U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Cooper P, Morris R. NCBI News, October 2010. 2010 Nov 10 [Updated 2010 Dec 29]. In: NCBI News [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 1991-2012.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/

NCBI News, October 2010

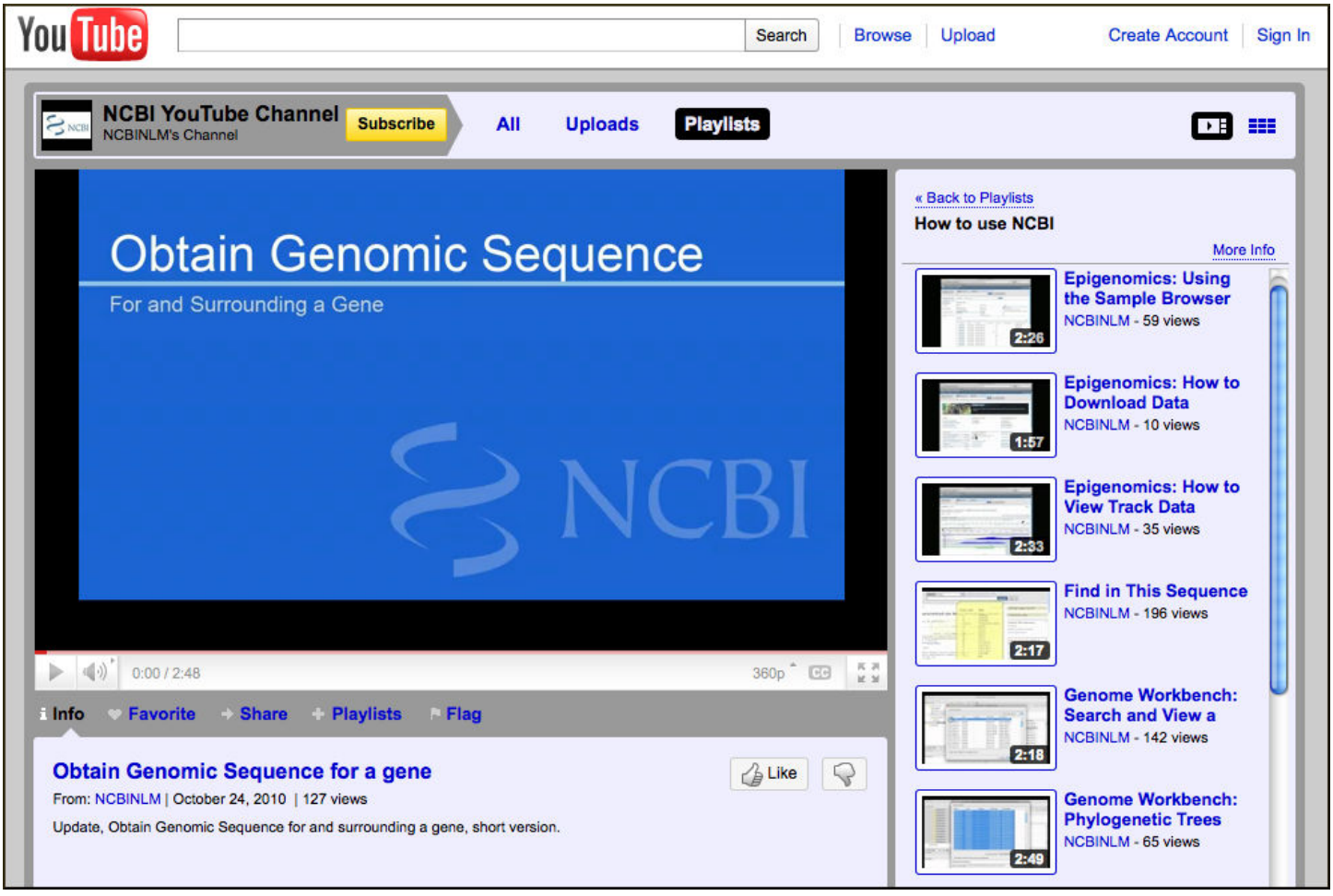
Peter Cooper, Ph.D.[1] and Rana Morris, Ph.D.[2]

Created: November 10, 2010; Updated: December 29, 2010.

New Databases and Tools

Instructional Videos on the NCBI YouTube Channel

The NCBI YouTube channel now has ten instructional videos that demonstrate how to use NCBI tools and databases. All ten are available through the Video Tutorials playlist. Topics include several How-to tasks from the NBCI Homepage, using the new Find-in-Sequence feature in the sequence databases, browsing the new Epigenomics resource, and using Genome Workbench.



Author Affiliations: 1 NCBI; Email: cooper@ncbi.nlm.nh.gov. 2 NCBI; Email: morrisrc@ncbi.nlm.nih.gov.

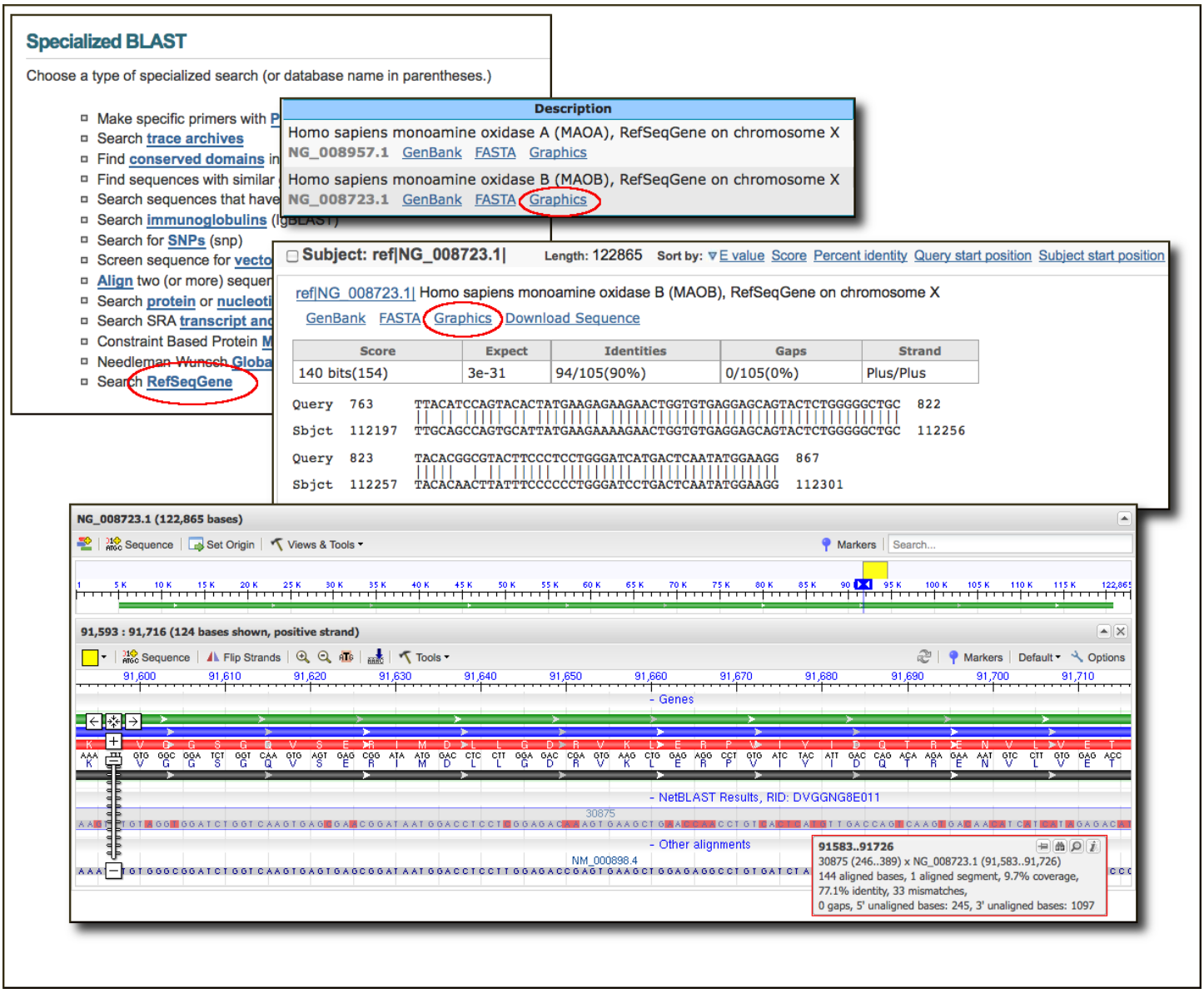
The RefSeqGene Project: a Stable Resource for Gene Annotations

NCBI produces RefSeqGene as a subset of NCBI's Reference Sequence (RefSeq) project. These genomic sequences are intended to serve as reference standards for well-characterized genes and offer stable genomic platforms with a permanent identifier and a core content that does not change. This provides a standard system for numbering exons and introns and defining the coordinates of variations and other features. Each RefSeqGene record offers gene-specific sequences for each gene as well as upstream and downstream flanking regions. All records are experimentally well-supported, come from a single genomic clone, and represent the allele present in the Genome Reference Consortium reference genome whenever possible. The RefSeqGene project is an active member of the Locus Reference Genomic (LRG) collaboration that accepts submission of variants, nominations for target genes, and provides input on curation.

Entrez nucleotide system incorporates RefSeqGene records where they may be selected by adding the term `refseqgene[keyword]` to any query. RefSeqGene records are also linked from their corresponding RefSeq transcript, genomic clone, protein sequence and gene records. The RefSeqGene homepage has documentation, links to related resources and tools – including the RefSeqGene BLAST service , described below – and a list of available RefSeq gene records by gene symbol. The RefSeqGene data are available for download from the RefSeq area of the FTP site.

RefSeqGene BLAST Service Now Available, Preview of Enhancements to BLAST

The NCBI BLAST web service now includes a specialized BLAST service that searches the NCBI RefSeqGene records. RefSeqGene BLAST also offers a preview of changes in format coming to the main BLAST services. These changes include a two-line format in the Descriptions section allowing more of the sequence title to be displayed. The second line has separate links to display formats in the sequence databases (GenBank, FASTA). Additional links to related data such as corresponding Gene records will be added to this second line in the future. Also on the second line is a link to display BLAST output in the graphical sequence viewer. This latter option allows the BLAST hits to be displayed in the context of the biological features annotated on the database record and provides a powerful new way to look at BLAST results.



Microbial Genomes

Twenty-one finished microbial genomes were released during September 2010. 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/. In addition 25 microbial whole genome shotgun sequencing projects were added. Orginal submitted files are available in ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria_DRAFT/ and RefSeq provisional versions are in ftp://ftp.ncbi.nih.gov/genomes/Bacteria_DRAFT/. All GenBank and RefSeq microbial genomes are incorporated in the NCBI integrated Entrez search and retrieval system.

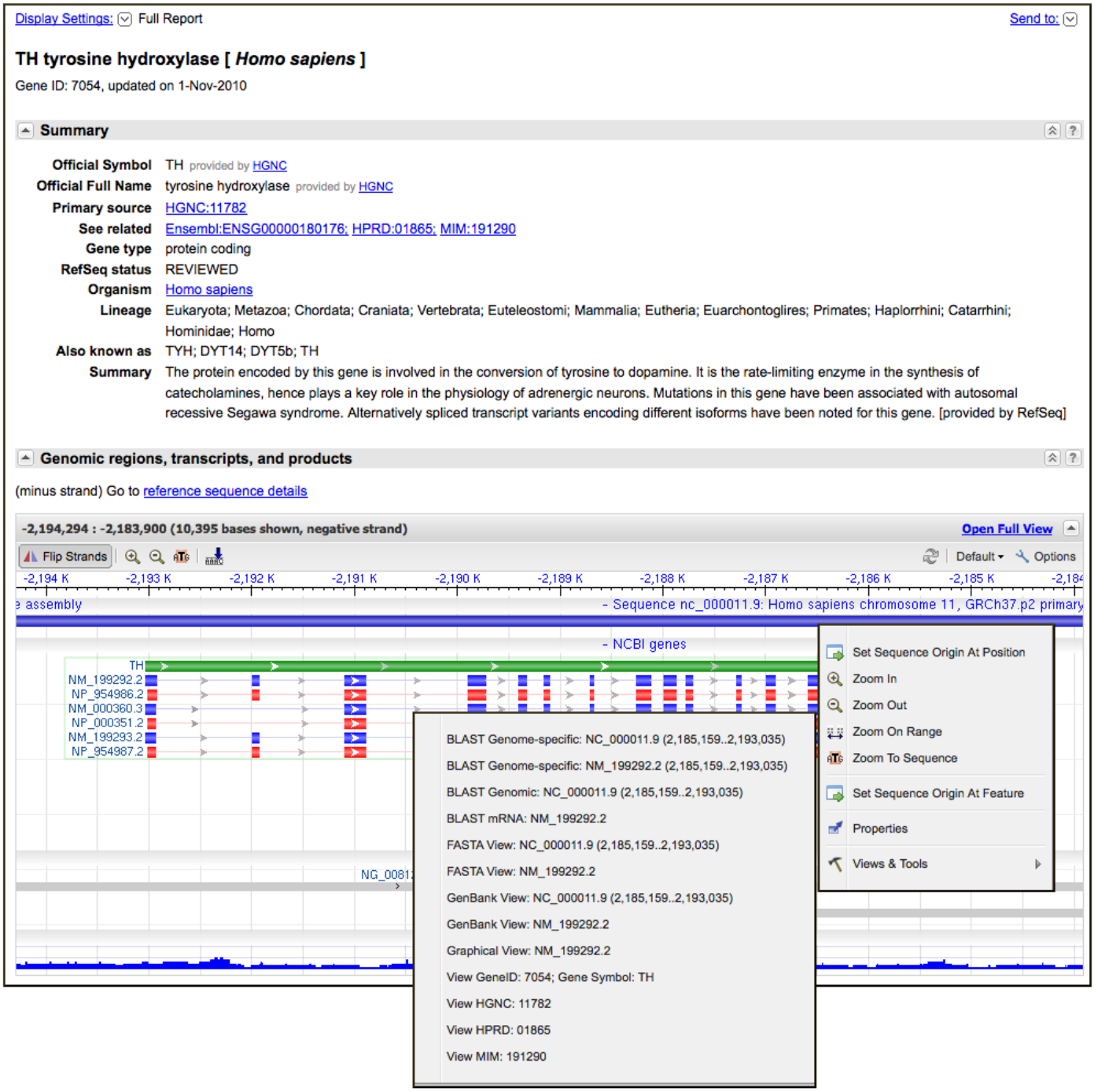
GenBank News

GenBank release 180.0 is available through the NCBI web and FTP sites. The current release includes information available as of October 15, 2010. Release notes describe the current state of data and upcoming changes.

Updates and Enhancements

Updated Gene Pages

The Entrez Gene database has moved to the updated interface introduced in PubMed last year. The new format has better controls and options for displaying and downloading records. More importantly, Gene record displays have several improvements including easier navigation, customizable display options, and better integration with closely related data in the NCBI system. Integration with the NCBI Reference Sequence transcripts, proteins, and genomic assemblies, as well as sequence variations from the dbSNP has been greatly improved through the graphical sequence viewer embedded in the gene record, as shown in the accompanying image.



Graphic Display in Primer-BLAST

Primer BLAST results now offer an alternative to the standard graphic that allows primer alignments to be displayed in the Graphical Sequence viewer. This new option is activated by the *Use new graphic view* option on the web form. Results in the graphical sequence viewer display the primer binding sites in the context of the biological features of the sequence such as the locations of exons, introns, coding sequences, and untranslated regions making it easier to assess the usefulness of particular primer pairs.

Genome Workbench 2.2.0

A new release of the NCBI Genome Workbench – the downloadable graphical sequence analysis, annotation, and display platform – is now available. This latest version, 2.2.20, includes several bug fixes described in the release notes. More information about Genome Workbench along with a link to instructions for downloading and installing the program are provided on the homepage.

Reference Human Genome, GRCh37, Updated to Patch 2

The reference human genome has been updated to patch 2 (GRCh37.p2). The new build is available in Entrez, the Map Viewer, BLAST, and on the FTP site. This update includes 70 patches. Patches are updated regional assemblies that either provide additional alternate assemblies for alleles that are not adequately represented in the current genome (Novel patches) or correct assembly errors in the current build (Fix patches). GRCh37.p2 comprises 52 novel patches and 18 fix patches. Fix patches be incorporated the next major genome build changing the tiling path while novel patches will be incorporated as alternate loci.

Giant Panda and Marmoset Added to Map Viewer

The NCBI Map Viewer and genomes FTP area (ftp://ftp.ncbi.nih.gov/genomes/) have new genome assemblies and their corresponding annotations for the giant panda (*Ailuropoda melanoleuca*, build 1.1), and the marmoset (*Callithrix jacchus*, build 1.1) build as well as updated annotations for the rhesus monkey (*Macaca mulatta*, build 1.2) and Sumatran orangutan (*Pongo abelii*, build 1.2). The giant panda genome reported in by the January 21, 2010 issue of Nature is a whole genome shotgun assembly produced from next generation (Illumina GA) sequencing reads with approximately 60x genome coverage. Contigs are not placed on chromosomes. However, the Map Viewer provides graphical displays of the contigs, genes, gene models, and aligned carnivore expressed sequences from GenBank and RefSeq. The NCBI marmoset genome and annotation is based on the whole genome shotgun assembly released by Washington University Genome Sequencing Center and the Baylor College of Medicine Human Genome Sequencing Center as Callithrix jacchus-3.2 in March 2009. The marmoset genome comprises 22 autosomes plus X and Y sex chromosomes with mapped genes, human and marmoset expressed sequences, and STS markers. Genomic BLAST services for both panda and marmoset allow similarity search results to be displayed in genomic context in the panda and marmoset sequence maps.

RefSeq

RefSeq Release 44 is now available through the Entrez system and can be downloaded from the FTP site. This full release incorporates genomic, transcript, and protein data available as of November 7, 2010 and includes 16,421,261 records from 11,354 different species and strains. The release notes describe changes since the last release. The RefSeq Homepage has more information on the RefSeq project.

Changes affecting E-utilities

GEO Database Name changes: geo to geoprofiles

Recently the name of the GEO Profiles database in E-utilities changed from 'geo' to 'geoprofiles'. While the old name (db=geo) will still function for a time, requests should be changed to use the new name (db=geoprofiles). ELink users should be aware that all linknames including 'geo' will no longer function. Instead, these names

should include 'geoprofiles' rather than 'geo'. For example, the linkname of links from Gene to GEO Profiles is now linkname=gene_geoprofiles.

Retirement of the Journals Database: journal data has been moved to nlmcatalog

The NCBI Journals Database will be retired in mid-December, approximately on December 13, 2010. The NCBI NLM Catalog will contain the detailed MEDLINE indexing information for the journals in PubMed and other NCBI databases. ESearch URLs for db=journals will automatically map to db=nlmcatalog. ESummary and EFetch will retrieve NLM Catalog XML. The NLM Technical Bulletin has more details on this change.

New DTDs

The PubMed E-Utility DTDs will be updated for 2011 in mid-December, approximately on December 13, 2010. The new DTDs are available from the following pages.

http://eutils.ncbi.nlm.nih.gov/corehtml/query/DTD/pubmed_110101.dtd

http://eutils.ncbi.nlm.nih.gov/corehtml/query/DTD/nlmmedlinecitationset_110101.dtd

SOAP Update

The NCBI E-Utility/SOAP Web site has been updated and includes a new WSDL and examples on usage. Please consult the EUtility/SOAP homepage for more information.

Announce Lists and RSS Feeds

Eighteen topic-specific mailing lists are available which provide email announcements about changes and updates to NCBI resources including dbGaP, BLAST, GenBank, and Sequin. The various lists are described on the Announcement List summary page: www.ncbi.nlm.nih.gov/Sitemap/Summary/email_lists.html. To receive updates on the *NCBI News*, please see: www.ncbi.nlm.nih.gov/About/news/announce_submit.html.

Twelve 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/.

Users can also stay updated on NCBI's resources on Facebook and Twitter: twitter.com/NCBI.

Send comments and questions about NCBI resources to info@ncbi.nlm.nih.gov, or call 301-496-2475 between the hours of 8:30 a.m. and 5:30 p.m. EST, Monday through Friday.