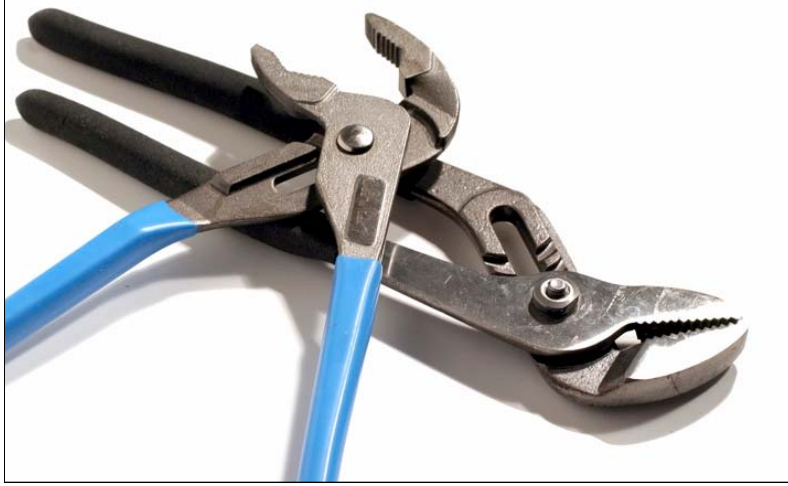
joanne@msl.ubc.ca

Bioinformatics

Common tools, useful databases, and tricks of the trade.



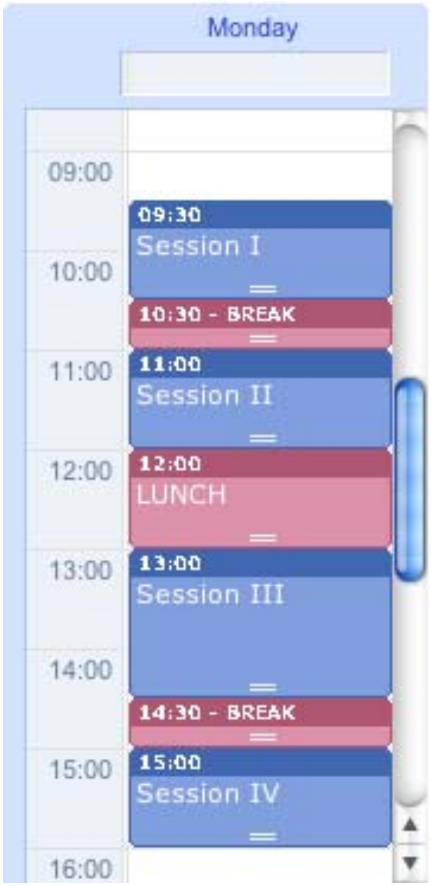
bioteach.ubc.ca/bioinfo2008

Workshop Schedule

- Laptops, available here for your use 9am - 4:30pm
- wireless login

  mslguest

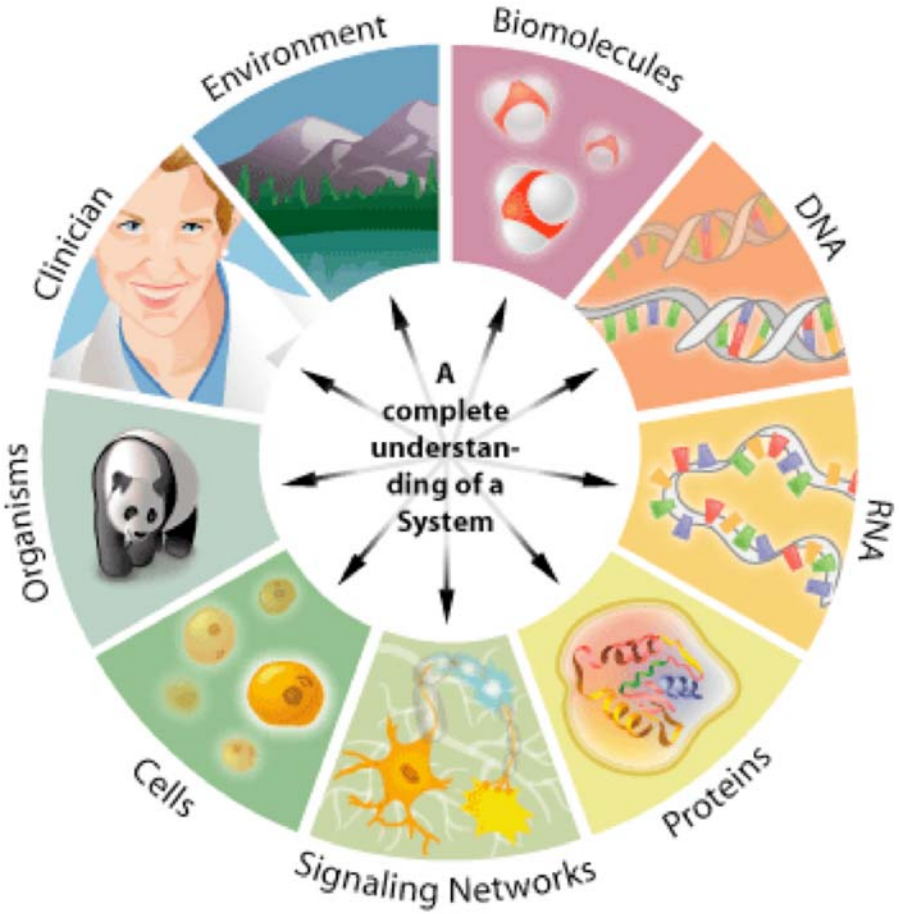
  4myguest
- Vancouver guide books available



Today's Topics

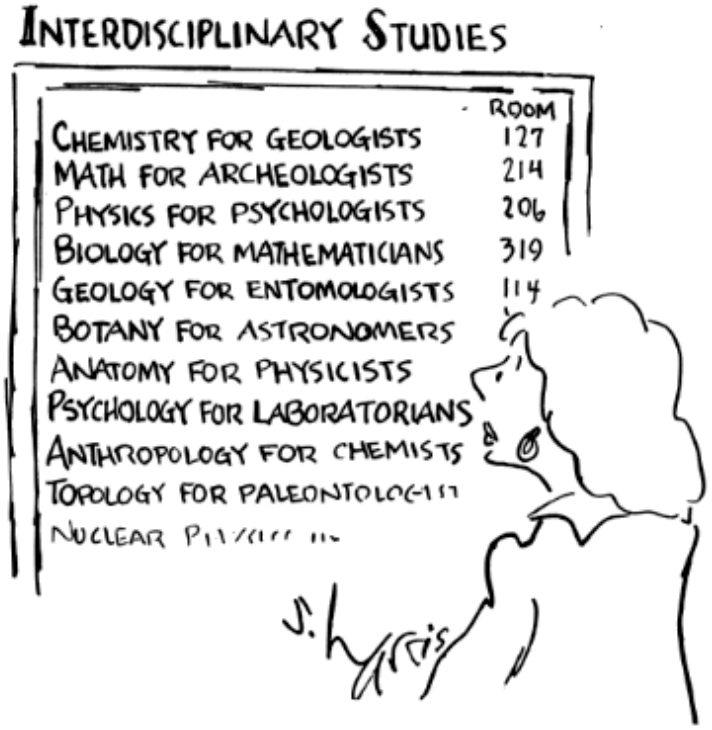
- **DNA Sequencing -** Generating Data & Emerging Technologies
- **Sequence Databases** - Public Resources at the NCBI
- **GUIDED TOUR** - Advanced Tips & Tricks for Searching Entrez
- **PRACTICAL EXERCISES** - Navigating Links, Retrieving Data with Entrez, and Searching PubMed

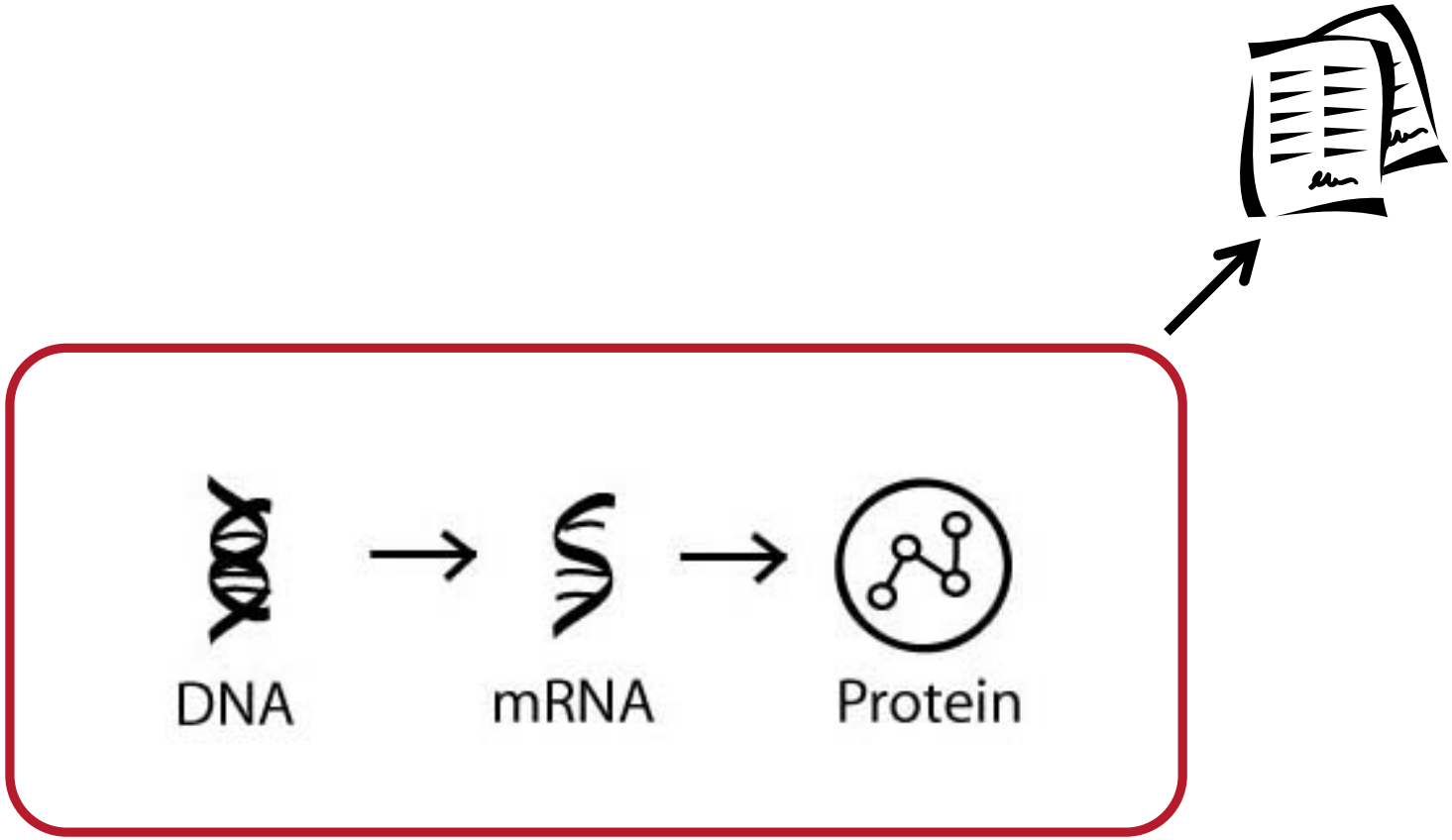
What is Bioinformatics?

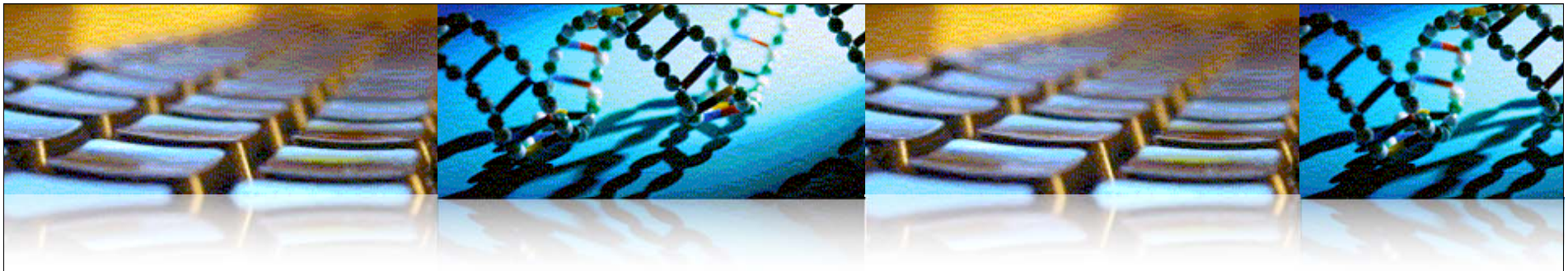


Bioinformatics for ~~Biologists~~

Government Employees

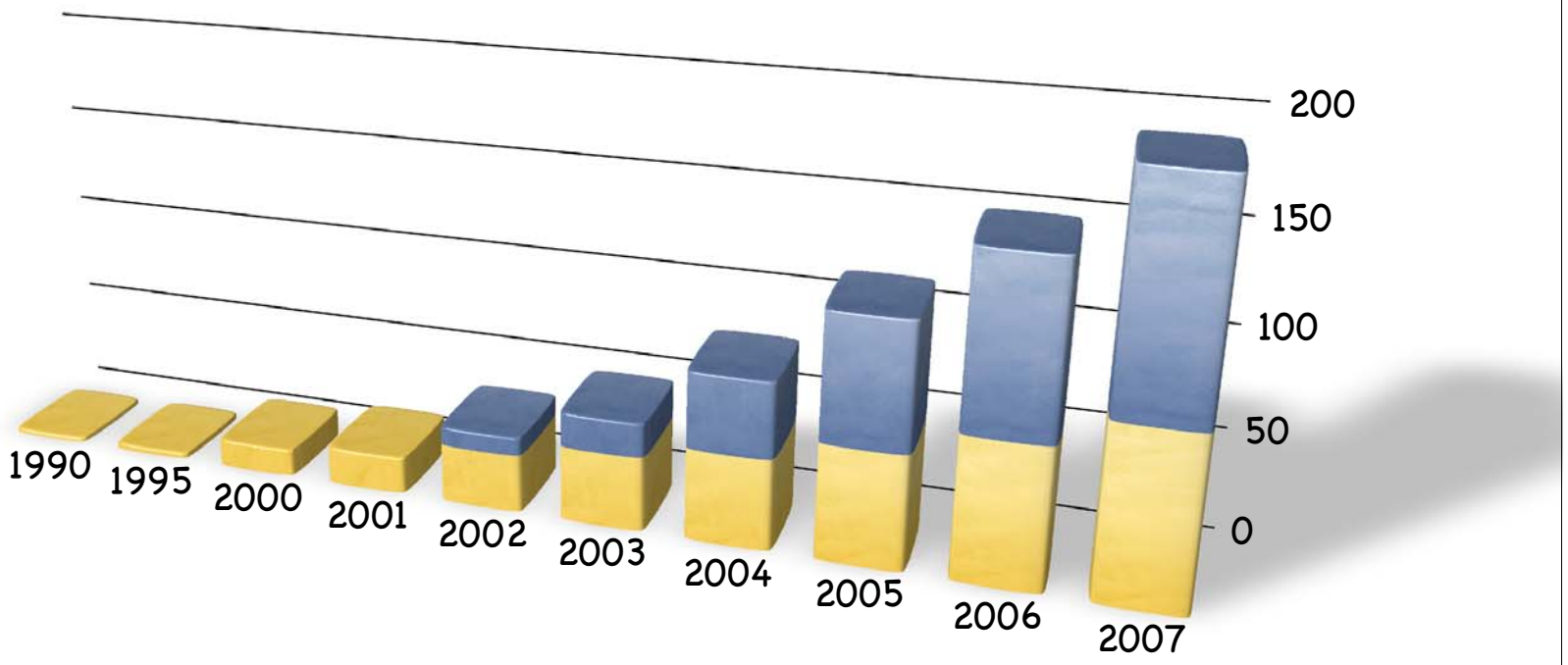



DNA
mRNA
Protein

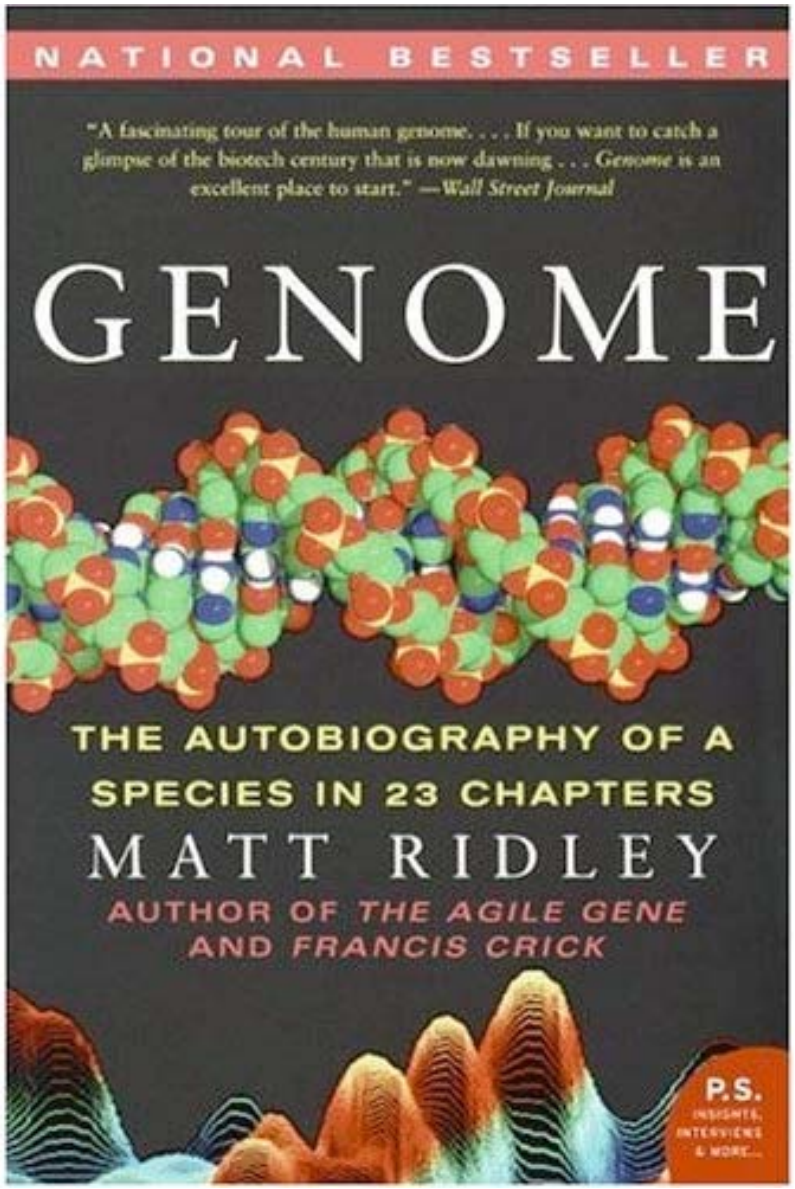


By the end of the morning session, you will define bioinformatics in your own words

Growth of GenBank

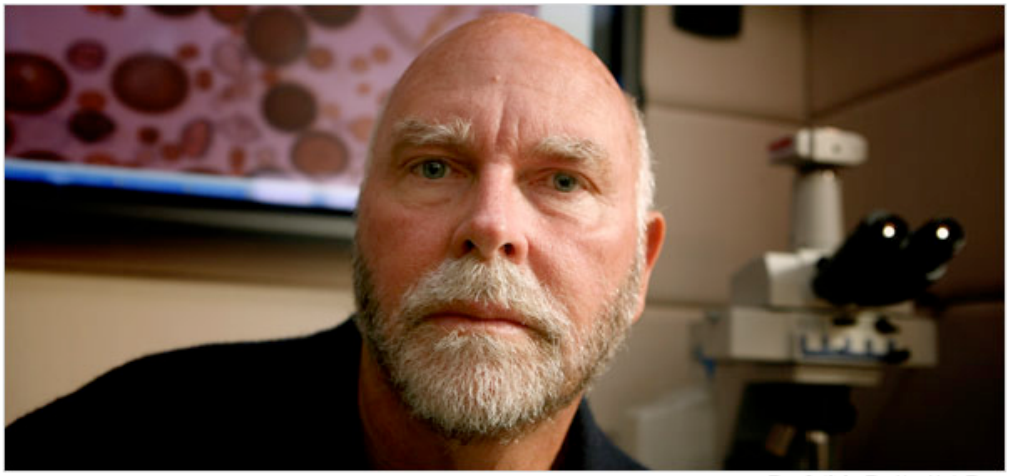


In 2005, International sequence databases exceed 100 gigabases


NATIONAL BESTSELLER
"A fascinating tour of the human genome. . . . If you want to catch a glimpse of the biotech century that is now dawning . . . Genome is an excellent place to start." —Wall Street Journal
GENOME
THE AUTOBIOGRAPHY OF A SPECIES IN 23 CHAPTERS
MATT RIDLEY
AUTHOR OF THE AGILE GENE AND FRANCIS CRICK
P.S.

Personalized Medicine?

In the Genome Race, the Sequel Is Personal

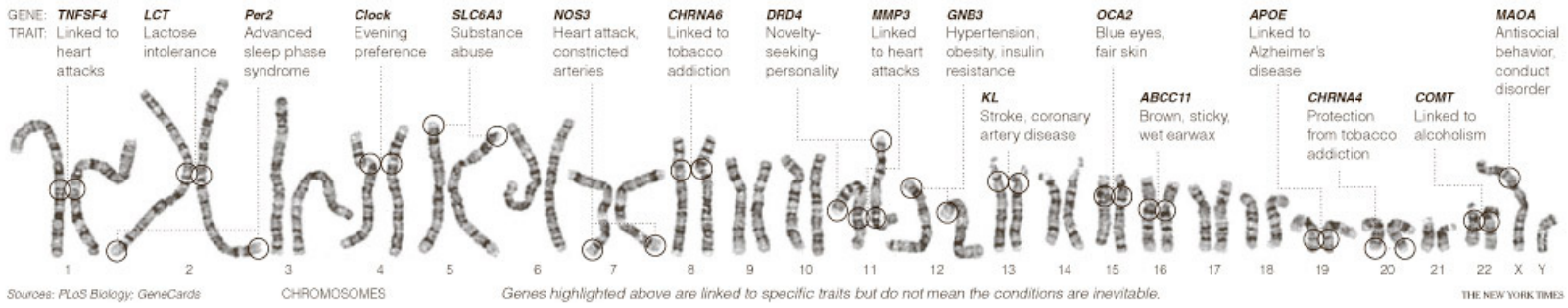


Thor Swift for The New York Times

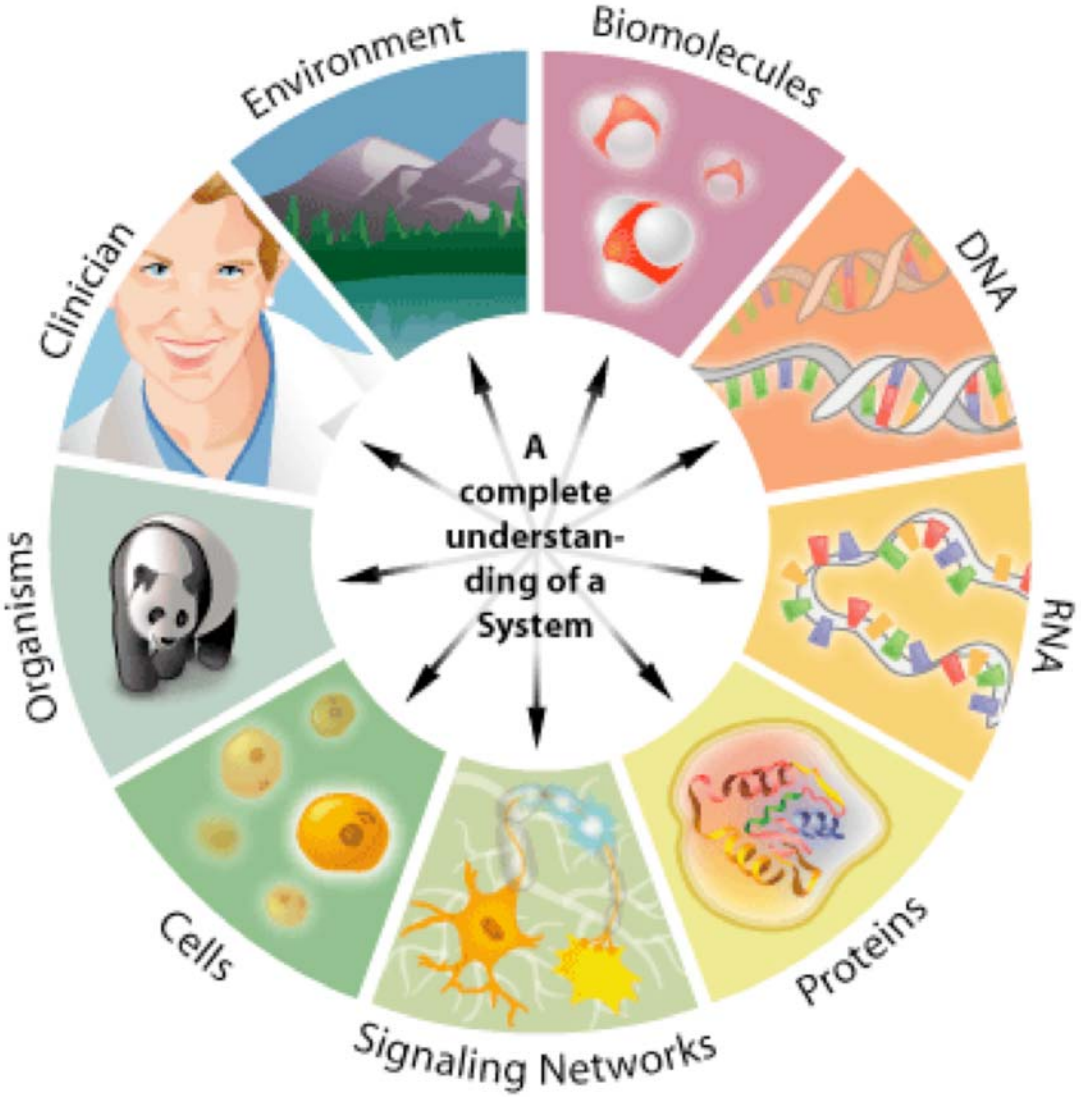
A team led by J. Craig Venter, above, has finished the first mapping of a full, or diploid, genome, made up of DNA inherited from both parents. The genome is Dr. Venter's own.

The New York Times — September 3, 2007

DECODING HIMSELF A team led by J. Craig Venter, above, has finished the first mapping of a full, or diploid, genome, made up of DNA inherited from both parents. The genome is Dr. Venter's own.

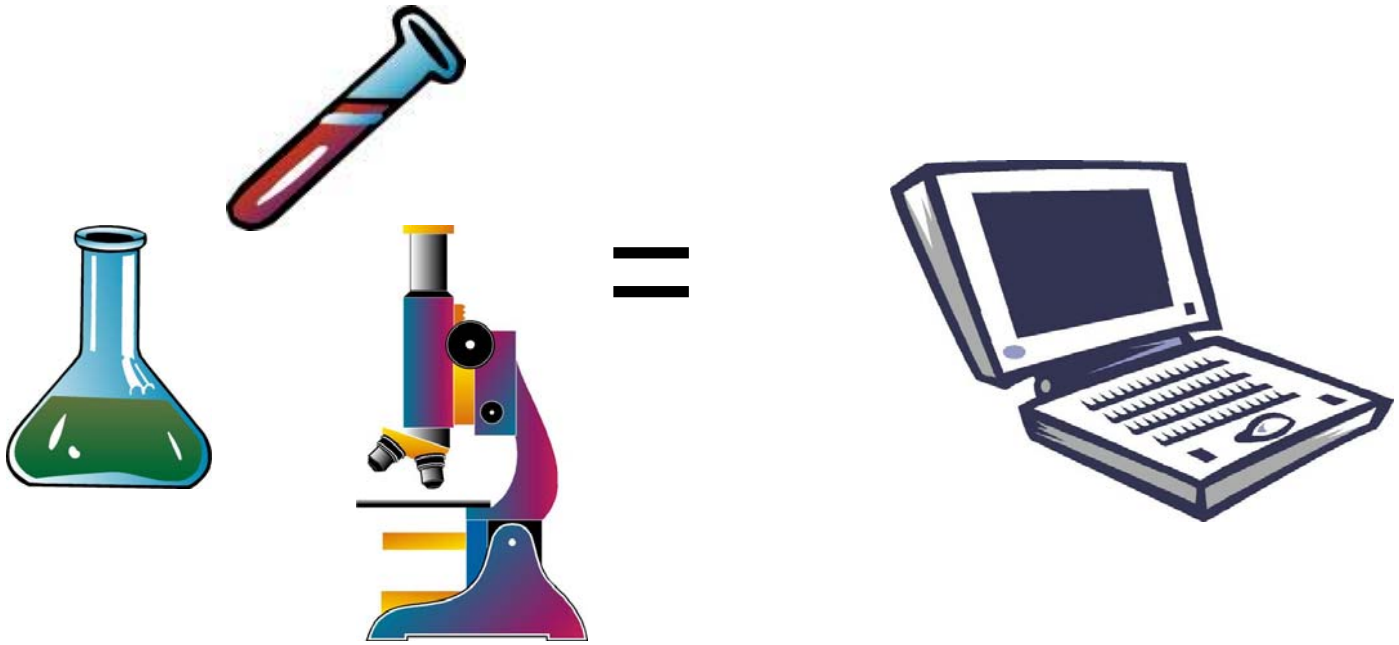


Sources: PLoS Biology; GeneCards — CHROMOSOMES — *Genes highlighted above are linked to specific traits but do not mean the conditions are inevitable.* — THE NEW YORK TIMES

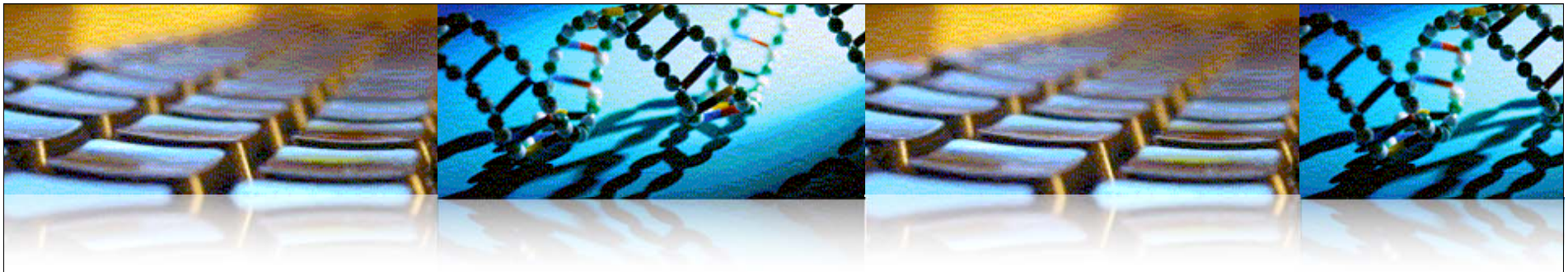

Environment
Biomolecules
Clinician
DNA
A complete understan-ding of a System
Organisms
RNA
Cells
Proteins
Signaling Networks

Linking Biological Information

- Nucleic acid & Protein Sequence data
- Sequence similarity implies homology or similar function
- Literature / Books
- Papers on related topics
- Macromolecular 3D Structure Data
- Similar protein fold implies homology
- Regulatory Pathways, Expression Data
- Two genes / enzymes in the same pathway
- Taxonomy
- Linkage of organisms by evolution



What is Bioinformatics?



Bioinformatics is an fast-paced interdisciplinary research field that involves the integration of computers, software tools, and databases in an effort to address biological questions.

Genomics refers to the analysis of all of the genes and transcripts included within the genome. Proteomics, on the other hand, refers to the analysis of the complete set of proteins or proteome.

Bioinformatics Questions

- What is encoded by the genome?
  - Genes, regulatory, and functional regions
- How is genome information expressed?
  - Function of genes and gene products (proteins)
  - Structure of proteins
- How can we interpret the information encoded in the genome?
  - Linking knowledge to the biological entities.
  - Systems biology approach
  - drugs, metabolites, ...
- How does the genome interact with its environment?

Summary

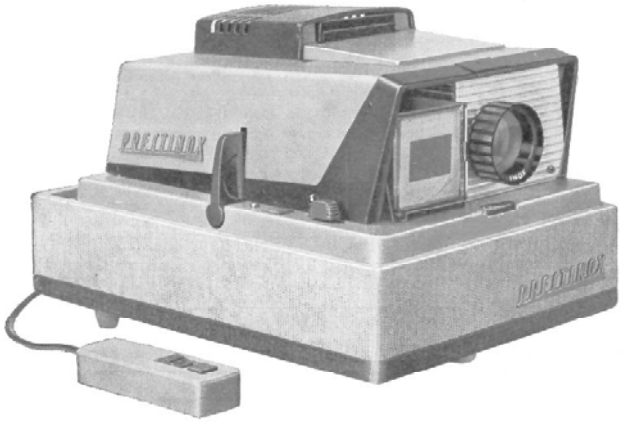
- An article called, "What is Bioinformatics?" is available from the Science Creative Quarterly.

http://www.scq.ubc.ca/what-is-bioinformatics/

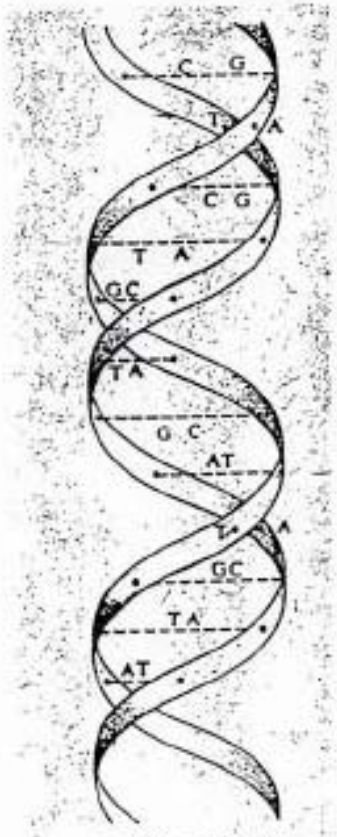
DNA Sequencing

Generating Data & Emerging Technologies

Technology

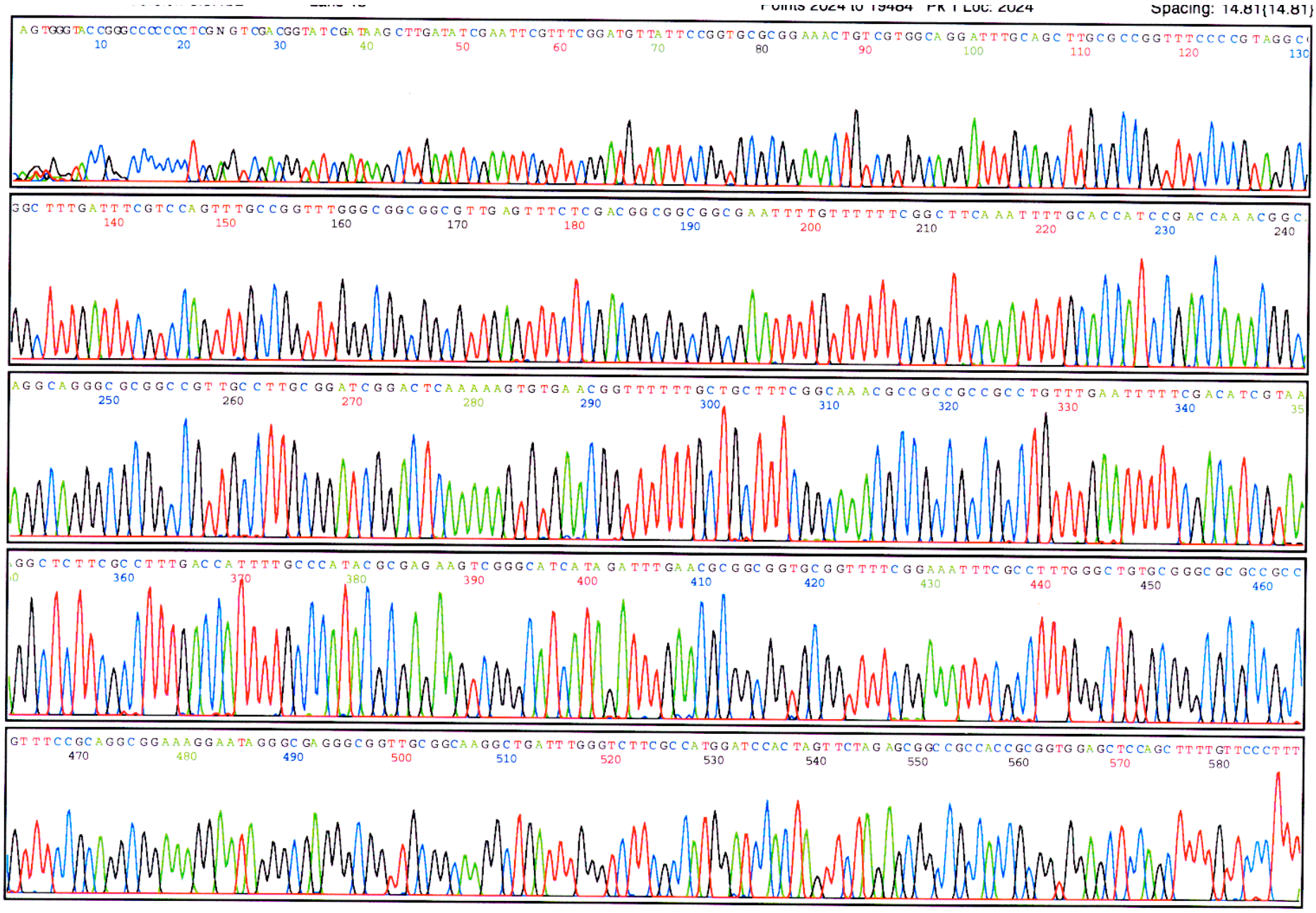


ggagcctcgg gaggtggtgg agtgacctgg ccccagtgct gcgtccttat cagccgagcc
ggtcccagct cttgctcctg cctgtttgcc tggaaatggc cacgcttctc cttctccttg
gggtgctggt ggtaagccca gacgctctgg ggagcacaac agcagtgcag acacccacct
ccggagagcc tttggtctct actagcgagc ccctgagctc aaagatgtac accacttcaa
taacaagtga ccctaaggcc gacagcactg gggaccagac ctcagcccta cctccctcaa
cttccatcaa tgagggatcc cctctttgga cttccattgg tgccagcact ggttcccctt
tacctgagcc aacaacctac caggaagttt ccatcaagat gtcatcagtg ccccaggaaa
cccctcatgc aaccagtcat cctgctgttc ccataacagc aaactctcta ggatcccaca
ccgtgacagg tggaaccata acaacgaact ctccagaaac ctccagtagg accagtggag
cccctgttac cacggcagct agctctctgg agacctccag aggcacctct ggacccctc
ttaccatggc aactgtctct ctggagactt ccaaaggcac ctctggaccc cctgttacca
tggcaactga ctctctggag acctccactg ggaccactgg acccctgtt accatgacaa
ctggctctct ggagccctcc agcggggcca gtggacccca ggtctctagc gtaaaactat
ctacaatgat gtctccaacg acctccacca acgcaagcac tgtgcccttc cggaacccag
atgagaactc acgaggcatg ctgccagtgg ctgtgcttgt ggccctgctg gcggtcatag
tcctcgtggc tctgctcctg ctgtggcgcc ggcggcagaa gcggcggact ggggccctcg
tgctgagcag aggcggcaag cgtaacgggg tggtggacgc ctgggctggg ccagcccagg
tccctgagga gggggccgtg acagtgaccg tgggagggtc cgggggcgac aagggctctg
ggttccccga tggggagggg tctagccgtc ggcccacgct caccactttc tttggcagac
ctggctctct ggagccctcc agcggggcca gtggacccca ggtctctagc gtaaaactat
ctacaatgat gtctccaacg acctccacca acgcaagcac tgtgcccttc cggaacccag
atgagaactc acgaggcatg ctgccagtgg ctgtgcttgt ggccctgctg gcggtcatag



a DNA molecule

= SNOT


Points 2024 to 19484 PK 1 Loc: 2024
Spacing: 14.81(14.81)

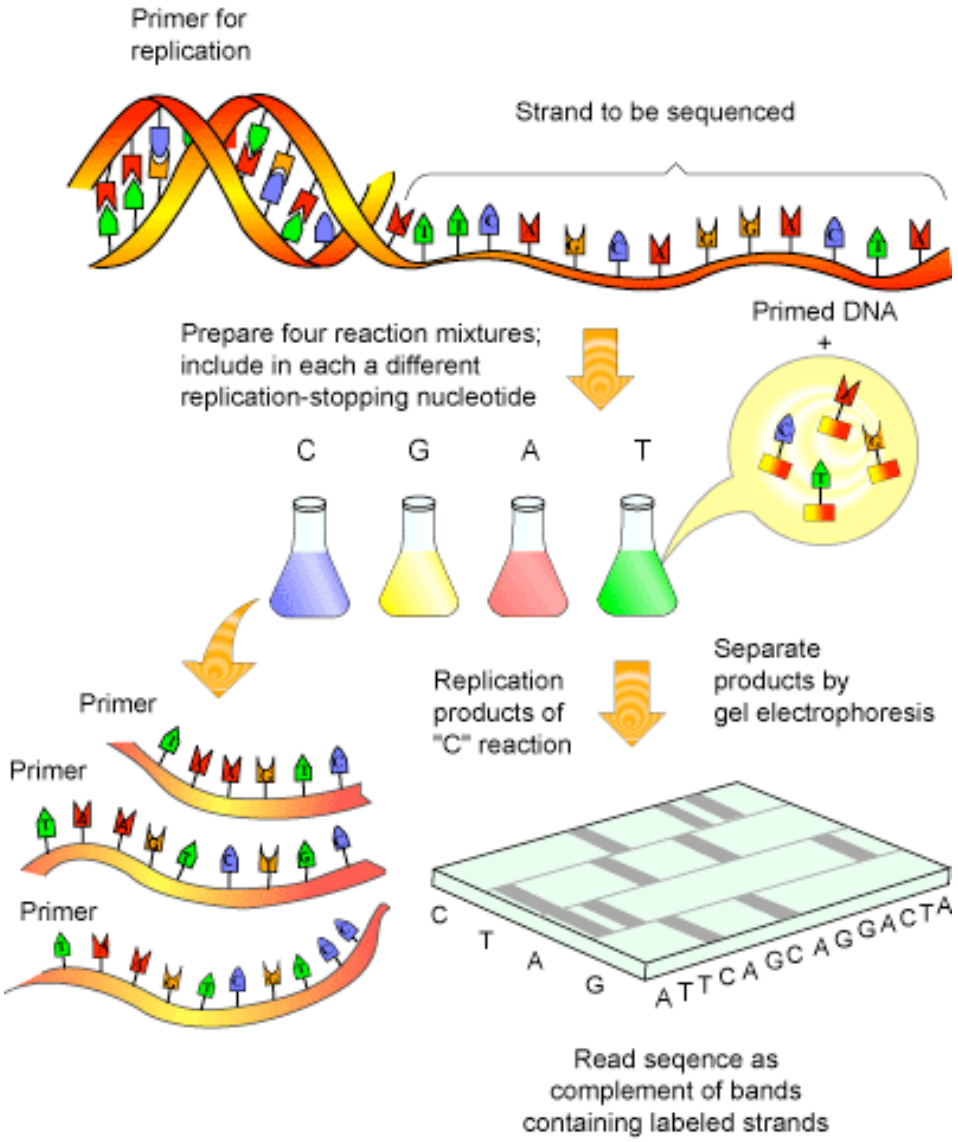


Figure 1. *The Sanger sequencing reaction. Single stranded DNA is amplified in the presence of fluorescently labelled ddNTPs that serve to terminate the reaction and label all the fragments of DNA produced. The fragments of DNA are then separated via polyacrylamide gel electrophoresis and the sequence read using a laser beam and computer.*

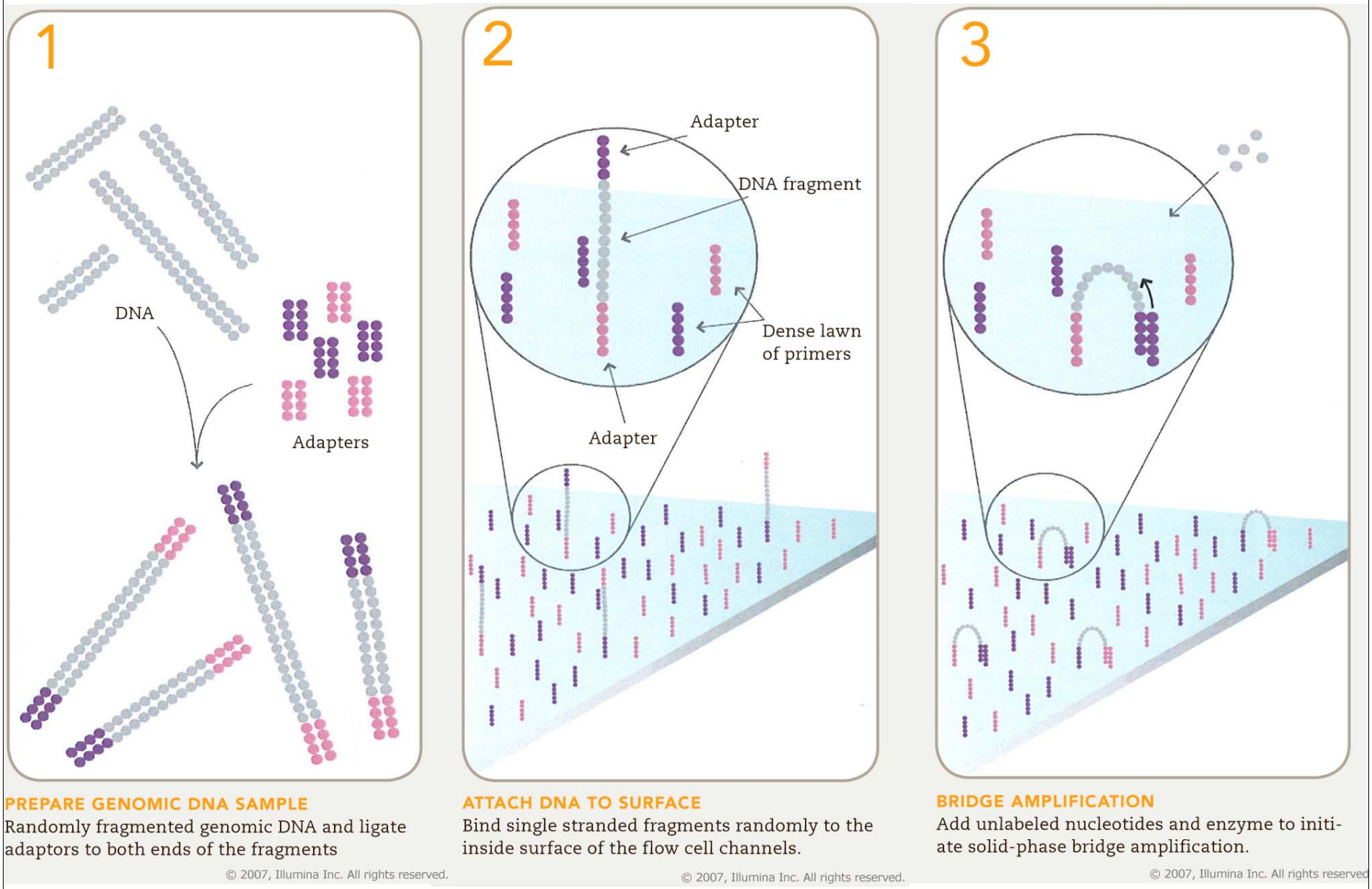
source: http://www.scq.ubc.ca/genome-projects-uncovering-the-blueprints-of-biology/

Every few years, a new technology comes along that dramatically changes how fundamental questions in biology are addressed. The impact of the technology is not always appreciated at first ...

- Stanley Fields

Solexa Technology

- DNA sequencing by synthesis
- approach built around very large number of short sequence reads
- key points:
  - ✓ solid phase amplification = no cloning necessary
  - ✓ reversible chemistry
  - ✓ data generated by imaging
  - ✓ read lengths 30-50 bp
  - ✓ < 1% cost, ultra high-throughput

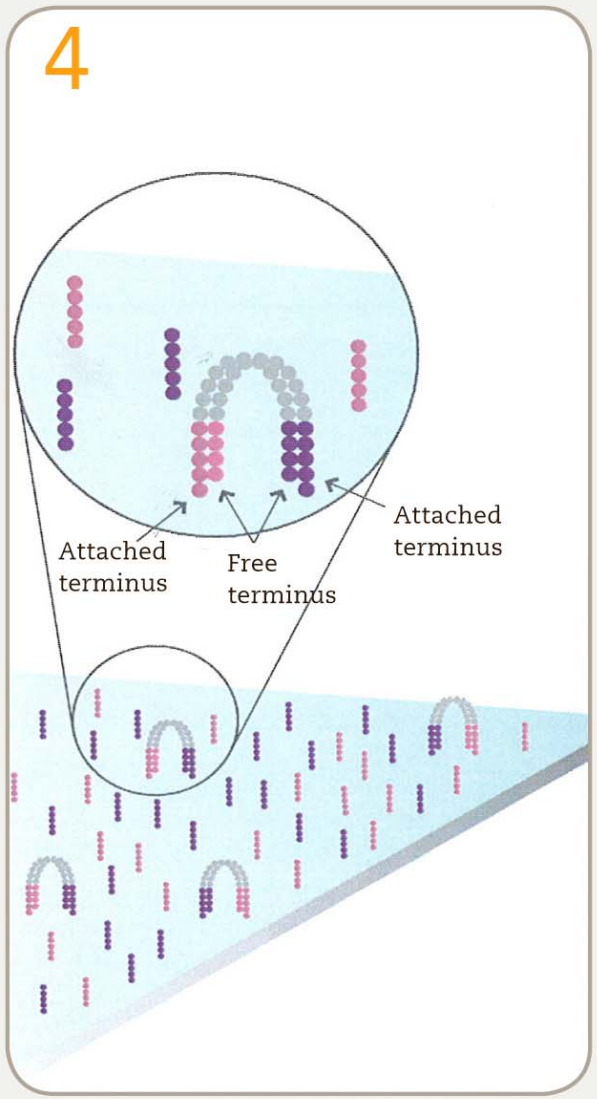


PREPARE GENOMIC DNA SAMPLE
Randomly fragmented genomic DNA and ligate adaptors to both ends of the fragments

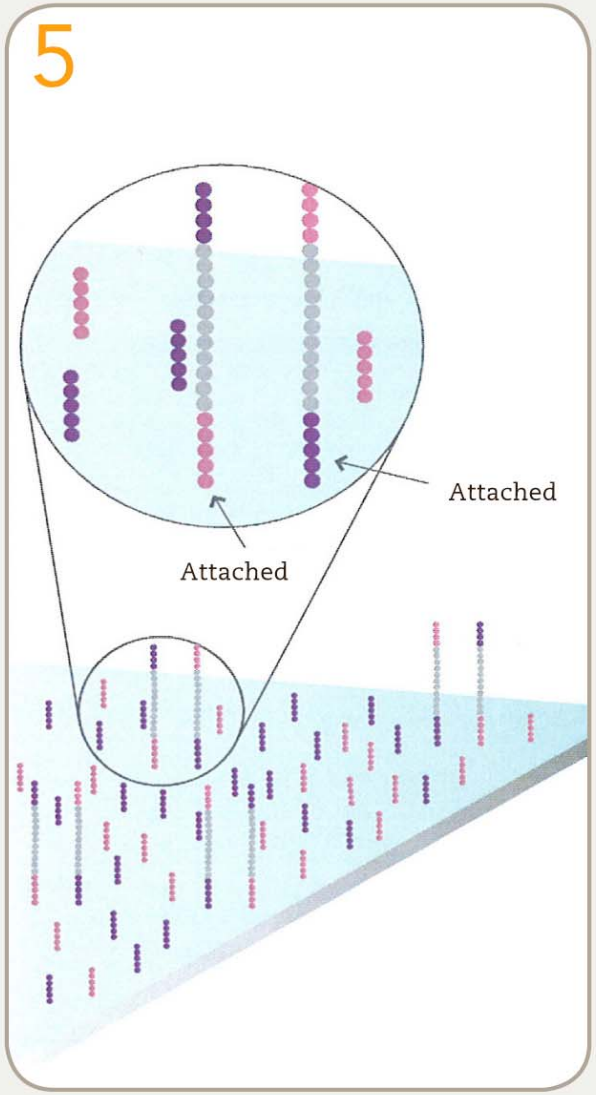
ATTACH DNA TO SURFACE
Bind single stranded fragments randomly to the inside surface of the flow cell channels.

BRIDGE AMPLIFICATION
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

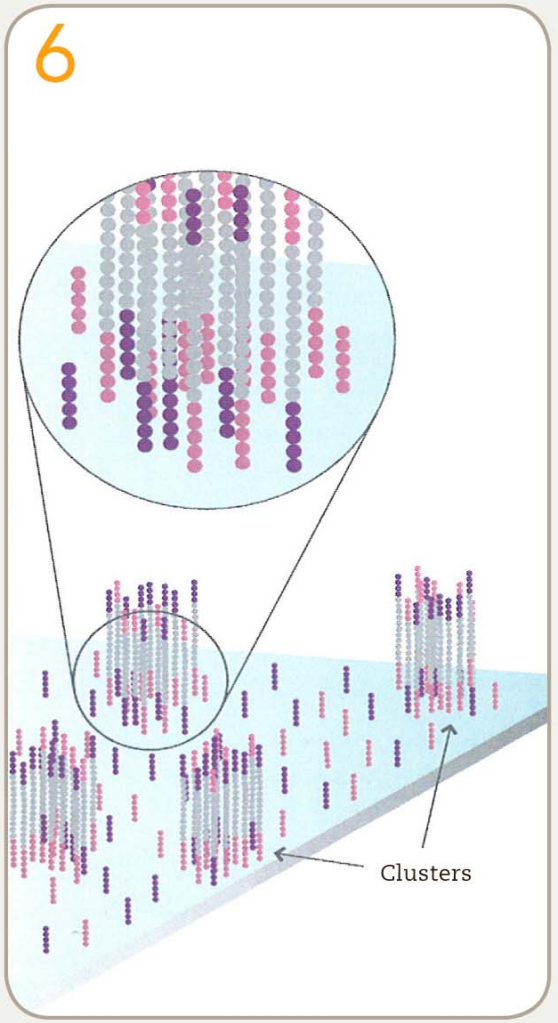
source: http://www.illumina.com/



FRAGMENTS BECOME DOUBLE STRANDED



DENATURE THE DOUBLE STRANDED MOLECULES

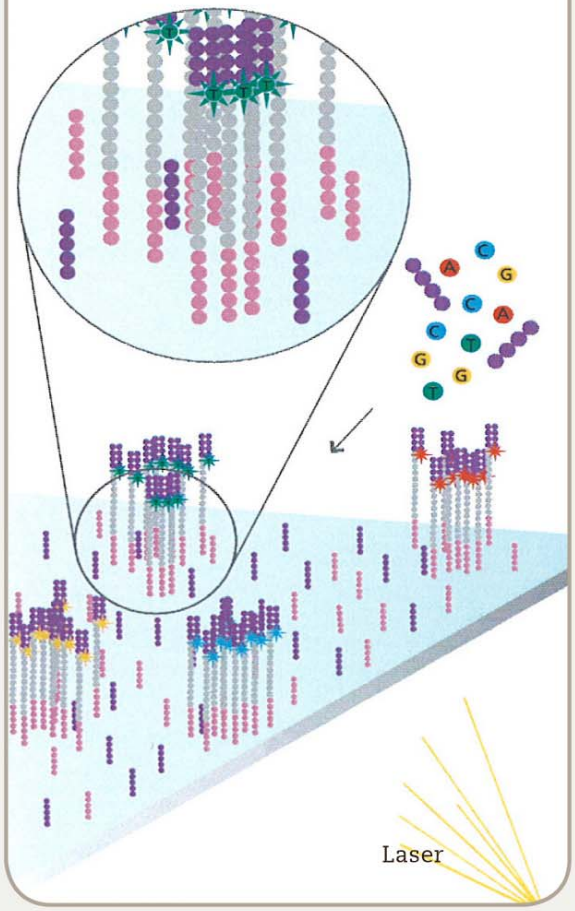


COMPLETION OF AMPLIFICATION

On completion, several million dense clusters of double stranded DNA are generated in each channel of the flow cell.

source: http://www.illumina.com/

7



FIRST CHEMISTRY CYCLE: DETERMINE FIRST BASE

To initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8

After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

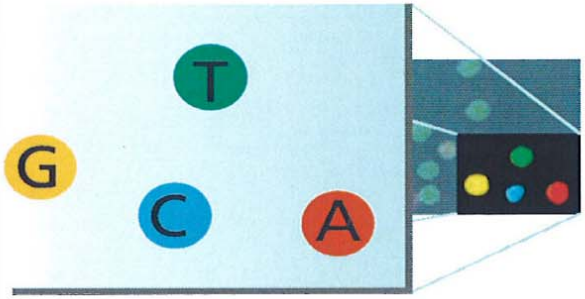
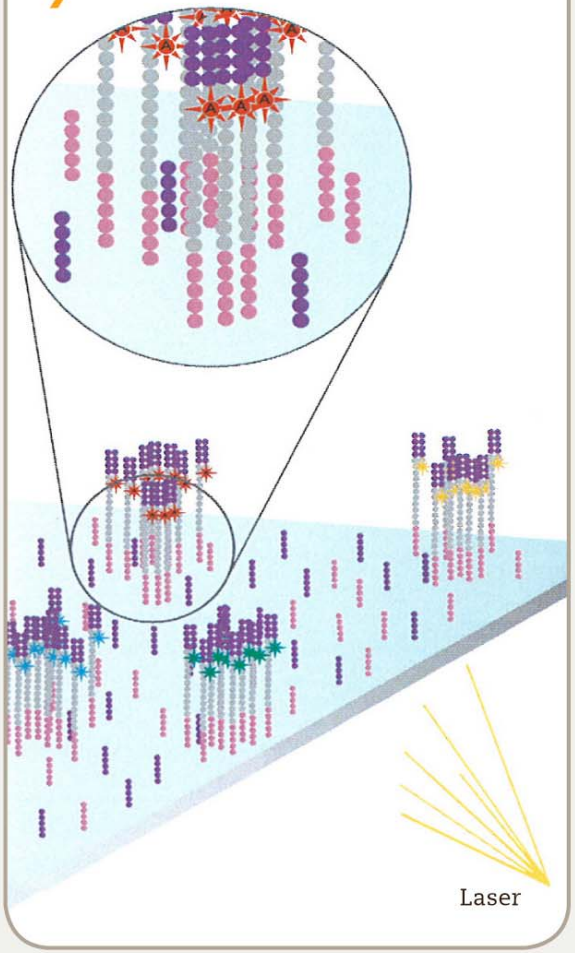


IMAGE OF FIRST CHEMISTRY CYCLE

9



SECOND CHEMISTRY CYCLE: DETERMINE SECOND BASE

To initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

source: http://www.illumina.com/

10

After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

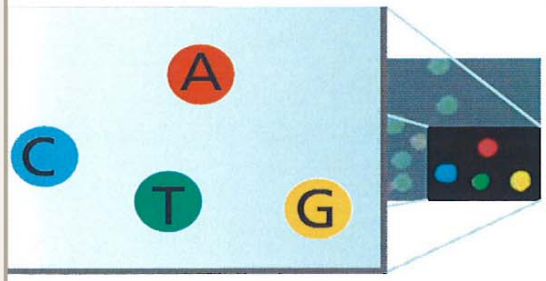
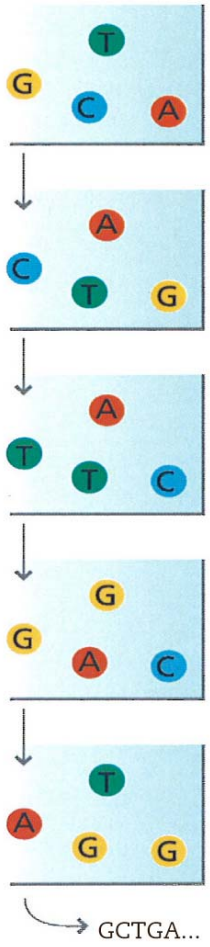


IMAGE OF SECOND CHEMISTRY CYCLE IS CAPTURED BY THE INSTRUMENT.

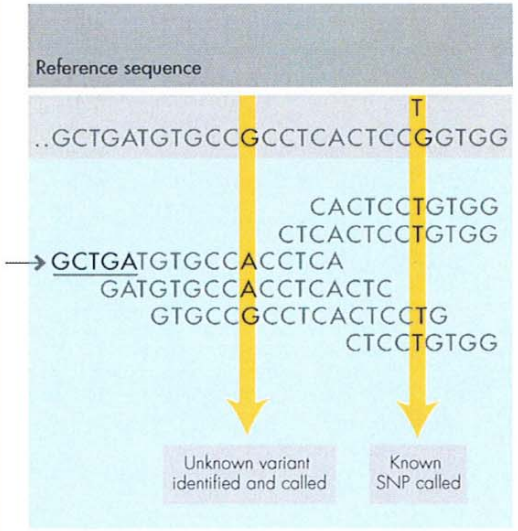
11

Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.



SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES

12

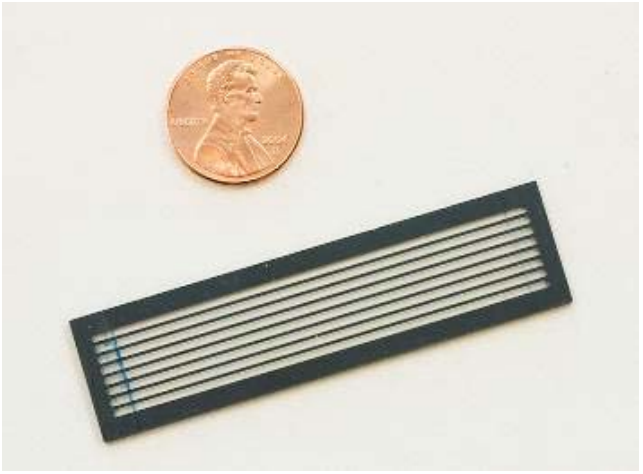


ALIGN THE NEW DATA TO A REFERENCE AND IDENTIFY SEQUENCE DIFFERENCES.

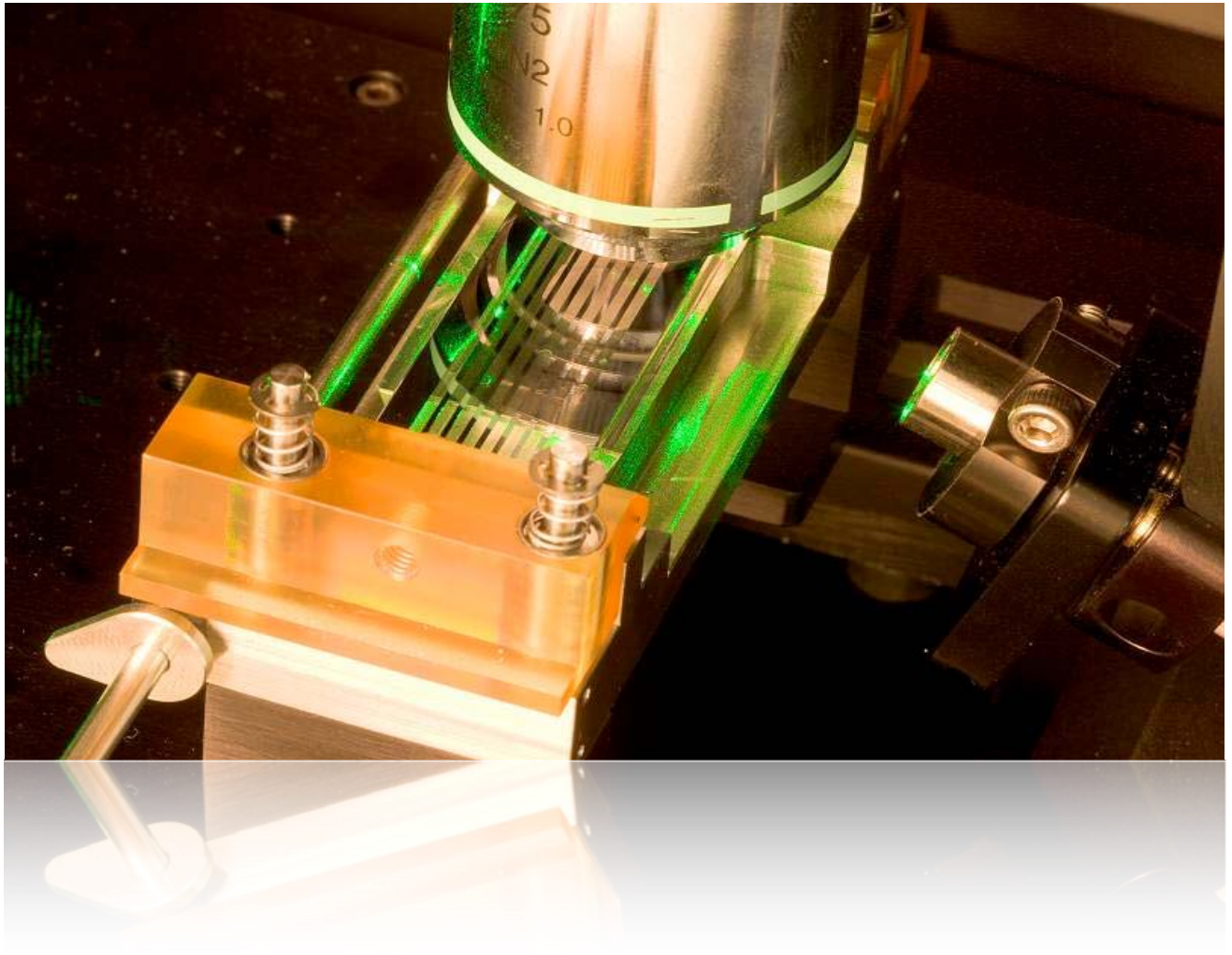
source: http://www.illumina.com/

Illumina/Solexa instrument

- Laser-based TIRF* Optics
- 4-colour Detection
- CCD camera
- 8-channel flow cell
- 1 Gb / run at launch

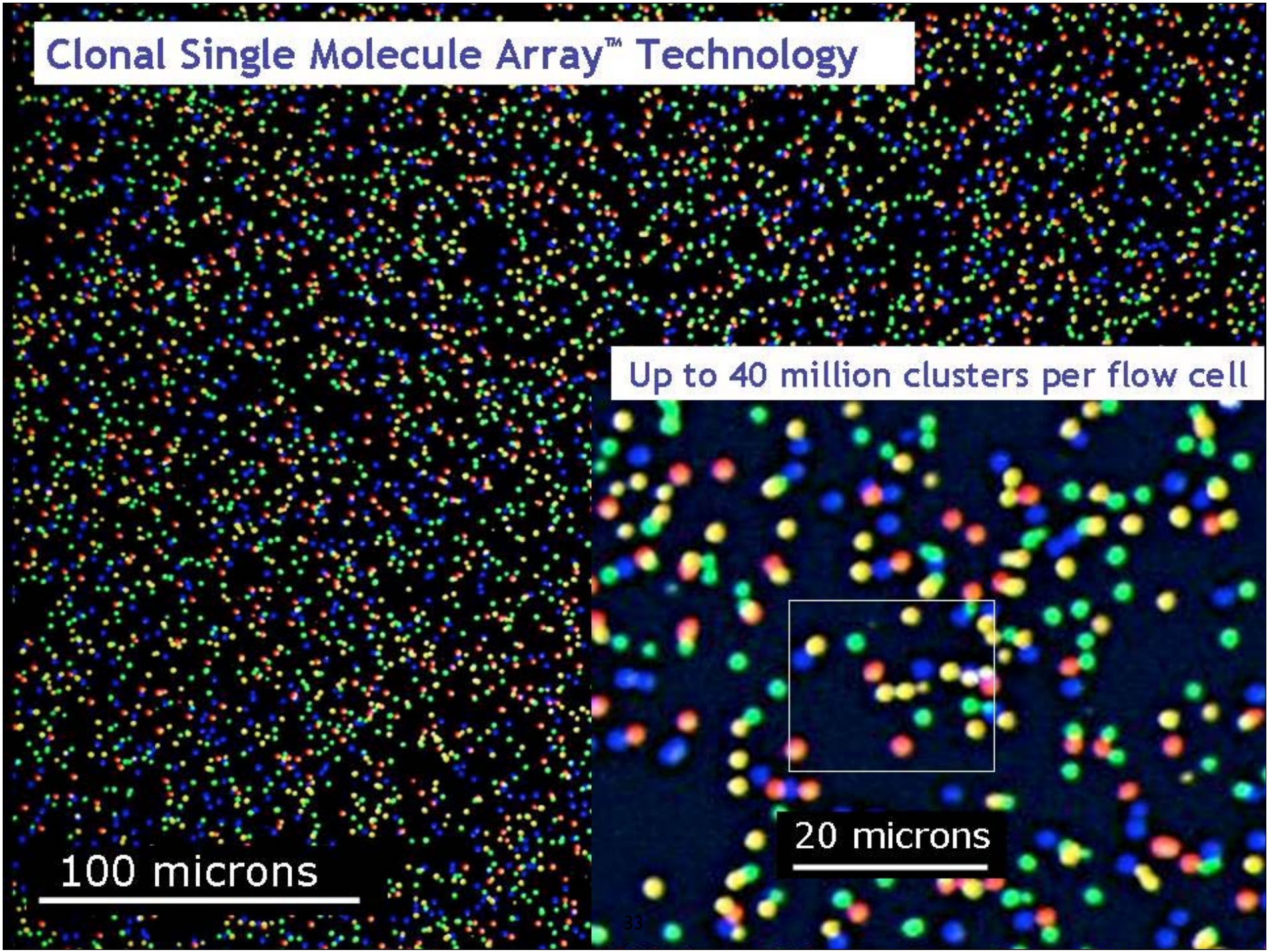


source: Dr. Steven Jones, GSC

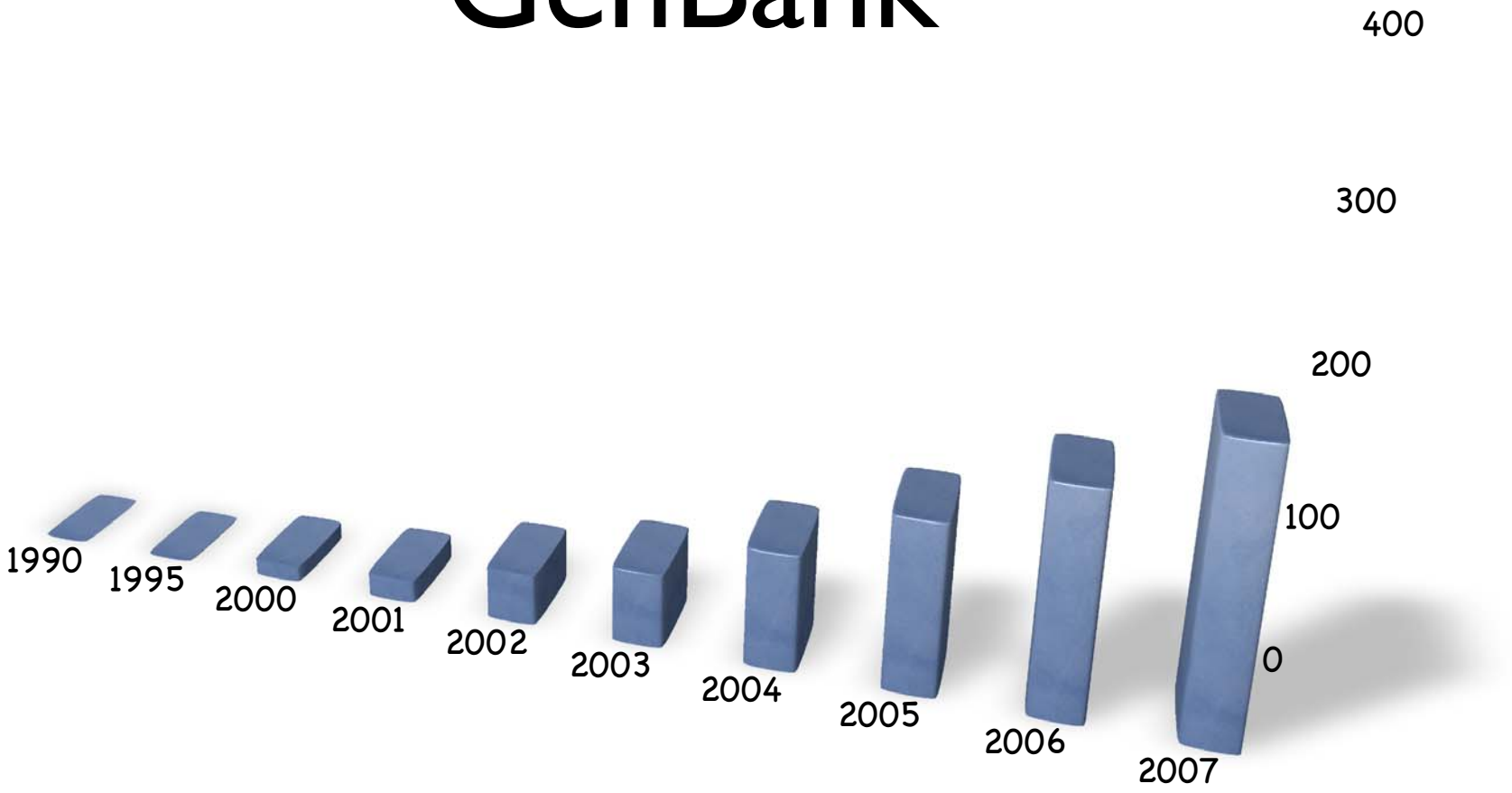


Data acquisition

source: Dr. Steven Jones, GSC


Clonal Single Molecule Array™ Technology
Up to 40 million clusters per flow cell
100 microns
20 microns

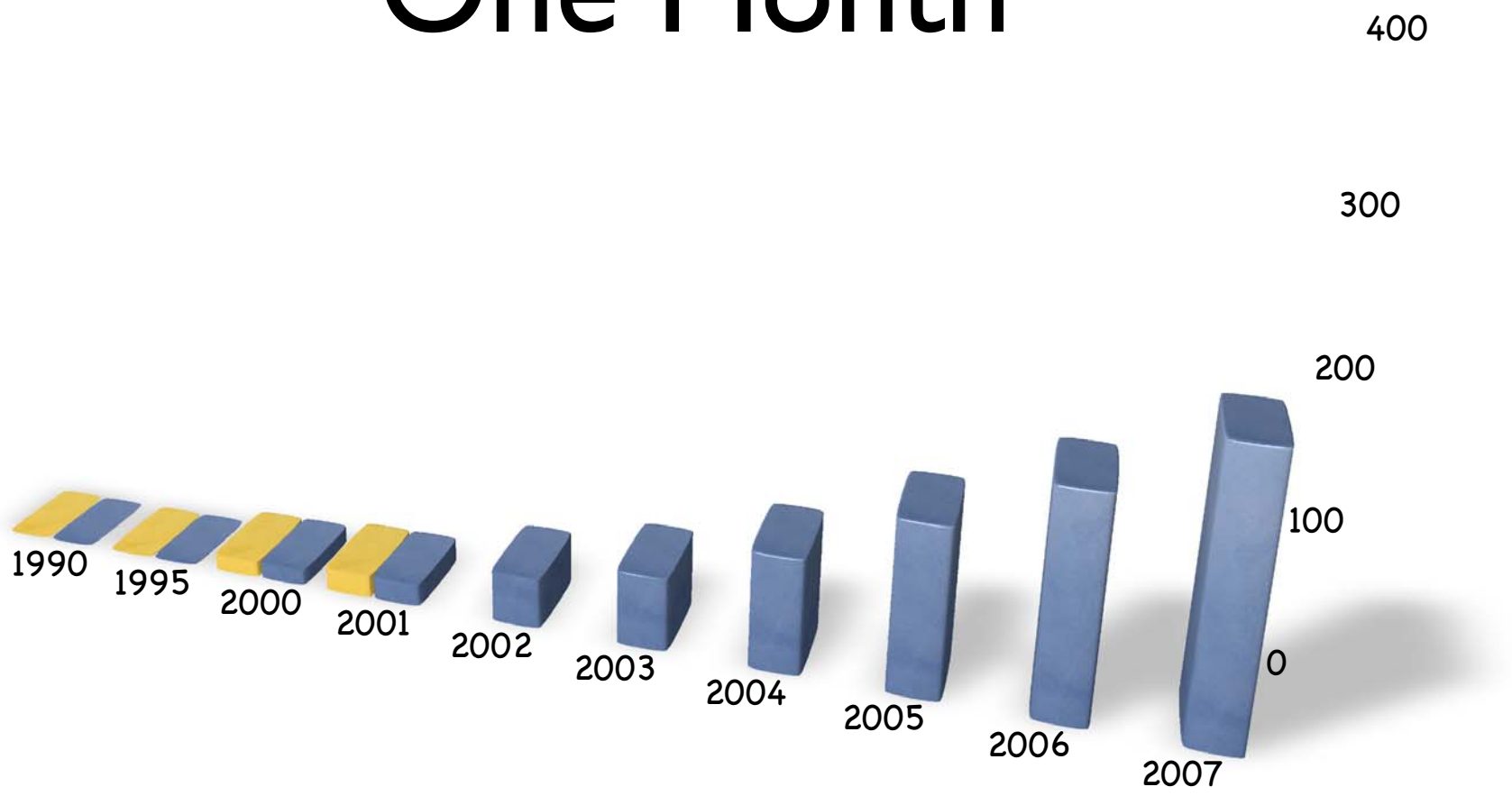
GenBank


400
300
200
100
0
1990
1995
2000
2001
2002
2003
2004
2005
2006
2007

Two Days

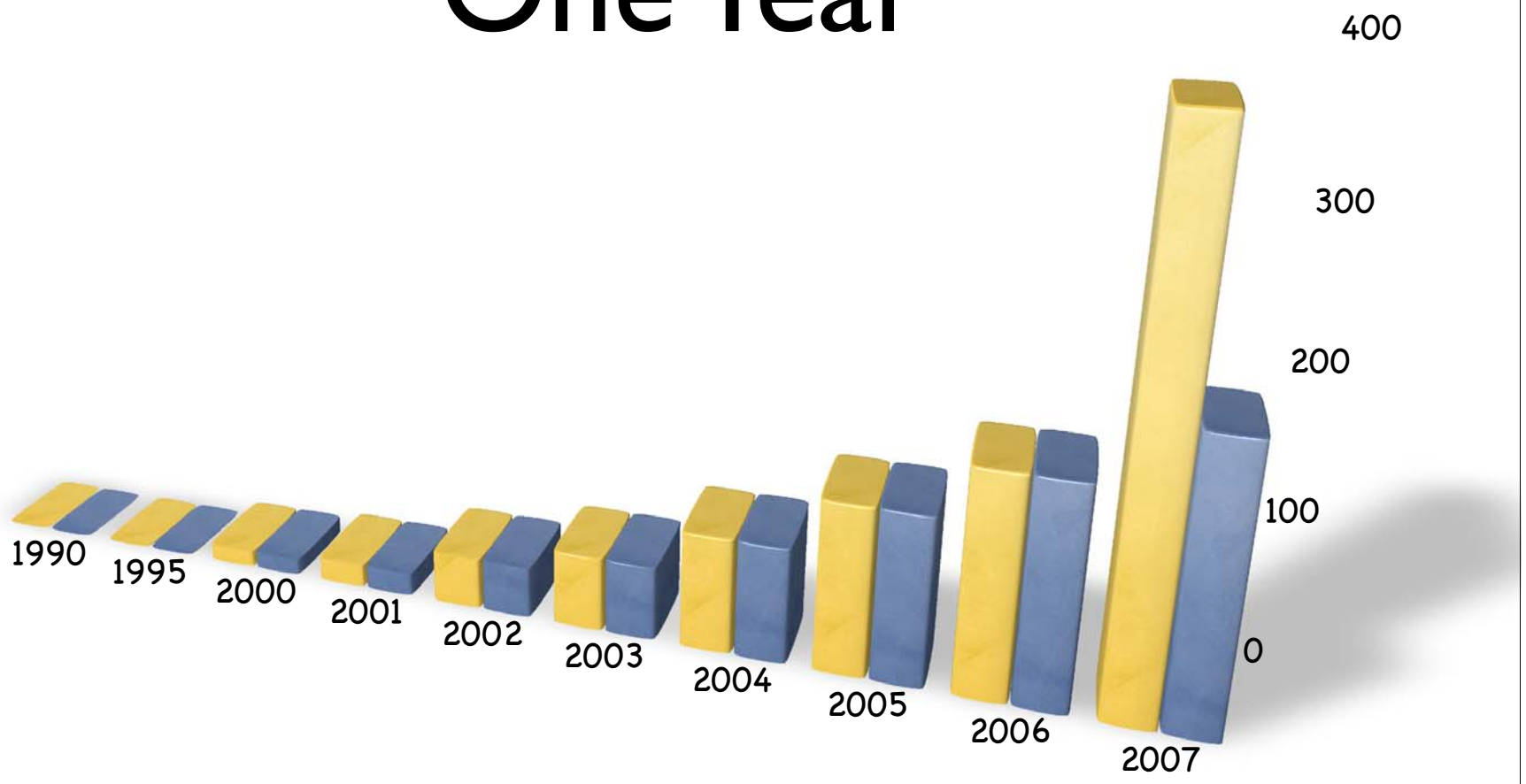
One machine, running for two days, can generate ~1 Gb data.

One Month

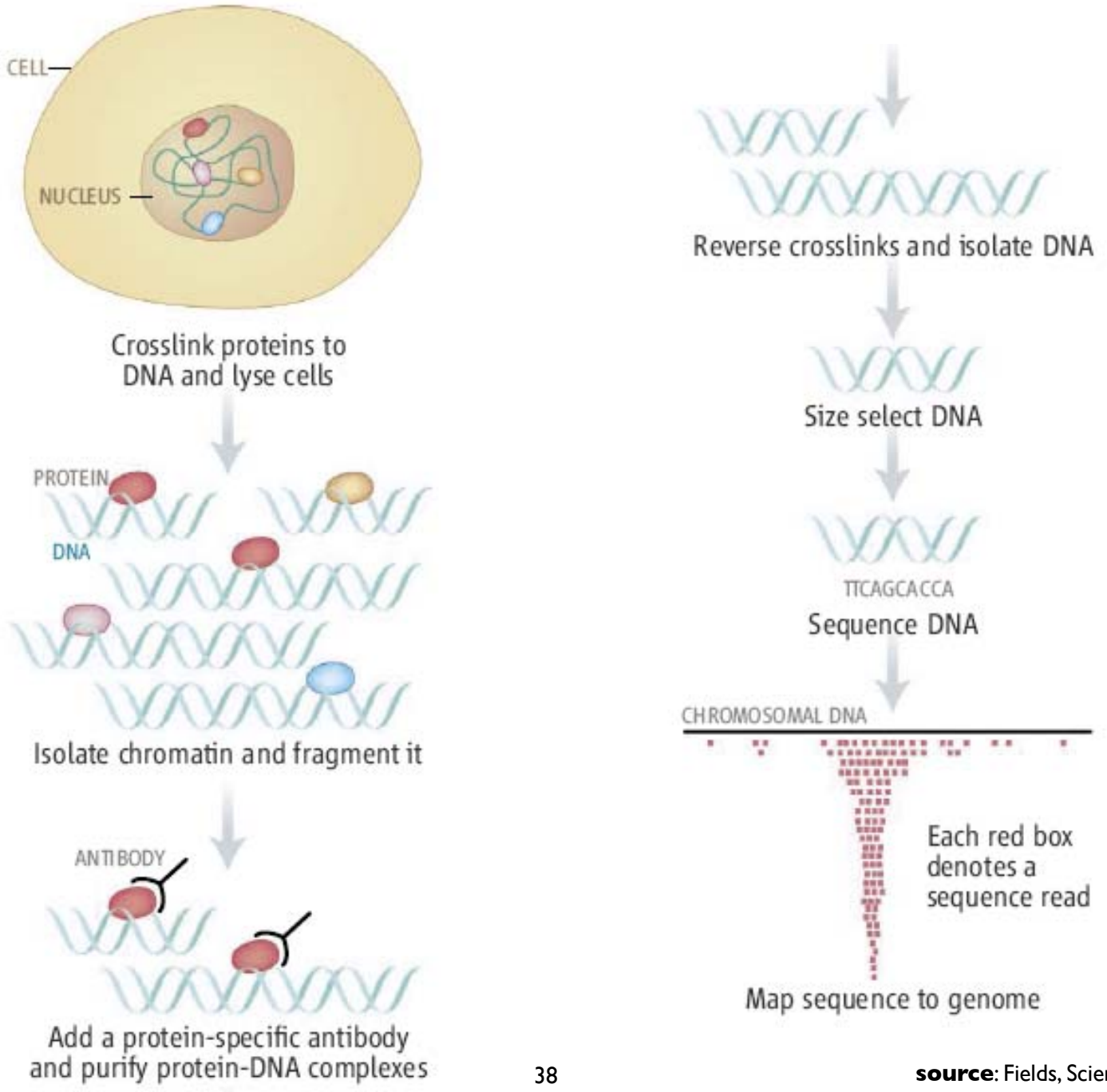


One machine, running for one month, could generate ~15 Gb data.

One Year



Two machines, running for one year, could generate ~365 Gb data.



source: Fields, Science 2007, p1441

Frontiers in Bioinformatics

- ultra high-throughput sequencing
  - DNA binding site identification
  - genome re-sequencing; SNPs, expression
    - ✓ low cost
    - ✓ whole genome
    - ✓ any genome

Credits & References

- Technology Spotlight on DNA Sequencing with Solexa Technology:

http://www.illumina.com/downloads/SS_DNAsequencing.pdf

- Dr. Steven Jones, GSC

several slides/images used with permission

- Stanley Fields, "Site-Seeing by Sequencing", Science, 8 June 2007

Sequence Databases

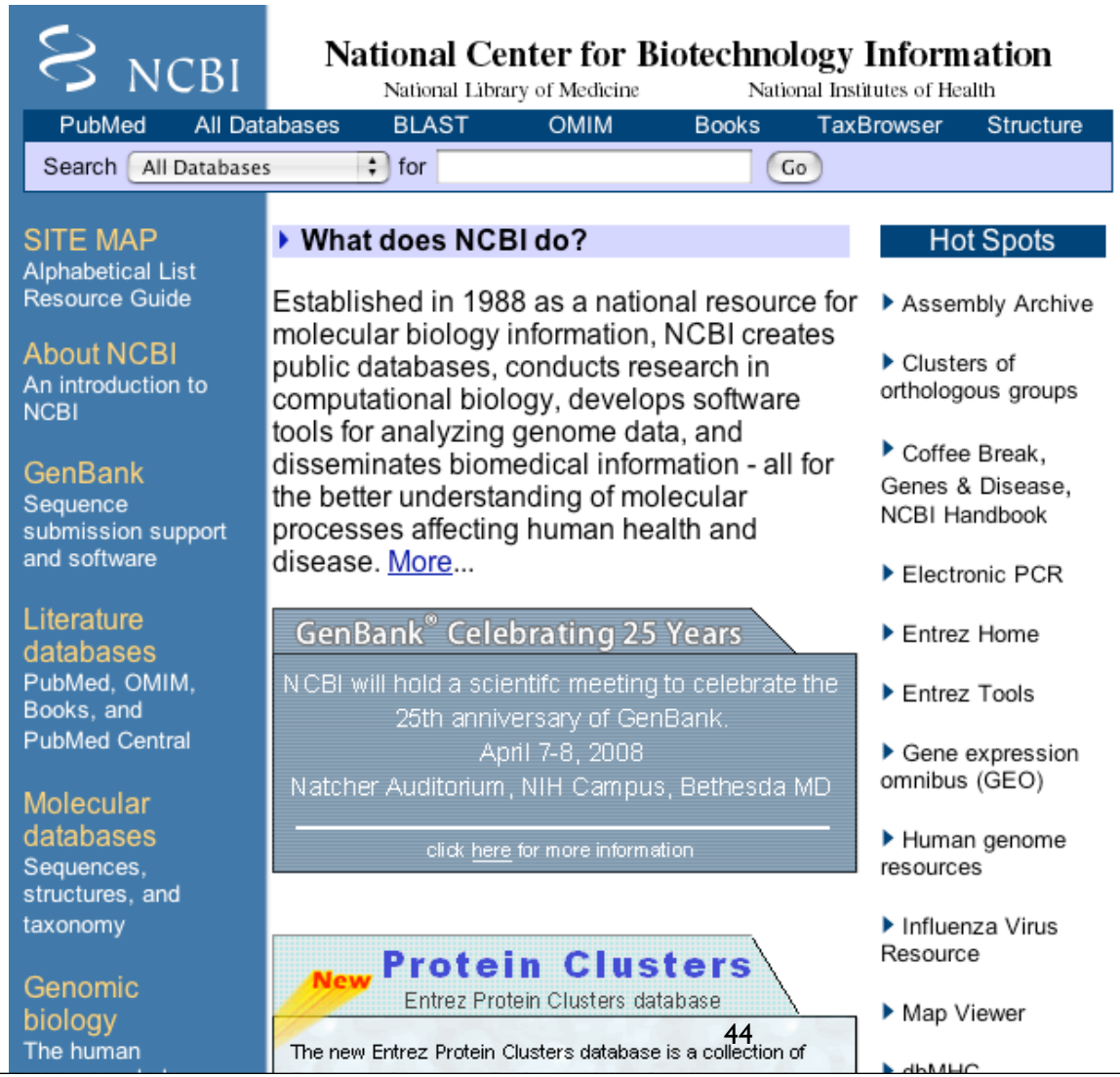
Public Resources at the NCBI

The National Center for Biotechnology Information

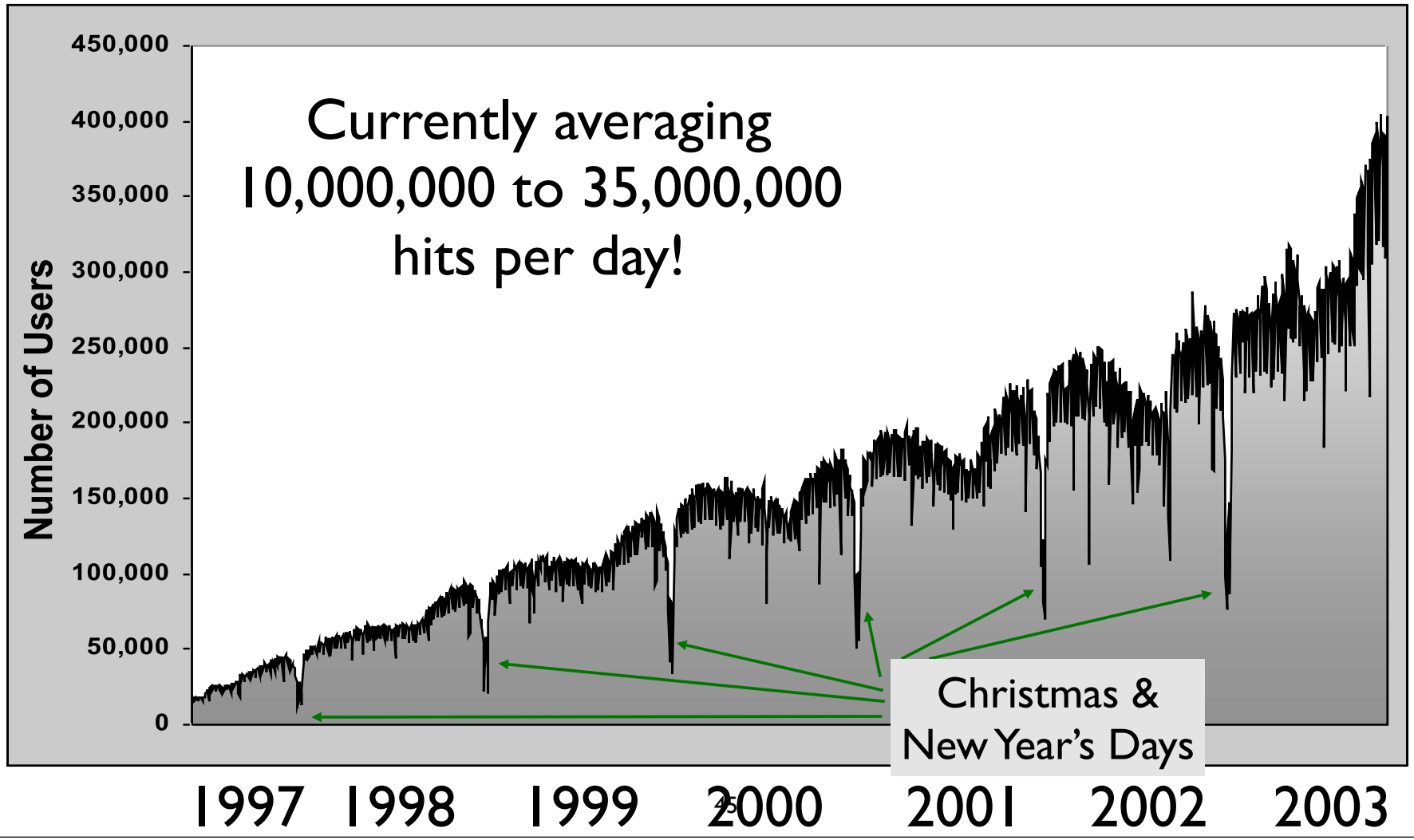
NCBI

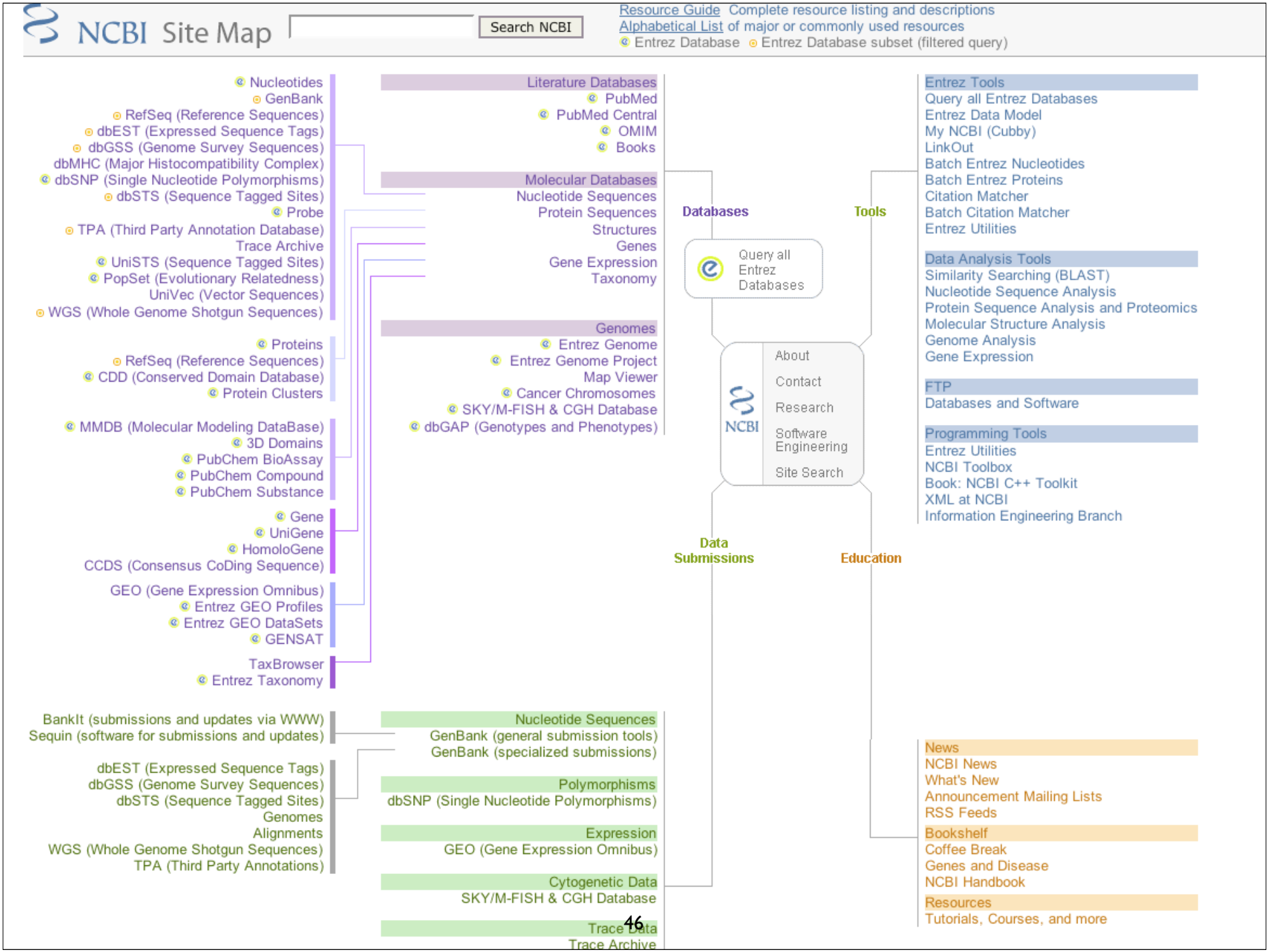
- **Created in 1988 as a part of the National Library of Medicine at NIH**
- Establish public databases
- Research in computational biology
- Develop software tools for sequence analysis
- Disseminate biomedical information

www.ncbi.nlm.nih.gov

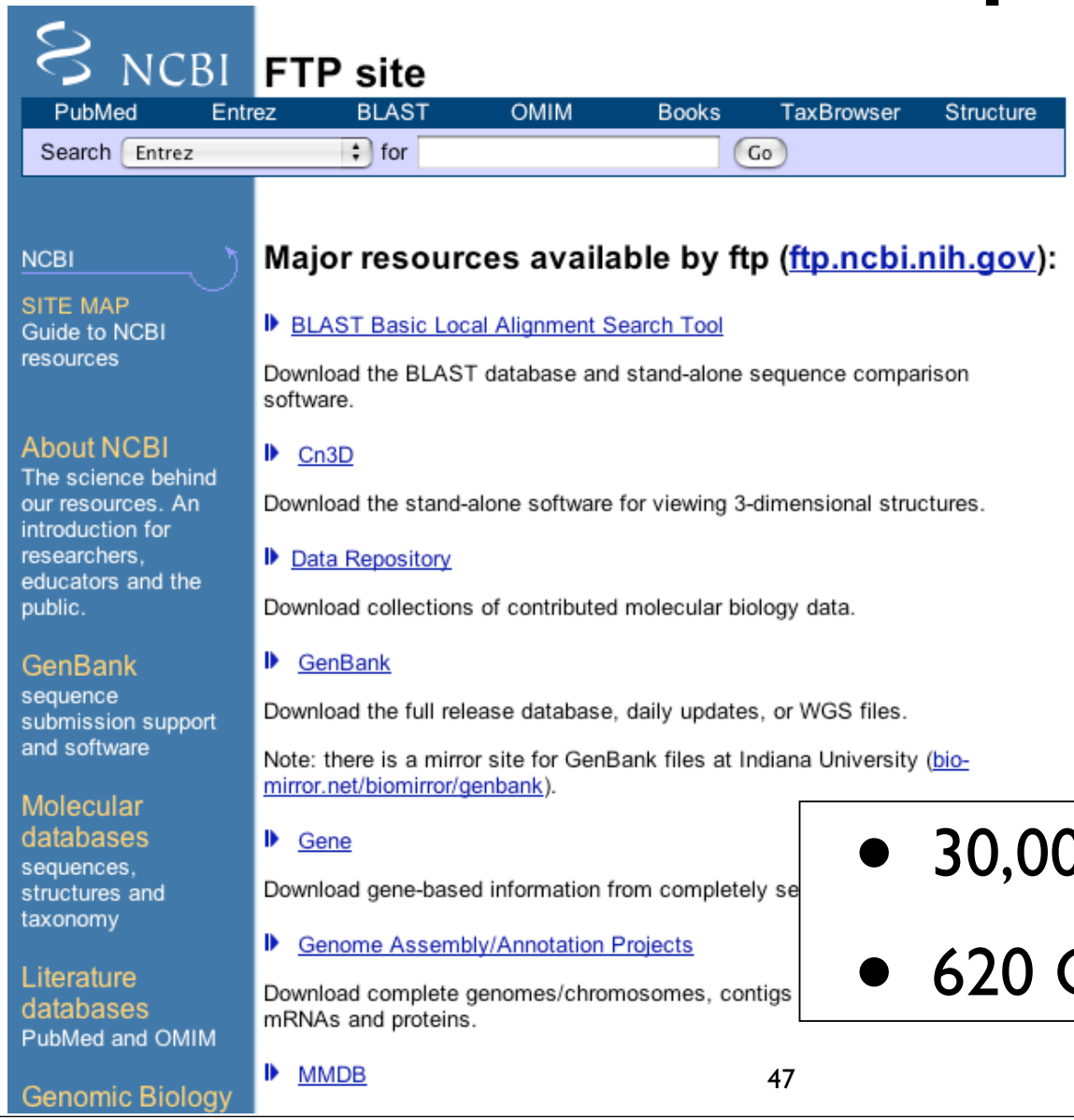


Number of Users and Hits Per Day




NCBI Site Map
Search NCBI
Resource Guide Complete resource listing and descriptions
Alphabetical List of major or commonly used resources
Entrez Database
Entrez Database subset (filtered query)
Nucleotides
GenBank
RefSeq (Reference Sequences)
dbEST (Expressed Sequence Tags)
dbGSS (Genome Survey Sequences)
dbMHC (Major Histocompatibility Complex)
dbSNP (Single Nucleotide Polymorphisms)
dbSTS (Sequence Tagged Sites)
Probe
TPA (Third Party Annotation Database)
Trace Archive
UniSTS (Sequence Tagged Sites)
PopSet (Evolutionary Relatedness)
UniVec (Vector Sequences)
WGS (Whole Genome Shotgun Sequences)
Proteins
RefSeq (Reference Sequences)
CDD (Conserved Domain Database)
Protein Clusters
MMDB (Molecular Modeling DataBase)
3D Domains
PubChem BioAssay
PubChem Compound
PubChem Substance
Gene
UniGene
HomoloGene
CCDS (Consensus CoDing Sequence)
GEO (Gene Expression Omnibus)
Entrez GEO Profiles
Entrez GEO DataSets
GENSAT
TaxBrowser
Entrez Taxonomy
Literature Databases
PubMed
PubMed Central
OMIM
Books
Molecular Databases
Nucleotide Sequences
Protein Sequences
Structures
Genes
Gene Expression
Taxonomy
Genomes
Entrez Genome
Entrez Genome Project
Map Viewer
Cancer Chromosomes
SKY/M-FISH & CGH Database
dbGAP (Genotypes and Phenotypes)
Databases
Query all Entrez Databases
Tools
Entrez Tools
Query all Entrez Databases
Entrez Data Model
My NCBI (Cubby)
LinkOut
Batch Entrez Nucleotides
Batch Entrez Proteins
Citation Matcher
Batch Citation Matcher
Entrez Utilities
Data Analysis Tools
Similarity Searching (BLAST)
Nucleotide Sequence Analysis
Protein Sequence Analysis and Proteomics
Molecular Structure Analysis
Genome Analysis
Gene Expression
FTP
Databases and Software
Programming Tools
Entrez Utilities
NCBI Toolbox
Book: NCBI C++ Toolkit
XML at NCBI
Information Engineering Branch
NCBI
About
Contact
Research
Software Engineering
Site Search
Data Submissions
BankIt (submissions and updates via WWW)
Sequin (software for submissions and updates)
dbEST (Expressed Sequence Tags)
dbGSS (Genome Survey Sequences)
dbSTS (Sequence Tagged Sites)
Genomes
Alignments
WGS (Whole Genome Shotgun Sequences)
TPA (Third Party Annotations)
Nucleotide Sequences
GenBank (general submission tools)
GenBank (specialized submissions)
Polymorphisms
dbSNP (Single Nucleotide Polymorphisms)
Expression
GEO (Gene Expression Omnibus)
Cytogenetic Data
SKY/M-FISH & CGH Database
Trace Data
Trace Archive
Education
News
NCBI News
What's New
Announcement Mailing Lists
RSS Feeds
Bookshelf
Coffee Break
Genes and Disease
NCBI Handbook
Resources
Tutorials, Courses, and more

The NCBI ftp site



- 30,000 files per day
- 620 Gigabytes per day

NCBI Databases & Services

- GenBank largest sequence database
- Free public access to biomedical literature
  - PubMed free Medline
  - PubMed Central full text online access
- Entrez integrated molecular & literature databases
- BLAST highest volume sequence search service
- VAST structure similarity searches
- Software and Databases

Types of Databases

Primary Databases

- ✓ Original submissions by experimentalists
- ✓ Content controlled by the submitter
- ✓ Examples: GenBank, SNP, GEO

Derivative Databases

- ✓ Built from primary data
- ✓ Content controlled by third party (NCBI)
- ✓ Examples: Refseq, TPA, RefSNP, UniGene, NCBI Protein, Structure, Conserved Domain

What is GenBank?
NCBI's Primary Sequence Database

- Nucleotide only sequence database
- Archival in nature
- Historical
- Reflective of submitter point of view (subjective)
- Redundant

GenBank Data

✓ Direct submissions (traditional records)

✓ Batch submissions (EST, GSS, STS)

✓ ftp accounts (genome data)

DDBJ

EMBL

CIB

getentry

SRS

EBI

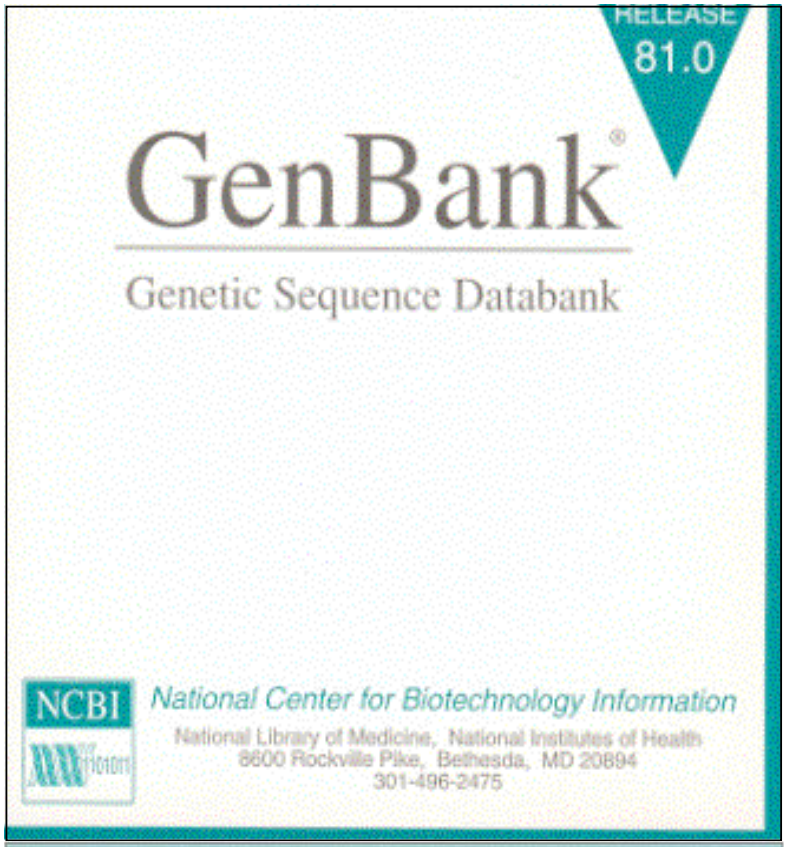
International Sequence Database Collaboration

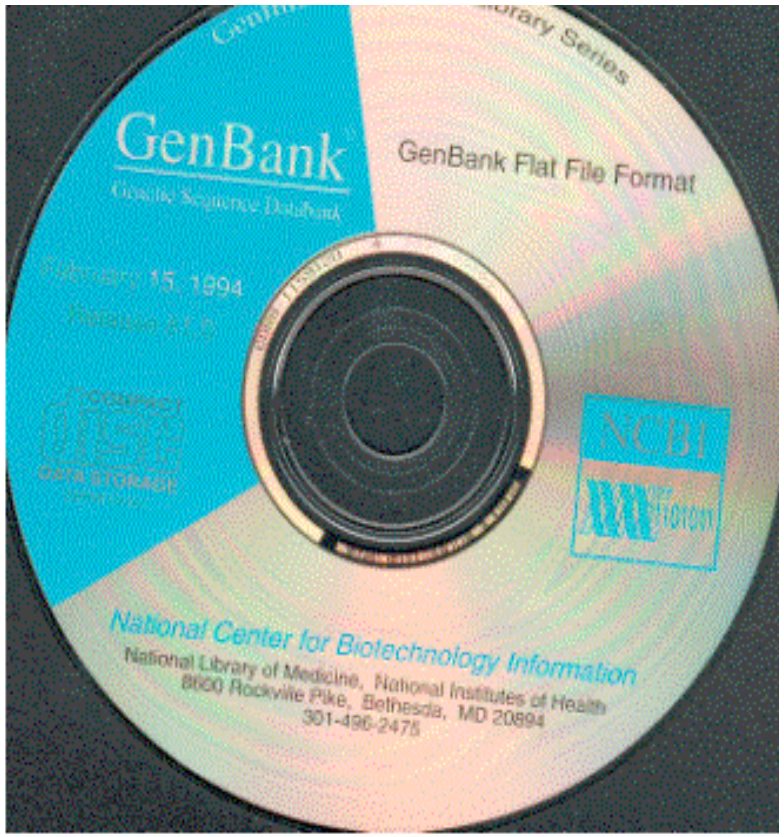
- submit anywhere
- daily updates

GenBank

Entrez

NCBI





GenBank: NCBI's Primary Sequence Database

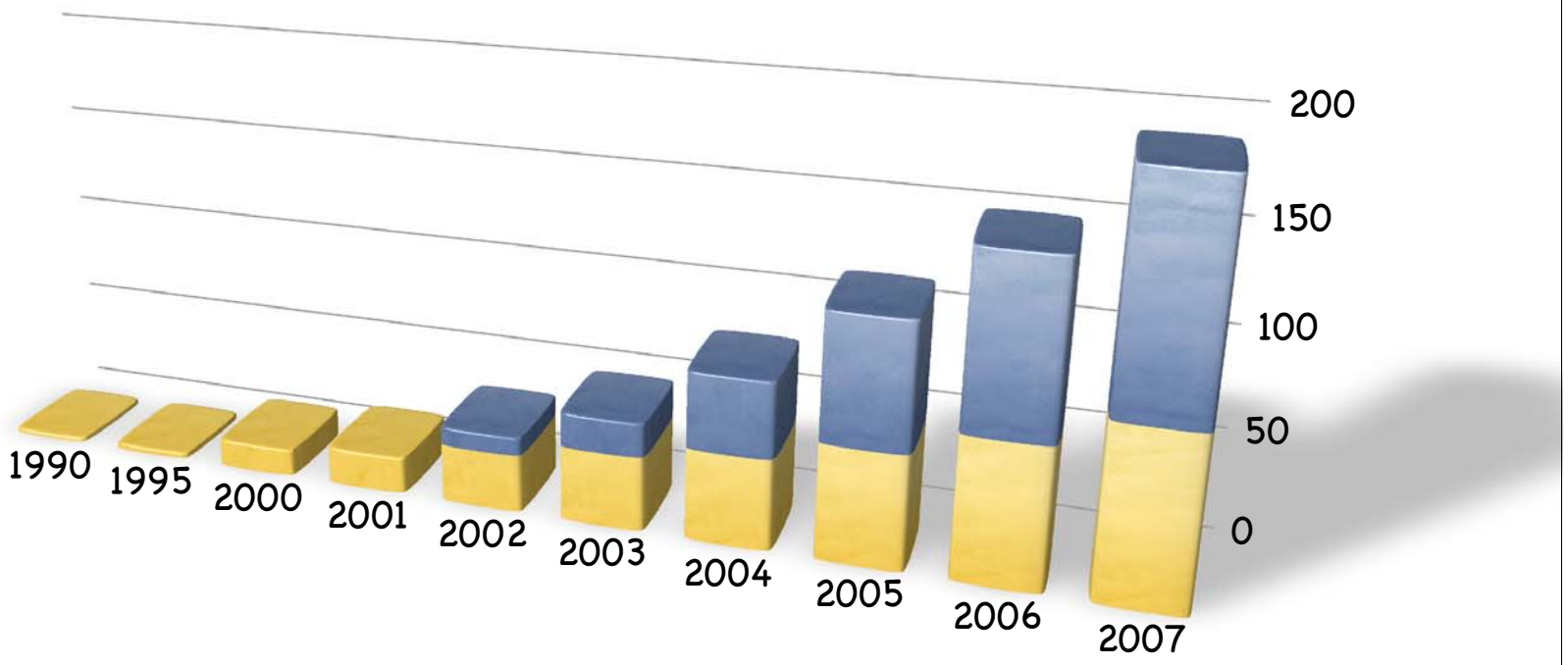
ftp://ftp.ncbi.nih.gov/genbank/

Release 161	August 2007
101,530,711	Records
181,489,883,388*	Total Bases

*includes WGS

- full release every two months
- incremental updates daily
- available only via ftp

Growth of GenBank



Current Release 163

Doubling time 12-14 months

Organization of GenBank

Records are divided into 18 Divisions.

12 Traditional

6 Bulk

- Traditional Divisions:

Direct Submissions (Sequin and BankIt)

Accurate

Well characterized

PRI Primate
PLN Plant and Fungal
BCT Bacterial and Archeal
INV Invertebrate
ROD Rodent
VRL Viral
VRT Other Vertebrate
MAM Mammalian
PHG Phage
SYN Synthetic(cloning vectors)
ENV Environmental Samples
UNA Unannotated

Organization of GenBank

Records are divided into 18 Divisions.

12 Traditional

6 Bulk

- BULK Divisions:

  Batch Submission (Email and FTP)

  Inaccurate

  Poorly characterized

```
EST Expressed Sequence Tag
GSS Genome Survey Sequence
HTG High Throughput Genomic
STS Sequence Tagged Site
HTC High Throughput cDNA
PAT Patent
```

Entrez query: gbdiv_xxx[Properties]

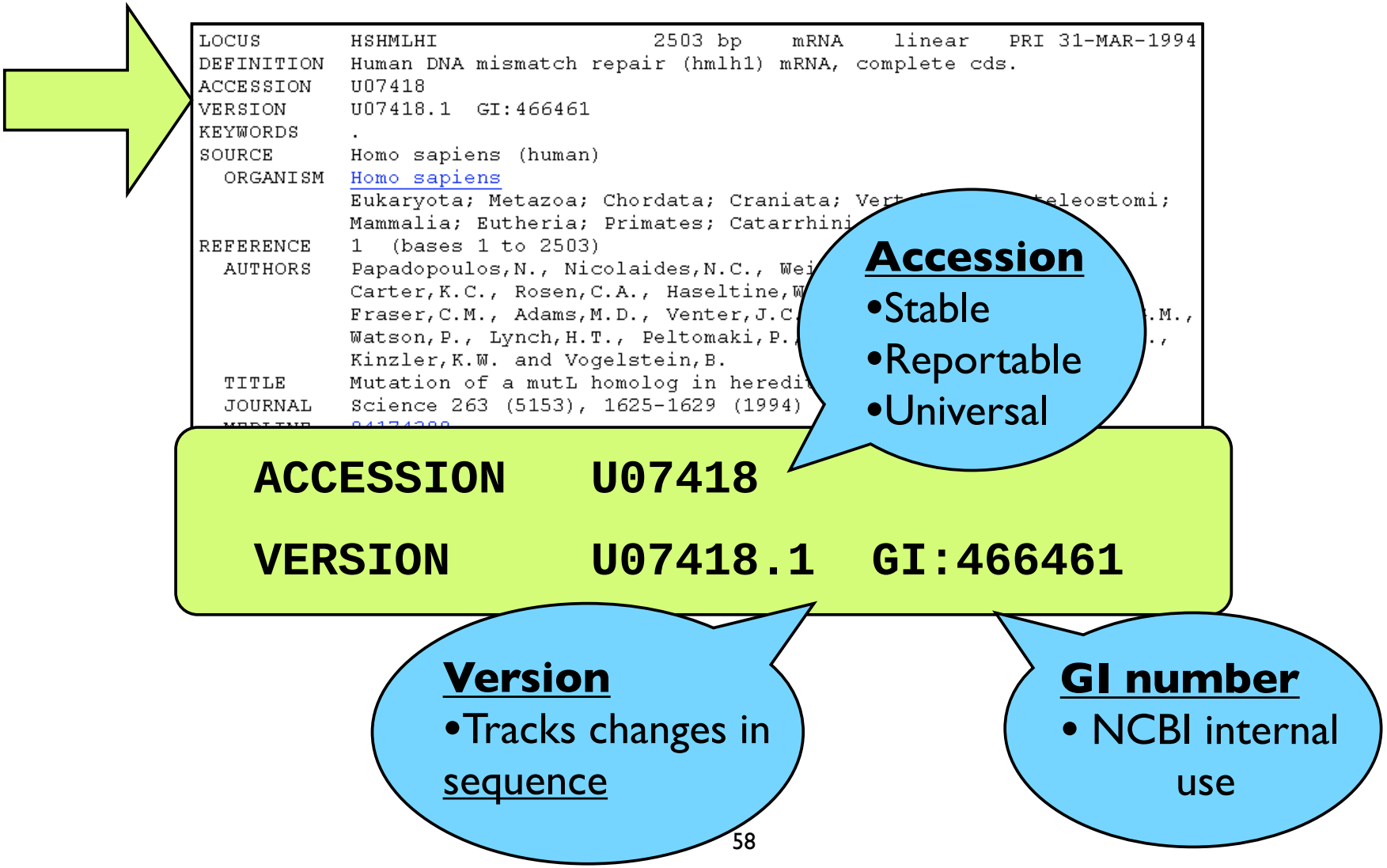
Header

Feature Table

Sequence

A Traditional GenBank Record

Traditional GenBank Record



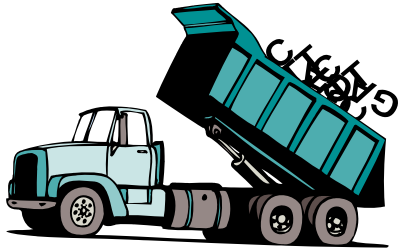
```
FEATURES             Location/Qualifiers
     source          1..2503
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="p21"
                     /tissue_type="gall bladder"
                     /dev_stage="adult"
     gene            1..2503
                     /gene="hmlh1"
     CDS             42..2312
                     /gene="hmlh1"
                     /function="DNA mismatch repair"
                     /note="human homolog of E. coli mutL gene product,
                     Swiss-Prot Accession Number P23367"
                     /codon_start=1
                     /protein_id="AAA17374.1"
                     /db_xref="GI:466462"
                     /translation="MSFVAGVIRRLDETVVNRIAAGEVIQRPANAIKEMIENCLDAKS
                     TSIQVIVKEGGLKLIQIQDNGTGIRKEDLDIVCERFTTSKLQSFEDLASISTYGFRGE
                     ALASISHVAHVTITTKTADGKCAYRASYSDGKLKAPPKPCAGNQGTQITVEDLFYNIA
                     TRRKALKNPSEEYGKILEVVGRYSVHNAGISFSVKKQGETVADVRTLPNASTVDNIRS
                     VFGNAVSRELIEIGCEDKTLAFKMNGYISNANYSVKKCIFLLFINHRLVESTSLRKAI
                     ETVYAAYLPKNTHPFLYLSLEISPQNVDVNVHPTKHEVHFLHEESILERVQQHIESKL
                     LGSNSSRMYFTQTLLPGLAGPSGEMVKSTTSLTSSSTSGSSDKVYAHQMVRTDSREQK
                     LDAFLQPLSKPLSSQPQAIVTEDKTDISSGRARQQDEEMLELPAPAEVAAKNQSLEGD
                     TTKGTSEMSEKRGPTSSNPRKRHREDSDVEMVEDDSRKEMTAACTPRRRIINLTSVLS
                     LQEEINEQGHEVLREMLHNHSFVGCVNPQWALAQHQTKLYLLNTTKLSEELFYQILIY
                     DFANFGVLRLSEPAPLFDLAMLALDSPESGWTEEDGPKEGLAEYIVEFLKKKAEMLAD
                     YFSLEIDEEGNLIGLPLLIDNYVPPLEGLPIFILRLATEVNWDEEKECFESLSKECAM
                     FYSIRKQYISEESTLSGQQSEVPGSIPNSWKWTVEHIVYKALRSHILPPKHFTEDGNI
                     LQLANLPDLYKVFERC"
```

well annotated

```
BASE COUNT      723 a    539 c    599 g    642 t
ORIGIN
        1 gttgaacatc tagacgtttc cttggctctt ctggcgccaa aatgtcgttc gtggcagggg
       61 ttattcggcg gctggacgag acagtggtga accgcatcgc ggcgggggaa gttatccagc
      121 ggccagctaa tgctatcaaa gagatgattg agaactgttt agatgcaaaa tccacaagta
      181 ttcaagtgat tgttaaagag ggaggcctga agttgattca gatccaagac aatggcaccg
      241 ggatcaggaa agaagatctg gatattgtat gtgaaaggtt cactactagt aaactgcagt
      301 cctttgagga tttagccagt atttctacct atggctttcg aggtgaggct ttggccagca
      361 taagccatgt ggctcatgtt actattacaa cgaaaacagc tgatggaaag tgtgcataca
      421 gagcaagtta ctcagatgga aaactgaaag cccctcctaa accatgtgct ggcaatcaag
      481 ggacccagat cacggtggag gacctttttt acaacatagc cacgaggaga aaagctttaa
      541 aaaatccaag tgaagaatat gggaaaattt tggaagttgt tggcaggtat tcagtacaca
      601 atgcaggcat tagtttctca gttaaaaaac aaggagagac agtagctgat gttaggacac
      661 tacccaatgc ctcaaccgtg gacaatattc gctccgtctt tggaaatgct gttagtcgag
      721 aactgataga aattggatgt gaggataaaa ccctagcctt caaaatgaat ggttacatat
      781 ccaatgcaaa ctactcagtg aagaagtgca tcttcttact cttcatcaac catcgtctgg
      841 tagaatcaac ttccttgaga aaagccatag aaacagtgta tgcagcctat ttgcccaaaa
      901 acacacaccc attcctgtac ctcagtttag aaatcagtcc ccagaatgtg gatgttaatg
      961 tgcaccccac aaagcatgaa gttcacttcc tgcacgagga gagcatcctg gagcgggtgc
     1021 agcagcacat cgagagcaag ctcctgggct ccaattcctc caggatgtac ttcacccaga
     1081 ctttgctacc aggacttgct ggcccctctg gggagatggt taaatccaca acaagtctga
     1141 cctcgtcttc tacttctgga agtagtgata aggtctatgc ccaccagatg gttcgtacag
     1201 attcccggga acagaagctt gatgcatttc tgcagcctct gagcaaaccc ctgtccagtc
     1261 agccccaggc cattgtcaca gaggataaga cagatatttc tagtggcagg gctaggcagc
     1321 aagatgagga gatgcttgaa ctcccagccc ctgctgaagt ggctgccaaa aatcagagct
     1381 tggaggggga tacaacaaag gggacttcag aaatgtcaga gaagagagga cctacttcca
     1441 gcaaccccag aaagagacat cgggaagatt ctgatgtgga aatggtggaa gatgattccc
     1501 gaaaggaaat gactgcagct tgtacccccc ggagaaggat cattaacctc actagtgttt
     1561 tgagtctcca ggaagaaatt aatgagcagg gacatgaggt tctccgggag atgttgcata
     1621 accactcctt cgtgggctgt gtgaatcctc agtgggcctt ggcacagcat caaaccaagt
     1681 tataccttct caacaccacc aagcttagtg aagaactgtt ctaccagata ctcatttatg
     1741 attttgccaa ttttggtgtt ctcaggttat cggagccagc accgctcttt gaccttgcca
     1801 tgcttgcctt agatagtcca gagagtggct ggacagagga agatggtccc aaagaaggac
     1861 ttgctgaata cattgttgag tttctgaaga agaaggctga gatgcttgca gactatttct
     1921 ctttggaaat tgatgaggaa gggaacctga ttggattacc ccttctgatt gacaactatg
     1981 tgcccccttt ggagggactg cctatcttca ttcttcgact agccactgag gtgaattggg
     2041 acgaagaaaa ggaatgtttt gaaagcctca gtaaagaatg cgctatgttc tattccatcc
     2101 ggaagcagta catatctgag gagtcgaccc tctcaggcca gcagagtgaa gtgcctggct
     2161 ccattccaaa ctcctggaag tggactgtgg aacacattgt ctataaagcc ttgcgctcac
     2221 acattctgcc tcctaaacat ttcacagaag atggaaatat cctgcagctt gctaacctgc
     2281 ctgatctata caaagtcttt gagaggtgtt aaatatggtt atttatgcac tgtgggatgt
     2341 gttcttcttt ctctgtattc cgatacaaag tgttgtatca aagtgtgata tacaaagtgt
     2401 accaacataa gtgttggtag cacttaagac ttatacttgc cttctgatag tattccttta
     2461 tacacagtgg attgattata aataaataga tgtgtcttaa cat
//
```

the sequence is the data

Primary vs. Derivative Databases



TATAGCCG

Sequencing Centers

ACGTGC

GenBank

Algorithms

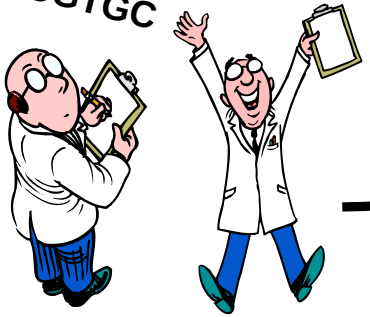
UniGene

Updated continually by NCBI

RefSeq: Annotation Pipeline

RefSeq: Entrez Gene and Genomes Pipelines

ACGTGC



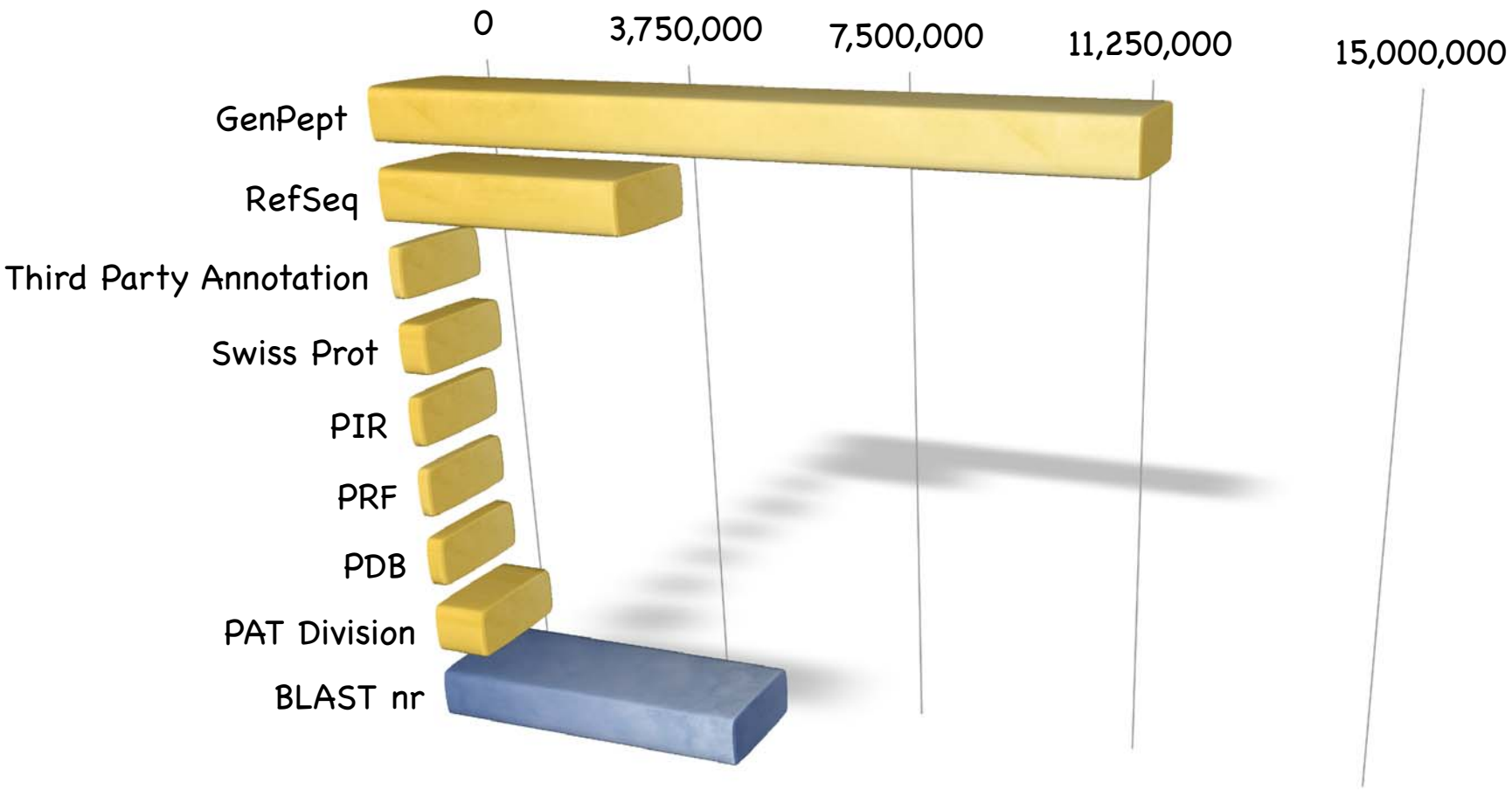
Labs

Curators

TATAGCCG
AGCTCCGATA
CCGATGACAA

Derivative Databases

Entrez Protein



GenPept

- GenBank CDS translations

```
FEATURES             Location/Qualifiers
     source          1..2484
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3p22-p23"
     gene            1..2484
                     /gene="MLH1"
     CDS             22..2292
                     /gene="MLH1"
                     /note="hom
                     Number P14242), S. cerevisiae MLH1 (GenBank Accession
                     Number U07187), E. col  MUTL (Swiss-Prot Accession Number
                     P23367), Salmonella t imurium MUTL (Swiss-Prot Accession
                     Number P14161) and S  ococcus pneumoniae (Swiss-Prot
                     Accession Number P141  "
                     /codon_start=1
                     /product="DNA mismatc  epair protein homolog"
                     /protein_id="AAC50285
                     /db_xref="GI:463989"
                     /translation="MSFVAGV  LDET VVNRIAAGEVIQRPANAIKEMIENCLDAKS
                     TSIQVIVKEGGLKLIQIQDNG  RKED LDIVCERFTTSKLQSFEDLASISTYGFRGE
                     ALASISHVAHVTITTKTADGK  RASY SDGKLKAPPKPCAGNQGTQITVEDLFYNIA
                     TRRKALKNPSEEYGKILE VVGRYSVHNA GISFSVKKQGETVADVRTLPNASTVDNIRS
```

```
>gi|463989|gb|AAC50285.1| DNA mismatch repair prote...
MSFVAGVIRRLDETVVNRIAAGEVIQRPANAIKEMIENCLDAKSTSIQVIV...
EDLDIVCERFTTSKLQSFEDLASISTYGFRGEALASISHVAHVTITTKTAD...
```

RefSeq

- The goal is to provide a comprehensive, standard dataset that represents sequence information for a species.
  - transcript, protein, assembled genomic (contigs), and chromosome records
  - known and predicted
  - reviewed
  - human, mouse, rat, fruit fly, zebrafish, arabidopsis, microbial genome (proteins), organelles, and more

RefSeq Accession Numbers

- mRNAs and Proteins

NM_123456	Curated mRNA
NP_123456	Curated Protein
NR_123456	Curated nc RNA
XM_123456	Predicted mRNA
XP_123456	Predicted Protein
XR_123456	Predicted nc RNA

- Genomic Records

NG_123456 Reference Genomic Sequence

- Chromosome

NC_123455 Microbial replicons, organelle, genomes, human chromosomes

- Assemblies

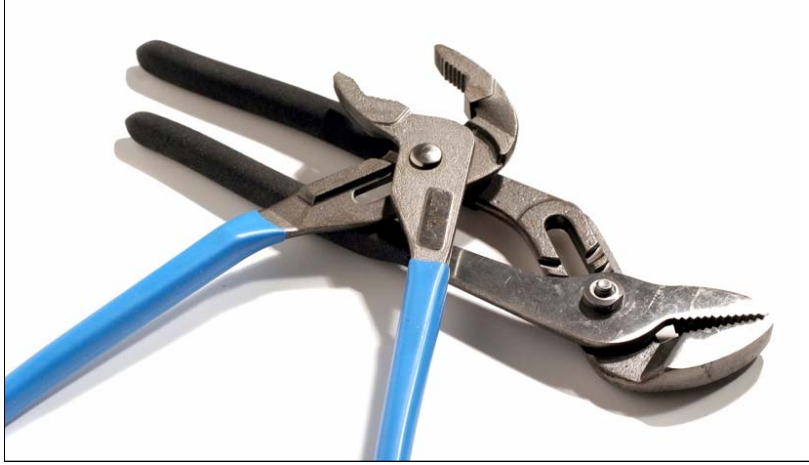
NT_123456	Contig
NW_123456	WGS Supercontig

Other NCBI Databases

Structure:	imported structures (PDB)	Cn3D viewer, NCBI curation
CDD:	conserved domain database	Protein families (COGs and KOGs); Single domains (PFAM, SMART, CD)
dbSNP:	nucleotide polymorphism	
Gene:	gene records	Unifies LocusLink and Microbial Genomes
HomoloGene:	homologs	neighboring function for Gene

Sequence Databases

GUIDED TOUR: Retrieving Data



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

⟷ WWW

⟷ Entrez

⟷ BLAST

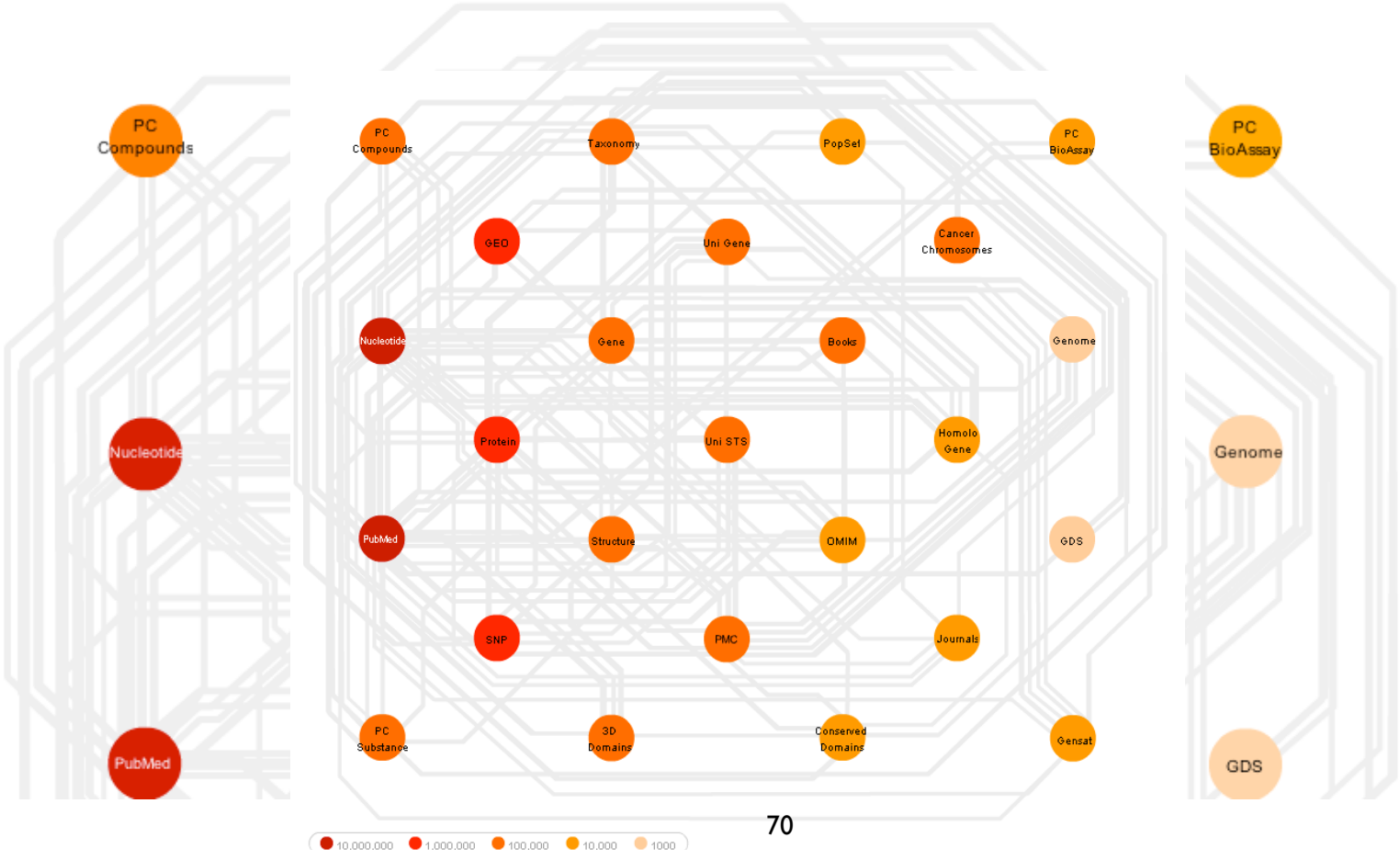
http://www.ncbi.nih.gov/Database/datamodel

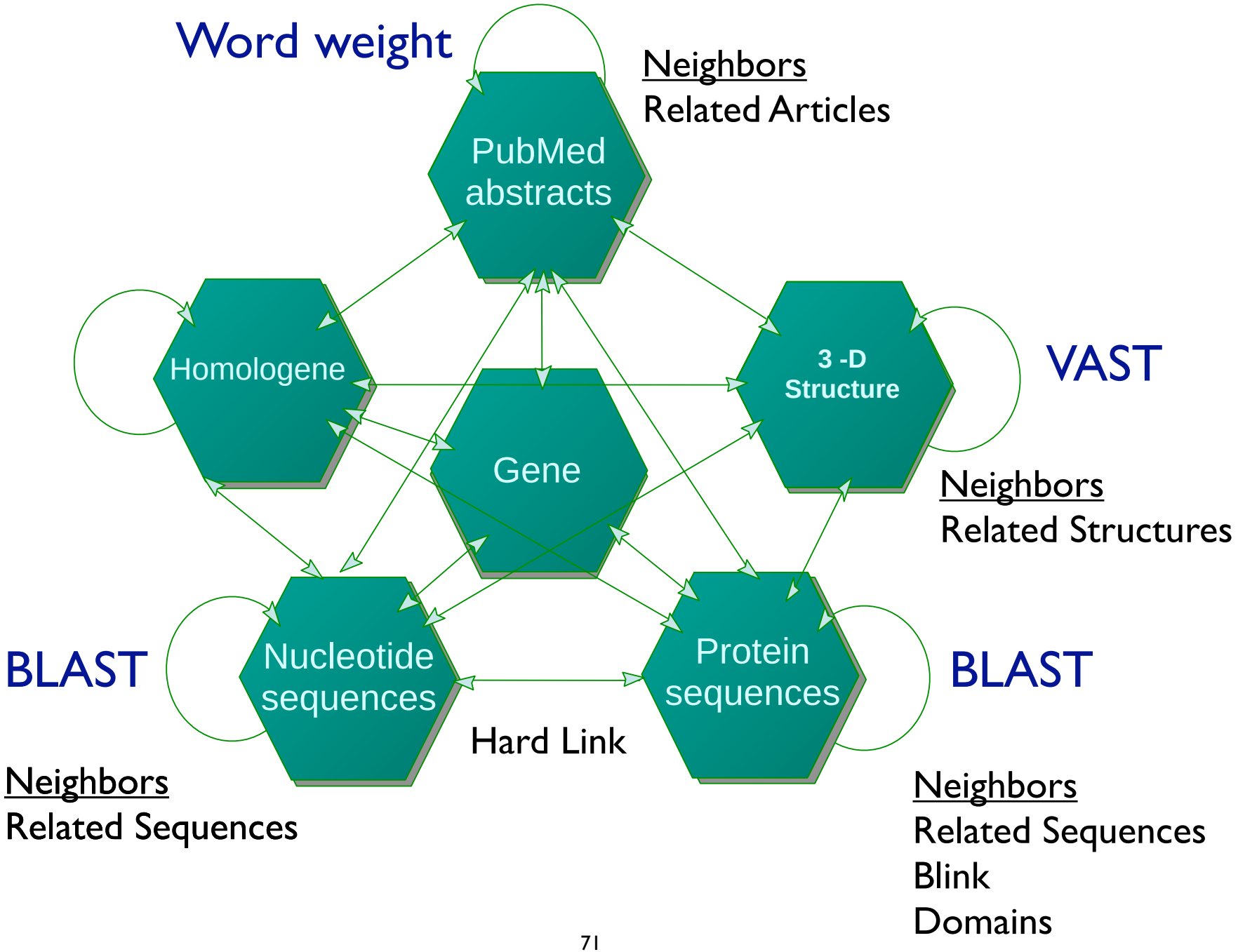
NCBI Model of Entrez Databases

Search Entrez for Go

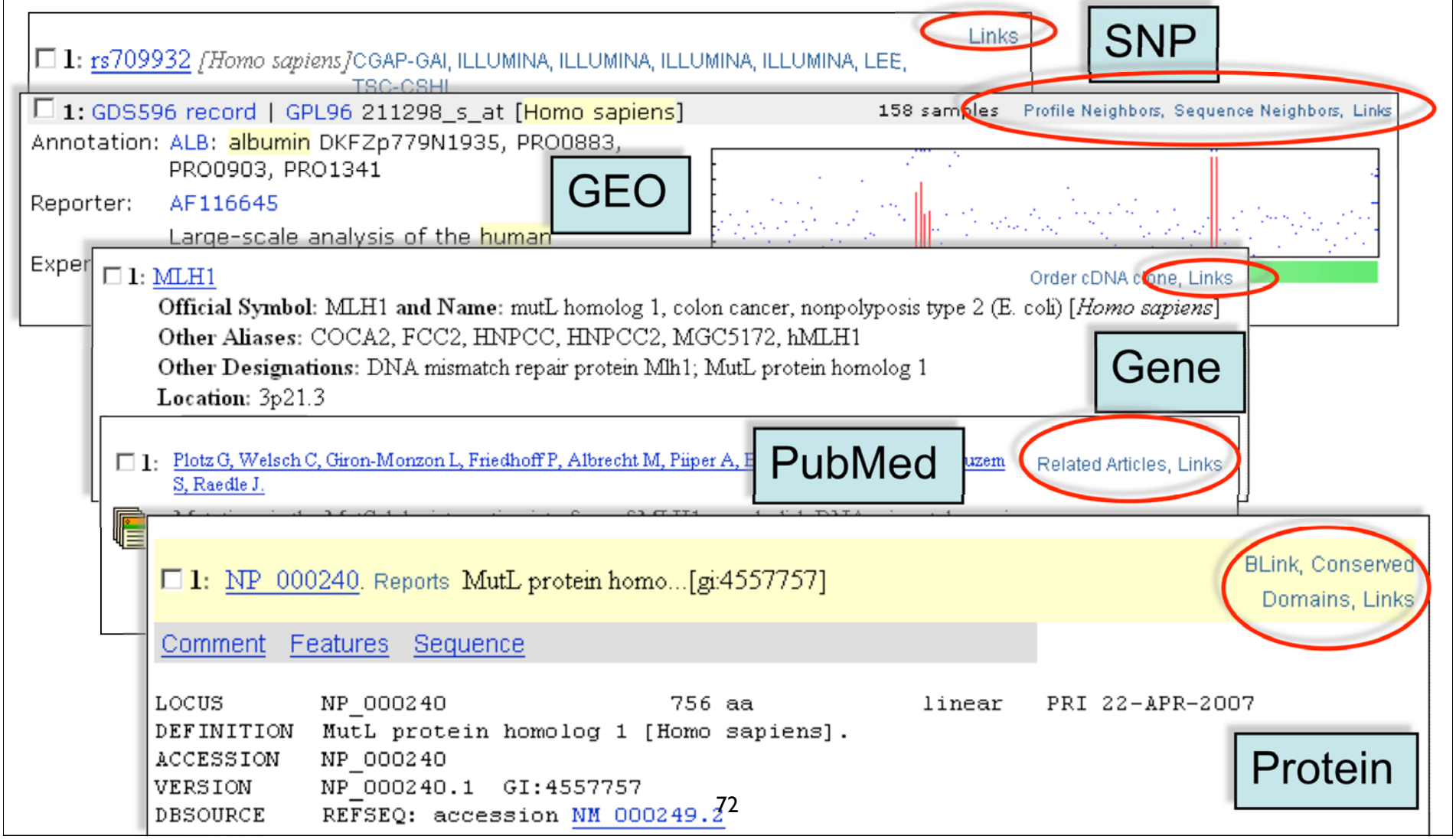
The diagram shows the Entrez databases and the connections between them. Each database is represented by a colored circle, where the color indicates the approximate number of records in the database. Mouse over a circle to see which databases are linked to the one selected, and how many links exist between those databases.

This diagram requires Flash for viewing.

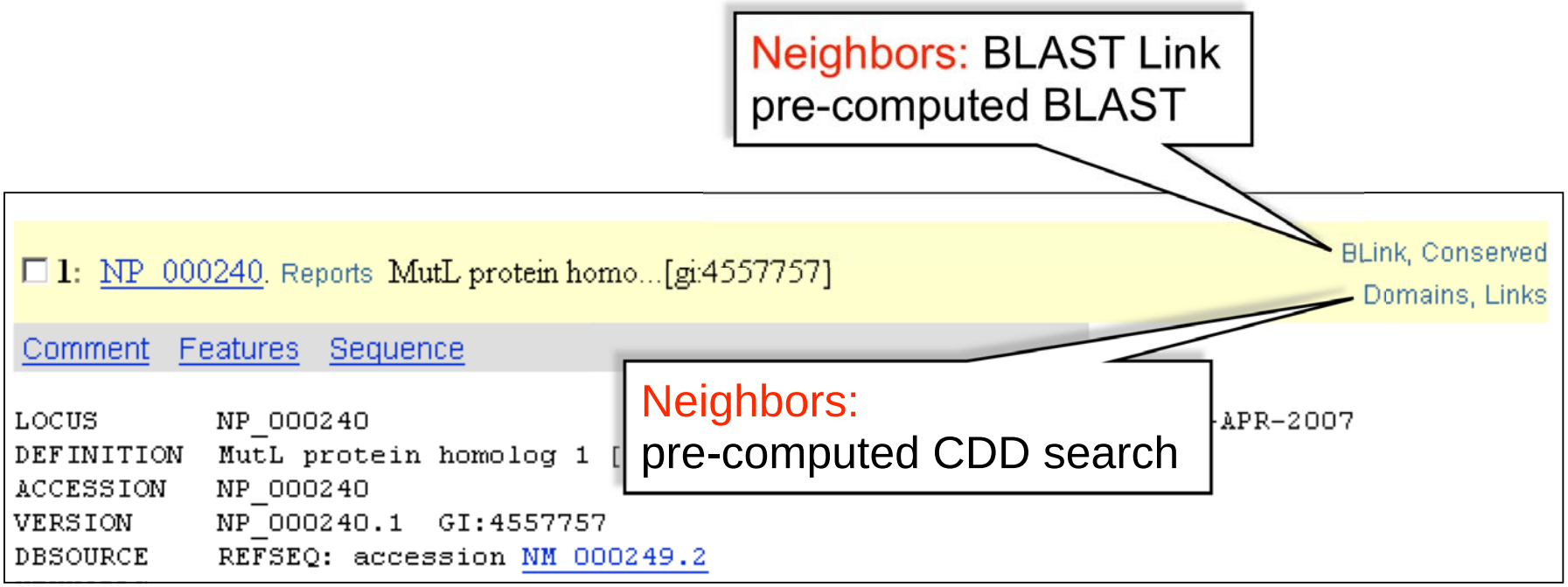



Word weight
Neighbors
Related Articles
PubMed abstracts
Homologene
Gene
3 -D Structure
VAST
Neighbors
Related Structures
BLAST
Nucleotide sequences
Protein sequences
BLAST
Hard Link
Neighbors
Related Sequences
Neighbors
Related Sequences
Blink
Domains

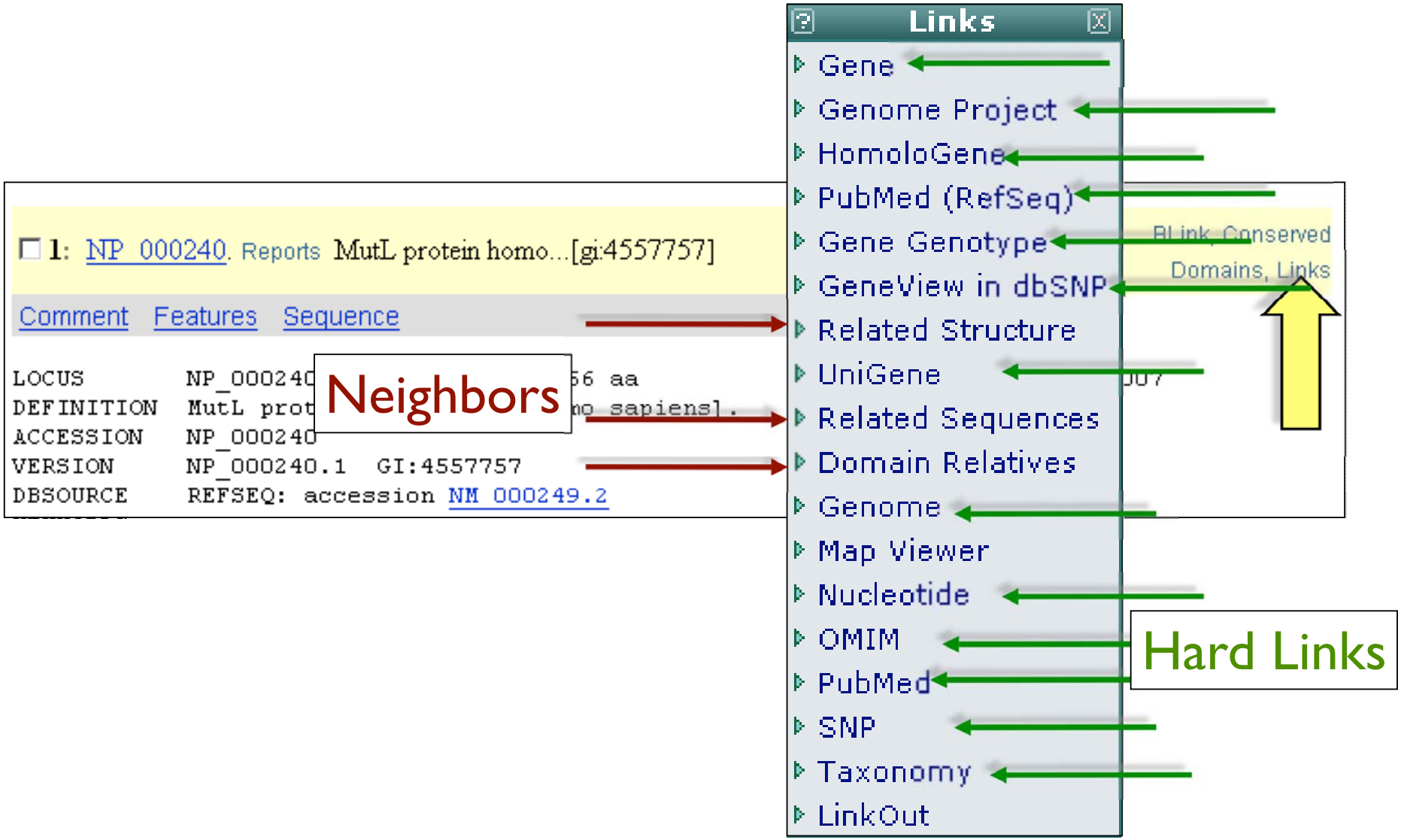
Neighbors in Entrez



Blink & Domains



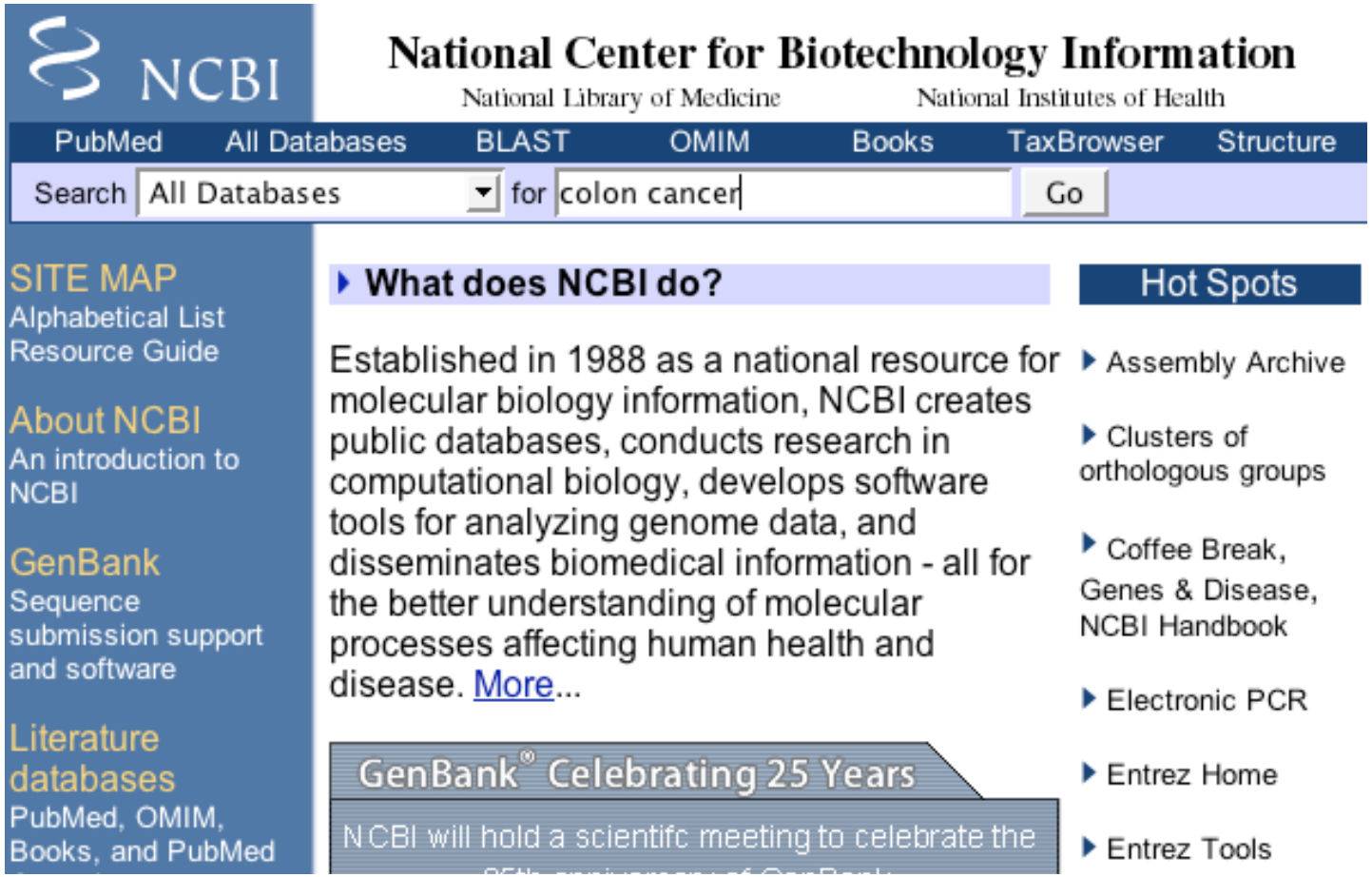
Links


Links
Gene
Genome Project
HomoloGene
PubMed (RefSeq)
Gene Genotype
GeneView in dbSNP
Related Structure
UniGene
Related Sequences
Domain Relatives
Genome
Map Viewer
Nucleotide
OMIM
PubMed
SNP
Taxonomy
LinkOut
Neighbors
Hard Links
1: NP_000240. Reports MutL protein homo...[gi:4557757]
Comment Features Sequence
LOCUS NP_000240
DEFINITION MutL prot
ACCESSION NP_000240
VERSION NP_000240.1 GI:4557757
DBSOURCE REFSEQ: accession NM_000249.2
BLink, Conserved Domains, Links

Database searching with Entrez

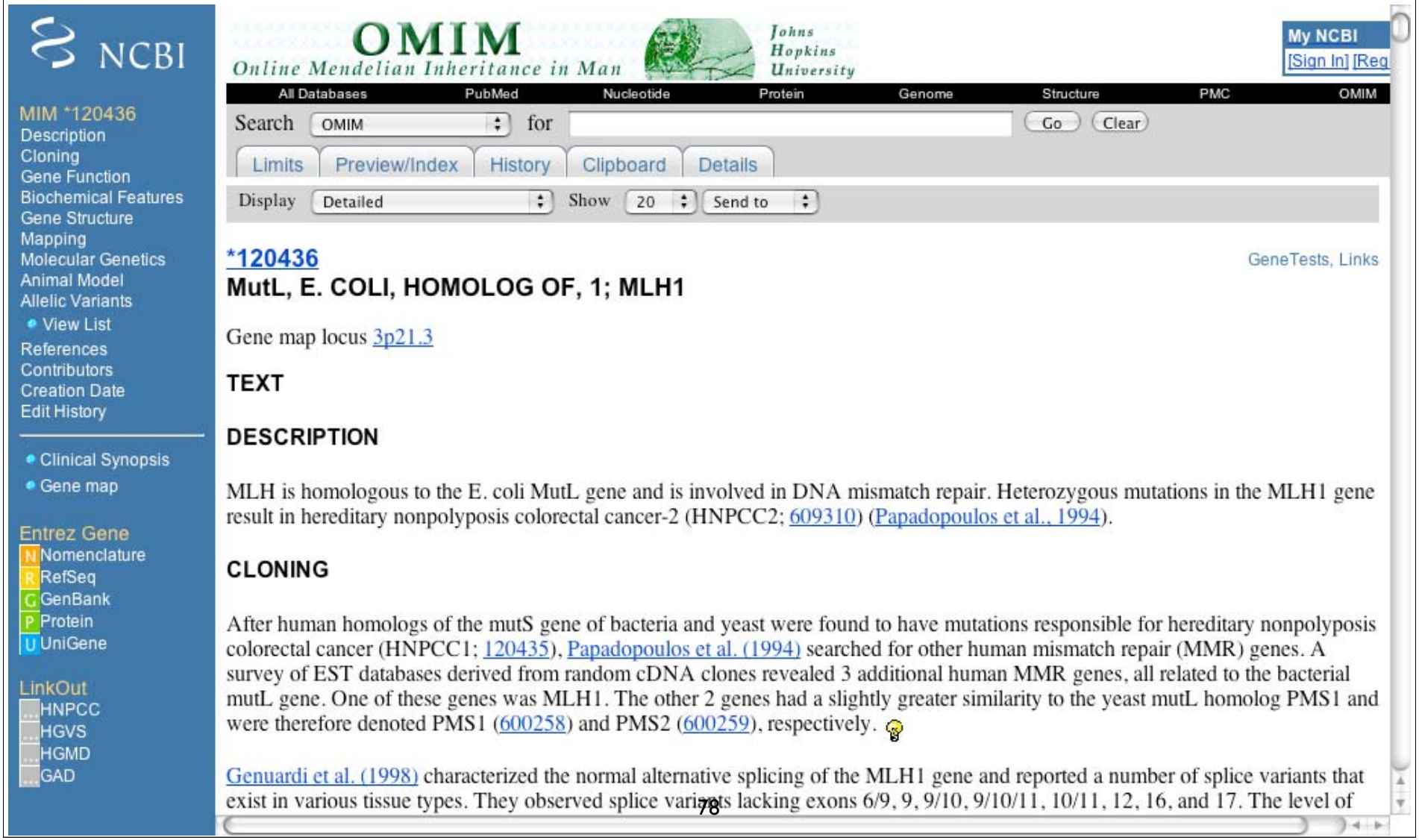
- **Scenario:** Let's find out more about the MutL gene
- Using limits and field restriction to find human MutL homolog
- Linking and neighboring with MutL

Start with a search for "colon cancer"

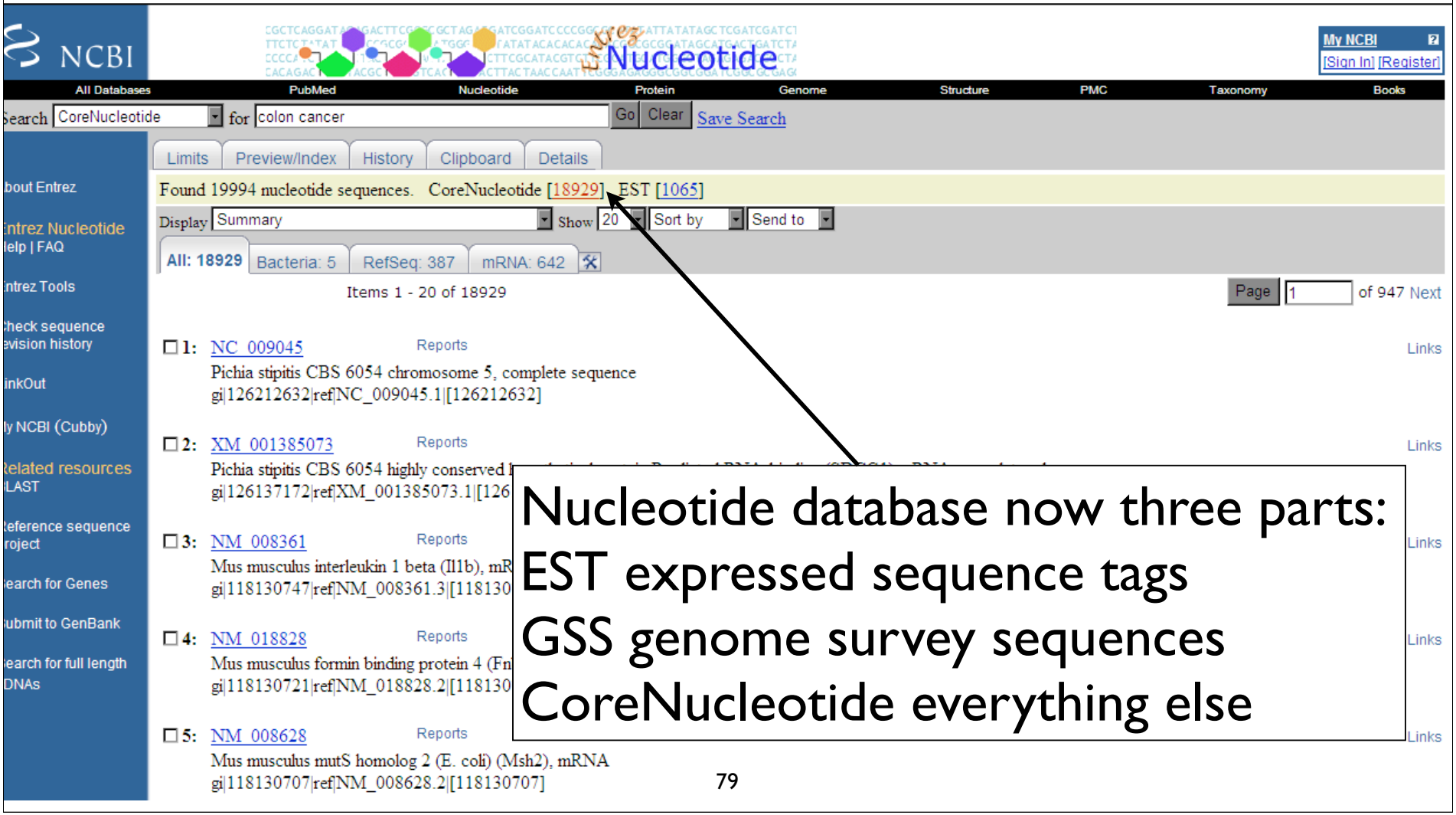



NCBI
Entrez, The Life Sciences Search Engine
HOME
SEARCH
SITE MAP
PubMed
All Databases
Human Genome
GenBank
Map Viewer
BLAST
Search across databases
colon cancer
GO
Clear
Help
- Result counts displayed in gray indicate one or more terms not found
58219 PubMed: biomedical literature citations and abstracts
7197 PubMed Central: free, full text journal articles
7 Site Search: NCBI web and FTP sites
894 Books: online books
374 OMIM: online Mendelian Inheritance in Man
none OMIA: online Mendelian Inheritance in Animals
19529 CoreNucleotide: Core subset of nucleotide sequence records
1156 EST: Expressed Sequence Tag records
none GSS: Genome Survey Sequence records
940 Protein: sequence database
6 Genome: whole genome sequences
2 Structure: three-dimensional macromolecular structures
none Taxonomy: organisms in GenBank
none SNP: single nucleotide polymorphism
493 Gene: gene-centered information
20 HomoloGene: eukaryotic homology groups
2 dbGaP: genotype and phenotype
160 UniGene: gene-oriented clusters of transcript sequences
6 CDD: conserved protein domain database
19 3D Domains: domains from Entrez Structure
34 UniSTS: markers and mapping data
2 PopSet: population study data sets
109008 GEO Profiles: expression and molecular abundance profiles
83 GEO DataSets: experimental sets of GEO data
123 Cancer Chromosomes: cytogenetic databases
4 PubChem BioAssay: bioactivity screens

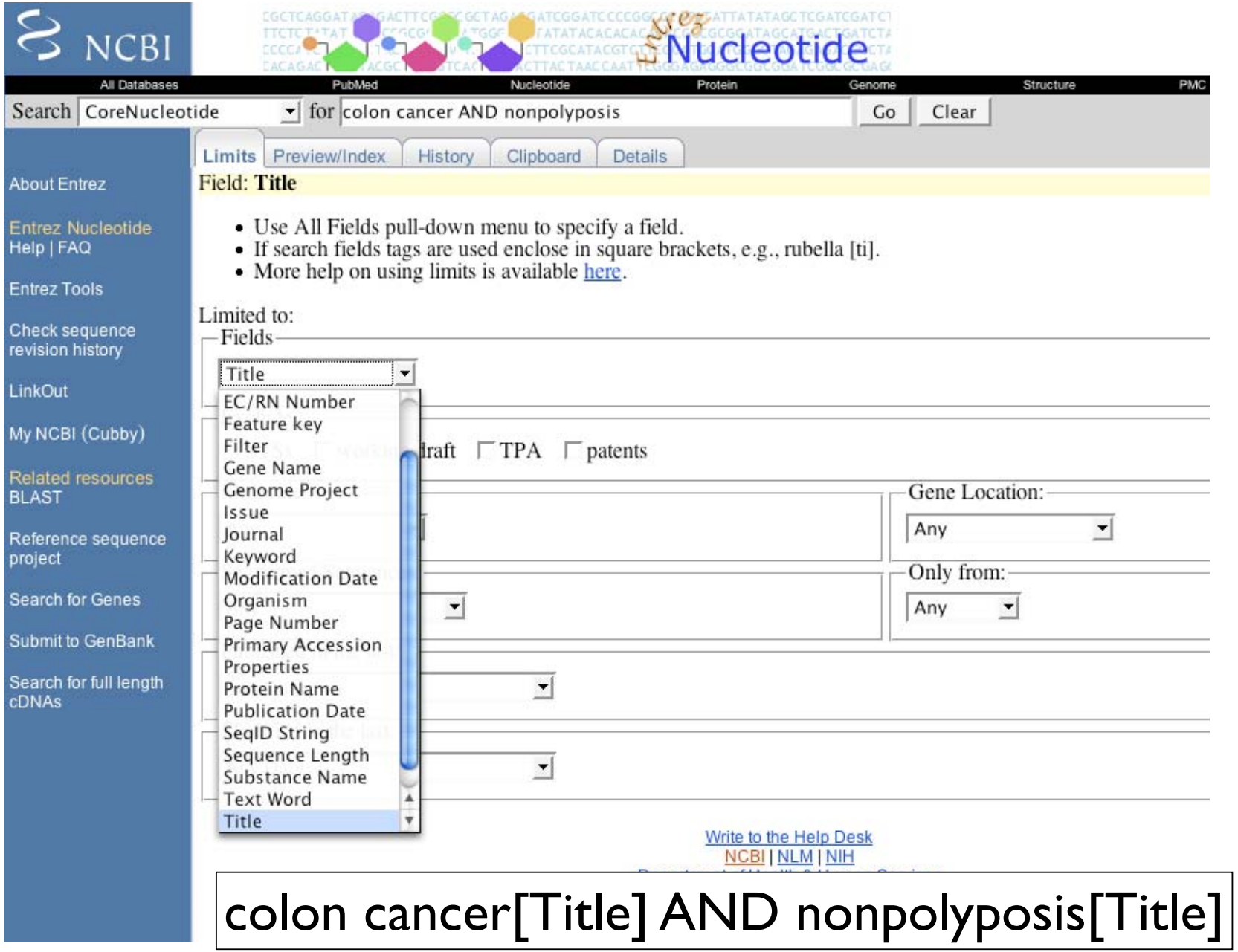
Human Disease Genes



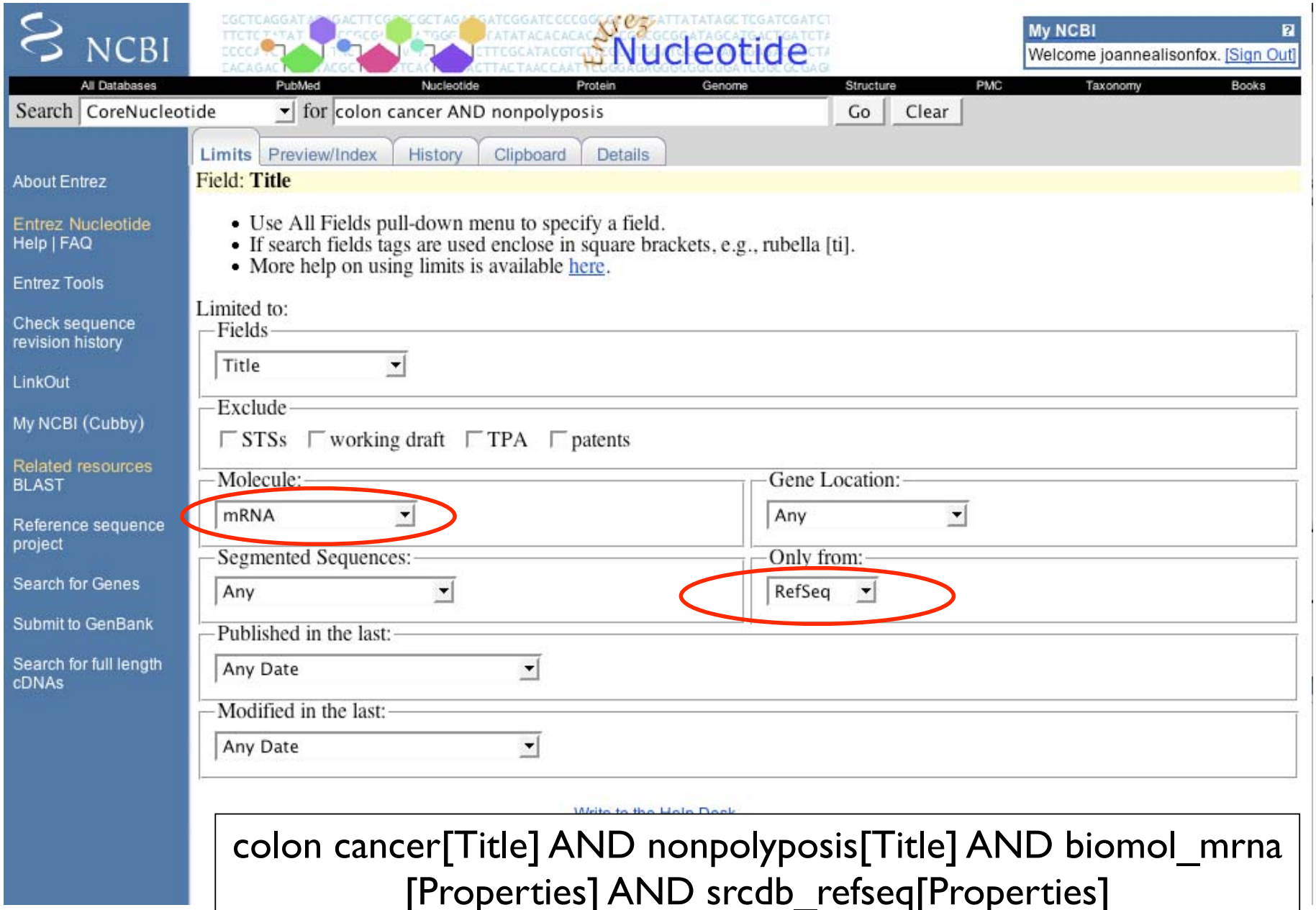
Search CoreNucleotide



Advanced Search Options

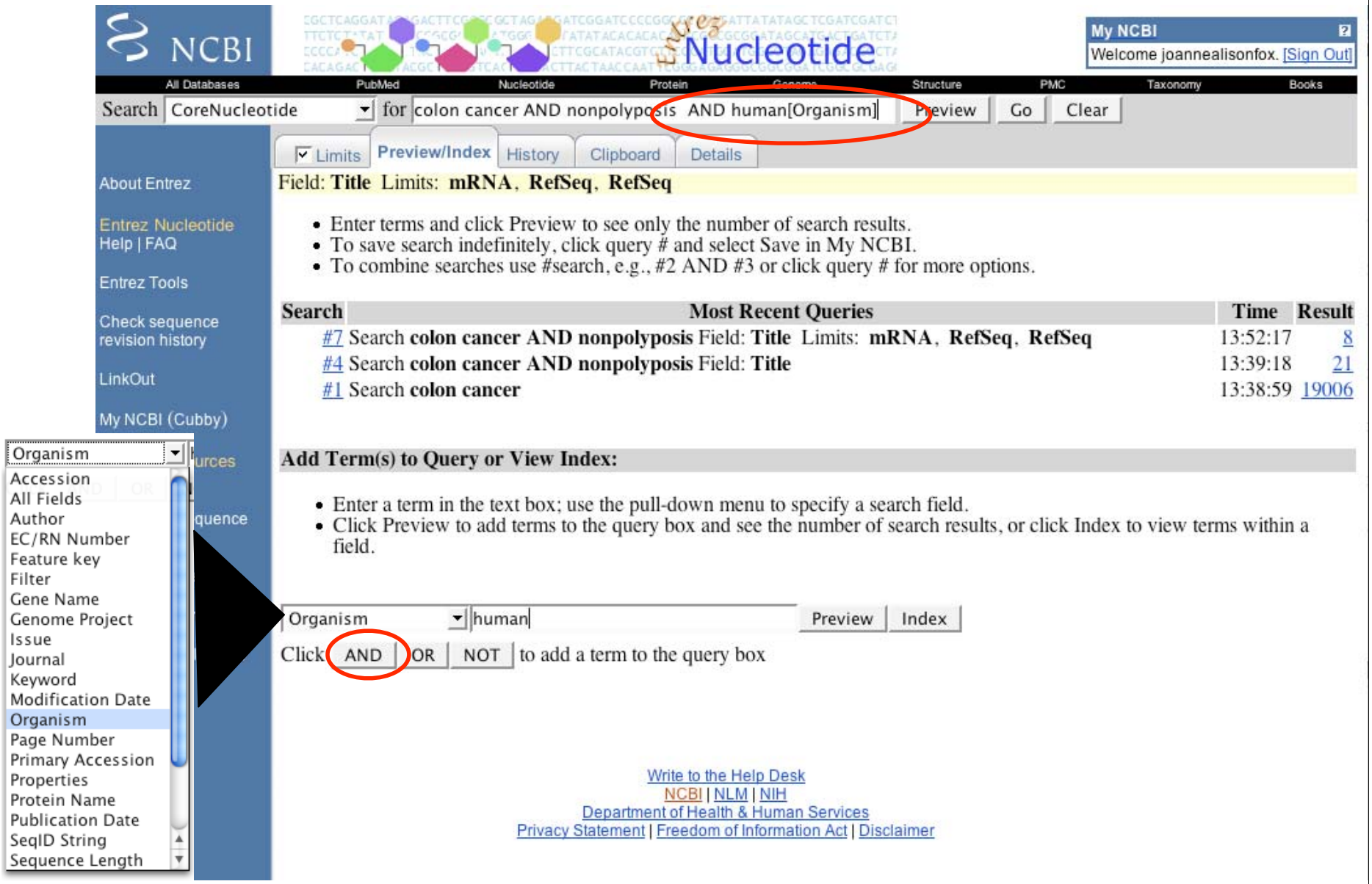


colon cancer[Title] AND nonpolyposis[Title]

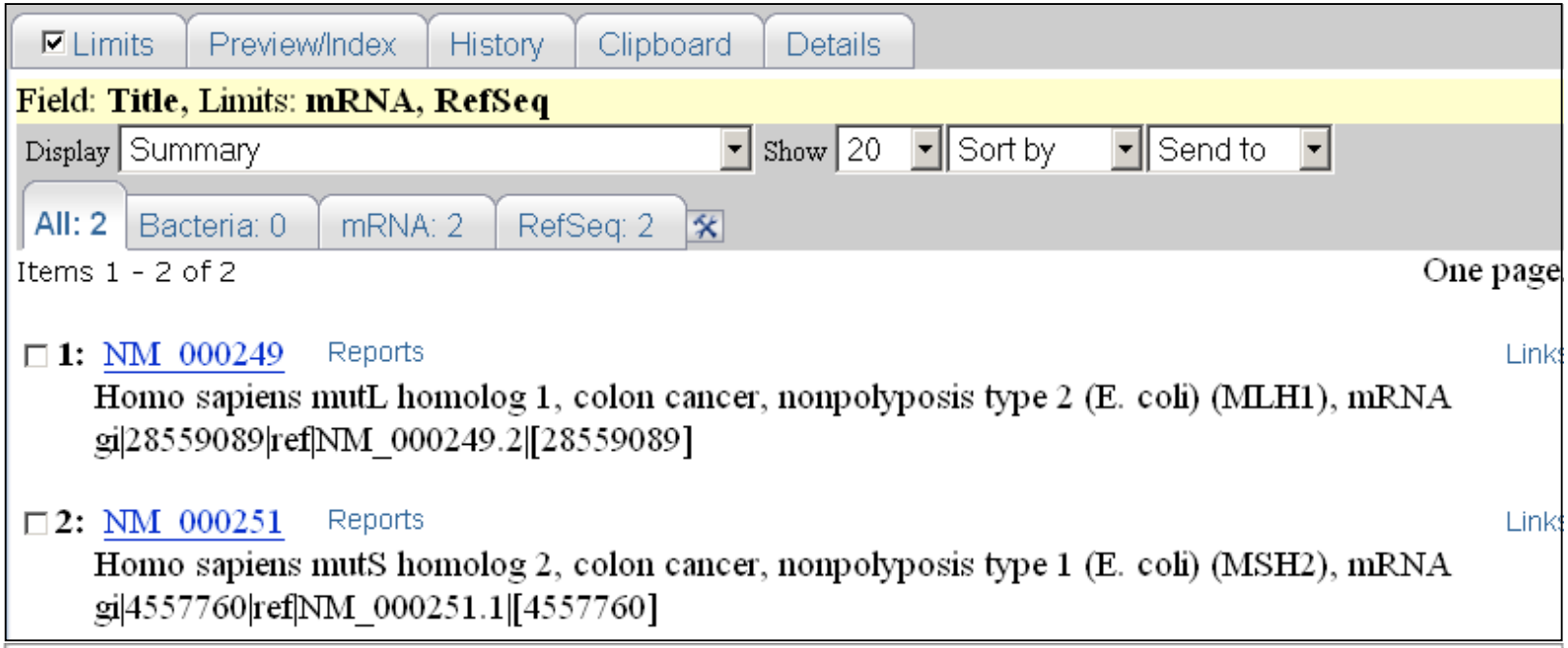


colon cancer[Title] AND nonpolyposis[Title] AND biomol_mrna [Properties] AND srcdb_refseq[Properties]

Advanced Search Options


NCBI
Entrez Nucleotide
My NCBI
Welcome joannealisonfox. [Sign Out]
All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search CoreNucleotide for colon cancer AND nonpolyposis AND human[Organism] Preview Go Clear
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Field: Title Limits: mRNA, RefSeq, RefSeq
Enter terms and click Preview to see only the number of search results.
To save search indefinitely, click query # and select Save in My NCBI.
To combine searches use #search, e.g., #2 AND #3 or click query # for more options.
Search Most Recent Queries Time Result
#7 Search colon cancer AND nonpolyposis Field: Title Limits: mRNA, RefSeq, RefSeq 13:52:17 8
#4 Search colon cancer AND nonpolyposis Field: Title 13:39:18 21
#1 Search colon cancer 13:38:59 19006
Add Term(s) to Query or View Index:
Enter a term in the text box; use the pull-down menu to specify a search field.
Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.
Organism
Accession
All Fields
Author
EC/RN Number
Feature key
Filter
Gene Name
Genome Project
Issue
Journal
Keyword
Modification Date
Organism
Page Number
Primary Accession
Properties
Protein Name
Publication Date
SeqID String
Sequence Length
Organism human Preview Index
Click AND OR NOT to add a term to the query box
Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer

Refining your Search



colon cancer[Title] AND nonpolyposis[Title] AND human[Organism] AND biomol_mrna [Properties] AND srcdb_refseq[Properties]

Useful Field Restrictions

- **[Title]:** Definition line in GenBank / GenPept format shown in Summary format

  `glyceraldehyde 3 phosphate dehydrogenase[Title]`

- **[Organism]:** NCBI's taxonomy. Organizing system for molecular databases

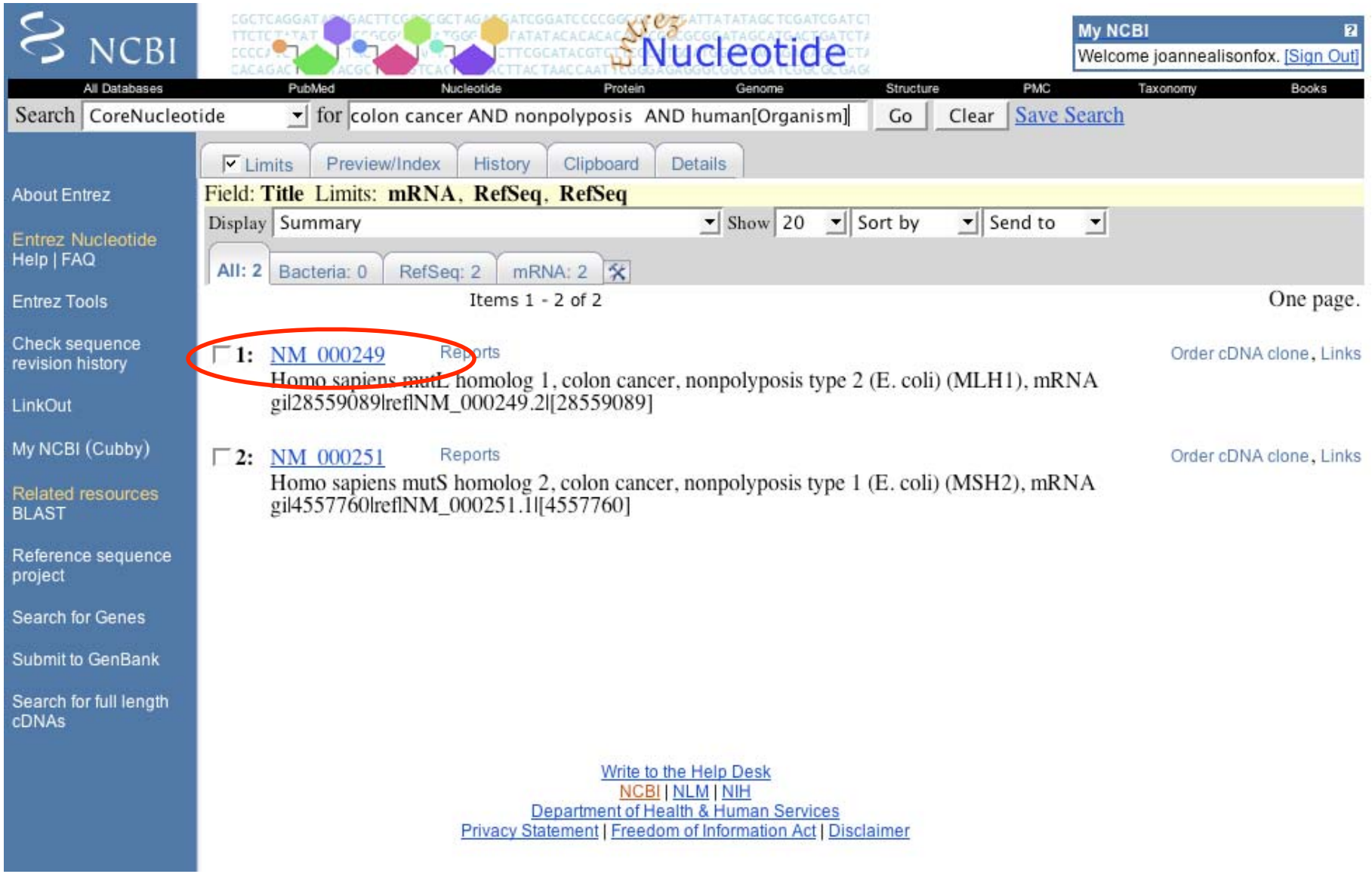
  `mouse[organism]; green plants[organism]; Streptomyces coelicolor [organism]`

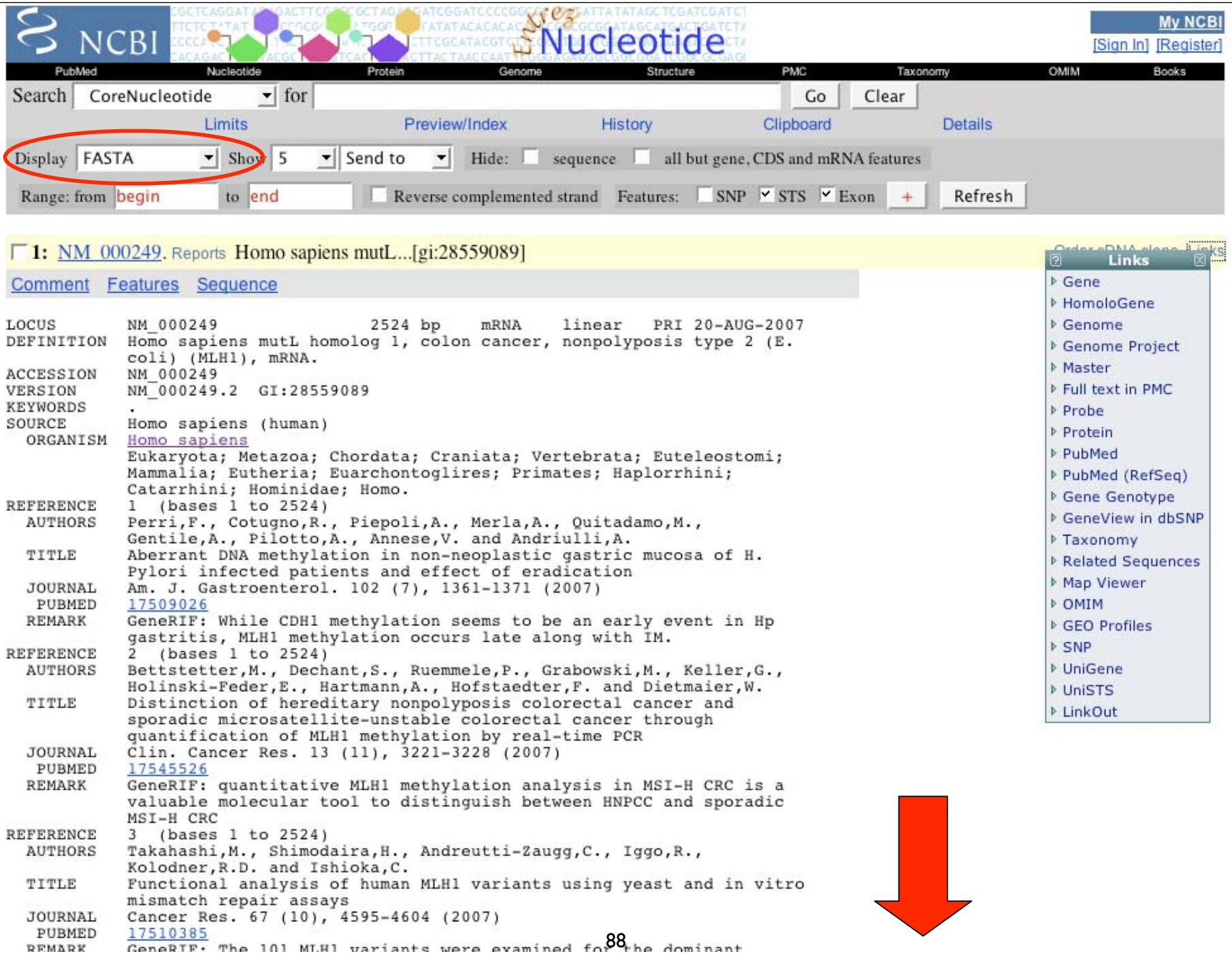
- **[Properties]:** molecule type, location, database source

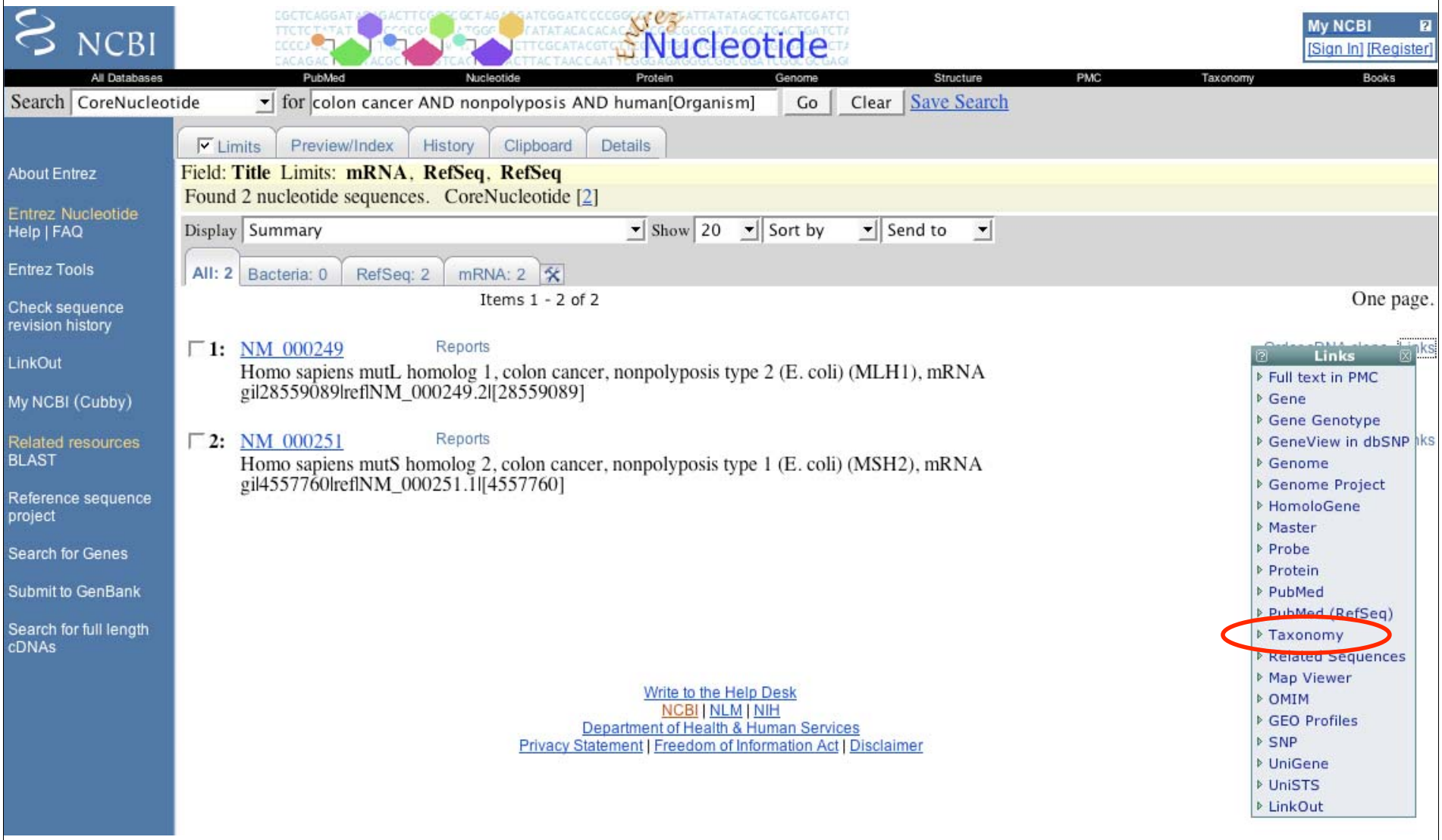
  `biomol_mrna[properties]; biomol_genomic[properties]; gene_in_mitochondrion[properties]; srcdb_pdb[properties]`

- **[Filter]:** subsets of data, Entrez links

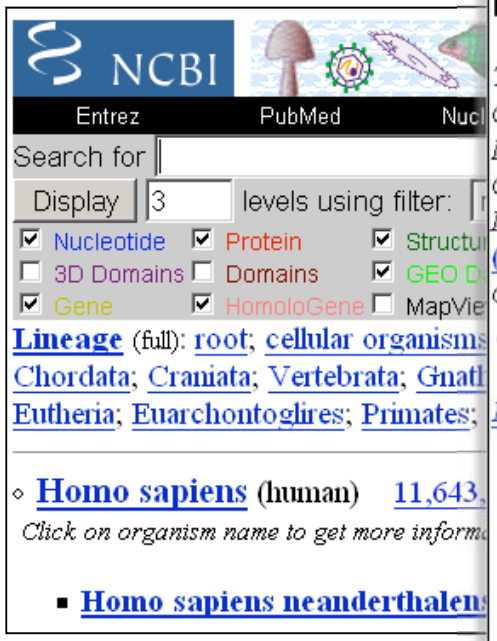
  `all[filter]; nucleotide mapview[filter]; nucleotide_omim[filter]`


NCBI
Entrez Nucleotide
My NCBI
Welcome joannealisonfox. [Sign Out]
All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search CoreNucleotide for colon cancer AND nonpolyposis AND human[Organism] Go Clear Save Search
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Field: Title Limits: mRNA, RefSeq, RefSeq
Display Summary Show 20 Sort by Send to
All: 2 Bacteria: 0 RefSeq: 2 mRNA: 2
Items 1 - 2 of 2
One page.
1: NM_000249 Reports
Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), mRNA
gi|28559089|ref|NM_000249.2|[28559089]
Order cDNA clone, Links
2: NM_000251 Reports
Homo sapiens mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) (MSH2), mRNA
gi|4557760|ref|NM_000251.1|[4557760]
Order cDNA clone, Links
Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer


NCBI
Nucleotide
My NCBI
[Sign In] [Register]
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Search CoreNucleotide for Go Clear
Limits Preview/Index History Clipboard Details
Display FASTA Show 5 Send to Hide: sequence all but gene, CDS and mRNA features
Range: from begin to end Reverse complemented strand Features: SNP STS Exon + Refresh
1: NM_000249. Reports Homo sapiens mutL...[gi:28559089]
Comment Features Sequence
LOCUS NM_000249 2524 bp mRNA linear PRI 20-AUG-2007
DEFINITION Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), mRNA.
ACCESSION NM_000249
VERSION NM_000249.2 GI:28559089
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2524)
AUTHORS Perri,F., Cotugno,R., Piepoli,A., Merla,A., Quitadamo,M., Gentile,A., Pilotto,A., Annese,V. and Andriulli,A.
TITLE Aberrant DNA methylation in non-neoplastic gastric mucosa of H. Pylori infected patients and effect of eradication
JOURNAL Am. J. Gastroenterol. 102 (7), 1361-1371 (2007)
PUBMED 17509026
REMARK GeneRIF: While CDH1 methylation seems to be an early event in Hp gastritis, MLH1 methylation occurs late along with IM.
REFERENCE 2 (bases 1 to 2524)
AUTHORS Bettstetter,M., Dechant,S., Ruemmele,P., Grabowski,M., Keller,G., Holinski-Feder,E., Hartmann,A., Hofstaedter,F. and Dietmaier,W.
TITLE Distinction of hereditary nonpolyposis colorectal cancer and sporadic microsatellite-unstable colorectal cancer through quantification of MLH1 methylation by real-time PCR
JOURNAL Clin. Cancer Res. 13 (11), 3221-3228 (2007)
PUBMED 17545526
REMARK GeneRIF: quantitative MLH1 methylation analysis in MSI-H CRC is a valuable molecular tool to distinguish between HNPCC and sporadic MSI-H CRC
REFERENCE 3 (bases 1 to 2524)
AUTHORS Takahashi,M., Shimodaira,H., Andreutti-Zaugg,C., Iggo,R., Kolodner,R.D. and Ishioka,C.
TITLE Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays
JOURNAL Cancer Res. 67 (10), 4595-4604 (2007)
PUBMED 17510385
REMARK GeneRIF: The 101 MLH1 variants were examined for the dominant
Links
Gene
HomoloGene
Genome
Genome Project
Master
Full text in PMC
Probe
Protein
PubMed
PubMed (RefSeq)
Gene Genotype
GeneView in dbSNP
Taxonomy
Related Sequences
Map Viewer
OMIM
GEO Profiles
SNP
UniGene
UniSTS
LinkOut


NCBI
Entrez Nucleotide
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search CoreNucleotide for colon cancer AND nonpolyposis AND human[Organism] Go Clear Save Search
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Field: Title Limits: mRNA, RefSeq, RefSeq
Found 2 nucleotide sequences. CoreNucleotide [2]
Display Summary Show 20 Sort by Send to
All: 2 Bacteria: 0 RefSeq: 2 mRNA: 2
Items 1 - 2 of 2
One page.
1: NM_000249 Reports
Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), mRNA
gi|28559089|ref|NM_000249.2|[28559089]
2: NM_000251 Reports
Homo sapiens mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) (MSH2), mRNA
gi|4557760|ref|NM_000251.1|[4557760]
Links
Full text in PMC
Gene
Gene Genotype
GeneView in dbSNP
Genome
Genome Project
HomoloGene
Master
Probe
Protein
PubMed
PubMed (RefSeq)
Taxonomy
Related Sequences
Map Viewer
OMIM
GEO Profiles
SNP
UniGene
UniSTS
LinkOut
Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer

Taxonomy



All molecular databases

Homo sapiens

Taxonomy ID: 9606
Genbank common name: **human**
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 2 (Vertebrate Mitochondrial)
Other names:
common name: **man**

Lineage (full)
cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homo/Pan/Gorilla group; Homo

Entrez records

Database name	Subtree links	Direct links
Nucleotide	11,643,469	11,642,134
Protein	392,990	392,989
Structure	9,472	9,472
Genome Sequences	51	51
Genome Projects	1	1
Popset	20,878	20,878
SNP	11,870,024	11,870,024
3D Domains	35,848	35,848
Domains	19	19
GEO Datasets	3,525	3,525
GEO Expressions	10,649,715	10,649,715
UniGene	124,179	124,179
UniSTS	322,789	322,789
PubMed Central	3,586	3,586
Gene	38,624	38,624
HomoloGene	20,167	20,167
Taxonomy	2	1

Genome Information

See the NCBI Genome homepage
Go to NCBI genomic BLAST page for Homo sapiens

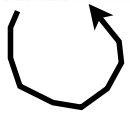
Genome view: 24 chromosomes

Names	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y

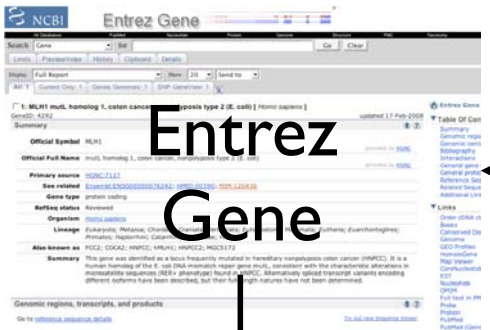
Goal: Find MLH1 homologs

- Tip: Use Entrez Gene as your hub to connect to everything else!

Protein

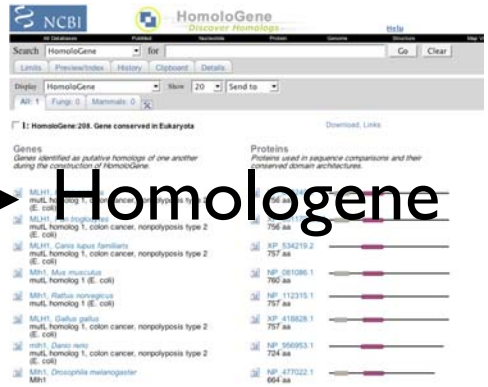


BLink



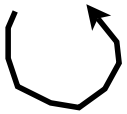
Entrez Gene

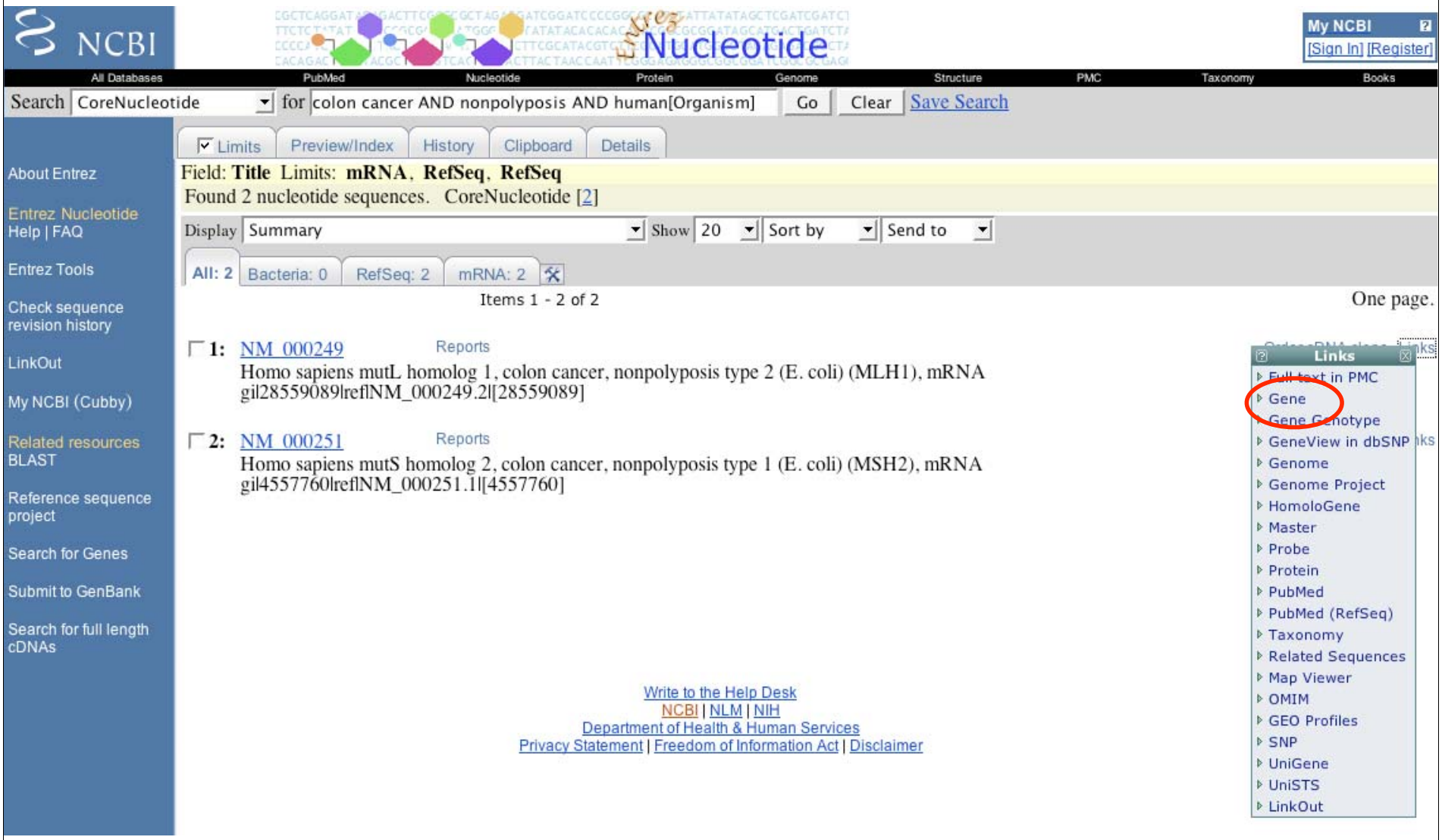
Other Entrez Databases



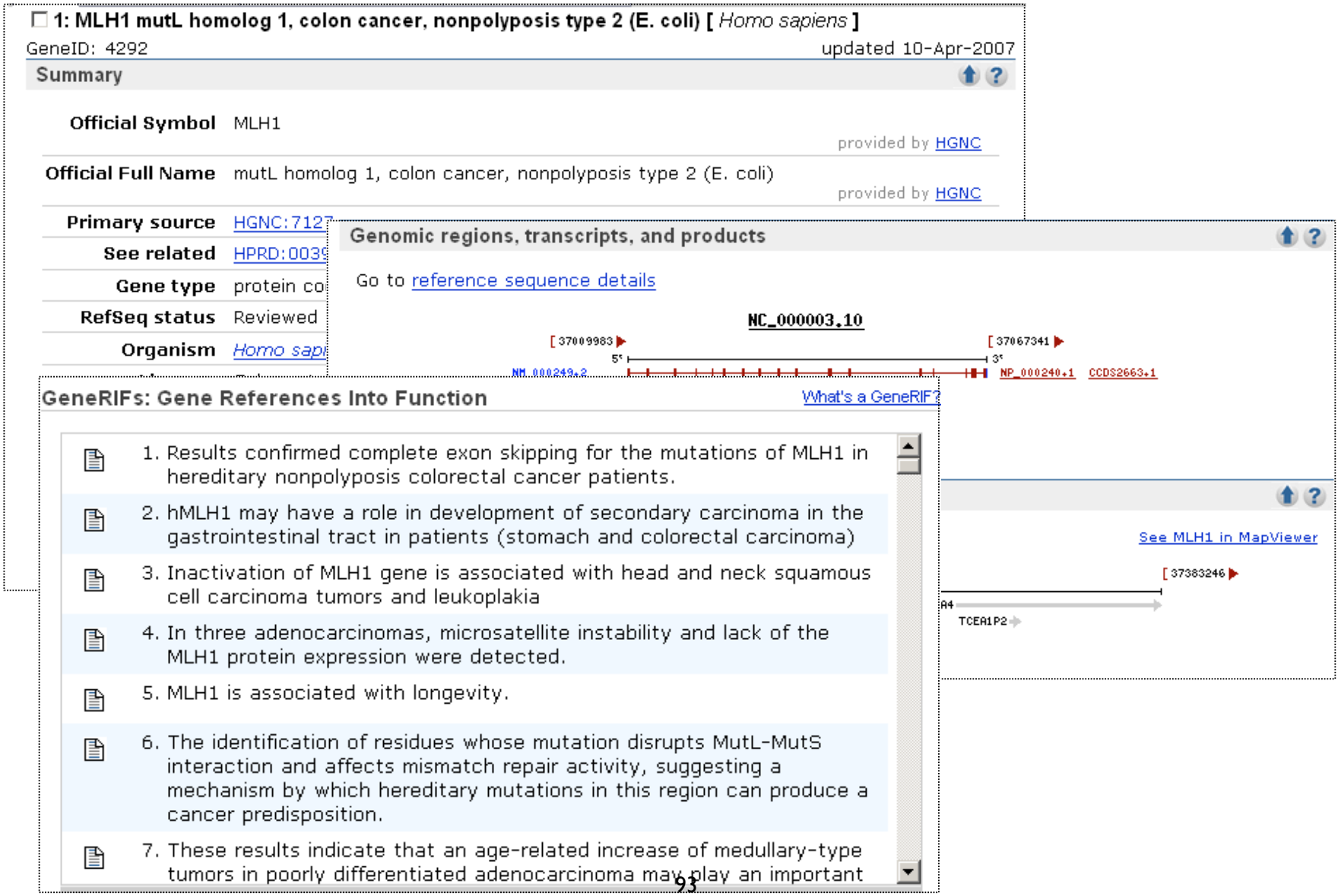
Homologene

Gene neighbors




NCBI
Nucleotide
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search CoreNucleotide for colon cancer AND nonpolyposis AND human[Organism] Go Clear Save Search
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Field: Title Limits: mRNA, RefSeq, RefSeq
Found 2 nucleotide sequences. CoreNucleotide [2]
Display Summary Show 20 Sort by Send to
All: 2 Bacteria: 0 RefSeq: 2 mRNA: 2
Items 1 - 2 of 2
One page.
1: NM_000249 Reports
Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), mRNA
gi|28559089|ref|NM_000249.2|[28559089]
2: NM_000251 Reports
Homo sapiens mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) (MSH2), mRNA
gi|4557760|ref|NM_000251.1|[4557760]
Links
Full text in PMC
Gene
Gene Genotype
GeneView in dbSNP
Genome
Genome Project
HomoloGene
Master
Probe
Protein
PubMed
PubMed (RefSeq)
Taxonomy
Related Sequences
Map Viewer
OMIM
GEO Profiles
SNP
UniGene
UniSTS
LinkOut
Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer

MLH1 Gene Record



Interactions + GO

Interactions

Description .....

Product	Interactant	Other Gene	Complex	Source	P
E2F1 interacts with the MLH1 promoter.					
NC_000003.9	NP_005216.1	E2F1		BIND	
E2F4 interacts with the MLH1 promoter region.					
NC_000003.9	NP_001941.2	E2F4		BIND	
NP_000240.1	NP_000048.1	BLM		HPRD	
MLH1 interacts with BLM.					
NP_000240.1	NP_000048.1	BLM		BIND	
NP_000240.1	NP_009225.1	BRCA1		HPRD	
The exonuclease HEX1 interacts with the mismatch repair protein hMLH1.					
NP_000240.1	NP_003677.3	EXO1		BIND	
The exonuclease hEXO1b interacts with the mismatch repair protein hMLH1.					
NP_000240.1	NP_006018.3	EXO1		BIND	
NP_000240.1	NP_569082.1	EXO1		HPRD	
NP_000240.1	NP_003916.1	MBD4		HPRD	
MLH1 and interacts with MED1.					
NP_000240.1	NP_003916.1	MBD4		BIND	
NP_000240.1	BAA92353.1	MLH3		HPRD	

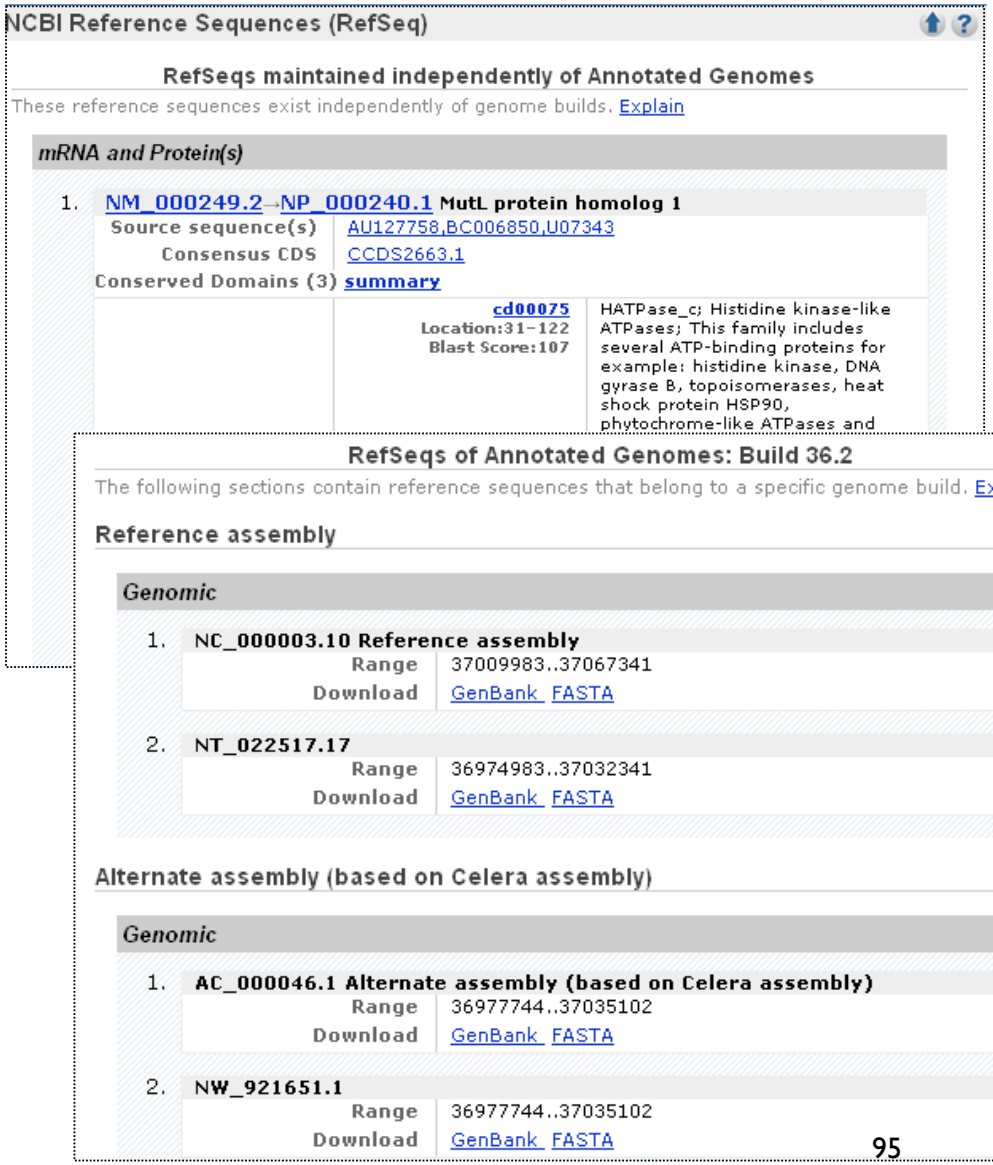
GeneOntology Provided by GOA

Function	Evidence	
ATP binding	IEA	
contributes_to MutSalpha complex binding	IDA	Pubmed
guanine/thymine mispair binding	IMP	Pubmed
guanine/thymine mispair binding	IEA	
mismatched DNA binding	IEA	
protein binding	IPI	Pubmed
contributes_to single-stranded DNA binding	IDA	Pubmed

Process	Evidence	
DNA damage response, signal transduction resulting in induction of apoptosis	IEA	
cell cycle	IEA	
male meiosis chromosome segregation	IEA	
meiotic recombination	IEA	
mismatch repair	IEA	
mismatch repair	TAS	Pubmed
negative regulation of mitotic recombination	IEA	
negative regulation of progression through cell cycle	IEA	

Component	Evidence	
MutLalpha complex	IEA	
condensed chromosome	IEA	
nucleus	IC	Pubmed
nucleus	IEA	
synaptonemal complex	IEA	

Sequences



NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. Explain

mRNA and Protein(s)

1. NM_000249.2→NP_000240.1 MutL protein homolog 1

Source sequence(s) AU127758,BC006850,U07343

Consensus CDS CCDS2663.1

Conserved Domains (3) summary

cd00075 Location:31–122 Blast Score:107 — HATPase_c; Histidine kinase-like ATPases; This family includes several ATP-binding proteins for example: histidine kinase, DNA gyrase B, topoisomerases, heat shock protein HSP90, phytochrome-like ATPases and

RefSeqs of Annotated Genomes: Build 36.2

The following sections contain reference sequences that belong to a specific genome build. Explain

Reference assembly

Genomic

1. NC_000003.10 Reference assembly
   - Range 37009983..37067341
   - Download GenBank FASTA
2. NT_022517.17
   - Range 36974983..37032341
   - Download GenBank FASTA

Alternate assembly (based on Celera assembly)

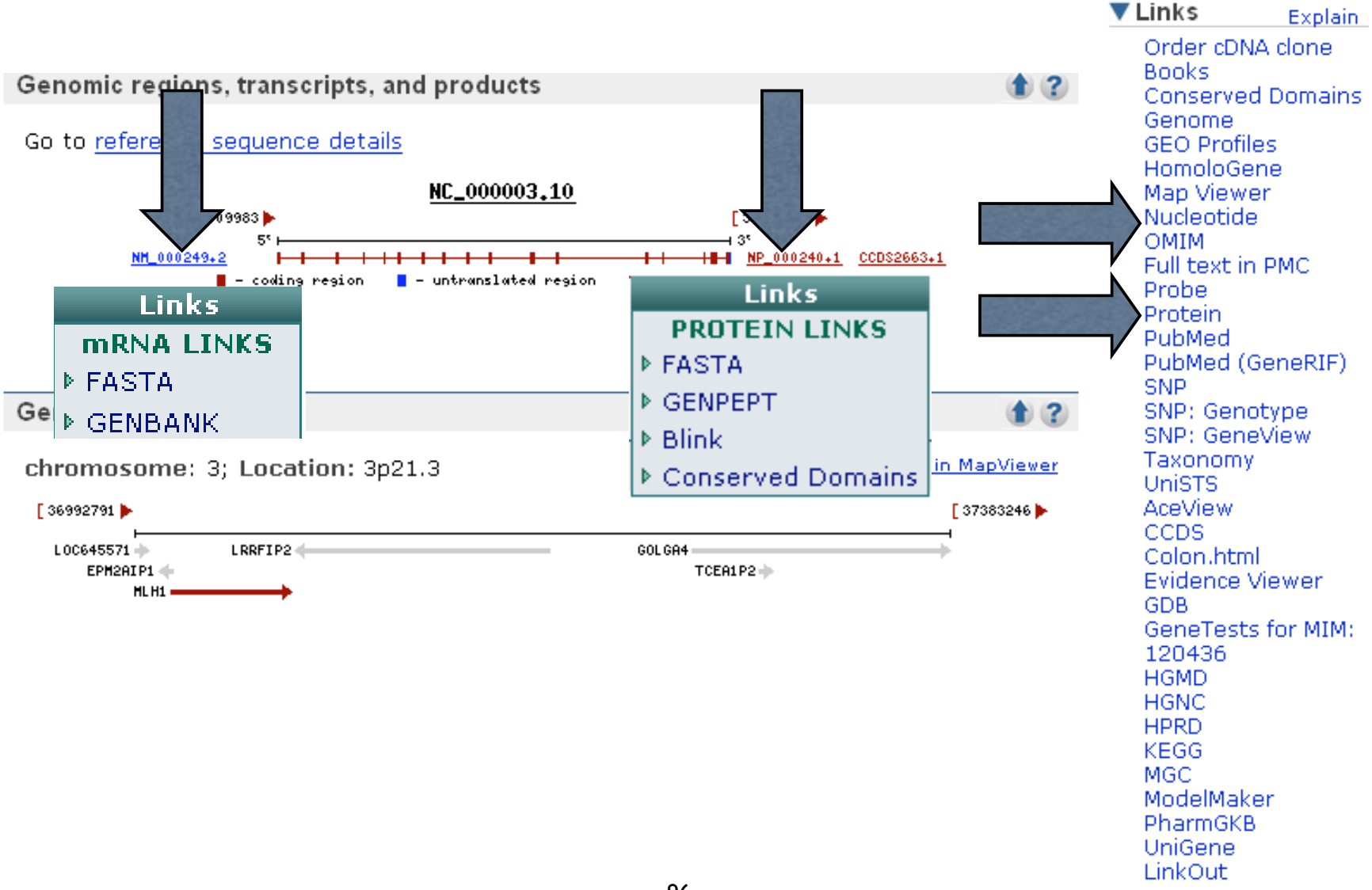
Genomic

1. AC_000046.1 Alternate assembly (based on Celera assembly)
   - Range 36977744..37035102
   - Download GenBank FASTA
2. NW_921651.1
   - Range 36977744..37035102
   - Download GenBank FASTA

Related Sequences

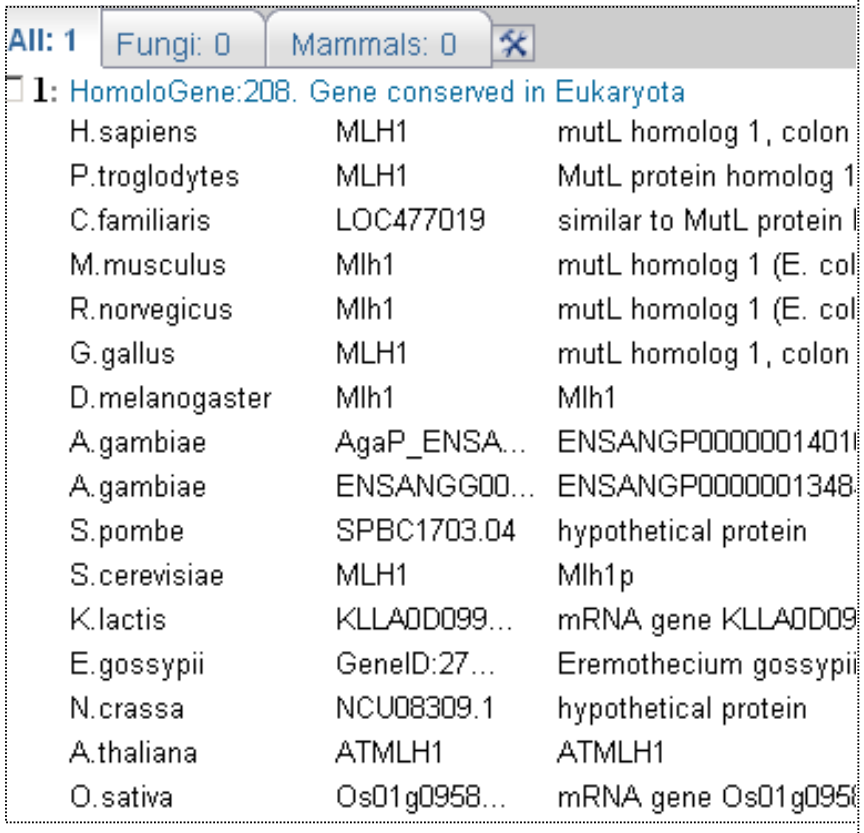
Nucleotide		Protein
Genomic	AC006583.31 (69181..100370, complement)	None
Genomic	AC011816.17 (143145..169313)	None
Genomic	AY217549.1	AAO22994.1
Genomic	AY344475.1	AAQ23474.1
Genomic	AY706914.1	AAU21566.1
Genomic	CH471055.1	EAW64483.1
		EAW64484.1
		EAW64485.1
Genomic	U17839.1	AAA85687.1
Genomic	U17840.1	AAA85687.1
Genomic	U17841.1	AAA85687.1
Genomic	U17842.1	AAA85687.1
Genomic	U17843.1	AAA85687.1
Genomic	U17844.1	AAA85687.1
Genomic	U17845.1	AAA85687.1
Genomic	U17846.1	AAA85687.1
Genomic	U17847.1	AAA85687.1
Genomic	U17848.1	AAA85687.1
Genomic	U17849.1	AAA85687.1
Genomic	U17850.1	AAA85687.1
Genomic	U17851.1	AAA85687.1
Genomic	U17852.1	AAA85687.1
Genomic	U17853.1	AAA85687.1
Genomic	U17854.1	AAA85687.1
Genomic	U17855.1	AAA85687.1
Genomic	U17856.1	AAA85687.1
Genomic	U17857.1	AAA85687.1
Genomic	U40978.1	AAA82079.1
mRNA	AB209848.1	BAD93085.1
mRNA	AF001359.1	AAB58936.1
mRNA	AK222810.1	BAD96530.1
mRNA	AU127758.1	None
mRNA	AY517558.1	AAT44531.1
mRNA	BC006850.1	AAH06850.1
mRNA	BX648844.1	None
mRNA	CR609870.1	None
mRNA	CR617505.1	None
mRNA	DQ648888.1	ABG49483.1
mRNA	DQ648889.1	ABG49484.1
mRNA	DQ648890.1	ABG49485.1
mRNA	DQ648891.1	ABG49486.1
mRNA	DQ648892.1	ABG49487.1
mRNA	DQ648893.1	ABG49488.1
mRNA	S77856.1	AAB34135.1
mRNA	U07343.1	AAC50285.1
mRNA	U07418.1	AAA17374.1

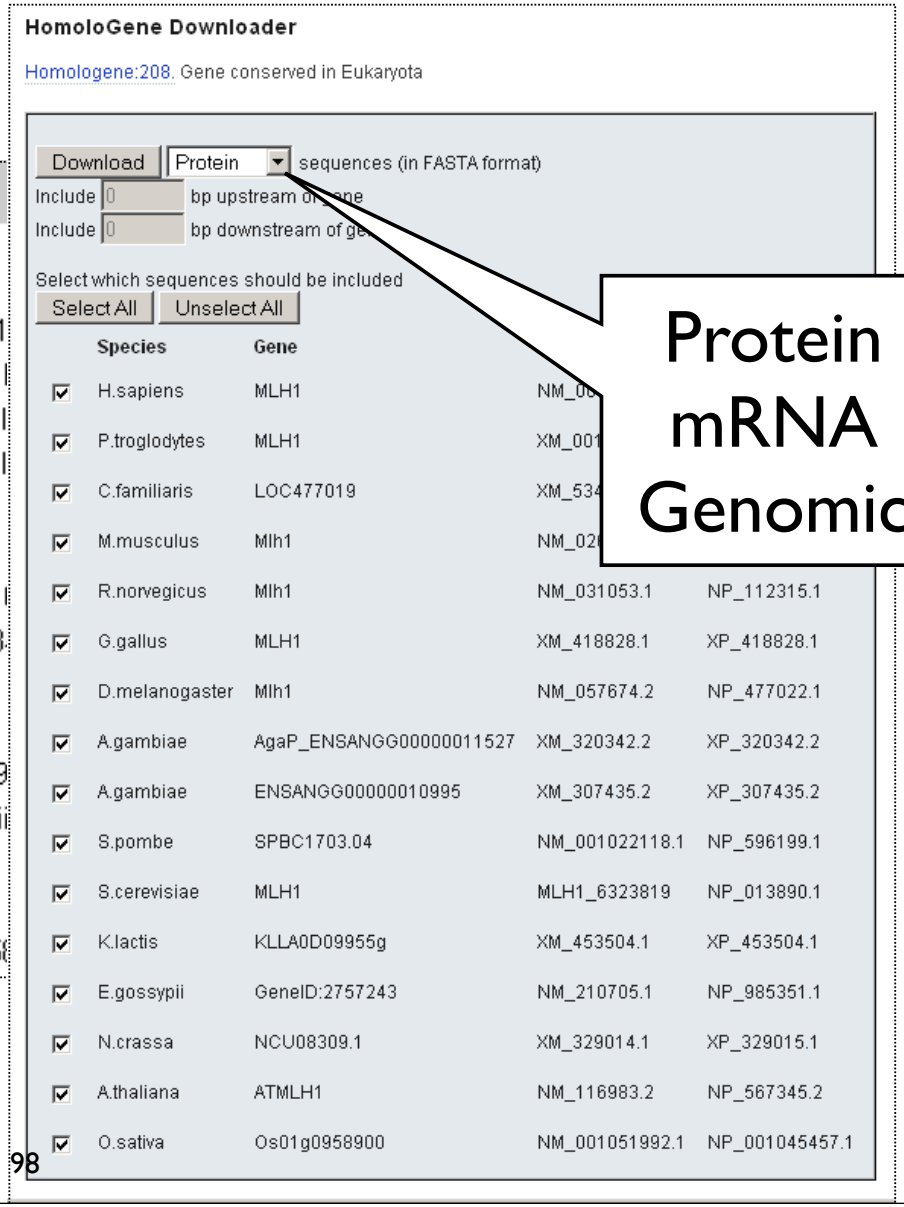
MLH1: Sequence Links




NCBI
Entrez Gene
My NCBI
Welcome joannealisonfox. [Sign Out]
All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books OMIM
Search Gene for Go Clear
Limits Preview/Index History Clipboard Details
Display Full Report Show 20 Send to
All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1
1: MLH1 mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) [Homo sapiens]
GeneID: 4292 updated 16-Sep-2007
Summary
Official Symbol MLH1 provided by HGNC
Official Full Name mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) provided by HGNC
Primary source HGNC:7127
See related Ensembl:ENSG00000076242; HPRD:00390; MIM:120436
Gene type protein coding
RefSeq status Reviewed
Organism Homo sapiens
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as FCC2; COCA2; HNPCC; hMLH1; HNPCC2; MGC5172
Summary This gene was identified as a locus frequently mutated in hereditary nonpolyposis colon cancer (HNPCC). It is a human homolog of the E. coli DNA mismatch repair gene mutL, consistent with the characteristic alterations in microsatellite sequences (RER+ phenotype) found in HNPCC. Alternatively spliced transcript variants encoding different isoforms have been described, but their full-length natures have not been determined.
Genomic regions, transcripts, and products
Go to reference sequence details
NC_000003.10
Entrez Gene Home
Table Of Contents
Summary
Genomic regions, transcripts...
Genomic context
Bibliography
Interactions
General gene information
General protein information
Reference Sequences
Related Sequences
Additional Links
Links Explain
Order cDNA clone
Books
Conserved Domains
Genome
GEO Profiles
HomoloGene
Map Viewer
CoreNucleotide
EST
Nucleotide
OMIM
Full text in PMC
Probe
Protein
PubMed
PubMed (GeneRIF)
SNP
SNP: Genotype

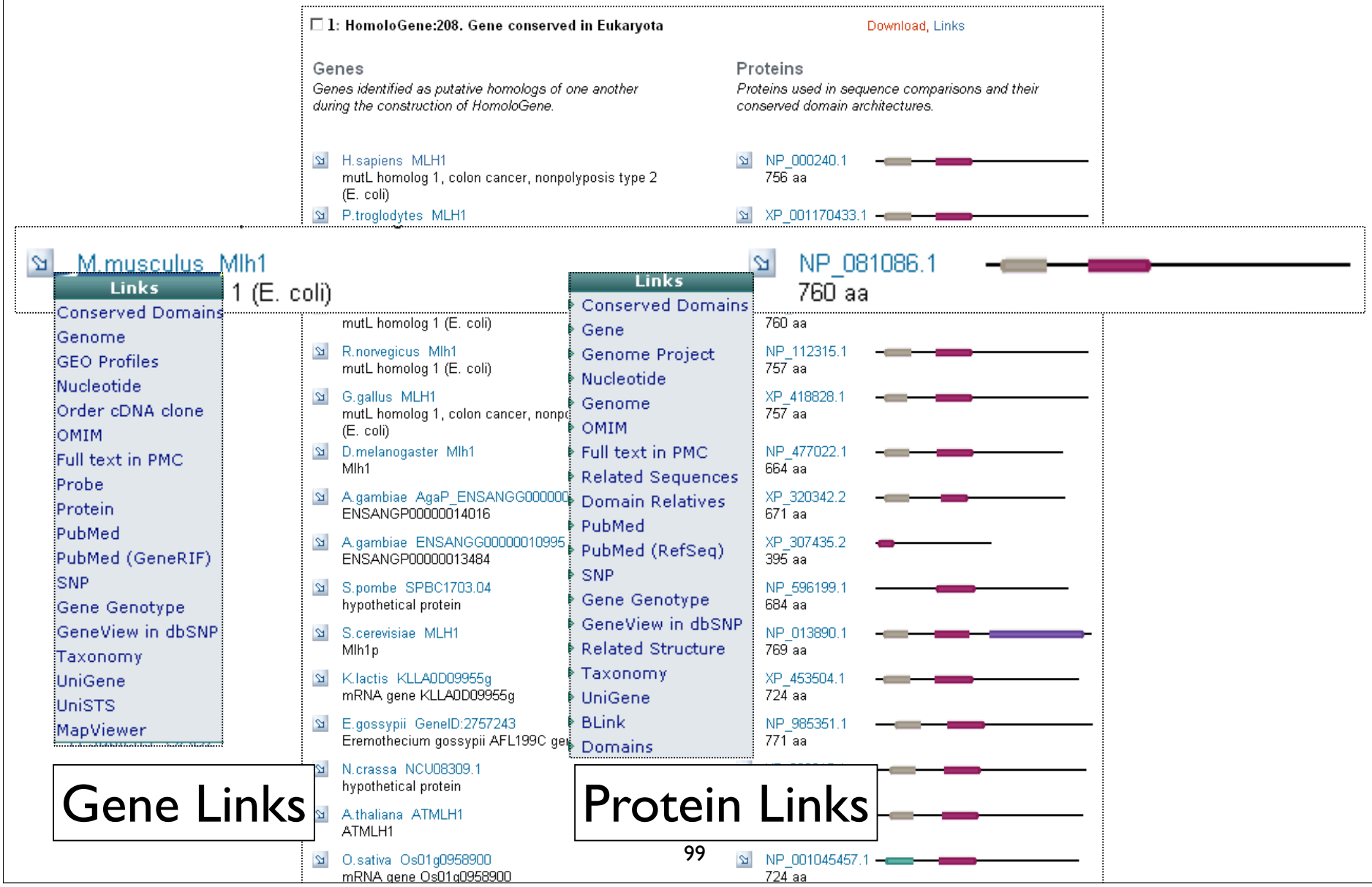
Finding Homologs:



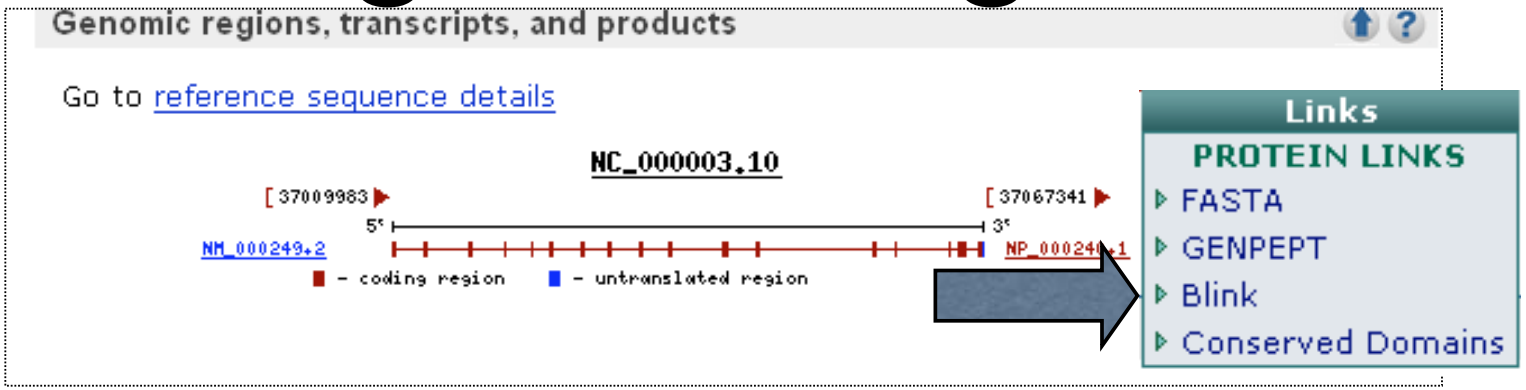


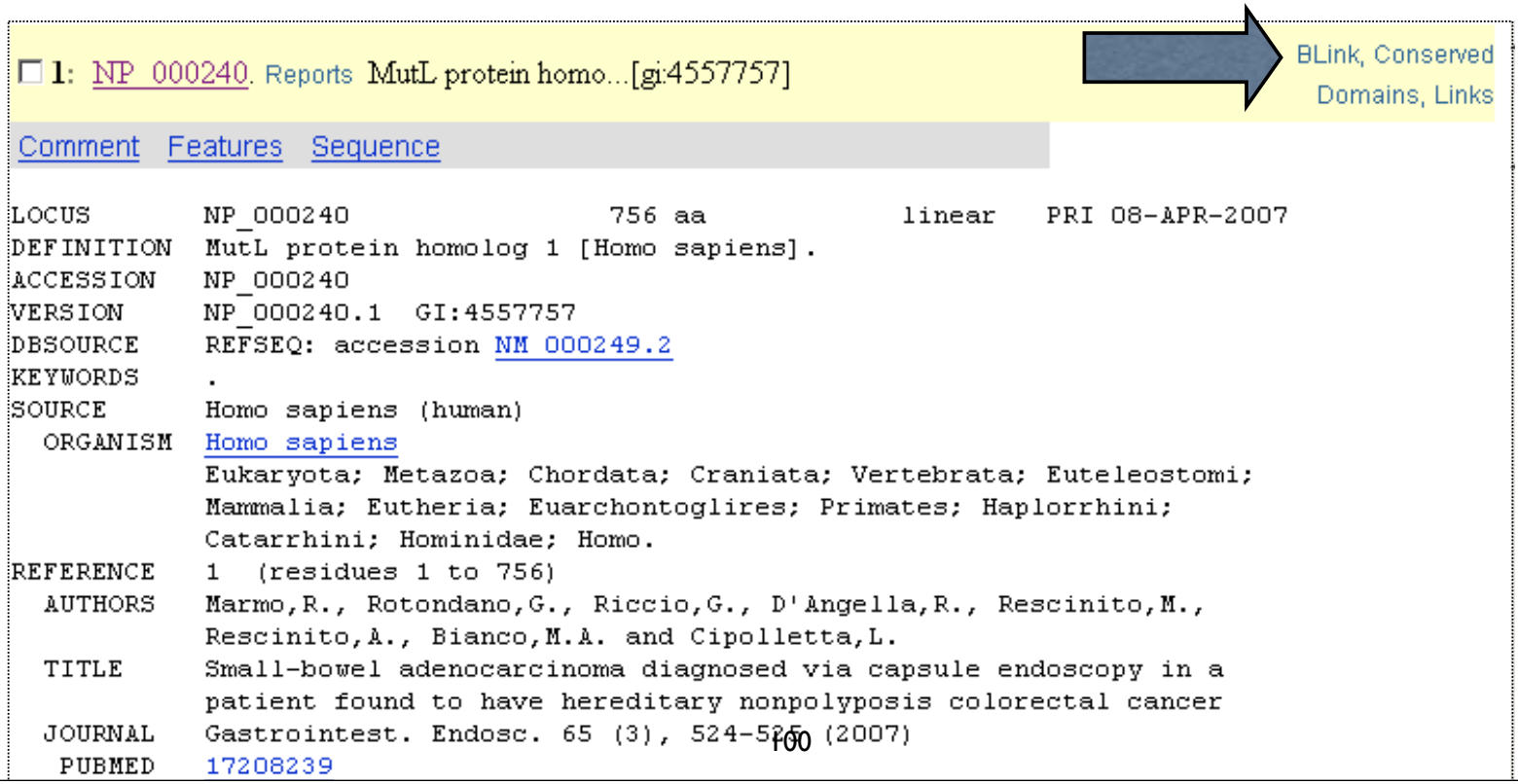
Protein
mRNA
Genomic

HomoloGene Cluster

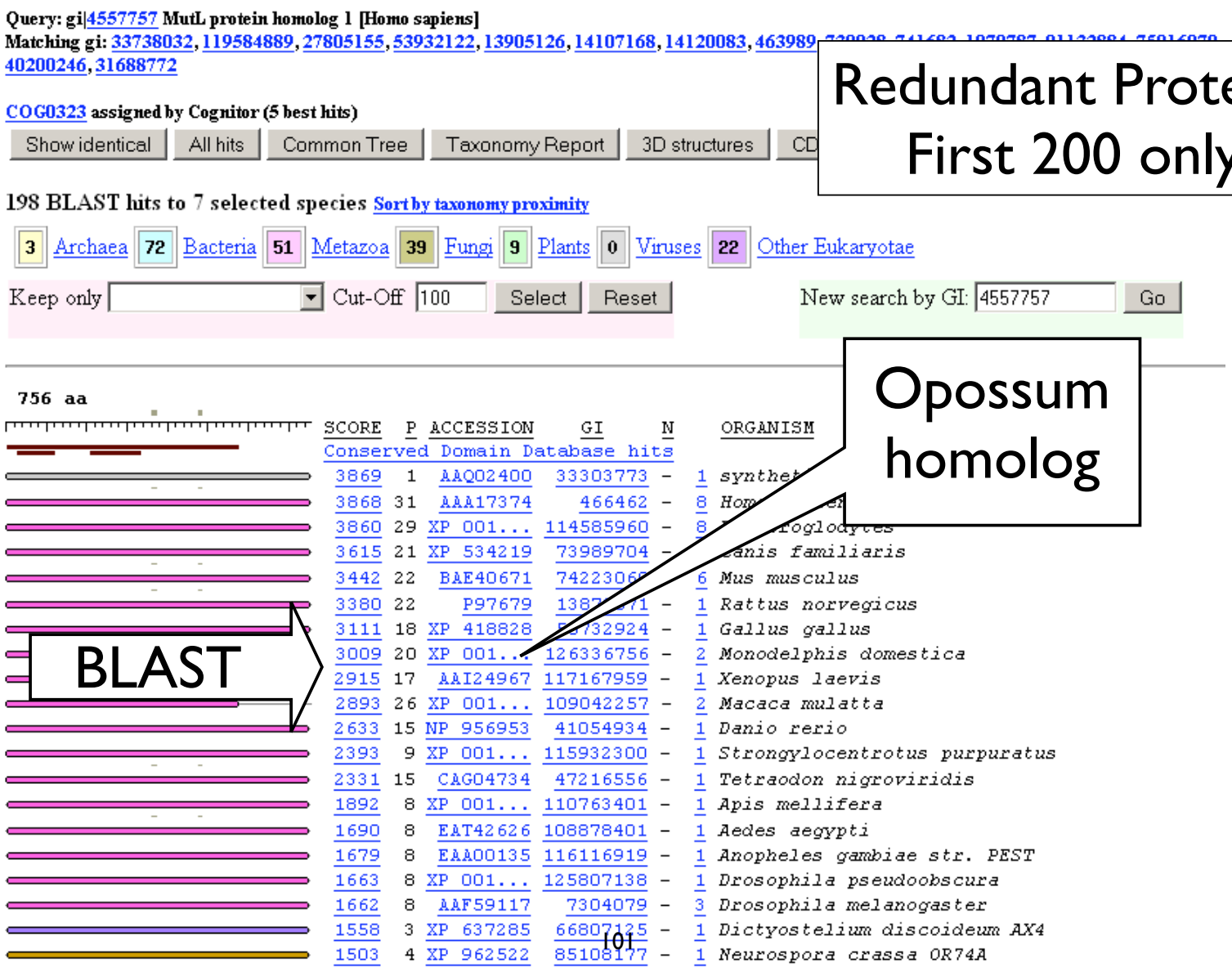


Finding Homologs 2: BLink





BLink: BLAST Link (Best Hits)

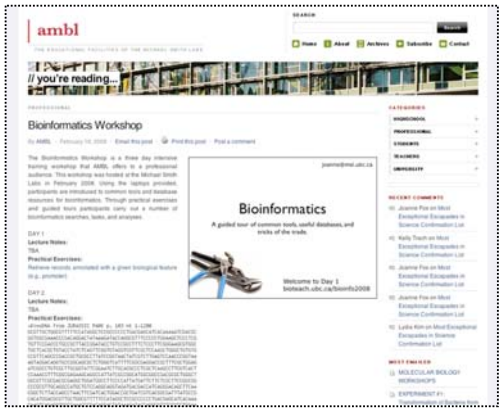


Sequence Databases

PRACTICAL EXERCISES: Navigating Links, Retrieving Data with Entrez, and Searching PubMed

I am studying the regulaton of cancer genes and would like to retrieve all human sequence records associated with cancer that contain a promoter region.

navigate to:
bioteach.ubc.ca/bioinfo2008



Let's compare our results

Follow link to practical exercise page at the NCBI where you'll find step-by-step instructions

Use the preview tab and feature keys

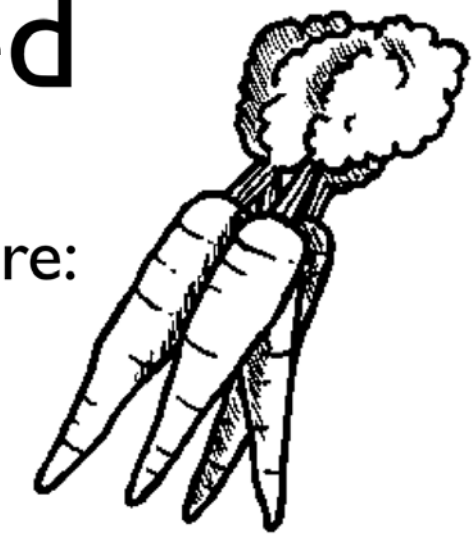
Strategy #1: search nt

Strategy #2: search entrez gene

Check your History

Search	Most Recent Queries	Result
#5	Search #3 NOT #1 (unique hits from Approach B: Entrez Gene to CoreNucleotide)	286
#4	Search #1 NOT #3 (unique hits from Approach A: straight to Entrez CoreNucleotide search)	183
#3	Search #2 AND promoter[Feature key] (limit Approach B search to records with promoter annotated)	329
#2	CoreNucleotide Links for Gene (Search human [Organism] AND cancer[Text Word] AND gene_nucleotide[Filter]) (Approach B: Entrez gene follow link to CoreNucleotide)	53844
#1	Search human[Organism] AND cancer[Text Word] AND promoter[Feature key] (Approach A: Entrez CoreNucleotide search)	226

Searching PubMed



- How many papers in PubMed are there:
  - about cancer?
  - about carrots?
- Using Entrez PubMed, can you see if there is any scientific links between carrots and cancer?
- How many papers are there about "carrots AND cancer"?
- What is the active chemical substance in carrots that may play a role in cancers?

NCBI

PubMed www.pubmed.gov

A service of the National Library of Medicine and the National Institutes of Health

My NCBI [Sign In] [Register]

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books

Search PubMed for cancer AND carrots Go Clear Save Search

Limits | Preview/Index | History | Clipboard | Details

About Entrez

Text Version

Entrez PubMed
Overview
Help | FAQ
Tutorials
New/Noteworthy
E-Utilities

PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Special Queries
LinkOut
My NCBI

Related Resources
Order Documents
NLM Mobile
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

Display Summary Show 20 Sort By Send to

All: 115 Review: 14

Items 1 - 20 of 115 Page 1 of 6 Next

1: Collins AR, Gaivao I. Related Articles, Links

DNA base excision repair as a biomarker in molecular epidemiology studies.
Mol Aspects Med. 2007 Jun 2; [Epub ahead of print]
PMID: 17659329 [PubMed - as supplied by publisher]

2: Young JF, Duthie SJ, Milne L, Christensen LP, Duthie GG, Bestwick CS. Related Articles, Links

Biphasic effect of falcarinol on caco-2 cell proliferation, DNA damage, and apoptosis.
J Agric Food Chem. 2007 Feb 7;55(3):618-23.
PMID: 17263451 [PubMed - indexed for MEDLINE]

3: Galeone C, Negri E, Pelucchi C, La Vecchia C, Bosetti C, Hu J. Related Articles, Links

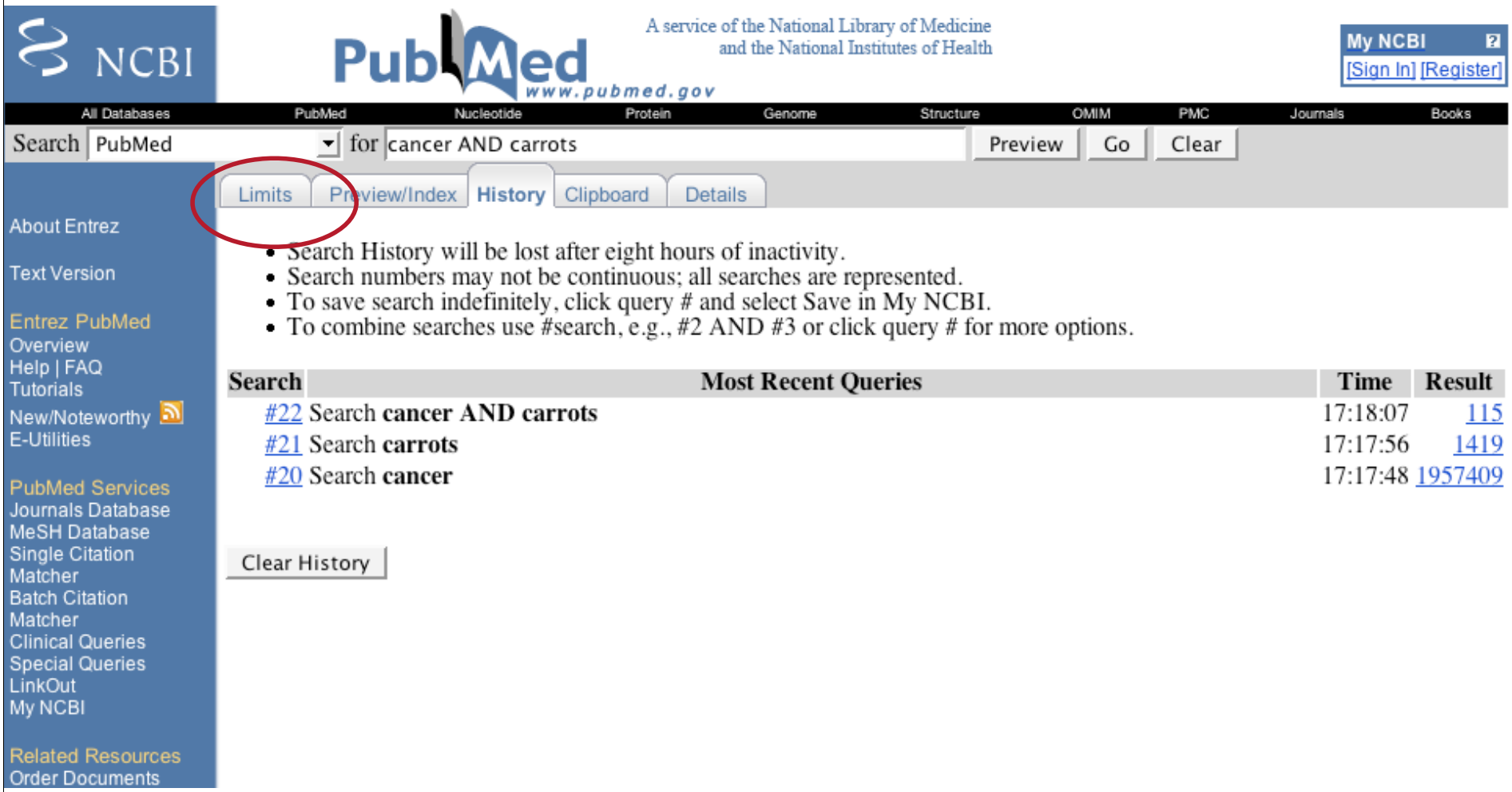
Dietary intake of fruit and vegetable and lung cancer risk: a case-control study in Harbin, northeast China.
Ann Oncol. 2007 Feb;18(2):388-92. Epub 2006 Oct 23.
PMID: 17060488 [PubMed - indexed for MEDLINE]

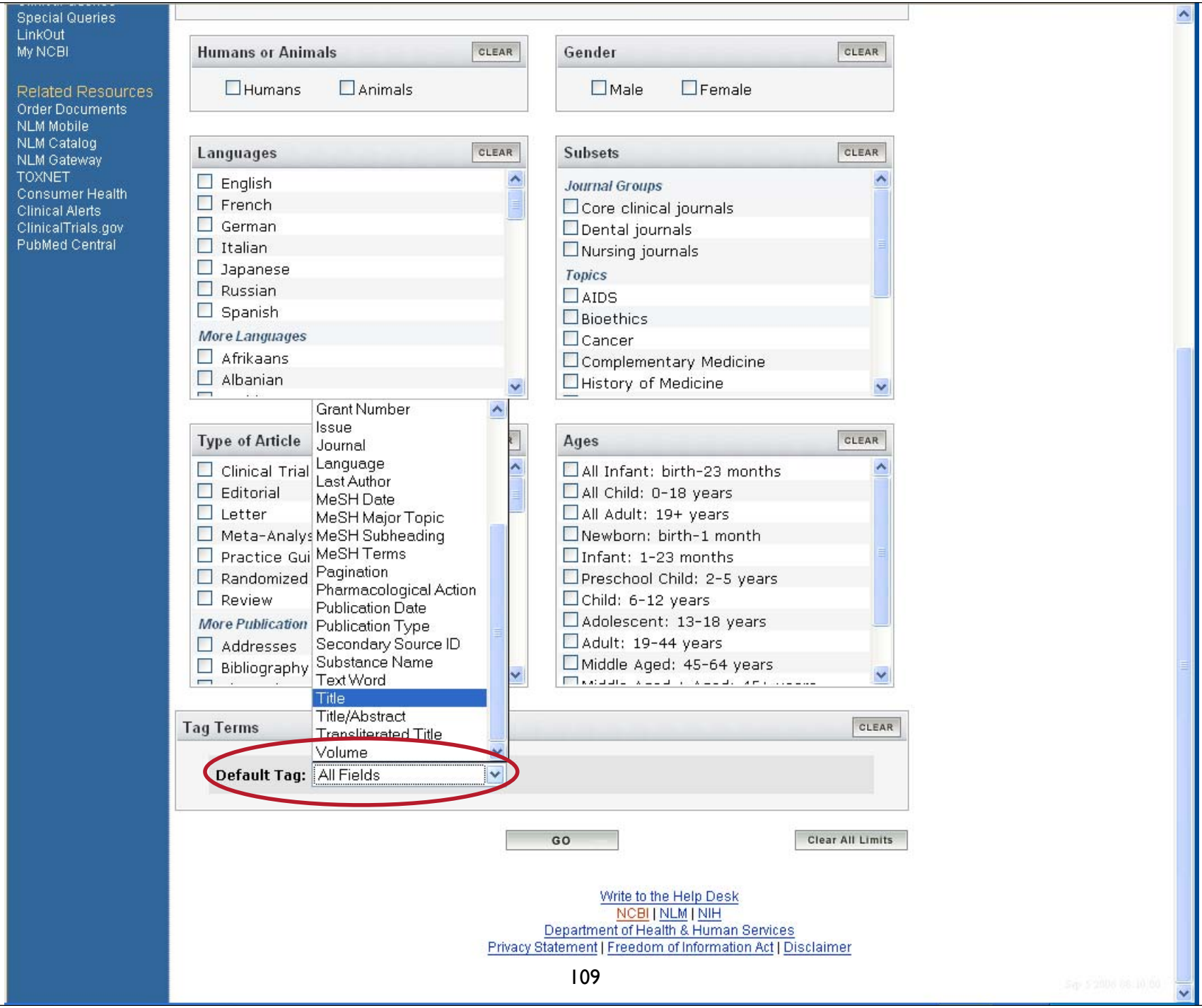
4: Roumanas ED, Garrett N, Blackwell KE, Freymiller E, Abemayor E, Wong WK, Beumer J 3rd, Fueki K, Fueki W, Kapur KK. Related Articles, Links

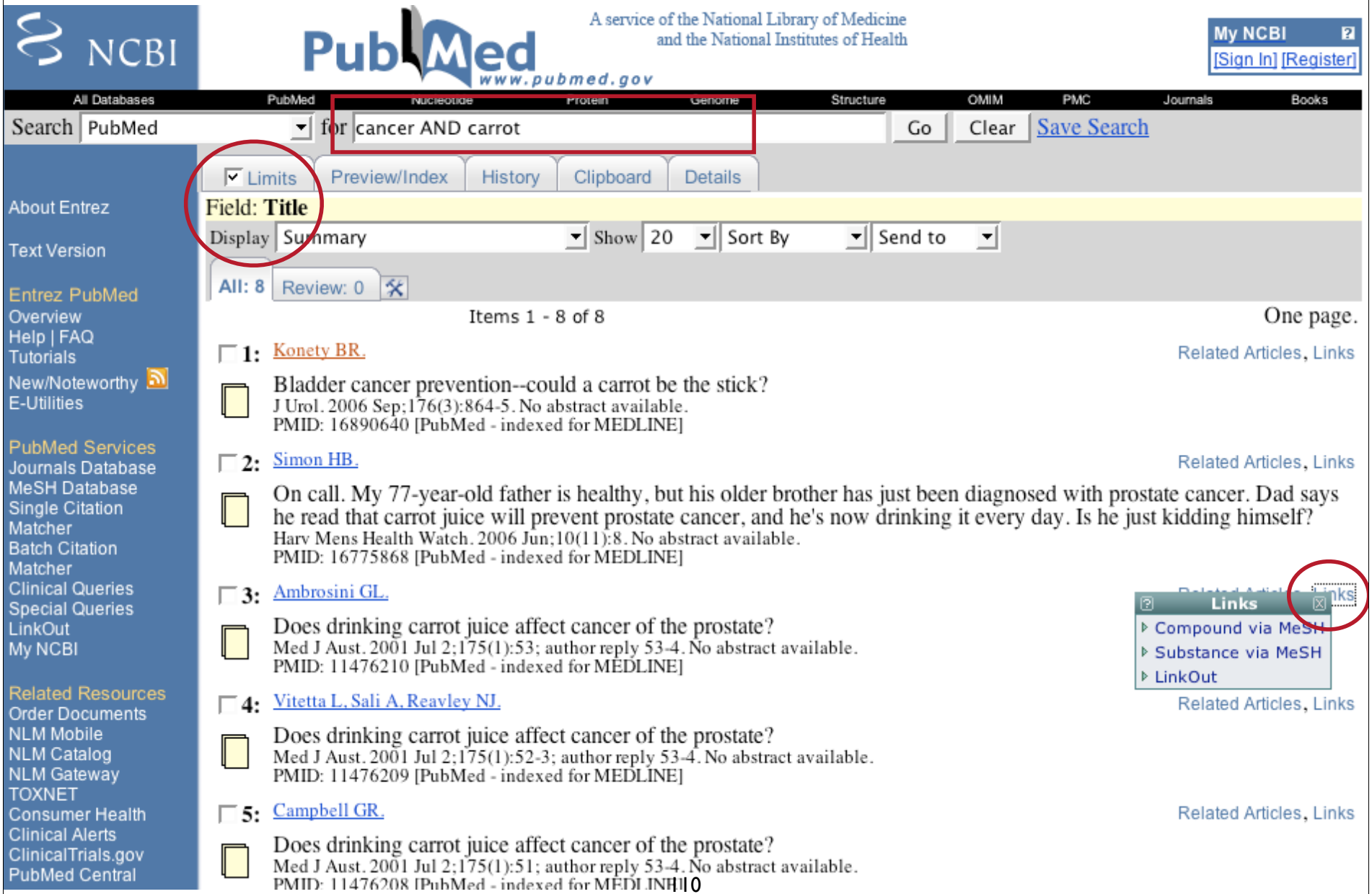
Masticatory and swallowing threshold performances with conventional and implant-supported prostheses after mandibular fibula free-flap reconstruction.
J Prosthet Dent. 2006 Oct;96(4):289-97.
PMID: 17052474 [PubMed - indexed for MEDLINE]

5: Simon HB. Related Articles, Links

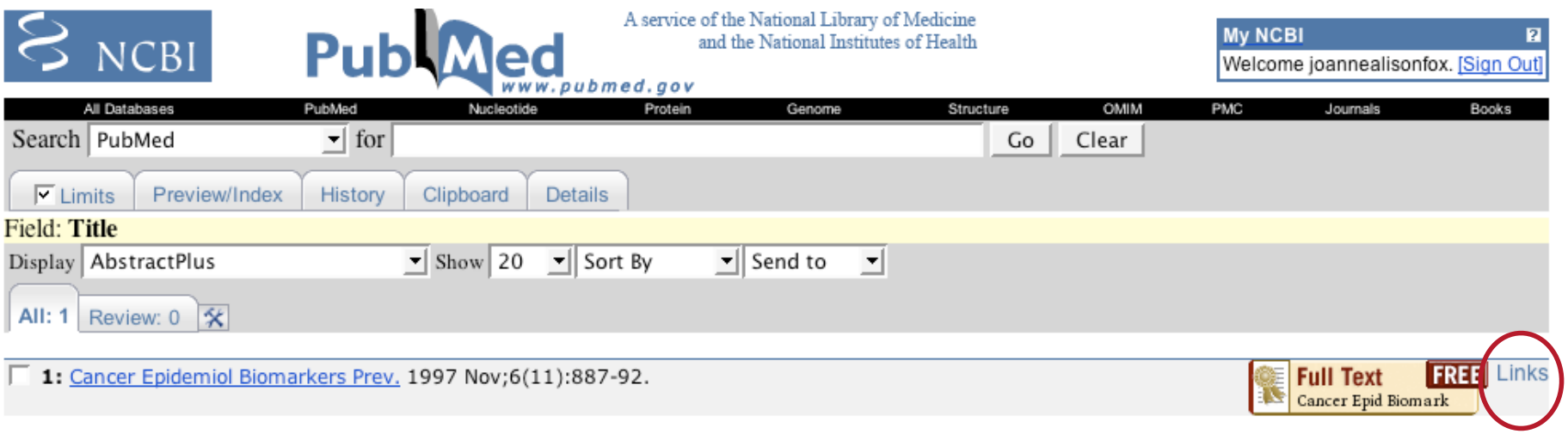
On call. My 77-year-old father is healthy, but his older brother has just been diagnosed with prostate cancer. Dad says he read that carrot juice will prevent prostate cancer, and he's now drinking it every day. Is he just kidding himself?
Harv Mens Health Watch. 2006 Jun;10(11):8. No abstract available.
PMID: 16775868 [PubMed - indexed for MEDLINE]


NCBI
PubMed
www.pubmed.gov
A service of the National Library of Medicine
and the National Institutes of Health
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for cancer AND carrots Preview Go Clear
Limits Preview/Index History Clipboard Details
About Entrez
Text Version
Entrez PubMed
Overview
Help | FAQ
Tutorials
New/Noteworthy
E-Utilities
PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Special Queries
LinkOut
My NCBI
Related Resources
Order Documents
Search History will be lost after eight hours of inactivity.
Search numbers may not be continuous; all searches are represented.
To save search indefinitely, click query # and select Save in My NCBI.
To combine searches use #search, e.g., #2 AND #3 or click query # for more options.
Search Most Recent Queries Time Result
#22 Search cancer AND carrots 17:18:07 115
#21 Search carrots 17:17:56 1419
#20 Search cancer 17:17:48 1957409
Clear History


Special Queries
LinkOut
My NCBI
Related Resources
Order Documents
NLM Mobile
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central
Humans or Animals
CLEAR
Humans
Animals
Gender
CLEAR
Male
Female
Languages
CLEAR
English
French
German
Italian
Japanese
Russian
Spanish
More Languages
Afrikaans
Albanian
Subsets
CLEAR
Journal Groups
Core clinical journals
Dental journals
Nursing journals
Topics
AIDS
Bioethics
Cancer
Complementary Medicine
History of Medicine
Type of Article
Clinical Trial
Editorial
Letter
Meta-Analys
Practice Gui
Randomized
Review
More Publication
Addresses
Bibliography
Grant Number
Issue
Journal
Language
Last Author
MeSH Date
MeSH Major Topic
MeSH Subheading
MeSH Terms
Pagination
Pharmacological Action
Publication Date
Publication Type
Secondary Source ID
Substance Name
Text Word
Title
Title/Abstract
Transliterated Title
Volume
Ages
CLEAR
All Infant: birth-23 months
All Child: 0-18 years
All Adult: 19+ years
Newborn: birth-1 month
Infant: 1-23 months
Preschool Child: 2-5 years
Child: 6-12 years
Adolescent: 13-18 years
Adult: 19-44 years
Middle Aged: 45-64 years
Tag Terms
CLEAR
Default Tag:
All Fields
GO
Clear All Limits
Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer


NCBI
PubMed
www.pubmed.gov
A service of the National Library of Medicine
and the National Institutes of Health
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for cancer AND carrot Go Clear Save Search
Limits Preview/Index History Clipboard Details
About Entrez
Text Version
Entrez PubMed
Overview
Help | FAQ
Tutorials
New/Noteworthy
E-Utilities
PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Special Queries
LinkOut
My NCBI
Related Resources
Order Documents
NLM Mobile
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central
Field: Title
Display Summary Show 20 Sort By Send to
All: 8 Review: 0
Items 1 - 8 of 8
One page.
1: Konety BR.
Related Articles, Links
Bladder cancer prevention--could a carrot be the stick?
J Urol. 2006 Sep;176(3):864-5. No abstract available.
PMID: 16890640 [PubMed - indexed for MEDLINE]
2: Simon HB.
Related Articles, Links
On call. My 77-year-old father is healthy, but his older brother has just been diagnosed with prostate cancer. Dad says he read that carrot juice will prevent prostate cancer, and he's now drinking it every day. Is he just kidding himself?
Harv Mens Health Watch. 2006 Jun;10(11):8. No abstract available.
PMID: 16775868 [PubMed - indexed for MEDLINE]
3: Ambrosini GL.
Related Articles, Links
Does drinking carrot juice affect cancer of the prostate?
Med J Aust. 2001 Jul 2;175(1):53; author reply 53-4. No abstract available.
PMID: 11476210 [PubMed - indexed for MEDLINE]
Links
Compound via MeSH
Substance via MeSH
LinkOut
4: Vitetta L, Sali A, Reavley NJ.
Related Articles, Links
Does drinking carrot juice affect cancer of the prostate?
Med J Aust. 2001 Jul 2;175(1):52-3; author reply 53-4. No abstract available.
PMID: 11476209 [PubMed - indexed for MEDLINE]
5: Campbell GR.
Related Articles, Links
Does drinking carrot juice affect cancer of the prostate?
Med J Aust. 2001 Jul 2;175(1):51; author reply 53-4. No abstract available.
PMID: 11476208 [PubMed - indexed for MEDLINE]


NCBI
PubMed
www.pubmed.gov
A service of the National Library of Medicine
and the National Institutes of Health
My NCBI
Welcome joannealisonfox. [Sign Out]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for cancer and carrots Go Clear Save Search
Limits Preview/Index History Clipboard Details
About Entrez
Text Version
Entrez PubMed
Overview
Help | FAQ
Tutorials
New/Noteworthy
E-Utilities
PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Special Queries
LinkOut
My NCBI
Related Resources
Order Documents
Field: Title
Display Summary Show 20 Sort By Send to
All: 3 Review: 0
Items 1 - 3 of 3
One page.
1: Longnecker MP, Newcomb PA, Mittendorf R, Greenberg ER, Willett WC.
Related Articles, Links
Intake of carrots, spinach, and supplements containing vitamin A in relation to risk of breast cancer.
Cancer Epidemiol Biomarkers Prev. 1997 Nov;6(11):887-92.
PMID: 9367061 [PubMed - indexed for MEDLINE]
2: Jacobsen BK.
Related Articles, Links
[Vegetables and prevention of cancer. Carrots are still good for you]
Tidsskr Nor Laegeforen. 1988 Oct 30;108(30):2744-6. Norwegian. No abstract available.
PMID: 3206486 [PubMed - indexed for MEDLINE]
3: Pisani P, Berrino F, Macaluso M, Pastorino U, Crosignani P, Baldasseroni A.
Related Articles, Links
Carrots, green vegetables and lung cancer: a case-control study.
Int J Epidemiol. 1986 Dec;15(4):463-8.
PMID: 3818153 [PubMed - indexed for MEDLINE]


NCBI
PubMed
www.pubmed.gov
A service of the National Library of Medicine and the National Institutes of Health
My NCBI
Welcome joannealisonfox. [Sign Out]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search PubMed for Go Clear
Limits Preview/Index History Clipboard Details
Field: Title
Display AbstractPlus Show 20 Sort By Send to
All: 1 Review: 0
1: Cancer Epidemiol Biomarkers Prev. 1997 Nov;6(11):887-92.
Full Text FREE Cancer Epid Biomark
Links
Intake of carrots, spinach, and supplements containing vitamin A in relation to risk of breast cancer.
Longnecker MP, Newcomb PA, Mittendorf R, Greenberg ER, Willett WC.
Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA.
Intake of fruits, vegetables, vitamin A, and related compounds are associated with a decreased risk of breast cancer in some studies, but additional data are needed. To estimate intake of beta-carotene and vitamin A, the authors included nine questions on food and supplement use in a population-based case-control study of breast cancer risk conducted in Maine, Massachusetts, New Hampshire, and Wisconsin in 1988-1991. Multivariate-adjusted models were fit to data for 3543 cases and 9406 controls. Eating carrots or spinach more than twice weekly, compared with no intake, was associated with an odds ratio of 0.56 (95% confidence interval 0.34-0.91). Estimated intake of preformed vitamin A from all evaluated foods and supplements showed no trend or monotonic decrease in risk across categories of intake. These data do not allow us to distinguish among several potential explanations for the protective association observed between intake of carrots and spinach and risk of breast cancer. The findings are, however, consistent with a diet rich in these foods having a modest protective effect.
PMID: 9367061 [PubMed - indexed for MEDLINE]
Related Links
Vitamins C and E, retinol, beta-carotene and dietary fibre in relation to breast cancer risk [Br J Cancer. 1997]
Population attributable risk for breast cancer: diet, nutrition, and physical exercise [J Natl Cancer Inst. 1998]
Fruits, vegetables, and micronutrients in relation to breast cancer m [Cancer Epidemiol Biomarkers Prev. 2004]
Intake of dietary fat and vitamin in relation to breast cancer risk in Korean women: a [J Korean Med Sci. 2003]
Dietary carotenoids and vitamins A, C, and E and risk of breast cancer. [J Natl Cancer Inst. 1999]
See all Related Articles...
Display AbstractPlus Show 20 Sort By Send to

NCBI

PubChem

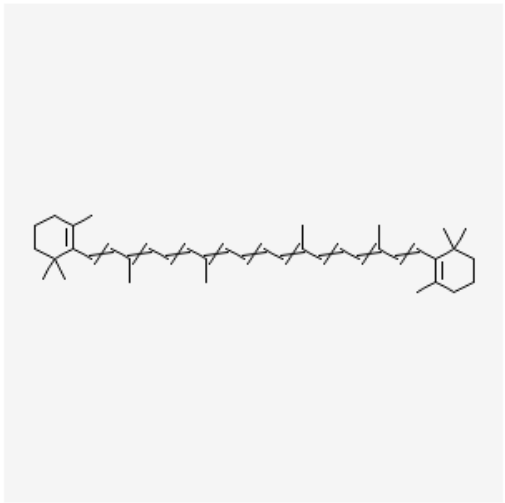
National Library of Medicine NLM

HOME | SEARCH | SITE MAP | PubMed | Entrez | Structure | GenBank | PubChem | Help

Search PubChem Substance GO

Substance Summary:

Compound Displayed PubChem



SID: 4266322

CID: 573

Related Substances:

Same: 11 Links

Same, Connectivity: 12 Links

Structure Search

Source: LipidMAPS (LMPR01070001)

MeSH | Synonyms | Properties | Descriptors | Exports

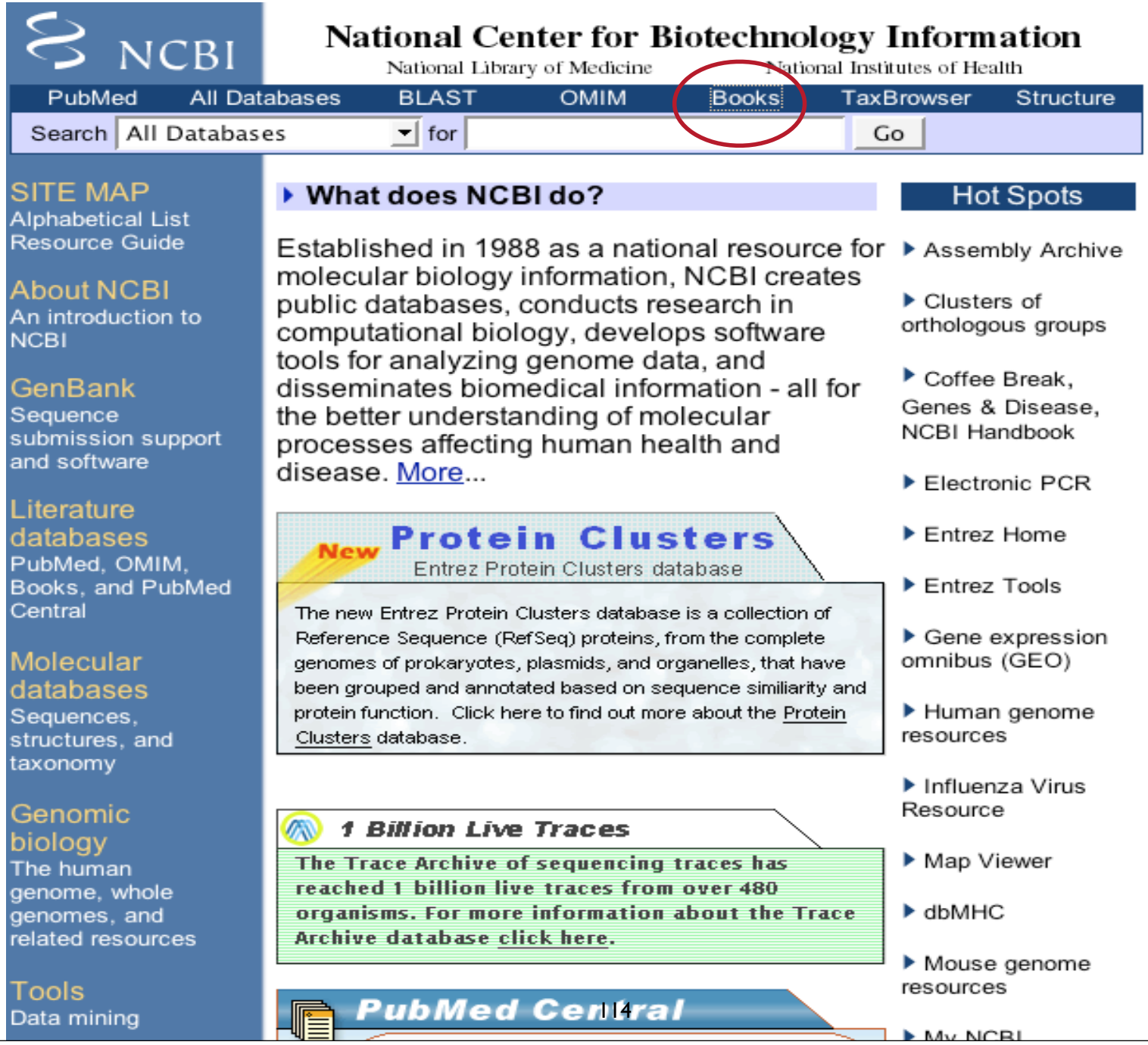
Medical Subject Annotations: (Total:1)

beta Carotene

A carotenoid that is a precursor of VITAMIN A. It is administered to reduce the severity of photosensitivity reactions in patients with erythropoietic protoporphyria (PORPHYRIA, ERYTHROPOIETIC). (From Reynolds JEF(Ed): Martindale: The Extra Pharmacopoeia (electronic version). Micromedex, Inc, Engewood, CO, 1995.)

Show MeSH Tree Structure

Pharmacological Action:
Antioxidants


NCBI
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
PubMed
All Databases
BLAST
OMIM
Books
TaxBrowser
Structure
Search
All Databases
for
Go
SITE MAP
Alphabetical List
Resource Guide
About NCBI
An introduction to NCBI
GenBank
Sequence submission support and software
Literature databases
PubMed, OMIM, Books, and PubMed Central
Molecular databases
Sequences, structures, and taxonomy
Genomic biology
The human genome, whole genomes, and related resources
Tools
Data mining
What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...
New
Protein Clusters
Entrez Protein Clusters database
The new Entrez Protein Clusters database is a collection of Reference Sequence (RefSeq) proteins, from the complete genomes of prokaryotes, plasmids, and organelles, that have been grouped and annotated based on sequence similiarity and protein function. Click here to find out more about the Protein Clusters database.
1 Billion Live Traces
The Trace Archive of sequencing traces has reached 1 billion live traces from over 480 organisms. For more information about the Trace Archive database click here.
PubMed Central
Hot Spots
Assembly Archive
Clusters of orthologous groups
Coffee Break, Genes & Disease, NCBI Handbook
Electronic PCR
Entrez Home
Entrez Tools
Gene expression omnibus (GEO)
Human genome resources
Influenza Virus Resource
Map Viewer
dbMHC
Mouse genome resources
My NCBI

NCBI **Bookshelf**

My NCBI
Welcome joannealisonfox. [Sign Out]

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | OMIM | Books

Search Books for beta carotene Go Clear Save Search

Limits | Preview/Index | History | Clipboard | Details

Display Books Show 20 Send to

All: 196 Figures: 2

- About Entrez
- Books
- Overview
- Using the books
- Information for authors and publishers
- Contact us
- Mailing list
- Project background
- FAQ
- My NCBI
- Privacy Policy

171 items in **Health Services/Technology Assessment Text (HSTAT)**
Bethesda (MD):
National Library of Medicine (US),
2003 Oct.

12 items in **Cancer Medicine.** 6th ed.
Kufe, Donald W.; Pollock, Raphael E.; Weichselbaum, Ralph R.; Bast, Robert C., Jr.; Gansler, Ted S.; Holland, James F.; Frei III, Emil, editors.
Hamilton (Canada): BC Decker Inc; c2003.

2 items in **Molecular Cell Biology.** 4th ed.
Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E.
New York: W. H. Freeman & Co.; c2000.

2 items in **Disease Control Priorities in Developing Countries** 2nd ed.
Dean T. Jamison, Joel G. Breman, Anthony R. Measham, George Alleyne, Mariam Claeson, David B. Evans, Prabhat Jha, Anne Mills, Philip Musgrove, editors
Washington (DC): IBRD/The World Bank and Oxford University Press; 2006

2 items in **Alternative Medicine and Rehabilitation: A Guide for Practitioners**
Wainapel, Stanley F.; Fast, Avital, editors
New York: Demos Medical Publishing; c2003

1 item in **The Organization of the Retina and Visual System**
Kolb, Helga; Fernandex, Eduardo; Nelson, Raph, editors
Salt Lake City (UT): Moran Eye Center, University of Utah; 2000

1 item in **Surgical Treatment**
Holzheimer, Rene G.; Mannick, John A., editors.
Munich: Zuckschwerdt Publishers; c2001.

1 item in **Parkinson's Disease: Diagnosis and Clinical Management**
Factor, Stewart A.; Weiner, William J.
New York: Demos Medical Publishing, Inc.; c2002

Navigation
About this book
Section 4: Cancer Epidemiology, Prevention, and Screening
29. Nutrition in the Etiology and Prevention of Cancer
Methodologic Issues in Diet, Nutrition, and Cancer Studies
Public Health Guidelines for Cancer Prevention
Summary of Research Efforts Focusing on Specific Cancers
→ Current Research
Survivorship: Diet and Nutritional Guidance During and Following Cancer Treatment
References

Search
Go
This book All books
PubMed

Current Research

Specific Foods, Nutrients, and Dietary Components Frequently Associated with Cancer Prevention

Many people at risk of cancer focus their attention upon specific foods or nutrients in part because of the extensive marketing of products and publicity generated by the popular press. This tendency is facilitated by the news media when science reporters publicize results of single studies or preliminary findings, often confusing readers with contradictory and conflicting results. The following section briefly summarizes data regarding selected food components or nutrients and may assist the medical practitioner in responding to specific inquires from individuals.
TOP

Vitamins

Vitamin A

Vitamin A is essential for the normal growth and development of epithelial tissues. Vitamin A deficiency is common in many parts of the developing world, but is extremely rare in Americans. Vitamin A is provided in the diet as retinol and its esters, primarily from milk and organ meats, and as β-carotene and a few other provitamin A carotenoids in yellow and leafy green vegetables. Interest in vitamin A and related compounds in the etiology, prevention, and treatment of cancer is rapidly expanding. A protective effect of consuming foods rich in vitamin A has been hypothesized for several types of cancer[1,2,11,16,18,188]; at this time, however, there is no clear evidence that vitamin A supplementation will decrease the risk of cancer in populations or individuals consuming a healthy diet. Although many studies in laboratory models indicate that vitamin A deficiency increases the susceptibility of
risk will be observed
A excess has not
s pharmacologic
s an important area of
s is lacking, the
a large body of

β Carotene

Foods rich in β-carotene, such as many fruits and vegetables, are associated with a lower risk of cancer. However, recent intervention trials with β-carotene clearly question the validity of that hypothesis that the benefits of a plant-based diet can be produced through β-carotene supplements. Two large intervention studies of β-carotene demonstrated a higher risk of lung cancer in smokers.[124,203] Although β-carotene is a potential antioxidant and source of vitamin A, supplements should be discouraged for cancer prevention and dietary sources should be encouraged.
TOP

Very few studies have investigated the role of vitamin E in human cancer [illegible] but several studies suggest a

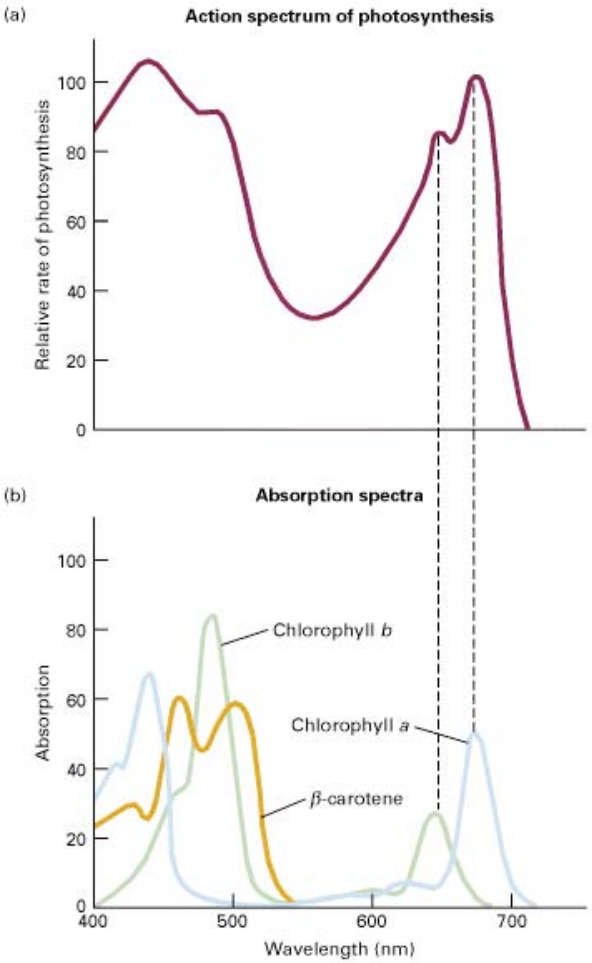


Figure 16-37. Photosynthesis at different wavelengths. (a) The action spectrum of photosynthesis in plants; that is, the ability of light of different wavelengths to support photosynthesis. (b) The absorption spectra for three photosynthetic pigments: chlorophyll *a*, chlorophyll *b*, and β-carotene. Each spectrum shows how well light of different wavelengths is absorbed by one of the pigments. A comparison of the action spectrum with the individual absorption spectra suggests that photosynthesis at 680 nm is primarily due to light absorbed in the

Navigation

About this book

16. Cellular Energetics: Glycolysis, Aerobic Oxidation, and Photosynthesis

16.1. Oxidation of Glucose and Fatty Acids to CO_2

16.2. Electron Transport and Oxidative Phosphorylation

16.3. Photosynthetic Stages and Light-Absorbing Pigments

16.4. Molecular Analysis of Photosystems

16.5. CO_2 Metabolism during Photosynthesis

PERSPECTIVES for the Future

PERSPECTIVES in the Literature

Testing Yourself on the Concepts

MCAT/GRE-Style Questions

References

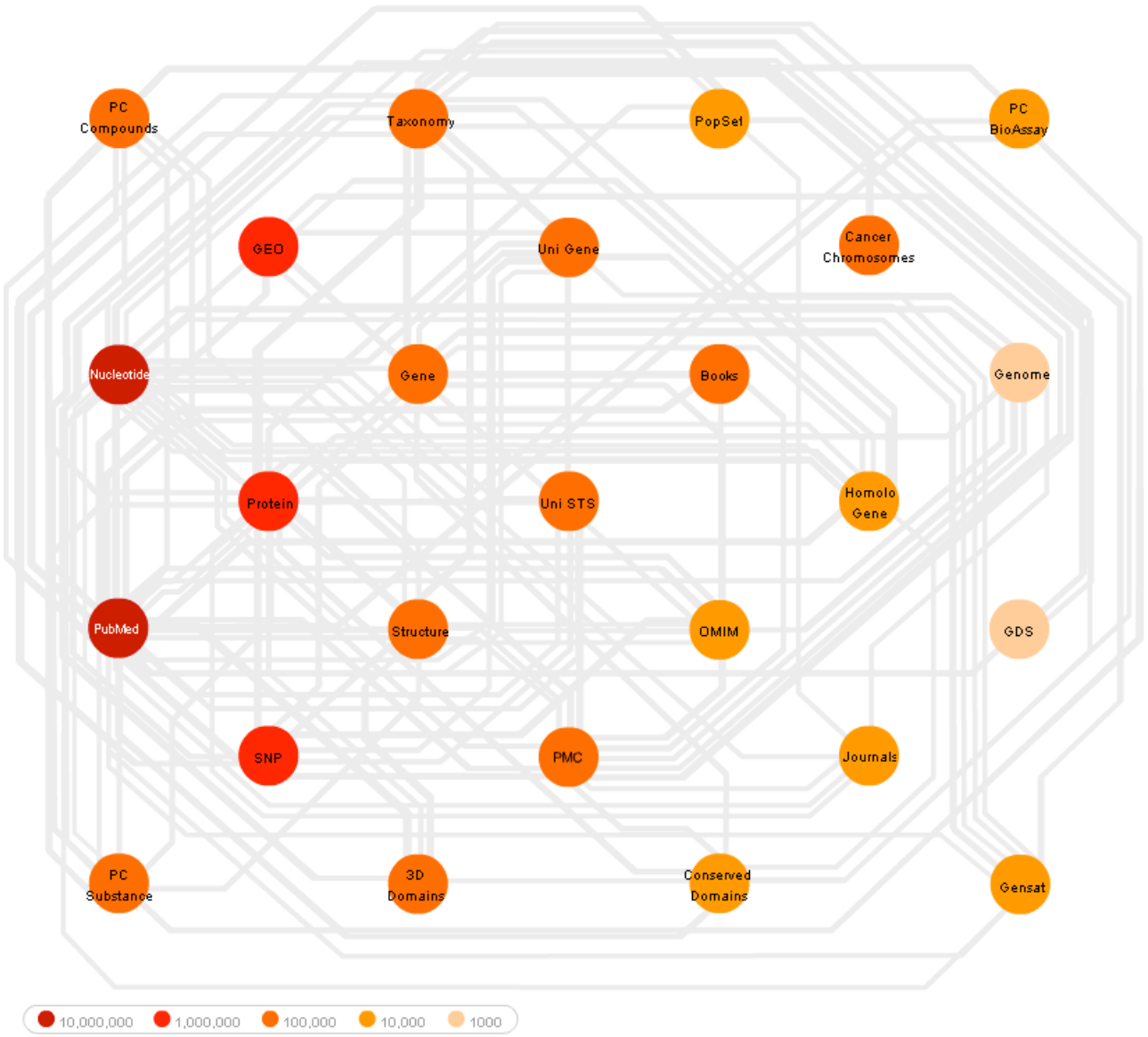
Search

Go

This book All books

PubMed

You can make up your own examples, to search Pubmed… the Bookshelf…


PC Compounds
Taxonomy
PopSet
PC BioAssay
GEO
Uni Gene
Cancer Chromosomes
Nucleotide
Gene
Books
Genome
Protein
Uni STS
Homolo Gene
PubMed
Structure
OMIM
GDS
SNP
PMC
Journals
PC Substance
3D Domains
Conserved Domains
Gensat
10,000,000
1,000,000
100,000
10,000
1000

Web References

- The "About Entrez" page at the NCBI

  http://www.ncbi.nlm.nih.gov/Database/index.html
- Model of Entrez Databases from NCBI

  http://www.ncbi.nih.gov/Database/datamodel/index.html
- PubMed Tutorial from NLM

  http://www.nlm.nih.gov/bsd/pubmed_tutorial/m1001.html

Credits

- Materials for this presentation have been adapted from the following sources:

  NCBI HelpDesk - Field Guide Course Materials

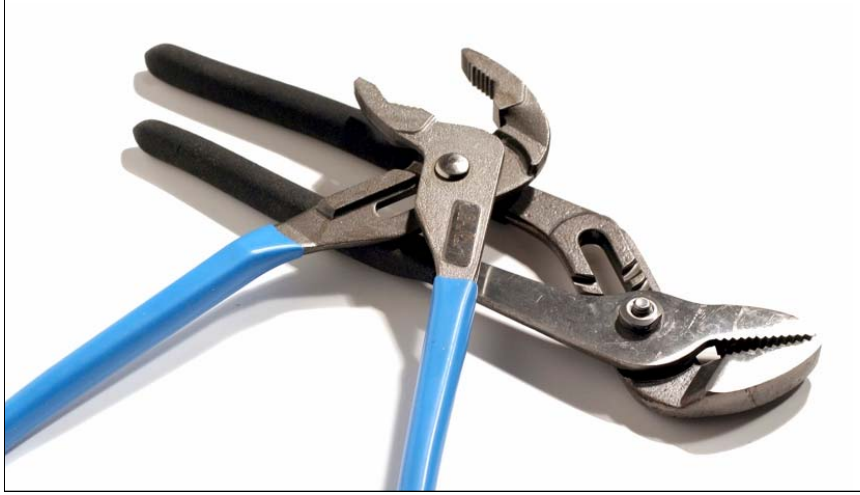
  Bioinformatics: A practical guide to the analysis of genes and proteins

- Questions? Please contact:

Dr. Joanne Fox
Michael Smith Laboratories
joanne@msl.ubc.ca

Let's start at 9:30am

BLAST background, guided tour & practical exercises



joanne@msl.ubc.ca