

差异表达基因的分析报告（NonoKO vs WT 差异表达）

姓名：李群

老师：吴飞珍

学号：20111510021

邮箱：20111510021@fudan.edu.cn

时间：2020年10月13日

数据来源

背景：

NONO是一种DNA/ rna结合蛋白，在小鼠胚胎干细胞(mESCs)细胞过渡过程中发挥着重要的调控作用。然而，其在神经元谱系承诺中的功能和其在这一过程中作用的分子机制大部分是未知的。尝试通过WT和KO来研究其功能。

该分析报告所用数据（两个重复）：

SRA	Day	Tag	Layout	Instrument
SRR8734708	Day 0	set1_WT_D0	PAIRED	Illumina HiSeq 2500
SRR8734718	Day 0	Set2_WT_D0	PAIRED	Illumina HiSeq 2500
SRR8734712	Day 0	set1_NonoKO_D0	PAIRED	Illumina HiSeq 2500
SRR8734722	Day 0	Set2_NonoKO_D0	PAIRED	Illumina HiSeq 2500

Organism: *Mus musculus*

All database accession code: PRJNA527295

Affiliations: Laboratory of Epigenetics, Institutes of Biomedical Sciences, Fudan University

Reference: Li W, Karwacki-Neisius V, Ma C, Tan L, Shi Y, Wu F, Shi YG. **Nono deficiency compromises TET1 chromatin association and impedes neuronal differentiation of mouse embryonic stem cells.** Nucleic Acids Res. 2020 May 21;48(9):4827-4838. doi: 10.1093/nar/gkaa213. PMID: 32286661; PMCID: PMC7229820.

分析过程

数据下载

在SRA数据库中查询表格中的数据，并记下数据的详细信息，包括测序类型、平台、采样处理方式等，下载方式多种，也可以通过==wget, prefetch和NCBI提供的接口==进行数据下载，该报告采用的方式为fastq-dump，由于服务器是公共服务器，可以分时间段进行下载、后台挂载下载、本地下载上传等方式获取数据。

代码如下：

```
[st28@ibs ~]$ mkdir report1          #创建文件夹 report1
[st28@ibs ~]$ cd report1/            #进入文件夹 report1，该报告分析过程在该文件夹下
进行
[st28@ibs report1]$ vim sra_list.txt  #创建文件sra_list.txt，将SRA number记录
[st28@ibs report1]$ cat sra_list.txt  #查看 sra_list.txt 文件
SRR8734708
SRR8734712
SRR8734718
SRR8734722
[st28@ibs report1]$ fastq-dump --help #查看帮助文档
...
--gzip                               Compress output using gzip: deprecated,
not recommended                     #在磁盘空间足够的情况下，不推荐使用 -
-gzip
...
[st28@ibs report1]$ while read line; do echo $line; fastq-dump --split-e $line
; done < sra_list.txt                #数据下载，也可以加nohup &挂载后台
SRR8734708
SRR8734712
SRR8734718
SRR8734722
[st28@ibs report1]$ mkdir rawData    #新建文件夹rawData
[st28@ibs report1]$ mv SRR87347* rawData/ #将测序文件放入rawData文件夹下
[st28@ibs report1]$ cd rawData/      #进入rawData文件夹
[st28@ibs rawData]$ ls               #查看文件夹
SRR8734708_1.fastq  SRR8734712_1.fastq  SRR8734718_1.fastq  SRR8734722_1.fastq
SRR8734708_2.fastq  SRR8734712_2.fastq  SRR8734718_2.fastq  SRR8734722_2.fastq
```

数据过滤

去除低质量的数据和接头等，采用trim_galore，推荐代码挂载后台nohup &。

```
[st28@ibs rawData]$ for i in *_1.fastq; do echo ${i:0:10}; trim_galore -q 30 -
stringency 3 -length 20 --phred33 -fastqc -clip_R1 3 -clip_R2 3 -e 0.1 -
dont_gzip -paired ${i:0:10}_1.fastq ${i:0:10}_2.fastq; done
#参数参考 trim_galore --help 帮助文档
```

随后进行数据质量查看（质量报告可以通过 *.html 文件进行查看），修剪文件报告文件可以通过 *_report.txt 查看

查看分析完之后文件夹下文件：

```
[st28@ibs rawData]$ ls
SRR8734708_1.fastq          SRR8734712_1_val_1.fq          SRR8734718_2_val_2_fastqc.zip
SRR8734708_1.fastq_trimming_report.txt  SRR8734712_2.fastq          SRR8734718_2_val_2.fq
SRR8734708_1_val_1_fastqc.html          SRR8734712_2.fastq_trimming_report.txt  SRR8734722_1.fastq
SRR8734708_1_val_1_fastqc.zip          SRR8734712_2_val_2_fastqc.html          SRR8734722_1.fastq_trimming_report.txt
SRR8734708_1_val_1.fq          SRR8734712_2_val_2_fastqc.zip          SRR8734722_1_val_1_fastqc.html
SRR8734708_2.fastq          SRR8734712_2_val_2.fq          SRR8734722_1_val_1_fastqc.zip
SRR8734708_2.fastq_trimming_report.txt  SRR8734718_1.fastq          SRR8734722_1_val_1.fq
SRR8734708_2_val_2_fastqc.html          SRR8734718_1.fastq_trimming_report.txt  SRR8734722_2.fastq
SRR8734708_2_val_2_fastqc.zip          SRR8734718_1_val_1_fastqc.html          SRR8734722_2.fastq_trimming_report.txt
SRR8734708_2_val_2.fq          SRR8734718_1_val_1_fastqc.zip          SRR8734722_2_val_2_fastqc.html
SRR8734712_1.fastq          SRR8734718_1_val_1.fq          SRR8734722_2_val_2_fastqc.zip
SRR8734712_1.fastq_trimming_report.txt  SRR8734718_2.fastq          SRR8734722_2_val_2.fq
SRR8734712_1_val_1_fastqc.html          SRR8734718_2.fastq_trimming_report.txt
SRR8734712_1_val_1_fastqc.zip          SRR8734718_2_val_2_fastqc.html
```

下载参考基因组、注释文件并构建索引

可以在ensembl和NCBI数据库中进行下载参考基因组和注释文件，构建索引可以使用HISAT2或Bowtie2，感谢老师提供的参考、注释和索引文件（`/public/share/Genomes/`），参考采用 `mm10`。

```
[st28@ibs Genomes]$ ls                                     #本次分析数据物种为小鼠，采用mm10参考基因组
mm10_Bowtie2Index  mm10_BowtieIndex  mm10_genes.gtf  Mus_musculus
```

比对到参考基因组

本次报告所用数据为双端测序数据，采用tophat比对，cufflinks进行组装，感谢吴老师提供的脚本 `go_FPKM_PE.sh`

```
[st28@ibs rawData]$ go_FPKM_PE.sh -h                      #可以通过帮助文档进行查看脚本的使用
Usage:
    go_FPKM_from_PE_fq.sh Threads Read1.fq Read2.fq transcriptome-index bowtie2-
index
[st28@ibs rawData]$ for i in *_1_val_1.fq; do echo ${i:0:10}; go_FPKM_PE.sh 10
${i:0:10}_2_val_2.fq /public/share/Genomes/mm10_genes.gtf
/public/share/Genomes/mm10_Bowtie2Index/genome; done
#批量运行脚本
```

查看比对结果，在输出结果的文件夹下找到对应的 `align_summary.txt` 文件，以SRR8734708为例，比对率（88.1%）：

```
[st28@ibs Tophat_Cufflinks_SRR8734718_1_val_1.fq_SRR8734718_2_val_2.fq]$ cat
align_summary.txt
Left reads:
    Input      : 32902461
    Mapped     : 29159466 (88.6% of input)
    of these:  3414216 (11.7%) have multiple alignments (220423 have
>20)
Right reads:
    Input      : 32902461
    Mapped     : 28809089 (87.6% of input)
    of these:  3400122 (11.8%) have multiple alignments (220429 have
>20)
```

```
88.1% overall read mapping rate.
```

```
Aligned pairs: 26816098
  of these: 3132101 (11.7%) have multiple alignments
            1578314 ( 5.9%) are discordant alignments
76.7% concordant pair alignment rate.
```

如果比对率不高，可重头进行思考（数据测序质量问题还是分析步骤出错，有无办法弥补），如果无法修正确保高比对率，最好的是在经费允许的情况下进行补测或重测。

FPKM与TPM的转换

FPKM: Fragments Per Kilobase of exon model per Million mapped fragments 即每千个碱基的转录每百万映射读取的fragments

TPM: TranscriptsPerKilobase of exonmodel per Million mapped reads 即每千个碱基的转录每百万映射读取的Transcripts

转换公式：

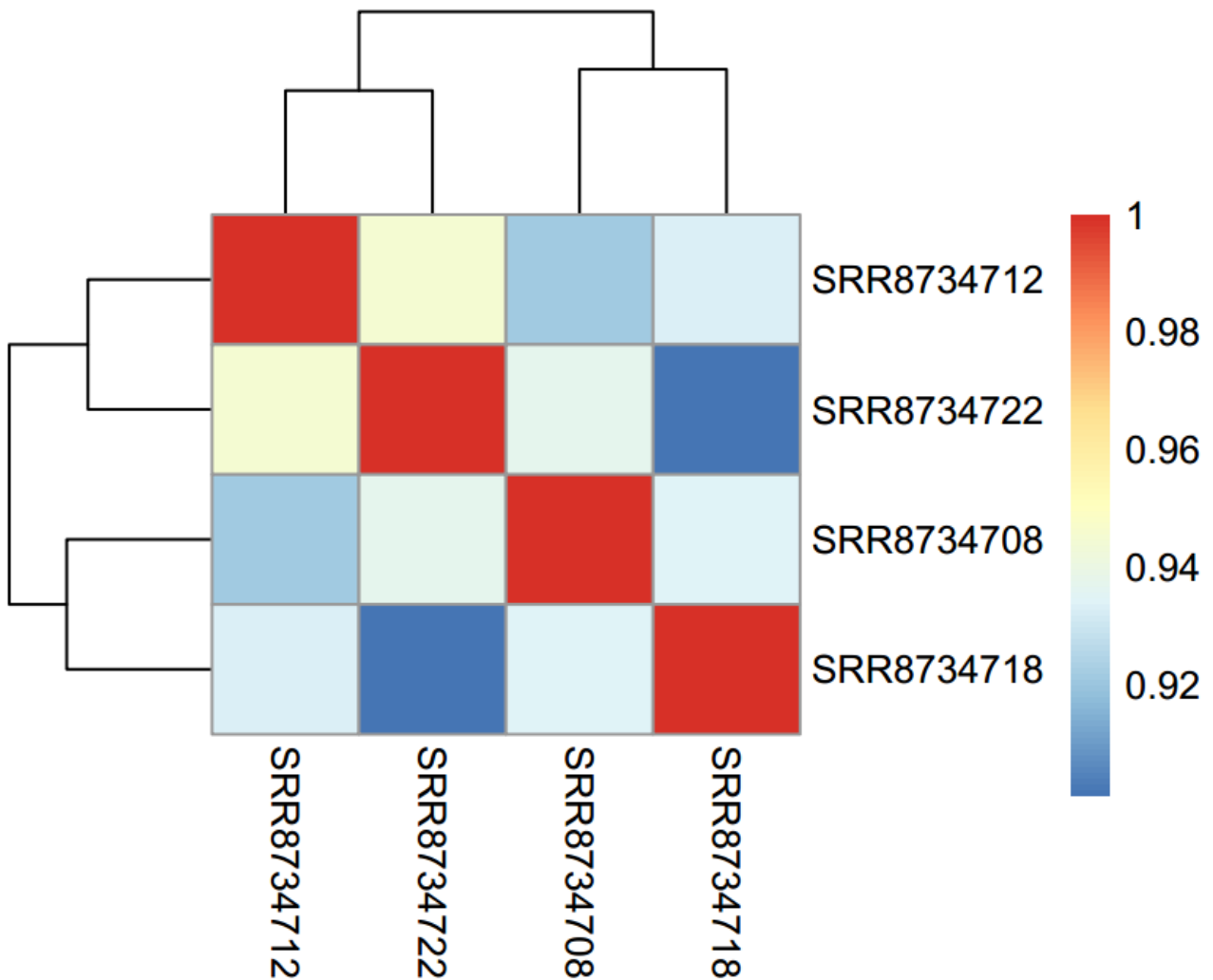
$$TPM = \frac{FPKM_i}{\sum FPKM_j} * 10^6$$

感谢吴老师提供的转换脚本 `FPKM2TPM.R`，提前将生成的FPKM文件放入文件夹`expFiles`下

提前安装好R包（"optparse","data.table","pheatmap"）

```
[st28@ibs expFiles]$
$ Rscript FPKM2TPM.R -f
SRR8734708.fpkm,SRR8734712.fpkm,SRR8734718.fpkm,SRR8734722.fpkm
The label of SRR8734708.fpkm was assigned as SRR8734708 !
The label of SRR8734712.fpkm was assigned as SRR8734712 !
The label of SRR8734718.fpkm was assigned as SRR8734718 !
The label of SRR8734722.fpkm was assigned as SRR8734722 !
null device
1
[st28@ibs expFiles]$ ls
Expression_TPM_correlation.pdf Expression_TPM.xls FPKM2TPM.R SRR8734708.fpkm
SRR8734712.fpkm SRR8734718.fpkm SRR8734722.fpkm
```

打开pdf文件查看样品之间的表达相似性



根据聚类结果：SRR8734712与SRR8734722为KO，SRR8734708与SRR8734718为WT分别聚类，结果较好。

差异表达基因

差异表达基因采用cuffdiff进行计算比较，结果输出在 **NONO_DiffGenes** 文件夹下

```
[st28@ibs rawData]$ cuffdiff -p 2 -o NONO_DiffGenes/  
/public/share/Genomes/mm10_genes.gtf -L KO,WT  
Tophat_Cufflinks_SRR8734712_1_val_1.fq_SRR8734712_2_val_2.fq/accepted_hits.bam,  
Tophat_Cufflinks_SRR8734722_1_val_1.fq_SRR8734722_2_val_2.fq/accepted_hits.bam  
Tophat_Cufflinks_SRR8734708_1_val_1.fq_SRR8734708_2_val_2.fq/accepted_hits.bam,  
Tophat_Cufflinks_SRR8734718_1_val_1.fq_SRR8734718_2_val_2.fq/accepted_hits.bam
```

[查看结果文件夹](#)

```
[st28@ibs NONO_DiffGenes]$ ls
bias_params.info      cds.fpkm_tracking      genes.fpkm_tracking
isoforms.fpkm_tracking  run.info
tss_groups.fpkm_tracking
cds.count_tracking    cds.read_group_tracking  genes.read_group_tracking
isoforms.read_group_tracking  splicing.diff
tss_groups.read_group_tracking
cds.diff              gene_exp.diff           isoform_exp.diff
promoters.diff        tss_group_exp.diff      var_model.info
cds_exp.diff          genes.count_tracking     isoforms.count_tracking
read_groups.info      tss_groups.count_tracking
```

gene_exp.diff 便是需要的差异表达文件

R语言代码

提前将TPM文件和差异表达文件拷贝到 `/public/home/st28/report1/rawData/RAnalysis` 文件夹下

```
#安装必要的R包
# Bioconductor, ggplot2, pheatmap, data.table, scales, org.Mm.eg.db, cowplot等其他依赖包
# 安装Bioconductor
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install(version = "3.9") #因为兼容性采用的3.9版本，目前最新版3.11 (2020-09-26)
# 安装其他包，在安装的过程中会自己安装其他依赖包，遇到没有该安装包或安装错误的时候，可以单独安装 (bioconductor、install和本地安装均可) 相关依赖
BiocManager::install(c('ggplot2', 'pheatmap', 'data.table', 'scales', 'org.Mm.eg.db', 'cowplot'))
#以下分析代码，参考吴老师提供的分析代码
setwd('/public/home/st28/report1/rawData/RAnalysis') #设置运行初始话文件夹
rm(list=ls()) #删除环境中已有变量
library(ggplot2) #加载ggplot2用于绘图
# 加载、提取数据并赋予标签
DEG=read.table("gene_exp.diff", header = T) #导入差异表达文件
DEG=DEG[,c(3,10,12)] #提取gene, logFC和p值
DEG=DEG[is.finite(DEG$log2.fold_change.),] #排除NA, NaN, 无穷大, 返回TRUE
DEG=DEG[abs(DEG$log2.fold_change.)>log2(1.5) & DEG$p_value<0.05,] #保留FC大于1.5, p值小于0.05的数据
names(DEG)=c("genes", "foldchange", "pvalue") #修改列名为 gene, foldchange和pvalue
DEG$regulation="down"
DEG$regulation[DEG$foldchange<0]="up" #增加一列regulation, 根据foldchange赋予up和down类别
#计算差异表达的时候我是KT vs WT所以此处令FC<0为up
# 统计上下调基因数量，主要使用的是ggplot2
```

```

tab = as.data.frame(table(DEG$regulation))
tab$Var1=factor(tab$Var1,levels=c("up","down"))
p=ggplot(tab,aes(x=Var1,y=Freq,label = Freq,fill=Var1))+geom_bar(stat =
"identity")
p=p+geom_text(position = position_dodge(0.9),vjust =
0,size=3)+ylim(0,max(tab$Freq)*1.1)
p=p+theme_classic(8)+xlab("differential expression")+ylab("Number of genes")
p=p+ggtitle("NONO-diffgenes")+theme(legend.position = "none")
p=p+theme(plot.title = element_text(hjust = 0.5))
p
ggsave(p,filename = "NONO_diffgene_number_barplot.pdf",width = 2.2,height =
2.2) #保存
# 绘制热图, 并保存
library(scales)
library(pheatmap)
dd=read.table("Expression_TPM.xls",header = T)
DEG1=DEG[order(abs(DEG$foldchange),decreasing = T),]
DEG1=DEG1[1:40,]
ddl=dd[dd$gene_id %in% DEG1$genes,]
row.names(ddl)=ddl$gene_id
ddl$gene_id=NULL
head(ddl)
#           SRR8734708 SRR8734712 SRR8734718 SRR8734722
#   Acta1    66.151933  1.94048393 194.288144  1.77682715
#   Acta2    64.656162  1.19114537 246.012581  0.61783471
#   Actc1     4.700736  0.10098259  35.848814  0.31898462
#   Antxr2    1.981463  0.07358183   8.479333  0.05007741
#   Bmp1     28.094466  2.15089616 152.900672  1.51706758
#   Bmp2      3.513508  0.11223066  29.559151  0.06167556
names(ddl)=c("WT1","KO1","WT2","KO2")
head(ddl)
#           WT1          KO1          WT2          KO2
#   Acta1    66.151933  1.94048393 194.288144  1.77682715
#   Acta2    64.656162  1.19114537 246.012581  0.61783471
#   Actc1     4.700736  0.10098259  35.848814  0.31898462
#   Antxr2    1.981463  0.07358183   8.479333  0.05007741
#   Bmp1     28.094466  2.15089616 152.900672  1.51706758
#   Bmp2      3.513508  0.11223066  29.559151  0.06167556
pdf(file = "top_40gene.pdf",width = 3,height = 5.8)
pheatmap(log(ddl+0.01),cutree_rows = 2,cutree_cols=2,fontsize_row =8)
dev.off()
# 差异表达基因富集 (GO和KEGG) , 并保存图表结果
library(clusterProfiler)
library(org.Mm.eg.db)
library(cowplot)
gene <- bitr(DEG$genes, fromType = "SYMBOL",
             toType = "ENTREZID",
             OrgDb = org.Mm.eg.db)
geneList <- bitr(dd$gene_id, fromType = "SYMBOL",

```

```

        toType = "ENTREZID",
        OrgDb = org.Mm.eg.db)
# 3.56% of input gene IDs are fail to map...
ego <- enrichGO(gene = gene$ENTREZID,
                universe = names(geneList$ENTREZID),
                OrgDb = org.Mm.eg.db,
                ont = "BP",
                pAdjustMethod = "BH",
                pvalueCutoff = 0.01,
                qvalueCutoff = 0.05,
                readable = TRUE)

kk <- enrichKEGG(gene = gene$ENTREZID,
                 organism = 'mmu',
                 pvalueCutoff = 0.05,
                 pAdjustMethod = "BH",
                 qvalueCutoff = 0.05)
p1 <- dotplot(ego, showCategory=10) + ggtitle("dotplot for GOBP")
p2 <- dotplot(kk, showCategory=10) + ggtitle("dotplot for KEGG")
pp=plot_grid(p1, p2, ncol=2)
ggsave(pp,filename = "DEG_enrichment.pdf",width = 12,height = 3.8)
write.table(DEG,file = "DEG.xls",sep="\t",quote = F,row.names = F)

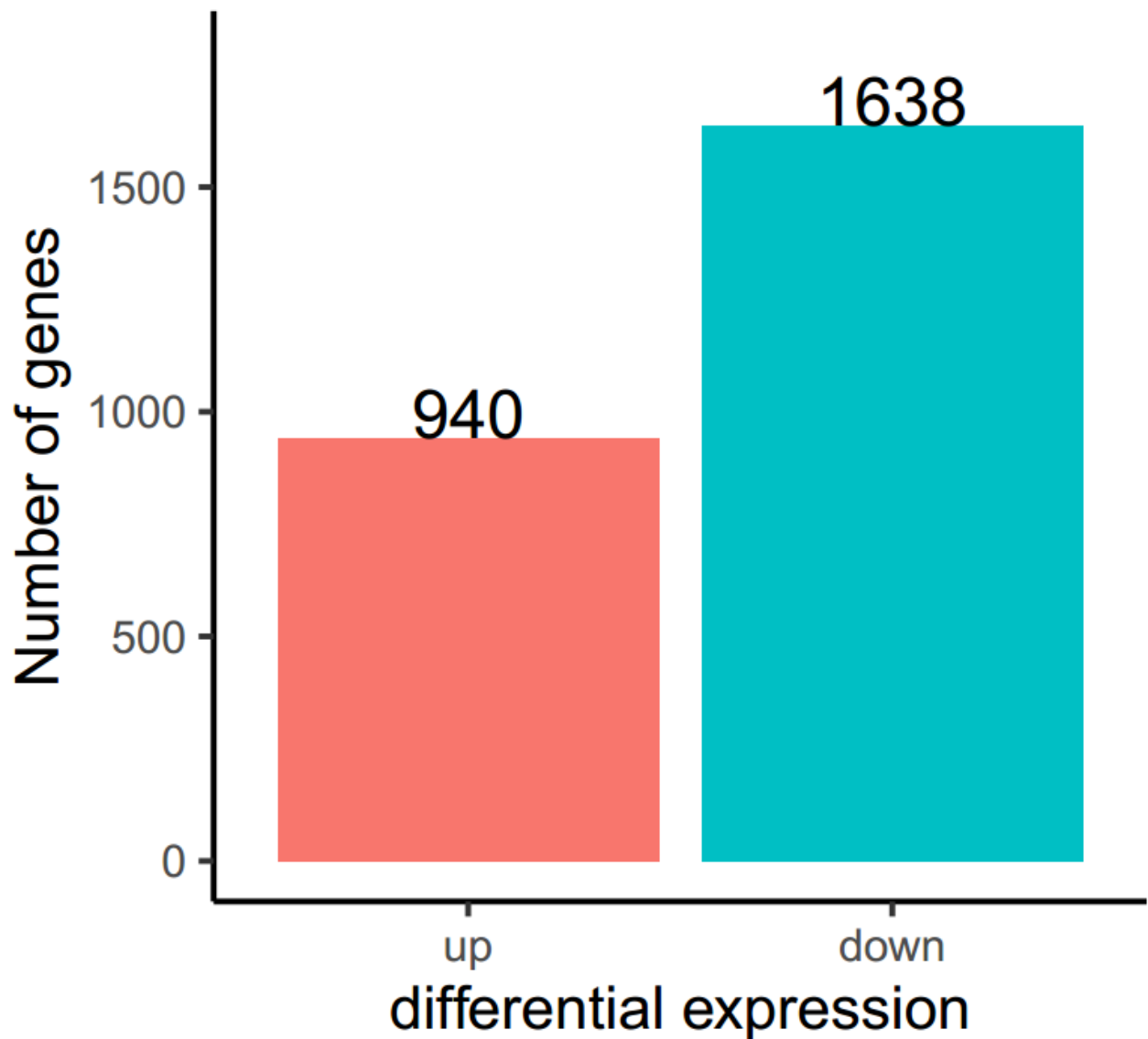
```

分析结果

差异表达基因上下调基因数量统计

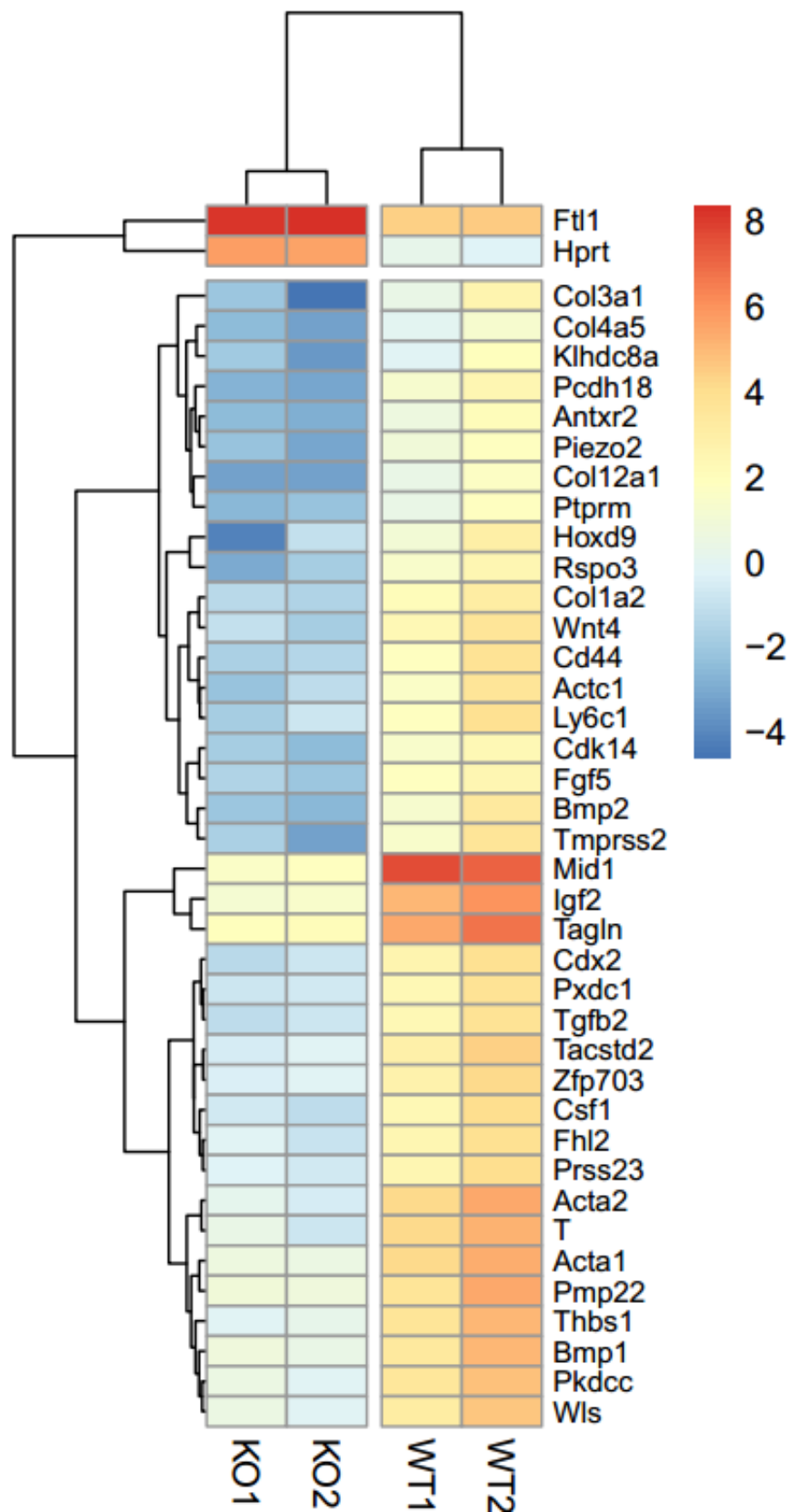
==下调1638个，上调940个基因== WT vs KO

NONO-diffgenes



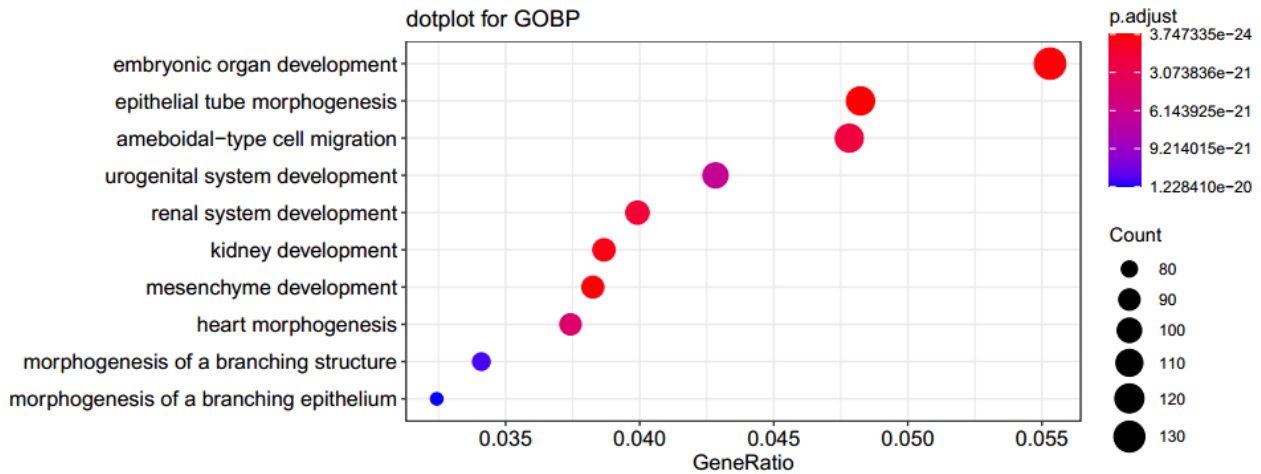
差异表达基因热图（top40 根据FC值）：

==关键发育相关基因 *mid1*, *igf2*, *wnt4*等下调。整体下调==,

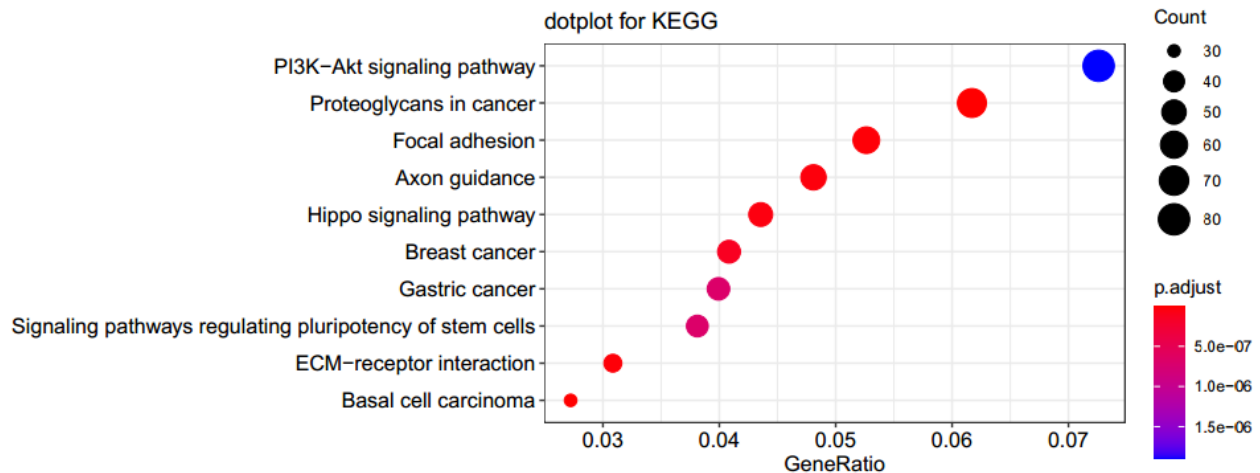


差异表达基因富集

==GO富集（与发育相关）==



==KEGG富集（与肿瘤相关）==



存在问题

敲除NONO后，关键发育相关基因 *mid1*, *igf2*, *wnt4*等下调，GO富集与发育相关，KEGG多于肿瘤相关，基本符合预期。但是分析过程存在以下问题：

1. 样本重复数目不够（大于等于3为宜）
2. 没有利用时间梯度数据
3. 没有利用回补实验数据
4. 差异表达基因数目过多，可进行转录因子等其他方向的预测筛选过滤

其他问题：

自己之前比较常用的是HISAT2-Stringtie-ballgown的分析流程，这次采用Tophat-Cufflinks-Cuffdiff流程，有一种不一样的感觉，学到了新的知识技巧，之前不习惯使用[%in%]这种方式循环赋值，本次报告使用了该方式，感觉很好，但是本次计算差异表达的时间相较于之前流程的时间较长，也对两种流程进行了对比，结果存在部分差异，重现性没有本报告的流程好。

感谢吴老师的分享。

最后的感受就是多学习，多进步。

附件

差异表达基因列表

> dd1

top40差异表达基因列表

	SRR8734708 (WT)	SRR8734712 (KO)	SRR8734718 (WT)	SRR8734722 (KO)
Acta1	66.1519330	1.940484e+00	194.2881440	1.776827e+00
Acta2	64.6561619	1.191145e+00	246.0125812	6.178347e-01
Actc1	4.7007358	1.009826e-01	35.8488144	3.189846e-01
Antxr2	1.9814628	7.358183e-02	8.4793329	5.007741e-02
Bmp1	28.0944660	2.150896e+00	152.9006717	1.517068e+00
Bmp2	3.5135080	1.122307e-01	29.5591506	6.167556e-02
Cd44	6.5612549	1.650077e-01	43.6100594	2.165396e-01
Cdk14	4.1978071	1.472880e-01	9.9242909	7.526878e-02
Cdx2	13.4700337	2.724905e-01	50.5207150	4.617529e-01
Col12a1	1.3919154	2.989396e-02	5.6689177	3.061835e-02
Col1a2	7.8612159	2.805825e-01	23.0578528	1.971897e-01
Col3a1	1.5229961	1.130129e-01	13.6248771	0.000000e+00
Col4a5	0.9382499	7.066095e-02	3.7134832	2.890888e-02
Csf1	10.8925158	5.297393e-01	58.9894538	3.126700e-01
Fgf5	6.3307378	2.000201e-01	11.0665988	1.165355e-01
Fhl2	11.6249256	8.519810e-01	49.8027806	3.982568e-01
Ftl1	77.7687646	3.622732e+03	93.2414938	4.333355e+03
Hoxd9	2.6922325	5.815661e-03	20.0780874	3.310563e-01
Hprt	1.2949007	3.096939e+02	0.7425953	2.776380e+02
Igf2	158.8835886	3.099140e+00	381.2636751	3.984557e+00
Klhdc8a	0.9072920	1.281523e-01	7.4184912	2.181745e-02
Ly6c1	6.1627387	1.451261e-01	48.0116296	4.344779e-01
Mid1	2097.5243196	4.925689e+00	1220.9889281	5.886125e+00
Pcdh18	3.6529514	5.234134e-02	11.7886172	3.567067e-02
Piezo2	2.5515621	1.063928e-01	5.8834382	3.352490e-02
Pkdcc	33.5337631	1.638894e+00	122.3086230	8.739140e-01
Pmp22	39.1555313	2.547372e+00	230.9086304	2.206820e+00
Prss23	11.0574392	7.505550e-01	58.2761256	5.300023e-01
Ptprm	1.4714017	6.451793e-02	6.3188664	9.895579e-02
Pxdc1	10.8845190	4.248187e-01	44.1959528	4.852000e-01
Rspo3	4.0822571	4.008057e-02	11.6655427	1.631729e-01
T	66.4572386	1.390009e+00	175.3347385	4.627446e-01
Tacstd2	16.8012649	5.691373e-01	79.6056543	8.660712e-01
Tagln	221.2229241	7.423137e+00	828.6860143	8.042690e+00
Tgfb2	10.0463691	3.068447e-01	44.8211514	4.703523e-01
Thbs1	37.9644183	8.498307e-01	150.9206714	1.232511e+00
Tmprss2	4.2315105	1.770847e-01	39.0819756	3.010043e-02
Wls	22.8489717	1.669864e+00	111.7465742	9.113347e-01
Wnt4	10.2283222	3.604768e-01	39.4922238	1.533446e-01
Zfp703	15.8542346	6.797623e-01	61.1969945	8.999231e-01

附差异表达基因全部内容

[Expression_TPM_liqun.xls](#)

[DEG_liqun.xls](#)

[gene_exp_liqun.diff](#)

R环境

```
# R版本环境
> sessionInfo()
R version 3.6.1 (2019-07-05)
Platform: x86_64-conda_cos6-linux-gnu (64-bit)
Running under: CentOS Linux 7 (Core)

# 安装的R包以及相关依赖
> .packages(all.available=T)
 [1] "AnnotationDbi"      "BH"                  "Biobase"             "BiocGenerics"
 [5] "BiocManager"       "BiocParallel"       "BiocVersion"        "DBI"
 [9] "DO.db"             "DOSE"               "GO.db"              "GOsemSim"
[13] "IRanges"           "MASS"              "Matrix"             "R6"
[17] "RColorBrewer"      "RSQLite"           "Rcpp"               "RcppArmadillo"
[21] "RcppEigen"         "S4Vectors"         "UpSetR"             "askpass"
[25] "assertthat"       "backports"         "bit"                "bit64"
[29] "blob"             "callr"             "class"              "cli"
[33] "cluster"          "clusterProfiler"   "colorspace"         "cowplot"
[37] "cpp11"            "crayon"            "curl"               "data.table"
[41] "desc"             "digest"            "dplyr"              "ellipsis"
[45] "enrichplot"       "europepmc"         "evaluate"           "fansi"
[49] "farver"           "fastmatch"         "fgsea"              "formatR"
[53] "futile.logger"    "futile.options"    "generics"           "getopt"
[57] "ggforce"          "ggplot2"           "ggplotify"          "ggraph"
[61] "ggrepel"          "ggridges"          "glue"               "graphlayouts"
[65] "gridExtra"        "gridGraphics"      "gtable"             "hms"
[69] "httr"             "igraph"            "isoband"            "jsonlite"
[73] "labeling"         "lambda.r"          "lattice"            "lifecycle"
[77] "magrittr"         "memoise"           "mgcv"               "mime"
[81] "munsell"          "nlme"              "nnet"               "openssl"
[85] "optparse"         "org.Mm.eg.db"      "pheatmap"           "pillar"
[89] "pkgbuild"         "pkgconfig"         "pkgload"            "plogr"
[93] "plyr"             "polyclip"          "praise"             "prettyunits"
[97] "processx"         "progress"          "ps"                 "purrr"
[101] "qvalue"           "reshape2"          "rlang"              "rprojroot"
[105] "rstudioapi"       "rvcheck"           "scales"             "snow"
[109] "spatial"          "stringi"           "stringr"            "survival"
[113] "sys"              "testthat"          "tibble"             "tidygraph"
[117] "tidyr"            "tidyselect"        "triebeard"          "tweenr"
[121] "urltools"         "utf8"              "vctrs"              "viridis"
[125] "viridisLite"      "withr"             "xml2"               "KernSmooth"
[129] "base"             "boot"              "codetools"          "compiler"
```

转录组分析Centos环境

安装的软件以及相关依赖

conda, trim_galore, Tophat, HISAT2, cufflinks等