

Fundamentos y Prospecciones del Paradigma de las Ciencias Ómicas en la Salud Humana (Primera Parte)

Guadalupe Prado Flores¹ y Arturo César García Casillas²

Resumen. *La vida es una manifestación emergente de la materia organizada. Del nivel material derivan cualidades energéticas, organizacionales, dinámicas, informáticas, sistémicas, reguladoras, reproducibles, relacionales, hologramáticas, autoreparables y unitarias. La congruencia entre la estructura molecular y la funcionalidad de sus componentes se considera la base constitutiva del proceso vital. En dicha plataforma estructural y operativa, el ADN constituye el fundamento de la información. Este ensayo pone el acento en la organización génica como fundamento de mecanismos universales de replicación, transcripción y traducción del código genético, hasta la construcción del proteoma. Dichos procesos secuenciales sostienen la unidad vital e igualmente son el origen evolutivo por el que se diversifican los organismos. De su revisión se genera la exploración al constructo teórico metodológico de las ciencias ómicas.*

Palabras Clave: *Gen, ADN, ARNs, Replicación, Transcripción, Traducción.*

¹ Profesor-Investigador, Departamento de Producción Agrícola y Animal, Universidad Autónoma Metropolitana, e-mail: gprado@correo.xoc.uam.mx

² Investigador Posdoctoral, Maestría Interinstitucional en Producción Pecuaria, Universidad de Colima, e-mail: cesargarciasillas@hotmail.com

Abstract. *Life is an emerging manifestation of organized matter. From the material level derived qualities, energetic, organizational, dynamic, informatic, systemic, regulatory, reproducible, relational, hologramatic, self-repairing and unitary qualities. The complex molecular level is considered the basis of the congruence between structure and functionality and is a constituent part of the total reality of the vital process. This essay emphasis on gene organization as the foundation of universal mechanisms of replication, transcription and translation of the genetic code to the construction of the proteome, which are sequential processes that sustain the vital unit and are also the evolutionary origin by which organisms are diversified. Its revision generated the theoretical methodological construct of the omic sciences.*

Keywords: *Gene, DNA, RNAs, Replication, Transcription, Translation.*

INTRODUCCIÓN

El ADN es un componente material de máxima organización. El proceso científico de investigación sobre la naturaleza y expresión de esta biomolécula ha generado un conocimiento multidisciplinario que abarca características del compuesto, comportamientos termodinámicos, sistémicos y reguladores que dan fundamento a los procesos de replicación, transcripción y traducción del propio ADN. Mediante estos mecanismos universales, la actividad celular sostiene la sobrevivencia, la herencia, la evolución y la permanencia de los organismos en los ecosistemas. Por la precisión de su naturaleza y la susceptibilidad a alteraciones, tanto la estructura como la expresión del genoma, pueden sufrir cambios con el significado de patologías monogénicas o patologías multifactoriales. Para acometer este acucioso análisis, las ciencias biológicas ejercitan los recursos más avanzados e integradores en los campos teóricos y metodológicos con el sustantivo apoyo de las ciencias y técnicas de la información.

ASPECTOS HISTÓRICOS

El descubrimiento del ácido desoxirribonucleico (ADN) por Friedrich Miescher en 1869 (Dahm, 2008), la descripción de su estructura por Watson y Crick en 1953 (Pray, 2008), la presencia de las células en los organismos por Matthias Jakob Schleiden y Theodor Schwann en 1838 (Zhegunov, 2012) y la información generada por la secuenciación del ADN humano en 2001-2003 (International Human Genome Sequencing Consortium, 2004) representan para genetistas, biólogos, biólogos moleculares, bioinformáticos, médicos y personal implicado en su estudio, un referente interno en las ciencias biológicas, así como para las actividades progresistas de las sociedades y los que se benefician de sus adelantos.

Desde el año 1938, la atracción por los fenómenos de la herencia, los genes, el genoma completo, y su definitiva direccionalidad en la vida humana, ha diseñado una trayectoria que es un hito en las ciencias biológicas. Su desarrollo responde a preguntas científicas, planteamientos teóricos y recursos metodológicos idóneos que analizan su presencia, compleja estructura, selectiva expresión y múltiples funciones. La plataforma que ha logrado la construcción de conocimiento y sus aplicaciones se afianza tanto en la investigación, la experimentación, la praxis de la biología molecular, como en las ciencias bioinformáticas. Estos avances permiten entender en gran medida la fisiología integral del genoma, su participación como agente de la evolución y el principal factor de la homeostasia (Bordbar *et al.*, 2011).

El ácido desoxirribonucleico

El modelo teórico del ADN se evidencia por su identificación analítica con equipos de alta resolución como cromatografía, espectrometría de masas y difracción por rayos X (Jaksch y Tay, 2014). Paralelamente, sus funciones se analizan, contrastan e interpretan desde rigurosos diseños

y múltiples experimentos, cuyos resultados finalmente son validados con métodos estadísticos.

Es conveniente tomar en cuenta la plataforma fundamental de la mecanística funcional de las moléculas y procesos mediante los cuales la célula ejercita su quehacer. La doble hélice complementaria del ADN (Figura 1) está formada por nucleótidos de adenina (A), timina (T) guanina (G) y citosina (C). Las bases púricas y pirimídicas están unidas entre sí por radicales fosfato y el azúcar desoxirribosa, en un orden preciso y en agrupaciones funcionales. Este orden secuencial constituye la estructura primaria de la molécula de ADN. Las hebras complementarias se acoplan entre sí por dos puentes de hidrógeno en el par A-T, y con tres enlaces en el par C-G. Esta doble cadena ensamblada se enrolla sobre octámeros de histonas. Dichas proteínas básicas atraen al ADN que posee carga negativa; por tanto, permanecen fuertemente unidos.

Figura 1. James Dewey Watson (1928) y Francis Harry Compton Crick (1916-2004) ante un modelo de lámina de la estructura del ADN



Fuente: A Signet Book-The New American Library, Nueva York 1969.

Su arquitectura es tridimensional; se compacta por mecanismos sucesivos formando nucleosomas, los que a su vez forman las hebras de cromatina, éstas se empaquetan hasta integrar los dominios subcromosómicos, que al plegarse dan lugar a los cromosomas (Cremer y Cremer, 2010). La organización del genoma nuclear significa 2850 millones de pares de bases (Mb), conformada por secuencias codificantes de proteínas, no codificantes, pseudogenes, secuencias únicas y repetidas del ADN, pero que son transcripcionalmente activas y con poca condensación, denominada eucromatina (Horvath *et al.*, 2001). El resto de las Mb corresponde a regiones de heterocromatina, altamente condensada. La primera está abierta a la inducción de la expresión génica, mientras que la segunda, más compactada, disminuye notablemente la expresión de la información que contiene (Ramírez, 2016).

Estructura del ADN

Este orden define la identidad de una especie y también le concede sus características particulares, dándole la individualidad. En cualquiera de los estadios de la filogenia, cada organismo contiene toda y la misma información genética en todas sus células, mientras que la expresión de esa estructura se hace manifiesta diferencialmente en los linajes celulares del organismo (Portin, 2014). El ADN condensa la información estructural y funcional del individuo.

El conocimiento de la estructura del ADN humano ha sido una empresa científica de la mayor relevancia. Con recursos de alta tecnología se hizo la secuenciación, de la cual se informó sobre 94% del genoma, en febrero de 2001. Tanto el "Proyecto del Genoma Humano" (PGH), de naturaleza pública, como el privado, *Celera Genomics*, dieron a conocer que el ADN en el *Homo sapiens* contiene 3.2×10^9 pares de bases y 23 000 genes que codifican para proteínas (Jasny y Kennedy, 2001; Venter *et al.*, 2001).

Se informó sobre la presencia de repeticiones de secuencias simples, microsatélites de 1-13 bases, minisatélites de 14 a 500 bases, duplicaciones de segmentos que pueden ser intercromosómicas o intracromosómicas, transposones polimórficos y más de 1.4 millones de polimorfismos de un solo nucleótido (SNPs) (Scott *et al.*, 2001). En el año 2012, se constituyó el *HapMap* con el objetivo de determinar los patrones comunes de variaciones (*International HapMap*, 2003). Con la información del PGH, del *SNP Consortium* y *HapMap* se realizaron los primeros estudios de asociación del genoma completo: *GWAS Genome-wide association studies*, con el objeto de buscar asociaciones significativas entre alguna patología y variantes genéticas (Bush y Moore, 2012).

La importancia de esta información radica en que se encontraron relaciones fundamentales entre combinaciones de factores genéticos y ambientales para patologías como diabetes (Saxena *et al.*, 2007), obesidad (Hinney *et al.*, 2007), lipidemias (Wallace *et al.*, 2008), cáncer (Thomas *et al.*, 2008), enfermedades cardiovasculares e inflamatorias (Torkamani *et al.*, 2008), entre otras. El entendimiento de su patogenia ayuda a un mejor diagnóstico, estrategias de profilaxis y mejores tratamientos, ya que en múltiples casos, su patrón de herencia es poligénico.

A la luz de tal información, se crearon nuevas herramientas para evaluar el total de nucleótidos a nivel masivo en poco tiempo. De esa manera, se revelaron numerosas variantes genéticas normales de los individuos. Estos avances permiten el desarrollo de nuevas tecnologías con las que se evalúan a gran escala el ADN, ARN, SNPs, proteínas, procesos de metilación (Mardis, 2011) y, en mayor progresión, las metodologías y disciplinas conocidas como las “ciencias ómicas” que contribuyen a identificar patologías mendelianas o multifactoriales (Evans, 2000).

La orientación de esta nueva manera de ver la patología conduce a una medicina personalizada, preventiva, predictiva y participativa (4P), apoyada por la farmacogenómica (Poland *et al.*, 2009). Dicha óptica en la investigación particulariza los factores que determinan los patrones de patología, discapacidad y mortalidad en el mundo (Ramírez, 2016).

Expresión del ADN

A partir de la estructura de la macromolécula se expresan las funciones de la herencia y todas las manifestaciones metabólicas de los organismos (Cremer y Cremer, 2010). De su estudio se han derivado sus implicaciones en procesos de proliferación celular, diferenciación, crecimiento, apoptosis e inmunidad celular, los cuales manifiestan la organización autónoma de los organismos (Ramírez, 2016). La disfunción de algunas piezas de la plataforma génica es causa frecuente de patologías raras y también de las comunes, llamadas patologías multifactoriales (Stolk *et al.*, 2008).

El gen

Wilhelm Johannsen acuñó el término *gen*, en el año 1909. A partir de esta primera apreciación ha evolucionado su concepción. Se le consideró como unidad de la herencia correspondiente a una región genómica estructural, formada por nucleótidos con una distribución específica (Johannsen, 2014). Su simple definición se amplió gracias al análisis que hicieron Griffiths y Stotz (2006) al incorporar los conceptos instrumental, nominal y post-genómico. En el año 2012, la Organización de Nomenclatura del Genoma Humano lo definió como “un segmento de ADN que contribuye a una función/fenotipo” (Ramírez, 2016).

La funcionalidad del ADN se realiza en buena medida por la expresión de los genes, que son organizaciones moleculares estrictas y discretas (Griffiths y Stotz, 2006). Se pensaba que cada gen originaba una proteína, conceptualización que se ha ampliado al reconocer que sólo 1.5% del genoma codifica para proteínas (Pennisi, 2001). Asimismo, se pensaba que las zonas que no constituían los genes tenían poca o ninguna función, ahora se reconoce que la mayor parte del genoma no se organiza en genes, y los estudios revelan las implicaciones fundamentales que llevan a cabo en la actividad celular (Zhegunov, 2012).

Los estudios actuales reconocen la complejidad y la sincronización de los genes y sus correspondientes funciones. Ellos se expresan en diferentes momentos del desarrollo y diferenciación, ciclo de vida, factores ambientales y género de los organismos (Jasny y Kennedy, 2001). Dicha expresión génica es un proceso especialmente dinámico y está regulado por eventos como la condensación de la cromatina, metilación del ADN, inicio de la transcripción, corte y empalme, estabilidad del ARNm, modificaciones variables transcripcionales, tráfico intracelular, control de la traducción y degradación de las proteínas (Wray *et al.*, 2003).

Estructura del gen

Los genes que codifican para proteínas constan de uno o varios segmentos promotores y una región subdividida en exones e intrones. Los promotores en eucariontes tienen dos partes fundamentales: el promotor basal y los elementos próximos al promotor. El promotor basal es fundamental en la regulación de la transcripción, y lo hace por medio de la ARN polimerasa II (ARN pol II) (Juven-Gershon *et al.*, 2008); ahí se ensambla el complejo de pre-iniciación junto con unas 40 proteínas que tienen actividad de factores de transcripción (Sandelin *et al.*, 2007). De los promotores humanos, 32% tiene las llamadas cajas TATA como secuencias consenso y también se encuentra el elemento iniciador (Inr) (Riethoven, 2010). Los exones representan 1.5% del genoma.

Los genes humanos codificadores de proteínas generalmente presentan en promedio 8.8 exones, cerca de 145 nucleótidos por exón, regiones 5'UTR y 3'UTR (Sakharkar *et al.*, 2004). En promedio cuentan con cerca de 27 kb. Más de 90% de los nucleótidos que conforman la región codificante se remueven en la maduración de los ARNm (Wray *et al.*, 2003). Los ARN maduros contienen además un Cap 7-metil guanosina y una cola de poli-A (Topisirovic *et al.*, 2011). Los intrones representan 24% del genoma, los grandes pueden tener hasta miles de nucleótidos; ahí se encuentran

secuencias consenso que regulan el corte y empalme convencional o corte y empalme alternativo (Sakharkar *et al.*, 2004). Las secuencias en *cis* de los intrones de los ARNm primarios se relacionan con la formación de diversos transcritos con la expresión génica, y pueden contener también genes de micro ARN (miARN) (Cuperus *et al.*, 2011; Varani, 2015).

Las regiones 5'UTR y 3'UTR son estratégicas en los genes porque regulan funciones transcripcionales, ofrecen eficiencia en la traducción, estabilidad en la estructura y facilitan la exportación del núcleo al citoplasma (Riethoven, 2010). Las formas cortas de los 3'UTR tienen mayor estabilidad que las largas. Se ha reportado que las regiones 3'UTR pueden presentar sitios de poliadenilación alternativa, lo cual produce gran diversidad de transcritos alternos a partir de un solo gen codificante de proteína (Di Giammartino *et al.*, 2011). Pueden encontrarse también en intrones y exones, lo cual aumenta la posible extensión del proteoma humano (Sakharkar *et al.*, 2004).

Los rasgos moleculares del ARN juegan un importante papel en la traducción, estabilidad de la estructura secundaria de los ARNm, corte y empalme, y favorecen la interacción entre ARNm y miARN (Riethoven, 2010; Cuperus *et al.*, 2011). Las zonas codificantes tienen otras características específicas, como cargas y polaridad que influyen en su estructura, plegamiento, conformación y función de las proteínas (Jasny y Kennedy, 2001). En promedio, un gen humano codificante de proteína genera entre dos o tres transcritos, no obstante, hay excepciones, como el gen de neurexina 3, por ejemplo, que puede producir, mediante corte y empalme alternativo, 1728 transcritos (Reissner *et al.*, 2013). Los ARN no codificantes (ARNnc) son muy abundantes, presuntivamente hay miles de ellos y tienen papeles importantes en los procesos biológicos; los hay pequeños, intermedios y largos (Costa, 2010).

Funciones del ADN

La vía directa del comportamiento del genoma consiste en que el ADN se replica; se transcribe para formar ARNs mensajeros que, a su vez, se traducen como proteínas. Este funcionamiento se considera vertebral y da lugar a respuestas divergentes en la actividad celular (Zhegunov, 2012). Adicionalmente, los descubrimientos actuales han dado información relevante de múltiples segmentos intergénicos, de los cuales se reconoce su funcionalidad como inductores, reguladores u otras actividades (Riethoven, 2010). Estos niveles de la actividad génica coexisten y se manifiestan en el tiempo del ser viviente, es decir, en su fenología y en el espacio en que los organismos viven su autopoiesis (Wray *et al.*, 2003). Puede considerarse como una red compleja de procesos interconectados y operativos.

Replicación del ADN

La replicación del genoma es un proceso activo de alta eficiencia y máxima exactitud. Es el suceso mediante el cual se preserva la información genética de generación en generación (Ramírez, 2016). Se realiza en el ciclo celular, antes que la célula se divida, duplica su ADN con un modelo semi-conservativo y complementario (Mechali, 2010). Puede decirse que tiene tres fases: iniciación, elongación y terminación (Dever y Green, 2012). Interviene un conjunto sincronizado de moléculas que reconocen sitios de acción, abren la doble hebra del ADN, la cortan, vuelven a unirla sin el supra-enrollamiento; otras enzimas adicionan oligos y polimerizan; otras más, dan la energía para sufragar la exigencia termodinámica (Mechali, 2010). La copia de 6 400 millones de bases en una célula diploide la realiza en un promedio de ocho horas continuas, y una vez iniciado el proceso, debe concluir, de otra manera, la célula inevitablemente muere. Su factor de error es 10^{-4} - 10^{-5} (Lee *et al.*, 2016).

La actividad catalítica en la replicación del ADN manifiesta una evolución de competencia máxima al integrar las acciones puntuales de los siguientes participantes: i) el complejo pre-replicativo que se modifica a complejo replicativo, y hasta posreplicativo, que al final impide que se reduplique el ADN antes de llegar a la mitosis; ii) cinasas implicadas en fosforilaciones, con lo cual la molécula tiene un cambio conformacional y un aumento de energía libre; iii) ciclinas del tipo D que actúan en la fase G1 del ciclo; iv) reguladores del gen supresor tumoral Rb; v) helicasas que abren la doble hebra del ADN; vi) topoisomerasas de tipo I y de tipo II que cortan y ligan al ADN; y vii) primasas que sirven como punto de anclaje para la síntesis del ADN y que forman parte del complejo de la polimerasa α (Masai *et al.*, 2010; Lim y Kaldis, 2013).

También participan las proteínas estabilizadoras de la hebra sencilla del ADN, el “balero replicativo”, el cual facilita el desplazamiento de la polimerasa e incrementa la eficacia de la replicación. Este complejo del balero replicativo, junto con el adaptador del balero y la ADN polimerasa, se le denomina replicasa (Lujan *et al.*, 2016). La terminación de la replicación se realiza por la acción de una proteína llamada RTP (Masai *et al.*, 2010). La polimerasa α , que tiene funciones de replicación de novo y reparación del ADN en mamíferos, contiene a las subunidades A, B, p49 y p58, cada una codificada por su correspondiente gen *polA*, *PRIM1*, *PRIM2*, con pesos moleculares de 180, 70, 49 y 58 kDa (Feng *et al.*, 2014).

Un pilar fundamental de la biología y genética molecular es el conocimiento estructural y de expresión del ADN en los genes y las regiones no codificantes, los dominios que los integran, los factores que inducen su expresión, la participación de enzimas, los agentes reguladores de su expresión, sus características y la relación de unos genes con otros (Masai *et al.*, 2010). Entre las nuevas ciencias, la genómica, contribuye a la identificación de alteraciones genéticas relacionadas con patologías comunes (Xiong *et al.*, 2015).

Transcripción del ADN

Se conoce que 70% del genoma se transcribe, pero sólo 1.5% lo hace para formar ARNm que codifica a proteínas (Pennisi, 2001), el resto son ARNnc (Costa, 2010). Este proceso constituye operaciones constantes para dar la estructura y funcionalidad a los organismos mediante la proliferación celular, diferenciación y regulación del metabolismo (Zhegunov, 2012).

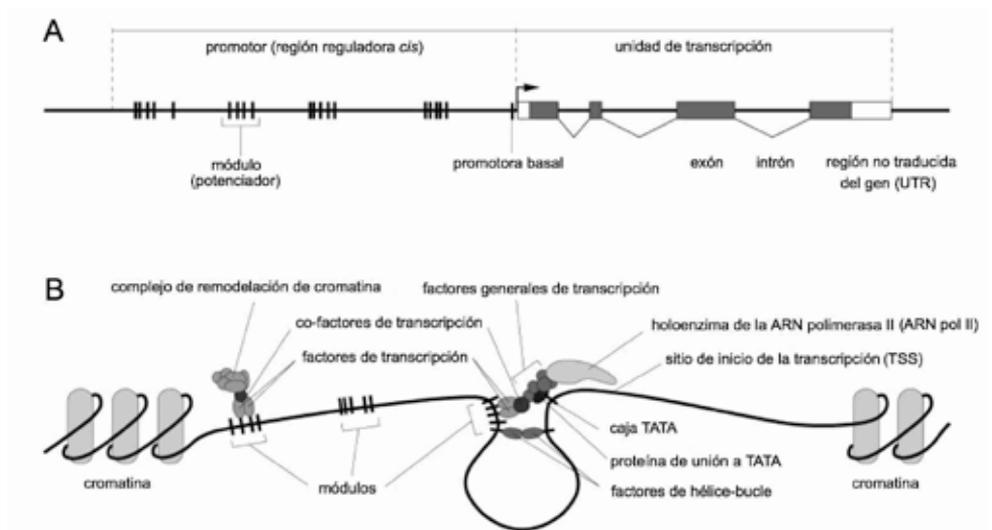
La transcripción (T) contiene cuatro etapas que corresponden a la iniciación, pausa, elongación y finalización del proceso; en cada una participan secuencias *cis*, que son conjuntos de nucleótidos y *trans* que generalmente son proteínas (Wray *et al.*, 2003). Estas secuencias nucleotídicas que actúan en *cis* son sitios de reconocimiento para proteínas que actúan en *trans*, las que aumentan o inhiben la expresión génica (Ramírez, 2016). La naturaleza de esas secuencias en *cis* y el estímulo definen la respuesta de los genes, de modo que no todas las células expresan los mismos genes al mismo tiempo, sino que son un reflejo espacio-temporal definido por la fenología y otros factores (Griffiths y Stotz, 2006). Las proteínas que participan se clasifican en tres grupos, relacionadas con su actividad en la activación, inicio y elongación de la transcripción.

Inicio

Se localiza el promotor que contiene su núcleo de promotor y elementos próximos a él. Asimismo, contiene elementos reguladores distales que incluyen a los módulos potenciadores, silenciadores y aisladores o regiones de control del *locus* (Ernst *et al.*, 2012). Se conocen dos tipos de núcleos del promotor: los centrados y los dispersos; en los primeros, hay los sitios de inicio de la transcripción (TSS) y en muchas ocasiones la caja TATA (Figura, 2), el iniciador, elementos de reconocimiento y otros componentes como elementos río abajo o río arriba del promotor (DPE y DCE) (Core *et al.*, 2014). En los núcleos dispersos se encuentran generalmente

islas CpG (Zeng *et al.*, 2014). La ARN pol II y los factores de transcripción (TFIIB, TFIID, TFIIE, TFIIF y TFIIH) se unen en este sitio y en conjunto forman el complejo de inicio (PIC) (Engel *et al.*, 2017).

Figura 2. Estructura y función del promotor



A) organización de un gen eucariótico generalizado. Muestra la posición relativa de la unidad de transcripción, región promotora basal (caja negra con flecha doblada) y sitios de unión del factor de transcripción (barras verticales)

B) promotor iniciando la transcripción. Incluye el complejo de holoenzima ARN polimerasa II (15 proteínas); caja TATA; proteína de unión a TATA; factores de transcripción; co-factores de transcripción y complejo de remodelación de cromatina.

Fuente: Modificado a partir de Wray *et al.* (2003).

Hay núcleos de promotores humanos que no contienen caja TATA y presentan combinación de Inr y DPE (Jasny y Kennedy, 2001). En un estudio de 10 000 promotores humanos se reconoció que el más abundante de los elementos era el Inr, ya que estaba contenido en casi la mitad de ellos, mientras que DPE y BRE formaron cada uno la cuarta parte y sólo una octava contenía caja TATA (Sandelin *et al.*, 2007; Core *et al.*, 2014). Hay estudios que evidencian la participación fundamental del Inr, ya que puede activar el inicio de la transcripción, solo o junto con la caja TATA o con DPE y MTE (Dikstein, 2011). Las investigaciones de Smale y Baltimore (1989) y Baumann *et al.* (2010) apuntan que mutaciones en el Inr afectan el inicio de la T; junto con los elementos anteriores se han identificado otros más como BRE y DPE que permiten la unión de los factores de transcripción (FT) (Jasny y Kennedy, 2001).

El inicio de la T requiere que la ARN pol II sea reclutada por el promotor, que se ensamble el PIC que contiene los FT y las secuencias que facilitan esa unión en el orden de reclutamiento que se llama la vía del ensamble secuencial (Baumann *et al.*, 2010). Los estudios acerca de la naturaleza enzimática de la ARN pol II señalan que es un complejo formado por 12 polipéptidos (Bartlett y Stirling, 2003). Su región Rpb1 es responsable de su actividad catalítica (Juven-Gershon *et al.*, 2008). El dominio carboxilo terminal (CTD) interviene en diferentes pasos, aumentando o modulando la eficiencia de la síntesis madura del ARNm (Baumann *et al.*, 2010). Este dominio se fosforila, y en esa condición influye en el reclutamiento de modificadores de histonas, remodelación de la cromatina, además participa en las diferentes etapas de la T y en la expresión de genes (Sandelin *et al.*, 2007). Los transcritos producidos experimentan la adición de Cap en el extremo 5', eliminación de intrones, unión de exones y adición de una cola de poli-A en el extremo 3'. Recientemente, se ha descrito la participación de un complejo mediador que funciona como un puente entre la unión del ADN con los FT (Borggreve y Yue, 2011).

Pausa

El evento de transición entre el inicio y la elongación de la T es clave en el proceso. Se realiza cerca del promotor, cuando comienza la transcripción de los ARNm y consiste en: i) desenrollar el ADN; ii) formar una burbuja de transcripción; iii) ya abierta la burbuja, la ARN pol II inicia la síntesis del ARNm; iv) la ARN pol II escapa del promotor, y v) formación del complejo de elongación (Sandelin *et al.*, 2007; Juven-Gershon *et al.*, 2008).

Elongación

Se ha revisado que hay dos fases: la temprana y la tardía. La fase temprana de la elongación depende de la participación del factor de elongación P-TEFb en la fosforilación de la serina 2 en la secuencia YSPTSPS del CTD de la enzima polimerasa (Fujinaga *et al.*, 2015). Al complejo de elongación se unen las proteínas Spt4/5, se adiciona el Cap 7-metilguanosina al extremo 5' de los transcritos de síntesis cuando tienen una longitud de unos 20 nucleótidos, se da la exportación del ARNm del núcleo, y se ha demostrado que el Cap recluta a un complejo de unión (CBC) para proseguir la T (Ramírez, 2016). Algunas causas que afectan la T son las siguientes: i) la presencia de factores de elongación cercanos a los promotores; ii) la probabilidad que se requieran factores que disocien la ARN pol II de los factores de inicio localizados en el promotor, y iii) la posibilidad que los nucleosomas cercanos al sitio de inicio inhiban la elongación del transcrito (Juven-Gershon *et al.*, 2008).

Terminación

Esta etapa tiene dos partes: i) pausa de la elongación en el sitio de terminación, y ii) desensamble del complejo de elongación (Fujinaga *et al.*,

2015). Se ha referido que este fenómeno es mediado por estructuras secundarias del tallo de los ARNm y otras secuencias de proteínas (Sandelin *et al.*, 2007). Los potenciadores, los silenciadores y los aisladores, así como las regiones de control del *locus* (LCR), son otras secuencias consenso que están involucradas en la regulación de la expresión génica (Dever y Green, 2012). Los potenciadores sirven para aumentar la concentración de factores que inducen la expresión o para reclutar a la ARN pol II (Li *et al.*, 2016). Los silenciadores regulan en forma negativa la expresión génica a distancia de los promotores de ciertos genes, que generalmente se refieren a la diferenciación e inhibición del ciclo celular (Rio, 2015). Recientemente, se han identificado otros silenciadores llamados ARN antígenos cortos y largos con función represiva (Kolovos *et al.*, 2012). Los aisladores regulan secuencias de ADN localizadas en diferentes dominios de la cromatina (Van Bortle y Corces, 2013). Ellos regulan la función alterada de los potenciadores o silenciadores y lo realizan con dos rasgos principales: i) bloquean la comunicación al romper la interacción entre potenciadores-promotor y ii) evitan la propagación de la cromatina represiva (Heger y Wiehe, 2014). Los LCR actúan en *cis* a gran distancia de la cromatina y llevan a un aumento de la expresión génica (Van Bortle y Corces, 2013).

Traducción del ARNm en proteínas

Este dinamismo consiste en la traducción de un alfabeto químico de bases púricas y pirimídicas del ARNm, en el orden secuencial de los aminoácidos (aa) que componen todas las proteínas normales o anómalas de una célula (Bass, 2015). Estas moléculas cuaternarias tienen actividades tanto estructurales, como funcionales, y son la expresión más elaborada del metabolismo celular (Pedley y Benkovic, 2017). Su síntesis se realiza en el citoplasma y participan organelos como los ribosomas, sistemas enzimáticos, moléculas pequeñas como los aa y los ARNm, ARNt, ARNr

y ARNnc (Lane y Fan, 2015). Su factor de error es muy bajo, entre 1×10^{-4} - 1×10^{-5} (Ramírez, 2016).

Los ARNm maduros presentan diferentes estructuras y secuencias consenso con funciones definidas: i) no tienen intrones (Lasda y Parker, 2014); ii) contienen la región 5' UTR (*untranslated region*) con un cap en el extremo (Schwerk y Savan, 2015); iii) su región 3' UTR contiene una cola de poli-A y diversos elementos reguladores involucrados en la estabilidad de los mensajeros y con interacción con miARN (Varani, 2015); iv) poseen su región codificante en los exones (Lasda y Parker, 2014). En esta región se encuentra el marco de lectura abierta (ORF *open reading frame*) o codón de inicio de la traducción, también llamado secuencia codificante (Somers *et al.*, 2013), y v) la secuencia codificante que tiene codones formados por tripletes de bases que son leídos como un aa específico. Estos codones interactúan con anticodones presentes en los ARNt (Riethoven, 2010). El codón de inicio es AUG y se reconocen tres codones de terminación que son UAA, UAG y UGA (Lujan *et al.*, 2016). Actualmente, se han detallado al menos 22 aa adicionales a los 20 anteriormente calificados como integrantes de las proteínas; inclusive, hay otros que no han sido cabalmente descritos (Ramírez, 2016).

Los sistemas enzimáticos

Participan las aminoacil-ARNt sintetasas específicas para cada aa, catalizando la unión de un aa específico al extremo 3' de su respectivo ARNt (Yao y Fox, 2013). Ahora se sabe que, junto con esta acción catalítica de unión, también participan en la biosíntesis de aa, replicación del ADN, corte y empalme del ARN y control del ciclo celular (Diodato *et al.*, 2014).

Los ARNt tienen unos 80 nucleótidos y su correcta colocación está dada por: i) el anticodón, que reconoce un codón específico en el ARNm que inicia su traducción; ii) el primer sitio de unión une con enlace covalente un aa específico, iii) el segundo sitio reconoce y se une a la ami-

noacil ARNt sintetasa. Esta unión requiere energía que es suministrada por ATP, el cual queda como AMP y iv) mediante un gen codificante, el ARNm se une al ribosoma y los ARNt transportan el aa para formar la proteína (Topisirovic *et al.*, 2011; Dever y Green, 2012).

Los ribosomas

Están adosados al retículo plasmático, dando una configuración rugosa. Su masa es de 4.3 MDa y están formados por ARNr y proteínas. En eucariontes sedimentan con un valor de 80S y tienen dos subunidades, la pequeña de 40S formada por 3 proteínas y un ARNr 18S; la subunidad grande 60S contiene 47 proteínas y tres tipos de ARN: 5S, 5.8S y 28S (Brown *et al.*, 2014). Contienen tres sitios: el A para el aminoacil-ARNt; el sitio P para el peptidil-ARNt, y el sitio E del ARNt acilado (Gupta y Warner, 2014).

La traducción tiene tres fases: inicio, elongación y término (Dever y Green, 2012). Las células eucariontes emplean diversos mecanismos para iniciar la traducción, esto depende de los estímulos o las condiciones anómalas que la célula recibe (Borggreffe y Yue, 2011). La mayoría de los ARNm son traducidos por el mecanismo Cap-dependiente (Baumann *et al.*, 2010). La cap es reconocida por 11 factores de inicio (eIFs) con el reconocimiento del codón AUG que traduce metionina, y la actividad de una helicasa que desdobra la estructura secundaria de la región 5' UTR (Topisirovic *et al.*, 2011). La elongación se da mientras el ribosoma 80S está ensamblado con el ARNm, con la participación del factor de elongación 1A (Hotokezaka *et al.*, 2002). Otros factores son las proteínas G que funcionan cuando hay hidrólisis de GTP a GDP y actúan pasando los aa del sitio A al P y luego al E (Brown *et al.*, 2014).

La terminación se da cuando se llega a un codón de terminación UAA, UGA y UAG. Esta etapa es catalizada por factores liberadores de la cadena polipeptídica: el eRF1 reconoce con alta fidelidad al codón de término e hidroliza el peptidil-ARNt en el centro peptidil transferasa ribosomal mediante el dominio amino terminal de eRF1 (Dever y Green, 2012). El dominio carboxilo facilita la interacción con el eRF3, que a su vez está involucrado en la interacción de eRF1 con la proteína de unión a la cola de poli-A (Lane y Fan, 2015). Esto aumenta la eficiencia de terminación que depende del GTP y se liberan las moléculas de ARNt durante la terminación, además que reconoce la presencia del codón de paro en el sitio A (Yao y Fox, 2013). Cuando hay alteraciones en las secuencias consenso o estructuras de la región 5' se afecta el inicio de la traducción, y cuando las alteraciones se dan en la región codificante, se puede originar una proteína trunca o cambia un aa por otro y esto origina consecuencias disfuncionales y desarrollo de patologías monogénicas y/o multifactoriales (Savic y McDermott, 2015; Lee *et al.*, 2016).

Es importante reconocer que esta compleja red de procesos se expresa con orden, sincronía y alta eficiencia, características que están dadas tanto por la estructura del ADN, como por la expresión del mismo. Cuando esta información génica se afecta por estímulos físicos, como la temperatura; químicos, como la presencia de un fármaco; nutricionales, como la alimentación, o perturbaciones como el estrés o radiaciones ionizantes, se modifica su expresión, la cual lleva implicados cambios en los procesos de transducción y traducción. En condiciones normales, intervienen los mecanismos de reparación del ADN que mantienen la funcionalidad correspondiente. Si los estímulos sobrepasan el umbral de reparación, sobrevienen procesos patológicos que pueden ser monogénicos o multifactoriales.

CONCLUSIONES

El conocimiento de la estructura del ADN y su funcionalidad se ha ampliado a niveles de especialización gracias a los grupos interdisciplinarios de científicos, a los recursos teóricos, a las metodologías de alta resolución, a los lenguajes y desarrollo de las ciencias informáticas y a las fundaciones nacionales e internacionales que han apoyado los costos. La publicación de artículos científicos con difusión mundial ha sido un catalizador que ha impulsado mayor productividad en este campo.

Los beneficios del conocimiento de estos procesos están en franca relación con la salud humana, se ha favorecido la comprensión de procesos patológicos y, por el momento, se cuenta con mayores recursos para hacer prognosis oportunas, diagnosis más asertivas y tratamientos menos invasivos.

Con base en tales conocimientos, las ciencias biológicas han generado redes de información de mayor precisión, por ejemplo, la nutrigénómica, la cual registra relaciones puntuales entre la genómica, epigenómica, transcriptómica y proteómica con el campo de la nutrición; la farmacogenómica, que nace de las relaciones ya señaladas con los fármacos. En este mismo modelo se conocen, entre otras, a la medicina genómica, metabolómica y lipidómica que se orientan tanto a la salud, como a la solución de problemas vinculados con ella.

BIBLIOGRAFÍA

- Bartlett, M. y D. Stirling, 2003, "A short history of the polymerase chain reaction", en *Methods Mol Biol* 226(1): 3-6.
- Bass, L., 2015, "Twenty years: a very short sequence in the RNA world", en *RNA* 21(4): 490-491.
- Baumann, M. *et al.*, 2010, "Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview", en *Mol Biotechnol* 45(3): 241-247.

- Bordbar, A. *et al.*, 2011, "A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology", en *BMC Syst Biol* 5(1): 180-185.
- Borggreffe, T. y X. Yue, 2011, "Interactions between subunits of the Mediator complex with gene-specific transcription factors", en *Semin Cell Dev Biol* 22(7): 759-768.
- Brown, A. *et al.*, 2014, "Structure of the large ribosomal subunit from human mitochondria", en *Science* 346(6210): 718-722.
- Bush, S. y J. Moore, 2012, "Chapter 11: Genome-wide association studies", en *PLoS Comput Biol* 8(12): e1002822- e1002832.
- Core, J. *et al.*, 2014, "Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers", en *Nat Genet* 46(12): 1311-1320.
- Costa, F., 2010, "Non-coding RNAs: Meet thy masters", en *Bioessays* 32(7): 599-608.
- Cremer, T. y M. Cremer, 2010, "Chromosome territories", en *Cold Spring Harb Perspect Biol* 2(3): a003889- a003897.
- Cuperus, T. *et al.*, 2011, "Evolution and functional diversification of MIRNA genes", en *Plant Cell* 23(2): 431-442.
- Dahm, R., 2008, "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research", en *Hum Genet* 122(6): 565-581.
- Dever, E. y R. Green, 2012, "The elongation, termination, and recycling phases of translation in eukaryotes", en *Cold Spring Harb Perspect Biol* 4(7): a013706-a013712.
- Di Giammartino, C. *et al.*, 2011, "Mechanisms and consequences of alternative polyadenylation", en *Mol Cell* 43(6): 853-866.
- Dikstein, R., 2011, "The unexpected traits associated with core promoter elements", en *Transcription* 2(5): 201-206.
- Diodato, D. *et al.*, 2014, "The Mitochondrial Aminoacyl tRNA Synthetases: Genes and Syndromes", en *Int J Cell Biol* 2014(1): 787956-787962.
- Engel, C. *et al.*, 2017, "Structural Basis of RNA Polymerase I Transcription Initiation", en *Cell* 169(1): 120-131.

- Ernst, W. *et al.* 2012, Promoter sequences. In Department of Commerce's. United States Patent and Trademark Office, United States Pat. No. PCT/EP2010/070537. US 20120258493 A1.
- Evans, A., 2000, "Designer science and the 'omic' revolution", en *Nat Biotechnol* 18(2): 127-132.
- Feng, Q. *et al.*, 2014, "An epigenomic approach to therapy for tamoxifen-resistant breast cancer", en *Cell Res* 24(7): 809-819.
- Fujinaga, K. *et al.*, 2015, "Visualization of positive transcription elongation factor b (P-TEFb) activation in living cells", en *J Biol Chem* 290(3): 1829-1836.
- Griffiths, E. y K. Stotz, 2006, "Genes in the postgenomic era", en *Theor Med Bioeth* 27(6): 499-521.
- Gupta, V. y J. Warner, 2014, "Ribosome-omics of the human ribosome", en *RNA* 20(7): 1004-1013.
- Heger, P. y T. Wiehe, 2014, "New tools in the box: an evolutionary synopsis of chromatin insulators", en *Trends Genet* 30(5): 161-171.
- Hinney, A. *et al.*, 2007, "Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants", en *PLoS One* 2(12): e1361-e1372.
- Horvath, E. *et al.*, 2001, "Lessons from the human genome: transitions between euchromatin and heterochromatin", en *Hum Mol Gen* 10(20): 2215-2223.
- Hotokezaka, Y. *et al.*, 2002, "Interaction of the eukaryotic elongation factor 1A with newly synthesized polypeptides", en *J Biol Chem* 277(21): 18545-18551.
- International HapMap, C., 2003, "The International HapMap Project", en *Nature* 426(6968): 789-796.
- International Human Genome Sequencing Consortium, 2004, "Finishing the euchromatic sequence of the human genome", en *Nature* 431(7011): 931-945.
- Jaksch, L. y S. Tay, 2014, "Editorial overview: Analytical biotechnology: new technologies for quantitative analysis of biological specimens and natural products", en *Curr Opin Biotechnol* 25(1): 12-19.

- Jasny, R. y D. Kennedy, 2001, "The human genome", en *Science* 291(5507): 1153-1161.
- Johannsen, W., 2014, "The genotype conception of heredity. 1911", en *Int J Epidemiol* 43(4): 989-1000.
- Juven, T. *et al.*, 2008, "The RNA polymerase II core promoter-the gateway to transcription", en *Curr Opin Cell Biol* 20(3): 253-259.
- Kolovos, P. *et al.*, 2012, "Enhancers and silencers: an integrated and simple model for their function", en *Epigenetics Chromatin* 5(1): 1-9.
- Lane, N. y T. Fan, 2015, "Regulation of mammalian nucleotide metabolism and biosynthesis", en *Nucleic Acids Res* 43(4): 2466-2485.
- Lasda, E. y R. Parker, 2014, "Circular RNAs: diversity of form and function", en *RNA* 20(12): 1829-1842.
- Lee, F. *et al.*, 2016, "Mapping DNA polymerase errors by single-molecule sequencing", en *Nucleic Acids Res* 44(13): e118-e127.
- Li, W. *et al.*, 2016, "Enhancers as non-coding RNA transcription units: recent insights and future perspectives" en Lim, S. y P. Kaldis, 2013, "Cdks, cyclins and CKIs: roles beyond cell cycle regulation", en *Development* 140(15): 3079-3093.
- Lujan, A. *et al.*, 2016, "DNA Polymerases Divide the Labor of Genome Replication", en *Trends Cell Biol* 26(9): 640-654.
- Mardis, R., 2011, "A decade's perspective on DNA sequencing technology", en *Nature* 470(7333): 198-203.
- Masai, H. *et al.*, 2010, "Eukaryotic chromosome DNA replication: where, when, and how?", en *Annu Rev Biochem* 79(1): 89-130.
- Mechali, M., 2010, "Eukaryotic DNA replication origins: many choices for appropriate answers", en *Nat Rev Mol Cell Biol* 11(10): 728-738.
- Pedley, M. y S. Benkovic, 2017, "A New View into the Regulation of Purine Metabolism: The Purinosome", en *Trends Biochem Sci* 42(2): 141-154.
- Pennisi, E., 2001, "The human genome", en *Science* 291(5507): 1177-1180.
- Poland, A. *et al.*, 2009, "Application of pharmacogenomics to vaccines", en *Pharmacogenomics* 10(5): 837-852.

- Portin, P., 2014, "The birth and development of the DNA theory of inheritance: sixty years since the discovery of the structure of DNA", en *J Genet* 93(1): 293-302.
- Pray, L., 2008, "Discovery of DNA structure and function: Watson and Crick", en *Nature Education* 1(1): 95-101.
- Ramírez, J. 2016. "Gen: unidad estructural y funcional de la herencia", en Pereyra L., (comp.), *Genómica estructural y funcional en las enfermedades multifactoriales*, 15-51. pp. LEEA México.
- Reissner, C. et al., 2013, "Neurexins", en *Genome Biol* 14(9): 213-220.
- Riethoven, J., 2010, "Regulatory regions in DNA: promoters, enhancers, silencers, and insulators", en *Methods Mol Biol* 674(2): 33-42.
- Rio, D., 2015, "Twenty years of RNA", en *RNA* 21(4): 718-720.
- Sakharkar, K. et al., 2004, "Distributions of exons and introns in the human genome", en *In Silico Biol* 4(4): 387-393.
- Sandelin, A. et al., 2007, "Mammalian RNA polymerase II core promoters: insights from genome-wide studies", en *Nat Rev Genet* 8(6): 424-436.
- Savic, S. y M. McDermott, 2015, "Clinical genetics in 2014: New monogenic diseases span the immunological disease continuum", en *Nat Rev Rheumatol* 11(2): 67-68.
- Saxena, R. et al., 2007, "Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels", en *Science* 316(5829): 1331-1336.
- Schwerk, J. y R. Savan, 2015, "Translating the Untranslated Region", en *J Immunol* 195(7): 2963-2971.
- Scott, C. et al., 2001, "Homologues of Twisted gastrulation are extracellular cofactors in antagonism of BMP signalling", en *Nature* 410(6827): 475-478.
- Smale, T. y D. Baltimore, 1989, "The 'initiator' as a transcription control element", en *Cell* 57(1): 103-113.
- Somers, J. et al., 2013, "A perspective on mammalian upstream open reading frame function", en *Int J Biochem Cell Biol* 45(8): 1690-1700.

- Stolk, P. *et al.*, 2008, "Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study", en *Eur J Epidemiol* 23(1): 67-74.
- Thomas, G. *et al.*, 2008, "Multiple loci identified in a genome-wide association study of prostate cancer", en *Nat Genet* 40(3): 310-315.
- Topisirovic, I. *et al.*, 2011, "Cap and cap-binding proteins in the control of gene expression", en *Wiley Interdiscip Rev RNA* 2(2): 277-298.
- Torkamani, A. *et al.*, 2008, "Pathway analysis of seven common diseases assessed by genome-wide association", en *Genomics* 92(5): 265-272.
- Van Bortle, K. y V. Corces, 2013, "The role of chromatin insulators in nuclear architecture and genome function", en *Curr Opin Genet Dev* 23(2): 212-218.
- Varani, G., 2015, "Twenty years of RNA: the discovery of microRNAs", en *RNA* 21(4): 751-752.
- Venter, C. *et al.*, 2001, "The sequence of the human genome", en *Science* 291(5507): 1304-1351.
- Wallace, C. *et al.*, 2008, "Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia", en *Am J Hum Genet* 82(1): 139-149.
- Wray, A. *et al.*, 2003, "The evolution of transcriptional regulation in eukaryotes", en *Mol Biol Evol* 20(9): 1377-1419.
- Xiong, Y. *et al.*, 2015, "RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease", en *Science* 347(6218): 1254806.
- Yao, P. y P. Fox, 2013, "Aminoacyl-tRNA synthetases in medicine and disease", en *EMBO Mol Med* 5(3): 332-343.
- Zeng, J. *et al.*, 2014, "Fundamental diversity of human CpG islands at multiple biological levels", en *Epigenetics* 9(4): 483-491.
- Zhegunov, G., 2012, "Cells and Organisms", en Zhegunov, G., (comp.) *The Dual Nature of Life*, 73-81, EEUU. The Frontiers Collection, Springer-Verlag ed. Berlin, Heidelberg, USA.

