Hi, my name is David. And I'll be taking you through our course exercise on variant calling and variant annotation. For this exercise, we shall be working with the Galaxy Europe platform because the tools that we want to use, that's DeepVariant for variant calling and SnpEff for variant annotation are already installed and ready to go within the Galaxy Europe platform.

So they encourage you to create accounts here if you don't have one already. On the right hand side, you can see that my history is empty. And this is because we need to load the datasets that we are going to use. These will be provided on the course platform. And they include a test BAM file from a SARS-CoV-2 sample and also the reference genome of SARS-CoV-2.

And you also need to load the indexes as well. So I'll go ahead and load those. And just say Start. So once this is finished, I just hit Close. And just hang on a minute until these turn to green because that's when it says that the data is ready to use.

OK, once the data is fully loaded, we can go ahead and search for DeepVariant here among the tools. And then just search for it down here. It takes a while to find, so need some patience there. But I've already found it here, DeepVariant. And here, it is asking for the reference genome.

So here, instead of built-in genome, we shall choose the genome from history, which is the SARS-CoV-2 reference. And then it is asking for a BAM file. And that will be our test BAM file.

And then it might also ask for a few options but I think we are good to go with this one. And so you can go ahead and execute.

Depending on how fast your connection is, the DeepVariant run is expected to take about 5 to 10 minutes. And now it's done on my end.

So the next thing that we are going to do, if you want, you can explore some of these outputs. For example, this is the HTML report just showing a summary of what variants they found.

So here, insertion and deletion means small indels. And then these are the SNPs. So this is what the majority of the variants were SNPs. Then you can actually also see the VCF file. So this is how it would look, as we talked about in the course.

But this one also has these genotypes. Then if you want to kind of download this file, you just expand it here and then say Download from down here. So the next thing that we are going to do is try to annotate these variants because having them in this VCF format is good. But we really need to know more information about them.

So as you can see, in the Info column here, there is a dot, which means, OK, they don't-- we are really lacking additional information that would help us to interpret them in the context of SARS-CoV-2. So we're going to go ahead and use-- here you can search for SNP there. And we are going to go ahead and use the SnpEff tool for annotating SARS-CoV-2 variants.

So there is-- yeah, so there is this SnpEff for annotating variants of SARS-CoV-2 instead of this general one. So once you select that, all we need to do is to give it the VCF that has just generated, which is already in our history. And it already has the SARS-CoV-2 genome loaded already.

Among the options for now, maybe we can select only use canonical transcripts. And you can try out others as you go along. And then afterwards, we just say Execute.

OK, so our SnpEff run is complete. And here, you can also see, it also gives an HTML report. But you can go ahead and see how the data looks like in the VCF format. So straight away, you can see that there are way more header lines, which are describing annotations that have been added by SnpEff.

And in the Info column, this here, you can see that before we had only a dot. But now we have a lot more information to go on. For example, it's telling us that this second variant is actually a missense variant. And it's giving us the protein change over here on the right hand side.

So if you're doing surveillance, you could bind this into other sites like NextStrain or GISAID and see if this variant has been found in other strains of SARS-CoV-2 that have been reported so far. And I think this is really super helpful among the steps of analysing your SARS-CoV-2 data. So I hope this was super helpful. And I hope that you followed along fine on your side.

And you can definitely play around with these files and get to know more about variant calling and variant annotation. There are way more tools out there other than DeepVariant and SnpEff for doing this kind of work. But for this exercise, these are the ones that we are just testing out.

But feel free to try out other tools, either here on Galaxy or the ones that require installation on your computer. So cheers and good luck with the rest of your research.