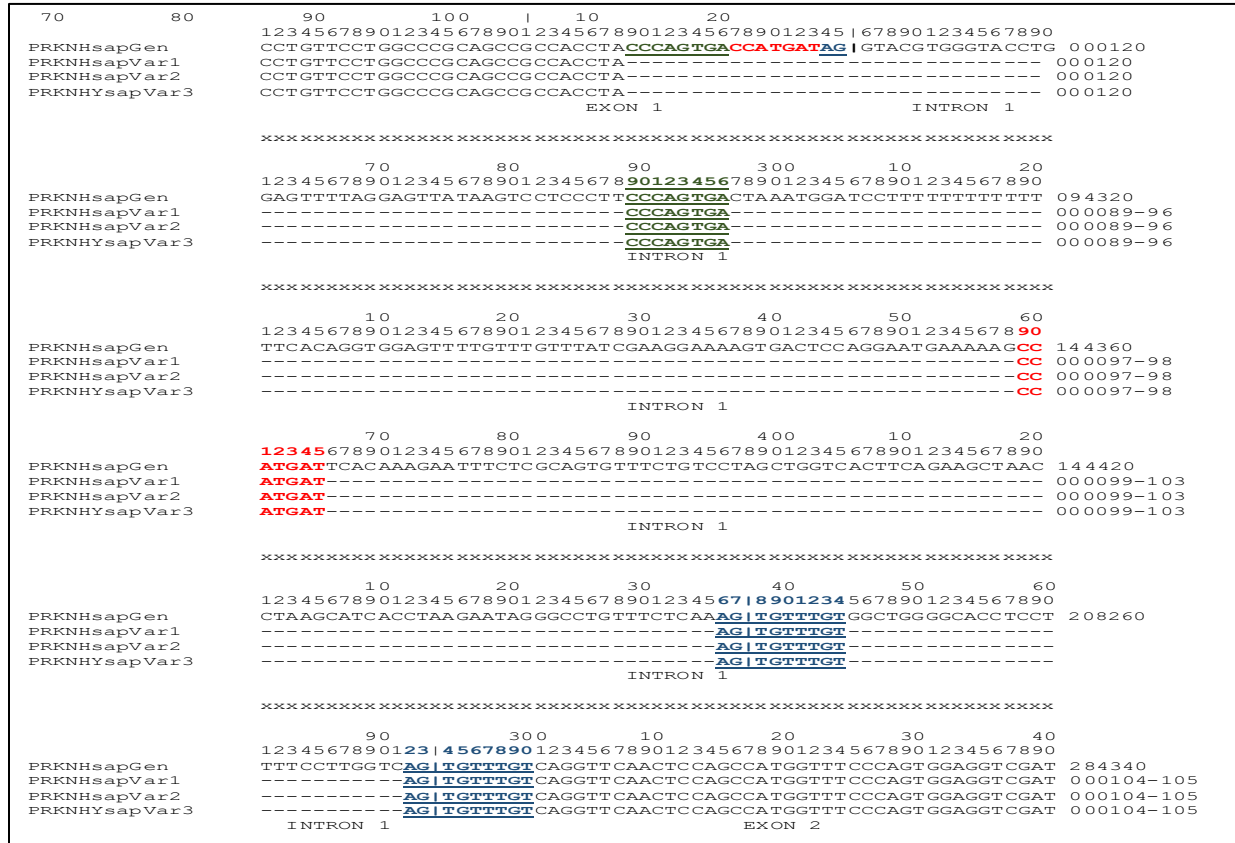
Sin embargo, al verificar la posición de cada región de identidad respecto a las posiciones reportadas por el GenBank en ese momento, los resultados indicaron una discordancia entre la posición reportada respecto a la región de identidad (Apéndice C.6). En relación con lo anterior, la posición de inicio del exón 1 indicado por el GenBank comenzaba a partir de la posición 1 hasta la posición 105 y así sucesivamente hasta su codón de paro. No obstante, este orden numérico no se cumplía para las variantes V1, V2 y V3, ya que estas presentaban posiciones diferentes en los exones reportados por el GenBank (Apéndice B).

En esta región común considerada como el primer exón para las tres variantes (posición 1 hasta la posición 105), los primeros 98 nucleótidos presentaron identidad con el exón del gen, pero los 17 nucleótidos restantes se distribuyeron en tres regiones del intrón 1. De estas, la segunda correspondía al codón de inicio de la traducción y a los dos primeros nucleótidos del siguiente codón, el cual estaba en la tercera región junto con los siete primeros nucleótidos del exón 2, referidos por el GenBank. (Figura 2 Apéndices C.6 y C.7).

Sin embargo, el alineamiento manual demostró que la posición colocada en el GenBank sí correspondía a lo esperado, por lo tanto, para esta etapa de la pasantía el exón 1 se localizó en las posiciones indicadas en el GenBank (posiciones 1-105) (Tabla 2 Anexo B). Así mismo, desde la posición 106 inicia el intrón 1 (Figura 3).

Figura 2

Alineamiento entre el primer exón y las variantes del ARNm de PRKN



Nota. Las letras verdes corresponden a los nucleótidos del exón 1, las rojas indican la región del codón de inicio de la traducción y los dos primeros nucleótidos del segundo codón; Se observa que la tercera región esta duplicada (letras azules), la primera región está dentro del intrón y la segunda corresponde al último nucleótido del segundo codón y el primer nucleótido del codón que continúa en el exón 2 que está separado por la barra azul.

Figura 3

Alineamiento manual entre el primer exón y las variantes del ARNm de PRKN

	EXON 1/ 5' UTR			
PRKNHsapGen	AGGATTAA	CCAGGAGAGCCGCTGGTGGGAGGCGCGGCTGGCGCCGCTGCGCGCATGGGCCTGTTCCCTG	0000	60
PRKNHsapVar1	AGGATTAA	CCAGGAGAGCCGCTGGTGGGAGGCGCGGCTGGCGCCGCTGCGCGCATGGGCCTGTTCCCTG	0000	60
PRKNHsapVar2	AGGATTAA	CCAGGAGAGCCGCTGGTGGGAGGCGCGGCTGGCGCCGCTGCGCGCATGGGCCTGTTCCCTG	0000	60
PRKNHsapVar3	AGGATTAA	CCAGGAGAGCCGCTGGTGGGAGGCGCGGCTGGCGCCGCTGCGCGCATGGGCCTGTTCCCTG	0000	60
	EXON 1/ 5' UTR	ORF	INTRON 1	
PRKNHsapGen	GCCCGCAGCCGCCACCTACCCAGTGAC	<u>ATGATAG</u>	GTACGTGGGTACCTGCCAGGTACAGCCTCTCTGCG	000120
PRKNHsapVar1	GCCCGCAGCCGCCACCTACCCAGTGAC	<u>ATGATAG</u>		
		<u>M I</u>		
PRKNHsapVar2	GCCCGCAGCCGCCACCTACCCAGTGAC	<u>ATGATAG</u>		
		<u>M I</u>		
PRKNHsapVar3	GCCCGCAGCCGCCACCTACCCAGTGAC	<u>ATGATAG</u>		
		<u>M I</u>		

Nota. En letra negra y subrayada se indican los primeros dos codones de inicio de la traducción y el primer nucleótido del tercer codón que continúa en el exón 2. Se considera que los nucleótidos río abajo de esta región corresponden al extremo 5' no traducido del ARNm.

Este proceso se realizó sucesivamente de manera manual, mediante el conteo del número de pb de cada exón dentro de las secuencias de cada variante. Para la variante uno, con este método, se determinó 12 exones, pero el GenBank reportaba 11; en tanto que la variante dos tenía 11 exones acorde a lo reportado por el GenBank, mientras que en la variante 3 se establecieron 9 exones, pero la base de datos reportaba 8 exones.

Esta discordancia entre la ubicación de exones y las secuencias usadas de cada variante se observó en los primeros alineamientos entre el gen y las tres variantes de ARNm de *PRKN* con los parámetros del modelo de pérdida de ADN y otros de forma aleatoria (Apéndice C.6 y C.7). Los resultados siempre mostraron una inconsistencia con el primer exón el cual contenía el extremo 5'UTR y el primer codón de metionina de inicio de la traducción en el intrón 1, lo cual era contrario a lo reportado en el GenBank en ese momento, ya que se reportaba como el exón 1 al segundo

exón. Lo anterior tuvo que ser demostrado mediante un alineamiento manual y no como un algoritmo (Apéndice B).

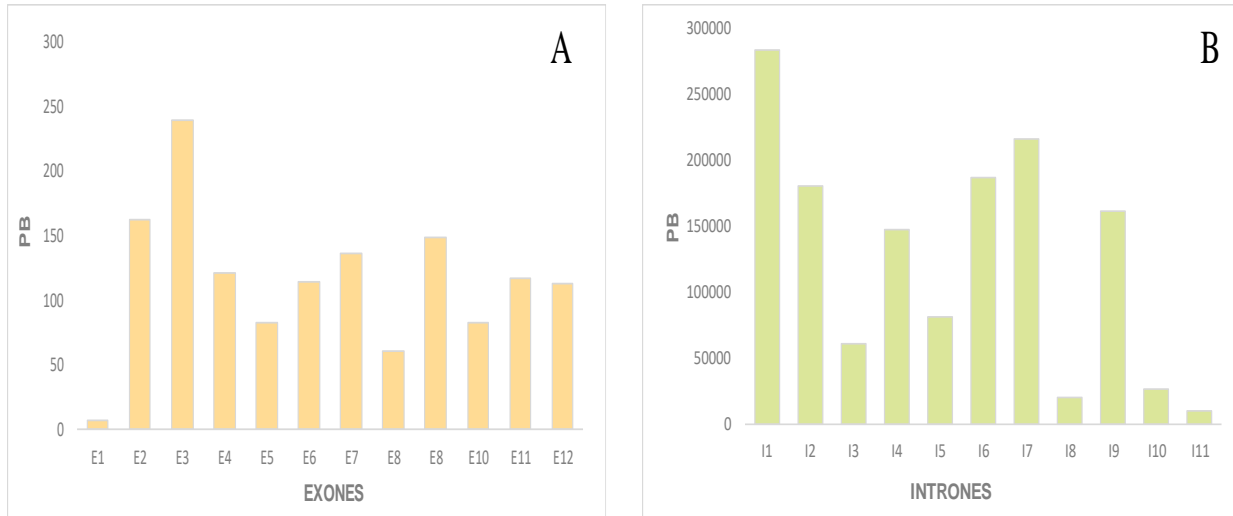
Estas inconsistencias entre la nomenclatura de las posiciones también se presentaron para los subsecuentes exones. Las 3 variantes contenían similitud hasta el exón 2 y a partir del exón 3 hubo inconsistencias en cada una de ellas, mostrando que los exones se interpolaban en extremo 3'→5' de unos con otros hasta el exón 12. Es decir, el final de los intrones de la Var 2 eran el inicio de los intrones de la Var 1. También se observó que la variante 3 tenía relación con las variantes 1 y 2, el exón 4 de la V3 era el final del exón 6 de la V2, el final del exón 10 de la V1 correspondía al final del exón 9 de la V2 y al final del exón 7 de la V3. El inicio del exón 11 de la V1 era el inicio del exón de la V3, y el inicio del exón 12 de la V1 es el inicio del exón 11 de la V2.

Para dar una solución a lo anterior, se tabularon las posiciones de los intrones y exones del gen, así como los exones de cada variante y se compararon con los reportados en el GenBank lo que corrobora la inconsistencia. Sin embargo, a marzo de 2023 el GenBank modificó las posiciones de los exones del gen y mostró el resultado inicial de la delimitación de los intrones y exones de esta pasantía (Apéndice B).

La graficación del resultado del límite de los intrones y exones en el gen *PRKN* (Apéndice B) mostró 3 tipos de intrones en función de su tamaño; los intrones 1, 2, 6, 7 y 9 tuvieron más de 150.000 pb. En tanto que los intrones 3, 4 y 5 su tamaño fue de 50.000 pb hasta menos de 150.000 pb y los intrones 8, 10 y 11 fueron menores a 27.000 pb. Los exones 5, 8 y 10 tuvieron un tamaño menor a 100 pb y los exones 2 y 3 fue mayor al valor anterior. El resto de los exones 1, 4, 6, 7, 9, 11 y 12 tuvieron un rango de 100 pb a menos de 150 pb (Figura 4A y 4B).

Figura 4

Gráfica del tamaño de los exones e intrones del gen PRKN



Nota. En la gráfica “A” se muestra el tamaño de los exones y en la gráfica “B” de los intrones del gen *PRKN*. Es de resaltar el tamaño de los intrones respecto a los exones del gen, además, de la tendencia nucleotídica mayor al extremo 5’ respecto al extremo 3’.

Por todo lo anterior, se puede considerar que la mayor cantidad de pb del gen se encuentra desde el exón 1 hasta el exón 7 incluyendo los intrones, mientras que la región minoritaria del gen es desde el exón 8 hasta el final del exón 12 con sus intrones.

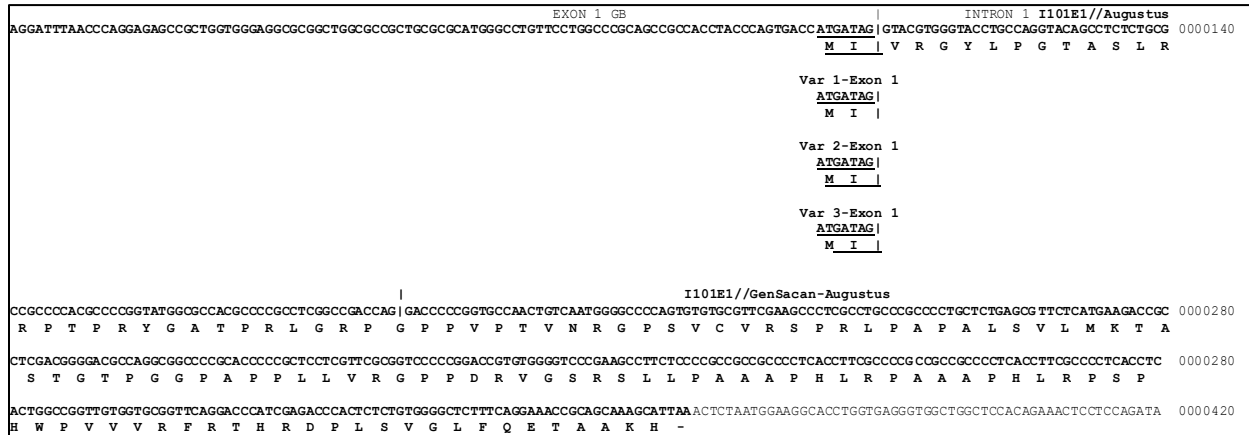
4.2 Validación *in silico* de exones en los intrones del gen *PRKN*

La determinación de posibles exones y/o ORFs o sitios activos de la traducción dentro de cada intrón con el programa AUGUSTUS obtuvo una predicción limitada de cada uno de los intrones, mostrando que únicamente el intrón 1 y 7 presentaron ORFs. Interesantemente el ORF 1 comienza desde el primer nucleótido del intrón 1 pero el codón de inicio es GTA para Valina, el cual está contiguo del último nucleótido del exón 1 que corresponde al tercer codón del primer ORF reportado para las 3 variantes. El ORF 1 del intrón 1 presenta 132 codones junto con el codón

de paro TAA localizado en el nucleótido 468 (Figura 5). Además, se determinó un segundo ORF dentro del intrón 1 conformado por cuatro exones y 5 sitios de poliadenilación (Apéndice D).

Figura 5

Primer Marco abierto de lectura del intrón 1 del gen PRKN



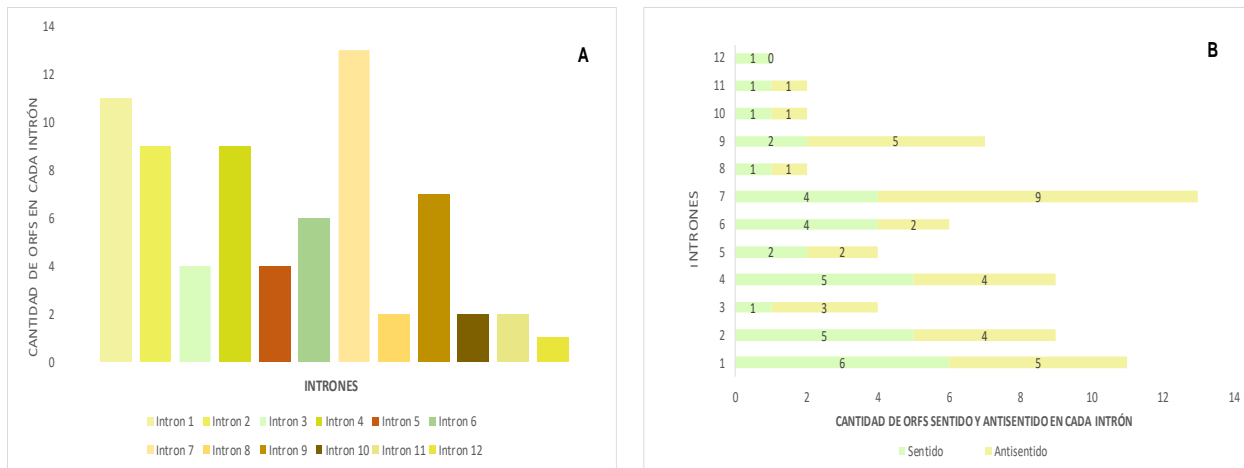
Nota. El primer exón es separado por el primer intrón por una barra (|), donde se muestra de manera subrayada el codón de inicio de la traducción, el segundo codón y el primer nucleótido del tercer codón. Inmediatamente se indica el ORF determinado con el Software AUGUSTUS y en la segunda barra se muestra el ORF encontrado con el software GenScan, el cual se sobrepone con el ORF determinado con el Software AUGUSTUS. El codón de paro se indica con una línea horizontal en la parte inferior del codón TAA.

A diferencia del intrón 1, el software AUGUSTUS determinó únicamente un ORF en el intrón 7, compuesto por un exón de 381 codones al incluir los codones de inicio y paro TGA de la traducción, además, de siete sitios potenciales de poliadenilación (Figura 6). Una de las características de este ORF la presencia de elementos repetidos en tándem; en los codones para los aminoácidos NAT y MQH, los cuales se repiten 19 y 20 veces respectivamente (Figura 6; Apéndice D).

de ORFs sentido y antisentido, fueron los intrones 5, 8, 10 y 11; únicamente el intrón 12 no presentó un ORF en dirección antisentido, por lo tanto, se identificaron 34 ORFs en dirección 5'→3' (Figura 7B).

Figura 7

Cantidad de ORFs sentidos y antisentidos del gen PRKN



Nota. La cantidad total de los ORFs determinados en los intrones del gen *PRKN* se muestra en el gráfico “A”. Mientras que la cantidad de ORFs en dirección sentido (5' → 3') y antisentido (3' → 5') se observan en el gráfico “B”; los números en las barras indican la cantidad de ORFs.

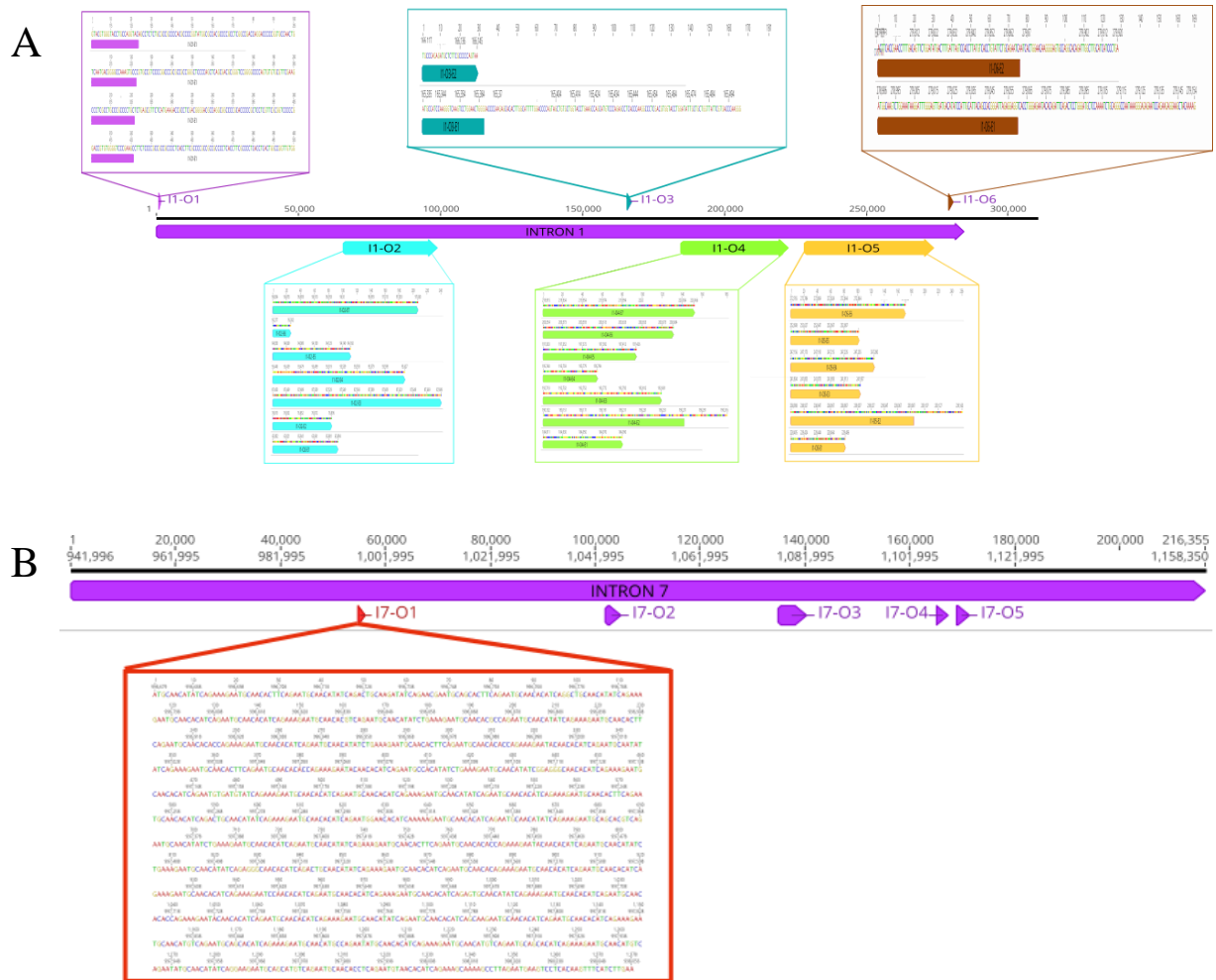
Otro aspecto por resaltar es que los ORFs en los intrones 10, 11 y 12 tuvieron 1 solo exón, el resto de los intrones tuvieron entre 3 a 9 exones por ORF, como es el caso del intrón 9. Posterior al codón de paro de todos los ORFs determinados se presentó una o más señales de poliadenilación. Por lo tanto, se determinó un total de 128 exones dentro de los intrones del gen *PRKN*. Cabe resaltar que la distribución de cada ORF fue de forma aleatoria dentro del intrón y no se determinó algún tipo de patrón independientemente del tamaño de cada intrón.

Lo anterior, se observa en la Figura 8 dónde se indica la distribución de los ORFs dentro de los intrones 1 y 7. Aunque el programa AUGUSTUS arrojó predicciones únicamente para estos

dos intrones, los resultados fueron similares a los obtenidos por el programa GenScan, el cual presentó una cobertura más amplia en los demás intrones (Figura 5 y 6; Apéndice D).

Figura 8

Distribución de los exones concomitantemente ORFs en los intrones 1 y 7



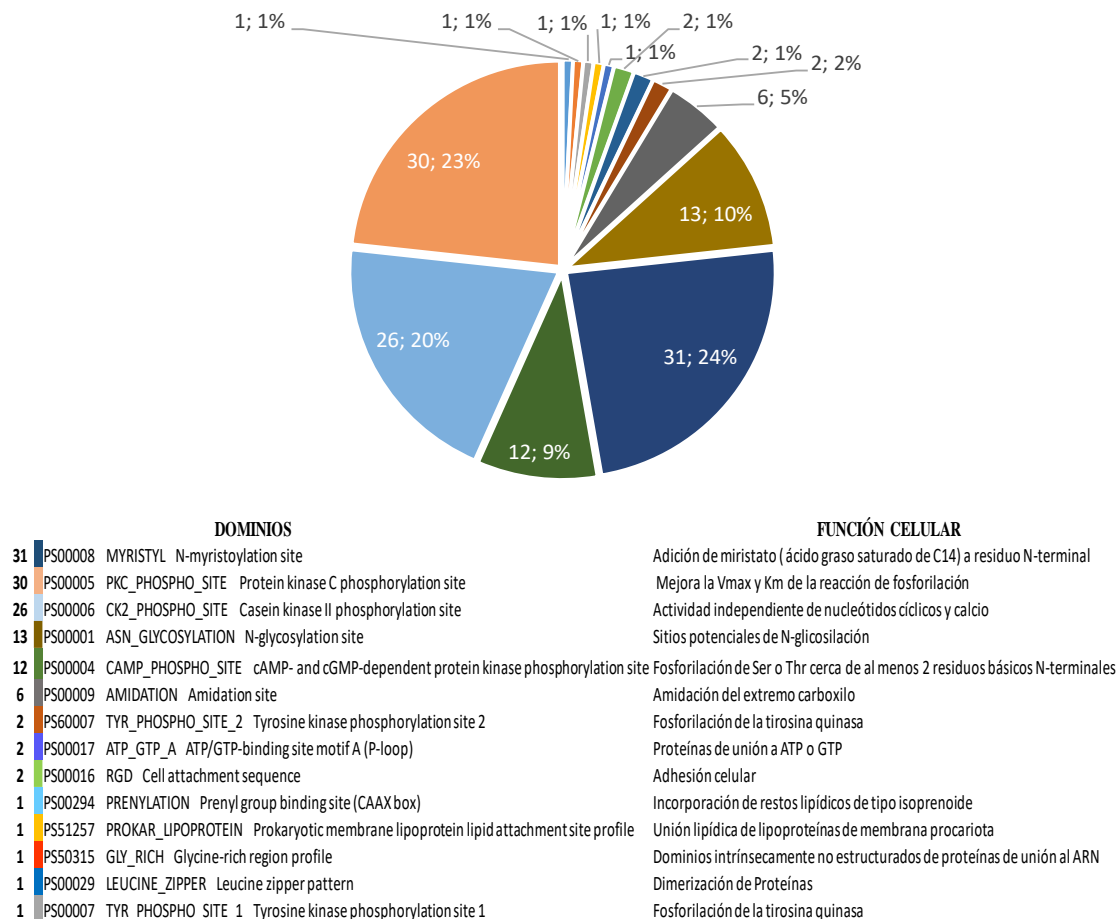
Nota. En el gráfico “A” se muestra la estructura de los 6 ORFs encontrados a partir del programa GeneScan, en el intrón 1. Al interior de cada recuadro, se encuentran las posiciones de los exones correspondientes a cada ORF. Mientras que en “B” se representan los ORFs encontrados a partir del programa GeneScan, en el intrón 7. En el recuadro rojo se observa la estructura del exón 1, del ORF 1, donde se identifica una posible repetición en tándem.

4.3 Características funcionales de los exones y marcos abiertos de lectura en los intrones del gen *PRKN*

El análisis de las secuencias de aminoácidos codificada para cada exón mostró respecto a la base de datos PROSITE que el 56.3 % de los ORFs codifican para elementos de regulación por fosforilación; mientras que el 27.3 % corresponde a modificaciones por la incorporación de un ácido graso o interacciones con la membrana plasmática o con la lámina basal. El 15 % codifica para sitios de N-Glicosilación o la amidación del extremo carboxilo, el 4.8 % para una región de amidación y el 1.6 % para dominios de dimerización o interacción con el ARN (Figura 9).

Figura 9

Caracterización de los dominios de los ORFs de los intrones del gen PRKN



Nota. El primer número corresponde a la cantidad de dominios encontrados y su respectivo color indicativo en la tabla superior. El siguiente número corresponde al código de los dominios en la base de datos PROSITE.

Teniendo en cuenta lo anterior, en el Apéndice E, todos los ORFs de los intrones contenían al menos un exón que codifica para un sitio de fosforilación, a excepción de los ORFs 4 y 1 de los intrones 4 y 8 respectivamente; todos los ORFs determinados dentro de los intrones presentaron un exón que codifica para la miristilación de la proteína, es decir, la adición de un ácido graso de 14 carbonos; mientras que el ORF 3 del intrón 7 presentó un exón que codifica para un sitio de unión de anclaje celular (Apéndice E).

4.3.1 Caracterización de secuencias específicas de unión para factores de la transcripción (SEUFT)

Mediante el software GenScan, se determinaron un total de 23 regiones promotoras de 40 pb río arriba de los ORFs de los intrones 1, 2, 3, 4, 5, 6, 7 y 9. Mientras que los otros ORFs determinados no presentaron estos elementos de sitios de unión a factores de la transcripción. Es de resaltar que, dada la cantidad de elementos promotores y el análisis bioinformático de cada uno de ellos, se limitó a colocar únicamente su posición, la cual se muestra en el Apéndice F. Los resultados con los softwares Softberry y CiiDER mostraban elementos promotores de otras especies diferentes a *H. sapiens* y por ello se descartaron del análisis de la pasantía.

5. Discusión

La secuenciación del genoma humano permitió comprender infinidad de genes y elementos regulatorios de la fisiología, y en algunos de ellos su papel en diversas enfermedades ligadas a la herencia como es la EP (Nalls et al., 2019). A pesar de este logro, una de las limitantes del uso de la información genómica reportada en la base de datos del GenBank, es que muchos genes han sido caracterizados por procedimientos bioinformáticos y se requiere la validez teórica y experimental tanto de la organización de las regiones que conformarán a sus promotores y la delimitación de intrones y exones, que darán paso a los ARNm generados por el procesamiento de ARN en caso de que tengan más de dos exones (Antunes, 2022). Esta limitante la pudimos observar en la primera fase de esta pasantía, donde la organización de intrones y exones para el gen *PRKN* propuestos por el GenBank no presentaban una consistencia en las posiciones.

Una ventaja bioinformática que ha contribuido a conocer la fisiología y propiedades de las proteínas, son las bases de datos proteicos, cuyos algoritmos matemáticos permiten la correlación de las nuevas regiones determinadas con las disponibles en las bases de datos, y con ello, lograr conocer la posible función celular de cada uno de los motivos y dominios que conforman a la proteína encontrada (Sayers et al., 2019).

Con base a lo anterior, una de las principales enfermedades de las que aún se desconoce su origen fisiológico es la EP (Antunes, 2022). Sin embargo, se ha demostrado que existen genes que contribuyen al desarrollo de la enfermedad en distintas edades del paciente, como lo es el gen *PRKN*, el cual participa en la mitogafia y sus mutaciones se relacionan con la EP temprana (Terešák et al., 2022; Zhu et al., 2022).

Teniendo en cuenta lo anterior, y con el fin de dar cumplimiento al objetivo del Grupo CAGE, en esta pasantía se determinó *in silico* que los intrones del gen *PRKN* de humano, presentan elementos que se pueden considerar exones y cuya traducción genera proteínas aún no descritas. La importancia de esto se explica a partir de lo reportado por el GenBank, donde se propone una organización de 12 exones y 11 intrones, sin embargo, los resultados de esta pasantía indican que se podrían generar nuevas organizaciones con importancia fisiológica.

Es de resaltar que este trabajo de haberse realizado de forma experimental habría requerido un período de tiempo de años para su ejecución. Sin embargo, aquí se demuestra la importancia de las bases datos públicas y del logro de tener el genoma humano a libre acceso, con lo cual, mediante palabras claves, se logró obtener la información referente al gen *PRKN* para su respectivo análisis, en tiempos operativos a la actualidad.

Al realizar la delimitación de cada exón reportado en la base de datos utilizada mediante criterios biológicos, en esta pasantía se demostró que en muchas de las predicciones publicadas en el GenBank no presentaban una concordancia biológica, es decir, las secuencias del ARNm de cada una de las variantes del gen *PRKN* no eran 100 % idénticas con su región genómica. Tal como se mostró, esta validación fue necesaria realizarla de manera manual ya que se determinó que los extremos 5' → 3' en la mayoría de los exones reportados, se repiten dentro del siguiente exón e inclusive, en la parte final de los intrones propuestos.

El ordenamiento inicial de la posición de cada intrón y exón reportados para el gen *PRKN* para esta pasantía, el cual era distinto al propuesto por el GenBank, fue corroborado de manera indirecta cuando esta base de datos cambió las posiciones de los elementos antes mencionados en marzo del 2023 y concuerdan con lo establecido en esta pasantía.

Se puede considerar que esta etapa fue fundamental para el desarrollo de la siguiente caracterización realizada en la investigación, ya que permitió establecer las regiones que fueron consideradas intrones para determinar la presencia de elementos promotores y de exones en cada uno de ellos.

Es de resaltar que, en un principio, se estableció que esta etapa se realizaría con las herramientas del paquete computacional Softberry y otros. Sin embargo, los resultados que mostraron este tipo de softwares incluían SEUFT de otras especies de organismos y se consideró que el colocar dichos resultados generaran sesgos en el conocimiento de dichas regiones y dada la cantidad obtenida es necesario un análisis detallado de cada uno de ellos para establecer su posible participación en el funcionamiento celular y en la EP.

Los resultados encontrados con los softwares AUGUSTUS y GenScan, permitieron establecer la presencia de 128 exones, algunos de ellos en forma individual o en grupos de más exones, que codificaban para Marcos Abiertos de Lectura aún no descritos para el gen *PRKN*. Cabe mencionar que el software AUGUSTUS mostró un menor rango de predicción que la generada por GenScan, ya que sólo se obtuvieron resultados para el intrón 1 y 7.

Por lo tanto, se puede afirmar que en esta pasantía se realizó el primer análisis de elementos de regulación de la transcripción, exones y señales de poliadenilación dentro de los intrones del gen *PRKN* en humanos. Los resultados generados por GenScan, mostraron que en todos los intrones se tienen ORFs, tanto en dirección sentido como antisentido, y que muchos de ellos contenían elementos que podrían generar un transcrito de ARNm con todos los elementos que los conforman: una región promotora para la ARN polimerasa II, que codifica para un extremo 5'; uno o más intrones que pueden ser escindidos por el procesamiento del ARN, los exones que

conformarán al ORF con un codón de inicio y de paro en la traducción realizada por los ribosomas y un extremo 3' no traducido con una o más señales de poliadenilación.

Con base a lo anterior, se puede considerar que la expresión de algunos de ellos estará en función del ciclo de vida del paciente y del tejido, e inclusive, la condición metabólica (Zhang et al., 2022). Este postulado se sustenta en literatura, donde se muestra el patrón de expresión del gen *PRKN* en los tejidos de adulto y en desarrollo embrionario; siendo mayor la expresión en la glándula adrenal, el cerebro, corazón, riñón, tejido graso, testículos y próstata tal como se muestra en la página del gen *PRKN* del NCBI.

Uno de los resultados sobresalientes de la caracterización de los exones dentro de los intrones del gen *PRKN*, fue el primer ORF del intrón 7, que está conformado por secuencias en tándem, las cuales podrían funcionar como microsatélites (Kariya et al., 2005; Tan et al., 2003). Se ha demostrado que el aumento o disminución de microsatélites afectan directa o indirectamente en enfermedades ligadas a la herencia como es la enfermedad de Huntington, entre otras (Bates et al., 2015; Price et al., 1998).

Por otro lado, se encontró un posible marco abierto de lectura en el intrón 1, el cual tiene su codón de inicio en la misma posición de inicio del intrón 1. Esta región limita con el codón de paro del exón 1, lo cual abre la posibilidad de tener un nuevo extremo amino en caso de tener alguna regulación por el procesamiento del pre-ARNm en esta parte del gen.

En esta pasantía se consideró la región 3' no traducida del exón 12, como el intrón 12. En esta, se estableció un exón que contiene un codón de paro y que, en caso de tener un procesamiento alternativo respecto al exón descrito en el GenBank, la posible traducción de este exón podría generar un nuevo extremo carboxilo, lo cual podría influir en la función de la proteína.

Finalmente, la traducción de los exones hallados en los intrones presentaba al menos una región de regulación por fosforilación y de otros motivos, para que la posible proteína traducida se asocie a membranas celulares, lo cual podría contribuir a establecer la función celular de la expresión de los ARNm del gen *PRKN* en las células de pacientes con EP, además de la mitogafia. Sumado a lo anterior, la presencia de una de las regiones de asociación con el ARNm, como es el motivo rico en glicinas o la presencia de un sitio de mistirilación, el cual genera un enlace covalente que añade un grupo de miristato a su residuo N-terminal a través de un enlace amida, permitirá que la proteína que contenga este dominio interactúe con los fosfolípidos de las membranas celulares o con proteínas asociadas a membranas.

Para el caso de los pacientes con la EP, este descubrimiento es fundamental ya que permitirá establecer nuevas asociaciones membranales que aún no han sido descritas. Se puede considerar que el trabajo realizado en esta pasantía abrirá nuevas opciones para comprender la asociación del gen *PRKN* con la EP y la mitogafia, entre otras, lo cual contribuirá a un mejor diagnóstico en los pacientes, además de los conocimientos en ciencia básica al ser aplicado en organismos modelo de laboratorio o de la vida silvestre.

6. Conclusiones

Los intrones del gen *PRKN* tienen elementos para la generación del ARNm que contiene un extremo 5'UTR, un ORF y un extremo 3'UTR con su respectiva señal de poliadenilación.

El uso de herramientas bioinformáticas es clave para la caracterización de secuencias de intrones y exones para el gen *PRKN* en humanos.

El conocimiento biológico es importante a la hora del uso de herramientas bioinformáticas para obtener mejores criterios e interpretaciones de los análisis con respecto a los resultados arrojados a partir de éstas.

Se debe efectuar la caracterización experimental de las variantes de procesamiento alternativo del ARNm del gen *PRKN* con procedimientos que permitan establecer las posibles combinaciones de exones dependiendo de la condición fisiológica del paciente.

El establecimiento de los posibles ORFs de los exones en los intrones del gen *PRKN* humano, requieren una validación experimental a nivel de su transcripción y traducción.

Es necesario determinar experimentalmente la posible localización membranal de las variantes del gen *PRKN* que contengan motivos de unión a membranas celulares.

Se requiere tanto la validación *in silico* y experimental de los SEUFT en los ORFs presentes en los intrones del gen *PRKN*.

Se da cumplimiento a las competencias establecidas para la realización de esta pasantía.

Referencias Bibliográficas

- Ahmad, A., Nkosi, D., y Iqbal, M. A. (2023). PARK2 Microdeletion or Duplications Have Been Implicated in Different Neurological Disorders Including Early Onset Parkinson Disease. *Genes*, 14(3), 600. <https://doi.org/10.3390/genes14030600>
- Ahmed, F., Kumar, M., y Raghava, G. P. S. (2009). Prediction of Polyadenylation Signals in Human DNA Sequences using Nucleotide Frequencies. *In Silico Biology*, 9(3), 135–148. <https://doi.org/10.3233/ISB-2009-0395>
- Al-kuraishy, H. M., Al-Gareeb, A. I., Alexiou, A., Papadakis, M., Alsayegh, A. A., Almohmadi, N. H., Saad, H. M., y Batiha, G. E. (2023). Pros and cons for statins use and risk of Parkinson's disease: An updated perspective. *Pharmacology Research y Perspectives*, 11(2), e01063. <https://doi.org/10.1002/prp2.1063>
- Alexoudi, A., Alexoudi, I., y Gatzonis, S. (2018). Parkinson's disease pathogenesis, evolution and alternative pathways: A review. *Revue Neurologique*, 174(10), 699–704. <https://doi.org/10.1016/j.neurol.2017.12.003>
- Antunes, A. S. L. M. (2022). Post-translational Modifications in Parkinson's Disease. In V. Corasolla Carregari (Ed.), *Advances in Experimental Medicine and Biology* (1st ed., pp. 85–94). Springer Berlin. https://doi.org/10.1007/978-3-031-05460-0_6
- Asai, H., Hirano, M., Kiriyama, T., Ikeda, M., y Ueno, S. (2010). Naturally- and experimentally-designed restorations of the *PARKIN* gene deficit in autosomal recessive juvenile parkinsonism. *Biochemical and Biophysical Research Communications*, 391(1), 800–805. <https://doi.org/10.1016/j.bbrc.2009.11.141>
- Barseghyan, H., Tang, W., Wang, R. T., Almalvez, M., Segura, E., Bramble, M. S., Lipson, A.,

- Douine, E. D., Lee, H., Délot, E. C., Nelson, S. F., y Vilain, E. (2017). Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. *Genome Medicine*, 9(1), 90. <https://doi.org/10.1186/s13073-017-0479-0>
- Bates, G. P., Dorsey, R., Gusella, J. F., Hayden, M. R., Kay, C., Leavitt, B. R., Nance, M., Ross, C. A., Scahill, R. I., Wetzel, R., Wild, E. J., y Tabrizi, S. J. (2015). Huntington disease. *Nature Reviews Disease Primers*, 1(1), 15005. <https://doi.org/10.1038/nrdp.2015.5>
- Bautista, N. (2021). *Día Mundial del Parkinson: Colombia se destaca en atención*. Ministerio de Salud y Protección Social. <https://minsalud.gov.co/Paginas/Dia-Mundial-del-Parkinson-Colombia-se-destaca-en-atencion.aspx>
- Bayne, A. N., y Trempe, J.-F. (2019). Mechanisms of PINK1, ubiquitin and *PARKIN* interactions in mitochondrial quality control and beyond. *Cellular and Molecular Life Sciences*, 76(23), 4589–4611. <https://doi.org/10.1007/s00018-019-03203-4>
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27(2), 573–580. <https://doi.org/10.1093/nar/27.2.573>
- Black, C. S., Whelan, T. A., Garside, E., MacMillan, A. M., Fast, N. M., y Rader, S. D. (2023). Spliceosome assembly and regulation: insights from analysis of highly reduced spliceosomes. *RNA*, 1(1), rna.079273.122. <https://doi.org/10.1261/rna.079273.122>
- Borsche, M., König, I. R., Delcambre, S., Petrucci, S., Balck, A., Brüggemann, N., Zimprich, A., Wasner, K., Pereira, S. L., Avenali, M., Deuschle, C., Badanjak, K., Ghelfi, J., Gasser, T., Kasten, M., Rosenstiel, P., Lohmann, K., Brockmann, K., Valente, E. M., ... Klein, C. (2020). Mitochondrial damage-associated inflammation highlights biomarkers in *PRKN*/*PINK1* parkinsonism. *Brain*, 143(10), 3041–3051. <https://doi.org/10.1093/brain/awaa246>

- Burge, C., y Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, 268(1), 78–94. <https://doi.org/10.1006/jmbi.1997.0951>
- Castelo Rueda, M. P., Gilmozzi, V., Riekschnitz, D. A., Di Segni, M., Silipigni, R., Pramstaller, P. P., Hicks, A. A., Pichler, I., y Zanon, A. (2022). Generation and characterization of induced pluripotent stem cell (iPSC) lines of two asymptomatic individuals carrying a heterozygous exon 7 deletion in *PARKIN* (*PRKN*) and two non- carriers from the same family. *Stem Cell Research*, 60, 102692. <https://doi.org/10.1016/j.scr.2022.102692>
- Elfferich, P., Verleun-Mooijman, M. C., Maat-Kievit, J. A., van de Warrenburg, B. P. C., Abdo, W. F., Eshuis, S. A., Leenders, K. L., Hovestadt, A., Zijlmans, J. C. M., Stroy, J.- P. M., van Swieten, J. C., Boon, A. J. W., van Engelen, K., Verschuuren-Bemelmans, C. C., Lesnik-Oberstein, S. A. J., Tassorelli, C., Lopiano, L., Bonifati, V., Dooijes, D., y van Minkelen, R. (2011). Breakpoint mapping of 13 large *PARKIN* deletions/duplications reveals an exon 4 deletion and an exon 7 duplication as founder mutations. *Neurogenetics*, 12(4), 263–271. <https://doi.org/10.1007/s10048-011-0302-9>
- Fakih, R., Sauv e, V., & Gehring, K. (2023). Feedforward activation of PRKN/parkin. *Autophagy*, 19(2), 729–730. <https://doi.org/10.1080/15548627.2022.2100615>
- Fang, Y.-Q., Mao, F., Zhu, M.-J., y Li, X.-H. (2019). Compound heterozygous mutations in *PARK2* causing early-onset Parkinson disease. *Medicine*, 98(5), e14228. <https://doi.org/10.1097/MD.00000000000014228>
- Faquih, T. O., Aziz, N. A., Gardiner, S. L., Li-Gao, R., de Mutsert, R., Milaneschi, Y., Trompet, S., Jukema, J. W., Rosendaal, F. R., van Hylckama Vlieg, A., van Dijk, K. W., y Mook-Kanamori, D. O. (2023). Normal range CAG repeat size variations in the *HTT* gene are associated with an adverse lipoprotein profile partially mediated by body mass index. *Human*

- Molecular Genetics*, 1(1), 1–12. <https://doi.org/10.1093/hmg/ddad020>
- Feigin, V. L., Norrving, B., y Mensah, G. A. (2017). Global Burden of Stroke. *Circulation Research*, 120(3), 439–448. <https://doi.org/10.1161/CIRCRESAHA.116.308413>
- Gasteiger, E. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13), 3784–3788. <https://doi.org/10.1093/nar/gkg563>
- Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., y Bryant, S. H. (2010). The NCBI BioSystems database. *Nucleic Acids Research*, 38(Database issue), 492–496. <https://doi.org/10.1093/nar/gkp858>
- Hoff, K. J., y Stanke, M. (2013). WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research*, 41(W1), W123–W128. <https://doi.org/10.1093/nar/gkt418>
- Jensen, I., Hendrich, C., Klietz, M., Berding, G., Höglinger, G. U., y Wegner, F. (2022). Case report: Early-onset Parkinson's disease with initial spastic paraparesis and hyperreflexia caused by compound heterozygous *PRKN*-gene exon 2 and 4 deletions. *In Frontiers in neurology*, 13, p. 969232). <https://doi.org/10.3389/fneur.2022.969232>
- Jian, X., Boerwinkle, E., y Liu, X. (2014). *In silico* tools for splicing defect prediction: a survey from the viewpoint of end users. *Genetics in Medicine*, 16(7), 497–503. <https://doi.org/10.1038/gim.2013.176>
- Johansen, K. K., Torp, S. H., Farrer, M. J., Gustavsson, E. K., y Aasly, J. O. (2018). A Case of Parkinson's Disease with No Lewy Body Pathology due to a Homozygous Exon Deletion in *Parkin*. *Case Reports in Neurological Medicine*, 2018, 6838965. <https://doi.org/10.1155/2018/6838965>
- Kariya, S., Hirano, M., Takahashi, N., Furiya, Y., y Ueno, S. (2005). Lack of association between

- polymorphic microsatellites of the VMAT2 gene and Parkinson's disease in Japan. *Journal of the Neurological Sciences*, 232(1–2), 91–94. <https://doi.org/10.1016/j.jns.2005.02.005>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., y Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, 28(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kluge, A., Bunk, J., Schaeffer, E., Drobny, A., Xiang, W., Knacke, H., Bub, S., Lückstädt, W., Arnold, P., Lucius, R., Berg, D., y Zunke, F. (2022). Detection of neuron-derived pathological α -synuclein in blood. *Brain*, 145(9), 3058–3071. <https://doi.org/10.1093/brain/awac115>
- Kretova, M., Selicky, T., Cipakova, I., y Cipak, L. (2023). Regulation of Pre-mRNA Splicing: Indispensable Role of Post-Translational Modifications of Splicing Factors. *Life*, 13(3), 604. <https://doi.org/10.3390/life13030604>
- Kumari, A., Sedehizadeh, S., Brook, J. D., Kozlowski, P., y Wojciechowska, M. (2022). Differential fates of introns in gene expression due to global alternative splicing. *Human Genetics*, 141(1), 31–47. <https://doi.org/10.1007/s00439-021-02409-6>
- Li, D., Aung-Htut, M. T., Ham, K. A., Fletcher, S., y Wilton, S. D. (2020). A Splice Intervention Therapy for Autosomal Recessive Juvenile Parkinson's Disease Arising from *PARKIN* Mutations. *International Journal of Molecular Sciences*, 21(19), 7282. <https://doi.org/10.3390/ijms21197282>
- Liu, Y., Fallon, L., Lashuel, H. A., Liu, Z., y Lansbury, P. T. (2002). The UCH-L1 gene encodes two opposing enzymatic activities that affect α -synuclein degradation and Parkinson's disease susceptibility. *Cell*, 111(2), 209–218. [https://doi.org/10.1016/S0092-8674\(02\)01012-7](https://doi.org/10.1016/S0092-8674(02)01012-7)

- Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., y Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(1), 636–641. <https://doi.org/10.1093/nar/gkz268>
- Martínez-Pérez, F., Becerra, A., Valdés, J., Zinker, S., y Aréchiga, H. (2002). A possible molecular ancestor for mollusk APGWamide, insect Adipokinetic Hormone, and crustacean Red Pigment Concentrating Hormone. *Journal of Molecular Evolution*, 54(6), 703–714. <https://doi.org/10.1007/s00239-001-0036-7>
- Martínez-Pérez, F., Durán-Gutiérrez, D., Delaye, L., Becerra, A., Aguilar, G., y Zinker, S. (2007). Loss of DNA: A plausible molecular level explanation for crustacean neuropeptide gene evolution. *Peptides*, 28(1), 76–82. <https://doi.org/10.1016/j.peptides.2006.09.021>
- Mirpour, S., Turkbey, E. B., Marashdeh, W., El Khouli, R., y Subramaniam, R. M. (2018). Impact of DAT-SPECT on Management of Patients Suspected of Parkinsonism. *Clinical Nuclear Medicine*, 43(10), 710–714. <https://doi.org/10.1097/RLU.0000000000002240>
- Mor-Shaked, H., Paz-Ebstein, E., Basal, A., Ben-Haim, S., Grobe, H., Heymann, S., Israel, Z., Namnah, M., Nitzan, A., Rosenbluh, C., Saada, A., Tzur, T., Yanovsky-Dagan, S., Zaidel-Bar, R., Harel, T., y Arkadir, D. (2021). Levodopa-responsive dystonia caused by biallelic *PRKN* exon inversion invisible to exome sequencing. *Brain Communications*, 3(3), fcab197. <https://doi.org/10.1093/braincomms/fcab197>
- Munk, S. H. N., Voutsinos, V., y Oestergaard, V. H. (2021). Large Intronic Deletion of the Fragile Site Gene *PRKN* Dramatically Lowers Its Fragility Without Impacting Gene Expression. *Frontiers in Genetics*, 12, 695172. <https://doi.org/10.3389/fgene.2021.695172>
- Nalls, M. A., Blauwendraat, C., Vallerga, C. L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan,

- M., Kia, D. A., Noyce, A. J., Xue, A., Bras, J., Young, E., von Coelln, R., Simón- Sánchez, J., Schulte, C., Sharma, M., Krohn, L., Pihlstrøm, L., Siitonen, A., ... Zhang, F. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology*, *18*(12), 1091–1102. [https://doi.org/10.1016/S1474-4422\(19\)30320-5](https://doi.org/10.1016/S1474-4422(19)30320-5)
- Nichols, E., Szeoke, C. E. I., Vollset, S. E., Abbasi, N., Abd-Allah, F., Abdela, J., Aichour, M. T. E., Akinyemi, R. O., Alahdab, F., Asgedom, S. W., Awasthi, A., Barker-Collo, S. L., Baune, B. T., Béjot, Y., Belachew, A. B., Bennett, D. A., Biadgo, B., Bijani, A., Bin Sayeed, M. S., ... Murray, C. J. L. (2019). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, *18*(1), 88–106. [https://doi.org/10.1016/S1474-4422\(18\)30403](https://doi.org/10.1016/S1474-4422(18)30403)
- Pereira, S. L., Grossmann, D., Delcambre, S., Hermann, A., & Grünewald, A. (2023). Novel insights into Parkin-mediated mitochondrial dysfunction and neuroinflammation in Parkinson's disease. *Current opinion in neurobiology*, *80*, 102720. Advance online publication. <https://doi.org/10.1016/j.conb.2023.102720>
- Pickrell, A. M., y Youle, R. J. (2015). The roles of PINK1, Parkin, and mitochondrial fidelity in parkinson's disease. *Neuron*, *85*(2), 257–273. <https://doi.org/10.1016/j.neuron.2014.12.007>
- Pradilla Ardila, G., Vesga Angarita, B. E., y León-Sarmiento, F. E. (2002). Estudio neuroepidemiológico en Aratoca, una área rural del oriente colombiano. *Revista Médica de Chile*, *130*(2), 1–5. <https://doi.org/10.4067/S0034-98872002000200009>
- Price, D. L., Sisodia, S. S., y Borchelt, D. R. (1998). Genetic Neurodegenerative Diseases: The Human Illness and Transgenic Models. *Science*, *282*(5391), 1079–1083. <https://doi.org/10.1126/science.282.5391.1079>

- Sakuwa, M., Adachi, T., Yoshida, K., Adachi, Y., Nakano, T., y Hanajima, R. (2021). An autopsy case of PARK2 due to a homozygous exon 2 deletion of *PARKIN* and associated with α -synucleinopathy. *Neuropathology*, 41(4), 293–300. <https://doi.org/https://doi.org/10.1111/neup.12735>
- Sanyal, J., Anirudhan, A., Banerjee, T. K., Guha, G., Ramakrishnan Veerabathiran, Murugesan, R., Ahmed, S. S. J. S., y Rao, V. R. (2020). PARK2 and PARK7 Gene Polymorphisms as Risk Factors Associated with Serum Element Concentrations and Clinical Symptoms of Parkinson's Disease. *Cellular and Molecular Neurobiology*, 40(3), 357–367. <https://doi.org/10.1007/s10571-019-00734-z>
- Sato, C., Barthélemy, N. R., Mawuenyega, K. G., Patterson, B. W., Gordon, B. A., Jockel-Balsarotti, J., Sullivan, M., Crisp, M. J., Kasten, T., Kirmess, K. M., Kanaan, N. M., Yarasheski, K. E., Baker-Nigh, A., Benzinger, T. L. S., Miller, T. M., Karch, C. M., y Bateman, R. J. (2018). Tau Kinetics in Neurons and the Human Central Nervous System. *Neuron*, 97(6), 1284-1298.e7. <https://doi.org/10.1016/j.neuron.2018.02.015>
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., y Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Research*, 47(D1), D94–D99. <https://doi.org/10.1093/nar/gky989>
- Seike, N., Yokoseki, A., Takeuchi, R., Saito, K., Miyahara, H., Miyashita, A., Ikeda, T., Aida, I., Nakajima, T., Kanazawa, M., Wakabayashi, M., Toyoshima, Y., Takahashi, H., Matsumoto, R., Toda, T., Onodera, O., Ishikawa, A., Ikeuchi, T., y Kakita, A. (2021). Genetic Variations and Neuropathologic Features of Patients with *PRKN* Mutations. *Movement Disorders*, 36(7), 1634–1643. <https://doi.org/10.1002/mds.28521>
- Shiraishi, Y., Okada, A., Chiba, K., Kawachi, A., Omori, I., Mateos, R. N., Iida, N., Yamauchi, H., Kosaki, K., y Yoshimi, A. (2022). Systematic identification of intron retention associated

- variants from massive publicly available transcriptome sequencing data. *Nature Communications*, 13(1), 5357. <https://doi.org/10.1038/s41467-022-32887-9>
- Shuvalova, L. D., Ereemeev, A. V., Bogomazova, A. N., Novosadova, E. V., Zerkalenkova, E. A., Olshanskaya, Y. V., Fedotova, E. Y., Glagoleva, E. S., Illarioshkin, S. N., Lebedeva, O. S., y Lagarkova, M. A. (2020). Generation of induced pluripotent stem cell line RCPCMi004-A derived from patient with Parkinson's disease with deletion of the exon 2 in *PARK2* gene. *Stem Cell Research*, 44, 101733. <https://doi.org/10.1016/j.scr.2020.101733>
- Sigrist, C. J. A., de Castro, E., Cerutti, L., Cucho, B. A., Hulo, N., Bridge, A., Bougueleret, L., y Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, 41(D1), D344–D347. <https://doi.org/10.1093/nar/gks1067>
- Stanke, M., Tzvetkova, A., y Morgenstern, B. (2006). AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biology*, 7(1), S11. <https://doi.org/10.1186/gb-2006-7-s1-s11>
- Tan, E. K., Shen, H., Tan, J. M. M., Lim, K. L., Fook-Chong, S., Hu, W. P., Paterson, M. C., Chandran, V. R., Yew, K., Tan, C., Yuen, Y., Pavanni, R., Wong, M. C., Puvan, K., y Zhao, Y. (2005). Differential expression of splice variant and wild-type *PARKIN* in sporadic Parkinson's disease. *Neurogenetics*, 6(4), 179–184. <https://doi.org/10.1007/s10048-005-0001-5>
- Tan, E. K., Tan, C., Shen, H., Chai, A., Lum, S.-Y., Teoh, M.-L., Yih, Y., Wong, M.-C., y Zhao, Y. (2003). Alpha synuclein promoter and risk of Parkinson's disease: microsatellite and allelic size variability. *Neuroscience Letters*, 336(1), 70–72. [https://doi.org/10.1016/S0304-3940\(02\)01178-3](https://doi.org/10.1016/S0304-3940(02)01178-3)
- Terešák, P., Lapao, A., Subic, N., Boya, P., Elazar, Z., y Simonsen, A. (2022). Regulation of

PRKN-independent mitophagy. *Autophagy*, 18(1), 24–39.

<https://doi.org/10.1080/15548627.2021.1888244>

Vizziello, M., Borellini, L., Franco, G., y Ardolino, G. (2021). Disruption of Mitochondrial Homeostasis: The Role of PINK1 in Parkinson's Disease. *Cells*, 10(11), 3022.

<https://doi.org/10.3390/cells10113022>

Walter, M. A., Rezaie, T., Hufnagel, R. B., y Arno, G. (2020). Ocular genetics in the genomics age. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 184(3),

860–868. <https://doi.org/10.1002/ajmg.c.31844>

Zanon, A., Riekschnitz, D., von Troyer, M., Volpato, C., Picard, A., Cantaloni, C., Di Segni, M.,

Silipigni, R., Pramstaller, P. P., Hicks, A. A., y Pichler, I. (2019). Generation of an induced pluripotent stem cell line (EURACi005-A) from a Parkinson's disease patient carrying a

homozygous exon 3 deletion in the *PRKN* gene. *Stem Cell Research*, 41, 101624.

<https://doi.org/10.1016/j.scr.2019.101624>

Zhang, T., Liu, Q., Gao, W., Sehgal, S. A., y Wu, H. (2022). The multifaceted regulation of mitophagy by endogenous metabolites. *Autophagy*, 18(6), 1216–1239.

<https://doi.org/10.1080/15548627.2021.1975914>

Zhu, W., Huang, X., Yoon, E., Bandres-Ciga, S., Blauwendraat, C., Billingsley, K. J., Cade, J. H.,

Wu, B. P., Williams, V. H., Schindler, A. B., Brooks, J., Gibbs, J. R., Hernandez, D. G.,

Ehrlich, D., Singleton, A. B., y Narendra, D. P. (2022). Heterozygous *PRKN* mutations are common but do not increase the risk of Parkinson's disease. *Brain*, 00(1), 1–15.

<https://doi.org/10.1093/brain/awab456>