

Más allá del dogma central de la biología molecular

• Dr. Víctor Javier Lara Díaz¹

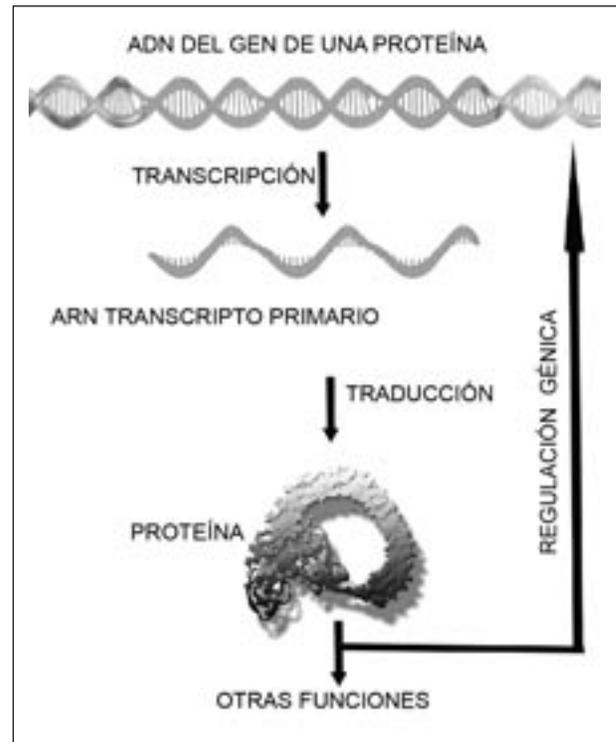
• **Palabras clave:** ADN, ARN, genética, genoma, genes, humano.

“Las suposiciones pueden ser peligrosas, especialmente en la ciencia. Generalmente se originan en la interpretación más plausible o cómoda de los hechos evidentes. Pero cuando su verdad no admite prueba inmediata, o sus defectos no son obvios, se transforman en un artículo de fe, y las nuevas observaciones se acomodan a sus predicciones. Eventualmente, si el volumen de información comprometida es insostenible, la ortodoxia sufre un colapso”, este es el comentario de John S. Mattick, en uno de sus artículos de divulgación sobre el tema que nos ocupa.

El dogma central de la biología molecular, postulado por Francis Crick, definió una vía general para la expresión de la información genética almacenada en el ADN, transcrita hacia mARNs temporales y decodificada en los ribosomas con la ayuda de ARNs adaptadores (tARNs) para producir proteínas, que a su vez llevarían a cabo todas las funciones catalíticas y estructurales de la célula. De acuerdo a este punto de vista, los ARNs jugarían más bien un papel accesorio, y la complejidad de un organismo dado estaría definida solamente por el número de proteínas codificadas por su genoma, de acuerdo a la hipótesis de “un gen – una proteína”.

Esta conclusión se derivó de estudios en *Escherichia coli* y otros procariontes y, en realidad, sigue siendo esencialmente cierta para los procariontes; su ADN consiste casi por entero de genes codificadores de proteínas, separados por secuencias laterales que regulan la expresión de los genes adyacentes, como sucede con el operón *-Lac*. El biólogo pionero Jacques Monod resumió la universalidad del dogma central diciendo “lo que fue cierto para *E. coli* sería cierto para el elefante.”(Ver Figura 1).¹

Figura 1. Actividad génica en procariontes



Los procariontes (bacterias y otros organismos unicelulares simples) tienen ADN que consiste casi por entero de genes codificadores de proteínas. Cuando están activos originan transcritos de ARN que son inmediatamente traducidos a proteínas, y éstas a su vez regulan la actividad génica y ejercen otras funciones celulares.

Este concepto simple se tornó complicado cuando se encontró que los genes de los eucariontes no eran bloques contiguos de secuencias codificadoras de proteínas, sino mosaicos de fragmentos de ADN codificadores de proteínas (exones) entremezclados con segmentos bastante largos de secuencias no codificadoras de proteínas (intrones), y por ende, al ser copiado inicialmente el ADN en el núcleo, los transcritos primarios (ptARN) de los genes codificadores de proteínas en los eucariontes proveían a la célula

¹ Director académico de la Especialidad Médica en Neonatología. Programa Multicéntrico, Escuela de Medicina del Tecnológico de Monterrey y Secretaría Estatal de Salud de Nuevo León.

Ilustraciones preparadas por: Alan Naranjo Palacios y Héctor Samuel Lara Gil.

eucarionte de un mecanismo para sintetizar más de un producto proteico a partir de un gen único, por el “splicing” o ensamblado alternativo.

Los proyectos de secuenciación de genomas han revelado un problema no esperado en nuestra comprensión de las bases moleculares de la complejidad del desarrollo de los organismos superiores: Organismos complejos tienen menor número de genes codificadores de proteínas de lo que se había anticipado. La mosca de la fruta *Drosophila melanogaster* y el nematodo *Caenorhabditis elegans* resultan tener solamente cerca del doble (~12-14,000) de genes codificadores de proteínas de lo encontrado en microorganismos como *Saccharomyces cerevisiae* (~6,200) y *Pseudomonas aeruginosa* (~5,500); los humanos parecen tener también alrededor del doble de los de la mosca de la fruta y la lombriz (~30,000), de acuerdo a J. Craig Venter y cols del International Human Genome Sequencing Consortium, 2001.²

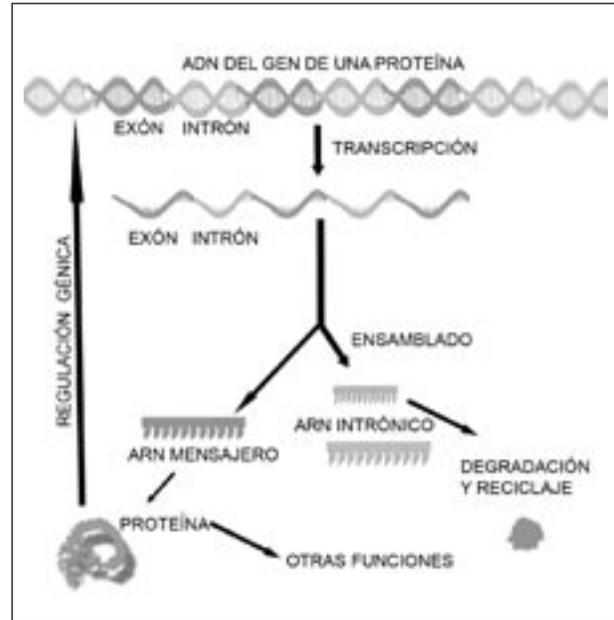
Hace más de veinte años, el descubrimiento de las propiedades catalíticas de la subunidad ARN de la Ribonucleasa P y la actividad *auto-splicing* (auto-ensamblado) de los intrones del grupo I sugirieron que las funciones del ARN iban más allá de un rol pasivo en la expresión de los genes codificadores de proteínas. Se pensó también que los ARN intrónicos resultantes del ensamblado, sin tener un propósito evidente, eran degradados y reciclados (ver Figura 2).^{1,3}

Aunque el repertorio de isoformas de proteínas expresadas en los organismos superiores se incrementa en gran manera por el fenómeno de *splicing* o ensamblado alternativo, la otra característica sorprendente en la evolución de los organismos superiores, largamente ignorada hasta ahora, es el enorme incremento en la cantidad de ARN no codificador de proteínas, que ha sido considerado como basura, chatarra (“junk”), o bien como remanente de tiempos remotos en la cadena evolutiva; en los humanos llega a ser de hasta el 98% de todo el producto genómico.

Expresado de manera simple, o el genoma de los organismos superiores está repleto de transcripciones inútiles, o estos ARNs no-codificadores de proteínas tienen alguna función específica. (Ver Tabla 1 y Figura 3).

Los proteomas de los organismos superiores son relativamente estables. Humanos y ratones compartimos 99% de los genes codificadores de proteínas y la diferenciación en éstos y otros eucariontes complejos

Figura 2. Concepto tradicional de la actividad genética en los eucariontes



En el ADN de los eucariontes (organismos complejos) los genes individuales comprenden secuencias de “exones” que codifican para segmentos de proteínas, separados por “intrones”, secuencias no-codificadoras. Cuando un gen está activo es transcrito por completo a ARN, luego el ARN intrónico es desacoplado y el ARN exónico restante es reensamblado (splicing) como mARN (ARN mensajero). La célula entonces traduce el mARN a proteína, mientras que destruye y recicla los ARN intrónicos que no tienen propósito aparente.

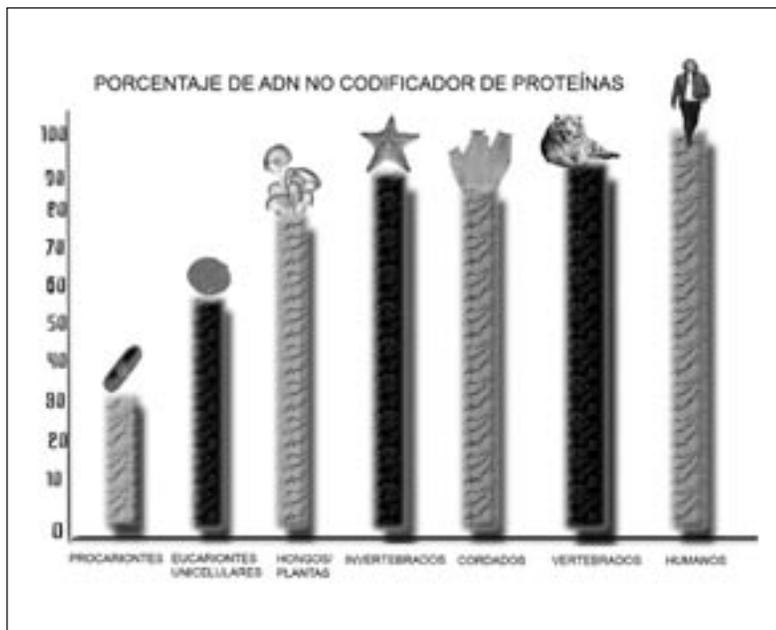
parece lograrse con base en reutilización modular y un sistema multitareas de diferentes partes del proteoma. Más aún, de aproximadamente 3,000,000 de diferencias entre genomas haploides de humanos individuales, sólo alrededor de 10,000 (0.3%) hasta un máximo de 30,000 (1%) ocurren en secuencias codificadoras de proteínas, y la mayoría de ellos como cambios silenciosos de una base en la tercera posición de un codón.^{2,3}

Así, la variación fenotípica entre individuos y especies podría estar basada en gran parte en diferencias entre secuencias no codificadoras de proteínas, y ser de este modo principalmente una situación de variación en la expresión genética. Esto implicaría que, aunque las variaciones en proteínas podrían contribuir, la fuente primaria de rasgos complejos y sus variaciones está embebida en esta arquitectura de control. De ser así, esto tendría implicaciones significativas para entender las bases de la diferenciación y el desarrollo, y las redes regulatorias subyacentes a la función neural, susceptibilidad a enfermedades y al cáncer, entre otros aspectos.

Tabla 1. Porcentajes del genoma total correspondientes a secuencias codificadoras de proteínas en varios genomas procariontes y eucariontes

| Organismo | Secuencias codificadoras(%) |
|--|-----------------------------|
| Homo sapiens | 1.4 |
| Drosophila melanogaster (mosca de la fruta) | 20 |
| Caenorhabditis elegans (lombriz de tierra) | 27 |
| Arabidopsis thaliana | 29 |
| Saccharomyces cerevisiae (levadura de cerveza) | 70 |
| Escherichia coli | 86 |
| Mycobacterium tuberculosis | 91 |
| Archa euglobus fulgidus (archae) | 92 |

Figura 3. Las secuencias no-codificadoras de proteínas constituyen sólo una pequeña fracción del ADN de los procariontes.



Entre los eucariontes, según se incrementa su complejidad, así sucede con la proporción de su ADN que es no-codificador. Estas secuencias no-codificadoras habían sido consideradas "junk" o basura, pero quizás sean las responsables de la complejidad de estos organismos.

Se ha asumido que este sistema de arquitectura de control está localizado principalmente en los genes promotores y facilitadores con acción *-cis*, que interactúan con los factores de transcripción y señalización; esto parece ser sólo una respuesta parcial, porque ignora los posibles roles de los ARNs no-codificadores, los cuales representan la vasta mayoría de la expresión genómica de los organismos superiores, que se nos presenta en dos formas: Intrones y otros ARNs no-codificadores de proteínas.

En los humanos, los intrones representan alrededor de 95% de los transcritos pre-mARN de los genes codificadores de proteínas, y generalmente son secuencias de considerable complejidad. Por otro lado, de lo que puede juzgarse de los resultados de reportes específicos y de análisis cinéticos de hibridación del hnARN (hn = Nuclear heterogéneo) otros ARNs no codificadores representan de la mitad a dos terceras partes de la transcripción de los organismos superiores, estos ARNs incluyen una plétora de transcritos antisentido e intergénicos, y pueden incluir muchas de las 65,000 a 75,000 unidades de transcripción del genoma humano. Si se asume que hasta dos terceras partes de todos los transcritos no contienen secuencias codificadoras de proteínas, el número real de "genes", definidos como aquéllos que producen transcritos primarios separados regularmente, sería de alrededor de 100,000.

Cuando se han examinado estos ARNs no-codificadores se ha encontrado que son reguladores del desarrollo y tienen efectos genéticos. Un buen ejemplo es el complejo A/B abdominal-bitórax de *Drosophila*, que abarca alrededor de 200 kb y expresa siete transcritos mayores que abarcan casi toda su longitud. Solamente tres de ellos contienen secuencias codificadoras de proteínas, pero todos tienen regulación espacial y temporal, y la interrupción o delección del ADN que los codifica tiene consecuencias fenotípicas específicas.

Si estos ARNs no-codificadores son funcionales, su papel más obvio sería el de interconexión, es decir, la producción de señales paralelas de acción *-trans* que permitirían que la actividad en un sitio se comunicara a otros distantes en tiempo real. Esto a su vez implicaría

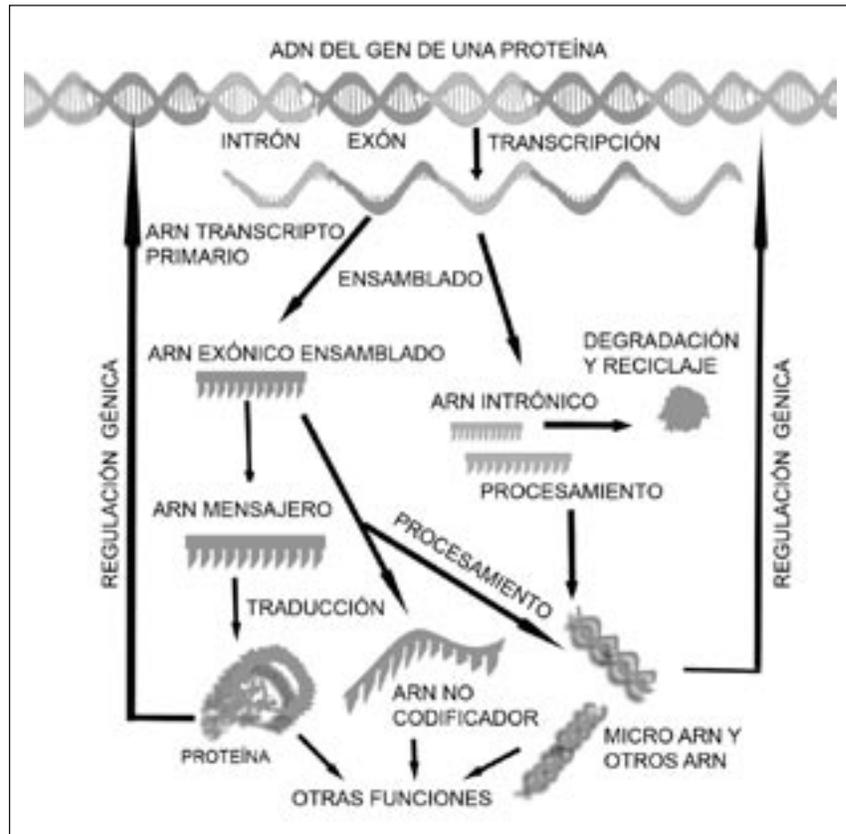
que grupos de actividad génica y otros niveles de control del sistema estarían directamente coordinados e integrados de manera programada vía señales eferentes de ARN (eARN), y sería fundamental para la operación del sistema total. Estos eARN podrían incluso actuar como memoria celular de los eventos recientes de transcripción, implicando un incremento notable en la interconectividad y funcionalidad del sistema en general, comparado con el que se lograría si solamente estuviese regulado por las proteínas y sus efectos, modificados por el ambiente y el metabolismo.

De ser cierto, no sorprendería que un sistema que utilizara una red de comunicación basada en ARN, al evolucionar, desarrollara muchos loci dedicados solamente a producir ARN.^{1,2,4} (Ver Figura 4).

Acerca de este punto, se ha desarrollado una teoría que postula que la diversidad fenotípica observada en los eucariontes resulta de la capacidad multitareas de un proteoma nuclear de tamaño limitado, equiparando esta capacidad multitareas a la de un sistema computacional avanzado, sugiriendo que estos eARNs han evolucionado para funcionar como moléculas de control endógeno de los procesos de interconexión, que permiten la comunicación directa gen-gen, mediante sofisticados mecanismos como la co-supresión, el silenciamiento transgénico, la interferencia de ARN, la improntación, la metilación, la compensación de dosis del cromosoma-X y la transvección; que a su vez involucran mecanismos como la remodelación de la cromatina, interacciones ARN-ADN, ARN-ARN y ARN-proteínas; y posibilitan de este modo la capacidad multitarea del genoma de los eucariontes, lo que les ha conferido una ventaja de orden evolutivo importante.^{4,5}

Cuando recién se descubrieron los intrones, se asumió que no eran funcionales, y que quizá eran remanentes del empalme prebiótico de cassettes exónicos codificadores de proteínas. Ahora está claro que los modernos intrones nucleares invadieron los genes de

Figura 4. Nuevo concepto de la actividad génica en los eucariontes



Algunos de los ARNs intrónicos y quizás algunos de los ARN exónicos ensamblados podrían jugar un papel regulatorio directo al interactuar con ADN, otras moléculas de ARN o proteínas. Al modificar la producción de proteínas a varios niveles, estos ARNs no-codificadores estarían proporcionando "instrucciones" genéticas a la célula, especificarían el plan de construcción y no sólo los componentes (proteínas) del sistema.

los eucariontes tardíamente en la evolución, después de la separación de la transcripción y la traducción. La fragmentación de los genes productores de proteínas ocasionada por los intrones parece haber conferido una ventaja al facilitar los rearrreglos modulares de los dominios eucariontes productores de proteínas, dando lugar al fenómeno de ensamblado alternativo. Los intrones nucleares se derivan de los intrones grupo II autoensamblantes de los procariontes, ofreciendo la posibilidad de actividad catalítica de ARN tanto *-cis* como *-trans*, como ocurre en el spliceosoma o ensambleosoma. Cualquier secuencia que adquiriese una nueva función, como la posibilidad de señalización *-trans* capaz de transmitir información en paralelo con las secuencias codificadoras de proteínas asociadas, adjudicaría por ello un cierto valor selectivo al posibilitar una red de interconexión en las células eucariontes. Esto no implica que todos los intrones hayan evolucionado a tener alguna función,

sino más bien que un número importante pudieran haberlo hecho. Tanto los intrones como los ARNs no codificadores de proteínas constituyen actualmente la mayoría de la producción del genoma de los eucariontes complejos, de hecho, hay una buena correlación entre la densidad de intrones y la complejidad del desarrollo.

También se aprecian patrones de conservación, aun a través de distancias evolutivas grandes, si bien esta conservación es menor que la observada en los exones (ver Figura 5).

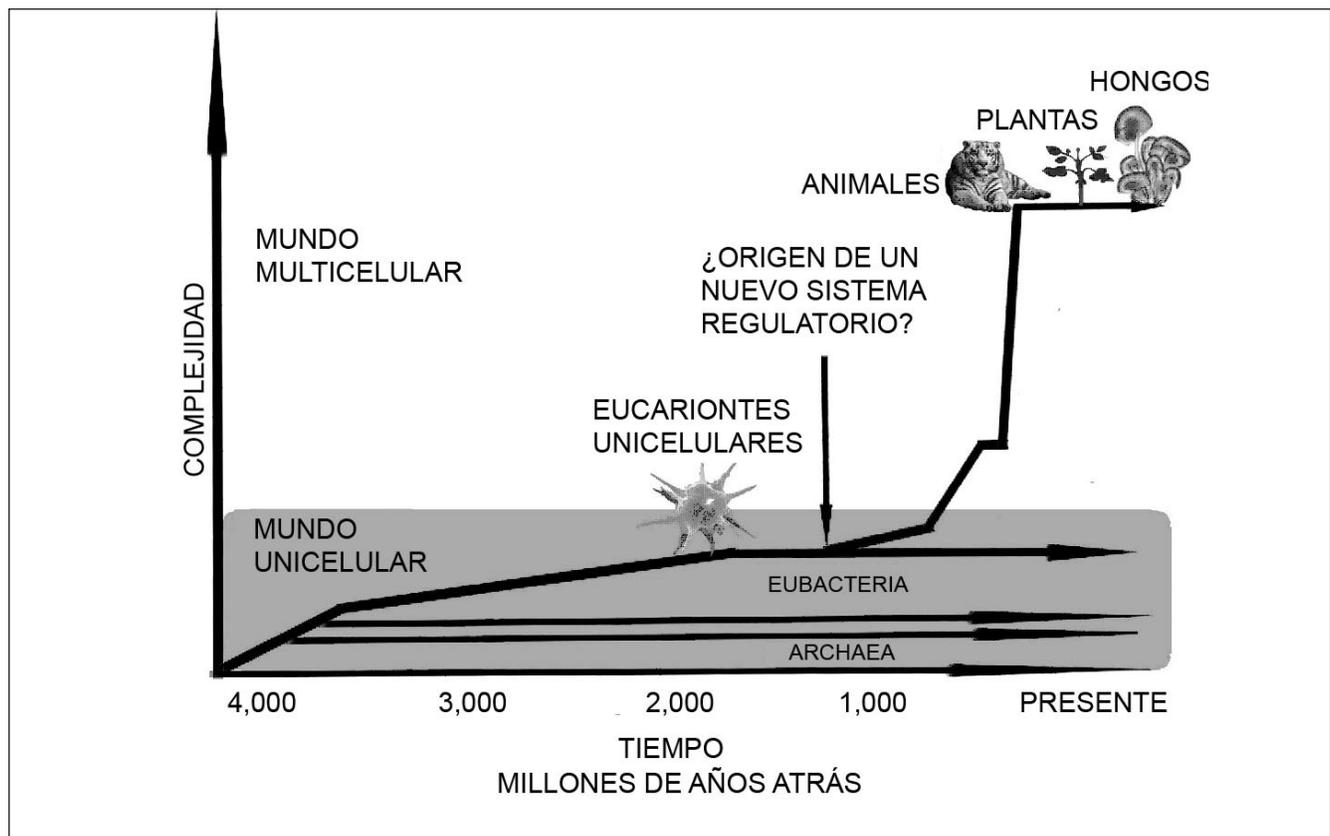
Existen numerosos ejemplos de que los intrones y otros ARNs no-codificadores de proteínas contienen información funcional; una subclase interesante de estos son los ARNs nucleolares pequeños o snoARNs producidos a partir de ARNs intrónicos derivados de genes que codifican proteínas ribosomales y nucleolares, así como de otros genes cuyos exones han perdido ya la capacidad de codificar proteínas. Otro ejemplo lo constituyen los ARNs temporales pequeños *lin-4* y *let-7* que controlan la temporalidad del

desarrollo en *C. elegans* a través de interacciones ARN-ARN que afectan la traducción y estabilidad de otros transcritos; *let-7* ha sido conservado a través de invertebrados y vertebrados. Estos ARNs pequeños se derivan de precursores mayores y tienen un tamaño de alrededor de 22 nucleótidos, similar a los ARN producidos por los procesos de iARN (interferencia de ARN).

Parece seguro predecir que la vasta mayoría de los ARNs no-codificadores no ha sido aún catalogada, puesto que la mayor parte de los rastreos genómicos han tenido un sesgo intrínseco contra ellos. Cada vez es más obvio que existe un papel central para la señalización y el metabolismo del ARN en la biología de los eucariontes.

Hay un número de fenómenos genéticos insuficientemente comprendidos (como la inhibición por ARN, la co-supresión, el silenciamiento transgénico, los efectos posicionales abigarrados, la improntación, la metilación del ADN, la compensación de dosis cromosómica y la transvección) que comparten características.

Figura 5. La vida unicelular, representada principalmente por organismos procariontes, controló la tierra por billones de años



Cuando apareció la vida multicelular, sin embargo, su complejidad aumentó de manera sorprendente. La evolución de un sistema de regulación genética adicional podría muy bien explicar tanto la evolución a la multicelularidad como la rápida diversificación y la creciente complejidad.

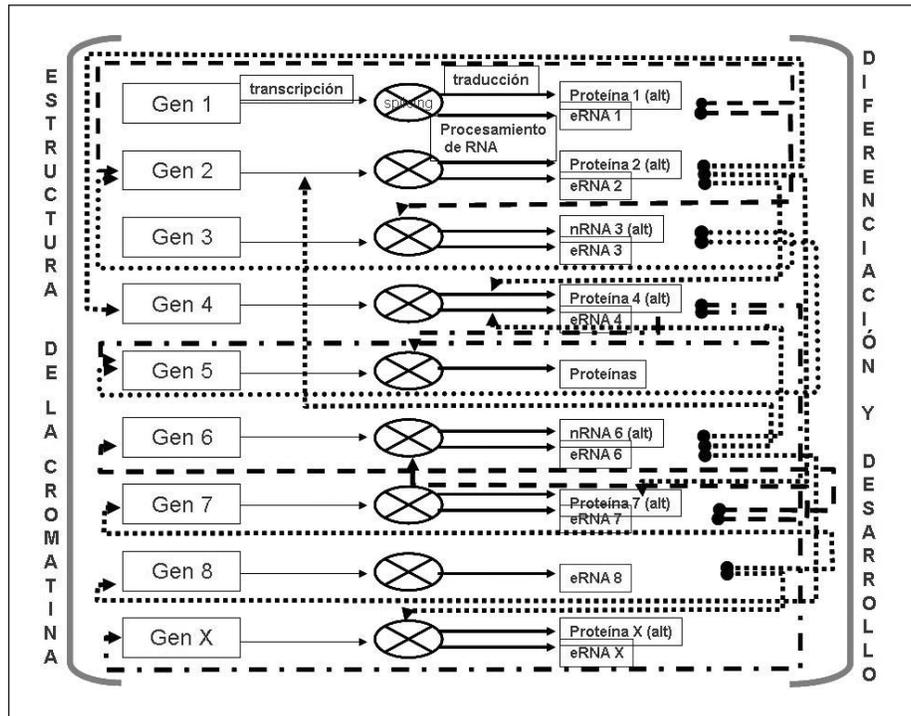
Sin abundar en detalles, se ha demostrado que en todos ellos existe señalización de ARN, ya sea mediante interacciones ARN-ARN, ARN-ADN, así como remodelación de la cromatina. La interferencia de ARN (iARN) y el silenciamiento de genes post-transcripcional en plantas y animales es mediado por ARNs de 21 a 22 nucleótidos, generados por la ARNasa III a partir de cadenas dobles más largas de ARN; esta longitud es similar a la requerida por la metilación del ADN dirigida por ARN.

Hay evidencia también de que el ARN regula la arquitectura de la cromatina (HP1, CDH, histona-acetiltransferasa MOF) y la metilación del ADN, por lo menos en plantas y probablemente en todos los eucariontes superiores. Además, un ARN no-codificador actúa como co-activador transcripcional de los receptores para esteroides, y esta acción también requiere de remodelación de la cromatina y del reclutamiento de histona-acetiltransferasas.

Visto como un todo, estas observaciones sugieren que una compleja red de señalización basada en ARN opera en los eucariontes superiores, lo que permite la comunicación directa gen a gen y la integración y regulación de la actividad génica a muchos niveles, incluyendo la estructura de la cromatina, la metilación del ADN, la transcripción, el empalme y procesamiento del ARN, la traducción del ARN, la estabilidad del ARN y la señalización de ARN en varias vías (ver Figura 6).

Si estas consideraciones son correctas, para entender la biología de los organismos superiores se requerirá entender no solamente el proteoma, que actualmente es el foco de mucha investigación, sino identificar también todos los ARNs no-codificadores y sus patrones de expresión, señalización y procesamiento, puesto que estas evidencias sugieren que, lejos de

Figura 6. Esquema detallado del papel propuesto para los eARN en los sistemas de control e interconexión de los eucariontes



Los genes, empaquetados en la cromatina, expresan transcritos primarios que son luego ensamblados de modo alternativo para producir mRNA y/o intrones, que pueden luego ser reprocesados para formar moléculas menores, como *-let-7*. Algunos genes ARN no-codificadores pueden producir ARNs tanto de intrones como de exones (nARN), éstos pueden luego actuar como señales o moléculas guía para integrar la actividad de este locus con la de otras partes de la red, vía efectos en la estructura de la cromatina, transcripción, ensamble, otros niveles de proceso de ARN, traducción de mRNA, estabilización de mRNA y otras formas de señalización mediadas por ARN dentro de la célula. La evidencia indica que muchas de estas interacciones dependen de homología (secuencia directa) e involucran uniones ARN-ADN, ARN-ARN y ARN-proteínas, otras podrían involucrar estructuras secundarias o terciarias de ARN o catálisis mediada por ARN. Este esquema no pretende ser comprensivo, sino solamente dar idea de la complejidad o potencial de estas redes de interacción informática entre los genes y la célula, y su papel en la diferenciación y el desarrollo.

ser basura o residuos evolutivos, los intrones y otros ARNs no-codificadores forman la arquitectura primaria de control que sostiene la diferenciación y el desarrollo de los organismos superiores.⁴

Conclusiones

- El ADN “basura o chatarra” regula la traducción de las isoformas de las proteínas.
- El ADN “basura o chatarra” influencia el comportamiento de los genes mediante la transcripción de múltiples tipos de ARN.
- El ADN “basura o chatarra” tiene tanto un papel reforzador como silenciador frente a genes adyacentes.
- El concepto simple expresado por el dogma central no parece ser aplicable a los organismos superiores.

- El ARN en sus múltiples formas es parte de un sofisticado mecanismo de control de la expresión génica y un integrante indispensable para lograr la diversidad entre los organismos.

Referencias bibliográficas:

1. Mattick J.S. The hidden genetic program of complex organisms, *Scientific American* 2004, 291(4); 60-67.
2. Venter, J. Craig, et al. The sequence of the human genome. *Science* 2001; 291: 1304-1351.
3. Szymanski M, Barciszewski J. Beyond the proteome: non-coding regulatory ARNs. *Genome Biology* 2002, 3(5): reviews 0005.1-0005.8.
4. Mattick JS. Non-coding ARNs: The architects of eukaryotic complexity. *EMBO reports* 2001, 2(1); 986-991.
5. Mattick J.S., Gagen M.J., The evolution of controlled multi-tasking gene networks: The role of introns and other non-coding ARNs in the development of complex organisms. *Mol Biol Evol* 2001, 18 (9); 1611-1630.

Correspondencia:
Dr. Víctor Javier Lara Díaz
Email: lara-diaz.vj@itesm.mx