

# Naturalment

## Article

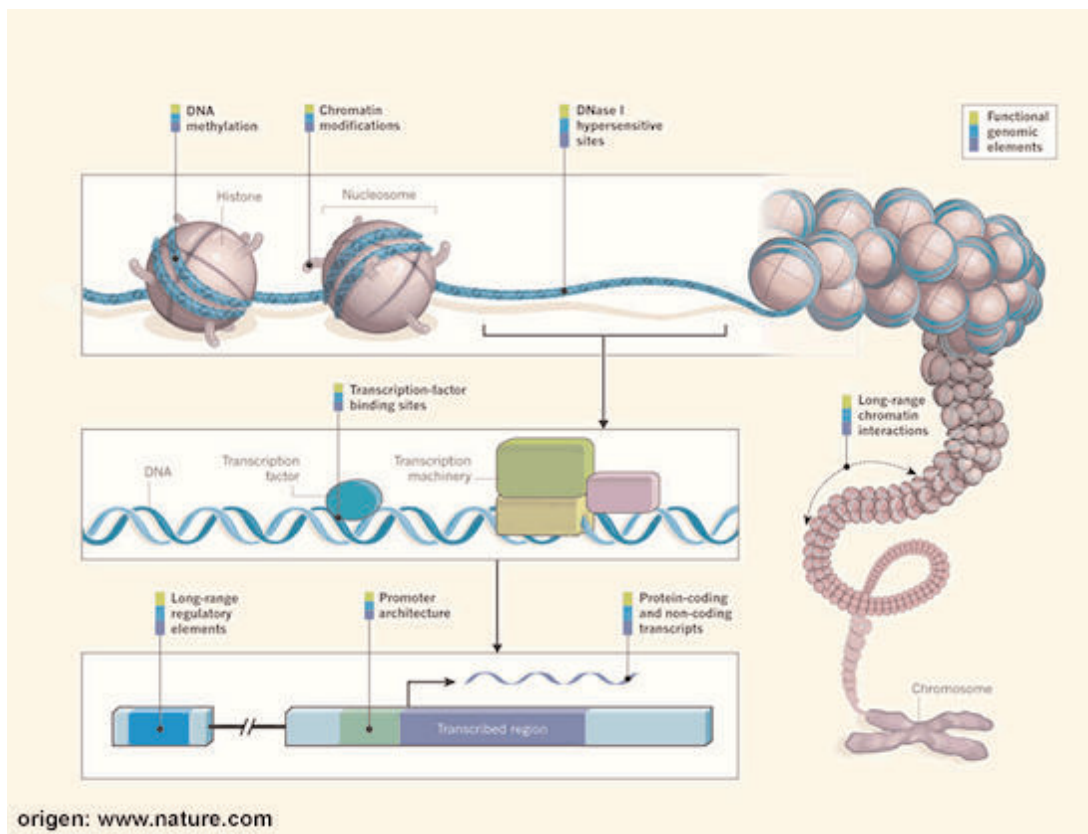
### ENCODE

#### Antecedents

Durant la dècada dels anys 80, els genetistes plantejaren poder seqüenciar els  $3 \times 10^9$  parells de bases del genoma. Es volia seqüenciar el màxim possible i per això el 1986 es va plantejar el PGH (Projecte Genoma Humà) que destinaria diners a la investigació i seqüenciació del genoma humà. El 1988 es va fundar HUGO (Human Genome Organization), que tenia l'objectiu de col·laborar amb una gran quantitat de països en la investigació i poder difondre més fàcilment els avenços que es fessin sobre el genoma.

gens. Altres objectius eren la creació de mapes genètics, la determinació de seqüències completes de diverses espècies i una millora significativa en el camp de la bioinformàtica.

El PGH va començar a l'octubre de 1990. El primer objectiu era desenvolupar nous mètodes automatitzats per clonar i seqüenciar l'ADN i així generar mapes físics i genètics del genoma. Als 3 anys, ja estava fet un mapa físic a gran escala dels 23 cromosomes de l'ésser humà.



El 1990 es va iniciar el projecte. Es destinarien 3.000 \$ milions i en 15 anys hauria d'estar seqüenciat tot el genoma humà. El projecte també pretenia la facilitació del processament de dades i la identificació dels

El Consorci Internacional de Seqüenciació del Genoma Humà pretenia representar la seqüenciació del genoma humà mitjançant l'ús de mapes físics, però un biòleg i empresari nord-americà anomenat

Craig Venter va decidir fundar una empresa privada, *Celera Genomics* que tenia l'objectiu de seqüenciar el genoma humà d'una manera més veloç i acabar abans que el Consorci. Per fer-ho, realitzà un esquema de seqüenciació escollint fragments del genoma a l'atzar, aquesta tècnica es va denominar *shotgun*. El Consorci i l'empresa de Craig Venter avançaven a la mateixa velocitat. A l'estiu de 2000 van treure el primer esborrany. No obstant això fins a 2003 no estaria acabada i declarada completa encara que quedessin alguns esclatxes per completar. Segons HUGO, la tècnica de *Celera Genomics* era imprecisa però molt ràpida, en canvi el Consorci encara que era més lenta era molt exacta.

Les dades més significatives van ser el nombre de gens que van sortir 30.000-35.000 en comptes dels 150.000 gens esperats. El 60% presenten homologies amb altres genomes seqüenciats prèviament. Només 94 famílies de 1278 eren de vertebrats, la qual cosa indica que hi ha més similitud amb els invertebrats. Els exons ocupaven 1.1 % del genoma, introns 24% i el 75% restant era ADN intergènic i el 44% del mateix derivaria d'elements transposables. Als 3 anys es va treure la seqüència definitiva del 99% de les regions eucromàtiques del genoma humà. El nombre total de gens que codifiquen proteïnes era de 20.000-25.000, menys dels que en el primer esborrany van sortir i que comparats amb altres espècies, en tenim menys.

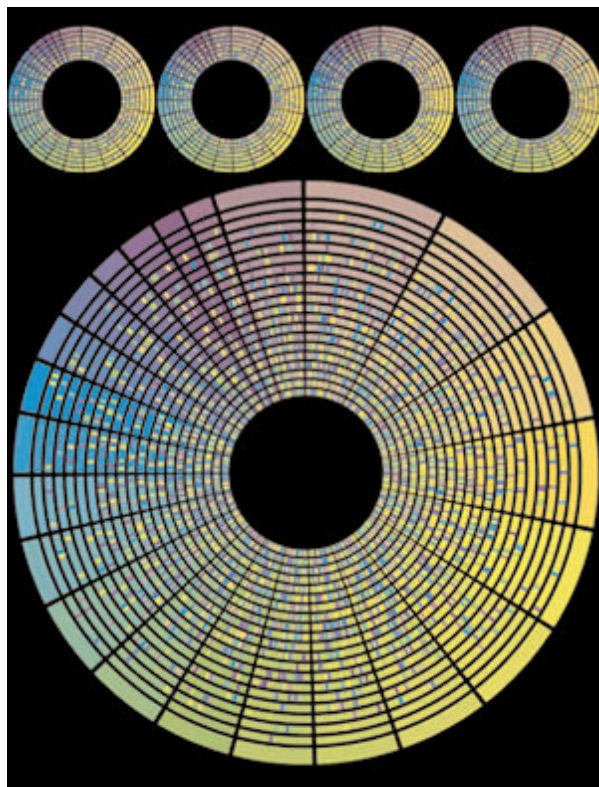
Cada cèl·lula del nostre cos s'expressa amb només alguns dels gens que conté. Per això una cèl·lula dels músculs és diferent d'una neurona o una de la pell. Al seu torn, els gens estan permanentment fabricant proteïnes per canviar les deteriorades o respondre a canvis en l'ambient. De manera que la clau està en l'expressió dels gens. Si examinem el nombre de gens i proteïnes veiem que hi ha deu vegades més de les últimes. És a dir, cada gen pot expressar-se de diferent manera i de mitjana pot fabricar 10 proteïnes. Un dels misteris que amaga el genoma és com tan pocs gens poden construir un organisme complex com l'ésser humà. És absurd pensar que l'ésser humà manté moltíssima quantitat de gens que no codifiquen a proteïnes i considerar que no són servibles, donant per fet que en no molestar, l'organisme mantingui tal malbaratament d'energia. A partir d'aquí els científics que es van plantejar què feien realment els gens que no codificaven a proteïnes, vulgarment conegut més com l'ADN escombraries i van decidir investigar i indagar més en aquests gens, pel que van decidir començar

el projecte ENCODE.

### El projecte ENCODE

Els científics saben que els gens juguen un paper important en qui som, però al mateix temps una bona quantitat d'informació sobre la seva funció segueix sent un gran misteri.

És per això que un equip internacional d'investigadors es va disposar a esbrinar quina és la feina que exerceixen algunes parts del genoma humà i el que



origen: mkweb.bcgsc.ca

signifiquen per al cos humà tal com el coneixem. És una base de dades massiva a l'abast de la població. El seu objectiu és catalogar les seqüències d'ADN funcionals que resideixen en el genoma, i aprendre quan i en quines cèl·lules actuen. També es pretenia conèixer com s'empaqueta l'ADN en els cromosomes i la seva relació amb el mode de lectura i de regulació.

### Característiques

El projecte es diu l'Enciclopèdia d'Elements de l'ADN (ENCODE, per les sigles en anglès). Quan el Projecte del Genoma Humà va fer la seqüència genètica del cos el 2003, es va establir l'ordre dels 3,000 milions de lletres del genoma, les quals poden ser considerades com "el llibre de la vida".

ENCODE tracta la interpretació i comprensió de les

variants genètiques que afecten les funcions bioquímiques, i les que estan associades amb malalties.

"Aquesta és la ciència d'aquest segle", va dir Ewan Birney, de l'Institut Europeu de Bioinformàtica i el coordinador d'anàlisi de Encode, en una conferència de premsa. "En aquest segle estarem treballant en com som els humans a partir d'aquest manual d'instruccions simples".

National Human Genome Research Institute



ENCODE ha començat a revelar els circuits, xarxes i coreografia en el genoma humà, va dir el doctor Eric Green, director de l'Institut Nacional d'Investigació del Genoma Humà nord-americà.

Després d'una fase pilot, els científics d'ENCODE aplicaren els seus nous mètodes a l'estudi del genoma i quan va arribar a la seva fi tots els resultats foren publicats en més de trenta articles científics a les revistes *NATURE*, *Genome Research* i *Genome Biology*. (veure secció de webs de la revista).

Les investigacions d'ENCODE conduïren a conèixer les funcions de un 80% del genoma, el que abans era conegut com ADN escombraries. Es donaren a conèixer més de 70.000 regions promotores. Els nous segments d'ADN investigats contenen llocs d'acoblament que compleixen diferents funcions, com ser llocs on altres proteïnes s'uneixen i activen o desactiven l'expressió dels gens. O bé, són fragments que es llegeixen i es transcriuen a ARN també poden influenciar l'activitat d'altres gens de vegades molt llunyans com els més de 400.000 facilitadors o afecten a la manera en què l'ADN es plega i s'empaqueta.

El projecte conté dades sobre el grau de metilació d'ADN i modificacions químiques d'histones que poden influir en la velocitat de transcripció d'ADN en molècules d'ARN ( histones són les proteïnes al voltant de la qual l'ADN s'enrotlla per formar la cromati-

na). ENCODE també examina les interaccions cromatina de llarg abast, com el looping, que alteren les proximitats relatives de diferents regions cromosòmiques en tres dimensions i també afecten la transcripció.

#### *Resultats del projecte ENCODE.*

Mostra que la majoria d'aquests trams de regions d'ADN en què s'uneixen les proteïnes i molècules d'ARN, en posicions des de les que cooperen entre sí per regular la funció i el nivell d'expressió de gens codificadors de proteïnes. A més, sembla que la transcripció generalitzada a partir d'ADN no codificant potencialment actua com un dipòsit per a la creació de noves molècules funcionals, com ara els ARN reguladors.

Els científics han determinat que hi ha quatre milions de llocs en el genoma on es produeixen determinats esdeveniments bioquímics, la majoria dels quals han estat descoberts amb ENCODE. Això vol dir que si una persona té el seu genoma seqüenciat, podria haver fins a quatre milions de diferències respecte d'una altra persona, va dir Michael Snyder, professor de la Universitat de Stanford, que és investigador principal d'ENCODE,

La majoria dels canvis genètics que poden causar malalties, o estan associats amb una malaltia, no es troben dins dels propis gens, sinó més aviat en aquests elements de control, va dir Snyder.

Aquests llocs són "interruptors" que determinen quins proteïnes i cèl·lules es van a formar, afirmà Richard Myers, de l'Institut de Biotecnologia HudsonAlpha i col·laborador d'ENCODE.

Des de 2005, les associacions que estudien el genoma (GWAS) han posat milers de punts d'una sola lletra (bases nitrogenades) en les que es poden diferenciar els genomes. Una variant, sembla estar associada amb el risc que es produeixi una malaltia. No obstant això, gairebé el 90 % d'aquestes variants queden fora de codificació de la proteïna del gen, de manera que els investigadors tenen poca idea de com podrien causar o influir en la malaltia.

El mapa creat per ENCODE revela que moltes de les regions unides a una malaltia inclouen potenciadors o altres seqüències funcionals i el tipus de cèl·lules és important. El grup de Manolis Kellis, un genetista computacional de l'Institut Tecnològic de Massachusetts, a Cambridge, considerava que algunes de les variants estan fortament associades amb

el lupus eritematós sistèmic, una malaltia en la qual el sistema immunitari ataca els propis teixits del cos. L'equip va notar que les variants identificades en GWAS tendien a ser en regions reguladores del genoma que eren actives en una línia de cèl·lules immunes, però no necessàriament en altres tipus de cèl·lules.

*Nou concepte de gen*

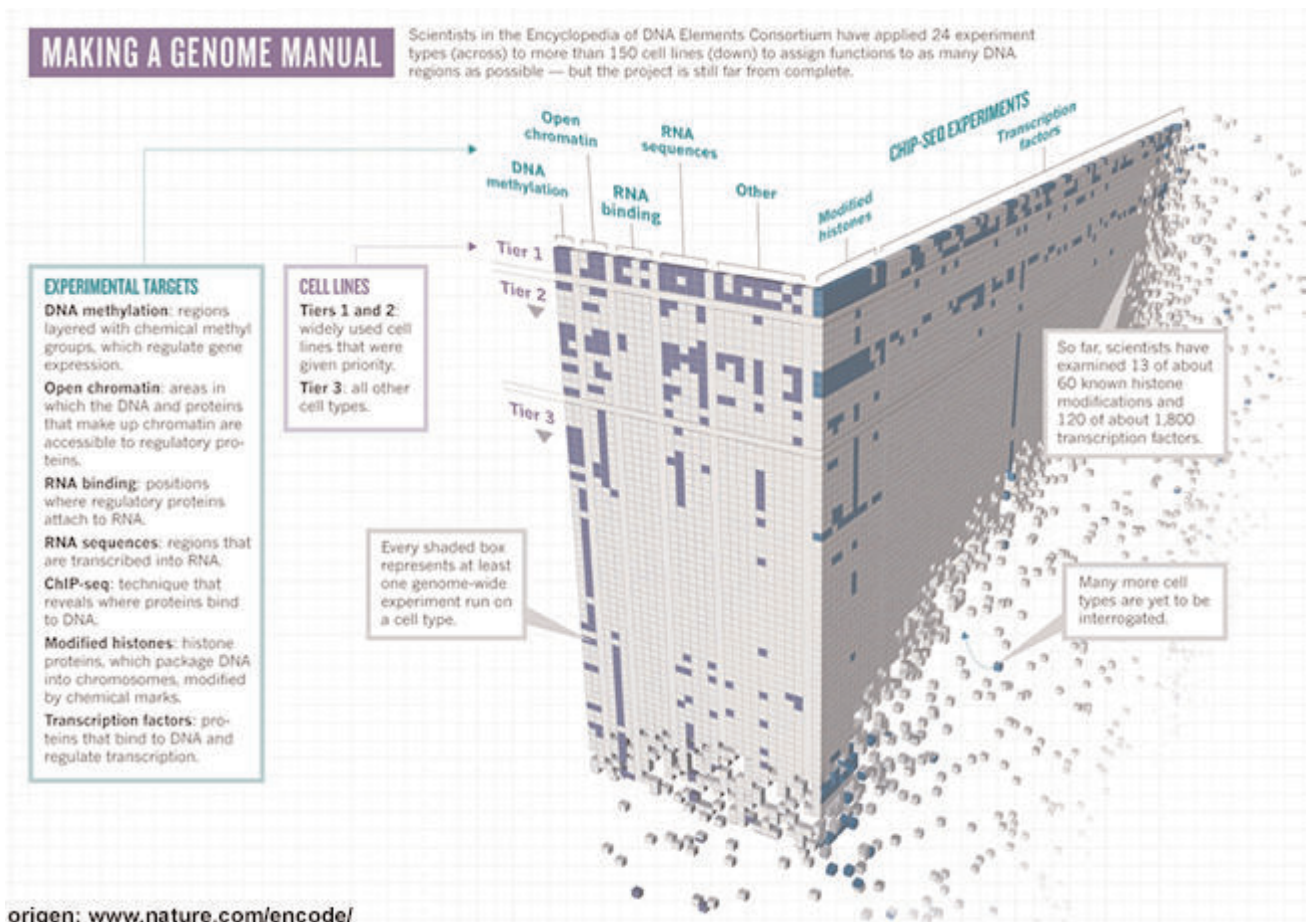
ENCODE ens porta a redefinir el concepte de gen. En alguns casos un gen està dispers per diferents llocs del genoma o la mateixa regió d'ADN pertany a més d'un gen. Potser el nou element bàsic no sigui el gen sinó el transcrit, la porció mínima d'ADN que passa a ARN. Alguns dubten encara que s'hagi eliminat el concepte ADN "escombraries".

**Utilitats**

ENCODE permet als investigadors identificar mutacions que són més susceptibles de causar una malaltia i les respostes de medicaments. En un estudi sobre les associacions amb malalties, els científics van trobar vincles amb les malalties del cor i diabetis tipus II. Els interruptors reguladors poden causar problemes amb els gens i conduir a una malaltia. Snyder que ha estat l'autor d'aquestes investigacions ha seqüenciat el seu propi genoma, creu que la seqüenciació del genoma personal es convertirà en una part important de la cura de la salut.

En el seu propi cas, la seqüenciació del genoma mostrava un risc de patir diabetis tipus II abans que el seu cos comencés a mostrar símptomes. Ja que ell sabia sobre el risc abans de temps, va ser capaç de començar a administrar els seus nivells de glucosa abans que apareguessin els símptomes de la malaltia.

Amb la nova informació d'ENCODE, Snyder podria començar a veure les variants que té que puguin afectar la funció del gen, però que no es troben dins del gen, sinó que resideixen en els interruptors de



control.

No obstant això, hi ha un conflicte sobre la quantitat d'informació d'una seqüència que ha de ser revelada a una persona. Després de tot, només perquè una persona té una connexió genètica a una malaltia no vol dir necessàriament que ell o ella desenvolupi aquesta condició i la informació podria causar una preocupació indeguda.

Una seqüència de genoma com a part d'un projecte de recerca té un cost inferior a 4.000 dòlars, però la interpretació adequada costa molt més, va dir Snyder.

Algunes empreses, com *23andMe*, l'ofereixen a un cost relativament baix, però només part de la seqüència del genoma.

Malgrat el que es coneix fins ara, la seqüenciació del genoma pot beneficiar les persones amb càncer. Aquesta tècnica va ser al tractament del creador d'Apple, Steve Jobs. Ell es va sotmetre a la seqüenciació dels gens del seu tumor de càncer i el seu ADN normal, pel que el seu equip mèdic podria triar fàrmacs que serien millor dirigits cap a la seva malaltia, segons la biografia *Steve Jobs*, de Walter Isaacson.

### Futur del projecte

No obstant això, amb milers de tipus cel·lulars a prova i un conjunt creixent d'eines per provar-los, el projecte podria arribar a no tenir fi. "Estem lluny d'haver acabat", diu el genetista Rick Myers, de l'Institut de Biotecnologia HudsonAlpha a Huntsville, Alabama. Això podria continuar per sempre i això preocupa a algunes persones. El projecte pilot d'ENCODE va costar uns \$ 55 milions, l'ampliació va ser d'uns \$ 130 milions, i el NHGRI va atorgar fins a \$ 123 milions. Alguns investigadors sostenen que encara no han vist un sòlid retorn de la inversió.

"És un projecte que no està conclòs, però és un gran pas endavant", va dir Snyder.

### Bibliografia

Revista Nature (2012). ENCODE. Guidebook to the human genome. *Nature*, vol 489, n. 7414. 45-113.

Nature. ENCODE. consultado [Febrero 2013] Disponible en [[www.nature.com/encode/](http://www.nature.com/encode/)]

The ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, vol. 489, vol 7414. 57-74.

Young, E. (2012). ENCODE: The rough guide to the human genome. *Discover*. Consultat: [Novembre 2012]. Disponible en [[blogs.discovermagazine.com](http://blogs.discovermagazine.com)]

