ENCICLOPEDIA DE ELEMENTOS DEL ADN (ENCODE): UN GRAN PASO HACIA LA COMPRENSIÓN DEL FUNCIONAMIENTO DEL GENOMA HUMANO

ESTUDIANTES: ALEJANDRA MARGARITA MORÓN GUERRERO
NATALIA PÁEZ ESCALLÓN

TUTOR: REGGIE GARCÍA ROBLES
UNIVERSIDAD EL BOSQUE

FACULTAD DE MEDICINA
INSTITUTO DE INVESTIGACIÓN EN NUTRICIÓN, GENÉTICA Y METABOLISMO
BOGOTÁ, DISTRITO CAPITAL

UNIVERSIDAD EL BOSQUE
FACULTAD DE MEDICINA
ENCICLOPEDIA DE ELEMENTOS DEL ADN (ENCODE)
INSTITUTO DE INVESTIGACIÓN EN NUTRICIÓN, GENÉTICA Y METABOLISMO
PREGRADO
ALEJANDRA MARGARITA MORÓN GUERRERO
NATALIA PÁEZ ESCALLÓN
REGGIE GARCÍA ROBLES, TUTOR

“La Universidad El Bosque, no se hace responsable de los conceptos emitidos por los investigadores en su trabajo, solo velará por el rigor cientifico, metodológico y ético del mismo en aras de la búsqueda de la verdad y la justicia”.

Queremos agradecer al Doctor Reggie García Robles por su colaboración, entrega y compromiso con la investigación.

Contenido

Glosario 9

Introducción 11

Desarrollo de la iniciativa ENCODE 16

Resumen de los hallazgos encontrados en el estudio ENCODE 22

Herramienta y análisis de bioinformática en el proyecto ENCODE 27

Discusión 29

Conclusión: importancia y perspectivas 32

Bibliografía 34

Figura 1. Organización del material genético nuclear 15

Figura 2. Flujograma Etapas del estudio ENCODE 21

Tabla 1. Descripción de las técnicas 25

RESUMEN: El proyecto de Enciclopedia de Elementos del ADN (ENCODE) es un consorcio internacional de investigación que surge a partir del Proyecto de Genoma Humano con el fin de interpretar los principales hallazgos que dejó este, especialmente sobre el denominado "ADN basura". El objetivo de ENCODE era identificar los elementos funcionales del genoma humano y comprender las interacciones y funcionamiento del genoma como un todo. El proyecto de ENCODE se dividió en dos fases. La primera fase se centró en el análisis del 1% del genoma caracterizado como funcional y la segunda fase se centró en analizar el 99% restante del genoma lo cual se realizó por medio de múltiples técnicas. A raíz de los descubrimientos de ENCODE se ha logrado una mejor comprensión de las enfermedades especialmente aquellas que tiene un componente genético generando un gran impacto sobre la salud y la enfermedad. Finalmente, ENCODE ha permitido tener una mayor compresión sobre la genética, dando respuestas y planteando cuestionamientos de importancia en medicina.

Palabras clave:

Genoma Humano**,** ADN intergenico, epigenomica, regulación de la expresión génica , Enciclopedia de Elementos del ADN (ENCODE).

ABSTRACT: The Encyclopedia of DNA elements Project is an international research consortium that emerges from the Human Genome Project with the

purpose to analyze the principal foundings that were left by this, especially on the so-called "junk DNA". ENCODE´s main objective was to identify all the functional elements of the human genome and understand the interactions and functioning of the genome as a whole. The ENCODE project was divided into two phases. The first phase focused on the analysis of the 1% of the genome characterized as functional and the second phase focused on analyzing the remaining 99% of the genome which was done by employing multiple techniques. From the discoveries of ENCODE, we manage to have a better understanding of the diseases especially those that have a genetic component which has a great impact on health and disease. Finally, ENCODE has allowed a mayor comprehension about genomics giving answers and raising questions about medicine.

Keywords

Human Genome, DNA, Intergenic, Epigenomics, Gene expression regulation, Encyclopedia of DNA elements (ENCODE).

Glosario

Ácido desoxirribonucleico (ADN): Material genético de la célula.

Acido ribonucleótido (ARN): Polímero de ribunocluteotidos.

ADN polimerasa: Enzima que cataliza la síntesis de ADN.

ARN polimerasa: Enzima que cataliza la síntesis del ARN.

Cromatina: Complejo fibroso del ADN eucariota y proteínas histonas.

Cromosoma: Portador de genes, consistentes en moléculas largas y ADN y proteínas asociadas.

ENCODE: Proyecto de Enciclopedia de Elementos del ADN.

Epigenética: Cambios heredable en el ADN que no implican alteraciones en las secuencias de nucleótidos y modifican la estructura y condensación de la cromatina por lo que afectan la expresión génica y el fenotipo.

Gen: Segmento de ADN que codifica una cadena poli peptídica o una molécula de ARN.

Genoma: Conjunto de instrucciones genéticas que se encuentran en una célula.

Genómica: El análisis sistemático de genomas celulares completos.

Histona: Miembro de una familia de proteínas que empaquetan en ADN a los cromosomas de las células eucariotas.

Locus: Lugar específico del cromosoma donde está localizado un gen u otra secuencia de ADN.

Nucleosoma: Unidad básica estructural de cromatina que consiste en ADN enrollado alrededor de un núcleo de histona.

Potenciador: Corta región de ADN que puede.

Pseudogenes: Copia no funcional de un gen.

Promotores: Secuencia de ADN donde se une la ARN polimerasa para iniciar la transcripción.

TUFs: Transcritos de función desconocida.

TSSs: Sitios de unión a la transcripción .

Introducción

El proyecto Genoma Humano cuyo objetivo era identificar la secuencia de todo el genoma humano, inició en 1990 y se completo en el 2003(1).Si bien el proyecto marcó un avance importante en el entendimiento del genoma y su complejidad, los hallazgos dejaron interrogantes, especialmente sobre el denominado “ADN basura” y cómo funciona la

información genética (2). A raíz de esto, nace el proyecto de Enciclopedia de Elementos del ADN (ENCODE) con el fin de interpretar toda la información que dejó el proyecto de Genoma Humano y las preguntas planteadas a partir de este(3). Por tal razón, en el año 2003 el Instituto Nacional de Investigación del Genoma Humano (NHGRI) puso en marcha y financió esta investigación con la participación anual del congreso de los Estados Unidos a través de un fondo federal que dispone el Presidente para los diferentes programas(4). Desde su inicio hasta la actualidad, el presupuesto de este proyecto ha variado cada año; se estima que para el año 2018 tenga un costo alrededor de los 400,000 millones de dólares, sin embargo, este monto ha disminuido en comparación con años anteriores(5). El principal objetivo de ENCODE ha sido construir una lista completa de las partes funcionales del genoma humano, incluyendo elementos que actúan a nivel de las proteínas y el ARN, y elementos regulatorios que controlan las células y las circunstancias en las que un gen está activo o inactivo. Esta funcionalidad del genoma esta relacionada con su organización dentro de nucleo y su interaccion con otras moleculas que podrian regular la expresion genica. Ver Figura 1.

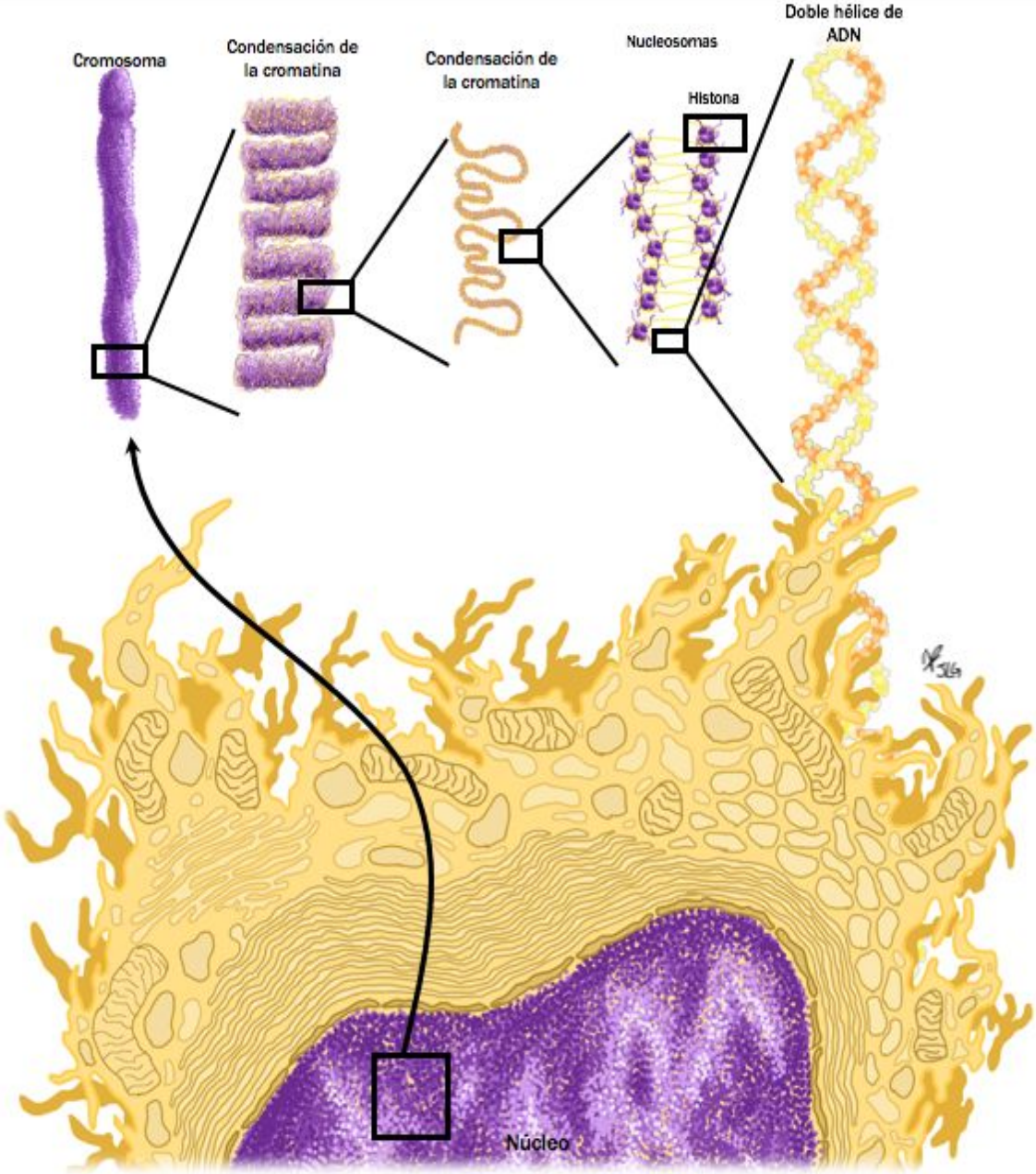
Finalmente, lo que pretende ENCODE es localizar en el genoma elementos funcionales del mismo y comprender las interacciones y funcionamiento del genoma como un todo (6).

En consecuencia, el proyecto ENCODE se dividió en dos fases. La primera fase piloto se centró en el análisis del ADN caracterizado inicialmente como funcional o regiones codificadoras de proteínas, las cuales corresponden a aproximadamente un 1,5 % del genoma. Finalizada esta primera fase, en el 2007 se dio inicio a la segunda etapa del proyecto y los resultados fueron publicados en el 2012. El objetivo de esta, fue analizar el 98,5 % restante del genoma humano (1). Así mismo, esta fase marcó una importante contribución en el desarrollo de la bioinformática puesto que se usaron varias técnicas para el análisis de la gran cantidad de datos generados del estudio de múltiples líneas celulares. Los resultados entre 2003 y 2007 contribuyeron para que ENCODE reportara que el 80,4 % del genoma humano tenía alguna función bioquímica conocida y que el denominado "ADN basura" en realidad guardaba una gran importancia en la regulación de los elementos codificantes. De esta forma, los nuevos avances en genética a partir de la iniciativa de ENCODE están generando una revolución de conocimiento con impacto en el campo de las ciencias y la salud (1).

Con esta revisión pretendemos hacer divulgación del conocimiento de manera concisa, precisa y sencilla del desarrollo y resultados de una iniciativa importante que pretende dar luz sobre el funcionamiento del genoma humano, surgido a partir de los resultados de otro proyecto histórico como lo fue el Proyecto de Genoma Humano. Para cumplir con este

objetivo se revisaron los articulos publicados por el consorcio en www.encodeproject.org durante el periodo 2003-2017 (972 publicaciones en total) y se escogieron aquellos con contenido general que permitieran describir de manera más global la iniciativa ENCODE, su impacto y resultados, así como el material disponible en las páginas de divulgación de ENCODE: NIH, ENCODE y UCSC.

Figura 1. Organización del material genético nuclear. El ADN interactúa con proteínas que permiten su oganización y condensación al interior del núcleo celular, es asi como por la interacción del ADN con las histonas se forman las unidades básicas del empaquetamiento denominadas nucleosomas. En las diversas células, el ADN se encuentra generalmente en diferentes grados de condensación que permiten la regulacion del acceso de la maquinaria transcripcional, y por ende la expresión génica. El estado de mayor compactación de la cromatina esta dado por la formación de los cromosomas.[1]



[1] Elaboración porpia desarrollada en sketchbook Pro por Joffre Alejandro Guzman Laguna

Desarrollo de la iniciativa ENCODE

El objetivo principal de ENCODE es anotar todos los genes que codifican para proteínas, pseudogenes, loci transcritos no codificantes en el genoma humano y caracterizar los productos de transcripción con el fin de proporcionar a la comunidad científica información de alta calidad y completas de elementos funcionales en el genoma humano (7). Los elementos funcionales son aquellos que se encuentran en las regiones no codificantes e incluyen los promotores, enhancers, silenciadores y genes de ARN no codificantes (8). Los elementos funcionales se definen como un segmento del genoma que codifica un producto definido (por ejemplo, una proteína o ARN no codificante) o con evidencia de actividad bioquímica que es reproducible(6). La búsqueda de elementos funcionales se ha centrado en la determinación experimental de la estructura de la cromatina y sitios de unión a la proteína y la detección computarizada de secuencias no codificantes de ADN conservadas (9).

La primera fase de ENCODE se centró en el análisis de 30 megabases (Mb) de ADN caracterizadas como funcionales o regiones codificantes de proteínas que corresponden a 44 regiones genómicas o aproximadamente al un por ciento del genoma. De estas 30 Mb, 50% residen en 14 regiones de las cuales existe un conocimiento biológico, y el otro 50% reside en las otras 30 regiones restantes. Estás 44 regiones se sometieron a análisis de trascripción con el fin de comprender mejor las moléculas de ARN codificantes, ya que los factores de transcripción están involucrados en muchas funciones celulares activando

moléculas de forma indirecta o directa (1). Las pruebas de chip-seq (*Chromatin inmunoprecipitation followed by sequencing)* fueron realizadas durante la fase piloto del proyecto para el análisis de 18 factores de trascripción y de los componentes generales de la maquinaria de transcripción. Durante esta fase, fue posible obtener secuencias de regiones genómicas que son ortólogas a las dianas de otras especies de vertebrados. Los datos generados de este estudio de comparación permitieron la identificación precisa de elementos evolutivamente conservados, haciendo posible la inferencia de sus funciones biológicas (1).

La segunda fase del proyecto ENCODE comenzó en el 2007 y los resultados fueron publicados en el 2012. El objetivo de esta fase, era analizar el 99% del genoma humano restante. Se utilizaron varias técnicas en varios tipos de linajes celulares. Durante los años 2003-2012 se produjeron 1,640 grupos de datos y se realizaron 24 tipos de experimentos en 147 linajes celulares(1). Estos estudios consistieron en la cuantificación de diferentes especies de ARN en células enteras y compartimientos celulares, mapear regiones codificantes de proteínas, modificaciones de histonas y sitios de unión de factor de transcripción por la técnica chip-seq, así como mapeo de sitios de metilación de ADN. Mediantes estos estudios fue posible demostrar que el 80,4% del genoma es funcional en al menos un tipo celular (1).

Se identificaron 20,687 genes codificantes para proteínas con un promedio de 6,3 transcritos por locus. Las secuencias codificantes encontrados en estos genes corresponden al 2,94% del genoma (1). Esto significa, que gracias al proyecto ENCODE se estimó que el 2,94 % de todo el genoma humano era codificante para proteínas, y que el 80,4% de secuencias gobiernan como los genes son regulados. Adicionalmente, el proyecto ENCODE también mapeó pseudogenes, los cuales se encuentran dispersos en el genoma como duplicados de genes que al parecer son transcripcionalmente inactivos (6).

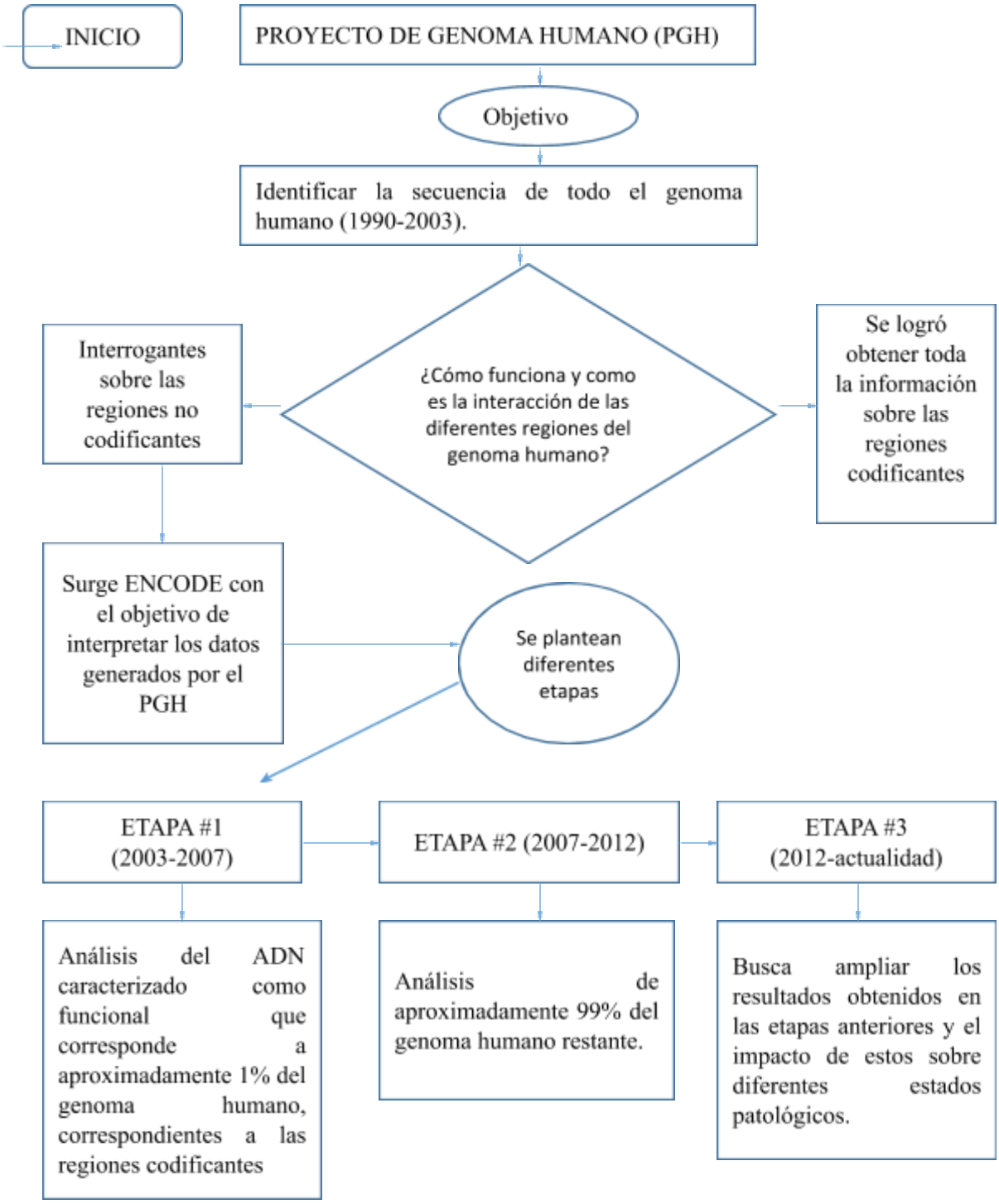
Así mismo, con el fin de identificar las regiones reguladoras del genoma, se realizó un mapeo de los sitios de unión de ADN para 119 proteínas, incluyendo los factores de transcripción y los componentes de ARN polimerasa, lo cual se llevó a cabo en 72 tipos celulares por técnica de chip-seq (7). Por otra parte, el proyecto ENCODE evaluó las localizaciones cromosómicas de las modificaciones de histonas en 46 tipos celulares, lo cual mostró un patrón variable de modificaciones a lo largo de los tipos celulares (7). Además de esto, el proyecto ENCODE mapeó 2,89 millones de sitios hipersensibles a la DNAasa I en 125 tipos de linajes celulares. Se encontró que aproximadamente el 75% de estos sitios DHS (sitios hipersensibles a DNAsa I) se encontraron en intrones o regiones intergénicas, lo cual sugirió que secuencias no codificantes ejercen funcionalidad en la regulación de la expresión génica (7).

Se ha logrado perfilar el estado de metilación del ADN de más de un millón de islas CpG en diversas líneas celulares y se ha proporcionado la determinación cuantitativa de la proporción de metilación CpG en cada sitio (10) ENCODE ha identificado los patrones específicos de modificaciones postraduccionales de histonas mediante la técnica de chip-seq, y las modificaciones de histonas se han relacionado con las regiones reguladoras como los promotores, potenciadores, dominios de transcripción y regiones silenciadas. Llevando a tener los mapas de las modificaciones de histonas en el genoma de diferentes líneas celulares (10).

La última fase del proyecto ENCODE inició en 2012 con el fin de dar mayor detalle de los resultados obtenidos en las dos fases anteriores, ampliar el conocimiento y el impacto de estos hallazgos en estados patológicos (1). Sin embargo, el planteamiento de esta fase dio pie a una amplia discusión en la comunidad académico – científica. El hallazgo de regiones bioquímicamente activas que cubren gran parte del genoma y que podría ser mayor comparado con las regiones conservadas evolutivamente, plantea la pregunta si regiones no conservadas pero bioquímicamente activas son verdaderamente funcionales. Sin embargo, algunos investigadores de ENCODE argumentan que el significado que se le asignó al término "funcional" es erróneo en la medida que no existe una definición universal de este y que ENCODE se apresuró en identificar todos los elementos funcionales del genoma humano. Así mismo, argumentan que ENCODE no tuvo en cuenta que la función bioquímica no se puede mantener indefinidamente sin selección natural y que podría no implicar una función biológica. Adicionalmente, no se puede considerar al genoma como

invulnerable a variantes patogénicas ni olvidar que la evolución no implica un propósito como si lo podría tener una función (1). Ver Figura 2.

Figura 2. Antecedentes y desarrollo histórico del proyecto ENCODE



Resumen de los hallazgos encontrados en el estudio ENCODE

Desde el 2007, ENCODE ha usado varias técnicas de laboratorio con el fin de identificar los elementos funcionales a través del genoma. Los elementos que fueron mapeados incluyen regiones de transcripción de ARN, regiones codificantes para proteínas, sitios de unión a factores de transcripción, regiones reguladoras de la estructura de la cromatina y sitios de metilación del ADN. Con el fin de integrar y comparar los resultados obtenidos de los diferentes laboratorios agruparon dos series de líneas celulares denominadas Nivel 1 y Nivel 2 (7). En primer lugar, el nivel uno fue un conjunto de líneas celulares ampliamente estudiadas el cual comprendía las células de eritroleucemia K562, una línea celular B-linfoblastoide y la línea de células madre embrionarias H1. Por otra parte, el nivel dos fue un conjunto de líneas celulares el cual incluía células de carcinoma cervical HeLa-S3, células de hepatoblastoma HepG2 y células endoteliales de vena umbilical humana (7).

En primer lugar, para producir un catálogo de las regiones codificantes para proteínas, ARN no-codificante y pseudogenes del genoma humano se utilizó una anotación manual y automatizada a traves del proyecto GENCODE (11). GENCODE es un proyecto liderado por el Instituto Sanger para la anotación integrada de las características de los genes. Se incluyen 20,687 genes codificantes para proteínas, con un promedio de 6,3 transcritos por locus por empalme alternativo (7). En total GENCODE reportó que el 2,94% del genoma corresponde a secuencias codificantes para proteínas. Además de esto, se anotaron 8,801 ARN pequeños y 9,640 ARN largos no codificantes. Así mismo, el proyecto de

GENCODE anotó que existe 11,224 pseudogenes , de los cuales 863 fueron trascritos y asociados con cromatina activa (7).

Por otra parte, se secuenció el ARN de diferentes líneas celulares y múltiples fracciones subcelulares para desarrollar un catálogo extenso de ARN. Para esto, se empleó la técnica cage-seq (*5′ cap-targeted RNA isolation and sequencing*) con el fin de identificar 62,403 sitios de inicio de la transcripción (TSSs). Se encontró, que el 44 % se encontraban en regiones 5` de genes y las restantes se encontraban en los exones o región 3`(7).

Por otra parte, la accesibilidad a la cromatina caracterizada por la hipersensibilidad a la DNAasaI (DHHs) es distintivo de regiones reguladoras de ADN. Se mapearon 2,89 millones DHHs por medio de la DNAasa-seq en 125 tipos celulares, de los cuales la gran mayoría se encuentran distales a los TSSs (7). Además, se mapearon 4,8 millones de regiones genomicas con reduccion en el numero de nucleosomas a través de 25 tipos celulares por medio de la técnica denominada FAIRE (*Formaldehyde assisted isolation of regulatory elements*), permitiendo identificar regiones con cromatina accesible como elementos regulatorios (7).

Adicionalmente, se evaluó el perfil de metilación del ADN por medio de la secuenciación por bisulfito. Con la utilización de la técnica se encontraron aproximadamente 1,2 millones de islas CpG en cada una de las 82 líneas celulares utilizadas y se pudo correlacionar que la metilación de la citosina en los dinucleótidos CG, está involucrada en la regulación epigenética de la expresión génica. Además, las CpG con metilación más variable se encuentran frecuentemente en regiones intergénicas en lugar que las regiones promotoras (7).

Tabla 1. *Descripción de las técnicas*

TÉCNICA	DESCRIPCIÓN
DNAseq *(DNA sequencing)*	Utilizada para hallar la secuencia de genes individuales, o genomas completos (6).
Chip-seq *(Chromatin inmunoprecipitation followed by sequencing)*	Utilizada para el análisis de cómo las proteínas interaccionan con el ADN. Se basa en la inmunoprecipitación de las proteínas unidas a la cromatina y secuenciación de las secuencias de ADN asociadas (6).
FAIRE-seq (*Formaldehyde assisted isolation of regulatory elements*)	Permite la identificación de secuencias de regiones de ADN en el genoma que se encuentran relacionadas actividad reguladora. Esta técnica permite el estudio y análisis de cualquier tipo de célula (6).
RNAseq *(RNA sequencing)*	Toma el ARNm de una muestra y lo secuencia, obteniendo niveles de transcripción para cada secuencia corta de ARNm. Permite obtener información de la secuencia durante la transcripción (6).
DNAasa-seq *(DNAasa I hypersensitive sites sequencing)*	La enzima DNAasa I corta la cromatina unida a proteínas no histonas y estos puntos

	de corte se secuencian para determinar aquellos sitios hipersensibles a la DNAasaI los cuales corresponden a los cromatina abierta (6).
CAGE-seq (5´cap-targeted RNA isolation and sequencing)	Cuantifica la expresión de ARN y mapea TSS en promotores. Esta técnica, consiste en la captura y secuenciación de los sitios de metilación en el extremo 5`del RNA los cuales suelen formarse en los sitios de inicio de la transcripción (6).
RRBS *(Reduced representation bisulphite sequencing)*	El bisulfito convierte las citosinas no metiladas en uracilo. Se utiliza una enzima de restricción que hace cortes a través de los dinucleótidos CpG que reduce el genoma en porciones específicas enriquecidas en CpG; el objetivo es determinar cualitativamente el estado de metilación de las citosinas (6).

Herramienta y análisis de bioinformática en el proyecto ENCODE

A raíz del crecimiento exponencial en la cantidad de datos y la necesidad de análisis e interpretación de estos mismos toma fuerza el uso y desarrollo de herramientas de bioinformática en el marco del proyecto ENCODE. Este es un campo de investigación

interdisciplinario enfocado en la recolección, organización, clasificación y análisis de datos a partir de modelos de cómputo especializados con el objetivo de encontrar una aplicación determinada de estos mismos (12).

En este sentido el desarrollo del proyecto ENCODE se ha convertido en un impulsor importante de la bioinformática. El proyecto de ENCODE continua creando un catálogo comprensible de los elementos del genoma humano. Los datos generados por los miembros del consorcio de ENCODE son presentados al Centro de Coordinación de Datos de ENCODE (DCC) cuya tarea principal es evaluar, procesar y validar los datos para la publicación de estos para el uso por parte de la comunidad científica (13). Con el objetivo de asegurar calidad y consistencia, ENCODE ha definido ciertos estándares para la recolección y procesamiento de cada tipo de datos los cuales se encuentran habilitados públicamente en el portal de ENCODE. Luego de ser revisados por el Centro de Coordinación de Datos, todos los datos procesados en ENCODE son liberados públicamente al "*UCSG Genome Browser database*" (14). El catálogo de bases de datos de ENCODE se encuentra disponible en la página web (http://encodeproject.org), creado por el DCC.

El nuevo portal de datos de ENCODE, publicado en Agosto del 2014, es la fuente principal de hallazgos ENCODE, análisis de datos, métodos, normas y metadatos experimentales. El usuario tiene a su disposición una página principal de estilo tablero, el cual incluye una actualización de noticias, menús y guías de inicio rápido que proporcionan instrucciones sobre las distintas formas de acceder a los datos mediante, búsqueda, visualización y

descarga (14). Dentro del menú existen diferentes categorías como datos, métodos, e información acerca de ENCODE. La sección de datos proporciona acceso de búsqueda inicial a los datos, así como un enlace de políticas de uso de estos. En la sección de métodos, incluye información sobre los estándares de datos, programas, experimentos y análisis definidos por el consorcio. La sección acerca de ENCODE da un contexto de la variedad y calidad de los experimentos que se encuentran en la base de datos, proporcionando una visión general del proyecto. Además, esta sección tiene la página de publicaciones la cual proporciona una entrada guiada a la navegación de las diferentes publicaciones. Por último, el menú de ayuda tiene una variedad de documentos de ayuda y tutoriales que facilitan la experiencia a los usuarios (14).

Discusión

El proyecto de ENCODE surge con el fin de contribuir en la identificación de los transcriptos ARN, los sitios de unión de regulación de la trascripción y los estados de cromatina de muchos tipos celulares. A raíz de esto, luego de múltiples estudios sobre la regulación genética y el metabolismo del ARN se han podido identificar elementos funcionales no codificantes, incluyendo promotores, enhancers, silenciadores y genes ARN

no codificantes. Estos elementos funcionales no codificantes están asociados con estructuras de cromatina distintas que muestran patrones distintivos de modificaciones de histonas, metilación de ADN, accesibilidad a DNAasa y ocupación de factores de transcripción. Gracias a ENCODE, se pudo encontrar que el tal llamado "ADN basura" era de gran importancia y que aquella gran fracción del genoma era bioquímicamente activo y funcional (15). El proyecto ENCODE marcó un paso importante en el enriquecimiento del entendimiento de la secuencia del genoma humano, y los diferentes hallazgos han demostrado la alta complejidad de la actividad molecular de este. A raíz de esto, ENCODE ha modificado conceptos tan claves como la definición de gen. Este concepto de gen ha evolucionado y se ha vuelto más complejo durante el transcurso de los años. Antiguamente, el gen era considerado como una unidad hereditaria de información en un cromosoma, el cual codifica para una proteína (16).

Actualmente, ENCODE define al gen como la unión de secuencias genómicas que codifican un conjunto coherente de productos funcionales potencialmente superpuestos. De esta forma, se sostiene que el genotipo determina el fenotipo, y esto a nivel molecular significa que las secuencias de ADN determinan las secuencias de las moléculas funcionales. Por tanto, una secuencia de ADN que codifica para una proteína o productos de ARN los cuales deben ser funcionales (16). El gen no puede ser comprendido como un elemento unitario y determinista que resulta en la expresión de una única secuencia del

genoma ya que los diferentes elementos del genoma interactúan entre sí de manera directa o indirecta. Esto nos lleva a replantear o completar el dogma central de la biología y afrontar nuestro genoma como un sistema complejo y considerar la relevancia de las redes biológicas (1).

ENCODE ha proporcionado información suficiente que sugiere que el genoma se transcribe más de lo que habiamos considerado inicialmente, de forma tal que la mayoría de sus bases se asocia por lo menos a un transcrito primario y muchos transcritos unen regiones distales para establecer *loci* proteinocodificantes. Así mismo, se han identificado transcritos no codificantes para proteínas o trascriptos de función desconocida (TUFs), los cuales parecen tener un potencial de codificación reducido y se distribuyen principalmente en las regiones intergénicas y junto a los limites del gen (17). Por otra parte, se ha demostrado que las secuencias reguladoras que rodean los sitios de inicio de la trascripción están distribuidas de manera simétrica y que el acceso a la cromatina y los patrones de modificación de histonas son altamente predictivos de la actividad y presencia de estos. Finalmente, se ha demostrado que los sitios hipersensibles a la DNAasaI distales tienen patrones característicos de modificaciones de histonas que los diferencias de los promotores (17).

Por otra parte, el proyecto encontró que la mayoría de genoma estaba representado por los trascritos primarios las cuales se superponen ampliamente e incluyen muchas regiones que no codifican para proteínas. Anteriormente, se creía que existía una cadena de trascriptos, sin embargo, ENCODE propone que existen unidades de transcripción simples. De tal

forma, se identificaron nuevos TSSs y se describen más ampliamente la disposición de las secuencias reguladoras alrededor de TSSs. Así mismo, ENCODE describió que aquellos enhancers que habían sido mapeados distalmente a los TSSs ahora se encontraban cercanos a los TSSs y a los transcriptos, lo cual sugiere un rol en la regulación proximal de la expresión (3).

En definitiva ENCODE está haciendo una caracterización completa de la función del genoma humano, una vez identificadas las partes ahora es indispensable saber cómo interactúan entre sí y funcionan como unidad, y luego el impacto de sus alteraciones en la patología humana(3). Es claro el impacto que puede tener la generación de este conocimiento en las ciencias aplicadas y en la salud humana respalda el esfuerzo impresionante que ha planteado esta iniciativa(3).

Conclusión: importancia y perspectivas

El proyecto de ENCODE ha enriquecido la anotación de la secuencia del ADN humano mediante la descripción de los elementos funcionales en el 60% de la secuencia, sin embargo, se sugiere que es necesario mapear más elementos con el fin de identificar el 40% restante (3). ENCODE ha permitido la atribución de funciones bioquímicas al 80% del genoma. Estos elementos identificados han sido relacionados con variantes en la secuencia que están relacionadas al desarrollo de enfermedades. A raíz de los descubrimientos que ha

logrado ENCODE se ha alcanzado una mejor comprensión de las enfermedades especialmente de aquellas que tienen un componente genético generando gran impacto sobre la salud y la enfermedad. En consecuencia, el proyecto proporciona nuevos conceptos sobre la organizacion y regulacion de los genes, y del genoma, lo cual es vital para las investigaciones del campo médico (18). Igualmente, se han generado nuevas inquietudes que requieren mayor profundización y para el futuro, ENCODE, pretende reunir toda la información obtenida con la de proyectos asociados con el fin de construir nuevas aproximaciones para comprender el funcionamiento del genoma humano en la salud y la enfermedad (18).

Gracias a ENCODE ahora existe una sólida evidencia de que existen millones de elementos regulatorios en regiones no codificantes que hacen parte del genoma humano. Por otra parte, el proyecto ENCODE ha brindado una visión global sobre el transcriptoma humano y ha encontrado que la región transcrita del genoma humano es más abundante que lo que se pensaba. Así mismo, toda los datos que ha generado ENCODE han permitido que se creen varias herramientas con el fin de aprovechar esta información de una mejor manera (19). Las principales conclusiones derivadas de el proyecto ENCODE se relacionan con la epigenética. En este sentido, se propone que en estudios futuros se realice un mapeo utilizando aquellos factores y tipos de células que no se hayan investigado antes. Para este propósito han surgido nuevos proyectos por ejemplo, el *International Epigenome Consortium* (IHEC) y *NIH Roadmap Epigenomics project* (20). Por lo tanto se espera que la mayor comprensión de la organización del genoma y epigenoma en estados normales y

patológicos continuarán aportando nuevas dianas farmacológicas y biomarcadores para la ciencia médica (21).

Uno de los objetivos en el futuro para ENCODE y proyectos relacionados será capturar los aspectos dinámicos de la regulación de los genes y de la organización del genoma. Otra finalidad que surge es identificar cómo se combinan los elementos genómicos para crear las redes de genes y las vías bioquímicas que llevan a cabo funciones complejas lo cual permite finalmente el desarrollo de órganos y tejidos que a la vez son necesarias para comprender los patrones de desarrollo de un fenotipo (22). Finalmente, hemos pasado de hablar de "ADN basura" ha afrontar los vacios de conocimiento en genética humana y el proyecto ENCODE ha cambiado nuestra visión de la genética, ha dado respuestas a interrogantes y planteado nuevos desafios que dan luces esperanzadoras a la medicina.

Bibliografía

(1) Moraes F, Geós A. A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. Biochemistry and Molecular Biology Education March 7, 2016;44(3):215-223.

(2) Pardo A. El genoma humano. Límites y perspectivas en el avance de la medicina. Archivos de Bronconeumología 2004 January 1,;40(3):133-138.

(3) Weinstock GM. ENCODE: more genomic empowerment. Genome Res 2007 Jun;17(6):667-668.

(4) National Human Genome Research Institute. Department of Health and Human

Services National Institutes of Health; Available at: https://www.genome.gov. Accessed 25/04/2018.

(5) National Human Genome Research Institute. NHGRI FY 2018 Justification of Estimates for Congressional Appropriations Commitee; Available at: https://www.genome.gov/about-nhgri/Budget-Financial-Information. Accessed Septiembre 5 de 2017.

(6) Qu H, Fang X. A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. Genomics Proteomics Bioinformatics 2013 Jun;11(3):135-141.

(7) The ENCODE project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012 Sep 06,;489(7414):57-74.

(8) Lomelin D, Jorgenson E, Risch N. Human genetic variation recognizes functional elements in noncoding sequence. Genome research 2010 Mar;20(3):311-319.

(9) Greenbaum JA, Parker SCJ, Tullius TD. Detection of DNA structural motifs in functional genomic elements. Genome research 2007 Jun;17(6):940-946.

(10) Yavartanoo M, Choi JK. ENCODE: A Sourcebook of Epigenomes and Chromatin Language. Genomics Inform 2013 Mar;11(1):2-6.

(11) Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome research 2012 Sep;22(9):1760.

(12) Dopazo Blázquez J. Bioinformática médica. In: Cardellach López F, Rozman Borstnar C, editors. Farreras Rozman. Medicina Interna. Decimoctava edición ed. España, S.L.U.: ELSEVIER; 2016. p. 1139-1143.

(13) The ENCODE project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol 2011 Apr;9(4):e1001046.

(14) Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, et al. ENCODE data at the ENCODE portal. Nucleic Acids Res 2016 Jan 04,;44(D1):726.

(15) Manolis Kellis, Barbara Wold, Michael P. Snyderd, Bradley E. Bernstein, Anshul Kundaje, Georgi K. Marinov, et al. Defining functional DNA elements in the human genome. Proceedings of the National Academy of Sciences of the United States of America 2014 Apr 29,;111(17):6131-6138.

(16) Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, et al. What is a gene, post-ENCODE? History and updated definition. Genome research 2007 Jun;17(6):669-681.

(17) Consortium TEP. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799-816 2007 Oct 1,;447(7146):799-816.

(18) The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012 Sep 06,;489(7414):57-74.

(19) Shen Li, Choi Inchan, Nestler Eric J, Won Kyoung-Jae. Human Transcriptome and Chromatin Modifications: An ENCODE Perspective. Genomics & Informatics. March 13, 2013;11(2):60-67.

(20) Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 2012 Sep 06 ;489(7414):83-90.

(21) Siggens L, Ekwall K. Epigenetics, chromatin and genome organization: recent

advances from the ENCODE project. J Intern Med 2014 Sep;276(3):201-214.

(22) Ecker JR, Bickmore WA, Barroso I, Pritchard JK, Gilad Y, Segal E. ENCODE explained: the Encyclopedia of DNA elements project dishes up a hearty banquet of data that illuminate the roles of the functional elements of the human genome. Here, five scientists describe the project and discuss how the data are influencing research directions across many fields. Nature 2012 Sep 6,;489(7414):52.

(23) Cooper G, Hausman R. La Célula Cooper. Quinta ed. Madrid, España: Marbán; 2011.