Educ. quím., **22**(1), 15-27, 2011. © Universidad Nacional Autónoma de México, ISSN 0187-893-X
Publicado en línea el 3 de noviembre de 2010, ISSNE 1870-8404

Genómica y medicina

*Irma Silva Zolezzi**

ABSTRACT (Genomics and Medicine)
During the last 20 years the world has witnessed how scientists deciphered the human genome sequence and developed strategies to extensively study genetic diversity, linkage disequilibrium (LD) structure and genetic expression in human populations. These remarkable advances have made possible the comprehensive assessment of genetic diversity in many biological relevant scenarios, which has helped increase our understanding about the contribution of genetic factors to complex traits and diseases. The increased knowledge in human genomics has paved the way for the development of "genomic medicine", a field based on the development of new diagnostic and therapeutic approaches for common multi-factorial diseases. Nevertheless, the advancement is still limited and a lot is yet to be discovered. Currently, new technological tools, such as next-generation sequencing are being generated, and innovative approaches like epigenomic profiling, are being developed to undertake the immense challenges of comprehensively characterizing the genetic bases of complex traits and diseases; as well as studying the interaction between genetic and environmental factors such as diet, pharmacological treatments, infectious agents and others. Although the translation of this knowledge to the clinical practice is still minimal, the genomics sciences are inexorably changing our understanding of the biology of a myriad of medical conditions. This review intends to cover the key concepts and most basic technology and experimental designs that have been instrumental to study the link between our "genome and medicine".

KEYWORDS: genomics, complex diseases, genetic diversity, microarrays, sequencing

La información genética

Todos los organismos vivos tienen un genoma con la información biológica indispensable para su desarrollo y función, la cual se hereda a la siguiente generación. El genoma está constituido por una o más moléculas de ácido desoxirribonucleico (ADN), un tipo de ácido nucleico, el cual es un polímero de compuestos químicos denominados nucleótidos, que están constituidos por una base nitrogenada, un azúcar pentosa, y de uno a tres grupos fosfato (figura 1). Existen dos tipos de nucleótidos de acuerdo con la base nitrogenada que se incorpora a la molécula: 1) Purinas: adenina (A) y guanina (G); y 2) Pirimidinas: citosina (C), timina (T), y uracilo (U) (figura 1). Este último, el uracilo, sólo está presente en otro ácido nucleico, estructural y funcionalmente distinto al ADN, conocido como ácido ribonucleico (ARN). En la estructura del ADN, los nucleótidos se unen covalentemente por enlaces fosfodiéster formando dos cadenas independientes, las cuales interaccionan entre sí, a través de puentes de hidrógeno entre pares de bases nitrogenadas, siempre dos entre A y T, y tres entre C y G. Lo anterior origina que la secuencia de cada cadena sea inversa y complementaria a la otra, y da lugar a la estructura de doble hélice característica del ADN (figura 2). El genoma contiene secuencias discretas con la información necesaria para generar moléculas con función biológica a las que denominamos genes, y en el caso del humano, el genoma está organizado en el núcleo en un conjunto de 23 pares de cromosomas (figura 3).

El Proyecto del Genoma Humano (PGH) fue un esfuerzo internacional cuyo objetivo fue determinar la secuencia de nucleótidos y los genes contenidos en el ADN humano. El PGH inició en los años noventa y se consideró terminado en el 2003, al haberse logrado la caracterización de la mayor parte de los más de 3,200 millones de pares de bases (bp) que contiene (International Human Genome Sequencing Consortium, 2004). El avance asociado a esta hazaña científica ha ocasionado una profunda evolución de la definición de gen. Originalmente el "gen" fue definido como "la unidad de herencia", unos años después con el establecimiento del dogma central de la biología molecular, ADN→ARN→Proteína, por Crick, se definió como "un segmento de ADN que codifica a una proteína"; y hoy, a la luz del descubrimiento de moléculas de ARN que no codifican a proteína (ncARN) pero que tienen un papel importante en la regulación de la expresión genética, se define como "un segmento de ADN que posee información necesaria para generar moléculas con función biológica (ARN o proteínas) (Gerstein, *et al.*, 2007). Hoy, gracias al PGH sabemos que el genoma humano posee alrededor de 22,280 genes que codifican proteínas (International Human Genome Sequencing Consortium, 2004), un número similar o incluso menor al de otras especies, como las vacas (*Bos taurus*), un tipo de pasto (*Brachypodium distachyon*) y el arroz (*Orzya sativa*), que poseen 22,000 (Elsik, *et al.*, 2009), 25,532 y 28,236 (The International Brachypodium Initiative, 2010), respectivamente. El

* Grupo de Genómica Funcional, Departamento de Ciencias Bioanalíticas. Centro de Investigación Nestlé, Lausana, Suiza.
Correo electrónico: irma.silvazolezzi@rdls.nestle.com

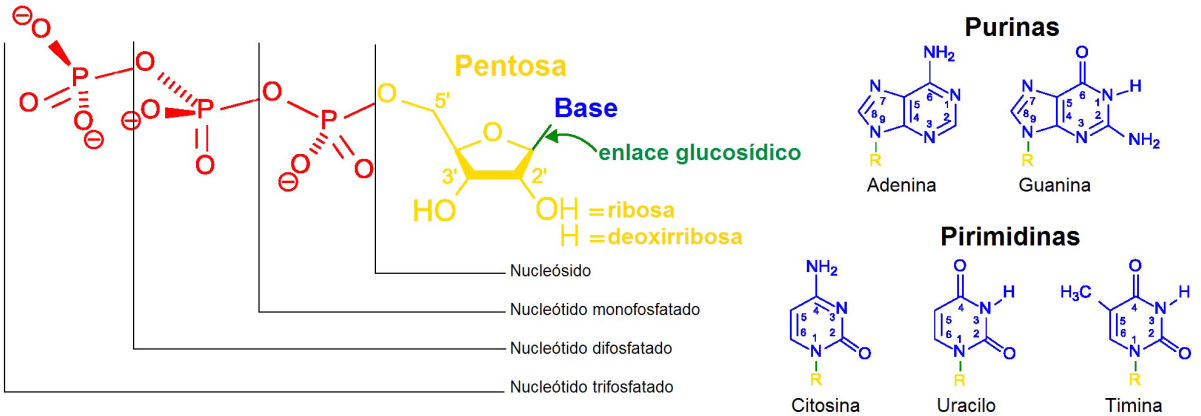

Figura 1. Los nucleótidos son los compuestos químicos que constituyen a los ácidos nucleicos. Cada nucleótido posee en su estructura un nucleósido formado por una pentosa (amarillo) unida a una base nitrogenada (azul) mediante un enlace glucosidico (verde), más uno a tres grupos fosfato (rojo). De acuerdo con la base nitrogenada (azul) que se incorpora a la molécula existen dos tipos de nucleótidos: purinas (adenina y guanina), y pirimidinas (citosina, timina, y uracilo). (Los colores pueden verse en la versión en línea en http://educacionquimica.info).

número total de genes casi se duplica, si se consideran los genes no-codificantes a proteína, los cuales en la base de datos del National Center for Biotechnological Information (NCBI) de EUA (agosto, 2010), eran 42,488 en el genoma humano y 33,086 en el de la vaca. Lo anterior refleja el alto contenido de información en regiones no-codificantes del genoma.

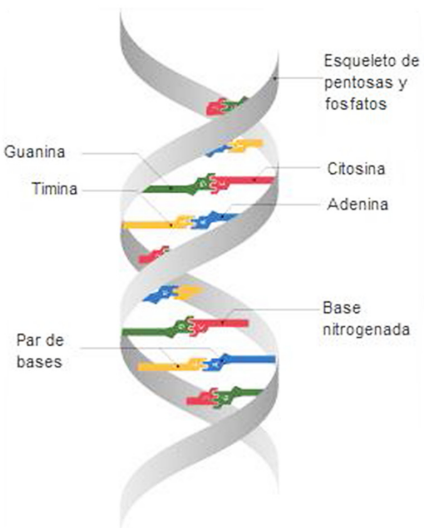

Figura 2. El ácido desoxirribonucleico (ADN) tiene una estructura de doble hélice alfa. El ADN está compuesto de diferentes subunidades. El esqueleto de la molécula está hecho de dos polímeros de nucleótidos unidos por las pentosas (desoxirribosas) a través de los grupos fosfato. La secuencia de ambas cadenas es complementaria y anti-paralela, y entre ellas interaccionan a través de puentes de hidrógeno entre las bases nitrogenadas presentes en su estructura, siempre adenina con timina (dos puentes de hidrógenos), y guanina con citosina (tres puentes de hidrógeno).

Los seres humanos nacemos, crecemos y morimos prácticamente con la misma información genética en nuestras células; además, heredamos 50% de esta información a nuestra progenie. Debido a su naturaleza estable y heredable, la información genética representa una fuente potencial de biomarcadores de amplia aplicación clínica, por lo que existe enorme interés en estudiar el papel de los factores genéticos en la enfermedad humana. El PGH ha sido un motor fundamental

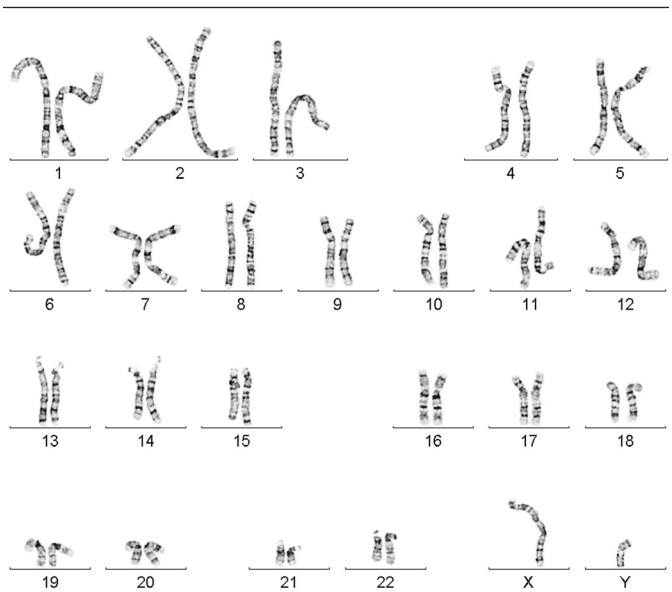

Figura 3. Cariotipo humano. Conjunto normal de cromosomas humanos de origen masculino teñidos con una técnica tradicional de bandeo, incluyendo 22 pares de autosomas (numerados del 1 al 22) y un par de cromosomas sexuales X y Y. Figura tomada del J. Craig Venter Institute.

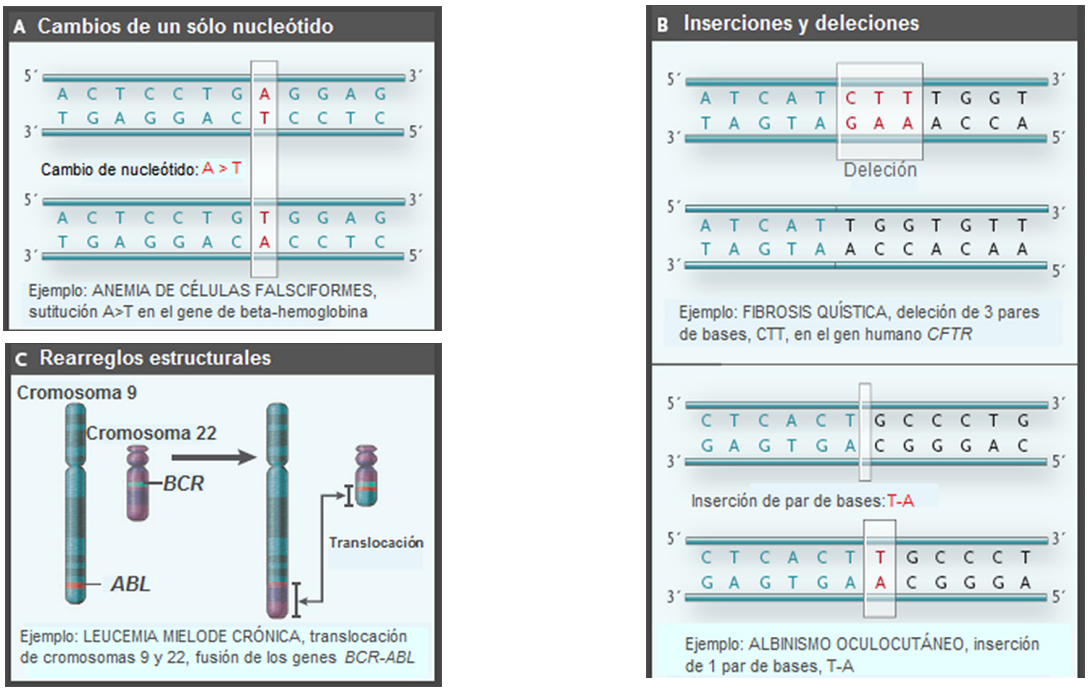

Figura 4. Variacion genética humana. Las variantes genéticas humanas se pueden agrupar en tres categorías principales: cambios de una sola base (Panel A), eventos de insercion y deleción pequeños y grandes (Panel B) y rearreglos estructurales (Panel C). La escala y consecuencias de estos cambios puede variar dramáticamente, dependiendo de donde y cuando aparece esta variación. Por ejemplo, el cambio de un solo par de bases puede tener graves consecuencias para la salud (ej. La substitución de una timina por una adenina en el gen de la beta-hemoglobina humana), mientras que un evento de translocación grande pero balanceado (en el que la información genética de un brazo completo del cromosoma puede intercambiar posición con el brazo completo de otro cromosoma) puede no tener consecuencias directas para la persona o célula afectada. Traducida de Feero, W.G. *et al.* (2010), con autorización.

para la investigación biomédica y clínica, y ha permitido el avance de la medicina genética, la cual se enfoca al diagnóstico y tratamiento de enfermedades causadas por defectos en genes específicos, que conocemos como enfermedades monogénicas (ej. hemofilia, fibrosis quística, fenilcetonuria, etc.). Sin embargo, este amplio desarrollo importante en medicina, ha beneficiado sólo a una pequeña proporción de la población dada la baja prevalencia de este tipo de enfermedades, aproximadamente 10 de cada mil nacimientos (http://www.who.int/genomics/public/geneticdiseases/en/).

Es así que en la última década, se han desarrollado metodologías y abordajes de investigación basados en el conocimiento genómico para intentar descubrir los factores genéticos que participan en la conformación de los rasgos y enfermedades "frecuentes o comunes" a las poblaciones humanas, como la diabetes o el cáncer, que son importantes causas de morbilidad y mortalidad, La mayoría de los rasgos y enfermedades comunes (ej. estatura, niveles de glucosa, cáncer, diabetes, enfermedades cardiovasculares, etc.) son "complejos o multifactoriales"; es decir, resultan de la interacción de factores genéticos y ambientales. Por lo anterior, es fundamental no sólo identificar a estos factores, sino también el papel de las interacciones entre genes y ambiente en el desarrollo y la severidad de los mismos. El creciente conocimiento acerca del papel del genoma humano en la enfermedad, ha sentado las bases científica y tecnológica para el desarrollo de la "medicina genómica", una nueva rama de la medicina basada en la aplicación del conocimiento genético en el diagnóstico y tratamiento de las enfermedades complejas (Feero, Guttmacher and Collins, 2010). Adicionalmente a la medicina, la potencial aplicación del conocimiento derivado del genoma humano ha sido de gran interés para la industria farmacéutica y la de alimentos. Para la primera, como una fuente de conocimiento e información para favorecer el desarrollo de nuevos fármacos y nuevos esquemas terapéuticos (ej. dosis ajustadas de acuerdo con el genotipo, nuevas aplicaciones para fármacos en uso, etc.), y para la segunda, como una oportunidad de posicionar ciertos alimentos, ingredientes específicos o bioactivos nutricionales/nutracéuticos, de acuerdo con su beneficio potencial en base a las características genéticas del consumidor para promover la salud y prevenir enfermedad. Lo anterior ha impulsado el desarrollo de la Farmacogenómica y la Nutrigenómica; ambas disciplinas utilizan estrategias "ómicas" (genómica, transcriptómica, proteómica, epigenómica, metabolómica, etc.), para estudiar las interacciones factores genéticos y factores ambientales y, en particular, medicamentos y dieta, respectivamente.

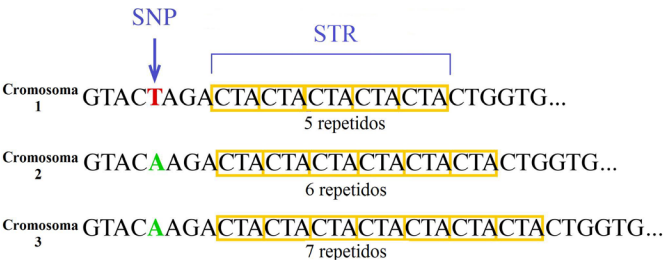

Figura 5. Variantes genéticas bi y multialélicas. Con base en el número de posibilidades en la que se presentan las variantes genéticas se clasifican en: bialélicas cuando se presentan generalmente en sólo dos formas comos los polimorfismos de un solo nucleótido o SNPs por sus siglas en inglés (ej. T>A) y multialélicas cuando se presentan en tres o más alternativas como los repetidos corto en tándem o STRs por sus siglas en inglés.

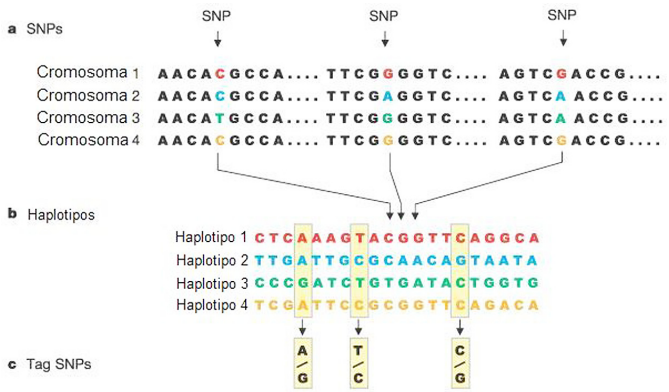

Figura 6. Haplotipos y SNPs etiqueta. De manera general la identificación de SNPs etiqueta ocurre en tres etapas: a) se caracterizan un conjunto de SNPs distribuidos en el genoma en muestras de ADN de diferentes individuos de una misma población; b) se reconstruyen computacionalmente las combinaciones de alelos de SNPs adyacentes que se heredan juntos (haplotipos), y c) se identifican los SNPs que permiten caracterizar todos los haplotipos identificados a los que llamamos "SNPs etiqueta". Al genotipificar los tres SNPs etiqueta que se muestran en esta figura, los investigadores pueden identificar cuales haplotipos (dos en el caso del humano) de los cuatro posibles tiene cada individuo de esa población. Tomada del sitio del HapMap (www.hapmap.org).

Variabilidad genética

Hoy sabemos que no existe una secuencia "normal" del genoma de la especie humana. Sin embargo, gracias a esfuerzos como el PGH contamos con secuencias "referencia" revisadas y anotadas, que usamos para estudiar la diversidad genética humana. Estas secuencias están en la base de datos de secuencias genéticas más importante del mundo (GenBank®). Esta información, así como varias herramientas para su visualización y análisis están disponibles de manera pública y gratuita a través de portales de internet (NCBI, Ensembl, UCSC, etc.). Algunas de las secuencias como la RefSeq de NCBI [1] no provienen de un individuo en particular sino del análisis de varias muestras anónimas de ADN; otras, como la HuRef del Instituto J. Craig Venter, son resultado de la secuenciación de un genoma individual (Levy, *et al.*, 2007).

Una mutación o variante genética, es una base o secuencia en el ADN que puede diferir entre individuos de la misma especie. Las variantes genéticas pueden clasificarse de manera sencilla en tres grupos: a) cambios de una sola base o nucleótido, b) eventos de inserción y deleción, y c) rearreglos estructurales (figura 4). Se le denomina alelo a cada variante o posibilidad de base o secuencia en la que se presenta una posición en el genoma; a su vez, cada variante de acuerdo con el número de formas en las que se presentan en el genoma, puede ser bialélica (ej. los polimorfismos de un solo nucleótido o snps), o multialélica (ej. repetidos en tándem de número variable o VNTRs) (figura 5). Cuando el alelo menos frecuente de una variante se encuentra en más del 1% de los cromosomas de una población se considera "común" y se le llama polimorfismo. Por su naturaleza todos los polimorfismos son mutaciones, pero el término "mutación" generalmente se usa para variaciones genéticas de "baja frecuencia o raras" (menos del 1%), como son casi todas las mutaciones causantes de enfermedades monogénicas. El conjunto de variaciones genéticas comunes y raras presentes en el genoma de un individuo constituyen su genotipo, el cual, en estrecha relación con los factores ambientales, da lugar a características individuales (normales o patológicas), es decir a su fenotipo. Las diferencias genéticas entre pares de humanos se han estimado en 0.5 a 1%, lo que corresponde a aprox.16 a 32 millones de bp diferentes entre individuos (Feero, Guttmacher and Collins, 2010; Frazer, *et al.*, 2009; Jakobsson, *et al.*, 2008).

Los SNPs son las variaciones genéticas más comunes; en las poblaciones humanas se han identificado al menos 10 millones de SNPs con frecuencias mayores que el 1% al menos en algún grupo humano (http://www.ncbi.nlm.nih.gov/snp/). Cada ser humano tiene en su genoma representado al alelo menor de al menos tres a cuatro millones de SNPs, de los cuales entre dos individuos se llegan a compartir aproximadamente 1.7 millones (Frazer, *et al.*, 2009). Aunque los SNPs son las variantes genéticas más estudiadas en medicina genómica, recientemente también se ha demostrado un papel importante para las variaciones de número de copia o CNVs (Stankiewicz and Lupski, 2010), que son variantes estructurales que consisten en segmentos de ADN de aproximadamente 1000 bp (1Kb) o más, que han sido eliminados, o insertados una o más veces en el genoma (Feuk, Carson, and Scherer, 2006). A los conjuntos de variaciones genéticas (ej. SNPs y/o CNVs) que se encuentran en una región definida del mismo cromosoma, y que se heredan juntas en una población, se les conoce como haplotipos. Su estudio y caracterización en las poblaciones humanas ha permitido identificar SNPs que al asociarse físicamente con otras variantes en el mismo

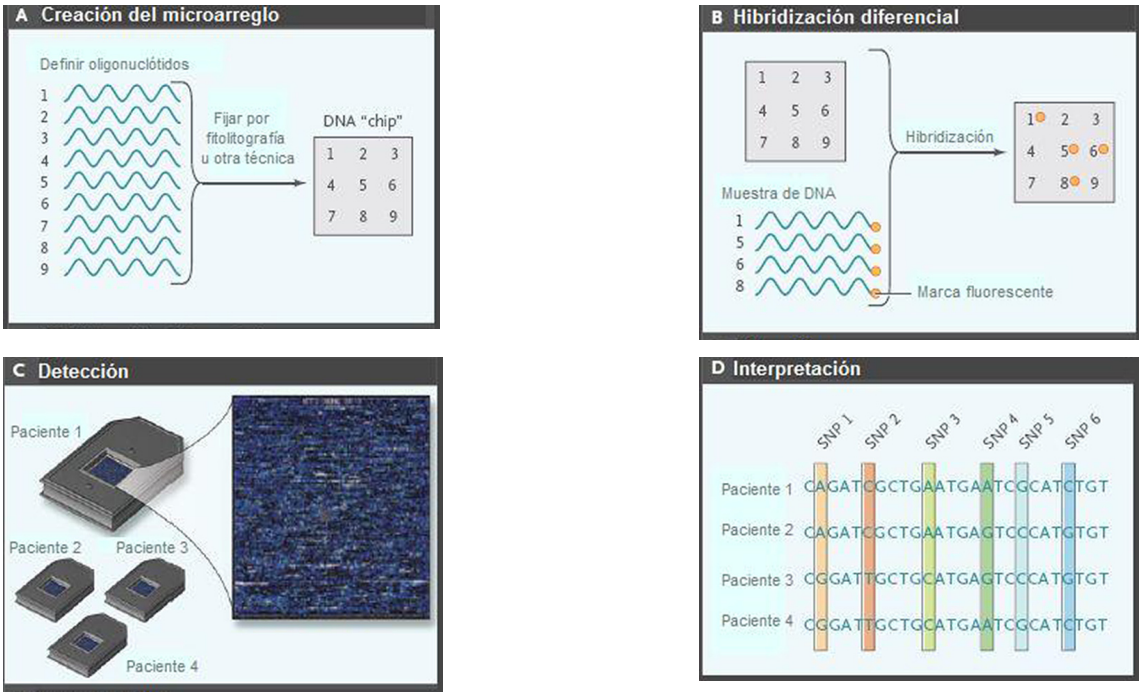

Figura 7. Tecnologías de microarreglos. Las tecnologías de microarreglos o chips genéticos son centrales a la mayoría de los más importantes avances científicos y clínicos de los últimos cinco años. Con la tecnología actual, usando un solo microarreglo se puede obtener información de más de un millón de variantes de secuencia única. Los microarreglos utilizados para detectar SNPs comparten algunas características: primero, la presencia de un arreglo microscópico estructurado de oligonucleótidos con secuencia definida en un sustrato sólido como un chip de vidrio o silicio (Panel A); segundo, la hibridización al arreglo de una muestra de ADN marcada con fluorescencia (Panel B); tercero, el subsecuente escaneo del arreglo para detectar la localización y la cantidad de unión específica de secuencia en la muestra (Panel C); y por último, el procesamiento computacional de los datos crudos de la imagen del arreglo para dar lugar a una lectura interpretable de los datos de SNPs (Panel D). Figura traducida de Feero, W.G. *et al.* (2010), con autorización.

haplotipo, pueden ser caracterizados selectivamente para interrogar simultáneamente al resto (figura 6), a lo que se le conoce como "etiquetado o *tagging*" de SNPs (Balding, 2006). La estructura de haplotipos es una característica poblacional; por lo tanto, los SNPs "etiqueta o *tag*" pueden ser diferentes en grupos poblacionales distintos (Frazer, *et al.*, 2009). En la última década, densos mapas de variabilidad genómica se han generado para algunas poblaciones, incluyendo la mexicana, con el objetivo de permitir la reconstrucción de haplotipos y la selección de SNPs etiqueta (The International HapMap Project, 2003; A haplotype map, 2005; Teo, *et al.*, 2009; Silva-Zolezzi, *et al.*, 2009). Los mapas de haplotipos más completos disponibles hoy en día son los de poblaciones con orígenes ancestrales del noroeste de Europa (CEU), del este de Asia (CHB y JPT) y del oeste de África (YRI), todas de la fase inicial del Proyecto Internacional del Mapa de Haplotipos (HapMap) (The International HapMap Project, 2003). Para estas poblaciones y las de estructura genética similar, existen ensayos basados en SNPs etiqueta, así como estrategias estadísticas y computacionales (ej. imputación) que permiten caracterizar ampliamente la diversidad genética (McCarthy, *et al.*, 2008). En el caso de poblaciones de origen europeo, existen plataformas tecnológicas que permite caracterizar 300,000 SNPs y con los resultados generar información digital sobre más de 4 millones de SNPs (Tian, *et al.*, 2008; Anderson, *et al.*, 2008).

Este conocimiento ha favorecido el desarrollo de tecnología para analizar desde unos cuantos hasta millones de SNPs y decenas de miles de CNVs en una muestra de ADN. La plataforma más utilizada con este fin es una tecnología conocida como microarreglo o "chip genético", el cual consiste en una matriz de oligonucleótidos secuencia-específicos (sondas) ordenados y unidos a una superficie sólida, que se unen específicamente o "hibridizan" al material genético a analizar previamente amplificado y marcado, generalmente con fluorescencia (figura 7). Posteriormente la señal fluorescente se cuantifica usando algún sistema de escaneo acoplado a un programa de análisis de imágenes para determinar cuáles son las variantes genéticas presentes en la muestra. Los microarreglos, inclusive los que evalúan millones de SNPs pueden procesarse por una persona, en unos cuantos días, por cientos de dólares; en contraste, generar la misma información hace menos de diez años, hubiese requerido de varias personas, meses de trabajo, y cientos de miles de dólares (Feero, Guttmacher, and Collins, 2010).

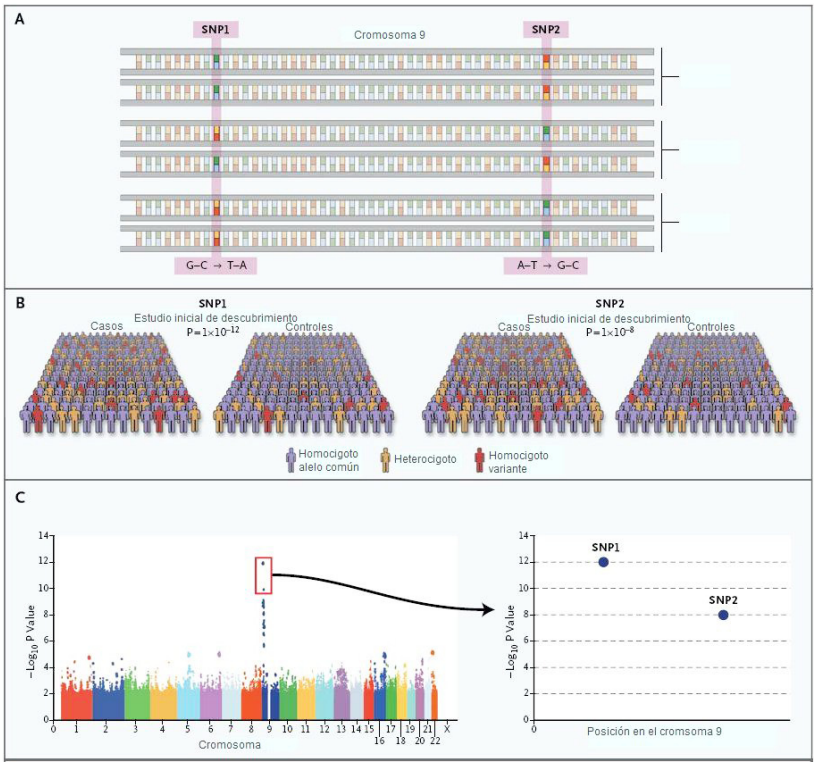

Figura 8. El estudio de asociación de genoma completo (GWAS). Los estudios de asociación de genoma completo se basan típicamente en un diseño caso-control en el que se genotipifican SNPs distribuidos a lo largo del genoma humano. El Panel A muestra un "locus" que corresponde a un fragmento pequeño del cromosoma 9. En el Panel B, la fuerza de la asociación entre cada SNP y la enfermedad se calculan en base a la prevalencia de cada SNP en el grupo de casos y en el de controles. En este ejemplo, los SNPs 1 y 2 del cromosoma 9 están significativamente asociados ($P<10^{-12}$ y 10^{-8}, respectivamente). La gráfica del Panel C muestra los valores de P (significancia estadística) para todos los SNPs genotipificados que pasaron todos los controles de calidad; cada cromosoma está representado con un color distinto. Los resultados señalan a un locus en el cromosoma 9, claramente marcado por dos SNPs adyacentes 1 y 2 (acercamiento en la gráfica a la derecha), y apoyado por otros SNPs en la misma región. Figura traducida de Manolio, T. A. (2010), con autorización.

Estudios de Asociación de Genoma Completo

Los estudios de asociación de genoma completo, también conocidos como GWAS por sus siglas en inglés, han revolucionado la investigación genómica en los últimos cinco años. Este abordaje ha sido posible en buena medida gracias a los microarreglos (figura 4), los cuales han facilitado la caracterización sistemática de una gran proporción de la variabilidad genética en estudios de casos y controles, estudios de cohorte y ensayos clínicos. El diseño al que se ha acudido mayoritariamente es el de casos y controles, el cual de manera muy general consiste en comparar la frecuencia de los factores estudiados (en este caso variantes genéticas) en un grupo de individuos afectados (ej. obesos) y un grupo de individuos no afectados (ej. delgados), para determinar si existe una diferencia significativa en la frecuencia de una o más de estas variantes entre ambos grupos (figura 8). Este abordaje posee fortalezas y limitaciones, que se han identificado y han sido evaluadas por diferentes grupos en el mundo (McCarthy, *et al.*, 2008). Una de sus principales fortalezas es que por su diseño libre de hipótesis, permite a la identificar regiones genéticas, que nunca se habrían sospechado como relevantes para las enfermedades estudiadas, y que por lo tanto probablemente no habrían sido incluidos en estudios basados en regiones específicas o candidato. Por otro lado, una limitación importante es la presencia de un número considerable de falsos positivos por efecto del análisis simultáneo de múltiples pruebas estadísticas (al menos una por SNP), así como las diferencias en frecuencias alélicas ocasionadas por diferencias en orígenes ancestrales o diferencias demográficas (estratificación poblacional) y no relacionadas al fenotipo en estudio.

Existen factores críticos en el diseño y análisis de GWAS, entre los que destacan tres: a) la selección cuidadosa de casos y controles para minimizar potenciales sesgos (ej. edad, sexo, etnicidad, etc.), b) la selección de los marcadores a caracterizar o genotipificar (ej. SNPs etiqueta, SNPs codificantes, etc.), y c) el control de calidad de los datos (McCarthy, *et al.*, 2008). Para descartar efectos de estructura genética y/o de factores ambientales es necesario validar las señales de asociación genética mediante su replicación con un alto nivel de certidumbre estadística ($P < 5 \times 10^{-01}$), en un número de individuos más grande al grupo en donde se identificó, y/o en diferentes poblaciones (McCarthy, *et al.*, 2008). Adicionalmente, es deseable generar evidencia del efecto funcional de la variación genética identificada (ej. modificación del patrón de expresión o de la función de un ARN o proteína codificada por un gen), lo que es un reto, ya que la mayoría de la asociaciones identificadas mediante este tipo de abordaje son indirectas, y son capturadas gracias al fenómeno de desequilibrio de ligamiento, es decir a la correlación entre pares de variantes genéticas dentro de una región cromosómica. Para identificar las variantes genéticas verdaderamente causales del efecto biológico relacionado al fenotipo, normalmente se recurre a estrategias de mapeo fino y/o re-secuenciación de las regiones identificadas en los GWAS. El mapeo fino consiste en analizar las regiones candidato identificadas en los GWAS utilizando

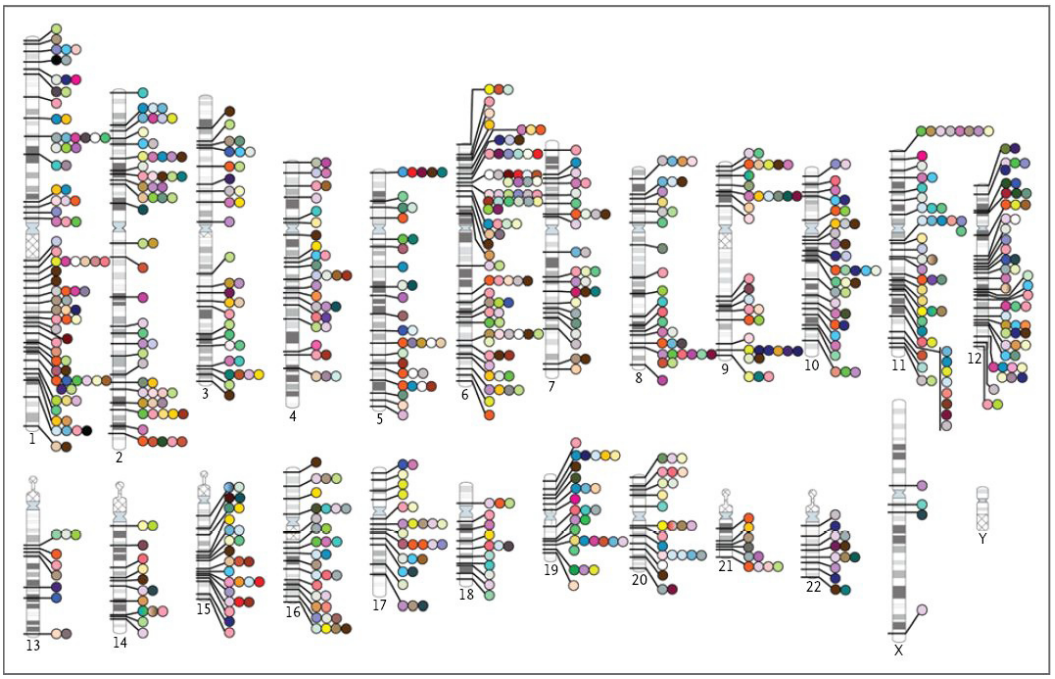

Figura 9. Estudios de asociación de genoma completo reportados hasta marzo de 2010. Los círculos indican la localización cromosómica de cerca de 800 SNPs significativamente asociados ($P < 5 \times 10^{-8}$) con una enfermedad o un rasgo y reportado en la literatura (545 estudios publicados hasta marzo de 2010 dieron lugar a las asociaciones anotadas). Cada enfermedad o rasgo esta codificado en un color distinto. Adaptada de la figura del Instituto Nacional de Investigación del Genoma Humano (http://www.genome.gov/gwastudies). Figura traducida de Manolio, T. A. (2010), con autorización.

mayor densidad de marcadores, en tanto que la resecuenciación se basa en analizar una a una cada base contenida en esas regiones.

A la fecha se han publicado más de 600 GWAS, en los que han sido identificadas aproximadamente 800 asociaciones significativas ($p < 5 \times 10^{-8}$) entre SNPs y, rasgos y/o patologías complejos a todo lo largo del genoma (figura 9) (Manolio, 2010). Estos resultados han sido importantes para el estudio de varias enfermedades, aunque con distinto grado de relevancia, ej. en la degeneración macular relacionada a la edad (DMRE), una enfermedad degenerativa que resulta en ceguera parcial o total en adultos mayores. Se han identificado cinco SNPs que explican más del 50% del riesgo genético en familiares de individuos afectados (Katta, Kaur, and Chakrabarti, 2009). Un escenario más común se ejemplifica en dos enfermedades autoinmunes: Chron y diabetes tipo 1, en las que se han identificado >30 SNPs, algunos con contribuciones importantes al riesgo y la mayoría con efectos pequeños (Manolio, 2010). Finalmente, para muchas enfermedades no se han encontrado respuestas claras, y en ocasiones para identificar alguna señal se ha requerido analizar a decenas de miles de individuos (ej. hipertensión) (Ehret, 2010), o de estrategias estadísticas sofisticadas como el meta-análisis (Ioannidis, Thomas, and Daly, 2009). En casi todos los fenotipos estudiados, las asociaciones tienen contribuciones pequeñas al riesgo y explican porcentajes bajos (5-10%) del riesgo genético total (Manolio, 2010). Lo anterior hace evidente que en el desarrollo de la medicina personalizada, estamos aún lejos de contar con perfiles genéticos predictivos, así como de contar con información suficiente para desarrollar políticas y lineamientos basados en evidencia que permitan llevar este conocimiento a la práctica clínica (Feero, Guttmacher, and Collins, 2010).

A pesar de la baja posibilidad de traducir de manera directa los resultados de los GWAS a un beneficio clínico, estos hallazgos han demostrado tener un valor fundamental en la identificación de nuevas vías y procesos biológicos subyacentes a la enfermedad humana. La mayor parte de este conocimiento, previo a la era de los GWAS, no se había sospechado como de potencial relevancia. Un ejemplo es la clara asociación de la vía del complemento en el desarrollo de la DMRE, y la evidencia de que la autofagia, un proceso catabólico de autodestrucción de componentes celulares, es un proceso relevante en el desarrollo de la enfermedad de Chron. La identificación de nuevos genes y vías biológicas asociados a patologías ha generado un gran número de candidatos para la investigación y el desarrollo de nuevos y mejores sistemas diagnóstico y tratamientos. Adicionalmente, se han identificado asociaciones en múltiples regiones genéticas que no permiten hacer hipótesis acerca de su relevancia funcional, y por lo tanto subrayan la existencia de lagunas en nuestro conocimiento. Por esto, es necesario desarrollar nuevas estrategias para identificar los mecanismos biológicos subyacentes a la

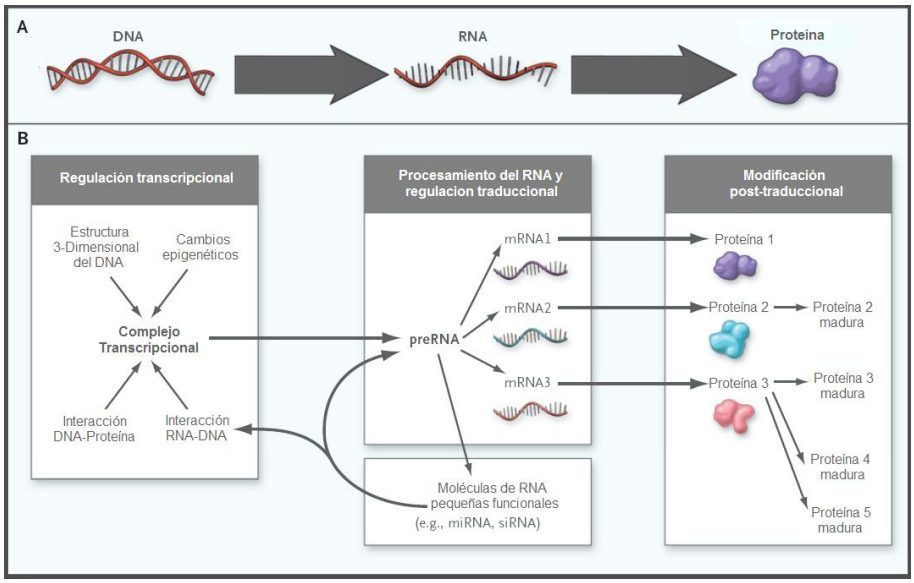

Figura 10. La creciente complejidad del dogma central de la biología molecular. El flujo de información genómica del ADN al ARN a la proteína continúa siendo la base para entender la función genómica (Panel A). Un solo gen puede dar lugar a un arreglo amplio de productos genéticos, dependiendo del ambiente en el que se expresa, lo cual resulta en una expansión del repertorio de alrededor de 20,000 genes codificantes a proteínas que existen en el genoma humano (Panel B). El evento inicial en la expresión genética, la transcripción, se regula a través de una compleja coreografía de eventos en la que están involucrados la estructura tridimensional del ADN, las modificaciones químicas covalentes o epigenéticas del esqueleto del ADN, e interacciones entre proteínas y ADN, y entre ARN y ADN. La traducción es también compleja y se encuentra estrictamente regulada por interacciones entre ARN mensajeros (mARN) y proteínas. El procesamiento de moléculas precursoras de ARN (pre-ARN) puede dar lugar a múltiples productos de ARN, incluyendo a moléculas de microARN (miARN) y de ARNs pequeños de interferencia (siARN). La modificaciones post-traduccionales (ej. doblamiento, corte y modificaciones químicas) de proteínas también contribuyen de manera importante a la diversidad resultante del genoma humano, a través de la modificación de proteínas inmaduras, lo que da lugar a un conjunto de productos protéicos relacionados. Figura y pie de figura traducidos de Feero, W.G. *et al.* (2010), con autorización.

enfermedad. Actualmente muchos recursos están siendo invertidos para descubrir las fuentes de riesgo heredable que hoy son desconocidas (Manolio, 2010), las cuales muy probablemente residen más allá de la secuencia de las bases del ADN, en el complejo universo de la regulación de la expresión genética.

El estudio de los perfiles de expresión genética

Un gran reto en la genómica humana es la caracterización de la compleja red estructural y funcional que se establece cuando la información genética de una célula, en un momento y circunstancia particular, es transcrita a ARN y es traducida a proteína (figura 10). Este proceso (transcripción) da lugar al conjunto de moléculas de ARN conocido como transcriptoma, y el conjunto de procesos coordinados de traducción de mARNs a proteínas y los eventos post-traduccionales (ej. glucosilación, esterificación, acetilación, etc.) da lugar al proteoma. En un organismo, transcriptoma y proteoma son cuerpos moleculares dinámicos que se transforman dependiendo del tipo celular, y de las condiciones particulares en las que se encuentran las células o tejidos (etapa del desarrollo, exposiciones a agentes endógenos y/o exógenos, etapa del ciclo circadiano, etc.). Lo anterior, contrasta con lo que sucede con el genoma de un organismo, el cual es estable y se mantiene prácticamente idéntico en casi todas las células a lo largo de su vida, con algunas excepciones como células sexuales y tumorales. El dinamismo del proteoma y el transcriptoma representa un reto para su estudio, pero el hecho de que estos llevan a cabo la mayoría de las funciones biológicas, los convierte en un recurso invaluable de información para entender la biología humana. Además, la identificación de biomarcadores proteicos es de fundamental relevancia por su alto potencial de utilidad clínica, gracias a la posibilidad de usar una gran variedad de técnicas convencionales no invasivas para su determinación, más sencillas que las que se utilizan para maracadores genéticos.

En paralelo al desarrollo de tecnología para analizar el genoma se han desarrollado herramientas para el análisis global del transcriptoma. Su aplicación ha resultado en generación de conocimiento de mucha utilidad en investigación biomédica y clínica, y ha revelado la importancia biológica en la regulación de la expresión genética de moléculas como los ARNs no codificantes (ncARNs) (Malecova, and Morris, 2010). En particular, se han identificado a los ARNs pequeños de inter-

ferencia y los microARNs (siARNs y miARNs) como de gran importancia en la enfermedad humana (Takizawa, *et al.*, 2010).

En medicina genómica la rama en la que el estudio del transcriptoma ha tenido mayor impacto es la oncogenómica (el estudio del genoma en relación con el cáncer). La aplicación de abordajes "ómicos" ha permitido a la identificación de perfiles de expresión genética para hacer sub-categorizaciones de utilidad clínica, ej. identificación tumores de buen pronóstico o de mal pronóstico, de acuerdo con la expectativa de sobrevida del paciente, a las aparición de metástasis o inclusive al beneficio de esquemas terapéuticos; algunos ejemplos notables existen en cáncer pulmonar (Baty, *et al.*, 2010; Santos, Blaya, and Raez, 2009), o en cáncer de mama (van de Vijver, *et al.*, 2002; Mook, *et al.*, 2010; Straver, *et al.*, 2010). La aplicación de esta estrategia también ha servido en algunos de estos escenarios para generar conocimiento acerca de genes clave para la oncogénesis (Gatza, *et al.*, 2010), lo anterior con mayor dificultad debido a la presencia de confusores, como la variación fisiológica que resulta de fenómenos generales como inflamación o hipoxia. Una oportunidad para lograr destacar aquello verdaderamente relevante para el desarrollo de la enfermedad, contra las alteraciones derivadas como resultado de la misma tanto en el escenario del cáncer como de otras enfermedades complejas, se vislumbra en estudio de las complejas interacciones biológicas, mediante el análisis conjunto de la variabilidad genética a nivel de ADN, los perfiles de expresión transcripcionales y proteicos, así como los mecanismos de regulación genética mediante estrategias basadas en el análisis de redes, en lo que se conoce como "Biología de Sistemas" (Gonzalez-Angulo, Hennessy, and Mills, 2010; Ramsey, Gold, and Aderem, 2010; Geschwind, Konopka, 2009).

Regulación de la expresión genética

Conocer cuáles son las características que definen la estructura de los genes y cómo se regula su expresión se ha convertido en un escenario francamente complejo. Hoy sabemos que múltiples regiones en el ADN, conocidas como secuencias motivo, participan en los mecanismos de regulación de la expresión genética. Estas secuencias pueden ubicarse en cualquier lugar con respecto al gen que regulan, dentro de su misma secuencia o hasta a millones de pares de bases (Mb) del mismo (Birney, *et al.*, 2007). En el control de la expresión genética participan una variedad de mecanismos moleculares mediados por interacciones ADN-proteína, ARN-ADN y ARN-ARN, así como modificaciones químicas sobre el ADN que no afectan la secuencia primaria, y modificaciones químicas sobre las proteínas que interactúan con el ADN. La regulación de la transcripción de ADN a ARN es sólo una parte del control de la expresión genética en humanos; otros fenómenos que también participan son el "*splicing* alternativo", que es el corte y edición de una molécula de ARN para generar versiones o isoformas de mARN; y el control de la traducción, que es el uso del mARN como molde para generar una proteína.

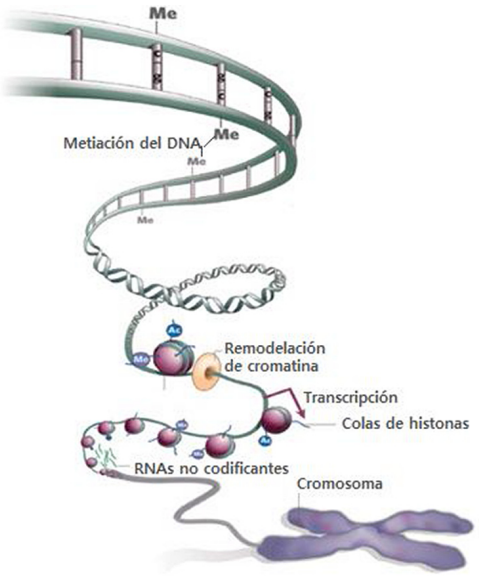

Figura 11. La regulación genética a través de mecanismos epigenómicos. La expresión genética puede ser regulada a través de modificaciones químicas al ADN o a la cromatina, como mediante la metilación del ADN, o la acetilación y la metilación de histonas, así como mediante la acción de los ncARNs.

La epigenómica estudia al conjunto de procesos que mediante modificaciones químicas sobre el ADN y cambios estructurales de la organización cromosómica tienen efecto sobre el control de la expresión genética, y los cuales son heredables aunque no alteren la secuencia genética. Los mecanismos epigenéticos incluyen a la metilación del ADN, la acetilación y metilación de histonas, así como a la actividad de los ncARN (figura 11). La regulación epigenética ocurre a diferentes niveles estructurales, afectando la posición de los cromosomas, los dominios subcromosómicos y las regiones genéticas (*loci*) dentro del espacio tridimensional del núcleo celular, lo cual es un importante factor en la regulación genética (Nunez, Fu, and Rosenfeld, 2009). La organización espacial del genoma se encuentra frecuentemente alterada en escenarios de enfermedad, como se ha podido evidenciar en el cáncer (Meaburn, *et al.*, 2009) y la epilepsia (Borden, and Manuelidis, 1988), y una gama de eventos epigenéticos se han identificado como relevantes en el envejecimiento (Calvanese, *et al.*, 2009) y en diferentes procesos de enfermedad (Jirtle, and Skinner, 2007). En el universo de eventos epigenéticos hay aquellos que son heredados (ej. los genes improntados), y otros adquiridos (ej. los inducidos en ciertos tejidos por efecto de la dieta); la frontera entre ambos hoy en día aún no es clara y la proporción de unos y otros es variable dependiendo de la especie, el tejido, la edad, el sexo, etc. (Petronis, 2010).

Debido a que los procesos epigenéticos responden a una gran variedad de estímulos ambientales como dieta, procesos infecciosos y exposición a xenobióticos (ej. fármacos y toxinas), nos remiten a un debate central en la biología: ¿nacemos

o nos hacemos? La existencia de una herencia independiente de la secuencia del ADN, relacionada a cambios moleculares inducidos por las exposiciones ambientales un individuo a las que se enfrenta un individuo desde la vida intrauterina hasta su muerte, y que adicionalmente plantea la existencia de fenotipos perdurables y potencialmente heredables clonal y/o transgeneracionalmente, ha originado que se vuelva a discutir acerca del Lamarckismo (una corriente de estudio que plantea la idea de que un organismo puede heredar a su progenie características adquiridas a lo largo de su vida) (Petronis, 2010). El enorme potencial de que los fenómenos epigenéticos expliquen una parte sustancial de la expresión y heredabilidad de rasgos y enfermedades complejos, ha convertido a la epigenómica en un tema central de investigación en la biomedicina . Lo anterior, por ejemplo ha derivado en el desarrollo de proyectos con el Human Epigenome Project (HEP), que tiene como objetivo identificar, caracterizar e interpretar los patrones de metilación a lo largo de todo el genoma en diferentes tejidos humanos.

Diagnóstico molecular

Aunque el desarrollo e implementación en el laboratorio clínico de métodos de diagnóstico molecular basados en información genética ha incrementado en la última década, se espera que esta área evolucione rápidamente en los siguientes años (Ostrowski, and Wyrwicz, 2009). Antes de la era genómica, el diagnóstico genético humano estaba centrado en la identificación de las mutaciones causantes de enfermedades monogénicas, para lo cual se utilizan ensayos diseñados específicamente para las regiones afectadas con las mutaciones. Actualmente, éstas siguen siendo las pruebas genéticas más comunes, y poco a poco se van incorporado otras pruebas, como algunas relacionadas al análisis de variantes genéticas de relevancia en farmacogenómica, así como pruebas genéticas para la identificación de microorganismos patógenos de importancia clínica (ej. los virus de papiloma humano y de hepatitis C). Hoy en día, hay mucho interés en desarrollar pruebas más informativas basadas en plataformas tecnológicas, que como los microarreglos, detectan simultáneamente decenas a millones de variantes genéticas, o decenas a miles de transcritos de ARN. Estas pruebas se utilizan ya en algunos escenarios clínicos, aunque no de manera generalizada; entre éstas existen algunos ensayos comerciales aprobados para uso diagnóstico por la Food & Drug Administration (FDA) de los EUA, ej. el AmpliChip Cytochrome P450 Genotyping Test (Roche) para pruebas farmacogenómicas, y el MammaPrint® (Agendia) para determinar un perfil de expresión pronóstico de recurrencia y metástasis en cáncer mamario. Tanto en EUA como en otros países existen otras pruebas aún no aprobadas por FDA, que se ofrecen para complementar el diagnóstico clínico, como el Oncotype DX para cáncer mamario en etapa temprana, y varias pruebas desarrolladas por la compañía deCODE Genetics, para estimar riesgos a desarrollar diabetes tipo 2, obesidad, cáncer de mama y próstata, glaucoma exfoliativo, infarto al miocardio, entre otros padecimientos. El uso de microarreglos también se ha extendido exitosamente al área de la citogenética molecular para la determinación con un alto grado de resolución de alteraciones cromosómicas, estructurales y numéricas de importancia clínica, como son el cromosoma Filadelfia en la leucemia mielocítica crónica o la trisomía 21 en Síndrome de Down. En citogenética esta nueva tecnología complementa, y en ocasiones reemplaza técnicas tradicionales, como el análisis de bandeo cromosómico y la hibridación *in situ* con fluorescencia (FISH) (Li, and Andersson, 2009). En un futuro cercano, se espera que microarreglos y otras plataformas genómicas que están en etapa de validación y/o esperando aprobación por la FDA se adopten exitosamente en la práctica clínica.

Una variante muy discutida y controversial son la pruebas genómicas que se ofrecen como un servicio "directo al cliente" o DTC por sus siglas en inglés. Existen tres compañías en EUA que ofrecen este tipo de servicio: 23andMe, Pathway Genomics y Navigenics, y otras más en diversas partes del mundo, como deCODE. Estas empresas ofrecen servicios de genotipificación utilizando microarreglos de alta capacidad, y generan reportes que contienen información sobre riesgos potenciales a padecer algunas enfermedades comunes, en particular aquellas para las que los recientes descubrimientos han generado más evidencia (ej. degeneración macular relacionada a la edad, diabetes tipo 2, artritis reumatoide, enfermedad de Chron, etc.). Es muy importante señalar que estas pruebas calculan riesgos con base en algoritmos estadísticos que usan información de estudios realizados mayoritariamente en población europea, y que generalmente sólo logran determinar un porcentaje pequeño del riesgo total a desarrollar dichos padecimientos (Janssens, *et al.*, 2008). Por lo anterior, hasta ahora estas pruebas carecen de verdadera utilidad clínica, particularmente para poblaciones en las que la asociación de estas variantes, así como el riesgo conferido por las mismas no se ha estimado en estudios epidemiológicos y ensayos clínicos controlados, como es el caso de la población mexicana. Dado su alto costo y poco potencial de utilidad clínica, es muy importante cuestionar y debatir acerca de su aplicación y, asimismo, es fundamental exigir y promover la legislación, regulación y normatividad para éstos y otros productos relacionados.

Nuevas avenidas y perspectivas

Las tecnologías utilizadas para secuenciar el ADN de un individuo se han desarrollado impresionantemente en los últimos 20 años y constituyen uno de los avances tecnológicos más importantes de nuestra era. Hasta hace unos cuantos años prácticamente toda la secuenciación, aun la más sofisticada y automatizada llevada a cabo en el mundo estaba basada en la técnica desarrollada por Sanger en los años 70 (Sanger, Nicklen, and Coulson, 1977). Esta técnica, conocida como "secuenciación dideoxi" aprovecha la estructura química de nucleótidos modificados, conocidos como dideoxinucleótidos (ddNTPs), los cuales son similares a los deoxinucleótidos que conforman al ADN, pero carecen de un grupo hidroxilo en el

carbono 3' de la ribosa. Esta modificación química ocasiona que, al incorporarse un ddNTP a una cadena creciente de ADN durante una síntesis enzimática, ésta se detenga, lo anterior porque carece del radical necesario que se forme el enlace químico (fosfodiéster) entre dos nucleótidos. Los secuenciadores automáticos basados en este método utilizan ddNTPs fluorescentes marcados diferencialmente para determinar la naturaleza y orden de los nucleótidos. Hasta la fecha, esta tecnología sigue siendo ampliamente utilizada y es el estándar de oro de la secuenciación; sin embargo, ha alcanzado su máxima optimización en costo y capacidad, y presenta la limitación de no permitir el análisis de regiones extensas (>1000bp por reacción). Por lo anterior, se han desarrollado métodos conocidos como "de nueva generación", los cuales tienen mucho mayor capacidad, y permiten analizar millones de pares bases (Gb) por reacción, con costos por base decenas de veces menores que la secuenciación tradicional (Shendure, and Ji, 2008). Recientemente se han desarrollado una amplia variedad de aplicaciones basadas en esta tecnología, como la resecuenciación de genomas completos, la caracterización de transcriptomas y el mapeo de regiones de interacción ADN-proteína, entre otras.

En 2003, el J. Craig Venter Institute y el premio Archon X en genómica plantearon otorgar un reconocimiento monetario de más de 10 millones de dólares al primero en lograr secuenciar un genoma humano individual por 1,000 dólares (Gross, 2007); hacia la segunda mitad de 2010 este objetivo aún no se ha logrado; sin embargo, éste pudiera estar más cerca de lo que imaginamos. La velocidad a la que ha avanzado esta tecnología es impresionante, sólo consideremos que el PGH se completó en 2003 después de 13 años con un costo de aproximado de 270 millones de dólares; para 2007 se publicaron las secuencias de dos genomas individuales, el de Craig Venter que se logró en cuatro años por 100 millones de dólares; y la de James Watson obtenida en sólo 4.5 meses por 1 millón de dólares (Wadman, 2008); por último, recientemente, Stephen R. Quake, científico de la Universidad de Stanford publicó la secuenciación de su genoma (Pushkarev, Neff, and Quake, 2009), la cual costó 50,000 dólares (http://news.stanford.edu/news/2009/august10/genome-081009.html).

A la luz de todo el desarrollo científico y tecnológico que ha dado lugar al universo de las ciencias "ómicas": genómica, transcriptómica, proteómica, metabolómica, etc., se vislumbra una era de generación de conocimiento científico que cambiará sustancialmente el escenario de las ciencias biomédicas y la medicina. Es entonces de fundamental importancia que tanto médicos como pacientes tengan acceso a suficiente información sobre los beneficios y limitaciones de este nuevo conocimiento, así como de los productos y servicios derivados del mismo, de tal manera que se promueva un uso lo más racional y ético posible de los mismos. El sueño de la medicina personalizada hoy en día es una realidad únicamente para un número muy limitado de pacientes con ciertas patologías; éste sólo será una realidad para una fracción considerable de la población si se llevan a cabo acciones que aumenten el conocimiento existente y den suficiente evidencia acerca de los beneficios de los recientes hallazgos. Hoy en día es necesario llevar a cabo ensayos clínicos controlados para medir la efectividad clínica de las intervenciones basadas en información genómica, e impulsar el desarrollo de las estrategias basadas en biología de sistemas para buscar entender las complejas redes de interacción entre genoma, transcriptoma y proteoma. Igualmente importante es reflexionar sobre el hecho de que la era genómica nos está enfrentando a una circunstancia en la que la alta capacidad de la tecnología nos lleva fácilmente a encontramos inundados de Giga a Terabytes de información genómica. Lo anterior llega a suceder en situaciones limitadas, en términos de la capacidad computacional y humana para el adecuado análisis e interpretación de esta información. Además, es necesario no perder de vista que por la naturaleza de su diseño, los resultados obtenidos mediante la aplicación de las tecnologías "ómicas" son generadores de hipótesis, lo que significa que generan muchas más preguntas que respuestas concretas, lo que representa un reto en la intención de hacer una adecuada traducción del conocimiento genómico en beneficios al cuidado de la salud humana.

Glosario

Alelo: Una o más versiones de un nucleótido o secuencia de nucleótidos en una región particular del genoma.

ARN no codificante: Segmentos de ARN que no son traducidos a secuencias de proteína, pero que pueden estar relacionados a la regulación de la expresión genética.

Bioactivo nutricional: Compuestos esenciales y no-esenciales que ocurren en la naturaleza y que son parte de los alimentos o suplementos dietéticos, que proveen beneficios al estado de salud, distintos a los valores nutricionales básicos de los mismos.

Biomarcador: Una característica biológica o química que puede ser medida y evaluada en materiales biológicos de manera objetiva para ser utilizada como indicador de procesos biológicos normales, procesos patológicos o exposiciones.

Desequilibrio de ligamiento: Asociación no azarosa de alelos genéticos en dos o más loci en una población. Se detecta cuando ciertas combinaciones de marcadores genéticos ocurren más o menos frecuentemente en una población que lo que se espera si se considera una formación azarosa de haplotipos a partir de las frecuencias alélicas de esos marcadores.

Epigenómica: Disciplina que estudia el conjunto las modificaciones químicas sobre el ADN y los cambios estructurales de la organización cromosómica, que tienen efecto sobre el control de la expresión genética, y que son heredables aunque no alteran la secuencia genética primaria.

Monogénico(a): Rasgo o enfermedad controlado por un solo gen y que puede verse afectado por mutaciones en el mismo.

Genotipificar: Determinar con una prueba de laboratorio el alelo o la combinación de alelos que un individuo presenta en su genoma.

Haploide: Número de cromosomas presentes en una célula sexual.

Haplotipo: Conjunto de variaciones en la secuencia de ADN que tienden a heredarse en conjunto. Se refiere a una combinación específica de alelos que se encuentran en el mismo cromosoma.

Heredabilidad: Proporción de la variación fenotípica que se atribuye a la variabilidad genética entre individuos. La variabilidad fenotípica puede deberse al efecto de factores genéticos y/o ambientales; entonces esta medida estima las contribuciones relativas de las diferencias entre factores genéticos y no genéticos a la varianza fenotípica total en una población.

Histona: Proteína básica rica en arginina y lisina presente en el núcleo de células eucariotas, que forma octámeros alrededor de los cuales se enrolla el ADN para dar lugar a los nucleosomas que a su vez forman una súper estructura denominada cromosoma.

P o valor de p: también se conoce como error tipo I o error alfa, es una medida de significancia estadística que corresponde a la probabilidad de rechazar la hipótesis nula cuando ésta es verdadera, es decir la probabilidad de que el resultado sea un "falso positivo".

Rasgo/Condición complejo: Característica de un ser vivo que resulta de la interacción entre múltiples genes, y entre factores genéticos y ambientales.

Secuencias motivo: región en el ADN para la que se asume un significado biológico definido.

SNP etiqueta: Variante genética que se encuentra en alto desequilibrio de ligamiento (en correlación) con otras y que sirve de aproximación para éstos cuando se utiliza en una plataforma de genotipificación.

Xenobióticos: Compuestos químicos presentes en un organismo que: a) no son producidos por el mismo, b) que normalmente no se espera que se encuentren o c) que se encuentran en cantidades muy por encima de lo esperado.

Ligas de interés

PGH: http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml
GWAS catalog: http://www.genome.gov/gwastudies/
HEP: http://www.epigenome.org/
NCBI: http://www.ncbi.nlm.nih.gov/
Ensembl: http://www.ensembl.org/
UCSC: http://genome.ucsc.edu/
HapMap: www.hapmap.org
HuRef: http://huref.jcvi.org/
23andMe: http://www.23andme.com/
DeCODE genetics: http://www.decode.com/
Navigenics: http://www.navigenics.com/

Bibliografía

A haplotype map of the human genome, *Nature*, **437** (7063), 1299-320, 2005.

Anderson, C.A., *et al.*, Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms, *Am J Hum Genet*, **83**(1) 112-9, 2008.

Balding, D. J., A tutorial on statistical methods for population association studies, *Nat Rev Genet*, **7**(10) 781-91, 2006.

Baty, F., *et al.*, *Gene profiling of clinical routine biopsies and prediction of survival in non-small cell lung cancer*, Am J *Respir Crit Care Med*, **181**(2) 181-8, 2010.

Birney, E., *et al.*, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature*, **447**(7146) 799-816, 2007.

Borden, J. and L. Manuelidis, Movement of the X chromosome in epilepsy, *Science*, **242**(4886) 1687-91, 1988.

Calvanese, V., *et al.*, The role of epigenetics in aging and age-related diseases, *Ageing Res Rev*, **8**(4) 268-76, 2009.

Gatza, M.L., *et al.*, A pathway-based classification of human breast cancer, *Proc Natl Acad Sci USA*, **107**(15) 6994-9, 2010.

Geschwind, D.H. and G. Konopka, Neuroscience in the era of functional genomics and systems biology, *Nature*, **461** (7266) 908-15, 2009.

Gonzalez-Angulo, A.M., B.T. Hennessy, and G.B. Mills, Future of personalized medicine in oncology: a systems biology approach, *J Clin Oncol*, **28**(16) 2777-83, 2010.

Gross, L., A new human genome sequence paves the way for individualized genomics, *PLoS Biol*, **5**(10) e266, 2007.

Ehret, G.B., Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension, *Curr Hypertens Rep*, **12**(1) 17-25, 2010.

Elsik, C. G., *et al.*, The genome sequence of taurine cattle: a window to ruminant biology and evolution, *Science*, **324**(5926) 522-8, 2009.

Feero, W. G., A. E. Guttmacher, and F. S. Collins, Genomic medicine–an updated primer, *N Engl J Med*, **362**(21) 2001-11, 2010.

Feuk, L., A. R. Carson, and S. W. Scherer, Structural variation in the human genome, *Nat Rev Genet*, **7**(2) 85-97, 2006.

Finishing the euchromatic sequence of the human genome, *Nature*, **431**(7011) 931-45, 2004.

Frazer, K. A., *et al.*, Human genetic variation and its contribution to complex traits, *Nat Rev Genet*, **10**(4) 241-51, 2009.

Genome sequencing and analysis of the model grass Brachypodium distachyon, *Nature*, **463**(7282) 763-8, 2010.

Gerstein, M. B., *et al.*, What is a gene, post-ENCODE? History and updated definition, *Genome Res*, **17**(6) 669-81, 2007.

Ioannidis, J. P., G. Thomas, and M. J. Daly, Validating, augmenting and refining genome-wide association signals, *Nat Rev Genet*, **10**(5) 318-29, 2009.

Jakobsson, M., et al., Genotype, haplotype and copy-number variation in worldwide human populations, *Nature*, **451**(7181) 998-1003, 2008.

Janssens, A.C., *et al.*, A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions, *Am J Hum Genet*, **82**(3) 593-9, 2008.

Jirtle, R.L. and M.K. Skinner, Environmental epigenomics and disease susceptibility, *Nat Rev Genet*, **8**(4) 253-62, 2007.

Katta, S., I. Kaur, and S. Chakrabarti, The molecular genetic basis of age-related macular degeneration: an overview, *J Genet*, **88**(4) 425-49, 2009.

Levy, S., *et al.*, The diploid genome sequence of an individual human, *PLoS Biol*, **5**(10) e254, 2007.

Li, M.M. and H.C. Andersson, Clinical application of microarray-based molecular cytogenetics: an emerging new era of genomic medicine, *J Pediatr*, **155**(3) 311-7, 2009.

Malecova, B. and K.V. Morris, Transcriptional gene silencing through epigenetic changes mediated by non-coding ARNs, *Curr Opin Mol Ther*, **12**(2) 214-22, 2010.

Manolio, T.A., Genomewide association studies and assessment of the risk of disease, *N Engl J Med*, **363**(2) 166-76, 2010.

McCarthy, M.I., *et al.*, Genome-wide association studies for complex traits: consensus, uncertainty and challenges, *Nat Rev Genet*, **9**(5) 356-69, 2008.

Meaburn, K.J., *et al.*, Disease-specific gene repositioning in breast cancer, *J Cell Biol*, **187**(6) 801-12, 2009.

Mook, S., *et al.*, Metastatic potential of T1 breast cancer can be predicted by the 70-gene MammaPrint signature, *Ann Surg Oncol*, **17**(5) 1406-13, 2010.

Nunez, E., X.D. Fu, and M.G. Rosenfeld, Nuclear organization in the 3D space of the nucleus-cause or consequence?, *Curr Opin Genet Dev*, **19**(5) 424-36, 2009.

Ostrowski, J. and L.S. Wyrwicz, Integrating genomics, proteomics and bioinformatics in translational studies of molecular medicine, *Expert Rev Mol Diagn*, **9**(6) 623-30, 2009.

Petronis, A., Epigenetics as a unifying principle in the aetiology of complex traits and diseases, *Nature*, **465**(7299) 721-7, 2010.

Pushkarev, D., N.F. Neff, and S.R. Quake, Single-molecule sequencing of an individual human genome, *Nat Biotechnol*, **27**(9) 847-52, 2009.

Ramsey, S.A., E.S. Gold, and A. Aderem, A systems biology approach to understanding atherosclerosis, *EMBO Mol Med*, **2**(3) 79-89, 2010.

Sanger, F., S. Nicklen, and A.R. Coulson, ADN sequencing with chain-terminating inhibitors, *Proc Natl Acad Sci U S A*, **74**(12) 5463-7, 1977.

Santos, E.S., M. Blaya, and L.E. Raez, Gene expression profiling and non-small-cell lung cancer: where are we now?, *Clin Lung Cancer*, **10**(3) 168-73, 2009.

Shendure, J. and H. Ji, Next-generation ADN sequencing, *Nat Biotechnol*, **26**(10) 1135-45, 2008.

Silva-Zolezzi, I., *et al.*, Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico, *Proc Natl Acad Sci USA*, **106**(21) 8611-6, 2009.

Stankiewicz, P. and J. R. Lupski, Structural variation in the human genome and its role in disease, *Annu Rev Med*, **61** 437-55, 2010.

Straver, M. E., *et al.*, The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer, *Breast Cancer Res Treat*, **119**(3) 551-8, 2010.

Takizawa, T., et al., Basic and clinical studies on functional ARN molecules for advanced medical technologies, *J Nippon Med Sch*, **77**(2) 71-9, 2010.

Teo, Y. Y., *et al.*, Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations, *Genome Res*, **19**(11) 2154-62, 2009.

The International HapMap Project, *Nature*, **426** (6968) 789-96, 2003.

Tian, C., *et al.*, Analysis and application of European genetic substructure using 300 K SNP information, *PLoS Genet*, **4**(1) e4, 2008.

van de Vijver, M. J., *et al.*, A gene-expression signature as a predictor of survival in breast cancer, *N Engl J Med*, **347**(25) 1999-2009, 2002.

Wadman, M., James Watson's genome sequenced at high speed, *Nature*, **452**(7189) 788, 2008.