Main task

Your task is to replicate a current computational workflow, then apply it using new data, and report your findings, which should be fully reproducible. You are required also to read the given publications for a better understanding of the topic (see "References").

The workflow in question is "Seurat - Guided Clustering Tutorial" (satijalab.org). It is an an introductory tutorial for the R package Seurat, which is designed for QC, analysis, and exploration of single-cell RNA-seq data (scRNA-Seq).

Instructions:

1. Read the given publications.
2. Replicate the workflow on your local infrastructure or the FH server. Comment on the replication in your own words, explain the main challenges and how you addressed them.
3. Expand the study by applying the workflow to new, publicly available data. For example, you can use the data provided by the company 10x Genomics (www.10xgenomics.com). Make sure to select an appropriate dataset for this analysis (e.g. bulk RNA-Seq data are not appropriate for this). Report your findings in a reproducible way.

Details

1. Preparation. As a preparation for the data analysis workflow, thoroughly read the given manuscripts. Some guiding questions are:
   - What are the advantages of the scRNA-Seq method compared to bulk RNA-Seq?
   - What are the basic steps of the scRNA-Seq analysis?
   - What are common problems and how are they typically solved?
   - What are the major challenges in integrating single-cell transcriptomic data across different conditions, technologies, and species? How they can be solved?
   - What problem does the workflow at hand address (the Seurat vignette linked above)?
2. Replication. To replicate the tutorial, you need to reproduce all figures presented in the workflow. Address at least the following questions:
   - Is a replication of the tutorial possible? Compare the tutorial against the rules/recommendations from Sandve et al. 2013.; comment on the clarity of the description and documentation.
   - How did you set up the required environment?
   - Explain all the steps of the vignette in your own words.
3. Expanding the work. Find a publicly available data set and apply the same workflow. You may need to adapt some of the code to make it work.
   - What challenges did you face when applying the workflow to a new data set?
   - What code modifications were required?
   - Are the results comparable to the results of the original tutorial, or do they deviate in some unexpected ways?
   - Discuss all the results and interpret them.

What will be assessed:

- A written summary (ca. 2-5 pages) covering Details points 1-3. Upload via Moodle until 01.12.2021, 23:59 (Abgabefrist).
- A reproducible vignettes, scripts or notebooks that precisely documents your work covering Detail points 2 and 3. Upload via Moodle until 01.12.2021, 23:59 (Abgabefrist).
- An oral exam (ca. 15-20 mins per group member) covering Detail points 1-3. At the FH, 13.12. and 15.12.2021 at 5pm (see Notes in Teams Channel "Fallstudie" for Details)
- You will be working in a group of 2 members. Each group should select a different dataset for the task "Expending the work". Use the provided Table in Notes in Teams Channel "Fallstudie" to enter the dataset your group is analyzing. In the case of an odd number of students, one group of 3 members will be formed. This group has an additional task: Apply quality control steps from Lucken and Theis 2019 (Fig. 2).

References

- Luecken and Theis 2019. Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular Systems Biology. 15(6):e8746. doi:10.15252/msb.20188746.
- Butler et al. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 36(5):411–420. doi:10.1038/nbt.4096.
- Sandve et al. 2013. Ten Simple Rules for Reproducible Computational Research. PLOS Computational Biology. 9(10):e1003285. doi:10.1371/journal.pcbi.1003285.

Grading scheme

The grading will be based on the written report: summary and vignette (50 points, 25 points each) and the oral exam (50 points); a positive evaluation (above 50%) of each of these two parts is required for a positive examination result (KO-criterion).

- 0-50 points: grade 5 (nicht genügend)
- 51-65 points: grade 4 (genügend)
- 66-80 points: grade 3 (befriedigend)
- 81-90 points: grade 2 (gut)
- 91-100 points: grade 1 (sehr gut)