

# Nuevos métodos de diagnóstico molecular

TRANSCRIPTÓMICA (MARNY MIR) pág. 160 EPIGENÓMICA pág. 165  
LA PROTEÓMICA COMO HERRAMIENTA DE DIAGNÓSTICO Y PRONÓSTICO: APLICACIÓN EN EL HEPATOCARCINOMA pág. 172

## Puntos clave

Los polimorfismos de nucleótido simple fueron elegidos como la nueva generación de marcadores genéticos, siendo imprescindibles en las investigaciones sobre la base genética de las enfermedades complejas.

La aparición de plataformas de genotipado de polimorfismos de nucleótido simple de alto rendimiento, y además coste-efectivas, supuso una revolución en la investigación genómica, haciendo posibles los estudios de asociación de genoma completo.

Los primeros resultados de los estudios de asociación de genoma completo abren nuevos frentes en la investigación de enfermedades humanas: identificación de la variante causal y estudio de otras formas de variabilidad genética, como los *copy number variants*.

Las nuevas técnicas de secuenciación permiten la lectura de regiones más largas del genoma en menor tiempo, por lo que tienen múltiples aplicaciones dentro del ámbito de la genómica.

## Genómica

CLARA RUIZ-PONTE Y CERES FERNÁNDEZ-ROZADILLA

Fundación Pública Galega de Medicina Xenómica. SERGAS. Grupo de Medicina Xenómica-USC. CIBERER.

Santiago de Compostela.

A Coruña. España.

En los últimos años se ha producido una revolución en las tecnologías de genotipado que ha hecho factible las aproximaciones a gran escala o a nivel de genoma completo en los estudios de asociación genética. Esto ha supuesto un gran avance en el conocimiento de la base genética de enfermedades complejas. Las tecnologías de genotipado de alto rendimiento han utilizado como marcadores genéticos los polimorfismos de nucleótido simple (SNP, de *single nucleotide polymorphism*) debido a su frecuencia, estabilidad, simplicidad y a su distribución en el genoma. Recientemente, están apareciendo nuevas tecnologías de secuenciación coste-efectivas, conocidas como secuenciación de nueva generación (NGS, de *next-generation sequencing*), que tienen múltiples aplicaciones, como el estudio de genomas completos.

## Polimorfismos de nucleótido simple y enfermedad compleja

Los SNP son las variantes genéticas más comunes en el genoma, y están accesibles a través de bases de datos públicas<sup>1-3</sup>. Consisten en el cambio de una sola base nucleotídica en el contexto de una secuencia genética. Para que una sustitución sea clasificada como SNP, la frecuencia del alelo menor (MAF, de *minor allele frequency*) debe ser > 1% en la población.

Se caracterizan por ser:

- Mayoritariamente bialélicos, y por tanto fáciles de analizar mediante sistemas automatizados.
- Muy estables, con tasas de mutación bajas.
- Muy frecuentes en el genoma (1 SNP/Kb), lo que posibilita la creación de mapas de alta densidad de SNP.

Por estas propiedades fueron propuestos como la nueva generación de marcadores genéticos,

siendo imprescindibles en las investigaciones sobre la base genética de las enfermedades complejas, entre ellas el cáncer.

El mecanismo etiopatogénico de las enfermedades complejas probablemente consiste en un trastorno poligénico en el que la combinación de alelos de diferentes genes, con penetrancia individual baja y gran prevalencia en la población, definirían la susceptibilidad individual a padecer la enfermedad, mientras que los factores ambientales modularían dicha susceptibilidad para determinar qué individuos desarrollarán finalmente la patología<sup>4</sup>. La identificación de estas variantes comunes se realiza mediante estudios de asociación caso-control, en los que se compara la frecuencia de una variante genética entre dos poblaciones (individuos enfermos frente a individuos sanos) con el objetivo de determinar la implicación de la misma en la susceptibilidad a la enfermedad<sup>5</sup>.

Tanto el Proyecto Genoma Humano como el HapMap aportaron datos muy valiosos sobre la frecuencia y herencia en bloques de los SNP<sup>1,6,7</sup>. Se observó que analizando un grupo bien seleccionado de SNP, denominados tagSNP, se podría capturar hasta el 90% de la variabilidad genética en población europea<sup>8</sup>. Esta información, junto con la aparición de nuevas tecnologías de genotipado basadas en *arrays*, en los que se puede analizar de forma simultánea hasta  $2 \times 10^6$  marcadores por muestra, hizo realidad los estudios de asociación de genoma completo<sup>9</sup> (GWAS, de *genome-wide association studies*) (fig. 1). Estos estudios representan un nuevo método para la identificación de alelos de bajo-moderado riesgo asociados a enfermedades comunes. A diferencia de los estudios de asociación clásicos en los que se plantea la hipótesis del gen candidato seleccionado según criterios biológicos, patológicos o moleculares, en el GWAS las hipótesis se generan a partir de los resultados<sup>10,11</sup>.

## Lectura rápida



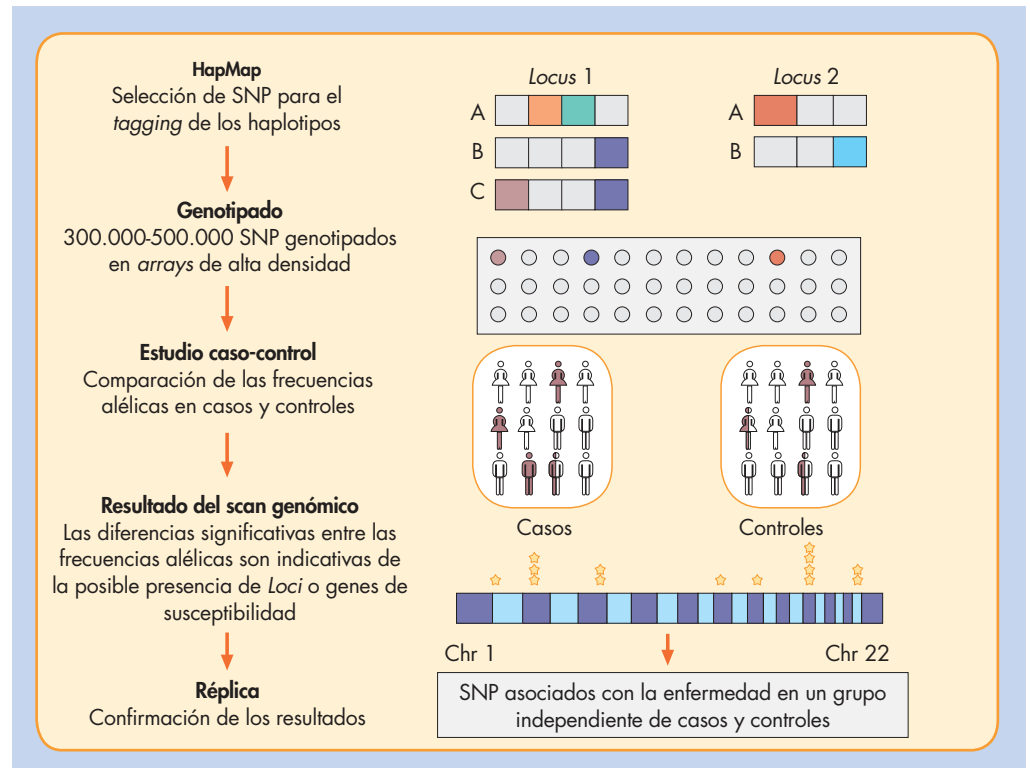
En los últimos años se ha producido una revolución en las tecnologías de genotipado, que ha hecho factibles las aproximaciones a gran escala o a nivel de genoma completo en los estudios de asociación genética. Esto ha supuesto un gran avance en el conocimiento de la base genética de enfermedades complejas.

Se postula que las enfermedades complejas siguen un modelo de herencia poligénica en el que la combinación de variantes genéticas comunes en la población, y con penetrancia individual baja, definirían la susceptibilidad individual a padecer la enfermedad.

Las variantes genéticas más comunes son los polimorfismos de nucleótido simple (SNP, de *single nucleotide polymorphism*), que han sido utilizados como marcadores por las distintas tecnologías de genotipado. Éstas han evolucionado hacia la caracterización simultánea de gran número de SNP/muestra, mediante lo que se denominan plataformas de alto rendimiento o *high-throughput*.

Los SNP fueron así seleccionados como la nueva generación de marcadores genéticos, siendo imprescindibles en la investigación genómica de las enfermedades complejas.

Los estudios de asociación genética caso-control en los que se compara la frecuencia de una variante genética entre dos poblaciones son muy útiles para identificar SNP de susceptibilidad a enfermedades complejas.



**Figura 1.** Estrategia GWAS (genome-wide association studies). Se estudia la distribución de los genotipos comparándolos en una población de individuos sanos y otra de afectados por la enfermedad. Adaptada de Mathew<sup>6</sup>. *Nat Rev Genet.* 2008.

## Plataformas de genotipado masivo de polimorfismos de nucleótido simple

La tecnología de genotipado ha evolucionado hacia la caracterización simultánea de gran número de SNP, mediante lo que se denominan plataformas de alto rendimiento o *high-throughput*<sup>12,13</sup>. Estas plataformas se han desarrollado a partir de la puesta a punto de nuevos protocolos de reacción química (la *single base extension*, o *SBE*, y la *allele-specific oligonucleotide ligation*, u *OLA*) y métodos de detección como el MALDI-TOF y la detección fluorescente, que permiten la identificación de alelos en varios SNP al mismo tiempo (lo que se conoce como reacciones *multiplex*). Las principales plataformas comerciales de este tipo son tres: MassARRAY-Sequenom<sup>®</sup>, Illumina<sup>®</sup> y Affymetrix<sup>®</sup> (fig. 2).

1. MassARRAY Sequenom<sup>®</sup>. Es la tecnología más flexible en cuanto a la capacidad de inclusión de SNP para el genotipado, aunque sólo ofrece la posibilidad de genotipar SNP elegidos por el usuario. La tecnología iPLEX<sup>®</sup> Gold se basa en la extensión de un *primer* en un único nucleótido; posteriormente la identidad del mismo se determina a través de su tiempo de

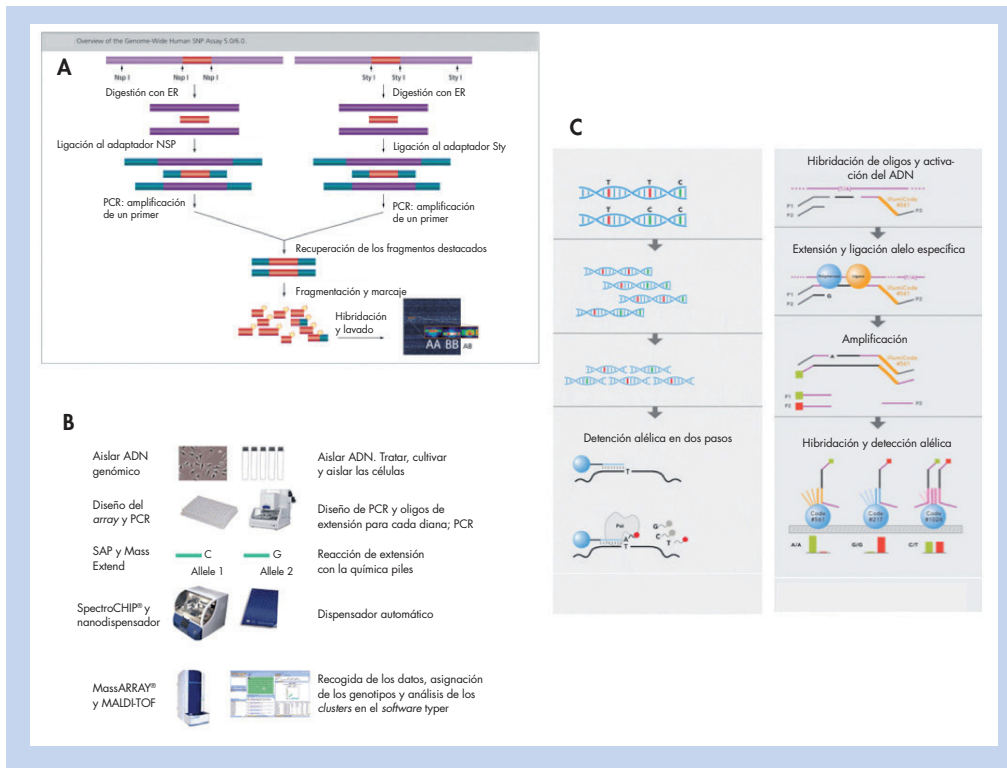
vuelo en un espectrómetro de masas de tipo MALDI-TOF.

2. Illumina<sup>®</sup>. Es una de las plataformas con mayor capacidad de SNP por reacción, pudiendo además analizar varias muestras en paralelo. Dentro de sus ofertas de genotipado se encuentran tanto los *arrays* a medida (*custom arrays* basados en las tecnologías GoldenGate<sup>®</sup> y VeraCode<sup>®</sup>), como los *arrays* para genomas completos (tecnología Infinium<sup>®</sup>), cuya cobertura ha ido aumentando, acercándose en estos momentos a los 2M de marcadores.

3. Affymetrix<sup>®</sup>. Se basa exclusivamente en *arrays* comerciales con marcadores ya preestablecidos. Además de los de genoma completo (Affy 6.0 y Human 20K CSNP Kit), existen otros con regiones diana específicas, como el Human Immune and Inflammation 9K SNP Kit, que incluye marcadores en genes relacionados con las rutas del sistema inmunitario.

## GWAS: interpretación de resultados y aproximaciones futuras

La base genética de muchas enfermedades complejas ha sido abordada con éxito mediante



**Figura 2.** Tecnologías high-throughput de genotipado de polimorfismos de nucleótido simple. A. Affymetrix®. B. MassARRAY-Sequenom®. C. Illumina®.

los estudios GWAS. Se han identificado y replicado asociaciones de las variantes genéticas relacionadas con enfermedades como la diabetes tipo 1, o los cánceres de colon y próstata<sup>14-19</sup>. Los resultados obtenidos son compatibles con un modelo de herencia poligénico. Sin embargo, estas variantes presentan, en general, unos riesgos asociados moderados (*odds ratio* [OR] = 1-1,25), son muy comunes (MAF > 10%) y la mayoría corresponde a tagSNP, los cuales no suelen tener una importancia funcional por ellos mismos, por lo que se desconoce la base molecular de la asociación encontrada<sup>19</sup>. En consecuencia, es necesario realizar una secuenciación masiva de estos *loci* para identificar la variante causal, así como promover estudios a gran escala en los que se analice un mayor número de muestras en *arrays* con mayor densidad de SNP. Otra aproximación es efectuar metaanálisis de GWAS, algunos de los cuales ya han identificado con éxito nuevos *loci* de susceptibilidad<sup>21-23</sup>, así como explorar otras formas de variabilidad genética.

Recientemente se ha descubierto que las variantes estructurales más comunes, que representan más del 15% de la variabilidad genética del genoma, son aquellas que afectan al número de copias de un fragmento de ADN<sup>24,25</sup>. Estas variantes, denominadas *copy number variants* (CNV), se definen como fragmentos de ADN (de 1 Kb hasta varias Mb) que presentan un número variable de copias en comparación con el

genoma de referencia. Las consecuencias funcionales de los CNV incluyen: alteración de la dosis génica y interrupción de genes o de elementos reguladores de la expresión. De ahí deriva su implicación en la etiología de enfermedades complejas y mendelianas<sup>26,27</sup>. Compañías como Affymetrix® e Illumina® han rediseñado sus *arrays* de GWAS para incluir sondas específicas de CNV, lo que permite estudiar las implicaciones de estas variantes en el desarrollo de enfermedades complejas. Aunque la hibridación genómica comparada (CGH, de *comparative genomic hybridization*) es la metodología más utilizada para estudiar CNV<sup>28,29</sup>, se espera que las nuevas tecnologías de secuenciación permitan una mayor definición de estas regiones génicas.

## Secuenciación de nueva generación

Desde la invención a finales de los setenta de los métodos de secuenciación de ADN, el método enzimático de Sanger ha sido el *gold standard* de la secuenciación. Este método automatizado permite la lectura de secuencias de ADN de una longitud de hasta 750 pb, con una altísima precisión de lectura. A pesar de la innegable aplicabilidad de la técnica, a partir del año 2003 empezaron a aparecer nuevas estrategias que combinaban la secuenciación de miles de

## Lectura rápida



Los proyectos Genoma Humano y HapMap aportaron datos muy valiosos sobre la frecuencia y herencia en bloques de los SNP.

Los resultados obtenidos en los GWAS abren nuevos frentes en la investigación de enfermedades humanas, que van desde la secuenciación masiva de los *loci* de susceptibilidad identificados, en búsqueda de la variante causal, hasta el estudio de otras formas de variabilidad genética.

Los *copy number variants* (CNV) son las variantes estructurales más comunes y representan más del 15% de la variabilidad genética del genoma. Es conocida su implicación en la etiología de enfermedades complejas y mendelianas por alteración de la dosis génica.

La aparición de las nuevas metodologías de secuenciación (NGS, de *next-generation sequencing*), que permiten la caracterización de regiones del genoma mucho más largas en periodos menores de tiempo y de forma coste-efectiva, está revolucionando la investigación genómica.

La NGS tiene múltiples aplicaciones, que incluyen desde la caracterización de *loci* de susceptibilidad identificados en los GWAS, hasta la secuenciación de genomas completos y estudios de ARN.



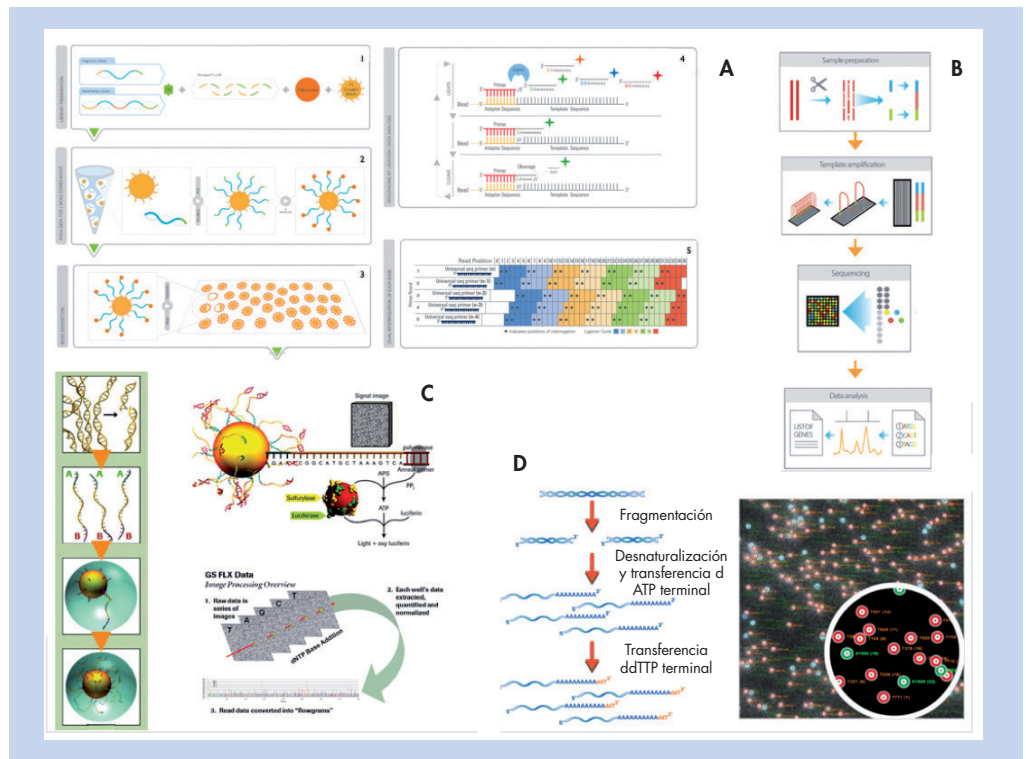
## Bibliografía recomendada

Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. *Trends Genet.* 2003;19:135-40.

El hallazgo de las propiedades de desequilibrio de ligamiento de los marcadores en el genoma supuso un evento importante en la genómica, ya que a través de la exploración de unos pocos marcadores se podían obtener datos de la mayoría de la variación genética de un individuo.

Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005;6:95-108.

Los estudios de asociación proporcionan una aproximación muy potente a la hora de determinar las variantes genéticas de baja susceptibilidad que predisponen a desarrollar una patología. Los estudios de asociación suponen una herramienta fundamental en este aspecto, ya que permiten la investigación de un gran número de genes y regiones a lo largo de todo el genoma.



**Figura 3.** Plataformas next-generation sequencing. **A.** SOLiD™. Consiste en una amplificación por reacción en cadena de la polimerasa (PCR) en emulsión seguida de la secuenciación por ligasas. La secuenciación incluye, además, la introducción de una codificación de las sondas por pares de bases, de forma que la identidad de cada base es comprobada por duplicado. **B.** Solexa. Se basa en una "bridge PCR" y una reacción de secuenciación con adición de terminadores reversibles en cada ciclo. **C.** 454 FLX Roche. Combina el uso de una PCR en emulsión y la pirosecuenciación. **D.** HeliScope™. No existe amplificación clonal. En su lugar, se usa un sistema de fluorescencia muy flexible que permite la detección del nucleótido incorporado a las cadenas en síntesis.

pequeños fragmentos de ADN con un ensamblaje posterior para crear lecturas de varios Mb. Estas nuevas técnicas fueron denominadas *next-generation sequencing*. La principal ventaja de estas nuevas plataformas es la capacidad de producción de grandes volúmenes de datos de forma coste-eficiente<sup>30</sup>.

En la actualidad, las plataformas NGS disponibles son principalmente 4: 454 FLX Roche, GS FLX Genome Analyzer o Solexa (Illumina), SOLiD™ (Applied Biosystems) y HeliScope™ (Helicos Biosciences) (fig. 3). A pesar de que difieren en cuanto a la metodología química usada y a la generación del *array* de secuencias, el esquema de trabajo es similar: la preparación de una librería de fragmentos, seguida de una amplificación y una ligación a adaptadores. El proceso de secuenciación alterna ciclos enzimáticos con un registro de fluorescencia que permite determinar la secuencia.

Los avances en las tecnologías NGS serán de especial utilidad no sólo para la caracterización de *loci* de susceptibilidad identificados en GWAS y asociados a enfermedades complejas, sino también para la resecuenciación de genomas, secuenciación de genomas *de novo*,

determinación de regiones de interacción proteína-ADN, análisis de metilación o análisis de distintos tipos de ARN (ARN-Seq). De hecho, cabe destacar la importancia de estas tecnologías en proyectos actuales como el 1000Genomes Project<sup>31,32</sup> y el proyecto de secuenciación del genoma Neanderthal<sup>33</sup>.

Además de las plataformas descritas, se espera que durante el 2010 salgan al mercado nuevas tecnologías con modificaciones del método, que ya se han bautizado como *next-next-generation* o tercera generación (Pacific Biosciences, Life/Visigen, LI-COR Biosciences)<sup>34</sup>.

## Bibliografía

GH [www.ghcontinuada.com](http://www.ghcontinuada.com)  
Encontrará enlaces a los resúmenes de esta bibliografía

- Importante ●● Muy importante
- Metaanálisis

1. Erichsen HC, Chanock SJ. SNPs in cancer research and treatment. *Br J Cancer.* 2004;90:747-51.
2. <http://www.ncbi.nlm.nih.gov/projects/SNP/> (dbSNP)

3. [http://www.ornl.gov/sci/techresources/Human\\_Genome/faq/snps.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml) (SNP fact sheet)
4. ●● De la Chapelle A. Genetic predisposition to colorectal cancer. *Nat Rev Cancer*. 2004;4:769-80.
5. ●● Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. *Nat Rev Cancer*. 2004;4:850-60.
6. Hunter DJ, Thomas G, Hoover RN, Chanock SJ. Scanning the horizon: what is the future of genome-wide association studies in accelerating discoveries in cancer etiology and prevention? *Cancer Causes Control*. 2007;18:479-84. <http://hapmap.ncbi.nlm.nih.gov/>
7. ●● Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. *Trends Genet*. 2003;19:135-40.
8. Mateswari CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nat Rev Genet*. 2008;9:9-14.
9. Gibbs JR, Singleton A. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. *PLoS Genet*. 2006;2:e150.
10. Jorgenson E, Witte JS. A gene-centric approach to genome-wide association studies. *Nat Rev Genet*. 2006;7:885-91.
11. Dearlove AM. High throughput genotyping technologies. *Brief Funct Genomic Proteomic*. 2002;1:139-50.
12. Ragoussis J. Genotyping technologies for genetic research. *Annu Rev Genomics Hum Genet*. 2009;10:117-33.
13. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42:105-16.
14. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet*. 2007;39:984-8.
15. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet*. 2007;39:989-94.
16. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447:1087-93.
17. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*. 2007;39:645-9.
18. Haiman CA, Le Marchand L, Yamamoto J, Stram DO, Sheng X, Kolonel LN, et al. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet*. 2007;39:954-6.
19. Cazier JB, Tomlinson I. General lessons from large-scale studies to identify human cancer predisposition genes. *J Pathol*. 2009;220:255-62.
20. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet*. 2008;40:1426-35.
21. McMahon FJ, Akula N, Shulze TG, Muglia P, Tozzi F, Detera-Wadleigh SD, et al; Bipolar Disorder Genome Study (BiGS) Consortium. Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nat Genet*. 2010;42:128-31.
22. Bouzigon E, Forabosco P, Koppelman GH, Cookson WO, Dizier MH, Duffy DL, et al. Meta-analysis of 20 genome-wide linkage studies evidenced new regions linked to asthma and atopy. *Eur J Hum Genet*. 2010;18:700-6. <http://projects.tcag.ca/variation/>
23. Estivill X, Armengol L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet*. 2007;3:1787-99.
24. ●● Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444-54.
25. Lupski JR. Genomic rearrangements and sporadic disease. *Nat Genet*. 2007;39 7 Suppl:S43-7.
26. Sharp AJ. Emerging themes and new challenges in defining the role of structural variation in human disease. *Hum Mutat*. 2009;30:135-44.
27. De Cid R, Riveira-Munoz E, Zeeuwen PL, Robarge J, Liao W, Dannhauser EN, et al. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet*. 2009;41:211-5.
28. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. 2008;26:1135-45.
29. Kuehn BM. 1000 Genomes Project promises closer look at variation in human genome. *JAMA*. 2008;300:2715. <http://www.1000genomes.org/page.php>
30. Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, et al. Sequencing and analysis of Neanderthal genomic DNA. *Science*. 2006;314:1113-8.
31. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010;11:31-46.

## Bibliografía recomendada

Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010;11:31-46.

*Las tecnologías de secuenciación de nueva generación basan sus protocolos en tres procedimientos básicos: la generación de librerías de pequeños fragmentos de ADN, su lectura por un escáner de fluorescencia, y el ensamblaje de los millones de readings para recrear secuencias finales de hasta varias Mb.*

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006;444:444-54.

*Las variaciones en el número de copia de fragmentos de ADN, o copy number variants, suponen una de las fuentes de variación más comunes en el genoma. Además, se ha demostrado que pueden alterar mecanismos como la dosis génica y afectar a la expresión, por lo que se cree que, al igual que los polimorfismos de nucleótido simple, podrían funcionar como loci de bajo-moderado riesgo en el desarrollo de las enfermedades complejas.*