Colloquium test for DSMGT01 Modern Genomic Technologies (spring 2021)

1. Questions for the after-alignment multiQC report (Alignment_MultiQC_Report.html):

1.1. How many samples were in the library?

1.2. Was the sequencing paired-end or single-end?

1.3. What type of the library is it for?
   a.) RNA-seq
   b.) Chip-seq
   c.) whole genome DNA
   d.) whole exome DNA

1.4. Are they human samples? If so, what are the sexes of individual samples?

1.5. Which sample have the best coverage?

1.6. If the region of interest is 45326818 bp how many bp is covered by at least 60 reads in the best covered sample?

1.7. What is the approximate median distance in bp between paired reads in the best covered sample?

1.8. Why is unique percentage of **Picard: Deduplication Stats** higher (~95%) then in the **FastQC: Sequence Counts** plot (~75%)?

2. Questions for the RNA-Seq customer report (RNA-seq.customer_report.pdf):

2.1. In the Alignment and splices section (from report), you will find the place marked in red. Select one of the following options:
   a.) pretty good for all
   b.) pretty good for some
   c.) not very good for some
   d.) not very good for all

2.2. In the **Mapped regions** section (from report), you will find the places marked in red. Select one of the following options:
   a.) ~22% ; ~23%
   b.) ~60% ; ~5%
   c.) ~72% ; ~47%
   d.) ~82% ; ~73%

2.3. It is typical for QuantSeq that the peak is at the end (around 100 %) of the graph (Figure 5). Why?

2.4. In the **Read count assignment to genes** section (from report), you will find the place marked in red. Select one of the following options:
   a.) from all reads
   b.) from all mapped reads
   c.) only from uniquely mapped reads
   d.) only from unmapped reads

2.5. What is the approximate average percentage of reads that map to both human and mouse (Figure 9)?

2.6. In the **DE analysis** section (from report), you will find the place marked in red. Select one of the following options:
   a.) gene editing
   b.) gene expression
   c.) gene splicing
   d.) sequencing quality

2.7. PCA (Figure 12) plot shows nice clustering of samples with the same condition. Can this be inferred also from heatmap plot (Figure 11)? How?

2.8. Do you think we need to use batch effect removal for the analysis (Figure 12)? Why?

2.9. Calculate the relative difference of average expression (in %) between wild type and mutated gene CCDC80 in this experiment (Figure 13).

2.10. Which two samples have the most similar expression (Figure 15)?

2.11. Which gene is most significantly overexpressed (Figure 15, Figure 13)?