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

NLM Citation: NCBI News [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 1991-2012. NCBI News, November 2015.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/

NCBI News, November 2015

dbSNP human build 146 available through Entrez and FTP

Wednesday, November 25, 2015

dbSNP human Build 146, based on the GRCh38.p2 and GRCh37.p13 assemblies, is now available on the integrated NCBI Entrez system and through FTP. Build 146 provides 150 million Reference SNP (RS) clusters, including 985,775 new RS clusters and allele frequency data from 1000 Genomes, GO-ESP, and ExAC projects. To see complete build statistics, visit the dbSNP summary page.

Tree Viewer 1.7.5 now available

Tuesday, November 24, 2015

NCBI Tree Viewer version 1.7.5 has several new features, improvements and bug fixes, including improved subtree de-selection function, fixed BLAST Tree View bugs, and a fixed mouse wheel zoom bug. To see the full list of updates, see the Tree Viewer release notes.

NCBI Tree Viewer is a tool for viewing your own phylogenetic tree data.

NCBI releases first five lectures of NCBI NOW on YouTube

Tuesday, November 24, 2015

Today, the first five lectures from the NCBI NOW workshop are available in a playlist on the NCBI YouTube channel. Last month, NCBI presented this online workshop (more information here) to 650 participants new to next generation sequencing (NGS) analysis.

Subscribe to the NCBI YouTube channel to receive notifications about our new videos, which range from quick tips to full webinar presentations.

December 2nd NCBI Minute webinar: Finding Genes in PubMed

Monday, November 23, 2015

Next Wednesday's NCBI Minute will show you how to quickly find literature about a gene of interest using PubMed. NLM staff will highlight the links between gene data and literature and help you leverage the vocabulary used to describe gene information in PubMed to build a better search.

Date and time: Dec 2, 2015, 12:00-12:15 PM EST

Registration URL: https://attendee.gotowebinar.com/register/6661858858940556801

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 find information about future webinars on this page.

The NCBI Minute is a series of short webinars that give a brief introduction to a specific topic or NCBI tool.

New video on the NCBI YouTube channel: "Explore Gene pages at NCBI: Variation and Expression"

Wednesday, November 18, 2015

The newest video on the NCBI YouTube channel, *Explore Gene Pages at NCBI: Variation and Expression*, provides a walkthrough of how to explore variation and expression data using the Sequence Viewer embedded on Gene pages. Use these additional track options to quickly enhance your understanding of your genes of interest.

Subscribe to the NCBI YouTube channel to receive alerts about new videos ranging from quick tips to full webinar presentations.

PubChem adds a legacy designation for outdated data

Tuesday, November 17, 2015

PubChem has introduced a "legacy" designation to help users quickly identify records that may have outdated information or links. The designation applies to projects or contributors that appear to be inactive, as well as to their individual records. The latest post on the PubChem blog explains more about the designation, its impact, and its future in PubChem.

NCBI to hold three-day genomics hackathon in January

Friday, November 13, 2015

From January 4th to 6th, NCBI will host a genomics hackathon focusing on advanced bioinformatics analysis of next generation sequencing data. This event is for students, postdocs and investigators already engaged in the use of pipelines for genomic analyses from next generation sequencing data.* Working groups of 5-6 individuals will be formed for twelve teams, in the following sections: Network Analysis of Variants, Structural Variation, RNA-Seq, Streaming Data and Metadata, and Neuroscience/Immunity. The working groups will build pipelines to analyze large datasets within a cloud infrastructure. Please see the application for specific team projects.

** Specific projects are available to other developers or mathematicians.*

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. This course will take place at the National Library of Medicine on the NIH main campus in Bethesda, Maryland.

Datasets

Datasets will come from the public repositories housed at 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 hackathon.

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 of the hackathon and describing participant processes, products and scientific outcomes will be submitted to an appropriate journal.

Application

To apply, complete this form (approximately 10 minutes to complete). Applications are due by **5pm ET on December 1**. Participants will be selected from a pool of applicants; prior students and prior applicants will be given priority in the event of a tie. Please note: applicants are judged based on the motivation and experience outlined in the form itself.

Accepted applicants will be notified on **December 4th by 2 pm ET**, and have until **5pm on December 7** to confirm their participation. 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 can be provided 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 with any questions.

Sequence Viewer 3.10.5 adds support for track sets with non-default options

Friday, November 06, 2015

Sequence Viewer 3.10.5 is now available with support for track sets with non-default display options. There are also a number of bug fixes, which are listed in the release notes.

Sequence Viewer is a graphical view of sequences and color-coded annotations on regions of sequences stored in the Nucleotide and Protein databases.

RefSeq Release 73 is now available

Friday, November 06, 2015

RefSeq Release 73 is now accessible online, on the FTP site, and through NCBI's programming utilities. This full release incorporates genomic, transcript and protein data available as of November 2, 2015 and includes 83,881,439 records, 54,766,170 proteins, 12,998,293 RNAs, and sequences from 55,966 organisms. More information can be found in the release notes.

For more information about the RefSeq project, please take a look at the RefSeq homepage.

Tree Viewer 1.7 now available

Thursday, November 05, 2015

NCBI Tree Viewer version 1.7 includes several new features, improvements and bug fixes, including a new rendering mechanism for displaying very large trees as image tiles. To see the full list of updates, see the Tree Viewer release notes.

NCBI Tree Viewer is a tool for viewing your own phylogenetic tree data.

Registration open for November 18 NCBI Minute, "The New ClinVar Submission Wizard"

Wednesday, November 04, 2015

Next Wednesday, November 18th, the NCBI Minute will be an introduction and demonstration of the new ClinVar Submission Wizard, a guided interface for direct data entry made for research laboratories that want to occasionally submit a small number of records.

Date & time: Nov 18, 2015, 12:00-12:15 pm EST

Registration URL: https://attendee.gotowebinar.com/register/7270975107138338562

Submission to ClinVar is usually done through the Variation Submission Portal, which is useful for groups who frequently submit large number of variants, but may not be convenient for infrequent submitters of small numbers of variants. The new Submission Wizard is designed to support all types of submissions to ClinVar, including structural variants, pharmacogenomics variants, somatic variants, as well as interpretations based on functional rather than clinical significance.

After the live presentation, this webinar will be uploaded to the NCBI YouTube channel. Any related materials will be accessible on the Webinars and Courses page; you can also find information about future webinars on this page.

ClinVar is the NCBI archive of submitted interpretations of variants relative to diseases and other phenotypes.

The NCBI Minute is a series of short webinars that give a brief introduction to a specific topic or NCBI tool.

Researchers identify potential alternative to CRISPR-Cas genome editing tools

Wednesday, November 04, 2015

An international team of CRISPR-Cas researchers has identified three new naturally-occurring systems that show potential for genome editing. The discovery and characterization of these systems is expected to further expand the genome editing toolbox, opening new avenues for biomedical research. The research, published October 22nd in the journal Molecular Cell, was supported in part by the National Institutes of Health.

"This work shows a path to discovery of novel CRISPR-Cas systems with diverse properties, which are demonstrated here in direct experiments," said Eugene Koonin, Ph.D., senior investigator at the National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), part of the NIH. "The most remarkable aspect of the story is how evolution has achieved a broad repertoire of biological activities, a feat we can take advantage of for new genome manipulation tools."

Enzymes from the CRISPR system are revolutionizing the field of genomics, allowing researchers to target specific regions of the genome and edit DNA at precise locations. "CRISPR" stands for Clustered Regularly Interspaced Short Palindromic Repeats, which are key components of a system used by bacteria to defend against invading viruses. Cas9 - one of the enzymes produced by the CRISPR system - binds to the DNA in a

highly sequence-specific manner and cuts it, allowing precise manipulation of a region of DNA. Enzymes such as Cas9 provide researchers with a gene editing tool that is faster, less expensive and more precise than previously developed methods.

The three newly-characterized systems share some features with Cas9 and Cpf1, a recently characterized CRISPR enzyme, but have unique properties that could potentially be exploited for novel genome editing applications. This study highlights the diversity of CRISPR systems, which can be leveraged to develop more efficient, effective, and precise ways to edit DNA.

The researchers took a novel bioinformatics approach to discover the new proteins, provisionally termed C2c1, C2c2, and C2c3, developing a series of computational approaches to search NIH genomic databases and identify new CRISPR-Cas systems. In addition to Koonin, the research team included Feng Zhang of the Broad Institute of MIT and Harvard and the McGovern Institute for Brain Research at MIT, Konstantin Severinov of Rutgers University – New Brunswick and the Skolkovo Institute of Science and Technology, Omar Abudayyeh, a graduate student at the Harvard- MIT Division of Health Sciences and Technology, and NCBI's Kira Makarova, Sergey Shmakov (also at Skolkovo Institute of Science and Technology), and Yuri Wolf.

"There are multiple ways to modify the search algorithm, so more exciting and distinct CRISPR-Cas mechanisms should be expected soon," said Severinov. "These new mechanisms will undoubtedly attract the attention of basic and applied scientists alike."

Initial experimental work exploring the function of these proteins reveals that they are substantially different from the well-characterized Cas9 protein, which has been widely used for genome editing.

With the analysis of C2c1, C2c2, and C2c3, the team was able to infer the intricate evolutionary pathway of these adaptive defense systems.

"The collaborative nature of this work highlights the power of bringing together top scientists with diverse strengths to innovate at the interface of computation, molecular biology and evolutionary biology," said Zhang.

The Koonin and Zhang groups also recently collaborated on a project that resulted in the characterization of Cpf1, a novel CRISPR nuclease that is expected to become an important genome editing tool.

Feng Zhang, of the Broad Institute and MIT, is supported by the National Institute of Mental Health (5DP-MH100706 and 1R01-MH110049) and by the National Institute of Diabetes and Digestive and Kidney Diseases (5R01DK097760-03).

Konstantin Severinov, of Rutgers University and the Skolkovo Institute of Science and Technology, is supported by National Institute of General Medical Sciences (GM10407).

About the National Center for Biotechnology Information (NCBI): NCBI creates public databases in molecular biology, conducts research in computational biology, develops software tools for analyzing molecular and genomic data, and disseminates biomedical information, all for the better understanding of processes affecting human health and disease. NCBI is a division of the National Library of Medicine. For more information, visit www.ncbi.nlm.nih.gov.

About the National Library of Medicine (NLM): The world's largest biomedical library, NLM maintains and makes available a vast print collection and produces electronic information resources on a wide range of topics that are searched billions of times each year by millions of people around the globe. It also supports and conducts research, development, and training in biomedical informatics and health information technology. Additional information is available at www.nlm.nih.gov.

About the National Institutes of Health (NIH): NIH, the nation's medical research agency, includes 27 Institutes and Centers and is a component of the U.S. Department of Health and Human Services. NIH is the

primary federal agency conducting and supporting basic, clinical, and translational medical research, and is investigating the causes, treatments, and cures for both common and rare diseases. For more information about NIH and its programs, visit www.nih.gov.

dbVar publishes October 2015 data release

Tuesday, November 03, 2015

The dbVar October 2015 data release has recently been published. This month's release has 109,463 new Variant regions, 165,519 new Variant calls, and 5 new studies, including the following:

- **Decker et al. 2015** (nstd115): The authors created the largest existing catalog of canine genome-wide variation and used it to identify somatic variation in the thousands-years-old parasitic cancer clone, canine transmissible veneral tumor (CTVT).
- **User submitted curated variants** (nstd51): A significant update to this collection of clinically relevant structural variants, curated by NCBI staff from PubMed, OMIM, and GeneReviews.
- **LSDB submitted variants** (nstd103): A new collection of clinically relevant structural variants submitted by public LSDBs to ClinVar and brokered to dbVar.

Follow the dbVar RSS feed for monthly releases.

Registration open for November 12th webinar, "PubMed for Scientists"

Monday, November 02, 2015

On November 12th, NCBI will present "PubMed for Scientists", a webinar that will show you how to search biomedical literature more efficiently with PubMed. NCBI staff will teach you how to search by author, explore a subject, use filters to narrow your search, find full text articles, and set up an email alert for new research on your topic. Finally, we will answer your questions about searching PubMed.

Date and time: Thursday, November 12, 2015 12:30 PM - 1:30 PM

Registration URL: https://attendee.gotowebinar.com/register/5594790520765285889

After the live presentation, the webinar will be uploaded to the NCBI YouTube channel. The webinar and any materials will also be accessible on the Webinars and Courses page by clicking the Archived Webinars & Courses tab. You can also check the Webinars & Courses page to find information about future webinars.