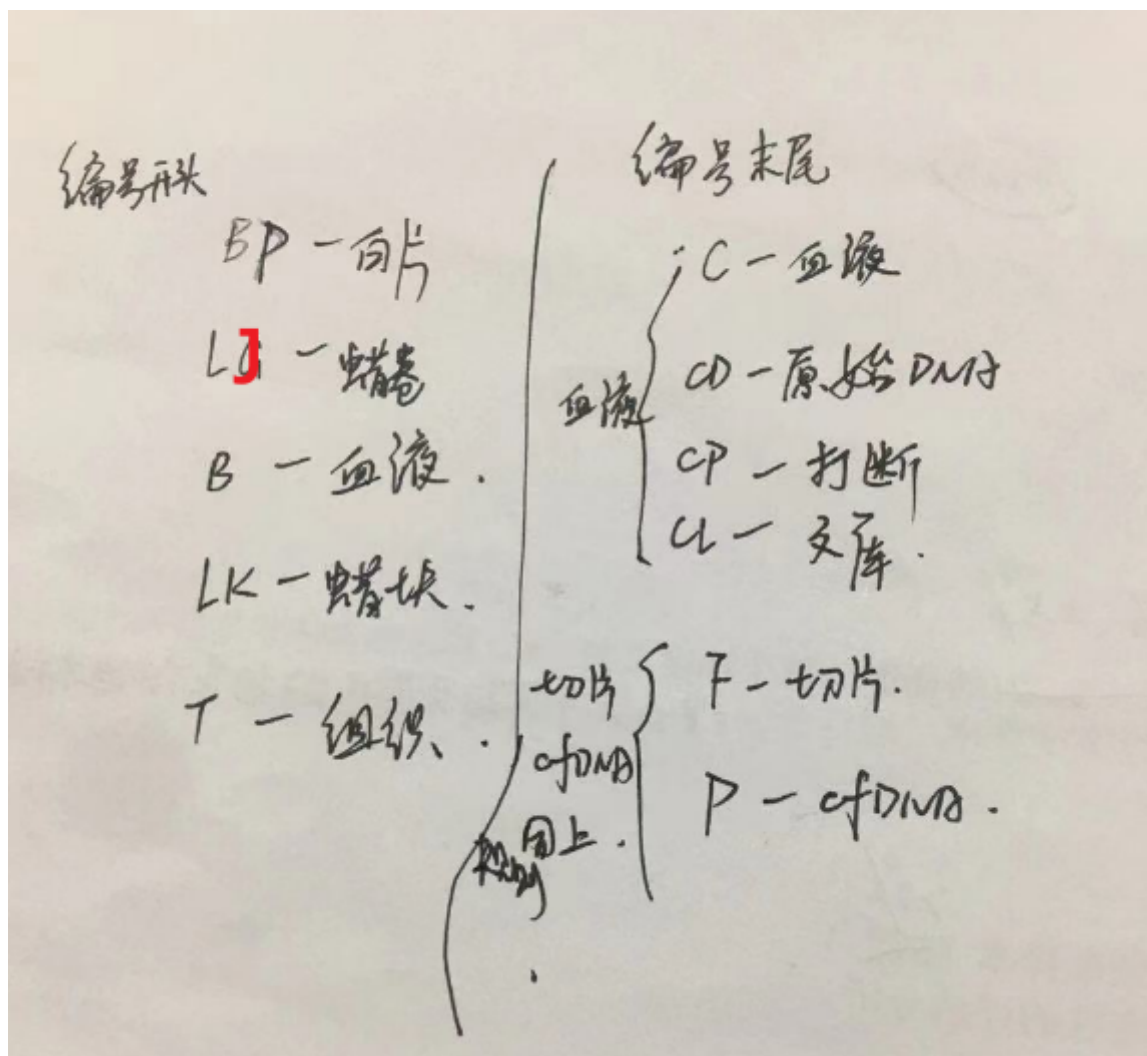


一.准备样本列表(id.txt)

1. YHB、YHH、YHT、YHF这些前缀去掉;

老版编号{ B: 血液 T: 新鲜组织 F: FFPE 切片等 H: 体液, 如腹水, 积液, 尿液等 }

新版编号:



2. 新版编号一般如果是全长编号 (18/19 长度如: B2107020040PL01S01/LJ2107020030FL01S01) 则不需要处理; 如果是短编号(如B2101100038PL01/B2103230008/LJ2103130093)则在**每个编号后面加***, 或者从样本中心表格中**查询完整编号**; 如果给到的是S开头的编号 (如S20012190) 则需要查询对应的生信分析编号 (如B2107020040PL01S01/LJ2107020030FL01S01)
3. 最后保存为如下内容 (注意字符串前后是否有空格) :

```
1 |1912727
2 |1911117
3 |2011593
4 |1911254
5 |1912466
6 |1912172
7 |1912381
8 |1911972
9 |1910260
0 |BP2104020034FL01S01
```

二.抓取样本所属panel(panel.txt)

```
#路径/mnt/vol2_ws/test/wbzhao/panel.sh
sh panel.sh id.txt panel.txt
#          input list output
```

2. 检查生成的panel.txt是否存在错误：一个编号来自多个panel的（例子如下）保留一个需要的panel

```
50gene 1820925
66gene 1823917
66gene 1823955
50gene 1820185
620geneV1 1821939
620geneV1/1822339 grep: /mnt/vol1_root/rendong/project/66gene/1822339 1822339 grep:
/mnt/vol1_root/rendong/project/66gene/1822339/: Is a directory
```

找不到编号（例子如下）的删掉

```
620geneV3 1911285
620geneV3 1910599
620geneV3 1911849
620geneV3 1910322
*/1234567 1234567/: No such file or directory
```

三.抓取xlsx表格

1.

```
#路径/mnt/vol2_ws/test/wbzhao/get_xlsx.sh
sh get_xlsx.sh /mnt/vol1_root/rendong/project panel.txt data
#                               indir      list      out
sh get_xlsx.sh /mnt/vol2_root/rendong/project panel.txt data
```

2. 检查抓到的文件数量与预期是否一致（如下图）：

```
1811006.xlsx
1910038.xlsx
1910043.xlsx
1910101.xlsx
1910149.xlsx
1910177.xlsx
1910211.xlsx
1910223.xlsx
```

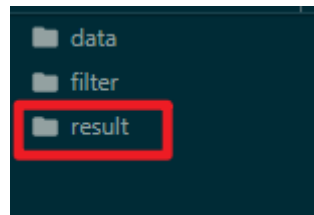
四.抓取cnv文件

1.

```
#路径/mnt/vol2_ws/test/wbzhao/cnv.sh
#如果数据在vol2_root则修改indir路径
perl  cnv.pl  /mnt/vol1_root/rendong/project  cnv_somatic  panel.txt
somatic
#                                     indir      outdir      list
type
perl  cnv.pl  /mnt/vol1_root/rendong/project  cnv_germline  panel.txt
germline
```

2. 检查生成的文件数量与预期是否一致

3. 最终使用的是输出文件夹中result文件夹中的数据（如下图）：



name	last modified		
1811006.somaticCNV.xls	3 days ago	1811006.germlineCNV.xls	3 days ago
1910038.somaticCNV.xls	3 days ago	1910038.germlineCNV.xls	3 days ago
1910043.somaticCNV.xls	3 days ago	1910043.germlineCNV.xls	3 days ago
1910101.somaticCNV.xls	3 days ago	1910101.germlineCNV.xls	3 days ago
1910149.somaticCNV.xls	3 days ago	1910149.germlineCNV.xls	3 days ago
1910177.somaticCNV.xls	3 days ago	1910177.germlineCNV.xls	3 days ago
1910211.somaticCNV.xls	3 days ago	1910211.germlineCNV.xls	3 days ago
1910222.somaticCNV.xls	3 days ago	1910222.germlineCNV.xls	3 days ago