**生信分析报告**

**项目标题： 预测甲基化调控因子 ;**

**单 号： BSCL240914 ;**

**分析人员： 黄礼闯 ;**

**分析类型： 生信分析 ;**

**委 托 人： 邱美婷 ;**

**受 托 人： 杭州铂赛生物科技有限公司 .**

# 1 分析流程

## 1.1 需求

通过软件预测甲基化调控因子（如METTL14）的靶基因，并通过数据库筛选于PCOS患者中表达水平具有显著差异性的基因，合并交集，并对该交集中的基因进行功能富集和KEGG通路富集分析，筛选PCOS患者中可能的METTL14甲基化调控基因及其相关通路；

## 1.2 实际流程

从 EpiFactors 获取表观遗传调控因子，筛出甲基化相关调控因子 (A 集合) 。 获取 PCOS GEO 数据，差异分析得到 DEGs，与 m6A-Atlas 数据库比对，发现可能存在甲基化修饰位点的基因 B 集合。 在 PCOS 中筛选出差异表达的甲基化调控因子 (C集合) ，与 B 集合关联分析，随后富集分析。

# 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: PCOS)

以 R 包 GEOquery (2.74.0) 获取 GSE277906 数据集。

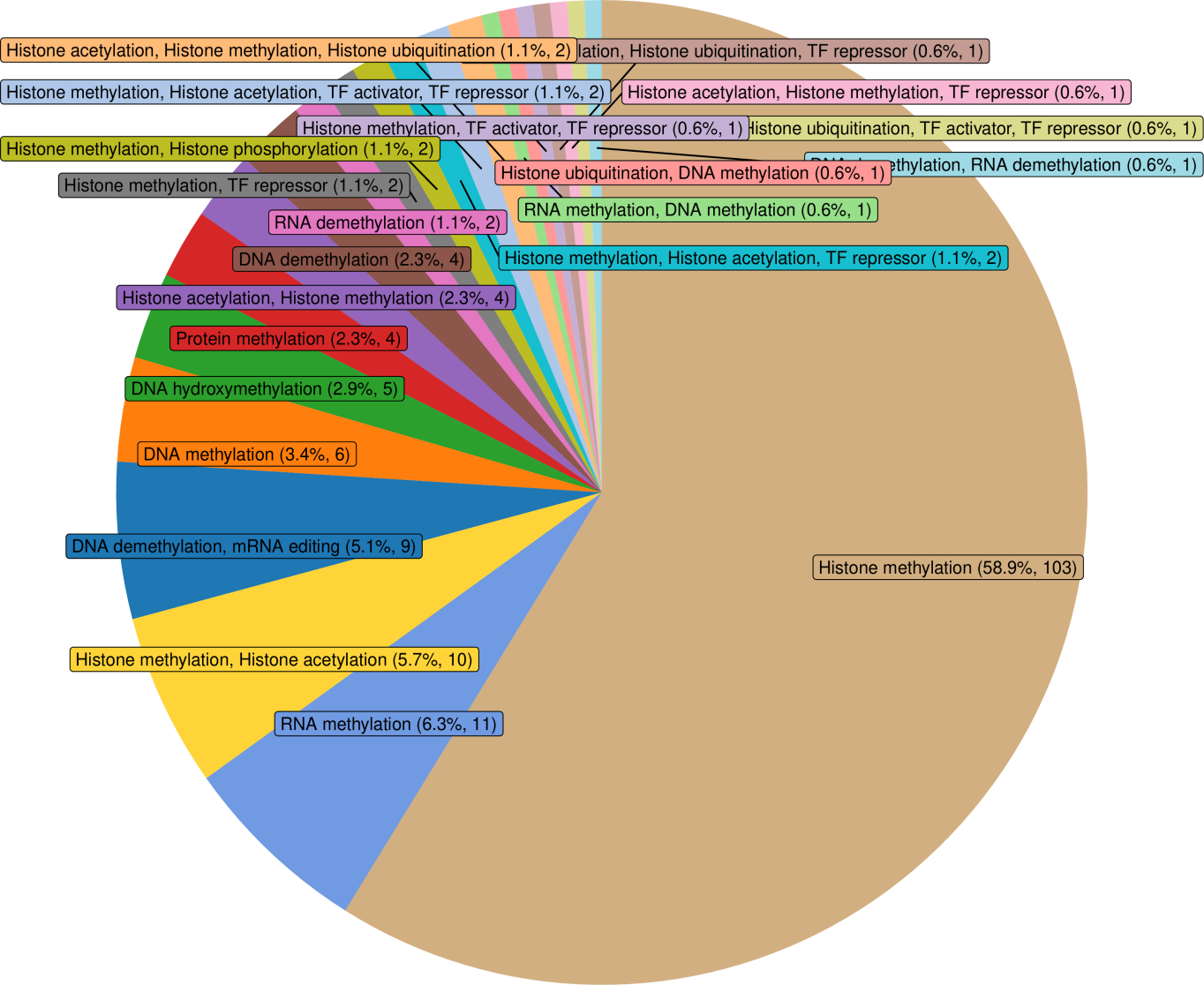
## 2.3 Limma 差异分析 (Dataset: PCOS)

以 R 包 limma (3.62.1) (2005, **IF:**, , )1 edgeR (4.4.0) (, **IF:**, , )2 进行差异分析。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)。分析方法参考 <https://bioconductor.org/packages/release/workflows/vignettes/RNAseq123/inst/doc/limmaWorkflow.html>。随后，以 公式 ~ 0 + group 创建设计矩阵 (design matrix) 用于线性分析。 使用 limma::lmFit, limma::contrasts.fit, limma::eBayes 差异分析对比组：pcos vs control。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。 对 GSE277906 的 mRNA 数据 (protein\_coding) 差异分析

# 3 分析结果

## 3.1 EpiFactors 表观遗传调控因子数据获取 (METHY)

从所有 表观调控因子 Fig. **[1](#Distribution-all-protein-of-epigenetic-regulators)** 中筛选出甲基化修饰调控因子，见 Tab. **[1](#METHY-regulators)**



**Fig.** **1** Distribution all protein of epigenetic regulators

**(File path: Figure+Table/Distribution-all-protein-of-epigenetic-regulators.pdf)**

**Tab.** **1** METHY regulators

| Id | HGNC s... | Status | HGNC ID | HGNC name | GeneID | UniPro......7 | UniPro......8 | Domain | MGI sy... |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 11 | AEBP2 | # | 24051 | AE bin... | 121536 | Q6ZN18 | AEBP2 ... | Pfam-B... | Aebp2 |
| 12 | AICDA | # | 13203 | Activa... | 57379 | Q9GZX7 | AICDA ... | APOBEC... | Aicda |
| 15 | ALKBH1 | New | 17911 | Nuclei... | 8846 | Q13686 | ALKB1 ... | PF13532 | Alkbh1 |
| 16 | ALKBH4 | New | 21900 | Alpha-... | 54784 | Q9NXW9 | ALKB4 ... | PF13532 | Alkbh4 |
| 17 | ALKBH5 | New | 25996 | AlkB h... | 54890 | Q6P6C2 | ALKB5 ... | PF13532 | Alkbh5 |
| 23 | APEX1 | # | 587 | APEX n... | 328 | P27695 | APEX1 ... | Exo en... | Apex1 |
| 24 | APOBEC1 | # | 604 | Apolip... | 339 | P41238 | ABEC1 ... | APOBEC... | Apobec1 |
| 25 | APOBEC2 | # | 605 | Apolip... | 10930 | Q9Y235 | ABEC2 ... | APOBEC... | Apobec2 |
| 26 | APOBEC3A | # | 17343 | Apolip... | 200315 | P31941 | ABC3A ... | APOBEC... | # |
| 27 | APOBEC3B | # | 17352 | Apolip... | 9582 | Q9UH17 | ABC3B ... | APOBEC... | Apobec3 |
| 28 | APOBEC3C | # | 17353 | Apolip... | 27350 | Q9NRW3 | ABC3C ... | APOBEC... | # |
| 29 | APOBEC3D | # | 17354 | Apolip... | 140564 | Q96AK3 | ABC3D ... | APOBEC... | # |
| 30 | APOBEC3F | # | 17356 | Apolip... | 200316 | Q8IUX4 | ABC3F ... | APOBEC... | # |
| 31 | APOBEC3G | # | 17357 | Apolip... | 60489 | Q9HC16 | ABC3G ... | APOBEC... | # |
| 32 | APOBEC3H | # | 24100 | Apolip... | 164668 | Q6NTF7 | ABC3H ... | APOBEC... | # |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/METHY-regulators.xlsx)**

## 3.2 GEO 数据获取 (PCOS)

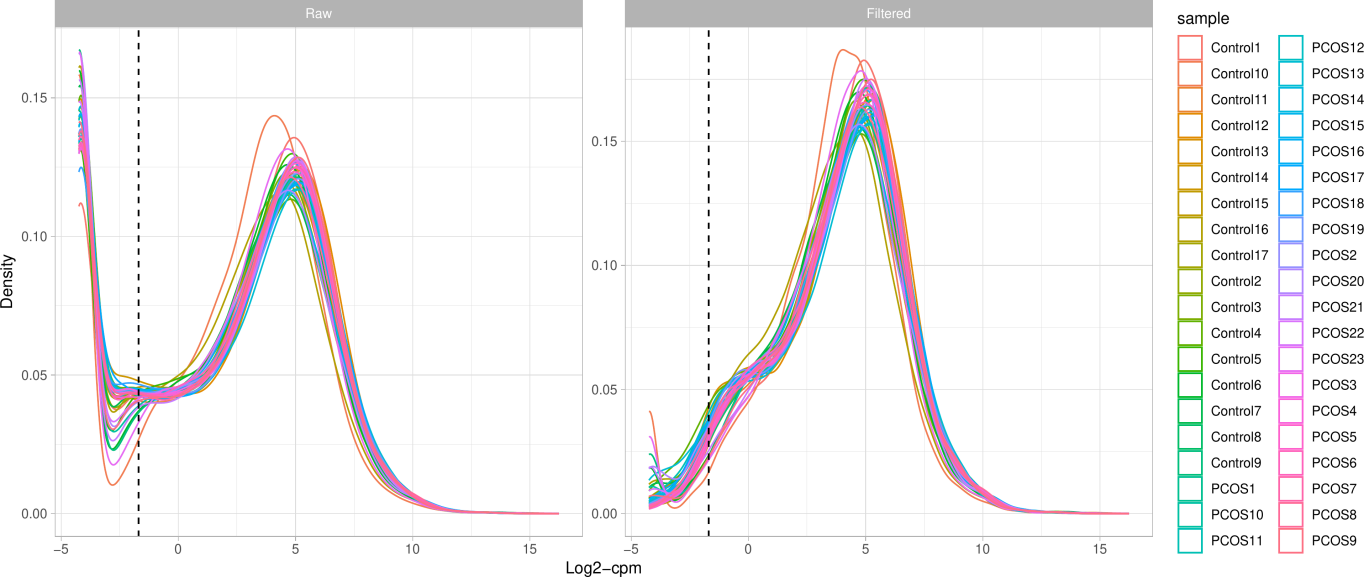
获取 GEO PCOS 数据，用于筛选差异表达基因。

* Data Source ID: GSE277906
* data\_processing: Illumina Casava1.7 software used for basecalling.
* data\_processing.1: Raw reads of fastq format were firstly processed using fastp and the low quality reads were removed to obtain the clean reads.
* data\_processing.2: The clean reads were mapped to the reference genome using HISAT2. FPKM of each gene was calculated and the read counts of each gene were obtained by HTSeq-count
* data\_processing.3: Assembly: GRCh38
* (Others): …

**(见Figure+Table/PCOS-GSE277906-content)**

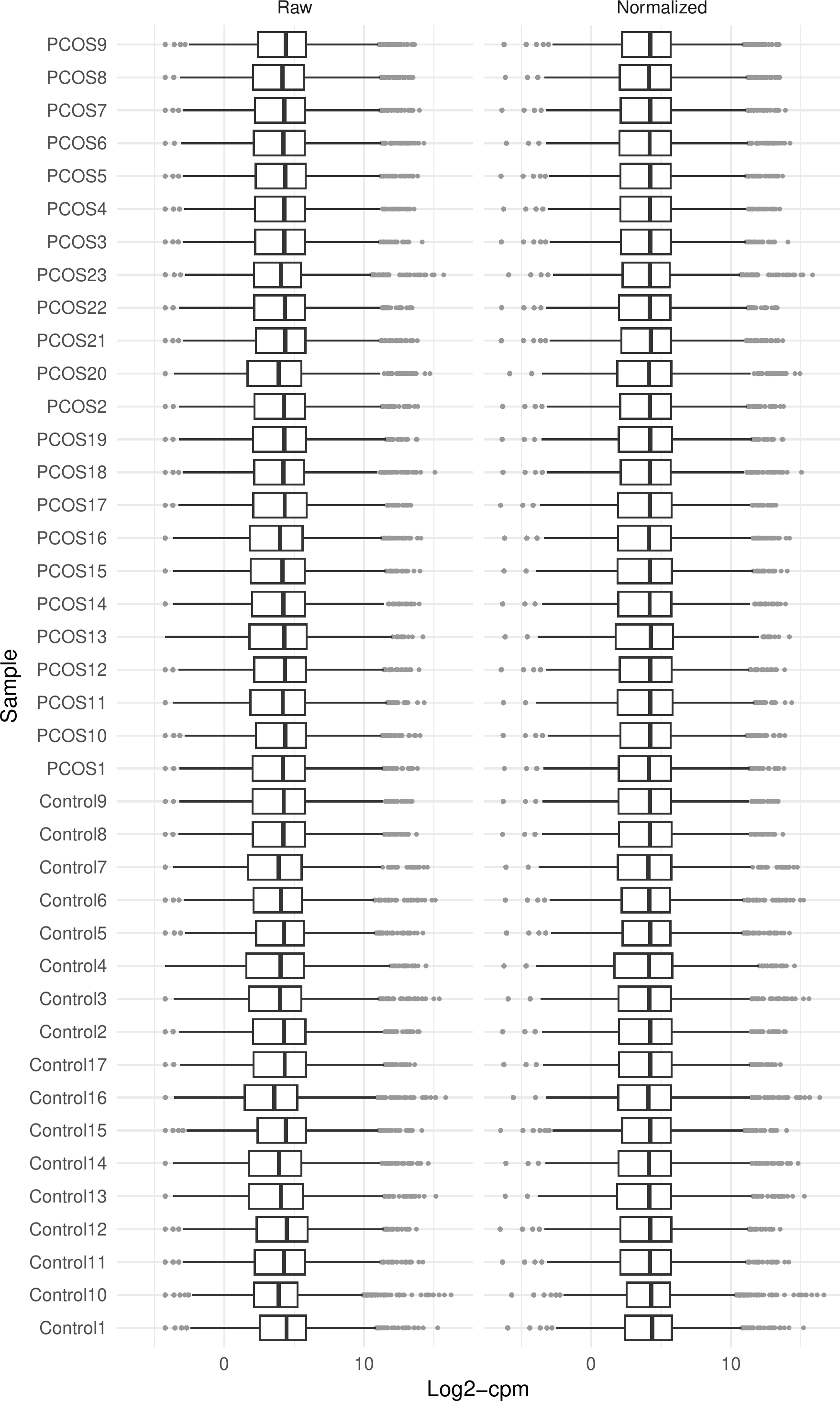
## 3.3 Limma 差异分析 (PCOS)

差异分析，得到 DEGs 见 Fig. **[4](#PCOS-pcos-vs-control)**



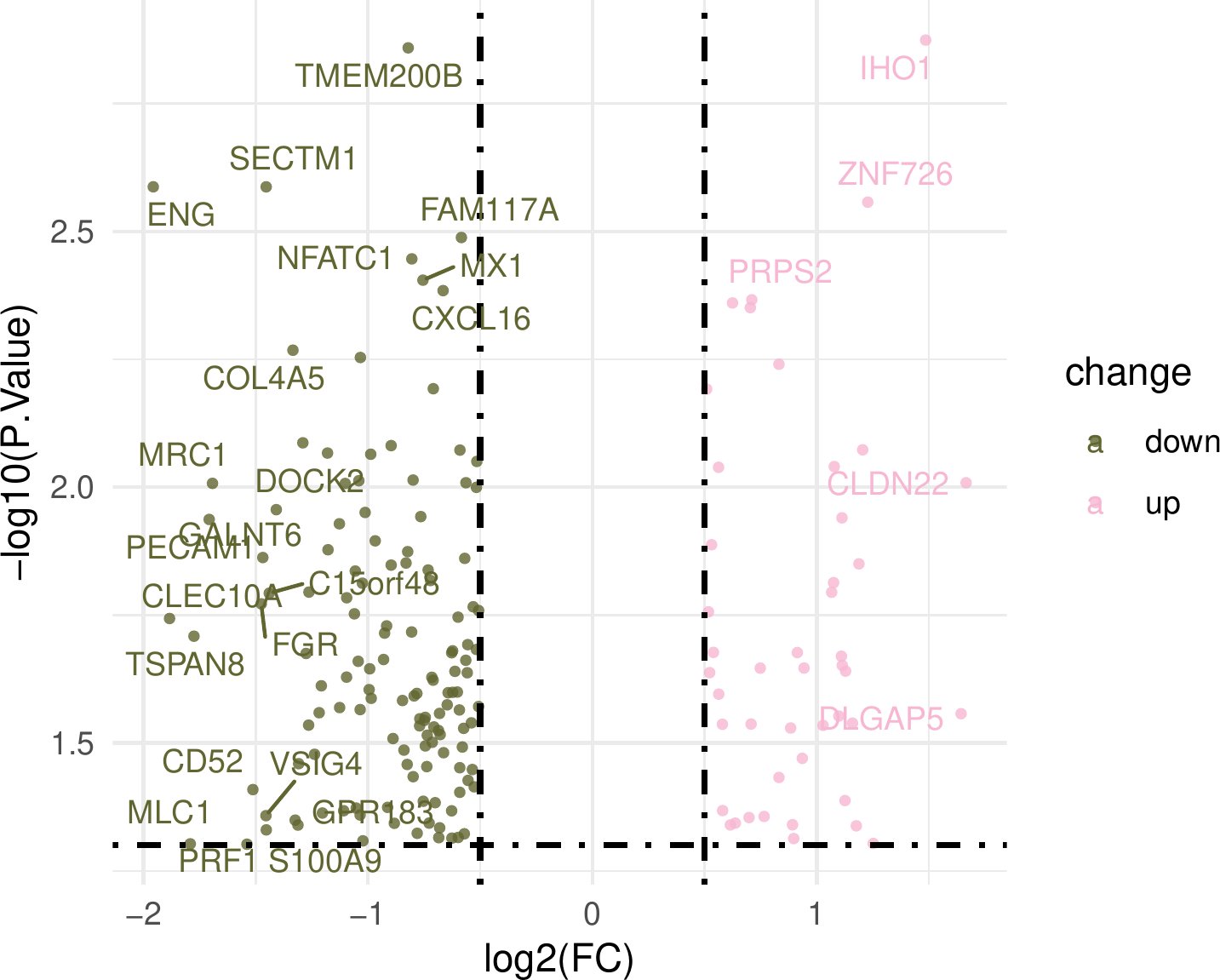
**Fig.** **2** PCOS Filter low counts

**(File path: Figure+Table/PCOS-Filter-low-counts.pdf)**



**Fig.** **3** PCOS Normalization

**(File path: Figure+Table/PCOS-Normalization.pdf)**



**Fig.** **4** PCOS pcos vs control

**(File path: Figure+Table/PCOS-pcos-vs-control.pdf)**

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

**(See: Figure+Table/PCOS-pcos-vs-control-content)**

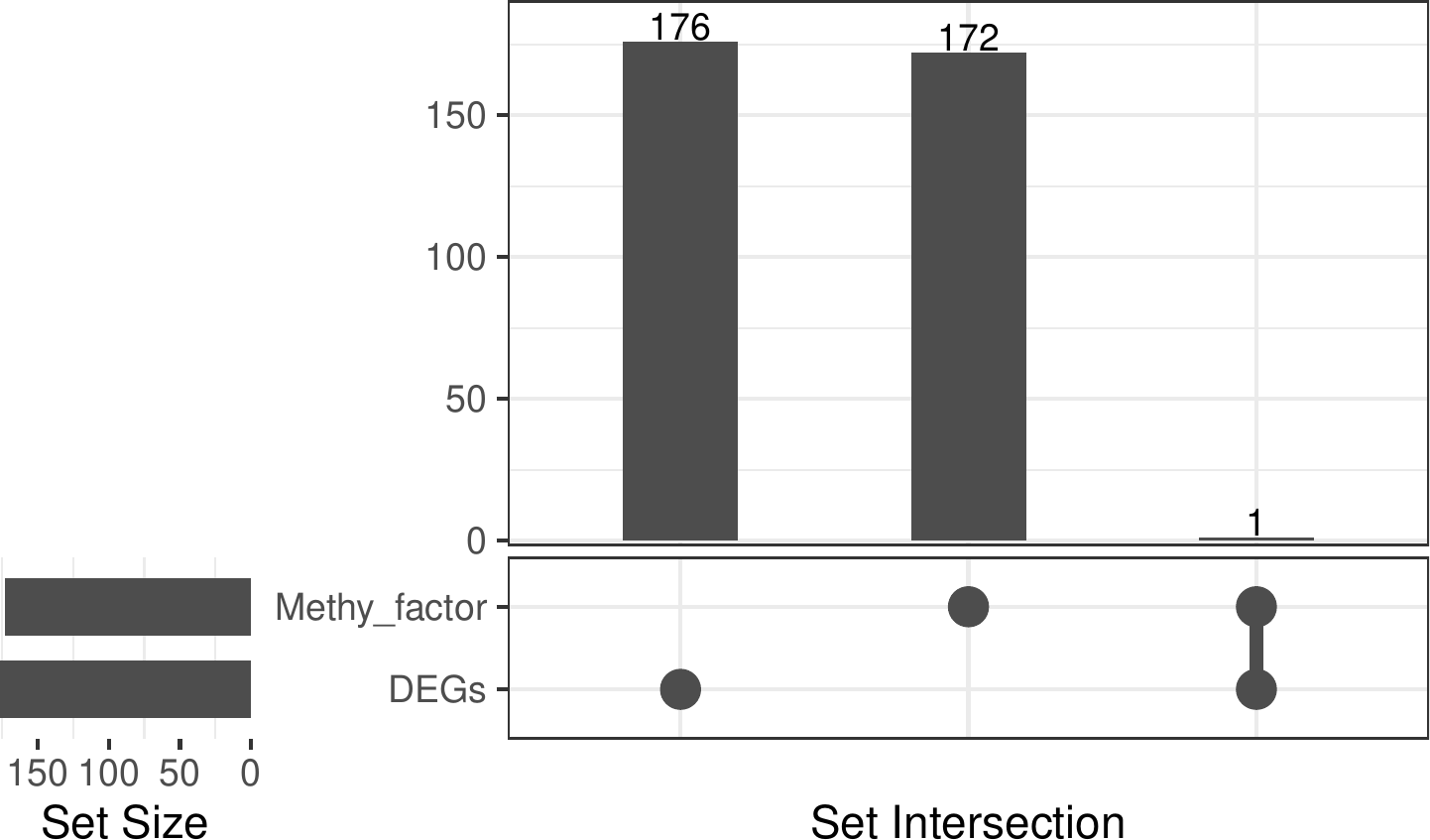
**Tab.** **2** PCOS data pcos vs control

| Rownames | Id | Gene D... | Coding... | Descri... | Pathway | Pathwa... | GO ID | GO term | Wiki ID |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PRPS2 | PRPS2 | 5634 | Protei... | Phosph... | Hsa000... | Pentos... | GO:000... | Magnes... |  |
| FXYD6 | FXYD6 | 53826 | Protei... | FXYD d... |  |  | GO:000... | Molecu... |  |
| MMP15 | MMP15 | 4324 | Protei... | Matrix... | Hsa04928 | Parath... | GO:000... | Metall... | WP5283... |
| CXCL16 | CXCL16 | 58191 | Protei... | C-X-C ... | Hsa040... | Cytoki... | GO:000... | Low-de... | WP5115... |
| MX1 | MX1 | 4599 | Protei... | MX dyn... | Hsa032... | Viral ... | GO:000... | GTPase... | WP5115... |
| HEBP2 | HEBP2 | 23593 | Protei... | Heme b... |  |  | GO:000... | Protei... |  |
| CD74 | CD74 | 972 | Protei... | CD74 m... | Hsa046... | Antige... | GO:000... | Golgi ... | WP4146... |
| CDC42EP2 | CDC42EP2 | 10435 | Protei... | CDC42 ... |  |  | GO:000... | Opioid... |  |
| GADD45B | GADD45B | 4616 | Protei... | Growth... | Hsa040... | MAPK s... | GO:000... | Protei... | WP4216... |
| PLEKHG4 | PLEKHG4 | 25894 | Protei... | Plecks... |  |  | GO:000... | Guanyl... |  |
| NFATC1 | NFATC1 | 4772 | Protei... | Nuclea... | Hsa040... | MAPK s... | GO:000... | Chroma... | WP2840... |
| BST2 | BST2 | 684 | Protei... | Bone m... | Hsa032... | Viral ... | GO:000... | Negati... | WP5115... |
| ANXA1 | ANXA1 | 301 | Protei... | Annexi... |  |  | GO:000... | Cornif... | WP98,W... |
| UNC93B1 | UNC93B1 | 81622 | Protei... | Unc-93... |  |  | GO:000... | Golgi ... |  |
| DMKN | DMKN | 93099 | Protei... | Dermokine |  |  | GO:000... | Protei... |  |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/PCOS-data-pcos-vs-control.xlsx)**

## 3.4 差异表达的 Methylation Factors

将差异表达基因与 Tab. **[1](#METHY-regulators)** 中的因子取交集， 见 Fig. **[5](#Intersection-of-Methy-factor-with-DEGs)** 。



**Fig.** **5** Intersection of Methy factor with DEGs

**(File path: Figure+Table/Intersection-of-Methy-factor-with-DEGs.pdf)**

* All\_intersection: PRDM6

**(See: Figure+Table/Intersection-of-Methy-factor-with-DEGs-content)**

**Tab.** **3** Intersection METHY epigenetic regulators

| Id | HGNC s... | Status | HGNC ID | HGNC name | GeneID | UniPro... | UniPro...1 | Domain | MGI sy... |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 510 | PRDM6 | # | 9350 | PR dom... | 93166 | Q9NQX0 | PRDM6 ... | SET PF... | Prdm6 |

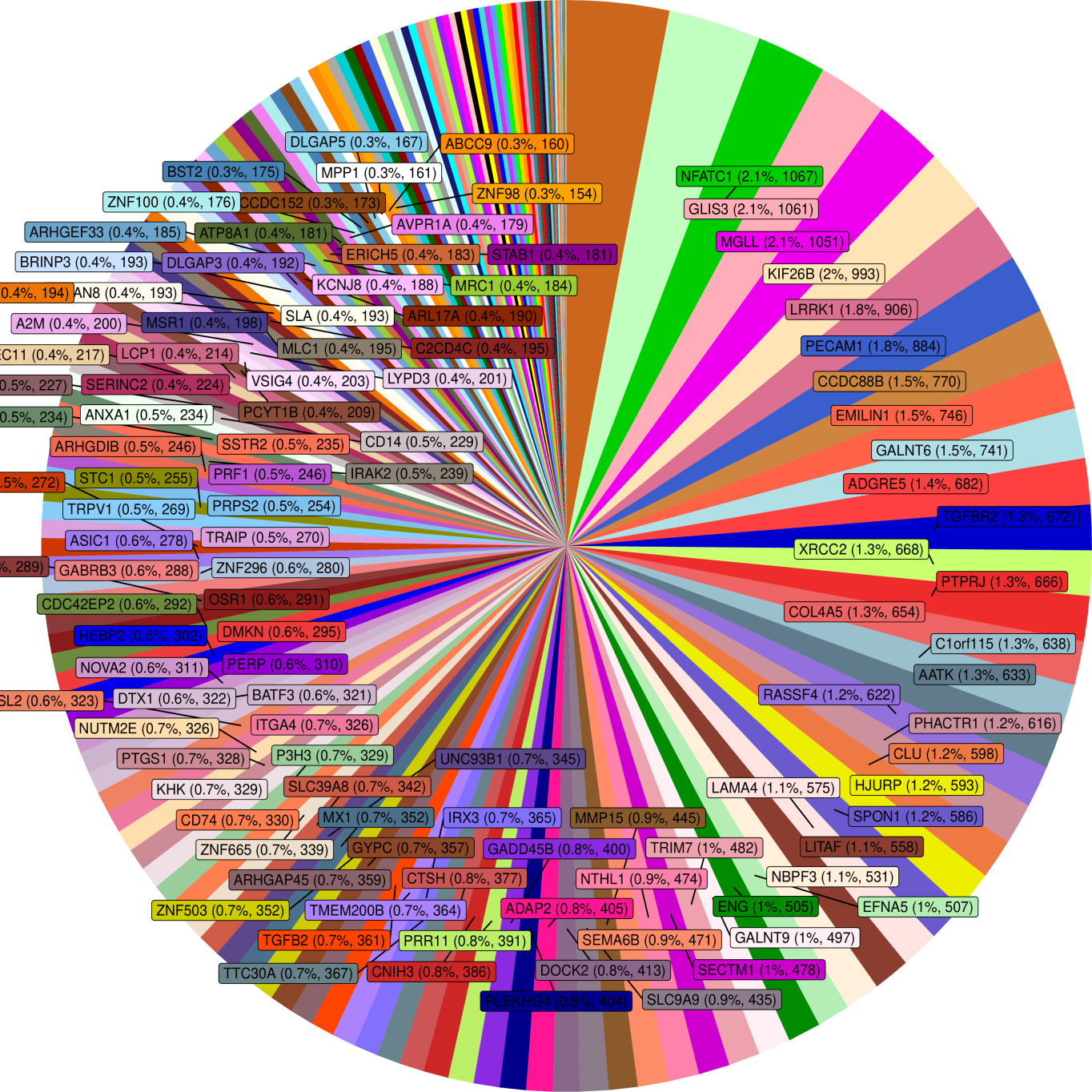
**(File path: Figure+Table/Intersection-METHY-epigenetic-regulators.xlsx)**

## 3.5 m6A-Atlas m6A 数据获取 (METHY)

据检索，所有的 DEGs (Fig. **[4](#PCOS-pcos-vs-control)**) 都存在甲基化修饰位点。 因此，所有 DEGs 都可能发生甲基化修饰 (Fig. **[6](#METHY-m6A-Atlas-search-results-distribution)** )

Note: The directory 'Figure+Table/METHY-m6A-Atlas-search-results' contains 2 files.  
  
1 1\_LowResolution.csv  
2 2\_HighResolution.csv

**(File path: Figure+Table/METHY-m6A-Atlas-search-results)**



**Fig.** **6** METHY m6A Atlas search results distribution

**(File path: Figure+Table/METHY-m6A-Atlas-search-results-distribution.pdf)**

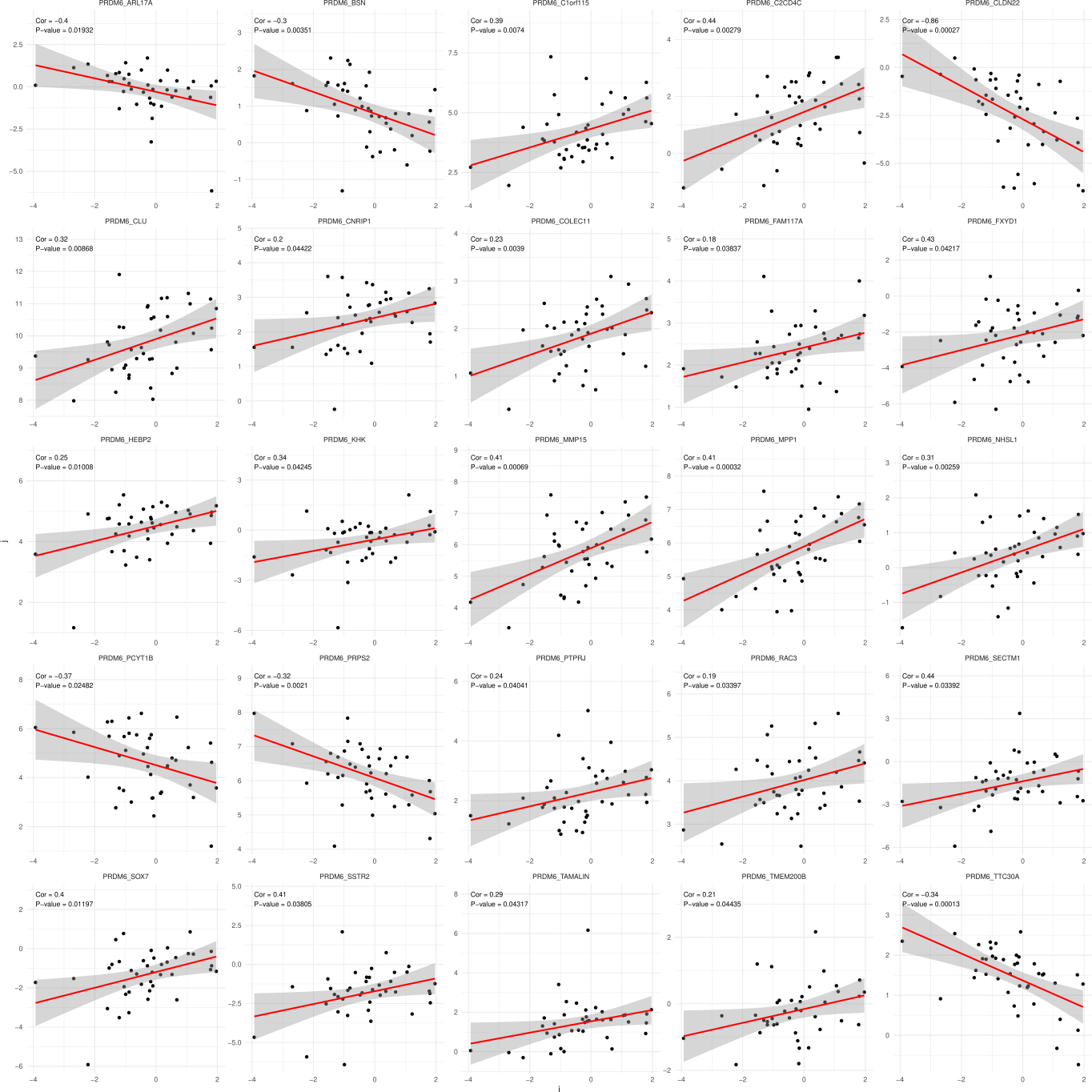
## 3.6 Methylation Factors 与 DEGs 关联分析

为了寻找 Fig. **[5](#Intersection-of-Methy-factor-with-DEGs)** 中发现的差异表达的 Methylation Factors 可能调控的 DEGs 修饰，将两个数据集作关联分析，结果见 Tab. **[4](#All-correlation-results)** 。 以 pvalue < 0.05 为条件筛选，见 Tab. **[5](#correlation-results-05)** ， Fig. **[7](#Significant-correlation)** 。 其中，pvalue < 0.001 的见Fig. **[8](#correlation-results-001)**。

**Tab.** **4** All correlation results

| From | To | Cor | Pvalue | Model | -log2(... | Signif... | Sign |
| --- | --- | --- | --- | --- | --- | --- | --- |
| PRDM6 | A2M | 0.1286... | 0.49596 | C(Cont... | 1.0117... | > 0.05 | - |
| PRDM6 | AARD | -0.195... | 0.35308 | C(Cont... | 1.5019... | > 0.05 | - |
| PRDM6 | AATK | -0.035... | 0.74928 | C(Cont... | 0.4164... | > 0.05 | - |
| PRDM6 | ABCC9 | -0.146... | 0.45304 | C(Cont... | 1.1422... | > 0.05 | - |
| PRDM6 | ADAMTSL2 | 0.1576... | 0.30011 | C(Cont... | 1.7364... | > 0.05 | - |
| PRDM6 | ADAP2 | 0.4636... | 0.053347 | C(Cont... | 4.2284... | > 0.05 | - |
| PRDM6 | ADGRE5 | -0.200... | 0.29662 | C(Cont... | 1.7533... | > 0.05 | - |
| PRDM6 | AFP | -0.215... | 0.38212 | C(Cont... | 1.3879... | > 0.05 | - |
| PRDM6 | ANXA1 | -0.309... | 0.050326 | C(Cont... | 4.3125... | > 0.05 | - |
| PRDM6 | AP3B2 | 0.0623... | 0.73583 | C(Cont... | 0.4425... | > 0.05 | - |
| PRDM6 | AQP7 | -0.118... | 0.55182 | C(Cont... | 0.8577... | > 0.05 | - |
| PRDM6 | ARHGAP45 | 0.0693... | 0.60709 | C(Cont... | 0.7200... | > 0.05 | - |
| PRDM6 | ARHGDIB | -0.013... | 0.93641 | C(Cont... | 0.0947... | > 0.05 | - |
| PRDM6 | ARHGEF33 | -0.301... | 0.057247 | C(Cont... | 4.1266... | > 0.05 | - |
| PRDM6 | ARL17A | -0.401... | 0.019322 | C(Cont... | 5.6936... | < 0.05 | \* |
| ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/All-correlation-results.xlsx)**



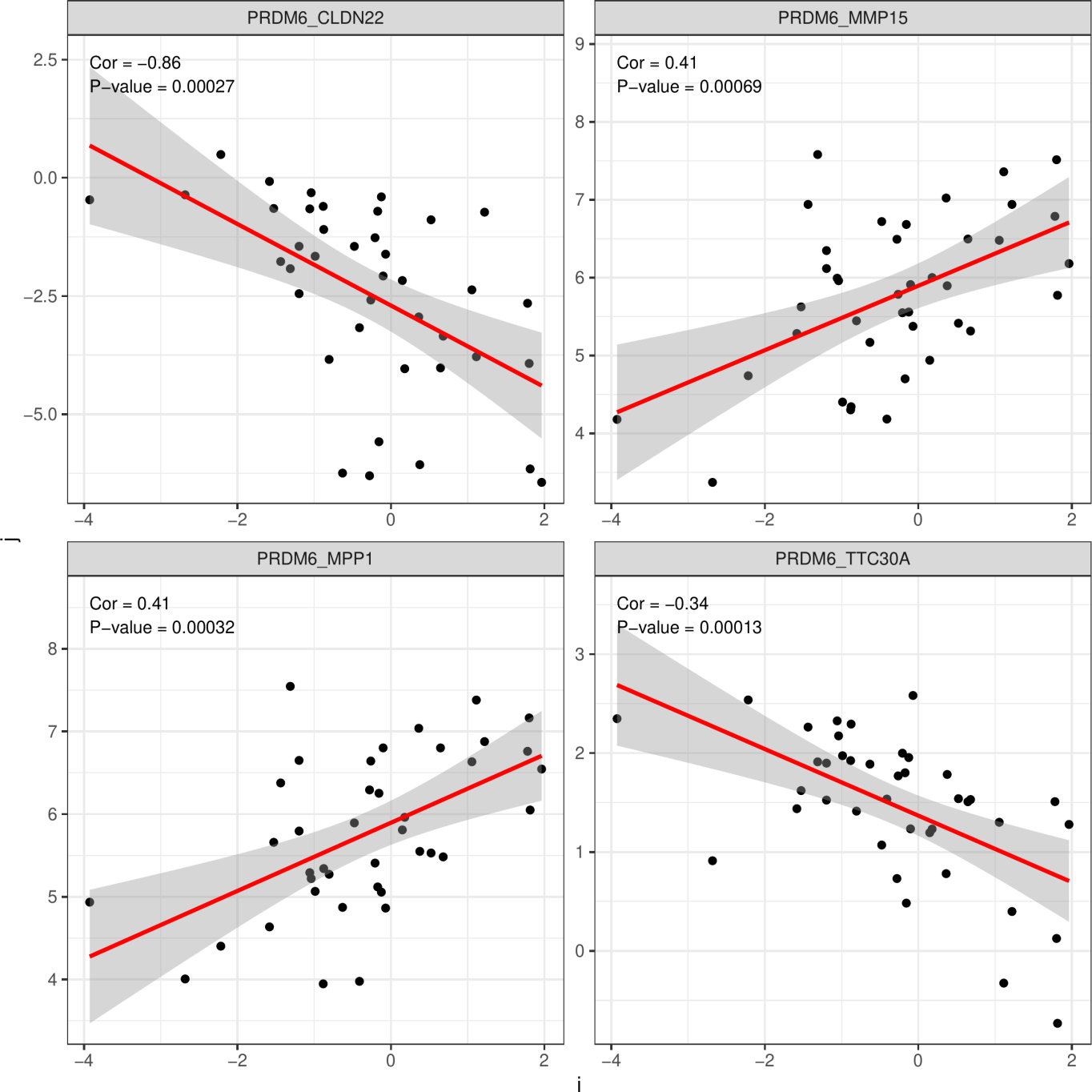
**Fig.** **7** Significant correlation

**(File path: Figure+Table/Significant-correlation.pdf)**

**Tab.** **5** Correlation results 05

| From | To | Cor | Pvalue | Model | -log2(... | Signif... | Sign |
| --- | --- | --- | --- | --- | --- | --- | --- |
| PRDM6 | ARL17A | -0.401... | 0.019322 | C(Cont... | 5.6936... | < 0.05 | \* |
| PRDM6 | BSN | -0.295... | 0.0035095 | C(Cont... | 8.1545... | < 0.05 | \* |
| PRDM6 | C1orf115 | 0.3889... | 0.0074019 | C(Cont... | 7.0778... | < 0.05 | \* |
| PRDM6 | C2CD4C | 0.4359... | 0.0027875 | C(Cont... | 8.4868... | < 0.05 | \* |
| PRDM6 | CLDN22 | -0.862... | 0.0002... | C(Cont... | 11.862... | < 0.001 | \*\* |
| PRDM6 | CLU | 0.3242... | 0.0086789 | C(Cont... | 6.8482... | < 0.05 | \* |
| PRDM6 | CNRIP1 | 0.2044... | 0.044218 | C(Cont... | 4.4992... | < 0.05 | \* |
| PRDM6 | COLEC11 | 0.2250... | 0.0038995 | C(Cont... | 8.0024... | < 0.05 | \* |
| PRDM6 | FAM117A | 0.1764... | 0.038368 | C(Cont... | 4.7039... | < 0.05 | \* |
| PRDM6 | FXYD1 | 0.4303... | 0.04217 | C(Cont... | 4.5676... | < 0.05 | \* |
| PRDM6 | HEBP2 | 0.2503... | 0.010081 | C(Cont... | 6.6322... | < 0.05 | \* |
| PRDM6 | KHK | 0.3434... | 0.042449 | C(Cont... | 4.5581... | < 0.05 | \* |
| PRDM6 | MMP15 | 0.4137... | 0.0006... | C(Cont... | 10.502... | < 0.001 | \*\* |
| PRDM6 | MPP1 | 0.4123... | 0.0003... | C(Cont... | 11.620... | < 0.001 | \*\* |
| PRDM6 | NHSL1 | 0.3122... | 0.0025927 | C(Cont... | 8.5913... | < 0.05 | \* |
| ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/correlation-results-05.xlsx)**



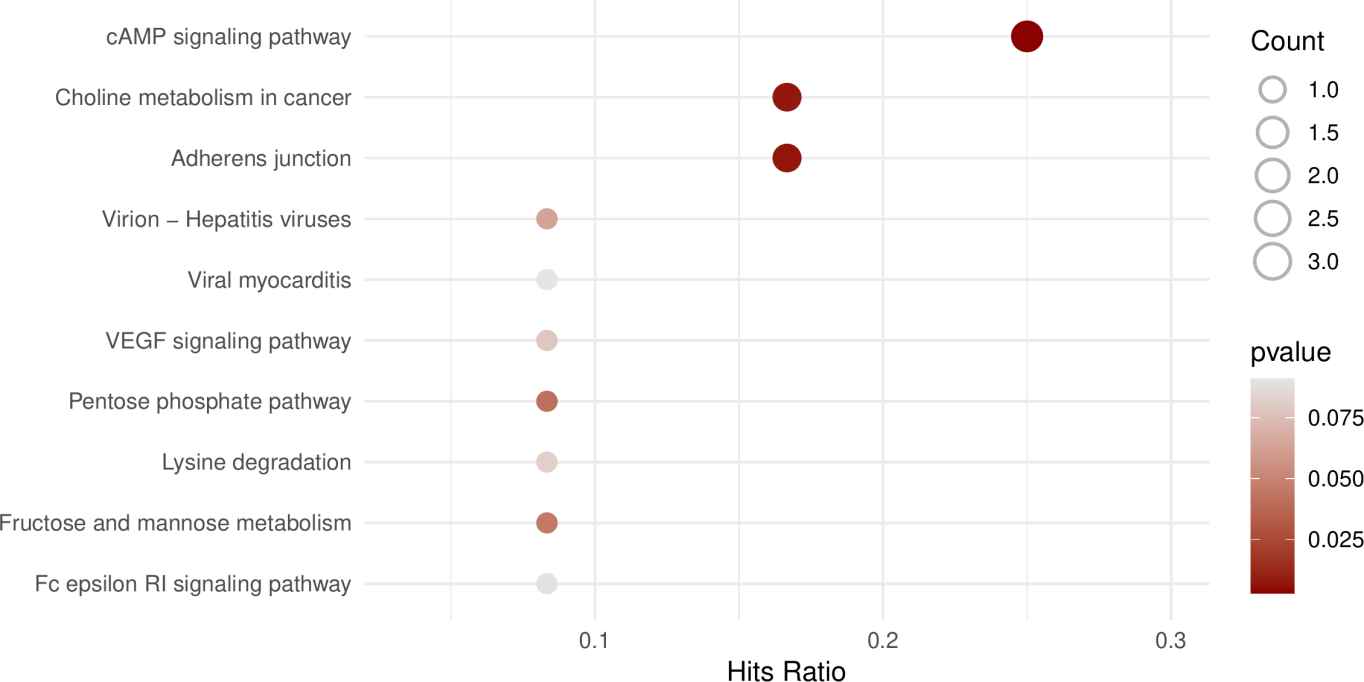
**Fig.** **8** Correlation results 001

**(File path: Figure+Table/correlation-results-001.pdf)**

## 3.7 富集分析 (SIGCOR\_05)

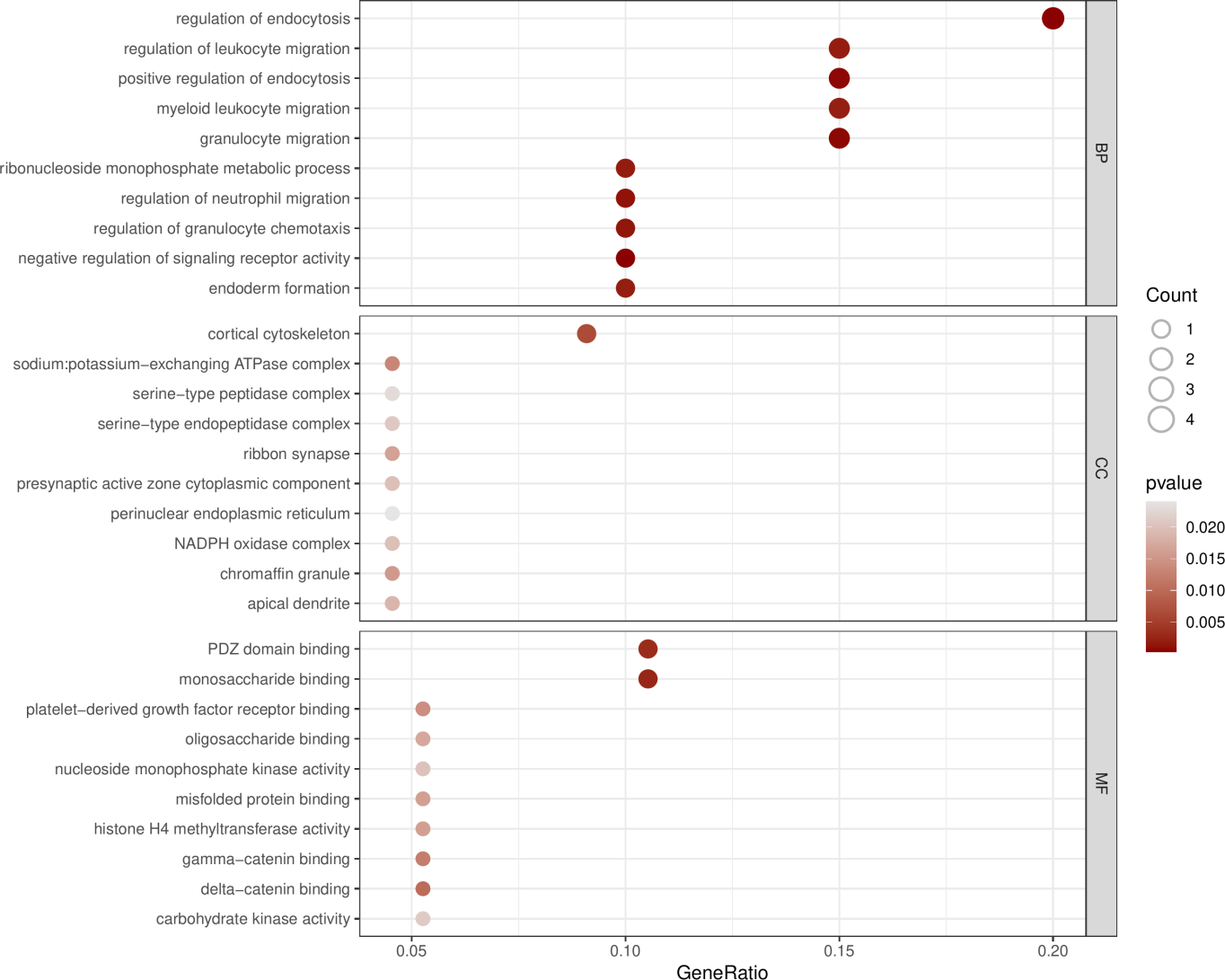
将 Tab. **[5](#correlation-results-05)** 中的基因富集分析 (包含 PRDM6)，

KEGG，GO 结果见 Fig. **[9](#SIGCOR-05-KEGG-enrichment)**， Fig. **[10](#SIGCOR-05-GO-enrichment)** 。 Fig. **[11](#SIGCOR-05-hsa04024-visualization)** 为 KEGG 中最为显著的 cAMP 通路，可能与 PCOS 中甲基化调控相关。 富集分析的数据表格见 Tab. **[6](#SIGCOR-05-KEGG-enrichment-data)**， Tab. **[7](#SIGCOR-05-GO-enrichment-data)** 。



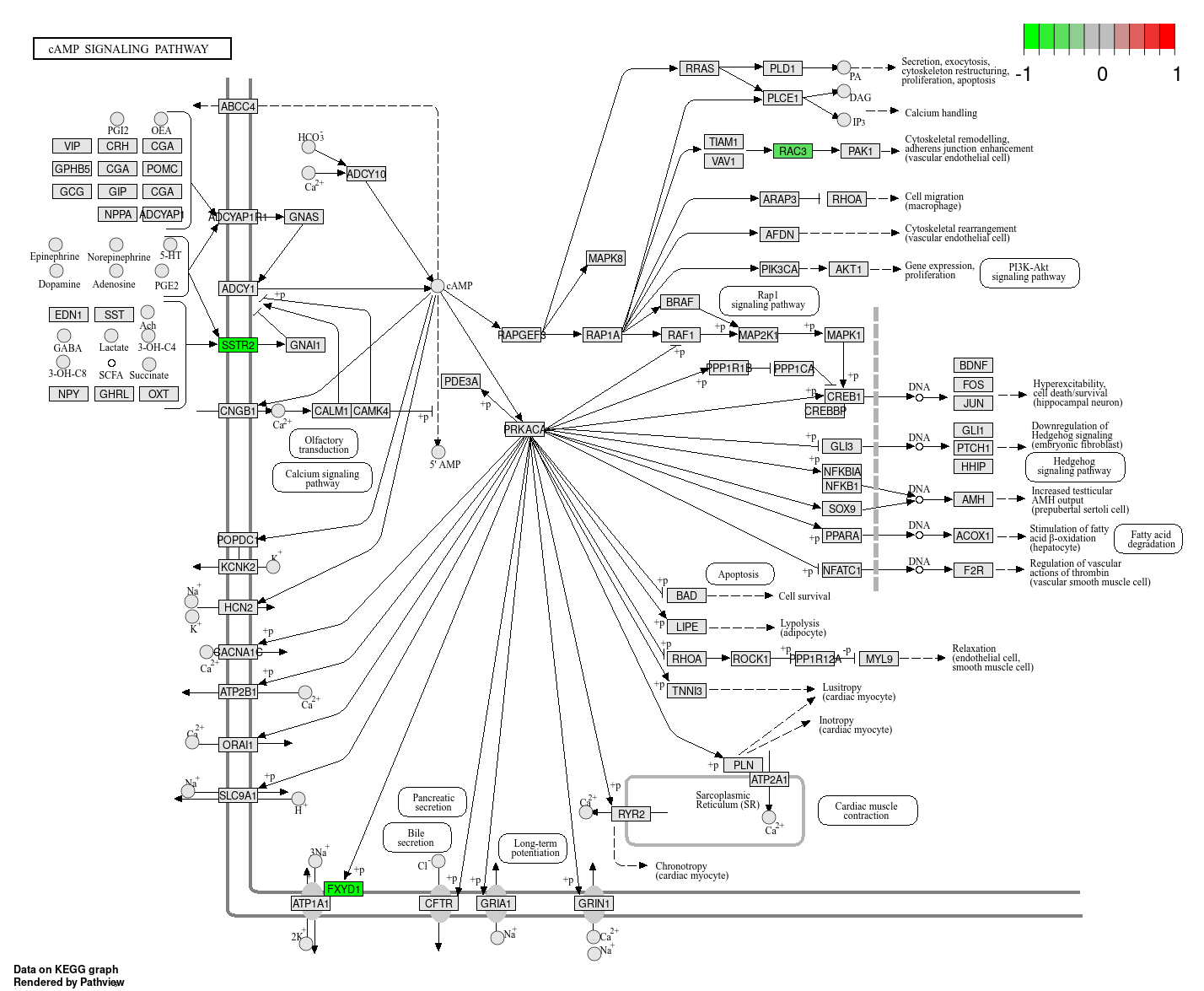
**Fig.** **9** SIGCOR 05 KEGG enrichment

**(File path: Figure+Table/SIGCOR-05-KEGG-enrichment.pdf)**



**Fig.** **10** SIGCOR 05 GO enrichment

**(File path: Figure+Table/SIGCOR-05-GO-enrichment.pdf)**



**Fig.** **11** SIGCOR 05 hsa04024 visualization

**(File path: Figure+Table/SIGCOR-05-hsa04024-visualization.png)**

* Interactive figure:
* Enriched genes: SSTR2, FXYD1, RAC3

**Tab.** **6** SIGCOR 05 KEGG enrichment data

| Category | Subcat... | ID | Descri... | GeneRatio | BgRatio | Pvalue | P.adjust | Qvalue | GeneID |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Enviro... | Signal... | Hsa04024 | CAMP s... | 3/12 | 226/8868 | 0.0030... | 0.1085... | 0.1009... | 5348/5... |
| Cellul... | Cellul... | Hsa04520 | Adhere... | 2/12 | 93/8868 | 0.0067... | 0.1085... | 0.1009... | 5795/5881 |
| Human ... | Cancer... | Hsa05231 | Cholin... | 2/12 | 99/8868 | 0.0075... | 0.1085... | 0.1009... | 9468/5881 |
| Metabo... | Carboh... | Hsa00030 | Pentos... | 1/12 | 31/8868 | 0.0411... | 0.2895... | 0.2693... | 5634 |
| Metabo... | Carboh... | Hsa00051 | Fructo... | 1/12 | 34/8868 | 0.0450... | 0.2895... | 0.2693... | 3795 |
|  |  | Hsa03272 | Virion... | 1/12 | 48/8868 | 0.0630... | 0.2895... | 0.2693... | 53842 |
| Enviro... | Signal... | Hsa04370 | VEGF s... | 1/12 | 60/8868 | 0.0782... | 0.2895... | 0.2693... | 5881 |
| Metabo... | Amino ... | Hsa00310 | Lysine... | 1/12 | 63/8868 | 0.0820... | 0.2895... | 0.2693... | 93166 |
| Organi... | Immune... | Hsa04664 | Fc eps... | 1/12 | 69/8868 | 0.0895... | 0.2895... | 0.2693... | 5881 |
| Human ... | Cardio... | Hsa05416 | Viral ... | 1/12 | 70/8868 | 0.0907... | 0.2895... | 0.2693... | 5881 |
| Metabo... | Global... | Hsa01230 | Biosyn... | 1/12 | 75/8868 | 0.0969... | 0.2895... | 0.2693... | 5634 |
| Organi... | Digest... | Hsa04971 | Gastri... | 1/12 | 76/8868 | 0.0981... | 0.2895... | 0.2693... | 6752 |
| Human ... | Cancer... | Hsa05212 | Pancre... | 1/12 | 77/8868 | 0.0994... | 0.2895... | 0.2693... | 5881 |
| Human ... | Cancer... | Hsa05210 | Colore... | 1/12 | 87/8868 | 0.1116... | 0.2895... | 0.2693... | 5881 |
| Organi... | Immune... | Hsa04610 | Comple... | 1/12 | 88/8868 | 0.1128... | 0.2895... | 0.2693... | 1191 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/SIGCOR-05-KEGG-enrichment-data.xlsx)**

**Tab.** **7** SIGCOR 05 GO enrichment data

| Ont | ID | Descri... | GeneRatio | BgRatio | Pvalue | P.adjust | Qvalue | GeneID | Count |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BP | GO:003... | Regula... | 4/20 | 307/18986 | 0.0002... | 0.0821... | 0.0605... | 1191/7... | 4 |
| BP | GO:200... | Negati... | 2/20 | 25/18986 | 0.0003... | 0.0821... | 0.0605... | 25927/... | 2 |
| BP | GO:009... | Granul... | 3/20 | 156/18986 | 0.0005... | 0.0821... | 0.0605... | 4354/5... | 3 |
| BP | GO:004... | Positi... | 3/20 | 159/18986 | 0.0005... | 0.0821... | 0.0605... | 1191/7... | 3 |
| BP | GO:190... | Regula... | 2/20 | 47/18986 | 0.0011... | 0.1115... | 0.0822... | 4354/5881 | 2 |
| BP | GO:007... | Regula... | 2/20 | 52/18986 | 0.0013... | 0.1115... | 0.0822... | 4354/5795 | 2 |
| BP | GO:000... | Ribonu... | 2/20 | 59/18986 | 0.0017... | 0.1115... | 0.0822... | 4354/5634 | 2 |
| BP | GO:000... | Regula... | 3/20 | 236/18986 | 0.0018... | 0.1115... | 0.0822... | 4354/5... | 3 |
| BP | GO:000... | Endode... | 2/20 | 61/18986 | 0.0018... | 0.1115... | 0.0822... | 4324/8... | 2 |
| BP | GO:009... | Myeloi... | 3/20 | 243/18986 | 0.0020... | 0.1115... | 0.0822... | 4354/5... | 3 |
| BP | GO:000... | Comple... | 2/20 | 67/18986 | 0.0022... | 0.1128... | 0.0832... | 1191/7... | 2 |
| BP | GO:005... | Positi... | 2/20 | 73/18986 | 0.0026... | 0.1224... | 0.0903... | 78989/... | 2 |
| BP | GO:000... | Nucleo... | 2/20 | 76/18986 | 0.0028... | 0.1224... | 0.0903... | 4354/5634 | 2 |
| BP | GO:000... | Endode... | 2/20 | 87/18986 | 0.0037... | 0.1481... | 0.1093... | 4324/8... | 2 |
| BP | GO:001... | Regula... | 2/20 | 91/18986 | 0.0040... | 0.1510... | 0.1114... | 25927/... | 2 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

**(File path: Figure+Table/SIGCOR-05-GO-enrichment-data.xlsx)**

# 4 总结

筛选的甲基化调控因子为 PRDM6，可能调控的基因见 Tab. **[5](#correlation-results-05)** ， 富集分析结果中，cAMP 通路最为显著，Fig. **[11](#SIGCOR-05-hsa04024-visualization)** 。

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