**生信分析报告**

**项目标题： 骨肉瘤ZDHHC ;**

**单 号： BSHQ241042 ;**

**分析人员： 黄礼闯 ;**

**分析类型： 生信分析 ;**

**委 托 人： 梁海东 ;**

**受 托 人： 杭州铂赛生物科技有限公司 .**

# 1 分析流程

* 生物信息学筛选OS中差异表达 (正常与OS组织) 并且与预后相关的ZDHHC基因
* 鉴定ZDHHC-X和脂质代谢合成相关枢纽基因
  + 将OS转移以及未转移样本中ZDHHC-X表达从低到高排序
  + 进行WGCNA分析，选择与脂质代谢合成相关性最高的模块
  + 将ZDHCC-X与筛选出的基因上传至GENEMANIA、STRING构建PPI网络，筛选出与ZDHHCC-X相互作用的脂质代谢合成相关蛋白
  + 在OS转移与未转移样本中分析相关蛋白的表达
  + 相关性分析：ZDHHC-X与相互作用蛋白在OS转移中的相关性

# 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 TCGA 数据获取 (Dataset: OS)

以 R 包 TCGAbiolinks (2.35.1) (2015, **IF:16.6**, Q1, Nucleic Acids Research)1 获取 TARGET-OS 数据集。

## 2.3 Survival 生存分析 (Dataset: OS)

以 R 包 edgeR (4.4.0) ()2 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。

## 2.4 GeneCards 基因获取 (Dataset: LIPID)

从 GeneCards 数据库 (2016, Current protocols in bioinformatics)3 获取 lipid metabolism 相关的基因集，得分 cut-off 为 1。

## 2.5 GEO 数据获取 (Dataset: GSE87624)

以 R 包 GEOquery (2.74.0) 获取 GSE87624 数据集。 以 GEOquery:::getRNASeqQuantResults 获取 RNA count 数据 (NCBI-generated data, 参考 <https://www.ncbi.nlm.nih.gov/geo/info/rnaseqcounts.html>) 以及基因注释。

## 2.6 Limma 差异分析 (Dataset: GSE87624\_ZDHHC)

以 R 包 edgeR (4.4.0) ()2 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)4 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：metastasis vs primary, primary vs normal\_bone。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。

## 2.7 STRINGdb PPI 分析 (Dataset: LIPID\_ZDHHC)

以 R 包 STEINGdb (2.18.0) (2021, **IF:16.6**, Q1, Nucleic Acids Research)5 构建 PPI 网络。数据版本为 12.0，互作类型为 full。以 Cytohubba (2014, BMC Systems Biology)6 的算法计算 MCC score (在 R 中计算) 。随后，以 ggraph 可视化网络 (2.2.1)。

## 2.8 ClusPro 蛋白质-蛋白质对接预测 (Dataset: ZDHHC\_LIPID)

以 R 包 biomaRt (2.62.0) (2009, **IF:13.1**, Q1, Nature protocols)7 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)8) (<https://cluspro.bu.edu/home.php>)，上传需要对接的蛋白结构 (PDB) 。

## 2.9 MusiteDeep 蛋白质转录后修饰位点预测 (Dataset: ZDHHC\_LIPID)

以 biomaRt (2009, **IF:13.1**, Q1, Nature protocols)7 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)9 预测 S-palmitoyl\_cysteine 修饰位点，设定 PTM 得分截断为 0.5。

## 2.10 GEO 数据获取 (Dataset: GSE32981)

以 R 包 GEOquery (2.74.0) 获取 GSE32981 数据集。

## 2.11 Limma 差异分析 (Dataset: GSE32981)

以 limma (3.62.1) (2005)4 差异分析。创建设计矩阵，对比矩阵，差异分析：Met vs Prim。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 0.1 的统计结果。

## 2.12 GEO 数据获取 (Dataset: GSE14827)

以 R 包 GEOquery (2.74.0) 获取 GSE14827 数据集。

## 2.13 Limma 差异分析 (Dataset: GSE14827)

使用 log2 和 limma::normalizeBetweenArrays 对数据标准化。 以 limma (3.62.1) (2005)4 差异分析。创建设计矩阵，对比矩阵，差异分析：Metastasis vs Primary。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 0.1 的统计结果。

## 2.14 GEO 数据获取 (Dataset: GSE18947)

以 R 包 GEOquery (2.74.0) 获取 GSE18947 数据集。

## 2.15 Limma 差异分析 (Dataset: GSE18947)

使用 log2 和 limma::normalizeBetweenArrays 对数据标准化。 以 limma (3.62.1) (2005)4 差异分析。创建设计矩阵，对比矩阵，差异分析：High\_metastatic\_potential vs Low\_metastatic\_potential。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 0.3 的统计结果。

## 2.16 GEO 数据获取 (Dataset: GSE21257)

以 R 包 GEOquery (2.74.0) 获取 GSE21257 数据集。

## 2.17 Limma 差异分析 (Dataset: GSE21257)

以 limma (3.62.1) (2005)4 差异分析。创建设计矩阵，对比矩阵，差异分析：Metastasis vs Primary。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 0.3 的统计结果。

## 2.18 GEO 数据获取 (Dataset: GSE9508)

以 R 包 GEOquery (2.74.0) 获取 GSE9508 数据集。

## 2.19 Limma 差异分析 (Dataset: GSE9508)

以 limma (3.62.1) (2005)4 差异分析。创建设计矩阵，对比矩阵，差异分析：Biopsy\_metastatic vs Biopsy\_non\_metastatic, Biopsy\_non\_metastatic vs Non\_malignant\_bone。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 0.3 的统计结果。

## 2.20 GEO 数据获取 (Dataset: GSE237033)

以 R 包 GEOquery (2.74.0) 获取 GSE237033 数据集。 以 GEOquery:::getRNASeqQuantResults 获取 RNA count 数据 (NCBI-generated data, 参考 <https://www.ncbi.nlm.nih.gov/geo/info/rnaseqcounts.html>) 缺失样本: GSM7593289, GSM7593290, GSM7593295, GSM7593296, GSM7593299, GSM7593301 (‘NCBI-generated data’ 缺失样本计数数据的原因包括运行未通过 50% 的对齐率或由于技术原因处理失败) 以及基因注释。

## 2.21 Limma 差异分析 (Dataset: GSE237033)

以 R 包 edgeR (4.4.0) ()2 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)4 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：metastasis vs primary。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 0.3 的统计结果。

## 2.22 GEO 数据获取 (Dataset: GSE234998)

以 R 包 GEOquery (2.74.0) 获取 GSE234998 数据集。 以 GEOquery:::getRNASeqQuantResults 获取 RNA count 数据 (NCBI-generated data, 参考 <https://www.ncbi.nlm.nih.gov/geo/info/rnaseqcounts.html>) 缺失样本: GSM7488324, GSM7488325, GSM7488327, GSM7488333, GSM7488335 (‘NCBI-generated data’ 缺失样本计数数据的原因包括运行未通过 50% 的对齐率或由于技术原因处理失败) 以及基因注释。

## 2.23 Limma 差异分析 (Dataset: GSE234998)

以 R 包 edgeR (4.4.0) ()2 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)4 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：primary\_met vs localized。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 0.3 的统计结果。

## 2.24 GEO 数据获取 (Dataset: GSE14359)

以 R 包 GEOquery (2.74.0) 获取 GSE14359 数据集。

## 2.25 Limma 差异分析 (Dataset: GSE14359)

使用 log2 和 limma::normalizeBetweenArrays 对数据标准化。 以 limma (3.62.1) (2005)4 差异分析。创建设计矩阵，对比矩阵，差异分析：Metastasis vs Primary。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 0.3 的统计结果。

## 2.26 GEO 数据获取 (Dataset: NORMAL\_GSE220538)

以 R 包 GEOquery (2.74.0) 获取 GSE220538 数据集。 以 GEOquery:::getRNASeqQuantResults 获取 RNA count 数据 (NCBI-generated data, 参考 <https://www.ncbi.nlm.nih.gov/geo/info/rnaseqcounts.html>) 缺失样本: GSM6806600, GSM6806605, GSM6806611, GSM6806616, GSM6806617 (‘NCBI-generated data’ 缺失样本计数数据的原因包括运行未通过 50% 的对齐率或由于技术原因处理失败) 以及基因注释。

## 2.27 Limma 差异分析 (Dataset: GSE220538)

以 R 包 edgeR (4.4.0) ()2 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)4 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：metastatic\_tumor vs primary\_tumor。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 0.3 的统计结果。

## 2.28 Limma 差异分析 (Dataset: TARGET\_OS)

以 R 包 edgeR (4.4.0) ()2 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)4 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：Metastasis\_\_NOS vs No\_Metastasis。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 0.3 的统计结果。

## 2.29 GEO 数据获取 (Dataset: NORMAL\_GSE218035)

以 R 包 GEOquery (2.74.0) 获取 GSE218035 数据集。

## 2.30 Limma 差异分析 (Dataset: NORMAL\_GSE218035)

以 limma (3.62.1) (2005)4 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：Osteosarcoma\_sample vs Normal\_adjacent\_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 0.3 的统计结果。

## 2.31 GEO 数据获取 (Dataset: NORMAL\_GSE99671)

以 R 包 GEOquery (2.74.0) 获取 GSE99671 数据集。

## 2.32 Limma 差异分析 (Dataset: NORMAL\_GSE99671)

以 R 包 edgeR (4.4.0) ()2 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)4 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 vs NORMAL。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 0.3 的统计结果。

## 2.33 Limma 差异分析 (Dataset: META\_GSE220538)

以 R 包 edgeR (4.4.0) ()2 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)4 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：metastatic\_tumor vs primary\_tumor。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 adj.P.Val 小于 0.05，|Log2(FC)| 大于 1 的统计结果。

## 2.34 STRINGdb PPI 分析 (Dataset: LIPID\_ZDHHC11)

以 R 包 STEINGdb (2.18.0) (2021, **IF:16.6**, Q1, Nucleic Acids Research)5 构建 PPI 网络。数据版本为 12.0，互作类型为 full。以 Cytohubba (2014, BMC Systems Biology)6 的算法计算 MCC score (在 R 中计算) 。随后，以 ggraph 可视化网络 (2.2.1)。

## 2.35 ClusPro 蛋白质-蛋白质对接预测 (Dataset: ZDHHC11\_LIPID)

以 R 包 biomaRt (2.62.0) (2009, **IF:13.1**, Q1, Nature protocols)7 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)8) (<https://cluspro.bu.edu/home.php>)，上传需要对接的蛋白结构 (PDB) 。

## 2.36 MusiteDeep 蛋白质转录后修饰位点预测 (Dataset: ZDHHC11\_LIPID)

以 biomaRt (2009, **IF:13.1**, Q1, Nature protocols)7 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)9 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 TCGA 数据获取 (OS)

获取 TARGET-OS 数据。

## 3.2 Survival 生存分析 (OS)

按 survminer::surv\_cutpoint 计算的 cutoff，将样本分为 Low 和 High 风险组。生存数据为TARGET-OS，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。根据 P value < 0.05, 共筛到 14 个特征。 分别为 ZDHHC8, ZDHHC15, ZDHHC2, ZDHHC7, ZDHHC3, ZDHHC24, ZDHHC21, ZDHHC13, ZDHHC22, ZDHHC23, ZDHHC11, ZDHHC11B, ZDHHC20P4, ZDHHC20P3。 Tab. **[1](#OS-Significant-Survival-PValue)** Fig. **[1](#OS-survival-curve-of-ZDHHC11)** 为 ZDHHC11 生存曲线。

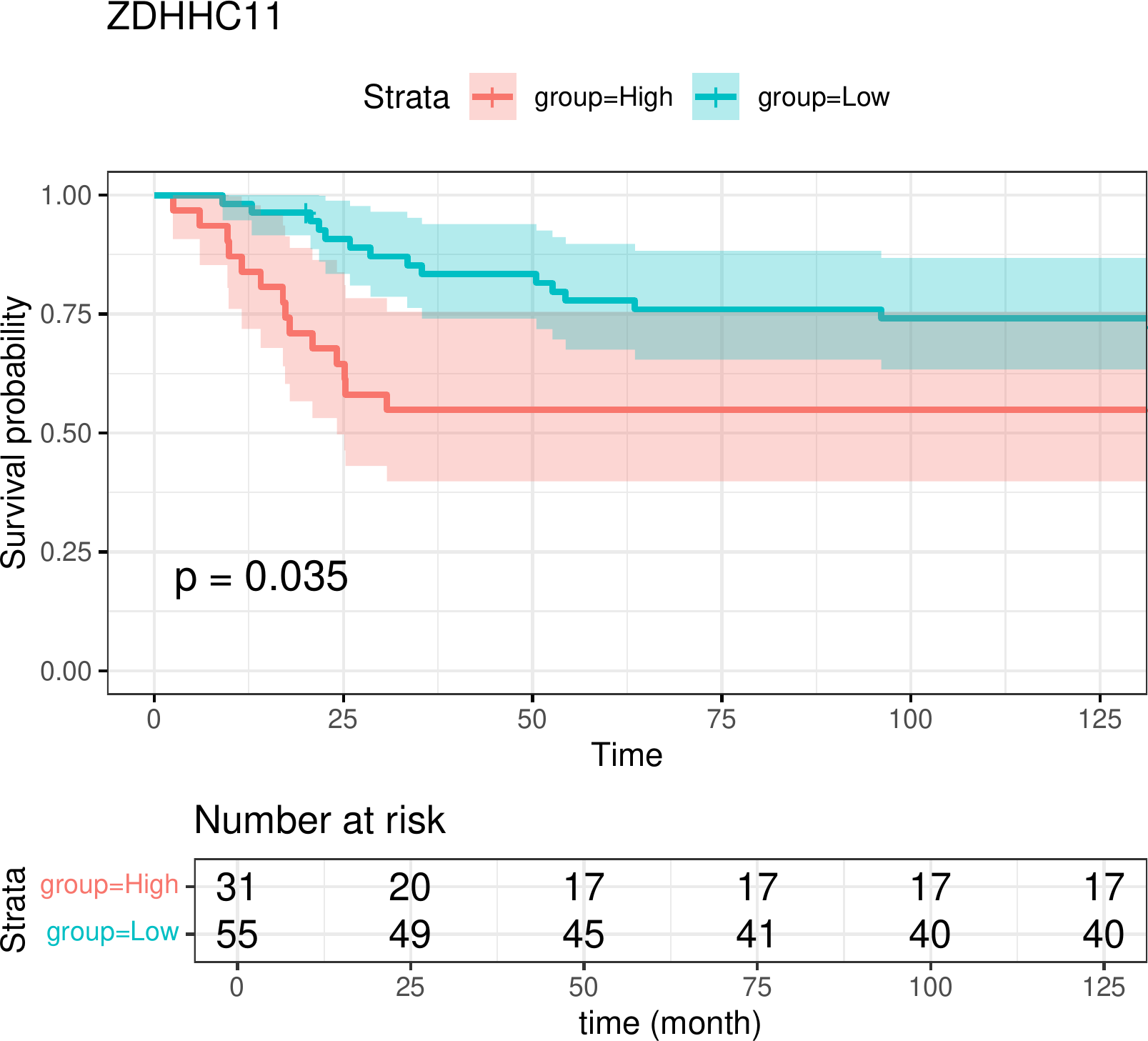
**Tab.** **1** OS Significant Survival PValue

| Name | Pvalue | Group low survival |
| --- | --- | --- |
| ZDHHC8 | 0.02109 | Low |
| ZDHHC15 | 0.003226 | Low |
| ZDHHC2 | 0.01969 | High |
| ZDHHC7 | 0.01439 | Low |
| ZDHHC3 | 0.0005188 | Low |
| ... | ... | ... |

**(File path: Figure+Table/3.2.0\_Survival\_生存分析\_(OS)/OS-Significant-Survival-PValue.csv)**

Note: The directory 'Figure+Table/OS-survival-curve-of-alls' contains 14 files.  
  
1 1\_ZDHHC8.pdf  
2 10\_ZDHHC23.pdf  
3 11\_ZDHHC11.pdf  
4 12\_ZDHHC11B.pdf  
5 13\_ZDHHC20P4.pdf  
6 ...

**(File path: Figure+Table/3.2.0\_Survival\_生存分析\_(OS)/OS-survival-curve-of-alls)**



**Fig.** **1** OS survival curve of ZDHHC11

**(File path: Figure+Table/3.2.0\_Survival\_生存分析\_(OS)/OS-survival-curve-of-ZDHHC11.pdf)**

## 3.3 GeneCards 基因获取 (LIPID)

从 GeneCards 搜索 lipid metabolism, 获取对应靶点数据，统计为 Functional Element (n=15) , Genetic Locus (n=2) , Protein Coding (n=3211) , Pseudogene (n=3) , RNA Gene (lncRNA) (n=59) , RNA Gene (miRNA) (n=104) , RNA Gene (scaRNA) (n=1) , RNA Gene (snoRNA) (n=2) , RNA Gene (snRNA) (n=2) , RNA Gene (tRNA) (n=11) 。共 3410 个靶点。

Tab. **[2](#LIPID-disease-related-targets-from-GeneCards)** 为 GeneCards 检索 (lipid metabolism) 得到的基因集。

**Tab.** **2** LIPID disease related targets from GeneCards

| Symbol | Description | Category | UniProt ID | GIFtS |
| --- | --- | --- | --- | --- |
| APOE | Apolipoprotein E | Protein Coding | P02649 | 63 |
| LDLR | Low Density Lipop... | Protein Coding | P01130 | 64 |
| PPARG | Peroxisome Prolif... | Protein Coding | P37231 | 66 |
| CETP | Cholesteryl Ester... | Protein Coding | P11597 | 60 |
| LIPC | Lipase C, Hepatic... | Protein Coding | P11150 | 59 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.3.0\_GeneCards\_基因获取\_(LIPID)/LIPID-disease-related-targets-from-GeneCards.xlsx)**

* The GeneCards data was obtained by querying: lipid metabolism
* Restrict (with quotes): TRUE
* Filtering by Score:: Score > 1

## 3.4 ZDHHC15 分析

### 3.4.1 GEO 数据获取 (GSE87624)

以 GEOquery 获取 GSE87624 的数据信息。

### 3.4.2 Limma 差异分析 (GSE87624\_ZDHHC)

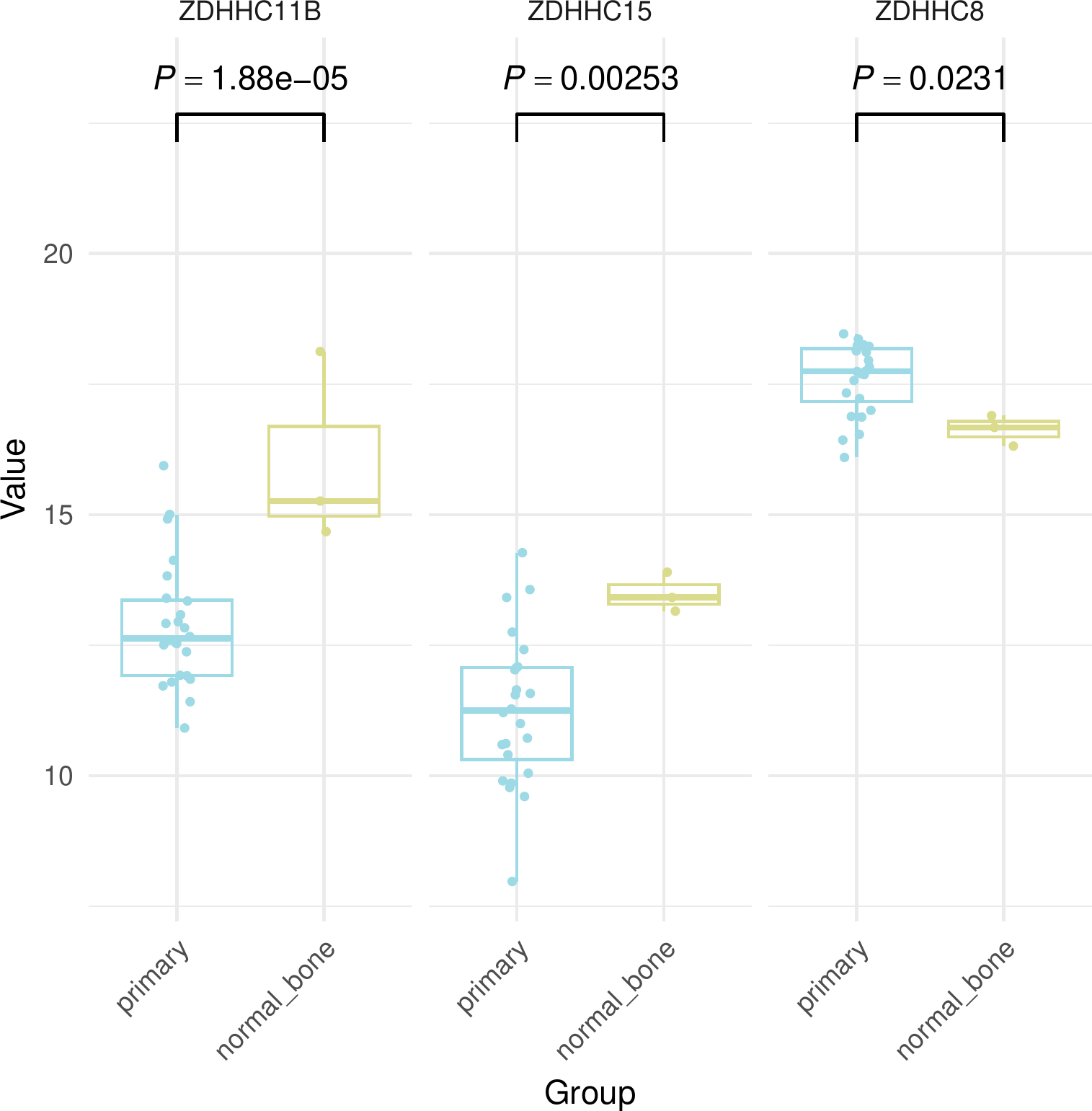
筛选数据集中的基因，以ZDHHC3, ZDHHC23, ZDHHC11B, …(n = 12)差异分析。以 edgeR 将GSE87624 RNA-seq 数据标准化 (详见方法章节)。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：metastasis vs primary, primary vs normal\_bone。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。各组差异分析 DEGs 统计：

* metastasis vs primary：up (n=1) , down (n=0) 。
* primary vs normal\_bone：up (n=1) , down (n=2) 。 所有上调 DEGs 共 2 个，所有下调 DEGs 共 2 个。所有非重复 DEGs 共 3 个。

聚焦于基因集 (ZDHHC8, ZDHHC15, ZDHHC2, …[n = 14], 来自于Survival 生存分析[Section: OS]) 的差异表达 (primary - normal\_bone)。 Fig. **[2](#GSE87624-ZDHHC-Box-Plot-Of-DEGs-normal)** 基因 ZDHHC8, ZDHHC15, ZDHHC11B 表达水平，以及对应的 limma 差异分析显著水平。 Tab. **[3](#Statistic-of-Focused-genes-normal)** 为聚焦分析的基因的统计附表。

筛选数据集中的基因，以ZDHHC3, ZDHHC23, ZDHHC11B, …(n = 12)差异分析。以 edgeR 将GSE87624 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：metastasis vs primary, primary vs normal\_bone。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。各组差异分析 DEGs 统计：

* metastasis vs primary：up (n=1) , down (n=0) 。
* primary vs normal\_bone：up (n=1) , down (n=2) 。 所有上调 DEGs 共 2 个，所有下调 DEGs 共 2 个。所有非重复 DEGs 共 3 个。 Fig. **[3](#GSE87624-ZDHHC-Box-Plot-Of-DEGs-metastasis)** 基因 ZDHHC15 表达水平，以及对应的 limma 差异分析显著水平。 Tab. **[4](#Statistic-of-Focused-genes-metastasis)** 为聚焦分析的基因的统计附表。



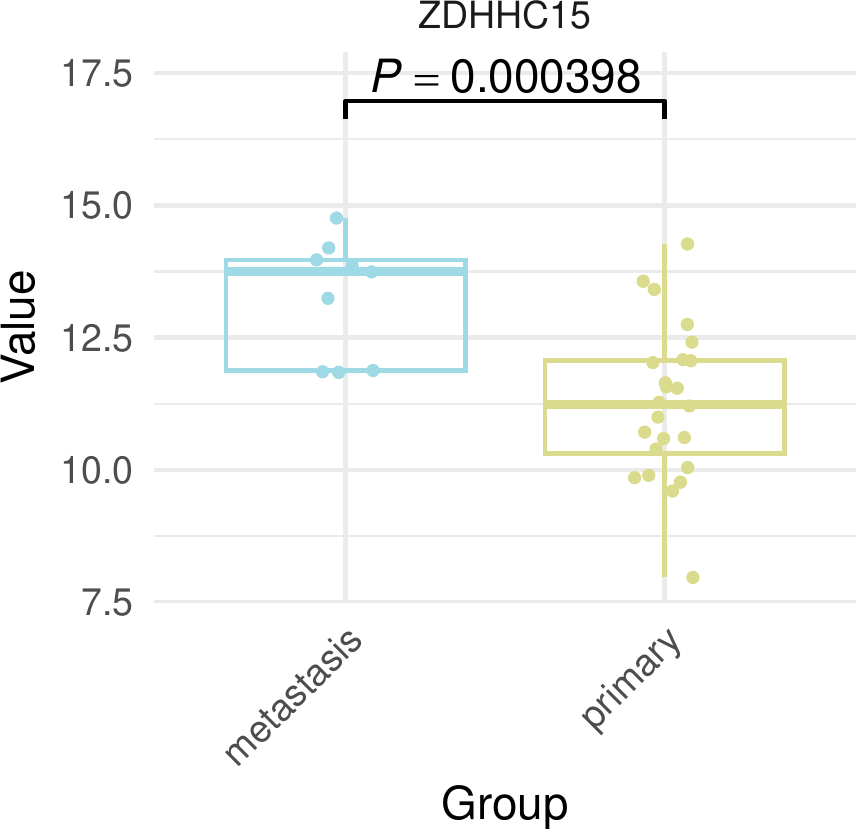
**Fig.** **2** GSE87624 ZDHHC Box Plot Of DEGs normal

**(File path: Figure+Table/3.4.2\_Limma\_差异分析\_(GSE87624\_ZDHHC)/GSE87624-ZDHHC-Box-Plot-Of-DEGs-normal.pdf)**

**Tab.** **3** Statistic of Focused genes normal

| Hgnc symbol | LogFC | Adj.P.Val | P.Value | Rownames |
| --- | --- | --- | --- | --- |
| ZDHHC11B | -2.848 | 0.0002257 | 1.881e-05 | 653082 |
| ZDHHC15 | -2.239 | 0.01517 | 0.002528 | 158866 |
| ZDHHC8 | 0.9797 | 0.09238 | 0.0231 | 29801 |

**(File path: Figure+Table/3.4.2\_Limma\_差异分析\_(GSE87624\_ZDHHC)/Statistic-of-Focused-genes-normal.csv)**



**Fig.** **3** GSE87624 ZDHHC Box Plot Of DEGs metastasis

**(File path: Figure+Table/3.4.2\_Limma\_差异分析\_(GSE87624\_ZDHHC)/GSE87624-ZDHHC-Box-Plot-Of-DEGs-metastasis.pdf)**

**Tab.** **4** Statistic of Focused genes metastasis

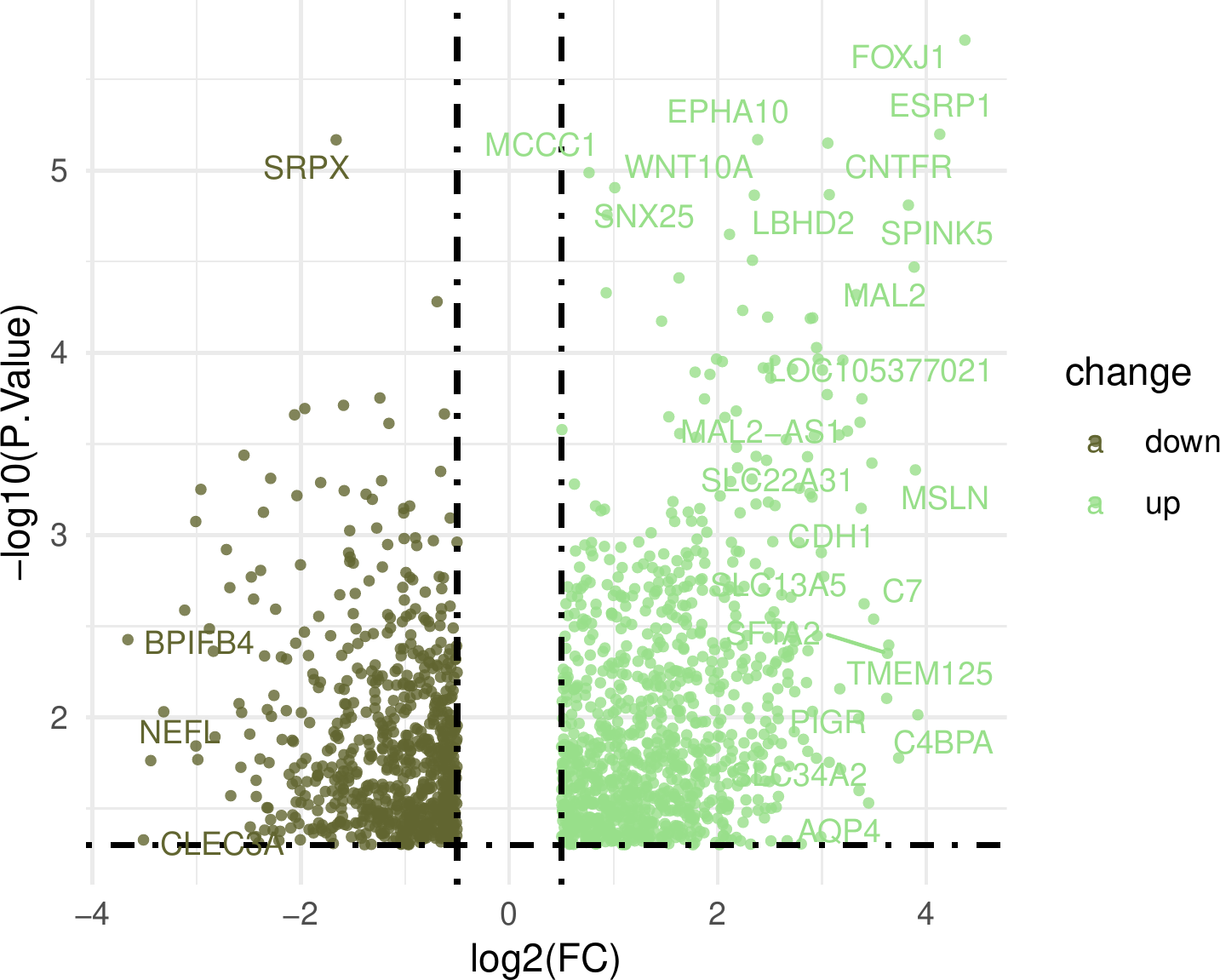
| Hgnc symbol | LogFC | Adj.P.Val | P.Value | Rownames |
| --- | --- | --- | --- | --- |
| ZDHHC15 | 2.044 | 0.004782 | 0.0003985 | 158866 |

**(File path: Figure+Table/3.4.2\_Limma\_差异分析\_(GSE87624\_ZDHHC)/Statistic-of-Focused-genes-metastasis.csv)**

### 3.4.3 Limma 差异分析 (GSE87624)

匹配 group 中包含“metastasis|primary”的描述，最终得到 33 例数据。样本分组：metastasis (n=9) , primary (n=24) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：metastasis vs primary。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DEGs 统计：up (n=1034) , down (n=740)

Fig. **[4](#GSE87624-metastasis-vs-primary)** 为 metastasis - primary 差异分析火山图。 Tab. **[5](#GSE87624-data-metastasis-vs-primary)** 为 metastasis - primary 差异分析统计表格。



**Fig.** **4** GSE87624 metastasis vs primary

**(File path: Figure+Table/3.4.3\_Limma\_差异分析\_(GSE87624)/GSE87624-metastasis-vs-primary.pdf)**

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

**(See: Figure+Table/3.4.3\_Limma\_差异分析\_(GSE87624)/GSE87624-metastasis-vs-primary-content)**

**Tab.** **5** GSE87624 data metastasis vs primary

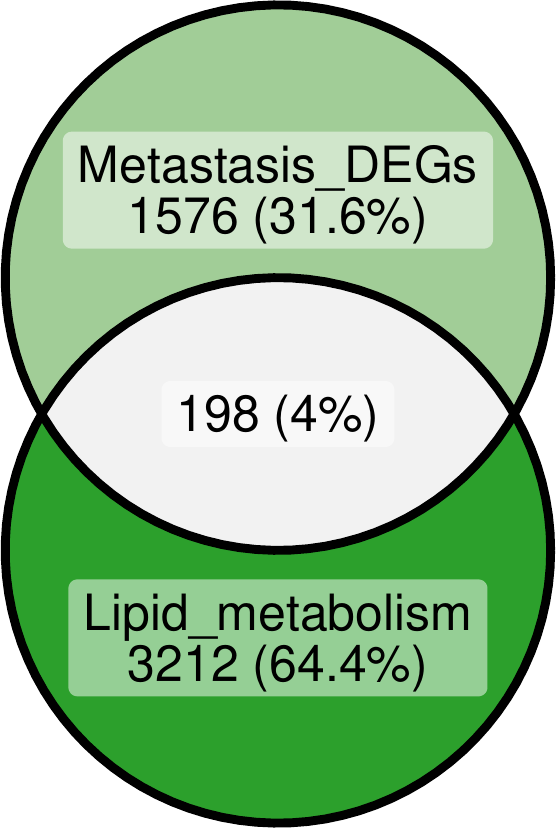
| Hgnc symbol | LogFC | P.Value | Rownames | GeneID |
| --- | --- | --- | --- | --- |
| FOXJ1 | 4.37 | 1.923e-06 | 2302 | 2302 |
| SRPX | -1.661 | 6.781e-06 | 8406 | 8406 |
| CNTFR | 3.055 | 7.068e-06 | 1271 | 1271 |
| MCCC1 | 0.7649 | 1.026e-05 | 56922 | 56920 |
| ESRP1 | 4.129 | 6.319e-06 | 54845 | 54840 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.4.3\_Limma\_差异分析\_(GSE87624)/GSE87624-data-metastasis-vs-primary.tsv)**

### 3.4.4 交集: Lipid\_metabolism + Metastasis\_DEGs (LIPID)

以下取交集： - 基因集 (来自于GeneCards 基因获取[Section: LIPID]) - 基因集 (metastasis - primary, 来自于Limma 差异分析[Section: GSE87624])

Fig. **[5](#Intersection-of-Lipid-metabolism-with-Metastasis-DEGs)** 将Lipid\_metabolism, Metastasis\_DEGs 取交集。



**Fig.** **5** Intersection of Lipid metabolism with Metastasis DEGs

**(File path: Figure+Table/3.4.4\_交集:\_Lipid\_metabolism\_+\_Metastasis\_DEGs\_(LIPID)/Intersection-of-Lipid-metabolism-with-Metastasis-DEGs.pdf)**

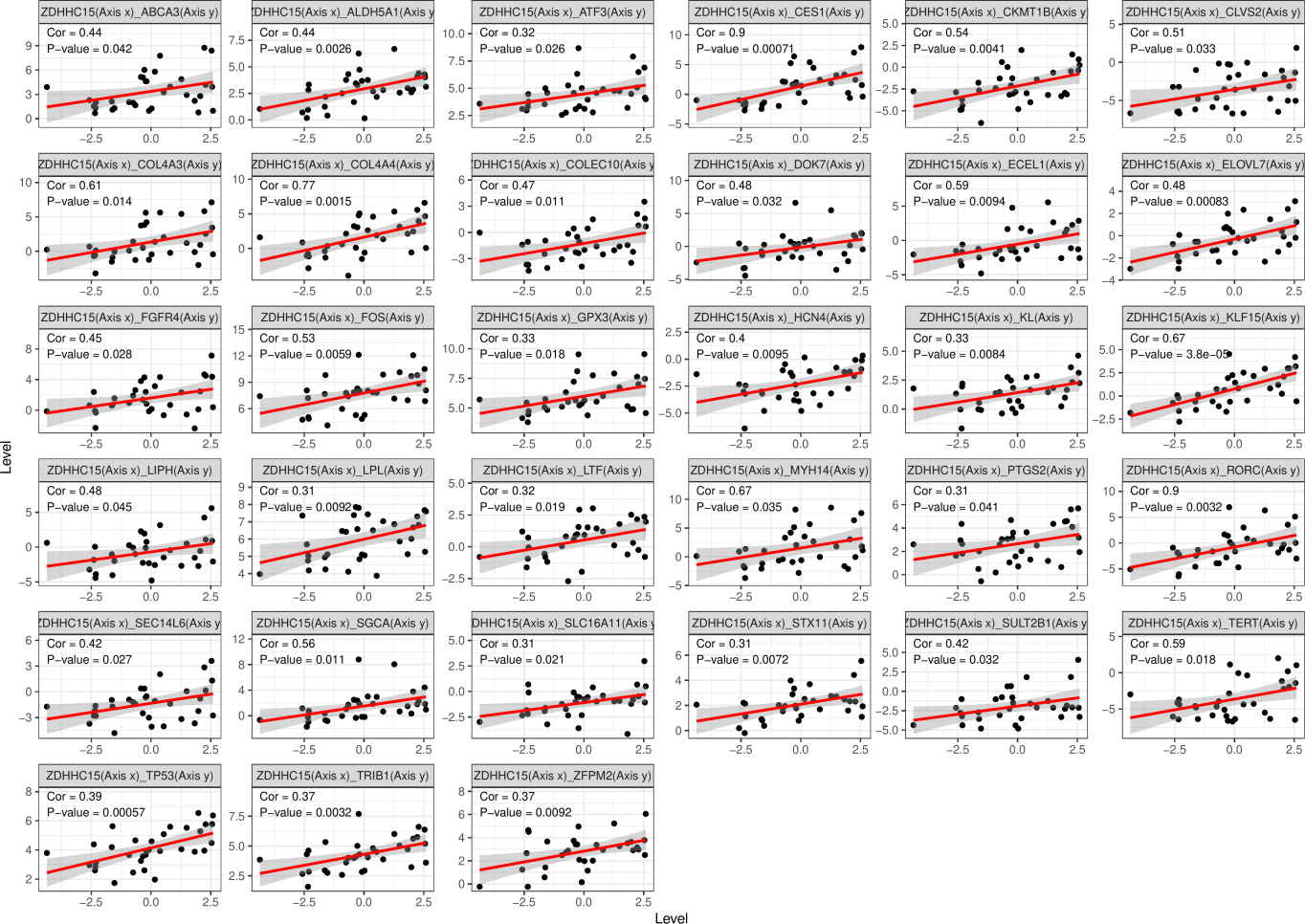
* All\_intersection: SLC52A2, LPL, ALDH5A1, PCSK9, LMNA, TP53, PLTP, PTGS2, MTHFR, ENHO, SLC27A2, ALB, LRP1, MBOAT7, PRKAA2, DHCR24, G6PD, PNPLA1, IL1B, BDNF, GPX8, CES1, MMP9, KL, FOS, SPP1, FN1, ACTB, VDR, NCOR1, CKB, CKMT1B, ITGA2B, FA2H, ACOT4, GLP1R, GPX3, CD4, TRIB1, FGFR4, LRP2, FLNA, KHK, AMY1B, EHHADH, ABCA3, MYH14, PXDN, GM2A, TGFBR1, ABCD2, C19orf12, MCCC1, SLC16A11, SLC6A8, RPGR, ATG14, GLB1, ITGB3, PEMT, KBTBD13, OSBPL10, HEXB, TPP1, LTBR, GALT, OTOF, NAGA, CTSK, LAMP3, JAZF1, GRN, TCIRG1, FBN1, VIM,…

**(See: Figure+Table/3.4.4\_交集:\_Lipid\_metabolism\_+\_Metastasis\_DEGs\_(LIPID)/Intersection-of-Lipid-metabolism-with-Metastasis-DEGs-content)**

### 3.4.5 关联分析 (GSE87624)

将基因 (ZDHHC15 -> 基因集 (SLC52A2, LPL, ALDH5A1, …[n = 198], 来自于Venn 交集[Section: LIPID]) ) 关联分析。共得到 33 个显著的基因对 (P < 0.05, |Cor| > 0.3)。

Fig. **[6](#GSE87624-significant-correlation-plots)** 为显著关联的基因的线型回归图。 Tab. **[6](#GSE87624-significant-correlation-analysis-data)** 为关联分析统计附表 (P-value cutoff: 0.05, Cor (关联系数) cutoff: 0.3)。 Fig. **[7](#GSE87624-Box-Plot-Of-DEGs)** 基因 ZDHHC15, RORC, CES1, COL4A4, KLF15, MYH14, COL4A3, ECEL1, TERT, SGCA, CKMT1B, FOS, CLVS2, DOK7, ELOVL7, LIPH, COLEC10, FGFR4, ALDH5A1, ABCA3, SEC14L6, SULT2B1, HCN4, TP53, ZFPM2, TRIB1, KL, GPX3, LTF, ATF3, PTGS2, LPL, SLC16A11, STX11 表达水平，以及对应的 limma 差异分析显著水平。



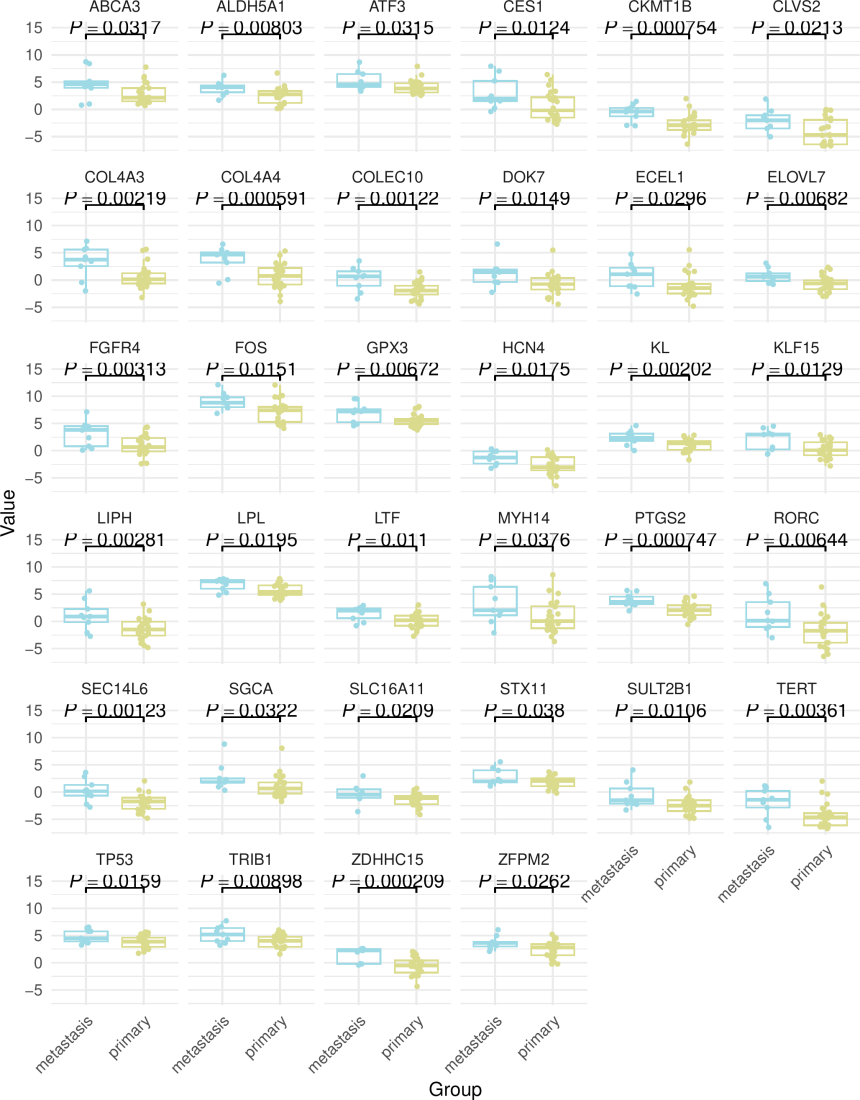
**Fig.** **6** GSE87624 significant correlation plots

**(File path: Figure+Table/3.4.5\_关联分析\_(GSE87624)/GSE87624-significant-correlation-plots.pdf)**

**Tab.** **6** GSE87624 significant correlation analysis data

| From | To | Cor | Pvalue | Model |
| --- | --- | --- | --- | --- |
| ZDHHC15 | RORC | 0.9013 | 0.003246 | C(GSM2335686 = 3.... |
| ZDHHC15 | CES1 | 0.9011 | 0.0007054 | C(GSM2335686 = 7.... |
| ZDHHC15 | COL4A4 | 0.7732 | 0.00149 | C(GSM2335686 = 5.... |
| ZDHHC15 | KLF15 | 0.675 | 3.772e-05 | C(GSM2335686 = 2.... |
| ZDHHC15 | MYH14 | 0.6698 | 0.03452 | C(GSM2335686 = 6.... |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.4.5\_关联分析\_(GSE87624)/GSE87624-significant-correlation-analysis-data.xlsx)**



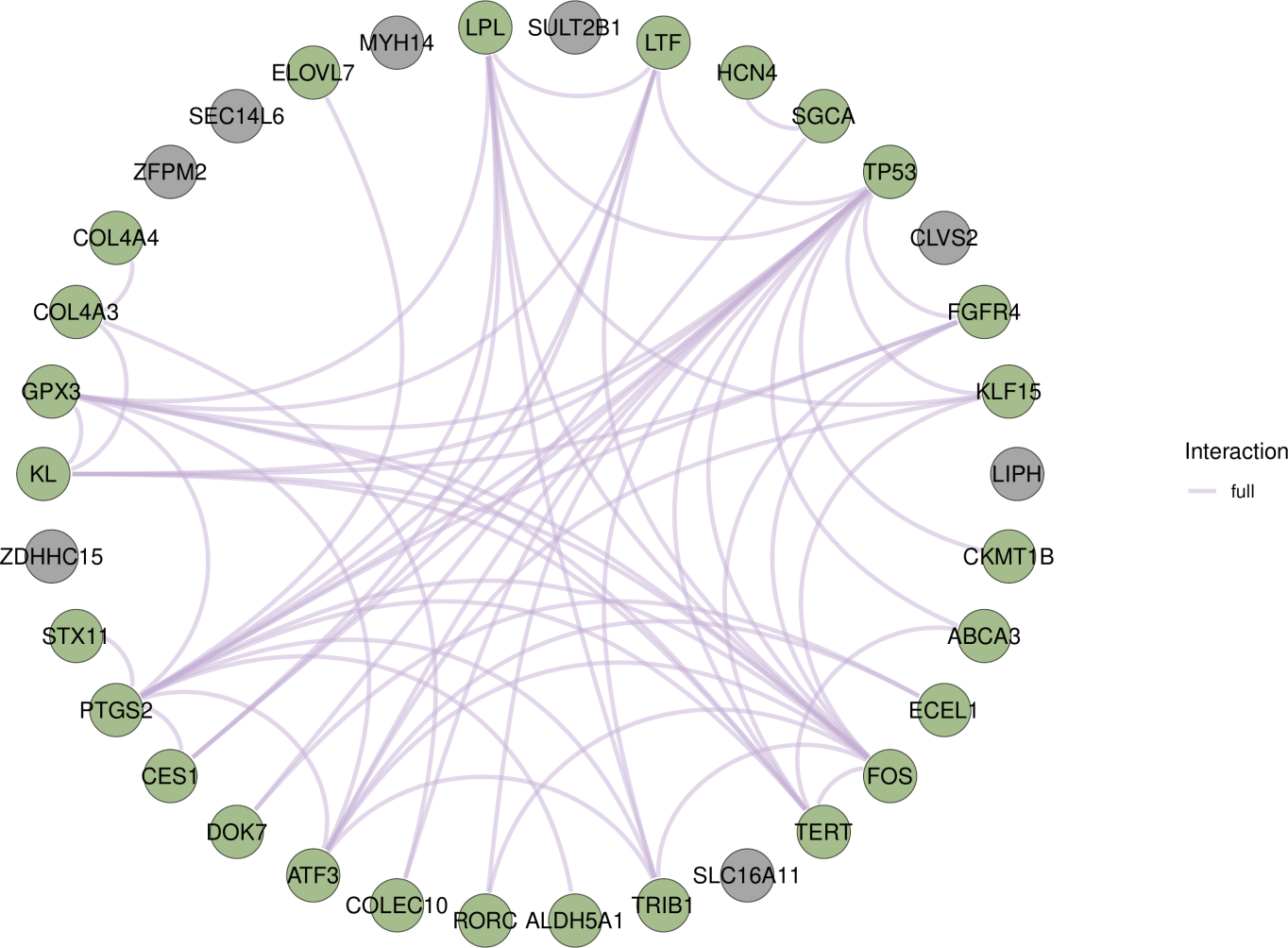
**Fig.** **7** GSE87624 Box Plot Of DEGs

**(File path: Figure+Table/3.4.5\_关联分析\_(GSE87624)/GSE87624-Box-Plot-Of-DEGs.pdf)**

### 3.4.6 STRINGdb PPI 分析 (LIPID\_ZDHHC)

对基因集 (from, to, 来自于关联分析[Section: GSE87624]) 进行STRINGdb PPI 分析。

Fig. **[8](#LIPID-ZDHHC-Top-MCC-score)** PPI (带有 Cytohubba (2014, BMC Systems Biology)6 MCC 得分) 网络图



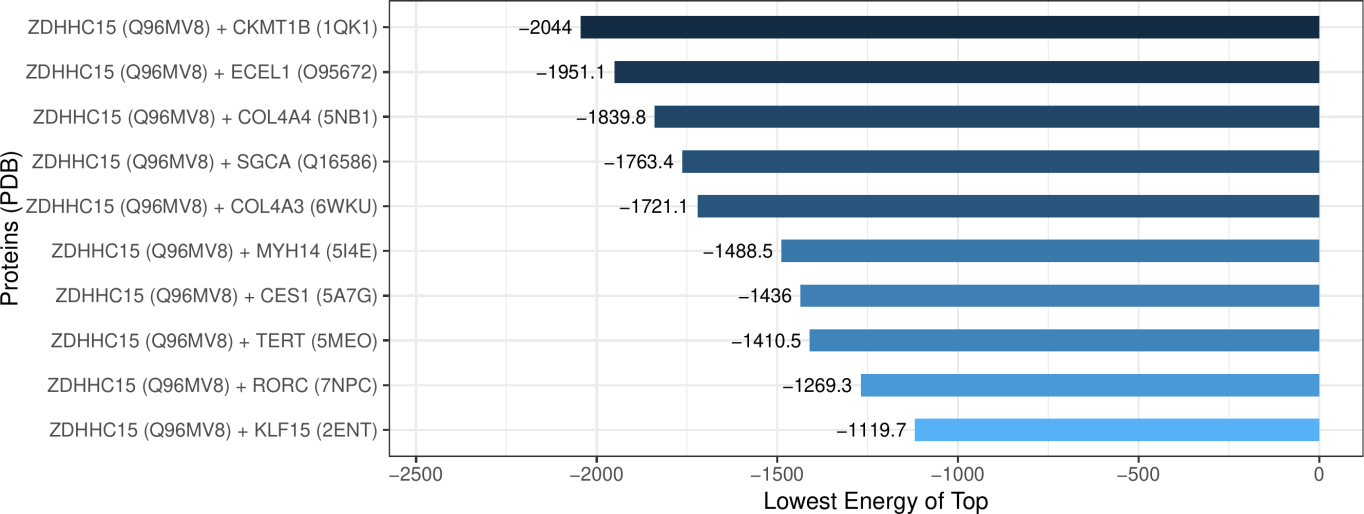
**Fig.** **8** LIPID ZDHHC Top MCC score

**(File path: Figure+Table/3.4.6\_STRINGdb\_PPI\_分析\_(LIPID\_ZDHHC)/LIPID-ZDHHC-Top-MCC-score.pdf)**

### 3.4.7 ClusPro 蛋白质-蛋白质对接预测 (ZDHHC\_LIPID)

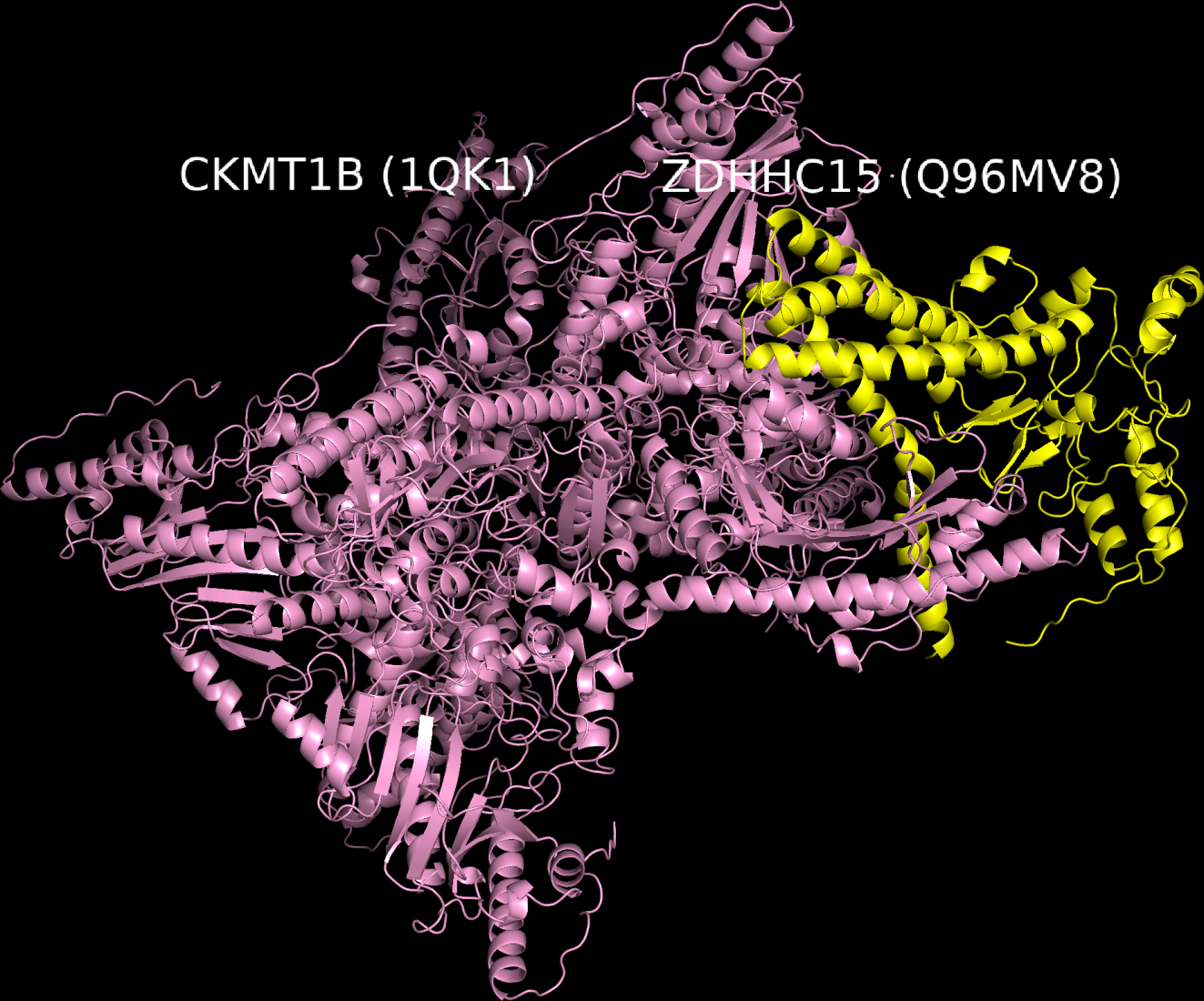
(取关联分析的关联系数排名 Top 10 的蛋白质对) 对基因集 (from, to, 来自于关联分析[Section: GSE87624]) 进行ClusPro 蛋白质-蛋白质对接预测。以 biomaRt 获取基因 Symbol 对应的蛋白结构 (PDB，详见方法章节)。选取分辨率最高 (即，resolution 值最小) 的 PDB 作为分子对接的蛋白结构。从 RCSB PDB 获取 PDB 文件。对于未从 PDB 数据库找到结构文件的，从数据库 AlphaFold 获取 ZDHHC15, ECEL1, SGCA 预测的蛋白结构 (根据 UniProtKB-Swiss-Prot ID，详见方法章节)。将 PDB 上传至 ClusPro 进行对接。

Fig. **[9](#ZDHHC-LIPID-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. **[10](#Protein-docking-of-ZDHHC15-CKMT1B)** 以 pymol 将蛋白质 (Top\_1\_ZDHHC15\_CKMT1B) 对接结果可视化。 Fig. **[11](#Protein-docking-of-ZDHHC15-ECEL1)** 以 pymol 将蛋白质 (Top\_2\_ZDHHC15\_ECEL1) 对接结果可视化。 Fig. **[12](#Protein-docking-of-ZDHHC15-COL4A4)** 以 pymol 将蛋白质 (Top\_3\_ZDHHC15\_COL4A4) 对接结果可视化。



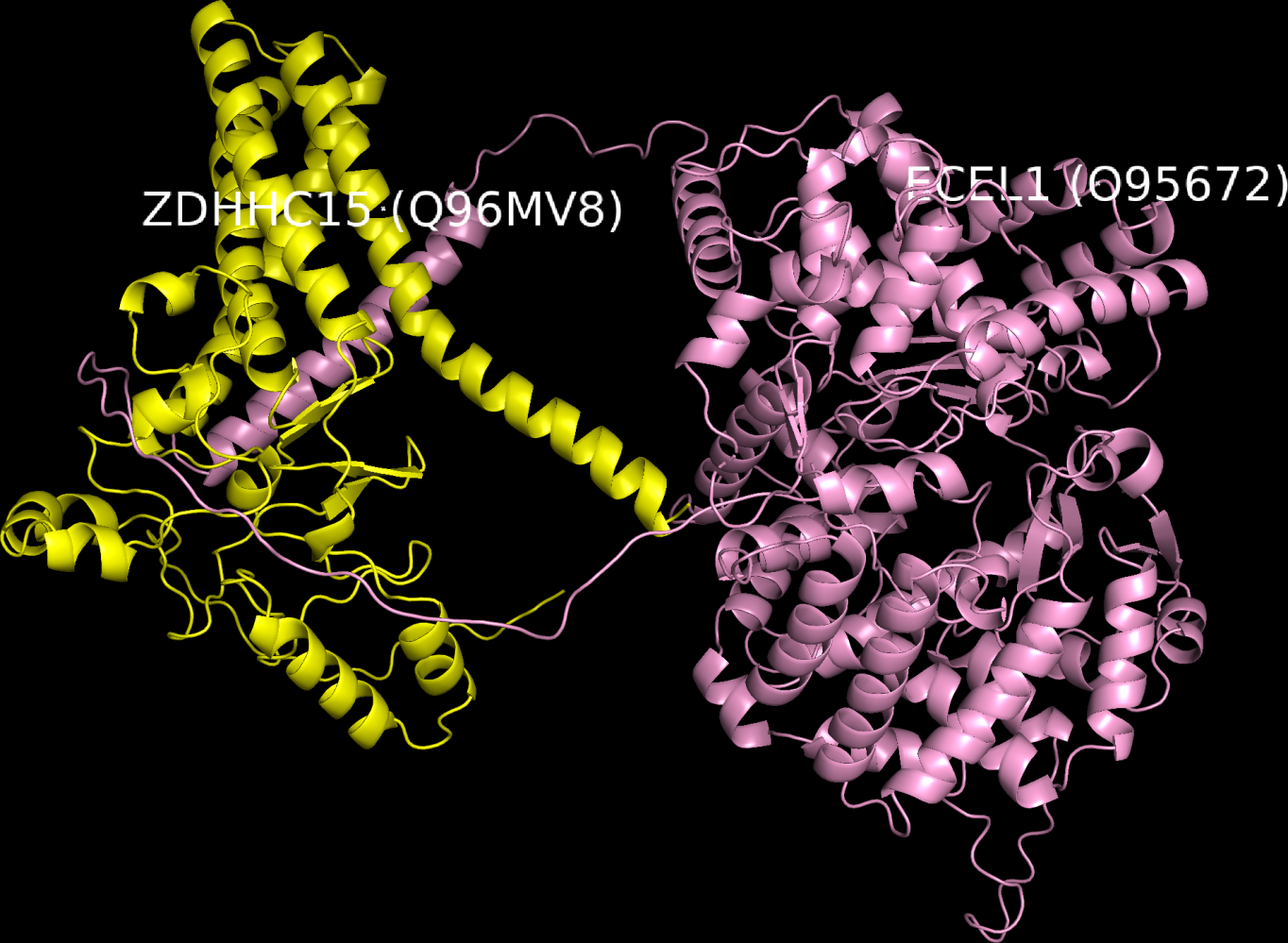
**Fig.** **9** ZDHHC LIPID Overview of protein docking results

**(File path: Figure+Table/3.4.7\_ClusPro\_蛋白质-蛋白质对接预测\_(ZDHHC\_LIPID)/ZDHHC-LIPID-Overview-of-protein-docking-results-.pdf)**



**Fig.** **10** Protein docking of ZDHHC15 CKMT1B

**(File path: Figure+Table/3.4.7\_ClusPro\_蛋白质-蛋白质对接预测\_(ZDHHC\_LIPID)/Protein-docking-of-ZDHHC15-CKMT1B.png)**



**Fig.** **11** Protein docking of ZDHHC15 ECEL1

**(File path: Figure+Table/3.4.7\_ClusPro\_蛋白质-蛋白质对接预测\_(ZDHHC\_LIPID)/Protein-docking-of-ZDHHC15-ECEL1.png)**



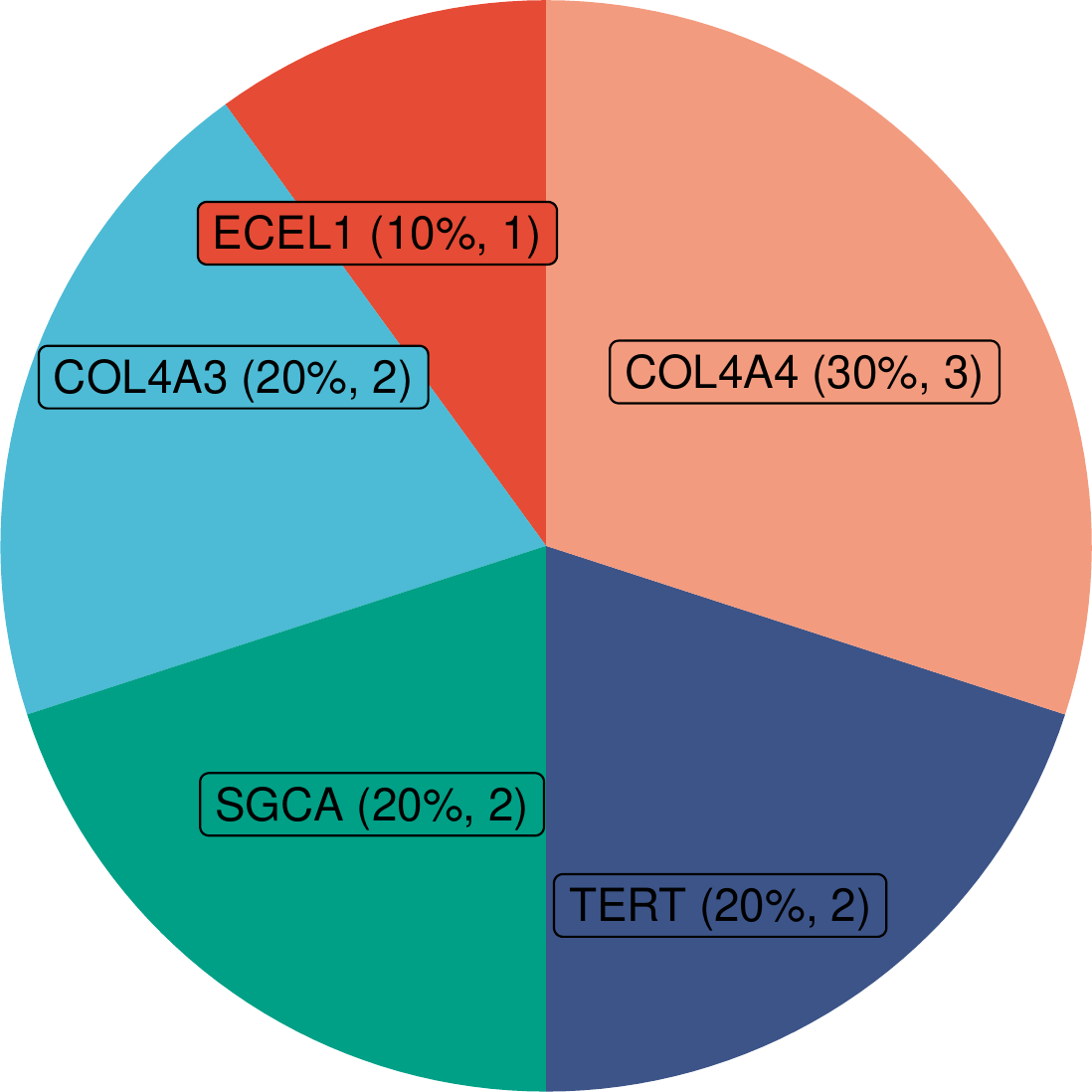
**Fig.** **12** Protein docking of ZDHHC15 COL4A4

**(File path: Figure+Table/3.4.7\_ClusPro\_蛋白质-蛋白质对接预测\_(ZDHHC\_LIPID)/Protein-docking-of-ZDHHC15-COL4A4.png)**

### 3.4.8 MusiteDeep 蛋白质转录后修饰位点预测 (ZDHHC\_LIPID)

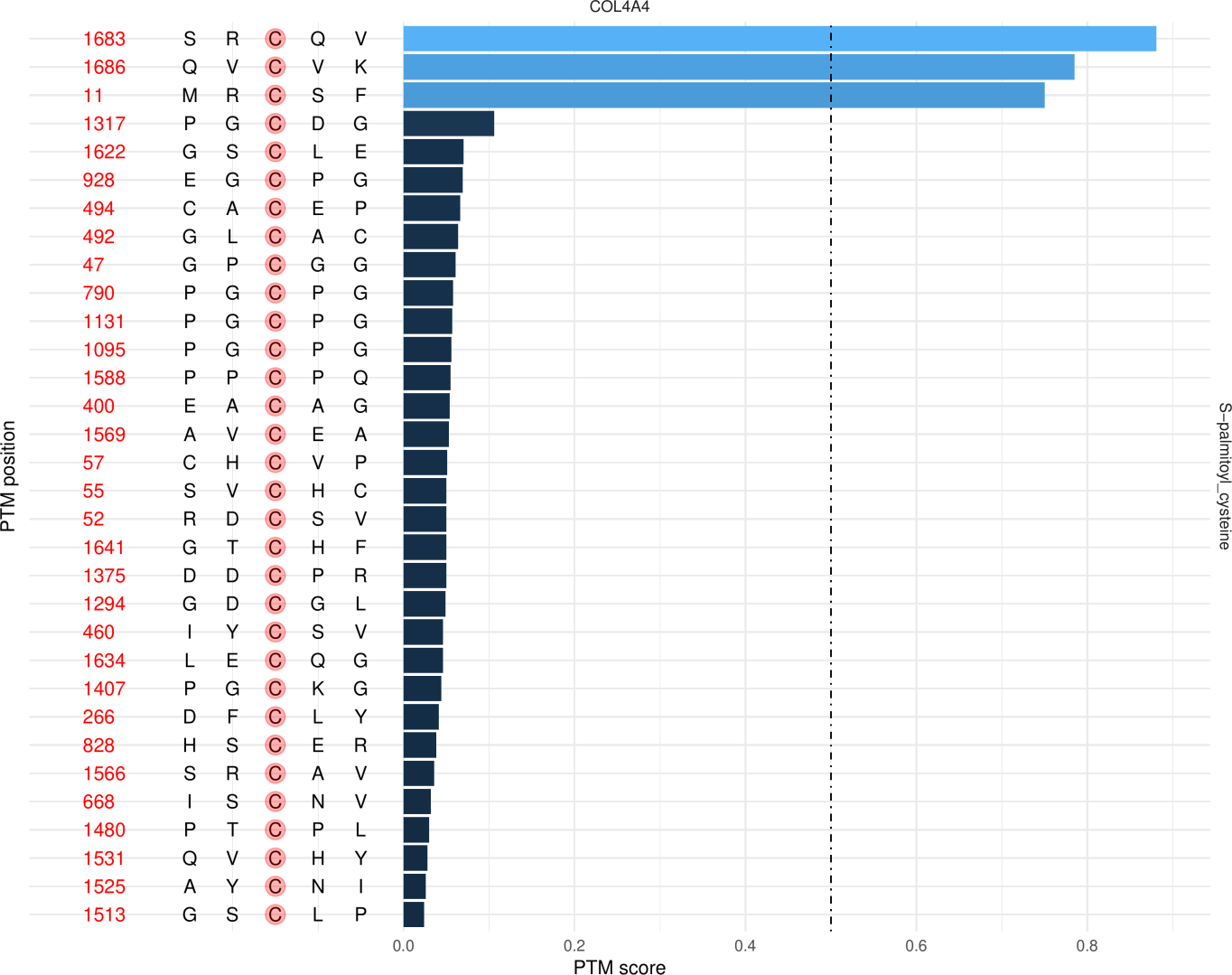
对基因集 (RORC, CES1, COL4A4, …[n = 10], 来自于关联分析[Section: GSE87624]) 进行MusiteDeep 蛋白质转录后修饰位点预测。以 biomaRt 获取蛋白质 (RORC, CES1, COL4A4, …(n = 10)) 的序列 (Peptide)。以 MusiteDeep 预测 S-palmitoyl\_cysteine 修饰位点。

Fig. **[13](#ZDHHC-LIPID-S-palmitoyl-cysteine-PTM-numbers)** 为预测到的 PTMs 数量饼图。 Fig. **[14](#ZDHHC-LIPID-COL4A4-PTM-score)** 为 COL4A4 的修饰位点以及得分可视化图。 Fig. **[15](#ZDHHC-LIPID-ECEL1-PTM-score)** 为 ECEL1 的修饰位点以及得分可视化图。 Fig. **[16](#GSE87624-Box-Plot-Of-DEGs-final)** 基因 ZDHHC15, ECEL1, COL4A4 表达水平，以及对应的 limma 差异分析显著水平。



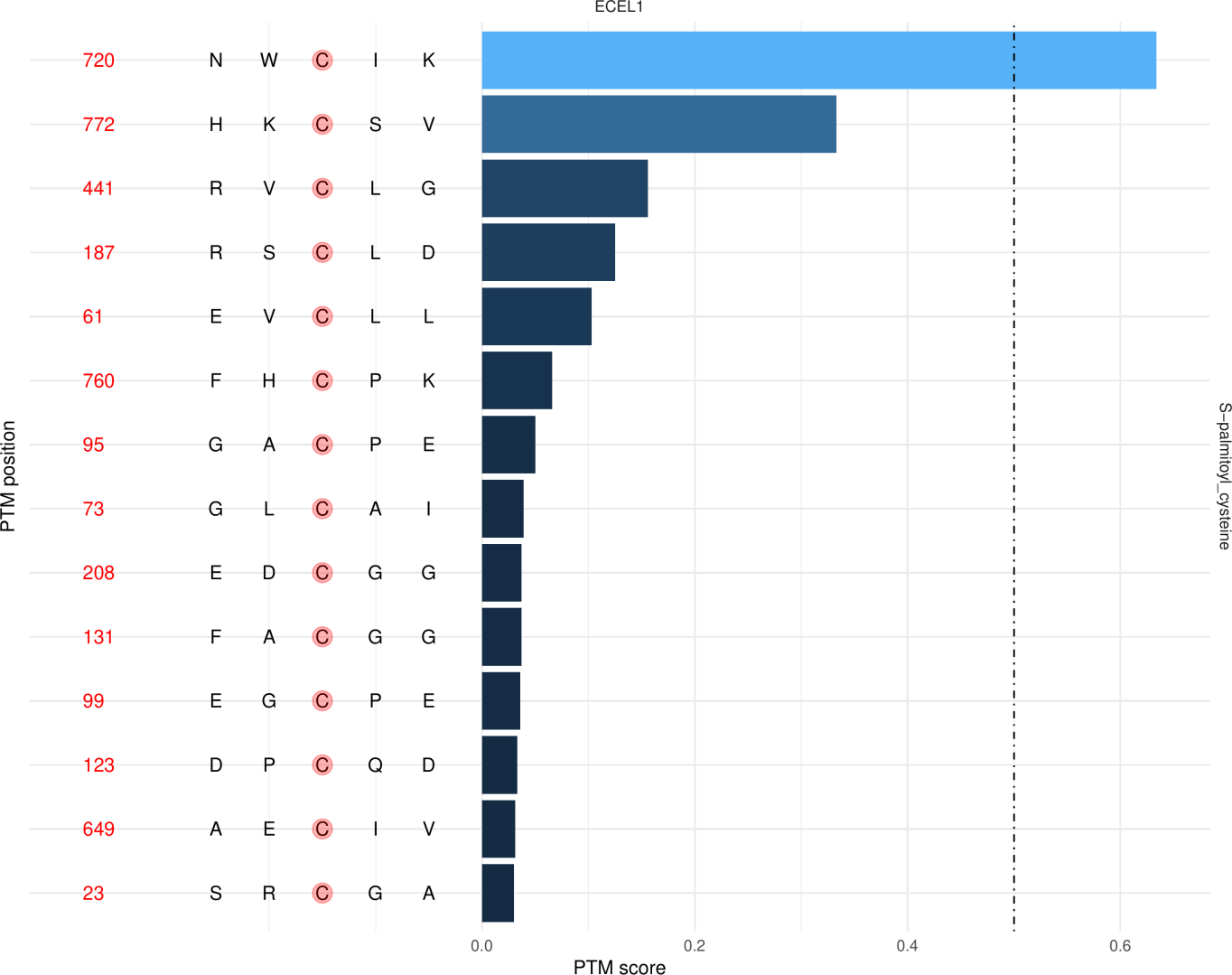
**Fig.** **13** ZDHHC LIPID S palmitoyl cysteine PTM numbers

**(File path: Figure+Table/3.4.8\_MusiteDeep\_蛋白质转录后修饰位点预测\_(ZDHHC\_LIPID)/ZDHHC-LIPID-S-palmitoyl-cysteine-PTM-numbers.pdf)**



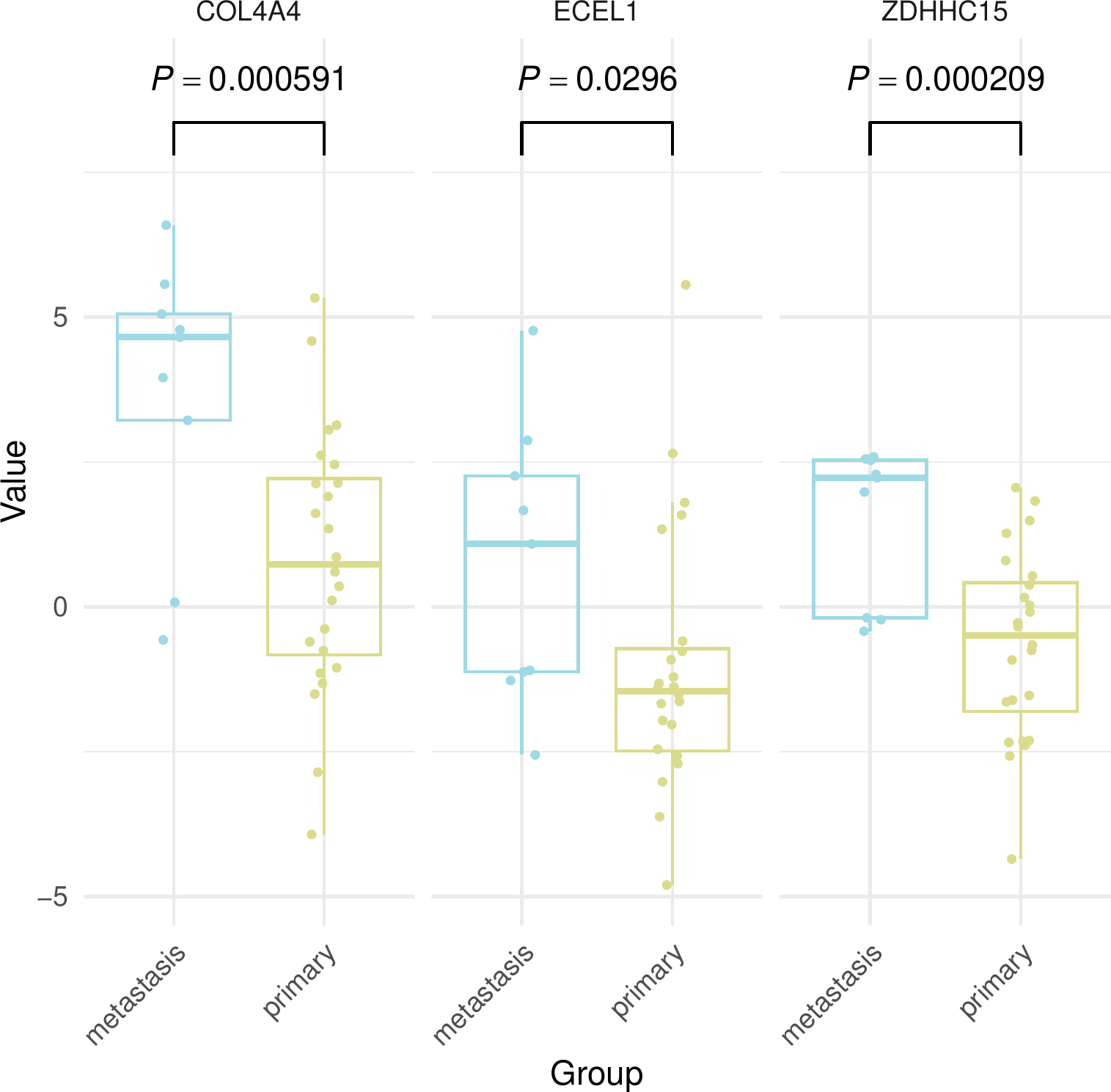
**Fig.** **14** ZDHHC LIPID COL4A4 PTM score

**(File path: Figure+Table/3.4.8\_MusiteDeep\_蛋白质转录后修饰位点预测\_(ZDHHC\_LIPID)/ZDHHC-LIPID-COL4A4-PTM-score.pdf)**



**Fig.** **15** ZDHHC LIPID ECEL1 PTM score

**(File path: Figure+Table/3.4.8\_MusiteDeep\_蛋白质转录后修饰位点预测\_(ZDHHC\_LIPID)/ZDHHC-LIPID-ECEL1-PTM-score.pdf)**



**Fig.** **16** GSE87624 Box Plot Of DEGs final

**(File path: Figure+Table/3.4.8\_MusiteDeep\_蛋白质转录后修饰位点预测\_(ZDHHC\_LIPID)/GSE87624-Box-Plot-Of-DEGs-final.pdf)**

## 3.5 ZDHHCs 在其他数据集中的表达 (Metastasis vs Primary)

### 3.5.1 GEO 数据获取 (GSE32981)

以 GEOquery 获取 GSE32981 的数据信息。

### 3.5.2 Limma 差异分析 (GSE32981)

筛选数据集中的基因，以ZDHHC1, ZDHHC2, ZDHHC16, …(n = 22)差异分析。样本分组：Met (n=11) , Prim (n=12) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Met vs Prim。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.5.3 GEO 数据获取 (GSE14827)

以 GEOquery 获取 GSE14827 的数据信息。

### 3.5.4 Limma 差异分析 (GSE14827)

筛选数据集中的基因，以“ZDHHC11 /// ZDHHC11B”, “ZDHHC15”, “ZDHHC19”, …(n = 58)差异分析。样本分组：Metastasis (n=9) , Primary (n=18) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Metastasis vs Primary。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.5.5 GEO 数据获取 (GSE18947)

以 GEOquery 获取 GSE18947 的数据信息。

### 3.5.6 Limma 差异分析 (GSE18947)

匹配 cell.line.ch1 中包含“Sosp-9607”的描述，最终得到 4 例数据。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：High\_metastatic\_potential vs Low\_metastatic\_potential。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.5.7 GEO 数据获取 (GSE21257)

以 GEOquery 获取 GSE21257 的数据信息。

### 3.5.8 Limma 差异分析 (GSE21257)

筛选数据集中的基因，以ZDHHC15, ZDHHC14, ZDHHC22, …(n = 29)差异分析。样本分组：Metastasis (n=34) , Primary (n=19) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Metastasis vs Primary。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.5.9 GEO 数据获取 (GSE9508)

以 GEOquery 获取 GSE9508 的数据信息。

### 3.5.10 Limma 差异分析 (GSE9508)

筛选数据集中的基因，以ZDHHC11, ZDHHC6, ZDHHC16, …(n = 23)差异分析。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Biopsy\_metastatic vs Biopsy\_non\_metastatic, Biopsy\_non\_metastatic vs Non\_malignant\_bone。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.5.11 GEO 数据获取 (GSE237033)

以 GEOquery 获取 GSE237033 的数据信息。

### 3.5.12 Limma 差异分析 (GSE237033)

筛选数据集中的基因，以ZDHHC18, ZDHHC3, ZDHHC23, …(n = 26)差异分析。以 edgeR 将GSE237033 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：metastasis vs primary。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.5.13 GEO 数据获取 (GSE234998)

以 GEOquery 获取 GSE234998 的数据信息。

### 3.5.14 Limma 差异分析 (GSE234998)

筛选数据集中的基因，以ZDHHC18, ZDHHC3, ZDHHC23, …(n = 26)差异分析。以 edgeR 将GSE234998 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：primary\_met vs localized。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.5.15 GEO 数据获取 (GSE14359)

以 GEOquery 获取 GSE14359 的数据信息。

### 3.5.16 Limma 差异分析 (GSE14359)

筛选数据集中的基因，以ZDHHC24, ZDHHC18, ZDHHC17, …(n = 13)差异分析。样本分组：Metastasis (n=8) , Others (n=2) , Primary (n=10) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Metastasis vs Primary。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.5.17 GEO 数据获取 (GSE220538)

以 GEOquery 获取 GSE220538 的数据信息。

### 3.5.18 Limma 差异分析 (GSE220538)

筛选数据集中的基因，以ZDHHC18, ZDHHC3, ZDHHC23, …(n = 26)差异分析。样本分组：metastatic\_tumor (n=10) , primary\_tumor (n=23) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：metastatic\_tumor vs primary\_tumor。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.5.19 Limma 差异分析 (TARGET\_OS)

(使用 TARGET-OS 数据集) 筛选数据集中的基因，以ZDHHC6, ZDHHC8, ZDHHC15, …(n = 31)差异分析。样本分组：Metastasis\_\_NOS (n=22) , No\_Metastasis (n=65) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Metastasis\_\_NOS vs No\_Metastasis。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

## 3.6 ZDHHCs 在其他数据集中的表达 (Primary vs Normal)

### 3.6.1 Limma 差异分析 (NORMAL\_GSE234998)

筛选数据集中的基因，以ZDHHC18, ZDHHC3, ZDHHC23, …(n = 26)差异分析。以 edgeR 将GSE234998 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：localized vs local\_control。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.6.2 GEO 数据获取 (NORMAL\_GSE218035)

以 GEOquery 获取 GSE218035 的数据信息。

### 3.6.3 Limma 差异分析 (NORMAL\_GSE218035)

筛选数据集中的基因，以ZDHHC6, ZDHHC8, ZDHHC15, …(n = 24)差异分析。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：Osteosarcoma\_sample vs Normal\_adjacent\_tissue。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

### 3.6.4 GEO 数据获取 (NORMAL\_GSE99671)

以 GEOquery 获取 GSE99671 的数据信息。

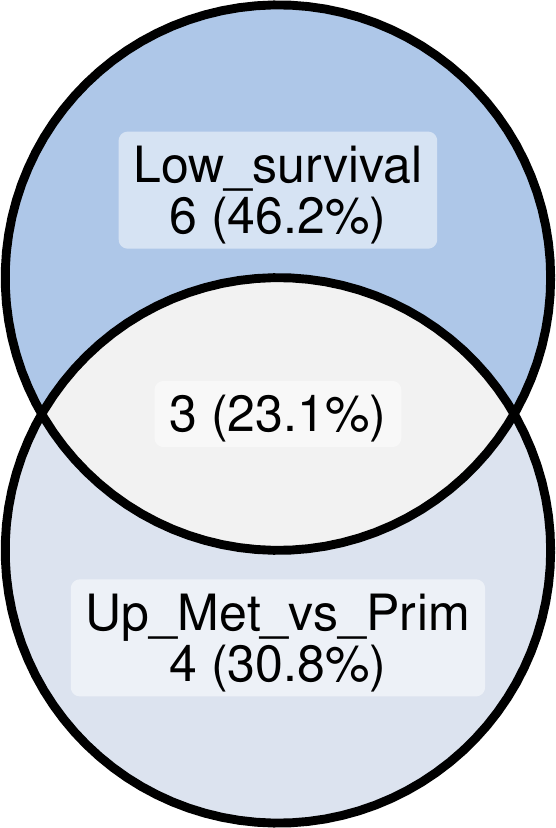
### 3.6.5 Limma 差异分析 (NORMAL\_GSE99671)

筛选数据集中的基因，以ZDHHC18, ZDHHC3, ZDHHC23, …(n = 24)差异分析。样本分组：NORMAL (n=18) , TUMOR (n=18) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：TUMOR vs NORMAL。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。

## 3.7 Collate 汇总分析结果 (ZDHHC)

筛选高表达差异基因 (Up\_Met\_vs\_Prim)，且生存分析为高表达预后不良。

Fig. **[17](#Intersection-of-Up-Met-vs-Prim-with-Low-survival)** 将Up\_Met\_vs\_Prim, Low\_survival 取交集。 Tab. **[7](#DEGs-of-Met-vs-Prim-in-datasets)** 为差异分析 DEGs 以及数据集来源汇总。



**Fig.** **17** Intersection of Up Met vs Prim with Low survival

**(File path: Figure+Table/3.7.0\_Collate\_汇总分析结果\_(ZDHHC)/Intersection-of-Up-Met-vs-Prim-with-Low-survival.pdf)**

* All\_intersection: ZDHHC11, ZDHHC11B, ZDHHC22
* Refer\_dataset\_Met\_vs\_Prim: GSE1435, GSE1482, GSE1894, GSE2125, GSE22053, GSE23499, GSE23703, GSE3298, GSE87624zdhhc, GSE950, target\_os
* Refer\_dataset\_survival: os

**(See: Figure+Table/3.7.0\_Collate\_汇总分析结果\_(ZDHHC)/Intersection-of-Up-Met-vs-Prim-with-Low-survival-content)**

**Tab.** **7** DEGs of Met vs Prim in datasets

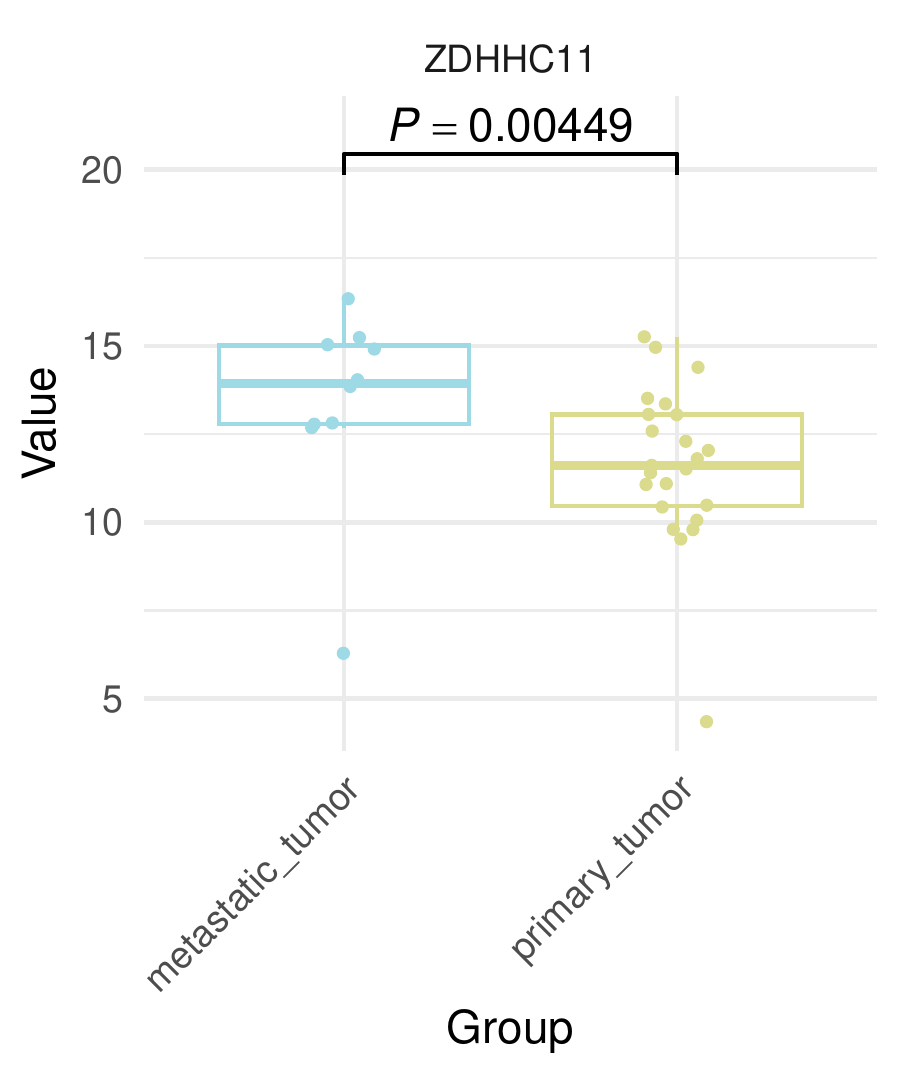
| Dataset | Symbol | LogFC | P.Value |
| --- | --- | --- | --- |
| GSE14359 | ZDHHC3 | 0.523 | 0.01438 |
| GSE14827 | ZDHHC2 | -0.3285 | 0.03034 |
| GSE18947 | ZDHHC11 | 2.006 | 0.0008553 |
| GSE18947 | ZDHHC2 | -0.9204 | 0.003782 |
| GSE18947 | ZDHHC9 | -1.125 | 0.005914 |
| ... | ... | ... | ... |

**(File path: Figure+Table/3.7.0\_Collate\_汇总分析结果\_(ZDHHC)/DEGs-of-Met-vs-Prim-in-datasets.csv)**

## 3.8 ZDHHC11 (重新分析)

### 3.8.1 Metastasis vs Primary (GSE220538)

Fig. **[18](#GSE220538-Box-Plot-Of-DEGs-zdhhc11)** 基因 ZDHHC11 表达水平，以及对应的 limma 差异分析显著水平。 Tab. **[8](#GSE220538-Statistic-of-Focused-genes-zdhhc11)** 为聚焦分析的基因的统计附表。



**Fig.** **18** GSE220538 Box Plot Of DEGs zdhhc11

**(File path: Figure+Table/3.8.1\_Metastasis\_vs\_Primary\_(GSE220538)/GSE220538-Box-Plot-Of-DEGs-zdhhc11.pdf)**

**Tab.** **8** GSE220538 Statistic of Focused genes zdhhc11

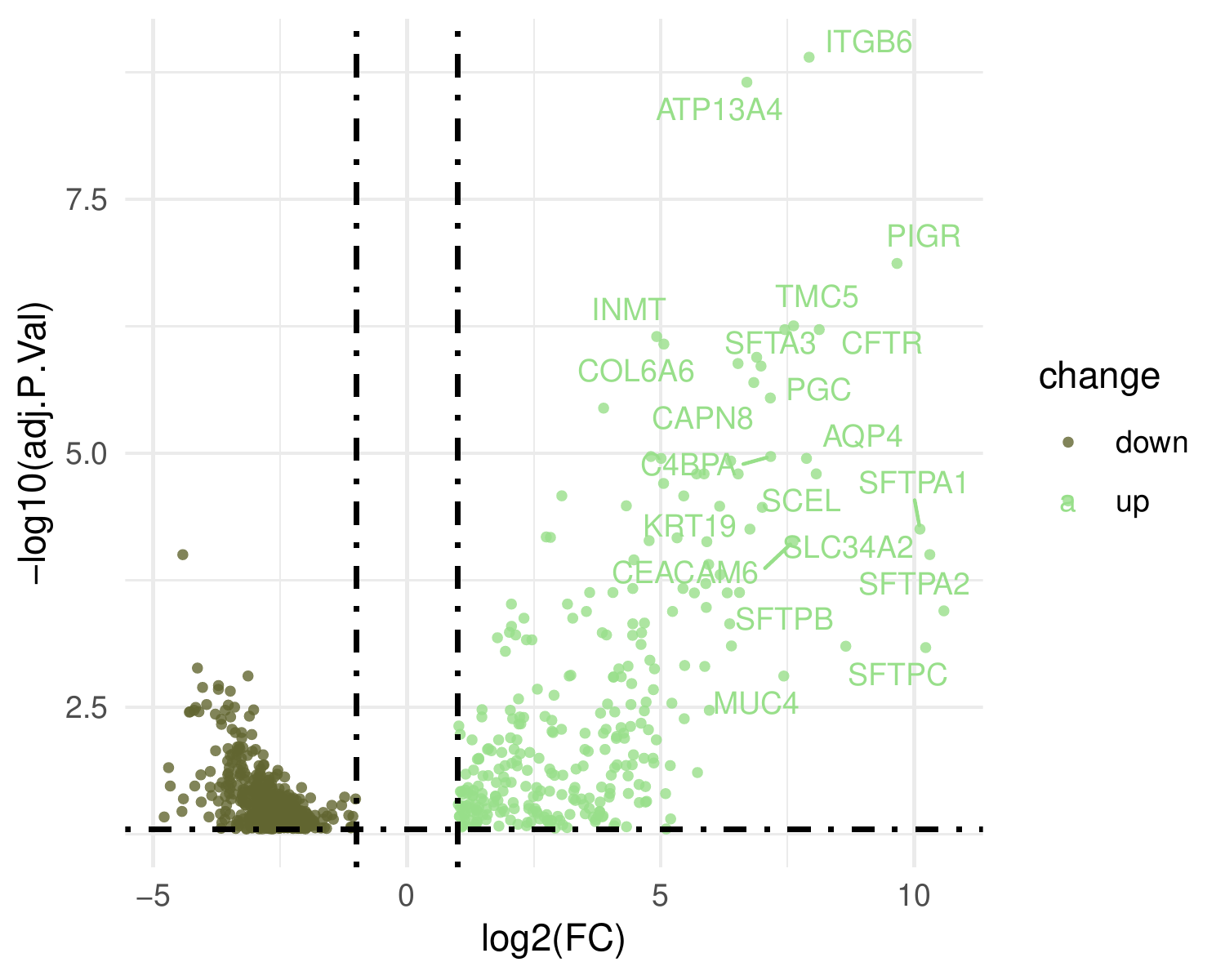
| Hgnc symbol | LogFC | Adj.P.Val | P.Value | Rownames |
| --- | --- | --- | --- | --- |
| ZDHHC11 | 2.171 | 0.03296 | 0.004495 | 79844 |

**(File path: Figure+Table/3.8.1\_Metastasis\_vs\_Primary\_(GSE220538)/GSE220538-Statistic-of-Focused-genes-zdhhc11.csv)**

### 3.8.2 Limma 差异分析 (META\_GSE220538)

样本分组：metastatic\_tumor (n=10) , primary\_tumor (n=23) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：metastatic\_tumor vs primary\_tumor。(若 A vs B，则为前者比后者，LogFC 大于 0 时，A 表达量高于 B)。上调或下调 DEGs 统计：up (n=300) , down (n=509)

Fig. **[19](#META-GSE220538-metastatic-tumor-vs-primary-tumor)** 为 metastatic\_tumor - primary\_tumor 差异分析火山图。 Tab. **[9](#META-GSE220538-data-metastatic-tumor-vs-primary-tumor)** 为 metastatic\_tumor - primary\_tumor 差异分析统计表格。



**Fig.** **19** META GSE220538 metastatic tumor vs primary tumor

**(File path: Figure+Table/3.8.2\_Limma\_差异分析\_(META\_GSE220538)/META-GSE220538-metastatic-tumor-vs-primary-tumor.pdf)**

* adj.P.Val cut-off: 0.05
* Log2(FC) cut-off: 1

**(See: Figure+Table/3.8.2\_Limma\_差异分析\_(META\_GSE220538)/META-GSE220538-metastatic-tumor-vs-primary-tumor-content)**

**Tab.** **9** META GSE220538 data metastatic tumor vs primary tumor

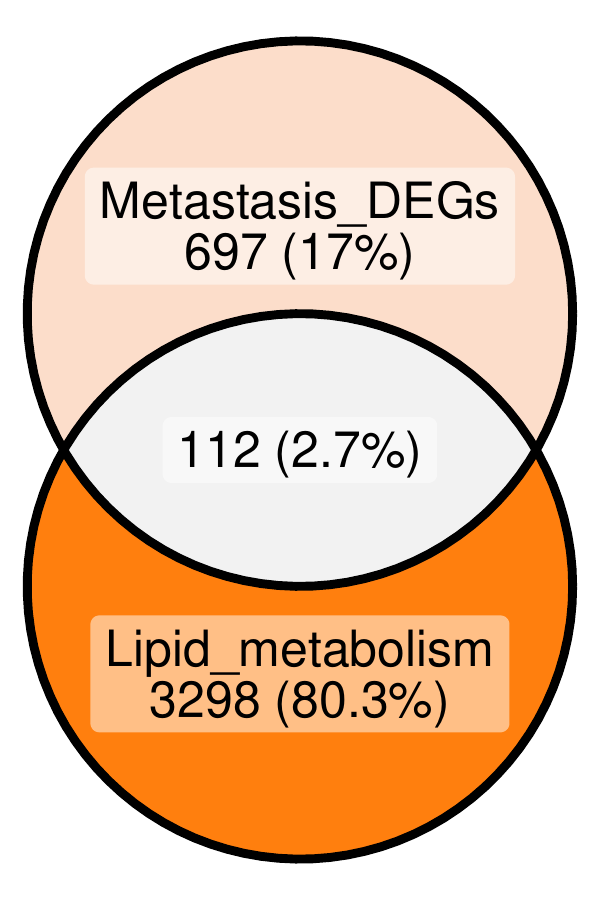
| Hgnc symbol | LogFC | Adj.P.Val | Rownames | GeneID |
| --- | --- | --- | --- | --- |
| ITGB6 | 7.924 | 1.267e-09 | 3694 | 3694 |
| ATP13A4 | 6.698 | 2.233e-09 | 84239 | 84240 |
| PIGR | 9.657 | 1.356e-07 | 5284 | 5284 |
| TMC5 | 7.618 | 5.571e-07 | 79838 | 79840 |
| CFTR | 8.124 | 6.071e-07 | 1080 | 1080 |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.8.2\_Limma\_差异分析\_(META\_GSE220538)/META-GSE220538-data-metastatic-tumor-vs-primary-tumor.tsv)**

### 3.8.3 交集: Lipid\_metabolism + Metastasis\_DEGs (LIPID\_METAGSE220538)

以下取交集： - 基因集 (来自于GeneCards 基因获取[Section: LIPID]) - 基因集 (metastatic\_tumor - primary\_tumor, 来自于Limma 差异分析[Section: META\_GSE220538])

Fig. **[20](#LIPID-METAGSE220538-Intersection-of-Lipid-metabolism-with-Metastasis-DEGs)** 将Lipid\_metabolism, Metastasis\_DEGs 取交集。



**Fig.** **20** LIPID METAGSE220538 Intersection of Lipid metabolism with Metastasis DEGs

**(File path: Figure+Table/3.8.3\_交集:\_Lipid\_metabolism\_+\_Metastasis\_DEGs\_(LIPID\_METAGSE220538)/LIPID-METAGSE220538-Intersection-of-Lipid-metabolism-with-Metastasis-DEGs.pdf)**

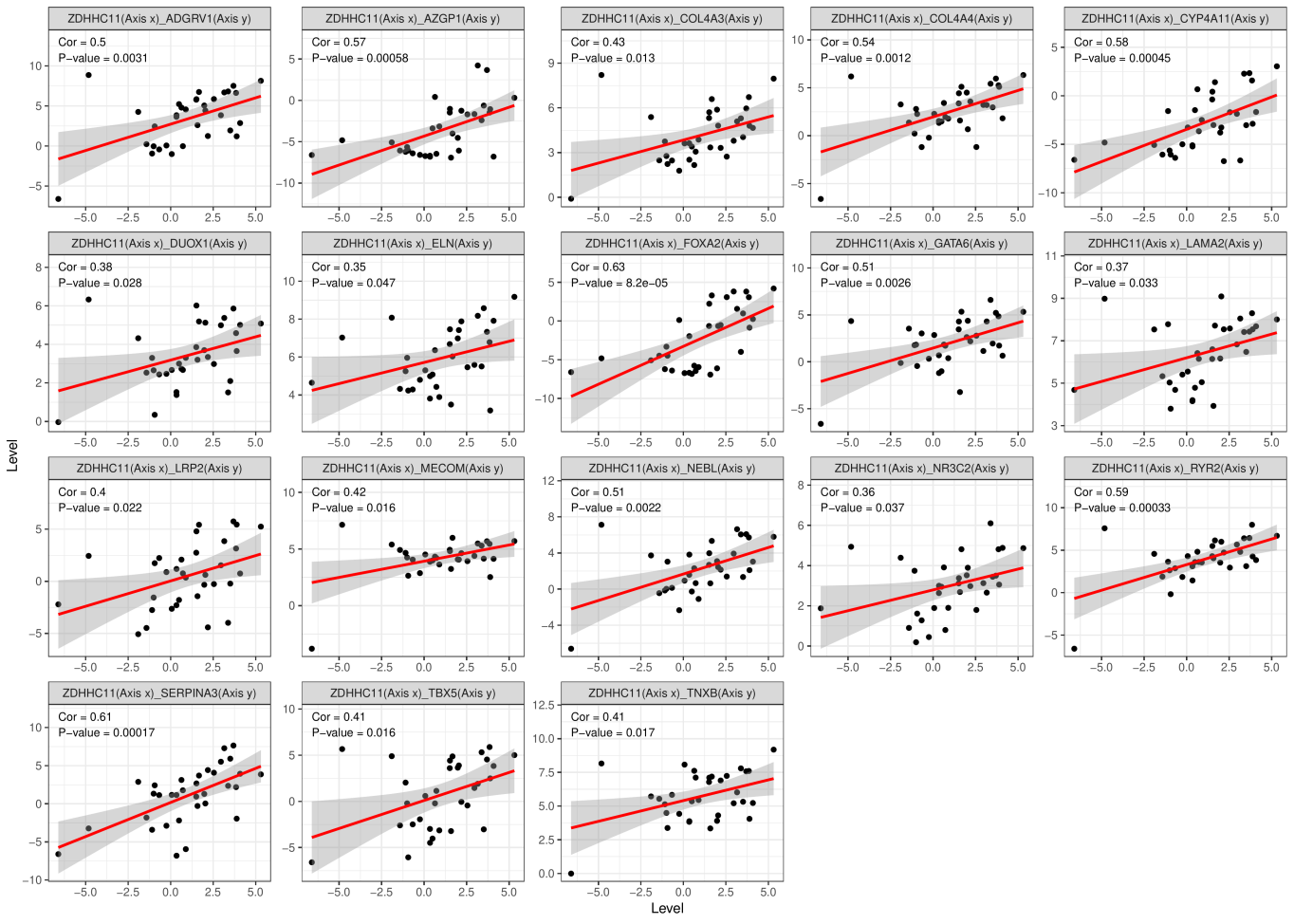
* All\_intersection: ACADS, ACADL, NR1H3, NPPA, SLC27A4, ACE, AGT, C3, ELN, CAV1, TNXB, FNDC5, AGER, SLC2A5, GPX8, NR3C2, DDIT3, EDN1, VWF, CAV2, SIRT6, ACOT4, CYP4A11, SLC5A1, LRP2, AMH, CFTR, DUOX1, AMY1B, ABCA3, MYH14, RPGR, DOLK, ABCA8, FDX2, NAA40, MMACHC, MYH11, AZGP1, CAPN3, PCDH15, RYR2, NOS1AP, LAMP3, SERPINA3, PLA2G4F, CYP3A5, MALAT1, EDNRB, LAMC2, ACOXL, GATA6, LAMA3, TBX5, ADGRV1, NEBL, PDE3B, KLF5, FHL1, COL4A3, LAMA2, COL4A4, TSFM, FREM2, ABHD14A, RARB, ENO3, SLC25A10, CYP4B1, MIR10B, TIMM22, UBL4A,…

**(See: Figure+Table/3.8.3\_交集:\_Lipid\_metabolism\_+\_Metastasis\_DEGs\_(LIPID\_METAGSE220538)/LIPID-METAGSE220538-Intersection-of-Lipid-metabolism-with-Metastasis-DEGs-content)**

### 3.8.4 关联分析 (META\_GSE220538)

将基因 (ZDHHC11 -> 基因集 (ACADS, ACADL, NR1H3, …[n = 112], 来自于Venn 交集[Section: LIPID\_METAGSE220538]) ) 关联分析。共得到 18 个显著的基因对 (P < 0.05, |Cor| > 0.3)。

Fig. **[21](#META-GSE220538-significant-correlation-plots)** 为显著关联的基因的线型回归图。 Tab. **[10](#META-GSE220538-significant-correlation-analysis-data)** 为关联分析统计附表 (P-value cutoff: 0.05, Cor (关联系数) cutoff: 0.3)。



**Fig.** **21** META GSE220538 significant correlation plots

**(File path: Figure+Table/3.8.4\_关联分析\_(META\_GSE220538)/META-GSE220538-significant-correlation-plots.pdf)**

**Tab.** **10** META GSE220538 significant correlation analysis data

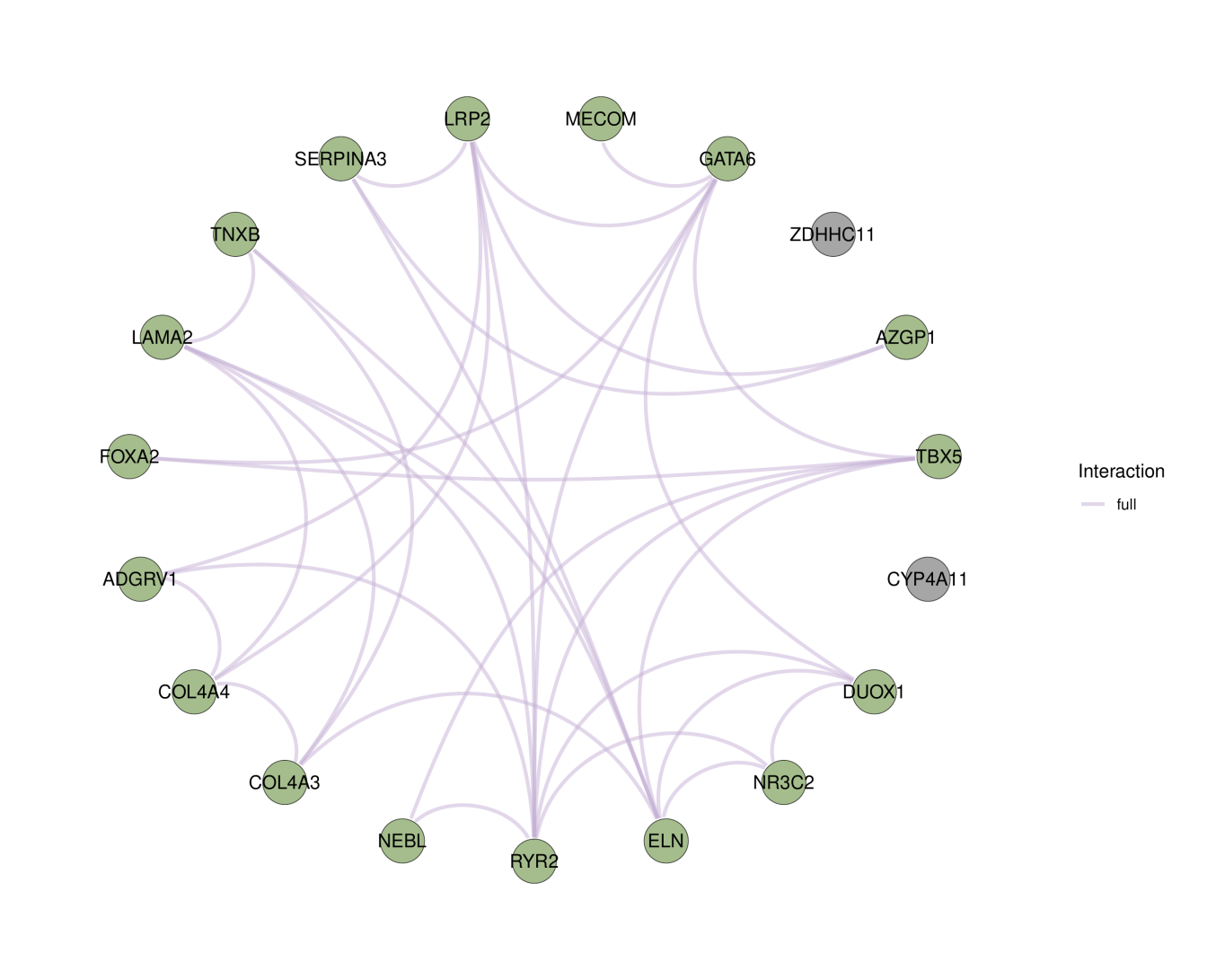
| From | To | Cor | Pvalue | Model |
| --- | --- | --- | --- | --- |
| ZDHHC11 | FOXA2 | 0.6311 | 8.231e-05 | C(-4.821775089577... |
| ZDHHC11 | SERPINA3 | 0.6098 | 0.0001652 | C(-3.236812588856... |
| ZDHHC11 | RYR2 | 0.5867 | 0.0003324 | C(7.5777698770498... |
| ZDHHC11 | CYP4A11 | 0.5765 | 0.0004455 | C(-4.821775089577... |
| ZDHHC11 | AZGP1 | 0.567 | 0.0005807 | C(-4.821775089577... |
| ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.8.4\_关联分析\_(META\_GSE220538)/META-GSE220538-significant-correlation-analysis-data.xlsx)**

### 3.8.5 STRINGdb PPI 分析 (LIPID\_ZDHHC11)

对基因集 (from, to, 来自于关联分析[Section: META\_GSE220538]) 进行STRINGdb PPI 分析。

Fig. **[22](#LIPID-ZDHHC11-Top-MCC-score)** PPI 网络图



**Fig.** **22** LIPID ZDHHC11 Top MCC score

**(File path: Figure+Table/3.8.5\_STRINGdb\_PPI\_分析\_(LIPID\_ZDHHC11)/LIPID-ZDHHC11-Top-MCC-score.pdf)**

### 3.8.6 ClusPro 蛋白质-蛋白质对接预测 (ZDHHC11\_LIPID)

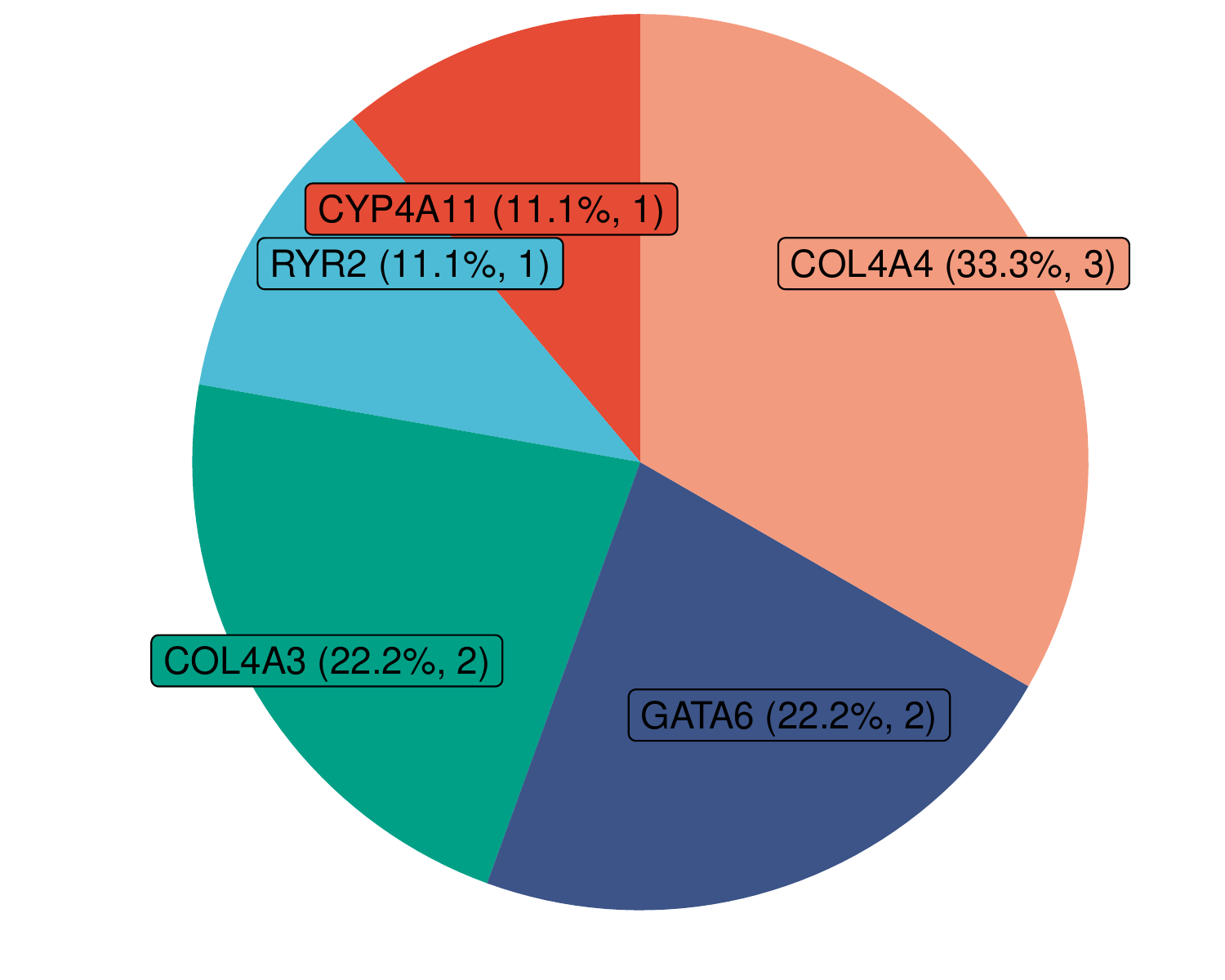
对基因集 (from, to, 来自于关联分析[Section: META\_GSE220538]) 进行ClusPro 蛋白质-蛋白质对接预测。以 biomaRt 获取基因 Symbol 对应的蛋白结构 (PDB，详见方法章节)。选取分辨率最高 (即，resolution 值最小) 的 PDB 作为分子对接的蛋白结构。从 RCSB PDB 获取 PDB 文件。对于未从 PDB 数据库找到结构文件的，从数据库 AlphaFold 获取 ZDHHC11, CYP4A11, GATA6, …(n = 4) 预测的蛋白结构 (根据 UniProtKB-Swiss-Prot ID，详见方法章节)。将 PDB 上传至 ClusPro 进行对接。

### 3.8.7 MusiteDeep 蛋白质转录后修饰位点预测 (ZDHHC11\_LIPID)

对基因集 (FOXA2, SERPINA3, RYR2, …[n = 10], 来自于关联分析[Section: META\_GSE220538]) 进行MusiteDeep 蛋白质转录后修饰位点预测。以 biomaRt 获取蛋白质 (FOXA2, SERPINA3, RYR2, …(n = 10)) 的序列 (Peptide)。以 MusiteDeep 预测 S-palmitoyl\_cysteine 修饰位点。

Fig. **[23](#ZDHHC11-LIPID-S-palmitoyl-cysteine-PTM-numbers)** 为预测到的 PTMs 数量饼图。

Fig. **[24](#ZDHHC11-LIPID-RYR2-PTM-score)** 为 RYR2 的修饰位点以及得分可视化图。 Fig. **[25](#ZDHHC11-LIPID-CYP4A11-PTM-score)** 为 CYP4A11 的修饰位点以及得分可视化图。

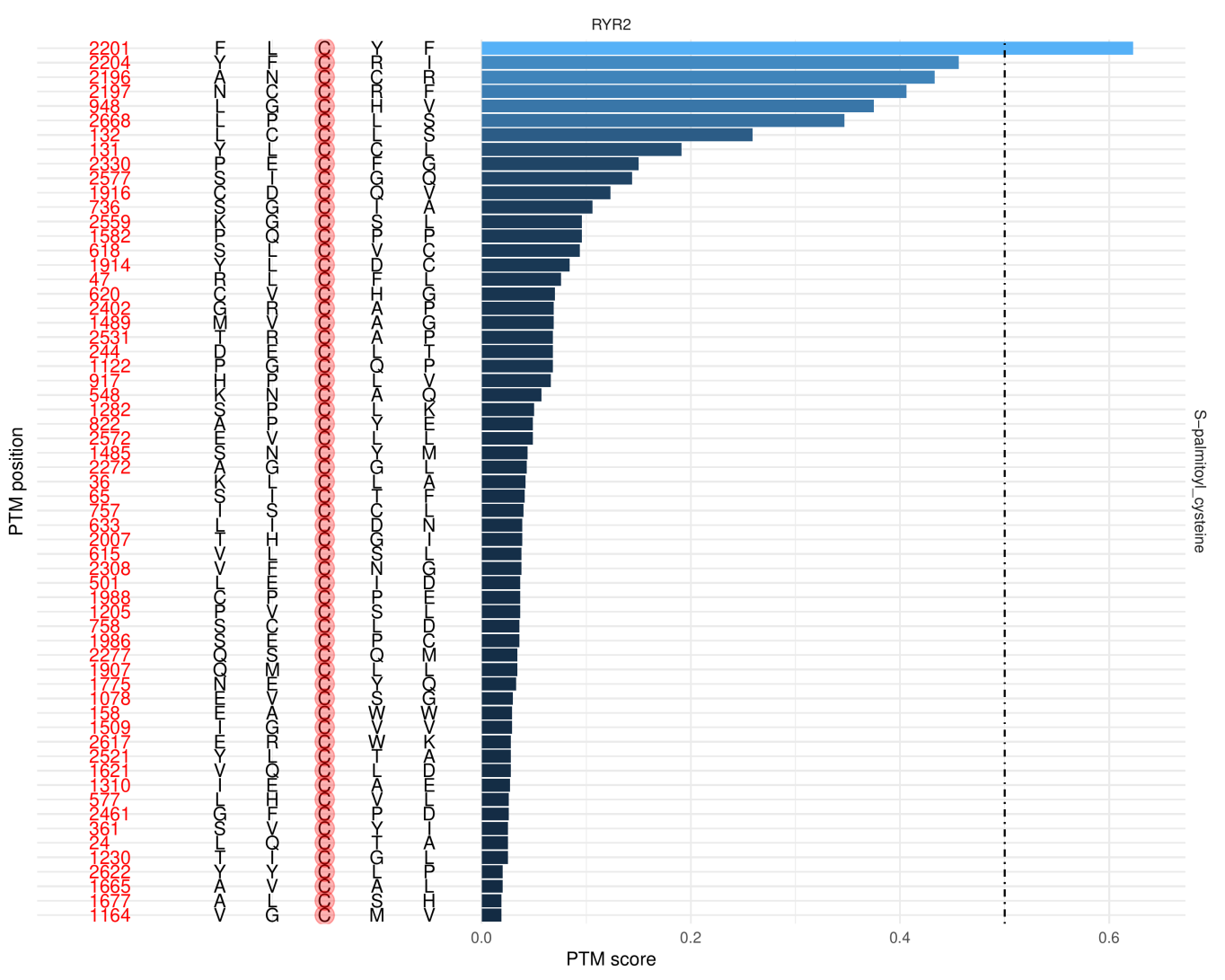


**Fig.** **23** ZDHHC11 LIPID S palmitoyl cysteine PTM numbers

**(File path: Figure+Table/3.8.7\_MusiteDeep\_蛋白质转录后修饰位点预测\_(ZDHHC11\_LIPID)/ZDHHC11-LIPID-S-palmitoyl-cysteine-PTM-numbers.pdf)**

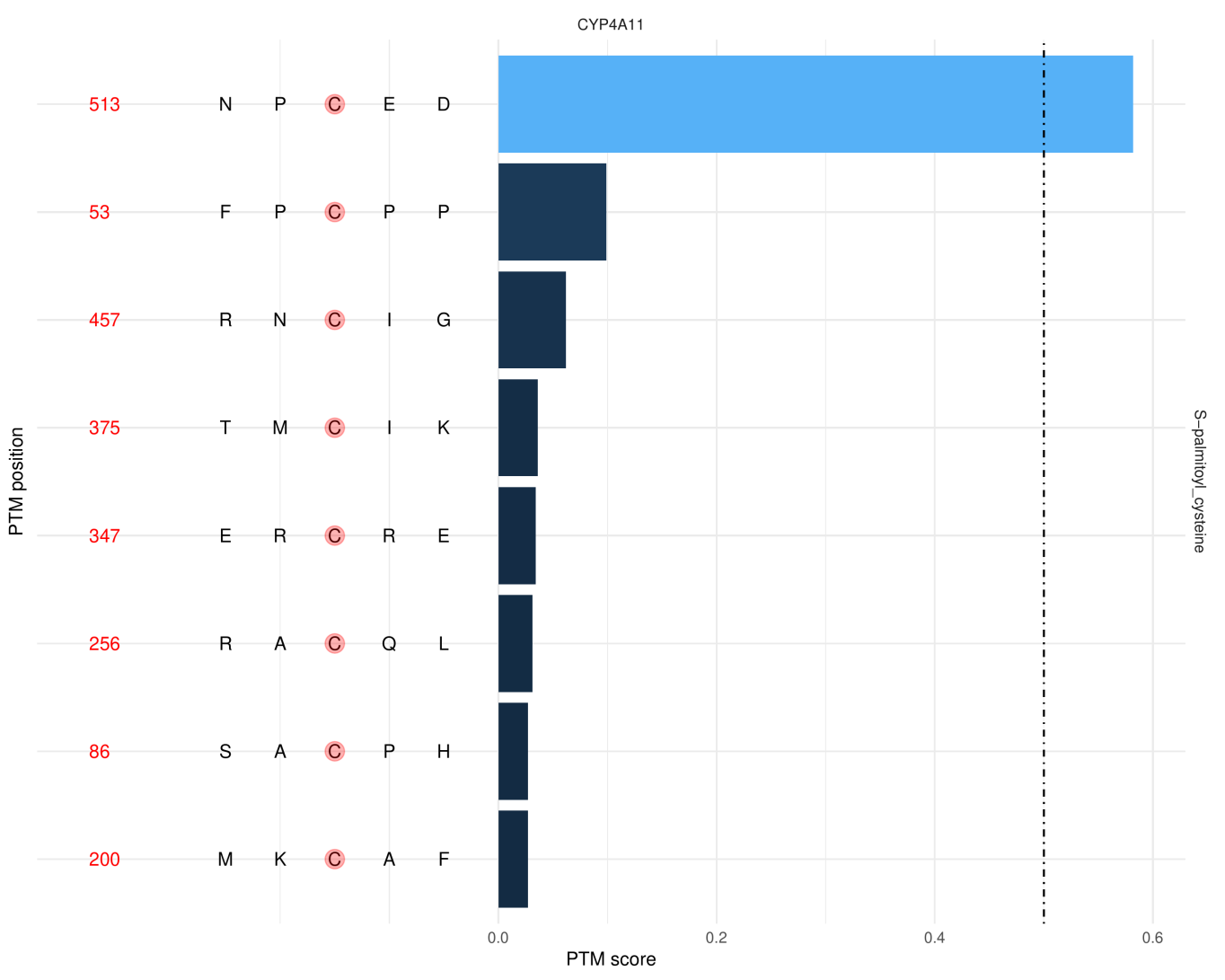
Note: The directory 'Figure+Table/ZDHHC11-LIPID-All-PTM-score' contains 10 files.  
  
1 1\_ADGRV1.pdf  
2 10\_SERPINA3.pdf  
3 2\_AZGP1.pdf  
4 3\_COL4A3.pdf  
5 4\_COL4A4.pdf  
6 ...

**(File path: Figure+Table/3.8.7\_MusiteDeep\_蛋白质转录后修饰位点预测\_(ZDHHC11\_LIPID)/ZDHHC11-LIPID-All-PTM-score)**



**Fig.** **24** ZDHHC11 LIPID RYR2 PTM score

**(File path: Figure+Table/3.8.7\_MusiteDeep\_蛋白质转录后修饰位点预测\_(ZDHHC11\_LIPID)/ZDHHC11-LIPID-RYR2-PTM-score.pdf)**



**Fig.** **25** ZDHHC11 LIPID CYP4A11 PTM score

**(File path: Figure+Table/3.8.7\_MusiteDeep\_蛋白质转录后修饰位点预测\_(ZDHHC11\_LIPID)/ZDHHC11-LIPID-CYP4A11-PTM-score.pdf)**

# 4 总结

## 4.1 ZDHHC11

* ZDHHC11 为 Metastases vs Primary (Fig. **[18](#GSE220538-Box-Plot-Of-DEGs-zdhhc11)** ) (根据多个数据集筛选，见 Tab. **[7](#DEGs-of-Met-vs-Prim-in-datasets)** ) 且生存分析显著 (Fig. **[1](#OS-survival-curve-of-ZDHHC11)** )。
* 与脂质代谢合成相关性的基因集取自 GeneCards (原分析要求的 WGCNA 无法得到这方面的结果)。
* ZDHHC-X与相互作用蛋白在OS转移中的相关性，见Fig. **[21](#META-GSE220538-significant-correlation-plots)**
* 先进行了关联分析，然后以 STRINGdb 寻找蛋白互作 (此前筛选的 ZDHHC11 与上述脂质蛋白)，未发现 ZDHHC11 与任何蛋白互作，见 Fig. **[22](#LIPID-ZDHHC11-Top-MCC-score)** 。
* 由于未从 STRINGdb 数据库发现互作蛋白，所以，以 cluspro 预测蛋白之间的互作，见 …
* 以 MusiteDeep 预测互作蛋白是否存在 S-palmitoyl\_cysteine 修饰，见Fig. **[23](#ZDHHC11-LIPID-S-palmitoyl-cysteine-PTM-numbers)**
* 综上，结合蛋白对接，关联分析，以及 S-palmitoyl\_cysteine 位点预测，与 ZDHHC11 结合的最佳蛋白可能为 …

# Reference

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

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

3. Stelzer, G. *et al.* The genecards suite: From gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics* **54**, 1.30.1–1.30.33 (2016).

4. 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).

5. 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).

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

7. 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).

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

9. 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).