



DIRECTORA

Purificación Ruiz Flaño

CONSEJO DE REDACCIÓN

Luis Español González
Rubén Esteban Pérez
Rafael Francia Verde
Juana Hernández Hernández
Luis Miguel Medrano Moreno
Patricia Pérez-Matute
Enrique Requeta Loza
Rafael Tomás Las Heras

CONSEJO CIENTÍFICO

José Antonio Arizaleta Urarte
(Instituto de Estudios Riojanos)
José Arnáez Vadillo
(Universidad de La Rioja)
Susana Caro Calatayud
(Instituto de Estudios Riojanos)
Eduardo Fernández Garbayo
(Universidad de La Rioja)
Rosario García Gómez
(Universidad de La Rioja)
José M^a García Ruiz
(Instituto Pirenaico de Ecología)
Javier Guallar Otazua
(Universidad de La Rioja)
Teodoro Lasanta Martínez
(Instituto Pirenaico de Ecología)
Joaquín Lasierra Cirujeda
(Hospital San Pedro, Logroño)
Luis Lopo Carramiñana
(Dirección General de Medio Natural del Gobierno de La Rioja)
Fernando Martínez de Toda
(Universidad de La Rioja)
Alfredo Martínez Ramírez
(Centro de Investigación Biomédica de La Rioja -CIBIR-)
Juan Pablo Martínez Rica
(Instituto Pirenaico de Ecología-CSIC)
José Luis Nieto Amado
(Universidad de Zaragoza)
José Luis Peña Monné
(Universidad de Zaragoza)
Félix Pérez-Lorente
(Universidad de La Rioja)
Diego Troya Corcuera
(Instituto Politécnico y Universidad Estatal de Virginia, Estados Unidos)
Eduardo Viladés Juan
(Hospital San Pedro, Logroño)
Carlos Zaldívar Ezquerro
(Dirección General de Medio Natural del Gobierno de La Rioja)

DIRECCIÓN Y ADMINISTRACIÓN

Instituto de Estudios Riojanos
C/ Portales, 2
26071 Logroño
publicaciones.ier@larioja.org

Suscripción anual España (1 número y monográfico): 15 €

Suscripción anual extranjero (1 número y monográfico): 20 €

Número suelto: 9 €

Número monográfico: 9 €

INSTITUTO DE ESTUDIOS RIOJANOS

ZUBÍA

REVISTA DE CIENCIAS

Núm. 31

ier

Gobierno de La Rioja
Instituto de Estudios Riojanos
LOGROÑO
2013

Zubía –N. 3 (1985)– . –Logroño : Instituto de Estudios Riojanos, 1985-v.; il.; 24 cm. Anual
D.L. Lo 56-1986
Es suplemento de esta publicación : Zubía. Monográfico, ISSN 0213-4306
Es continuación de : Berceo. Ciencias
ISSN 0213-4306 = Zubía
5/6

Reservados todos los derechos. Ni la totalidad ni parte de esta publicación pueden reproducirse, registrarse ni transmitirse, por un sistema de recuperación de información, en ninguna forma ni por ningún medio, sea electrónico, mecánico, fotoquímico, magnético o electroóptico, por fotocopia, grabación o cualquier otro, sin permiso previo por escrito de los titulares del copyright.

© Logroño 2013
Instituto de Estudios Riojanos
C/ Portales, 2
26001-Logroño, La Rioja (España)

© Diseño de cubierta e interior: ICE Comunicación

© Cubierta: Chorlito dorado europeo (*Pluvialis apricaria*), fotografía de Javier Robres.
Contracubierta: Huellas de dinosaurio en Enciso, fotografía de Félix Pérez-Lorente.

Producción gráfica: kbcreativos.com

ISSN 0213-4306
Depósito Legal LO-56-1986

Impreso en España - Printed in Spain

ÍNDICE

I. GÁMEZ CARMONA, J. SERRADILLA RODRÍGUEZ, C. M^a AGUILAR GÓMEZ, J. ROBRES CABEZÓN, Ó. GUTIÉRREZ JIMÉNEZ

Anuario ornitológico de La Rioja 2009-2012

Ornithological yearbook of La Rioja 2009-2012 7-198

J. M. ÁLVAREZ BERMEJO, O. ARGÁIZ GIL, I. DÍAZ-MARTÍNEZ, I. EGUILUZ GARCÍA, M. HERREROS MARTÍNEZ, J. LUCAS LÓPEZ, C. MARTÍNEZ GARCÍA, J. MARTÍNEZ PÉREZ, A. PÉREZ LLANOS, F. PÉREZ-LORENTE, P. VEGA GARCÍA

El yacimiento paleoicnológico de Guilera. Redescubrimiento, situación, definición y análisis de sus icnitas

The Guilera tracksite; rediscovery, location, and ichnology 199-210

F. PÉREZ-LORENTE

Las huellas de dinosaurio más modernas de Enciso (La Rioja, España). Yacimiento de Las Riscas

The latest dinosaur footprints of Enciso (La Rioja, España). Las Riscas site 211-222

M. ÍÑIGUEZ MARTÍNEZ, B. EZCURRA GARCÍA, Á. BREA-HERNANDO, J. CABELLO

Secuenciación masiva de DNA y aplicación práctica al diagnóstico de la hipercolesterolemia familiar

Next generation DNA sequencing and application to familial hypercholesterolemia

diagnosis 223-236

VARIA

237-247

I. CANTARERO GONZÁLEZ, M. CUERVO ZAPATEL

Jornadas gastronómicas de la verdura: semana verde en Alfaro, La Rioja

Gastronomic days of vegetables: Alfaro's green week, La Rioja 239-247

SECUENCIACIÓN MASIVA DE DNA Y APLICACIÓN PRÁCTICA AL DIAGNÓSTICO DE LA HIPERCOLESTEROLEMIA FAMILIAR

MARÍA ÍÑIGUEZ MARTÍNEZ^{1*}

BEGOÑA EZCURRA GARCÍA¹

ÁNGEL BREA-HERNANDO²

JUAN CABELLO¹

RESUMEN

En los últimos años se han desarrollado las tecnologías de secuenciación de DNA de alto rendimiento capaces de generar millones de secuencias de DNA en poco tiempo y con costes cada vez más reducidos. Estas técnicas han revolucionado el mundo de la genómica contribuyendo al descubrimiento de importantes hitos científicos como replantear el concepto de gen o caracterizar nuevos genes y genomas. Gracias a su capacidad han emergido nuevas aplicaciones todas ellas trasladables al diagnóstico molecular de enfermedades. Una de estas aplicaciones, la secuenciación del exoma, ha sido utilizada en este trabajo para identificar mutaciones implicadas en el desarrollo de hipercolesterolemia familiar (HF) en un paciente con sospecha de padecer de HF de etiología molecular desconocida.

Palabras clave: *Genómica, Ultrasecuenciación, exomas*

High-throughput DNA sequencing technologies have been developed in the last few years. They can generate millions of DNA sequences in a short time and with lower costs. These techniques have revolutionized the genomic area and have contributed to significant discoveries as redefining the concept of the gene or characterizing new genes and genomes. Thanks to their capabilities, new applications have risen, all of them applicable to molecular diagnosis of diseases. One of these applications, exome sequencing, has been used in this work to identify mutations involved in Familial Hipercholesterolemia (FH) in a patient affected by HF of unknown molecular etiology.

* E-mail: miniguez@riojasalud.es

1. Plataforma de Genómica, Centro de Investigación Biomédica de La Rioja (CIBIR), C/ Piqueras 98, 26006, Logroño, La Rioja, España.

2. Unidad de Lípidos, Servicio de Medicina Interna, Hospital San Pedro, C/ Piqueras 98, 26006, Logroño, La Rioja, España.

Keywords: *Genomics, Next-generation sequencing, Exome*

1. SECUENCIACIÓN MASIVA Y APLICACIONES

Determinar la secuencia del DNA es la forma más exhaustiva de obtener información acerca del genoma de cualquier organismo vivo. Con este objetivo, en 1975 Frederick Sanger desarrolló el método de secuenciación de DNA conocido como método de Sanger, el cual ha dominado la investigación en genómica durante los últimos 30 años y ha logrado hitos tan significativos como la secuenciación del genoma humano completo. Mediante esta tecnología se han identificado desordenes genéticos y se ha hecho posible la detección de mutaciones somáticas como método diagnóstico clínico (Taylor *et al*, 2011; Sosman *et al*, 2012).

En el año 2005 se comercializó la primera plataforma de pirosecuenciación masiva en paralelo (454 Genome Sequencer de Roche) y comenzó la nueva era del análisis genómico de alto rendimiento conocido como ultrasecuenciación, secuenciación masiva o secuenciación de nueva generación. Estas tecnologías han transformado el campo de la genética, han reducido sustancialmente el coste de secuenciar grandes regiones genómicas y han permitido realizar experimentos que hasta el momento no eran técnicamente abordables.

En la actualidad existen tres principales proveedores de plataformas de ultrasecuenciación en el mercado: Illumina con sus Genome Analyzer, HiSeq y Myseq, el sistema 454 de Roche y el sistema SOLiD de Life Technologies. Las tres tecnologías comparten características comunes que han sido ampliamente descritas en la bibliografía (Shendure J *et al.*, 2008; Voelkerding *et al.*, 2009). Todas realizan una secuenciación masiva en paralelo de millones de moléculas de DNA amplificadas clonalmente que se encuentran separadas espacialmente en una celda de flujo o en una placa picotituladora. Este diseño tecnológico es un cambio de modelo respecto al método de secuenciación de Sanger que está basado en la separación electroforética de productos con terminador marcado con colorante producidos en reacciones de secuenciación individuales y que es capaz de procesar 96 o 384 secuencias de DNA en paralelo. En la secuenciación masiva, la secuenciación se realiza mediante ciclos repetidos de extensiones de nucleótidos ejecutados por una polimerasa o por ciclos iterativos de ligación de oligonucleótidos. Como proceso masivo en paralelo la ultrasecuenciación genera desde cientos de megabases a cientos de gigabases de secuencias de nucleótidos en una única carrera de análisis. Ello hace que la secuenciación masiva necesite del apoyo indispensable de la bioinformática para la recolección, procesamiento y organización de estas grandes cantidades de datos y lo convierta en información útil.

Dada la escala y la eficiencia de secuenciación alcanzada en la actualidad por estas tecnologías, se ha realizado un progreso impredecible en diferentes áreas, desde el análisis de genomas hasta el estudio de cómo las proteínas interaccionan con los ácidos nucleicos. Ello ha hecho que los investigadores

estén utilizando el poder de las tecnologías de ultrasecuenciación para abordar un diverso rango de problemas biológicos que están en aumento.

Su diverso espectro de aplicaciones incluye:

- El análisis genómico de genomas enteros de cualquier organismo vivo, desde microorganismos hasta el ser humano, incluyendo el reino vegetal. Mediante secuenciación masiva se han caracterizado nuevos genomas de organismos nunca secuenciados por completo y se han caracterizado de nuevo genomas ya analizados con el objetivo de buscar variaciones de secuencia entre muestras e identificar nuevos genes asociados a enfermedades raras y comunes (Goh *et al*, 2012). Esta tecnología hace viable el estudio de la evolución de especies, la genética de poblaciones y la genómica comparativa. Asimismo han aparecido nuevas aplicaciones como el estudio de la diversidad microbiana en muestras clínicas y medioambientales a través del análisis metagenómico (Holmes *et al*, 2012; Hazen *et al*, 2012).

- Mediante la ultrasecuenciación del RNA es posible analizar la expresión génica y estudiar su regulación. El transcriptoma es el conjunto de RNAs mensajeros expresado en un organismo en una determinada circunstancia, como pueden ser un momento, tejido o célula concreto. Al contrario que el genoma caracterizado por su estabilidad, el transcriptoma cambia dinámicamente bajo el efecto de diversos factores. Gracias a la secuenciación masiva se ha avanzado en la descripción y cuantificación de transcriptomas. Se ha conseguido una caracterización más compleja de los transcritos de RNA mejorando el mapeo de sitios de inicio de la transcripción, análisis específicos de hebra, descubrimiento de fusiones entre genes, y detección de eventos de splicing alternativo (Wang *et al*, 2009; Ozsolak *et al*, 2011). Asimismo mediante la secuenciación masiva de RNA es posible identificar y cuantificar pequeños RNAs, moléculas muy importantes en la regulación de la expresión génica y cuya implicación en multitud de procesos, así como en el desarrollo de enfermedades, se está poniendo de manifiesto en los últimos tiempos (Anglicheau *et al*, 2010; Pritchard *et al*, 2012).

- El potencial de la ultrasecuenciación también permite estudiar alteraciones epigenéticas principalmente a través del estudio de cambios en la metilación del DNA o modificaciones de histonas mediante ultrasecuenciación (Sukuki *et al*, 2013). De esta manera se hace factible el estudio de la regulación de procesos tan importantes como la diferenciación celular y la aparición o desarrollo del cáncer.

- Mediante el proceso conocido como ChIPseq es posible identificar sitios de unión de proteínas al DNA y estudiar las interacciones de proteína-DNA o modificaciones de la cromatina (Furey, 2012).

Combinando los datos obtenidos mediante todos estos métodos de ultrasecuenciación (DNA-seq, RNA-seq, ChIP-seq) se logrará por tanto una mejor comprensión de toda la información contenida en un genoma. Con esta finalidad se están desarrollando grandes proyectos colaborativos a nivel mundial cuyos frutos quedarán a disposición de la comunidad científica y médica en bases de datos.

Uno de ellos es el proyecto ENCODE (Encyclopedia Of DNA Elements) que analiza de manera intensiva el genoma humano (<http://encodeproject.org/ENCODE/>). Hasta la fecha ha realizado grandes aportaciones como mostrar todos los transcritos primarios y maduros, así como la localización de las principales modificaciones de histonas, los sitios de unión a factores de transcripción, sitios de inicio de la transcripción, sitios hipersensibles a DNAsa, etc, todo ello unido a datos de expresión génica, de replicación y del número de copia de esas regiones. Gracias a este proyecto se ha cambiado la visión tradicional según la cual un gen da lugar a uno o varios transcritos alternativos que codifican una proteína o sus isoformas y se ha descubierto que los genes son más complejos de lo que hasta ahora se pensaba. Una misma región genómica puede codificar distintas proteínas y además dar lugar a otros transcritos en ambas cadenas. Esto ha llevado a replantear el concepto de gen que se definiría como la unión de secuencias genómicas que codifican un conjunto coherente de productos funcionales, potencialmente solapantes. Esta nueva definición probablemente conllevará el aumento del número total de genes del genoma humano (Maher B, 2012; ENCODE Project Consortium, 2012).

Comprender la relación entre variaciones en el DNA y enfermedad ha sido uno de los mayores objetivos de la investigación en genética. Sin embargo, la identificación de variaciones específicas en enfermedades complejas sigue siendo un reto. Una aproximación para resolver este desafío es catalogar la variación genética a lo largo del genoma e intentar asociar esas variantes a un fenotipo particular. Con este objetivo se ha creado el proyecto 1000 genomas, cuyo objetivo principal es catalogar la variación genética en la especie humana y relacionarla con el riesgo de incidencia de algunas enfermedades (1000 Genomes Project Consortium, 2010; 1000 Genomes Project Consortium, 2012)

La mejora en la detección de nuevas variantes por medio de la secuenciación masiva también puede aplicarse a la identificación de nuevas mutaciones somáticas en cáncer. Con esta idea se ha creado un grupo de investigación colaborativo denominado Consorcio Internacional del Genoma del Cáncer (<http://icgc.org/>). A él pertenecen 9 países con el objetivo común de estudiar los cambios genéticos, transcriptómicos y epigenéticos presentes en 50 tipos de cáncer con gran importancia clínica y social a nivel mundial. Con los datos obtenidos en estos estudios provenientes de cientos de muestras se espera conseguir un conocimiento más profundo de la patogénesis del cáncer que redunde en nuevas terapias y nuevos métodos diagnósticos (International Cancer Genome Consortium, 2010). En este proyecto, España es la encargada de la secuenciación genómica y epigenómica de la leucemia linfática crónica a través del Proyecto genoma Leucemia linfática crónica – España del cual ya se han publicado los primeros hallazgos (Andrew et al, 2013).

También se están realizando esfuerzos para analizar y caracterizar los microorganismos que colonizan el cuerpo humano y su potencial alteración en enfermedades crónicas (Weinstock, 2012). Para aportar luz sobre este

tema nació el Proyecto Microbioma Humano cuyo objetivo es caracterizar las comunidades microbianas de diferentes sitios del cuerpo humano (piel, boca, tracto vaginal, intestino, y cavidad nasal) para encontrar correlaciones entre cambios en el microbioma y la salud o enfermedad humana (<http://www.hmpdacc.org/>) (Aagaard *et al*, 2013).

Aunque quedan numerosos desafíos por resolver en cuanto a la aplicación de la secuenciación masiva al ámbito clínico, la evolución y el refinamiento de las tecnologías de ultrasecuenciación proporcionarán la base para desarrollar una nueva era de medicina personalizada basada en el análisis genómico personalizado como el que se muestra en el presente trabajo. Se trata de un caso de prueba de concepto de aplicación clínica de la secuenciación masiva del exoma realizado en la Plataforma de Genómica del CIBIR. Esta unidad creada a mediados de 2010 proporciona servicio de ultrasecuenciación a investigadores y clínicos de ámbito nacional. Está concebida como un servicio de apoyo a la investigación para usuarios internos y externos, y para desarrollar esta labor cuenta con un equipo de secuenciación masiva de Illumina con capacidad para secuenciar hasta 95 gigabases de DNA en cada carrera. Con este equipo de secuenciación masiva se puede abordar prácticamente cualquier proyecto dentro del ámbito de la genómica, destacando:

- Secuenciación *de novo* de genomas completos.
- Re-secuenciación en busca de variantes genéticas.
- Análisis estructurales de DNA.
- Perfiles de expresión génica de mRNAs y microRNAs a escala global.
- Estudios transcriptómicos
- Interacciones DNA-proteína y técnicas de inmunoprecipitación de cromatina ChIP.
- Estudios de metagenómica y epigenómica.

2. APLICACIÓN CLÍNICA DE LA SECUENCIACIÓN DE EXOMA PARA EL DIAGNÓSTICO DE HIPERCOLESTEROLEMIA FAMILIAR

La hipercolesterolemia familiar (HF) es una enfermedad hereditaria que se manifiesta desde el nacimiento. Se caracteriza por un aumento de colesterol en sangre y un elevado riesgo de desarrollar infarto de miocardio en edades tempranas (Sjouke *et al*, 2011). Es una enfermedad de alta prevalencia en la población general y en La Rioja se estima que hay unas 15.000 personas con un nivel de colesterol elevado por causas genéticas. Sin un tratamiento adecuado la esperanza de vida de esta población es de 20 a 30 años inferior a la de la población general y un 70% de las personas con HF no tratada sufrirán un infarto antes de los 60 años. Por ello es de vital importancia su diagnóstico precoz para proporcionar un tratamiento adecuado y temprano que evitaría muchas muertes frecuentes.

El colesterol es transportado en la sangre unido a proteínas formando unas partículas conocidas como lipoproteínas de baja densidad o LDL. Des-

pués, a través de receptores específicos que se localizan principalmente en el hígado, el LDL- colesterol es eliminado de la sangre. La hipercolesterolemia familiar es debida a alteraciones en la estructura y función de estos receptores de LDL de la membrana celular. La forma mas común de la enfermedad es producida por mutaciones en el gen que codifica el receptor de las LDL que hace que el receptor no se una de manera normal a las partículas LDL-colesterol, por lo que estas permanecen en circulación (Hobbs *et al*, 1992). La consecuencia de este trastorno es una reducción importante del número de receptores funcionales para las LDL a nivel hepático, lo que ocasiona un incremento considerable de colesterol LDL en sangre que favorece su depósito en las arterias y aumenta la probabilidad de padecer un infarto de miocardio (Goldstein *et al*, 2001). Se han descrito hasta 200 mutaciones diferentes en el gen que codifica para el receptor de LDL y en los últimos años también se han encontrado casos de HF por mutaciones en el gen para la apolipoproteína B que es la proteína encargada de unir el LDL-colesterol al receptor LDL (García-Alvarez *et al*, 2003).

Las manifestaciones clínicas, el pronóstico de la enfermedad y el riesgo cardiovascular dependen del número de receptores afectados. Cada persona tiene dos alelos que codifican estos receptores, uno heredado del padre y el otro de la madre, por lo tanto la enfermedad tiene dos formas de presentación según se afecten uno o los dos alelos del gen. La forma heterocigota, más leve, la padecerán sujetos que tienen alterada la mitad de su dotación de receptores LDL. La forma homocigota es rara y debido a la ausencia total de receptores LDL se detectan hallazgos clínicos desde los primeros años de vida (arterioesclerosis coronaria anterior a los 20 años) (Garg *et al*, 2007). Asimismo, el comienzo y la severidad de la enfermedad también varían considerablemente dependiendo del tipo de mutación y de qué gen esté afectado (varios estudios han mostrado diferencias significativas en las concentraciones de colesterol entre varios tipos de mutaciones así como en el riesgo cardiovascular). Por todo ello es importante diagnosticar en cada paciente qué tipo de alteración presenta para lo cual se hace indispensable el diagnóstico genético.

El diagnóstico clínico de la HF se basa en la existencia de elevados niveles de colesterol total y LDL-colesterol, antecedentes familiares de hipercolesterolemia, aparición de depósitos de colesterol en tejidos extravasculares, e historia personal y familiar de enfermedad coronaria prematura. Sin embargo hay una elevada variabilidad interindividual tanto clínica como bioquímica que dificulta el diagnóstico inequívoco de esta enfermedad. Aunque se han desarrollado varias herramientas diagnósticas entre las cuales se incluyen el sistema de puntuación del Programa MEDPED un buen número de veces no se logra un diagnóstico definitivo de HF hasta que no se confirma la mutación causante.

Actualmente se dispone de un test genético para el diagnóstico de las mutaciones anteriormente descritas denominado Lipochip® (Civeira *et al*, 2008). Sin embargo, pese al reconocimiento de estas alteraciones existe entre un 17 y un 33% de pacientes con expresión fenotípica y características clínicas y analíticas similares a la HF que dan negativo en el test del Lipochip®, es decir, no presentan alteraciones en las partes del genoma conocidas como causante de una HF (Garg *et al*, 2007).

En este estudio se ha utilizado la secuenciación masiva de exoma para identificar qué alteración genética puede estar provocando el desarrollo de HF en un paciente con expresión fenotípica de HF de etiología desconocida en el cual no ha sido identificado el gen responsable. La secuenciación masiva de exoma completo es una técnica que permite detectar en un mismo ensayo mutaciones puntuales en todas las regiones codificantes del genoma humano (Bamshad *et al*, 2011; Zhang *et al*, 2011). Se ha estimado que mutaciones en dichas regiones suponen el 85% de las mutaciones responsables de enfermedad. Por lo tanto con esta técnica es posible identificar nuevos genes asociados a la HF. La resolución de la secuenciación masiva es muy superior a la del Lipochip[®], ya que no se limita a las mutaciones conocidas sino que permite analizar los genes de manera global, y abarcando toda su secuencia independientemente de cual sea esta. Ello permite estudiar toda la variabilidad genética que pueda existir, y por ello se convierte en la tecnología más apropiada para el diagnóstico de enfermedades con heterogeneidad génica y para la identificación de nuevos genes asociados a enfermedades genéticas (Choi *et al*, 2009).

3. MATERIALES Y MÉTODOS

Sujetos de estudio

En este caso clínico se ha estudiado el exoma completo de un paciente con sospecha clínica de padecer HF y el de un familiar sano. El paciente fue diagnosticado de HF basándose en los criterios clínicos del Programa MED-PED. En concreto, se trataba de un individuo caucásico atendido en la Unidad de Lípidos del Hospital San Pedro de Logroño con expresión fenotípica y clínica de HF que había dado negativo en las pruebas genéticas disponibles para su detección.

Extracción de sangre y purificación del DNA de leucocitos

Tras la obtención del consentimiento informado de los dos sujetos del estudio se procedió a la extracción de muestras de sangre total. En ambos casos se realizó en ayunas entre las 8:00 y 9:00 de la mañana para evitar variabilidad circadiana y homogeneizar al máximo posible las condiciones entre muestras.

Inmediatamente después de recoger la sangre en tubos Vacutainer con EDTA se procedió al aislamiento de la fracción rica en leucocitos mediante centrifugación a 2500 rpm durante 10 minutos. Seguidamente se purificó el DNA mediante el kit de Qiagen QIAamp DNA Blood mini siguiendo las instrucciones del fabricante.

Secuenciación de exoma completo

Una vez aislado el DNA de ambos sujetos del estudio se procedió a la preparación de las genotecas o librerías de DNA de cada muestra siguiendo los protocolos de la Plataforma Illumina.

Seguidamente empleando técnicas de hibridación de DNA como el método híbrido de selección de secuencias diana se capturaron los exones de todos los genes del genoma humano. Para ello se utilizó el kit TruSeq Exome Enrichment también suministrado por Illumina que permite la captura del exoma completo de cada una de las muestras.

Posteriormente las librerías de exoma fueron validadas en el sistema de electroforesis automatizado Experion de BioRad y cuantificadas mediante nanodrop y PCR cuantitativa.

Por último las librerías de exoma fueron secuenciadas en el ultrasecuenciador Genome Analyzer IIX de Illumina del CIBIR mediante una carrera single-read de 150 ciclos.

Análisis bioinformático

Los datos crudos de la secuenciación fueron tratados siguiendo el proceso descrito a continuación: a) Transformación de los datos qseq a FASTAQ y demultiplexación; b) eliminación de las secuencias de adaptadores incluidas en la preparación de las genotecas; c) alineamiento de las secuencias con las secuencias del genoma humano GRCh37 versión 64 mediante el programa STAR; d) Búsqueda de mutaciones en el genome browser IGV del Broad Institute; e) Identificación de mutaciones descritas en las bases de datos Ensembl (<http://www.ensembl.org>) y dbSNP (<http://www.ncbi.nlm.nih.gov/SNP>); f) Análisis funcional incluyendo predicciones de consecuencias en la codificación de proteínas, Sorting Tolerant From Intolerant (SIFT) y PolyPhen.

4. RESULTADOS Y DISCUSIÓN

La profundidad media de lectura para la secuencia del exoma completo fue de 50x, con el 70% del exoma cubierto al menos 20x y el 55,8 % de la secuencia diana cubiertas 50x o más.

Aunque en el genoma del paciente se hallaron más de 3,5 millones de variaciones de secuencia, la mayoría de ellas pudieron ser excluidas de su consideración como mutaciones causantes de enfermedad por ser comunes a individuos sanos. Seguidamente se restringió la búsqueda inicial al estudio de las mutaciones localizadas en genes implicados en el metabolismo y transporte de lípidos y colesterol y cuya alteración ha sido descrita que produce desórdenes lipídicos mendelianos (Tabla 1).

A continuación se filtraron las mutaciones que no causaban cambio de aminoácido, denominadas mutaciones sinónimas, y que por tanto no afectarían a la función de la proteína codificada por esos genes. Asimismo también se descartaron las mutaciones intrónicas pues la mayor parte de los desordenes mendelianos son causados por sustituciones raras de un único nucleótido en regiones codificantes, reguladoras o en las uniones de splicing.

De este modo, en el paciente estudiado, se encontraron un total de 24 mutaciones que afectaban a 11 de los 19 genes candidatos. 17 eran mutaciones no sinónimas que producían cambio de aminoácido y 7 se localizaban en regiones reguladoras 5' UTR y 3' UTR (Tabla 2).

El cribado continuó con la exclusión de la lista de las mutaciones compartidas con el control familiar sano, así como de las mutaciones presentes en alta frecuencia en la población. Para este último paso se examinó el índice MAF (Global Minor Allele Frequency) que indica la frecuencia a la cual el alelo menos común ocurre en la población. Este índice da un valor menor de 0,01 para las variantes raras, un valor comprendido entre 0,01 y 0,05 para las variantes de frecuencia intermedia y un índice superior a 0,05 para las variantes comunes en la población.

Asimismo se evaluó también si la mutación hallada podía producir un efecto deletéreo en la función de la proteína codificada por el gen alterado. Para ello se examinaron dos indicadores distintos que predicen el efecto final de la mutación:

- SIFT: puntuaciones inferiores a 0,05 indican que la función de la proteína puede estar afectada.

- Polyphen: Valores cercanos a 1 es más probable que sean deletéreos.

Ateniéndonos al esquema de trabajo descrito anteriormente, de las 17 mutaciones no sinónimas encontradas en el paciente, solo 7 mutaciones heterocigotas no se hallaban en el familiar control. 2 de ellas eran mutaciones nuevas que afectaban a los genes APOB y AP2B1. Las otras 5 variantes descubiertas eran mutaciones ya descritas, de las cuales 2 eran de baja prevalencia en la población y además podían tener un efecto deletéreo en la función de la proteína. Estas variantes alteraban los genes LDLRAP1 y APOB. La variante rs41291058 que afectaba al gen LDLRAP1 ha sido ya descrita en otros estudios con pacientes afectados de HF. La variante rs1042031 que presentaba el exón 29 del gen APOB puede afectar a la unión del LDL al receptor del LDL generando una disminución en el catabolismo de las partículas de LDL y contribuyendo por tanto al aumento de los niveles de LDL plasmáticos.

De las 7 mutaciones encontradas en regiones reguladoras, se hallaron 3 variantes nuevas de las cuales solo una no estaba en el sujeto control sano. Esta variante afectaba al gen LDLRAD2 perteneciente a la familia del LDLR e implicado en la homeostasis del colesterol. El resto de variantes descubiertas habían sido ya descritas y todas se localizaron también en el familiar control sano.

5. RESUMEN Y CONCLUSIONES

En un paciente con sospecha de sufrir HF al cual no se le había encontrado defecto genético alguno asociado a su enfermedad se le han identificado mutaciones, algunas nunca caracterizadas, en genes cuya alteración ha sido descrita como implicada en el desarrollo de HF. En concreto, se han detectado mutaciones en los genes ApoB y LDLRAP1 que alteran la composición proteica de las proteínas que estos genes codifican. Ello podría conllevar un efecto deletéreo en su función que desemboque en hipercolesterolemia.

Además en este paciente se han localizado 3 mutaciones nuevas, susceptibles también de ser estudiadas en mayor profundidad en un grupo mayor de casos y controles, que en combinación con las anteriores podrían ser

también las causantes del fenotipo hipercolesterolémico del paciente. Estas variantes se han encontrado en los genes APOB, AP2B1 y LDLRAD2.

Ninguna de estas mutaciones había podido ser diagnosticada en este paciente por medio del LipoChip lo que confirma la gran utilidad de la secuenciación de exoma en el diagnóstico de la hipercolestoremia en particular, y de otros desórdenes genéticos en general.

AGRADECIMIENTOS

Al Instituto de Estudios Riojanos (IER) por parte de la financiación recibida en el marco de los proyectos de Estudios Riojanos.

BIBLIOGRAFÍA

- 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. *Nature*. 467(7319):1061-73.
- 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*. 491(7422):56-65.
- Aagaard, K., Petrosino, J., Keitel, W., Watson, M., Katancik, J., Garcia, N., Patel, S., Cutting, M., Madden, T., Hamilton, H., Harris, E., Gevers, D., Simone, G., McInnes, P., Versalovic, J. (2013). The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB J*. 27(3):1012-22.
- Agirre, E., Eyra, E. (2011). Databases and resources for human small non-coding RNAs. *Hum Genomics*. 5(3):192-9.
- Anglicheau, D., Muthukumar, T., Suthanthiran, M. (2010). MicroRNAs: small RNAs with big effects. *Transplantation*. 27;90(2):105-12.
- Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., Shendure, J. (2001). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 12(11):745-55.
- Chiou, K.R., ChRNAg, M.J. (2012). Common mutations of familial hypercholesterolemia patients in Taiwan: Characteristics and implications of migrations from southeast China. *Gene*. 498(1):100-6.
- Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakka-lo, lu, A., Ozen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., Lifton, R.P. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*. 106(45):19096-101.
- Civeira, F., Ros, E., Jarauta, E., Plana, N., Zambon, D., Puzo, J., Martinez de Esteban, J.P., Ferrando, J., Zabala, S., Almagro, F., Gimeno, J.A., Masana, L., Pocovi, M. (2008). Comparison of genetic versus clinical diagnosis in familial hypercholesterolemia. *Am J Cardiol*. 102(9):1187-93.

- ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*. 489(7414), 57–74.
- Etxebarria, A., Palacios, L., Stef, M., Tejedor, D., Uribe, K.B., Oleaga, A., Irigoyen, L., Torres, B., Ostolaza, H., Martin, C. (2012). Functional characterization of splicing and ligand-binding domain variants in the LDL receptor. *Hum Mutat*.33(1):232-43.
- Furey, T.S. (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet*. 13(12):840-52.
- García-Alvarez, I., Castillo, S., Mozas, P., Tejedor, D., Reyes, G., Artieda, M., Cenarro, A., Alonso, R., Mata, P., Pocovi, M., Civeira, F. (2003).Differences in clinical presentation between subjects with a phenotype of familial hypercholesterolemia determined by defects in the LDL-receptor and defects in Apo B-100. *Rev Esp Cardiol*. 2003. 56(8):769-74.
- Garg, A., Simha, V. (2007).Update on dyslipidemia. *J Clin Endocrinol Metab*. 92(5):1581-9.
- Goh, G., Choi, M. (2012). Application of Whole Exome Sequencing to Identify Disease-Causing Variants in Inherited Human Diseases. *Genomics & informatics*.10(4):214-219.
- Goldstein, J.L., Hobbs, H.H., Brown, M.S. (2001). Familial Hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. The Metabolic and Molecular Basis of Inherited Disease. 8th ed. New York,NY. McGraw-Hill;;2863-2913.
- Hazen, T.C., Rocha, A.M., Techtmann SM. (2013). Advances in monitoring environmental microbes. *Curr Opin Biotechnol*. 24(3):526-33.
- Hobbs, H.H., Brown, M.S., Goldstein, J.L. (1992). Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. *Hum Mutat*.;1(6):445-66.
- Holmes, E., Li, J.V., Marchesi, J.R., Nicholson, J.K. (2012). Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab*. 7;16(5):559-64.
- International Cancer Genome Consortium. (2010). International network of cancer genome projects. *Nature*. 464(7291):993-8.
- Maher, B. (2012). ENCODE: The human encyclopaedia. *Nature*. 489(7414): 46-8.
- Ozsolak, F., Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*. 12(2):87-98.
- Pritchard, C.C., Cheng, H.H., Tewari, M. (2012). MicroRNA profiling: approaches and considerations. *Nat Rev Genet*. 13(5):358-69.
- Ramsay, A.J., Quesada, V., Foronda, M., Conde, L., Martínez-Trillos, A., Villamor, N., Rodríguez, D., Kwarciak, A., Garabaya, C., Gallardo, M., López-Guerra, M., López-Guillermo, A., Puente, X.S., Blasco, M.A., Campo, E., López-Otín, C. (2013). POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nature Genetics* 45, 526–530.
- Shendure J, Ji H. (2008). Next-generation DNA sequencing. *Nat Biotechnol*.;26(10):1135-1145.
- Sjouke, B., Kusters, D.M., Kastelein, J.J., Hovingh, G.K. (2011). Familial hypercholesterolemia: present and future management. *Curr Cardiol Rep*. 13(6):527-36.

- Sosman, J.A., Kim, K.B., Schuchter, L., Gonzalez, R., Pavlick, A.C., Weber, J.S., McArthur, G.A., Hutson, T.E., Moschos, S.J., Flaherty, K.T., Hersey, P., Kefford, R., Lawrence, D., Puzanov, I., Lewis, K.D., Amaravadi, R.K., Chmielowski, B., Lawrence, H.J., Shyr, Y., Ye, F., Li, J., Nolop, K.B., Lee, R.J., Joe, A.K., Ribas, A. (2012). Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N Engl J Med.* 366:707–714.
- Suzuki, M., Grealley, J.M. (2013). Genome-wide DNA methylation analysis using massively parallel sequencing technologies. *Semin Hematol.* 50(1):70-7.
- Taylor, B.S., Ladanyi, M. (2011) Clinical cancer genomics: how soon is now? *J Pathol.* (2011). 223:318–326.
- Voelkerding, K.V., Dames, S.A., Durtschi, J.D. (2009). Next-generation sequencing: from basic research to diagnostics. *Clin Chem.* 55(4):641-658.
- Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 10(1):57-63.
- Weinstock, G.M. (2012). Genomic approaches to studying the human microbiota. *Nature.* 489(7415):250-6.
- Zhang, J., Chiodini, R., Badr, A., Zhang, G. (2011). The impact of next-generation sequencing on genomics. *J Genet Genomics.* 38(3): 95–109.

ANEXO

Gen	Locus génico	Tipo de desorden lipídico
ABCA1	9q31.1	Enfermedad de Tangier: Bajos niveles de HDL
ABCG5	2p21	Sitosterolemia: Altos niveles de LDL
ABCG8	2p21	Sitosterolemia: Altos niveles de LDL
APOA1	11q23-q24	Deficiencia en ApoA-I : Bajos niveles de LDL
APOA5	11q23	Deficiencia en ApoA-V: Altos niveles de VLDL y quilomicrones
APOB	2p24	Hipobetalipoproteinemia familiar: Bajos niveles de LDL
APOC2	19q13	Deficiencia familiar ApoC-II: Altos niveles de quilomicrones
APOE	19q13	Disbetalipoproteinemia familiar: Altos niveles de VLDL y quilomicrones.
CETP	16q13	Deficiencia en proteína CETP: Altos niveles de HDL
LCAT	16q22	Deficiencia en Lecitin-colesterol acyltransferasa: Bajos niveles de HDL
LDLR	19p13	Hipercolesterolemia familiar: Altos niveles de LDL
LDLRAP1	1p36-p35	Hipercolesterolemia recesiva autosómica: Altos niveles de LDL
LIPC	15q22	Deficiencia en lipasa hepática familiar: Altos niveles de VLDL
LPL	8p21	Deficiencia en Lipoproteinlipasa: Altos niveles de quilomicrones
MTTP	4q24	Abetalipoproteinemia: Bajos niveles de LDL
PSCK9	1p32	Hipercolesterolemia autosómica dominante: Altos niveles de LDL
LRP2	2q31	Miembro de la familia del receptor del LDL. Polimorfismos en LRP2 se asocian a niveles de colesterol plasmáticos.
LDLRAD2	1p36	Low Density Lipoprotein Receptor Class A Domain Containing 2 perteneciente a la familia del LDLR.

Tabla1. Desordenes lipídicos de herencia mendeliana

GEN	CAMBIO DE NUCLEÓTIDO Y AMINOÁCIDO	EXÓN	EFFECTO FUNCIONAL POSICIÓN EN PROTEÍNA	CONDICIÓN	DESCRIPCIÓN	MAF	SIFT	POLYPHEN	CONTROL SANO
ABCA1	C/T	3'UTR	chr9:107,544,700	Homocigoto	Mutación conocida rs363717	0,151	Variante en región reguladora		Homocigoto
	AAA [K] / AGA[R]	35	1587	Homocigoto	Mutación conocida rs2230808	0,41	0,28 Tolerada	0,0022 Benigna	Homocigoto
	AGG[R] / AAG[K]	7	219	Homocigoto	Mutación conocida rs2230806	0,42	0,92 Tolerada	0 Benigna	Homocigoto
ABCG5	NO								
ABCG8	NO								
APOA1	NO								
APOA5	G/A	3'UTR	chr11:116,660,686	Homocigoto	Mutación conocida rs2266788	0,114	Variante en región reguladora		Homocigoto
	C/T	5'UTR	chr11:116,662,579	Homocigoto	Mutación conocida rs651821	0,159	Variante en región reguladora		Homocigoto
APOB	AGT[S] / ATT[N]	29	4338	Homocigoto	Mutación conocida rs1042034	0,338	1 Tolerada	0 Benigna	Homocigoto
	GAA[E] / AAA[K]	29	4181	Heterocigoto	Mutación conocida rs1042031	0,152	071 Tolerada	0,312 Benigna	No
	ATT[I] / GTT[V]	26	2313	Homocigoto	Mutación conocida rs584542	0,028	0,69 Tolerada	0 Benigna	Homocigoto
	TAT[Y] / TGT[C]	26	1422	Homocigoto	Mutación conocida rs568413	0	0,27 Tolerada	0 Benigna	Homocigoto
	GTC[V] / ATC[I]	15	730	Heterocigoto	Mutación conocida rs12691202	0,022	0,28 Tolerada	0,987 Probablemente dañina	No
	GCT[A] / TCT[S]	14	618	Heterocigoto	nueva				No
	ACC[T] / ATC[I]	4	98	Heterocigoto	Mutación conocida rs1367117	0,206	0,02 Deleterea	0,825 Posiblemente dañina	Heterocigoto
APOC2	NO								
APOE	NO								
CETP	NO								
LCAT	NO								
LDLRAP1	CGG[R]/TGG[W]	7	238	Heterocigoto	Mutación conocida rs41291058	0,02	0,02 Deleterea	0,846 Posiblemente dañina	No
LIPC	AAT[N] / AGT[S]	5	215	Heterocigoto	Mutación conocida rs6083	0,412	0,14 Tolerada	0,013 Benigna	No
	TTC[F] / TTA[L]	7	356	Homocigoto	Mutación conocida rs3829462	0,07	0,36 Tolerada	0,155 Benigna	Homocigoto
LPL	A/T	5'UTR	chr8:19,796,870	Heterocigoto	Nueva		Variante en región reguladora		No
MTTP	GAG[E] / GAC[D]	4	98	Heterocigoto	Mutación conocida rs2306986	0,126	0,41 Tolerada	0,009 Benigna	Heterocigoto
PCSK9	C/T	3'UTR	chr1:55,529,828	Homocigoto	Mutación conocida rs662145	0,272	Variante en región reguladora		Homocigoto
CEPT1	C/A	5'UTR	chr1:111,690,270	Heterocigoto	Nueva		Variante en región reguladora		Heterocigoto
AP2B1	ATT[I] / TTT[F]	13	525	Heterocigoto	Nueva				No

Tabla 2. Mutaciones patogénicas encontradas mediante secuenciación de exoma en el paciente estudiado.



ZUBÍA

31

Gobierno de La Rioja
www.larioja.org



**Instituto
de Estudios
Riojanos**