**circRNA差异分析结果：**

**=====================================文件=====================================**

**|-- DEG\_anno.stat.xls** **差异基因的注释信息统计结果**

**|-- DEG.stat.xls** **不同差异分组中差异circRNA数目(up/down)的统计结果**

**=====================================文件夹１==================================**

**|-- BMK\_1\_All\_DEG 所有差异表达circRNA表达量和聚类分析的统计结果**

**|-- All.DEG\_final\_anno.xls 所有差异circRNA表达量+差异信息+Host基因各个数据库注释结果(所有差异circRNA整合文件，信息较全面,方便查看，比较重要)**

|-- #Gene：差异circRNA Host gene ID；

|-- ID：差异circRNA ID；

|-- L\*：样品表达量；

|-- \*\_vs\_\*.Pvalue/FDR：差异Pvalue/FDR值,所有组合都有列出,若无,则用"--"代替；

|-- \*\_vs\_\*.log2FC：差异log2FC值,所有组合都有列出,若无,则用"--"代替；

|-- \*\_vs\_\*.regulated：差异上下调(up/down),若无,则用"--"代替；

|-- Host基因在COG/KOG/GO/KEGG/NR/eggNOG/Swissport各个数据库注释结果。

**|-- All.DEG\_final.pdf 所有差异circRNA表达量聚类热图（PDF格式）**

**|-- All.DEG\_final.png 所有差异circRNA表达量聚类热图（PNG格式）**

**|-- All.DEG\_final\_target.xls**  **所有差异circRNA表达信息+差异信息+Host基因信息整合**

|-- Gene：差异circRNA来源基因ID

|-- ID：差异circRNA id

|-- L\*：差异circRNA 表达量值，所有组合都有列出；若无，则用"--"代替；

|-- \*\_vs\_\*.PValue：差异circRNA Pvalue，所有组合都有列出；若无，则用"--"代替；

|-- \*\_vs\_\*.log2FC：差异circRNA log2FC，所有组合都有列出；若无，则用"--"代替；

|-- \*\_vs\*.regulated：差异circRNA 上下调，所有组合都有列出；若无，则用"--"代替；

**|-- All.DEG\_final.xls 所有差异circRNA表达量+差异信息（重要文件）**

|-- ID： 差异circRNAID；

|-- L\*：样品表达量；

|-- \*\_vs\_\*.Pvalue/FDR：差异Pvalue/FDR值，所有组合都有列出；若无，则用"--"代替；

|-- \*\_vs\_\*.log2FC：差异log2FC值，所有组合都有列出；若无，则用"--"代替；

|-- \*\_vs\_\*.regulated：差异上下调(up/down)，若无，则用"--"代替。

**|-- Veen.pdf 不同组合差异circRNA数目Venn图（2≤差异组合数目＜9），PDF格式**

**|-- Veen.png** **不同组合差异circRNA数目Venn图（2≤差异组合数目＜9），PNG格式**

**=====================================文件夹２==================================**

**|-- BMK\_\*\_\*\_vs\_\*** 表示条件1样品和条件2样品的差异表达分析结果目录

\*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circRNA的表达量/log2FC/Pvalue/FDR**

**|-- \*\_vs\_\*.DEG\_final.Target.xls 差异circRNA文件增加Host基因ID（重要文件）**

**|-- \*\_vs\_\*.DEG\_final.xls 差异circRNA信息（重要文件）**

**|-- \*\_vs\_\*.heatmap.pdf 差异表达circRNA聚类热图（PDF格式）**

**|-- \*\_vs\_\*.heatmap.png 差异表达circRNA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circRNA及circRNA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log10（circRNA+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：log2(FPKM)，即两样品中表达量均值的对数值；纵坐标为M值：log2(FC)，即两样品间circRNA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circRNA,红色的点代表显著上调的circRNA, 黑色的点代表表达差异不显著的circRNA。

**|-- \*\_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circRNA,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circRNA,红色的点代表上调差异表达circRNA，黑色代表非差异表达circRNA。

**注释文件夹BMK\_2\_Anno\_enrichment如下：**

**BMK\_2\_Anno\_enrichment** **差异circRNA的Host基因注释和富集结果**

**注释文件夹下包括三个文件夹：**

**BMK\_1\_Annotation Host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circRNA Host基因 Count+表达量+Pvalue/FDR+log2FC+regulated+各个数据库注释结果（重要文件）**

**|-- \*\_vs\_\*.GO\_enrichment.stat.xls 差异circRNA Host基因GO二级节点注释统计**

|-- #GO\_classify1 GO二级节点ID;

|-- GO\_classify2 GO三级节点名称；

|-- All 注释到该节点的所有基因的数目；

|-- DE 注释到该节点的差异circRNA Host基因的数目。

**|-- \*\_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circRNA Host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circRNA Host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================================**

采用R包clusterProfiler进行GO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 差异circRNA Host基因GO富集条形图(PDF格式)**

**|-- \*\_vs\_\*\_(BP/CC/MF)\_enrich\_barplot.png 差异circRNA Host基因GO富集条形图(PNG格式)**

**|-- \*\_vs\_\*\_(BP/CC/MF)\_enrich\_dotplot.pdf 差异circRNA Host基因GO富集气泡图(PDF格式)**

**|-- \*\_vs\_\*\_(BP/CC/MF)\_enrich\_dotplot.png 差异circRNA Host基因GO富集气泡图(PNG格式)**

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

**|-- \*\_vs\_\*\_(BP/CC/MF)\_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circRNA Host基因的数目/所有差异circRNA Host基因的数目；

|--BgRatio 富集到此节点上的所有基因的数目/所有基因的数目；

|--enrich\_factor 富集因子；

|--pvalue Pvalue；

|--qvalue Qvalue；

|--geneID 富集到此条目上的差异circRNA Host基因的ID；

|--gene\_symbol 对应的Name。

**================================BMK\_3\_KEGG\_enrich================================**

采用R包clusterProfiler进行KEGG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circRNA Host基因KEGG富集条形图**

横坐标代表差异表达circRNA Host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circRNA Host基因KEGG富集点图**

**|-- \*\_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circRNA Host基因数目/所有差异circRNA Host基因数目；

|--BgRatio 富集到此通路上的所有基因的数目/所有基因的数目；

|--enrich\_factor 富集因子；

|--pvalue Pvalue；

|--qvalue Qvalue；

|--geneID 富集到此通路上的差异circRNA Host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。