

■ 传统蛋白质组学研究

目前基于质谱仪器的蛋白质组学研究多数利用EBI的IPI、NCBI的NR和SWISSPROT作为数据库对蛋白进行检索鉴定。因此，在已获取这三类数据库的实验鉴定结果后，我们可以实现以下分析：

● 数据库交叉注释批量查询

将IPI号转化成：
UniProtKB/Swiss-Prot, UniProtKB/TrEMBL, Refseq, GI, ENSEMBL, UniGene, Entrez Gene(Gene ID)和Symbol等。
将SP号转化成：
EMBL, IPI, Refseq, GI, ENSEMBL, UniGene, Entrez Gene(Gene ID), Symbol, IntAct, GeneCards, MIM等。

● 数据库功能注释查询

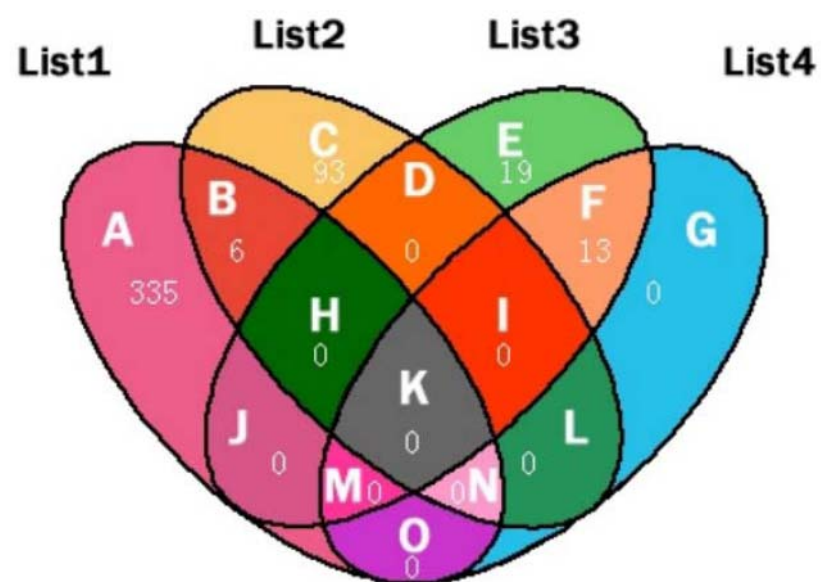
GO(Gene Ontology) :Biological Process ,Subcellular Location ,Molecular Function ;
蛋白修饰信息，包括：磷酸化、糖基化、甲基化、乙酰化等
相互作用蛋白，包括：String , IPA等
通路信息，包括：KEGG, Biocarta, Protein Lounge, Pathway_Interaction_DB等。

● ID 号比较

基于多批数据集之间的比较，采用数据交叉表和venn图方式表示数据集之间的交盖性。

	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8
No. 1	0	658	773	1101	687	629	622	768
No. 2	0	0	947	1193	773	748	737	963
No. 3	0	0	0	1682	972	886	884	1239
No. 4	0	0	0	0	1068	967	948	1389
No. 5	0	0	0	0	0	765	768	1156
No. 6	0	0	0	0	0	0	736	1054
No. 7	0	0	0	0	0	0	0	1002
No. 8	0	0	0	0	0	0	0	0

表一 数据集之间交盖度



图一. 数据集之间venn图

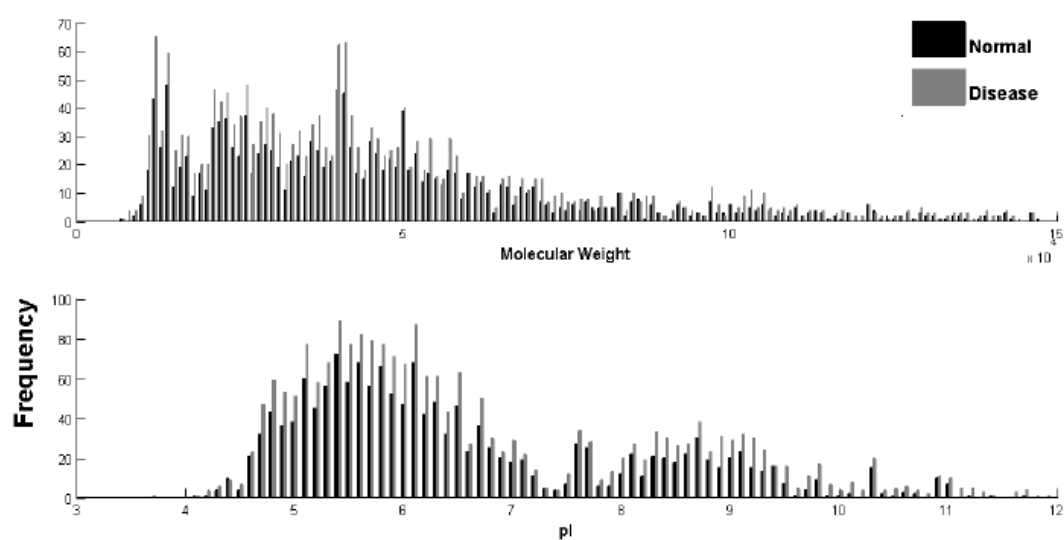
● 蛋白序列分析

许多数据库可能没有提供相关蛋白的功能注释，我们可以利用蛋白序列预测相关的功能。

序列相关的预测涉及：

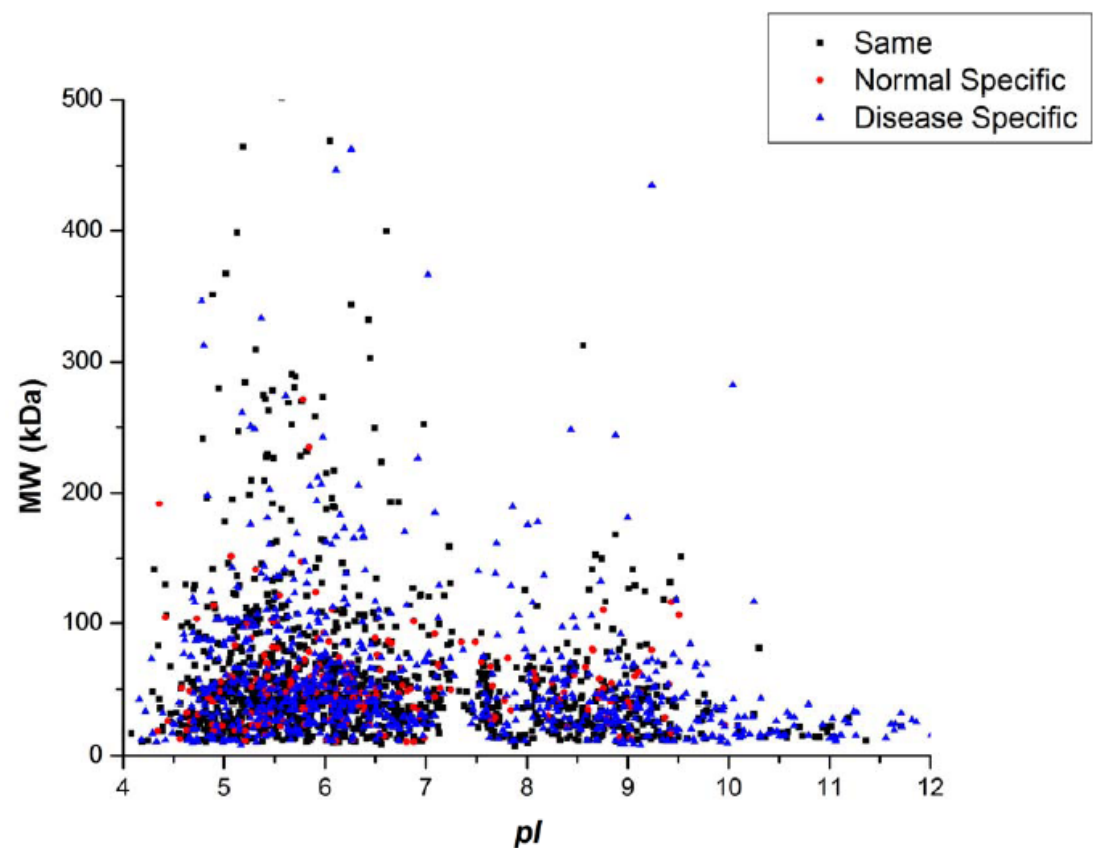
膜蛋白和跨膜区段预测；亚细胞定位预测； 信号肽预测； 分泌蛋白预测；

翻译后修饰预测：磷酸化位点、N-糖基化位点、O-糖基化位点、阴阳位点（即磷酸化和O-糖基化竞争位点）预测。



图二. 鉴定蛋白的理化性质分布图

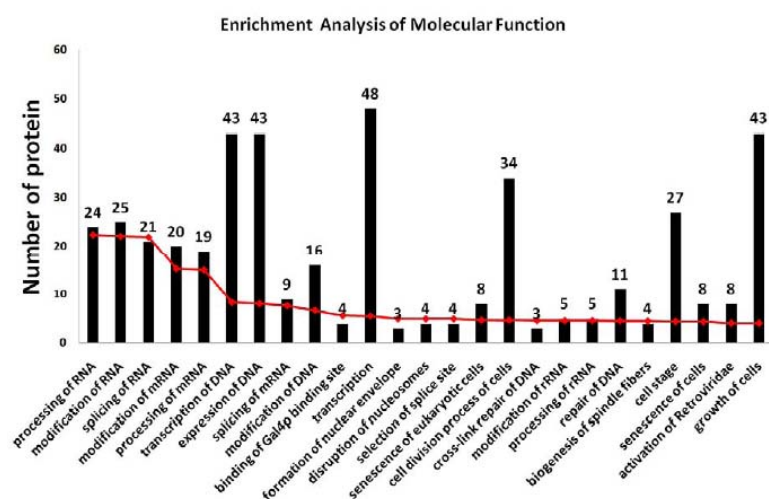
根据蛋白序列计算出相关理化性质，包括分子量、等电点、疏水性、酸性氨基酸数目、碱性氨基酸数目和蛋白长度，并构建对应的分布图。



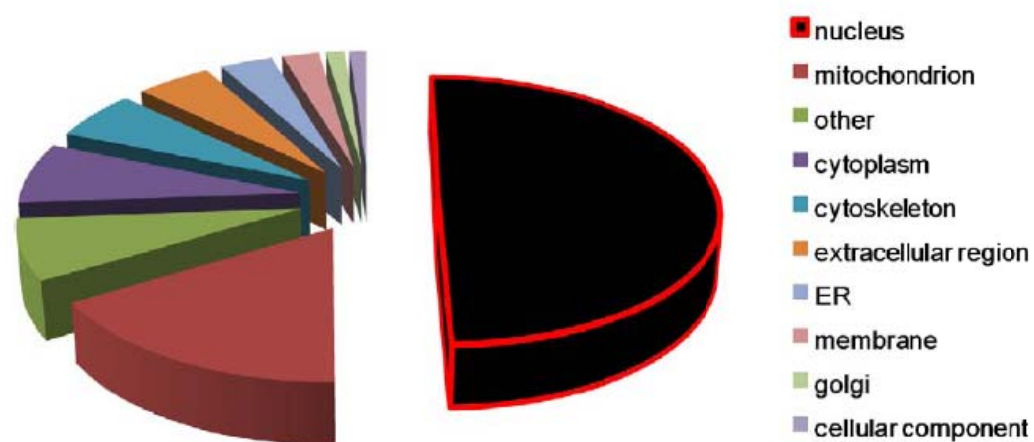
图三. 分子量和等电点在样本之间的比较

● 基因富集度计算 (Gene Ontology)

在蛋白组学中，对大规模的鉴定结果经常用pie图或bar图表示蛋白的某个分类，如亚细胞定位、生物功能、生物通路等。有时也会考虑这些分子在哪些疾病分子交盖比例过高。



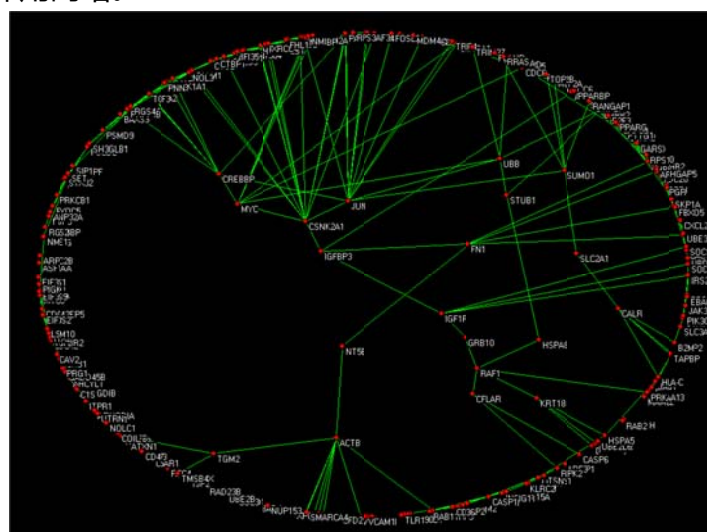
图四. Bar图表示鉴定蛋白在分子功能上的富集度



图五. Pie图表示鉴定蛋白在亚细胞定位上的分类

● 网络分析

通过蛋白的差异谱或实际鉴定谱，可以在各种相互作用数据库中找到对应的相互作用蛋白，并构画出相互作用网络。



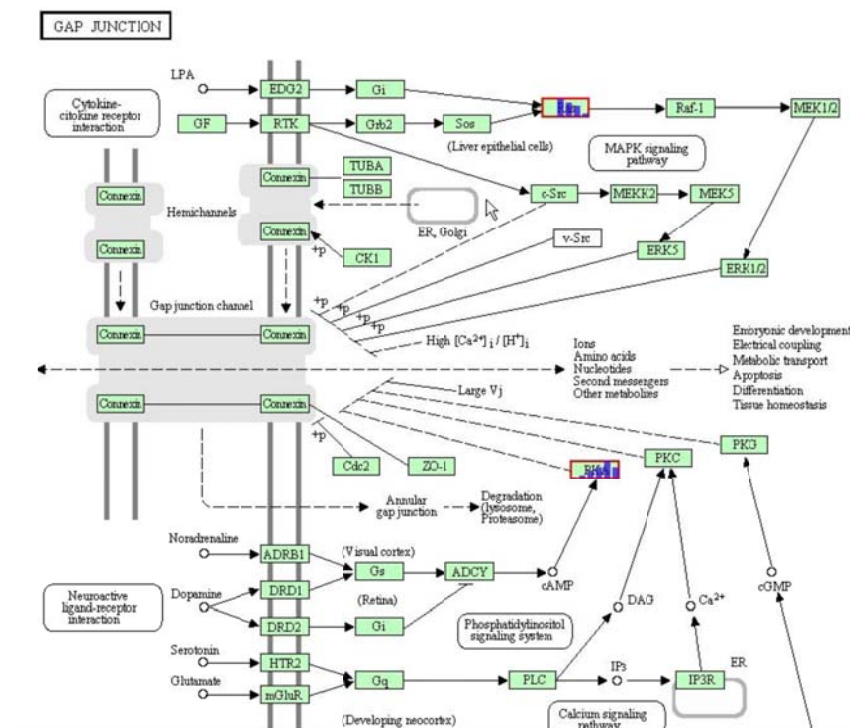
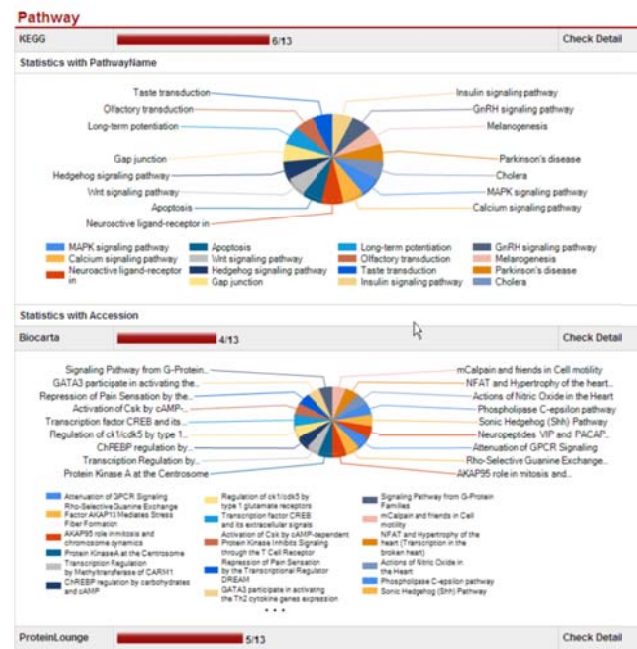
图六. 鉴定蛋白构建相互作用网络

● 通路分析

对已鉴定蛋白可以mapping到通路数据库中，给出mapping的统计结果并图示化。

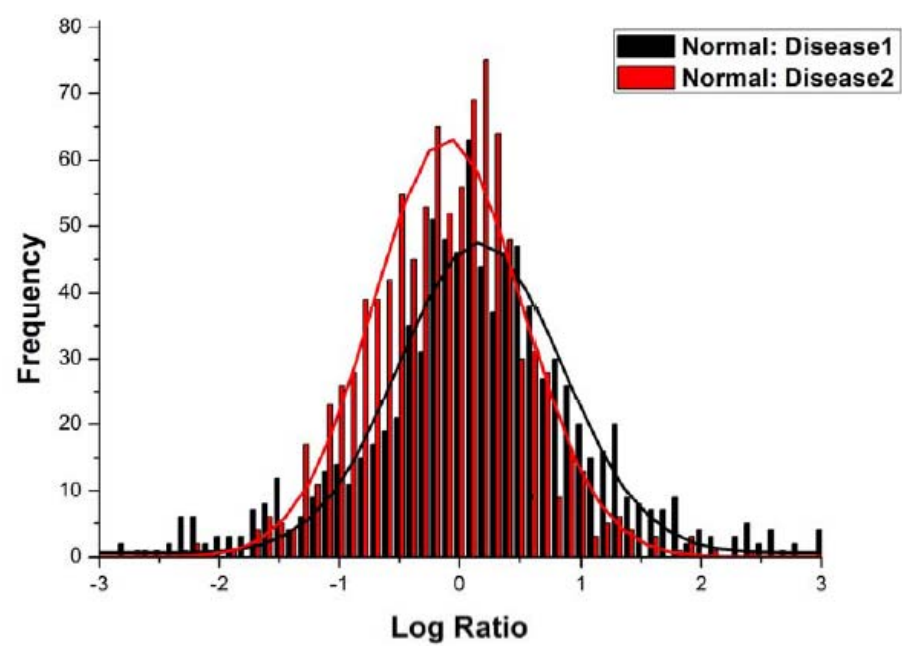
通路数据库来源：

KEGG



●差异蛋白筛选

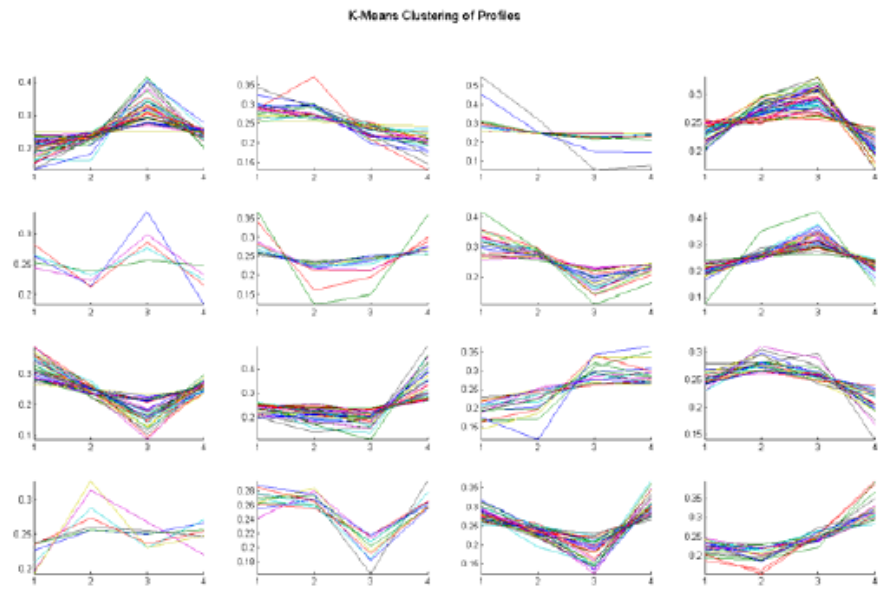
在两两比较的比较蛋白质组学中，要找出差异的蛋白列表，可以用正态分布拟合方法筛选。



图九. 蛋白相对定量的数据通过log转换拟合成正态分布 ,再通过正态分布的95%和99%置信线对表达差异的数据进行筛选。

●共表达模式挖掘

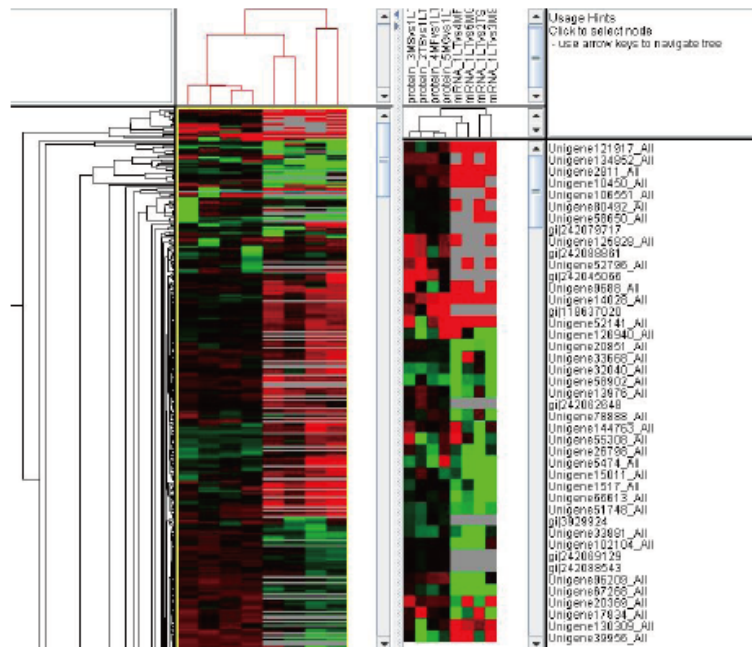
在四标或八标的Itraq定量中，可以用Kmean和SOM等非监督方法，挖掘共表达趋势的蛋白类别。



图十. 用Kmean对蛋白相对定量的数据共表达模式筛选

●分层聚类挖掘

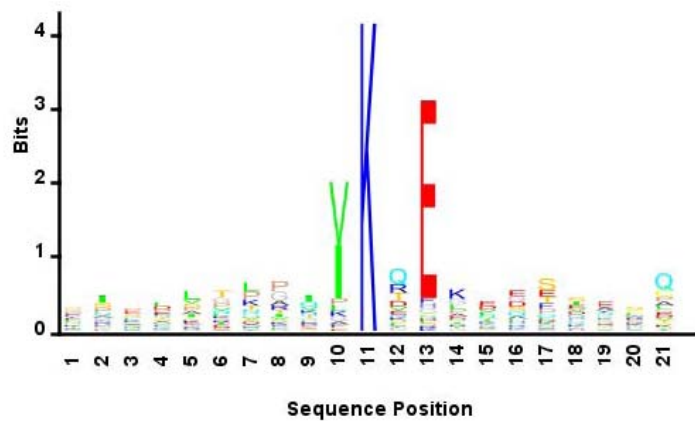
在四标或八标的Itraq定量中，展示表达趋势相近的蛋白。



图十一. 用聚类热图展示蛋白相对定量数据中表达趋势相近的蛋白。

■ 修饰蛋白质组学研究

- 从质谱数据进行数据库搜索，确定蛋白质翻译后修饰的位点和相对量的变化
- 蛋白特定的修饰鉴定后，想挖掘修饰位点附近氨基酸组成的状况，可以用聚类图或氨基酸比例图展现。

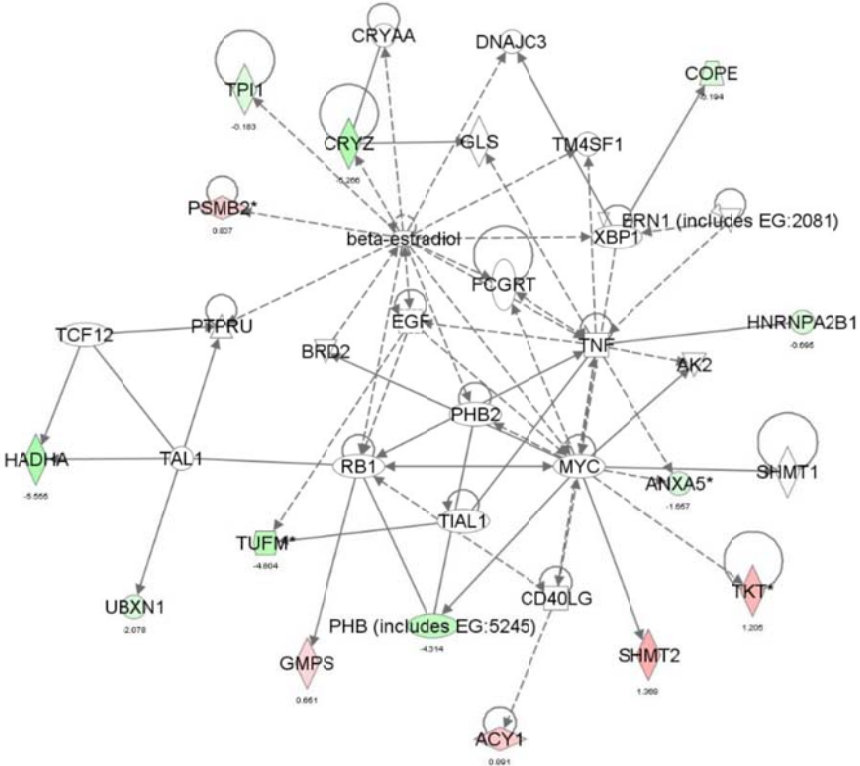


图十二. 用概率图形式展示靶标修饰氨基酸附近的氨基酸组成情况

■高级分析

●IPA 分析

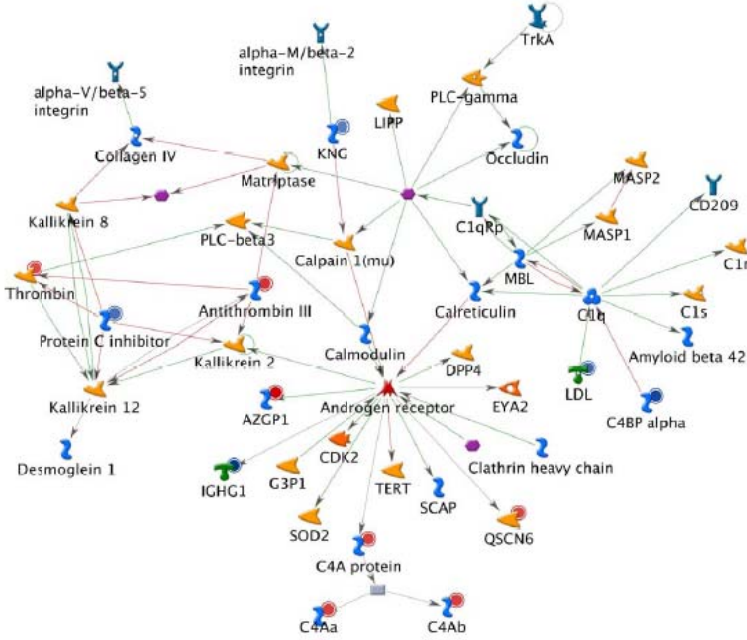
IPA是一种公认的生物数据分析工具，非常适用于疾病相关的数据分析。它的主要特色在于根据实际数据找到疾病分子的网络和通路，甚至在其基础之上构建自己的网络和通路。



图十三. IPA捕捉到差异分子形成网络

●Metacore 分析

是IPA的竞争对手，基本上能实现IPA的所有分析。



图十四. 构建相互作用网络

