

## BMK\_1\_Statistics\_Visualization 差异基因的筛选情况

|-- \*\_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,当物种是人则有该文件)

|-- ENSG\_ID 基因 ID  
|-- log2FC 差异 log2FC  
|-- regulated 上下调(up/down)  
|-- DescriptionCosmic 中的描述信息  
|-- Tumour\_Types(Somatic) 体细胞突变的肿瘤类型  
|-- Tumour\_Types(Germline) germline 突变的肿瘤类型

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

|-- ENSG\_ID 基因 ID  
|-- log2FC 差异 log2FC  
|-- regulated 上下调(up/down)  
|-- Family 转录因子家族

|-- \*\_vs\_\*.heatmap.pdf/png 差异表达基因聚类热图

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基因的聚类结果。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  $\log_{10}(\text{gene}+0.000001)$ 。

|-- \*\_vs\_\*.MA.pdf/png 差异表达 MA 图

差异表达基因 MA 图中每一个点代表一个 gene。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 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/png 差异表达火山图

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: GO 注释系统包含三个主要分支,即:生物学过程(Biological Process,以下简称 BP)、分子功能(Molecular Function,即 MF)和细胞组分(Cellular Component,即 CC)。

## BMK\_2\_Anno\_enrichment 差异基因功能注释及富集分析

BMK\_1\_Anno 差异基因注释分析

|-- \*\_vs\_\*.annotation.xls 差异基因在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.list 差异基因所注释的 GO 条目，第一列为基因，第二列为 symbol，第三列及以后为该基因所注释的 GO 条目。

|-- \*\_vs\_\*.GO.pdf/png GO 注释分类统计图

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\_Enrichment 差异表达基因富集分析

|-- \*\_vs\_\*\_\*\_enrich\_barplot.pdf/png 差异表达基因 GO/KEGG 富集柱状图

横坐标为注释在该条目中的感兴趣基因数，纵坐标每一个 BP/CC/MF/KEGG 条目。颜色代表超几何检验的校正后的 p 值。

|-- \*\_vs\_\*\_\*\_enrich\_dotplot.pdf/png 差异表达基因 GO/KEGG 富集点图

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基因数的比例，纵坐标每一个 BP/CC/MF/KEGG 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\_\*\_\*\_enrich.list.xls 差异表达基因 GO/KEGG 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基因的 ID  
|--gene\_symbol 对应的 Name

|-- \*\_vs\_\*.Phase.tiff/png 差异表达基因 KEGG 富集图

注：图中每一个行代表一个 KEGG 通路。横坐标为富集因子，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。点的颜色代表 qvalue，点的大小代表注释在该通路中差异基因的数目，点的形状代表该通路中差异基因的类型（如：只包含上调基因，只包含下调基因，同时包含上下调基因）。

|-- \*\_vs\_\*.KEGG.list 差异表达基因 KEGG 富集文件

该文件前几列同上述富集文件，最后两列分别为 gene\_number 即注释在该通路中的差异基因数，Contained 为该通路中包含的差异基因类型。

## BMK\_3\_KEGG\_map                  差异表达基因所注释的通路图

|-- 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\_3\_GSEA                  GSEA 分析

|-- result                  BP/CC/MF/KEGG 功能富集的结果图

|-- \*\_vs\_\*.preranked.pdf/png                  GSEA preranked 图

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\_\*.runningScore.pdf/png                  GSEA runningScore 图

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\_\*.GSEA.xls          GO/KEGG 的 GSEA 分析结果

|-- ID                          GO 条目/KEGG 通路 ID  
|-- Description              描述信息  
|-- setSize                  所包含的基因个数  
|-- enrichmentScore        富集打分  
|-- NES                      NES 打分  
|-- pvalue                   pvalue  
|-- qvalue                   qvalue  
|-- rank                      ES 打分最大时所在的基因位置  
|-- core\_enrichment        核心富集的基因 ID