

To our CGC Bioinformatics Workshop Participants,

Drs. Rachel Karchin, Alex Wagner and Deb Ritter are so excited to welcome you to the 2022 Bioinformatics Workshop at the Cancer Genomics Consortium Annual Meeting! The workshop will be held Sunday July 31st, 9:00 - 10:30 am CST.

https://www.cancergenomics.org/meetings/2022_pre-meeting_workshops.php

We have filled our 1.5hrs to the brim with active learning and we hope you come away from the workshop knowing how to run single, batch and full VCF files through the OpenCRAVAT annotation platform! For some, it might be the first time you learn where the Terminal is on your Mac! For others, you might be helping US with pip install of OpenCRAVAT, or sharing a nuance of applying annotated data to variant classification.

To make the most of this workshop, we ask the following in advance:

- (1) Create an account on the OpenCRAVAT platform, here: <https://opencravat.org/> . Go to “Get Started” on the left side and follow instructions to create an account. This will let you run batch variants through the web-interface.
- (2) Download at least 1 of the 5 example VCF files for batch processing that are in this directory:
<https://drive.google.com/drive/folders/1q4XjLEBsu4BYiOVnecz5qpVwHRHXOdEP?usp=sharing>
- (3) Download the Variant Report Workflow, Batch Process and Genie Cohort Analysis slides, so that you have them to refer to.
 - Variant Report Workflow:
https://docs.google.com/presentation/d/1tQuv478epbzhra_ztPRG6NEEI0J6ouQ5/edit?usp=sharing&oid=115007761664968985806&rtpof=true&sd=true
 - Batch Process:
<https://docs.google.com/presentation/d/15fLtlIlgXxtc1qoO3GuGj5PQCuMPGpg/edit?usp=sharing&oid=115007761664968985806&rtpof=true&sd=true>
 - AACR GENIE Cohort Analysis:
<https://docs.google.com/presentation/d/1y99hcYhp0IHfk5rog8iqBqOog-KdEJ1S/edit?usp=sharing&oid=103262517578037509440&rtpof=true&sd=true>
- (4) Download a Project Genie MAF file (~400MB) and cut it in half by doing the following:
 - Create an account with Synapse through SageBionetwork
 - <https://www.synapse.org/> – use Google account or register with another email
 - Once you have an account, login and go to Project GENIE in Synapse: <https://www.synapse.org/#!Synapse:syn32689317>
 - This goes to the page to download the file “data_mutations_extended.txt”
 - Click the link next to the **yellow “Access” Key symbol** in the middle of the page.
 - Agree to terms and conditions in the pop-up box, and close the box.

- REFRESH THE SCREEN
- The Access should now look like a **green unlocked lock**. The download button on the right should be blue and clickable.
- Download the file using the blue download button. You are downloading variants in the Mutation Annotation Format (MAF) file format.
- You will need to convert it to the TSV format for OpenCravat and can do so using awk on your Terminal (in Mac) by using the command below. Open Terminal by going to Applications -> Utilities -> Terminal.app. You will need to either navigate to the directory where the file is downloaded, or go to Downloads. You can do that in Terminal by these commands:
 - `cd ~/Downloads/`
 - If you downloaded it to a specific directory like “CGC-2022” on your desktop, you’d need to go to that directory path:
 - `cd ~/Desktop/CGC-2022`
- When you are in the same directory on Terminal as the downloaded file, then run the awk command below. This command should all be on one line (there is no line break between print and \$5)
 - `cat data_mutations_extended.txt | awk 'BEGIN{FS="\t"};{print $5"\t"$6"\t"$8"\t"$12"\t"$14"\t"$17}' > data_mutations_extended.tsv`
- Finally, we’re going to halve this file to make processing go quicker during the session.
 - `cat data_mutations_extended.tsv | head -n 600000 > data_mutations_half.tsv`

● (5) Download Open Cravat in advance to your own computer.

- Go here: <https://open-cravat.readthedocs.io/en/latest/quickstart.html> and follow the instructions to download. (I used the Mac Installer and it installed like butter in about 10-15 min).
- When it’s done installing, launch the Graphical User Interface by going to your Terminal in Mac (and SSH client in Windows, many use Putty). The Terminal is in Applications -> Utilities -> Terminal.app. Type “oc gui” and hit enter. It should launch a screen that looks exactly like the web version of Open Cravat.
- Click on the tab that says “Store” and install COSMIC, ESP6500 and UniProt and an output converter.
 - Click on the category Variants and choose COSMIC (it is about 2 GB), and click the button to install.
 - Unclick Variants and click the category for Allele Frequency and choose ESP6500 (it is 111 MB) and click the button to install
 - Unclick Variants and click the category for Literature and choose UniProt (it is 773KB) and click the button to install
 - Unclick Literature and click the category for Input/Output, choose TSV reporter and install (this will convert output to tab format you can open in an excel file too).

That's it! If you can create an account and download these materials in advance, you will be able to follow along with the workshop, ask questions, throw emojis in the chat (for our virtual audience) and not need to depend on the conference internet to download first. *If you are in person and do not have a computer, have a Windows computer and are unfamiliar with SSH/command line processing or are unable to download these, we have a few options! We have a smaller prepped file you can download after agreeing to terms and conditions from the Synapse platform for the AACR GENIE dataset. Please email Deb Ritter dritter@bcm.edu to get the form and link. Alternatively, we can pair you with a buddy to work with in the workshop and all materials will be available to run through, as well as the recording, after the workshop.*

Sincerely,

Drs. Deb Ritter, Alex Wagner and Rachel Karchin