

+-----+  
| 转录组项目信息分析结果说明文档 |  
+-----+

目录结构及文件说明:

\*\*\*\*\*

```
-- BMK_4_Unigene_Anno          #Unigene 功能注释分析结果目录
|  |-- All_Database_annotation.xls  #功能注释整合文件 (excel 格式)
|  |-- Function_Annotation.stat.xls  #Unigene 在各个数据库注释结果统计文件
|  |-- BMK_1_annotation          #各个数据库详细注释结果目录
|  |-- BMK_2_statistic           #各个数据库统计图目录
|  |-- BMK_3KEGG_map             #注释到的 KEGG 通路图目录
|  |      |-- ko00010.png
|  |      |-- ... ..
|  |      `-- ko04626.png
```

=====文件夹=====

BMK\_1\_annotation :各个数据库详细注释结果目录,里面包含里各个数据库比对结果的详细信息。

BMK\_2\_statistic:基因在各个数据库统计结果和作图结果目录

BMK\_3\_Kegg\_map:基因所在的所有通路图。

#####

#文件: Function\_Annotation.stat.xls

#描述: Unigene 在各个数据库注释结果统计文件

#第一列 Anno\_Database: 注释的数据库

#第二列 Annotated\_Number: 此数据库注释的 Unigene 个数

#第三列 300<=length<1000: 此数据库注释的长度在 300 到 1000bp 的 Unigene 个数

#第四列 length>=1000: 此数据库注释的长度大于等于 1000bp 的 Unigene 个数

注: 第二列数字不等于第三列和第四列之和, 因为没有考虑小于 300bp 的 Unigene 个数

#####

# 文件: All\_Database\_annotation.xls

# 描述: 功能注释、富集和统计整合文件 (6 个 sheets)

# 字段解释:

#####sheet

# INTegrated\_Function.anno: 所有基因的注释信息

# #GeneID: 基因名称

# COG\_class: COG 库注释的功能分类

# COG\_class\_annotation: COG 库功能分类描述

# GO\_annotation: GO 注释

# KEGG\_annotation: KEGG 库注释

# KOG\_class: KOG 库注释的功能分类

# KOG\_class\_annotation: KOG 库功能分类描述

# Swissprot\_annotation: Swissprot 库注释

# NR\_annotation: NR 库注释

```

# Pfam_annotation: Pfam 库注释
#####sheet
# Function_anno.stat: 注释结果统计
# #Anno_Database: 注释用的数据库
# Annotated_Number: 注释到的基因数目
# 300<=length<1000: 注释到的基因长度在 300 到 1000 的数目
# length>=1000: 注释到的基因长度大于等于 1000 的基因数目
#####sheet
# GO.list: 每个基因注释到 GO 数据库上的编号
#####sheet
# GO_tree.stat: 对于各个 GO 功能分类, 注释到的基因统计
# #GO_Function: GO 功能分类
# Unigene_number: 注释到的基因数目
# Unigene_ID: 注释到的基因名称
#####sheet
# Kegg.pathway: Kegg 上注释到的基因信息
# #pathway: pathway 名称
# pathway_id: pathway 编号 (一个 ko 编号表示一个通路)
# Gene_number: 注释到某 pathway 上的基因数目
# Gene_id: 注释到某 pathway 上的基因名称
# KOs: KO 分类信息 (一个代谢通路上有哪些 KO 分类, 它是蛋白质 (酶) 的一个分类体系, 序列高度相似, 并且在同一条通路上有相似功能的蛋白质被归为一组, 然后打上 K 标签)
#####sheet
# Kegg.ko: 每个基因的 kegg 注释信息
# #Gene_id: 基因编号
# KO|e_value|Database_Genes|Anno: 四列值, 分别表示基因的注释到的 K 编号, blastx 比对上基因 E 值, 比对上的基因 ID 以及它的功能注释
#####

```