**标记与非标记定量蛋白质组分析流程**

**一、蛋白项目分类**

蛋白项目类型可分为：TMT一次标记项目，iTRAQ一次标记项目，label free项目，DIA项目，磷酸化项目，含有内标校正的项目，TMT多组标记项目，iTRAQ多组标记项目，以及外来数据。

流程作业将面对的类型分为三种情况：

1．single\_mark：一次标记项目(包含：TMT一次标记项目，iTRAQ一次标记项目)

2．multi\_mark:：重复标记项目(包含：TMT多组标记项目，iTRAQ多组标记项目）

3．label free：非标记项目(包含：label free项目，DIA项目)

**说明1**：因为磷酸化会在以上三种类型中出现，内标校正会在1. 2. 类型中出现，流程将会自动检测，做出相应处理。磷酸化项目以["Modifications"]为检测标记，内标校正以["Mix"]为检测标记。

**说明2**：外来数据可根据分析要求按类型构造标准输入。

**二、标准输入**

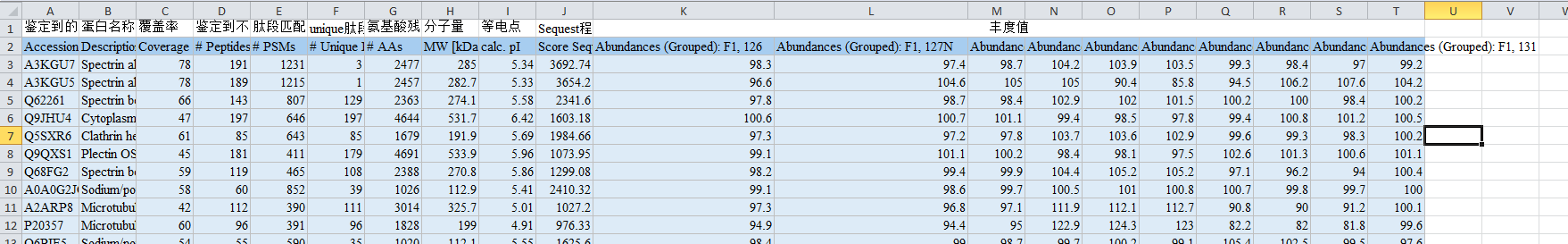
标准输入文件：

**1. 蛋白定量文件** (Protein quantitation.xlsx 需提供"**.xlsx**"格式，且文件名以包含"**Protein quantitation**"为标识)

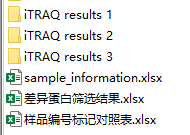
例如文件名：



文件内容：



多组标记时，需将整个文件夹上传华为云



打开文件夹中的内容如下：

C:\Users\LM\AppData\Roaming\Tencent\Users\934903250\TIM\WinTemp\RichOle\3ZT8HC2)YEFX{ZS{Y0F)PLC.png

**2．样品标记对照表**（需提供"**.xlsx**"格式，且文件名以"**对照表**"为标识）

文件名：



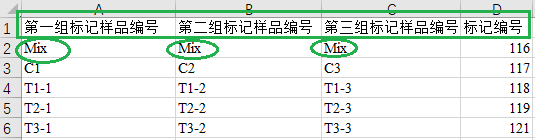
文件内容：

**单组标记：**



**多组标记：**

绿色框中的为程序识别关键，需要严格按照例子来写，如下：

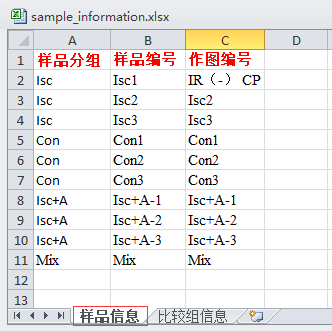
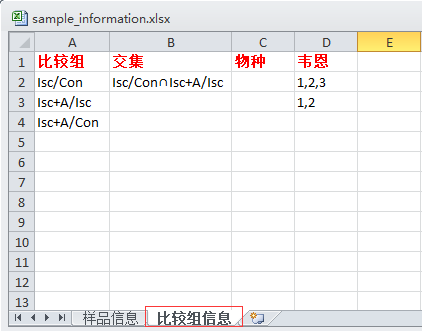


**3．sample\_information** (sample\_information.xlsx，此文件集中提供样品信息，和分析要求)

文件名：



文件内容：

说明：上图示例中标红字体均不可变，其中“比较组信息”表格中“交集”，“韦恩”列为可选参数，“物种”列为保留参数，没有分析要求参数位置保留内容可以不填写。

**流程作业每种情况都有相应的标准输入**

对应情况如下：

|  |  |  |
| --- | --- | --- |
| **面对类型** | **标准输入** | **处理与否** |
| single\_mark | 1,2,3 | 无需处理 |
| multi\_marke | 1,2,3 | 无需处理 |
| label free | 1,3 | 将蛋白定量文件名改为"数据矩阵",sheet名改为"数据矩阵"，且保证表头只有一行, 若有空值(NA)以0表示。 |

**三、流程作业**

标准输入文件集中上传华为云工作目录

配置临时环境: export PATH=/public/hstore5/software/anaconda3/bin:$PATH

进入环境: source activate prap

退出环境： source deactivate

若已在~/.bashrc中添加环境变量，则直接使用runprap 进入环境。（endparp退出环境。）

**第一部分：**

**step 1: 差异筛选/统计分析**

prap -r 1.5 -f single\_mark

prap -r 1.5 -f multi\_mark -i 绝对路径

prap –screen -r 1.5 -f multi\_mark -i绝对路径

prap2 -r 1.5 -f ‘label free’

注： prap 命令主要参数有“筛选差异倍数”[-r | --ratio]：1.2 | 1.5 | 2 ，“面对的数据类型”[-f | --form] single\_mark | multi\_mark | label free。

其他参数见prap [-h | --help] 。

**step 2: 图形展示 1**(火山图，聚类热图，韦恩图，上下调数量柱状图)

unionplot -f 1.5

注：unionplot命令必要参数是[-f ]差异筛选倍数，缺失时默认值为1.5。其他参数取默认值。

若单独作图参见：unionplot [-h | --help]。

**图形展示2**(相关性图，蛋白分子质量分布，肽段数鉴定到的蛋白数分布，肽段对蛋白的覆盖度饼图)

**脚本目录：**

**/public/hstore5/proteome/pipeline/script/R\_Plotting\_Pipeline\_\*\***

cp plotting\_v\*.R cp runplotting.sh 到当前项目目录下，

sh runplotting.sh

**step 3: 准备背景文件（功能富集分析背景文件（reviewed+unreviewed））**

step-sql-all -s mmu -p 1

(-p 1 为真核KOG, -p 0 为原核COG)

step-sql2 -s mmu （1.0版背景文件准备命令，只有reviewed库，没有COG/KOG）

cp /背景文件/\* ./

注：step-sql命令提供唯一必要参数[-s] KEGG物种简写

**第二部分：互作分析**

**脚本路径：**

**/public/hstore5/proteome/pipeline/script/enrich\_ppi\_script\_2019/PPI\_network**

**需cp 脚本文件夹到项目路径中**

**cp -r /public/hstore5/proteome/pipeline/script/enrich\_ppi\_script\_2019/PPI\_network ./**

**Step 4: 准备蛋白互作文件（protein2protein.xls）**

p2p -taxid 10090

注：p2p命令提供为唯一必要参数[-taxid]：物种的NCBI-taxid编号

**step 5: 生成ppi作图数据信息**

perl PPI\_network/ppi\_network.pl -i ./ -g protein2protein.xls -o ppi

**step 6: 生成ppi图**

perl PPI\_network/ppi\_network\_html.pl -i ppi/ -o ppi/

**Step7 处理背景文件（**将uniprot id 🡪 uniport id : gene name）

pc

**第三部分：富集分析**

**脚本路径：**

**/public/hstore5/proteome/pipeline/script/enrich\_ppi\_script\_2019/enrichment\_of\_different\_expressed\_gene**

**step 8:** **准备Unigene.KEGG.Classification.xls**

perl /public/hstore5/proteome/pipeline/script/enrich\_ppi\_script\_2019/enrichment\_of\_different\_expressed\_gene/format\_keggbackgroud.pl -i gene\_kegg.backgroud.xls -s mmu -o Unigene.KEGG.Classification.xls

**step 9: 准备Unigene.GO.classification.stat.xls**

cp /public/hstore5/proteome/pipeline/script/enrich\_ppi\_script\_2019/enrichment\_of\_different\_expressed\_gene/category.xls ./

perl /public/hstore5/proteome/pipeline/script/enrich\_ppi\_script\_2019/enrichment\_of\_different\_expressed\_gene/format\_gobackgroud.pl -i gene\_go.backgroud.xls -c category.xls -s Unigene -o ./

**step 10:** **富集分析 (选择物种)**

perl /public/hstore5/proteome/pipeline/script/enrich\_ppi\_script\_2019/enrichment\_of\_different\_expressed\_gene/5.2.enrich\_go\_kegg.pl -infile \*-vs-\*-diff-\*.xls -go\_bg gene\_go.backgroud.xls -category category.xls -kegg\_bg gene\_kegg.backgroud.xls -html /public/hstore5/metabonomics/LM-database/kegg/mmu -anno\_kegg gene\_anno-kegg.backgroud.xls -outdir enrich/ -go\_lv2 Unigene.GO.classification.stat.xls -kegg\_lv2 Unigene.KEGG.Classification.xls

完成！