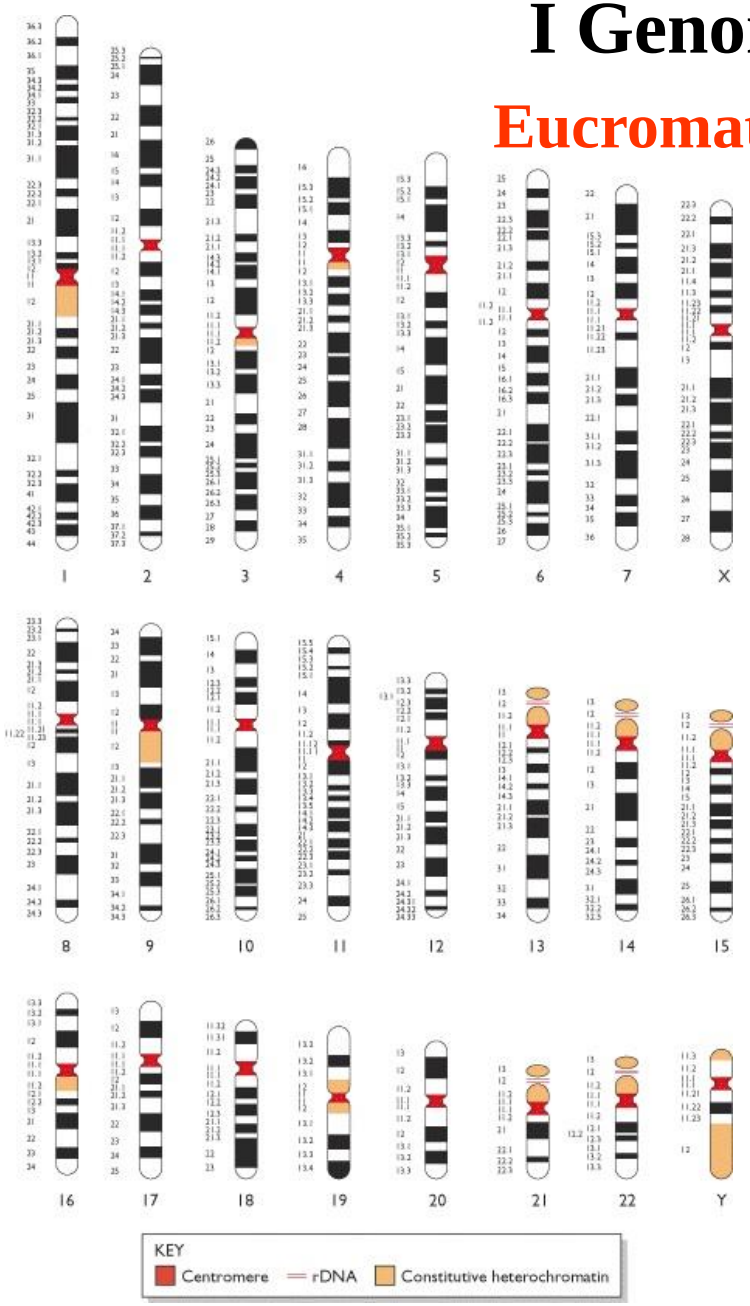
I Genomi degli Eucarioti:

Eucromatina ed Eterocromatina



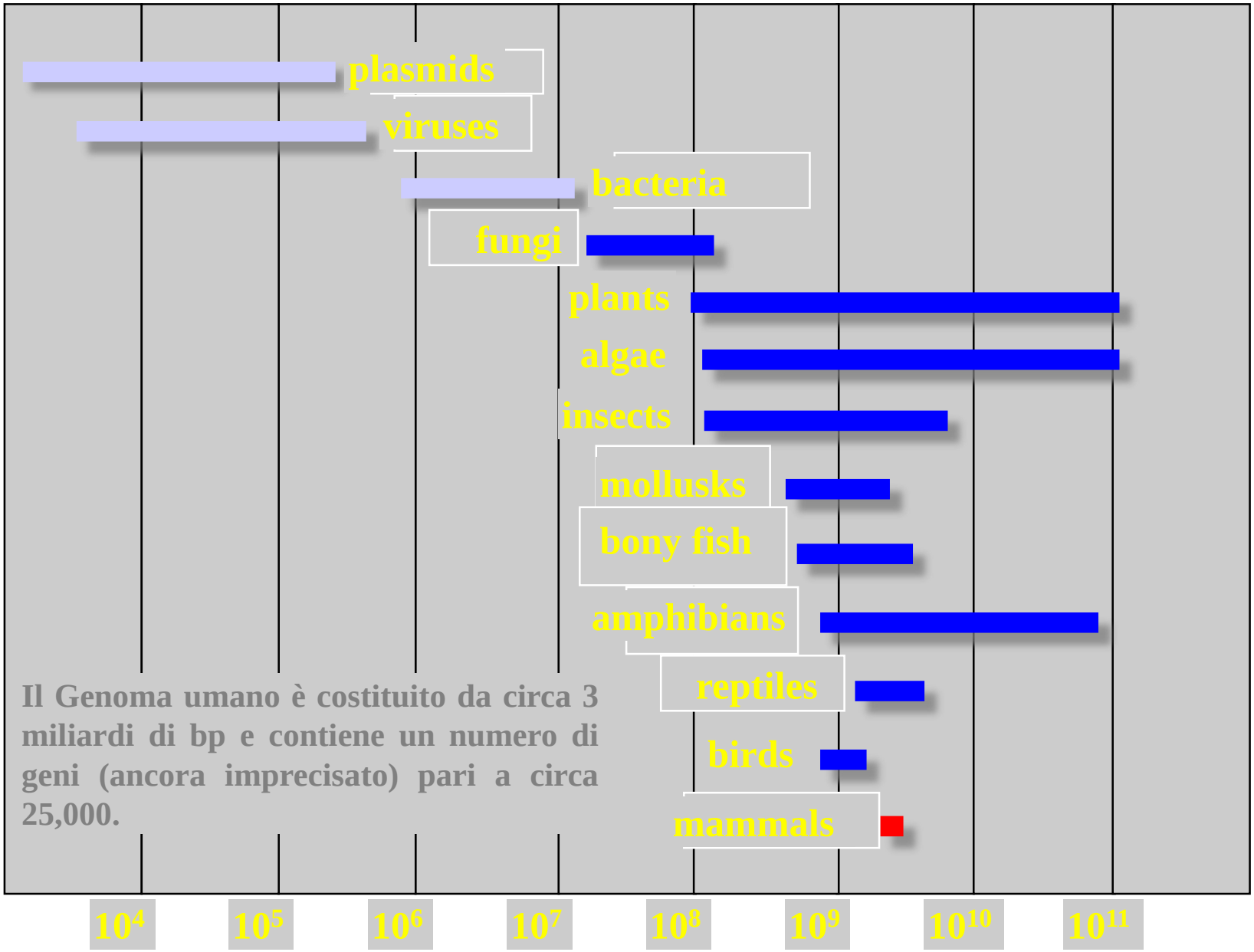
Eucromatina: regioni cromosomiche non condensate, attivamente trascritte e ad alta densità genica.

Eterocromatina: (facoltativa o costitutiva): cromatina mediamente o altamente condensata e generalmente non trascritta, ad alta percentuale di sequenze ripetute e contenuto di geni relativamente basso. Comprende regioni telomeriche e centromeriche.

In una tipica cellula di mammifero il contenuto di eterocromatina è pari a circa il 10%. in *Drosophila melanogaster* costituisce il 34% del genoma totale.

Brown TA, Genomi 2, Fig. 2.8)

Dimensioni dei Genomi Eucariotici



Il Genoma umano è costituito da circa 3 miliardi di bp e contiene un numero di geni (ancora imprecisato) pari a circa 25,000.

Paradosso del Valore C

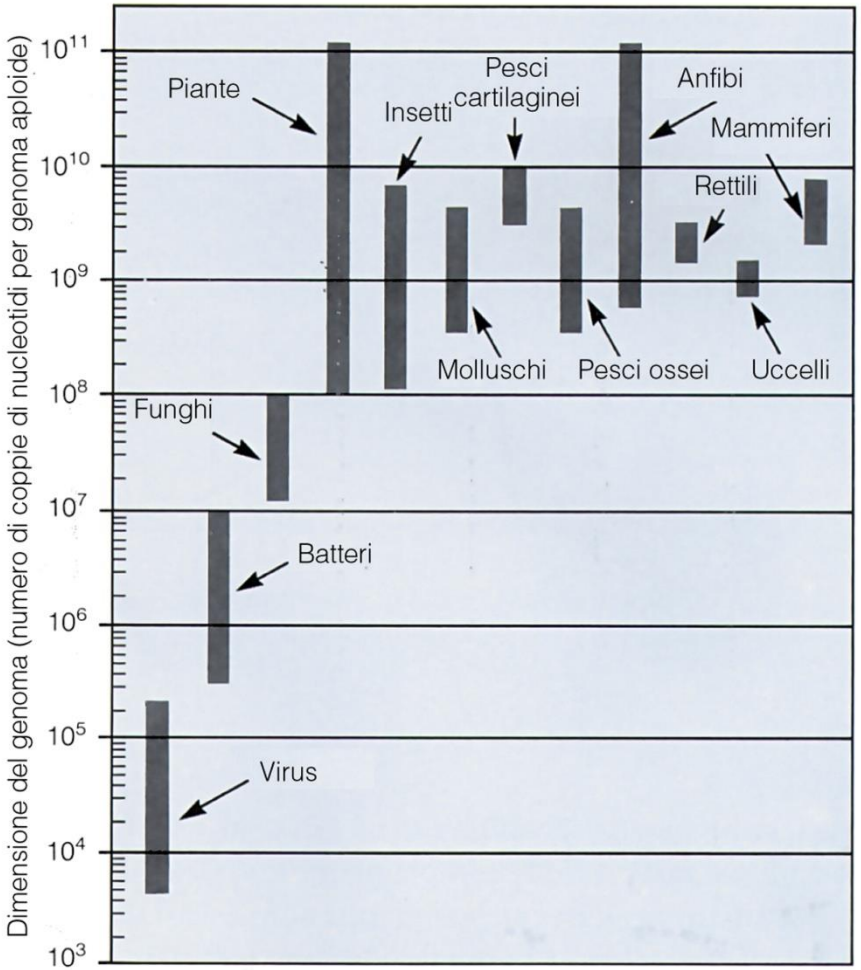


Figura 16-11 Correlazione fra dimensione del genoma e tipo di organismo. Per ogni gruppo di organismi la barra rappresenta la variabilità approssimativa della dimensione del genoma, misurata come quantità di coppie di nucleotidi per genoma aploide.

Organismo	Dimensione del genoma (Mb)
Procarioti	
Mycoplasma genitalium	0,58
Escherichia coli	4,64
Bacillus megaterium	30
Eucarioti	
Funghi	
Saccharomyces cerevisiae (lievito)	12,1
Aspergillus nidulans	25,4
Protozoi	
Tetrahymena pyriformis	190
Invertebrati	
Caenorhabditis elegans (nematode)	100
Drosophila melanogaster (moscerino della frutta)	140
Bombyx mori (baco da seta)	490
Strongylocentrotus purpuratus (riccio di mare)	845
Locusta migratoria (locusta)	5000
Vertebrati	
Fugu rubripes (pesce palla)	400
Homo sapiens (uomo)	3000
Mus musculus (topo)	3300
Piante	
Arabidopsis thaliana	100
Oryza sativa (riso)	565
Pisum sativum (pisello)	4800
Zea mays (mais)	5000
Triticum aestivum (grano)	17.000
Fritillaria assyriaca (fritillaria)	120.000

Tratto da Brown (1998).

Paradosso del valore C

La mancanza di correlazione tra la complessità genetica/morfologica di un organismo e le dimensioni del suo genoma è definita Paradosso del valore C.

Paradosso del valore C

Non esiste una stretta correlazione tra il numero di basi e l'informazione genetica in esse contenuta

DNA ripetitivo (satellite);

Geni discontinui;

Compattezza di alcuni genomi eucariotici

Proprietà del genoma	***S.cerevisiae***	***D.melanogaster***	***H. sapiens***
Densità genica (numero medio di geni per Mb)	479	79	11
Introni per gene (media)	0,04	3	9
% del genoma occupata dalle ripetizioni intersperse	3,4%	12%	44%

I Genomi degli Eucarioti:Compattezza

I genomi degli eucarioti hanno una densità genica molto ridotta.

In media, i geni codificanti per proteine occupano solo il 2-4% dell'intero genoma. La scarsa compattezza del genoma nucleare è dovuta alla struttura discontinua dei geni, con introni che nei mammiferi possono raggiungere dimensioni intorno a 20-30 kbp (ed oltre) e alla presenza di elementi ripetuti.

I geni eucariotici sono monocistronici, tuttavia strutture simili agli operoni batterici sono state descritte in *C. elegans.*

I Genomi degli Eucarioti:Compattezza

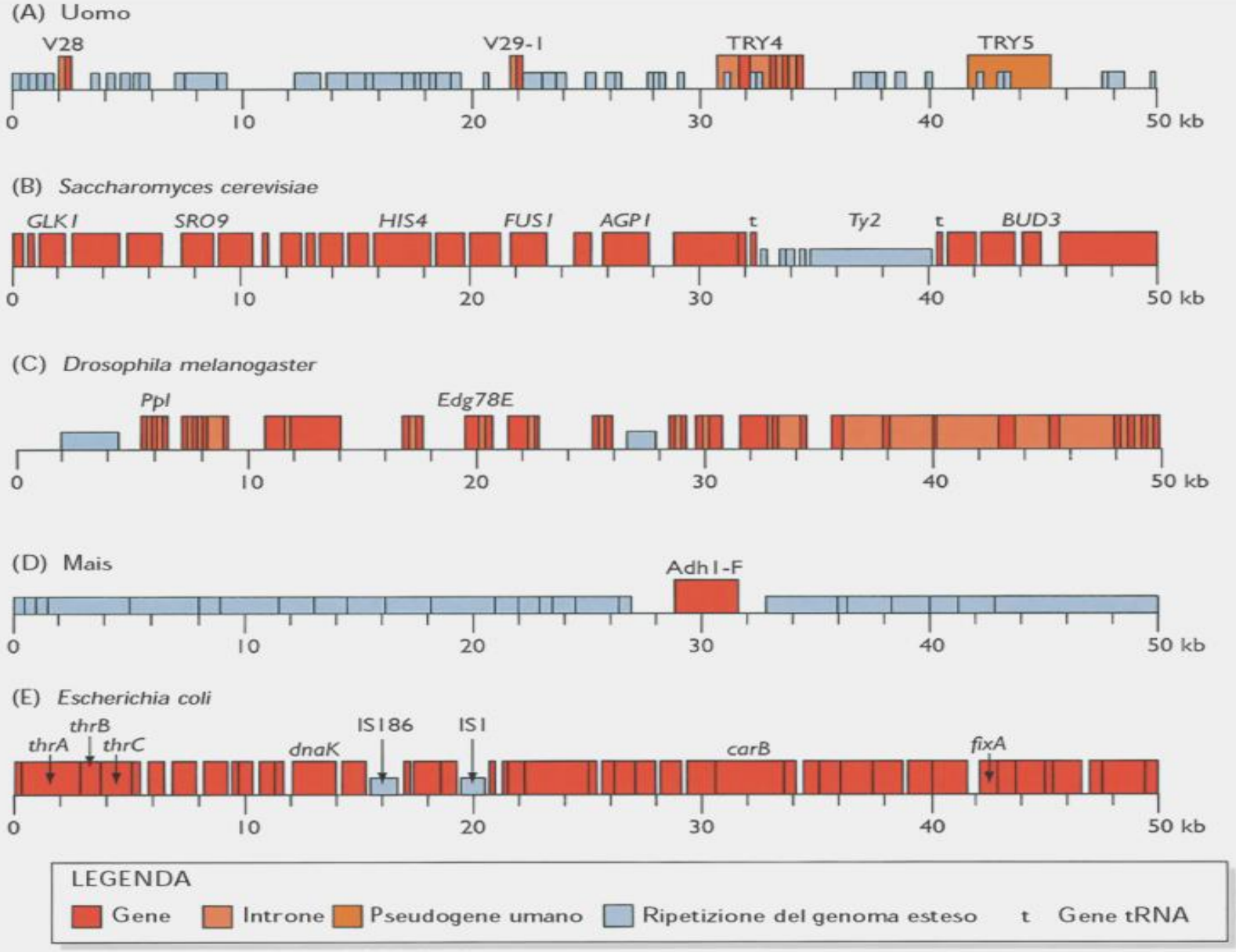
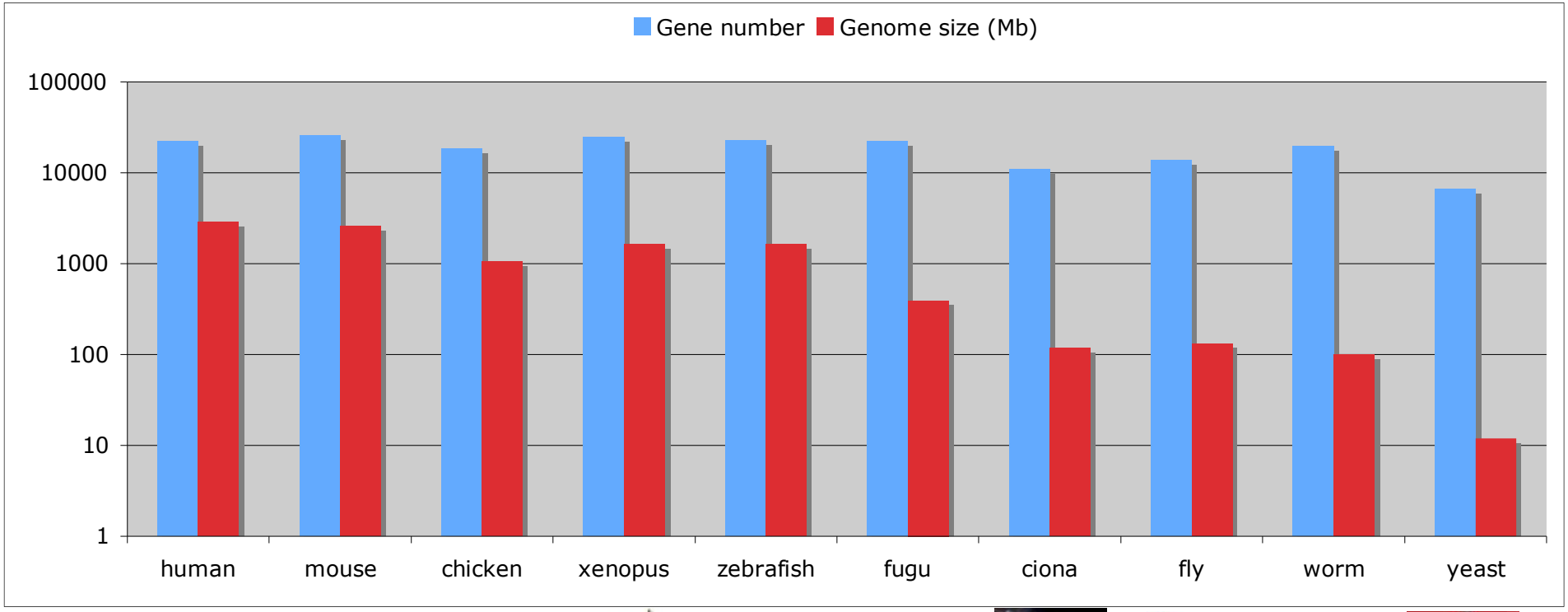


Figura 2.2 Confronto tra il genoma umano, di lievito, del moscerino della frutta, di mais e di *E. coli*.

Assenza di correlazione tra numero di geni e dimensione del genoma negli eucarioti

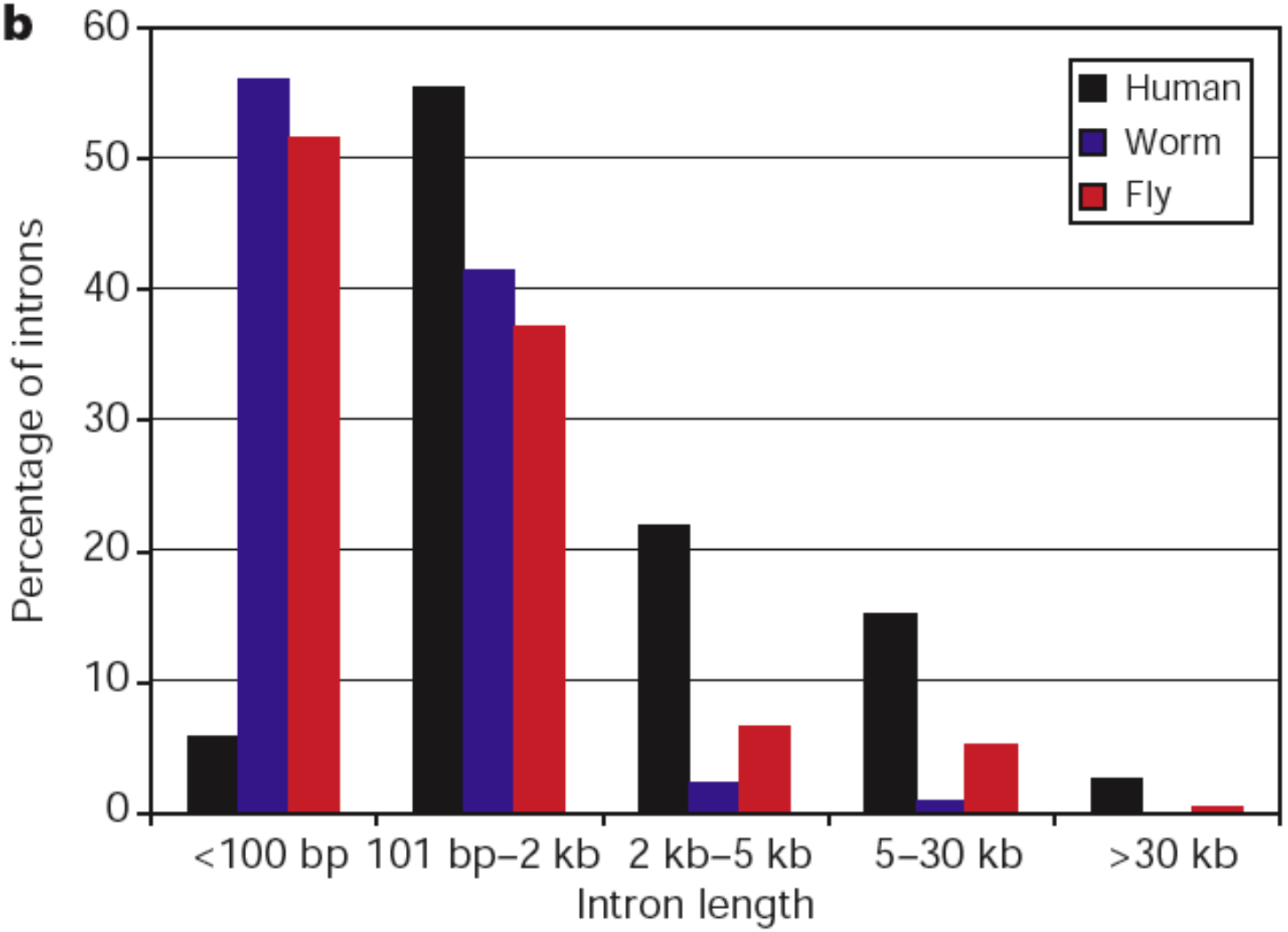


Number of genes in prokaryotes (up to 8000)

Genome size in prokaryotes (up to 9 Mb)

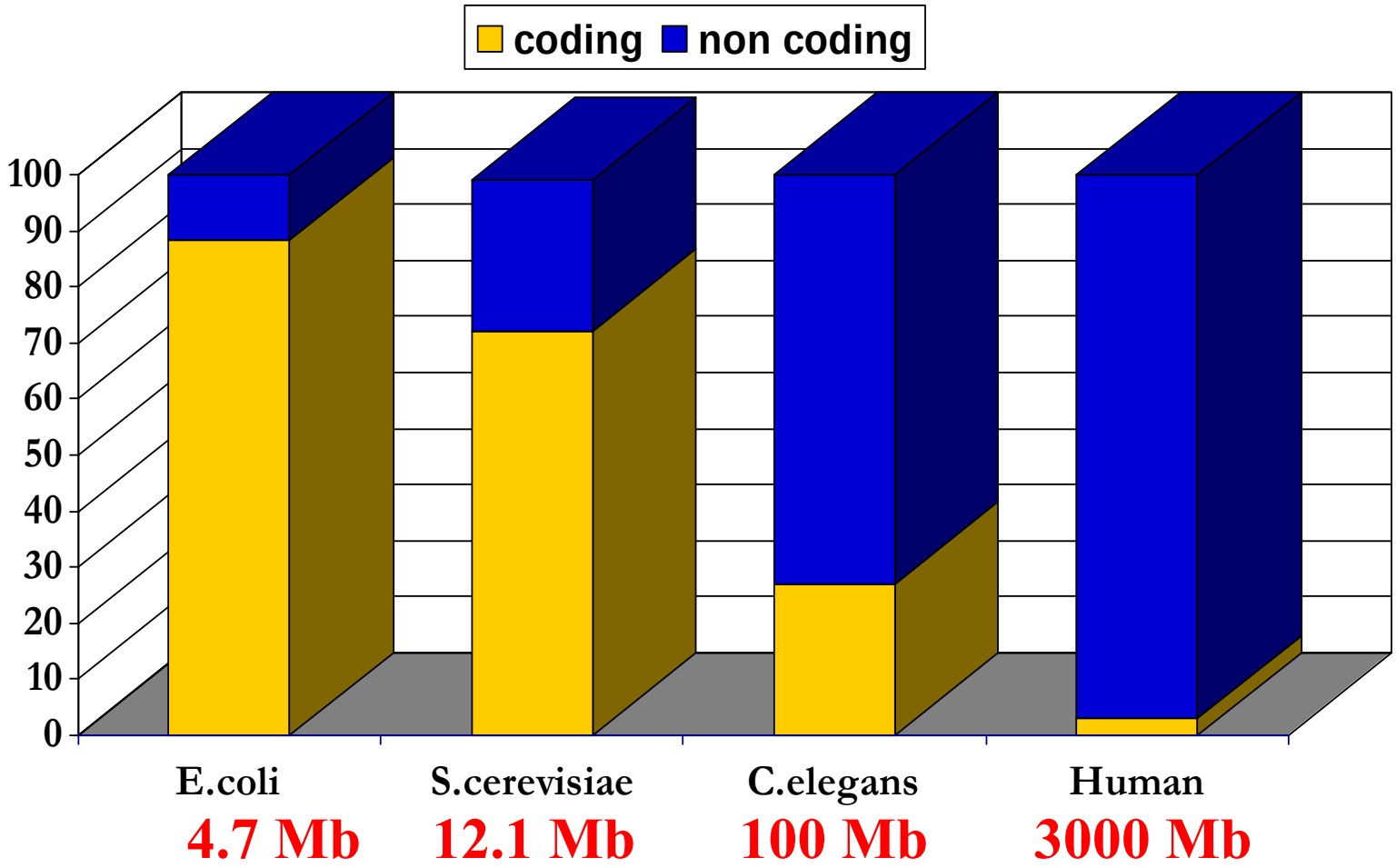
La struttura dei geni eucariotici: introni

IHGSC, Nature 2001 409:860-921, Tab. 35



I geni umani contengono introni mediamente più lunghi dei geni di *C.elegans* o *Drosophila*.

La porzione non codificante dei genomi eucariotici



L'annotazione funzionale delle porzioni non-codificanti del genoma è una delle sfide principali dell'era post-genomica.

International Human Genome Sequencing Consortium

Nature ottobre 2004

articles

Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium*

*A list of authors and their affiliations appears in the Supplementary Information

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human genome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

International Human Genome Sequencing Consortium

Nature ottobre 2004

- 3 miliardi di nucleotidi
- 34 milioni di nucleotidi rappresentano le sequenze codificanti proteine (1,2%)
- 20.000 - 25.000 geni codificanti proteine
- 35.000 trascritti codificanti proteine

The ENCODE Project:
ENCyclopedia Of DNA Elements

Researchers Expand Efforts to Explore Functional Landscape of the Human Genome

Bethesda, Md., Tues., Oct. 9, 2007 – The National Human Genome Research Institute (NHGRI), part of the National Institutes of Health (NIH), today announced grants totaling more than $80 million over the next four years to expand the ENCyclopedia Of DNA Elements (ENCODE) project, which in its pilot phase yielded provocative new insights into the organization and function of the human genome.

ENCODE 11 anni dopo

NEWS & VIEWS

FORUM: Genomics

ENCODE explained

The Encyclopedia of DNA Elements (ENCODE) 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. SEE ARTICLES P.57, P.75, P.83, P.91, P.101 & LETTER P.109

Serving up a genome feast

JOSEPH R. ECKER

Starting with a list of simple ingredients and blending them in the precise amounts needed to prepare a gourmet meal is a challenging task. In many respects, this task is analogous to the goal of the ENCODE project[1], the recent progress of which is described in this issue[2–7]. The project aims to fully describe the list of common ingredients (functional elements) that make up the human genome (Fig. 1). When mixed in the right proportions, these ingredients constitute the information needed to build all the types of cells, body organs and, ultimately, an entire person from a single genome.

The ENCODE pilot project[8] focused on just 1% of the genome — a mere appetizer — and its results hinted that the list of human genes was incomplete. Although there was scepticism about the feasibility of scaling up the project to the entire genome and to many hundreds of cell types, recent advances in low-cost, rapid DNA-sequencing technology radically changed that view[9]. Now the ENCODE consortium presents a menu of 1,640 genome-wide data sets prepared from 147 cell types, providing a six-course serving of papers in *Nature*, along with many companion publications in other journals.

One of the more remarkable findings described in the consortium's 'entrée' paper (page 57)[2] is that 80% of the genome contains elements linked to biochemical functions, dispatching the widely held view that the human genome is mostly 'junk DNA'. The authors report that the space between genes is filled with enhancers (regulatory DNA elements), promoters (the sites at which DNA's transcription into RNA is initiated) and numerous previously overlooked regions that encode RNA transcripts that are not translated into proteins but might have regulatory roles. Of note, these results show that many DNA variants previously correlated with certain diseases lie within or very near non-coding functional DNA elements, providing new leads for linking genetic variation and disease.

The five companion articles[3–7] dish up diverse sets of genome-wide data regarding the mapping of transcribed regions, DNA binding of regulatory proteins (transcription factors) and the structure and modifications of chromatin (the association of DNA and proteins that makes up chromosomes), among other delicacies.

Djebali and colleagues[3] (page 101) describe ultra-deep sequencing of RNAs prepared from many different cell lines and from specific compartments within the cells. They conclude that about 75% of the genome is transcribed at some point in some cells, and that genes are highly interlaced with overlapping transcripts that are synthesized from both DNA strands. These findings force a rethink of the definition of a gene and of the minimum unit of heredity.

> *"These findings force a rethink of the definition of a gene and of the minimum unit of heredity."*

Moving on to the second and third courses, Thurman *et al.*[4] and Neph *et al.*[5] (pages 75 and 83) have prepared two tasty chromatin-related treats. Both studies are based on the DNase I hypersensitivity assay, which detects genomic regions at which enzyme access to, and subsequent cleavage of, DNA is unobstructed by chromatin proteins. The authors identified cell-specific patterns of DNase I hypersensitive sites that show remarkable concordance with experimentally determined and computationally predicted binding sites of transcription factors. Moreover, they have doubled the number of known recognition sequences for DNA-binding proteins in the human genome, and have revealed a 50-base-pair 'footprint' that is present in thousands of promoters[5].

The next course, provided by Gerstein and colleagues[6] (page 91) examines the principles behind the wiring of transcription-factor networks. In addition to assigning relatively simple functions to genome elements (such as 'protein X binds to DNA element Y'), this study attempts to clarify the hierarchies of transcription factors and how the intertwined networks arise.

Beyond the linear organization of genes and transcription factors, chromosomes lies a more complex (and still poorly understood) network of chromosome loops and twists through which promoters and more distant elements, such as enhancers, can communicate their regulatory information to each other. In the final course of the ENCODE genome feast, Sanyal and colleagues[7] (page 109) map more than 1,000 of these long-range signals in each cell type. Their findings begin to overturn the long-held (and probably oversimplified) prediction that the regulation of a gene is dominated by its proximity to the closest regulatory elements.

One of the major future challenges for ENCODE (and similarly ambitious projects) will be to capture the dynamic aspects of gene regulation. Most assays provide a single snapshot of cellular regulatory events, whereas a time series capturing how such processes change is preferable. Additionally, the examination of large batches of cells — as required for the current assays — may present too simplified a view of the underlying regulatory complexity, because individual cells in a batch (despite being genetically identical) can sometimes behave in different ways. The development of new technologies aimed at the simultaneous capture of multiple data types, along with their regulatory dynamics in single cells, would help to tackle these issues.

A further challenge is identifying how the genomic ingredients are combined to assemble the gene networks and biochemical pathways that carry out complex functions, such as cell-to-cell communication, which enable organs and tissues to develop. An even greater challenge will be to use the rapidly growing body

ENCODE
Encyclopedia of DNA Elements
nature.com/encode

11 Years Ago

The draft human genome

OUR GENOME UNVEILED

Unless the human genome contains a lot of genes that are opaque to our computers, it is clear that we do not gain our undoubted complexity over worms and plants by using many more genes. Understanding what does give us our complexity — our enormous behavioural repertoire, ability to produce conscious action, remarkable physical coordination (shared with other vertebrates), precisely tuned alterations in response to external variations of the environment, learning, memory … need I go on? — remains a challenge for the future.

David Baltimore
From *Nature* 15 February 2001

GENOME SPEAK

With the draft in hand, researchers have a new tool for studying the regulatory regions and networks of genes. Comparisons with other genomes should reveal common regulatory elements, and the environments of genes shared with other species may offer insight into function and regulation beyond the level of individual genes. The draft is also a starting point for studies of the three-dimensional packing of the genome into a cell's nucleus. Such packing is likely to influence gene regulation … The human genome lies before us, ready for interpretation.

Peer Bork and Richard Copley
From *Nature* 15 February 2001

more than 2 million putative enhancers without known targets, revealing the enormous expanse of the regulatory genome landscape that is yet to be explored. Chromosome-conformation-capture methods that detect long-range physical associations between distant DNA regions are attempting to bridge this gap. Indeed, Sanyal and colleagues[7] applied these techniques to survey such associations across 1% of the genome.

The ENCODE data start to paint a picture of the logic and architecture of transcriptional networks, in which DNA binding of a few high-affinity transcription factors displaces nucleosomes and creates a DHS, which in turn facilitates the binding of further, lower-affinity factors. The results also support the idea that transcription-factor binding can block DNA methylation (a chemical modification of DNA that affects gene expression), rather than the other way around — which is highly relevant to the interpretation of disease-associated sites of altered DNA methylation[11].

The exquisite cell-type specificity of regulatory elements revealed by the ENCODE studies emphasizes the importance of having appropriate biological material on which to test hypotheses. The researchers have focused their efforts on a set of well-established cell lines, with selected assays extended to some freshly isolated cells. Challenges for the future include following the dynamic changes in the regulatory landscape during specific developmental pathways, and understanding chromatin structure in tissues containing heterogeneous cell populations.

Wendy A. Bickmore *is in the Medical Research Council Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.*
e-mail: wendy.bickmore@igmm.ed.ac.uk

Non-coding but functional

INÊS BARROSO

The vast majority of the human genome does not code for proteins and, until now, did not seem to contain defined gene-regulatory elements. Why evolution would maintain large amounts of 'useless' DNA had remained a mystery, and seemed wasteful. It turns out, however, that there are good reasons to keep this DNA. Results from the ENCODE project[2–8] show that most of these stretches of DNA harbour regions that bind proteins and RNA molecules, bringing these into positions from which they cooperate with each other to regulate the function and level of expression of protein-coding genes. In addition, it seems that widespread transcription from non-coding DNA potentially acts as a reservoir for the creation of new functional molecules, such as regulatory RNAs.

What are the implications of these results for genetic studies of complex human traits and disease? Genome-wide association studies (GWAS), which link variations in DNA sequence with specific traits and diseases, have in recent years become the workhorse of the field, and have identified thousands of DNA variants associated with hundreds of complex traits (such as height) and diseases (such as diabetes). But association is not causality, and identifying those variants that are causally linked to a given disease or trait, and understanding how they exert such influence, has been difficult. Furthermore, most of these associated variants lie in non-coding regions, so their functional effects have remained undefined.

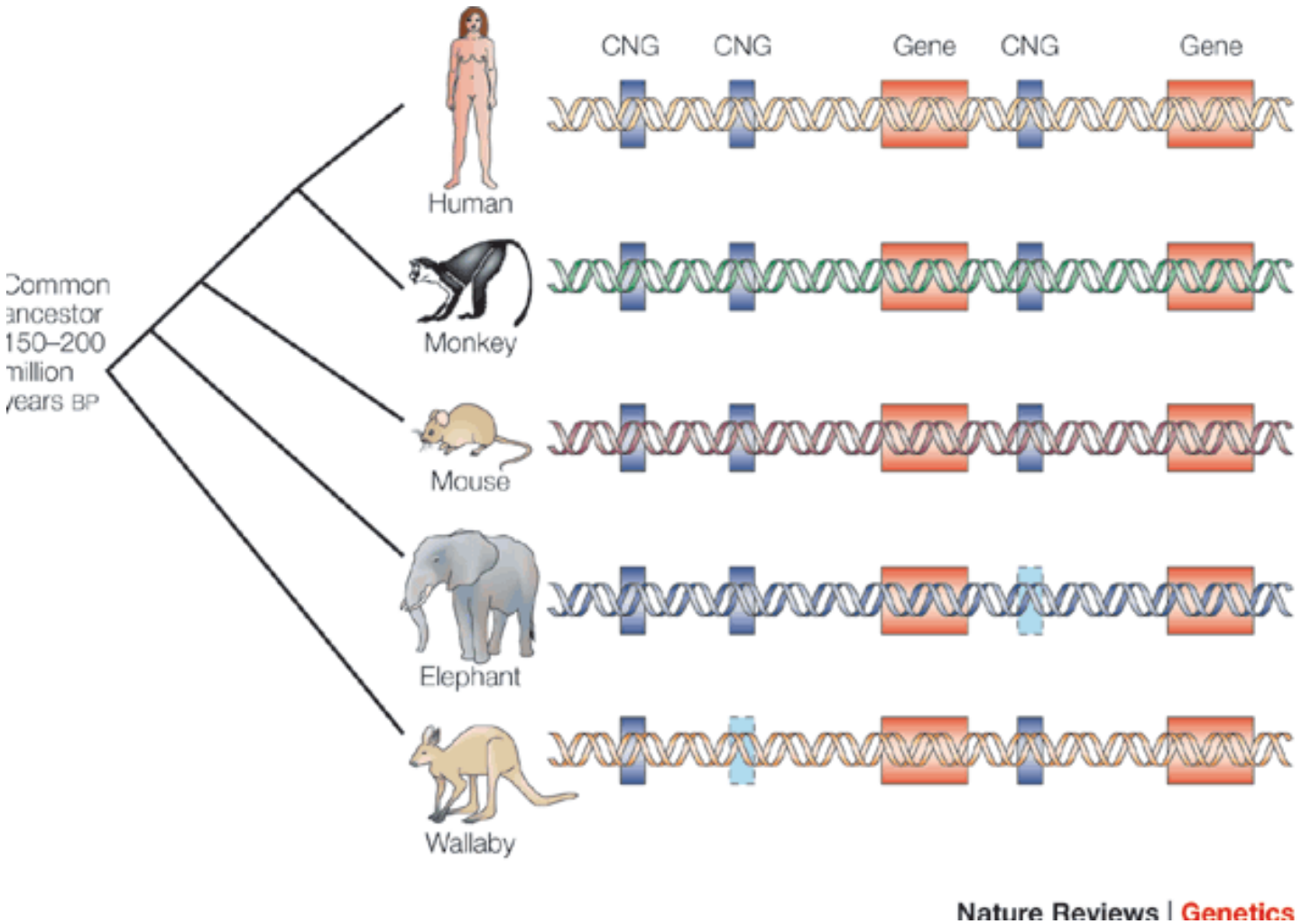
> *"The results imply that sequencing studies focusing on protein-coding sequences risk missing crucial parts of the genome."*

The ENCODE project provides a detailed map of additional functional non-coding units in the human genome, including some that have cell-type-specific activity. In fact, the catalogue contains many more functional non-coding regions than genes. These data show that results of GWAS are typically enriched for variants that lie within such non-coding functional units, sometimes in a cell-type-specific manner that is consistent with certain traits, suggesting that many of these regions could be causally linked to disease. Thus, the project demonstrates that non-coding regions must be considered when interpreting GWAS results, and it provides a strong motivation for reinterpreting previous GWAS findings. Furthermore, these results imply that sequencing studies focusing on protein-coding sequences (the 'exome') risk missing crucial parts of the genome and the ability to identify true causal variants.

However, although the ENCODE catalogues represent a remarkable tour de force, they contain only an initial exploration of the depths of our genome, because many more cell types must yet be investigated. Some of the remaining challenges for scientists searching for causal disease variants lie in: accessing data derived from cell types and tissues relevant to the disease under study; understanding how these functional units affect genes that may be distantly located[7]; and the ability to generalize such results to the entire organism.

Inês Barroso *is at the Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK, and at the University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Cambridge, UK.*
e-mail: ib1@sanger.ac.uk

Conserved non genic sequence



Emmanouil et al. (2005) Nature Reviews Genetics

Conserved non genic sequence

Table 1 | **Numbers and characteristics of exons, conserved exons and CNGs for each human chromosome**

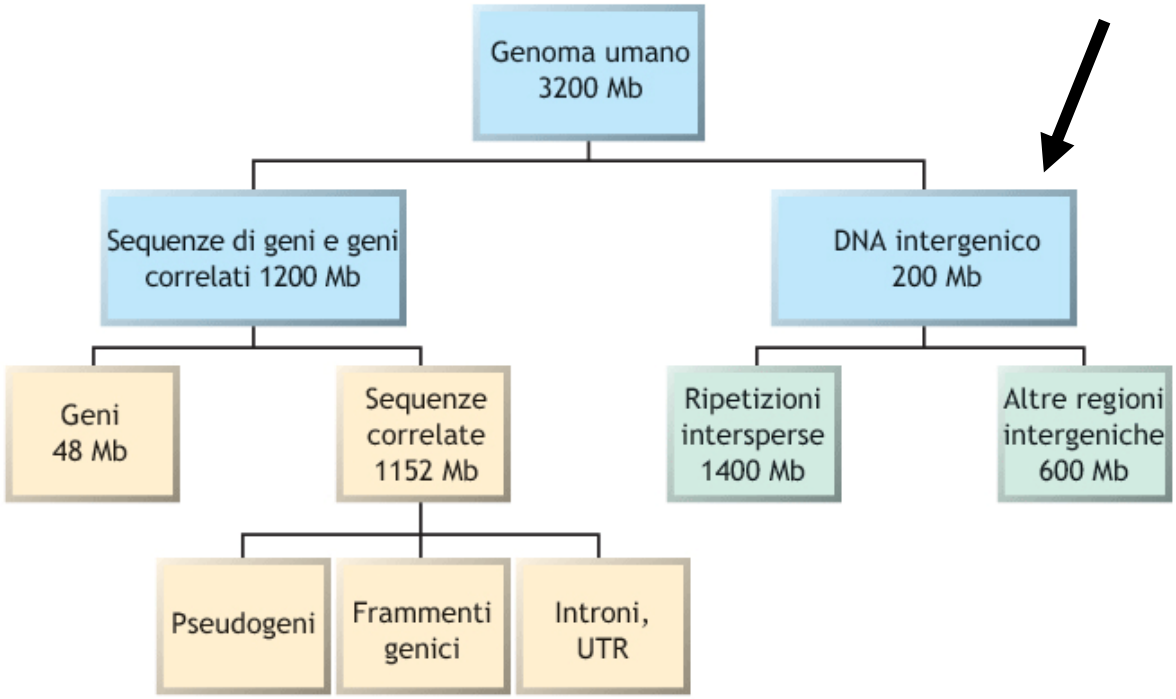
Chromosome	Length (in nucleotides)	Number of exons	Average exon length	Number of conserved exons	Average conserved exon length	Number of CNGs	Average CNG length
1	246,047,941	38,056	231.96	20,713	292.22	27,890	151.45
2	243,415,958	35,858	207.5	18,029	264.5	32,427	160.67
3	199,289,050	24,871	219.23	13,241	281.05	26,257	154.33
4	191,721,959	18,963	219.78	9,090	289.36	22,184	154.62
5	181,014,922	21,503	226.7	10,675	297.7	24,521	155.13
6	170,911,576	22,183	226.59	10,519	303.66	19,813	151.25
7	158,545,518	22,792	218.1	10,616	284.14	18,895	153.64
8	146,308,819	18,545	218.81	7,742	297.22	16,300	153.46
9	136,372,045	17,593	226.35	8,893	293.69	14,679	153.78
10	135,037,215	21,078	213.06	9,456	280.01	16,123	153.08
11	134,482,954	21,904	232.7	12,086	298.7	16,471	151.84
12	132,018,379	19,876	224.9	10,890	282.46	13,451	148.81
13	113,027,980	10,963	213.2	4,466	295.7	11,696	157.87
14	105,261,216	13,362	234.4	7,015	301.8	11,859	153.48
15	100,256,656	15,378	222.24	8,261	282.79	10,620	153.66
16	90,036,932	17,316	224.8	9,032	283	9,670	159.58
17	81,740,266	19,389	236.71	11,557	285.36	9,127	147.63
18	76,115,139	10,005	202.13	3,921	274.32	9,331	156.12
19	63,806,651	16,257	272.6	8,540	334.7	2,264	156.11
20	63,691,868	10,348	232.43	4,904	314.3	7,084	151.13
21	46,976,097	5,141	224.65	1,837	326.6	2,978	150.02
22	49,376,972	8,069	252.69	3,894	331.3	2,159	142.65
X	153,692,391	15,790	235.91	8,244	298.63	23,295	153.2
Y	50,286,555	2,610	218.9	602	284	993	220.5

CNG, conserved non-genic sequence.

Emmanouil et al. (2005) Nature Reviews Genetics

2000 Mb

FIGURA 4.66 Organizzazione del genoma umano: 3.200.000.000 di basi, solo una porzione composta da 48.000.000 di basi forma prodotti maturi.



De Leo - Fasano - Ginelli
Biologia e Genetica II Ed.
EdiSES

Il DNA RIPETUTO

RIPETUTO IN TANDEM

SATELLITE, tipico delle sequenze centromeriche (a-satellite, monomero di 171 bp)

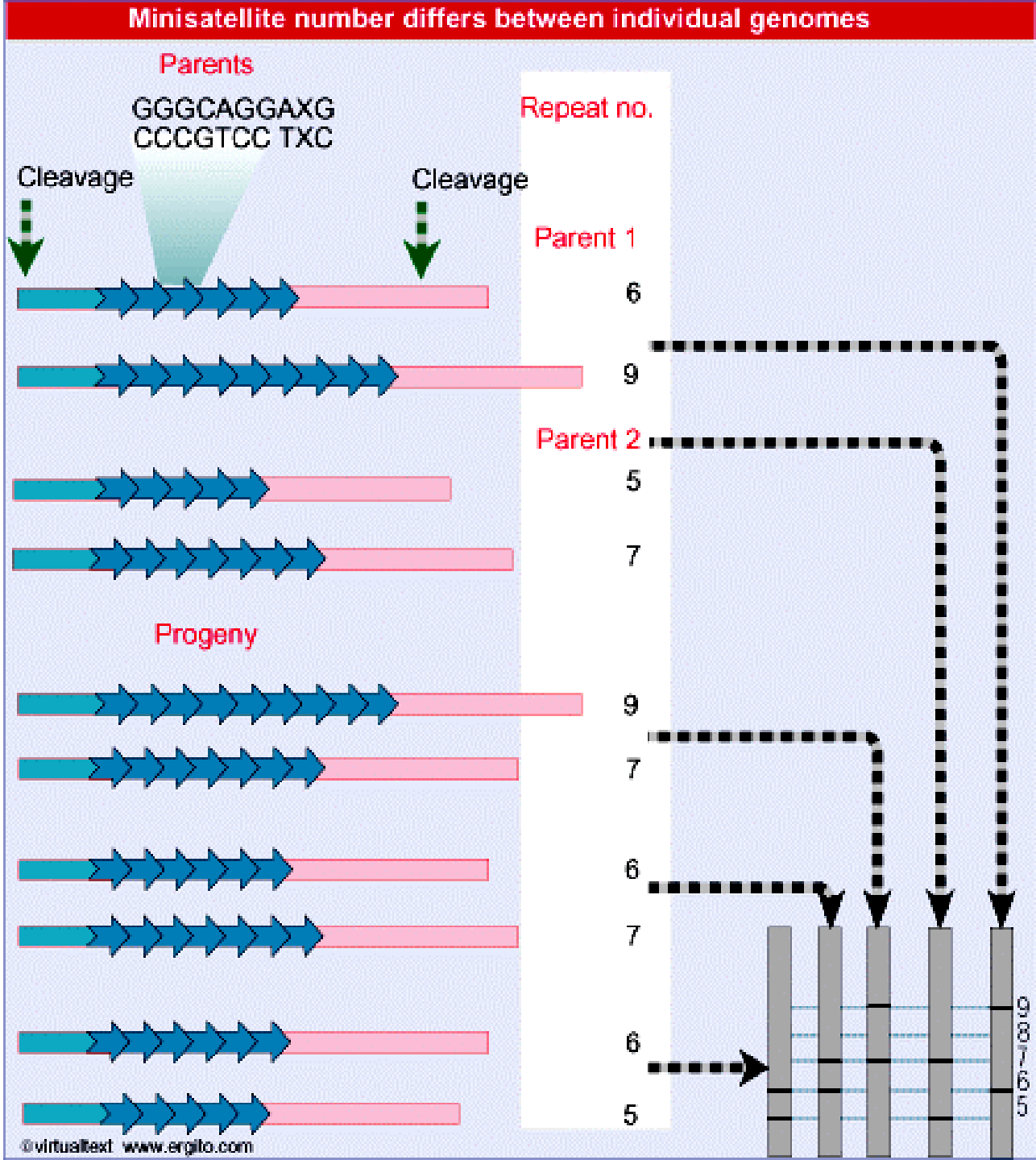
MINISATELLITE, monomero 6-64 bp, altamente polimorfico.
Utilizzato per esami di fingerprint del DNA.
Es.DNA telomerico (TTAGGG)

MICROSATELLITE, 2-4 bp ripetuti in tandem.
Espansioni di triplette sono responsabili di alcune patologie (Distrofia Miotonica)

Tabella I1.3-1 **Classi principali del DNA ripetuto in cluster**

Classe	Dimensioni dell'unità ripetuta (cb)	Principale localizzazione cromosomica
DNA satellite (spesso blocchi lunghi da 100 a diverse Mb)		
Satelliti 2 e 3	5	La maggior parte dei cromosomi (probabilmente tutti)
Satellite 1 (ricco in AT)	25-48	Eterocromatina centromerica della maggior parte dei cromosomi e altre regioni eterocromatiche
α (DNA alfoide)	171	Eterocromatina centromerica di tutti i cromosomi
β (famiglia *Sau*3A)	68	In particolare l'eterocromatina centromerica dei cromosomi 1, 9, 13, 14, 15, 21, 22 e Y
DNA minisatellite (spesso blocchi lunghi da 0,1 a 20 kb)		
Famiglie telomeriche	6	Tutti i telomeri
Famiglie ipervariabili	9-24	Tutti i cromosomi, spesso vicino ai telomeri
DNA microsatellite (spesso blocchi inferiori alle 150 cb)	1-4	Tutti i cromosomi

DNA fingerprinting



Il DNA RIPETUTO

INTERSPERSO

SINE, brevi elementi nucleari ripetuti (pseudogene processato di RNA7SL)
Alu (300bp, 1.000.000 copie nel genoma umano)

LINE, lunghi elementi nucleari ripetuti
L1 (6,1Kb a lunghezza completa, 200.000-500.000 copie)

Tabella I1.3-2 **Classi di DNA ripetuto intersperso**

Classe	Famiglia	Dimensioni unità ripetuta	N° copie	% genoma
SINE	*Alu*	0,3 kb lunghezza completa	1.300.000 ca	10,7% ca
	altre	dimensione media 0,13 kb	1.500.000 ca	12,5% ca
LINE	LINE-1 (*Kpn*)	6,1 kb lunghezza completa, ma le dimensioni medie sono 0,8 kb	1.900.000 ca	17,3%
	altre		1.370.000	13,3%
LTR	ERV	dimensione media 1,3 kb	1.240.000	14,7%
	altre		1.280.000	13,8%
Trasposoni a DNA	MER-1 (Charlie)	dimensione media 0,25 kb ca	1.213.000	11,5%
	altre		1.130.000	11,4%

De Leo - Fasano - Ginelli
Biologia e Genetica II Ed.
EdiSES

Porzione Trascritta del genoma eucariotico

- **geni codificanti per proteine, in copia singola**

- **geni codificanti per proteine, organizzati in famiglie geniche**

- **geni per rRNA, tRNA ed istoni, organizzati in unità ripetute in tandem**

- **geni per ncRNA**

(A)

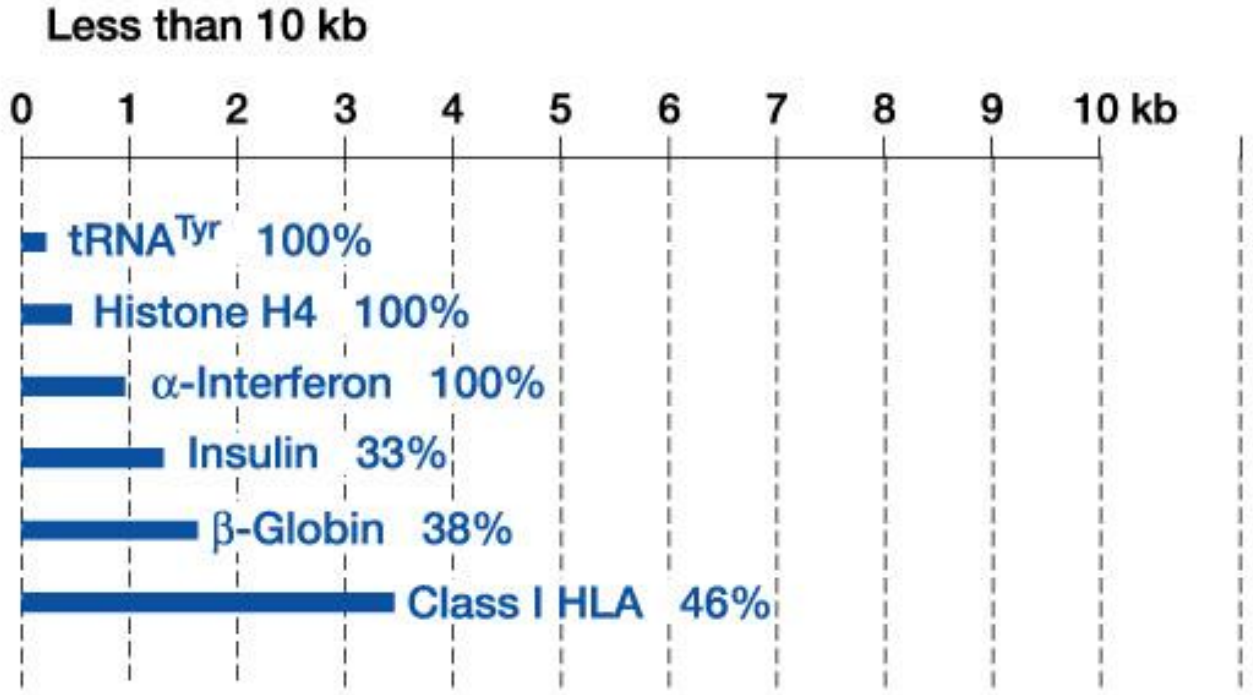


Figure 9-7 part 1 of 3 Human Molecular Genetics, 3/e. (© Garland Science 2004)

(B)

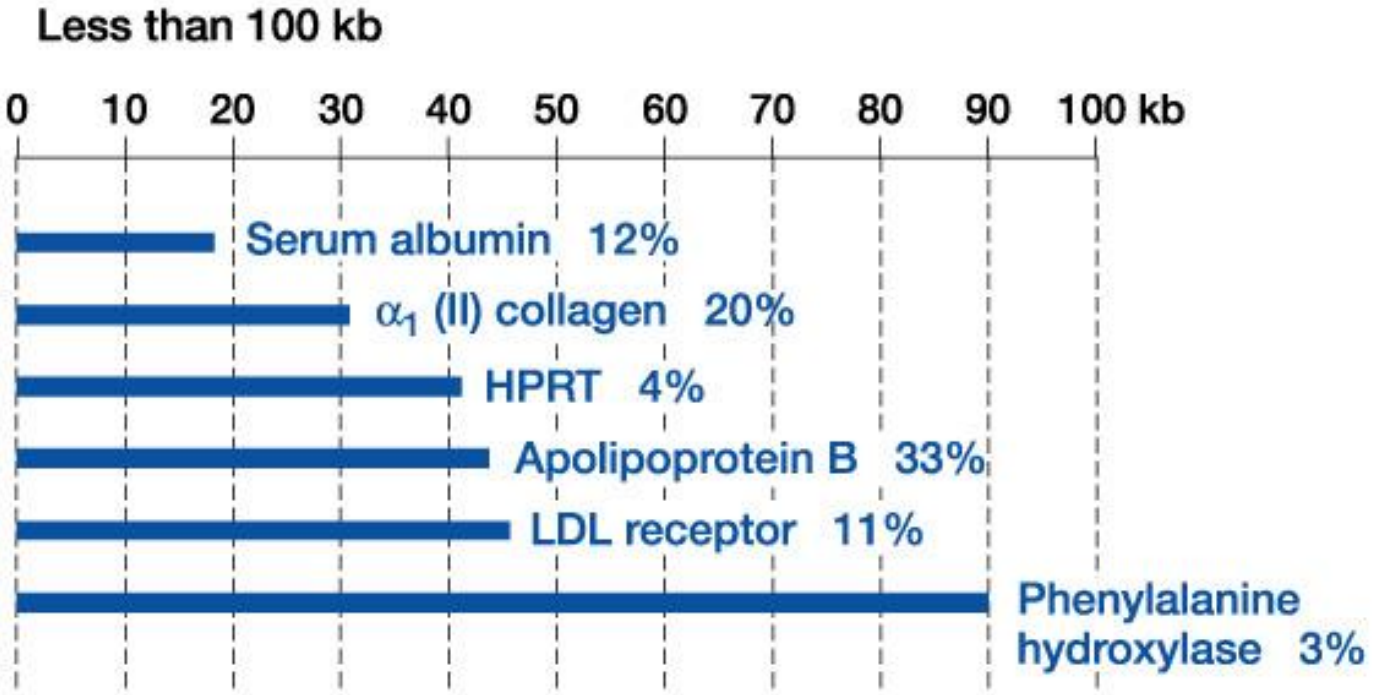


Figure 9-7 part 2 of 3 Human Molecular Genetics, 3/e. (© Garland Science 2004)

(C)

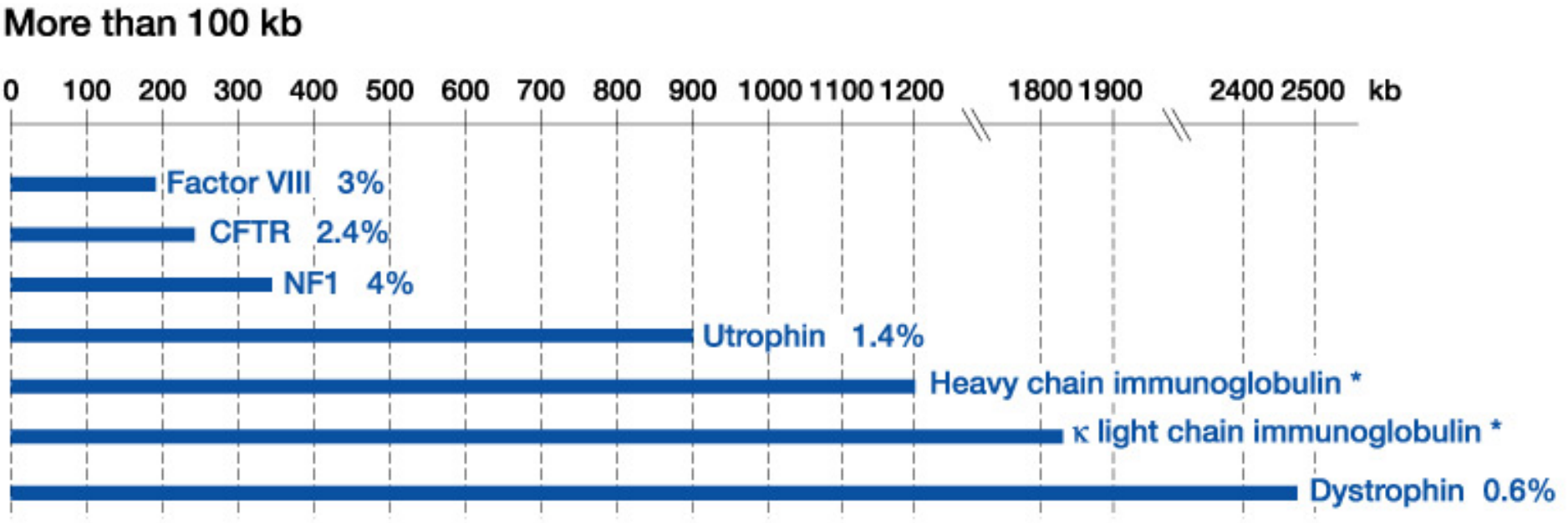
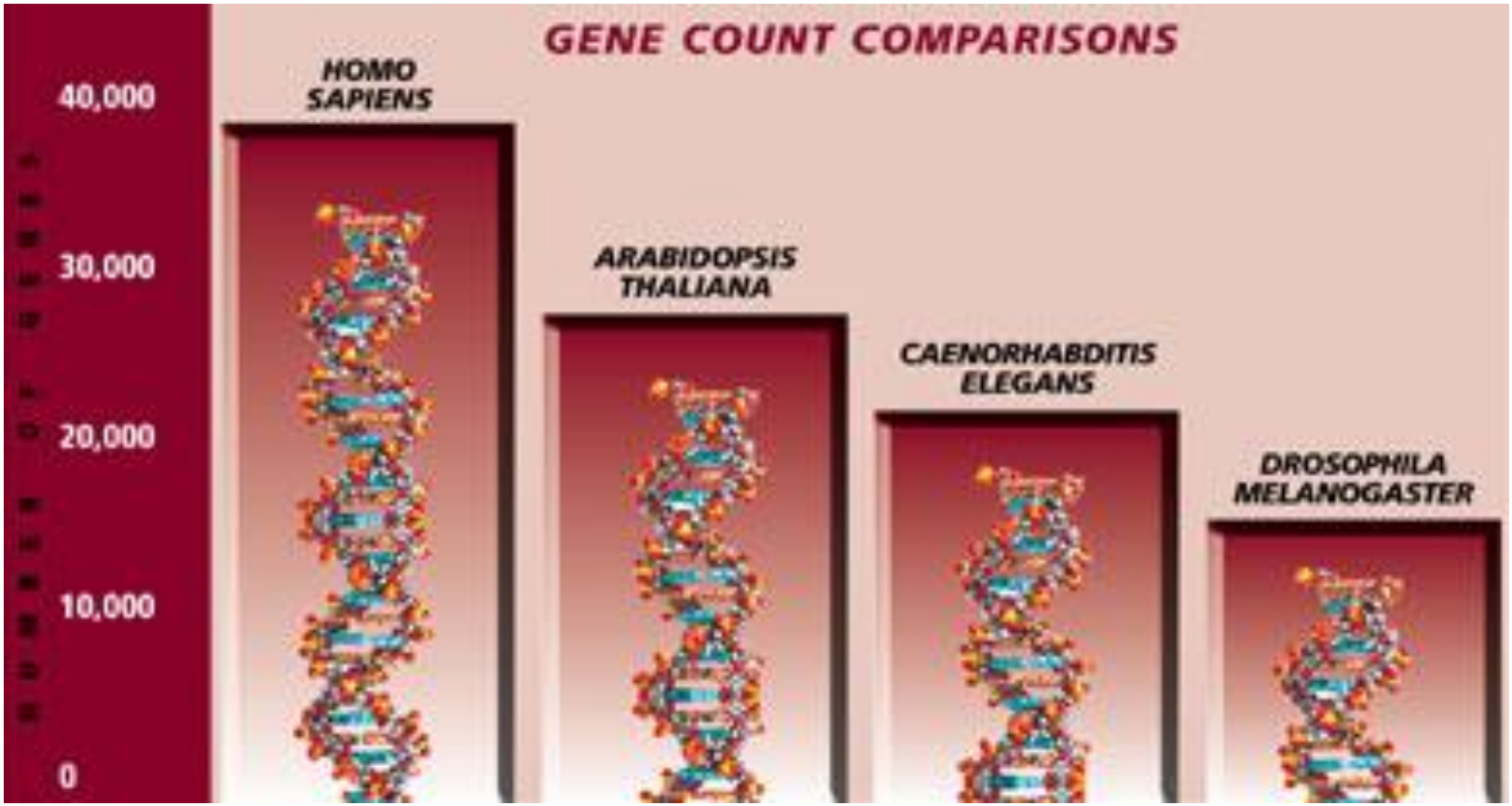


Figure 9-7 part 3 of 3 Human Molecular Genetics, 3/e. (© Garland Science 2004)



Precedente al 2002

La complessità di un organismo non correla necessariamente con il numero di geni

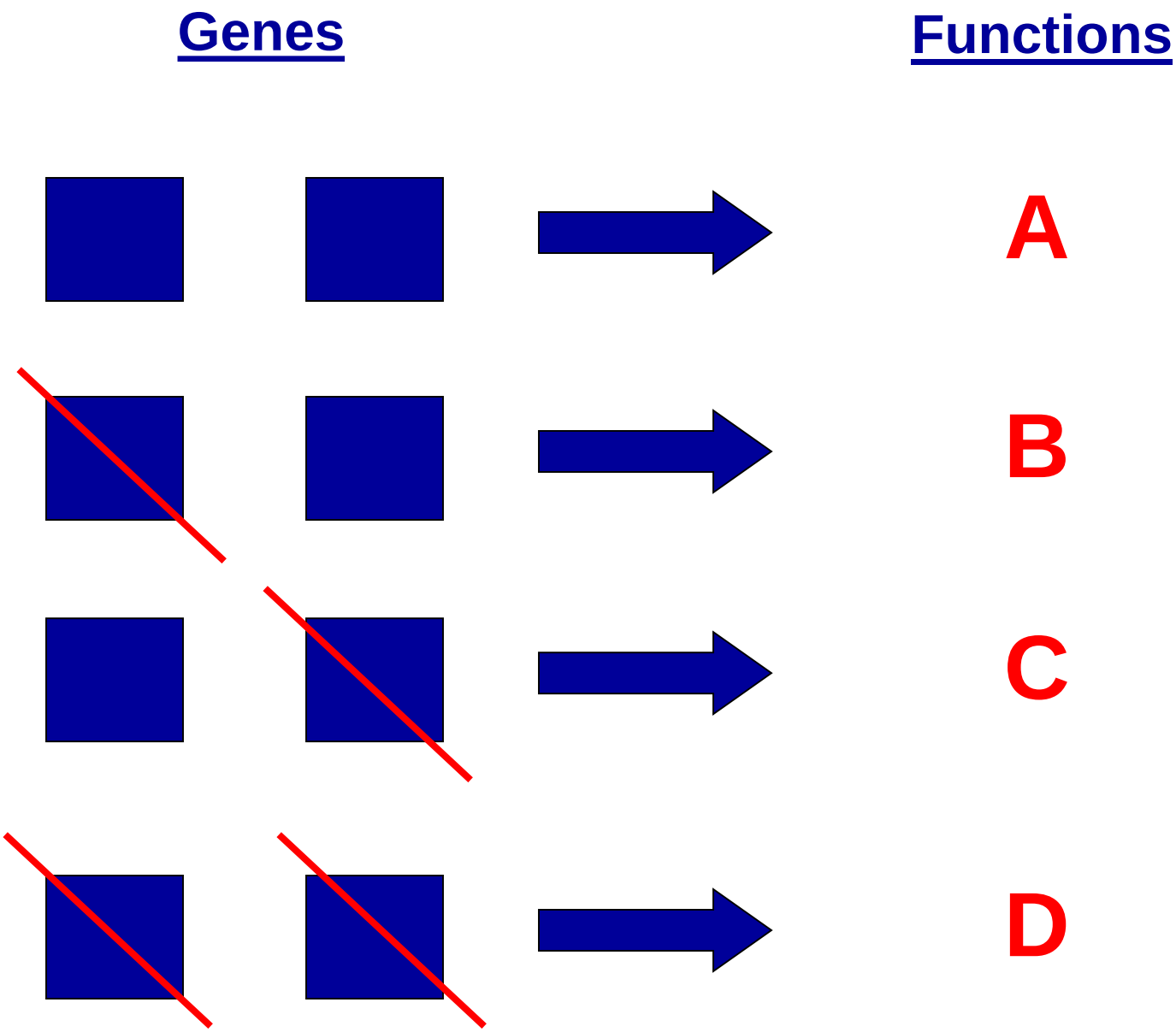
Come misurare la Complessità biologica ?

La complessità biologica può essere "misurata" in diversi modi, ad es. sulla base della diversità di tipi cellulari, della complessità dei circuiti del cervello,……o del n° teorico di stati dell'espressione genica.

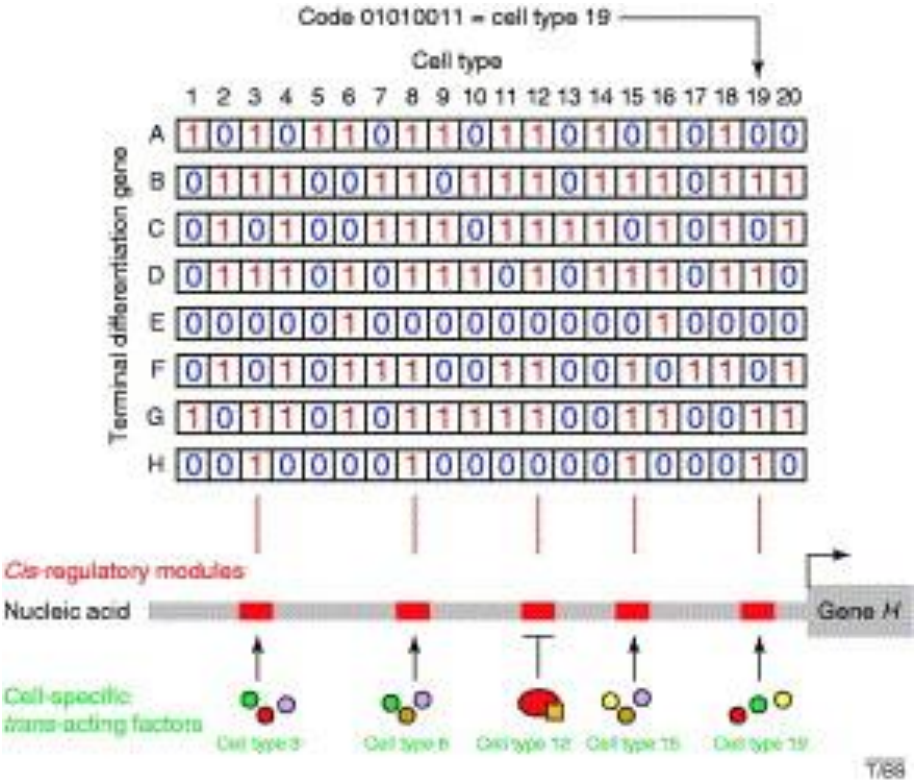
Ipotizzando N geni umani e supponendo che ciascuno possa essere presente in due soli stati, ON o OFF, il numero di possibili stati sarebbe pari a 2^N.

25,000 geni nel genoma umano

Complessità = $2^{25,000}$


Genes
Functions
A
B
C
D

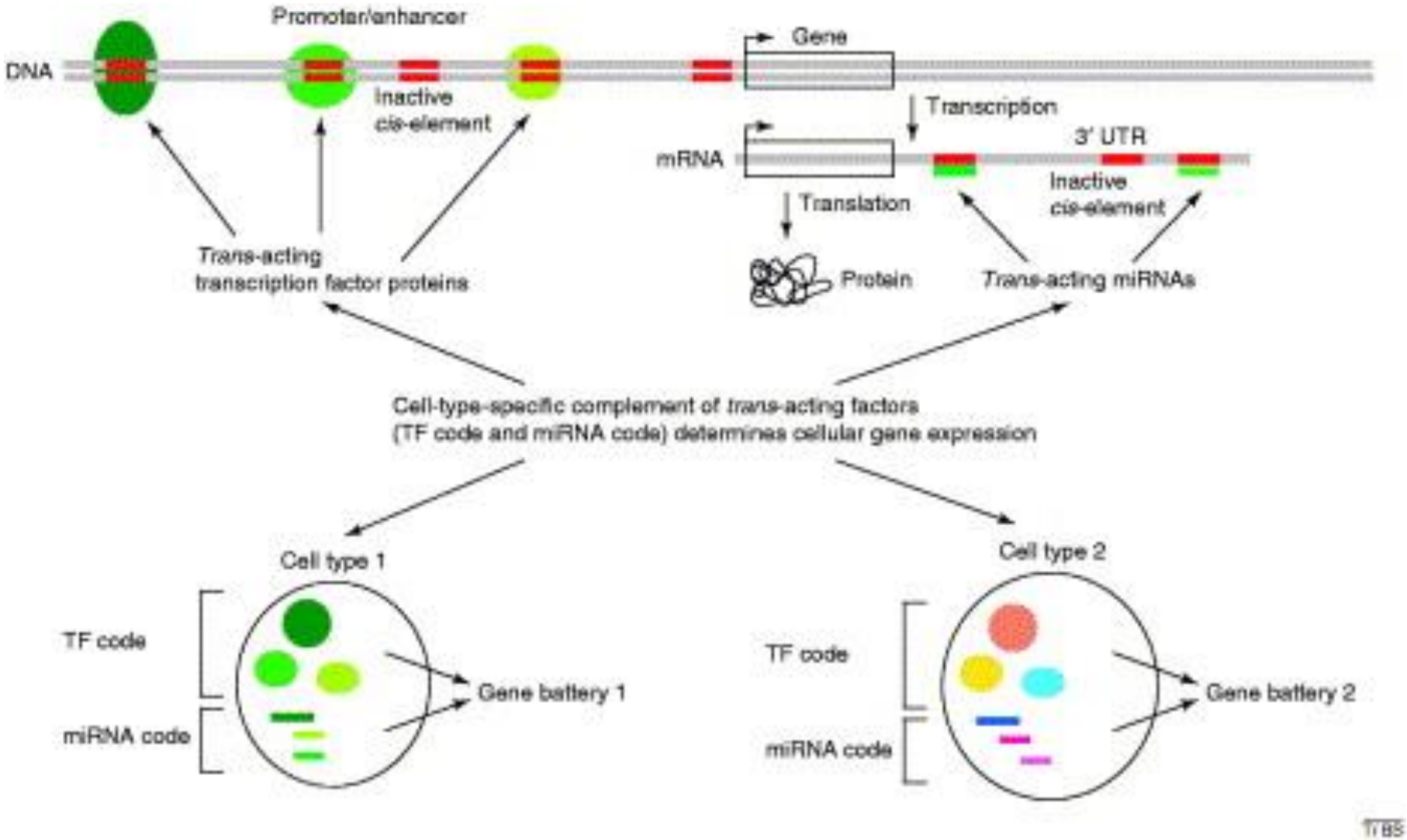
Create cellular complexity by differential genes expression



This form of combinatorial coding endows an organism with *n* genes to create, in theory, 2^n different cell-specific gene batteries.

This rationalization provides an easy explanation for the fact that the absolute number of genes in a genome does not correlate with organismal complexity

Hobert (2004) TIBS



Hobert (2004) TIBS

Evoluzione del genoma umano

Duplicazione genica

Rimescolamento degli esoni

Famiglie geniche

Le famiglie geniche possono essere generate attraverso diversi meccanismi:

- **poliploidizzazione del genoma**
- **duplicazione di segmenti genomici (famiglia dei geni omeotici)**
- **duplicazione di un singolo gene (geni per α e β globine)**
- **retrotrascrizione**

Duplicazione genica

Produzione di due copie identiche di un gene

Delle due copie, una continua a svolgere la propria funzione, l'altra può andare incontro a diversi destini

Il gene duplicato mantiene la stessa funzione del gene ancestrale (istoni)

Gene redundancy

Il gene duplicato, non essendo sottoposto alla stessa pressione selettiva del gene ancestrale, può accumulare mutazioni casuali

1. **L'accumulo di mutazioni porta all'inattivazione del gene duplicato, trasformandolo in pseudogene (pseudogeni delle α e β globine)**

2. **L'accumulo di mutazioni fa sì che il gene duplicato possa acquisire una nuova funzione utile per l'organismo**

(le nuove funzioni acquisite possono diventare specie-specifiche)

Duplicazioni intra-geniche

Nuove funzioni geniche possono essere acquisite mediante riarrangiamento di segmenti genici codificanti per domini proteici strutturali

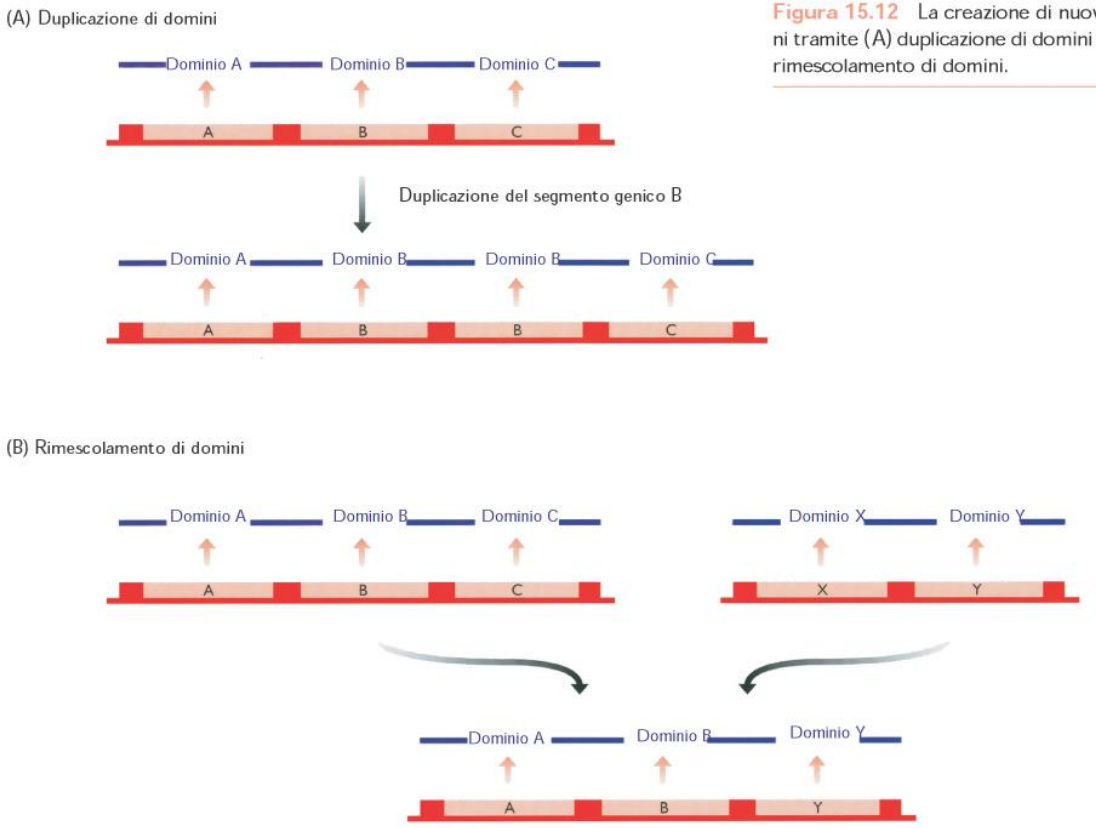


Figura 15.12 La creazione di nuovi geni tramite (A) duplicazione di domini e (B) rimescolamento di domini.

2 meccanismi:

- duplicazione dei domini

- rimescolamento dei domini

Porzione Trascritta del genoma eucariotico

- geni codificanti per proteine, in copia singola

- geni codificanti per proteine, organizzati in famiglie geniche

- geni per rRNA, tRNA ed istoni, organizzati in unità ripetute in tandem

Ridondanza genetica del genoma nucleare

- geni per ncRNA

I geni organizzati in famiglie geniche sono tra loro omologhi, e derivano da un evento di duplicazione genica o di retrotrasposizione mediata da RNA.

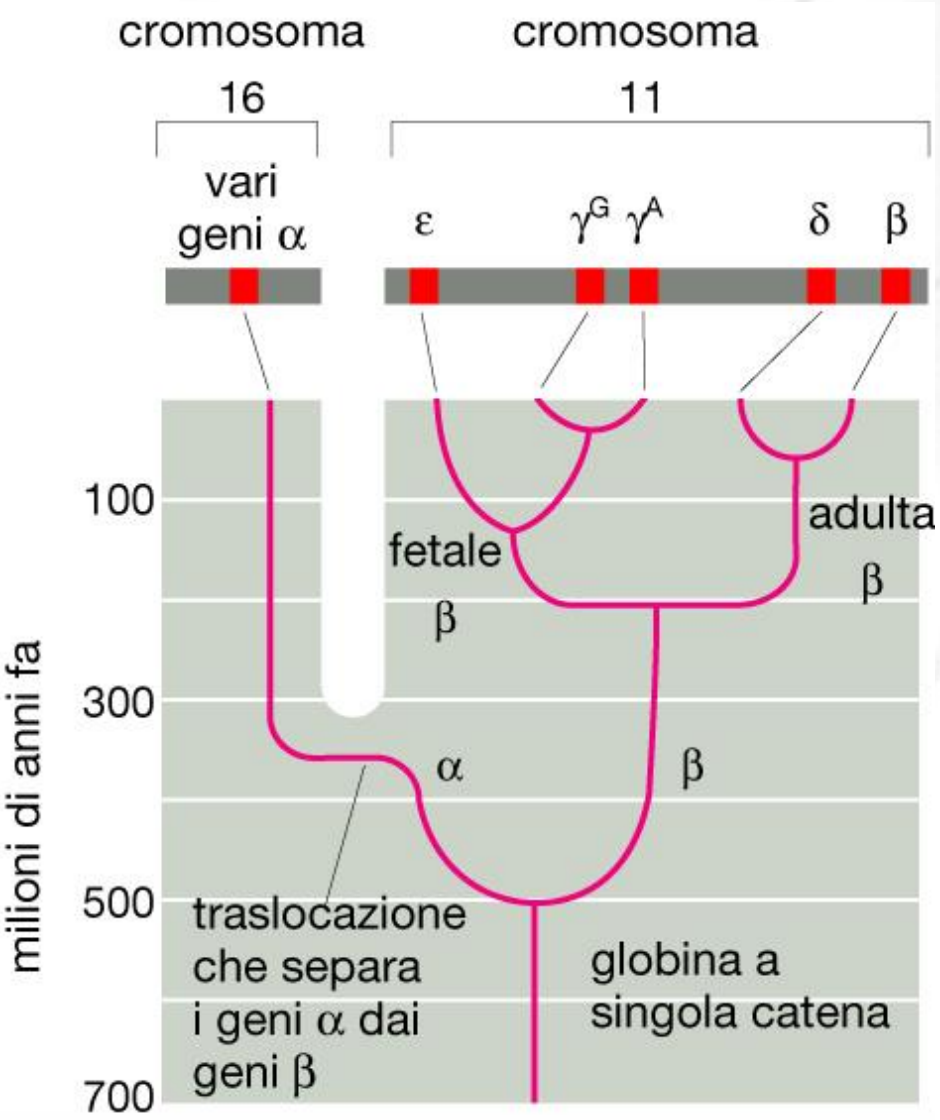
I membri di una famiglia genica all'interno di uno stesso genoma sono detti paraloghi, e normalmente si specializzano acquisendo funzioni distinte.

Hb-A (2α - 2β)

HbA-2 (2α – 2δ)

HbF (2α - 2γ)

Hb embrionale (2ζ – 2ε)



Evoluzione genica per duplicazione: i geni delle globine

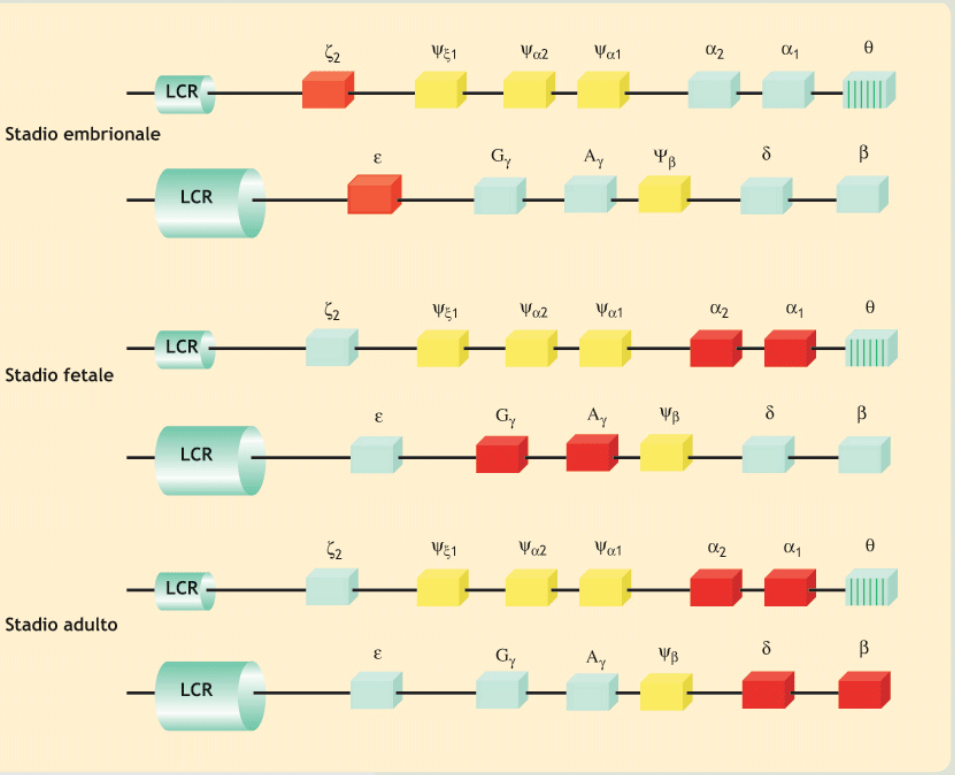


FIGURA I1.3-5 **Organizzazione genomica dei geni alfa e beta globinici nei rispettivi raggruppamenti sui cromosomi 16 e 11.**

De Leo - Fasano - Ginelli
Biologia e Genetica II Ed.
EdiSES

GENOMICA dei microRNA

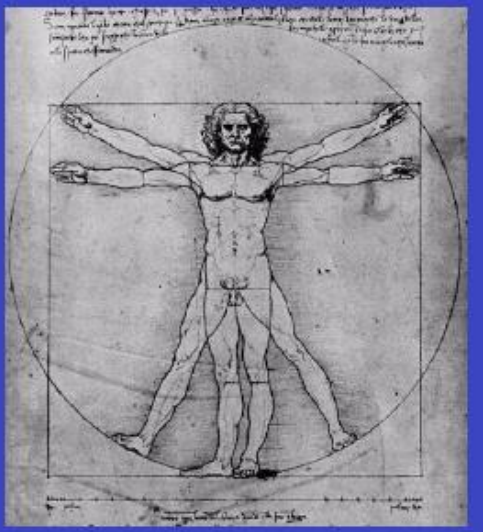
- **Le sequenze nucleotidiche dei miRNA si sono conservate nel corso dell'evoluzione;**

- **I loro geni possono essere localizzati sia in regioni intergeniche che all'interno di geni di seconda classe (di solito in introni o in esoni non tradotti);**

- **Molte volte troviamo cluster di geni.**

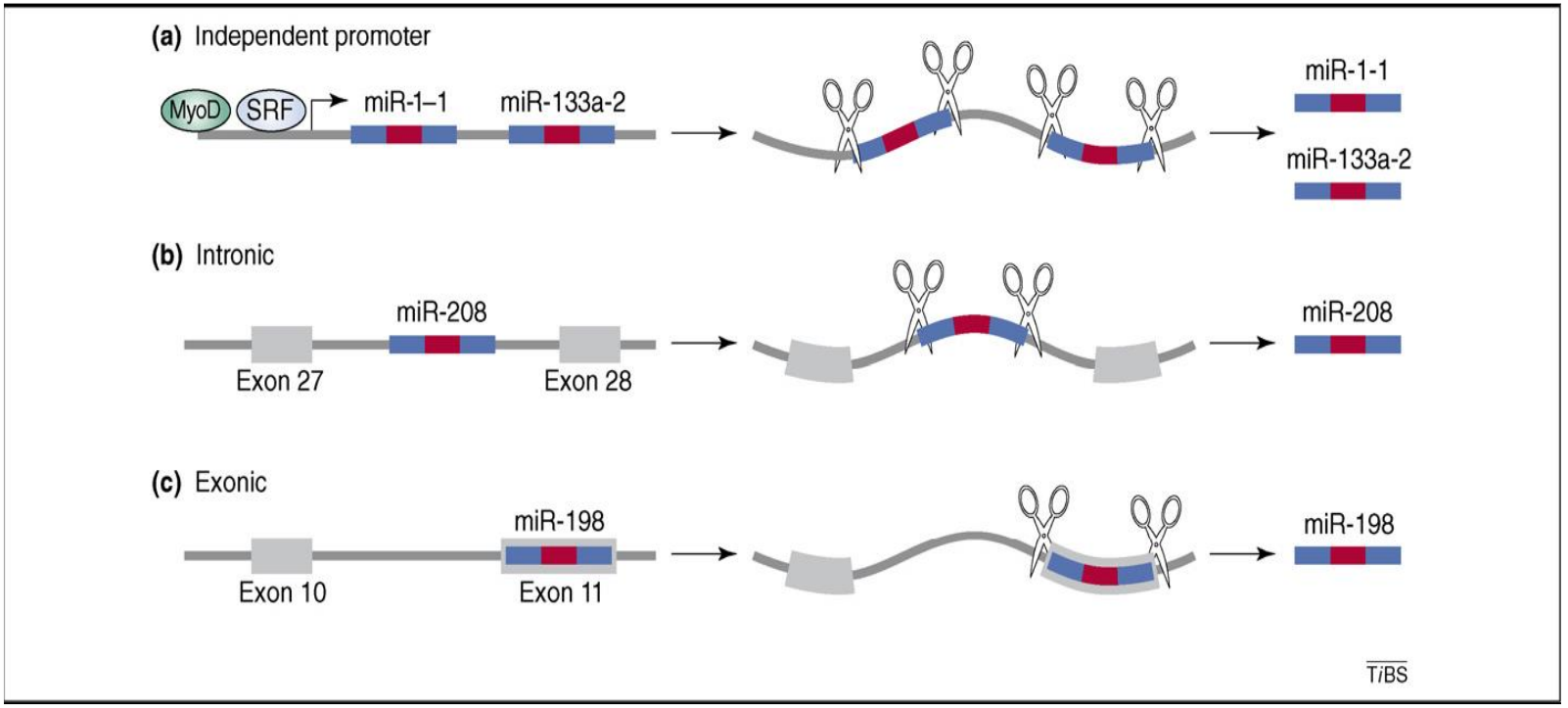
Evolutionary Conservation of microRNA sequences

Lim et al. compared microRNA sequences from *C. elegans* to the human genome, and found that over 1/3 of these genes have homologs in humans.



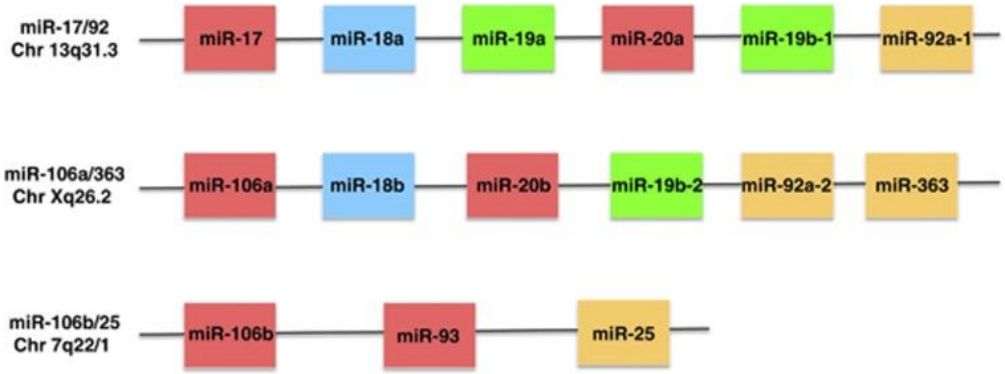
Hundreds of conserved microRNAs

LOCALIZZAZIONE GENOMICA



Zhao Y and Srivastava D, 2007

ORGANIZZAZIONE GENOMICA



miR-17 family	**hsa-miR-17-5p**	C**AAAGUG**CUUACAGUGCAGGUAGU
	hsa-miR-20a-5p	U**AAAGUG**CUUAUAGUGCAGGUAG
	hsa-miR-20b-5p	C**AAAGUG**CUCAUAGUGCAGGUA
	hsa-miR-106a-5p	C**AAAGUG**CUAACAGUGCAGGUA
	hsa-miR-106b-5p	U**AAAGUG**CUGACAGUGCAGAU
	hsa-miR-93-5p	C**AAAGUG**CUGUUCGUGCAGGUAG
miR-18 family	**hsa-miR-18a-5p**	U**AAGGUG**CAUCUAGUGCAGAUA
	hsa-miR-18b-5p	U**AAGGUG**CAUCUAGUGCAGUUA
miR-19 family	**hsa-miR-19a-3p**	U**GUGCAA**AUCUAUGCAAAACUGA
	hsa-miR-19b-3p	U**GUGCAA**AUCCAUGCAAAACUGA
	hsa-miR-19b-3p	U**GUGCAA**AUCCAUGCAAAACUGA
miR-92 family	**hsa-miR-92a-3p**	U**AUUGCA**CUUGUCCCGGCCUG
	hsa-miR-92a-3p	U**AUUGCA**CUUGUCCCGGCCUG
	hsa-miR-25-3p	C**AUUGCA**CUUGUCUCGGUCUGA
	hsa-miR-363-3p	A**AUUGCA**CGGUAUCCAUCUGUAA

OPEN

Cell Death and Differentiation (2013) 20, 1603–1614

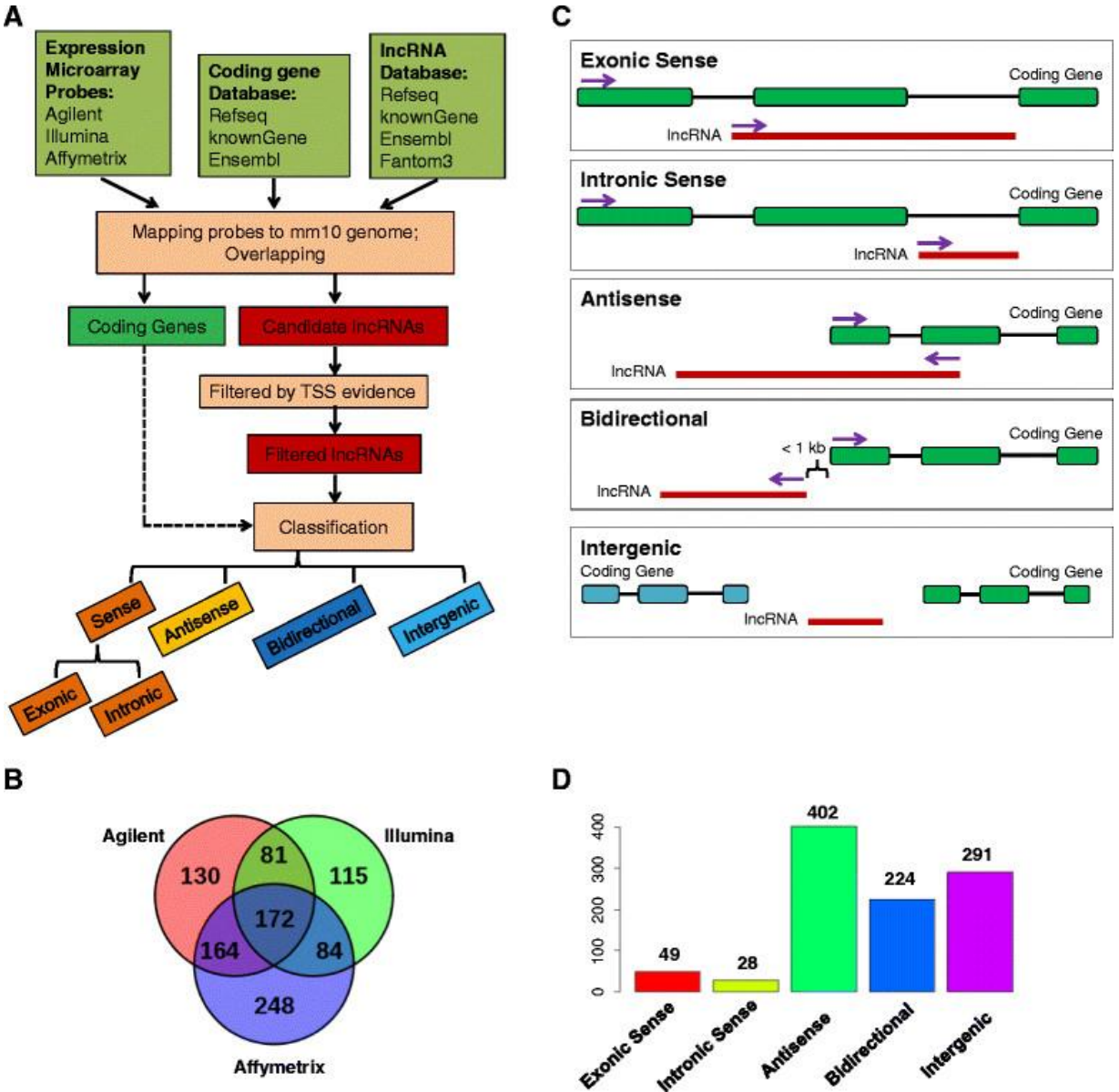
www.nature.com/cdd

Review

The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease

E Mogilyansky[1] and I Rigoutsos*[,1]

Organizzazione Genomica dei lncRNAs



http://www.hugo-international.org/index.php

Human Genome Organization