

Biologia Cellulare Molecolare Avanzata

“Functional & regulatory genomics”

1. Complexity of eukaryotic genomes.
2. Basic concepts of gene transcription and regulation
3. Transcriptomes
4. Coding, noncoding and alternative splicing

5. Chromatin as the dynamic environment of genomes.
6. Functional states of chromatin and chromosome territories
7. Epigenetics and gene imprinting
8. Mechanisms and pathways of gene control
9. Evolution of alternative splicing
10. RNA elements that regulate RNA fate and life
11. Other noncoding RNAs and primordial RNA functions

<http://biologia.i-learn.unito.it/>

<http://www.personalweb.unito.it/michele.debortoli>

Michele De Bortoli
Professor of Molecular Biology
University of Turin

Ogni lezione contiene:

1. ripasso concetti di base (dai corsi del triennio)
2. definizioni di concetti e spiegazioni della metodologia e sperimentazione che ha permesso di chiarirli (slides)
3. studio di uno (o due) lavori scientifici sull'argomento

Il materiale didattico disponibile è costituito da:

1. le slides proiettate a lezione
2. una (o due) reviews che illustrano l'argomento
3. il (o i) lavoro (i) studiato a lezione

Gli studenti devono:

1. conoscere i concetti di **base**
2. conoscere i concetti **specifici** trattati
3. avere compreso metodologie ed approcci **sperimentali**

Prove di profitto:

- 1) discussione di un **articolo** scientifico recente su uno degli argomenti trattati (durante il corso, a coppie)
- 2) esame **orale** con discussione sull'intero programma

Alla pagina del corso:

<http://biologia.i-learn.unito.it/>

Troverete:

pdf delle lezioni
pdf degli articoli utilizzati per le lezioni
pdf di rassegne utili per lo studio

Ogni lezione ed il materiale relativo sarà contenuta in una cartella compressa (.zip).

Se il vostro OS non decomprime, potete scaricare da web gratuito il software "7-zip", ottimo.

Altre indicazioni utili

These background are needed:

1. - Basic Molecular Biology & Genetics

DNA replication

Transcription

Post-transcriptional RNA processing

Translation

Post-translational protein modification

Gene expression regulation (basic mechanisms)

Basics of protein structure and molecular representations

Example: Chapters 1 through 10 from “Essential Cell Biology” 2°
or 3° Edition - Alberts et al., Garland, 2004 (2°), 2009 (3°)
Italian version - Zanichelli (2005)

Basic recombinant DNA methodology:

1. DNA replication (in vivo and in vitro)
2. PCR, rt-PCR and real-time PCR
3. Basic DNA cloning in plasmids and other vectors
4. Libraries, clones, colonies, storage, propagation, analysis.
5. DNA sequencing, restriction, Southern blot
6. RNA analysis, Northern blot, Rnase protection

Basic bioinformatics:

1. Database organization
2. Finding gene and protein sequences
3. Basic alignment protocols

Durante la lezione, compariranno dei piccoli spot come questo:

Sex determination in *D. melanogaster*

Questi rappresentano il lavoro di **revisione di base** che dovete fare voi

10-100 milioni di specie viventi sulla Terra

Stupefacente varietà nei singoli particolari

Stupefacente costanza nei meccanismi fondamentali

Quanti genomi sono stati sequenziati o sono in corso di sequenziamento ?

Complete updates for genomic projects (comparative):

<http://www.ncbi.nlm.nih.gov/Genomes/>

Animated tutorials on the Human Genome Project:

<http://www.genome.gov/Pages/EducationKit/>

(free downloads or on-line view)

How to sequence a genome

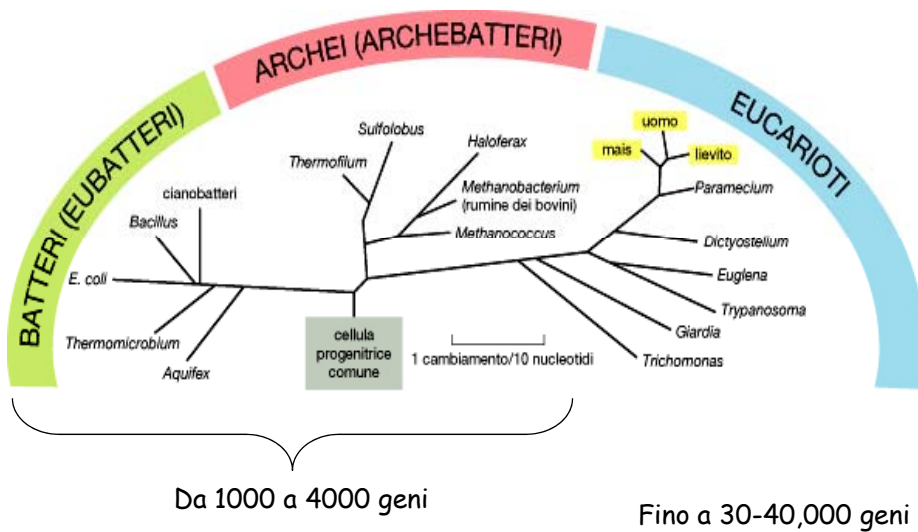
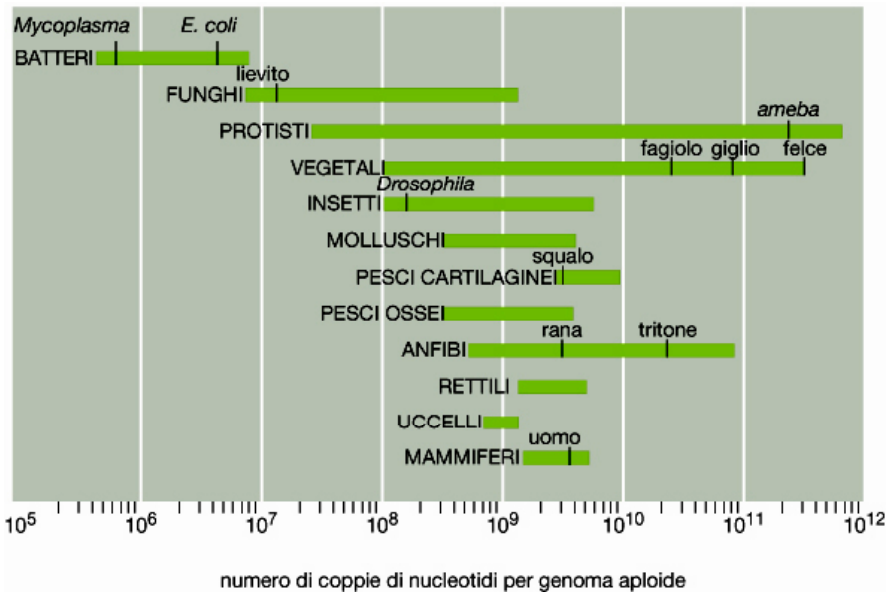
Hundreds of genomes sequenced

NCBI site [Genomic Data](#)

<http://www.ncbi.nlm.nih.gov/projects/genome/guide/>

Organismo	grandezza del genoma (Mb)	N° cromosomi
Batteriofago lambda	0,05	1
Adenovirus	0,15	1
E. coli	4,6	1
S. cerevisiae	12,1	16
D. melanogaster	180	4
Uomo	3.200	23
Mais	2.365	10
Triticum Aestivum	17.000	
Salamandra	90.000 (stima)	12

No linear relationship with gene numbers



Quante funzioni (geni) sono necessari per la vita "minima"?

Tabella 2.9 Esempi di organizzazione genomica nei procarioti

Specie	Organizzazione genomica		
	Molecole di DNA	Dimensioni (Mb)	Numero di geni
<i>Escherichia coli</i> K-12	Una molecola circolare	4,639	4397
<i>Vibrio cholerae</i> El Tor N16961	Due molecole circolari		
	Cromosoma principale	2,961	2770
	Megaplasmide	1,073	1115
<i>Deinococcus radiodurans</i> R1	Quattro molecole circolari		
	Cromosoma 1	2,649	2633
	Cromosoma 2	0,412	369
	Megaplasmide	0,177	145
	Plasmide	0,046	40
<i>Borrelia burgdorferi</i> B31	Sette o otto molecole circolari, undici molecole circolari		
	Cromosoma lineare	0,911	853
	Plasmide circolare cp9	0,009	12
	Plasmide circolare cp26	0,026	29
	Plasmide circolare cp32*	0,032	Non conosciuto
	Plasmide lineare lp17	0,017	25
	Plasmide lineare lp25	0,024	32
	Plasmide lineare lp28-1	0,027	32
	Plasmide lineare lp28-2	0,030	34
	Plasmide lineare lp28-3	0,029	41
	Plasmide lineare lp28-4	0,027	43
	Plasmide lineare lp36	0,037	54
	Plasmide lineare lp38	0,039	52
	Plasmide lineare lp54	0,054	76
	Plasmide lineare lp56	0,056	Non conosciuto

Gene structure in Prokaryotes

Classificazione del DNA eucariotico

Geni codificanti proteine

- geni solitari (in copia unica)
- geni duplicati e diversificatisi (famiglie di geni funzionali e *pseudogeni* non funzionali)

Geni ripetuti (che codificano rRNA, tRNA, rRNA 5S e istoni)

DNA ripetitivo

- DNA a sequenza semplice (satellite)
- DNA a ripetitività intermedia (elementi genetici mobili)
- Trasposoni
- Retrotrasposoni virali
- Elementi dispersi lunghi (LINES; retrotrasposoni non virali)
- Elementi dispersi corti (SINES; retrotrasposoni non virali)

DNA spaziatore non classificato

H. sapiens

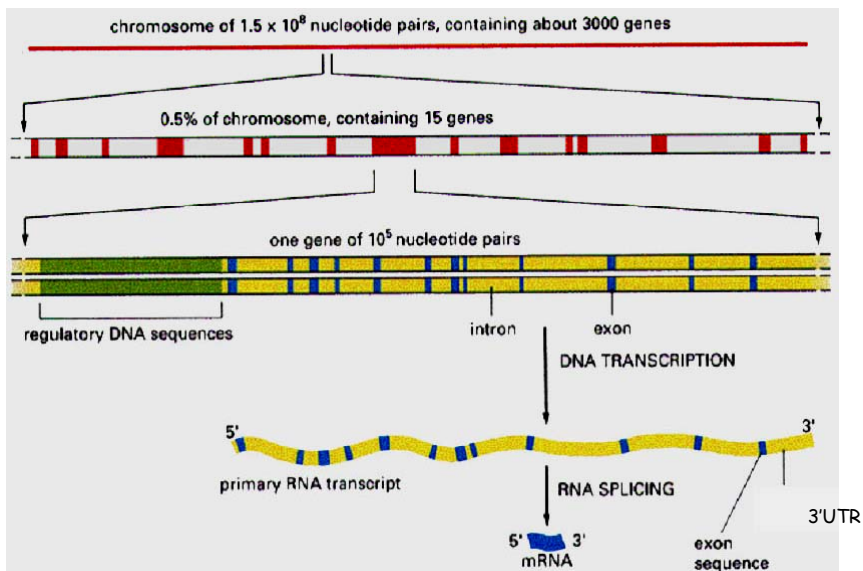
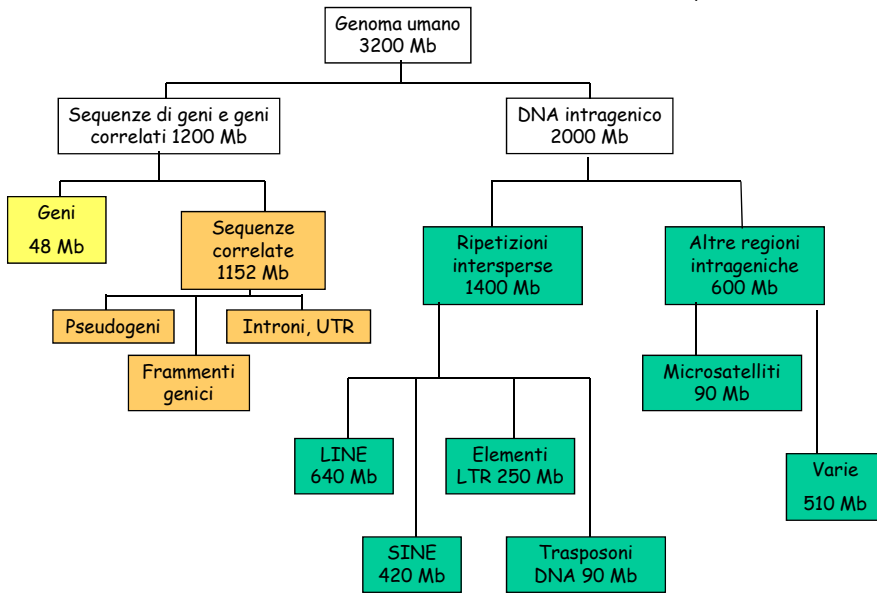
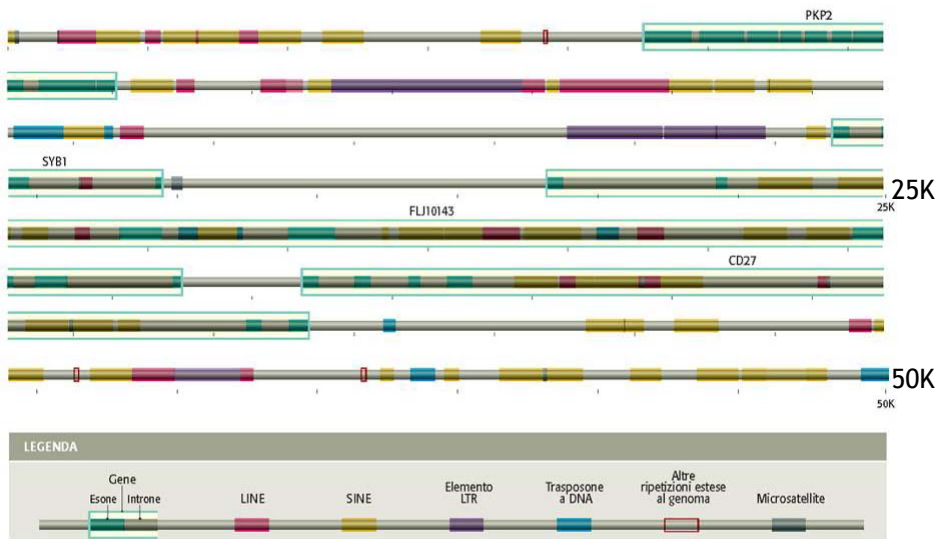


Figura 7.12 Un tratto del genoma umano.

Questa mappa mostra la posizione dei geni, dei segmenti genici, delle ripetizioni estese all'intero genoma e dei microsatelliti in un segmento da 50 kb del cromosoma 12 umano.



Geni

- codificanti proteine (mRNA)
- codificanti ncRNA (RNA non codificante)
 - ribosomali e tRNA
 - snRNA
 - miRNA (micro RNA)
 - antisenso
 - altri

Ripetitive

- intersperse
- tandem

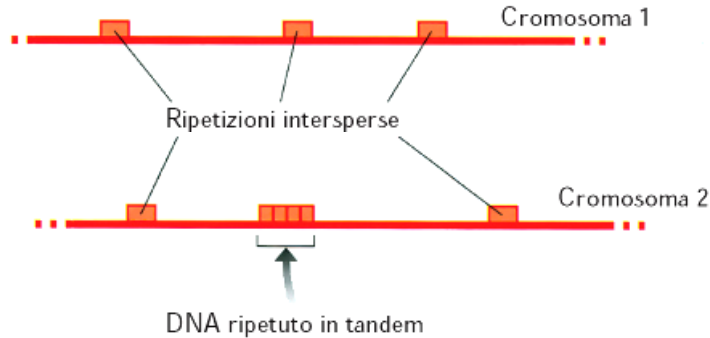


Figura 1.21 I due tipi di DNA ripetitivo: ripetizioni intersperse e DNA ripetuto in tandem.



T.A. Brown
Genomi
EdiSES

Ripetizioni codificanti

Table 19 Number of tRNA genes in various organisms

Organism	Number of canonical tRNAs	SeCys tRNA
Human	497	1
Worm	584	1
Fly	284	1
Yeast	273	0
<i>Methanococcus jannaschii</i>	36	1
<i>Escherichia coli</i>	86	1

Number of tRNA genes in each of six genome sequences, according to analysis by the computer program tRNAscan-SE. Canonical tRNAs read one of the standard 61 sense codons; this category excludes pseudogenes, undetermined anticodons, putative suppressors and selenocysteine tRNAs. Most organisms have a selenocysteine (SeCys) tRNA species, but some unicellular eukaryotes do not (such as the yeast *S. cerevisiae*).

Ripetizioni codificanti

The human genome contains:

2,000 genes
encoding **5S rRNA** in one cluster on chromosome 1

280 copies
of the **transcription unit encoding 28S, 5,8S and 18S rRNA**,
organized in 5 clusters of 50-70 units on chromosomes
13, 14, 15, 21, 22.

Ripetizioni non codificanti

Tabella 1.3 Microsatelliti nel genoma umano

Lunghezza della unità ripetitiva	Numero approssimativo delle copie nel genoma umano
1	120.000
2	140.000
3	37.500
4	105.000
5	56.000
6	49.000
7	27.000
8	35.500
9	27.500
10	27.500
11	28.000

Da IHGSC (2001).

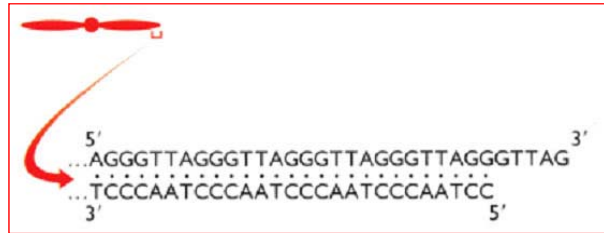
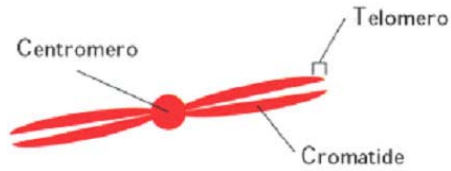
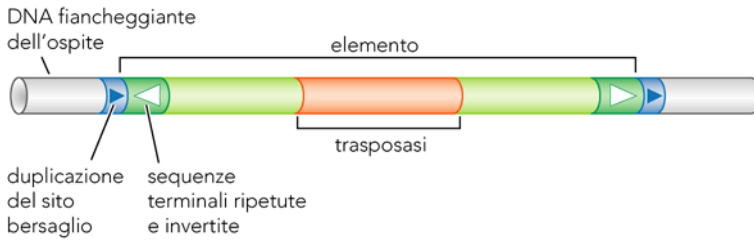


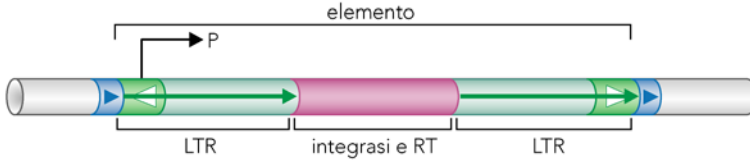
Tabella 1.2 Tipi di ripetizioni estese a tutto il genoma nell'uomo

Tipo di ripetizione	Sottotipo	Numero approssimativo delle copie nel genoma umano
SINE	Alu	1.558.000
	MIR	1.090.000
	MIR3	393.000
		75.000
LINE		868.000
	LINE-1	516.000
	LINE-2	315.000
	LINE+3	37.000
Elementi LTR		443.000
	Classe I ERV	112.000
	Classe II ERV(K)	8.000
	Classe III ERV(L)	83.000
	MaLR	240.000
Trasposoni DNA		294.000
	hAT	195.000
	Tc-I	75.000
	PiggyBac	2.000
	Non classificato	22.000

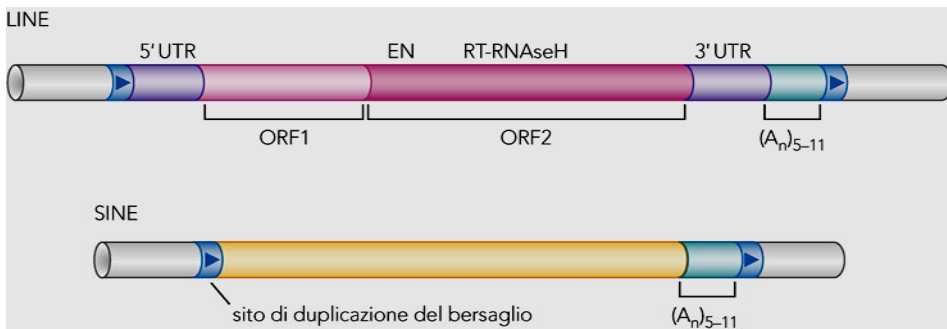
a trasposoni a DNA



b retrotrasposoni tipo virus/retrovirus



c retrotrasposoni poli-A



Trasposizione e retrotrasposizione

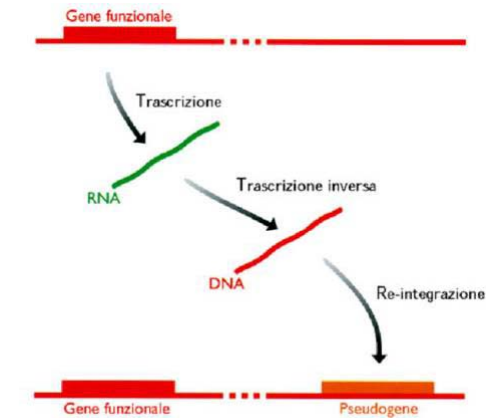
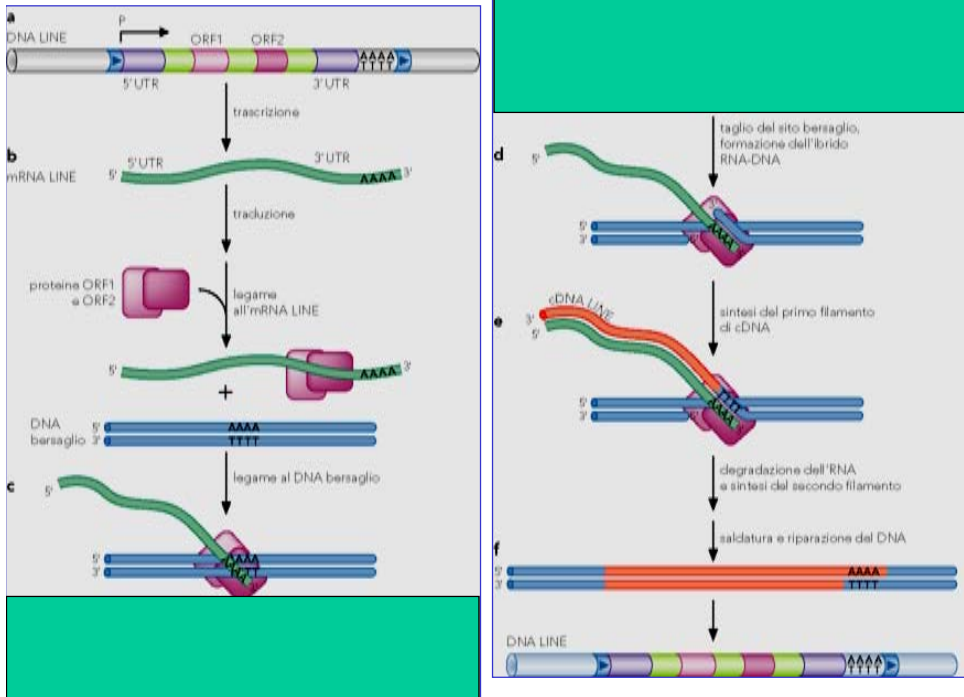
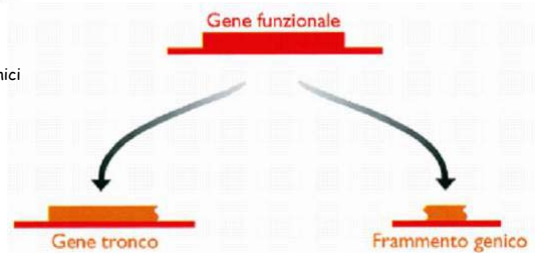


Figura 1.19 L'origine di un pseudogene maturato.

e di frammenti genici



Non c'è relazione lineare stretta tra lunghezza del genoma e complessità degli organismi

Nei Vertebrati, c'è una sostanziale coerenza nel numero di geni, che tuttavia sono pochi rispetto ai geni di organismi molto meno complessi

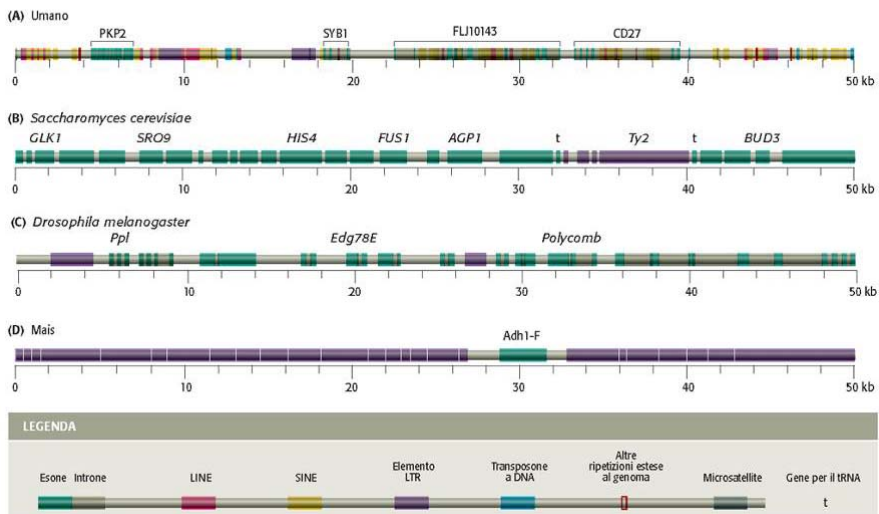
I genomi hanno contenuto molto variabile di sequenze ripetitive e non codificanti

Meno del 3% circa dei genomi dei Vertebrati codifica proteine

Caratteristica	Lievito	Moscerino della frutta	Uomo
Densità genica (numero medio per Mb)	479	76	11
Introni per gene (media)	0,04	3	9
Percentuale del genoma occupata dalle ripetizioni intersperse	3,4%	12%	44%

Specie	Dimensioni (Mb)	N° di geni
<i>Saccharomyces cerevisiae</i>	12,1	5.800
<i>Schizosaccharomyces pombe</i>	12,5	4.900
<i>Caenorhabditis elegans</i>	97	19.000
<i>Arabidopsis thaliana</i>	125	25.500
<i>Drosophila melanogaster</i>	180	13.600
<i>Oryza sativa</i>	466	40.000
<i>Gallus gallus</i>	1200	20-23.000
<i>Homo sapiens</i>	3200	30-40.000

Figura 7.15 Confronto tra genoma umano, di lievito, del moscerino della frutta e di mais. (A) Il segmento di 50 kb del cromosoma 12 umano mostrato precedentemente, è confrontato con segmenti di 50 kb derivanti da genomi di (B) *S. cerevisiae*; (C) *Drosophila melanogaster*; (D) mais.



Gene functions

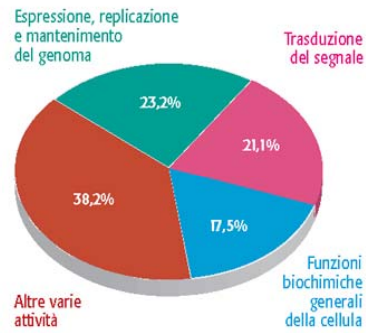


Figura 7.16 Classificazione del catalogo dei geni umani. Il diagramma a torta mostra la suddivisione dei geni umani codificanti proteine finora identificati. Mancano approssimativamente 13.000 geni, le cui funzioni sono sconosciute. La porzione corrispondente a varie altre attività include, tra gli altri, le proteine coinvolte nei processi di trasporto biochimico e nel ripiegamento, le proteine immunologiche e le proteine strutturali.

7.17 - Confronto tra i cataloghi genici di *S. cerevisiae*, *A. thaliana*, *C. elegans*, *D. melanogaster* e *H. sapiens*. I geni sono raggruppati in base alle loro funzioni, dedotta dai domini proteici che sono specificati da ciascun gene.

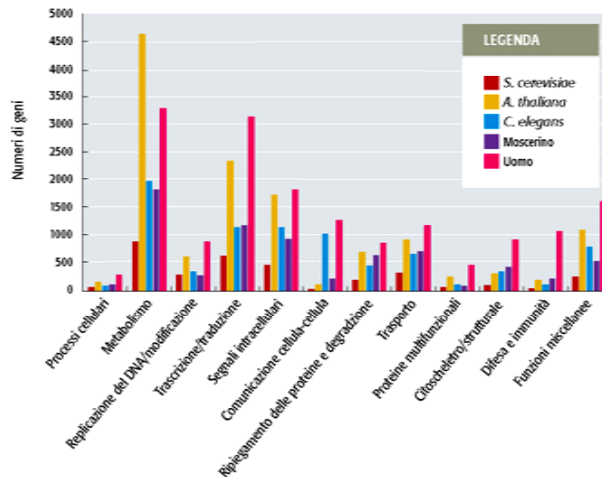
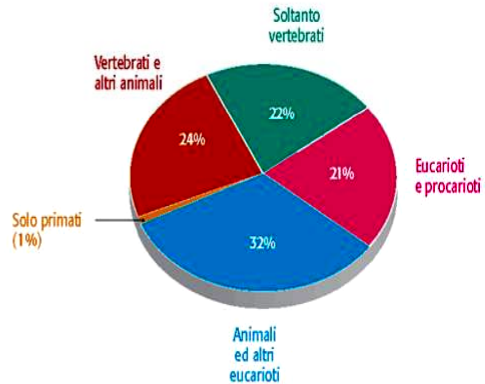


Figura 7.18 Relazione tra il catalogo genico umano e quello di altri gruppi di organismi. Il grafico a torta classifica il catalogo genico umano in base alla distribuzione dei singoli geni negli altri organismi. Il grafico mostra, per esempio, che il 22% del catalogo genico umano è costituito da geni che sono specifici per i vertebrati, mentre un altro 24% comprende geni specifici per i vertebrati e altri animali.



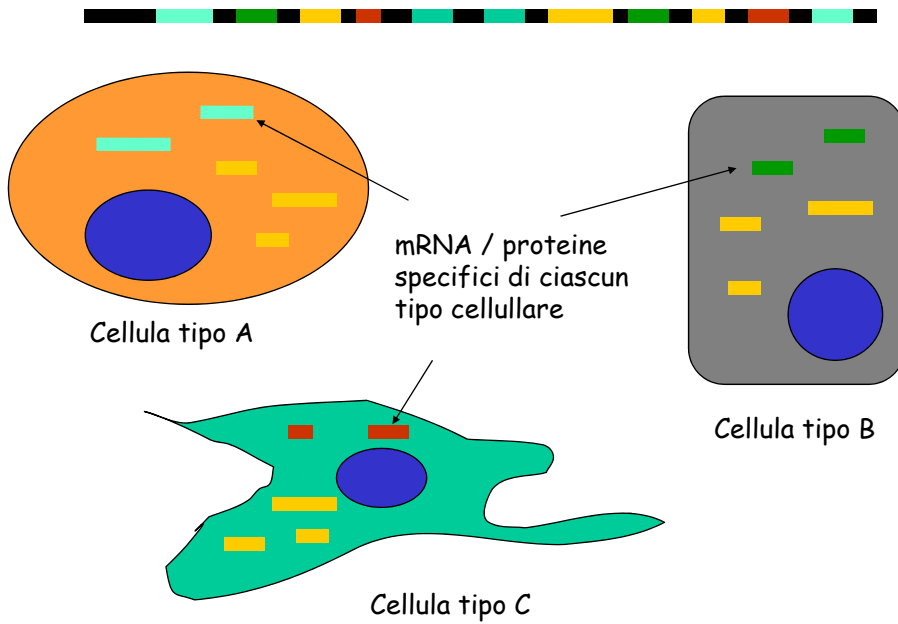
T.A. Brown
Genomi, III Ed.
EdiSES

Genome expression

Development, growth, apoptosis, homeostasis, and all cell activities depend on the qualitative and quantitative control of gene expression.

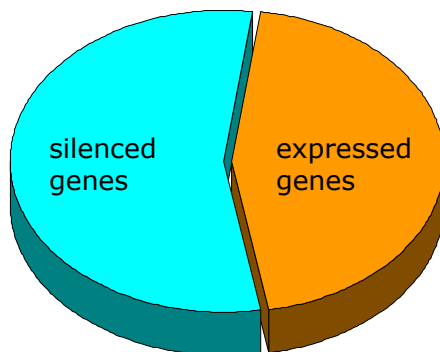
How could we estimate or measure the entire genome expression in a cell, tissue or organism ?

Il DNA di ogni cellula è identico: tutte le cellule contengono l'intera infor



In any cell type, a large part of the genes are kept in a silenced form.

...in any differentiated cells:



The primary control is at the genomic level, i.e. transcription

H. sapiens genome: 30,000 genes estimated

alternative splicing

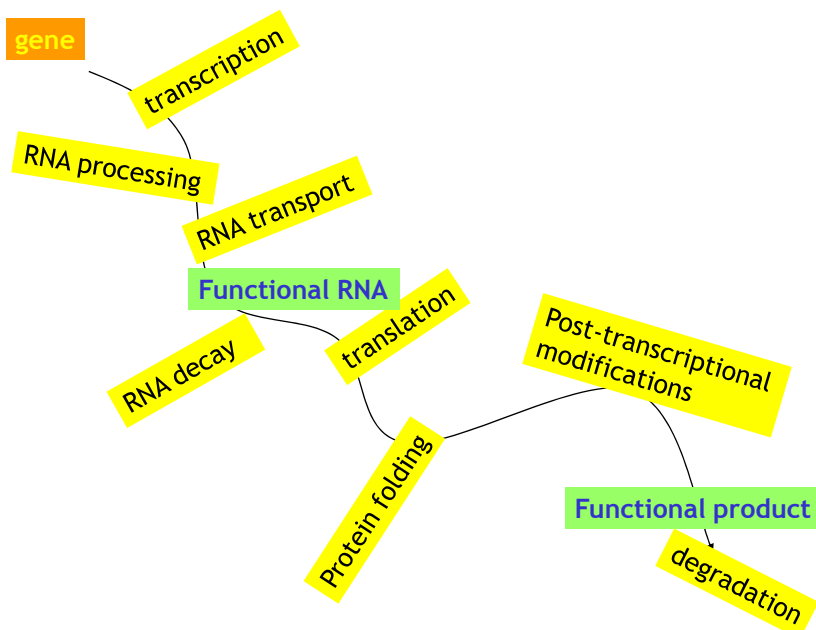
100 - 150,000 proteins estimated



We call **genetic program** the set of genes that a cell uses for expressing a function (e.g. differentiation, cell division, apoptosis, response to a drug, etc...)

We call **gene expression profile** the sum of genes that are expressed in a cell or tissue, weighted by their individual levels of expression.

Transcription is the most important level of control



Large-scale **analysis of gene expression**

measuring mRNAs

-More mature and technically OK

Gene-by-gene methods to measure gene expression (mRNA)

measuring proteins

- Several approaches available, still under development

Gene-by-gene methods to measure gene expression (proteins)

Why mRNA?

- mRNAs represent a closer mirror of genome activity than proteins
- Homogeneous chemistry makes handling robust
- Complete knowledge of probes and control of hybridization conditions guarantees specificity
- The amount of any mRNA species is not necessarily proportional to the amount of the encoded protein
- One gene can encode for different proteins by alternative splicing

Why proteins?

- Proteins are the real object of gene expression
- Separation and quantitation may be more reliable than for mRNAs
- Not all measured protein may represent “functional” protein.
- Dis-homogeneous chemistry makes it difficult to find procedures equally good for all proteins
- Specificity of interactions is hardly controlled, since it is based on a sum of different chemical interactions

Large-scale analysis of gene expression limits

(with microarray methodology → until deep-sequencing methods)

Theoretical limit: measuring all genes

Practical limit: measuring all known genes

Practice: measuring a relevant part of genes

How to measure the activity of all genes (genome-wide) in cells/tissues (mRNA).

Sequencing methods

(EST)

SAGE (LongSAGE, CAGE)

Sequence approach

direct re-sequencing (deep sequencing)

DNA microarrays, oligonucleotide microarrays.

Spotted arrays

In situ synthesized oligo arrays

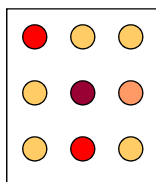
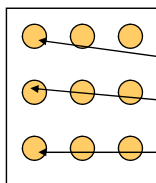
Hybridization approach

Bead-arrays®

Hybridization to microarrays

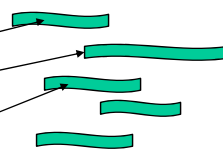
La tecnica di ibridazione su fase solida permette di verificare la presenza di molti geni contemporaneamente

Sonde complementari a diversi geni vengono depositate su un filtro

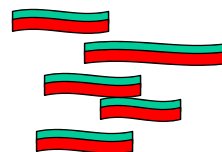


Gene A RNA A
Gene B RNA B
Gene C RNA C
eccetera eccetera ...

L'RNA o il DNA vengono estratti dalle cellule



appositamente marcati



...e ibridizzati alle sonde sul filtro



