NCBI News, August 2013

Human genome annotation release 105 with new splice variants

Tuesday, August 27, 2013

The NCBI recently finished a re-annotation of three complete (GRCh37.p13, CHM1_1.1, and HuRef) assemblies and one partial (CRA_TCAGchr7v2) assembly of the human genome. Annotated genomic, transcript, and protein records are available through the integrated Entrez system (Nucleotide, Protein, Gene) and may be downloaded through FTP or the Aspera protocol. The new annotation provides additional splice variants for many human genes. Genes may now be annotated with both known Reference Sequences (NM_ style accessions) and gene models (XM_ style accessions.) RefSeq models are generated from mRNA, protein, and RNAseq data. Twelve thousand genes are now annotated with both known and model RefSeqs on the GRCh37.p13 assembly, approximately doubling the number of splice variants represented.

This is NCBI's last full annotation of the GRCh37 assembly. The next full annotation release for human will include GRCh38. See the Genome Reference Consortium site for information on the upcoming human genome build.

## dbSNP Build 138, phase III, now available

Thursday, August 22, 2013

dbSNP build 138 phase III update is now available. This update includes data for mouse, *Arabidopsis thaliana*, honeybee, *C. elegans*, and rice. Build 138 provides more than 505 million submitted and 226 million reference variants for 131 species. To see complete build statistics visit the SNP summary page. You may access build 138 SNP data through the integrated NCBI Entrez system and download data through FTP or Aspera protocol.

Sequence Viewer 2.27: new features, improvements, and help documentation

Wednesday, August 21, 2013

Sequence Viewer 2.27 (Release Notes) is now appearing on the NCBI site (nucleotide, protein, gene, SNP) and available for embedding in outside pages. Version 2.27 has important new features and improvements including new tiling path and contig (scaffold) tracks for assembled records, drag and drop reordering and one click removal of tracks within the graphical view. The Sequence Viewer also now has a reorganized and improved help documentation site.
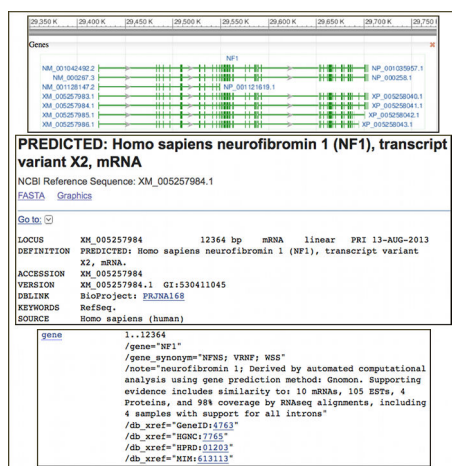
New splice variants for the human NF1 gene (Gene ID: 4763). Top panel. A graphical view of chromosome 17 showing the seven splice variants for NF1, the last three are predictions based on mRNA, protein, and RNAseq data. Lower panels. A Nucleotide database view of splice variant X2 (XM_005257984) showing the gene feature with supporting evidence.
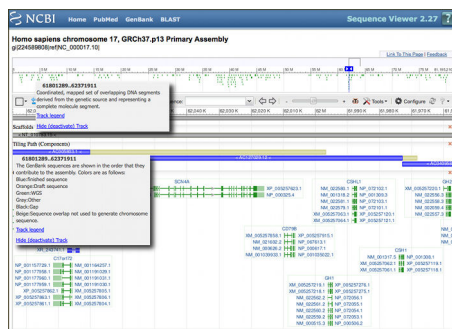
Sequence viewer 2.27 showing a region of chromosome 17 near the SCN4A gene. The new Tiling Path and Scaffolds tracks show how the region is assembled. Tracks now can be dragged to new positions using the mouse and may be dismissed by clicking the red 'x' at the right hand margin.

>10,000 tests now listed in the NIH Genetic Testing Registry

Wednesday, August 21, 2013

The NIH Genetic Testing Regsitry (GTR) is a free online resource that provides centralized access to comprehensive genetic test information voluntarily submitted by test providers. As of August 19, 2013, more than 10,000 tests for over 3,350 conditions have been submitted by laboratories.

For information on how to submit data to GTR, see the documentation describing the GTR Submission Process.

GenBank Release 197.0 is Available

Friday, August 16, 2013

The new release for GenBank is now available via FTP, as well as in the Nucleotide database and BLAST services.

Release 197.0 (08/15/2013) 167,295,840 non-WGS, non-CON records which were comprised of 154,192,921,011 basepairs of sequence data. In addition, there were 124,812,020 WGS records containing 500,420,412,665 basepairs of sequence data.

During the 60 days between the close dates for GenBank Releases 196.0 and 197.0, the non-WGS/non-CON portion of GenBank grew by 1,593,690,899 basepairs and by 1,555,676 sequence records and the WGS component of GenBank grew by 46,590,660,345 basepairs and by 12,323,984 sequence records.

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

- BCT = 3 new files, now a total of 106
- CON = 7 new files, now a total of 215
- ENV = 1 new file, now a total of 62
- EST = 1 new file, now a total of 474
- GSS = 5 new files, now a total of 278
- PLN = 1 new file, now a total of 63
- PRI = 1 new file, now a total of 46
- TSA = 4 new file, now a total of 145
- VRL = 1 new file, now a total of 26
- VRT = 1 new file, now a total of 31

For downloading purposes, please keep in mind that these GenBank flatfiles are roughly 607 GB (sequence files only).

For additional release information, see the Release Notes and README files in individual directories.