

Credits to NCBI Folks

Database resources of the National Center for Biotechnology Information

Eric W. Sayers*, Tanya Barrett, Dennis A. Benson, Stephen H. Bryant, Kathi Canese, Vyacheslav Chetverinn, Deanna M. Church, Michael DiRuccio, Ron Edgar, Scott Federhen, Michael Feolo, Lewis Y. Geer, Wolfgang Heinberg, Yuri Kapustin, David Landsman, David J. Lipman, Thomas L. Madden, Donna R. Maglott, Vadim Miller, Irena Mizrahi, James Ostell, Kim D. Pruitt, Gregory D. Schuler, Edwin Sequeira, Stephen T. Sherry, Martin Shumway, Karl Sirotkin, Alexandre Souvorov, Grigory Starchenko, Tatiana A. Tatusova, Lukas Wagner, Eugene Yaschenko and Jian Ye

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 9609 Rockville Pike, Bethesda, MD 20894, USA

Received September 16, 2008; Revised October 1, 2008; Accepted October 2, 2008

ABSTRACT
In addition to maintaining the GenBank® nucleotide sequence database, the National Center for Biotechnology Information (NCBI) provides analysis and retrieval resources for the data in GenBank and other biological data made available through the NCBI web site. NCBI resources include Entrez, the Entrez Programming Utilities, eQuery, PubMed, PubMed Central, BioProject, and the NCBI Taxonomy.

INTRODUCTION
The National Center for Biotechnology Information (NCBI) at the National Institutes of Health was created in 1988 to develop information systems for molecular biology. In addition to maintaining the GenBank® (1) nucleotide sequence database, which receives data through the international collaboration with DNA Database of Japan (DDBJ) and European Molecular Biology Laboratory (EMBL) as well as from the scientific community...

http://nar.oxfordjournals.org/cgi/content/full/37/suppl_1/D5

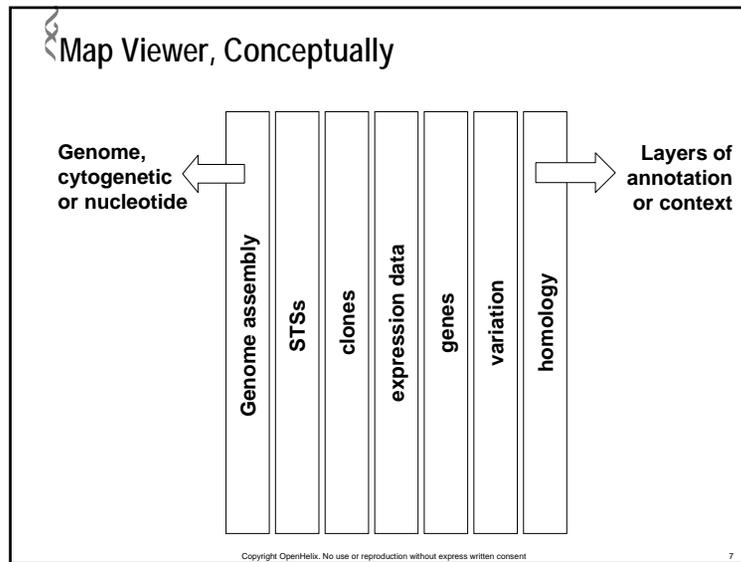
- Map Viewer is only one of the many projects at NCBI
- Learn more about the NCBI offerings in the annual Nucleic Acids Research database issue

Copyright OpenHelix. No use or reproduction without express written consent. 5

Map Viewer Tools

- Compare maps in different coordinate systems (genetic maps vs sequence maps)
- Robust query interface (basic, advanced, text or sequence searches)
- Configure the display for your needs
- Reports, and download maps and annotations
- Customize the transcript sequence (Model Maker)
- Detailed descriptions and links for the objects you see

Copyright OpenHelix. No use or reproduction without express written consent. 6



Sample Image of Map Viewer

Master Map: Genes On Sequence

Chromosome: 7

Region Displayed: 0-159,011 bp

Map	Gene	Symbol	Link	E	Cyto	Description
LOC100128036	LOC100128036		or	dsrtran	mRNA	7p15.2 similar to ubiquitin-conjugating enzyme 2
CCMG2	CCMG2		or	dsrtran	best RefSeq	7p13 cerebral cavernous malformation 2
LOC729527	LOC729527		or	dsrtran	best RefSeq	7p13-p11 GDP-association inhibitor 2 protein
BTBD1	BTBD1		or	dsrtran	best RefSeq	7q11.23 brain domain (beta)-like 2
NEFC1	NEFC1		or	dsrtran	best RefSeq	7q11.23 neurofil cytosolic factor 1C protein
FOXL2	FOXL2		or	dsrtran	best RefSeq	7q11.23 Brn3-like 2
CRBOT	CRBOT		or	dsrtran	best RefSeq	7q21.1 carnitine O-octanoyltransferase
GAATAD1	GAATAD1		or	dsrtran	best RefSeq	7q21.1-q22 GATA zinc finger domain containing
GRB11	GRB11		or	dsrtran	best RefSeq	7q21 guanine nucleotide binding protein (G)
MGC112925	MGC112925		or	dsrtran	best RefSeq	7q22.1 similar to Wilms tumor suppressor phosphonuclein-3 isoform, candidate
CAV1	CAV1		or	dsrtran	best RefSeq	7q31.1 caveolin 1, caveolar protein, 22kDa
RNF148	RNF148		or	dsrtran	best RefSeq	7q31.3 ring finger protein 148
LOC100131317	LOC100131317		or	dsrtran	best RefSeq	7q31.3 Sac72 RNA polymerase II CTD phi
LOC100131312	LOC100131312		or	dsrtran	mRNA	7q32.1 hypodermal protein LOC100131312
WDR91	WDR91		or	dsrtran	best RefSeq	7q33 WD repeat domain 91
M30AM	M30AM		or	dsrtran	mRNA	7q34 malate-gluconate lyase (alpha-glucon)

Maps **Links**

Copyright OpenHelix. No use or reproduction without express written consent. 8

Sample Search: Human TP53

■ Diagram shows location of hits and numbers of results (red)

■ Click the Gene filter to simplify results for this example

■ Text format output of the same results as diagram

Copyright OpenHelix. No use or reproduction without express written consent

Closer Look at Text Results

Search results for query "TP53 AND gene[ob_type]" : 25 hits

Chr	Assembly	Match	Map element	Type	Maps
2	reference	represses TP53 dependent G2 arrest mediator candidate	RFFEM	GENE	Genes_cyto Genes_seq
2	Celera	represso, TP53 dependent			
6	HuRef	represso, TP53 dependent			
6	reference	p53-induced protein P1			
6	Celera	p53-induced protein P1			
7	reference	TP53AIP1, TP53 target			
7	Celera	TP53AIP1, TP53 target			
8	reference	ribonucleotide reductase			
8	Celera	ribonucleotide reductase			
8	HuRef	p53-inducible cell-survival regulator of TP53			
12	reference	all matches			
12	Celera	all matches	LETMD1	GENE	Genes_cyto Genes_seq
12	HuRef	p53-inducible cell-survival factor regulator of TP53	TRAP1	GENE	craGenes Genes_seq
12	reference	all matches	LETMD1	GENE	craGenes Genes_seq
12	HuRef	p53-inducible cell-survival factor regulator of TP53	TRAP1	GENE	Genes_seq
12	reference	all matches	LETMD1	GENE	Genes_seq
17	reference	Breast cancer-related regulator, TP53	RCPB	GENE	Genes_cyto
17	Celera	TP53, tumor protein p53 (Li-Fraumeni syndrome)	TP53	GENE	
17	reference	TP53, ENS00000141510	TP53	GENE	Genes_cyto
17	Celera	TP53, tumor protein p53 (Li-Fraumeni syndrome)	TP53	GENE	craGenes Genes_seq

■ Assembly, different groups provide genome sequences

■ Reference assembly will be shown in our example

Click Map element = TP53
Type = Gene to go to the Map Viewer location

Copyright OpenHelix. No use or reproduction without express written consent

Map Viewer Overview

■ Default maps shown; you can add/remove any maps

Copyright OpenHelix. No use or reproduction without express written consent

Map Viewer, Zoom to Show 100k View

■ From previous results, chose zoom to show 100k region

■ Image shows the TP53 region (pink highlight)

■ For info on Maps, click the Map names

■ Links offer additional information as well

Copyright OpenHelix. No use or reproduction without express written consent

TP53 Links to Entrez Gene

■ Gene symbol links to Entrez Gene

NCBI Entrez Gene

1: TP53 tumor protein p53 [Homo sapiens]

Official Symbol: TP53

Official Full Name: tumor protein p53

Primary Source: MIM:113258

Gene type: protein coding

RefSeq status: HOMOLOG

Organism: HOMO SAPIENS

Summary: This gene encodes tumor protein p53, which responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or change gene expression. p53 protein is expressed at low levels in normal cells and at a high level in a variety of transformed cell lines, where it's believed to contribute to transcriptional regulation. p53 is a DNA-binding protein containing transcription activation, DNA-binding, and oligomerization domains. It is postulated to bind to a p53-binding site and activate expression of downstream genes that inhibit growth and/or invasion, and thus function as a tumor suppressor. Mutations of p53 that frequently occur in a number of different human cancers fail to bind the consensus DNA-binding site, and hence cause the loss of tumor suppressor activity. Alterations of the gene occur not only as somatic mutations in human malignancies, but also as germline mutations in some cancer-prone families with Li-Fraumeni syndrome. Multiple p53 variants due to alternative promoters and multiple alternative splicing have been found. These variants encode distinct

Copyright OpenHelix. No use or reproduction without express written consent

Entrez Gene, Overview

1: TP53 tumor protein p53 [Homo sapiens]

HIV-1 protein interactions

Phenotypes: Adrenal cortical cancer

Homology: Adrenal cortical cancer

NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

mRNA and Protein(s)

1. NM_000546.4 NP_000537.3 tumor protein p53 isoform a

Related Sequences

Nucleotide	Protein
Genomic: AB118156.1	BAD1809.1
Genomic: AC007422.13 (4-3751)	None
Genomic: AC007308.2 (67932-42741)	None
Genomic: AF155121.1	AAD28535.1

Additional Links

- MIM 191170
- PharmGKB PA36679
- GeneTests for MIM: 191170
- Genes and Disease p53.html
- Genes and Disease Colon.html
- HPD 01859
- p53 Mutation Database p53 Mutation Database
- IARC TP53 Mutation Database IARC TP53 Mutation Database
- UMC Locus Specific Databases UMC Locus Specific Databases
- UCSC IUCS
- Database of Germline p53 Mutations Database of Germline p53 Mutations
- UniGene 15164461

Copyright OpenHelix. No use or reproduction without express written consent

Genes_seq LinkOut Options, TP53

OMIM HGNC sv

OMIM: Online Mendelian Inheritance in Man
HGNC: HUGO gene nomenclature committee
sv: sequence viewer

Symbol Report: TP53

Homo sapiens chromosome 17 genomic contig reference assembly

Approved Symbol: TP53

Approved Name: tumor protein p53 a

HGNC ID: HGNC:11398

Approved Status: Approved

Chromosome: 17p13.1

Provisional Symbol: TP53P1

Provisional Name: TP53P1

Accession: P53

Copyright OpenHelix. No use or reproduction without express written consent

Genes_seq LinkOut Options, TP53

pr dl ev

pr: Entrez Protein records
dl: download sequences
ev: evidence viewer

NCBI Evidence Viewer

Homo sapiens TP53

Key for display of mRNAs aligning in this region:

- Chromatin sequence (C)
- no-del exon, single (M)
- no-del exon, contigging (M)
- C = contig, M = model mRNA, E = RefSeq mRNA, G = GenBank mRNA
- E = exon repeat last genomic build, G = up-dated since last genome build

EST density key (E):

- 1-5 EST
- 6-20 ESTs
- 21-99 ESTs
- 100+ ESTs

Aligning 52 models, with an average of 7 exons each. Please be patient.

13 exons and 1 gene found in the genomic region spanning 19990 bp.

View genomic context

Copyright OpenHelix. No use or reproduction without express written consent

Map Viewer Agenda

- Introduction and Credits
- Basic Searches
- Map Options
- Additional Searches
- BLASTing a Genome
- Summary
- Exercises

Map Viewer: <http://www.ncbi.nlm.nih.gov/mapview/>

Maps and Options to Add/Remove Tracks

- Maps and Options
- Change default maps, or add new ones

Add Additional Detail Using Maps and Options Tool

Add Additional Detail Using Maps and Options Tool

Add Additional Detail Using Maps and Options Tool

2 new maps added (*M* = Morbid map, *V* = Variation map)

Can put them in any order you like

Copyright OpenHelix. No use or reproduction without express written consent

Different Species, Different Maps, Same Software

Software functions are the same; data types may vary

Copyright OpenHelix. No use or reproduction without express written consent

Map Viewer Agenda

- Introduction and Credits
- Basic Searches
- Map Options
- Additional Searches
- BLASTing a Genome
- Summary
- Exercises

Map Viewer: <http://www.ncbi.nlm.nih.gov/mapview/>

Map Viewer Advanced Search

Human shown, others vary

Limit to certain data types

More refined searches

Copyright OpenHelix. No use or reproduction without express written consent

Advanced Search: Kinase, Human Chromosome 3

Advanced Search: Kinase, Human Chromosome 3

Search for: Find Clear

Homo sapiens (human) advanced search

Search in fields:
 no restriction
 ID of mapped object
 symbol
 name of mapped object
 MIM number
 name of phenotype

Type of mapped object: select all clear
 Clone Contig Gene Gene Prediction
 Marker SNP STS Sequence Transcript

Chromosome(s): select all clear
 1 2 3 4 5 6
 7 8 9 10 11 12
 13 14 15 16 17 18
 19 20 21 22 X Y
 MT Unknown

Assembly: select all clear
 reference CRA_TOAGchr7v2
 Celera DR53
 HbRef c22_H2
 c5_H2 c6_COX

Search results for query "kinase[TITLE]": 1480 hits shown (out of 4424 found)
 Hits shown: 1 - 100

Chr	Assembly	Match	Map element	Type	Maps
3	reference	All matches			
		Protein tyrosine kinase 9-like (A6-related protein) (predicted)	Rn_218510	Rn_EST_Cl	Rn_UncG
		Myosin light polypeptide kinase (predicted)	Rn_203004	Rn_EST_Cl	Rn_UncG
		Inositol hexaphosphate kinase 2	Rn_112474	Rn_EST_Cl	Rn_UncG
		Tyrosine kinase, non-receptor, 2	Rn_28332	Rn_EST_Cl	Rn_UncG
		Protein kinase 5, delta	Rn_28272	Rn_EST_Cl	Rn_UncG

Copyright OpenHelix. No use or reproduction without express written consent. 33

Clone Finder – Access Tool From Homepage

Clone Finder – Access Tool From Homepage

NCBI Home GenBank BLAST

Map Viewer Home Homo sapiens

Search by Position **Search by Feature**

Region: All From: Any Feature name: is is

Assembly: reference To: Gene is is

Clone SNP Marker Transcript

Specify Region: Chromosome: 17 From: 7,498Kb To: 7,568Kb

Range: 2,015 K 2,040 K 2,115 K

What is Clone Finder?
 Clone Finder is a tool developed to facilitate the identification of clones within a given genomic region.
 • Regions on a given assembly can be defined either by providing a specific location, or searching for a feature (gene, STS, clone, etc) of interest.
 • Once a region is defined, the libraries of interest can be specified and the region viewed in the browser (see Figure 1).
 • Information concerning the assembly and gene annotation are provided at the top of the display.
 • Concordant and discordant clone placements for each library are then displayed.
 • PopUp boxes provide additional information and links for each feature.
 • Tabulated clone placement data can also be viewed and downloaded.

concordant clones displayed as features on histograms

Tools (17) (14) (3)

Copyright OpenHelix. No use or reproduction without express written consent. 34

Clone Finder Search Results

Clone Finder Search Results

Homo sapiens Clone Finder
 Assembly: reference (Celera) Coordinates: Chromosome 17 NC_000017.8 7,498,000-7,568,000 bp

Download

Annotation

Concordant clones

Discordant clones

Concordant clone histograms

Discordant clone histograms

Copyright OpenHelix. No use or reproduction without express written consent. 35

Jumping to a Chromosomal Region

Jumping to a Chromosomal Region

NCBI Context Help

Map Viewer Home

Go to Region

This area allows you to quickly jump to a region on a chromosome. Select chromosome and assembly and specify chromosome coordinates. Specify coordinate as a number optionally followed by a unit (case insensitive).
 • M - for Mega-basepairs (1,000,000 basepairs)
 • K - for Kilo-basepairs (1,000 basepairs)
 • You can also use decimal point (.) and comma (,) separators.
 The following examples specify the same coordinate:
 • 2,450M
 • 2,456K
 • 2,456,000
 • 2456000

Specify Region

Homo sapiens Build 36.3

Search by Position

Chromosome: 17

Assembly: reference

Range: From: 4,550K To: 4,750K

Region displayed

BLAST The Human Genome

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

Region Displayed: 4,550K-4,750K bp

Download View Sequence/Evidence

Copyright OpenHelix. No use or reproduction without express written consent. 36

Map Viewer Agenda

- Introduction and Credits
- Basic Searches
- Maps Options
- Additional Searches
- BLASTing a Genome
- Summary
- Exercises

Map Viewer: <http://www.ncbi.nlm.nih.gov/mapview/>

Copyright OpenHelix. No use or reproduction without express written consent

37

BLASTing a Genome from Map Viewer Homepage

The screenshot shows the NCBI Map Viewer homepage. The search bar contains 'Homo sapiens'. The search results are displayed in a table with columns for 'Scientific name', 'Common name', 'Build', and 'Tools'. The 'Homo sapiens (human)' entry is highlighted with a red box. A red arrow points to this entry.

■ BLAST: can choose a single species

■ Sometimes groups (notice Protozoa, Fungi, Bacteria, others)

Copyright OpenHelix. No use or reproduction without express written consent

38

BLASTing just the Human Genome, Part I

The screenshot shows the NCBI BLAST Human Sequences page. The search bar contains 'NM_000546.2'. The 'Database' dropdown is set to 'genome (all assemblies)'. The 'Expect' value is 0.01. The 'Program' is 'blastn'. The 'Advanced options' section is expanded, showing 'Alignments' and 'Display' options.

■ BLAST link next to *Homo sapiens* in species list

■ Default parameters in this example

Copyright OpenHelix. No use or reproduction without express written consent

39

BLASTing just the Human Genome, Part II

The screenshot shows the NCBI BLAST results page for 'Homo sapiens (human)'. The search results show a list of alignments with columns for 'Query', 'Hit', 'Score', 'E value', and 'Map Viewer'. The 'Map Viewer' column contains links to view the location of the hits in the genome.

■ BLAST results page gives alignments, links

■ Also provides the option to view location in Map Viewer

Copyright OpenHelix. No use or reproduction without express written consent

40

Map Viewer Agenda

- Introduction and Credits
- Basic Searches
- Map Options
- Additional Searches
- BLASTing a Genome
- Summary
- Exercises

Map Viewer: <http://www.ncbi.nlm.nih.gov/mapview/>

Map Viewer – Search, Retrieve & Display Your Data

Context for genomic data

Extensive links

Putative exons (graphic view):

Your model: 6-7-10-11-16

Copyright OpenHelix. No use or reproduction without express written consent

Lots of Options

Basic or advanced searches

BLAST searches, access maps

Use Maps & Options to add other maps

Copyright OpenHelix. No use or reproduction without express written consent

Map Viewer Agenda

- Introduction and Credits
- Basic Searches
- Map Options
- Additional Searches
- BLASTing a Genome
- Summary
- Exercises

Map Viewer: <http://www.ncbi.nlm.nih.gov/mapview/>

