User Guide

EasySeq

Index

I. What should your data look like?

II. Workflow

I. What should your data look like?

1. For Differential expression, GSEA, ssGSEA, GO:

- Gene Expression File: **.csv or .tsv or .xlsx or .xls file**(.csv or .tsv could be more stable). Column name(header) should be sample name and first column should always be gene id. (Fig. 1a). Gene id could be **Ensembl ID** or **EntrezID** or **Gene Symbol** or **Ensembl_Gene Symbol** .
- Metadata: **tab separated .txt file**, header should be "**sample**" and "**group**" (Case sensitive) (Fig. 1b)

a.

First column should always be gene id.

	A	B	C	D
1	gene_id	sample1	sample2	sample3
2	ENSMUSG00000000001_Gnai3	5224	6084	6313
3	ENSMUSG00000000003_Pbsn	0	0	0
4	ENSMUSG00000000028_Cdc45	1454	1766	1780
5	ENSMUSG00000000031_H19	0	0	0
6	ENSMUSG00000000037_Scml2	0	0	0

Sample name

b.

```
sample  group
sample1 group_A
sample2 group_A
sample3 group_B
```

Header should always be "sample" and "group"

Figure 1. Differential expression, GSEA, ssGSEA, GO data example.
(a) Expression matrix format (open in Excel).
(b) Metadata format (open in Windows Notepad).

2. For SingleCell:

- Path to filtered_feature_bc_matrix : Fill in a directory, best for filtered_feature_bc_matrix from 10X platform, but .rds or .h5 or counts matrix and metadata(.csv or .tsv) are also acceptable.
  - For **filtered_feature_bc_matrix**, target folder should contain 3 files : **barcodes.tsv.gz**, **features.tsv.gz** and **matrix.mtx.gz**
  - For **.rds** or **.h5**, target folder should contain **only 1** file, which will be your ***.rds** or ***.h5**
  - For **counts matrix** and **metadata**, target should contain **only 1 counts file** named ***_counts.csv** (or .tsv) and may contain **only 1 metadata file** named ***_metadata.csv** (or .tsv). Column name of count matrix should be cell barcode and first column should be gene id. (Fig. 2a). The first column of metadata should be the cell barcode, and the others are not specific. (Fig. 2b). If you need to apply an existing ident to counts data, directly change the column name to "**assigned.ident**". (Fig. 2b)

3. For Visium:

- Path to filtered_feature_bc_matrix : Fill in a directory, only filtered_feature_bc_matrix are accepted
- Path to spatial data : Fill in a directory, target folder should contain following files >>>
  - aligned_fiducials.jpg
  - detected_tissue_image.jpg
  - scalefactors_json.json
  - Thumbs.db
  - tissue_hires_image.png
  - tissue_lowres_image.png
  - tissue_positions_list.csv

a.

gene_id	AAACCCAGTCTTAGTG_V1	AAACCCAGTTGTGGCC_V1	AAACGAAAGGTGCATG_V1
OR4F5	0	0	0
OR4F29	0	0	0
OR4F16	0	0	0
SAMD11	0	0	2
NOC2L	1	0	0
KLHL17	0	0	0
PLEKHN1	0	0	0
PERM1	0	0	0
HES4	0	0	1
ISG15	0	0	0
AGRN	1	0	1
RNF223	0	0	0
C1orf159	0	0	0
TTLL10	0	0	0
TNFRSF18	0	0	0

b.

cell	orig.ident	nCount_RNA	nFeature_RNA	percent.mt	cluster_id_res0.3
AAACCCACACCTTCGT_V1	V1	1550	814	17.5483871	1
AAACCCAGTCTTAGTG_V1	V1	2000	943	0.45	1
AAACCCAGTTGTGGCC_V1	V1	14502	4084	4.261481175	0
AAACGAAAGGTGCATG_V1	V1	5287	1973	7.603555892	0
AAACGAAGTTGCATAC_V1	V1	1364	749	1.173020528	1
AAACGCTTCCTCACTG_V1	V1	1796	1038	12.08240535	2
AAAGGGCAGTATGACA_V1	V1	8489	3019	8.76428319	0
AAAGGGCCAGCAGTGA_V1	V1	1017	592	3.539823009	1
AAAGGTACATTGGGAG_V1	V1	10843	3161	2.914322604	0
AAAGGTAGTAGAGTTA_V1	V1	13434	3274	1.793955635	1

Change target column name to "assigned.ident"

cell	assigned.ident	nCount_RNA	nFeature_RNA	percent.mt	cluster_id_res0.3
AAACCCACACCTTCGT_V1	V1	1550	814	17.5483871	1
AAACCCAGTCTTAGTG_V1	V1	2000	943	0.45	1
AAACCCAGTTGTGGCC_V1	V1	14502	4084	4.261481175	0
AAACGAAAGGTGCATG_V1	V1	5287	1973	7.603555892	0
AAACGAAGTTGCATAC_V1	V1	1364	749	1.173020528	1
AAACGCTTCCTCACTG_V1	V1	1796	1038	12.08240535	2
AAAGGGCAGTATGACA_V1	V1	8489	3019	8.76428319	0
AAAGGGCCAGCAGTGA_V1	V1	1017	592	3.539823009	1
AAAGGTACATTGGGAG_V1	V1	10843	3161	2.914322604	0
AAAGGTAGTAGAGTTA_V1	V1	13434	3274	1.793955635	1

Figure 2. SingleCell counts matrix and metadata example.
(a) Counts matrix format of SingleCell (open in Excel).
(b) Metadata format of SingleCell and method of applying an existing ident to counts data (open in Excel).

4. For Multi-sample SingleCell:

- Path to Data List: **tab separated .txt file**, header should be "**sample_name**" and "**data_path**" (Case sensitive) (Fig. 3). "data_path" should be a directory to target folder.

```
sample_name     data_path
test1   C:\Users\Desktop\test\filtered_feature_bc_matrix
test2   C:\Users\Desktop\test\HNSCCxxx\filtered_feature_bc_matrix
```

Figure 3. Multi-sample SingleCell data list example (open in Windows Notepad).

5. Others:

- Target Gene File: **.csv or .tsv or .xlsx or .xls or tab separated .txt file** with only one column named **"id"** (Case sensitive). Only gene symbols are accepted (Fig. 4a). Expression barplot will be drown for each gene and correlation will be performed on every combination of target genes.
- Path to Target Gene Set: **.csv or .tsv or .xlsx or .xls or tab separated .txt file,** header should be your gene set name or gene name(a single gene can also be accepted, mainly used in SingleCell and Visium). One column will be treated as one gene set and only gene symbols are accepted (Fig. 4b).

a.

	A
1	id
2	SERPINE1
3	TGFBI
4	MMP10
5	LAMC2

b.

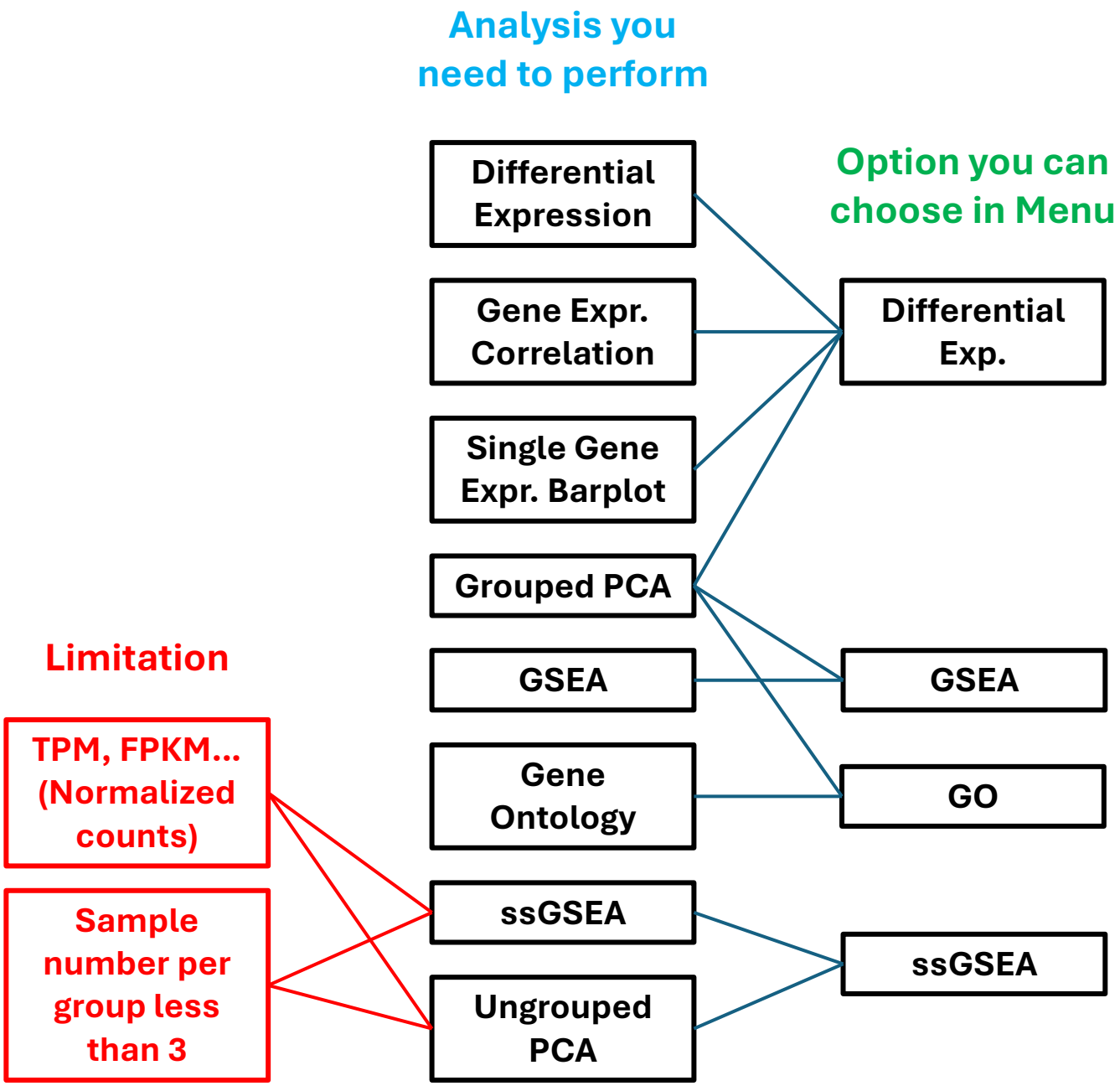
header

	A	B	C	D
1	your_gene_set1	your_gene_set2	ITGA5	BRCA1
2	SERPINE1	TP53	ITGA5	BRCA1
3	TGFBI	EGFR		
4	MMP10	BRCA1		
5	LAMC2	APOE		
6	P4HA2	CFTR		
7	PDPN	IL6		
8	ITGA5	VEGFA		
9	LAMA3	CDKN2A		
10	CDH13	SLC2A1		
11	TNC			
12	MMP2			
13	EMP3			
14	INHBA			
15	LAMB3			
16	VIM			

Figure 4. Target gene file and target gene list example.
(a) Target gene file (open in Excel).
(b) Target gene list (open in Excel).

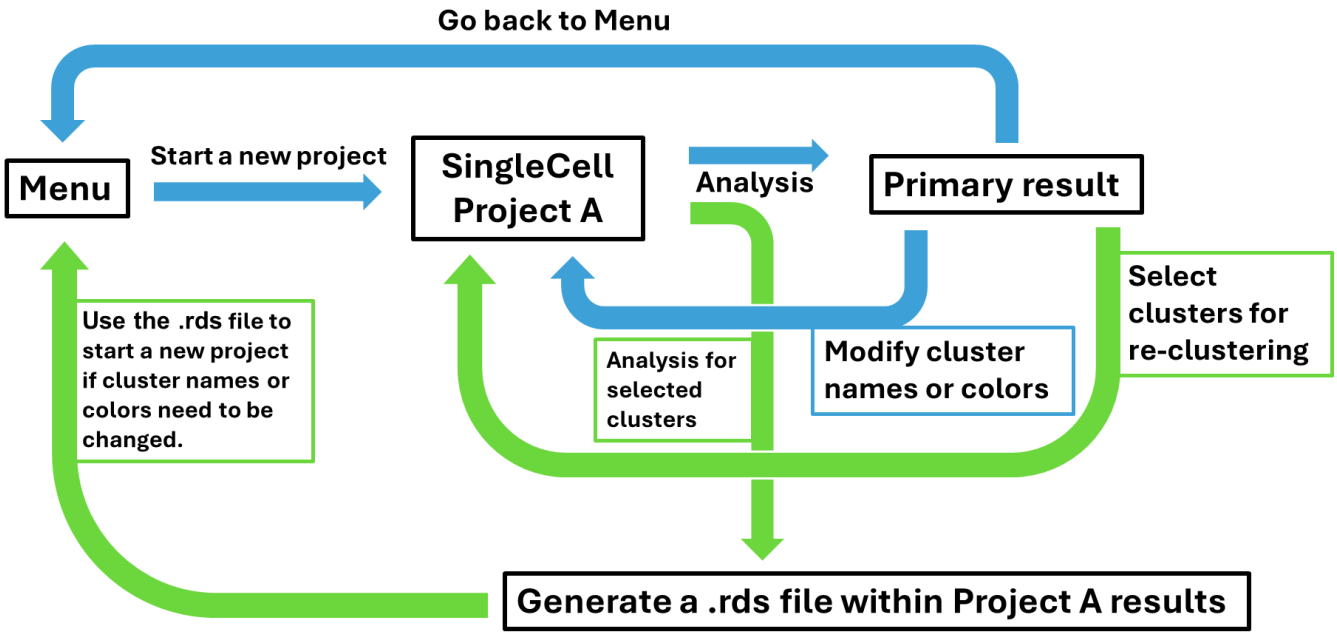
II. Workflow

1. Basic Analysis:



NOTE:
If your data are in TPM, FPKM, CPM format (or other normalized data), or the sample number per group is less than 3 (DESeq2 not accepted), you can only perform ssGSEA in this app.

2. SingleCell and Visium:



NOTE:
After the initial run, you will only be able to view the Primary Results (blue arrow). From there, you have three options:
(1) Accept Results and Return to Menu.
(2) Modify Colors or Cluster Names (for all clusters).
(3) Select Clusters for Re-clustering.

Important: Options 2 and 3 cannot be performed simultaneously. If you choose to modify colors or cluster names (Option 2), all clusters must be selected. If you select specific clusters for re-clustering (Option 3), you cannot change their colors or names.
If you need to modify the color or cluster name for selected clusters after re-clustering, return to the Menu and start a new project using the .rds file generated from previous project, which will be located in the folder named "Recluster for Selected Clusters".