

2002



REAL ACADEMIA DE MEDICINA Y CIRUGÍA DE MURCIA

SESIÓN PÚBLICA Y SOLEMNE
DE LA INAUGURACIÓN DEL CURSO 2002



DISCURSO DOCTRINAL SOBRE

‘La secuencia del siglo’

POR EL ILMO. SR. D. FRANCISCO J. MURILLO ARAUJO,
ACADÉMICO DE NÚMERO



MURCIA, 24 DE ENERO DE 2002

FMA

REAL ACADEMIA DE MEDICINA
Y CIRUGÍA DE MURCIA

SESIÓN PÚBLICA Y SOLEMNE
DE LA INAUGURACIÓN DEL CURSO 2002

DISCURSO DOCTRINAL SOBRE

‘La secuencia del siglo’

FRANCISCO J. MURILLO ARAUJO

*Académico de Número de la Academia de Medicina y Cirugía de Murcia
Catedrático de Genética de la Universidad de Murcia*

Excmo. Sr. Presidente de la Comunidad Autónoma de la Región de Murcia,
Excmo. Sr. Presidente de la Academia de Medicina y Cirugía de Murcia,
Excmas. e Ilmas. Autoridades,
Ilmos. Sres. Académicos,
Sras. y Sres.:

El azar ha *determinado* que, a pocos meses de la publicación de la secuencia casi completa del genoma humano, le corresponda a este humilde practicante (amante) de la Genética el honor y la responsabilidad de la lección de apertura de curso de la Real Academia de Medicina y Cirugía de Murcia. ¿Cómo sustraerse a esta señal del destino y no dedicar mi discurso a glosar al menos algunos aspectos de tal acontecimiento, considerado un hito, no sólo del devenir de la ciencia sino de la historia general de la humanidad?

Para empezar, no me resisto a traducir los primeros párrafos de uno de los dos artículos originales sobre la secuencia del genoma humano, el aparecido en la prestigiosa revista *Nature*, en su número del 15 de febrero de 2001. Titulado *Initial sequencing and analysis of the human genome*, y firmado por el *International Human Genome Sequencing Consortium*, contiene 61 páginas de abigarrado texto, datos y figuras. Contiene, además, varias láminas desplegadas que incluyen esquemas de todos los cromosomas humanos, con diversas anotaciones, incluidos los genes presentes en cada uno de ellos; genes, por cierto, predichos en su mayoría por programas informáticos y, por tanto, todavía hipotéticos. He aquí la traducción libre de esos primeros párrafos, con algunas acotaciones propias.

«El nuevo descubrimiento de las leyes de Mendel en las primeras semanas del siglo XX encendió la mecha de una indagación científica para comprender la naturaleza y el contenido de la información genética que ha impulsado la Biología de los últimos cien años. Los progresos alcanzados se distribuyen de forma natural en cuatro períodos, que se corresponden, aproximadamente, con los cuatro cuartos del siglo. Durante el primer cuarto se estableció la base celular de la herencia, los cromosomas. En el segundo se definió su base molecular, la hélice doble de DNA. El tercer cuarto de siglo llevó a desentrañar la base informativa de la herencia, con el descubrimiento del mecanismo biológico mediante el que las células expresan la información contenida en los genes, añadiéndose a ello la invención de las técnicas del DNA recombinante, técnicas que permiten a los científicos hacer (en el tubo de ensayo) lo mismo que las células, reproducir y expresar a voluntad cualquier tramo de DNA (clonación molecular), así como determinar su secuencia de nucleótidos (nombre que reciben las cuatro 'letras químicas' en las que está escrito el mensaje genético).

El último cuarto de siglo se ha distinguido por una labor incansable de análisis, primero de genes individuales y luego de genomas completos, dando así nacimiento al nuevo campo de la Genómica. Los frutos de esa labor incluyen ya las secuencias genómicas completas de 599 virus y viroides, 185 orgánulos celulares (mitocondrias y cloroplastos), 31 eubacterias, 7 arqueobacterias, un hongo, dos animales y una planta.

Ahora presentamos aquí los resultados de un esfuerzo de colaboración de 20 grupos, de Estados Unidos, Reino Unido, Japón, Francia, Alemania y China, para generar un primer borrador de la secuencia del genoma humano. El borrador de la secuencia representa el 94%, aproximadamente, de todo el genoma. La secuencia se ha obtenido en un período de tiempo relativamente corto, pasando de cubrir el 10% a cubrir más del 90% en unos quince meses. La secuencia se ha ido poniendo a disposición pública, sin restricción alguna y actualizada diariamente, a lo largo de todo el proyecto

Diversos aspectos de la secuencia del genoma humano resultan de interés. Constituye el genoma más largo (del orden de 3.000 millones de letras) secuenciado hasta ahora de forma extensa, siendo 25 veces mayor que ningún otro secuenciado previamente y 8 veces mayor que la suma de todos ellos. Se trata del

primer genoma secuenciado de un vertebrado. Y, lo más singular, es el genoma de nuestra propia especie.»

Es al hilo de estos párrafos que quisiera compartir con ustedes algunas consideraciones. Tan histórico artículo comienza precisamente por un homenaje a la historia, la de la Genética y la de la Biología molecular en general. Entre el público común no es rara la idea de que el progreso de la ciencia ha ido unido normalmente a insólitos personajes, cuya extraordinaria inteligencia les llevó a desvelar en un momento, de un golpe (el famoso ¡eureka! de Arquímedes), revolucionarias teorías que convulsionaron sus disciplinas, estableciendo nuevos paradigmas. En realidad, pocas veces ha ocurrido así. La ciencia, prudentemente conservadora, progresa en general pasito a pasito. No puede ser de otra forma, dada la propia naturaleza del quehacer científico, críticamente celoso de la acumulación de pruebas experimentales antes de dar por establecida teoría alguna. Aquellos que aparecen como períodos explosivos de avance de una disciplina científica no son, en la mayoría de los casos, más que la aceleración final que sigue lógicamente a un tiempo previo de largas y pausadas exploraciones, a menudo en varias disciplinas distintas. Es el prudente análisis lógico del resultado de experimentos controlados y repetibles el que distingue a la Ciencia, considerada en sentido estricto, de otras actividades, de índole artística, filosófica o religiosa, que se afanan, noblemente en muchos casos, bien que infiltradas de intereses espúreos en algunos otros, en explicar cuanto nos rodea. Es de ese cauteloso comportamiento del que emana la especial *veracidad* de las explicaciones científicas y su enorme poder de predicción y aplicación. Ciertamente es, también, que de tal obligado comportamiento derivan las limitaciones de la Ciencia, en especial la que se refiere a sus objetos de estudio posibles en cada momento. Ello da pie a la consabida y despreciativa frase «la Ciencia no lo explica todo». Aparte de que en algún caso oculte la frustración del ignorante, o el interés malicioso del especulador, tal frase es injusta por desenfocada. Sería más ajustado decir que «no todo puede ser objeto del análisis científico» (afortunadamente, añadirían el ignorante y el especulador; desafortunadamente, añadiría una persona civilizada).

Es cierto, sin embargo, que a lo largo de la historia de la ciencia se han producido algunos grandes *saltos*, anticipaciones conceptuales en el tiempo de tal magnitud que, casi inevitablemente, fueron recibidas con escepticismo, sino con franco menosprecio. Tal es el caso de Mendel, al que el trascendental artículo de Nature rinde justo homenaje en su primera frase. A mediados del siglo XIX, anticipándose muchos años a su tiem-

po, y tras una serie de experimentos de una elegancia pocas veces igualada, Mendel reconoció por primera vez la existencia de los genes, estableció su carácter de entidades físicas discretas, y explicó su forma de transmisión de generación en generación. Cabe decir que aunque los experimentos y conclusiones de Mendel se repitieron justo a primeros del siglo XX, la existencia de los genes como portadores de la información hereditaria no fue aceptada por toda la comunidad científica en varios años, otro ejemplo más de la parsimonia de la ciencia. Todavía en 1917, Thomas H. Morgan debió escribir un famoso texto que derribara las últimas dudas sobre la existencia y el comportamiento de los genes. Trabajando en el Instituto Tecnológico de California, con la famosa mosca del vinagre y junto a un grupo de brillantes colaboradores, Morgan protagonizaría el primer período al que hace referencia el artículo del genoma humano, demostrando que los genes se disponen linealmente en los cromosomas y estableciendo un procedimiento para cartografiarlos.

No me extenderé mucho más en los descubrimientos del recién pasado siglo que han precedido y hecho posible la hazaña del genoma humano. Pero no puedo dejar de destacar el papel de las investigaciones que llevaron a identificar al ácido desoxirribonucleico (el DNA) como la materia de la que están hecho los genes, y la tremenda repercusión del modelo de estructura propuesto por James Watson y Francis Crick para dicha molécula (la famosa hélice doble) a mediados del siglo. Watson y Crick dieron a conocer su modelo sobre la base de pocos datos experimentales, pero la doble hélice recibió una inmediata y entusiasta aceptación, dada la elegante sencillez con la que explicaba las tres *capacidades fundamentales reconocidas en la información genética*. En primer lugar, su capacidad de duplicación, evidente en la propia transmisión de los caracteres hereditarios de generación en generación, que el modelo resolvía con una especial complementariedad química entre las dos cadenas de la hélice doble. En segundo lugar, su capacidad para dar forma (*informar*) al organismo. Sobre esta capacidad informativa ya se sabía que descansaba en la facultad de cada gen de determinar la constitución de una proteína distinta, siendo así que la estructura y funcionamiento peculiares de un organismo dependen del juego específico de estas versátiles herramientas, las proteínas, con que cuenta dicho organismo. El modelo de Watson y Crick permitía cualquier combinación de las cuatro unidades (o letras) químicas en el DNA. En moléculas de la longitud adecuada, esta falta de restricción permite un número casi infinito de combinaciones, de *significados*, de proteínas distintas. Finalmente, el modelo de Watson y Crick explicaba con facilidad la capacidad de variación, de mutación de los genes. Las mutaciones serían el resultado de alteraciones en la secuencia de letras, debidas, por ejemplo, a cambios

químicos espontáneos o inducidos por agentes ambientales, o a errores de la maquinaria de replicación del DNA.

El fenómeno de la mutación es de una enorme trascendencia biológica. La evolución de los seres vivos, su diversificación en el asombroso número de especies que hoy conocemos, se deben, en última instancia, a la capacidad de cambio del DNA. Sin ella, no es posible imaginar nuestra presencia en este planeta, ni siquiera la del primer ser vivo. La mutación es también la fuente última de la diversidad de caracteres observable en los individuos de una misma especie o población, incluida la humana, la razón de que nos distingamos en el color de la piel o de los ojos, y, en buena medida, en nuestras habilidades físicas y mentales. No tengo tiempo para extenderme en este tema, pero me gustaría resaltar que si hay algo que la Genética, y la Biología en general, nos enseña con claridad es la bondad selectiva de la variación, la superior capacidad adaptativa de las poblaciones diversas frente a las *razas puras*. En realidad, no existen tales razas en la naturaleza.

Con motivo de la publicación de la secuencia del genoma humano, los medios de comunicación general se han hecho eco de comentarios de algunos de los investigadores implicados en el proyecto que resaltaban el hecho de no haber encontrado diferencias en el genoma de individuos de diferentes etnias. En realidad, se trata de una simplificación. La verdad, más compleja pero en la misma dirección, es que lo que iguala a las diferentes etnias es el hecho de que en ninguna de ellas se encuentran dos individuos genéticamente idénticos. Somos todos iguales porque todos somos únicos. No hay un genoma humano, sino miles de millones de genomas humanos. Ni siquiera dos gemelos monocigóticos son tan genéticamente idénticos como puede parecer a simple vista. Cada uno de nosotros está formado por unos 10 trillones de células. Todas derivan de las divisiones sucesivas de una única célula original: el cigoto. El genoma presente en el cigoto se duplica en cada división celular, de manera que cada célula del adulto contiene una copia completa del mismo. Aparte de otras causas de mutación, sabemos que la propia maquinaria encargada de duplicar el genoma en cada división celular no está exenta de errores, y comete uno, más o menos al azar, cada 1000 millones de letras. Una precisión enorme, dirán ustedes, muy superior, desde luego, a la del más esmerado copista. Pero en un genoma de la magnitud del nuestro, ello significa la introducción de uno o unos pocos cambios en cada división celular, y nuestro desarrollo embrionario requiere ¡200 billones de divisiones celulares!. Así pues, ni siquiera dos células de un mismo individuo adulto contienen exactamente la misma secuencia genómica. Se producen, además, otros ciertos fenómenos que generan variación celular hereditaria, que los expertos denominan global-

mente *epigenéticos*, y que son responsables, por ejemplo, de que la famosa oveja Dolly fuera ya varios años vieja cuando nació. Si a ello unimos el efecto definitivo del medio ambiente, la educación, el entrenamiento, en muchas de nuestras habilidades físicas e intelectuales, salta a la vista que la ilusión de alguno de eternizarse en un clon de sí mismo, cuestiones éticas aparte, es una imposibilidad biológica. Larga vida, pues, a una sociedad humana diversa.

Pero como en la filosofía china dualista, la mutación ofrece otra cara, inevitable, más amarga. El cambio en la secuencia de letras de un gen puede ser causa de una alteración grave de su significado; es decir, de la estructura de la proteína determinada por dicho gen, o provocar una alteración en el momento o tejido preciso en que dicho gen debe expresarse, apareciendo así una enfermedad genética. Las enfermedades genéticas son causa importante de mortalidad y morbilidad en el hombre. Con ligeras oscilaciones entre distintas poblaciones, del orden del 3% de los recién nacidos sufren de algún trastorno genético más o menos grave, añadiéndose a ello que varias de las enfermedades más frecuentes de aparición normal en el adulto (cáncer, diabetes, enfermedades coronarias, ciertos desórdenes psiquiátricos, etc.) sabemos hoy que se deben en mayor o menor grado a la constitución genética de los individuos afectados, aunque en estas enfermedades incidan también agentes causales no genéticos, sino ambientales.

Volvamos al artículo de la revista *Nature*, en concreto ahora a sus autores. Se trata del trabajo conjunto de centenares de investigadores y técnicos de 20 grupos distintos de varios países, en lo que ha constituido uno de los mayores esfuerzos de cooperación entre científicos, si no el mayor, de nuestra historia. La ciencia se asocia con frecuencia a la competición, y así ocurre realmente a menudo. No puede ser de otra forma. La misma curiosidad y ¿por qué no? la misma ambición de reconocimiento, son humanos, mueve a todos los científicos. Cualquiera de ellos que está al tanto de su disciplina sabe de la siguiente altura a escalar, de la siguiente frontera a cruzar, del siguiente enigma importante a resolver. Aunque a todos resulta incómoda, ¿a qué investigador no se le oye decir alguna vez que le han *pisado* un artículo?; la competencia es una compañera inevitable, y deseable por su efecto estimulador, de la labor científica. No cabe más que lamentarse de que, en ocasiones, conduzca a algunos a una apresurada falta de rigor, o sencillamente al fraude. Pero también existe la cooperación entre científicos, el intercambio de ideas, de técnicas, de materiales, de propuestas de investigación conjunta, y la secuenciación del genoma humano ha ofrecido un ejemplo tan glorioso de ello que mueve incluso a la emoción.

El artículo menciona la puesta a disposición pública de la secuencia, con acceso, sin restricciones, a todos, sean del sector académico o comercial. No es un comentario banal, ni mucho menos, sino una justa y orgullosa reivindicación. Es bien conocido que cuando el consorcio internacional responsable del Proyecto Genoma Humano llevaba ya varios años de trabajo, debió enfrentarse a la competencia de la empresa privada americana Celera, la cual anunció su intención de alcanzar el mismo objetivo, por un procedimiento distinto, antes incluso que el consorcio internacional y con fines comerciales. Que tal empresa hubiera tenido éxito hubiera sido ciertamente injusto, por varios motivos. Está, por un lado, el hecho de que Celera podía aprovechar, sin restricciones, los progresos diarios del consorcio público, pero no a la inversa. Por otro, está la razonable declaración de las Naciones Unidas: «el genoma humano es, en sentido simbólico, patrimonio común de la Humanidad» (basta pensar en la pléyade de investigadores, genéticos, bioquímicos, microbiólogos, desde Mendel, sin cuyas contribuciones no hubiera existido la posibilidad de secuenciar el genoma humano). Es bien sabido que Celera y el consorcio público internacional llegaron a un acuerdo, propiciado por el propio Presidente de Estados Unidos, que la determinación de la secuencia fue anunciada en una rueda de prensa conjunta de Francis Collins, director del proyecto público, y Craig Venter, presidente de la empresa privada, y que el trabajo de Celera apareció simultáneamente al de *Nature* en la revista *Science*, su eterna competidora, por cierto, en cuanto a recoger las contribuciones científicas de mayor relieve en cualquier disciplina.

¿Y qué dice esa secuencia que ha servido de apropiado colofón a todo un siglo de oro de la Genética y la Biología molecular? ¿Qué deja translucir ese mensaje que influye tan poderosamente en nuestros cuerpos, nuestra mente y nuestra conducta individual y colectiva, ese texto que debe permitir nuevas visiones de nuestro propio origen e historia como especie o abrir nuevas vías para diagnosticar y combatir la enfermedad? Bueno, la secuencia es un espléndido galimatías. Ya hace tiempo que conocemos la lengua universal (todos los seres vivos usan la misma) de los genes, que contamos con la clave que utiliza la célula para traducir una sucesión de letras en un tramo de DNA en una proteína útil (la famosa clave genética que *nuestro* Severo Ochoa contribuyó a desentrañar). Aplicando esta clave y poderosos procedimientos informáticos, se descubre que no más del 2% de la secuencia parece tener sentido. El resto resulta ser un guirigay, denominado por algunos *DNA basura*, atestado además de numerosas repeticiones, monótonas multiplicaciones de secuencias muy pequeñas, copias deformadas e inactivas de genomas de virus, o copias de otros tipos de DNA parásito. Tal guirigay no se encuentra

sólo entre un gen y el siguiente, sino que incluso interrumpe continuamente a los tramos con sentido, a los propios genes.

En realidad, eso es algo que ya sabíamos hace años, cuando los genes, humanos o de otros organismos, se analizaban, trabajosamente, de uno en uno. Conocíamos, por ejemplo, el caso extremo del gen de la proteína llamada distrofina, que cuando está alterado por mutación provoca dos tipos de distrofia muscular, la de Duchenne y la de Becker. Su longitud total alcanza los dos millones y medio de letras (tanto o más grande que el genoma completo de algunas bacterias), pero la del tramo con sentido de este gen, múltiples veces interrumpido por larguísimos galimatías, no ocupa más que unos pocos miles de ellas. Yo mismo he comparado esta situación con la que se produciría si *El Quijote* estuviera escrito de forma que a la frase «En un lugar de la Mancha» siguiera un tropel de mil letras tomadas al azar de nuestro alfabeto antes de que apareciera la lógica continuación «de cuyo nombre no quiero acordarme», y así sucesivamente. Engorrosa situación para disfrutar de la lectura de tan excelsa novela, que la célula, sin embargo, leería casi de corrido. Para fabricar la proteína determinada por un gen, la célula hace primero una copia de todo el tramo correspondiente del DNA, y luego la trabaja con extraordinaria precisión. Cortando aquí y pegando allí, desecha los tramos sin sentido y une en el orden apropiado los que sí lo tienen. Conocemos muchos detalles de la maquinaria encargada de este asombroso proceso, pero carecemos de una respuesta adecuada a la pregunta obvia de por qué el genoma (el nuestro y el del resto de los organismos, salvo las bacterias) está organizado de una manera aparentemente tan absurda.

La comparación de nuestro genoma con el de otros organismos pone de manifiesto que les sobrepasamos largamente en la proporción de *DNA basura*. Por ello, a pesar de que nuestro genoma es muchísimo más largo que el mayor conocido, no parece que el número real de nuestros genes, unos 35.000, llegue siquiera a doblar al del humilde y famoso gusano *Caenorhabditis*, y resulte ser no mayor del 25% del de una mala hierba, la planta *Arabidopsis*. Así pues, salvo que nuestro genoma resulte ser especialmente opaco a los procedimientos informáticos de localización de genes, nos enfrentamos a un enigma paradójico. La suprema complejidad estructural y de funcionamiento del ser humano, piénsese en nuestro sistema nervioso o en nuestra exclusiva capacidad para el lenguaje, la abstracción o el razonamiento lógico, no tiene la esperada correspondencia en el número de nuestras unidades genéticas informativas.

Desde luego, la paradoja es sólo aparente y vislumbramos ya varias vías de solución.

La propia secuencia del genoma descubre que nuestros genes, si bien menos numerosos de lo esperado, poseen en bastantes casos una mayor complejidad interna. Ello significa, sin entrar en farragosos detalles, que muchas de nuestras herramientas celulares, de nuestras proteínas, son más versátiles en cuanto a las operaciones moleculares que pueden llevar a cabo y en cuanto al número de otras proteínas con las que pueden asociarse en procesos de cooperación funcional. Por otro lado, sabemos que la maquinaria que ya he citado, la que a modo de sastre corta y cose tramos de letras, puede operar sobre una misma pieza para hacer varios trajes distintos, cambiando cada vez los tramos que permanecen o no en el producto final. Quizás nuestra complejidad específica descanse, al menos en parte, en que nuestro sastre celular es tan imaginativo como el ahora retirado Yves Saint Laurent.

Finalmente, la propia relación entre cantidad (número de unidades informativas) y calidad (complejidad estructural y de funcionamiento) no se ajusta de forma sencilla a la realidad. El análisis comparativo del modo de actuar de ciertos genes entre distintas especies, en especial de algunos genes que juegan un papel crítico durante el desarrollo embrionario, ponen de manifiesto la importancia de fenómenos como la reiteración de una misma operación molecular y las interacciones con otras operaciones moleculares distintas. Algunas metáforas ayudarán a entender esto. La mayor complejidad de un potente ordenador, capaz por ejemplo de enfrentarse a un gran maestro del ajedrez, no deriva necesariamente de que la naturaleza o el número de sus distintos componentes sean diferentes de los de un pequeño ordenador personal, sino del número de veces que se reiteran los componentes y el número y variedad de las conexiones entre ellos. Tras la enorme complejidad (no hablemos de belleza) de la Alhambra granadina, de la Mezquita cordobesa, o de la portada de nuestra catedral, no se esconden materiales, bastante humildes por cierto, ni elementos arquitectónicos básicos muy distintos de los que podían encontrarse en modestos alojamientos o iglesias de la misma época, sino inesperadas reiteraciones y novedosas combinaciones entre los mismos.

Las operaciones subyacentes a estos fenómenos (el grado de imaginación de nuestro sastre o el juego de reiteraciones e interacciones de nuestras proteínas) no son describibles mediante la mera inspección de la secuencia de letras del DNA. Pero contar con esta secuencia es fundamental para diseñar nuevas herramientas (como los *chips* de DNA, sobre los que no puedo extenderme) y métodos analíticos (como los programas informáticos que realiza provechosas comparaciones entre los genomas y proteínas de distintas especies) que nos permitirán estudiar de forma global qué genes se expresan en ca-

da momento, cómo, y en qué tejidos (la nueva disciplina de la Transcriptómica) así como el correspondiente juego de proteínas que aparecen, y sus interacciones (Proteómica). Avanzaremos así, sin duda, en la comprensión de nuestro desarrollo y funcionamiento normales. Pero esperamos también que, aplicadas a situaciones patológicas, tales nuevas herramientas nos permitan progresar en la comprensión, el diagnóstico y la lucha contra las enfermedades genéticas. Queda, pues, mucho por hacer. Un siglo para alcanzar la secuencia. Parece que necesitaremos otro para entenderla por completo y aprovecharla. Hay que volver al laboratorio.

Muchas gracias.