OEFinder: A user interface to identify and visualize ordering effects in single-cell RNA-seq data

Graphical User Interface Manual

Ning Leng, Jeea Choi, Christina Kendziorski and Ron Stewart

Table of Contents

1. Installation

OEFinder is a statistical method to identify ordering effect (OE) genes in single cell RNA-seq data (https://github.com/lengning/OEFinder). This manual is a guideline for using the OEFinder interface functions, which will allow a user to run OEFinder without coding in R. Files can be inputted in xls, txt, tab, or csv format.

We provide two graphical user interface (GUI) implementations of OEFinder. The two versions of GUIs were implemented using R/shiny and R/RGtk2 respectively. The user may choose either of the implementations. The shiny implementation may require less effort in installation.

1.1 Install OEFinder

1.1.1 OEFinder shiny implementation

The OEFinder shiny implementation requires installation of shiny package and a modern web browser (Chrome, Firefox or Safari). To in stall shiny, start R and type:

```
Install.packages('shiny')
```

A user may run the OEFinder shiny implementation via GitHub, which doesn't need extra implementations. A user could also save the OEFinder shiny implementation source codes to a local directory and run from there. To do so, download the OEFindershiny_0.0.1.zip and unzip it.

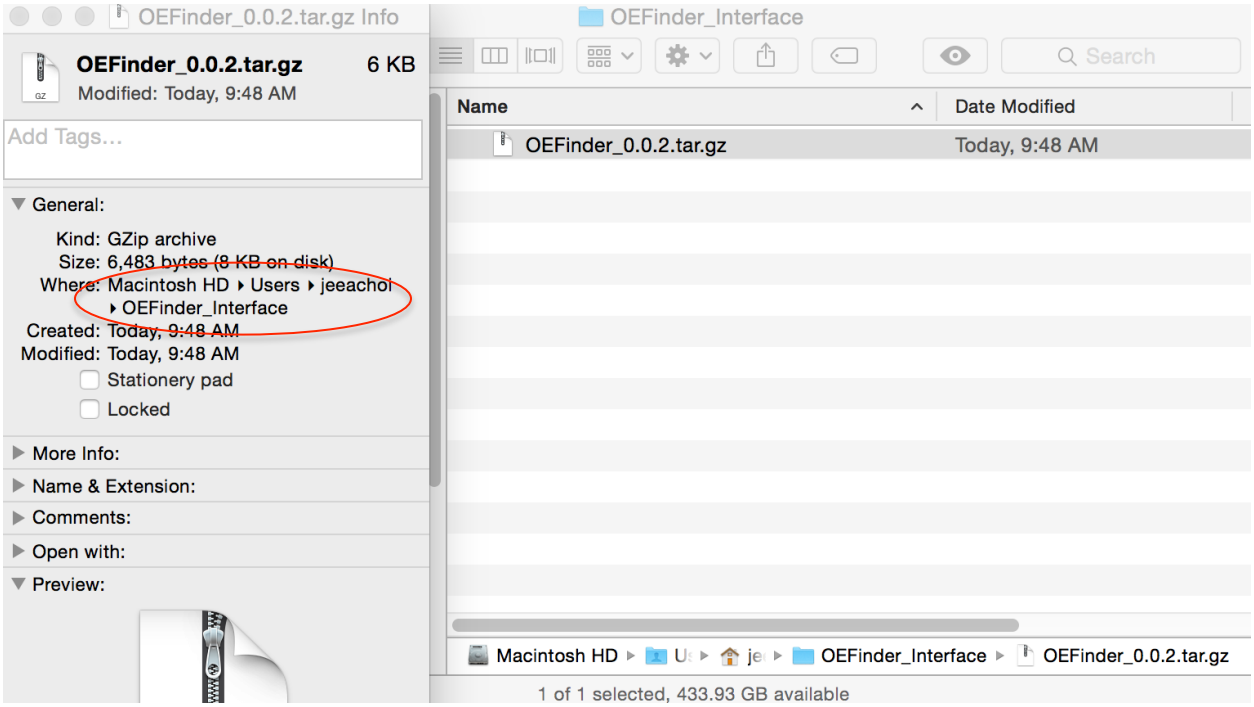
1.1.2 OEFinder RGtk2 implementation

The OEFinder RGtk2 implementation requires the installation of the OEFinderGtk package, and also requires the installation of the C package GTK+ and R package RGtk2. Both packages are available online for free. Detailed installation instructions are specified below. OEFinder requires R version >= 3.1.3.

The OEFinderGtk package is available at https://github.com/lengning/OEFinder/OEFinderGtk/

- Linux and Mac users please download the OEFinderGtk_0.0.4.tar.gz file
- Windows users please download the OEFinderGtk_0.0.4.zip file.

Download OEFinderGtk package to a folder. We will refer to this folder later in this manual as *YOUR_PATH*. For example, My *YOUR_PATH* directory is: /Users/jeeachoi/OEFinder_Interface/ (see screenshot below)



To install OEFinder in R:

Start R and type:

install.packages("*YOUR_PATH*/ OEFinderGtk_0.0.4.tar.gz", repos=NULL, type="source")

Windows users type:

install.packages("*YOUR_PATH*/ OEFinderGtk_0.0.4.zip", repos=NULL, type="source")

If you get an error regarding installation about RGtk2 and/or Gtk+, please go through section 1.2 and 1.3. Otherwise you may skip these two sections.

1.2 Install RGtk2

Start R and type

install.packages(RGtk2)

library(RGtk2)

If you get an error about Gtk+, please install (re-install) Gtk+ following the steps in section 1.3. If you don't get an error massage here, you may skip section 1.3.

1.3 Install Gtk+

Gtk+ downloads are available at http://www.gtk.org/download/index.php

Linux Users:

The GTK+2.X version is suggested for simple installation on linux. GTK+3.X will require higher versions of libraries. More detailed instructions can be found at gtk.org.

To start, type in your bash shell:

```
sudo apt-get install libgtk2.0-dev
sudo apt-get install glade
```

To check, type

```
gtk-demo
```

Windows Users:

a. Download and unzip the all-in-one GTK bundle of any version

b. Copy the complete bin/ folder of the bundle to *YOUR_PATH*

c. Open cmd.exe and set path into the bin/ folder in *YOUR_PATH*

d. Run commends to install

```
pkg-config --cflags gtk+-2.0
```

e. Check to see if the demo works

```
gtk-demo
```

2. Running OEFinder

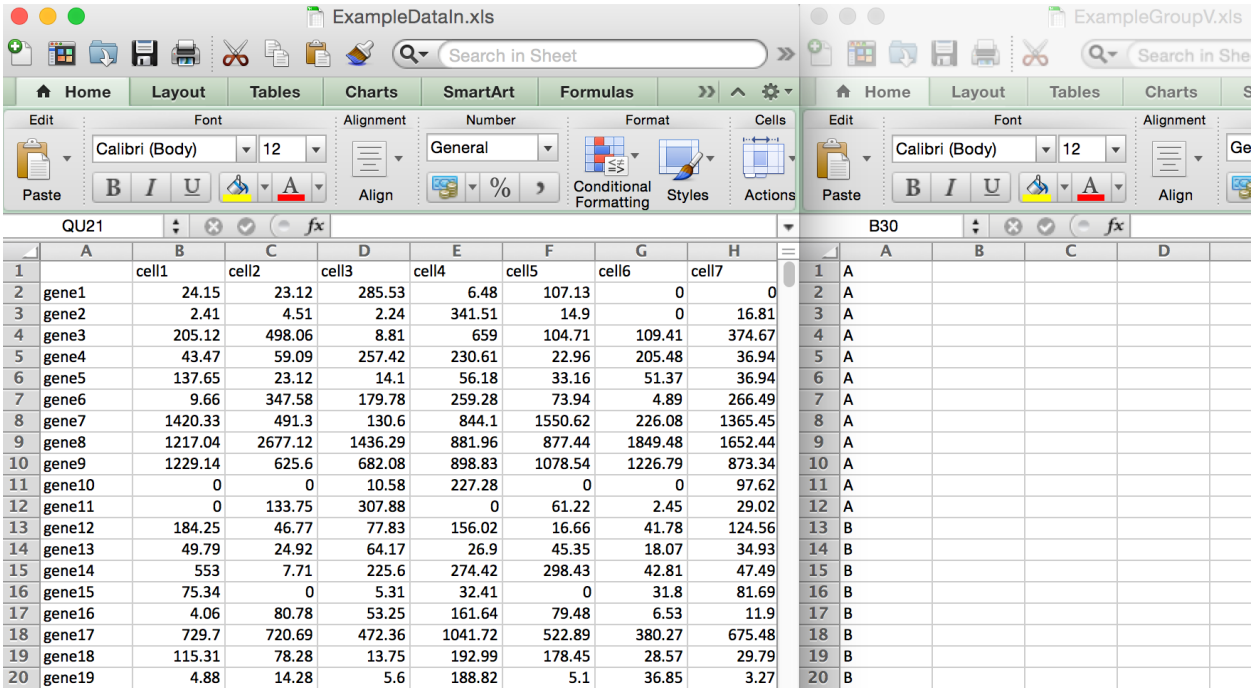
Input requirement:

The input file formats supported by OEFinder are .txt, .tab, .csv, or .xls. In your input file, the rows should be the genes and the columns should be the cells. In other words, your first row stores the cell names and the first column shows your gene names.

If the capture site ID of the data set is available, the user may input row number of the capture site IDs (A-H) as grouping vector. An example is shown below. If the capture site ID information is not available, OEFinder will group cells based on their input order. Number of groups can be specified by the user.

Example data set in .xls format:

Expression data (ExampleDataIn.xls) and grouping vector (ExampleGroupV.xls):



	A	B	C	D	E	F	G	H
1		cell1	cell2	cell3	cell4	cell5	cell6	cell7
2	gene1	24.15	23.12	285.53	6.48	107.13	0	0
3	gene2	2.41	4.51	2.24	341.51	14.9	0	16.81
4	gene3	205.12	498.06	8.81	659	104.71	109.41	374.67
5	gene4	43.47	59.09	257.42	230.61	22.96	205.48	36.94
6	gene5	137.65	23.12	14.1	56.18	33.16	51.37	36.94
7	gene6	9.66	347.58	179.78	259.28	73.94	4.89	266.49
8	gene7	1420.33	491.3	130.6	844.1	1550.62	226.08	1365.45
9	gene8	1217.04	2677.12	1436.29	881.96	877.44	1849.48	1652.44
10	gene9	1229.14	625.6	682.08	898.83	1078.54	1226.79	873.34
11	gene10	0	0	10.58	227.28	0	0	97.62
12	gene11	0	133.75	307.88	0	61.22	2.45	29.02
13	gene12	184.25	46.77	77.83	156.02	16.66	41.78	124.56
14	gene13	49.79	24.92	64.17	26.9	45.35	18.07	34.93
15	gene14	553	7.71	225.6	274.42	298.43	42.81	47.49
16	gene15	75.34	0	5.31	32.41	0	31.8	81.69
17	gene16	4.06	80.78	53.25	161.64	79.48	6.53	11.9
18	gene17	729.7	720.69	472.36	1041.72	522.89	380.27	675.48
19	gene18	115.31	78.28	13.75	192.99	178.45	28.57	29.79
20	gene19	4.88	14.28	5.6	188.82	5.1	36.85	3.27

	A
1	A
2	A
3	A
4	A
5	A
6	A
7	A
8	A
9	A
10	A
11	A
12	A
13	B
14	B
15	B
16	B
17	B
18	B
19	B
20	B

If you want to use the OEFinder shiny implementation via GitHub, in R type:

```
library(shiny)
runGitHub('OEFinder', 'lengning')
```

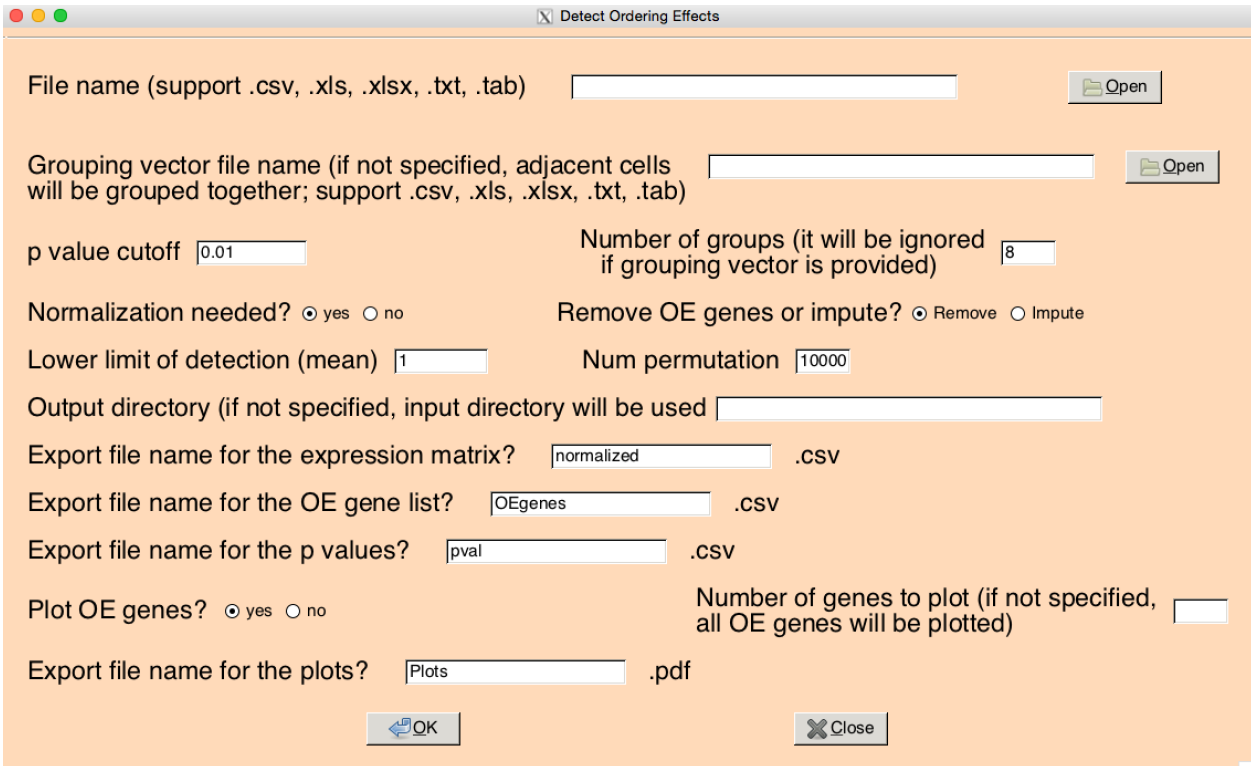
If you downloaded the OEFinder shiny zip package and unzipped it under YOUR_PATH, in R type:

```
library(shiny)
runApp('YOUR_PATH/OEFindershiny_0.0.1')
```

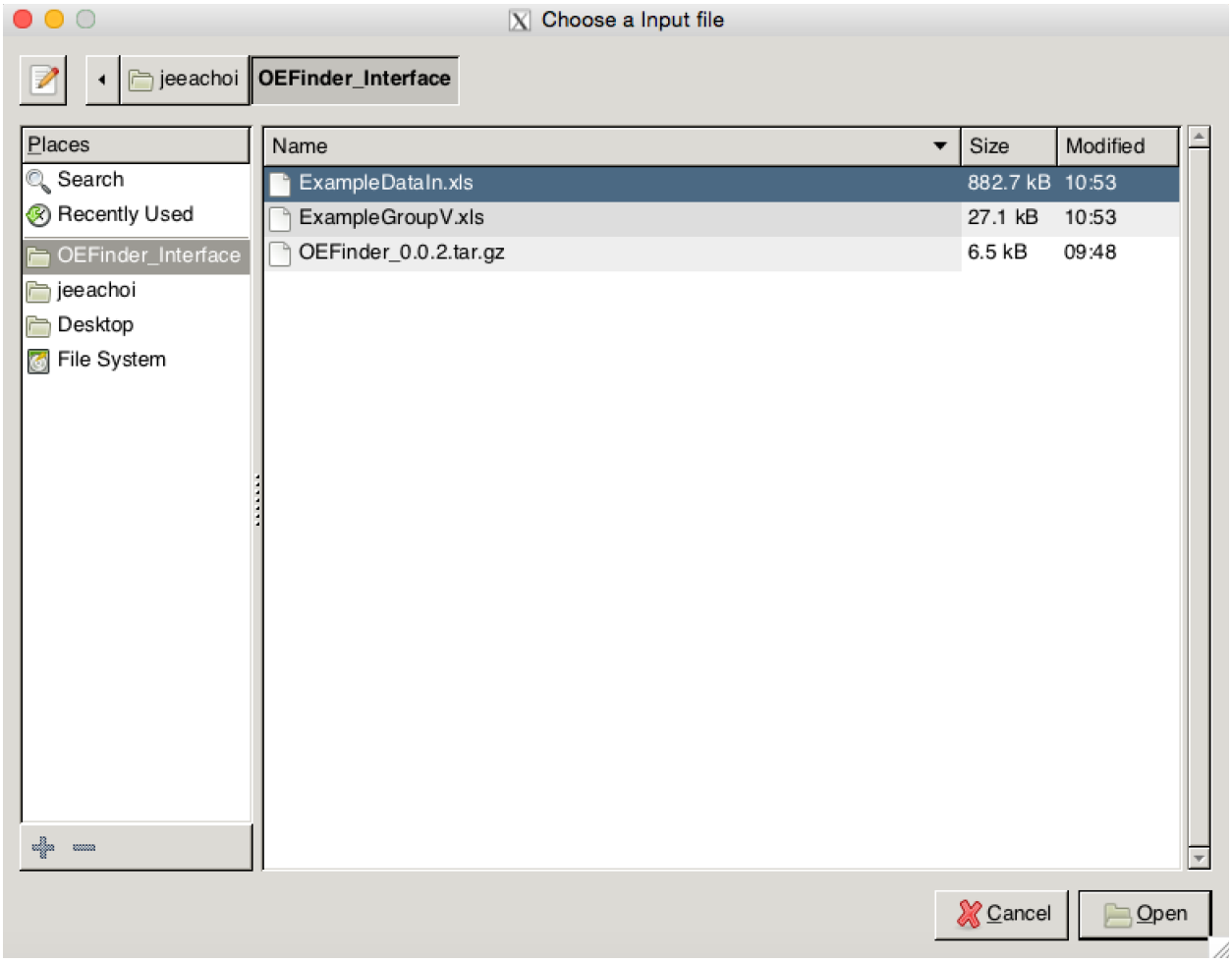
If you installed the OEFinderGtk implementation, in R type:

```
library(OEFinderGtk)
OEFinder()
```

A window will pop up (RGtk2 implementation is shown below, shiny implementation is similar):

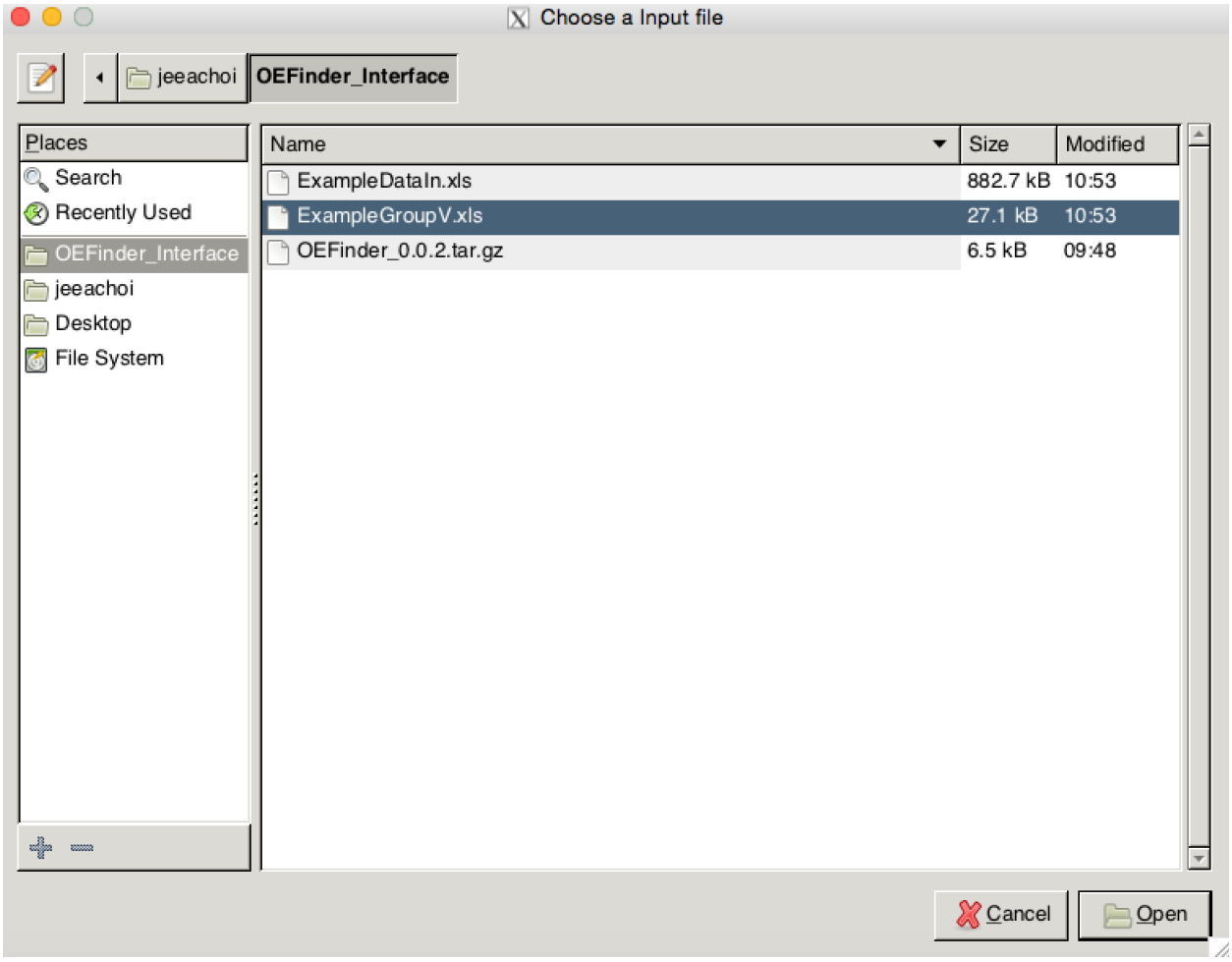


To select the input file, click the upper right “Open” button. A window will pop up and ask for an input file (shown on the next page).



Select ExampleDataIn.xls and click open. Then you will return to the interface window.

If the grouping vector file is not available, you may leave the second box empty. If it is available, click the open button next to the second box. Another window will pop up. Select ExampleGroupV.xls and click open.



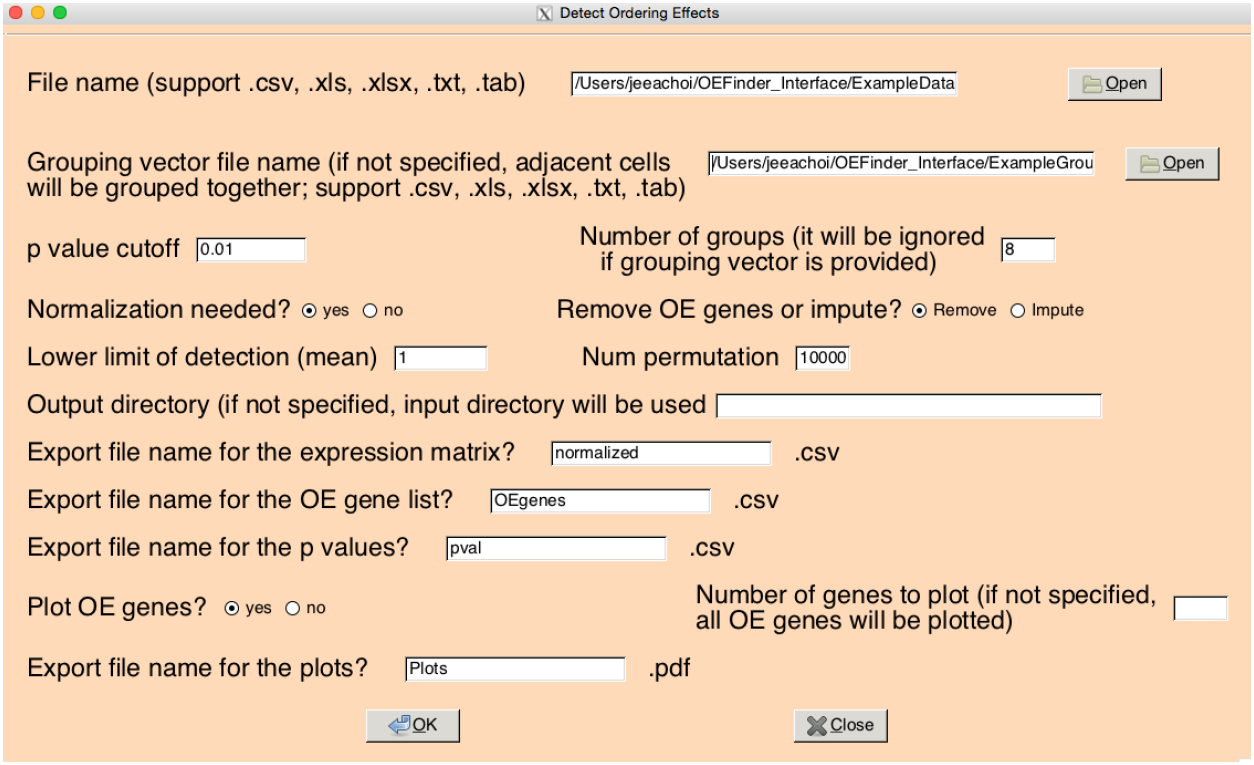
To run the analysis with default settings, the user may click the 'OK' button in the main interface.

A user can customize:

i. The gene expression input file.
ii. The grouping vector input file.
iii. OE cutoff *p*-value: default is 0.01.
iv. Number of groups: default is 8 even-sized groups. Cells will be grouped based on their order in the expression input file. This parameter will be ignored if the grouping vector input file is provided.
v. Normalization: whether normalization is needed. Please select 'Yes' if the input matrix is unnormalized.
vi. Handling OE genes: either remove from the output expression file or impute with adjusted values.
vii. Low limit of detection (mean); if the mean normalized gene expression is less than M, the gene will be removed prior to analysis. Default is 1.

viii. Number of permutation: number of permutation to perform when calculating the permutation p values. Default is 10000.
ix. Output directory. If it is empty, in shiny implementation the default is your home directory ("~/"); in RGtk2 implementation the default is the input directory.
x. Output file name for the normalized expression, with OE genes removed or adjusted.
xi. Output file name for the detected OE genes.
xii. Output file name for the OEFinder permutation p values.
xiii. Whether plot OE genes or not.
xiv. Number of top OE genes to plot. If it is not specified, all OE genes will be included in the output plots.
xv. Output file name for the OE gene plots.

Note: OEFinder applies the Median-by-Ratio normalization method introduced in Anders and Huber (2010) prior to OE detection when it is needed.



Outputs

Three (Four) files will be generated:

(1) normalized.csv: Normalized expression matrix with genes in row and cells in column. OE genes are excluded if you choose 'Remove' option.
(2) OEgenes.csv: OE genes are listed with their corresponding *p*-values. Only detected OE genes are included.
(3) pval.csv: *p*-values of all genes. Genes are sorted by their *p*-values.
(4) Plots.pdf: This file will be generated only when the user chooses to output plots of OE genes. In each plot, x-axis shows cells and y-axis shows normalized expression. Genes are sorted by their p-values.

3. Problem shooting

More details of the OEFinder implementation can be found at

https://github.com/lengning/OEFinder

If you have additional questions not addressed in this manual regarding the OEFinder interface, please contact us at nleng@morgridge.org.

Reference

Leng, N., J. Choi, L-F. Chu, J.A. Thomson, C. Kendziorski, and R.M. Stewart. (2015). OEFinder: A user interface to identify and visualize ordering effects in single-cell RNA-seq data, *submitted*

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biology, 11, R106.