



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

ΠΡΟΓΡΑΜΜΑ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ ΑΕΙ ΓΙΑ ΤΗΝ ΕΠΙΚΑΙΡΟΠΟΙΗΣΗ ΓΝΩΣΕΩΝ ΑΠΟΦΟΙΤΩΝ ΑΕΙ (ΠΕΓΑ)

«Οι σύγχρονες τεχνικές βιο-ανάλυσης στην υγεία, τη γεωργία, το περιβάλλον και τη διατροφή»

**ΠΡΟΓΡΑΜΜΑ ΧΑΡΤΟΓΡΑΦΗΣΗΣ
ΤΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ ΤΟΥ
ΑΝΘΡΩΠΟΥ**

HUMAN GENOME PROJECT (HGP)

Ιστορική αναδρομή

Μάιος 1985



Robert Sinsheimer
Πρύτανης του Παν. της
Καλιφόρνιας, Santa Cruz



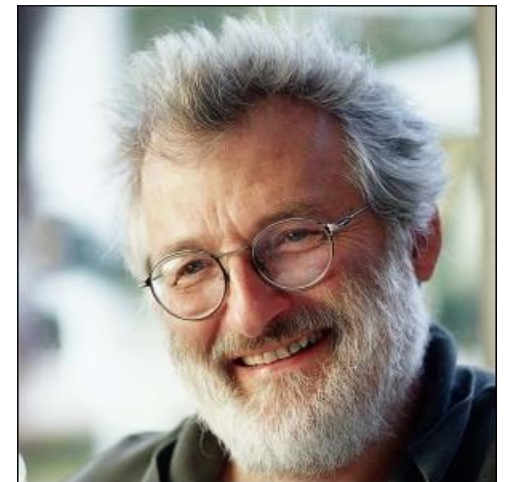
David Botstein



Walter Gilbert

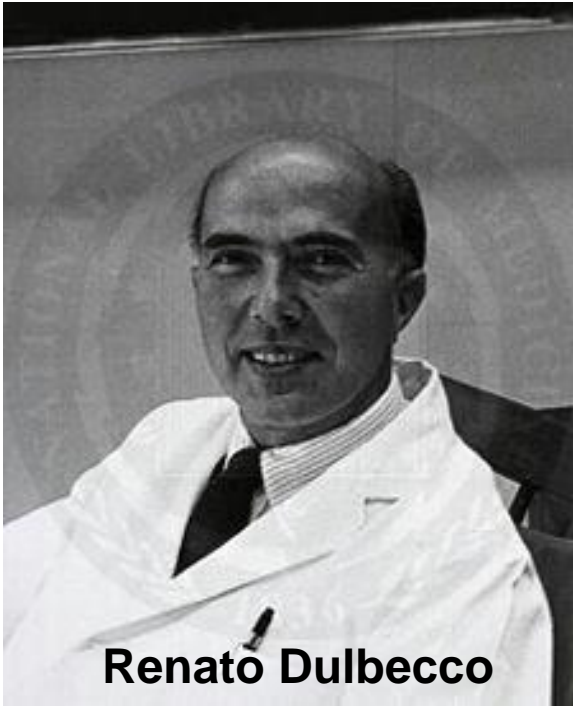


Lee Hood



John Sulston

Ιστορική αναδρομή



Renato Dulbecco
Nobel Prize, 1975, in
Physiology-Medicine

Σεπτέμβριος 1985

Ομιλία στο Cold Spring Harbour

Perspective

A Turning Point in Cancer Research: Sequencing the Human Genome

RENATO DULBECCO

ONE OF THE GOALS OF CANCER RESEARCH IS TO ASCERTAIN the mechanisms of cancer. Efforts in this direction have been made by using model systems of limited complexity, such as cancer cells in vitro and oncogenic viruses. The use of cell cultures avoided the complexity of the whole animal but not the

7 MARCH 1986

The author is in the Monoclonal Antibody Laboratory of the Armand Hammer Cancer Center, the Salk Institute, La Jolla, CA 92037.

Ιστορική αναδρομή



Μάρτιος 1986

Santa Fe meeting

Εκτίμηση του κόστους, της ωφελιμότητας, της δυνατότητας επίτευξης και του χρόνου του προγράμματος



3 φάσεις

Ανάπτυξη τεχνολογίας

Χαρτογράφηση

Αλληλούχηση

Charles De Lisi

**Διευθυντής του Τμήματος
Έρευνας για την Υγεία & το
Περιβάλλον στο Υπουργείο
Ενέργειας**

Ιστορική αναδρομή

1986: «Molecular Biology of Homo Sapiens»
Cold Spring Harbor



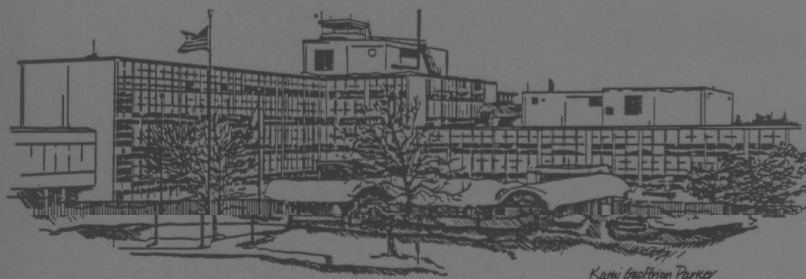
Οργανωτής: James Watson



Department of Energy
Office of Health and Environmental Research

SEQUENCING THE HUMAN GENOME

Summary Report of the Santa Fe Workshop
March 3-4, 1986



Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

Los Alamos National Laboratory is operated by the University of California for the
United States Department of Energy under contract W-7405-ENG-36.

OPINION MEETINGS THAT CHANGED THE WORLD



Santa Fe 1986: Human genome baby-steps

The 1980s saw plenty of discussion on sequencing the human genome. But, according to Charles DeLisi, one conference was crucial for converting an idea to reality.

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



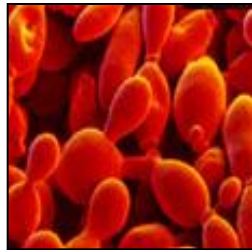
J. Watson

Διευθυντής στο National Human Genome Research Institute

Οργανισμοί-μοντέλα



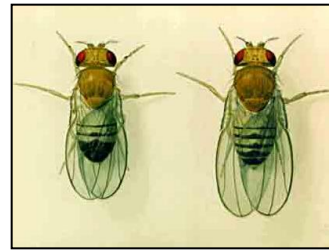
E. coli



S. cerevisiae



C. elegans



D. melanogaster



M. musculus

WORKSHOP ON INTERNATIONAL COOPERATION FOR THE HUMAN GENOME PROJECT

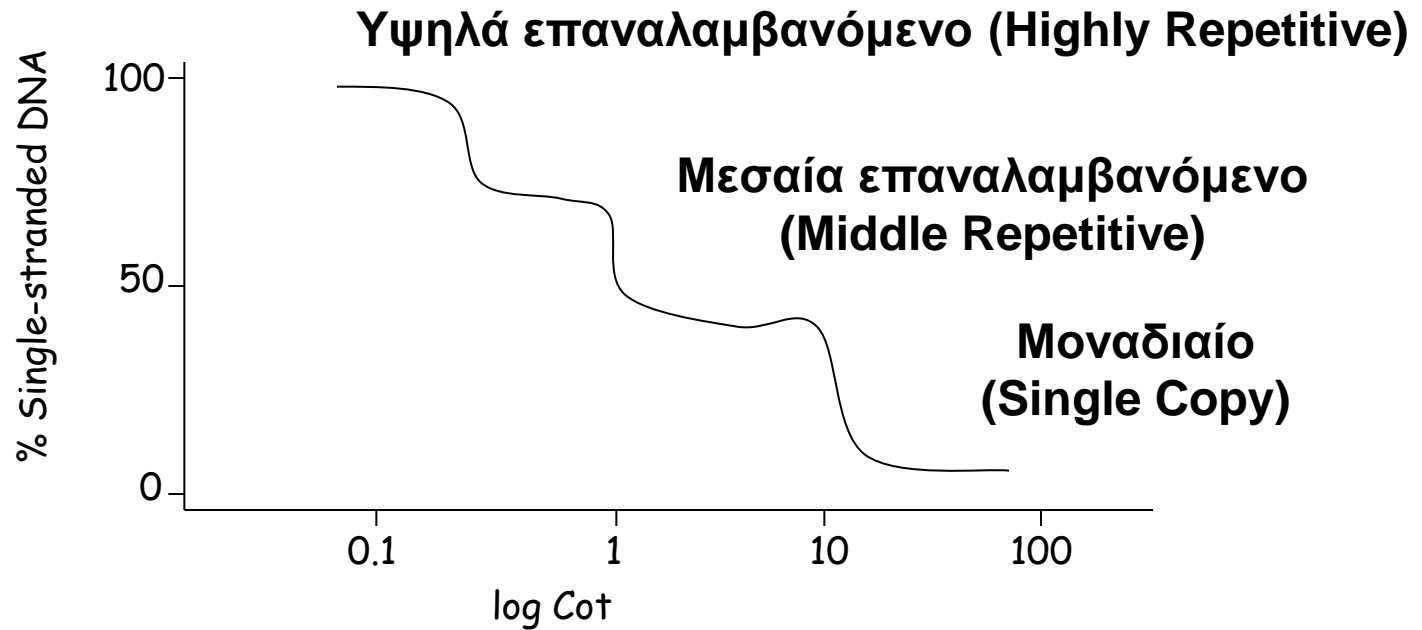
VALENCIA DECLARATION ON THE HUMAN GENOME PROJECT

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)

Γιατί να αλληλουχηθούν και οι επαναλαμβανόμενες αλληλουχίες;

Πολυκύτταροι ευκαρυωτικοί

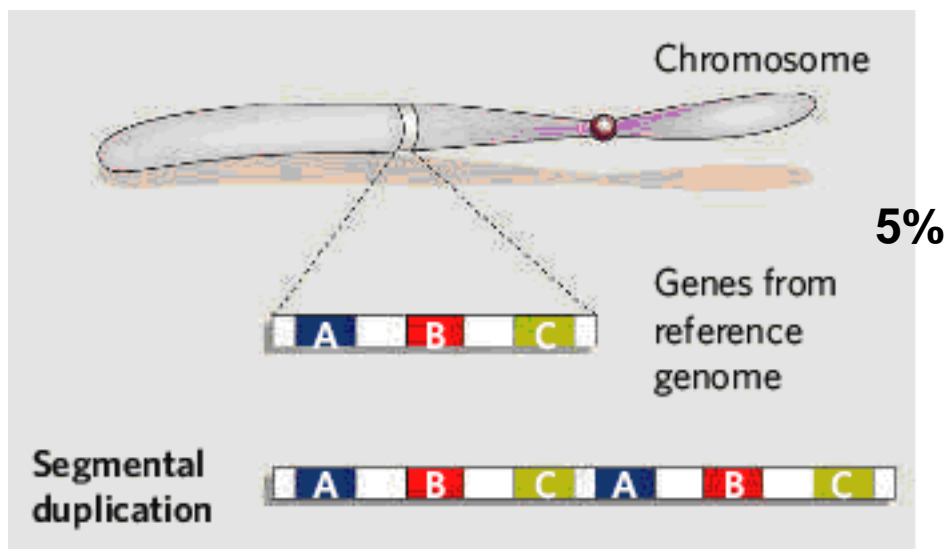


Επαναλαμβανόμενες αλληλουχίες του ανθρώπινου γονιδιώματος

	Μήκος	Αριθμός αντιγράφων	% του γονιδιώματος
LINES: <u>L</u> ong <u>I</u> nterspersed <u>E</u> lements	6-8kb	850.000	21%

SINES: <u>S</u> ong <u>I</u> nterspersed <u>E</u> lements (Alu repeats...)	100-300bp	1.500.000	13%
---	-----------	-----------	-----

Segmental duplications
(διπλασιασμοί χρ. τμημάτων)
(> 90% ομοιότητα, >1kb μήκος).



Αλληλούχηση μόνο των cDNAs;



Sydney Brenner

1990



Craig Venter

1991 (NIH)

**~600 ESTs
>50% νέα
Πατέντα???**

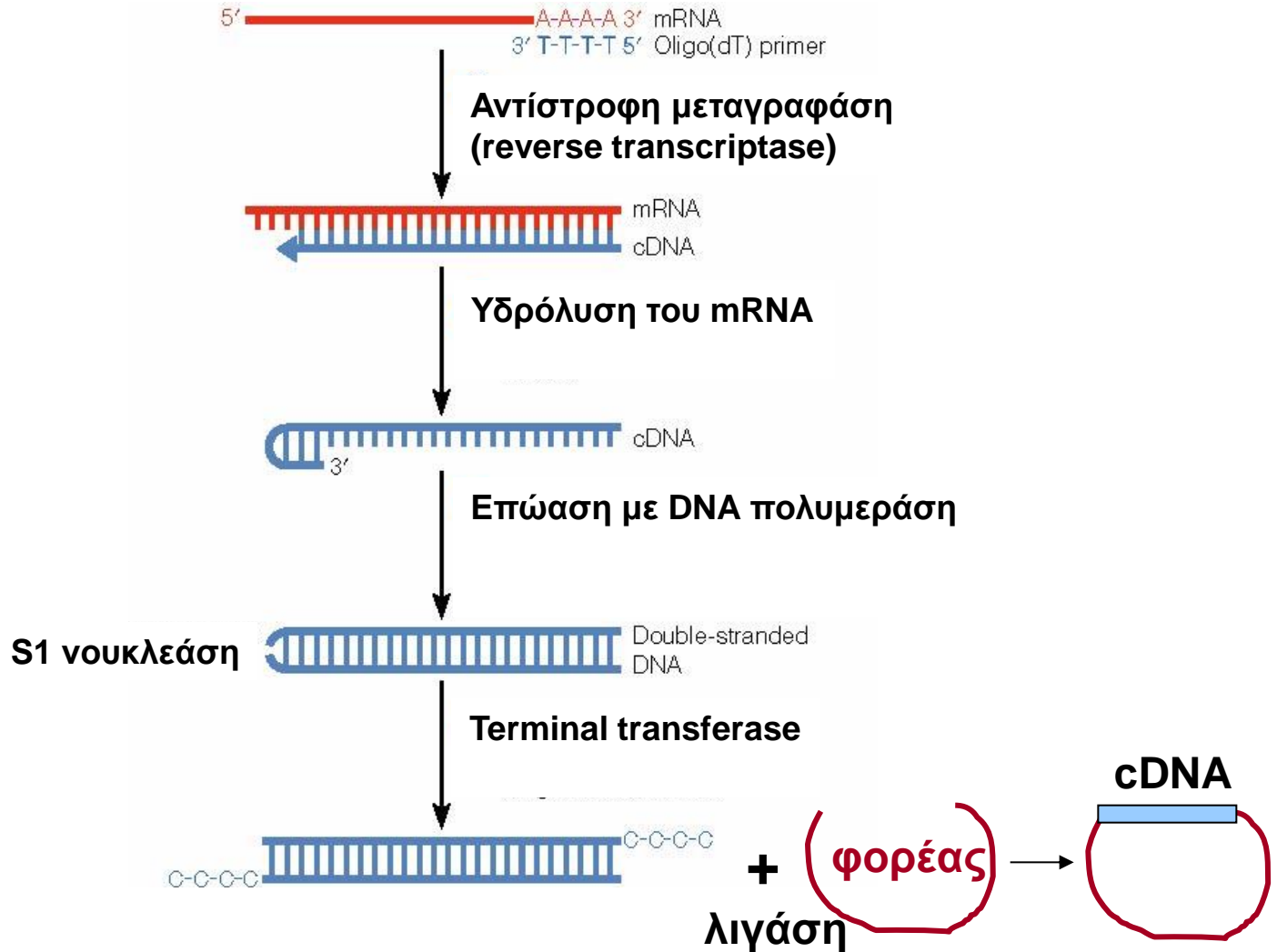
**TIGR
ΠΑΤΕΝΤΑ**

**GSC (Genome Sequencing Center, U. of Washington)
Ελεύθερα στις db**

* ΒΙΒΛΙΟΘΗΚΗ cDNA

Ιστός π.χ.
εγκέφαλος

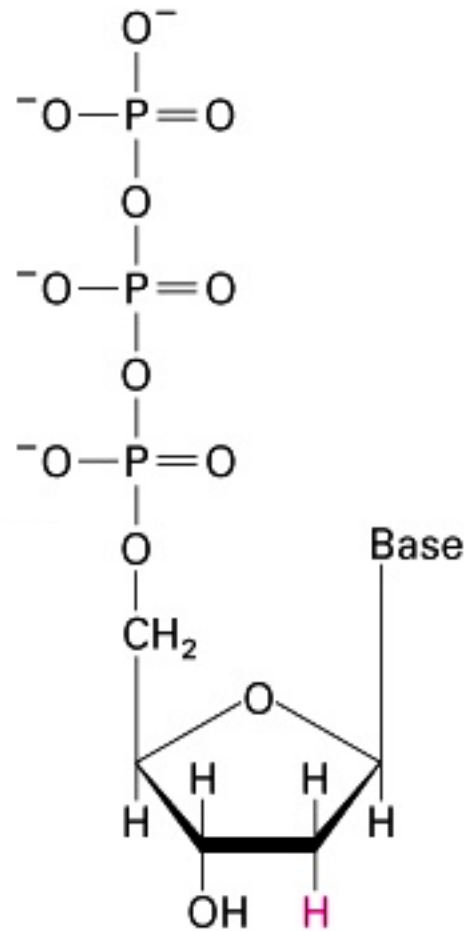
→ Απομόνωση mRNAs



Αλληλούχηση

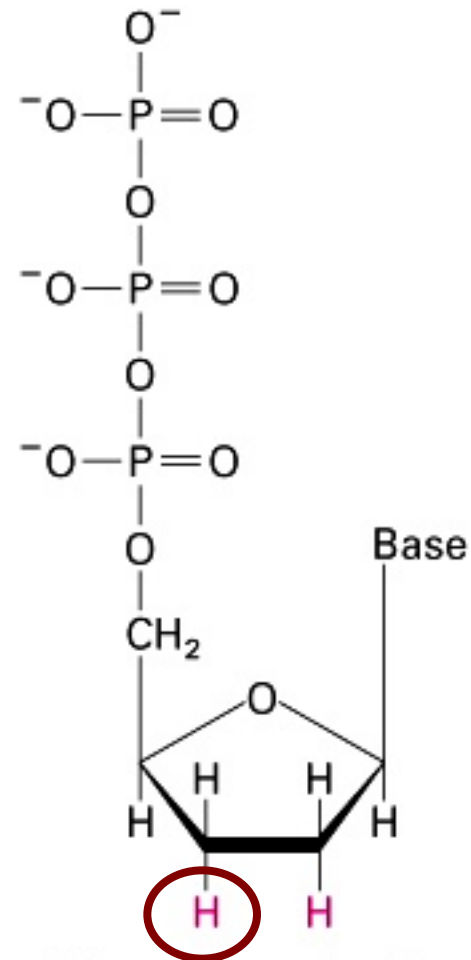


Fred Sanger
13 August, 1918



Deoxyribonucleoside triphosphate

dNTP

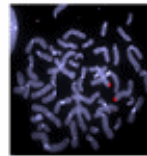


Dideoxyribonucleoside triphosphate

ddNTP

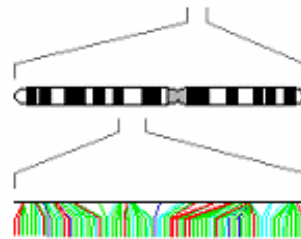
STRATEGIES FOR SEQUENCING THE HUMAN GENOME

BY MAPPED CLONES

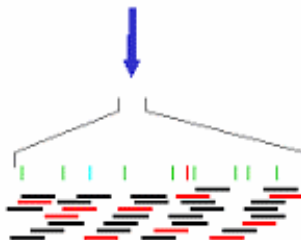


BY WHOLE GENOME SHOTGUN

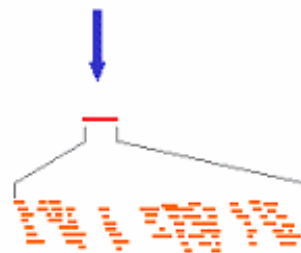
1. Construction of maps of ordered landmarks (genetic markers, genes): provides long-range map and organisation into individual chromosomes.



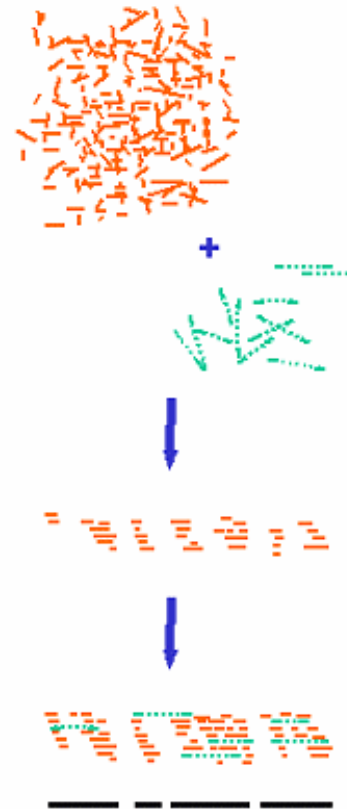
2. Physical maps of overlapping clones anchored to the landmark maps.



3. Selection of tile path (clones in red)



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



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.

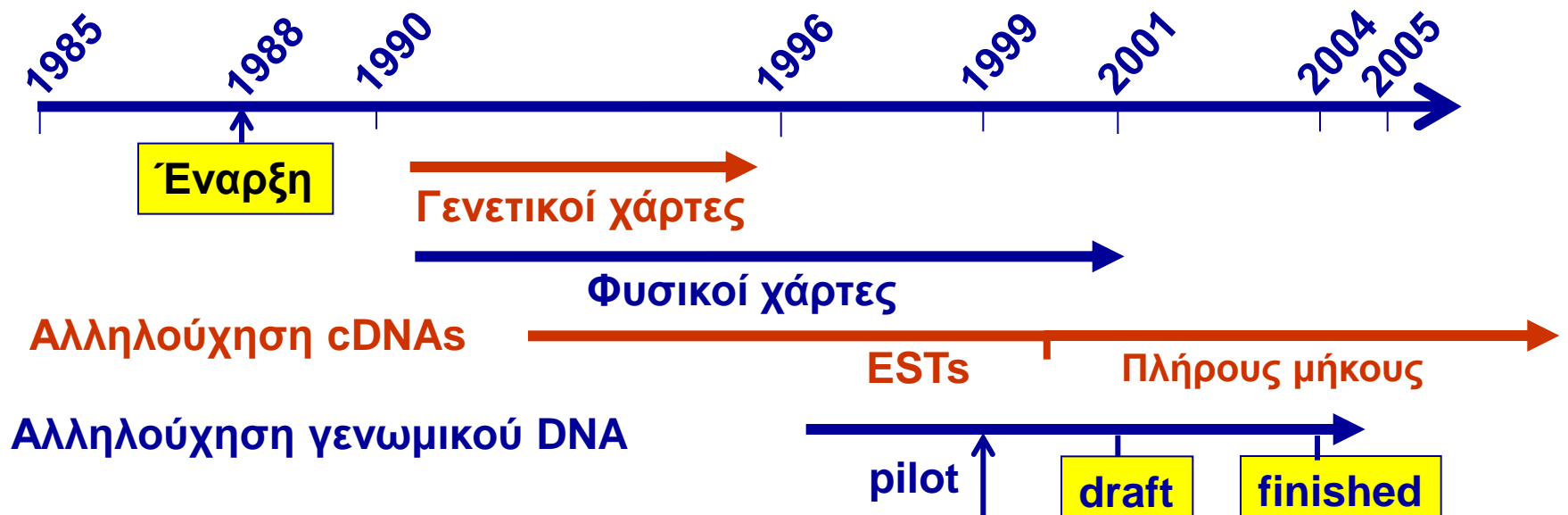


J. Graig Venter



The automated production line for sample preparation at the Whitehead Institute, Center for Genome Research.

ΠΟΡΕΙΑ ΤΟΥ HGP



chr. 22
Κάλυψη: 4x - 5x
90% του γονιδιώματος

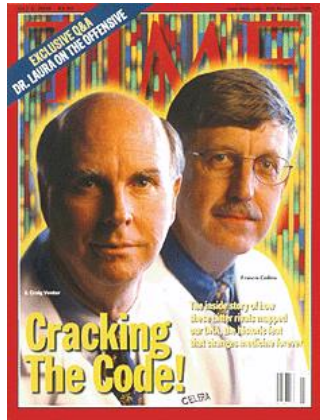
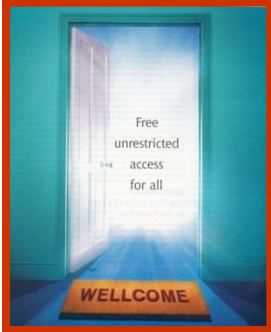


Κάλυψη: 8x - 9x,
99% του γονιδιώματος
Ακρίβεια: 99,9%

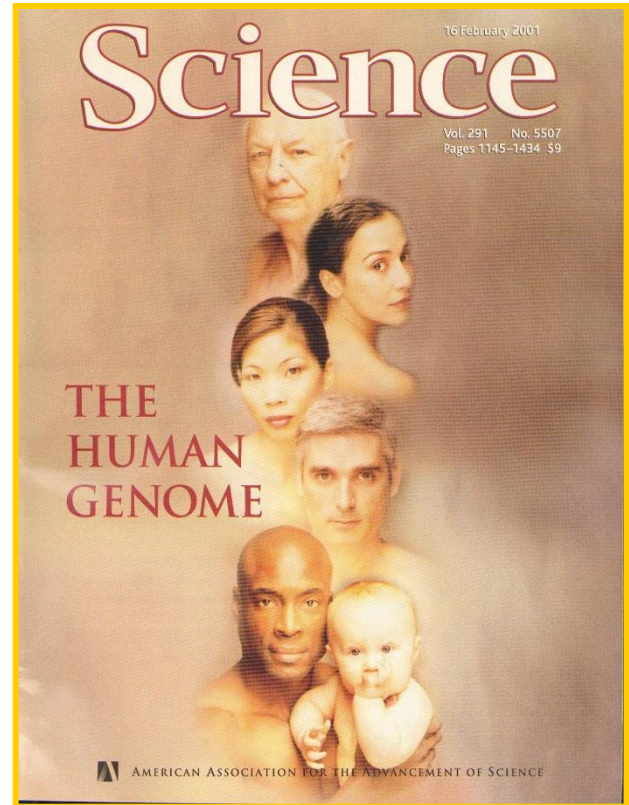
Vol 409, Feb 15, 2001



**Human Genome Project
HGP
FREE ACCESS**



Vol 291, Feb 16, 2001



**CELERA GENOMICS
Access to db upon
subscription**



Τώρα για πρώτη φορά διαθέτουμε μια ιστορική ανθολογία του εαυτού μας, ένα μέρος της οποίας μεταβιβάζεται από γενιά σε γενιά επί ένα δισεκατομμύριο χρόνια. Μόλις τώρα μάθαμε να διαβάζουμε την ιστορία και είναι σίγουρο πως θα μας συναρπάσει για πολλές ακόμα δεκαετίες.
Eric Lander, Whitehead Institute

Θα παραξενεύουν πολύ αν η θεραπεία του καρκίνου δεν μεταμορφωνόταν μέσα στα επόμενα 20 χρόνια
Mike Stratton, Cancer Genome Project

Το γονίδιο είναι, με διαφορά, το πιο περίτεχνο πρόγραμμα
Bill Gates, Microsoft

Είναι το ανεπανάληπτο επίτευγμα, όχι μόνο του καιρού μας αλλά ολόκληρης της ανθρώπινης ιστορίας. Αυτό το λέω επειδή το Human Genome Project, δυνητικά μπορεί να επηρεάσει τη ζωή κάθε ανθρώπου στον πλανήτη μας.
Michael Dexter, Wellcome Trust

AGGTTAGATTATGCCCCGAGGGCGCCCCAGCCGAAATTTTTTAATGCAGGTTTAATAGTTTAGAGC
CTGTGGGCTTCCATGGCTTGGTTCTGCTGTTCTTCACTGGGGACTTGGGGGACCCTGGGAGCTTC
TGATGGGGCCTGTCTCCACCTCTGTAAATCCAAGGAGTCAGATGACAAATCTGTCATTTTCGGGCC
ACACACTCCCCTGAGGAAAGGGCCTTGCAGGAGGGCAGAGCAGCTTGCTGGGCATGGCAGGGAGT
GGAGAAGGGCAGGGGGCGCAGAGCAGGAGCAGCTTCCTGCCTCTGGGTGGGGACAGTGATCCCCA
CTGGGGACTGGCAAAGCCCCATGCTCTCTGTTCAACCCTGGATGGGTGGCACCTGGGGGCAGGCAT
GGGGCCTGCAGGAGCCCCTGTGTGCCAGCCCTCCCCTGCCAGCATCCCATCTCCCAGGAGGCCCC
CAGGGCAGGTAAGTGCCACCTCCCGCCCTCAGCCCTCAGCCCTTCCTTGGACGAACGCCTCC
CACTCCCGGAACCACTGCACTCCCGCCCTCAGCCCTCAGCCCTTCCTTGGACGAACGCCTCC
CTTGTCAACTGGCAGGCCCTCCCGCCCTCAGCCCTCAGCCCTTCCTTGGACGAACGCCTCC
TCAGCAAGGATTTCCCAAATCCCGCCCTCAGCCCTCAGCCCTTCCTTGGACGAACGCCTCC
ATACTCAAGGTGAAAAATCCTCCCGCCCTCAGCCCTCAGCCCTTCCTTGGACGAACGCCTCC
AGAGGACGGTGGTAACATTCCTCCCGCCCTCAGCCCTCAGCCCTTCCTTGGACGAACGCCTCC
GAGATTCTACTTTTTGAGATTCCTCCCGCCCTCAGCCCTCAGCCCTTCCTTGGACGAACGCCTCC
CGGCTTACTGCAACCTCTCCCGCCCTCAGCCCTCAGCCCTTCCTTGGACGAACGCCTCC
GGGATTACAGGTGTGTGCCCTCCCGCCCTCAGCCCTCAGCCCTTCCTTGGACGAACGCCTCC
ATGTTGGCCAGGCTGGTCTTGAACCTCCTGACCTCAGGTGATCCGCCACCTCGACCTCTCAAAG
TGCTGGGATAACAGGCATGAACCACTGCGCCCGGCCTGGGAGATGCTAATTTTCTCCGGTTGAAT
AGAATGTGCCTATCTGCTCAGAGAGGCAGCTCTCCTTCTGACAGGAGCATTTTCTTTTTTCGAGAT
GGGGGGTGGTCTCACTCTGTGCCAGGCGGGAGTGCAGCGGCGCAATCATGGCTCACTGCAGCCT
TGACCTCCTGGGCTCAAGTGATCCTCCCACCTCAGCCTCCTGAGTAGGTTGGACCACAGGTGCAI
ACCACTAGGCCAGCCCTGACAGTCTCTTTTTCGTTTTGTGTTCTGAGACAGGGTCTCACTCTATI
GCCAGGCTGCGGTGCAGTGGCATGATCACGGCTCACTGCAGCCTCAACCTCCCAGGCTTAGGTC
ATCCTCCCAACTCACTCAGCCCTCCAGGTAGCGGGACTACAGGTACACATCACCATGCCTGGCT
AATTTTTGTATTGTTTTGTAGAGATGGGGTTTTCGCCATGTTGGCCAAGTTGGTCTTGAACCTCCTGC

1990

50.000-100.000 bp/έτος

2002

10.000 bp/sec (24h, 7 ημέρες)

Κόστος:

1990: \$ 10 /bp = \$ 3 δις

2002: \$ 0,1 /bp = \$ 3 εκ.

TCCTCGGCCTGGAAGGC
TCAGTCTCTATTTACCT
TCTTACAAAACACCAG
ACACACCCATCACTCAG
CAACTGAGACCACAAGC
GTGCAGTGGCGCGATCT
CAGCCTCTTGAGTAGCT
AGAGACGGGGTTCGCC



ONCE YOU UNFOLD
ONE OF THESE THINGS,
IT'S NEVER THE SAME.

Joe Heller
625 N BAY PRESS LN 2
JOE@HELLERBOY.COM

MAP OF THE HUMAN GENOME

Ethical
Questions

Medical
Dilemmas

Legal
Tangles

Privacy
concerns

Moral
Issues

Insurance
Applications

© 1997 Joe Heller

PHASE TWO: INTERPRETATION

SEDMAN with Laizer

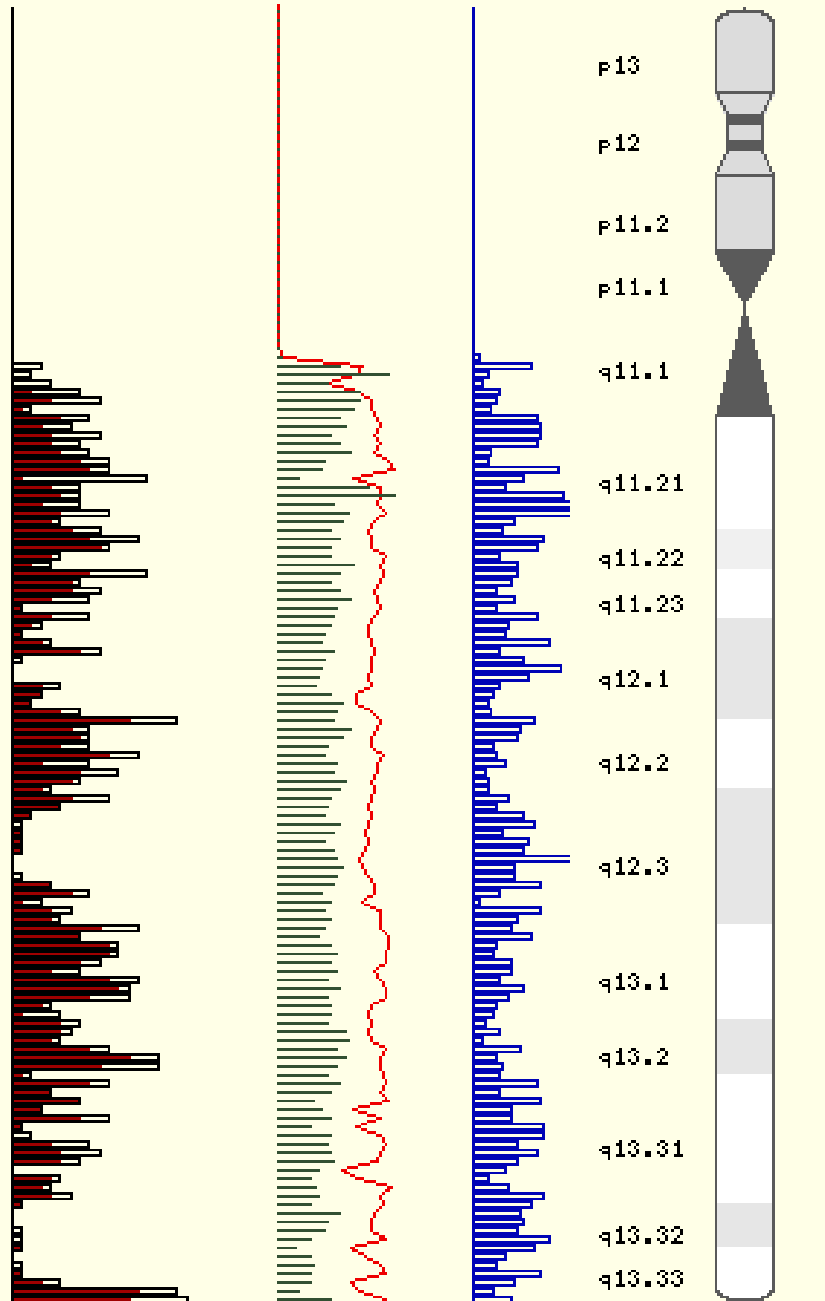


Known Genes
Total Genes

% GC
Repeats

SNPs

Chromosome 22



**Ανάλυση γενωμικών
αλληλουχιών
για την ταυτοποίηση νέων
γονιδίων, *in silico***

<http://www.ensembl.org>

Chr 10



Chromosome band

q23.1

84.26 Mb

85.26 Mb

DNA(contigs)



Markers

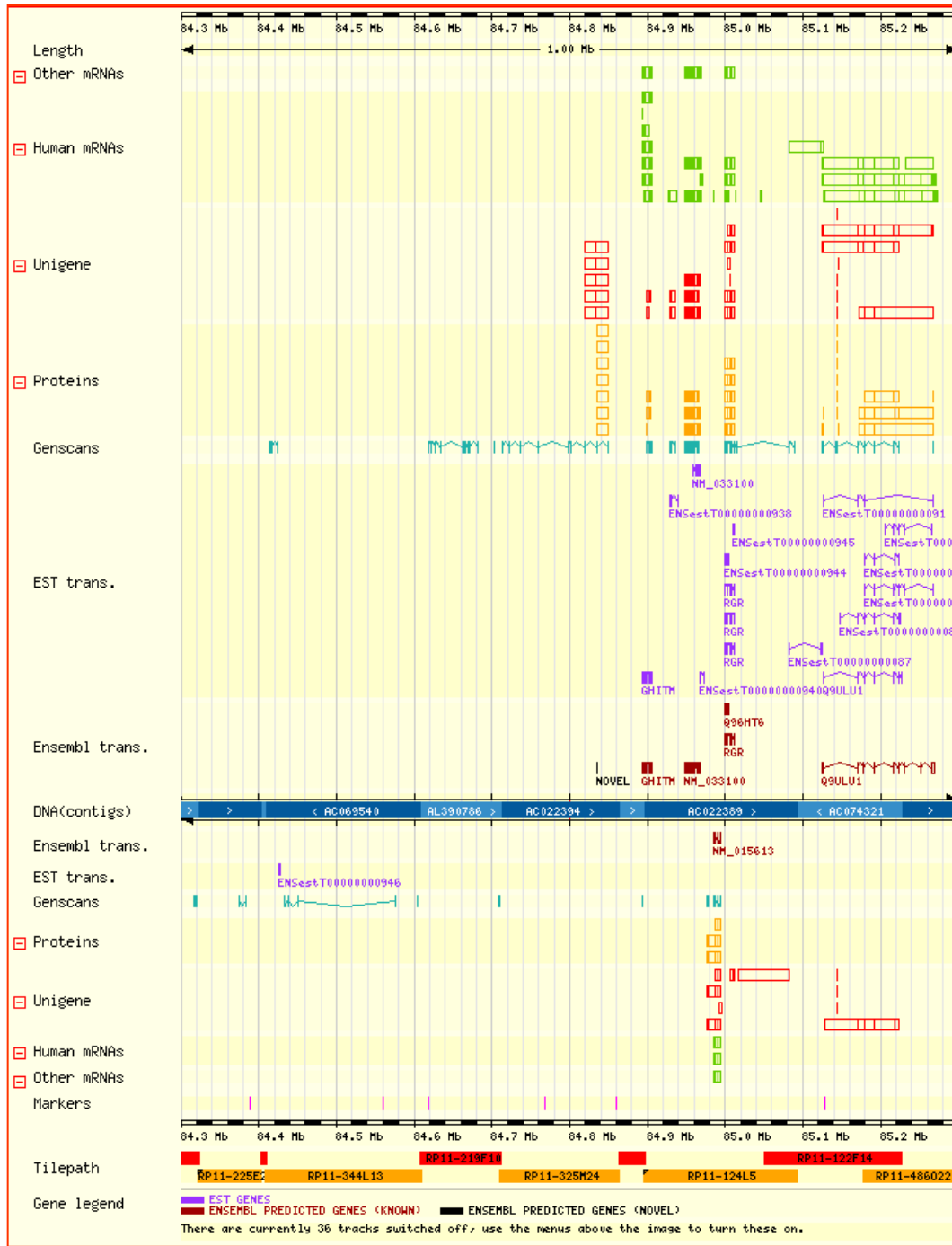
Markers: D10S551, AFM240uf10, D10S551, AFM240uf10, D10S551, AFM191va9, D10S1686, D10S1174, AFMa196zb1, D10S1658, AFMb299ze9, D10S1717, D10S2175

Genes



Gene legend

ENSEMBL PREDICTED GENES (KNOWN) ENSEMBL PREDICTED GENES (NOVEL)



Length

74028514 74028614

101 bp

Amino acids

N T A Q I N G P N P L F * S * P S F K A P L L L H K P F L D H S S P
K H C T N K K W S Q S S I L I L T I L Q G P F A S T Q A L P R P F Q S R
T L H K * H V P I L Y F N P N H P S R P L C F Y T S P S * T I P V Q

Sequence

ACACTGCACAAA TAAATGGTCCCAATCCCTCTATTTTAA TCC TAACCATCCTTCAGGCCCCCTTTGCTTC TACACACAGCCC TTCC TAGACCATTCCAGTCCA

DNA(contigs)

AL359074 >

Sequence

TGTGACGTGTTTATTATCCAGGGTTAGGAGATRAAAATTAGGATTGGTAGGAAGTTCCGGGGAAACGAGATGTGTTCCGGGAGGATCTGGTAAGGTCAGGTT

Amino acids

C Q V F L H D W D E I K I R V M R * P G K A E V C A R G L G N W D L
F V A C I F P G L G R N * D * G D K L A G K S R C L G K R S W E L G S
V S C L Y I T G I R * K L G L W G E L G R Q K * V L G E * V H G T W

Ensembl trans.

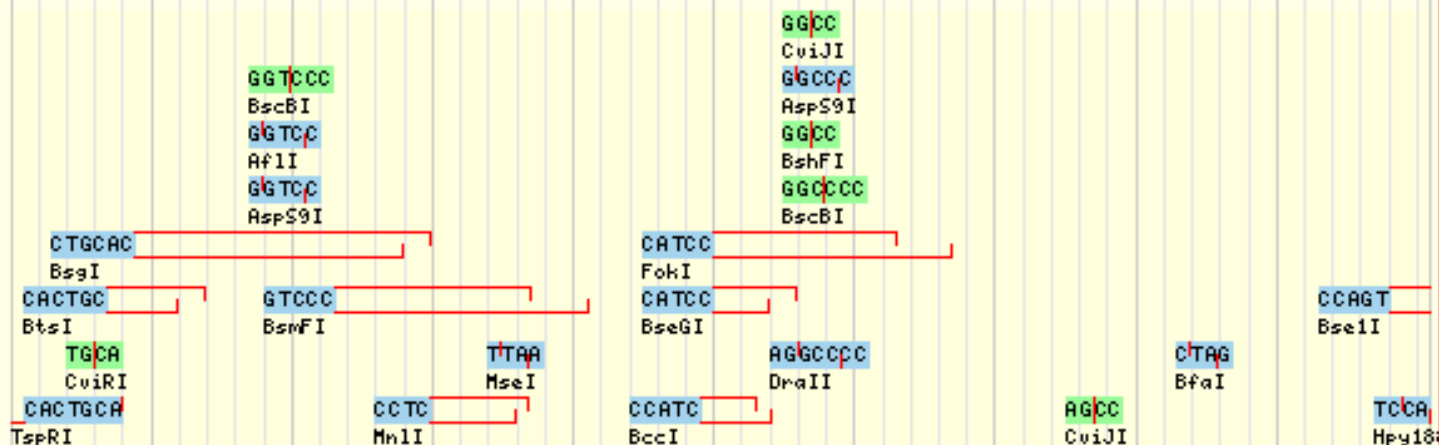
PPP3CB

EST trans.

PPP3CB

Genscans

Restr.Enzymes



74028514 74028614

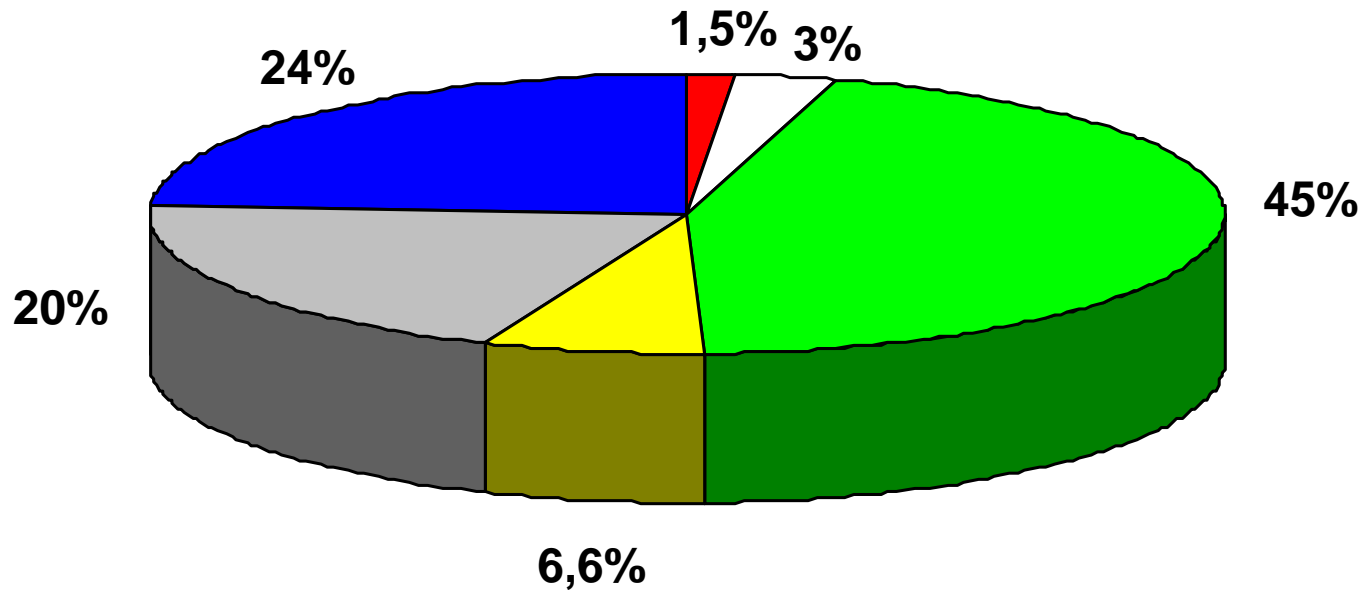
Tilepath







RP11-137L10

Gene legend

EST GENES
ENSEMBL PREDICTED GENES (KNOWN) ENSEMBL PREDICTED GENES (NOVEL)

Το γονιδίωμα με αριθμούς



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

ΠΟΣΑ ΓΟΝΙΔΙΑ ΚΩΔΙΚΟΠΟΙΟΥΝ ΠΡΩΤΕΪΝΕΣ;

Κινητική επαναδιάταξης
40.000

Νησίδια CpG
80.000

ESTs
35.000-120.000

Συγκριτικές μελέτες με *Tetraodon nigroviridis*
28.000-34.000

Κατ' αναλογία με χρ. 20, 21, 22
31.500-45.000



Ανάλυση 2004
20.000-25.000

Ανάλυση 2001
30.000-40.000



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

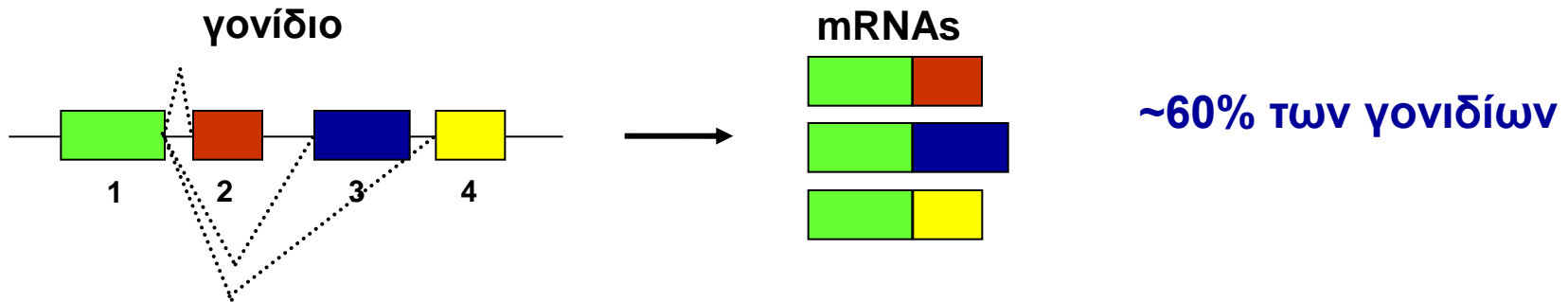
3000Mb



~25.000

Η πολυπλοκότητα της γενετικής πληροφορίας

1. Εναλλακτική συναρμογή

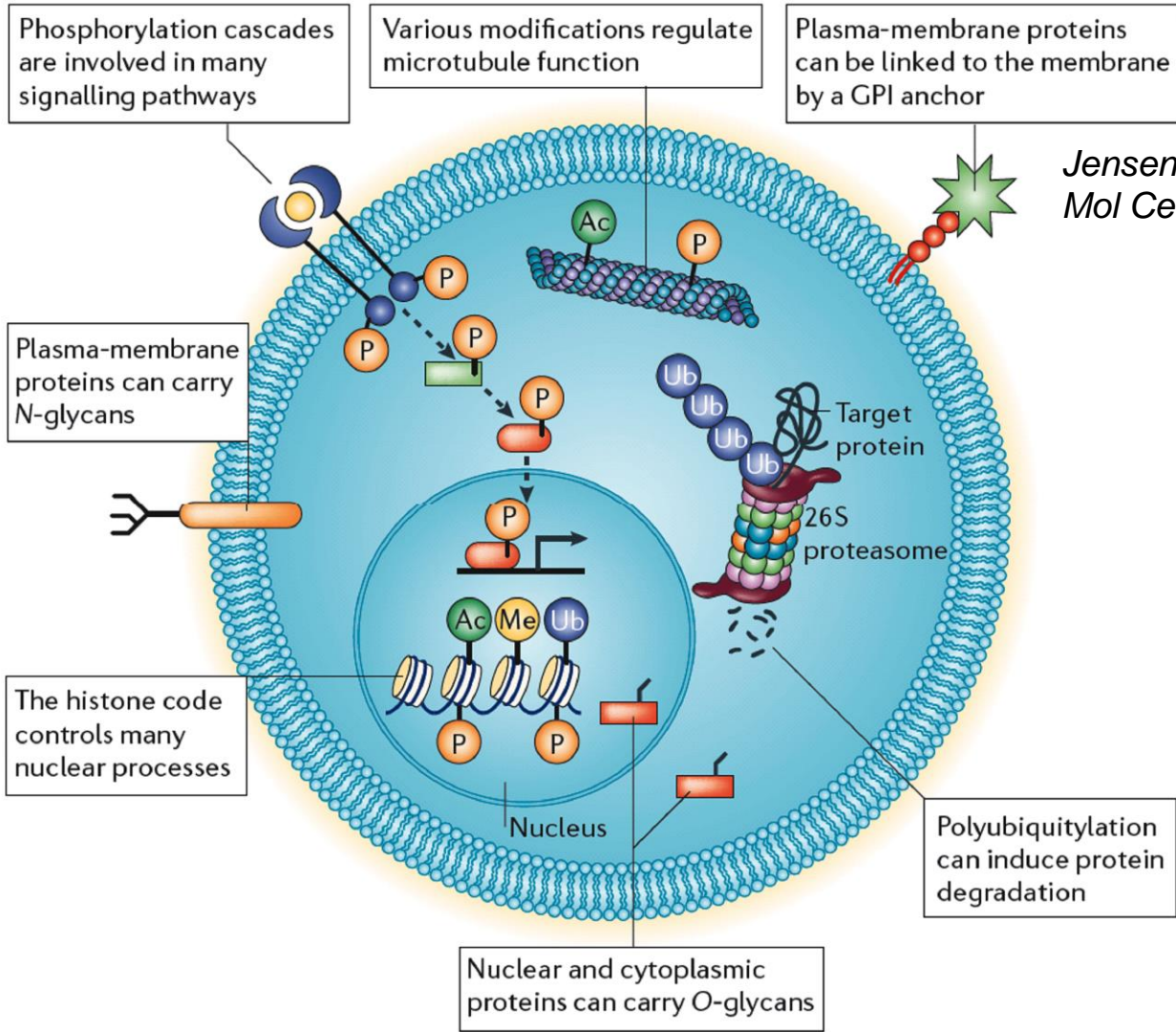


- Chr 22: 642 μετάγραφα αντιστοιχούν σε 245 γονίδια
 - 2,6 μετάγραφα/ γονίδιο
- Chr 19: 1859 μετάγραφα αντιστοιχούν σε 544 γονίδια
 - 3,2 μετάγραφα/ γονίδιο

70% επηρεάζουν την κωδική αλληλουχία

- *C. elegans*
 - 22% γονιδίων με εναλλακτική συναρμογή
 - 1.34 μετάγραφα / γονίδιο

2. Μετα-μεταφραστικές τροποποιήσεις



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

Γονίδια = C



Μετάγραφα ~ 3 x C



Πρωτεΐνες = 10 x C (?)

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

Σχόλια για τον αριθμό των γονιδίων... (2001)

Είναι μία προοπτική που μας ντροπιάζει. Όποιος μελετά το γονιδίωμα σύντομα αρχίζει να αισθάνεται ότι δεν είναι παρά ένα εφήμερο μέσο για τη δημιουργία περισσότερου DNA
Robert Waterston

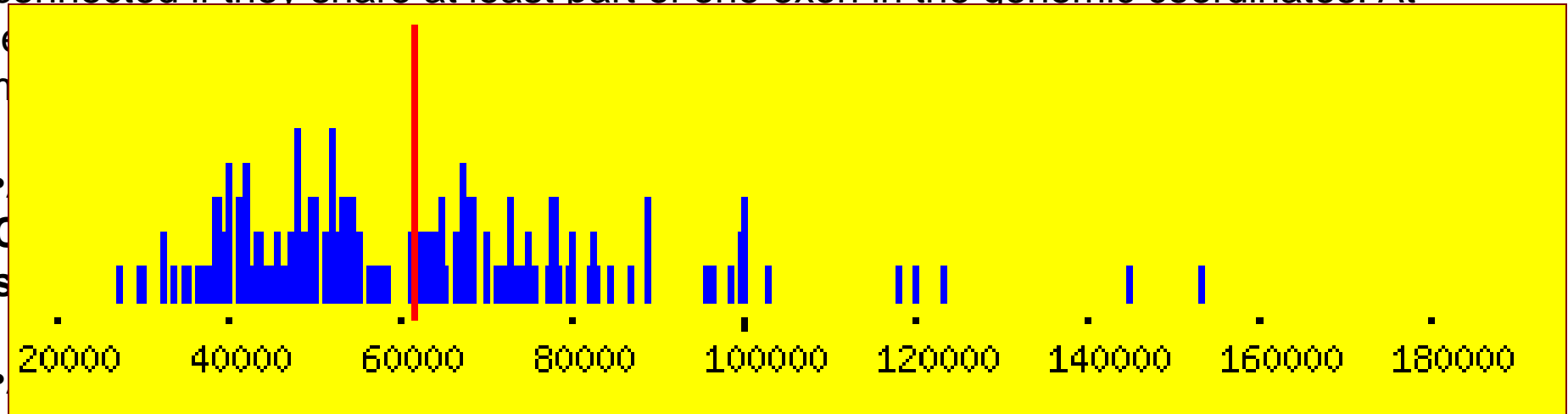
Για να δημιουργηθεί ο Αϊνστάιν χρειάστηκαν μόλις 12.000 περισσότερα γονίδια από ένα σκουλήκι και γύρω στις 17.000 περισσότερα γονίδια από μία φρουτόμυγα
Daily Telegraph

Δεν αισθάνομαι να μειονεκτώ επειδή δεν διαθέτω πολύ περισσότερα γονίδια από μία μύγα-οι μύγες είναι πολύ περίπλοκα όντα, έχουν τέσσερα φτερά και ξέρουν να πετούν, ενώ εγώ δεν ξέρω
Martin Bobrow, καθ. Ιατρικής Γενετικής

Gene Sweepstake

The Gene Sweepstake will run between 2000 and 2003. The rules are:

- It costs \$1 to make a bet in 2000, \$5 in 2001 and \$20 in 2002.
- Bets are for one number. Closest number wins, and in case of ties, the pot is split
- A gene is a set of connected transcripts. A transcript is a set of exons via transcription followed (optionally) by pre-mRNA splicing. Two transcripts are connected if they share at least part of one exon in the genomic coordinates. At



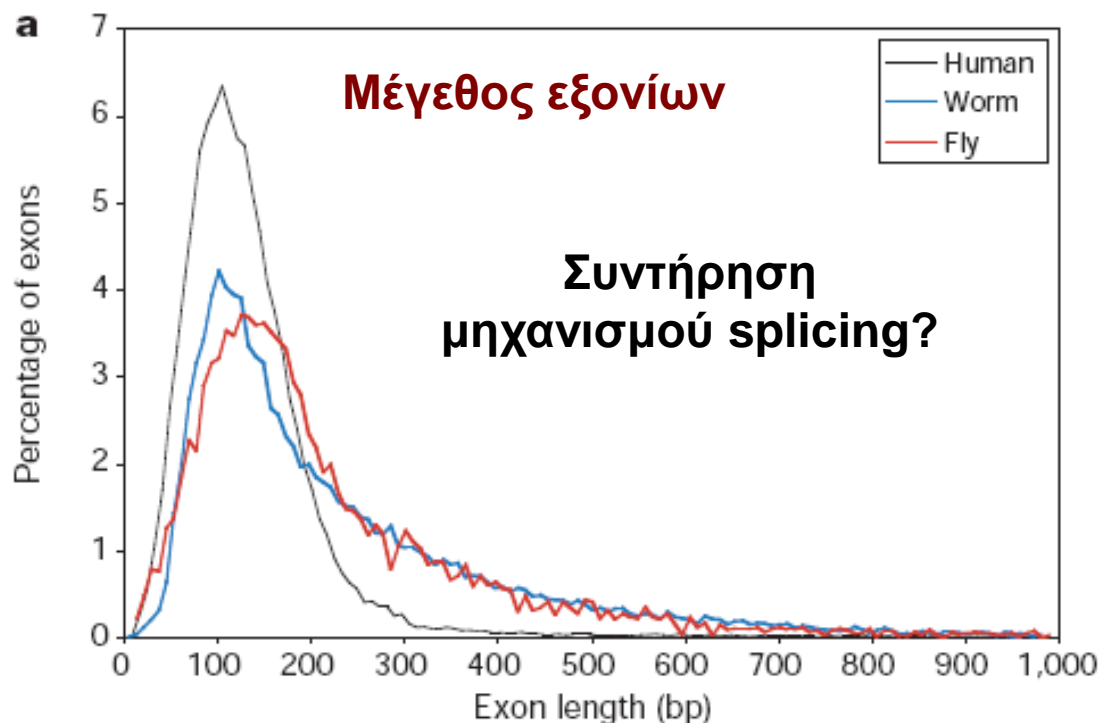
• People betting should write their name, email and number in the Gene Sweepstake book, held at Cold Spring Harbor (contact [David Stewart](#)).

• One bet per person, per year. Year defined as a calendar year.

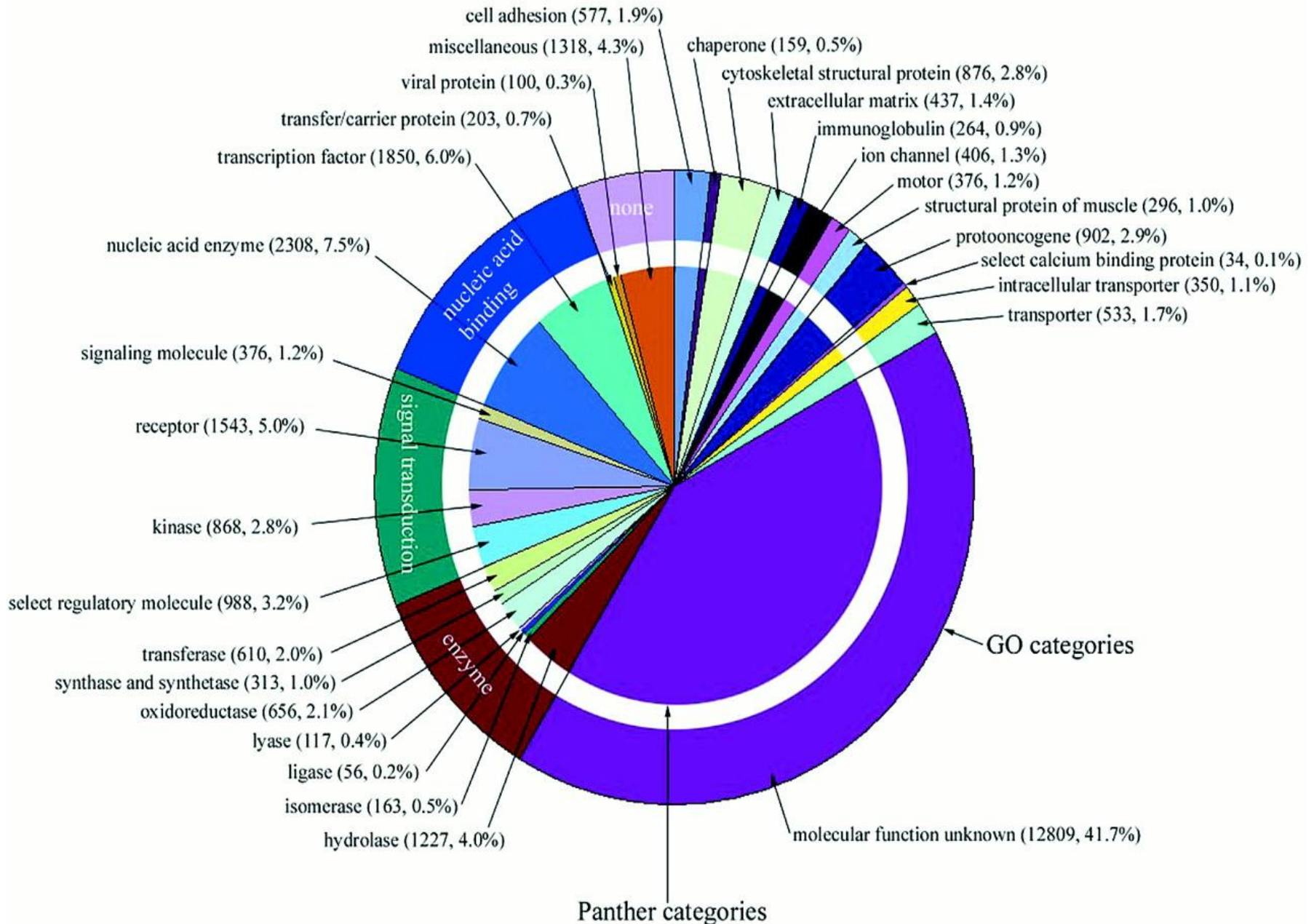
• No pencil bets (ie, you can't change your number)

Χαρακτηριστικά γονιδίων του ανθρώπου

	Median	Mean
Internal exon	122 bp	145 bp
Exon number	7	8.8
Introns	1,023 bp	3,365 bp
3' UTR	400 bp	770 bp
5' UTR	240 bp	300 bp
Coding sequence (CDS)	1,100 bp 367 aa	1,340 bp 447 aa
Genomic extent	14 kb	27 kb

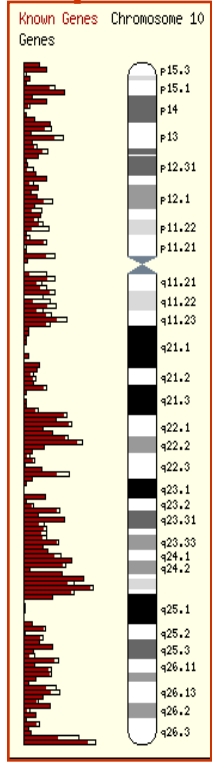


Κατηγοριοποίηση γονιδίων που κωδικοποιούν πρωτεΐνες



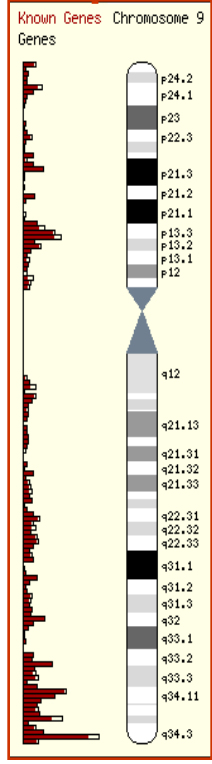
Μέση γονιδιακή πυκνότητα ~7,5 γονίδια / Mb

Χρ. 10



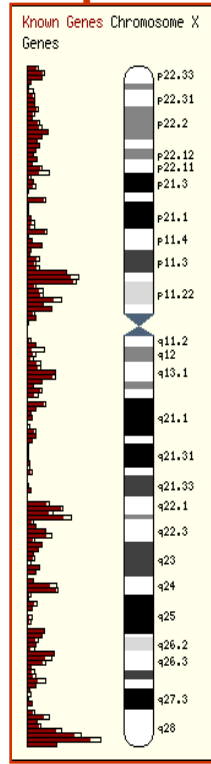
6,5 / Mb

Χρ. 9



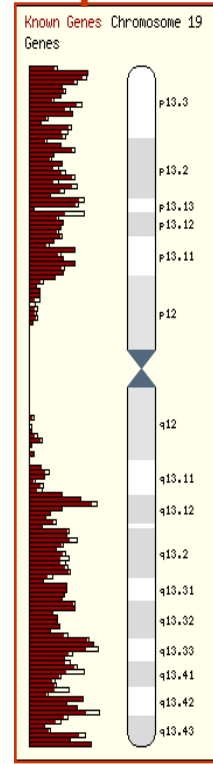
6,5 / Mb

Χρ. X



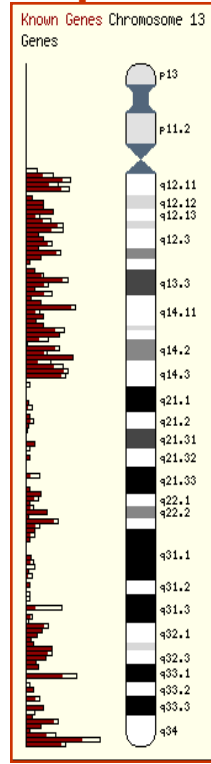
6 / Mb

Χρ. 19



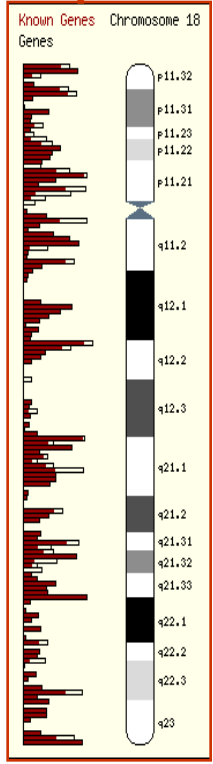
22 / Mb

Χρ. 13



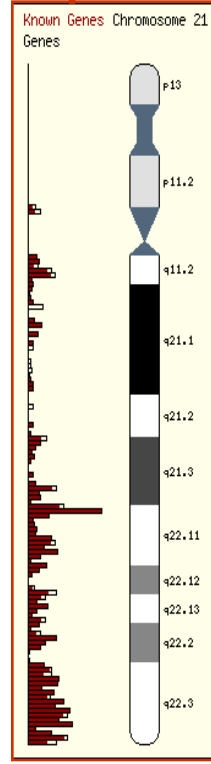
3,5 / Mb

Χρ. 18



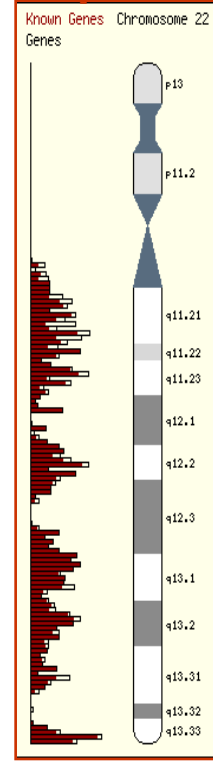
4 / Mb

Χρ. 21



5 / Mb

Χρ. 22



11 / Mb

ΓΟΝΙΔΙΑ ΠΟΥ ΚΩΔΙΚΟΠΟΙΟΥΝ RNA

