Differential gene expression using DEasy

Diego Sierra R.

2025-02-05

DEasy!

`DEasy` is a shiny app created to make Differential gene expression (DGE) analyses easier, with `DEasy` is possible make DGE analysis from raw counts using `DESeq2` or `edgeR`, those are the most used and better documented packages to perform DGE analyses.

How to use it?

One of the advantages of DEasy is the you can run DE analyses without prior knowledge programming or without solid R bases, you just need internet connection and clicking on: DEasy!

Preparing the data   You can access to DEasy using web browser for free, this is useful if you want to use the app without using R, There are not major differences, essentially the performance its the same. in any case, first you have prepare the data:

To make a differential gene expression analysis is required having a set of samples and two groups to compare. This is the **Sample information file**. Take on mind the next advice:

- DEasy can take .xslx or .csv files

- The first column have to contain all your samples

- All columns must have the same length.

- The second and third (optional) columns must contain the groups that are going to be compared.

- The column names should not start with symbols

The next table shows an example where the 20 samples belong to two main groups according to their biological sex:

	A	B	C
1	ID_sample	SEX	DIET
2	Sample_FG2	Female	Vertebrate
3	Sample_FG4	Female	Vertebrate
4	Sample_Fmix1	Female	Mix
5	Sample_Fmix2	Female	Mix
6	Sample_Fmix3	Female	Mix
7	Sample_Fmix4	Female	Mix
8	Sample_FT1	Female	Invertebrate
9	Sample_FT2	Female	Invertebrate
10	Sample_FT3	Female	Invertebrate
11	Sample_FT4	Female	Invertebrate
12	Sample_MG2	Male	Vertebrate
13	Sample_MG3	Male	Vertebrate
14	Sample_MG4	Male	Vertebrate
15	Sample_Mmix2	Male	Mix
16	Sample_Mmix3	Male	Mix
17	Sample_Mmix4	Male	Mix
18	Sample_MT1	Male	Invertebrate
19	Sample_MT2	Male	Invertebrate
20	Sample_MT3	Male	Invertebrate
21	Sample_MT4	Male	Invertebrate

The second table required to perform differential expression analysis is the **Raw counts file,** a table with all the transcripts or genes that where found on each one of the samples, take on mind these advises:

- DEasy can take .xslx or .csv files

- The first column must contain all the gene or transcript names

- The number of samples must be equal to the number of samples on your **Sample information file**

- The ID related to each of the samples must be the same between **Sample information** file and the **Raw counts file**

- The counts must be raw counts, DEasy use specific normalizing method according to the package chosen internally (`DEseq2` or `edgeR`).

- All columns must have the same length.

The next table show the head of a **Raw counts file**

	A	B	C	D	E	F	G	H	I
1	Reference	Sample_FG2	Sample_FG4	Sample_Fmix1	Sample_Fmix2	Sample_Fmix3	Sample_Fmix4	Sample_FT1	Sample_
2	TRINITY_DN10022_c0_g1_i1	0	0	1	1	0	0	0	
3	TRINITY_DN10032_c0_g1_i1	6	14	123	9	18	11	2	
4	TRINITY_DN10055_c0_g1_i1	11	2	1	12	0	8	15	
5	TRINITY_DN10066_c0_g1_i1	2	1	3	7	0	6	0	
6	TRINITY_DN10100_c0_g1_i1	0	0	1	0	2	1	1	
7	TRINITY_DN10102_c0_g1_i1	1	9	30	12	0	12	5	
8	TRINITY_DN10111_c0_g1_i1	0	0	3	1	0	0	0	
9	TRINITY_DN10116_c0_g1_i1	2	4	3	16	9	29	14	
10	TRINITY_DN10121_c0_g1_i1	20	2	0	4	0	0	0	
11	TRINITY_DN10128_c0_g1_i1	4	1	10	6	1	14	0	
12	TRINITY_DN10149_c0_g1_i1	8	3	2	2	9	31	26	
13	TRINITY_DN10154_c0_g1_i1	0	1	3	2	0	1	2	
14	TRINITY_DN10171_c0_g1_i1	0	0	1	2	0	0	0	

You can get the example data and try yourself here: data example

DESeq2 performs an internal normalization procedure using the geometric mean, which is calculated for each gene across all samples. The counts for a gene in each sample is then divided by this mean: see more Documentation for DESeq2, on the other hand, edgeR use TMM method to compute normalization factors that represent sample-specific biases considering library size: see more Documentation.

The panels ~DEasy~ have sidebar panel where you can upload the **Raw counts file** and **Sample information file**. you can upload just *.xlsx* or *.csv* files. This app can deal with files with a maximum size of 500MB.

Once you provide the file with the correct format you can review your data on the preview tab panel, by default ~DEasy~ take the second column as the group (Treatments) that are going to be compared.

DEasy

Upload files

Sample imformation

Browse... | Samples_information.csv
Upload complete

Raw counts

Browse... | ara_counts.xlsx
Upload complete

Parameters

Differential expression package to use:

DESeq2 ▾

Treatment

SEX ▾

Add interaction (Optional)

None ▾

p-value

0,05 ⬍

Threshold

2 ⬍

Color palette

Set1 ▾

⌛ Run

Preview ▾ | Results | Plots | Help

Show 20 ⌄ entries

	SEX	DIET
Sample_FG2	Female	Vertebrate
Sample_FG4	Female	Vertebrate
Sample_Fmix1	Female	Mix
Sample_Fmix2	Female	Mix
Sample_Fmix3	Female	Mix
Sample_Fmix4	Female	Mix
Sample_FT1	Female	Invertebrate
Sample_FT2	Female	Invertebrate
Sample_FT3	Female	Invertebrate
Sample_FT4	Female	Invertebrate
Sample_MG2	Male	Vertebrate
Sample_MG3	Male	Vertebrate
Sample_MG4	Male	Vertebrate
Sample_Mmix2	Male	Mix

Showing 1 to 20 of 20 entries   Previous 1 Next

The next section are the parameters, you can select `DESeq2` or `edge2`, to performance the analysis, if you have a second group on your sample information file and you want to test the interaction between treatment, for example ~SEX + DIET you can select "Add interaction", by default DEasy chose "none" interaction, in other words It just considerate the the first treatment.

DEasy

Upload files

Sample imformation

Browse... | Samples_information.csv
Upload complete

Raw counts

Browse... | ara_counts.xlsx
Upload complete

Parameters

Differential expression package to use:

DESeq2 ▾

Treatment

SEX ▾

Add interaction (Optional)

None ▾

p-value

0,05 ⬍

Threshold

2 ⬍

Color palette

Set1 ▾

⌛ Run

Preview ▾ | Results | Plots | Help

Show 20 ⌄ entries

	Sample_FG2	Sample_FG4	Sample_Fmix1	Sample_Fmix2	Sample_Fmix3	Sample_Fmix4	Sample_FT1	Sample_FT2	Sample_FT3
TRINITY_DN10022_c0_g1_i1	0	0	1	1	0	0	0	0	1
TRINITY_DN10032_c0_g1_i1	6	14	123	9	18	11	2	15	21
TRINITY_DN10055_c0_g1_i1	11	2	1	12	0	8	15	2	1
TRINITY_DN10066_c0_g1_i1	2	1	3	7	0	6	0	3	13
TRINITY_DN10100_c0_g1_i1	0	0	1	0	2	1	1	0	4
TRINITY_DN10102_c0_g1_i1	1	9	30	12	0	12	5	25	43
TRINITY_DN10111_c0_g1_i1	0	0	3	1	0	0	0	0	2
TRINITY_DN10116_c0_g1_i1	2	4	3	16	9	29	14	16	31
TRINITY_DN10121_c0_g1_i1	20	2	0	4	0	0	0	0	8
TRINITY_DN10128_c0_g1_i1	4	1	10	6	1	14	0	13	29
TRINITY_DN10149_c0_g1_i1	8	3	2	2	9	31	26	9	16
TRINITY_DN10154_c0_g1_i1	0	1	3	2	0	1	2	3	0
TRINITY_DN10171_c0_g1_i1	0	0	1	2	0	0	0	0	0
TRINITY_DN10197_c0_g1_i1	3	0	2	2	0	2	0	3	8

Showing 1 to 20 of 52,863 entries   Previous 1 2 3 4 5 ... 2,644 Next

Then you can set an alfa value (0.05 p-value by default) and a threshold where you con, once you have defined the parameters you have to clink on the **Run** button and go the **Results** tab to watch the results, this process usually take some minutes so you have time to go for a coffee. Once the results main table is ready you can explore the pages and look for significance difference on each transcript, or you can download the table and use other tools like Microsoft Excel to explore the data.

Note: Shinny is a lazy web framework, so you have to click the **Results** panel after click on the **Run button** to start processing de data.

DEasy

Upload files

Sample imformation
Browse... | Samples_information.csv
Upload complete

Raw counts
Browse... | ara_counts.csv
Upload complete

Parameters

Differential expression package to use:
edgeR

Treatment
SEX

Add interaction (Optional)
None

p-value
0,05

Threshold
0

Color palette
Set1

⧖ Run

Preview ▾ Results Plots Help

Show 20 entries

	logFC	logCPM	F	PValue	FDR	Difference
TRINITY_DN10102_c0_g1_i1	0.146090798129588	1.87362724916002	0.147545441838039	0.704479092657714	0.868247464010991	Not significant
TRINITY_DN10116_c0_g1_i1	-0.662040060391329	1.57424694050511	2.79382371830687	0.108406273838577	0.445465135875192	Not significant
TRINITY_DN10128_c0_g1_i1	0.0987815621365912	1.23387852761789	0.0423075589903203	0.838874423505363	0.930571205985087	Not significant
TRINITY_DN10235_c0_g1_i1	-0.372470139326779	6.62432024827775	1.16926987685261	0.290944650328599	0.619554756886028	Not significant
TRINITY_DN10372_c0_g1_i1	-0.317250520869557	4.58478980657648	2.21569205060416	0.150423160119299	0.494605999436451	Not significant
TRINITY_DN10432_c0_g1_i1	0.220078692618645	7.37918738319185	0.743394610647002	0.397624015354547	0.688315201514528	Not significant
TRINITY_DN10455_c0_g1_i1	0.589819635684909	2.43272381948779	3.89175533367209	0.0608671262057408	0.349530670408242	Not significant
TRINITY_DN10628_c0_g1_i1	-0.215827516068989	2.30580056459168	0.263952353269434	0.612399876293697	0.814708554872297	Not significant
TRINITY_DN10944_c0_g1_i1	-1.27726639815896	3.72023908870835	1.41978273321032	0.24579093898937	0.584853698601025	Not significant

Showing 1 to 20 of 7,293 entries   Previous 1 2 3 4 5 ... 365 Next

⬇ Download

A progressing bar should appear indicating the progress.

DEasy

Upload files

Sample imformation
Browse... | Samples_information.csv
Upload complete

Raw counts
Browse... | ara_counts.csv
Upload complete

Parameters

Differential expression package to use:
DESeq2

Treatment
SEX

Add interaction (Optional)
None

p-value
0,05

Threshold
0

Color palette
Set1

⧖ Run

Preview ▾ Results Plots Help

Show 20 entries

	logFC	logCPM	F	PValue	FDR	Difference
TRINITY_DN10102_c0_g1_i1	0.146090798129588	1.87362724916002	0.147545441838039	0.704479092657714	0.868247464010991	Not significant
TRINITY_DN10116_c0_g1_i1	-0.662040060391329	1.57424694050511	2.79382371830687	0.108406273838577	0.445465135875192	Not significant
TRINITY_DN10128_c0_g1_i1	0.0987815621365912	1.23387852761789	0.0423075589903203	0.838874423505363	0.930571205985087	Not significant
TRINITY_DN10235_c0_g1_i1	-0.372470139326779	6.62432024827775	1.16926987685261	0.290944650328599	0.619554756886028	Not significant
TRINITY_DN10372_c0_g1_i1	-0.317250520869557	4.58478980657648	2.21569205060416	0.150423160119299	0.494605999436451	Not significant
TRINITY_DN10432_c0_g1_i1	0.220078692618645	7.37918738319185	0.743394610647002	0.397624015354547	0.688315201514528	Not significant
TRINITY_DN10455_c0_g1_i1	0.589819635684909	2.43272381948779	3.89175533367209	0.0608671262057408	0.349530670408242	Not significant
TRINITY_DN10628_c0_g1_i1	-0.215827516068989	2.30580056459168	0.263952353269434	0.612399876293697	0.814708554872297	Not significant
TRINITY_DN10944_c0_g1_i1	-1.27726639815896	3.72023908870835	1.41978273321032	0.24579093898937	0.584853698601025	Not significant

Showing 1 to 20 of 7,293 entries   Previous 1 2 3 4 5 ... 365 Next

⬇ Download

Processing... ✕

The last component on this app is the **Plot** tab section, the first plot its a Multidimentional scaling plot

The second plot its a heatmap, that shows you how similar some samples according their differential gene expression, The clusting method behind this heatmap is Ward.d2 to find compact and spherical clusters.

The third plot it's a MA-plot where you can visualize how much differentially expressed are your genes between your treatments, and their proportion again the number of genes who are not considered differentialy expressed.
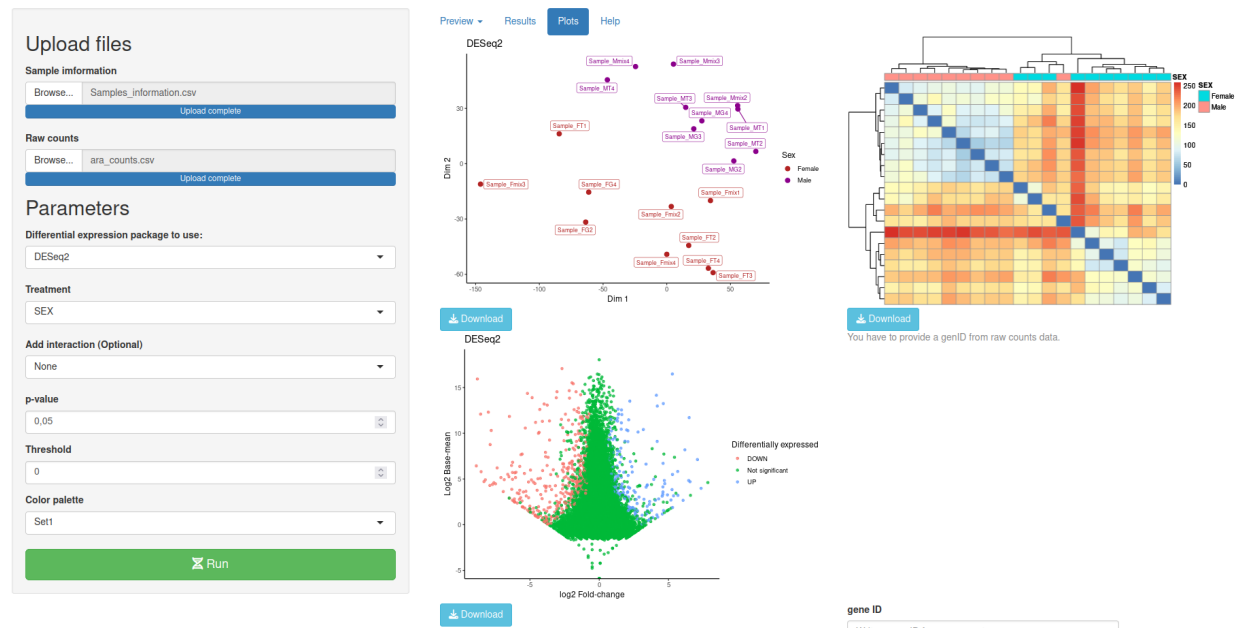
Once you have found an interesting gene or transcript you can use the last tool on the left bottom corner to generate a basic plot where you can visualize differences on a specific gene.

That's how `DEasy` can help you to explore and analyze data from Differential gene expression experiments, if you want more information you can check the GitHub repository here: DEasy GitHub, also you can run this example using data from *tests* folder on GitHub.

I hope you can find `DEasy` useful.