Gaceta Mexicana de Oncología. 2013;**12(6)**:426-432

SOCIEDAD MEXICANA DE ONCOLOGÍA, A.C.
GACETA MEXICANA DE ONCOLOGÍA
www.elsevier.es

ARTÍCULO DE REVISIÓN

Organización estructural y funcional del genoma humano: variación en el número de copias predisponentes de enfermedades degenerativas

Víctor Manuel Valdespino-Gómez[a,*], Patricia Margarita Valdespino-Castillo[b] y Víctor Edmundo Valdespino-Castillo[c]

[a] *Departamento de Atención a la Salud, Universidad Autónoma Metropolitana-Unidad Xochimilco, México D.F., México*
[b] *Instituto de Ecología, Universidad Nacional Autónoma de México, México D.F., México*
[c] *Unidad Médica de Atención Ambulatoria, Instituto Mexicano del Seguro Social, Campeche, Camp., México*

PALABRAS CLAVE
Organización estructural y funcional del genoma humano; Variaciones en el número de copias; Enfermedades degenerativas; México.

Resumen Los procesos de salud y enfermedad humana pueden ser analizados desde niveles macroscópicos, microscópicos y nanóscopicos. El avance en los diferentes aspectos del conocimiento biológico molecular de finales del siglo XX y principios del siglo XXI permiten progresar en su análisis.

Los proyectos Genoma Humano, HapMap y 1000 Genomas han avanzado en el análisis estructural del genoma humano. Recientemente, el Proyecto del Consorcio Enciclopedia de los Elementos del DNA (ENCODE) ha iniciado el análisis de los principales componentes funcionales relacionados con el estudio del genoma humano, explorando las variantes genómicas y epigenómicas que se presentan en la mayoría de tipos celulares del organismo humano. Los avances en el entendimiento de la regulación de la expresión génica logradas por ENCODE se encuentran revolucionando la comprensión de los procesos moleculares de la salud y de muchas enfermedades.

En este artículo abordamos los principales conceptos de la organización estructurales del genoma humano y los primeros conceptos modernos de su organización funcional en condiciones de salud, y comentamos algunas de las principales alteraciones en el número de copias de regiones particulares del genoma, que predisponen a algunas enfermedades degenerativas complejas.

* Autor para correspondencia: Andrés Molina Enríquez N° 361, Colonia Ampliación Sinatel, C.P. 09479. Delegación Iztapalapa, México D.F., México. Teléfono: (55) 5674 3439. *Correo electrónico*: vvaldespinog@yahoo.com.mx (Víctor Manuel Valdespino-Gómez).

KEYWORDS
Structural and functional genome organization; Copy number variations; Degenerative diseases; Mexico.

Human genomic structural and functional alterations: copy number variation in degenerative diseases predisposition

Abstract Biological processes in human health and disease can be studied by macroscopic, microscopic and nanoscopic scales. The scientific biological advances obtained at the end of XX century and at the beginning XXI century allow their analysis progress.
Human Genome, HapMap and 1000 Genomas Projects have advanced in the structural analysis of the human genome. Recently the Encyclopedia of DNA Elements Project has initiated the analysis of the main human genome functional components, exploring the genomic and epigenomic variants that take place in most cellular types of the human body. The progress in elucidating the regulation of gene expression, obtained by ENCODE Project, are shaking up the understanding of the molecular process involved in human health and disease.
Here, we describe the main concepts of genome structure and the modern concepts related to functional genome organization, and review some of the main copy number variations that predispose some complex genetic diseases.

Introducción

El estudio de las diferentes razas humanas relacionadas con indicadores de salud, demuestra que existen diferencias genéticas y epigenéticas en sus poblaciones. Los datos del análisis del genoma humano nos permiten entender la variabilidad genómica en las poblaciones y facilitan las vías para los avances en el progreso de la medicina y la biotecnología. Los análisis comparativos de los genomas completos de diferentes poblaciones humanas y de especies cercanas, han identificado dos grandes grupos de variantes genéticas relacionadas con la enfermedad, las primeras relacionadas con enfermedades monogénicas/cromosómicas por variantes genéticas raras que afectan predominantemente una característica fenotípica (por ejemplo, fibrosis quística, enfermedad de Huntington), y el segundo grupo, enfermedades poligénicas multifactoriales más frecuentes, generalmente enfermedades crónico-degenerativas, que implican complejas características fenotípicas (por ejemplo, diabetes, cáncer, enfermedades cardiacas, enfermedades autoinmunes).

Los diferentes estudios de asociación pangenómica de las variantes genéticas estructurales relacionadas con enfermedades poligénicas multifactoriales han logrado solo un éxito limitado en su entendimiento, particularmente porque además de la obtención de la información estructural al secuenciar el genoma humano, se requiere entender la organización funcional y la regulación de la expresión génica.

En el campo de la genética humana, los estudios biológicos experimentales de las últimas 2 décadas del siglo XX y de la primera del presente, han logrado 3 avances puntuales técnico-conceptuales importantes: la factibilidad de estudiar los genomas completos de diversos organismos a través del uso de micromatrices o microarreglos de DNA (los cuales pueden rastrear cientos de miles de reacciones moleculares en paralelo para detectar conjuntos de genes específicos o para medir la actividad génica en diferentes células), la identificación del código epigenético (cambios químicos menores del DNA que son provocados por los señalamientos del entorno extracelular) que modula la expresión genética[1], y muy recientemente la integración de los códigos genético y epigenético humanos para entender más profundamente la regulación de la expresión genómica.

Avances en el estudio de la genómica estructural

En el año 2003, el Proyecto Genoma Humano (HGP) después de 14 años de investigación publicó el primer borrador de la secuencia nucleotídica del DNA humano (secuencia del 97% de la eucromatina del genoma humano haploide). En esa fecha, el estudio del HGP cubrió incompletamente el análisis de la totalidad del DNA, ya que algunas regiones de la heterocromatina no fueron analizadas (aproximadamente el 8% del genoma total), y fue el siguiente año cuando se completó el estudio de la secuencia en un 99.99%. Cuatro años después, Venter y colaboradores publicaron la secuencia del genoma diploide de un individuo[2]. La secuencia del genoma humano permitió la expansión de diferentes campos de la medicina molecular y mejoró el entendimiento de la evolución humana. La secuencia del DNA humano ha sido guardada en bases de datos disponibles para consulta en Internet, y el análisis global para identificar, por ejemplo, regiones codificantes (genes) o no-codificantes, requiere el empleo de herramientas bioinformáticas.

La Bioinformática (la unión entre la informática y la biología) como disciplina, permite la organización y el análisis de megadatos genómicos, su aplicación inicial generó la base de datos denominada *GenBank*, y posteriormente centenares de bases de registro, análisis y predicción, las cuales permiten realizar asociaciones y correlaciones. Como veremos adelante la generación de una gran variedad de determinaciones de variables genéticas y epigenéticas en condiciones de salud y enfermedad, hacen que el análisis bioinformático se haya convertido en una herramienta indispensable para establecer asociaciones de variables genómicas con característica celulares fenotípicas y permitir predicciones en los fenómenos biológicos.

Todos los seres humanos contienen secuencias genómicas únicas, por lo que la secuencias publicadas por el HGP no representan la secuencia exacta del genoma de cada

individuo, sino más bien corresponden a la referencia de un grupo pequeño de donadores anónimos. La información obtenida por el HGP ha logrado avances notables en el inicio del entendimiento molecular de la diversidad genética, Una vez completada la secuenciación del genoma humano por el HGP, los siguientes estudios de la variación del DNA humano continuaron a través del *HapMap Project* (HMP), el cual determinó mapas de haplotipos o "haps" asociados a diferentes poblaciones (grupos de secuencias génicas adyacentes), por medio de la identificación metodológica de polimorfismos de un solo nucleótido o *single-nucleotide polymorphism* (SNPs), junto con algunos haplotipos relacionados al riesgo mayor o menor de desarrollar algunas enfermedades crónico-degenerativas. La denominación de SNPs corresponde a variaciones en las secuencias del DNA que se presentan en aproximadamente en el 1% de la población, mientras que el concepto de "mutaciones", corresponde a las variaciones que se presentan con absoluta menor frecuencia. La mayoría de estos SNPs corresponden a variaciones neutrales y sólo el 1% de ellos, presentan repercusión funcional. Los diferentes patrones de polimorfismos de DNA humano se pueden asociar con variaciones en la resistencia a enfermedades específicas y a respuestas metabólicas diferentes a medicamentos específicos. El proyecto *HapMap* a través de 3 fases de estudio (2002-2009), determinó los principales haplotipos entre los genomas de poblaciones europeas, asiáticas y africanas, identificados a partir de patrones de grupos de SNPs. Más tarde, el proyecto *1000 Genomes* (2008) obtuvo un catálogo más detallado de las variaciones genéticas humanas a partir del análisis de 1,000 donadores anónimos de diferentes grupos étnicos (variaciones estructurales, variaciones en el número de copias, retroelementos, SNPs, etc.).

Los haplotipos, las mutaciones, los SNPs y otras variantes estructurales en el genoma humano, como el número de copias de segmentos o regiones específicas contribuyen a la diversidad genética en las poblaciones humanas, y algunas de ellas, se han asociado a diferentes enfermedades. Todas estas variantes modifican estructuralmente las regiones codificantes (de proteínas) y las regiones no-codificantes. La continuación de los estudios del genoma humano han sido coordinados por el *Encyclopedia of DNA Elements Project* (ENCODE). Recientemente, un numeroso grupo de investigaciones coordinadas por el Consorcio Enciclopedia de los Elementos del DNA, destinadas a encontrar todos los elementos funcionales en el genoma humano, publicaron sus resultados en 30 artículos simultáneos[3]. El Consorcio ENCODE, empleó diferentes tecnologías para la identificación de cambios genéticos y epigenéticos que funcionan como marcas regulatorias de la expresión génica (reguloma), y demostraron que las regiones no-codificantes del DNA humano (más del 80% del genoma total) participan en la regulación funcional de la expresión génica, y que esta regulación se encuentra controlada por múltiples sitios localizados cerca y distante de las regiones codificadoras del gen. Las variaciones en estos sitios se asocian al desarrollo de diferentes enfermedades[3-6].

1.1 Organización estructural del genoma humano

El genoma humano está constituido por una secuencia de 3.2×10^9 nucleótidos, organizado y compactado en 23 pares de cromosomas (22 pares de autosomas y 2 cromosomas sexuales). El genoma humano globalmente puede ser analizado a partir de sus elementos estructurales y de sus elementos funcionales. Los primeros corresponden por ejemplo a regiones codificantes de proteínas y a regiones no-codificantes; los segundos a los componentes que participan en interacciones, regulación y función biológica. Dentro del genoma humano se encuentran 23,000 genes o regiones codificantes de proteínas (similar al número de genes en otros mamíferos y en algunas plantas); constituidos por exones e intrones en una proporción de secuencias de 1:24). El genoma humano a diferencia de genomas de otros organismos cuenta con mayor cantidad de segmentos o regiones duplicadas; las secuencias repetitivas del genoma humano corresponden a diferentes tipos de regiones: transposones, regiones intergénicas, seudogenes, repeticiones cortas, repeticiones largas y repeticiones en tándem.

La larga cadena de 3.2 meganucleótidos o megabases del DNA humano se encuentra enrollada por un complejo de proteínas organizadas, que en conjunto constituyen la cromatina. Los componentes de esta larga cadena de DNA del genoma humano pueden ser estudiados a través de diferentes enfoques organizacionales: conformando cromosomas, en regiones DNA codificantes y no-codificantes, en zonas constitutivas de genes y seudogenes, etc.

El DNA humano cuenta con regiones no repetitivas donde se localizan los genes que codifican proteínas y los genes que codifican diferentes RNAs "clásicamente" funcionales (por ejemplo, RNA ribosomal, RNA de transferencia). Sólo el 1% del total del genoma humano, corresponde a regiones exónicas, mientras que el 24% lo abarcan las regiones intrónicas. El DNA que codifica proteínas es el componente más ampliamente estudiado del genoma humano. El DNA codificante está constituido por 20,687 genes aproximadamente; en promedio de 500 a 1,000 genes se ubican en cada cromosoma; un gen en promedio está constituido por 10 a 50 kilonucleótidos o kilobases (kn o kb), pero entre ellos existe gran variabilidad. Las regiones codificantes de los genes corresponden a los exones. En el genoma humano se han identificado 180,000 exones (aproximadamente 10 exones por cada gen), los cuales en conjunto corresponden aproximadamente a 30 megabases de extensión (1% del genoma). Los intrones (regiones intragénicas no-codificantes) corresponden al 24% del genoma humano, son secuencias de DNA cuya longitud es de 10 a 100 veces mayor comparativamente al de los exones; un tipo de ellos son los 5'-UTRs o los 3'-UTRs. Vecinos a los exones/intrones se han identificado a las regiones no-codificantes conocidas como reguladoras (8%-20% del genoma humano) -aunque la identificación de éstas como veremos adelante, se ha ampliado significativamente de acuerdo a los estudios de ENCODE-. Los grupos de proteínas con mayores porcentajes de regiones codificantes dentro del exoma (todos los exones del genoma), son los factores de transcripción (2,067 correspondiente al 12%), las transferasas (8.8%), otras proteínas que se unen al DNA (8.5%), proteínas transportadoras (6.4%), proteínas receptoras (6.3%), moléculas de señalamientos (5.6%), y enzimas moduladoras (5%); muchas proteínas no han sido clasificadas (4061 que corresponden al 23.6%). La secuenciación nucleotídica del exoma identifica las variantes de las secuencias codificantes de proteínas (por ejemplo, mutaciones autosómicas o recesivas de genes)[7]; pero este tipo de se-

cuenciación es sólo una pequeña parte de la secuenciación del genoma completo.

Las regiones repetitivas del genoma (DNA-RR) abarcan el 98% de su totalidad, e incluyen regiones de RNA no-codificantes, seudogenes, intrones, regiones no traducidas del RNAm (UTRs), secuencias regulatorias, secuencias repetitivas (intergénicas) y secuencias relacionadas con elementos movibles o transposones. Globalmente, las regiones de DNA-RR corresponden a regiones denominadas pseudogenes (aproximadamente 14,000) y a regiones relacionadas con RNAs no-codificantes y micro/mini RNAs (aproximadamente 18,400). Las regiones del DNA-RR dentro del genoma humano pueden distribuirse en conjuntos o tándems, y en repeticiones interpuestas entre regiones definidas como genes o exones/intrones y seudogenes. El 8% aproximadamente del genoma humano consiste en secuencias repetitivas de DNA denominadas repeticiones en tándem; éstas son muy variables dentro de los individuos (se emplean como marcadores en análisis forense del DNA); las secuencias repetitivas de menos de 10 nucleótidos se denominan microsatélites, y aquellas entre 10-60 nucleótidos de longitud se denominan minisatélites[8].

Los transposones son el componente más grande del genoma humano (45%), los cuales son elementos movibles dentro del DNA y corresponden a secuencias que se pueden replicar e insertar copias de ellos en diferentes localizaciones del genoma ("genes saltarines"), algunos de ellos parecen retrovirus endógenos, los cuales se encuentran integrados permanentemente y son heredados en la duplicación celular. Existen 2 tipos de transposones, los de clase 1 o retrotransposones y los de clase 2 o DNA-transposones. Los retrotransposones en sus procesos de duplicación y desplazamiento a otros sitios del genoma son inicialmente transcritos a RNA, y pueden ser clasificados en repeticiones terminales largas (LTRs) y repeticiones terminales no-largas (ambos corresponden al 8.3% del genoma, en elementos intercalados largos (LINEs) que corresponden a 20% del genoma humano, y en elementos intercalados cortos (SINEs) que corresponden al 13% del genoma humano). Los DNA-transposones son más escasos (3%) y no emplean al RNA como elemento intermediario para duplicarse y desplazarse. El retrotransposon *Alu* es el transposón más frecuente intercalado en el genoma humano, cuenta con 300 pb y corresponde aproximadamente al 11% de este (más de un millón de repeticiones), recientemente se identificó que participa en el direccionamiento ribosomal de las proteínas. Como es conocido, la transferencia de genes *ex vivo* se realiza frecuentemente empleando sistemas de virus recombinantes, sin embargo algunos sistemas que emplean transposones pueden ser más eficientes en condiciones particulares (por ejemplo, el sistema *Sleeping Beauty* en linfocitos T)[9].

Los seudogenes son secuencias de DNA, relacionados o parecidos con genes conocidos, los cuales han perdido su capacidad para codificar proteínas por la acumulación de mutaciones múltiples. Los seudogenes son generalmente no funcionales, y forman los "fósiles genómicos" (genes de nuestros ancestros, que han sido silenciados); por lo que parecen servir como material conservado del proceso evolutivo del individuo.

Los RNAs no-codificantes (RNAsnc) participan en el procesamiento del RNAm y en la regulación de la síntesis de las proteínas, el genoma humano contiene aproximadamente 7,000 regiones de RNAsnc.

Los genomas humanos entre los individuos varían en su secuencia nucleotídica en menos de 0.1%. La complejidad de la organización del genoma humano ha sido mejor entendida gracias al progreso de las técnicas en el análisis de la secuenciación del DNA. En las últimas décadas los secuenciadores de DNA de segunda generación (empleando tubos capilares) y tercera generación (secuenciación masiva en paralelo), han favorecido la secuenciación rápida (billones de nucleótidos por semana) de grandes segmentos de DNA y RNA y su costo ha disminuido (se calcula que para el año 2014, el costo de la secuenciación del genoma humano será de 1,000 dólares). Además de lograrse secuenciaciones nucleotídicas más rápidas y económicas, han sido desarrolladas otras técnicas que complementan la exploración genómica, entre ellas los microarreglos de DNA y RNA, la secuenciación empleando SNPs (que exploraran marcadores polimórficos y no polimórficos a resoluciones de 10-20 kb), la hibridación genómica comparativa (CGH) y el cariotipo virtual. Las diferentes sondas en estos ensayos (por ejemplo, cDNA, BAC clones, oligonucleótidos) pueden identificar regiones genómicas asociadas e involucradas en enfermedades específicas, determinar el número de copias de regiones de DNA (sondas no-polimórficas) e incluso identificar condiciones como de pérdida de la heterocigosidad (LOH) de los alelos.

El DNA no-codificante varía grandemente entre los genomas de las especies. En los organismos eucariontes la proporción de DNA no-codificante comparativamente a la del DNA-codificante es absolutamente mayor, así en el genoma humano la proporción es de 98:2. Mucho de este DNA no-codificante participa en la regulación transcripcional y traduccional de los genes que codifican proteínas. Varias de las secuencias no-codificantes determinan el sitio donde los factores de transcripción se unen al DNA-codificante. Las principales funciones de las secuencias no-codificantes de DNA corresponden a regiones que codifican RNAs (son regiones no-codificantes para proteínas, pero sí para RNA no-codificante). Los RNAsnc participan principalmente en la regulación génica, de ellos, probablemente los más importantes son los microRNAs (los cuales participan en la regulación de traducción del 30% de los genes que codifican proteínas en los mamíferos).

Las regiones que codifican proteínas, RNA, intrones, regiones reguladoras, y regiones no-codificantes contienen zonas conservadas en el tiempo (millones de años) relacionadas con la presión evolutiva y la selección positiva; muchas de regiones de DNA-RR regulan la estructura de los cromosomas o algunas de sus áreas (telómeros y centrómeros) y corresponden a regiones de origen de la replicación del DNA; también regulan la estabilidad de los RNAm, la estructura de la cromatina (modificaciones postraduccionales de las histonas y metilación del DNA) y la recombinación del DNA.

1.2 Variación en el número de copias de regiones particulares en el genoma humano

La variación en el número de copias de regiones particulares (CNVs) en el genoma humano corresponde a una variación estructural de escala intermedia, que consiste en un

número anormal de una o más secciones del DNA. La información genética humana es diploide, contiene 2 copias en los cromosomas (a excepción de los cromosomas X y Y en los varones). Los tipos de CNVs más frecuentes en una o varias regiones del DNA comprenden a deleciones, duplicaciones, inversiones y translocaciones. Estas variaciones oscilan en tamaño, de pocos cientos de bases a decenas de megabases (frecuentemente alrededor de 1,000 bp), las cuales conducen a rearreglos genómicos (dependiendo del tamaño, orientación, porcentaje, etc.). Particularmente las CNVs mayores a 10 Mgbp se denominan variaciones de escala mayor (por ejemplo, la trisomía del cromosoma 21 que supera a 100 Mgbp) y las CNVs menores a 1 kb, que se denominan *indels* (insersiones-deleciones). Las CNVs pueden estar limitadas a un gen o a un conjunto de genes contiguos, lo cual influye en la variabilidad fenotípica, y en la susceptibilidad de diversas enfermedades. El aumento en el número de copias de un gen particular, incrementa la expresión de la proteína que codifica, como sucede con el gen de la quimiocina *CCL3L1*, que participa en la inmunoregulación de la respuesta inflamatoria, y que se asocia con menor susceptibilidad para la infección con HIV; por el contrario un bajo número de copias del gen *FCGR3B*, receptor de membrana de neutrófilos, células NK y macrófagos, puede aumentar la susceptibilidad a desarrollar lupus eritematoso sistémico. El efecto funcional del CNVs depende del fenotipo y contexto celular y de las condiciones ambientales que rodean al individuo. Se ha estimado que en el genoma humano de individuos no relacionados, las CNVs se presentan en el 0.4%[10].

1.3 Variación en el número de copias en condiciones evolutivas adaptativas y en enfermedades del desarrollo y degenerativas

Las CNV pueden ser heredadas o ser causadas por mutación *de novo*. Generalmente, las CNVs pueden ser causadas por rearreglos estructurales del genoma, como deleciones, duplicaciones, inversiones y traslocaciones; las cuales se producen durante la meiosis por recombinación homóloga no-alélica, o por alteraciones en la replicación o reparación del DNA en la unión de extremos no-homólogos.

Algunos tipos de CNVs pueden corresponder a condiciones evolutivas adaptativas, por ejemplo en el aumento de más de 6 copias del gen de la amilasa *AMY1* en las células de las glándulas salivales de los humanos, lo cual mejora la digestión de los alimentos compuestos por polisacáridos, comparado con la presencia de dos copias en chimpancés. Otro ejemplo es la CNVs de los genes de la ß-defensinas, las cuales funcionan como antibióticos naturales en la piel.

Las CNVs en diferentes sitios del genoma de células germinales y somáticas humanas favorecen la susceptibilidad para desarrollar enfermedades complejas como el autismo, la esquizofrenia, la degeneración macular de la retina, la enfermedad de Crohn, el lupus eritematoso sistémico, la esclerosis lateral amiotrófica y la infección por HIV-SIDA. Estas diferentes alteraciones en CNVs se asocian entre un 10% a 50% de la predisposición a desarrollar diversas enfermedades del neurodesarrollo, neurodegenerativas y autoinmunes, principalmente[10-13]. En algunos tipos de cánceres, las CNVs se asocian a la iniciación y progresión tumoral, particularmente en algunos tipos de leucemias y linfomas y en algunos tumores sólidos[14,15].

La identificación en el cambio del número de copias de segmentos de DNA (deleciones, ganancias, o amplificaciones) puede ser detectada en los ensayos de exploración genómica empleando sondas no-polimórficas. Ejemplos de alteraciones en la CNVs son la inactivación por deleción o por LOH de los genes supresores tumorales y la amplificación (ganancia en más de 10 veces) de algunos oncogenes en células tumorales (por ejemplo, Her2Neu).

Los métodos clásicos que exploran la variación del número de copias de regiones del DNA en el genoma humano corresponden a los patrones denominados "huella génica" por determinaciones de microsatélites y minisatélites (*DNA fingerprinting*), por hibridación *in situ* con sondas fluorescentes, y CGH. La exploración moderna de las CNVs, tanto en regiones codificantes y no-codificantes, se realiza por estudios de genotipificación/secuenciación de "siguiente generación", como polimorfismos de fragmentos obtenidos por corte de enzimas de restricción, polimorfismos de fragmentos amplificados al azar, utilización de oligonucleótidos alelo-específicos, hibridación del DNA en microarreglos o empleando perlas nanométricas y por cariotipo virtual[11,16,17]. La diversidad en el número de copias de distintas regiones del DNA comprende el 12% de la variabilidad del genoma humano[18].

Avances en el estudio de la genómica funcional

La genómica funcional se enfoca a los aspectos dinámicos de la transcripción, traslación e interacciones proteína-proteína; su meta es entender las relaciones entre el genoma de un organismo con su fenotipo. Los microarreglos de RNA y el análisis serial de expresión génica (SAGE) permiten la identificación de la variación comparativa de los transcritos en distintos tipos celulares y en diferentes condiciones de un mismo tipo celular.

Las grandes funciones del DNA en el genoma son la replicación y la transcripción; sin embargo cada una de ellas se encuentra regulada por mecanismos complejos. En un organismo, la transcripción, la traducción y el funcionamiento de sus proteínas determinan su fenotipo. El genoma como el mayor elemento jerárquico en estos procesos, se encuentra localmente programado para la sobrevivencia celular, sin embargo los señalamientos externos son elementos suficientes para modificar puntualmente sus tiempos y formas de expresión génica. La respuesta a los señalamientos externos es codificada por patrones diferenciales de proteínas (frecuentemente factores de transcripción) que interaccionan con secuencias de DNA directa o indirectamente y modifican la activación de los genes codificantes de proteínas o de RNAs.

El código epigenómico es un regulador importante de la expresión génica. Las tecnologías en la exploración de los cambios moleculares epigenéticos del DNA han progresado marcadamente en las últimas 2 décadas, como es el caso de la inmunoprecipitación de la cromatina acoplada a la secuenciación del DNA[1]. Como hemos mencionado el genoma humano contiene muchas secuencias reguladoras de la expresión génica que conforman aproximadamente el 80% del

DNA total. La regulación de la expresión génica es mediada por códigos epigenéticos representados por modificaciones no covalentes de las histonas (por ejemplo, acetilación en sus colas N-terminales), y de patrones de metilación del DNA (de regiones promotoras de genes), que modifican el empaquetamiento de la cromatina (normalmente los diferentes tipos de células mantienen activo aproximadamente el 2% de su genoma a través de la cromatina abierta), facilitando o impidiendo la asociación de proteínas reguladoras y de factores de transcripción a las secuencias nucleotídicas de promotores génicos. Estos cambios funcionan como etiquetas bioquímicas llamadas marcadores epigenéticos, que actúan como interruptores para controlar cuales genes serán leídos por la maquinaria transcripcional[19]. La cadena del DNA se comporta dentro de la regulación de su expresión, como una molécula polígama de una gran cantidad de factores genéticos y epigenéticos.

Como hemos mencionado, el Consorcio ENCODE logró obtener un mapa global del reguloma para determinar qué partes del DNA pertenecen a este y qué regiones participan en transcripción génica (regiones reguladoras, asociación con los diferentes factores de transcripción, modificaciones estructurales de la cromatina y químicas de las histonas)[4,5,20]. Considerando los mecanismos genómicos y epigenómicos de la expresión genética, ENCODE empleó para lograr su objetivo, múltiples metodologías que exploran simultáneamente los principales componentes genéticos y epigenéticos. Entre las tecnologías empleadas más importantes, destacan la inmunoprecipitación de la cromatina combinada con la secuenciación del DNA (ChIP-seq), que revela el lugar donde las proteínas se unen al DNA; el empleo de ensayos con la enzima DNasa 1 combinada con la secuenciación del DNA (DNase-seq), la DNasa corta sitios hipersensitivos que corresponden a regiones de cromatina abierta, donde secuencias específicas son unidas a factores de transcripción y a la maquinaria proteínica de transcripción; la secuenciación de los diferentes RNAs (RNA-seq) que identifica transcritos con corte-empalme diferencial, los RNAs no-codificantes, mutaciones postranscripcionales y fusión de transcritos; diferentes tipos de ensayos para identificar el estado de metilación de los dinucleótidos CpG en las secuencias del DNA, como CAGE/RRBS, y el aislamiento de elementos reguladores por formaldehído (FAIRE-seq)[3].

El proyecto ENCODE empleó en su estudio 24 tipos de tecnologías genómicas experimentales de segunda y tercera generación (algunas previamente mencionadas) para explorar más de 150 tipos celulares. Su estudio identificó y secuenció todos los RNAs transcritos en el genoma humano, los sitios de unión al DNA de cerca de 120 factores de transcripción, más de 70,000 regiones promotoras y 400,000 regiones aumentadoras o *enhancers*; también identificó las regiones metiladas del DNA (las cuales corresponden generalmente a los genes que no expresan), y los patrones de las modificaciones químicas de 13 tipos de histonas (los cuales ayudan a empacar el DNA en los cromosomas, conduciendo al aumento o supresión de la expresión génica)[21]. Este Consorcio demostró que cada tipo celular emplea diferentes combinaciones y permutaciones en los mecanismos regulatorios de expresión génica para constituir su propio fenotipo. Las conceptos más importantes concluidos por el Consorcio ENCODE han permitido la construcción de un mapa de identificación de 3,000,000 regiones que participan en la regulación de la expresión génica (muchos de ellos previamente identificados), la identificación de secuencias específicas de DNA que se unen a proteínas que participan en el reguloma, la obtención de un bosquejo preliminar de redes de factores de transcripción que se une al DNA para promover o inhibir la expresión génica, y determinaron que más del 75% del total del genoma humano es capaz de transcribir (distintos tipos de RNAs son transcritos en diferentes momentos y tipos de células), dichos transcritos corresponden a moléculas funcionales que junto con otros diferentes factores de transcripción son entrelazados a múltiples genes, provocando cambios en su nivel de expresión. Los resultados del proyecto ENCODE demuestran que las diferentes regiones no-codificantes del DNA deben ser necesariamente consideradas, para reinterpretar el efecto de las variaciones estructurales del DNA determinadas en los estudios de asociación pangenómicos previos, ya que aproximadamente el 90% de las variaciones génómicas se presentan fuera de las regiones codificadoras de genes. En los próximos años ENCODE ampliará su estudio a otros tipos celulares, y a la exploración del efecto de variantes de histonas y de otros diferentes factores de transcripción[3,4,6,22].

Conclusiones

La comprensión de la organización estructural del genoma humano ha requerido del trabajo progresivo de muchos grupos biomédicos de investigación por más de 30 años. Hemos entrado a la etapa del estudio de la Genómica Funcional en estos últimos años. En la regulación de la expresión génica participan mecanismos genómicos y epigenómicos, cuyo análisis será empleado para mejorar el entendimiento de conceptos de salud como el desarrollo embrionario, la multicelularidad, el envejecimiento y alteraciones de salud como las enfermedades crónico-degenerativas.

Las variaciones en el número de copias de regiones particulares son uno de los tipos de variaciones estructurales más frecuentes en el genoma humano; estas favorecen condiciones evolutivas adaptativas, variabilidad en la respuesta a agentes externos y predisposición a algunas enfermedades del desarrollo y degenerativas.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Financiamiento

Los autores no recibieron patrocinio para llevar a cabo este estudio.

Referencias

1. Marx V. Reading the second genomic code. Nature 2012;491:144-147.
2. Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. Plos Biology 2007;5:e254.
3. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-72.

4. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012;337:1190-1195.
5. Stamatoyannopoulos JA. What does our genome encode? Gen Res 2012;22:1602-1161.
6. Schadt E, Chang R. A GPS for navigating DNA. Science 2012;337:1179-1180.
7. Singleton AB. Exome sequencing: a transformative technology. Lancet Neurol 2011;10:942-946.
8. Lewin B. Genes IX. Boston: Jones and Bartlett Publishers; 2008. p. 640-666.
9. Hackett PB, Largaespada DA, Cooper LJ. A transposon and transposase system for human application. Mol Ther 2010;18:674-683.
10. Choy KW, Setlur SR, Lee C, et al. The impact of human copy number variation on a new era of genetic testing. BJOG 2010;117:391-398.
11. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med 2010;61:437-455.
12. Wain LV, Armour JA, Tobin MD. Genomic copy number variation, human health and disease. Lancet 2009;374:340-350.
13. Girirajan S, Campbell CD, Eicher EE. Human copy number variation and complex genetic disease. Annu Rev Genet 2011;45:203-226.
14. Raphael BJ. Chapter 6: Structural variation and medical genomics. PLoS Comput Biol 2012;8(12)e10002821.
15. Ueno T, Emi M, Sato H, et al. Genome-wide copy number analysis in primary breast cancer. Expert Opin Ther Targets 2012;16Suppl1:S31-35.
16. Vissers LE, Stankiewicz P. Microdeletion and microduplications syndromes. Methods Mol Biol 2012;838:29-75.
17. Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. J Hum Genet 2012;57:6-13.
18. Van Binsbergen E. Origins and breakpoint analyses of copy number variations: up close and personal. Cytogenet Genome Res 2011;135:271-276.
19. Li G, Ruan X, Auerbach RK, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 2012;148:84-98.
20. Neph S, Vierstra J, Stergachis AB, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 2012;489:83-90.
21. Maher B. The human encyclopedia. Nature 2012;489:41-43.
22. Barroso I. Non-coding but functional. ENCODE explained. Nature 2012;489:54.