**Gene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**  **差异LncRNA数目统计结果**

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

**|-- BMK\_1\_All\_DEG**

**|-- All.DEG\_final\_anno.xls** **所有差异基因表达信息+差异信息+注释信息(整合文件,信息全面,使用方便,比较重要)**

|-- #ID：差异基因ID

|-- \*：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

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

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

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

横坐标代表样品名称及样品的聚类结果，纵坐标代表的差异基因及其聚类结果。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（gene+0.000001），表达量的高低可根据色度条判断。

**|-- All.DEG\_final.xls** **所有差异gene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≤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≤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3==================================**

**|-- BMK\_PPI 差异gene蛋白互作网络分析**

**|-- \*\_vs\_\*.DEG.detail.xls 互作关系详细信息文件（直接从STRING数据库中调取）**

|-- #Query\_id1

|-- Type

|-- Query\_id2

|-- Subject\_id1

|-- Subject\_id2

|-- Mode

|-- Score

**|-- \*\_vs\_\*.ppi.cytoscapeInput.sif** **可直接导入cytoscape中的文件**

**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** **该组合中所有基因的表达量/log2FC/Pvalue/FDR（重要文件）**

**|-- \*\_vs\_\*.DEG\_final.xls** **差异基因信息（经过FC/FDR or Pvalue筛选后的结果）（重要文件）**

**|-- \*\_vs\_\*.DEG\_cosmic.xls** **癌症基因功能注释（重要文件）**

|-- ENSG\_ID 基因ID

|-- log2FC 差异log2FC

|-- regulated 上下调(up/down)

|-- Description Cosmic中的描述信息

|-- Tumour\_Types(Somatic) 来自上代的突变

|-- Tumour\_Types(Germline) 体细胞突变

**|-- \*\_vs\_\*.DEG\_TF.xls** **转录因子注释（重要文件）**

|-- ENSG\_ID

|-- log2FC

|-- regulated

|-- Family 转录因子家族

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

**|-- \*\_vs\_\*.heatmap.png 差异表达基因聚类热图（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mRNA及mRNA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（mRNA+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gene。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gene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gene,红色的点代表显著上调的gene, 黑色的点代表表达差异不显著的gene。

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

**注释文件夹BMK\_2\_Anno\_enrichment如下：**

**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**

**BMK\_4\_GSEA GSEA分析结果**

**详细内容如下:**

**=================================BMK\_1\_Annotation================================**

**|-- \*\_vs\_\*.annotation.xls 差异基因Count+FPKM+Pvalue/FDR+log2FC+regulated+各个数据库注释结果（重要文件）**

**|-- \*\_vs\_\*.GO\_enrichment.stat.xls 差异基因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================================**

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** **差异表达基因GO富集图**

横坐标为GeneRatio即注释在该条目中的感兴趣基因占所有差异基因数的比例,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靶基因的数目/所有差异靶基因的数目

|--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靶基因的ID

|--gene\_symbol 对应的Name

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

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gene 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gene 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gene的数目/所有差异gene的数目；

|--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gene的ID；

|--gene\_symbol 对应的Name。

**|-- \*\_vs\_\*.Phase.png** **差异表达基因KEGG通路富集散点图（PNG格式）**

图中每一个图形表示一个KEGG通路，通路名称见右侧图例。横坐标为富集因子（Enrichment Factor），表示差异基因中注释到该通路的基因比例与所有基因中注释到该通路的基因比例的比值。富集因子越大，表示差异表达基因在该通路中的富集水平越显著。纵坐标为-log10(Q-value)，其中Q-value为多重假设检验校正之后的P-value。因此，纵坐标越大，表示差异表达基因在该通路中的富集显著性越可靠。

**|-- \*\_vs\_\*.Phase.tiff 差异表达基因KEGG通路富集散点图（TIFF格式）**

**|--** **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。

**=================================BMK\_4\_GSEA====================================**

GSEA的分析结果，选择富集的前5个GO节点/pathway做富集图。

**|-- \*\_vs\_\*\_(BP/CC/MF)\_GO\_\*.preranked.pdf GSEA preranked图（PDF格式）**

**|-- \*\_vs\_\*\_(BP/CC/MF)\_GO\_\*.preranked.png GSEA preranked图（PNG格式）**

图中横坐标代表排序后的基因集的位置信息，纵坐标代表打分信息。每一条竖线代表感兴趣基因集合中的基因，竖线的高低代表打分的高低。

**|-- \*\_vs\_\*\_(BP/CC/MF)\_GO\_\*.runningScore.pdf GSEA runningScore图（PDF格式）**

**|-- \*\_vs\_\*\_(BP/CC/MF)\_GO\_\*.runningScore.png GSEA runningScore图（PNG格式）**

图中横坐标代表排序后的基因集的位置信息，纵坐标为动态的富集打分。图形顶部的竖线代表感兴趣基因集合中的基因。绿色的曲线代表每个位置上基因集合的富集打分。

**|-- \*\_vs\_\*\_KEGG\_pathway\_ko\*.preranked.pdf**

**|-- \*\_vs\_\*\_KEGG\_pathway\_ko\*.preranked.png**  **同GO**

**|-- \*\_vs\_\*\_KEGG\_pathway\_ko\*.runningScore.pdf**

**|-- \*\_vs\_\*\_KEGG\_pathway\_ko\*.runningScore.png** **同GO**

**|-- \*\_vs\_\*\_(BP/CC/MF)\_GSEA.xls GO节点GSEA分析结果（此文件已经按照qvalue列升序排序）**

|-- ID GO节点ID

|-- Description 节点描述

|-- setSize 基因个数

|-- enrichmentScore 富集分数

|-- NES

|-- pvalue

|-- qvalue

|-- rank

|-- core\_enrichment 基因ID

**|-- \*\_vs\_\*\_KEGG\_pathway\_GSEA.xls KEGG通路GSEA富集结果（同GO）**