DMirNet User Guide

Version 1.0

6-6-2018

Table of Contents

List of Figures

List of Table

1 Introduction

1.1 Application overview

DMirNet is a Shiny-based application that leverages the DMirNet framework [1] for inferring direct miRNA-mRNA association from expression profiles. By taking advantage of the reactive programming framework of Shiny, our Shiny-based DMirNet provides a GUI-based interactive web application and makes dry-lab experiments for exploring miRNA-mRNA associations simple and intuitive. To explore direct miRNA-mRNA associations, DMirNet incorporates four direct correlation estimation methods, namely Space, CorND, Corpcor, and IDA. Moreover, DMirNet allows users to run bootstrapping and Ensemble aggregation procedures of the methods to obtain more reliable and robust results. For increased efficiency of the dry-lab experiments, DMirNet supports parallel processing via MPI and reuse of intermediate results. In terms of results analysis, DMirNet allows users to select highly ranked miRNA-mRNA pairs from the experimental results and to validate the predictions using a given ground-truth dataset.

1.2 Features

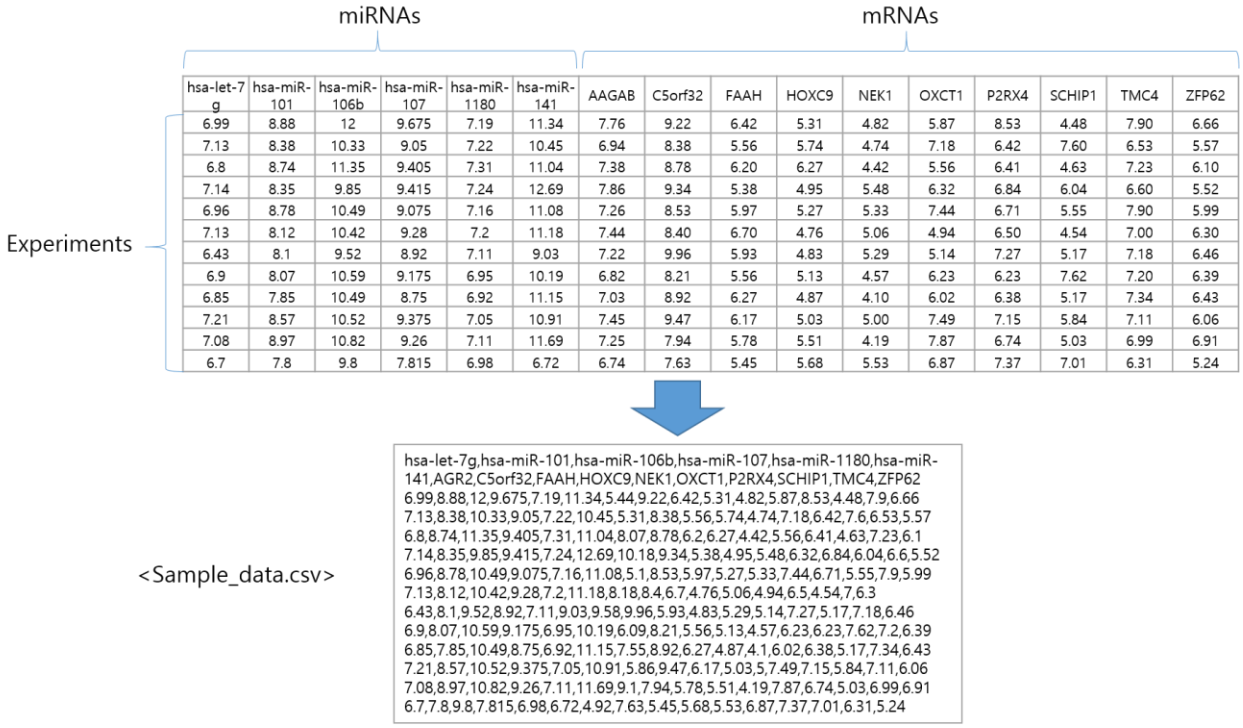
- Supports four direct correlation estimation methods:
  - 'Space' estimates non-zero partial correlations assuming overall sparsity of the partial correlation matrix [2].
  - 'Corpcor' estimates partial correlations by suppressing the effect of a set of controlling random variables [3].
  - 'CorND' infers direct effects from an observed correlation matrix using network deconvolution method [4] on Spearman's correlation coefficients.
  - 'IDA' estimates the multiset of possible total causal effects from observation data [5].
- Supports bootstrapping of direct correlation estimation methods.
- Supports Ensemble aggregation of inferred direct correlation results.
- Supports selection of highly ranked pairs from results of direct estimation methods and Ensemble aggregations.
- Supports evaluation of the inferred results by comparing the selected top-ranked miRNA-mRNA pairs with a given ground-truth dataset or experimentally confirmed dataset.
- Displays the output files in a table format that allows searching and filtering the data.
- Saves the output files in an organized directory structure.

2 Getting Started

2.1 Input Data Preparation

DMirNet takes as input data a file containing pre-processed matched miRNA and mRNA expression profiles in CSV format (Figure 1). For more reliable results, pre-processing of the input gene expression profiles should include normalization and differential gene expression analysis.

The dataset should not include NA (Non-Accessible) values. If there are any missing values in the dataset, an additional process of missing value imputation is needed.



hsa-let-7g	hsa-miR-101	hsa-miR-106b	hsa-miR-107	hsa-miR-1180	hsa-miR-141	AAGAB	C5orf32	FAAH	HOXC9	NEK1	OXCT1	P2RX4	SCHIP1	TMC4	ZFP62
6.99	8.88	12	9.675	7.19	11.34	7.76	9.22	6.42	5.31	4.82	5.87	8.53	4.48	7.90	6.66
7.13	8.38	10.33	9.05	7.22	10.45	6.94	8.38	5.56	5.74	4.74	7.18	6.42	7.60	6.53	5.57
6.8	8.74	11.35	9.405	7.31	11.04	7.38	8.78	6.20	6.27	4.42	5.56	6.41	4.63	7.23	6.10
7.14	8.35	9.85	9.415	7.24	12.69	7.86	9.34	5.38	4.95	5.48	6.32	6.84	6.04	6.60	5.52
6.96	8.78	10.49	9.075	7.16	11.08	7.26	8.53	5.97	5.27	5.33	7.44	6.71	5.55	7.90	5.99
7.13	8.12	10.42	9.28	7.2	11.18	7.44	8.40	6.70	4.76	5.06	4.94	6.50	4.54	7.00	6.30
6.43	8.1	9.52	8.92	7.11	9.03	7.22	9.96	5.93	4.83	5.29	5.14	7.27	5.17	7.18	6.46
6.9	8.07	10.59	9.175	6.95	10.19	6.82	8.21	5.56	5.13	4.57	6.23	6.23	7.62	7.20	6.39
6.85	7.85	10.49	8.75	6.92	11.15	7.03	8.92	6.27	4.87	4.10	6.02	6.38	5.17	7.34	6.43
7.21	8.57	10.52	9.375	7.05	10.91	7.45	9.47	6.17	5.03	5.00	7.49	7.15	5.84	7.11	6.06
7.08	8.97	10.82	9.26	7.11	11.69	7.25	7.94	5.78	5.51	4.19	7.87	6.74	5.03	6.99	6.91
6.7	7.8	9.8	7.815	6.98	6.72	6.74	7.63	5.45	5.68	5.53	6.87	7.37	7.01	6.31	5.24

```
hsa-let-7g,hsa-miR-101,hsa-miR-106b,hsa-miR-107,hsa-miR-1180,hsa-miR-
141,AGR2,C5orf32,FAAH,HOXC9,NEK1,OXCT1,P2RX4,SCHIP1,TMC4,ZFP62
6.99,8.88,12,9.675,7.19,11.34,5.44,9.22,6.42,5.31,4.82,5.87,8.53,4.48,7.9,6.66
7.13,8.38,10.33,9.05,7.22,10.45,5.31,8.38,5.56,5.74,4.74,7.18,6.42,7.6,6.53,5.57
6.8,8.74,11.35,9.405,7.31,11.04,8.07,8.78,6.2,6.27,4.42,5.56,6.41,4.63,7.23,6.1
7.14,8.35,9.85,9.415,7.24,12.69,10.18,9.34,5.38,4.95,5.48,6.32,6.84,6.04,6.6,5.52
6.96,8.78,10.49,9.075,7.16,11.08,5.1,8.53,5.97,5.27,5.33,7.44,6.71,5.55,7.9,5.99
7.13,8.12,10.42,9.28,7.2,11.18,8.18,8.4,6.7,4.76,5.06,4.94,6.5,4.54,7,6.3
6.43,8.1,9.52,8.92,7.11,9.03,9.58,9.96,5.93,4.83,5.29,5.14,7.27,5.17,7.18,6.46
6.9,8.07,10.59,9.175,6.95,10.19,6.09,8.21,5.56,5.13,4.57,6.23,6.23,7.62,7.2,6.39
6.85,7.85,10.49,8.75,6.92,11.15,7.55,8.92,6.27,4.87,4.1,6.02,6.38,5.17,7.34,6.43
7.21,8.57,10.52,9.375,7.05,10.91,5.86,9.47,6.17,5.03,5.7,7.49,7.15,5.84,7.11,6.06
7.08,8.97,10.82,9.26,7.11,11.69,9.1,7.94,5.78,5.51,4.19,7.87,6.74,5.03,6.99,6.91
6.7,7.8,9.8,7.815,6.98,6.72,4.92,7.63,5.45,5.68,5.53,6.87,7.37,7.01,6.31,5.24
```

Figure 1. Input Data Preparation

2.2 Prerequisite

Download and install ***R*** from https://cran.r-project.org/. (Note: DMirNet requires ***R*** version 3.4.3 or greater.)

The following R packages must be installed to run Shiny_DMirNet. Run the following scripts to install the packages:

install.packages("shiny")
install.packages("checkpoint")

2.3 Installing/Launching DMirNet

There are two ways of installing DMirNet as follows.

- Quick install/launch instruction: To run DMirNet locally from GitHub, run the script that follows. (The following R code will launch DMirNet on most systems.)
  shiny::runGitHub('Shiny_DMirNet','dmirnet')
- Download the source code and launch DMirNet locally: Download the zip file of Shiny_DMirNet from https://github.com/dmirnet/Shiny_DMirNet, and extract the source files. Then run one of the source files – global.R, ui.R, or server.R. To run the source file, type the following script:
  source("<path of the source file>/global.R")

2.4 Directory Structure of DMirNet output files

The DMirNet application includes four tabs, namely 'Experiments,' 'Ensemble,' 'Result Analysis,' and 'Validation.' For each experiment, outputs from functions in each tab are stored in a corresponding directory. The organization of the directory structure of DMirNet output files is illustrated in Figure 2 and Table 1.

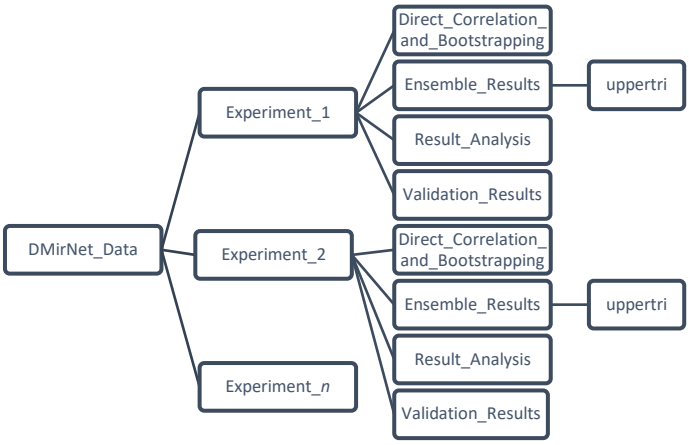


Figure 2. Output Files Directory Structure

Table 1: Description of output files directory structure

Output file directory structure description		
Directory Name	**Sub-directories**	**Description**
DMirNet_Data	• Experiment_1, • Experiment_2, • Experiment_*n*	The root directory of the output files. "Experiment_*n*" includes results from one experiment. The number of experiments (*n*) is automatically increased whenever a user creates a new experiment.
Experiment_*n*	• Direct_Correlation_and_Bootstrapping • Ensemble_Results • Result_Analysis • Validation_Results	"Experiment_*n*" contains the sub-directories of the output files of an experiment.
Direct_Correlation_and_Bootstrapping	• uppertri	Directory that contains the output file of direct correlation estimation methods and their upper triangular results
Uppertri	~	Directory that contains the upper triangular format of the output files from direct correlation estimation methods. These files are used for fast execution of Bootstrapping or Ensemble Aggregation.
Ensemble_Results	~	Directory that contains the output file of running Ensemble Aggregation
Result_Analysis	~	Directory that contains the output file of running Result Analysis
Validation_Results	~	Directory that contains the output file of running Validation

3 Using DMirNet

3.1 Performing an Experiment

DMirNet supports a pipeline following the DMirNet framework [1]. Its GUI harbors four tabs that facilitate dry-lab experiments (Figure 3).

- Experiment tab: To execute the direct correlation estimation methods on a given dataset, with or without bootstrapping
- Ensemble tab: To execute an ensemble aggregation of the results of direct correlation methods
- Result Analysis tab: To select highly ranked pairs from direct correlation estimation method results and Ensemble aggregation results
- Validation tab: To evaluate the inferred results using known miRNA-mRNA pairs

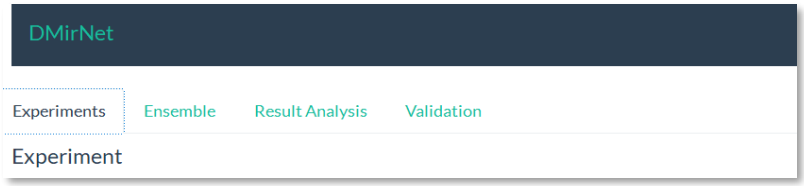


Figure 3. DMirNet Tabs

3.1.1 Experiments Tab

The Experiment tab is the initial stage of an experiment. It performs direct correlation methods on a selected dataset with or without bootstrapping.

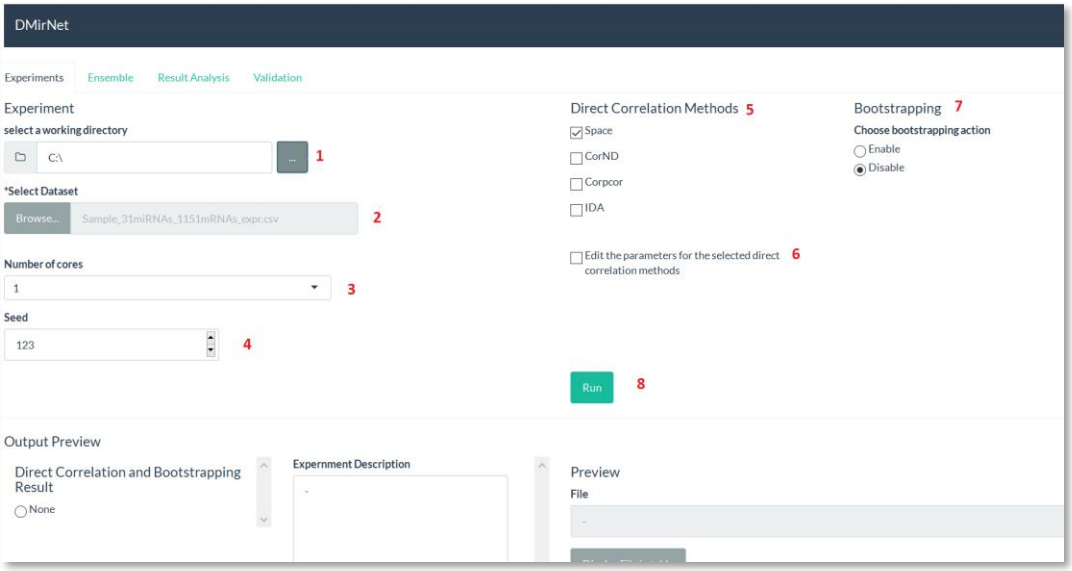


Figure 4. Experiments Tab

Experimental settings:

1. **Select a working directory and experiment directory**: Select a working directory to save experimental results. DMirNet application generates a sub-directory in the working directory for each experiment, as needed. There are two ways of setting an experiment directory.
    - To run a new experiment: If there is no existing experiment directory in the working directory, DMirNet generates a new directory named "Experiment_Results_1." If experiment directories already exist, select "New experiment" among the radio button options (Figure 5).
    - To continue an experiment: Select the root directory of an existing experiment. The root directory should be the parent directory of "DMirNet_Data" folder. Then, select an existing directory for the experiment from the available radio button options (Figure 5). The structure of an experiment directory is described in section 2.4.

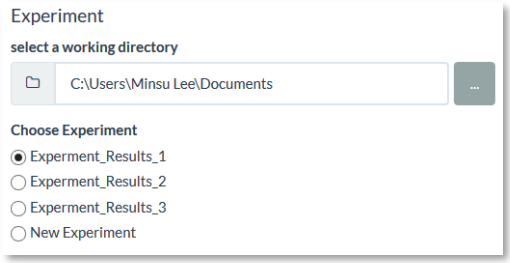


Figure 5. Selecting a Working Directory

2. **Select a Dataset**: Select a dataset on which to perform the experiment. The dataset should be in ".csv" format, and it should not include NA (Non-Accessible) values.
3. **Number of Cores**: Select the number of cores that can be used in parallel for the experiment to improve its performance.
4. **Seed:** Set a seed for reproducible results. Its value should be a positive integer.
5. **Direct Correlation Methods**: Select a direct correlation method with which to perform the experiment.
    - To explore direct miRNA-mRNA associations from expression profiles, DMirNet provides four direct correlation analysis methods, namely 'Corpcor,' 'Space,' 'CorND,' and 'IDA.' The result of each method is normalized linearly between -1 and +1.
    - Note: The output of 'Space' is a sparse correlation matrix that includes numerous "NA" elements that indicate non-edge. This might induce biases in the ensemble aggregation results. The total number of edges detected can be determined by tuning the value of lambda. The total number of edges detected decreases as lambda increases [2].
6. Select "*Edit the parameters for the selected direct correlation methods*" to edit the default value of the direct correlation methods. Descriptions of the parameters can be found in the corresponding references:
    - Space [2]: space.joint() https://cran.r-project.org/web/packages/space/space.pdf
    - Corpcor [3]: pcor.shrink() https://cran.r-project.org/web/packages/corpcor/corpcor.pdf

- CorND [4]: ND() http://compbio.mit.edu/nd/code/ND.m
- IDA [5]: idaFast() https://cran.r-project.org/web/packages/pcalg/pcalg.pdf.

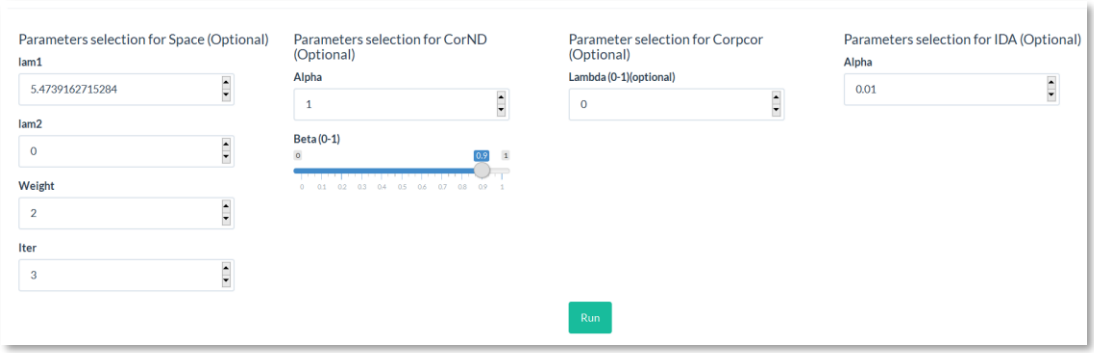


Figure 6. Editing the Default Parameters of Direct Correlation Methods

7. **Bootstrapping**: Click on 'Enable' to enable bootstrapping of the experiment. Select the number of iterations, the sampling rate, and a bootstrapping method, as shown below.

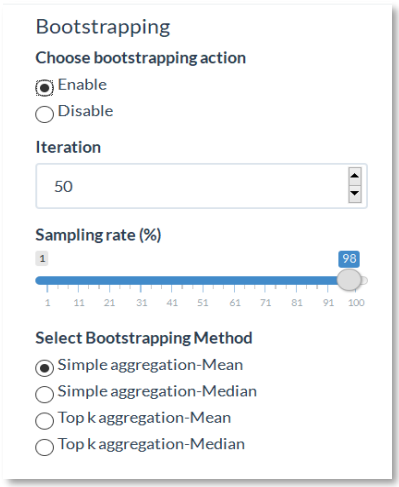


Figure 7. Bootstrapping Parameters

8. To start running the experiment, click on the "Run" Button.

3.1.2 Running an Ensemble Aggregation

An ensemble aggregation is performed on multiple output files from the result of multiple direct correlation methods' run.

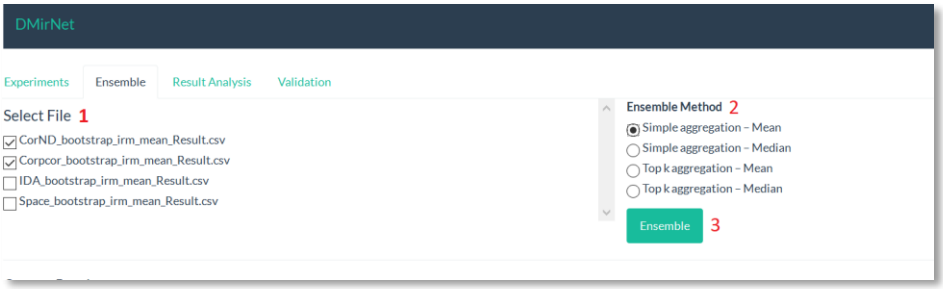


Figure 8. Ensemble Aggregation

Description of inputs:

1. **Select File**: Select multiple files, from the list of files, on which to perform the ensemble aggregation.
2. **Ensemble Method**: Select the desired ensemble aggregation method.
3. Click on "Run" to run the ensemble aggregation.

3.1.3 Running a Result Analysis

Result Analysis can be performed on the output files from a direct correlation experiment run or an ensemble aggregation run.

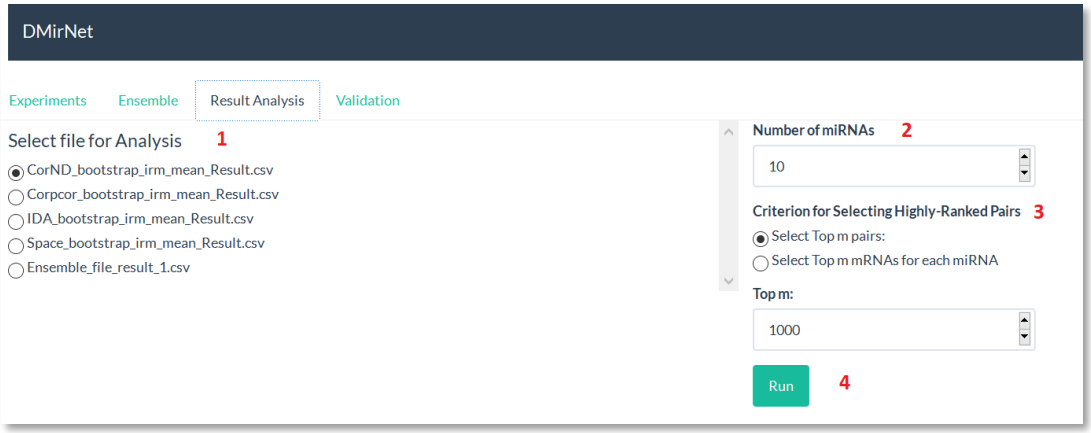


Figure 9. Result Analysis

Description of inputs:

1. **Select File for Analysis**: Select an output file from the list of files.
2. **Number of miRNAs**: Set the number of miRNAs in input dataset.
3. **Criterion for Selecting Highly-Ranked Pairs**: Choose a criterion for selecting highly ranked pairs, and set the number of **Top m**.
4. Click on "Run" to run the result analysis of the selected file.

3.2 Running Validation

Validation is available to validate the output files from running a result analysis.

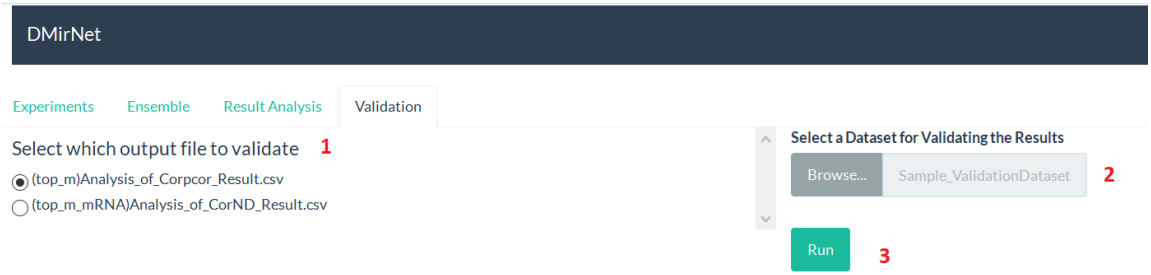


Figure 10. Validation

Description of inputs:

1. **Select which output file to validate**: Select an output file from the list of files on which to perform the validation.
2. **Select a Dataset for Validating the Results**: Choose a dataset (CSV file format) for result validation that includes known miRNA and mRNA association pair.
3. Click on "Run" to run the validation.

<Sample_ValidationDataset.csv>

```
miRNA,Gene
hsa-miR-218, MAFG
hsa-miR-125b,4E-BP1
hsa-miR-335,7SK
hsa-miR-26b,A26C1B
hsa-let-7b,A2BP1
hsa-miR-30a,A2BP1
hsa-miR-34a,A2D
hsa-miR-34a,A2LG
hsa-miR-34a,A2LP
hsa-miR-122,A2M
hsa-miR-128,A2M
hsa-miR-98,A2M
hsa-miR-34a,A2RP
hsa-miR-193b,A4GALT
hsa-miR-7,AA534526
hsa-miR-215,AA598803_RC
hsa-miR-7,AA707320
hsa-miR-16,AAAS
```

Figure 11. File Format of Validation Dataset

3.3 Output Result Preview

After performing an experiment, a preview of the output files and their description is displayed on the output preview section of each tab.

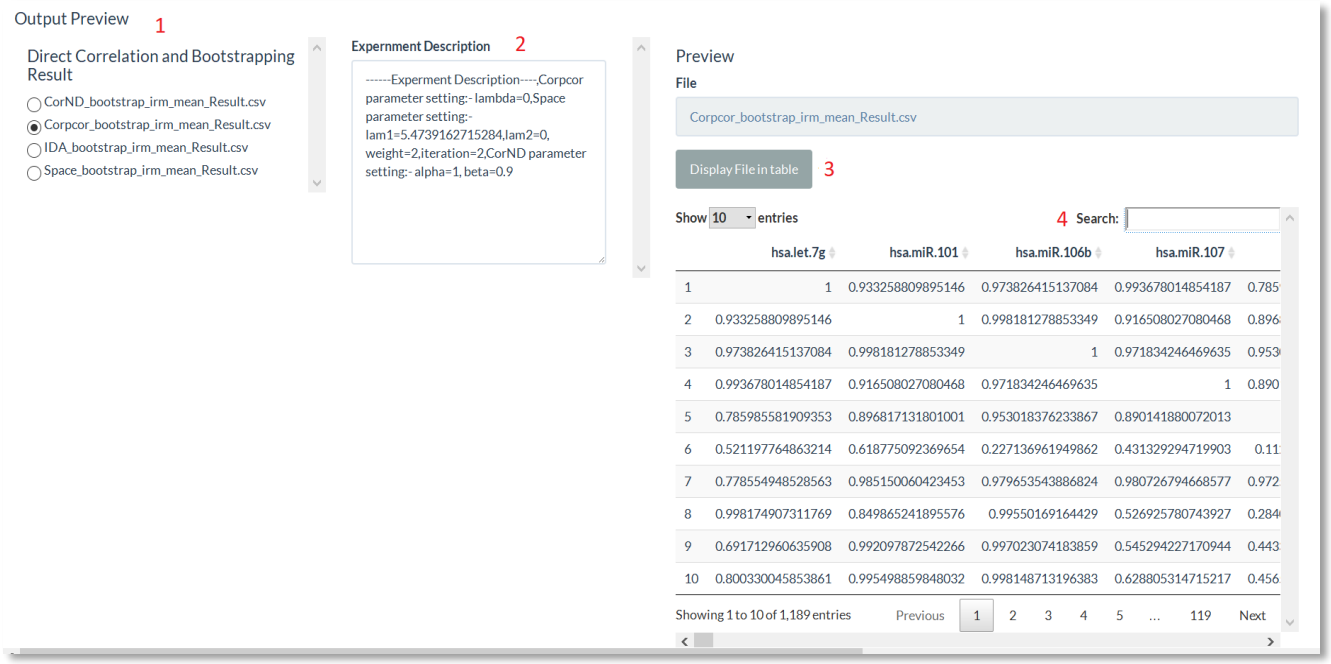


Figure 12. Output Result View

Description:

1. **Output Preview**: Lists all the output files from the running experiment.
2. **Experiment Description**: Description of the input values used in the experiment.
3. **Display File in table:** To view the selected output file in table format. The table can be sorted, searched, and filtered.
4. **Search:** Type a keyword to search any value in the table. You can also perform filtering by typing a filtering value in the input.

4 References

[1] Lee,M and Lee,HJ. (2016) DMirNet: Inferring direct microRNA-mRNA associa-tion networks, *BMC Systems Biology*, 10(suppl 5), 125.

[2] Peng,J., Wang,P., Zhou,N., and Zhu,J. (2009) Partial correlation estimation by joint sparse regression models. *J Am Stat Assoc – Theory and Methods*, 104(486), 735–46.

[3] Schäfer,J. and Strimmer,K. (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statist Appl Genet Mol Biol*, 4, 32.

[4] Feizi,S., Marbach,D., Médard,M., and Kellis,M. (2013) Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat Biotechnol*, 31, 726–33.

[5] Maathuis,M.H., Colombo,D., Kalisch,M., and Bühlmann,P. (2010). Predicting causal effects in largescale systems from observational data. *Nature Methods*, 7, 247–248.