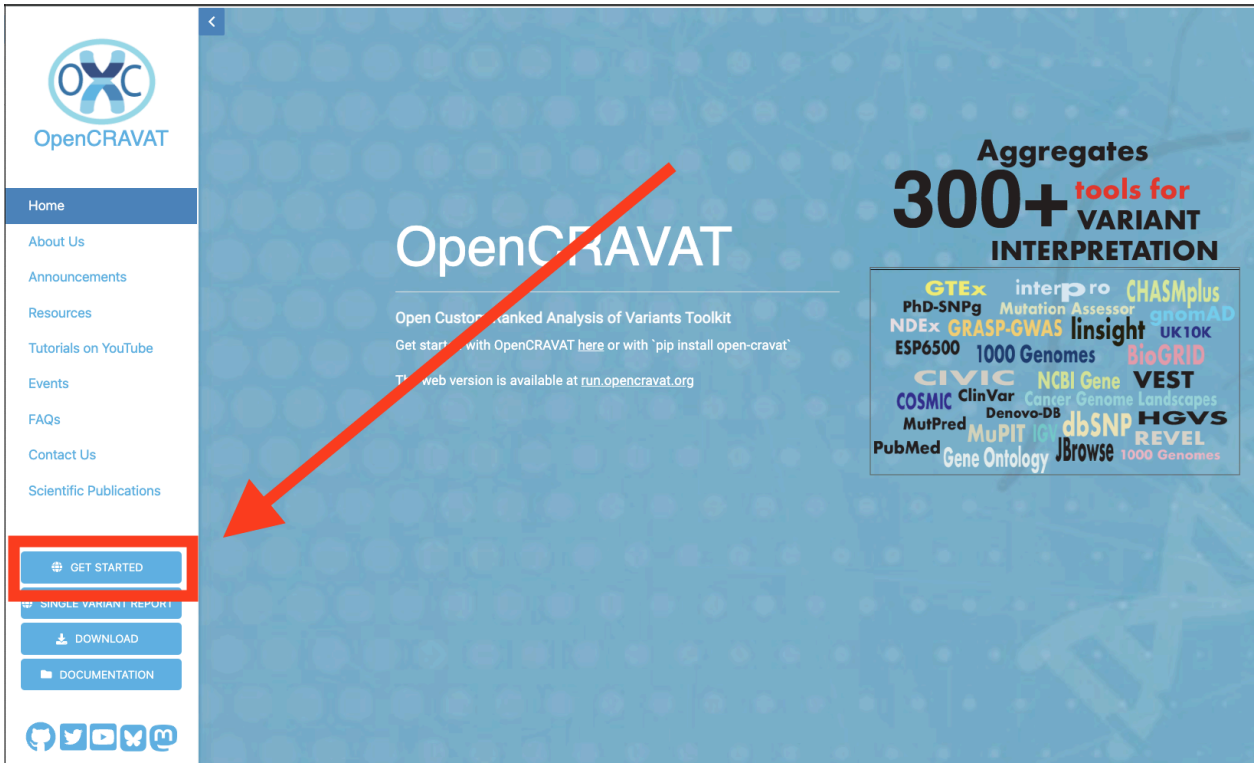
CSHL SeqTech 2024
Variant Annotation Lab (40 minutes)

In this guided lab, we will walk through using OpenCravat to look at variant annotations for several clinically-relevant annotation services.

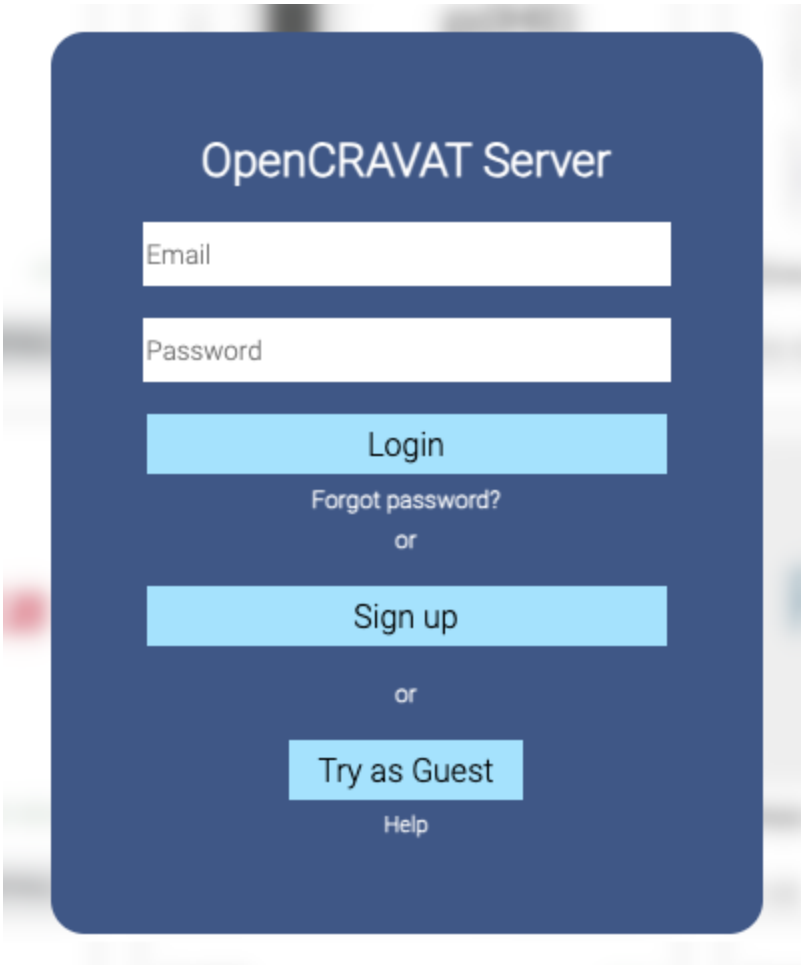
1. Navigate in your browser to opencravat.org.

2. Click on the “Get Started” button

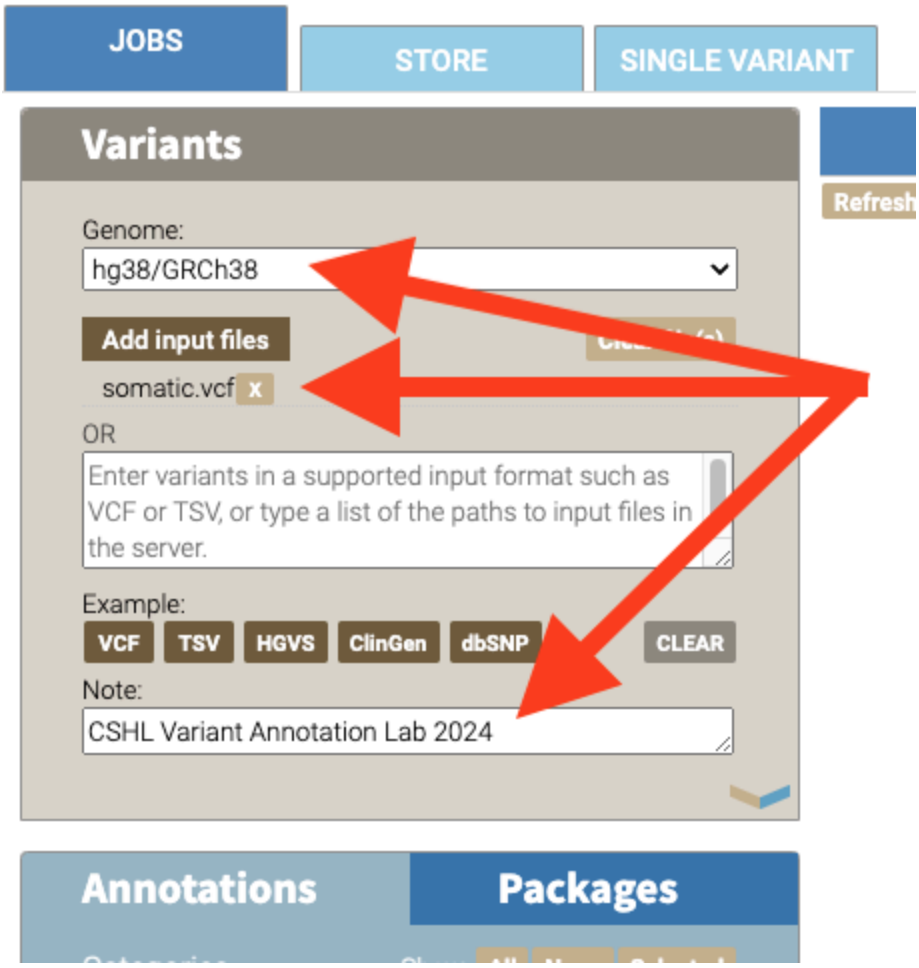


3. If you already have an account on OpenCravat, then log in. Otherwise, just click “Try as Guest”, and the system will automatically generate you a random username and password.

   Note: **Guest accounts are deleted after 7 days** so as not to clutter their servers. If you want to be able to save the records from this lab to access them longer than that period, please either make an account now, or else update the email and password of the guest account within 7 days.

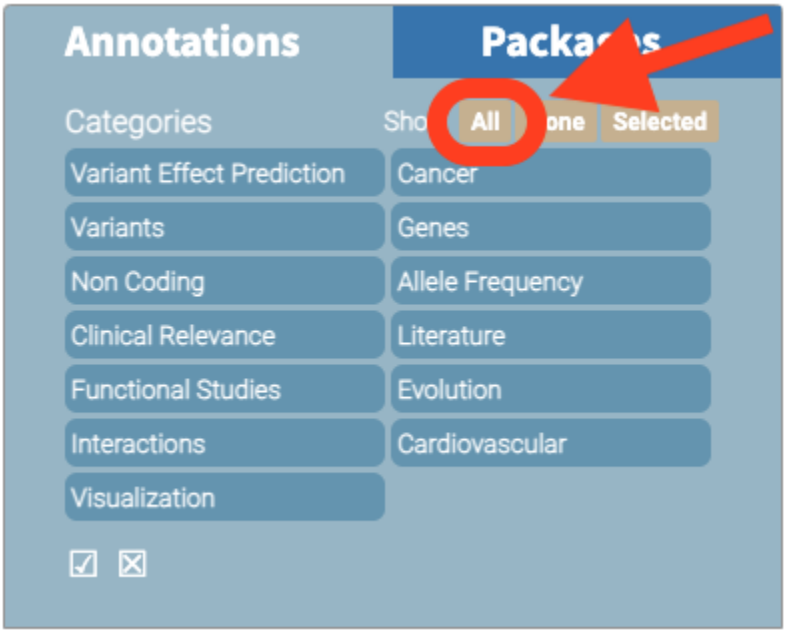


4. When you see the “Signup Successful” message, click “Ok”.
5. For this lab, we will be using the file somatic.vcf, which you can find here.
6. In the `JOBS` box, select `hg38/GRCh38` in the `Genome` dropdown menu, use the `Add input files` button to upload `somatic.vsf`, and add a descriptive note for your own future reference, as shown below.



7. We will be inspecting the variants in this file using a variety of clinically useful annotation services, but just a small subset of the services available on OpenCravat. On the Annotations tab, select Show: All, so see all the annotation services, and check the

boxes for the following annotations:



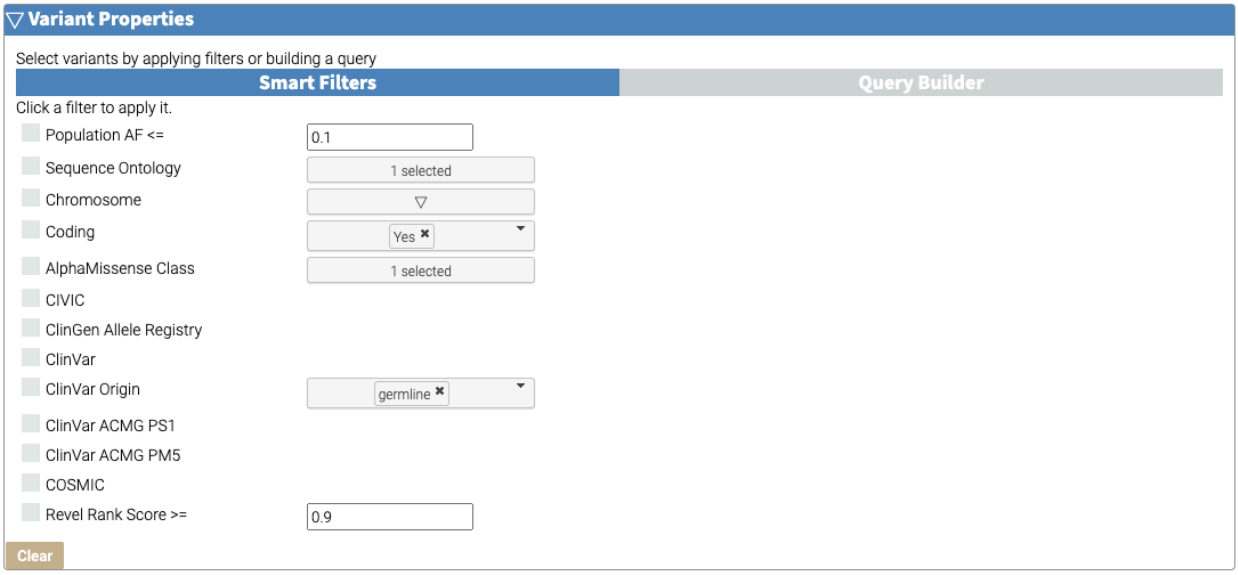
*Note: annotation services on OpenCravat are listed left-to-right, THEN Top-to-bottom for some reason.

- ☐ AlphaMissense
- ☐ CADD
- ☐ CIViC
- ☐ ClinGen Allele Registry
- ☐ ClinPred
- ☐ ClinVar
- ☐ ClinVar ACMG
- ☐ COSMIC
- ☐ COSMIC Gene
- ☐ dbSNP
- ☐ DGIdb: The Drug Interaction Database
- ☐ gnomAD
- ☐ gnomAD3
- ☐ LitVar
- ☐ OMIM
- ☐ PubMed
- ☐ REVEL
- ☐ SpliceAI
- ☐ UK10K Cohorts

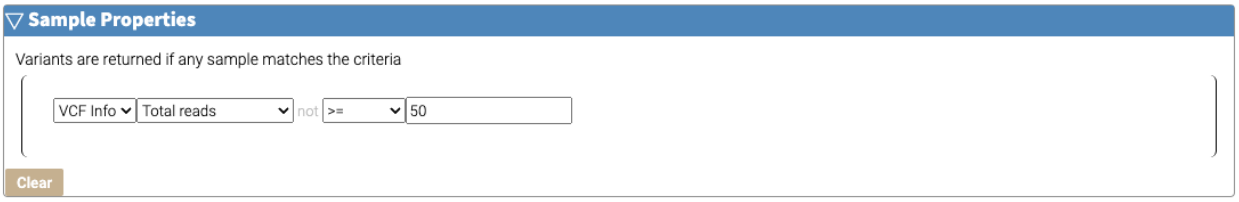
8. Then press the big Orange `ANNOTATE` button, and wait the 3-300 seconds it may take for OpenCravat to annotate your document depending on our web connection and how busy their servers are. It's ready when the value in the `Status` column is `Open Result Viewer.`

9. The `SUMMARY` tab gives us just that, a summary of this analysis, including the number of variants, annotation services called, the success status of the run, etc.

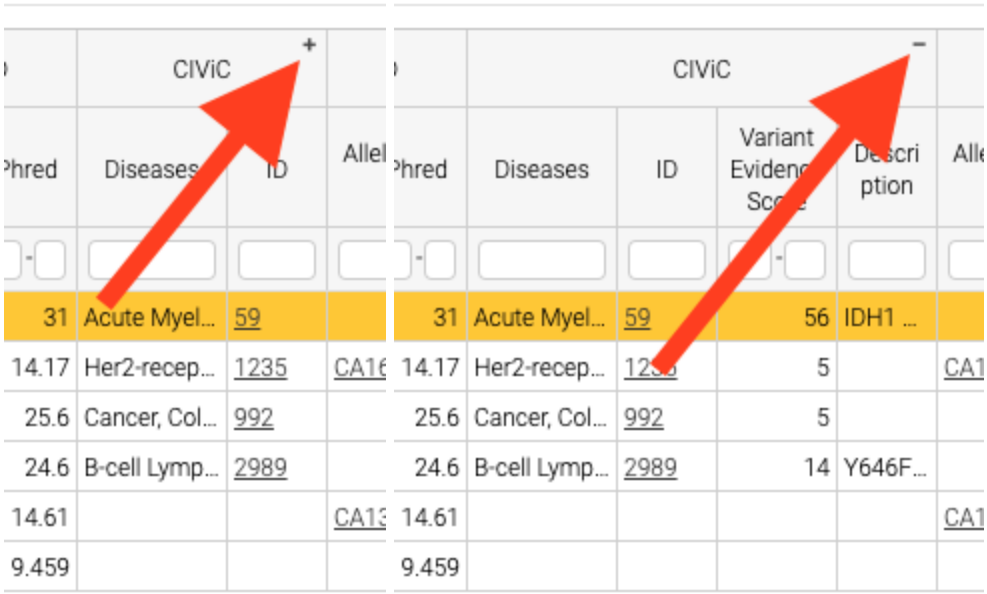
10. Let's take a look at the `FILTER` tab. We are not using these tools right now because our VCF file is very small, but they are very important when annotating a full VCF, as you can, for example, filter the variants by Population Allele Frequency (AF) to filter our variants that are very common, and therefore likely benign, or only those variants that are in protein coding regions.



You can also use sample filters to, eg, filter out any variants with less than some number of total reads.



11. Click on the VARIANT tab to explore the variants in this VCF from the perspective of our different annotation services. Many columns in the top row will have a + button to expand the information fields available from an annotation service. You can then press the − button to hide the extra fields again.



Try expanding the fields related to `AlphaMissense`, and then hiding the extra fields again.

12. You can also click on column headers in the second row to sort the variants by that field. Click once for Ascending order, twice for Descending order, and a third time to clear the sort.

13. Try sorting the variants here according to `Position` in ascending order and note the chromosome order of the variants. Next, sort by `Chrom` in ascending order, and take note of the variant order.
    - Is there anything odd about the order now?
    - Why do you think it's sorting that way?

14. One feature that OpenCravat is very useful for is that it links to the same variant across many different knowledgebases. Such links in the table are underlined. Since we selected them as annotation services to call, pick one of the variants, and then look at the same variant in (where available) `CIViC`, the `ClinGen Allele Registry`, and `COSMIC`. Note that you may have to expand columns to find the linked field.

15. Take a look at the `AlphaMissense` columns, and hover your mouse over `Score`. You can often hover your mouse over a cell value to be shown a description of that column.
    - What do higher and lower `AlphaMissense Score` values mean?

- Is this value (and the corresponding `Class`) important for annotating the variants in the current VCF? Why or why not?

16. Pick one of the variants listed, and trace its annotations across the spectrum of resources we've used enlisted for our analysis. Make note of 3 new things you've learned about the variant in question from different resources.

17. If you have time, go back to the `Submit` page on OpenCravat, and input the VCF file you've been using during this course, and re-run the annotation analysis. This should take 1-5 minutes, and try to to get through the following exercises (each starting from the whole VCF with no filters enabled):
    - How can you filter the variants to just those that are coding region frameshift variants?

    - How can you filter the vcf so as to view just the non-synonymous variants?

    - Going by AlphaMissense, How many variants are rated as likely_pathogenic, ambiguous, and likely_benign

        i. How many variants are not rated by AlphaMissense?

        ii. Looking at the variants, what variants are or are not rated by AlphaMissense?

Play around with your VCF and see if you can find any variants with interesting annotations to share!