easyMF User Mannual

(version 1.0)

easyMF is a user-friendly web platform that aims to facilitate biological discovery from large-scale transcriptome data through matrix factorization (MF). It offers several functional tools for gene expression matrix generation, expression matrix factorization, and metagene-based exploratory analysis including sample clustering, signature gene identification, and functional gene discovery.It should be noted that the metagenes in PCA can be ranked by the extent to which they explain the variation in the data, the features in both ICA and NMF are assumed to have equal weight. Therefore, we noticed the user that the choice of MF techniques should be based on the purpose of one study with a preferred choice: PCA finds sources of separation in the data, which resulting in identification of dominated variation; ICA learns variation that are statically independent, which resulting in more accurate literature-derived association; NMF is well suited to transcriptional data, which is typically non-negative itself.

- easyMF project is hosted on https://github.com/cma2015/easyMF.
- easyMF docker image is available in https://hub.docker.com/r/malab/easymf.
- easyMF demo server can be accessed via http://easymf.omicstudio.cloud.
- The following part shows installation of easyMF docker image and detailed documentation for each function in easyMF.

0. Metagene-based Deep Mining Using AM

Amplitude matrix (AM), a matrix with genes in rows and metagenes in columns, describes gene-level relationships. In current version of easyMF, users can make use of AM for functional gene discovery.

This module consists of one function: **Functional Gene Discovery**.

Functions/Tools	Description	Inputs	Outputs	Time (test data)	Program	References
Functional Gene Discovery	Calculate gene score and rank genes based on the probability of their association with a specific biology function	Amplitude matrix; A set of genes with a specific characteristic	Gene score and rank; Area under the self-ranked curve (AUSR) plot	~ 10s	In-house scripts	Fehrmann *et al.*, 2015

1. Functional Gene Discovery

Functional gene discovery can be used to calculate gene score and rank genes based on the probability of their association with a specific biology function.

Inputs

- **Amplitude matrix**: An amplitude matrix of AM coefficients with genes in rows and metagenes in columns. Here is an example:

	Metagene 1	**Metagene 2**	**Metagene 3**	**...**	**Metagene n**
Zm00001d053636	0.080	-0.889	1.504	...	2.029
Zm00001d053632	1.338	0.729	-0.113	...	-0.049
...	...	...	...	...	...
Zm00001d053635	-1.674	0.036	-0.047	...	-0.494

- **Functional genes**: A set of genes associated with a specific biology function, such as enriched in a phenotype of interest. If users select **Upload a file with functional gene IDs from local disk**, a newline-delimited file containing gene IDs needs to be provided; if users select **Enter functional gene IDs**, gene IDs need to be separated by comma. Here are two examples:

A newline-delimited file containing gene IDs for **Upload a file with functional gene IDs from local disk**:

```
Zm00001d053636
Zm00001d053632
Zm00001d053630
...
Zm00001d053635
```

Comma-separated gene IDs for **Enter functional gene IDs**:

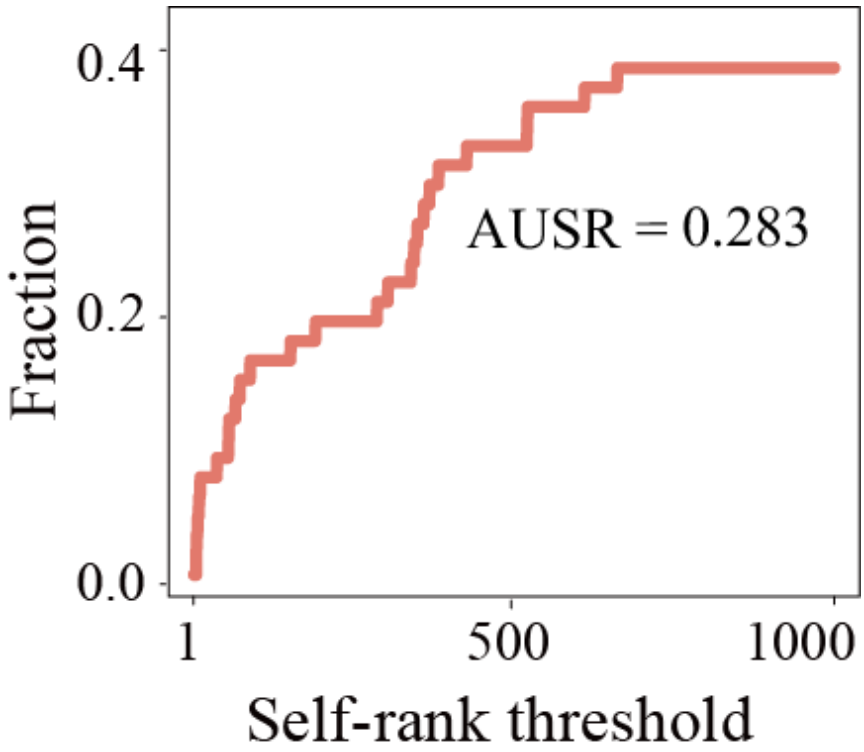
```
Zm00001d053636,Zm00001d053632,Zm00001d053630,...,Zm00001d053635
```

Outputs

- **Gene score and rank**: Summary of gene prioritization results containing **Gene ID**, **Score**, **Rank**, and **Annotation**. The higher ranking of a gene, the more related to the biological function. Here is an example:

Gene ID	Score	Rank	Annotation
Zm00001d053636	1	1	Label
Zm00001d053632	0.888	3	Label
Zm00001d004839	1	1	Unlabel
...	...	...	...
Zm00001d053635	0.92	2	Unlabel

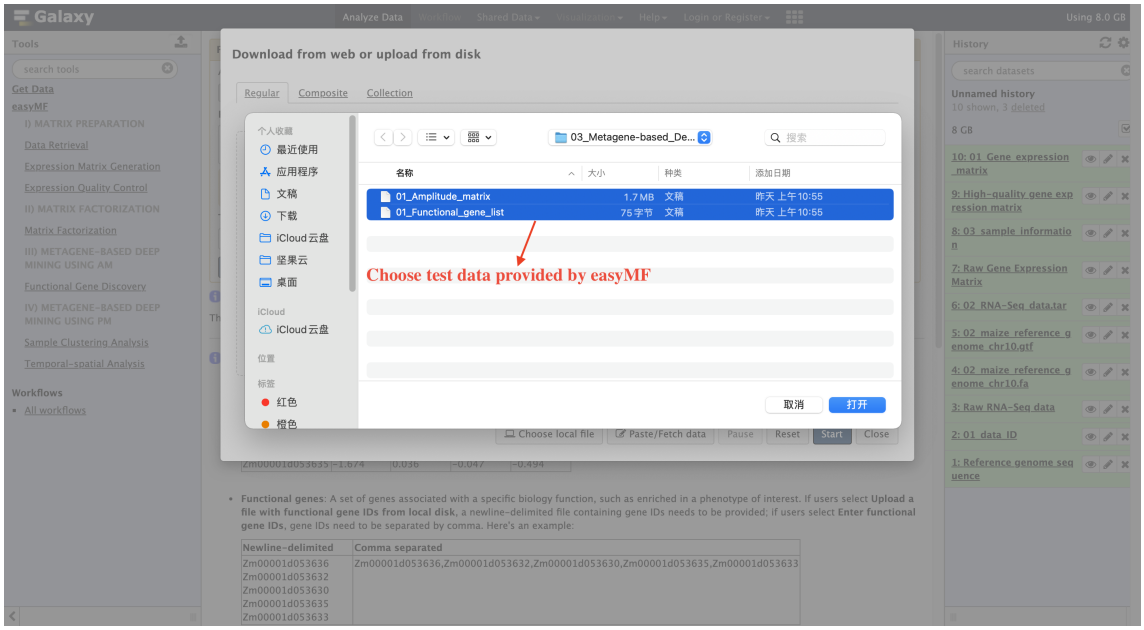
- **Area under the self-ranked curve (AUSR) plot**: A plot of ratio (*Ra*) along the *y* axis versus self-rank along the *x* axis, where rank represents the ranks of all positive genes, *Ra*(*l*) represents the ratio of ranks lower than a pre-defined level of *l*.



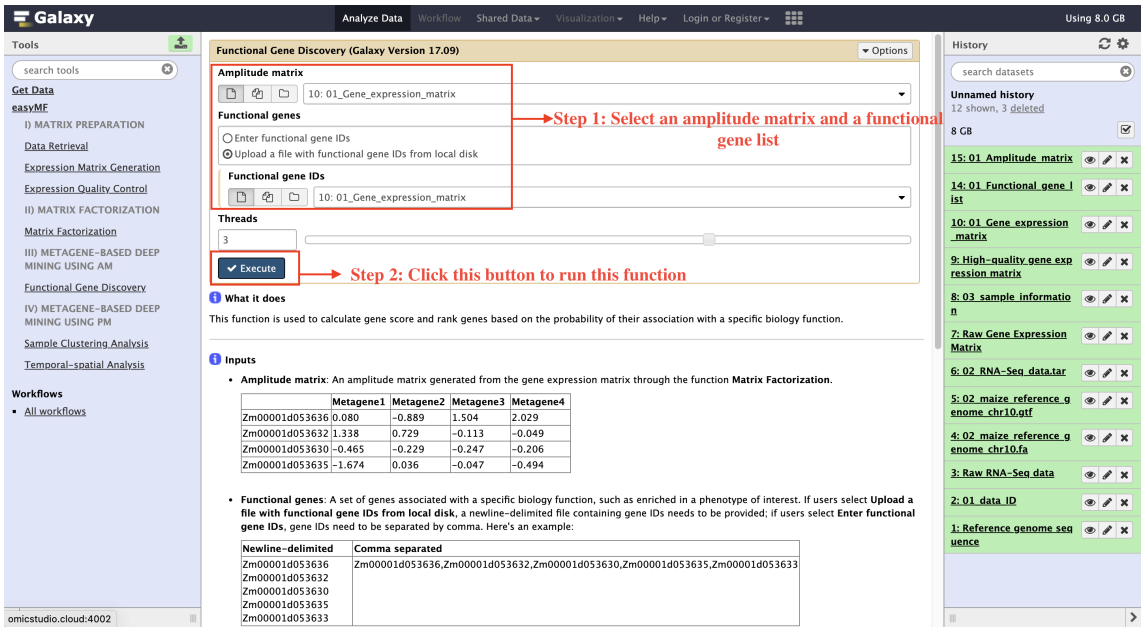
How to use this function

- Test data for this function are in directory `Test_data/03_Metagene-based_Deep_Mining_Using_AM` including.
- The following screenshots show us how to implement functional gene discovery using easyMF.

  Step 1: upload test data in directory `Test_data/03_Metagene-based_Deep_Mining_Using_AM` to history panel;



Step 2: input the corresponding files and appropriate parameters, then run the function.



Running time

This step will cost ~ 10s for the test data.