

Qué es el genoma humano

Las proteínas protegen, empaquetan y regulan la actividad biológica del ADN. Los humanos tenemos 46 cromosomas en cada célula, organizados en 23 pares. Recibimos 23 cromosomas de nuestra madre y 23 de nuestro padre.

Juan R. Riesgo

El 26 de junio de 2000 se anunció la secuenciación casi completa (al 99%) y el ensamblaje del genoma humano por parte de la compañía de biotecnología Celera. Al mismo tiempo, el Proyecto Genoma Humano (*Human Genome Project*, HGP), un consorcio internacional financiado públicamente y dirigido por Francis Collins anunció la terminación del *scaffold* o armazón del genoma al 85% y su secuenciación parcial. Este hecho corona los esfuerzos de más de diez años del consorcio y de aproximadamente dos años de Celera, compañía presidida por J. Craig Venter. En febrero de 2001 las revistas *Nature* y *Science* publicaron los resultados de estos grupos, así como el primer análisis del genoma humano.

Las noticias, que si bien por esperadas no fueron menos importantes, me hicieron pensar inmediatamente en dos incidentes: la famosa frase de Neil Armstrong: “éste es un pequeño paso para un ser humano, pero un paso gigantesco para la humanidad” y la opinión muchas veces vertida por el científico es-

tadunidense James Watson, ganador del Premio Nobel por dilucidar, junto con Francis Crick, la estructura del ácido desoxirribonucleico o ADN, en el sentido de que se requerían urgentemente formas rápidas de secuenciación del ADN para poder leer todo el genoma humano, al menos antes de que él muriese. Armstrong se

refería a la conquista de la Luna por el hombre, pero el hecho de conocer el libro de toda nuestra información genética —que ahora tenemos que interpretar y entender— es tan o más monumental que aquel hecho, si bien no se anunció al mundo de manera tan espectacular. Y sí, el deseo de James Watson se cumplió, pues estuvo presente durante el anuncio del 26 de junio.

A través de la historia, el hombre siempre ha tratado de estudiarse y entenderse a sí mismo cada vez mejor y con mayor detalle, así como a su medio ambiente. En este sentido, la secuenciación del genoma humano es un paso más en tal dirección, comparable al trabajo de los anatomistas que disecaban cadáveres humanos para describir, estudiar y entender el funcionamiento de nuestro organismo. De la misma manera como se describieron el corazón, los pulmones, el sistema circulatorio, etc. —a la mente vienen los incomparables estudios y dibujos anatómicos de Leonardo da Vinci y de Miguel Ángel Buonarroti—, ahora

se ha terminado de leer el libro de todas nuestras instrucciones genéticas. Es en este afán humanista y en la naturaleza inquisitiva que tanto caracterizan a nuestra especie en el que se inscribe este nuevo paso.

¿Qué es el genoma humano? Es una especie de vademécum, de enciclopedia de toda la información genética que poseemos como especie. Tal vez podamos hacer un símil con la construcción y el funcionamiento de una casa que seríamos nosotros. Para poder construir pisos y techos, levantar paredes, hacer instalaciones eléctricas y sanitarias se requiere información sobre los materiales, y sobre su disponibilidad, y en algunos casos sobre la manera de elaborarlos a partir de materias primas; igualmente necesarios son la forma de hacer la mezcla de cemento para pegar los ladrillos, el modo de hacer aplanados con yeso, conocer de pinturas y el modo de aplicarlas; además, hay que tener conocimientos de física, de química, de hidráulica, de diseño, etc. Toda esta información existe en libros, y continuando nuestra metáfora, debemos usarla en el entorno natural en donde se está construyendo la casa: utilizar los materiales locales, ajustarse al terreno en donde se piensa edificar y, siguiendo ciertas normas, establecer una especie de diálogo con el medio ambiente de donde nos proveemos de materiales y en donde estamos construyendo, para poder llevar a buen término el proyecto. En esta metáfora, el genoma ocuparía el lugar de los libros y textos que nos indicarían cómo llevar a cabo la construcción.

Estos textos están escritos en un idioma: el idioma de los genes, y tienen su sustrato físico en una sustancia química: el ADN. El sustrato físico equivaldría al papel y a la tinta de los libros y de los textos. El ADN es una molécula muy larga y delgada formada por dos hebras o ejes entrelazados que forman una doble hélice. Estas hebras están formadas por subunidades repetidas de manera lineal, llamadas nucleótidos, constituidos a su vez por tres partes: una base nitrogenada, un azúcar de cinco carbonos, llamado desoxirribosa, y un grupo fosfato. Las bases nitrogenadas pueden ser de dos clases, purinas y pirimidinas, de acuerdo con su estructura química, y existen dos tipos dentro de cada clase: adenina y guanina, para las purinas, y citosina y timina, para las pirimidinas (Figura 1). Así, cada nucleótido está conformado por una de las

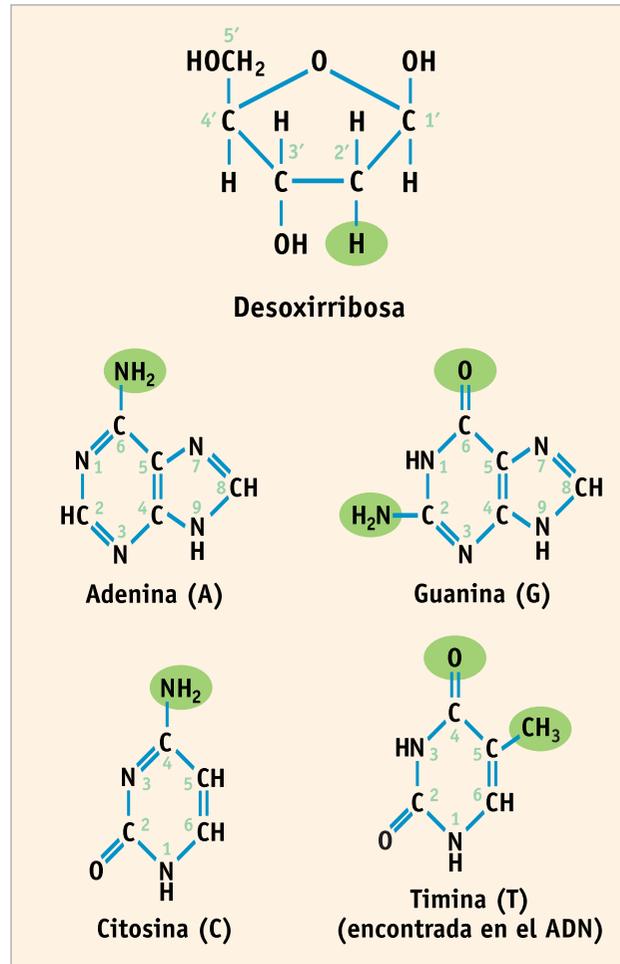


Figura 1. Moléculas componentes del ADN. La molécula del ADN está compuesta por varios tipos de moléculas que se unen entre sí por medio de distintos tipos de enlaces químicos. Las moléculas componentes incluyen bases púricas (adenina y guanina), bases pirimídicas (citosina y timina), el azúcar desoxirribosa, de la que toma su nombre el ADN, y el ácido fosfórico.

cuatro posibles bases nitrogenadas (abreviadas A, G, C y T, respectivamente), junto con una molécula de desoxirribosa y un grupo fosfato invariantes (Figura 2). Dado que sólo son las bases nitrogenadas las que pueden variar, en el lenguaje del ADN sólo existen cuatro letras: A, G, C y T.

Los nucleótidos se entrelazan entre sí por medio de los grupos fosfato y las desoxirribosas, de modo tal

que forman una especie de hilo que tiene en un mismo lado todas las bases nitrogenadas, como en un collar de abalorios. Las bases nitrogenadas están por un lado unidas al armazón de grupos fosfato y desoxirri-

bosas, y por el otro están libres para interactuar. Dos hebras constituidas así y enlazadas entre sí por uniones químicas débiles, llamadas puentes de hidrógeno, que unen los extremos libres de las bases nitrogenadas, constituyen una molécula de ADN (Figura 3). La unión de las bases por medio de puentes de hidrógeno se realiza de la siguiente manera: siempre se une una purina con una pirimidina, de modo tal que la adenina se une siempre con la timina y la guanina con la citosina (A con T y G con C). No se pueden unir de otro modo porque la adenina y la timina forman entre sí dos puentes de hidrógeno, mientras que la citosina y la guanina forman tres (Figura 3).

Es esta característica de las bases nitrogenadas de los nucleótidos, de unirse formando parejas invariantes, lo que le da a la molécula de ADN una propiedad muy

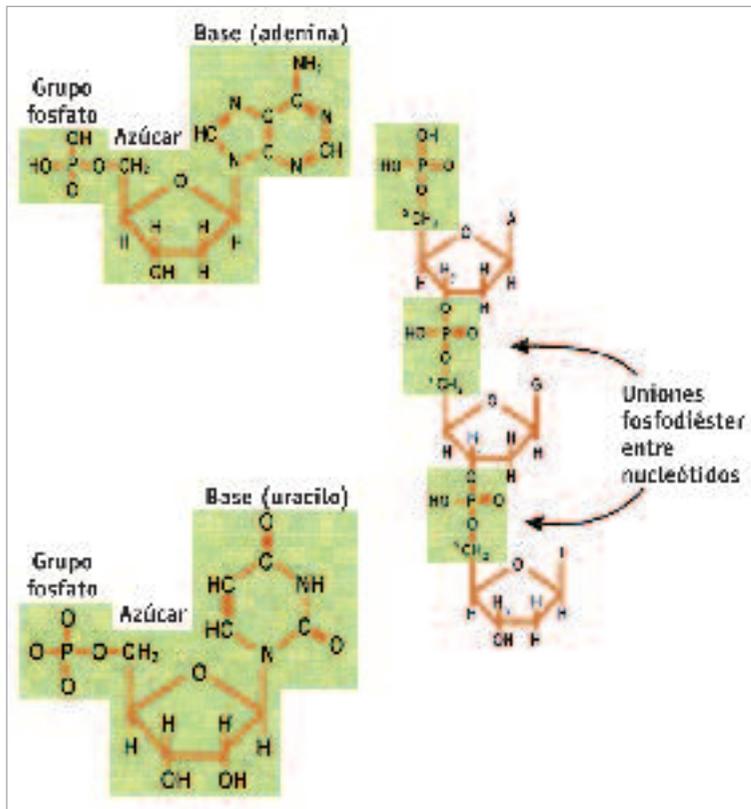


Figura 2. Estructura de nucleótidos y polinucleótidos. Una base purínica o pirimidínica se une a un azúcar y a un fosfato para formar un nucleótido. Aquí se muestra la estructura de un nucleótido de purina que contiene adenina y uno de pirimidina que contiene uracilo. Los nucleótidos en la molécula del ADN se unen entre sí a través de sus fosfatos y sus azúcares (desoxirribosas) mediante un tipo especial de enlace químico denominado enlace fosfodiéster.

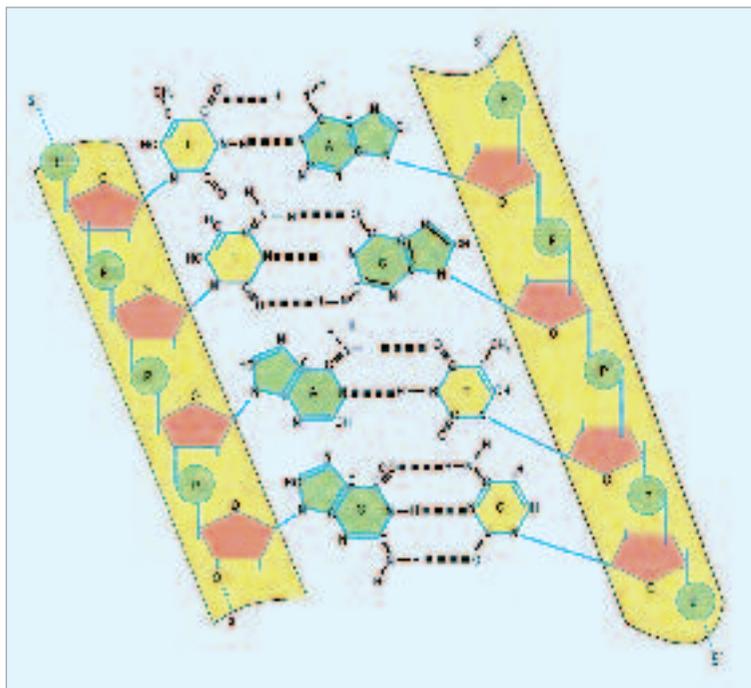


Figura 3. Interacciones entre nucleótidos en el ADN. En cada una de las dos hebras del ADN los nucleótidos se unen entre sí por medio de uniones fosfodiéster que forman el esqueleto de la hebra y del cual protruyen las bases purínicas y pirimidínicas hacia el centro de la molécula. Nótese que las dos hebras del ADN interactúan entre sí de manera tal que la base purínica de una hebra interactúa con la base pirimidínica de la otra, estableciéndose entre ellas puentes de hidrógeno que unen átomos de oxígeno con átomos de nitrógeno, y que las cadenas de nucleótidos tienen polaridades opuestas.

importante: cada hebra de las dos que la conforman puede servir de molde para la elaboración de la otra (Figura 4). De esta manera las hebras que conforman una molécula de ADN son complementarias. Es decir, al conocer el orden en que los nucleótidos se han enlazado en una hebra (por ejemplo, GAATTC) podemos conocer exactamente cuál será el ordenamiento de los nucleótidos de la otra hebra (en nuestro ejemplo, CTTAAG). Esta propiedad de la estructura de la molécula de ADN sirve muy bien para duplicar la información genética, pues al separarse por medio de enzimas las dos hebras que la constituyen, cada una sirve de molde para la elaboración de otra hebra, de donde al final se obtienen dos moléculas de ADN idénticas, cada una con una hebra que sirvió de molde y con otra que se formó a partir de la primera. Esto es precisamente lo que ocurre con todo el ADN de las células cada vez que se dividen, y así se explica que cada una de las dos células hijas tenga la misma información genética que la célula original.

El ADN de nuestras células se encuentra dentro de una estructura llamada núcleo. Las dos hebras de la molécula de ADN se van retorciendo sobre sí mismas para formar una doble hélice (Figura 4). Ésta es la estructura característica del ADN que describieron James Watson y Francis Crick. Dentro del núcleo, sin embargo, estas dobles hélices de ADN están rodeadas de proteínas. Juntos, el ADN y las proteínas asociadas, conforman una estructura que se llama cromosoma (hay dos excepciones a lo anterior: existe también ADN no nuclear con algunos organelos celulares como las mitocondrias y los cloroplastos, y existen también moléculas de ADN de pequeño tamaño que normalmente son circulares y que pueden estar en el citoplasma o en el núcleo de las células, pero que no forman parte de los cromosomas; estas moléculas de ADN se denominan episomas o plásmidos).

Las proteínas protegen, empaquetan y regulan la actividad biológica del ADN. Los humanos tenemos 46 cromosomas en cada una de nuestras células, organizados en 23 pares. Recibimos 23 cromosomas de nuestra madre y 23 de nuestro padre. Durante la formación de las células especializadas de la reproducción, los gametos, es decir, los cromosomas de esas células, sufren un proceso llamado meiosis, que tiene por resultado la reducción de la cantidad de cromosomas. Ya maduros, los gametos sólo contienen la mitad de la cantidad normal de cromosomas; es decir, para nuestra especie, 23 cromosomas. Al unirse los dos gametos, el femenino y el masculino, se reconstituye el número normal de cromosomas, 23 pares o 46 cromosomas. Por tanto, cada célula de nuestro cuerpo tiene dos copias de cada cromosoma, o dos juegos completos de todos nuestros genes, o genoma: uno que heredamos de nuestro padre y otro que heredamos de nuestra madre.

El orden en el que se han enlazado los nucleótidos que conforman las hebras del ADN constituye la información genética. Es decir, la

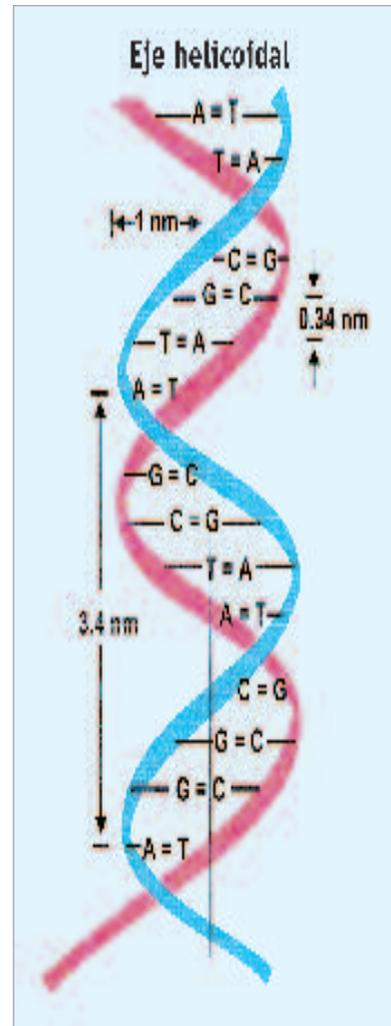
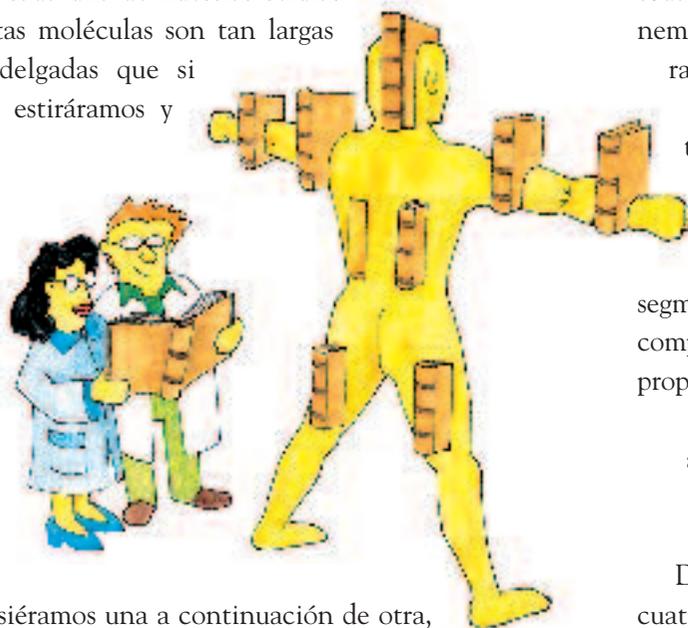


Figura 4. Estructura del ADN. Obsérvese que el ADN está constituido por dos hebras de nucleótidos que forman una doble hélice. En cada una de estas hebras los nucleótidos, definidos por la base que poseen, siguen una secuencia determinada en la que se codifican las características fenotípicas de la especie en cuestión.

secuencia que se obtiene de ir “leyendo” cada nucleótido de acuerdo con su lugar en la hebra —de igual modo que al leer nosotros una palabra vemos en qué lugar está cada una de las letras que la conforman— nos va a decir qué tipo de mensaje está codificado. Por supuesto, la información que contenga una secuencia como GAATTC no será la misma que la que contenga otra secuencia como TATAGC, aunque ambas sean del mismo tamaño y utilicen las mismas letras, de igual manera que CARO no es lo mismo que ROCA. Todo el genoma humano (el conjunto de todos los genes que poseemos como especie) está codificado en las moléculas de ADN que poseemos en cada una de nuestras células.

Estas moléculas son tan largas y delgadas que si las estiráramos y



pusiéramos una a continuación de otra, todas las moléculas de ADN de una de nuestras células medirían cerca de dos metros de longitud. Tomando en cuenta que un núcleo promedio sólo mide de dos a tres micras de longitud (una micra es la milésima parte de un milímetro), es claro que la gran cantidad de ADN que poseemos se encuentra muy bien empaquetada en los núcleos celulares. Ahora sabemos que el genoma humano tiene aproximadamente 3 mil millones de nucleótidos, que llenarían aproximadamente 500 directorios telefónicos de 500 páginas cada uno, o un solo DVD (video disco digital), repartidos en 23 cromosomas. Estos 23 cromosomas representan un genoma (nosotros tenemos dos copias en cada una de nuestras células, excepto en los gametos), y un genoma mediría estirado aproximada-

mente un metro. Esto es justamente lo que están terminando de elaborar el Proyecto Genoma Humano y la compañía Celera: conocer y ensamblar toda esta secuencia correctamente. Es como tener una oración gigantesca de 3 mil millones de letras. Embebidas dentro de esta secuencia se encuentran palabras que codifican para todos nuestros genes. A su vez, cada gen que poseemos puede codificar finalmente para una proteína o para otro tipo de ácido nucleico, llamado ácido ribonucleico, o ARN, que es de cadena sencilla. Estos resultados se publicaron en el otoño del año 2000. Finalmente tenemos los libros, volviendo a nuestra metáfora de la construcción, que contienen todas las instrucciones para hacer la casa. Es decir, tenemos el libro que especifica a nuestra especie. Ahora es necesario entender e interpretar este libro.

Hagamos algunas aclaraciones acerca de la sintaxis del idioma del ADN: la secuencia es continua, como una gargantuesca oración que nos dejaría sin aliento muchísimo antes del final, y dentro de ella, embebidos, y frecuentemente segmentados, se encuentran los genes. Si hacemos una comparación con las letras del alfabeto, podríamos proponer que un segmento del genoma se vería así:

actdbacieghiiiiROgtdvbgfsabnammmnuloka
huiypacCAmkolkijhjaauaaahkn

Dentro de esta secuencia están en mayúsculas cuatro letras: RO y CA, que unidas harían la palabra ROCA. Pero ¿cómo saber que lo que codifica esa secuencia es ROCA y no otra cosa? (Las mayúsculas las he puesto para facilitar la comprensión.)

Existen signos de puntuación: por ejemplo, la corrida de cuatro íes (iiii) que están justo antes de RO significaría que las letras que siguen son importantes porque son parte de una palabra (gen), y que esta palabra inicia justo a continuación de la última *i*. Esta palabra continúa sin interrupción hasta encontrar *gt*, que indica que la palabra está segmentada (de nuevo, para simplificar las cosas, he dividido la palabra en sílabas), y que se debe seguir leyendo hasta encontrar *ac*, que vendría a significar que justo después de *ac* se halla el siguiente segmento de la palabra, CA. Las señales *gt* y *ac* en este contexto significan que debemos

unir RO con CA para formar la palabra, y que debemos desechar la secuencia *mk*, que significaría que ése es el final de la palabra. Ésta es parte de la información, pero también dentro de la secuencia hay otros mensajes: se especifica cuándo debe leerse ROCA, en dónde debe leerse y quiénes deben leer ROCA. Es decir, no todas las células del cuerpo expresan todos los genes que tenemos. Por ejemplo, una neurona no expresa el gen que codifica para la queratina, que es la proteína mayoritaria del pelo, así como una célula pilosa no expresa las proteínas que sirven para fabricar los mensajeros que utilizan las neuronas para comunicarse entre sí. Además, los genes no necesariamente se expresan todo el tiempo, sino que muchas veces sólo lo hacen durante cierta etapa del desarrollo: el feto humano en desarrollo dentro de la madre expresa una hemoglobina distinta de la que expresa cuando ya ha nacido. La hemoglobina fetal es más afín al oxígeno que la hemoglobina que se sintetiza después, y que también produce la madre, lo que posibilita que el oxígeno pase de la hemoglobina de la madre a la del feto. Todas estas señales de control también se encuentran embebidas en la secuencia, y pueden localizarse antes (por ejemplo: *baci*), en medio, es decir, entre las señales *gt* y *ac*, o después, es decir después de las letras *mk* de nuestro ejemplo.

Estas señales de control se conocen de manera parcial, de modo que aún no es posible tener el número exacto de nuestros genes porque aún no los hemos localizado a todos dentro de esta gigantesca oración. Por supuesto, lo primero que se necesita es tener la secuencia completa (oración) para poder después analizarla. Lo que estamos terminando de descubrir es precisamente la secuencia del genoma, la oración completa. Tenemos el diccionario y los instructivos, pero no conocemos todas las palabras. El siguiente paso es someter esa secuencia a diversos algoritmos computacionales capaces de reconocer las palabras embebidas en la secuencia, así como los códigos de control, algo que ciertamente representa otro esfuerzo monumental. De hecho, una rama de la biología y de la informática, la bioinformática, se dedica a estudiar esos aspectos de los genes: diseñar programas de cómputo que sean capaces de descubrir las palabras embebidas en la oración. La bioinformática también

tuvo un papel muy revelante en la secuenciación del genoma: para poder leerlo todo, se tuvo que partir al genoma en partes muy pequeñas, que se podían secuenciar, para luego almacenar la información, y finalmente, como en un rompecabezas, colocar todos los fragmentos de secuencia en el lugar adecuado. Se dice que ensamblar el genoma humano es el rompecabezas más complejo que ha resuelto la especie humana hasta hoy día, un reto que hace empequeñecer el famoso nudo gordiano o el laberinto de Minos.

El 15 de febrero de 2001 se dio a conocer públicamente los resultados de este esfuerzo de secuenciación. El genoma humano se compone aproximadamente de 3 mil 200 millones de pares de bases. Según estudios de bioinformática como los esbozados antes, se cree que tenemos entre 30 y 40 mil genes. Sin embargo, dada la experiencia en predecir genes a partir de la secuencia en el genoma de la mosca de la fruta, *Drosophila melanogaster*, es posible que el número total de genes aumente entre 10 y 20%, es decir, hasta cerca de 50 mil genes; pero esto es así porque los algoritmos que se utilizan hoy en día para predecir genes a partir de la secuencia no son perfectos, y dan una subestimación del número real.

Si hacemos una comparación con alguien que empieza a aprender otro idioma, diríamos que sabemos ya muchas palabras y que conocemos algunos de sus usos (cuáles son algunos de los códigos de control, cuándo se usan, bajo qué circunstancias y por quiénes), pero desconocemos muchas otras, de manera que el cuadro que tenemos ahora es parcial. Sólo tenemos un dominio parcial del idioma; somos un alumno adelantado en la escuela de idiomas, pero el aprendizaje aún no termina.

Juan R. Riesgo es licenciado en biología por la UNAM, maestro y doctor en esa ciencia por la Universidad de Yale, y posdoctorado en la Universidad de Zurich. Pertenece al Sistema Nacional de Investigadores. Desde 1998 se desempeña como investigador en el Centro de Neurobiología de la UNAM, en Querétaro, cuyo tema central es el estudio de la vía de señalización de la cinasa de Jun y la transducción olfatoria y visual de la mosca de la fruta, *Drosophila melanogaster*, en la que se emplea una mezcla de biología molecular, celular, genética, bioquímica, fisiología y biología conductual para identificar y caracterizar los diversos genes que componen las cadenas transductoras.