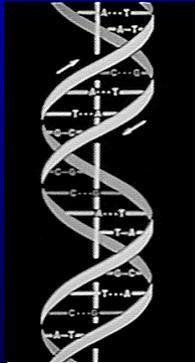


Structural & Functional Genomics: The Information Landscape

NCBI



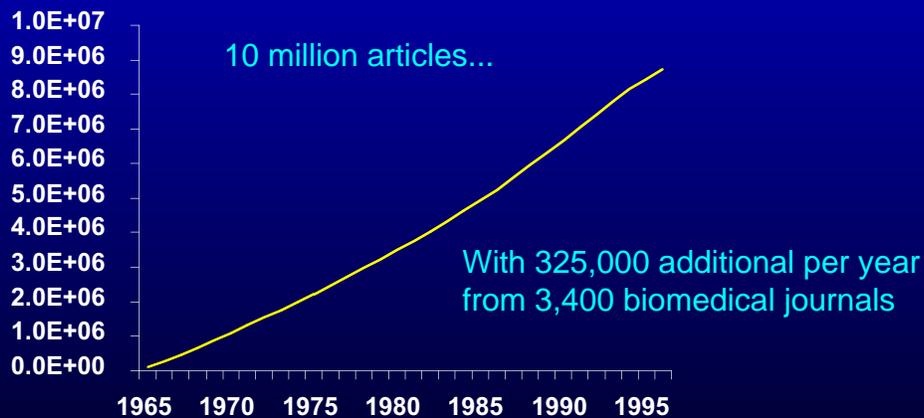
Mark S. Boguski, M.D., Ph.D.
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
Bethesda, Maryland



Current Topics in Genome Analysis, 4 November 1997

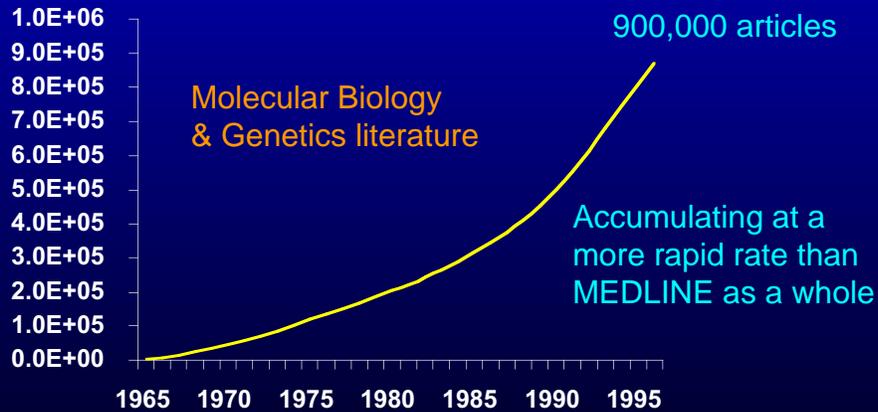
Growth of Biomedical Information (1) MEDLINE®

NCBI



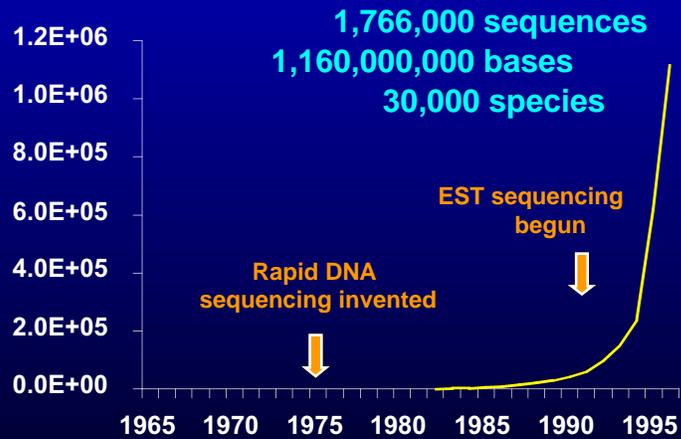
Growth of Biomedical Information (2) “G5” MeSH subset of MEDLINE

NCBI



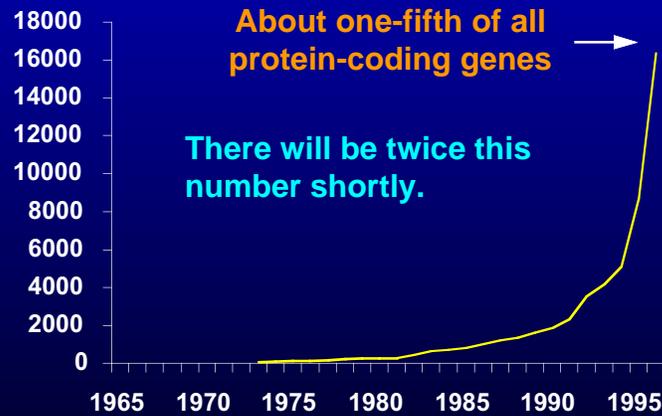
Growth of Biomedical Information (3) GenBank DNA sequences

NCBI



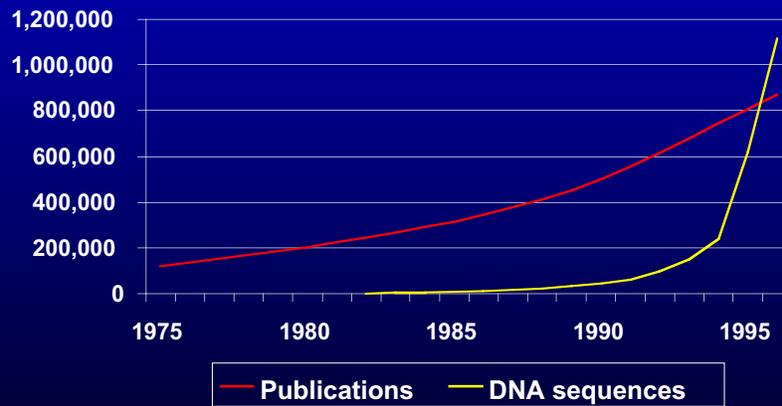
Growth of Biomedical Information (4) Mapped Human Genes

NCBI



The Cross-over to Functional Genomics (1)

NCBI



The Cross-over to Functional Genomics (2)

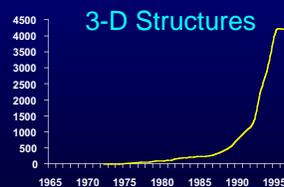
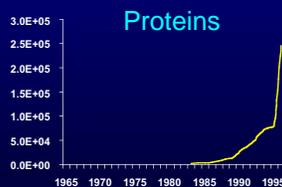
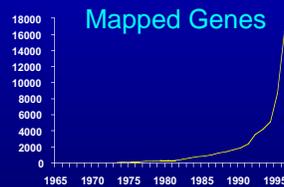
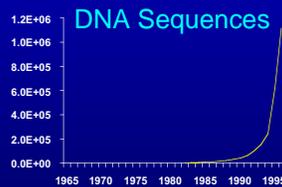
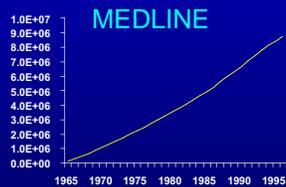
NCBI

“In the past we have had functions in search of sequences. In the future, pathology and physiology will become ‘functionators’ for the sequences.”

Daniel C. Tosteson, Dean
Harvard Medical School
March 26, 1997

Growth of Biomedical Information 1965-1996

NCBI



The National Center for Biotechnology Information

NCBI



Created by Congress
in 1988 with a
mandate to...

Create automated systems for
knowledge about molecular
biology, biochemistry and
genetics

Perform research into advanced
methods of analyzing and
interpreting molecular
biology data

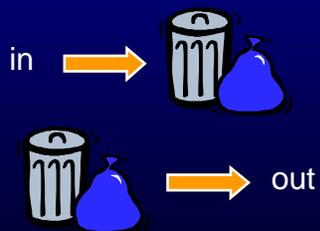
Enable biotechnology
researchers and medical
care personnel to use the
systems and methods
developed

Informatics on the World Wide Web

NCBI

“[There are] 150 million Web pages now in
existence....We can expect a billion Web
pages by 2000. **Some of them will even be
worth reading.**”

WIRED Magazine
March 1997



World Wide Web: Information Cornucopia?

NCBI

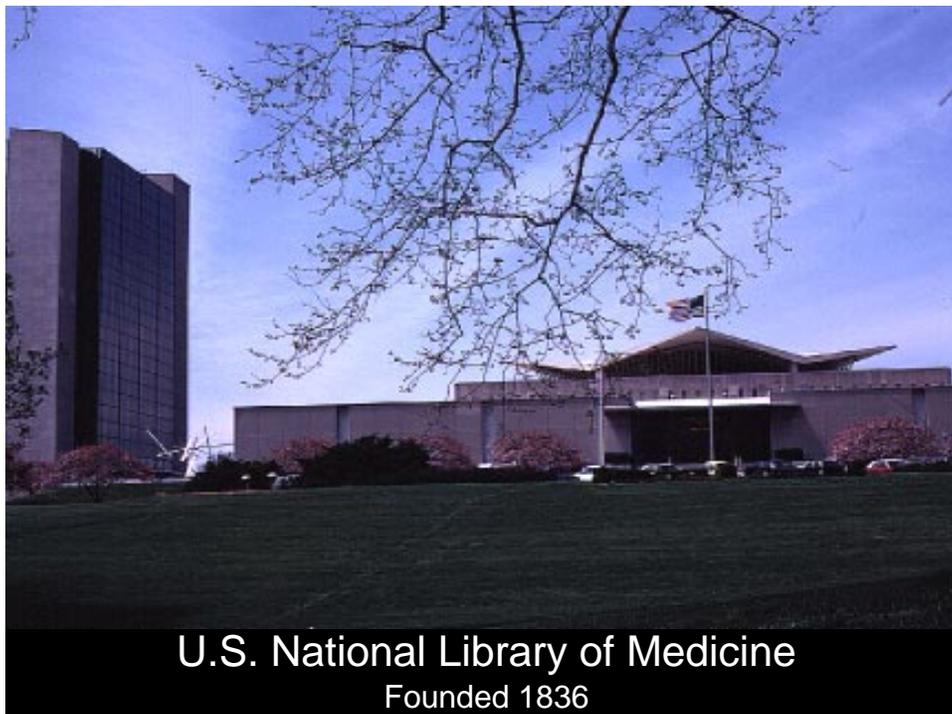
Among 30 million Web pages in cyberspace, touted as the road to real-time, up-to-the-minute information resources, 5 million pages have been neither checked nor updated in the past year, and 75,000 have languished untouched since 1994.



OR



The Wall Street Journal
March 1996



U.S. National Library of Medicine
Founded 1836

The National Center for Biotechnology Information

NCBI

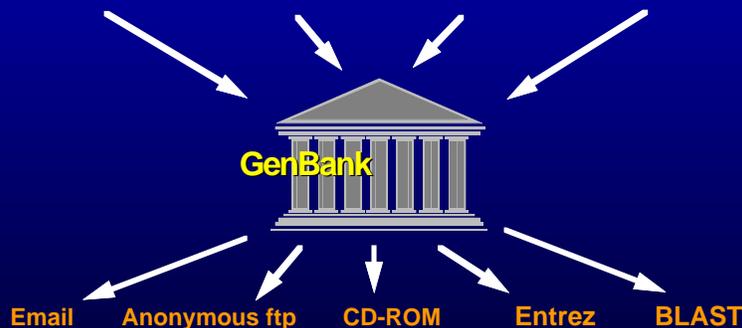
- Builders and providers of **GenBank®**, **BLAST**, **Entrez** and many other data and software resources
- NCBI is also a center for **basic research** and **training** in computational biology and bioinformatics



Building and Distributing GenBank®

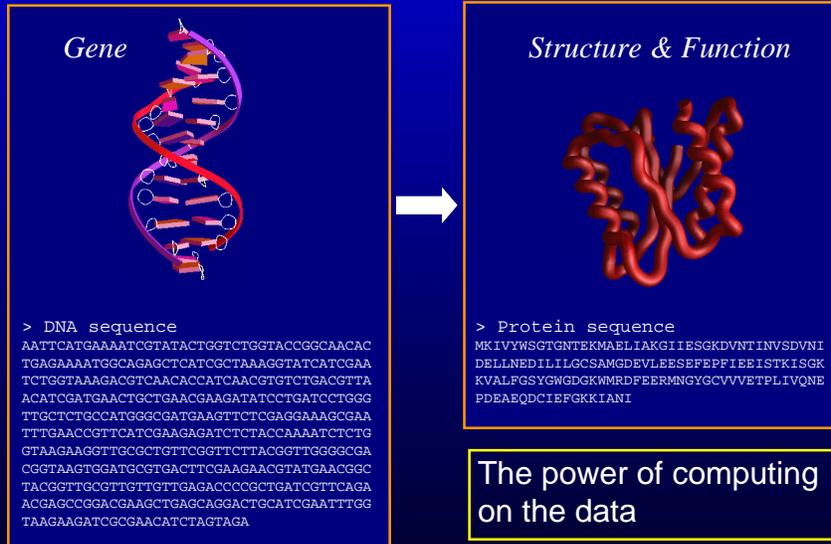
NCBI

European & Japanese collaborators Author-direct submission Journal scanning High-throughput sequencing centers

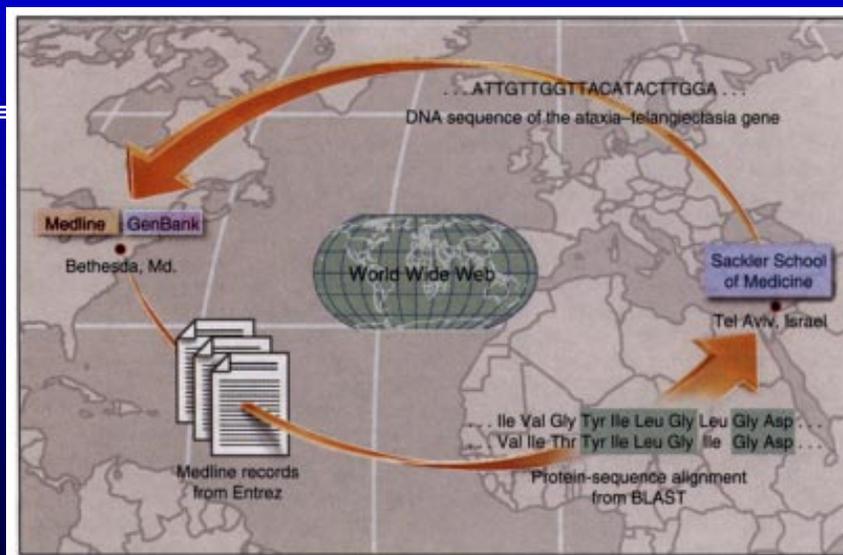


NCBI services 1.6 million web hits and 200,000 intellectual queries per day from 32,000 individual users

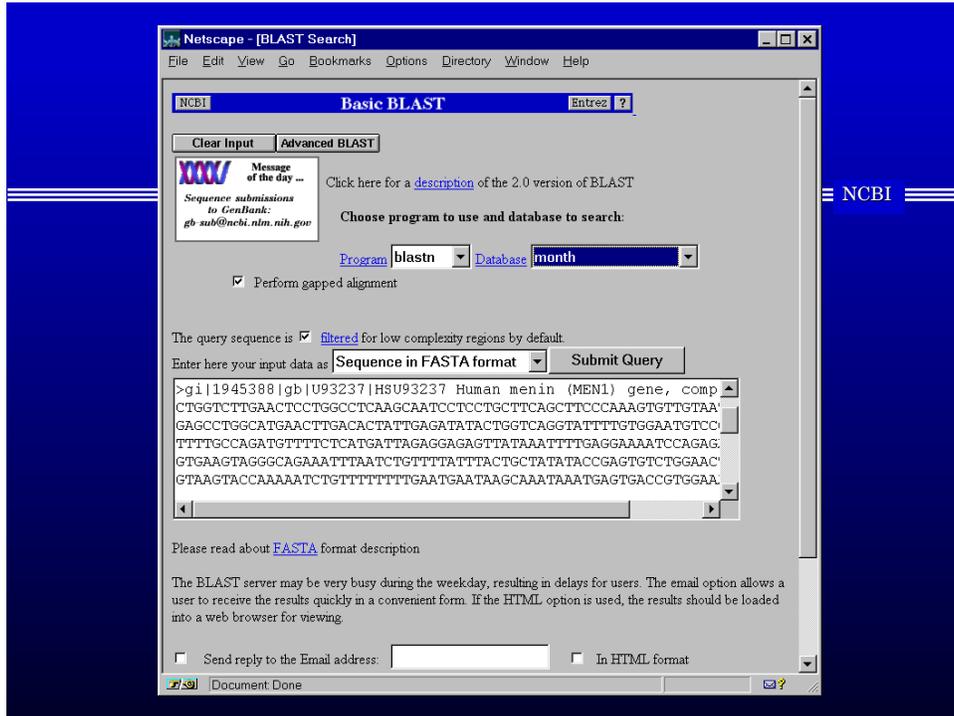
Biotechnology Information



Ataxia telangiectasia: 18 years and 5 minutes



New England Journal of Medicine 333:645-7; 1995



Comparative Analysis of Genes

NCBI

Cell, Vol. 75, 1027-1038, December 3, 1993, Copyright © 1993 by Cell Press

The Human Mutator Gene Homolog MSH2 and Its Association with Hereditary Nonpolyposis Colon Cancer

Richard Fishel,* Mary Kay Lescoe,* M. R. S. Rao, § Neal G. Copeland, † Nancy Jenkins, † Judy Garber, ‡ Michael Kane, § and Richard Kolodner §

*Department of Microbiology and Molecular Genetics
 Markey Center for Molecular Genetics
 University of Vermont Medical Center

Homology to bacterial and yeast genes sheds new light on human disease process

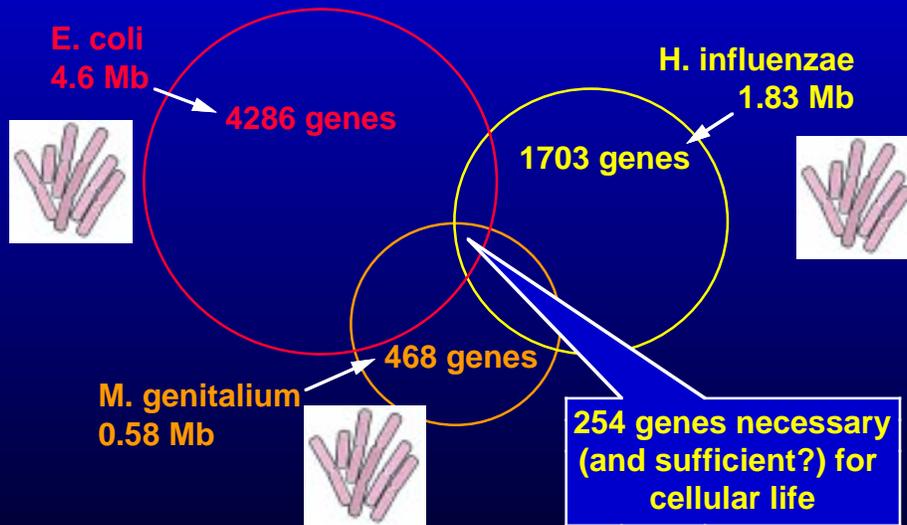
```

Human 638 RHACVEVQDEIAFIPNDVYFEKDKQMFHIITGPNMGGKSTYIRQTVIVLMAQIGCFVPC 697
Yeast 657 RHPVLEMQDDISFISNDVTLESKGDFLIITGPNMGGKSTYIRQVGVISLMAQIGCFVPC 716
E.coli 584 RHPVVEQVLNEPFIANPLNLSPPQR-MLIITGPNMGGKSTYMRQTALIALMAVIGSYVPA 642
  
```

portion of DNA mismatch repair protein sequence

Comparative analysis of genomes

NCBI



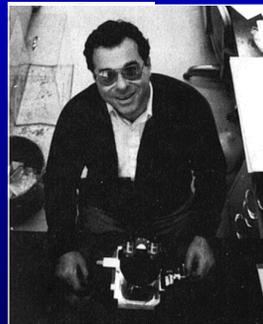
*Mushegian & Koonin, Proc. Natl. Acad. Sci. 93:10268, 1996

Comparative Analysis of Genomes

NCBI



“What is true for *E. coli*
is also true for elephant.”
Jacques Monod, c. 1961



“What is true for yeast
is also true for human.”
David Botstein, 1988

Netscape - [Human cDNAs that complement *S. cerevisiae* mutants]

Location: <http://www.ncbi.nlm.nih.gov/Bassett/cerevisiae/ComplementNew.html>

NCBI XREFdb

Sequence similarity between Human cDNAs that complement *S. cerevisiae* mutants and their correlate yeast proteins.

As of July 15, 1996, this table contains 71 *S. cerevisiae* mutant/complementing human cDNA pairs.

If you know of a human cDNA which complements a yeast mutant, and the pair is not yet present on this table, please report the cDNA/mutant pair, along with a reference if possible, to bassett@ncbi.nlm.nih.gov and this table will be updated.

Human Gene	GenBank Accession#	BLASTX P-value	Yeast Mutant	SwissProt Accession#	Reference
ARF5	M57567	1.0e-91	arf2	P19146	Lee FJ, et al. (1992)
ARF5	M57567	1.4e-91	arf1	P11076	Lee FJ, et al. (1992)
ARF6	M57763	3.1e-85	arf2	P19146	Lee FJ, et al. (1992)
ARF6	M57763	4.0e-84	arf1	P11076	Lee FJ, et al. (1992)

**Bassett et al. (1997)
Nat. Gen. 15:439-44**

Sequence Conservation among Human and Rodent mRNAs

NCBI

93%

85%

86%

Mouse Rat Human

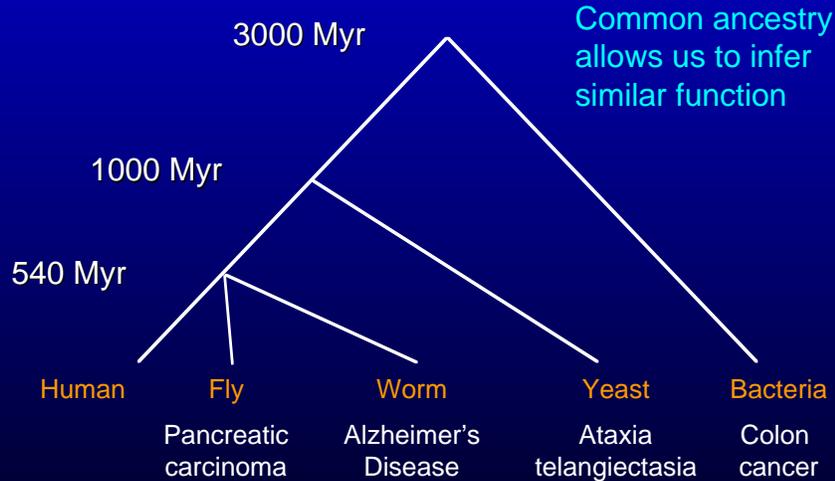
Coding sequence

5' UTR 3' UTR

Makalowski et al. (1996) Genome Res. 6:846

Molecular Evolution

NCBI



“Homology...”



NCBI

... is the central concept for all of biology. Whenever we say that a mammalian hormone is the ‘same’ hormone as a fish hormone, that a human gene sequence is the ‘same’ as a sequence in a chimp or a mouse, that a HOX gene is the ‘same’ in a mouse, a fruit fly, a frog, and a human -- even when we argue that discoveries about a worm, a fruit fly, a frog, a mouse, or a chimp have relevance to the human condition -- we have made a bold and direct statement about homology. The aggressive confidence of modern biomedical science implies that we know what we are talking about.”

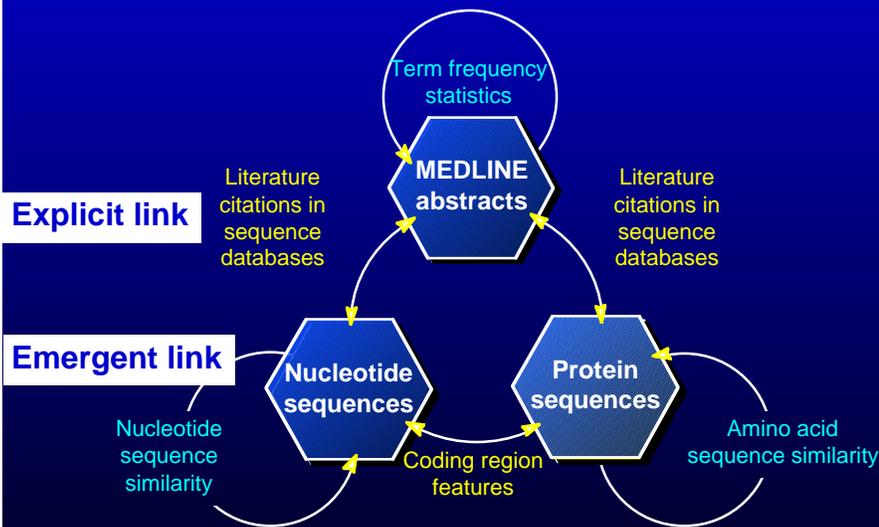


David B. Wake



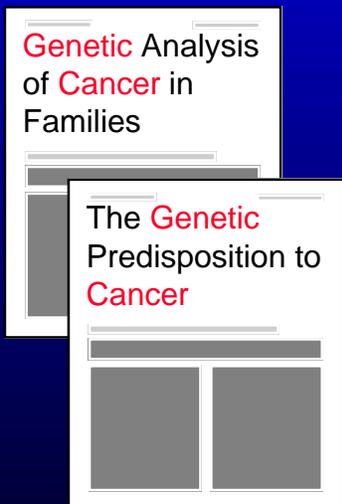
Entrez (1992)

NCBI



MEDLINE® Text Neighboring

NCBI



- Common terms could indicate similar subject matter
- Statistical method
- Weights based on term frequencies within document and within the database as a whole
- Some terms are better than others

Entrez (1994)

NCBI

Sequence similarity

Nucleotide sequence
ACGATGGTGGATGGGT
TTCCTATTATTATCCCT
GGAAGCTAAGGATATAGG
CGCTGATGTGAGGTGGT
TCGGTCTATCTGCATGC
TAGCATGGATATTGATCG
TGGCTTATAGGCTAGTCG

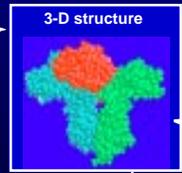
Protein sequence
MVLIVLAIIVLISDKRI
VTREGSWQIPCMNVNER
KRKKTDEDDHIVLILGSP
ILLNNASAIIVLPEDESA
SDSGPLIILKRKEKRWKL
LALAMAREENSPNCTGGT
PLIKRESAEDSEDLRRED

Sequence similarity

Cross-references



Text Similarity



Structural Similarity

Entrez (1996)

NCBI

MEDLINE



Nucleotide Sequences

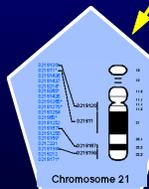
...TGG...
...CTATTAT...
...GAAGCTAAGGAT...
...CGCTGATGTGAGGTG...
...TCGGTCTATCTGC...
...TAGCATGGATATTG...
...TGGCTTATAGGCT...
...CTGATGTGAGC...

Protein Sequences

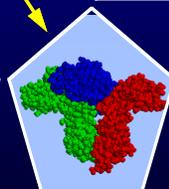
...ILA...
...REGSWQIP...
...RKKGREGDHIVL...
...ILLNNAWASVLPES...
...SDSGPLIILHEREK...
...LALAMAREENSPN...
...PLIKRESAEDSED...
...KKTDEDDHIVI...

Links

Genomes



Structures



Netscape - [Genomes Organism Representation]

File Edit View Go Bookmarks Options Directory Window Help

NCBI **Entrez** Genomes

Search Field: All Fields Mode: Automatic

Search Reset

Genomes Help

Prominent Organisms

- [1] [Helicobacter pylori](#) NEW
- [23] [Homo sapiens](#)
chromosomes: [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [20](#), [21](#), [22](#), [X](#)
- [21] [Mus musculus](#)
chromosomes: [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#), [9](#), [10](#), [11](#), [12](#), [13](#), [14](#), [15](#), [16](#), [17](#), [18](#), [19](#), [X](#), [Y](#)
- [5] [Drosophila melanogaster](#)
chromosomes: [1](#), [2](#), [3](#), [4](#), [Y](#)
- [6] [Caenorhabditis elegans](#)
chromosomes: [I](#), [II](#), [III](#), [IV](#), [V](#), [X](#)
- [16] [Saccharomyces cerevisiae](#)
chromosomes: [I](#), [II](#), [III](#), [IV](#), [V](#), [VI](#), [VII](#), [VIII](#), [VIII](#), [IX](#), [X](#), [XI](#), [XII](#), [XIII](#), [XIV](#), [XV](#), [XVI](#)
- [1] [Escherichia coli](#)
- [1] [Haemophilus influenzae](#)
- [1] [Mycoplasma genitalium](#)
- [1] [Mycoplasma pneumoniae](#)
- [3] [Methanococcus jannaschii](#)

GenBank Genomes Division

[Nucleotides] [Proteins] [3D Structures] [Taxonomy] [PubMed]

Document Done

The Human Genome Project, 2005

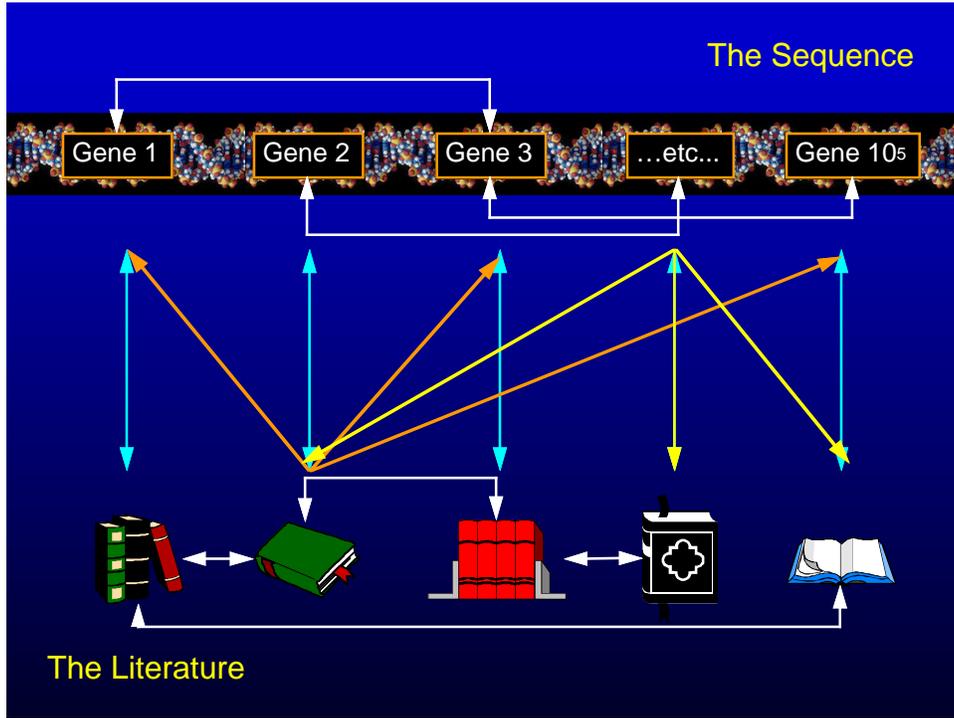
NCBI



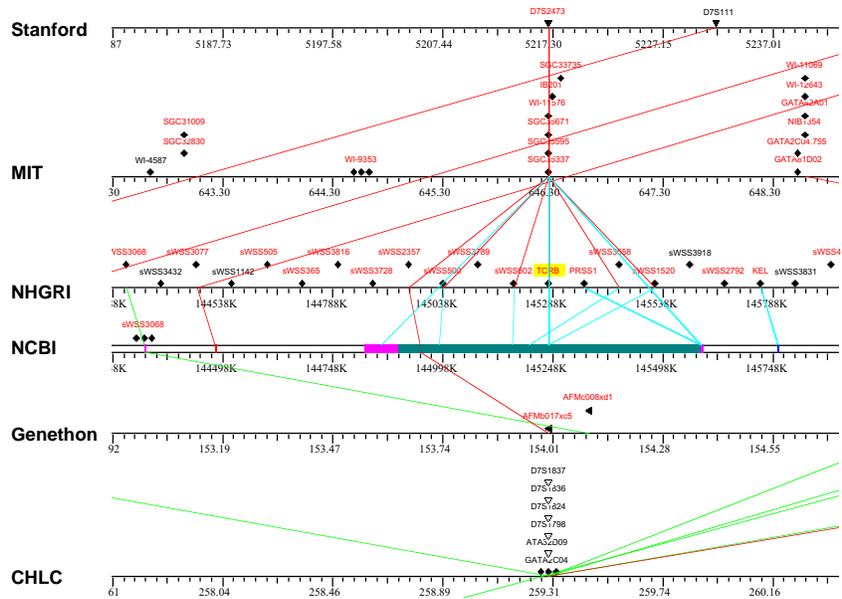
“The final, irreducible product of the Human Genome project will be...

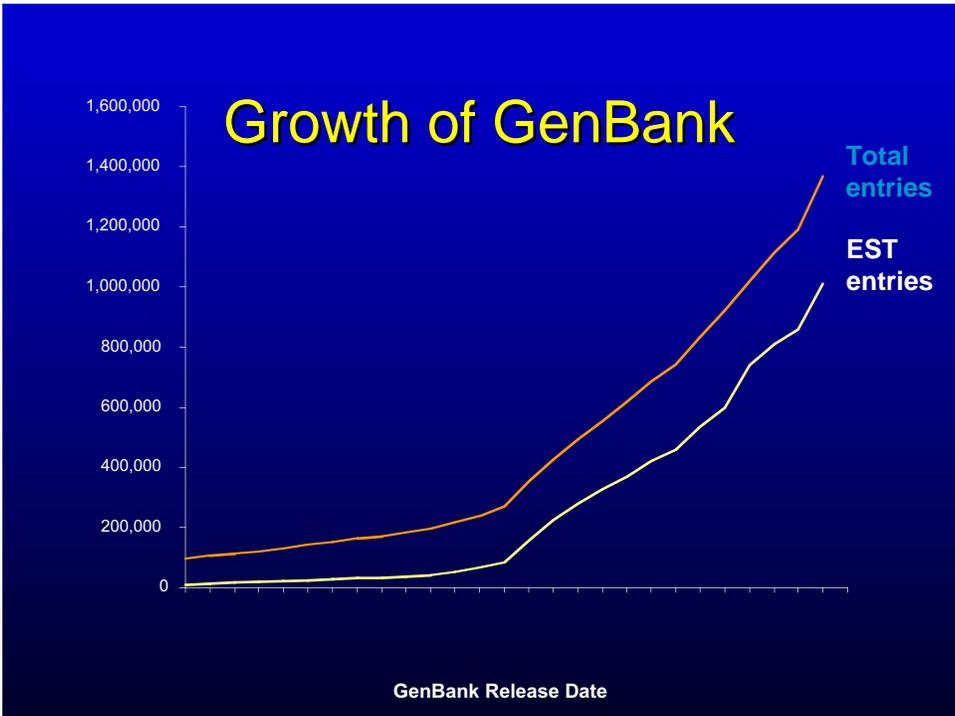
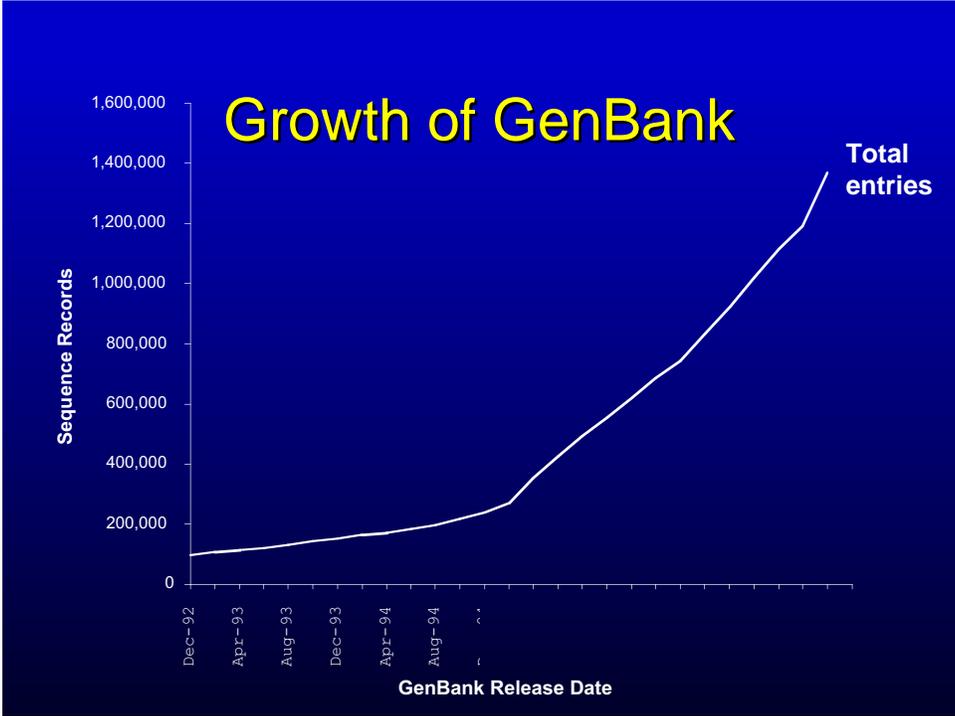
- The genetic map
- The sequence (map)
- The peer-reviewed scientific literature.”

—Maynard V. Olson, 1995



Chromosome 7: Local Views of T-Cell Receptor β Locus





What are “Expressed Sequence Tags” (ESTs)?

NCBI

- Partial, inaccurate cDNA sequences obtained by rapid survey sequencing of various cDNA libraries derived from various tissues, developmental stages, tumors, etc.
- There are now >750,000 human ESTs and > 200,000 mouse ESTs available
- **Ref:** Trends in Biochem. Sci. 20:295; 1995
- **URL:** <http://www.ncbi.nlm.nih.gov/dbEST/>

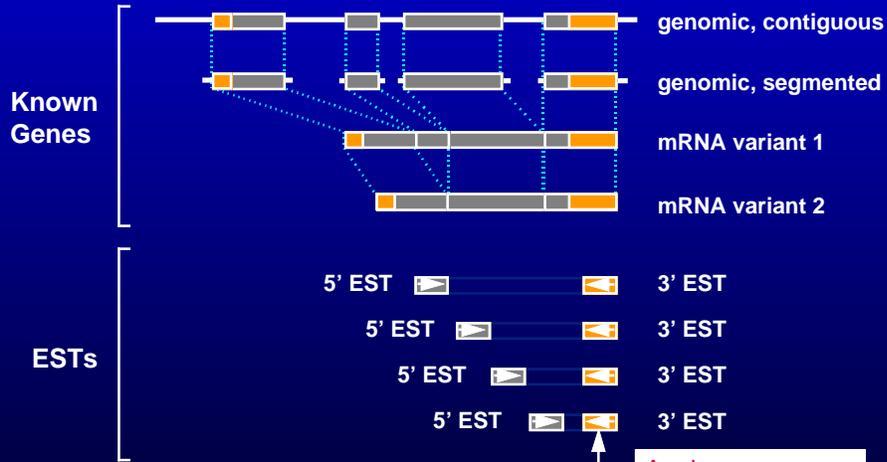
Evolution of EST Research Applications

NCBI

- **Gene Discovery (1991-present)**
Species-jumping & Comparative Genomics
- **Gene Mapping (1995-present)**
 - 16,000 human genes mapped to date
 - Mouse gene map in progress
 - Rat gene map planned
- **Gene Expression (1997-)**
Microarrays of cDNA clones & large-scale expression analysis

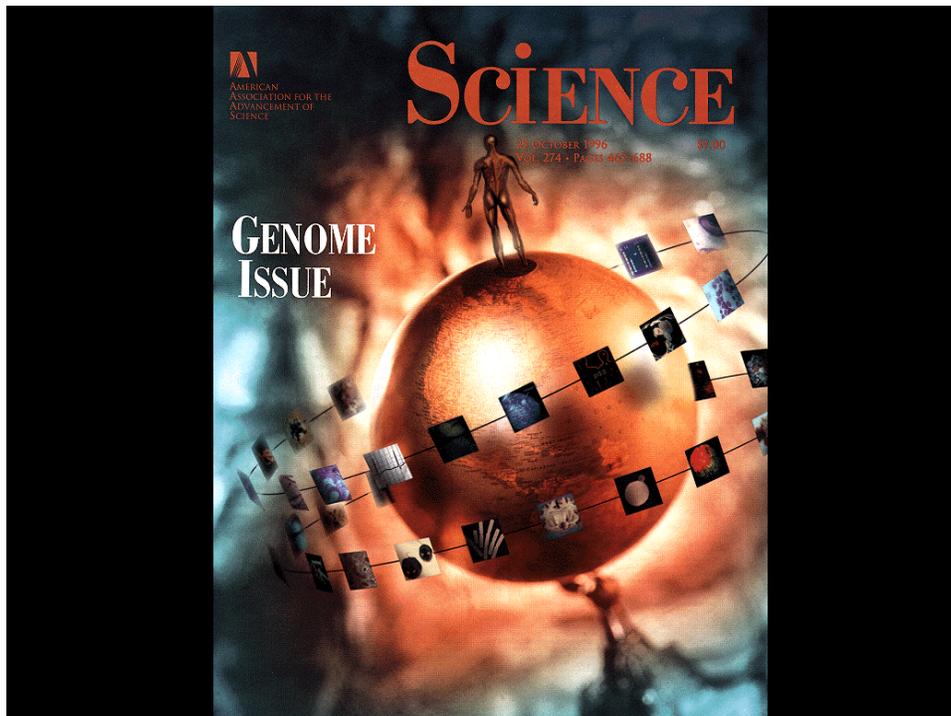
One Gene, Many Sequences

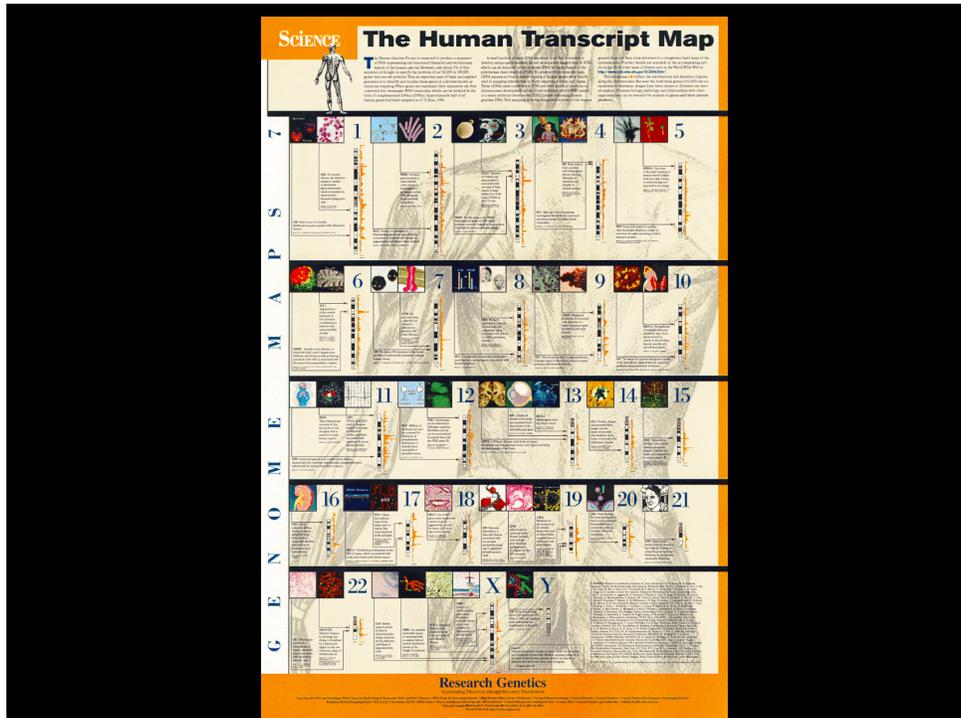
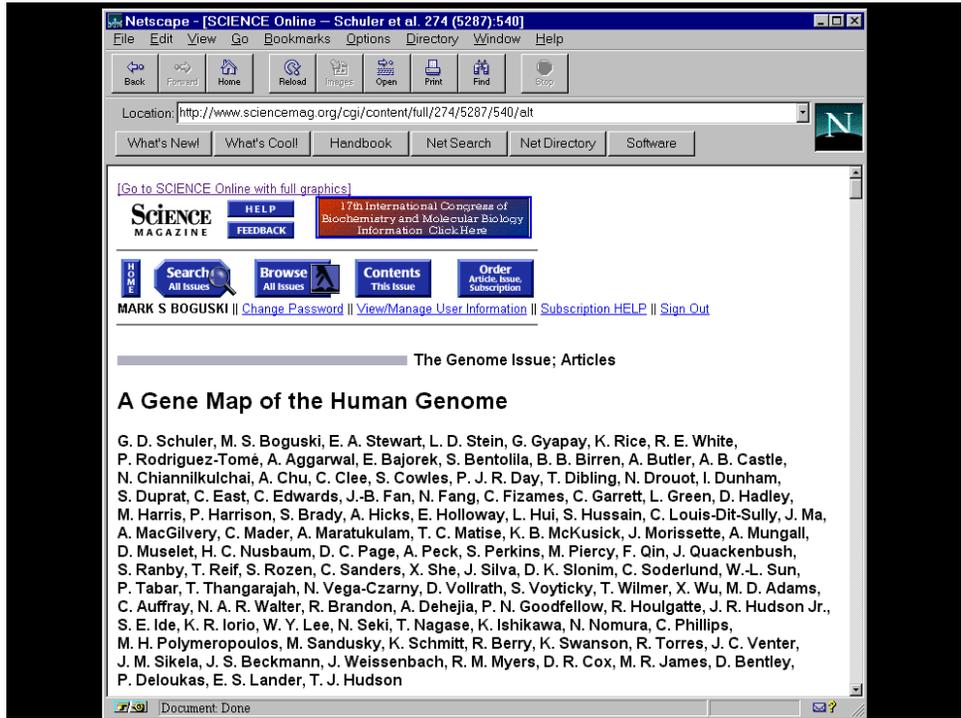
NCBI



Ref: Nature Genetics 10:369-371; 1995

Anchor sequence used for STS development





Netscape - [The Human Transcript Map]

File Edit View Go Bookmarks Options Directory Window Help

Location: <http://www.ncbi.nlm.nih.gov/SCIENCE96/>

What's New! What's Cool! Handbook Net Search Net Directory Software

SCIENCE The Human Gene Map
GENOME MAPS 7 NCBI

A Gene Map of the Human Genome

The Human Genome Project is expected to produce a sequence of DNA representing the functional blueprint and evolutionary history of the human species. However, only about 3% of this sequence is thought to specify the portions of our 50,000 to 100,000 genes that encode proteins. Thus an important part of basic and applied genomics is to identify and localize these genes in a process known as transcript mapping. When genes are expressed, their sequences are first converted into messenger RNA transcripts, which can be isolated in the form of complementary DNAs (cDNAs). Approximately half of all human genes had been sampled as of 15 June, 1996.

A small portion of each cDNA sequence is all that is needed to develop unique gene markers, known as sequence tagged sites or STSs, which can be detected in chromosomal DNA by assays based on the polymerase chain reaction (PCR). To construct a transcript map, cDNA sequences from a master catalog of human genes were distributed to mapping laboratories in North America, Europe, and Japan. These cDNAs were converted to STSs and their physical locations on chromosomes determined on one of two radiation hybrid (RH) panels or a yeast artificial chromosome (YAC) library containing human genomic DNA. This mapping data was integrated relative to the human genetic

Genome Maps 1997
[Show Featured Genes](#)



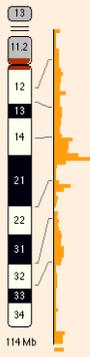
Document Done

Netscape - [Chromosome 13]

File Edit View Go Bookmarks Options Directory Window Help

SCIENCE The Human Gene Map
GENOME MAPS 7 NCBI

Chromosome 13



118 cM



BRCA2 Mammogram showing breast cancer
IMAGE CREDIT: Pat Connelly, Miami Valley Hospital, Dayton, OH, USA



RB1 Childhood tumors of the retina are associated with inactivation of the retinoblastoma gene
IMAGE CREDIT: K. Sathff, SCIENCE

Document Done

Netscape - [BRCA2]

File Edit View Go Bookmarks Options Directory Window Help

NCBI ? The Human Gene Map

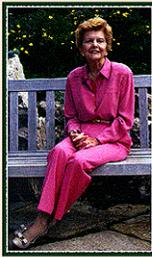
NLM FEATURED GENE

Breast Cancer

Breast cancer is often detected when there are visible changes in the breast, such as a lump, thickening, swelling, dimpling, skin irritation, distortion, retraction, scaliness, pain, tenderness of the nipple, or nipple discharge.

Breast cancer is the second major cause of cancer death in American women. An estimated 44,560 lives (44,300 women and 260 men) will be lost to breast cancer in the U.S. in 1996. About 184,300 new invasive cases among women will be diagnosed this year. About 1,400 new cases of breast cancer will be diagnosed among American men during this same period. Breast cancer incidence rates among women increased about 4x year between 1982 and 1987, but recently leveled off at about 110 cases per 100,000. Most of the recent increase in rates is believed to be due to marked increases in mammography use, allowing the detection of early-stage breast cancers, frequently before they become clinically apparent.

The risk of breast cancer increases with age. The risk is higher in a woman who has: a personal or family history of breast cancer; some forms of benign breast disease; an early beginning of menstruation; late menopause; lengthy exposure to cyclic estrogen; never had children, or had the first live birth at a later age; lower socioeconomic status and lower education level. To date, knowledge about risk factors has not translated into practical ways to prevent breast cancer. Since women may not be able to alter their personal risk factors, the best opportunity at present for reducing mortality is through early detection. It's recommended that asymptomatic women aged 40-49 should have a screening mammogram every 1-2 years; and women aged 50 and over, every year. In addition, a clinical breast exam is recommended every three years for women 20-40, and every year for



Betty Ford
Former breast cancer patient and activist on behalf of expanded research and education.

"While we all work toward a cure, education, research and increased access to treatment remain our best allies in the fight against breast cancer."

Document Done

Netscape - [BRCA2]

File Edit View Go Help

become an important part

For the public

For More Information:

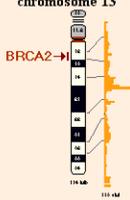
- Cancer Information Service, National Cancer Institute, National Institutes of Health, 1-800-4-CANCER, <http://www.cic.nci.nih.gov>.
- American Cancer Society 1-800-ACS-2345 (within each state with a divisional office), <http://www.cancer.org>.

For Physicians

For Scientists

IMAGE CREDIT: Pat Connelly, Miami Valley Hospital, Dayton, OH, USA

The BRCA2 gene maps to chromosome 13



Database Records

- [Breast cancer entry in OMIM](#)
- [BRCA2 entry in OMIM](#)
- [BRCA2 mRNA sequence in GenBank](#)

Address comments and suggestions to info@ncbi.nlm.nih.gov

Document Done

Netscape - [OMIM ENTRY 600185]

File Edit View Go Bookmarks Options Directory Window Help

OMIM Home Search Comments

***600185 BREAST CANCER 2, EARLY-ONSET; BRCA2**

Alternative titles; symbols

BREAST CANCER, TYPE 2

TABLE OF CONTENTS

- [TEXT](#)
- [ALLELIC VARIANTS](#)
 - [View List of allelic variants](#)
- [REFERENCES](#)
- [CONTRIBUTORS](#)
- [CREATION DATE](#)
- [EDIT HISTORY](#)
- [CLINICAL SYNOPSIS](#)

Database Links

[MEDLINE](#) [Protein](#) [DNA](#) [HGMD](#) [Gene Map](#) [GDE](#)

Gene Map Locus: 13q12.3

Note: pressing the  symbol will find the citations in the NCBI MEDLINE subset whose text most closely matches the text of the preceding OMIM paragraph, using the Entrez MEDLINE neighboring function.

TEXT

Document Done

Link to OMIM -- Online Mendelian Inheritance in Man

Gene Map User Feedback

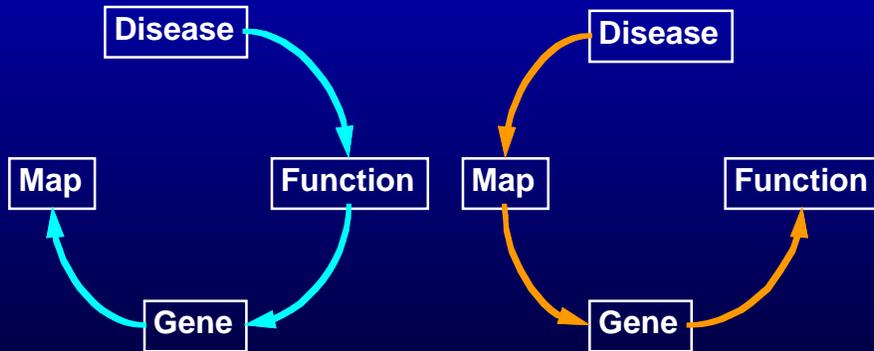
NCBI

From: user@aol.com
Date: Tue, 29 Oct 1996 17:51:22 -0500
To: info@ncbi.nlm.nih.gov
Subject: awesome

I am a 6th grade student on Long iSLAND AND i AM LEARNING ALL ABOUT THIS AND i THINK IS IS MAGNIFECENT THAT YOU GUYS CAN DO ALL THIS

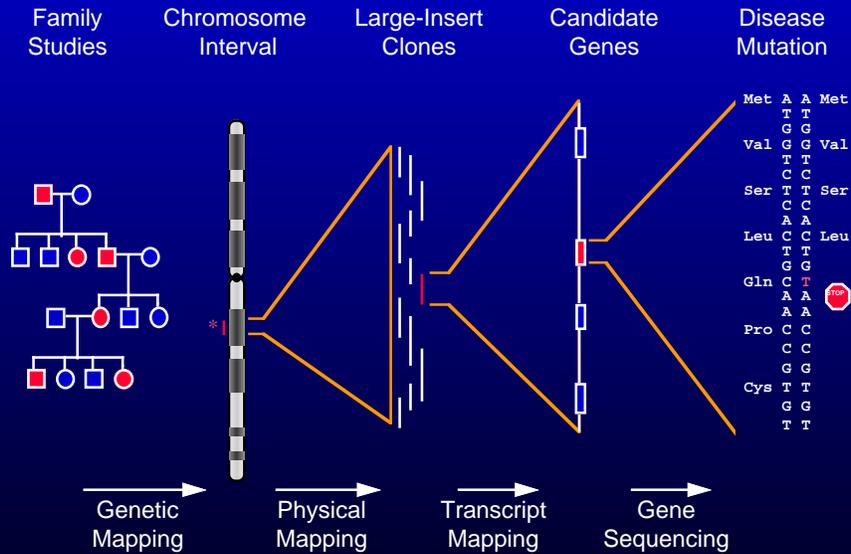
Functional Cloning and Positional Cloning

NCBI



Standard Positional Cloning

NCBI



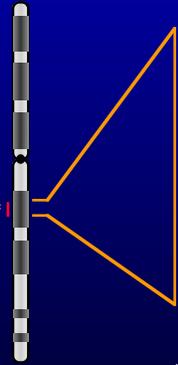
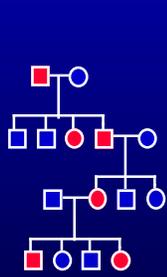
“Positional Candidate” Approach

NCBI

Family Studies

Chromosome Interval

Candidate Gene



Met A A Met
T T G
G G G Val
T C T C T
Ser T C C T Ser
C A C A C
Leu C C C T G Leu
T G T G T
Gln C C A A A
A A A A A
Pro C C C C C
G G G G G
Cys T T G T
T T T T T

Recent example (July '97):

Parkinson's Disease

Genetic Mapping

Mutation Detection

Portage - [Jobq] Map Search

NCBI The Human Gene Map

Map Search

Chromosome **11**

GO: 11 12 13 14 15 16 17 18 19 20 21 22 X Y

Search Interval

Enter the names of the Genbank markers which define the desired search interval.

Select Markers

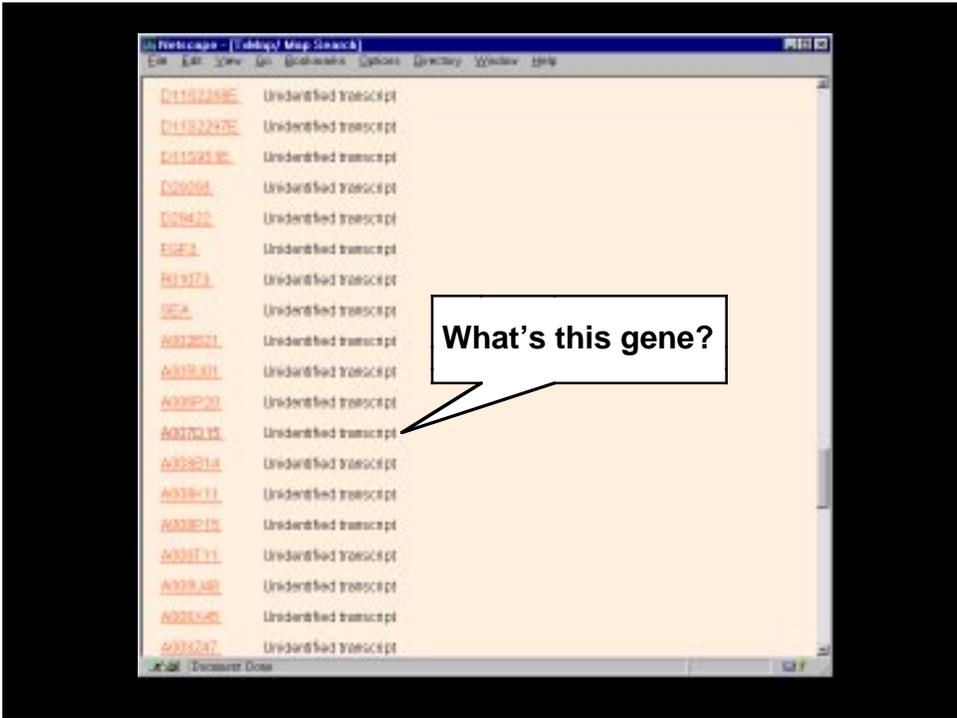
Interval:

144 cDNA markers in this interval

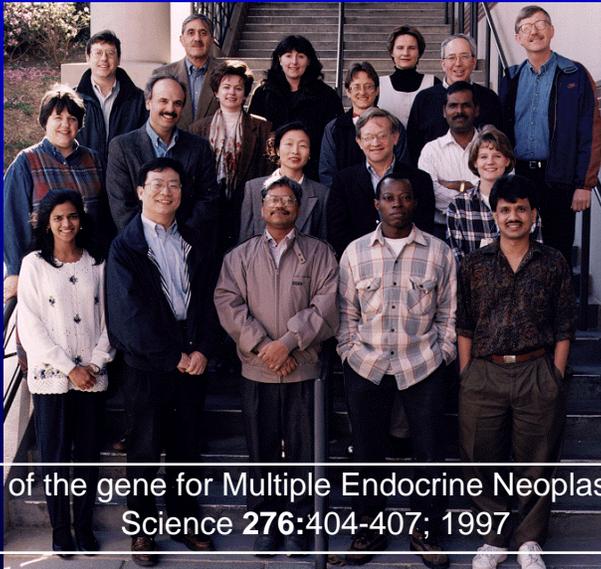
Between D11S913 and D11S1113 (70-71 cM)

- D11S1094E CHOLINE KINASE
- D11S1104E DNA-BINDING PROTEIN SMURF-2
- D11S1173A GLUTATHIONE S-TRANSFERASE P
- D11S1203Z Undeclared transcript
- D11S1201Z Undeclared transcript

Between D11S913 and D11S1307 (70-72 cM)



The MEN1 Group

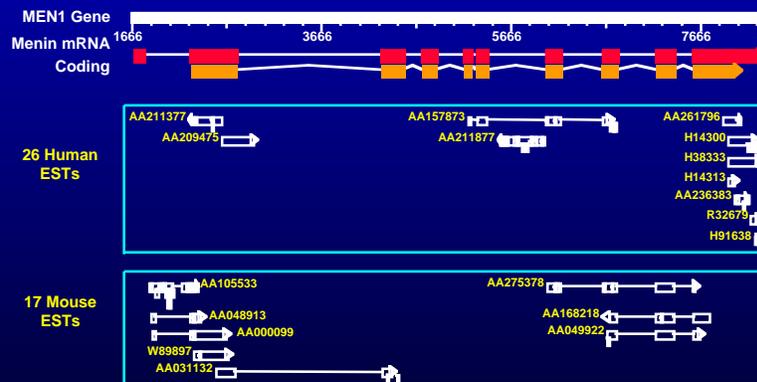


Cloning of the gene for Multiple Endocrine Neoplasia, Type I
Science 276:404-407; 1997

Absent: Lance Liotta, Bruce Roe

PowerBLAST* analysis of the MEN1 gene: human and mouse EST coverage

NCBI



*J. Zhang & T. Madden (1997) Genome Res. 7:649-56

National Cancer Institute Cancer Genome Anatomy Project

NCI CGAP

PROJECT GOAL

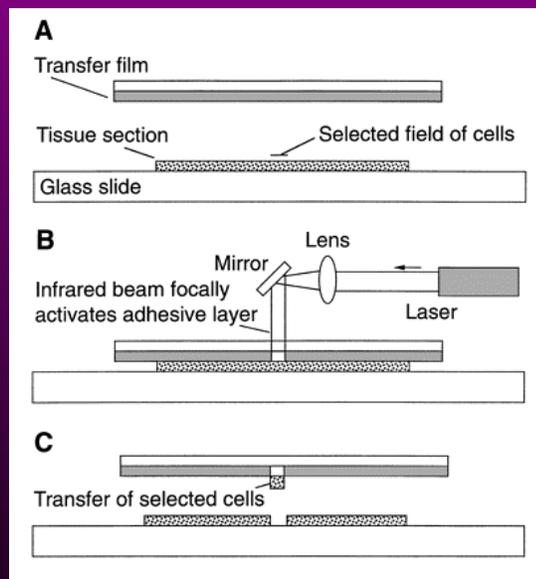
To achieve a comprehensive molecular characterization of normal, precancerous, and malignant cells

- Tumor Gene Index
- Transcript "Profiling"
- Physical DNA resources
- All data & material made available without restrictions

NCI Tumor Gene Index cDNA Libraries

NCI CGAP

- Initial target tissues are breast, prostate, colon, lung and ovarian tumors
- Microdissected as well as “bulk” specimens
- Normalized and non-normalized libraries
- Interplay between gene discovery and pathophysiologic analysis of cancer progression
- R & D of methods to assess gene expression in archival, embedded tissue

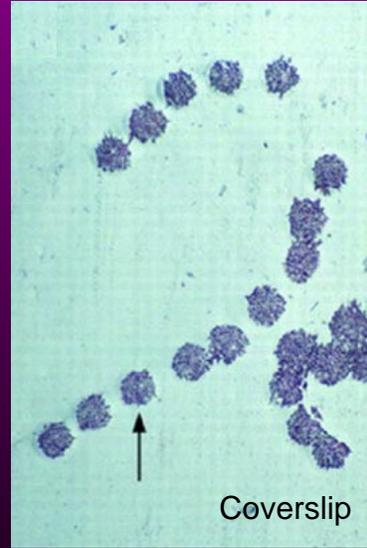
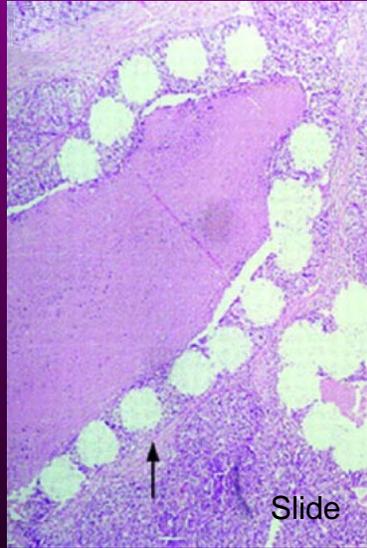


Laser Capture Microdissection

Emmert-Buck et al.
Science 274:998; 1996

NCI CGAP

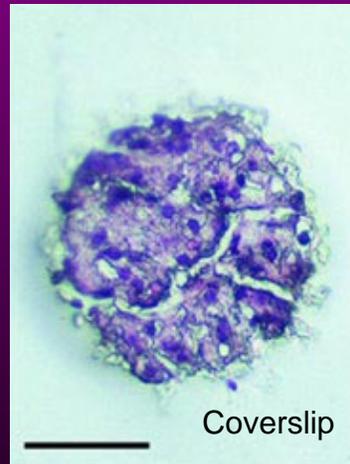
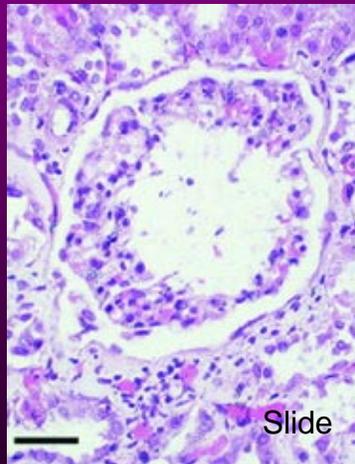
Laser Capture Microdissection: in situ breast carcinoma



NCI CGAP

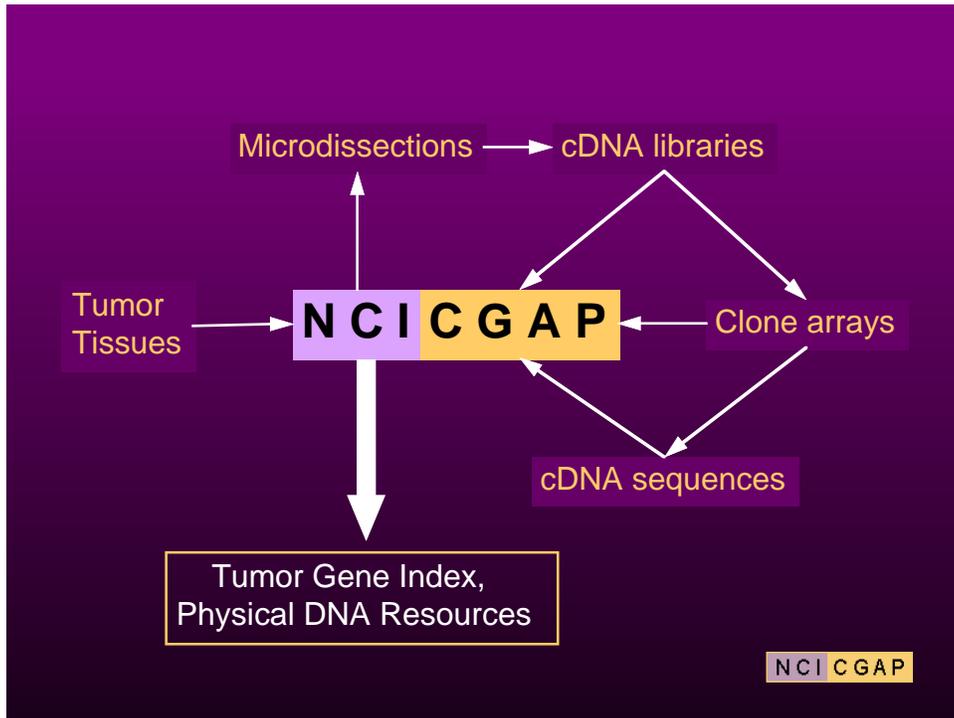
Emmert-Buck et al. (1996) Science 274:998

Laser Capture Microdissection: glomerulus, normal kidney



NCI CGAP

Emmert-Buck et al. (1996) Science 274:998



National Cancer Institute

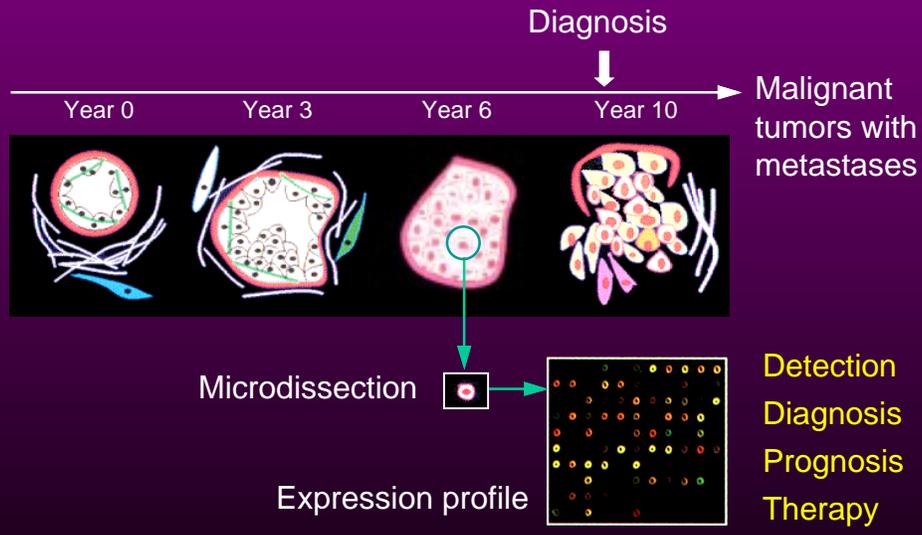
CGAP

Now appearing at a web browser near you!

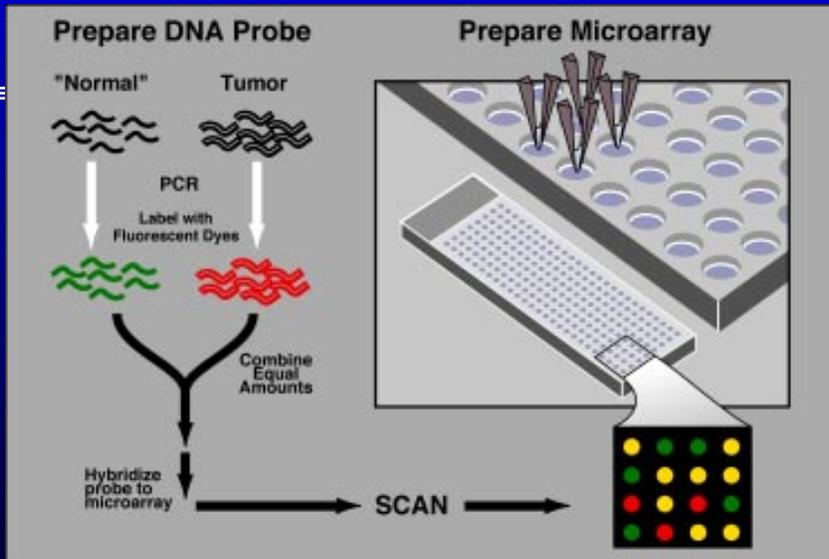
CGAP Web Site

Carcinogenesis and Tumor Progression

NCI CGAP

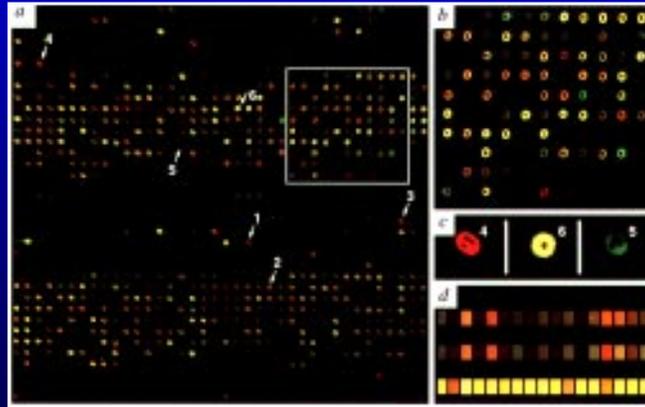


Gene Expression & Glass Microarrays



Large-scale Analysis of Gene Expression

NCBI

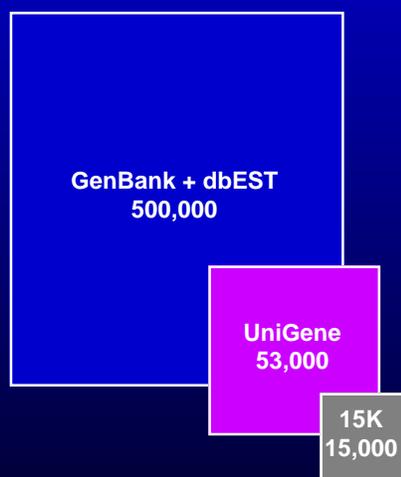


Use of a cDNA microarray to analyze gene expression patterns in human cancer

J. DeRisi, L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, and J.M. Trent, *Nature Genetics* 14, 457-460 (1996)

15,000 cDNA Clones to Array

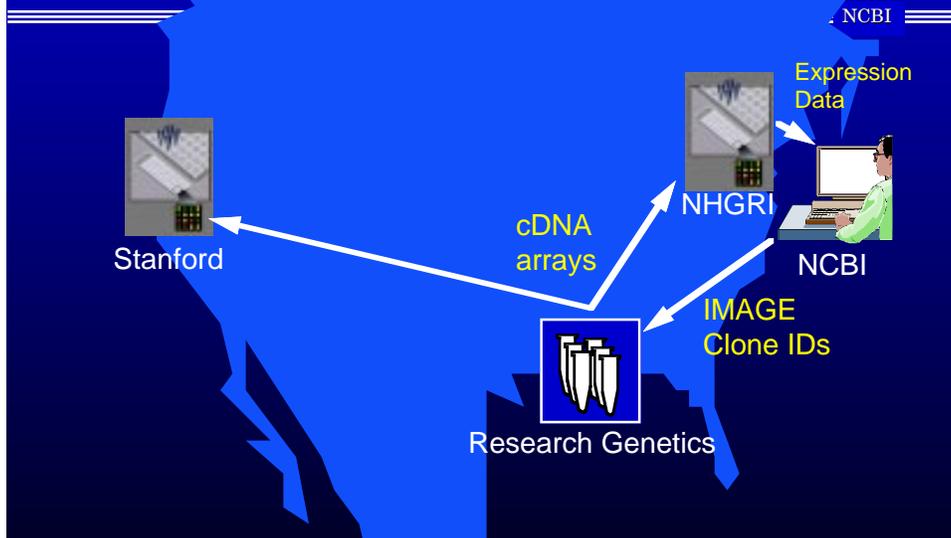
NCBI



Selection criteria:

1. Correspondence to functionally cloned human gene
2. Significant protein similarity
3. Included on the Human Transcript Map
4. Of specific research interest
5. Physical cDNA clone available

The "15K" Collaboration



15K Set: Portion of the FASTA file

NCBI

```
>M63960 Human protein phosphatase-1 catalytic subunit mRNA, complete cds
/cds=(29,1021) /gb=M63960 /gi=190515 /len=1367
GGGCAAGGAGCTGCTGGCTGGACGGCGGCATGTCCGACAGCGAGAAGCTCAACCTGGACT
CGATCATCGGGCGCCTGCTGGAAGTGCAGGGCTCGCGGCCTGGCAAGAATGTACAGCTGA
CAGAGAACGAGATCCGCGGTCTGTGCCTG...etc.
>W37493 zc10g02.s1 Homo sapiens cDNA, 3' end /contact=Wilson,RK
/clone=321938 /clone_end=3' /gb=W37493 /gi=1319087 /len=348
GAGAATCCANCTTTGACCTTTATTCAAGAGACCAGATGGGTTGCCCCAGGATCCGGCTGC
CAGCCTGAGGCCAAGCACGGCTGGAGACCCACGACCTGGCCTGCCGTTGCCCTGAGCTGC
AGCCTCGGGCCCCAGGATCCTGCTCACAGT...etc.
>H42556 yo63c10.r1 Homo sapiens cDNA, 5' end /contact=Wilson,RK
/clone=182610 /clone_end=5' /gb=H42556 /gi=918608 /len=544
GTGTGACCAGACATGCAACCGNCATCTATGGTTTCTACGNATGNAGTGNCAAGCAGNACG
NCTNACAACATCAAACCTGTGGNAAAACCTTCACTGACTGNCTTCAACTGNCTGNCCCA
TCGCGNCCATAGTGGACGTAAGATCTTCTGNCTGNCCACGGAGGCCTGTTCCCCGGA
CCTGNCAGTTCTATGGNAGCAGATTCGG...etc.
```

Sample 15K Cluster Report

NCBI

TITLE: Human protein phosphatase-1 catalytic subunit mRNA,
complete cds

CLONE: 488948

FLAGS: Gene SwissProt Mapped

CLUST: Hs.1001

GENES: J04759 X70848 M63960

3'EST: AA115517 AA004413 H97499 W02143 W37493

5'EST: AA115516 N42323 N41606 H42559 H42556

//

15K Cluster Report: Sample 2

NCBI

TITLE: ESTs, Highly similar to HYPOTHETICAL 34.9 KD PROTEIN IN
FRE2-JEN1 INTERGENIC REGION [Saccharomyces cerevisiae]

CLONE: 310438

FLAGS: SwissProt Mapped

CLUST: Hs.10018

3'EST: R44498 N33766 T93144 N48585 N99970

5'EST: N47271 R14393 T56201 R68523 W30909

//

15K Cluster Report: Sample 3

NCBI

TITLE: ESTs
CLONE: 302998
FLAGS: Mapped
CLUST: Hs.24297
3'EST: H97384 AA147620 N24156 AA157374 H41351 R98414
H91639 N51774 R28488 H14277 H38296 N90077
5'EST: W38527 AA147612 AA157873 H91638 R32679
H38333 H14300 N36190
//

This is the MEN1 Gene

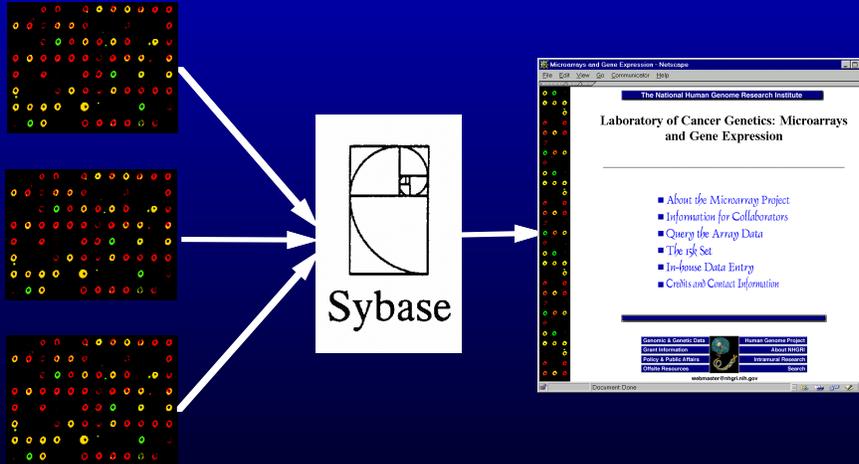
Informatics Issues in Large-Scale Studies of Gene Expression

NCBI

- 4 Resources (arrayed genes, probes)
- 4 Laboratory information management system
- 4 Information retrieval & query systems

Expression Array Database

NCBI



The screenshot shows a Netscape browser window displaying a UniGene Cluster Report. The report is titled 'Damage-specific DNA binding protein 1 (127 kD)' and lists various identifiers and associated data. A yellow box with the text 'UniGene Cluster Report' is overlaid on the right side of the browser window.

Title	Damage-specific DNA binding protein 1 (127 kD)
Cl_id	1466
Clone	376234
Flags	Gene Mapped SwissProt
Gsymb	DDB1
Cymap	11q12-q13
Txmap	11:66-66 11:67-67 11:63-63
Clust	Hs.2165
Type	1
PI/Row/Col	72/G/12/
Genes	U18299 U32986 L40326
3'EST	R24264 N94513 T32406 T32988 N66406 H53855 T50093 R53076 T33118 T33789 N80854 T67517 T54158 T62193 T62836 W48654 H90167 H90867 H12892 AA179709 AA040781 W72213 W72401 AA022905 AA024833 N78589 N78833 T48843 H63963 H13838 T70287 N21651 W84440 T03687 AA148958 AA129271 R71571
5'EST	R24373 H53480 R53165 T53766 T62002 T62685 H90211 H90274 H12891 AA180528 W17195 AA041217 AA022973 AA024789 W23963 T48842 H63690 W51866 H13884 T69516 T80862 W76527 W77900 W84366 W07183 W07519 AA148957 AA129306
Sequence	AA040781 AA041217

NCBI

PubMed nucleotide query - Netscape

File Edit View Go Communicator Help

NCBI Entrez Nucleotide QUERY BLAST Entrez ?

Other Formats: FASTA Graphic

LOCUS AA041217 500 bp mRNA EST 01-FEB-1997

DEFINITION zf07c06.r1 Soares fetal heart NbHH19W Homo sapiens cDNA clone
376234 5' similar to PIR:S37614 S37614 DNA-binding protein - green monkey ;.

ACCESSION AA041217

NID g1517451

KEYWORDS EST.

SOURCE human.

ORGANISM [Homo sapiens](#)
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 500)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1305 Std Error: 0.00

<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wgetorg?id=9606>

NCBI

GenBank record

Data Mining in Gene Expression Arrays

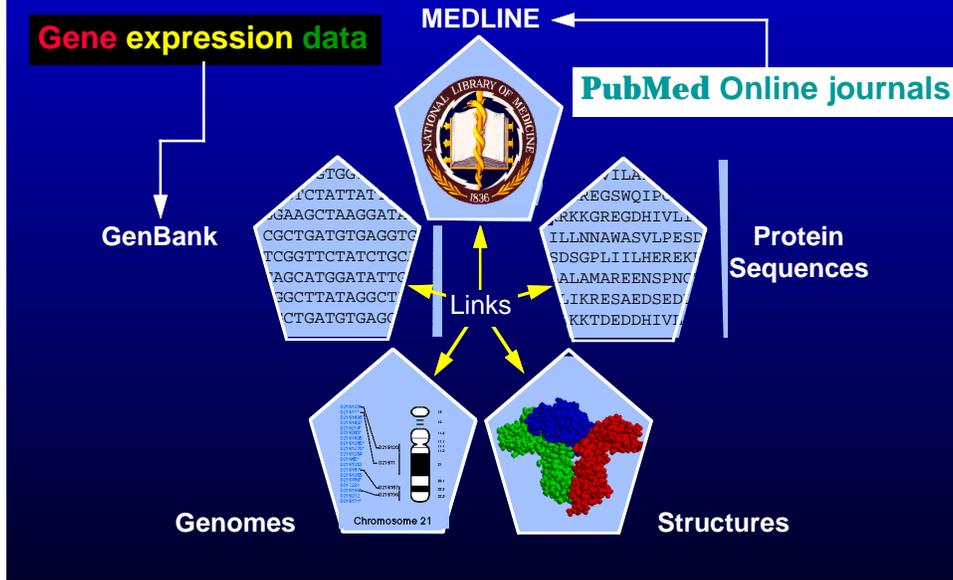
NHGRI / NCBI

- Finding associations
 - Which genes tend to be expressed together?
- Finding sequential patterns
 - Which genes are expressed in succession, i.e. in a pathway?
- Clustering the data
 - Does this set of genes have any common features?
- Classifying the data
 - Has this expression pattern been observed in other experiments?
- Predicting values
 - Can we extrapolate to other conditions under which these genes may be expressed?



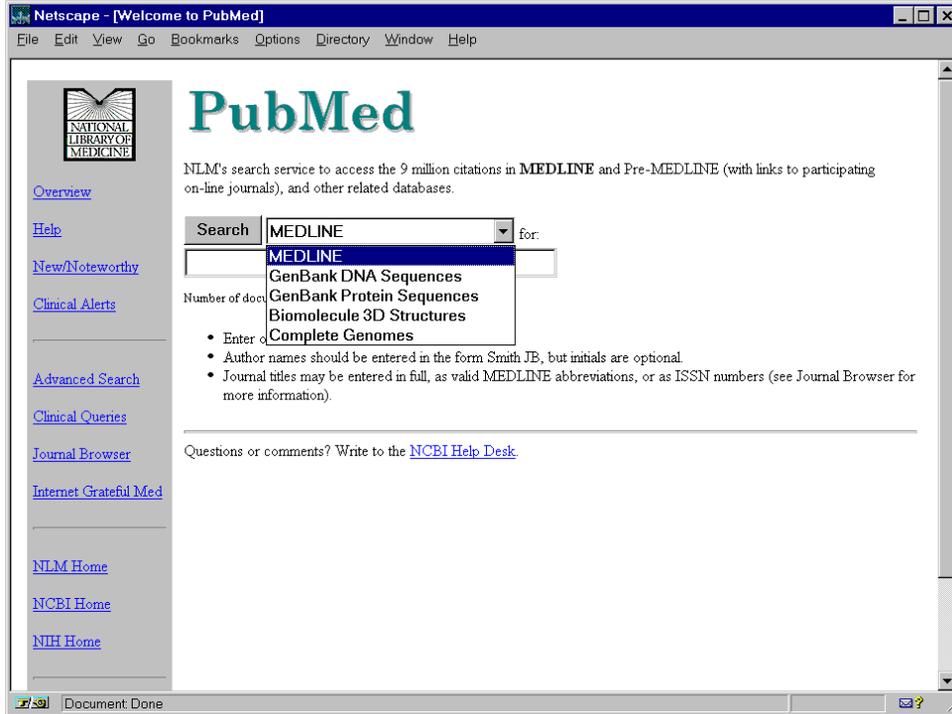
Data Mining in Gene Expression Arrays

NCBI



PubMed

- An information retrieval system, based on Entrez “neighboring” technology
- Includes all 10 million articles in MEDLINE
- Available without charge via the World Wide Web
- Links to full-text, online journals available from various publishers



Netscape - [Entrez MEDLINE query]

File Edit View Go Bookmarks Options Directory Window Help

NCBI PubMed PubMed QUERY PubMed ?

Other Formats: MEDLINE

Links: Related Articles **Science**

Science 276 (5321): 2045-2047 (Jun 27 1997)

Click here.

Mutation in the alpha-synuclein gene identified in families with Parkinson's disease

Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL

M. H. Polymeropoulos, C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, J. Rubenstein, R. Boyer, R. L. Nussbaum, Laboratory of Genetic Disease Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-1430,

Parkinson's disease (PD) is a common neurodegenerative disorder with a lifetime incidence of approximately 2 percent. A pattern of familial aggregation has been documented for the disorder, and it was recently reported that a PD susceptibility gene in a large Italian kindred is located on the long arm of human chromosome 4. A mutation was identified in the alpha-synuclein gene, which codes for a presynaptic protein thought to be involved in neuronal plasticity, in the Italian kindred and in three unrelated families of Greek origin with autosomal dominant inheritance for the PD phenotype. This finding of a specific molecular alteration associated with PD will facilitate the detailed understanding of the pathophysiology of the disorder.

PMID: 9197268, MUID: 97342853

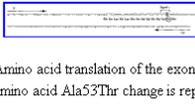
Save the above report in Macintosh Text format.

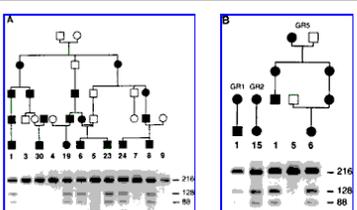
Document Done

Netscape - [SCIENCE Online - Polymeropoulos et al. 276 (5321):2045]

File Edit View Go Bookmarks Options Directory Window Help

the creation of a novel Tsp45 I restriction site (Fig. 1). Mutation analysis for the G209A change in the Italian kindred showed complete segregation with the PD phenotype with the exception of individual 30, who is affected but not carrying this mutation (Fig. 2A). This individual apparently inherited a different PD mutation from his father because we have shown that he shares a genetic haplotype with his unaffected maternal uncle, individual 3, for genetic markers in the PD linkage region.

 Fig. 1. DNA sequence of the PCR product used for mutation detection. Oligonucleotide primers are shown by arrows and the numerals 3 and 13. Intron sequence is shown in lower case and exon sequence in upper case. Amino acid translation of the exon is shown below the DNA sequence. The circled base represents the G209A change in the mutant allele. The resulting amino acid Ala53Thr change is represented by the circled amino acid. [View Larger Version of this Image (7K GIF file)]

 Fig. 2. Mutation analysis of the G209A change is shown in a subpedigree of the Italian kindred (A) and the three (GR1, GR2, GR5) Greek PD kindreds (B). Filled symbols represent affected individuals. Numerical identifiers denote the individuals immediately above. Tsp45 I digestion of PCR products is shown at the bottom of the figure, and fragment sizes are indicated on the right in base pairs. [View Larger Versions of these Images (43 + 27K GIF file)]

The frequency of this variation was studied in two general population samples, one consisting of 120 chromosomes of the parents of the CEPH (Centre d'Etude du Polymorphisme Humain) reference families, and the other consisting of 194 chromosomes of unrelated individuals from the blood bank in

Document Done

Full text of article available online

Netscape - [SCIENCE Online - Polymeropoulos et al. 276 (5321):2045]

File Edit View Go Bookmarks Options Directory Window Help

by a combination of factors, such as the relatively short life-span of rodents, the need for interaction with other cellular components not present in the rat, absence of a critical environmental trigger in the rodents, or a requirement for heterozygous status for the production of a phenotype.



Fig. 4. Sequence alignments of α -synuclein homologues in different species. Accession numbers for the sequences used were as follows: *Homo sapiens* Swiss-Prot P37840, *Rattus norvegicus* Swiss-Prot P37377, *Bos taurus* Swiss-Prot P33567, *Serinus canaria* GenBank L33860, and *Torpedo californica* Swiss-Prot P37379. Numbering on top of the alignments is according to the human sequence. Amino acid 53, which is the site of the Ala53Thr change, is circled. [\[View Larger Version of this Image \(33K GIF file\)\]](#)

Studies of early onset AD have previously documented that missense mutations can cause an adult onset neurodegenerative disorder. Of the 31 mutations described so far in the loci for presenilin 1 and 2, 30 were missense and 1 was a splice variant (14). Missense mutations in the prion protein have also been implicated in the amyloid production seen in Gerstmann-Strausstrinker disease (15). Studies in these neurodegenerative disorders have implicated mutations in proteins in initiating and propagating neuronal lesions leading to disease (16). The identification of mutations in these proteins provides a clue that should lead to the understanding of the etiology and pathogenesis of PD.

Although the mutation identified in the α -synuclein gene is unlikely to account for all sporadic and familial cases of PD, it may account for a significant proportion of those early-onset families with PD characterized by a mutation in the α -synuclein gene. The mutation we have described is directly related to only a small fraction of the total number of mutations in the α -synuclein gene. This provides a clue that should lead to the understanding of the underlying pathways resulting in the symptoms of PD.

REFERENCES AND NOTES

1. J. Parkinson, *An Essay on the Shaking Palsy* (Whittingham and Rowland, London, 1817); W. R. Gowers, *A Manual of Diseases of the Nervous System* (Blakiston, Philadelphia, PA, ed. 2, 1891), pp. 6366-6657.
2. A. M. Lazarrini, et al., *Neurology* **44**, 499 (1994).
3. L. I. Golbe, et al., *Ann. Neurol.* **27**, 276 (1990) [[Medline](#)]; M. H. Polymeropoulos, et al., *Science* **274**, 1197 (1996) [[Abstract/Full Text](#)].
4. K. Ueda, et al., *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11282 (1993) [[Medline](#)].
5. X. Chen, et al., *Genomics* **26**, 425 (1995) [[Medline](#)]; Y. Shitasaki, et al., *Cytogenet. Cell. Genet.* **71**, 54 (1995).
6. E. Leroy et al., in preparation; C. Lavedan et al., in preparation; M. Polymeropoulos et al., in preparation.

Document: Done

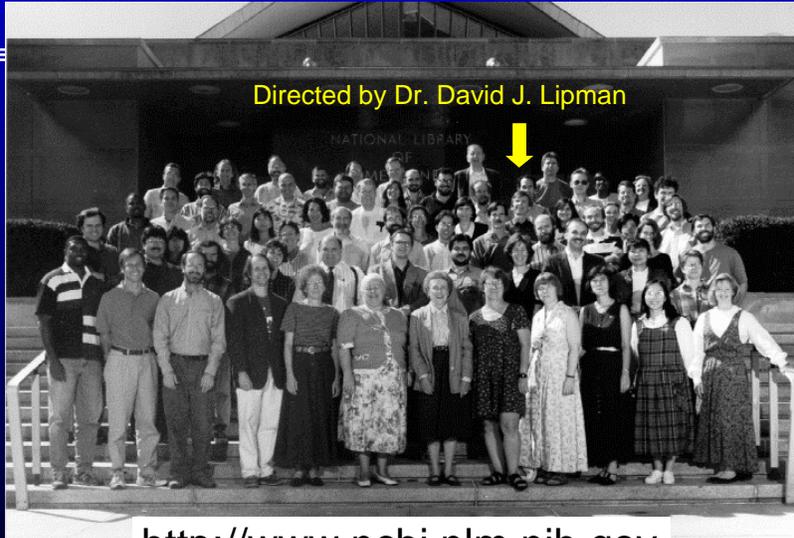
Citations linked back to Entrez/PubMed system

Acknowledgements

NCBI

<u>UniGene</u>		<u>15K cDNA Microarray</u>	
Greg Schuler	NCBI	Paul Mettzer	NHGRI
David Lipman	NCBI	Jeff Trent	NHGRI
Webb Miller	Penn State U.	Mike Bittner	NHGRI
Alejandro Schaeffer	NCBI	Yidong Chen	NHGRI
		Olga Ermolaeva	NHGRI/NCBI
<u>Transcript Mapping</u>		Lou Staudt	NCI
Tom Hudson	MIT/Whitehead	Pat Brown	Stanford U.
David Cox	Stanford U.	Mike Eisen	Stanford U.
Jean Weissenbach	Genethon	Jim Hudson	Res. Genetics
David Bentley	Sanger Centre	<u>NCI-CGAP</u>	
Michael James	Oxford	Bob Strausberg	NCI
et al.		Carol Dahl	NCI
		Lance Liotta	NCI
<u>Digital Differential Display</u>		Mike Emmert-Buck	NCI
John Spouge	NCBI	David Krizman	NCI
		David Lipman	NCBI
		Ken Katz	NCBI
		Carolyn Tolstoshev	NCBI

The National Center for Biotechnology Information



Directed by Dr. David J. Lipman

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