

Proteomics Service Report

Aksomics

2018-01-25

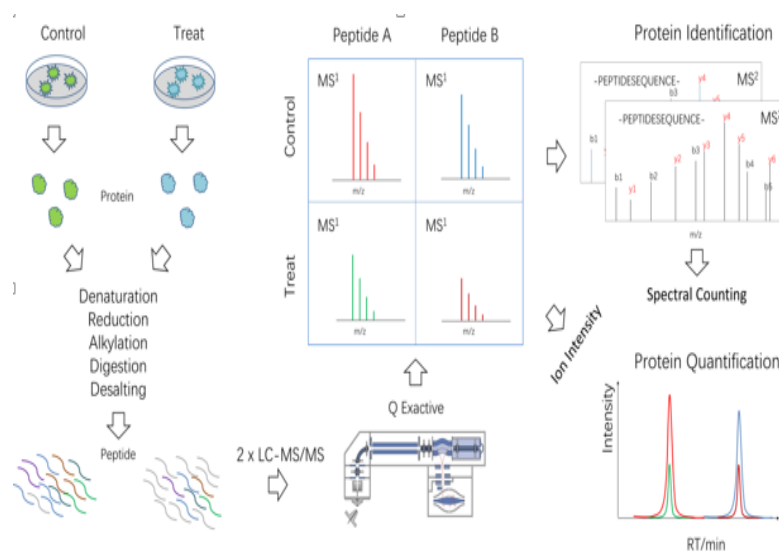
目录

1	简介	2
2	实验方法	4
2.1	仪器及试剂	4
2.2	实验流程	6
2.3	MaxQuant 分析和 LFQ 定量	9
2.4	统计检验和生物信息学分析	10
3	实验结果	10
3.1	定量与分组	10
3.2	蛋白丰度分布	11
3.3	定量重复性	11
3.4	差异倍数展示	13
4	生信分析	14
4.1	聚类分析	14
4.2	GO 分析	15
4.3	KEGG pathway 分析	17
4.4	蛋白互作	17

1 简介

采用稳定同位素标记定量的方法，能够精确的对蛋白表达水平进行定量分析。然而，标记定量的方法也存在某些不足之处：包括，额外的样品处理过程，昂贵的标记试剂，不完全标记导致样品复杂化，低丰度多肽难以检测，可标记样品数量有限等。作为替代，非标记定量技术（LFQ, label-free quantification）能够避免此类缺陷。LFQ 定量通常可采用两种方式：（1）谱图计数法（Spectral counting），使用多肽被鉴定次数来代表对应多肽的相对含量，进而计算出对应蛋白的相对表达量；（2）母离子强度（precursor ion intensity），提取鉴定到的多肽在一级质谱（MS1）中的色谱峰面积代表其相对含量，然后计算出对应蛋白的相对表达量。和标记定量技术不同在于：标记定量样品预先混合并同时 LC-MS/MS 检测，而 LFQ 技术要求每个样品独立进行制备和 LC-MS/MS 检测，因此准确性和通量弱于标记定量技术。

如下图:Label-free 定量技术原理



2 实验方法

2.1 仪器及试剂

- 重要仪器
 - Q Exactive (Thermo Scientific)
 - EASY-nLC1200 (Thermo Scientific)
 - Thermo mixer
 - Freeze dryer/vacuum concentrator
 - Microplate Reader
- 主要试剂
 - *Sigma-Aldrich (St. Louis, MO, USA)*
 - TCEP (tris (2-carboxyethyl) phosphine)
 - IAA (iodoacetamide)
 - SDC (sodium deoxycholate)
 - Tris.HCl
 - C18 columns (3M)
 - FA (Formic acid, LC-MS)
 - PMSF
 - Ammonium bicarbonate (ABC)
 - *J.T.Baker (PA, USA)*

- ACN (Acetonitrile, LC-MS)
 - H₂O (LC-MS)
- *Promega (Madison, WI, USA)*
 - Trypsin (sequence grade)
- *Sangon Biotech (Shanghai, China)*
 - NP-40 (Nonidet P 40)
 - NaCl
 - SDS
 - Acetone
- *Kangchen Bio-tech (Shanghai, China)*
 - Protease inhibitor cocktail
- *RIPA Buffer (modified)*
 - Contents: 25 mM Tris • HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 1% SDS
- *ABC-SDC (-20°C)*
 - 100mM ABC, 1% SDC, pH 8.5
- *Pierce BCA Protein Assay Kit*
 - A : B = 50 : 1, 160 μ L BCA + 20 μ L sample, 37°C incubate 30min, read at 562nm
 - For calibration curve, sequentially dilute BSA (2 mg/mL) to 5 concentration

- *TCEP (tris (2-carboxyethyl) phosphine)*
 - 1 M TCEP in H₂O, store at -80°C
- *IAA (iodoacetamide)*
 - 0.5 M IAA in H₂O, store at -80°C, away from light
- *Acetone*
 - Pre-chilled to -20°C
- *H₂O, formic acid, acetonitrile (HPLC grade)*
 - Buffer A: 0.1% FA H₂O, 2% ACN;
 - Buffer B: 0.1% FA, 70% ACN;

2.2 实验流程

2.2.1 样品裂解

RIPA 裂解液配制: 25 mM Tris • HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate:1% SDS

1. 将 RIPA 裂解液与蛋白酶抑制剂、PMSF 混合, 置于冰上预冷, 得到工作液。
2. 细胞样品加入 1000 L 工作液充分混匀, 冰浴超声 ~5 min, 至充分溶解。
3. 14000 g, 4°C 离心 15 min, 转移上清至新的 ep 管中。

2.2.2 BCA 定量

4. 按 50: 1 混合 regent A、B, 加至 96 孔板, 每孔 160 L, 5 个标准点, 一个空白。
5. 加入 20 L 样品 (稀释 5 – 10 倍) 或标准蛋白 (BSA, 5 个浓度梯度) 。

6. 37°C 震荡 30min, 用 562nm 波长检测吸光度。
7. 根据标准蛋白计算标准曲线, 并计算相应样品的蛋白浓度。

2.2.3 丙酮沉淀

8. 每个样品取 100 g 总蛋白, 并用裂解液稀释至 ~1mg/mL。
9. Acetone 预冷至 -20°C, 加入 4 – 6 倍体积 acetone 到样品中, 于冰上震荡 30 min, 沉淀蛋白 (或者 -20°C O/N)。
10. 10,000 g 转速, 4°C 离心 10 min, 小心去除上清。
11. 加入 200 L 预冷的 80% acetone 润洗沉淀 2 次。

2.2.4 蛋白重溶、还原、烷基化以及酶解

12. 加入 200 L 含 1% SDC 的 100mM ABC, 震荡混匀并简单离心。
13. 水浴超声 5 - 30 min 充分溶解蛋白沉淀。
14. 加入 TCEP 至 5mM, 55 °C 孵育 10 min, 还原二硫键。
15. 冷却样品至室温, 加入 IAA 至 10mM 避光反应 15 min, 烷基化已还原的二硫键。
16. 用 resuspension buffer 溶解 trypsin 至 0.5 g/ L, 室温孵育 5 min。
17. 加入 2 g trypsin 溶液至每个样品中, trypsin: 蛋白 = 1: 50。
18. 充分混匀, 简单离心后 37°C 震荡孵育过夜。

2.2.4.1 去除 SDC

19. 加入 TFA 至混合样品中 (终浓度 2%, pH < 2), 沉淀 SDC。

20. 高速离心 10 min, 转移上清至新的 ep 管中。
21. 加入 $n \times 100 \text{ L}$ 2% TFA 充分混匀, 提取共沉淀的多肽。
22. 重复提取 2 次。
23. 合并 3 个上清组分, 高速离心 10-20 min, 取上清至新 ep 管, 得多肽样品。

2.2.5 多肽脱盐

24. 配制 Buffer A: 0.1% FA, H_2O , 2% ACN; Buffer B: 0.1% FA, 70% ACN;
25. 加入 500 L ACN 平衡 C18 脱盐柱。
26. 加入 1 mL 含 0.1% FA 的 H_2O (Buffer A) 洗去残余的 ACN, 活化脱盐柱。
27. 将样品加至脱盐柱中, 低速离心并收集流出液, 重复过柱一次, 收集流出液 A。
28. 用 1 mL Buffer A 洗柱 2 次, 除去残余盐分。
29. 加入 400 L 0.1%FA, 70% ACN (Buffer B) 洗脱多肽, 收集流出液 B。
30. 用收集的流出液 A 重复整个脱盐过程一次 (25 - 29)。
31. 合并两次流出液 B, 真空干燥。
32. 加入 buffer A 重溶多肽至 $1 \text{ g} / \text{L}$ 用于 LC-MS/MS 检测, 或 -80°C 冻存。

2.2.6 nano-UPLC 分离

每个组分取 2 g 多肽经 nano-UPLC 液相系统 EASY-nLC1200 进行分离，并使用在线质谱仪 (Q-Exactive) 进行检测。分析采用 100 m ID × 15 cm 反相色谱柱 (Reprosil-Pur 120 C18-AQ, 1.9 μ m, Dr. Math)。流动相 A 液为 0.1% 甲酸乙腈水溶液 (乙腈为 2%)，B 液为 0.1% 甲酸乙腈水溶液 (乙腈为 80%)。色谱柱以 100% 的 A 液平衡。样品由自动进样器直接上样到色谱柱，再经色谱柱分离，流速 300 nL/min，梯度时长 120 min。流动相 B: 8 - 35% 持续 92 min, 35 - 45% 持续 20 min, 45 - 100% 持续 2 min, 100% 持续 2 min, 100 - 2% 持续 2min, 2% 持续 2min。

2.2.7 LC-MS/MS

质谱分析时长: 120 min/sample, 正离子检测模式, 母离子扫描范围: 350 - 1600 m/z。DDA 采集方式为: 每次全扫描 (full scan) 后采集 20 个碎片图谱 (MS2 scan, HCD)。MS1 在 m/z @200 时分辨率为 70,000, MS2 在 m/z @200 时分辨率为 17,500; MS1 AGC 为 3E+6, MS2 AGC 为 1E+5, 最大离子注入时间 (Max IT): MS1, 50 ms, MS2, 45ms。标准化碰撞能量 (NCE) 为 28 %, 隔离窗口为 2 m/z, 动态排除时间 40 s。

2.3 MaxQuant 分析和 LFQ 定量

原始.raw 文件使用 MaxQuant (1.5.6.0) 处理。蛋白数据库来自于 UNIPROT 数据库 (Uniprot_mouse_2016_09)。蛋白序列及其反向 decoy 序列同时用于 MaxQuant 搜库。定量类型为包含 match between run 的非标定量 (LFQ); Trypsin 设为特异性内切酶, 最大 3 个漏切位点; Oxidation [M] 和 Acetyl [protein N-term] 设为可变修饰, Carbamidomethyl [C] 为固定修饰, 最大可变修饰数 3。多肽和蛋白水平 FDR 均为 0.01, 未发生可变修饰的 unique 多肽用于定量。

随后对 4 个样品进行标准化: 性质相似的生物样品中绝大部分蛋白表达水平应该是不变的, 只有少数蛋白会出现差异表达; 根据这个原理, 将各组样品标准化, 使各组样品总蛋白或中位数一致 LFQ 定量结果经 log 转化后, 缺失值使用 Perseus 软件从正态分布随机抽样进行填补, 非缺失值数小于样品重复数 (replicates) 的蛋白 group 舍去。

2.4 统计检验和生物信息学分析

随后对标准化后的定量结果进行统计学分析，得到对应的差异表达蛋白。本次实验没有生物学重复，不进行统计学检验。将表达倍数差异 (ratio A/B > 2 or ratio A/B<1/2,) 的蛋白定义为差异显著，并进行后续 GO、KEGG pathway、蛋白互作分析和展示。

3 实验结果

3.1 定量与分组

根据标准曲线和吸光度 OD562nm 之间的线性关系计算得出相应样品的总蛋白量；Tag 与 Group ID 表示生物学分组情况。

表 1: 分组情况

Sample ID	Sample name
1	Blank.1
2	Model.1
3	Positive.1
4	Treat.1

本次实验多肽和蛋白水平 FDR 均控制在 0.01，差异倍数选择鉴定及定量结果统计如下：将表达倍数差异 (ratio A/B > 2 or ratio A/B<1/2,) 的蛋白定义为差异显著

附件: [all protein](#),[sig protein](#)

表 2: 蛋白数量概述

多肽总数	蛋白总数 (Group)	可定量蛋白数
18336	2594	2467

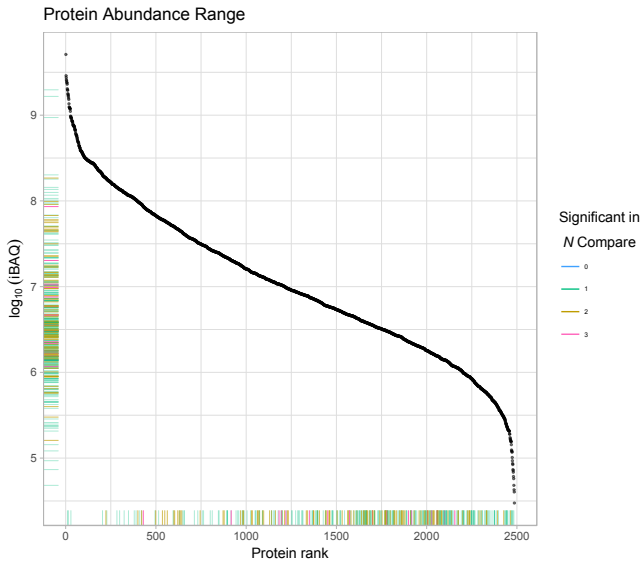
表 3: 差异蛋白数量

比较	数量
Blank-Model	180
Model-Positive	139
Positive-Treat	206

3.2 蛋白丰度分布

复杂生物样本中各种蛋白的表达丰度差异极大，例如在血浆中，不同蛋白间差异可达 10 个数量级，对蛋白的定量带来极大的挑战。本次定量动态范围超过 6 个数量级，y 轴表示蛋白绝对量 (iBAQ)，不同颜色表示该蛋白是否差异表达。[附件：Summary /rank](#)

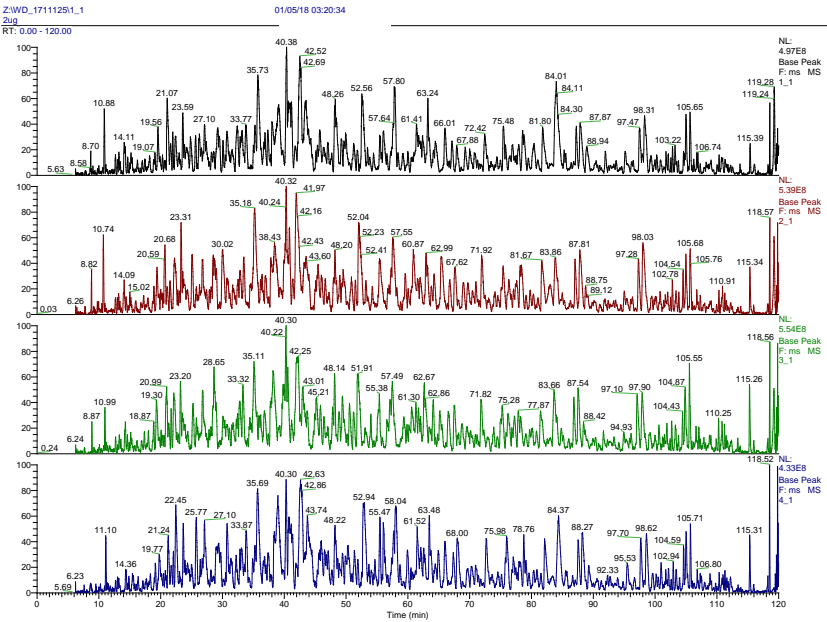
如下图: 蛋白丰度分布图



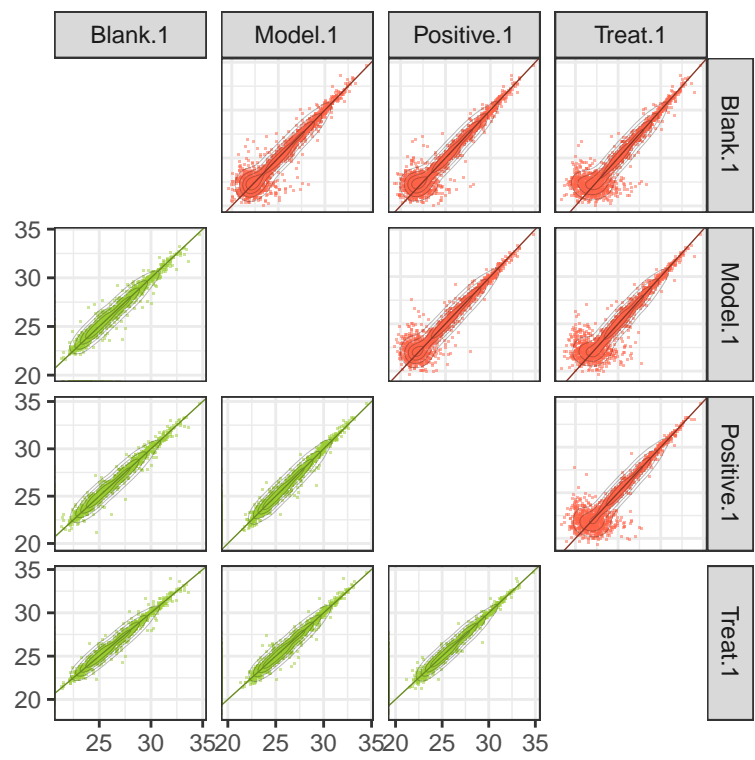
3.3 定量重复性

下图展示了 4 个样品总离子强度 (TIC) 及色谱保留时间 (RT) 的稳定性，TIC 和 RT 的稳定性是 label-free 定量可靠性的前提保证。

如下图：总离子流图 (TIC)



如下图：4 个样本间定量重复性分析

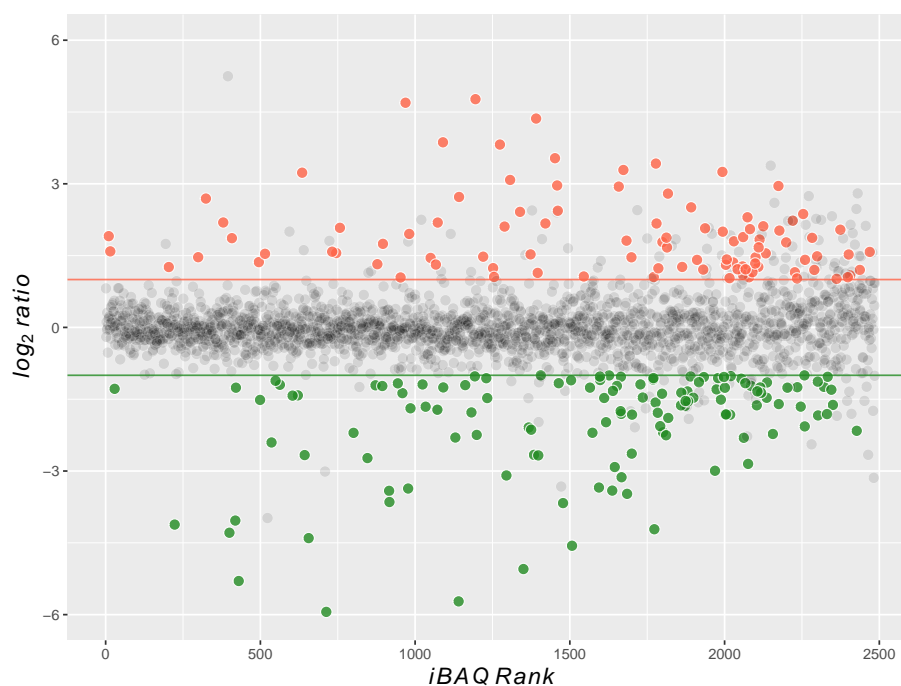


3.4 差异倍数展示

由于无生物学重复，无法进行火山图展示，只能进行差异倍数的展示。

附件: FC

如下图: Blank-Model 蛋白差异倍数展示。差异显著的蛋白分别以深绿色（下调）和橙红色（上调）表示。无统计学差异蛋白以淡绿色表示。

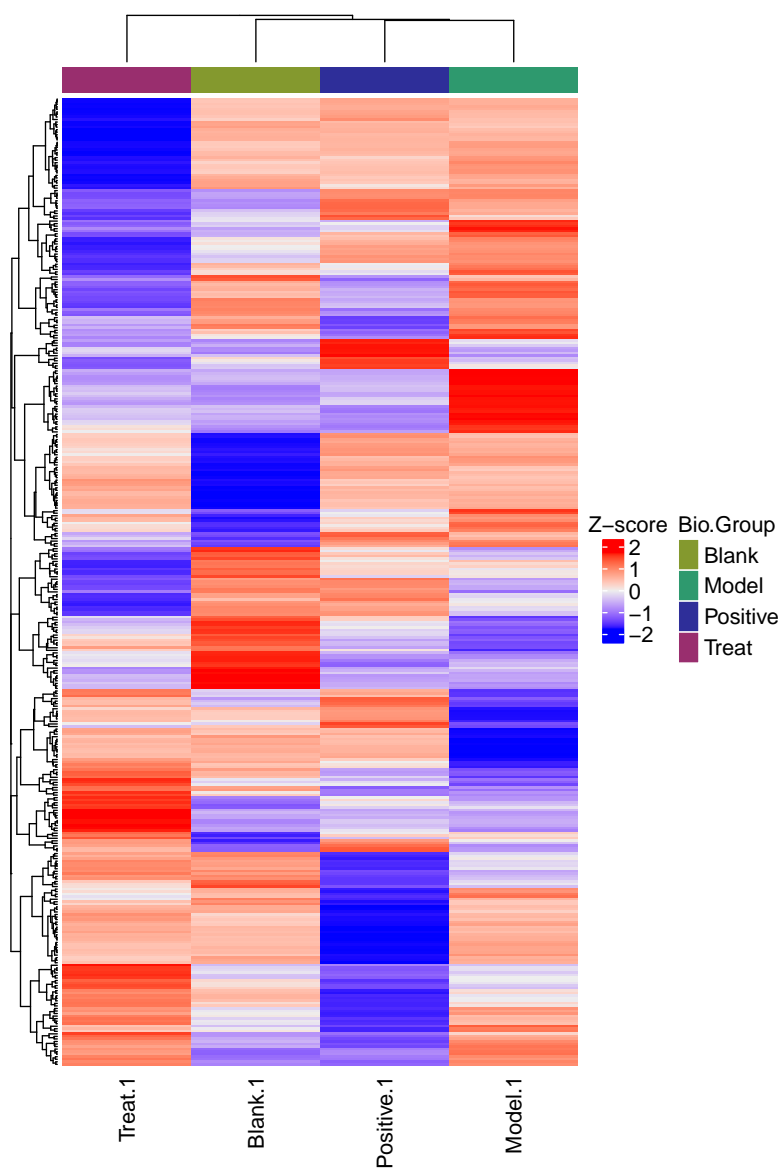


4 生信分析

4.1 聚类分析

为检验所选取的差异表达蛋白质或特征差异表达蛋白质的合理性和准确性,我们利用筛选出的蛋白质,对各组样本进行层次聚类。原始定量值先进行 Z-score 变换后再进行聚类分析和展示。(R-package: ComplexHeatmap)
附件: [cluster](#)

如下图: 差异蛋白层次聚类分析, 顶端颜色条代表样品分组情况(例如, 实验组、对照组, 对应图例: bio.group); 低端为对应样品名称; 右侧图例 Z-score 以颜色代表蛋白的相对表达水平; 右侧蛋白名对应于差异蛋白清单(significant.csv)中 symbol 列。



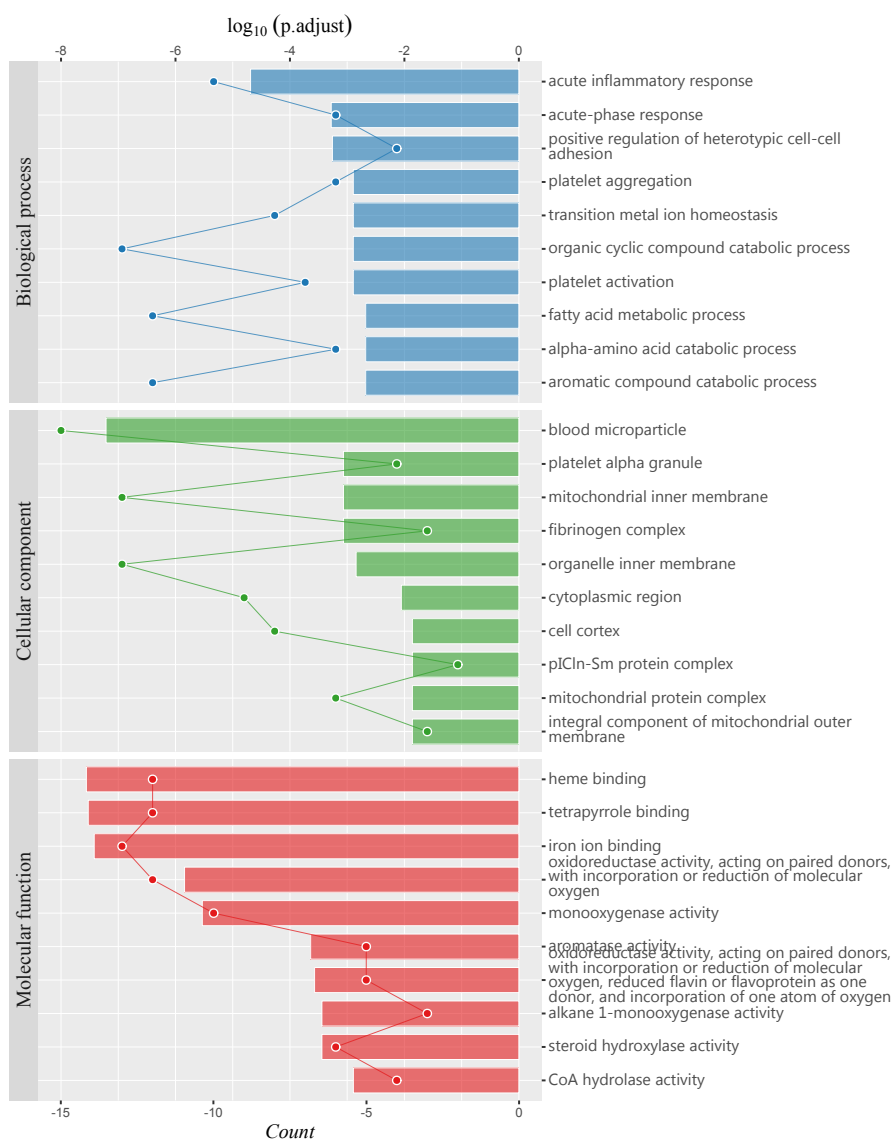
4.2 GO 分析

基因本体 (Gene Ontology) 是一个标准化的基因功能分类体系，提供了一套动态更新的标准化词汇表，并以此从三个方面描述生物体中基因和基因产物的属性：参与的生物过程 (Biological Process, BP)，分子功能 (Molecular Function, MF) 和细胞组分 (Cellular Component, CC)。我们

选择差异表达的蛋白进行 GO 富集分析，并以图片展示前 10 条 GO 条目。

(R-package: clusterProfiler) 附件: [enrichGO*.csv](#), [GO.tiff](#)

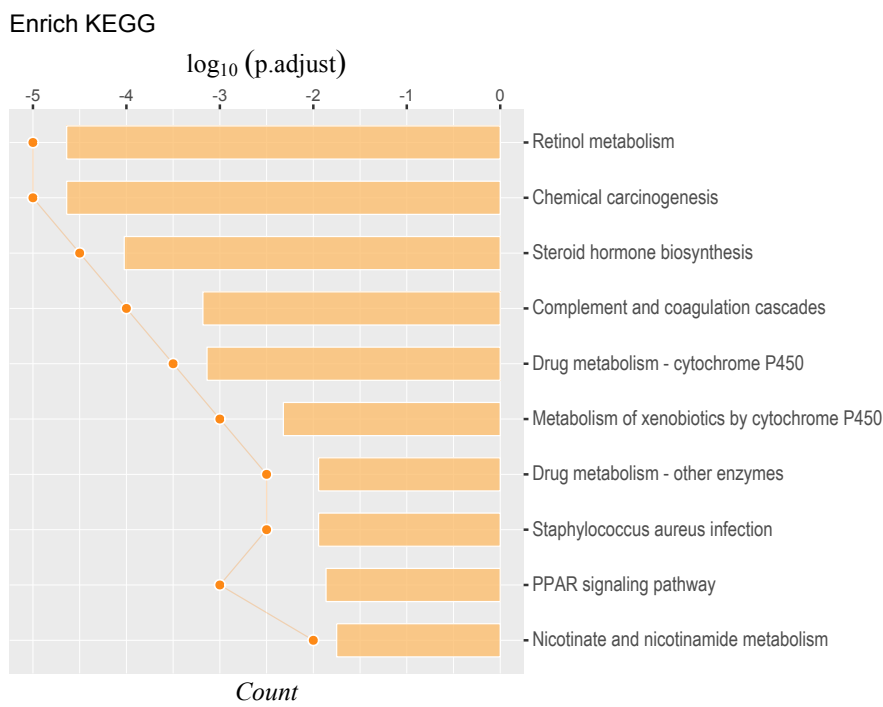
如下图:Blank-Model 差异蛋白 GO 富集分析。Bar: $-\log(p.adjust)$, point: Count。



4.3 KEGG pathway 分析

在生物体中，蛋白质并不独立行使其功能，而是不同蛋白质相互协调完成一系列生化反应以行使其生物学功能。因此，通路分析有助于更系统、全面地了解细胞的生物学过程、性状或疾病的发生机理、药物作用机制，等等。KEGG 是常用于通路研究的数据库之一。针对挑选出的差异表达蛋白质进行 KEGG 通路注释，分析并确定差异表达蛋白质参与的最主要的代谢和信号转导途径。和 GO 分析类似，我们选择差异表达的蛋白进行 KEGG 富集分析。[参考附件：KEGG](#)

如下图：Blank-Model 差异蛋白 KEGG 富集分析。Bar: $-\log(p.adjust)$, point: Count

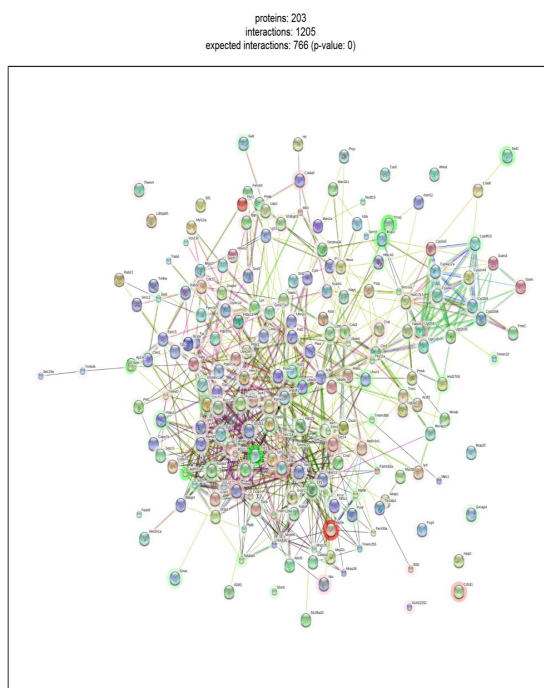


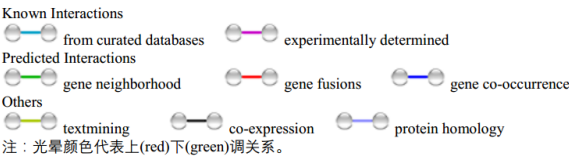
4.4 蛋白互作

蛋白互作网络是系统生物学的一个重要组成部分，可用于筛选和获取功能基因组数据，为研究蛋白质的结构、功能和进化等特征提供一个直观的分析平台。对蛋白互作网络进行预测能够为后期的生物学实验提供有价值的

参考。STRING 数据库整合多方面信息，包括实验数据，计算预测和文本挖掘等：from curated databases, experimentally determined; gene neighborhood, gene fusions, gene co-occurrence; text mining, co-expression, protein homology。 (R-package: STRINGdb) 附件： [PPI](#)

如下图:STRING 蛋白互作分析。运用 STRINGdb 对 Blank-Model 差异表达的蛋白进行蛋白互作分析发现，213 个基因间存在显著富集的相互作用，PPI enrichment p-value :0。Known Interactions





```
sessionInfo()
```

```
## R version 3.4.3 (2017-11-30)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 7 x64 (build 7601) Service Pack 1
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=Chinese (Simplified)_People's Republic of China.936
## [2] LC_CTYPE=Chinese (Simplified)_People's Republic of China.936
## [3] LC_MONETARY=Chinese (Simplified)_People's Republic of China.936
## [4] LC_NUMERIC=C
## [5] LC_TIME=Chinese (Simplified)_People's Republic of China.936
##
## attached base packages:
## [1] splines parallel stats graphics grDevices utils datasets
## [8] methods base
##
## other attached packages:
## [1] Cairo_1.5-9 ggrepel_0.7.0 ropis_1.10.0
## [4] dplyr_0.7.4 readr_1.1.1 bookdown_0.5
## [7] kableExtra_0.7.0 huxtable_2.0.0 knitr_1.18
## [10] siggenes_1.52.0 multtest_2.34.0 extrafont_0.17
## [13] RColorBrewer_1.1-2 stringi_1.1.6 gplots_3.0.1
## [16] ggplot2_2.2.1 reshape2_1.4.3 Biobase_2.38.0
## [19] BiocGenerics_0.24.0
##
## loaded via a namespace (and not attached):
## [1] gtools_3.5.0 lattice_0.20-35 colorspace_1.3-2
## [4] htmltools_0.3.6 stats4_3.4.3 viridisLite_0.2.0
## [7] yaml_2.1.16 survival_2.41-3 rlang_0.1.6
## [10] pillar_1.1.0 glue_1.2.0 bindrcpp_0.2
## [13] plyr_1.8.4 bindr_0.1 stringr_1.2.0
```

```
## [16] munsell_0.4.3      gtable_0.2.0      rvest_0.3.2
## [19] caTools_1.17.1     evaluate_0.10.1   highr_0.6
## [22] Rttf2pt1_1.3.5     Rcpp_0.12.14      KernSmooth_2.23-15
## [25] scales_0.5.0       backports_1.1.2   gdata_2.18.0
## [28] png_0.1-7          hms_0.4.0         digest_0.6.14
## [31] grid_3.4.3         rprojroot_1.3-2   tools_3.4.3
## [34] bitops_1.0-6       magrittr_1.5       lazyeval_0.2.1
## [37] tibble_1.4.1       rtticles_0.4.1    extrafontdb_1.0
## [40] pkgconfig_2.0.1    MASS_7.3-48       Matrix_1.2-12
## [43] xml2_1.1.1         assertthat_0.2.0  rmarkdown_1.8
## [46] httr_1.3.1         R6_2.2.2          compiler_3.4.3
```