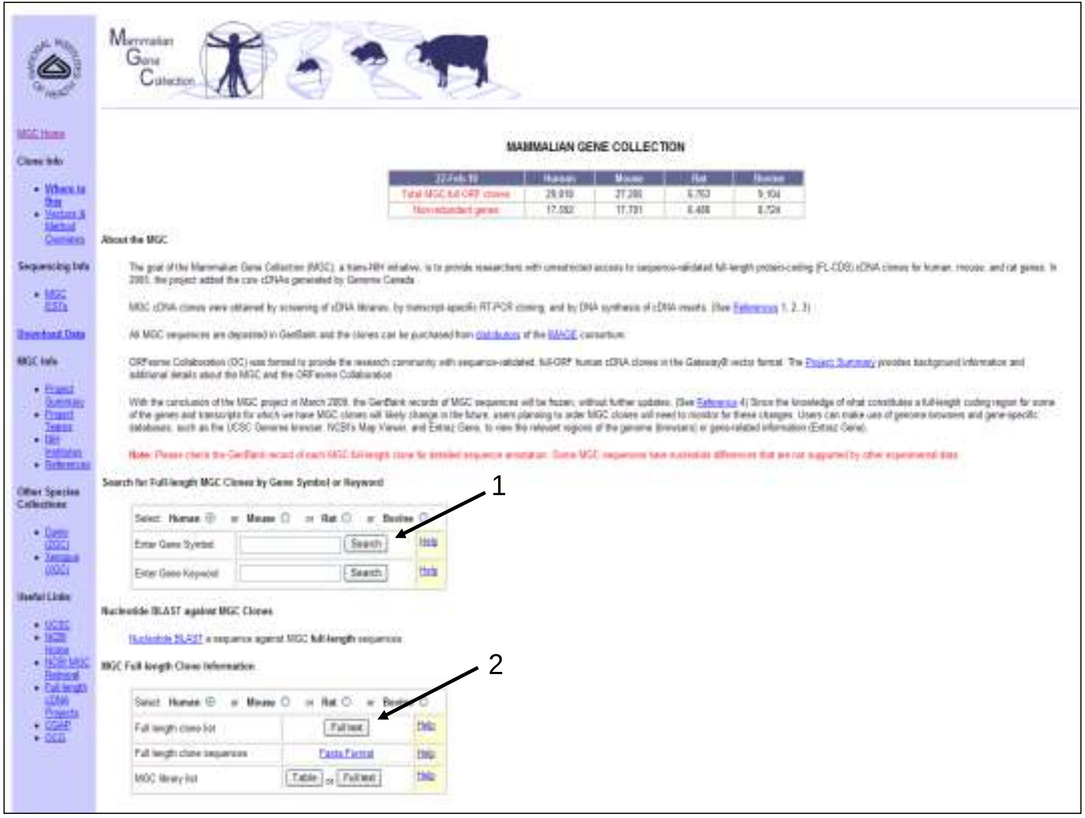
A Guide to Finding Mammalian Gene Collection (MGC) Clones and Evaluating Their Sequence

Part A. MGC Clone Search. A variety of ways exist to determine whether MGC cDNA clones are available for human, mouse, and rat genes and transcripts of interest[1]. Here we describe three approaches. We illustrate by searching for MGC cDNA clones for protein-coding transcripts from the human gene SERPINA1, encoding α-1-anti-trypsin protein.

(As per HUGO-defined convention, all letters of human gene names are capitalized, but only the first letter of mouse and rat gene names is capitalized. Entries into the search engines described below are case insensitive.)

Approach 1. The MGC homepage (Figure 1) provides several search tools. You can search for individual full-protein coding (full-cds) human, mouse, rat, or bovine clones from this page using gene names or key words. Entering SERPINA1 into the **Enter Gene Symbol** box (Figure 1, arrow 1) opens a page that shows two MGC clones are available, BC011991 and BC015642 (Figure 2), together with names of the libraries from which they were isolated and links to associated vector and source tissue information.

Figure 1 **MGC Website Home Page**



MAMMALIAN GENE COLLECTION

22-Feb-10	Human	Mouse	Rat	Bovine
Total MGC full ORF clones	29,919	27,200	6,753	9,104
Nonredundant genes	17,582	17,791	6,408	8,728

About the MGC

The goal of the Mammalian Gene Collection (MGC), a trans-NIH initiative, is to provide researchers with unrestricted access to sequence-validated full-length protein-coding (FL-CDS) cDNA clones for human, mouse, and rat genes. In 2005, the project added the cow cDNAs generated by Genome Canada.

MGC cDNA clones were obtained by screening of cDNA libraries, by transcript-specific RT-PCR cloning, and by DNA synthesis of cDNA inserts. (See References 1, 2, 3)

All MGC sequences are deposited in GenBank and the clones can be purchased from distributors of the IMAGE consortium.

ORFeome Collaboration (OC) was formed to provide the research community with sequence-validated, full-ORF human cDNA clones in the Gateway® vector format. The Project Summary provides background information and additional details about the MGC and the ORFeome Collaboration.

With the conclusion of the MGC project in March 2009, the GenBank records of MGC sequences will be frozen, without further updates. (See Reference 4) Since the knowledge of what constitutes a full-length coding region for some of the genes and transcripts for which we have MGC clones will likely change in the future, users planning to order MGC clones will need to monitor for these changes. Users can make use of genome browsers and gene-specific databases, such as the UCSC Genome browser, NCBI's Map Viewer, and Entrez Gene, to view the relevant regions of the genome (browsers) or gene-related information (Entrez Gene).

Note: Please check the GenBank record of each MGC full-length clone for detailed sequence annotation. Some MGC sequences have nucleotide differences that are not supported by other experimental data.

Search for Full-length MGC Clones by Gene Symbol or Keyword

Select: Human ● or Mouse ○ or Rat ○ or Bovine ○		
Enter Gene Symbol	[ ] Search	Help
Enter Gene Keyword	[ ] Search	Help

Nucleotide BLAST against MGC Clones

Nucleotide BLAST a sequence against MGC full-length sequences

MGC Full-length Clone Information

Select: Human ● or Mouse ○ or Rat ○ or Bovine ○		
Full length clone list	Full text	Help
Full length clone sequences	Fasta Format	Help
MGC library list	Table or Full text	Help

[1] All newly isolated MGC, XGC, and ZGC clones are assigned a "BC" accession when their sequence is submitted to GenBank, but only a subset of these candidate clones has a full-CDS. Once a candidate clone is confirmed by full-length sequencing to have a full-CDS (without changes altering the phase of reading frame, the position of the translational start ATG codon, or the position of the stop codon) it is then assigned the Keyword "MGC" in a new GenBank record.

The MGC homepage also provides lists of MGC cDNA clones, as well as lists of MGC cDNA libraries (Figure 1, arrow 2). More detailed lists of MGC clones are provided at the MGC ftp site, described in **Part C**, below.

(The left side of the MGC Homepage provides links to the XGC and ZGC pages, with similar search functions for *Xenopus tropicalis* and *Xenopus laevis* clone or *Danio rerio* (zebrafish) clones.)

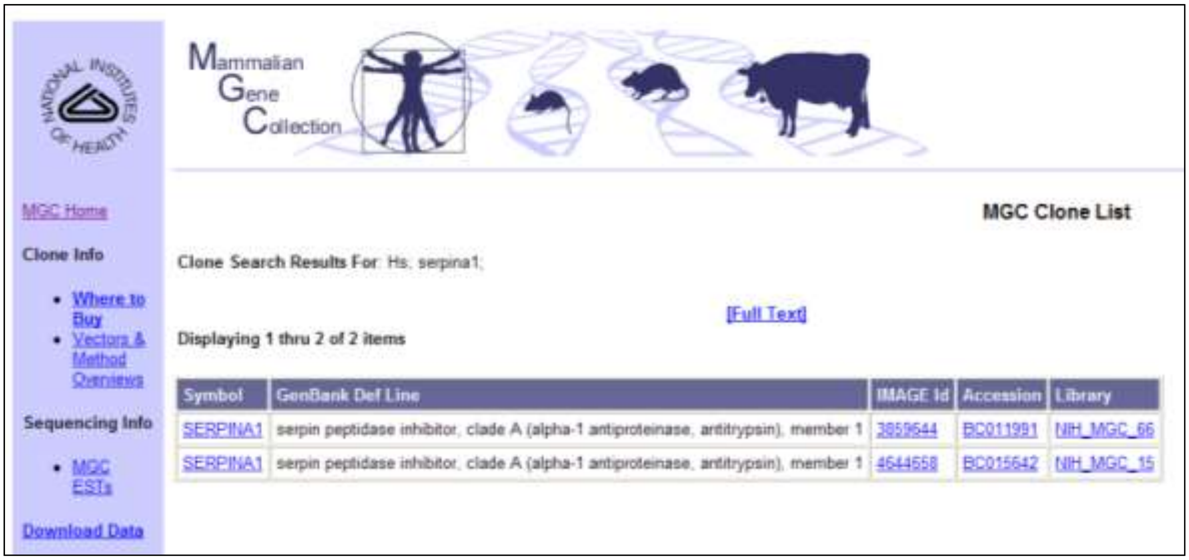
Clicking on the link for BC011991 leads to the GenBank record for this clone in Entrez Nucleotide (Figure 3a). This page provides many details about MGC cDNA clones, including information on how the clone was obtained, the cloning vector, tissue source, nucleotide sequence, and expected translated amino acid sequence. Near the top of the page, the **Definition** line (Figure 3a, arrow 1) notes that this clone includes a complete protein-coding sequence (cds).

(If this MGC clone had been prepared with a synthetic DNA insert, the **Definition** line also would indicate whether the natural protein-coding sequence was cloned with or without a stop codon, as illustrated by BC167860 and BC140303.)

The existence of RefSeq alternative splicing isoforms for this gene is noted on the right side of this page (Figure 3a, arrow 2), with links to information on the alternative splice isoforms. Further down on the right side, **All links from this record** shows links to other resources. These include **Order cDNA Clone** (Figure 3a, arrow 3), which connects to a clone order page with links to several commercial distributors of these two clones.

Other links under **All links from this record** connect to data and analysis tools, such as **dbSNP**, **GEO**, **Map Viewer**, **PubMed**, **UniGene**, **UniSTS**, **OMIM**, and **LinkOut**. **LinkOut** connects to gene-related research materials from commercial vendors, including antibodies, peptides, and siRNA reagents.

Figure 2 **Result of Search for SerpinA1 Gene on MGC Website**


NATIONAL INSTITUTES OF HEALTH
Mammalian Gene Collection
MGC Home
Clone Info
Where to Buy
Vectors & Method Overviews
Sequencing Info
MGC ESTs
Download Data

MGC Clone List

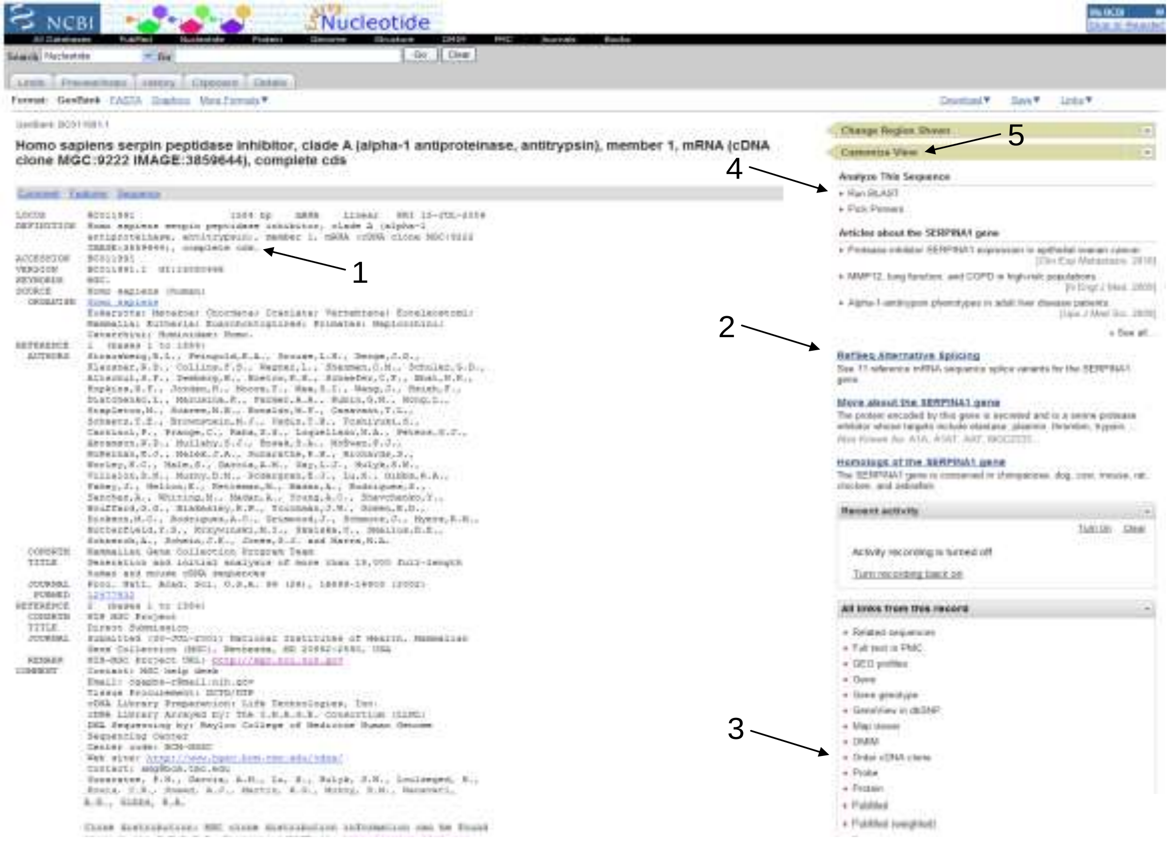
Clone Search Results For: Hs. serpina1;

[Full Text]

Displaying 1 thru 2 of 2 items

Symbol	GenBank Def Line	IMAGE Id	Accession	Library
SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	3859644	BC011991	NIH_MGC_66
SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	4644658	BC015642	NIH_MGC_15

Figure 3a GenBank Record for BC011991



GenBank: BC011991.1

Homo sapiens serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1, mRNA (cDNA clone MGC:9222 IMAGE:3859644), complete cds

```
LOCUS       BC011991                1364 bp    mRNA    linear   PRI [illegible]
DEFINITION  Homo sapiens serpin peptidase inhibitor, clade A (alpha-1
            antiproteinase, antitrypsin), member 1, mRNA (cDNA clone MGC:9222
            IMAGE:3859644), complete cds.
ACCESSION   BC011991
VERSION     BC011991.1  GI:[illegible]
KEYWORDS    MGC.
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 1364)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            [illegible]
  CONSRTM   Mammalian Gene Collection Program Team
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   [illegible]
REFERENCE   2  (bases 1 to 1364)
  CONSRTM   NIH MGC Project
  TITLE     Direct Submission
  JOURNAL   Submitted ([illegible]) National Institutes of Health, Mammalian
            Gene Collection (MGC), Bethesda, MD [illegible], USA
  REMARK    NIH-MGC Project URL: [illegible]
COMMENT     Contact: MGC help desk
            Email: [illegible]
            Tissue Procurement: [illegible]
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: [illegible]
            Contact: [illegible]
            [illegible]
            Clone distribution: MGC clone distribution information can be found
```

Articles about the SERPINA1 gene

- Protease inhibitor SERPINA1 expression in epithelial ovarian cancer [Clin Exp Metastasis. 2010]
- MMP12, lung function, and COPD in high-risk populations [N Engl J Med. 2009]
- Alpha-1-antitrypsin phenotypes in adult liver disease patients [Ups J Med Sci. 2009]

» See all...

RefSeq Alternative Splicing

See 11 reference mRNA sequence splice variants for the SERPINA1 gene

More about the SERPINA1 gene

The protein encoded by this gene is secreted and is a serine protease inhibitor whose targets include elastase, plasmin, thrombin, trypsin, ...

Also Known As: A1A, A1AT, AAT, MGC23330...

Homologs of the SERPINA1 gene

The SERPINA1 gene is conserved in chimpanzee, dog, cow, mouse, rat, chicken, and zebrafish.

Recent activity

Turn On Clear

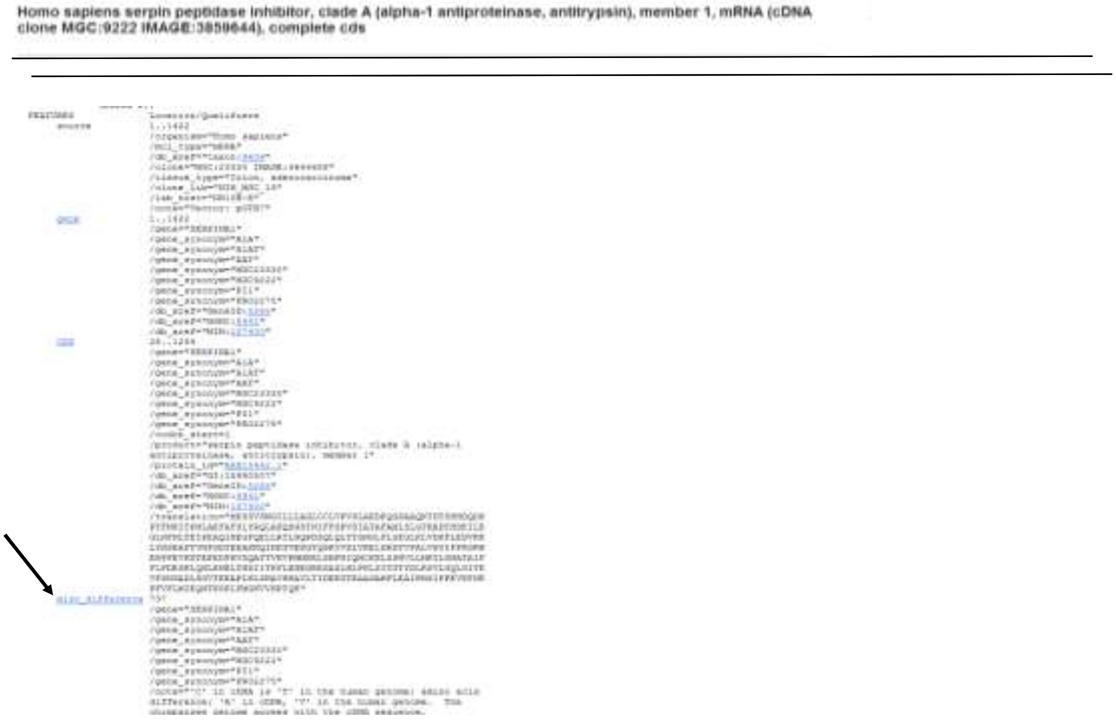
Activity recording is turned off

Turn recording back on

All links from this record

- Related sequences
- Full text in PMC
- GEO profiles
- Gene
- Gene genotype
- GeneView in dbSNP
- Map viewer
- OMIM
- Order cDNA clone
- Probe
- Protein
- PubMed
- PubMed (weighted)

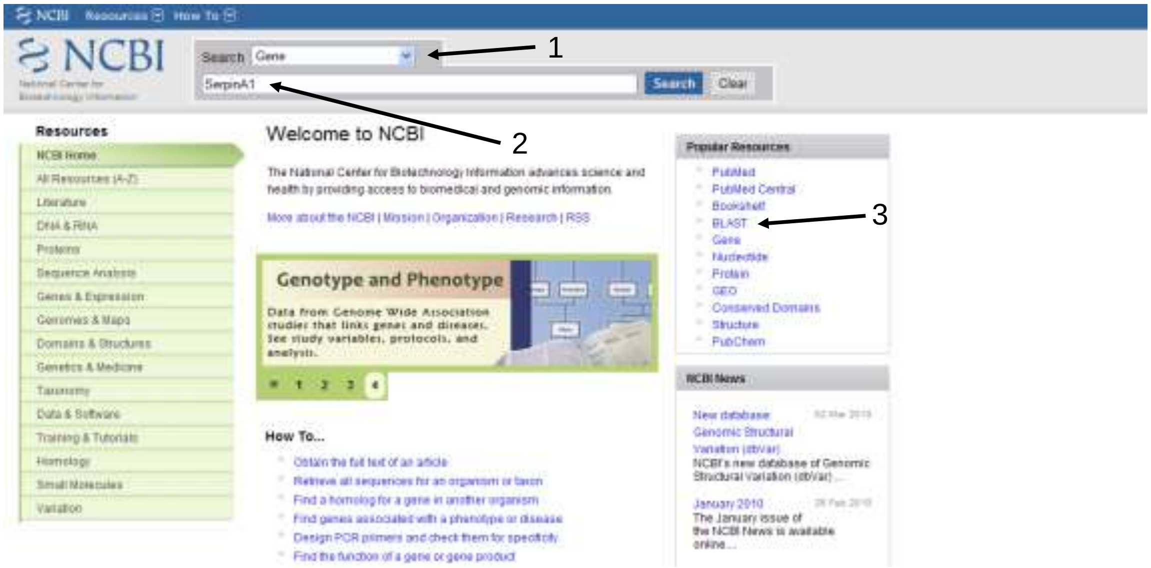
Figure 3b GenBank Record for BC011991 (continued)



Homo sapiens serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1, mRNA (cDNA clone MGC:9222 IMAGE:3859644), complete cds

```
FEATURES             Location/Qualifiers
     source          1..1422
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:[illegible] IMAGE:[illegible]"
                     /tissue_type="Colon, adenocarcinoma"
                     /clone_lib="NIH_MGC_10"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1..1422
                     /gene="SERPINA1"
                     /gene_synonym="A1A"
                     /gene_synonym="A1AT"
                     /gene_synonym="AAT"
                     /gene_synonym="MGC23330"
                     /gene_synonym="MGC9222"
                     /gene_synonym="PI1"
                     /gene_synonym="PRO2275"
                     /db_xref="GeneID:5265"
                     /db_xref="HGNC:8941"
                     /db_xref="MIM:107400"
     CDS             [illegible]..1268
                     /gene="SERPINA1"
                     /gene_synonym="A1A"
                     /gene_synonym="A1AT"
                     /gene_synonym="AAT"
                     /gene_synonym="MGC23330"
                     /gene_synonym="MGC9222"
                     /gene_synonym="PI1"
                     /gene_synonym="PRO2275"
                     /codon_start=1
                     /product="serpin peptidase inhibitor, clade A (alpha-1
                     antiproteinase, antitrypsin), member 1"
                     /protein_id="[illegible]"
                     /db_xref="GI:[illegible]"
                     /db_xref="GeneID:5265"
                     /db_xref="HGNC:8941"
                     /db_xref="MIM:107400"
                     /translation="[illegible]"
     misc_difference [illegible]
                     /gene="SERPINA1"
                     /gene_synonym="A1A"
                     /gene_synonym="A1AT"
                     /gene_synonym="AAT"
                     /gene_synonym="MGC23330"
                     /gene_synonym="MGC9222"
                     /gene_synonym="PI1"
                     /gene_synonym="PRO2275"
                     /note="'C' in cDNA is 'T' in the human genome; amino acid
                     difference: 'A' in cDNA, 'V' in the human genome.  The
                     [illegible] genome agreed with the cDNA sequence."
```

Figure 4 NCBI Homepage



The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

Data from Genome Wide Association studies that links genes and diseases. See study variables, protocols, and analysis.

- Obtain the full text of an article
- Retrieve all sequences for an organism or taxon
- Find a homolog for a gene in another organism
- Find genes associated with a phenotype or disease
- Design PCR primers and check them for specificity
- Find the function of a gene or gene product

New database Genomic Structural Variation (dbVar)
NCBI's new database of Genomic Structural Variation (dbVar) ...

January 2010
The January issue of the NCBI News is available online...

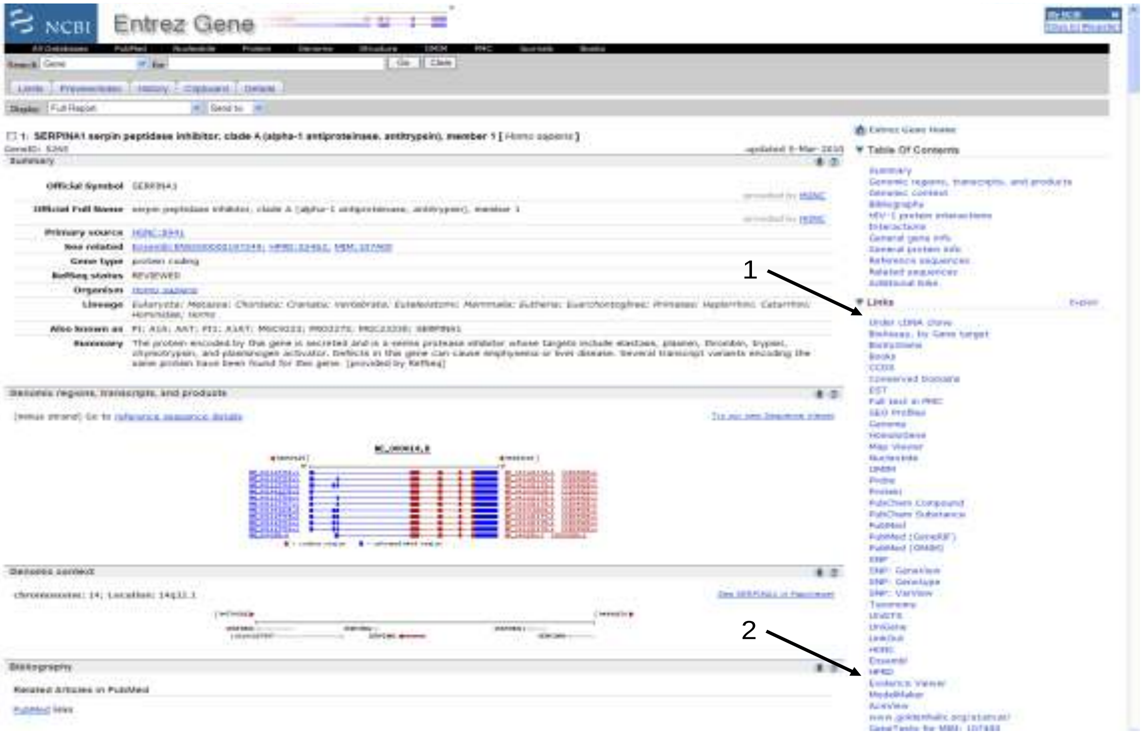
Approach 2. A second way to find an MGC clone, such as for the human gene SERPINA1, is to start at the NCBI homepage. First, choose **Gene** in the database **Search** dropdown menu (Figure 4, arrow 1), then type the gene symbol and the organism (serpina1 [SYMBOL] AND human [ORGN]) or enter the gene id for human SERPINA1 (5265) into the box (Figure 4, arrow 2) and click on **Search**. This will take you to the Entrez Gene page for human SERPINA1. On the right side of this page, under **LINKS** (Figure 5, arrow 1), **Order cDNA Clone** is displayed; this link leads to the same clone ordering information described above. In addition, the Gene page provides a wealth of other valuable information about the SerpinA1 gene.

Approach 3. You can create custom lists of MGC clones by performing searches from the NCBI Homepage using different combinations of search terms. For example, to search for all MGC full-cds clones, entering "MGC [KEYWORD] AND human [ORGANISM]" (Figure 6, arrow) will produce a list all ~30,000 human clones in the MGC. Adding other qualifying terms further restricts the search: "MGC [KEYWORD] AND human [ORGANISM] AND "peptidase inhibitor" yields a list of human MGC clones for peptidase inhibitors. Suggestions for searching the NCBI databases are given at Entrez Help.

Related resources:
"Tips for Finding cDNA Clones" is an NCBI page with details on locating MGC clones. The UCSC Genome Bioinformatics website offers a training page with many helpful tutorials and guides to using the UCSC genome browser, including a tutorial, "Fishing for Genes in the UCSC Browser," with advice on finding information on MGC clones.

Part B. Evaluating MGC Clone Sequence Integrity. The annotation of MGC clone sequences in GenBank records is based on comparing the cDNA clone sequence to the genome sequences. For example, the GenBank record for the SERPINA1 MGC clone BC015642, under **FEATURES,** notes in the **misc_difference** category (Figure 3b, arrow) that this clone contains a T>C change at position 737, with valine encoded by the genome and alanine by the cDNA; it also states that "the chimpanzee genome agrees with the cDNA sequence, suggesting that this difference is unlikely to be due to an artifact."

Figure 5 Gene Page for SerpinA1



1: SERPINA1 serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1 [Homo sapiens]

GeneID: 5265 updated 8-Mar-2010

Summary

Official Symbol SERPINA1 provided by HGNC

Official Full Name serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1 provided by HGNC

Primary source HGNC:8941

See related Ensembl:ENSG00000197249; HPRD:02462; MIM:107400

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 PI; A1A; AAT; PI1; A1AT; MGC9222; PRO2275; MGC23330; SERPINA1

Summary The protein encoded by this gene is secreted and is a serine protease inhibitor whose targets include elastase, plasmin, thrombin, trypsin, chymotrypsin, and plasminogen activator. Defects in this gene can cause emphysema or liver disease. Several transcript variants encoding the same protein have been found for this gene. [provided by RefSeq]

Genomic regions, transcripts, and products

(minus strand) Go to reference sequence details

NC_000014.8

Genomic context

chromosome: 14; Location: 14q32.1

Bibliography

Related Articles in PubMed

PubMed links

Entrez Gene Home

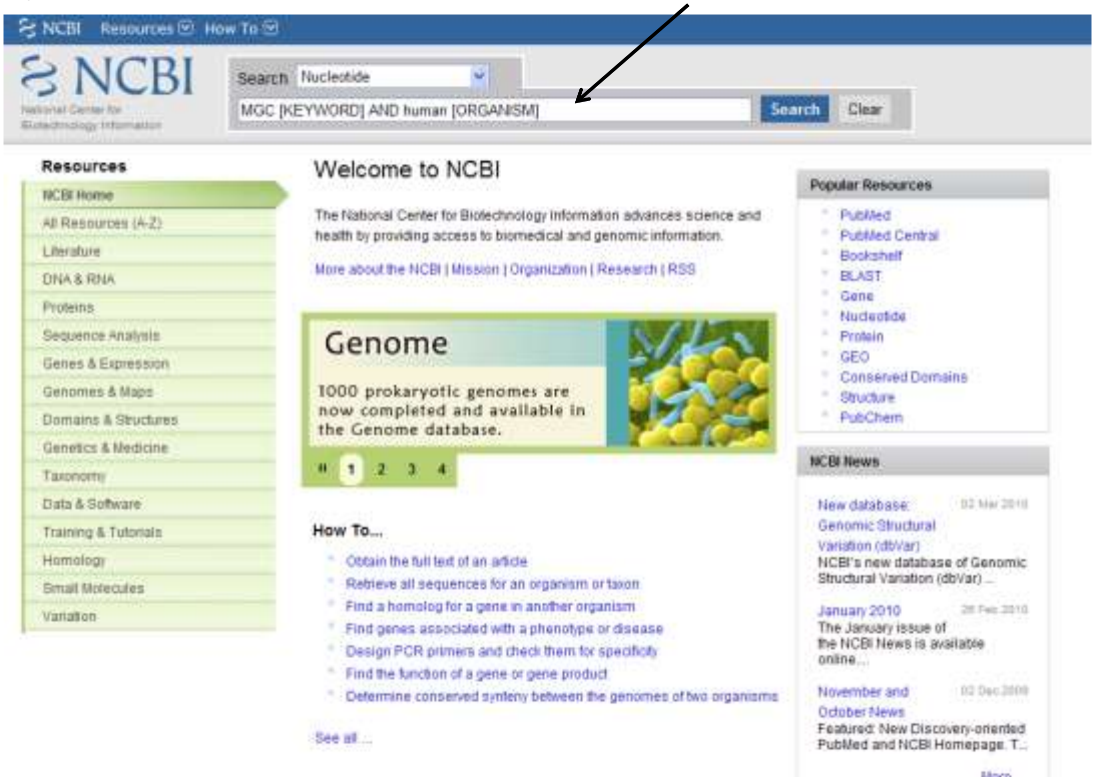
Table Of Contents

Summary
Genomic regions, transcripts, and products
Genomic context
Bibliography
HIV-1 protein interactions
Interactions
General gene info
General protein info
Reference sequences
Related sequences
Additional links

Links

Order cDNA clone
BioAssay, by Gene target
BioSystems
Books
CCDS
Conserved Domains
EST
Full text in PMC
GEO Profiles
Genome
HomoloGene
Map Viewer
Nucleotide
OMIM
Probe
Protein
PubChem Compound
PubChem Substance
PubMed
PubMed (GeneRIF)
PubMed (OMIM)
SNP
SNP: GeneView
SNP: Genotype
SNP: VarView
Taxonomy
UniSTS
UniGene
LinkOut
HGNC
Ensembl
HPRD
Evidence Viewer
ModelMaker
AceView
www.goldenhelix.org/xenobiotic/
GeneTests for MIM: 107400

Figure 6 A Custom Search for MGC Clones



MGC [KEYWORD] AND human [ORGANISM]

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

More about the NCBI | Mission | Organization | Research | RSS

Genome

1000 prokaryotic genomes are now completed and available in the Genome database.

How To...

- Obtain the full text of an article
- Retrieve all sequences for an organism or taxon
- Find a homolog for a gene in another organism
- Find genes associated with a phenotype or disease
- Design PCR primers and check them for specificity
- Find the function of a gene or gene product
- Determine conserved synteny between the genomes of two organisms

See all ...

Popular Resources

- PubMed
- PubMed Central
- Bookshelf
- BLAST
- Gene
- Nucleotide
- Protein
- GEO
- Conserved Domains
- Structure
- PubChem

NCBI News

New database: Genomic Structural Variation (dbVar) 02 Mar 2010
NCBI's new database of Genomic Structural Variation (dbVar) ...

January 2010 26 Feb 2010
The January issue of the NCBI News is available online...

November and October News 02 Dec 2009
Featured: New Discovery-oriented PubMed and NCBI Homepage. T...

More...

Annotations in the MGC clone GenBank records were frozen on March 23, 2009 at the conclusion of the MGC project. As a result, subsequent revisions of the human, mouse, and rat genomes that result in changes to annotations of the protein-coding sequence (CDS) of a gene will not be reflected in the GenBank records for these MGC clones. New annotations might, for example, reposition the ATG translation start codon so it now lies upstream (5′) to the start ATG noted in the GenBank record.

Prior to ordering an MGC clone, therefore, users should compare the MGC cDNA sequence against the most recent genome build, to determine whether more recent data have altered the cds annotation of the MGC clone. A variety of tools are available to perform such analysis, four of which are described here.

Tool 1. A detailed comparison of MGC clone sequences to the current reference genome and RefSeq transcripts is available from Evidence Viewer. The Entrez Gene page for SERPINA1, for example, provides links (Figure 5, arrow 2) to Evidence Viewer, which displays an extensive list of SERPINA1 RefSeq transcripts, MGC clones, and other clones, including alternative isoforms, together with details on nucleotide mismatches and indels associated with each.

Tool 2. A second way to display differences between an MGC clone sequence and the current version of the genome is to do a BLAST search of the MGC clone against the databases which harbor genome and RefSeq sequences. Although BLAST can be accessed from the NCBI homepage (Figure 4, arrow 3), a convenient BLAST link is also provided on the upper right side of GenBank record (Figure 3a, arrow 4). Clicking on the **BLAST** link for BC015642.2 takes the user to the BLAST page, where one can choose to compare this clone sequence against the human genome and transcript databases. Doing so reveals a single nt difference in the alignment between the MGC clone and its RefSeq transcript homolog, NM_001127707.1, at 737. Thus the annotated difference at position 737, noted on the MGC clone page is current.

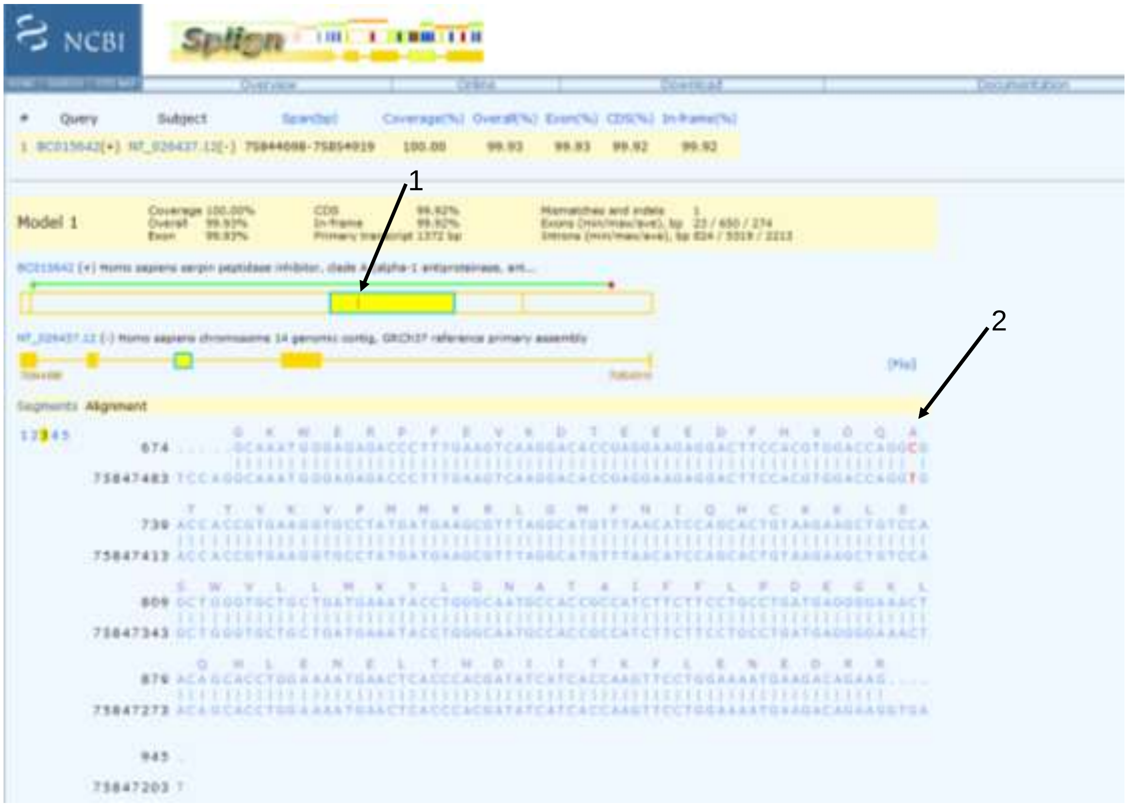
Tool 3. SPLIGN provides a third way to align a cDNA clone sequence to the current version of a reference genome, with exon-by-exon graphics and sequence displays. If you Enter BC015642 in the **cDNA** box and select ***homo sapiens*** in the pull-down menu under **Genome, SPLIGN** generates a display of the exon structure of this transcript, showing a single mismatch in exon 3 (Figure 7, arrow 1). Clicking on exon 3 reveals its sequence, with the T>C change at position 737 highlighted in red (Figure 6, arrow 2).

Tool 4. The UCSC Genome Browser provides detailed graphical views and alignment statistics of MGC clone sequences against their respective genomes. Entering the gene name SERPINA1 or the clone accession number BC015642 into the **position or search term** box of the UCSC Genome Browser Gateway page (Figure 8, arrow) and clicking **submit** leads to a schematic view of clones for SERPINA1 aligned against the human genome, including all relevant MGC clones (if you had first activated the **MGC Genes** track). A red line in exon 3 of the schematic line-figure of BC015642 signals a sequence difference from the reference genome.

Clicking on the line-figure of BC015642 connects you to its clone GenBank record, with summary statistics on alignments of the clone sequence to the Genome Reference Consortium (GRC) human genomic sequence and to multiple RefSeq mRNAs. This clone page shows that the exon structure of BC015642 shares 100% identity with multiple RefSeq RNAs and that its sequence is 99.93% identical to the reference human genome sequence on chromosome 14. Clicking on "BC015642:1-1371," under "mRNA (alignment details)," displays a single CDS nt difference at position 737.

The web pages noted above provide many other kinds of information and useful links, besides the ones described. For example, at the upper right of an MGC clone GenBank record, such as for BC015642 , under **Customize View** (Figure 3a, arrow 5) a user can activate the **Features added by NCBI** (followed by **Update View**) to display the 37 SNPs listed in dbSNP that overlap with this cDNA sequence.

Figure 7 Splign Alignment of BC015642 vs. Reference Human Genome



#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	BC015642(+)	NT_026437.12(-)	75844098-75854019	100.00	99.93	99.93	99.92	99.92

Model 1

Coverage	100.00%	CDS	99.92%	Mismatches and indels	1
Overall	99.93%	In-frame	99.92%	Exons (min/max/ave), bp	23 / 650 / 274
Exon	99.93%	Primary transcript	1372 bp	Introns (min/max/ave), bp	824 / 5519 / 2213

BC015642 (+) Homo sapiens serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, ant...

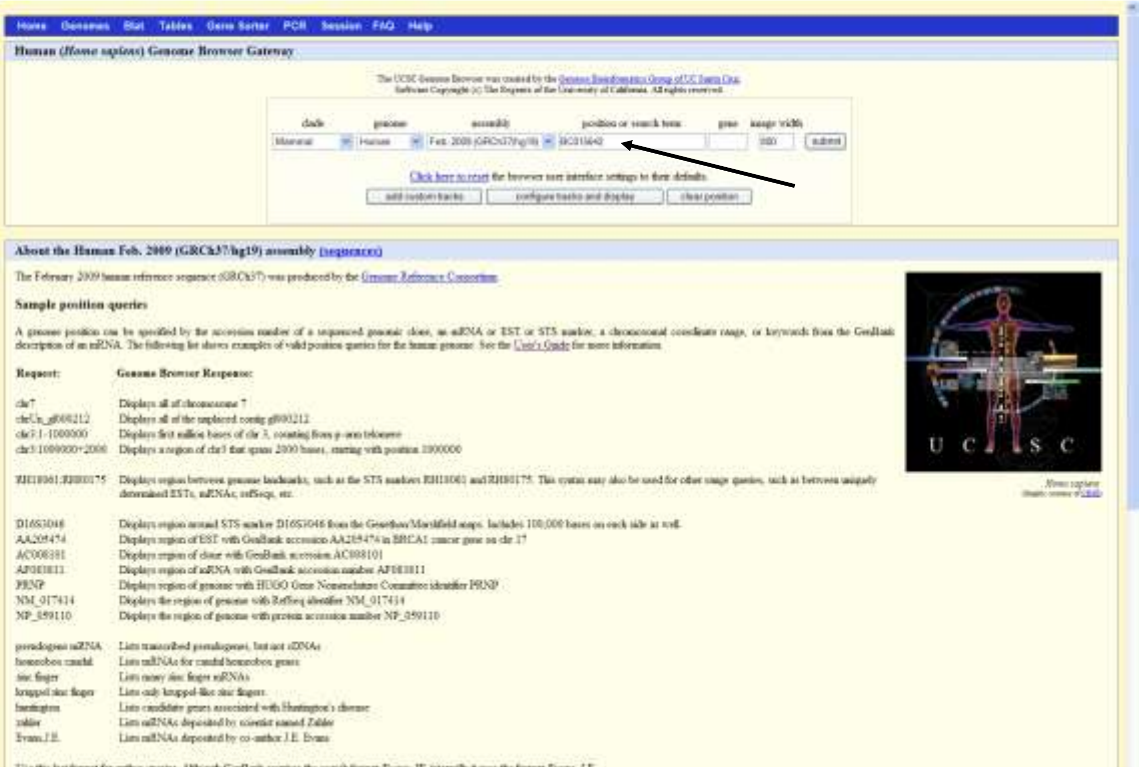
NT_026437.12 (-) Homo sapiens chromosome 14 genomic contig, GRCh37 reference primary assembly

75844098 75854019

Segments Alignment

```
12345                G  K  W  E  R  P  F  E  V  K  D  T  E  E  E  D  F  H  V  D  Q  A
             674 .....GCAAATGGGAGAGACCCTTTGAAGTCAAGGACACCGAGGAAGAGGACTTCCACGTGGACCAGGCG
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
        75847483 TCCAGGCAAATGGGAGAGACCCTTTGAAGTCAAGGACACCGAGGAAGAGGACTTCCACGTGGACCAGGTG

                  T  T  V  K  V  P  M  M  K  R  L  G  M  F  N  I  Q  H  C  K  K  L  S
             739 ACCACCGTGAAGGTGCCTATGATGAAGCGTTTAGGCATGTTTAACATCCAGCACTGTAAGAAGCTGTCCA
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        75847413 ACCACCGTGAAGGTGCCTATGATGAAGCGTTTAGGCATGTTTAACATCCAGCACTGTAAGAAGCTGTCCA

                 S  W  V  L  L  M  K  Y  L  G  N  A  T  A  I  F  F  L  P  D  E  G  K  L
             809 GCTGGGTGCTGCTGATGAAATACCTGGGCAATGCCACCGCCATCTTCTTCCTGCCTGATGAGGGGAAACT
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        75847343 GCTGGGTGCTGCTGATGAAATACCTGGGCAATGCCACCGCCATCTTCTTCCTGCCTGATGAGGGGAAACT

                   Q  H  L  E  N  E  L  T  H  D  I  I  T  K  F  L  E  N  E  D  R  R
             876 ACAGCACCTGGAAAATGAACTCACCCACGATATCATCACCAAGTTCCTGGAAAATGAAGACAGAAG.....
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        75847273 ACAGCACCTGGAAAATGAACTCACCCACGATATCATCACCAAGTTCCTGGAAAATGAAGACAGAAGGTGA

             945 .

        75847203 T
```

Figure 8 USCS Genome Browser Gateway Page


Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help
add custom tracks
configure tracks and display
clear position
submit
U C S C
Homo sapiens
Graphic courtesy of CBSE

Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade	genome	assembly	position or search term	gene	image width
Mammal	Human	Feb. 2009 (GRCh37/hg19)	BC015642		800

Click here to reset the browser user interface settings to their defaults.

About the Human Feb. 2009 (GRCh37/hg19) assembly (sequences)

The February 2009 human reference sequence (GRCh37) was produced by the Genome Reference Consortium.

Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the User's Guide for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gl000212	Displays all of the unplaced contig gl000212
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061;RH80175	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175. This syntax may also be used for other range queries, such as between uniquely determined ESTs, mRNAs, refSeqs, etc.
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
AA205474	Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101	Displays region of clone with GenBank accession AC008101
AF083811	Displays region of mRNA with GenBank accession number AF083811
PRNP	Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414	Displays the region of genome with RefSeq identifier NM_017414
NP_059110	Displays the region of genome with protein accession number NP_059110
pseudogene mRNA	Lists transcribed pseudogenes, but not cDNAs
homeobox caudal	Lists mRNAs for caudal homeobox genes
zinc finger	Lists many zinc finger mRNAs
kruppel zinc finger	Lists only kruppel-like zinc fingers
huntington	Lists candidate genes associated with Huntington's disease
zahler	Lists mRNAs deposited by scientist named Zahler
Evans,J.E.	Lists mRNAs deposited by co-author J.E. Evans

Part C. Information on MGC Clones Posted at the NCBI-MGC ftp site:
ftp://ftp.ncbi.nih.gov/repository/MGC/MGC_project/supplementary_tables/
(A Glossary of Abbreviations is provided below)

Supplementary Table A: Lists targets and results of MGC PCR-Rescue (RT-PCR cloning) and DNA synthesis: Rows: Assigned transcripts. **Columns:** Organism, rescue_uid (PCR rescue ID), laboratory, geneid, target nuc_acc.ver, target nuc_gi, target prot_gi, target cds_length, mgc_acc , mgc_acc_other_gene, non_mgc_acc .

Supplementary Table B: Lists clones produced by the MGC Project which are Not Full-CDS: Rows: information on each transcript accession. **Columns:** organism, taxid, MGCid, imageid, nuc_acc, nuc_acc.ver, nuc_gi, nuc_len, prot_acc, prot_acc.ver, prot_gi, prot_length, geneid, gene symbol, gene name, gene type.

Supplementary Table C: Total MGC Full-CDS Collection: Rows: information on each transcript accession. **Columns:** organism, taxid (species code), MGCid, imageid, nuc_acc, nuc_acc.ver, nuc_gi, nuc_len, prot_acc, prot_acc.ver, prot_gi, prot_len, geneid, gene symbol, gene name, gene type.

Supplementary Table D: List of protein-coding genes missing from the MGC full-cds collection: Rows: Genes that have RefSeq transcript with prefix NM_ or XM_, ordered by organism, and geneid. **Columns:** organism, geneid, gene_symbol, gene_name, longest_CDS_length_of_NM_RefSeq, NM_RefSeq_with_longest_CDS, list of RefSeq_accession.ver

Supplementary Table E: List of OMIM Disease Genes Represented in MGC: Rows: information on each transcript accession. **Columns:** OMIM_id; OMIM_DESCR; human_GeneID; human_MGC_accession.ver; HomoloGene_id; mouse_GeneID; mouse_MGC_accesion.ver; rat_GeneID;rat_MGC_accession.ver

Glossary of Abbreviations in Supplementary Tables (above):
rescue_uid (PCR rescue ID)
geneid (gene ID)
target nuc_acc.ver (RefSeq transcript accession, with version number after decimal point)
nuc_gi (gi number that corresponds to the RefSeq transcript)
nuc_len (length of entire cloned cDNA, including linker sequences and, in some cases, 5' and/or 3' UTR sequences
target prot_gi (protein gi number corresponding to RefSeq transcript)
mgc_acc (accession of full-cds clone for intended RefSeq transcript)
mgc_acc_other_gene (accession of full-cds clone from gene other than the target gene
non_mgc_acc (accession of a partial-cds or no-cds clone)

Individual MGC, XGC, & ZGC Project Data
ftp://ftp.ncbi.nih.gov/repository/MGC/MGC_project/MGC_project_data/
All MGC, Xenopus Gene Collection, and Zebrafish Gene Collection clones and their related sequences are included and grouped by organism, with two files for each: "MGC full-CDS" clones "MGC Other" clones. "MGC Other" clones include partial CDS clones; short-CDS clones (CDS < 50% of the longest RefSeq transcript CDS); and clones for transcripts lacking annotated CDS, including possible pseudogenes. Files include information about the clone and sequence in tab-delimited format; fasta nucleotide sequences; GenBank flatfile for nucleotide record; protein fasta sequences; and GenBank flatfile for protein record.

Part D. Gene Expression Data Related to MGC Clones. Gene expression profiles for MGC clones can be accessed directly through the GenBank clone page links to UniGene, located on the lower right side of the page (under **All links from this record**). For BC015642, this link leads to a UniGene page that displays both GEO and EST profiles of SERPINA1 gene expression.

mRNA, EST, SAGE, and microarray expression profiles for MGC genes are also displayed on the UCSC Genome Browser, when these tracks are activated.

AceView provides gene expression information, including tissues of origin for cDNA clones and average expression levels of transcripts.

GeneNote provides gene-specific microarray expression profiles. See, for example, the results for SERPINA1.

Other sources of gene expression profiles include: CGAP Tools; MGC-GXD; GENSAT; BioGPS; and DAVID.

A rapid way to survey all NCBI databases for information about a gene or a specific cDNA clone (once its accession is identified) is to search "All Databases" for a gene name or clone accession, from the NCBI home page. This will reveal on one page whether, for example, expression data are available in GEO, markers exist in UniSTS, or TaqMan probes are listed in Probe.