**生信分析报告**

**项目标题： 卵巢癌耐药性 ;**

**单 号： BSXL240708 ;**

**分析人员： 黄礼闯 ;**

**分析类型： 生信分析 ;**

**委 托 人： 宛迎春 ;**

**受 托 人： 杭州铂赛生物科技有限公司 .**

# 1 分析流程

* ~~不同RUNX2水平癌症患者总生存期的荟萃分析（根据文献情况及分析结果选择是否保留）(预期：RUNX2高表达预后不好)。~~
  + 未完成
* 基于ENCORI平台的卵巢癌和癌旁组织中RUNX2水平的比较
* ~~显示TMA（组织微阵列）中RUNX2蛋白的IHC评分的箱形图。~~
  + 无合适数据集
* 不同RUNX2水平的卵巢癌患者的总生存期。
* GEO数据集验证RUNX2 mRNA在耐药细胞系中的表达
* 通过STRING、GeneMANIA等网站或者数据库，寻找与RUNX2互作的ZDHHCs蛋白； 根据实际预测结果确定，暂定为ZDHHC5
* 使用SwissPalm和TermiNator在线工具预测RUNX2可能受S-棕榈酰化调节。

# 2 材料和方法

## 2.1 数据分析平台

在 Linux pop-os x86\_64 (6.9.3-76060903-generic) 上，使用 R version 4.4.2 (2024-10-31) (<https://www.r-project.org/>) 对数据统计分析与整合分析。

## 2.2 GEO 数据获取 (Dataset: GSE289067)

以 R 包 GEOquery (2.74.0) 获取 GSE289067 数据集。

## 2.3 Limma 差异分析 (Dataset: GSE289067)

以 R 包 edgeR (4.4.0) ()1 对数据预处理。以 edgeR::filterByExpr 过滤 count 数量小于 10 的基因。以 edgeR::calcNormFactors，limma::voom 转化 count 数据为 log2 counts-per-million (logCPM)。 以 limma (3.62.1) (2005)2 差异分析。分析方法参考 <https://bioconductor.org/packages/release/workflows/vignettes/RNAseq123/inst/doc/limmaWorkflow.html>。创建设计矩阵，对比矩阵，差异分析：Tumor\_tissue vs Control\_tissue。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 P.Value 小于 0.05，|Log2(FC)| 大于 0.5 的统计结果。

## 2.4 TCGA 数据获取 (Dataset: OV)

以 R 包 TCGAbiolinks (2.35.1) (2015, **IF:16.6**, Q1, Nucleic Acids Research)3 获取 TCGA-OV 数据集。

## 2.5 Survival 生存分析 (Dataset: TCGA\_OV)

以 R 包 edgeR (4.4.0) ()1 对数据预处理。以 edgeR::filterByExpr 过滤 count 数量小于 10 的基因。以 edgeR::calcNormFactors，limma::voom 转化 count 数据为 log2 counts-per-million (logCPM)。使用标准化过的基因表达数据。 以 R 包 survival (3.7.0) 生存分析，以 R 包 survminer (0.5.0) 绘制生存曲线。

## 2.6 GEO 数据获取 (Dataset: GSE173579)

以 R 包 GEOquery (2.74.0) 获取 GSE173579 数据集。 以 GEOquery::getRNASeqData 获取 RNA count 数据以及基因注释。

## 2.7 Limma 差异分析 (Dataset: GSE173579)

以 R 包 edgeR (4.4.0) ()1 对数据预处理。以 edgeR::filterByExpr 过滤 count 数量小于 10 的基因。以 edgeR::calcNormFactors，limma::voom 转化 count 数据为 log2 counts-per-million (logCPM)。 以 limma (3.62.1) (2005)2 差异分析。分析方法参考 <https://bioconductor.org/packages/release/workflows/vignettes/RNAseq123/inst/doc/limmaWorkflow.html>。创建设计矩阵，对比矩阵，差异分析：Carboplatin\_high\_Resistant\_SKOV3 vs Non\_resistance\_SKOV3。使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 拟合线形模型。以 limma::topTable 提取所有结果，并过滤得到 P.Value 小于 0.05，|Log2(FC)| 大于 0.5 的统计结果。

## 2.8 STRINGdb PPI 分析 (Dataset: ZDHHC)

以 R 包 STEINGdb (2.18.0) (2021, **IF:16.6**, Q1, Nucleic Acids Research)4 构建 PPI 网络。数据版本为 12.0，互作类型为 full。以 Cytohubba (2014, BMC Systems Biology)5 的算法计算 MCC score (在 R 中计算) 。随后，以 ggraph 可视化网络 (2.2.1)。

## 2.9 ClusPro 蛋白质-蛋白质对接预测 (Dataset: ZDHHC)

以 R 包 biomaRt (2.62.0) (2009, **IF:13.1**, Q1, Nature protocols)6 获取基因 Symbol 对应的蛋白结构 PDB (<https://www.rcsb.org/>) 数据库 ID。以 R 包 bio3d (2.4.5) 获取 PDB ID 对应的注释 (蛋白结构分辨率, resolution) 。首要以 resolution 选取用于分子对接的蛋白结构 (resolution 越小，分辨率越高) 。以 RCSB API (<https://www.rcsb.org/docs/programmatic-access/web-apis-overview>) 获取蛋白 PDB 文件。以 R 包 UniProt.ws (2.46.1) 获取基因 (symbol) 的 UniProtKB-Swiss-Prot ID (Entry ID)，随后，以 Entry ID 从数据库 AlphaFold (<https://alphafold.ebi.ac.uk/>) 获取数据库 PDB 中不包含的蛋白结构 (预测的结构)。 登录 ClusPro ((2017, **IF:13.1**, Q1, Nature protocols)7) (<https://cluspro.bu.edu/home.php>)，上传需要对接的蛋白结构 (PDB) 。

## 2.10 MusiteDeep 蛋白质转录后修饰位点预测 (Dataset: RUNX2)

以 biomaRt (2009, **IF:13.1**, Q1, Nature protocols)6 获取蛋白质 (hsa) 的序列 (biomaRt::getSequence 获取 peptide)。 以 Python 工具 MusiteDeep (2020, **IF:16.6**, Q1, Nucleic Acids Research)8 预测 S-palmitoyl\_cysteine 修饰位点，设定 PTM 得分截断为 0.5。

# 3 分析结果

## 3.1 GEO 数据获取 (GSE289067)

以 GEOquery 获取 GSE289067 的数据信息。

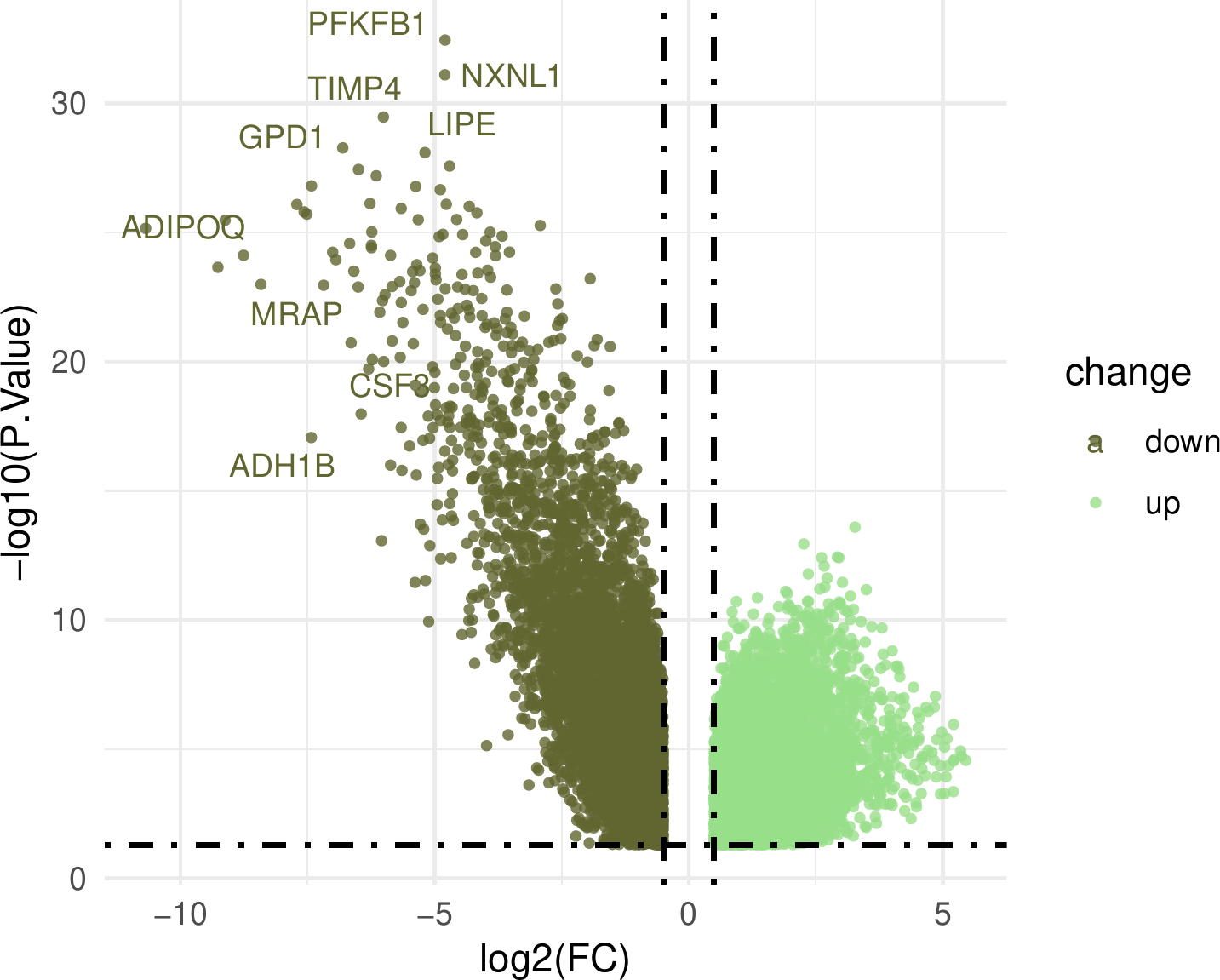
* Data Source ID: GSE289067
* data\_processing: The pipeline encompassed several key steps, beginning with initial quality assessment and adapter trimming of raw reads. Processed reads were then mapped to the GRCh38 primary assembly (GRCh38.primary\_assembly.genome.fa, obtained from the Gencode website) using STAR aligner102. The alignment was guided by the Gencode v41 primary annotation (gencode.v41.primary\_assembly.annotation.gtf).
* data\_processing.1: Following alignment, Salmon103 was employed to quantify transcript abundances, with the –rf-stranded (reverse/forward) parameter specified to account for the stranded nature of the TruSeq library preparation, ensuring correct strand-specific quantification of transcripts. Throughout the pipeline, several quality control measures were implemented.
* data\_processing.2: These included FastQC for assessment of sequence quality metrics, featureCounts for quantification of reads mapped to different genomic features and biotypes (e.g., protein-coding genes, lncRNAs, miRNAs), and DESeq2 for Principal Component Analysis (PCA) and sample similarity analysis to identify potential batch effects or outliers. Additionally, Picard MarkDuplicates was used to identify PCR and optical duplicates (duplicates were marked but not removed), Samtools generated alignment statist…
* data\_processing.3: Assembly: GRCh38
* (Others): …

**(见Figure+Table/3.1\_GEO\_数据获取\_(GSE289067)/GSE289067-GSE289067-content)**

## 3.2 Limma 差异分析 (GSE289067)

匹配 group 中包含“Tumor\_tissue|Control\_tissue”的描述，最终得到 139 例数据。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Tumor\_tissue vs Control\_tissue。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

Fig. **[1](#GSE289067-Tumor-tissue-vs-Control-tissue)** 为 Tumor\_tissue - Control\_tissue 差异分析火山图。 Tab. **[1](#GSE289067-data-Tumor-tissue-vs-Control-tissue)** 为 Tumor\_tissue - Control\_tissue 差异分析统计表格。 Fig. **[2](#GSE289067-Box-Plot-Of-DEGs)** 基因 RUNX2 表达水平，以及对应的 limma 差异分析显著水平。 Tab. **[2](#DEGs-of-ZDHHCs)** 差异表达的 ZDHHCs。 Fig. **[3](#GSE289067-Box-Plot-Of-DEGs-ZDHHC)** 基因 ZDHHC18, ZDHHC23, ZDHHC16, ZDHHC13, ZDHHC15, ZDHHC8P1, ZDHHC12, ZDHHC11B, ZDHHC14 表达水平，以及对应的 limma 差异分析显著水平。



**Fig.** **1** GSE289067 Tumor tissue vs Control tissue

**(File path: Figure+Table/3.2\_Limma\_差异分析\_(GSE289067)/GSE289067-Tumor-tissue-vs-Control-tissue.pdf)**

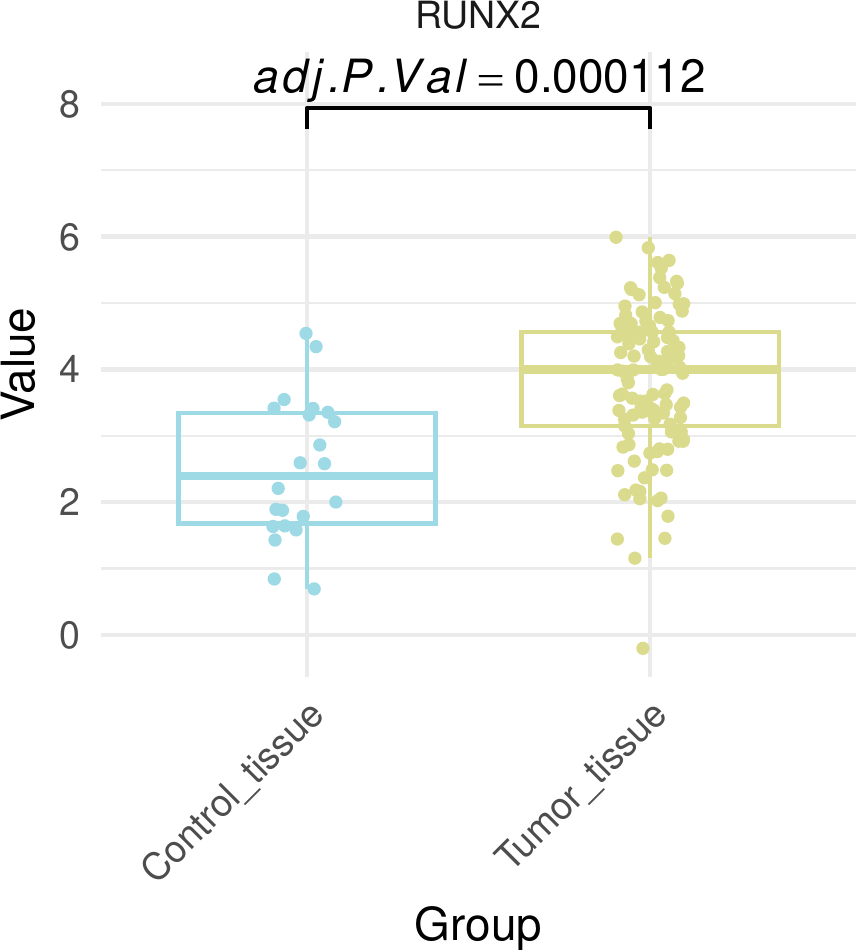
* P.Value cut-off: 0.05
* Log2(FC) cut-off: 0.5

**(See: Figure+Table/3.2\_Limma\_差异分析\_(GSE289067)/GSE289067-Tumor-tissue-vs-Control-tissue-content)**

**Tab.** **1** GSE289067 data Tumor tissue vs Control tissue

| Hgnc symbol | LogFC | P.Value | Rownames | AveExpr |
| --- | --- | --- | --- | --- |
| PFKFB1 | -4.796 | 3.527e-33 | PFKFB1 | -0.4323 |
| NXNL1 | -4.8 | 7.781e-32 | NXNL1 | -5.996 |
| TIMP4 | -6.01 | 3.36e-30 | TIMP4 | 1.167 |
| GPD1 | -6.809 | 5.246e-29 | GPD1 | 2.559 |
| LIPE | -5.191 | 8.015e-29 | LIPE | 3.3 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.2\_Limma\_差异分析\_(GSE289067)/GSE289067-data-Tumor-tissue-vs-Control-tissue.csv)**



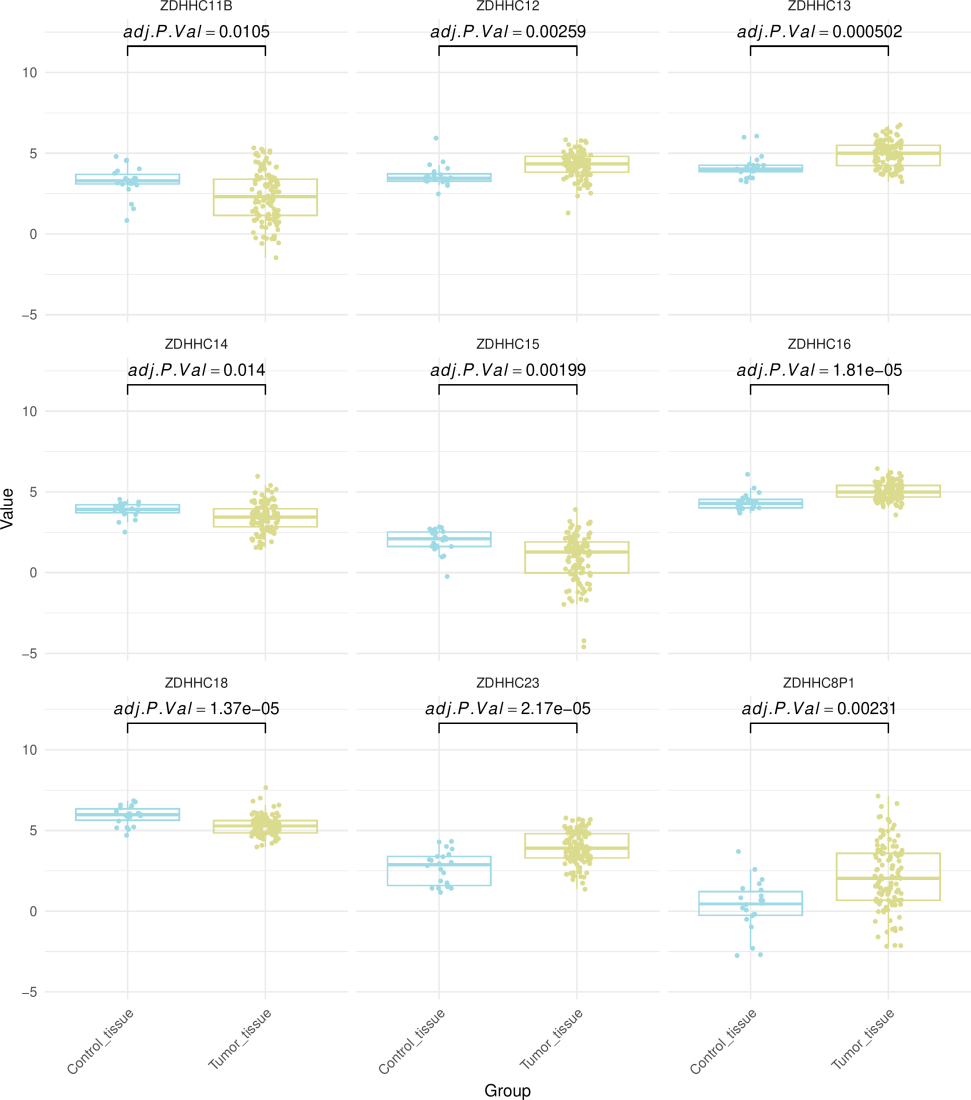
**Fig.** **2** GSE289067 Box Plot Of DEGs

**(File path: Figure+Table/3.2\_Limma\_差异分析\_(GSE289067)/GSE289067-Box-Plot-Of-DEGs.pdf)**

**Tab.** **2** DEGs of ZDHHCs

| Hgnc symbol | LogFC | P.Value | Rownames | AveExpr |
| --- | --- | --- | --- | --- |
| ZDHHC18 | -0.6534 | 1.953e-06 | ZDHHC18 | 5.377 |
| ZDHHC23 | 1.311 | 3.317e-06 | ZDHHC23 | 3.757 |
| ZDHHC16 | 0.6486 | 2.684e-06 | ZDHHC16 | 4.918 |
| ZDHHC13 | 0.7838 | 0.000125 | ZDHHC13 | 4.824 |
| ZDHHC15 | -1.031 | 0.0006157 | ZDHHC15 | 1.08 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.2\_Limma\_差异分析\_(GSE289067)/DEGs-of-ZDHHCs.csv)**



**Fig.** **3** GSE289067 Box Plot Of DEGs ZDHHC

**(File path: Figure+Table/3.2\_Limma\_差异分析\_(GSE289067)/GSE289067-Box-Plot-Of-DEGs-ZDHHC.pdf)**

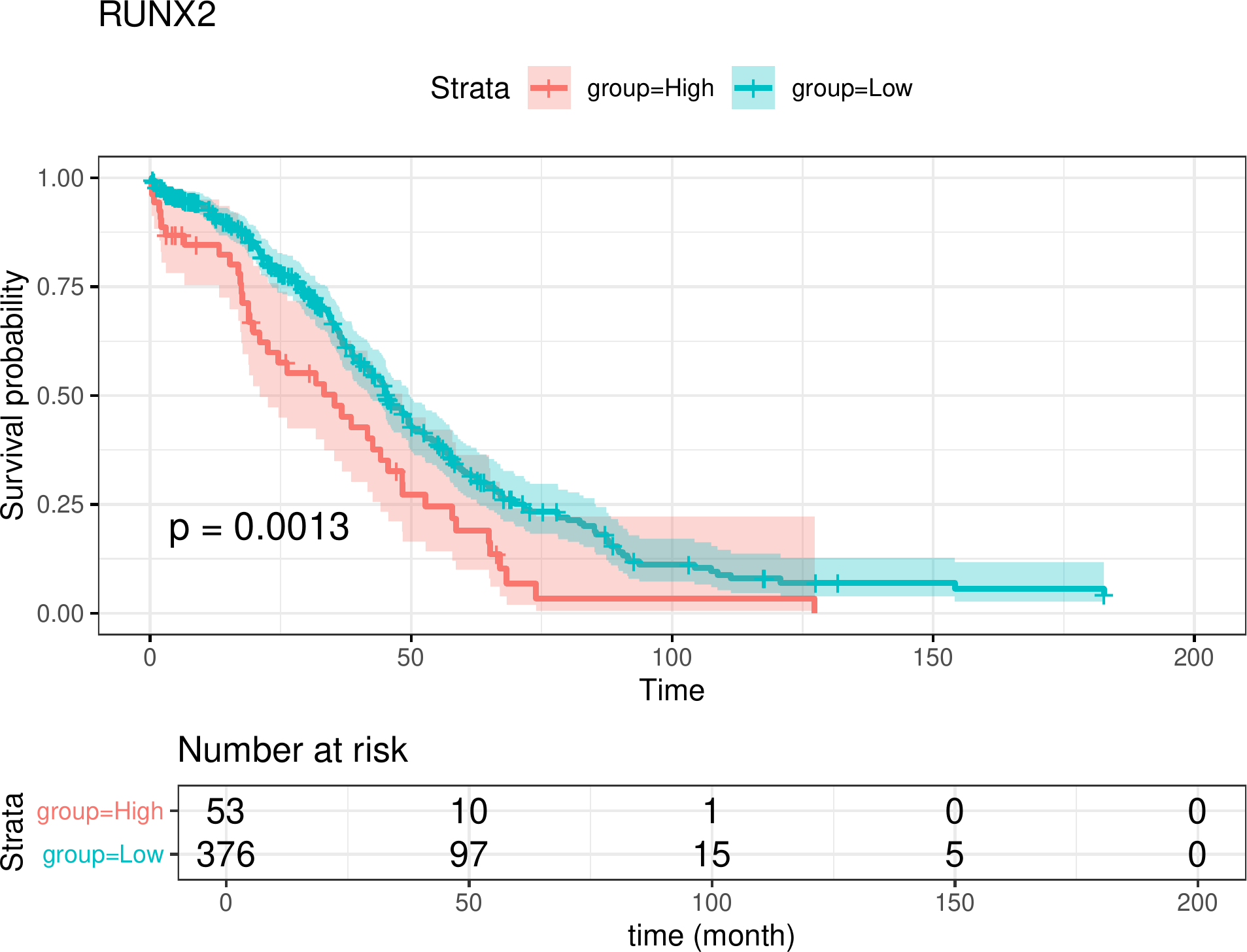
## 3.3 TCGA 数据获取 (OV)

获取 TCGA-OV 数据。

## 3.4 Survival 生存分析 (TCGA\_OV)

按 survminer::surv\_cutpoint 计算的 cutoff，将样本分为 Low 和 High 风险组。生存数据为TCGA-OV，使用标准化过的基因表达数据。根据元数据信息 (即临床数据) ，去除了生存状态未知的样例。

Fig. **[4](#TCGA-OV-Survival-plots)** 为 RUNX2 生存曲线。



**Fig.** **4** TCGA OV Survival plots

**(File path: Figure+Table/3.4\_Survival\_生存分析\_(TCGA\_OV)/TCGA-OV-Survival-plots.pdf)**

## 3.5 GEO 数据获取 (耐药细胞系) (GSE173579)

以 GEOquery 获取 GSE173579 的数据信息。

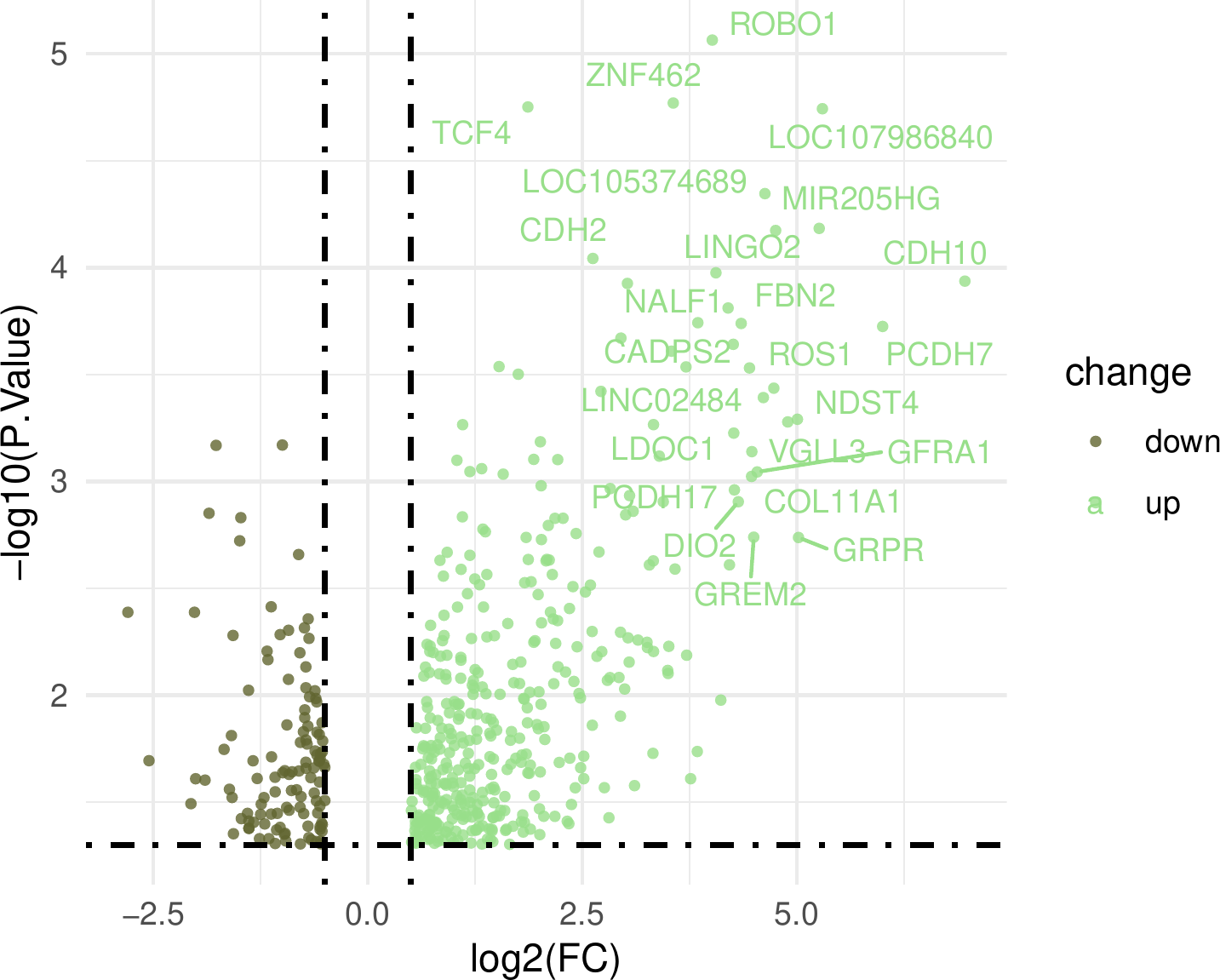
* Data Source ID: GSE173579
* data\_processing: Illumina Casava1.8 software used for basecalling.
* data\_processing.1: Sequenced reads were trimmed for adaptor sequence, and masked for low-complexity or low-quality sequence(using fastx\_trimmer), then mapped to hg19 whole genome using Bowtie2
* data\_processing.2: Read count extraction and nomalization were performed using edgeR.
* data\_processing.3: Genome\_build: hg19
* (Others): …

**(见Figure+Table/3.5\_GEO\_数据获取\_(耐药细胞系)\_(GSE173579)/GSE173579-GSE173579-content)**

## 3.6 Limma 差异分析 (耐药细胞系) (GSE173579)

匹配 group 中包含“high\_Resistant|Non\_resistance”的描述，最终得到 6 例数据。以 edgeR 将GSE173579 RNA-seq 数据标准化 (详见方法章节)。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Carboplatin\_high\_Resistant\_SKOV3 vs Non\_resistance\_SKOV3。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

Fig. **[5](#GSE173579-Carboplatin-high-Resistant-SKOV3-vs-Non-resistance-SKOV3)** 为 Carboplatin\_high\_Resistant\_SKOV3 - Non\_resistance\_SKOV3 差异分析火山图。 Tab. **[3](#GSE173579-data-Carboplatin-high-Resistant-SKOV3-vs-Non-resistance-SKOV3)** 为 Carboplatin\_high\_Resistant\_SKOV3 - Non\_resistance\_SKOV3 差异分析统计表格。 Fig. **[6](#GSE173579-Box-Plot-Of-DEGs)** 基因 RUNX2 表达水平，以及对应的 limma 差异分析显著水平。



**Fig.** **5** GSE173579 Carboplatin high Resistant SKOV3 vs Non resistance SKOV3

**(File path: Figure+Table/3.6\_Limma\_差异分析\_(耐药细胞系)\_(GSE173579)/GSE173579-Carboplatin-high-Resistant-SKOV3-vs-Non-resistance-SKOV3.pdf)**

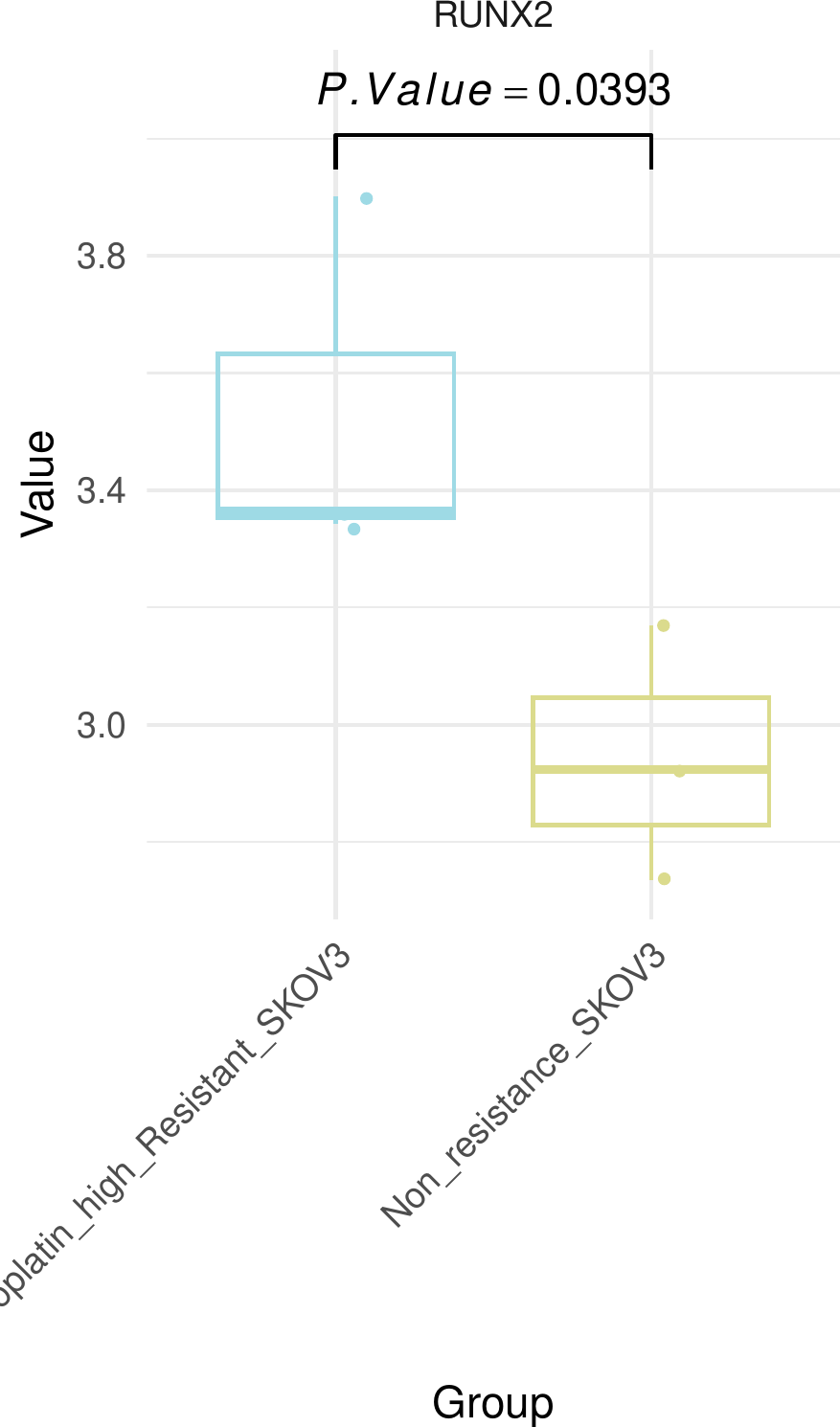
* P.Value cut-off: 0.05
* Log2(FC) cut-off: 0.5

**(See: Figure+Table/3.6\_Limma\_差异分析\_(耐药细胞系)\_(GSE173579)/GSE173579-Carboplatin-high-Resistant-SKOV3-vs-Non-resistance-SKOV3-content)**

**Tab.** **3** GSE173579 data Carboplatin high Resistant SKOV3 vs Non resistance SKOV3

| Hgnc symbol | LogFC | P.Value | Rownames | GeneID |
| --- | --- | --- | --- | --- |
| TCF4 | 1.866 | 1.772e-05 | 6925 | 6925 |
| ROBO1 | 4.015 | 8.622e-06 | 6091 | 6091 |
| ZNF462 | 3.56 | 1.698e-05 | 58499 | 58500 |
| CDH2 | 2.623 | 9.056e-05 | 1000 | 1000 |
| CNTNAP2 | 3.026 | 0.0001185 | 26047 | 26050 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.6\_Limma\_差异分析\_(耐药细胞系)\_(GSE173579)/GSE173579-data-Carboplatin-high-Resistant-SKOV3-vs-Non-resistance-SKOV3.tsv)**



**Fig.** **6** GSE173579 Box Plot Of DEGs

**(File path: Figure+Table/3.6\_Limma\_差异分析\_(耐药细胞系)\_(GSE173579)/GSE173579-Box-Plot-Of-DEGs.pdf)**

## 3.7 STRINGdb PPI 分析 (ZDHHC)

Fig. **[7](#ZDHHC-Top-MCC-score)** PPI (带有 Cytohubba (2014, BMC Systems Biology)5 MCC 得分) 网络图



**Fig.** **7** ZDHHC Top MCC score

**(File path: Figure+Table/3.7\_STRINGdb\_PPI\_分析\_(ZDHHC)/ZDHHC-Top-MCC-score.pdf)**

## 3.8 ClusPro 蛋白质-蛋白质对接预测 (ZDHHC)

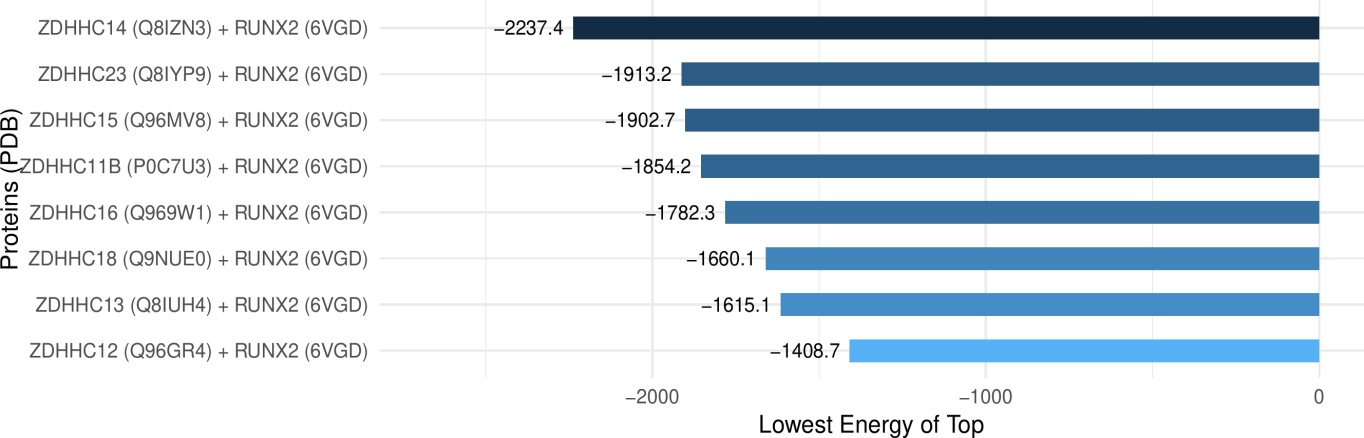
以 biomaRt 获取基因 Symbol 对应的蛋白结构 (PDB，详见方法章节)。选取分辨率最高 (即，resolution 值最小) 的 PDB 作为分子对接的蛋白结构。从 RCSB PDB 获取 PDB 文件。对于未从 PDB 数据库找到结构文件的，从数据库 AlphaFold 获取 ZDHHC18, ZDHHC23, ZDHHC16, …(n = 9) 预测的蛋白结构 (根据 UniProtKB-Swiss-Prot ID，详见方法章节)。将 PDB 上传至 ClusPro 进行对接。 Tab. **[4](#ZDHHC-docking-layouts)** 为蛋白质对接所使用的 PDB 来源附表 (如果是 PDB ID，则 pdb 文件直接来源于 PDB 数据库；如果是 UniProtKB-Swiss-Prot ID，则根据该 ID，从 AlphaFold 获取预测的结构文件。

Fig. **[8](#ZDHHC-Overview-of-protein-docking-results-)** 为每组对接结果的最小能量柱状图 (请参考 <https://cluspro.bu.edu/help.php>)。 Fig. **[9](#Protein-docking-of-ZDHHC14-RUNX2)** 以 pymol 将蛋白质 (Top\_1\_ZDHHC14\_RUNX2) 对接结果可视化。 Fig. **[10](#Protein-docking-of-ZDHHC23-RUNX2)** 以 pymol 将蛋白质 (Top\_2\_ZDHHC23\_RUNX2) 对接结果可视化。 Fig. **[11](#Protein-docking-of-ZDHHC15-RUNX2)** 以 pymol 将蛋白质 (Top\_3\_ZDHHC15\_RUNX2) 对接结果可视化。

**Tab.** **4** ZDHHC docking layouts

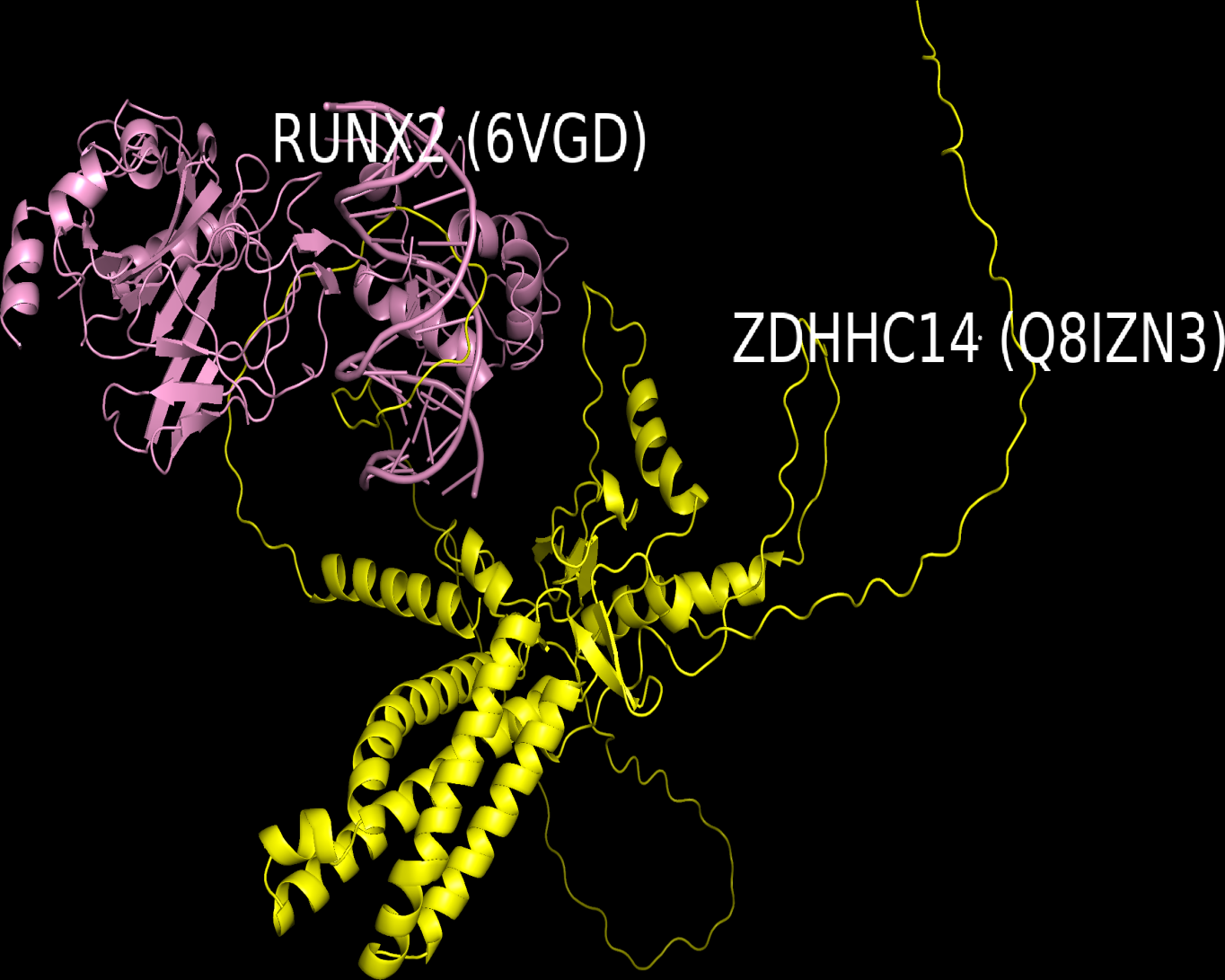
| Name | From id | To id |
| --- | --- | --- |
| ZDHHC18 RUNX2 | Q9NUE0 | 6VGD |
| ZDHHC23 RUNX2 | Q8IYP9 | 6VGD |
| ZDHHC16 RUNX2 | Q969W1 | 6VGD |
| ZDHHC13 RUNX2 | Q8IUH4 | 6VGD |
| ZDHHC15 RUNX2 | Q96MV8 | 6VGD |
| ... | ... | ... |

**(File path: Figure+Table/3.8\_ClusPro\_蛋白质-蛋白质对接预测\_(ZDHHC)/ZDHHC-docking-layouts.csv)**



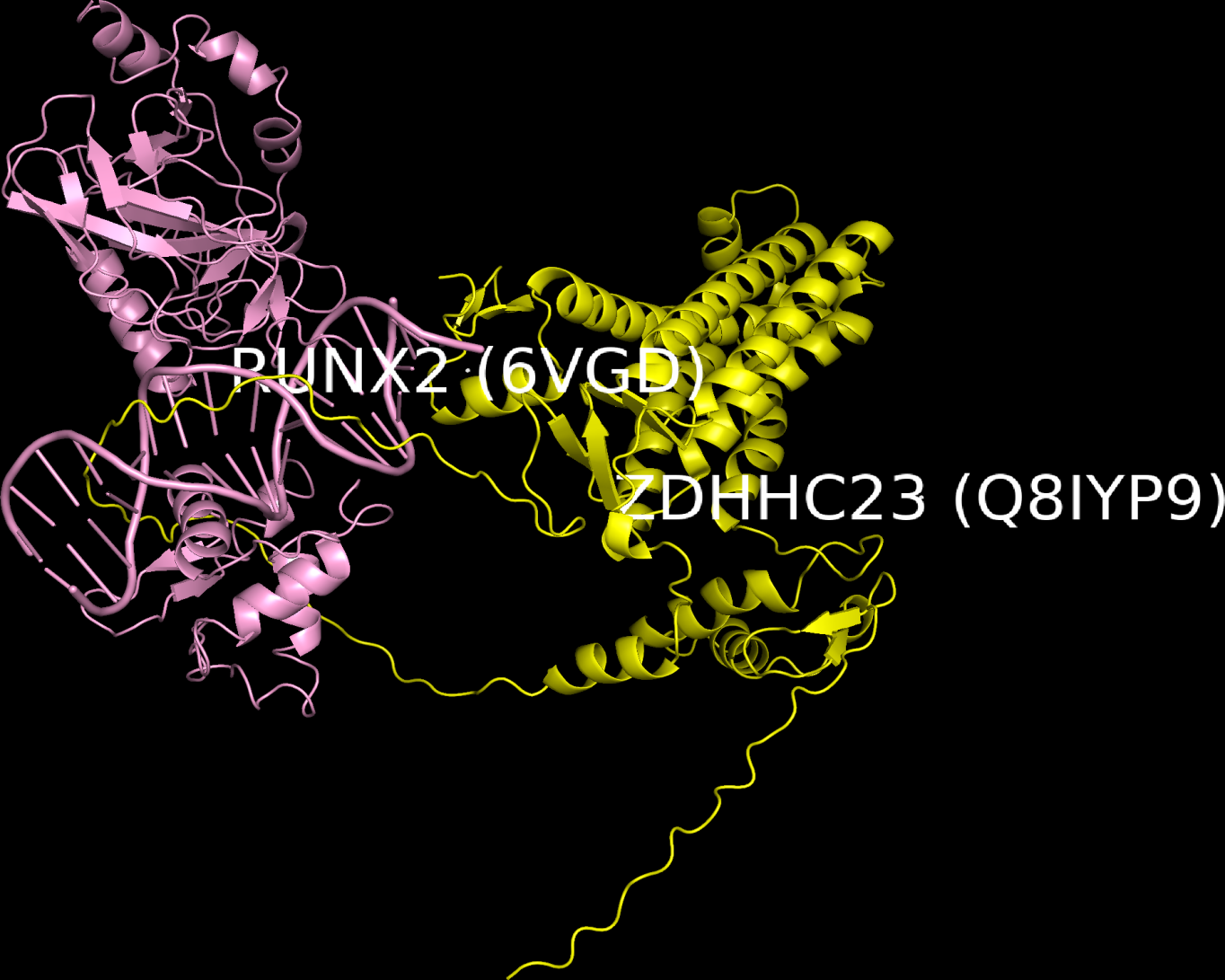
**Fig.** **8** ZDHHC Overview of protein docking results

**(File path: Figure+Table/3.8\_ClusPro\_蛋白质-蛋白质对接预测\_(ZDHHC)/ZDHHC-Overview-of-protein-docking-results-.pdf)**



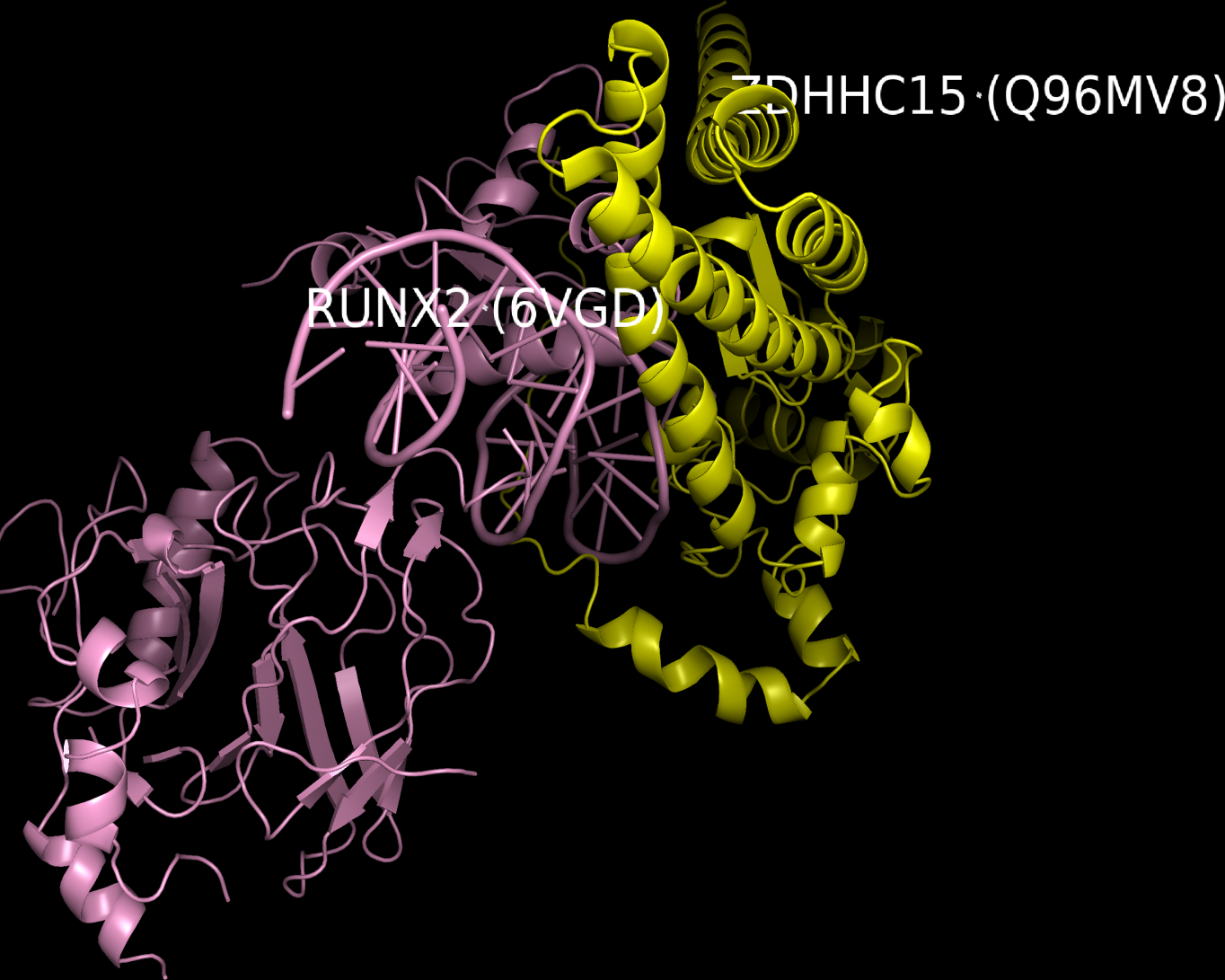
**Fig.** **9** Protein docking of ZDHHC14 RUNX2

**(File path: Figure+Table/3.8\_ClusPro\_蛋白质-蛋白质对接预测\_(ZDHHC)/Protein-docking-of-ZDHHC14-RUNX2.png)**



**Fig.** **10** Protein docking of ZDHHC23 RUNX2

**(File path: Figure+Table/3.8\_ClusPro\_蛋白质-蛋白质对接预测\_(ZDHHC)/Protein-docking-of-ZDHHC23-RUNX2.png)**



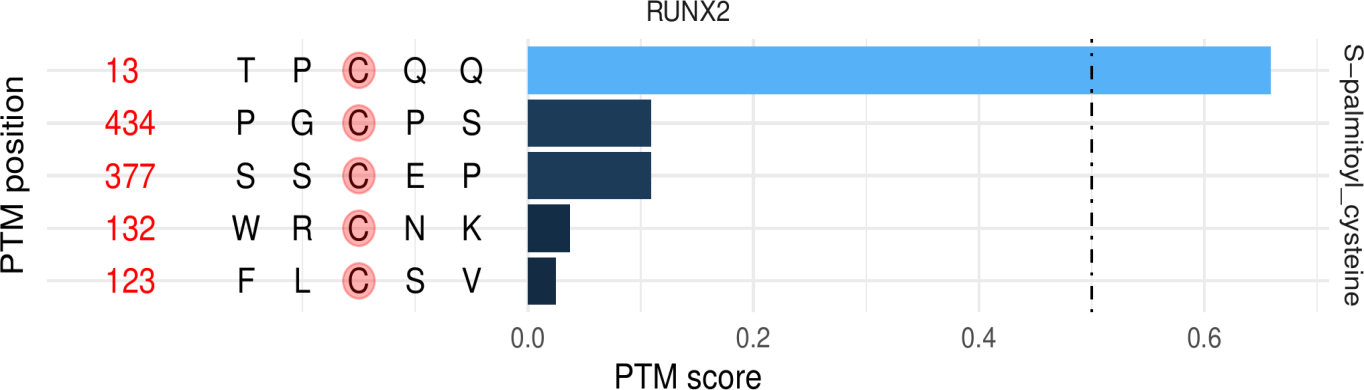
**Fig.** **11** Protein docking of ZDHHC15 RUNX2

**(File path: Figure+Table/3.8\_ClusPro\_蛋白质-蛋白质对接预测\_(ZDHHC)/Protein-docking-of-ZDHHC15-RUNX2.png)**

## 3.9 MusiteDeep 蛋白质转录后修饰位点预测 (RUNX2)

以 biomaRt 获取蛋白质 (RUNX2) 的序列 (Peptide)。以 MusiteDeep 预测 S-palmitoyl\_cysteine 修饰位点。

Fig. **[12](#RUNX2-RUNX2-PTM-score)** 为 RUNX2 的修饰位点以及得分可视化图。 Tab. **[5](#RUNX2-prediction-PTM-of-S-palmitoyl-cysteine)** 以 MusiteDeep 预测的修饰位点以及得分附表。



**Fig.** **12** RUNX2 RUNX2 PTM score

**(File path: Figure+Table/3.9\_MusiteDeep\_蛋白质转录后修饰位点预测\_(RUNX2)/RUNX2-RUNX2-PTM-score.pdf)**

**Tab.** **5** RUNX2 prediction PTM of S palmitoyl cysteine

| Sequence name | PTM type | Position | Residue | PTM score |
| --- | --- | --- | --- | --- |
| RUNX2 | S-palmitoyl cysteine | 13 | C | 0.659 |
| RUNX2 | S-palmitoyl cysteine | 377 | C | 0.109 |
| RUNX2 | S-palmitoyl cysteine | 434 | C | 0.109 |
| RUNX2 | S-palmitoyl cysteine | 132 | C | 0.037 |
| RUNX2 | S-palmitoyl cysteine | 123 | C | 0.025 |

**(File path: Figure+Table/3.9\_MusiteDeep\_蛋白质转录后修饰位点预测\_(RUNX2)/RUNX2-prediction-PTM-of-S-palmitoyl-cysteine.csv)**

# 4 总结

* 卵巢癌和癌旁组织中RUNX2水平的比较，见Fig. **[2](#GSE289067-Box-Plot-Of-DEGs)**
* 不同RUNX2水平的卵巢癌患者的总生存期，见 Fig. **[4](#TCGA-OV-Survival-plots)**
* GEO数据集验证RUNX2 mRNA在耐药细胞系中的表达，见 Fig. **[6](#GSE173579-Box-Plot-Of-DEGs)** ，
* 通过STRING、GeneMANIA等网站或者数据库，寻找与RUNX2互作的ZDHHCs蛋白，见 Fig. **[7](#ZDHHC-Top-MCC-score)** (未发现互作数据)。随后，尝试以 ClusPro 对 RUNX2 与 ZDHHCs 对接， 见 Fig. **[8](#ZDHHC-Overview-of-protein-docking-results-)**
* 预测RUNX2可能受S-棕榈酰化调节，Fig. **[12](#RUNX2-RUNX2-PTM-score)**

未完成：

* 不同RUNX2水平癌症患者总生存期的荟萃分析（根据文献情况及分析结果选择是否保留）(预期：RUNX2高表达预后不好)。
* 显示TMA（组织微阵列）中RUNX2蛋白的IHC评分的箱形图。(无合适数据集)

# Reference

1. Chen, Y., McCarthy, D., Ritchie, M., Robinson, M. & Smyth, G. EdgeR: Differential analysis of sequence read count data users guide. 119.

2. Smyth, G. K. Limma: Linear models for microarray data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds. Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. & Dudoit, S.) 397–420 (Springer-Verlag, 2005). doi:[10.1007/0-387-29362-0\_23](https://doi.org/10.1007/0-387-29362-0_23).

3. Colaprico, A. *et al.* TCGAbiolinks: An r/bioconductor package for integrative analysis of tcga data. *Nucleic Acids Research* **44**, (2015).

4. Szklarczyk, D. *et al.* The string database in 2021: Customizable proteinprotein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* **49**, D605–D612 (2021).

5. Chin, C.-H. *et al.* CytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8**, S11 (2014).

6. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomaRt. *Nature protocols* **4**, 1184–1191 (2009).

7. Kozakov, D. *et al.* The cluspro web server for protein-protein docking. *Nature protocols* **12**, 255–278 (2017).

8. Wang, D. *et al.* MusiteDeep: A deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Research* **48**, W140–W146 (2020).