

**DETERMINACIÓN DE VARIANTES GENÉTICAS ASOCIADAS A
DISCAPACIDAD COGNITIVA EN EL SUROCCIDENTE COLOMBIANO
MEDIANTE SECUENCIACIÓN DE NUEVA GENERACIÓN**

JOSÉ MARÍA SATIZÁBAL SOTO M.D., M.Sc. Ph.D.

**UNIVERSIDAD DEL VALLE
FACULTAD DE SALUD
ESCUELA DE CIENCIAS BÁSICAS
MAESTRÍA EN CIENCIAS BIOMÉDICAS - 7670
SANTIAGO DE CALI
2017**

**DETERMINACIÓN DE VARIANTES GENÉTICAS ASOCIADAS A
DISCAPACIDAD COGNITIVA EN EL SUROCCIDENTE COLOMBIANO
MEDIANTE SECUENCIACIÓN DE NUEVA GENERACIÓN**

JOSÉ MARÍA SATIZÁBAL SOTO M.D., M.Sc. Ph.D.

**Trabajo de grado como requisito parcial para optar al título de:
MAGISTER EN CIENCIAS BIOMÉDICAS CON ÉNFASIS EN GENÉTICA
MÉDICA**

Director

ADALBERTO SÁNCHEZ GÓMEZ Ph.D.

**UNIVERSIDAD DEL VALLE
FACULTAD DE SALUD
ESCUELA DE CIENCIAS BÁSICAS
MAESTRÍA EN CIENCIAS BIOMÉDICAS - 7670
SANTIAGO DE CALI
2017**

NOTA DE ACEPTACIÓN

Presidente del jurado

Jurado

Jurado

Santiago de Cali, Agosto de 2017

AGRADECIMIENTOS

Agradezco principalmente a mi familia, a mi esposa Lina Johanna Morena, a mis hijos Mauricio Andrés y Christian Camilo. A mi director Adalberto Sánchez Gómez y a mis compañeros del departamento de Ciencias Fisiológicas. A la Universidad del Valle y principalmente a la Fundación para la Prevención de la Discapacidad -FUNDIS- por la financiación parcial de este proyecto.

CONTENIDO

Pág.

RESUMEN.....	9
ABSTRACT.....	11
1. INTRODUCCIÓN.....	13
1.1. DISCAPACIDAD COGNITIVA.....	13
1.2. EL GENOMA HUMANO.....	14
1.3. VARIANTES EXÓMICAS.....	17
1.4. VARIANTES ASOCIADAS A DISCAPACIDAD COGNITIVA.....	18
1.5. SECUENCIACIÓN DE NUEVA GENERACIÓN (NGS): SECUENCIACIÓN DE EXOMA COMPLETO.....	20
2. OBJETIVOS.....	26
2.1. OBJETIVO GENERAL.....	26
2.2 OBJETIVOS ESPECÍFICOS.....	26
3. MATERIALES Y MÉTODOS.....	27
3.1. TOMA DE MUESTRAS.....	27
3.2. SECUENCIACIÓN DE NUEVA GENERACIÓN (NGS).....	27
3.3. ANÁLISIS GENÓMICO Y BIOINFORMÁTICO.....	28
4. RESULTADOS.....	30
4.1. VARIANTES EXÓMICAS ASOCIADAS A DISCAPACIDAD COGNITIVA....	30
4.2. ANÁLISIS GENÓMICO Y BIOINFORMÁTICO.....	42
5. DISCUSIÓN.....	46
6. CONCLUSIONES Y PERSPECTIVAS.....	51
BIBLIOGRAFÍA.....	52

LISTA DE TABLAS

Pág.

Tabla 1. Información acerca de la cobertura y lectura total de bibliotecas reportada por Illumina para cada muestra secuenciada.....	28
Tabla 2. Descripción de genes cuyas mutaciones fueron detectadas en la secuenciación de exoma completo.....	30
Tabla 3. Número de transiciones y transversiones totales encontradas en la secuenciación de exoma completo.....	35
Tabla 4. Variantes exómicas encontradas por gen, en la secuenciación de exoma completo.....	36
Tabla 5. Parámetros topológicos de la red de genes asociados a discapacidad cognitiva, extraídos del programa Cytoscape 3.2 usando la herramienta <i>Network Analyzer</i>	44
Tabla 6. Categorías GO-Procesos biológicos destacados de la red de genes asociados a discapacidad cognitiva, extraídas del programa Cytoscape 3.1 y Quick GO.....	45

LISTA DE FIGURAS

Pág.

- Figura 1. Representación del flujo de la información genética. En primer lugar el ADN es transcrito, con sus intrones y exones, a pre-ARN; posteriormente en un proceso de maduración de dicho pre-ARN, se eliminan los intrones y se empalman los exones dando lugar al ARN mensajero.....15
- Figura 2. Flujograma para el descubrimiento y genotipo de las variaciones a partir de los resultados de la secuenciación de nueva generación.....22
- Figura 3. Estrategias (a-f) para identificar genes de enfermedades a partir de secuenciación de exomas. Los pedigrís indican el modelo de herencia relacionado a la estrategia; Los símbolos llenos representan a los individuos afectados, los símbolos vacíos representan individuos presumiblemente sanos y los portadores están representados con un punto. El rectángulo rebordea a individuos que están secuenciados. Los círculos debajo de cada pedigrí simbolizan conjuntos de variantes genéticas identificadas en los exomas. Los círculos sólidos representan variantes de los individuos afectados, mientras que los círculos discontinuos representan variantes no afectadas,.....24
- Figura 4. Cantidad de mutaciones clasificadas por efectos causados encontradas en la secuenciación de exoma completo.....35
- Figura 5. Redes de interacción proteína-proteína construidas con base en las proteínas asociadas a discapacidad cognitiva según la secuenciación de exoma completo con el programa Cytoscape 3.2. A) Red completa B) Sub-red mostrando los dos nodos con mayor cantidad de interacciones.....44

ABREVIATURAS

AD. Alzheimer disease
ADN. Ácido desoxirribonucleico
ALD. Adrenoleucodistrofia
AMN. Adrenomieloneuropatía
ARN. Ácido ribonucleico
ARNm. Ácido ribonucleico mensajero
BRCA1. Breast cancer 1
DL. Desequilibrio de ligamiento
DYRK1A. Dual specificity tyrosine-phosphorylation-regulated kinase 1
FUNDIS. Fundación para la discapacidad
GEO. Gene expression omnibus
GO. Gene Ontology
ICBF. Instituto Colombiano de Bienestar Familiar
NGS. Next generation sequencing
nsID. Non syndromic Intellectual disability
PDC. Complejo multienzimático de la piruvato deshidrogenasa
RPGRIP1L. RPGRIP1 like
SD. Síndrome de Down
sID. Síndromic Intellectual disability
SNP. Polimorfismo de nucleótido único
SNV. Single nucleotide variant
TBK1. TANK binding kinase 1
TDAH. Trastorno por Déficit de Atención e Hiperactividad
TMEM216. Transmembrane protein 216
UTRs. Untranslated region
VNC. Variaciones del número de copias

RESUMEN

La discapacidad cognitiva se caracteriza por una limitación significativa en el funcionamiento cognitivo y adaptativo, así como una edad de aparición anterior a los 18 años de edad. Esta es una condición que afecta a una gran cantidad de personas no solo en Colombia sino también en el mundo, por lo que el estudio de sus causas a nivel genómico utilizando las nuevas herramientas tecnológicas que se han desarrollado es más que apropiado. En este trabajo se utilizó precisamente uno de estos métodos, la secuenciación de nueva generación, específicamente la secuenciación de exomas completos de pacientes con discapacidad cognitiva con el fin de descubrir qué variantes exómicas presentaban y poder crear una base de datos con información de la estructura genómica de personas del sur-occidente colombiano. Para esto se tomaron muestras de sangre de 15 pacientes de la Fundación para la Discapacidad que presentaban discapacidad cognitiva en cierto grado con el correspondiente consentimiento informado y se mandaron a secuenciar los exomas de cada muestra. Posteriormente se organizaron los resultados en una tabla de Excel, y se realizó un análisis bioinformático utilizando el programa Cytoscape 3.2 para construir una red de interacciones, y las bases de datos BioGRID para obtener dichas interacciones y QUICK GO para obtener las categorías GO específicamente procesos biológicos más destacados de la red. Como resultado de la secuenciación exómica se hallaron en total 184 variantes presentes en a 108 genes, 50% de dichas variantes fueron de tipo *missense* y más del 70% fueron transiciones. Dos de las variantes que se encontraron en uno de los pacientes, en los genes ABCD1 y DLAT, no han sido previamente reportadas en la literatura. A pesar de que según las predicciones de Polyphen y Sift dichas variantes no causan daño en la proteína resultante, existe evidencia de que mutaciones en estos dos genes causan enfermedades que conllevan discapacidad cognitiva en algún grado, por lo que se sugiere en estudios

posteriores estudiar más a fondo las consecuencias de las dos variantes nuevas aquí reportadas. De acuerdo al análisis bioinformático se encontró que de los genes que resultaron afectados con alguna variante, los que más presenta interacciones físicas proteína-proteína es BRCA1 y TMEM216. Además entre los procesos biológicos más destacados de la red se encontraron la sinapsis, la regulación de la muerte neuronal y la migración celular. Este trabajo además de aportar información nueva de variantes exómicas en pacientes con discapacidad cognitiva del sur-occidente colombiano, abre puertas a la creación de una base de datos nacional que recopile las variantes de diferentes regiones del país haciendo uso de la secuenciación de nueva generación.

Palabras clave: Discapacidad cognitiva, Secuenciación de nueva generación, Secuenciación de exomas completos, Variantes exómicas

ABSTRACT

Cognitive impairment is characterized by a significant limitation in cognitive and adaptive functioning, as well as an age of onset prior to 18 years of age. This is a condition that affects a large number of people not only in Colombia but worldwide, so the study of its causes at the genomic level using the new technological tools that have been developed is more than appropriate. In this work we used precisely one of these methods, the Next Generation Sequencing, specifically the sequencing of complete exomes of patients with cognitive disabilities in order to discover what exomic variants they present and to be able to create a database with information of the genomic structure of people from the south-west of Colombia. For this purpose, blood samples were taken from 16 patients from the Foundation for Disability who had cognitive impairment to some degree with the corresponding informed consent, the samples were then sent to be sequenced. Subsequently, the results were organized in an Excel table, and a bioinformatic analysis was performed using the Cytoscape 3.2 program to build a network of interactions, and the BioGRID databases to obtain such interactions and QUICK GO to obtain GO categories specifically the biological processes more prominent from the network. As a result of the exomic sequencing, a total of 184 variants were present in 108 genes, 50% of these variants were missense and more than 70% were transitions. Two of the variants found in one of the patients, in the ABCD1 and DLAT genes, have not been previously reported in the literature. Although according to the predictions of Polyphen and Sift such variants do not cause damage in the resulting protein, there is evidence that mutations in these two genes cause diseases that lead to cognitive impairment to some degree, so it is suggested in later studies to study more in depth the consequences of the two new variants reported here. According to the bioinformatic analysis it was found that of the genes that were affected with some variant, those that presented more

physical interactions of the protein-protein type were BRCA1 and TMEM216. In addition, the synapses, regulation of neuronal death and cell migration were found among the most important biological processes in the network. This work, in addition to providing new information on exomic variants in patients with cognitive disabilities in the south-west of Colombia, opens the door to the creation of a national database that compiles the variants of different regions of the country making use of the Next Generation Sequencing.

Keywords: Cognitive disability, Next generation sequencing, Whole exome sequencing, Exomic variants

1. INTRODUCCIÓN

1.1. DISCAPACIDAD COGNITIVA

La discapacidad cognitiva o discapacidad intelectual, por mucho tiempo fue mal llamada Retraso mental (1), y se refería a una persona cuyo funcionamiento tanto a nivel cognitivo como adaptativo está por debajo de cierto umbral. Sin embargo con el paso del tiempo se evidenció la necesidad de cambiar la terminología de Retraso mental, ya que conllevaba un carácter peyorativo que reducía la comprensión de las personas con limitaciones intelectuales a una categoría diagnóstica nacida desde perspectivas psicopatológicas. Hoy en día el término correcto es discapacidad cognitiva o discapacidad intelectual y debe concebirse desde un enfoque que le dé el primer lugar a la persona como a cualquier individuo de la sociedad. Según la cartilla del ICBF (Instituto Colombiano de Bienestar Familiar) un niño con discapacidad cognitiva presenta diferencias en su ritmo de aprendizaje y desarrollo. Además se considera que existen tres elementos clave en la definición de discapacidad cognitiva: capacidades, entorno y funcionamiento.

Entre las causas de la discapacidad cognitiva se encuentran tanto genéticas como ambientales o multifactoriales (2). En cuanto a las genéticas se cree que son las que se presentan en la mayoría de casos de pacientes con discapacidad cognitiva, e incluyen alteraciones en el número de cromosomas o también cuando existen mutaciones puntuales en uno o más genes. En total para el año 2014 existían más de 500 genes propuestos como candidatos a causar discapacidad cognitiva con una alta penetrancia cuando mutan, otorgando una heterogeneidad fenotípica tanto en severidad como en síntomas asociados. Las causas ambientales hacen por su parte, referencia a factores del entorno o contexto que afectan el desarrollo

del niño, llevando a cambios epigenéticos que también se han visto involucrados en el desarrollo de discapacidad cognitiva.

La discapacidad intelectual según Redin y colaboradores (3) es un desorden del neurodesarrollo reportado en aproximadamente del 1.5 al 2% de niños y adolescentes, definiéndose como una limitación significativa tanto en el funcionamiento intelectual como el comportamiento adaptativo con síntomas visibles antes de los 18 años. Usualmente las personas con discapacidad cognitiva se clasifican en discapacidad severa (<35), moderada (35-49) o leve (50-70) dependiendo de los valores de coeficiente intelectual. Además de esto, es importante mencionar que puede o no estar asociada a un síndrome y se clasifican por lo tanto en discapacidad cognitiva sindrómica (SID por sus siglas en inglés: Syndromic Intellectual Disability) y no sindrómica (nsID por sus siglas en inglés: Non-syndromic Intellectual Disability), las cuales comparten algunas características neurológicas como Trastornos del Espectro Autista, epilepsia, Trastorno por Déficit de Atención e Hiperactividad (TDAH) y anomalías comportamentales (4).

1.2. EL GENOMA HUMANO

El genoma humano es la colección completa de ácido desoxirribonucleico (ADN). Las moléculas de ADN están a su vez conformadas por 3000 millones unidades de las cuatro bases nucleotídicas llamadas Adenina (A), Timina (T), Guanina (G) y Citosina (C). El ADN además presenta diferentes niveles de compactación, siendo los cromosomas, el mayor nivel; precisamente el genoma humano está distribuido en 23 pares de cromosomas dentro del núcleo de las células del cuerpo, 22 de ellos autosomas y el par sexual X y Y. (5). Dicho genoma presenta aproximadamente unos 30 mil genes con información para formar proteínas. A partir de la secuenciación del genoma humano en el año 2003, se logró evidenciar su estructura compleja, encontrando que el ADN no solo presentaba zonas con información para codificar proteínas, sino también presentaba microsátélites,

zonas de secuencias repetidas y reguladoras entre otras. Es de gran importancia comprender la estructura de un gen y cómo es el flujo de la información genética, como se muestra en la figura 1. Los genes están formados por unidades llamadas exones e intrones, sin embargo solo los exones contienen la información necesaria para sintetizar proteínas, llegando a ser aproximadamente el 1.5% del total del genoma humano. En un proceso conocido como transcripción, una ARN-polimerasa forma una cadena de Ácido ribonucleico (ARN) a partir de un molde de ADN, dicha cadena en una primera instancia tiene tanto la información de intrones como los exones, sin embargo tras el proceso de maduración de ARN, específicamente del splicing, se eliminan los intrones y se empalman los exones. Este ARN conocido como ARN mensajero, es el que finalmente sirve como guía para la síntesis de proteínas.

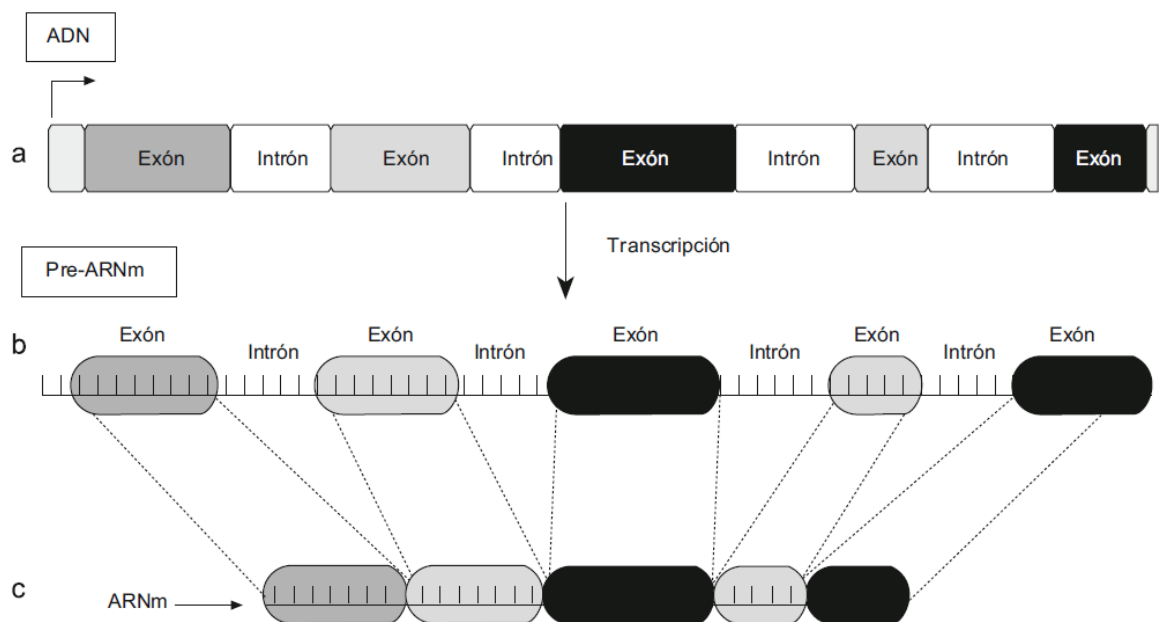


Figura 1. Representación del flujo de la información genética. En primer lugar el ADN es transcrito, con sus intrones y exones, a pre-ARN; posteriormente en un proceso de maduración de dicho pre-ARN, se eliminan los intrones y se empalman los exones dando lugar al ARN mensajero.

Fuente: (6) Martínez-Frías, 2010

De manera interesante cabe destacar que los exones pueden ser clasificados en: exones UTR (Por sus siglas en inglés: Untranslated regions) 5', exones codificantes y exones UTR 3'. Se habla de exoma cuando se refiere a todos los exones de un organismo.

En cuanto a la traducción del ARNm a proteína, es un proceso que se da en el citoplasma de las células e involucra ribosomas, ARNs de transferencia y diferentes factores protéicos que regulan el proceso. En el ARNm está formado por ribonucleótidos, y se habla de codón a una tripleta de dichos ribonucleótidos; cada uno de estos codones puede codificar para cierto aminoácido, de hecho varios codones pueden codificar para un mismo aminoácido. Las diferentes combinaciones de nucleótidos que forman tripletas están ya organizadas en lo que se conoce como código genético. Cualquier cambio que se produzca en la combinación de los nucleótidos puede modificar el código genético del ADN y alterar la expresión de una proteína. Por ejemplo, si un triplete TCA que codifica el aminoácido serina, se produce un cambio de la T por una C, el aminoácido que ahora codifica el triplete CCA es una prolina, que es muy diferente de la serina, por lo que se altera la proteína. Sin embargo, no todas las mutaciones son patológicas, ni tienen el mismo efecto en la proteína (6). Entre los efectos que puede tener una mutación se puede encontrar el cambio de sentido o *missense*, mutación sin sentido y mutaciones silenciosas entre otras. Las mutaciones *missense* son los cambios puntuales de un nucleótido, que resultan en el cambio de un aminoácido por otro distinto en la proteína resultante; las mutaciones sin sentido producen una señal de parada de la traducción, obteniendo como resultado una proteína truncada y no funcional. Las mutaciones silenciosas por su parte no producen cambios en la proteína resultante, ya que a pesar de que el cambio de nucleótido genera un codón diferente al original, este codifica para el mismo aminoácido, por lo que no se ve alterada la función ni la estructura de la proteína (7)

Dichas mutaciones en el genoma humano pueden producirse ya sea en las células somáticas o en las células germinales. Precisamente si una mutación ocurre en la línea germinal es mucho más trascendental debido a que esta información va a pasar con ese mismo error a las siguientes generaciones. Se puede hablar de mutaciones naturales e inducidas, las primeras son espontáneas y ocurren en grandes cantidades, siendo la mayoría correctamente reparadas por la maquinaria de la misma célula, mientras que las mutaciones inducidas son daños en el ADN provocados artificialmente con radiación UV, sustancias químicas u otros agentes mutagénicos (8).

Se pueden distinguir tres grupos grandes de mutaciones de acuerdo a la extensión de material genético afectado: Genéticas o puntuales, cromosómicas estructurales y cromosómicas numéricas o genómicas (8). En cuanto a las mutaciones puntuales son aquellas que producen alteraciones en la secuencia de nucleótidos de un gen y existen varios tipos como las sustituciones, inserciones o deleciones entre otros. Las sustituciones pueden ser transiciones si el cambio de nucleótido es de una purina por otra purina o una pirimidina por otra pirimidina, o pueden ser transversiones si el cambio de nucleótido es de una purina por una pirimidina o viceversa. Las mutaciones cromosómicas estructurales por su parte pueden agruparse en dos grupos, las que suponen pérdida o duplicación de segmentos completos de un cromosoma, y las que suponen variaciones en la distribución de segmentos de los cromosomas. Por último, las mutaciones cromosómicas numéricas son las que alteran el número de los cromosomas propios de la especie y pueden dividirse en euploidías y aneuploidías.

1.3. VARIANTES EXÓMICAS

Durante el proceso de replicación del ADN y por otras múltiples causas, el ADN puede presentar variantes, pequeños errores de un solo nucleótido, que rápidamente tratan de ser reparados por la maquinaria interna de cada célula. Sin

embargo, algunas de estas variantes no son corregidas y pueden aumentar la diversidad genética. Cada ser humano es único, aún con un 99.9% de similitud, el 0.1% restante es responsable de muchas de las diferencias que definen a cada individuo, además esta variación genómica puede o no tener impacto en la salud. En el genoma humano pueden encontrarse variantes en toda su extensión, no solo en las regiones exómicas, sino también en regiones del ADN que no codifican proteínas. El exoma del genoma humano está formado por unos 180.000 exones que constituyen aproximadamente el 1,5% del genoma total, pero a pesar de que sea un porcentaje tan pequeño, las mutaciones que tienen lugar en esta zona, tienen una probabilidad mucho mayor de expresarse que las de cualquier otra zona. Se calcula que el exoma contiene el 85% de las mutaciones que son capaces de generar una enfermedad o un trastorno. De hecho la secuenciación de exomas ha sido una estrategia eficaz para determinar la base genética de decenas de enfermedades hereditarias y para identificar mutaciones *de novo* implicadas en trastornos del desarrollo como la esquizofrenia, la discapacidad intelectual, la epilepsia o el autismo (9 - 10).

La secuenciación del exoma permite identificar mutaciones poco comunes que causan enfermedades a escala de todo el genoma. Esto contrasta de forma importante con los estudios tradicionales de asociación del genoma completo, que son más eficientes para identificar alteraciones comunes que son importantes en una enfermedad. Cuando mutaciones genéticas raras pueden causar la misma enfermedad individualmente, como es el caso en muchas condiciones complejas, los cambios en las secuencias relevantes pueden ser más difíciles de encontrar mediante estudios de asociación del genoma completo.

1.4. VARIANTES ASOCIADAS A DISCAPACIDAD COGNITIVA

Para el diagnóstico de la discapacidad cognitiva se utilizan evaluaciones estandarizadas de inteligencia y comportamiento adaptativo (habilidad para

realizar actividades de la vida diaria) que terminan en la clasificación del paciente en una de las categorías que se tienen. Sin embargo actualmente con el desarrollo de la tecnología y el conocimiento, se pueden utilizar técnicas diferentes y mucho más precisas para diagnosticar la discapacidad cognitiva como lo es la secuenciación completa del exoma. Esta técnica ha probado ser una de las herramientas más poderosas para identificar la etiología de enfermedades aparentemente monogénicas, en particular desórdenes del neurodesarrollo y desvelar nuevas enfermedades o genes candidatos a biomarcadores. Recientemente se ha reportado una alta tasa de diagnósticos (68%) en investigaciones de 41 pacientes con discapacidad cognitiva y un fenotipo metabólico ampliamente definido según Evers y colaboradores (11)

En los últimos años y conforme a la tecnología de secuenciación de nueva generación (NGS) entró en auge, se han realizado una gran cantidad de estudios que involucran la NGS como medio para hallar variables asociadas a diversas enfermedades (12) (13). En el caso puntual de la discapacidad cognitiva no ha sido diferente, y de hecho se ha logrado reunir información clave para detectar genes biomarcadores. En el estudio de Rump y colaboradores (14) se investigó la utilidad de la NGS como aproximación diagnóstica para establecer un diagnóstico molecular en un grupo altamente heterogéneo de pacientes con discapacidad intelectual variada y microcefalia, demostrando que es una poderosa herramienta para el fin propuesto. Otro estudio realizado por Nesbitt y colaboradores (15) reportó la identificación de novo de una variante con pérdida de función en el gen *SOX5* identificada mediante secuenciación de exoma completo; este gen ha sido previamente asociado con discapacidad intelectual, de lenguaje y motora. Por otro lado el estudio de Luco y colaboradores (16) identificó mutaciones en el gen *Dual specificity tyrosine-phosphorylation-regulated kinase 1* (*DYRK1A*) encontradas en individuos con discapacidad intelectual sindrómica mediante el uso de NGS, señalando este gen como biomarcador y de importancia para diagnosticar pacientes. Con la secuenciación de exomas completos y el panorama que ha

mostrado es claro que se seguirá teniendo un profundo efecto sobre la genética y el diagnóstico de la discapacidad intelectual (17). Además de esto, con la incorporación de la secuenciación de exomas en el laboratorio de diagnóstico, el ritmo y el rendimiento tanto del diagnóstico como del descubrimiento de genes deberían seguir aumentando (17). Así mismo, las investigaciones en torno a tratamientos eficientes y personalizados para los pacientes deberían aumentar por igual.

1.5. SECUENCIACIÓN DE NUEVA GENERACIÓN (NGS): SECUENCIACIÓN DE EXOMA COMPLETO

El exoma de un individuo es toda la región del ADN que codifica para proteínas y representa aproximadamente el 1.5% del genoma humano total, sin embargo contiene casi el 85% de las variantes asociadas con enfermedades. Es por esto que resulta mucho más efectivo en relación a su costo, realizar una secuenciación de exoma completo, en vez de una secuenciación de genoma completo, produciendo una cantidad de datos mucho más pequeña, manejable y más fácil de analizar. La secuenciación de exomas detecta variantes encontradas en exones, con la capacidad de expandir su “target” hacia regiones no traducidas (UTR) y microARNs para una visión más integral de la regulación de genes.

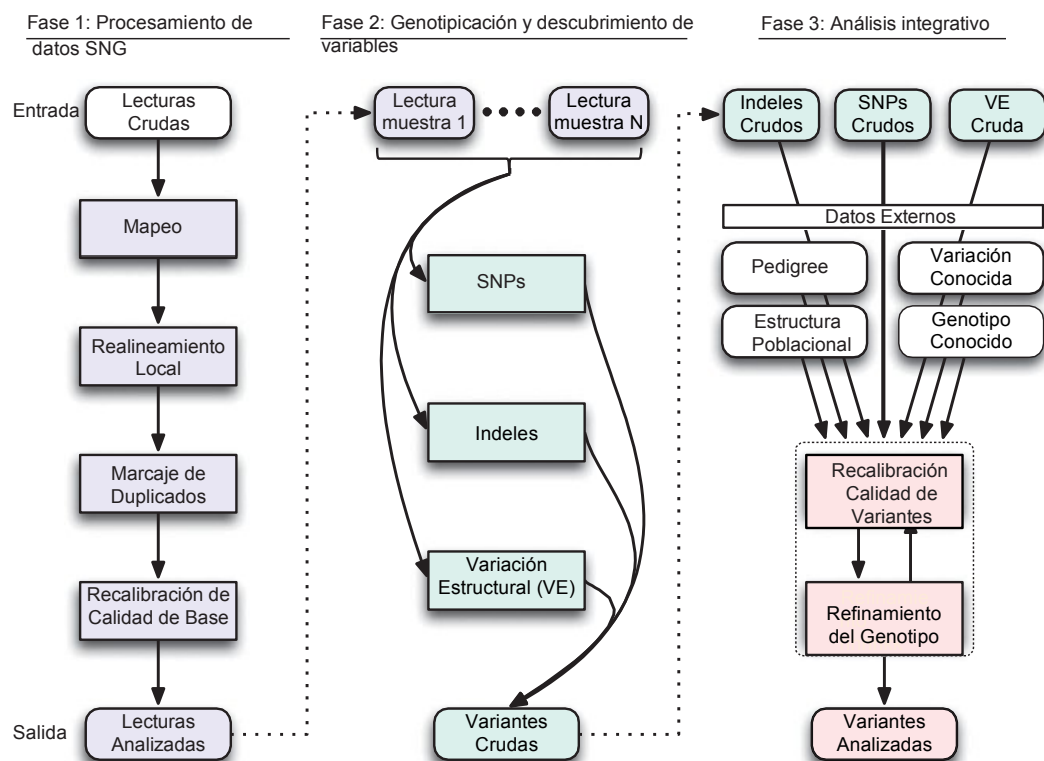
Los rápidos avances en métodos de secuenciación de alto rendimiento facilitan la búsqueda de variantes genéticas codificantes, que causan enfermedades imposibles de detectar en estudios de asociación a nivel genético (18). Esta metodología consiste básicamente de tres pasos: fragmentación del ADN de interés, amplificación y secuenciación de fragmentos; y el ensamblaje de los fragmentos (19).

Estudios recientes utilizando NGS han identificado con éxito los genes mutados relacionados a trastornos mendelianos a partir de la secuenciación del exoma

completo, esta técnica permite conocer las secuencias de todos los exones que se expresan en un genoma (20). Una de las grandes ventajas, comparado con la secuenciación del genoma completo, es su bajo costo (21). Antes de iniciar el proceso se debe hacer la selección del ADN que codifica para proteínas o conocerlo, en el humano hay alrededor de 180.000 exones, aproximadamente el 1% del genoma total o 30 millones de pares de bases (22).

En la primera fase del proceso (Fig. 2), los datos crudos provenientes de algún método de NGS, como Sanger, Roche 454, Illumina y Life SOLiD, se transforman a una sola representación genérica, mapeada a su correcto origen genómico y consistentemente alineadas. Los algoritmos de mapeo sitúan lecturas con un alineamiento inicial sobre el genoma de referencia, luego los duplicados moleculares son removidos, los alineamientos iniciales se refinan por realineamiento local y luego se determina empíricamente un adecuado modelo de error por base. La segunda fase se caracteriza por identificar todos los sitios con evidencia estadística de un alelo alterno presente entre las muestras incluyendo polimorfismos de nucleótido único (SNP), indeles pequeños y variaciones del número de copias (VNC). La fase final reúne datos de variables técnicas, sitios conocidos de variación, genotipos por individuos, desequilibrio de ligamiento (DL), estructura de la familia y de la población con las variantes crudas de la fase dos para separar sitios polimórficos verdaderos de errores del sistema (23).

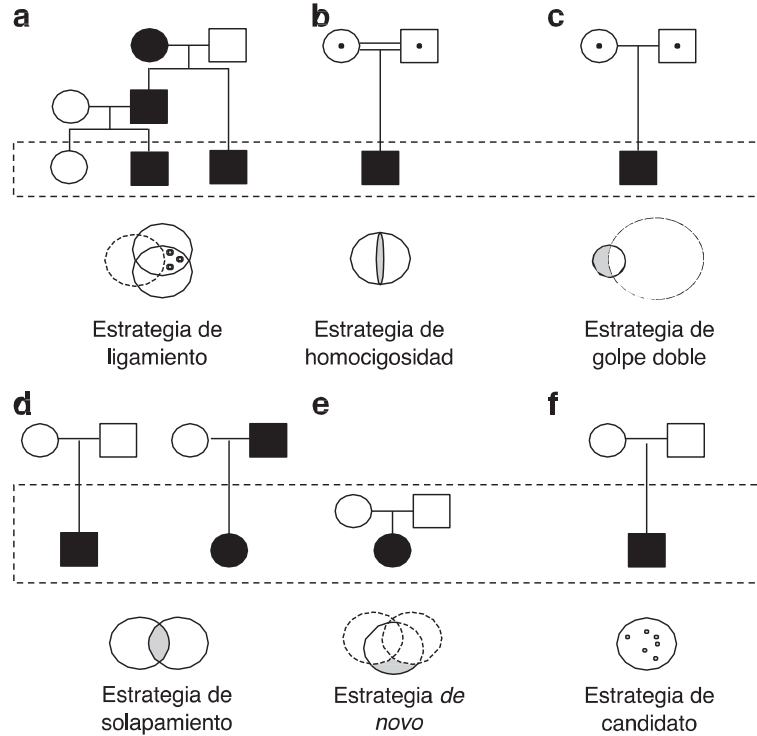
Figura 2. Flujograma para el descubrimiento y genotipo de las variaciones a partir de los resultados de la secuenciación de nueva generación



Fuente: Modificado de DePristo et al 2010 (23).

Dado que el número de variantes por exoma oscila entre 20.000 y 50.000 (24), para reducir el número de falsos positivos, las variantes son filtradas basadas en el criterio de calidad, como el número total de lecturas independientes y el porcentaje de lecturas que presentan la variación, posteriormente las variaciones externas a la región codificante pueden ser filtradas, al igual que las variantes sinónimas, esto reduce el número de variantes potenciales que causan enfermedades a 5000. La reducción más sustancial se da por la exclusión de variantes conocidas (publicadas previamente y almacenada en bases de datos). Este paso reduce del 90 a 95% el número de variantes potenciales, quedando entre 150 a 500 variantes privadas no sinónimas como potenciales patogénicas (25). Otras estrategias deben ser usadas para encontrar la variante. Una de ellas es la de *ligamiento* (Fig. 3.a), implementada en una familia con un desorden monogénico heredado, donde múltiples miembros afectados se secuencian para

identificar la variación compartida, también se secuencian un familiar no afectado para eliminar variación benéfica (20, 26). La estrategia de *homocigosis* (Fig. 3.b) en un desorden heredado recesivamente y presunta consanguinidad, con los supuestos que la enfermedad es causada por variantes homocigóticas heredadas de ambos padres y que esta variante se ubica en una gran extensión de una región homocigótica no se necesita de miembros familiares adicionales, pues estos pueden ser mapeados por microarreglos SNPs (27). La estrategia de *golpe doble* (Fig. 3.c) se usa cuando un solo paciente está disponible sin miembros familiares adicionales y se sospecha que el desorden fue heredado recesivamente (pero sin ninguna indicación de consanguinidad), se secuencian al exoma del paciente seleccionando los genes homocigóticos, así como las variantes heterocigóticas (28, 29). Por otro lado, la estrategia de *solapamiento* (Fig. 3.d) se usa cuando hay ausencia de heterogeneidad genética y se buscan mutaciones en un solo gen en múltiples pacientes no relacionados (25, 30); mientras que la estrategia *de novo* (Fig. 3.e) se utiliza en enfermedades altamente heterogénicas producto de mutaciones durante la meiosis, por lo que se requiere la secuenciación de los padres del paciente para eliminar todas las variaciones heredadas (31). En último lugar está la estrategia del *candidato* (Fig. 3.f) en la que a un solo individuo afectado predominantemente, sin disponibilidad de familiares, se le predice el impacto de la variación sobre la función y estructura de una proteína importante en mutaciones de parada, con desplazamiento del marco de lectura y de sitios de empalme (23). Mientras que en variantes con sustitución de nucleótidos además del impacto en la proteína también la conservación evolutiva del nucleótido variante pues las mutaciones patogénicas tienden a conservarse más que las benéficas (31).



N

Figura 3. Estrategias (a-f) para identificar genes de enfermedades a partir de secuenciación de exomas. Los pedigrís indican el modelo de herencia relacionado a la estrategia; Los símbolos llenos representan a los individuos afectados, los símbolos vacíos representan individuos presumiblemente sanos y los portadores están representados con un punto. El rectángulo rebordea a individuos que están secuenciados. Los círculos debajo de cada pedigrí simbolizan conjuntos de variantes genéticas identificadas en los exomas. Los círculos sólidos representan variantes de los individuos afectados, mientras que los círculos discontinuos representan variantes no afectadas,

Fuente: Modificado de Gllissen et al 2012 (24).

Con la secuenciación de exoma completo para hallar posibles variantes en las zonas codificantes del ADN, es necesario tener en cuenta que se debe asumir que precisamente dichas variantes relacionadas con una condición, están en los exones y no hacen parte de la zona reguladora o intrónica del ADN.

Teniendo en cuenta que en Colombia se siguen usando como referencia de variantes, las publicadas en bases de datos de otros países, la idea de comenzar a construir una base de datos propia, con recolección de muestras de pacientes colombianos permitiría tener la información de la estructura genómica de la población, y se facilitaría una comparación mucho más efectiva y acertada en los estudios que se realicen. Precisamente cuando se compara una muestra genómica colombiana con una base de datos extranjera se puede perder información clave, ya que las variantes reportadas en esas bases de datos son propias de otras poblaciones o si tienen muestras de América Latina pero no específicamente de Colombia o del sur-occidente colombiano. El desarrollo del presente trabajo es de gran importancia para crear una base de datos autóctona del sur-occidente colombiano, abriendo puertas hacia su ampliación al resto del país.

2. OBJETIVOS

2.1. OBJETIVO GENERAL

Identificar variantes genéticas asociadas con discapacidad cognitiva en pacientes de Cali, Valle del Cauca mediante secuenciación de exoma completo

2.2 OBJETIVOS ESPECÍFICOS

- Determinar las variantes genéticas asociadas a la discapacidad cognitiva
- Construcción de una base de datos con la información de variantes genéticas, autóctona de Colombia
- Relacionar las variantes encontradas en cada paciente, con la discapacidad cognitiva

3. MATERIALES Y MÉTODOS

3.1. TOMA DE MUESTRAS

Cada uno de los 15 pacientes pertenecientes a la población de Cali, Valle de Cauca, que hicieron parte de este trabajo, fueron contactados desde la Fundación para la Discapacidad FUNDIS. Todos los pacientes presentaban un grado de discapacidad cognitiva asociado a diferentes causas. Debido precisamente a la dificultad de lograr un gran número de pacientes dispuestos a dar su consentimiento para toma de sangre, se reportan aquí los casos que dieron efectivamente su aprobación (Anexo 1). Para la obtención del DNA, cada paciente fue sometido a punción digital y se recogieron gotas de sangre en papel de filtro hasta su desecación. Los discos de papel filtro fueron resuspendidos en buffer fosfato y se utilizó el estuche DNeasy de la compañía QIAGEN. Cada extracción fue cuantificada y verificada en calidad para someterse al proceso de secuenciación.

3.2. SECUENCIACIÓN DE NUEVA GENERACIÓN (NGS)

Alícuotas de DNA para cada una de las 15 muestras fueron sometidas a alistamiento con el estuche TruSeq Exome Library preptm y posteriormente las librerías obtenidas se normalizaron para la secuenciación utilizando el estuche TruSeq Rapid Exome™. Los estuches anteriormente mencionados son suministrados por la compañía Illumina de San Diego, California, USA. Los fragmentos normalizados con los adaptadores para secuenciación fueron cargados en un equipo HiSeq2500. La correspondiente secuenciación de exoma completo de las muestras se realizó durante los años 2015, 2016 y 2017. En la tabla 1 se muestran especificaciones del proceso de secuenciación.

Tabla 1. Información acerca de la cobertura y lectura total de bibliotecas reportada por Illumina para cada muestra secuenciada

Plataforma	Paciente	Año	Cobertura	Lectura total de bibliotecas Nextera™ Illumina
Illumina	1	2016	35X	34259140
Illumina	2	2015	35.7X	32926882
Illumina	3	2015	33.4X	30550106
Illumina	4	2016	32.9X	32899412
Illumina	5	2016	17.8X	19169284
Illumina	6	2016	N/A	N/A
Illumina	7	2016	54.3X	60810220
Illumina	8	2016	52.8X	59325332
Illumina	9	2016	51.7X	53616392
Illumina	10	2016	32.9X	27729348
Illumina	11	2016	38X	35573552
Illumina	12	2016	35.2X	38293920
Illumina	13	2016	32.1X	36870506
Illumina	14	2017	41X	36502044
Illumina	15	2017	54X	51279266

3.3. ANÁLISIS GENÓMICO Y BIOINFORMÁTICO

Cada uno de los 108 genes que presentó variantes según la secuenciación de exoma completo fue descrito de acuerdo a su nombre completo, locus, número de exones y proceso biológico en el que se ha visto involucrado. Estos datos fueron extraídos del GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) y organizados en una tabla de Excel®. Además de esto, con dichos genes se realizó un análisis a través del software Cytoscape 3.2 donde se construyó una red con dichos genes y se extrajeron los procesos biológicos asociados destacados. Para la construcción de la red se buscaron en primer lugar las interacciones proteína-proteína de cada gen en la base de datos BioGRID (<https://thebiogrid.org/>); con esta información también se realizó un diagrama de barras utilizando Excel®. A partir de la red construida se extrajeron las categorías GO (Gene Ontology), utilizando el plugin BiNGO de Cytoscape 3.2 y la base de datos Quick GO del Instituto Europeo de Bioinformática EBI (<https://www.ebi.ac.uk/QuickGO/>). Esta información se organizó

en una tabla en orden decreciente, comenzando por el proceso biológico más destacado de la red. Por su parte, los parámetros topológicos de la red fueron extraídos con la herramienta *Network Analyzer* de Cytoscape 3.2.

4. RESULTADOS

4.1. VARIANTES EXÓMICAS ASOCIADAS A DISCAPACIDAD COGNITIVA

En la secuenciación de exoma completo realizada en este trabajo a 15 pacientes que presentaban discapacidad cognitiva se encontró un total de 184 variantes distribuidas en 108 genes. La descripción de cada uno de estos genes, incluyendo el locus, cromosoma donde se ubica y proceso biológico en el que interviene se encuentra en la tabla 2.

Tabla 2. Descripción de genes cuyas mutaciones fueron detectadas en la secuenciación de exoma completo

Gen	Nombre completo	Locus	Ex.	Proceso Biológico
HPSE2	Heparanase 2 (inactive)	10q24.2	18	Glycosaminoglycan catabolic process
ATP7B	ATPase copper transporting beta	13q14.3	30	Copper ion export
POF1B	Premature ovarian failure, 1B	Xq21.1	17	Actin cytoskeleton organization
EVC	EvC ciliary complex subunit 1	4p16.2	31	Smoothed signaling pathway
DSPP	Dentin sialophosphoprotein	4q22.1	5	Ossification
CYP21A2	Cytochrome P450 family 21 subfamily A member 2	6p21.33	10	Steroid biosynthetic process
SCNN1A	Sodium channel epithelial 1 alpha subunit	12p13.31	14	Sodium ion homeostasis
ABCC6	ATP binding cassette subfamily C member 6	16p13.11	33	Response to drug
RPGRIP1L	RPGRIP1 like	16q12.2	36	Cerebellum development
CST3	Cystatin C	20p11.21	4	Brain development
CBS	Cystathionine-beta-synthase	21q22.3	26	Homocysteine metabolic process
G6PD	Glucose-6-phosphate dehydrogenase	Xq28	14	Lipid metabolic process
MTHFR	Methylenetetrahydrofolate reductase	1p36.22	13	Methionine metabolic process
DPYD	Dihydropyrimidine dehydrogenase	1p21.3	26	Purine nucleobase catabolic process
DBT	Dihydrolipoamide branched chain transacylase E2	1p21.2	14	Branched-chain amino acid catabolic process
F5	Coagulation factor V	1q24.2	25	ER to Golgi vesicle-

				mediated transport
FMO3	Flavin containing monooxygenase 3	1q24.3	10	Oxidation-reduction process
CFH	Complement factor H	1q31.3	25	Regulation of complement activation
GNPAT	Glyceronephosphate O-acyltransferase	1q42.2	17	Synapse assembly
ACAD9	Acyl-CoA dehydrogenase family member 9	3q21.3	19	Mitochondrial respiratory chain complex I assembly
KLKB1	Kallikrein B1	4q35.2	17	Positive regulation of fibrinolysis
IL7R	Interleukin 7 receptor	5p13.2	8	Immune response
HEXB	Hexosaminidase subunit beta	5q13.3	15	Astrocyte cell migration
PRSS1	Protease, serine 1	7q34	7	Cobalamin metabolic process
CA2	Carbonic anhydrase 2	8q21.2	7	Positive regulation of synaptic transmission, GABAergic
ADAMTs13	ADAM metalloproteinase with thrombospondin type 1 motif 13	9q34.2	32	Glycoprotein metabolic process
ANKK1	Ankyrin repeat and kinase domain containing 1	11q23.2	10	Protein phosphorylation
ACADS	Acyl-CoA dehydrogenase, C-2 to C-3 short chain	12q24.31	11	Fatty acid beta-oxidation
HNF1A	HNF1 homeobox A	12q24.31	9	Glucose homeostasis
BBS2	Bardet-Biedl syndrome 2	16q13	18	Hippocampus development
APOE	Apolipoprotein E	19q13.32	6	AMPA glutamate receptor clustering
DGCR6	DiGeorge syndrome critical region gene 6	22q11.21; 22q11	6	Animal organ morphogenesis
PRODH	Proline dehydrogenase 1	22q11.21	15	Positive regulation of cell death
ARSA	Arylsulfatase A	22q13.33	9	glycosphingolipid metabolic process
CPT2	Carnitine palmitoyltransferase 2	1p32.3	5	fatty acid beta-oxidation
KEL	Kell blood group, metallo-endopeptidase	7q34	20	regulation of axon diameter
HR	HR, lysine demethylase and nuclear receptor corepressor	8p21.3	19	histone H3-K9 demethylation
COL5A1	Collagen type V alpha 1 chain	9q34.3	67	blood vessel development
UCP3	Uncoupling protein 3	11q13.4	7	aging
SCNN1G	Sodium channel epithelial 1 gamma subunit	16p12.2	13	sodium ion transmembrane transport
GMPPB	GDP-mannose pyrophosphorylase B	3p21.31	9	GDP-mannose biosynthetic process
FOXC1	Forkhead box C1	6p25.3	1	Cerebellum development

PRF1	Perforin 1	10q22.1	3	apoptotic process
MYLK2	Myosin light chain kinase 2	20q11.21	13	neuromuscular synaptic transmission
JPH2	Junctophilin 2	20q13.12	8	calcium ion homeostasis
FBN1	Fibrillin 1	15q21.1	66	embryonic eye morphogenesis
PRSS56	Protease, serine 56	2q37.1	13	camera-type eye development
AGXT	alanine-glyoxylate aminotransferase	2q37.3	11	L-alanine catabolic process
TPMT	Thiopurine S-methyltransferase	6p22.3	10	methylation
LOXHD1	Lipoxygenase homology domains 1	18q21.1	52	Sensory perception of sound
ABCB4	ATP binding cassette subfamily B member 4	7q21.12	31	brain development
TLR4	Toll like receptor 4	9q33.1	4	T-helper 1 type immune response
MEF2A	Myocyte enhancer factor 2A	15q26.3	21	dendrite morphogenesis
ADA	Adenosine deaminase	20q13.12	12	embryonic digestive tract development
GBE1	1,4-alpha-glucan branching enzyme 1	3p12.2	16	glycogen biosynthetic process
ATP6V0A4	ATPase H+ transporting V0 subunit a4	7q34	24	sensory perception of sound
PTH	Parathyroid hormone	11p15.3	4	bone resorption
MYF6	Myogenic factor 6	12q21.31	3	skeletal muscle tissue development
BRCA1	BRCA1, DNA repair associated	17q21.31	24	G2 DNA damage checkpoint
GLDC	Glycine decarboxylase	9p24.1	25	cellular response to leukemia inhibitory factor
ECM1	Extracellular matrix protein 1	1q21.2	10	positive regulation of angiogenesis
SPTA1	Spectrin alpha, erythrocytic 1	1q23.1	53	axon guidance
ACKR1	Atypical chemokine receptor 1 (Duffy blood group)	1q23.2	2	defense response
TOR1A	Torsin family 1 member A	9q34.11	5	synaptic vesicle transport
PAH	Phenylalanine hydroxylase	12q23.2	14	neurotransmitter biosynthetic process
CDH15	Cadherin 15	16q24.3	14	adherens junction organization
EPOR	Erythropoietin receptor	19p13.2	8	brain development
TRMU	tRNA 5-methylaminomethyl-2-thiouridylate methyltransferase	22q13.31	12	mitochondrial tRNA thio-modification
CTH	Cystathionine gamma-lyase	1p31.1	13	cysteine metabolic process
CPN1	Carboxypeptidase N subunit 1	10q24.2	9	insulin processing
LIG4	DNA ligase 4	13q33.3	5	central nervous system development

RPGRIP1	Retinitis pigmentosa GTPase regulator interacting protein 1	14q11.2	25	neural precursor cell proliferation
ITPA	Inosine triphosphatase	20p13	12	chromosome organization
PPOX	Protoporphyrinogen oxidase	1q23.3	19	heme biosynthetic process
EPHX2	Epoxide hydrolase 2	8p21.2-p21.1	22	inflammatory response
ELAC2	ElaC ribonuclease Z 2	17p12	24	mitochondrial tRNA processing
SEMA4A	Semaphorin 4A	1q22	18	angiogenesis
NPHS2	NPHS2, podocin	1q25.2	8	actin cytoskeleton reorganization
ASB10	Ankyrin repeat and SOCS box containing 10	7q36.1	7	intracellular signal transduction
ZFPM2	Zinc finger protein, FOG family member 2	8q23.1	9	Negative regulation of cell death
AARS	Alanyl-tRNA synthetase	16q22.1	21	cerebellar Purkinje cell layer development
AARS2	Alanyl-tRNA synthetase 2, mitochondrial	6p21.1	23	mitochondrial alanyl-tRNA aminoacylation
ABCA7	ATP binding cassette subfamily A member 7	19p13.3	47	cholesterol efflux
ABCD1	ATP binding cassette subfamily D member 1	Xq28	11	fatty acid beta-oxidation
ABHD12	Abhydrolase domain containing 12	20p11.21	17	adult walking behavior
ACO2	Aconitase 2	22q13.2	18	citrate metabolic process
CYP2W1	Cytochrome P450 family 2 subfamily W member 1	7p22.3	10	aflatoxin B1 metabolic process
CYP7B1	Cytochrome P450 family 7 subfamily B member 1	8q12.3	8	B cell chemotaxis
D2HGDH	D-2-hydroxyglutarate dehydrogenase	2q37.3	20	response to cobalt ion
DAO	D-amino acid oxidase	12q24.11	12	D-serine metabolic process
DARS	Aspartyl-tRNA synthetase	2q21.3	17	aspartyl-tRNA aminoacylation
DARS2	Aspartyl-tRNA synthetase 2, mitochondrial	1q25.1	18	tRNA aminoacylation
DCAF17	DDB1 and CUL4 associated factor 17	2q31.1	19	protein ubiquitination
DLAT	Dihydrolipoamide S-acetyltransferase	11q23.1	14	Pyruvate metabolic process
DMPK	DM1 protein kinase	19q13.32	16	nuclear envelope organization
DMXL2	Dmx like 2	15q21.2	47	vacuolar acidification
DNAJC13	DnaJ heat shock protein family (Hsp40) member C13	3q22.1	57	endosome organization
SYT14	Synaptotagmin 14	1q32.2	13	NOT calcium ion-regulated exocytosis of neurotransmitter
TAF15	TATA-box binding protein associated factor 15	17q12	17	positive regulation of transcription, DNA-

				templated
TARBP1	TAR (HIV-1) RNA binding protein 1	1q42.2	34	enzyme-directed rRNA 2'-O-methylation
TBCE	Tubulin folding cofactor E	1q42.3	18	peripheral nervous system neuron axonogenesis
TBCK	TBC1 domain containing kinase	4q24	33	actin cytoskeleton organization
TBK1	TANK binding kinase 1	12q14.2	24	regulation of neuron death
TECPR2	Tectonin beta-propeller repeat containing 2	14q32.31	20	autophagy
TENM4	teneurin transmembrane protein 4	11q14.1	36	central nervous system myelin formation
TFG	TRK-fused gene	3q12.2	12	ER to Golgi vesicle-mediated transport
TGM6	Transglutaminase 6	20p13	13	peptide cross-linking
TMEM216	Transmembrane protein 216	11q12.2	6	cilium assembly

En la figura 4 se describen los diferentes efectos causados por las variantes encontradas, organizadas por la cantidad total. El efecto de cambio de sentido o *missense* fue el más común con más del 50% (93 variantes), seguido de las variantes sinónimas el cual estuvo representado con un 13% (24 variantes).

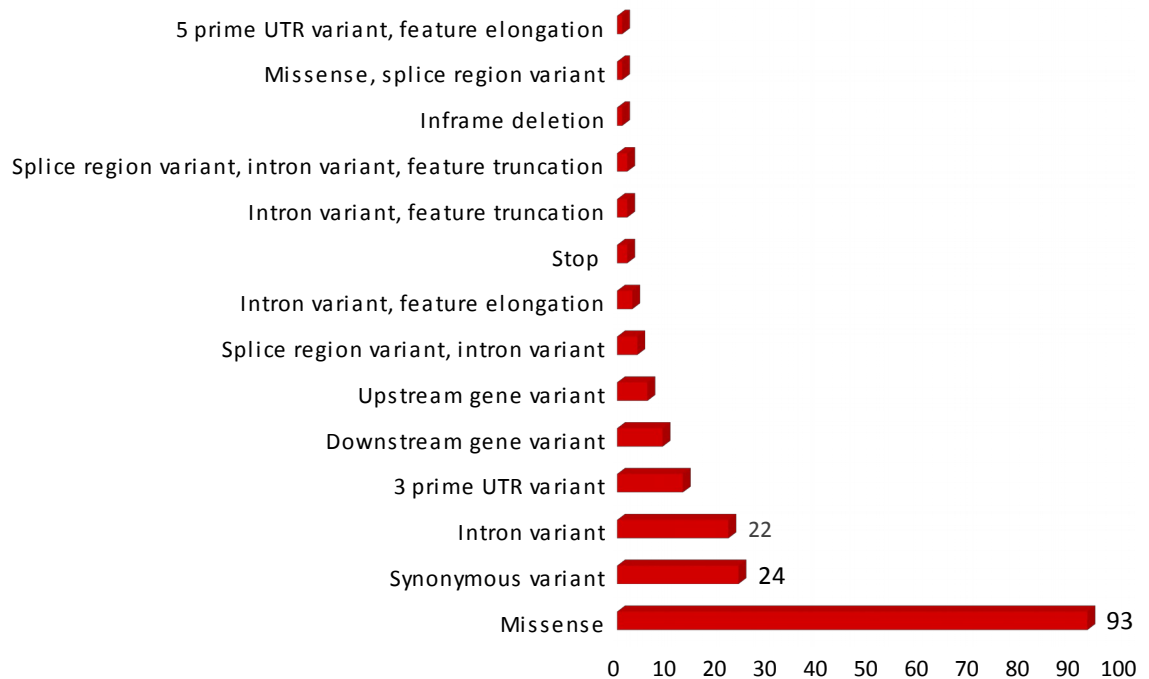


Figura 4. Cantidad de mutaciones clasificadas por efectos causados encontradas en la secuenciación de exoma completo

Se presentó una alta tasa de transiciones (76.3%) con respecto a las transversiones (23.7%) como era de esperarse, ya que es mucho más común encontrar variantes de este tipo en el genoma en general (Tabla 3).

Tabla 3. Número de transiciones y transversiones totales encontradas en la secuenciación de exoma completo

Tipo de mutación	Cantidad	%Total
Transición	132	76.3
Transversión	41	23.7
Total	173	100.0

En la tabla 4 se muestran los resultados de la secuenciación de exoma completo realizada a 16 pacientes del sur-occidente colombiano, especificando cada variante encontrada por gen, el cromosoma en el que se encuentra, la posición, efecto, cambio de aminoácido y número RS de ClinVar. Estos resultados son de gran importancia para la creación de una base de datos con las variantes propias de la población del sur-occidente colombiano, abriendo la posibilidad a realizar un estudio ampliado con muestras de todo el país y así crear una base de datos nacional.

Uno de los resultados más destacables de la secuenciación exómica fue el hallazgo de 2 variantes no registradas previamente en la literatura para los genes *ATP binding cassette subfamily D member 1* ABCD1 y *Dihydrolipoamide S-acetyltransferase* DLAT, señaladas en color rojo en la tabla 4. Al ser sometidas a análisis con los paquetes SIFT y PolyPhen, dichas variantes registran un efecto en expresión proteica normal. Además de esto, cabe destacar que los genes EVC, G6PD, PRSS56, y AGXT presentaron variantes en más de un paciente.

Tabla 4. Variantes exómicas encontradas por gen, en la secuenciación de exoma completo

Pacien.	Gen	Variante	Crom.	Posición	Efecto	A.A	ClinVar RS
1	HPSE2	G>G/A	10	100242490	stop	R/*	rs267606866
	ATP7B	A>A/C	13	52535985	missense	M/R	rs121907998
	POF1B	C>C/T	X	84563194	missense	R/Q	rs75398746
2	EVC	G>A	4	5755524	missense	R/Q	rs35953626
	DSPP	A>T	4	88533540	missense	R/W	rs36094464
	CYP21A 2	T>A	6	32007593	missense	V/E	rs6476
	SCNN1A	A>G	12	6458350	missense	W/R	rs5742912
	ABCC6	C>T	16	16251599	missense	R/Q	rs2238472
	RPGRIP 1L	C>T	16	53720436	missense	A/T	rs61747071
	CST3	C>T	20	23618427	missense	A/T	rs1064039
	CBS	A>G	21	44483184	missense	I/T	rs5742905
	G6PD	T>C	X	153763492	missense	N/D	rs1050829
		C>T	X	153764217	missense	V/M	rs1050828
3	MTHFR	G>A	1	11856378	missense	A/V	rs1801133
	DPYD	G>A	1	98348885	missense	C/R	rs1801265
	DBT	T>C	1	100672060	missense	G/S	rs12021720
	F5	T>C	1	169519049	missense	R/Q	rs6025
	FMO3	G>A	1	171076966	missense	E/K	rs2266782
	CFH	C>T	1	196659237	Synonymous variant	Y/H	rs1061170
	GNPAT	A>G	1	231408091	missense	D/G	rs11558492
	ACAD9	G>A	3	128622922	missense	A/T	rs115532916
	KLKB1	G>A	4	187158034	missense	N/S	rs3733402
	ILR7	T>C	5	35861068	missense	T/I	rs1494558
		G>A	5	35871190	missense	I/V	rs1494555
	HEXB	T>C	5	73981270	missense	S/L	rs820878
	PRSS1	C>T	7	142458412	missense	A/V	rs202003805
		A>T	7	142458451	missense	N/I	rs111033566
		A>G	7	142458526	missense	N/S	rs144422014
	CA2	A>G	8	86392989	missense	N/D	rs2228063
	ADAMTs 13	C>G	9	136301982	missense	Q/E	rs2301612
	ANKK1	G>A	11	113270828	missense	E/K	rs1800497
	ACADS	G>A	12	121176083	missense	G/S	rs1799958
	HNF1A	A>G	12	121437382	missense	G/S	rs1169305
HDP	T>C	12	122295335	missense	A/T	rs1154510	
BBS2	C>T	16	56548501	missense	N/S	rs4784677	

	APOE	T>C	19	45411941	missense	C/R	rs429358
	DGCR6	C>T	22	18901004	missense	Q/R	rs450046
	PRODH	C>T	22	rs2904552	missense	R/H	rs2904552
	ARSA	G>C	22	51064039	missense	T/S	rs743616
	G6PD	T>C	X	153763492	missense	N/D	rs1050829
4	CPT2	C>C/T	1	53668099	missense, splice region variant	S/L	rs74315294
	EVC	G>G/A	4	5755524	missense	R/Q	rs35953626
	KEL	A>A/G	7	142640113	downstream gene variant		rs8176038
	HR	T>T/C	8	21976710	missense	T/A	rs7014851
	COL5A1	G>G/A	9	137642654	missense	G/S	rs61735045
	UCP3	G>G/A	11	73716889	stop	R/*	rs10489431
	SCNN1G	G>G/A	16	23200921	missense	G/S	rs5736
5	GMPPB	C>C/T	3	4975926	upstream gene variant		rs397509422
	FOXC1	C>C/T	6	1611569	missense	P/S	rs79691946
	PRF1	C>C/T	10	72360648	missense	R/H	rs35418374
	MYLK2	C>C/A	20	30408160	missense	A/E	rs121908108
	JPH2	C>C/T	20	42744802	missense	G/S	rs140740776
6	FBN1	G>C/C	15	48720526	Intron variant	N/A	N/A
		C>T/T	15	48807637	missense	C/Y	N/A
7	FMO3	G>G/A	1	171080080	missense	V/M	rs1736557
	PRSS56	C>C/G	2	233390199	missense	P/A	rs61744404
	AGXT	C>C/T	2	241808314	missense	P/L	rs34116584
	TPMT	T>T/C	6	18130918	missense	Y/C	rs1142345
		C>C/T	6	18139228	missense	A/T	rs1800460
	LOXHD1	G>G/T	18	44104697	Synonymous variant	R	rs75949023
8	ABCB4	C>C/T	7	87060844	missense	R/Q	rs45575636
	TLR4	C>C/T	9	120475602	missense	T/I	rs4986791
	MEF2A	C>C/T	15	100230605	missense	P/L	rs121918529
	ADA	C>C/T	20	43280227	downstream gene variant		rs73598374
9	GBE1	G>G/A	3	81627151	missense	R/C	rs80338672
	ATP6V0 A4	A>A/G	7	138417791	missense	M/T	rs3807153
	PTH	G>G/T	11	13514053	Synonymous variant	R	rs6256
	MYF6	G>G/T	12	81101832	missense	A/S	rs28928909
	BRCA1	G>G/A	17	41197752	Synonymous variant	Y	rs80356977
10	GLDC	C>C/T	9	6589230	Synonymous variant	R	rs121964976

11	ECM1	C>C/T	1	150482172	downstream gene variant		rs121909115
	SPTA1	G>G/T	1	158624528	missense	A/D	rs35948326
	DARC	C>C/T	1	159175494	upstream gene variant		rs34599082
	PRSS56	C>G/G	2	233390199	missense	P/A	rs61744404
	AGXT	C>C/T	2	241808314	missense	P/L	rs34116584
	GMPPB	C>C/A	3	49761081	upstream gene variant		rs142336618
	TOR1A	C>C/G	9	132580901	missense	D/H	rs1801968
	PAH	G>G/T	12	103288604	missense	S/R	rs62516151
	CDH15	C>C/T	16	89246680	missense	R/W	rs121434540
	EPOR	T>T/C	19	11488727	downstream gene variant		rs62638745
	TRMU	G>G/T	22	46731689	missense	A/S	rs11090865
12	CTH	G>G/T	1	70904800	missense	S/I	rs1021737
	AGXT	C>C/T	2	241808314	missense	P/L	rs34116584
	CPN1	C>C/T	10	101829514	missense	G/D	rs61751507
	LIG4	G>G/A	13	108863591	missense	T/I	rs1805388
	RPGRIP1	G>G/T	14	21790040	missense	A/S	rs10151259
	ITPA	C>C/A	20	3193842	missense	P/T	rs1127354
	CBS	A>A/G	21	44483184	missense	I/T	rs5742905
13	PPOX	C>C/G	1	161138933	downstream gene variant		rs12735723
	EPHX2	G>G/A	8	27373865	missense	R/Q	rs751141
	CA2	A>A/G	8	86392989	missense	N/D	rs2228063
	AIPL1	GGGTGGC TCTGCA>G GGTGGCT CTGCA/G	17	6328870	inframe deletion	PAEP P/P	rs281865195
	ELAC2	G>G/A	17	12915009	missense	S/L	rs4792311
14	SEMA4A	G>G/A	1	156146640	missense	R/Q	rs41265017
	NPHS2	G>G/A	1	179544941	missense	P/L	rs74315344
	ASB10	C>C/T	7	150884003	missense	R/H	rs104886488
	ZFPM2	A>A/G	8	106814279	missense	S/G	rs28374544
15	AARS	A>A/G	16	70287177	upstream gene variant	None	N/A
	AARS2	T>C/C	6	44268371	upstream gene variant	None	N/A
		T>C/C	6	44269023	upstream gene variant	None	N/A
		T>C/C	6	44274797	intron variant	None	N/A
		T>C/C	6	44275011	missense	I/V	N/A
	ABCA7	A>A/G	19	1047161	Synonymous variant	G	N/A

		A>A/G	19	1051214	Synonymous variant	V	N/A
		C>C/T	19	1052005	Synonymous variant	A	N/A
		T>T/A	19	1052017	Synonymous variant	G	N/A
		A>A/G	19	1054060	Synonymous variant	L	N/A
		G>G/A	19	1055191	missense	R/Q	N/A
		G>C/C	19	1056492	missense	G/A	N/A
		A>A/G	19	1059004	intron variant	None	N/A
		G>T/T	19	1065018	missense	A/S	N/A
		C>C/T	19	1065044	Synonymous variant	G	N/A
ABCD1		G>G/A	X	153006137	missense	V/I	N/A
		T>T/A	X	153006141	missense	V/E	N/A
		G>A/A	20	25288632	Synonymous variant	R	N/A
	ABHD12	GGCCCCC >G/G	20	25371112	intron variant, feature truncation	None	N/A
		C>G/G	20	25371125	intron variant	None	N/A
		G>GGCCT CCGCC/GG CCTCCGC C	20	25371378	5 prime UTR variant, feature elongation	None	N/A
ACO2		C>C/T	22	41911525	Synonymous variant	L	N/A
CYP2W1		G>G/A	7	1024091	missense	R/H	N/A
CYPB7B1		GA>GA/G	8	65528842	splice region variant, intron variant, feature truncation	None	N/A
	D2HGDH	T>G/G	2	242674959	intron variant	None	N/A
		A>A/G	2	242680430	intron variant	None	N/A
		T>T/C	2	242684115	intron variant	None	N/A
		C>C/T	2	242688292	intron variant	None	N/A
		C>C/T	2	242688366	intron variant	None	N/A
		C>C/T	2	242690745	missense	A/V	N/A
		A>A/G	2	242707101	intron variant	None	N/A
	DAO	C>C/A	12	109277492	intron variant	None	N/A
		T>T/G	12	109278747	intron variant	None	N/A
		G>G/GA	12	109293109	intron variant, feature elongation	None	N/A
DARS		T>T/C	2	136740900	intron variant	None	N/A

	DARS2	T>T/C	1	173797451	intron variant	None	N/A
	DCAF17	C>T/T	2	172291616	Synonymous variant	V	N/A
		G>T/T	2	172329458	intron variant	None	N/A
		A>G/G	2	172330393	Synonymous variant	Q	N/A
		C>T/T	2	172336541	splice region variant, intron variant	None	N/A
		G>A/A	2	172338194	3 prime UTR variant	None	N/A
		T>G/G	2	172338300	3 prime UTR variant	None	N/A
		C>T/T	2	172339964	3 prime UTR variant	None	N/A
		T>A/A	2	172340049	3 prime UTR variant	None	N/A
		T>C/C	2	172340419	3 prime UTR variant	None	N/A
		C>G/G	2	172340736	3 prime UTR variant	None	N/A
		A>G/G	2	172341327	3 prime UTR variant	None	N/A
		DLAT	C>C/T	11	111896324	missense	A/V
	GT>GT/G		11	111897028	splice region variant, intron variant, feature truncation	None	N/A
	A>A/T		11	111897050	intron variant	None	N/A
	T>T/C		11	111908162	missense	V/A	N/A
	AG>A/A		11	111916001	intron variant, feature truncation	None	N/A
	G>G/A		11	111916647	missense	D/N	N/A
	C>C/T		11	111930576	intron variant	None	N/A
	A>A/G		11	111933088	intron variant	None	N/A
	DMPK	A>A/C	19	46282503	downstream gene variant	None	N/A
	DMXL2	A>A/G	15	51748564	splice region variant, synonymous variant	L	N/A
	DNAJC13	G>T/T	3	132218623	missense	A/S	N/A
	SYT14	A>G/G	1	210337043	3 prime UTR variant	None	N/A
		G>A/A	1	210335066	3 prime UTR variant	None	N/A
		T>C/C	1	210335028	3 prime UTR	None	N/A

				variant		
	GTTAATT> G/G	1	210334665	3 prime UTR variant	None	N/A
TAF15	C>C/T	17	34171827	Synonymous variant	Y	N/A
	G>G/GT	17	34165601	intron variant, feature elongation	None	N/A
TARBP1	G>G/T	1	234603386	Synonymous variant	L	N/A
	T>T/C	1	234569192	Synonymous variant	A	N/A
	T>T/C	1	234563060	splice region variant, intron variant	None	N/A
	A>A/G	1	234541801	Synonymous variant	C	N/A
TBCE	G>G/A	1	235590559	splice region variant, intron variant	None	N/A
	C>C/T	1	235543470	splice region variant, intron variant	None	N/A
TBCK	A>A/G	4	107151587	Synonymous variant	C	N/A
	A>A/G	4	107133919	Synonymous variant	G	N/A
	A>A/G	4	107115960	Intron variant	None	N/A
TBK1	T>T/A	12	64875787	Synonymous variant	I	N/A
	A>A/AT	12	64858100	Intron variant, feature elongation	None	N/A
TECPR2	A>A/T	14	102965437	3 prime UTR variant	None	N/A
	A>A/G	14	102965435	3 prime UTR variant	None	N/A
	G>G/A	14	102894593	missense	V/I	N/A
TENM4	G>A/A	11	78443586	Synonymous variant	I	N/A
	G>A/A	11	78372536	Synonymous variant	L	N/A
TFG	T>T/C	3	100467018	downstream gene variant	None	N/A
	A>A/G	3	100438784	intron variant	None	N/A
TGM6	C>T/T	20	2413126	intron variant	None	N/A
	G>A/A	20	2380396	intron variant	None	N/A
	A>G/G	20	2375262	missense	M/V	N/A
TMEM21 6	G>G/C	11	61165741	downstream gene variant	None	N/A

		C>C/CA	11	61165731	downstream gene variant	None	N/A
--	--	--------	----	----------	-------------------------	------	-----

4.2. ANÁLISIS GENÓMICO Y BIOINFORMÁTICO

Con el diagrama de barras obtenido a partir de la búsqueda de interacciones proteína-proteína de cada gen evaluado en este trabajo, se encontró un total de 3748 interacciones (Figura 5). La proteína con mayor número de interacciones fue *Breast cancer 1* BRCA1 (702 interacciones), seguido de *Transmembrane protein 216* TMEM216 (244 interacciones), *TANK binding kinase 1* TBK1 (158 interacciones) y *RPGRIP1 like* RPGRIP1L (157 interacciones).

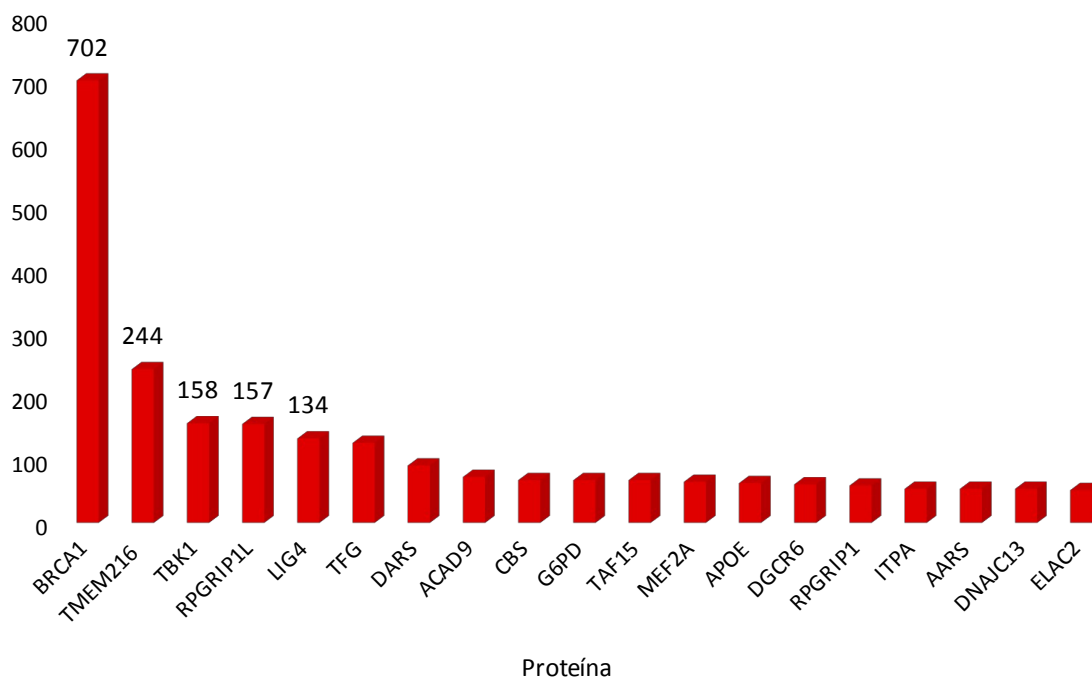


Gráfico 1. Número de interacciones proteína-proteína para cada uno de los genes evaluados, extraídas de la base de datos BioGRID. Se presentan solo los genes con mayor número de interacciones.

En cuanto a la red de interacciones proteína-proteína realizada con base en las proteínas codificadas por los genes que presentaron variantes exómicas (Figura 6A), tuvo un total de 2758 interacciones. Se puede destacar la presencia de los

dos nodos más grandes, es decir, con más interacciones proteicas, BRCA1 y TMEM216 (Figura 6B).

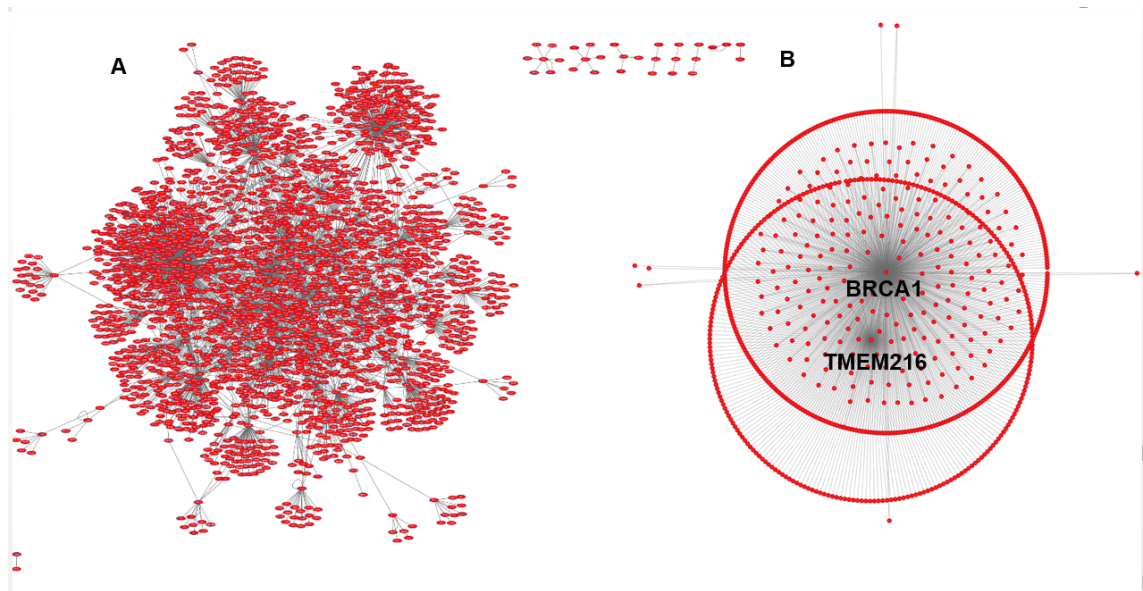


Figura 5. Redes de interacción proteína-proteína construidas con base en las proteínas asociadas a discapacidad cognitiva según la secuenciación de exoma completo con el programa Cytoscape 3.2. A) Red completa B) Sub-red mostrando los dos nodos con mayor cantidad de interacciones.

Los parámetros topológicos de la red mostraron un coeficiente de agrupamiento de 0.009, 10 componentes conectados, una centralización de la red de 0.207 y heterogeneidad de 5.547 (Tabla 5)

Tabla 5. Parámetros topológicos de la red de genes asociados a discapacidad cognitiva, extraídos del programa Cytoscape 3.2 usando la herramienta *Network Analyzer*

Parámetro topológico	Valor
Coefficiente de agrupamiento	0.009
Componentes conectados	10
Diámetro de la red	10

Radio de la red	1
Centralización de la red	0.207
Número promedio de vecinos	2.498
Número de nodos	2758
Densidad de la red	0.001
Heterogeneidad de la red	5.547

De las categorías GO específicamente de procesos biológicos más destacados de la red se pueden destacar la sinapsis (P-valor 8,79E-28), la regulación positiva de la muerte celular (P-valor 1,26E-23), Regulación negativa de la muerte neuronal (P-valor 8,62E-23) y de la migración celular (P-valor 8,73E-23) y la regulación positiva de la transmisión sináptica (P-valor 1,01E-22) (Tabla 6). Cada uno de estos procesos biológicos que se dan normalmente en cualquier individuo sano es de gran importancia para el mantenimiento de las habilidades cognitivas normales, por lo que se asume que una falla producida en proteínas asociadas con estos procesos ocasionaría una afectación grave en la cognición.

Tabla 6. Categorías GO-Procesos biológicos destacados de la red de genes asociados a discapacidad cognitiva, extraídas del programa Cytoscape 3.1 y Quick GO.

GO_ID	Descripción del proceso	P-valor (Bonferroni)
9987	Fosforilación oxidativa	1,56E-46
44237	Cadena respiratoria de transporte de electrones	2,35E-43
8152	Proceso de oxidación-reducción	2,81E-36
44260	Proceso catabólico de proteínas dependiente de ubiquitina través de la vía de clasificación cuerpos multivesiculares	6,56E-36
44238	Transporte de electrones mitocondrial, NADH a ubiquinona	2,57E-33
43170	Cascada MAPK	1,31E-32
16043	Sinapsis	8,79E-28
48518	Regulación positiva de la muerte celular	1,26E-23
48523	Regulación negativa de la muerte neuronal	8,62E-23
48519	Regulación negativa de la migración celular	8,73E-23
48522	Regulación positiva de la transmisión sináptica	1,01E-22
44267	Desacetilación de histonas	2,63E-22
34641	Síntesis de ATP acoplado al transporte de electrones	2,96E-22

44419	Patogénesis	5,45E-22
6139	Transporte de electrones mitocondrial, Ubiquinol a citocromo c	8,87E-22

5. DISCUSIÓN

El uso de NGS, específicamente de secuenciación de exoma completo de pacientes con discapacidad cognitiva en este trabajo, logró la identificación de 108 genes afectados, presentando un total de 184 variantes exómicas. Además como se mencionó anteriormente se hallaron 2 nuevas variantes, en los genes ABCD1 y DLAT que no han sido previamente reportadas en la literatura, lo cual es un gran aporte a la información que se tiene de la estructura genómica de la población Colombiana, específicamente del sur-occidente colombiano. Todas las variantes aquí reportadas servirán para la creación de una base de datos autóctona, que abrirá el camino a nuevos estudios ampliados que cubran más regiones del país hasta reunir información de pacientes con discapacidad cognitiva de todo el país.

Específicamente la variante nueva encontrada en el gen ABCD1 fue una transición G>G/A con cambio de aminoácido (*missense*) Valina por Isoleucina, ubicada en la posición 153006137 del cromosoma X. El gen ABCD1 codifica un transportador peroxisomal que permite el paso de ácidos grasos desde el citosol hasta el interior del peroxisoma para su consiguiente degradación por β -oxidación. Este gen está asociado al desorden ligado al cromosoma X Adrenoleucodistrofia (ALD), el cual se caracteriza por la acumulación de cadenas muy largas de aminoácidos en el plasma y tejidos afectados (32). Las manifestaciones de esta enfermedad ocurren primeramente en el córtex adrenal, la mielina del sistema nervioso central y las células Leydig de los testículos. Moser y colaboradores (33) explicaron 7 fenotipos

posibles de ADL incluyendo la forma cerebral en la niñez, adrenomieloneuropatía (AMN), cerebral adulta, adolescente, insuficiencia adrenal sin enfermedad neurológica, asintomática y heterocigotos. Los adultos presentan una forma mucho más crónica de la enfermedad (AMN) la cual involucra la médula espinal y los nervios periféricos aunque los hemisferios cerebrales pueden afectarse también. Estas características según el estudio de Edwin y colaboradores (34) fueron causales de la discapacidad cognitiva del 60% de pacientes evaluados con AMN.

La segunda variante nueva encontrada en el presente trabajo fue la presente en el gen DLAT, específicamente una transición C>C/T con cambio de aminoácido (*missense*) Alanina por Valina, ubicada en la posición 111896324 del cromosoma 11. El gen DLAT codifica el componente E2 del complejo multienzimático de la piruvato deshidrogenasa (PDC), el cual reside en la membrana interna mitocondrial y cataliza la conversión del piruvato en acetil coenzima A (35). La deficiencia en piruvato deshidrogenasa es una causa principal de acidosis láctica primaria y disfunción neurológica en infancia y niñez; a pesar de que esta deficiencia casi siempre es causada por mutaciones encontradas en el gen que codifica para la sub-unidad E1-alfa del complejo PDC, a partir del año 2005 con la publicación de Head y colaboradores (36) donde se reportaron dos casos de personas con la deficiencia en piruvato deshidrogenasa cuya causa residía en mutaciones en el gen DLAT que ocasionaban una proteína E2 malfuncional. Las características clínicas encontradas en estos uno de los pacientes de dicho estudio fueron ataxia con retraso motor grueso y fino, apraxia oculomotora y movimientos distónicos en cara, manos, pies e hiperreflexia. Por otro lado el segundo paciente desarrolló signos mucho más prominentes como retraso en el crecimiento, discapacidad cognitiva no tan severa y distonia generalizada. Otro estudio más reciente de McWilliam y colaboradores (37) presentó evidencias adicionales del fenotipo mostrado en el estudio de Head y colaboradores (36) y su relación con mutaciones en el gen DLAT. Dicho estudio describió el caso de dos

hermanas con síntomas iniciales de distonia y deficiencia de la piruvato deshidrogenasa causada por defectos en la subunidad E2.

Por otro lado, según el análisis bioinformático, los genes que codifican para las proteínas BRCA1 y TMEM216 son los que más número de interacciones proteína-proteína presentan, lo cual implica que cualquier mutación que cause un daño en la proteína y su función, afectará diversas rutas metabólicas y procesos en los que se ve implicada. El gen BRCA1 se encuentra ubicado en el cromosoma 17 y codifica una fosfoproteína nuclear con el rol de mantener la estabilidad genómica, además de actuar como un supresor de tumor. Acerca de este gen se conoce ampliamente su implicación en cáncer de seno y ovario, sin embargo, recientemente se ha aclarado mucho más cuál es la función de éste en el desarrollo del cerebro y de hecho, su contribución en el déficit cognitivo asociado a Alzheimer (AD). En el trabajo de Pao y colaboradores (38) se explica que BRCA1, a pesar de tener grandes implicaciones como supresor de tumores, presenta su máxima expresión en el neuroepitelio embrionario cuando los progenitores neurales proliferan más. Dichos autores inactivaron el gen BRCA1 en un cerebro en desarrollo y encontraron que el fenotipo resultante se caracterizaba por agenesia severa de múltiples estructuras cerebrales laminadas, afectando notablemente al neocórtex, hipocampo, cerebelo y bulbo olfatorio. Cabe resaltar la importancia de estructuras como el hipocampo para todo el proceso de aprendizaje y memoria que se ve afectado en personas con discapacidad cognitiva. Otro de los estudios que demuestra la implicación del gen BRCA1 y la discapacidad cognitiva es el realizado por Suberbielle y colaboradores (39) en el que se encontraron niveles reducidos de dicho gen en cerebros de pacientes con AD, concluyendo que BRCA1 es regulado por la actividad neuronal al mismo tiempo que protege el genoma neuronal en cuanto a su integridad y funciones cognitivas.

En cuanto al gen TMEM216, codifica una proteína transmembrana que se requiere para la ciliogénesis específica de ciertos tejidos. Mutaciones en este gen se han

visto asociadas al síndrome Meckel-Gruber tipo 2 y al síndrome de Joubert 2, el cual también se conoce como síndrome de cerebelo-oculorenal 2. Pacientes con el síndrome de Joubert 2 sufren de enfermedades neurológicas que se manifiestan en retardo psicomotor, hipotonía y ataxia entre otros; precisamente en el estudio realizado por Edvardson y colaboradores (40) se identificó la mutación R12L en el gen TMEM216 en 13 pacientes con el síndrome de Joubert 2. Entre otros síntomas, todos los pacientes presentaron un funcionamiento cognitivo que fue gradualmente empeorando, de un rango ligero a severo.

Por otro lado, las categorías GO específicamente los procesos biológicos destacados de la red construida en Cytoscape 3.2, evidenciaron entre otros, 5 procesos principales altamente implicados en el buen mantenimiento de un proceso cognitivo normal, la sinapsis, la regulación positiva de la transmisión sináptica, la regulación de la muerte neuronal y la migración celular (41). La sinapsis por su parte es la comunicación entre neuronas, de hecho en cuanto las neuronas terminan su migración en el cerebro en desarrollo, extienden sus axones y dendritas para establecer contacto con otras neuronas. A medida que el cerebro va completando su desarrollo, se forman millones de conexiones sinápticas y se forma todo el circuito neuronal (42). De hecho, aunque el cerebro alcanza aproximadamente el 90% de tamaño entre los 2 y 5 años de edad (43), las conexiones sinápticas continúan formándose por mucho más tiempo, interviniendo en los procesos cognitivos de la persona.

En cuanto al proceso biológico de la muerte neuronal, se encuentra asociado a enfermedades neurodegenerativas como la esclerosis múltiple, el Alzheimer, Parkinson en las cuales se ven afectadas seriamente las funciones cognitivas (44). En los últimos años se ha identificado gran cantidad de mutaciones y polimorfismos en pacientes afectados de enfermedades neurodegenerativas. El descubrimiento de estas alteraciones genéticas ha ayudado a comprender mejor la patogenia de estas enfermedades (45).

Otro de los resultados obtenidos en este trabajo se relaciona con la alta cantidad de variantes de tipo transicionales en comparación con las transversiones. De hecho esto concuerda con lo reportado en la literatura, Rosenberg y colaboradores (46) explican que un aspecto fundamental de la mutación puntual del ADN es la observación de que los cambios de nucleótidos transicionales ocurren con mayor frecuencia que los cambios de transversión. Al parecer este sesgo se debe principalmente a la estructura bioquímica de las bases de nucleótidos y las propiedades químicas de la asociación de bases complementarias (47). Sin embargo otro aspecto interesante es que según un estudio realizado por Prendergast y colaboradores (48) la tasa de transiciones sinónimas se encontró fuertemente correlacionada positivamente con la estructura de la cromatina, mientras que la tasa de transversiones no mostró tendencia evidente de asociación con dicha estructura.

En la secuenciación se encontraron también una gran cantidad de variantes de tipo *missense* en comparación con otros tipos, esto también concuerda con lo reportado en estudios anteriores. Según Kryukov y colaboradores (49) varios estudios recientes han reportado un exceso significativo de variantes raras *missense* en genes candidatos o vías metabólicas en individuos con valores extremos de fenotipos cuantitativos.

Finalmente vale la pena resaltar que la secuenciación tanto de exoma completo como de exomas ha facilitado en gran manera la detección de novo de SNVs (Single nucleotide variant) en enfermedades genéticas raras en todo el mundo, por lo que su aplicación como herramienta diagnóstica en la población colombiana es de gran utilidad (50). La amplia disponibilidad de tecnología de NGS ha impulsado el estudio de mutaciones de novo en salud y enfermedad. De hecho, estudios piloto realizadas con enfermedades de desarrollo neurológico de inicio temprano han indicado que las mutaciones de novo pueden desempeñar una parte mucho

más importante de lo que se suponía anteriormente, lo que explica en parte por qué estas enfermedades con tan poco *fitness* permanecen frecuentes en la población humana (51).

6. CONCLUSIONES Y PERSPECTIVAS

Con la secuenciación del exoma completo se logran identificar mutaciones poco comunes vinculadas a enfermedades de una manera eficiente, y teniendo en cuenta que el exoma es la región codificante del genoma y precisamente es donde se encuentran la mayoría de mutaciones, esta técnica ha sido de gran ayuda para los estudios genómicos que abordan la etiología de ciertas enfermedades.

Es de gran importancia resaltar que este trabajo aporta dos variantes nuevas asociadas con discapacidad cognitiva en pacientes del sur-occidente colombiano, además de otras 182 variantes haciendo uso de la secuenciación de nueva generación y aportando información para la creación de una base de datos de la estructura genómica propia de la región, con posibilidades de ampliar la investigación hacia más regiones del país y tomar una cantidad mayor de muestras que permitan una representación más fiel del país y completar una base de datos nacional. Por otro lado se podría realizar posteriormente un estudio que asocie los efectos de las variantes con la expresión del gen que posea dichas variantes, de esta forma se podría obtener información más integral de cómo se ve afectada la síntesis de proteínas y su impacto en el metabolismo del paciente.

BIBLIOGRAFÍA

1. Verdugo M. Análisis de la definición de discapacidad intelectual de la asociación americana sobre retraso mental de 2002. Siglo Cero: REDIS 2003; 34(205): 5-19
2. Fernández C, Méndez JL (dir). La discapacidad intelectual en la sociedad. Percepción e integración social [Trabajo de grado]. [Barcelona]: Universitat de Barcelona; 2013
3. Redin C, Gérard B, Lauer J, Herenger Y, Muller J, et al. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. J Med Genet 2014; 51: 724–736
4. Agha Z, Iqbal Z, Azam M, Ayub H, Vissers LELM, et al. Exome Sequencing Identifies Three Novel Candidate Genes Implicated in Intellectual Disability. PLoS ONE 2014; 9(11): e112687
5. Makalowski W. The human genome structure and organization. Acta Biochim Pol 2001; 48(3): 587-598
6. Martínez-Frías ML. Estructura y función del ADN y de los genes. I Tipos de alteraciones de la función del gen por mutaciones. Semergen 2010; 36(5): 273-277.
7. Loewe, L. Genetic mutation. Nature Education 2008; 1(1):113.
8. Clancy, S. Genetic mutation. Nature Education 2008; 1(1):187.
9. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nature Genet 2011; 43: 860–863.
10. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, et al. Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nature Genet 2011; 43, 864–868.

11. Evers C, Staufner C, Granzow M, Paramasivam N, Hinderhofer K, et al. Impact of clinical exomes in neurodevelopmental and neurometabolic disorders. *Mol Genet Metabol* 2017; 121(4): 297-307
12. Burillo-Sanz S, Montes-Cano MA, García-Lozano JR, Ortiz-Fernández L, Ortego-Centeno N, et al. Mutational profile of rare variants in inflammasome-related genes in Behçet disease: A Next Generation Sequencing approach. *Sci Rep* 2017; 7(1): 8453
13. Millholland JM, Li S, Fernandez CA, Shuber AP. Use of next-generation deep sequencing to improve FGFR3 mutation detection in the urine of patients with bladder cancer. *J Clin Oncol* 2012; 30(5_suppl):28
14. Rump P, Jazayeri O, van Dijk-Bos KK, Johansson LF, van Essen AJ, et al. Whole-exome sequencing is a powerful approach for establishing the etiological diagnosis in patients with intellectual disability and microcephaly. *BMC Medical Genomics* 2016; 9:7
15. Nesbitt A, Bhoj EJ, McDonald Gibson K, Yu Z, Denenberg E, et al. Exome sequencing expands the mechanism of SOX5-associated intellectual disability: A case presentation with review of sox-related disorders. *Am J Med Genet A* 2015; 167A(11): 2548-2554
16. Luco SM, Pohl D, Sell E, Wagner JD, Dyment DA, et al. Case report of novel DYRK1A mutations in 2 individuals with syndromic intellectual disability and a review of the literature. *BMC Med Genet* 2016; 17:15
17. Topper S, Ober C, Das S. Exome Sequencing and the Genetics of Intellectual Disability. *Clinical genetics*. 2011; 80(2):117-126
18. Nho K, Corneveaux JJ, Kim S, Lin H, Risacher SL, Shen L, et al. Whole-exome sequencing and imaging genetics identify functional variants for rate of change in hippocampal volume in mild cognitive impairment. *Mol Psychiatry*. NIH Public Access 2013; 18(7): 781.
19. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008; 26(10): 1135-1145.

20. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet.* Nature Publishing Group 2009; 42(1): 30-35.
21. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods.* Nature Publishing Group 2010; 7(2): 111-118.
22. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics The promise of whole-exome sequencing in medical genetics. *J Hum Genet.* Nature Publishing Group 2013; 59(1): 5-15.
23. Depristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43 (5): 491-498.
24. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Nature Publishing Group* 2012; 20(5):490-497.
25. Hoischen A, van Bon BWM, Rodríguez-Santiago B, Gilissen C, Vissers LELM, de Vries P, et al. De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat Genet.* Nature Research 2011; 43(8): 729-731.
26. Krawitz PM, Schweiger MR, Rödelsperger C, Marcelis C, Kölsch U, Meisel C, et al. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat Genet.* Nature Research 2010; 42(10): 827-829.
27. Becker J, Semler O, Gilissen C, Li Y, Bolz HJ, Giunta C, et al. Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. *Am J Hum Genet* 2011; 88(3): 362-371.
28. Pierce SB, Walsh T, Chisholm KM, Lee MK, Thornton AM, Fiumara A, et al. Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *Am J Hum Genet* 2010; 87(2): 282-288.

29. Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, et al. Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am J Hum Genet* 2010; 87(3): 418-423.
30. Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 2010; 42(6): 483.
31. Vissers LELM, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, et al. A de novo paradigm for mental retardation. *Nat Genet. Nature Research* 2010; 42(12): 1109-1112.
32. Musolino PL, Gong Y, Snyder JMT, Jimenez S, Lok J et al. Brain endothelial dysfunction in cerebral adrenoleukodystrophy. *Brain* 2015; 138(11): 3206-3220
33. Moser HW, Loes DJ, Melhem ER, Raymond GV, Bezman L, et al. X-linked adrenoleukodystrophy: overview and prognosis as a function of age and brain magnetic resonance imaging abnormality: a study involving 372 patients. *Neuropediatrics* 2000; 31: 227-239
34. Edwin D, Speedie L, Naidu S, Moser H. Cognitive impairment in adult-onset adrenoleukodystrophy. *Mol Chem Neuropathol* 1990; 12(3): 167-176
35. Ciara E, Rokicki D, Halat P, Karkucińska-Więckowska A, Piekutowska-Abramczuk D, et al. Difficulties in recognition of pyruvate dehydrogenase complex deficiency on the basis of clinical and biochemical features. The role of next generation sequencing. *Mol Genet Metab Rep* 2016; 7:70-76
36. Head RA, Brown RM, Zolkipli Z, Shahdadpuri R, King MD, et al. Clinical and genetic spectrum of pyruvate dehydrogenase deficiency: dihydrolipoamide acetyltransferase (E2) deficiency. *Ann Neurol* 2005; 58(2): 234-241
37. McWilliam CA, Ridout C, Brown RM, McWilliam RC, Tolmie J, et al. Pyruvate dehydrogenase E2 deficiency: A potentially treatable cause of episodic dystonia. *Eur J Paediatr Neurol* 2010; 14(2010): 349-353
38. Pao GM, Zhu Q, Perez-Garcia CG, Chou SJ, Suh H, Gage FH, et al. Role of BRCA1 in brain development. *PNAS* 2014; 111(13): E1240-E1248

39. Suberbielle E, Djukic B, Evans M, Kim D, Taneja P, et al. DNA repair factor BRCA1 depletion occurs in Alzheimer brains and impairs cognitive function in mice. *Nat. Commun* 2015; 6:8897
40. Edvardson S, Shaag A, Zenvirt S, Erlich Y, Hannon GJ, et al. Joubert Syndrome 2(JBTS2) in Ashkenazi Jews is associated with a TMEM216 mutation. *Am. J. Hum. Gene* 2010; 86(1): 93-97
41. Hatten ME. The role of migration in central nervous system neuronal development. *Curr Opin Neurobiol* 1993; 3(1): 38-44
42. Tau GZ, Peterson BS. Normal Development of Brain Circuits. *Neuropsychopharmacology* 2010; 35(1): 147-168
43. Dekaban AS. Changes in brain weights during the span of human life: relation of brain weights to body heights and body weights. *Ann Neurol* 1978; 4: 345–356
44. Singh S, Mishra A, Mishra SK, Shukla S. ALCAR promote adult hippocampal neurogenesis by regulating cell-survival and cell death-related signals in rat model of Parkinson's disease like-phenotypes. *Neurochem Int* 2017; 108: 388-396
45. Lladó A, Gaig C, Molinuevo JL. Genética de las enfermedades neurodegenerativas más prevalentes. *Med Clin (Barc)* 2006; 126(17): 662-670
46. Rosenberg MS, Subramanian S, Kumar S. Patterns of Transitional Mutation Biases Within and Among Mammalian Genomes. *Molecular Biology and Evolution* 2003; 20(6): 988–993
47. Topal MD. Complementary base pairing and the origin of substitution mutations. *Nature* 1976; 263: 285-289
48. Prendergast JGD, Campbell H, Gilbert N, Dunlop MG, Bickmore WA, et al. Chromatin structure and evolution in the human genome. *BMC Evolutionary Biology* 2007; 7:72
49. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. *Am. J. Hum. Genet* 2007; 80(4): 727-739

50. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Rev. Genet* 2011; 12: 745–755
51. Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Genet* 2012; 13: 565-575