

陈云杰测序数据分析 十十

2023-12-14

LiChuang Huang



@ 立效研究院

Contents

1 摘要	1
1.1 差异基因分析	1
1.2 进一步分析	1
2 前言	1
3 材料和方法	1
3.1 方法	1
4 分析结果	1
5 结论	1
6 附：分析流程	1
6.1 元数据	1
6.2 差异分析	2
6.2.1 火山图	2
6.2.2 差异基因	2
7 附：进一步分析	3
7.1 N-M vs M 组数据	3
7.2 N-M vs M 上调、Z-O-M vs N-M 上调、Z-K-M vs Z-O-M 下调	3
7.3 N-M vs M 上调、Z-O-M vs N-M 上调交集	4
7.3.1 KEGG	4
7.3.2 GO	6
Reference	7

List of Figures

1 Three sets intersection	4
2 Two sets KEGG enrichment	5
3 Two sets GO enrichment	6

List of Tables

1 Sample metadata	2
2 Source NM vs M data	3
3 Two sets KEGG enrichment data	5
4 Two sets GO enrichment data	7

1 摘要

1.1 差异基因分析

结果见 6.2

1.2 进一步分析

- 取 N-M vs M 上调、Z-O-M vs N-M 上调、Z-K-M vs Z-O-M 下调基因交集可视化。
 - 见 7.2
- 取 N-M vs M 上调、Z-O-M vs N-M 上调交集基因富集分析。
 - 见 7.3

2 前言

3 材料和方法

3.1 方法

Mainly used method:

- R package `ClusterProfiler` used for gene enrichment analysis.¹
- `Limma` and `edgeR` used for differential expression analysis.^{2,3}
- 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 元数据

Table 1 (下方表格) 为表格 sample metadata 概览。

(对应文件为 `Figure+Table/sample-metadata.csv`)

注：表格共有 12 行 2 列，以下预览的表格可能省略部分数据；表格含有 12 个唯一 ‘sample’。

1. sample: 样品名称
2. group: 分组名称

Table 1: Sample metadata

sample	group
Count_.M-1	M
Count_.M-2	M
Count_.M-3	M
Count_.N-M-1	N_M
Count_.N-M-2	N_M
Count_.N-M-3	N_M
Count_.Z-O-M-1	Z_O_M
Count_.Z-O-M-2	Z_O_M
Count_.Z-O-M-3	Z_O_M
Count_.Z-K-M-1	Z_K_M
Count_.Z-K-M-2	Z_K_M
Count_.Z-K-M-3	Z_K_M

6.2 差异分析

6.2.1 火山图

‘Volcano plot’ 数据已全部提供。

(对应文件为 **Figure+Table/Volcano-plot**)

注：文件夹 Figure+Table/Volcano-plot 共包含 3 个文件。

1. 1_Z_K_M - N_M.pdf
2. 2_Z_O_M - N_M.pdf
3. 3_Z_K_M - Z_O_M.pdf

6.2.2 差异基因

按照 $|\log_2(FC)| > 0.03$, P-value < 0.05 筛选 DEGs:

‘DEGs’ 数据已全部提供。

(对应文件为 **Figure+Table/DEGs**)

注：文件夹 Figure+Table/DEGs 共包含 3 个文件。

1. 1_Z_K_M - N_M.csv
2. 2_Z_O_M - N_M.csv
3. 3_Z_K_M - Z_O_M.csv

7 附：进一步分析

7.1 N-M vs M 组数据

注：N-M vs M 的数据取自测序公司的分析结果，并非上述分析 (6.2) 所得，方法上可能有差异。

N-M vs M 上调取： $|\log_2(\text{FC})|$ ($\log_2\text{FoldChange}$) > 0.03 , adjusted P-value (padj) < 0.05

Table 2 (下方表格) 为表格 Source NM vs M data 概览。

(对应文件为 `Figure+Table/Source-NM-vs-M-data.xlsx`)

注：表格共有 4324 行 14 列，以下预览的表格可能省略部分数据；表格含有 4324 个唯一 'gene_id'。

1. symbol: 基因或蛋白符号。
2. pvalue: 显著性 P。

Table 2: Source NM vs M data

gene_id	N-M_re...	M_read...	log2Fo...	pvalue	padj	N-M-1	N-M-2	N-M-3	M-1
ENSG00...	3365.97	168.56	4.3215	4.6753...	1.1787...	27.981...	26.642...	23.652...	1.2021...
ENSG00...	4199.47	92.13	5.5128	2.7715...	3.4937...	60.455...	42.834...	24.984...	0.8167...
ENSG00...	15.07	2400.45	-7.3158	3.9538...	3.3227...	0.0438...	0.0320...	0.0384...	12.583...
ENSG00...	147.49	3264.66	-4.4688	7.0628...	4.4515...	1.4207...	3.7049...	3.4416...	61.869...
ENSG00...	65955.39	5078.61	3.699	1.2619...	6.3627...	204.29...	291.06...	173.27...	19.678...
ENSG00...	5062.63	397.68	3.6705	2.825e-34	1.187e-30	19.294...	25.215...	16.449...	1.7147...
ENSG00...	10668.43	1186.39	3.1686	6.4503...	2.3231...	29.626...	47.622...	31.720...	3.9347...
ENSG00...	30.33	523.44	-4.1091	1.2134...	3.824e-25	0.3787...	0.3508...	0.3640...	6.2172...
ENSG00...	249.53	2683.41	-3.4276	2.3189...	6.4959...	1.7546...	1.6100...	2.2335...	14.101...
ENSG00...	63.41	1341.53	-4.4039	6.0081...	1.5147...	0.2705...	0.4045...	0.4096...	7.6234...
ENSG00...	72.37	1935.8	-4.7441	1.1484...	2.6321...	1.5756...	2.5555...	3.7379...	31.653...
ENSG00...	1902.56	153.47	3.6331	3.0773...	6.4651...	5.2568...	4.4495...	3.7893...	0.3906...
ENSG00...	471.51	4670.83	-3.3081	1.0975...	2.1284...	8.2448...	6.5732...	6.3058...	49.982...
ENSG00...	1439.36	113.34	3.6665	6.7856...	1.2219...	6.7058...	10.766...	5.2679...	0.6395...
ENSG00...	40.59	1629.34	-5.3345	3.5876...	6.0297...	0.2881...	1.2878...	2.4561...	33.419...
...

7.2 N-M vs M 上调、Z-O-M vs N-M 上调、Z-K-M vs Z-O-M 下调

Figure 1 (下方图) 为图 Three sets intersection 概览。

(对应文件为 `Figure+Table/Three-sets-intersection.pdf`)

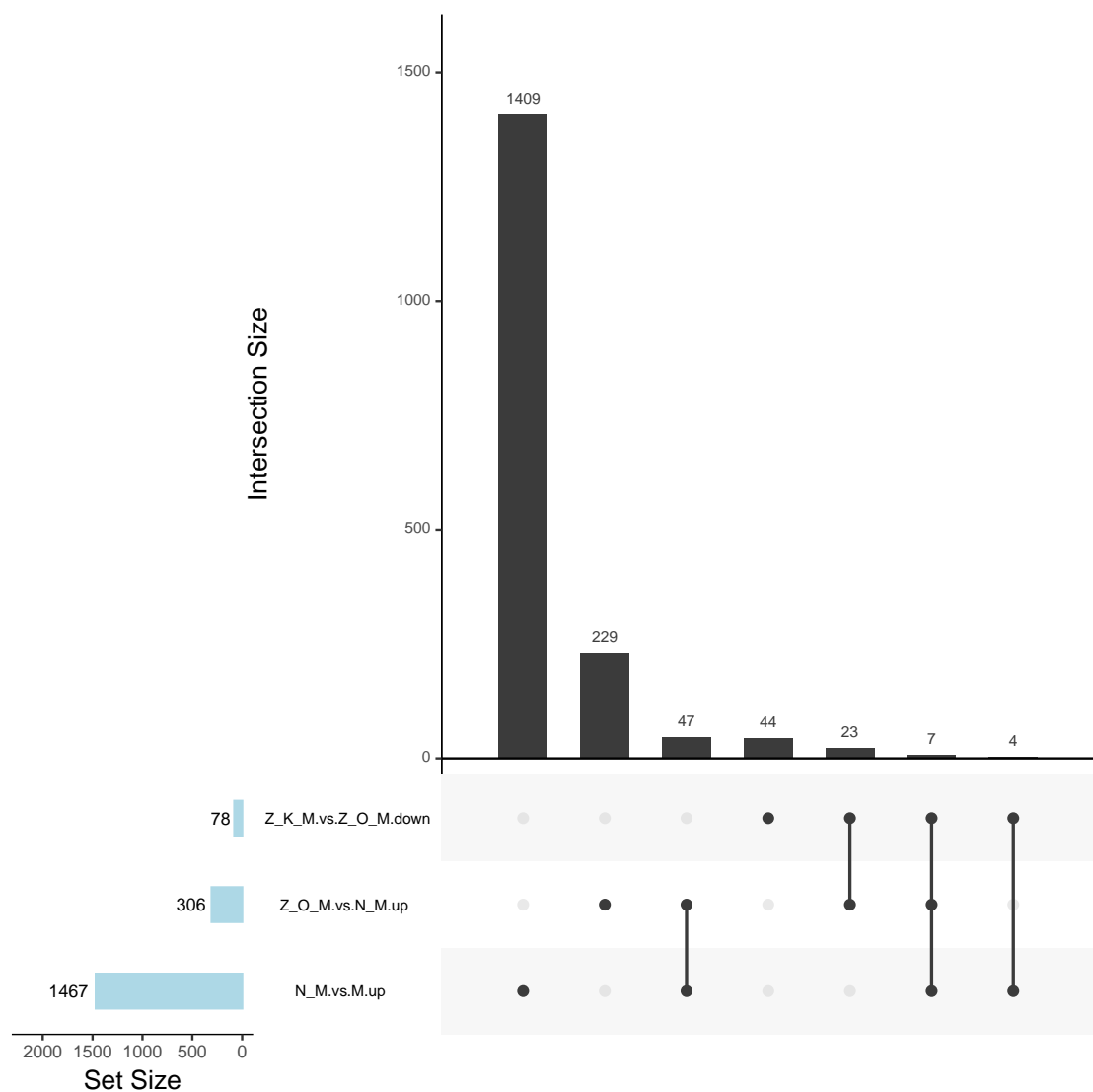


Figure 1: Three sets intersection

All_intersection :

POGLUT3, ALCAM, GUCY1B1, GCNT4, CTSO, CNTN3, ITGA8

(上述信息框内容已保存至 Figure+Table/Three-sets-intersection-content)

7.3 N-M vs M 上调、Z-O-M vs N-M 上调交集

7.3.1 KEGG

Figure 2 (下方图) 为图 Two sets KEGG enrichment 概览。

(对应文件为 Figure+Table/Two-sets-KEGG-enrichment.pdf)

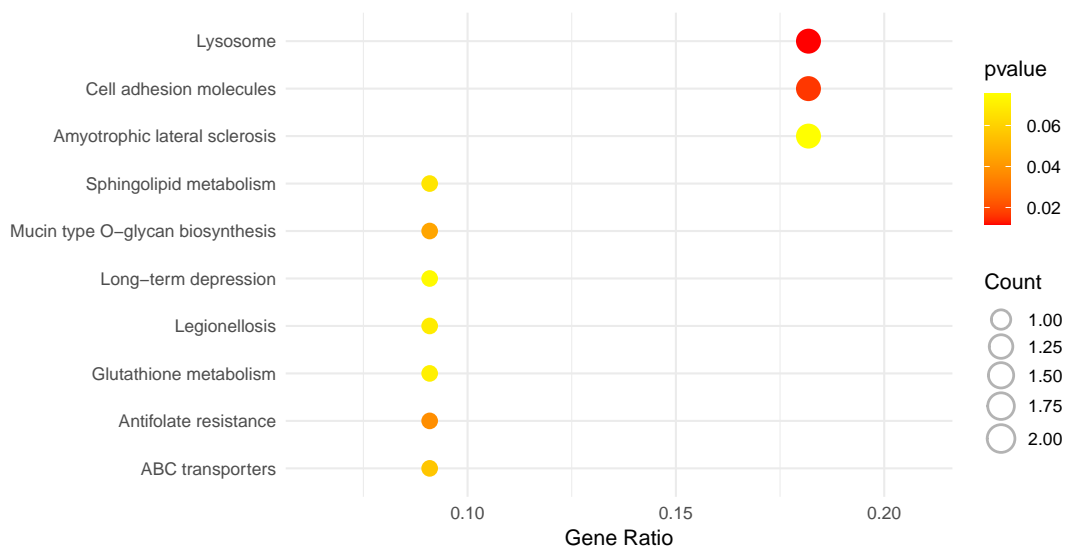


Figure 2: Two sets KEGG enrichment

Table 3 (下方表格) 为表格 Two sets KEGG enrichment data 概览。

(对应文件为 **Figure+Table/Two-sets-KEGG-enrichment-data.csv**)

注：表格共有 51 行 9 列，以下预览的表格可能省略部分数据；表格含有 51 个唯一 ‘ID’。

1. pvalue: 显著性 P。

Table 3: Two sets KEGG enrichment data

ID	Descri...	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
hsa04142	Lysosome	2/11	132/8644	0.0116...	0.2356...	0.2285...	1519/2581	2
hsa04514	Cell a...	2/11	158/8644	0.0163...	0.2356...	0.2285...	214/8516	2
hsa01523	Antifo...	1/11	30/8644	0.0375...	0.2356...	0.2285...	9429	1
hsa00512	Mucin ...	1/11	36/8644	0.0448...	0.2356...	0.2285...	51301	1
hsa02010	ABC tr...	1/11	45/8644	0.0558...	0.2356...	0.2285...	9429	1
hsa00600	Sphing...	1/11	54/8644	0.0666...	0.2356...	0.2285...	2581	1
hsa05134	Legion...	1/11	56/8644	0.0690...	0.2356...	0.2285...	834	1
hsa00480	Glutat...	1/11	57/8644	0.0702...	0.2356...	0.2285...	493869	1
hsa04730	Long-t...	1/11	60/8644	0.0737...	0.2356...	0.2285...	2983	1
hsa05014	Amyotr...	2/11	364/8644	0.0756...	0.2356...	0.2285...	834/49...	2
hsa00140	Steroi...	1/11	62/8644	0.0761...	0.2356...	0.2285...	1646	1
hsa04924	Renin ...	1/11	69/8644	0.0844...	0.2356...	0.2285...	2983	1
hsa05204	Chemic...	1/11	70/8644	0.0856...	0.2356...	0.2285...	1646	1
hsa04623	Cytoso...	1/11	75/8644	0.0914...	0.2356...	0.2285...	834	1
hsa04918	Thyroi...	1/11	75/8644	0.0914...	0.2356...	0.2285...	493869	1

ID	Descri...	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
...

7.3.2 GO

Figure 3 (下方图) 为图 Two sets GO enrichment 概览。

(对应文件为 [Figure+Table/Two-sets-GO-enrichment.pdf](#))

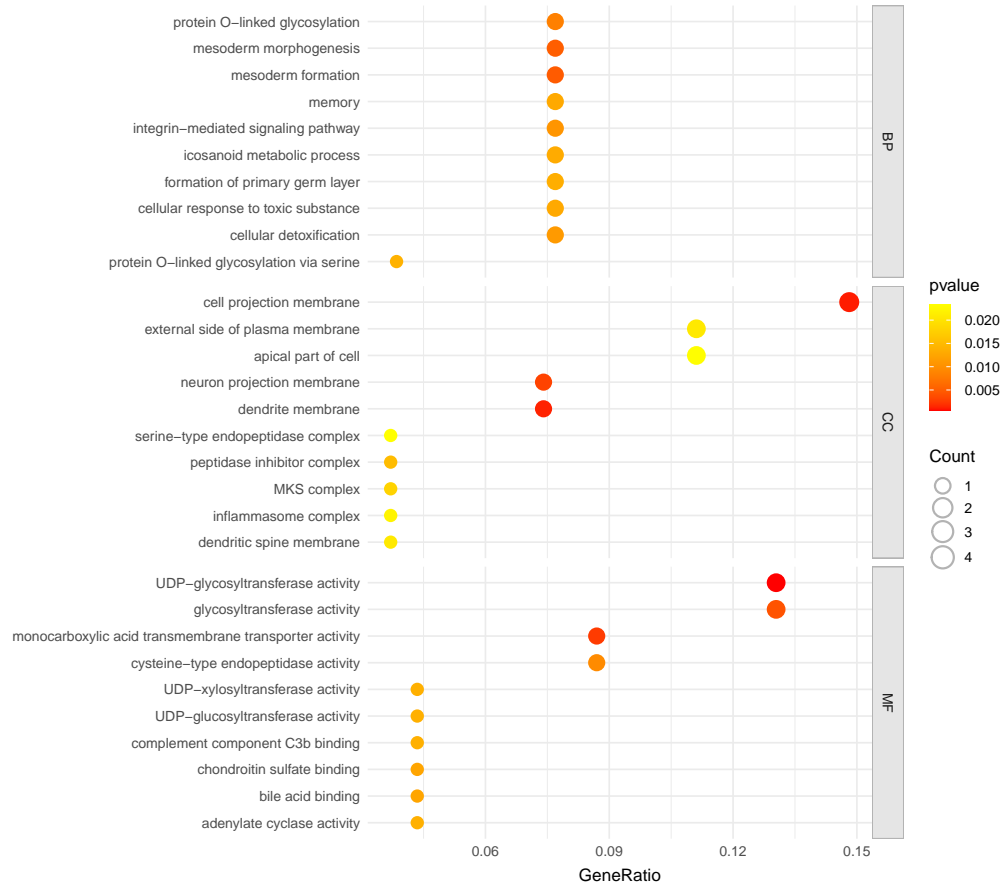


Figure 3: Two sets GO enrichment

Table 4 (下方表格) 为表格 Two sets GO enrichment data 概览。

(对应文件为 [Figure+Table/Two-sets-GO-enrichment-data.xlsx](#))

注：表格共有 717 行 10 列，以下预览的表格可能省略部分数据；表格含有 3 个唯一 ‘ont’。

1. pvalue: 显著性 P。
2. ont: One of "BP", "MF", and "CC" subontologies. The Cellular Component (CC), the Molecular Function (MF) and the Biological Process (BP).

Table 4: Two sets GO enrichment data

ont	ID	Descri...	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:000...	mesode...	2/26	75/18614	0.0048...	0.2203...	0.1821...	8516/6862	2
BP	GO:004...	mesode...	2/26	77/18614	0.0051...	0.2203...	0.1821...	8516/6862	2
BP	GO:000...	protei...	2/26	99/18614	0.0083...	0.2203...	0.1821...	51301/...	2
BP	GO:000...	integr...	2/26	112/18614	0.0106...	0.2203...	0.1821...	8516/5764	2
BP	GO:199...	cellul...	2/26	115/18614	0.0111...	0.2203...	0.1821...	9429/4...	2
BP	GO:000...	memory	2/26	124/18614	0.0128...	0.2203...	0.1821...	8516/5764	2
BP	GO:009...	cellul...	2/26	124/18614	0.0128...	0.2203...	0.1821...	9429/4...	2
BP	GO:000...	icosan...	2/26	125/18614	0.0130...	0.2203...	0.1821...	1646/834	2
BP	GO:000...	format...	2/26	127/18614	0.0134...	0.2203...	0.1821...	8516/6862	2
BP	GO:001...	protei...	1/26	10/18614	0.0138...	0.2203...	0.1821...	143888	1
BP	GO:003...	polyke...	1/26	10/18614	0.0138...	0.2203...	0.1821...	1646	1
BP	GO:003...	aminog...	1/26	10/18614	0.0138...	0.2203...	0.1821...	1646	1
BP	GO:004...	doxoru...	1/26	10/18614	0.0138...	0.2203...	0.1821...	1646	1
BP	GO:014...	dendri...	1/26	10/18614	0.0138...	0.2203...	0.1821...	5764	1
BP	GO:200...	positi...	1/26	10/18614	0.0138...	0.2203...	0.1821...	5764	1
...

Reference

1. Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
2. Ritchie, M. E. *et al.* Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47 (2015).
3. Chen, Y., McCarthy, D., Ritchie, M., Robinson, M. & Smyth, G. EdgeR: Differential analysis of sequence read count data user's guide. 119.