

NCBI News, March 2009

Peter Cooper, Ph.D.¹ and Dawn Lipshultz, M.S.²

Created: February 12, 2009.

Featured Resource: The New Entrez Sequence View has an Emphasis on Discovery

NCBI now has an updated sequence view for nucleotide and protein records that provides better display options, links to analysis tools, and an emphasis on the discovery of related information in other NCBI databases (Figure 1). The new view retains the standard format of the sequence but has more obvious links to display FASTA, other formats, and specific subregions and features. The right-hand discovery column features direct links to analysis tools, and, most importantly, explicit links to valuable related information such as better-annotated sequences, and more information-rich databases such as Gene, HomoloGene, and PubMed.

New display links, regions and feature options

The format row at the top of the new view has direct links to display the sequence in FASTA format and to display the record and features in the NCBI Graphical Viewer (NCBI News, Aug 2008). Other less specialized reports such as ASN.1 and XML are available through the separate “More formats” list. A “Download” link to save the records in various formats to a local file and a “Save” link to store the records in the NCBI Clipboard or through MyNCBI in “My Collections” for later use are located at the right-hand end of this row along with the traditional “Links” menu that provides links to related information in other databases. Several of the more useful items from the “Links” menu are listed explicitly as Discovery column items described in the “Discovery Links” section below. The expandable display controls, “Change Region Shown” and “Customize View”, are located below the Links menu on the right-hand-side. The “Change Region Shown” control provides a convenient mechanism for displaying specific regions of the sequence. The “Customize View” control specifies the number and kinds of annotated features and the DNA strand that is shown.

Direct entry to analysis tools

The NCBI Discovery column begins just below the display controls with direct access to sequence analysis tools. The NCBI primer designing tool, Primer-BLAST (NCBI News,

¹ NCBI; Email: cooper@ncbi.nlm.nih.gov. ² NCBI; Email: lipshultz@ncbi.nlm.nih.gov.

Format: [GenBank](#) [FASTA](#) [Graphics](#) [More Formats](#) [Download](#) [Save](#) [Links](#)

GenBank: AF013988.1

Homo sapiens serine protease mRNA, complete cds

[Features](#) [Sequence](#)

LOCUS AF013988 1451 bp mRNA linear PRI 20-MAY-2008
 DEFINITION Homo sapiens serine protease mRNA, complete cds.
 ACCESSION AF013988
 VERSION AF013988.1 GI:2318114
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM [Homo sapiens](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
 Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1451)
 AUTHORS Little,S.P., Dixon,E.P., Norris,F., Buckley,W., Becker,G.W.,
 Johnson,M., Dobbins,J.R., Wyrick,T., Miller,J.R., MacKellar,W.,
 Hepburn,D., Corvalan,J., McClure,D., Liu,X., Stephenson,D.,
 Clemens,J. and Johnstone,E.M.
 TITLE Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated
 from Alzheimer's disease brain
 J. Biol. Chem. 272 (40), 25135-25142 (1997)
 PUBMED [9312124](#)
 REFERENCE 2 (bases 1 to 1451)
 AUTHORS Little,S.P., Johnstone,E.M. and Norris,F.
 TITLE Direct Submission
 JOURNAL Submitted (15-JUL-1997) CNS Division, Eli Lilly and Company, Lilly
 Corporate Center, Indianapolis, IN 46285, USA
 FEATURES
 source Location/Qualifiers
 1..1451
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /strain="human"
 /db_xref="taxon:[9606](#)"
 /chromosome="19"
 /map="19q13.3"
 /tissue_type="Alzheimer's disease brain tissue"
 147..881
 /note="Zyme; protease bears homology to Kallikrein class
 and can be localized to microvessels and microglia;
 chymotrypsin-like"
 /codon_start=1
 /product="serine protease"
 /protein_id="[AAB66483.1](#)"
 /db_xref="GI:2318115"
 /translation="MKKLMVLSLIAAAWAEQNKLVHGGPCDKTSHFYQAAALYTSGH
 LLCGGVLIHPLWVLTAAHCKKPNLQVFLGKHNLRQRESSQEQQSSVVRVAVIHPDYDAAS
 HDQDMLRLRLARPAKLSLQPLERDCSANTTSCILIGWGKTADGDFPDTIQCAVI
 HLVSRECEHAYPGQITQNMCLCAGDEKYGKDCSQGDSGGPLVCGDHLRGLVSWGNIPI
 GSKEKPGVYTNVCRYTNWIKTIQAK"

[CDS](#)

ORIGIN
 1 gtcgaccac gcgtccgct ggctggctcg ctctctctct gggacacaga ggtggcagg
 61 cagcacacag agggacctac gggcagctgt tccttcccc gactcaagaa tccccggagg
 121 cccggaggcc tgcagcagga gggccatga agaagctgat ggtggtgctg agtctgattg
 181 ctgcagcctg ggcagaggag cagaataagt tggcgcattg cggacctgc gacaagacat
 241 ctaccacctc ccaagctgcc ctctacacct cgggccactt gctctgtgtg ggggtcttca
 301 tccatccact gtgggtctct acagctgccc actgcaaaaa accgaatctt caggctcttc
 361 tggggaagca taacctctcg caaagggaga gtccccagga gcagagttct gttgtccggg
 421 ctgtgatcca cctgactat gatgcccga gccatgacca ggacatcatg ctgttgccgc
 481 tggcaagccc agcaaacctc tctgaactca tccagccctc tcccctggag agggactgct
 541 cagcccaaac caccagctgc cacatctctg gctggggcaa gacagcagat ggtgatttcc
 601 ctgacaccat ccagtgtgca tacatccacc tgggtgtccc tgaggagtg gaggatgctt
 661 accctggcca gatcaccag aacatgttgt gtgctgggga tgagaagtac gggaggatt
 721 cctgcccagg tgattctggg ggtccgctgg tatgtggaga ccacctccga ggcttctgtg
 781 catggggtaa cctccctgt ggatcaaaag agaagccag agtctacacc aacgtctgca
 841 gatacacgaa ctggatccaa aaaaccattc aggcacaagt accctgacat gtgacatcta
 901 cctcccagcc taccaccoca ctggctggtt ccagaaactc tctcacctag accttgctcc
 961 cctctctctc ctgcccagct ctgacctgca tgccttaata acgcagcgac gtgagggtcc
 1021 tgattctccc tggttttacc ccagctccat ccttgcatca ctggggagga cgtgatgatg
 1081 gaggacttgg gtctctgctc ttacccccac cactaagaga ataccggaaa atcccttcta
 1141 ggcatctctc ctcccacacc cttccacacg tttgatttct tctgcagag gccagccacc
 1201 gtgtctgtaa tccagctcc gctgcttact gtctggttcc ccttgggatg tacctttctt
 1261 cactgcagat ttctcacctg taagatgaag ataaggatga tacagtctcc ataaggcagt
 1321 ggctgttgga aagatttaag gtttcacacc tatgacatac atggaaatgc acctgggcca
 1381 ccatgcactc aataaagaat gaattttatt atgaaaaaaa aaaaaaaaaa aaaaaaaaaa
 1441 agggcggccg c

//

[Change Region Shown](#)

[Customize View](#)

[Pick Primers](#)
 Design and test primers for this sequence using Primer-BLAST.

Articles about KLK6

- ▶ Prognostic value of kallikrein-related peptidase 6 protein expr [Cancer Sci. 2008]
- ▶ Co-expression of KLK6 and KLK10 as prognostic factors for su [Br J Cancer. 2008]
- ▶ Kallikrein 6 is a mediator of K-RAS-dependent migra [Biol Chem. 2008] » See all...

Reference sequences

- ▶ mRNA
- ▶ Protein

More about the KLK6 gene
 Kallikreins are a subgroup of serine proteases having diverse physiological functions. Growing evidence suggests that many kallikreins are i...
 Also Known As: Bssp, Kik7, MGC9355, NE...

Homologs of KLK6
 The KLK6 gene is conserved in dog, cow, mouse, and rat.

Order cDNA Clone
 The NIH MGC Collection contains a sequence-verified cDNA clone for KLK6.

Recent Activity

All links from this record

- ▶ Gene
- ▶ Gene Genotype
- ▶ GeneView in dbSNP
- ▶ Probe
- ▶ Protein
- ▶ PubMed
- ▶ PubMed (Weighted)
- ▶ Taxonomy
- ▶ Related Sequences
- ▶ Map Viewer
- ▶ OMIM
- ▶ GEO Profiles
- ▶ SNP

Figure 1. The new Entrez sequence view of a GenBank record for a human mRNA (AF013988) submitted in 1997. Links to alternate display formats are at the top of the record. Expandable display controls are at the top of the right-hand Discovery Column. One-click submission to design primers is available through the “Pick Primers” link. The items evident in the Discovery Column provide an instant update for the molecular biology, nomenclature, and literature relevant to this gene and its products.

Nov 2008), is linked here now. Additional tools will be added in later versions of the viewer including a direct link to run a BLAST database search. In all cases these analysis

tools are set in advance to provide the most relevant and up-to-date results appropriate to the context. For example, following the “Pick Primers” link from a human mRNA sequence in the viewer loads the Primer-BLAST form already set up to perform a specificity check against the appropriate background database, the human genome transcripts in this case. Moreover if the search is run with an older GenBank sequence as the query, the equivalent NCBI Reference Sequence (RefSeq) will be substituted to improve specificity checking. Direct access to analysis tools will streamline many visits to the NCBI Website by providing one-click, live access to the computing power of the NCBI.

Discovery Links

As shown in the figure, the links in the Discovery column expose several highly relevant and useful sets of related information. These include selected relevant articles in PubMed (“Articles about KLK6”), links to corresponding mRNA and protein Reference Sequences, a link to the corresponding Gene record (“More about the KLK6 gene”), and links to homologs in other species provided by HomoloGene (“Homologs of KLK6”). For the record shown in the figure, these Discovery column items automatically update the biology and nomenclature for this older GenBank record, identifying it at a glance as a transcript of the human kallikrein-related peptidase gene, KLK6, and providing in a single click the Reference Sequences for the three of known splice variants of this gene. Each of these has additional biological annotations directly on the record plus its own discovery column. A link to “Order cDNA Clone” provides access to molecular reagents for this transcript. The “Articles about ...” link is an enhanced set of human-reviewed references about KLK6 combining the citations from the Online Mendelian Inheritance in Man (OMIM) article for this gene with the linked articles from the NCBI Gene record. These constitute an essential set of literature about the biological roles of the KLK6 gene and its products. Finally, the “More about the KLK6 gene” link provides direct access to the Gene record, a gateway to the human genome and all molecular biology information about the KLK6 gene.

Summary

The new Entrez sequence provides intuitive display controls, direct access to live analysis, and to the rich pre-compiled information available through Entrez Gene, OMIM and Homologene. The current version and future improvements move the sequence databases towards a condition where even older sequences become self-annotating and are automatically updated through the analysis performed at the NCBI evident in the Discovery column. These enhancements should make the NCBI Entrez system a more efficient experience for visitors and easier to use as a Discovery system.

New Databases and Tools

Genome Build

Build 1.1 of *Hydra magnapapillata* is available in the Genomes database and on the NCBI Map Viewer. The Map Viewer page is: www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=6085

Microbial Genomes

Twenty-eight finished microbial genomes were released between January 14 and February 6. The original sequence data files submitted to GenBank/EMBL/DDBJ are available on the FTP site: <ftp://ncbi.nih.gov/genbank/genomes/Bacteria/>. The RefSeq provisional versions of these genomes are also available: <ftp://ncbi.nih.gov/genomes/Bacteria/>.

GenBank News

GenBank release 170.0 is available via web and FTP. The current release includes information available as of February 13, 2009. With this release, the new DBLINK linetype is now legal for GenBank sequence records, and it will begin to appear in GenBank Update files, soon after GenBank 170.0 is made available. Release notes are on the on the ftp site: <ftp://ncbi.nih.gov/genbank/gbrel.txt>

Updates and Enhancements

Clone Finder

The Clone Finder tool, designed to identify genomic clones on a genome assembly, has been updated with an improved search interface that makes it easier to search by feature. Clone Finder also offers the ability to filter clones on results pages using feature information, and more informative pop-up menus on clone features. The performance of graphical view is improved and now integrates the library table view. The tool will now allow users to download all clones, or only clones from a given library in Excel format.

Documentation on Clone Finder can be found on the following Web page:
www.ncbi.nlm.nih.gov/projects/mapview/static/clonfinder_documentation.shtml

UniVec

UniVec database build 5.1 is now available. UniVec is a non-redundant database of vector sequences used in conjunction with the VecScreen tool to identify vector sequence contamination in nucleotide sequences. The number of sequences in UniVec has increased by 2% for build 5.1. The vector BLAST database has also been updated to contain full-length versions of all sequences from GenBank that were used in the current UniVec build.

PubMed

The PubMed Summary page now displays information about free articles from publishers. The new information is in addition to the PubMed Central links that appear for full-text PMC articles. For more information, see the *NLM Technical Bulletin* article: www.nlm.nih.gov/pubs/techbull/jf09/jf09_pm_free_article.html . The current issue of the *Technical Bulletin* also has an informative article about shared settings in the My NCBI tool.

Exhibits

NCBI will have an exhibit booth at the Experimental Biology Annual Meeting on April 18-22 in New Orleans, Louisiana.

Announce Lists and RSS Feeds

Fifteen topic-specific mailing lists are described on the Announcement List summary page. Announce lists provide email announcements about changes and updates to NCBI resources. www.ncbi.nlm.nih.gov/Sitemap/Summary/email_lists.html

Seven RSS feeds are now available from NCBI including news on PubMed, PubMed Central, NCBI Bookshelf, LinkOut, HomoloGene, UniGene, and NCBI Announce. Please see: www.ncbi.nlm.nih.gov/feed/

Comments and questions about NCBI resources may be sent to NCBI at: info@ncbi.nlm.nih.gov, or by calling 301-496-2475 between the hours of 8:30 a.m. and 5:30 p.m. EST, Monday through Friday.