

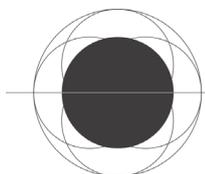


Genómica estructural y Genómica funcional

Susana Hospital Marchal

Contenido

Introducción	4
Genética y Genómica.....	6
Genómica Estructural.....	10
Conceptos	10
Gen	10
Genoma.....	10
Fenotipo y genotipo.....	11
Cromosomas y cromatina.....	12
Ácidos nucleicos	17
ADN.....	18
ARN.....	28
Código genético.....	31
El lenguaje del genoma humano.....	33
Plegamiento del ADN y elementos estructurales adicionales (epigenéticos)	36
Mutaciones y polimorfismos.....	36
Genómica Funcional	42
El dogma central de la Biología Molecular.....	42
Transcripción.....	43
Traducción.....	46



Replicación.....	49
Recombinación	51
Regulación génica	53
Regulación pretranscripcional.....	54
Regulación transcripcional.....	58
Regulación postranscripcional.....	61
Regulación en la replicación	61
Regulación a través de receptores nucleares.....	62
El patrón de la herencia genética.....	63
Epigenética	66
Abreviaciones	68
Bibliografía.....	69
Anexo 1. Ciencias ómicas.....	71
Anexo 2. Breve glosario de términos.....	72
Anexo 3. Nomenclatura.....	73
Anexo 4. Animaciones y ejercicios.....	77

Introducción

¿Cómo la Genómica puede ayudar a mejorar la salud de las personas?

Nos encontramos delante de un nuevo escenario que plantea un nuevo posicionamiento de los profesionales de la salud que atendemos día a día a los pacientes.

Normalmente, la práctica habitual de la medicina establece una serie de enfermedades y tratamientos a partir de los síntomas. Cuando ya han aparecido los síntomas. El centro es la enfermedad, su diagnóstico y tratamiento, en base a protocolos uniformes.

Para los tratamientos, después de observar la respuesta del paciente mediante pruebas ensayo-error, en ocasiones se acierta, pero en otras, el fracaso terapéutico inicia un nuevo tratamiento hasta así acertar con el adecuado. Además, pueden aparecer efectos adversos secundarios que coexisten con los tratamientos, interacciones farmacológicas cuando existe polimedicación, disminuyendo o exacerbando el efecto terapéutico buscado, o con una respuesta a fármacos deficiente según los hábitos del paciente y su interacción con el medio ambiente. Y todo ello responde a una variabilidad interindividual para la misma enfermedad y tratamientos.

Así, ya en la Antigua Grecia Hipócrates sugirió *“Es más importante saber qué tipo de persona tiene una enfermedad que el tipo de enfermedad que tiene una persona”*.

Los avances tecnológicos y científicos desde entonces, de manera casi artesanal en el siglo XIX, persistente y testaruda en el siglo XX, y brutalmente exponencial en las últimas décadas gracias a los avances de la biotecnología, están haciendo posible que esa afirmación/intuición de la filosofía clásica sea una realidad ya en nuestros días.

A principios del siglo XX, el médico inglés Archibald Garrod sugirió que las personas eran diferentes, no solamente por sus características físicas externas, su apariencia, sino también a nivel bioquímico.

En 1999 se acuñó el término de Medicina Personalizada por dos periodistas del Wall Street Journal y desde 2008 se empezó a utilizar el término Medicina de Precisión. No habiéramos llegado a estos términos sin el descubrimiento de la

estructura del ADN, nuestro material genético, aquello de lo que están formados nuestros genes, y el orden en que están dispuestos, tras completar el estudio de la secuencia del genoma humano el año 2003. Antes, muchos investigadores han contribuido con sus trabajos a construir este conocimiento que, junto con avances biotecnológicos importantes, nos plantean este nuevo escenario hacia la Medicina de Precisión que permite una medicina más segura, más eficiente, preventiva y predictiva. Siendo las personas distintas, se plantea una medicina individualizada o, como mejor gusta describirla a algunos autores, tal como veremos, una medicina estratificada, puesto que responderá a grupos de personas con características equivalentes. Y dónde la persona tomará un papel central, preocupándose activamente de su salud. Llegamos entonces a la Medicina de las 4 “P”, Medicina Personalizada (Estratificada), Preventiva, Predictiva y Participativa. Porque hablaremos de participantes (que no de pacientes, porque hay personas sanas y enfermas).

Y la Genómica está en la base de todo ello, el conocimiento de nuestro material genético, a partir del cual construir la arquitectura de nuestro funcionamiento molecular. Serán necesarios también datos de la Transcriptómica, de la Proteómica, de la Metabolómica, de la Epigenómica, y otras ciencias ómicas que completarán la Genómica. En el **Anexo 1** se indican brevemente la contribución de las diferentes ciencias ómicas. Y todos los individuos, sanos y enfermos, proporcionarán datos genéticos, clínicos, ambientales y de comportamiento que, gracias a la informática médica, porque se generarán gran cantidad de datos, nos permitirá avanzar en la Medicina de Precisión.

Según el Profesor Ramón Cacabelos, “la Genómica ha venido a revolucionar conceptos inmovilistas de la biología y dogmas médicos que no resisten el impacto de la modernidad”

La verdadera Medicina Preventiva se origina y se originará en la Medicina de Precisión, en contraposición a la Medicina Tradicional en la que la única práctica preventiva se refiere a hacer chequeos de salud que son en realidad diagnósticos precoces. Y con la Genómica se puede, se podrá, dar diagnósticos antes de que aparezcan los síntomas en base al genotipo, y mejorar los tratamientos de lo que ya sabemos sobre farmacogenética.

También conocemos que los factores ambientales, nuestra dieta y hábitos influyen en nuestra salud. La genética en muchas ocasiones, como veremos, no tendrá la última palabra, pero sí nos orientará en saber cómo estos factores externos nos pueden ayudar, y cuáles de éstos pueden ejercer un efecto negativo

atendiendo a potenciales vulnerabilidades escritas en nuestra genética. Para ello, las dianas moleculares deberán estar bien estudiadas y definidas, de lo que nos ocuparemos en este capítulo y los siguientes de la Genómica.

Y la farmacéutica, el farmacéutico, desde el día a día en la oficina de farmacia, a pie de calle, podrá sumarse a este cambio desde el conocimiento, formándonos en estos ámbitos de la biología molecular que explican la causa de la enfermedad, que afinan tratamientos, que fundamentan comportamientos saludables en la vida de las personas y que mejoran, en definitiva, su salud y calidad de vida.

Ahora es el turno de la Genómica. Como si de un edificio se tratara, se trata de construir los cimientos.

Genética y Genómica

Y la “instrumental” de la Genómica es la Genética.

La **Genética** se refiere al estudio de los **genes individuales** y su papel en la herencia. Se centra en las características o rasgos que se transmiten de padres a hijos, de una generación a otra. En cambio, la **Genómica** estudia el **material genético completo**, cómo interactúan los genes entre ellos, qué consecuencias bioquímicas tienen para la célula o el organismo completo el efecto conjunto de los genes. La Genética nace 40 o 50 años y está sólidamente establecida. En cambio, la Genómica es más reciente, se han podido observar ciertos efectos y queda todavía mucho camino por recorrer.

Los primeros estudios que investigaron el material genético de la vida empezaron con Gregor Mendel, quien publicó los resultados de su investigación en 1865. A través del uso de guisantes en experimentos de cruzamiento, Mendel reportó la herencia de las características o rasgos ocurridos vía unidades (que más tarde fueron descritas como genes).

Así, enunció las leyes de la herencia, y el enunciado de la primera es:

“Al cruzar dos razas puras, la descendencia será heterocigótica y dominante”. Se representa en la Figura 1.

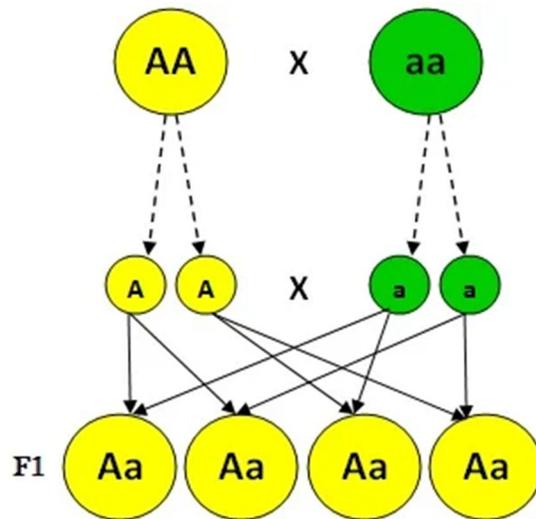
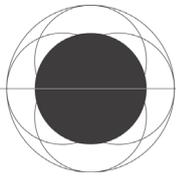


Figura 1. Primera ley de Mendel.

Fuente: Biobook

Y la segunda:

“Al cruzar dos razas híbridas, la descendencia será homocigótica e híbrida al 50%”.
Se representa en la Figura 2.

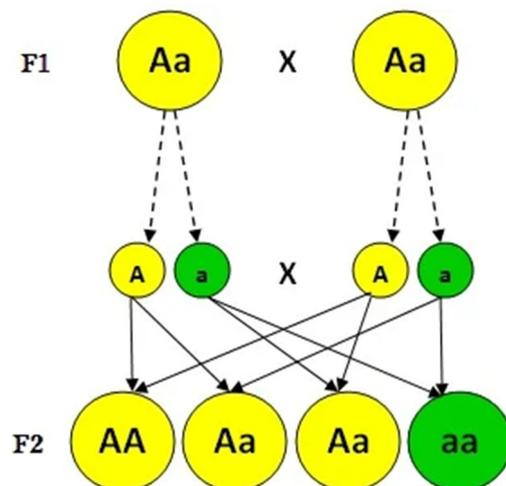


Figura 2. Segunda ley de Mendel.

Fuente: Biobook

Introdujo la idea de los rasgos dominantes y recesivos. Sus experimentos determinaron que hay unas partículas indivisibles de información que transmiten los rasgos hereditarios.

El año 1900 se redescubre la obra de Mendel y se crea la palabra gen. El avance científico que se ha llevado a cabo se puede dividir en cuatro fases principales¹ que corresponden aproximadamente a los cuatro cuartos de siglo. En la primera fase se determinó la estructura celular de la herencia: los cromosomas. En la segunda se descubrió la estructura molecular: la doble hélice de ADN (el año 1941 Avery demuestra que el ADN es el portador de la información genética y en el año 1953, Watson, Crick, Wilkins i Franklin descubren la estructura del ADN). En la tercera se encontró la estructura informativa (es decir, el código genético), con el descubrimiento mediante el cual las células leen la información contenida en los genes y, con la invención de distintas técnicas, se posibilitó la ingente tarea de descifrar la secuencia del genoma humano, señalando ahora sí, el punto de partida de la cuarta fase de la **genética**. Se trataba de la era de la **genómica**, con el análisis de genomas completos de cualquier organismo, también el del ser humano.

Cómo indica el Profesor Cacabelos², podemos estructurar el conocimiento del genoma humano, des de un punto de vista práctico, en la realización de una serie de fases:

A.- **Genómica estructural**: esta se ocupará de la estructura del genoma, saber cuántos genes hay, dónde se encuentran y cómo se organizan dentro del genoma.

B.- **Genómica funcional**: saber para qué sirven los genes, cómo se relacionan entre ellos, cómo funcionan y a qué da lugar su activación o desactivación (en esta fase toman cuerpo la transcriptómica, la proteómica y la metabolómica).

C.- **Patogenómica**: saber cómo se estropean los genes y cómo se altera su expresión dando lugar a enfermedades.

D.- **Toxicogenómica**: saber cómo interacciona el genoma con el medio ambiente y entender los efectos que los tóxicos medioambientales pueden ejercer sobre el genoma, así como los mecanismos que el genoma tiene para defender al organismo frente a agresiones externas e internas.

E.- **Genómica terapéutica**: saber cómo podemos reparar los defectos de nuestro genoma para prevenir enfermedades, corregir disfunciones o tratar específicamente diferentes patologías que afecten al ser humano. En esta fase de

desarrollo del conocimiento genómico caben la terapia génica, la farmacogenómica, la nutrigenómica.

Apasionante el campo de trabajo que se abre con el desarrollo de la Genómica. Ahora bien, es un desafío impresionante y que se encuentra en pleno desarrollo sin completar. Sabemos un poco de la estructura del genoma, muy poco de su función y aún menos del resto. Ahora bien, no hay demora para adentrarnos en lo que sabemos, y no hay excusa para que la medicina y la salud de las personas se beneficien de ello.

Genómica estructural

Entramos en la parte de la Genómica que estudia la estructura del genoma, los genes, donde se encuentran y cómo se organizan dentro del genoma.

Conceptos

Gen

Como pilar de la genómica estructural, definiremos que es un **gen**. Es la unidad básica de la herencia³. Es una secuencia ordenada de nucleótidos en la molécula de ADN que contiene la información necesaria para la síntesis de una macromolécula con función celular específica, como las proteínas. Los genes se transmiten de los padres a la descendencia y contienen la información necesaria para precisar sus rasgos. Los genes están dispuestos, uno tras otro, en estructuras llamadas cromosomas. Un cromosoma contiene una única molécula de ADN, sólo una parte de la cual corresponde a un gen individual, como veremos. Los seres humanos tenemos entre 20.000 genes y 27.000 genes organizados en nuestros cromosomas que van a codificar más de 100.000 proteínas. El gen codificador de proteínas humanas promedio tiene aproximadamente 3000 letras de largo, pero nuestros genes tienen una amplia gama de tamaños. El más corto tiene solo 500 letras y el más largo tiene 2,3 millones.

Genoma

Es el conjunto de instrucciones genéticas que se encuentran en la célula. Es TODO el material genético que, a nivel molecular, se concreta en la molécula de ADN, como veremos. Todos los organismos vivos tienen su propio genoma. Cada genoma contiene la información necesaria para construir y mantener ese organismo a lo largo de toda su vida. En los seres humanos, el **genoma** consiste en 46 cromosomas agrupados en 23 pares, que se encuentran en el núcleo, así como un pequeño cromosoma circular que se encuentra en las mitocondrias. En los cromosomas se localizan los genes. Ahora bien, los genes no son el genoma. De hecho, es una pequeñísima parte, el 1%. Del resto conocemos una parte que, aunque sin llegar a conocerlo completamente, sí nos permite poder aplicar aquello que sabemos.

Fenotipo y genotipo

Cuando hablamos de rasgos observables de un individuo, tales como altura, color de los ojos, el grupo sanguíneo, otros datos analíticos, nos referimos al **fenotipo** del individuo. También a la presencia o ausencia de una enfermedad. Los fenotipos están influenciados por factores genéticos y por factores ambientales.

Cuando hablamos de **genotipo** nos referimos a la colección de genes de un individuo. El genotipo es la versión de la información genética que un individuo tiene.

Algunos rasgos observables (fenotipo) son determinados en gran medida por el genotipo, mientras que otros están determinados en gran medida por factores ambientales, como lo que uno come, si se hace ejercicio, si se fuma, etc. Es más, también puede influir el azar en el fenotipo, así como los estímulos a los que ha estado expuesto el individuo¹.

Podríamos resumir estos conceptos con la fórmula:

$$\text{genotipo} + \text{medio} + \text{estímulos} + \text{azar} = \text{fenotipo}$$

Como ejemplo de lo que hemos dicho, podemos indicar que en los seres humanos hay un gen denominado *BRCA1* que cuando está alterado incrementa el riesgo de padecer cáncer mamario. Pero no todas las mujeres que presentan este gen mutado desarrollan cáncer. Estos genes susceptibles a estímulos y al azar se dice que tienen una “penetrancia” parcial o incompleta; es decir, que es un gen que, a pesar de haber sido heredado, no tiene la capacidad de “penetrar” o incidir en una característica concreta. Este gen también se puede expresar de forma variable, es decir, a pesar de haberse heredado puede dar lugar a una variedad de cáncer más agresiva o no llegar a desarrollar la enfermedad. No sabemos qué provoca la diferencia de resultados entre estas situaciones, pero es una combinación de la edad, la exposición al sol, y la buena y mala suerte. No podemos tener en cuenta sólo el genotipo para predecir con certeza el resultado.

Cromosomas y cromatina

Un **cromosoma**³ es un paquete ordenado de material genético de estructura ADN que se encuentra en el núcleo de la célula. Los diferentes organismos tienen diferentes números de cromosomas. En el ser humano el ADN no es una sola molécula larga. Si fuera así, que no lo es, y la estiráramos, su longitud sería 2,2 metros. En realidad, se divide en una serie de segmentos longitudinales desiguales. En ciertos puntos del ciclo de la vida de una célula, esos segmentos pueden ser esos paquetes compactos conocidos como cromosomas. Durante estas etapas, los cromosomas parecen tener forma de X. Concretamente es en la división celular durante la mitosis o la meiosis. Como decíamos, los humanos tenemos 46 cromosomas agrupados en 23 pares (22 llamados autosómicos o autosomas, y un par de cromosomas sexuales, X e Y). Cada par de cromosomas ha sido heredado de cada uno de nuestros progenitores. El número de cromosomas, que se encuentran en pares homólogos (dotación diploide, $2n$) es constante para todas las células somáticas, mientras que las células germinales maduras poseen un sólo cromosoma de cada par (dotación haploide, n). Cada par de cromosomas homólogos posee características morfológicas parecidas y en ambos sus genes contienen información para los mismos caracteres, aunque no necesariamente la información será idéntica, ya que uno tiene origen materno, y otro, paterno.

Denominamos **cariotipo**³ a la colección de cromosomas de un individuo, como recoge la imagen de la Figura 3. Consiste en la disposición ordenada de los cromosomas según la forma y el tamaño, atendiendo a los criterios de mayor a menor. La designación de los cromosomas como “primero” o “último” es arbitraria; el primer cromosoma se le designa como primero porque es el más largo. El cromosoma 1 (el cromosoma humano más grande) tiene la mayor cantidad de genes (3.141) y el cromosoma Y la menor (231).

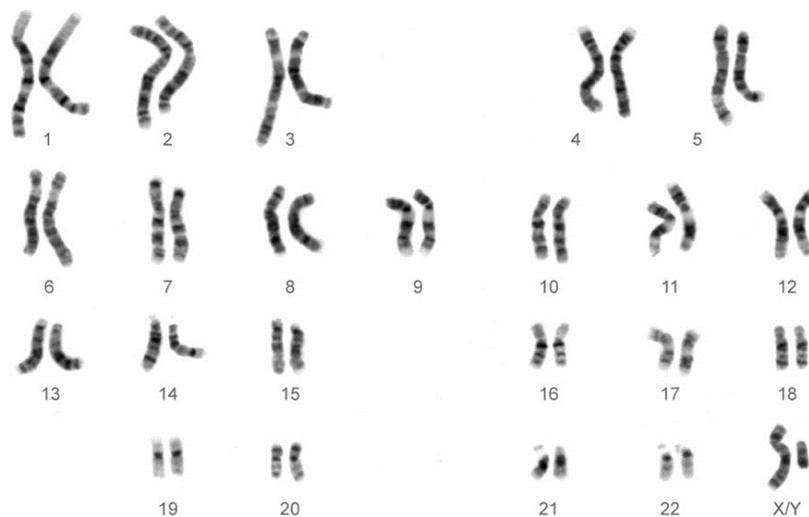


Figura 3. Colección de pares de cromosomas humanos³.

El término cariotipo también se refiere a una técnica de laboratorio que produce la imagen de los cromosomas de un individuo. Se utiliza para buscar números o estructuras anormales de los cromosomas. Es una de las técnicas más utilizadas para detectar anomalías cromosómicas, como el síndrome de Down. Un ejemplo de aplicación de la obtención del cariotipo sería las técnicas de diagnóstico precoz prenatal. Si los pares homólogos son iguales y su morfología se ciñe a la descrita, se concluye que la muestra se corresponde a un cariotipo normal.

Para “leer” un conjunto de cromosomas, los científicos utilizan tres características clave para identificar sus similitudes y diferencias:

- 1.- El tamaño.
- 2.- El patrón de bandas: El tamaño y ubicación de las bandas hace que cada cromosoma sea único.
- 3.- La posición del centrómero, región estrecha de un cromosoma que lo separa en un brazo corto (p) y un brazo largo (q).

En la Figura 4 se indican estas características.

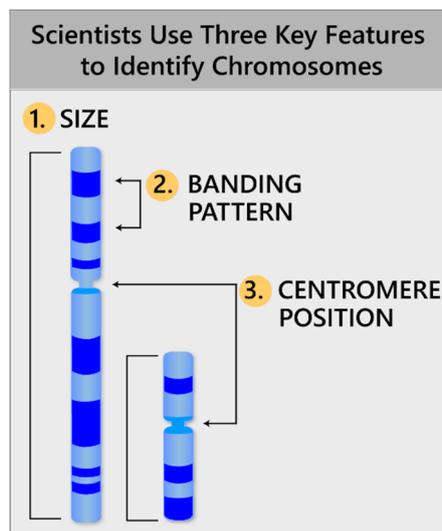


Figura 4. Características de los cromosomas para su descripción.

Fuente: Learn Genetics-Utah

Por ejemplo, 9q34.12 indica cromosoma 9, brazo largo, región 3, banda 4, subbanda 1-2.

El sitio web que nos da información sobre cada uno de los cromosomas es:

<https://medlineplus.gov/genetics/chromosome/>

Los cromosomas son estructuralmente muy sofisticados, conteniendo elementos necesarios para su funcionalidad. De hecho, semejante longitud lineal de ADN es como imposible que pueda caber en el interior del núcleo. Para ello, el ADN está superenrollado y se encuentra empaquetado gracias a unas proteínas llamadas histonas, las más abundantes, que, junto con el ADN forman la **cromatina** (o ADN empaquetado). El empaquetamiento permite condensar muchísimo el espacio que ocupa el material genético. Los cambios en la estructura de la cromatina se producen cuando el ADN se duplica y durante, lo que veremos, se denomina expresión génica. Así, se distinguen, a este respecto, dos tipos de cromatina; la eucromatina, que presenta un empaquetamiento menor, y la heterocromatina, que es la porción de cromatina más condensada.

Se sabía desde hacía tiempo de la **cromatina**, que estaba formada por dos tipos de sustancias químicas: proteínas y ácidos nucleicos. Nadie conocía ni

comprendía la estructura química de la cromatina, pero sí que estos dos componentes estaban íntimamente mezclados. Hoy esta estructura ya nos parece comprensible.

Y solíamos pensar que las histonas actuaban básicamente como maletas que guardaban y sostenían el ADN, pero está muy claro que las histonas están sometidas a regulación y tienen mucho que ver con la activación y desactivación de los genes, como veremos más adelante. Se puede pensar en ellas como maletas que están controladas y determinan cuándo se abre la maleta y sale un gen. Así que resulta que tienen funciones muy importantes, no sólo estructurales, sino también en la regulación de la función del gen por su expresión. En la Figura 5 se dibujan estas estructuras “desplegadas”³.

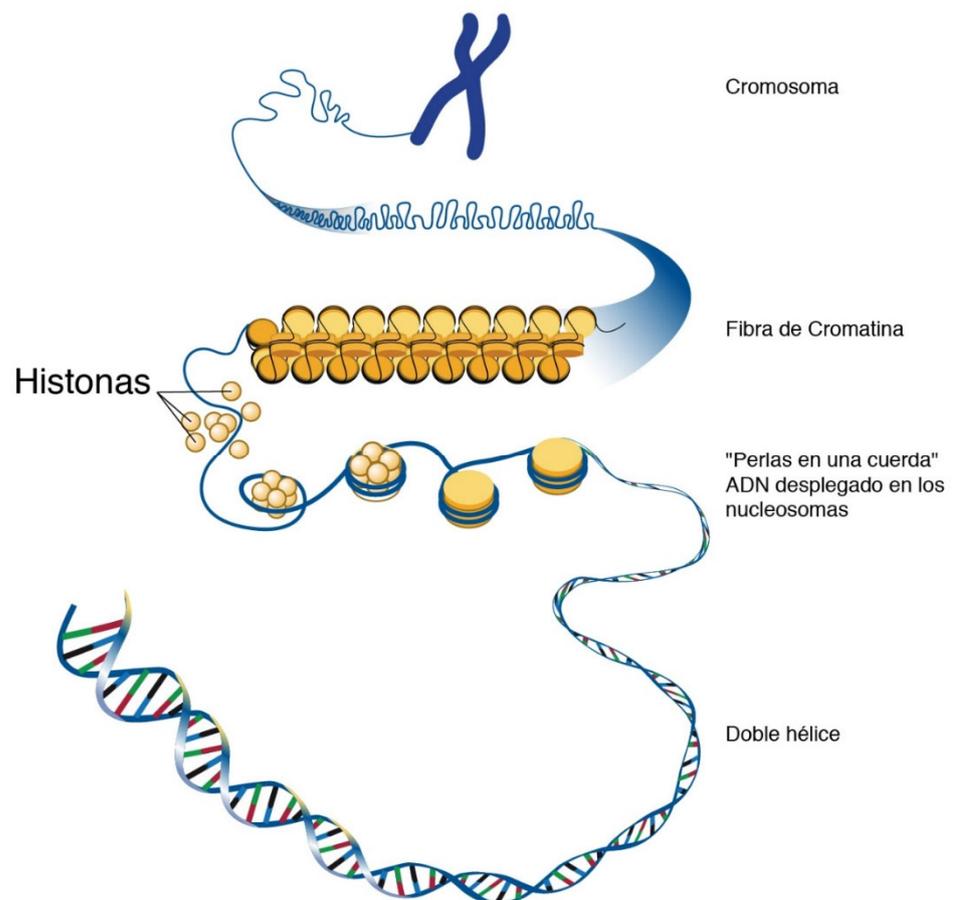


Figura 5. “Desdoblamiento” de un cromosoma mostrando las distintas estructuras asociadas³.

En la Figura 5 también se representan los **nucleosomas** que son unidades fundamentales de enrollamiento del ADN, formadas por 146 pares de bases de ADN que se unen a las histonas como si se trataran de cuentas de un collar, excepto que estas cuentas tienen el ADN rodeándolas en lugar de pasar a través de ellas, como ocurre en el caso de un collar. Estos 146 pares de bases se envuelven alrededor de un núcleo de 8 histonas.

En los extremos de los cromosomas encontramos los **telómeros** que son secuencias especiales de ADN y que tienen secuencias repetitivas que son reconocidas como el final de éstos, y que impiden que el cromosoma se rompa o se dañe³. Cuando la célula se divide, estos se hacen más cortos. Funcionan como un reloj para saber la edad que tiene la célula, y así puede limitar las veces que la célula se divide sin perder algunas de las partes importantes del ADN del cromosoma. Ahora sabemos que las células cancerosas mantienen sus telómeros largos, por lo que el reloj molecular desaparece y las células pueden seguir dividiéndose. Así engañan al cuerpo humano haciéndole pensar que aún debe mantener su división. En la Figura 6³ se identifican estas partes del cromosoma.

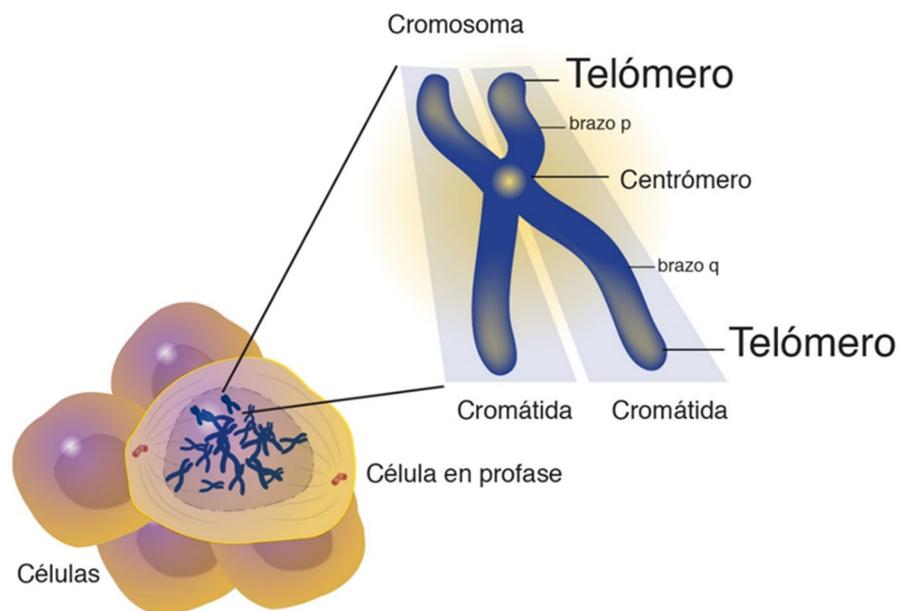


Figura 6. Partes de un cromosoma³.

En el cromosoma hablamos de cada una de las dos mitades idénticas de un cromosoma duplicado, denominada **cromátida**. Durante la división celular

(mitosis), en primer lugar, se duplica el cromosoma para que cada una de las células hijas reciba una dotación cromosómica completa. Después de la duplicación del ADN, el cromosoma pasa a estar compuesto por dos estructuras idénticas, llamadas cromátidas hermanas, que se unen por la zona del **centrómero**. Estas partes del cromosoma también se indica en la Figura 6.

Ácidos nucleicos

En la búsqueda de las sustancias químicas que componen las células, los bioquímicos descubrieron unas cuantas moléculas de los sistemas vivos. Pero desconocían cual era la molécula que contiene el código hereditario. Si que se conocía la cromatina, como hemos indicado, formada por proteínas y una nueva clase de moléculas de las células, los ácidos nucleicos.

El bioquímico Friedrich Miescher hacia el año 1869, fragmentando células y separando las sustancias químicas que se liberaban, quedó intrigado por un tipo de sustancia que precipitaba en forma de filamentos densos y enmarañados procedentes de glóbulos blancos extraídos de pus humana. Nombró a esta molécula nucleína porque se concentraba en el núcleo de la célula. Como se trataba de una sustancia química ácida, más tarde se le dio el nombre de ácidos nucleicos¹. Más tarde, hacia 1920, se conoció que se presentaban de dos formas: ADN y ARN. En la Figura 7 aparece una foto de la nucleína.



Figura 7. Nucleína.

Fuente: Alamy

ADN y ARN eran largas cadenas constituidas por cuatro componentes, o bases, enganchados en forma de cadena o esqueleto. De este esqueleto sobresalían las cuatro bases, como las hojas del tallo de una enredadera tipo hiedra. En el ADN, se sabía que las cuatro “hojas” eran la adenina (A), la guanina (G), la citosina (C) y la timina (T). En el ARN, en lugar de timina había un uracilo (U). A parte de estos datos, no se sabía nada ni de la estructura ni de la función del ADN y ARN.

Hoy conocemos perfectamente su estructura, después de los trabajos que desarrollaron con gran ingenio diversos investigadores, puesto que dicha elucidación no podía ser por observación directa, porque las técnicas utilizadas para ello descomponían su estructura (por ejemplo, los RX).

ADN

El **ADN** quiere decir ácido desoxirribonucleico. Es un polinucleótido, es decir, está formado por la unión de nucleótidos dispuestos en forma de doble hélice, como una escalera retorcida. Cada cadena de la doble hélice tiene una espina dorsal en la cual alternan un azúcar (desoxirribosa) y un grupo fosfato. A su vez, el grupo fosfato que se une al azúcar puede ser un monofosfato, un difosfato o un trifosfato. A cada azúcar se une una de las cuatro bases nitrogenadas: adenina (A) y guanina (G) (bases púricas), y timina (T) y citosina (C) (bases pirimidínicas). El ADN se lee como un código. Cada nucleótido toma el nombre de la base que contiene. El orden de las letras en este código permite que el ADN funcione de diferentes maneras.

La unión entre los pares de bases corresponde al peldaño de la escalera. Las dos cadenas se mantienen juntas gracias a los puentes de hidrógeno entre las bases complementarias, es decir, la adenina (A) con la timina (T), y la citosina (C) con la guanina (G). Si tenemos una A en una hebra, en la otra hebra habrá una T, en cambio, si tenemos una C en una hebra, en la otra habrá una G (A-T y C-G). A esta relación restrictiva entre las bases se le denomina complementariedad y hace que las cadenas de nucleótidos de ADN sean complementarias entre sí. Dado que una base púrica se aparea siempre con una base pirimidínica, la cantidad de bases púricas será siempre igual a la de pirimidínicas. En la Figura 8 se indica esta estructura³.

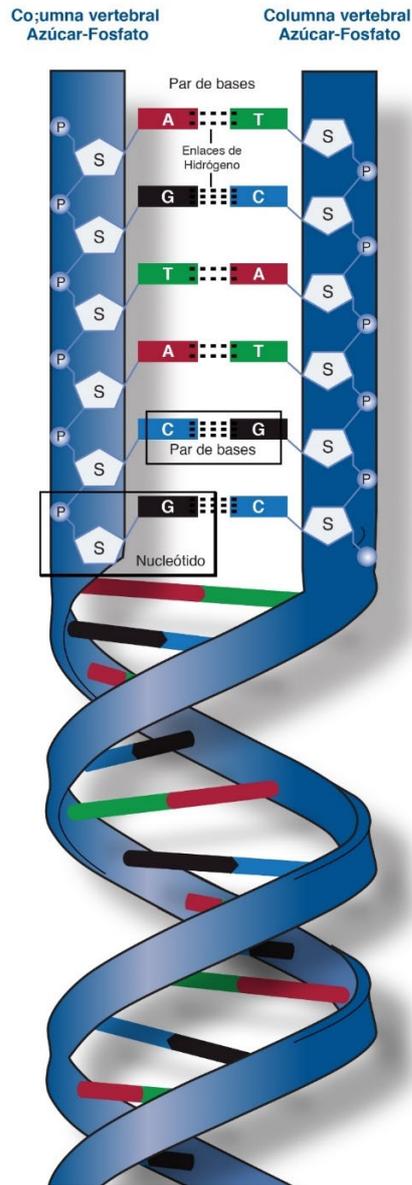


Figura 8. Estructura del ADN³.

La unión entre los nucleótidos es muy estable, por eso estos enlaces son muy difíciles de romper y se necesitan enzimas específicos para hacerlo.

Cada cadena de polinucleótidos tiene direccionalidad, es decir, tiene dos extremos que son distintos entre sí. En el extremo 5', sobresale el grupo fosfato unido al carbono 5' del primer nucleótido. En el otro extremo, llamado 3', está expuesto el hidroxilo unido al carbono 3' del último nucleótido. La hebra antisentido empieza en el extremo 3', también llamada cadena molde, y la hebra sentido empieza en el extremo 5', tal y como indica la Figura 9. Las cadenas de ADN generalmente se escriben en la dirección 5' a 3', lo que significa que el nucleótido del extremo 5' es el primero y el nucleótido del extremo 3' es el último. Como veremos, es en la transcripción que la ARN polimerasa se une a la cadena molde o antisentido.

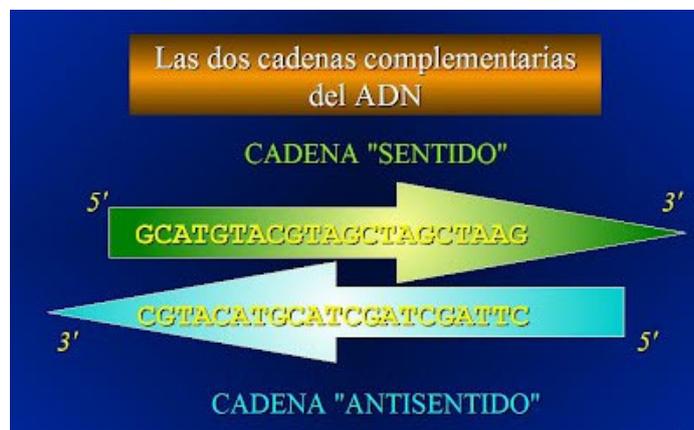


Figura 9. Representación de las cadenas complementarias del ADN.⁵

Las dos cadenas de la hélice corren entonces en direcciones opuestas, lo que significa que el extremo 5' de una cadena se une al extremo 3' de su correspondiente, según indica la Figura 10. Esto se conoce como orientación antiparalela y es importante al copiar el ADN.

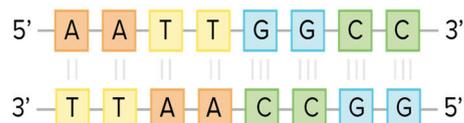
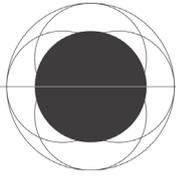


Figura 10. Orientación antiparalela del ADN.

Fuente: Khan Academy



En la Figura 11 se indica en detalle los enlaces entre las bases nitrogenadas.

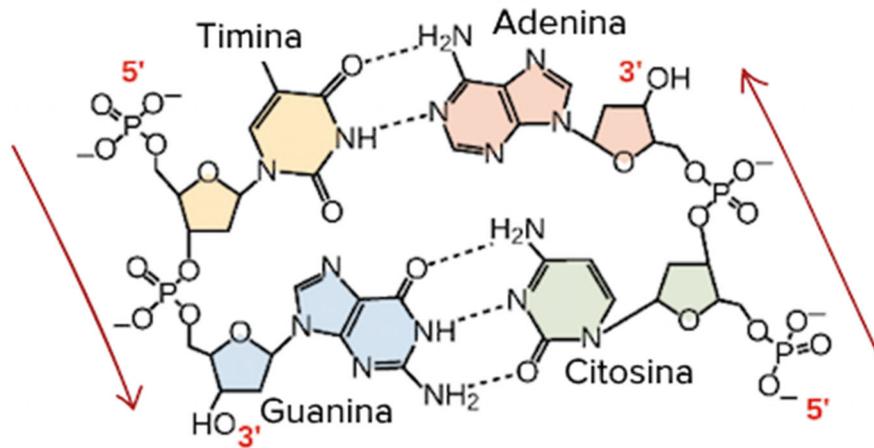


Figura 11. Uniones entre bases nitrogenadas en el ADN.

Fuente: Khan Academy

Podemos contar el ADN y la cantidad de ADN usando los pares de bases. Un fragmento de ADN con una función específica puede estar formado por 10.000 pares de bases o, lo que es equivalente, 10 kilobases. El par de bases es la unidad para medir el ADN. También se utilizará para medir el ARN.

El ADN nuclear o genómico contiene⁴:

. 6 a 7 billones ($6-7 \times 10^9$) de nucleótidos.

. 6 a 7 millones ($6-7 \times 10^6$) de kilobases (10^3) de ADN

O también⁴,

. 3.095.677.412 pares de bases (pb).

En la Figura 12 se representan las dimensiones de la cadena de ADN.

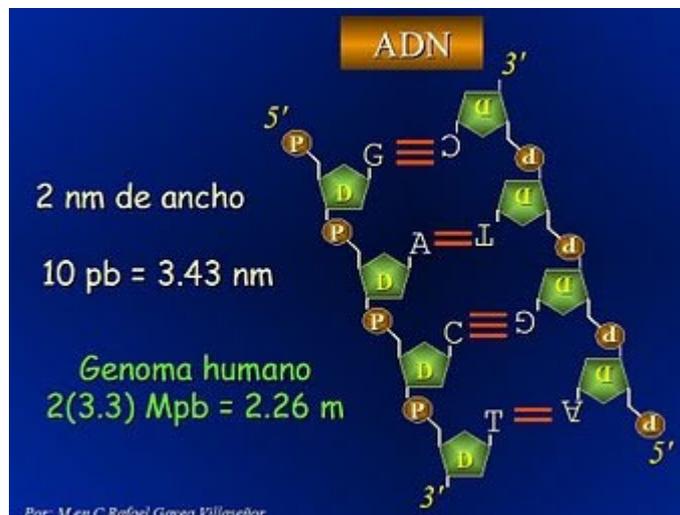


Figura 12. Dimensiones de la cadena de ADN⁵.

En el ADN nuclear, cada persona tiene dos copias de cada gen, una heredada de cada progenitor.

En 2003 se publicó la secuencia completa del genoma humano nuclear, resultado del *Human Genome Project* (HGP). Esto quiere decir que se consiguió conocer el orden en el están dispuestas las más de 3.000 millones de pares de bases. En términos científicos probablemente representa el hito más importante en la historia de la ciencia des de El Renacimiento². Los más de 3.000 millones de pares de bases que daban lugar a los 20.000-30.000 genes del genoma humano representaban sólo el 2% del genoma total. Se conoció que casi todos los seres humanos tienen los mismos genes dispuestos aproximadamente en el mismo orden y más del 99,9% de la secuencia de una persona es idéntica a la de otra. Es decir, nos diferenciamos en un 0,1 % de nuestra secuencia de ADN. En la Figura 13³ se representa esta similitud y diferencia, suficiente para cambiar la forma y la función de una proteína que se fabrica y cuando o dónde se fabrica. Afectan al color de los ojos, el cabello, la piel. Y lo más importante, las variaciones en el genoma también afectan el riesgo de desarrollar enfermedades y las respuestas a los medicamentos, entre otros mecanismos.

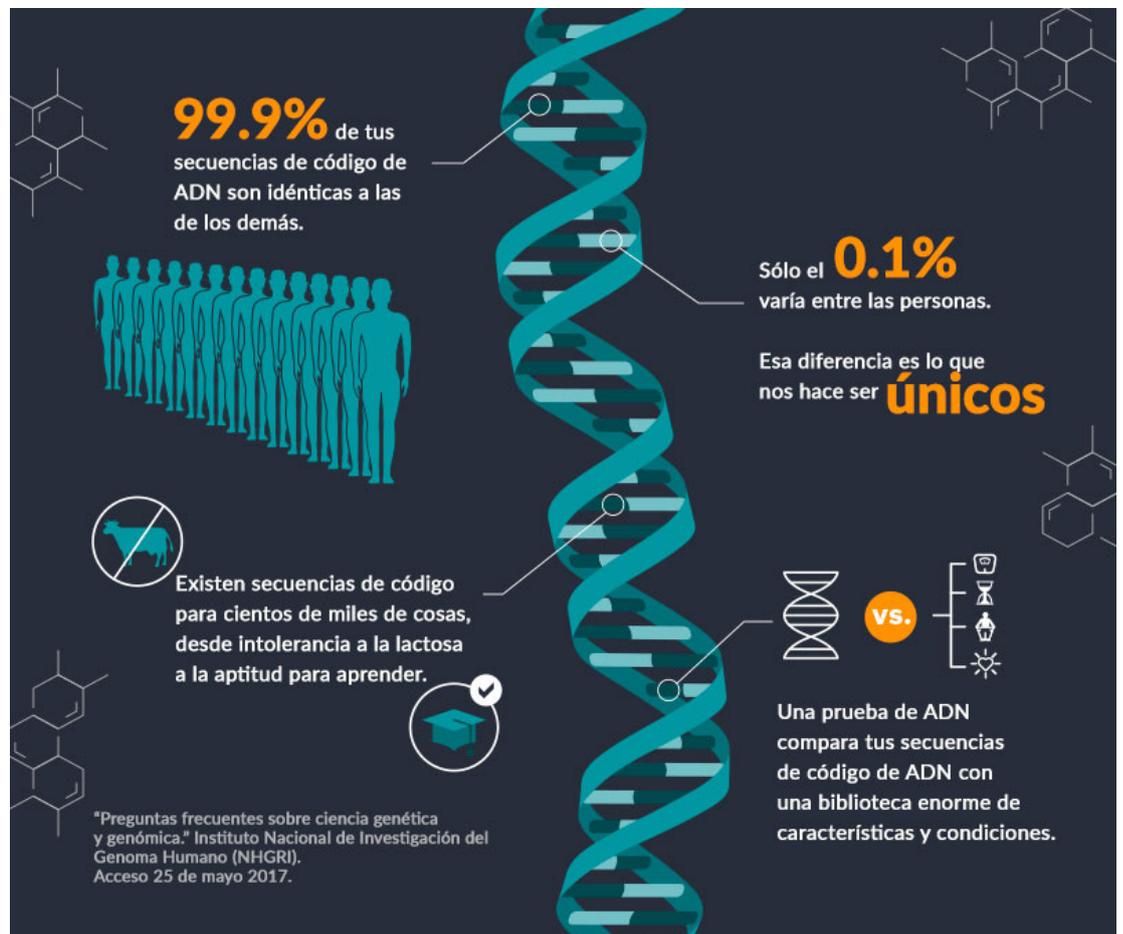


Figura 13. Representación de algunos aspectos del significado de la composición del ADN³.

La mayor parte del genoma (el 98%) no codifica ningún producto en absoluto. Sin embargo, es necesaria. Es por lo que desde el mismo HGP se inició el ENCODE (del inglés *"the ENCyclopedia Of DNA Elements"*). ENCODE es uno de los proyectos que se diseñó para empezar a entender la función del genoma, también de aquella parte no correspondiente a genes.

Se distinguen distintas categorías de ADN en función de su repetitividad y carácter codificante⁶, tal y como se esquematiza en la Figura 14.

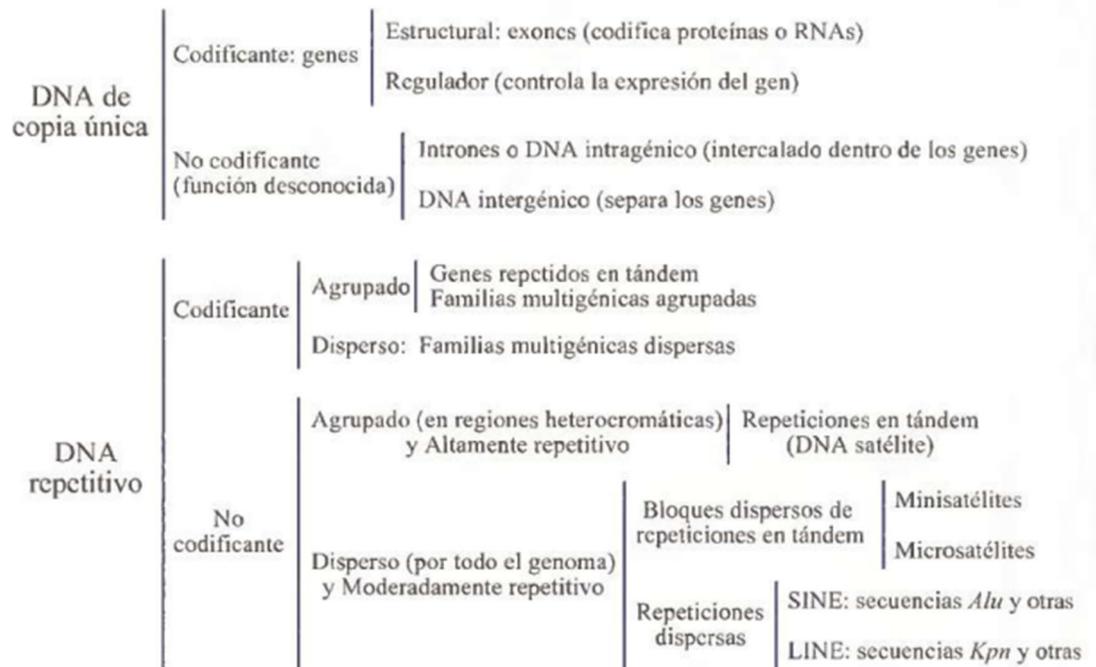


Figura 14. Distintas categorías de ADN⁶.

El **ADN** de copia única, **simple** o no repetitivo constituye la mayor parte del genoma. Parte de este ADN constituye las secuencias de genes que codifican proteínas y ARNs (**ADN simple codificante**), y la otra es la responsable del control de la expresión de esas secuencias, mientras que el resto, mayoritario, es **ADN simple no codificante**, cuya función no siempre es conocida.

El **ADN repetitivo** son secuencias repetidas que tienen tamaños diversos y cada una se encuentra de forma idéntica o casi idéntica muchas veces en el genoma. Su distribución puede ser en forma dispersa por todo el genoma, entremezcladas con secuencias de copia única o bien en forma agrupada, localizada en regiones concretas del cromosoma. El número de copias de las repeticiones varía desde unos cientos o miles (ADN moderadamente repetitivo) hasta cientos de miles (ADN altamente repetitivo). Una parte del **ADN repetitivo** es **codificante**, sin embargo, para el **ADN repetitivo no codificante** no se conoce siempre una función clara.

Algunos ejemplos de **ADN repetitivo codificante** son familias génicas clásicas o conservadas, con genes repetidos en tándem, que codifican histonas, ARNr, ARNt, etc. Todas ellas se expresan. También se encuentran las familias multigénicas con

genes agrupados, que codifican las globulinas o el HLA-1, entre otras proteínas. En éstas, solo se expresan algunas de las repeticiones. Por último, están las familias multigénicas con genes dispersos, como las que codifican la ferritina, aldolasa, actina, entre otras, y casi siempre son todas funcionales.

Como ejemplo de **ADN repetitivo no codificante agrupado** podemos indicar el **ADN satélite** que se encuentra en regiones en torno al centrómero y en los telómeros (con secuencias teloméricas TTAGGG, repetidas entre 250 y 1000 veces), tal y como se representa en la Figura 15. Se ha sugerido que ayudan a estabilizar el cromosoma y tendrían un rol en el control del envejecimiento.

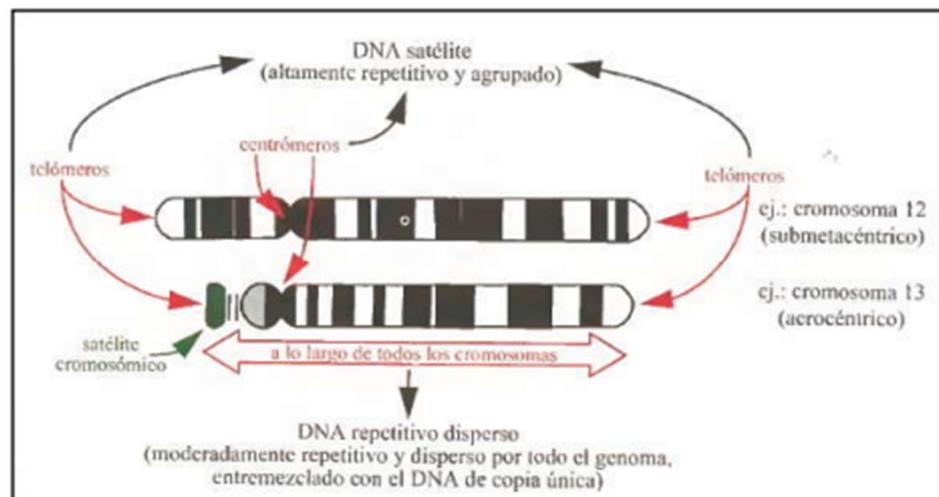


Figura 15. Dibujo de la localización de ADNs satélite en el cromosoma 12 y 13⁶.

En abril de 2022 se han publicado unos trascendentes estudios que se citan en la reseña periodística de S. Romero⁷. Se trata de un nuevo hito. Después de 20 años del mapeo del genoma de referencia humano (nombrado GRCh38, acrónimo del inglés "*Genome Reference Consortium build 38*", que tuvo su origen en el Proyecto del Genoma Humano, y que se ha actualizado con algunas incorporaciones posteriores), en abril de 2022 se describe el mapeo del genoma CHM13, elaborado por el consorcio Telómero a Telómero T2T (nombrado T2T-CHM13). Este consorcio está formado por un equipo internacional de investigadores. Gracias a una nueva tecnología de secuenciación del ADN, se han podido desenredar tramos repetitivos de ADN que fueron redactados en la versión anterior GRCh38, que en realidad se trataba de un borrador, según indican los investigadores. Se ha conseguido secuenciar el 8% que faltaba, que no

se había omitido por falta de importancia, sino por limitaciones tecnológicas, que hoy se han superado. Se ha secuenciado, por tanto, el genoma completo, sin lagunas. Se han agregado 200 millones de pares de bases de las más de 3000 millones totales. También se corrigen miles de errores estructurales. Las lagunas que ahora cubre la nueva secuencia abarcan la totalidad de los brazos cortos de cinco cromosomas humanos y cubren algunas regiones más complejas del genoma. Entre ellas, secuencias de ADN repetitivas que se encuentran en los telómeros y los centrómeros que coordinan la separación de los cromosomas replicados durante la división celular. También se revelan duplicaciones de segmentos no detectadas anteriormente. CHM13 carece de un cromosoma Y. Conocemos la que nos aporta GRCh38. Otra limitación es que se trata de un genoma humano, cuando se mejoraría el mapeo si tuviéramos el mapeo de todos los genomas humanos. Este es el nuevo objetivo, lo que se denomina el pangenoma humano.

En la tabla de la Figura 16 se comparan el genoma humano GRCh38 y T2T-CHM13.

Summary	GRCh38p13	CHM13v1.1	±%
Assembled bases (Gbp)	2.92	3.05	+4.5%
Unplaced bases (Mbp)	11.42	0	-100.0%
Gap bases (Mbp)	120.31	0	-100.0%
# Contigs	949	24	-97.5%
Ctg NG50 (Mbp)	56.41	154.26	+173.5%
# Issues	230	46	-80.0%
Issues (Mbp)	230.43	8.18	-96.5%
Gene Annotation			
# Genes	60,090	63,494	+5.7%
protein coding	19,890	19,969	+0.4%
# Exclusive genes	263	3,604	
protein coding	63	140	
# Transcripts	228,597	233,615	+2.2%
protein coding	84,277	86,245	+2.3%
# Exclusive transcripts	1,708	6,693	
protein coding	829	2,780	
Segmental duplications (SDs)			
% SDs	5.00%	6.61%	
SD bases (Mbp)	151.71	201.93	+33.1%
# SDs	24097	41528	+72.3%
RepeatMasker			
% Repeats	50.03%	53.94%	
Repeat bases (Mbp)	1,516.37	1,647.81	+8.7%
LINE	626.33	631.64	+0.8%
SINE	386.48	390.27	+1.0%
LTR	267.52	269.91	+0.9%
Satellite	76.51	150.42	+96.6%
DNA	108.53	109.35	+0.8%
Simple repeat	36.5	77.69	+112.9%
Low complexity	6.16	6.44	+4.6%
Retroposon	4.51	4.65	+3.3%
rRNA	0.21	1.71	+730.4%

Figura 16. Comparación entre el genoma humano GRCh38 y T2T-CHM13⁸.

Existe otro fragmento de ADN en la célula³, el **ADN mitocondrial** que contiene 37 genes de un total de unos 16500 pares de bases (200.000 veces inferior al genoma nuclear haploide), que son esenciales para la función mitocondrial normal. Es un ADN pequeño y circular. Se representa en la Figura 16. El ADN mitocondrial solo se hereda de la madre. Es un ADN diferente del núcleo, que se hereda de ambos progenitores. Codifica diferentes proteínas que son específicas de la mitocondria. Si existe un defecto en algunas de las bases del ADN mitocondrial, es decir, una mutación, se tiene una enfermedad mitocondrial con la consecuente incapacidad de producir suficiente energía en órganos como el músculo y el cerebro, o el riñón. A veces una enfermedad se hereda por línea materna, y no de ambos padres. Entonces podemos sospechar que se trata de una enfermedad mitocondrial. Para ello, se estudia el árbol genealógico para conocer si esa afectación está presente en la línea materna de las anteriores generaciones.

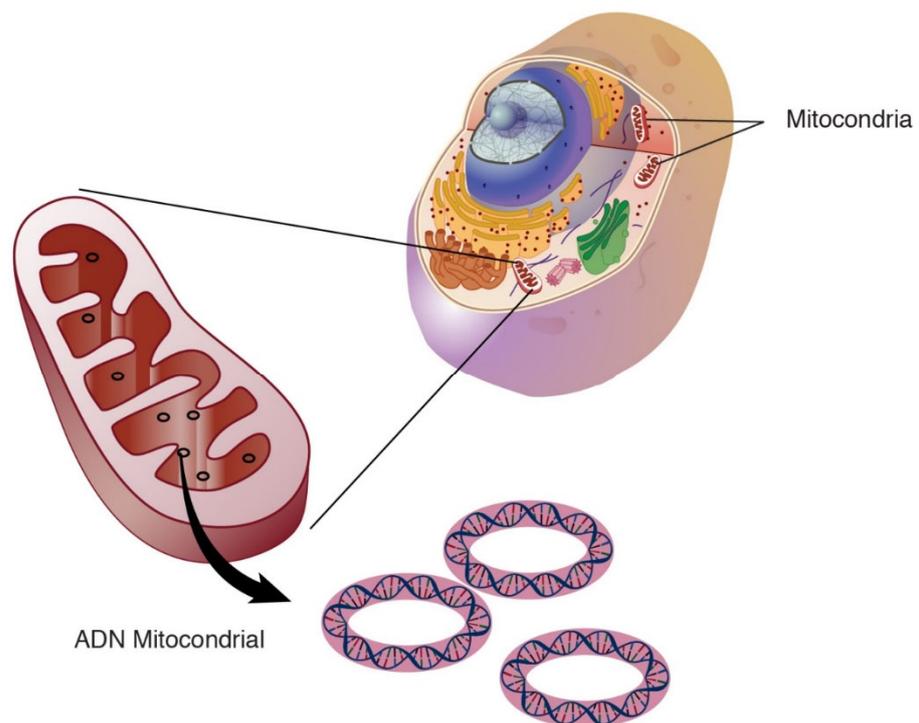


Figura 17. Representación del ADN mitocondrial³.

A diferencia del nuclear, el genoma de la mitocondria está formado por ADN codificante casi en su totalidad, tal y como indica la Figura 18, en la que también se indican otras características diferenciales.

	NUCLEAR	MITOCONDRIAL
Tamaño	6,4 Gb = 6.400 Mb = 6.400.000 kb = $6,4 \cdot 10^9$ pb	16,5 kb $\times m$
Tipo de DNA	Lineal, bicatenario	Circular, bicatenario
Número de cromosomas	23 distintos, 2 copias: 46 en total	Uno, con varias copias en cada mitocondria; entre 200 y 2.000 copias por célula
Presencia de nucleosomas	Sí	No
Proteínas asociadas	Alrededor del 50% de la masa, histonas y otras	Poco numerosas o ausentes
DNA codificante	Entre 2 y 5%	~ 93%
DNA repetitivo	Abundante (~ 40%)	Prácticamente nada
Número de genes	~ 25.000	37
Densidad de genes	~ 1 gen cada 128 kb	1 gen cada 0,45 kb
Presencia de intrones en los genes	En la mayoría	Ausentes

Figura 18. Características diferenciales entre el ADN nuclear y el ADN mitocondrial.

Fuente: Consejo Mexicano de Neurociencias.

T2T-CHM13 también incluye el mapeo del genoma mitocondrial⁷.

ARN

EL **ARN** es el ácido ribonucleico. Su estructura es muy similar a la del ADN³. A diferencia del ADN, el ARN es de cadena sencilla. Una hebra de ARN tiene un eje constituido por un azúcar (ribosa) y grupos fosfato de forma alterna. Unidos a la ribosa se encuentra una de las cuatro bases nitrogenadas adenina (A), uracilo (U), citosina (C) o guanina (G). En el ARN, la timina (T) es sustituida por uracilo (U). En la Figura 19 se indican las diferencias estructurales de un nucleótido de ADN y ARN.

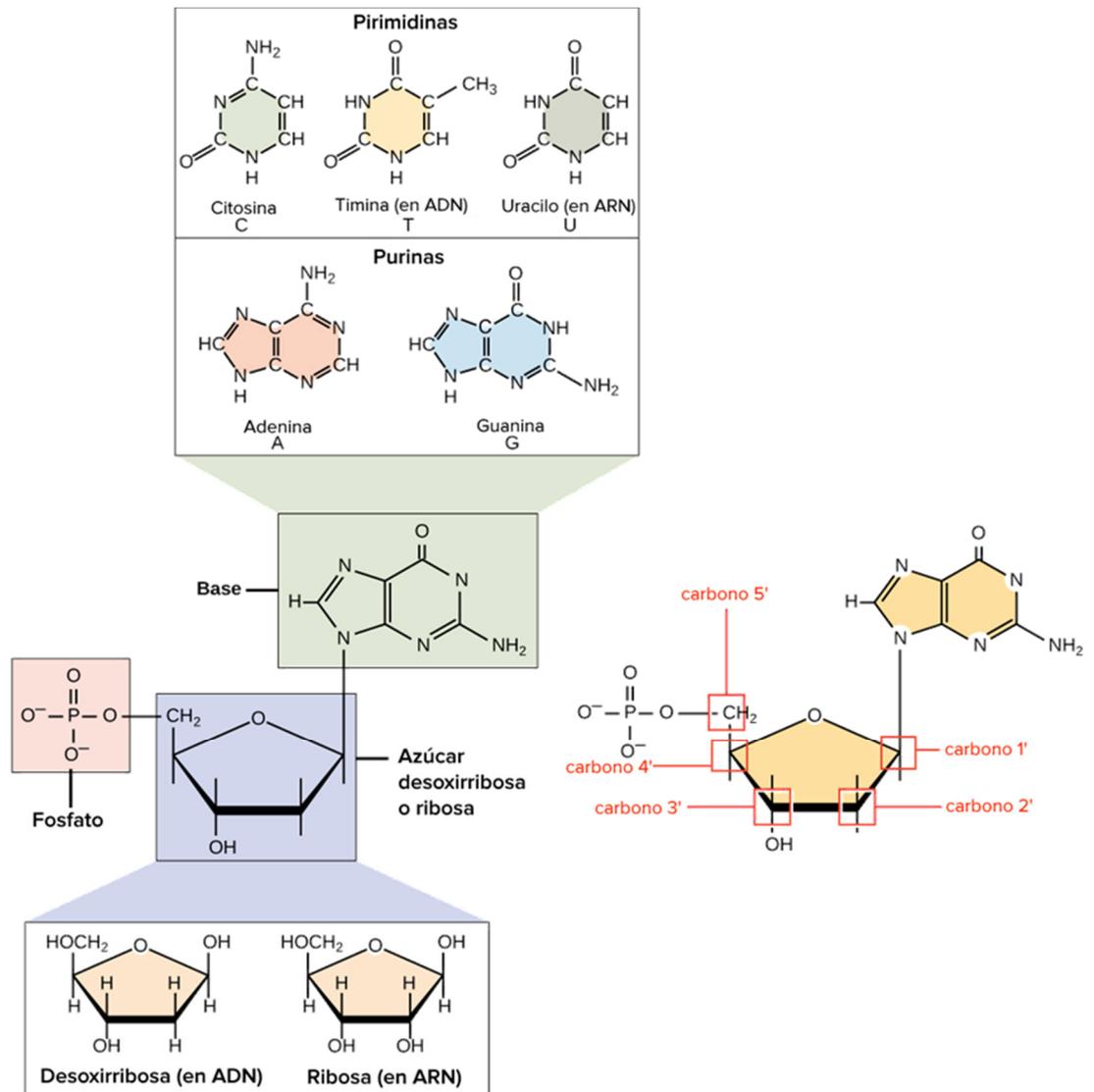
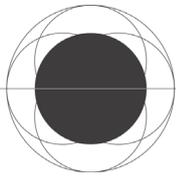


Figura 19. Diferencias estructurales de un nucleótido de ADN y ARN.

Fuente: Khan Academy

En la Figura 20 se representan las diferentes cadenas de ADN y ARN.

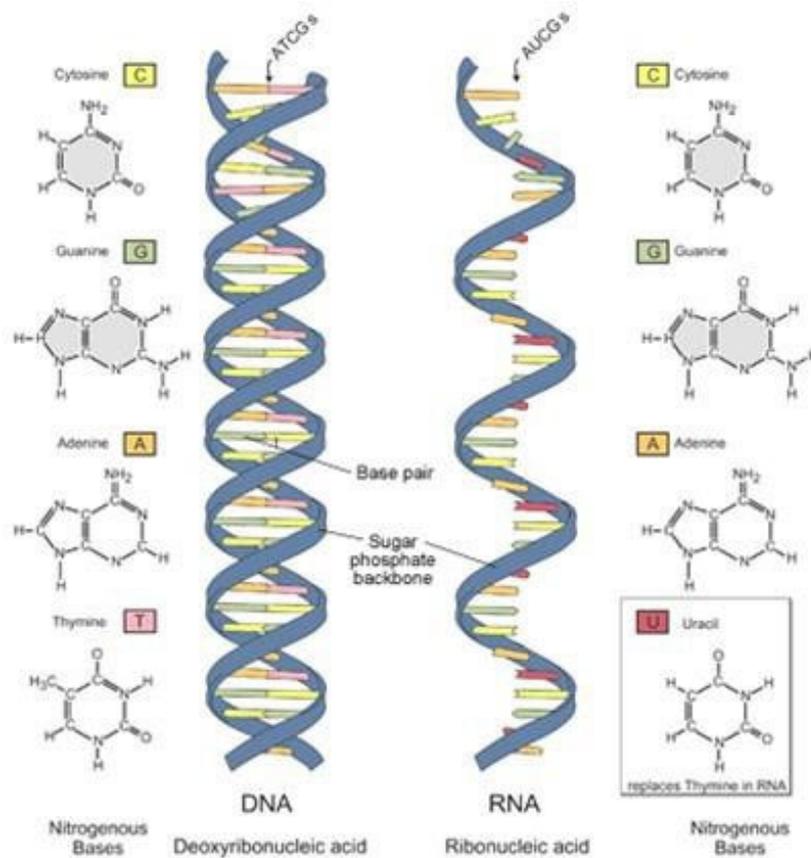
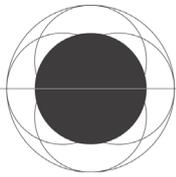


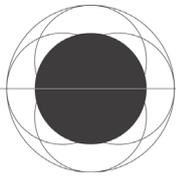
Figura 20. Cadenas de ADN y ARN.

Fuente: Human Genome Research Institute.

Hay diferentes tipos de ARN en la célula. Se pueden clasificar en dos grupos ARN codificante y ARN no codificantes. **ARN mensajero** (ARNm) es el ARN codificante, que más tarde dará lugar a las proteínas, y los ARN no codificantes no conducen a proteínas. Son el **ARN ribosomal** (ARNr), el **ARN de transferencia** (ARNt), ambos involucrados en el proceso de traducción. Más tarde, se han descrito unos ARN de pequeño tamaño (**microARN o miARN**) que están involucrados en la regulación de la expresión génica, como indicaremos más adelante.

ADN codificante. El código genético.

¿Cómo transfiere el ADN la información a la célula para que ésta pueda realizar las distintas funciones? Como hemos indicado, los nucleótidos se disponen en el ADN exponiendo las bases nitrogenadas adenina (A), citosina (C), guanina (G) y timina (T) como elemento distintivo entre un nucleótido y otro. Y, como veremos más adelante cuando hablemos de transcripción, el ADN pasará la información al ARN, siguiendo el mismo orden de bases y de acuerdo con el modelo de complementariedad. En el ARN tenemos uracilo (U) en lugar de timina (T). A través de la combinación de estas 4 letras (A, C, G, U), se transmitirá la información a la célula para desarrollar los elementos estructurales que llevan a cabo las distintas funciones. Esos elementos estructurales son las proteínas. Pues bien, para comprender cómo tiene lugar este proceso, se consiguió elucidar que la combinación de cada tres letras (que se conoce con el nombre de codón) incorpora un aminoácido en la estructura de la proteína sintetizada. Hay 20 aminoácidos para la construcción de proteínas y cuatro bases nitrogenadas o “letras” (A, C, G, U) que, en combinación de tres en tres codificaran uno u otro aminoácido, tal y como indica la Figura 21. La elucidación de la combinación de las tres letras que codifican un aminoácido permitió conocer el **código genético** universal, ya que es el mismo en todos los organismos vivos.



		Segunda letra					
		U	C	A	G		
Primera letra	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Alto UAG Alto	UGU } Cys UGC } UGA Alto UGG Trp	U C A G	
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G	
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G	
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G	
						Tercera letra	

Figura 21. Código genético.

Fuente: Khan Academy

La secuencia de ARNm se agrupa en 64 tipos de codones de los cuales podemos distinguir:

- . 61 codones que codifican los 20 aminoácidos
- . 1 codón de inicio de la traducción (AUG), que codifica metionina
- . 3 codones de parada, tres secuencias de tres bases que no codifican un aminoácido (UAA, UAG, UGA), que indican final de traducción.

Ejemplo: ACU, codifica treonina
CAU, codifica histidina
GGU, codifica glicina

La Figura 22 es una representación circular del código genético.



Figura 22. Código genético⁵.

Es un código genético degenerado, que quiere decir que diferentes combinaciones de bases nitrogenadas codifican para el mismo aminoácido.

El lenguaje del genoma humano

En las bacterias, los genes están dispuestos formando una cadena de ADN ininterrumpida y continua, que empieza por un triplete de inicio (ATG) y sigue hasta el triplete final de parada. Los genes bacterianos no contienen módulos separados, e internamente no están separados por ningún espaciador¹. Se podría establecer una analogía de esto que hemos explicado con la palabra

estructura

En las bacterias no hay particiones, ni relleno, ni interrupciones. El gen estaría contenido en el genoma así

estructura

En los animales y en los virus de animales, en cambio, un gen acostumbra a estar dividido en partes y está interrumpido por largos segmentos de ADN de relleno, como hemos indicado anteriormente. En la analogía, en el genoma humano, la palabra estaría partida por segmentos intermedios de ADN

es...tru...c...t...ur...a

Los largos segmentos de ADN indicados en forma de puntos suspensivos no contendrían ninguna información que codifique una proteína. Cuando este gen se utilice para crear un mensaje, -cuando el ADN se utiliza para sintetizar ARN mensajero, como veremos-, los segmentos de rellenos serían escindidos del ARNm y el ARN se volvería a empalmar sin los segmentos intermedios, de manera que

es...tru...c...t...ur...a

quedaría simplificada a

estructura

Este proceso se llama “**splicing**” o **corte** o **empalme** del ARN. Se obtiene así el ARN maduro.

¿Porque sucede así?

Teniendo los genes divididos en módulos, una célula puede formar una cantidad impresionante de mensajes a partir de un solo gen.

En la analogía, los fragmentos de

es...tru...c...t...ur...a

se podrían empalmar para dar

cura

truca

o de

g...e...n...om...a

podrían dar

gen

gnom

om

Este proceso se llama **empalme alternativo**, generación de nuevos ARN maduros a partir de un mismo gen, obtenido mediante la delección de uno o varios exones en los procesos de corte y empalme.

Por otro lado, los genes modulares tienen una ventaja: los módulos individuales de genes diferentes se podrían mezclar y emparejar entre ellos para formar genes completamente nuevos. En la analogía de las dos palabras:

es...c...e...n...a

Los segmentos intermedios de relleno son los **intrones**.

Los fragmentos o módulos que se conservan se denominan **exones**.

En los genes humanos, los intrones no son la excepción. Son la norma. Hay intrones humanos que son muy largos, de centenares de millares de bases de ADN.

Como hemos indicado anteriormente, los genes propiamente dichos están separados entre ellos por largos segmentos de ADN intercalados que se denominan ADN intergénico. Se cree que tanto el ADN intergénico como los intrones tienen unas secuencias que permiten regular los genes si conviene. Siguiendo la analogía de antes,

Esta....es.....la ...(...) ... es...tru...c...t...ur...a...de.....nuestro.....gen...om...a;

las palabras equivaldrían a los genes, los puntos suspensivos largos que hay entre palabras representarían los segmentos de ADN intergénico, mientras que los puntos suspensivos más cortos del interior de las palabras serían los intrones. Los paréntesis, y el punto y coma -signos de puntuación-, serían regiones del ADN que regulan los genes.

Con ello podemos concluir que la larga molécula de ADN que contiene material genético codificante se estructura de tal manera que permite gran interrelación entre los distintos segmentos. Esto explica que los humanos teniendo un número de genes (entre 20000 y 27000) del orden de magnitud similar al de un gusano

(éste tiene 20.000 genes), podemos llevar a cabo una gran variedad de funciones, entre otras sintetizar miles de proteínas (más de 100.000).

De forma más precisa, las anotaciones del genoma T2T-CHM13 recientemente publicado⁷ indican que este incluye 63494 genes y 233615 transcripciones, de los cuales se predice que 19969 son genes (86245 transcripciones) que codificarán proteínas. La mayoría de los genes que no codifican proteínas, están asociados a elementos repetitivos.

Plegamiento del ADN y elementos estructurales adicionales (epigenéticos)

El **tipo de plegamiento de la cromatina** también es un elemento estructural del material genético. Según cómo se encuentre, puede significar el grado de expresión del ADN, es decir, que esté en condiciones de ser expresado o no. Se puede establecer una relación directa entre el grado de condensación de la cromatina y la ausencia de transcripción. Se detallará más adelante.

Existen otros elementos estructurales que se pueden asociar a la cromatina. Éstos son, por ejemplo, la **adición de grupos metilo** o la **adición de grupos acetilo**. Estos elementos estructurales intervienen en la regulación de la expresión del ADN y se denominan elementos epigenéticos, pues están por encima del material genético propiamente dicho, sin cambiar la secuencia del ADN. Estos se estudian en la epigenética, ciencia emergente que está teniendo cada vez más importancia en la modulación de la expresión genética a través de los factores ambientales o externos. Más adelante también abordaremos estos elementos estructurales.

Mutaciones y polimorfismos

Estamos hablando de cambios permanentes en la secuencia del ADN o mutaciones respecto a la secuencia natural o salvaje. Las mutaciones varían de tamaño; pueden afectar un par de bases hasta un gran segmento de un cromosoma que incluye múltiples genes.

Las mutaciones pueden dar lugar a una manifestación fenotípica, ya sea positiva o negativa o, no, pueden ser silentes. Ésta últimas son las más frecuentes. Si provocan una alteración fenotípica negativa (fenotipo clínico) se denominan mutaciones patogénicas.

Se pueden clasificar de dos formas diferentes según su origen:

- Mutaciones **hereditarias**. Son las mutaciones que se heredan de padres a hijos y están presentes durante toda la vida de la persona en prácticamente todas las células del cuerpo. Se producen en la línea germinal en los óvulos y el esperma, y pueden transmitirse a la descendencia.
- Mutaciones **adquiridas** (o **somáticas**). Ocurren en algún momento durante la vida de una persona y están presentes solo en ciertas células. Estos cambios pueden ser causados por factores ambientales como la radiación ultravioleta del sol, o puede ocurrir si se comete un error, ya que el ADN se copia durante la división celular. Estas mutaciones no se transmiten a la siguiente generación.

La mayoría de las mutaciones genéticas que causan enfermedad son poco frecuentes en la población general. Sin embargo, otros cambios genéticos ocurren con mayor frecuencia.

También se pueden clasificar en función si afectan a un gen (mutaciones génicas), a un cromosoma (mutaciones cromosómicas) o al número total de cromosomas (mutaciones genómicas)⁷.

1.- Mutaciones génicas.

Nos referiremos a aquellas alteraciones genéticas que ocurren en más del 1 % de la población y se denominan **polimorfismos**. Son suficientemente comunes como para considerarse una variación normal en el ADN. Aunque muchos polimorfismos no tienen efectos negativos en la salud (a diferencia de las mutaciones propiamente dichas que siempre tienen un efecto patológico), algunas de estas variaciones pueden originar susceptibilidad a desarrollar ciertos trastornos.

Dicha variación puede ser de varios tipos:

- Variaciones en las que **se sustituye una base** de una secuencia del genoma por otra distinta. Se dan en un porcentaje de la población superior al 1%. Son los polimorfismos de nucleótido único (SNP, del inglés “*Single Nucleotide Polymorphism*”), tal y como indica la Figura 23.

Polimorfismo de un solo nucleótido (SNP)

Individuo 1

Cr 2 ..CGATATTCC**T**ATCGAATGTC..
copia1 ..GCTATAAGG**A**UAGCTTACAG..

Cr 2 ..CGATATTCC**C**ATCGAATGTC..
copia2 ..GCTATAAGG**G**TAGCTTACAG..

Individuo 2

Cr 2 ..CGATATTCC**C**ATCGAATGTC..
copia1 ..GCTATAAGG**G**TAGCTTACAG..

Cr 2 ..CGATATTCC**C**ATCGAATGTC..
copia2 ..GCTATAAGG**G**TAGCTTACAG..

Figura 23. Polimorfismo de un solo nucleótido (SNP)³.

Son las más frecuentes ya que constituyen el 90% de las variaciones genómicas humanas, apareciendo un SNP cada varias centenas a un millar de pares de bases y se encuentran dispersas por todo el genoma. Así, se conocen más de 5 millones registradas en bases de datos. La consecuencia fenotípica para un determinado individuo dependerá de la localización en el gen de los citados SNPs, que se pueden resumir según la posición de la variación de la siguiente manera:

a.- Que se sitúe en la región codificadora del gen (**exones**).

Se distinguen dos tipos:

- o Los SNP **sinónimos** son aquellos en los que no se produce un cambio de aminoácido en la proteína que codifica, pero que

puede tener consecuencias funcionales (afectando a la estabilidad del transcrito o al “*splicing*”)

- Los SNP **no sinónimos** sí que van a producir un cambio en la cadena de aminoácido que puede afectar a la estabilidad y estructura de la proteína y su afinidad por los sustratos.

Dentro de esta categoría se encuentran:

- Mutaciones de sentido erróneo o **cambio de sentido** (“*missense*”): se produce un codón que codifica para un aminoácido diferente, pudiendo afectar a la funcionalidad de la proteína resultante. Sin embargo, no todas las mutaciones de este tipo van a producir consecuencias funcionales.
- Mutaciones **sin sentido** (“*nonsense*”): se produce el cambio del codón original por un codón de parada, que va a producir una proteína truncada (normalmente inactiva).

b.- Que se sitúe el SNP en la **región reguladora** del gen, lo que puede conllevar que se vea afectada la capacidad de enlace de los factores de transcripción al gen y, en definitiva, que se vean afectados los niveles normales de expresión del citado gen y la consiguiente cantidad de proteína expresada.

c.- Que se localicen los SNPs en las **regiones no-codificadoras** del genoma (regiones intergénicas e intrones), en cuyo caso está por conocer el impacto que pueden tener sobre el fenotipo del individuo, aunque su estudio es el de mayor interés como marcadores en determinados procesos patológicos de origen genético.

- Variaciones por **inserción o delección de una o más pares de bases** en la secuencia de ADN, siendo las más frecuentes las de 1 a 3 pares de bases. En la Figura 24 se representan variaciones por inserción o delección de un par de bases.

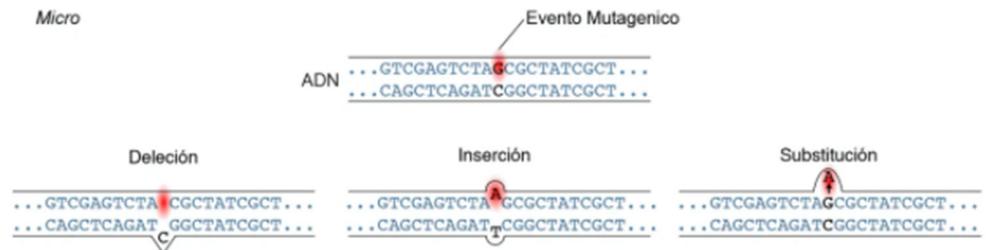


Figura 24. Representación de variaciones por delección, inserción y sustitución de un par de bases⁷.

- Variaciones **por repeticiones en tándem**: que a su vez se clasifican en microsatélites, minisatélites y satélites. Se representan en la Figura 25.

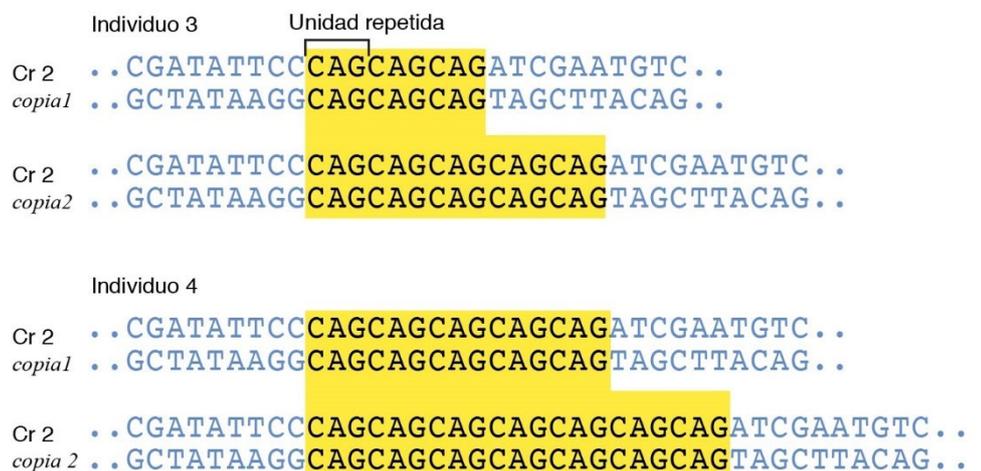


Figura 25. Polimorfismos en tándem de repeticiones cortas³.

- Variaciones en el **número de copias** (CNVs, del inglés “*Copy Number Variations*”): segmento de ADN de 1 kilobase (kb) o mayor, cuyo número de copias es variable si se compara con un genoma de referencia. Se englobarían las duplicaciones, multiplicaciones y delecciones de genes enteros. Se pierde el gen o, por el contrario, se producen varias copias de él, lo que se traduce en ausencia o excesivas cantidades de enzima.

2.- Mutaciones cromosómicas.

Los cromosomas pueden sufrir delecciones, duplicaciones, inversiones, sustituciones o cambios de lugar en los genes. Al estar involucrados muchos más genes, las consecuencias suelen ser peores. De hecho, las mutaciones

cromosómicas normalmente dan lugar a organismos que no son viables. En la figura 26 se representan algunos ejemplos.

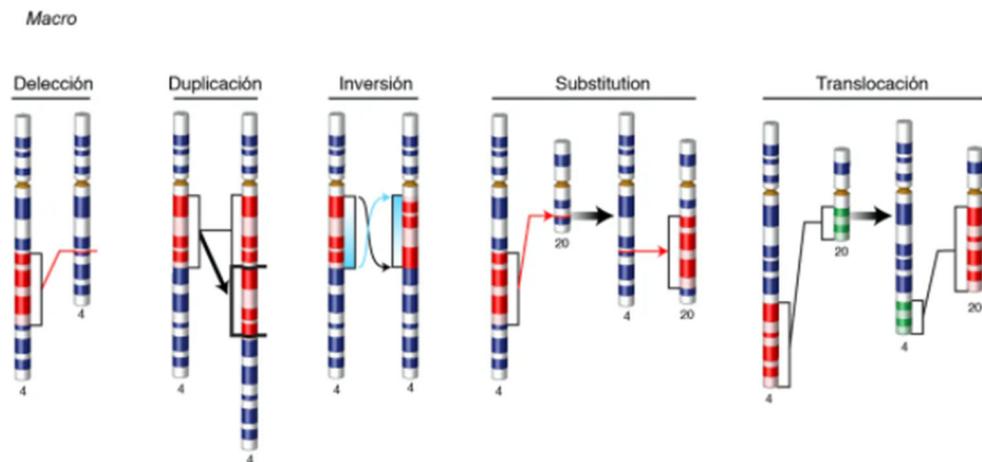


Figura 26. Ejemplos de mutaciones cromosómicas⁹.

3.- Mutaciones genómicas.

Hacen referencia a alteraciones en el número de cromosomas y no afectan solo a un gen o a un cromosoma, sino a todo el genoma. Pueden ser de distintos tipos:

- **Poliploidía:** es un tipo de mutación genómica en la que hay un aumento en el número total de “juegos de cromosomas”. Una mutación poliploide sería aquella que hiciera que el individuo no tuviera 23 pares de cromosomas, sino que tuviera, por ejemplo, 23 tripletes. Estas mutaciones son muy extrañas y no serían viables.
- **Haploidía:** es un tipo de mutación genómica en la que hay una disminución en el número total de “juegos de cromosomas”. Un individuo pasaría a tener 23 cromosomas simplemente. También son mutaciones muy raras y no son viables.
- **Aneuploidía:** es un tipo de mutación genómica en la que un cromosoma concreto está duplicado, es decir, está de más, o ha desaparecido. Hay un incremento o disminución en el número total de cromosomas, pero no afecta al conjunto entero. Pueden ser monosomías, como el síndrome de Turner, o trisomías, como el síndrome de Down (en el juego de cromosomas 21, hay un cromosoma extra). Las personas portadoras de éstas sí que es posible que nazcan, aunque su vida estará determinada por ella.

Genómica funcional

A continuación, abordaremos aquella parte de la Genómica que estudia para qué sirven los genes, cómo se relacionan entre ellos, cómo funcionan y a qué da lugar su activación o desactivación (en esta fase toman cuerpo la transcriptómica, la proteómica y la metabolómica). Distintos procesos estarán implicados: transcripción, traducción, replicación, recombinación, regulación génica.

El dogma central de la biología molecular: transcripción y traducción

La elucidación del dogma central de la biología molecular nos ha permitido conocer cómo el ADN traspa la información a la célula. Se esquematiza en la Figura 27. Esto se conoce con el nombre de expresión génica que es el proceso mediante el cual la información contenida en un gen se utiliza para dirigir el montaje de una molécula de proteína. Es la manera de como fluye la información desde el núcleo hasta el citoplasma para la síntesis de proteínas. En la Figura 27 se también se indica su relación con las “ciencias ómicas”. El resultado definitivo constituirá el fenotipo del individuo.

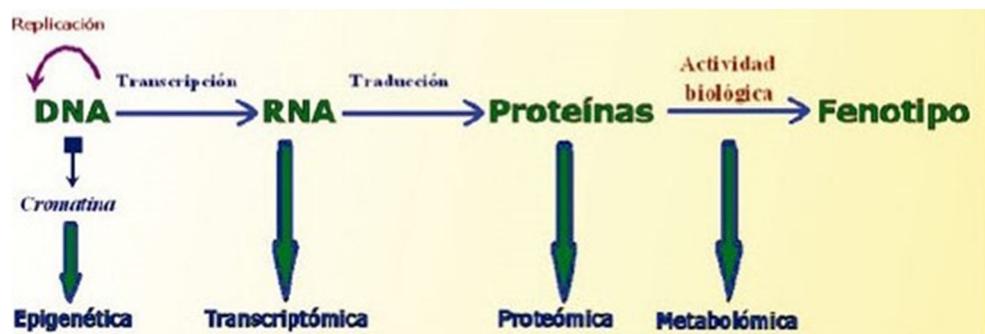


Figura 27. Dogma central de la biología molecular y su relación con las “ciencias ómicas”¹⁰.

Se necesitan dos pasos principales: la **transcripción** y la **traducción**. La expresión de los genes sigue esta secuencia.

A nivel molecular se podría esquematizar tal y como indica la Figura 28.

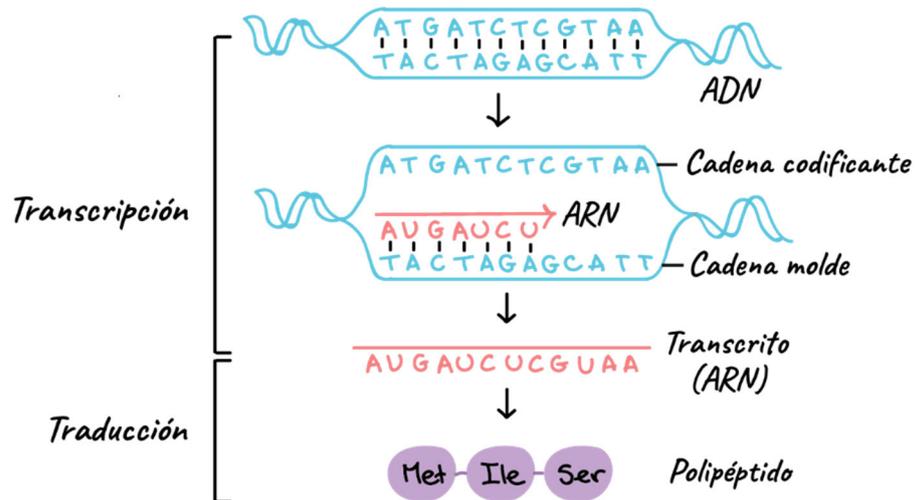
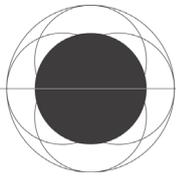


Figura 28. Transcripción y traducción.

Fuente: Khan Academy

Transcripción

El primer proceso que tiene lugar es la **transcripción** mediante la cual la información genética que contiene el gen es transcrita a un ácido ribonucleico.

Para ello, una vez la célula ha escogido el gen a partir del cual construir una proteína, la célula hace una copia de la información, de su ADN, en forma de ácido ribonucleico mensajero (**ARNm**). Los enzimas que la realizan son las ARN polimerasas.

La molécula de ARNm se va construyendo a partir del molde de ADN siguiendo la complementariedad de bases (Figura 29). Es decir, donde en el ADN hay una citosina (C), en el ARNm habrá una guanina (G). Si en el ADN hay una adenina (A), en el ARNm la timina (T) será sustituida por uracilo (U).

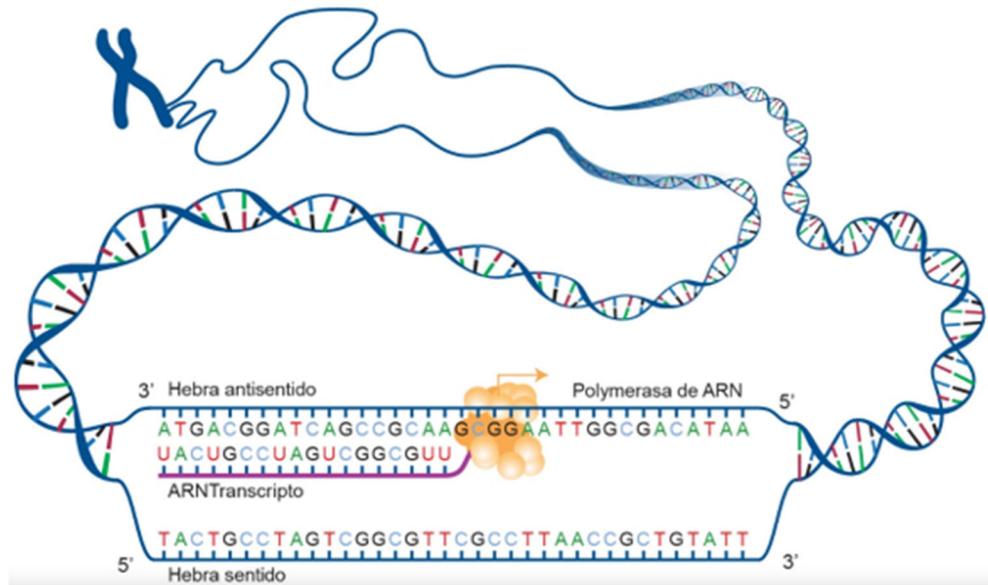
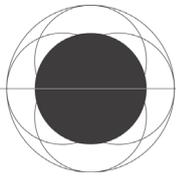


Figura 29. Representación de la transcripción del ADN³.

La transcripción utiliza una de las dos hebras expuestas de ADN como plantilla; esta hebra se conoce como hebra molde o antisentido. El producto de ARN es complementario a la hebra molde y es casi idéntico a la otra hebra de ADN, llamada hebra no molde o sentido. Como hemos comentado, hay una diferencia importante: en el ARN recién hecho, todos los nucleótidos T han sido sustituidos por nucleótidos U. El sitio en el ADN del que se transcribe el primer nucleótido se conoce como el sitio +1, o sitio de iniciación. Los nucleótidos que están antes del sitio de iniciación se marcan con números negativos y los que están después se marcan con números positivos.

La transcripción tiene tres **etapas**: iniciación, elongación y terminación.

a.- **Iniciación**: la ARN polimerasa se une a la secuencia de ADN llamada **promotor**, que se encuentra al inicio del gen. A continuación, separa las cadenas de ADN para proporcionar el molde de cadena sencilla.

b.- **Elongación**: la ARN polimerasa produce una molécula de ARN a partir de la molécula de ADN siguiendo el principio de la complementariedad de bases. Se inicia por el extremo 3' de la molécula de ADN y va en dirección 5'. La nueva cadena de ARN se construirá en sentido de 5' a 3'.

c.- Terminación: Las secuencias llamadas terminadores indican que se ha completado el transcrito de ARN. Estas secuencias provocan que el transcrito sea liberado de la ARN polimerasa. Las bases UAA, UAG y UGA son codones de terminación.

Algunas particularidades en el proceso de transcripción en seres humanos:

Como comentábamos anteriormente, los genes están formados por exones e intrones. Los exones son la porción del gen que codifica aminoácidos. Estas regiones están alternadas por una o más secuencias de ADN que llamábamos intrones, que no codifican. Así, las cadenas de ARNm transcritas incluirán las secuencias intrónicas y exónicas de su ADN patrón. Se obtendrá el ARNm primario.

Entonces, una vez concluido el proceso de transcripción, la cadena de ARNm primario deberá experimentar un procesamiento adicional de maduración antes de que pueda dirigir la traducción, el paso siguiente, que se denomina **empalme** o "**splicing**", como hemos indicado anteriormente. En este proceso, los intrones se cortan y eliminan, y las piezas restantes, los exones, se vuelven a unir, formándose el ARN maduro. También se añadirá un cabeza 5' y una cola de poli-A en sus extremos (Figura 30).

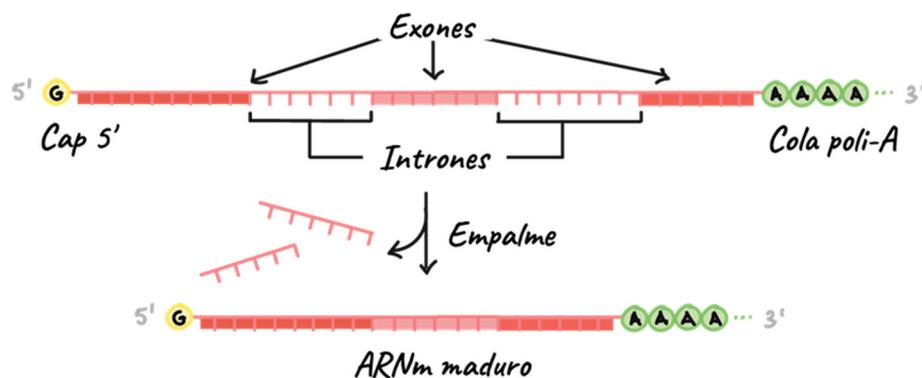


Figura 30. "Splicing" o empalme del ARNm

Fuente: Khan Academy.

El número, tamaño y posición de los intrones varía de un gen a otro.

Este ARNm maduro ya será capaz de atravesar la membrana del núcleo para dirigirse al citoplasma para continuar con el proceso.

Es interesante conocer que, al juntar diferentes combinaciones de exones, nuestras células pueden producir diferentes ARNm del mismo gen. Este proceso, conocido como “**empalme alternativo**”, permite que nuestras células utilicen la información de nuestros genes de diferentes formas. Por ello, podemos producir más proteínas que genes tenemos.

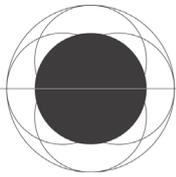
En cada extremo del ARNm hay regiones no traducidas (UTR). Los UTR se ensamblan a partir de lo que se consideran exones, aunque no codifican directamente proteínas. Sin embargo, contienen secuencias que son importantes en el proceso de formación de proteínas. Por ejemplo, muchas UTR contienen secuencias que ayudan a que el ribosoma se adhiera y se desprenda, influyen en la cantidad de proteína que se produce y afectan la vida útil del ARNm. Algunos también contienen señales de "localización", etiquetas especiales que mantienen un ARNm dentro de un área específica de la célula. Las UTR varían en tamaño desde alrededor de 100 hasta algunos miles de nucleótidos¹

Traducción

Una vez la información del ADN ha sido copiada o transcrita en forma de ARNm y, en eucariotas, ha experimentado el procesamiento o maduración, este sale del núcleo y entra en el citoplasma. Entonces se inicia la **traducción** mediante la cual el ARNm maduro se “**decodifica**” para construir una proteína que contiene una serie de aminoácidos específicos. En el ribosoma se encuentra la maquinaria celular que lee la secuencia de ARNm cada tres letras de una vez.

Este proceso se realiza en los ribosomas y también interviene el ARN de transferencia (ARNt) que es el encargado de transportar los aminoácidos hasta el ribosoma para formar la cadena polipeptídica, y adquiere una forma de trébol.

La síntesis de proteínas transcurre en cuatro etapas:



a.- Iniciación (Figura 31): Todos los ARNm comienzan con el codón AUG, que corresponde al aminoácido metionina, que no se conserva ya que cuando ha concluido la traducción es desprendido de la cadena. En el ribosoma existen dos regiones, la correspondiente al denominado sitio aminoacil (sitio A) y la correspondiente al sitio peptidil (sitio P). En el sitio P se colocarán los ARNt con la cadena polipeptídica en crecimiento y en el sitio A, los ARNt con el nuevo aminoácido que se vaya a incorporar en la cadena. Al comienzo de la traducción, el ribosoma se posiciona colocando en el sitio P el ARNt con el codón de inicio.

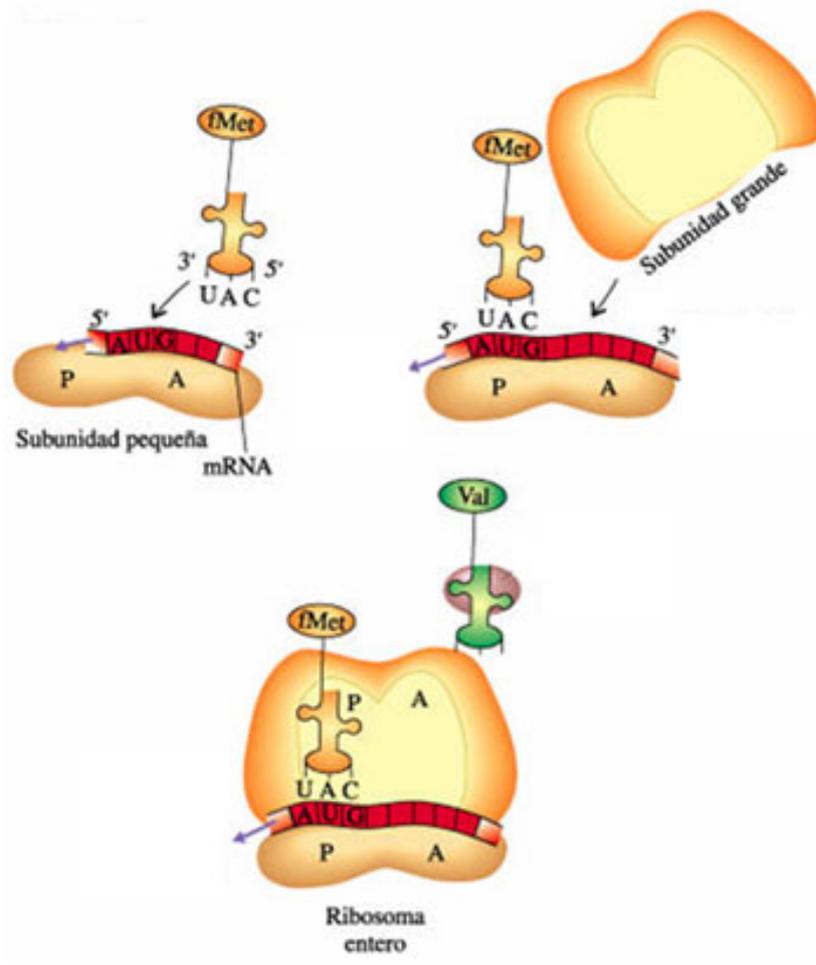


Figura 30. Iniciación de la traducción¹¹.

b.- Elongación (Figura 31): a continuación, llega el siguiente ARNt con un anticodón complementario del codón de ARNm colocado en el sitio A. Un enzima específico se encargará de establecer el enlace peptídico entre ambos

aminoácidos, formándose el dipéptido unido al sitio A.

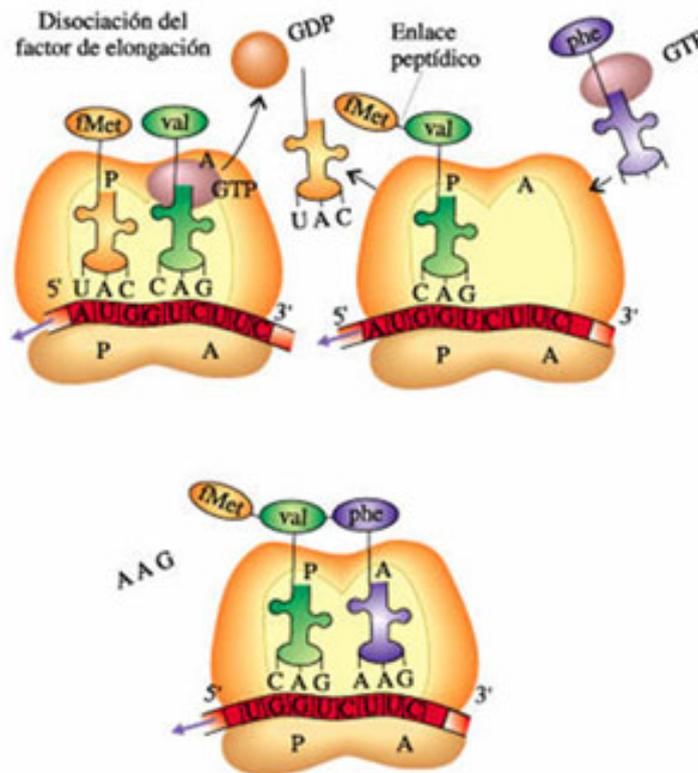


Figura 31. Elongación en la traducción¹¹.

c.- Traslocación: el ribosoma se desplaza a lo largo del ARNm, de forma que el ARNt con el dipéptido pasa al sitio P, expulsando el ARNt vacío (Figura 32).

d.- Terminación (Figura 32): el proceso de elongación se repite hasta que el codón de parada (en este caso el codón UGA) entra en el sitio A. La elongación finaliza y el complejo se disocia liberando el polipéptido.

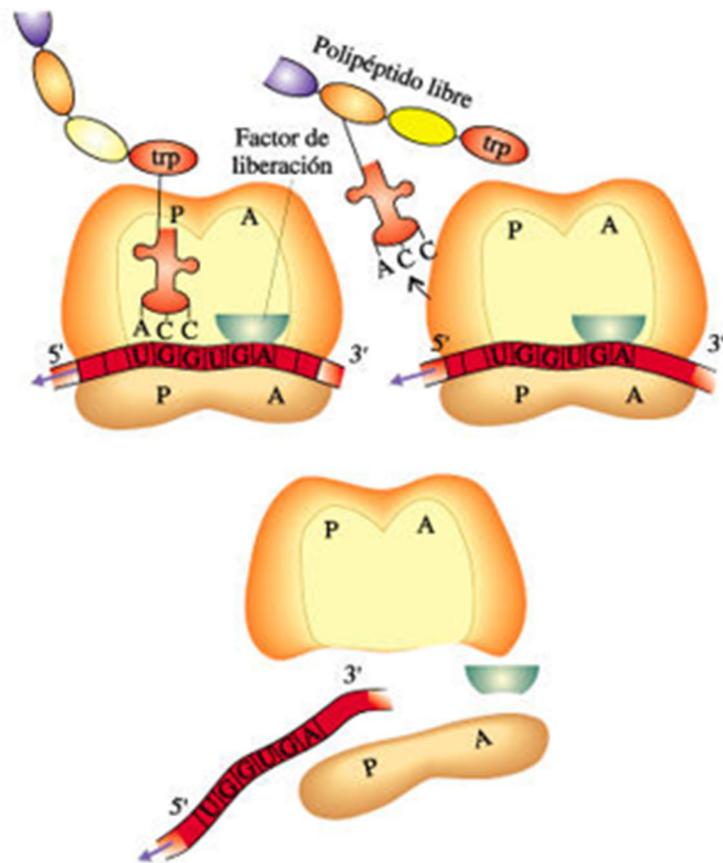
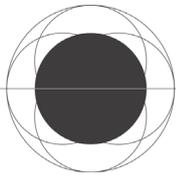


Figura 32. Traslocación y terminación de la traducción¹¹.

Replicación

La replicación del ADN es el proceso mediante el cual se duplica una molécula de ADN. Cuando una célula se divide (en la mitosis o en la meiosis), en primer lugar, debe duplicar su genoma para que la nueva célula contenga un juego completo de cromosomas. Esto quiere decir que cada cadena de la doble hélice del ADN puede hacer una copia de ella misma, duplicando la cadena de nucleótidos.

La replicación del ADN¹² es probablemente uno de los trucos más impresionantes que hace el ADN. Cada célula contiene todo el ADN que necesita para fabricar las demás células. De hecho, empezamos siendo una sola célula y terminamos con billones de células. Y durante ese proceso de división celular, toda la información de una célula tiene que ser copiada; y tiene que ser copiada a la perfección. Por tanto, el ADN es una molécula que puede ser replicada para hacer copias casi

perfectas de sí misma. Y eso es sorprendente teniendo en cuenta que hay casi tres mil millones de pares de bases de ADN para ser copiadas.

La replicación del ADN requiere de la participación de diferentes enzimas. La elongación de la nueva hebra se lleva a cabo por un enzima perteneciente al grupo de las **ADN polimerasas**. Replicar todo el ADN de una sola célula humana lleva varias horas. En primer lugar, las dos hebras se separan y se inicia la replicación del ADN en múltiples ubicaciones para acelerar el proceso y en ambas direcciones a una velocidad de aproximadamente 50 nucleótidos por segundo. Estas múltiples ubicaciones se denominan burbujas de replicación (Figura 33).

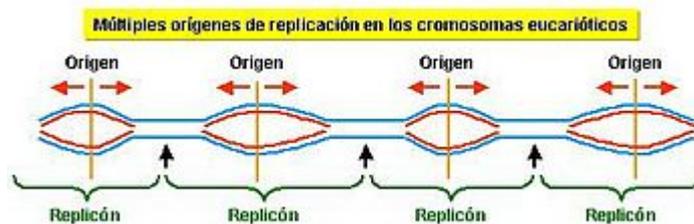


Figura 33. Burbujas de replicación del ADN.

Fuente: Wikipedia.

En los extremos de la burbuja es donde empieza la replicación en una estructura que se denomina horquilla de replicación. Así, la doble hélice de ADN parental se divide en dos hebras simples, que después se copian. Para el inicio de la síntesis de la nueva hebra se usa un ARN cebador. Una de las hebras que se va copiando se llama hebra conductora, y la nueva copia de esta se sintetiza en dirección $5' \rightarrow 3'$, y la otra hebra se denomina hebra retardada (Figura 34).

Los cromosomas apareados de los progenitores masculino y femenino se alinean de forma que secuencias similares del ADN pueden girarse y entrecruzarse. Este entrecruzamiento produce un intercambio de material genético, el cual es causa importante de la variabilidad genética que se observa en la descendencia. Así, los descendientes tienen combinaciones genéticas diferentes a la de sus progenitores³.

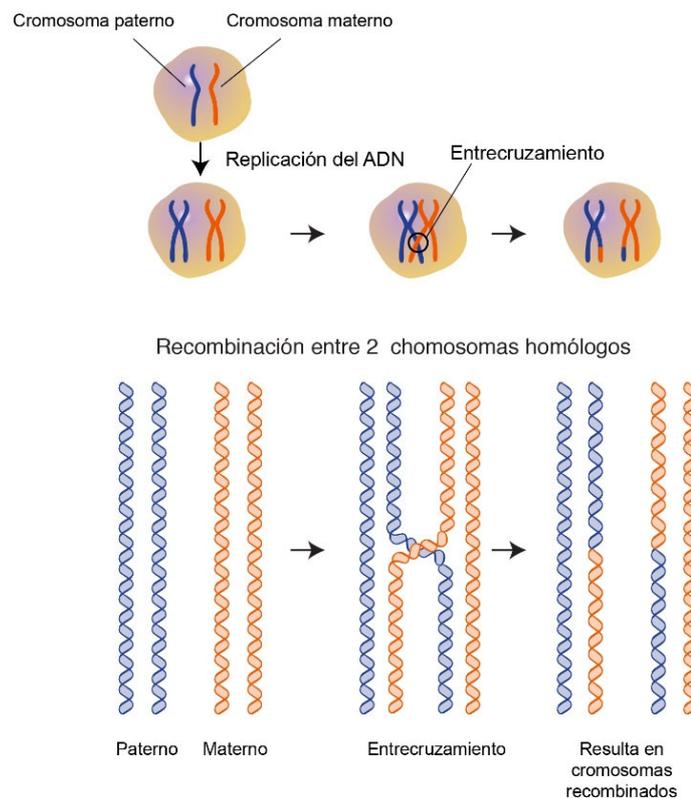


Figura 35. Recombinación homóloga³.

La recombinación genética permite también que, en determinadas células, como las células B y T del sistema inmune, se pueda generar una tremenda variabilidad de anticuerpos y de receptores de células T, importantes para la respuesta inmune.

Regulación génica

Nuestro organismo fue una vez una sola célula (un óvulo fecundado). Ahora tenemos billones de células, cada una con una función y una estructura especializada. Tenemos células del músculo cardíaco, células pulmonares, nerviosas, cutáneas y muchas más. Aunque parezcan distintas, todas ellas contienen los mismos genes. La expresión diferencial de genes es la responsable de crear distintos tipos de células, que se disponen en tejidos, que coordinan su actividad para formar un individuo. Las células óseas son distintas de las células sanguíneas no porque contengan genes diferentes, sino porque expresan genes distintos. Las diferentes células de nuestro organismo expresan conjuntos únicos de genes porque responden a distintas señales que, en el ser humano, proceden de otras células, es decir, de un ambiente interno. Se calcula que cada célula utiliza sólo el 10% de sus genes; o sea, se puede tener un gen y no utilizarlo, permanece en silencio. Del mismo modo, cada individuo expresa parte de sus genes de forma distinta, desde el silencio más absoluto hasta la formación de productos diferenciados. Con lo cual, lo que estamos afirmando es que ese proceso anteriormente propuesto como el “dogma” de la biología molecular es un proceso influenciado, maleable, polifacético. No es un proceso rígido e impenetrable sino ampliamente variable sobre el que diversas fuerzas o estímulos van a incidir y regular la dirección en que se mueva.

La expresión génica se puede controlar a través de distintos mecanismos y en todos los niveles del procesamiento del ADN (transcripción, traducción, replicación), a nivel del ARNm, en la activación o inactivación de proteínas (post-traducción)¹². Pero además hay niveles adicionales que tienen relación con la estructura dinámica de la cromatina y marcas moleculares de la cadena de ADN. En la Figura 36 se esquematizan el modelo simple y el modelo complejo del “dogma” central de la Biología Molecular. El modelo complejo incluye algunos elementos de la regulación de la expresión génica.

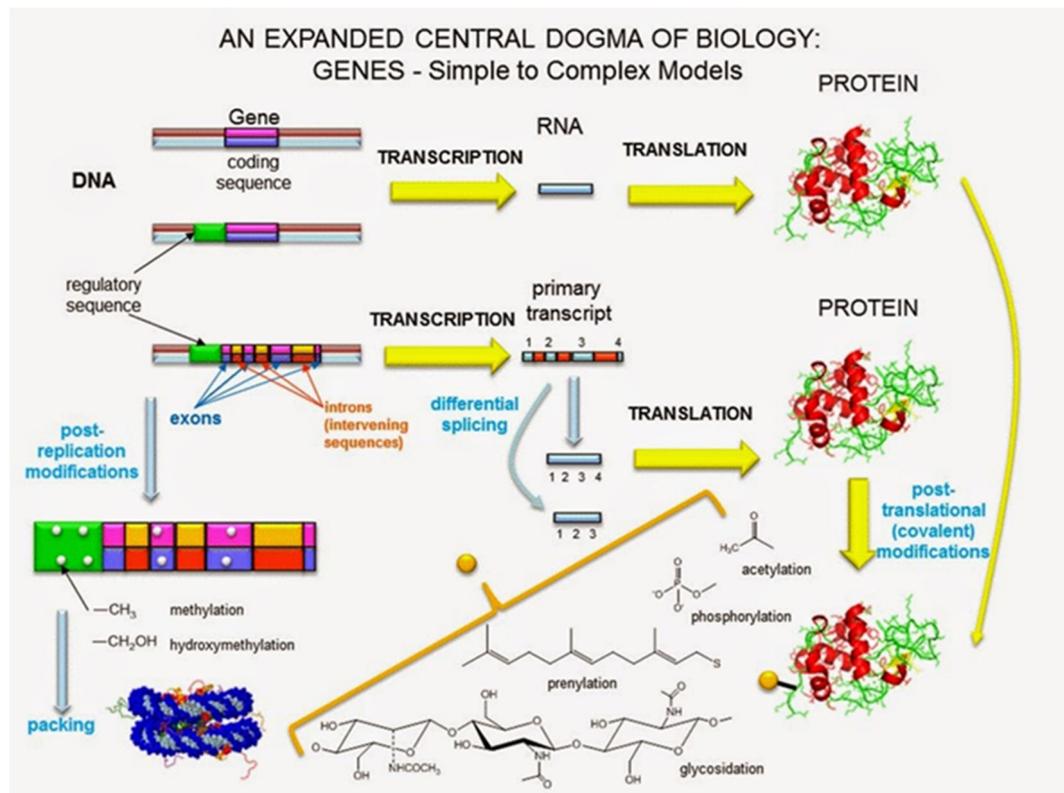


Figura 36. Modelo simple y complejo del “dogma” central de la biología molecular¹³.

A continuación, se indican los mecanismos de regulación genética en los distintos niveles de la expresión genética:

a.- Regulación pretranscripcional.

a.1.- A nivel de reestructuración o acceso al ADN, de forma que la **estructura de la cromatina** define la expresión genética. Recordemos que el ADN en el núcleo está densamente empaquetado junto con proteínas (histonas), formando la cromatina. El enzima encargado de la transcripción del ADN, la ARN polimerasa, no puede acceder al ADN cuando se establece esta interacción física entre el ADN y las histonas. Por tanto, previa a la transcripción, se produce una modificación de la estructura de la cromatina.

- . Un grupo de proteínas encargadas de la modificación de la cromatina son los **complejos de remodelado de la cromatina** que, mediante una serie de reacciones **dependientes de ATP**, afectan a la estructura del nucleosoma de una u otra manera. Estos complejos remodeladores

desempeñan un papel fundamental en la expresión génica regulando el acceso de la maquinaria de la transcripción. Una regulación incorrecta de estos se pone de manifiesto en ciertas patologías o pueden ser requeridos para la actividad de otros genes supresores de tumores.

. Un segundo grupo principal de proteínas funcionan **añadiendo grupos** acetilo o metilo a las histonas. Estos procesos se llaman acetilación y metilación, respectivamente. También se han descrito procesos de fosforilación, ubiquitinación, ADP-ribosilación y sumoilación de histonas, con unos efectos locales muy importantes sobre la cromatina. Estas modificaciones se dan sin alterar la secuencia de nucleótidos del ADN sino a nivel estructural por encima de la estructura del ADN. Por ello se denominan modificaciones epigenéticas.

Por un lado, tenemos la **acetilación de las histonas**. Las histonas acetil transferasas (HAT) son las encargadas de adicionar grupos acetilo en los residuos de lisina de las histonas cargados positivamente. Entonces, el número de carga positivas de las histonas se reduce. El resultado es una menor atracción electrostática entre las histonas y el ADN cargado negativamente. La asociación entre nucleosomas se debilita y la cromatina se hace menos densa. Por tanto, la acetilación de las histonas conduce al desenrollamiento de la cromatina (**euromatina**) haciendo así accesibles los enzimas de la transcripción. La cromatina se “recondensa” por la acción de las histonas desacetilasas (HDAC) que eliminan el grupo acetilo añadido por las HAT. Por tanto, la desacetilación de histonas conduce a una estructura condensada o cerrada de la cromatina (**heterocromatina**) y a una menor transcripción de esos genes. Si las HAT son la llave que abre la puerta a la transcripción, la HDAC son la llave que la cierra.

Así, la acetilación de histonas se asocia generalmente con un control positivo, es decir, activación de genes. En la Figura 37 se representa la cromatina como estructura cerrada (desacetilada), asociada a transcripción inhibida, y abierta (acetilada), asociada a transcripción permitida.

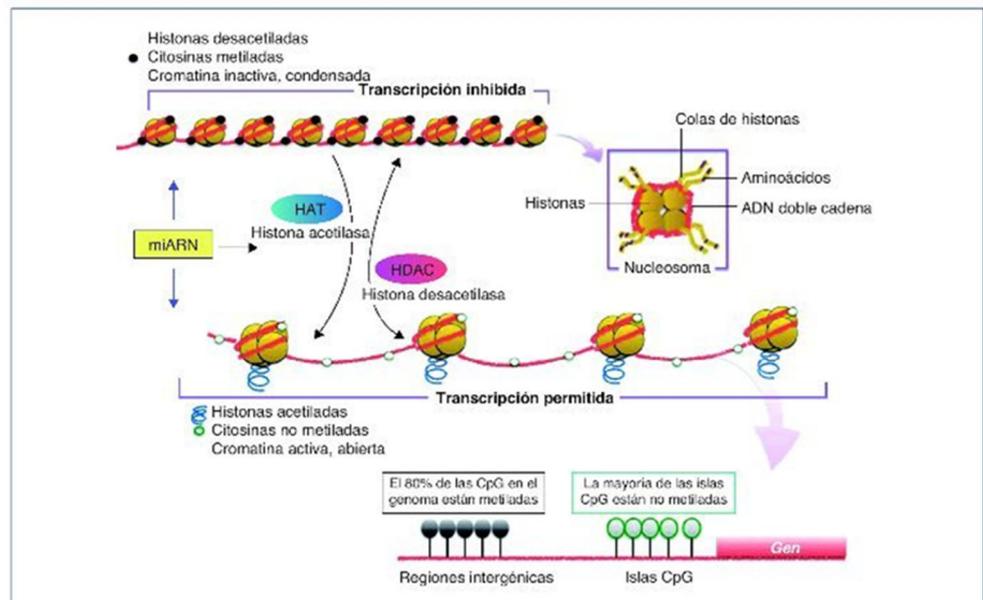


Figura 37. Representación de diversas modificaciones epigenéticas¹⁴.

La metilación de histonas puede estar relacionada con activación o inactivación, dependiendo de qué histonas se alteren y dónde se añada el grupo metilo en la proteína. La metilación puede echar el freno de mano o quitarlo, según las condiciones.

El patrón de modificaciones químicas que ocurren en las histonas varía de un tipo celular a otro y se transmiten a las células hijas. Así, las células musculares son distintas de las células nerviosas en parte porque heredaron distintos tipos de histonas modificados, no distintos genes.

a.2.- **Marcas moleculares de la cadena de ADN.** Consiste en la unión de grupos **metilo** en la cadena de ADN.

A nivel molecular se trata de una unión covalente de un grupo metilo al carbono 5' de citosinas del ADN¹⁵. En animales la metilación se da exclusivamente en el contexto del dinucleótido CpG (citosinas seguidas de guanina, -citosina-fosfato-guanina dinucleótido). En el genoma humano, los dinucleótidos CpG habitualmente se concentran en regiones llamadas islas CpG que, preferentemente, se encuentran en las regiones promotoras del gen, estando generalmente no metiladas. Las CpG localizadas fuera de las islas, por lo general están metiladas y se cree que participan en la estabilidad del cromosoma. En la Figura 37 también se representan las metilaciones y no metilaciones de las CpG.

Los patrones de metilación se establecen durante el desarrollo embrionario. Durante el desarrollo preimplantacional se produce una desmetilación masiva del ADN y una reorganización de las marcas de las histonas, seguida de una metilación *de novo* del ADN tras la implantación del embrión, que determina el patrón epigenético del nuevo individuo. Esta reprogramación epigenética parece ser necesaria para la totipotencia celular, para la expresión génica y para el desarrollo temprano de los linajes celulares en el embrión.

La metilación del ADN la llevan a cabo unos enzimas llamados DNMTs (del inglés, “*DNA Methyltransferases*”) que catalizan la transferencia de un grupo metilo desde la S-adenosil-L-metionina (SAME) hasta la posición 5’ del anillo pirimidínico de la citosina. ¿Cómo puede afectar la metilación del ADN en la expresión de los genes? La información contenida en la metilación de la citosina es interpretada por una familia de proteínas que comparten un dominio de unión a ADN metilado llamado MBD (del inglés “*Methyl-CpG Binding Domain*”). La proteína MBD pueden reclutar proteínas adicionales para el locus o la ubicación particular de un cromosoma, como las enzimas desacetilasas y otras proteínas de la remodelación de la cromatina, y esto modifica las histonas, formando heterocromatina condensada que es transcripcionalmente silenciosa.

La metilación del ADN altera de manera estable la expresión de los genes. Por eso, se mantiene en el proceso de división de la célula en un tejido concreto. Es decir, el patrón de metilación pasa de la célula madre a la célula hija. También la metilación del ADN está implicada en la impronta genómica y en la inactivación del cromosoma X.

La metilación anormal del ADN se ha implicado en las carcinogénesis o desarrollo de cáncer, por lo que podemos ver que la regulación normal de la metilación del ADN es un mecanismo regulador crítico para nuestras células.

Como conclusión de la regulación de la expresión genética a nivel pretranscripcional, podemos decir que va a estar muy relacionada con el nivel de empaquetamiento del ADN y las histonas en el complejo conocido como cromatina. Por tanto, la función de la cromatina no sólo será organizar el empaquetamiento del material genético en el núcleo de la célula. La cromatina en realidad va a ser una estructura muy dinámica que afecta a todos los procesos metabólicos relacionados con el ADN y en la regulación de la expresión génica.

b.- Regulación transcripcional.

b.1.- **Los factores de transcripción.** El inicio del proceso de transcripción, que hace que un gen comience a actuar, es un proceso complejo en el que participan muchas unidades proteicas que, adecuadamente ensambladas entre sí y de ellas con el gen, le dan la señal para que inicie su actividad¹⁶, como veremos a continuación.

La regulación de la expresión genética a nivel de la transcripción es un mecanismo de control ubicuo en muchos procesos biológicos, como el crecimiento, la diferenciación celular y el metabolismo. Durante la transcripción se requiere de factores que regulan la fuerza del promotor, los cuales se denominan factores de transcripción (**FT**). Los factores de transcripción son proteínas que reconocen y se unen al ADN, reconociendo una secuencia específica, pero no forman parte de la ARN polimerasa, que es el enzima que va a llevar a cabo la transcripción uniéndose al promotor para iniciar la transcripción (Figura 38). El promotor contiene una secuencia de bases específica denominada caja TATA. A los promotores se une la misma proteína: proteína de unión a TATA (TBP). Estos son factores basales o generales de transcripción. Se unen directamente al ADN. TBP es un factor basal de transcripción (**FT general**).

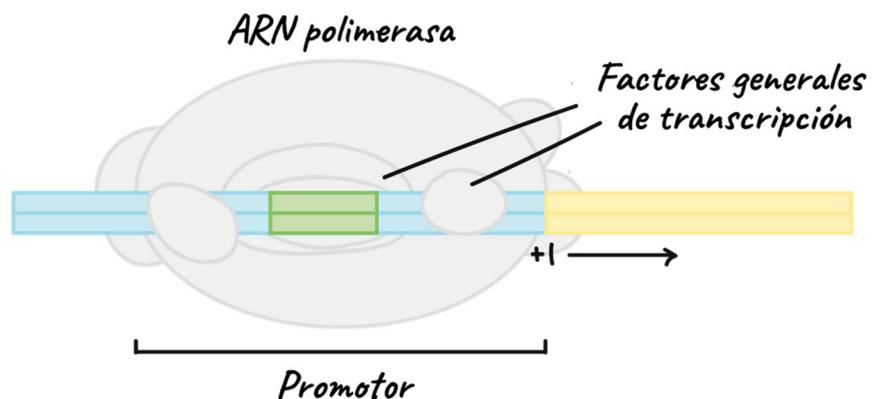


Figura 38. ARN polimerasa se une al promotor al inicio de la transcripción.

Fuente: Khan Academy

Los factores generales de transcripción (FT general) no pueden por sí mismos incrementar o disminuir la tasa de transcripción. ¿Cómo se puede controlar entonces la transcripción? La respuesta está en las interacciones entre las secuencias reguladoras distintas del promotor y otras proteínas reguladoras además de la TBP, que son los factores de transcripción regulatorios o específicos (**FT específicos**).

Las secuencias reguladoras pueden estar cerca del promotor y son únicas para genes específicos. De este modo, componen un mecanismo de control preciso sobre la transcripción.

Algunas secuencias reguladoras están lejos del promotor y se denominan intensificadores. Estos están asociados a distintos genes. Los intensificadores son un acelerador, un elemento de control positivo. Cuando las proteínas reguladoras (FT específicos) se unen a los intensificadores, empieza la transcripción.

Otras secuencias reguladoras de estructura similar a los intensificadores tienen una función contraria. Estas secuencias son silenciadores. Cuando las proteínas reguladoras (FT específicos) se unen a los silenciadores, la transcripción se apaga.

En base a la estructura del dominio de unión al ADN, los FT específicos se han podido clasificar en familias, como los factores con dedos de zinc, hélice-giro-hélice, cremallera de leucina, entre otros¹⁶.

El descubrimiento de los intensificadores y los silenciadores replanteó a los biólogos la definición de gen que incluye también las secuencias reguladoras para la expresión, además de la que codifica un polipéptido o una molécula de ARN funcional.

Y el control de la expresión diferencial de genes se encuentra en la presencia de proteínas reguladoras exclusivas (FT específicos) que hacen que una célula muscular sea una célula muscular y una célula ósea, sea una célula ósea.

Así, los FT específicos están implicados en la regulación de la transcripción genética tejido-específica en respuesta a diversos estímulos. Estos estímulos, que pueden ser extracelulares, pueden activar los FT específicos.

En el inicio, los FT específicos reclutan complejos de remodelado de la cromatina y HAT. El resultado es un “licuado” de la estructura de la cromatina, abriendo una amplia franja que incluye la región del promotor, al que se dirigen los FT generales.

Además, otras proteínas llamadas coactivadores participan en el inicio de la transcripción. No se unen al ADN (como TBP) pero unen las proteínas implicadas en iniciar la transcripción (TBP) y los FT específicos. Se forma así el complejo basal de la transcripción, siendo el resultado una maquinaria molecular que se sitúa en el lugar de inicio y es capaz de empezar la transcripción (Figura 39).

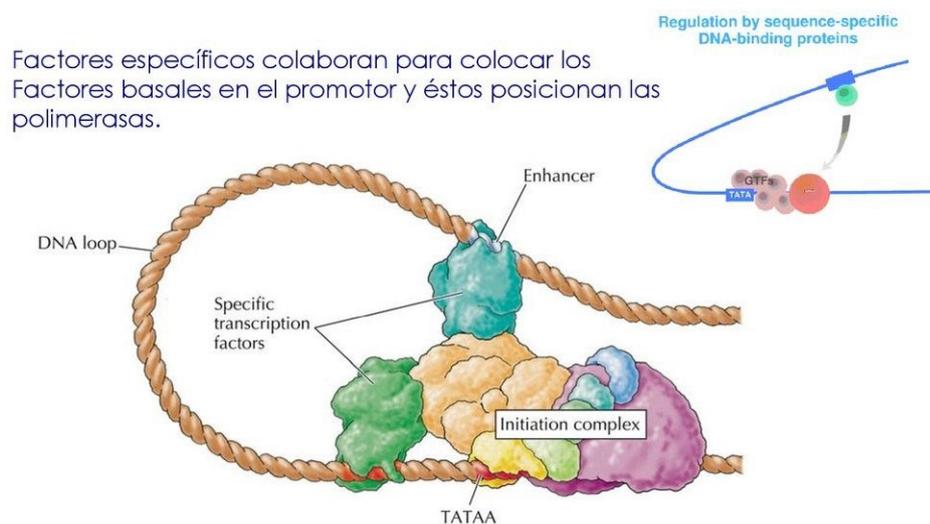


Figura 39. Complejo basal de la transcripción.

Fuente: Departamento de Genética. Universidad Autónoma de Honduras.

b.2.- **Procesamiento del ARNm.** Durante el procesamiento del ARN, cuando se produce el corte y empalme ("**splicing**"), los intrones son eliminados de los transcritos primarios de ARN mientras el mensaje está todavía en el núcleo. Pero, además, los cambios en la expresión génica son posibles porque se pueden eliminar exones seleccionados, además de intrones (empalme alternativo). Como resultado, del mismo transcrito primario de ARN pueden obtenerse ARNm maduros y procesados con varias combinaciones distintas de exones transcritos. Así, los genes pueden dirigir la producción de uno o más polipéptidos o ARN relacionados¹².

b.3.- Una vez completado el ajuste, los ARM maduros se exportan al citoplasma y pueden entrar en juego nuevos **mecanismos reguladores relacionados con la vida útil del ARNm**. Esta está controlada en muchos casos por la unión de una hebra simple de minúsculos ARN denominados microARN o miARN. Este fenómeno se conoce con el nombre de interferencia de ARN. El ARNm unido a estos miARN no se traduce (Figura 40).

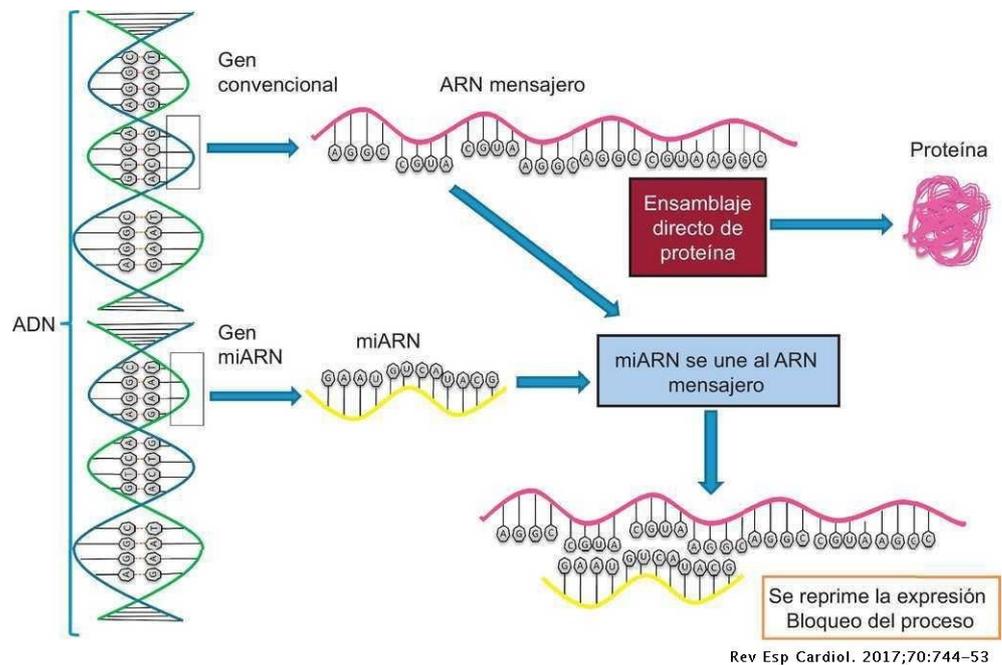


Figura 40. Representación del efecto de los miARN en la expresión de los genes.

b.4.- Hay **otro tipo de ARN no codificante** que puede tener un papel en la regulación génica. Son los ARN pequeño nuclear o ARNs. Su tamaño es de aproximadamente 150 nucleótidos. Ayudan a regular los factores de transcripción. También ayudan a mantener los telómeros.

c.- **Regulación postranscripcional.** Se refiere a la actividad de la proteína. Las proteínas pueden someterse a una variedad de modificaciones, tales como ser cortadas o etiquetadas con grupos químicos (acetilación, fosforilación, glicosilación). Se representan en la Figura 36. Estas modificaciones pueden ser reguladas y pueden afectar la actividad o el comportamiento de la proteína. Los mecanismos reguladores que tienen lugar al final del flujo de información desde el ADN al ARN y las proteínas implican un intercambio entre rapidez y uso de recursos, porque la transcripción y traducción requiere energía y nutrientes¹².

d.- **Regulación en la replicación.** También se pueden dar mecanismos de regulación génica en la replicación o duplicación del ADN. Como hemos indicado anteriormente, en este proceso intervienen proteínas del grupo de las ADN polimerasas. Este grado adicional de complejidad - el hecho de que el ADN codifique una proteína que permite al ADN replicarse- es importante porque proporciona un principio de regulación determinante. La replicación del ADN se

puede activar o desactivar por medio de otras señales o reguladores como, por ejemplo, la edad o el estado nutricional de la célula, que permite a las células hacer copias de ADN solo cuando están en condiciones de dividirse. Si este sistema no funciona, si los reguladores se descontrolan, no hay nada que haga que una célula se replique sin parar. Este es el peor mal de los genes defectuosos: el cáncer.

e.-**Regulación** de la expresión génica **a través** de la intervención de **receptores nucleares**¹⁰. Los receptores nucleares que controlan la expresión de los genes son proteínas que pueden asociarse a diversos ligandos (glucocorticoides, mineralocorticoides, hormonas sexuales, hormonas tiroideas, ácidos grasos poliinsaturados, entre otros), normalmente de carácter lipofílico. Se agrupan en receptores endocrinos, receptores huérfanos (debido a que se desconocen sus ligandos) y receptores “huérfanos adoptados” (se han identificado sus ligandos). Los receptores más conocidos y estudiados son los llamados receptores activadores de la proliferación peroxisomal (PPARs, del inglés “*Peroxisome Proliferator Activated Receptors*”), los receptores X de ácido retinoico (RXR), los receptores para ácido retinoico (RAR), y los receptores X hepáticos (LXR).

Un receptor nuclear típico contiene un dominio de interacción con el ADN (DBD), otro para interactuar con el ligando (LBD) y otros para interactuar con activadores funcionales de la transcripción (AF). Los receptores nucleares se unen a sectores del ADN conocidos como elementos de respuesta (RE), a activadores y/o inhibidores y a ligandos específicos (Figura 41).



Figura 41. Estructura general de un receptor nuclear¹⁰.

Forman homodímeros o heterodímeros que interactúan con el DNA. La unión del receptor al ADN es determinada por la presencia de ligandos específicos. El

resultado final de este complejo proceso produce la activación o la represión de la expresión de un gen. Adicionalmente, los receptores nucleares actúan normalmente asociados a cofactores, que pueden activar o reprimir la expresión del gen regulado por el receptor. También pueden determinar una remodelación de la cromatina (cambian su estructura espacial).

El patrón de la herencia genética

Aquí nos referimos a cómo se produce la transmisión de rasgos externos hereditarios o fenotípicos, como puede ser la transmisión de una enfermedad. Esta transmisión responde a unos patrones, como por ejemplo la herencia mendeliana que se refiere a la transmisión de un único gen mediante un patrón dominante, recesivo o ligado al cromosoma X. Pero algunas enfermedades hereditarias no siguen el patrón clásico de herencia mendeliana por lo que se han definido otros patrones de transmisión de caracteres hereditarios, en especial referidos a la herencia multifactorial y a la herencia mitocondrial. La herencia multifactorial está basada en los efectos conjuntos de los genes y los factores ambientales. La herencia mitocondrial extracelular solo es transmitida por la madre. Se describen a continuación los patrones de la herencia genética para las enfermedades hereditarias.

-Herencia mendeliana monogénica.

. Herencia autosómica dominante. Cuando el gen responsable de la patología se encuentra en uno de los 22 cromosomas no sexuales o autosomas y, además, una copia simple del gen mutado es suficiente para que se exprese la enfermedad. La copia alterada del gen procede de uno de los progenitores. Normalmente se manifiesta en todas las generaciones de una misma familia. Unos ejemplos serían el enanismo o acondroplasia, la enfermedad de Huntington

(Figura 41).

Enfermedad de Huntington

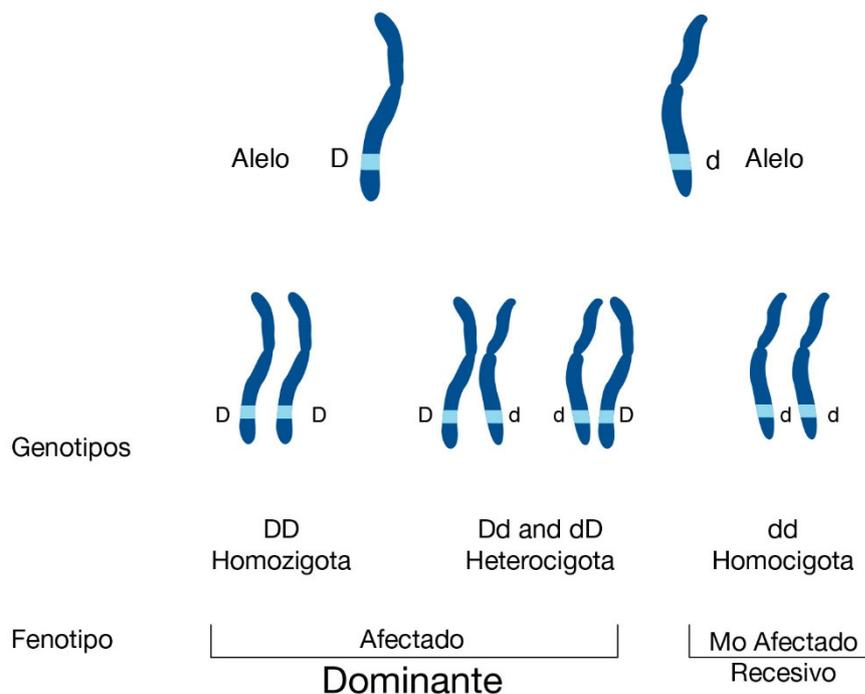


Figura 41. Ejemplo de herencia autosómica dominante³.

. Herencia autosómica recesiva. Cuando el gen responsable de la patología se encuentra en uno de los 22 cromosomas no sexuales o autosomas, sin embargo, son necesarias dos copias del gen mutado para que se exprese la enfermedad. Las enfermedades con este tipo de herencia no suelen manifestarse en todas las generaciones y suelen saltarse una generación. Un ejemplo sería la fibrosis quística, la fenilcetonuria, intolerancias a los azúcares (galactosa, fructosa, sacarosa y lactosa), deficiencias de los factores I, II, V, XII y XIII, anemia falciforme, talasemia, ...

. Herencia ligada al cromosoma X dominante. Cuando el gen alterado domina sobre el normal, por lo que una única copia de este es suficiente para que se desarrolle la enfermedad y, además, se encuentra en el cromosoma X. Afecta tanto a hombres como a mujeres, pero es más probable que las mujeres sufran la enfermedad. Un ejemplo es el raquitismo hipofosfatémico.

. Herencia ligada al cromosoma X recesiva: Cuando el gen mutado sigue encontrándose en el cromosoma X, pero es recesivo sobre el sano, lo que conlleva que se necesiten dos copias del gen para que se dé la enfermedad. Se produce con mayor frecuencia en hombres dado que tiene un solo cromosoma X, y si heredan el gen mutado sufrirán la enfermedad. En cambio, en mujeres, como tienen dos copias del cromosoma X, sufrirán la enfermedad si heredan las dos copias del gen mutado, y si heredan una sola copia, sólo serán portadoras. Un ejemplo es la distrofia muscular de Duchenne y otro es la hemofilia. A y B, el daltonismo, deficiencia de glucosa-6-fosfatodeshidrogenasa (G6PD)

-Herencia multifactorial

. Multialélico: existen varios alelos posibles en el mismo locus, cada individuo tiene dos posibles alelos, y la transmisión se realiza como el patrón monogénico.

. Multifactorial: para un determinado carácter, hay una serie de genes implicados. Su estudio es matemático y complejo. Contribuyen los factores ambientales, el peso del individuo, cardiopatías, epilepsia. Ejemplos de enfermedades multifactoriales son el paladar hendido, esquizofrenia, enfermedades cardiovasculares, diabetes, gota, dislocación de la cadera, estrabismo, psoriasis, etc.

-Herencia mitocondrial: No es una herencia mendeliana, es una herencia materna. Existen enfermedades hereditarias debidas a mutaciones en genes mitocondriales. Las citopatías mitocondriales son generalmente deletéreas con síntomas pleiotrópicos (múltiples), ya que el déficit va a implicar a varios órganos: insuficiencia pancreática exocrina, déficit muscular, hepático, renal y enfermedades gastrointestinales.

Una enfermedad causada por una alteración en un gen mitocondrial se transmite únicamente por mujeres, a todos sus descendientes, hombres y mujeres, y son de expresividad variable. Algunos ejemplos serian, la atrofia óptica de Leber, miopatías mitocondriales, el síndrome de Pearson...

Epigenética

Mención especial en un apartado aparte merece este campo emergente de la ciencia que estudia los cambios hereditarios causados por activación y desactivación de los genes sin ningún cambio en la secuencia de ADN subyacente del organismo. Estos cambios se denominan marcas epigenéticas. Nos hemos referido a ellas en el apartado de regulación de la expresión génica ya que afectan a la expresión de los genes pues permiten la activación o represión de estos jugando un papel relevante en números procesos biológicos.

Se estima que en el ser humano existen 200 tipos de células diferentes. Todas ellas con el mismo genoma. Existe un control por encima de la secuencia de ADN que decidirá qué genes se expresarán. La epigenética tendrá un papel primordial en la identidad celular. Cada tipo celular tendrá un epigenoma distinto. Pero también, mediante mecanismos epigenéticos, la expresión genética de las células podrá ser modulada por factores externos e internos (medio ambiente, edad temprana, edad longeva, ciertas patologías), lo que no ocurre con la secuencia del ADN, que no puede cambiar. Estos cambios epigenéticos son susceptibles de ser modificados, lo que abre un amplio horizonte de posibilidades terapéuticas hoy en día insospechadas¹⁴.

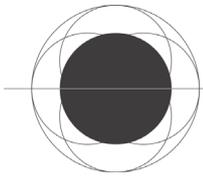
El epigenoma es la colección de todas las marcas epigenéticas en el ADN de una sola célula. Como hemos indicado, las marcas epigenómicas son diferentes en distintos tipos de células. Por ejemplo, una célula de la sangre puede tener diferentes marcas o modificaciones respecto a una célula del hígado. Así, las modificaciones epigenómicas, o sea, la colección completa de todas las marcas epigenéticas en el ADN de las células de la sangre de un individuo debe ser más similar a las marcas en el ADN de las células de la sangre de otro individuo, que la colección de todas las marcas en el ADN de las células del hígado. Esta es una forma de definir un tipo particular de célula. Estas características epigenéticas van a ser dinámicas durante el desarrollo. Y, ahora podemos decir que las diferencias individuales en el epigenoma nos hacen a todas las personas individuales, y vamos a ver cambios aún mayores en un estado de enfermedad. Así, si comparamos una célula normal y todas sus marcas epigenéticas, con células cuando se presenta una enfermedad, sus marcas epigenéticas serán distintas. Y podemos utilizar estas diferencias para averiguar los mecanismos de dicha enfermedad³.

Se ha desarrollado el Proyecto Epigenoma Humano (HEP) cuyo objetivo es identificar, catalogar e interpretar los patrones de metilación del ADN de todos los

genes humanos, en todos los tejidos importantes². La metilación es un parámetro flexible que puede cambiar la función de los genes por influencia exógena. Constituye el principal eslabón entre genes, enfermedad y ambiente. Se considera que juega un papel decisivo en la etiología de virtualmente todas las patologías humanas. Las citosinas diferencialmente metiladas dan lugar a distintos patrones específicos para cada tejido y estado de enfermedad, como hemos indicado. Estas posiciones de metilación variable (MVPs) son los marcadores epigenéticos más comunes y, como los SNPs, nos proporcionan avances significativos en la capacidad de entender y diagnosticar la enfermedad humana. Otro cambio epigenético con remodelación de la cromatina incluye la acetilación, fosforilación, ubiquitinación y ADP-ribosilación de las histonas, lo que modifica el acceso a la maquinaria de transcripción. Este cambio también se puede ver afectado por la acción del medio ambiente. Y otras modificaciones epigenéticas son realizadas por ARNs no codificantes (los miARN, por ejemplo) que se estima que regulan el 30% del genoma humano¹⁴. Recientemente se ha conocido cómo ciertos miARN podrían ejercer un papel modulador-regulador interactuando con las metiltransferasas, contribuyendo a la metilación del ADN, o por destrucción de ciertos ARNm que hubieran podido eludir la represión. También se ha estudiado como factores externos como la dieta pueden modificar la formación de miARN específicos que pueden bloquear o inducir un crecimiento canceroso.

Como resumen podemos decir que el epigenoma se refiere a las modificaciones del ADN (metilación, por ejemplo), interacciones proteína-ADN, las modificaciones de las histonas y la organización de la cromatina que, en conjunto, influyen en la expresión génica, la regulación del genoma y la estabilidad del genoma.

En definitiva, la epigenética va a estar presente en los aspectos que desde la Genómica vamos a tratar: la Genómica terapéutica, en su rama farmacogenética y en la rama nutrigenética.



Abreviaciones

A: Adenina

ADN: Ácido desoxirribonucleico

ARN: Ácido ribonucleico

ARNm: Ácido ribonucleico mensajero

ARNr: Ácido ribonucleico ribosómico

ARNt: Ácido ribonucleico de transferencia

ARNsn: Ácido ribonucleico pequeño nuclear

C: Citosina

CNv: Variación en el número de copias

DNMT: ADN metiltransferasa

ENCODE: del inglés, "Encyclopedia of DNA Elements"

FT: factor de transcripción

G: Guanina

HGP: del inglés "the Human Genome Project"

PPAR: del inglés, "peroxylome proliferator activated receptor"

SAME: S-adenosil metionina

SNP: del inglés, "Single Nucleotid Polymorphism"

TBP: Proteína de unión a TATA

Bibliografía

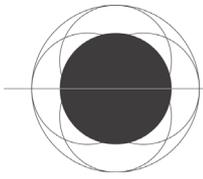
- 1.- Mukherjee S. El gen, una història íntima. 1ª edición. Barcelona. Ediciones La Campana. 2017.
- 2.- Cacabelos R. Del genoma humano a los retos futuros de la Medicina Genómica. Gen-T. P. 2013; 9: 89-114.
- 3.- National Human Genome Research Institute (NHGRI). <https://www.genome.gov/>. Consulta octubre 2021.
- 4.- Baiget M. Curso “Fem camí cap a la Farmacogenètica” del Col.legi de Farmacèutics de Barcelona. Noviembre 2016.
- 5.-Govea Villaseñor MCR. <https://biol1c201.blogspot.com/2018/06/semana-14.html>. Consulta octubre 2021
- 6.- <https://blocs.xtec.cat/arnaubiogeo/files/2013/04/Genoma-eucariota.pdf>. Consulta octubre 2021
- 7.- <https://www.muyinteresante.es/ciencia/articulo/logran-secuenciar-por-primera-vez-un-genoma-humano-completo-541648793581/amp>. Consulta abril 2022
- 8.- https://www.researchgate.net/publication/353345283_The_complete_sequence_of_a_human_genome_preprint. Consulta abril 2022.
- 9.- Bertran Prieto P. <https://medicoplus.com/ciencia/tipos-mutaciones>. Consulta octubre 2021
- 10.- Sanhuenza J, Valenzuela B J.A. Nutrigenómica: revelando los aspectos moleculares de una nutrición personalizada. Rev Chil Nutr. 2012; 39: 71-85.
- 11.- <http://www.facultad.efn.uncor.edu/webs/departamentos/divbioeco/anatocom/Biologia/Celula/Metabolismo/sinproteinas.htm>. Consulta octubre 2021.
- 12.- Freedman S. Biología. 3ª edición. Madrid. Editorial Pearson Educación S.A. 2009.

13.- Feduchi, Blasco, Romero, Yáñez. Bioquímica. Conceptos esenciales. 2010. Editorial Médica Panamericana.

14.- De la Torre M. Introducción a la epigenética, nuevo paradigma en nefrología. NefroPlus, 2017; 9 (1): 1-103.

15.- López Serra, L. Proteína de unión a DNA metilado y cáncer: familia MBD (*Methyl-CpG Binding Domain*). 2007. Tesis Doctoral. Universidad Autónoma de Madrid.

16.- Arratia J, Aguirre J. Los factores de transcripción tipo Myb, una familia de reguladores de la diferenciación celular conservada en los organismos eucariontes. Revista Especializada en Ciencias Químico-Biológicas. 2013; 16 (2): 98-108.



Anexo 1

LAS CIENCIAS ÓMICAS?

Las "ómicas" son las ciencias que estudian, a nivel molecular, los distintos elementos (ej. genes, proteínas, metabolitos) que integran los sistemas biológicos (ej. células, tejidos, organismos), y las interacciones entre ellos y con elementos externos al individuo (ej. el ambiente). Por ejemplo, **la Genómica** estudia los genes, **la Proteómica** las proteínas, **la Epigenómica** las modificaciones reversibles del ADN o proteínas, etc.

¿Cuál es su aplicación en la práctica clínica?

Aunque cada una de ellas, de manera aislada, ofrece mucha información, la combinación del conocimiento generado a partir de las diferentes ciencias ómicas contribuye a mejorar la prevención, el diagnóstico, el tratamiento y el seguimiento personalizado de cada paciente. Además, nos permite aproximarnos a una medicina cada vez más personalizada basada en las características individuales de cada paciente.

ÁREAS DE MAYOR IMPACTO

- Enfermedades cardiovasculares
- Cáncer
- Enfermedades inmunes e infecciosas
- Enfermedades neurológicas
- Enfermedades metabólicas y nutrición
- Enfermedades raras
- Envejecimiento saludable

GENÓMICA
Estudio del conjunto del material genético presente en un organismo

TRANSCRIPTÓMICA
Estudio de los perfiles de expresión de los ARN mensajeros, los micro-ARNs y otros ARN no codificantes

PROTEÓMICA
Estudio del set completo de proteínas expresadas en un organismo en un tiempo determinado y particular de cada tipo celular o tisular

ARN

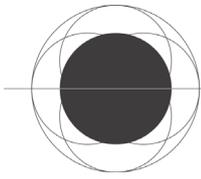
METABOLÓMICA
Identificación y cuantificación de productos metabólicos de pequeño tamaño (metabolitos) de un sistema biológico (célula, tejido, fluido biológico y órgano)

EPIGENÓMICA
Estudio de los elementos que controlan la expresión genica sin modificar la secuencia de nucleótidos de ADN

CITÓMICA
Estudio de la colección de procesos celulares complejos y dinámicos subyacentes a los procesos fisiológicos, así como de la diversidad funcional y estructural de conjunto de células de un organismo

OIRAS
Existen tantas ciencias ómicas como elementos biológicos susceptibles de ser estudiados.

Observatorio para la Opción de las Ciencias Ómicas
Instituto Roche



Anexo 2

ADN	El nombre químico de la molécula que lleva las instrucciones genéticas. Consiste en 2 hebras que se enrollan una alrededor de la otra para formar una doble hélice. Unida a cada azúcar hay 1 de 4 bases: adenina (A), citosina (C), guanina (G) y timina (T). Las 2 hebras se mantienen unidas por enlaces entre las bases con una correspondencia determinada
ARN	Molécula formada por un polirribonucleótido de longitud variable que contiene uracilo en vez de timina. Hay 3 tipos: ARN mensajero (ARNm), ARN ribosomal (ARNr) y ARN transferente (ARNt)
Alelo	Cada una de las versiones de un polimorfismo o gen. Un individuo típicamente hereda 2 alelos para cada polimorfismo, 1 de cada padre. Si los 2 alelos son iguales, el individuo es homocigótico para ese polimorfismo. Si son diferentes, es heterocigoto
Autosoma	Uno de los cromosomas no sexuales
Cromatina	Material formado por ácidos nucleicos y proteínas que se observa en el núcleo de la célula en interfase
Cromosoma	Paquete organizado de ADN que se encuentra en el núcleo de la célula. Cada organismo tiene un número de cromosomas diferente. Los humanos tienen 23 pares de cromosomas
Codón	Una secuencia de 3 bases de ADN o ARN que especifica 1 solo aminoácido en la traducción
Epigenético	Mecanismo de regulación de la expresión (transcripción y traducción) de genes que no depende de cambios en las bases del ADN, sino que opera a un nivel superior
EWA	Siglas en inglés de estudio de epigenoma completo. Se refiere a los estudios que analizan la metilación de epigenoma completo
Exón	Zona codificante de un gen
Fenotipo	Característica o rasgo observable de un individuo, fruto de la interacción entre su genotipo y el ambiente en que este se expresa. Se distinguen los fenotipos finales de enfermedad cardiovascular (infarto, ictus, etc.) y los fenotipos intermedios (hipertensión, dislipemias, etc.)
Gen	Unidad de herencia que ocupa una posición concreta en el genoma (<i>locus</i>) y tiene una estructura determinada
Genoma	Conjunto cromosómico básico que contiene toda la información genética del individuo
Genotipo	Suma de los alelos de un individuo para una posición determinada
GRS	Puntuación de riesgo genético. Puede ser ponderada o no ponderada en función de si valora el efecto de cada polimorfismo o solo los considera de manera aditiva
GWAS	Estudio de asociación de genoma completo
Histona	Proteínas pequeñas de carácter básico, ricas en lisina y arginina, que se unen al ADN en la cromatina
Locus	Lugar que un gen ocupa en el genoma
Metilación	Adición de restos metilo (-CH ₃) al ADN, en forma de bases metiladas
MicroARN	Fragmento muy pequeño de ARN no codificante y con importante función reguladora
Mutación	Cualquier cambio de base introducido en la secuencia de ADN. A veces se utiliza específicamente para indicar que la frecuencia alélica de la variación genética es muy baja (< 1%)
Nucleótido	Molécula constituida por una base nitrogenada, una pentosa y un grupo de ácido fosfórico. Unidad básica de la que se compone un ácido nucleico
Telómero	Extremo final de los cromosomas. En humanos, el ADN de los telómeros está compuesto por repeticiones en tándem de la secuencia TTAGGG

Anexo 3

Nomenclatura

Genes

La nomenclatura de los genes inicialmente parecía simple para organismos sencillos, como bacterias, levaduras. Se designaban, en principio, siguiendo dos criterios: no poner el nombre de su descubridor y nombrarlos con una abreviación (de tres letras) descriptiva de la función del gen. Normalmente con letras en MAYÚSCULA y *cursiva*. El problema vino en organismos superiores. Tres letras no parecían suficientes, pero se mantenía la tendencia de que el nombre fuera descriptivo. Respecto al número del gen, también hay varios criterios. A veces los genes tienen varias copias en el genoma, y a medida que se van identificando se numeran de forma correlativa. Otras veces tienen una función similar y se suelen bautizar correlativamente. (En ocasiones el número hace referencia al peso molecular de la proteína que codifica). Y si el número no quiere decir nada, se pone 1.

Un recurso para consultar información sobre genes es <https://www.genecards.org/>. Se encuentran recogida la información de genes (20692) que codifican proteínas.

También se recoge información sobre genes en <https://www.ncbi.nlm.nih.gov/gene>

<https://www.genenames.org/about/guidelines/> recoge las directrices actuales para nombrar los genes.

Nucleótidos

En la numeración de nucleótidos de la secuencia de ADN de referencia, la A del codón de iniciación (ATG) se considera el nucleótido +1. No hay ningún nucleótido que se enumere con el cero. Los nucleótidos que se encuentran en posición corriente arriba del +1 deben ir precedidos por el signo menos; por ejemplo, -1, -2, -3, etc. Los nucleótidos situados corriente abajo del +1 no van precedidos de ningún signo.

Mutaciones

Hay cierta confusión en la terminología. El término mutación se utiliza en ocasiones para indicar un cambio y en otras para indicar un cambio que causa cierta enfermedad. De forma similar, el término polimorfismo se utiliza tanto para describir un cambio que no causa enfermedad como para señalar un cambio que se encuentra con una frecuencia $\geq 1\%$ de la población. Para evitar esa confusión, se pueden emplear términos tales como variaciones de secuencia, alteración o variante alélica.

Según el tipo de nucleótido sustituido, las mutaciones puntuales se denominan: *a)* transición, que consiste en la sustitución de una pirimidina por otra o de una purina por otra, de tal forma que un par G-C es reemplazado por un par A-T o viceversa, y *b)* transversión, cuando una purina es reemplazada por una pirimidina o viceversa, en cuyo caso un par A-T es reemplazado por un par T-A o C-G.

Para nombrar las mutaciones se recomienda utilizar una serie de reglas elaboradas por el Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC), que se detallan en su página web (www.hgvs.org/mutnomen/recs) y resumimos a continuación:

- . La regla más importante es que las variantes deben ser descritas en el nivel más básico posible, es decir a nivel de ADN. La descripción siempre debe referirse a una secuencia de referencia, ya sea la secuencia genómica o la codificante de ADN.

- . Para genes/proteínas, sólo se utilizarán los símbolos recomendados por el HGNC (disponibles en: www.genenames.org/guidelines.html). Cuando se hace referencia a nivel de ARN o proteína en un determinado texto, la primera vez que se menciona debe utilizarse el siguiente formato: en primer lugar, ADN y entre paréntesis proteína o ARN. Ejemplo: c.48G>C (p.Trp16Cys).

- . Cuando se describen varios cambios, es conveniente proporcionar una tabla con la lista de los hallazgos, utilizando columnas separadas para ADN, ARN y proteínas, e indicar si los cambios fueron deducidos de forma experimental o teórica.

. Para hacer referencia a un nucleótido, deberá precederle la letra g en minúscula si se trata del ADN genómico, la c minúscula si se refiere a un ADN codificante (ADN-c), la m minúscula si se refiere a un ADN mitocondrial, la r minúscula si se refiere a una secuencia de ARN ribosomal y la p minúscula si se refiere a una proteína. Para evitar posibles confusiones entre el número del nucleótido y la g, la c, etc., se intercalará un punto. Por ejemplo g.576, significa que se trata del nucleótido que está situado en la posición 576 del ADN genómico.

Sufijo	Tipo de secuencia
g.	secuencia genómica
m.	secuencia mitocondrial
c.	secuencia de ADN codificante para proteína
n.	secuencia de ADN no codificante
r.	secuencia de ARN
p.	secuencia de la proteína

Para la descripción de cambios específicos:

- . El símbolo > indica una sustitución a nivel de ADN; por ejemplo, c.75A>T.
- . El símbolo _ indica un cambio que afecta a varios residuos, que separa el primero y el último de los residuos afectados; por ejemplo, c.77_79delACT.
- . Con del se indica una delección; por ejemplo, c.77delA.
- . Con dup se indica una duplicación; por ejemplo, c.dupA.
- . Con ins se indica una inserción; por ejemplo, c.75_76insG.

Los cambios de nucleótido deben empezar por el número del nucleótido seguido del cambio que tiene lugar en esa posición; por ejemplo, 576G>C significa que el nucleótido de la posición 576 que en la secuencia de referencia es una G ha sido reemplazado por una C.

Al describir dos variantes de secuencia en un mismo individuo:

- . Dos cambios en la secuencia en diferentes alelos se representan entre corchetes, separados por el signo +; por ejemplo, c.[76A>C]+[87delG].
- . Dos variantes de secuencia en el mismo alelo se ponen entre corchetes separados por punto y coma; por ejemplo, c.[76A>C; 83G>C].
- . Dos variantes de secuencia en alelos desconocidos se ponen entre corchetes separadas por (+); por ejemplo, c.[76A>C(+)]83G>C].

Para designar la localización de una variación en un intrón, se puede utilizar la numeración de los nucleótidos correspondientes al ADN. Por ejemplo, c.1232+1G>C significa la sustitución de G por C del primer nucleótido del intrón que sigue a la posición 1232 del ADN.

Debido a la gran cantidad de posibles variantes que pueden existir de un único SNP, su nomenclatura es bastante confusa. En la mayoría de los casos cuando se hace referencia a un SNP es sobre la base de su número rs (RefSNP) de la base de datos dbSNP <https://www.ncbi.nlm.nih.gov/snp/>. Así, el número rs es un número de acceso utilizado por investigadores y clínicos para referirse a SNP específicos. Este número de variación es inequívoca.

Alelos

Las diferentes versiones de un gen (alelo) se designan escribiendo, a continuación del nombre del gen, asterisco (*) y un número. Normalmente la variación salvaje original se le asigna el número 1, y las demás variantes, números correlativos. Existen genes muy polimórficos, de las cuales se han descritos muchas variantes, como el gen CY2D6.

Una base de datos donde consultar variantes del CYP450 es <https://www.pharmvar.org/gene/> donde se encuentran recogidas todas sus variantes.

La base de datos <https://www.omim.org/> recoge la información de 16.000 genes relacionados con el fenotipo.

Anexo 4

Animaciones y ejercicios

Animaciones

1.- Animación del gen³: <https://www.youtube.com/watch?v=LVk0xmtlbos&t=29s>

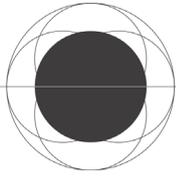
2.- Animación del código genético³. <https://www.genome.gov/es/genetics-glossary/Codigo-genetico>

3.- Mapa interactivo de un gen:
<https://learn.genetics.utah.edu/content/basics/geneanatomy/>

4.- Cosas que sí sabemos del ADN:
<https://learn.genetics.utah.edu/content/basics/dnathings>

5.- Animación interactiva de las diferencias entre ADN y ARN:
<https://learn.genetics.utah.edu/content/basics/rna/>

6.- Animación del dogma central de la biología molecular:
<https://learn.genetics.utah.edu/content/basics/transcribe/>



Ejercicios

1.- En base a la información que encontrarás en la página web

<https://medlineplus.gov/genetics/chromosome/>

. Elabora una relación del número de genes de cada cromosoma.

. ¿Hay una relación directa entre el tamaño del cromosoma y el número de genes que contiene?

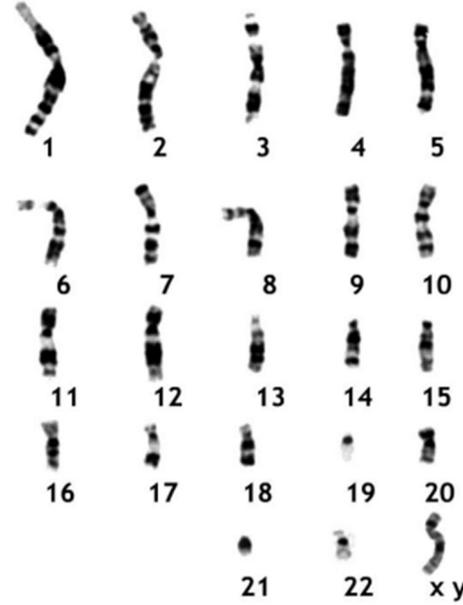
2.- Confecciona un cariotipo.

Cut n^o Paste Karyotyping Activity

Cut out chromosomes here

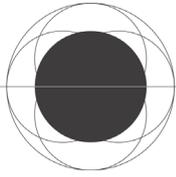


Paste chromosomes here with their match

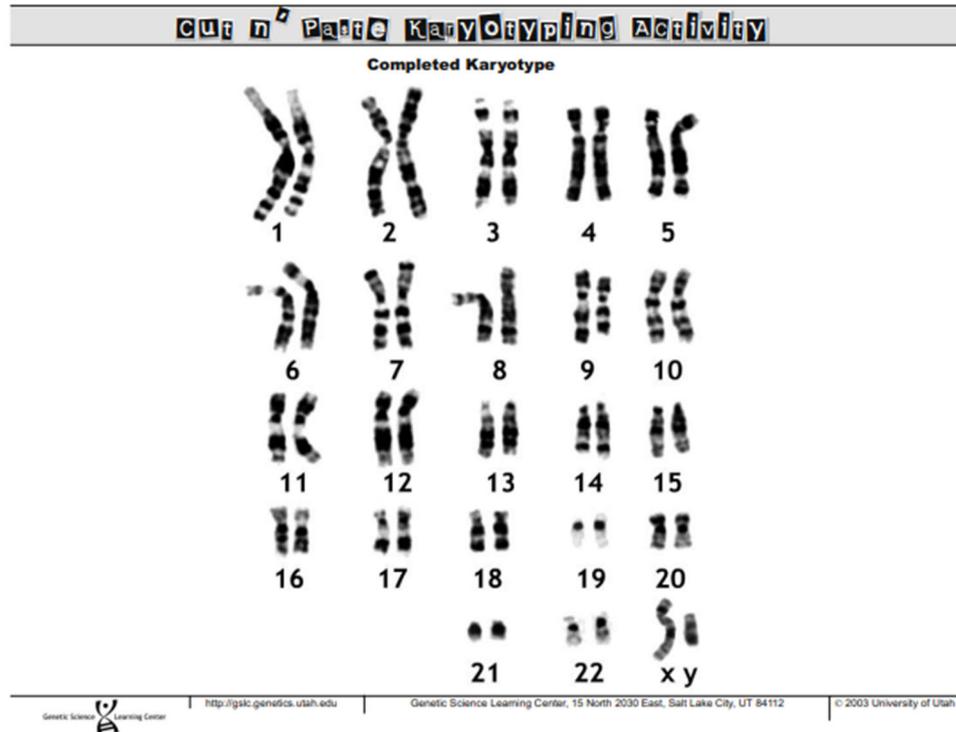


1 2 3 4 5
6 7 8 9 10
11 12 13 14 15
16 17 18 19 20
21 22 x y

Genetic Science Learning Center
<http://gslc.genetics.utah.edu>
Genetic Science Learning Center, 15 North 2030 East, Salt Lake City, UT 84112
© 2003 University of Utah



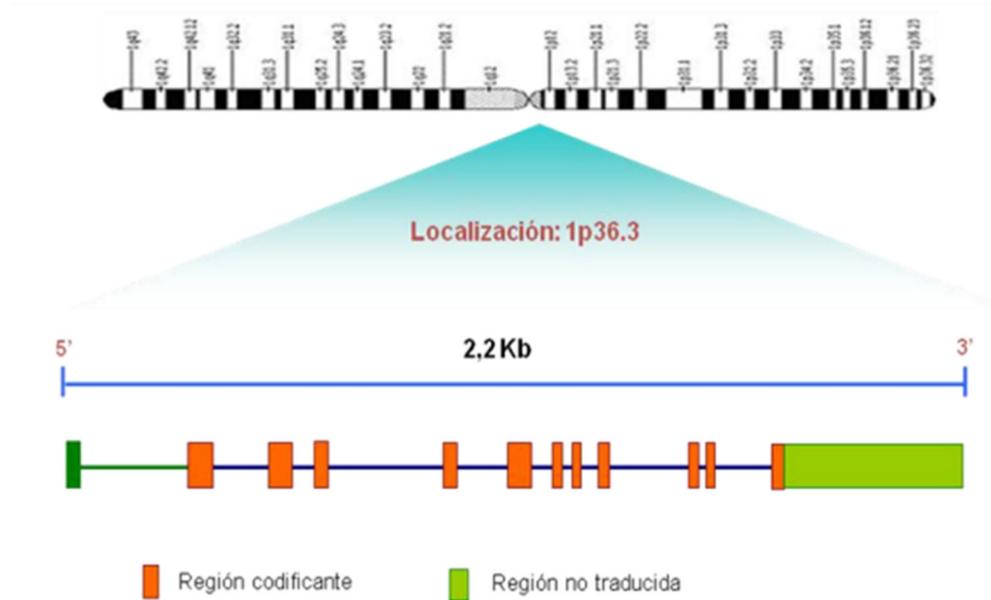
Solución:



. ¿Has encontrado algún tipo de anomalía cromosómica?

. ¿Dirías que la muestra se corresponde con un cariotipo normal?

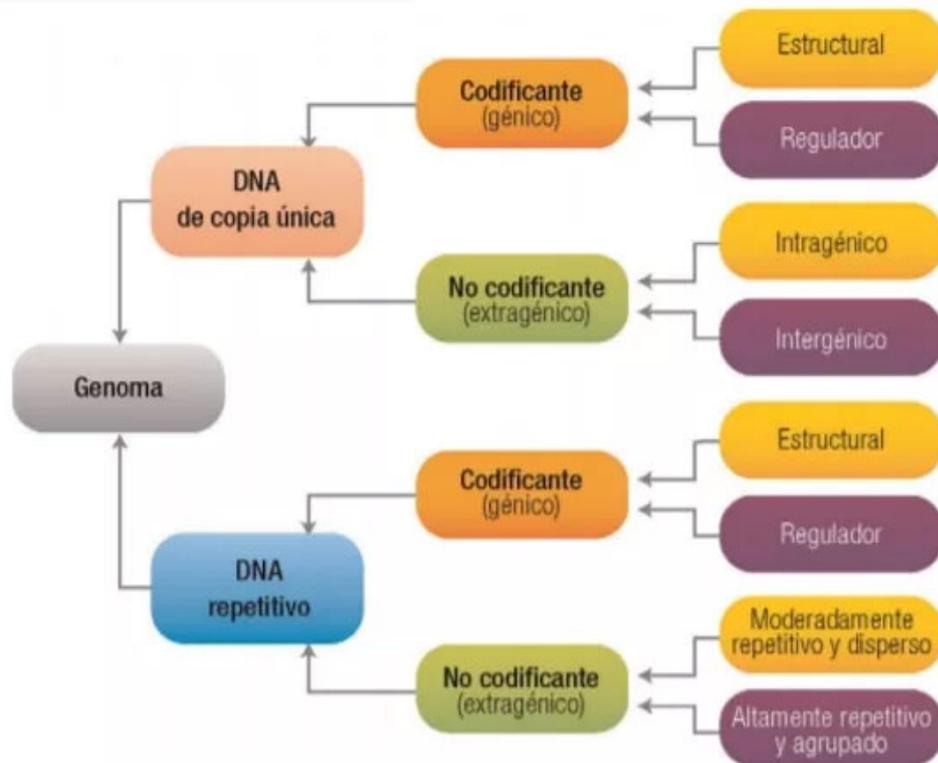
3.- Mapa genético del gen *MTHFR*. El gen *MTHFR* está localizado en la región 36.3 del brazo corto del cromosoma 1.



Fuente: Tesis Doctoral R.P. Arturo. 2009

- . ¿Puedes contabilizar el número de exones que tiene?
- . ¿Y el número de intrones?
- . ¿Cómo se denomina la región no traducida?

4.- En los distintos tipos de ADN que se indican en el siguiente esquema, ¿dónde colocarías los genes de histonas, los promotores, los telómeros, los genes de las globulinas, los intrones, los centrómeros y los exones?



Fuente: Instituto superior de Neurociencias. Consejo Mexicano de Neurociencias.

5.- Indica el tipo de mutación a la que corresponden estas nomenclaturas:

