

In this pipeline, you submit a spreadsheet of gene-level transcriptomic (or other omics) profiles of a collection of biological samples. Each sample is also annotated with a numeric phenotype (e.g., drug response) or binary phenotype (e.g., metastatic status). This pipeline scores each gene by the correlation between its “omic” value (e.g., expression) and the phenotype, and reports the top phenotype-related genes. Gene prioritization can be done in a Knowledge Network-guided mode (using ProGENI, PMID:28800781), and with optional use of bootstrapping to achieve robust prioritization.

Pipeline Selection

1. Click on “Start A New Pipeline” on the homepage or on “Analysis Pipelines” at the top.
2. Hover the cursor over “Gene Prioritization” and click on the “Start Pipeline” button next to it.

Data Files

3. **“Select species”**: This quickstart demo run uses human transcriptomic data, so keep the default setting of “Human”.
4. **Upload transcriptomic spreadsheet**:
 - a. Click on the “Use Demo Data” button. The file [“demo_GP.genomic.txt”](#) will immediately begin loading into the data table.
 - b. Once the demo file appears in the data table, make sure the checkbox to the right of the filename is selected (it should be checked automatically when you use demo data).
 - c. Click “Next” at the bottom right corner.
5. **Upload phenotype spreadsheet**
 - a. Here, again click on the “Use Demo Data” button. The file [“demo_GP.phenotypic.txt”](#) will immediately begin loading into the data table.
 - b. Once the demo file appears in the data table, make sure the checkbox to the right of the filename is selected (it should be checked automatically when you use demo data).
 - c. Click “Next”

Parameters

6. On the new page that appears, keep the ‘primary prioritization method’

Demo File Downloads

Before you get started with this tutorial, [Download demo files](#) and [unzip them](#).

Currently our platform accepts tsv formatted files only.

About the Demo File:

Transcriptomic file

This genomic spreadsheet preprocessed for this demo contains basal gene expression profiles (all genes) of 491 cancer cell lines originally obtained from CCLE [PMID: 22460905].

This is a big file (~74 MB), so please be patient as it uploads! You can use a spreadsheet software such as Excel to view it locally if you are curious.

at its default value ('Absolute Pearson Correlation') and click 'Next'.

Knowledge Network

7. On the new page ('Do you want to use the Knowledge Network?'), select 'Yes', keep the default settings, click 'Next'.

Bootstrapping

8. On the new page ('Do you want to use bootstrapping?'), select 'Yes', keep the default settings otherwise and click 'Next'.

Review & Submission

9. The new page is a summary of various choices you've made until now, and you can go ahead and click 'Submit Job'. The new page should say 'Running!' and even if it shows a spinning wheel, click on 'Go to Results Page'.

Results and Visualization

10. You should now see a listing of all jobs you've run, with the latest one at the top. (It should have the name "gene_prioritization_<date>"). Wait for a few minutes until the status column for this entry turns to a green checkmark. Click on the row and then on the "View Results" button in the Info Panel on the right.
11. The main panel shows a grid view of genes (rows) and phenotypes (columns) – drug treatments in this sample data set – with colored cells indicating an association detected between that gene and phenotype pair. You can sort this view by a specific column (by right-clicking on the column name) to find the top genes for a drug.
12. Using the left panel, you can select how many top scoring genes (per phenotype) you want to include in the grid view. You may also select any subset of phenotypes you wish the grid to display.

Downloading Results

13. Click on 'Results' at the top of the page. You should see a listing of the various jobs you've run, with the current job at the top. Click on it and then click the Download Icon in the right panel. Wait a few moments until a .zip archive is downloaded.
14. In the downloaded archive, there is a tsv (tab-separated values) text file for each phenotype that lists the genes sorted by their relevance to the phenotype. There is also a file that has such sorted lists of relevant genes for every phenotype, it is a tsv file called 'all_phenotypes...'. Currently the genes are listed in this file by their Ensembl gene ids (ENSG...) and you can convert these to human readable gene names using the gene_name_mapping.tsv file. The file README-GP has more information about the downloaded results.

About the Demo File:

Phenotypic file

The sample phenotypic file for this demo contains values of 24 different phenotypes for each cell line in the transcriptomics file you used above.

Each phenotype here is the cell line's cytotoxicity in response to a drug, measured as IC50 values. *This is a much smaller spreadsheet than the transcriptomic file, so it should upload fast.*

Using the Knowledge Network

In a nutshell, using the Knowledge Network allows you to identify genes that both directly and indirectly (through their network neighbors) affect the phenotype. By selecting this option, you will use a method called ProGENI (PMID:28800781).

If you don't wish to use the Knowledge Network, select "No" and go to the next step.

Advanced Information

Additional information about making Knowledge Network and Bootstrapping selections is available in the Info Panel on the right side of each page.

If it's not visible click on the "?" to the right of each page.

Binary Phenotypes

To run the pipeline on a different dataset with 'binary' phenotypes, e.g., if the sample belongs to a patient in the test versus control group, you need to repeat steps 1-10 while changing step 6 to 't-test'.

Running Multiple Jobs Using the Same Data Files

When re-running the pipeline on the same data files you don't need to upload the files again in steps 4, 5. Instead, those files will automatically be available in the table. Click on the checkbox to select them and then click "Next".

