**生信分析报告**

**项目标题： 高级别浆液性卵巢癌 ;**

**单 号： BSXN240936 ;**

**分析人员： 黄礼闯 ;**

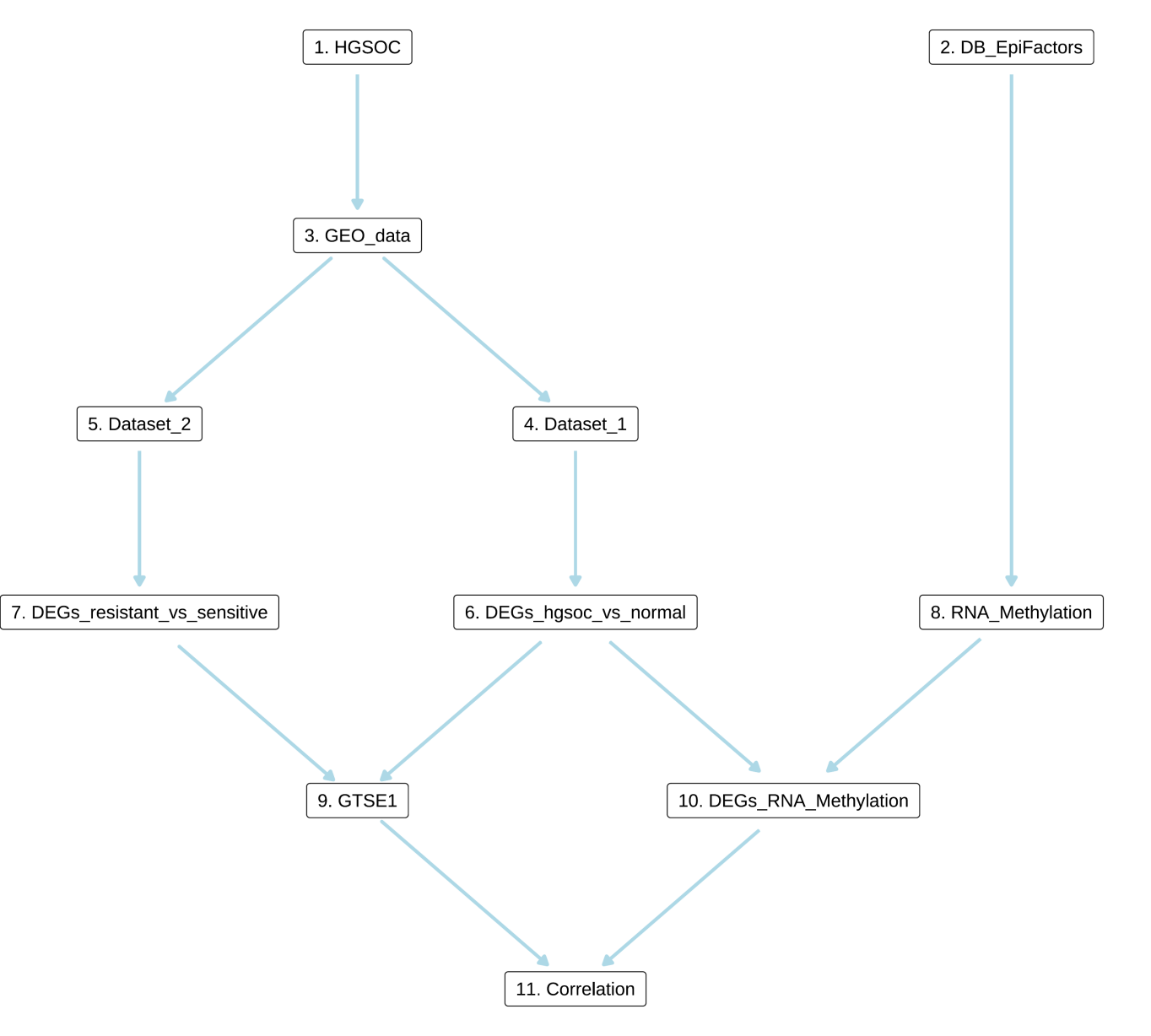
**分析类型： 生信协助 ;**

**委 托 人： 李扬 ;**

**受 托 人： 杭州铂赛生物科技有限公司 .**

# 1 分析流程

* GTSE1 在 HGSOC 中高表达促进肿瘤增值与迁移。 GTSE1 与化疗耐药有关。在卵巢癌 (HGSOC) 中的表达，在化疗耐药卵巢癌数据集中的表达。
* GTSE1 的 mRNA 序列上的 m6A 甲基化，GTSE1可能主要被ALKBH5、IGF2BP1、METTL14、METTL3、YTHDF1、YTHDF2。 GTSE1与甲基化相关酶的表达进行相关性分析



**Fig.** **1** Route

**(File path: Figure+Table/1.0\_分析流程\_{#abstract}/Route.pdf)**

# 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 GSE 数据搜索 (Dataset: HGSOC)

使用 Entrez Direct (EDirect) <https://www.ncbi.nlm.nih.gov/books/NBK3837/> 搜索 GEO 数据库 (esearch -db gds)，查询信息为: ((high-grade serous ovarian cancer[Description] AND (normal[Description]) AND ((6:1000[Number of Samples]) AND (GSE[Entry Type]) AND (Homo Sapiens[Organism]))。 以正则匹配，滤除包含 ‘single cell’ 或 ‘scRNA’ 的数据例。仅查询临床样本信息，因此滤除匹配到 ‘cells’, ‘cell type’ 或 ‘cell line’ 的实验数据例。 此外，去除了以特定 Marker 细胞类型为研究对象的数据例 (CD4、CD8 T 细胞等)。仅获取类型包含 ‘Expression profiling by high throughput sequencing’ 或 ‘Expression profiling by array’ 的数据例。

## 2.3 GEO 数据获取 (Dataset: HGSOC)

以 R 包 GEOquery (2.74.0) 获取 GSE146553 数据集。

## 2.4 Limma 差异分析 (Dataset: HGSOC)

使用 log2 和 limma::normalizeBetweenArrays 对数据标准化。 以 limma (3.62.1) (2005)1 差异分析。创建设计矩阵，对比矩阵，差异分析：hgsoc 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 0.5 的统计结果。

## 2.5 GSE 数据搜索

使用 Entrez Direct (EDirect) <https://www.ncbi.nlm.nih.gov/books/NBK3837/> 搜索 GEO 数据库 (esearch -db gds)，查询信息为: ((high-grade serous ovarian cancer[Description] AND (Drug resistance[Description]) AND ((6:1000[Number of Samples]) AND (GSE[Entry Type]) AND (Homo Sapiens[Organism]))。

## 2.6 GEO 数据获取 (Dataset: HGSOC\_RES)

以 R 包 GEOquery (2.74.0) 获取 GSE141630 数据集。

## 2.7 Limma 差异分析 (Dataset: HGSOC\_RES)

以 limma (3.62.1) (2005)1 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\_resistant vs sensitive。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 EpiFactors 表观遗传调控因子数据获取 (Dataset: M6A)

从数据库 EpiFactors (2023, **IF:16.6**, Q1, Nucleic acids research)2 获取表观遗传调控蛋白的数据。

# 3 分析结果

## 3.1 GSE 数据搜索 (HGSOC)

以 Entrez Direct (EDirect) 搜索 GEO 数据库 (检索条件见方法章节) 。

**Tab.** **1** High grade serous ovarian cancer EDirect query

| GSE | Title | Summary | Taxon | GdsType | N samples | PubMedIds |
| --- | --- | --- | --- | --- | --- | --- |
| GSE209964 | Gene count... | Transcript... | Homo sapiens | Expression... | 128 | PRJNA863245 |
| GSE181955 | Gene expre... | We report ... | Homo sapiens | Expression... | 8 | PRJNA754050 |
| GSE146556 | Global DNA... | This Super... | Homo sapiens | Methylatio... | 139 | PRJNA610846 |
| GSE146553 | Global DNA... | Comparison... | Homo sapiens | Expression... | 55 | PRJNA610853 |
| GSE135886 | LncRNA and... | Ovarian ca... | Homo sapiens | Expression... | 18 | PRJNA560410 |
| ... | ... | ... | ... | ... | ... | ... |

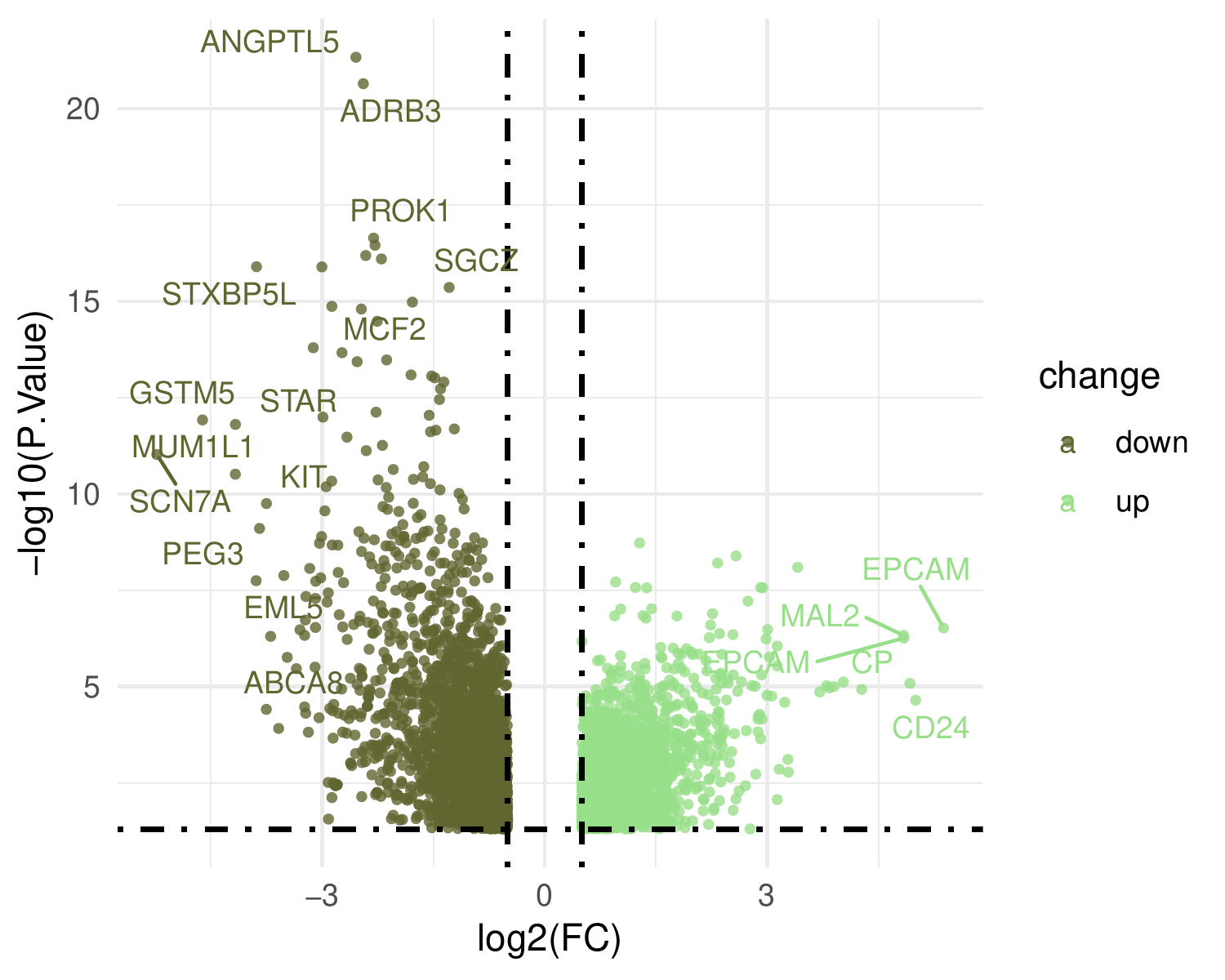
**(File path: Figure+Table/3.1\_GSE\_数据搜索\_(HGSOC)/High-grade-serous-ovarian-cancer-EDirect-query.xlsx)**

## 3.2 GEO 数据获取 (HGSOC)

## 3.3 Limma 差异分析 (HGSOC)

筛选 group 为 “normal”, “hgsoc”，最终得到 43 例数据。样本分组：hgsoc (n=40) , normal (n=3) , Others (n=0) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：hgsoc 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)。上调或下调 DEGs 统计：up (n=1992) , down (n=2347)

Fig. **[3](#HGSOC-Box-Plot-Of-DEGs)** 基因 GTSE1 表达水平，以及对应的 limma 差异分析显著水平。

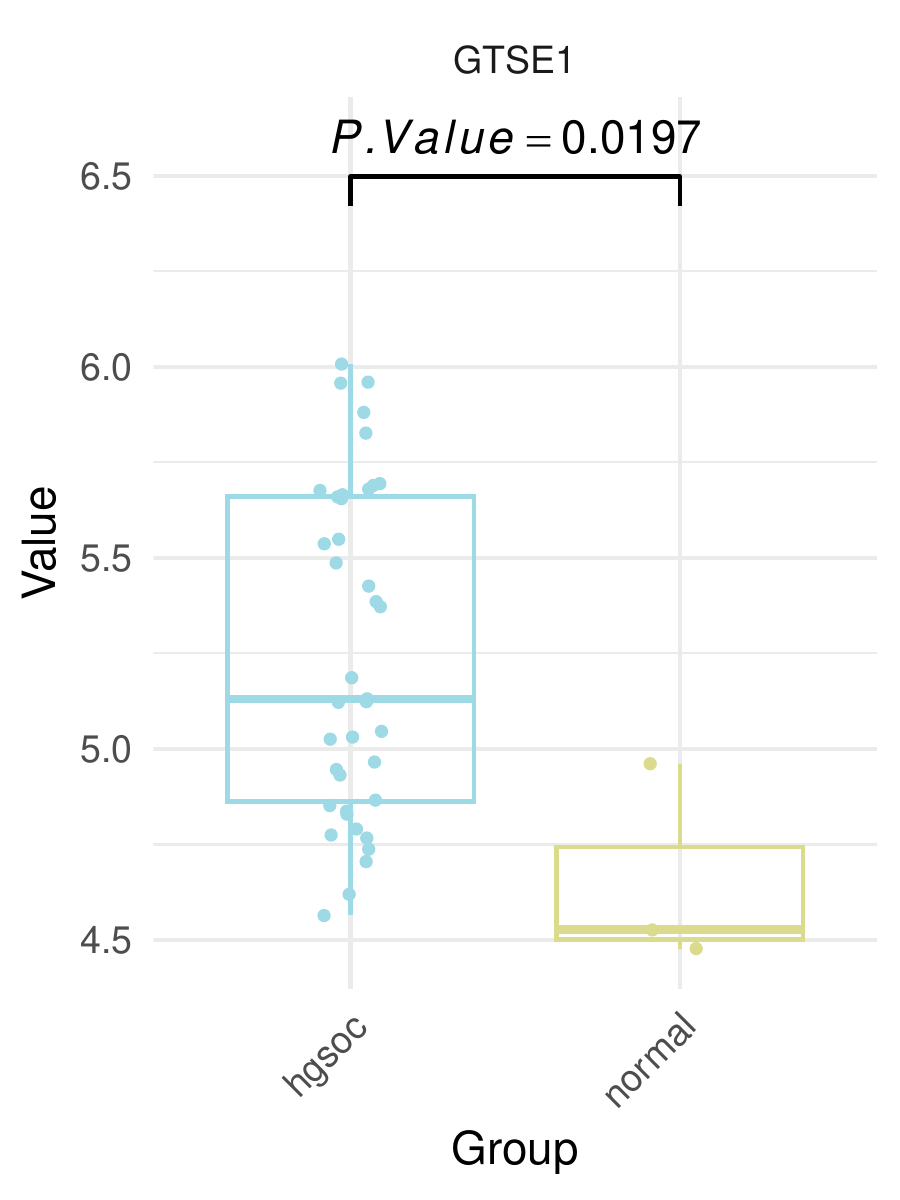


**Fig.** **2** HGSOC hgsoc vs normal

**(File path: Figure+Table/3.3\_Limma\_差异分析\_(HGSOC)/HGSOC-hgsoc-vs-normal.pdf)**

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

**(See: Figure+Table/3.3\_Limma\_差异分析\_(HGSOC)/HGSOC-hgsoc-vs-normal-content)**



**Fig.** **3** HGSOC Box Plot Of DEGs

**(File path: Figure+Table/3.3\_Limma\_差异分析\_(HGSOC)/HGSOC-Box-Plot-Of-DEGs.pdf)**

## 3.4 GSE 数据搜索 (HGSOC\_RES)

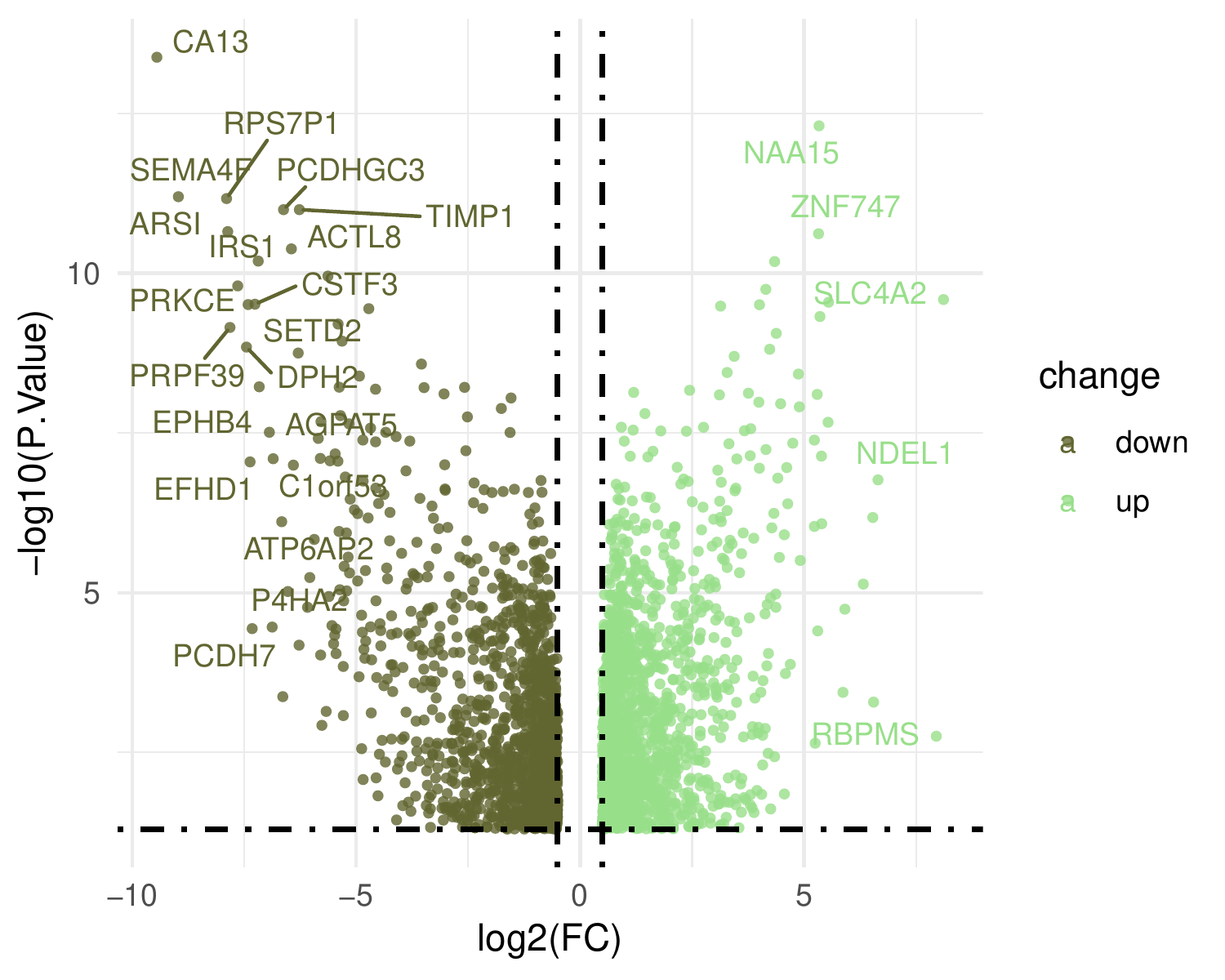
## 3.5 GEO 数据获取 (HGSOC\_RES)

以 GEOquery 获取 GSE141630 的数据信息。

## 3.6 Limma 差异分析 (HGSOC\_RES)

样本分组：carboplatin\_resistant (n=9) , cisplatin\_resistant (n=3) , sensitive (n=3) 。以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 。差异分析：carboplatin\_resistant vs sensitive。(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=1608) , down (n=1331) (注：cisplatin\_resistant 未显著，所以未分析)

Fig. **[5](#HGSOC-RES-Box-Plot-Of-DEGs)** 基因 GTSE1 表达水平，以及对应的 limma 差异分析显著水平。

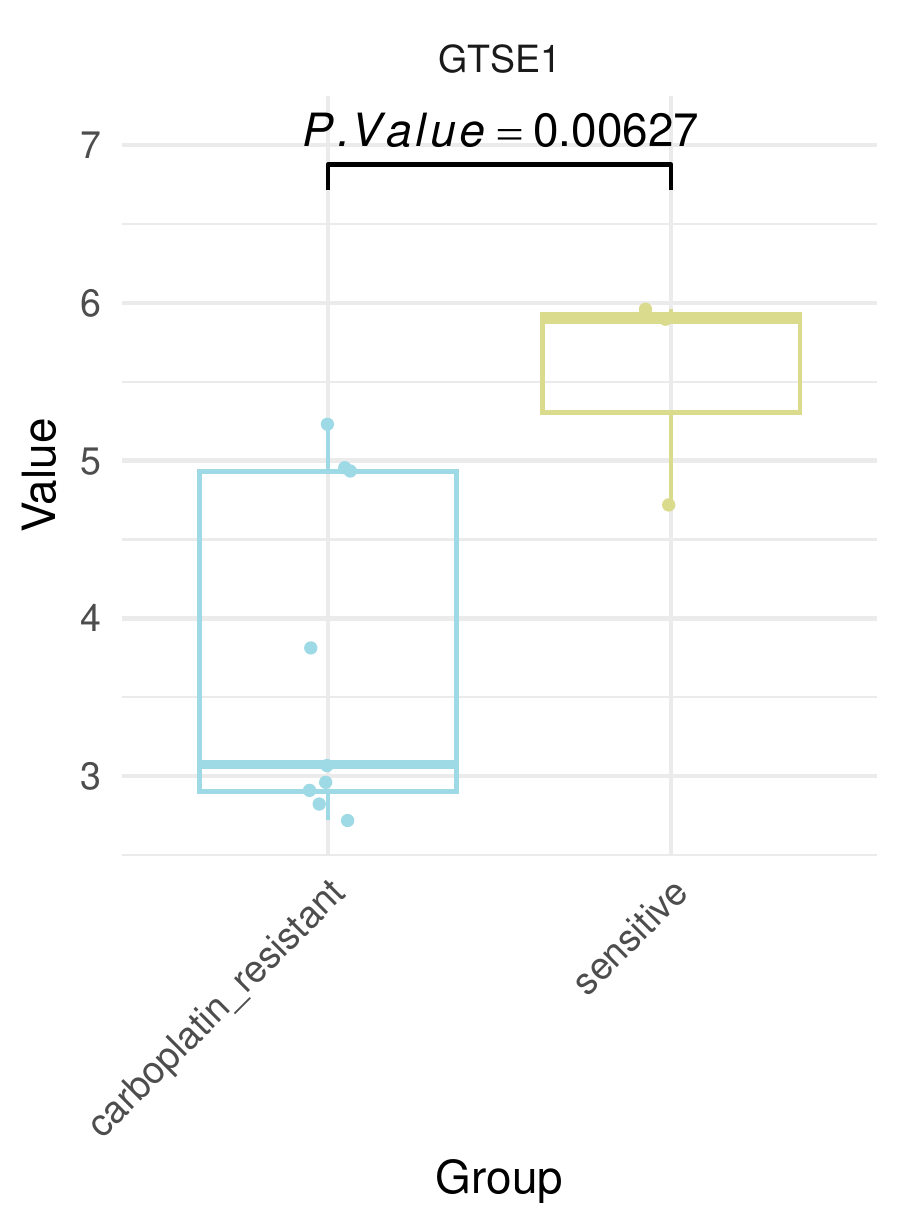


**Fig.** **4** HGSOC RES carboplatin resistant vs sensitive

**(File path: Figure+Table/3.6\_Limma\_差异分析\_(HGSOC\_RES)/HGSOC-RES-carboplatin-resistant-vs-sensitive.pdf)**

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

**(See: Figure+Table/3.6\_Limma\_差异分析\_(HGSOC\_RES)/HGSOC-RES-carboplatin-resistant-vs-sensitive-content)**



**Fig.** **5** HGSOC RES Box Plot Of DEGs

**(File path: Figure+Table/3.6\_Limma\_差异分析\_(HGSOC\_RES)/HGSOC-RES-Box-Plot-Of-DEGs.pdf)**

## 3.7 EpiFactors 表观遗传调控因子数据获取 (M6A)

获取 EpiFactors 中的表观遗传调控蛋白。匹配 Modification 中包含“RNA methylation”的描述，最终得到 12 例数据。

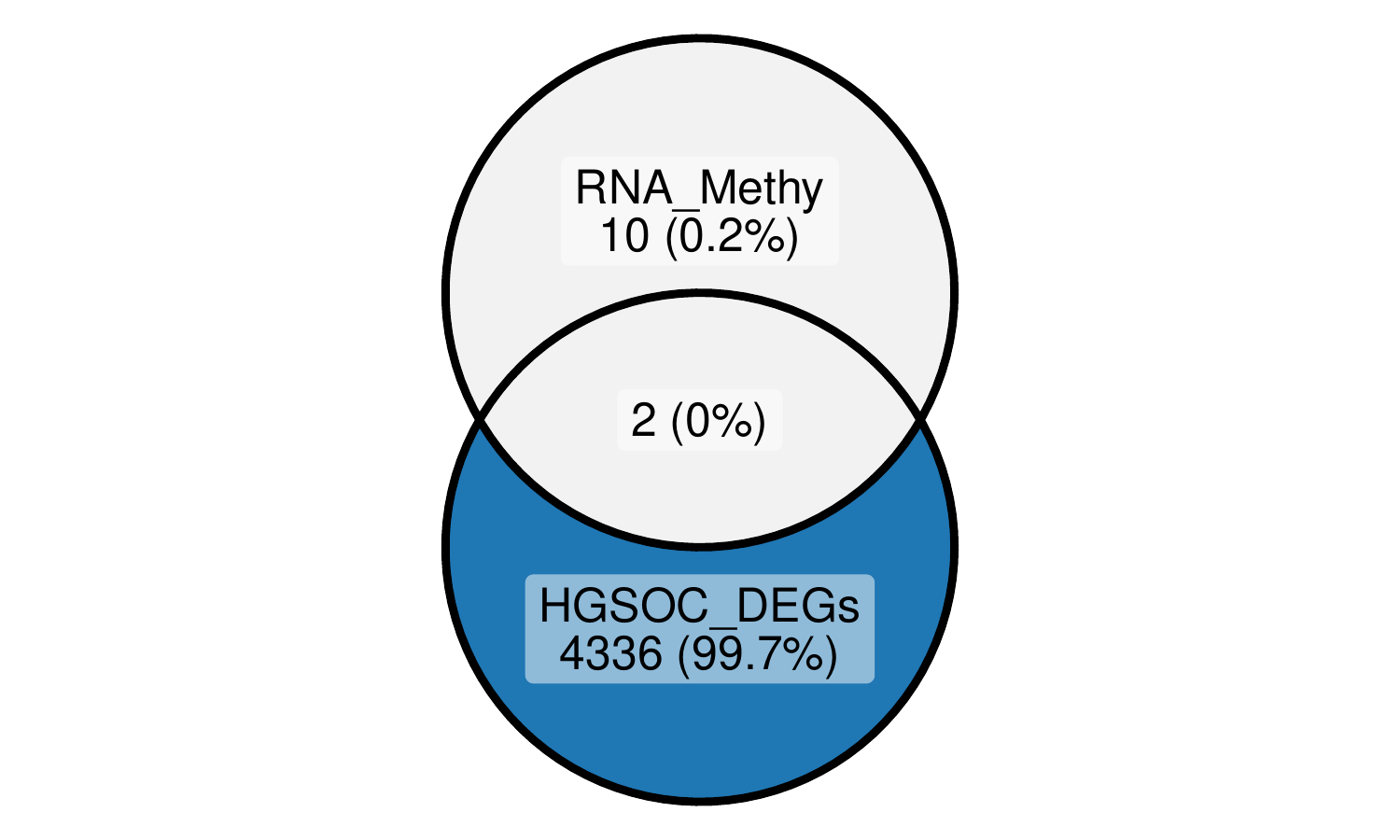
**Tab.** **2** M6A all protein of epigenetic regulators

| Id | HGNC s... | Status | HGNC ID | HGNC name | GeneID | UniPro......7 | UniPro......8 | Domain | MGI sy... |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 96 | CBLL1 | New | 21220 | E3 ubi... | 79872 | Q75N03 | HAKAI ... | PF18408 | Cbll1 |
| 387 | METTL14 | New | 29330 | Methyl... | 57721 | Q9HCE5 | MET14 ... | PF05063 | Mettl14 |
| 388 | METTL16 | New | 28480 | Methyl... | 79066 | Q86W50 | MET16 ... | PF05971 | Mettl16 |
| 390 | METTL3 | New | 17560 | Methyl... | 56339 | Q86U44 | MTA70 ... | PF05063 | Mettl3 |
| 391 | METTL4 | New | 24730 | Methyl... | 64863 | Q8N3J2 | METL4 ... | PF05063 | Mettl4 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/3.7\_EpiFactors\_表观遗传调控因子数据获取\_(M6A)/M6A-all-protein-of-epigenetic-regulators.xlsx)**

## 3.8 交集: HGSOC\_DEGs + RNA\_Methy (SIG)

Fig. **[6](#Intersection-of-HGSOC-DEGs-with-RNA-Methy)** 以下取交集： - 基因集 (hgsoc - normal, 来自于Limma 差异分析(Section: HGSOC)) - 基因集 (CBLL1, METTL14, etc., 来自于EpiFactors 表观遗传调控因子数据获取(Section: M6A))



**Fig.** **6** Intersection of HGSOC DEGs with RNA Methy

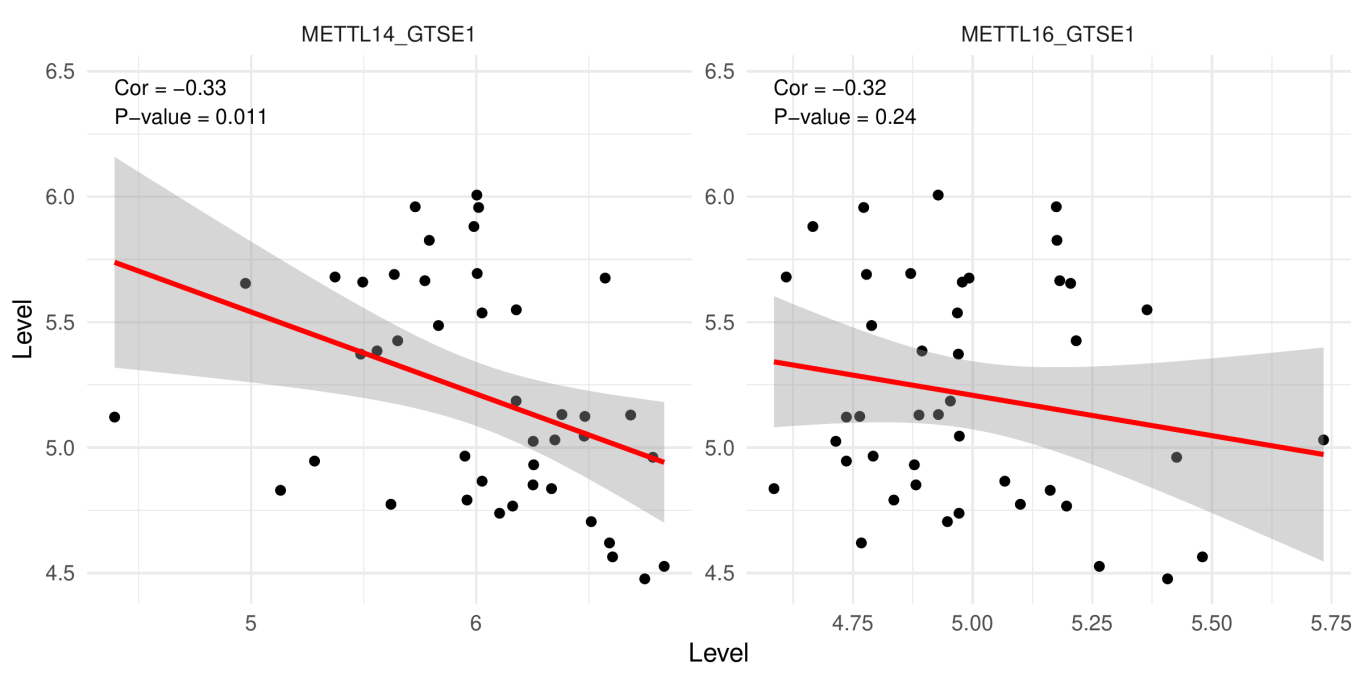
**(File path: Figure+Table/3.8\_交集:\_HGSOC\_DEGs\_+\_RNA\_Methy\_(SIG)/Intersection-of-HGSOC-DEGs-with-RNA-Methy.pdf)**

* All\_intersection: METTL14, METTL16

**(See: Figure+Table/3.8\_交集:\_HGSOC\_DEGs\_+\_RNA\_Methy\_(SIG)/Intersection-of-HGSOC-DEGs-with-RNA-Methy-content)**

## 3.9 关联分析 (HGSOC)

将基因 (METTL14, METTL16 -> GTSE1) 关联分析。



**Fig.** **7** Linear regression

**(File path: Figure+Table/3.9\_关联分析\_(HGSOC)/Linear-regression.pdf)**

# 4 总结

Fig. **[5](#HGSOC-RES-Box-Plot-Of-DEGs)** 与 Fig. **[3](#HGSOC-Box-Plot-Of-DEGs)** ， 结果表明 GTSE1 在癌症或耐药数据集中均差异表达，且癌症高表达，耐药数据集相对以于癌症低表达， 表明耐药组已无法通过调控该基因发挥抗癌作用(耐药)。

随后，筛选出的差异表达基因 METTL14，与 GTSE1 显著关联，见 Fig. **[7](#Linear-regression)** 。

# Reference

1. 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).

2. Marakulina, D. *et al.* EpiFactors 2022: Expansion and enhancement of a curated database of human epigenetic factors and complexes. *Nucleic acids research* **51**, D564–D570 (2023).