📖 **README.zh.md**

Transmart Loader

操作指南

1. 下载/创建临床数据, 将 *数据文件* 与 *map文件* 放到服务器上某个路径.
2. 修改并保存 *clinical.params*. (*.params的定义在 *参数列表* 部分说明)
3. 导入临床数据

```
./load_clinical.sh clinical.params
```

4. 检查 gpl 是否已经存在于 tranSMART

```
./check_gpl.sh
```

- 如果gpl已存在于列表, 跳过下面三个步骤

5. (可选) 从transmart dataset 下载注释数据
6. (可选) 修改并保存 *annoataion.params* .
7. (可选) 导入注释

```
./load_annotation.sh annotation.params
```

8. 修改并保存 *expression.params* .
9. 导入分子表达数据

```
./load_expression.sh expression.params
```

Done!

执行脚本

- load_clinial.sh
- load_expression.sh
- load_annotation.sh
- chk_gpl.sh

参数文件

- clinical.params
- expression.params
- annotation.params

参数列表

clinical.params

```
# data
DATA_LOCATION="/home/transmart/datasets/RanchoGSE4698/clinical"
COLUMN_MAP_FILE="Acute_Lymphoblastic_Leukemia_Kirschner_Schwabe_GSE4698_Mapping_File.txt"

# info
STUDY_ID="GSE4698"
TOP_NODE="\\Public Studies\\Acute Lymphoblastic Leukemia_Kirschner_Schwabe_GSE4698"

# security
SECURITY_REQUIRED="N"

# not using
WORD_MAP_FILE=x
RECORD_EXCLUSION_FILE=x
```

Field Name	Meaning
DATA_LOCATION	数据文件夹路径.
COLUMN_MAP_FILE	map文件名.
STUDY_ID	Study id.
TOP_NODE	Top node. (TOP_NODE=\\TOP_NODE_PREFIX\\STUDY_NAME)
SECURITY_REQUIRED	是否保密?
WORD_MAP_FILE	Word map file.
RECORD_EXCLUSION_FILE	Record exclusion file.

expression.params

```
# data
DATA_LOCATION="/home/transmart/datasets/RanchoGSE4698/expression"
DATA_FILE_PREFIX="Acute_Lymphoblastic_Leukemia_Kirschner_Schwabe_GSE4698_Gene_Expression_Data"
MAP_FILENAME=\
"Acute_Lymphoblastic_Leukemia_"\
"Kirschner_Schwabe_GSE4698_"\
"Subject_Sample_Mapping_File.txt"\

# info
STUDY_ID="GSE4698"
TOP_NODE="\\Public Studies\\Acute Lymphoblastic Leukemia_Kirschner_Schwabe_GSE4698"
SOURCE_CD=""

# security
SECURITY_REQUIRED="N"
```

Field Name	Meaning
DATA_LOCATION	分子表达量数据文件夹.
DATA_FILE_PREFIX	数据文件的前缀.
MAP_FILENAME	Map 文件名.
STUDY_ID	Study id.
TOP_NODE	Top node. (TOP_NODE=\\TOP_NODE_PREFIX\\STUDY_NAME)

Field Name	Meaning
SOURCE_CD	需要包含的SOURCE_CD. 与map文件中的_SOURCE_CD_ 字段相关. 默认值: STD
SECURITY_REQUIRED	是否保密?

annotation.params

```
# data
DATA_LOCATION="/home/transmart/datasets/EtriksGSE43696/annotation"
SOURCE_FILENAME="GPL6480.txt"

# info
ANNOTATION_TITLE="Agilent-014850 Whole Human Genome Microarray 4x44K G4112F (Probe Name Version)"
GPL_ID="GPL6480"

# col numbers
PROBE_COL=2
GENE_SYMBOL_COL=3
GENE_ID_COL=4
ORGANISM_COL=5

# header?
SKIP_ROWS=0
```

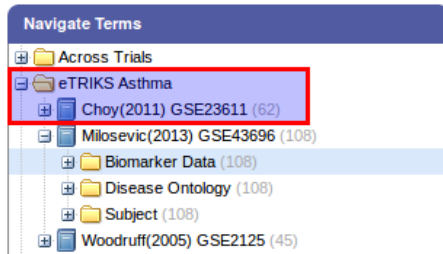| Field Name	Meaning
DATA_LOCATION	注释数据文件夹路径.
SOURCE_FILENAME	数据文件名.
ANNOTATION_TITLE	注释标题. (从下载压缩包的params文件复制)
GPL_ID	GPL id.
PROBE_COL	Column index of the probe ID.
GENE_SYMBOL_COL	Column index of the gene symbol.
GENE_ID_COL	Column index of the gene ID.
ORGANISM_COL	Column index of the organism.
SKIP_ROWS	Number of rows to skip. Note: This script does not assume a header row is present. If a header row exists, this should be set to one.

UI Explain

- TOP_NODE_PREFIX
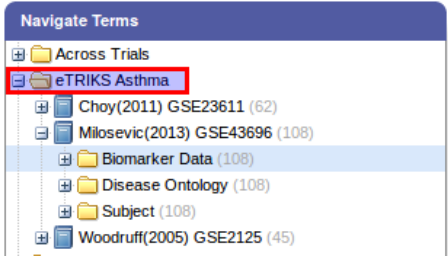
- STUDY_NAME
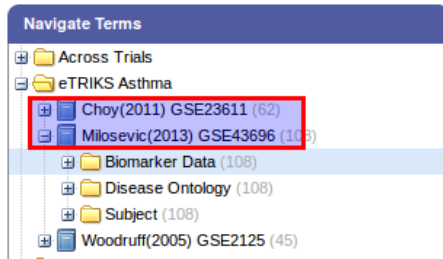
- TOP_NODE (TOP_NODE=\\TOP_NODE_PREFIX\\STUDY_NAME)

- CATEGORY_CD