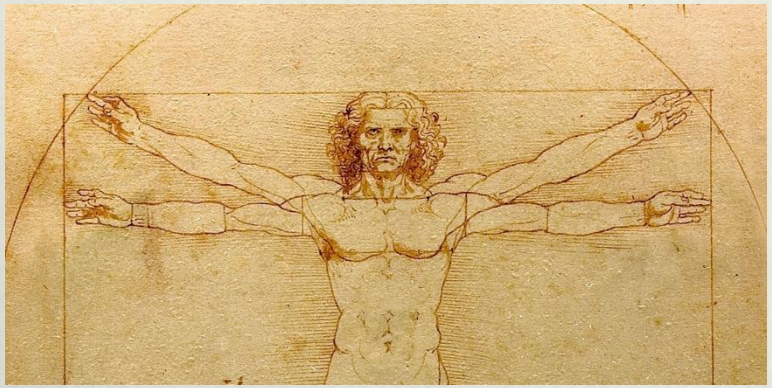
Introdución

Genómica Funcional
Máster en Genética y Evolución

El tamaño del genoma y la función



Homo sapiens

¿Quien tiene el genoma mas grande?

Allium cepa

El tamaño del genoma

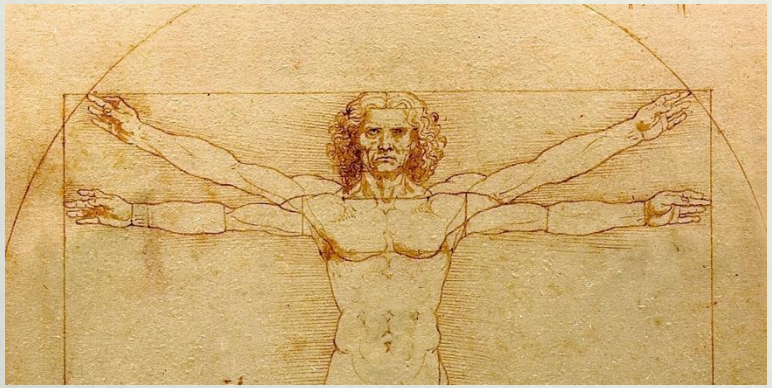
El tamaño del genoma es la cantidad de DNA de una copia de un genoma *haploide*

En eucariotas con genoma diploide se define el tamaño de genoma como la mitad de la cantidad de DNA de una celular diploide

El tamaño del genoma se puede medir en pg (picogramos) o en numero de bases:

Tamaño del genoma(bp) = (0.978×10^9) x contenido en DNA (pg)
Contenido en DNA (pg) = tamaño del genoma (bp) / (0.978×10^9)
1 pg = 978 Mb

El tamaño del genoma y la función



¿Quien tiene el genoma mas grande?

Homo sapiens

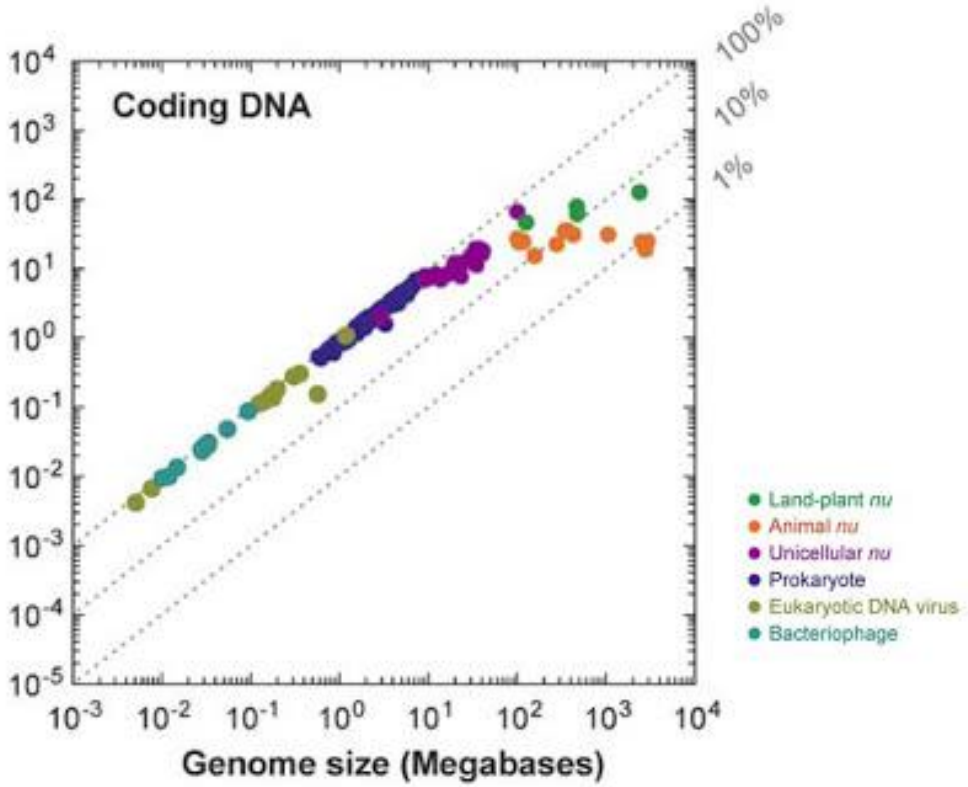
Allium cepa

Aprox. 3,200,000,000 bp

Aprox. 16,600,000,000 bp

¿Todo el genoma es funcional?

¿Correlación entre numero de genes y tamaño del genoma?



En eucariotas (pero no en procariotas) el tamaño del genoma no es proporcional al número de genes y "la complejidad" del organismo

http://www.genomicron.evolverzone.com/2007/09/genome-size-and-gene-number/

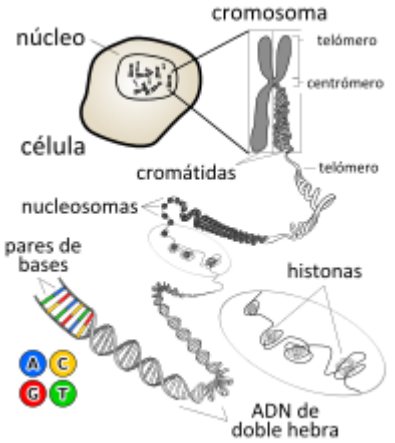
Objetivos de la genómica (funcional)

(i) Analizar el contenido del genoma
(ii) Analizar el funcionamiento del genoma
(iii) Desentrañar la relación entre la secuencia de DNA, el ambiente y el fenotipo
(iv) Determinar la forma en la que evolucionan los genomas

¿Qué tipos de análisis tenemos que llevar a cabo?

- Determinar la secuencia del genoma
- Determinar y caracterizar todos los elementos genómicos existentes
- Anotar todos los elementos genómicos en el genoma, es decir, detectar su ubicación en el genoma (especial enfoque en los genes)
- Averiguar la función de todos los elementos genómicos
- Determinar la expresión del genoma en diferentes tejidos o bajo diferentes condiciones patológicas, es decir el transcriptoma y proteoma
- Analizar la información epi-genética

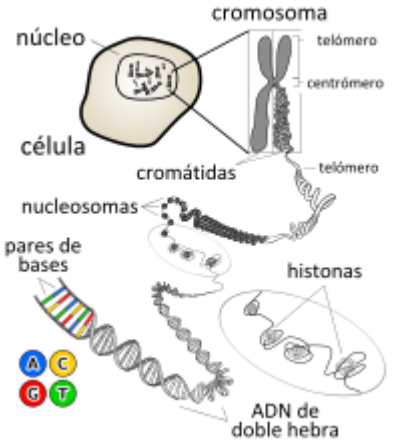
Un proyecto genómico



Secuenciación

5’-
GCAGTGGTTCACAGACGAGCAGGCACCGGAATCACCTACAAGGCGTGCTGGC
AAAACACCGATTGCCGGGTCCCACCCCCAGAGTTTCTGATTCAGTACGGGAGT
CTCGCTCCGTTGCCCAGGCTGGAGTGCAATGGCGTGATCTCGGCTTACTACAA
CCTCCTCCTCCCGGGCTCAAGAGATTTTCCTGCGTCAGCCTTCTGAGTAGCTGG
GATTACAGGCACACGCCACCACACCCAGCTAATTTTTGCATTTTTAGTAGAGG
CGGGTGAAACCCCGTCTCCACTAAAAGAGATCACGGGGTCTCGATCTCCTGAA
CTCGTGATCCACCTGCCTTGGCCTCCCAAAGTGTTGGGATTACATGCATAAGC
CACCATGCCAAGCCAAGAATTTGCATTTCTAACAATCTGCCAGGGGATGCCGA
TGCTGCTTGTCTGGAATTCACACTTTGAGAGCCACTGTGTGGAGTCATCTGGT
AGCTAAGTTTTAGTAAAGGCTTTCTGGTGTACTAAGTCATTGAGAAAATGATT
AGATTGTAATAACTGGTAACAAATCCTTTTACAAGTTAAATTGTTTTCAATCTTT
TATTTTCAATAAA-3’

Un proyecto genómico



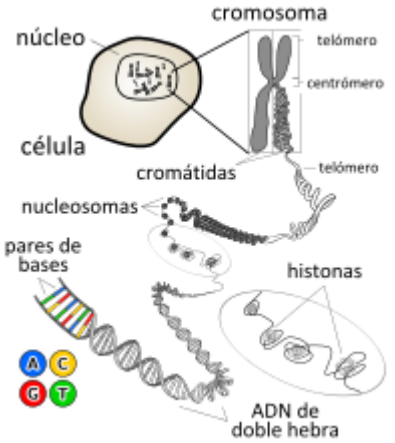
Secuenciación

5’-
GCAGTGGTTCACAGACGAGCAGGCACCGGAATCACCTACAAGGCGTGCTGGC
AAAA**CACCGATTGCCGGGTCCCACCCCCAGAGTTTCTGATTCAGTACGGGAG**
TCTCGCTCCGTTGCCCAGGCTGGAGTGCAATGGCGTGATCTCGGCTTACTAC
AACCTCCTCCTCCCGGGCTCAAGAGATTTTCCTGCGTCAGCCTTCTGAGTAGCT
GGGATTACAGGCACACGCCACCACACCCAGCTAATTTTTGCATTTTTAGTAGA
GGCGGGTGAAACCCCGTCTCCACTAAAAGAGATCA**CGGGGTCTCGATC**TCCT
GAACT**CGTGATCCACCTGCCTTGGCCTCCCAAAGTGTTGGGATTACATGCATA**
AGCCACCATGCCAAGCCAAGAATTTGCATTTCTAACAATCTGCCAGGGGATGC
CGATGCTGCTTGTCTGGAATTCACACTTTGAGAGCCACTGTGTGGAGTCATCT
GGTAGCTAAGTTTTAGTAAAGG**C**TTTCTGGTGTACTAAGTCATTGAGAAAATG
ATTAGATTGTAATAACTGGTAACAAATCCTTTTACAAGTTAAATTGTTTTCAATC
TTTTATTTTCAATAAA-3’

Anotación

Genes
Elementos regulatorios
Variación / SNP
Repetido

Objetivos



Secuenciación

5’-
GCAGTGGTTCACAGACGAGCAGGCACCGGAATCACCTACAAGGCGTGCTGGC
AAAA**CACCGATTGCCGGGTCCCACCCCCAGAGTTTCTGATTCAGTACGGGAG**
TCTCGCTCCGTTGCCCAGGCTGGAGTGCAATGGCGTGATCTCGGCTTACTAC
AACCTCCTCCTCCCGGGCTCAAGAGATTTTCCTGCGTCAGCCTTCTGAGTAGCT
GGGATTACAGGCACACGCCACCACACCCAGCTAATTTTTGCATTTTTAGTAGA
GGCGGGTGAAACCCCGTCTCCACTAAAAGAGATCA**CGGGGTCTCGATC**TCCT
GAACT**CGTGATCCACCTGCCTTGGCCTCCCAAAGTGTTGGGATTACATGCATA**
AGCCACCATGCCAAGCCAAGAATTTGCATTTCTAACAATCTGCCAGGGGATGC
CGATGCTGCTTGTCTGGAATTCACACTTTGAGAGCCACTGTGTGGAGTCATCT
GGTAGCTAAGTTTTAGTAAAGG**C**TTTCTGGTGTACTAAGTCATTGAGAAAATG
ATTAGATTGTAATAACTGGTAACAAATCCTTTTACAAGTTAAATTGTTTTCAATC
TTTTATTTTCAATAAA-3’

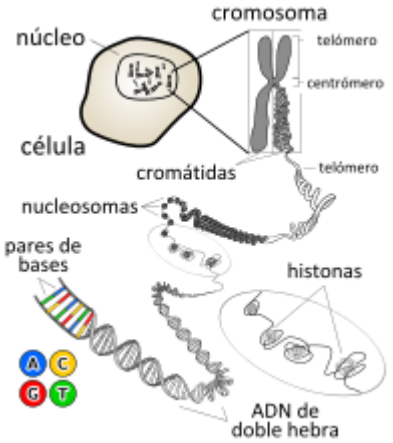
Anotación

Genes
Elementos regulatorios
Variación / SNP
Repetido

Expresión del genoma

Transcriptoma
Proteoma

Un proyecto genómico



Secuenciación

5’-
GCAGTGGTTCACAGACGAGCAGGCACCGGAATCACCTACAAGGCGTGCTGGC
AAAA**CACCGATTGCCGGGTCCCACCCCCAGAGTTTCTGATTCAGTACGGGAG**
TCTCGCTCCGTTGCCCAGGCTGGAGTGCAATGGCGTGATCTCGGCTTACTAC
AACCTCCTCCTCCCGGGCTCAAGAGATTTTCCTGCGTCAGCCTTCTGAGTAGCT
GGGATTACAGGCACACGCCACCACACCCAGCTAATTTTTGCATTTTTAGTAGA
GGCGGGTGAAACCCCGTCTCCACTAAAAGAGATCA**CGGGGTCTCGATC**TCCT
GAACT**CGTGATCCACCTGCCTTGGCCTCCCAAAGTGTTGGGATTACATGCATA**
AGCCACCATGCCAAGCCAAGAATTTGCATTTCTAACAATCTGCCAGGGGATGC
CGATGCTGCTTGTCTGGAATTCACACTTTGAGAGCCACTGTGTGGAGTCATCT
GGTAGCTAAGTTTTAGTAAAGG**C**TTTCTGGTGTACTAAGTCATTGAGAAAATG
ATTAGATTGTAATAACTGGTAACAAATCCTTTTACAAGTTAAATTGTTTTCAATC
TTTTATTTTCAATAAA-3’

Regulación de la expresión génica

Factores de transcripción
Metilación del DNA
Regulación postranscripcional

Expresión del genoma

Transcriptoma
Proteoma

Anotación

Genes
Elementos regulatorios
Variación / SNP
Repetido

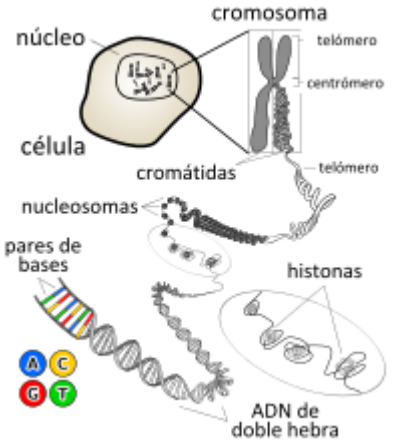
Un proyecto genómico

Fenotipo

Tejido, estado fisiopatológico (cáncer, diabetes, Alzheimer, etc.)

Secuenciación

5’-
GCAGTGGTTCACAGACGAGCAGGCACCGGAATCACCTACAAGGCGTGCTGGC
AAAA**CACCGATTGCCGGGTCCCACCCCCAGAGTTTCTGATTCAGTACGGGAG**
TCTCGCTCCGTTGCCCAGGCTGGAGTGCAATGGCGTGATCTCGGCTTACTAC
AACCTCCTCCTCCCGGGCTCAAGAGATTTTCCTGCGTCAGCCTTCTGAGTAGCT
GGGATTACAGGCACACGCCACCACACCCAGCTAATTTTTGCATTTTTAGTAGA
GGCGGGTGAAACCCCGTCTCCACTAAAAGAGATCA**CGGGGTCTCGATC**TCCT
GAACT**CGTGATCCACCTGCCTTGGCCTCCCAAAGTGTTGGGATTACATGCATA**
AGCCACCATGCCAAGCCAAGAATTTGCATTTCTAACAATCTGCCAGGGGATGC
CGATGCTGCTTGTCTGGAATTCACACTTTGAGAGCCACTGTGTGGAGTCATCT
GGTAGCTAAGTTTTAGTAAAGG**C**TTTCTGGTGTACTAAGTCATTGAGAAAATG
ATTAGATTGTAATAACTGGTAACAAATCCTTTTACAAGTTAAATTGTTTTCAATC
TTTTATTTTCAATAAA-3’



Regulación de la expresión génica

Factores de transcripción
Metilación del DNA
Regulación postranscripcional

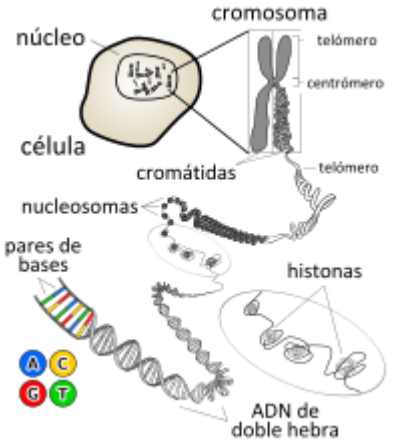
Expresión del genoma

Transcriptoma
Proteoma

Anotación

Genes
Elementos regulatorios
Variación / SNP
Repetido

Un proyecto genómico



Fenotipo

Tejido, estado fisiopatológico (cáncer, diabetes, Alzheimer, etc.)

Secuenciación

5’-
GCAGTGGTTCACAGACGA GGCACCGGAATCACCTACAAGGCGTGCTGGC
AAAA**CACCGATTGCCGGGTCC CCCAGAGTTTCTGATTCAGTACGGGAG**
TCTCGCTCCGTTGCCCAGGCTGGA CAATGGCGTGATCTCGGCTTACTAC
AACCTCCTCCTCCCGGGCTCAAGAGAT CTGCGTCAGCCTTCTGAGTAGCT
GGGATTACAGGCACACGCCACCACACCCAC ATTTTTGCATTTTTAGTAGA
GGCGGGTGAAACCCCGTCTCCACTAAAAGAGA **GGGGTCTCGATC**TCCT
AACT**CGTGATCCACCTGCCTTGGCCTCCCAAAGTC GGATTACATGCATA**
CACCATGCCAAGCCAAGAATTTGCATTTCTAACAA CCAGGGGATGC
CG GCTGCTTGTCTGGAATTCACACTTTGAGAGCCACTGTG AGTCATCT
GGT TAAGTTTTAGTAAAGG**C**TTTCTGGTGTACTAAGTCATTG AAATG
ATTAG TGTAATAACTGGTAACAAATCCTTTTACAAGTTAAATTGTTT ATC
TTTTATT AATAAA-3’

Regulación de la expresión génica

Factores de transcripción
Metilación del DNA
Regulación postranscripcional

Expresión del genoma

Transcriptoma
Proteoma

Anotación

Genes
Elementos regulatorios
Variación / SNP
Repetido

Los resultados de ENCODE: mas del 80% del genoma tiene funciones bioquímicas

ARTICLE

doi:10.1038/nature11247

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions. Many discovered candidate regulatory elements are physically associated with one another and with expressed genes, providing new insights into the mechanisms of gene regulation. The newly identified elements also show a statistical correspondence to sequence variants linked to human disease, and can thereby guide interpretation of this variation. Overall, the project provides new insights into the organization and regulation of our genes and genome, and is an expansive resource of functional annotations for biomedical research.

- The vast majority (80.4%) of the human genome participates in at least one biochemical RNA- and/or chromatin-associated event in at least one cell type. Much of the genome lies close to a regulatory event: 95% of the genome lies within 8 kilobases (kb) of a DNA–protein interaction (as assayed by bound ChIP-seq motifs or DNase I footprints), and 99% is within 1.7 kb of at least one of the biochemical events measured by ENCODE.

Los ecos de ENCODE: el evangelio libre de evolución

On the Immortality of Television Sets: "Function" in the Human Genome According to the Evolution-Free Gospel of ENCODE

Dan Graur[1,*], Yichen Zheng[1], Nicholas Price[1], Ricardo B.R. Azevedo[1], Rebecca A. Zufall[1], and Eran Elhaik[2]

[1]Department of Biology and Biochemistry, University of Houston

[2]Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health

*Corresponding author: E-mail: dgraur@uh.edu.

Abstract

A recent slew of ENCyclopedia Of DNA Elements (ENCODE) Consortium publications, specifically the article signed by all Consortium members, put forward the idea that more than 80% of the human genome is functional. This claim flies in the face of current estimates according to which the fraction of the genome that is evolutionarily conserved through purifying selection is less than 10%. Thus, according to the ENCODE Consortium, a biological function can be maintained indefinitely without selection, which implies that at least 80 − 10 = 70% of the genome is perfectly invulnerable to deleterious mutations, either because no mutation can ever occur in these "functional" regions or because no mutation in these regions can ever be deleterious. This absurd conclusion was reached through various means, chiefly by employing the seldom used "causal role" definition of biological function and then applying it inconsistently to different biochemical properties, by committing a logical fallacy known as "affirming the consequent," by failing to appreciate the crucial difference between "junk DNA" and "garbage DNA," by using analytical methods that yield biased errors and inflate estimates of functionality, by favoring statistical sensitivity over specificity, and by emphasizing statistical significance rather than the magnitude of the effect. Here, we detail the many logical and methodological transgressions involved in assigning functionality to almost every nucleotide in the human genome. The ENCODE results were predicted by one of its authors to necessitate the rewriting of textbooks. We agree, many textbooks dealing with marketing, mass-media hype, and public relations may well have to be rewritten.

Los ecos de ENCODE: definición de función y tamaño del genoma

Is junk DNA bunk? A critique of ENCODE

W. Ford Doolittle[1]

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Canada B3H 4R2

Edited by Michael B. Eisen, Howard Hughes Medical Institute, University of California, Berkeley, CA, and accepted by the Editorial Board February 4, 2013 (received for review December 11, 2012)

Do data from the Encyclopedia Of DNA Elements (ENCODE) project render the notion of junk DNA obsolete? Here, I review older arguments for junk grounded in the C-value paradox and propose a thought experiment to challenge ENCODE's ontology. Specifically, what would we expect for the number of functional elements (as ENCODE defines them) in genomes much larger than our own genome? If the number were to stay more or less constant, it would seem sensible to consider the rest of the DNA of larger genomes to be junk or, at least, assign it a different sort of role (structural rather than informational). If, however, the number of functional elements were to rise significantly with C-value then, (*i*) organisms with genomes larger than our genome are more complex phenotypically than we are, (*ii*) ENCODE's definition of functional element identifies many sites that would not be considered functional or phenotype-determining by standard uses in biology, or (*iii*) the same phenotypic functions are often determined in a more diffuse fashion in larger-genomed organisms. Good cases can be made for propositions *ii* and *iii*. A larger theoretical framework, embracing informational and structural roles for DNA, neutral as well as adaptive causes of complexity, and selection as a multilevel phenomenon, is needed.