nature methods

Perspective

https://doi.org/10.1038/s41592-023-01785-3

Initial recommendations for performing, benchmarking and reporting single-cell proteomics experiments

In the format provided by the authors and unedited

Description of a single-cell proteomics experiment

This README is an example of how to describe a single-cell experiment. You can use this document as a template for your own project. The content of this document highly depends on the project and experiment. Make sure to add, remove and modify information to match your project.

Description of the project

Provide a brief description of the purpose of the study and its design.

Example:
We analyzed peripheral blood mononuclear cells (PMBC) from (n=20) treated patients and (n=15) control patients (control) to investigate the cell-type specific response of the treatment. We analyzed 1,600 single cells per patient either using a TMTpro-labeling and isobaric carrier (n = 1,500) or label-free quantification (n = 100). Treatment and control samples were randomized as shown in Table 1 of the article doi: https://doi.org/10.1101/2022.05.xx.xxxxxxx.

List of analysis files

Provide a list of the analyses files (scripts, notebooks, log files, parameter files), their location in a repo (e.g. GitLab or GitHub) with a brief description of each file. Make sure to describe the repository hierarchy.

Example:
The raw files have been processed with MaxQuant and subsequent analyses were performed in R. The files to replicate the analysis have been deposited on GitHub (or GitLab): https://github.com/YourLab/ProjectName
The repository is organized as follow:

- ↳ README # add this readme file in the repository.
- ↳ parameters/
  The folder contains the parameter and log files to replicate GUI-based software outputs
  - ↳ version.txt: a text file listing the software version and supplementary information needed to replicate the raw data processing with MaxQuant.
  - ↳ TMT/
    - ↳ parameters.txt: parameters used to run MaxQuant on the raw TMT data
  - ↳ LFQ/
    - ↳ parameters.txt: parameters used to run MaxQuant on the raw LFQ data
- ↳ analyses/
  The folder contains one or multiple notebooks (e.g. Rmd or Jupyter) to run data analysis. If multiple files are provided, use explicit names.

- ↳ read_data.Rmd: a notebook that describes how to retrieve the data and load it into R.
- ↳ read_data.pdf: a PDF document compiled after running the read_data.Rmd notebook.
- ↳ data_QC.Rmd: a notebook that performs various quality controls on the data.
- ↳ data_QC.pdf: a PDF document compiled after running the data_QC.Rmd notebook.
- ↳ data_processing.Rmd: a notebook that performs data processing on the MaxQuant tables.
- ↳ data_processing.pdf: a PDF document compiled after running the data_processing.Rmd notebook.
- ↳ downstream_analysis.Rmd: a notebook that performs downstream analysis on the processed data to answer the biological question at hand.
- ↳ downstream_analysis.pdf: a PDF document compiled after running the downstream_analysis.Rmd notebook.

Description of data files

Provide an overview of the data files (raw files and intermediate tables) as a list of data filenames and location, as different files can be located in different locations (github, PRIDE, MassIVE, ...), as well as a brief description of each file. Make sure to describe the folder hierarchy.

Example:
The raw MS data and the results were deposited in MassIVE (ID: MSV0000XXXXX). The repository is structured as follows:

- ↳ README # add this readme file in the repository.
- ↳ Sample_metadata.csv: the metadata table describing all the samples acquired in this experiment.
- ↳ raw/
  - ↳ tmt/
    The folder contains 100 .raw files as acquired by the MS instrument, and 100 mzML files converted from the raw files. Each file contains TMT data with 15 single cells per file.
  - ↳ lfq/
    The folder contains 100 .raw files as acquired by the MS instrument, and 100 mzML files converted from the raw files. Each file contains LFQ data (hence 1 single cell per file).
- ↳ search/
  - ↳ tmt/
    The folder contains the spectrum identification results for the TMT data, stored using the mzIndentML format.
  - ↳ lfq/
    The folder contains the spectrum identification results for the LFQ data, stored using the mzIndentML format.

↳ outputs/
This folder contains intermediate and processed data outputs.
- ↳ MaxQuant_output/
  This folder contains the tables exported by MaxQuant (for instance). The folder content will vary depending on the software used.
  - ↳ tables.pdf: a PDF generated by MaxQuant that describes the output tables
  - ↳ TMT/
    - evidence.txt: a table combining all the information about the identified peptides in the TMT data.
    - peptides.txt: a table containing information on the identified peptides in the TMT data.
    - proteinGroups.txt: a table containing information on the identified proteins in the TMT data. Each single row contains the group of proteins that could be reconstructed from a set of peptides.
  - ↳ LFQ/
    - evidence.txt: a table combining all the information about the identified peptides in the LFQ data.
    - peptides.txt: a table containing information on the identified peptides in the LFQ data.
    - proteinGroups.txt: a table containing information on the identified proteins in the LFQ data. Each single row contains the group of proteins that could be reconstructed from a set of peptides.
- ↳ R/
  Output data can be stored as serialized objects that can be loaded from R or python. For R, this could be the output of the scp package. For python, this could be the output of the sceptre library.
  - ↳ processed_data.rda: a serialized R object containing the processed data used for downstream analyses.
- ↳ tables/
  The folder contains output data as tables. Optionally, you could provide mzTab files.
  - ↳ peptides_processed.txt: a table with peptides x single cells at 1% FDR. The first 2 columns list the corresponding protein identifiers and peptide sequences and each subsequent column corresponds to a single cell.
  - ↳ proteins_processed.txt: a table with proteins x single cells at 1% FDR, imputed and batch corrected.
  - ↳ differential_abundance_test.csv: a table reporting the outcome of the differential abundance test. Columns contain: protein ID, log2 fold change, p-value, FDR adjusted p-value.

↳ misc/
The folder contains other files required for running the data analysis
- ↳ human.fasta: the human SwissProt database (downloaded July 30, 2018; 20,373 entries) used to perform peptide identification by MaxQuant.

- ↳ signatures.gmt: a GMT file containing selected signatures for gene set enrichment.

Description of the metadata table

Provide a description of the metadata table that provides the experimental design.

Example
An example of a sample metadata table (for 3 TMT acquisitions and 13 LFQ acquisitions) can be found here.

The sample metadata table contains the following columns:

- ID: a unique identifier associated to each acquired sample
- Sample type: one of “Carrier”, “Reference”, “Empty”, “Negative control” or “Single cell”.
- Patient ID: the ID of the patient from which the cell was isolated. The ID starts with a “C” when the patient is part of the control group and “T” when the patient is part of the treatment group. This only applies for single-cell samples.
- Condition: either “TRT” or “CTL”, indicating whether the cell belongs to a patient from the treated or control group, respectively. This only applies for single-cell samples.
- File name: the name of the file containing the MS spectral data. The name convention is date (YYYYMMDD) + time (HHMMSS) + sample preparation batch (plate 1 or 2).
- Date and Time of acquisition: the date is formatted as YYYY-MM-DD and the time is formatted as HH:MM:SS.
- Mass tag: the chemical acronym of the isobaric tag used before sample pooling. A “-” means no label was used (LFQ).
- FSC-A: the signal intensity area of the forward scatter as measured by flow cytometry.This only applies for single-cell samples.
- SSC-A: the signal intensity area of the side scatter as measured by flow cytometry. This only applies for single-cell samples.
- CD34:APC-Cy7-A: the signal intensity area of the APC-Cy7 emission fluorescence as measured by flow cytometry. The fluorophore is attached to an anti-CD34 antibody. This only applies for single-cell samples.
- Sample preparation batch: the sample preparation batch for extracting, labeling and pooling peptides.
- LC batch: the liquid chromatography batch. Each batch corresponds to an LC column.
- MS instrument name: the name of the instrument used for mass spectrometry acquisition.