

蛋白质数据绘制火山图

2024-04-10

LiChuang Huang



@ 立效研究院

Contents

1	摘要	1
2	前言	1
3	材料和方法	1
3.1	材料	1
3.2	方法	1
4	分析结果	1
5	结论	1
6	附：分析流程	1
6.1	差异分析和火山图绘制	1
6.1.1	Model vs Control	1
6.1.2	Other Group	3
	Reference	4

List of Figures

1	PRO Model vs Control DEPs	2
---	-------------------------------------	---

List of Tables

1	PRO data Model vs Control DEPs	3
2	RAW quantification	4

1 摘要

需求：蛋白质数据差异分析和火山图绘制

结果：

- Model vs Control 火山图见 Fig. 1，对应数据见 Tab. 1
- 除了 Model vs Control, 其他组别数据也已提供，见 6.1.2

注：分析所使用的数据来源于 <http://101.66.242.136:5212/s/r7MTk?path=%2F> 中的 ‘txt/peptides.txt’

2 前言

3 材料和方法

3.1 材料

3.2 方法

Mainly used method:

- R package Limma and edgeR used for differential expression analysis^{1,2}.
- R version 4.3.2 (2023-10-31); Other R packages (eg., dplyr and ggplot2) used for statistic analysis or data visualization.

4 分析结果

5 结论

6 附：分析流程

6.1 差异分析和火山图绘制

6.1.1 Model vs Control

Figure 1 (下方图) 为图 PRO Model vs Control DEPs 概览。

(对应文件为 Figure+Table/PRO-Model-vs-Control-DEPs.pdf)

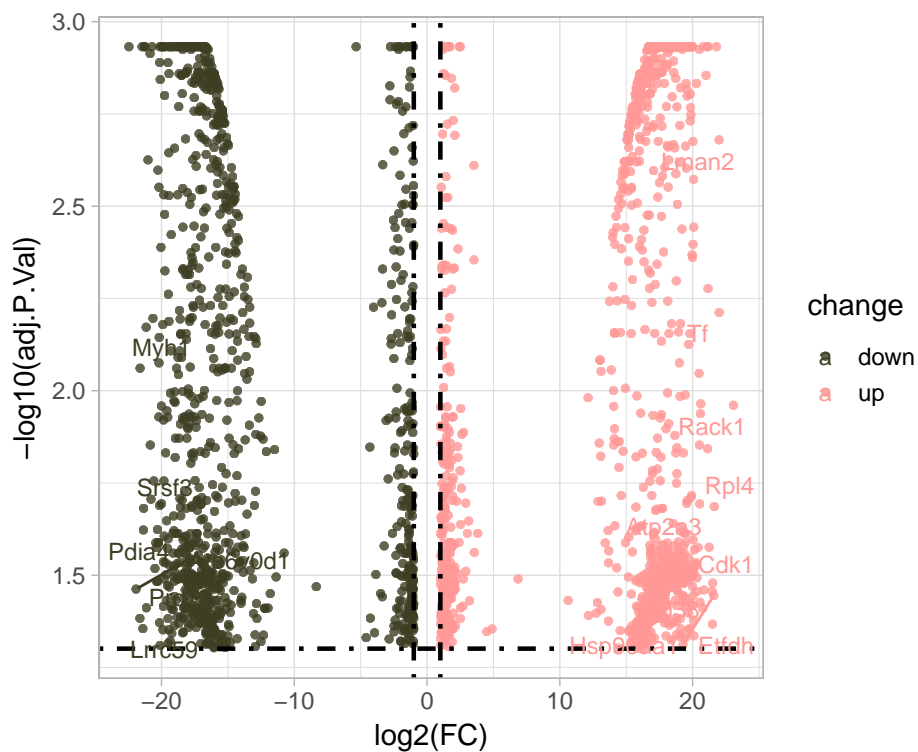


Figure 1: PRO Model vs Control DEPs

adj.P.Val cut-off :

0.05

Log2(FC) cut-off :

1

(上述信息框内容已保存至 Figure+Table/PRO-Model-vs-Control-DEPs-content)

Table 1 (下方表格) 为表格 PRO data Model vs Control DEPs 概览。

(对应文件为 Figure+Table/PRO-data-Model-vs-Control-DEPs.csv)

注：表格共有 2131 行 11 列，以下预览的表格可能省略部分数据；含有 1221 个唯一 'gene_name'。

1. logFC: estimate of the log2-fold-change corresponding to the effect or contrast (for ‘topTableF’ there may be several columns of log-fold-changes)
2. AveExpr: average log2-expression for the probe over all arrays and channels, same as ‘Amean’ in the ‘MarrayLM’ object
3. t: moderated t-statistic (omitted for ‘topTableF’)
4. P.Value: raw p-value
5. B: log-odds that the gene is differentially expressed (omitted for ‘topTreat’)
6. gene_name: GENCODE gene name

Table 1: PRO data Model vs Control DEPs

rownames	id	gene_name	unipro...	Proteins	logFC	AveExpr	t	P.Value	adj.P.Val
27560	27559	Myh1	Q5SX40	sp Q5S...	-2.136...	7.3178...	-21.38...	2.3613...	0.0011...
19391	19390	Ighm	A0A075...	tr A0A...	-2.009...	7.4110...	-19.35...	4.5213...	0.0011...
8591	8590	Hmgb1	A0A0J9...	tr A0A...	2.4353...	7.7990...	16.608...	1.2218...	0.0011...
25892	25891	Hspa5	Q3U9G2	tr Q3U...	-1.099...	8.8830...	-17.93...	7.4249...	0.0011...
534	533	Lmnbl	P14733	sp P14...	1.7203...	6.2488...	15.712...	1.7491...	0.0011...
3609	3608	Mmp9	Q3TTU7	tr Q3T...	2.4949...	5.5360...	14.767...	2.6098...	0.0011...
11691	11690	Ca3	P16015	sp P16...	-2.389...	5.8003...	-14.52...	2.8998...	0.0011...
31046	31045	Hbbt1	A8DUP7	tr A8D...	-5.345...	4.7019...	-14.37...	3.1058...	0.0011...
12790	12789	Rps14	Q3UJS5	tr Q3U...	1.1565...	7.2399...	14.517...	2.9129...	0.0011...
14192	14191	Lcp1	Q3U9M7	tr Q3U...	1.6660...	6.4834...	14.031...	3.6264...	0.0011...
5699	5698	Ceacam1	Q3LFS5	tr Q3L...	-1.532...	6.4558...	-13.90...	3.8427...	0.0011...
8092	8091	Tf	Q921I1	sp Q92...	-1.233...	8.0005...	-14.77...	2.5981...	0.0011...
16104	16103	Atp2a1	Q8R429	sp Q8R...	-1.601...	7.8387...	-13.95...	3.7611...	0.0011...
35	34	Alb	P07724	sp P07...	-1.155...	8.6863...	-15.00...	2.3523...	0.0011...
16588	16587	Rcc2	Q8BK67	sp Q8B...	1.2754...	6.6191...	13.589...	4.4527...	0.0011...
...

6.1.2 Other Group

‘All volcano plots’ 数据已全部提供。

(对应文件为 Figure+Table/All-volcano-plots)

注：文件夹 Figure+Table/All-volcano-plots 共包含 6 个文件。

1. 1_Model - Control.pdf
2. 2_Low - Model.pdf
3. 3_Middle - Model.pdf
4. 4_High - Model.pdf
5. 5_High - Middle.pdf
6. ...

‘ALL data DEPs’ 数据已全部提供。

(对应文件为 **Figure+Table/ALL-data-DEPs**)

注：文件夹 Figure+Table/ALL-data-DEPs 共包含 6 个文件。

1. 1_Model - Control.csv
2. 2_Low - Model.csv
3. 3_Middle - Model.csv
4. 4_High - Model.csv
5. 5_High - Middle.csv
6. ...

Table 2 (下方表格) 为表格 RAW quantification 概览。

(对应文件为 **Figure+Table/RAW-quantification.csv**)

注：表格共有 33174 行 27 列，以下预览的表格可能省略部分数据；含有 10461 个唯一 ‘Proteins’。

Table 2: RAW quantification

Proteins	Mass	id	Intensity	Intens.....5	Intens.....6	Intens.....7	Intens.....8	Intens.....9	...
tr Q6Z...	2031.9895	0	177160	0	0	0	31867	0	...
sp O70...	3186.471	1	35665	0	0	13434	22230	0	...
tr Q5J...	2047.0586	2	250350	0	67437	0	0	79893	...
sp A2A...	1888.8625	3	7268400	91029	140760	1236600	3638500	0	...
sp P63...	1520.6752	4	14545000	2117500	1889000	173690	747070	2088200	...
sp Q9R...	1943.9912	5	38486	0	0	7531.9	10726	0	...
sp P97...	1503.7205	6	347290	18450	1878.3	80010	188040	19633	...
tr Q3V...	1313.6939	7	56784	0	0	0	31003	2961.5	...
tr Q8C...	1735.8853	8	1491800	43990	19935	93206	68419	228030	...
tr Q0P...	970.49444	9	295650	0	0	65125	101240	0	...
sp Q9Z...	2716.4119	10	366870	64761	19275	50407	53458	10873	...
sp A2A...	1051.6026	11	2913000	66586	63995	527650	1114700	1594.5	...
tr A0A...	1415.762	12	699510	54119	32114	71040	81738	0	...
sp O70...	936.51411	13	1152200	50337	56782	298300	406900	61515	...
tr Z4Y...	1731.8614	14	42505	0	0	0	30772	9345.3	...
...

Reference

1. Ritchie, M. E. *et al.* Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47 (2015).

2. Chen, Y., McCarthy, D., Ritchie, M., Robinson, M. & Smyth, G. EdgeR: Differential analysis of sequence read count data user's guide. 119.