

**多组学联合项目报告**

**多组学联合分析项目报告**

****

**iTRAQ蛋白质组**

**与iTRAQ磷酸化蛋白质组联合分析**

**生物信息学分析报告**

**项目名称：**

**委托单位：**

**项目编号：**

**报告时间：**

目录

[1 联合分析概述 Introduction 4](#_Toc44925224)

[2 联合分析流程 Flowchart 5](#_Toc44925225)

[3 生物信息学分析 Bioinformatics analysis 6](#_Toc44925226)

[3.1 表达水平分析 6](#_Toc44925227)

[3.1.1 蛋白关联与关联数量统计 6](#_Toc44925228)

[3.1.2 关联蛋白的表达差异数量统计 7](#_Toc44925229)

[3.2 功能分析 10](#_Toc44925230)

[3.2.1 单组学层面GO功能富集比对分析 11](#_Toc44925231)

[3.2.2 两类重要联合蛋白的GO功能注释&富集 12](#_Toc44925232)

[3.2.3 单一组学层面的KEGG通路富集比较分析 15](#_Toc44925233)

[3.2.4 两类重要联合蛋白的KEGG通路注释&富集 15](#_Toc44925234)

[3.2.5 KEGG通路整合图 18](#_Toc44925235)

[3.2.6 互作网络整合分析 19](#_Toc44925236)

[3.3 激酶分析（限人、小鼠、大鼠） 20](#_Toc44925237)

[3.3.1 激酶注释 21](#_Toc44925238)

[3.3.2 激酶的表达变化分析 23](#_Toc44925239)

[3.3.3 激酶与底物关系分析 24](#_Toc44925240)

[4 生物信息学分析方法 Bioinformatic Methods 27](#_Toc44925241)

[4.1 聚类分析 27](#_Toc44925242)

[4.2 注释分析 27](#_Toc44925243)

[4.3 富集分析 27](#_Toc44925244)

[4.4 蛋白质相互作用网络分析 27](#_Toc44925245)

[4.5 激酶注释 27](#_Toc44925246)

[5 参考文献 References 28](#_Toc44925247)

[6 组学期刊投稿指南（供参考）Publication Recommendation 28](#_Toc44925248)

[7 部分合作发表文章 Cooperative Projects Papers 30](#_Toc44925249)

# 联合分析概述 Introduction

**多组学分析的意义：**传统单一层面的组学研究比如基因组，转录组，蛋白组，代谢组等研究手段能够在一定程度上对特揭示遗传信息、蛋白功能或代谢通路进行阐释。但是，仅通过单一组学数据很难对复杂的生物网络调控进行更加系统性更全面的解释，不足以解释遗传信息表达调控的传递链条。因此，多组学数据整合成为了组学研究的发展趋势，其优势如下：（1）弥补单一组学分析时，由于数据噪音、缺失等因素带来的数据问题。（2）多组学数据资源之间可以进行相互验证，减少单一组学分析带来的假阳性。（3）多组学数据联合分析更有利于从不同层面、不同角度，系统解析生物学模型的多层次机制或机制表型的联系。

**蛋白质组与修饰蛋白质组联合分析的意义：**生命活动的功能执行者是蛋白质，蛋白质功能的调控功能发挥主要受其自身表达水平、以及翻译后修饰状态等方面的影响，并最终影响生物的表型变化。因此，基于蛋白组与修饰蛋白质组的联合分析，是系统探究蛋白层面功能调控的有效研究手段。

**蛋白质组与修饰蛋白质组联合分析解决的科学问题：**在联合分析中，通过差异蛋白与差异修饰**表达趋势分析**、蛋白与修饰**功能对比分析、特定功能蛋白分析（如激酶）**这三个方面的分析，拟解析如下科学问题：（1）比较蛋白表达、蛋白修饰表达的表达模式异同，揭示蛋白在不同层面上受生物学处理后的扰动变化程度。（2）筛选由蛋白、修饰层面共同或分别调控的重要通路（或功能条目），揭示其在相关生物学变化中的潜在调控作用。（3）分析关键功能蛋白质（如激酶）的上下游，探讨关键调控蛋白在介导生物学变化中的重要作用。**最终，通过联合分析实现对蛋白层面变化趋势的系统了解，筛选出影响生物表型变化的潜在关键蛋白质分子或修饰位点，提出分子生物学变化机制或模型，并为后续筛选出重点蛋白或蛋白修饰位点，进行深入实验分析及应用奠定基础。**

# 联合分析流程 Flowchart

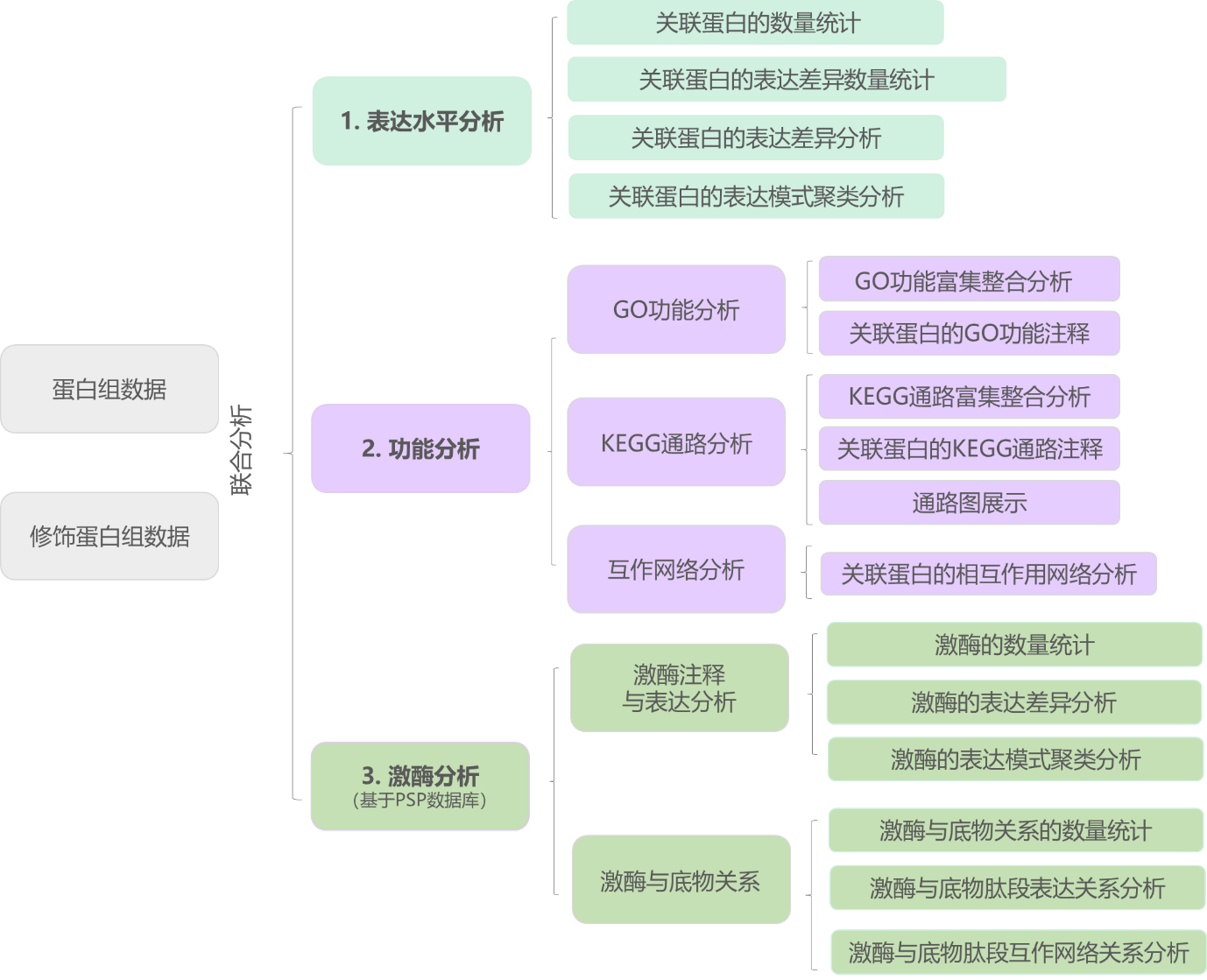


图 生物信息学分析内容

# 生物信息学分析 Bioinformatics analysis

## 表达水平分析

### 蛋白关联与关联数量统计

**关联蛋白：**同一蛋白在蛋白组、修饰蛋白组层面均鉴定到并具有可用于统计分析的定量信息。对关联蛋白在蛋白表达层面、蛋白修饰层面的表达差异，进行统计比较。统计结果见表**“关联蛋白的数量统计表”**和**“关联蛋白的数量统计韦恩图”**。

如下表，分别展示了每个比较组中，蛋白的**数量统计结果**，包括：蛋白质组层面的可定量蛋白，与显著差异表达蛋白；磷酸化蛋白质组层面的可定量修饰蛋白，与显著差异表达的修饰蛋白；两个组学层面的关联蛋白，仅蛋白水平有显著差异的关联蛋白，仅磷酸化水平均有显著差异的关联蛋白，以及在蛋白水平与磷酸化水平均有显著差异的关联蛋白。进一步，以韦恩图形式展示数量统计结果，如下图。

表1 关联蛋白的数量统计表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Quantified proteins** | **Differential expressed proteins** | **Quantified phosphoproteins** | **Differential expressed phosphoproteins** |
|  |  |  |  |  |

说明：Quantified proteins，可定量蛋白；Differential expressed proteins，显著差异蛋白，即在蛋白表达水平有显著差异的蛋白；Quantified phosphoproteins，可定量磷酸化蛋白（修饰蛋白中可定量磷酸化肽段≥1）； Differential expressed phosphoproteins，显著差异磷酸化蛋白，即在磷酸化肽段水平有显著差异的磷酸化蛋白（蛋白中差异磷酸化肽段≥1）；Correlated proteins，关联蛋白，即在蛋白表达水平与蛋白磷酸化肽段水平均有定量信息；Correlated proteins(Differential expressed only in protein level)，仅蛋白水平有显著差异的关联蛋白；Correlated proteins(Differential expressed only in phosphoproteins level) ，仅磷酸化水平均有显著差异的关联蛋白；Correlated proteins(Differential expressed both in protein and phosphoprotein level)，在蛋白水平、磷酸化水平均有显著差异的关联蛋白。

[venn]

关联蛋白的数量统计韦恩图

输出文件：

1. 关联蛋白数量统计表
2. 关联蛋白数量统计韦恩图

### 关联蛋白的表达差异数量统计

根据关联蛋白在蛋白表达水平以及磷酸化修饰水平的不同差异特征，对关联蛋白进行分类。根据表达趋势变化，可**将关联蛋白分为不同模式**，包括在蛋白与磷酸化水平均无差异的关联蛋白，仅在蛋白表达水平或磷酸化修饰水平有显著变化的关联蛋白，蛋白表达与磷酸化水平均发生显著变化的关联蛋白等，具体如下表。

表2 关联蛋白的表达差异数量统计表

|  |  |  |  |
| --- | --- | --- | --- |
| **Comparisons** | **groupvs** | | |
|  | **Unchanged (Proteins)** | **Upregulated (Proteins)** | **Downregulated (Proteins)** |
|  |
| **Unchanged (Phosphoproteins )** |  |  |  |
| **Upregulated (Phosphoproteins )** |  |  |  |
| **Downregulated (Phosphoproteins )** |  |  |  |

说明： Comparisons，比较组。表格中行表示关联蛋白在蛋白表达水平的差异变化(Protein)，列表示关联蛋白在磷酸化表达水平的差异变化(Phosphoprotein)。 Unchanged，蛋白水平无显著差异的关联蛋白；Upregulated，显著上调的关联蛋白；Downregulated，显著下调的关联蛋白。

输出文件：

1. 关联蛋白的表达差异数量统计表

利用火山图直观展示所有可关联蛋白在两个组学中的差异变化，用不同颜色标注不同趋势类型的关联蛋白。结果如下图，关联蛋白仅在蛋白表达水平有差异，标注为蓝色；关联蛋白仅在修饰肽段水平有差异，标注为黄色；关联蛋白在蛋白表达水平、修饰肽段水平均有差异，标记为紫色。

[volcano]

关联蛋白的表达差异火山图

说明：横坐标为差异倍数（以2为底的对数变换），纵坐标为差异的显著性P-value（以10为底的对数变换），左上区为显著差异下调，右上区为显著差异上调。其中关联蛋白仅在蛋白表达水平有差异标注为蓝色，关联蛋白仅在修饰肽段水平有差异标注为黄色，关联蛋白在蛋白表达水平、修饰肽段水平均有差异标记为紫色。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Protein ID** | **Protein Name** | **Gene Name** | **Protein FC (B\_vs\_A)** | **p-value** | **Phosphorylated Site** | **Amino acid** | **Phosphorylated peptide FC (B\_vs\_A)** | **p-value** |
|  |  |  |  |  |  |  |  |  |

说明： Protein ID，蛋白ID；Protein Name，蛋白名称；Gene Name，基因名称；Protein FC，蛋白差异表达倍数；p-value，差异倍数统计检验； Phosphorylated Site（N），Amino acid (X)，Phosphorylated peptide FC，该蛋白第N位X氨基酸磷酸化修饰水平差异倍数。

输出文件：

1. 关联蛋白的表达差异火山图
2. 关联蛋白差异表达详情表

为了分析比较组中关联蛋白在蛋白水平与磷酸化水平表达变化的**分布特征**，以及比较蛋白与磷酸化差异的**整体表达模式的异同**，将关联蛋白进行聚类分析。

如下图，蛋白组表达展示在左侧，磷酸化组表达展示在右侧，颜色深浅表明差异倍数（fold change）大小。蛋白质组与磷酸化组颜色越接近，说明它们的表达趋势更接近；颜色差异越大，说明蛋白质组与磷酸化组的表达趋势差异更大。

[cluster]

关联蛋白的表达模式聚类热图

说明：图中每一行表示一个关联蛋白，每一列表示一个比较组，其中左边为关联蛋白在蛋白水平的表达倍数差异，右边为关联蛋白在磷酸化水平（修饰肽段水平）的表达倍数差异。图中不同颜色表明上下调，颜色深浅为表达倍数差异大小，其中红色为表达上调，蓝色为表达下调。

输出文件：

1. 关联蛋白的表达模式聚类热图

## 功**能分析**

功能维度的联合分析主要目标包括：（1）两个层面在功能、通路方面的异同，即哪些功能和通路是蛋白表达与蛋白修饰协同参与的，哪些功能以各自单独调控为主；（2）两个层面如何相互调控和影响，具有怎样的分子网络关系。

深入挖掘蛋白质参与的生物功能，需要借助于生物信息学的方法进行蛋白功能挖掘，如GO、KEGG、PPI等。功能联合分析的内容包括：（1）单一组学层面GO、KEGG注释及富集的整合比较分析；（2）两类重要关联蛋白的GO、KEGG的比较分析；（3）关联蛋白的KEGG通路整合分析；（4）关联蛋白的互作网络分析。

### 单组学层面GO功能富集比对分析

首先，将蛋白组、磷酸化组各自分别进行GO注释和功能分析；然后，将两者富集的结果整合展示在同一张图中。可直观比较各GO条目在蛋白表达层面、修饰层面的富集程度的异同，进而反映蛋白表达、蛋白修饰在GO功能维度的协同关系及差异调控特征。如下图。

[go\_enrich1]

单组学层面GO功能富集比对图

说明：图中纵坐标表示GO功能条目；圆圈表示差异蛋白，圆圈大小代表差异蛋白数量，三角表示差异的磷酸化蛋白，三角大小代表差异磷酸化蛋白数量，横坐标为rich factor，表示注释到该功能的差异蛋白或差异磷酸化蛋白数目与注释到该功能的所有蛋白数目的比值。标识点颜色表示对富集比值的Fisher统计检验P值。Rich factor值越大，P值越小，说明该功能越显著富集。

输出文件：

1. 单组学层面GO功能富集比对图

### 两类重要联合蛋白的GO功能注释&富集

在上述关联蛋白的表达差异数量统计结果中，有两类生物学意义较明确，也是较多研究者更为关注的蛋白：仅在磷酸化水平有差异的关联蛋白，以及在蛋白与磷酸化水平均有差异的关联蛋白。专门对这两类蛋白分别进行GO功能注释与富集分析，可以帮助进一步了解这两类关联蛋白所参与功能的侧重性，有助于后续实验进一步挑选和探究的相关生物功能和机理，GO功能注释图与富集图如下。

[go\_level2-1]

关联蛋白（蛋白表达无差异、磷酸化表达显著差异）的GO功能注释柱状图

[go\_level2-2]

关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）的GO功能注释柱状图

说明：图中纵坐标表示GO 二级功能注释信息（GO Level2），包含生物过程（Biological Process），分子功能（Molecular Function）和细胞组分（Cellular Component），依次以紫色，红色，橙色予以区分；横坐标表示每个功能分类下的关联蛋白质数目。

[go\_enrich2]

关联蛋白（蛋白表达无差异、磷酸化表达显著差异）的GO功能富集柱状图

说明：图中横坐标表示GO 二级功能注释信息（GO Level2），包含生物过程（Biological Process），分子功能（Molecular Function）和细胞组分（Cellular Component），纵坐标表示该功能分类注释到的蛋白数量。柱子上方数字为rich factor，表示注释到该功能的关联蛋白（蛋白表达无差异、磷酸化表达显著差异）数目与注释到该功能的所有蛋白数目的比值。柱子颜色表示对富集比值的Fisher统计检验P值。Rich factor值越大，P值越小，说明该功能越显著富集。

[go\_enrich3]

关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）的GO功能富集柱状图

说明：图中横坐标表示GO 二级功能注释信息（GO Level2），包含生物过程（Biological Process），分子功能（Molecular Function）和细胞组分（Cellular Component），纵坐标表示该功能分类注释到的蛋白数量。柱子上方数字为rich factor，表示注释到该功能的关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）数目与注释到该功能的所有蛋白数目的比值。柱子颜色表示对富集比值的Fisher统计检验P值。Rich factor值越大，P值越小，说明该功能越显著富集。

输出文件：

1. 关联蛋白（蛋白表达无差异、磷酸化表达显著差异）的GO功能注释柱状图
2. 关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）的GO功能注释柱状图
3. 关联蛋白（蛋白表达无差异、磷酸化表达显著差异）的GO功能富集柱状图
4. 关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）的GO功能富集柱状图
5. GO功能富集详情表

### 单一组学层面的KEGG通路富集比较分析

首先，将蛋白组、磷酸化组各自分别进行KEGG通路注释和功能分析；然后，将两者富集的结果整合展示在同一张图中。可直观比较各KEGG通路在蛋白表达层面、修饰层面的富集程度的异同，进而反映蛋白表达、蛋白修饰在通路维度的协同关系和差异调控特征。如下图。

[kegg\_enrich1]

单组学层面KEGG通路富集比对图

说明：图中纵坐标表示KEGG通路条目；圆圈表示差异蛋白，圆圈大小代表差异蛋白数量，三角表示差异的磷酸化蛋白，三角大小代表差异磷酸化蛋白数量，横坐标为rich factor，表示注释到该功能的差异蛋白或差异磷酸化蛋白数目与注释到该功能的所有蛋白数目的比值。标识点颜色表示对富集比值的Fisher统计检验P值。Rich factor值越大，P值越小，说明该功能越显著富集。

输出文件

1. 单组学层面KEGG通路富集比对图

### 两类重要联合蛋白的KEGG通路注释&富集

同样，对仅在磷酸化水平有差异的关联蛋白，以及在蛋白与磷酸化水平均有差异的两类关联蛋白，专门分别进行KEGG通路注释与富集分析，可以帮助进一步了解这两类关联蛋白所参与通路的侧重性，有助于后续实验进一步挑选和探究的相关通路及调控机理，KEGG功能注释图与富集图如下。

[top\_map1]

关联蛋白（蛋白表达无差异、磷酸化表达显著差异）的KEGG注释柱状图

[top\_map2]

关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）的KEGG功能注释柱状图

说明：图中纵坐标表示KEGG注释通路信息；横坐标表示每个通路分类下的关联蛋白质数目。

[kegg\_enrich2]

关联蛋白（蛋白表达无差异、磷酸化表达显著差异）的KEGG通路富集柱状图

[kegg\_enrich3]

关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）的KEGG通路富集柱状图

说明：图中纵坐标表示KEGG通路功能条目，横坐标表示该功能分类注释到的蛋白数量。柱子右侧数字为rich factor，表示注释到该功能的关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）数目与注释到该功能的所有蛋白数目的比值。柱子颜色表示对富集比值的Fisher统计检验P值。Rich factor值越大，P值越小，说明该功能越显著富集。

输出文件

1. 关联蛋白（蛋白表达无差异、磷酸化表达显著差异）的KEGG注释柱状图
2. 关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）的KEGG功能注释柱状图
3. 关联蛋白（蛋白表达无差异、磷酸化表达显著差异）的KEGG功能富集柱状图
4. 关联蛋白（蛋白表达显著差异、磷酸化表达显著差异）的KEGG功能富集柱状图
5. KEGG功能富集详情表

### KEGG通路整合图

为了直观比较在**某一特定通路中**，蛋白在两个组学层面的表达关系，将两个组学的KEGG通路图进行整合展示。

[pathway]

两组学差异蛋白的KEGG通路图示例

说明：图中每一个方框分为左右两部分，左边代表蛋白表达、右边代表磷酸化表达。红色表示上调, 蓝色表示下调，灰色表示无差异，紫色表示同一个修饰蛋白中同时存在上调、下调的修饰肽段。小圆圈表示小分子代谢物，大圆框代表其他通路。

输出文件

1. /KEGG通路整合图 （文件夹）

### 互作网络整合分析

将蛋白组与磷酸化组的数据进行整合的分子网络分析，能够帮助我们揭示蛋白与磷酸化蛋白之间潜在的协同功能关系与调控关系。例如，某个激酶发生表达量或磷酸化水平的改变，可能会导致其下游底物蛋白发生磷酸化修饰水平的改变。另一方面，可以通过网络节点的连接度，来寻找对整个网络系统影响最大的核心蛋白，帮助我们锁定最关键的蛋白进行后续功能验证。

本分析取关联蛋白中**差异蛋白**与**差异磷酸化蛋白**，共同绘制相互作用网络：不同形状的节点表示来源于不同组学的蛋白，不同颜色表示表达上下调，节点大小表示连接度大小，线表示两蛋白之间发生互作关系，如下图：

[ppi]

图 关联蛋白的相互作用网络

说明：图中圆圈表示仅在蛋白水平有显著差异；三角表示仅在磷酸化水平有显著差异；正方形表示在蛋白水平、磷酸化水平均有显著差异；图形加灰色外框表示存在“有无”差异。图中颜色表示差异变化，其中红色表示上调、蓝色表示下调、紫色表示蛋白中既有表达上调又有表达下调的修饰肽段。

图中线表示蛋白质与蛋白质之间的相互作用，图形大小表明该蛋白质连接度（即与某蛋白直接相互作用的蛋白质数目）。**通常来讲，连接度越大，该蛋白质发生变化时整个系统受到的扰动就越大，更可能是维持系统平衡和稳定的关键，为后续重点研究的候选蛋白质。**

将**连接度最高**的差异表达蛋白，进行连接度及蛋白表达、修饰表达信息等的展示，见下表：

表3 PPI互作分析表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Protein ID | Protein Name | Gene Name | PPI degree | Expressed regulated | Phosphopeptide regulated |
|  |  |  |  |  |  |

输出文件：

1. 关联蛋白的相互作用网络图
2. PPI互作分析表

## 激酶分析（限人、小鼠、大鼠）

蛋白激酶是细胞中调控生物学变化的一类重要功能蛋白，其往往通过催化下游蛋白（底物）发生磷酸化修饰而参与信号转导调控。因此，筛选激酶并进行深入分析具有重要的生物学意义，是磷酸化修饰研究通常最为关注的一类蛋白。对于激酶的分析，我们不仅要探究激酶的蛋白表达水平变化，也要关注激酶本身的磷酸化修饰变化对其活性的影响，例如：蛋白激酶家族通过逐级磷酸化方式，引起多级激酶（如MAP3K、MAP2K、MAPK）磷酸化级联激活，放大信号，引起细胞反应[1]。

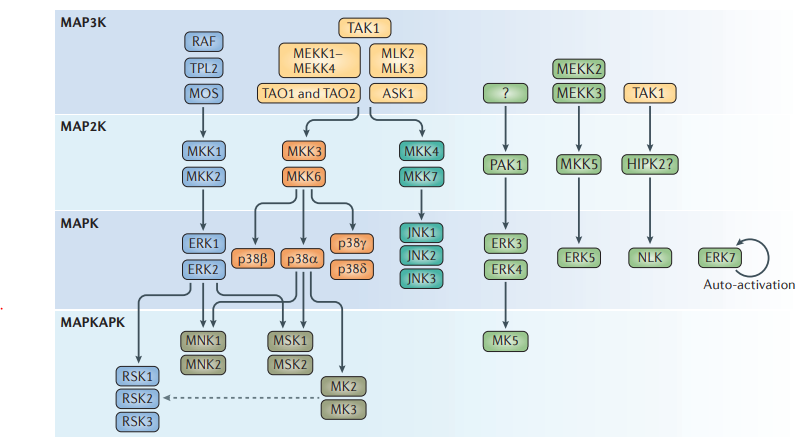


图 Mapk信号级联[1]

### 激酶注释

PhosphoSitePlus（PSP, http://www.phosphosite.org）是一个收录了蛋白质磷酸化位点信息的数据库，保存了大量人类、小鼠、大鼠的蛋白质磷酸化位点信息，主要来源于高通量筛选或已发表文献的信息。同样，激酶及激酶已知底物的信息也被收录至该数据库中，因此可以帮助我们进行激酶注释及激酶底物的注释和预测[2]。

为了筛选该研究中发挥作用的关键激酶，首先利用PhosphoSitePlus中的激酶底物数据库，对本项目鉴定到的蛋白和磷酸化蛋白分别进行激酶注释，并对激酶的数量进行分类统计，包括：（1）蛋白质组中，激酶及差异表达激酶的数量；（2）磷酸化修饰组中，激酶及差异修饰激酶的数量；（3）两个组学关联上的激酶，及差异表达激酶的数量，数据如下表。

表4 激酶的数量统计表

|  |  |  |  |
| --- | --- | --- | --- |
| **Quantified kinases** | **Differential expressed kinases** | **Quantified phosphorylated kinases** | **Differential expressed phosphorylated kinases** |
|  |  |  |  |

说明：Quantified kinases，可定量激酶，即注释为激酶的可定量蛋白；Differential expressed kinases，差异表达激酶，即注释为激酶的差异表达蛋白；Quantified phosphorylated kinases，可定量磷酸化激酶，即注释为激酶的可定量磷酸化蛋白；Differential expressed phosphorylated kinases ，差异表达磷酸化激酶，即注释为激酶的差异磷酸化蛋白；Correlated kinases，关联激酶，即在蛋白表达水平与蛋白磷酸化肽段水平均有定量信息；Correlated kinases(Differential expressed only in protein level)，仅蛋白水平有显著差异的关联激酶；Correlated kinases(Differential expressed only in phosphoproteins level) ，仅磷酸化水平均有显著差异的关联激酶；Correlated kinases(Differential expressed both in protein and phosphoprotein level)，在蛋白水平、磷酸化水平均有显著差异的关联激酶。

[venn\_kinase]

图 激酶的数量统计韦恩图

输出文件

1. 关联激酶数量统计表
2. 激酶的数量统计韦恩图

### 激酶的表达变化分析

为了区分激酶是否在两个组学层面均有差异变化，或者仅在一个组学层面存在差异变化，以火山图形式进行直观展示。如下图，激酶仅在蛋白表达水平有差异时，被标注为蓝色；激酶仅在修饰肽段水平有差异时，被标注为黄色；激酶在蛋白表达水平、修饰肽段水平均有差异表达时，被标记为紫色。

[volcano\_kinase]

关联激酶的表达差异火山图

说明：横坐标为差异倍数（以2为底的对数变换），纵坐标为差异的显著性P-value（以10为底的对数变换），左上区为显著差异下调，右上区为显著差异上调。其中激酶仅在蛋白表达水平有差异标注为蓝色，激酶仅在修饰肽段水平有差异标注为黄色，激酶在蛋白表达水平、修饰肽段水平均有差异标记为紫色。

为了分析比较组中，关联激酶在蛋白水平与磷酸化水平表达变化的**分布特征**，以及比较激酶的蛋白与磷酸化在整体水平上的**差异表达模式**的异同，将关联激酶进行聚类分析。

如下图，激酶的蛋白组表达展示在左侧，激酶的磷酸化组表达展示在右侧，颜色深浅表明差异倍数（fold change）大小。蛋白质组与磷酸化组颜色越接近，说明它们的表达趋势更接近；颜色差异越大，说明蛋白质组与磷酸化组的表达趋势差异更大。

[cluster\_kinase]

关联激酶的表达模式聚类分析

说明：图中每一行表示一个关联激酶，每一列表示一个比较组，其中左边为激酶在蛋白水平的表达倍数差异，右边为激酶在磷酸化水平的表达倍数差异。图中不同颜色表明上下调，颜色深浅为表达倍数差异大小，其中红色为表达上调，蓝色为表达下调。

输出文件：

1. 关联激酶的表达模式聚类分析

### 激酶与底物关系分析

组学结果的高度复杂性，往往会对数据挖掘造成较大的困扰。多组学的优势之一，是可以通过将多个组学的结果进行交叉相互验证，排除假阳性结果，进而帮助进行有效的筛选。例如，如果某个激酶的表达或修饰水平增加，同时其激酶底物的修饰水平也相应增加，则该激酶可进行后续重点关注和研究，进而揭示重要基因之间的相互联系，为寻找关键激酶靶点、阐述磷酸化分子信号机制提供信息。

首先，通过PhosphoSitePlus数据库，将数据结果中的激酶以及底物进行注释，并将激酶-底物关系进行对应。然后比较激酶的表达水平或激酶的修饰水平，与其对应底物的修饰水平的一致性。

下表分别展示了本组学结果中，激酶的蛋白表达差异，与该激酶对应的底物的磷酸化表达信息。

表5 激酶(蛋白表达水平)与底物关系表

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **KIN\_ACC\_ID** | **KINASE** | **GENE** | **KINASE\_FC (protein)** | **P-value (protein)** | **Phosphorylated**  **Site-FC**  **(备注：位置信息(氨基酸，磷酸化肽段差异倍数，Pvalue))** | **SUB\_ACC\_ID** | **SUBSTRATE** | **SUB\_GENE** | **Phosphorylated**  **Site-FC**  **(备注：位置信息(氨基酸，磷酸化肽段差异倍数，Pvalue))** |
|  |  |  |  |  |  |  |  |  |  |

说明：KIN\_ACC\_ID， 激酶ID；KINASE，激酶名称；GENE，激酶基因名；KINASE\_FC (protein)，激酶蛋白表达差异倍数；P-value (protein)，激酶差异表达统计检验；Phosphorylated Site-FC，激酶磷酸化修饰肽段差异表达信息，（位置信息(氨基酸，磷酸化肽段差异倍数，Pvalue))；SUB\_ACC\_ID， 激酶对应底物ID；SUBSTRATE，底物名称；SUB\_GENE，底物基因名；Phosphorylated Site-FC，底物磷酸化修饰肽段差异表达信息，（位置信息(氨基酸，磷酸化肽段差异倍数，Pvalue))。如果没有定量信息，则以空格表示。

接着，根据激酶及其预测底物关系筛选出本项目定量数据中存在的激酶及底物，绘制激酶与底物肽段关系弦图，直观反映本结果中激酶与底物的对应关系及表达情况。

[chord]

激酶与底物肽段关系弦图

说明：由外至内第一圈表示激酶（红圈）与底物（绿圈）；第二圈表示磷酸化肽段表达差异倍数；第三圈表示蛋白表达差异倍数；第四圈表示激酶或底物基因名，不同色块表示不同激酶以及底物基因；内圈连线表示激酶与底物对应关系，连线颜色与相应激酶对应。如果蛋白表达或磷酸化修饰没有定量信息，则不展示颜色。

输出文件：

1. 激酶(蛋白表达水平)与底物关系表
2. 激酶与底物肽段关系弦图

# 生物信息学分析方法 Bioinformatic Methods

## 聚类分析

首先对目标蛋白质集合的定量信息进行归一化处理（归一化到（-1,1）区间）。然后，使用Complexheatmap R包（R Version 3.4）同时对样品和蛋白质的表达量两个维度进行分类（距离算法：欧几里得，连接方式：Average linkage），并生成层次聚类热图。

## 注释分析

利用Blast2GO对目标蛋白质集合进行GO注释，过程大致可以归纳为序列比对（Blast）、GO条目提取（Mapping）、GO注释（Annotation）和InterProScan补充注释（Annotation Augmentation）等四个步骤。

利用KAAS (KEGG Automatic Annotation Server) 软件，对目标蛋白质集合进行KEGG通路注释。

## 富集分析

采用Fisher精确检验（Fisher’s Exact Test），比较各个GO分类或KEGG通路在目标蛋白质集合和总体蛋白质集合中的分布情况，对目标蛋白质集合进行GO注释或KEGG通路注释的富集分析。

## 蛋白质相互作用网络分析

基于IntAct (<http://www.ebi.ac.uk/intact/main.xhtml>)或者STRING（http://string-db.org/）数据库中的信息查找目标蛋白质之间的直接和间接相互作用关系，并使用CytoScape软件（版本号：3.2.1）生成相互作用网络并对网络进行分析。

## 激酶注释

基于PSP（<https://www.phosphosite.org//homeAction.action>）数据库对本项目中有定量信息的蛋白进行激酶注释与底物注释。

# 参考文献 References

[1] Hazzalin CA, Mahadevan LC. MAPK-regulated transcription: a continuously variable gene switch? **Nat Rev Mol Cell Biol.** 2002 Jan;3(1):30-40. Review.

[2] Hornbeck PV, Zhang B. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. **Nucleic Acids Res.** 2015. Jan;43(Database issue):D512-20.

# 组学期刊投稿指南（供参考）Publication Recommendation

* **第一档期刊（影响因子9 分以上）**

组学期刊：Nature Genetics，Genome Research， Microbiome，Molecular Systems Biology等

综合性期刊： Cell，Nature，Science，Nature Communications，PNAS，Cell Metabolism，Molecular Cell，Cell Research等

医学专业期刊：Nature Medicine，Cancer Cell，Cancer Research，European Heart Journal，Circulation，Journal of the American College of Cardiology，Gut，Gastroenterology，Hepatology，Diabetes Care，Molecular Psychiatry，Immunity 等

农林专业期刊：Nature Plants，Molecular Plant，Plant Cell等

**组学思路**：在这类高水平期刊中，组学驱动的研究主要包括：1.大队列研究，筛选临床生物标志物以及进行疾病分子分型等。2. 组学+后续功能机制研究。3. 结合时下的热点，如肠道微生物进行相关研究，阐明肠道微生物与疾病的相关性及相互作用机制。

* **第二档期刊（影响因子4~9分）**

组学期刊：Molecular & Cellular Proteomics，Metabolism，Metallomics，Gigascience，Proteomics & Bioinformatics等

综合期刊： Scientific Data， Scientific Reports，Cell Reports，Progress in Lipid Research，Frontiers in Microbiology，mSystems等

医学专业期刊：Advances in Cancer Research，International Journal of Cancer，Diabetologia，Thyroid，Diabetes，Oncogene，mBio等

农林领域专业期刊：New Phytologist，Plant Biotechnology Journal，Plant Physiology，Plant Journal，Plant Cell and Environment，Food Chemistry，Journal of Experimental Botany等

**组学思路**：发表在该分数段期刊的组学文章，1. 对于蛋白组学，修饰组学而言，要求组学+功能验证与机制研究（如表达量验证，基因敲除，点突变，位点特异性抗体验证，蛋白活性验证，后续功能实验，动物模型等）。2. 对于脂质组学是个例外，研究基础浅，新颖性高，纯脂质组学也能发表文章，无需后续验证。

* **第三档期刊（影响因子4分以下）**

组学期刊：BMC Genomics， Proteomics，Journal of Proteomics，Journal of Proteome Research，Metabolomics等

综合期刊：Plos One等

农林领域专业期刊: BMC Plant Biology，Journal of Agricultural and Food Chemistry，Frontiers in Plant Science等

**组学思路**：这类文章一般是最常见也是最简单的组学研究思路。1. 蛋白质组学方向：组学+差异蛋白WB或PRM验证。2.蛋白质翻译后修饰组学方向：则以修饰组学数据分析为主。3. 代谢组学方向：组学（+差异代谢物MRM验证）。

# 部分合作发表文章 Cooperative Projects Papers

* **医学：**

[1] Lysozyme-Assisted Photothermal Eradication of Methicillin-Resistant Staphylococcus aureus Infection and Accelerated Tissue Repair with Natural Melanosome Nanostructures. **ACS Nano**

[2]One-Carbon Metabolism Supports SAdenosylmethionine and Histone Methylation to Drive Inflammatory Macrophages. **Molecular Cell**

[3] Early Prediction of Developing Type 2 Diabetes by Plasma Acylcarnitines: Population-Based Study. **Diabetes Care**

[4] Hepatocellular Carcinoma-Associated Protein TD26 Interacts and Enhances Sterol Regulatory Element-Binding Protein 1 Activity to Promote Tumor Cell Proliferation and Growth. **Hepatology**

[5] Polyunsaturated fatty acids metabolism, purine metabolism and inosine as potential independent diagnostic biomarkers for major depressive disorder in children and adolescents. **Mol Psychiatry**

[6] CLOCK Acetylates ASS1 to Drive Circadian Rhythm of Ureagenesis. **Mol Cell**

[7] EGFR modulates monounsaturated fatty acid synthesis through phosphorylation of SCD1 in lung cancer. **Molecular Cancer**

[8] Protein phosphatase 2A has an essential role in promoting thymocyte survival during selection. **Proc Natl Acad Sci U S A**

[9] O-GlcNAcylation destabilizes the active tetrameric PKM2 to promote the Warburg effect. **Proc Natl Acad Sci U S A**

[10] Proteomic analysis of solid pseudopapillary tumor of the pancreas reveals dysfunction of the endoplasmic reticulum protein processing pathway. **Molecular & Cellular Proteomics**

* **植物：**

[1] Natural alleles of a proteasome α2 subunit gene contribute to thermotolerance and adaptation of African rice. **Nat Genetics**

[2]Maize Oxalyl-CoA Decarboxylase1 Degrades Oxalate and Affects the Seed Metabolome and Nutritional Quality. **Plant Cell**

[3] OsSPL3, a SBP-Domain Protein, Regulates Crown Root Development in Rice. **Plant Cell**

[4] RRM Transcription Factors Interact with NLRs and Regulate Broad-Spectrum Blast Resistance in Rice. **Mol Cell**

[5] Global analysis of sumoylation function reveals novel insights into development and appressorium-mediated infection of the rice blast fungus. **New Phytologist**

[6] Quantitative Phosphoproteomic and Metabolomic Analyses Reveal GmMYB173 Optimizes Flavonoid Metabolism in Soybean under Salt Stress. **Molecular & Cellular Proteomics**

[7] Label-Free Quantitative Proteomics of Lysine Acetylome Identifies Substrates of Gcn5 in Magnaporthe oryzae Autophagy and Epigenetic Regulation. **mSystems**

[8] Phototrophy and starvation-based induction of autophagy upon removal of Gcn5-catalyzed acetylation of Atg7 in Magnaporthe oryzae. **Autophagy**

[9] Integrative Transcriptome and Proteome Analysis Identifies Major Metabolic Pathways Involved in Pepper Fruit Development. **J Proteome Res**

[10] Proteomics integrated with metabolomics: analysis of the internal causes of nutrient changes in alfalfa at different growth stages. **BMC Plant Biology**

* **动医/动科/食品等其他：**

[1] Fecal Microbiota Transplantation Beneficially Regulates Intestinal Mucosal Autophagy and Alleviates Gut Barrier Injury. **mSystems**

[2] Quantitative Phosphoproteomic Analysis among Muscles of Different Color Stability using Tandem Mass Tag Labeling. **Food Chem**

[3] Comparative Analysis of Whey N-Glycoproteins in Human Colostrum and Mature Milk Using Quantitative Glycoproteomics. **J Agric Food Chem**

[4] N-glycosylation proteomic characterization and cross-species comparison of milk fat globule membrane proteins from mammals. **Proteomics**

[5] Metabolomics Reveals that Crossbred Dairy Buffaloes Are More Thermotolerant than Holstein Cows under Chronic Heat Stress. **Journal of agricultural and food chemistry**

[6] The Synergistic Effect of Exogenous Glutamine and Rifampicin Against Mycobacterium Persisters. **Frontiers in Microbiology**

