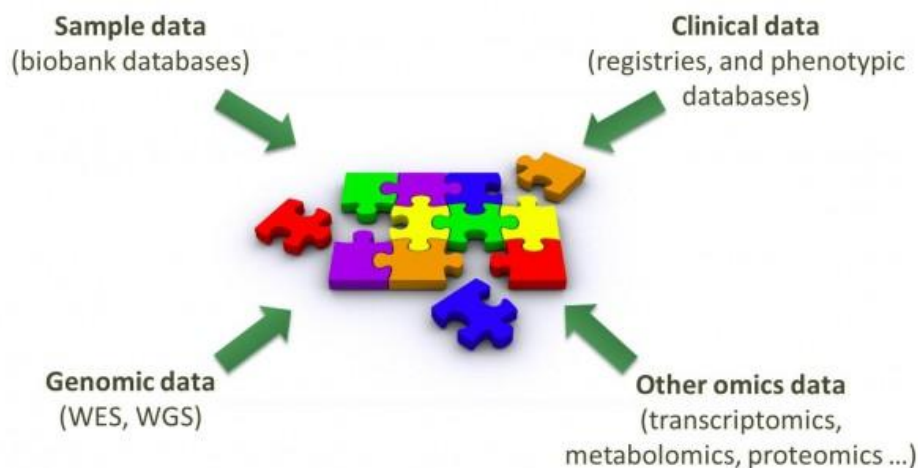


# 生物信息学： 组学时代的生物信息数据挖掘和理解

2020年秋



# 有关信息

- 授课教师：宁康，张礼斌，陈鹏
  - Email: ningkang@hust.edu.cn
  - Office: 华中科技大学东十一楼504室
  - Phone: 87793041, 18627968927
- 课程网页
  - <http://www.microbioinformatics.org/Bioinformatics.html>
  - QQ群:



# 课程安排

## (生物信息中的算法设计与概率统计模型)



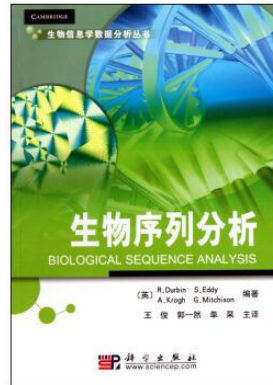
- 生物背景和课程简介
- 生物信息学和生物数据挖掘
  - 生物数据的格式及其意义
    - 序列数据
    - 树状数据
    - 网络数据
    - 表达数据等
  - 生物数据库及其用法
  - 生物信息基本算法
    - 双序列联配
    - 多序列联配
    - 基因组组装算法
    - 基因预测和功能注释
    - 系统发育树构建
    - 蛋白质结构预测
    - 生物调控网络解析
  - 组学数据分析方法
    - 基因组变异分析
    - 基因表达和比较分析
    - 非编码RNA分析
    - 蛋白组分析
    - 宏基因组分析
  - 系统生物学与交叉科学
- 面向生物大数据挖掘的深度学习

研究对象：  
生物序列，  
进化树，  
生物网络，  
基因表达

...

方法：  
生物计算与生物信息

# 教材及参考书目



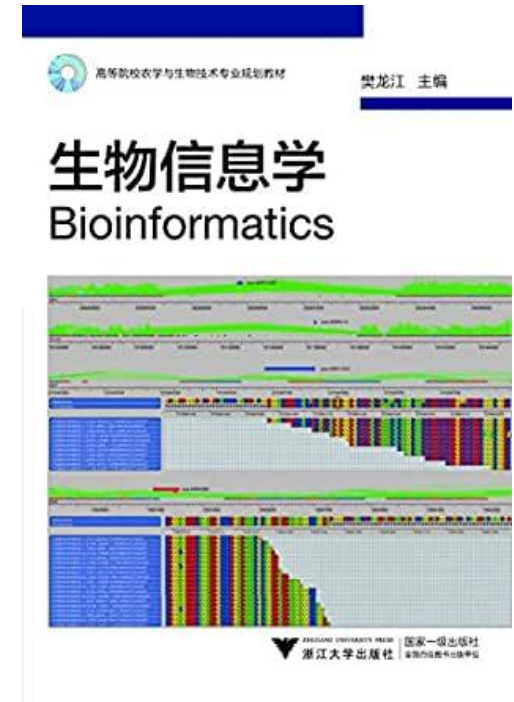
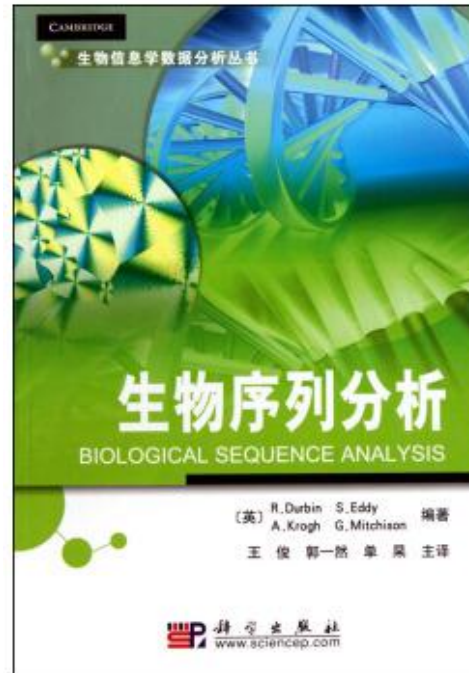
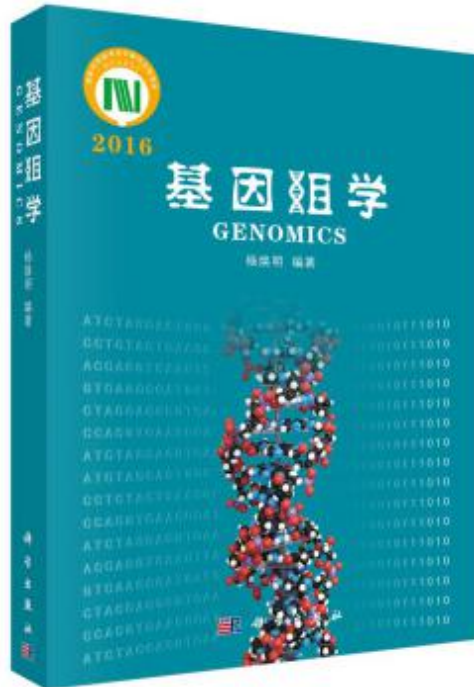
- **教学参考书:**

- 《生物序列分析》（第1版）. 科学出版社. 2010年8月出版. R. Durbin等编著，王俊等主译.

- **课外文献阅读:**

- 《生物信息学》（第1版）. 浙江大学出版社. 2017年3月出版. 樊龙江主编.
- 《基因组学》（第1版）. 科学出版社. 2016年10月出版. 杨焕明主编.

# References



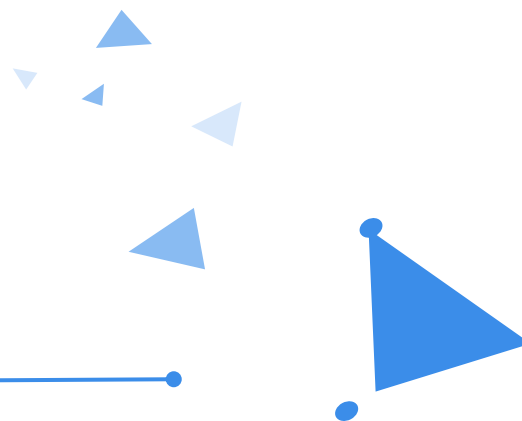
# Slides credits

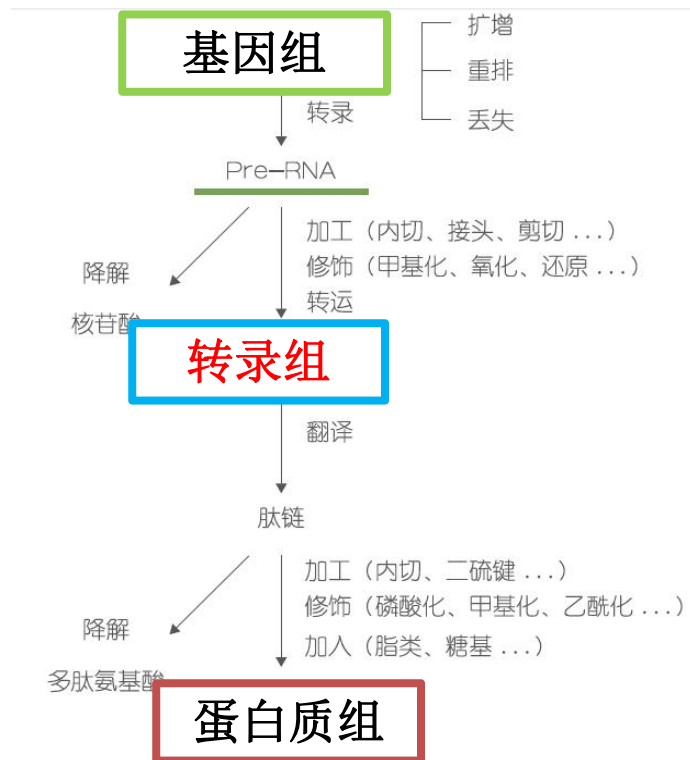
- 生物信息学研究方法概述：北京大学生物信息中心**
- 生物统计学：卜东波@中国科学院计算技术研究所，邓明华@北京大学**
- 神经网络与深度学习：邱锡鹏@复旦大学**
- Introduction to Computational Biology and Bioinformatics: Xiaole Shirley Liu Lab@Harvard University**
- Combinatorial Methods in Computation Biology: Ken Sung Lab@NUS**
- Deep Learning in the Life Sciences: MIT**

# 01

*Part One*

蛋白质组学简介







## 质谱主要组成部分



### 常见类型 (如下)

✓ 电喷雾电离-  
ESI

✓ 基质辅助激光

✓ 四级杆-Quadrupole

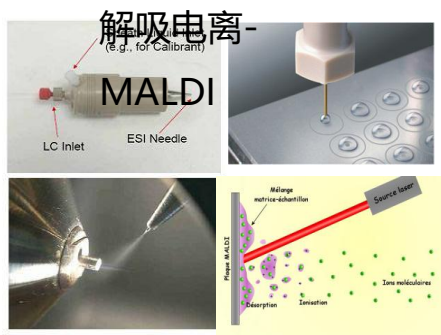
✓ 离子阱-Ion trap

✓ 飞行时间-TOF

✓ 直接检测器

✓ 电子倍增器

✓ 闪烁检测器



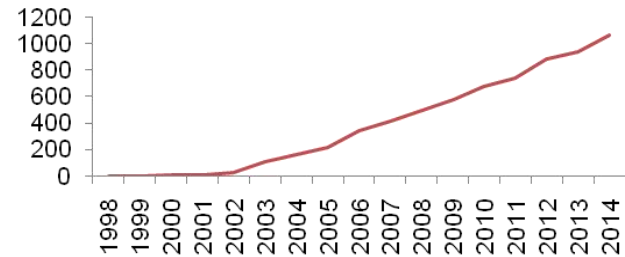
### Orbitrap Fusion



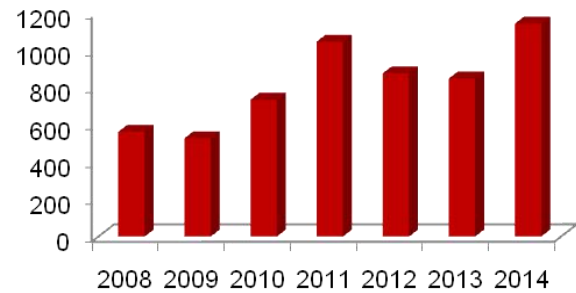
# 基金和政策情况



2014年6月“中国人类蛋白质组计划（CNHPP）”全面启动实施：主要目标是以我国重大疾病的防治需求为牵引，发展蛋白质组研究相关设备及关键技术，绘制人类蛋白质组生理和病理精细图谱、构建人类蛋白质组“百科全书”，全景式揭示生命奥秘，为提高重大疾病防治水平提供有效手段，为我国生物医药产业发展提供原动力。



PubMed中国蛋白质组学领域文章发表数目统计



国家自然科学基金蛋白质组资助金额逐年上升

定性  
蛋白质组

定性

Shotgun

定量蛋白  
质组

相对

iTRAQ/TMT、Label free、SILAC

绝对

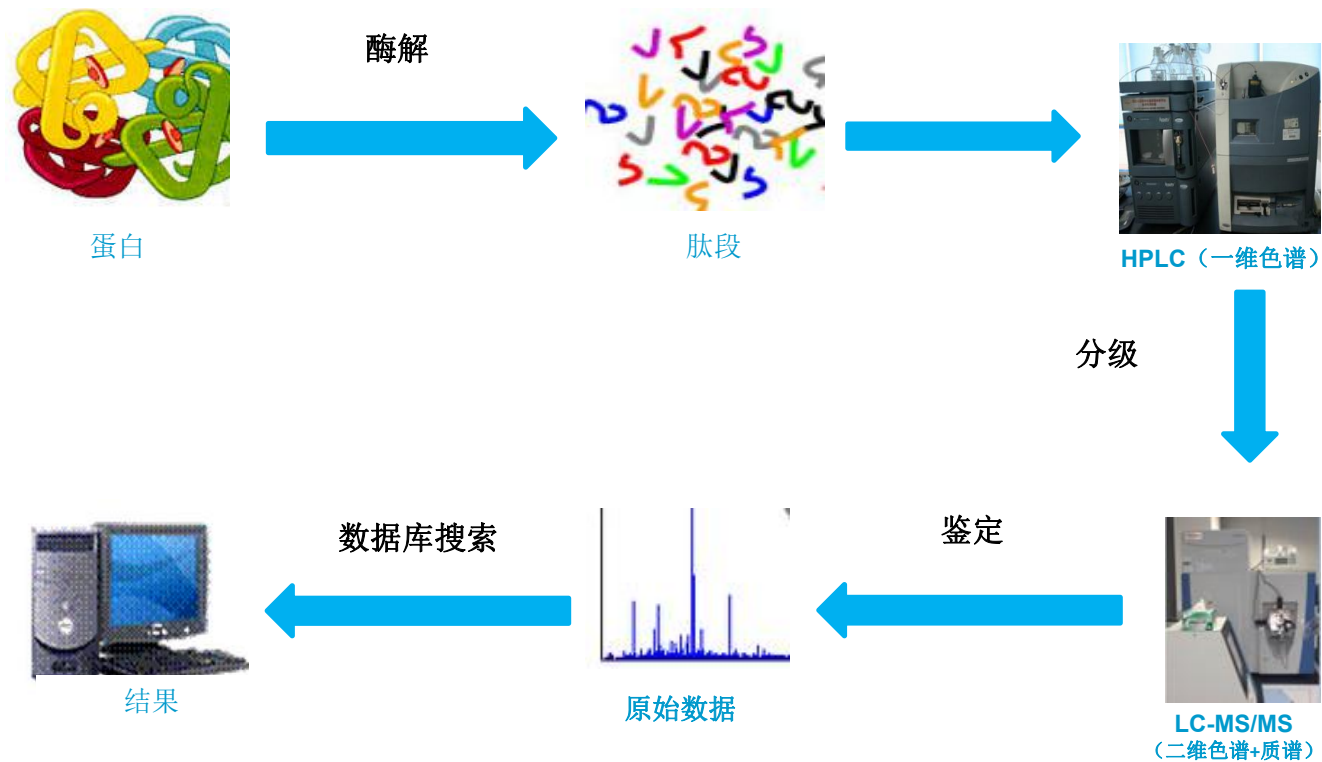
PRM/MRM

翻译后修  
饰蛋白组

探索

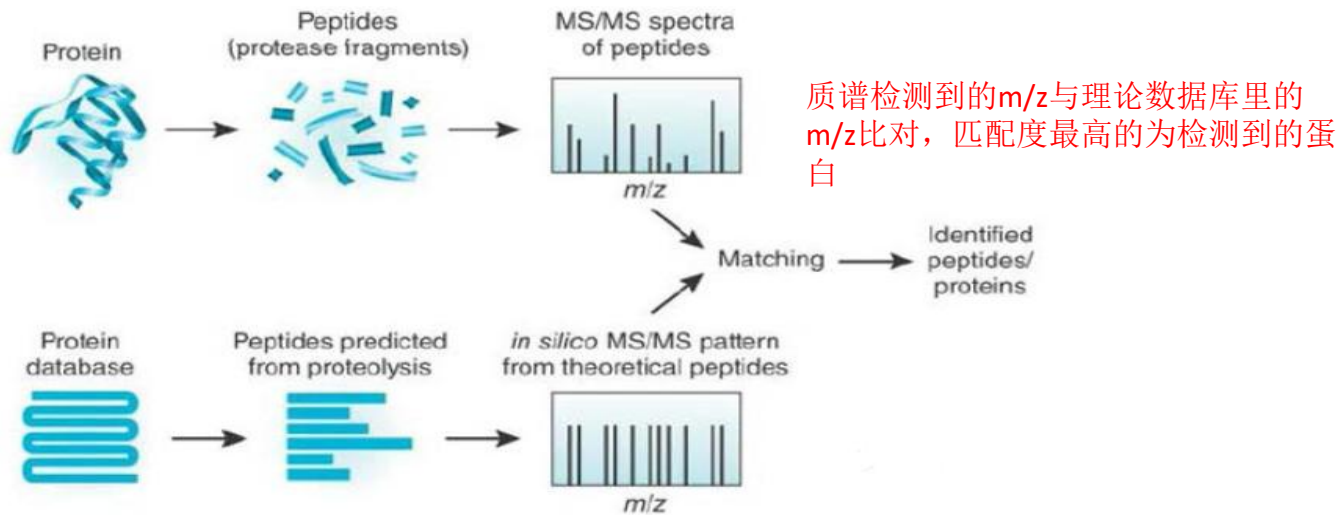
修饰富集结合定量技术

## 基于质谱的蛋白质组学实验基本流程



## 蛋白质定性原理

### 鸟枪法



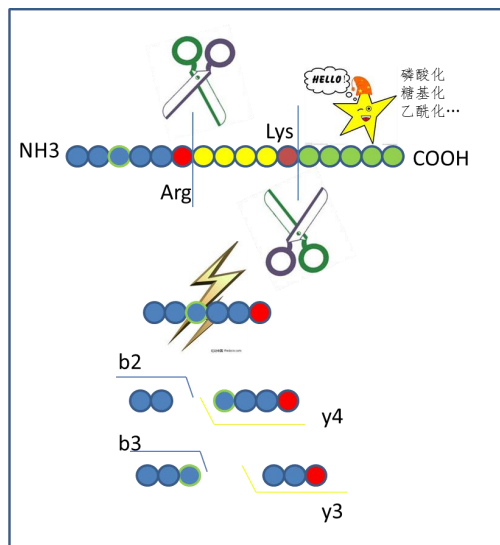
检测物种没有蛋白数据库怎么办？

方案一：搜相近物种的数据库

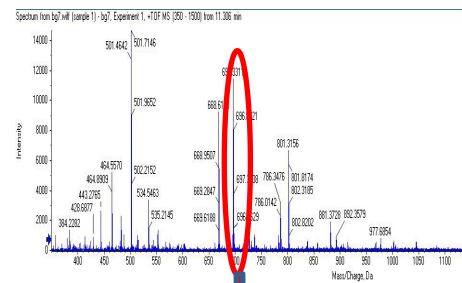
方案二：搜上一级数据库

方案三：基因组或转录组测序数据构建数据库

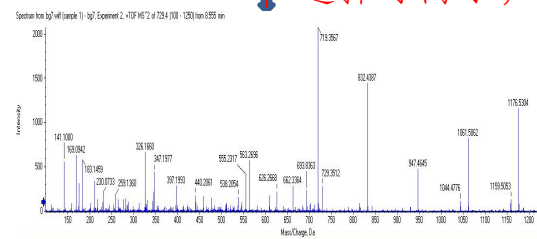
## 蛋白质定性工作原理图



## MS1



## MS2



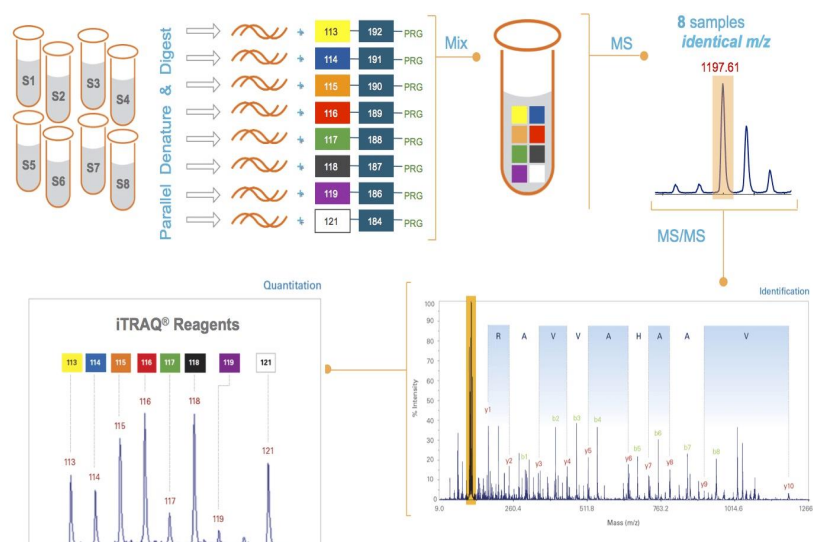
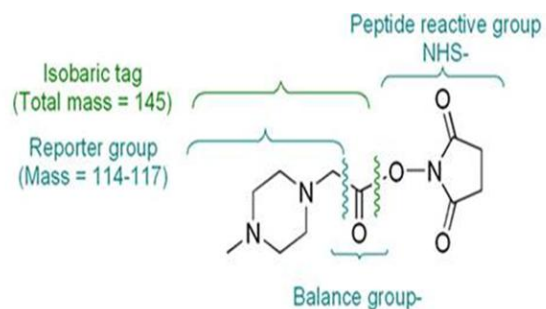
选择母离子，继续进行碎裂

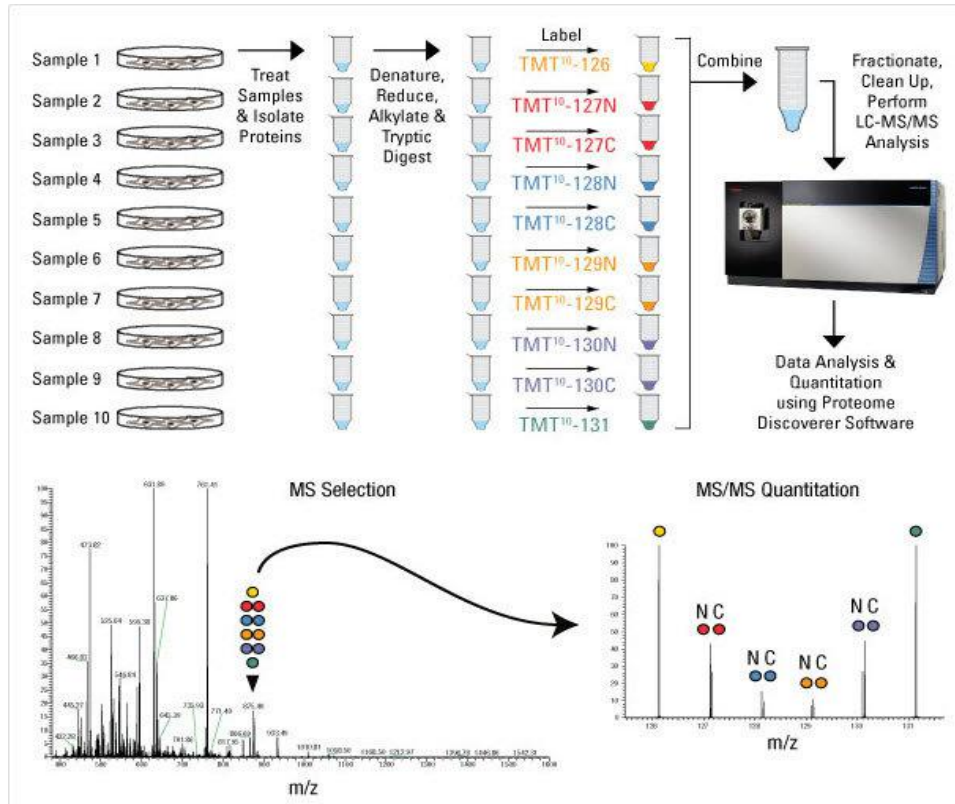
MSKEKFER**TKPHVNVGTIGHVDHGK****TTL**TAAITTVLAKTYGGAARAFDQIDNAPEEKARGITINTSHVEYDTPTRHYAHVDCPGHADYVKNMITGAAQMDGAILVVAATDGPMPQTR**EHILLGR**QVGPYIIVFL  
 NKCDMVDDEEL**LELVEMEVR**ELLSQYDFPGDDTPIVRGSALEK**ALEGDAEWEAK**ILELAGFLDSYIPEPERAIDKPFLLPIEDVFSISGRGTVVIGRVERGIK**VGEEVEIVGIKETQK**STCTGVEMFRKLLDEGR  
 AGENVGVLRLRGIKREEIER**GQVLAK**PGTIKPHTKFSEVYILSKDEGGRHPTFFKGYRPGFYFR**TTDVTGTIEL**PEGVEMVMPGDNIKRMVTLIHP**IAMDDGLRFAIREGGRTVGAGVVARVLS**

## iTRAQ基本原理

iTRAQ 技术是由美国应用生物系统公司(Applied Biosystems Incorporation, ABI)2004年开发的同位素标记相对和绝对定量(isobaric tags for relative and absolute quantitation, iTRAQ)技术, 是目前差异蛋白质定量分析方法通量最高, 系统误差最小, 功能最强大的分析方法之一。可以同时比较2-8组样品的蛋白质表达差异。

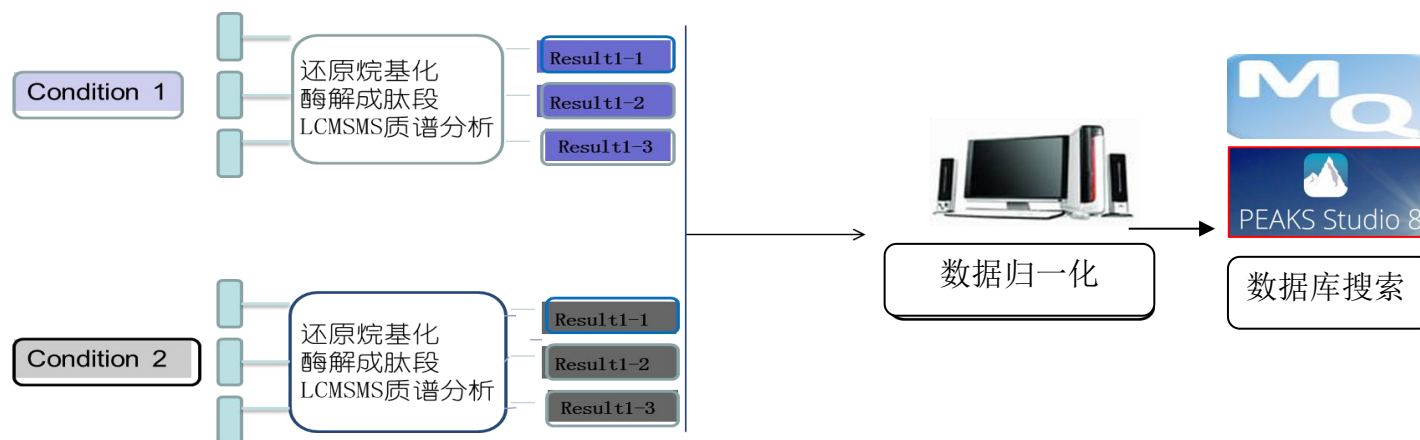
## iTRAQ® Reagents Workflow





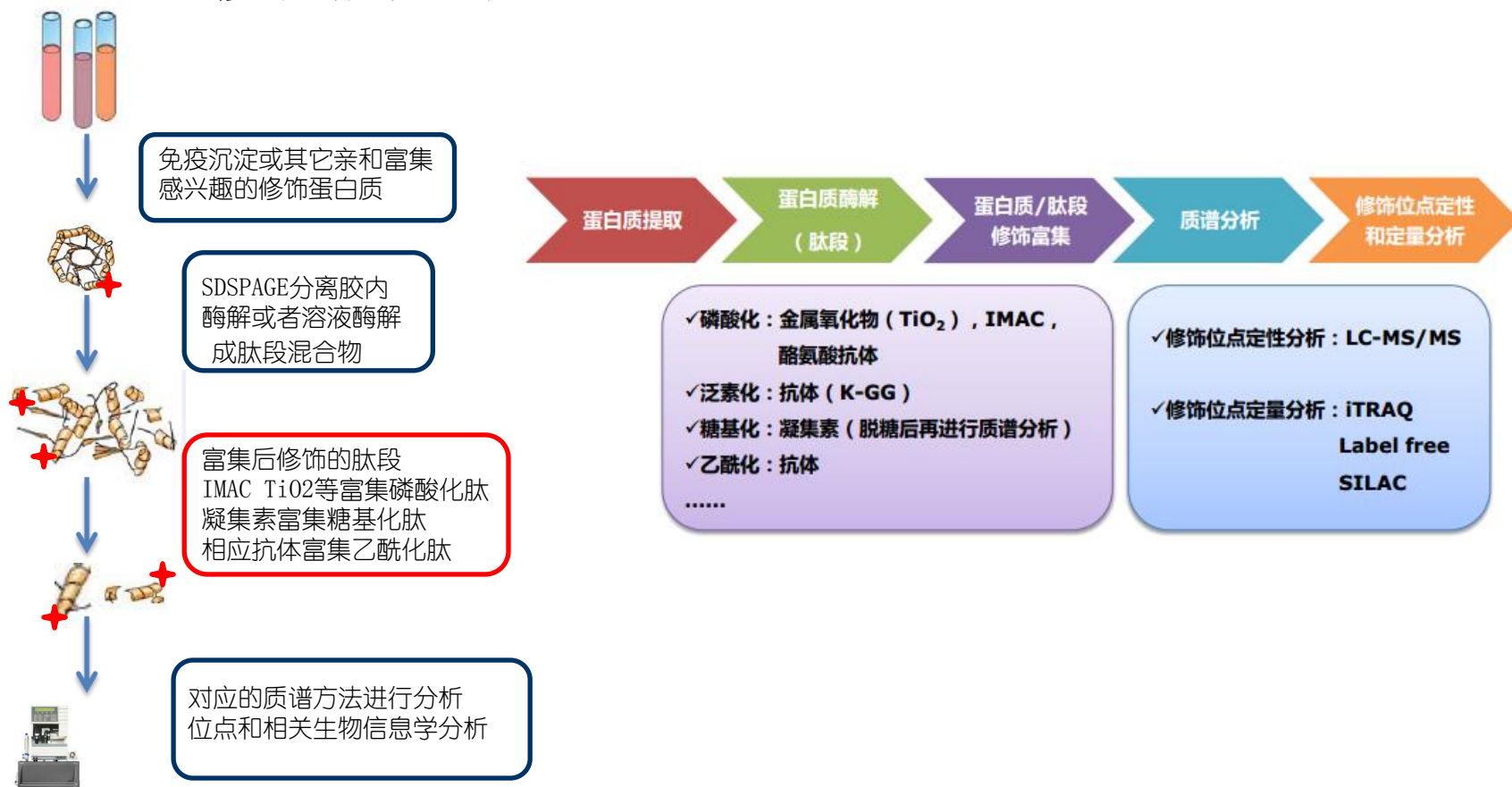


## 非标记定量--Label free



- 操作简单
- 系统误差大
- 机时较长
- 数据分析工作量大

# 修饰蛋白质组学



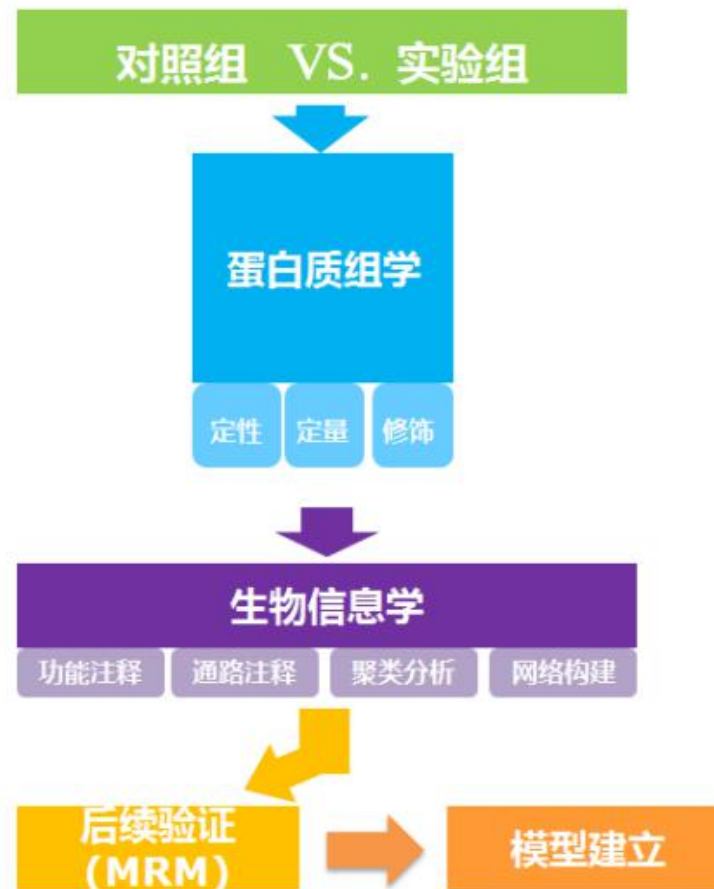
## 农业应用

- 植物逆境（抗盐，高温，干旱，低温等）
- 植物保护（病虫，机械损伤，农药等）
- 遗传发育（器官组织发育的不同时期等）
- 动物育种（不同品种，抗病，感染等）
- 食品科学（发酵时期，食品风味，冷藏时间，加工手段）
- 微生物机理研究（药物处理，环境影响等）
- 环境科学
- .....

## 医学应用

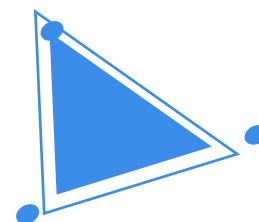
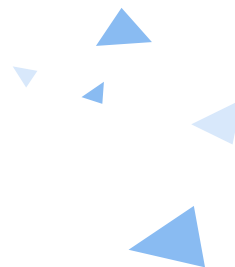
- 疾病标志物biomarker
- 药物靶点
- 机理研究
- 药理研究
- 信号通路
- .....

## 定量蛋白质组学实验思路



# 02 *Part Two* 案例分析

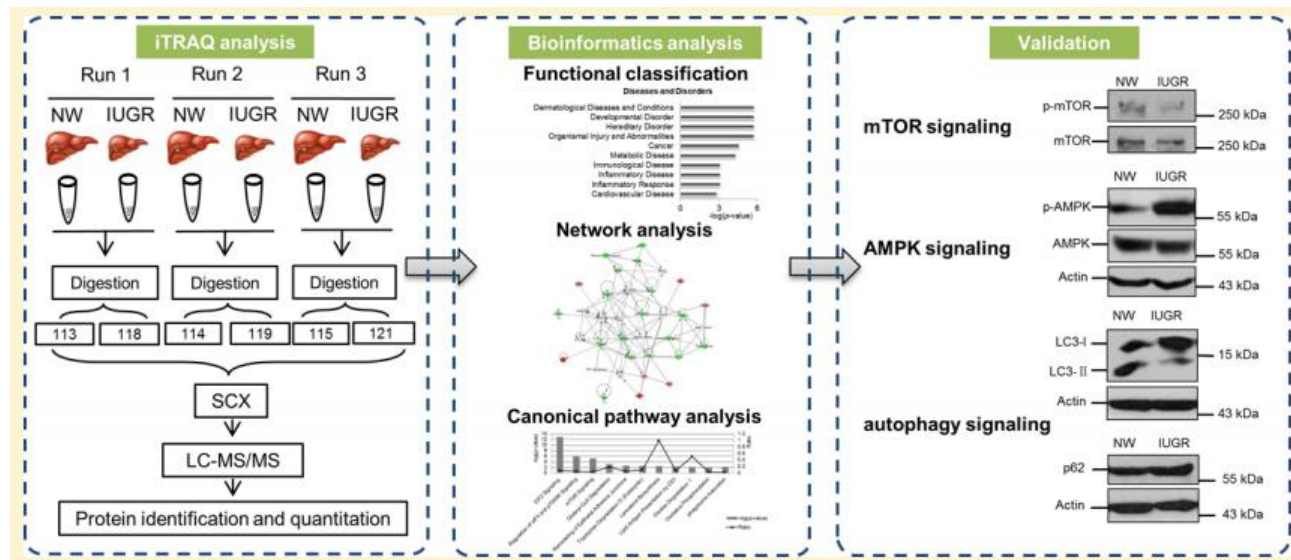
---



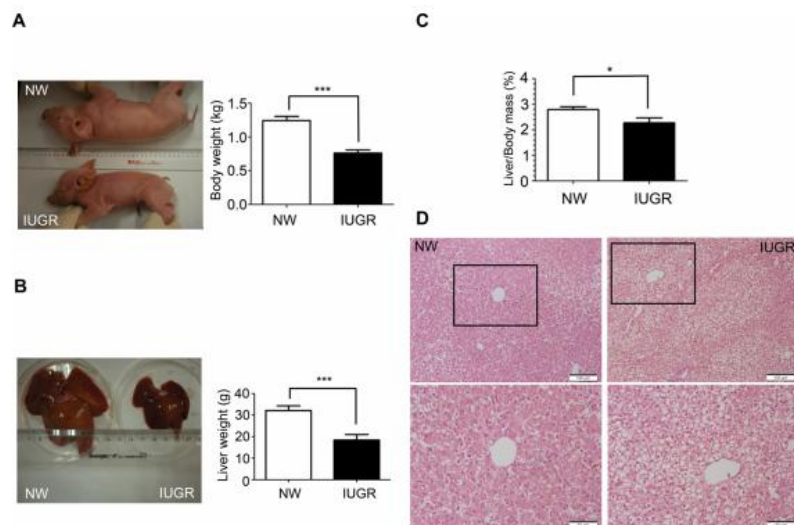
# 案例1

## Global Liver Proteome Analysis Using iTRAQ Reveals AMPK–mTOR–Autophagy Signaling Is Altered by Intrauterine Growth Restriction in Newborn Piglets

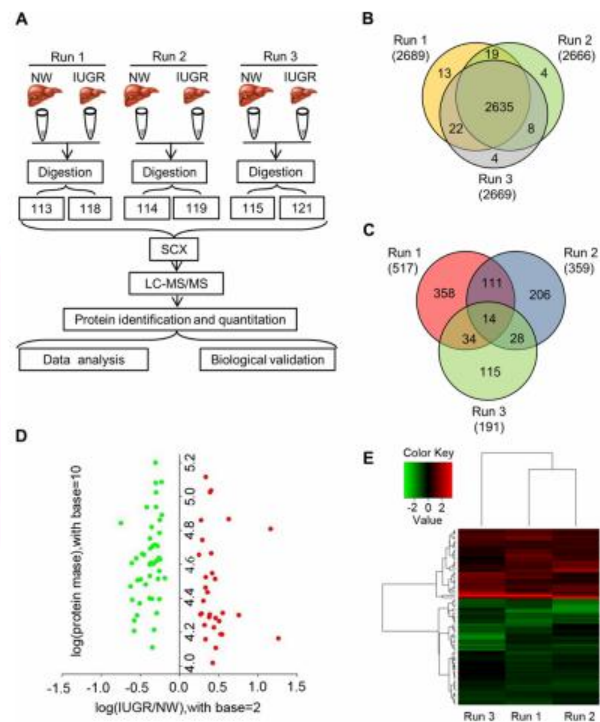
实验思路



## 实验结

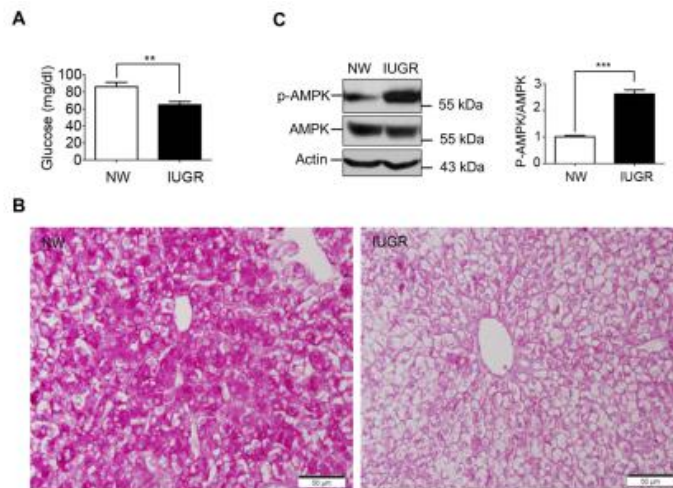


NW&IUGR形体和肝脏的比较

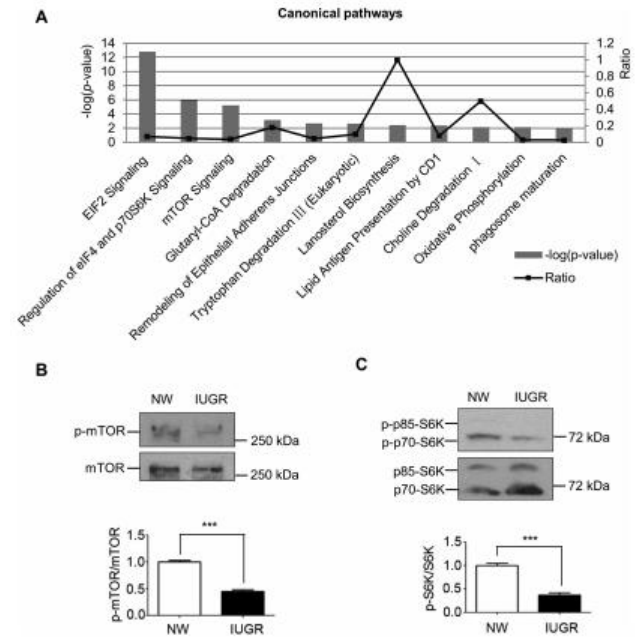


NW&IUGR差异蛋白分析





IUGR 中AMPK信号途径被激活

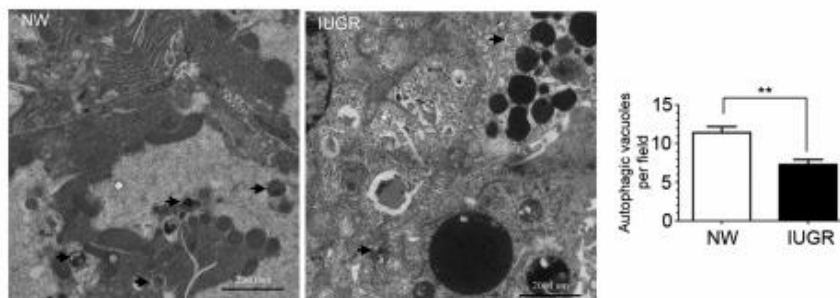


IUGR中mTOR信号途径受到抑制

**A**

Phagosome maturation			
Gene Symbol	First Run	Second Run	Third Run
NAPA	0.824*	0.755*	0.856*
RAB7A	0.769*	0.708*	0.970
TUBA1A	0.749*	0.669*	0.889*

**B**

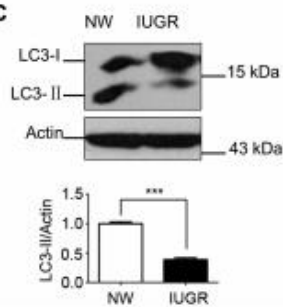


IUGR 中自我吞噬受到抑制

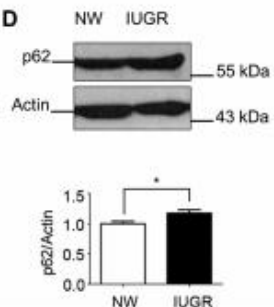
**E**

Gene Symbol	First Run	Second Run	Third Run
PP2Ac	0.740	0.729*	0.829*

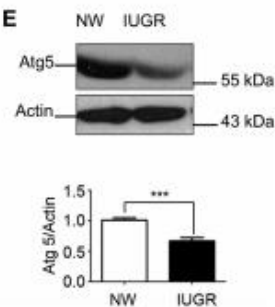
**C**



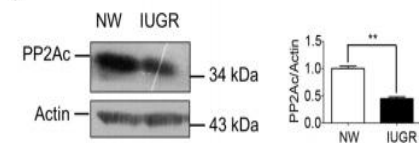
**D**



**E**



**F**



## 案例2

# SCIENTIFIC REPORTS



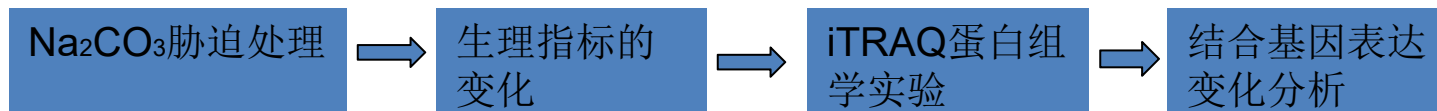
IF= 5.228

OPEN

## **Na<sub>2</sub>CO<sub>3</sub>-responsive mechanisms in halophyte *Puccinellia tenuiflora* roots revealed by physiological and proteomic analyses**

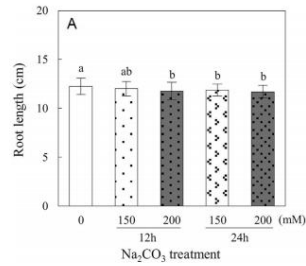
Received: 13 May 2016  
Accepted: 15 August 2016

### 实验思路

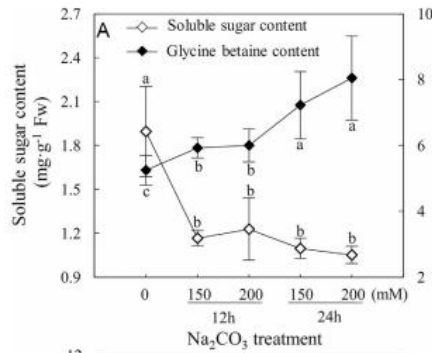


## 实验结果

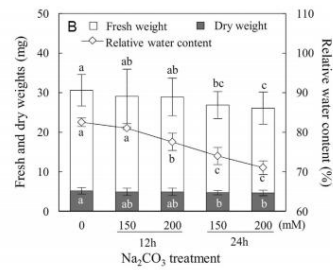
A



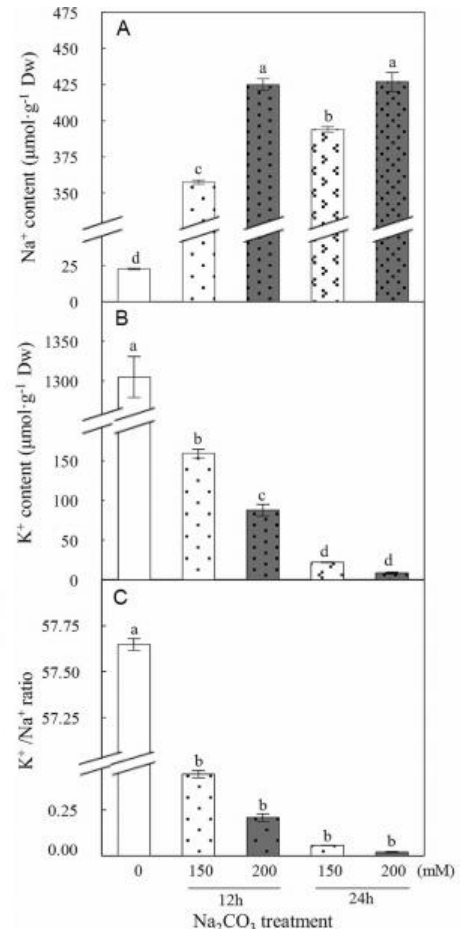
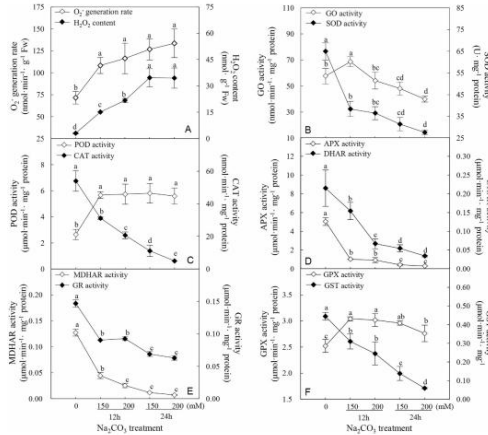
C

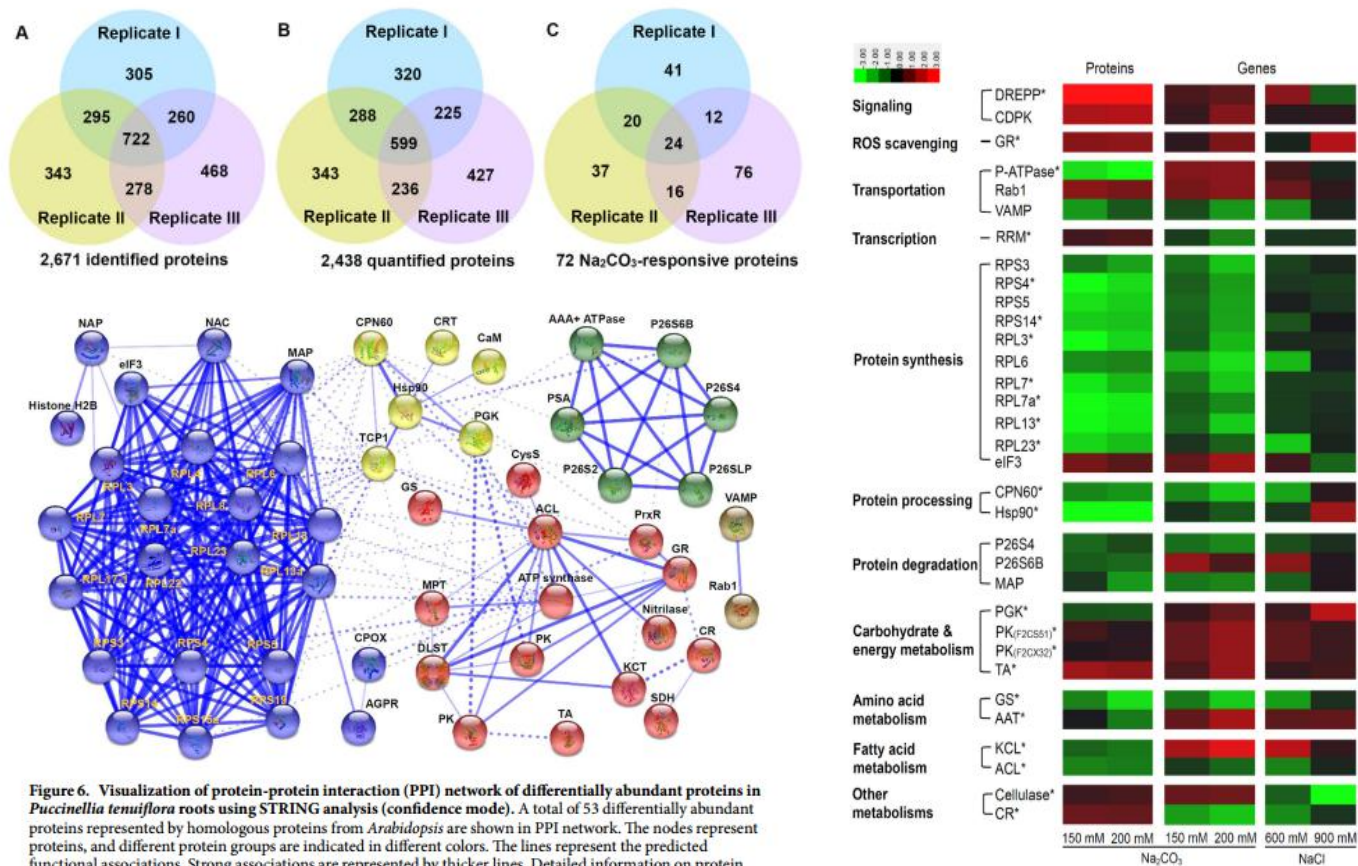


B



D

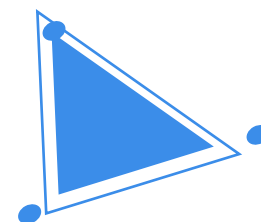




# 03

*Part Three*

样品采集和实验设计要  
求



## ●迅速性原则

样本质量是实验中影响实验结果的最关键因素，因此用于研究的样本，在采集、制备、贮存、运输过程中应尽可能地做到迅速，最大限度的缩短从样本采集到实验的时间，减少非实验因素对样品产生影响。

## ●低温原则

所取样本离体后，如有分离步骤立即在4度或冰上等低温下进行，分离好的样品立即置于液氮、干冰或-80℃冰箱中，并保证在实验前始终处于-70℃以下，以及避免蛋白的降解和蛋白的继续变化。

## ●样本无杂质污染或干扰

（动物组织无血迹残留）

## ●取样部位要一致

## ●包装完好，标记明确

## ●取样的质量或体积参考送样指导

## 实验设计中应注意的问题

### 重复性问题

- 蛋白组学：  
至少3个生物学重复

### 技术选择问题

- 组学技术作为大范围筛选的工具，不适用于对某个特定蛋白的检测，如果想要检测特定物质建议采用靶向技术（蛋白质组的PRM技术）。

### 后期验证问题

- 验证方法多样性，多种方法交叉验证
- 验证范围针对性，针对某一方向深入细致的分析
- 验证角度全面性，多个方面综合考虑



# Proteomics

