

LncRNA 差异分析结果：

=====文件=====

|-- DEG_Cis_anno.stat.xls 差异 LncRNA 的顺式靶基因注释统计结果

|-- DEG_Trans_anno.stat.xls 差异 LncRNA 的反式靶基因注释统计结果 (样品数目不少于 5)

|-- DEG.stat.xls 差异 LncRNA 数目统计结果

=====文件夹 1 =====

|-- BMK_1_All_DEG 所有差异 LncRNA 统计整合结果

|-- All.DEG_final_Cis_anno.xls 所有差异 LncRNA/表达量/顺式靶基因/顺式靶基因注释

整合结果 (整合文件, 信息全面, 使用方便, 比较重要)

|-- #Gene : 顺式靶基因 ID ;
|-- ID : 差异 LncRNA ID ;
|-- 样本 ID : 表达量列 ;
|-- *_vs_*.Pvalue/FDR : 差异分组的 Pvalue/FDR ;
|-- *_vs_*.log2FC : 差异分组的 log2FC ;
|-- *_vs_*.regulated : regulated (差异类型: 上下调)。

|-- All.DEG_final.pdf 所有差异 LncRNA 表达量聚类热图 (pdf)

|-- All.DEG_final.png 所有差异 LncRNA 表达量聚类热图 (png)

横坐标代表样品名称及样品的聚类结果, 纵坐标代表的差异 lncRNA 及 lncRNA 的聚类结果。图中不同的列代表不同的样品, 不同的行代表不同的基因。颜色代表了基因在样品中的表达量水平 $\log_{10}(\text{lncRNA}+0.000001)$, 表达量的高低可根据色度条判断。

|-- All.DEG_final_Trans_anno.xls 所有差异 LncRNA/表达量/反式靶基因/反式靶基因注释

整合结果 (样品数目不少于 5)

|-- All.DEG_final.xls 所有差异 LncRNA 表达量/Pvalue 或者 FDR/regulated 整合结果

|-- Veen.pdf 不同组合差异基因数目 Venn 图 ($2 \leq$ 差异组合数目 < 9), PDF 格式

|-- Veen.png 不同组合差异基因数目 Venn 图 ($2 \leq$ 差异组合数目 < 9), PNG 格式

=====**文件夹 2**=====

|-- **BMK_*_*_vs_*** 条件 1 样品和条件 2 样品的差异表达分析结果目录

*_vs_*表示差异分组的样本编号，如 BMK_2_L01_L02_vs_L03_L04 表示 L01，L02 与 L03，L04 做差异比较。

在注释统计表中不存在的内容表示没有对该数据库做注释，不存在的文件夹表示没有对该项内容做分析。

文件夹 2 下的内容包括**差异分析的相关图表文件夹**以及**注释和富集结果文件夹**

图表文件夹 BMK_1_Statistics_Visualization 内容如下：

|-- *_vs_*.all.xls 该组合中所有 lncRNA 的表达量/log2FC/Pvalue/FDR

|-- *_vs_*.DEG_final.Cis_Target.xls 差异 lncRNA 增加顺式靶基因 ID

|-- *_vs_*.DEG_final.Trans_Target.xls 差异 lncRNA 增加反式靶基因 ID (**当样品数目不少于 5 时**)

|-- *_vs_*.DEG_final.xls 差异 lncRNA 的信息 (经过 FC/FDR or Pvalue 筛选后的结果) (**重要文件**)

|-- *_vs_*.heatmap.pdf 差异表达 lncRNA 聚类热图 (PDF 格式)

|-- *_vs_*.heatmap.png 差异表达 lncRNA 聚类热图 (PNG 格式)

横坐标代表样品名称及样品的聚类结果，纵坐标代表的差异 lncRNA 及 lncRNA 的聚类结果。图中不同的列代表不同的样品，不同的行代表不同的基因。颜色代表了基因在样品中的表达量水平 $\log_{10}(\text{lncRNA} + 0.000001)$ 。

|-- *_vs_*.MA.pdf 差异表达 MA 图 (PDF 格式)

|-- *_vs_*.MA.png 差异表达 MA 图 (PNG 格式)

差异表达基因 MA 图中每一个点代表一个 lncRNA。横坐标为 A 值： $\log_2(\text{FPKM})$ ，即两样品中表达量均值的对数值；纵坐标为 M 值： $\log_2(\text{FC})$ ，即两样品间 lncRNA 表达量差异倍数的对数值，用于衡量表达量差异的大小。图中绿色的点代表显著下调的 lncRNA，红色的点代表显著上调的 lncRNA，黑色的点代表表达差异不显著的 lncRNA。

|-- *_vs_*.Volcano.pdf 差异表达火山图 (PDF 格式)

|-- *_vs_*.Volcano.png 差异表达火山图 (PNG 格式)

差异表达火山图中的每一个点表示一个 lncRNA，横坐标表示某一个基因在两样品中表达量差异倍数的对数值；纵坐标表示 pvalue 的负对数值。横坐标绝对值越大，说明表达量在两样品间的表达量倍数差异越大；纵坐标值越大，表明差异表达越显著，筛选得到的差异表达基因越可靠。图中绿色的点代表下调差异表达 lncRNA，红色的点代表上调差异表达 lncRNA，黑色代表非差异表达 lncRNA。

注释文件夹如下：

Cis_Anno_enrichment 差异 LncRNA 的顺式靶基因注释和富集结果

Trans_Anno_enrichment 差异 LncRNA 的反式靶基因注释和富集结果（当样品数目不少于 5 时）

注释文件夹下包括三个文件夹：

BMK_1_Annotation 靶基因注释和 GO 二级分类结果

BMK_2_GO_enrich GO 富集分析结果

BMK_3_KEGG_enrich KEGG 富集分析结果

详细内容如下：

=====BMK_1_Annotation=====

|-- *_vs_*.annotation.xls 差异基因顺/反式靶基因注释信息和差异 LncRNA 信息整合文件（重要文件）

|-- *_vs_*.GO_enrichment.stat.xls GO 二级节点注释分类统计结果文件

|-- *_vs_*.GO.pdf GO 注释分类统计图（PDF 格式）

|-- *_vs_*.GO.png GO 注释分类统计图（PNG 格式）

横坐标为 GO 分类，纵坐标左边为基因数目所占百分比，右边为基因数目。

此图展示的是在差异表达基因背景和全部基因背景下 GO 各二级功能的基因富集情况，体现两个背景下各二级功能的地位，具有明显比例差异的二级功能说明差异表达基因与全部基因的富集趋势不同，可以重点分析此功能是否与差异相关。

=====BMK_2_GO_enrich=====

此部分分析是采用 R 包 clusterProfiler 对 LncRNA 的顺/反式靶基因分别进行生物学过程、分子功能和细胞组分的富集分析。富集分析采用超几何检验方法来寻找与整个基因组背景相比显著富集的 GO 条目。

此目录下的 BP/CC/MF 分别是 Biological_Process，Cellular_Component 和 Molecular_Function 的简写。

|-- *_vs_*(BP/CC/MF)_enrich_barplot.pdf

|-- *_vs_*(BP/CC/MF)_enrich_barplot.png 差异表达 lncRNA 顺/反式靶基因 GO 富集条形图

|-- *_vs_*(BP/CC/MF)_enrich_dotplot.pdf

|-- *_vs_*(BP/CC/MF)_enrich_dotplot.png 差异表达 lncRNA 顺/反式靶基因 GO 富集气泡图

横坐标为 GeneRatio 即注释在该条目中的感兴趣基因占有所有差异 lncRNA 顺式靶基因数的比例，纵坐标每一个 BP/CC/MF 条目。点的大小代表该通路中注释的差异表达 lncRNA 顺式靶基因数，点的颜色代表超几何检验的校正后的 p 值。

|-- *_vs_*_Biological_Process_enrich.list.xls 富集结果文件

|--ID GO 节点；
|--Description GO 节点名称；
|--GeneRatio 富集到此节点上的差异 lncRNA 靶基因的数目/所有差异 lncRNA 靶基因的数目；
|--BgRatio 富集到此节点上的所有靶基因的数目/所有靶基因的数目；
|--enrich_factor 富集因子；
|--pvalue Pvalue；
|--qvalue Qvalue；
|--geneID 富集到此通路上的差异 lncRNA 的靶基因 ID；
|--gene_symbol 对应的 Name。

=====BMK_3_KEGG_enrich=====

此部分分析是采用 R 包 clusterProfiler 对 lncRNA 的顺/反式靶基因进行 KEGG 通路富集分析。富集分析采用超几何检验方法来寻找与整个基因组背景相比**显著富集的 KEGG 通路**。

|-- *_vs_*_KEGG_pathway_enrich_barplot.pdf

|-- *_vs_*_KEGG_pathway_enrich_barplot.png 差异表达 lncRNA 顺/反式靶基因 KEGG 富集条形图

(PNG 格式)

横坐标代表差异表达 lncRNA 顺式靶基因注释在通路中的基因数，纵坐标代表通路，柱的颜色代表校正后的 p 值。

|-- *_vs_*_KEGG_pathway_enrich_dotplot.pdf

|-- *_vs_*_KEGG_pathway_enrich_dotplot.png 差异表达 lncRNA 顺式靶基因 KEGG 富集气泡图

(PNG 格式)

|-- *_vs_*_KEGG_pathway_enrich.list.xls 富集结果文件

```
|--ID      ko_id ;
|--Description  ko 名称 ;
|--GeneRatio   富集到此通路上的差异 LncRNA 靶基因的数目/所有差异 LncRNA 靶基因的数目 ;
|--BgRatio     富集到此通路上的所有靶基因的数目/所有靶基因的数目 ;
|--enrich_factor 富集因子 ;
|--pvalue      Pvalue ;
|--qvalue      Qvalue ;
|--geneID      富集到此通路上的差异 LncRNA 的靶基因的 ID ;
|--gene_symbol 对应的 Name。
```

|-- BMK_1_Keggmap KEGG 通路图

|-- ko*.html (html 格式)

|-- ko*.png KEGG 通路图

相对于对照组来说，红色框标记的酶与上调基因有关，绿色框标记的酶与下调基因有关。蓝色框标记的酶与上调和下调基因均有关，框内的数字代表酶的编号（EC number），而整个通路由多种酶催化的复杂生化反应构成，此通路图中与差异表达基因相关的酶均用不同的颜色标出，研究人员可以根据自己的研究对象间的差异，重点研究某些代谢通路相关基因的差异表达情况，通过通路解释表型差异的根源。