

2. EL PROYECTO GENOMA HUMANO

Las condiciones previas: ideas, mapas y un cerebro global

Nos ha costado un siglo y un capítulo entender lo que es el genoma. Ahora que sabemos que nuestro cuerpo tiene un mensaje para nosotros, hemos emprendido la tarea de descifrarlo. Nuestro genoma es una fuente de información de proporciones monumentales, un reto para los analistas y los descodificadores. No es el Grial, ni la solución a todos nuestros males, pero nos puede ayudar a comprender infinidad de aspectos de nosotros mismos y de la naturaleza en general. Para saber exactamente dónde está toda esta información hace falta orientarse; necesitamos un mapa. Este mapa está prácticamente acabado: lo ha dibujado el Proyecto Genoma Humano.

Desde los años 70 disponemos de otros mapas menos detallados que nos dan información aproximada sobre en qué cromosoma está un gen, o qué genes están cerca unos de otros. Parece poco, pero es una gran ayuda. Para según qué estudios, no es necesario tener todos los datos; vea, si no, lo útiles que resultan los mapas turísticos donde sólo se indican los lugares más típicos de una ciudad. Si necesita un conocimiento más detallado, siempre está a tiempo de comprar un mapa más completo.

En el genoma, por ahora, no podemos escoger. Durante años hemos viajado con la ayuda de mapas cada vez más detallados, completándolos sobre la marcha al más puro estilo de los viajeros de tiempos pasados. Estamos empezando a ver el final de la era cartográfica: el mapa más preciso posible está al alcance de la mano. Este mapa contiene todos los

detalles del universo que es el genoma. Igual que un terreno geográfico, nuestros genes cambian con el paso del tiempo. Un mapa es una foto de algo en movimiento: un mapa de su ciudad de hace cincuenta o cien años puede tener un valor histórico, pero le servirá de poco para orientarse. El genoma también es variable, pero a una escala tan grande en el tiempo que no nos hemos de preocupar por esto. La etapa siguiente no es menos emocionante: tendremos que interpretar los signos de este mapa. Qué ciudades están abandonadas, qué otras están en construcción, cuáles son los centros de comunicación sobre los que se asienta todo el resto, qué nos dice el paisaje entre un pueblo y otro.

Recorta, pega y colorea

Si me permite la analogía, podemos decir que los científicos implicados en el Proyecto Genoma Humano han pasado los últimos quince años ocupados en una especie de “recorta, pega y colorea”, en un magnífico ejemplo de aplicación de los conceptos de la educación primaria a la investigación pionera. Veamos cómo.

Hemos hablado antes del sexo de las bacterias, de cómo intercambian fragmentos de DNA entre ellas de manera que pueden transmitir algunas características no sólo a sus hijos sino también a sus vecinos. Esto es habitual, pero no necesariamente simple. En principio un DNA extraño no tiene por qué traer nada bueno y las bacterias no aceptan pasivamente cualquier DNA que se les introduzca. Como mecanismo de defensa han evolucionado unas proteínas que reconocen una secuencia de bases, por ejemplo: ...GAATTC... y cortan la molécula de DNA entre la G y la A. El genoma de esta bacteria concreta no contiene esta secuencia de DNA (o la oculta, pero no nos perdamos en los detalles). De lo contrario la bacteria destruiría su propio material genético en una especie de suicidio molecular. En cambio, si un extraño le introduce una molécula de DNA que contenga esta secuencia de bases,

la bacteria la cortará. Esta señal es un reclamo para unas proteínas que detectan DNA roto y lo acaban de destruir. El efecto es parecido al de la sangre para atraer a los tiburones. A lo mejor este fragmento de DNA podría ser muy útil para la bacteria receptora: quizá le ofrecía la posibilidad de resistir un antibiótico, o de extraer energía del petróleo pero, aunque en conjunto las bacterias nos parezcan promiscuas y flexibles, cada una de ellas individualmente es muy conservadora. Salvo por error o casualidad, las bacterias restringen la entrada de DNA extraño: no destruyen el DNA de sus parientes cercanos, porque es parecido al suyo y probablemente no contiene la secuencia señal, pero sí el de los desconocidos.

Las proteínas que restringen la entrada de DNA se descubrieron en los años 70 y fueron llamadas “enzimas de restricción”. “Enzima” es el nombre que se da a un tipo de proteína que lleva a cabo reacciones químicas: una enzima reconoce una sustancia y la transforma. En el ejemplo que hemos visto, la secuencia GAATTC es reconocida por una enzima de restricción producida por la bacteria *Escherichia coli*, un habitante de nuestros intestinos que nos causa problemas cuando se encuentra en otros lugares. La enzima de restricción purificada de *Escherichia coli* es llamada con sus iniciales: *Eco*, seguidas de un número de identificación: *EcoRI*. Sólo las bacterias producen enzimas de restricción: otros organismos han desarrollado diferentes sistemas de defensa ante la invasión de DNA extraños. Nosotros, por ejemplo, tenemos un sistema inmunitario que detecta la presencia en la sangre de virus, bacterias y otros invasores, pero que no reconoce el DNA que llevan dentro sino las proteínas que los recubren por fuera.

Hace casi treinta años que utilizamos estas enzimas de restricción para cortar el DNA de manera controlada: para un biólogo molecular, es la diferencia entre rasgar al azar y cortar por la línea de puntos. Podemos cortar exactamente el fragmento que nos interesa, pegarlo en una de las fotocopadoras a las que nos referíamos en el capítulo anterior y así

obtener todas las copias que necesitemos. ¿Con qué lo pegamos? Con un adhesivo molecular, claro.

Las bacterias producen enzimas que cortan el DNA, y algunos virus producen enzimas que reparan estos cortes: como el gato y el ratón a escala molecular. Los virus son mucho más pequeños que las bacterias y utilizan los recursos de sus huéspedes para vivir. Pero primero han de superar las defensas bacterianas. Uno de los mecanismos para evitar el daño a su material genético es repararlo rápidamente, antes de que los *tiburones* detecten la señal y lo destruyan. Para ello existen enzimas, llamadas ligasas, que reconocen los extremos de una molécula de DNA y los unen. Todos los seres vivos, nosotros incluidos, tienen algún mecanismo para reparar el DNA roto, pero las ligasas que se usan como herramienta en los laboratorios son las producidas por virus.

La combinación de estas dos herramientas, las tijeras (enzimas de restricción) y el adhesivo (ligasa), ha impulsado la evolución de la biotecnología. Este avance no estuvo libre de debate en su momento: la incertidumbre inicial relacionada con la introducción de DNA de organismos como la mosca en bacterias y virus llevó a que los científicos se autoimpusieran en 1974 una moratoria de entre dos y cuatro años para efectuar experimentos con organismos patógenos, especialmente los virus causantes de tumores, que pudieran implicar riesgos para la salud de los investigadores o el público. En una reunión en Asilomar (Estados Unidos), 140 biólogos moleculares (con la ayuda de algunos abogados) discutieron las condiciones en que se podrían reanudar los experimentos. Estó implicó el diseño de los laboratorios modernos de biología molecular, con sistemas de seguridad físicos y biológicos. Al mismo tiempo, señaló el inicio de la preocupación del público por la seguridad de los experimentos de microbiología y genética. También en este momento se planteó por primera vez la pregunta “Una bacteria que contiene un gen de mosca, ¿sigue siendo una bacteria?”, que hoy es una parte central en el debate sobre la introducción de genes humanos en otros mamíferos (para producir insulina humana

en la leche de vaca, por ejemplo) o la posibilidad de trasplantar órganos de mamíferos a seres humanos.

Nuestra capacidad de manipular las enzimas de restricción y las ligasas para adaptarlas a las necesidades del laboratorio ha representado para la biología un avance comparable a lo que la anestesia fue para la cirugía: la posibilidad de traspasar una barrera hasta entonces infranqueable. Gracias a ello podemos plantearnos experimentos imposibles de imaginar antes: la manipulación precisa y controlada del material genético.

Las dos primeras partes de esta analogía, referidas a cómo se corta y cómo se pega el DNA, estaban más o menos indicadas en el primer capítulo: son las herramientas que utilizamos para clonar genes. La tercera parte es menos evidente. ¿Cómo se colorea el DNA? ¿Para qué lo queremos de colores?

El DNA no tiene un color que lo identifique, como lo pueden tener las naranjas, el azufre o el curaçao. Disuelto en agua es transparente y cuando lo tenemos en grandes cantidades y seco es translúcido, tirando a marrón claro, como una mancha vieja de café. Recuerde que, tratándose de DNA, grandes cantidades quiere decir unas cuantas milésimas de gramo: por encima de esto ya estamos a escala industrial. Ni seco ni disuelto, su color no es importante para la investigación. Si queremos colorear el DNA es para conocer el orden de las bases que forman una molécula concreta, lo que llamamos su secuencia.

Supongamos que el ordenador de la Universidad de California en Santa Cruz, uno de los centros donde está almacenada la secuencia de nuestro genoma, guardara los datos en orden alfabético: una lista con los millones de cada una de las cuatro bases nitrogenadas (A, C, G y T). Este dato no nos diría nada sobre qué genes puede haber en el genoma: el número de combinaciones es tan grande que necesitamos saber en qué orden están las bases para que la información tenga sentido. Conocer el orden de las bases en una molécula de DNA es lo que llamamos *secuenciar* esta molécula. A principios de los años 70 se hizo evidente que para profundi-

zar en la nueva biología molecular basada en el DNA era necesario idear un método para secuenciarlo.

Es interesante comprobar cuántas veces se ha dado en la investigación científica la publicación simultánea e independiente de un mismo resultado. Casos como el de Darwin y Wallace para la teoría de la evolución o el de Newton y Leibniz para el cálculo infinitesimal parecen indicar que, de alguna manera, el momento histórico es el adecuado cuando la necesidad de una explicación coincide con la capacidad para encontrarla. Hay muchos otros ejemplos menos conocidos, entre ellos la publicación simultánea en 1977 de dos métodos casi idénticos para secuenciar el DNA, por Allan Maxam y Walter Gilbert en Harvard (Estados Unidos) y Fred Sanger y Alan Coulson en Cambridge (Reino Unido). Mediante reacciones químicas que distinguían entre las cuatro bases y utilizando un marcaje radiactivo se podía hacer una radiografía de un fragmento de DNA que alcanzaba a ordenar unas 250 bases. Usted ha visto estas radiografías, casi seguro. Intente recordar la última vez que vio en la televisión alguna noticia referente a genética. Si no en la última, seguramente en la penúltima aparecía un investigador, tal vez dos, sosteniendo contra la luz una placa radiográfica llena de líneas horizontales, como un esquema de un rascacielos: una secuencia de DNA.

Durante casi quince años, los métodos de secuenciación con radiactividad fueron el pan de cada día para miles de investigadores. Los primeros genes que saltaron a las páginas de los periódicos en los años 80 fueron secuenciados mediante uno de estos métodos que hoy nos parecen tan lejanos como la televisión en blanco y negro.

A finales de los años 80 una nueva estrategia mejoró la calidad de vida de los investigadores y también la calidad de sus resultados: la secuenciación automática, llevada a cabo por máquinas que utilizan un lector láser para distinguir las cuatro bases. Para ello, se une a cada base mediante un proceso químico una sustancia fluorescente, se aplica el método básico y los datos se analizan directamente en un ordenador.

El láser provoca que cada base brille con un color diferente, según el colorante fluorescente que le hayamos unido. Usted también ha visto estas secuencias en la televisión y en algún suplemento dominical de los periódicos: ¿le suena una pantalla de ordenador llena de líneas de colores, verde, amarillo, azul y rojo? La secuenciación automática permite trabajar sin radiactividad, los resultados son más rápidos y exactos y las lecturas son más largas, desde unas 300 bases por reacción hace quince años hasta casi 1.000 ahora. A un precio, por supuesto: las máquinas, los reactivos y el mantenimiento cuestan una cantidad de dinero que obliga a los laboratorios más modestos a contratar estos servicios a centros más grandes. Parece claro que se ha acabado hacer los experimentos de genética en la cocina de casa (aunque, como veremos en la segunda parte del libro, es posible que en el futuro podamos volver a hacerlos).

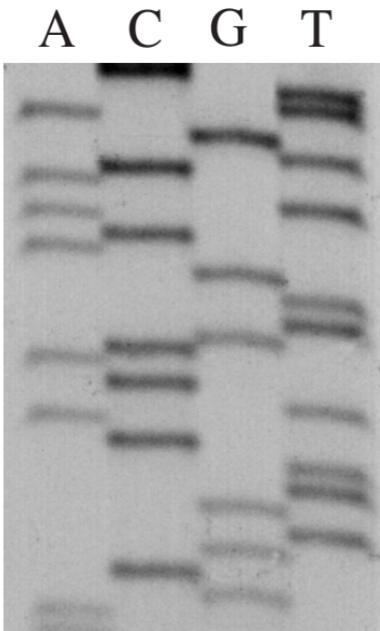


FIGURA 5. Secuencia manual. Los cuatro carriles corresponden a una única muestra, y cada carril indica la posición de una de las cuatro bases en la secuencia. Las líneas horizontales se obtienen gracias a la radiactividad con que se ha marcado anteriormente cada fragmento de DNA. Si empieza a leer por abajo, uniendo los fragmentos de la escalera, verá que esta secuencia es:
 ...AAGCGTGTTTCATCAGTTC...
 Secuencia cedida por Virginia Nunes y Lúdia Feliubadaló. Vea un ejemplo animado de esta técnica en:

<http://vector.cshl.org/resources/resources.html>

Los nuevos aliados: la PCR, la informática e internet

Recortamos el DNA, lo pegamos en una fotocopiadora y lo coloreamos para ver su secuencia. Para estudiar los genes nos ha sido necesaria otra herramienta, una ampliadora que nos permite obtener cantidades suficientes de DNA para trabajar a partir de muy poco sin tener que pasar por el proceso de cortar y pegar (que, aunque sobre el papel parezca muy simple, en la vida real es bastante más lento y escurridizo). Además, esta estrategia no necesita de la colaboración de virus ni bacterias (nuestras fotocopiadoras más habituales).

Me refiero a la reacción en cadena de la polimerasa, familiarmente conocida como PCR. Si un gen en el genoma es una aguja en un pajar, ¿por qué no copiar la aguja varios millones de veces, hasta tener un montón de agujas iguales? Kary Mullis ideó un experimento para amplificar exponencialmente un fragmento de DNA. Armado de un cronómetro y baños de agua a diferentes temperaturas, rompió para siempre la barrera entre la investigación básica y la aplicada: la PCR es una herramienta que convierte a las universidades en empresas y a las empresas en centros de investigación.

Antes de la PCR, el ejemplo de amplificación exponencial más popular era el legendario premio que el inventor del ajedrez pidió a su gobernante por idear un pasatiempo tan perfecto: un grano de arroz por la primera casilla del tablero, dos por la segunda, cuatro por la tercera y el doble cada vez hasta la casilla 64. Intuitivamente no parece gran cosa, pero una vez calculado este número es imposible de satisfacer, ni siquiera por un mandamás persa de la Antigüedad. Así el ajedrez y las matemáticas aplicadas tienen un origen legendario común, y a su lado la PCR es una modesta amplificación exponencial de entre 30 y 40 ciclos: muchos menos que los 64 del ingenioso ajedrecista, pero suficientes como para cambiarle la cara al mundo. La idea de la PCR fue desarrollada, mejorada y puesta en el mercado por Cetus, la empresa donde el doctor Mullis trabajaba. Cetus ganó millones y a él le fue concedido el premio Nobel de Química en 1993. El

doctor Mullis es un premio Nobel atípico: con el dinero del premio ha dedicado estos últimos años a hacer *surf*. En 1999 se le acabó el dinero y volvió a trabajar, como asesor de varias empresas de biotecnología.

La PCR se usa para casi todo, desde las pruebas de paternidad o el diagnóstico de enfermedades hasta la identificación de nuevas especies de algas. En la segunda parte del libro veremos algunos usos de la PCR en campos que nadie habría pensado hace pocos años. La PCR sirve incluso para que la agente Scully identifique extraterrestres en un capítulo de *Expediente X* que sorprendió agradablemente a los biólogos moleculares que vieron cómo su rutina diaria era convertida en ciencia ficción sin perder la base científica.

En el clonaje de genes, la PCR es una lupa que podemos pasear por el genoma para ver con detalle las zonas que nos interesan. Tiene sus limitaciones: hay regiones del genoma difíciles de amplificar, normalmente se requiere algún conocimiento previo de la zona que se estudia y el tamaño de los fragmentos que queremos amplificar es limitado, pero se trata de una herramienta que ha impulsado la investigación de una manera definitiva: en biología hay un antes y un después de la PCR.

La generalización del clonaje y la secuenciación en los laboratorios de todo el mundo nos ha inundado de información. Tanta información es sobrehumana, literalmente: no hay persona capaz de procesarla. Esto ha llevado al desarrollo de la bioinformática: programas especializados en el análisis de datos relacionados con el DNA. Gracias a ellos podemos comparar secuencias o deducir si una secuencia concreta codifica una proteína o si forma parte del 95% del genoma que no se traduce a proteína. Imagine la diferencia entre las antiguas máquinas de escribir y los modernos procesadores de textos; esto no es nada comparado con lo mucho que la bioinformática ha facilitado la vida a los biólogos.

Otro fruto de nuestra época que ha venido como anillo al dedo para permitir el avance del estudio del genoma es inter-

net. Cualquiera que tenga un ordenador conectado a la red puede acceder a una cantidad ingente de datos biológicos, bibliografía y programas (al final del libro encontrará algunas direcciones). Gracias a esta comunicación los datos producidos en un laboratorio pueden ser puestos inmediatamente a disposición de todo el mundo. Para centralizar esta información se han desarrollado bases de datos, algunas generales y otras más específicas, donde los investigadores pueden enviar sus resultados y encontrar información que les pueda interesar. Muchas de estas bases de datos están interconectadas de modo que se puede pasar fácilmente de una secuencia de DNA a una familia de proteínas, a una enfermedad de origen genético y a una publicación donde esta enfermedad está descrita. Hoy, gran parte del trabajo de un biólogo molecular se hace sentado delante de un ordenador. Fíjese que no todas las bases de datos son de acceso gratuito; las hay financiadas con capital privado que requieren cuotas de acceso. Algunos centros de investigación, universidades y empresas farmacéuticas han contratado licencias para acceder a estas bases de datos.

Finalmente, el Proyecto Genoma Humano

Con las herramientas a punto, ahora nos puede parecer que no había otra opción que emprender el estudio en profundidad del genoma humano. Pero en los años 80 gran parte de la comunidad científica no pensaba así.

De alguna manera, existían pequeños proyectos genoma desde finales de los 70, descoordinados y con objetivos limitados. Una propuesta del DoE americano (la agencia federal de energía) de financiar la secuenciación del genoma humano de manera exhaustiva y sistemática fue recibida con una sensación de “ya era hora” por los principales implicados en el desarrollo de la biología molecular. Varias discusiones para aclarar los detalles acabaron con la propuesta de que el proyecto dependiera principalmente del NIH (la agencia

federal de la salud) y no del DoE. Así se aseguraban tanto la orientación científico-médica del trabajo como un control más público que si fuera un proyecto militar. Irónicamente, gran parte de la secuenciación del genoma se ha llevado a cabo en el mismo laboratorio de Los Alamos donde en los años 40 se trabajó en el Proyecto Manhattan que acabó con la construcción de la primera bomba atómica.

Con la incipiente tecnología de secuenciación de la época, este proyecto fue calificado como “Santo Grial” por algunos. Otros temían que la financiación destinada al genoma humano limitase la investigación en otros campos. Uno de los primeros impulsores del proyecto fue el premio Nobel Renato Dulbecco, que en 1986 publicó un artículo sobre este tema en la revista *Science*. Otro de los que vieron en seguida la utilidad del proyecto fue el también premio Nobel James Watson, quien gozaba de un merecido prestigio como codescubridor de la estructura del DNA, la famosa doble hélice, y un gran poder de persuasión. En 1986 se anunció la creación de la Iniciativa del Genoma Humano, con el objetivo de secuenciar todo el genoma y ofrecer un catálogo de todos los genes. Poco después, en 1990, se redefinió el trabajo con la incorporación de laboratorios de Europa y Japón, y se emprendió el Proyecto Genoma Humano, bajo la dirección del doctor Watson. La meta era tener el genoma secuenciado en el año 2005. En 1992 el doctor Watson dio por acabado su ciclo de cuatro años al frente del proyecto por discrepancias respecto a la comercialización de los descubrimientos relacionados con el genoma: como veremos en el capítulo 4, a principios de los años 90 el NIH solicitó patentes sobre una gran cantidad de secuencias obtenidas en sus centros de investigación. James Watson es conocido como *honest Jim* por su carácter contrario a la falsa modestia y la diplomacia. A pesar de ello, incluso sus detractores le reconocen una capacidad científica que lo ha mantenido en la cumbre durante medio siglo. Desde su puesto como presidente del Cold Spring Harbor Laboratory continúa siendo una influencia de primer orden en la investigación genética.

Su puesto lo ocupó Francis Collins, de una generación más joven y con dos clonajes históricos a sus espaldas: los genes implicados en la fibrosis quística y la neurofibromatosis. La mejora de la tecnología permitió revisar los objetivos y acercar la fecha de finalización del proyecto al año 2003, coincidiendo con el cincuentenario del descubrimiento de la doble hélice. El primer borrador se presentó a la prensa en junio del año 2000 y a la comunidad científica en febrero de 2001, y es posible que el resultado final esté a punto antes de la fecha programada inicialmente.

Un trabajo que implica a miles de investigadores de varios países y que mueve presupuestos millonarios necesita una coordinación muy delicada. La Organización del Genoma Humano (HUGO) fue creada en 1989 por los principales impulsores del Proyecto Genoma Humano para ejercer esta tarea de coordinación; es una especie de Naciones Unidas en que la representación no es nacional sino científica.

El trabajo de secuenciación del genoma se ha repartido para optimizar el esfuerzo, de manera que a diferentes grupos les ha sido asignado un fragmento de cromosoma. El Consorcio de Secuenciación del Genoma Humano está formado por veinte instituciones, cinco de las cuales desempeñan un papel centralizador: el Sanger Centre (donde se ha secuenciado un tercio del total), en el Reino Unido y el Baylor College of Medicine, el Whitehead Institute for Biomedical Research/MIT, el Genome Sequencing Center de la Universidad de Washington y el Joint Genome Institute, estos últimos en los Estados Unidos. Además, dos centros han proporcionado el soporte informático necesario para que los datos fueran ordenados y accesibles: el National Center for Biotechnology Information (NCBI, en los Estados Unidos) y el European Bioinformatics Institute (EBI, en el Reino Unido). Al final del libro encontrará todas estas direcciones. Cualquiera de ellas es una puerta hacia el genoma y todas incluyen material didáctico de primera calidad.

¿Cómo se ha secuenciado el genoma? El DNA que ha

secuenciado el Proyecto Genoma Humano se obtuvo a partir de muestras de sangre de 12 donantes anónimos: la secuencia que leemos es una mezcla de varias personas. La mayor parte del trabajo se ha realizado sobre el DNA de un hombre anónimo, mezclado con muestras de otras personas de orígenes étnicos variados. El otro grupo de investigadores que ha efectuado este trabajo, en la empresa Celera Genomics (que veremos con más detalle dentro de unos párrafos), utilizó una mezcla de DNA de dos hombres y tres mujeres de orígenes étnicos diversos, escogidos entre 21 donantes anónimos, todo ello perfectamente certificado por lo que respecta a la confidencialidad.

El Proyecto Genoma Humano no sólo se propone leer el genoma, sino leerlo unas diez veces para estar seguros de que no hay ningún error. Los centros coordinadores actualizan sus bases de datos cada 24 horas a partir de los datos que les envían los laboratorios del consorcio. Esto quiere decir que las secuencias están disponibles inmediatamente, pero antes de usarlas hay que fijarse en su grado de exactitud: una secuencia que no esté repasada un número suficiente de veces puede tener errores que lleven a conclusiones equivocadas.

El resultado de esta multitudinaria secuenciación es un mapa para cada cromosoma, 24 largas listas de bases que incluyen genes, regiones reguladoras y un 95% de DNA de función desconocida. Conocer la secuencia del DNA no implica conocer los genes, tan sólo facilita la búsqueda posterior. El siguiente paso se llama “anotación” y consiste, literalmente, en poner notas al lado de cada secuencia con lo que se sabe de ella: si es un gen, si se conoce su función, si se ha encontrado un gen homólogo en otros animales, cualquier dato que pueda ayudar a quien quiera estudiar esa secuencia. Encontrar los datos que se necesitan para un estudio en medio de toda la montaña de datos superfluos es un trabajo duro: la “minería genética” requiere mucha preparación para saber reconocer un filón a partir de indicios poco prometedores. Una parte del trabajo es rutina de pico y pala, para lle-

gar rápido a la zona que se quiere estudiar más a conciencia, pero luego es necesario un conocimiento muy preciso de lo que se busca y de las herramientas informáticas necesarias para extraer los datos. La expresión inglesa usada para referirse a la búsqueda en bases de datos es “*data mining*”, que tiene esta connotación de encontrar un filón entre la ganga, mientras que “extracción de datos” sugiere algo más fácil, como abrir la nevera para extraer una botella de zumo de naranja. Con la secuencia de nuestro genoma y el de otros organismos disponibles en las bases de datos, los mineros genéticos van a tener trabajo durante años.

Chromosome 17, Bases 45246031-45369256, Size 123226

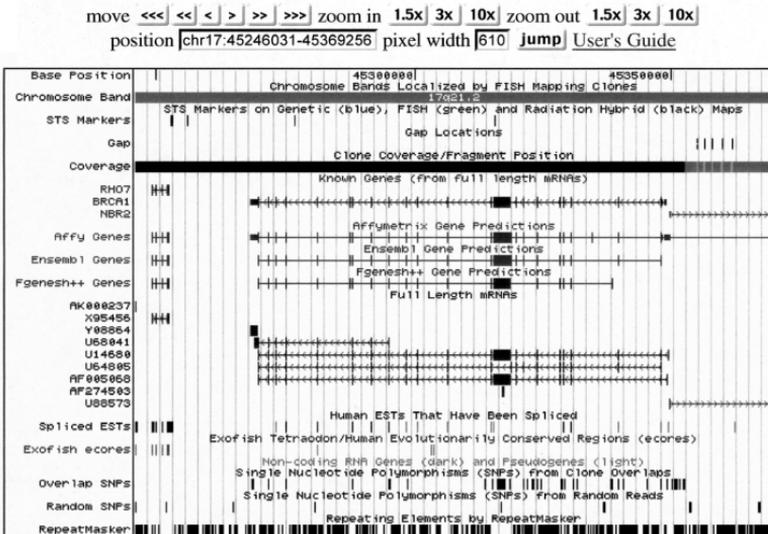


FIGURA 6. Este es el aspecto que tiene el genoma en las bases de datos. Para este ejemplo hemos buscado el gen *BRCA1*, implicado en algunos casos de cáncer de mama y ovario hereditarios. En un sitio web llamado “Golden Path” (“sendero de oro”, <http://genome.cse.ucsc.edu/goldenPath/octTracks.html>) puede teclear el nombre de un gen y obtener información acerca de su secuencia, los SNPs que se han encontrado y otros datos.

Si ha seguido la historia del Proyecto Genoma Humano a través de los medios de comunicación, le habrá sorprendido que el número de genes que se predecían en nuestro genoma ha ido variando con el tiempo. Hasta hace poco se solía dar un número redondo: 100.000 genes en total. Este número no sólo tenía un gran valor simbólico sino que representaría el número más alto que se podría encontrar en la naturaleza, de acuerdo con lo que cabe esperar de un organismo tan complejo como el hombre. Con la aparición de las bases de datos privadas y las empresas de biotecnología, el número de genes predichos se incrementó de manera considerable: tener más genes aumentaba el valor potencial de la empresa.

Sin embargo, la secuenciación completa de los dos primeros cromosomas (los más pequeños: el 21 y el 22) obligó a corregir las previsiones. El cromosoma 22 era considerado rico en genes y, una vez que estuvo secuenciado, se le asignó un número de genes de 545 (este número se ha corregido a 561 al cabo de un año). En el cromosoma 21, considerado pobre en genes, se han predicho 225. Entre los dos representan un 2% del genoma y una proyección de estos resultados arroja un número de genes cercano a los 40.000. Hace dos años que conocemos este dato, por lo que no es especialmente sorprendente que el número estimado de genes después de la presentación del primer borrador de la secuencia del genoma se cifre en unos 30.000.

Hay varias maneras de predecir el número de genes que se podrán encontrar, según qué parámetros se consideren más importantes y qué programas informáticos se utilicen. Además, hemos visto que el concepto de gen es muy impreciso, porque esta palabra no deja de ser un instrumento de trabajo: una definición que incluyera todos los casos sería demasiado complicada para ser una verdadera definición y más bien habría que llamarla “discusión”. Para acabar de complicar esta pregunta, los bioinformáticos que se dedican a la detección de genes en las secuencias de DNA han tomado prestado un término de la astrofísica y hablan con familiaridad (y cierta irritación, todo hay que decirlo) de la

“materia negra”. ¿A qué se refieren? Se refieren a que sólo sabemos buscar cuando somos capaces de reconocer que hemos encontrado lo que buscábamos. Aplicado a los ordenadores, quiere decir que, para que encuentren genes, primero les tenemos que decir cómo es un gen. Por lo tanto, los genes que son demasiado diferentes de los que ya conocemos se escapan a la detección, aunque veamos sus efectos. Los tenemos delante y no sabemos verlos.

Una de las características de la biología molecular es que hay mucho margen para el sentido del humor. No es extraño ver muestras de ello en los encuentros de especialistas: de alguna manera se tienen que desdramatizar las largas jornadas laborales y la elusividad de los resultados. Como no podía ser menos, el genoma humano no se ha librado de ello. El año 2000 se puso en marcha una porra en la que los participantes apuestan un dólar (cinco en el año 2001 y veinte en el 2002) a un número de genes. En la actualización más reciente de esta página el promedio es de 61.710 genes, con apuestas que oscilan entre los 27.462 y los 153.478. En un congreso, a celebrar el año 2003, una vez que la secuencia del genoma humano esté leída, repasada y anotada con la suficiente exactitud, se dará a conocer el número de genes y el nombre del ganador. Éste se llevará todo el dinero del bote y una copia dedicada del libro *La doble hélice*, de James Watson.

Pero conocer el número de genes es anecdótico. Incluso con el mapa completo encima de la mesa, conocer la posición y la secuencia de un gen no es lo mismo que conocer su función. Además, un mismo gen puede dar lugar a varias proteínas, que pueden tener distintas funciones. ¿Cómo? Combinando fragmentos de un mismo gen de diferentes maneras, en un proceso conocido como “ayuste” (o más a menudo en su forma inglesa *splicing*). En el capítulo anterior hemos visto que un gen está formado por una secuencia codificante interrumpida por largos trechos de secuencia no codificante (que puede tener un papel regulador, pero no codifica aminoácidos). En el capítulo 9 veremos que una medida de la

complejidad de un organismo es la presencia y el tamaño de estas secuencias no codificantes llamadas “intrones”. Algunos genes sólo tienen una manera de combinar sus fragmentos de secuencia codificante, los llamados “exones”. Otros genes pueden combinar estos fragmentos de diferentes maneras, en diferentes momentos del desarrollo o en diferentes tejidos. Se calcula que más del 35% de los genes humanos utilizan este ajuste alternativo, y esta cifra puede llegar a más del 60%. Esto da lugar a proteínas que pueden ser parecidas, incluso iguales en muchos tramos, pero que desempeñan papeles diferentes en la célula. Un ejemplo de cómo un número limitado de genes puede dar lugar a infinidad de proteínas lo tiene en el sistema inmunitario: los genes que codifican anticuerpos combinan unos pocos elementos, y con ellos pueden reconocer cualquier elemento extraño que circule por su sangre, incluso (o especialmente) si no lo habían encontrado nunca antes.

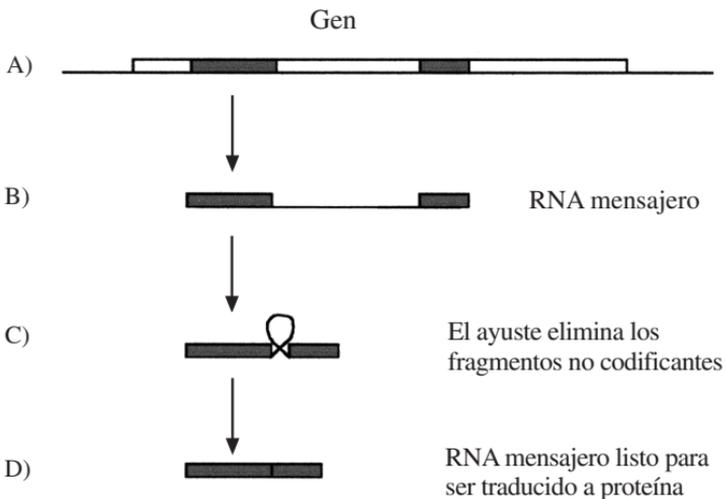


FIGURA 7. Ayuste (o *splicing*, en inglés).

Recuerde que casi todo lo que pasa en una célula, desde su forma hasta sus interacciones, es mayoritariamente obra de proteínas. Por tanto, un paso lógico en la era del genoma es la puesta en marcha de un Proyecto Proteoma Humano. Su objetivo es conocer todas las proteínas que se pueden encontrar en nuestro cuerpo y las interacciones que se dan entre ellas. El título de una reunión fundacional en abril de 2001 lo dice todo: “El Proyecto Proteoma Humano. Los genes fueron fáciles”. Recientemente se ha presentado una base de datos de proteínas encontradas en tejidos diferentes, procedentes de una donante fallecida de un infarto. En total, unas 150.000 proteínas que se calcula que son codificadas por unos 18.000 genes; como ve, un gen puede codificar varias proteínas diferentes. Está claro que la secuenciación del genoma es sólo el principio.

El Proyecto Genoma Humano es un inicio, no un fin. Hace años que conocemos los genomas de la levadura o la mosca, pero los avances en el estudio de estas especies aún requieren inteligencia, trabajo y suerte. Entre unos y otros hemos puesto en la criatura muchas esperanzas y tenemos poca paciencia para esperar a que madure lo suficiente, pero no hay duda de que los biólogos de las próximas décadas usarán esta información como si fuera un diccionario: sin él no se pueden escribir novelas, pero aún con él es necesario un gran esfuerzo y una dosis de genialidad.

No todo el mundo está igual de convencido de la utilidad de este trabajo. Dentro y fuera de la comunidad científica hay quien denuncia un supuesto enfoque determinista en el estudio del genoma. Normalmente este enfoque está más presente en el discurso de los detractores que en el de los investigadores que estudian el genoma. Otros científicos tachan el proyecto de burocrático y rutinario, sin interés científico. La gran cantidad de talento que se ha destinado a este campo en las últimas dos décadas y los avances en el conocimiento que estamos viendo contradicen esta opinión. Fuera de la comunidad científica, algunos grupos han visto la investigación sobre el genoma como un ejemplo de capitalismo, machis-

mo, colonialismo y varias calamidades más. Queda muy lejos del objetivo de este libro el analizar estos argumentos; a medida que los resultados positivos lleguen a la sociedad es de esperar que las enmiendas a la totalidad pierdan terreno. Este debate ha sido una constante de los últimos años y sólo acaba de empezar.

Porque no todo es ciencia en el Proyecto Genoma Humano: las diferentes instituciones y organizaciones participantes dedican alrededor del 5% de su presupuesto al estudio de las implicaciones éticas, sociales y legales del estudio del genoma, un área conocida como ELSI (del inglés *Ethical, legal and social issues*). Independientemente de los comités ELSI de los centros participantes, la Organización del Genoma Humano dispone de un comité de ética, con más libertad de acción al no depender del presupuesto de las organizaciones directamente implicadas en la secuenciación del genoma. El comité de ética del HUGO está formado hoy por trece personas de reconocido prestigio en sus especialidades, ya sean la biología, la medicina, el derecho o la ética, y su presidenta es la doctora Bartha Maria Knoppers, de la Facultad de Derecho de la Universidad de Montreal. Sus documentos no son vinculantes, porque no cuentan con el respaldo de ningún gobierno, pero tienen la autoridad que los propios científicos han depositado en el HUGO y son adoptados como propios por la Organización.

Probablemente el aspecto más olvidado del proyecto sea el referido a la comunicación pública de los trabajos. Las páginas web de los centros participantes incluyen información básica para el público no especializado, pero hasta hace poco no se ha realizado un esfuerzo decidido para informar a la sociedad. Esto ha provocado que, algunas veces, las interpretaciones del trabajo no estén muy de acuerdo con la realidad.

El Proyecto Genoma Humano es un trabajo llevado a cabo por laboratorios financiados en gran parte con dinero público, aunque algunas empresas farmacéuticas y especialmente una fundación privada (el Wellcome Trust) han patrocinado una

parte importante del mismo. Las herramientas que hemos visto al principio del capítulo también han impulsado la aparición de multitud de empresas, pequeñas y grandes, que ofrecen servicios relacionados con este tema. Las empresas de biotecnología aparecen en los periódicos cada día, y ninguna más a menudo que Celera Genomics. En 1998, el doctor Craig Venter, un colaborador del Proyecto Genoma Humano, ideó un método más rápido para hacer el trabajo. Los organismos financiadores del proyecto consideraron que esta estrategia no sería viable y rechazaron su propuesta, lo cual le llevó a fundar su propia empresa, Celera Genomics. Su estrategia es más rápida pero menos sistemática y propone sólo unas cuatro lecturas del genoma en lugar de las diez que se propone el Proyecto Genoma Humano. A pesar de las críticas, la prueba de que este sistema funciona fue la publicación el año 2000 del genoma de la mosca del vinagre, *Drosophila melanogaster*, el organismo más complejo secuenciado hasta la fecha. En el primer trimestre de 2001 los responsables del Proyecto Genoma Humano reconocieron tácitamente que esta estrategia es válida al utilizarla para ultimar la secuenciación del genoma del ratón, si bien es cierto que el Proyecto Genoma Humano utilizó esta técnica en la última fase de su trabajo, después de la primera fase en la cual se avanzaba más lentamente a cambio de tener un resultado más exacto. Por otro lado, Celera utilizó los mapas elaborados por el Proyecto Genoma Humano para ensamblar partes de su secuencia: cada grupo tenía herramientas que el otro valoraba.

Esta competencia inesperada obligó a acelerar el trabajo para no ceder la parte de prestigio que corresponde a los pioneros. La solución a un problema de magnitud bíblica sólo podía ser salomónica: en mayo de 2000 se presentó en rueda de prensa un borrador del genoma humano, de manera simultánea por los doctores Collins y Venter. De alguna manera, esta presentación era un reconocimiento de que ambos tienen asegurado un lugar en los libros de historia y que, honores aparte, aún queda mucho trabajo por hacer. A mediados de febrero de 2001 han aparecido dos artículos en revistas espe-

cializadas, uno en *Nature* y otro en *Science*, donde cada grupo ha presentado este borrador de manera más formal a la comunidad científica internacional. Esta publicación no está libre de polémica, por la difícil negociación que ha llevado a que Celera acepte publicar gratuitamente unos datos de los que espera obtener beneficios económicos: las revistas especializadas no publican ningún trabajo si los datos, en este caso las secuencias de DNA, no están a la vista de cualquiera que los quiera comprobar. El trato entre Celera y *Science* permite el acceso gratuito a su base de datos, con algunas restricciones, a los investigadores que trabajan en centros públicos (y con más restricciones a los que lo hacen en empresas privadas). Por si Celera cambia de opinión en un futuro la revista *Science* guarda una copia de esta base de datos. Esta situación es inédita, y no es extraño que las reacciones ante este trato hayan sido mayoritariamente adversas. Los editores de *Science* argumentan que es mejor tener esta secuencia así que no tenerla y que hay que adaptarse a los cambios, mientras que los detractores del pacto argumentan que no se puede tener al mismo tiempo el prestigio académico de publicar en una revista seria y restringir el acceso a los datos que se publican. Como suele suceder, todos tienen parte de razón.

Los que han visto este trabajo como una competición se preguntan: ¿quién ha ganado? La competencia entre equipos rivales es habitual en la investigación y no es raro que la misma revista publique simultáneamente dos o tres trabajos semejantes. Nadie gana ni nadie pierde; simplemente, más resultados están a la vista de quien los quiera estudiar. La carrera por secuenciar el genoma ha tenido su parte de rivalidad personal entre los líderes de los grupos y para algunos ha representado un modelo de la sociedad: la iniciativa pública contra la privada, donde cada cual ha puesto sus simpatías en uno u otro campo en función de sus convicciones. Con los resultados encima de la mesa, los dos trabajos son diferentes pero semejantes y no hay manera de declarar un ganador. Ni necesidad, realmente.

Los otros proyectos genoma

Por si no fuera suficiente con uno, existen varios proyectos que persiguen objetivos paralelos y complementarios al Proyecto Genoma Humano. Las técnicas, la base teórica y las herramientas son fundamentalmente las mismas, pero la orientación y el tipo de información que se espera obtener difieren.

Una pregunta que seguramente usted se había hecho a propósito del Proyecto Genoma Humano es: ¿de qué nos va a servir una secuencia concreta cuando es evidente que todos somos diferentes?

Esta misma cuestión también se la han planteado varios equipos de científicos y, entre ellos, el equipo liderado por Luigi Luca Cavalli-Sforza ha intentado completar un registro de la diversidad genética humana. Para este estudio se intentan obtener muestras de poblaciones variadas, preferiblemente aisladas o minoritarias, con la idea de preservar los alelos que no están presentes en las poblaciones más mezcladas o mayoritarias. Veremos en el capítulo 6 que todos somos muy parecidos pero, aun así, hay variantes de cada gen que sólo se encuentran en algunas poblaciones. Así explicado no parece un tema especialmente conflictivo, pero en la práctica este proyecto ha encontrado una fuerte oposición.

Por un lado, las organizaciones que financian la investigación en EE.UU. se oponen a este registro porque reconoce de una manera explícita que hay diferencias genéticas entre grupos humanos. La tensión racial, los problemas legales asociados a la distinción de razas y el temor a un uso discriminatorio de esta información han bloqueado este estudio en Estados Unidos. Por otro lado, en África y América Central y del Sur, a algunas comunidades se les ha hecho creer que si los científicos sacaban una *foto* de su DNA les robarían el alma. La analogía no es exagerada: es muy fácil convencer a alguien de que su DNA oculta misterios que más vale no revelar a otros. Hay que reconocer que algunos grupos, como

los aborígenes australianos, pueden quejarse, con razón, del beneficio nulo obtenido de sus colaboraciones con la ciencia de hace 20 o 30 años. La reciente incorporación de aborígenes a la comunidad científica debe tender puentes para reparar esta desconfianza. En otros países, como la India o la China, las leyes no permiten la exportación de muestras de DNA. Estos países y otros han iniciado registros de DNA de sus poblaciones minoritarias. En opinión del doctor Cavalli-Sforza, no es importante quién haga este trabajo, o si está repartido entre centros de todo el mundo, pero sí lo es que alguien lo haga. Mientras tanto, y pese a su innegable interés, el Proyecto de la Diversidad del Genoma Humano (*Human Genome Diversity Project, HDGP*) está gravemente muerto.

Un estudio que ha empezado recientemente y goza de muy buena salud es el de los SNPs (pronúnciese “snips”). SNP es la abreviación de *Single Nucleotide Polymorphism*, y se refiere a las diferentes formas (“polimorfismos”) que puede presentar una base concreta del genoma (un “único nucleótido”). Un SNP es un punto en el genoma que difiere de una persona a otra: por ejemplo, donde una persona tiene una A, otra tiene una C, y se asume que cada uno de estos cambios se ha dado una única vez durante la evolución. Algunos SNPs están dentro de genes y son la causa de la variabilidad que se puede ver en cualquier característica, desde el color del pelo hasta la respuesta a los antiinflamatorios. Otros están en zonas no codificantes del genoma y no sabemos si influyen de alguna manera en los genes próximos.

El interés de este estudio es doble. Por un lado, los SNPs nos pueden ayudar a comprender cuestiones como la diferente respuesta a los medicamentos en personas que padecen una misma enfermedad. Por otro lado, un mapa de SNPs facilitaría la identificación de genes relacionados con enfermedades complejas.

Esta utilidad principalmente farmacológica explica que el mayor impulsor de este trabajo sea un consorcio formado por las diez empresas farmacéuticas más activas en la investiga-

ción, además de cinco universidades, dos empresas de comunicación (IBM y Motorola) y el Wellcome Trust, la mayor institución filantrópica británica. Muestras de DNA que incluyen diferentes grupos étnicos son comparadas entre ellas y con otras muestras de lugares variados. Así se obtienen secuencias que, a veces, pueden ser diferentes en sólo una base. Si este cambio está presente en un número suficiente de personas se considera que es un SNP y se introduce en una base de datos. Las secuencias son accesibles gratuitamente a través del portal de *The SNP Consortium*. Este consorcio ha puesto a disposición del público algo más de un millón de SNPs: un SNP cada 3.000 bases. Hay más bases de datos dedicadas a los SNPs. Una de ellas llamada dbSNP, mantenida por el NCBI y donde se recogen SNPs descubiertos por investigadores de todo el mundo, dentro y fuera del Consorcio, tenía en su última actualización 2.558.364 SNPs, aunque no todos ellos verificados por igual. El SNP Map Working Group (con la colaboración de The SNP Consortium) ha publicado 1.400.000 SNPs simultáneamente a la secuencia del Proyecto Genoma Humano. Un mapa de SNPs es de una resolución muy conveniente para los investigadores, ya que no es tan detallado como para mostrar cada base individualmente pero indica los puntos variables, los que pueden dar información.

Otra manera de obtener información de las diferencias es el Proyecto Genoma del Cáncer, que se propone comparar DNA de células normales y de células cancerosas para identificar todos los genes implicados en el cáncer de una manera u otra. El resultado será un mapa de puntos débiles de la célula ante el cáncer y ha de ayudar a idear tratamientos más eficaces. Esta investigación se inició en febrero de 2000 en el Sanger Centre y sería imposible si no se dispusiera de la secuencia del genoma humano. Desde cierto tiempo antes existía un Proyecto de Anatomía Genómica del Cáncer, que recoge en una base de datos un catálogo de los RNA mensajeros que se pueden encontrar en diferentes tejidos cancerosos. Capturar al mensajero es una estrategia bélica de gran

eficacia, y en la guerra contra el cáncer todo vale. La combinación de los datos de estas dos bases de datos ha de facilitar enormemente la investigación.

Fuera del genoma humano hay multitud de otros proyectos, en marcha o terminados: los genomas de unas 60 especies han sido completamente secuenciados, entre ellos los genomas de bacterias, virus, levaduras, la mosca del vinagre, dos plantas (una especie de mostaza llamada *Arabidopsis thaliana* y el arroz) e incluso un gusano (*Caenorhabditis elegans*). Alrededor de cien más están en camino, entre ellos los genomas de los animales y plantas de mayor valor comercial. La secuenciación del genoma del ratón, fruto de la colaboración de entidades públicas y empresas privadas, estará finalizada durante este año 2001. Mientras se escribían estas líneas, Celera anunció que tenía a punto su versión de este mismo genoma, pero en este caso no se espera otra publicación simultánea de resultados. ¿Por qué nos interesan estos genomas? ¿Qué nos pueden decir el ratón o la mosca del vinagre que no nos hayan dicho ya?

Algunos de estos genomas los queremos conocer para descubrir sus puntos débiles: los virus y las bacterias que nos causan enfermedades están mostrando sus armas a los equipos de secuenciadores. Sus genomas nos muestran cómo escapan a los antibióticos y a las células del cuerpo que se ocupan de eliminar a los intrusos: algunos se disfrazan para no ser reconocidos como extraños, otros encuentran maneras de desactivar o destruir los medicamentos con que les atacamos. Identificando estas estrategias podremos combatirlos más precisamente.

Otros genomas nos pueden hablar de nosotros mismos: la evolución ha dado lugar a unas pocas estrategias que se repiten en organismos tan diferentes como la mosca del vinagre, el sapo y nosotros mismos. ¿Estrategias de qué? De copiar el DNA, repararlo y empaquetarlo, de producir enzimas y proteínas estructurales o de construir una célula con todos sus componentes y compartimentos, por ejemplo. Los genes implicados en estos procesos son diferentes, a veces incluso

mucho, entre nosotros y la levadura, pero es posible agruparlos de acuerdo con su papel en el organismo. Un grupo de investigadores especialistas en organismos tan diversos como la levadura de la cerveza, la mosca del vinagre y el ratón ha iniciado un estudio comparativo de estas genomas, y es de esperar que a medida que nuevas genomas sean secuenciadas se incorporen más investigadores a este grupo para ayudar a dibujar los esquemas del funcionamiento de los seres vivos. Comparando nuestro genoma con los genomas más *simples* podemos ver qué vías metabólicas se han conservado, qué estrategias han aparecido y qué genes son importantes para el desarrollo y la función de cada una de las partes de nuestro cuerpo.

Hemos visto que el Proyecto Genoma Humano es un inicio, una herramienta elaborada gracias a la colaboración de una comunidad multinacional. Las instrucciones para usarla no son más complicadas que las de un vídeo doméstico, siempre que se tengan presentes dos conceptos. Uno, que lo que vemos no es el alma de la humanidad, sino el material base a partir del cual se construyen nuestras piezas: una pieza defectuosa puede dificultar la construcción de un individuo, pero el ambiente, la educación y el azar contribuyen a hacernos lo que somos. Y dos, que nuestro genoma es dinámico y único: la secuencia que vemos en las bases de datos es muy parecida a la nuestra, incluso igual en muchas partes, pero con un promedio de una diferencia cada mil bases. Además, esta secuencia es dinámica: a lo largo de la vida se producen cambios en nuestro DNA que pueden influir en nuestro aspecto, nuestro envejecimiento, nuestra muerte y la información que heredan nuestros descendientes. El mapa del genoma es una foto de nuestra evolución.