

NCBI News, October 2013

Human CCDS release 14 is now available in the Gene database

Tuesday, October 29, 2013

The Consensus Coding Sequence (CCDS) update for Homo sapiens annotation release 105 was released this week. The new CCDS data is available in the [CCDS web site](#) and [FTP site](#). In addition, this update is now reflected in [relevant Gene database records](#).

The [Consensus CDS \(CCDS\) project](#) is a collaborative effort (groups from NCBI, EBI, Sanger and UCSC) to identify a core set of consistently annotated, high quality human and mouse protein coding regions. For this update, the NCBI, Ensembl, and Sanger (Havana) annotations of the most updated human reference genome ([GRCh38.p13 assembly](#), [NCBI annotation release 105](#), [Ensembl annotation release 73](#)) were analyzed.

[CCDS release 14](#) includes a total of 28,694 CCDS IDs that correspond to 18,673 GeneIDs. This update adds 978 new CCDS IDs, and adds 74 Genes into the human CCDS set.

New NCBI Insights Blog Post: Joining PubMed Commons - A step-by-step guide

Wednesday, October 23, 2013

[PubMed Commons](#) is a new system that enables researchers to share their opinions about scientific publications indexed in the PubMed database. Participation in PubMed Commons requires users with My NCBI accounts to join before they can view or add comments. A new [NCBI Insights Blog post](#) describes how to join PubMed Commons.

For more information, please see:

[PubMed Commons Homepage](#)

["Joining PubMed Commons: A Step-by-step Guide"](#)

GenBank Release 198.0 is Available

Tuesday, October 22, 2013

The new release for [GenBank](#) is now available via [FTP](#), as well as in the [Nucleotide database](#) and [BLAST services](#). *Please note* that delivery of this release missed the normal target date of October 15th due to a partial shutdown of the United States government which impacted NCBI operations. When the shutdown ended on October 17th, we expedited release processing and delivered 198.0 only a week later than usual. Our apologies for the delay.

Release 198.0 (10/17/2013) 168,335,396 non-WGS, non-CON records which were comprised of 155,176,494,699 basepairs of sequence data. In addition, there were 130,203,205 WGS records containing 535,842,167,741 basepairs of sequence data.

During the 60 days between the close dates for GenBank Releases 197.0 and 198.0, the non-WGS/non-CON portion of GenBank grew by 983,573,688 basepairs and by 1,039,556 sequence records and the WGS component of GenBank grew by 35,421,755,076 basepairs and by 5,391,185 sequence records.

The total number of sequence data files increased by 25 with this release, with the divisions that expanded in file number:

- BCT = 6 new files, now a total of 112
- CON = 11 new files, now a total of 226
- ENV = 3 new files, now a total of 65
- INV = 1 new file, now a total of 36
- GSS = 5 new files, now a total of 278
- PAT = 2 new files, now a total of 197
- PLN = 1 new file, now a total of 64
- VRL = 1 new file, now a total of 27

For downloading purposes, please keep in mind that these GenBank flatfiles are roughly 613 GB (sequence files only).

Upcoming Change: As of the December 2013 GenBank release, new CON-division WGS scaffolds will have a new accession format.

Prior to this date, WGS scaffolds constructed from WGS contigs were labeled with a '2+6' accession number format with two leading alphabetic characters followed by six digits. For example, AABR00000000.

The new accession format for newly-processed records will mirror that of the underlying WGS contigs:

- 4 letter WGS project code
- 2 digit assembly-version number
- "S" (for 'scaffold')
- Six or seven digits

For example, AABR06S000001 and AABR06S112651.

We do not currently plan to update existing records with the new accession format, but only ones that are newly-processed beginning with the 199.0 GenBank release.

For additional release information, see the [Release Notes](#) and README files in individual directories.

PubMed Commons is now live!

Tuesday, October 22, 2013

NCBI has released [PubMed Commons](#), currently in pilot phase, which is a new system that enables researchers to share their opinions about scientific publications indexed in the PubMed database. This is intended to be a forum for open and constructive criticism and discussion of scientific issues. A [new NCBI Insights Blog post](#) provides more information and explains how researchers can join in!

For more information, please see:

[PubMed Commons Homepage](#)

[NCBI Insights Blog post: "PubMed Commons - a new forum for scientific discourse"](#)

NCBI Staff will be attending the ASHG 2013 Meeting

Monday, October 21, 2013

The [2013 National Meeting for the American Society for Human Genetics \(ASHG\)](#) will be held from October 22nd through the 26th in Boston, MA. NCBI Staff members will be displaying Posters, presenting a workshop and attending the NCBI Booth to answer questions, participate in community dialog, and gain feedback from users.

Wednesday evening at 6pm in room 102 - Dr. Peter Cooper will be hosting a free workshop on "[Discovering Biological Data at the NCBI](#)".

- *This workshop will show how to use the NCBI Entrez system to perform searches and find related molecular data starting with a list of reviewed human genes.*
- **Abstract:** The National Center for Biotechnology Information (NCBI) is the premier repository for biological information in the U.S. and is the primary archive for submitter-provided data through resources such as the Sequence Read Archive (SRA), GenBank, GEO, dbSNP, dbVar and dbGaP. Resources at NCBI use the Entrez system to search various databases and display records. This workshop will give a basic introduction to using the Entrez system to perform searches and find related data starting with a list of reviewed human genes. Specific tasks covered include finding reference sequences, mapping variations, identifying homologous genes, exploring expression studies, and using MyNCBI to save searches and manage data.

Several staff members will be available at the [NCBI Booth #755](#):

- Wednesday, October 23: 10:00 am – 6:00 pm
- Thursday, October 24: 10:00 am – 4:30 pm
- Friday, October 25: 10:00 am – 2:30 pm

Some of the NCBI Posters that will be presented:

Wednesday - Oct 23, 2013 11:30am-12:30pm

1686W: "Representation of Medical Variation at NCBI: ClinVar, Gene, and MedGen."

- D. Maglott, M. Landrum, J. Lee, W. Rubinstein, K. Katz, W. Jang, D. Hoffman, S. Chitipiralla, M. Ovetsky, J. Garner, R. Tully, L. Phan, D. Shao, R. Maiti, R. Villamarin, S. Gorelenkov, S. Sherry, D. M. Church

740W: "Pharmacogenetics at NCBI."

- A. J. Malheiro, W. Rubinstein, B. Kattman, J. Lee, D. Maglott, V. Hem, M. Ovetsky, G. Song, K. Katz, C. Wallin, R. Villamarin, J. Ostell

Thursday - Oct 24, 2013 10:30am-11:30am

1441T: "Web-based tools to support the clinical genetics lab."

- D. M. Church, L. Kalman, V. Ananiev, N. Bouk, C. Chen, A. Doubintchik, M. Halavi, M. Landrum, P. Meric, L. Phan, D. Shao, D. Slotta, J. Trow, M. Ward, D. R. Maglott

1543T: "Variation data services at NCBI: archives, tools, and curation for research and medicine."

- S. Sherry, K. Address, V. Ananiev, C. Chen, D. Church, M. Feolo, J. Garner, T. Hefferon, D. Hoffman, B. Holmes, M. Kholodov, A. Kitts, J. Lee, J. Lopez, D. Maglott, R. Maiti, L. Phan, G. Riley, W. Rubinstein, D. Rudnev, Y. Shao, E. Shekhtman, K. Sirotkin, D. Slotta, R. Tully, R. Villamarin-Salomon, Q. Wang, M. Ward, H. Zhang, C. Xiao

2619T: "ClinVar: Improving Access to Clinically Relevant Variants for the Research and Clinical Genomics Communities."

- M. J. Landrum, J. Lee, G. Riley, R. Tully, S. Chitipiralla, M. Halavi, D. Hoffman, J. B. Holmes, W. Jang, K. Katz, M. Ovetsky, A. Sethi, R. Villamarin, D. M. Church, W. S. Rubinstein, D. R. Maglott

Thursday - Oct 24, 2013 11:30am-12:30pm

1546T: "The database of Genotypes and Phenotypes: dbGaP."

- M. feolo, R. Bagoutdinov, S. Dracheva, L. Hao, Y. Jin, M. Kimura, M. Lee, J. Mena, N. Popova, S. Pretel, N. Sharopova, S. Stefanov, A. Stine, A. Sturcke, K. T. Tryka, Z. Wang, M. Xu, L. Ziyabari, S. T. Sherry

1594T: "Change can be good: updating the human reference genome assembly."

- V. A. Schneider, P. Flicek, T. Graves, T. Hubbard, D. M. Church for the Genome Reference Consortium

2628T: "The NIH Genetic Testing Registry: 2013 status report on genetic testing."

- W. S. Rubinstein, B. L. Kattman, A. J. Malheiro, J. M. Lee, D. R. Maglott, V. Hem, M. Ovetsky, G. Song, C. Wallin, K. S. Katz, R. Villamarin-Salomon, C. Fomous, J. M. Ostell

Organism BLAST pages now use top-level RefSeq genomic records instead of scaffold records

Monday, October 21, 2013

The organism BLAST pages are being updated to use top-level (chromosome + unplaced and unlocalized scaffolds) RefSeq genomic records instead of scaffold records. This change has also been made for the human and mouse G+T BLAST databases. Reporting hits in chromosome coordinates is more useful for public reporting and also makes it easier to relate the results to data on other sites.

For more information, see this BLAST News Story on the "[Update to organism BLAST databases](#)".