

## U87 in vivo SELEX

文件路径：/mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL

服务器：159.226.118.212

账号：liuzhe

密码：liuzhe123

前期准备：数据质量控制：

文件路径：

/mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-Basic\_operation/StatisticsInformation/PartII

fastqc --noextract \*.fq.gz

(备注：

因为本实验没有加接头序列，并且根据FASTQC的结果可知：并没有接头。所以我们无需去掉Illumina的接头序列)

第一步：使用HISAT2把PE的测序数据比对到人类参考基因组hg19

文件路径：

/mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-Basic\_operation/Step1\_Mapping/PartII

程序代码：

/mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-Basic\_operation/Step1\_Mapping/PartII/Ctrl-P1/Ctrl-P1\_rep1/run.sh

gunzip -c MixC-GATCAG\_HCLJ5CCXY\_L8\_1.fq.gz > MixC-GATCAG\_HCLJ5CCXY\_L8\_1.fq

gunzip -c MixC-GATCAG\_HCLJ5CCXY\_L8\_2.fq.gz > MixC-GATCAG\_HCLJ5CCXY\_L8\_2.fq

hisat2 -5 10 -3 30 -p 5 --dta /home/liuzhe/Genome/hg19/hg19/genome -1 MixC-GATCAG\_HCLJ5CCXY\_L8\_1.fq -2 MixC-GATCAG\_HCLJ5CCXY\_L8\_2.fq -S Ctrl-P1\_rep1.sam

(备注：

根据上面的数据质量控制结果，我们发现5'端前10个碱基以及3'端前30个碱基的数据质量不是很好，因此，我们hisat2中加上两个参数：“-5 10 -3 30”。

hisat2版本：2.0.4

)

| PartII   |              |                                      |               |
|--|--------------|--------------------------------------|---------------|
| Sample name  | Description  | #Rawdata                             | #Mapdata Rate |
| Sample5  | Ctrl-P1_rep1 | read1:1985079<br>7; read2:19850797   | 22.96%        |
| Sample6  | Ctrl-P1_rep2 | read1:3061191<br>1; read2:30611911   | 72.91%        |
| Sample7  | Ctrl-P2_rep1 | read1:1032021<br>27; read2:103202127 | 50.66%        |
| Sample8  | Ctrl-P2_rep2 | read1:3134402<br>6; read2:31344026   | 51.67%        |
| Sample9  | Ctrl-P3_rep1 | read1:2098058<br>4; read2:20980584   | 84.34%        |
| Sample10   | Ctrl-P3_rep2 | read1:2901711<br>1; read2:29017111   | 84.02%        |
| Sample11   | P0_rep1      | read1:3474237<br>4; read2:34742374   | 19.40%        |
| Sample12   | P0_rep2      | read1:2995402<br>7; read2:29954027   | 5.48%         |
| Sample13   | TMZ-P1_rep1  | read1:2435960<br>9; read2:24359609   | 87.14%        |
| Sample14   | TMZ-P1_rep2  | read1:2419692<br>7; read2:24196927   | 76.28%        |
| Sample15   | TMZ-P2_rep1  | read1:2359738<br>2; read2:23597382   | 72.34%        |
| Sample16   | TMZ-P2_rep2  | read1:3050946<br>6; read2:30509466   | 78.12%        |
| Sample17   | TMZ-P3_rep1  | read1:2627481<br>6; read2:26274816   | 87.98%        |
| Sample18   | TMZ-P3_rep2  | read1:2571974<br>7; read2:25719747   | 88.05%        |
| File direction: /mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-Basic_operation/StatisticsInformation/PartII/ |              |                                      |               |
| Software: FastQC v0.10.1   |              |                                      |               |

### 补充：数据可视化 ( UCSC Genome Browser )

samtools view -h -S Ctrl-P1\_rep1.sam -b -o Ctrl-P1\_rep1\_unsorted.bam

samtools sort Ctrl-P1\_rep1\_unsorted.bam Ctrl-P1\_rep1.sort

bedtools genomecov -bg -split -ibam Ctrl-P1\_rep1.sort.bam > Ctrl-P1\_rep1.bdg

LC\_COLLATE=C sort -k1,1 -k2,2n Ctrl-P1\_rep1.bdg > Ctrl-P1\_rep1.sorted.bdg

bedGraphToBigWig Ctrl-P1\_rep1.sorted.bdg

/media/ibm\_disk/work/database/UCSC\_refseq\_annotation/hg19\_known\_gene/hg19.chrom.sizes Ctrl-P1\_rep1.bw

( 可视化Track :

track type=bigWig db=hg19 name="Ctrl-P1\_rep1" visibility=2

bigDataUrl="http://159.226.118.212/home/liuzhe/projects\_cas/xuelab/JL\_GSC\_projects/RNA-seq/PartII\_U87/Ctrl-P1\_rep1.bw"

track type=bigWig db=hg19 name="Ctrl-P1\_rep2" visibility=2

bigDataUrl="http://159.226.118.212/home/liuzhe/projects\_cas/xuelab/JL\_GSC\_projects/RNA-seq/PartII\_U87/Ctrl-P1\_rep2.bw"

track type=bigWig db=hg19 name="Ctrl-P2\_rep1" visibility=2

bigDataUrl="http://159.226.118.212/home/liuzhe/projects\_cas/xuelab/JL\_GSC\_projects/RNA-seq/PartII\_U87/Ctrl-P2\_rep1.bw"

track type=bigWig db=hg19 name="Ctrl-P2\_rep2" visibility=2

```
bigDataUrl="http://159.226.118.212/home/liuzhe/projects_cas/xuelab/JL_GSC_projects/RNA-seq/PartII_U87/Ctrl-P2_rep2.bw"
track type=bigWig db=hg19 name="Ctrl-P3_rep1" visibility=2
bigDataUrl="http://159.226.118.212/home/liuzhe/projects_cas/xuelab/JL_GSC_projects/RNA-seq/PartII_U87/Ctrl-P3_rep1.bw"
track type=bigWig db=hg19 name="Ctrl-P3_rep2" visibility=2
bigDataUrl="http://159.226.118.212/home/liuzhe/projects_cas/xuelab/JL_GSC_projects/RNA-seq/PartII_U87/Ctrl-P3_rep2.bw"
track type=bigWig db=hg19 name="TMZ-P1_rep1" visibility=2
bigDataUrl="http://159.226.118.212/home/liuzhe/projects_cas/xuelab/JL_GSC_projects/RNA-seq/PartII_U87/TMZ-P1_rep1.bw"
track type=bigWig db=hg19 name="TMZ-P1_rep2" visibility=2
bigDataUrl="http://159.226.118.212/home/liuzhe/projects_cas/xuelab/JL_GSC_projects/RNA-seq/PartII_U87/TMZ-P1_rep2.bw"
track type=bigWig db=hg19 name="TMZ-P2_rep1" visibility=2
bigDataUrl="http://159.226.118.212/home/liuzhe/projects_cas/xuelab/JL_GSC_projects/RNA-seq/PartII_U87/TMZ-P2_rep1.bw"
track type=bigWig db=hg19 name="TMZ-P2_rep2" visibility=2
bigDataUrl="http://159.226.118.212/home/liuzhe/projects_cas/xuelab/JL_GSC_projects/RNA-seq/PartII_U87/TMZ-P2_rep2.bw"
track type=bigWig db=hg19 name="TMZ-P3_rep1" visibility=2
bigDataUrl="http://159.226.118.212/home/liuzhe/projects_cas/xuelab/JL_GSC_projects/RNA-seq/PartII_U87/TMZ-P3_rep1.bw"
track type=bigWig db=hg19 name="TMZ-P3_rep2" visibility=2
bigDataUrl="http://159.226.118.212/home/liuzhe/projects_cas/xuelab/JL_GSC_projects/RNA-seq/PartII_U87/TMZ-P3_rep2.bw"
)
```

## 第二步：转录本组装并过滤转录本

文件路径：

/mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-Basic\_operation/Step2\_AssembleTranscript/PartII/cufflinks

输入文件：

Ctrl-P1\_rep1.sort.bam;Ctrl-P1\_rep2.sort.bam;Ctrl-P2\_rep1.sort.bam;Ctrl-P2\_rep2.sort.bam;Ctrl-P3\_rep1.sort.bam;Ctrl-P3\_rep2.sort.bam;TMZ-P1\_rep1.sort.bam;TMZ-P1\_rep2.sort.bam;TMZ-P2\_rep1.sort.bam;TMZ-P2\_rep2.sort.bam;TMZ-P3\_rep1.sort.bam;TMZ-P3\_rep2.sort.bam;gencode.v19.annotation.gtf

输出文件：

程序代码：

```
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o Ctrl-P1_rep1 Ctrl-P1_rep1.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o Ctrl-P1_rep2 Ctrl-P1_rep2.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o Ctrl-P2_rep1 Ctrl-P2_rep1.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o Ctrl-P2_rep2 Ctrl-P2_rep2.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o Ctrl-P3_rep1 Ctrl-P3_rep1.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o Ctrl-P3_rep2 Ctrl-P3_rep2.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o TMZ-P1_rep1 TMZ-P1_rep1.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o TMZ-P1_rep2 TMZ-P1_rep2.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o TMZ-P2_rep1 TMZ-P2_rep1.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o TMZ-P2_rep2 TMZ-P2_rep2.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o TMZ-P3_rep1 TMZ-P3_rep1.sort.bam
cufflinks -p 12 --library-type fr-firststrand -G gencode.v19.annotation.gtf -o TMZ-P3_rep2 TMZ-P3_rep2.sort.bam
```

(备注：

mergel以上文件：

```
stringtie --merge -G gencode.v19.annotation.gtf -T 1 -o invivoselexU87.gtf Ctrl-P1_rep1.gtf Ctrl-P1_rep2.gtf Ctrl-P2_rep1.gtf
Ctrl-P2_rep2.gtf Ctrl-P3_rep1.gtf Ctrl-P3_rep2.gtf TMZ-P1_rep1.gtf TMZ-P1_rep2.gtf TMZ-P2_rep1.gtf TMZ-P2_rep2.gtf TMZ-
P3_rep1.gtf TMZ-P3_rep2.gtf -p 12
```

参数解释：-F <min\_fpk> minimum input transcript FPKM to include in the merge  
(default: 1.0)

stringtie版本：1.2.3

)

文件路径：

/mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-Basic\_operation/Step2\_AssembleTranscript/PartII

输入文件：

Ctrl-P1\_rep1.sort.bam;Ctrl-P1\_rep2.sort.bam;Ctrl-P2\_rep1.sort.bam;Ctrl-P2\_rep2.sort.bam;Ctrl-P3\_rep1.sort.bam;Ctrl-P3\_rep2.sort.bam;TMZ-P1\_rep1.sort.bam;TMZ-P1\_rep2.sort.bam;TMZ-P2\_rep1.sort.bam;TMZ-P2\_rep2.sort.bam;TMZ-P3\_rep1.sort.bam;TMZ-P3\_rep2.sort.bam;gencode.v19.annotation.gtf

输出文件：

Ctrl-P1\_rep1.gtf;Ctrl-P1\_rep2.gtf;Ctrl-P2\_rep1.gtf;Ctrl-P2\_rep2.gtf;Ctrl-P3\_rep1.gtf;Ctrl-P3\_rep2.gtf;TMZ-P1\_rep1.gtf;TMZ-P1\_rep2.gtf;TMZ-P2\_rep1.gtf;TMZ-P2\_rep2.gtf;TMZ-P3\_rep1.gtf;TMZ-P3\_rep2.gtf

程序代码：

```

stringtie Ctrl-P1_rep1.sort.bam -G gencode.v19.annotation.gtf -o Ctrl-P1_rep1.gtf -p 12
stringtie Ctrl-P1_rep2.sort.bam -G gencode.v19.annotation.gtf -o Ctrl-P1_rep2.gtf -p 12
stringtie Ctrl-P2_rep1.sort.bam -G gencode.v19.annotation.gtf -o Ctrl-P2_rep1.gtf -p 12
stringtie Ctrl-P2_rep2.sort.bam -G gencode.v19.annotation.gtf -o Ctrl-P2_rep2.gtf -p 12
stringtie Ctrl-P3_rep1.sort.bam -G gencode.v19.annotation.gtf -o Ctrl-P3_rep1.gtf -p 12
stringtie Ctrl-P3_rep2.sort.bam -G gencode.v19.annotation.gtf -o Ctrl-P3_rep2.gtf -p 12
stringtie TMZ-P1_rep1.sort.bam -G gencode.v19.annotation.gtf -o TMZ-P1_rep1.gtf -p 12
stringtie TMZ-P1_rep2.sort.bam -G gencode.v19.annotation.gtf -o TMZ-P1_rep2.gtf -p 12
stringtie TMZ-P2_rep1.sort.bam -G gencode.v19.annotation.gtf -o TMZ-P2_rep1.gtf -p 12
stringtie TMZ-P2_rep2.sort.bam -G gencode.v19.annotation.gtf -o TMZ-P2_rep2.gtf -p 12
stringtie TMZ-P3_rep1.sort.bam -G gencode.v19.annotation.gtf -o TMZ-P3_rep1.gtf -p 12
stringtie TMZ-P3_rep2.sort.bam -G gencode.v19.annotation.gtf -o TMZ-P3_rep2.gtf -p 12

```

(备注:

mergel以上文件:

```

stringtie --merge -G gencode.v19.annotation.gtf -T 1 -o invivoselexU87.gtf Ctrl-P1_rep1.gtf Ctrl-P1_rep2.gtf Ctrl-P2_rep1.gtf
Ctrl-P2_rep2.gtf Ctrl-P3_rep1.gtf Ctrl-P3_rep2.gtf TMZ-P1_rep1.gtf TMZ-P1_rep2.gtf TMZ-P2_rep1.gtf TMZ-P2_rep2.gtf TMZ-
P3_rep1.gtf TMZ-P3_rep2.gtf -p 12

```

参数解释: -F <min\_fpk> minimum input transcript FPKM to include in the merge  
(default: 1.0)

stringtie版本: 1.2.3

)

第三步: 计算表达值并根据表达水平并根据表达水平过滤数据

(1) 计算表达值FPKM

文件路径: /mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-

Basic\_operation/Step3\_CalculateExpression/PartI/Known/GSC\_GSC+TMZ\_hisat2\_out

输入文件: Ctrl-P1\_rep1.sort.bam;Ctrl-P1\_rep2.sort.bam;Ctrl-P2\_rep1.sort.bam;Ctrl-P2\_rep2.sort.bam;Ctrl-P3\_rep1.sort.bam;Ctrl-  
P3\_rep2.sort.bam;TMZ-P1\_rep1.sort.bam;TMZ-P1\_rep2.sort.bam;TMZ-P2\_rep1.sort.bam;TMZ-P2\_rep2.sort.bam;TMZ-  
P3\_rep1.sort.bam;TMZ-P3\_rep2.sort.bam;targets;gencode.v19.annotation.gtf

输出文件:

merge\_known.fpkm.csv;merge\_known.count.csv;JL.RData

程序代码:

RCode.R

(2) 根据表达值过滤数据FPKM>1

文件路径: /mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-

Basic\_operation/Step3\_CalculateExpression/PartI/Known/GSC\_GSC+TMZ\_hisat2\_out

输入文件: ref.fpkm.csv

输出文件: ref.fpkm.filterbyExp.csv

程序代码: python PyFilterByExp.py ref.fpkm.csv ref.fpkm.filterbyExp.csv

(3) 基因注释FPKM

文件路径: /mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-

Basic\_operation/Step3\_CalculateExpression/PartI/Known/GSC\_GSC+TMZ\_hisat2\_out

输入文件: summary.txt ; ref.fpkm.filterbyExp.csv

输出文件: geneExpression.fpkm.filterbyExp.txt

程序代码: python PyAnnotation.py summary.txt ref.fpkm.filterbyExp.csv geneExpression.fpkm.filterbyExp.txt

(6) 基因注释Raw Read

文件路径: /mnt/pub/work/liuzhe/projs-cas-ibp/xuelab/JL/Step1-

Basic\_operation/Step3\_CalculateExpression/PartI/Known/GSC\_GSC+TMZ\_hisat2\_out

输入文件: summary.txt ; ref.count.csv

输出文件: geneExpression.count.txt

程序代码: python PyAnnotation.py summary.txt ref.count.csv geneExpression.count.txt