

La nueva perspectiva molecular del gen en la era posgenómica

The new molecular gene perspective in the postgenomic era

Pedro Martínez Gómez 

Acceso Abierto

Correspondencia:

pmartinez@cebas.csic.es
Departamento de Mejora vegetal,
CEBAS-CSIC, PO Box 164, E-30100
Espinaro, Murcia, España.

Sometido: 16-12-2021
Aceptado para publicación:
06-05-2022
Publicado en línea: 01-07-2022

Palabras clave:

Dogma central de la Biología
Molecular; gen; genética;
genómica; posgenómica.

Key words:

Central Dogma of Molecular
Biology; gen; genetics;
genomics; postgenomic.

Citación:

Martínez Gómez P. La nueva
perspectiva molecular del gen en la
era posgenómica. *Magna Scientia
UCEVA* 2022;2:1 69-78.
<https://doi.org/10.54502/msuceva.v2n1a7>

Resumen

El Proyecto ENCODE (Encyclopedia of DNA Elements) fue considerado como una continuación del Proyecto Genoma Humano (PGH) que tenía como objetivo identificar todos los elementos funcionales en el genoma y profundizar en el análisis de la expresión del gen y su complejidad. A pesar de los cientos de miles de proteínas presentes en el ser humano únicamente 20.000 genes habían sido descritos. El objetivo principal del proyecto ENCODE era determinar el papel del resto del componente del genoma, excluyendo las regiones codificantes o genes. Sin embargo, partir de ENCODE, en la nueva era posgenómica, se evidenciaron nuevos fenómenos moleculares relacionados con el genoma y localizados en el núcleo de la célula (incluyendo las variaciones de copia del genoma, los genes de fusión, los fenómenos de pleiotropía, la herencia epigenética, la epitranscriptómica, las epimutaciones, los daños del ADN, la transmisión transgeneracional de información ambiental o la agrupación del ADN en una cuádruple hélice) o no relacionados con el genoma y localizados en el citoplasma celular (incluyendo la herencia mediada por material extra-genómico, las modificaciones postraduccionales de proteínas, la presencia de glucógenos y la regulación de ARNt nuclear, cloroplástico y mitocondrial) que cuestionan el concepto de gen y el Dogma Central de la Biología Molecular (DCBM). Estos nuevos fenómenos que discutiremos a continuación han supuesto una nueva perspectiva molecular del gen y del DCBM.

Abstract

The ENCODE Project (Encyclopedia of DNA Elements) was considered as a continuation of the Human Genome Project (HGP) that aimed to identify all the functional elements in the genome and deepen the analysis of gene expression and its complexity. Despite the hundreds of thousands of proteins present in humans, only 20000 genes had been described. The main objective of the ENCODE project was to determine the role of the rest of the genome component, excluding the coding regions or genes. However, starting with ENCODE, in the new post-genomic era, new molecular phenomena related to the genome and located in the nucleus of the cell (including genome copy variations, fusion genes, pleiotropy phenomena, epigenetic inheritance, epitranscriptomics, epimutations, DNA damage, transgenerational transmission of environmental information or DNA quadruple helix assembly) or not related to the genome and located in the cell cytoplasm (including inheritance mediated by extraneous material).-genomic, post-translational modifications of proteins, the presence of glycogens and the regulation of nuclear, chloroplastic and mitochondrial tRNA) that question the concept of gene and the Central Dogma of Molecular Biology (DCBM). These new phenomena that we will discuss below, have allowed to a new molecular gene perspective and DCBM.



Copyright: ©2022 para el autor. Este es un artículo de acceso abierto distribuido bajo los términos y condiciones de la licencia internacional Creative Commons Atribución-No comercial- Sin Derivados 4.0 (CC BY-NC-ND 4.0) <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es/>

Introducción

La secuenciación del genoma completo del ser humano dentro del Proyecto Genoma Humano (PGH) en 2001, se constituía como el principal objetivo de la genómica que abriría las puertas a la secuenciación del genoma completo del resto de organismos vivos y al descifrado de las claves de la herencia. Sin embargo, los resultados del PGH, propiciaron una serie de incógnitas acerca de la naturaleza y expresión del gen molecular como fragmento del ADN que supusieron de inmediato, el comienzo de un nuevo proyecto para realizar una mejor caracterización a modo de enciclopedia de todos los elementos del ADN, no solo los transcritos. El conocimiento de la secuencia de ADN del genoma completo de un ser vivo, no era suficiente para conocer las bases de la herencia de los caracteres. Era necesaria la realización de nuevos estudios al margen del genoma. El PGH, no cumplió las expectativas de los científicos y abrió una nueva era posgenómica más compleja para descifrar las claves de la herencia que comenzó con el proyecto ENCODE (Encyclopedia of DNA Elements). La posgenómica, propició también un cambio de perspectiva en la descripción del gen molecular poniendo el foco de atención en el ARN a la vez que en el ADN, y evidenció otros muchos fenómenos experimentales que lo cuestionaría. Este cambio de perspectiva biológica estuvo también propiciado por la introducción en la posgenómica de nuevas técnicas de análisis masivo de genomas y transcriptomas que producen millones de datos por experimento en la nueva ciencia denominada Big Data.

En este artículo de reflexión sobre la nueva perspectiva molecular del gen en la era posgenómica, analizaremos los nuevos desarrollos metodológicos que han dado lugar a esta nueva perspectiva y que han generado nuevos fenómenos experimentales.

Análisis masivo de genomas y transcriptomas

Dentro de la postgenómica tenemos que destacar como principal rasgo característico desde el punto de vista de la experimentación, la incorporación de nuevos instrumentos específicos de análisis masivo tanto de DNA como de ARN que presentan, sin embargo, un elevado costo. Estas nuevas aproximaciones metodológicas, así como la creación de bases de datos con miles de millones de datos, ofrece una nueva perspectiva en la genética como ciencia de laboratorio.

En los años 70, se inició la carrera de la secuenciación del ADN mediante el método Sanger, que usaba dideoxinucleótidos que interrumpían la polimerización del ADN dando lugar a fragmentos de diferente longitud a partir de los que se deducía la secuencia [1]. Posteriormente, a finales de los 90's, se emplearon dideoxinucleótidos marcados con diferentes colorantes, lo que agilizó la lectura del ADN; permitiendo leer secuencias de hasta 1000 bases de forma automática. Esta primera generación de secuenciadores tenía un costo promedio de 1 dólar por base, y fueron empleados en las primeras fases del PGH [2] (ver figura 1).

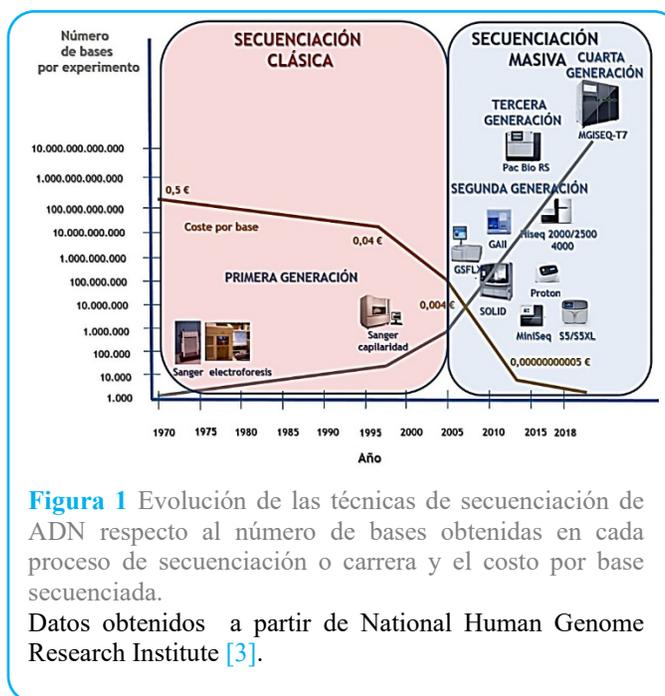


Figura 1 Evolución de las técnicas de secuenciación de ADN respecto al número de bases obtenidas en cada proceso de secuenciación o carrera y el costo por base secuenciada.

Datos obtenidos a partir de National Human Genome Research Institute [3].

Durante los primeros años del siglo XXI, aparecieron las técnicas de secuenciación masiva (high-throughput en inglés) llamadas también NGS (Next Generation Sequencing en inglés). Estas nuevas tecnologías basadas en la generación de miles de reacciones de secuenciación en paralelo inmovilizadas en una superficie sólida lo que disminuye los requerimientos de reactivos y la necesidad de clonar los fragmentos de ADN, abaratando así el coste, que pasó a 0.1 dólar por millones de bases en 2013 [4]. La primera versión de estos nuevos secuenciadores de segunda generación, se llevó a cabo mediante la plataforma GSFLX Titanium (Roche), utilizando la pirosecuenciación del ADN a finales del siglo pasado [5]. Esta plataforma es capaz de generar secuencias de hasta 700 bases, lo que la hacía adecuada para la secuenciación de genomas *de novo* antes de la aparición de los secuenciadores de tercera generación (figura 1). Otra

versión de la secuenciación de segunda generación, fue desarrollada en 2005 por SOLID/AB™, que utiliza cuatro tipos de octámeros, los cuales emitían un tipo de fluorescencia que sería registrada para posteriormente ser comparada y ensamblada en la secuencia final. Este tipo de plataformas en la actualidad, están en desuso debido a su reducido rendimiento, elevado costo y sobre todo, dificultad en la interpretación de resultados para análisis posteriores.

Una tercera versión de las técnicas de segunda generación, corresponde a la plataforma Solexa™. Derivadas de esta tecnología, las plataformas desarrolladas a partir de Solexa por Illumina GAI® y posteriormente HiSeq 2000/2500®, Hiseq X y Hiseq4000®[6], son más económicas que la pirosecuenciación y son más fiables en la secuenciación de las regiones homopoliméricas, pero la pequeña longitud de sus lecturas (entre 80 y 250 bases), las limita a ser aplicadas a casos de resecuenciación y de expresión de ARN. Estas plataformas son en la actualidad, las más utilizadas debido a su robustez, de lecturas por ensayo y el costo de secuenciación por Mpb es de 10 € (de los más elevados), pero el costo del equipo está alrededor de 6000 €, siendo de los más económicos. También se debe señalar que los prototipos de segunda generación, son más económicos y asequibles a laboratorios individuales. En la actualidad, sin embargo estas plataformas de secuenciación de bajo coste están en desuso debido a su falta de fiabilidad y robustez a pesar de su buen rendimiento (figura 1).

Más recientemente, en 2011, llega al mercado la tercera generación de secuenciadores, basados en la tecnología SMRT™ (Single Molecule Real Time Sequencing), destacando la plataforma PacBio™ desarrollada por la empresa Pacific Bioscience [7]. Estas técnicas están limitadas únicamente a la secuenciación del ADN y consisten en la secuenciación a tiempo real de una molécula de ADN adherida a una superficie sólida, gracias al uso de dNTP con diferentes fluoróforos unidos. Además, más recientemente en 2018, se ha desarrollado la primera plataforma de secuenciación masiva de ADN con tecnología china denominada DNBseq™ basada también al igual que PacBio® en análisis de fluorescencia de fragmentos de ADN en un secuenciador desarrollado por el BGI (Beiging Genome Institut) de Pekín, China llamado MGISEQ-T7®.

Este secuenciador, considerado de cuarta generación, es capaz de secuenciar hasta 60 genomas de humanos en un

día con capacidad de secuenciación diaria de 6 Terabites (TB) [8] (figura 1).

Observación, medición y creación de bases de datos: la nueva ciencia del Big Data

Las nuevas metodologías de secuenciación masiva producen datos obtenidos mediante fenómenos físico-químicos conectados, aunque no son fruto de la observación directa ni del ADN ni del ARN. Tenemos que entender pues la ciencia molecular obtenida de las nuevas tecnologías postgenómicas como un fenómeno codificado y una serie de aspectos físico-químicos en los que no hay una síntesis observable del proceso estudiado con un flujo de datos computacionales. Este escenario presenta importantes retos desde la perspectiva del pluralismo metodológico y su validez. La acumulación de los datos de la secuenciación a lo largo de estos últimos años ha desencadenado el desarrollo de bases de datos inmensas a nivel genómico y transcriptómico.

Las nuevas metodologías de análisis masivo producen millones de datos depositados en estas bases de datos que es necesario analizar y utilizar en la interpretación de los fenómenos. Incluso algunos autores como Callebaut [9], caracterizaron esta biología de datos intensivos (Big Data) como un nuevo tipo de ciencia diferente de la llamada biología que se transformarían entonces en una ciencia de gestión de la información.

Podemos clasificar estas bases de datos genómicas como bases de datos primarias que incluyen las secuencias de nucleótidos procedentes de resultados experimentales (NCBI GenBank, Entrez, EMBL o DDBJ) y las secuencias múltiples anotadas con la proteína a la que codifican (Uniprot y NCBI Protein Databases); bases de datos secundarias con estructuras de proteínas (Prosite, Pfam e Interpro) o huellas de polimorfismos de ADN (PRINTS); y bases de datos terciarias que incluyen los datos fenotípicos como PhenomicDB® [10]. Respecto a las bases de datos primarias que son las más importantes en Postgenómica, en la figura 2, se puede observar la evolución del número de bases y secuencias incluidas en la base de datos primarias (de ADN) del NCBI (National Center of Biotechnology Information) del Instituto Nacional de Salud de EEUU (National Institutes of Health, NIH, en inglés).

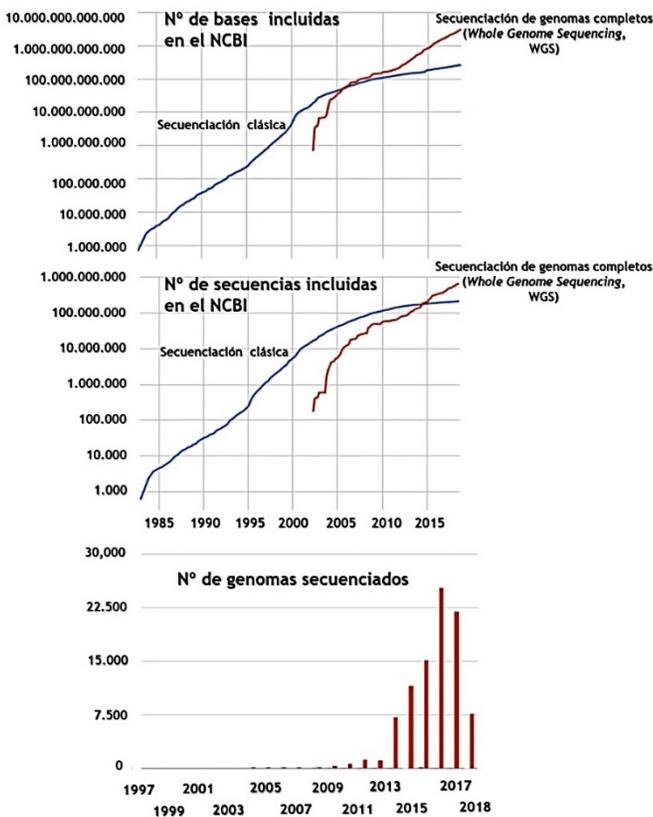


Figura 2 Evolución del número de bases y secuencias incluidas en la base de datos NCBI (National Center of Biotechnology Information) del Instituto Nacional de Salud de EEUU [11] y del número de especies secuenciadas según el Genome Online Database (GOLD) [12].

En los últimos 30 años, se ha pasado de 1 millón de secuencias en la base de datos a casi 8 billones de secuencias (8.000.000.000.000), observándose un crecimiento exponencial similar al del número de secuencias almacenadas, de unas 8000 en 1985 a los actuales 1.3 millones tanto mediante secuenciación clásica como mediante secuenciación de genomas completos a partir de 2003. Las tres bases de datos genómicas más importantes son la mencionada norteamericana del NCBI (National Center of Biotechnology Information, www.ncbi.nlm.nih.gov), la Japonesa KEGG (Kyoto Encyclopedia of Genes and Genomes [13]) y más recientemente, la europea (European Genome-Phenome Archive [14]). Estas bases de datos genómicas en la actualidad, siguen completándose. Además, existen otras bases de datos más recientes. Una de las más importantes es la del Proyecto Herencia mendeliana en el Ser Humano (OMIM), que

incluye un catálogo de genes humanos y de los trastornos y rasgos genéticos con libre acceso a información sobre las enfermedades humanas mendelianas y sobre más de 15000 genes. Se centra especialmente en la relación entre genotipo y fenotipo a la luz de las nuevas reflexiones postgenómicas derivadas del proyecto ENCODE. En 2016, se creó también la base de datos DASHR (database of small human noncoding RNAs) para aglutinar los miARNs identificados en humanos [15].

Más recientemente, también a nivel postgenómico, cabe destacar una base de datos celular que pretende revolucionar el conocimiento de la genética del ser humano a nivel celular, el Human Cell Atlas (HCA, atlas de células en humanos) [16], donde se pretende hacer un mapa de todas las células del ser humano. La base tecnológica es la secuenciación a nivel de ARN de células individuales [17]. Este proyecto está financiado por el National Institute of Health (NIH) de EEUU dentro del programa NIH Human Biomolecular Atlas Program (HuBMAP). Además, la iniciativa Genomic Data Commons del NHI de EEUU [18], que contiene más de 30000 genomas secuenciados [19]. Finalmente, otras bases de datos postgenómicas muy recientes, son el Human Protein Atlas (Atlas de Proteínas en Humanos o también llamado Proyecto proteoma Humano) comenzado en 2016, pretende la identificación de proteínas en humanos mediante inmunohistoquímica y su localización a nivel celular [20] y el Human Epigenetic Consortium (IHEC) que pretende incluir datos epigenéticos derivados de la metilación del ADN y la modificación de histonas en humanos [21].

Por otro lado, a nivel de interconexión de bases de datos, se cuenta en la actualidad, con aportaciones cada vez más importantes. La iniciativa más significativa en este sentido comenzó en el año 2008 con el repositorio denominado BioGRID (Biological General Repository for Interaction Datasets) para integrar en especies modelos proteínas e interacciones genéticas incluyendo ADN y ARN [22]. En su última actualización de 2017, contenía 1072173 interacciones genéticas y de proteínas, además de 38559 modificaciones postraduccionales anotadas provenientes de un total de 48114 publicaciones [23]. Más recientemente, Nuño Cabete et al. [24], han desarrollado una base de datos en levaduras que integra diferentes aproximaciones ómicas a nivel transcripcional, postranscripcional y traduccional.

El proyecto ENCODE

El Proyecto ENCODE es un proyecto de investigación desarrollado por el Instituto Nacional de Investigación del Genoma Humano de Estados Unidos (NHGRI, National Human Genome Research Institute en inglés) desde el año 2003 [25]. Este proyecto se concibió como una continuación o aclaración del PGH para identificar todos los elementos funcionales en el genoma humano y analizar la expresión del gen molecular y su complejidad. Este proyecto trata de dar respuesta desde el ámbito de la experimentación a las dudas teóricas surgidas tras el PGH y la secuenciación completa de diferentes organismos vivos. El PGH evidenció cómo, a pesar de los cientos de miles de proteínas presentes en el ser humano, únicamente habían sido descritos poco más de 20000 genes que codificaban para proteínas. El objetivo principal del proyecto ENCODE era pues determinar el papel del resto de componentes del genoma [26].

El proyecto se llevó a cabo por un consorcio mundial de grupos de investigación, y los datos generados a partir del mismo, se pueden consultar en de bases de datos públicas. El proyecto está en continua actualización (su cuarta fase comenzó en el año 2017 y continua activa). Al entrar en la web del proyecto [27], se evidencia una gran gama de referencias bibliográficas derivadas de los numerosos experimentos llevados a cabo. Los principales hallazgos de ENCODE, derivan del cambio de perspectiva en el análisis de la herencia y expresión de caracteres en organismo vivos; donde el centro de gravedad de estos procesos se pone en el estudio del ARN más que en el del ADN [28], indicando también que la mayoría del ADN en el genoma humano y del resto de los organismos vivos, se transcribe en moléculas funcionales de ARN que no codifican directamente para proteínas. Además, los experimentos dentro del proyecto ENCODE, se basan en el uso innovador de las tecnologías de secuenciación masiva (o de alto rendimiento) tanto de ADN como de ARN. En total, ENCODE ha generado más de 15 billones de bytes de datos brutos [29]. Al unir cuidadosamente una gran variedad de datos, se observó que el genoma humano está vivo con interruptores que activan y desactivan los genes y controlan cuándo y dónde se producen los ARNs.

El transcriptoma completo de un ser vivo, está constituido por el conjunto de todos los ARNs. A partir del proyecto ENCODE, los primeros resultados experimentales demostraron cómo el ADN puede ser codificante y transcrito al ARN mensajero (el correspondiente a los

exones del gen molecular) y representa un 1.5 % (con un tamaño de cientos o miles de nucleótidos, nt) del total de la secuencia del genoma humano, además del ADN codificante y no expresado procedente de intrones (con tamaño de cientos de nt) que supone un 24.5%. Sin embargo, en su mayor parte, el ADN se expresa en ARN llamado no codificante porque no se traduce directamente a proteínas (denominado inicialmente ADN basura o junk DNA), que representa un 74% del total. Este ARN no codificante, puede ser de tipo estructural como el de transferencia (ARNt) (con un tamaño de 75-90nt) (5% del total) y el ribosómico (ARNr) [compuesto de varias subunidades de 121nt (subunidad 5S);156nt (subunidad 5.8S); 5070nt (subunidad 28S) y 1869nt (subunidad 18S)] que puede representar hasta un 66% del RNA total. A su vez, se incluye el ARN no codificante y regulador (3% del total) que incluye el grupo de los microARNs [pequeños ARN de interferencia (ARNsi) (de 20nt); micro ARN (miARN) (20-25nt); small nucleolar ARN (ARNsno) (de 60 a 300nt) o piwi-interacting ARN (piARNA)] y el grupo de los long non coding ARN (lncARN) [30] (figura 3A).

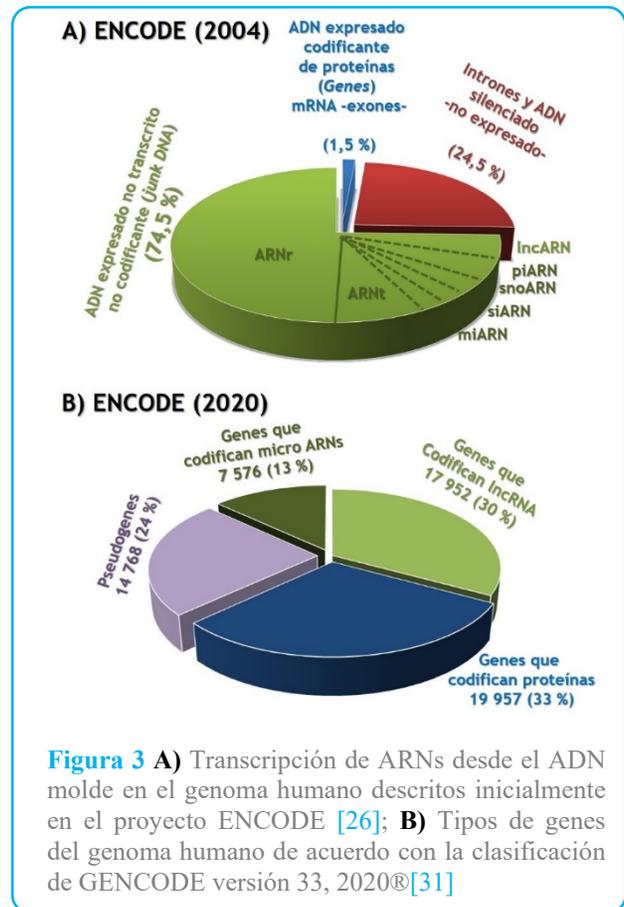


Figura 3 A) Transcripción de ARNs desde el ADN molde en el genoma humano descritos inicialmente en el proyecto ENCODE [26]; B) Tipos de genes del genoma humano de acuerdo con la clasificación de GENCODE versión 33, 2020@[31]

Esta caracterización inicial, realizada durante las primeras fases del proyecto ENCODE, dio lugar a una caracterización posterior mucho más detallada de los genes que conforman el genoma humano. De acuerdo con esta nueva clasificación, el genoma humano posee 60253 genes que incluyen a su vez, 19957 genes que codifican proteínas (lo que se corresponderían a genes moleculares de la genómica) (33%) además de 17952 genes que codifican lncARN (30%); 7576 genes que codifican microARNs (incluyendo miARN, sARN, siARN, snoARN o piARN) (13%), y 14768 pseudogenes que derivan de otros genes ya conocidos, cuyas funciones son distintas y pueden haber perdido su funcionalidad o haberla cambiado radicalmente (24%) (figura 3B) [31].

Nuevas evidencias experimentales de la Postgenómica relacionadas con el genoma

A partir de la nueva perspectiva de ENCODE y del uso de metodologías de análisis masivo, podemos describir una serie de nuevas evidencias experimentales de la postgenómica relacionadas con el genoma (figura 4; tabla 1). En primer lugar, se describen una serie de variaciones estructurales del ADN que afectan al concepto de gen en cuanto a unidad de herencia. Por un lado, las variaciones copia-número (copy-number variations) descritas experimentalmente por primera vez en 2005. Estas variaciones, afectan la expresión del genoma de una forma ajena a lo que sería el gen molecular, debido a que incluyen una serie de copias del gen que se expresan o no en los diferentes tejidos en un momento dado [32].

Además, otra importante evidencia experimental postgenómica consiste en los genes de fusión como modificaciones del ADN heredables al margen del gen molecular. Existen muchos ejemplos conocidos de fusiones de genes, así como los conocidos como genes "encriptados", lo que significa que los genes a menudo, se encuentran en partes que se pueden observar como segmentos separados alrededor del genoma. Distintas partes del mismo gen molecular, pueden ubicarse en diferentes cromosomas que al unirse forman estos genes de fusión procedentes de diversas regiones del genoma. Esta inestabilidad genómica es causa de enfermedades como el cáncer y en la mayoría de los casos se basa en la ruptura de dos cromosomas que intercambian fragmentos de ADN entre ellos, generando estos genes de fusión que disparan la multiplicación celular [33].

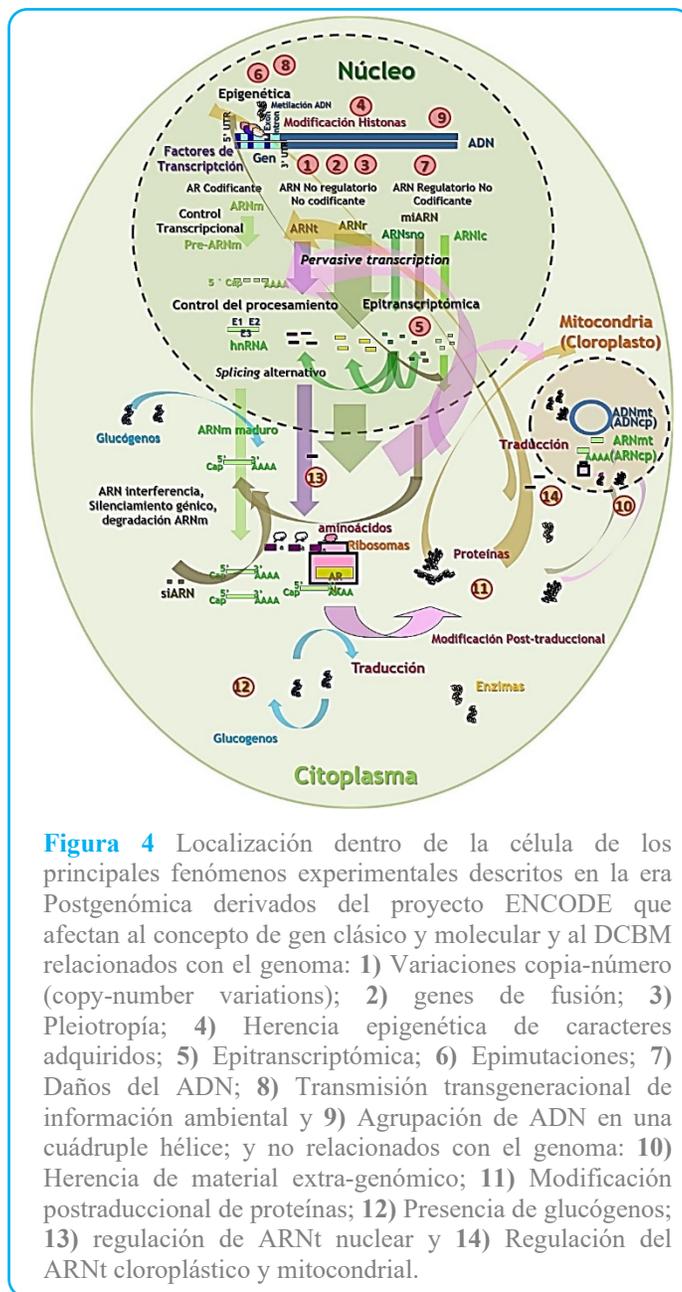


Figura 4 Localización dentro de la célula de los principales fenómenos experimentales descritos en la era Postgenómica derivados del proyecto ENCODE que afectan al concepto de gen clásico y molecular y al DCBM relacionados con el genoma: 1) Variaciones copia-número (copy-number variations); 2) genes de fusión; 3) Pleiotropía; 4) Herencia epigenética de caracteres adquiridos; 5) Epitranscriptómica; 6) Epimutaciones; 7) Daños del ADN; 8) Transmisión transgeneracional de información ambiental y 9) Agrupación de ADN en una cuádruple hélice; y no relacionados con el genoma: 10) Herencia de material extra-genómico; 11) Modificación postraduccional de proteínas; 12) Presencia de glucógenos; 13) regulación de ARNt nuclear y 14) Regulación del ARNt cloroplástico y mitocondrial.

Por otro lado, cabe destacar también los cambios pleiotrópicos mediante los cuales, un solo gen molecular es responsable de efectos fenotípicos en caracteres distintos y no relacionados que se pueden heredar de manera no mendeliana de una generación a otra; cuestionando el concepto de gen molecular y también la unidireccionalidad del flujo de información genética establecido en el DCBM [32].

Tabla 1 Fenómenos experimentales relacionadas y no relacionadas con el genoma evidenciados en la postgenómica que cuestionan el concepto de gen y el DCBM

Fenómeno experimental posgenómico	Referencia
Fenómenos relacionados con el genoma	
Variaciones copia-número	[31]
Genes de fusión	[32]
Pleiotropía	[33]
Herencia epigenética de caracteres adquiridos	[33]
Epitranscriptómica	[36]
Epimutaciones	[40]
Transmisión transgeneracional de información ambiental	[41]
Daños del ADN	[42]
Agrupación del ADN en una cuádruple hélice	[43]
Fenómenos no relacionados con el genoma	
Herencia mediada por material extra-genómico	[45]
Modificación postraduccional de proteínas	[46]
Presencia de glucógenos	[49]
Regulación de ARNt nuclear	[50]
Regulación de ARNt cloroplástico	[51]
Regulación de ARNt mitocondrial	[52]

Pero sin duda, los diferentes mecanismos de regulación epigenética, son los fenómenos experimentales relacionados con el genoma más estudiados y que han tenido una mayor importancia en el desarrollo de la Postgenómica. El mecanismo de regulación genética más conocido es el de la modificación de las histonas. El ADN se enrolla dos veces alrededor de cada nucleosoma (formado por un tipo de proteínas llamado histonas) entrando en estrecho contacto con estas proteínas. Estas histonas son proteínas muy conservadas tanto en el reino animal como vegetal. Mediante la modificación de estas histonas (acetilación, metilación, fosforilación o ubiquitinación) se hace más o menos accesible el genoma. El código de histonas se utiliza para describir las modificaciones de las histonas y la estructura de la cromatina [34]. Otro mecanismo alternativo epigenético consiste en la metilación del ADN. Esta metilación es una modificación de naturaleza química, también heredable, que se produce mediante la adición de grupos metilos (un átomo de carbono y tres de hidrógeno, CH₃). Esta metilación modifica (silencia) la función del gen molecular sin alterar su secuencia. La mayoría de las metilaciones se producen en la base citosina creando una

5-metil citosina, y suelen reducir la expresión génica. Los patrones de metilación del ADN son característicos de los diferentes tipos de células y pueden variar con el tiempo [35].

Más recientemente se ha descrito la regulación epigenética del ARN, lo que se denomina epitranscriptómica [36]. Esta regulación química del ARN y no mediada por el ADN, está conservada y se puede heredar modulando el flujo de información genética a nivel de estabilidad, degradación y empalme (*splicing*) del ARN. El mecanismo de acción descrito, es el de la metilación de la adenina del tipo m⁶A. Si bien la epitranscriptómica se describió inicialmente en ARNm [37], con posterioridad, se ha descrito también en la regulación del ARN (*long coding*) [38]. Más recientemente, se han desarrollado protocolos de análisis masivo de estas modificaciones epigenéticas del ARN a través de micromatrices [39]. Zaccara et al. [40], también introdujeron el término epiproteoma para definir las proteínas fruto de esta regulación epigenética del ARN. Además, las epimutaciones, o mutaciones de bases de ADN que están modificadas epigenéticamente mediante metilación, descritas por Jiang et al. [41], son efectos moleculares que también cuestionan el concepto de gen molecular dentro de los fenómenos epigenéticos. Finalmente, se ha descrito la herencia epigenética transgeneracional de caracteres adquiridos, donde el efecto del medio se puede transmitir a una generación posterior. Estos cambios epigenéticos son reversibles y, a diferencia de las mutaciones, se desvanecen gradualmente del genoma durante el curso de las generaciones y, por lo tanto, no se heredan infinitamente [42].

Otras evidencias postgenómicas relacionadas con el genoma, son los daños en el ADN producidos por fenómenos químicos en el núcleo de la célula que también cuestionan al gen molecular como unidad de herencia [43]. Además, es necesario incluir el fenómeno de agrupación del ADN descrito en células humanas que forma una cuádruple hélice de ADN en forma retorcida tras la unión de citosinas de dos hebras diferentes de ADN bautizado como *i-motif*. Estos nudos de ADN parecen ser estructuras transitorias que afectan a la expresión del ADN [44]. Esta estructura cuádruple del ADN, cuestiona la estructura de doble hélice propuesta como única y universal para la conformación del ADN de los seres vivos a la vez que cuestiona el DCBM, al establecer nuevas pautas de expresión génica ajenas a cualquier enzima.

Nuevas evidencias experimentales de la Postgenómica no relacionadas con el genoma

Por otro lado, a partir de la nueva perspectiva de ENCODE y del uso de metodologías de análisis masivo, se pueden elucidar una serie de nuevas evidencias experimentales de la postgenómica no relacionadas con el genoma (tabla 1; figura 4). En primer lugar, cabe anotar que no todo el material genético se encuentra en el núcleo de la célula. Fuera del núcleo, los animales y las plantas (organismos eucariotas) contienen en sus células, docenas de orgánulos denominados mitocondrias (unas 100 en el ser humano) cada una de las cuales, alberga unas cuantas copias (5 en el caso de humanos) de su propio genoma. Además, en el caso de las plantas, estos orgánulos independientes que portan información genética, son los cloroplastos. Desde la década de 1920, se sabe de la existencia de estos materiales de ADN extranucleares que se heredan de una manera no mendeliana, conocimiento que contribuyó en los años 60 a la Teoría Endosimbiótica Seriada (Serial Endosymbiosis Theory en inglés) desarrollada por la bióloga norteamericana Lyn Margulis [45]. Sin embargo, la influencia de este material extranuclear en la herencia, ha sido descrita a nivel experimental y molecular mucho más recientemente. Lolle et al. [46], en la especie vegetal *Arabidopsis thaliana*, señalaron a este material extranuclear como responsable de muchas características de esta especie vegetal. A partir de esta fecha, numerosos casos experimentales han demostrado el efecto de este ADN extranuclear en la herencia; bien de forma directa o bien interactuando con otros ARNs provenientes del ADN nuclear.

Por otro lado, las proteínas son los actores principales del funcionamiento y la dinámica celular. Sus propiedades físicas y químicas determinan las actividades específicas que pueden llevar a cabo dentro de la célula. La diversidad de formas de estas proteínas es mucho mayor que la directamente codificada en el genoma. Esta gran variabilidad se debe a diferentes mecanismos como el mencionado empalme alternativo del ARN mensajero durante la transcripción. Además, tiene mucha importancia un fenómeno más recientemente descrito consistente en las modificaciones postraduccionales de estas proteínas que modifican su estructura y efecto [47]. Si bien estas modificaciones pueden estar catalizadas por enzimas [48], en muchas ocasiones, estas modificaciones ocurren mediante reacciones químicas a nivel celular no

relacionadas con el genoma que pueden transmitirse a la descendencia. Más recientemente, han sido observados en la postgenómica, otros efectos moleculares como la influencia del contenido celular de glucógenos en la transmisión de caracteres en humanos en la herencia. Estas moléculas de reserva presentes en las células germinales, se transmiten a la descendencia y tienen una influencia en la herencia de los caracteres al margen del gen molecular [48].

Finalmente, se han descrito las regulaciones de ARNt nuclear [49], cloroplástico [50] y mitocondrial [51], como mecanismos que afectan a la eficiencia en la traducción del ARNm y no está relacionado con el genoma. Se producen fenómenos de demetilación e hidrólisis de las cadenas de ARN del ARNt, que terminan afectando su función en la expresión del ARNm y en la estabilidad de las funciones celulares. Diferentes moléculas del citoplasma de la célula, pueden afectar al ARNt actuando como moléculas de señalización y regulación de la transcripción en seres vivos al margen del gen molecular. Estos efectos moleculares también se transmiten a la descendencia.

Conclusión

La práctica experimental en la posgenómica, es una actividad cada vez más compleja con múltiples variables en cuanto al objeto de estudio genético, la aplicación de metodologías y la organización de los laboratorios interconectados con plataformas tecnológicas. El investigador utiliza la observación y medición de los resultados del experimento a través de plataformas tecnológicas de secuenciación masiva de ADN y ARN como instrumento básico para el logro empírico de los objetivos científicos planteados. Esta observación en postgenómica, debe entenderse como un hecho codificado y dinámico producido en unos instrumentos basados en complicadas interacciones físico-químicas. Las nuevas metodologías de análisis masivo producen millones de datos depositados en inmensas bases de datos, en lo que algunos autores denominan una nueva ciencia de Big Data, que es necesario analizar y utilizar en la interpretación de los fenómenos observados mediante el uso de nuevos algoritmos matemáticos para su análisis. La medición de datos en la Posgenómica debe realizarse como un flujo de datos computacionales que únicamente se pueden analizar a través de sofisticadas herramientas bioinformáticas específicamente desarrolladas para este fin en la nueva disciplina científica de la bioinformática. Esta nueva era es más pluralista al comprometerse con

una concepción del gen postgenómico mucho más plural y diverso. Además, la variación del material hereditario puede ocurrir no solo como consecuencia de la recombinación genética y las mutaciones, sino que otros fenómenos biológicos también producen variabilidad y son heredables, incluyendo la epigenética, las mutaciones adaptativas o los glucógenos en un nuevo contexto de interacción del medio con el gen postgenómico.

Consentimiento de publicación

El autor leyó y aprobó el manuscrito final.

Conflicto de interés

El autor declara no tener conflicto de interés. Este documento solo refleja sus puntos de vista y no el de la institución a la que pertenece.

Perfil de autoría

Pedro Martínez Gómez

Master of Science en Mejora Genética por el IAMZ de Zaragoza, España y Dr. Ingeniero Agrónomo por la Universidad de Murcia, España; estancia Doctoral en la Universidad de California-Davis, EEUU, jefe del Departamento de Mejora Vegetal, Grupo de Mejora Genética de Frutales del Centro de Edafología y Biología Aplicada (CEBAS), Consejo Superior de Investigaciones Científicas (CSIC)

en Espinardo-Murcia, España. Editor Jefe de la revista *Scientia Horticulturae* (revista Q1 de Elsevier) y miembro del Editorial Board of *Plant Article Transfer Group* de Elsevier.

Sus temas de interés comprenden la genética, genómica y la transcriptómica de *Prunus*. Sus líneas de investigación abarcan la mejora genética del albaricoquero y del almendro; la resistencia a virus en frutales del género *Prunus*; los marcadores moleculares aplicados a la mejora genética en frutales del género *Prunus*; el letargo invernal y época de floración en frutales del género *Prunus* y la calidad del fruto en frutales del género *Prunus*. Es coobtentor de 5 variedades de albaricoquero, 2 de almendro y 1 de ciruelo. Además, ha publicado 140 trabajos en revistas científicas del SCI, director de 12 tesis doctorales, 13 TFG y TFM. En agosto de 2006, recibió en Seúl el Premio "Miklos Faust International Award for Young Pomologists", convocado por la ASHS y la ISHS; en febrero de 2012, recibió en Teherán, el Premio "Khwarizmi International Award" y la "Medalla de la FAO" por su contribución a la investigación agraria en países en vías de desarrollo. En enero de 2019, recibió en La Habana-Cuba, el Premio CITMA 2018 por su colaboración en la conservación y explotación de recursos fitogenéticos de Cuba.



Referencias

- [1] Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* 1975;94:441–8. [https://doi.org/10.1016/0022-2836\(75\)90213-2](https://doi.org/10.1016/0022-2836(75)90213-2)
- [2] García-Sancho M. A New Insight into Sanger's Development of Sequencing: From Proteins to DNA, 1943–1977. *Journal of the History of Biology* 2010;43:265–323. <https://doi.org/10.1007/s10739-009-9184-1>
- [3] National Human Genome Research Institute (NIH). MINC Toolkit. *The Forefront of Genomics* 2022. <https://www.genome.gov/minc>.
- [4] Brown TA. *Genomes 4*. Other Titles: *Genomes | Genomes Four Description: 4th.* | New York, NY : Garland Science, [2017] | Preceded By: Garland Science; 2018. <https://doi.org/10.1201/9781315226828>.
- [5] Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* 1996;242:84–9. <https://doi.org/10.1006/abio.1996.0432>
- [6] Illumina. Powering the heroes on the front lines with sequencing solutions to address a pandemic. *Illumina Innovation Technologies* 2022. <https://www.illumina.com/>.
- [7] Thermo Fisher Scientific. Featured education & support 2022. <https://www.thermofisher.com/co/en/home.html>.
- [8] PacBio. Sequence with confidence. Your Partner in an Evolving Pandemic 2022. <https://www.pacb.com/>.
- [9] MGI. High speed, high flexibility, and ultra-high throughput turbocharge your sequencing. *DNBSEQ-T7 Brochure* 2022. https://en.mgi-tech.com/products/instruments_info/5/.
- [10] Callebaut W. Scientific perspectivism: A philosopher of science's response to the challenge of big data biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 2012;43:69–80. <https://doi.org/10.1016/j.shpsc.2011.10.007>
- [11] Selzer PM, Marhöfer RJ, Koch O. *Applied Bioinformatics*. Cham: Springer International Publishing; 2018. <https://doi.org/10.1007/978-3-319-68301-0>.
- [12] National Library of Medicine (NIH). Welcome to NCBI. *National Library of Medicine (NIH)* 2022. <https://www.ncbi.nlm.nih.gov/>
- [13] Gold JGI. *Genomes Online Databases*. *Genomes Online Databases* 2022. <https://gold.jgi.doe.gov/simplesrch>
- [14] KEGG. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Kyoto Encyclopedia of Genes and Genomes* 2022. <https://www.genome.jp/kegg/>
- [15] European Genome Phenome Archive (EGA). EGA Consortium. What Is in the EGA? *Studies in the EGA by Diseases* 2022. <https://ega-archive.org/>
- [16] Leung YY, Kuksa PP, Amlie-Wolf A, Valladares O, Ungar LH, Kannan S, et al. DASHR: database of small human noncoding RNAs. *Nucleic Acids Research* 2016;44:D216–22. <https://doi.org/10.1093/nar/gkv1188>
- [17] Human Cell Atlas. *Human Cell Atlas*. About Human Cell Atlas 2022. <https://www.humancellatlas.org/>

- [18] Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. The Human Cell Atlas: from vision to reality. *Nature* 2017;550:451–3. <https://doi.org/10.1038/550451a>
- [19] National Cancer Institute (NIH). Genomic data commons. The NCI's Genomic Data Commons (GDC) 2022. <https://gdc.cancer.gov/>
- [20] Ríos Insuna D, Gómez Ullate Oteiza D. Big data. Conceptos, tecnologías y aplicaciones. vol. 1. 1st ed. Madrid, España: Consejo Superior de Investigaciones Científicas : Los Libros de la Catarata; 2019.
- [21] Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein Science* 2018;27:233–44. <https://doi.org/10.1002/pro.3307>
- [22] International Human Epigenetic Consortium (IHEC). Sharing epigenomes globally. *Nature Methods* 2018;15:151–151. <https://doi.org/10.1038/nmeth.4630>
- [23] Stark C, Breitkreutz B-J, Chatr-aryamontri A, Boucher L, Oughtred R, Livstone MS, et al. The BioGRID Interaction Database: 2011 update. *Nucleic Acids Research* 2011;39:D698–704. <https://doi.org/10.1093/nar/gkq1116>
- [24] Chatr-aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Research* 2017;45:D369–79. <https://doi.org/10.1093/nar/gkw1102>
- [25] Nuño-Cabanes C, Ugidos M, Tarazona S, Martín-Expósito M, Ferrer A, Rodríguez-Navarro S, et al. A multi-omics dataset of heat-shock response in the yeast RNA binding protein Mip6. *Scientific Data* 2020;7:69. <https://doi.org/10.1038/s41597-020-0412-z>
- [26] ENCODE. The ENCODE (Encyclopedia of DNA Elements) project. *Science* (1979) 2004;306:636–40. <https://doi.org/10.1126/science.1105136>
- [27] ENCODE. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74. <https://doi.org/10.1038/nature11247>
- [28] ENCODE. ENCODE Portal. ENCODE Stanford University 2022. <https://www.encodeproject.org/>
- [29] GENCODE. Statistics about the current GENCODE Release (version 40). GENCODE Human 2022. <https://www.genecodegenes.org/human/stats.html>
- [30] Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, et al. Fine-scale structural variation of the human genome. *Nature Genetics* 2005;37:727–32. <https://doi.org/10.1038/ng1562>
- [31] Akiva P, Toporik A, Edelheit S, Peretz Y, Diber A, Shemesh R, et al. Transcription-mediated gene fusion in the human genome. *Genome Research* 2006;16:30–6. <https://doi.org/10.1101/gr.4137606>
- [32] Nadeau JH, Topol EJ. The genetics of health. *Nature Genetics* 2006;38:1095–8. <https://doi.org/10.1038/ng1006-1095>
- [33] Amaral PP, Dinger ME, Mercer TR, Mattick JS. The Eukaryotic Genome as an RNA Machine. *Science* (1979) 2008;319:1787–9. <https://doi.org/10.1126/science.1155472>
- [34] Riddihough G, Zahn LM. What Is Epigenetics? *Science* (1979) 2010;330:611–611. <https://doi.org/10.1126/science.330.6004.611>
- [35] Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE. The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biology* 2012;13:175. <https://doi.org/10.1186/gb-2012-13-10-175>
- [36] Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* 2015;518:560–4. <https://doi.org/10.1038/nature14234>
- [37] Li X, Xiong X, Yi C. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nature Methods* 2017;14:23–31. <https://doi.org/10.1038/nmeth.4110>
- [38] Zaccara S, Ries RJ, Jaffrey SR. Reading, writing and erasing mRNA methylation. *Nature Reviews Molecular Cell Biology* 2019;20:608–24. <https://doi.org/10.1038/s41580-019-0168-5>
- [39] Jiang C, Mithani A, Belfield EJ, Mott R, Hurst LD, Harberd NP. Environmentally responsive genome-wide accumulation of de novo Arabidopsis thaliana mutations and epimutations. *Genome Research* 2014;24:1821–9. <https://doi.org/10.1101/gr.177659.114>
- [40] Klosin A, Casas E, Hidalgo-Carcedo C, Vavouri T, Lehner B. Transgenerational transmission of environmental information in *C. elegans*. *Science* (1979) 2017;356:320–3. <https://doi.org/10.1126/science.aah6412>
- [41] Chen L, Liu P, Evans TC, Ettwiller LM. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science* (1979) 2017;355:752–6. <https://doi.org/10.1126/science.aai8690>
- [42] Zeraati M, Langley DB, Schofield P, Moye AL, Rouet R, Hughes WE, et al. I-motif DNA structures are formed in the nuclei of human cells. *Nat Chem* 2018;10:631–7. <https://doi.org/10.1038/s41557-018-0046-3>
- [43] Schwartz W. Lynn Margulis, Origin of eukaryotic cells. evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth. XXII u. 349 S., 89 Abb., 49 Tab. New Haven-London 1970: Yale University. Zeitschrift Für Allgemeine Mikrobiologie 1973;13:186–186. <https://doi.org/10.1002/jobm.19730130220>
- [44] Lolle SJ, Victor JL, Young JM, Pruitt RE. Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis. *Nature* 2005;434:505–9. <https://doi.org/10.1038/nature03380>
- [45] Walsh CT, Garneau-Tsodikova S, Gatto GJ. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angewandte Chemie International Edition* 2005;44:7342–72. <https://doi.org/10.1002/anie.200501023>
- [46] Yang A, Cho K, Park H-S. Chemical biology approaches for studying posttranslational modifications. *RNA Biology* 2018;15:427–40. <https://doi.org/10.1080/15476286.2017.1360468>
- [47] Harmel R, Fiedler D. Features and regulation of non-enzymatic post-translational modifications. *Nature Chemical Biology* 2018;14:244–52. <https://doi.org/10.1038/nchembio.2575>
- [48] Wang S-CM, Muscat GEO. Nuclear receptors and epigenetic signaling: Novel regulators of glycogen metabolism in skeletal muscle. *IUBMB Life* 2013;65:657–64. <https://doi.org/10.1002/iub.1181>
- [49] Agris PF, Narendran A, Sarachan K, Väre VYP, Erusyal E. The Importance of Being Modified, 2017, p. 1–50. <https://doi.org/10.1016/bs.enz.2017.03.005>
- [50] Mohanta TK, Khan AL, Hashem A, Allah EFA, Yadav D, Al-Harrasi A. Genomic and evolutionary aspects of chloroplast tRNA in monocot plants. *BMC Plant Biology* 2019;19:39. <https://doi.org/10.1186/s12870-018-1625-6>
- [51] Ribas de Pouplana L. The mitochondrial tRNA conundrum. *Nature Reviews Molecular Cell Biology* 2020;21:361–361. <https://doi.org/10.1038/s41580-020-0220-5>