NCBI Human Genome Resources

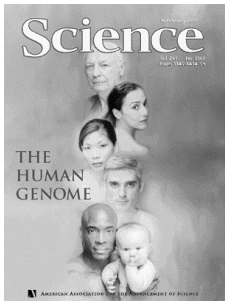
American Society for Cell Biology

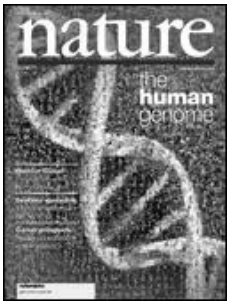
NCBI

Dec 10, 2001

NCBI Booth: 666

The Draft Human Genome





NCBI

What Data is Available?

- NCBI assembled annotated **genomic contigs**
  - Genome project data
  - Other primary data
- **Reference sequences** - mRNA, proteins, transcripts
- Genome Scan **gene models**
- Mapped **variation data**
- **Integrated maps** - RH, genetic, cytogenetic, and sequence
- Clustered and mapped **expressed sequences**
- **Links** to outside data sources

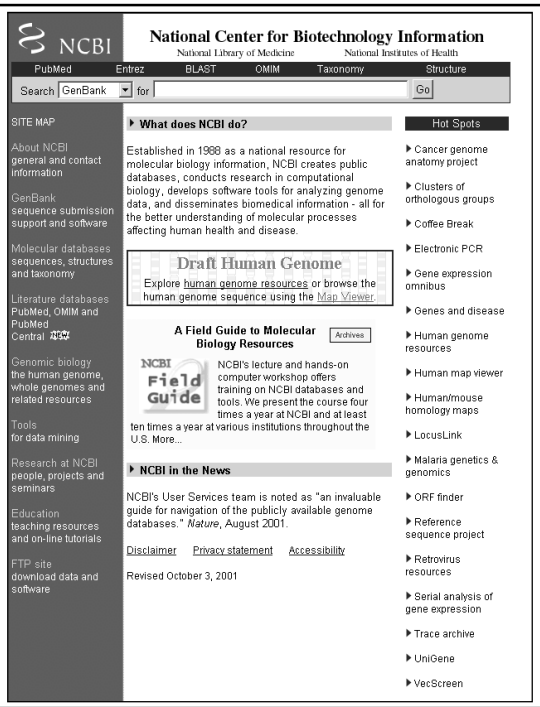
NCBI

How to access it?

Type of query	Resource
Sequence Similarity	Human genome BLAST
Gene name	LocusLink
Map Location	Map Viewer
Database ID	UniGene

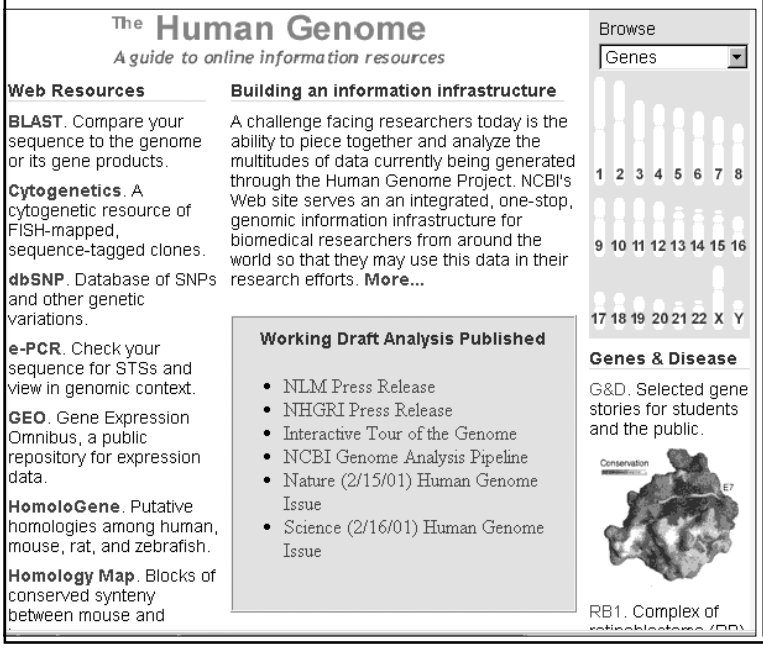
NCBI

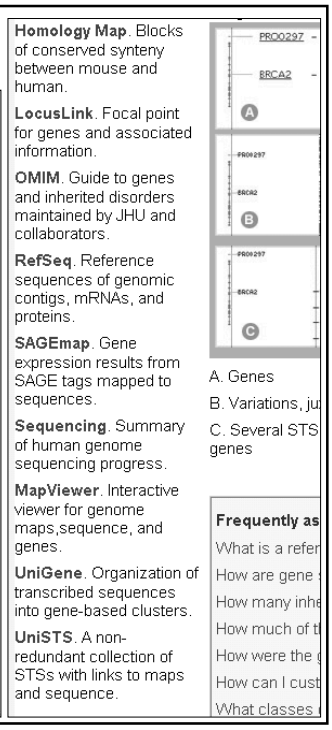
NCBI Homepage

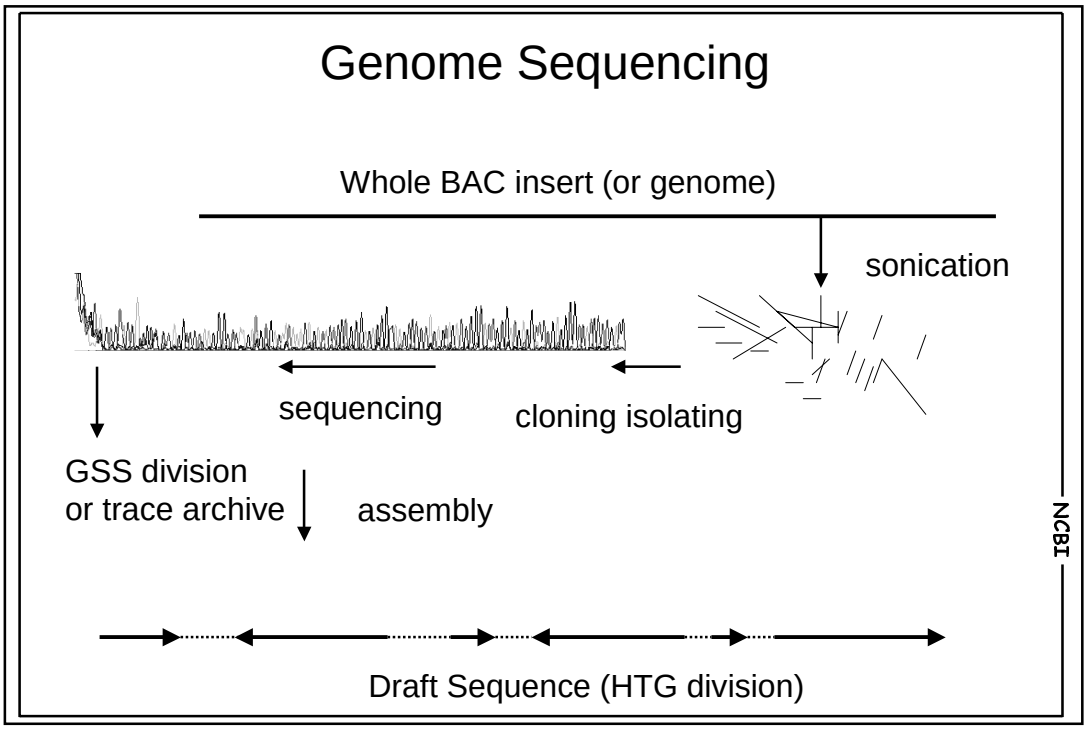


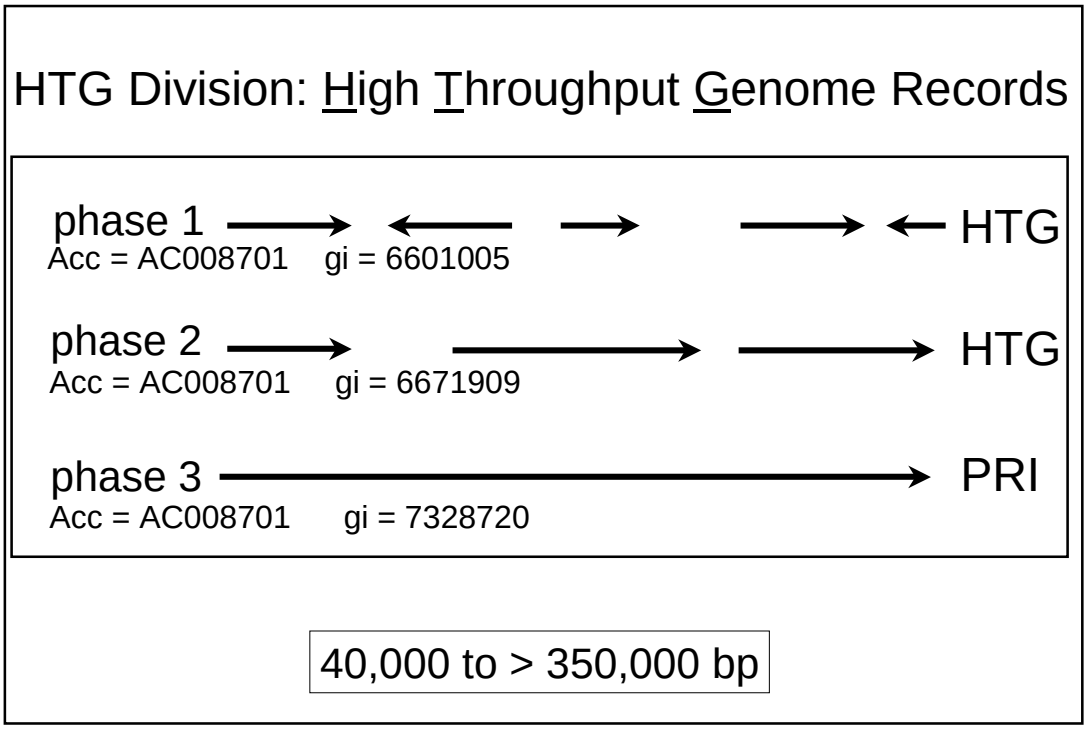
NCBI

Human Genome Resources






Genome Sequencing
Whole BAC insert (or genome)
sonication
sequencing
cloning isolating
GSS division
or trace archive
assembly
NCBI
Draft Sequence (HTG division)


HTG Division: High Throughput Genome Records
phase 1
Acc = AC008701 gi = 6601005
HTG
phase 2
Acc = AC008701 gi = 6671909
HTG
phase 3
Acc = AC008701 gi = 7328720
PRI
40,000 to > 350,000 bp

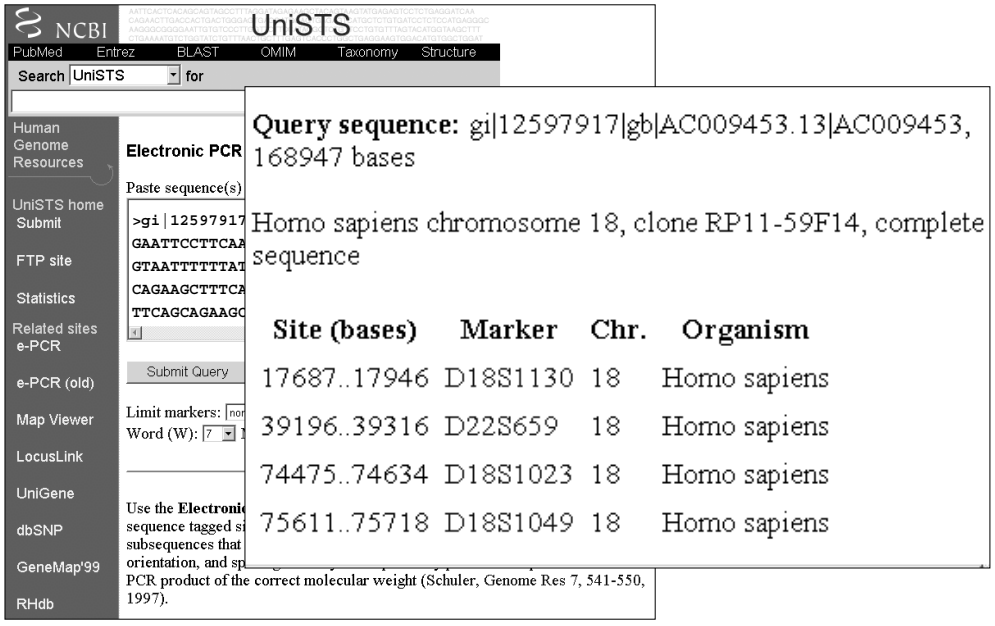
STS Division : Sequence Tagged Sites

- **Segment of gene, EST , mRNA or genomic DNA of known position (microsatellite)**
- **PCR with STS primers gives unique product (one per genome)**
- **Basis of Radiation Hybrid Mapping**
  - **UniGene**
  - **Genome Assembly**
- **Related resource: Electronic PCR**

http://www.ncbi.nlm.nih.gov/genome/sts/epcr.cgi

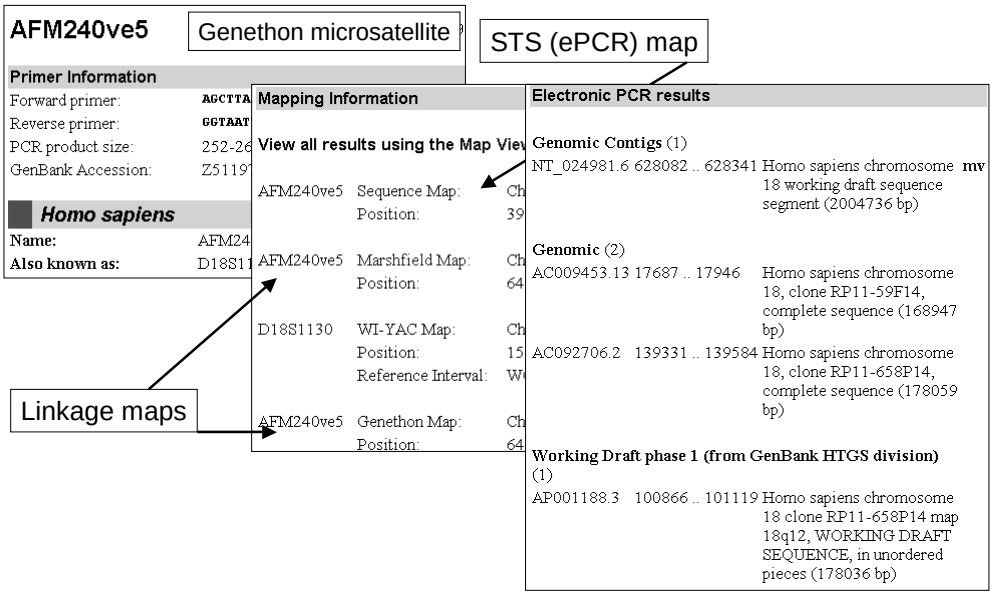
NCBI

Electronic PCR

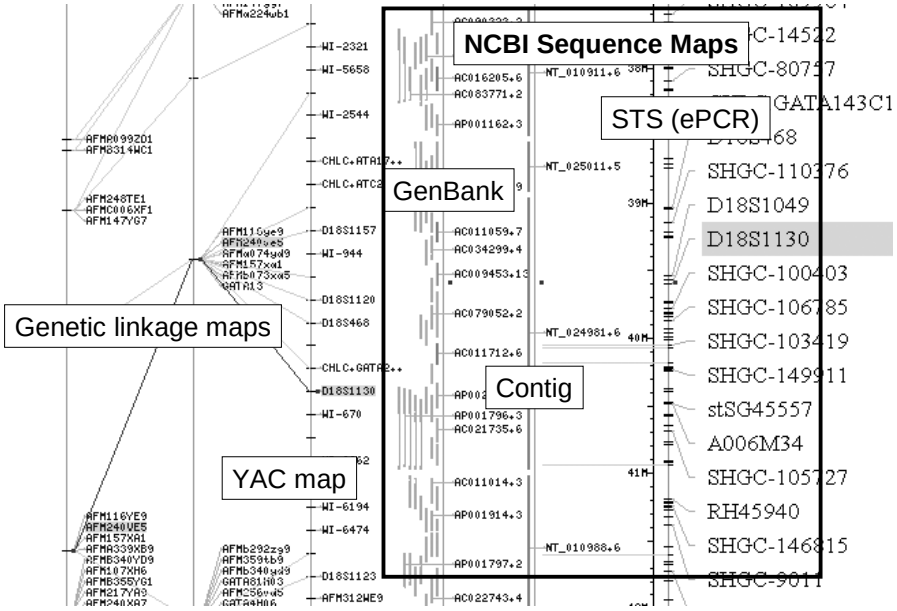


NCBI

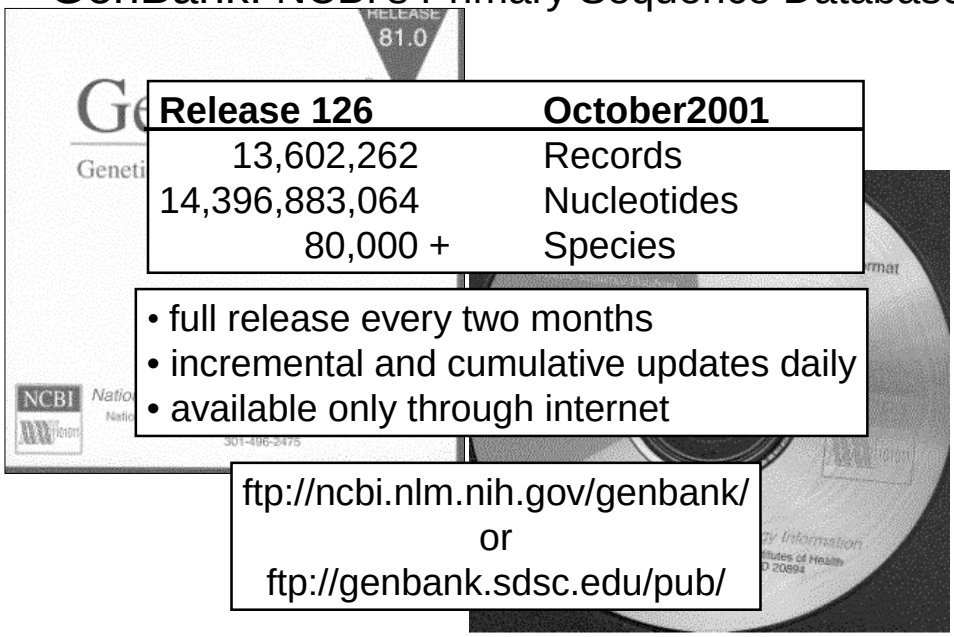
Results for AFM240ve5



Map Viewer: ePCR Results



GenBank: NCBI's Primary Sequence Database



Release 126	October2001
13,602,262	Records
14,396,883,064	Nucleotides
80,000 +	Species

- full release every two months
- incremental and cumulative updates daily
- available only through internet

ftp://ncbi.nlm.nih.gov/genbank/
or
ftp://genbank.sdsc.edu/pub/

NCBI

NCBI Derivative Sequence Databases: RefSeq

NCBI Reference Sequences

mRNAs and Proteins

`NM_123456` **Curated mRNA**
`NP_123456` **Curated Protein**
`XM_123456` **Predicted Transcript**
`XP_123456` **Predicted Protein**

Gene Records

`NG_123456` **Reference Genomic Sequence**

Assemblies

`NT_123456` **Contig (Mouse and Human Genomes)**
`NC_123455` **Chromosome (Microbial Genomes)**

NCBI

GenBank Sequences human CFTR

Show: 100 Items 1-83 of 83 One page.

☐ **1:** M86631 Related Sequences, PubMed, Taxonomy
Homo sapiens (clone ST-18-5(9/16)) cystic fibrosis transmembrane conductance regulator (CFTR) gene, 3' end intron 17B; complete exon 18; complete intron 18
gi|180296|gb|M86631.1|HUMCFTR[180296]

☐ **2:** S64699 Related Sequences, OMIM, Protein, PubMed, Taxonomy
Homo sapiens cystic fibrosis transmembrane conductance regulator isoform 36 (CFTR) mRNA, partial cds
gi|408285|gb|S64699.1|S64699[408285]

☐ **3:** AL121762 Related Sequences, Taxonomy
Human DNA sequence from clone RP4-610C12 on chromosome 20 Contains the 3' end of a novel gene, a putative novel gene, a pseudogene similar to part of the cystic fibrosis transmembrane conductance regulator (CFTR), four CpG islands, ESTs, STSs and GSSs, complete sequence
gi|8574423|emb|AL121762.13|HSJ610C12[8574423]

☐ **4:** AH006034 OMIM, Protein, PubMed, Taxonomy
Human cystic fibrosis transmembrane conductance regulator (CFTR) gene
gi|306537|gb|AH006034.1|SEG_HUMCFTRA[306537]

☐ **5:** M55131 Related Sequences, ProbeSet, OMIM, Protein, PubMed, Taxonomy
Human cystic fibrosis transmembrane conductance regulator (CFTR) gene, exon 24
gi|306536|gb|M55131.1|HUMCFTRA26[306536]

☐ **6:** M55130 Related Sequences, Protein, PubMed, Taxonomy
Human cystic fibrosis transmembrane conductance regulator (CFTR) gene, exon 23

NCBI

Curated RefSeq Records: NM_, NP_

```
LOCUS       NM_000492    6159 bp    mRNA            PRI      26-JUL-1999
DEFINITION  Homo sapiens cystic fibrosis transmembrane conductance
ACC
```

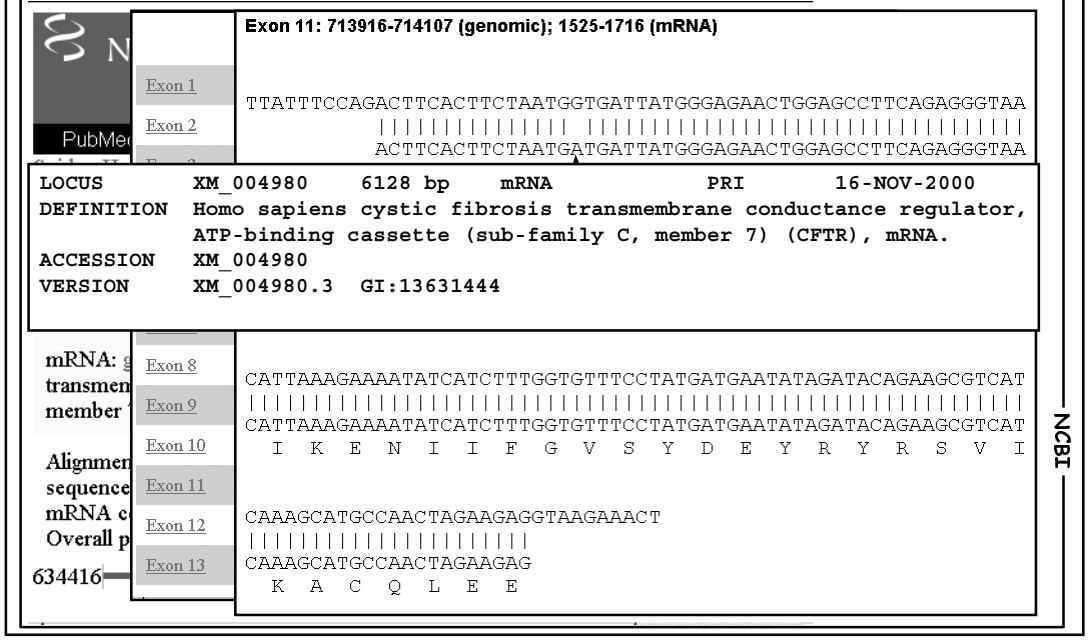
```
LOC                                                                  999
DEF
ACC
PIL
VER
DBS
```

```
COMMENT   REFSEQ: This reference sequence was derived from M55131.
          PROVISIONAL RefSeq: This is a provisional reference sequence
          record that has not yet been subject to human review. The final
          curated reference sequence record may be somewhat different from
          this one.
```

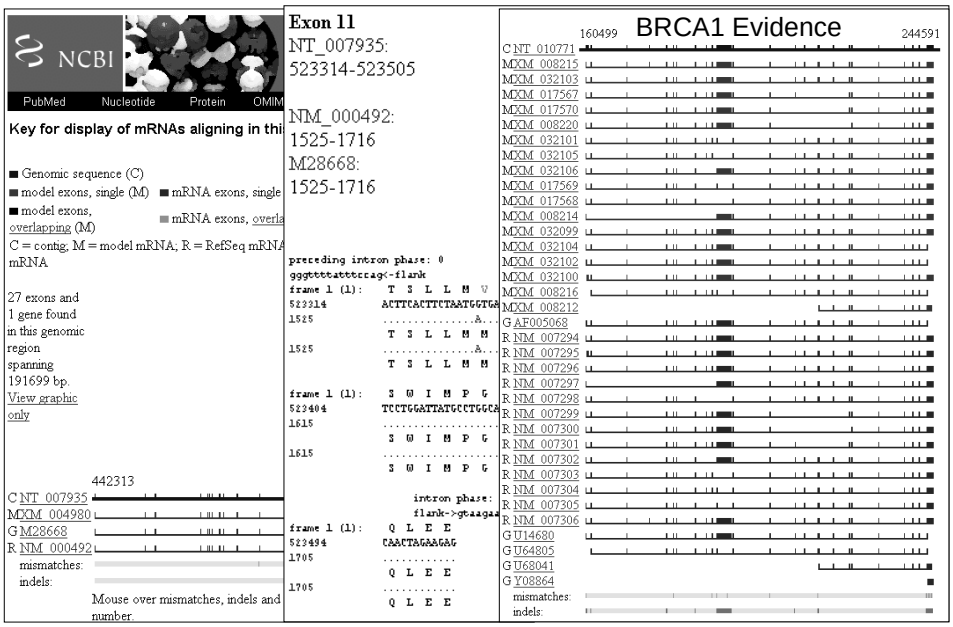
```
REFSEQ: This reference sequence was derived from M28668.1,
M55131.1.
On Feb 17, 2000 this sequence version replaced gi:4502784.
Summary: Cystic fibrosis transmembrane conductance regulator is
member 7 of the ATP-binding cassete sub-family C. The protein
functions as a chloride channel and controls the regulation of
other transport pathways. Mutations in this gene cause the
autosomal recessive disorder, cystic fibrosis (CF) and congenital
bilateral aplasia of the vas deferens (CBAVD). Alternative splice
variants have been described, many of which result from mutations
in the CFTR gene.
COMPLETENESS: full length.
```

Reviewed

Alignment Generated Transcripts: XM_, XP_



The Evidence Viewer: CFTR and BRCA1

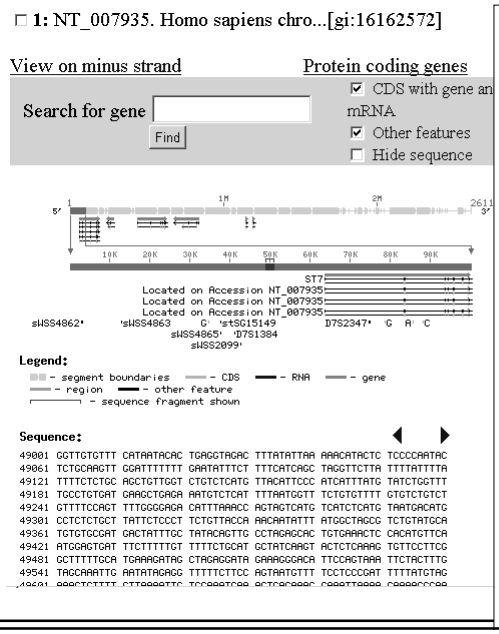


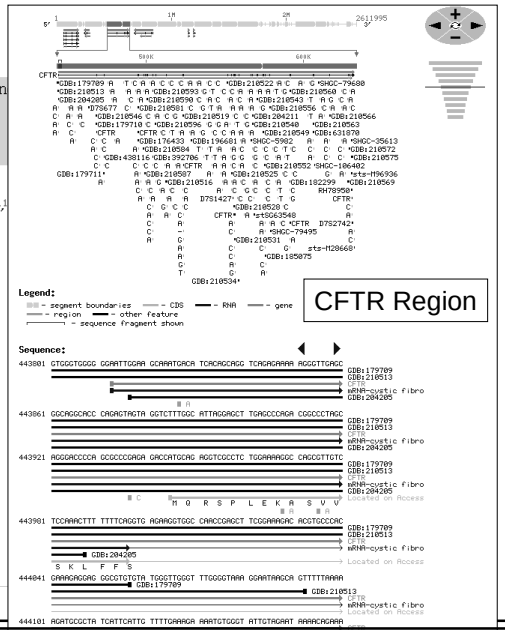
RefSeq Human Contig: NT_

```
LOCUS                                                                    -2000
DEFI
     mRNA            complement(join(1255889..1257642,1258986..1259091,
                     1259690..1259862,1271619..1271708,1281957..1282112,
CONTIG      join(AC073042.3:1155..2680,gap(100),AC074390.2:119526..151445,
            gap(100),AC074390.2:1..5245,gap(100),
            complement(AC074390.2:17705..23645),gap(100),
            AC074390.2:97658..119425,AC073042.3:106479..121155,
            AC074390.2:164226..165036,AC073042.3:70628..79503,gap(100),
            AC073042.3:4627..6382,gap(100),AC073042.3:2781..4526,gap(100),
            complement(AC073042.3:183627..209083),gap(100),
            AC073042.3:79604..88622,gap(100),AC073042.3:139234..160437,
            gap(100),complement(AC073042.3:64..8319),gap(100),
            complement(AC073042.3:39354..45372),..(100),
            complement(AC073042.3:21461..24064),gap
            AC074390.2:156347..160294,gap(100),
            complement(AC074390.2:5346..10750),gap(100),
            complement(AC074390.2:153911..156246),gap(100),
            complement(AC074390.2:23746..32402),gap(100),
            complement(AC074390.2:151546..153810),gap(100),
            complement(AC074390.2:57277..75275),gap(100),
            complement(AC074390.2:75376..97557),gap(100),
```

Reordering draft sequence

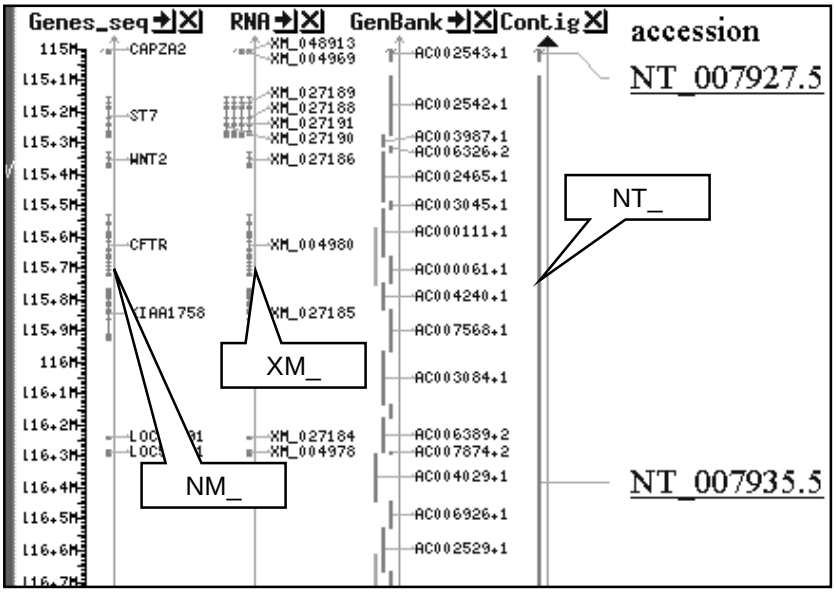
Graphical Sequence View of Contig





NCBI

Map View of RefSeqs



NCBI

RefSeq Genome Records: NG_

- 1: NG_000004 PubMed, Protein, Related Sequences, Taxonomy
  Homo sapiens genomic cytochrome P450, subfamily IIIA (niphedipine oxidase) (CYP3A) on chromosome 7
  gi|13937343|ref|NG_000004.1|[13937343]

- 2: NG_000007 PubMed, Protein, Related Sequences, Taxonomy
  Homo sapiens genomic beta globin region (HBB@) on chromosome 11
  gi|13907843|ref|NG_000007.1|[13907843]

NCBI

Other NCBI Derivative Databases

UniGene -	gene oriented expressed sequence clusters
LocusLink -	central resource and interface for known genes

NCBI

EST Division: Expressed Sequence Tags

```
>IMAGE:275615 5' mRNA sequence
GACAGCATTCGGGCCGAGATGTCTCGCTCCGTGGCCTTAGCTGTGCTCGCGCTACTCTCTCTTTCTGGCC
TGGAGGTATCCAGCGTACTCCAAAGATTCAGGTTTACTCACGTCATCCAGCAGAGAATGGAAAGTCAAAT
TTCCTGAATTGCTATGTGTCTGGGTTTCATCCATCCGACATTGAAGTTGACTTACTGAAGAATGGAGAGA
GAATTGAAAAAGTGGAGCATTCAGACTTGTCTTTCAGCAAGGACTGGTCTTTCTATCTCTTGTACTACAC
TGAATTCACCCCCACTGAAAAAGATGAGTATGCCTGCCGTGTTGAACCATGTNGACTTTGTCACAGNCCC
AAGTTNAGTTTAAGTGGGNATCGAGACATGTAAGGCAGGCATCATGGGAGGTTTTGAAGNATGCCGCNTT
TTGGATTGGGATGAATTCCAAATTTCTGGTTTGCTTGNTTTTTTAATATTGGATATGCTTTTG
```

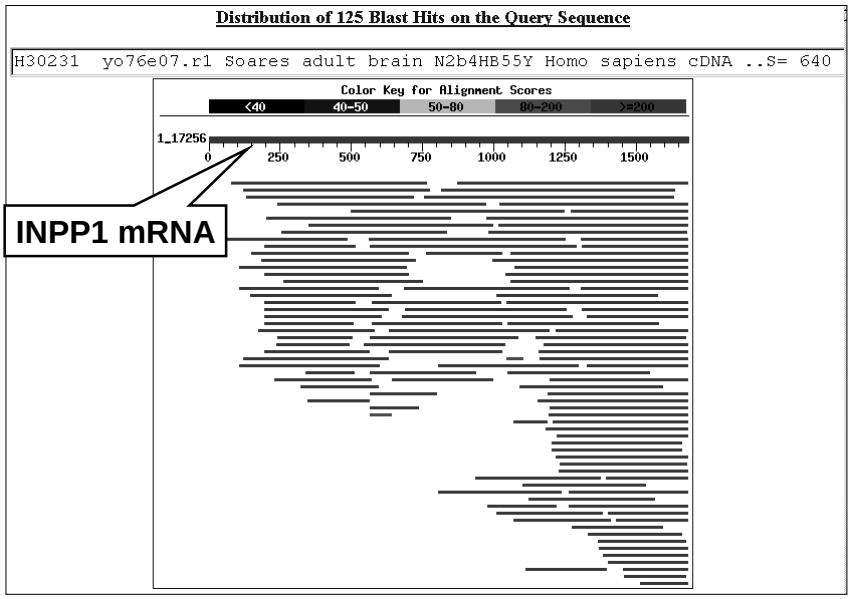
```
>IMAGE:275615 3', mRNA sequence
NNTCAAGTTTTATGATTTATTTAACTTGTGGAACAAAAATAAACCAGATTAACCACAACCATGCCTTACT
TTATCAAATGTATAAGANGTAAATATGAATCTTATATGACAAAATGTTTCATTCATTATAACAAATTTCC
AATAATCCTGTCAATNATATTTCTAAATTTTCCCCCAAATTCTAAGCAGAGTATGTAAATTGGAAGTTAA
CTTATGCACGCTTAACTATCTTAACAAGCTTTGAGTGCAAGAGATTGANGAGTTCAAATCTGACCAAGAT
GTTGATGTTGGATAAGAGAATTCTCTGCTCCCCACCTCTANGTTGCCAGCCCTC
```

NCBI

make cDNA library

80-100,000 unique cDNA clones in library

EST hits INPP1 mRNA



NCBI

What is UniGene?

A gene-oriented view of sequence entries

- **MegaBlast based automated sequence clustering**
- **Nonredundant set of gene oriented clusters**
- **Each cluster a unique gene**
- **Information on tissue types and map locations**
- **Includes well-characterized genes and novel ESTs**
- **Useful for gene discovery and selection of mapping reagents**

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

NCBI

Hs UniGene Statistics

```
    64,436     mRNAs + gene CDSs
 1,137,040     EST, 3'reads
 1,215,789     EST, 5'reads
+  597,535     EST, other/unknown
----------
 3,014,800    total sequences in clusters

Final Number of Clusters (sets)
===============================
96,574   sets total

20,213   sets contain at least one known gene
95,519   sets contain only EST
19,158  sets contain both genes and ESTs
```

80% uncharacterized transcripts

UniGene Build 144
Nov. 20th, 2001

NCBI

UniGene Collections *Sept 26, 2001*

		Sequences	Clusters
Animals			
Homo sapiens	human	3,014,800	96,574
Mus musculus	mouse	1,825,043	89,242
Rattus norvegicus	rat	298,003	59,265
Danio rerio	zebrafish	56,938	10,642
Bos taurus	cow	87,310	7,367
Xenopus laevis	frog	58,133	11,984
Plants			
Arabidopsis thaliana	thale cress	131,068	25,997
Oryzia sativa	rice	47,841	12,836
Triticum aestivum	wheat	31,826	2,744
Hordeum vulgare	barley	34,812	4,041
Zea mays	maize (corn)	69,231	7,161

NCBI

Cluster Hs.32309
Links and Homology

UniGene Cluster Hs.32309 INPP1

Inositol polyphosphate-1-phosphatase

SEE ALSO

LocusLink: 3628
OMIM: 147263
HomoloGene: Hs.32309

SELECTED MODEL ORGANISM PROTEIN SIMILARITIES
organism, protein and percent identity and length of aligned region

H. sapiens:	sp:P49441 - INPP_HUMAN INOSITOL POLYPHOSPHATE 1-PHOSPHATASE	**100 % / 398 aa**
M. musculus:	sp:P49442 - INPP MOUSE INOSITOL POLYPHOSPHATE 1-PHOSPHATASE	**80 % / 396 aa**
C. elegans:	pir:T27862 - T27862 hypothetical protein ZK430.2 - Caenorhabditis elegans	**29 % / 198 aa**

NCBI

Cluster Hs.32309
Mapping Data

UniGene Cluster Hs.32309 INPP1

Inositol polyphosphate-1-phosphatase

SEE ALSO
LocusLink:
OMIM:
HomoloGene:

SELECTED MC
organism, prote
H. sapiens: s
M. musculus: s

MAPPING INFORMATION

Chromosome:	2
Cytogenetic Position:	2q32
Whitehead map:	WI-9155, Chr.2, YAC contig WC2.15
UniSTS entries:	WI-9155 Genomic Context: Map View
UniSTS entries:	L08488 Genomic Context: Map View
UniSTS entries:	SHGC-8914 Genomic Context: Map View

C. elegans: pir:T27862 - T27862 hypothetical protein ZK430.2 - Caenorhabditis elegans **29 % / 198 aa**

NCBI

Cluster Hs.32309
Expression Data

UniGene Cluster Hs.32309 INPP1

Inositol polyphosphate-1-phosphatase

SEE ALSO
LocusLink:
OMIM:
HomoloGene:

SELECTED MO
organism, prote
***H. sapiens*:** s
***M. musculus*:** s
***C. elegans*:** pir:T2
Caeno

MAPPING INFORMATION
Ch
Cy
Wh
Un
Un
Un

EXPRESSION INFORMATION

cDNA sources: Brain, CNS, Colon, Eye, Germ Cell, Heart, Kidney, Liver, Lung, Pancreas, Pooled, Prostate, Stomach, Testis, Tonsil, Uterus, Whole embryo, amnion_normal, bladder_tumor, bone marrow, brain, cervix, colon, denis_drash, epid_tumor, head_neck, head_normal, kidney, liver, lung, lung_normal, marrow, muscle (skeletal), nervous_normal, ovary, placenta_normal, prostate, prostate_normal, stomach

BodyMap: GS 6746

SAGE : Gene to Tag mapping

NCBI

Cluster Hs.32309
Sequences

mRNA/GENE SEQUENCES (4)

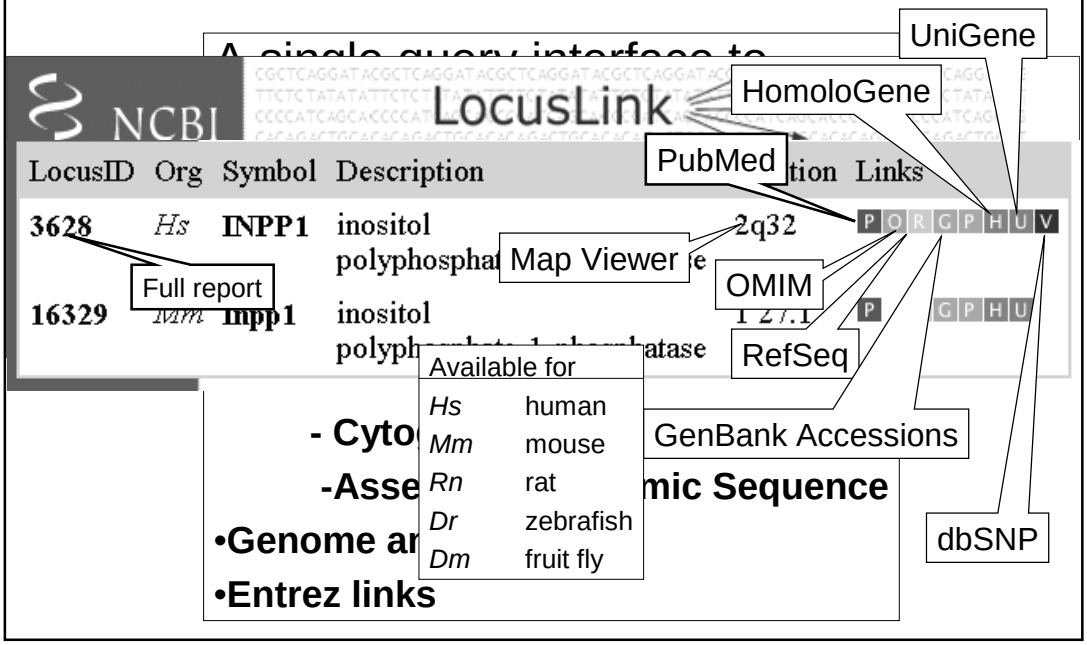
AF141325	Homo sapiens inositol polyphosphate 1-phosphatase (INPP1) gene, complete cds	P
XM_002279	Homo sapiens inositol polyphosphate-1-phosphatase (INPP1), mRNA	A
NM_002194	Homo sapiens inositol polyphosphate-1-phosphatase (INPP1), mRNA	P A
L08488	Human inositol polyphosphate 1-phosphatase mRNA, complete cds	P A S

EST SEQUENCES (10 of 136)[Show all ESTs]

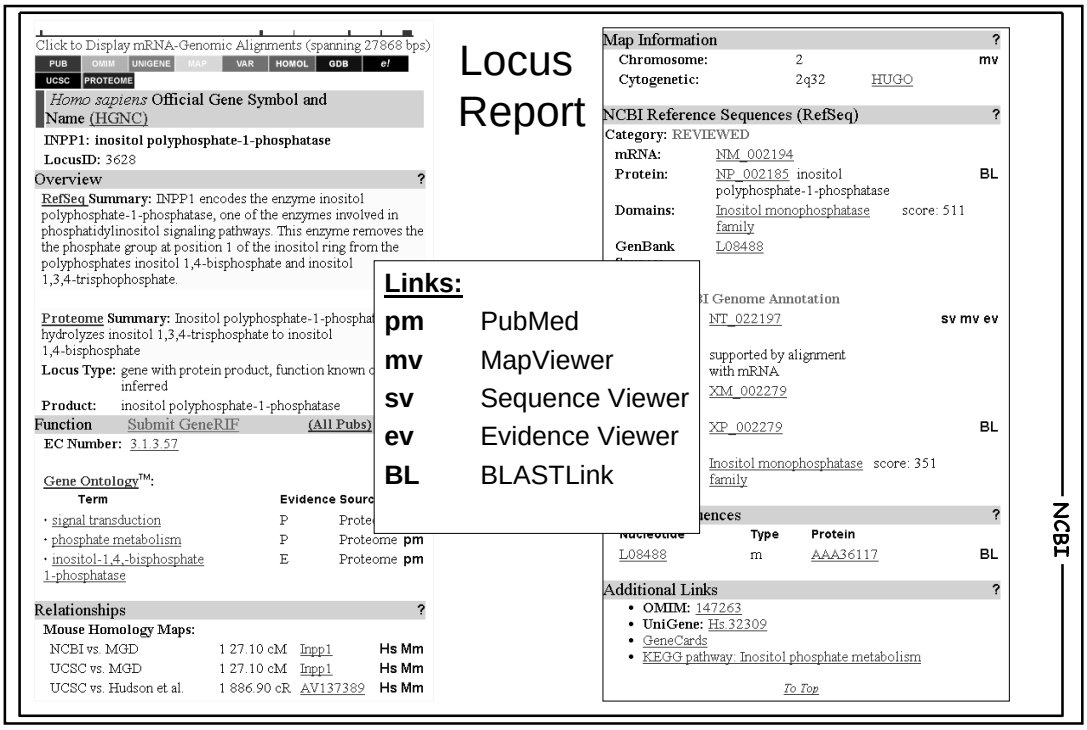
AI634687	cDNA clone IMAGE:2288044	Uterus	3' read 1.8 kb	P
H52036	cDNA clone IMAGE:180803	Brain	5' read 1.8 kb	S
H52141	cDNA clone IMAGE:180803	Brain	3' read 1.8 kb	P
AI244189	cDNA clone IMAGE:1865821	Kidney	3' read 1.7 kb	P
H30231	cDNA clone IMAGE:183876	Brain	5' read 1.7 kb	S
H26976	cDNA clone IMAGE:183876	Brain	3' read 1.7 kb	P
R85218	cDNA clone IMAGE:180546	Brain	5' read 1.7 kb	P
R85219	cDNA clone IMAGE:180546	Brain	3' read 1.7 kb	P
AA115192	cDNA clone IMAGE:491608	Uterus	5' read 1.6 kb	P
AA115193	cDNA clone IMAGE:491608	Uterus	3' read 1.6 kb	P A S

NCBI

LocusLink



Locus Report



NCBI

Polymorphisms from DB SNP

From SNP Database:

```
Submitter Handle:            CMGMM
Submitter Batch ID:          Martens-260201
Release Date:                Apr 10 2001  3:59PM
Molecular type:              Genomic
No. of Chromosomes sampled:  144
Synonym defined:
Organism:                    Homo sapiens
Submitter Method ID:         METHOD-Seq
Citation:
   DNA sequence diversity in the human inositol polyphosphate 1-phosphatase
   gene (INPP1) in lithium-treated bipolar patients

   View citation details
```

```
NCBI Assay ID:        2979072
Submitter SNP ID:     inpp10084
Synonyms:
LOCUSID:              3628
Submitter STS ID:
STS Accession:        not available
GenBanK Accession:    AF141325
Comment:
   Alternative SNP Name:682A>G

Gene Name:            INPP1
Length:               473
Linkout_URL:          http://www.dr-martens.no/research/lithium/inpp1/sup_inpp1_showinfo.asp?id=inpp10084

Flanking Sequence Information:

5' Assay:   ACAACTCAGG CAGCTGCATG CATATAAGTA ACTTTCTGAA ATAAACTAGC TACCTGATAT
            CTATTCCTAG ACTTAAAGGT GTTAAGATTC ACCATGTTAT GATGAATTGT GTATTTCTTC
            GTATCATTAA GTTTGAAAGT GACATGAGTA ATGAAAAACA AATGTGATAA TCATCACCAA
            TTGAAATTTT TCTTCTACAG GTGGAAAGGA CAGTGCTATT GGGGCCTTTC TTACAT

Observed:   A/G

3' Assay:   CCAACATGCA TTCACTACAG CTCACCATCT CTAGAAGAAA CGGCAGTGAA ACACACACTG
            GAAACACCGG CTCTGAGGCA GCATTCTCCC CCAGTTTTTC AGCCGTAATT AGTACAAGTG
            AAAAGGAGAC TATGAAAGCT GCATTGTCAC GTGTGTGTGG AGATCGCATA TTTGGGGCAG
            CTGGGGCTGG TTATAAGAGC CTATGTGTTG TCCAAGGCCT CGTTGACATT TACATC
```

NCBI

Advanced Protein Neighbors: BLink INPP1

BLAST | PubMed | Nucleotide | Protein | Genome | Structure | Taxonomy | Help

PubMed | Entrez | BLAST | OMIM | Taxonomy | Structure

Go back

- cellular organisms
  - Bacteria
    - Proteobacteria
      - gamma subdivision
        - Pasteruellaceae group **Actinobacillus actinomycetemcomitans** [1 BLAST hits]
        - Pseudomonaceae/Moraxellaceae group **Pseudomonas aeruginosa** [1 BLAST hits]
      - alpha subdivision **Rickettsia conorii** [1 BLAST hits]
  - Eukaryota
    - Bilateria
      - Pseudocoelomata **Caenorhabditis elegans** [2 BLAST hits]
        - Coelomata
          - Deuterostomia
            - Eutheria
              - Primates **Homo sapiens** [4 BLAST hits Query]
              - Rodentia
                - Murinae
                - Rattus **Rattus norvegicus** [1 BLAST hits]
                - Mus **Mus musculus** [3 BLAST hits]
              - Cetartiodactyla **Bos taurus** [1 BLAST hits]
          - Protostomia **Drosophila melanogaster** [4 BLAST hits]

NCBI

Related Structures

BLAST PubMed Nucleotide Protein Genome Structure Taxonomy Help

Query: gi|4504703 inositol polyphosphate-1-phosphatase [Homo sapiens]
Matching gi: 186426, 1352464, 5678815, 15930108, 11428852
Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo

Best hits | Common Tree | Taxonomy Report | 3D structures | CDD-Search | GI list

2 BLAST hits to 2 unique species Sort by taxonomy proximity

0 Archaea 0 Bacteria 2 Metazoa 0 Fungi 0 Plants 0 Viruses 0 Other Eukaryotae

Keep only Cut-Off 100 Select Reset

rat

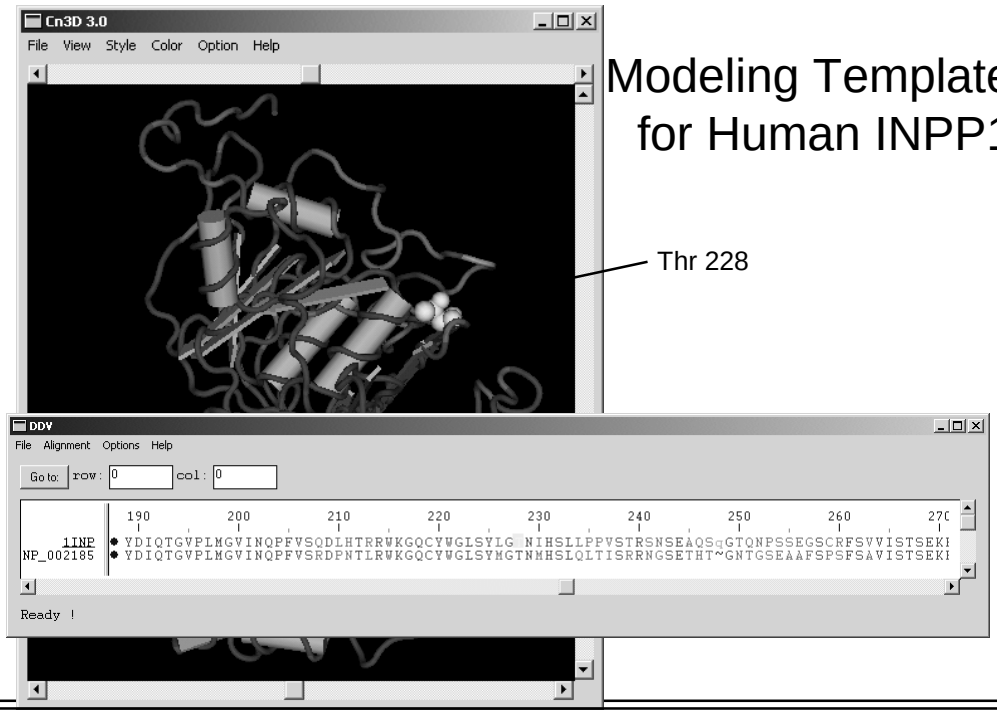
cow

399 aa

SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
1682	•	1INP	999485	Inositol Polyphosphate 1-Phosphatase (1-
196	•	1JP4A	15825834	Chain A, Crystal Structure Of An Enzyme

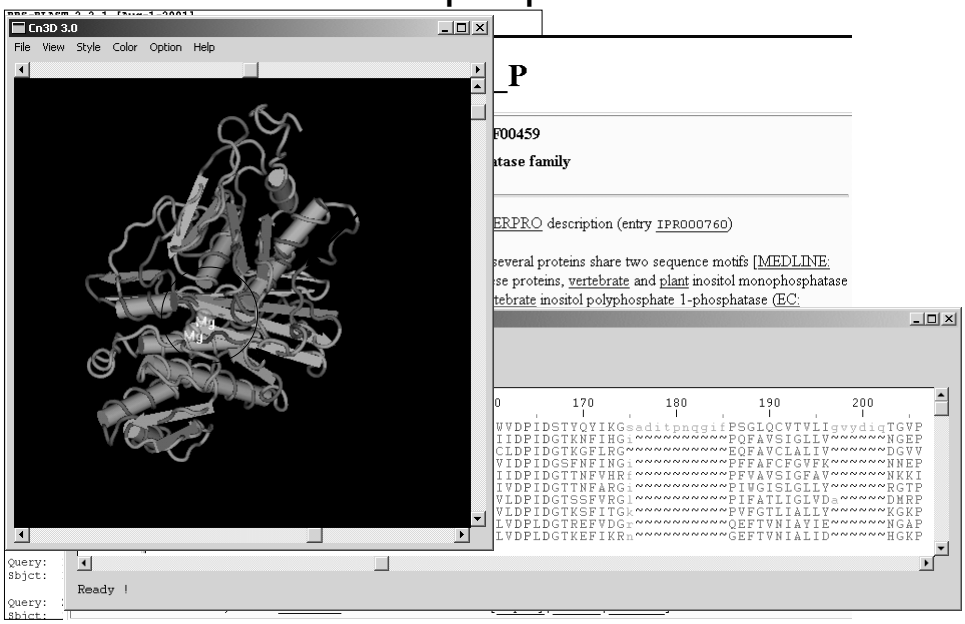
NCBI

Modeling Template for Human INPP1



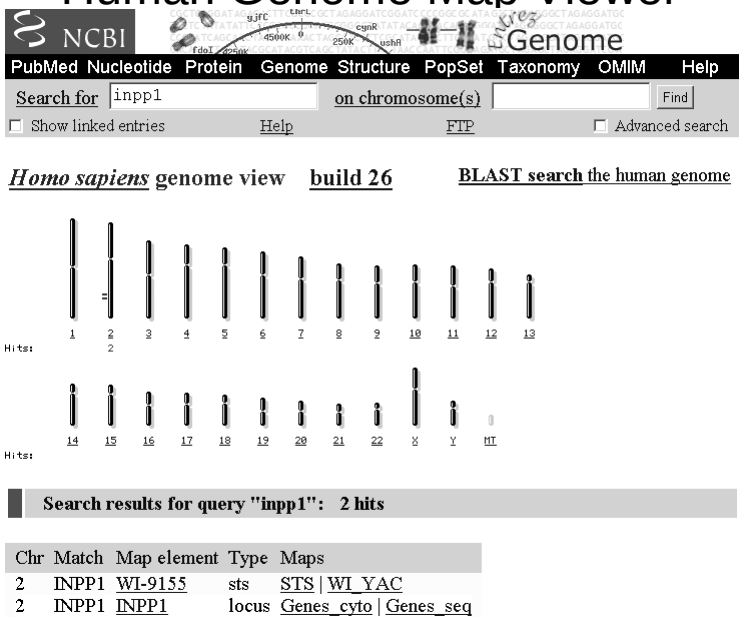
NCBI

CDD: Inositol monophophastase Domain



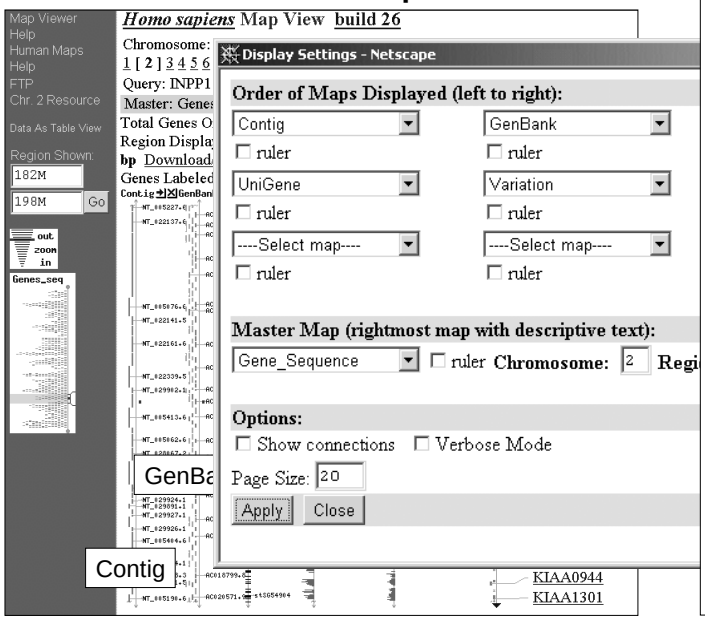
NCBI

Human Genome Map Viewer



NCBI

Map View INPP1



SEQUENCE MAPS
- *Clone*
- *Contig*
- *UniGene (EST)*
- *GenBank*
- *GenomeScan*
- *Gene_Sequence*
- *STS*
- *Variation*

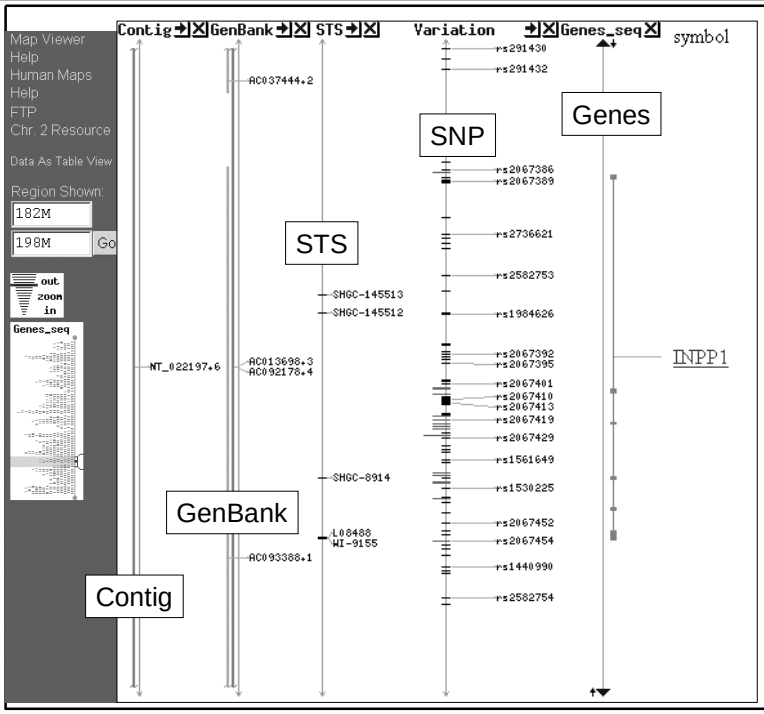
CYTOGENETIC MAPS
- *Ideogram*
- *FISHClone*
- *Genes_Cytogenetic*
- *Mitelman*
- *Morbid*

GENETIC_LINKAGE_MAPS
- *Genethon*
- *Marshfield*

RADIATION_HYBRID_MAPS
- *GeneMap99_G3*
- *GeneMap99_GB4*
- *NCBI_RH*
- *Stanford_G3*
- *Whitehead_RH*

NCBI

INPP1 Zooming In



NCBI

Finding the Human Prestin Gene

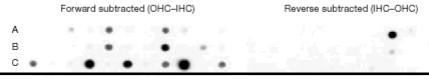
Prestin is the motor protein of cochlear outer hair cells

Jing Zheng*, Weixing Shen*, David Z. Z. He*, Kevin B. Long†, Laird D. Madison† & Peter Dallos*

** Auditory Physiology Laboratory (The Hugh Knowles Center), Departments of Neurobiology and Physiology and Communciation Sciences and Disorders, Northwestern University, Evanston, Illinois 60208, USA*
† Center for Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Medical School, Chicago, Illinois 60611, USA

The outer and inner hair cells of the mammalian cochlea perform different functions. In response to changes in membrane potential, the cylindrical outer hair cell rapidly alters its length and stiffness. These mechanical changes, driven by putative molecular motors, are assumed to produce amplification of vibrations in the cochlea that are transduced by inner hair cells. Here we have identified an abundant complementary DNA from a gene, designated *Prestin*, which is specifically expressed in outer hair cells. Regions of the encoded protein show moderate sequence similarity to pendrin and related sulphate/anion transport proteins. Voltage-induced shape changes can be elicited in cultured human kidney cells that express prestin. The mechanical response of outer hair cells to voltage change is accompanied by a 'gating current', which is manifested as nonlinear capacitance. We also demonstrate this nonlinear capacitance in transfected kidney cells. We conclude that prestin is the motor protein of the cochlear outer hair cell.

Cochlear hair cells are non-neuronal epithelial cells that transduce acoustic signals. Outer hair cells (OHCs) are responsible for the exquisite sensitivity and frequency-resolving capacity of the normal mammalian hearing organ[1]; they provide local mechanical amplification (the 'cochlear amplifier') in the form of feedback. In contrast, inner hair cells convey auditory information to the brain. Outer hair cells have cylindrical somata of constant diameter and variable length. When their membrane potential is altered[2–5], somatic shape changes of up to 5% occur; the cell shortens when depolarized and lengthens when hyperpolarized[3,5]. Length changes do not depend on either ATP or Ca^{2+} (ref. 6), and they can be elicited with unchanging amplitude at microsecond rates up to high audio frequencies[7,8]. Motile responses are accompanied by charge movement, which is reflected in nonlinear capacitance[9,10], akin to the translocation of gating charges of voltage-gated ion channels[11]. This nonlinear capacitance is widely used as a 'signature' of the electromotile process[10,12]. Motility is also accompanied by a change in the axial stiffness of the cell[13]. By virtually any test, electromotility and electrically-induced stiffness changes are correlated (D.Z.Z. and P.D., unpublished data) and we collectively describe them as



NCBI

Standard BLAST Output

Distribution of 6 Blast Hits on the Qu

Mouse-over to show defline and scores. Click to show

Color Key for Alignment Sc
<40 | 40-50 | 50-80

1_6605
0 250 500 750 1000 1250 1

Distribution of 17 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments

Color Key for Alignment Scores
<40 | 40-50 | 50-80 | 80-200 | >=200

1_7380
0 250 500 750 1000 1250 1500 1750 2000

Results against nr(nt)

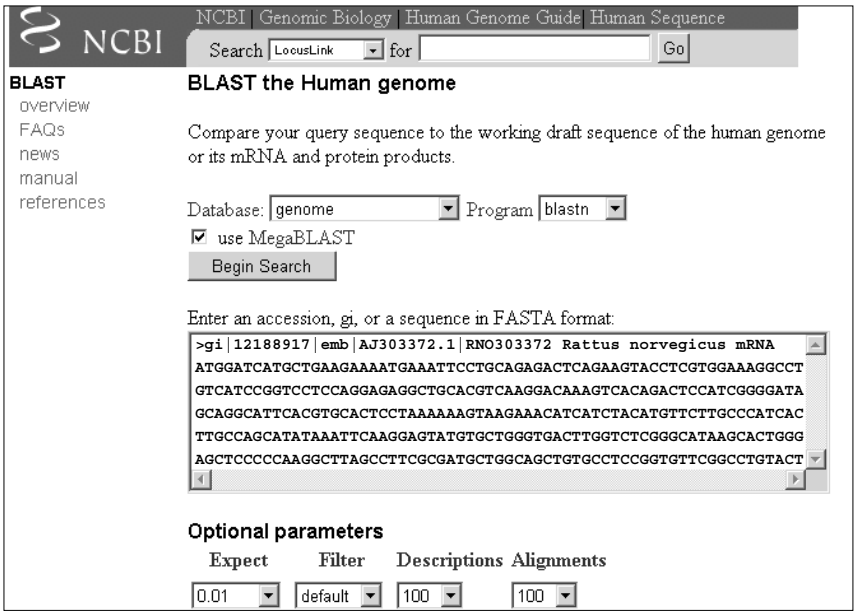
```
Sequences producing significant alignments:

gi|10445387|gb|AC005064.3|AC005064  Homo sapiens clone CT

                                        Alignments

>gi|10445387|gb|AC005064.3|AC005064 Homo sapiens clone CT
          Length = 119484

 Score =  230 bits (116), Expect = 2e-57
 Identities = 143/152 (94%)
 Strand = Plus / Minus


Query: 736  agtaccgtggctgtgctgcagaatgttaaaaacctcaacgtgtgt
            ||||| || |||||| ||||||||||||||||||||||||||||
Sbjct: 6079 agtacagttgctgtgttgcagaatgttaaaaacctcaacgtgtgt


Query: 796  ctcatggtttttggtttgctgttgggtggcaaggagtttaatgag
            || ||||||||||||||||||||||||||||||||||||||||||
Sbjct: 6019 ctgatggtttttggtttgctgttgggtggcaaggagtttaatgag


Query: 856  ctgccagcgcccattcctctagagttctttgc 887
             |||| ||||| |||||| ||||||||||||
Sbjct: 5959 ttgccggcgcctattcctttagagttctttgc 5928
```

Results against HTG

```
                                                        Score     E
Sequences producing significant alignments:             (bits)  Value

gi|15488554|gb|AC093701.1|  Homo sapiens chromosome UNK clon...   230  3e-57
gi|14485286|gb|AC092065.1|AC092065  Homo sapiens chromosome ...   161  2e-36

                                        Alignments

>gi|15488554|gb|AC093701.1| Homo sapiens chromosome UNK clone RP11-1252L15, *** SEQUENCING IN
                PROGRESS ***, 23 unordered pieces
          Length = 193104

 Score =  230 bits (116), Expect = 3e-57
 Identities = 143/152 (94%)
 Strand = Plus / Minus


Query: 736    agtaccgtggctgtgctgcagaatgttaaaaacctcaacgtgtgttccctaggcgtcggc 795
              ||||| || |||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 151624 agtacagttgctgtgttgcagaatgttaaaaacctcaacgtgtgttccctaggcgtcggg 151565


Query: 796    ctcatggtttttggtttgctgttgggtggcaaggagtttaatgagagatttaaagagaaa 855
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 151564 ctgatggtttttggtttgctgttgggtggcaaggagtttaatgagagatttaaagagaaa 151505


Query: 856    ctgccagcgcccattcctctagagttctttgc 887
               |||| ||||| |||||| ||||||||||||
Sbjct: 151504 ttgccggcgcctattcctttagagttctttgc 151473
```

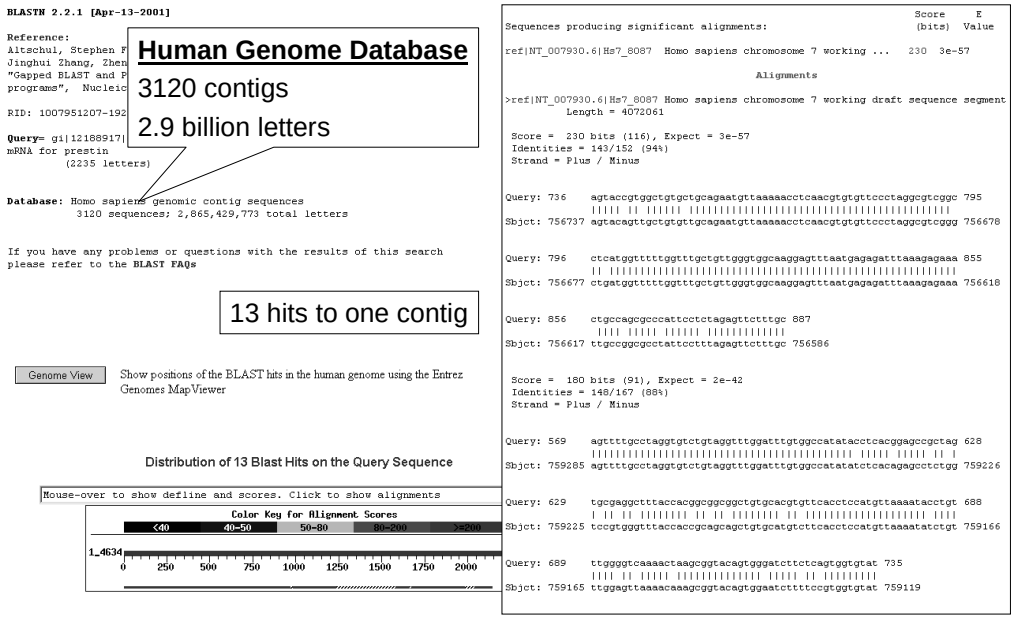
NCBI

Searching the Human Genome



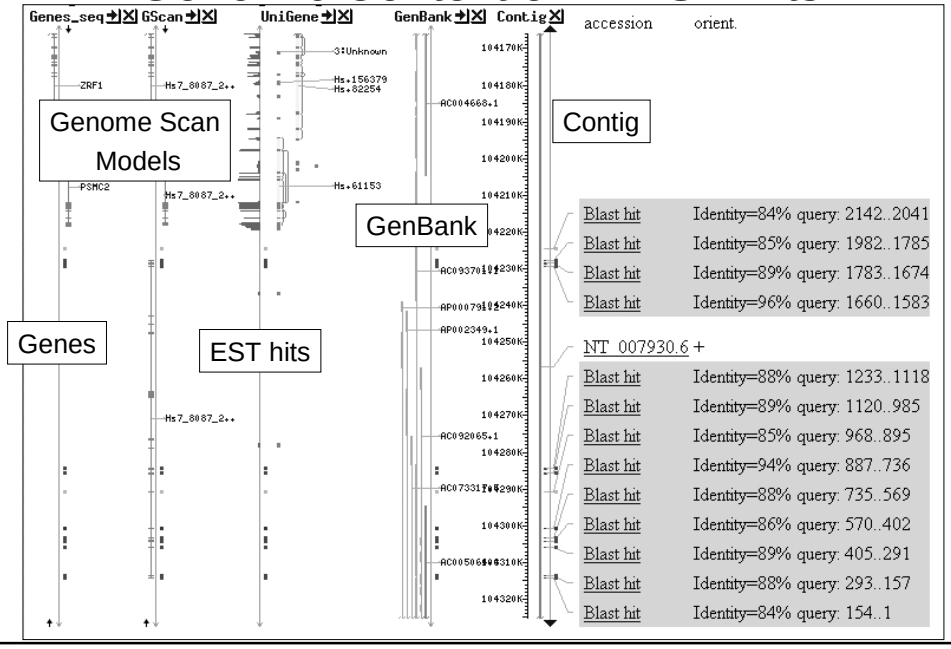
NCBI

BLAST Results



NCBI

Genomic Context of BLAST Hits



NCBI

Fanconi Renal Syndrome

Finding a candidate gene

NCBI

Fanconi Renal Syndrome

1: Am J Hum Genet 2001 Jan;68(1):264-268 Related Articles, OMIM, Books, LinkOut

The University of Chicago Press

Genetic and physical mapping of the locus for autosomal dominant renal Fanconi syndrome, on chromosome 15q15.3.

Lichter-Konecki U, Broman KW, Blau EB, Konecki DS.

Center for Medical Genetics, Marshfield Medical Research Foundation, Marshfield, WI, USA.

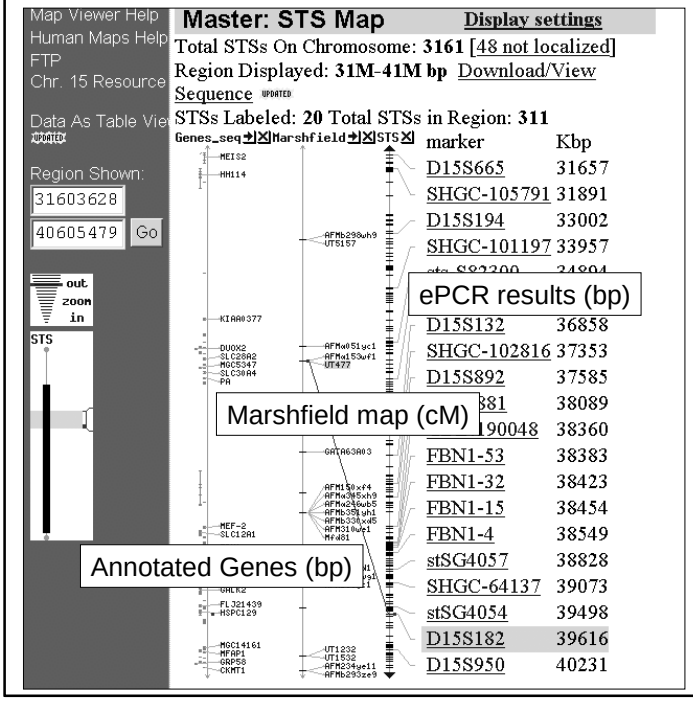
Autosomal dominant renal Fanconi syndrome is a genetic model for the study of proximal renal tubular transport pathology. We were able to map the locus for this disease to human chromosome 15q15.3 by genotyping a central Wisconsin pedigree with 10 affected individuals. ic simple sequence repeat markers, a r arker D15S659 on chromosome 15q15.3. Linkage and haplotype analysis for an additional 24 markers flanking D15S659 narrowed the interval to approximately 3 cM, with the two highest single-point LOD scores observed being 4.44 and 4.68 (for D15S182 and D15S537, respectively). Subsequently, a complete bacterial artificial chromosome contig was constructed, from the High Throughput Genomic Sequence Database, for the region bounded by D15S182 and D15S143. The identification of the gene and gene product altered in autosomal dominant renal Fanconi syndrome will allow the study of the physiology of proximal renal tubular transport.

PMID: 11090339 [PubMed - indexed for MEDLINE]

D15S182 and D15S143

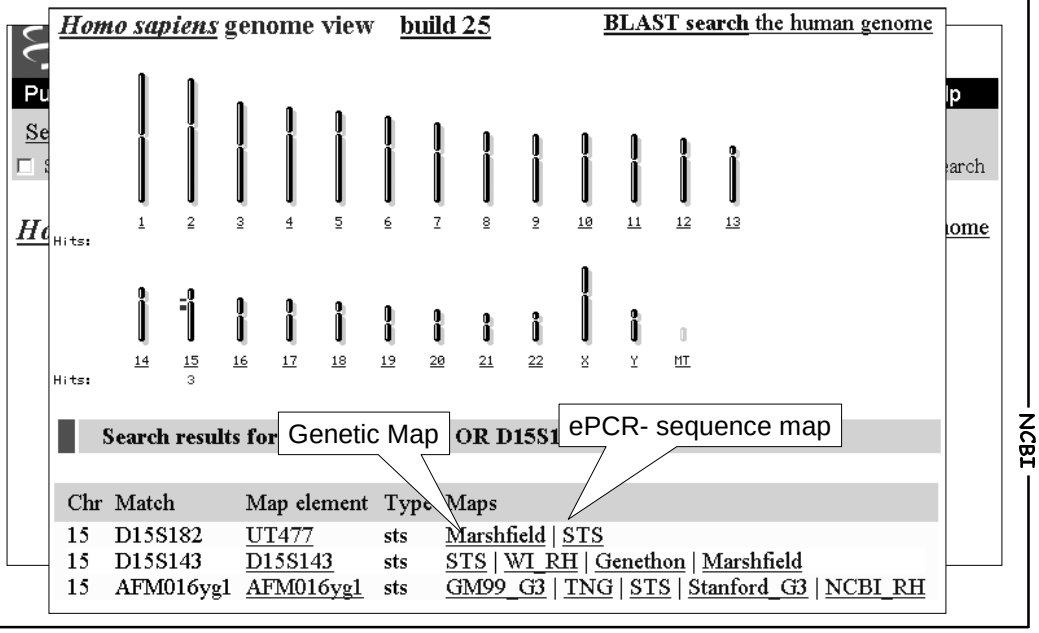
NCBI

Map Viewer Display

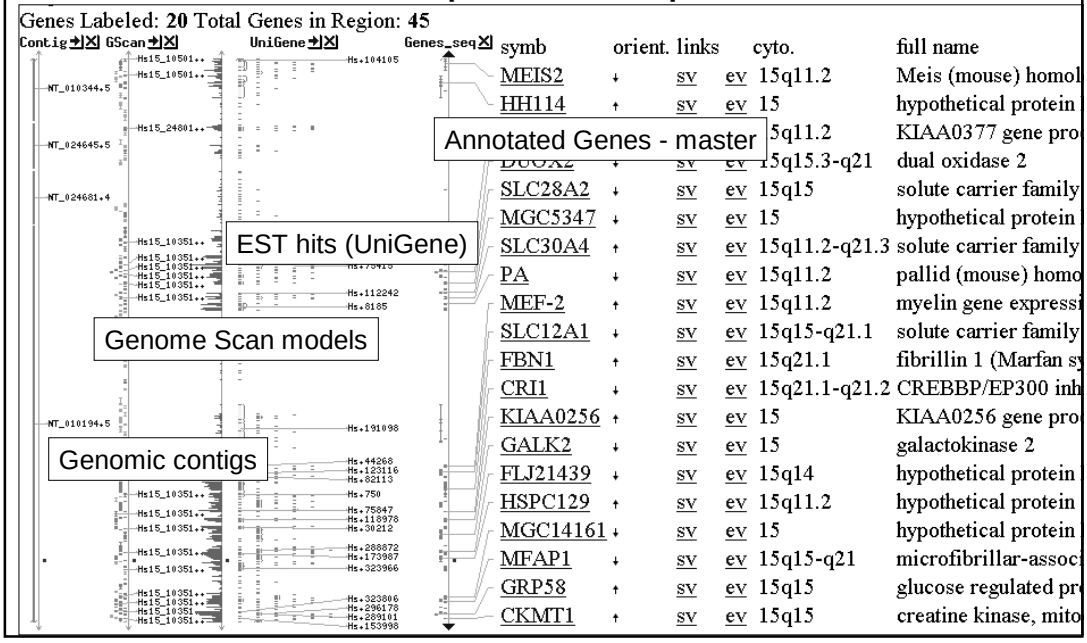


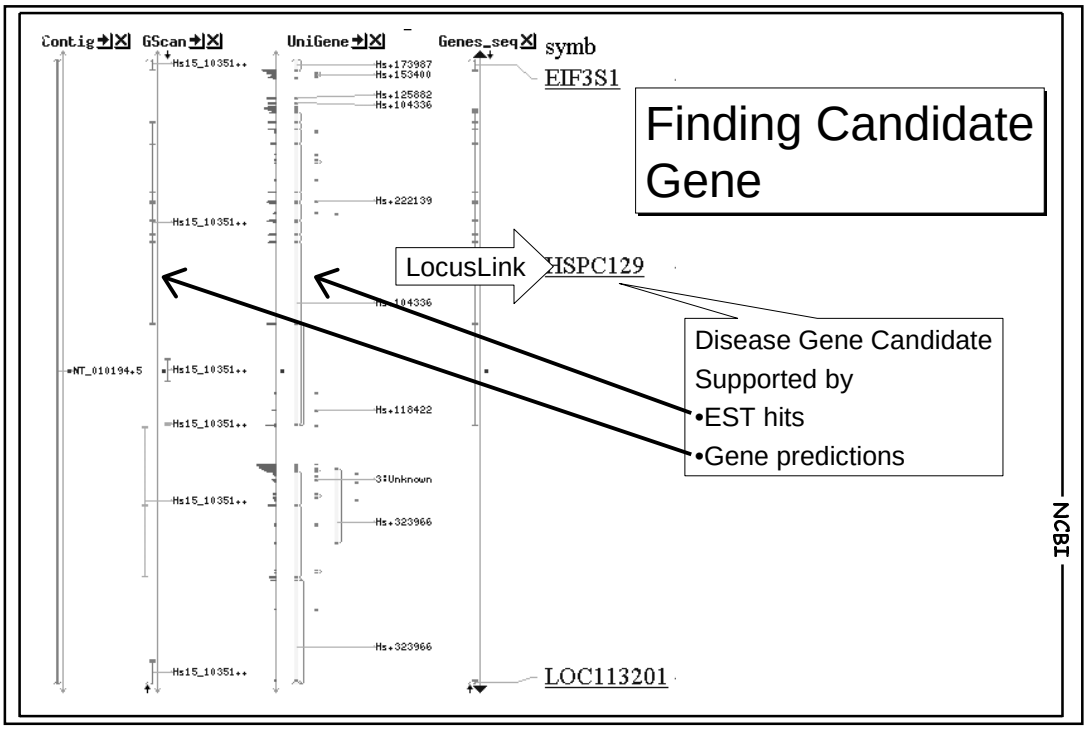
NCBI

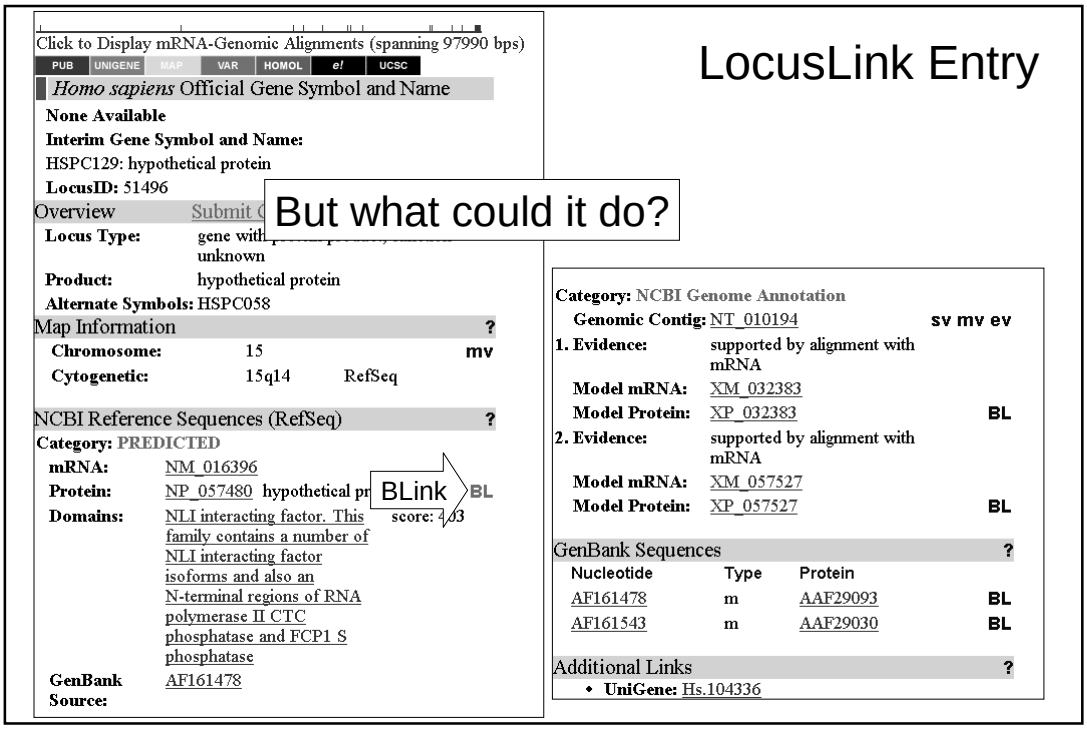
Finding Map Location



Sequence Maps




Contig
GScan
UniGene
Genes_seq
symb
EIF3S1
Finding Candidate Gene
LocusLink
HSPC129
Disease Gene Candidate
Supported by
•EST hits
•Gene predictions
NCBI
LOC113201


LocusLink Entry
Click to Display mRNA-Genomic Alignments (spanning 97990 bps)
PUB UNIGENE MAP VAR HOMOL el UCSC
Homo sapiens Official Gene Symbol and Name
None Available
Interim Gene Symbol and Name:
HSPC129: hypothetical protein
LocusID: 51496
But what could it do?
Overview
Locus Type: gene with unknown
Product: hypothetical protein
Alternate Symbols: HSPC058
Map Information
Chromosome: 15 mv
Cytogenetic: 15q14 RefSeq
NCBI Reference Sequences (RefSeq)
Category: PREDICTED
mRNA: NM_016396
Protein: NP_057480 hypothetical pr
BLink
BL
Domains: NLI interacting factor. This family contains a number of NLI interacting factor isoforms and also an N-terminal regions of RNA polymerase II CTC phosphatase and FCP1 S phosphatase
GenBank Source: AF161478
Category: NCBI Genome Annotation
Genomic Contig: NT_010194 sv mv ev
1. Evidence: supported by alignment with mRNA
Model mRNA: XM_032383
Model Protein: XP_032383 BL
2. Evidence: supported by alignment with mRNA
Model mRNA: XM_057527
Model Protein: XP_057527 BL
GenBank Sequences
Nucleotide Type Protein
AF161478 m AAF29093 BL
AF161543 m AAF29030 BL
Additional Links
UniGene: Hs.104336

BLink Results

0 Archaea | 0 Bacteria | 63 Metazoa | 27 Fungi | 40 Plants | 1 Viruses | 5 Other Eukaryotae

Keep only [ ] Cut-Off 100 Select Reset

466 aa

SCORE	P	ACCESSION	GI	PROTEIN DESCRIPTION
2378	26	XP 032383	14785185	hypothetical protein [Homo sapiens]
1825	26	AAF29030	6841354	HSPC058 [Homo sapiens]
1489	26	BAA91664	7022613	unnamed protein product [Homo sapiens]
737	7	AAF60646	7331958	contains similarity to several yeast and human
630	3	CAB87659	7573353	putative protein [Arabidopsis thaliana]
630	3	T48545	11282308	hypothetical protein F14F18.30 - Arabidopsis th
630	3	NP 196747	15239800	putative protein [Arabidopsis thaliana]
471	3	BAB63547	15289850	contains ESTs AU092190(C11346),AU062476(C11346)
449	3	BAB19125	11761135	putative HSPC058 [Oryza sativa]
406	18	AAF17484	6572958	NLI-interacting factor isoform R5; NLI/Ldb1/CLI
406	18	AAF17482	6572954	NLI-interacting factor isoform T2; NLI/Ldb1/CLI
405	26	BAA21667	2289786	HYA22 [Homo sapiens]
405	26	NP 005799	5031775	HYA22 protein [Homo sapiens]
403	3	AAD28548	4731912	development protein DG1148 [Dictyostelium disco
399	4	CAA97541	1360322	ORF YLR019w [Saccharomyces cerevisiae]
391	4	S64841	2131751	hypothetical protein YLR019w - yeast (Saccharom
[illegible]	[illegible]	NP 013119	6323047	Psr2p [Saccharomyces cerevisiae]
[illegible]	18	AAF17481	6572952	NLI-interacting factor isoform T1; NLI/Ldb1/CLI
388	26	JC5707	7512494	HYA22 protein - human
388	4	CAA97454	1360175	ORF YLL010c [Saccharomyces cerevisiae]
388	4	NP 013091	6323019	Psr1p [Saccharomyces cerevisiae]

Hit to yeast

Entrez: Saccharomyces RefSeq Protein

```
LOCUS       NP_013119     397 aa                    PLN       11-AUG-2001
DEFINITION  Psr2p [Saccharomyces cerevisiae].
ACCESSION   NP_013119
PID         g6323047
VERSION     NP_013119.1  GI:6
DBSOURCE    REFSEQ: accession
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cer
            Eukaryota; Fungi;
            Saccharomycetacea
REFERENCE   1  (residues 1 to
  AUTHORS   Goffeau,A., Barre
            Feldmann,H., Gali
            Louis,E.J., Mewes
            Oliver,S.G.
  TITLE     Life with 6000 ge
  JOURNAL   Science 274 (5287
  MEDLINE   97002444
REFERENCE
  AUTHORS
            Ansorge,W., Benes
            Dusterhoft,A., En
            Heumann,K., Heuss
            Kotter,P., Louis,
            al.
  TITLE     The nucleotide se
  JOURNAL   Nature 387 (6632
  MEDLINE   97313267
REFERENCE   3  (residues 1 to
  AUTHORS   Saccharomyces Gen
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV
            University, Sacch
            USA
```

```
COMMENT     REFSEQ: This reference sequence was provided by the Saccharomyces
            Genome Database (SGD).
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..397
                     /organism="Saccharomyces cerevisiae"
                     /strain="S288C"
                     /db_xref="taxon:4932"
                     /chromosome="XII"
     Protein         1..397
                     /product="Psr2p"
     CDS             1..397
                     /gene="PSR2"
                     /db_xref="GeneID:YLR019W"
                     /db_xref="SGD:S0004009"
                     /codon_start=1
                     /protein_id="NC_001144.2:180287..181480"
ORIGIN
        1 mgfianilcc ssdtskthrq rqppetnhnr nrnrkhssnk aqtqgrkqka tpngdkmqys
       61 tpeillsssd sgsnagsktm qengnsgngk laplsrdhsn nsydeekeye dynegdvemt
      121 evnnageeee eddeakekqd hvvheynvda drnssindea ppqqglyqvg qedmnpqyva
      181 sspdndlnli ptteedfsdl thlqpdqyha pgydtllppk lqefqqkkcl ildldetlvh
      241 ssfkymhsad fvlpveiddq vhnvyvikrp gvdeflnrvs qlyevvvfta svsryanpll
      301 dtldpngtih hrlfreacyn yegnyiknls qigrplseti ildnspasyi fhpqhavpis
      361 swfsdthdne lldiiplled lssgnvldvg svldvti
//
```

Saccharomyces Genome Database

Yeast Homologue Function

PSR2 BASIC INFORMATION | PSR2 RESOURCES

Standard Name
Systematic Name
Feature Type
GO Annotations
Molecular Funct
Biological Proc
Cellular Compo
Description
Phenotype
Position
External Links
Primary SGDID

1: J Biol Chem 2000 Jun 23;275(25):19352-19360 — Related Articles, Books, LinkOut

FREE full text article at www.jbc.org

Psr1p/Psr2p, two plasma membrane phosphatases with an essential DXDX(T/V) motif required for sodium stress response in yeast.

Human Function?

Regulator of membrane pump expression in renal tubules?

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom.

Regulation of intracellular ion concentration is an essential function of all cells. In this study, we report the identification of two previously uncharacterized genes, PSR1 and PSR2, that perform an essential function under conditions of sodium ion stress in the yeast Saccharomyces cerevisiae. Psr1p and Psr2p are highly homologous and were identified through their homology with the endoplasmic reticulum membrane protein Nem1p. Localization and biochemical fractionation studies show that Psr1p is associated with the plasma membrane via a short amino-terminal sequence also present in Psr2p. Growth of the psr1psr2 mutant is severely inhibited under conditions of sodium but not potassium ion or sorbitol stress. This growth defect is due to the inability of the psr1psr2 mutant to properly induce transcription of ENA1/PMR2, the major sodium extrusion pump of yeast cells. We provide genetic evidence that this regulation is independent of the phosphatase calcineurin, previously implicated in the sodium stress response in yeast. We show that Psr1p contains a DXDX(T/V) phosphatase motif essential for its function in vivo and that a Psr1p-PtA fusion purified from yeast extracts exhibits phosphatase activity. Based on these data, we suggest that Psr1p/Psr2p, members of an emerging class of eukaryotic phosphatases, are novel regulators of salt stress response in yeast.

Mouse Genome Resources - Sequencing

NCBI — Mouse Genome Resources

PubMed | Entrez | BLAST | OMIM | Taxonomy | Structu

Search LocusLink for mouse Go

Genomic Biology

Resources
LocusLink
MapViewer

Homology Information
Human/Mouse Homology Map

Sequences
Mouse Genome Sequencing
Trace Archive
Human Genome Sequencing
Reference mRNA Sequences
Unigene

Mouse Genome Resources

Welcome to the mouse genome resources page. This homepage will bring together information on diverse mouse-related resources from multiple centers: sequence, mapping and clone information as well as pointers to strain and mutant resources. We encourage your suggestions. Stay tuned for more information.

What's New...

30 July 2001 Genetic and Radiation Hybrid maps available in the MapViewer.
More...
30 Apr 2001 Add your annotations to LocusLink records using GeneRIF.
More...
12 Jan 2001 Whole Genome Shotgun sequencing reads generate by the Mouse Sequencing Consortium are available from the Trace Archive
More...
2 May 2000 Homologene is a new resource which displays calculated and curated homologies for human, mouse, rat and zebrafish
More

Trans NIH M Initiative

Mouse Genor Monthly

Strain Resources
IMSR
The Jackson Laboratory
Harwell Mo Resource
EMMA
ORNL
Genealogy Chart

Mapping Informatio
Mouse BAC Fingerprints

NCBI — Mouse Genome Sequencing

Chrs: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X Y

Search Contigs for Go Mouse

Mouse Genome Resources

Access to Data
BLAST
FTP Site
Trace Archive
Mouse Genome Data
Current Progress
Clone Registry
Genome Sequencing Centers

Mouse Genome Sequencing

Welcome to the Mouse Genome Sequencing Page.

Mouse Genome Sequencing Resources currently available:

Finished Sequence Assembled into Contigs:
Annotated with: STS markers
SNPs

Mouse specific sequences available for BLAST:

- curated NT contigs (non-redundant finished sequence)
- HTGS sequence (phase 0,1,2&3 BACs)
- Two WGS assemblies
- All traces in the Trace Archive
- Reference mRNAs
- Reference Proteins
- ESTs
- BAC end sequences

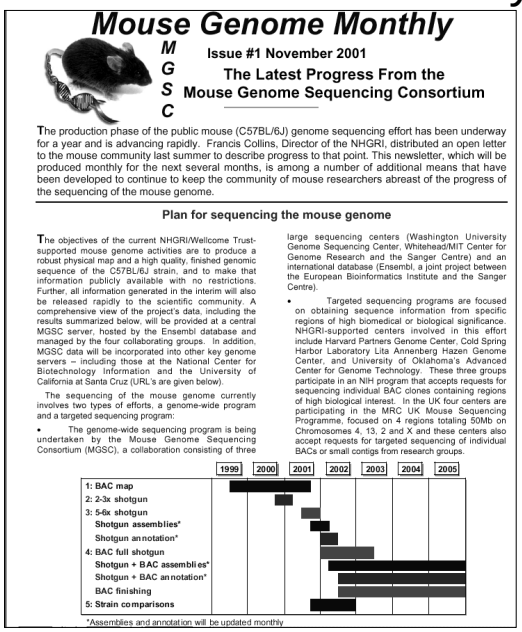
Coming Soon: Assembly of Draft Contigs
Annotation of known genes

6 December 2001
Finished Sequence
75.5 Mb
2.5 % of genome
Draft Sequence
514.2 Mb
17.1 % of genome
Whole Genome Shotgun (WGS)
30,043,512 reads
More Statistics...

Trans-NIH BAC Sequencing Program- apply to get your BAC sequenced!

NCBI

Mouse Genome Monthly



Mouse Genome Monthly

M G S C

Issue #1 November 2001

The Latest Progress From the Mouse Genome Sequencing Consortium

The production phase of the public mouse (C57BL/6J) genome sequencing effort has been underway for a year and is advancing rapidly. Francis Collins, Director of the NHGRI, distributed an open letter to the mouse community last summer to describe progress to that point. This newsletter, which will be produced monthly for the next several months, is among a number of additional means that have been developed to continue to keep the community of mouse researchers abreast of the progress of the sequencing of the mouse genome.

Plan for sequencing the mouse genome

The objectives of the current NHGRI/Wellcome Trust-supported mouse genome activities are to produce a robust physical map and a high quality, finished genomic sequence of the C57BL/6J strain, and to make that information publicly available with no restrictions. Further, all information generated in the interim will also be released rapidly to the scientific community. A comprehensive view of the project's data, including the results summarized below, will be provided at a central MGSC server, hosted by the Ensembl database and managed by the four collaborating groups. In addition, MGSC data will be incorporated into other key genome servers – including those at the National Center for Biotechnology Information and the University of California at Santa Cruz (URL's are given below).

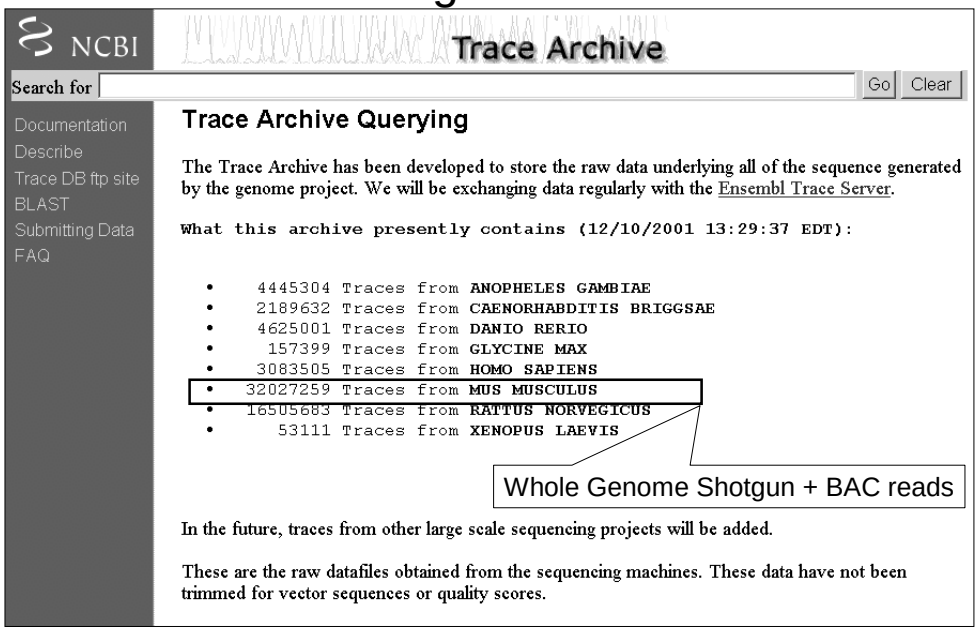
The sequencing of the mouse genome currently involves two types of efforts, a genome-wide program and a targeted sequencing program:

- The genome-wide sequencing program is being undertaken by the Mouse Genome Sequencing Consortium (MGSC), a collaboration consisting of three large sequencing centers (Washington University Genome Sequencing Center, Whitehead/MIT Center for Genome Research and the Sanger Centre) and an international database (Ensembl, a joint project between the European Bioinformatics Institute and the Sanger Centre).
- Targeted sequencing programs are focused on obtaining sequence information from specific regions of high biomedical or biological significance. NHGRI-supported centers involved in this effort include Harvard Partners Genome Center, Cold Spring Harbor Laboratory Lita Annenberg Hazen Genome Center, and University of Oklahoma's Advanced Center for Genome Technology. These three groups participate in an NIH program that accepts requests for sequencing individual BAC clones containing regions of high biological interest. In the UK four centers are participating in the MRC UK Mouse Sequencing Programme, focused on 4 regions totaling 50Mb on Chromosomes 4, 13, 2 and X and these centers also accept requests for targeted sequencing of individual BACs or small contigs from research groups.

*Assemblies and annotation will be updated monthly

NCBI

Shotgun Reads



NCBI

Trace Archive

Search for Go Clear

Documentation
Describe
Trace DB ftp site
BLAST
Submitting Data
FAQ

Trace Archive Querying

The Trace Archive has been developed to store the raw data underlying all of the sequence generated by the genome project. We will be exchanging data regularly with the Ensembl Trace Server.

What this archive presently contains (12/10/2001 13:29:37 EDT):

- 4445304 Traces from **ANOPHELES GAMBIAE**
- 2189632 Traces from **CAENORHABDITIS BRIGGSAE**
- 4625001 Traces from **DANIO RERIO**
- 157399 Traces from **GLYCINE MAX**
- 3083505 Traces from **HOMO SAPIENS**
- 32027259 Traces from **MUS MUSCULUS**
- 16505683 Traces from **RATTUS NORVEGICUS**
- 53111 Traces from **XENOPUS LAEVIS**

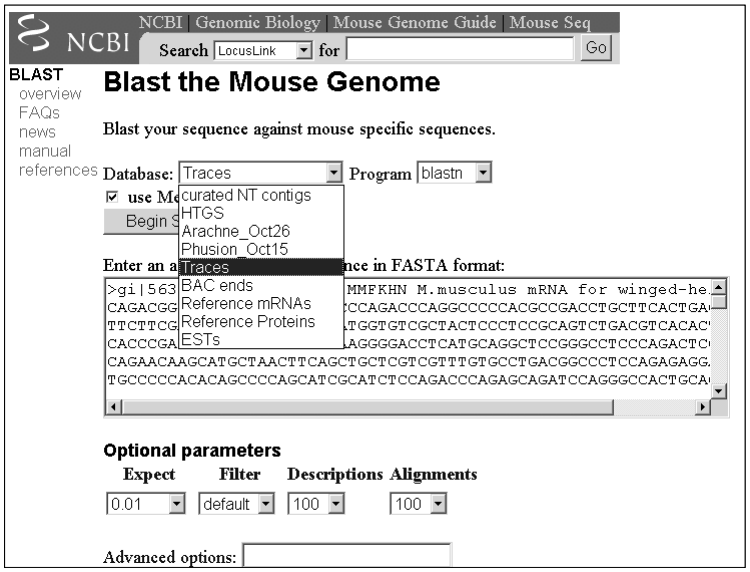
Whole Genome Shotgun + BAC reads

In the future, traces from other large scale sequencing projects will be added.

These are the raw datafiles obtained from the sequencing machines. These data have not been trimmed for vector sequences or quality scores.

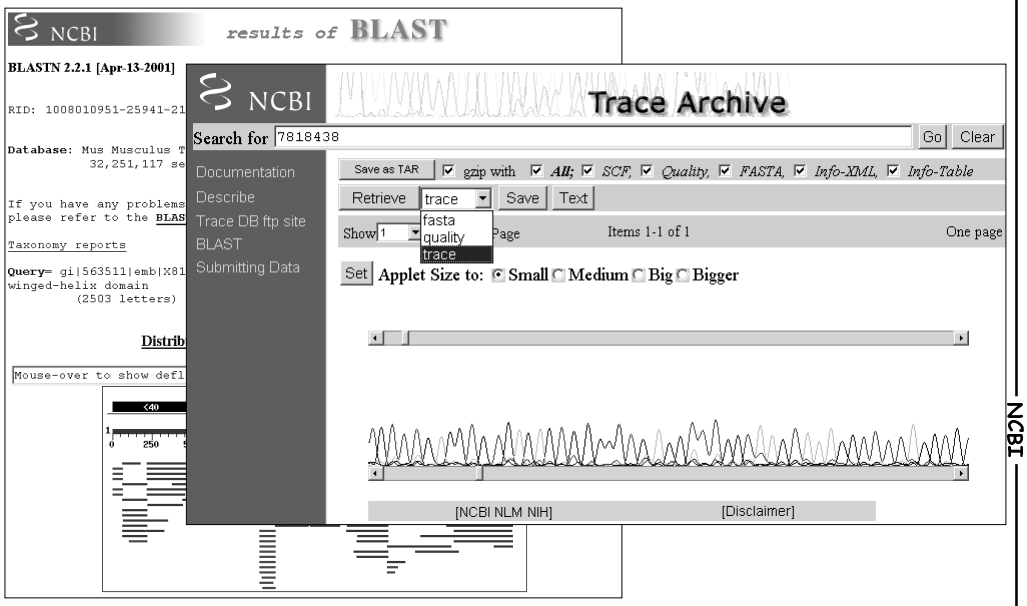
NCBI

Mouse Genome BLAST Page



NCBI

Whole Genome Shotgun



NCBI

Service Addresses

•*General Help*	`info@ncbi.nlm.nih.gov`
•*Updates to records*	`update@ncbi.nlm.nih.gov`
•*Questions about BLAST*	`blast-help@ncbi.nlm.nih.gov`
•*Sequin submissions*	`gb-sub@ncbi.nlm.nih.gov`
•*Batch Submissions*	`batch-sub@ncbi.nlm.nih.gov`

E-mail Servers

•*BLAST Server*	`blast@ncbi.nlm.nih.gov`
•*Query Server*	`query@ncbi.nlm.nih.gov`

NCBI