|  |
| --- |
| 蛋白质组学服务分析报告 |
| Label-free定量蛋白质组学 |
|  |

委托单位：康成生物

合同编号：1711125

分析人员：黄强/苗豪菲

报告日期：2018-01-12

目录

[简介 3](#_Toc503516073)

[实验方法 3](#_Toc503516074)

[仪器及试剂 3](#_Toc503516075)

[重要仪器 3](#_Toc503516076)

[主要试剂 4](#_Toc503516077)

[实验流程 5](#_Toc503516078)

[样品裂解 5](#_Toc503516079)

[BCA定量 6](#_Toc503516080)

[丙酮沉淀 6](#_Toc503516081)

[蛋白重溶、还原、烷基化以及酶解 6](#_Toc503516082)

[去除SDC 7](#_Toc503516083)

[多肽脱盐 7](#_Toc503516084)

[nano-UPLC分离 7](#_Toc503516085)

[LC-MS/MS 8](#_Toc503516086)

[MaxQuant 分析和LFQ定量 8](#_Toc503516087)

[统计检验和生物信息学分析 8](#_Toc503516088)

[实验结果 8](#_Toc503516089)

[定量与分组 8](#_Toc503516090)

[蛋白丰度分布 9](#_Toc503516091)

[定量重复性 10](#_Toc503516092)

[差异倍数展示 12](#_Toc503516093)

[生信分析 15](#_Toc503516094)

[聚类分析 15](#_Toc503516095)

[GO分析 17](#_Toc503516096)

[KEGG pathway分析 20](#_Toc503516097)

[蛋白互作 21](#_Toc503516098)

# 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进行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。

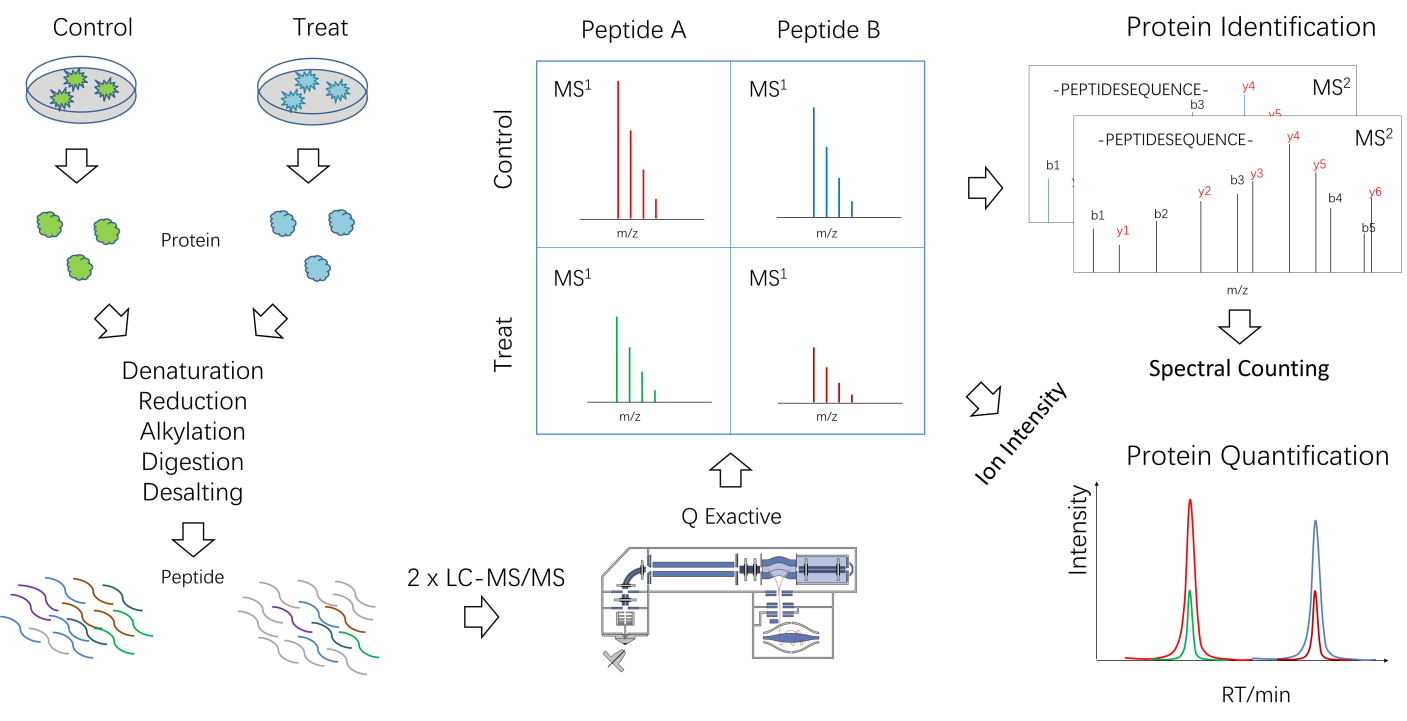


图1. Label-free定量技术原理。

# 实验方法

## 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℃)

* 100mM ABC, 1% SDC, pH 8.5

#### Pierce BCA Protein Assay Kit

* A : B = 50 : 1, 160 μL BCA + 20 μL sample, 37℃ incubate 30min, read at 562nm
* For calibration curve, sequentially dilute BSA (2 mg/mL) to 5 concentration levels

#### TCEP ( tris (2-carboxyethyl) phosphine )

* 1 M TCEP in H2O, store at -80℃

#### IAA ( iodoacetamide )

* 0.5 M IAA in H2O, store at -80℃, away from light

#### Acetone

* Pre-chilled to -20℃

#### H2O, formic acid, acetonitrile (HPLC grade)

* Buffer A: 0.1% FA H2O, 2% ACN;
* Buffer B: 0.1% FA, 70% ACN;

## 实验流程

### 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1000μL工作液充分混匀，冰浴超声 ~5 min，至充分溶解。
3. 14000 g，4℃ 离心15 min，转移上清至新的ep管中。

### BCA定量

1. 按50：1混合regent A、B，加至96孔板，每孔160 μL，5个标准点，一个空白。
2. 加入20 μL样品 (稀释5 – 10倍) 或标准蛋白 (BSA，5个浓度梯度) 。
3. 37℃ 震荡30min，用562nm波长检测吸光度。
4. 根据标准蛋白计算标准曲线，并计算相应样品的蛋白浓度。

### 丙酮沉淀

1. 每个样品取100 μg总蛋白，并用裂解液稀释至 ~1mg/mL。
2. Acetone预冷至 -20℃，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℃ O/N）。
3. 10,000 g转速，4℃离心10 min，小心去除上清。
4. 加入200 μL预冷的80% acetone润洗沉淀2次。

### 蛋白重溶、还原、烷基化以及酶解

1. 加入200 μL 含1% SDC 的100mM ABC，震荡混匀并简单离心。
2. 水浴超声5 - 30 min 充分溶解蛋白沉淀。
3. 加入TCEP至5mM，55 ℃ 孵育10 min，还原二硫键。
4. 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。
5. 用resuspension buffer溶解trypsin至0.5 μg/μL，室温孵育5 min。
6. 加入2 μg trypsin溶液至每个样品中，trypsin：蛋白 = 1：50。
7. 充分混匀，简单离心后37℃ 震荡孵育过夜。

### 去除SDC

1. 加入TFA至混合样品中 (终浓度2%，pH < 2) ，沉淀SDC。
2. 高速离心10 min，转移上清至新的ep管中。
3. 加入n\*100 μL 2% TFA充分混匀，提取共沉淀的多肽。
4. 重复提取2次。
5. 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。

### 多肽脱盐

1. 配制Buffer A：0.1% FA ，H2O，2% ACN；Buffer B： 0.1% FA，70% ACN;
2. 加入500 μL ACN平衡C18脱盐柱。
3. 加入1 mL 含0.1% FA的H2O (Buffer A) 洗去残余的ACN，活化脱盐柱。
4. 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。
5. 用1 mL Buffer A洗柱2次，除去残余盐分。
6. 加入400 μL 0.1%FA，70% ACN (Buffer B) 洗脱多肽，收集流出液B。
7. 用收集的流出液A重复整个脱盐过程一次 (25 - 29)。
8. 合并两次流出液B，真空干燥。
9. 加入buffer A 重溶多肽至1 μg / μL用于LC-MS/MS检测，或-80℃ 冻存。

### nano-UPLC分离

每个组分取2 μg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 μm ID × 15 cm 反相色谱柱（Reprosil-Pur 120 C18-AQ， 1.9um,，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。

### 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。

### 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。

### 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。

## 实验结果

### 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。

|  |  |
| --- | --- |
| Sample ID | Sample name |
| 1 | Blank |
| 2 | Model |
| 3 | Positive |
| 4 | Treat |

表1：分组情况

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

表2: 实验结果统计

|  |  |  |  |
| --- | --- | --- | --- |
| 多肽总数 | 蛋白总数 (Group) | 可定量蛋白数 | 差异蛋白蛋白数 (Group)  Blank-Model Model-Positive Positive-Treat |
| 18336 | 2594 | 2467 | 177 144 213 |

*附件：all protein ，sig protein*

### 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*

**

图2.蛋白丰度分布图

### 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。



图3.总离子流图（TIC



图4.4个样本间定量重复性分析

### 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



图5.空白组和模型组蛋白差异倍数展示。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。



图6.阳性组和模型组蛋白差异倍数展示。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。



图7.阳性组和给药组蛋白差异倍数展示。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。

## 生信分析

### 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*



图8. 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。

### 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*



图9. 空白组和模型组差异蛋白GO富集分析。Bar: -log( p.adjust), point: Count。



图10. 阳性组和模型组差异蛋白GO富集分析。Bar: -log( p.adjust), point: Count



图11. 阳性组和给药组差异蛋白GO富集分析。Bar: -log( p.adjust), point: Count

### 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*



图12. 空白组和模型组差异蛋白KEGG富集分析。Bar: -log( p.adjust), point: Count



图13.阳性组和给药组差异蛋白KEGG富集分析。Bar: -log( p.adjust), point: Count

### 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*



图14. STRING 蛋白互作分析。运用STRINGdb对空白组和模型组差异表达的蛋白进行蛋白互作分析发现，177个基因间存在显著富集的相互作用，PPI enrichment p-value :0。

Known Interactions

database edge from curated databases experiment edge experimentally determined

Predicted Interactions

neighborhood edge gene neighborhood fusion edge gene fusions cooccurrence edge gene co-occurrence

Others

textmining edge textmining coexpression edge co-expression homology edge protein homology

注：光晕颜色代表上(red)下(green)调关系。



图15. STRING 蛋白互作分析。运用STRINGdb对模型组和阳性组差异表达的蛋白进行蛋白互作分析发现，145个基因间存在显著富集的相互作用，PPI enrichment p-value :0。

Known Interactions

database edge from curated databases experiment edge experimentally determined

Predicted Interactions

neighborhood edge gene neighborhood fusion edge gene fusions cooccurrence edge gene co-occurrence

Others

textmining edge textmining coexpression edge co-expression homology edge protein homology

注：光晕颜色代表上(red)下(green)调关系。



图16. STRING 蛋白互作分析。运用STRINGdb对阳性组和给药组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

database edge from curated databases experiment edge experimentally determined

Predicted Interactions

neighborhood edge gene neighborhood fusion edge gene fusions cooccurrence edge gene co-occurrence

Others

textmining edge textmining coexpression edge co-expression homology edge protein homology

注：光晕颜色代表上(red)下(green)调关系。