NCBI News, November 2016

Evidence Viewer has been retired

Tuesday, November 22, 2016

Evidence Viewer, which was designed to show biological evidence supporting curated gene models, has been retired. Current evidence supporting annotated gene structures is included in the "Genomic regions, transcripts and products" section of the Gene database for organisms annotated using NCBI's Eukaryotic Genome Annotation Pipeline.

The RNA-Seq exon coverage tracks in in the graphical Sequence Viewer show the aggregate exon and intron coverage and individual intron features based on RNA-Seq data in the SRA database. For an example, take a look at the information displayed on the Human EGFR gene page.

NCBI, NLM, NHGRI to hold on-campus hackathon January 9-11

Thursday, November 17, 2016

From January 9th to 11th, the NCBI, with involvement from several NIH institutes, will host a biomedical data science hackathon at the National Library of Medicine. To apply for this hackathon, complete this application (approximately 10 minutes to complete). Applications are due **December 7th by 4 PM ET**.

This hackathon will primarily focus on:

- Medical informatics,
- Advanced bioinformatics analysis of next generation sequencing data,
- And metadata.

This event is for students, postdocs and investigators or other researchers already engaged in the use of medical informatics data or pipelines for genomic analyses from next generation sequencing data. However, there are some projects available to other non-scientific developers, mathematicians or librarians. The event is open to anyone selected for the hackathon who is able to travel to NIH.

Organization

There will be 5-7 teams of 5-6 individuals. These teams will build pipelines and tools to analyze large datasets within a cloud infrastructure.

The potential subjects for this iteration are:

- HL-7 compliance of myfamilyhealthportrait and GTR,
- Integrating cbioportal with TCGA and dbGaP metadata,
- Variant and gene screening from PubMed,
- Code discovery in PubMed Central,
- Integration of PubChem with other chemical datasets,
- Auto-updating dictionaries for natural language processing,
- And several others.

Please see the application for specific and evolving team projects. Again, some projects are available to other non-scientific developers, mathematicians or librarians.

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, primarily those 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 may be submitted to an appropriate journal such as the F1000Research hackathons channel.

Application

To apply, complete this application (approximately 10 minutes to complete). Applications are due **December 7th, 2016 by 4 PM ET**.

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.

The first round of accepted applicants will be notified on December 9th by 5 pm ET, and have until December 12th at 4 PM 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.

Notes

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 with any questions.

Genome Workbench 2.11.5 now available

Wednesday, November 16, 2016

The latest version of Genome Workbench includes a number of new features, fixes and improvements like the use of encrypted HTTPS protocol, multiple feature table loading, and improved exporting.

For a full list of changes, please see the Genome Workbench release notes.

November 17th webinar: NCBI Resources for Agricultural Research

Tuesday, November 08, 2016

On November 17th, NCBI will present a workshop for researchers interested in agriculturally important organisms.

Date and time: Thursday, November 17, 2016 1:00 – 2:00 PM EST

Registration URL: https://attendee.gotowebinar.com/register/42857142156965378

In the first part of the webinar, participants will learn to effectively use the NCBI website and BLAST to find relevant data including sequence, variation, gene and expression

information. The second part of the webinar will focus on accessing large-scale genomics datasets.

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.

RefSeq release 79 now available

Monday, November 07, 2016

RefSeq release 79 is now accessible online, via FTP and through NCBI's programming utilities. This full release incorporates genomic, transcript, and protein data available as of October 31, 2016 and contains 111,024,999 records, including 73,099,060 proteins, 16,967,019 RNAs, and sequences from 64,277 organisms. The release is provided in several directories as a complete dataset and also as divided by logical groupings.

As announced in March 2016, NCBI has implemented the removal of GI numbers from some presentations of nucleotide and protein sequence records. The affected presentations are the (default) flat file (GenBank and GenPept) views, along with FASTA views. This change affects only the web views of these two file formats. Note that this change is not reflected in this ftp release. However, GI sequence identifiers will be removed from flatfile and FASTA formats in the March 2017 RefSeq FTP release.

For more details on GI numbers, please see these announcements:

- The Future of Existing GI Numbers at NCBI
- GI numbers will be removed from sequence record presentations

New NCBI Insights post: Identifying and Correlating Chemical Names and Synonyms

Monday, November 07, 2016

Identifying and Correlating Chemical Names and Synonyms is the latest post on the NCBI Insights blog. This blog post briefly describes the PubChem system of identifying and correlating varying names for one chemical.

New NCBI Insights blog post: Clearing Up Confusion with Human Gene Symbols and Names

Monday, November 07, 2016

The newest blog post on NCBI Insights shows you how to use NCBI resources to find and reference official gene names or symbols, as well as synonyms or aliases that refer to the same gene. The blog post also includes a video showing how to correctly import gene

symbol data into Excel, avoiding the problems caused by the autocorrect and autoformat functions in spreadsheet applications.

Permanent redirect to HTTPS will occur on November 10, 2016

Wednesday, November 02, 2016

Starting on November 10th, NCBI will begin a permanent redirect to HTTPS. More specifically, all HTTP traffic for GET and HEAD requests will be redirected. All other requests will be rejected.

The HTTPS at NCBI page provides further guidance for NCBI web API users.

New video on YouTube: The New MSA Viewer

Tuesday, November 01, 2016

The newest video on the NCBI YouTube channel introduces the new multiple sequence alignment (MSA) viewer for amino acid and nucleotide sequences. This short video demonstrates MSA Viewer's basic functions.

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