ΠΡΟΓΡΑΜΜΑ ΧΑΡΤΟΓΡΑΦΗΣΗΣ ΤΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ ΤΟΥ ΑΝΘΡΩΠΟΥ

HUMAN GENOME PROJECT (HGP)

Ιστορική αναδρομή

Μάιος 1985

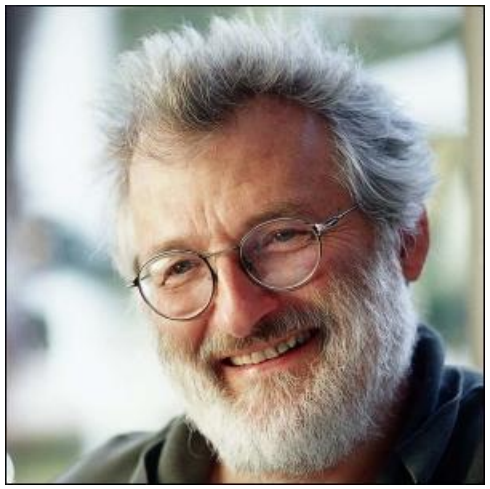
Robert Sinsheimer

Πρύτανης του Παν. της Καλιφόρνιας, Santa Cruz

David Botstein

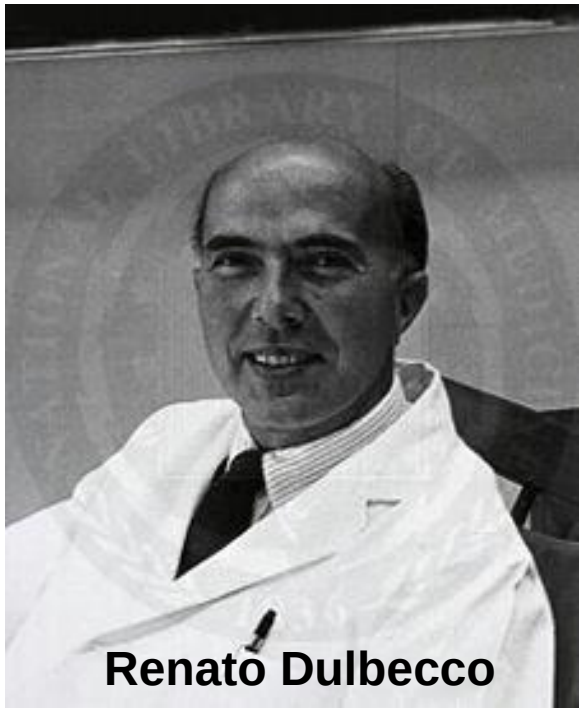
Walter Gilbert

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

Charles De Lisi
Διευθυντής του Τμήματος Έρευνας για την Υγεία & το Περιβάλλον στο Υπουργείο Ενέργειας

Μάρτιος 1986

Santa Fe meeting

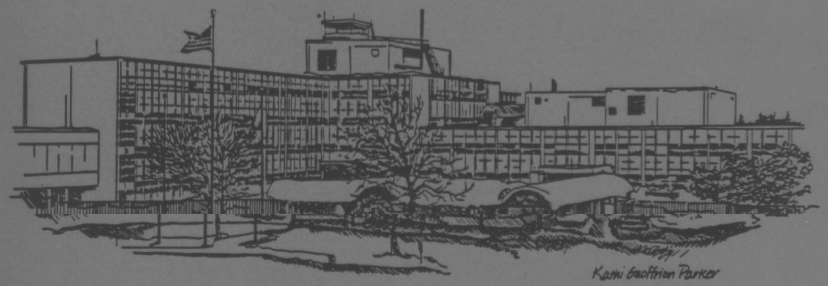
Εκτίμηση του κόστους, της ωφελιμότητας, της δυνατότητας επίτευξης και του χρόνου του προγράμματος

3 φάσεις
Ανάπτυξη τεχνολογίας
Χαρτογράφηση
Αλληλούχηση

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 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.

NATURE Vol 455 16 October 2008

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.

Ιστορική αναδρομή

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

Οργανωτής: James Watson

Sequencing the genome now is like Lewis and Clark going to the Pacific one millimetre at a time. If they had done that, they would still be looking. David Botstein, Whitehead Institute, Cold Spring Harbor Symposium on the Molecular Biology of Homo sapiens, June 1986

I am surprised consenting adults have been caught in public talking about it [sequencing the genome]… it makes no sense, Robert Weinberg, Whitehead Institute, in The New Scientist, Mar. 5, 1987, p.35

Of course we are interested in having the sequence, but the important question is the route we take to getting it, Maxine Singer, director, Carnegie Institution of Washington, in Science 232:1600, 1986

Sequencing the human genome is like pursuing the holy grail, Walter Gilbert, Harvard University, at several national meetings, March 1987-August 1987

The sequence of the human genome would be perhaps the most powerfull tool ever developed to explore the mysteries of human development and disease, Leroy Hood and Lloyd Smith, California Institute of technology, in Issues in Science and technology, 3:37, 1987

The main reason that research in other species is so strongly supported by Congress is its applicability to human beings. Therefore, the obvious answer as to whether the human genome should be sequenced is 'Yes. Why do you ask?' Daniel Koshland, Editor, Science, 236:505, 1987

Human Genome Project (HGP)

ΕΝΑΡΞΗ: 1988

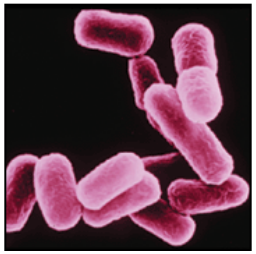
ΣΤΟΧΟΙ:

- 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

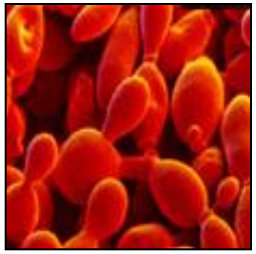
J. Watson

Διευθυντής στο National Human Genome Research Institute

Οργανισμοί-μοντέλα

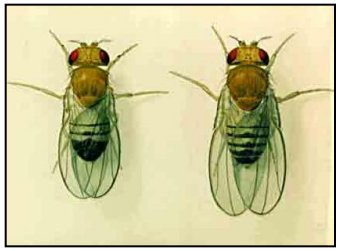


E. coli



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.

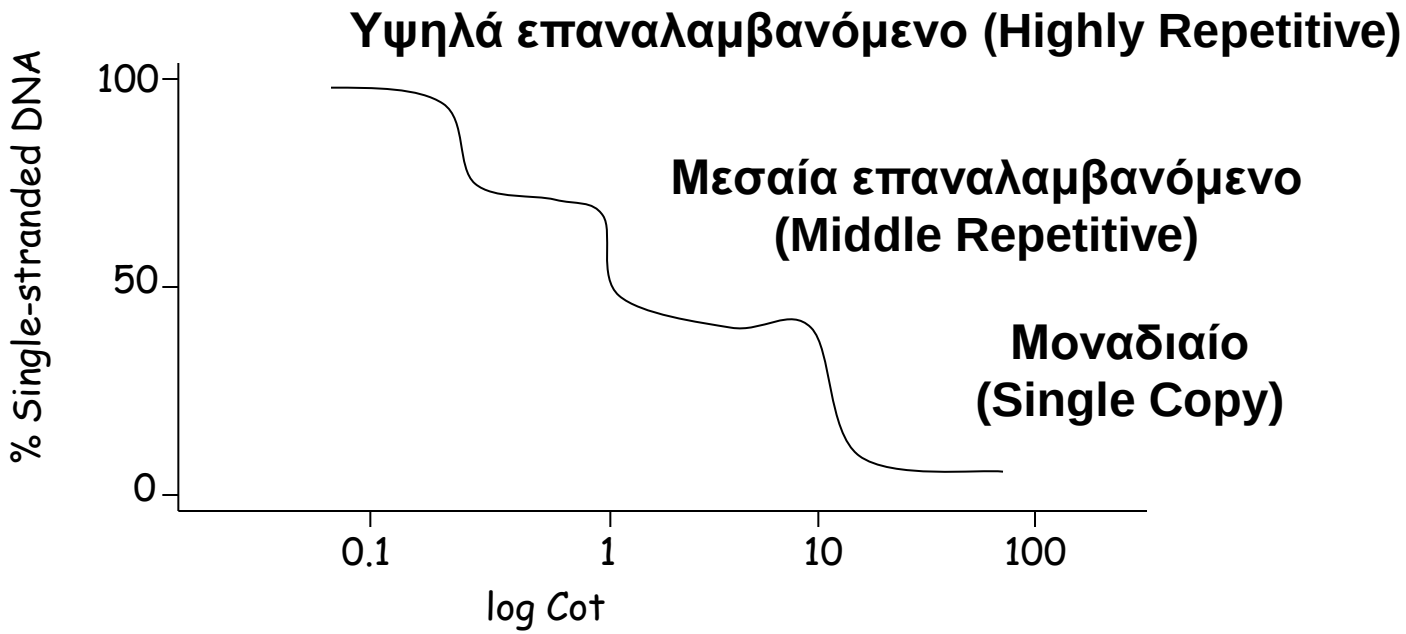
6. The participants encourage continued effort to develop compatible genomic data bases and networks and measures to ensure world-wide access to these resources.

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



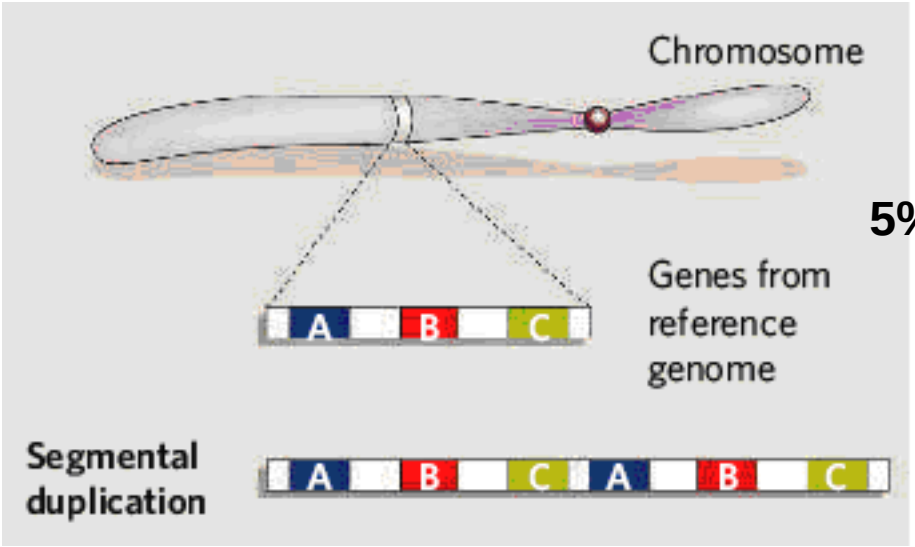
Επαναλαμβανόμενες αλληλουχίες του ανθρώπινου γονιδιώματος

	Μήκος	Αριθμός αντιγράφων	% του γονιδιώματος
LINES: Long INterspersed Elements	6-8kb	850.000	21%
SINES: Song INterspersed Elements (Alu repeats…)	100-300bp	1.500.000	13%

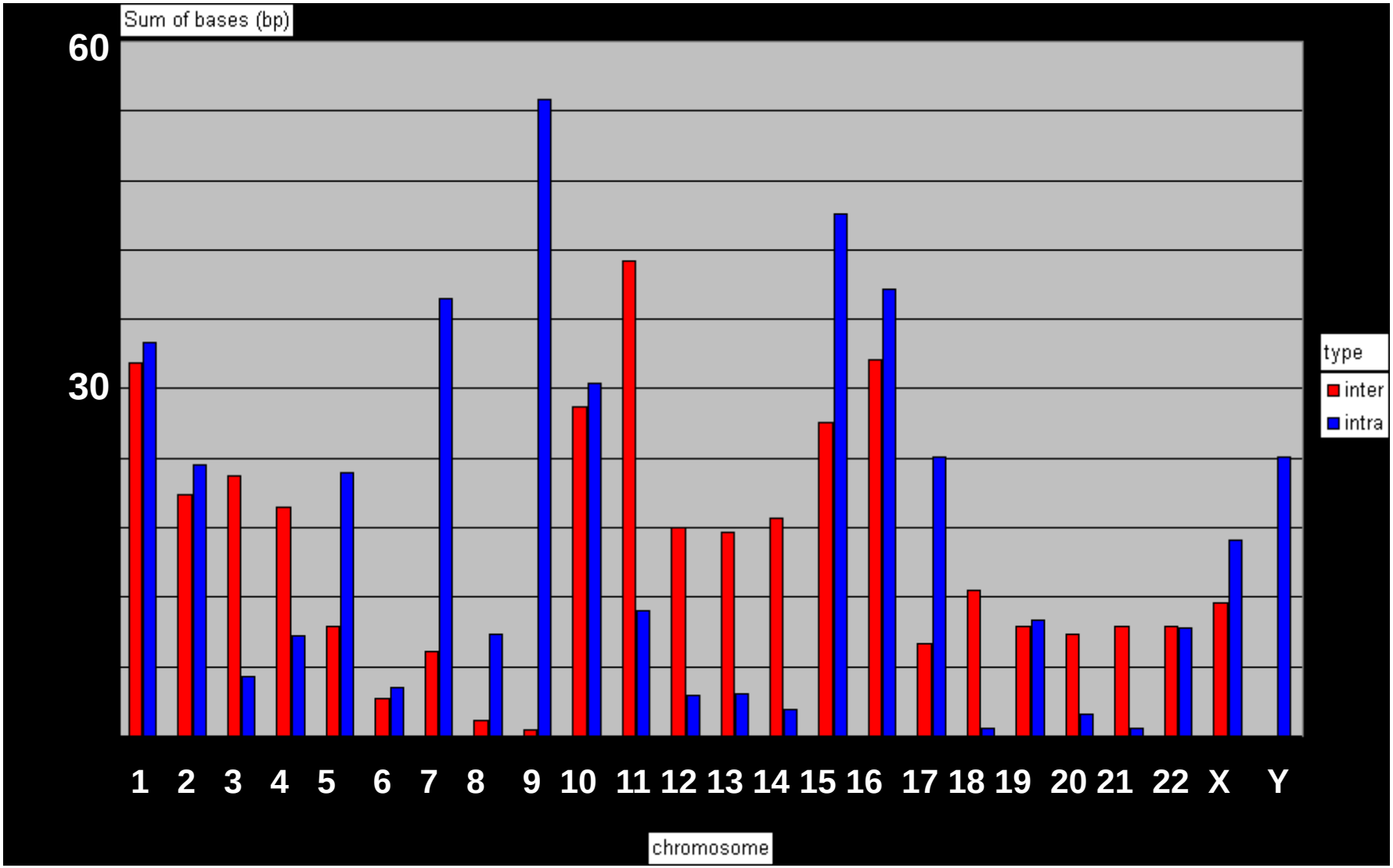
Segmental duplications (διπλασιασμοί χρ. τμημάτων)

(> 90% ομοιότητα, >1kb μήκος).

5%



Segmental Duplication Database



http://humanparalogy.gs.washington.edu

Αλληλούχηση μόνο των cDNAs;

Sydney Brenner

1990

Craig Venter

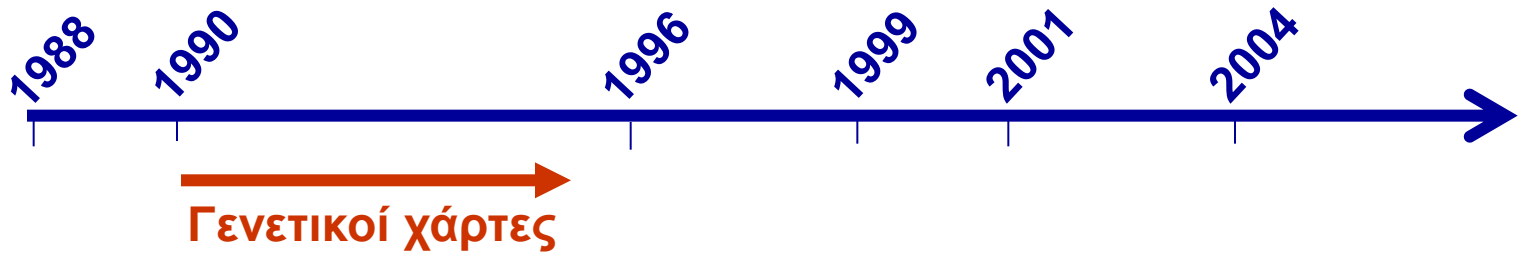
1991 (NIH)

~600 ESTs
>50% νέα
Πατέντα???

TIGR
πατέντα

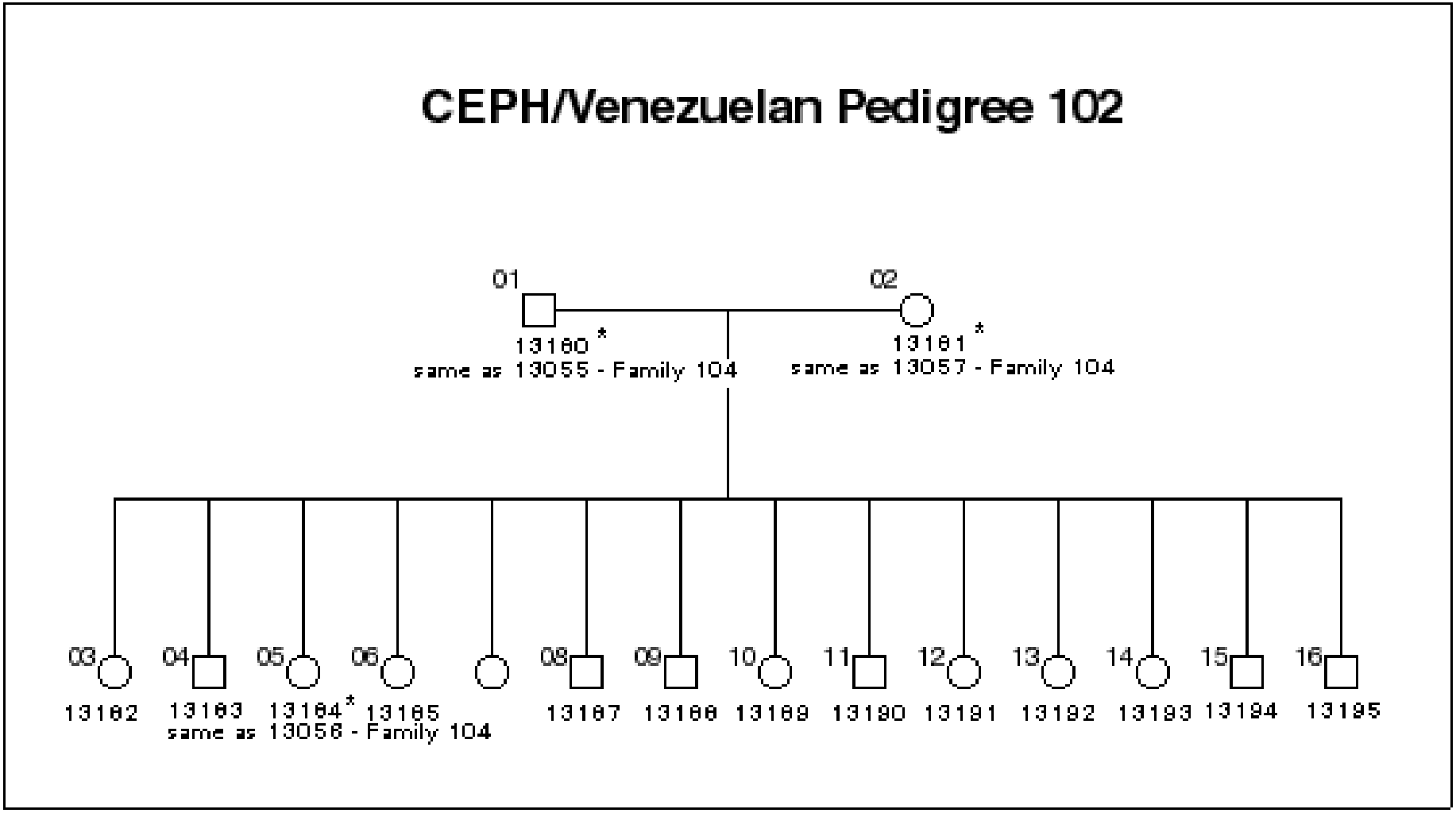
GSC (Genome Sequencing Center, U. of Washington)
Ελεύθερα στις db

ΠΟΡΕΙΑ ΤΟΥ HGP

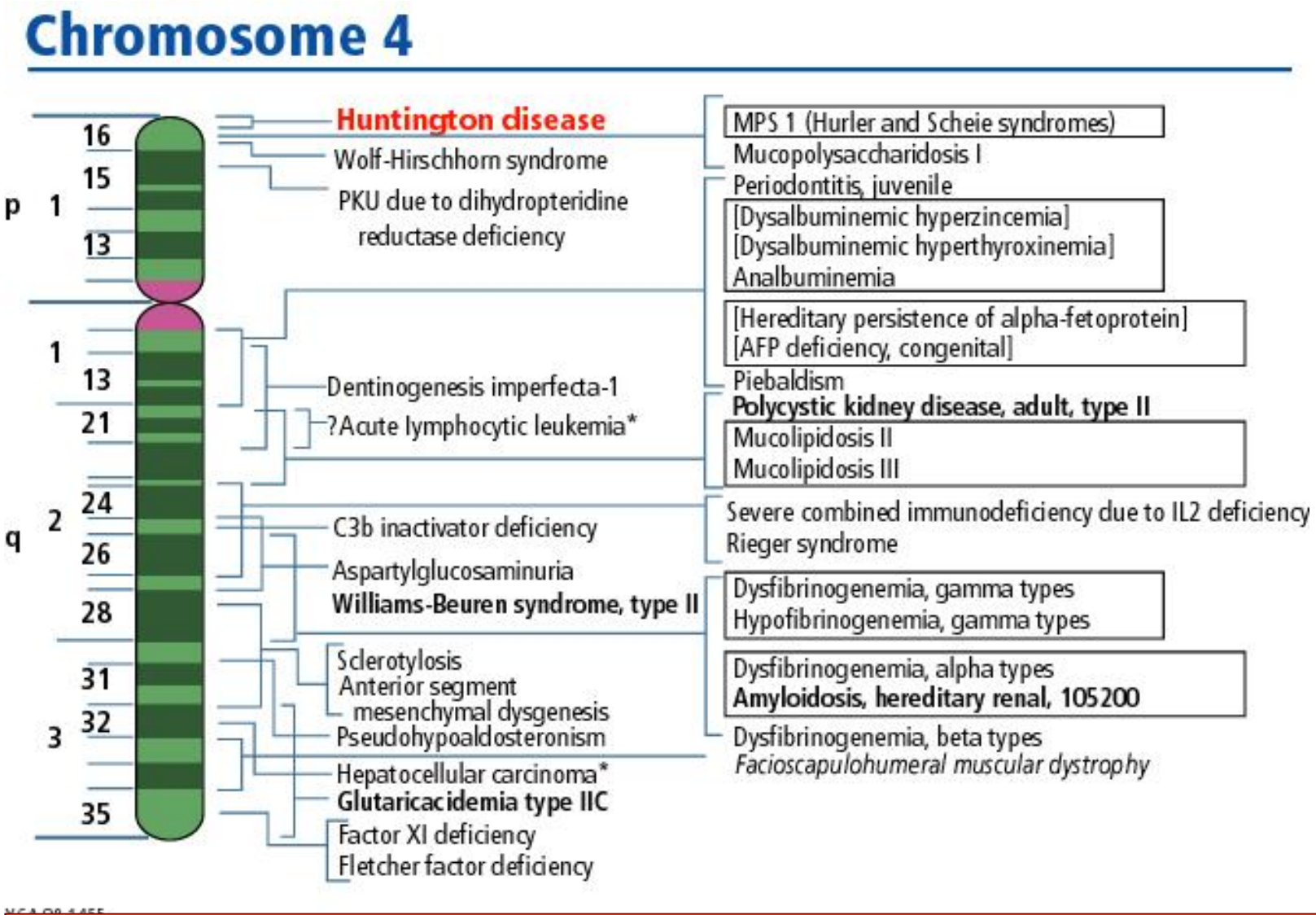


Οικογένειες CEPH

Πολλές γενιές
Πολλά άτομα

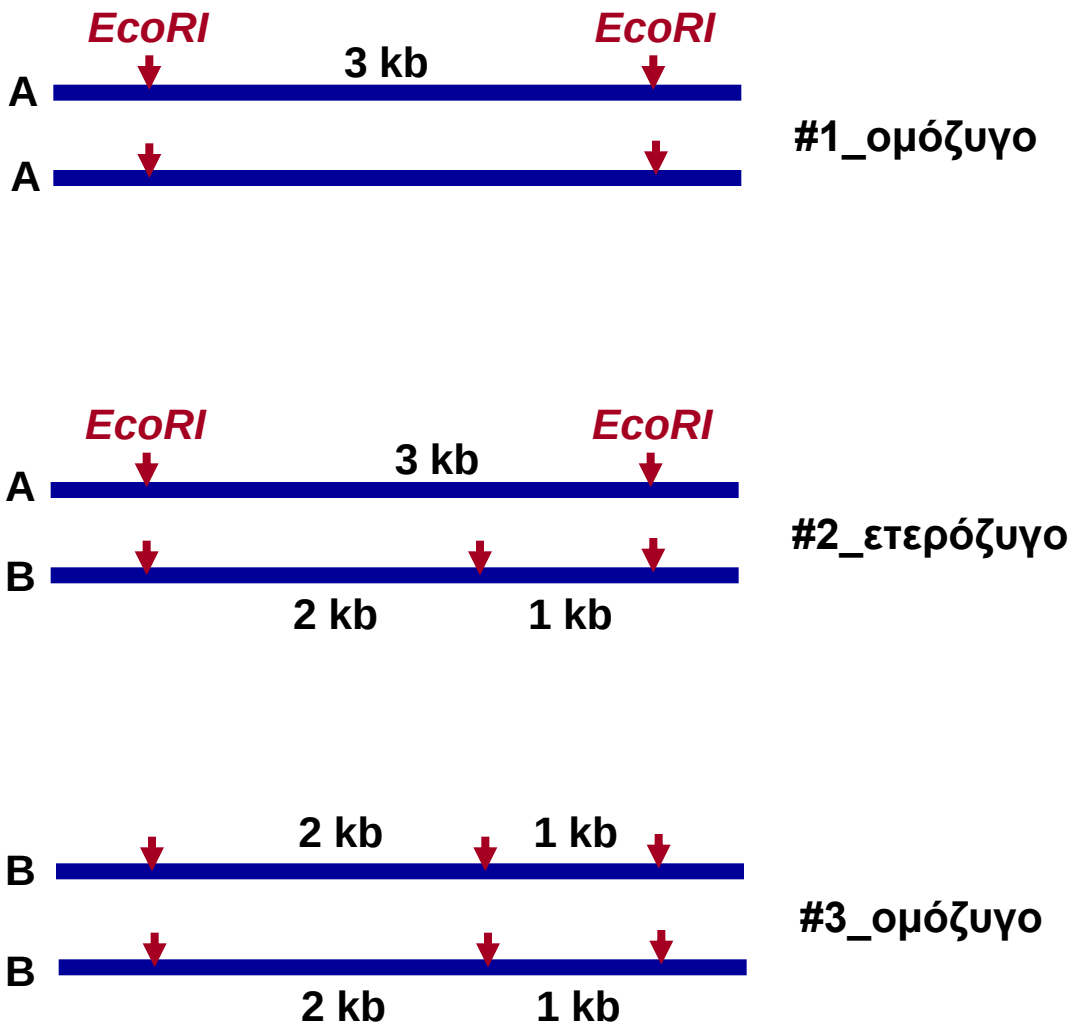


Γενετικός χάρτης με γονίδια του χρωμοσώματος 4 του ανθρώπου

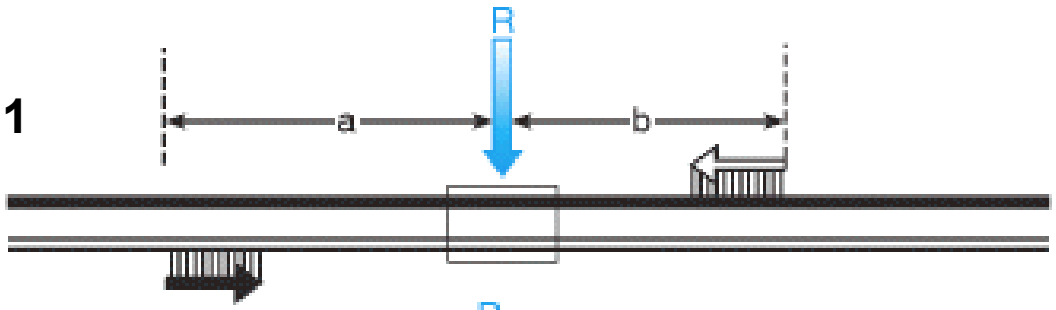


Χρειάζονται περισσότεροι μοριακοί γενετικοί δείκτες (markers)

RFLP (restriction site length polymorphisms), 1975-



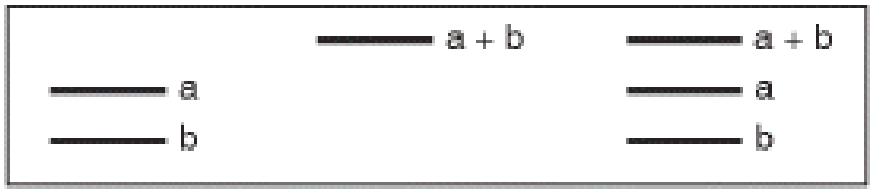
Ανίχνευση με PCR-πέψη



Πολλαπλασιασμός

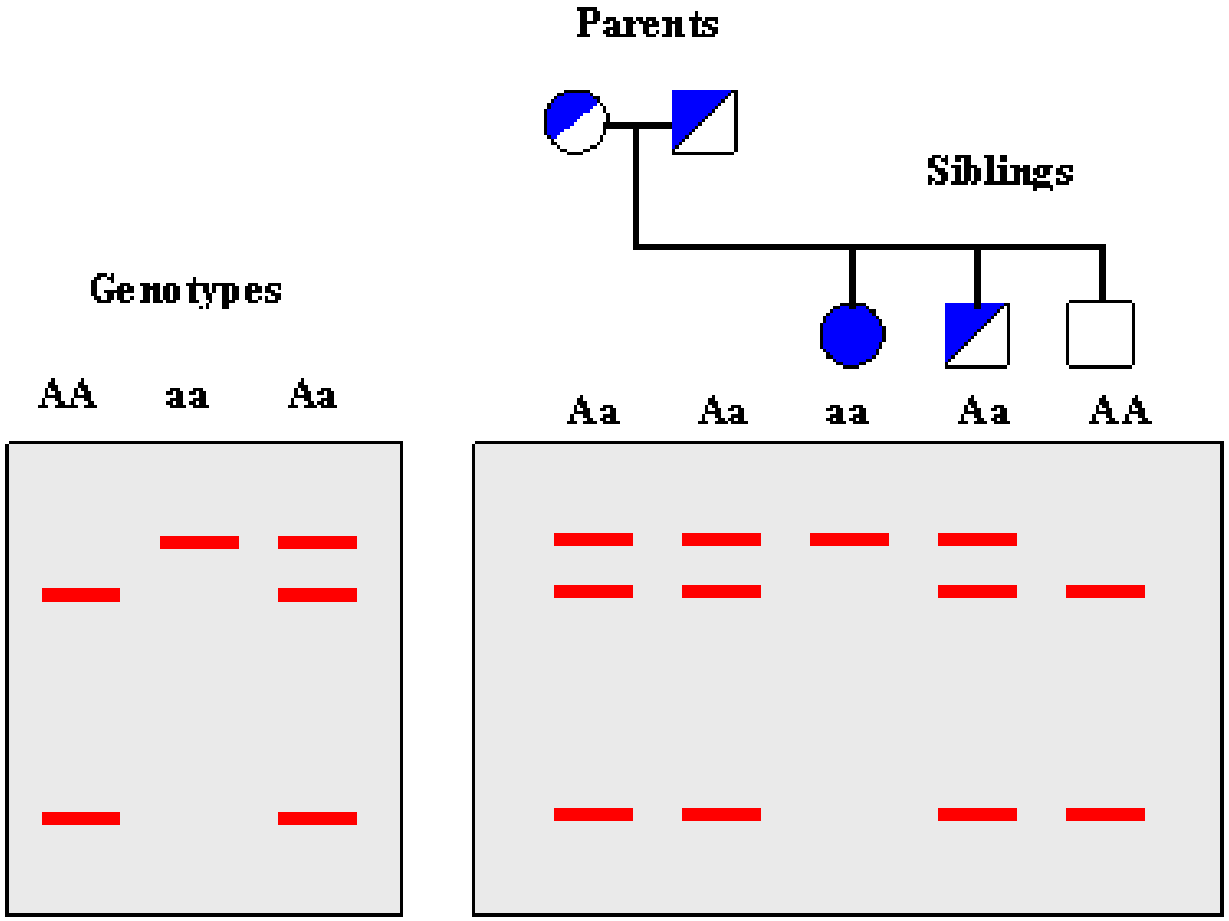
Πέψη του προιόντος της PCR
με το ένζυμο περιορισμού

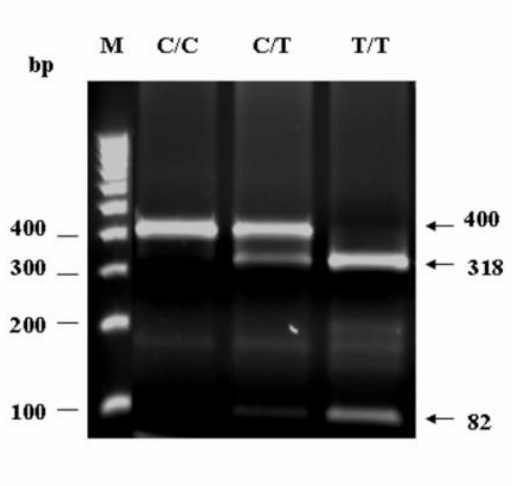
Διαχωρισμός προϊόντων της πέψης
με ηλεκτροφόρηση σε πήκτωμα



Αλληλόμορφα	1, 1	2, 2	1, 2

Κληρονόμηση RFLP





Cell, Vol. 51, 319–337, October 23, 1987, Copyright © 1987 by Cell Press

A Genetic Linkage Map of the Human Genome

Helen Donis-Keller,* Philip Green,* Cynthia Helms,* Samuel Cartinhour,* Barbara Weiffenbach,* Karen Stephens,* Tim P. Keith,* Donald W. Bowden,* Douglas R. Smith,* Eric S. Lander,† David Botstein, Gita Akots,* Kenneth S. Rediker,* Thomas Gravius,* Valerie A. Brown,* Marcia B. Rising,* Carol Parker,* Jody A. Powers,* Diane E. Watt,* Erick R. Kauffman,* Angela Bricker,* Pamela Phipps,* Hans Muller-Kahle,* Thomas R. Fulton,* Siu Ng,* James W. Schumm,* Jeffrey C. Braman,* Robert G. Knowlton,* David F. Barker,* Steven M. Crooks,* Steven E. Lincoln,† Mark J. Daly,† and Jeff Abrahamson†

* Department of Human Genetics
Collaborative Research, Inc.
Bedford, Massachusetts 01730

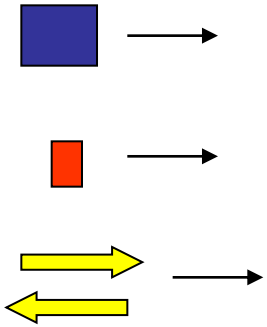
†Whitehead Institute for Biomedical Research
9 Cambridge Center
Cambridge Massachusetts 02142

inheritance. Over the next 75 years, complete genetic linkage maps proved to be essential tools for studying the properties of mutations. Genetic markers gained new importance with the advent of recombinant DNA, since cloned markers provide starting points for cloning closely linked genes by chromosomal walking (Bender et al., 1983). Unfortunately, the construction of complete genetic linkage maps has traditionally required the isolation of hundreds of single-gene mutations with easily scored phenotypes, followed by extensive interbreeding of mutant stocks to ascertain the map position of the mutations. Such an effort has only been practical in a few intensively studied genetic systems, such as Escherichia coli, Saccharomyces cerevisiae, Drosophila melanogaster, Caenorhabditis elegans, Zea mays, and Mus musculus. In humans, despite great interest and occasional successes in detecting linkage (e.g., Mohr, 1954), construction of a genetic map seemed impractical.

Several years ago, Botstein et al. (1980) argued that it was feasible to construct a complete linkage map of the

~400 RFLPs

Μικροδορυφόροι (microsatellites)



Μοναδιαία αλληλουχία

Επανάληψη 1-5bp

Εκκινητές PCR

(GT)n
(GAC)n
(GATA)n
……

1/ 20 Kb στο ανθρώπινο γονιδίωμα

Παράδειγμα : μικροδορυφόρος Α

10

.....ACAACCCTAC CTGGCACTGC ATTGTGGGCC GAAACTTCGG GAGTTATGTG ACACATGAAC

Εκκινητής 1 → CCAAACACTT CATCTACTTC TACCTGGGTC GGGTGGCCAG TCTTCTGTTC AAATCTGGTT

AAGAGCATGG ACTGTGCCAA ACACCCAGTG ACCCATCCAA AAACAAGGAC TGCATCCAAA

GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

TYCCAAATAC CAGAGACTGA ATCTTCAGCC TTGCTAAGGG AACACCTCGT TTGAATCTGT

TGTGTTGTGT ACAGGGCTTC ATTCTCTGTA CAAGTCTGTG GTTATAAAAT TAGTAAAACC

GCTTACATTT GTATTTATTT TCTAGTCCAT ACTTCTGTAC CCTGAGCGGC CGCTGGAGCC.....

Εκκινητής 2

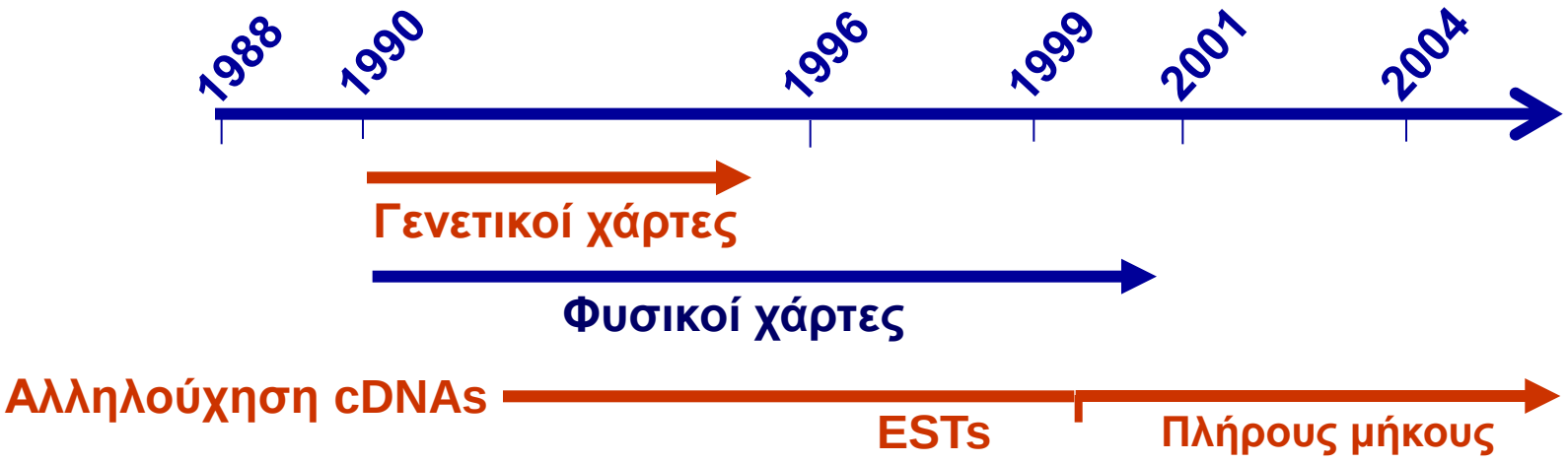
Ανίχνευση με PCR

The 1993–94 Généthon human genetic linkage map

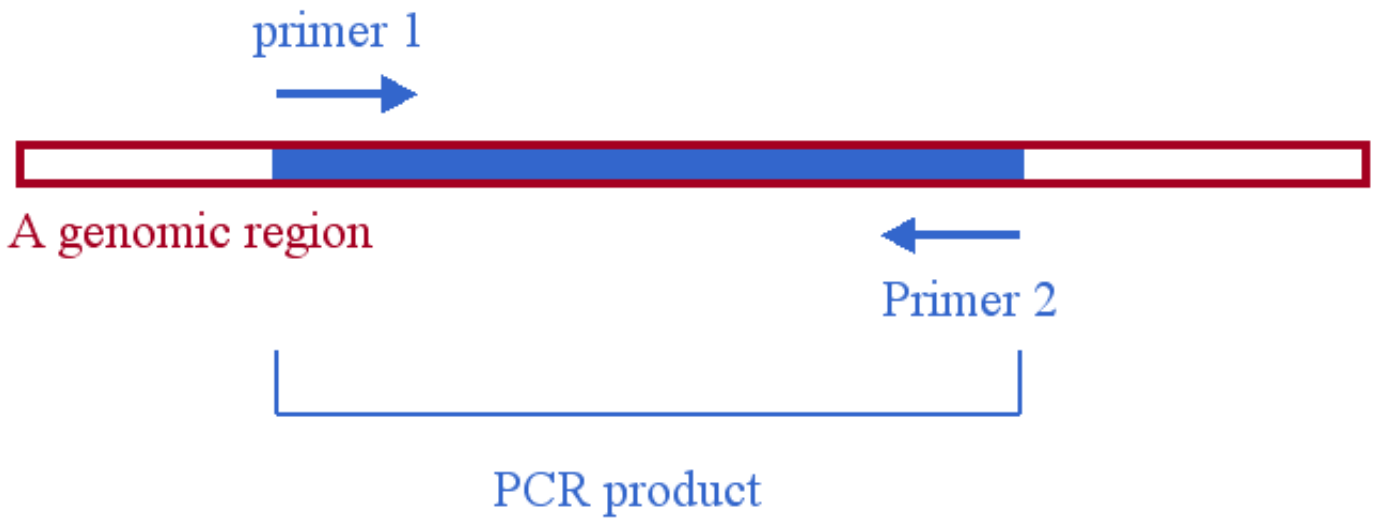
Gabor Gyapay[1,2], Jean Morissette[1,3], Alain Vignal[1], Colette Dib[1], Cécile Fizames[1], Philippe Millasseau[1,2], Sophie Marc[1], Giorgio Bernardi[4], Mark Lathrop[5] & Jean Weissenbach[1,6]

In 1992, we described a second-generation genetic linkage map of the human genome. Using 1,267 new microsatellite markers, we now present a new genetic linkage map containing a total of 2,066 (AC)$_n$ short tandem repeats, 60% of which show a heterozygosity of over 0.7. Statistical linkage analysis based on the genotyping of eight large CEPH families placed these markers in the 23 linkage groups. The map includes 1,266 intervals and spans a total distance of 3690 centiMorgans (cM). A total of 1,041 markers could be ordered with odds ratios greater than 1000:1. About 56% of this map is at a distance of 1 cM or less from one of its markers.

ΠΟΡΕΙΑ ΤΟΥ HGP



* Sequence Tagged Site



STS is defined as:

primer 1

primer 2

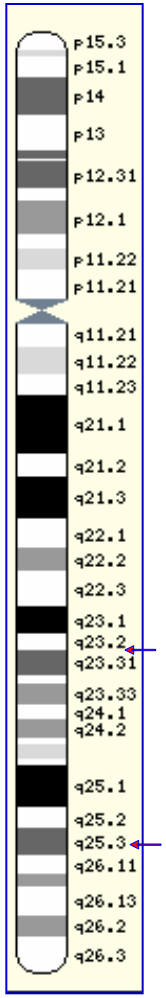
PCR product size

Sequence Tag Site

- **Μοναδιαία**
- **Γνωστής αλληλουχίας**
- **Χαρτογραφημένα (π.χ. μπορεί να είναι πολυμορφικοί δείκτες, ESTs)**

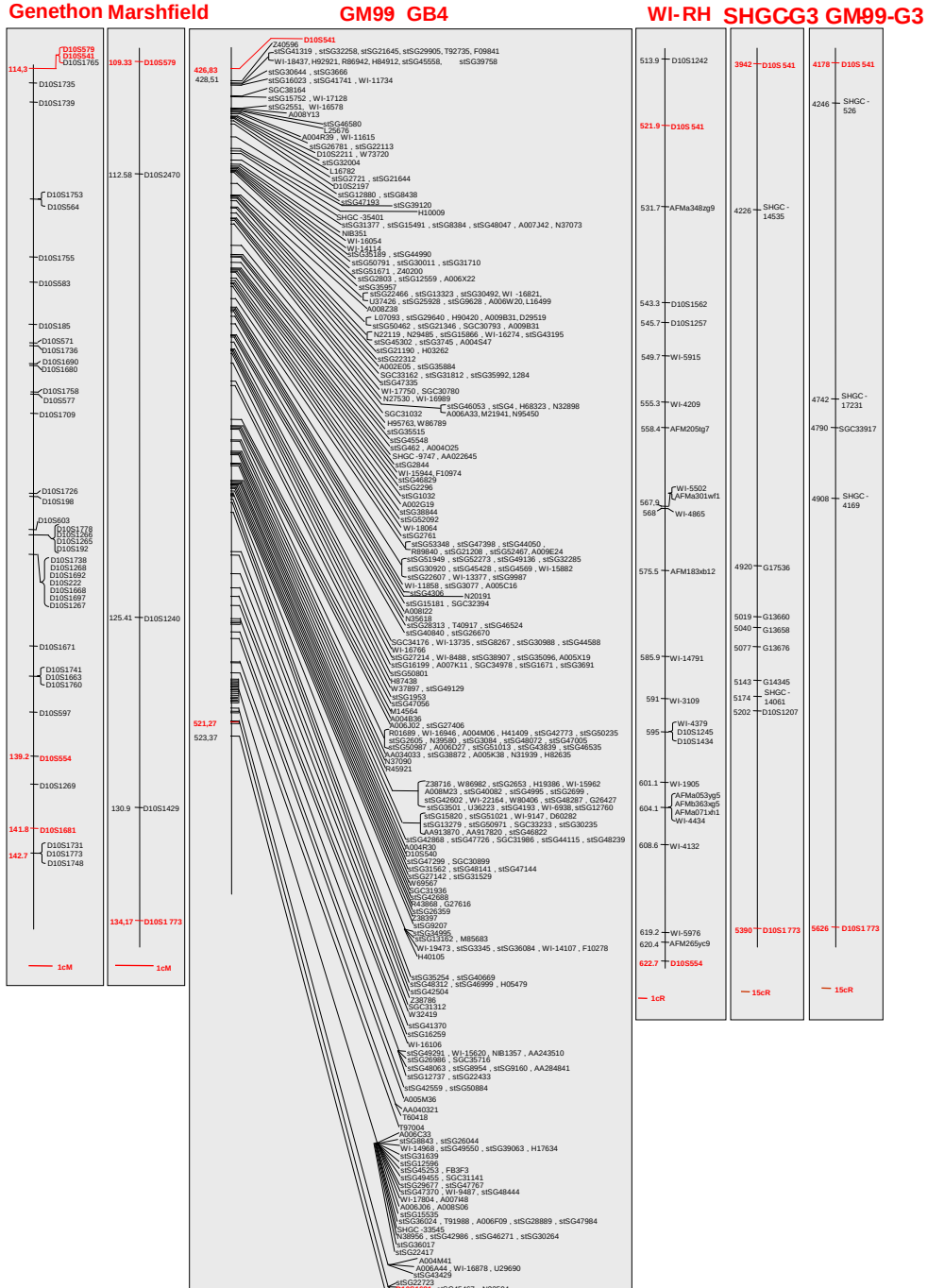
Φυσικός χάρτης γενωμικών κλώνων

Επιλογή ανιχνευτών

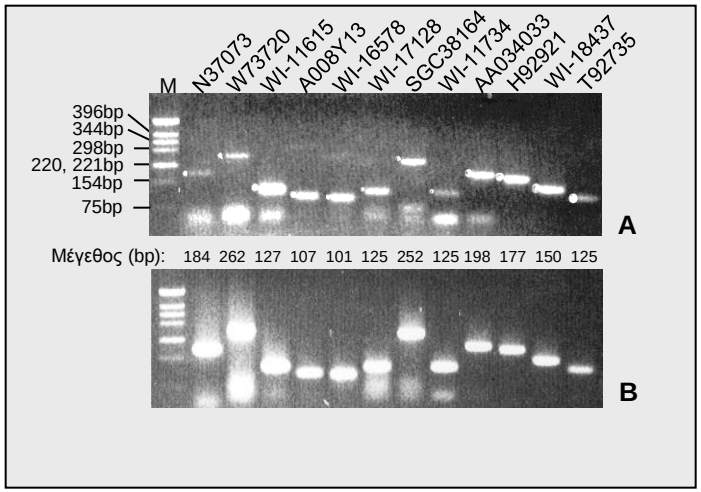


D10S541

D10S1731



Παρασκευή ανιχνευτών



PCR σε γενωμικό DNA

Επαναληπτική PCR

PCR ραδιοσήμανσης

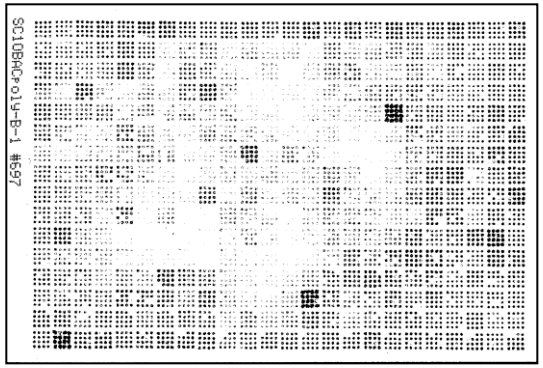
Υβριδοποίηση γενωμικής βιβλιοθήκης

Μεγεθος ενθέματος: 100-200kb
1 αντίγραφο/κύτταρο
15000-30000 για 1x κάλυψη του γονιδιώματος

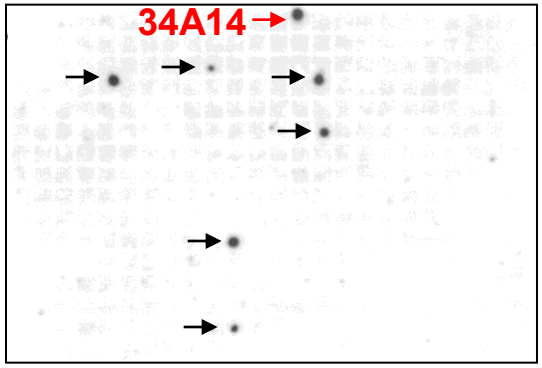
Pool: ~10 STSs

12cm

8cm

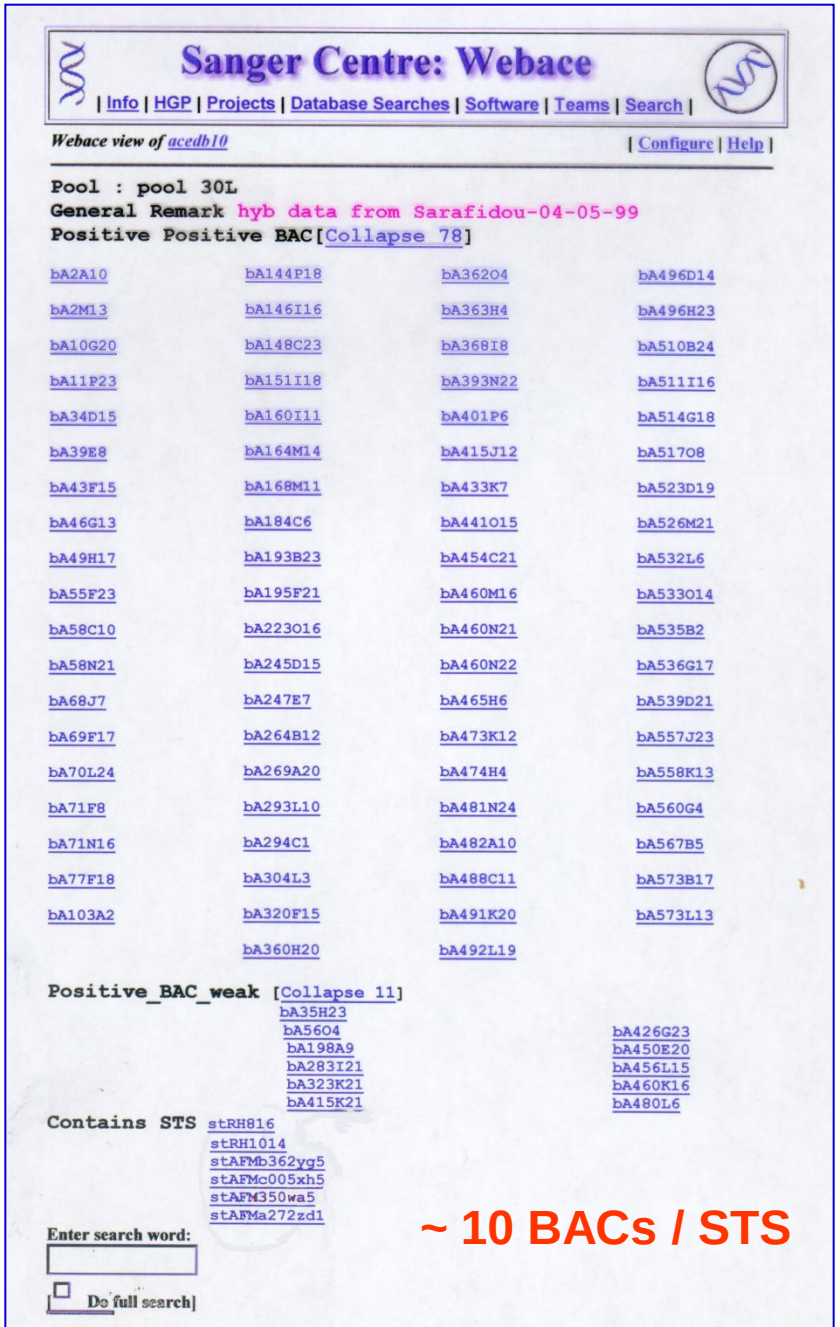


X36



~220.000 BACs
x 25 γονιδίωμα

Κατάθεση αποτελεσμάτων σε βάση δεδομένων



Sanger Centre: Webace

| Info | HGP | Projects | Database Searches | Software | Teams | Search |

Webace view of acedb10 [Configure | Help]

Pool : pool 30L
General Remark hyb data from Sarafidou-04-05-99
Positive Positive BAC[Collapse 78]

bA2A10	bA144P18	bA362O4	bA496D14
bA2M13	bA146I16	bA363H4	bA496H23
bA10G20	bA148C23	bA368I8	bA510B24
bA11P23	bA151I18	bA393N22	bA511I16
bA34D15	bA160I11	bA401P6	bA514G18
bA39E8	bA164M14	bA415J12	bA517O8
bA43F15	bA168M11	bA433K7	bA523D19
bA46G13	bA184C6	bA441O15	bA526M21
bA49H17	bA193B23	bA454C21	bA532L6
bA55F23	bA195F21	bA460M16	bA533O14
bA58C10	bA223O16	bA460N21	bA535B2
bA58N21	bA245D15	bA460N22	bA536G17
bA68J7	bA247E7	bA465H6	bA539D21
bA69F17	bA264B12	bA473K12	bA557J23
bA70L24	bA269A20	bA474H4	bA558K13
bA71F8	bA293L10	bA481N24	bA560G4
bA71N16	bA294C1	bA482A10	bA567B5
bA77F18	bA304L3	bA488C11	bA573B17
bA103A2	bA320F15	bA491K20	bA573L13
	bA360H20	bA492L19	

Positive_BAC_weak [Collapse 11]

bA35H23	
bA56O4	bA426G23
bA198A9	bA450E20
bA283I21	bA456L15
bA323K21	bA460K16
bA415K21	bA480L6

Contains STS stRH816
stRH1014
stAFMb362yg5
stAFMc005xh5
stAFM350wa5
stAFMa272zd1

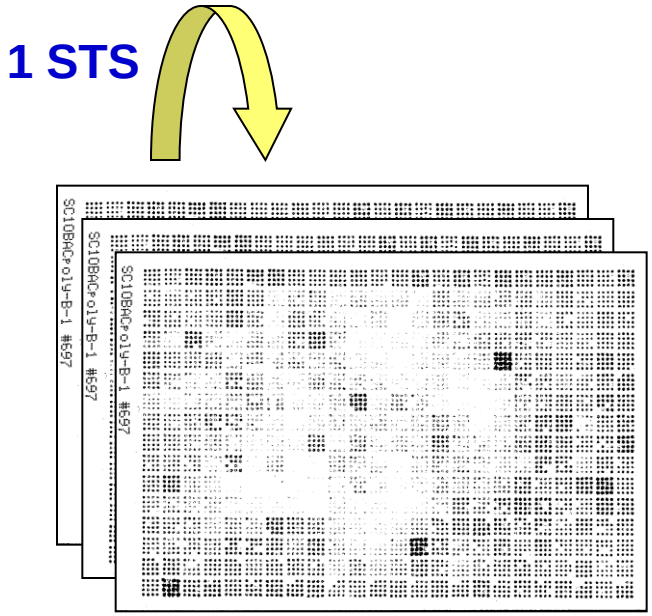
Enter search word:

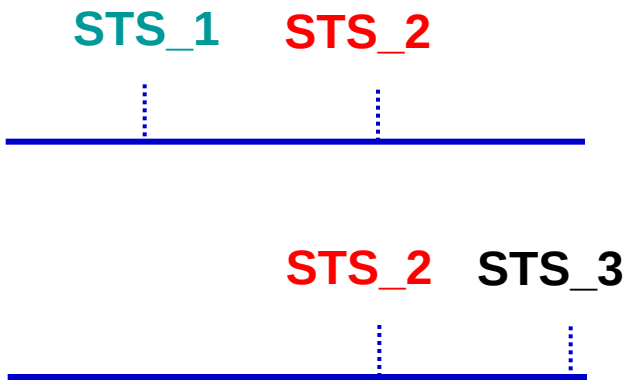
Do full search]

~ 10 BACs / STS

Αντιστοίχηση κλώνων BACs με STSs

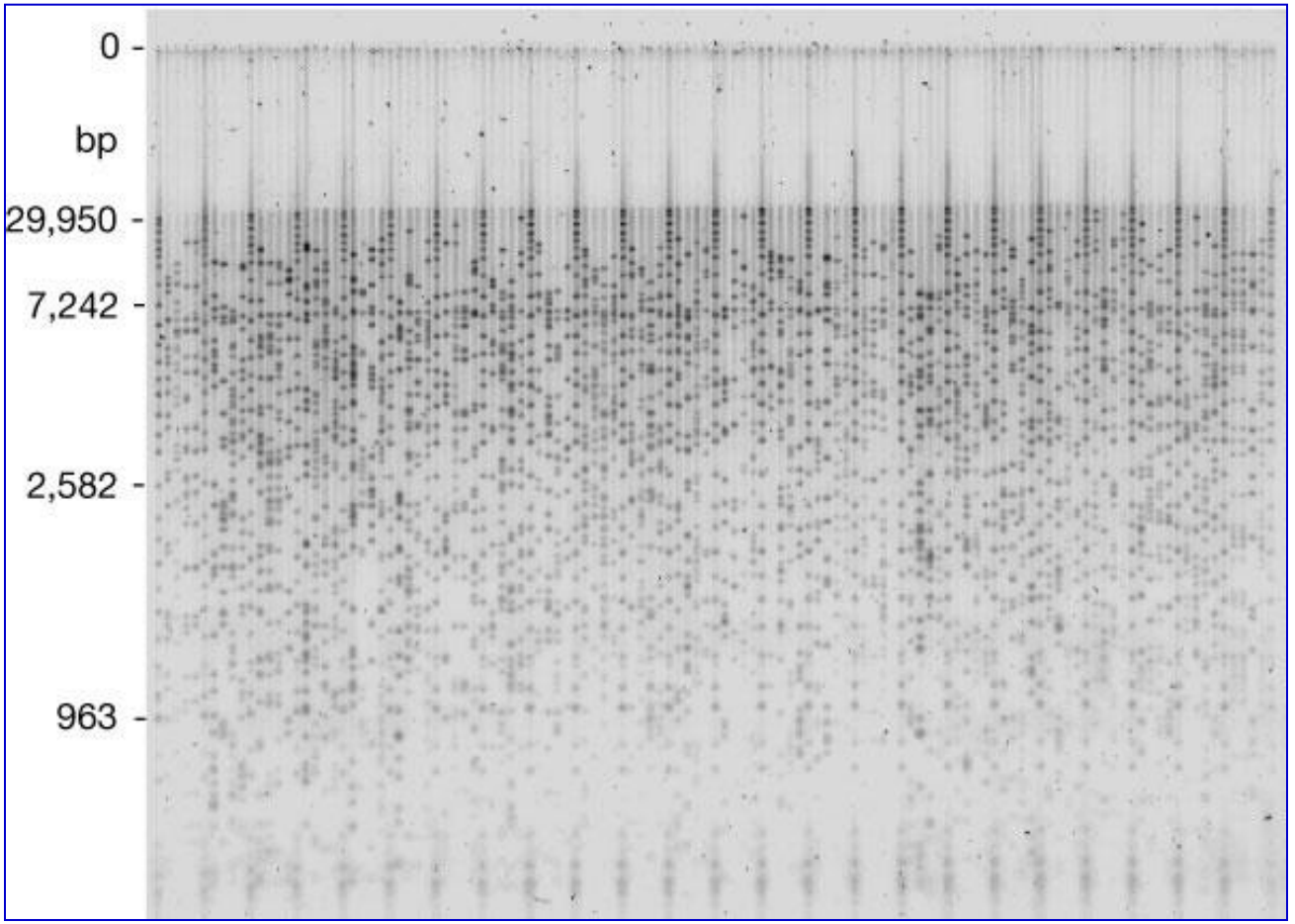
Χρωμοσωμο-ειδική βιβλιοθήκη



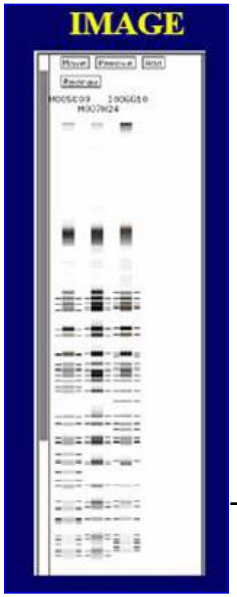


Fingerprinting των κλώνων BAC

(The International Human Genome Mapping Consortium)



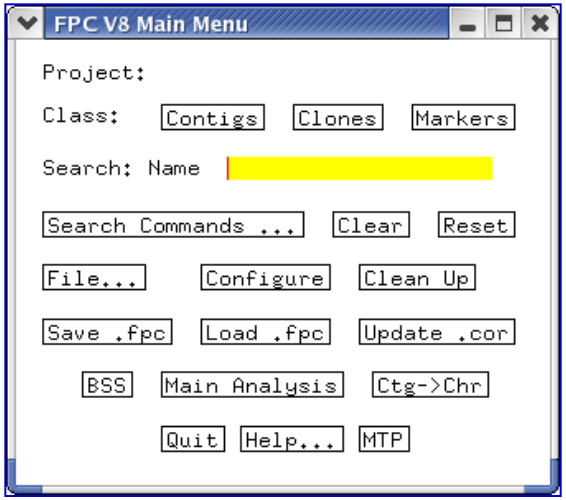
Τεμαχισμός με *HindIII*



→ **FPC db human map**

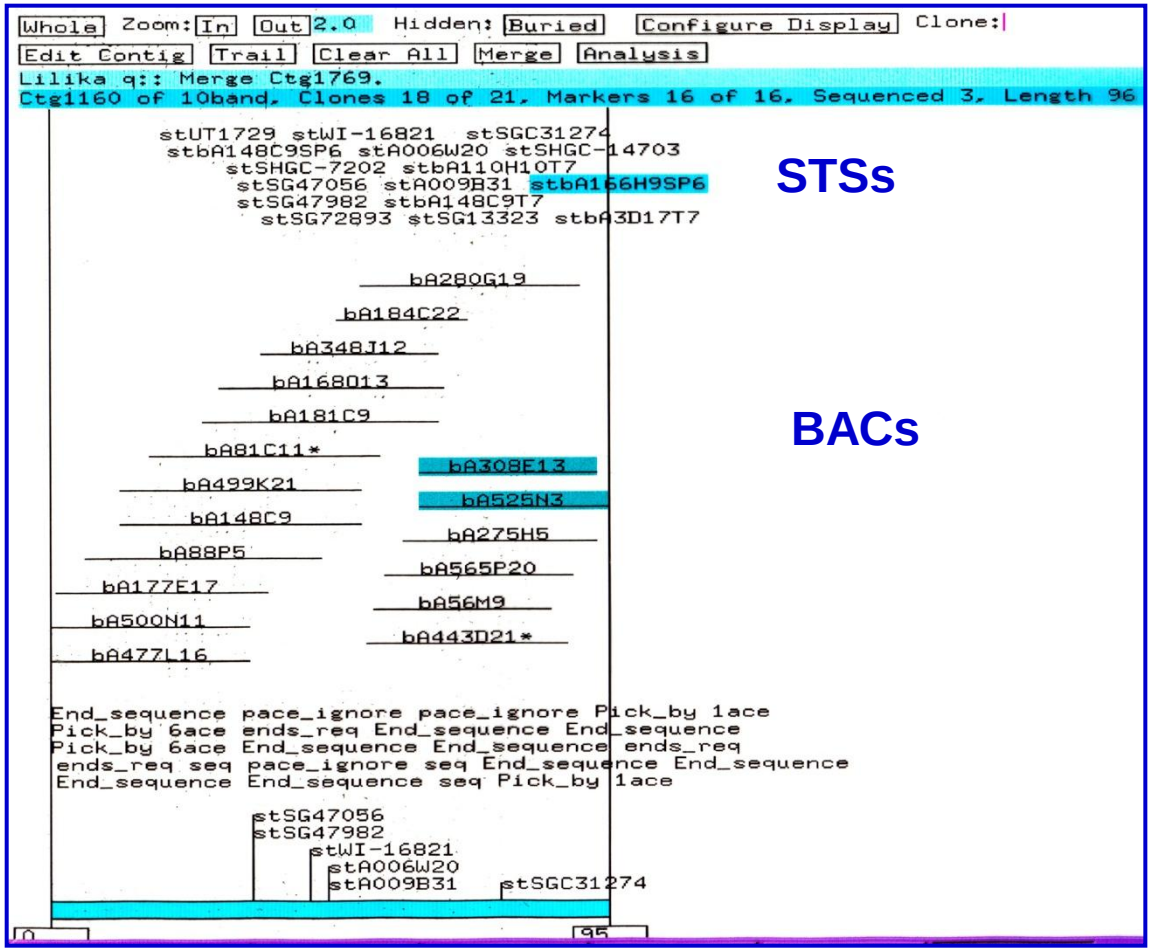
Sulston J *et al*, 1989

FPC db

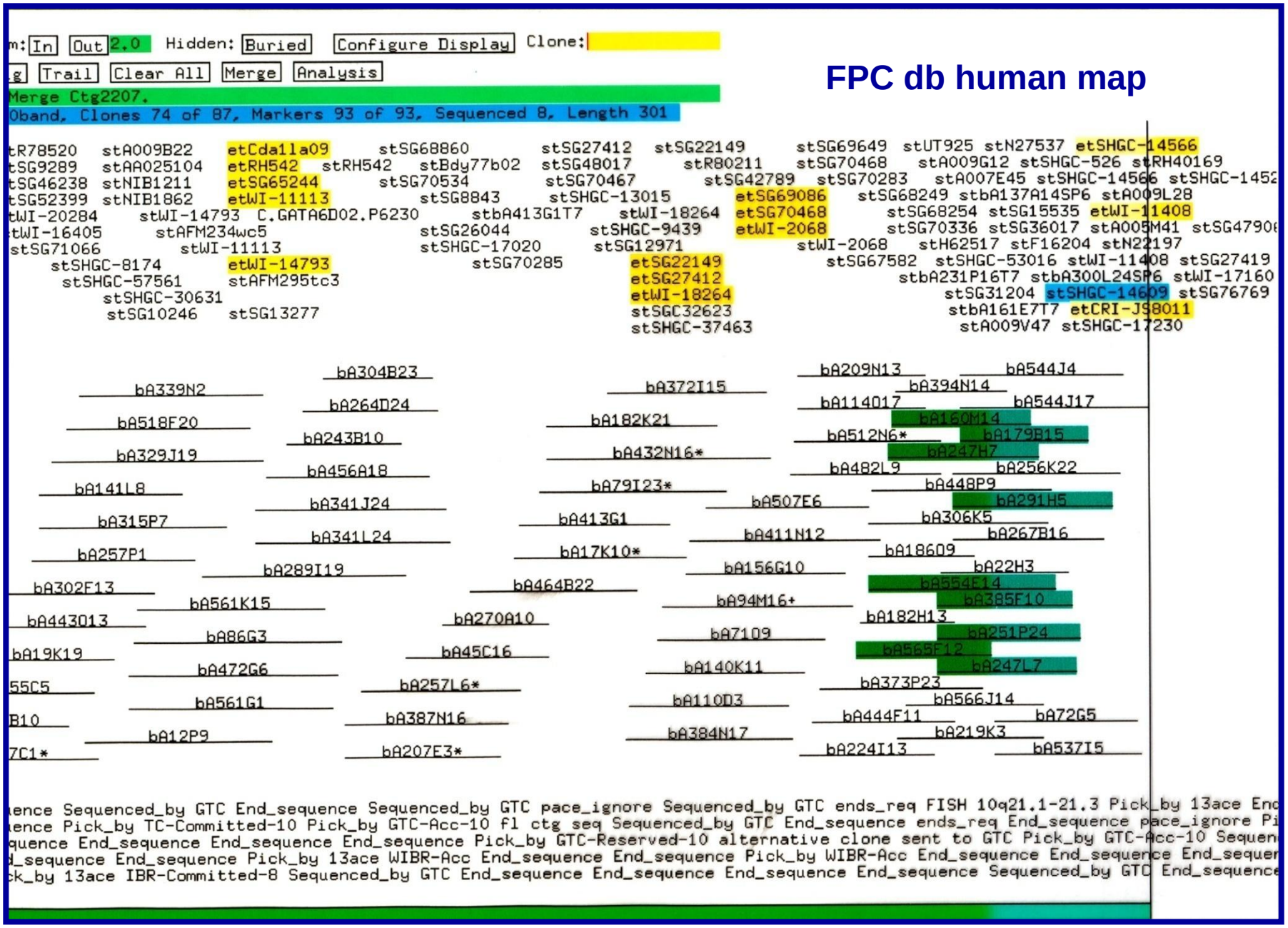


Soderlund C *et al*, 1997 & 2000

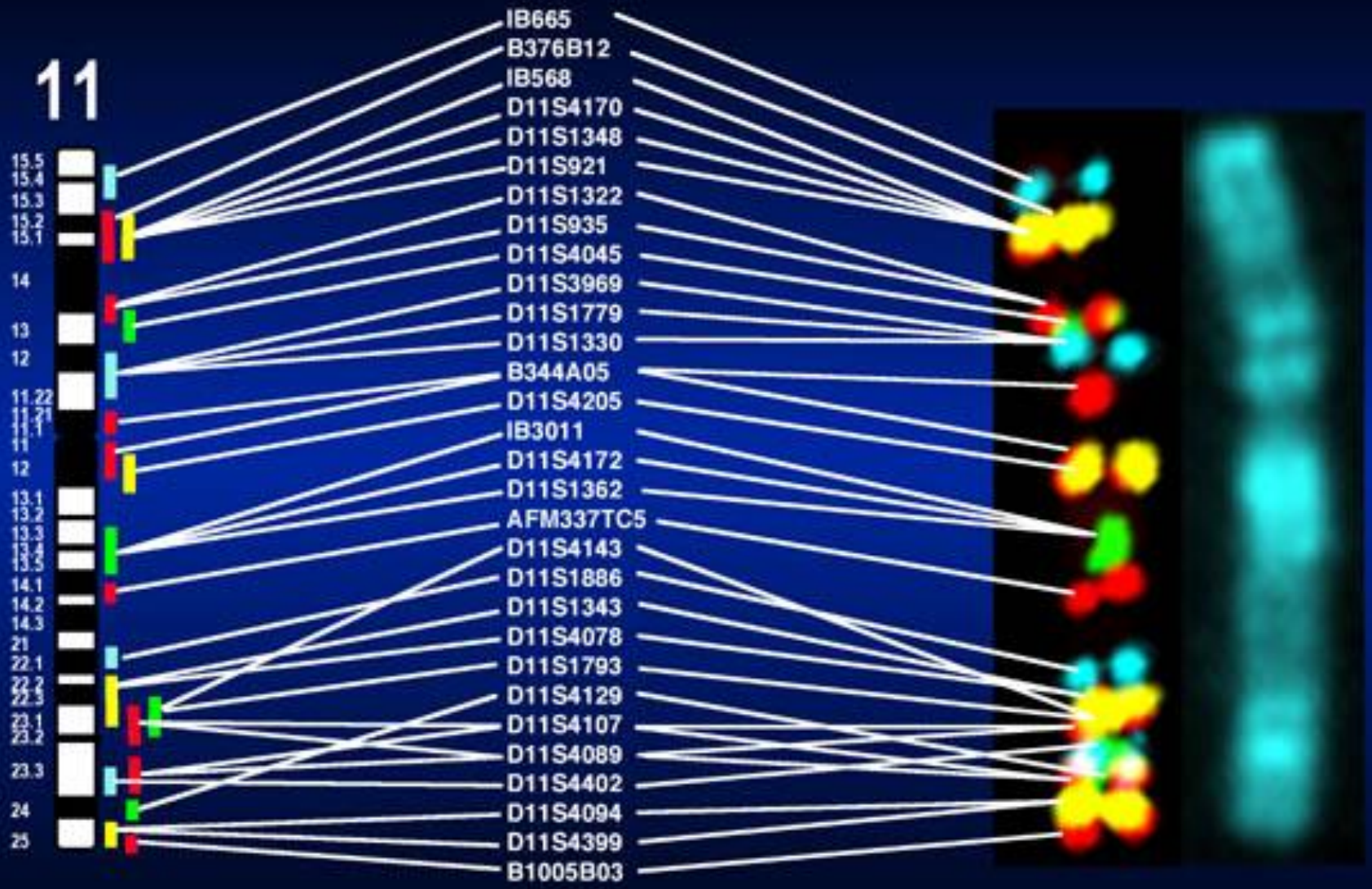
Contigs



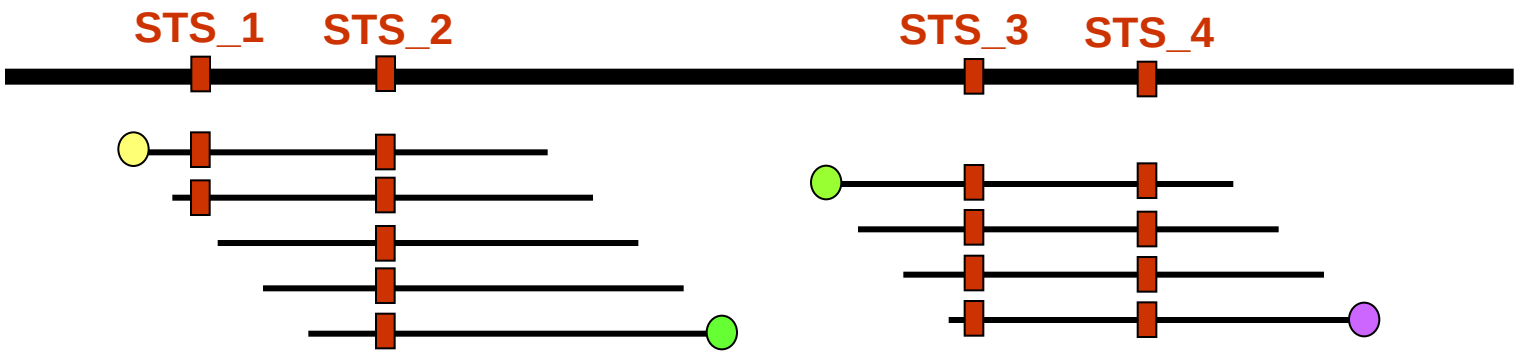
Φυσικός χάρτης γενωμικών κλώνων BAC

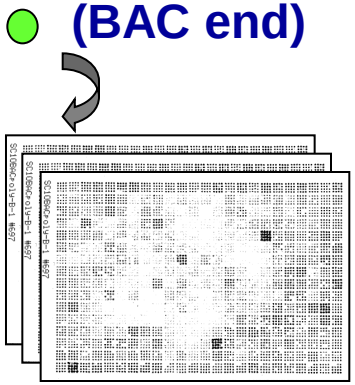


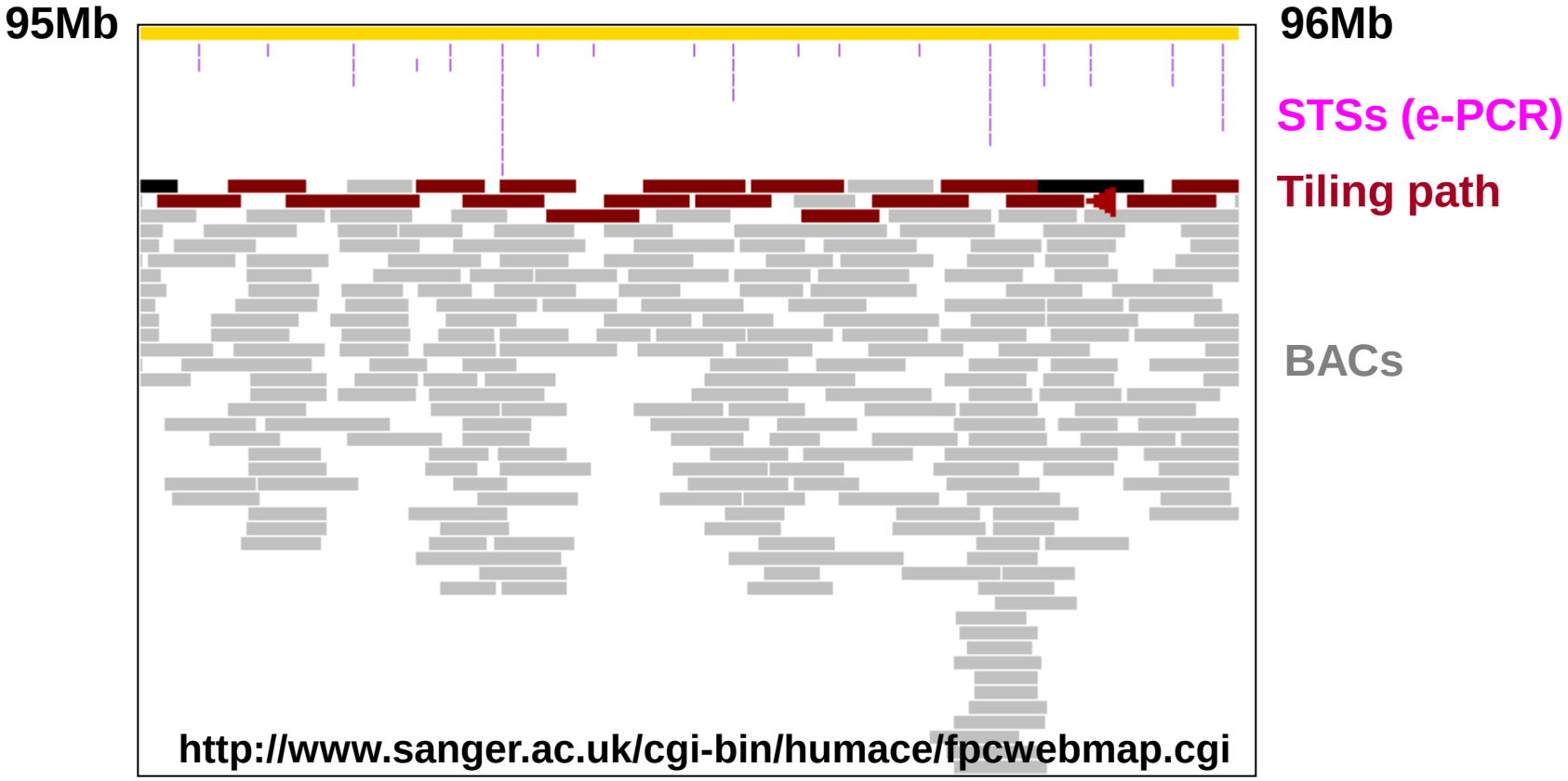
Ανάλυση FISH κλώνων



Χρωμοσωμικό περπάτημα

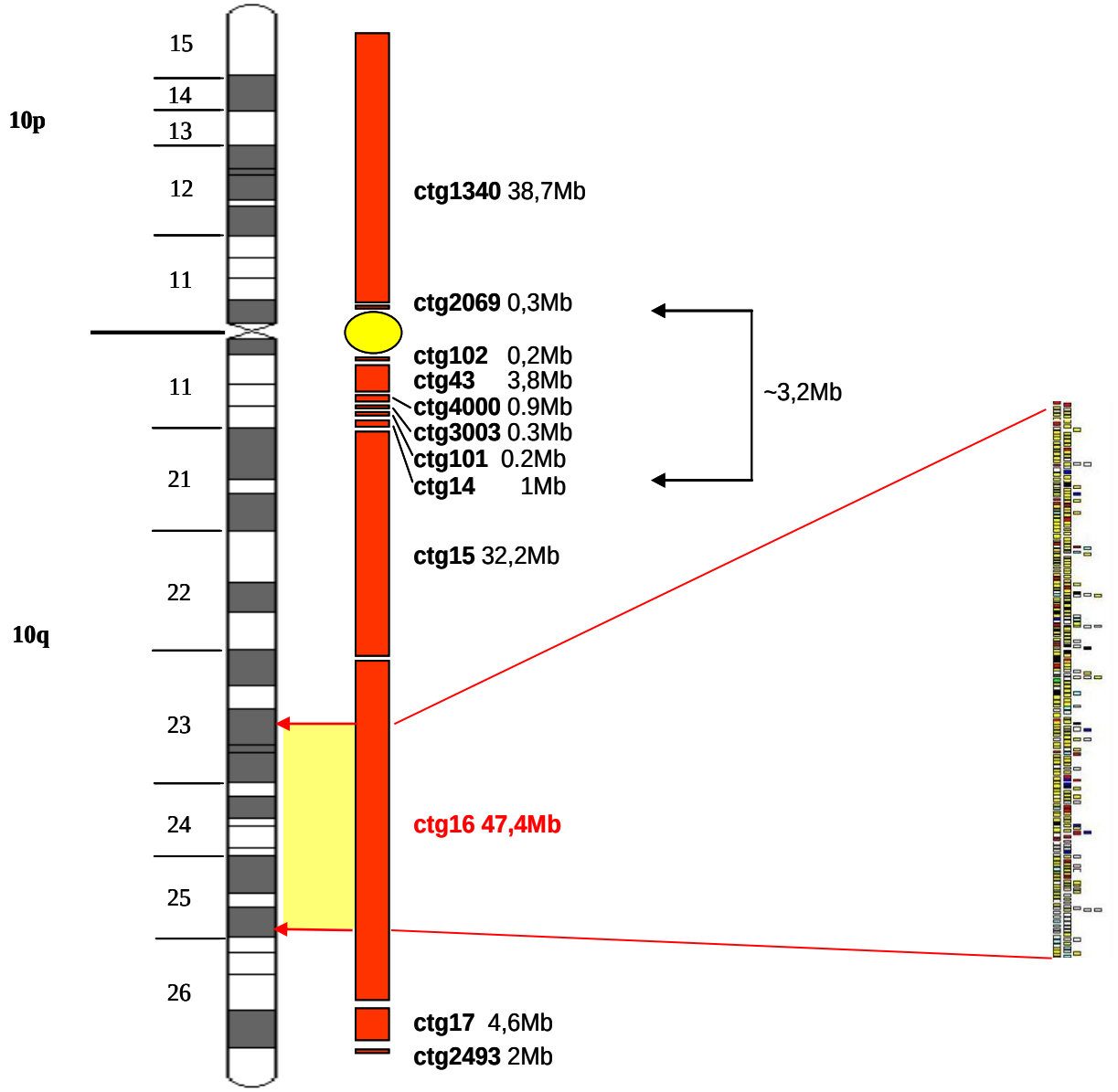






Sanger clone name:	**bA437J2**
International name:	**RP11-437J2**
Species name:	**Homo sapiens**
Chromosome:	**10**
Genbank accession:	**AL157396**
Sequencing pipeline status:	**Analysis**
Sequenced by:	**Sanger Centre**
Library:	**RPCI-11.2**
Library description:	**Human Male BAC Library**
Sanger internal clone prefix:	**bA**
External clone prefix:	**RP11**
Vector:	**pBACe3.6**
Clone type:	**BAC**

Ο φυσικός χάρτης



227 BACs προς αλληλούχηση

Bentley *et al*, 2001. The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature* 409, 942-943)

articles

A physical map of the human genome

The International Human Genome Mapping Consortium*

* *A partial list of authors appears at the end of this paper. A full list is available as Supplementary Information.*

The human genome is by far the largest genome to be sequenced, and its size and complexity present many challenges for sequence assembly. The International Human Genome Sequencing Consortium constructed a map of the whole genome to enable the selection of clones for sequencing and for the accurate assembly of the genome sequence. Here we report the construction of the whole-genome bacterial artificial chromosome (BAC) map and its integration with previous landmark maps and information from mapping efforts focused on specific chromosomal regions. We also describe the integration of sequence data with the map.

The International Human Genome Sequencing Consortium (IHGSC) used a hierarchical mapping and sequencing strategy to construct the working draft of the human genome. This clone-based approach involves generating an overlapping series of clones that covers the entire genome. Each clone is 'fingerprinted' on the basis of the pattern of fragments generated by restriction enzyme digestion[1,2]. Clones are then selected for shotgun sequencing and the whole genome sequence is reconstructed by map-guided assembly of overlapping clone sequences[3].

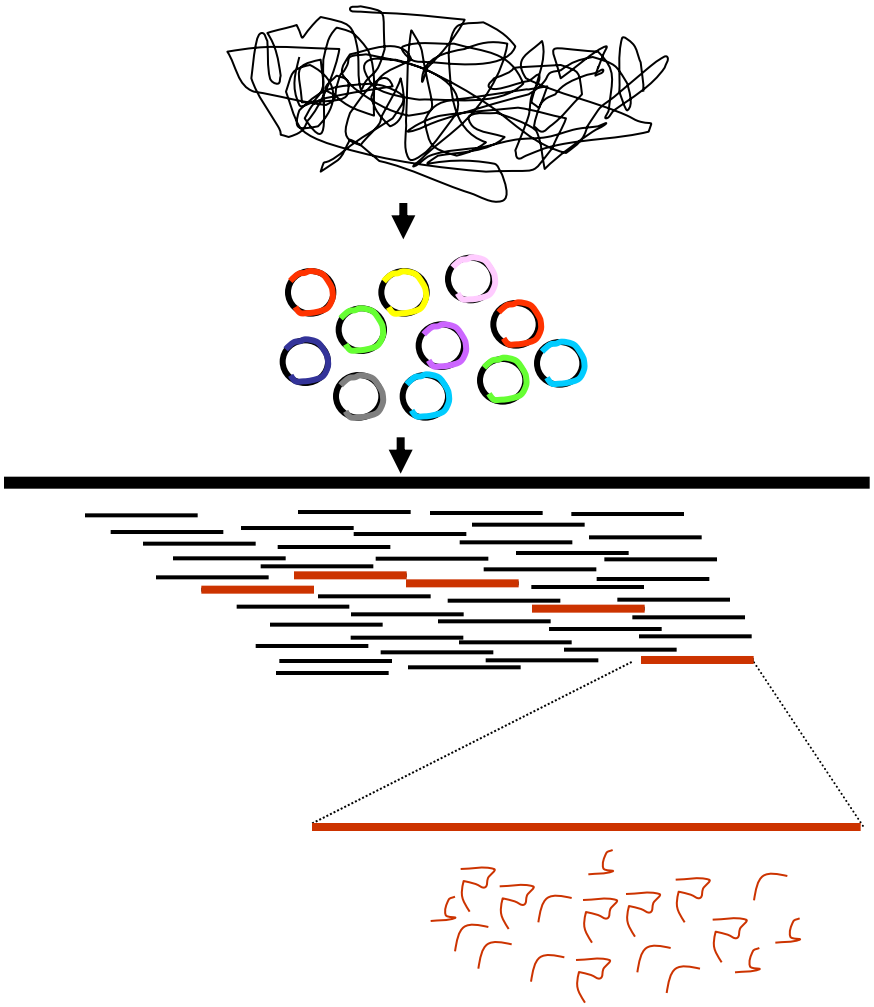
The availability of the whole-genome clone-based map assisted the sequencing of the human genome in many respects. The fingerprinted BAC map made it possible to select clones for sequencing that would ensure comprehensive coverage of the

Construction of the whole-genome BAC map

The pilot phase of the sequencing project began in 1995, at which time efforts were renewed to develop clone-based maps covering specific regions of the genome. To construct these regional maps, we screened PAC and BAC clones for STS markers, fingerprinted the positive clones, integrated them into the existing maps, and selected the largest, intact clones with minimal overlap for sequencing.

To keep pace with the ramping up of the sequencing effort in 1998, the ongoing efforts to construct the whole-genome BAC map were increased approximately tenfold. The whole-genome BAC map was constructed in several steps. First we collected fingerprint data for a large sample of random clones from a genome-wide BAC library. We then assembled the BAC map first by using the

Αλληλούχηση βάσει χρωμοσωμικού χάρτη (ιεραρχική shotgun αλληλούχηση)



γενωμικό DNA

βιβλιοθήκη

contigs

Minimal Tiling Path

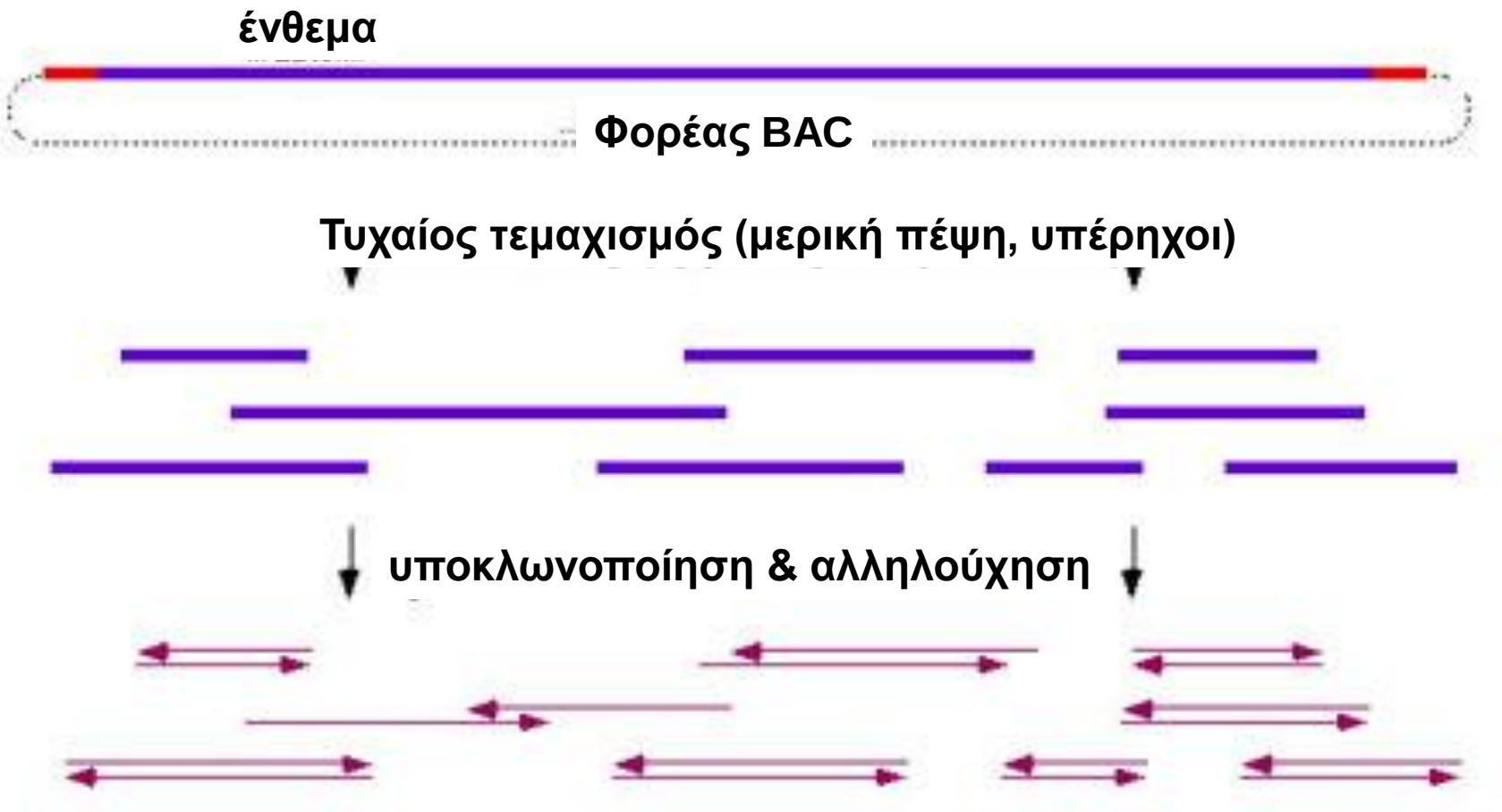
επιλεγμένος κλώνος

shotgun

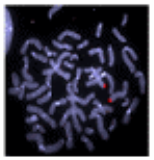
…AGGCTTGAACGTTGCAACCAGCTGAC
TTGCAACCAGCTGACGTTAACCGTAGGCTA…

…AGGCTTGAACGTTGCAACCAGCTGACGTTAACCGTAGGCTA…

Αλληλούχηση shotgun ενός κλώνου BAC

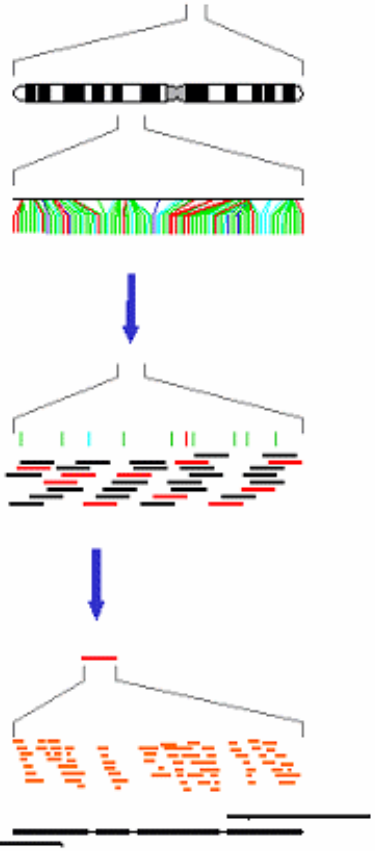


STRATEGIES FOR SEQUENCING THE HUMAN GENOME



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).

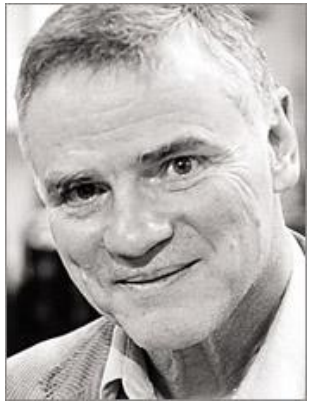


BY WHOLE GENOME SHOTGUN

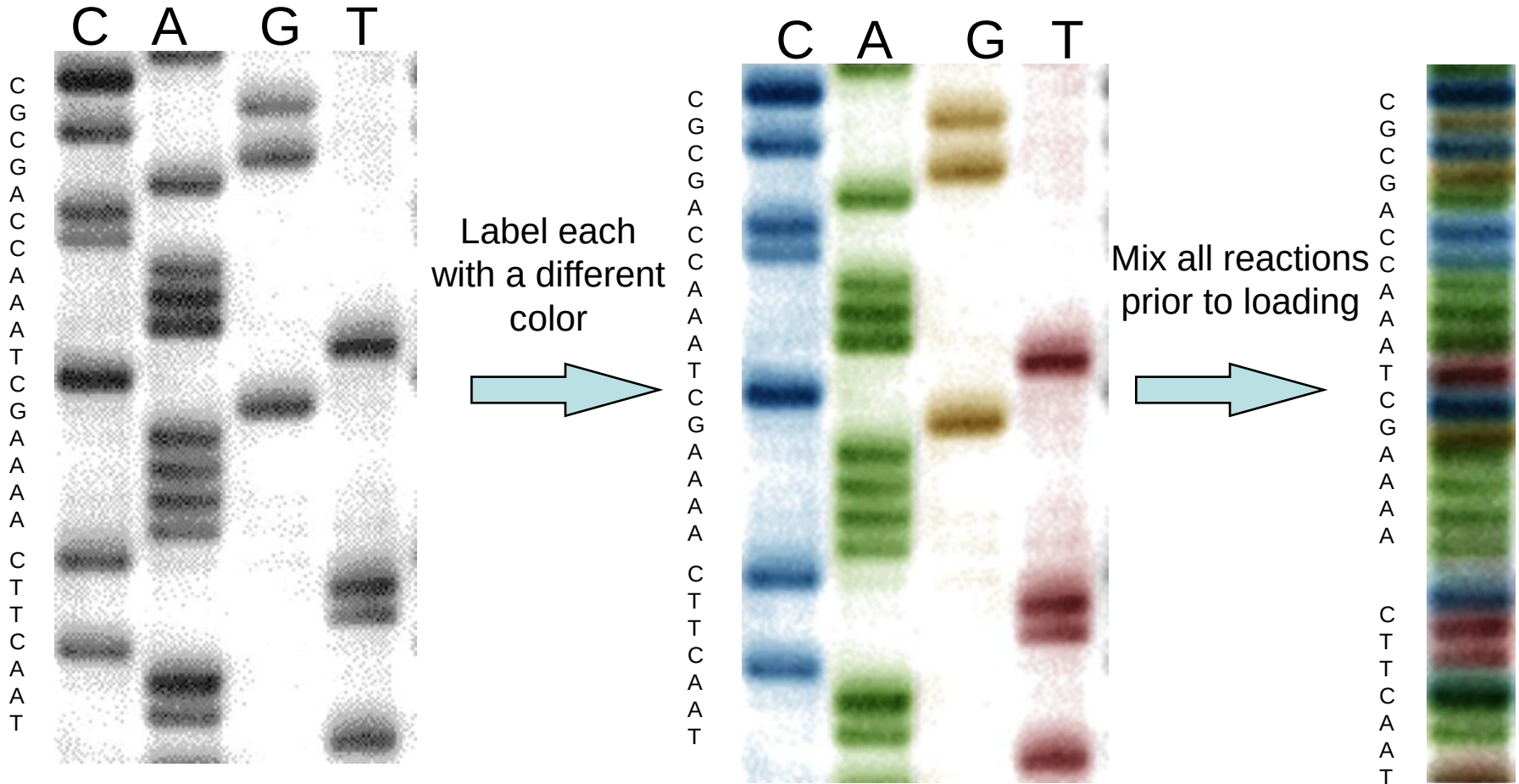
1. Shotgun sequencing of short-insert clones
2. Paired end sequencing of large-insert clones
3. Assembly of seed contigs (unitigs)
4. Incorporation of other sequences, and integration of long- range data.

J. Graig Venter

1986: Αντικατάσταση της ραδιενεργής σήμανσης με φθορίζουσα

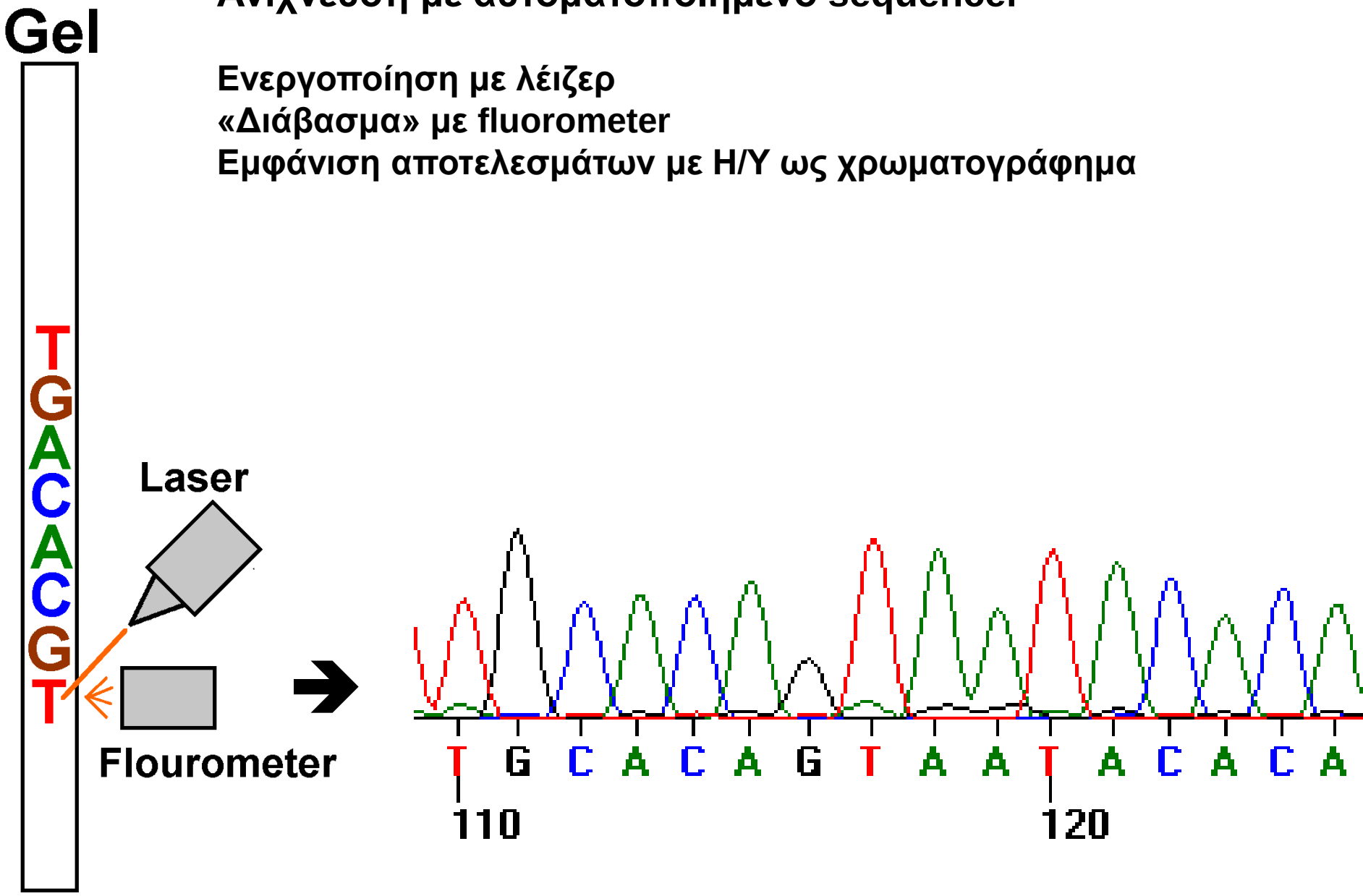


Με 4 διαφορετικές χρωστικές, όλες οι αντιδράσεις μπορούν να ηλεκτροφορηθούν μαζί

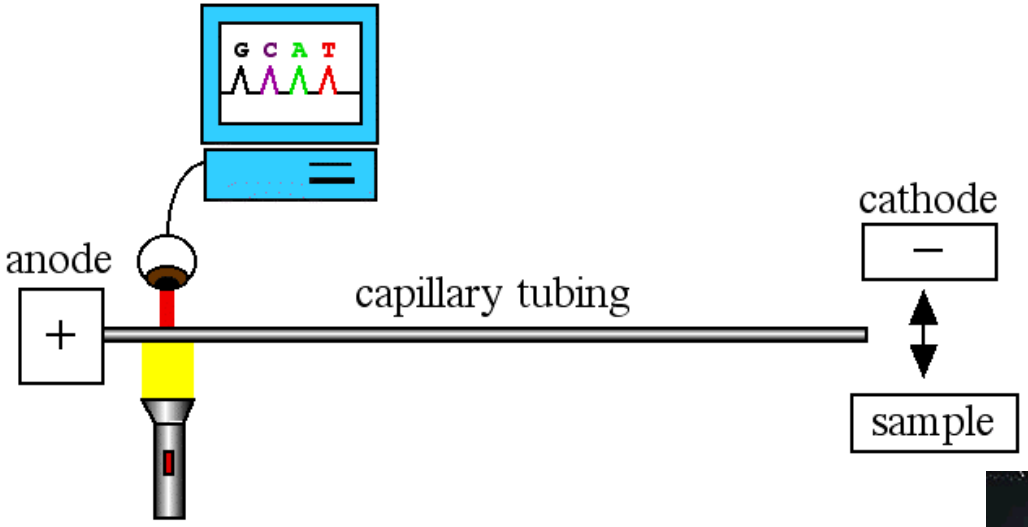


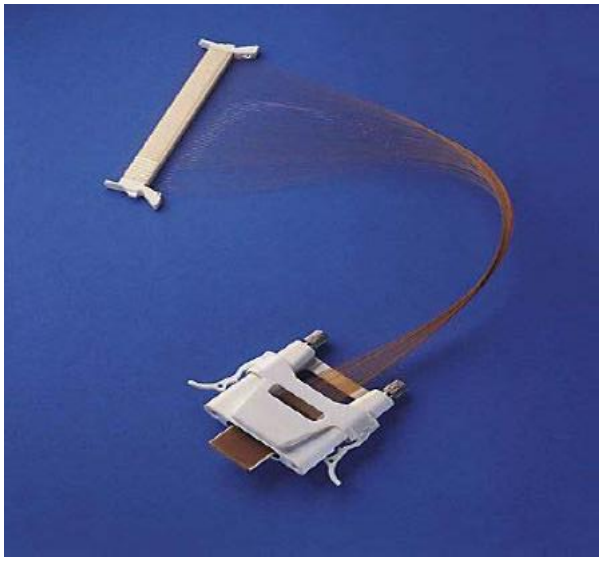
Ανίχνευση με αυτοματοποιημένο sequencer

Ενεργοποίηση με λέιζερ
«Διάβασμα» με fluorometer
Εμφάνιση αποτελεσμάτων με Η/Υ ως χρωματογράφημα

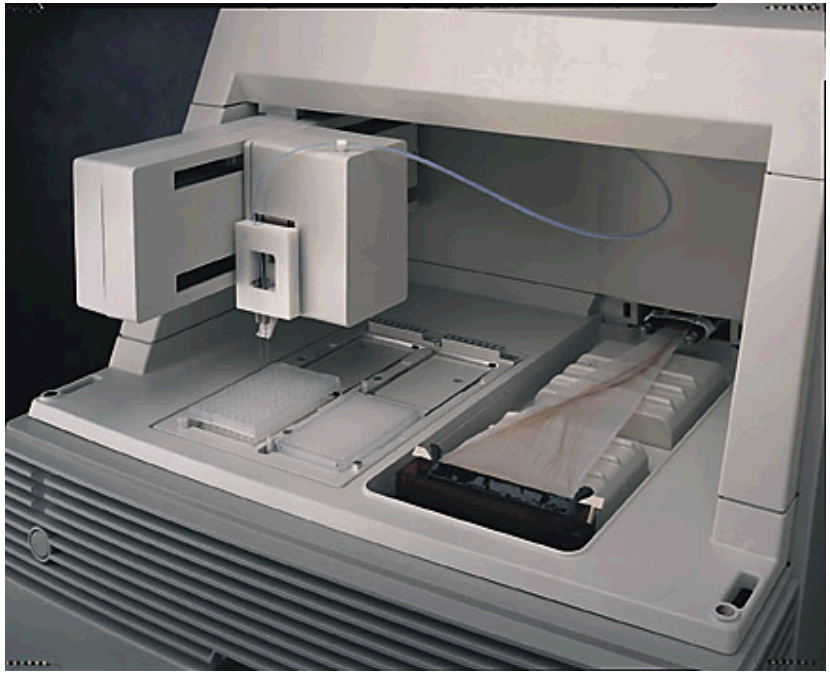


Sequencer με τριχοειδή



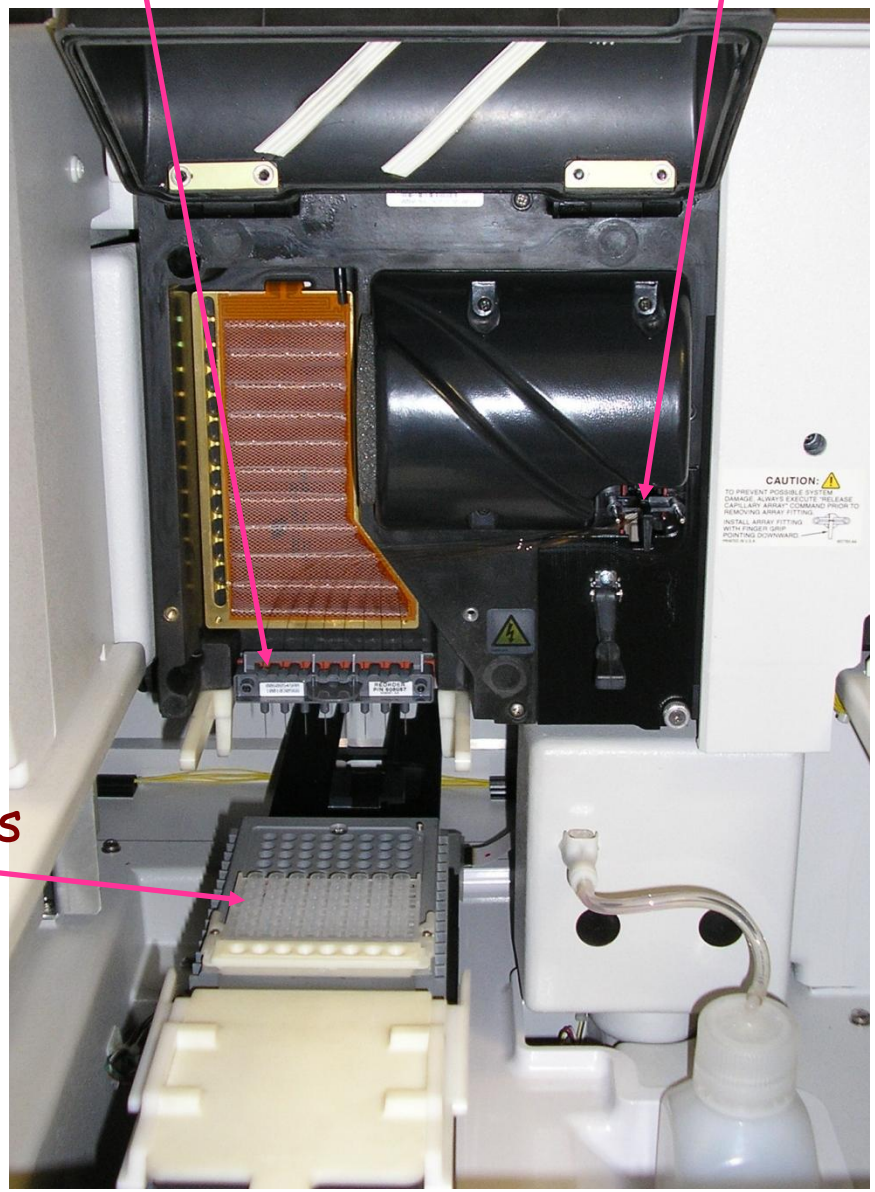


τριχοειδή

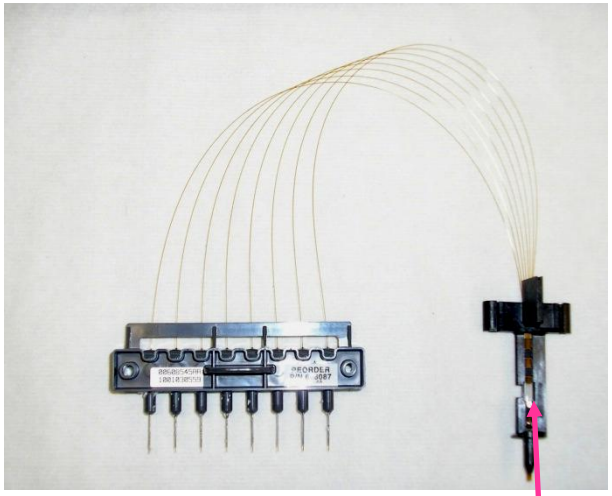


capillary

Laser / detector



samples



Detection window

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

ΠΟΡΕΙΑ ΤΟΥ HGP

1985 1988 1990 1996 1999 2001 2004 2005

Έναρξη

Γενετικοί χάρτες

Φυσικοί χάρτες

Αλληλούχηση cDNAs

ESTs

Πλήρους μήκους

Αλληλούχηση γενωμικού DNA

pilot

draft

finished

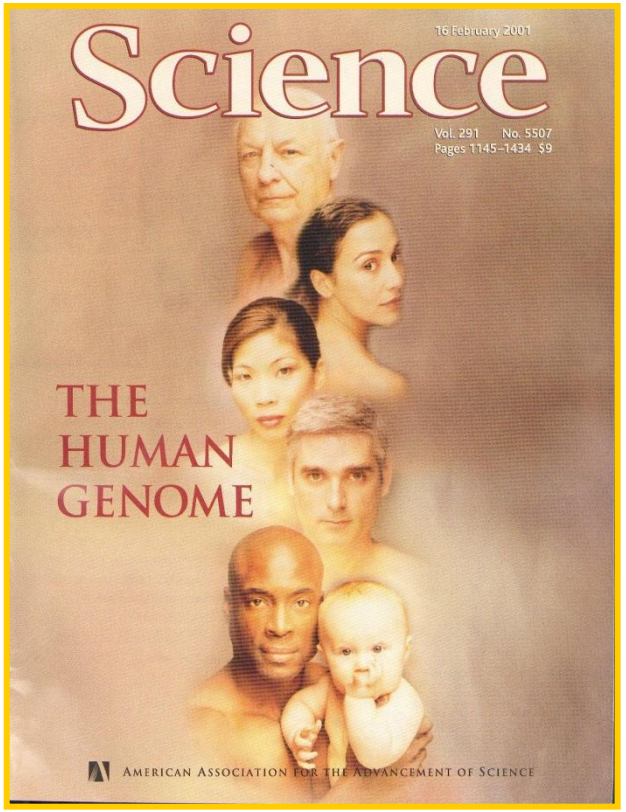
chr. 22

Κάλυψη: 4x - 5x
90% του γονιδιώματος

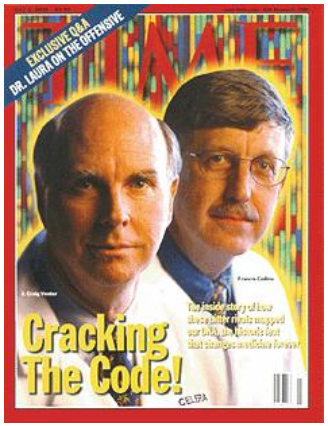
Κάλυψη: 8x - 9x,
99% του γονιδιώματος
Ακρίβεια: 99,9%

Vol 409, Feb 15, 2001

Vol 291, Feb 16, 2001



Human Genome Project
HGP
FREE ACCESS



CELERA GENOMICS

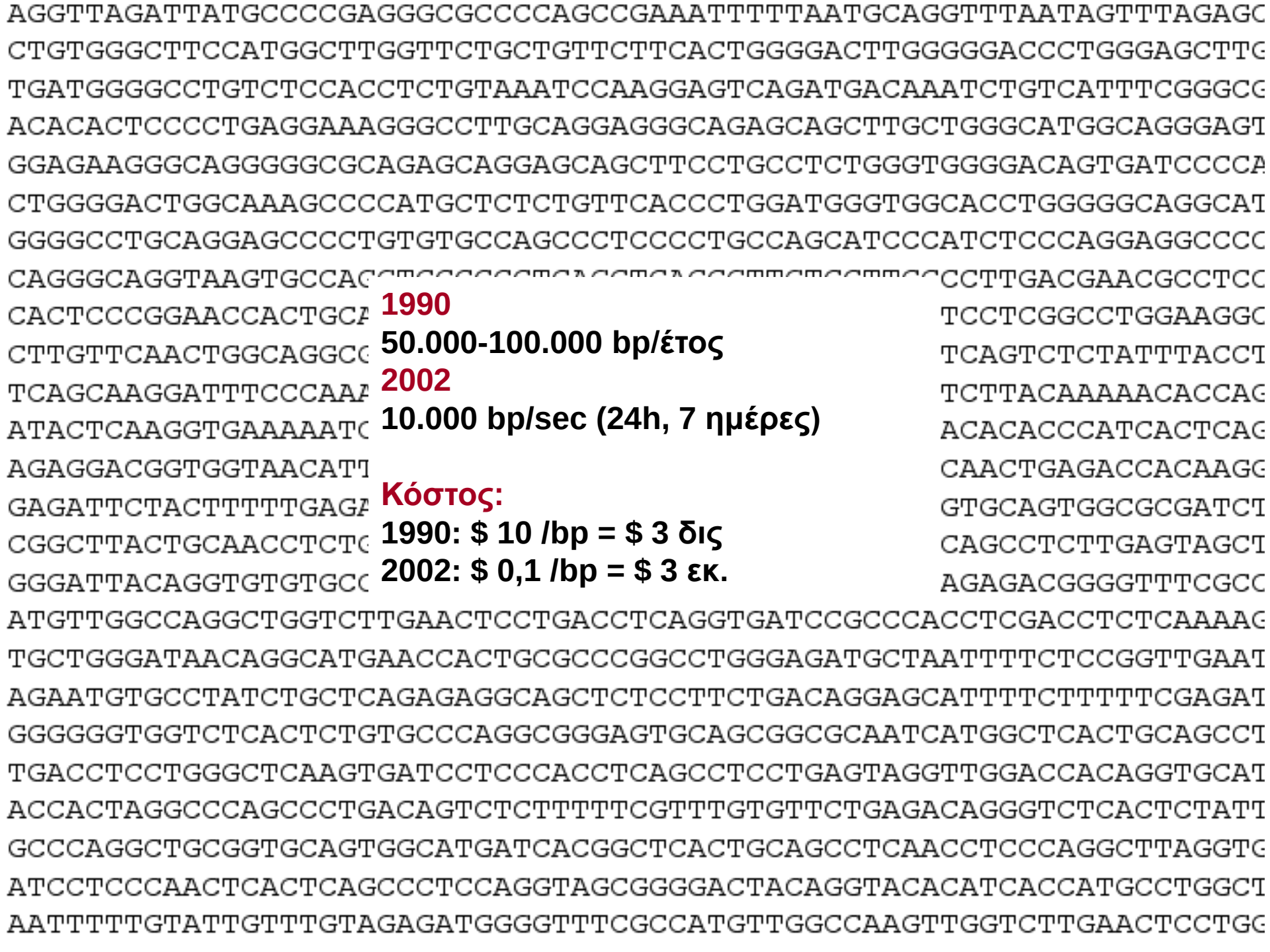
Access to db upon subscription

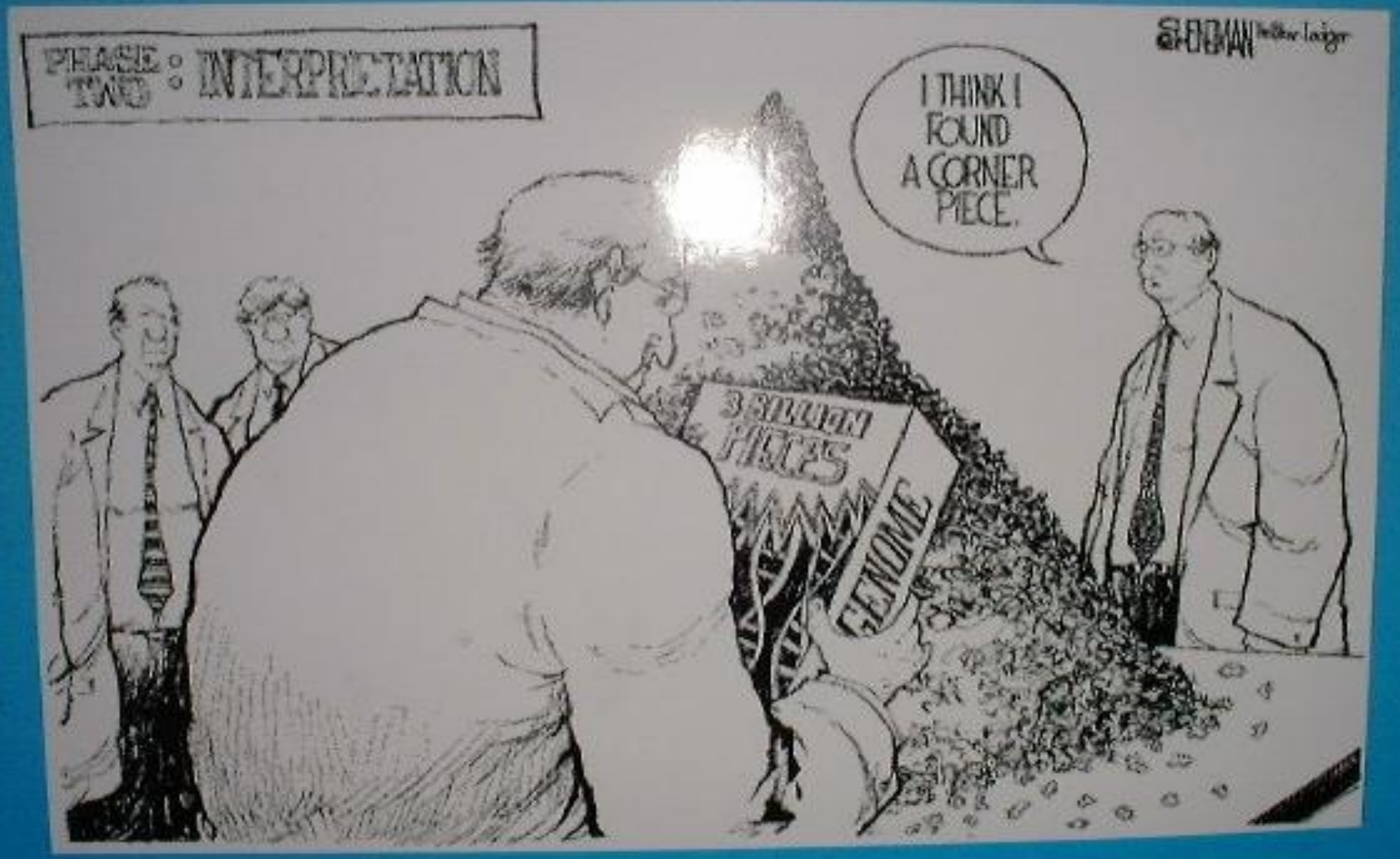
Τώρα για πρώτη φορά διαθέτουμε μια ιστορική ανθολογία του εαυτού μας, ένα μέρος της οποίας μεταβιβάζεται από γενιά σε γενιά επί ένα δισεκατομμύριο χρόνια. Μόλις τώρα μάθαμε να διαβάζουμε την ιστορία και είναι σίγουρο πως θα μας συναρπάσει για πολλές ακόμα δεκαετίες.
Eric Lander, Whitehead Institute

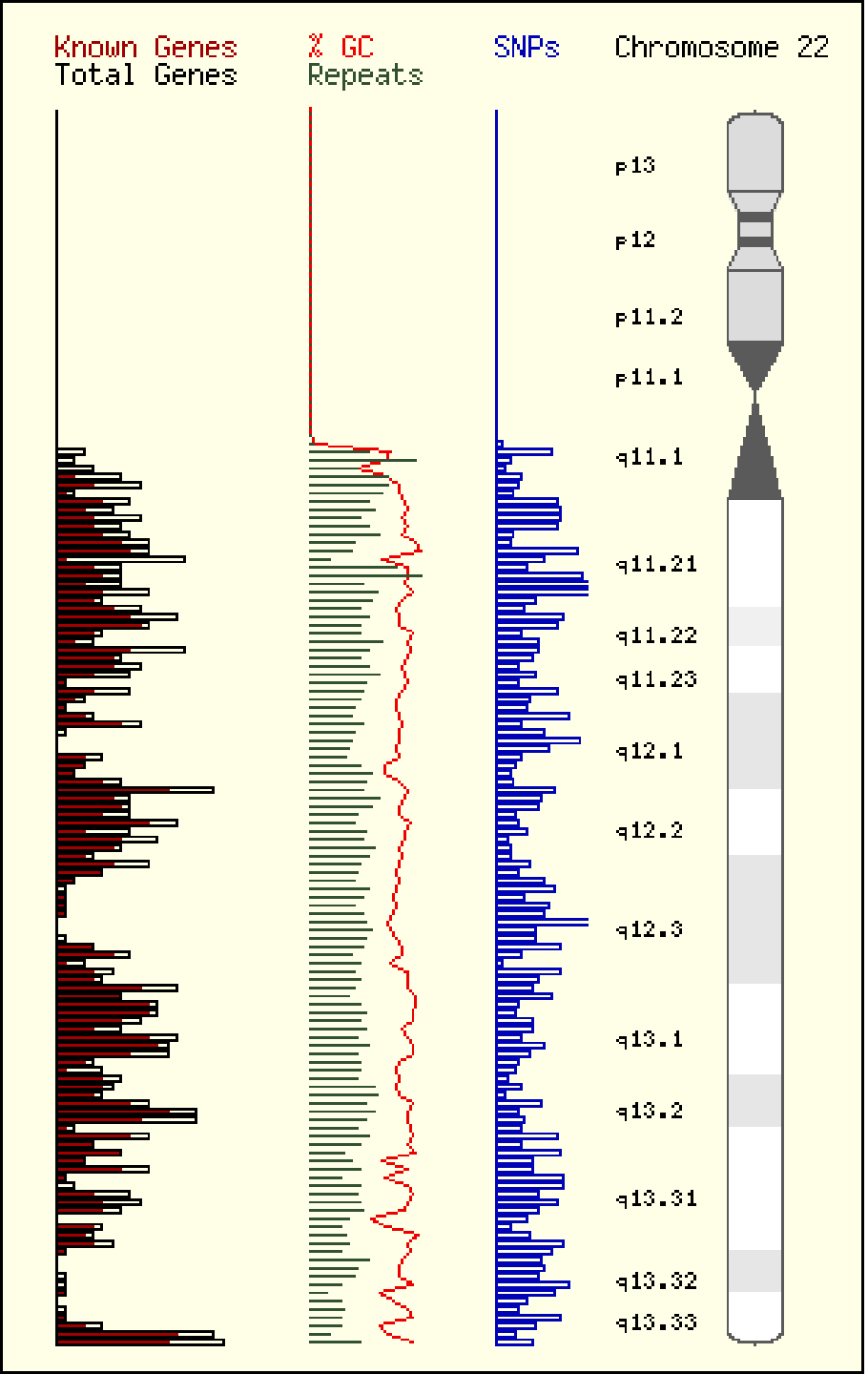
Το γονίδιο είναι, με διαφορά, το πιο περίτεχνο πρόγραμμα
Bill Gates, Microsoft

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

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

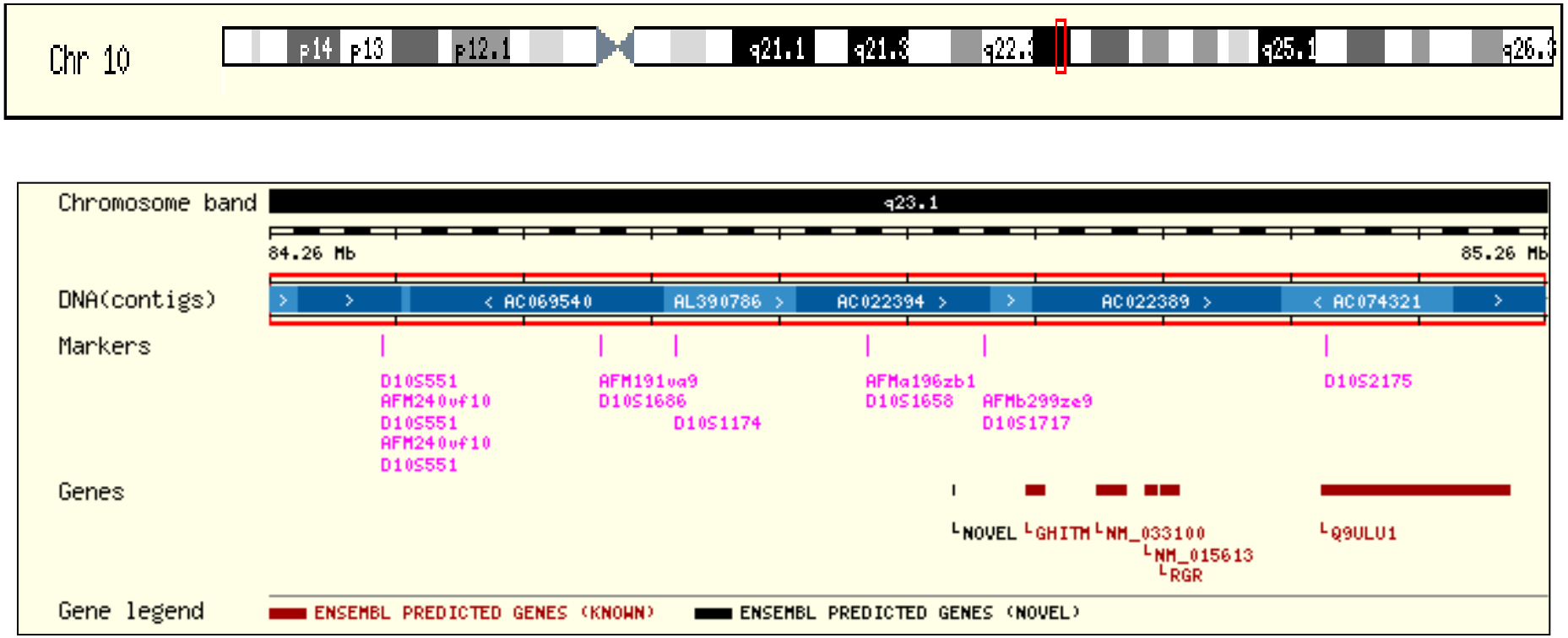

1990
50.000-100.000 bp/έτος
2002
10.000 bp/sec (24h, 7 ημέρες)
Κόστος:
1990: $ 10 /bp = $ 3 δις
2002: $ 0,1 /bp = $ 3 εκ.

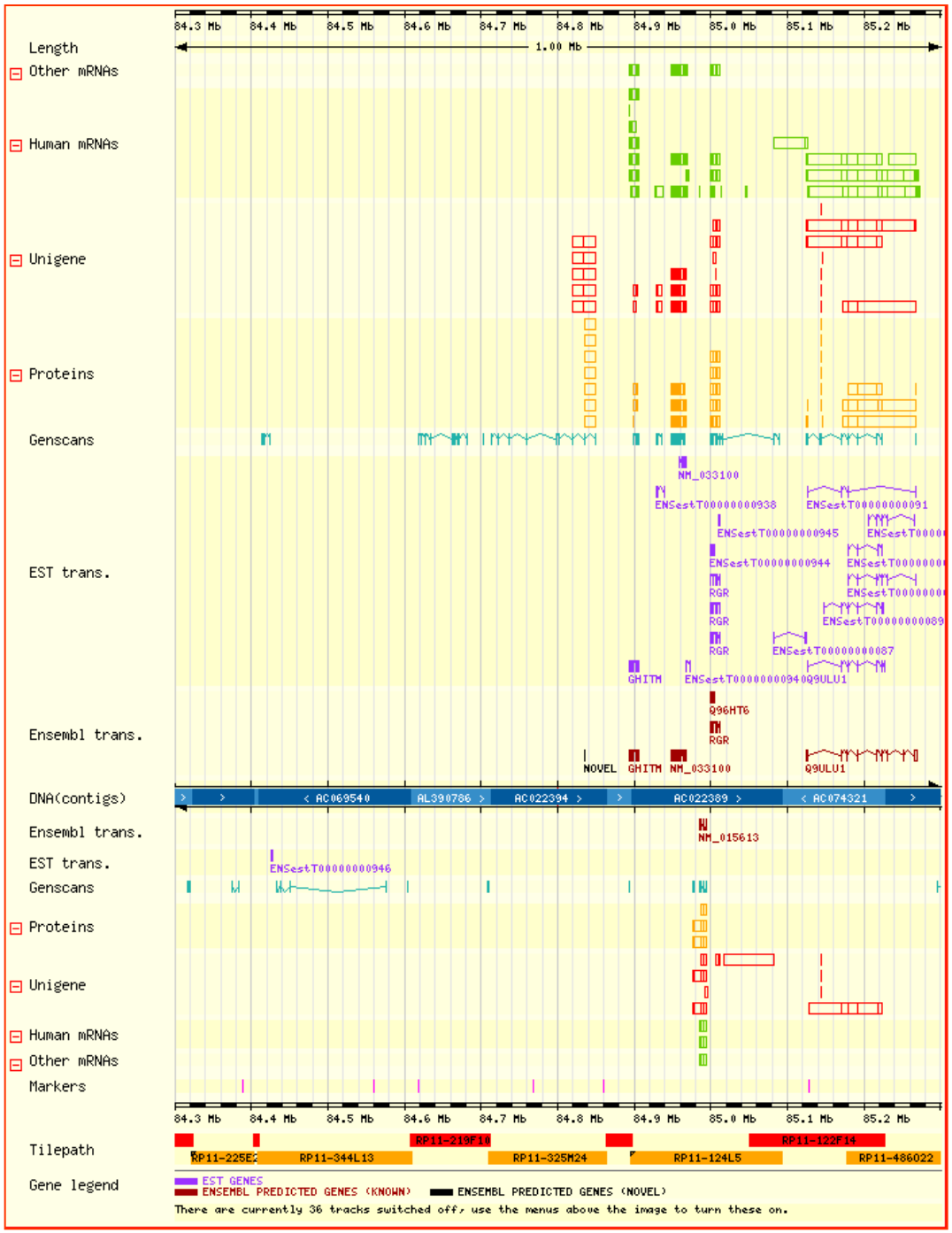

PHASE TWO: INTERPRETATION
I THINK I FOUND A CORNER PIECE.
3 BILLION PIECES
GENOME

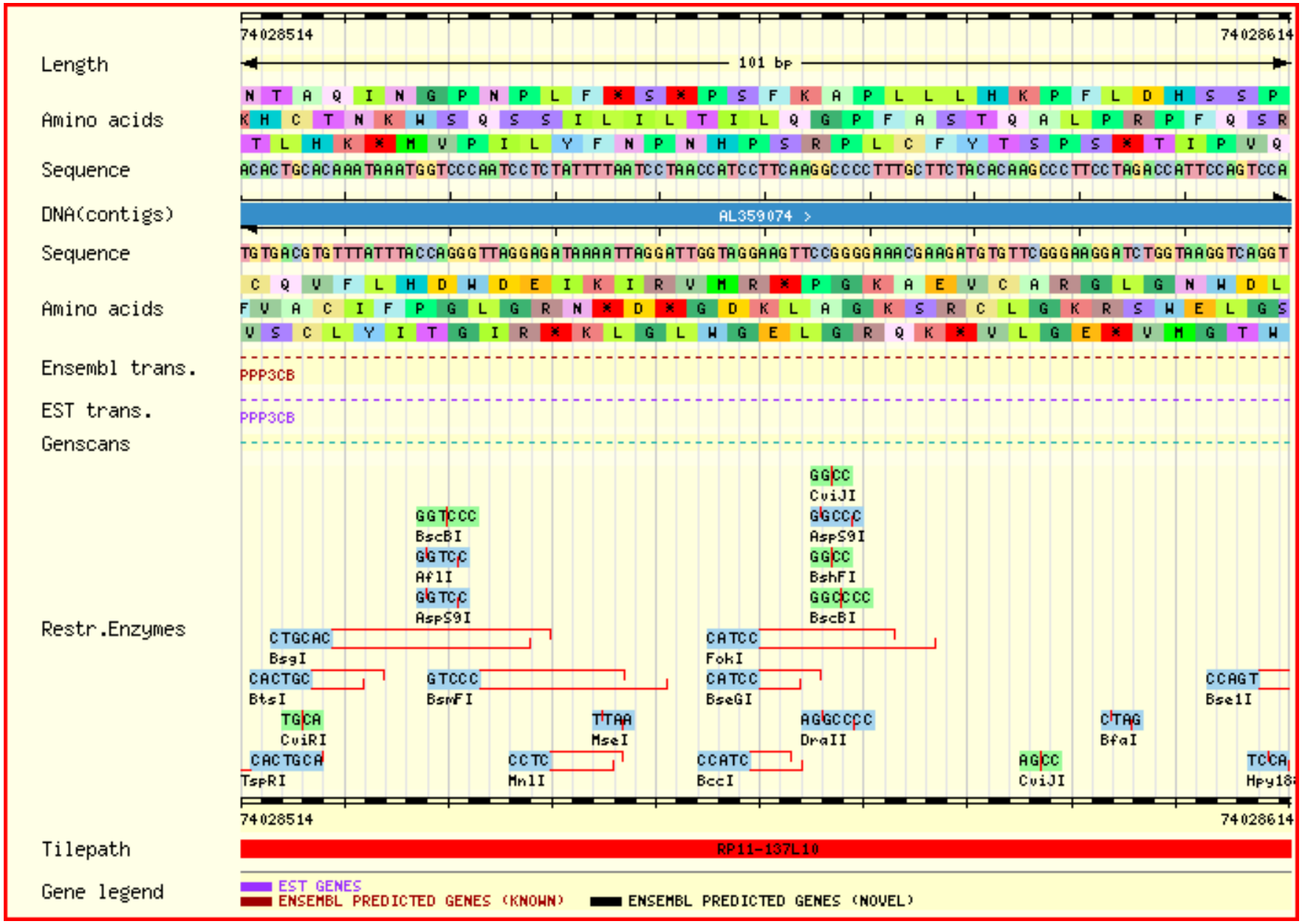


Ανάλυση γενωμικών αλληλουχιών για την ταυτοποίηση νέων γονιδίων, *in silico*

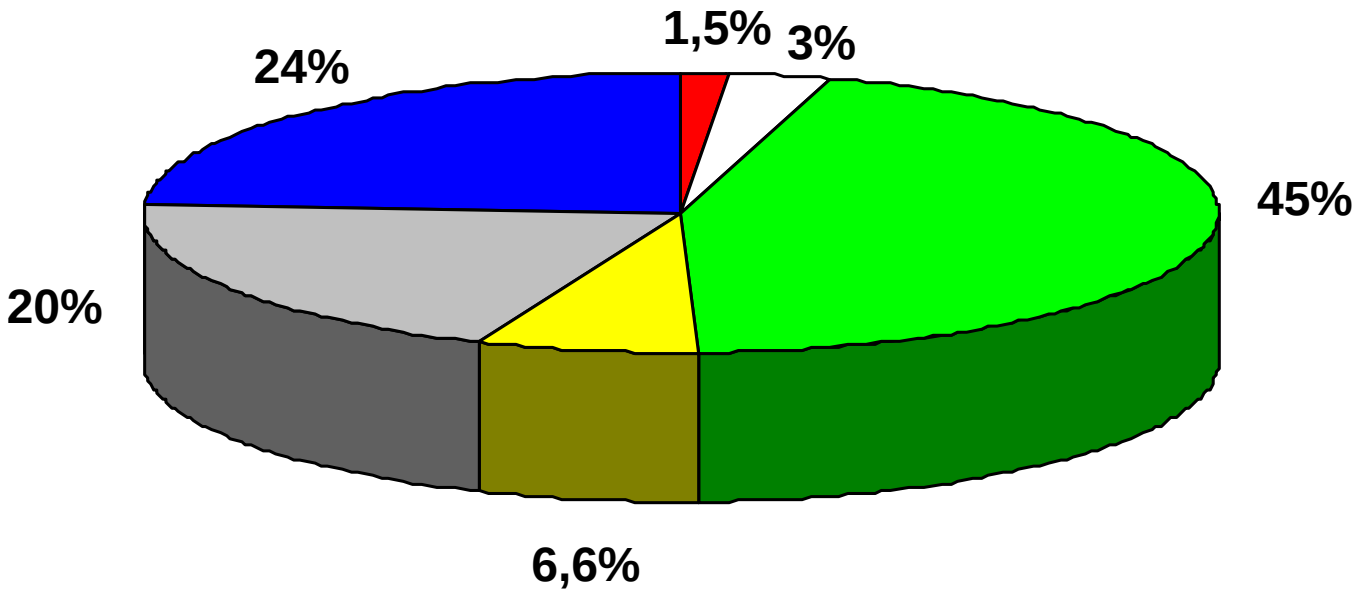
http://www.ensembl.org.


Chr 10
p14
p13
p12.1
q21.1
q21.3
q22.3
q25.1
q26.3
Chromosome band
q23.1
84.26 Mb
85.26 Mb
DNA(contigs)
AC069540
AL390786
AC022394
AC022389
AC074321
Markers
D10S551
AFM240vf10
D10S551
AFM240vf10
D10S551
AFM191va9
D10S1686
D10S1174
AFMa196zb1
D10S1658
AFMb299ze9
D10S1717
D10S2175
Genes
NOVEL
GHITM
NM_033100
NM_015613
RGR
Q9ULU1
Gene legend
ENSEMBL PREDICTED GENES (KNOWN)
ENSEMBL PREDICTED GENES (NOVEL)


84.3 Mb
84.4 Mb
84.5 Mb
84.6 Mb
84.7 Mb
84.8 Mb
84.9 Mb
85.0 Mb
85.1 Mb
85.2 Mb
Length
1.00 Mb
Other mRNAs
Human mRNAs
Unigene
Proteins
Genscans
NM_033100
ENSestT00000000938
ENSestT00000000091
ENSestT00000000945
ENSestT00000000944
RGR
ENSestT00000000089
ENSestT00000000087
GHITM
ENSestT00000000940
Q9ULU1
EST trans.
Q96HT6
RGR
Ensembl trans.
NOVEL
GHITM
NM_033100
Q9ULU1
DNA(contigs)
AC069540
AL390786
AC022394
AC022389
AC074321
Ensembl trans.
NM_015613
EST trans.
ENSestT00000000946
Genscans
Proteins
Unigene
Human mRNAs
Other mRNAs
Markers
84.3 Mb
84.4 Mb
84.5 Mb
84.6 Mb
84.7 Mb
84.8 Mb
84.9 Mb
85.0 Mb
85.1 Mb
85.2 Mb
Tilepath
RP11-219F10
RP11-122F14
RP11-225E2
RP11-344L13
RP11-325M24
RP11-124L5
RP11-486022
Gene legend
EST GENES
ENSEMBL PREDICTED GENES (KNOWN)
ENSEMBL PREDICTED GENES (NOVEL)
There are currently 36 tracks switched off; use the menus above the image to turn these on.


74028514
74028614
Length
101 bp
Amino acids
Sequence
DNA(contigs)
AL359074 >
Sequence
Amino acids
Ensembl trans.
PPP3CB
EST trans.
PPP3CB
Genscans
Restr.Enzymes
GGTCCC
BscBI
GGTCC
AflI
GGTCC
AspS9I
GGCC
CviJI
GGCCC
AspS9I
GGCC
BshFI
GGCCCC
BscBI
CTGCAC
BsgI
CACTGC
BtsI
TGCA
CviRI
CACTGCA
TspRI
GTCCC
BsmFI
TTAA
MseI
CCTC
MnlI
CATCC
FokI
CATCC
BseGI
AGGCCCC
DraII
CCATC
BccI
CTAG
BfaI
AGCC
CviJI
CCAGT
Bse1I
TCCA
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

Published online 6 May 2004
[DOI: 10.1126/science.1098119]

Previous Article • Table of Contents • Next Article

Submitted on March 19, 2004
Accepted on April 27, 2004

Ultraconserved Elements in the Human Genome

Gill Bejerano [1*], Michael Pheasant [2], Igor Makunin [2], Stuart Stephen [2], W. James Kent [1], John S. Mattick [2], David Haussler [3*]

[1] Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA 95064, USA.
[2] ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia.
[3] Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA.

[*] To whom correspondence should be addressed.
Gill Bejerano , E-mail: jill@soe.ucsc.edu
David Haussler , E-mail: haussler@soe.ucsc.edu

There are 481 segments longer than 200 bp that are absolutely conserved (100% identity with no insertions or deletions) between orthologous regions o
Nearly all of these segments are also conserved in the chicken and dog genomes, with an average of 95% and 99% identity, respectively. Many are also
ultraconserved elements of the human genome are most often located either overlapping exons in genes involved in RNA processing or in introns or nea
transcription and development. Along with more than 5,000 sequences of over 100bp that are absolutely conserved among the three sequenced mamm
elements whose functions and evolutionary origins are yet to be determined, but which are more highly conserved between these species than proteins, a
ontogeny of mammals and other vertebrates.

ΠΟΣΑ ΓΟΝΙΔΙΑ ΚΩΔΙΚΟΠΟΙΟΥΝ ΠΡΩΤΕΪΝΕΣ;

Κινητική επαναδιάταξης
40.000

Νησίδια CpG
80.000

ESTs
35.000-120.000

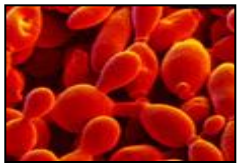
Συγκριτικές μελέτες με *Tetraodon nigroviridis* 28.000-34.000

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

Ανάλυση 2004
20.000-25.000

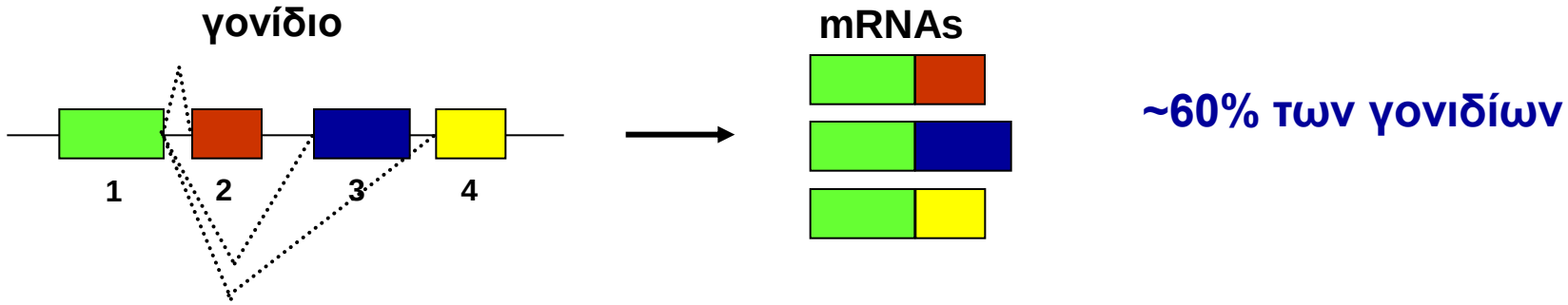
Ανάλυση 2001
30.000-40.000

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



~60% των γονιδίων

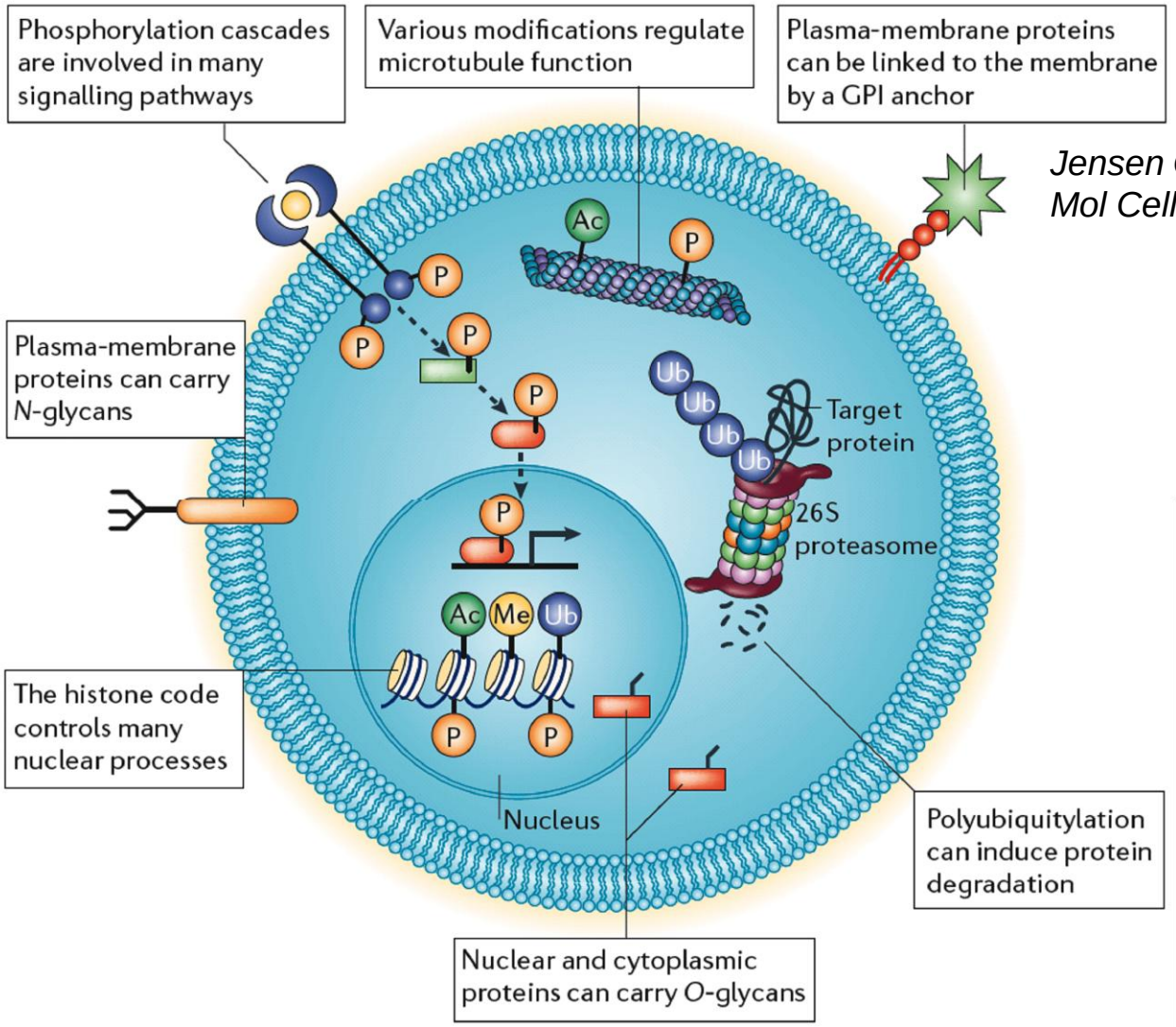
- Chr 22: 642 μετάγραφα αντιστοιχούν σε 245 γονίδια
  - 2,6 μετάγραφα/ γονίδιο
- Chr 19: 1859 μετάγραφα αντιστοιχούν σε 544 γονίδια

  -- 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. Επινόηση 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

Science 19 May 2000:
Vol. 288. no. 5469, pp. 1146 - 1147

NEWS OF THE WEEK

HUMAN GENOME PROJECT:

And the Gene Number Is ...?

Elizabeth Pennisi

COLD SPRING HARBOR, NEW YORK--Even though a draft sequence of the human genome is nearing completion, biologists still don't know how many genes it contains. Indeed, the range of estimates seems to be growing rather than shrinking. The question lies at the core of our understanding of genetic complexity. If genomes are the books of life, then genes are the words that tell the story of each organism. Biologists have long assumed that microorganisms are short stories and complex organisms such as humans, great tomes.

Place your bet. Uncertainty over the number of human genes has sparked a debate--and a betting pool

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 least one transcript must be expressed outside of the nucleus and one transcript must encode a protein (*see footnotes*).

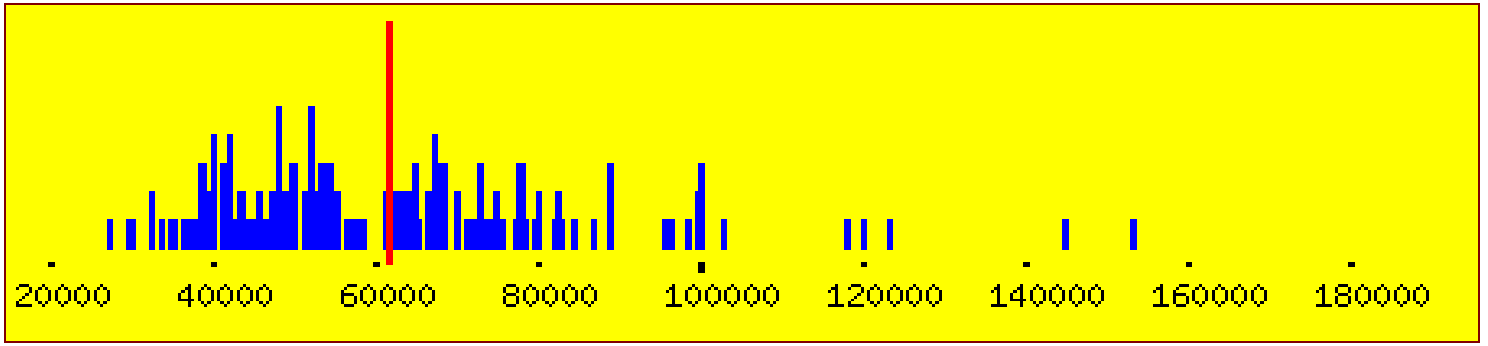
•Assessment of the **method used to determine the gene will occur by voting at Cold Spring Harbor Genome Meeting 2002. Researchers will be invited to submit their methods to the community at this time.**

•Assessment of the gene number will occur on the 2003 CSHL Genome meeting

•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)

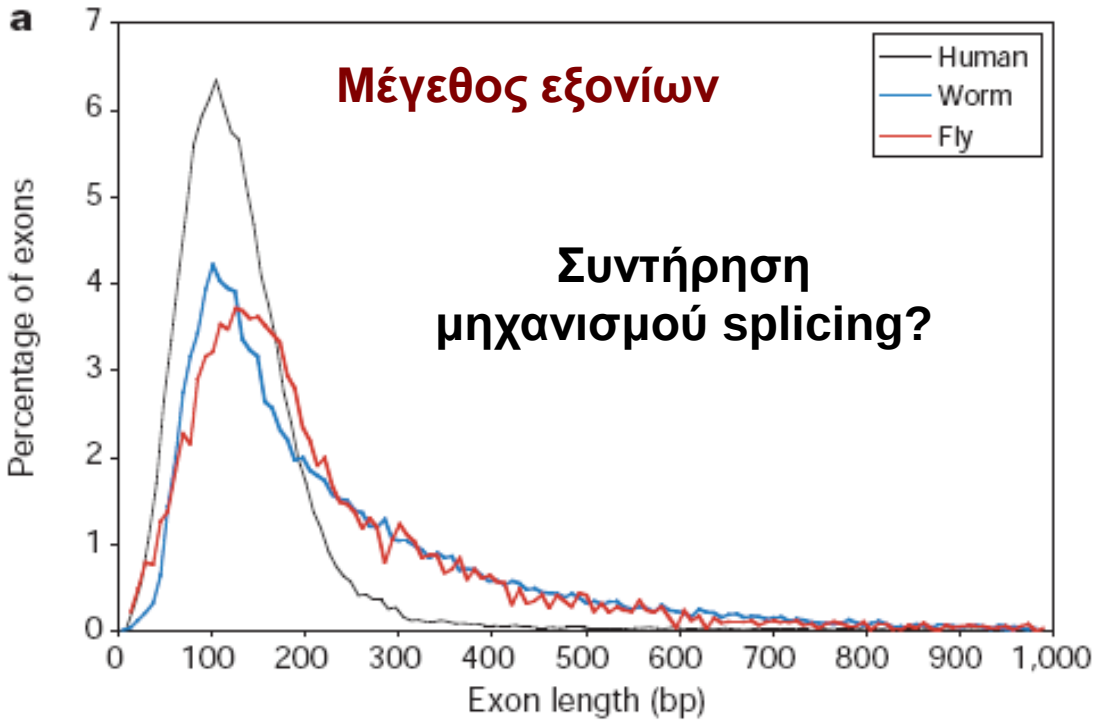


Πώς γίνεται η πρόβλεψη γονιδίων;

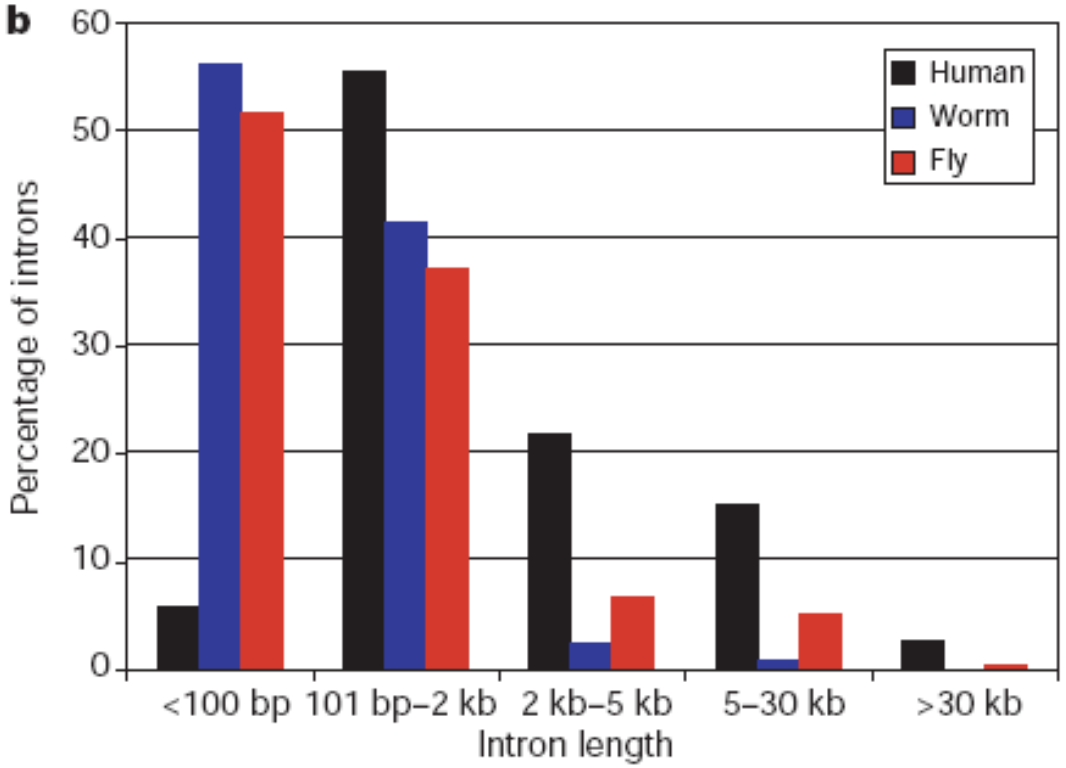
- **Άμεση ένδειξη μεταγραφής από ESTs & cDNAs**
- **Ομοιότητα αλληλουχίας με ήδη αναγνωρισμένα γονίδια**
- **Ab initio (εξαρχής) αναγνώριση με αλγόριθμους που αναγνωρίζουν θέσεις συρραφής, κωδικοποιητικές προτιμήσεις, μήκος εξονίων-ιντρονίων**

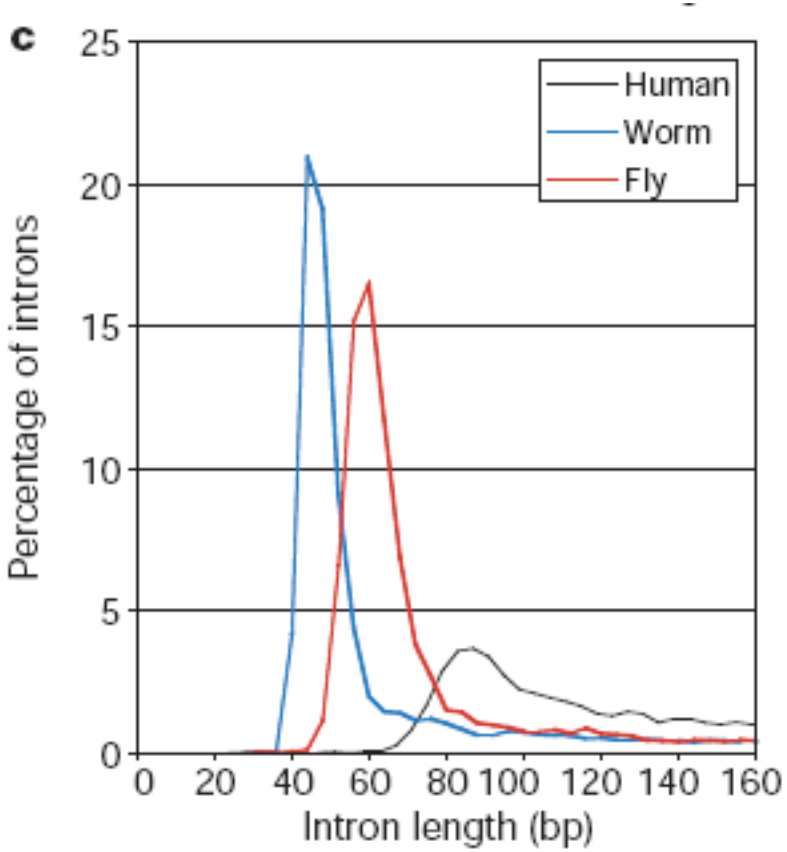
Χαρακτηριστικά γονιδίων του ανθρώπου

	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	1,100 bp	1,340 bp
(CDS)	367 aa	447 aa
Genomic extent	14 kb	27 kb

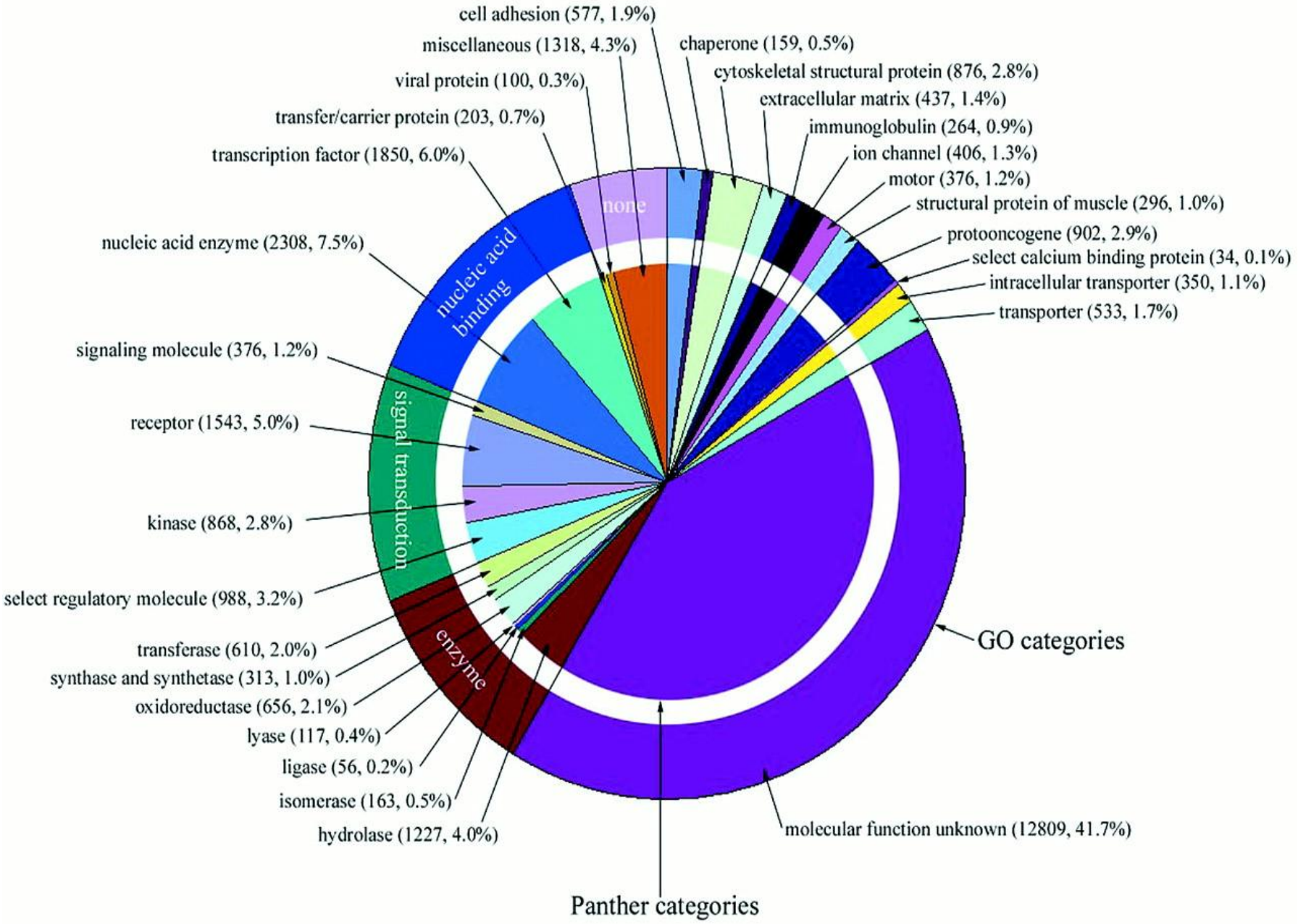


Μέγεθος ιντρονίων



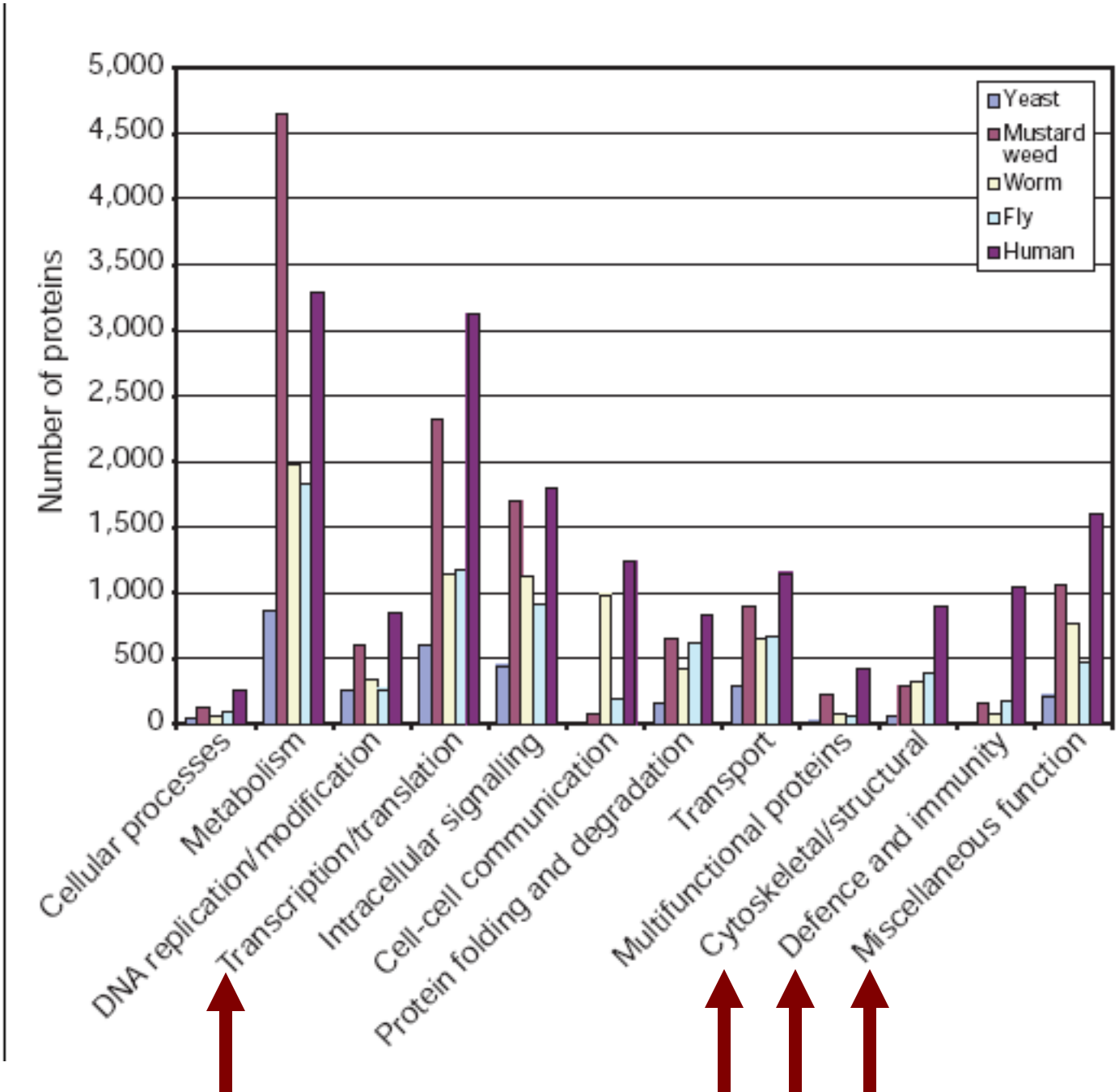


Κατηγοριοποίηση γονιδίων που κωδικοποιούν πρωτεΐνες



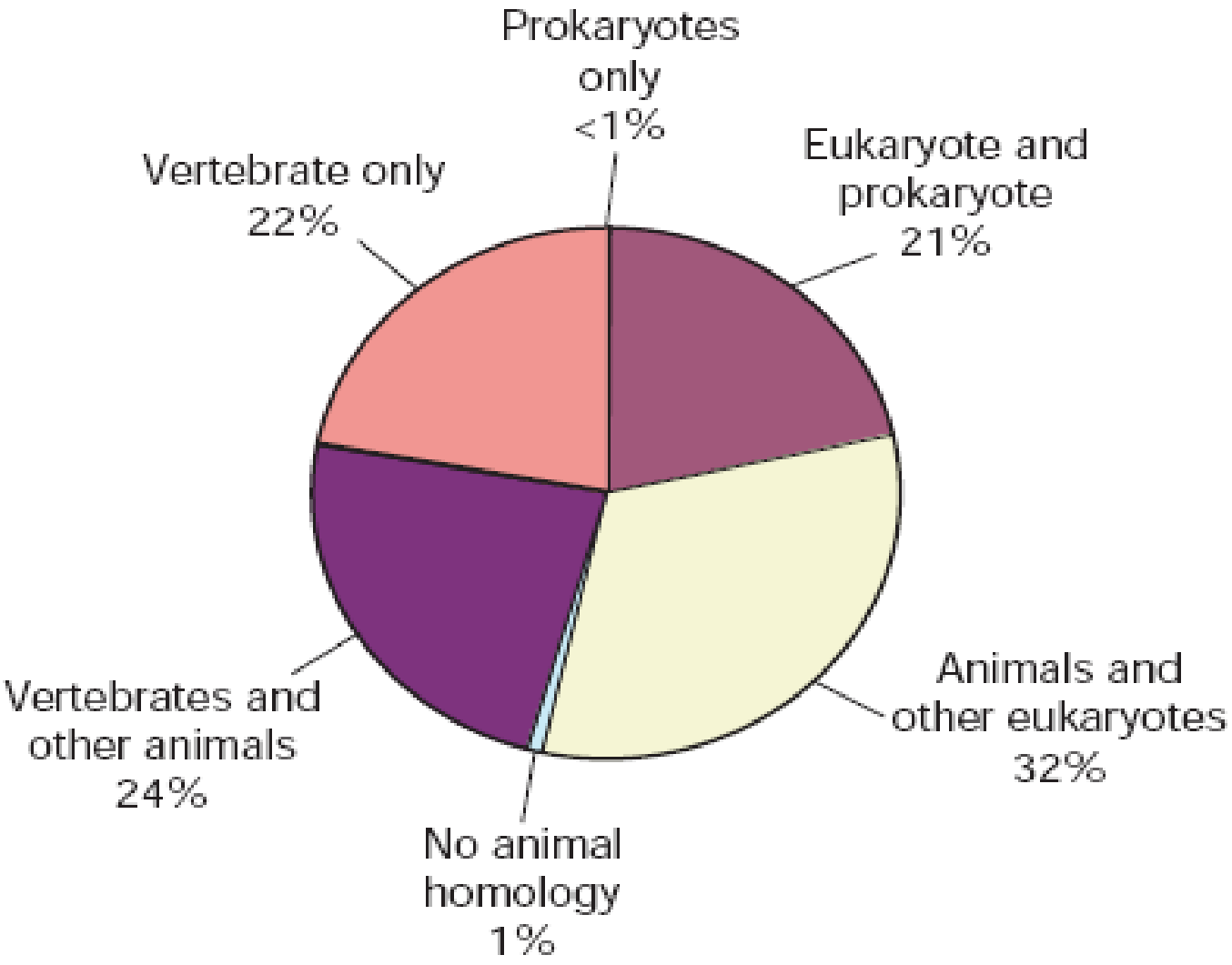
J. C. Venter et al., Science 291, 1304 -1351 (2001)

Και σύγκρισή τους με σακχαρομύκητα, Arabidopsis, σκουλήκι & μύγα



Συντήρηση κατά την εξέλιξη των πρωτεϊνών του ανθρώπου

Οριζόντια μεταφορά (ενδοκυτταρικά ένζυμα, μεταβολισμό ξενοβιοτικών, απάντηση στο στρες)



ομολογία

IPI ανθρώπου → 61% → D. melanogaster

→ 43% → C. elegans

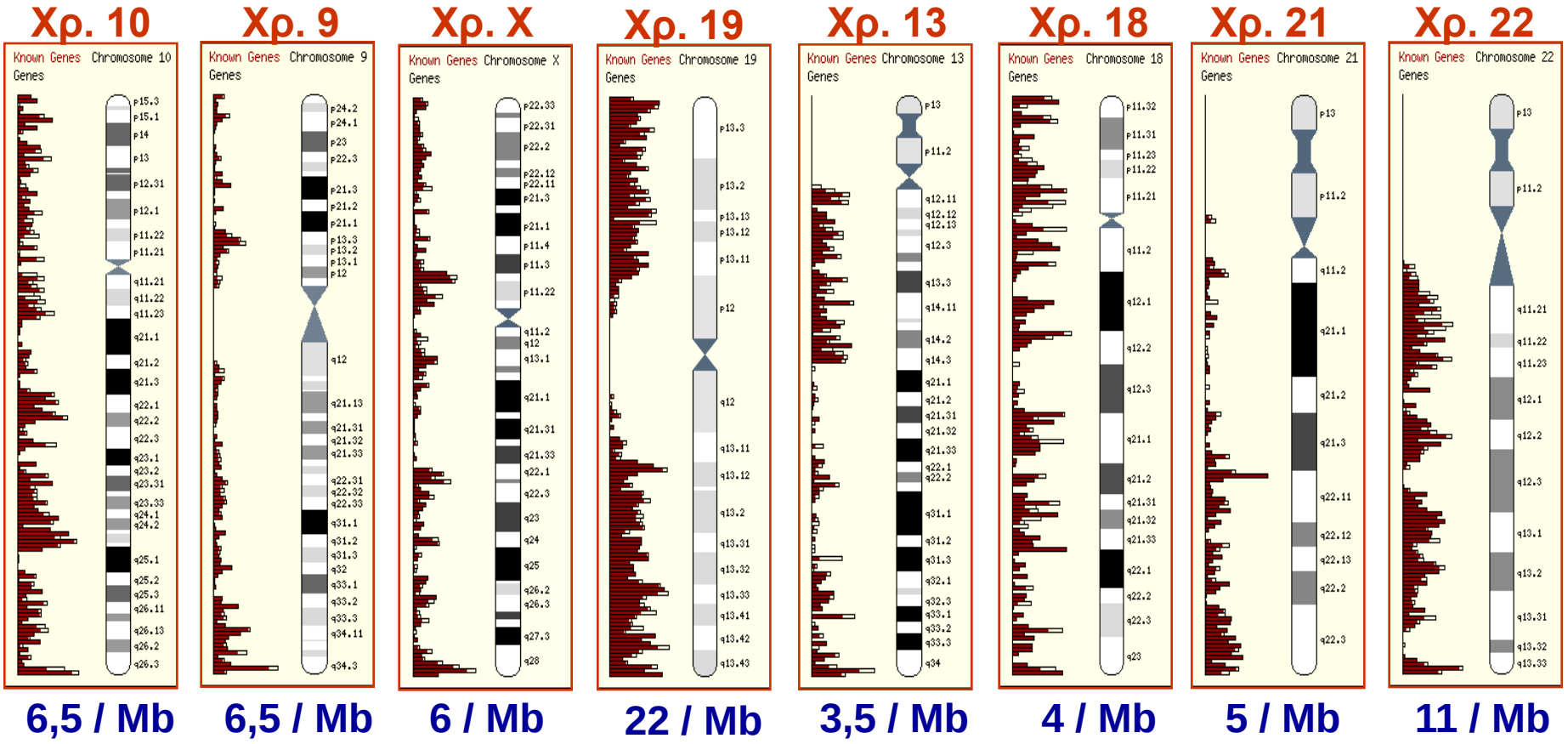
→ 46% → yeast

Ομάδες πρωτεϊνών 1-1-1-1 (πιθανά ορθόλογα στους 4 οργανισμούς)

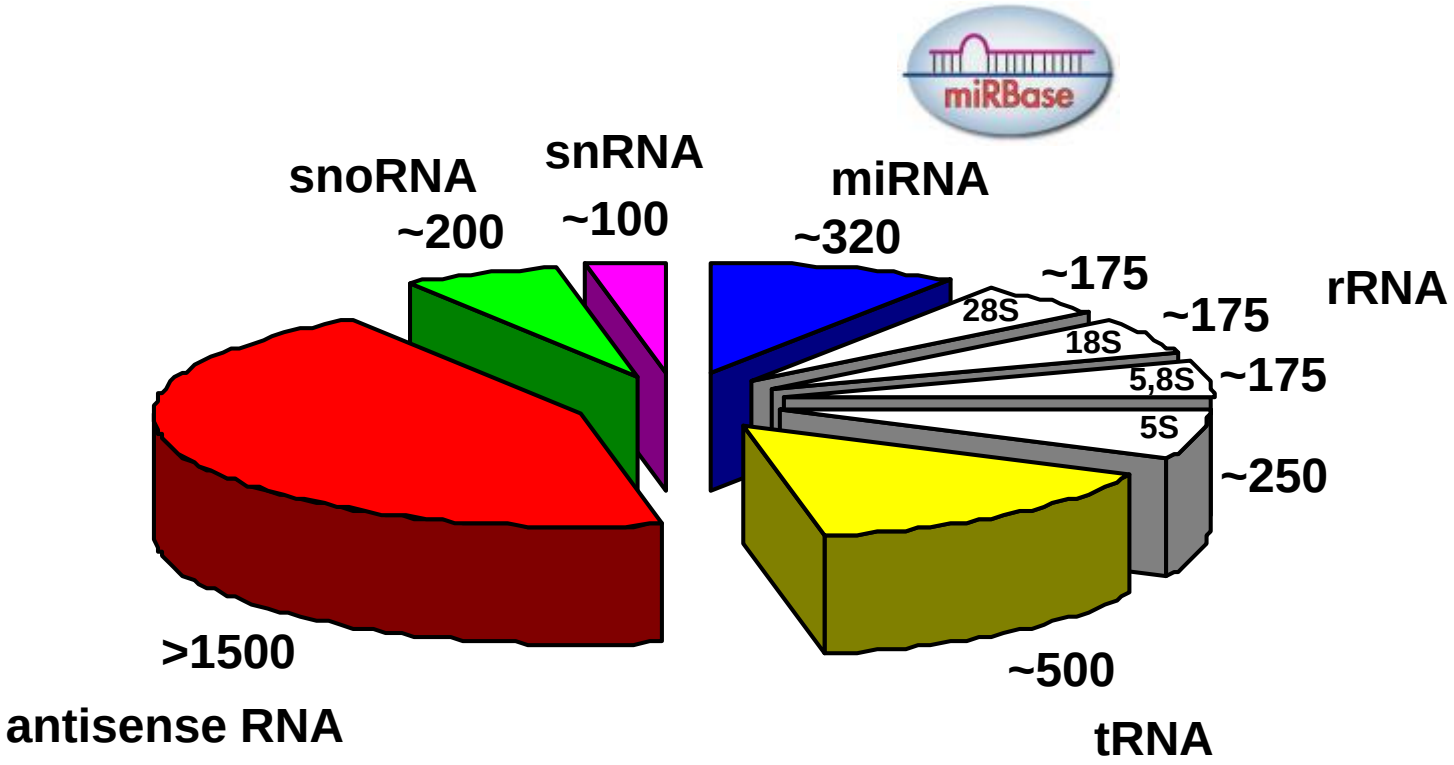
3129 – 1445 – 1503 - 1441

(μεταβολισμό, αντιγραφή, επιδιόρθωση, μετάφραση)

Μέση γονιδιακή πυκνότητα ~7,5 γονίδια / Mb



ΓΟΝΙΔΙΑ ΠΟΥ ΚΩΔΙΚΟΠΟΙΟΥΝ RNA



Identification of hundreds of conserved and nonconserved human microRNAs

Isaac Bentwich[1,2], Amir Avniel[1,2], Yael Karov[1,2], Ranit Aharonov[1,2], Shlomit Gilad[1,2], Omer Barad[1], Adi Barzilai[1], Paz Einat[1], Uri Einav[1], Eti Meiri[1], Eilon Sharon[1], Yael Spector[1] & Zvi Bentwich[1]

MicroRNAs are noncoding RNAs of ~22 nucleotides that suppress translation of target genes by binding to their mRNA and thus have a central role in gene regulation in health and disease[1–5]. To date, 222 human microRNAs have been identified[6], 86 by random cloning and sequencing, 43 by computational approaches and the rest as putative microRNAs homologous to microRNAs in other species. To prove our hypothesis that the total number of microRNAs may be much larger and that several have emerged only in primates, we developed an integrative approach combining bioinformatic predictions with microarray analysis and sequence-directed cloning. Here we report the use of this approach to clone and sequence 89 new human microRNAs (nearly doubling the current number of sequenced human microRNAs), 53 of which are not conserved beyond primates. These findings suggest that the total number of human microRNAs is at least 800.

We developed microRNA discovery tools that detect microRNAs missed by existing methods, which detect only conserved hairpins. Our approach (**Fig. 1a**) comprises the following steps: (i) computationally scanning the entire human genome for hairpin structures; (ii) annotating all hairpins for conserved, repetitive and protein-coding regions; (iii) scoring hairpins by thermodynamic stability and structural features, using a method (PalGrade) that detects a large percentage

Antisense Transcription in the Mammalian Transcriptome

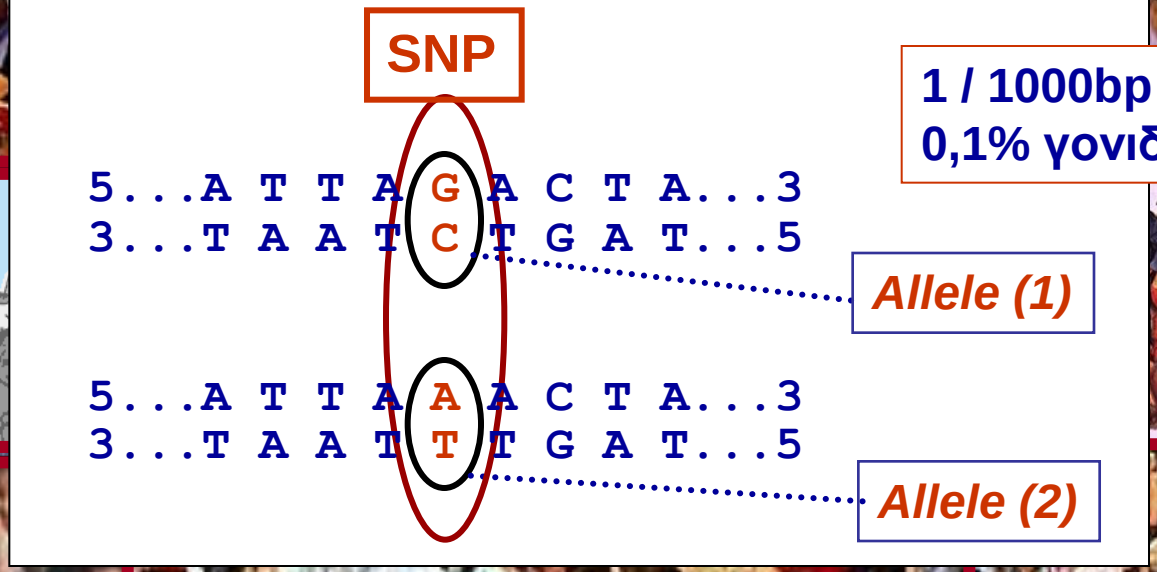
RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) and the FANTOM Consortium

Antisense transcription (transcription from the opposite strand to a protein-coding or sense strand) has been ascribed roles in gene regulation involving degradation of the corresponding sense transcripts (RNA interference), as well as gene silencing at the chromatin level. Global transcriptome analysis provides evidence that a large proportion of the genome can produce transcripts from both strands, and that antisense transcripts commonly link neighboring "genes" in complex loci into chains of linked transcriptional units. Expression profiling reveals frequent concordant regulation of sense/antisense pairs. We present experimental evidence that perturbation of an antisense RNA can alter the expression of sense messenger RNAs, suggesting that antisense transcription contributes to control of transcriptional outputs in mammals.

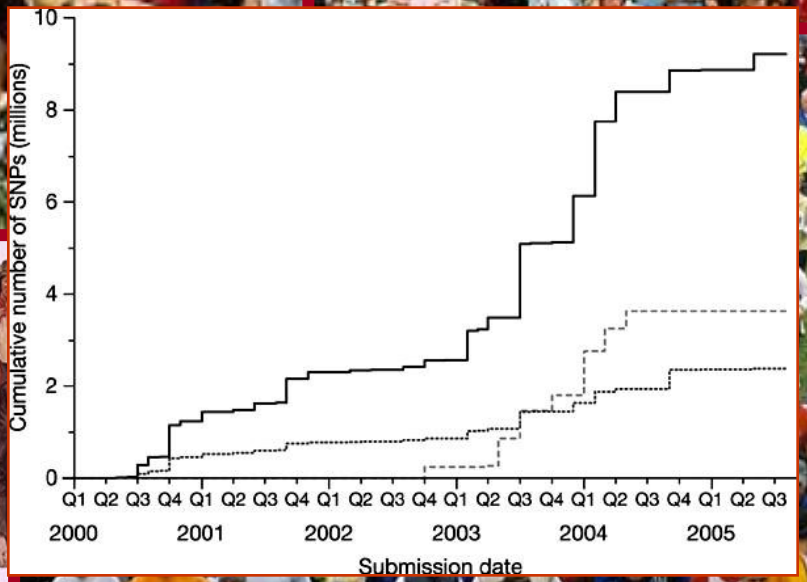
2 SEPTEMBER 2005 VOL 309 SCIENCE

ΓΕΝΕΤΙΚΗ ΠΟΙΚΙΛΟΤΗΤΑ

SNPs (single nucleotide polymorphisms)



1 / 1000bp
0,1% γονιδιώματος



A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms

The International SNP Map Working Group*

* *A full list of authors appears at the end of this paper.*

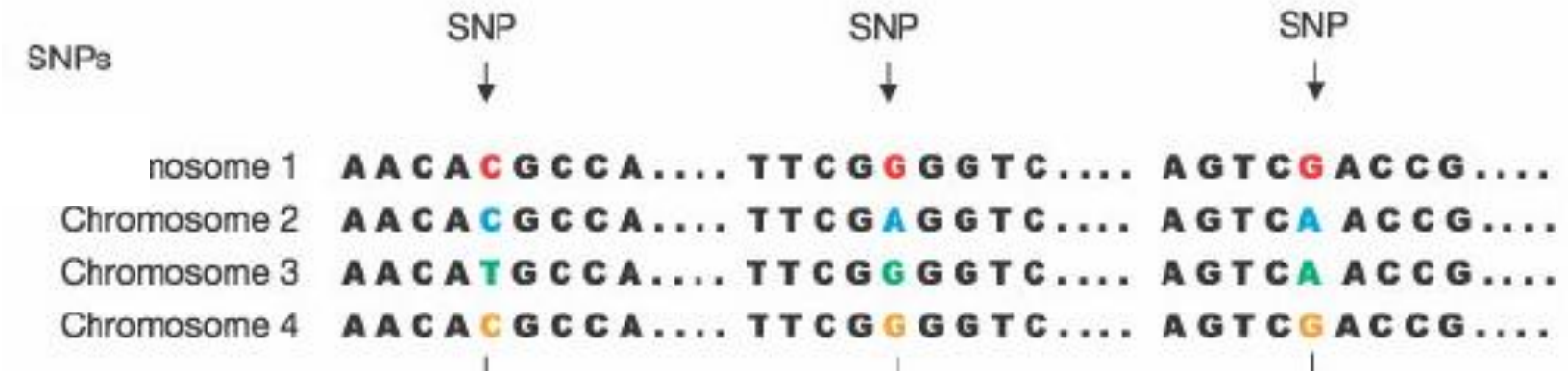
We describe a map of 1.42 million single nucleotide polymorphisms (SNPs) distributed throughout the human genome, providing an average density on available sequence of one SNP every 1.9 kilobases. These SNPs were primarily discovered by two projects: The SNP Consortium and the analysis of clone overlaps by the International Human Genome Sequencing Consortium. The map integrates all publicly available SNPs with described genes and other genomic features. We estimate that 60,000 SNPs fall within exon (coding and untranslated regions), and 85% of exons are within 5 kb of the nearest SNP. Nucleotide diversity varies greatly across the genome, in a manner broadly consistent with a standard population genetic model of human history. This high-density SNP map provides a public resource for defining haplotype variation across the genome, and should help to identify biomedically important genes for diagnosis and therapy.

International HapMap Project

(Χάρτης απλοτύπων)

Απλότυπος: ένας συγκεκριμένος συνδυασμός αλληλομόρφων σε ένα χρωμόσωμα

Συγκρίνοντας απλοτύπους από πολλά άτομα, παρατηρούμε κοινά πρότυπα



- **Ομάδες γειτονικών SNPs στο ίδιο χρωμόσωμα κληρονομούνται μαζί (blocks) - ανισορροπία σύνδεσης**

- **Το πρότυπο των SNPs σε ένα block είναι ο απλότυπος**

- **Είναι δυνατόν να επιλεγούν και να ελεγχθούν συγκεκριμένα SNPs ώστε να γίνουν αναλύσεις συσχέτισης με συγκεκριμένο φαινότυπο.**

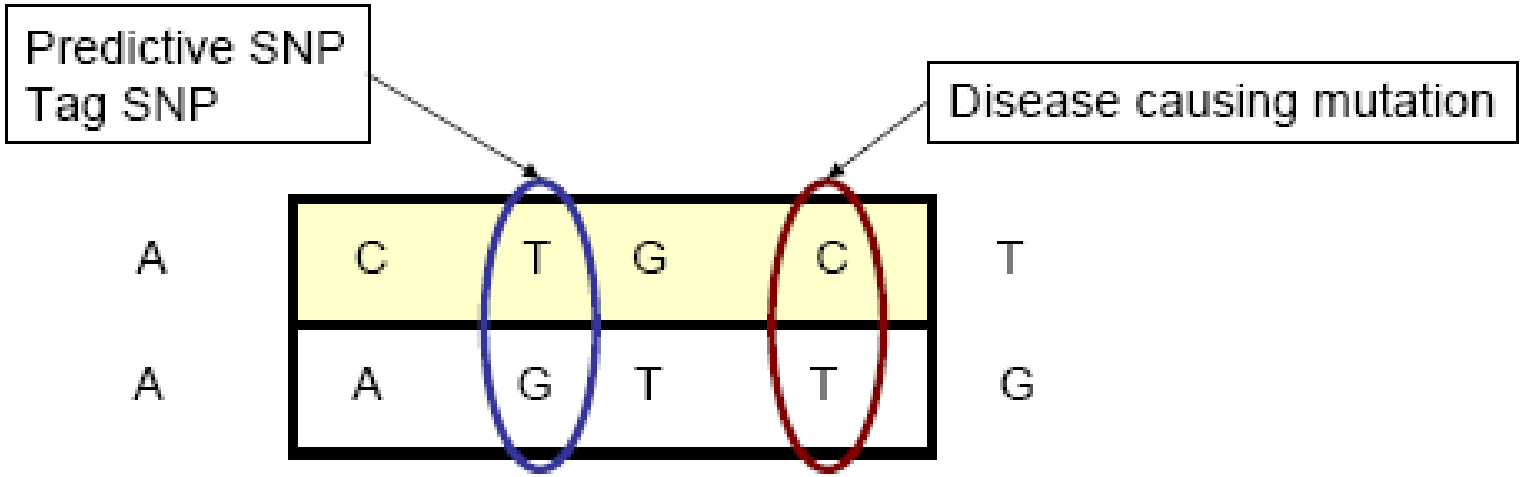
- **Για άτομα που έχουν ένα συγκεκριμένο SNP σε μία θέση, μπορούμε να προβλέψουμε τα SNPs σε γειτονικές θέσεις**

- **Ο HapMap είναι ο χάρτης των blocks των SNPs που συν-κληρονομούνται καθώς και των επιλεγμένων SNPs (tag SNPs) που ταυτοποιούν αυτά τα blocks**

~10 εκατομμύρια SNPs στο ανθρώπινο γονιδίωμα

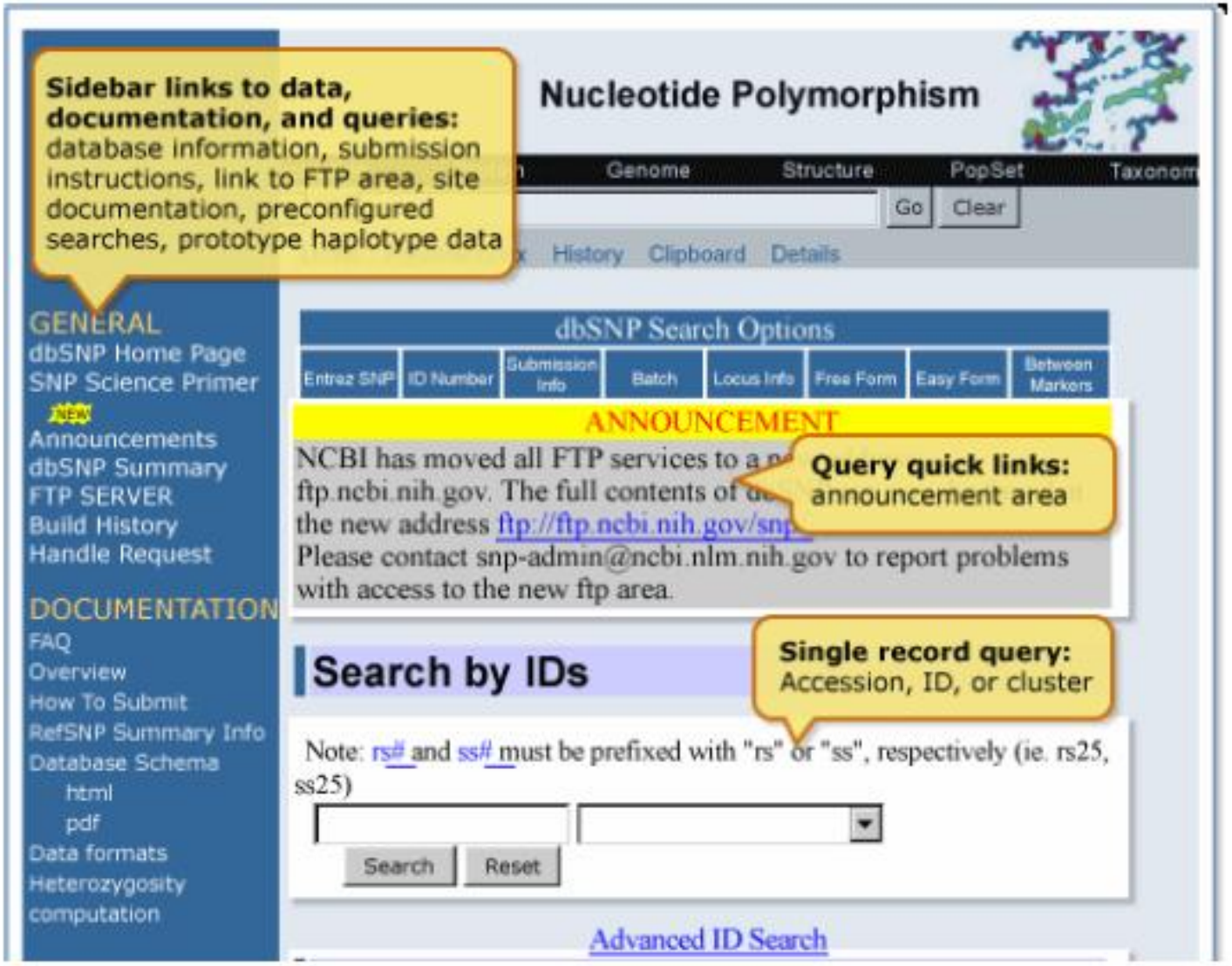
↓

~ 500,000 tag SNPs

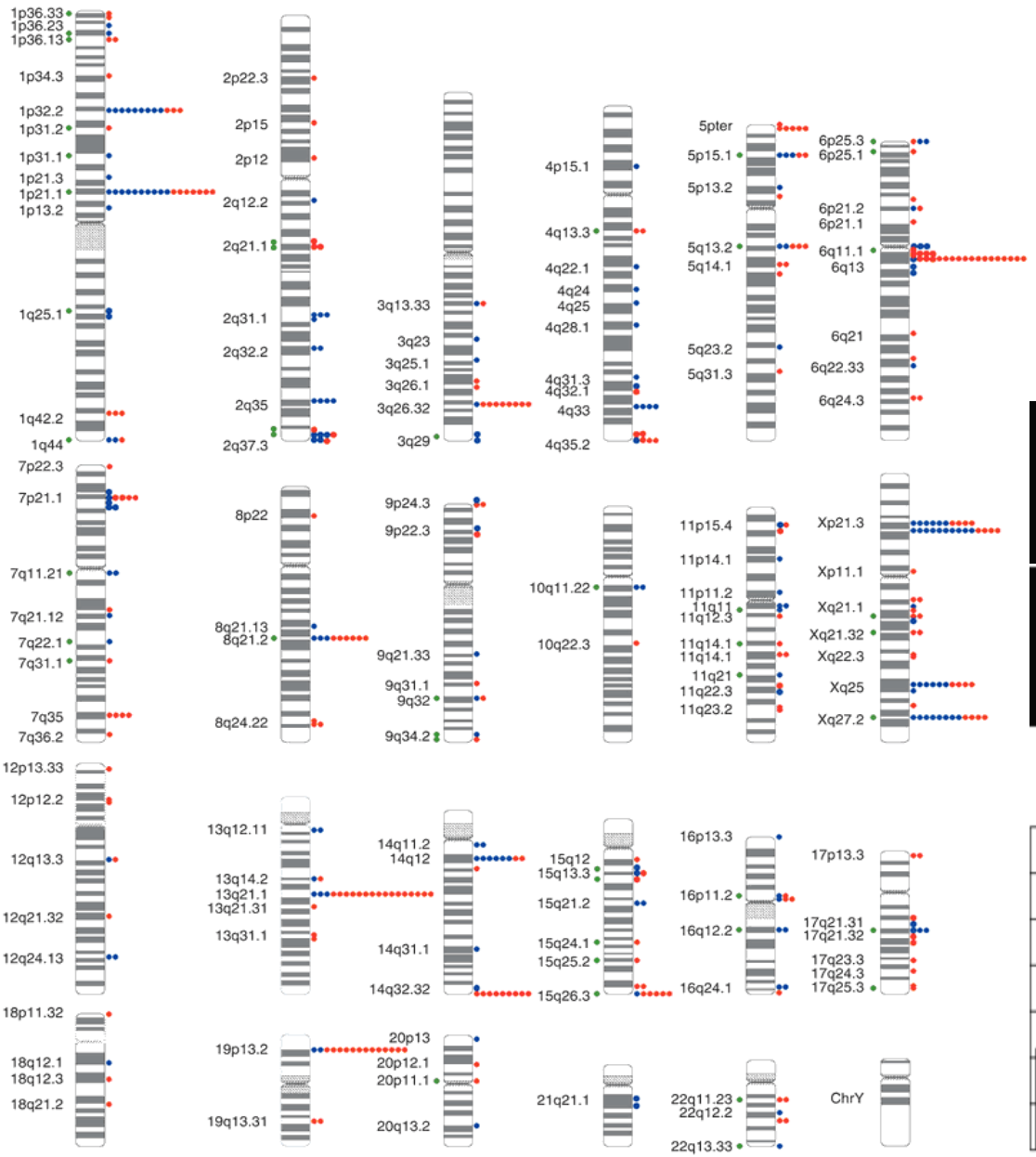


Μελέτες συσχέτισης σε επίπεδο γονιδιώματος
(χωρίς επιλογή υποψηφίων γονιδίων)

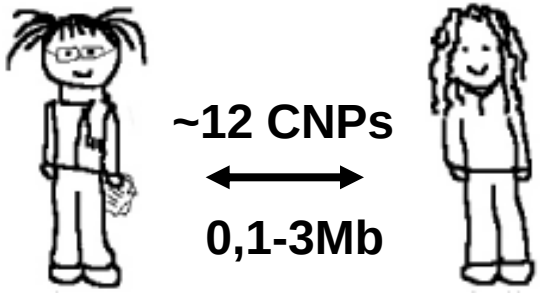
http://www.ncbi.nlm.nih.gov/SNP/



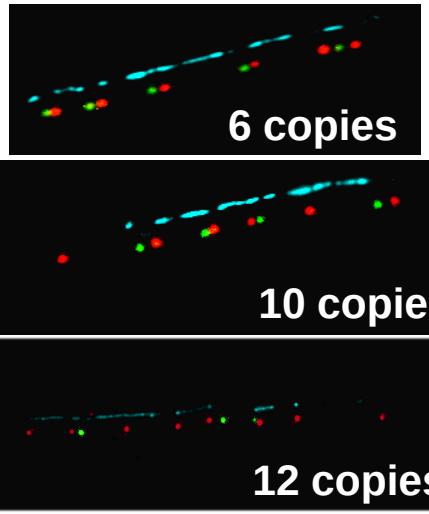
CNPs (copy number variation)



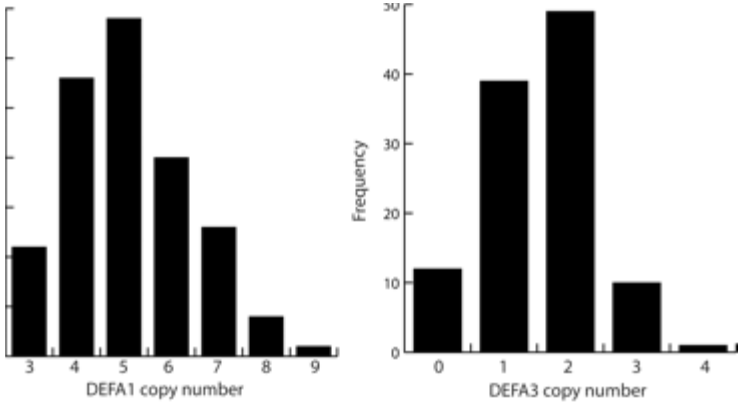
Iafrate et al Nat Genet 2004, Aldred et al Human Mol Genet 2005



AMYLASE GENES



DEFENSIN GENES



nature Vol 444|23 November 2006|doi:10.1038/nature05329

ARTICLES

Global variation in copy number in the human genome

Richard Redon[1], Shumpei Ishikawa[2,3], Karen R. Fitch[4], Lars Feuk[5,6], George H. Perry[7], T. Daniel Andrews[1], Heike Fiegler[1], Michael H. Shapero[4], Andrew R. Carson[5,6], Wenwei Chen[4], Eun Kyung Cho[7], Stephanie Dallaire[7], Jennifer L. Freeman[7], Juan R. González[8], Mònica Gratacòs[8], Jing Huang[4], Dimitrios Kalaitzopoulos[1], Daisuke Komura[3], Jeffrey R. MacDonald[5], Christian R. Marshall[5,6], Rui Mei[4], Lyndal Montgomery[1], Kunihiro Nishimura[2], Kohji Okamura[5,6], Fan Shen[4], Martin J. Somerville[9], Joelle Tchinda[7], Armand Valsesia[1], Cara Woodwark[1], Fengtang Yang[1], Junjun Zhang[5], Tatiana Zerjal[1], Jane Zhang[4], Lluis Armengol[8], Donald F. Conrad[10], Xavier Estivill[8,11], Chris Tyler-Smith[1], Nigel P. Carter[1], Hiroyuki Aburatani[2,12], Charles Lee[7,13], Keith W. Jones[4], Stephen W. Scherer[5,6] & Matthew E. Hurles[1]

Copy number variation (CNV) of DNA sequences is functionally significant but has yet to be fully ascertained. We have constructed a first-generation CNV map of the human genome through the study of 270 individuals from four populations with ancestry in Europe, Africa or Asia (the HapMap collection). DNA from these individuals was screened for CNV using two complementary technologies: single-nucleotide polymorphism (SNP) genotyping arrays, and clone-based comparative genomic hybridization. A total of 1,447 copy number variable regions (CNVRs), which can encompass overlapping or adjacent gains or losses, covering 360 megabases (12% of the genome) were identified in these populations. These CNVRs contained hundreds of genes, disease loci, functional elements and segmental duplications. Notably, the CNVRs encompassed more nucleotide content per genome than SNPs, underscoring the importance of CNV in genetic diversity and evolution. The data obtained delineate linkage disequilibrium patterns for many CNVs, and reveal marked variation in copy number among populations. We also demonstrate the utility of this resource for genetic disease studies.

Vol 439|16 February 2006|doi:10.1038/nature04489

nature

LETTERS

Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans

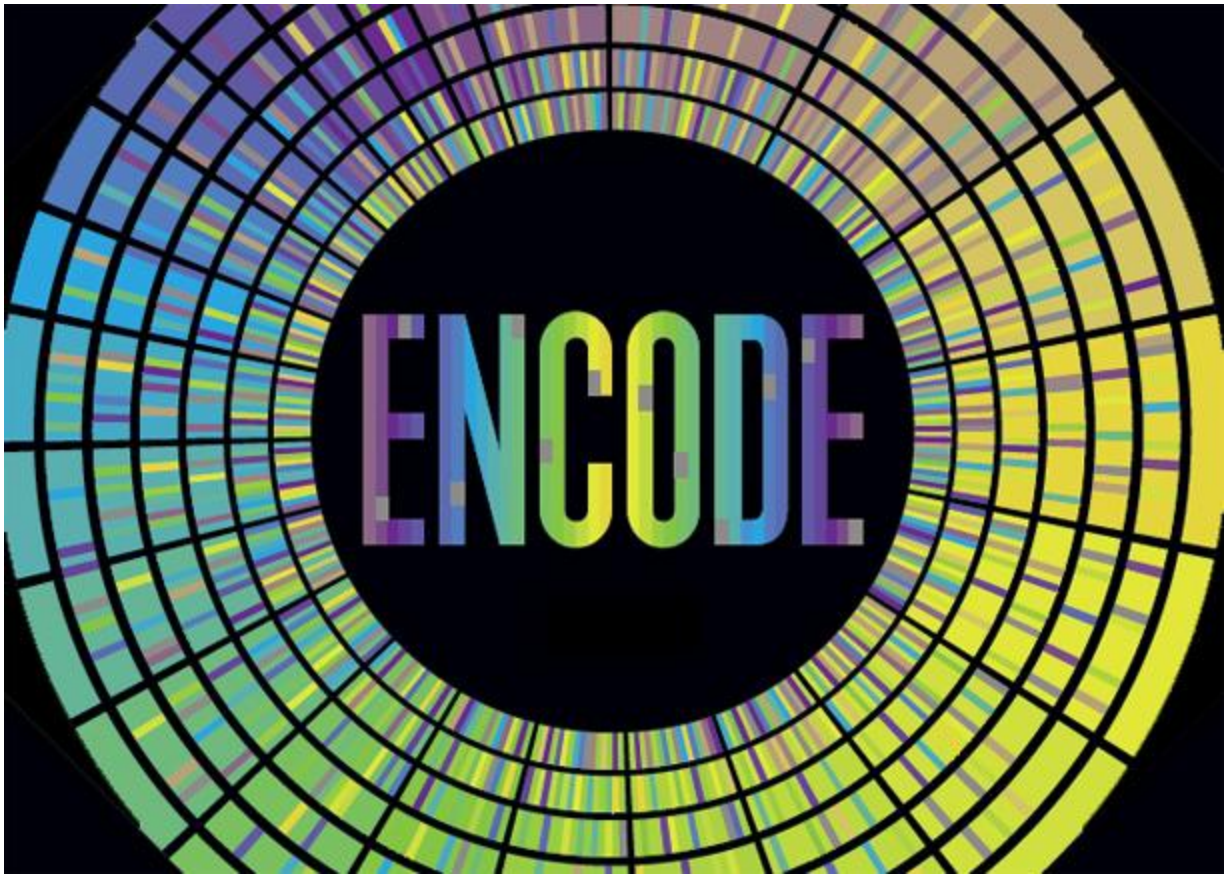
Timothy J. Aitman[1], Rong Dong[1]*, Timothy J. Vyse[2]*, Penny J. Norsworthy[1]*, Michelle D. Johnson[1], Jennifer Smith[3], Jonathan Mangion[1], Cheri Roberton-Lowe[1,2], Amy J. Marshall[1], Enrico Petretto[1], Matthew D. Hodges[1], Gurjeet Bhangal[3], Sheetal G. Patel[1], Kelly Sheehan-Rooney[1], Mark Duda[1,3], Paul R. Cook[1,3], David J. Evans[3], Jan Domin[3], Jonathan Flint[4], Joseph J. Boyle[5], Charles D. Pusey[3] & H. Terence Cook[5]


nature
THE CHIMPANZEE
GENOME


nature
The mouse genome


Science
C. elegans
Sequence to Biology


Science
24 March 2000
The
Drosophila
Genome



The Encyclopedia of DNA Elements

Σκοπός: **ο χαρακτηρισμός όλων των λειτουργικών στοιχείων του γονιδιώματος του ανθρώπου**

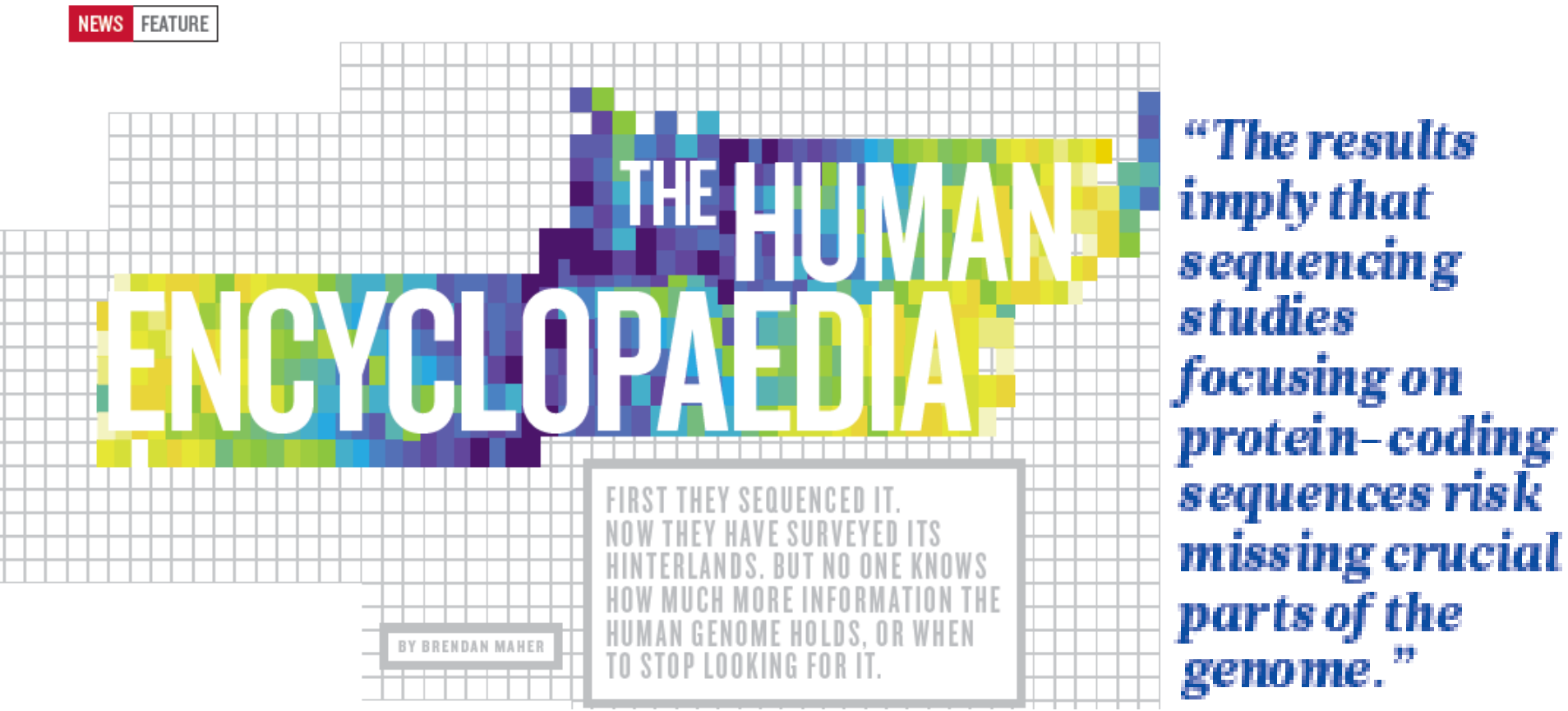
would stretch to Pluto and at least 30 kilometres long.

ENCODE was designed to pick up where the Human Genome Project left off. Although that massive effort revealed the blueprint of human biology, it quickly became clear that the instruc-

scape. Many biologists suspected that the information responsible for the wondrous complexity of humans lay somewhere in the 'deserts' between the genes. ENCODE, which started in 2003, is a massive data-collection effort designed to populate this terrain. The aim is to catalogue the 'functional' DNA sequences that lurk there, learn when and in which cells they are active and trace their effects on how the genome is packaged, regulated and read.

Σεπτέμβριος 2012: 30 δημοσιεύσεις με τα αποτελέσματα του προγράμματος (ελεύθερα προσβάσιμα)

Ταυτοποίηση λειτουργικών στοιχείων στο 80% του γονιδιώματος του ανθρώπου



BY THE NUMBERS

32 INSTITUTES

THOMAS POROSTOCKY; SOURCE: MEETINGZONE

The ENCODE project involved hundreds of people from around the world, and a lot of editing, disk space and phone calls.

442 CONSORTIUM MEMBERS

DATA

1,649 EXPERIMENTS

11,972 FILES ANALYSED

15 TB DISK SPACE USED

ENCODE Wiki

741 WIKI CONTENT PAGES

18,500 PAGE EDITS SINCE 2008

248,140 VIEWS

TELECONFERENCING MAY 2008 TO JUNE 2012

675 CALLS MADE

13 PARTICIPANTS PER CALL

45m MINUTES PER CALL PER PARTICIPANT

292 PERSON-DAYS SPENT ON CONFERENCE CALLS

TOTAL COST OF TELECONFERENCING = **£49,310.54**

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

ncbi.nlm.nih.gov

National Center for Biotechnology Information

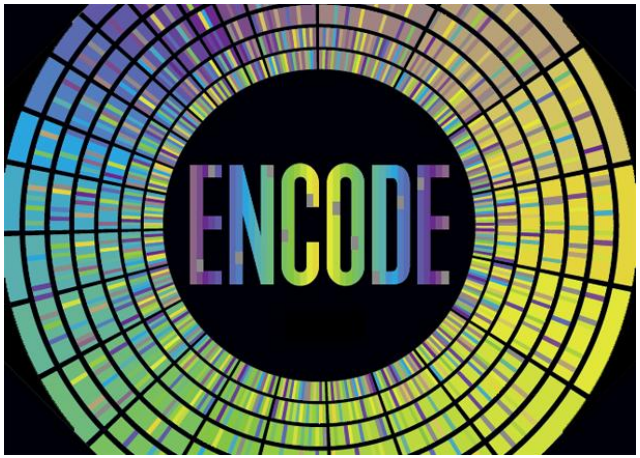
- **Gene**
- **Blast**

- **SNP**

By gene (e.g. BRCA1), by rs (e.g rs12427078)
Limits (function class)

http://www.ensembl.org

Π.χ. Ποια γονίδια εντοπίζονται σε μία κρίσιμη περιοχή (π.χ. D10S597-D10S554)
Πληροφορίες για ένα γονίδιο (π.χ. BRCA2)



Nature.com/encode