Γονιδιωματική (Genomics)

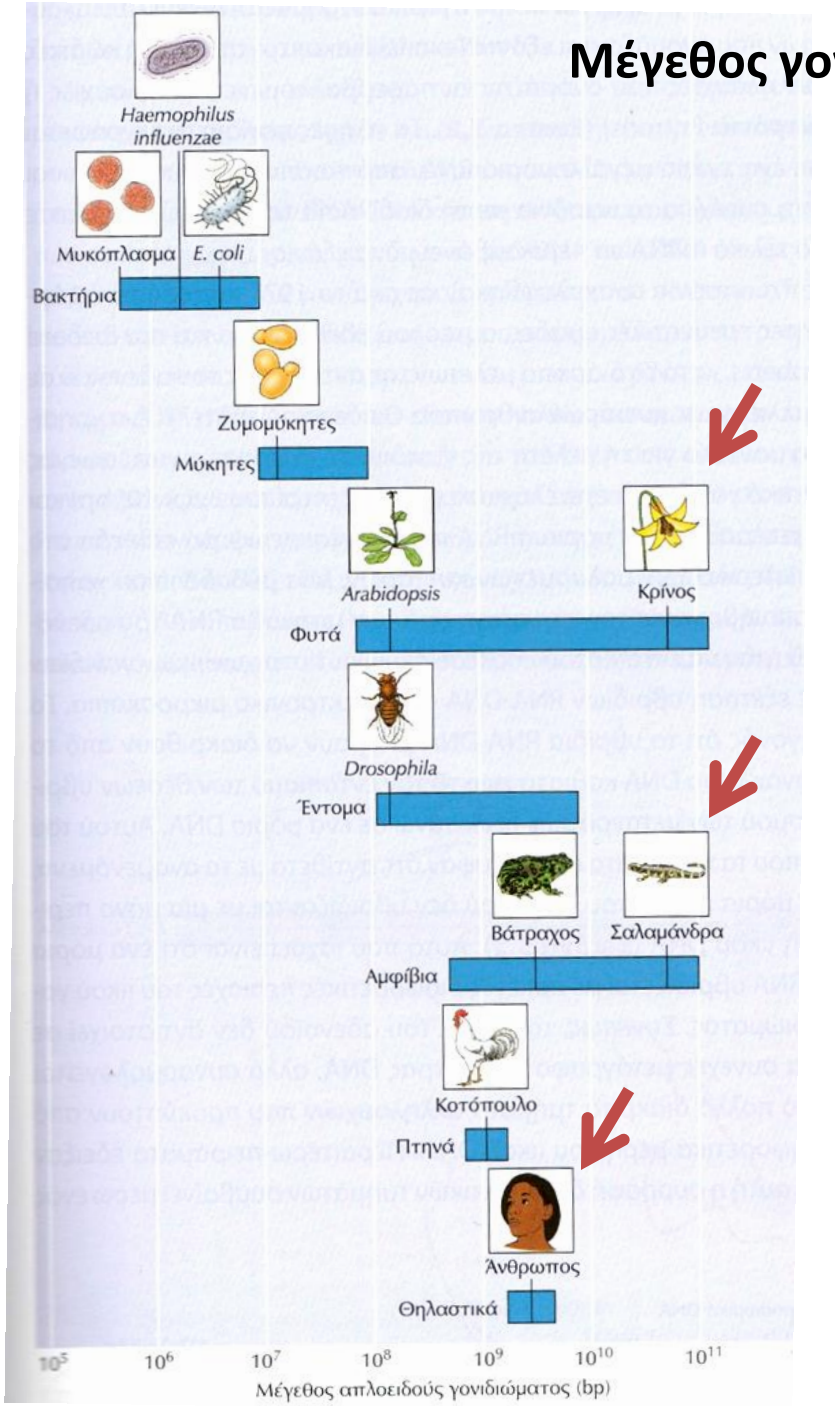
Θεολογία Σαραφίδου

Επικ. Καθηγήτρια Μοριακής Γενετικής
Τμήμα Βιοχημείας και Βιοτεχνολογίας
Πανεπιστήμιο Θεσσαλίας

Γονιδιωματική: Αλληλούχηση και λειτουργική μελέτη ολόκληρου του γονιδιώματος

Λειτουργική γονιδιωματική

Μέγεθος γονιδιώματος

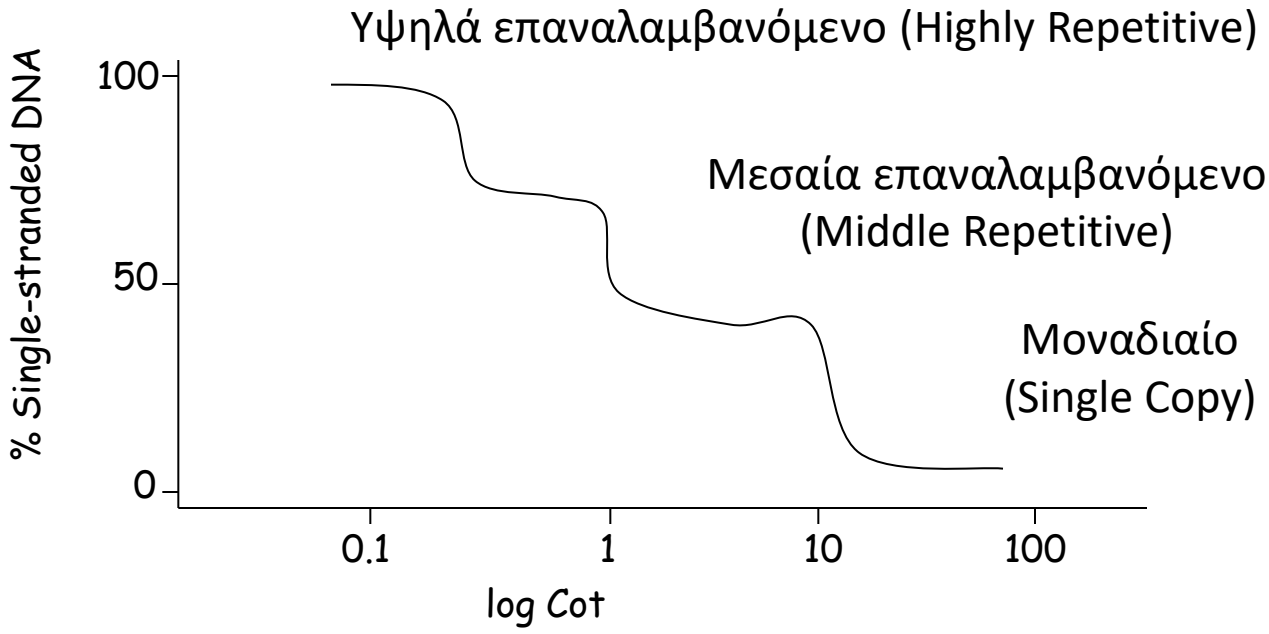


Το μέγεθος δεν σχετίζεται με την πολυπλοκότητα του οργανισμού

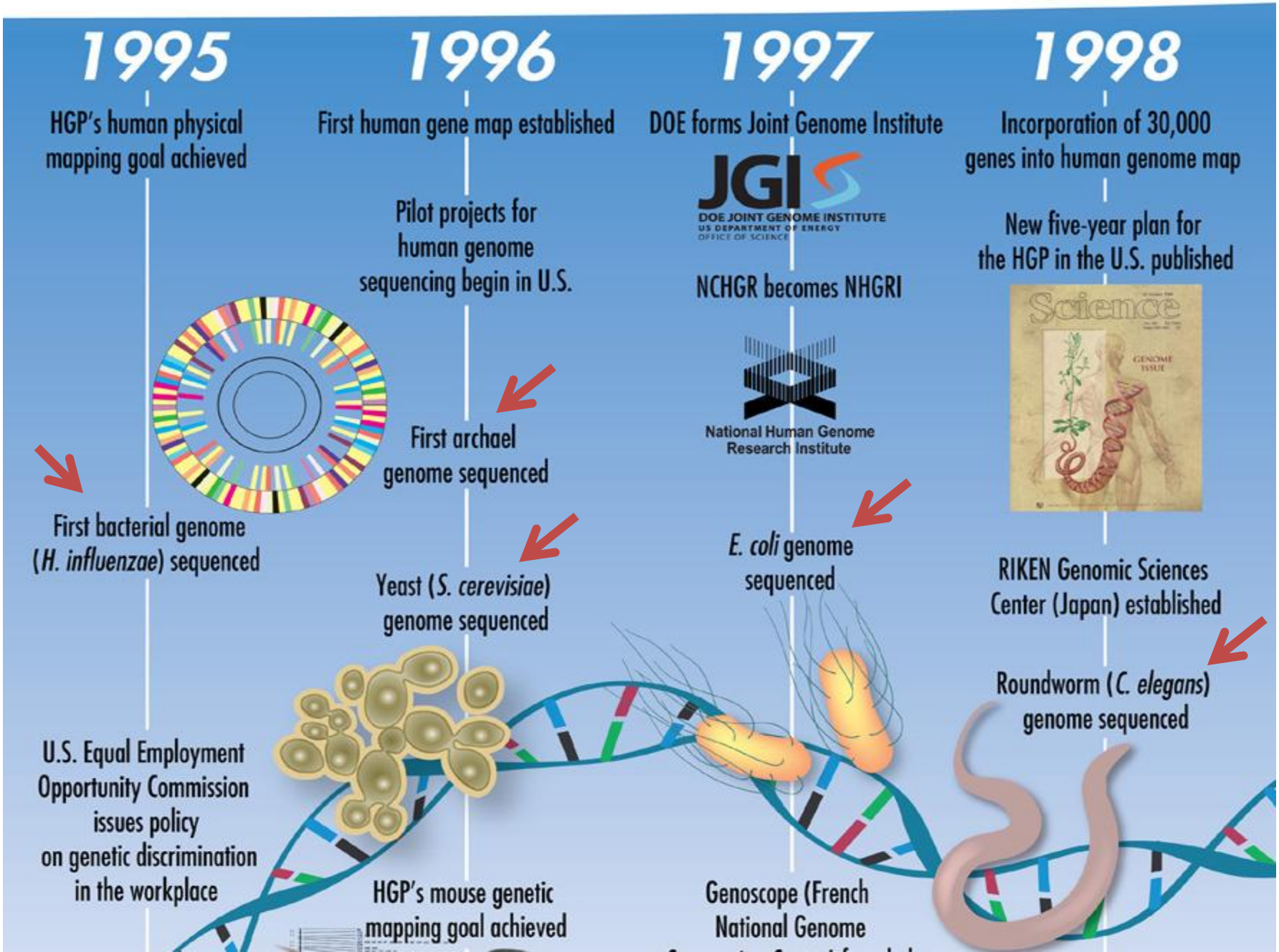
Επαναλαμβανόμενο DNA

Εκτίμηση % επαναλαμβανόμενων αλληλουχιών με επαναδιάταξη αποδιατεταγμένων θραυσμάτων DNA → Ο ρυθμός επανασύνδεσης εξαρτάται από τη σχετική συγκέντρωση των δύο κλώνων DNA

Πολυκύτταροι ευκαρυωτικοί



Αλληλούχηση ολόκληρων γονιδιωμάτων



1999

Full-scale human sequencing begins

Sequence of first human chromosome (chromosome 22) completed

2000

Draft version of human genome sequence completed

President Clinton and Prime Minister Blair support free access to genome information

Fruit fly (*D. melanogaster*) genome sequenced

Mustard cress (*A. thaliana*) genome sequenced

2001

Draft version of human genome sequence published

10,000 full-length human cDNAs sequenced

Mammalian Gene Collection

2002

Draft version of mouse genome sequence completed and published

Draft version of rat genome sequence completed

Draft version of rice genome sequence completed and published

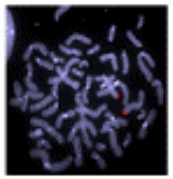
2003

Finished version of human genome sequence completed

HGP ends with all goals achieved

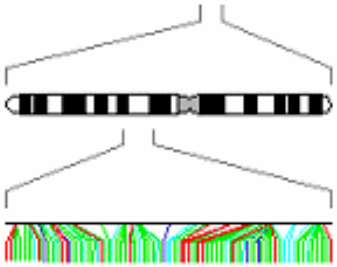
to be continued..

Στρατηγικές αλληλούχησης γονιδιωμάτων



BY MAPPED CLONES

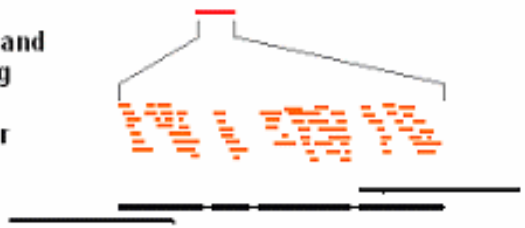
1. Construction of maps of ordered landmarks (genetic markers, genes): provides long-range map and organisation into individual chromosomes.



2. Physical maps of overlapping clones anchored to the landmark maps.

3. Selection of tile path (clones in red)

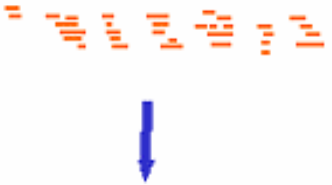
4. Shotgun sequencing and assembly (for working draft); subsequent directed finishing (for reference sequence).



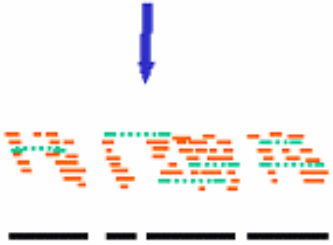
BY WHOLE GENOME SHOTGUN

1. Shotgun sequencing of short-insert clones

2. Paired end sequencing of large-insert clones



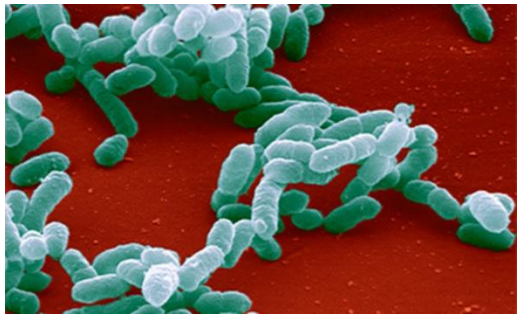
3. Assembly of seed contigs (unitigs)



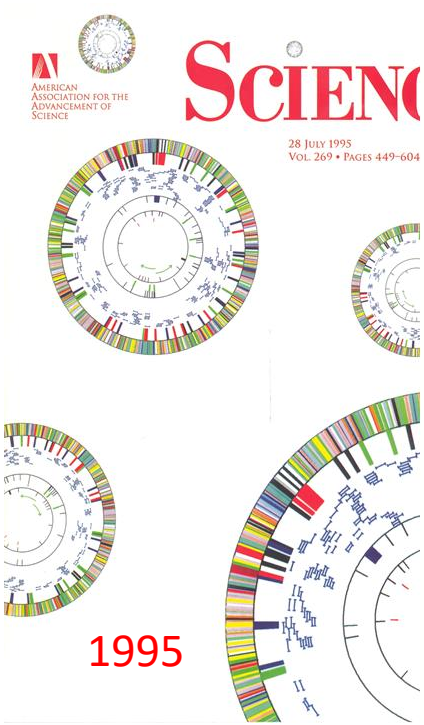
4. Incorporation of other sequences, and integration of long- range data.

Προκαρυωτικά γονιδιώματα

Haemophilus influenzae: 1,8Mb



Λοίμωξη αναπνευστικού



1995

Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd

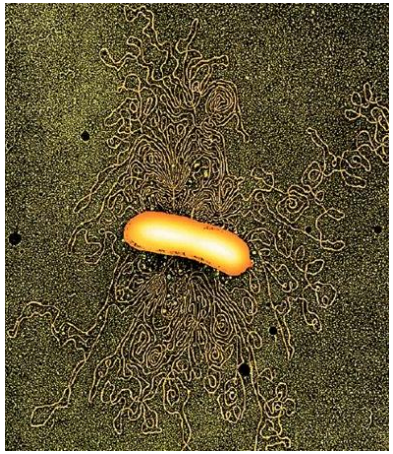
Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, Jean-Francois Tomb, Brian A. Dougherty, Joseph M. Merrick, Keith McKenney, Granger Sutton, Will FitzHugh, Chris Fields,* Jeannine D. Gocayne, John Scott, Robert Shirley, Li-Ing Liu, Anna Glodek, Jenny M. Kelley, Janice F. Weidman, Cheryl A. Phillips, Tracy Spriggs, Eva Hedblom, Matthew D. Cotton, Teresa R. Utterback, Michael C. Hanna, David T. Nguyen, Deborah M. Saudek, Rhonda C. Brandon, Leah D. Fine, Janice L. Fritchman, Joyce L. Fuhrmann, N. S. M. Geoghagen, Cheryl L. Gnehm, Lisa A. McDonald, Keith V. Small, Claire M. Fraser, Hamilton O. Smith, J. Craig Venter†

An approach for genome analysis based on sequencing and assembly of unselected pieces of DNA from the whole chromosome has been applied to obtain the complete nucleotide sequence (1,830,137 base pairs) of the genome from the bacterium *Haemophilus influenzae* Rd. This approach eliminates the need for initial mapping efforts and is therefore applicable to the vast array of microbial species for which genome maps are unavailable. The *H. influenzae* Rd genome sequence (Genome Sequence DataBase accession number L42023) represents the only complete genome sequence from a free-living organism.

1.743 πιθανά ORFs
54 tRNAs
6rRNAs

} αντιστοιχούν στο ~90% του γονιδιώματος

Escherichia coli: 4,6Mb



Τα περισσότερα στελέχη συμβιωτικά στο έντερο ζώων

ARTICLE

1997 **The Complete Genome Sequence of *Escherichia coli* K-12**

Frederick R. Blattner,* Guy Plunkett III,* Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, Ying Shao

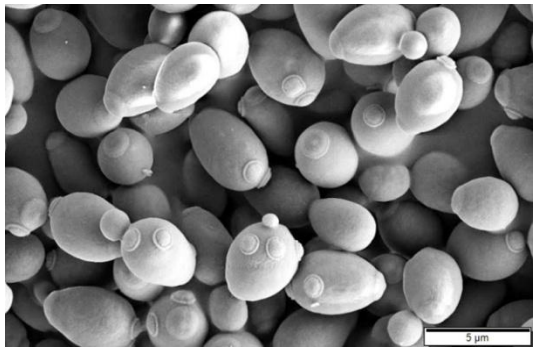
The 4,639,221–base pair sequence of *Escherichia coli* K-12 is presented. Of 4288 protein-coding genes annotated, 38 percent have no attributed function. Comparison with five other sequenced microbes reveals ubiquitous as well as narrowly distributed gene families; many families of similar genes within *E. coli* are also evident. The largest family of paralogous proteins contains 80 ABC transporters. The genome as a whole is strikingly organized with respect to the local direction of replication; guanines, oligonucleotides possibly related to replication and recombination, and most genes are so oriented. The genome also contains insertion sequence (IS) elements, phage remnants, and many other patches of unusual composition indicating genome plasticity through horizontal transfer.

The first 1.92 Mb (*13*, *14*), positions 2,686,777 to 4,639,221 [in base pairs (bp)], was sequenced from our overlapping set of 15- to 20-kb MG1655 lambda clones (*15*) by means of radioactive chemistry and was deposited in GenBank between 1992 and 1995. Subsequently, we switched to dye-terminator fluorescence sequencing (Applied Biosystems). In addition to greater speed and lower cost, this new technology avoided electrophoretic compression arti-

- 4.288 γονίδια (κατά μέσο όρο 1 γονίδιο/kb)
- Πολύ μικρό ποσοστό επαναλαμβανόμενου DNA
- Μικρές δια-γονιδιακές περιοχές
- ~90% γονιδιώματος κωδικοποιεί πρωτεΐνες

Το γονιδίωμα του πρώτου ευκαρυωτικού οργανισμού

Saccharomyces cerevisiae → Το πιο απλό ευκαρυωτικό γονιδίωμα: 1,2 x 10^7 bp



6000 γονίδια
- 5885 mRNAs
- 140 rRNAs
- 275 tRNAs
- 40 snRNAs

Κωδικές αλληλουχίες ~ 70% γονιδιώματος

4% των γονιδίων έχουν (μικρά) ιντρόνια

Γονιδιώματα ασπονδύλων

C. elegans → 9,7 x 10^7 bp
19.000 γονίδια (με μ.ο. 5 ιντρόνια το καθένα)
Κωδικές αλληλουχίες → ~ 25% του γονιδιώματος

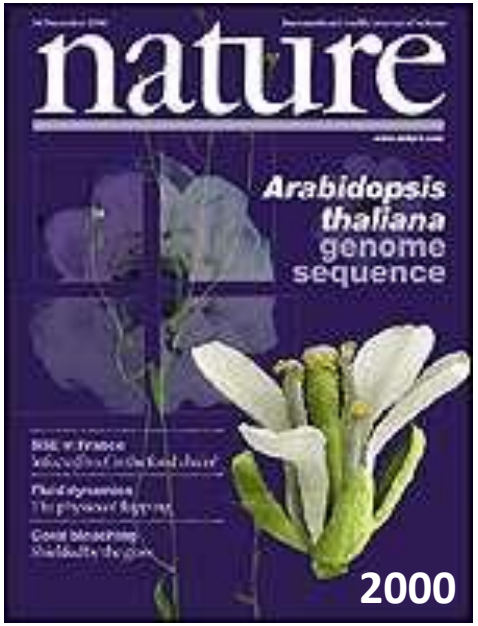
D. melanogaster → 1,8 x 10^8 bp

Το 1/3 → ετεροχρωματίνη, Αλληλούχηση ευχρωματίνης

14.000 γονίδια (με μ.ο. 4 ιντρόνια)

Περίπου διπλάσια από τον σακχαρομύκητα!
Λιγότερα από τον *C. elegans*!

Κωδικές αλληλουχίες → ~ 13% του γονιδιώματος



Γονιδιώματα φυτών

Arabidopsis thaliana: 1,25 x 10^8 bp (σχετικά μικρό γονιδίωμα)

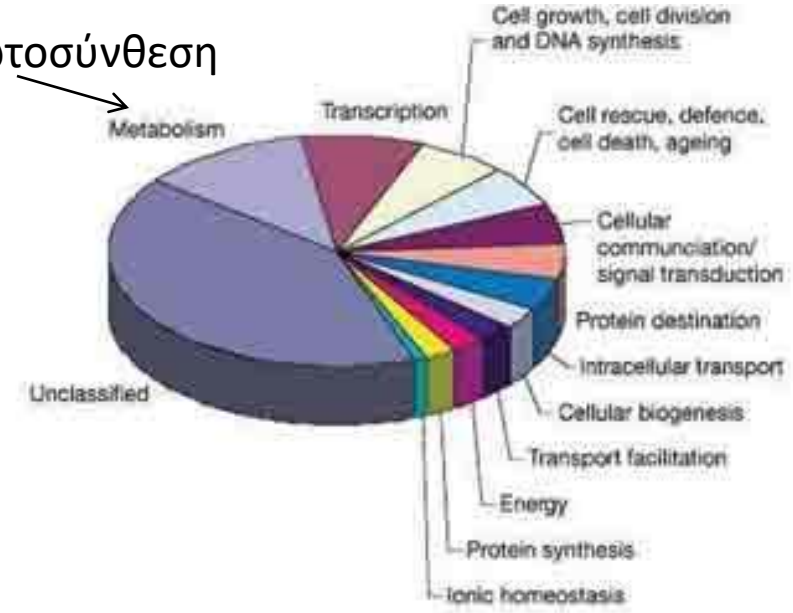
26.000 γονίδια! (με μ.ο. 4 ιντρόνια)
Διπλασιασμοί στο 60% γονιδιώματος (10.000/26.000 γονίδια)

Κωδικές αλληλουχίες → ~ 25% του γονιδιώματος

Συνολικό μήκος ιντρονίων ~ συνολικό μήκος εξονίων
10% γονιδιώματος → μεταθετά στοιχεία

Λειτουργική κατηγοριοποίηση πρωτεϊνών

φωτοσύνθεση



…και άλλα γονιδιώματα φυτών
Μεγαλύτερο αριθμό γονιδίων από τα ζώα

Το γονιδίωμα του ανθρώπου

Human Genome Project (HGP)

ΕΝΑΡΞΗ: 1988

ΣΤΟΧΟΙ:

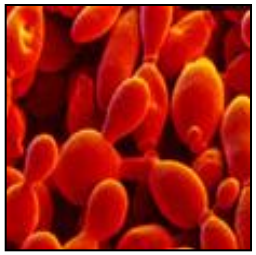
- Κατασκευή χρωμοσωμικών χαρτών και αλληλούχηση
- Ταυτοποίηση των γονιδίων που σχετίζονται με Μεντελικά & πολυπαραγοντικά νοσήματα
- Διερεύνηση της γενετικής ποικιλότητας
- Ανάλυση της γονιδιωματικής δομής & λειτουργίας του ανθρώπου και οργανισμών-μοντέλων

J. Watson

Διευθυντής στο National Human Genome Research Institute

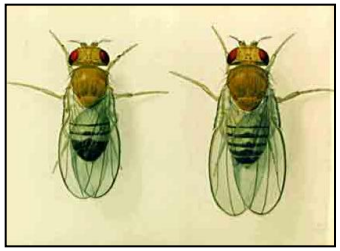
Οργανισμοί-μοντέλα

E. coli



S. cerevisiae

C. elegans



D. melanogaster

M. musculus

WORKSHOP ON INTERNATIONAL COOPERATION FOR THE HUMAN GENOME PROJECT

VALENCIA DECLARATION ON THE HUMAN GENOME PROJECT

1. The members of the workshop believe that knowledge gained from mapping and sequencing the human genome can have great benefit for human health and wellbeing. Towards these ends, participating scientists acknowledge their responsibility to help ensure that genetic information be used only to enhance the dignity of the individual. They also encourage public debate on ethical, social, legal, and commercial implications of the use of genetic information.

2. The members endorse the concept of international collaboration for the project and urge the widest possible participation of countries throughout the world, within the resources and interests of each country.

3. The participants strongly encourage parallel studies of genomes of selected animal, plant and micro-organism models in order to achieve a deeper understanding of the human genome.

4. The workshop urges coordination of research and information on complex genomes among nations and across disciplines and species.

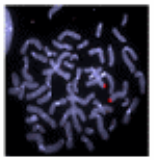
5. The workshop believes that information resulting from mapping and sequencing of the human genome should be in the public domain and made freely available to scientists of all countries.

6. The participants encourage continued effort to develop compatible genomic data bases and networks and measures to ensure world-wide access to these resources.

7. The workshop endorses The Human Genome Organization (HUGO) as the lead body, in collaboration with other non-governmental and government organizations, to promote the goals and objectives addressed in this declaration.

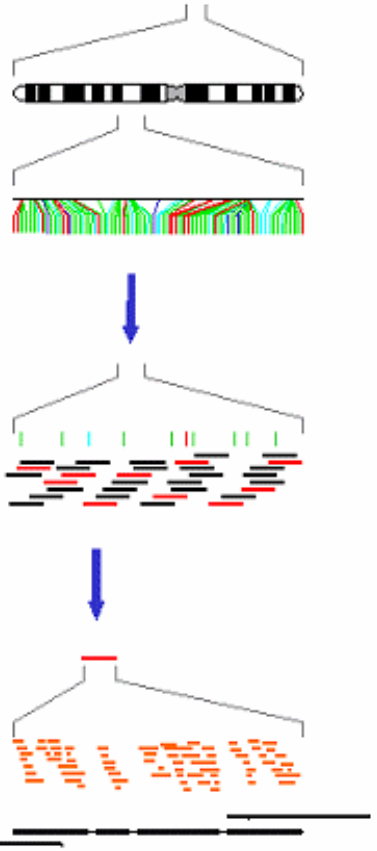
October 24-26, 1988
VALENCIA (Spain)

STRATEGIES FOR SEQUENCING THE HUMAN GENOME



BY MAPPED CLONES

1. Construction of maps of ordered landmarks (genetic markers, genes): provides long-range map and organisation into individual chromosomes.
2. Physical maps of overlapping clones anchored to the landmark maps.
3. Selection of tile path (clones in red)
4. Shotgun sequencing and assembly (for working draft); subsequent directed finishing (for reference sequence).

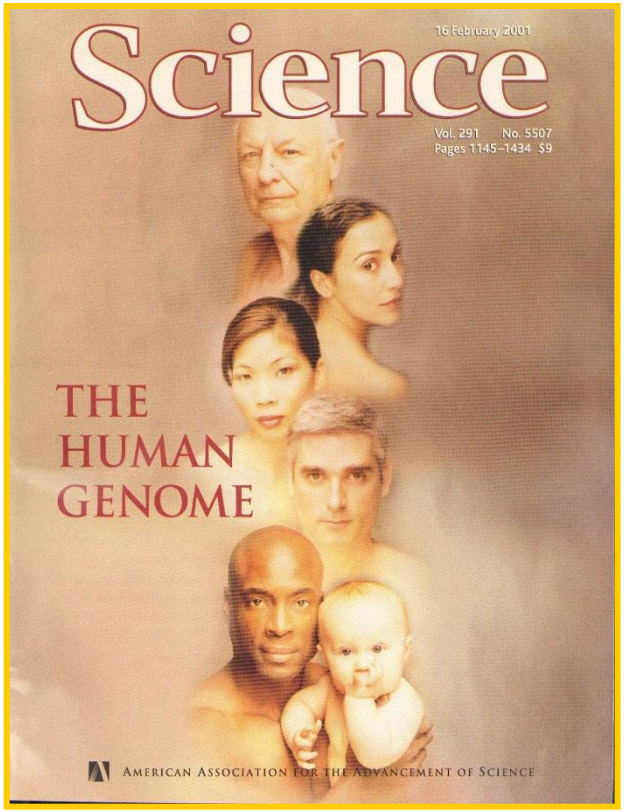


BY WHOLE GENOME SHOTGUN

1. Shotgun sequencing of short-insert clones
2. Paired end sequencing of large-insert clones
3. Assembly of seed contigs (unitigs)
4. Incorporation of other sequences, and integration of long- range data.

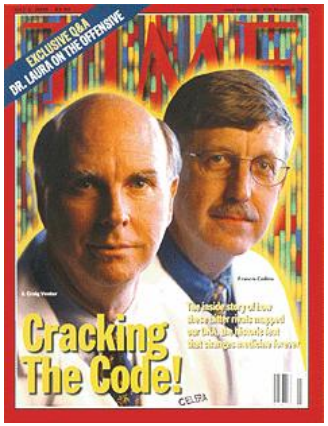
J. Graig Venter

Vol 409, Feb 15, 2001



Vol 291, Feb 16, 2001

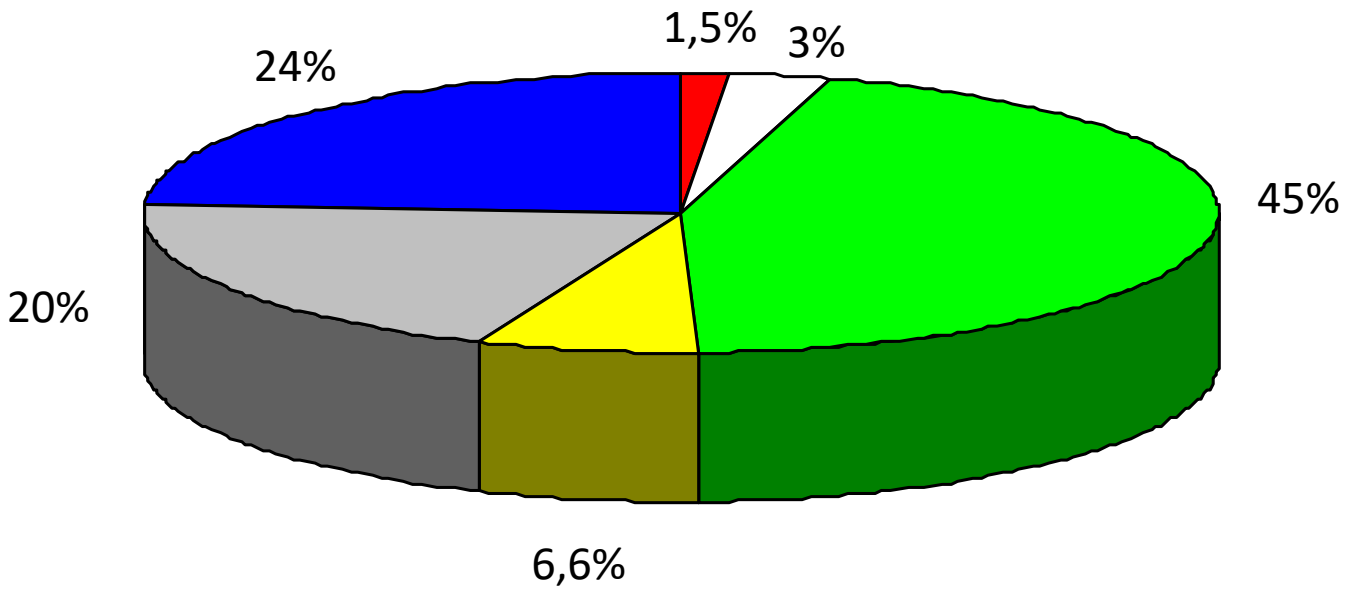
Human Genome Project
HGP
FREE ACCESS



CELERA GENOMICS

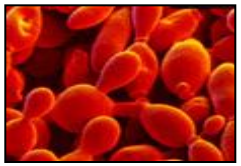
Access to db upon subscription

Το γονιδίωμα με αριθμούς



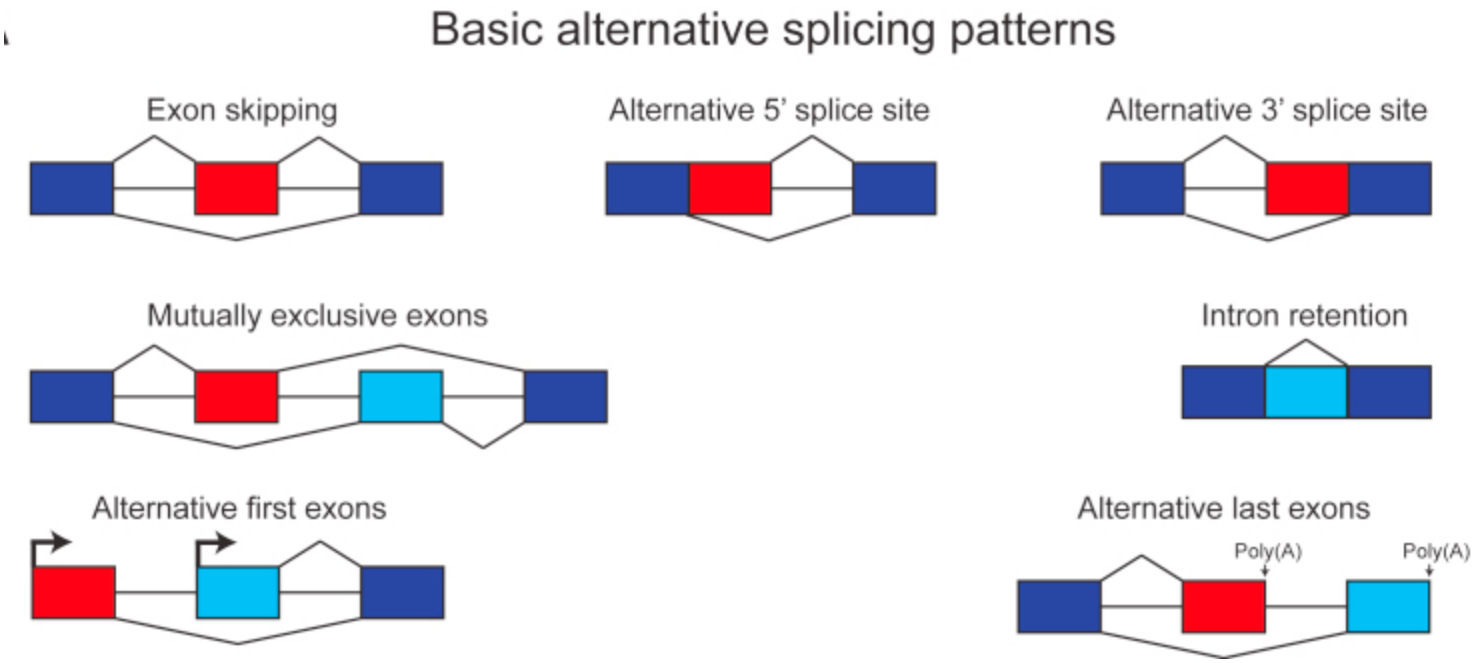
- Κωδικές περιοχές (υψηλή συντήρηση)
- Μη κωδικές περιοχές (υψηλή συντήρηση)
- Ιντρόνια
- Επαναλαμβανόμενες αλληλουχίες από μεταθετά στοιχεία
- Ετεροχρωματίνη
- Μη συντηρημένες περιοχές

Ο ΑΡΙΘΜΟΣ ΤΩΝ ΓΟΝΙΔΙΩΝ ΕΙΝΑΙ ΑΝΑΛΟΓΟΣ ΜΕ ΤΗΝ ΠΟΛΥΠΛΟΚΟΤΗΤΑ?

Μέγεθος γονιδιώματος		Αριθμός γονιδίων
14Mb		~6.000
100Mb		~19.000
140Mb		~13.000
115Mb		~40.000
430Mb		~26.000
3000Mb		~20.000

Η πολυπλοκότητα της γενετικής πληροφορίας

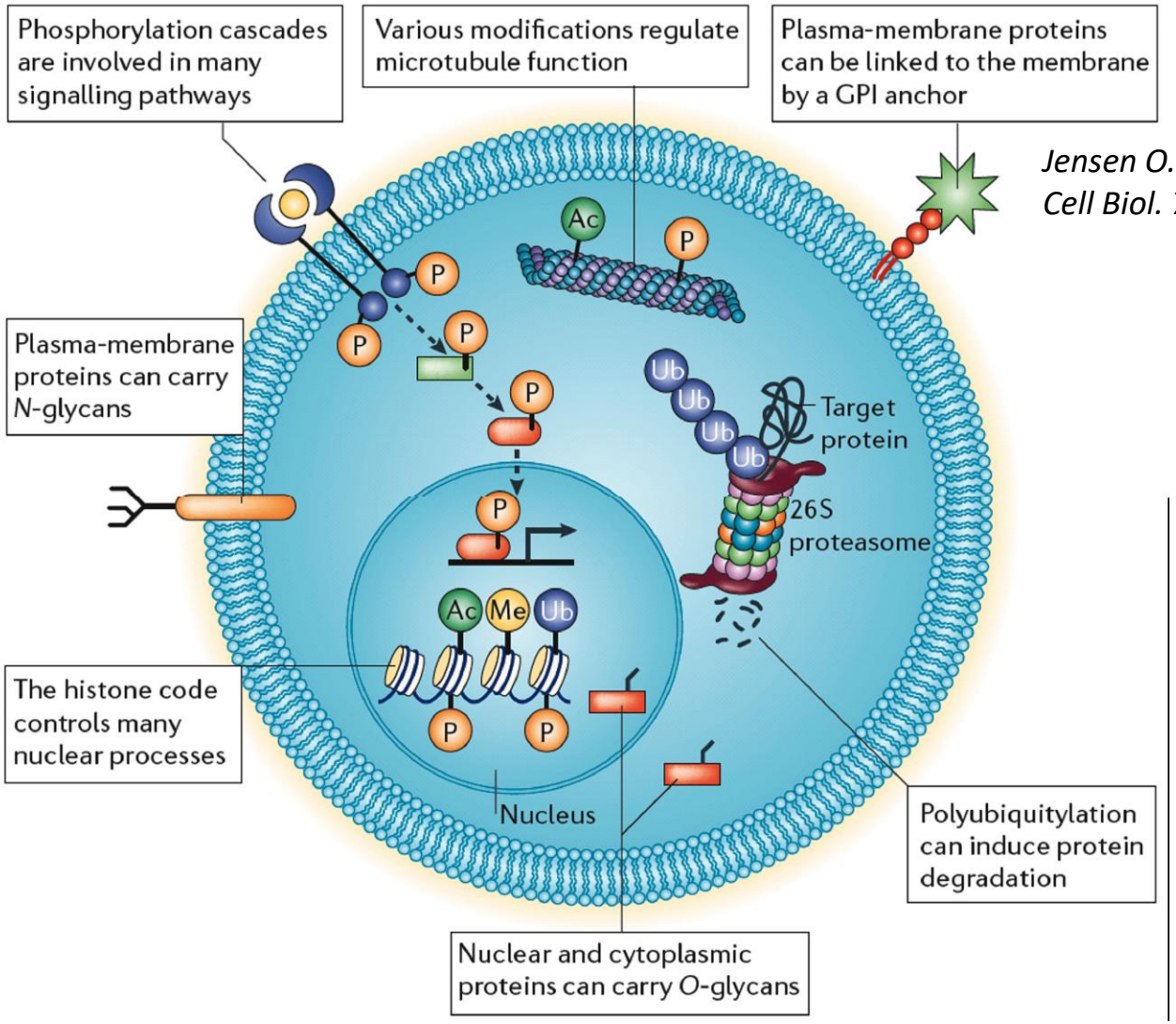
1. Εναλλακτική συναρμογή



Σχεδόν όλα τα γονίδια που έχουν πολλαπλά εξόνια
Ρυθμίζεται με βάση τον κυτταρικό τύπο/στάδιο ανάπτυξης

Χ γονίδια → 3-5 Χ μετάγραφα

2. Μετα-μεταφραστικές τροποποιήσεις



Jensen O. 2006. Nature Review Mol Cell Biol. 7,391-403.

Γονίδια = C

↓

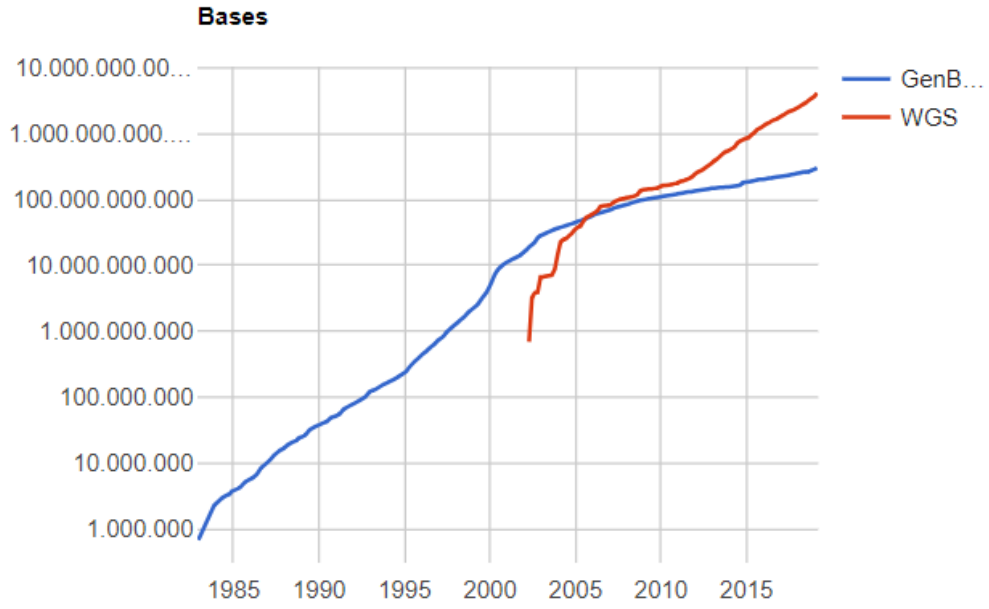
Μετάγραφα ~ 3-5 x C

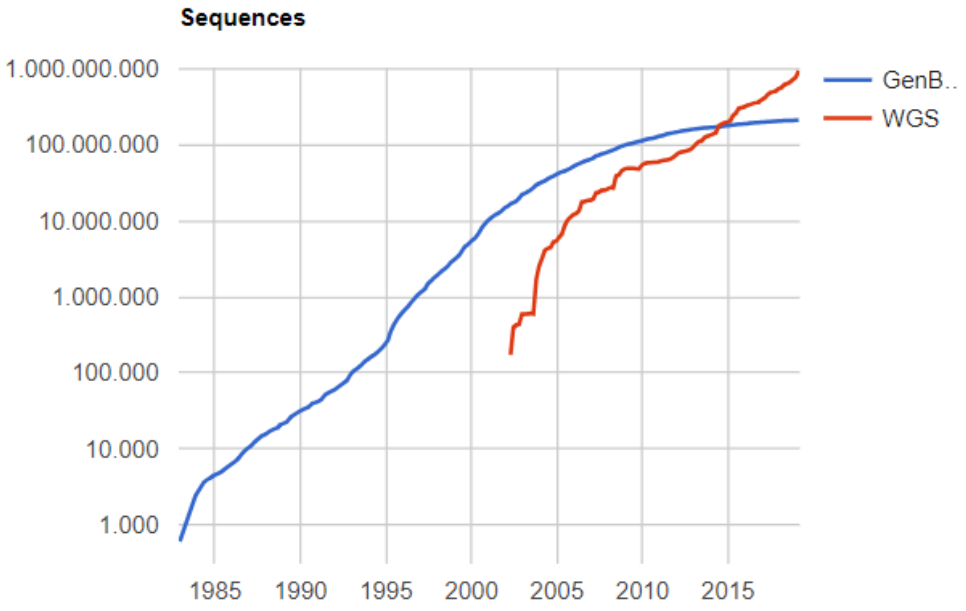
↓

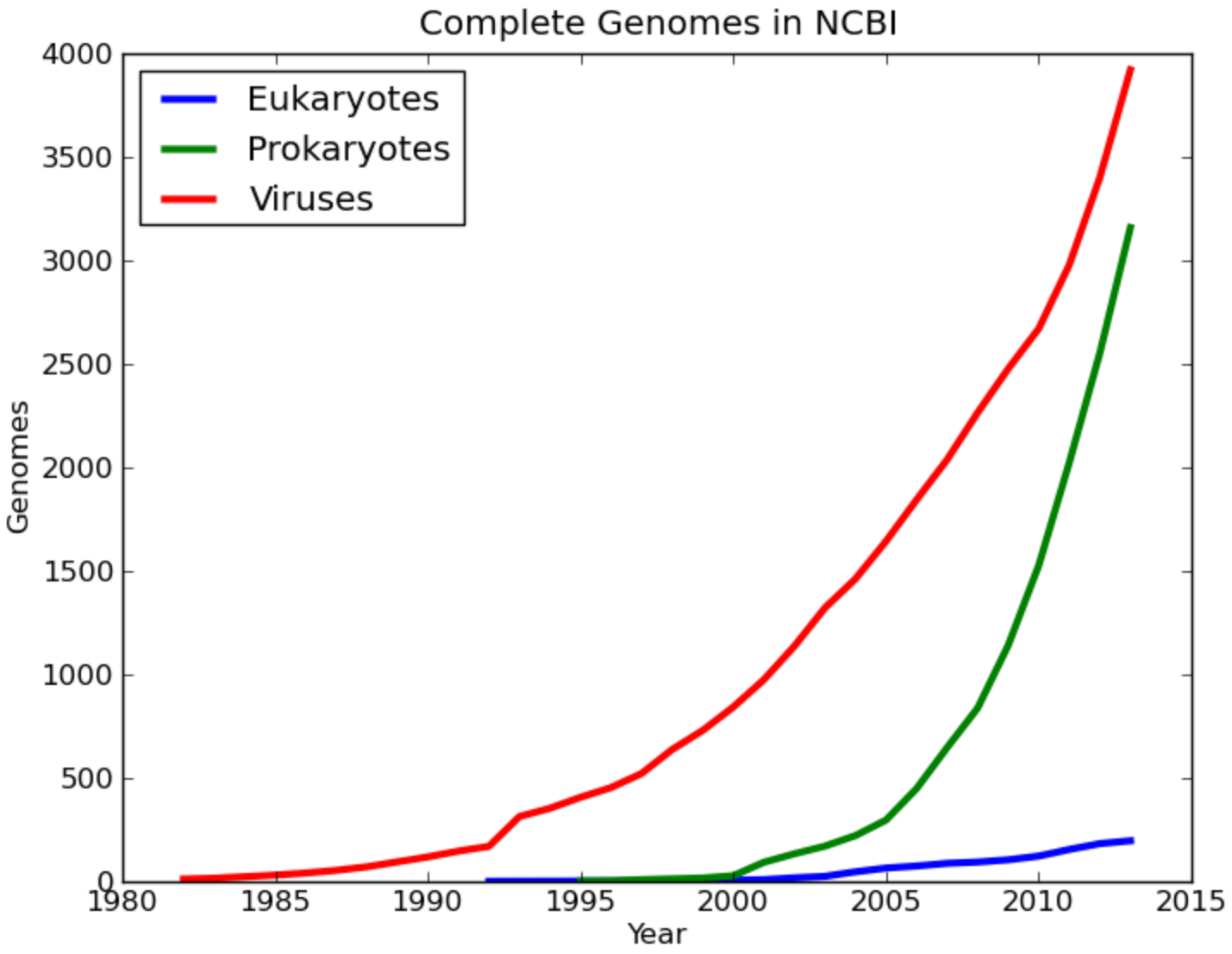
Πρωτεΐνες = 10 x C (?)

3. Επινόηση στην αρχιτεκτονική οργάνωση περιοχών (domains) των πρωτεϊνών

GenBank and WGS Statistics






Complete Genomes in NCBI
Eukaryotes
Prokaryotes
Viruses
Genomes
Year
0
500
1000
1500
2000
2500
3000
3500
4000
1980
1985
1990
1995
2000
2005
2010
2015

https://www.ensembl.org/index.html

Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Tools

All tools

BioMart >

Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >

Search our genomes for your DNA or protein sequence

Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes

-- Select a species --

- View full list of all Ensembl species
- Edit your favourites

Favourite genomes

Human
GRCh38.p12

Still using GRCh37?

Mouse
GRCm38.p6

Zebrafish
GRCz11

Laurasiatheria
- Alpaca
- American black bear
- Cat
- Cow
- Dingo
- Dog
- Dolphin
- Donkey
- Ferret
- Goat
- Hedgehog
- Horse
- Leopard
- Megabat
- Microbat
- Panda
- Pig
- Polar bear
- Red fox

-- Select a species --

Bacteria
(44,000 genomes)

Vertebrates
(154 genomes)

Fungi
(1,000 genomes)

e!

Protists
(189 genomes)

Plants
(57 genomes)

Metazoa
(71 genomes)

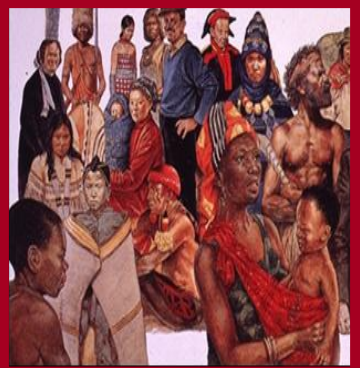
ΓΕΝΕΤΙΚΗ ΠΟΙΚΙΛΟΤΗΤΑ



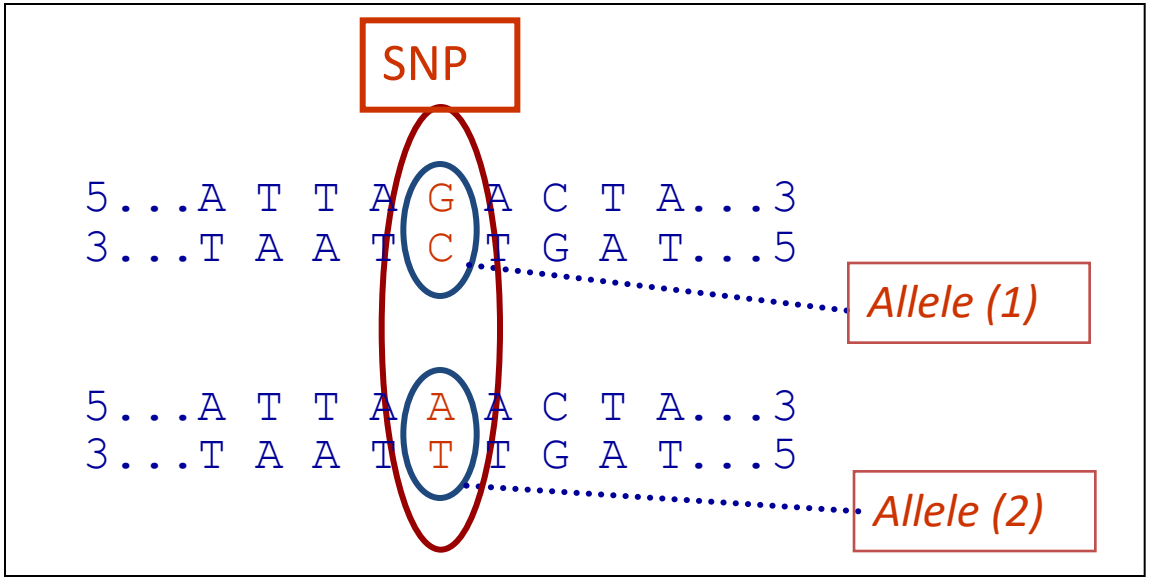








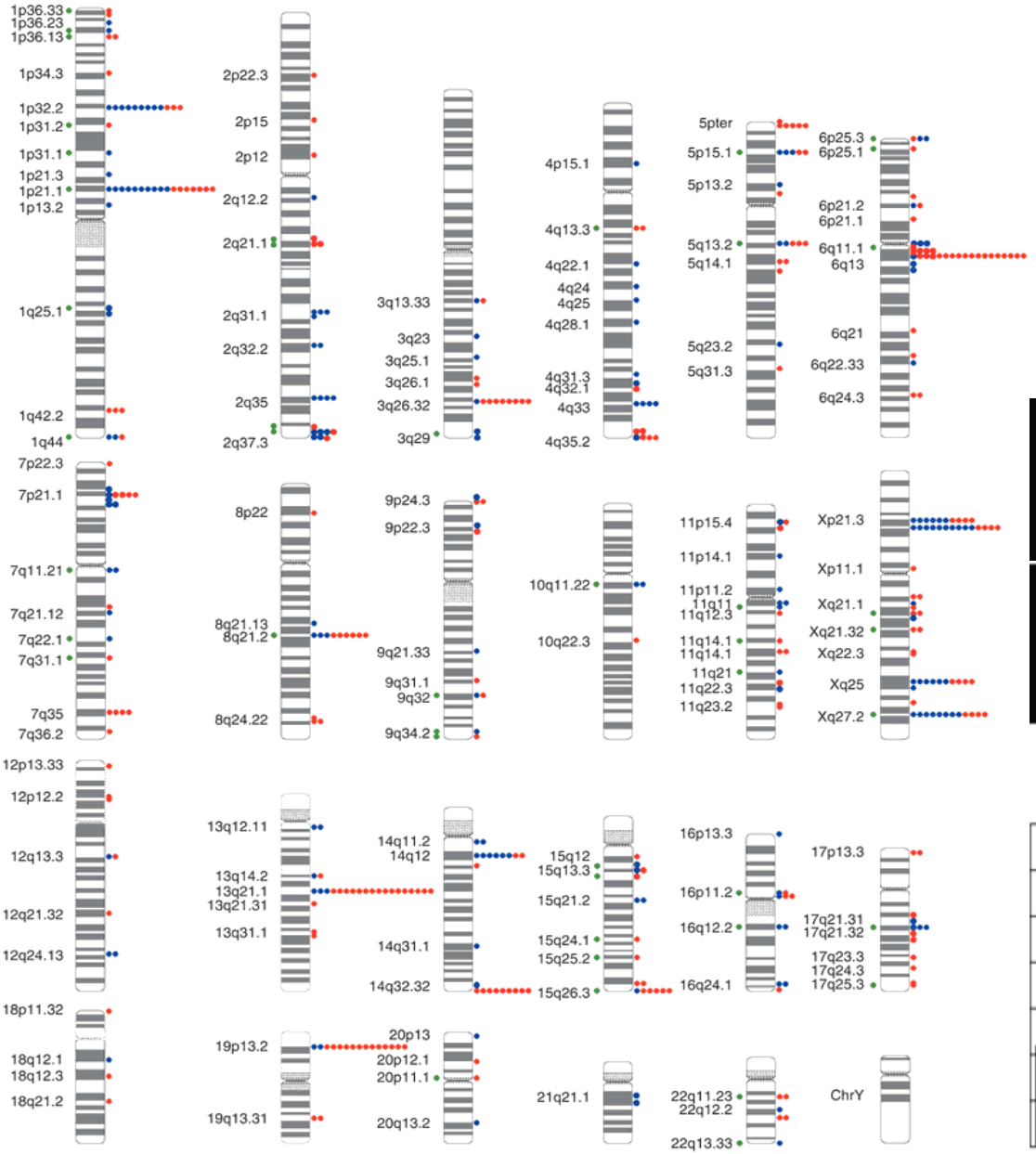

SNPs (single nucleotide polymorphisms)



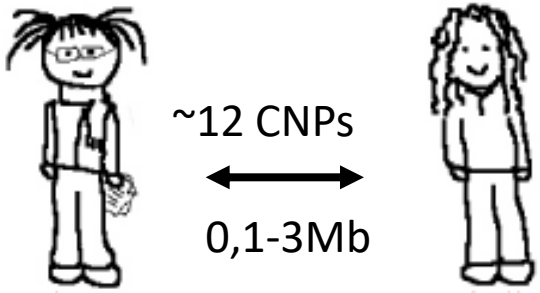
~10 εκατομμύρια SNPs στο ανθρώπινο γονιδίωμα

~ 500,000 tag SNPs

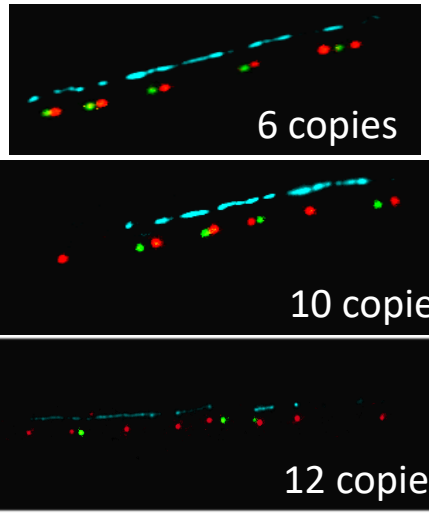
CNPs (copy number variation)



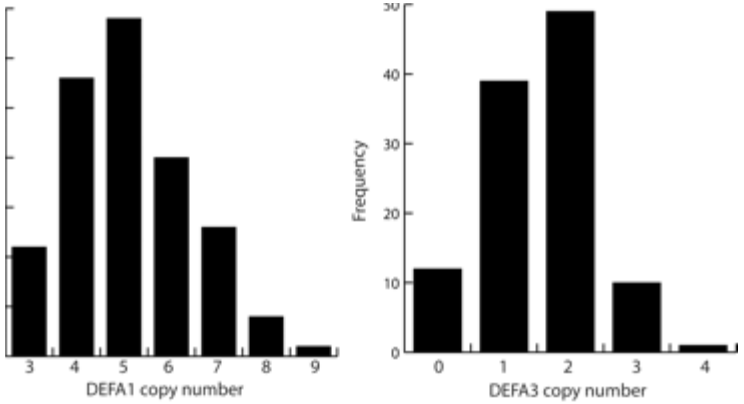
Iafrate et al Nat Genet 2004, Aldred et al Human Mol Genet 2005



AMYLASE GENES



DEFENSIN GENES



Cow assembly and gene annotation

Statistics

Summary

Assembly	ARS-UCD1.2, INSDC Assembly GCA_002263795.2, Apr 2018
Base Pairs	2,715,837,454
Golden Path Length	2,715,837,454
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Sep 2018
Genebuild released	Dec 2018
Genebuild last updated/patched	Oct 2018
Database version	95.12

Gene counts

Coding genes	21,867
Non coding genes	5,211
Small non coding genes	3,351
Long non coding genes	1,488
Misc non coding genes	372
Pseudogenes	492
Gene transcripts	43,947

Other

Genscan gene predictions	46,441
Short Variants	98,843,582
Structural variants	18,942

Goat assembly and gene annotation

Statistics

Summary

Assembly	ARS1, INSDC Assembly GCA_001704415.1, Aug 2016
Base Pairs	2,922,813,246
Golden Path Length	2,922,813,246
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Sep 2017
Genebuild released	Mar 2018
Genebuild last updated/patched	Mar 2018
Database version	95.1

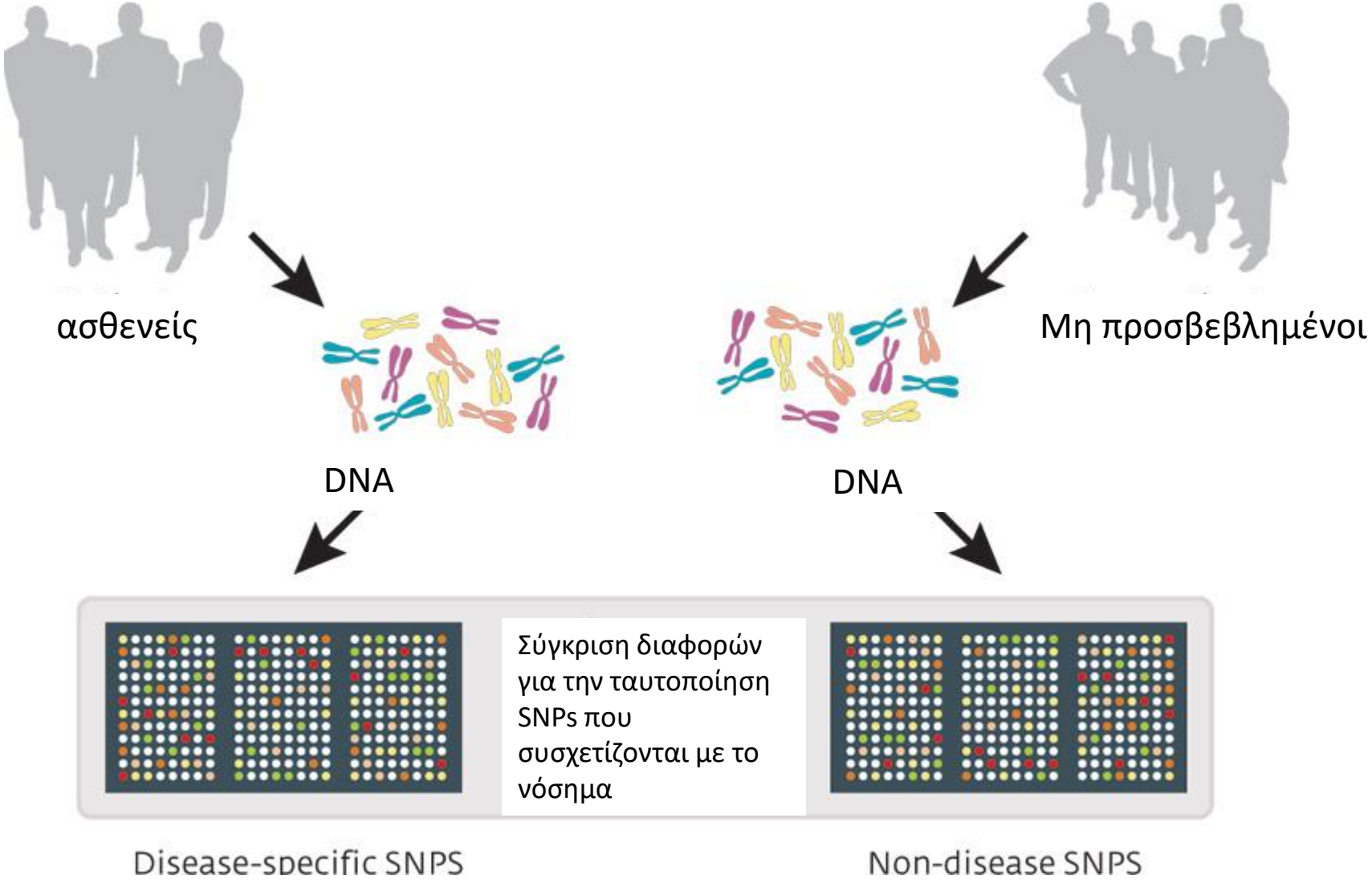
Gene counts

Coding genes	21,361
Non coding genes	5,688
Small non coding genes	2,623
Long non coding genes	2,713
Misc non coding genes	352
Pseudogenes	222
Gene transcripts	41,794

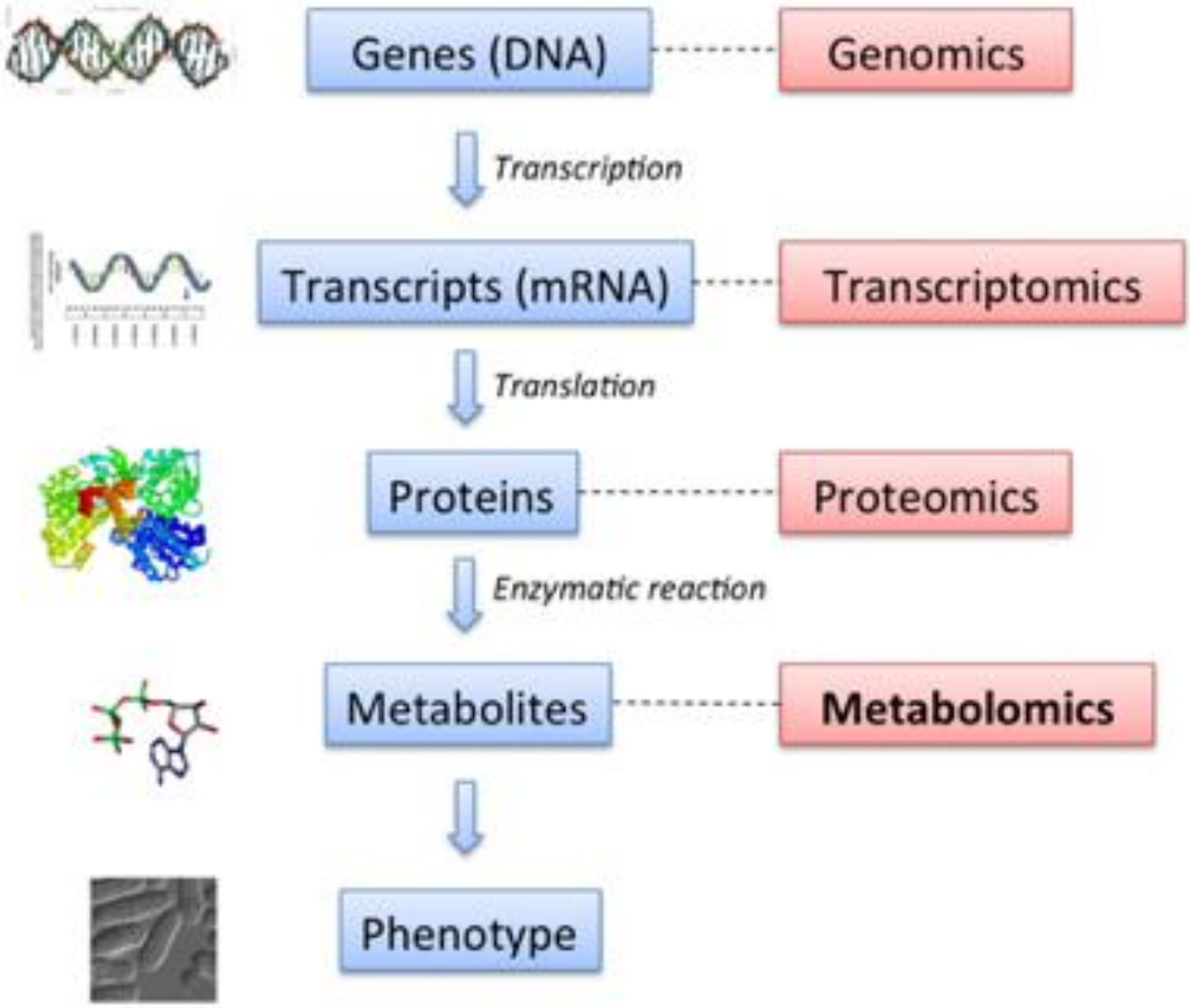
Other

Genscan gene predictions	58,693
Short Variants	34,116,678

Ανάλυση συσχέτισης σε επίπεδο γονιδιώματος (GWAS)



Επίπεδα βιολογικής πληροφορίας

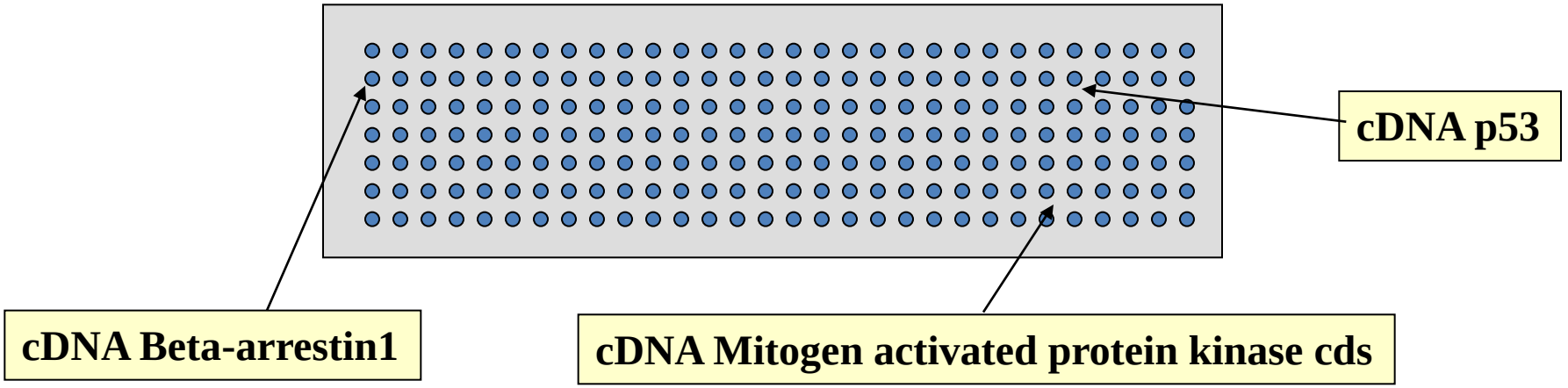


Μεταγραφωματική (Transcriptomics)

Ταυτόχρονη μελέτη πολλών/όλων των διαφορετικών mRNAs

Μικροσυστοιχίες

- **Τι είναι μία συστοιχία;**

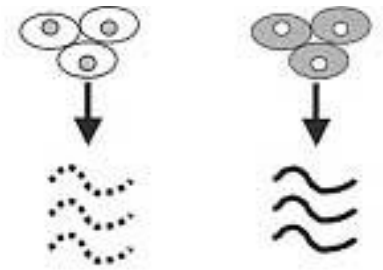


Ή συνθετικά ολιγονουκλεοτίδια (>1 για κάθε γονίδιο)

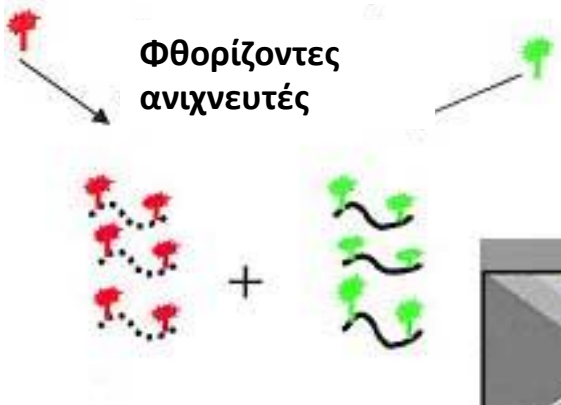
Απομόνωση RNA

Δείγμα α

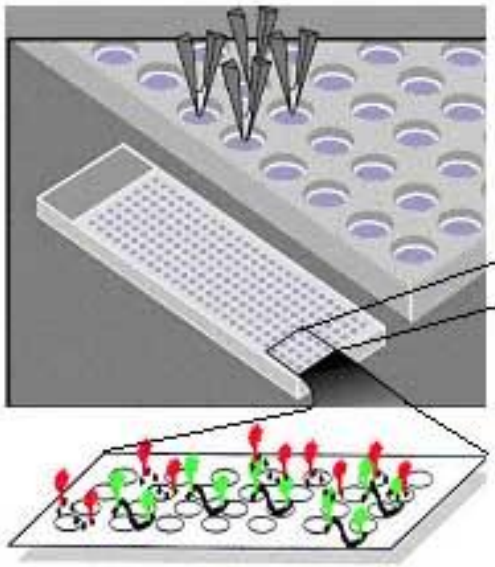
Δείγμα β



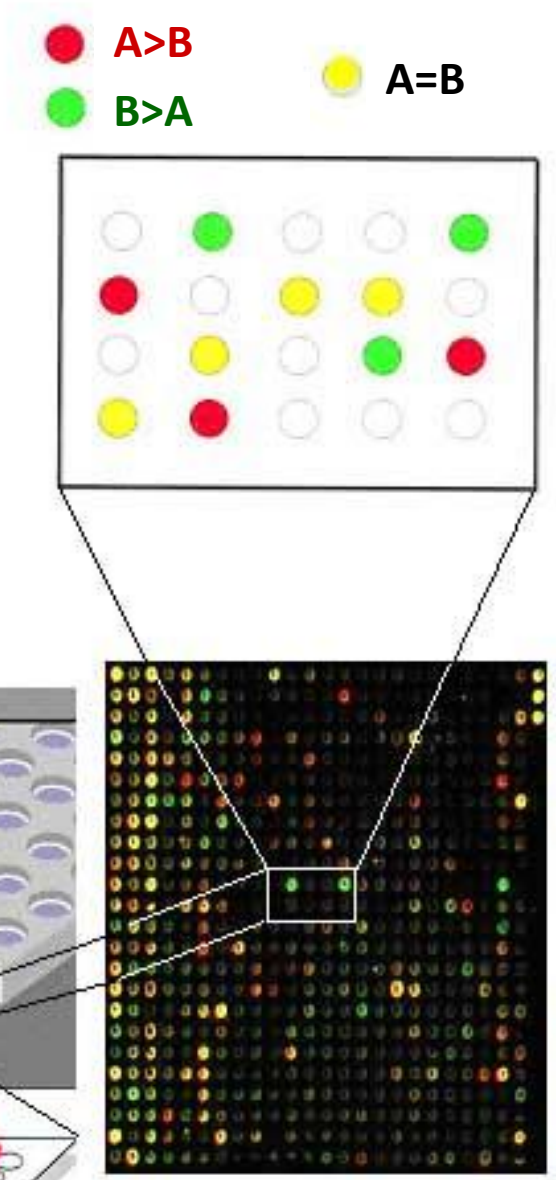
Δημιουργία cDNA
Σήμανση του ανιχνευτή



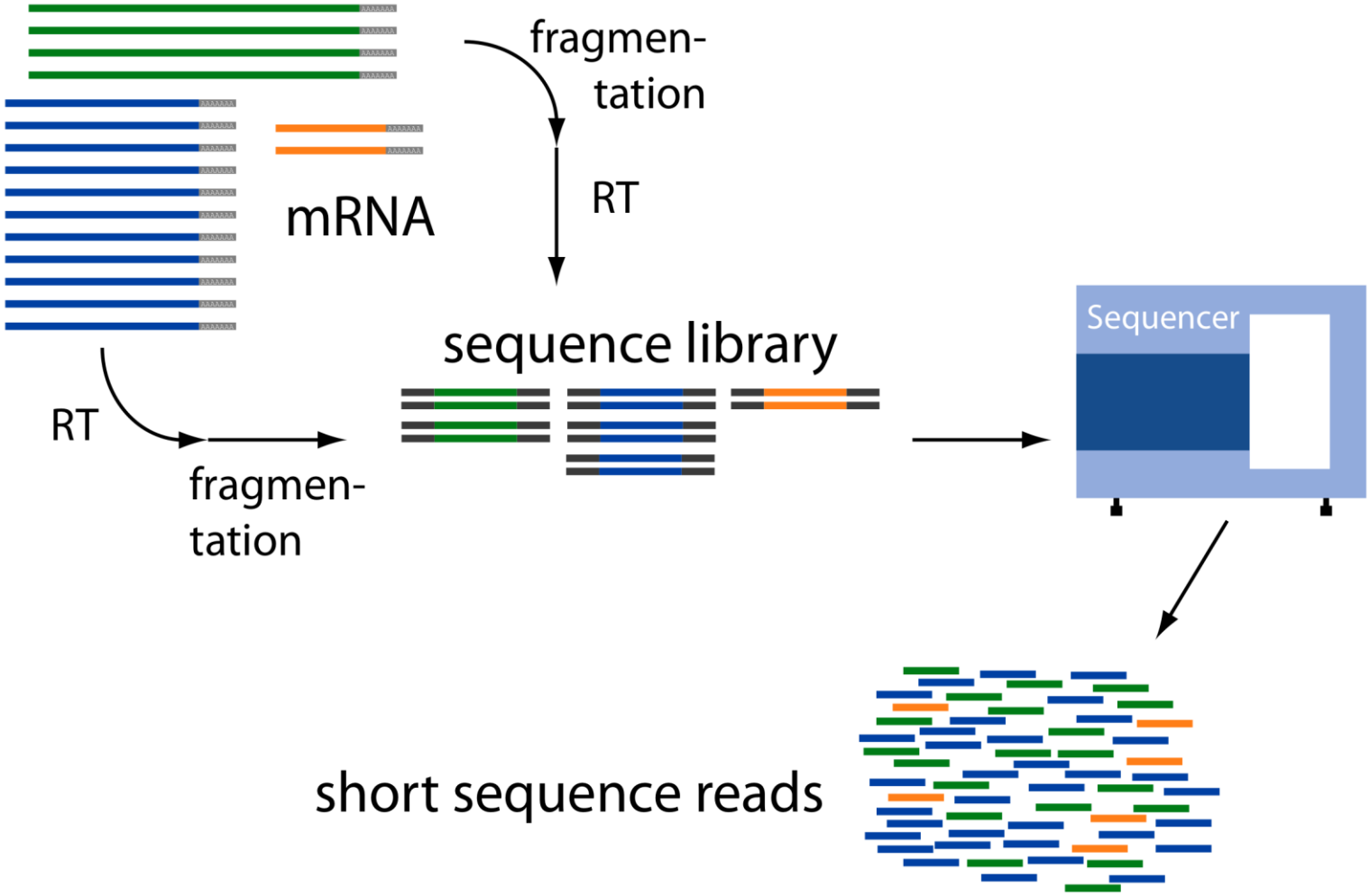
Υβριδοποίηση με την συστοιχία



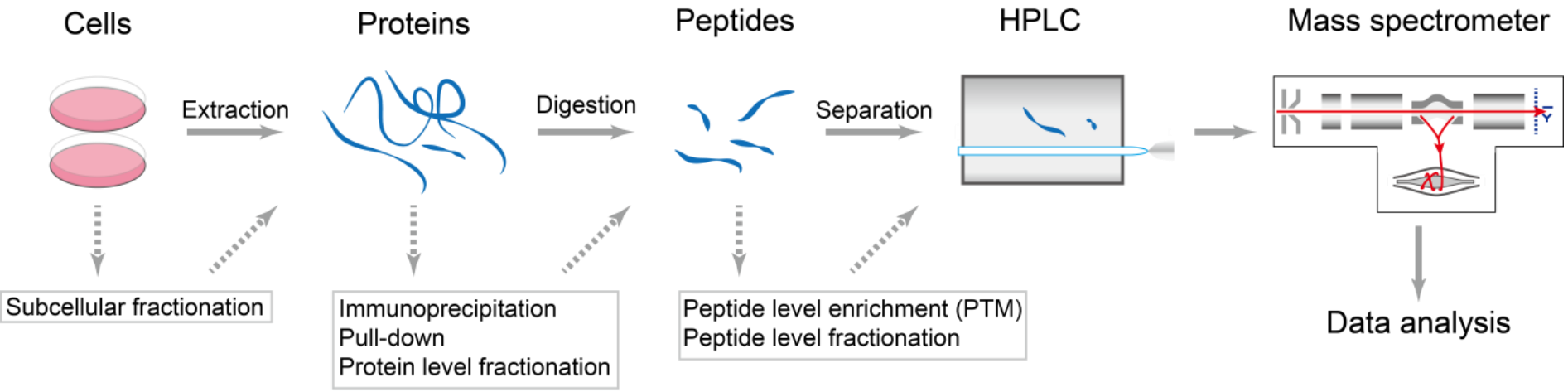
Απεικόνιση



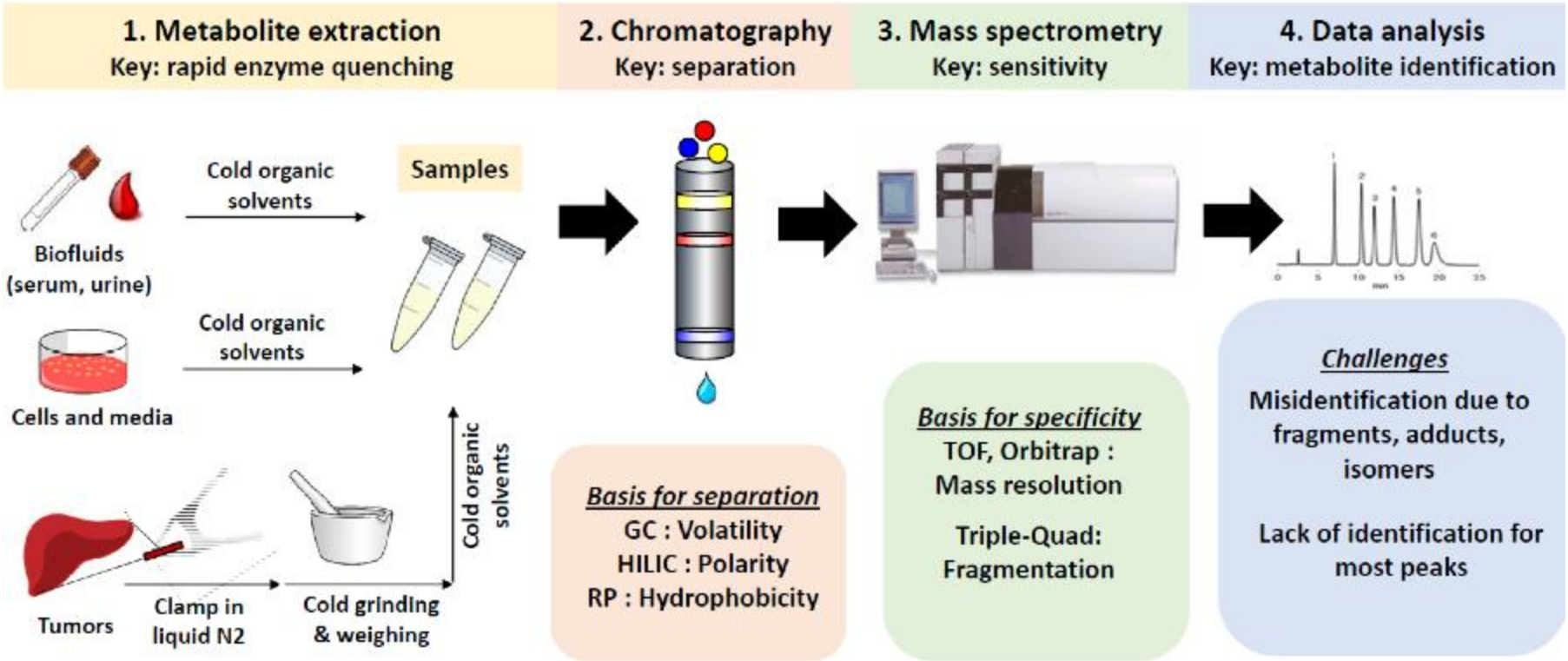
Αλληλούχηση RNA

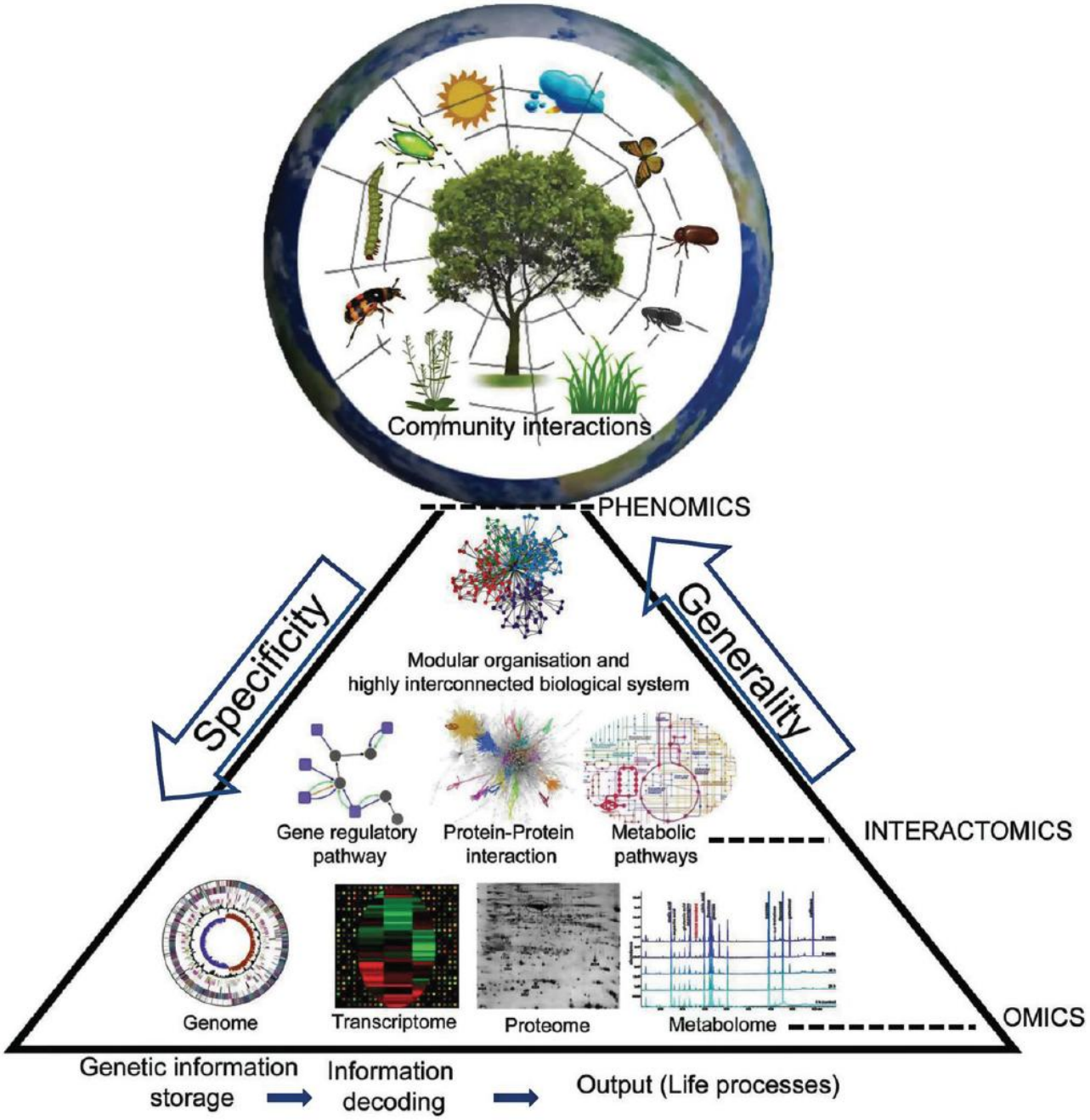


Πρωτεϊνωματική (Proteomics)



Μεταβολομική (Metabolomics)




Community interactions
PHENOMICS
Specificity
Generality
Modular organisation and
highly interconnected biological system
Gene regulatory
pathway
Protein-Protein
interaction
Metabolic
pathways
INTERACTOMICS
Genome
Transcriptome
Proteome
Metabolome
OMICS
Genetic information
storage
Information
decoding
Output (Life processes)