

# HALLAZGOS RECIENTES SOBRE LA ESTRUCTURA Y FUNCIÓN DEL GEN

<sup>1</sup>BORNACELLI A., MD, CARABALLO L., MD, PhD

*Instituto de Investigaciones Inmunológicas  
Universidad de Cartagena - Cartagena, Colombia*

## RESUMEN

Definir un gen es una tarea difícil y cada vez más complicada. Dado su interés en muchas disciplinas, algunos pensamos que debe reflejar un concepto global que incluya la mayoría de los elementos estructurales y funcionales que lo componen. En esta revisión se analizan los hallazgos que han ayudado a construir un nuevo concepto de gen, enfatizando en los descubrimientos recientes que han obligado a modificar la definición tradicional de gen, resumida como “secuencia de ADN que codifica la información para la síntesis de proteínas o ARN”. Entre esos hallazgos están el “splicing” alternativo, la recombinación somática en los genes de las inmunoglobulinas, la regulación de la expresión génica, ya sea inter- o intra-cromosómica, por secuencias reguladoras, por mecanismos epigenéticos o por ARN; la naturaleza de los transcritos según el sentido de la transcripción, los transcritos quiméricos, la repetición de exones y la herencia extragenómica. Integrando estos fenómenos y su impacto funcional podemos definir los genes como un conjunto de secuencias de ácidos nucleicos que determinan la expresión y regulación de una característica heredable.

**Palabras clave:** genes, regulación genética, función del gen, genómica

## RECENT FINDINGS ON THE STRUCTURE AND FUNCTION OF THE GENE

### ABSTRACT

Defining a gene, a matter of interest of many disciplines, is a difficult and increasingly complicated task. Some authors think it should reflect a global concept that includes most of the structural and functional elements that compose it. In this paper, we analyze the findings that have helped to build a new concept of the gene, focusing on recent discoveries that have forced us to change the traditional definition of gene, summarized as “AND A sequence that encodes information for the synthesis of proteins or RNA”. Among these findings are alternative splicing, somatic recombination of immunoglobulin genes, regulation of gene expression at inter or intra-chromosomal levels, regulation by regulators sequences, epigenetic mechanisms or RNA; the sense and nature of transcripts, quimerism transcript, exon repetition and extragenomic inheritance. Taking into account these findings and their functional impacts we can define genes as a set of nucleic acids sequences that determine the regulation and expression of an inheritable feature.

**Keywords:** genes, gene regulation, gene expression, genomic.

• \*Correspondencia: xxxxxxxxx  
Fecha de recepción: 1 de junio de 2009 - Fecha de aceptación: enero de 2010

## INTRODUCCIÓN

Uno de los conceptos que con frecuencia se ha modificado en biología es el de gen y cada vez es más difícil describirlo con precisión. En general se acepta que gen es “una secuencia de ácido desoxiribonucleico (ADN) que codifica la información para la síntesis de proteínas o ácido ribonucleico (ARN)”, pero los nuevos descubrimientos en biología celular y molecular indican que esta definición debe revisarse [1].

La gran conservación de secuencias génicas entre organismos distantes filogenéticamente, la ausencia de correlación entre la complejidad de los organismos y el número de genes codificadores de proteínas (paradoja del valor C), así como los variados y refinados procesos de regulación de la expresión génica han modificado radicalmente el concepto de información genética. E. Pennisi lo comenta de esta manera: “Los genes, piedra angular del desarrollo y funcionamiento de los organismos, no pueden explicar por sí solos qué hace a las vacas vacas y maíz al maíz. Los mismos genes se han manifestado en organismos tan diferentes como ratón y medusa. Es más, nuevos hallazgos de varios investigadores han puesto en claro que es el exquisito control por el genoma de la actividad de cada gen y no los genes per se lo que más importa” [2]. En esta revisión discutiremos algunos aspectos que han llevado a la visión contemporánea de la estructura y funcionamiento de los de genes.

### 1. Evolución del concepto de gen

Las primeras teorías sobre la herencia fueron enunciadas por Hipócrates (460-370 a.C.) para quien existía una especie de semillas repartidas por el cuerpo que se transmitían a los hijos en el momento de la concepción, haciendo que éstos se parecieran a sus padres. Es claro que en ese momento solo se concibe una idea acerca de que la información genética se transmite por herencia. Aristóteles rechazó estas teorías y propuso que el semen de los machos podía contener partículas heredadas de generaciones pasadas. Varios siglos después el monje austriaco Gregor Mendel, trabajando con guisantes, percibiría la existencia de elementos que se transmiten de una generación a la otra y que se redistribuyen en cada generación.

En 1869, Federico Miescher obtuvo un precipitado grisáceo a partir de glóbulos blancos tratados con ácido clorhídrico y lo denominó “nucleína” actualmente conocida

como ADN. En 1882 el alemán Walther Flemming descubre una estructura que absorbe fuertemente los tintes de anilina a la que denomina consecuentemente cromatina, la cual hace parte de los cromosomas [3], pero todavía no se integran estos conocimientos; es alrededor de 1892 cuando se logra una idea menos general de lo que es la herencia y se enuncia que ésta es transmitida por una sustancia de constitución molecular.

Posterior al redescubrimiento de las leyes de Mendel, Thomas Hunt Morgan y colaboradores logran integrar el conocimiento acerca de los cromosomas y lo que se conoce como gen al establecer la teoría cromosómica de la herencia, que propone que los factores hereditarios se disponen de forma lineal en los cromosomas.

Varios investigadores corroboraron esta idea y además propusieron que aquellos genes que se encontraban en el mismo cromosoma se heredaban juntos, por lo que se les llamó genes ligados. Esto se sustentaba en las definiciones de Wilhelm Ludvig Johannsen, quien además de proponer que cada porción del cromosoma que controla una característica se llamara gen (del griego “dar a luz”), introdujo los términos genotipo y fenotipo. Estos avances conceptuales se consolidaron cuando Oswald Avery y sus colaboradores, utilizando bacterias muestran que el material que transmite la información genética de la célula es el ADN. Esta molécula cobra entonces gran importancia y en 1953 James Watson y Francis Crick elaboran su modelo agregando elementos que sustentan mejor la visión de los genes y su participación en la transmisión de la herencia.

Un aspecto adicional muy importante ha sido la evolución de la idea de George Beadle y Edward Tatum respecto a que un gen codifica para una enzima (un gen-una enzima), la cual se amplió estableciendo que el resultado de la codificación de un gen es la producción de una proteína [1]. Este pensamiento lo modifican Richard Roberts y Phillips Sharp, quienes trabajando en la expresión de genes de adenovirus descubren que un gen puede codificar para varias proteínas y se empiezan a buscar nuevos elementos que contribuyan a la diversidad proteica. De esta manera cobran importancia las interacciones entre genes y otros aspectos que comentaremos a continuación. A pocos años de iniciarse la secuenciación del genoma humano se propone que un gen es un segmento cromosómico completo responsable de un producto funcional [4].

## 2. Descubrimientos que han modificado el concepto de gen

En la década del 70 con el descubrimiento de la recombinación somática en los genes de las inmunoglobulinas [5] y con los análisis iniciales del “*Splicing*” alternativo por Roberts y Sharp [6] comenzaron a entenderse varios mecanismos sobre el funcionamiento de los genes; entre otros, la regulación de la expresión génica, la composición molecular de algunos genes y la herencia “*extra genómica*” (Tabla 1).

### 2.1. “*Splicing*” alternativo

El “*Splicing*” o empalme alternativo en el ARN mensajero (ARNm) es el proceso donde se cortan los intrones y se empalman los exones del preARNm para generar un ARNm que se traducirá a una proteína. Este es uno de los fenómenos que más ha inducido a cambiar la idea de gen, especialmente porque en una revisión realizada por Modrek y colaboradores. Se estimó que al menos del 40-60% de los genes humanos tienen formas de “*splicing*” alternativo, lo que sugiere que es uno de los procesos más significativos que generan complejidad funcional del genoma humano [7].

En el trabajo de J. Johnson, quien monitoreó el splicing en 10.000 genes humanos multi-exónicos de 52 tejidos diferentes utilizando “*microarrays*”, se demostró que el 74% son empalmados alternativamente lo que genera la gran diversidad de proteínas existente aunque solo un 3% del genoma sea codificador, lo que terminó por derribar la noción de que un gen codificaba solamente una proteína [8].

Wang y colaboradores analizaron las isoformas generadas por “*splicing*” alternativo en 15 tejidos humanos diferentes y los transcriptomas por líneas celulares con la secuenciación del ADN complementario (ADNc), describiendo que alrededor del 92% de los genes humanos presentan “*splicing*” alternativo [9] y recientemente Barash y colaboradores agregan importante y novedosa información sobre este fenómeno y su especificidad en células y tejidos [10].

Para el empalme alternativo P. Shepard y K. J. Hertel [11] describieron la necesidad de que haya flexibilidad del spliceosoma para identificar los exones que han de ser empalmados en un determinado pre-ARNm, lo cual depende de parámetros como la fuerza del sitio que va a ser cortado, los reguladores del empalme, la arquitectura de los exones e intrones y el proceso de síntesis del mismo

pre-mARN. Algunos estudios muestran que los patrones de empalme alternativo son dependientes de la especie, en una publicación donde se comparó la conservación de las isoformas generadas por este mecanismo en 166 pares de genes ortólogos en el genoma humano y de ratón, se encontró que la mitad de los genes analizados son específicos para cada especie [12].

Barash y colaboradores estudiaron este proceso en cuatro tejidos de ratón encontrando que el patrón de empalme es diferente en tejidos adultos y tejido embrionario, y además elaboraron un modelo informático que permite predecir cómo será el empalme en diferentes tejidos dependiendo de la secuencia de ADN y de los elementos *cis* y *trans* que regulen el proceso. Este código de empalmes, como lo denominan los autores, abre nuevas expectativas acerca de cómo es la regulación de este fenómeno y necesariamente influirá en la conceptualización del gen [10]. El conocimiento de cómo las células producen ARNm alternativos es esencial para comprender el funcionamiento y regulación del genoma [13], además este tipo de estudio induce a futuras investigaciones que ayudarán a conocer mejor la fisiología molecular de muchas enfermedades y de la evolución misma.

### 2.2. Recombinación somática

Este es un fenómeno específico de los genes que codifican para las inmunoglobulinas (o también llamados anticuerpos) y lleva a la generación de la diversidad de anticuerpos necesaria para responder a casi cualquier antígeno. Ocurre durante la maduración de los linfocitos B y consiste en que estos genes presentan re-arreglos y recombinaciones que codifican los dominios variables y a veces los constantes de las inmunoglobulinas.

El genoma contiene múltiples versiones ligeramente distintas que se combinan entre sí al azar dando como resultado la diversidad de anticuerpos. La primera evidencia de la recombinación somática fue descrita por Hozumi y Tonegawa [5], quienes trabajando en células de mieloma en ratones Balb/c encontraron que genes distantes codificadores para las regiones constante y variable de la cadena Kappa de las inmunoglobulinas se unen produciendo distintas formas de anticuerpos durante la diferenciación de los linfocitos. Este hallazgo, aunque no aplicable a todos los genes, ha ayudado a entender cómo unos pocos generan una gran diversidad proteica.

### 2.3. Regulación de la expresión génica

La expresión de los genes está regulada por diversos mecanismos coordinados en tiempo y espacio [14], en lo que influye la organización dinámica de los cromosomas, la cual varía durante las etapas del ciclo celular y entre diferentes tipos de células, determinando cambios en la replicación y la transcripción del ADN [15].

#### 2.3.1. Descubrimiento de las secuencias reguladoras

Las secuencias reguladoras juegan un papel fundamental en el control de la expresión de los genes y, en consecuencia, su descubrimiento ha sido uno de los hechos más influyentes en la definición estructural de los genes, ayudando además a comprender su funcionamiento. Su identificación la iniciaron Queen y Baltimore, en 1983 [16], mientras trabajaban en la expresión del gen que codifica para la cadena ligera  $\kappa$  (*kappa*) de las inmunoglobulinas, en células de mieloma murino. Allí descubrieron que una secuencia ubicada a 2,6 Kilobases (Kb), corriente abajo de la región promotora, es necesaria para la transcripción.

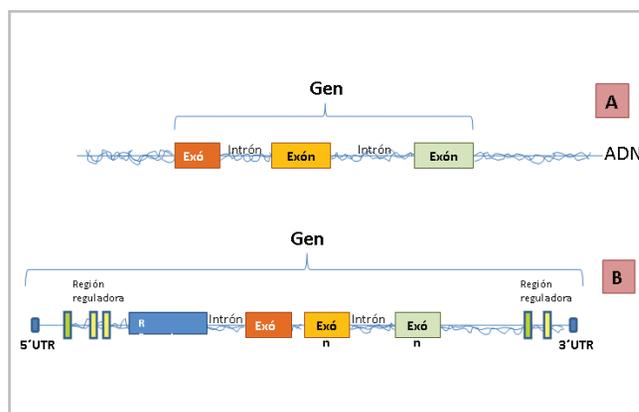
Las secuencias reguladoras son de diversa naturaleza; pueden estar cerca o lejos de la región promotora, pueden activar o reprimir la expresión de los genes dependiendo de los factores de transcripción que unen y de las interacciones con otras secuencias. Un tipo particular son los “*insulators*” (secuencias insulares o de insulación), secuencias que aíslan genes de una regulación promiscua por secuencias activadoras o represoras en regiones cercanas, y dependiendo de su ubicación respecto al gen que aíslan, actúan como activadoras o represoras, modificando la estructura de la cromatina y creando regiones fuera del nucleosoma. Una revisión reciente sobre las propiedades de estas secuencias propone que los “*insulators*” son elementos derivados de las regiones promotoras que se han conservado evolutivamente por sus propiedades específicas [17].

Saber que las secuencias reguladoras son un componente de la estructura de los genes complica los intentos de delimitar el gen, ya que dichas secuencias pueden estar cerca o lejos de éste, en el mismo cromosoma o incluso en otro. Esta dificultad la superan algunos autores separando el concepto estructural del regulatorio, y denominar al primero gen y al segundo “*genon*” [18].

#### 2.3.2. Regulación intra-cromosómica

Este tipo de regulación puede ser intra-génica, cuando

una región reguladora ubicada en un mismo gen modula la unión de factores de transcripción a las regiones promotoras, o intergénica cuando dicha región se encuentra en otro gen pero en el mismo cromosoma. Estos procesos desafían concepciones previas del gen, tanto desde el punto de vista funcional como estructural al presentarlo, no como un elemento aislado, sino como una estructura dinámica capaz de interactuar con secuencias cercanas o distantes ubicadas en un mismo cromosoma, las cuales regulan, ya sea de manera positiva o negativa, su expresión. A nivel estructural se amplía la idea de un gen como un segmento de ADN constituido por intrones y exones, para concebirlo como una estructura que además contiene secuencias reguladoras y regiones promotoras (ver figura 1).

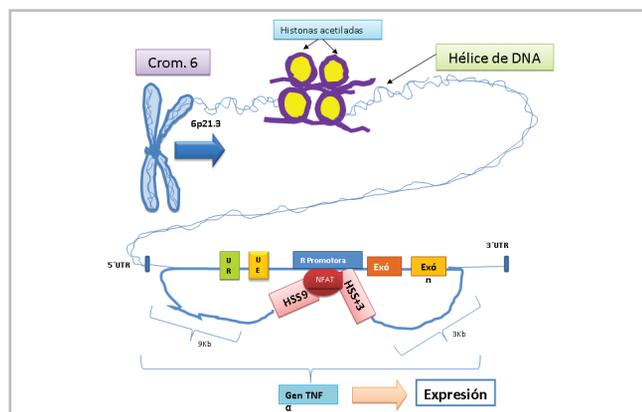


**Figura 1.** Esquema de un gen. A, estructura simple concebida como secuencia de ADN constituida por exones e intrones; B, estructura que incluye regiones reguladoras y promotoras.

El grupo de Spilianakis [19] describe la regulación intra-cromosómica como un fenómeno frecuente que modula la expresión de distintas citocinas durante la respuesta inmunitaria dependiente de linfocitos T de tipo Th2. Un ejemplo es la activación del gen del Factor de Necrosis Tumoral alfa (TNF $\alpha$ ), que requiere dos elementos de hipersensibilidad a DNAasa, llamados HSS 9 y HSS 3, ubicados a 9 Kb corriente arriba, y a 3 Kb corriente abajo del primer exón del gen TNF $\alpha$  respectivamente, altamente conservados en ratón y en humanos.

Dichos segmentos se unen al NFAT (factor activador nuclear de células T, del inglés-*Nuclear Factor of Activated T cell-*) funcionando como secuencias aumentadoras (o amplificadoras) de la transcripción, mediada por la región promotora del TNF $\alpha$ . Los autores demostraron que la activación de este gen en las células T ocurre de manera intracromosómi-

ca en forma de bucle mostrando un complejo en estrecha proximidad entre las secuencias aumentadoras y la región promotora [20] (ver figura 2). Otro ejemplo lo tenemos en el gen de la cadena *Kappa* de las inmunoglobulinas del ratón, donde a través de técnicas de captura de conformación de cromosomas (3C), se observó cómo una secuencia aumentadora (“*enhancer*”) intrónica interacciona con dos secuencias “*enhancers*” ubicadas corriente abajo del sitio de inicio de la transcripción y la región promotora, estimulando la expresión del gen [21].

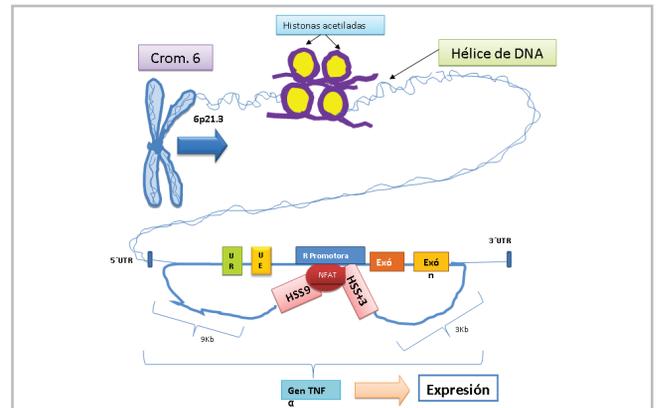


**Figura 2.** Modelo de regulación intra-cromosómica del gen del TNF . Las secuencias *Hss9* y *Hss3* actúan como secuencias aumentadoras (“*enhancers*”) sobre la expresión del gen TNF , interactuando en forma de lazo con el complejo factor de transcripción (NFAT) y la región promotora en el ratón.

**2.3.3. Regulación inter-cromosómica**

Este proceso ocurre cuando un gen es regulado por secuencias que se encuentran en otro cromosoma. Un ejemplo lo muestra el trabajo de Spilianakis y colaboradores, quienes observaron en ratones las interacciones entre el gen que codifica para el interferón gamma ( $IFN\gamma$ ) (ubicado en el cromosoma 10) y el locus génico para las interleucinas Th2 (Interleucinas-4, -5 y -13: IL4, IL5, IL13; ubicados en el cromosoma 11).

Estos autores sugieren que la interacción es mediada por dos regiones conservadas denominadas CNS1 y CNS2 ubicadas en el cromosoma 10, la primera a 5 kb corriente arriba del sitio de inicio de la transcripción del gen que codifica para el  $IFN\gamma$ , y la segunda a 18 kb corriente abajo. Estas regiones al parecer interaccionan con la región promotora (Rad 50) del gen que codifica para la IL5 y con una zona de hipersensibilidad a DNAasa denominada RHS6 en el cromosoma 11, aumentando la expresión del gen del  $IFN\gamma$  [19] (ver figura 3).



**Figura 3.** Interacciones inter-cromosómicas entre el gen del *IFN* (ubicado en el cromosoma 10) y el locus para las interleucinas Th2 (IL 4, IL5, IL13, ubicado en el cromosoma 11) en el ratón.

**2.3.4. Regulación epigenética**

Incluye las modificaciones de la cromatina e histonas que intervienen en la expresión de genes sin modificar la secuencia de DNA. Entre ellas se encuentran la metilación de la citosina del ADN, en la que se transfiere un grupo metilo a la posición C-5 de la citosina por acción de ADN(Citosina)-5-metiltransferasa, en di nucleótidos CpG(dinucleótico covalentemente unido con un enlace fosfodiéster) [22].

Las histonas son modificadas químicamente por distintas vías, incluyendo acetilaciones, metilaciones, ubiquitinizaciones y fosforilaciones, que favorecen la conformación de una cromatina abierta y disponible para la transcripción (eucromatina) o, por el contrario, la adición de grupos metilos que inducen una forma densa de la cromatina (heterocromatina) inactiva para transcribirse.

Se ha propuesto, además, que estos estados modificados se extienden a lo largo de los cromosomas y se transmiten durante la replicación [23]. A pesar de que las modificaciones epigenéticas no contemplan cambios directos sobre el gen, deben tenerse en cuenta cuando se analizan los aspectos funcionales de este.

**2.3.5. Regulación por ARN**

El control de la expresión génica por el ARN amplía las propiedades funcionales del gen. Por una parte, al existir transcritos de ARN que no se traducen en una proteína, se disipa la idea de los genes como secuencias de ADN que siempre llevan a codificar proteínas. De otro lado, al conocerse una regulación de la expresión génica por parte de pequeños transcritos de ARN, el concepto de regulación adquiere otra dimensión al tener en cuenta un conjunto

complejo de interacciones del orden ADN-pequeños transcritos. En estas interacciones participan los pequeños ARN de transferencia (siARN por sus siglas en inglés *small interfering RNA*), los micro ARN (miARN), los pequeños ARN nucleares y los ARNs asociados a la familia de proteínas argonautas y especialmente a una de sus subclases, llamadas proteínas piwi (piARN) [24].

El mecanismo básico de interferencia para los siARNs es la inducción de la destrucción de un ARNm blanco mientras que para los miARNs es la inducción de un silenciamiento temporal del ARNm diana [25]. Los piARN (mediante las proteínas argonautas Ago 1 y Ago 2 que forman complejos ribonucleoproteicos al asociarse a miARN y siARN) se unen a un ARNm reprimiendo su expresión; sin embargo, no todas las clases de proteínas argonautas aparecen asociadas a siARN o miARN. Otro elemento importante son los pequeños ARN nucleares (snARN) que regulan el empalme del pre mARN. Los piARNs fueron descritos por Richard Carthew [24], quien detectó que además de los tipos conocidos de ARN habían evidencias sobre otro tipo, contribuyente a la regulación genética en células de líneas germinales, asociados a proteínas Piwi y con una actividad ADN helicasa dependiente de ATP, similar a la proteína RecQ1, relacionada con el silenciamiento de genes [26] [27].

La inhibición de la expresión de genes por transcritos no codificantes es debida en parte a su interacción con las regiones promotoras, lo que parece ocurrir ampliamente y contribuye a la variabilidad de proteínas y por consiguiente a la diferenciación entre individuos.

Un ejemplo lo aportan Martianov y colaboradores, quienes encontraron que en células inactivas la región promotora mayor del gen que codifica para la enzima Dihidrofolato-reductasa es reprimida por un ARN no codificante, transcrito de la región promotora menor, el cual forma un complejo con el factor de transcripción TFIID y disocia el complejo de preiniciación en la región promotora mayor [28].

### 2.3.6. Naturaleza y sentido de los transcritos

Es sabido que los genes se transcriben en ambas direcciones, una cadena en un sentido y la otra en sentido contrario; sin embargo, publicaciones recientes describen que una misma cadena puede transcribirse simultáneamente en ambos sentidos. En una publicación reciente [29] se describe que los transcritos generados en el sentido tradicional

5'→3' traducen proteínas, mientras que los transcritos en sentido contrario (anti-sentido) se asocian con funciones reguladoras (ARNs largos y cortos reguladores).

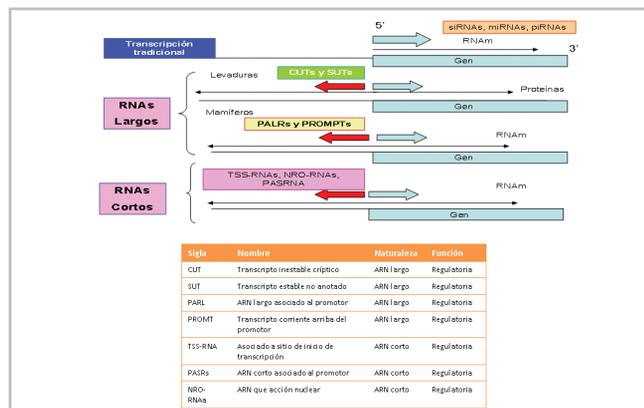
Se sabe además que en las levaduras [30] se transcriben ARNs largos no codificantes inestables denominados CUTs (del inglés-*cryptic unstable transcript*-) corriente arriba de la región promotora [31], similar a los transcritos denominados PROMPT (del inglés-*promoter upstream transcripts*-) en células humanas, descritos por Preker [32].

Estos últimos son ARNs largos, poliadenilados y altamente inestables que se producen corriente arriba del sitio de inicio de la transcripción (figura 4), y se transcriben en ambos sentidos respecto al gen corriente abajo. Hallazgos similares los obtuvieron Seila y colaboradores [33], quienes investigaron la expresión de ARNs pequeños (20 - 90 nucleótidos de largo) localizados cerca del sitio de inicio de la transcripción de genes que codifican para proteínas en células madres embrionarias de ratón.

Ellos detectaron los ARNs en todas las células estudiadas y además en las células madres sin la enzima *Dicer*, necesaria para el procesamiento de los microARNs, lo que sugiere que no son productos de esta vía. También se encontró que la transcripción de estos ARNs, denominados TSS-RNAs (del inglés-*transcription start site-RNAs*-), tiene orientación tanto corriente arriba del sitio de inicio de la transcripción como a 50 nucleótidos corriente abajo del mismo y no depende de la ubicación de otros genes vecinos o de genes con múltiples sitios de inicio de la transcripción. Estos avances científicos han retroalimentado el conocimiento sobre la funcionalidad de los genes al descubrir que una misma cadena de ADN puede transcribirse simultáneamente en ambos sentidos y mostrar que además de ARNs pequeños existen ARNs largos con funciones reguladoras (ver figura 4).

### 2.4. Transcritos quiméricos

Otro punto importante son los transcritos originados por fusión de genes y denominados transcritos quiméricos, los cuales se derivan de la transcripción de dos genes consecutivos en una sola molécula de ARNm. Para esto el empalme en el pre-mARN incluye un sitio donador en el extremo 5' del gen corriente arriba y un sitio aceptor en el extremo 3' del gen corriente abajo, removiendo de esta manera la región de terminación de la transcripción en el gen corriente arriba ("*upstream*"), la región de regulación e inicio de



**Figura 4.** Se muestra que en levaduras y en mamíferos se producen transcritos bidireccionales, corriente arriba de la región promotora del gen. Estos pueden ser moléculas de ARN largas o cortas; los transcritos en sentido contrario se asocian a funciones reguladoras mientras los transcritos con sentido se asocian a proteínas funcionales.

los exones de ambos genes [34]. El grupo de Parra estima que el 4-5% de los genes se transcriben de esta manera, siendo este uno de los mecanismos generadores de mayor variedad de proteínas [35]. Según ellos, las proteínas empiezan a ser codificadas por una secuencia de ADN asociada y luego corre hacia el gen que va a completar su diferenciación; lo que constituiría una transcripción combinada que contribuye a la diversidad de las proteínas. Algunos autores opinan que existen proteínas que son codificadas por exones de una parte del genoma combinados con otros distantes, lo que hace parte de la interacción inter-cromosómica y también genera diversidad proteica. Esto modifica también la definición de gen porque la proteína no sería codificada por un gen sino por varios. También se han descrito micro-ARN quiméricos que son precursores de ARNm [36].

**Tabla 1.** Resumen de hallazgos sobresalientes sobre la estructura y función del gen.

Referencia	Hallazgo	Impacto
6, 7, 8, 9, 10, 11, 12, 13	Empalme alternativo	Un gen genera varios transcritos y puede codificar varias proteínas.
5	Recombinación somática	Pocos genes pueden generar un gran repertorio de anticuerpos por recombinación génica.
14	Descubrimiento de las secuencias reguladoras	Amplía el concepto anatómico de los genes. La estructura de un gen incluye sus secuencias reguladoras.
20	Regulación intra-cromosómica	Un gen no es un elemento aislado, para su expresión interacciona con secuencias reguladoras ubicadas cerca o lejos de su región promotora, pero en el mismo cromosoma.
19	Regulación inter-cromosómica	Los genes son elementos dinámicos que interaccionan entre sí para una función, aún estando en cromosomas distintos.
22, 23	Regulación epigenética	La función de los genes depende de las modificaciones y regulaciones a nivel de la cromatina e histonas.
24, 25, 26, 27	Pequeños transcritos no codificadores-regulación por ARN	Los genes, además de generar productos codificadores, generan transcritos no codificadores que tienen función reguladora de la expresión génica.
29, 30, 31, 32	Naturaleza de los transcritos	Una misma cadena de ADN puede transcribir múltiples productos de forma bidireccional. Los transcritos antisentido se asocian con funciones reguladoras mientras que los transcritos con sentido codifican proteínas funcionales; existen además ARNs largos reguladores.
34, 35, 36	Transcritos quiméricos o por fusión de genes.	Las proteínas pueden ser producidas por fusión de varios genes, lo que genera una mayor variedad de proteínas y de transcritos no codificadores que surgen del mismo proceso.
37, 38	Herencia extra-genómica	La herencia puede ser transmitida por una molécula de ARN.
40, 41	Repetición de exones	Es variable la expresión de un gen, ya que no solo genera una proteína sino que puede generar varias al mismo tiempo por, re-arreglo en los exones del ARNm.
4, 10, 13, 42	Proyecto genoma humano	Ofrece información en bases de datos públicas junto al mejoramiento en técnicas de biología celular y molecular.

la transcripción en el gen corriente abajo (“downstream”), la región intergénica y exones (dependiendo del patrón de empalme) para al final generar un transcripto maduro con

**2.5. Herencia extra-genómica**

En un estudio efectuado por Lolle y colaboradores [37] se presenta un desafío a la idea de que el ADN transmite

información de una generación a la siguiente. Los autores muestran que plantas *Arabidopsis* homocigotas para un alelo mutante recesivo del gen HOTHEAD (*hth*) pueden heredar secuencias alélicas específicas que no estaban presentes en el genoma de sus padres, pero sí en el de generaciones previas, siendo este posiblemente un mecanismo general para la herencia del ADN extra-genómico.

Los autores sugieren que la reversión en el fenotipo de las plantas se debe a la existencia de una ARN caché, denominado en esta forma porque actuaría de manera similar como lo hace la memoria caché de un computador, guardando la información normal de un ADN ancestral que no se encuentra en las plantas padres, pero que puede heredarse a las plantas hijas con una frecuencia que oscila entre el 8 y el 16%; al estudiar el genotipo de las plantas que revertían la mutación encontraron que el 100% de las plantas eran heterocigotos (HTH/*hth*) (ver figura 5). Este fenómeno también puede suceder en animales, como lo describe Rassoulzadegan en investigaciones parecidas. Ratones descendientes de padres que poseían un gen mutado no heredaron la mutación, ya que tenían información del gen normal en la secuencia de ARN; esta información corrigió la información anormal [38].

## 2.6. Repetición de exones

Este mecanismo, que al igual que el empalme alternativo genera diversidad de proteínas, fue descrito en dos genes de ratón, el gen SA (codificante de la Sintetasa de ácidos grasos de cadena media), donde fue específico para cada tejido [39] y en el gen de la carnitina-octanoil-transferasa (COT) [40].

Este es un evento raro, descrito en menos de 10 genes, en los cuales uno o más exones están duplicados en tándem en el ARNm pero no en el gen, se denomina RREO (*rearrangements or repetition in exon order*) por sus siglas en inglés, y ocurre predominantemente en las regiones codificantes del ARNm, pero al parecer no incluye al exón 1 [41]. La repetición de exones es alelo-específica, ya que en un ARNm de una persona heterocigota solo el alelo susceptible tiene exones repetidos.

Por otra parte, la repetición se presenta solo en cis, modalidad en la cual los exones localizados en un pre-ARNm son empalmados para generar un ARNm maduro de manera contigua y lineal de orientación 5'→3', en el mismo orden como se encuentran en el gen [34].

## 2.7. Secuenciación del genoma humano

El mejoramiento de las técnicas de secuenciación y la mayor disponibilidad de bases de datos públicas han permitido conocer cada vez más genes y más transcritos, y plantear definiciones de gen que cambian en relación directa con los nuevos conocimientos.

En 2003 Snyder y Gerstein plantearon que los genes son segmentos cromosómicos completos responsables de un producto funcional y que pueden ser identificados con los siguientes criterios: marco de lectura abierto, secuencias características como los exones, secuencias de conservación, evidencia de transcripción y posibilidad de inactivación, ya sea por mutación o por ARNs de interferencia (ARNi) [4].

El continuo interés en identificar nuevos transcritos tanto codificantes como no codificantes de proteínas, ha llevado a que la definición de gen se amplíe, considerándolo como la unión de secuencias genómicas que codifican un conjunto de productos funcionales superpuestos [42]. Al cumplirse los primeros diez años de la secuenciación del genoma humano persisten numerosos interrogantes sobre las aplicaciones prácticas, especialmente las clínicas, de la gran diversidad de datos obtenidos, pero con seguridad

se han logrado aportes de aplicación inmediata al manejo conceptual del gen.

## 3. ¿Una nueva definición de gen?

Como se desprende de esta revisión, es bastante limitado referirse a los genes solo como segmentos de ADN que codifican proteínas. Hemos visto ejemplos de transmisión genética mediada por ARN y además genes de ARN que no codifican para proteínas, sino que hacen parte del andamiaje necesario para la regulación génica, como sucede con los ARNi. Por otro lado, si consideramos las secuencias reguladoras como parte del gen, tendría que aceptarse que este puede estar disperso en un cromosoma e inclusive en cromosomas distintos. Tratando de agrupar estos aspectos podríamos definir el gen como un conjunto de secuencias de ácidos nucleicos que determinan la expresión y regulación de una característica heredable. Sin embargo, expertos en la materia han propuesto separar el componente funcional del regulatorio en la definición de gen, considerando gen a la secuencia de nucleótidos codificante y "genon" a las secuencias y mecanismos regulatorios de la expresión génica [18].

## BIBLIOGRAFÍA

- Pearson, H., *Genetics: what is a gene?* Nature, 2006. 441(7092): p. 398-401.
- Pennisi, E., *Searching for the genome's second code.* Science, 2004. 306(5696): p. 632-5.
- Paweletz, N., *Walther Flemming: pioneer of mitosis research.* Nat Rev Mol Cell Biol, 2001. 2(1): p. 72-5.
- Snyder, M. and M. Gerstein, *Genomics. Defining genes in the genomics era.* Science, 2003. 300(5617): p. 258-60.
- Hozumi, N. and S. Tonegawa, *Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions.* Proc Natl Acad Sci U S A, 1976. 73(10): p. 3628-32.
- Roberts, R.J., *Intervening sequences excised in vitro.* Nature, 1978. 274(5671): p. 530.
- Modrek, B. and C. Lee, *A genomic view of alternative splicing.* Nat Genet, 2002. 30(1): p. 13-9.
- Johnson, J.M., et al., *Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays.* Science, 2003. 302(5653): p. 2141-4.
- Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. 456(7221): p. 470-6.
- Barash, Y., et al., *Deciphering the splicing code.* Nature, 2010. 465(7294): p. 53-9.
- Shepard, P.J. and K.J. Hertel, *Conserved RNA secondary structures promote alternative splicing.* Rna, 2008. 14(8): p. 1463-9.
- Nurtdinov, R.N., et al., *Low conservation of alternative splicing patterns in the human and mouse genomes.* Hum Mol Genet, 2003. 12(11): p. 1313-20.
- Tejedor, J.R. and J. Valcarcel, *Gene regulation: Breaking the second genetic code.* Nature, 2010. 465(7294): p. 45-6.
- Alberts, B., *Molecular Biology of The cell.* Fifth ed. 2008. 411-499.
- Dekker, J., et al., *Capturing chromosome conformation.* Science, 2002. 295(5558): p. 1306-11.
- Queen, C. and D. Baltimore, *Immunoglobulin gene transcription is activated by downstream sequence elements.* Cell, 1983. 33(3): p. 741-8.
- Raab, J.R. and R.T. Kamakaka, *Insulators and promoters: closer than we think.* Nat Rev Genet, 2010. 11(6): p. 439-46.
- Scherrer, K. and J. Jost, *Gene and genon concept: coding versus regulation. A conceptual and information-theoretic analysis of genetic storage and expression in the light of modern molecular biology.* Theory Biosci, 2007. 126(2-3): p. 65-113.
- Spilianakis, C.G., et al., *Interchromosomal associations between alternatively expressed loci.* Nature, 2005. 435(7042): p. 637-45.
- Tsytsykova, A.V., et al., *Activation-dependent intrachromosomal interactions formed by the TNF gene promoter and two distal enhancers.* Proc Natl Acad Sci U S A, 2007. 104(43): p. 16850-5.

21. Liu, Z. and W.T. Garrard, *Long-range interactions between three transcriptional enhancers, active  $\kappa$  gene promoters, and a 3' boundary sequence spanning 46 kilobases*. Mol Cell Biol, 2005. 25(8): p. 3220-31.
22. Lan, F. and Y. Shi, *Epigenetic regulation: methylation of histone and non-histone proteins*. Sci China C Life Sci, 2009. 52(4): p. 311-22.
23. Burgess-Beusse, B., et al., *The insulation of genes from external enhancers and silencing chromatin*. Proc Natl Acad Sci U S A, 2002. 99 Suppl 4: p. 16433-7.
24. Carthew, R.W., *Molecular biology. A new RNA dimension to genome control*. Science, 2006. 313(5785): p. 305-6.
25. Sontheimer, E.J., *Assembly and function of RNA silencing complexes*. Nat Rev Mol Cell Biol, 2005. 6(2): p. 127-38.
26. Lau, N.C., et al., *Characterization of the piRNA complex from rat testes*. Science, 2006. 313(5785): p. 363-7.
27. Aravin, A.A., G.J. Hannon, and J. Brennecke, *The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race*. Science, 2007. 318(5851): p. 761-4.
28. Martianov, I., et al., *Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript*. Nature, 2007. 445(7128): p. 666-70.
29. He, Y., et al., *The antisense transcriptomes of human cells*. Science, 2008. 322(5909): p. 1855-7.
30. Carninci, P., *Molecular biology: The long and short of RNAs*. Nature, 2009. 457(7232): p. 974-5.
31. Wyers, F., et al., *Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase*. Cell, 2005. 121(5): p. 725-37.
32. Preker, P., et al., *RNA exosome depletion reveals transcription upstream of active human promoters*. Science, 2008. 322(5909): p. 1851-4.
33. Seila, A.C., et al., *Divergent transcription from active promoters*. Science, 2008. 322(5909): p. 1849-51.
34. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing*. Annu Rev Biochem, 2003. 72: p. 291-336.
35. Parra, G., et al., *Tandem chimerism as a means to increase protein complexity in the human genome*. Genome Res, 2006. 16(1): p. 37-44.
36. Smalheiser, N.R., *EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues*. Genome Biol, 2003. 4(7): p. 403.
37. Lolle, S.J., et al., *Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis*. Nature, 2005. 434(7032): p. 505-9.
38. Rassoulzadegan, M., et al., *RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse*. Nature, 2006. 441(7092): p. 469-74.
39. Frantz, S.A., et al., *Exon repetition in mRNA*. Proc Natl Acad Sci U S A, 1999. 96(10): p. 5400-5.
40. Caudevilla, C., et al., *Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver*. Proc Natl Acad Sci U S A, 1998. 95(21): p. 12185-90.
41. Dixon, R.J., et al., *A genome-wide survey demonstrates widespread non-linear mRNA in expressed sequences from multiple species*. Nucleic Acids Res, 2005. 33(18): p. 5904-13.
42. Gerstein, M.B., et al., *What is a gene, post-ENCODE? History and updated definition*. Genome Res, 2007. 17(6): p. 669-81.