

¿QUÉ HAY EN EL ADN NO CODIFICANTE?

José Luis Gómez Skarmeta

Centro Andaluz de Biología del Desarrollo, CSIC-Universidad Pablo de Olavide, Sevilla.

Summary

Title: What is in the non-coding DNA?

More than 95% of our DNA is non-coding. This does not mean that 98% of the genome has no role at all, on the contrary. But, what is it there for? Part of this DNA contains cis-regulatory sequences that control when, where and how much a gene is transcribed. Cis-regulatory sequences can activate (enhancers) or suppress (silencers) transcription. These enhancers or silencers can be very far from the promoter they act on, or even in introns of adjacent genes. Some of them can control several neighboring genes at the same time. These sequences are therefore called Locus Control Regions. In addition, certain cis-regulatory sequences can also prevent the action of enhancers or silencers on promoters when they lie in between. These types of sequences are called insulators, and play essential roles in generating different regulatory landscapes for neighboring genes. Despite their essential function in gene regulation, and in contrast to coding sequences, their language is mostly unknown. This prevents their identification in the genome just based on their DNA sequences. Therefore, the goal for the next decade is to unravel this language and identify all cis-regulatory elements in the human genome and their implication in human diseases associated with mutations in non-coding DNA.

Conceptos

El genoma humano contiene unos 30.000 genes. Se consideran genes las regiones del genoma que codifican para proteínas, esto es, las regiones que se transcriben para generar ARN mensajeros que luego se traducen en proteínas en los ribosomas. De media, los genes tienen un tamaño de 3.000 nucleótidos o bases. Teniendo en cuenta que el total del genoma asciende a 3.164,7 millones de bases, la suma de todos los genes supone sólo un 2% del genoma. Por tanto, el 98% del genoma humano es ADN no codificante, o lo que es lo mismo, no contiene información relevante para la síntesis de proteínas.

En su gran mayoría, se desconoce la función del ADN no codificante, aunque se sabe que en él se localiza el ADN regulador. Éste consiste en regiones del ADN (también llamadas regiones reguladoras) que controlan cuánto, cuándo y dónde se produce el proceso de transcripción de un gen. Este control de la transcripción o, lo que es lo mismo, de la expresión génica, llevado a cabo por las regiones reguladoras es fundamental para la diferenciación celular. Así, a pesar de que todas las células de un individuo contengan el mismo genoma, lo que hace que un tipo celular sea

distinto de otro es el conjunto específico de genes activos que se expresan en cada célula. Para ello, en cada tipo celular se transcribe sólo una fracción de esos 30.000 genes. De este modo, tipos celulares muy distintos como, por ejemplo, una célula del páncreas y otra del corazón, activan un subconjunto de genes altamente especializado y en gran medida diferente entre ambas células y del de otros tipos celulares.

A pesar del papel fundamental del ADN regulador en el control de la expresión génica, su identificación en el genoma es extremadamente difícil. El principal motivo es que su código o lenguaje es desconocido. Esto contrasta con las regiones codificantes cuyo código genético es perfectamente conocido (por ejemplo ATG codifica para el aminoácido metionina), gracias en parte al trabajo del premio Nobel Severo Ochoa que ayudó a descifrarlo hace ya más de cuatro décadas. De hecho, la estimación actual sobre el número de genes en el genoma humano de 30.000, se basa en parte en la capacidad de predecir la existencia de genes según dicho código genético. Esto es imposible para las regiones reguladoras, lo que hace difícil actualmente predecir dónde se localizan esas regiones en el genoma y cuál es la información que contienen.

Una región reguladora está formada por un fragmento de ADN no codificante de tamaño variable (entre unas pocas bases y cientos de ellas) al que se unen, dependiendo del tamaño, unos pocos o un gran número de factores de transcripción. En los casos en que la unión de dichos factores a la región reguladora favorece la transcripción de un gen, se le denomina **potenciador** o *enhancer*. Si por el contrario la impide, a la región reguladora se la denomina **silenciador** o *silencer*.

Los genes contienen regiones de ADN llamadas **promotores**, que controlan el inicio de la transcripción, ya que son los lugares donde se une la ARN polimerasa para iniciar la transcripción. Las regiones reguladoras modulan la expresión génica interactuando con los promotores de los genes. Así, los *enhancers* o los *silencers* potencian o previenen el reclutamiento de la ARN polimerasa al promotor de un gen, favoreciendo o bloqueando la transcripción génica. A las regiones reguladoras se las denomina elementos en **cis** (o cis-reguladores), ya que están en la misma cadena de ADN que los genes. A los factores de transcripción que se unen a los elementos cis-reguladores se les denomina reguladores en **trans**. Así, la configuración-cis de un gen consiste en el conjunto de regiones reguladoras que operan sobre él.

Un aspecto característico de las regiones reguladoras es su modularidad. En general, cada región reguladora controla la expresión de un gen en un reducido número de tejidos u órganos. Por tanto, genes que se expresan en muchos tejidos y órganos, y en distintos momentos del desarrollo embrionario, contienen múltiples elementos cis-reguladores. Cada uno de ellos es un módulo independiente para controlar la expresión en unos pocos territorios. De esta forma, los genes con patrones de expresión sencillos (un tejido muy definido o simplemente en todas las células del organismo) suelen tener muy pocos elementos reguladores y localizados en regiones próximas al promotor. Por el contrario, los genes con patrones de expresión complejos contienen muchos elementos cis-reguladores dispersos a lo largo del ADN no-codificante tanto en la vecindad del gen como a distancias de hasta cientos de kilobases. A estos elementos cis-reguladores localizados en regiones genómicas lejanas del promotor se les denomina elementos distales. En muchos casos, este tipo de genes (a menudo implicados en el desarrollo temprano

del embrión), están flanqueados por grandes regiones de ADN no-codificante en las cuales no reside ningún otro gen. Estas regiones, también conocidas como **desiertos génicos**, están plagadas de regiones cis-reguladoras esenciales para controlar la expresión del gen en múltiples tejidos y en distintos momentos.

Si las regiones reguladoras se localizan en zonas intergénicas ¿cómo se organiza el genoma para que unas regiones actúen sobre un determinado gen y no sobre el vecino?

Para ello existen las regiones reguladoras denominadas **aisladores**. Estas tienen dos funciones:

- prevenir que un *enhancer* actúe sobre un promotor cuando el **aislador** se localiza entre estos dos elementos.
- impedir la expansión de la heterocromatina sobre un locus determinado, y con ello el silenciamiento de su expresión.

Los aisladores son, por tanto, elementos esenciales del genoma para generar compartimentos con entornos reguladores diferentes para genes vecinos. De esta forma, dos genes contiguos en el genoma, si están separados por aisladores, pueden estar bajo la influencia de un conjunto de elementos cis-reguladores totalmente diferentes, o como se dice en la jerga genómica, estos genes “ven” paisajes reguladores diferentes.

Un caso especial que conviene mencionar son los **complejos génicos**. Estos están constituidos por grupos de varios genes agrupados en una región del genoma que codifican proteínas estructuralmente relacionadas. Estos genes suelen tener una expresión génica muy coordinada en el espacio y en el tiempo, que deriva de la existencia de elementos cis-reguladores compartidos entre ellos. Estos elementos cis-reguladores compartidos controlan la expresión de varios genes del locus de una forma coordinada, y por ello se denominan *Locus Control Regions* (LCR). De hecho, es la existencia de estos LCR lo que mantiene a los genes formando complejos y lo que evita su dispersión cromosómica a lo largo de la evolución.

CUADRO 1

DEFINICIONES DE LOS DIFERENTES CONCEPTOS

ADN regulador (o regiones reguladoras): fragmento de ADN no codificante de tamaño variable (entre unas pocas bases y cientos de ellas) al que se unen, dependiendo del tamaño, unos pocos o un gran número de factores de transcripción.

Cis-regulador: región reguladora que se encuentra en la misma cadena de ADN que los genes a los que regula.

Regulador en trans: factor de transcripción que se une a los elementos cis-reguladores.

Enhancer o potenciador: región reguladora que favorece la transcripción de un gen.

Silencer o silenciador: región reguladora que impide la transcripción de un gen.

Modularidad: las regiones reguladoras controlan la expresión de los genes en distintos tejidos y momentos, mediante elementos cis-reguladores. Cuando los genes se expresan en muchos tejidos y órganos, y en distintos momentos del desarrollo embrionario, contienen múltiples elementos cis-reguladores. Cada uno de ellos constituye un módulo independiente para controlar la expresión en unos pocos territorios.

Regiones distales: elementos cis-reguladores localizados en regiones genómicas alejadas del promotor del gen.

Aislador: elemento del genoma que genera compartimentos con entornos reguladores diferentes para genes vecinos.

Locus Control Regions (LCR): elementos cis-reguladores compartidos que controlan la expresión de forma muy coordinada, de varios genes agrupados en una región del genoma que codifican proteínas estructuralmente relacionadas.

Conclusión

En resumen, nuestro conocimiento del genoma humano es hoy en día muy reducido, y en la próxima década gran parte del esfuerzo se dedicará a descifrar el significado de este 98% del ADN no-codificante. Esto nos permitirá poder identificar y

comprender las funciones de las distintas regiones reguladoras asociadas a los distintos genes y con ello estudiar el efecto que tienen la infinidad de mutaciones existentes en el ADN no codificante, tanto en las enfermedades humanas como en los genes asociados a éstas.

