

Supplementary Material

Here we outline how to run the presented pipeline on a small exemplary dataset locally on a laptop to do some first initial testing of the pipeline and get familiar with the setup. We copied most of the steps from the protocol as outlined in the manuscript but removed some steps to shorten it. Also, we added exemplary screenshots to evaluate the results that should be obtained when running the protocol on the exemplary data and discuss some of the results shortly.

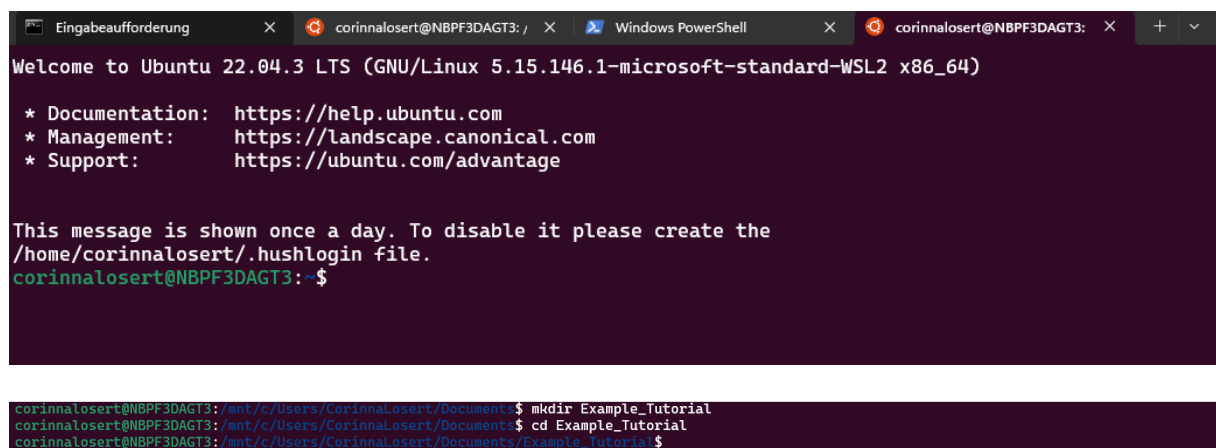
The exact configuration files used for this example can be found here:

https://github.com/heiniglab/mofa_workflow/tree/main/scripts/configurations/config_examples/Example_Data

1. Preparations: Technical Setup and Installation

NOTE: To run this program, have wget, git and Apptainer preinstalled on the device. A guide for the installation of Apptainer on different systems (Linux, Windows, Mac) is given here: <https://apptainer.org/docs/admin/main/installation.html>. Installation information on git can be found here: <https://git-scm.com/book/en/v2/Getting-Started-Installing-Git>

1.1 Open the console and choose or create a folder in which all the analysis code and outputs will be stored. Navigate to the folder typing the command: 'cd *path_to_folder*' in the terminal.



```
Eingabeaufforderung X corinnalosert@NBPF3DAGT3: / X Windows PowerShell X corinnalosert@NBPF3DAGT3: X + v
Welcome to Ubuntu 22.04.3 LTS (GNU/Linux 5.15.146.1-microsoft-standard-WSL2 x86_64)

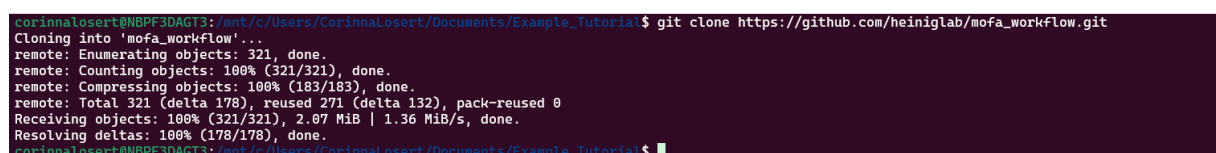
 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

This message is shown once a day. To disable it please create the
/home/corinnalosert/.hushlogin file.
corinnalosert@NBPF3DAGT3:~$

corinnalosert@NBPF3DAGT3:/mnt/c/Users/CorinnaLosert/Documents$ mkdir Example_Tutorial
corinnalosert@NBPF3DAGT3:/mnt/c/Users/CorinnaLosert/Documents$ cd Example_Tutorial
corinnalosert@NBPF3DAGT3:/mnt/c/Users/CorinnaLosert/Documents/Example_Tutorial$
```

Example: usage of an Ubuntu console on windows and navigation to newly generated folder for executing analysis

1.2 Download or clone the code repository from Github (https://github.com/heiniglab/mofa_workflow) or by typing 'git clone https://github.com/heiniglab/mofa_workflow.git' in the terminal window.



```
corinnalosert@NBPF3DAGT3:/mnt/c/Users/CorinnaLosert/Documents/Example_Tutorial$ git clone https://github.com/heiniglab/mofa_workflow.git
Cloning into 'mofa_workflow'...
remote: Enumerating objects: 321, done.
remote: Counting objects: 100% (321/321), done.
remote: Compressing objects: 100% (183/183), done.
remote: Total 321 (delta 178), reused 271 (delta 132), pack-reused 0
Receiving objects: 100% (321/321), 2.07 MiB | 1.36 MiB/s, done.
Resolving deltas: 100% (178/178), done.
corinnalosert@NBPF3DAGT3:/mnt/c/Users/CorinnaLosert/Documents/Example_Tutorial$
```

Example: console output when cloning repository

1.3 Download the image that contains all the required installations from zenodo by typing: 'wget https://zenodo.org/records/11192947/files/mofa_image.sif' in the terminal window.

```
corinnalosert@NBPF3DAGT3:/mnt/c/Users/Corinnalosert/Documents/Example_Tutorial$ wget https://zenodo.org/records/11192947/files/mofa_image.sif
--2024-05-16 16:03:30-- https://zenodo.org/records/11192947/files/mofa_image.sif
Resolving zenodo.org (zenodo.org)... 188.185.79.172, 188.184.103.159, 188.184.98.238, ...
Connecting to zenodo.org (zenodo.org)|188.185.79.172|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3220754432 (3.0G) [application/octet-stream]
Saving to: 'mofa_image.sif'

mofa_image.sif          7%[====>                ] 240.87M  1.07MB/s   eta 45m 31s
```

Example: output in console while downloading image and example data from zenodo.

1.4 Generate a folder in which all the result data will be stored by typing 'mkdir results' in the terminal window.

1.5 Generate a folder in which all the input data to be used in the analysis will be added by typing 'mkdir input_data' in the terminal window.

```
corinnalosert@NBPF3DAGT3:/mnt/c/Users/Corinnalosert/Documents/Example_Tutorial$ mkdir results
corinnalosert@NBPF3DAGT3:/mnt/c/Users/Corinnalosert/Documents/Example_Tutorial$ mkdir input_data
```

Example: generate 'input_data' and 'results' folder

1.6 (Step adjusted from Protocol to download and add example data): Navigate to the input_data folder by typing the command 'cd input_data' in the command line.

Download the example 'h5ad' data from zenodo by typing in the console: 'wget https://zenodo.org/records/11192947/files/Prepared_Example_Data.h5ad'

Download the example example 'Prepared_Sample_Meta_Data.csv' by typing in the console: 'wget https://zenodo.org/records/11192947/files/Prepared_Sample_Meta_Data.csv'

Download the example 'Prepared_Pathway_Data.csv' by typing in the console: 'wget https://zenodo.org/records/11192947/files/Prepared_Pathway_Data.csv'.

Navigate back to the folder containing the container image typing the command: 'cd ..' in the terminal

```
corinnalosert@NBPF3DAGT3:/mnt/c/Users/Corinnalosert/Documents/Example_Tutorial/input_data$ wget https://zenodo.org/records/11192947/files/Prepared_Example_Data.h5ad
--2024-05-16 16:16:10-- https://zenodo.org/records/11192947/files/Prepared_Example_Data.h5ad
Resolving zenodo.org (zenodo.org)... 188.184.103.159, 188.185.79.172, 188.184.98.238, ...
Connecting to zenodo.org (zenodo.org)|188.184.103.159|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2385351075 (2.2G) [application/octet-stream]
Saving to: 'Prepared_Example_Data.h5ad'

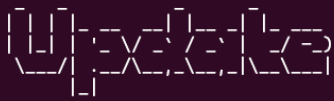
Prepared_Example_Data.h5ad  0%[                      ]  5.07M  762KB/s   eta 52m 7s
```

Example: downloading data from zenodo in the 'input_data' folder

1.7 (Step adjusted from Protocol to run on local machine instead of cluster): Execute the container that will start a JupyterLab session by typing the following command in the terminal:

‘apptainer exec mofa_image.sif jupyter-lab’. Copy the URL that is returned by the command to the browser which will open up a jupyter-lab session (more information on jupyter-lab can be found within the [software documentation](#)¹⁶).

```
corinnalosert@NBPF3DAGT3:/mnt/c/Users/CorinnaLosert/Documents/Example_Tutorial$ apptainer exec mofa_image.sif jupyter-lab
WARNING: Insecure writes have been enabled via environment variable 'JUPYTER_ALLOW_INSECURE_WRITES'! If this is not intended, remove the variable or set its value to 'False'.
[W 2024-05-16 16:22:08.610 ServerApp] A '_jupyter_server_extension_points' function was not found in jupyter_lsp. Instead, a '_jupyter_server_extension_paths' function was found and will be used for now. This function name will be deprecated in future releases of Jupyter Server.
[I 2024-05-16 16:22:08.744 ServerApp] Extension package jupyter_server_ydoc took 0.1135s to import
[W 2024-05-16 16:22:08.748 ServerApp] A '_jupyter_server_extension_points' function was not found in nbclassic. Instead, a '_jupyter_server_extension_paths' function was found and will be used for now. This function name will be deprecated in future releases of Jupyter Server.
[W 2024-05-16 16:22:08.749 ServerApp] A '_jupyter_server_extension_points' function was not found in notebook_shim. Instead, a '_jupyter_server_extension_paths' function was found and will be used for now. This function name will be deprecated in future releases of Jupyter Server.
[I 2024-05-16 16:22:08.750 ServerApp] jupyter_lsp | extension was successfully linked.
[I 2024-05-16 16:22:08.753 ServerApp] jupyter_server_fileid | extension was successfully linked.
[I 2024-05-16 16:22:08.756 ServerApp] jupyter_server_terminals | extension was successfully linked.
[I 2024-05-16 16:22:08.758 ServerApp] jupyter_server_ydoc | extension was successfully linked.
[I 2024-05-16 16:22:08.761 ServerApp] jupyterlab | extension was successfully linked.
[I 2024-05-16 16:22:08.764 ServerApp] nbclassic | extension was successfully linked.
[I 2024-05-16 16:22:09.227 ServerApp] notebook_shim | extension was successfully linked.
[I 2024-05-16 16:22:09.408 ServerApp] notebook_shim | extension was successfully loaded.
[I 2024-05-16 16:22:09.410 ServerApp] jupyter_lsp | extension was successfully loaded.
```



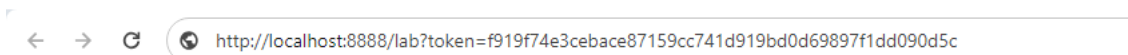
Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions.

https://jupyter-notebook.readthedocs.io/en/latest/migrate_to_notebook7.html

Please note that updating to Notebook 7 might break some of your extensions.

```
[I 2024-05-16 16:22:09.442 ServerApp] nbclassic | extension was successfully loaded.
[I 2024-05-16 16:22:09.443 ServerApp] Serving notebooks from local directory: /mnt/c/Users/CorinnaLosert/Documents/Example_Tutorial
[I 2024-05-16 16:22:09.443 ServerApp] Jupyter Server 2.12.1 is running at:
[I 2024-05-16 16:22:09.443 ServerApp] http://localhost:8888/lab?token=f919f74e3cebase87159cc741d919bd0d69897f1dd090d5c
[I 2024-05-16 16:22:09.443 ServerApp] http://127.0.0.1:8888/lab?token=f919f74e3cebase87159cc741d919bd0d69897f1dd090d5c
[I 2024-05-16 16:22:09.443 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2024-05-16 16:22:09.519 ServerApp]
```

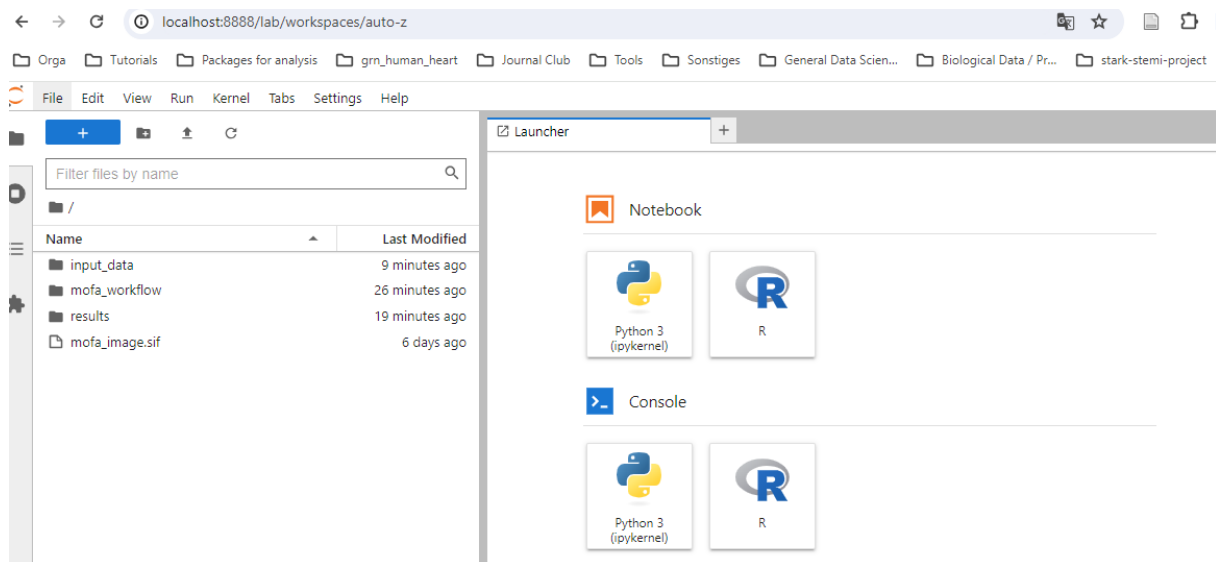
Example: output of terminal when starting the jupyter-lab session



Example: copying the URL from above and entering it in the browser

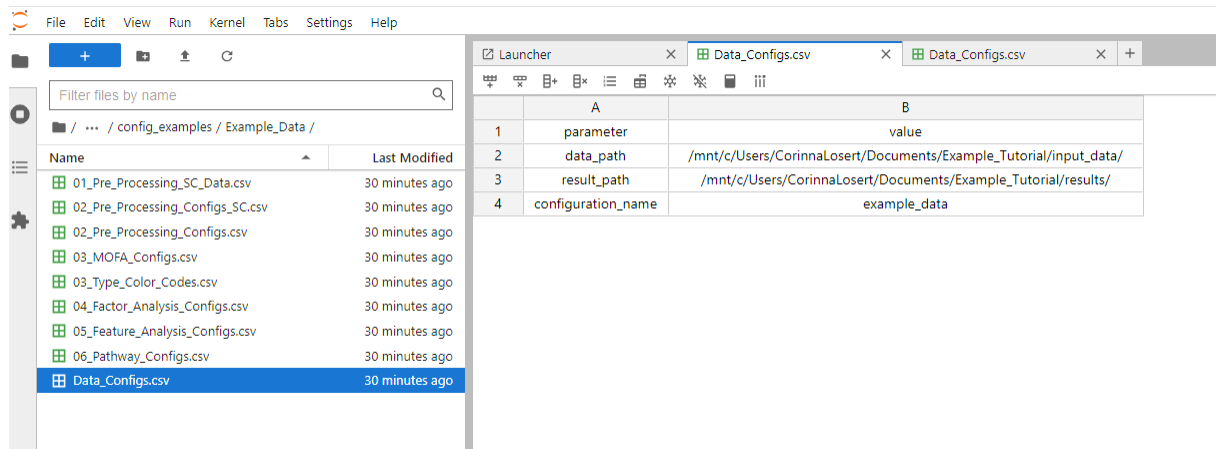
2. Initialization

2.1. In the Jupyter-Lab Session use the navigation menu on the left side. Navigate to the ‘configurations’ folder by double-clicking on the folders: ‘mofa_workflow’, ‘scripts’, ‘configurations’. Within the folder open the file: ‘Data_configs.csv’ by double-clicking on it.



Example: jupyter-lab interface. Use navigation menu on the left for navigating into the 'mofa_workflow' folder

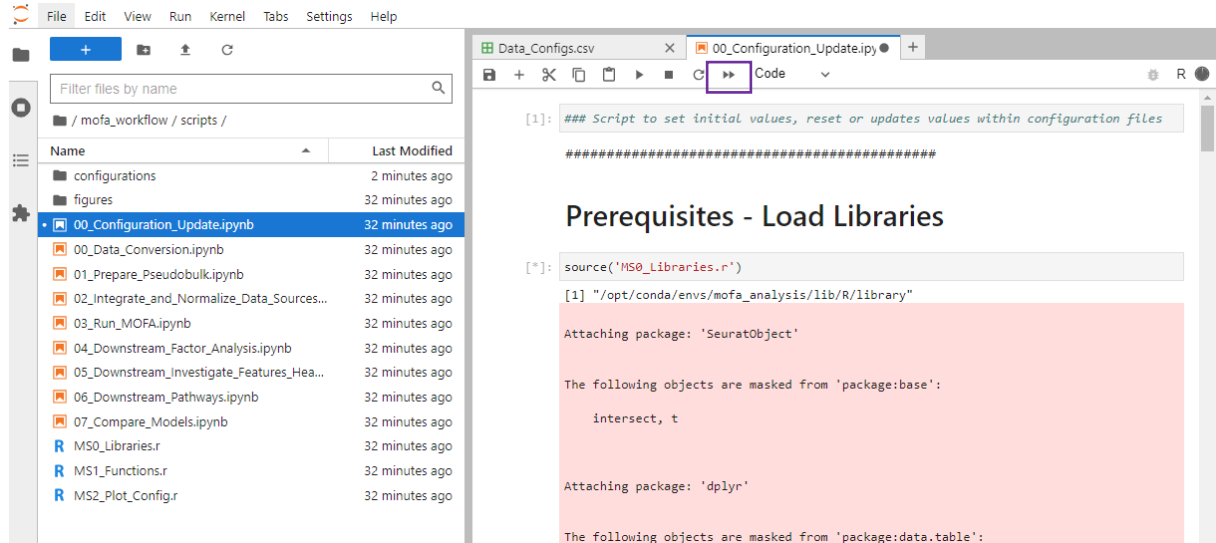
2.2. In the 'value' column add the paths to the folders of the 'input_data' (data_path) and 'results' (result_path) folder. Add a name that will be added as file extension to all saved files in the 'value' column for the 'configuration_name')). Save the changes by clicking on 'File' > 'Save CSV File' in the menu on the top.



Example: Data_Config.csv file with added paths to files and chosen name for example data configuration

2.3. Use the navigation menu on the left side to navigate to the scripts folder by clicking on 'scripts'. Open the initialization notebook by double-clicking on '00_Configuration_Update.ipynb'. Execute the script by clicking on the 'Restart kernel and run

all cells' button in the top and clicking 'Restart' in the pop-up.



Example: running jupyter notebook after clicking on 'Execution' button in the top

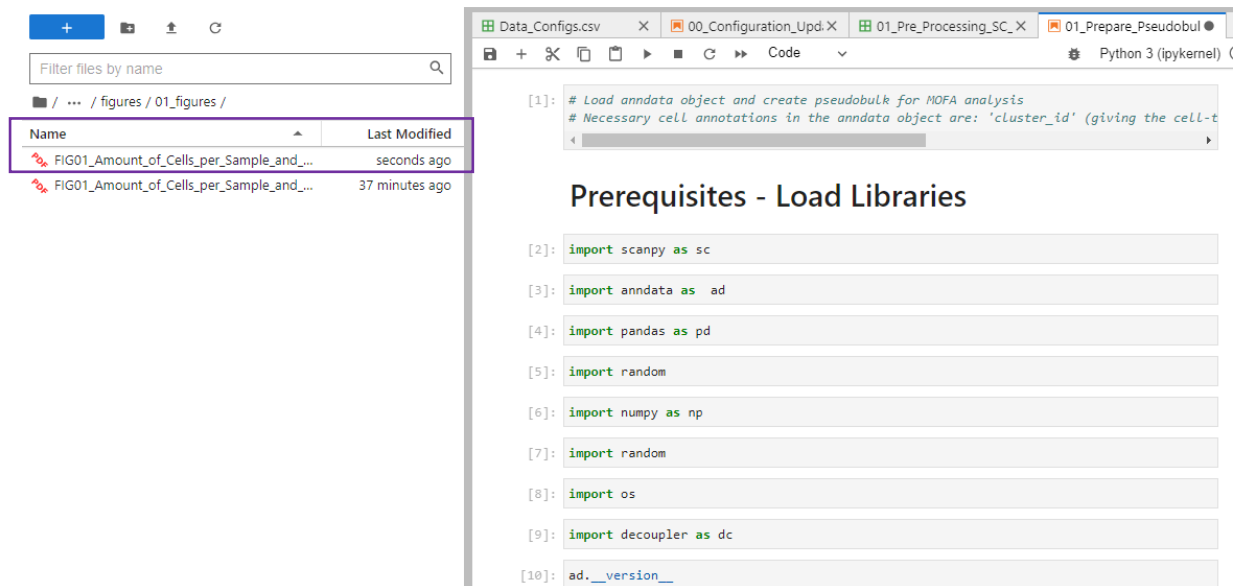
3. Data Pre-Processing and Harmonization

3.1. Pre-Processing – Convert sc Data to Pseudobulk

3.1.1. Open the notebook '01_Prepare_Pseudobulk.ipynb' by double-clicking on it. Execute the script by clicking on the 'Restart kernel and run all cells' button in the top and clicking 'Restart' in the pop-up.

3.1.2. Use the navigation menu on the left side to navigate to the figures folder by double-clicking first on 'figures' and then on '01_figures'. Open the newly generated plot 'FIG01_Amount_of_Cells_Overview' by double clicking on it.

NOTE: In case of a successful execution of the notebook the file 'FIG01_Amount_of_Cells_Overview' will either be updated by the notebook or generated newly. The 'Last Modified' column can indicate when the file was generated to evaluate whether it is a new or old file.



Example: added figure after execution of script within '01_figures' folder highlighted in purple.

3.1.3. Investigate the plot to identify cell-type clusters with very low number of cells per sample.



Example: resulting plot on the example data. Shows that for most cell-types we have a substantial number of cells per sample. Only 'adipocyte' and 'mast cells' seem to contain samples for which the number of measured cells is rather low.

3.1.4. Use the navigation menu on the left side to navigate back to the configurations folder by clicking on '...' and then double clicking on 'configurations'. Open the file: '02_Preprocessing_Configs_SC.csv' by double-clicking on it.

3.1.5. Adjust the values in the 'cell_expr_thres1' column to '30;5' and in the 'cell_expr_thres2' column to '20;10'.

	A	B	C	D	E	F
1	configuration_name	data_name	data_type	cell_expr_thres1	cell_expr_thres2	cell_type_exclusion
2	example_config	Prepared_Example_Data	h5seurat	30:5	20:10	cell_type_exclusion

Example: values of the '02_Preprocessing_Configs' set for this example run

3.1.6. Save the changes by clicking on 'File' and then 'Save CSV File' in the navigation on the top.

3.2. Pre-Processing – Harmonize and Integrate other Omics Data Sources

3.2.1. Open the file '02_Preprocessing_Configs.csv' by double clicking on it and adjust the pre-processing configuration for each of the datasets that will be included and is stored in the 'data_input' folder (one row per dataset).

	A	B	C	D	E	F
1	configuration_name	data_name	file_type	data_type	remove_sample_ids	sample_filtering_t
2	example_data	Prepared_Example_Data	.h5ad			1
3	example_data	Prepared_Pathway_Data	.csv			1
4	example_data	Prepared_Sample_Meta_Data	.csv			1
5						

Example: config file setup for the example data as generated automatically. We need to remove the rows for 'Prepared_Pathway_Data' and 'Prepared_Sample_Meta_Data' from the file. As we already downloaded them to the input folder the initialization script assumes that they also should be used for the MOFA model, but they shouldn't.

3.2.2. Adjust the other parameters in the columns accordingly depending on which pre-processing steps should be applied.

NOTE: the file as we use it for the example data with all configurations can be found when navigating to 'config_examples/Example_Data'. We copy the configurations from here and add them in the file that we opened up above.

	A	B	C	D	E	F
1	configuration_name	data_name	file_type	data_type	remove_sample_ids	sample_filtering_th
2	example_config	Prepared_Example_Data	h5seurat	sc		1

Example: configuration '02_Pre_Processing_Configs.csv' after adjusting the configurations for this example

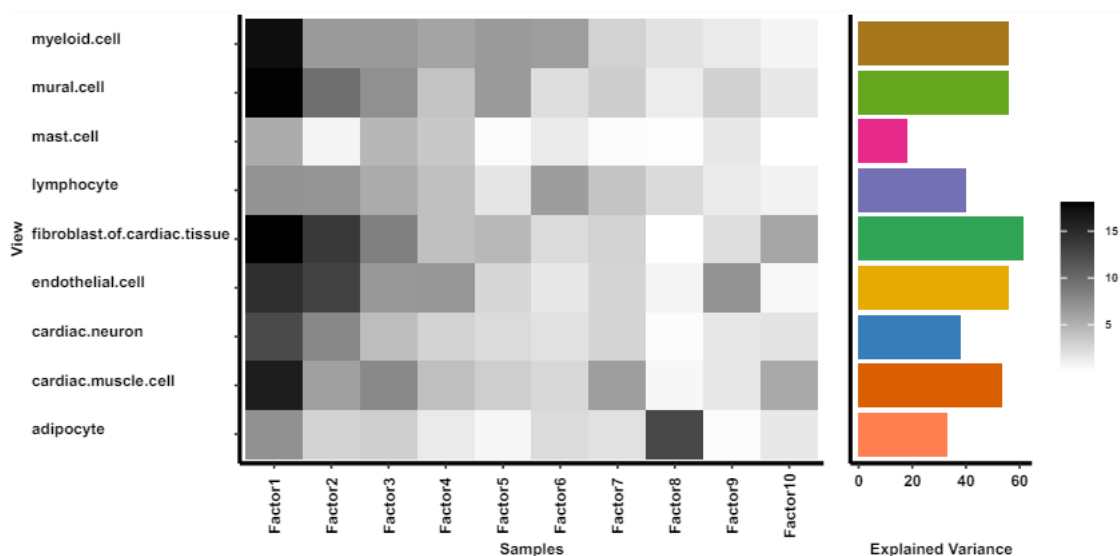
3.2.3. Save the changes by clicking on 'File' > 'Save CSV File'

3.2.4. Use the navigation menu on the left side to navigate to the 'scripts' folder by clicking on 'scripts'. Open the notebook '02_Integrate_and_Normalized_Data_Sources.ipynb' by double-clicking on it. Execute the script by clicking on the 'Restart kernel and run all cells' button in the top and clicking 'Restart' in the pop-up.

4. Run MOFA

4.1. Use the navigation menu on the left side to navigate to the 'scripts' folder by clicking on 'scripts'. Open the notebook '03_Run_MOFA.ipynb' by double-clicking on the file. Execute the script by clicking on the 'Restart kernel and run all cells' button in the top.

4.2. Navigate to the '/figures/03_figures' folder by double-clicking on 'figures' and then '03_figures'. Open the generated plot 'FIG03_Overview_Variance_Decomposition_'mofa_result_name' and investigate the model result.



Example: variance decomposition output plot for the example data. We can observe multiple factors that capture shared variance not only of one cell type but multiple cell-types, like for example Factor1. We can also see that for example Factor8 seems to be capture variance that is mainly inherent to adipocytes.

5. Downstream Analysis

5.1. Factor Interpretation

5.1.1. Prepare a '.csv' file ('Prepared_Sample_Meta_Data.csv') that contains all the meta-data (covariates) of the samples that will be analyzed in association with the generated factors. Copy the file to the 'input_data' folder by using 'Drag&Drop' dropping the file in the folder overview of the 'input_data' folder.

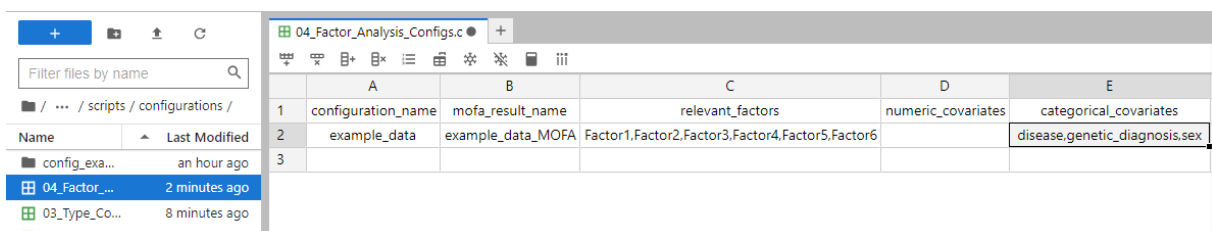
[We have already downloaded this file to the folder from zenodo in the beginning of this protocol for the example data, so this step is not necessary. The example file can be inspected in the input_data folder. We can navigate to this folder clicking on the ‘folder’ symbol in the left folder by double-clicking on ‘input_data’]

disease	sex	development_stage	self_reported_ethnicity	genetic_diagnosis	sample_id
dilated ca	female	fifth decade human stage	European	PLN	DP2
normal	male	sixth decade human stage	Asian	control	H3
dilated ca	male	seventh decade human stage	European	TTN	DT4
normal	female	sixth decade human stage	European	control	H5
dilated ca	male	eighth decade human stage	European	TNNT2	DO1

Example: of the structure of the downloaded ‘Sample_Meta_Data.csv’ File

5.1.2. In Jupyter-Lab use the navigation menu on the left to navigate back to the configurations folder by clicking on ‘...’ and then double-clicking on ‘configurations’. Open the file: ‘04_Factor_Analysis.csv’ by double-clicking on it.

5.1.3. In the ‘categorical_covariates’ column add the name of all numeric columns in the ‘Prepared_Sample_Meta_Data.csv’ file that will be investigated in relation to the MOFA factors separated by comma (we choose ‘disease,sex,genetic_diagnosis’ here).



	A	B	C	D	E
1	configuration_name	mofa_result_name	relevant_factors	numeric_covariates	categorical_covariates
2	example_data	example_data_MOFA	Factor1,Factor2,Factor3,Factor4,Factor5,Factor6		disease,genetic_diagnosis,sex
3					

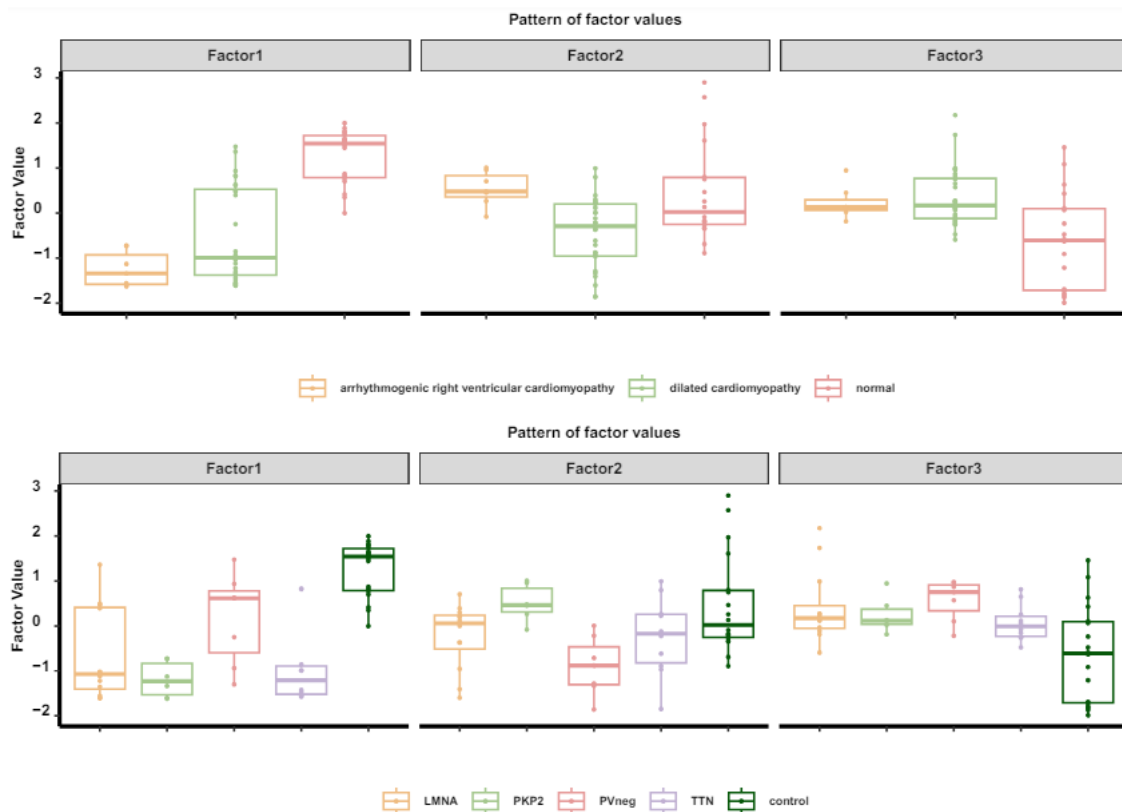
Example: added covariates to the configuration file

5.1.4. In the ‘numeric_covariates’ column add the name of all categorical columns in the ‘Prepared_Sample_Meta_Data.csv’ file that will be investigated in relation to the MOFA factors separated by comma (for the example data we keep this column empty as there are no numeric covariates available).

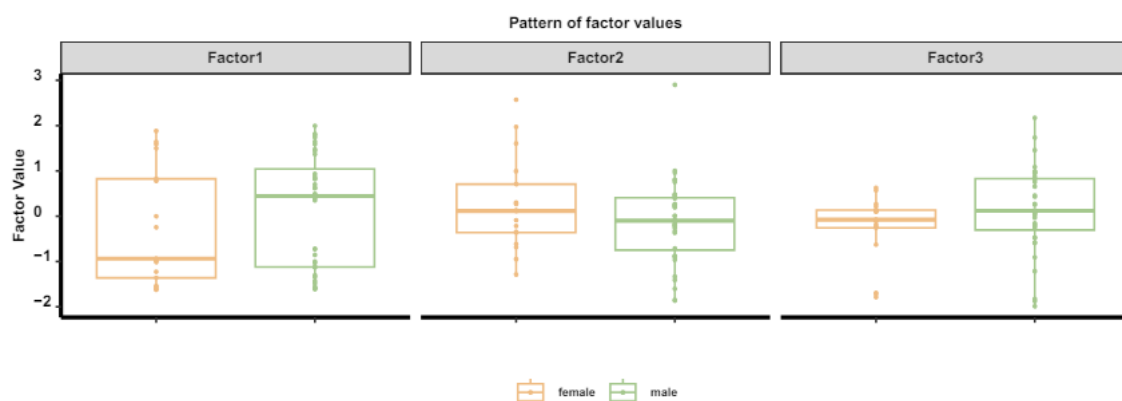
5.1.5. Save the changes by clicking on ‘File’ > ‘Save CSV File’

5.1.6. Use the navigation menu on the left to navigate to the ‘scripts’ folder by clicking on ‘scripts’. Open the notebook ‘04_Downstream_Factor_Analysis.ipynb’ by double-clicking on it. Execute the script by clicking on the ‘Restart kernel and run all cells’ button in the top and then on ‘Restart’ in the pop-up.

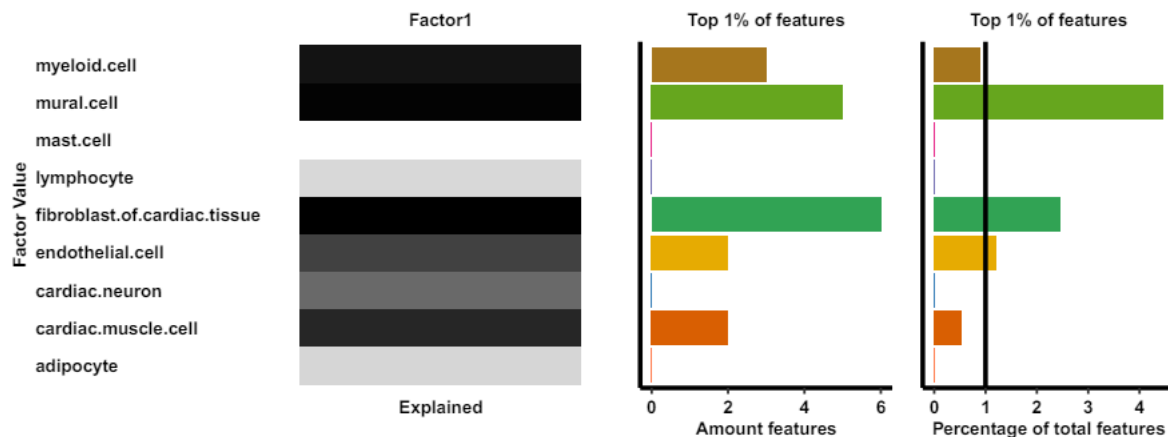
5.1.7. Use the navigation menu on the left to navigate to the ‘/figures/04_figures’ folder by double-clicking on ‘figures’ and then ‘04_figures’. Open the generated plot by double clicking on them and investigate the factors for interesting patterns and associations:
‘FIG04_Factor_Association_with_numeric_features_‘mofa_result_name’.pdf.
‘FIG04_Factor_Association_with_categorical_features_‘mofa_result_name’.pdf.
‘FIG04_Top_Feature_Overview_per_Factor_‘mofa_result_name’.pdf.



Example: FIG04_Factor_Association_with_categorical_features: disease and genetic_diagnosis. For the given dataset where we have included patients that suffer from different types of heart diseases (arrhythmogenic right ventricular cardiomyopathy, dilated cardiomyopathy and normal – which corresponds to healthy patients) and have different types of genetic diagnosis ('LMNA', 'PKP2', 'PVneg', 'TTN', 'control') we can observe that Factor1 clearly distinguishes the different groups and especially the healthy from the disease patients in the upper plot. Based on that one could conclude that there are large differences in the measured gene expression profiles across multiple genes and cell-types between healthy and diseases patients that are aggregated by 'Factor1'.



Example: FIG04_Factor_Association_with_categorical_features: sex. As a contrary example to the disease state: when investigating the factor values divided by the sex of the patient, we can observe that there are no differences between female and male patients.



Example: FIG04_Top_Feature_Overview_per_Factor. With this plot we can investigate which features from the different dimensions are most relevant for defining the pattern that we can observe in Factor1. The plot shows for which of the different input views there are most highly ranking features (barplot left) also in relation to the total amount of features of a view (barplot right). Here we can observe that for example mural cells and fibroblasts of cardiac tissues seem to make up a high amount of the highest-ranking features on the factor.

NOTE: It makes sense to also investigate for example the top 5% or 10% of highest ranking features especially if there is a low amount of features as in this example.

5.2. Feature Analysis

5.2.1. Use the navigation menu on the left to navigate back to the configurations folder by clicking on '...' and then double clicking on 'configurations'. Open the file: '05_Feature_Analysis_Configs.csv' by double-clicking on it.

5.2.2. In the column 'faceting_variable' add a column name of a categorical column in the 'Prepared_Sample_Meta_Data.csv' that will be used to group the samples in the plot (we use: 'genetic_diagnosis' here)

	A	B	C	D	E	F	G	H
1	configuration_name	mofa_result_name	factor	top_variable_thres	faceting_variable	type	plot_width	plot_h
2	example_data	example_data_MOFA	Factor1	0.005	genetic_diagnosis		8.07	5

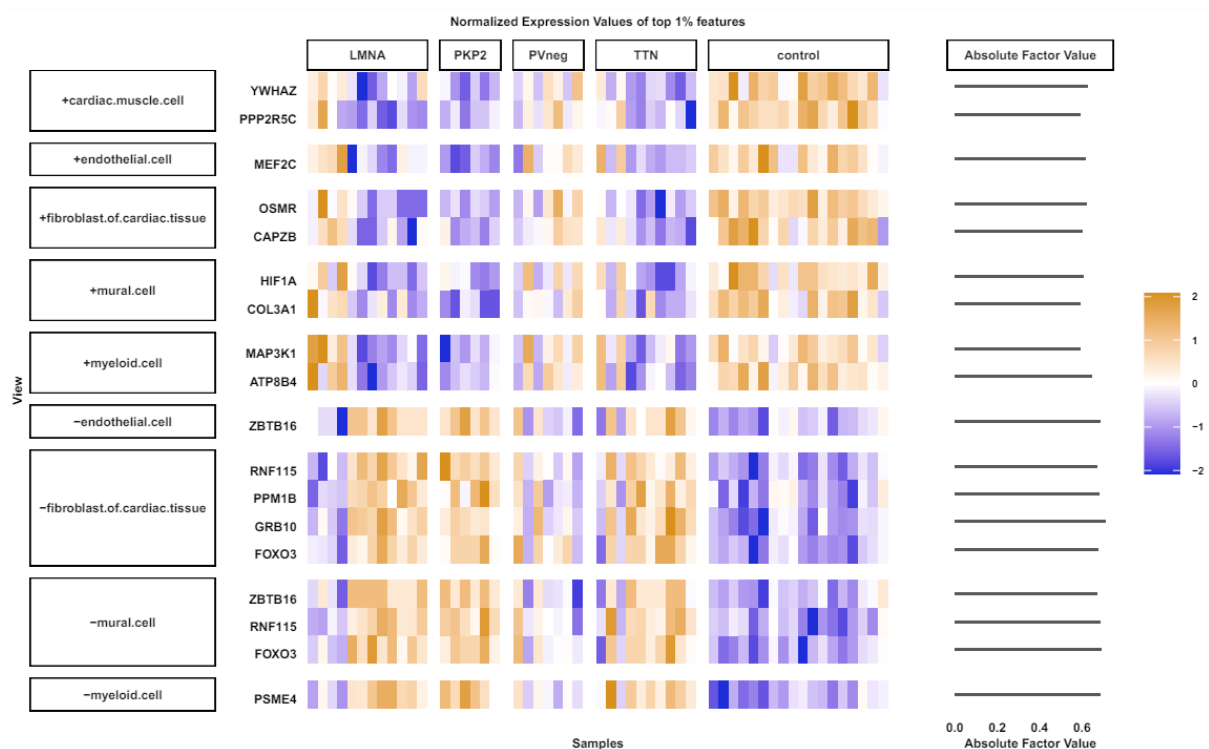
Example: configuration file 05_Feature_Analysis_Configs for example data

5.2.3. Save the changes by clicking on 'File' > 'Save CSV File'

5.2.4. Use the navigation menu on the left to navigate to the 'scripts' folder by clicking on scripts. Open the notebook '05_Downstream_Investigate_Features_Heatmap.ipynb' by double-clicking on it. Execute the script by clicking on the 'Restart kernel and run all cells' button in the top and 'Restart' in the pop-up.

5.2.5. Use the navigation menu on the left to navigate to the '/figures/05_figures' folder by double-clicking first on 'figures' and then on '05_figures'. Open and investigate the generated plot 'FIG05_Heatmap_Feature_Overview__mofa_result_name'.pdf by double clicking on the

file.



Example: Heatmap showing the top-ranking genes defining Factor1 estimated on the example data. We can observe that similarly to factor we find clear differences in the expression values between the patients of different genetic diagnosis on these genes.

5.3. Pathway Analysis

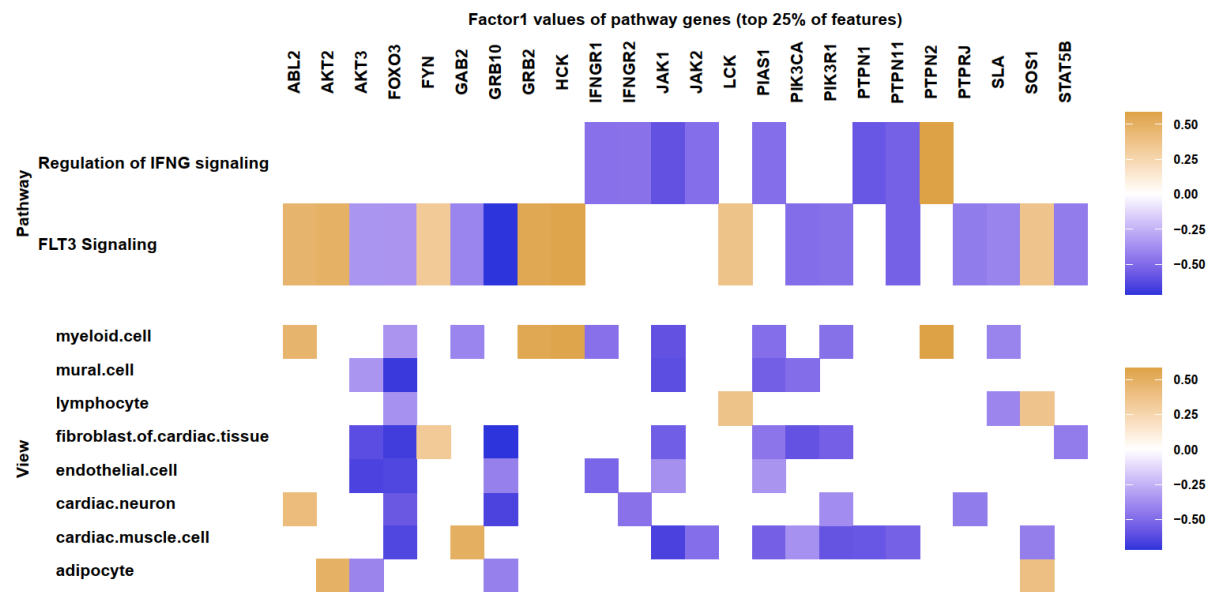
5.3.1. Prepare a '.csv' file ('Prepared_Pathway_Data.csv') that contains a list of pathways that will be tested for enrichment. Copy the file to the 'input_data' folder by using 'Drag&Drop' dropping the file in the folder overview of the 'input_data' folder.

[For this example this dataset is provided on zenodo and was already downloaded in the beginning to the 'input_data' folder, so this step is not necessary]

5.3.2. Use the navigation menu to navigate to the 'scripts' folder by clicking on 'scripts'. Open the notebook '06_Downstream_Pathways.ipynb' by double-clicking on it. Execute the script by clicking on the 'Restart kernel and run all cells' button in the top.

5.3.3. Use the navigation menu on the left to navigate to the '/figures/06_figures' folder by first double-clicking on 'figures' and then '06_figures'. Open the generated plot 'FIG06_Pathways_and_Genes_mofa_result_name' by double-clicking on it and investigate the visualized pathways.

NOTE: how the visualized pathways are selected can be configured via the configuration file. For more details refer to the documentation of the parameters.



Example: negatively enriched pathways on Factor1. Within the heatmap the feature factor weight of each gene on the factor is shown. In the top aggregated for all the different cell-types and below per cell-type.