

RNA-Seq分析流程V0.1.0版结果说明

一级结构

每个目录均有 `projectName_xxx` 组成

分析目录下共有 $n+3$ 个子目录，其中 n 为分析的样本数，如下图所示中的项目名+样本名的目录即为单个样本分析结果目录，剩下的三个目录分别为“项目名+DIFF”，“项目名+COEXP”，“项目名+WORKDIR”

- 项目名+DIFF：存放差异表达分析结果
- 项目名+COEXP：存放共表达模块分析结果
- 项目名+WORKDIR：存放多样本分析中间数据及部分结果

```
[chenyl@bj-ngs-run pipeline_test]$tree -L 1
.
├── AcellsTest_2S-A
├── AcellsTest_2S-D
├── AcellsTest_COEXP
├── AcellsTest_DIFF
├── AcellsTest_DIFFpheatmap.pdf
├── AcellsTest_S-A
├── AcellsTest_S-D
└── AcellsTest_WORKDIR
```

tree1

二级结构

单样本

下面分为：

- QC结果
- 比对结果
- reconstruct结果
- 变异检测结果
- 变异检测注释结果

- workdir-中间数据目录

```
└─ AcellTest_2S-A
   └─ ALIGN
   └─ ANNO
   └─ QC
   └─ RECONSTRUCT
   └─ VARIANT
   └─ WORKDIR
```

tree2.1

WORKDIR

主要包括差异表达分析中间数据

```
└─ AcellTest_WORKDIR
   └─ deseq2
   └─ logs
```

tree2.2

COEXP

```
└─ AcellTest_COEXP
   └─ AllGeneCluster.csv
   └─ AllGeneCoExpress.pdf
   └─ DESeqCoExpressHeatMap.pdf
   └─ DESeqGeneCluster.csv
   └─ DESeqGeneCoExpress.pdf
   └─ Rplots.pdf
```

tree2.3

- `*Cluster.csv` : 表示全基因或差异表达基因共表达模块分类结果，结果第一列表示基因名，第二列表示基因所属类别
- `*CoExpress.pdf` : 表示全基因或者差异表达基因表达模块分类图示
- `*HeatMap.pdf` : 表示差异表达基因表达模块热图

三级结构

展示结果如下：

```
— AcellsTest_S-D
  — ALIGN
    — hisat2
  — ANNO
  — QC
    — S-D.clean.R1.fastq.gz
    — S-D.clean.R2.fastq.gz
    — S-D.cleanUp.R1.fastq.gz
    — S-D.cleanUp.R2.fastq.gz
    — S-D_H75N5DMXX_L1_1.fq_fastqc
    — S-D_H75N5DMXX_L1_1.fq_fastqc.html
    — S-D_H75N5DMXX_L1_1.fq_fastqc.zip
    — S-D_H75N5DMXX_L1_2.fq_fastqc
    — S-D_H75N5DMXX_L1_2.fq_fastqc.html
    — S-D_H75N5DMXX_L1_2.fq_fastqc.zip
  — RECONSTRUCT
    — stringtie
  — VARIANT
  — WORKDIR
    — fastqc.trim.log
    — gatk
    — hisat2
    — logs
    — stringtie
```

重要结果

- 单样本质控文件：[项目名_样本名/QC/*html](#)
- 单个样本比对情况：[项目名_样本名/WORKDIR/stringtie/样本名/gene_abund.tab](#)
- 差异表达基因list：[项目名_DIFF/deseq2/样本1,样本2-样本3,样本4/deseq2_res.tab](#)
- 差异表达基因热图：[项目名_DIFF/deseq2/样本1,样本2-样本3,样本4/pheatmap.pdf](#)
- 差异表达基因共表达模块（基因类别）：[项目名_COEXP/DESeqGeneCluster.csv](#)
- 差异表达基因共表达模块（图示模块）：[项目名_COEXP/DESeqGeneCoExpress.pdf](#)
- 差异表达基因共表达模块热图：[项目名_COEXP/DESeqCoExpressHeatMap.pdf](#)

说明

1. 需要说明差异基因的结果，及模块基因的结果如果要进行功能富集，需要手动使用[DAVID](#)进行处理。