**基迪奥生物科技有限公司**

**趋势分析结果说明文档**

目录

[1 趋势分析方法原理详细描述 2](#_Toc413835913)

[2 基因趋势分析结果文件说明 2](#_Toc413835914)

[2.1 趋势0的分析结果（文件夹） 3](#_Toc413835915)

[2.2 所有趋势基因RPKM统计表（all\_rpkm\_annot.xls） 5](#_Toc413835916)

[2.3 所有趋势基因对应的趋势基因集统计表（All-Profile.xls） 6](#_Toc413835917)

[2.4 各pathway基因在各个趋势中分布的统计（all\_pathway.xls）； 6](#_Toc413835918)

[2.5 0趋势内的所有基因整体展示图 7](#_Toc413835919)

[2.6 按基因数目展示所有趋势 8](#_Toc413835920)

[2.7 按趋势-P值展示所有趋势 8](#_Toc413835921)

[3 miRNA趋势分析结果文件说明 9](#_Toc413835922)

[3.1 所有趋势miRNA TPM统计表（all\_TPM\_annot.xls） 9](#_Toc413835923)

[3.2 所有趋势miRNA对应的趋势miRNA集统计表（all\_profile.xls） 10](#_Toc413835924)

[3.3 0趋势内的所有miRNA整体展示图 10](#_Toc413835925)

[3.4 按miRNA数目展示所有趋势 11](#_Toc413835926)

[3.5 按趋势-P值展示所有趋势 12](#_Toc413835927)

[3 STEM的参数 12](#_Toc413835928)

[4 RNA-seq与miRNA趋势关联 12](#_Toc413835929)

[4.1 RNA-seq与miRNA趋势关联分析方法原理详细描述 12](#_Toc413835930)

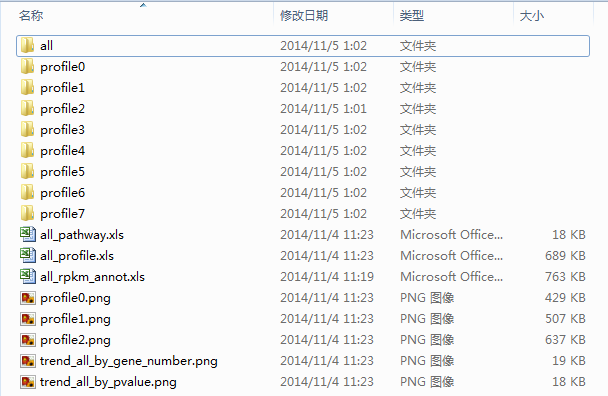
[4.2 结果文件说明 13](#_Toc413835931)

[4.3 附录-斯皮尔曼等级相关 13](#_Toc413835932)

# 趋势分析方法原理详细描述

趋势分析是针对多个连续型样本（至少3个）的特点（样本间包含特定的时间、空间或处理剂量大小顺序）而对基因或miRNA的表达模式（在多阶段中表达曲线的形状）进行聚类的方法。然后从聚类结果中挑选符合一定生物学特性（如表达量持续上升）的基因集或miRNA。我们使用软件STEM—Short Time-series Expression Miner(<http://www.cs.cmu.edu/~jernst/stem>) 输入一个包含每个样品中的基因或miRNA表达量(按生物学逻辑将样品顺序排好)的文件，然后选择参数，进行趋势分析。

# 基因趋势分析结果文件说明

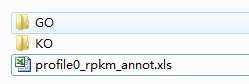


**文件及其含义**

|  |  |
| --- | --- |
| 1．all | 所有趋势合并的分析结果（文件夹） |
| profile0 | 趋势0的分析结果（文件夹） |
| profile1 | 趋势1的分析结果（文件夹） |
| 2．all\_pathway.xls | 各pathway的基因在各趋势中分布的统计总表 |
| 3．All-Profile.xsl | 所有差异基因对应的趋势基因集统计表 |
| 4．all\_rpkm\_annot.xls | 所有差异基因RPKM统计表 |
| 5．profile0.png | 0趋势内的所有基因整体展示图 |
| profile1.png | 1趋势内的所有基因整体展示图 |
| profile2.png | 2趋势内的所有基因整体展示图 |
| 6．trend\_all\_by\_gene\_number.png | 按基因数目展示所有趋势 |
| 7．trend\_all\_by\_significance.png | 按趋势-P值展示所有趋势 |
|  |  |

**详细说明**

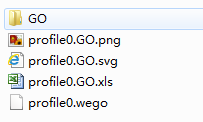
## 趋势0的分析结果（文件夹）



打开该文件夹，路径：根目录/RNA\_Trend\_analysis/profile0

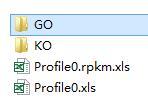


|  |  |
| --- | --- |
| 1．GO | 0趋势中差异基因GO功能注释(文件夹) |
| 2．KO | 0趋势中差异基因KEGG pathway注释（文件夹） |
| 3．Profile0.rpkm\_annot.xsl | 0趋势中差异基因RPKM统计表 |



打开该文件夹，

路径：根目录/RNA\_Trend\_analysis /profile0/GO



SVG、PNG文件：0号趋势GO分析整体图

Xls文件：0号趋势GO分类详细列表

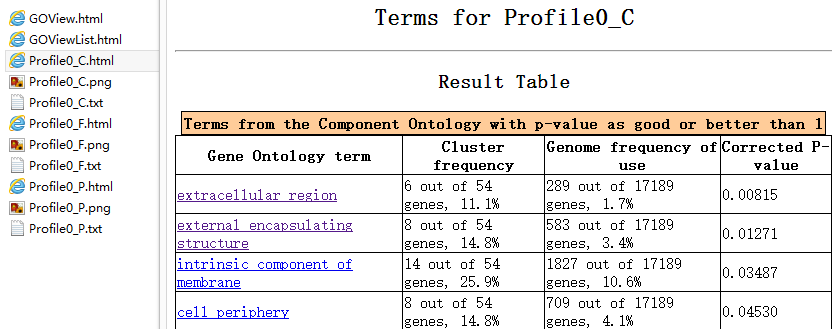
GO文件夹：0号趋势网页版列表及细胞组分（CellularComponent）、生物进程（BiologicalProcess）、分子功能（MolecularFunction）展示图

GO结果查看

打开该文件夹GO，

路径：RNA\_Trend\_analysis/profile0/GO/GO

/profile0/GO



0号趋势中可以注释到

GO数据库的基因及

各个term中的基因数

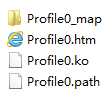
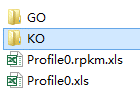
所有基因中可以注释到GO数据库的基因及各个term中的基因数

KO文件夹：0号趋势各pathway网页版及展示图

ko、path文件：0号趋势KEGG pathway分析统计表

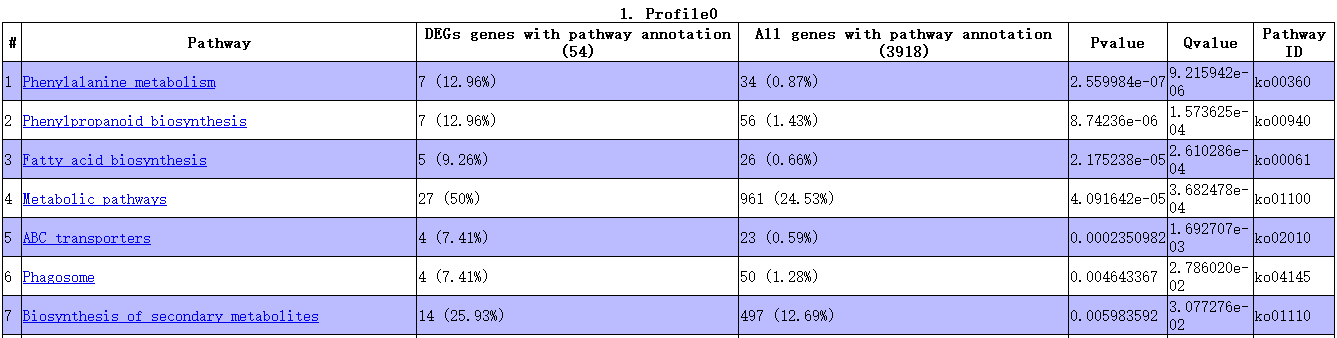
profile0.html：网页版列表

pathway结果查看



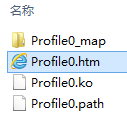
打开该文件夹，

路径：RNA\_Trend\_analysis/profile0/KO



0号趋势中可以注释到KEGG数据库的基因及该pathway中的基因数

所有基因中可以注释到KEGG数据库的基因及各个pathway中的基因数



打开该文件夹KO，

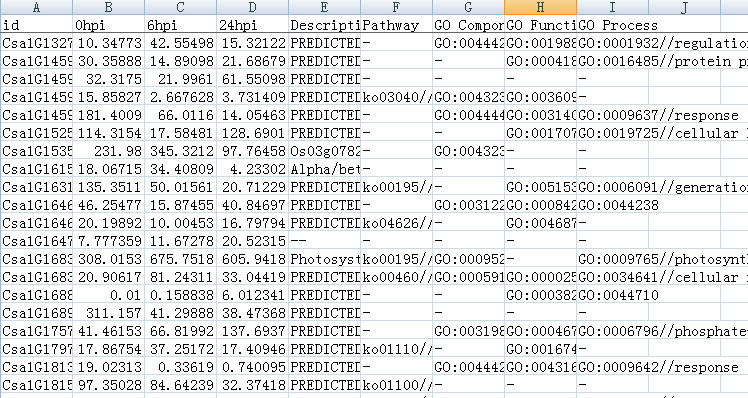
路径：RNA\_Trend\_analysis/profile0/KO/profile0.html

/profile0/GO

## 所有趋势基因RPKM统计表（all\_rpkm\_annot.xls）

路径：根目录/ RNA\_Trend\_analysis/all\_rpkm\_annot.xls

包含所有趋势中基因的表达量以及注释；



## 所有趋势基因对应的趋势基因集统计表（All-Profile.xls）

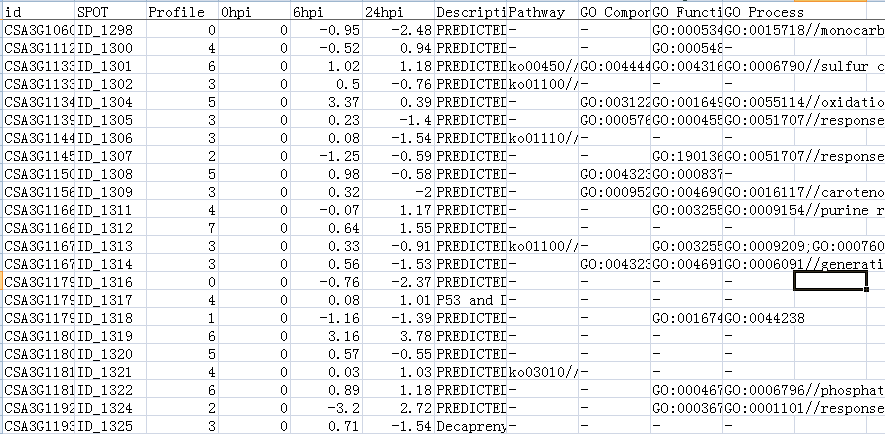
路径：根目录/ RNA\_Trend\_analysis/All-Profile.xls

第1列：gene ID;

第2列：点的序号（无特定意义）

第3列：基因所在的趋势ID（例如，基因属profile 0 ，则这里的趋势ID就是0）

第3~最后：分别为基因在各个点归一化后表达量值以及基因的功能注释；



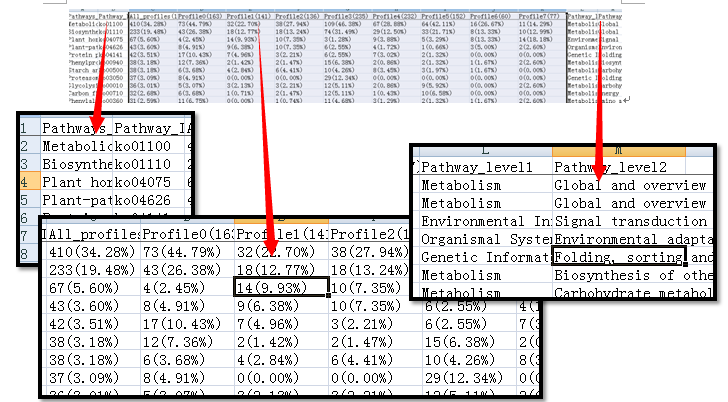
## 各pathway基因在各个趋势中分布的统计（all\_pathway.xls）；

路径：根目录/ RNA\_Trend\_analysis/all\_pathway.xls

第1,2列：pathway的名称和ID（备注：我们针对KEGG数据库Pathway分类的第三级进行分类）

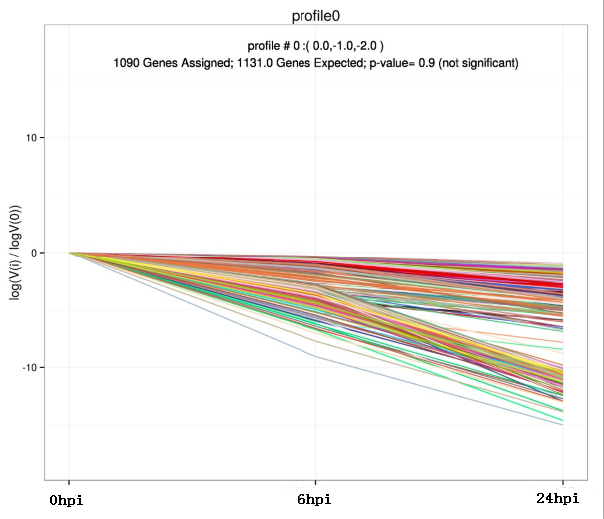
第3~X列：分别为该pathway的基因在所有趋势以及各个子趋势中的分布个数，以及这些基因在对应趋势中所占的比例；

最后两列：这个pathway（第三级）所属的第1级和第二级pathway。



## 0趋势内的所有基因整体展示图

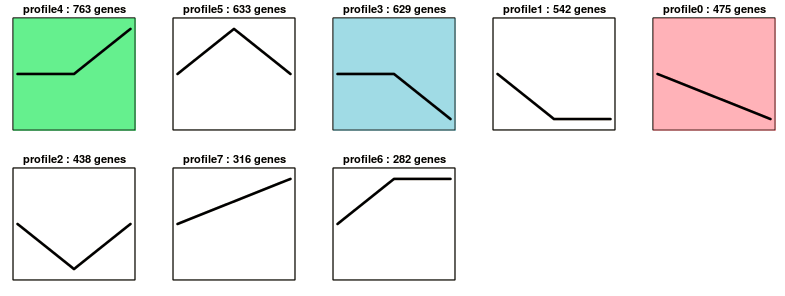
路径：根目录/ RNA\_Trend\_analysis/Profile0.png



说明：从这张图中，可直观看出属于该趋势的基因的表达变化情况，以及聚集的紧密程度，图中每一条线代表一个基因，横坐标为样品的编号，纵坐标为。

## 按基因数目展示所有趋势

路径：根目录/ RNA\_Trend\_analysis/trend\_all\_by\_gene\_number.png

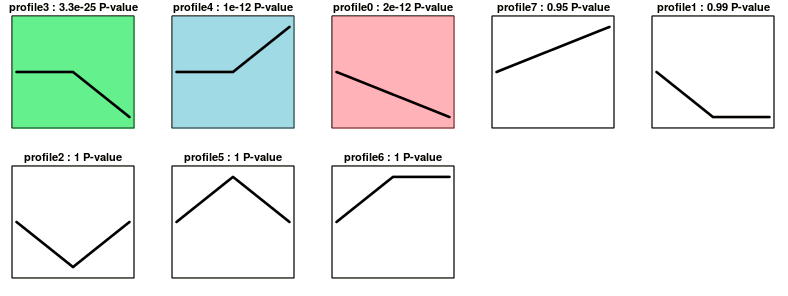


说明：

* 图上方为趋势的ID和趋势中的基因数量；
* 带颜色的趋势块：显著富集的趋势，不同的颜色是软件为了将不同的趋势区分开，而设置的，没有特殊意义
* 不带颜色的趋势块：非显著富集的趋势

## 按趋势-P值展示所有趋势

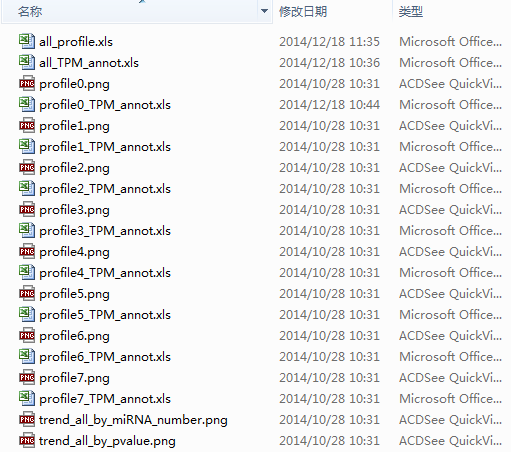
路径：根目录/ RNA\_Trend\_analysis/trend\_all\_by\_significance.png



说明：

* 图上方为趋势的ID和趋势中的基因数量
* 带颜色的趋势块：显著富集的趋势，趋势相似的趋势块颜色相同
* 不带颜色的趋势块：非显著富集的趋势

# miRNA趋势分析结果文件说明



**文件及其含义**

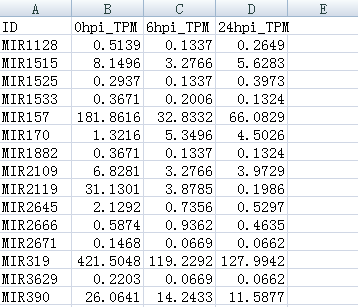
|  |  |
| --- | --- |
| 1．all\_profile.xls | 所有差异miRNA对应的趋势miRNA集统计表 |
| 2．all\_TPM\_annot.xls | 所有差异miRNA的TPM（表达量）统计表 |
| 3．profile0\_TPM\_annot.xls | 0趋势内所有miRNA的TPM（表达量）统计表 |
| 4．profile0.png | 0趋势内的所有miRNA整体展示图 |
| profile1.png | 1趋势内的所有miRNA整体展示图 |
| 5．trend\_all\_by\_miRNA\_number.png | 按miRNA数目展示所有趋势 |
| 6．trend\_all\_by\_significance.png | 按趋势-P值展示所有趋势 |

**详细说明**

## 所有趋势miRNA TPM统计表（all\_TPM\_annot.xls）

路径：根目录/ miRNA\_Trend\_analysis/all\_TPM\_annot.xls

包含所有趋势中miRNA的表达量；



## 所有趋势miRNA对应的趋势miRNA集统计表（all\_profile.xls）

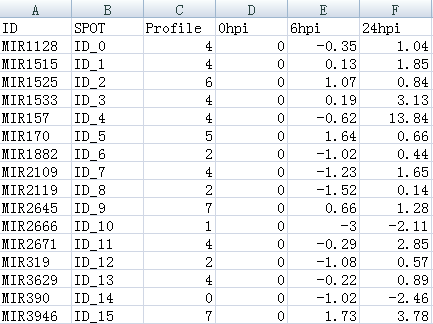
路径：根目录/ miRNA\_Trend\_analysis/all\_profile.xls

第1列：gene ID;

第2列：点的序号（无特定意义）

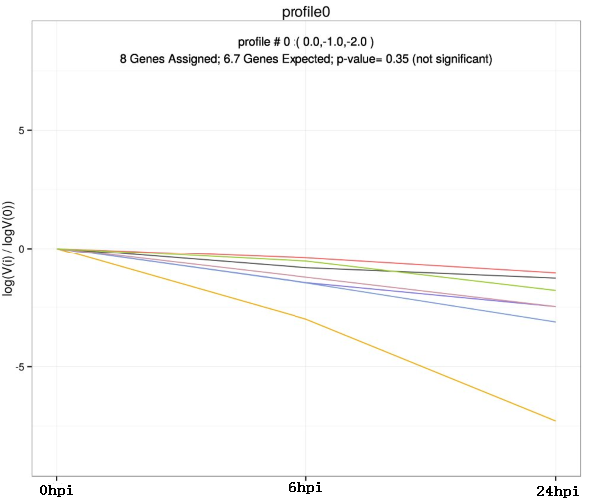
第3列：miRNA所在的趋势ID（例如，miRNA属profile 0 ，则这里的趋势ID就是0）

第3~最后：分别为miRNA在各个点归一化后表达量值；



## 0趋势内的所有miRNA整体展示图

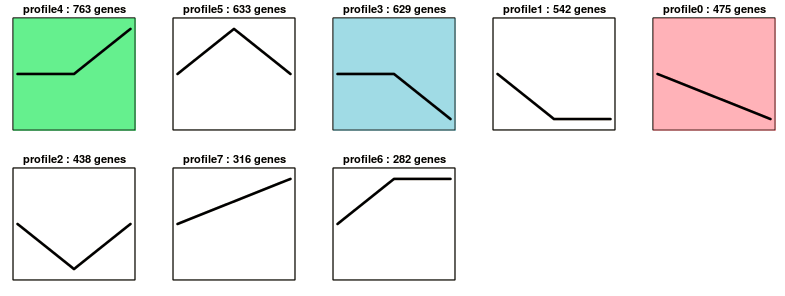
路径：根目录/ miRNA\_Trend\_analysis/Profile0.png



说明：从这张图中，可直观看出属于该趋势的miRNA的表达变化情况，以及聚集的紧密程度，图中每一条线代表一个miRNA，横坐标为样品的编号，纵坐标为。

## 按miRNA数目展示所有趋势

路径：根目录/ miRNA\_Trend\_analysis/trend\_all\_by\_miRNA\_number.png

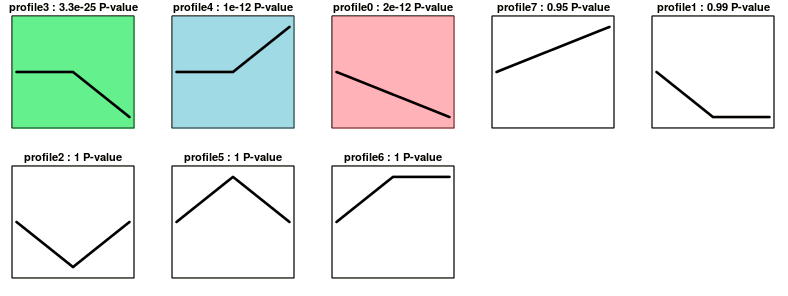


说明：

* 图上方为趋势的ID和趋势中的miRNA数量；
* 带颜色的趋势块：显著富集的趋势，不同的颜色是软件为了将不同的趋势区分开，而设置的，没有特殊意义
* 不带颜色的趋势块：非显著富集的趋势

## 按趋势-P值展示所有趋势

路径：根目录/ miRNA\_Trend\_analysis/trend\_all\_by\_significance.png



说明：

* 图上方为趋势的ID和趋势中的miRNA数量
* 带颜色的趋势块：显著富集的趋势，趋势相似的趋势块颜色相同
* 不带颜色的趋势块：非显著富集的趋势

# STEM的参数

均一化参数：Log normalize data

Maximum Unit change in Model profiles between Time points: 1

Maximum Number of Model Profiles: 最终得到的趋势个数（例如设定为8，则最终得到8个趋势）

其他参数：全部使用软件默认值

# RNA-seq与miRNA趋势关联

## RNA-seq与miRNA趋势关联分析方法原理详细描述

我们使用软件STEM—Short Time-series Expression Miner(<http://www.cs.cmu.edu/~jernst/stem>)得到差异基因和差异miRNA的两组趋势结果，我们可将两组趋势关联分析。挑选具有靶向关系的差异基因和差异miRNA，运用斯皮尔曼等级相关法计算这些miRNA-靶基因对在其两组趋势中的相关性，得到相关系数。因miRNA常对其靶向mRNA进行负调控，我们将斯皮尔曼等级相关系数小于等于-0.5的miRNA-靶基因对列出。得到可能的负调控miRNA-靶基因对。

## 结果文件说明



**文件及其含义**

|  |  |
| --- | --- |
| 1．miRNA\_Trend\_analysis | 差异miRNA趋势分析结果（文件夹） |
| 2．RNA\_Trend\_analysis | 差异基因趋势分布结果（文件夹） |
| 3．rho\_miRNA\_RNA\_annot .xls | 有靶向关系的差异基因与差异miRNA趋势关联分析统计表 |

**详细说明**

miRNA\_Trend\_analysis（文件夹）与RNA\_Trend\_analysis（文件夹）详细说明参见文件中readme.docx文档；

**有靶向关系的差异基因与差异miRNA趋势关联分析统计表（rho\_miRNA\_RNA\_annot .xls）**

第A、B列：miRNA-靶基因对（miRNA-mRNA-paired），A列为miRNA的ID号，B列为其靶基因ID号；

第C列：斯皮尔曼等级相关系数（rho），rho≤-0.5的全列出；

第D列：miRNA\_ID;

第E列：miRNA\_趋势类型(与差异miRNA趋势分析结果中profile编号对应)；

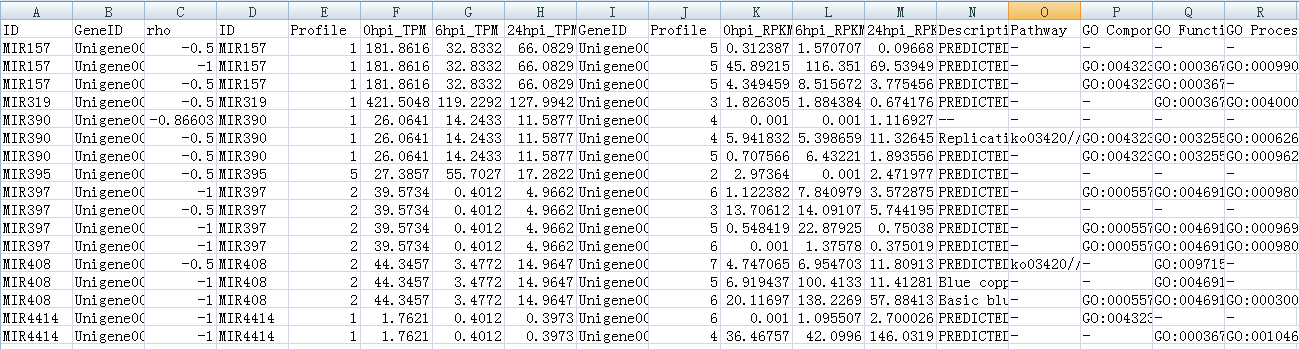
第F、G、H列：miRNA表达量(TPM)；

第I列：mRNA\_ID;

第J列：mRNA\_趋势类型(与差异基因趋势分析结果中profile编号对应)；

第K、L、M列：mRNA表达量(RPKM)；

第N、O、P、Q、R列：mRNA基因注释（依次为Nr、KEGG、GO注释）



## 附录-斯皮尔曼等级相关

斯皮尔曼等级相关（Spearman’s correlation coefficient for ranked data）主要用于解决称名数据和顺序数据相关的问题。适用于两列变量，而且具有等级[线性关系](http://baike.baidu.com/view/91595.htm)的资料。由英国心理学家、统计学家斯皮尔曼根据[积差相关](http://baike.baidu.com/view/8542741.htm)的概念推到而来，一些人把斯皮尔曼等级相关看做积差相关的特殊形式。

c:\users\user\appdata\roaming\360se6\USERDA~1\Temp\9D82D1~1.JPG

n为等级个数

d为二列成对变量的等级差数

简单点说，就是无论两个变量的数据如何变化，符合什么样的分布，我们只关心每个数值在变量内的排列顺序。如果两个变量的对应值，在各组内的排序顺位是相同或类似的，则具有显著的相关性。

举个例子，例如表1的数值，用斯皮尔曼等级相关计算相关系数，将发生如下变化。

表1 斯皮尔曼等级排列

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 样本编号 | 样本1 | 样本2 | 样本3 | 样本4 | 样本5 | 样本6 | 样本7 | 样本8 |
| A表达量 | 0.6 | 0.7 | 1 | 2.1 | 2.9 | 3.2 | 5.5 | 6.7 |
| A 排序等级 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| D表达量 | 6.0E-3 | 2.8E-2 | 1 | 1.7E3 | 4.2E4 | 1.1E5 | 2.5E7 | 1.8E8 |
| D排序等级 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| d(等级差) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

备注：排序等级就是这个数值在组内从小到大排列的序位号。

利用斯皮尔曼等级相关计算A、D基因表达量的相关性，结果是：

*r*=1，p-value = 4.96e-05

这里斯皮尔曼等级相关的显著性显然高于皮尔森相关。这是因为虽然两个基因的表达量是非线性关系，但两个基因表达量在所有样本中的排列顺序是完全相同的，因为具有极显著的斯皮尔曼等级相关性。