Running the JGI Integration Workflow

Overview

The general stages of the workflow are:

- Loading and normalizing multi-omics datasets (e.g., transcriptomics and metabolomics)
- Integrating data (quantitative values) and metadata (sample info) between datasets
- Selecting features for analysis via statistical tests
- Building and visualizing a correlation network of integrated features
- Running multi-omics factor analysis

Running the Notebook

Note: This tutorial assumes you've already completed all stages of the workflow and environment setup detailed in *setup.pdf.*

1. You should see the JupyterLab interface rendered in your browser tab.
2. Double-click the workflow notebook `/integration_workflow.ipynb` in the left menu navigator to bring it into the workspace.
3. Double-click the configuration file in `/input_data/config/project_config.yml` to bring it into the workspace
4. Begin a workflow with all default parameters:
    - Ensure that the kernel is "JGI Integration" by checking the top right corner of the workspace. If not, click the kernel name and select "JGI Integration" from the dropdown menu.
    - Run all notebook cells in order, either with the "play" button in the top menu bar or with the keyboard shortcut in each cell (ctrl/cmd/shift-click).
    - Each cell performs a workflow step (data loading, normalization, integration, correlation analysis, etc.) and prints some information to the standard output for review.
    - Review outputs to see where plots and tables are saved to the `/input_data` directory; some key results or previews are also displayed in the notebook.
5. Run a workflow with customized parameters:
    - Update parameters in the `/input_data/config/project_config.yml` file (in the JupyterLab interface) using the guides provided in the `/jgi_integration/docs/*_parameters_explained.md` files.
    - Rerun the notebook as described above from the beginning to re-load the updated configuration settings and generate new outputs.