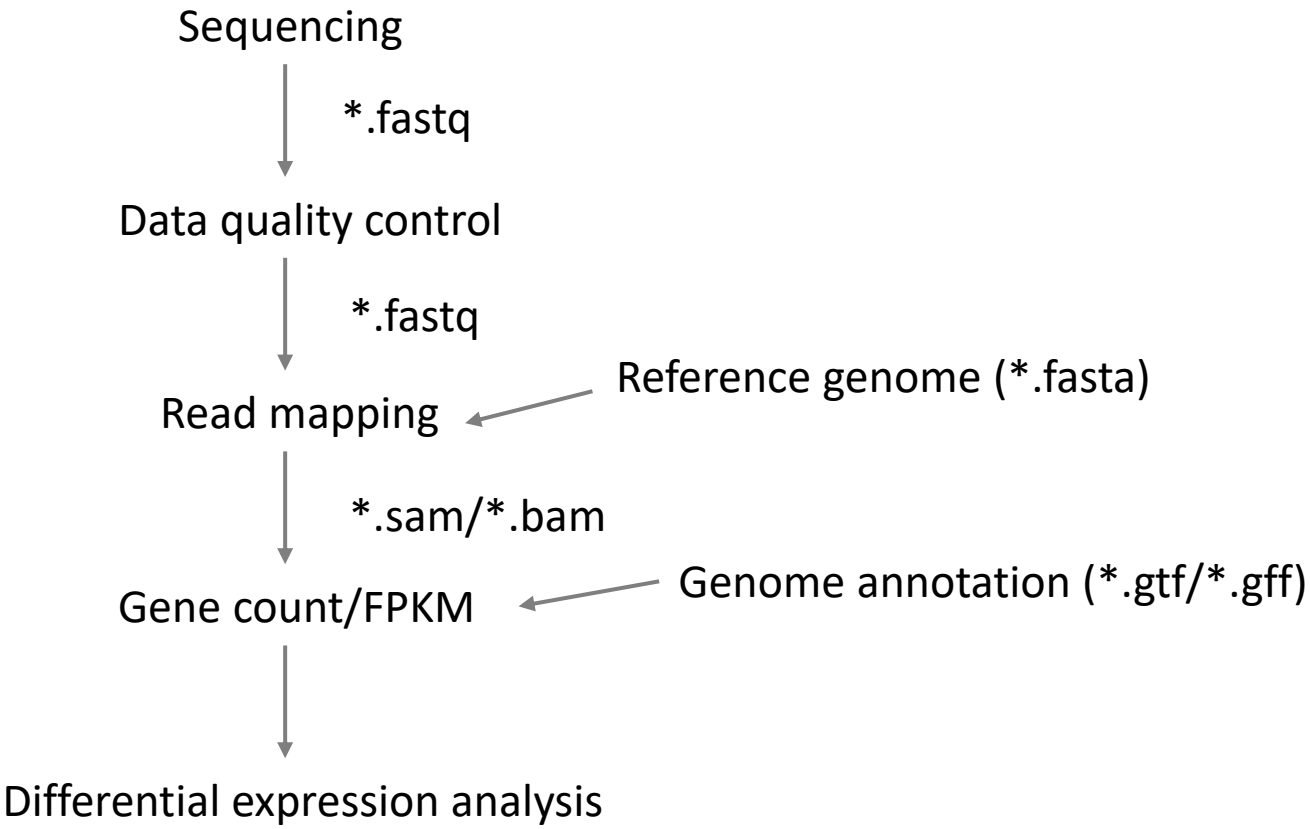
RNA-seq基本分析流程

2020.10.27

崔泽嘉

RNA-seq分析流程


Sequencing
*.fastq
Data quality control
*.fastq
Reference genome (*.fasta)
Read mapping
.sam/.bam
Genome annotation (*.gtf/*.gff)
Gene count/FPKM
Differential expression analysis

数据获取

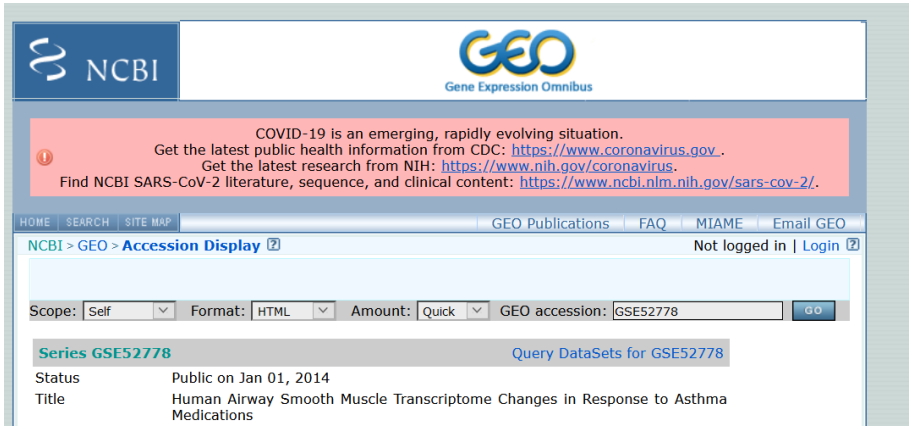
OPEN ACCESS Freely available online

PLOS | ONE

RNA-Seq Transcriptome Profiling Identifies *CRISPLD2* as a Glucocorticoid Responsive Gene that Modulates Cytokine Function in Airway Smooth Muscle Cells

Blanca E. Himes[1,2,3]*, Xiaofeng Jiang[4], Peter Wagner[4], Ruoxi Hu[4], Qiyu Wang[4], Barbara Klanderman[2], Reid M. Whitaker[1], Qingling Duan[1], Jessica Lasky-Su[1], Christina Nikolos[5], William Jester[5], Martin Johnson[5], Reynold A. Panettieri Jr.[5], Kelan G. Tantisira[1], Scott T. Weiss[1,2], Quan Lu[4]*

positive control of gene expression, the FPKM values for four housekeeping genes (i.e., *B2M*, *GABARAP*, *GAPDH*, *RPL19*) were obtained. Each had high FPKM values that did not differ significantly by treatment status [Figure S11]. The NIH Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to perform gene functional annotation clustering using Homo Sapiens as background, and default options and annotation categories (Disease: OMIM_DISEASE; Functional Categories: COG_ONTOLOGY, SP_PIR_KEYWORDS, UP_SEQ_FEATURE; Gene_Ontology: GOTERM_BP_FAT, GOTERM_CC_FAT, GOTERM_MF_FAT; Pathway: BBID, BIOCARTA, KEGG_PATHWAY; Protein_Domains: INTERPRO, PIR_SUPERFAMILY, SMART) [28]. The RNA-Seq data is available at the Gene Expression Omnibus Web site (http://www.ncbi.nlm.nih.gov/geo/) under accession GSE52778.



数据获取

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52778

Platforms (1)	GPL11154 Illumina HiSeq 2000 (Homo sapiens)	
Samples (16) Less...	GSM1275862 N61311_untreated	SRR1039508
	GSM1275863 N61311_Dex	SRR1039509
	GSM1275864 N61311_Alb	
	GSM1275865 N61311_Alb_Dex	
	GSM1275866 N052611_untreated	SRR1039512
	GSM1275867 N052611_Dex	SRR1039513
	GSM1275868 N052611_Alb	
	GSM1275869 N052611_Alb_Dex	
	GSM1275870 N080611_untreated	
	GSM1275871 N080611_Dex	
	GSM1275872 N080611_Alb	
	GSM1275873 N080611_Alb_Dex	
	GSM1275874 N061011_untreated	
	GSM1275875 N061011_Dex	
	GSM1275876 N061011_Alb	
	GSM1275877 N061011_Alb_Dex	

数据获取

1、prefetch SRR1039508

2、wget https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos1/sra-pub-run-5/SRR1039508/SRR1039508.1

*.sra → *.fastq

fastq-dump -O ./ --gzip --split-3 *.sra

注:
-O: 输出文件路径;
--split-3: 将双端测序分为两份,放在不同的文件, 对于一方有而一方没有的reads会单独放在一个文件里

fasterq-dump *.sra

NCBI SRA Toolkit

https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software

NCBI SRA Toolkit

Below are the latest releases of various tools and release checksum file.

SRA Toolkit

Compiled binaries/install scripts of June 29, 2020, version 2.10.8:

- CentOS Linux 64 bit architecture - non-sudo tar archive
- Ubuntu Linux 64 bit architecture - non-sudo tar archive
- Cloud - apt-get install script - for Debian and Ubuntu - requires sudo permissions
- Cloud - yum install script - for for CentOS - requires sudo permissions
- MacOS 64 bit architecture
- MS Windows 64 bit architecture
- md5 checksums

安装参考：https://www.jianshu.com/p/c29ae5fe6f99

数据质量控制

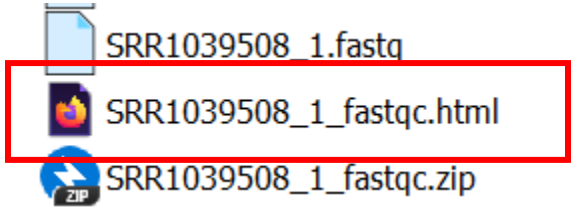
FastQC是一款基于Java的软件，官网：
http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

```
#获取fastqc
wget http://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc_v0.11.7.zip
#解压
unzip fastqc_v0.11.7.zip
#更改文件权限
chmod 777 fastqc
#输出帮助文档
fastqc -h
```

数据质量控制

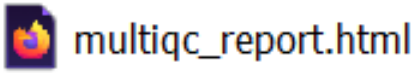
fastqc -o outdir -t threads *.fastq

```
Started analysis of SRR1039508_1.fastq
Approx 20% complete for SRR1039508_1.fastq
Approx 40% complete for SRR1039508_1.fastq
Approx 60% complete for SRR1039508_1.fastq
Approx 80% complete for SRR1039508_1.fastq
Approx 100% complete for SRR1039508_1.fastq
Analysis complete for SRR1039508_1.fastq
```



multiqc *.zip

整合质控报告



数据质量控制

Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

数据质量控制

1、Trimmomatic
https://www.jianshu.com/p/a8935adebaae
#双端测序数据
java -jar trimmomatic-0.39.jar PE $seq1 $seq2 [-phred33 | -phred64]
seq1.clean.fq.gz seq1.unpaired.fq.gz
seq2.clean.fq.gz seq2.unpaired.fq.gz
SLIDINGWINDOW:5:15 LEADING:5 TRAILING:5 MINLEN:50

2、Trim Galore
https://www.jianshu.com/p/7a3de6b8e503
trim_galore -q 20 [--phred33 | --phred64]
--length 20 --paired $seq1 $seq2 --gzip -o outputdir --stringency 3 --length 20
注：
-q：设定Phred quality score阈值，默认为20；
--length：设定输出reads长度阈值，小于设定值会被抛弃；
--stringency：设定可以忍受的前后adapter重叠的碱基数
[--phred33 | --phred64]：碱基质量体系的选择

数据质量控制

Illumina 1.8 + : phred33

Basic Statistics

Measure	Value
Filename	[illegible]
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	22935521
Sequences flagged as poor quality	0
Sequence length	63
%GC	50

Basic Statistics

Measure	Value
Filename	[illegible]
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	22795331
Sequences flagged as poor quality	0
Sequence length	30-90
%GC	50

Illumina 1.3/1.5 : phred64

其他质控软件

Cutadapter：
https://zhuanlan.zhihu.com/p/34999944

Fastp：
https://www.jianshu.com/p/6f492058da5b

…

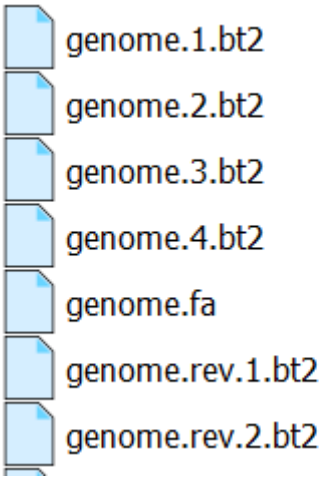
质控完别忘了再质检一遍！！！
质控完别忘了再质检一遍！！！
质控完别忘了再质检一遍！！！

序列比对

Bowtie2, BWA, STAR, TopHat, HISAT…

```
#Bowtie2下载
wget https://nchc.dl.sourceforge.net/project/bowtie-bio/bowtie2/2.3.5.1/bowtie2-2.3.5.1-linux-x86_64.zip
#解压
unzip bowtie2-2.3.5.1-linux-x86_64.zip
#进入相应文件夹
cd bowtie2-2.3.5.1
#安装
make
#运行检测
./bowtie
```

序列比对



Bowtie2
http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml

#对参考基因组构建索引
nohup bowtie2-build genome.fa genome &

#比对
bowtie2 -p 5 --very-sensitive -x Bowtie2Index_dir/genome [--phred33 | --phred64]
-1 seq1.fastq.gz -2 seq2.fastq.gz -S sample.sam
注：
-p：线程数
--very-sensitive：-D 20 -R 3 -N 0 -L 20 -i S,1,0.50
-x: 基因组索引路径
-1 -2：质控后的双端测序文件
-S：输出sam格式文件

序列比对

```
22935521 reads; of these:
  22935521 (100.00%) were paired; of these:
    6990228 (30.48%) aligned concordantly 0 times
    11414132 (49.77%) aligned concordantly exactly 1 time
    4531161 (19.76%) aligned concordantly >1 times
    ----
    6990228 pairs aligned concordantly 0 times; of these:
      1981754 (28.35%) aligned discordantly 1 time
    ----
    5008474 pairs aligned 0 times concordantly or discordantly; of these:
      10016948 mates make up the pairs; of these:
        5648290 (56.39%) aligned 0 times
        3783996 (37.78%) aligned exactly 1 time
        584662 (5.84%) aligned >1 times
87.69% overall alignment rate
```

比对结果解释：
http://www.manongjc.com/article/45738.html

序列比对

参考基因组（*.fasta）

```
>1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

人类参考基因组序列下载：
Ensembl，UCSC，NCBI

示例：Ensembl（GRCh38）
ftp://ftp.ensembl.org/pub/release-101/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.toplevel.fa.gz

序列比对

人类基因组版本对照关系：

GRCh36 <=> hg18
GRCh37 <=> hg19
GRCh38 <=> hg38

参考基因组下载方式参考：
https://my.oschina.net/u/4580290/blog/4620761

序列比对

SAM文件格式：
SAM的全称是sequence alignment/map format；

SAM 格式主要包括两大部分:
1.标头注释部分（header section）
2.比对结果部分（alignment section）

```
@HD     VN:1.0  SO:unsorted
@SQ     SN:10   LN:135534747
@SQ     SN:11   LN:135006516
@SQ     SN:12   LN:133851895
@SQ     SN:13   LN:115169878
@SQ     SN:14   LN:107349540
@SQ     SN:15   LN:102531392
@SQ     SN:16   LN:90354753
@SQ     SN:17   LN:81195210
@SQ     SN:18   LN:78077248
@SQ     SN:19   LN:59128983
```

```
SRR1039508.49   99      1       150768950       42      63M     =       150769191       304     CTATGTGAAA
ATCTCCAGCCTGTACCTGTACAGCATCAGCCCTGGGACAACACAGTCAGGGGC HJJJJIIJJJJJJJJJJJJJJJHIJJJJIIJIJJJJJJJJJJJJJJJJIJJI
JJJGHJJJHHF AS:i:0  XN:i:0  XM:i:0  XO:i:0  XG:i:0  NM:i:0  MD:Z:63 YS:i:0  YT:Z:CP
SRR1039508.49   147     1       150769191       42      63M     =       150768950       -304    ACACCAACTC
CCTTCCAAAGTGCATCGTTACACTGCACCATCGTGGAAGAAATGGAAGAGCAG BFDHHHEIIJJJIHJJJJJJJJJJHJJJJIJHJJJIGJJJJJJJJJJJJJJJ
JJJJJJJJJJH AS:i:0  XN:i:0  XM:i:0  XO:i:0  XG:i:0  NM:i:0  MD:Z:63 YS:i:0  YT:Z:CP
SRR1039508.50   81      5       41861431        42      63M     =       41853534        -7960   ACTGCAGTTA
GTCCTTTTACTCCAGTTTTCAGTAAAGCATCTATAAGATTCTCTGGAATTCCA EJJIJIIJJJJIJIIIJHFJIJIIJJJGJIJJJJIHGIGIJGGDJJJJIJJ
IJIHJIJJIHF AS:i:0  XN:i:0  XM:i:0  XO:i:0  XG:i:0  NM:i:0  MD:Z:63 YS:i:0  YT:Z:DP
SRR1039508.50   161     5       41853534        42      63M     =       41861431        7960    TCCACTTCTA
ATTCACCAGATAAGTACTGTCGTTCAAATTCTGCATTTTCTCCCACATATGAA HJIEHGIJIIIIHJIGIJDGIIIJDFHIJJJJFHIIJIFHIJIJJIIIIJJJ
IIJIJIJJJJI AS:i:0  XN:i:0  XM:i:0  XO:i:0  XG:i:0  NM:i:0  MD:Z:63 YS:i:0  YT:Z:DP
```

参考：https://www.jianshu.com/p/2aad7fc4f14a

计数/表达量

计数前需要进行格式转换。
工具：Samtools

```
#下载
wget  https://github.com/samtools/samtools/releases/download/1.9/samtools-1.9.tar.bz2
#解压缩
tar jxvf samtools-1.9.tar.bz2
#安装
cd samtools-1.9
./configure --prefix=（绝对路径）
make
make install
#查看帮助文档
./samtools –help
```

参考： https://www.jianshu.com/p/6b7a442d293f

计数/表达量

#将sam文件转为bam文件
samtools view –Sb sample.sam > sample.bam

#对bam文件进行排序
samtools sort -O bam -o sample.sorted.bam sample.bam

#查看bam文件
samtools view sample.bam | less

```
SRR1039508.49   99      1       150768950       42      63M     =       150769191       304     CTATGTGAAA
ATCTCCAGCCTGTACCTGTACAGCATCAGCCCTGGGACAACACAGTCAGGGGC HJJJJIIJJJJJJJJJJJJJJJHIJJJJIIJIJJJJJJJJJJJJJJJJIJJI
JJJGHJJJHHF AS:i:0  XN:i:0  XM:i:0  XO:i:0  XG:i:0  NM:i:0  MD:Z:63 YS:i:0  YT:Z:CP
SRR1039508.49   147     1       150769191       42      63M     =       150768950       -304    ACACCAACTC
CCTTCCAAAGTGCATCGTTACACTGCACCATCGTGGAAGAAATGGAAGAGCAG BFDHHHEIIJJJIHJJJJJJJJJJHJJJJIJHJJJIGJJJJJJJJJJJJJJJ
JJJJJJJJJJH AS:i:0  XN:i:0  XM:i:0  XO:i:0  XG:i:0  NM:i:0  MD:Z:63 YS:i:0  YT:Z:CP
SRR1039508.50   81      5       41861431        42      63M     =       41853534        -7960   ACTGCAGTTA
GTCCTTTTACTCCAGTTTTCAGTAAAGCATCTATAAGATTCTCTGGAATTCCA EJJIJIIJJJJIJIIIJHFJIJIIJJJJGJIJJJJIHGIGIJGGDJJJJIJJ
IJIHJIJJIHF AS:i:0  XN:i:0  XM:i:0  XO:i:0  XG:i:0  NM:i:0  MD:Z:63 YS:i:0  YT:Z:DP
SRR1039508.50   161     5       41853534        42      63M     =       41861431        7960    TCCACTTCTA
ATTCACCAGATAAGTACTGTCGTTCAAATTCTGCATTTTCTCCCACATATGAA HJIEHGIJIIIIHJIGIJDGIIIJDFHIJJJJFHIIJIFHIJIJJIIIIJJJ
IIJIJIJJJJI AS:i:0  XN:i:0  XM:i:0  XO:i:0  XG:i:0  NM:i:0  MD:Z:63 YS:i:0  YT:Z:DP
```

计数/表达量

#统计比对率
samtools flagstat sample.sorted.bam > sample.stat.txt

```
45871042 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
40222752 + 0 mapped (87.69% : N/A)
45871042 + 0 paired in sequencing
22935521 + 0 read1
22935521 + 0 read2
31890586 + 0 properly paired (69.52% : N/A)
36326358 + 0 with itself and mate mapped
3896394 + 0 singletons (8.49% : N/A)
262958 + 0 with mate mapped to a different chr
183663 + 0 with mate mapped to a different chr (mapQ>=5)
```

计数/表达量

```
#HTSeq获取
wget
https://pypi.python.org/packages/fd/94/b7c8c1dcb7a3c3dcbde66b8d29583df4fa0059d88cc3592f62d15ef539a2/HTSeq-0.9.1.tar.gz#md5=fc71e021bf284a68f5ac7533a57641ac
#解压
tar zxvf HTSeq-0.9.1.tar.gz
cd HTSeq-0.9.1
#安装
python setup.py build
python setup.py install

#使用
htseq-count  -f bam  -a 10  -t  exon  -i  gene_name  sample.sorted.bam  genes.gtf  >
sample_count.txt   2>sample-htseq.log
```

参考：
https://www.cnblogs.com/triple-y/p/9338890.html

计数/表达量

基因组注释文件（*.gtf/*.gff）下载:
Ensembl，UCSC，NCBI

示例：Ensembl（GRCh38）
wget ftp://ftp.ensembl.org/pub/release-101/gtf/homo_sapiens/Homo_sapiens.GRCh38.101.gtf.gz

*.gtf格式：

```
1	processed_transcript	exon	11869	12227	.	+	.	exon_id "ENSE00002234944"; exon_number "1";
gene_biotype "pseudogene"; gene_id "ENSG00000223972"; gene_name "DDX11L1"; gene_source "ensembl_havana"; transcript_id "ENST
00000456328"; transcript_name "DDX11L1-002"; transcript_source "havana"; tss_id "TSS15145";
1	processed_transcript	transcript	11869	14409	.	+	.	gene_biotype "pseudogene"; gene_id "
ENSG00000223972"; gene_name "DDX11L1"; gene_source "ensembl_havana"; transcript_id "ENST00000456328"; transcript_name "DDX11
L1-002"; transcript_source "havana"; tss_id "TSS15145";
1	transcribed_unprocessed_pseudogene	exon	11872	12227	.	+	.	exon_id "ENSE00002234632"; e
xon_number "1"; gene_biotype "pseudogene"; gene_id "ENSG00000223972"; gene_name "DDX11L1"; gene_source "ensembl_havana"; tra
nscript_id "ENST00000515242"; transcript_name "DDX11L1-201"; transcript_source "ensembl"; tss_id "TSS192935";
```

参考：https://blog.csdn.net/u011262253/article/details/89363809

注释文件要与参考基因组文件版本保持一致！

计数/表达量

Htseq输出：

```
A1BG    18
A1BG-AS1        90
A1CF    1
A2M     22673
A2M-AS1 94
A2ML1   0
A2ML1-AS1       0
A2ML1-AS2       0
A2MP1   1
A3GALT2 0
A4GALT  1569
A4GNT   3
AAAS    743
AACS    512
AACSP1  0
AADAC   6
AADACL2 1
AADACL3 0
```

其他软件：feature Counts，Cufflinks

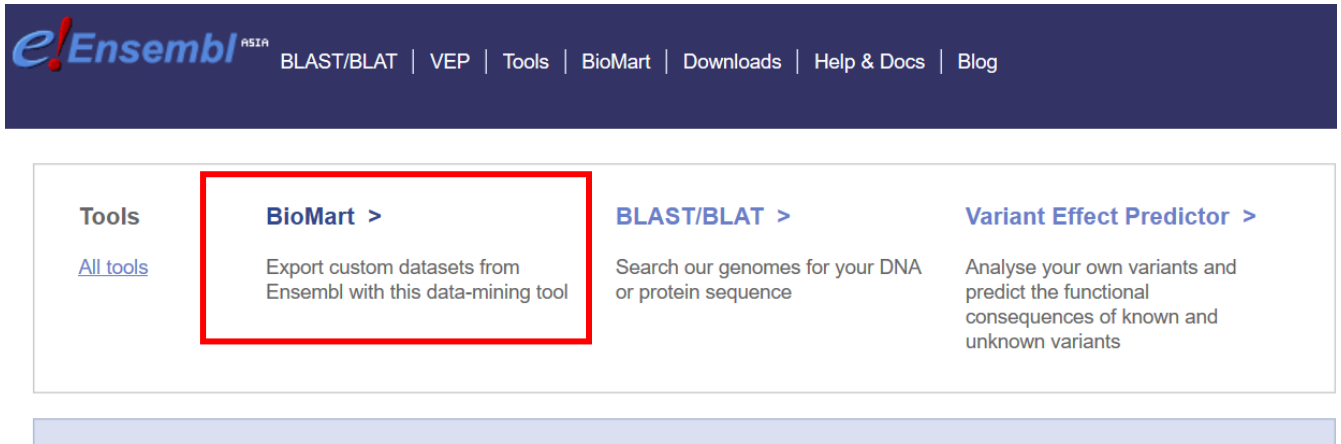
基因名称的转换

1、R

```
#导入依赖的包
library("AnnotationDbi")
library("org.Hs.eg.db")
#进行名称转换
res$symbol <- mapIds(org.Hs.eg.db, data, keytype="ENSEMBL“,
column="SYMBOL")
```

2、在线工具

http://asia.ensembl.org/biomart/martview/46238e04117169c8ed832f2d1fdbc74e



conda

```
#下载
wget https://repo.anaconda.com/archive/Anaconda3-5.2.0-Linux-x86_64.sh
#安装
sh Anaconda3-5.2.0-Linux-x86_64.sh

#利用conda安装生信工具
conda install -c bioconda sra-tools
conda install -c bioconda bowtie2
conda install -c bioconda samtools
conda install -c bcbio htseq
```

基因表达差异分析

```r
##安装DESeq2
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("DESeq2")
##导入DESeq2
library(DESeq2)
##读入gene count数据
raw_count <- read.csv("raw_count.csv")
##取出样本count值
count_data <- raw_count[,2:5]
##把第一列设置为行名
row.names(count_data) <- raw_count[,1]
##DESeq2构建表达矩阵
condition <- factor(c("trt","trt","untrt","untrt"),levels = c("trt","untrt"))
col_data <- data.frame(row.names = colnames(count_data),condition)
dds <- DESeqDataSetFromMatrix(countData = count_data, colData = col_data, design= ~ condition)
##将所有样本基因表达量之和小于1的基因过滤掉
dds_filter <- dds[rowSums(counts(dds))>1, ]
##使用DESeq函数进行差异分析
dds_out <- DESeq(dds_filter)
res <- results(dds_out)
```

```
"gene_name","trt1","trt2","untrt1","untrt2"
"A1BG",18,6,18,23
"A1BG-AS1",115,105,90,110
"A1CF",0,0,1,0
"A2M",17398,30450,22673,37152
```

基因表达差异分析

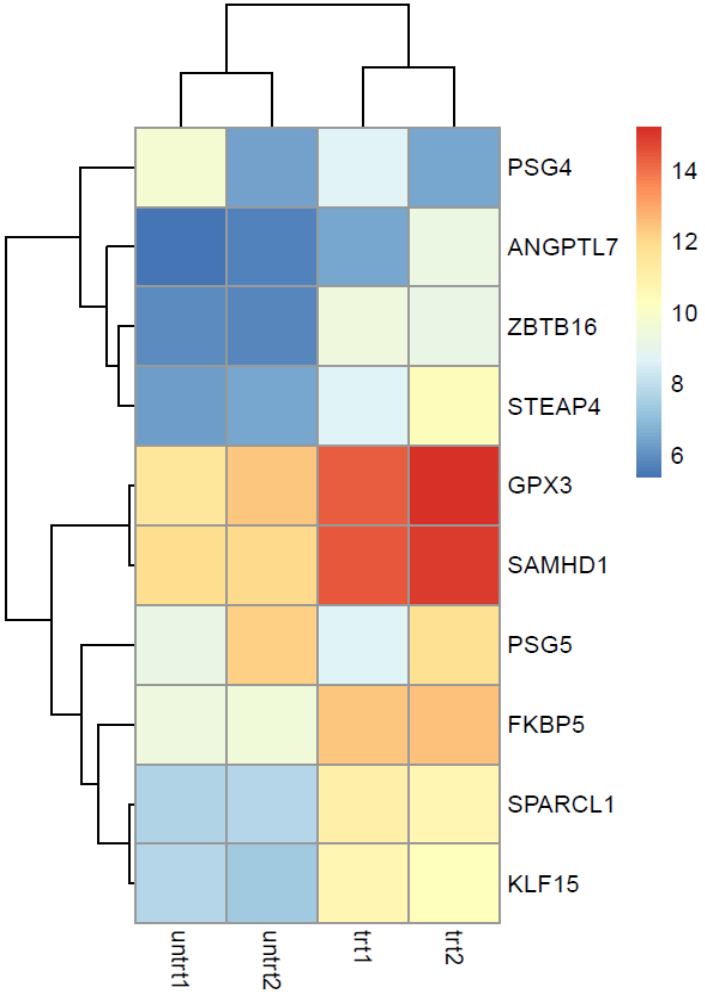
```
> summary(res)

out of 26003 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 550, 2.1%
LFC < 0 (down)     : 705, 2.7%
outliers [1]       : 0, 0%
low counts [2]     : 9075, 35%
(mean count < 11)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results

#设定阈值，筛选差异基因，保存结果
res <- res[order(res$padj),]
diff_gene <- subset(res, padj < 0.05 & (log2FoldChange > 1 | log2FoldChange < -1))
write.csv(diff_gene,file= "DEG_trt_vs_untrt.csv")
```

基因表达差异分析

```
###基因聚类热图
library(genefilter)
library(pheatmap)
##将数据进行log2转换，并进行归一化
rld <- rlogTransformation(dds_out,blind = F)
##选择方差最大的前10个基因
topVarGene <- head(order(rowVars(assay(rld)),
                         decreasing = TRUE),10)
mat <- assay(rld)[topVarGene, ]
##热图展示
pheatmap(mat)
```



PCA
火山图
MA-plot
GO, KEGG富集分析(enrichGO, enrichKEGG)等

流程选择

Bowtie2 + samtools + htseq + DeSeq2

STAR/HISAT+ samtools + htseq/feature Counts + DeSeq2

Tophat + cufflinks (Bowtie2 + samtools)

可参考：

1. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012,7(3):562-578.

2. http://combine-australia.github.io/RNAseq-R/