

# SEBBM DIVULGACIÓN

## LA CIENCIA AL ALCANCE DE LA MANO



### La secuenciación de genomas personales... ¿secuenciar, para qué?

Ricardo Ramos

Unidad de Genómica, Fundación Parque Científico de Madrid

#### Biografía

*Ricardo Ramos Ruiz es doctor en CC. Químicas por la Universidad Autónoma de Madrid desde el año 1991. Su carrera investigadora desarrollada en diversos laboratorios del CSIC, la Universidad Autónoma de Madrid y la Universidad de Utrecht en Holanda se ha centrado fundamentalmente en el proceso de expresión génica. En el año 2004 se incorporó a la Unidad de Genómica del Parque Científico de Madrid para potenciar el Servicio de Apoyo a la Investigación "Antonia Martín Gallardo", de la cual es ahora Responsable Técnico. Su laboratorio ofrece soporte técnico y científico en PCR a tiempo real, secuenciación convencional y en las nuevas tecnologías de secuenciación masiva, gracias a una estrecha colaboración con distintos centros de investigación del CSIC de Madrid. Dirige personalmente un grupo de 5 científicos y colabora con el trabajo de más de un centenar de grupos de investigación distribuidos por toda España.*

<http://www.sebbm.es/>

HEMEROTECA:

[http://www.sebbm.es/ES/divulgacion-ciencia-para-todos\\_10/la-ciencia-al-alcance-de-la-mano-articulos-de-divulgacion\\_29](http://www.sebbm.es/ES/divulgacion-ciencia-para-todos_10/la-ciencia-al-alcance-de-la-mano-articulos-de-divulgacion_29)

SEBBM  
SEBBM  
Sociedad Española  
de Bioquímica y  
Biología Molecular

#### Resumen

**En los últimos años se han producido unos enormes avances en las técnicas de investigación en Genómica, que permiten la secuenciación de genomas individuales. La utilidad de los genomas personales es que descubren variantes génicas asociadas a la susceptibilidad a distintas enfermedades o a la respuesta frente a los estímulos.**

#### Summary

**In the last years an amazing progress in genomics research has allowed the development of techniques which allow the sequencing of complete personal genomes. Personal genome sequencing is a powerful tool to analyze individual diversity which may explain the different risk to acquire diseases or respond to environmental stimuli.**

Con la llegada del siglo XXI la Genómica observa una auténtica revolución. Su causa: la primera descripción completa de un genoma de mamífero, nada menos que el Genoma Humano. Se recogía el trabajo de un ambicioso consorcio internacional diseñado al efecto, pero no por esperada su publicación fue menos espectacular. Podemos considerar que este hito sentó las bases de la Genómica actual, y supuso un vuelco principalmente a tres niveles.

En primer lugar (aun con las críticas que pueda plantear) introdujo a la empresa privada a un nivel semejante a los consorcios públicos

de rango internacional. Lo más destacable es que se empieza a apreciar que la información sobre nuestro genoma es importante más allá del ámbito académico: se pueden realizar pruebas clínicas y forenses basadas en el genoma, buscar biomarcadores, identificar agentes terapéuticos...; es decir, la información sobre el genoma es "diagnosticable".

El segundo nivel es el científico. Ha permitido conocer cómo se organizan los genes, cómo es ese "DNA basura" que no codifica proteínas (hoy sabemos que esconde funciones esenciales), y ha permitido describir cómo está estructurado nuestro genoma hasta el último detalle.

Finalmente, el tercer avance es el tecnológico, tanto por las herramientas que se crearon para el desarrollo del proyecto, como por la llegada de nuevas técnicas que utilizan información extraída directamente del genoma.

En apenas cinco años se desarrollaron secuenciadores automáticos capaces de producir decenas de miles de bases diarias, equipos de PCR a tiempo real como un sistema óptimo de la medida de la expresión génica y *microarrays* de alta densidad, en los que se puede interrogar un genoma a escala global. Estas últimas tecnologías utilizan como reactivos específicos secuencias cortas de bases (oligonucleótidos) que reconocen las regiones de interés localizadas a lo largo del genoma (la secuencia que codifica una proteína, una mutación patogénica, etc.). Basta fijar un gen objetivo para localizar en un catálogo

el reactivo que lo identifica o mide su expresión. Como se conoce el genoma de principio a fin, no hay límites y el catálogo es completo. La escena estaba servida para llegar a la segunda generación de la Genómica: la secuenciación masiva, en la cual se automatiza el proceso de secuenciación, aplicándose en paralelo y a una escala enormemente superior. El proceso supone fragmentar un genoma completo y añadir en cada extremo unas señales biológicas (de nuevo, unos oligonucleótidos) llamados adaptadores. Estos adaptadores sirven para capturar, enriquecer, purificar y finalmente secuenciar cada uno de los fragmentos iniciales. Lo más espectacular es que todo el proceso se puede realizar en miles de millones de moléculas a la vez. Comparado con los miles de bases de un secuenciador automático "convencional", la productividad alcanza los miles de millones (Gigabases, Gb), de la magnitud de nuestro genoma completo. Gracias al desarrollo conjunto de técnicas informáticas adaptadas, podemos transformar esa ingente cantidad de nucleótidos en una secuencia inteligible y ordenada, lista para ser analizada.

Ahora bien: si ya conocemos el genoma y cómo son nuestros genes: ¿qué sentido tiene volver a secuenciarlo? La respuesta es que no existe tal genoma humano único, sino que cada individuo tiene sus propias particularidades y cambios frente a lo que podríamos considerar una secuencia "consenso". Se calcula que todos los seres humanos compartimos más de un 99,9 % de nuestro genoma, pero el pequeño porcentaje de diferencia se traduce en que presentamos varios millones

de posiciones no coincidentes respecto a cualquier otro ser humano. Son esas variantes las que dan sentido al estudio del genoma de cada individuo. Se han logrado asociar ya casi 8.000 variantes a un riesgo añadido de desarrollar una patología en comparación a la variante "saludable". La forma en que actúan dichas variantes puede ser muy diversa: inducir modificaciones en proteínas efectoras, cambiar la expresión de ciertos genes o afectar a otros mecanismos aún desconocidos. También parte del efecto de los fármacos (su efectividad y sus efectos secundarios) está definida por variantes en ciertos genes. Incluso el beneficio que obtenemos de determinados alimentos se puede llegar a trazar a variantes en el genoma. La secuenciación de genomas individuales es la base de la medicina personalizada: definir riesgos, predecir respuestas, escoger tratamientos o mejorar tasas de éxito se empiezan a ver en el horizonte.

Sin embargo, no podemos considerar que dispongamos de una herramienta definitiva. La genética por sí sola no explica toda nuestra biología. La actividad génica se regula también por mecanismos no hereditarios "epigenéticos". La acción de un gen no puede considerarse de forma aislada, sino en base a la proteína que codifica, que actúa en asociación y dentro de una compleja red de interacciones con otras proteínas en un compartimento celular definido. Genómica, Transcriptómica, Proteómica, Metabolómica... son conceptos inter-relacionados que debemos integrar y asociar para comprender cómo los genes regulan

nuestra vida, y empezar a comprender cómo aprovecharnos de ellos para mejorar nuestra salud y nuestro estilo de vida.

## Referencias

1. Información sobre el proyecto genoma Humano en: [http://www.ornl.gov/sci/techresources/Human\\_Genome/project/about.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml)
2. Información sobre proyectos de análisis de la variabilidad del genoma humano: Proyecto 1000 genomas: <http://www.1000genomes.org/> Proyecto HapMap: <http://hapmap.ncbi.nlm.nih.gov/> Proyecto de secuenciación en cáncer: <http://cancergenome.nih.gov/>
3. Bases de datos con información completa del genoma humano y de otras especies: NCBI: <http://www.ncbi.nlm.nih.gov/> EMBL: <http://www.ensembl.org/index.html>

**Figura. La evolución de la tecnología de secuenciación. A) Secuenciación clásica "Sanger". Se basa en el uso de un oligonucleótido radiactivo que se extiende en presencia de terminadores. Las moléculas se resuelven en un gel de poliacrilamida de cuatro carriles, utilizando uno por cada base. El rendimiento alcanzaba > 1.000 bases por carrera. B) Secuenciación automática. Utiliza terminadores fluorescentes y capilares en lugar de geles. La señal de cada base se detecta por técnicas ópticas. El rendimiento puede alcanzar 50.000 bases por carrera. C) y D) Secuenciación masiva o Ultra-secuenciación. Se realiza una secuenciación automática en paralelo por técnicas de piro-secuenciación (C) o secuenciación por síntesis (D). El rendimiento puede alcanzar 500 millones de bases y más de 50.000 millones de bases por carrera, respectivamente.**

