

ADN, DISCO DURO DE LA VIDA

DNA, life's hard drive

Edwin Francisco Herrera Paz¹

¹Doctor en medicina y cirugía, magister en genética humana. Profesor de genética, fisiología e inmunología, Facultad de Medicina, Universidad Católica de Honduras, Campus San Pedro y San Pablo, San Pedro Sula, Honduras.

RESUMEN. A lo largo de los últimos dos siglos la medicina se vio nutrida con los descubrimientos bioquímicos que impulsaron el entendimiento de los mecanismos fisiopatológicos y facilitó el desarrollo de la terapéutica. En cambio, en el presente siglo entramos a la era de la genómica y del “*big data*”, por lo que el estudio de las funciones del ADN como dispositivo de almacenamiento de información es esencial para la comprensión de la nueva medicina genómica personalizada, de precisión. En la presente revisión, se analiza el ADN como un dispositivo informático con tres funciones: almacenamiento, expresión y transmisión de la información acumulada a lo largo de la filogenia en forma de secuencias de nucleótidos. Se describe cada una de estas funciones comparándolas con la información manejada por una computadora o una sociedad, y se brindan ejemplos de patologías que surgen ante el fallo de alguna de las funciones. La revisión bibliográfica es amplia e incluye los artículos más relevantes, tanto históricos como del estado del arte, correspondientes a cada tema.

Palabras clave: Ácidos nucleicos, Bioinformática, Enfermedades hereditarias, Genómica, Genética médica.

INTRODUCCIÓN

Los dos últimos siglos se caracterizaron por los avances en el estudio de la bioquímica, principalmente de las proteínas, pero también de otras moléculas de vida, lo que impulsó la medicina como nunca en la historia.¹ Entre los grandes avances, de gran utilidad a la medicina clínica y la terapéutica médica, se incluyen la descripción de la estructura y función de receptores moleculares, tanto nucleares como de membrana,^{2,3} que fue seguido por el diseño de innumerables fármacos estimuladores o bloqueadores;⁴ las bombas iónicas dependientes de energía y los canales iónicos, incluyendo los regulados por voltaje, que contribuyeron a la comprensión de procesos como la contracción muscular y la conducción nerviosa,⁵ y el estudio de las enzimas que reveló un micro-universo de ciclos energéticos, de degradación y reciclaje, y de biosíntesis de sustancias.^{6,7} Sin embargo, tanto en medicina como en el resto de las ciencias biomédicas, el desarrollo ha sido en apariencia contrario al que dicta el sentido común. Por ejemplo, en las ingenierías se estudian los fenómenos individuales básicos. Para construir una casa o un puente debemos comenzar por los planos; para la invención de las modernas computadoras fue necesario el desarrollo de un nuevo tipo de álgebra en el siglo XIX y de los circuitos integrados décadas después.⁸ La tecnología de los sistemas artificiales requiere del dominio de fenómenos básicos que pueden ser utilizados en la generalidad, y se construye de abajo hacia arriba, desde los esquemas, hasta los circuitos o los materiales, terminando en las grandes estructuras.

En las ciencias biomédicas el sentido cambia, y esto no debe sorprendernos, porque al investigador médico le toca desentrañar lo que la naturaleza ya diseñó, utilizó y probó. Y por supuesto, estudiamos primero lo macro, lo que es evidente, y no profundizamos en los fenómenos básicos sino hasta que se desarrolla una nueva tecnología que nos permite analizar y estudiar rangos espaciales y temporales cada vez más pequeños. Como ejemplo, descubrimos primero los fenómenos básicos de la circulación sanguínea,⁹ pero no fue hasta el invento del microscopio siglos después que se pudo constatar que el organismo humano incluido el sistema cardiovascular está compuesto de células,¹⁰ de cuya integridad y buen funcionamiento depende en gran medida todo el sistema. Y no fue sino hasta el desarrollo de técnicas moleculares que nos dimos cuenta de que las proteínas y su adecuado funcionamiento están en la raíz del asunto. Finalmente, fue hasta muy recientemente que se descubrió que más allá de las proteínas y en la base de toda estructura viviente se encuentra la información, que en lugar de discos duros y circuitos integrados utiliza el ADN como sustrato.¹¹ Todo dispositivo informático tiene dos caras, la información propiamente dicha, y un sustrato para almacenarla y desde donde pueda ser leída. Pero la información como tal es independiente del sustrato. Esta noción adquirió popularidad con la serie de películas *The Matrix*, basadas en la idea de que la realidad puede ser completamente simulada.¹² Por ejemplo, el texto que usted está leyendo puede ser almacenado en una computadora en forma de ceros y unos, o puede estar impreso en papel. En ambos casos la información útil es la misma. En otro ejemplo una secuencia de un gen puede encontrarse escrita con nucleótidos en el ADN de cada célula de un organismo, pero también puede estar escrita con letras que representan esos nucleótidos en la pantalla de una computadora. Esa información podría servir en un futuro para sintetizar la proteína co-

Recibido para publicación el 10/2017, aceptado el 11/2017

Dirección para correspondencia: Dr. Edwin Francisco Herrera Paz.
19 calle, 3era y 4ta avenida N.O. #191, Col. Moderna, San Pedro Sula, Honduras.

Correo electrónico: eherrera@unicah.edu.

dificada por el gen, o incluso organismos completos. Entonces la información habrá viajado hacia el futuro por otros medios sin necesidad de su sustrato original, que es el ADN. La información genética entonces puede viajar en el tiempo para incluso resucitar especies extintas.¹³ Todo es cuestión de desarrollar la tecnología adecuada. En esta revisión se estudia el ADN, no solo como el compuesto químico que es, sino como el sustrato regular que mantiene la información de la vida, poniendo énfasis en los organismos humanos. Además, en algunas secciones se compara la información contenida en el ADN con otros tipos de información, como la cultural y la de los sistemas computacionales. Y como es propio de las ciencias computacionales más que de la medicina y la biología, se examina, más allá de la información misma, la manera en la que esta se maneja.

Función de almacenamiento de la información biológica

Una sociedad primitiva carece de una forma adecuada de almacenar información excepto en los cerebros mismos, y esta se transmite por tradición oral. Una vez se inventó la escritura, la información cultural y tecnológica podía ser grabada y luego leída por los iguales o por las siguientes generaciones. Almacenar la información facilitó los procesos de aprendizaje, impulsó la cohesión de grupo y catapultó la inventiva y por ende la evolución de las sociedades humanas.¹⁴ El almacenamiento de la información cultural evolucionó en la elaboración de textos, que podían ser guardados en grandes centrales llamadas bibliotecas. En la actualidad, la central se ha hecho global, se encuentra en el mundo virtual de la red y contiene todo el conocimiento universal. De igual manera, los organismos vivos poseen una central de información almacenada en el núcleo celular (en el caso de los eucariotas). Así como la información universal se encuentra disponible para cada ser humano que cuente con una conexión a internet, la información para la construcción de un ser humano se encuentra disponible para cada célula de nuestro organismo. El núcleo celular es una biblioteca con una sola obra llamada genoma. La estructura de doble hélice del ADN fue dilucidada el siglo pasado por James Watson y Francis Crick.¹⁵ Los genomas incluido el humano están compuestos por estas moléculas, que son ristas dobles de desoxirribonucleótidos en fila india unidos entre sí por enlaces covalentes llamados fosfodiéster. En realidad, el genoma no es un solo libro. Como una enciclopedia, se encuentra fragmentado en 23 libros llamados cromosomas. Esta fragmentación tiene motivos prácticos, y similares a aquellos por los que una enciclopedia se divide en tomos: son más fáciles de manejar. Además, el ADN desnudo, sin las estructuras proteicas presentes en los cromosomas, carece de funciones mecánicas que son vitales para la célula como la recombinación genética en la meiosis 1 y la cariocinesis.¹⁶ Cada célula humana contiene dos genomas completos, cada uno con 3 mil millones de nucleótidos (nt) en cada cadena de ADN (como el ADN es de doble cadena también de les denomina pares de bases, o pb, en el lenguaje de la genómica, término que se utilizaremos con frecuencia), que son las letras de la enciclopedia. Estas dos enciclopedias contienen la misma información en la mayor parte de su extensión, pero

una es heredada de la madre y la otra del padre, y, por ende, una pequeña parte es diferente entre ambas. Se necesitan los dos genomas para el funcionamiento adecuado del organismo.

Por otro lado, en las ciencias físicas la información es representada mediante unidades propias: El bit, contracción de "dígito binario". El término, acuñado por John Tukey,¹⁷ hace referencia a una unidad o dispositivo capaz de almacenar uno de dos elementos o estados entre los que se puede escoger. Con un bit podríamos representar un 0 o un 1, positivo o negativo, blanco o negro, arriba o abajo etc., según la manera en la que codifiquemos la información. Un bit es la unidad mínima y en una computadora se puede representar mediante los dos posibles estados de un circuito, encendido o apagado, o por las dos posibles orientaciones de una partícula magnética. La información necesita de dos componentes: uno o varios lenguajes (analógico o digital), y un código (digital) que es una serie de reglas de asignación de elementos (a pesar de la idea que se tiene de los lenguajes como códigos, no lo son por sí mismos).¹⁸ Para poder asignar más de dos elementos a un estado no nos sirve un bit. Necesitamos unidades mayores que podemos obtener mediante filas. Por ejemplo, si unimos dos bits, entonces podremos tener cuatro estados a los cuales podríamos asignar cuatro elementos diferentes. Si el estado de un cable encendido/apagado lo representamos como 0/1, entonces la asignación más sencilla es la de los números binarios del 0 al 3: 00, 01, 10 y 11. Pero también a cada combinación le podríamos asignar un color, o un sabor etc., y entonces estaríamos construyendo el código. Por ejemplo, podríamos hacer la siguiente asignación: 00=azul; 01=rojo; 10=verde, y 11=amarillo. Las computadoras entienden el lenguaje de ceros y unos (cables encendidos/apagados), es decir, binario, mientras que los humanos solo entendemos el lenguaje analógico de letras, palabras y números, por lo que un código que facilite la comunicación se hace necesario.

A diferencia de una palabra en los lenguajes cotidianos cuya longitud es variable, una palabra en lenguaje binario de las computadoras actualmente es fija, midiendo 8 bits y tomando el nombre de byte. No siempre un byte ha tenido ocho bits, e históricamente este número ha cambiado dependiendo del número de bits necesarios para codificar un carácter.¹⁹ En el espacio de un byte de información se puede almacenar una combinación de ceros y unos específica de entre 256 posibles. Entonces, para que la computadora pueda comunicarse con nosotros debe existir un código, un conjunto de reglas de equivalencia entre cada uno de 256 caracteres cotidianos, incluyendo el espacio vacío y algunos procesos simples, y las 256 combinaciones de ceros y unos. Ese código se denomina ASCII (*American Standard Code for Information Exchange*) extendido.²⁰ Por ejemplo, el número 0 es representado por la combinación 00110000, la letra "a" por 01100001 y el espacio vacío por 00000000. Un disco duro de computadora contiene líneas secuenciales de elementos informativos (bits) en cada uno de los cuales podemos almacenar una de dos opciones. De la misma forma, una cadena de ADN es una línea secuencial de elementos informativos construida a partir de desoxirribonucleótidos que son de cuatro tipos: Guanina (G), citocina (C), Adenina (A) y timina (T) y, por lo tanto, en cada posición del

ADN no tenemos solo dos elementos para escoger sino cuatro, lo que representa dos bits de información. Así, necesitamos solo cuatro nucleótidos en cadena para representar uno de los $(4^4) = 256$ caracteres posibles de un byte. Podríamos entonces construir un código arbitrario en donde a cada combinación específica de cuatro nucleótidos se le pudiera asignar uno de 256 caracteres o comandos de computadora.

Si contáramos con un dispositivo capaz de editar (escribir) el ADN y otro capaz de secuenciar (leer) rápidamente, podríamos utilizar estas moléculas para grabar y leer imágenes, sonidos y texto en tiempo real substituyendo los medios de almacenamiento actuales, pero aún estamos lejos de eso. Nos toma aun varios días leer los 3 mil millones de pb de un ADN humano con la tecnología de secuenciación de siguiente generación,²¹ y apenas comenzamos a agilizar las formas de edición con los sistemas CRISPR-CAS que prometen ser invaluable para la medicina.²² No obstante, el ADN representa una manera barata y eficiente de almacenar información y tan solo un gramo puede contener hasta 700 Terabytes (TB), por lo que diferentes equipos trabajan para encontrar maneras más eficientes de codificación.²³ Un problema serio que enfrentan los estudios genómicos y otros campos de la ciencia y el conocimiento en la actualidad es el escaso espacio de almacenamiento informático para la creciente cantidad de datos que se generan cada año, a lo que se le ha dado el nombre inglés de “*big data*”. Cada año la ciencia genera una ingente cantidad de datos. Solo en las ramas biomédicas, la velocidad de generación lograda gracias a los avances actuales en la secuenciación genómica de extensas muestras poblacionales, los estudios de asociación genómica completa (*Genome Wide Associations Studies*, o *GWAS*) y otras tecnologías “ómicas”, sobrepasan la capacidad de análisis, por lo que es preciso el almacenamiento para su procesamiento futuro. Se calcula que el contenido de información mundial en todas las ramas del conocimiento llegará a los 163 Zettabytes (1ZB=1 billón de MB) para el 2025,²⁴ y el ADN como dispositivo de almacenamiento es prometedor.²⁵ El primer trabajo que codificó en secuencias de ADN una obra artística, que representa los genitales femeninos—llamada *Microvenus*—fue realizado por Joe Davis en los 90s.²⁶ Más recientemente, Church et al. almacenaron un libro de 53,426 palabras y 11 imágenes en oligonucleótidos de 96 nt, que luego leyeron exitosamente con un error de solo 10 bits entre 5.27 millones,²⁷ y Shipman et al. escribieron con éxito en el genoma de *E. Coli* las secuencias en las que se codificó una película corta e imágenes, utilizando el sistema inmune adaptativo CRISPR CAS de las mismas bacterias.²⁸ Por su parte, Goldman et al. almacenaron cinco archivos que contenían los 154 sonetos de Shakespeare, un artículo científico en PDF, una fotografía a colores de Instituto Europeo de Bioinformática en JPEG, el discurso de 26 segundos de Martin Luther King “Tengo un sueño”, y un código Hoffman, en oligonucleótidos de 117 nt cada uno. Los autores refieren haber sintetizado el ADN y luego recuperado la información en un 100%.²⁹ Calcular la capacidad de almacenamiento de un genoma del tamaño del humano es relativamente sencillo. Carl Sagan en su libro “Los dragones del Edén” estimó, erradamente, en 20 mil millones de bits, o aproximadamente 2.5 Gi-

gabytes (GB), el contenido informativo de un solo cromosoma.³⁰ Por tanto, si consideramos los 23 cromosomas entonces según sus cálculos el genoma humano tendría una capacidad de alrededor de 57.5 GB, suficiente para almacenar una biblioteca completa. La realidad es diferente. Si el genoma humano tiene 3 mil millones de pb y un byte puede ser almacenado en cuatro pb consecutivas, dividiendo obtenemos una capacidad de 750 millones de bytes, o lo que es lo mismo, 750 Megabytes (MB). ¡Una cantidad pequeña para los estándares comerciales actuales! En contraposición, un teléfono inteligente moderno tiene muchas veces más capacidad de almacenamiento. El error de Sagan derivó de sobreestimar el número de nucleótidos en un cromosoma, que para la época se calculaba en 5 mil millones, y, además, en sus cálculos utilizó bytes de seis bits. Resulta entonces sorprendente que este espacio de almacenamiento tan reducido sea suficiente para contener las instrucciones para la formación de un ser humano. Esas instrucciones escritas en forma de secuencias de nucleótidos pueden codificar, incluso, los diferentes caminos de desarrollo que puede tomar el organismo dependiendo de las condiciones ambientales actuales e intergeneracionales, la llamada “plasticidad fenotípica”.^{31,32} Se hace evidente que en los organismos biológicos la información se encuentra compactada de una manera mucho más eficiente que en nuestros dispositivos electrónicos, porque la naturaleza se ha encargado de mejorarla y comprimirla en su filogenia, a lo largo de muchos millones de años de prueba y error mediante el mecanismo evolutivo.

Función de expresión de la información

El genoma humano al igual que en el resto de los organismos eucariotas posee diferentes tipos de secuencias que, si bien son indistinguibles desde el punto de vista bioquímico, tienen enormes diferencias funcionales e informáticas. La división más básica es entre el ADN codificante y el no codificante. El primero, está compuesto por las regiones exónicas de genes que codifican los diferentes dominios de las proteínas, y por genes de ARN ribosomal (ARNr) y de transferencia (ARNt). El segundo, está constituido por las regiones no traducidas en 5' y 3',³³ los intrones,³⁴ la mayor parte de los micro y minisatélites,^{35,36} pseudogenes procesados y no procesados,^{37,38} ADN centromérico y telomérico, y otras secuencias con función estructural y reguladora.³⁹

En los primeros tiempos del desarrollo de la genómica, se pensaba que la gran mayoría de las secuencias no codificantes no tenían valor alguno denominándoseles “ADN basura”,⁴⁰ pero desarrollos posteriores adjudicaron funciones a muchas de ellas. Más aun, se observó evidencia de selección natural de ambos tipos, direccional y purificadora, actuando sobre el ADN no codificante lo que es propio de secuencias con alguna función importante.⁴¹ En cada uno de los dos tipos de secuencias, la información se expresa de manera diferente.

Expresión de la información en el ADN codificante

El ADN de las regiones codificantes, es decir los genes, funciona como las palabras en un libro que al ser leídas determinan las secuencias de aminoácidos en una proteína. La evi-

dencia apunta a que hay alrededor de 19,000 genes funcionales que codifican para proteínas en el genoma humano, un número no mucho mayor que el de organismos menos complejos como el *Caenorabditis elegans*,⁴² un nemátodo ampliamente utilizado como modelo para el estudio del desarrollo embrionario. El primer paso en el proceso de síntesis de una proteína es la transcripción del gen, es decir, la copia de la secuencia del gen en moléculas de ARN mensajero (ARNm). Cada molécula de ARN así formada se denomina transcrito primario y necesita ser procesada antes de ser traducida. En el procesamiento de las moléculas se eliminan las regiones intrónicas y se ensamblan de nuevo los exones, lo que se denomina corte y empalme. En algunos genes, los exones no se ensamblan en las secuencias originales sino en formas alternativas que al momento de la traducción originarán nuevas proteínas: el corte y empalme alternativo. Como consecuencia, existen genes que codifican más de una proteína, lo que aumenta la diversidad del proteoma humano sin aumentar el contenido de ADN, un método eficiente de compresión de la información.⁴³⁻⁴⁶ Además del corte y empalme, a las moléculas de ARNm se les adiciona un nucleótido metilado en su extremo 5', y una cola de poli-Adeninas en 3' como señalización para iniciar la traducción.^{47,48} Una vez maduros los ARNm salen del núcleo celular para ser traducidos en el citoplasma. El segundo paso de la síntesis proteica es la traducción del ARNm, que implica la lectura de unidades compuestas cada una por tres nucleótidos consecutivos, llamadas codones. De la misma forma que el código ASCII se desarrolló temprano en la era de las modernas computadoras personales y promete mantenerse por mucho tiempo, la manera en que los codones codifican sus aminoácidos correspondientes debió haber evolucionado temprano en la vida en la Tierra y se mantiene hasta nuestros días: el llamado código genético estándar, compartido por todas las formas de vida, con pequeñas variantes en algunas especies que son la excepción.^{49,50} Así, las características universales de la vida en nuestro planeta incluyen los mismos cinco tipos de nucleótidos para formar los largos polímeros de ADN o ARN, los 20 tipos de aminoácidos para formar proteínas, y la codificación para traducir codones en aminoácidos. La maquinaria bioquímica de la traducción está compuesta por los ribosomas y un conjunto complejo de enzimas. Las subunidades ribosomales mayor y menor se ensamblan sobre el ARNm dejando el codón de iniciación, que siempre es un AUG, listo para ser traducido y por ende toda proteína se inicia con el aminoácido metionina. A partir de ahí, el complejo traduccional lee los codones de forma continua y coloca los aminoácidos a una velocidad variable que depende de diferentes factores,⁵¹ hasta que encuentra un codón de parada. Los codones de parada son tres y no codifican para ningún aminoácido, por lo tanto, en ese momento el ribosoma se desensambla y la cadena polipeptídica queda libre.

¿En qué parte de este proceso se encuentra el corazón del código genético? El código es un diccionario o traductor. Los aminoácidos no llegan solos al sitio donde se está llevando a cabo la traducción, sino cargados en el extremo 3' de moléculas de ARNt. Cada una de estas, identifica un codón específico mediante una secuencia complementaria llamada anticodón, y

deja el aminoácido correspondiente para ser incorporado a la cadena polipeptídica naciente. Así, el corazón del código genético son las enzimas que, previo a la traducción, realizan la carga de cada uno de los aminoácidos en alguno de los ARNt cognados que le corresponden, y no en otros, haciéndolo con extrema precisión. En total son 20 enzimas, una por cada tipo de aminoácido y se denominan, en conjunto, aminoacil-ARNt sintetasas.^{52,53} Una vez terminada la síntesis de la proteína esta debe ser plegada hasta adquirir su estructura terciaria, es decir, su forma o proyección en el espacio, que es en última instancia la que le confiere la mayor parte de su función. El plegamiento es en su mayoría un proceso espontáneo, aunque ayudado por proteínas llamadas chaperonas.⁵⁴ La proteína se pliega porque en su forma lineal es inestable. Los aminoácidos hidrofóbicos buscan esconderse del medio hídrico circundante y los hidrofílicos salen a flote. Luego, los de carga positiva formarán puentes de hidrógeno con los de carga negativa y en un tiempo variable dependiendo del tamaño de la molécula y otros factores, el polipéptido habrá adquirido su forma final o estado nativo.⁵⁵ El plegamiento cesa cuando la molécula posee su mínima energía interna, que es su estado más estable. Anomalías en estos mecanismos de plegamiento dan lugar a patologías como el Alzheimer.⁵⁶⁻⁵⁸ En algunos casos, se hacen necesarias modificaciones adicionales como la formación de homo y heterodímeros, la N-glicosilación y la adición de radicales. Esto se realiza en el retículo endoplásmico y el aparato de Golgi, y requiere de información adicional cuya descripción escapa al alcance de esta revisión. El trabajo de una o de un conjunto coordinado de proteínas dará lugar a los fenotipos, y es sobre estos que actuará la selección natural, permitiendo la supervivencia de los individuos que posean aquellas moléculas que realicen mejor su función en beneficio del organismo. Entonces, así como las imágenes o texto en la pantalla de la computadora son la interfaz con el usuario de la información almacenada en el disco duro, los fenotipos son la interfaz entre la información del ADN y el ambiente. De esta forma, indirectamente, en el devenir de las generaciones se seleccionan las secuencias mejor adaptadas, que den lugar a mejores fenotipos, mientras las que dan lugar a fenotipos deletéreos originarán enfermedad o muerte, disminuyendo la supervivencia o el éxito reproductivo del que las porta.⁵⁹ Igual que el ingeniero en software escoge los programas que efectuarán con más eficiencia un trabajo de cómputo (menos tiempo, menos líneas de código y menos memoria), la selección natural escoge las secuencias de ADN que mejor realicen su trabajo, es decir, se adapten al ambiente, por intermedio de los fenotipos. Sin embargo, hay que aclarar que la selección natural no es la única fuerza actuando sobre los fenotipos. Así como la tecnología evoluciona en una sociedad por la necesidad de herramientas, pero también en parte por una serie de eventos fortuitos, la evolución de las secuencias de ADN también responde a factores aleatorios, como las mezclas poblacionales y los cuellos de botella propios de la historia demográfica de las poblaciones, que puede ser reconstruida en estudios biodemográficos utilizando técnicas de genética de poblaciones aplicadas sobre marcadores biológicos, o incluso culturales tales como los apellidos.⁶⁰⁻⁶³ Entre la lectura de la

secuencia de los exones en el gen hasta la adquisición de la forma final de la proteína, se pierde información. Es decir, si conocemos la secuencia exacta del gen podemos inferir la forma final más probable de la proteína, pero no al contrario. Para el caso, un oligopéptido de tres aminoácidos puede tener una entre $20^3 = 8,000$ configuraciones diferentes. Por otro lado, la cadena de nueve nucleótidos, es decir tres codones, que codifica estos tres aminoácidos puede tener $61^3 = 226,981$ configuraciones diferentes, donde 61 es el número de codones que codifican aminoácidos. Esto representa una reducción de $226,981 / 8,000 \approx 28$ veces el contenido informativo en una cadena de tres aminoácidos. Esto sucede porque en el código genético un aminoácido es codificado por un codón (tres nucleótidos), pero existen 61 codones codificantes y solo 20 aminoácidos. Entonces, muchos de los codones son sinónimos. A pesar de que esto puede parecer a simple vista un desperdicio de información, la redundancia es útil. La redundancia, presente al menos en dos niveles, le brinda robustez al código genético. En un primer nivel, cuando se introducen errores en las copias de ADN (mutaciones) los codones sinónimos reducen la probabilidad de que haya un cambio de aminoácido. Así, si un codón CCG sufre una mutación y el último nucleótido es cambiado a CCA, el aminoácido codificado siempre será prolina y la secuencia de la proteína no se verá afectada. Por lo tanto, esta robustez está determinada por la sinonimia de los codones que comparten el primer y segundo nucleótidos, pero no el tercero.

En el segundo nivel, si una mutación origina un cambio de aminoácido, pero el nuevo tiene propiedades fisicoquímicas similares al sustituido (por ejemplo, que ambos sean hidrófobos), entonces la secuencia de la proteína sí cambia, pero es probable que el plegamiento siga siendo igual y por ende la forma y la función se conserven. Para el caso, si un codón CAU muta a CGU, se cambiará el aminoácido histidina por arginina, ambos hidrofílicos básicos. Esto sucede porque los codones de aminoácidos con las mismas propiedades suelen compartir el segundo nucleótido, y en algunos casos el primero, como en el ejemplo expuesto. Así, en el código genético se intercambia contenido informativo a favor de robustez, lo cual conduce a la conservación de la función de las proteínas a pesar de las mutaciones (para una discusión sobre la evolución y ordenamiento del código genético y su redundancia ver referencia 64). Esto, obedece al hecho de que en los organismos vivos debe existir un balance entre la generación de variabilidad (mutaciones) necesaria para la adaptación al cambio, y la resistencia a ese cambio (robustez), que brinda estabilidad.⁶⁵

Expresión de la información en el ADN no codificante

La segunda forma de expresión es menos evidente y se encuentra codificada, paradójicamente, en el ADN no codificante. Tanto las moléculas de ADN como de ARN poseen, además de su secuencia y características lineales básicas, una estructura terciaria dependiente de la secuencia. Así como la secuencia de una proteína determina directamente su forma y por ende su función gracias a las fuerzas que actúan sobre aminoácidos individuales, los ácidos nucleicos realizan diversas funciones que dependen de la proyección espacial de las cadenas. Los

ARN forman estructuras bidimensionales que dependen de regiones complementarias que se asocian mediante puentes de hidrógeno formando dobles cadenas y estructuras parecidas a ganchos de pelo. Posteriormente las fuerzas internas plegarán la molécula en tres dimensiones. Este es el caso de los ARNt y ARNr.^{66,67} Una secuencia de ADN particular tiene una configuración espacial propia debido a las fuerzas que actúan sobre los diferentes nucleótidos. Esta forma tridimensional le permite formar patrones a los que se unen proteínas de unión al ADN, y la interacción ADN-proteína tiene funciones de diferente índole de las cuales sobresalen las estructurales y de regulación. El reconocimiento de estos patrones es objeto de intensa investigación por métodos de biología molecular, pero también por métodos informáticos como el de redes neuronales profundas.⁶⁸ Más aun, ciertas secuencias le confieren a la molécula de ADN la capacidad de plegarse en diversas formas secundarias diferentes a la estructura canónica simple de doble hélice dextrógira llamada B-ADN propuesta por Watson y Crick. Estas formas han sido llamadas "motivos G4", y no se encuentran distribuidas aleatoriamente en el genoma, sino de forma preferencial en regiones promotoras de genes y otras secuencias importantes como los telómeros.⁶⁹ Los patrones o motivos formados por algunas secuencias podrían incluso estar involucradas en la programación epigenética de la célula,⁷⁰ que veremos más adelante. Ejemplos de secuencias con función estructural son las destinadas al empaquetamiento del ADN en los octámeros de histonas, y las llamadas "regiones de unión al andamio", a las que se unen proteínas involucradas en la estructura y función de los cromosomas.⁷¹ El ADN del centrómero está formado por una secuencia repetida que se une con proteínas centroméricas para formar el cinetocoro, estructura a la que se adhiere el huso mitótico, esencial en la cariocinesis;⁷² y el ADN telomérico se une a un complejo proteico que forma el telosoma.⁷³ Cuando hablamos de "secuencias de regulación" nos referimos a una de las funciones más importantes en los organismos eucariotas, especialmente los metazoarios. En esencia, un organismo complejo debe contar con extensos mecanismos de regulación para la expresión de genes, es decir, para la producción de proteínas, tanto en espacio como en tiempo. En espacio, porque cada linaje celular produce las proteínas que necesita, y no otras; y en tiempo, porque las proteínas producidas por la célula varían dependiendo de factores como el instante del ciclo celular en que se encuentra, o respondiendo a las necesidades del organismo. Para que esta regulación se lleve a cabo se necesita una serie de proteínas de regulación, llamadas factores en trans,⁷⁴ y los sitios del ADN donde estas se unirán para cumplir su función de activar o desactivar, llamados factores en cis.⁷⁵ La interacción de estos factores apaga y enciende genes, además de regular la cantidad de proteína producida. Más aun, se ha demostrado que algunas secuencias poseen dos funciones, los llamados exones *enhancer* (mejoradores), que funcionan codificando proteínas en algunos tejidos, y regulando la expresión de genes vecinos en otros.⁷⁶ La activación y desactivación de genes en la célula por este método funciona como un complejo programa de computadora. Una vez que se activa un programa, por ejemplo, cuando se activa un gen en una etapa específica

del desarrollo embrionario, o por una señalización externa, este puede desencadenar cascadas de activación génica que ponen en marcha el resto del programa.

Un ejemplo interesante de este tipo de cascadas de activación es la diferenciación del embrión masculino. Pensemos en cada célula de los múltiples órganos involucrados en la virilización, como una computadora individual que correrá una subrutina específica que forma parte de un gran programa, que es activado por una sola señal. Cada computadora (célula) cumplirá una función específica en el proceso, pero para que resulte un individuo del sexo masculino, todas deben actuar coordinadamente. Así, el embrión masculino cuenta con un interruptor que pone en marcha el programa, el gen SRY, que se activa a la sexta semana de gestación, regulando un segundo gen corriente debajo de la cascada llamado SOX9, que a su vez regulará la expresión de una gran cantidad de genes lo que conducirá a la diferenciación de la gónada en los componentes del testículo.⁷⁷ Uno de estos componentes, la célula de Leydig, es programada para secretar testosterona, que a su vez activará otros programas en las células blanco, lo que llevará a la formación de los genitales y otras características de la masculinización.⁷⁸ Otro ejemplo de genes que actúan como subrutinas de programas mayores son los pleiotrópicos, que se expresan en diferentes tejidos pudiendo cumplir funciones diferentes en cada uno de ellos. Así, cuando esta subrutina (gen pleiotrópico) presenta anomalías de sintaxis (mutaciones deletéreas) sus consecuencias se observan en el funcionamiento o la estructura de los múltiples órganos y tejidos donde el gen se expresa.^{79,80} Pero es el ADN con función reguladora el que realiza ese "llamado de subrutinas". No es extraño entonces que un genio de la computación como Alan Turing, ya en 1952, en un único artículo sobre el tema, intentara describir la morfogénesis en base a sustancias que difunden formando gradientes y reaccionan, llamados morfógenos, que crean inestabilidades que alejan al embrión en desarrollo de su simetría inicial dándole su forma.⁸¹ Los algoritmos de difusión-reacción de Turing constituyen hoy en día uno de los pilares fundamentales para el estudio de los procesos responsables de la embriogénesis y de la formación de patrones en los organismos vivos.⁸²

Niveles de complejidad biológica y expresión multinivel

Los metazoarios complejos como los humanos estamos contruidos en niveles de complejidad biológica. Cuando tenemos un grupo de elementos especializados trabajando al unísono, entonces el conjunto forma un elemento de nuevo nivel superior, con propiedades emergentes que no pueden ser explicadas enteramente mediante una simple descripción de los elementos constitutivos básicos. A su vez, los elementos de este segundo nivel se especializarán y trabajarán en conjunto para formar un tercer nivel, y así sucesivamente. Así, las células son ya organismos de alta complejidad puesto que están formados por organelas especializadas, que a su vez están formadas por agregaciones moleculares complejas. Las células se unen para formar tejidos, estos forman órganos que a su vez forman sistemas, que a su vez forman seres humanos,

que forman comunidades complejas (una descripción de las características de los sistemas biológicos complejos y su evolución se da en la referencia 14). Los procariotas son organismos relativamente simples cuyo ADN, empaquetado en uno o dos cromosomas circulares o en plásmidos, es en su mayoría codificante conteniendo pocas secuencias no codificantes, incluyendo pseudogenes.⁸³ Por otra parte, los genomas de los eucariotas multinivel no solo controlan la producción de sustancias necesarias para la supervivencia de la célula mediante los llamados "genes de mantenimiento" (*housekeeping*),⁸⁴ sino que deben darle a esta su naturaleza particular, con el objeto de que cumpla su función en el gran contexto del organismo. Una célula adulta especializada se caracteriza por expresar un alto número de receptores moleculares que le permiten censar su medio y "escuchar" los mensajes de otros sitios del organismo y del medioambiente.⁸⁵ Los receptores son activados por ligandos producidos por otras células. La unión ligando-receptor desencadena cascadas de señalización que terminan en la activación de factores en cis, como factores de transcripción, que se unen al ADN regulando la expresión de genes.⁸⁶ Resulta natural entonces que una gran proporción del genoma se encuentre destinado a la regulación multinivel, al sistema de mensajería que hace que la célula cumpla sus funciones en la comunidad celular que es el organismo y que diferentes órganos y sistemas se comuniquen entre sí. Existe clara evidencia de que el grado de complejidad de los organismos biológicos se correlaciona con una mayor cantidad de ADN destinado a la regulación.³⁹ Se necesita más ADN regulador para una mayor diversidad en las funciones y morfología celular, y para que esta diversidad se acople formando un organismo. En última instancia, el tipo y número de receptores moleculares expresados por una célula, así como la respuesta fisiológica específica ante su estimulación, dependen de su linaje. Un linfocito tendrá receptores de citoquinas cuya estimulación causará la producción de otras citoquinas; una neurona tendrá receptores de neurotransmisores, y una célula de Sertoli tendrá receptores de andrógenos y de la hormona folículo estimulante cuyas acciones sinérgicas resultarán en la producción de factores de maduración espermática,⁸⁷ para poner algunos ejemplos. Comparemos por un momento al organismo multicelular con una sociedad moderna llena de computadoras, donde cada uno tiene la suya en casa. Cada computadora viene equipada con todos los programas básicos necesarios, y además tiene acceso al conocimiento universal a través de internet. A pesar de ello, ningún usuario utilizará todos los programas, ni estudiará toda la información universal. Una persona ejecutará aquellas aplicaciones que le sirvan en su carrera o en su diario vivir, e ignorará el resto. Un estudiante de medicina estudiará sus materias, que serán diferentes a las de un estudiante de ingeniería. Además, buscará y utilizará la información cuando la necesite. Es imposible para un ser humano acceder a todo el conocimiento universal que le brinda su computadora todo el tiempo, además de ser una pérdida de tiempo y recursos. En las sociedades los individuos se especializan y utilizan solo la información que necesitan. De la misma forma, un tipo dado de célula no utilizará toda la información genética almacenada en su ADN. En su lugar, expresará solo los

genes que necesita para su funcionamiento mientras los demás se mantienen apagados. Este tipo de programación, llamada epigenética, se da en cada paso sucesivo de la diferenciación y especialización celular.⁸⁸ La programación epigenética ocurre mediante complejos mecanismos de regulación de los cuales el principal es la metilación de genes, pero también incluye la acetilación y el remodelamiento de la cromatina. Un gen cuyo promotor ha sido metilado en regiones especiales denominadas islas de CpG usualmente deja de expresarse. Una vez que un gen se metila, la célula hereda la metilación a sus descendientes de forma irreversible por intermedio de enzimas ADN metil transferasas (ADMT). Existen otras formas de regulación por metilación, por ejemplo, en regiones pobres en islas CpG, pero son complejas y apenas comienzan a comprenderse.⁸⁹ Durante las primeras etapas del desarrollo del embrión las células son pluripotenciales,⁹⁰ definidas como aquellas capaces de originar las tres capas embrionarias. Podemos comparar estas células con niños recién nacidos que tienen la potencialidad de cumplir con cualquier actividad laboral en su adultez. Pero a medida que el niño crece se va orientando a una ocupación específica, y muchas de las otras habilidades se van perdiendo o ignorando. Tanto la célula como el humano utilizan la información que necesitan para su desempeño y adecuado funcionamiento, y en ambos casos la programación epigenética juega un papel fundamental. Los patrones de metilación determinan la maduración y diferenciación celular pero también contribuyen a la diferenciación entre individuos, incluso entre gemelos monocigóticos, que comparten idénticos genomas.⁹¹ La programación epigenética se ha implicado en procesos de plasticidad cerebral en respuesta a estímulos ambientales,⁹² y en el riesgo a la obesidad, principalmente en personas expuestas a una transición alimentaria,⁹³ entre otros caracteres que presentan plasticidad fenotípica. La programación en etapas tempranas de la vida en respuesta a condiciones del microambiente intrauterino, como la disponibilidad de nutrientes, la presencia de toxinas, de disruptores endócrinos, o de agentes infecciosos, puede afectar la vida adulta aumentando el riesgo de ciertos procesos mórbidos. Más aun, la programación puede modificar los gametos, por lo que las respuestas pueden ser heredadas incluso a la siguiente generación, lo que constituye una forma rápida de adaptación intergeneracional.^{94,95} Una vez que una célula pluripotencial o multipotencial se compromete en un determinado linaje el proceso se vuelve irreversible y la desmetilación ya no es posible, excepto en dos situaciones: 1) La gametogénesis, donde hay una reprogramación para formar un futuro cigoto totipotencial,⁹⁶ y 2) es posible reprogramar *in vitro* células adultas diferenciadas usando combinaciones adecuadas de factores de transcripción, convirtiéndolas en células pluripotenciales inducidas,⁹⁷ tecnología prometedora en el campo de la medicina regenerativa, especialmente cuando se asocia con la edición genética por CRISPR CAS9.⁹⁸

Función de transferencia de la información

No basta con almacenar la información y encontrar la manera de expresarla. Es necesario poder hacer copias de esa información para diseminarla y que perdure en el tiempo. Consi-

deremos la información cultural en una población humana. Esta se puede transferir de dos formas. La primera, de padres a hijos donde aquellos les enseñan a estos su lengua, sus costumbres, sus oficios etc. La segunda forma es a través de los pares, ya sean estos hermanos, primos, compañeros de escuela, amigos cercanos y otros contactos de redes sociales extendidas. El primer tipo de transferencia es vertical, y el segundo horizontal. De la misma manera, la información genética puede pasar de un organismo a otro mediante transferencia vertical, heredado de padres a hijos, pero también de forma horizontal, cuando un organismo pasa genes a otro de alguna forma.

Transferencia horizontal

Las bacterias pueden transferir de forma horizontal a sus congéneres, e incluso a miembros de otras especies, fragmentos de ADN denominados plásmidos que contienen genes que resultan necesarios en condiciones especiales, como por ejemplo de resistencia a antibióticos o a metales pesados. Estos a su vez, pueden ser copiados e insertados al genoma bacteriano mediante elementos con capacidad de autocopia llamados transposones.⁹⁹ El ADN de los mamíferos cuenta con una gran cantidad de secuencias que en principio han sido transferidas desde virus por retrotranscripción. Estas, los retrotransposones, a su vez cuentan con la capacidad de autocopiarse e insertarse en otros sitios del genoma. En conjunto, las secuencias capaces de ser transmitidas horizontalmente se denominan elementos móviles.¹⁰⁰ La transferencia horizontal de genes desde procariotas simbiotas a metazoarios huéspedes o a eucariotas unicelulares, o entre diferentes especies de metazoarios, ha sido intensa y continúa hasta la fecha, contribuyendo a la evolución de los genes y a la diversificación bioquímica. Los genes transferidos, que muchas veces carecen de función en el organismo receptor, con el tiempo pueden ser “domesticados” y evolucionar hasta ocupar nuevos nichos funcionales.^{101,102} Se ha demostrado que la transferencia de estos elementos, o genes saltarines, a los ancestros antiguos y recientes de los humanos modernos ha influido ampliamente en el proceso de diferenciación que ha conducido a nuestra especie.^{103,104} Más aun, la creación de nuevas copias de estos genes y su inserción aleatoria en el genoma durante la diferenciación celular que se da en la neurogénesis embrionaria puede contribuir a la enorme variabilidad neuronal del sistema nervioso humano. Es posible que la mala regulación de este proceso esté implicada en algunos trastornos como el desorden por estrés postraumático, alcoholismo, neurodegeneración, envejecimiento, esquizofrenia, y desórdenes del desarrollo neuronal como el síndrome de Rett.¹⁰⁵

En las antiguas formas de vida bacteriana sobre la tierra que datan de unos 3.8 mil millones de años predominaba, por mucho, la transferencia horizontal. La competencia por la supervivencia en ese entonces debió ser fenomenal, y las bacterias se alimentaban unas de otras no solo para obtener energía, sino información por medio de nuevos genes. Se considera que este tipo de transferencia genética fue intensa antes, durante y después de la aparición de LUCA (acrónimo de *Last universal common ancestor*), el presunto último ancestro común de bacterias y archa, que a su vez originaron a los eucariotas.^{106,107} A

medida que los organismos se volvieron más complejos y se fueron adaptando a diferentes ambientes, la transferencia vertical fue predominando, pero la transferencia horizontal continuó siendo un importante impulsor de la evolución.

Transferencia vertical y fidelidad de las copias

Las células humanas transmiten su material genético a su descendencia de manera vertical por medio de la replicación, es decir, la copia de todo el contenido de ADN durante la fase S del ciclo celular, que luego será repartido en partes iguales en las células hijas durante la mitosis. En cambio, el organismo humano como un todo transmite la información genética de manera vertical a su prole reduciendo a la mitad el contenido genético durante la meiosis. Ambas tienen la misma función: que la información sea pasada a la siguiente generación como una carrera de relevo. Una de las características que debe tener una transmisión de información de buena calidad es una alta fidelidad de las copias. Por ejemplo, si se copia la información de un disco duro a otro es deseable que esta sea exacta, sin errores. De la misma forma, cuando todo el ADN de una célula es copiado para que cada célula hija reciba un contenido idéntico, es deseable que no existan errores, es decir mutaciones. Existen diversos mecanismos para evitar errores de copiado. En primer lugar, la molécula de ADN polimerasa, la enzima encargada de copiar el ADN, i.e., de leer cada desoxirribonucleótido de la cadena molde, identificar el desoxirribonucleótido trifosfato (dNTP) libre complementario e incorporarlo a la cadena catalizando el enlace fosfodiéster, tiene además otra función. Si coloca un dNTP equivocado reconoce el error y luego retrocede, con su acción de exonucleasa hidroliza el dNTP incorrecto, y luego coloca el apropiado. Este proceso se conoce como corrección de pruebas y aumenta la fidelidad de las copias en unas 100 veces.¹⁰⁸ En segundo lugar, cada célula cuenta con diversos sistemas de reparación del ADN, que son complejos proteicos que intentan reconstruir la secuencia original en caso de errores de la ADN polimerasa. Las mutaciones deletéreas en los genes de las proteínas que componen estos sistemas se encuentran implicadas en la génesis del cáncer en general y de tipos específicos,¹⁰⁹⁻¹¹¹ enfermedades progéricas,¹¹² y en la intolerancia a la radiación ultravioleta, i.e., xeroderma pigmentoso,¹¹³ entre otras patologías. Por ello, a simple vista, pareciera un hecho sorprendente y contradictorio que los sistemas de reparación no hayan evolucionado para ser perfectos y disminuir a cero la tasa de error en el copiado. Sin embargo, volvemos al tema de la estabilidad versus el cambio. Las tasas de sustituciones nucleotídicas *de novo* en el esperma humano se han calculado en el orden de alrededor de 10^{-8} por sitio por

generación. Esta tasa aumenta con la edad del hombre, puesto que los errores se acumulan en cada división de las células madre espermáticas.¹¹⁴ Las mutaciones germinales relacionadas con la edad paterna se han implicado en patologías como la esquizofrenia y el autismo,¹¹⁵ pero también en la generación de variabilidad útil para la evolución. Desde el punto de vista evolutivo, la generación de variabilidad por las mutaciones tiene por lo menos dos ventajas: 1) Origina una división genética incipiente necesaria para la especialización y la evolución hacia la complejidad.¹⁴ Así, el genoma de un individuo no es necesariamente representativo de la población, y al conjunto de todas las variantes genómicas de esa población se denomina pan-genoma.¹¹⁶ 2) Da lugar a nuevas variantes genéticas, algunas de las cuales podrían ser más eficientes o mejor adaptadas a los ambientes siempre cambiantes, por lo que constituyen la fuente primaria de la evolución. Se ha observado que las tasas de mutación más altas en la línea germinal masculina corresponden a regiones metiladas en las islas de CpG.¹¹⁷ Como se explicó en otro apartado, la metilación puede actuar como un mecanismo rápido de adaptación intergeneracional por medio del silenciamiento de genes en los gametos. El hecho de que las regiones metiladas también experimenten una alta tasa de mutación en la línea germinal es novedoso, y podría evidenciar un mecanismo de adaptación genética intergeneracional rápida a los ambientes cambiantes mediante mutagénesis adaptativa dirigida,¹¹⁸ reivindicando, en parte, la teoría del uso y el desuso de Lamarck.¹¹⁹ Así, la falta de fidelidad de las copias, en promedio, es dañina para el individuo, pero beneficiosa para el grupo puesto que aumenta la diversidad necesaria para el mecanismo evolutivo. Las tasas de mutación permitidas por los sistemas de reparación en las células germinales probablemente reflejan el equilibrio entre la selección del individuo y la selección de grupo, es decir, se trata de un rasgo en el que la evolución está actuando a dos niveles diferentes de complejidad biológica (para una descripción detallada de la teoría de selección de grupo en la evolución humana se recomienda la referencia 120).

CONCLUSIÓN

Hemos recorrido las tres funciones de los dispositivos especializados en el manejo de la información: almacenamiento, expresión y transmisión, y hemos revisado cómo el ADN realiza estas funciones. Lejos de ser una revisión exhaustiva, este trabajo tiene como objetivo principal despertar la inquietud de los estudiantes y profesionales de la salud en un área total para el estudio de la nueva medicina genómica de precisión.

REFERENCIAS

- Singh P, Batra HS, Naithani M. History of biochemistry .Bull Indian Inst Hist Med [Internet].2004[citado el 10 de septiembre de 2017]; 34(1):75-86. Disponible en: http://www.ccras.nic.in/sites/default/files/viewpdf/fjmh/BIIHM_2004/75%20to%2086.pdf
- Tata JR. Signalling through nuclear receptors. *Nat Rev Mol Cell Biol*. 2002;3(9):702-10. DOI:10.1038/nrm914
- Pierce, KL, Premont RT, Lefkowitz RJ. Signalling: seven-transmembrane receptors. *Nat Rev Mol Cell Biol*. 2002;3(9):639-50. DOI:10.1038/nrm908
- Hill SJ. G-protein-coupled receptors: past, present and future. *Br J Pharmacol*. 2006;147(S1):S27-S37. DOI:10.1038/sj.bjp.0706455
- Unwin N. The structure of ion channels in membranes of excitable cells. *Neuron* [Internet]. 1989 [citado el 15 de septiembre de 2017]; 3:665-676. Disponible en: <http://www2.mrc-lmb.cam.ac.uk/groups/nu/pdf/neuron89.pdf>
- Lyle WG, Kober PA. The Importance of Enzymes and Enzyme Reactions in Medicine and Surgery. *Ind Eng Chem*. 1914;6(10):855-56. DOI:10.1021/ie50070a025
- Christakopoulos P, Topakas E. Editorial note: advances in enzymology and enzyme engineering. *Comput Struct Biotechnol J*. 2012;2:e201209001. DOI:10.5936/csbj.201209001
- Barnett JH. Applications of boolean algebra: Claude Shannon and circuit design [Internet]. Colorado (US): Department of Mathematics and Physics; 2011. [citado el 19 de septiembre de 2017]. Disponible en: https://www.maa.org/sites/default/files/images/upload_library/46/Pengelly_projects/Project-9/Project_9-Shannon_with_figures.pdf
- Ribatti D. William Harvey and the discovery of the circulation of the blood. *J Angiogenesis Res*. 2019;1(1):3. DOI:10.1186/2040-2384-1-3
- Gest H. The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society. *Notes Rec R Soc Lond* 2004;58(2):187-201. DOI:10.1098/rsnr.2004.0055
- Travers A, Muskhelishvili G. DNA structure and function. *FEBS J*. 2015;282(12):2279-95. DOI:10.1111/febs.13307
- Bostrom N. Are you living in a computer simulation? *Philos Q* [Internet]. 2003 [citado el 20 de septiembre de 2017]; 53(211):243-255. Disponible en: <https://www.simulation-argument.com/simulation.pdf>
- Shapiro B. Mammoth 2.0: will genome engineering resurrect extinct species? *Genome Biol* 2015;16(1):228. DOI:10.1186/s13059-015-0800-4
- Herrera-Paz EF. Superorganismo universal. Una teoría de la evolución hacia la complejidad. Charleston, USA: Createspace; 2014.
- Watson JD, Crick FH. Molecular structure of nucleic acids. *Nature*. 1953;171(4356):737-38. DOI:10.1038/171737a0
- Kleckner N, Zickler D, Jones GH, Dekker J, Padmore R, Henle J, et al. A mechanical basis for chromosome function. *Proc Natl Acad Sci USA*. 2004;101(34):12592-97. DOI:10.1073/pnas.0402724101
- Mallows C. John Tukey at Bell Labs. *Statist Sci*. 2003;18(3):332-5. DOI:10.1214/ss/1076102420
- Love N. Are languages digital codes? *Language Sciences* [Internet]. 2007 [no hay fecha de cita]; 29(5): 690-709. Disponible en: <https://doi.org/10.1016/j.langsci.2007.01.008>
- Blaauw GA, Brooks Jr FP, Buchholz W. Anecdotes News and Notices. *Ann Hist Comput* [Internet]. 1981 [no hay fecha de cita];3(1):72. Disponible en: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4392913>
- Ibáñez Á. Internet y el correo electrónico en español. *ACTA* [Internet]. [citado el 20 de septiembre de 2017]. [s.f.]: 31-41. Disponible en: <https://www.acta.es/medios/articulos/internet/005029.pdf>
- Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333-51. DOI:10.1038/nrg.2016.49
- Waddington SN, Privilizzi R, Karda R, O'Neill HC. A broad overview and review of CRISPR-Cas technology and stem cells. *Curr Stem Cell Rep*. 2016;2(1):9-20. DOI:10.1007/s40778-016-0037-5
- Garg A, Choudhary M. Analysing and obtaining the most efficient DNA computing algorithm. *Journal of Computational Intelligence in Bioinformatics (JCBI)*. 2015;8(1):1-6. <http://dx.doi.org/10.1155/2013/160687>
- Cave A. What will we do when the world's data hits 163 zettabytes in 2025? [Internet]. *Forbes.com*. 2018 [citado el 20 de septiembre de 2017]. Disponible en: <https://www.forbes.com/sites/andrewcave/2017/04/13/what-will-we-do-when-the-worlds-data-hits-163-zettabytes-in-2025/#43d245df349a>
- O'Driscoll A, Sleator RD. Synthetic DNA: the next generation of big data storage. *Bioengineered*. 2013; 4(3):123-125. <https://doi.org/10.4161/bioe.24296>
- Davis, J. *Microvenus*. *Art J*. 1996;55(1):70-4. DOI:10.2307/777811
- Church GM, Gao Y, Kosuri S. Next-generation digital information storage in DNA. *Science*. 2012;337(6102):1628 DOI:10.1126/science.1226355
- Shipman SL, Nivala J, Macklis JD, Church GM. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*. 2017;547(7663):345-49. DOI:10.1038/nature23017
- Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, Sipos B, et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*. 2013;494(7435):77-80. DOI:10.1038/nature11875
- Sagan C. *The Dragons of Eden: Speculations on the Nature of Human Intelligence*. Nueva York, USA: Random House; 1977.
- Kerker K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet*. 2008;40(7):904-8. DOI:10.1038/ng.174
- Wadhwa PD, Buss C, Entringer S, Swanson JM. Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. *Semin Reprod Med*. 2009;27(5):358-68. DOI:10.1055/s-0029-1237424
- Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell*. 2009;101(5):251-62. DOI:10.1042/BC20080104
- Pathy, L. Genome evolution and the evolution of exon-shuffling—a review. *Gene*. 1999;238(1):103-114. [https://doi.org/10.1016/S0378-1119\(99\)00228-0](https://doi.org/10.1016/S0378-1119(99)00228-0)
- Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet*. 2004;5(6):435-45. DOI:10.1038/nrg1348
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. 2010;464(7289):704-12. DOI:10.1038/nature08516
- Zhang Z, Carriero N, Gerstein M. Comparative analysis of processed pseudogenes in the mouse and human genomes. *Trends Genet*. 2004;20(2):62-7. DOI:10.1016/j.tig.2003.12.005
- Zhang Z, Gerstein M. Large-scale analysis of pseudogenes in the human genome. *Curr Opin Genet Dev*. 2004;14(4):328-35. DOI:10.1016/j.gde.2004.06.003
- Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays*. 2007;29(3):288-99. DOI:10.1002/bies.20544
- Ohno, S. So much "junk" DNA in our genome. *Evolution of Genetic Systems*. Brookhaven Symp Biol. 1972;23:366-70. <http://www.junkdna.com/ohno.html>
- Andolfatto P. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature*. 2005;437(7062):1149-52. DOI:10.1038/nature04107
- Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, et al. Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Hum Mol Genet*. 2014;23(22):5866-78. DOI:10.1093/hmg/ddu309
- Buckingham ME. The Control of muscle gene expression: A review of molecular studies on the processing of primary transcripts. *Br Med Bull*. 1989;45(3):608-29.
- Breitbart RE, Andreadis A, Nadal-Ginard B. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu Rev Biochem*. 1987;56(1):467-95. DOI:10.1146/annurev.bi.56.070187.002343
- Smith CW, Patton JG, Nadal-Ginard B. Alternative splicing in the control of gene expression. *Annu Rev Genet*. 1989;23(1):527-77. DOI:10.1146/annurev.ge.23.120189.002523
- Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet*. 2010;11(5):345-55. DOI:10.1038/nrg2776
- Mitchell SF, Walker SE, Algire MA, Park EH, Hinnebusch AG, Lorsch JR. The 5'-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. *Mol Cell*. 2010;39(6):950-62. DOI:10.1016/j.molcel.2010.08.021
- Munroe D, Jacobson A. mRNA poly (A) tail, a 3' enhancer of translational initiation. *Mol Cell Bio*. 1990;10(7):3441-55. DOI:10.1128/MCB.10.7.3441
- Gardini S, Chelli S, Baroni S, Di Lascio G, Mangiavacchi G, Micheletti N, et al. On nature's strategy for assigning genetic code multiplicity. *PLoS one*.

- 2016; 11(2):e0148174. DOI: 10.1371/journal.pone.0148174
50. Koonin EV, Novozhilov AS. Origin and evolution of the universal genetic code. *Annu Rev Genet.* 2017;51:45-62. DOI:10.1146/annurev-genet-120116-024713
 51. Zhang G, Ignatova Z. Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS one.* 2009;4(4):e5036. DOI:10.1371/journal.pone.0005036
 52. Ibbá M, Söll D. Aminoacyl-tRNA synthesis. *Annu Rev Biochem.* 2000;69(1):617-50. DOI:10.1146/annurev.biochem.69.1.617
 53. Guo M, Yang XL, Schimmel P. New functions of aminoacyl-tRNA synthetases beyond translation. *Nat Rev Mol Cell Biol.* 2010;11(9):668-74. DOI:10.1038/nrm2956
 54. Fink AL. Chaperone-mediated protein folding. *Physiol Rev.* 1999;79(2):425-49. DOI: 10.1152/physrev.1999.79.2.425
 55. Aftabuddin MD, Kundu S. Hydrophobic, hydrophilic, and charged amino acid networks within protein. *Biophys J [Internet].* 2007 [citado el 23 de septiembre de 2017]; 93(1): 225-231. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1914426/> DOI:10.1529/biophysj.106.098004
 56. Englander SW, Mayne L. The nature of protein folding pathways. *Proc Natl Acad Sci U S A* 2014;111(45):15873-80. DOI:10.1073/pnas.1411798111
 57. Díaz-Villanueva JF, Díaz-Molina R, García-González V. Protein folding and mechanisms of proteostasis. *Int J Mol Sci.* 2015;16(8):17193-230. DOI:10.3390/ijms160817193
 58. Balchin D, Hayer-Hartl M, Hartl FU. In vivo aspects of protein folding and quality control. *Science.* 2016;353(6294):aac4354. DOI:10.1126/science.aac4354
 59. Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, et al. The strength of phenotypic selection in natural populations. *The Am Nat* 2001;157(3):245-61. DOI:10.1086/319193.
 60. Herrera-Paz EF. La genética de poblaciones y el origen de la diversidad humana. *Rev Med Hondur [Internet].* 2013 [citado el 24 de septiembre de 2017]; 81(1): 40-45. Disponible en: <http://www.bvs.hn/RMH/pdf/2013/pdf/Vol81-1-2013-10.pdf>
 61. Herrera-Paz EF. Apellidos e isonimia en las comunidades garífunas de la costa atlántica de Honduras. *Rev Med Inst Mex Seguro Soc [Internet].* 2013 [citado el 24 de septiembre de 2017];51(2):150-157. Disponible en: <http://www.medigraphic.com/pdfs/imss/im-2013/im132g.pdf>
 62. Herrera-Paz EF. Aislamientos genéticos y costumbres endogámicas en tres municipios rurales de Honduras. *Rev Med Inst Mex Seguro Soc [Internet].* 2016 [citado el 24 de septiembre de 2017]; 54(4):504-513. Disponible en: <http://www.medigraphic.com/pdfs/imss/im-2016/im164n.pdf>
 63. Herrera-Paz EF. Biodemography research and the history of Central American and North-West South American populations. En: Ubelaker DH, Colantonio S, editores. *Anthropological research in Latin America.* Washington D.C., USA: Smithsonian Institution Scholarly Press; 2018. En imprenta.
 64. Koonin EV, Novozhilov AS. Origin and evolution of the genetic code: the universal enigma. *IUBMB life.* 2009;61(2):99-111. DOI:10.1002/iub.146
 65. Lenski RE, Barrick JE, Ofria C. Balancing robustness and evolvability. *PLoS biology.* 2006;4(12):e428. DOI:10.1371/journal.pbio.0040428
 66. Kim SH, Sussman JL, Suddath FL, Quigley GJ, McPherson A, Wang AH J, et al. The general structure of transfer RNA molecules. *Proc Natl Acad Sci USA.* 1974;71(12):4970-74. <http://www.jstor.org/stable/63951>
 67. Wimberly BT, Brodersen DE, Clemons Jr WM, Morgan-Warren RJ, Carter AP, Vonnrhein C, et al. Structure of the 30S ribosomal subunit. *Nature.* 2000; 407(6802):327-39. DOI: 10.1038/35030006
 68. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nature. biotechnology* 2015; 33(8):831-38. DOI: 10.1038/nbt.3300
 69. Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet.* 2012;13(11):770-80. DOI:10.1038/nrg3296
 70. Whitaker JW, Chen Z, Wang W. Predicting the human epigenome from DNA motifs. *Nat Methods.* 2015;12(3):265-72. DOI:10.1038/nmeth.3065
 71. Gasser SM, Amati BB, Cardenas ME, Hofmann JFX. Studies on Scaffold Attachment Sites and Their Relation to Genome Function. *Int Rev Cytol [Internet].* 1990 [citado el 26 de septiembre de 2017]; 119:57-96. Disponible en: [https://doi.org/10.1016/S0074-7696\(08\)60649-X](https://doi.org/10.1016/S0074-7696(08)60649-X)
 72. Cheeseman IM, Drubin DG, Barnes G. Simple centromere, complex kinetochore: linking spindle microtubules and centromeric DNA in budding yeast. *J Cell Biol.* 2002;157(2):199-203. DOI:10.1083/jcb.200201052
 73. Liu D, O'Connor MS, Qin J, Songyang Z. Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *J Biol Chem.* 2004;279(49):51338-42. DOI:10.1074/jbc.M409293200
 74. Jones NC, Rigby PW, Ziff EB. Trans-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. *Genes Dev.* 1988;2(3):267-81. DOI:10.1101/gad.2.3.267
 75. Bolouri H, Davidson EH. Modeling DNA sequence-based cis-regulatory gene networks. *Dev Biol.* 2002;246(1):2-13. DOI:10.1006/dbio.2002.0617
 76. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, et al. Coding exons function as tissue-specific enhancers of nearby genes. *Genome res.* 2012;22(6):1059-68. DOI:10.1101/gr.133546.111
 77. Sekido R, Lovell-Badge R. Sex determination and SRY: down to a wink and a nudge? *Trends Genet.* 2009;25(1):19-29. DOI:10.1016/j.tig.2008.10.008
 78. Freeman ER, Bloom DA, McGuire EJ. A brief history of testosterone. *J Urol.* 2001;165(2):371-73. DOI:10.1097/00005392-200102000-00004
 79. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, et al. Abundant pleiotropy in human complex diseases and traits. *The Am J Hum Genet.* 2011;89(5):607-18. DOI:10.1016/j.ajhg.2011.10.004
 80. Paaby AB, Rockman MV. The many faces of pleiotropy. *Trends Genet.* 2013;29(2):66-73. DOI:10.1016/j.tig.2012.10.010
 81. Turing AM. The Chemical Basis of Morphogenesis. 1953. *Philos Trans R Soc Lond B Biol Sci [Internet].* 1952 [citado el 2 de octubre de 2017]; 237(641):37-72. Disponible en: <http://www.dna.caltech.edu/courses/cs191/paperscs191/turing.pdf>
 82. Green JB, Sharpe J. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development.* 2015;142(7):1203-11. DOI:10.1242/dev.114991
 83. Lawrence JG, Hendrix RW, Casjens S. Where are the pseudogenes in bacterial genomes?. *Trends microbial.* 2001;9(11):535-40. [https://doi.org/10.1016/S0966-842X\(01\)02198-9](https://doi.org/10.1016/S0966-842X(01)02198-9)
 84. Chang CW, Cheng WC, Chen CR, Shu WY, Tsai ML, Huang CL, et al. Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis. *PLoS one.* 2011;6(7):e22859. DOI:10.1371/journal.pone.0022859
 85. Uings IJ, Farrow SN. Cell receptors and cell signalling. *Mol Pathol [Internet].* 2000 [citado el 5 de octubre de 2017]; 53(6): 295-299. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1186983/>
 86. Kholodenko BN. Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol.* 2006;7(3):165-76. DOI:10.1038/nrm1838
 87. Walker WH, Cheng J. FSH and testosterone signaling in Sertoli cells. *Reproduction* 2005;130(1):15-28. DOI:10.1530/rep.1.00358
 88. Cantone I, Fisher AG. Epigenetic programming and reprogramming during development. *Nat Struct Mol Biol.* 2013;20(3):282-89. DOI:10.1038/nsmb.2489
 89. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012;13(7):484-92. DOI:10.1038/nrg3230
 90. Martí M, Mulero L, Pardo C, Morera C, Carrió M, Laricchia-Robbio L, et al. Characterization of pluripotent stem cells. *Nat protoc.* 2013;8(2):223-53. DOI:10.1038/nprot.2012.154
 91. Bell JT, Spector TD. A twin approach to unraveling epigenetics. *Trends Genet.* 2011;27(3):116-25. DOI:10.1016/j.tig.2010.12.005
 92. Fagioli M, Jensen CL, Champagne FA. Epigenetic influences on brain development and plasticity. *Curr Opin Neurobiol.* 2009;19(2):207-12. DOI:10.1016/j.conb.2009.05.009
 93. van Dijk SJ, Molloy PL, Varinli H, Morrison JL, Muhlhauser BS, Buckley M. Epigenetics and human obesity. *Int J Obes (Lond).* 2015;39(1):85-97. DOI:10.1038/ijo.2014.34.
 94. St-Cyr S, McGowan PO. Programming of stress-related behavior and epigenetic neural gene regulation in mice offspring through maternal exposure to predator odor. *Front Behav Neurosci.* 2015;9:145. DOI:10.3389/fnbeh.2015.00145
 95. Fernandez-Twinn DS, Constância M, Ozanne SE. Intergenerational epigenetic inheritance in models of developmental programming of adult disease. *Semin Cell Dev Biol.* 2015;43:85-95. DOI:10.1016/j.semcdb.2015.06.006
 96. Sasaki H, Matsui Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet.* 2008;9(2):129-40. DOI:10.1038/nrg2295
 97. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 2006;126(4):663-76. <https://doi.org/10.1016/j.cell.2006.07.024>
 98. Hockemeyer D, Jaenisch R. Induced pluripotent stem cells meet genome editing. *Cell Stem Cell.* 2016;18(5):573-86. DOI:10.1016/j.stem.2016.04.013
 99. MacLean RC, San Millan A. Microbial evolution: towards resolving

- the plasmid paradox. *Curr Biol.* 2015;25(17):R764-7. DOI:10.1016/j.cub.2015.07.006
100. Kazazian HH. Mobile elements: drivers of genome evolution. *Science.* 2004;303(5664):1626-32. DOI:10.1126/science.1089670
 101. Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* 2015;16(1):50. DOI:10.1186/s13059-015-0607-3
 102. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. *Nat Rev Genet.* 2015;16(8):472-82. DOI:10.1038/nrg3962
 103. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 2009;10(10):691-703. DOI:10.1038/nrg2640
 104. Britten RJ. Transposable element insertions have strongly affected human evolution. *Proc Natl Acad Sci U S A.* 2010;107(46):19945-48. DOI:10.1073/pnas.1014330107
 105. Erwin JA, Marchetto MC, Gage FH. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat Rev Neurosci.* 2014;15(8):497-506. DOI:10.1038/nrn3730
 106. Jheeta S. Horizontal gene transfer and its part in the reorganisation of genetics during the LUCA epoch. *Life (Basel).* 2013;3(4):518-23. DOI:10.3390/life3040518
 107. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, et al. The physiology and habitat of the last universal common ancestor. *Nat Microbiol.* 2016;1(9):16116. DOI:10.1038/nmicrobiol.2016.116
 108. Reha-Krantz LJ. DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochim Biophys Acta.* 2010;1804(5):1049-63. DOI:10.1016/j.bbapap.2009.06.012
 109. Sehgal R, Sheahan K, O'Connell PR, Hanly AM, Martin ST, Winter DC. Lynch syndrome: an updated review. *Genes.* 2014;5(3):497-507. DOI:10.3390/genes5030497
 110. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, et al. Inherited DNA-repair gene mutations in men with metastatic prostate cancer. *N Engl J Med.* 2016;375(5):443-53. DOI:10.1056/NEJMoa1603144
 111. Jeggo PA, Pearl LH, Carr AM. DNA repair, genome stability and cancer: a historical perspective. *Nat Rev Cancer.* 2016;16(1):35-42. DOI:10.1038/nrc.2015.4
 112. Karikkineth AC, Scheibye-Knudsen M, Fivenson E, Croteau DL, Bohr VA. Cockayne syndrome: clinical features, model systems and pathways. *Ageing Res Rev.* 2017;33:3-17. DOI:10.1016/j.arr.2016.08.002
 113. Koch SC, Simon N, Ebert C, Carell T. Molecular mechanisms of xeroderma pigmentosum (XP) proteins. *Q Rev Biophys.* 2016;49:e5. DOI:10.1017/S0033583515000268
 114. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. *Nature.* 2012;488(7412):471-5. DOI:10.1038/nature11396
 115. Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet.* 2012;13(8):565-75. DOI:10.1038/nrg3241
 116. Vos M, Eyre-Walker A. Are pangenomes adaptive or not? *Nat Microbiol.* 2017;2(12):1576. DOI:10.1038/s41564-017-0067-5
 117. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, et al. Timing, rates and spectra of human germline mutation. *Nat genet.* 2016;48(2):126-33. DOI:10.1038/ng.3469
 118. Skinner MK, Guerrero-Bosagna C, Haque MM. Environmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations. *Epigenetics.* 2015;10(8):762-71. DOI:10.1080/15592294.2015.1062207
 119. Wang Y, Liu H, Sun Z. Lamarck rises from his grave: parental environment-induced epigenetic inheritance in model organisms and humans. *Biol Rev Camb Philos Soc.* 2017;92(4):2084-2111. DOI:10.1111/brv.12322.
 120. Wilson O. *La conquista social de la Tierra.* Barcelona, España: Editor.

ABSTRACT. Throughout the last two centuries medicine was nourished with biochemical discoveries that promoted the understanding of physiopathological mechanisms and facilitated the development of therapeutics. On the other hand, in the present century we entered the era of genomics and big data, so the study of the functions of DNA as an information management device is essential for the understanding of the new precision, personalized genomic medicine. In this review, DNA is analyzed as a computer device with three functions: storage, expression and transmission of the information accumulated throughout the phylogeny in the form of nucleotide sequences. Each of these functions is described by comparing it with the information handled by a computer or a society, and examples of pathologies arising from the failure to properly perform such functions are provided. The bibliographic compilation is extensive and includes the most relevant articles corresponding to each topic, both historical and of the state of the art.

Keyword: Bioinformatics, Genomics, Hereditary diseases, Medical genetics, Nucleic acids