Exercises

Always copy your commands into your protocol file

1 FASTQC

1. QC all „4students“ data sets. Attention, the file structure is NOT the same in every file
2. Briefly write a few sentences to interpret the findings and were required copy/paste figures/screenshots of the key QC plots .

2 Annotation

1. Annotate the file atgl.vcf using VEP and hg19 as reference
2. How many variants are observed? How many are missense variants? How many are stop gained/lost mutations? how many are in splice sites?
3. Filter for mutations in the gene PNPLA2 (reference transcript only) and evaluate the findings. Play around with the data, try to filter for missense variants, damaging variants according to Polyphen, etc.

3 Nanopore Sequencing Experiment

1. **Provide the following statistics**

	Total reads	total bp	mean	median	min	max	N25	N50	N75
Barcode01									
Barcode02									
Barcode03									
Barcode04									
Barcode05									
Barcode06									

1. **Do fastqc and pauvre plots for all 5 samples and upload the html results to your results folder.**

2. **Metagenomics approaches: Which taxas are recognizes by Centrifuge in Barcode05 and 06? To which percentages?**

3. **Select a couple of <u>bacteria</u> of your choice. Find der reference genome sequence and run an alignment. How well is the genome covered?**

4. **After how many reads/bp/run time would have it been possible to correctly recognize the taxas? Using the file time stamp copy some subsets of reads into a folder and rerun Centrifuge.**