



及因生物
— TGENE —

免疫肽组学结题报告



及因（上海）生物科技有限公司

Tgene Biotech (Shanghai) Co.,Ltd.

- 1 样品信息
- 2 实验材料
- 3 样品前处理
- 4 LC-MS/MS检测与数据库检索分析
- 5 实验结果
 - 5.1 肽段检索结果
 - 5.2 数据质控及可信分析
 - 5.3 数据处理后的数据情况
 - 5.4 个性化分析
 - 5.4.1 差异统计
 - 5.4.2 差异肽段火山图
 - 5.4.3 差异肽段表达水平聚类分析
 - 5.4.4 差异肽段motif分析

免疫肽组学分析报告

项目编号：

客户名称：xxx

报告时间：

1 样品信息

| NO. | 样品名称 | 样品类型 |
|-----|---------|------|
| 1 | Control | XXX |
| 2 | Control | XXX |
| 3 | Control | XXX |
| 4 | Case | XXX |
| 5 | Case | XXX |
| 6 | Case | XXX |

2 实验材料

2.1 主要试剂和耗材

| 试剂或耗材 | 品牌 |
|----------------|----------------------|
| 乙腈Acetonitrile | Merck |
| 甲酸Formic acid | Thermo fisher |
| C18 Zip-Tip除盐柱 | Agilent Technologies |

2.2 主要仪器

| 仪器 | 型号 | 品牌 |
|---------|-----------------------|-------------------|
| 冷冻离心机 | Fresco 17 | Thermo Scientific |
| 真空冷冻干燥机 | Savant DNA120 | Thermo Scientific |
| 纳升液相系统 | Easy-nLC1200 | Thermo Scientific |
| 高分辨质谱仪 | Orbitrap Fusion Lumos | Thermo Scientific |

3 样品前处理

3.1 MHC抗体偶联beads

将溶解在Antibody coupling buffer 中的MHC-I或MHC-II抗体和洗涤后的protein A/G beads 混合，室温摇晃孵进行偶联以制作MHCI-beads或MHCII-beads；

3.2 样本裂解

根据样本实际类型是组织还是细胞将在以下两个操作中二选一进行。

3.2.1 组织裂解

1. 组织研磨：使用研钵将组织研磨成细粉；
2. 组织裂解：加入CHAPS裂解液，4度慢速摇裂解；
3. 蛋白定量：离心取上清裂解液进行BCA法蛋白定量。

3.2.2 细胞裂解

1. 细胞裂解：在细胞沉淀物中加入CHAPS裂解液，4度慢速摇裂解；
2. 蛋白定量：离心取上清裂解液进行BCA法蛋白定量。

3.3 MHC-II结合肽免疫亲和富集

1. 将蛋白裂解液加入MHCII-beads中，4度慢速转动孵育，收集上清蛋白后续的MHC-I结合肽的免疫亲和富集；
2. 使用wash buffer进行清洗；
3. 使用洗脱液进行洗脱。

3.4 MHC-I结合肽的免疫亲和富集

1. 将3.3第一步中收集的上清蛋白裂解液加入MHCI-beads中，4度慢速转动孵育；
2. 使用wash buffer进行清洗；
3. 使用洗脱液进行洗脱。

4 LC-MS/MS检测与数据库检索分析

4.1 LC-MS/MS检测

多肽样品经离心干燥后，重新溶解于Nano-LC 流动相A（0.1%甲酸/水）中装瓶上样，进行在线LCMS分析。溶解后的样品以合适的体积上样到nanoViper C18预柱上（3μm， 100Å），然后20ul体积冲洗脱盐。液相为Easy nLC 1200纳升液相系统（ThermoFisher, USA），样品在预柱上脱盐保留后再经分析柱分离，分析柱规格是C18反相色谱柱（Acclaim PepMap RSLC, 75 μm×25 cm C18-2 μm 100 Å），实验所用梯度为90 min内流动相B（80%乙腈，0.1%甲酸）由5%升高至38%。质谱采用ThermoFisher Orbitrap Fusion Lumos系统（ThermoFisher, USA）结合纳升喷雾Nano Easy-Spray离子源（ThermoFisher, USA），喷雾电压为2.2 kV，离子传输管加热温度为300 °C。质谱扫描方式为信息依赖的采集工作模式下（DDA，Data Dependent Analysis），一级质谱扫描分辨率为60000，扫描范围350-1250m/z，最大注入时间为100ms。每个DDA循环为3s，采集电荷为2+到5+的二级图谱，二级质谱离子最大注入时间为50ms。碰撞室能量（高能碰撞诱导解离，HCD）设定为30eV，适用于所有前体离子，动态排除设置为60秒。

4.2 数据库检索分析

质谱采集到的原始图谱文件，采用Peaks studio 11软件进行数据加工处理和检索分析，检索参数设置如下：非酶解，一级质谱质量容差为10ppm，二级质谱为0.05Da。具体情况见下列表格。

| 选项 | 参数 |
|---------------|-----------------|
| Software | PEAKS Studio 11 |
| MS1 tolerance | 10ppm |
| MS2 tolerance | 0.05Da |

| 选项 | 参数 |
|----------------------|---------------|
| Enzyme | none |
| Missed cleavage | none |
| Static modification | none |
| Dynamic modification | Oxidation (M) |

5 实验结果

5.1 肽段检索结果

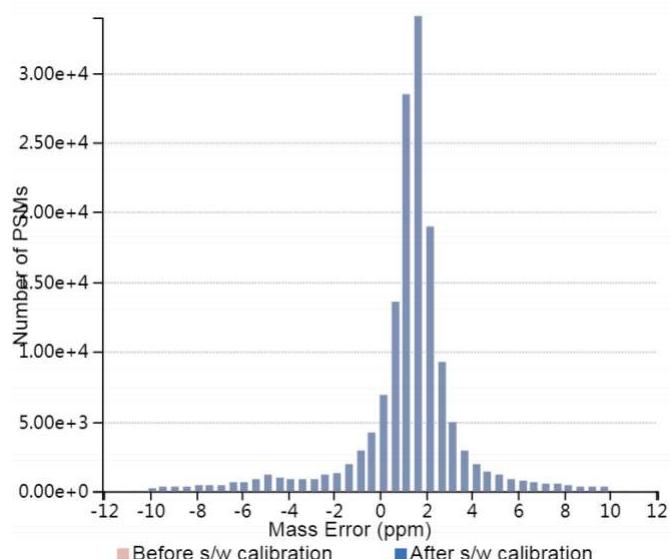
(1) mouse 中肽段水平的检索结果为: data_unique.csv (./mouse/data) ; HELPX 中肽段水平的检索结果为: data_unique.csv (./HELPX/data)

(2) 所用数据库序列文件为: uniprot-Mouse-HELPX (./)

5.2 数据质控及可信分析

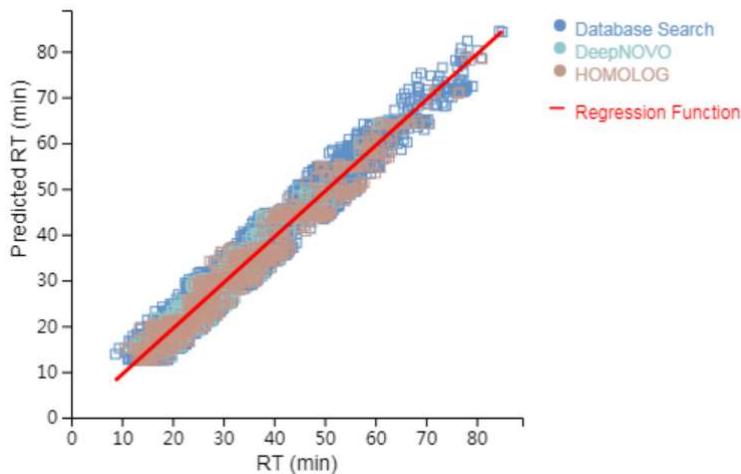
以下为肽段鉴定过程中质控 (QC) 信息, 包括一级质量误差分布、谱图数肽段数分布、肽段鉴定覆盖度分布等:

5.2.1 一级质量误差分布



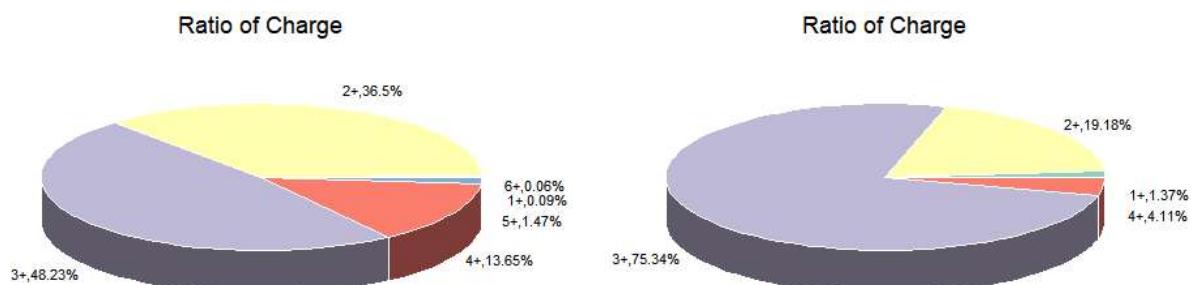
图片说明: 横坐标表示质量误差的 ppm 值, 纵坐标表示具有该误差值的肽匹配图谱的数量, 不同颜色的柱状图分别代表校准前后所有匹配到的肽段的一级质量误差分布情况, 可以观察到大多数肽段的质量误差在 10 ppm 以内, 校准前后无差异, 表明质谱数据的质量和精确度较高, 不会因为质量偏差过大而影响到蛋白的定性定量分析。

5.2.2 肽保留时间与预测保留时间的散点图



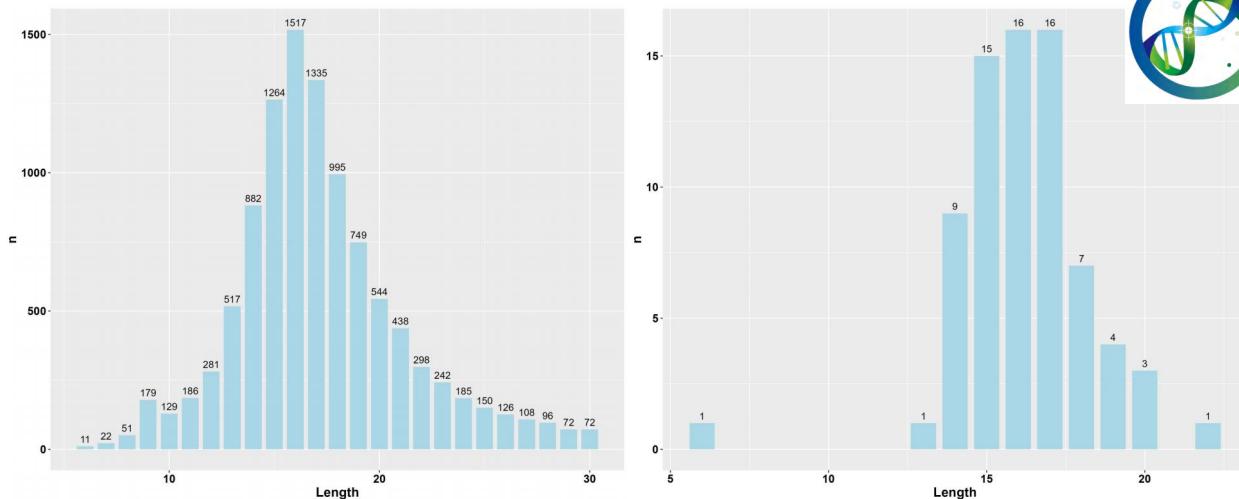
图片说明：横坐标表示实际的肽段保留时间，纵坐标表示预测的肽段保留时间，图上的散点颜色代表肽段的来源（database search：使用数据库搜索方法鉴定的肽段；homolog：与数据库中蛋白序列具有相似序列的肽段；DeepNOVO：除前两种肽段外通过DeepNOVO算法从质谱数据中预测出的多肽序列），理论上所有的数据点应该紧密地围绕这条斜率为1的直线。

5.2.3 肽段电荷分布



左图为mouse,右图为HELPX

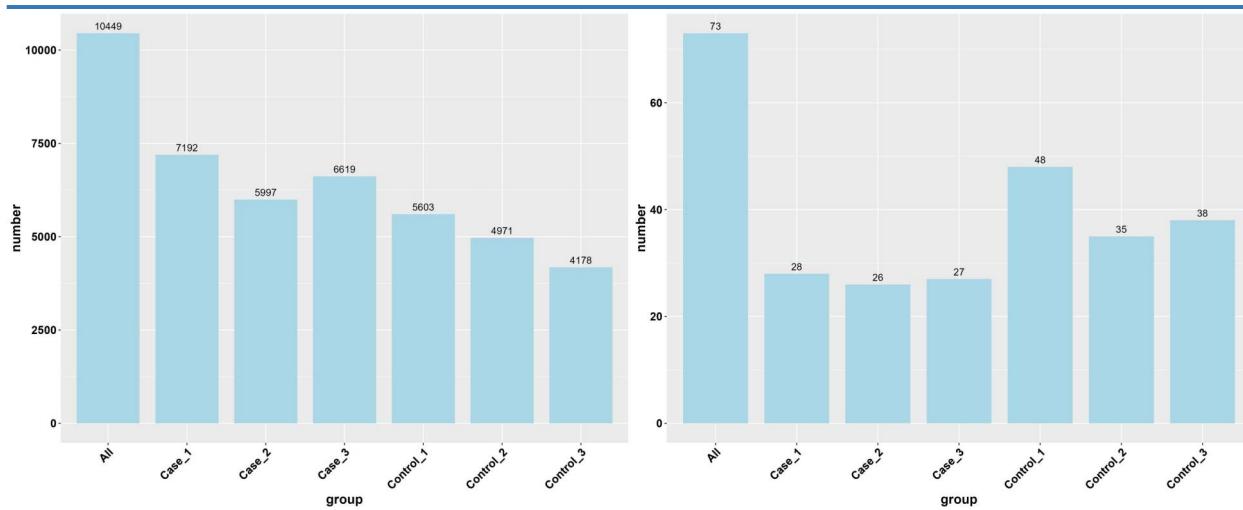
5.2.4 肽段长度分布



左图为mouse,右图为HELPX

图片说明：横坐标为肽段长度，纵坐标为对应的肽段数量。

5.2.5 多肽鉴定总数及在各样品数目分布：



左图为mouse,右图为HELPX

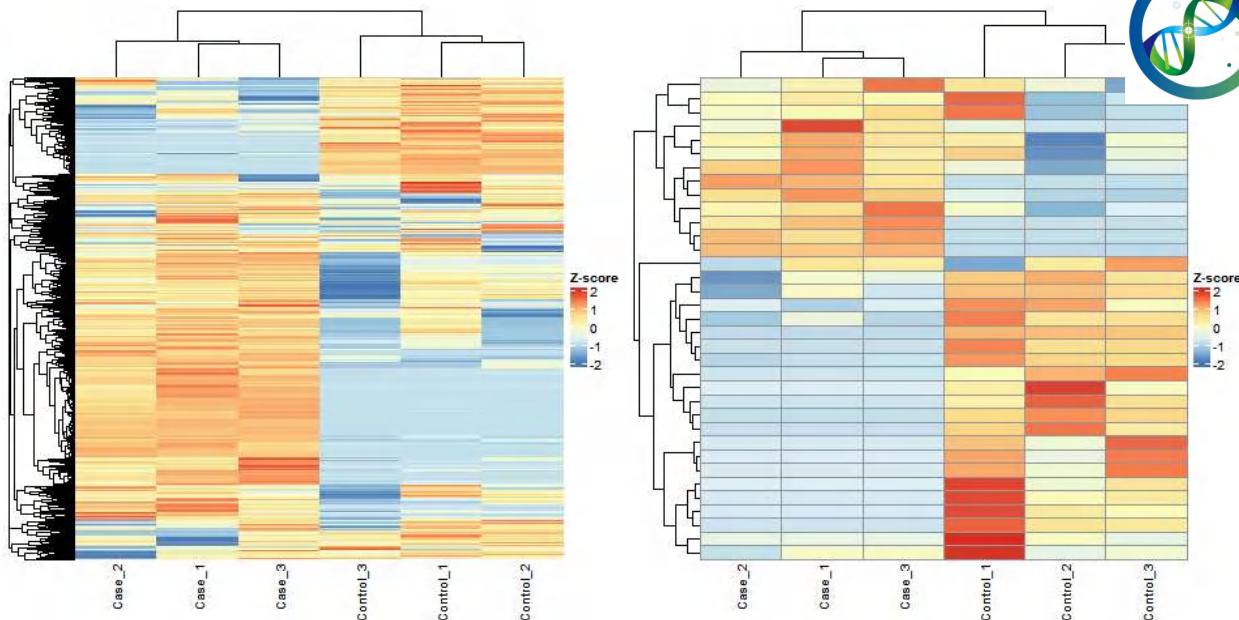
图片说明：横坐标为样本，纵坐标为对应的样本鉴定到的肽段数量。

5.3 数据处理后的数据情况

利用数据库检索得到原始数据表格后，经过空值过滤、填充等一些数据处理过程，对样品定量到的可信肽段的信息进行整体评估，包括肽段强度分布，样品间关系等。

5.3.1 丰度热图

层次聚类 (Hierarchical Clustering) 是聚类算法的一种，通过计算不同类别数据点间的相似度来创建一棵有层次的嵌套聚类树。聚类热图对数据和样品进行聚类，可以用来进行实验数据的质量控制。一般地，同一组样品能通过聚类出现在同一簇中。

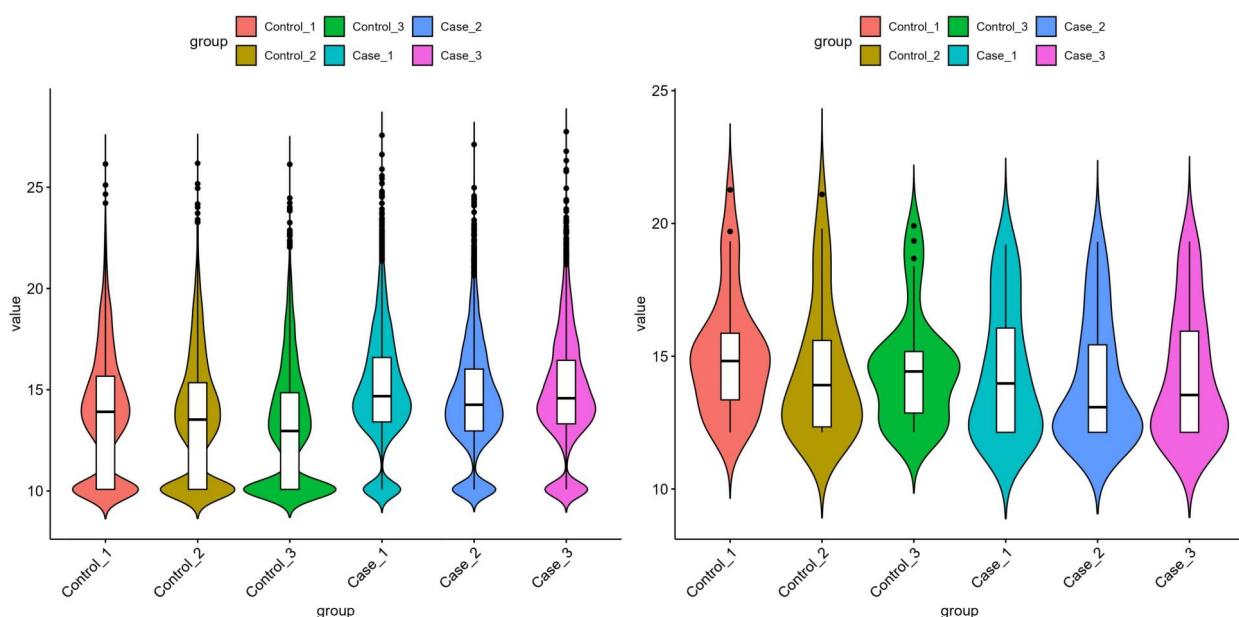


左图为mouse,右图为HELPX

图片说明：横坐标代表样本，纵坐标代表蛋白，颜色反映了肽段的丰度（z-score），颜色越红表示肽段的丰度越高，越蓝说明丰度越低。

5.3.2 丰度小提琴图

将每个样品所有多肽的强度进行箱线图和密度分布图展示，可以看到每个样品所有多肽强度的中位数、动态范围等信息，也了解样品间分布的差异。详细见下图，value值为log2强度值处理展示。

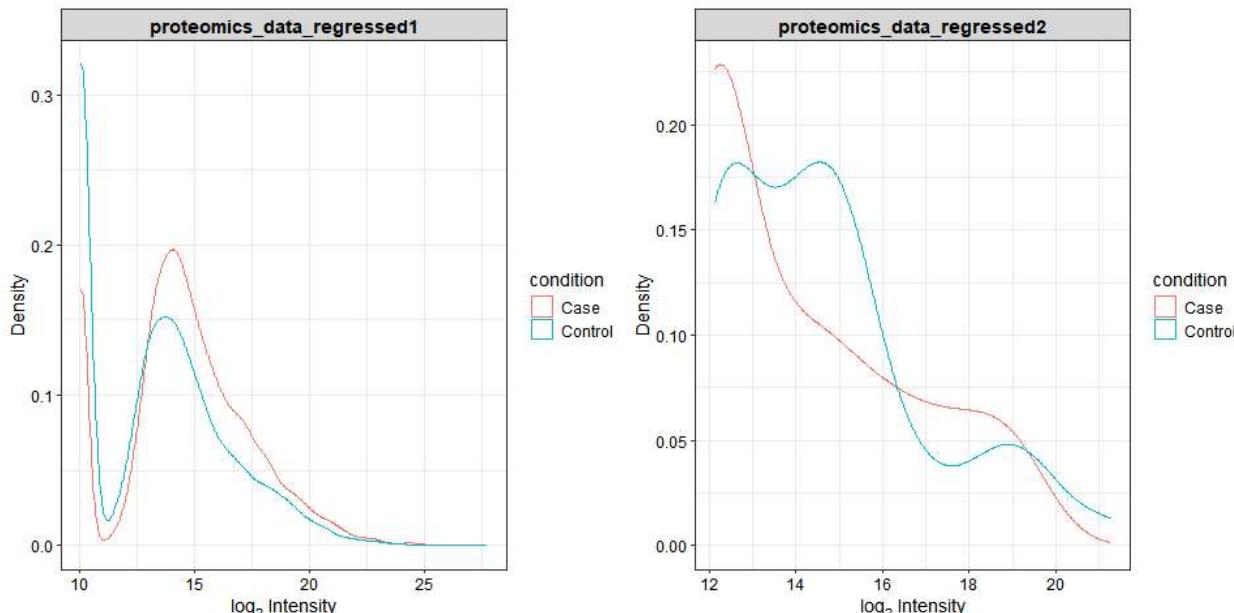


左图为mouse,右图为HELPX

图片说明：横坐标为样品名，纵坐标value值为log2强度值处理展示。小提琴图（ViolinPlot）是一种常用于数据可视化的图表类型，它结合了箱形图和密度图的特点，能够直观地展示数据的分布情况。横坐标代表样本，纵坐标代表肽段丰度（log2处理后），可同时展示数据的分布情况和统计信息。

5.3.3 核密度图

每条曲线表示该样品表达值的概率分布情况，即疏密程度，“峰”越高，数据越密集，密度越高。

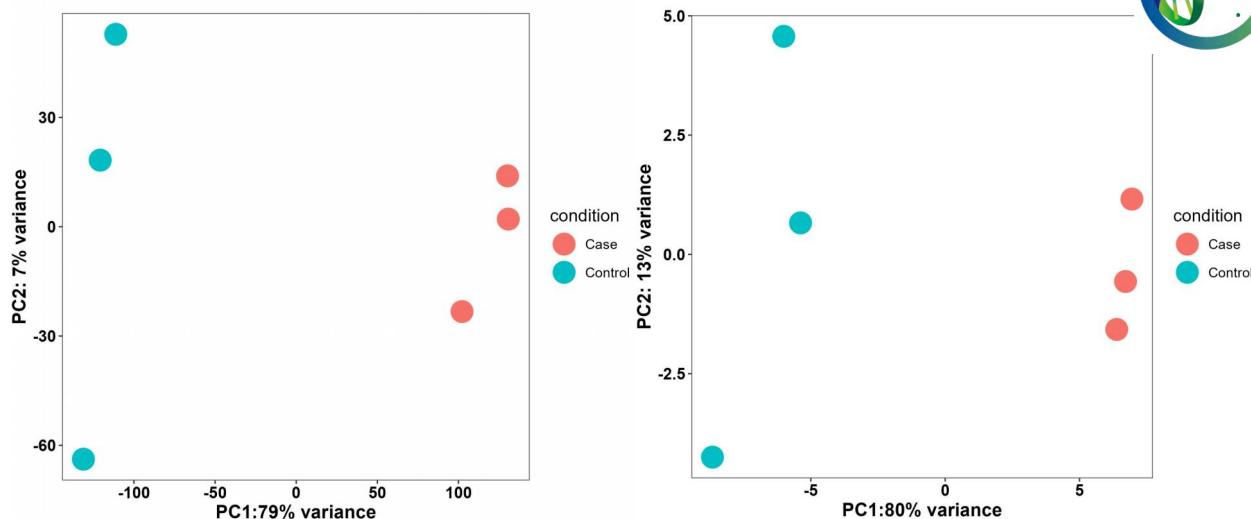


左图为mouse,右图为HELX

图片说明：核密度图（Kernel Density Plot）是一种用于估计数据分布的图形工具，它通过平滑数据点生成一个连续的概率密度函数，从而显示数据的分布情况。核密度图比直方图更为光滑，因为它不会依赖于具体的分箱选择。核密度估计是基于核函数（通常是高斯核函数）计算的，核函数会在每个数据点周围放置一个平滑的曲线，然后将这些曲线相加得到整体的密度估计。核密度图能够很好地显示数据的分布特征，如数据的多峰性和尾部行为。

5.3.4 PCA分析

利用可信多肽的表达量进行主成分分析（PCA分析），从不同维度展现样品间的关系。图中每个点代表一个样本，用不同颜色区分组别。如果两个样本之间差异显著，那么这两个坐标点在得分图上的位置相对较远，反之亦然。

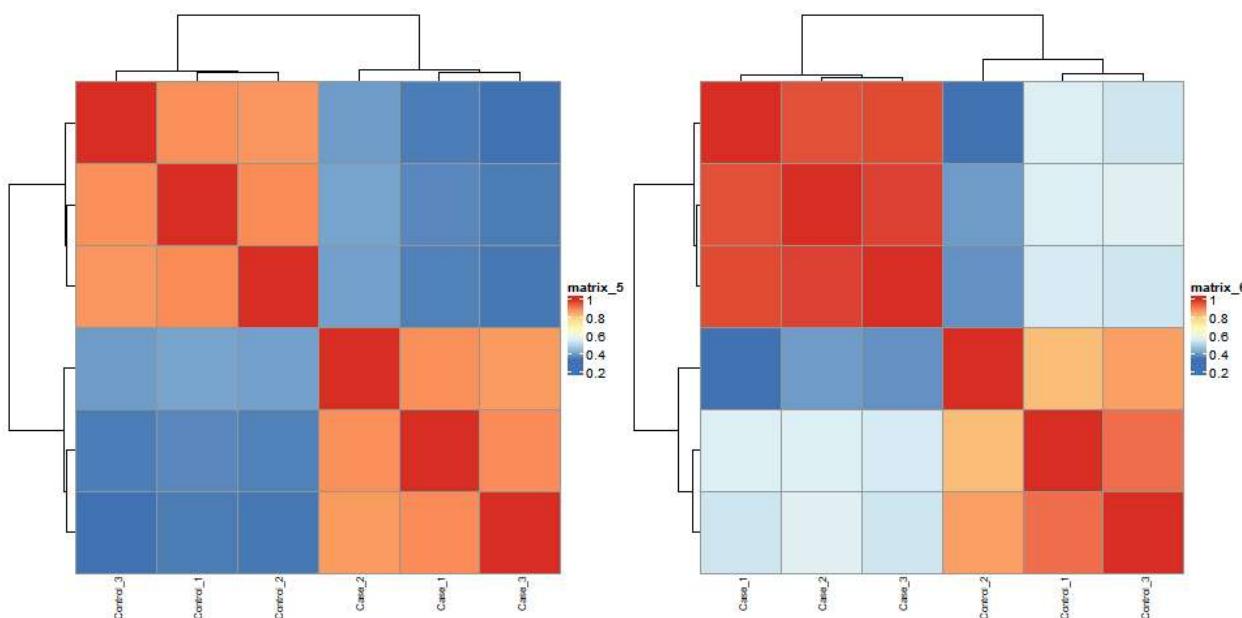


左图为mouse,右图为HELPX

图片说明：PCA得分图，横坐标pc1为第一主成分解释率，纵坐标pc2为第二主成分解释率，图中每个点代表一个样品，可以直观看出不同样品之间距离，据此可判别样品相似性或差异性。

5.3.5 样品相关性分析

对可信肽段进行样品相关性分析，通过度量样品之间的相关程度，反映样品之间的分组情况和重复性问题。皮尔森相关性热图使用颜色的深浅来表示变量之间相关性的强度，颜色越深表示相关性越强。



左图为mouse,右图为HELPX

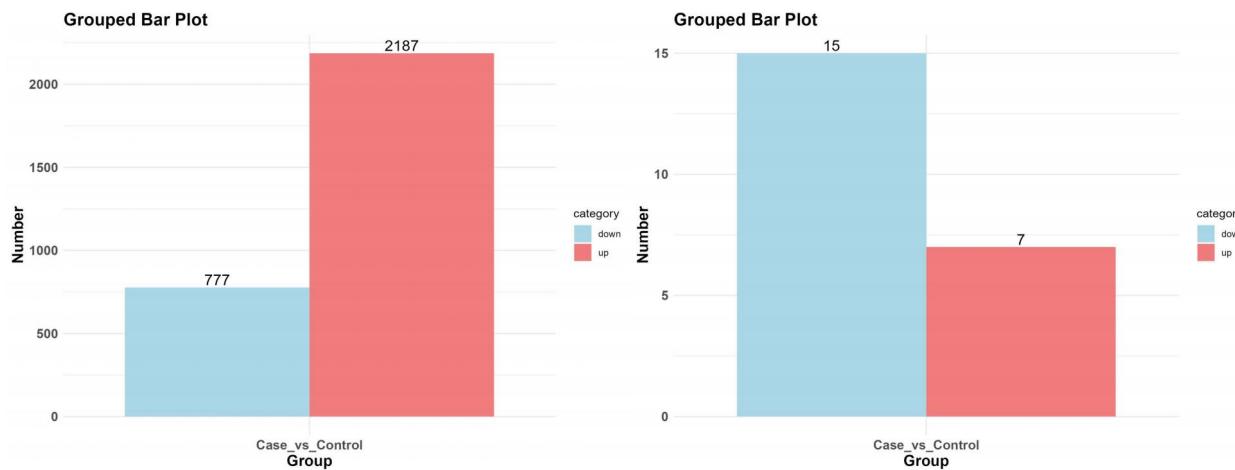
图片说明：皮尔逊相关系数的值介于-1到1之间，其中1表示完全正相关，-1表示完全负相关，而0表示没有线性相关。该热图的横纵坐标均为样本名，颜色代表相关性系数r，具体大小可参考color bar。相关性热图反映了组间样本的一致性，以及组间的差异。

5.4 个性化分析

数据经log转化后，使用两个参数共同评估组间的肽段表达差异性，即log2 Fold change (FC，倍数变化，由log2FC计算而来， $\log_2(\text{Foldchange}) = \text{实验组均值}-\text{对照组均值}$) 和经limma计算得到的p-value。本报告中，差异筛选条件为：p-adjust < 0.05 和 $|\log_2\text{foldchange}| > 0.585$

5.4.1 差异统计

各比较组显著差异上、下调肽段数目分布：



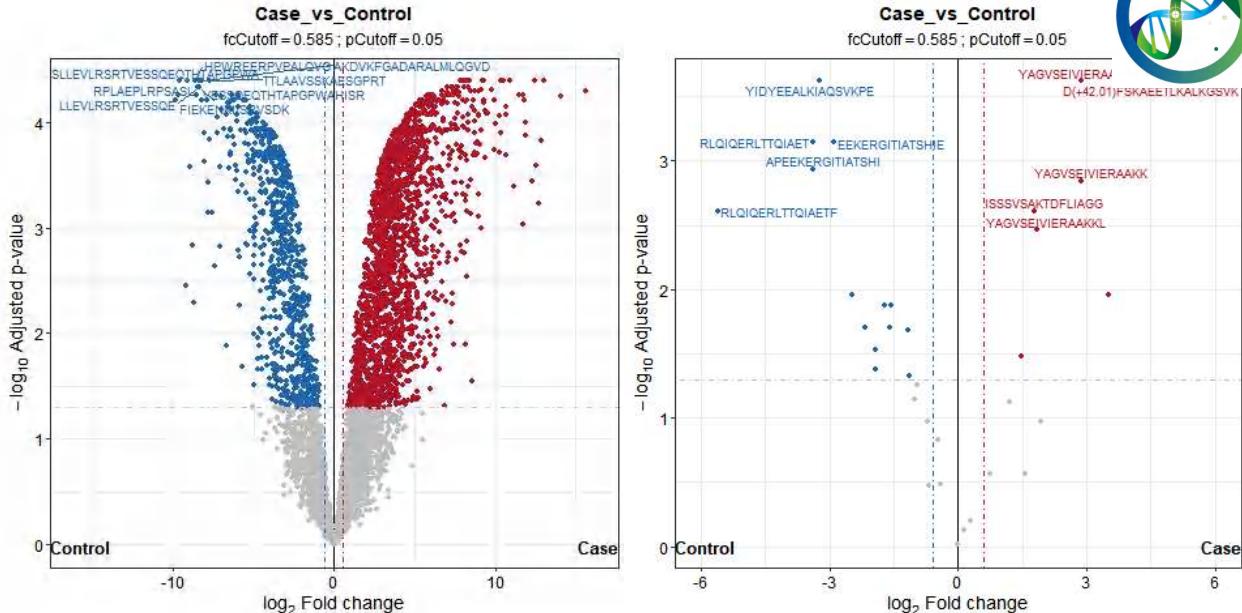
左图为mouse,右图为HELUX

图片说明：横坐标为不同的比较组，纵坐标为显著差异蛋白数量，红色代表上调的蛋白数量，浅蓝色代表下调蛋白数量。

5.4.2 差异肽段火山图

火山图横坐标为比较组比值取log2，使其比值成对称分布，即 $\log_2(\text{ratio}) > 0$ 为高表达肽段， $\log_2(\text{ratio}) < 0$ 为低表达肽段。各比较组火山图如下：

Case_vs_Control



左图为mouse,右图为HELPX

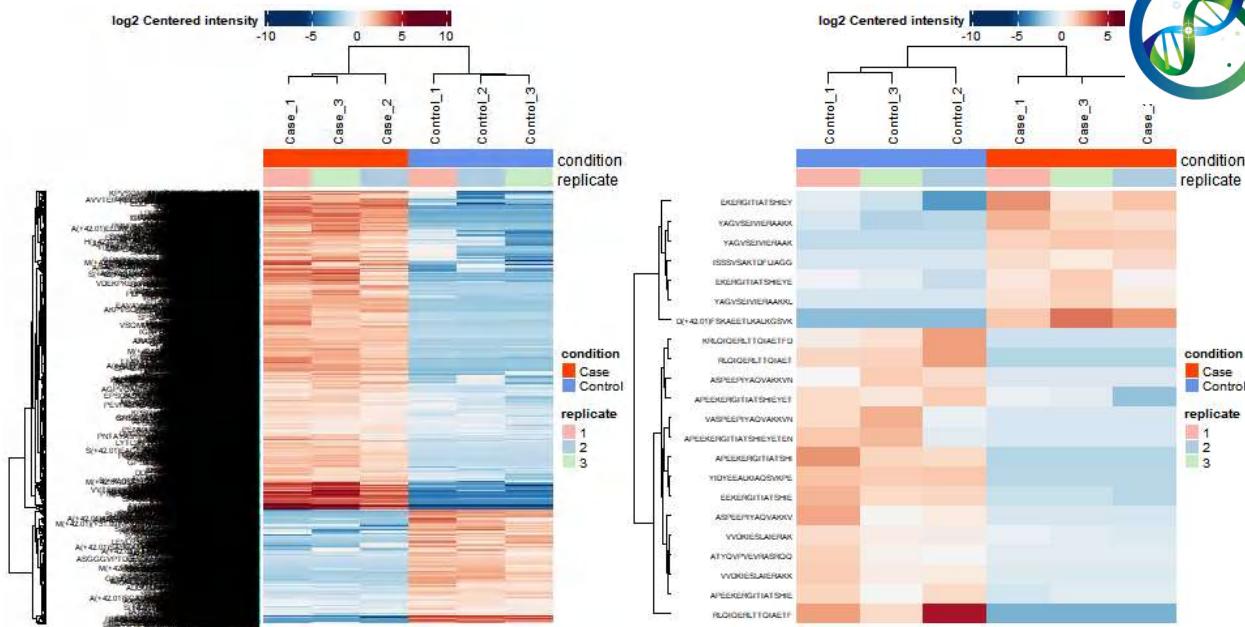
图片说明：横坐标代表log2Fold change值（表示蛋白在两个样本之间的倍数变化，是计算蛋白表达差异的常用指标），纵坐标代表其矫正后的p值的 $-\log_{10}$ 值，按照 $|\log_2 \text{Fold change}| > 0.585$ 且 $p.\text{adj} < 0.05$ 进行差异蛋白筛选，灰色的为不具有显著差异的蛋白，红色的为显著上调的蛋白，蓝色的为显著下调的蛋白。

(1)mouse 中不同比较组详细差异肽段列表：Results of different protein (./mouse/3.Different_expression/)

(2)HELPX 中不同比较组详细差异肽段列表：Results of different protein (./HELPX/3.Different_expression/)

5.4.3 差异肽段表达水平聚类分析

利用热图展示每个样品所有差异肽段的丰度分布，可以直观展示差异肽段在组间的变化，每一列代表一个样品，每一行代表一个肽段，颜色表示丰度信息（丰度值经过 \log_2 处理）。



左图为mouse,右图为HELPX

图片说明：横坐标为不同样本，纵坐标代表不同差异蛋白，样本名称下方的第一行颜色代表其分组，第二行颜色代表其同组内平行样本，蛋白所对应的颜色反映了蛋白的丰度情况，颜色越红表示蛋白的丰度越高，越蓝说明丰度越低。

5.4.4 差异肽段motif分析

Motif是一段典型的序列或者一个结构。一般来说，我们称为基序。一般情况下是指构成任何一种特征序列的基本结构。基于motif序列的提取，我们可以预测潜在的结合位点等等，有助于我们进一步理解各生物学过程中涉及的生物学意义。

(1)mouse 中 motif 分析所使用的差异肽段序列数据存放于： pep.fasta
 (../mouse/data); motif 分析详细结果请看： Results of different protein
 (../mouse/MEME%20Results.html)

(2)HELPX 中 motif 分析所使用的差异肽段序列数据存放于： pep.fasta
 (../HELPX/data); motif 分析详细结果请看： Results of different protein
 (../HELPX/MEME%20Results.html)



图片说明：图片为示例图片。motif logo由每个位置的一堆字母组成。字母的相对大小表示它们在序列中的频率。每个字母的高度与该位置的相应碱基的出现频率成正比，常以bits为单位。每个位置的字母按照保守性从大到小排列，可以方便的从顶端的字母识别保守序列。