

tsv-utils之根据映射合并数值: melt

一、tsv-utils melt介绍

功能描述:

`tsv-utils melt` 根据 `bin` 文件以及元素 (`elements`) 的数值映射表合并新的数值。

命令行接口:

```
1 $ tsv-utils melt
2
3 Usage: tsv-utils melt [options] <bin-tab> <matrix>
4
5 Options:
6   -r round value to nearest intege;
7   -d delimiter for bins, default [:]
```

可选参数:

```
1 -r 四舍五入为整数;
2 -d bin文件的间隔符;
```

二、使用场景实例及其用法

1. 宏转录组数据分析: 根据基因在每个样本的丰度信息以及计算不同水平功能信息(比如 `KEGG KO`)

示例演示

示例文件: `ko.txt`,

`bin` 文件, 2 列, 第二列为元素文件

```
1 $ cat ko.txt | head -n 6
```

```
1 #catalog      members
2 K01948 TRINITY_g108531_i1_1
3 K15578 TRINITY_g436639_i1_1
4 K00360 TRINITY_g317023_i1_2,TRINITY_g83207_i1_1
5 K01672 TRINITY_g220886_i1_1
6 K01743 TRINITY_g299867_i1_1
```

丰度文件,可以为多个样本, Salmon结果文件, 需要提取第1,4列,计算TPM数值,添加'#'符号.

```
1 $ cat quant.sf | head -n 6
```

	Name	Length	EffectiveLength	TPM	NumReads
1	TRINITY_g1_i1_1	354	40.677	0.000000	0.000
2	TRINITY_g2_i1_1	339	30.778	0.000000	0.000
3	TRINITY_g2_i1_2	273	12.091	0.000000	0.000
4	TRINITY_g3_i1_1	456	139.056	0.000000	0.000
5	TRINITY_g4_i1_1	186	40.789	0.000000	0.000

运行命令:

准备输入文件

```
1 $ cut -f1,4 quant.sf | tsv-utils view -c - | tee quant.txt | head -n6
```

```
1 #Name    TPM
2 TRINITY_g1_i1_1 0.000000
3 TRINITY_g2_i1_1 0.000000
4 TRINITY_g2_i1_2 0.000000
5 TRINITY_g3_i1_1 0.000000
6 TRINITY_g4_i1_1 0.000000
```

计算 KEGG KO 丰度信息, 使用 `-d ','` 元素分隔符

```
1 $ tsv-utils melt -d ',' ko.txt quant.txt | head -n 10
```

```
1 #features    TPM
2 K01948  0.00000
3 K15578  0.00000
4 K00360  0.00000
5 K01672  0.00000
6 K01743  0.00000
7 K02567  0.00000
8 K04561  0.00000
9 K00262  0.00000
10 K00362  1.77565
```

使用 `-r` 参数, 将数值转换为整形.

```
1 $ tsv-utils melt -r -d ',' ko.txt quant.txt | head -n 10
2
3 #features    TPM
4 K01948  0
5 K15578  0
6 K00360  0
7 K01672  0
8 K01743  0
9 K02567  0
10 K04561  0
11 K00262  0
12 K00362  2
```

注意事项: 输入数值型数值可以为多个样本数据矩阵.

本文材料为 **BASE (Biostack Applied bioinformatic SEies)** 课程 **Linux Command Line Tools for Life Scientists** 材料, 版权归 上海逻捷信息科技有限公司 所有。

