

Mitos y verdades del Genoma Humano

Paula Cramer y Norberto D. Iusem

Departamento Fisiología, Biología Molecular y Celular-FCEN-UBA e Instituto de Fisiología, Biología Molecular y Neurociencias (IFIByNE)-UBA-CONICET

Comenzaremos esta mini-revisión sobre el Genoma Humano con algunas respuestas a preguntas básicas para adentrarnos luego con más soltura en algunos conceptos clave.

¿Qué es el genoma, en general?

El genoma es el patrimonio genético que en cada especie biológica se hereda de generación en generación y que está almacenada en su ADN (con excepción de los virus cuyo ácido nucleico es ARN). En organismos eucariotas (células con núcleo), está íntimamente asociado a las histonas, formando la cromatina.

¿Y qué es la genómica?

La genómica es la disciplina que estudia los genomas de las distintas especies, intentando establecer la relación entre la estructura y la función de los elementos de ADN que los componen.

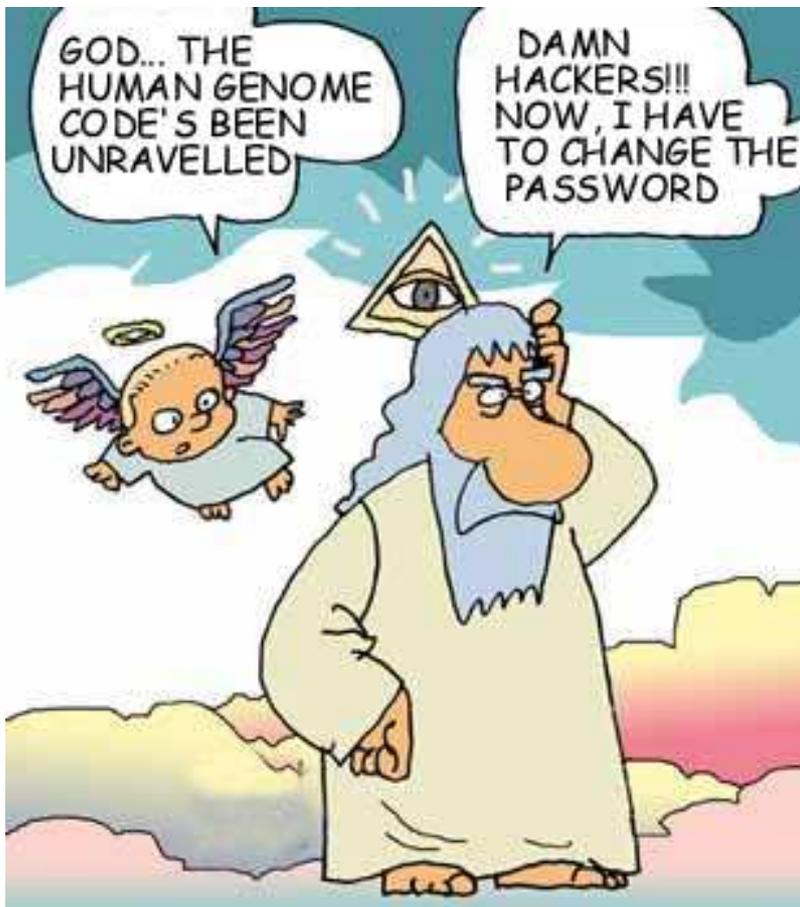
¿Y para qué sirve la genómica?

Advirtamos primero que el conocer el orden de cada uno de los nucleótidos que componen una especie (lo que se llama “secuencia”) no garantiza entender ni predecir sus funciones, es sólo la “punta del iceberg”. Lo que sí nos permite es comparar distintas especies relacionadas y extraer conclusiones sobre su evolución; describir las variaciones genéticas dentro de los individuos de una misma especie (*single nucleotide polymorphisms*, abreviado *SNPs*), ya sean inocuas o perjudiciales porque causan enfermedades; predecir la existencia de ciertos genes además de confirmar lo que ya sabemos sobre los genes ya conocidos, y por qué no, tener delante de nuestros ojos la evidencia de todo lo que aún no sabemos.

Debido a las limitaciones técnicas y de costos para obtener secuencias de ADN, los primeros genomas *secuenciados* eran relativamente pequeños, es decir, de menos de 20.000 nucleótidos de longitud. Así, en los años 80 se secuenciaron los genomas de organelas tales como mitocondrias y cloroplastos, o de ciertos virus. En los 90's se logró secuenciar el genoma completo de algunos procariontes y de un eucariota genómicamente sencillo: la levadura de la cerveza. Sobre el filo del siglo XX, se logró secuenciar el genoma completo de *Drosophila melanogaster*, la mosca de fruta, modelo en Genética desde el siglo XIX, de un centenar de millones de nucleótidos de longitud. Y ya en el siglo XXI, se anunció el primer borrador de la

secuencia del Genoma Humano (Lander et al. 2001), para lo cual hicieron falta 13 años, cerca de 3.000 millones de dólares y un consorcio internacional de laboratorios (conformado por EEUU, Reino Unido, Japón, China, Francia y Alemania). La interpretación de la secuencia obtenida, incluyendo la predicción de genes, fue posible gracias a la “bioinformática”, que utiliza poderosos programas de computación diseñados para desentrañar la información genética -y la accesoria- contenida en el genoma.

Afortunadamente, los formidables desarrollos tecnológicos han acelerado enormemente (actualmente en cuestión de días) la secuenciación de ADN, de manera que hoy contamos con las secuencias genómicas de más de 4.000 especies biológicas abarcando todo el espectro evolutivo, y la lista sigue creciendo.



¿Qué sabemos hoy del genoma humano entonces?

A grandes rasgos, nuestro genoma está compuesto por genes (o elementos funcionales, que se transcriben a ARN) y el mal llamado "*junk DNA*" o "ADN chatarra" (al que nos referiremos más adelante).

Entre los componentes funcionales, hay unos 22.000 **genes con información para sintetizar proteínas**, los cuales incluyen las secuencias realmente *codificantes* (exones) que finalmente

se *traducen* en proteínas, así como todas aquellas regiones del gen que no son codificantes pero que son necesarias para su correcta *expresión*, como los promotores, los estimuladores o *enhancers* y los intrones.

También están los genes para fabricar ARNs sin potencial codificante de proteínas, por ejemplo los “clásicos” **ARN ribosomales** y **ARN de transferencia**, necesarios para la traducción de ARN mensajero (mARN) en proteínas; y genes para producir **ARN pequeños** como los que participan del *splicing* de mARNs (los snARNs) o en la modificación química de otros ARNs nucleares (los snoARNs). En nuestro genoma también hay un número aún no exactamente determinado de genes que originan ARNs, que tampoco son codificantes (ncARN), y que intervienen en una apropiada y fina regulación de la expresión génica: los muy pequeños o **microARNs**, de los que se cree que hay unos 2.000 en humanos (<http://www.mirbase.org/blog/category/releases/>) y los muy largos, (**long ncARNs**), de hasta decenas de kilobases (miles de pares de nucleótidos), entre los que se cuenta el responsable de la inactivación de uno de los dos cromosomas X en las células de individuos de sexo femenino.

Además, es llamativo que cerca de la mitad de nuestro genoma consiste en elementos de ADN repetidos, que pueden encontrarse en tándem (vecinos unos con otros) o dispersos. Entre ellos, están los **ADN mini y microsatélites**, que son secuencias no transcritas muy cortas (de 5 a algunas decenas de nucleótidos) repetidas, y los intrigantes **transposones**, que son elementos muy abundantes, muy a menudo con capacidad de codificar proteínas (precisamente las responsables del “salto”), **potencialmente transcritos, móviles**, de distintos orígenes evolutivos y características estructurales. Estos últimos se clasifican en dos grandes grupos: 1) los transposones “a ADN” y 2) los retrotransposones -LINEs, SINEs y tipo retrovirus-. Estos últimos requieren transcribirse a ARN y luego esos ARNs se *retrotranscriben* a cADNs que se insertan en regiones al azar en el genoma, pudiendo la variabilidad genómica resultante tener consecuencias globales beneficiosas o adversas. Pero no entremos en pánico por los efectos negativos: si bien los transposones son muy abundantes, la frecuencia de su “salto” es muy baja, debido a mutaciones o a causas epigenéticas (modificaciones químicas sobre el ADN o la cromatina, heredables), que escapan al alcance de este artículo.

Finalmente, también están las secuencias correspondientes a los **centrómeros** (que tienen un rol estructural en la división celular) y los **telómeros** (extremos de los cromosomas, que requieren una maquinaria de replicación especial), también formados por unidades repetidas. En este contexto, se sabe que las secuencias repetidas, cualesquiera que sean, pueden causar

recombinación desigual –anómala- entre cromosomas homólogos cuando las células germinales entran en meiosis, alterando el genoma de la descendencia.

En la Tabla 1 se muestran las proporciones estimadas del genoma que ocupan todos los elementos mencionados, independientemente del grado de impacto que tienen en el funcionamiento de una célula.

Tipo de elemento de ADN	%
genes para proteínas	33
transposones o elementos móviles	45
genes de ncARN en total	1,3
Centrómeros	0,75
Telómeros	0,01
repeticiones no transcritas	3
el resto (se supone regiones intergénicas)	17

Tabla 1. Abundancia relativa de distintos tipos de elementos de ADN en el Genoma Humano. Las cinco primeras filas corresponden a elementos de función conocida. Las últimas dos filas, a secuencias de función aún desconocida.

Claramente, los elementos móviles ocupan la mayor parte de nuestro genoma (para un detalle de su composición ver la Tabla 2), mientras que los genes para proteínas ocupan un porcentaje importante (un tercio, aproximadamente) solamente si incluimos sus regiones regulatorias, mientras que si las excluimos y tenemos en cuenta solamente la información para codificar proteínas, ocupan apenas un 1 % del genoma.

Elemento móvil	% del genoma
LINEs	21
SINEs	13
Tipo retrovirus	8
Transposones “a ADN”	3

Tabla 2. Detalle de la proporción del genoma ocupada por los elementos móviles (LINEs: Long Interspersed Nuclear Elements, SINEs: Short Interspersed Nuclear Elements).

Como se puede ver en la Tabla 1, la mayor parte de nuestro genoma no codifica para funciones específicas, al menos conocidas al momento. En este contexto, el respetado genetista Susumu Ohno acuñó en 1972 el término "**junk ADN**" para referirse a ellas, que fue

infelizmente traducido como "ADN basura". Sin embargo, el no poder atribuirles una función no significa que no tengan rol o impacto alguno. El genial Sydney Brenner (premio Nobel de Medicina en 2002) llamó la atención a distinguir entre "chatarra" y "basura": mientras que la basura es una molestia de la que hay que deshacerse, los trastos o chatarra son potencialmente útiles en un futuro o en nuevo contexto (un estrés, por ejemplo). Por eso proponemos referirnos en castellano al "ADN chatarra".

Detengámonos ahora un poco en discutir las implicancias de llevar a cuestras en nuestro genoma esta carga de "ADN chatarra". Lo primero que hay que destacar es que la mayoría de los elementos móviles mencionados se encuentran inactivos al presente (sólo un 0,05% estaría activo, Mills et al 2007), ya sea porque han mutado y hoy constituyen fósiles moleculares, o como consecuencia de marcas epigenéticas silenciadoras que les han impuesto los genomas que los hospedan. ¿Pero qué sucede sin embargo con aquellos que siguen activos? La conclusión más inmediata es imaginarse que estos elementos son parásitos de nuestro genoma y que lamentablemente nos pueden causar enfermedades. Efectivamente, se han encontrado entre las causas moleculares de algunas patologías humanas la presencia de elementos móviles. Podemos citar como ejemplo el caso de la inserción de un elemento **Alu** (el tipo de retrotransposón SINE más abundante, que abarca tanto como un 5% del genoma) dentro de algunas de las variantes génicas de los genes BRCA1 y BRCA2 -las cuales están asociadas a ciertos cánceres de mama y ovario- y que resultan en la expresión de una proteína aberrante (Teugels et al 2005; para una revisión de otras enfermedades causadas por transposones, como retinitis, consultar Chen 2005 y Babatz & Burns 2013).

Sin embargo, la otra conclusión que podemos extraer es que, si como producto de la movilización de transposones surgen nuevas variantes génicas, éstos contribuyen a la plasticidad de nuestro genoma. Este es el caso del gen de la enzima alfa-amilasa salival humana. En la historia evolutiva de los mamíferos, este gen que se expresa en el páncreas se duplicó varias veces, pero hace unos 20 millones de años, un retrotransposón se movilizó e insertó en la vecindad de una de las copias del gen, lo cual le otorgó la capacidad de expresarse en las glándulas salivales (Samuelson et al. 1996). Esto ocurrió exclusivamente en la rama de los primates que dio lugar a los homínidos (chimpancés, gorilas y orangutanes) e implicó un evidente valor adaptativo: la capacidad de degustar el agradable sabor dulce de la glucosa y sacarosa, lo que probablemente ayudó a reconocer y acaparar esta fuente de alimento nutritivo. Este es solamente un ejemplo bien estudiado de la influencia benéfica de los elementos móviles en nuestro genoma, pero hay otros casos que también ilustran que los transposones han aportado regiones promotoras, *enhancers* o exones, generando novedosos

alelos en la evolución (Oliver & Keene 2011), aunque algunos autores contemporáneos advierten la necesidad de verificar estas hipótesis experimentalmente, entre ellos científicos argentinos (de Souza et al 2013).

¿Qué interrogantes siguen en pie?

Como la secuencia del genoma humano obtenida originalmente provino del ADN de un puñado de individuos, una cuestión importante era saber cuán diferentes somos unos humanos de los otros, y si hay regiones del genoma o elementos particulares en los que se encuentran las mayores diferencias. Para responder esto recientemente se llevó a cabo el proyecto de los 1.000 genomas (<http://www.1000genomes.org/>), que como su nombre indica consistió en secuenciar los genomas completos de ese número de personas de diversas procedencias geográficas. Por otro lado, el proyecto ENCODE (ENCyclopedia Of DNA Elements) es un consorcio internacional que pretende generar y luego compilar toda la información funcional de que se disponga sobre el genoma. Este proyecto viene a responder algunos de los interrogantes planteados aquí, por ejemplo, qué elementos serían funcionales y cuáles no, qué funciones cumplirían (si son regiones regulatorias, si codifican proteínas o ARNs no codificantes) y cuál es la influencia de las modificaciones epigenéticas sobre la expresión génica. Este emprendimiento ha generado un gran debate en la comunidad científica internacional. Sus seguidores, que están convencidos de que la mayor parte del genoma tiene alguna función específica, esperan que este proyecto brinde una pintura completa y acabada de nuestro paisaje genómico y de su funcionalidad y, usando analogías cartográficas, predicen que será como un gran GPS de nuestro genoma. Por el contrario, sus detractores, más rigurosos a la hora de definir “función” de elementos genómicos, argumentan que probablemente la mayor parte de nuestro genoma no esté compuesta por elementos funcionales y advierten que el proyecto es demasiado ambicioso y que difícilmente cumpla con sus promesas. Queda por verse si los resultados darán la razón a unos o a otros, o si el tiempo hará decantar alguna postura intermedia.

Conclusiones

Recapitulando, nuestro genoma está formado por 1) componentes de función bioquímica conocida que ocupan un porcentaje minoritario de nuestra información genética y 2) una proporción muy elevada de elementos eventualmente móviles. La mayoría de éstos últimos se han convertido en fósiles moleculares, pero seguramente han tenido un rol en la plasticidad de nuestro genoma en el transcurrir de nuestra historia evolutiva, aunque pueden en el presente y

en pequeña escala seguir teniendo efectos negativos en nuestra salud. La fracción del genoma interpretada como ocupada por “regiones intergénicas”, verdadero “DNA chatarra” si es que queremos seguir usando esta analogía, sigue disminuyendo a medida que se descubren cada vez más genes de ARN no codificante (localizados entre genes previamente conocidos) que ayudan a la célula a apagar genes circunstancialmente innecesarios.

Todavía tenemos mucho que aprender sobre qué significan los tres mil millones de nucleótidos del Genoma Humano, sus modificaciones químicas adquiridas en cada tipo celular y tejido de acuerdo a estímulos externos (terreno de la epigenética), y la influencia en la expresión génica que ejerce la dinámica cromatina. Futuros avances en técnicas de Biología Molecular y en Bioinformática seguramente contribuirán al desafío, ayudando a profundizar en el conocimiento que tenemos sobre la evolución humana y a entender y prevenir enfermedades.

Referencias

Babatz TD, Burns KH. Functional impact of the human mobilome. *Curr Opin Genet Dev*. 2013 Mar 22. doi:pii: S0959-437X(13)00025-7. 10.1016/j.gde.2013.02.007. [Epub ahead of print]

Chen JM, Chuzhanova N, Stenson PD, et al. Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. *Hum Mutat*. 2005; 25(3): 318.

de Souza FS, Franchini LF, Rubinstein M. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol*. 2013; 30(6): 1239-51.

The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 2004; 306 (5696): 636-640.

Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409(6822): 860-921.

Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? *Trends Genet*. 2007; 23(4): 183-91.

Ohno S. So much junk DNA in our genome. *Brookhaven Symp Biol*. 1972; 23: 366-70.

Oliver KR, Greene WK. Transposable elements and viruses as factors in adaptation and evolution: an expansion and strengthening of the TE-Thrust hypothesis. *Mob DNA*. 2011; 2(1):8

Samuelson LC, Phillips RS, Swanberg LJ. Amylase gene structures in primates: retroposon insertions and promoter evolution.

Mol Biol Evol. 1996;13(6): 767-79.

Teugels E, De Brakeleer S, Goelen G, Lissens W, Sermijn E, De Grève J. De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. *Hum Mutat.* 2005; 26(3): 284.