



## NCBI News, April 2016

### dbSNP build 147 data for human, chicken, soybean and more are available

*Friday, April 29, 2016*

dbSNP Build 147 is accessible on the web and via [FTP](#). This release includes data for human, chicken, tilapia, mallard, sheep, date palm and soybean. Build 147 provides over 745 million submitted variants and 250 million reference variants for 7 organisms. To see complete build statistics, visit the [SNP summary page](#).

dbSNP, the NCBI Short Genetic Variations database, catalogs short variations in nucleotide sequences from a wide range of organisms.

### Eukaryotic Genome Annotation Pipeline now directly annotates top-level sequences, not scaffolds

*Wednesday, April 27, 2016*

The [Eukaryotic Genome Annotation Pipeline](#) has been modified to directly annotate RefSeq assemblies' top-level sequences (chromosomes, and unplaced and unlocalized scaffolds) instead of scaffolds. This change, included in [software release 7.0](#), improves the annotation of features spanning gaps between adjacent scaffolds, and applies to all upcoming annotation releases, including human (scheduled for May 2016).

The consequence is that for genomes assembled to the level of chromosomes, the annotation is no longer reported on placed scaffolds, and is only available on chromosomes. Specific changes include:

#### In Nucleotide:

- GenBank, Graphics and ASN views of RefSeq placed scaffolds no longer show any annotation.
- ASN view of RefSeq chromosomes now include the annotation.

#### On the FTP site:

- GFF files are now only provided for top-level sequences.
- Files in the CHR\_\* directories for nuclear chromosomes no longer include annotation on placed scaffolds.
- Masked spans ([masking\\_coordinates.gz](#)) are now in top-level coordinates.
- Comparison of current to previous annotation are now in top-level coordinates.

Here are examples from a [recent annotation of the platypus genome](#) that illustrate the change:

- [Chromosome](#)
- [Placed scaffold record](#)

- [FTP directory](#)

To see all organisms annotated by the Eukaryotic Genome Annotation Pipeline, click [here](#).

## May 4th NCBI Minute: Linking PubMed and ClinicalTrials.gov

*Tuesday, April 26, 2016*

Next Wednesday, May 4th, NCBI will present a short tutorial that will teach you two ways to filter PubMed searches for publications linked to clinical trials in [clinicaltrials.gov](#); you'll also learn how to use the ClinicalTrials database to get more information on trials of interest.

**Date and time:** May 4, 2016 12:00pm EDT

**Registration link:** <https://attendee.gotowebinar.com/register/8673331823519860737>

After registering, you will receive a confirmation email with information about attending the webinar. After the live presentation, the webinar will be uploaded to the [NCBI YouTube channel](#). Any related materials will be accessible on the [Webinars and Courses page](#); you can also learn about future webinars on this page.

ClinicalTrials.gov is a registry and results database of publicly and privately supported clinical trials.

## New NCBI video on YouTube: "Sequence Viewer: Display dbVar Supporting Calls"

*Monday, April 25, 2016*

The newest video on the NCBI YouTube channel, [Sequence Viewer: Display dbVar Supporting Calls](#), demonstrates a new feature for dbVar tracks in [Sequence Viewer](#). You can now toggle the track display to show or hide supporting variant calls, or children, for the parent structural variant.

Subscribe to the [NCBI YouTube channel](#) to receive alerts about new videos ranging from quick tips to full webinar presentations.

## Webinars on April 29 present BLAST, human variation & medical genetic records

*Friday, April 22, 2016*

On April 29th, NCBI will host two webinars, *A Practical Guide to NCBI BLAST* and *NCBI Human Variation and Medical Genetic Resources*. Each webinar will provide an overview of the respective resources and show you how to use them.

### ***A Practical Guide to NCBI BLAST***

This webinar highlights important features and demonstrates the practical aspects of using the NCBI BLAST service, the most popular sequence similarity service in the world.

You will learn about useful but under-used features of the service, including: access from the Entrez sequence databases; the new genome BLAST service quick finder; the integration and expansion of Align-2-Sequences; organism limits and other filters; re-organized databases; formatting options and downloading options; and TreeView displays.

You will also learn how to use other important sequence analysis services associated with BLAST including Primer BLAST, an oligonucleotide primer designer and specificity checker; the multiple protein sequence alignment tool, COBALT; and SmartBLAST, a new tool for rapid protein identification. These aspects of BLAST provide easier access and results that are more comprehensive and easier to interpret.

**Date and time:** Apr 29, 2016 1:30-2:30pm

**Registration link:** <https://nih.webex.com/nih/onstage/g.php?d=629972037&t=a>

## **NCBI Human Variation and Medical Genetic Resources**

Through this webinar, you will learn to use and access resources associated with human sequence variations and phenotypes associated with specific human genes and phenotypes. The webinar will emphasize the Gene, MedGen and ClinVar resources to search by gene, phenotype and variant respectively. You will learn how to map variation from dbSNP and dbVar onto genes, transcripts, proteins, and genomic regions and how to find genetic tests in GTR. You will also gain experience using additional tools and viewers including PheGenI, a browser for genotype associations, the Variation Viewer and the 1000 Genomes Browser. These provide useful ways to search for, map and browse variants as well as upload and download data in genomic context.

**Date and time:** Apr 29, 2016 2:45-4:00pm

**Registration link:** <https://nih.webex.com/nih/onstage/g.php?d=626275627&t=a>

After registering for each webinar, you will receive a confirmation email with information about attending. After the live presentation, the webinar will be uploaded to the [NCBI YouTube channel](#). Any related materials will be accessible on the [Webinars and Courses page](#) as well as a schedule of future webinars.

## **NCBI to assist UC Davis in June hackathon**

*Tuesday, April 19, 2016*

From June 13th to 15th, NCBI will assist the University of California Davis in hosting a biomedical data science hackathon at the [School of Veterinary Medicine](#) in Davis, CA, focusing on advanced bioinformatics analysis of next generation sequencing data and metadata. This event is for students, postdocs, investigators and other researchers already engaged in the use of pipelines for genomic analyses from next-generation sequencing data or metadata.\*

\*Some projects are available to other non-scientific developers, mathematicians or librarians.

Researchers and/or data scientists from the west coast of the United States are especially encouraged to apply, but the event is open to anyone selected for the hackathon, and able to travel to Davis. Working groups of 5-6 individuals will be formed into five or six teams. These teams will build pipelines and tools to analyze large datasets within a cloud infrastructure. The potential subjects for this iteration are:

1. Medical informatics
2. Cancer immunogenicity
3. Workflow languages
4. Sequencing contamination
5. Metagenomics
6. Metadata
7. Closing bacterial genomes

Please see the [application](#) for specific team projects.

## Organization

After a brief organizational session, teams will spend three days analyzing a challenging set of scientific problems related to a group of datasets. Participants will analyze and combine datasets in order to work on these problems.

## Datasets

Datasets will come from the public repositories housed at the NCBI. During the course, participants will have an opportunity to include other datasets and tools for analysis. Please note, if you use your own data during the course, we ask that you submit it to a public database within six months of the end of the event.

## Products

All pipelines and other scripts, software and programs generated in this course will be added to a [public GitHub repository designed for that purpose](#). A manuscript outlining the design and usage of the software tools constructed by each team will be submitted to an appropriate journal.

## Application

To apply, complete [this form](#) (approximately 10 minutes to complete). Applications are due **May 5, 2016 by 5PM Eastern**. Participants will be selected from a pool of applicants based on the experience and motivation they provide on the form. Prior participants and applicants are especially encouraged to reapply.

Accepted applicants will be notified on May 9, 2016 by 2PM Eastern, and have until May 12, 2016 at 9AM Eastern to confirm their participation. If you confirm, please make sure it is highly likely you can attend, as confirming and not attending bars other data scientists from attending this event. Please include a monitored email address, in case there are follow-up questions.

**Note:** Participants will need to bring their own laptop to this program. A working knowledge of scripting (e.g., Shell, Python) is necessary to be successful in this event. Employment of higher level scripting or programming languages may also be useful. Applicants must be willing to commit to all three days of the event. No financial support for travel, lodging or meals is available for this event. Also note that the course may extend into the evening hours on Monday and/or Tuesday. Please make any necessary arrangements to accommodate this possibility.

Please contact [ben.busby@nih.gov](mailto:ben.busby@nih.gov) with any questions. Finally, if you are interested in having NCBI facilitate a regional hackathon hosted at your institution, please fill out [this form](#).

## New NCBI video on YouTube: Navigating the NIH Manuscript Submission Process

*Monday, April 18, 2016*

The newest video on the NCBI YouTube channel, [Navigating the NIH Manuscript Submission Process](#), gives you detailed help with submitting, reviewing and approving your manuscript in the [NIH Manuscript Submission \(NIHMS\)](#) system. The NIHMS system supports manuscript depositing into PubMed Central (PMC) as required by the public access policies of NIH and other participating funding agencies.

Subscribe to the [NCBI YouTube channel](#) to receive alerts about new videos ranging from quick tips to full webinar presentations.

U.S. Department of Health & Human Services

About Help

**NIH** Manuscript Submission System

**Sign In**

1 NIH Researchers  

HHMI Researchers 

2 Publishers and Others 

Sign-In Help  
Forgot your sign-in route? [Request E-mail Reminder](#)

The NIH Manuscript Submission (NIHMS) system supports the deposit of manuscripts into PubMed Central (PMC), as required by the public access policies of NIH and other participating funders.

[Learn More](#)

### 3 Manuscript Submission Process



## Articles in Nucleic Acids Research Database 2016 Issue discuss NCBI databases, updates and future plans

Tuesday, April 12, 2016

The 23rd annual edition of the [Nucleic Acids Research Database Issue](#) features several papers from NCBI staff that describe the current state of our databases, recent updates and future plans to improve their use.

The NCBI database articles in NAR are also available from [PubMed](#). To read an article, click on the PMID listed below:

- "The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection" by Daniel J. Rigden, Xose M. Fernandez-Suarez and Michael Y. Galperin (PMID: [26740669](#))
- "Database resources of the National Center for Biotechnology Information" by NCBI Resource Coordinators (PMID: [26615191](#))
- "The International Nucleotide Sequence Database Collaboration" by Guy Cochrane, Ilene Karsch-Mizrachi, Toshihisa Takagi and INSDC (PMID: [26657633](#))
- "[GenBank](#)" by Karen Clark, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell and Eric W. Sayers (PMID: [26590407](#))

- "[Assembly](#): a resource for assembled genomes at NCBI" by Paul A. Kitts, Deanna M. Church, Françoise Thibaud-Nissen, Jinna Choi, Vichet Hem et al. (PMID: [26578580](#))
- "Reference sequence ([RefSeq](#)) database at NCBI: current status, taxonomic expansion, and functional annotation" by Nuala A. O'Leary, Matthew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad et al. (PMID: [26553804](#))
- "ClinVar: public archive of interpretations of clinically relevant variants" by Melissa J. Landrum, Jennifer M. Lee, Mark Benson, Garth Brown, Chen Chao et al. (PMID: [26582918](#))
- "PubChem Substance and Compound databases" by Sunghwan Kim, Paul A. Thiessen, Evan E. Bolton, Jie Chen, Gang Fu et al. (PMID: [26400175](#))

## Maximizing PubChem: webinar on April 20th will cover new and future features

*Wednesday, April 06, 2016*

In two weeks, NCBI staff will discuss features recently added to [PubChem](#), as well as upcoming changes to the resource.

**Date and Time:** April 20, 2016 1:00 PM ET

**Registration link:** <https://attendee.gotowebinar.com/register/2150693495841803266>

After registering, you will receive a confirmation email with information about attending the webinar. After the live presentation, the webinar will be uploaded to the [NCBI YouTube channel](#). Any related materials will be accessible on the [Webinars and Courses page](#); you can also learn about future webinars on this page.

PubChem, which provides information on the biological activities of small molecules, has been under active development. The resource is organized as three linked databases within the NCBI's [Entrez](#) information retrieval system: PubChem Substance, PubChem Compound, and PubChem BioAssay. PubChem also provides a fast chemical structure similarity search tool.