

## 伯远生物Ribo-seq结题报告

项目编号	M00002
项目名称	Ribo-seq测序技术服务
受托方	武汉伯远生物科技有限公司
结题日期	2025-03-04

### 物种信息

物种名称	拉丁学名	参考基因组
拟南芥	<i>Arabidopsis thaliana</i>	Arabidopsis_thaliana

# 1 背景及分析流程

## □ 1.1 背景介绍

信使RNA (mRNA) 转化为蛋白质以及将所得蛋白质折叠成活性形式是几乎每个细胞过程的先决条件，并不是所有mRNA都能与核糖体结合翻译出蛋白质，因此需要翻译组学方法鉴定正在翻译中的mRNA。mRNA的翻译通常从募集由核糖体小亚基、三元复合物和起始因子形成的起始前复合物 (preinitiation complex) 开始。在真核生物中，通过真核起始因子eIF4E识别5' m7G帽，将前起始复合物募集到mRNA的5'末端。然后，起始前复合体扫描5'非翻译区域 (5' UTR)，直到识别起始位点。然而，起始位点可能绕过5' UTR 中的 AUG或者选择5' UTR 中的其他 AUG 样三联体作为起始密码子。一旦起始位点被识别，大多数起始因子被释放，核糖体大亚基加入起始复合物。在释放起始因子eIF5B后，起始核糖体进入延伸循环，真核延伸因子 eEF1 将带氨基酸的转移 RNA (tRNA) 递送至核糖体 A 位点，eEF2 催化核糖体易位。当核糖体到达终止密码子时翻译终止，与起始位点选择类似，终止密码子识别的保真度受顺式元件和反式调节因子的控制[1]。核糖体印记测序(Ribosome profile sequencing)，使用NGS测序鉴定核糖体保护的mRNA片段，可用于对体内翻译和偶联、共翻译事件进行全局分析[2]。

## □ 1.2 实验原理

核糖体分析依赖于核糖体足迹 (ribosome footprints, RFPs; 又称ribosome protected fragments, RPFs) 的测序，核糖体足迹是 mRNA 的短片段 (通常为约30个核苷酸 [nt])。首先，用翻译抑制剂处理细胞或组织，使得细胞翻译活动停止。再用核酸酶处理核糖体新生肽链复合物，降解没有核糖体覆盖的mRNA片段，去除核糖体。RFPs片段被核糖体物理包裹并屏蔽，不被核酸酶消化，它们被转化为DNA片段库，检测合格后进行文库制备[3]。构建好的文库经过文库质控合格后，使用Illumina Platform进行测序，Ribo-Seq测序策略一般使用SE 150(单端150bp测序)。

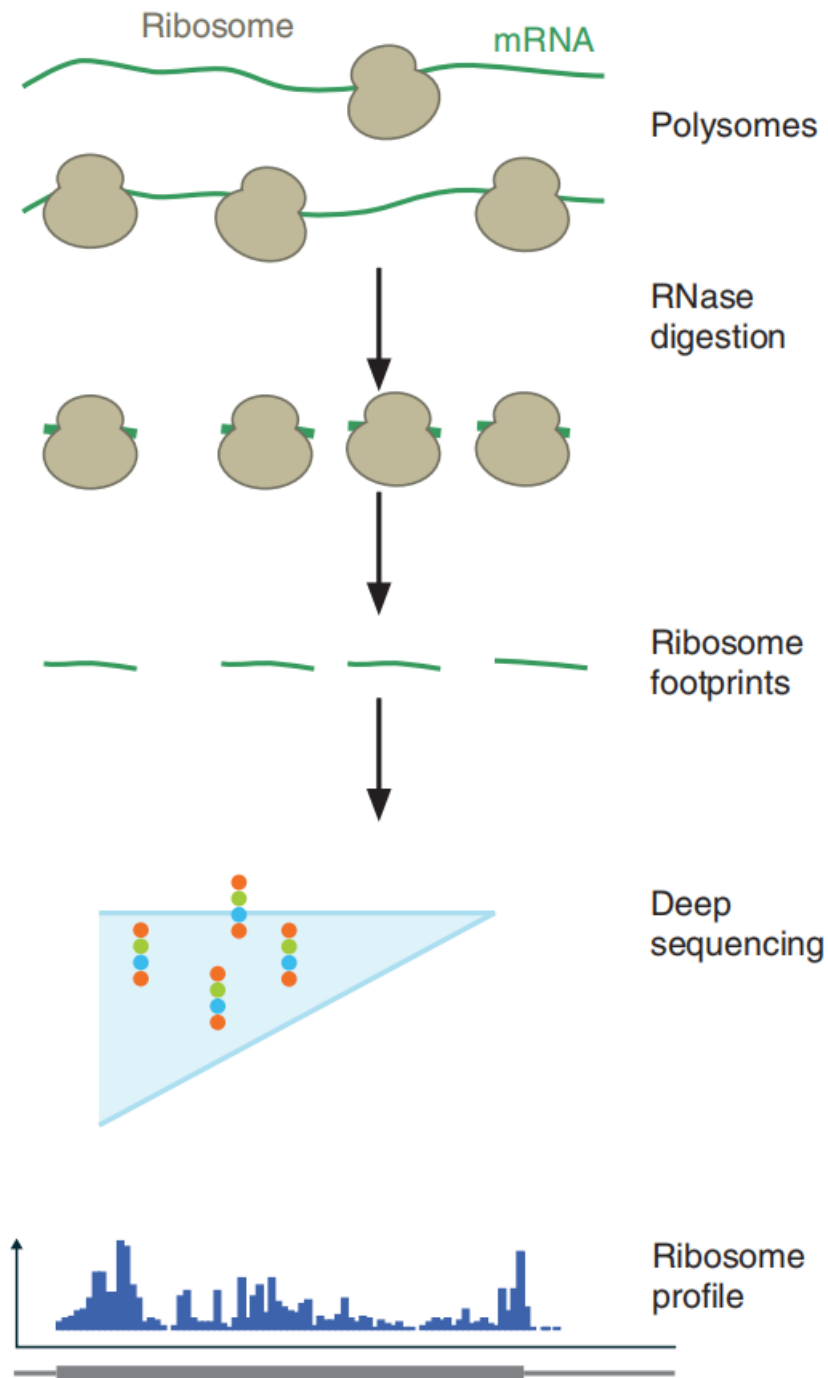


图1.1 Ribo-Seq实验原理[4]

### □ 1.3 生信分析流程

信息分析主要分为7部分：1、数据预处理。去接头序列、污染序列、低质量碱基，获得 clean data 序列，并进行相关数据统计；2、根据提供的rRNA参考序列，去除clean data的rRNA序列，之后定位到参考基因组上，得到 bam文件，并去除重复序列，保留唯一比对的序列；3、分析RPFs（核糖体保护片段）；4、鉴定样本中ORF特别是上游的uORF；5、RPFs表达量统计；6、差异表达基因分析及功能富集分析；7、与mRNA联合分析，计算各组翻译效率与差异以及各基因转录和翻译变化。

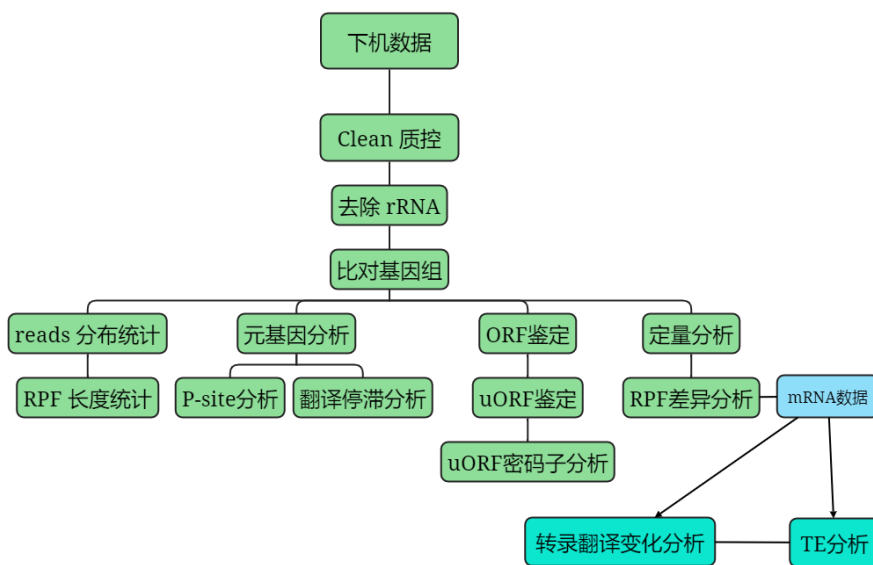


图1.2 Ribo-Seq分析流程

## 1.4 样本信息

进行分析的样本的信息，由客户或实验人员提供样本详细的处理操作信息。

表1.1 样本信息

sampleID	sampleName	description
1	wt_1	wt
2	wt_2	wt
3	h3_1	h3
4	h3_2	h3

## 2 数据处理

### 2.1 fastq文件说明

Illumina等高通量测序平台得到的原始图像数据文件，经碱基识别（Base Calling）转化为原始测序序列（Sequenced Reads），我们称之为 Raw Reads。Raw Reads 以 FASTQ 格式存储，包含序列以及对应的测序质量信息。Fastq 数据示例如下：

```

@IGE:001:HY3LSDMX1:4:1101:15763:1078 1:N:0:AACGTGAT
AATAAGATCGGAAGAGCACACGTCTGAACTCCAGTCACAACGTGATATCTCGTATGCCGTCTTCTGCTTGAAAAGGGGGG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
} 每四行是一个reads

@IGE:001:HY3LSDMX1:4:1101:15763:1078 1:N:0:AACGTGAT
AATAAGATCGGAAGAGCACACGTCTGAACTCCAGTCACAACGTGATATCTCGTATGCCGTCTTCTGCTTGAAAAGGGGGG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
} reads 标题
  reads 序列
  reads 识别码
  reads 质量
  
```

图2.1 FASTQ格式文件说明

每条序列由4行组成，第1行是序列名称；第3行也是序列名称，一般省略，只保留“+”；第2行

是序列；第4行是序列的测序质量。第四行中每个字符对应的ASCII值减去33，即为第二行对应碱基的测序质量值： $Q = -10\log_{10}(E)$ 。

测序错误率与对应字符换算示例如下所示：

表2.1 测序错误率与测序质量值对应关系

测序错误率	测序质量值	字符	ASCII码值
5%	13	.	46
1%	20	5	53
0.10%	30	?	63
0.01%	40	!	73

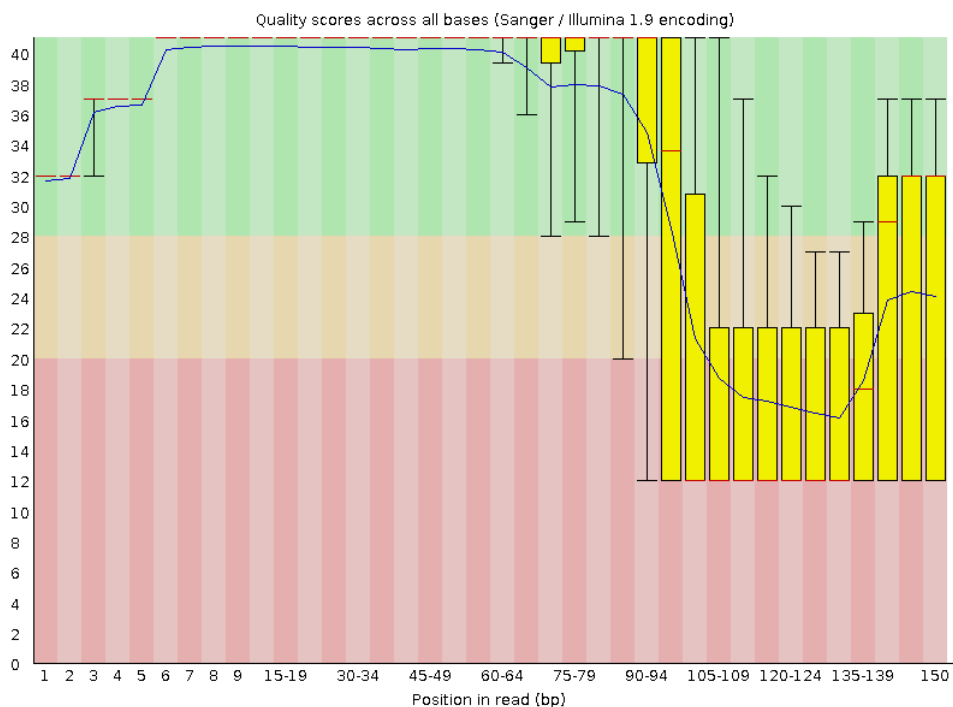
## □ 2.2 Raw Data 数据评估

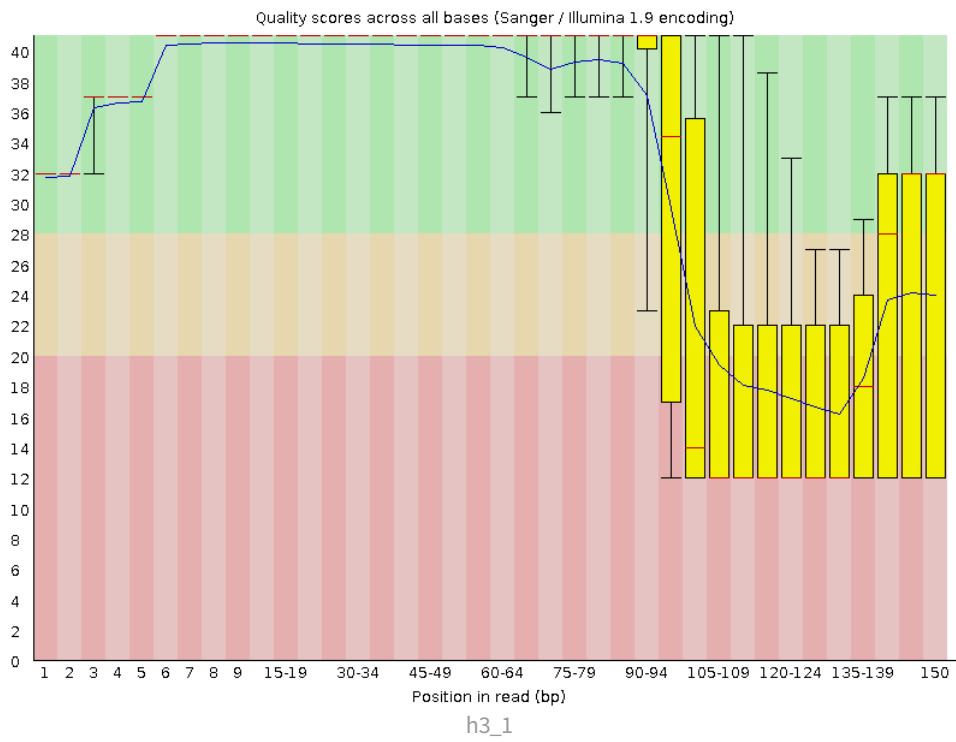
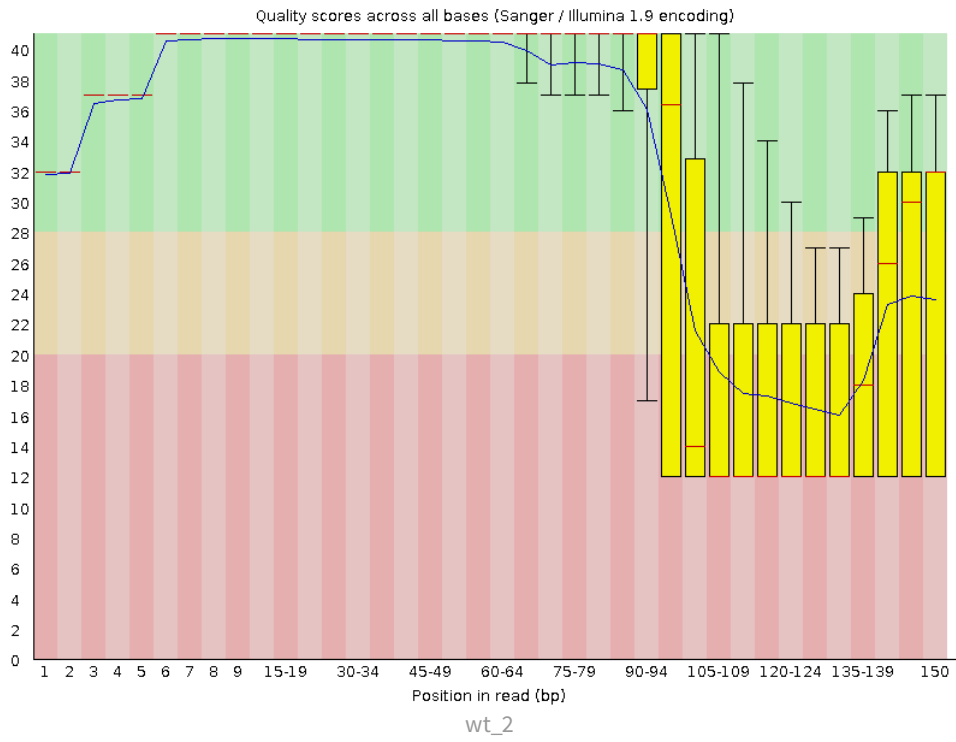
对原始数据，使用软件 FastQC (version: 0.11.5) 质控处理，详细结果见 ./sup/01.qc/raw，结果如下所示。

### ■ 2.2.1 碱基质量

X轴是read中的碱基位置，Y轴是碱基质量；盒形图中间的红线表示中位数 (median value)；黄色部分代表四分位距 (25-75%)；上下分割线代表90%和10%的上下临界值；蓝色的线代表碱基质量的平均值。

碱基质量 (Q值)， $-10 \cdot \log_{10}(p)$ ，p为测错的概率。所以一条read某位置出错概率为0.01时，其 quality就是20，通常认为Q20反映了数据的质量。Y轴将质量值被划分三部分：绿色 (高质量)，橘黄色 (中等质量) 和红色 (低质量)。由于在每一个测序反应开始后，碱基的信号质量会逐渐降低，因此在每个读长最后的碱基通常都会处于橘黄色的中等质量区。





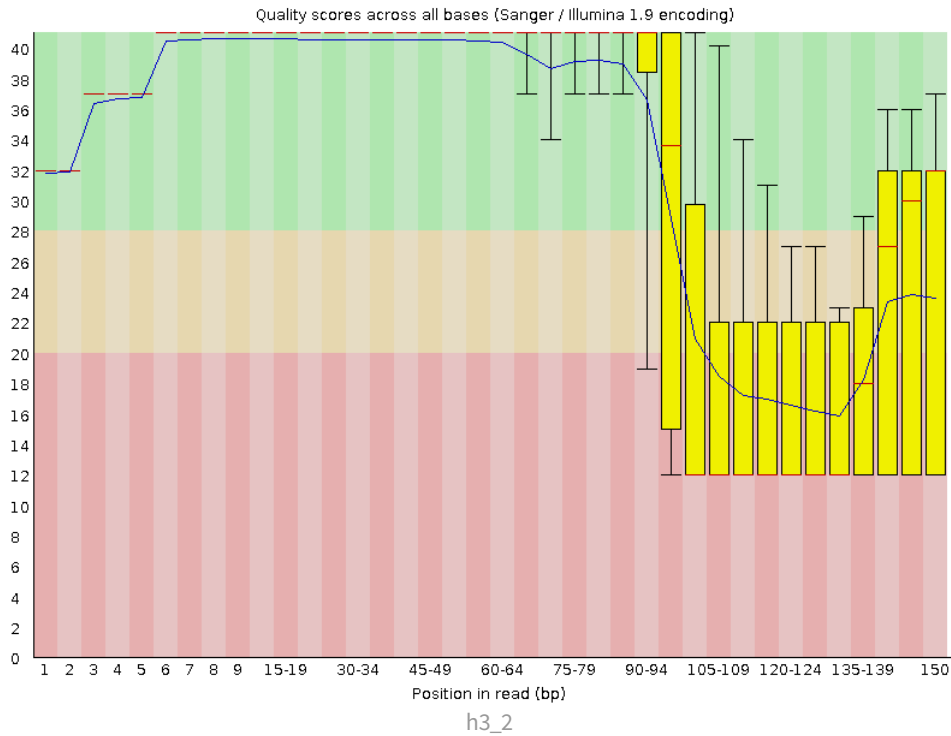
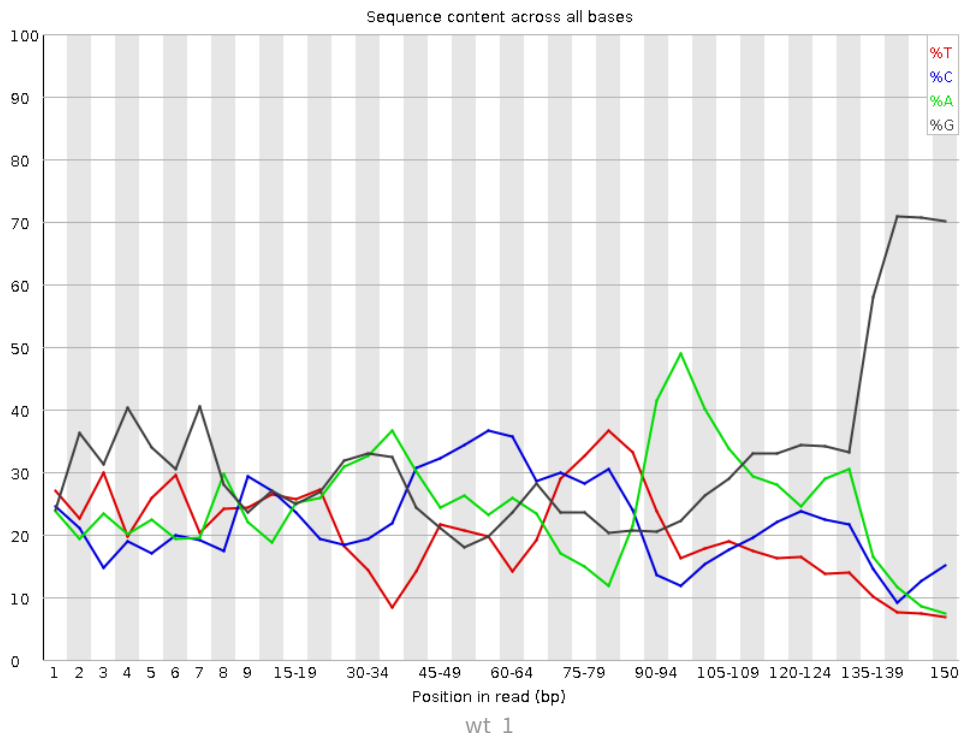
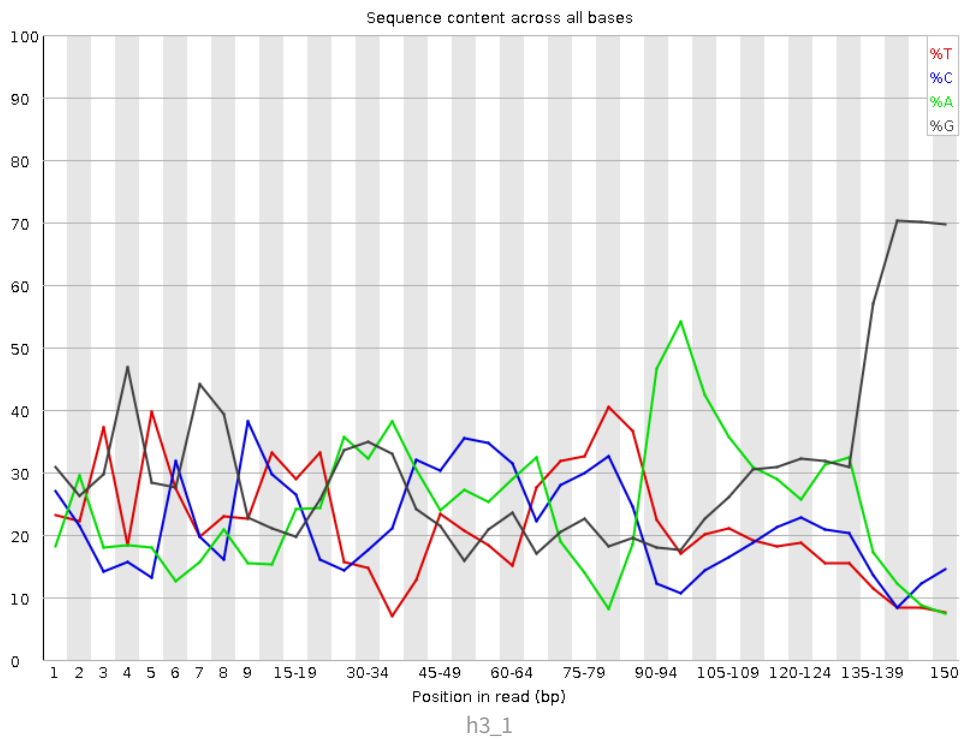
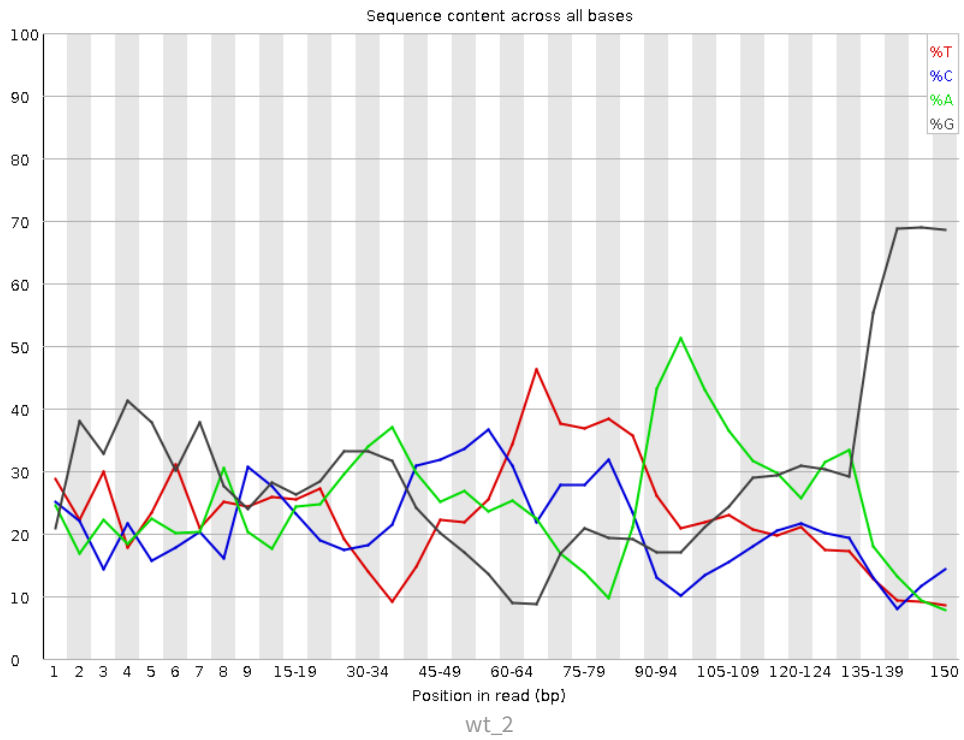


图2.2 原始数据碱基质量分析

## 2.2.2 碱基分布

横坐标为read中的碱基位置，纵坐标为对应位点上单个碱基所占的比例。不同颜色代表不同的碱基类别。正常情况下四种碱基的出现频率应该是接近的，而且没有位置差异。因此好的样本中四条线应该平行且接近。当部分位置碱基的比例出现bias时，即四条线在某些位置纷乱交织，往往提示我们有overrepresented sequence的污染。当所有位置的碱基比例一致的表现出bias时，即四条线平行但分开，往往代表文库有bias (建库过程或本身特点)，或者是测序中的系统误差。





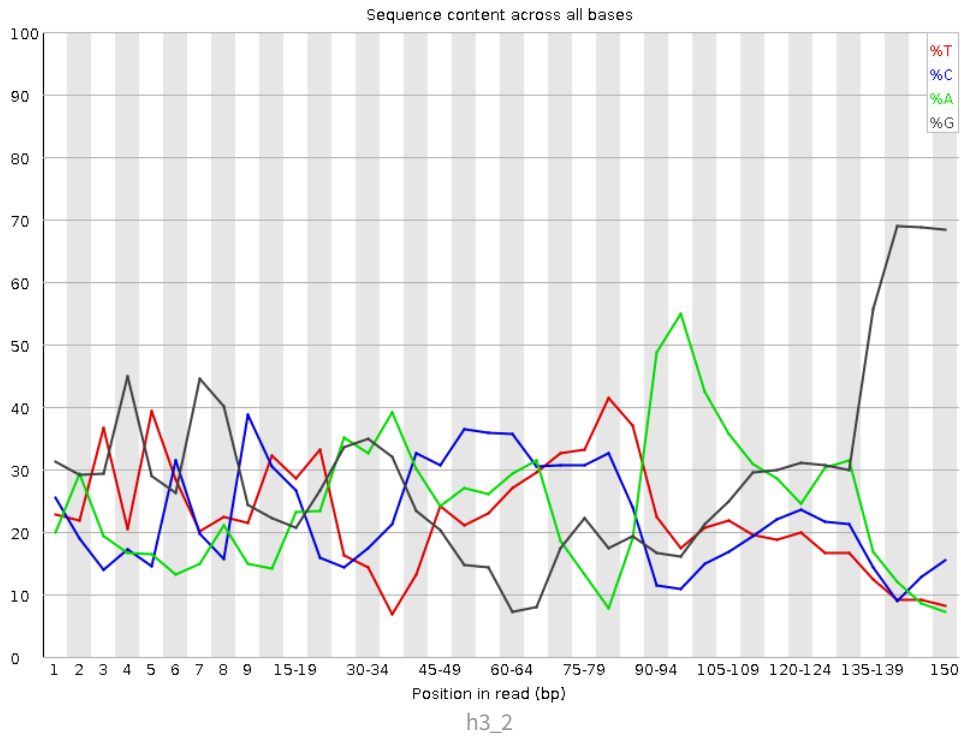
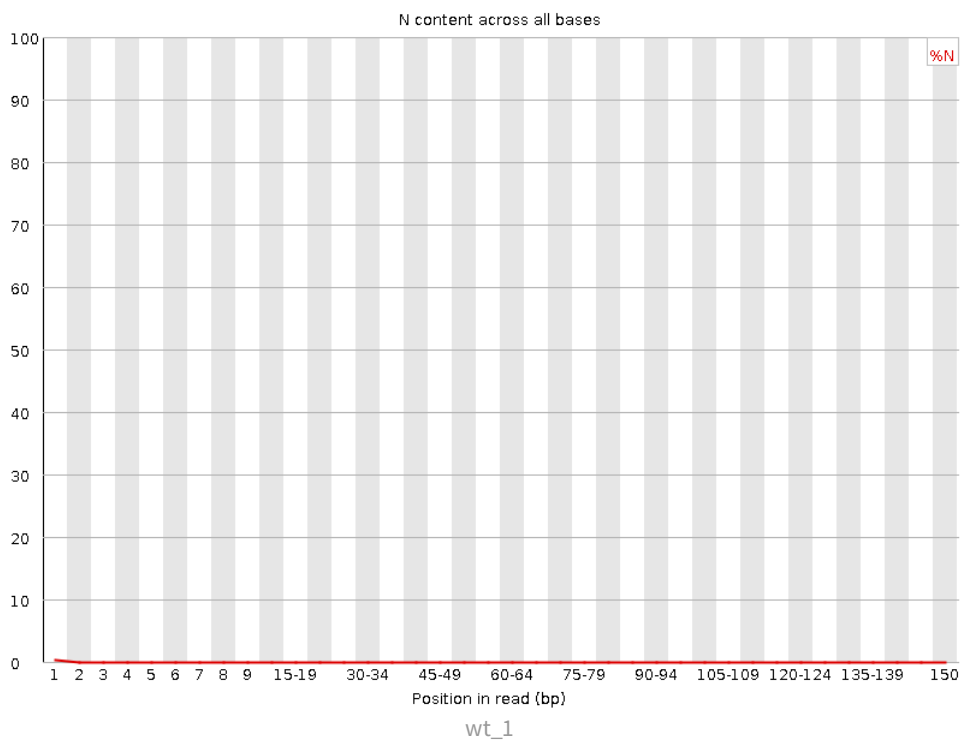
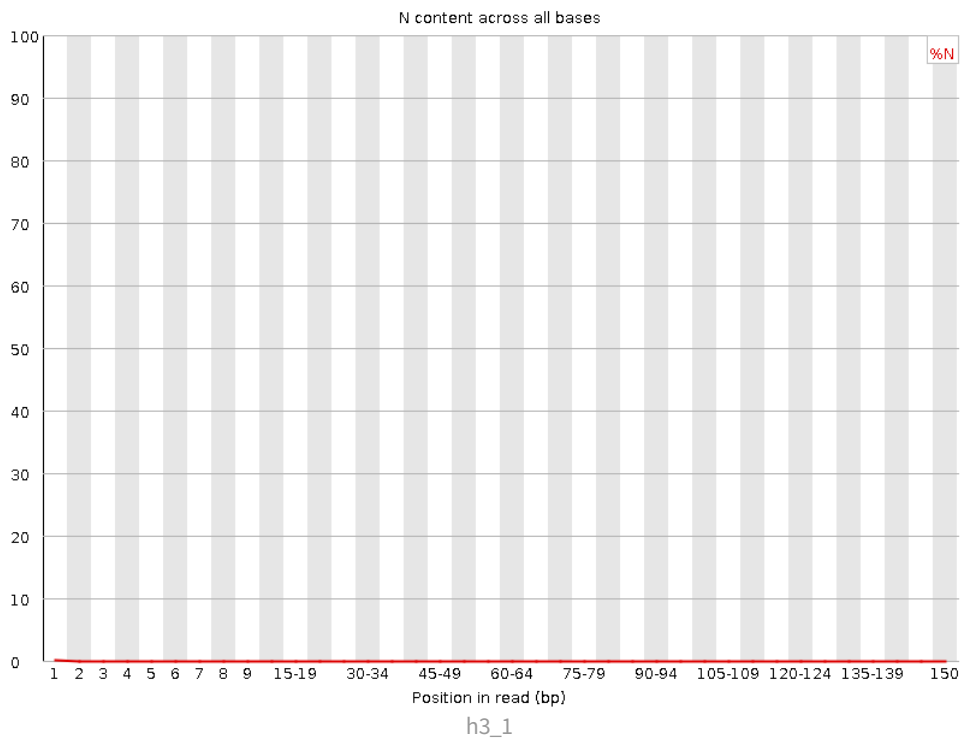
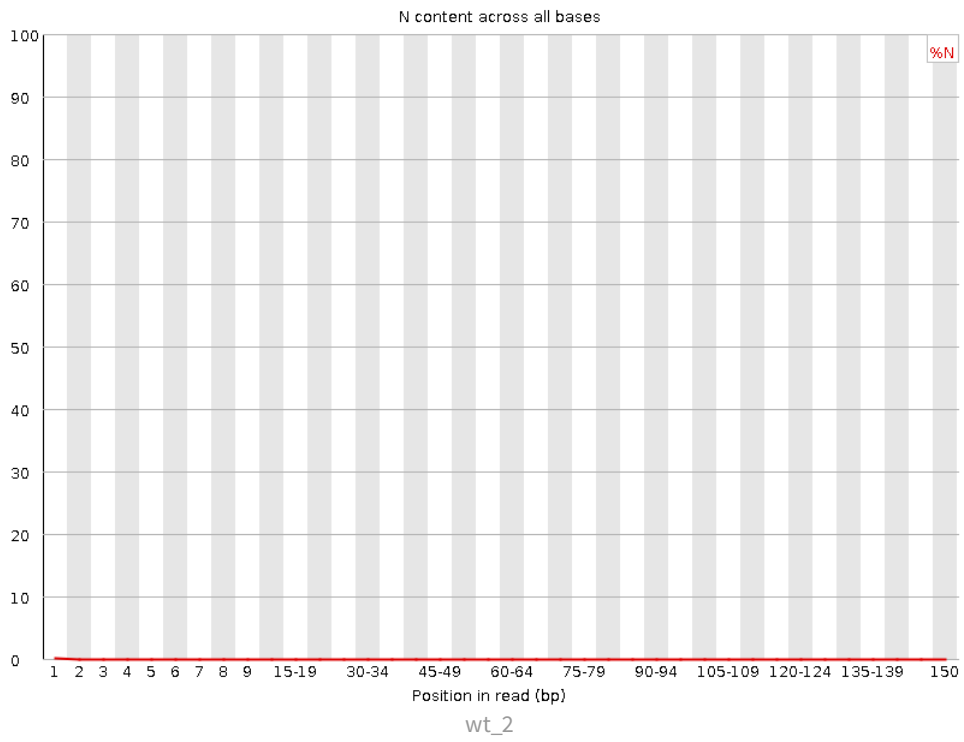


图2.3 原始数据碱基分布图

### 2.2.3 N碱基含量

横坐标为序列长度，纵坐标为N碱基的比例。当测序仪无法识别具体是哪种碱基时，就会给出N，N比例越小越好。当某个位点N碱基的比例大于5%时，会给出警告信息，大于20%时，会给出错误信息。





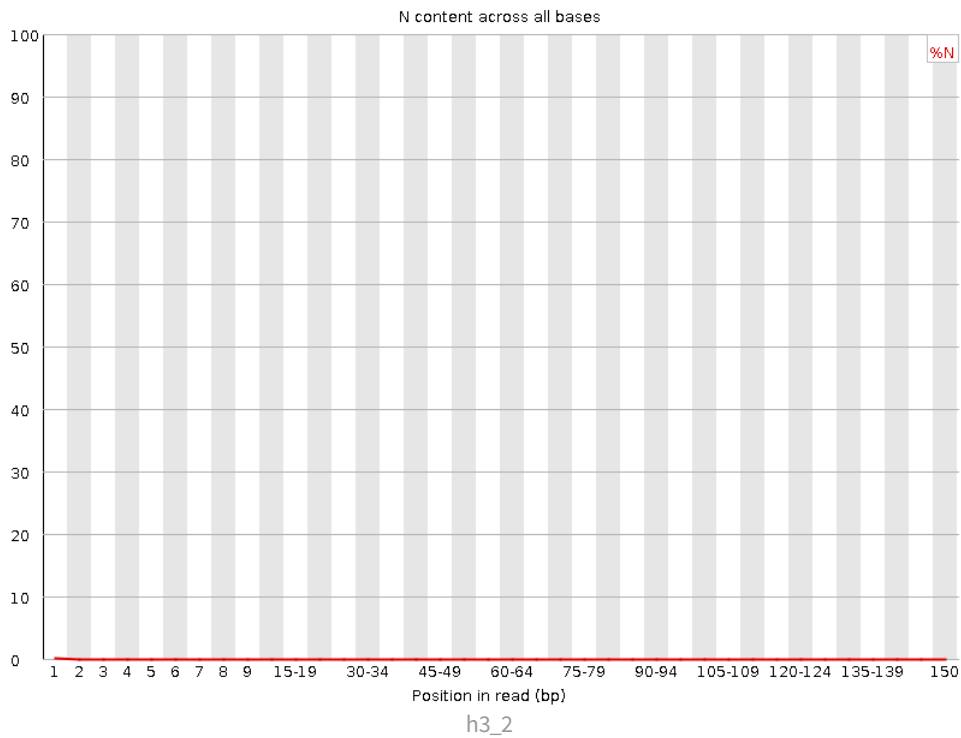


图2.4 原始数据N碱基含量图

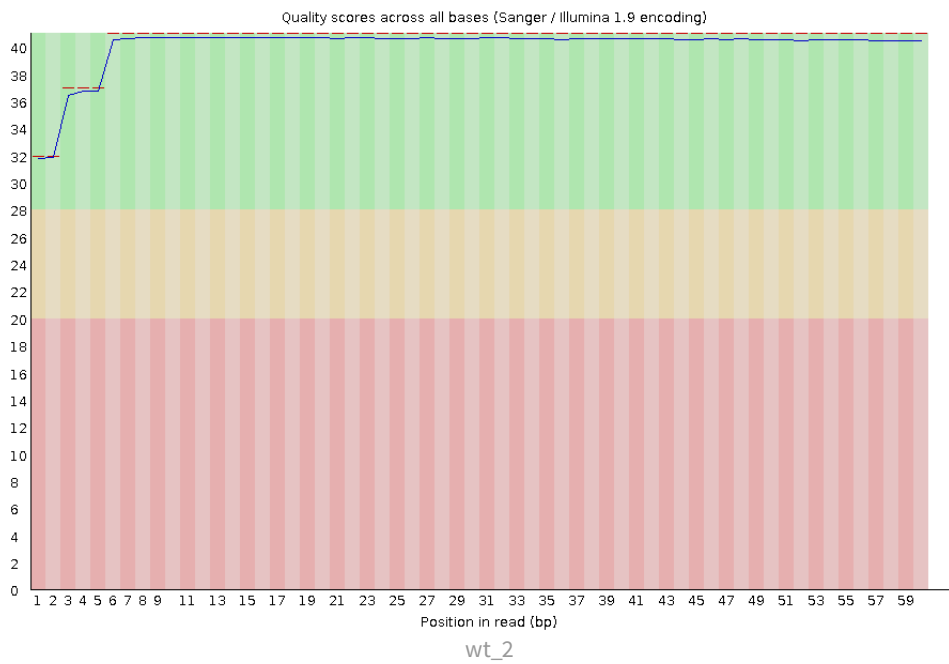
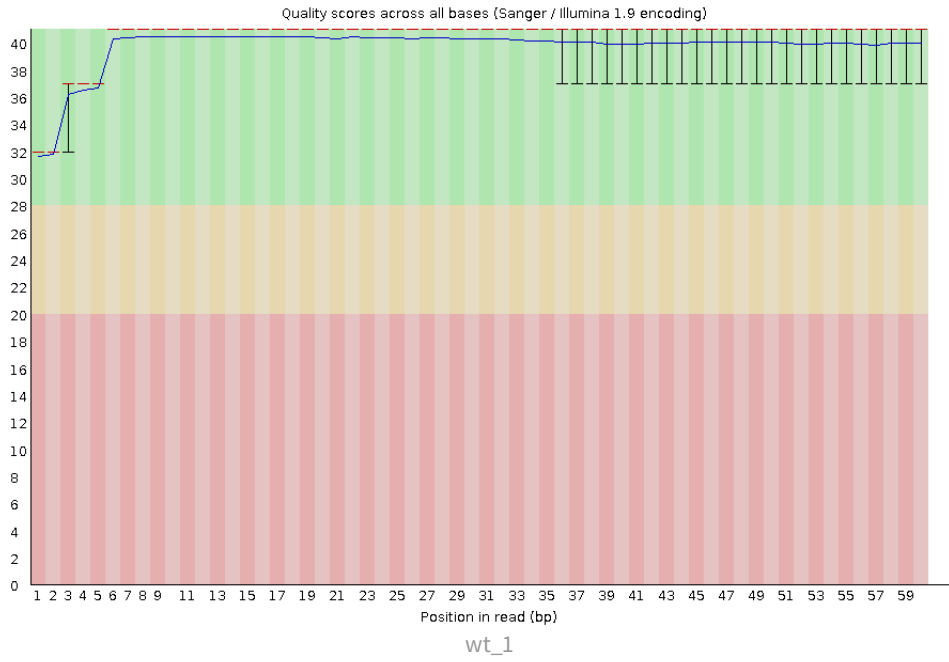
## □ 2.3 clean Data 数据评估

使用fastp软件[5]进行数据过滤，截掉末尾的adapter序列，保留最短长度15bp。对过滤数据，使用软件 FastQC (version: 0.11.5) 质控处理，详细结果见 ./sup/01.qc/clean，结果如下所示。

### ■ 2.3.1 碱基质量

X轴是read中的碱基位置，Y轴是碱基质量；盒形图中间的红线表示中位数 (median value)；黄色部分代表四分位距 (25-75%)；上下分割线代表90%和10%的上下临界值；蓝色的线代表碱基质量的平均值。

碱基质量 (Q值)， $-10 \cdot \log_{10}(p)$ ，p为测错的概率。所以一条read某位置出错概率为0.01时，其quality就是20，通常认为Q20反映了数据的质量。Y轴将质量值被划分三部分：绿色 (高质量)，橘黄色 (中等质量) 和红色 (低质量)。由于在每一个测序反应开始后，碱基的信号质量会逐渐降低，因此在每个读长最后的碱基通常都会处于橘黄色的中等质量区。



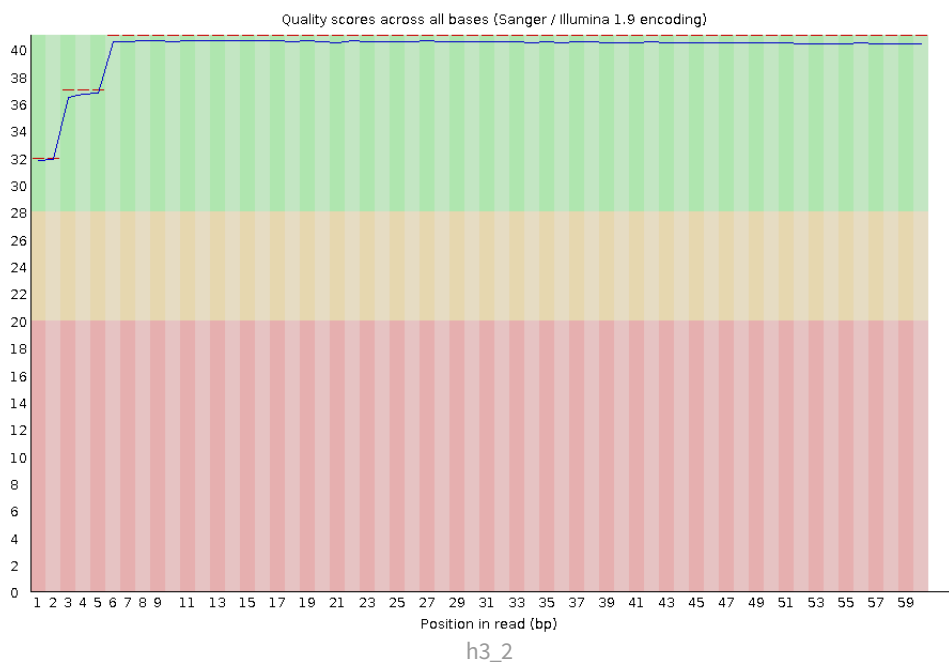
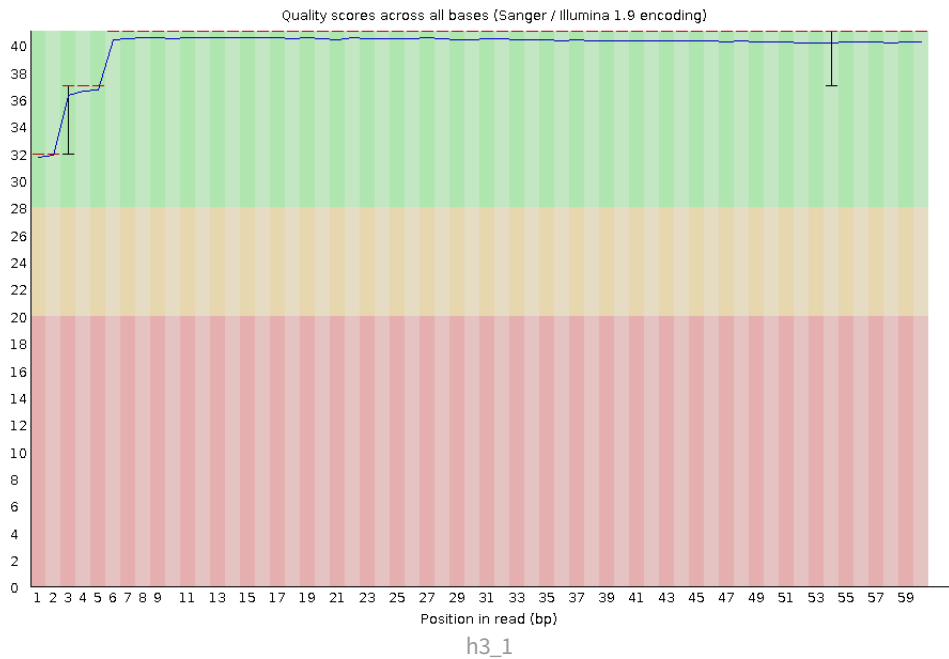
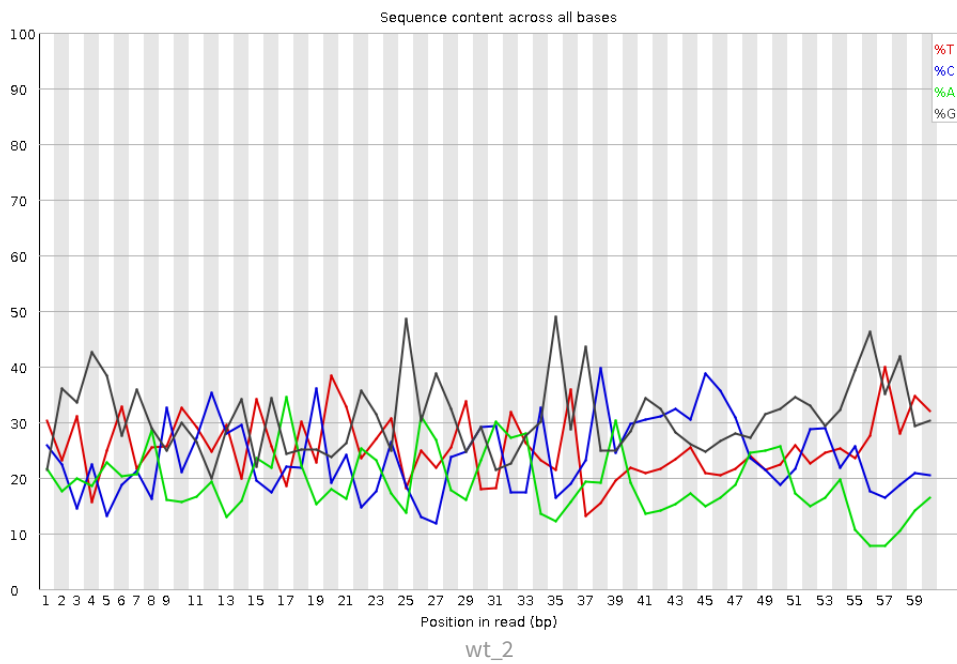
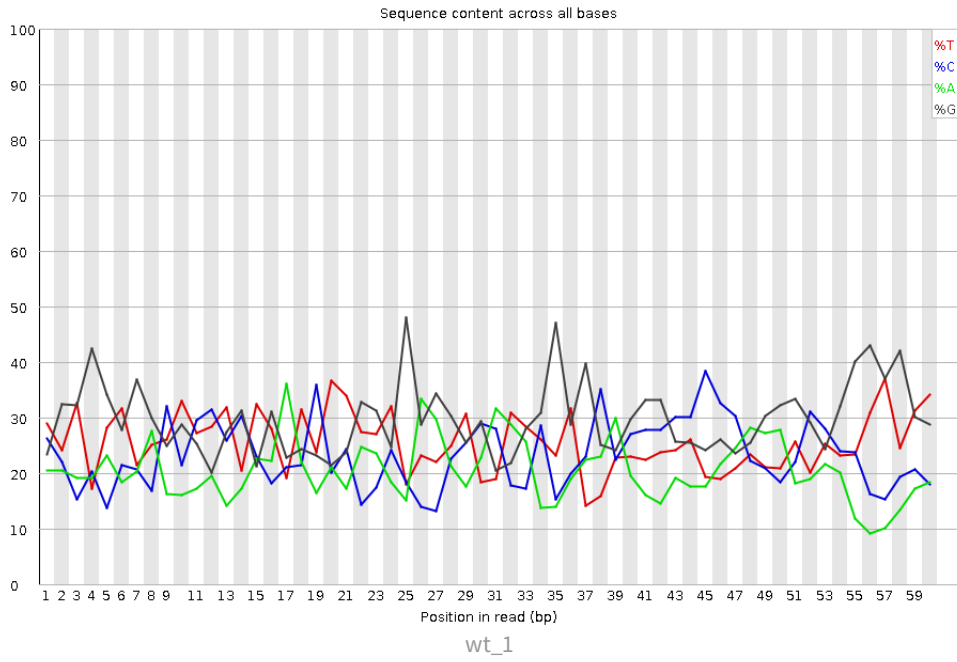


图2.5 过滤数据碱基质量分析

### 2.3.2 碱基分布

横坐标为read中的碱基位置，纵坐标为对应位点上单个碱基所占的比例。不同颜色代表不同的碱基类别。正常情况下四种碱基的出现频率应该是接近的，而且没有位置差异。因此好的样本中四条线应该平行且接近。当部分位置碱基的比例出现bias时，即四条线在某些位置纷乱交织，往往提示我们有overrepresented sequence的污染。当所有位置的碱基比例一致的表现出bias时，即四条线平行但分开，往往代表文库有bias (建库过程或本身特点)，或者是测序中的系统误差。



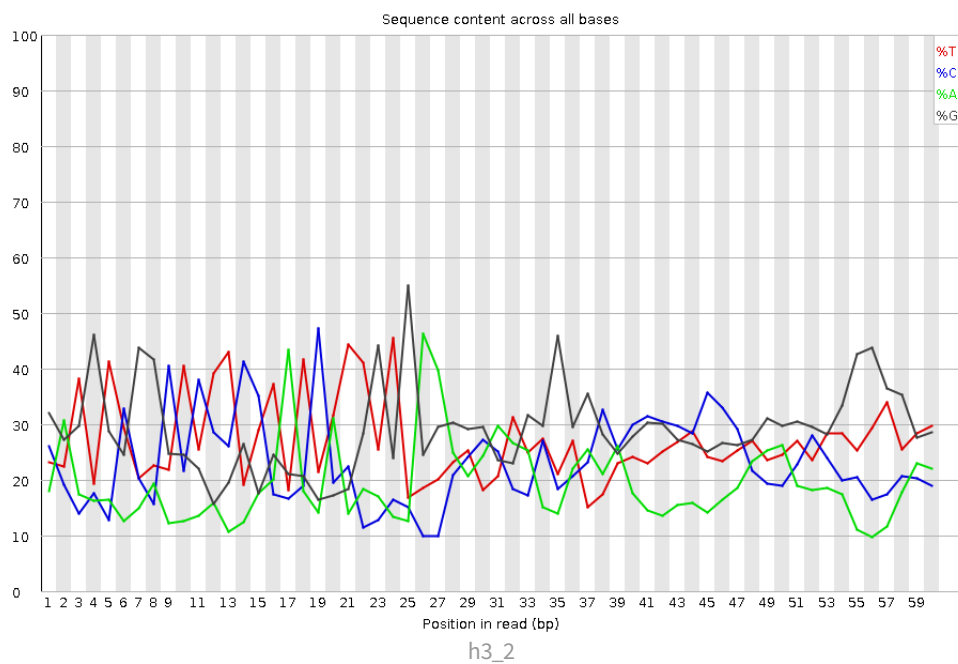
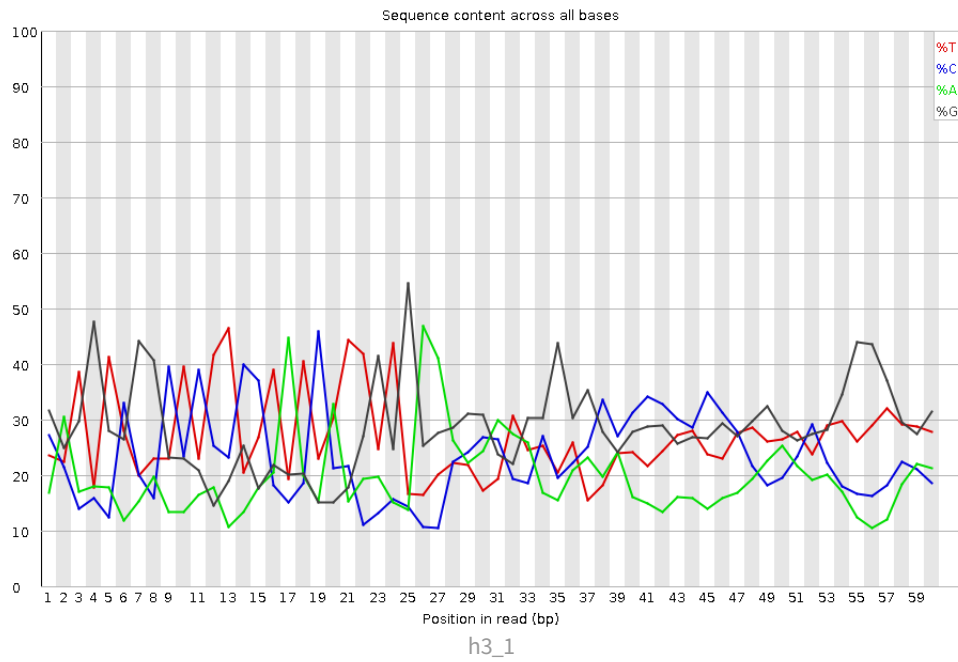
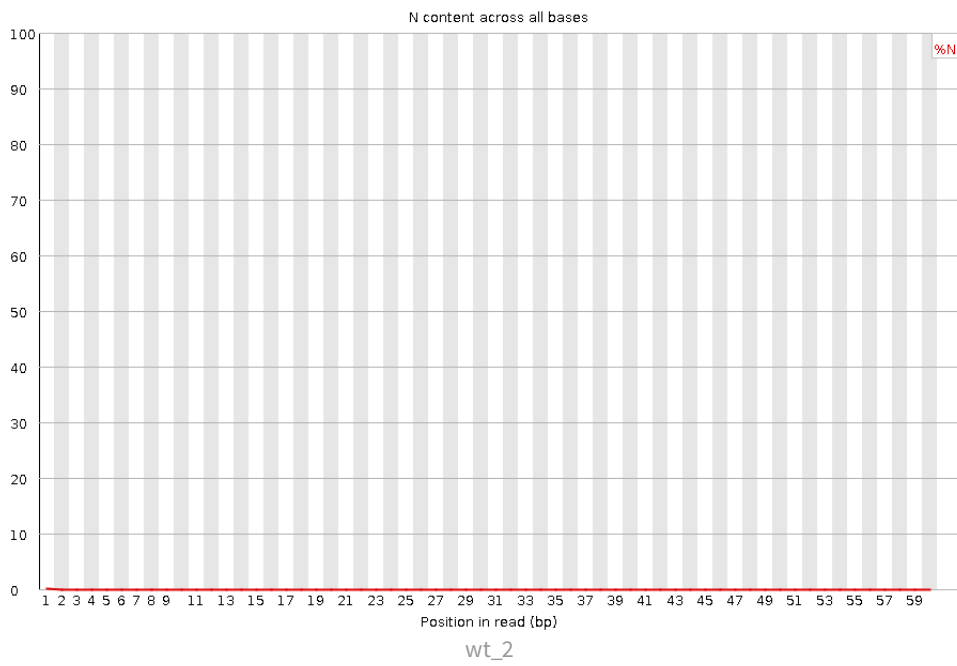
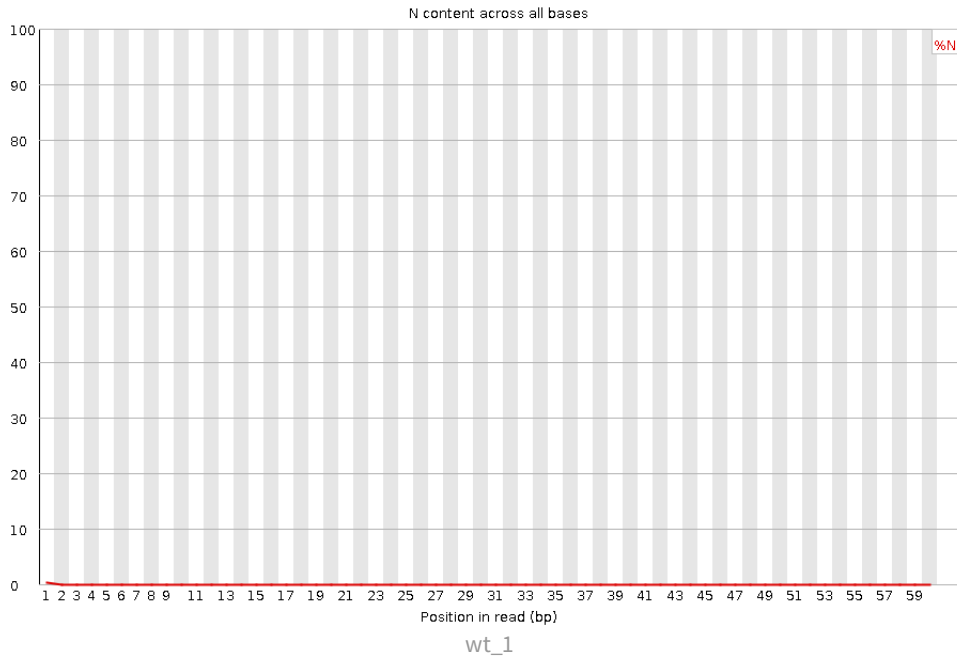
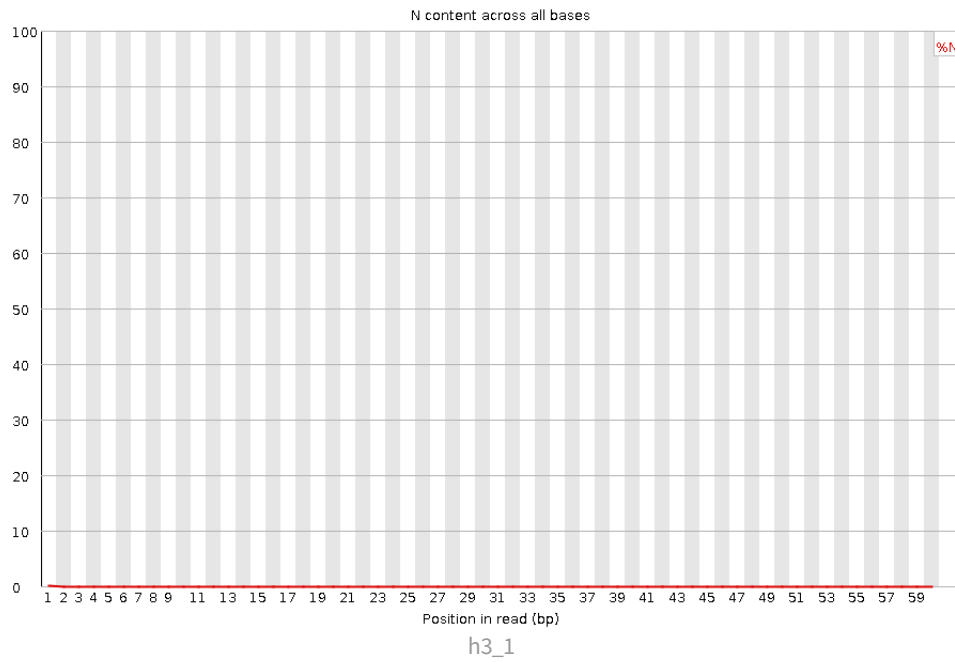


图2.6 原始数据碱基分布图

### 2.3.3 N碱基含量

横坐标为序列长度，纵坐标为N碱基的比例。当测序仪无法识别具体是哪种碱基时，就会给出N，N比例越小越好。当某个位点N碱基的比例大于5%时，会给出警告信息，大于20%时，会给出错误信息。





●

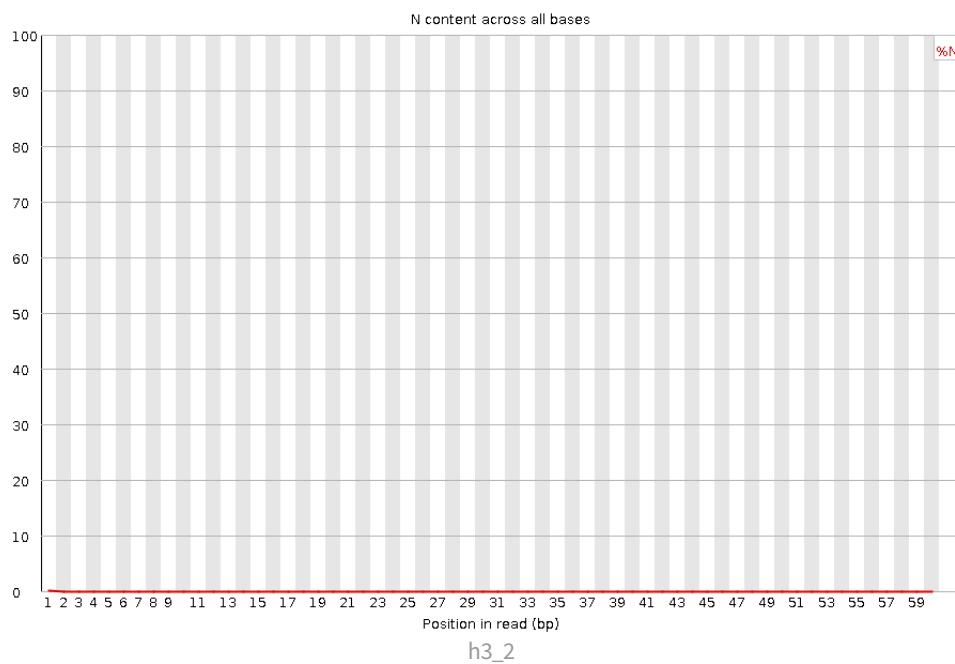


图2.7 过滤数据N碱基含量图

## □ 2.4 Clean 结果统计

对原始数据和过滤处理得到高质量的数据（clean reads）进行数据统计，得到数据基本信息。统计结果如下表。

- 1 背景及分析流程
- 2 数据处理
- 3 比对分析
- 4 RPF分析
- 5 翻译停滞

表2.2 clean data数据统计表

2 翻译注释

分析

6 ORF分析

7 表达量分

析

8 差异表达

分析

9 与转录组

联合分析

10 差异基因

富集分析

11 备注

12 参考文献

sample	rawReads	rawBase(G)	cleanReads	cleanBase(G)	cleanQ20(%)	clean
1	37912738	5.69	34872569(91.98%)	1.02(17.93%)	99.66	98.8
2	29446028	4.42	27292871(92.69%)	0.78(17.65%)	99.74	99.0
1	28920085	4.34	25088318(86.75%)	0.72(16.59%)	99.56	98.5
2	27369933	4.11	24625016(89.97%)	0.71(17.27%)	99.79	99.2

## 量评估Q&A

错误率较高，错误率过滤阈值一般是多少？

单碱基的测序错误率应该低于1%（即Q20，对应字符"5"），末端最高在5%（即Q13，对应字符"."）。

reads是否合格？

面进行判断：

示：碱基质量整体分布在Q20以上，末端可能在Q20以下；

TCG分布较为均衡，A和T含量接近，C和G含量接近，在reads开始部分可能有由于引物偏好性而出现不平

N碱基不在reads中间部分明显出现。

clean后的reads满足上述条件，即可进入后续分析。

## 3 比对分析

### □ 3.1 去除rRNA

测序得到的reads中有很多rRNA序列，为了去除它们，我们使用软件bowtie（version :1.3.1）[7]将测序数据与参考rRNA序列进行比对，只保留没有比对成功的序列（即非rRNA序列）进行下一步分析。

### □ 3.2 比对参考基因组

过滤rRNA之后，我们将测序数据与参考基因组比对，使用软件为STAR（version:2.7.11a）[8]。我们将对比对得到的唯一比对reads用于后续分析。

表3.1 STAR比对情况

sample	after_remove_rRNA_reads	uniqMapReads	mapReads
h3_1	28280575	3293940(11.65%)	26562172(93.92%)
h3_2	22717032	2560558(11.27%)	21662908(95.36%)
wt_1	19566782	4613070(23.58%)	18541843(94.76%)
wt_2	18977066	4130649(21.77%)	17971346(94.70%)

1. sample: 样本名
2. after\_remove\_rRNA\_reads: 去除rRNA后reads总数
3. uniqMapReads: 唯一比对无重复reads数量
4. mapReads: 比对上参考基因组reads总数

Q: 基础分析都需要基因组什么文件？

A: 对应物种的参考基因组文件（fasta格式）及基因结构注释文件（gff文件）及GOKEGG注释文件（文本格式）。

Q: 造成比对率较低的原因可能有哪些?

A: 当比对率较低时主要可能有2个原因:

- 1.由于参考基因组组装不好, 或者所测物种与参考基因组的物种的亲缘关系较远;
- 2.由于样品的特殊前处理或者相对于参考基因组此样品本身的变异太大, 导致比对率相对较低。

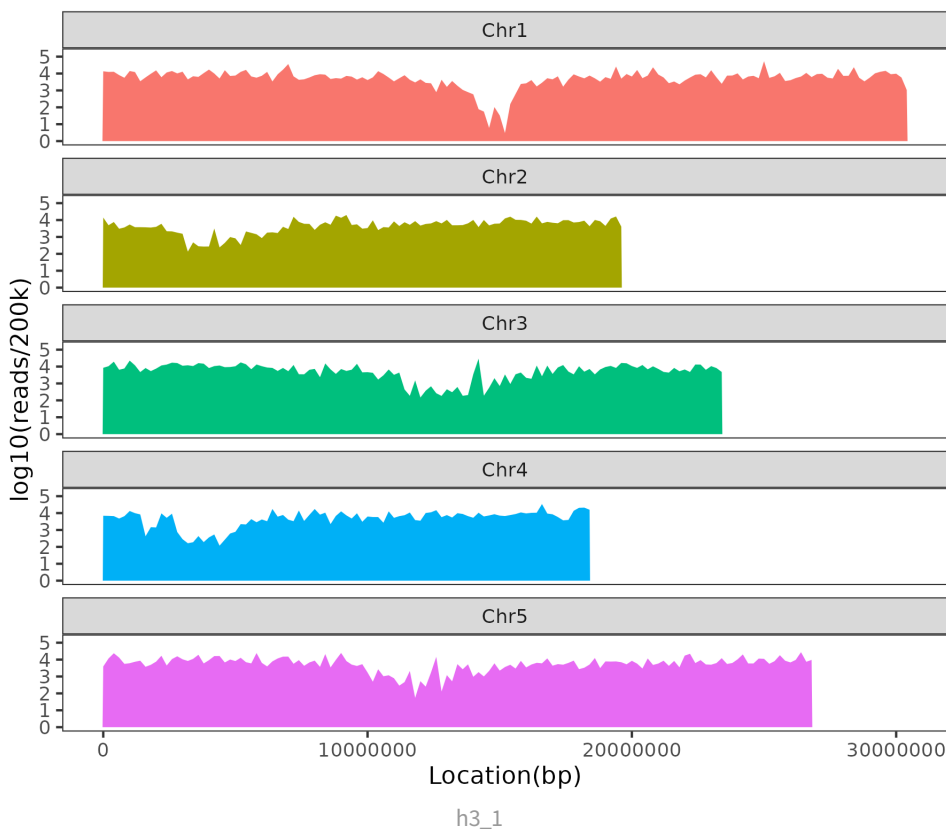
Q: 重复reads和非uniq比对的reads是什么, 为什么要去掉?

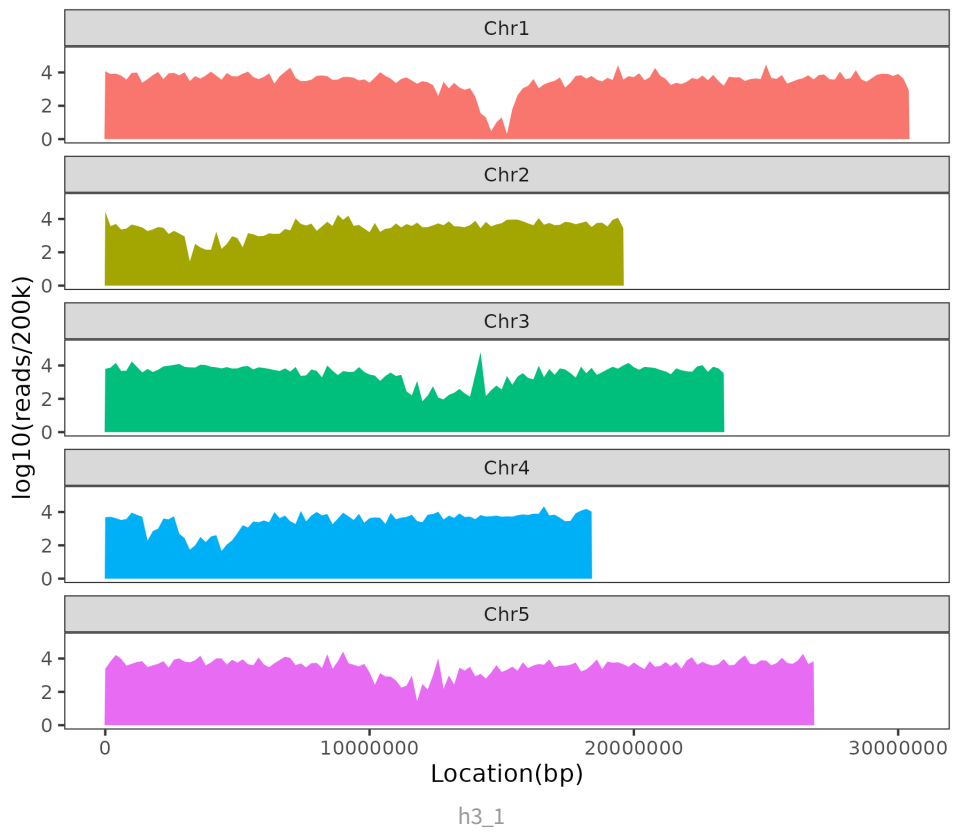
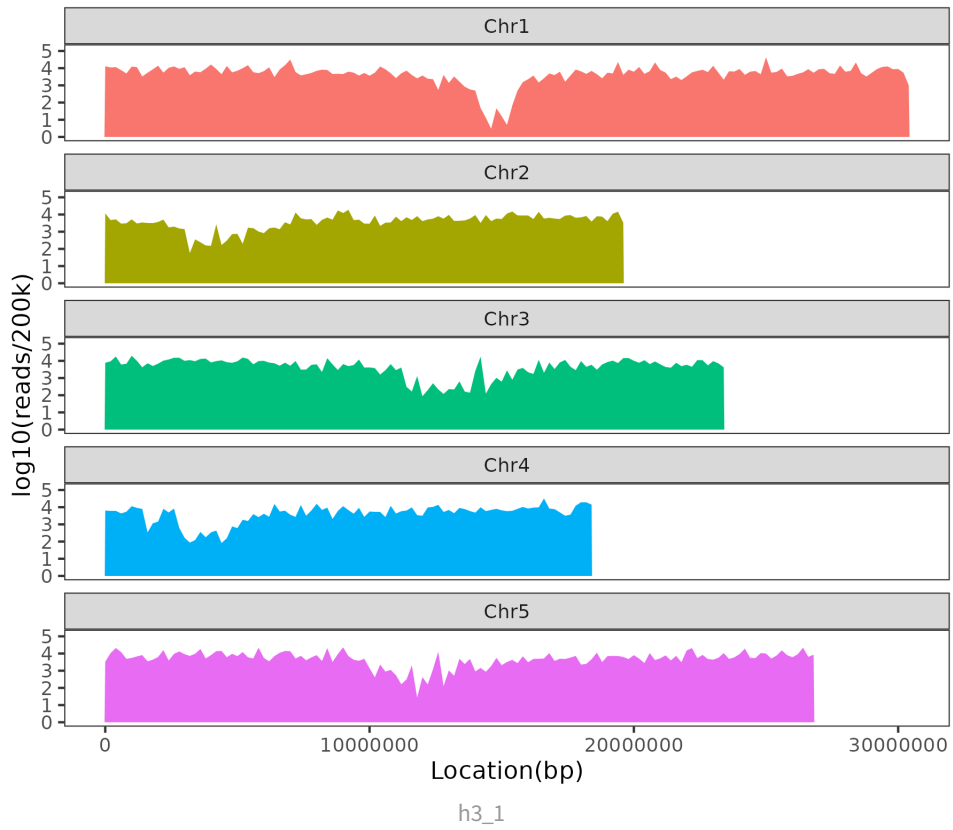
A: 重复reads是指reads的序列一样, 多是由于PCR扩增引起的, 这些reads不能代表样本真实状态, 所以需要去掉。

非uniq比对的reads是指reads可以比对到基因组上多个位置, 因为参考基因组上可能存在重复片段和同源区域, 比对到这些区域的reads我们并不能确定reads到底来自什么地方, 所以也会去掉。

### □ 3.3 reads在染色体上分布

为了直观的展示染色体长度和Reads总数以及Reads覆盖情况, 将比对到基因组上各条染色体的Reads进行唯一比对和去重处理后进行统计, 以200k长度为1个bin计算各个bin内部比对到参考基因组上的Reads数目的均值(作图时取对数)。通常情况下染色体越长定位到该染色体的Reads数目越多, DNA项目的分布较为均匀, RNA项目由于获得的多是exon区域, 分布有一定间隔。如果染色体数目过多, 则只会展示部分染色体。reads在整个基因组上的分布详细结果见./sup/02.map/readsOnGenome, 结果如图所示。





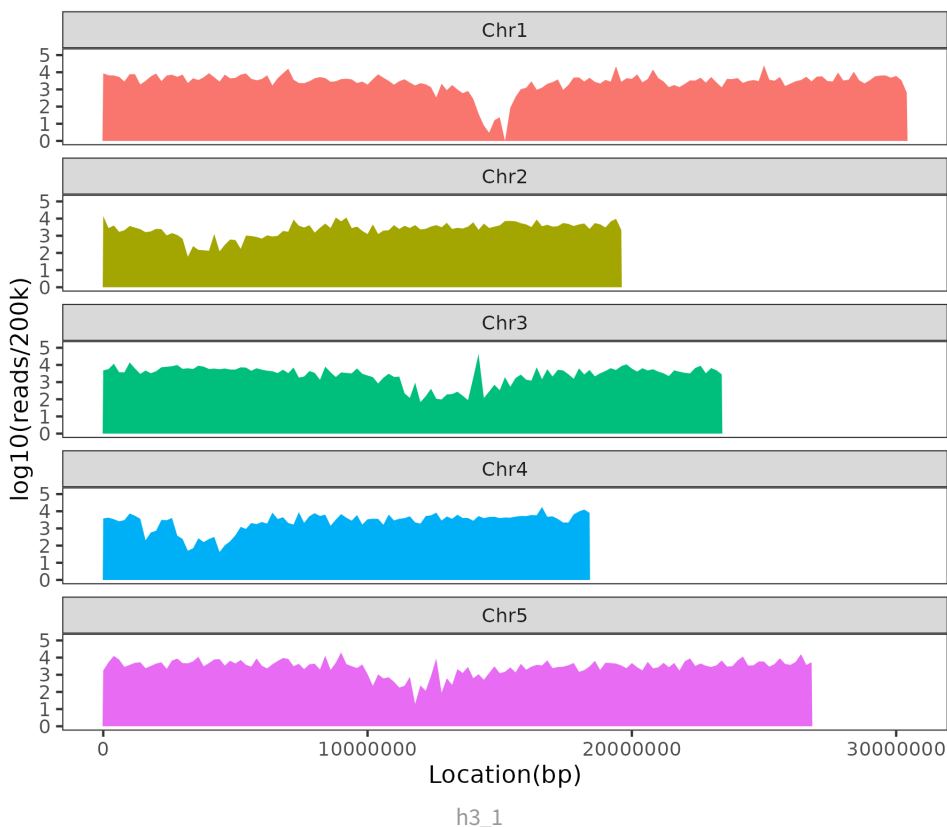


图3.1 reads on genome

### □ 3.4 reads 区域分布

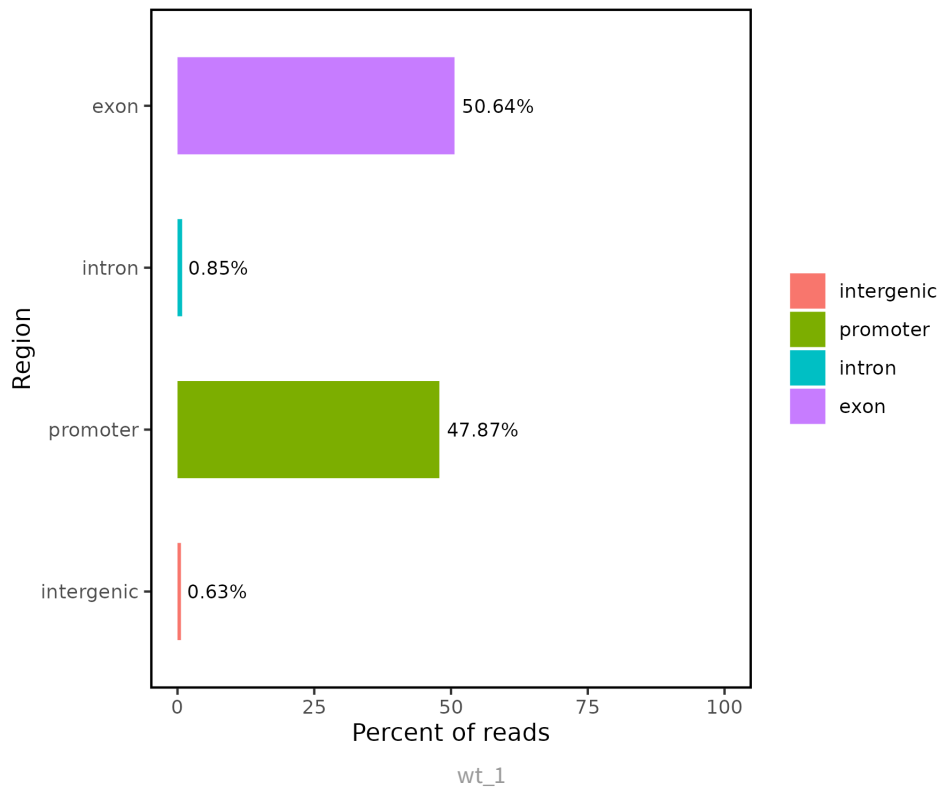
统计reads在基因组上各类元件的分布，RNA项目reads主要分布在CDS和intron区，详细结果见./sup/02.map/mapRegion，结果如图所示。

表3.2 基因组各区域分布统计

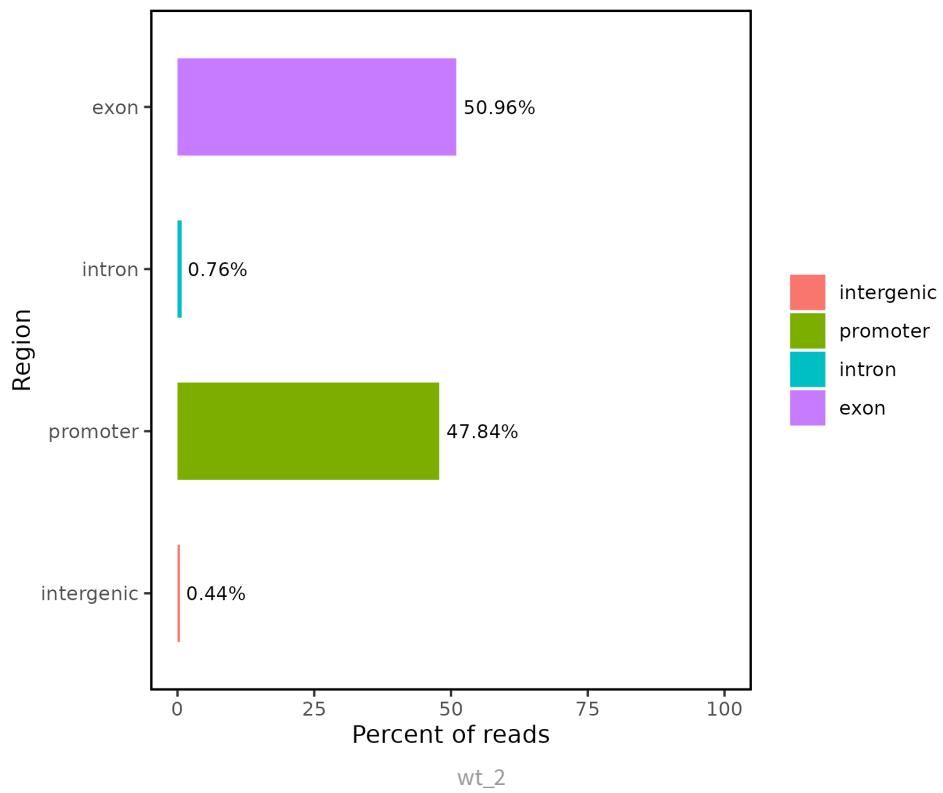
Sample	exon	intergenic	intron	promoter
wt_2	50.96%	0.44%	0.76%	47.84%
wt_1	50.64%	0.63%	0.85%	47.87%
h3_1	50.81%	0.53%	0.73%	47.93%
h3_2	50.66%	0.57%	0.73%	48.03%

•

reads region distribution



reads region distribution



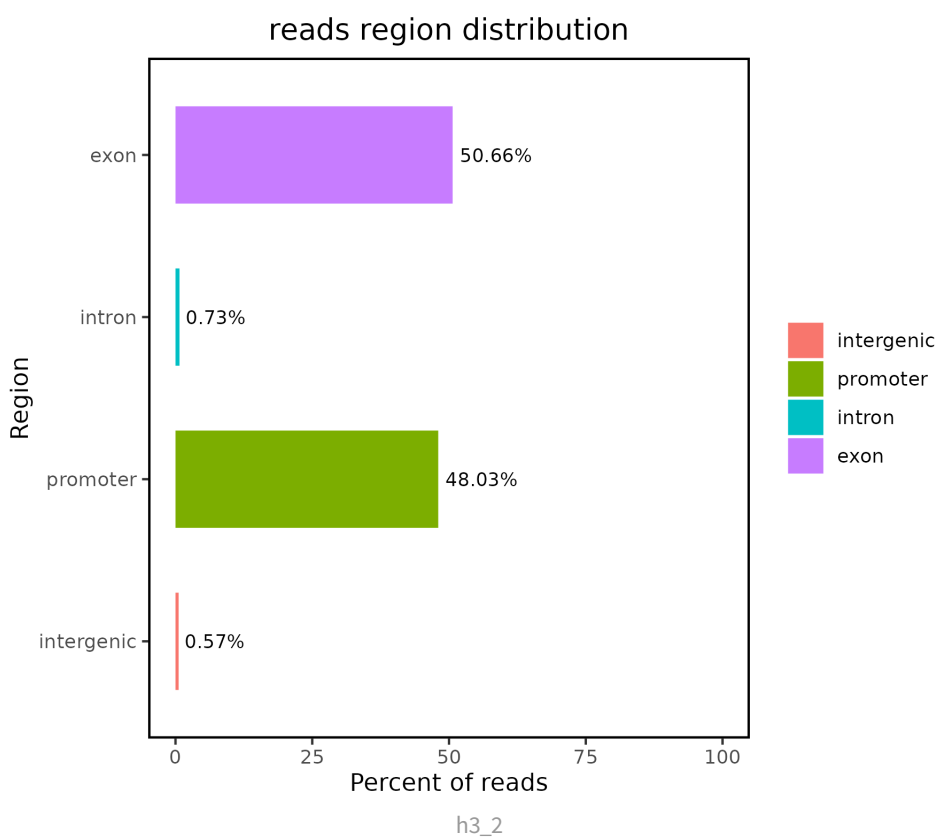
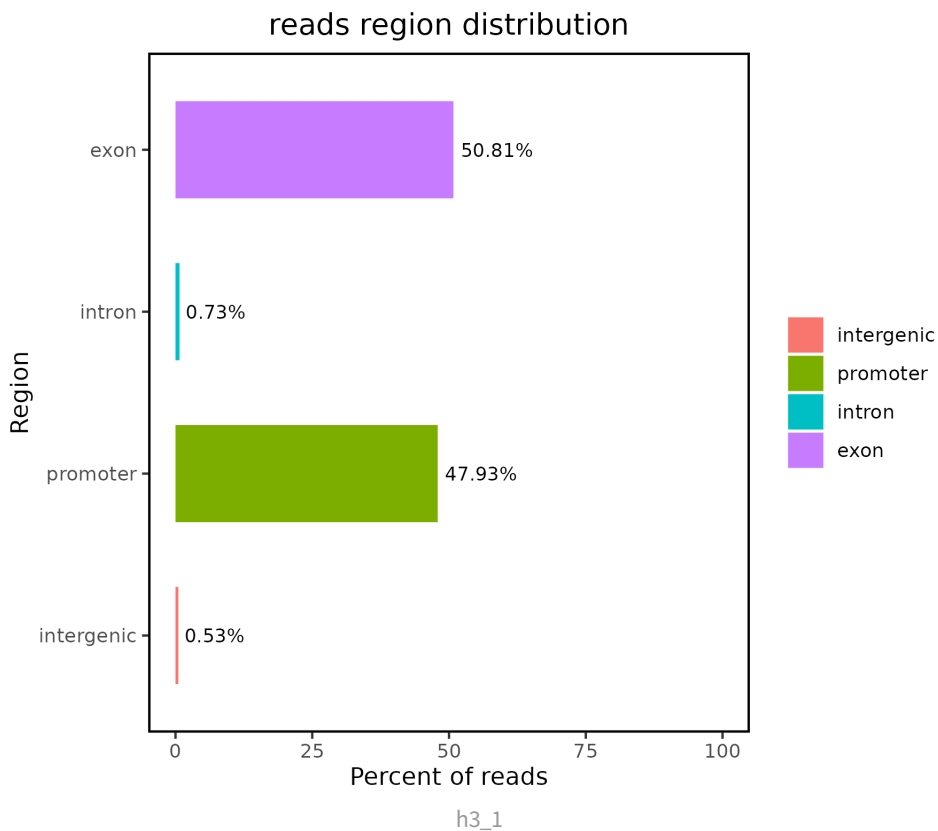


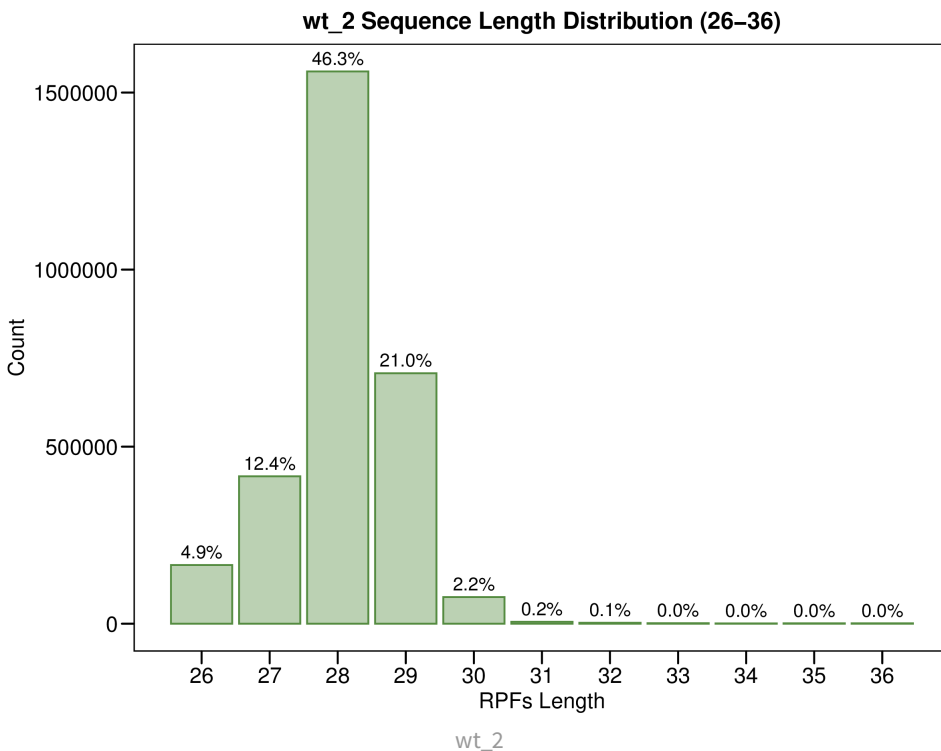
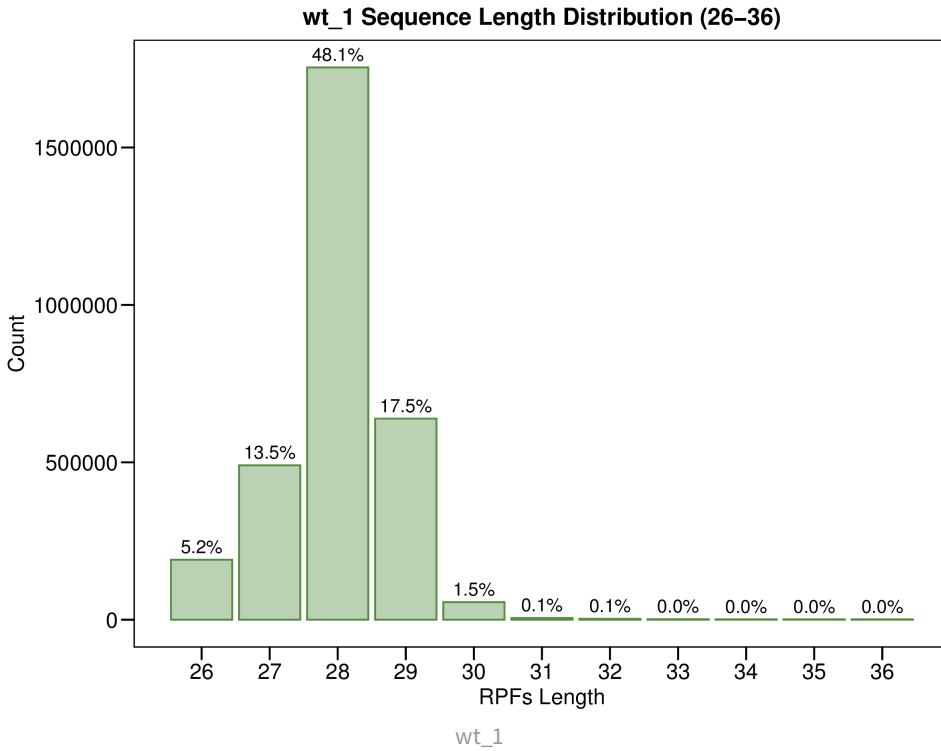
图3.2 reads region

## 4 RPF分析

### □4.1 RPF长度统计

RPF长度普遍在28-30 nt左右，我们统计每个样本RPF长度并绘制直方图，如果RPF长度集中在过长

或过短的区间，可能代表实验结果不理想。Frame0：从mRNA的第一个核苷酸开始翻译，每三个核苷酸为一组，形成一个阅读框；Frame1：从mRNA的第二个核苷酸开始翻译，同样每三个核苷酸为一组，形成另一个阅读框；Frame2：从mRNA的第三个核苷酸开始翻译，同样每三个核苷酸为一组，形成第三个阅读框。



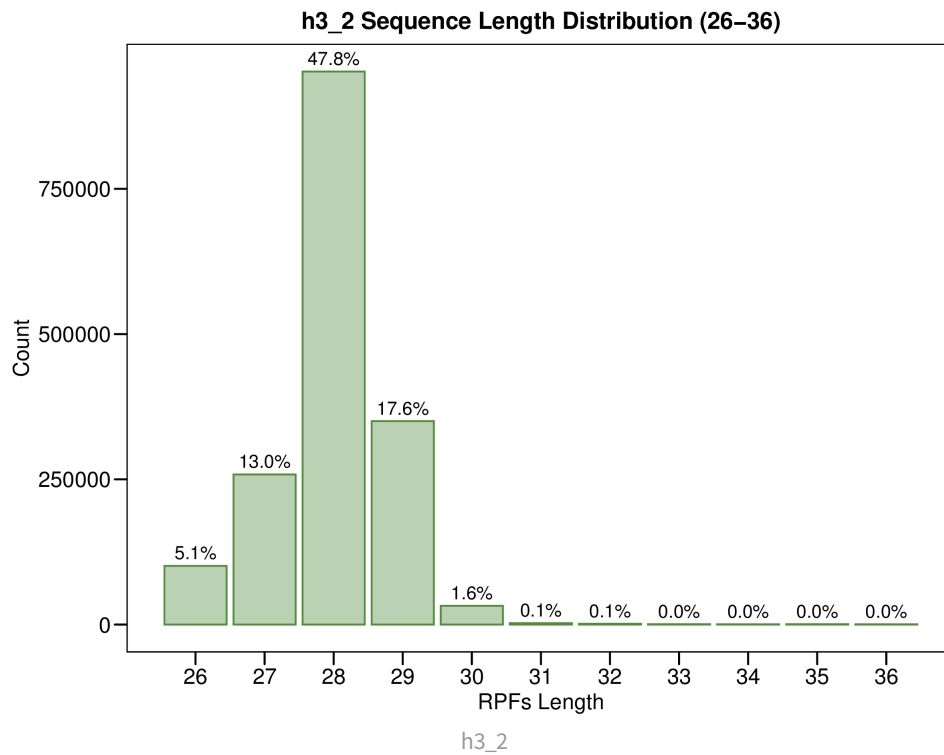
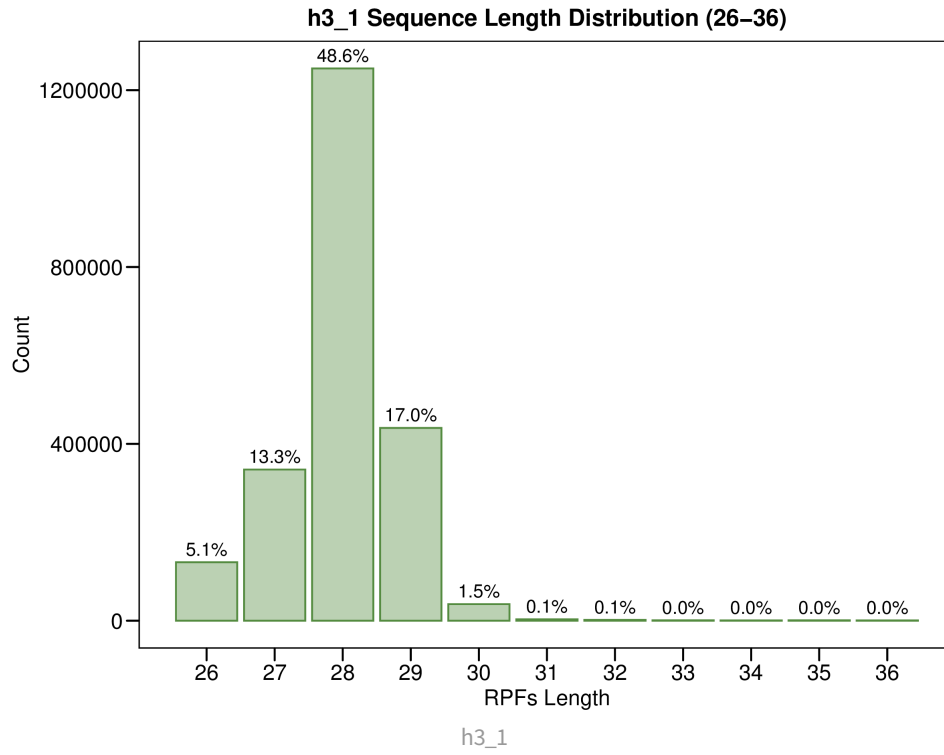


图4.1 RPF长度直方图

表4.1 RPF长度统计

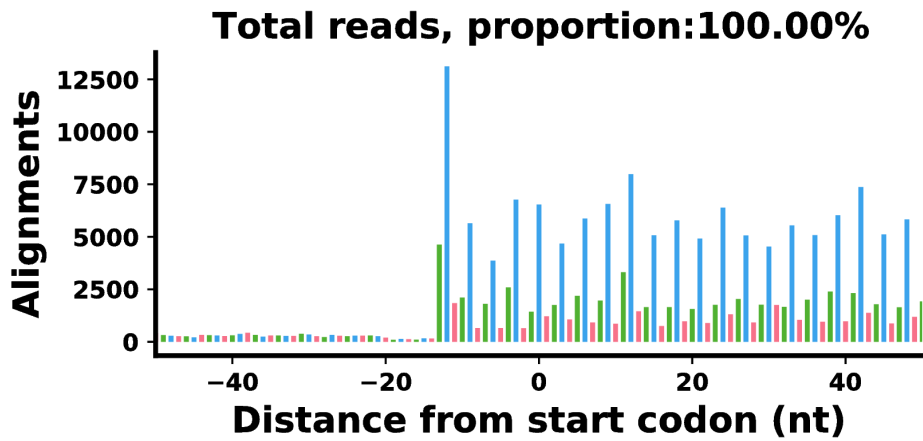
sample	Lengths	proportion	f0_sum	f1_sum	f2_sum	f0_percent
h3_1	29	15.87%	18762	5961	489	74.42%
h3_1	28	45.46%	67470	1815	1959	94.70%
h3_1	21	3.51%	14217	1025	2346	80.83%
h3_1	20	3.61%	13808	2585	1129	78.80%
h3_1	25	2.87%	3252	576	680	72.14%
h3_1	30	1.37%	1582	266	100	81.21%
h3_1	34	0.03%	21	3	7	67.74%
h3_2	21	3.72%	11021	838	2078	79.08%
h3_2	28	44.48%	50554	1370	1539	94.56%
h3_2	29	16.36%	14536	4775	347	73.94%
h3_2	20	3.75%	10512	2167	989	76.91%
h3_2	25	2.88%	2444	453	589	70.11%
h3_2	30	1.51%	1402	220	93	81.75%
wt_1	28	44.64%	61660	1620	2848	93.24%
wt_1	20	3.36%	9432	1728	1510	74.44%
wt_1	29	16.26%	17217	5658	511	73.62%
wt_1	21	3.14%	8595	675	2643	72.15%
wt_1	25	2.94%	3115	537	728	71.12%
wt_1	30	1.43%	1463	253	100	80.56%
wt_1	32	0.09%	74	12	22	68.52%
wt_1	37	0.03%	40	4	5	81.63%
wt_1	39	0.03%	27	6	3	75.00%
wt_1	40	0.03%	36	3	4	83.72%
wt_2	29	19.91%	18532	6229	485	73.41%
wt_2	28	43.91%	53499	1573	2706	92.59%
wt_2	20	3.05%	8302	1593	1079	75.65%
wt_2	21	3.29%	9146	614	2733	73.21%
wt_2	25	2.74%	2673	465	635	70.85%
wt_2	30	2.12%	1873	275	104	83.17%
wt_2	37	0.03%	24	4	6	70.59%
wt_2	40	0.03%	21	8	1	70.00%

1. sample:样本名称
2. Length:RPF reads长度
3. proportion:该长度reads占比对上reads的百分比
4. f0\_sum:frame0的数量
5. f1\_sum:frame1的数量
6. f2\_sum:frame2的数量
7. f0\_percent: frame0所占百分比

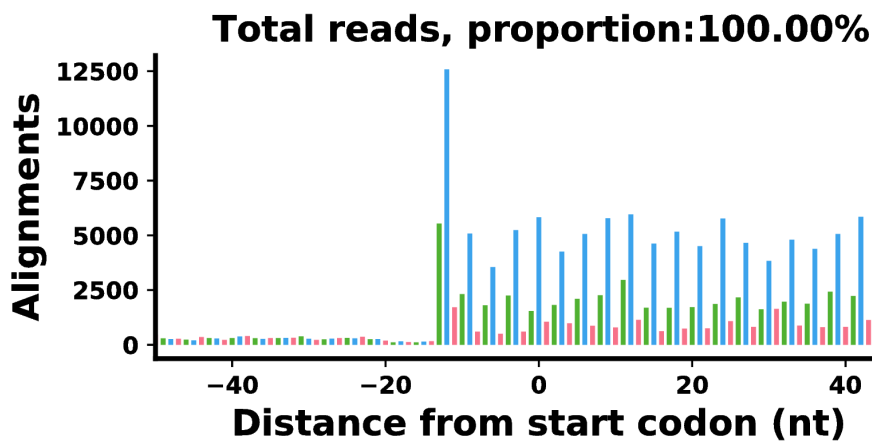
## □ 4.2 P-site分析

使用软件RiboCode (version:1.2.13) [9]中的 `metaplots` 命令, 选择最有可能来自翻译中核糖体的 RPF reads 的长度范围, 并通过已有注释中基因上比对到的 RPF reads 进行 meta-gene analysis 识别不同长度的 RPF 的 P-site。具体来说, 对于具有特定长度的每组 RPF 读段, 计算从它们的 5' 端到注释的起始和终止密码子的距离并总结为直方图。由于 Ribo-seq 亚密码子分辨率, 核糖体分析揭示了 RPF reads 中 80S 核糖体的肽基位点 (P-site) 的精确位置。因此, 由翻译核糖体产生的 RPFs 应沿 ORF 表现出 3-nt 周期性, 且大多数 reads 应该在 frame0 处富集。详细结果见 [./sup/03.rpf/psite](#)。

下图横坐标为以预测的 P-site 为中心左右 50 bp RPF reads 的分布, 纵坐标为该位置比对到的 reads 数量。蓝色代表属于 frame0 的 reads, 红色代表 frame1, 绿色代表 frame2。



wt\_1\_start\_periodicity-16



wt\_2\_start\_periodicity-16

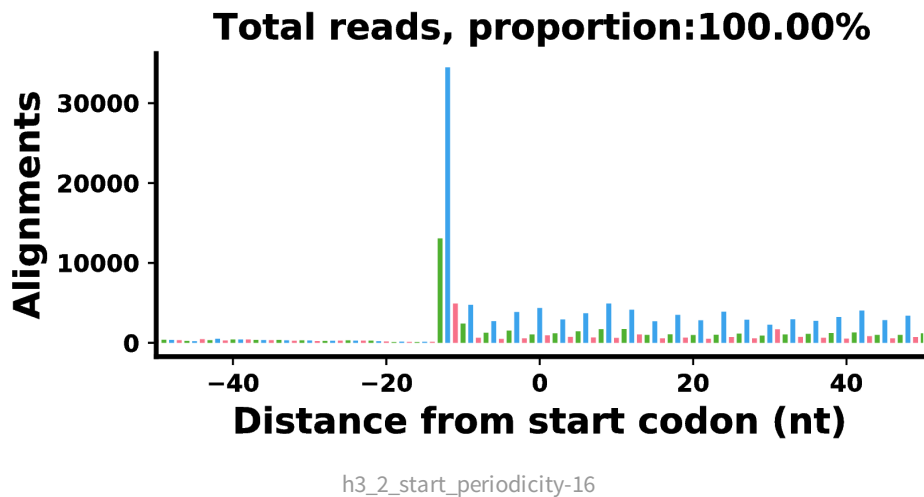
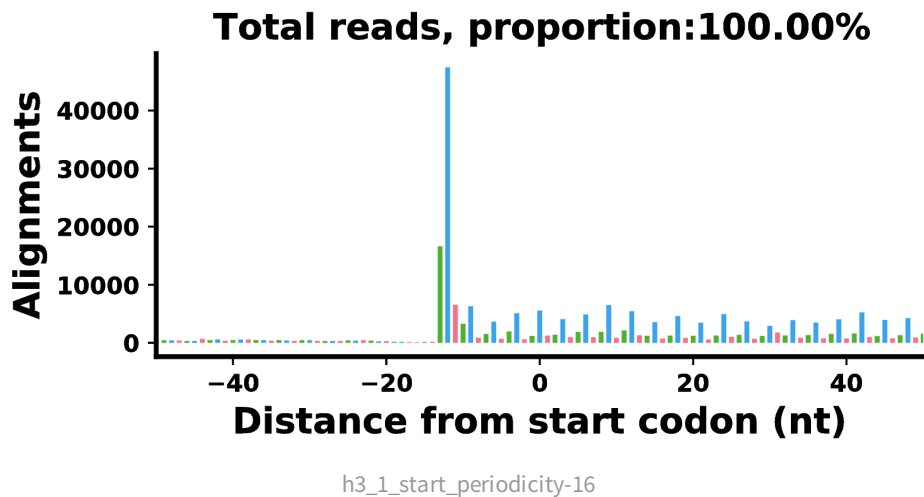


图4.2 P-site区域RPFs数量

Q: P/E-site是什么意思?

A: 核糖体在 mRNA 上的特定位置会被分为三个主要的位点，分别是 A 位点（Acceptance site，又称接受位点）、P 位点（Peptidyl site，又称肽基位点）、和 E 位点（Exit site，又称退出位点）。这些位点对应于核糖体在 mRNA 上合成蛋白质时的不同步骤和状态。A 位点（Acceptance site）：A 位点位于核糖体的 tRNA 结合部位，用于接受下一个氨基酸 tRNA 分子。当核糖体处于翻译的状态时，tRNA 的氨基酸会与 mRNA 上的密码子配对，形成氢键，然后 tRNA 将氨基酸运送至 A 位点，待下一个氨基酸的 tRNA 到来。P 位点（Peptidyl site）：P 位点是核糖体上的另一个重要位置，用于将已经合成的肽链连接到新的氨基酸。在翻译过程中，tRNA 上的氨基酸与 mRNA 上的密码子配对后，核糖体会将 tRNA 上的氨基酸与前一个 tRNA 上的氨基酸形成肽键，然后 tRNA 从 A 位点移动到 P 位点。E 位点（Exit site）：E 位点是核糖体上的另一个结合位置，用于 tRNA 在翻译过程结束后的释放。当 tRNA 上的氨基酸与 mRNA 上的密码子配对后，核糖体会将 tRNA 上的氨基酸与前一个 tRNA 上的氨基酸形成肽键，然后 tRNA 从 P 位点移动到 E 位点，待 tRNA 释放后，核糖体移动到 mRNA 上的下一个密码子进行继续翻译。

## 5 翻译停滞分析

### □5.1 元基因分析

我们使用软件RiboMiner (version:0.2) [10]对各样本进行元基因分析 (Metagene analysis)，可以检查核糖体占有率是否发生变化。如果编码区出现RPFs异常富集，说明此处翻译复合体没有顺利延

伸，出现了翻译停滞。

各样本转录本元基因分析，横坐标为标准化后转录本位置，纵坐标为该位置reads密度。

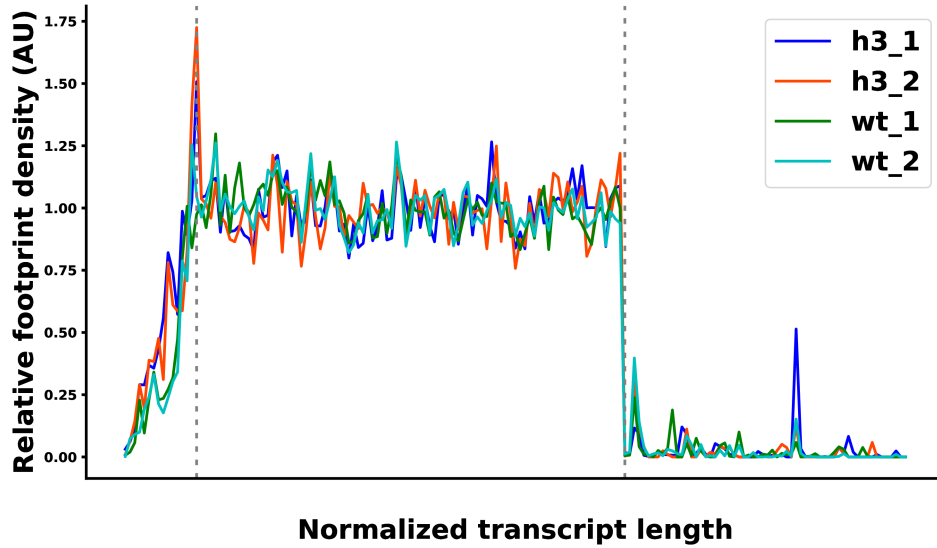


图5.1 转录本元基因分析

各样本极性分数分布图，横坐标为核糖体极性分数，纵坐标为相关基因的数量。如果峰值不在0处，出现在负值说明核糖体向5`端移动，出现在正值说明核糖体向3`端移动。

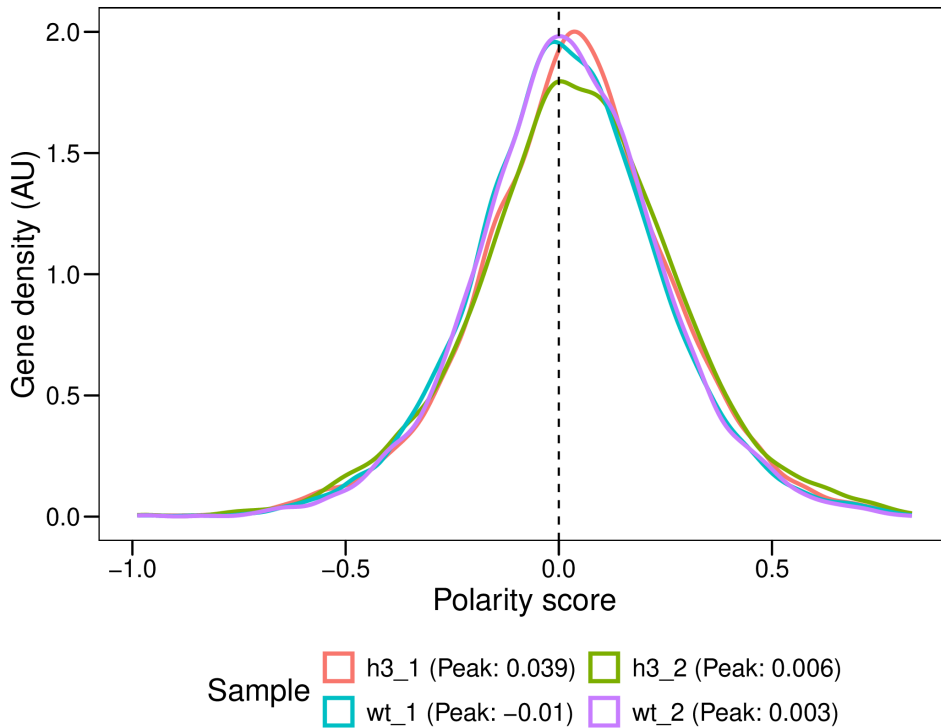


图5.2 极性分数分布图

为了准确体现各样本在起始/终止密码子处核糖体密度分布，我们绘制了这两个区域核糖体足迹分布图，0坐标代表起始/终止密码子，纵坐标为该位置reads密度。

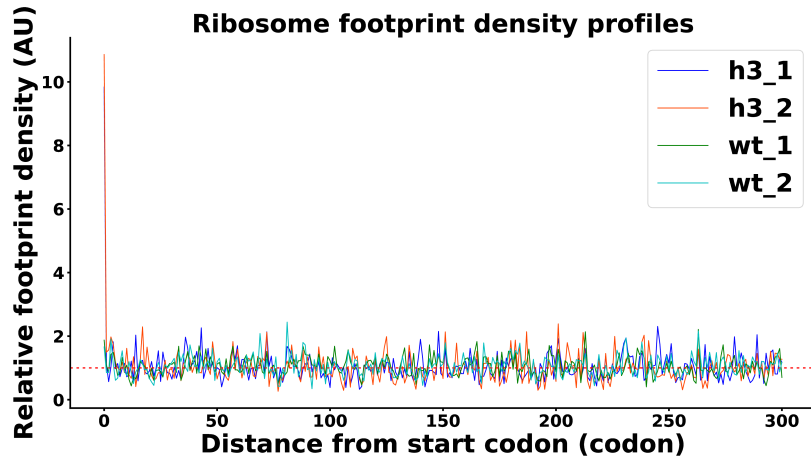


图5.3 起始密码子核糖体足迹分布

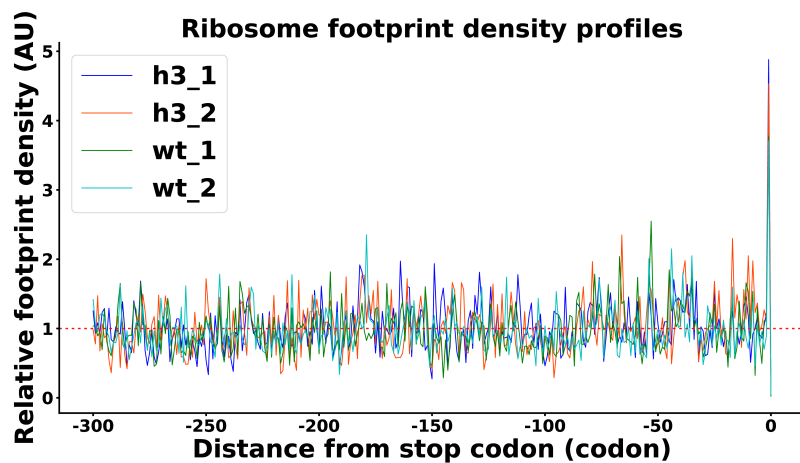


图5.4 终止密码子核糖体足迹分布

## □ 5.2 翻译停滞基因

上一步的分析显示了出现翻译停滞的区域，我们使用软件RiboMiner比较各个样本每个转录本翻译停滞区域密码子上的reads密度，选择reads比率大于1.5作为翻译停滞基因，详细结果见./sup/04.meta/stallgene。分析结果如下：

h3\_wt

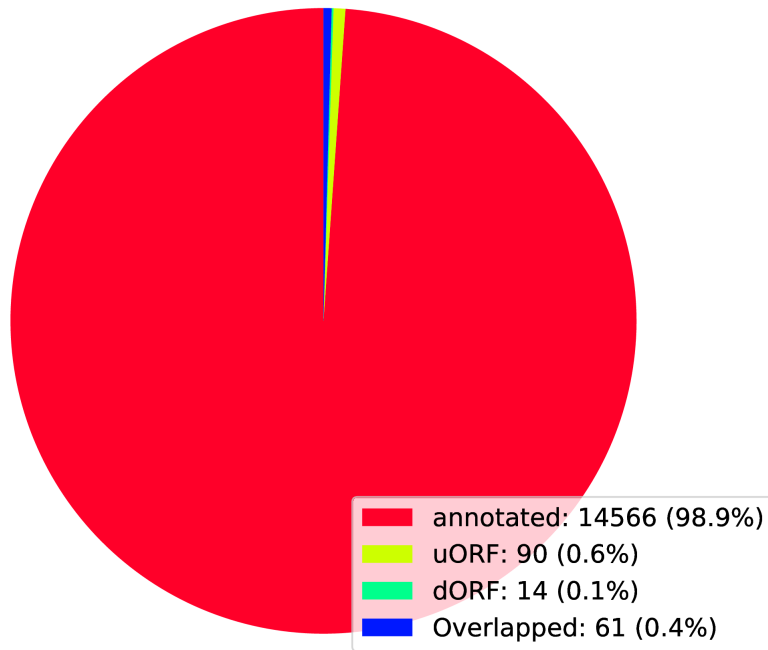
### h3\_wt\_stall\_gene\_list

gene	trans_id	codon	log2FoldChange	pvalue	h3_1	h3_2
AT4G22970	ma-NM_001203875.2	70	-9.645199100192523	0.04385686766079078	0	0
AT1G05380	ma-NM_001331546.1	1	8.112036990677465	0.04694559493088919	85.05	30.64864
AT1G05570	ma-NM_001331556.1	1	9.107876942944895	0.025320129031905884	52.594595	194.6
AT1G05570	ma-NM_001331556.1	18	-8.822763641952754	0.030276879120537673	0	0
AT2G18760	ma-NM_001335606.1	1	22.25619480813978	3.303357639186922e-6	394.333333	0
AT3G48190	ma-NM_001339355.1	73	22.87567475647021	1.7483324197864117e-6	0	531
AT5G08110	ma-NM_001342981.1	1	22.42116399823187	2.794392270118046e-6	334.571429	0
AT5G41790	ma-NM_001344421.1	100	-8.192005493355731	0.046665269750094246	0	0
AT5G42400	ma-NM_001344445.1	67	-8.923597747602827	0.02793460429531282	0	0
AT3G01460	ma-NM_111012.4	1	9.13351202842542	0.024286209469923697	91.774648	155.1428
AT3G13380	ma-NM_112183.3	18	10.045686662178843	0.013156764923197352	158.181818	309.3333
AT4G13750	ma-NM_117449.4	1	8.97712197580054	0.027952937511878568	38.380282	190.1162

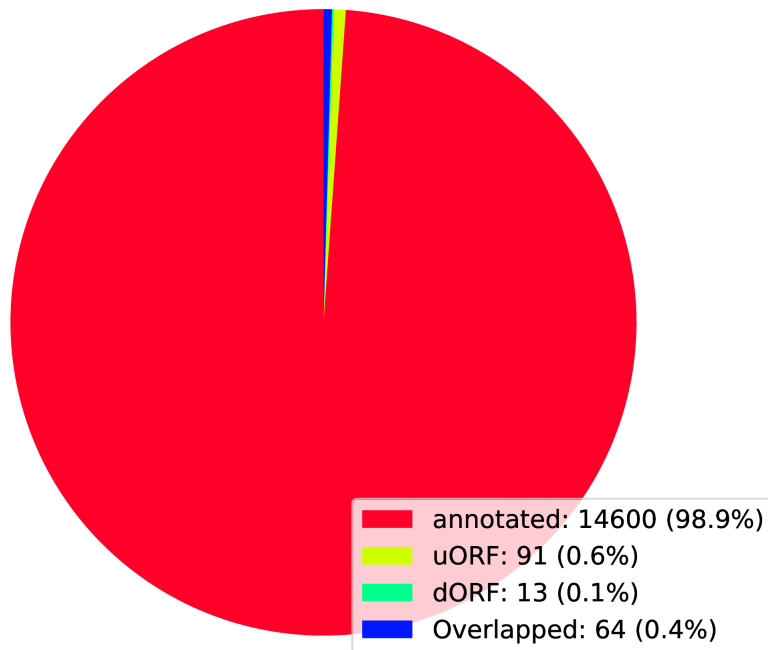
1. gene:基因ID
2. trans\_id:转录本ID
3. codon: 该转录本第几个密码子出现了差异富集
4. log2FoldChange: 实验组与对照组之间log2化的差异倍数
5. pvalue: 差异的显著性
6. status: 实验组相较于对照组该位置上下调情况
7. 其他: DESeq2结果

## 6 ORF分析

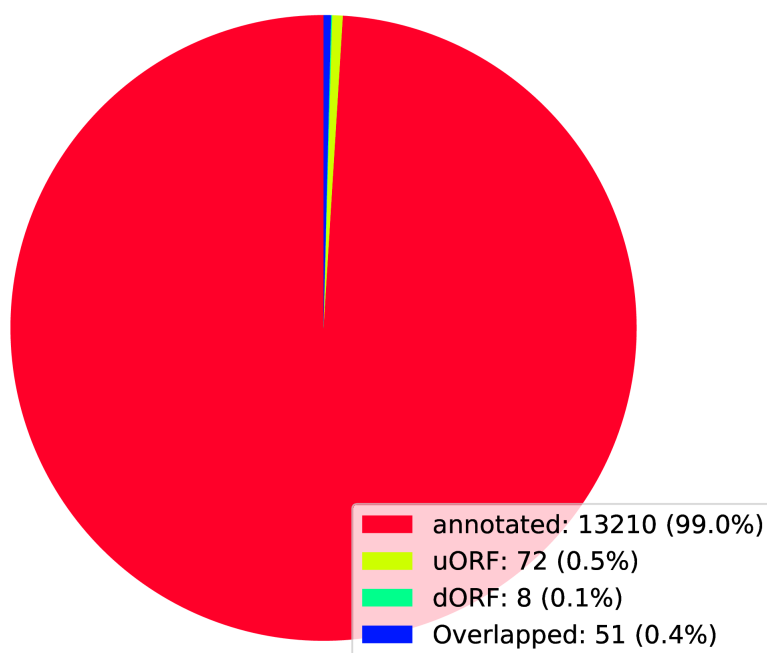
我们使用软件RiboCode根据PRF鉴定ORF，ORF的结果可以分成7类：1.annotated，与已经注释到的ORF重合，且起始密码子相同；2.uORF，在已注释CDS上游且不与CDS重合；3.dORF，在已注释CDS下游且不与CDS重合；4.Overlap\_uORF，在已注释CDS上游且与CDS重合；5.Overlap\_dORF，在已注释CDS下游且与CDS重合；6.Internal，在已注释CDS的内部ORF，但在不同的ORF中；7.novel，来自非编码基因或编码基因的非编码转录物。详细结果请见./sup/05.orf/。sup/05.orf/gtf/内gtf文件为鉴定的ORF结果。



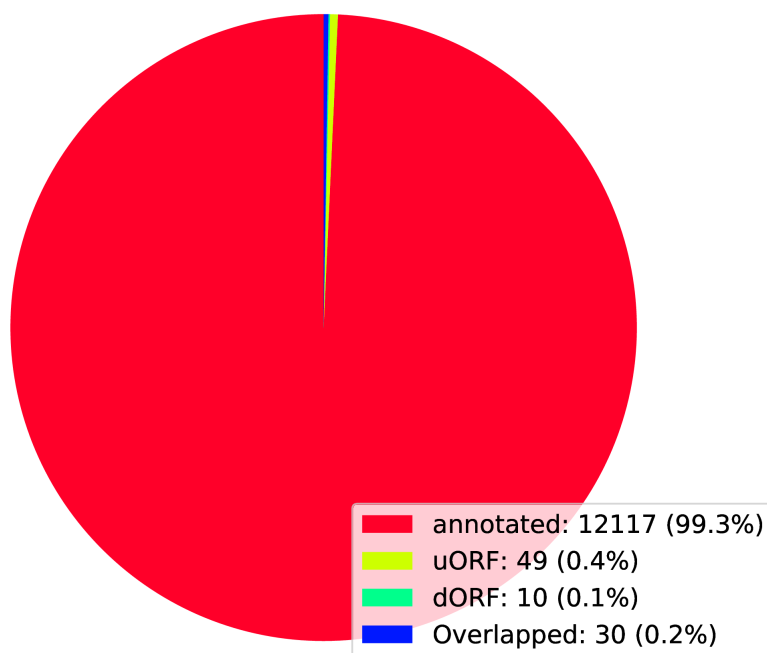
wt\_1\_ORFs



wt\_2\_ORFs



h3\_1\_ORFs



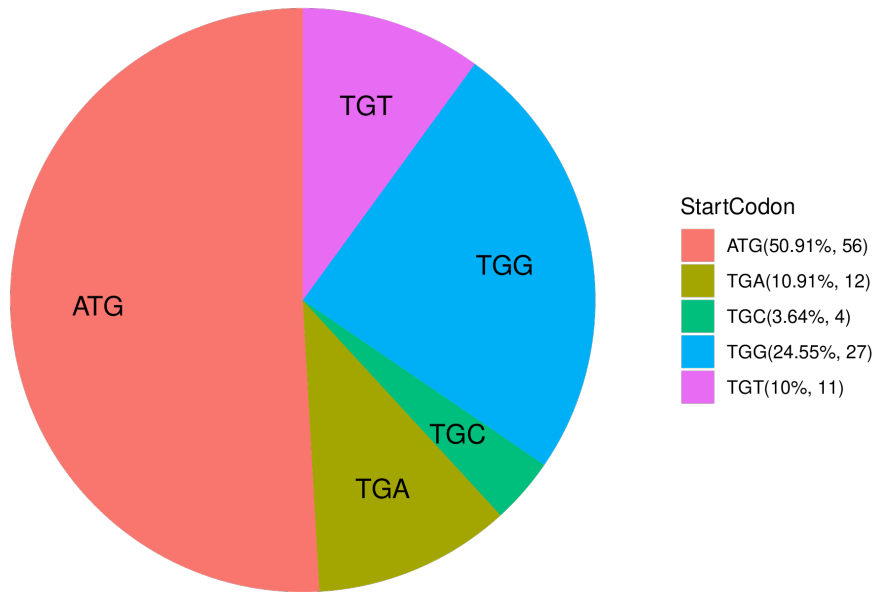
h3\_2\_ORFs

图6.1 ORF鉴定结果统计

## □ 6.1 uORF分析

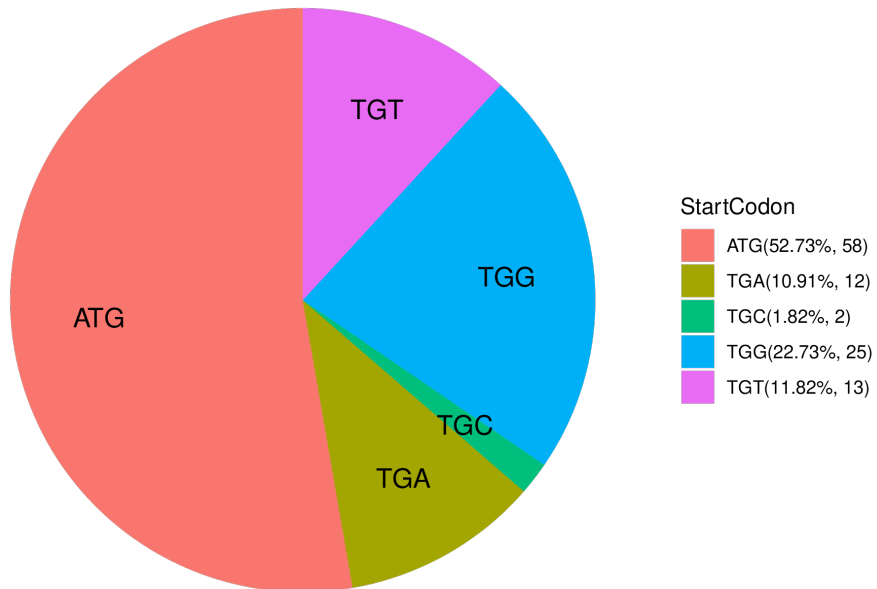
在真核生物体内，uORF可能起着相似的翻译调节机制，它与下游的mORF竞争翻译起始复合物或影响mORF的复苏。而在uORF中可能存在并未以传统ATG为起始密码子而以其他碱基序列为密码子的情況，导致了翻译提前发生。

wt\_1\_uORF Start Codon Frequency



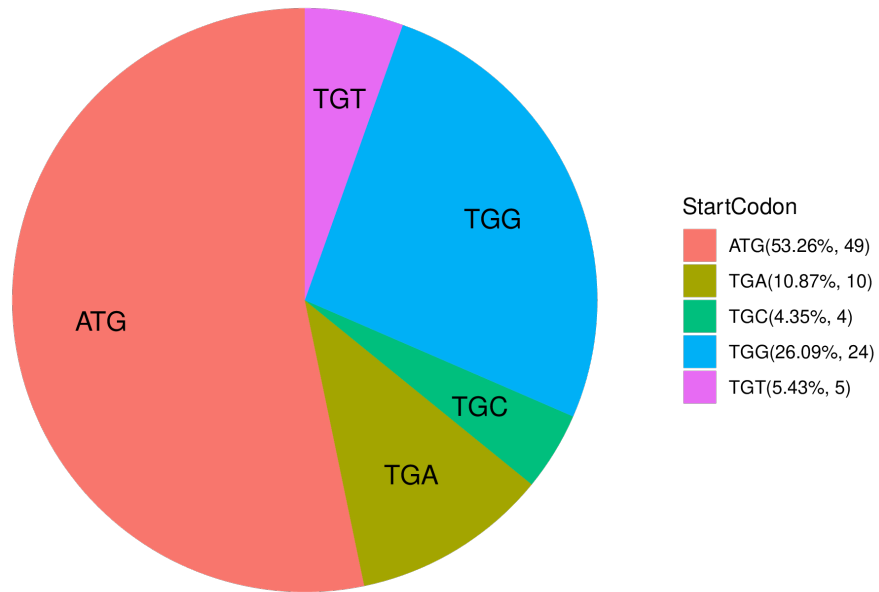
wt\_1\_uORF

wt\_2\_uORF Start Codon Frequency



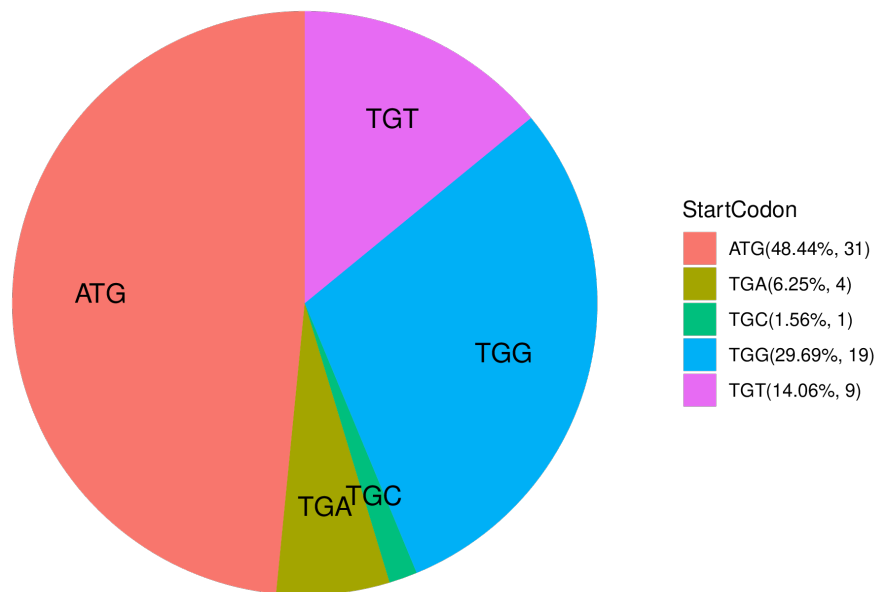
wt\_2\_uORF

h3\_1\_uORF Start Codon Frequency



h3\_1\_uORF

h3\_2\_uORF Start Codon Frequency



h3\_2\_uORF

图6.2 uORF起始密码子统计

每个uORF起始密码子情况请见下表:

wt\_1 wt\_2 h3\_1 h3\_2

### wt\_1\_uORF\_start\_codon\_list

ORF_ID	StartCodon	ORF_type	transcript_id	gene_id	chrom	ORF_
AT1G09340_3015482_3015691_69	ATG	uORF	ma-NM_001331830.1	AT1G09340	Chr1	30154
AT1G10200_3347733_3347623_36	TGG	uORF	ma-NM_001331891.1	AT1G10200	Chr1	33477
AT1G02205_419397_419552_51	ATG	uORF	ma-NM_001331328.1	AT1G02205	Chr1	41939
AT1G02205_419397_419552_51	ATG	uORF	ma-NM_180601.2	AT1G02205	Chr1	41939
AT1G02205_419665_419898_77	ATG	uORF	ma-NM_001331328.1	AT1G02205	Chr1	41966
AT1G02205_419397_419552_51	ATG	uORF	ma-NM_001331328.1	AT1G02205	Chr1	41939
AT1G02205_419397_419552_51	ATG	uORF	ma-NM_180601.2	AT1G02205	Chr1	41939
AT1G14290_4881781_4881506_91	TGA	uORF	ma-NM_001332110.1	AT1G14290	Chr1	48817
AT1G17380_5957070_5956837_77	TGT	uORF	ma-NM_001332300.1	AT1G17380	Chr1	59570
AT1G19180_6622312_6622575_87	ATG	uORF	ma-NM_202133.1	AT1G19180	Chr1	66223
AT1G20020_6942851_6942985_44	ATG	uORF	ma-NM_001332422.1	AT1G20020	Chr1	69428
AT1G21780_7652193_7652345_50	ATG	uORF	ma-NM_001084105.1	AT1G21780	Chr1	76521

1. ORF\_ID:本次分析时ORF编号
2. StartCodon: 起始密码子
3. ORF\_type: ORF类型
4. transcript\_id: 转录本ID
5. gene\_id: 基因ID
6. chrom: 染色体
7. ORF\_gstart: ORF在基因组起始坐标
8. ORF\_gstop: ORF在基因组终止坐标

## 7 表达量分析

### □7.1 样本表达量分析

根据比对的reads，保留uniq比对（在基因组上仅比对到一个位置），去掉重复的reads(reads片段的序列相同)，然后使用featureCounts(v2.0.1)[11]对样本进行表达量分析。详细结果见 ./sup/06.exp。

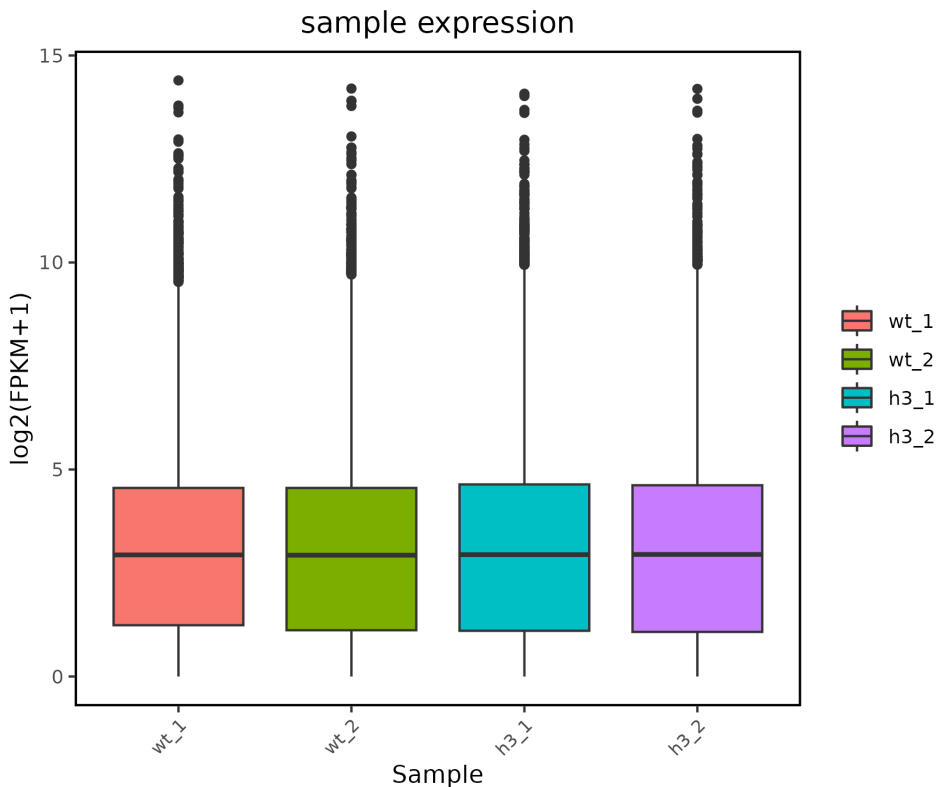


图7.1 样本表达量分布

## 7.2 PCA分析

根据基因在各个样本中的RPKM (Reads Per Kilobase per Million mapped reads,每千个碱基的转录每百万映射的reads) 的概念对每一个基因进行定量, 通过RPKM计算样本的PCA分析。详细结果见 [./sup/06.exp](#)。

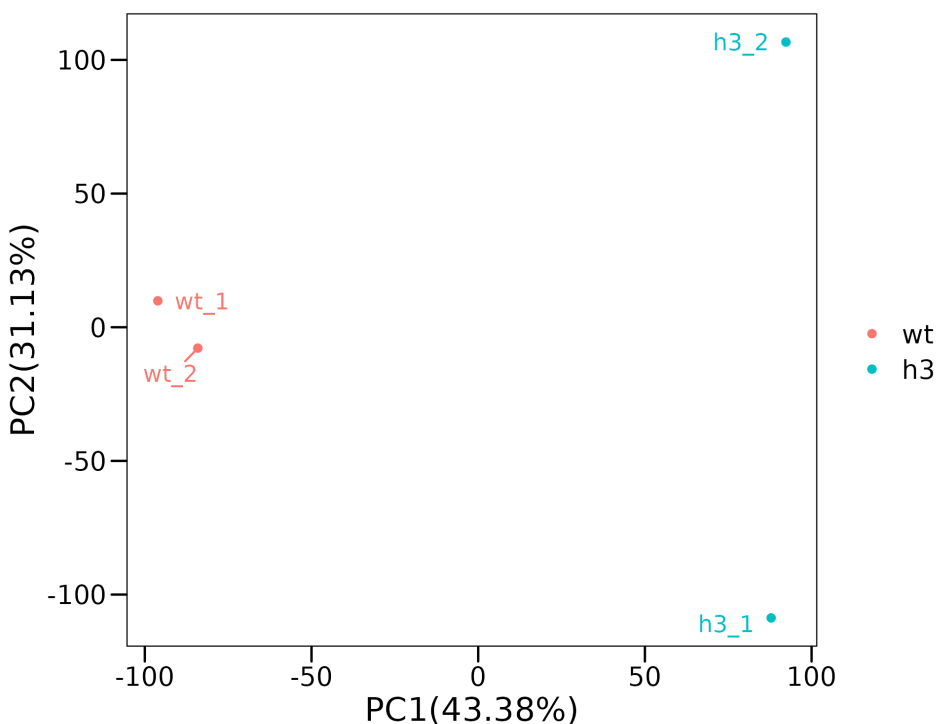


图7.2 样本PCA分析

## 7.3 样本相关性

根据基因在各个样本中的RPKM的概念对每一个基因进行定量，通过RPKM计算样本的相关性分析。详细结果见 ./sup/06.exp。

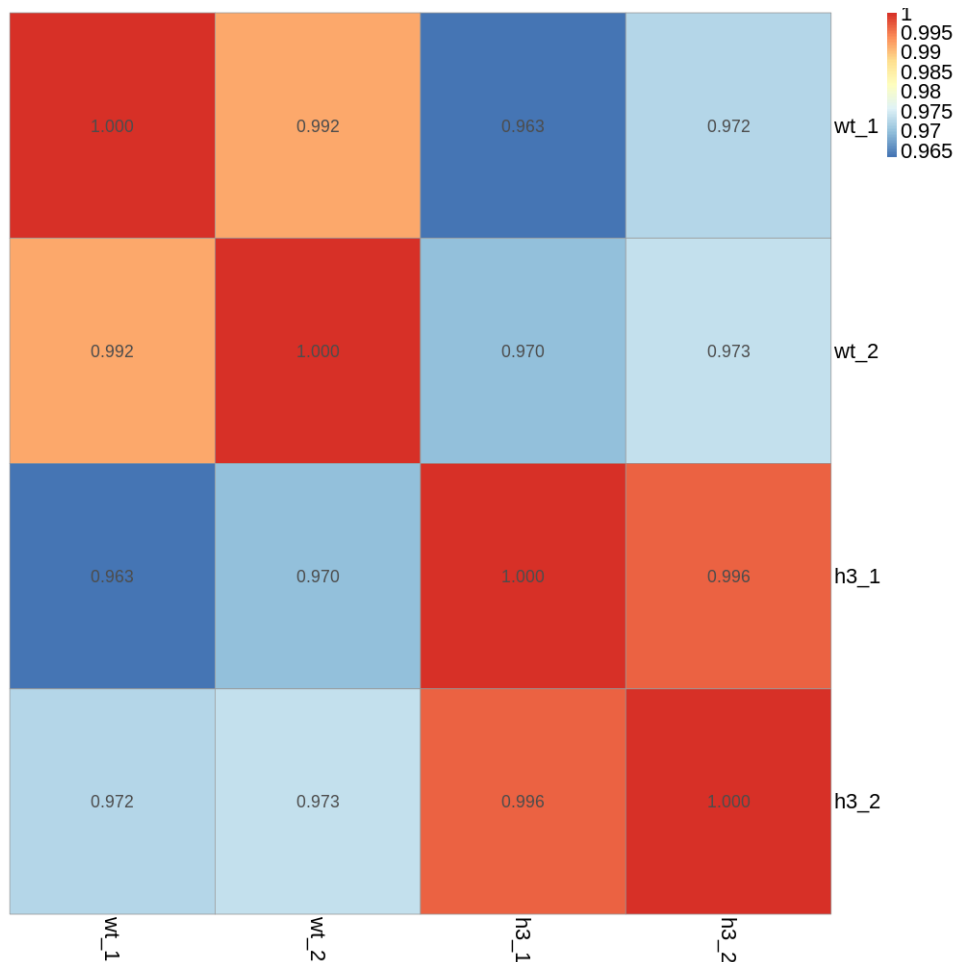


图7.3 样本相关性

## 7.4 表达量Q&A

Q: 基因表达水平如何计算?

A: Ribo-Seq分析时一般是使用RPKM (Reads Per Kilobase per Million mapped reads,每千个碱基的转录每百万映射的reads) 表征基因表达量。

Q: 样品间的相关性有何意义?

A: 样品间的相关性反应了样品间的相似程度，即不同处理或组织的样品在表达水平方面的相似度。使用Pearson相关性计算，相关系数越接近1，样品间的相似度越高，样品间的差异基因也越少。一般生物学重复间的样品的相关系数较高。

## 8 差异表达分析

统计各个基因的reads数，根据reads，使用DESeq2[12]方法，分析各组的差异表达基因。参数一般选择FC(fold change) = 2, pValue = 0.05。差异基因列表见sup/07.deg/\*\_vs\_\*文件夹下各组的\*\_ann\_deg.txt 文件。

## □ 8.1 火山图

各组样本差异表达分析的火山图。

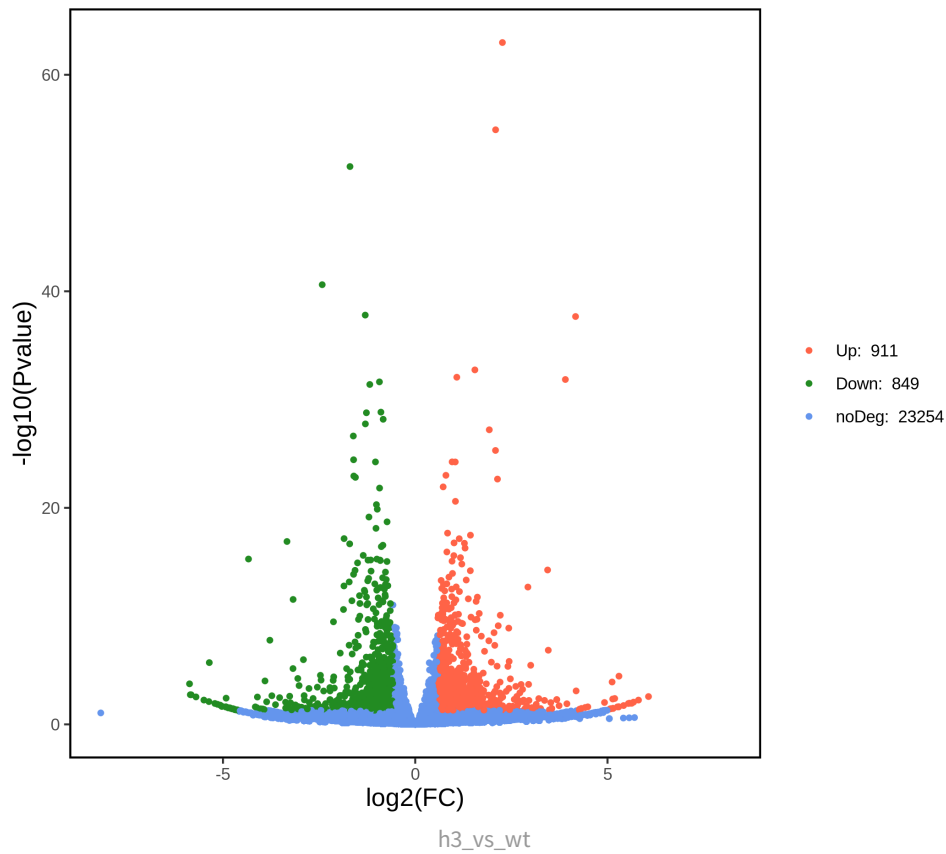


图8.1 火山图

## □ 8.2 差异基因统计

对各组差异基因进行汇总统计，详细结果见 ./sup/07.deg。

表8.1 差异基因统计表

group	Up	Down	Total
h3_vs_wt	911	849	1760
totalDeg	911	849	1760

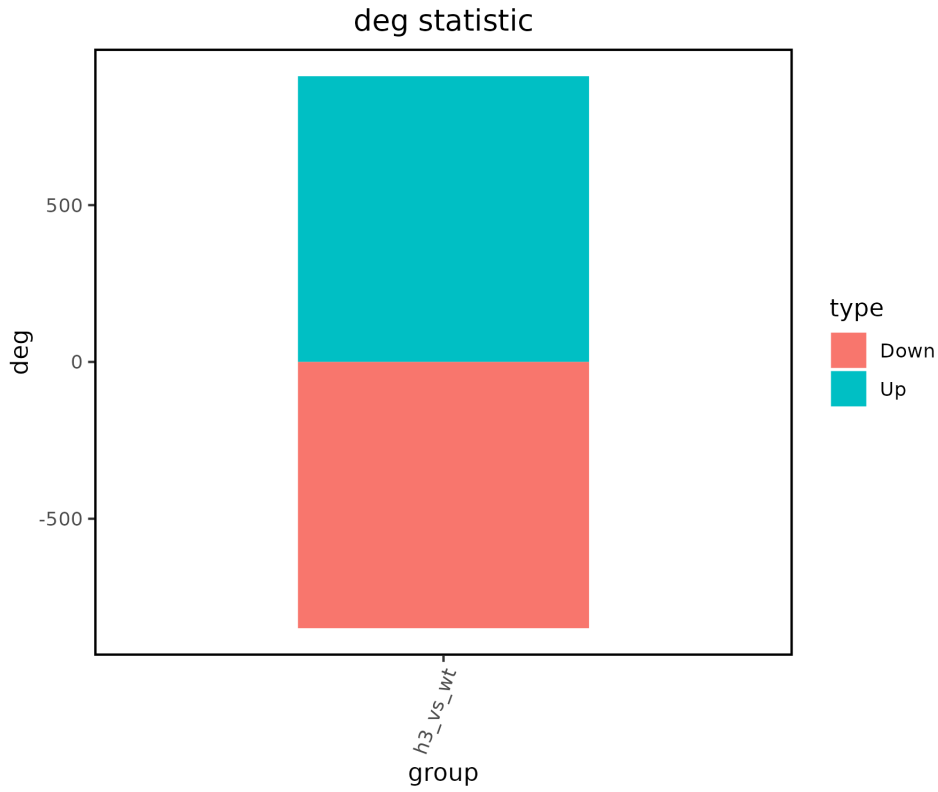


图8.2 差异基因统计

### □ 8.3 差异基因样本相关性分析

根据差异表达的基因在各个样本中的RPKM（Reads Per Kilobase per Million mapped reads）的概念对每一个差异表达基因进行定量，通过RPKM计算样本的相关性分析。详细结果见 [./sup/07.deg](#)。

•

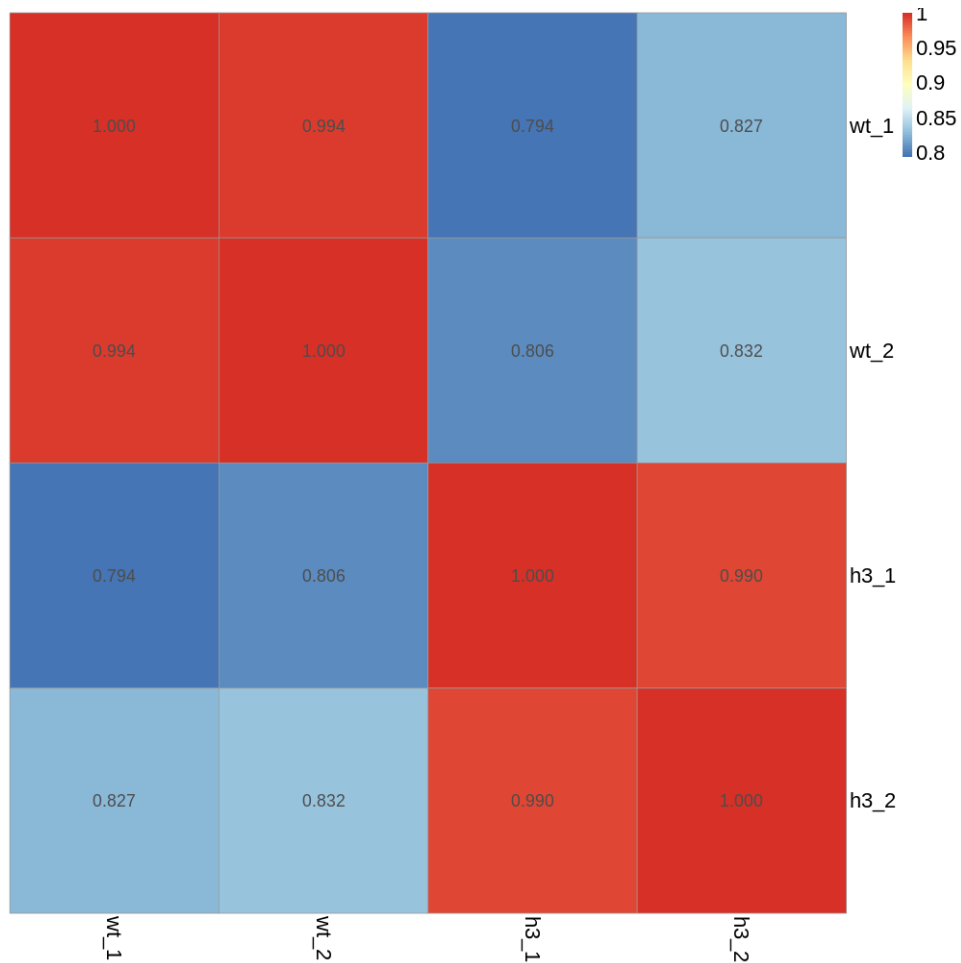


图8.3 差异基因样本相关性

## □ 8.4 差异基因样本PCA分析

根据差异表达的基因在各个样本中的RPKM (Reads Per Kilobase per Million mapped reads) 的概念对每一个差异表达基因进行定量，通过RPKM计算样本的PCA分析。详细结果见 [./sup/07.deg](#)。

•

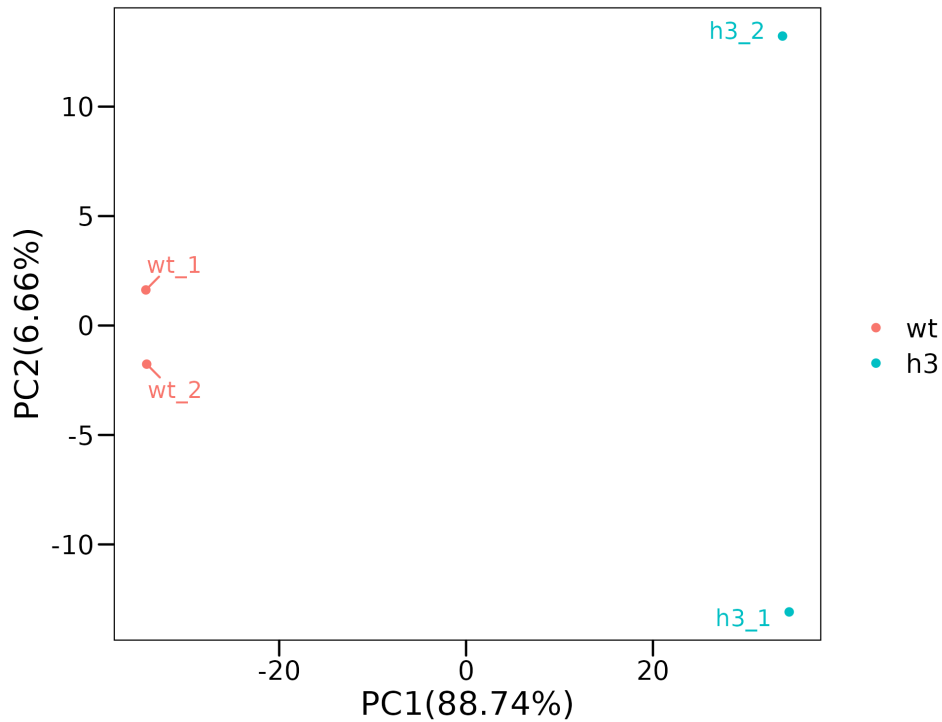


图8.4 差异基因样本PCA分析

## □ 8.5 差异基因样本heatMap分析

选取差异基因，对这些基因在各样本中的表达情况进行heatMap分析，详细结果见 [./sup/07.deg](#)。

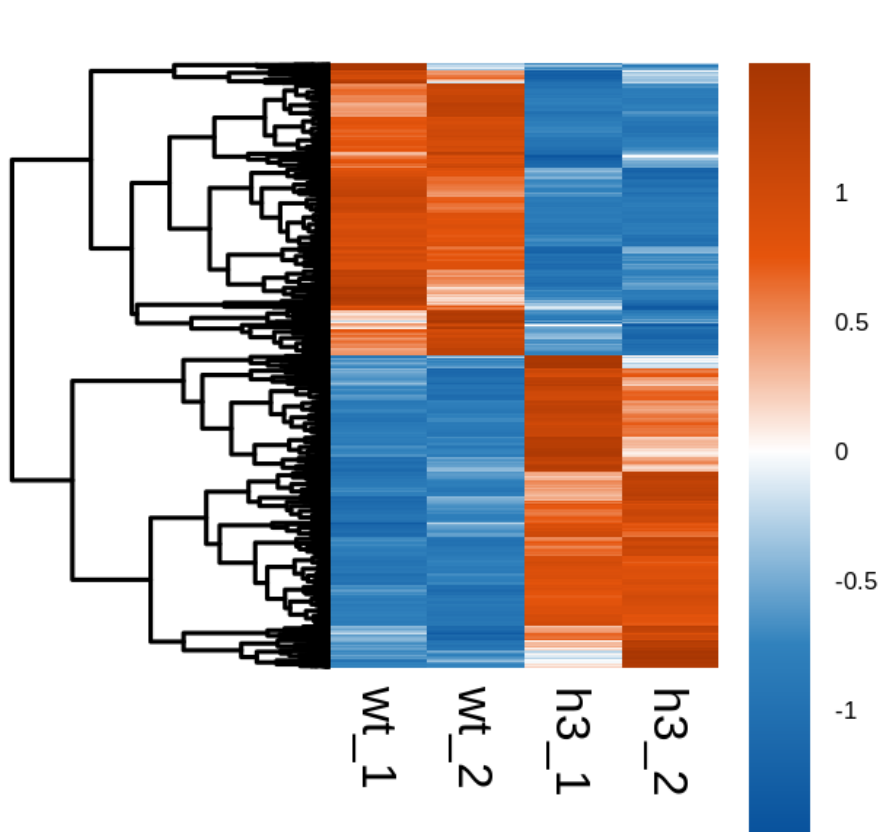


图8.5 差异基因样本heatMap分析

## □ 8.6 差异表达分析Q&A

Q: 用什么软件分析差异表达基因?

A: 对于有重复的样本, 使用DESeq2方法分析, 无重复的样本, 使用edgeR方法进行差异表达分析。

Q: 如何判断一个基因是否是差异表达基因?

A: 差异基因标准一般是 $\log_2FC \geq 2$ , 则认为该差异基因是上调, 反之, 若 $\log_2FC \leq -2$ , 认为该差异基因是下调, 同时需要满足  $pvalue \leq 0.05$ 。

Q: 用什么数据进行基因差异表达分析?

A: 在做差异分析时, 一般使用reads count数据, 通过DESeq或者TMM标准化后, 进行差异分析。在进行差异分析时, DESeq和TMM的标准化效果最好, FPKM/RPKM的标准化效果较差, 所以, 不推荐使用FPKM/RPKM进行差异分析。

## 9 与转录组联合分析

### 9.1 翻译效率分析

我们比较蛋白组数据和转录组数据时, 往往会发现转录本丰度不等于蛋白质丰度, 翻译效率 (Translation Efficiency, TE) 的改变是基因翻译水平调控的主要形式, 直接影响蛋白质的产量, 其计算公式为:  $TE = (\text{RPKM in RiboProfiling}) / (\text{RPKM in RNAseq})$ 。完整结果请见 [./sup/08.te](#), 各比较组翻译效率分析结果为\*\_TE.ann.txt, 各比较组翻译效率差异基因如下表所示:

h3\_vs\_wt\_diff

**h3\_vs\_wt diff TE genes list**

ID	TEstatus	log2FC_TE	pvalue	p.adjust	Name
AT1G64790	down	-6.013126872578892	8.845240655164248e-186	2.2511137467393006e-182	AT1G64790
AT3G28900	down	-7.200276632159319	6.045667871075857e-116	7.69311236594403e-113	AT3G28900
AT1G74550	down	-3.170062872113195	2.0608076921863752e-32	1.748251858871442e-29	AT1G74550
AT4G23800	down	-3.074612387211685	7.459396622919407e-27	4.746041101332472e-24	AT4G23800
AT5G44180	down	-5.403629606061854	4.6804750134315336e-24	2.38236178183665e-21	AT5G44180
AT1G12840	down	-1.5666369116776793	1.319531121942205e-19	5.59701117557152e-17	AT1G12840
AT5G46700	down	-3.1039366902318877	2.3003274400748845e-17	8.36333333570083e-15	AT5G46700
AT2G17695	down	-6.435321374441748	5.981237852432756e-16	1.9027812918051705e-13	AT2G17695
AT1G21160	up	3.3256158158740106	4.214093103130005e-15	1.1916518830517625e-12	AT1G21160
AT3G10720	down	-2.572524854546513	6.282498432577117e-14	1.5988958510908762e-11	AT3G10720
AT1G03140	up	3.729770365680059	1.4206440862555306e-13	3.2868538177457494e-11	AT1G03140
AT1G14650	down	-3.436389480566657	7.724384320296391e-13	1.512196776550332e-10	AT1G14650

1. ID: 基因ID
2. TEstatus: 比较组中翻译效率上下调情况
3. log2FC\_TE: 翻译效率log2化倍数
4. pvalue: 差异显著性p值
5. p.adjust: BH校正后差异显著性p值
6. Name: 基因名称
7. GO: 基因对应GO条目
8. KEGG: 基因对应KEGG通路

## 9. Nr: NR数据库注释

各组样本差异翻译效率的火山图。

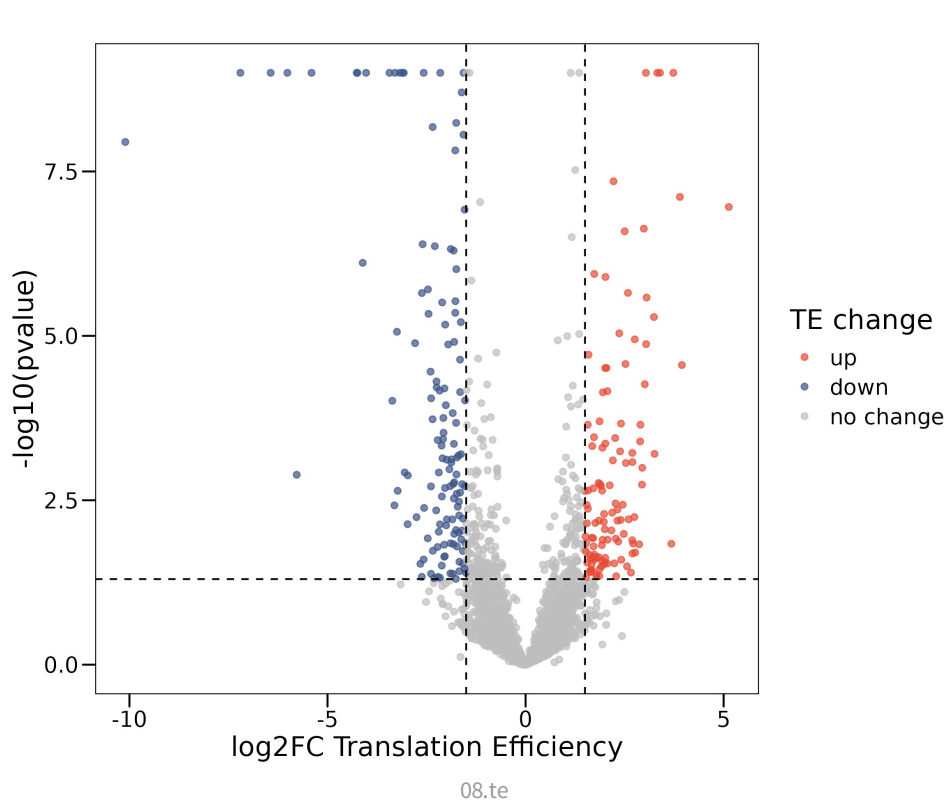


图9.1 TE差异基因火山图

## □9.2 转录与翻译差异基因交集

我们将mRNA表达差异基因和RPF表达差异基因取交集，绘制venn图。

为了进一步表现翻译组和转录组数据整合结果，我们根据mRNA和RPF差异倍数绘制转录翻译九象限图。下图横坐标为转录差异基因log<sub>2</sub>FC值，纵坐标为翻译差异基因log<sub>2</sub>FC值。从左往右第一条竖线左边为转录下调基因，第二条竖线右边为转录上调基因，两条竖线中间为转录没有差异的基因；从上往下第一条横线上面为翻译上调基因，第二条横线下为翻译下调基因，两条横线中间为翻译没有差异基因。基因可以分为5类：

1. 仅在mRNA水平上表达改变的基因（transcription only），基因仅在转录水平上存在差异，但它们的RPF水平没有显著差异；
2. 仅在TE水平上表达变化的基因（translation only），基因的两个等位基因仅在RPF水平上存在差异，但在转录水平上没有显著差异；
3. 反向变化基因（opposite change），转录和翻译水平上出现反向变化；
4. 同向变化基因（homodirectional），转录和翻译水平上出现同向变化；
5. 剩下的无差异基因。

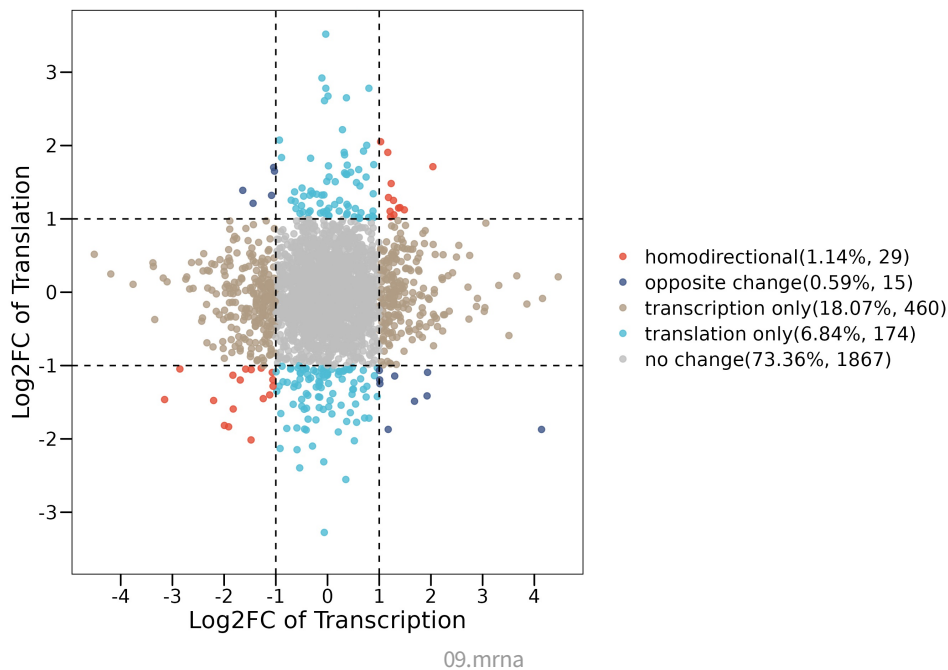


图9.2 转录翻译九象限图

### h3\_vs\_wt\_log2fc\_diff

h3_vs_wt diff RPF and mRNA list						
ID	status	mRNA_log2FC	RPF_log2FC	Name	TF	
AT1G07260	homodirectional	-1.8236245481263984	-1.5923560628481055	AT1G07260	--	GO:0004134,4-s
AT1G19780	homodirectional	1.179083180515525	1.2915980492979413	AT1G19780	--	GO:0000166,nuc
AT1G49490	homodirectional	-1.4757142192786172	-1.0555553567967992	AT1G49490	--	GO:0005199,str
AT1G53480	homodirectional	-1.4787174487018375	-2.014045053341304	AT1G53480	--	GO:0008150,bic
AT1G67330	homodirectional	1.2759202376500542	1.2518267844989206	AT1G67330	--	GO:0000139,Go
AT1G67870	homodirectional	-1.2431108702264588	-1.4508734249800177	AT1G67870	--	GO:0008150,bic
AT1G68940	homodirectional	-1.068680285455804	-1.0937670528266714	AT1G68940	--	GO:0004842,ubi
AT1G72280	homodirectional	-1.8301584713784889	-1.1319428270316263	AT1G72280	--	GO:0005515,pro
AT2G03630	homodirectional	1.486341978910953	1.123467185129733	AT2G03630	--	--
AT2G30830	homodirectional	-1.118203467840868	-1.4001746755600275	AT2G30830	--	GO:0008150,bic
AT2G35658	homodirectional	-1.9941616589089712	-1.816636710092048	AT2G35658	--	GO:0003674,mo
AT2G36090	homodirectional	1.284407545344614	1.0555263420563008	AT2G36090	--	GO:0005515,pro

1. ID: 基因ID
2. status: 该基因转录翻译变化情况
3. mRNA\_log2FC: mRNA中比较组log2差异倍数
4. RPF\_log2FC:RPF中比较组log2差异倍数
5. Name: 基因名称
6. GO: 基因对应GO条目
7. KEGG: 基因对应KEGG通路
8. Nr: NR数据库注释

Q: 用什么软件分析差异基因?

A: 使用核糖体差异分析软件Xtail (version:1.1.5) [13]进行翻译效率差异分析以及转录翻译变化分析。

Q: 筛选差异基因标准是什么?

A: 我们根据mRNA和RPF的log2FC变化情况对基因进行分类, 为了获得较多的基因可以放松筛选条件, 以 $abs(log2FC) > 1$ 来筛选; 如果需要更严格筛选, 可以调整阈值为1.5或2。

## 10 差异基因富集分析

功能聚类分析介绍:

GO 功能分析

基因本体论 (Gene Ontology) 简称GO, 是一个国际化的基因功能分类体系, 包含生物学领域知识体系本质的表示形式, 本体通常由一组类 (或术语或概念) 组成, 分为3大类: 分子功能 (Molecular Function)、细胞组分 (Cellular Component)、生物过程 (Biological Process)。对研究的物种使用相关GO注释数据库或利用blast, 得到每个基因对应的GO term, 根据基因注释信息, 统计每个基因所在的GO term, 根据每个GO term的基因数目, 以及背景中此GO term的基因数目, 用Fisher Exact Test分析每个GO term的显著性。

KEGG 功能分析

京都基因与基因组百科全书 (Kyoto encyclopedia of genes and genomes) 简称KEGG, 是系统分析基因功能、基因组信息的数据库, 它整合了基因组学、生物化学以及系统功能组学的信息, 有助于研究者把基因及表达信息作为一个整体网络进行研究。对差异基因进行KEGG功能富集分析, 研究的物种使用相关KEGG注释数据库或利用blast, 得到每个基因对应的KEGG pathway。根据基因注释信息, 统计每个基因所在的KEGG pathway, 根据每个KEGG pathway的基因数目, 以及背景中此KEGG pathway的基因数目, 用Fisher Exact Test分析每个KEGG pathway的显著性。

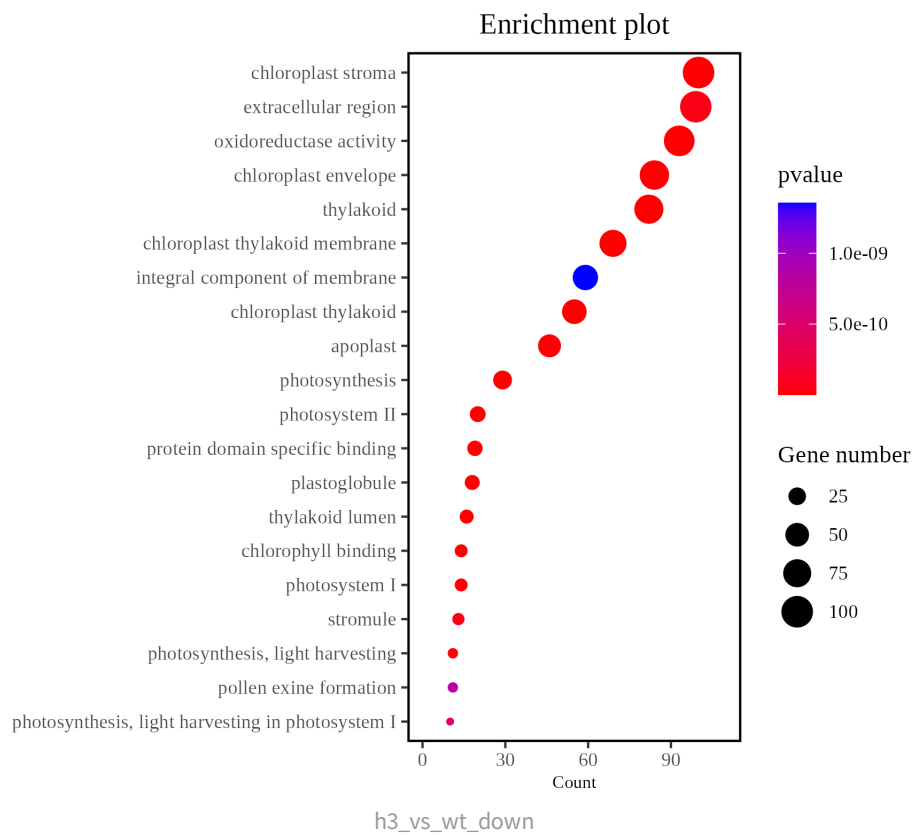
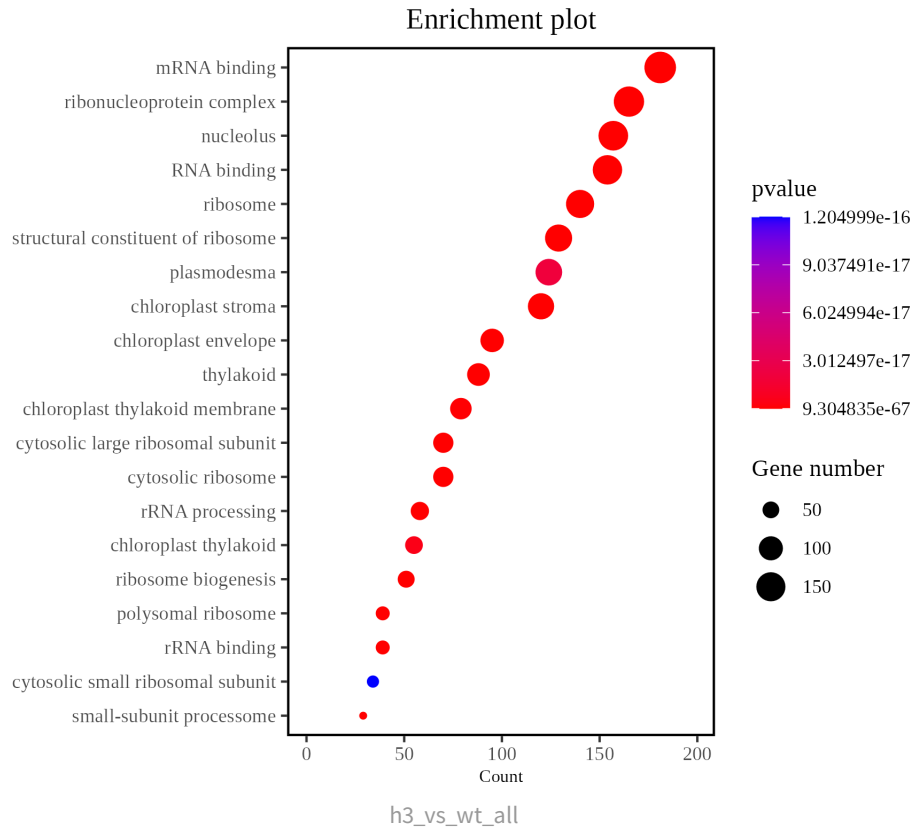
### □ 10.1 RPF表达差异基因

我们对RPF表达差异基因进行GO与KEGG富集分析。基因注释结果请见./sup/10.gokegg/rpf/。

#### ■ 10.1.1 GO 功能分析

GO 功能聚类气泡图展示。纵坐标是GO Term 名称, 横坐标是对应GO Term 中检出的基因占背景基因的个数, 颜色代表显著性pvalue。

•



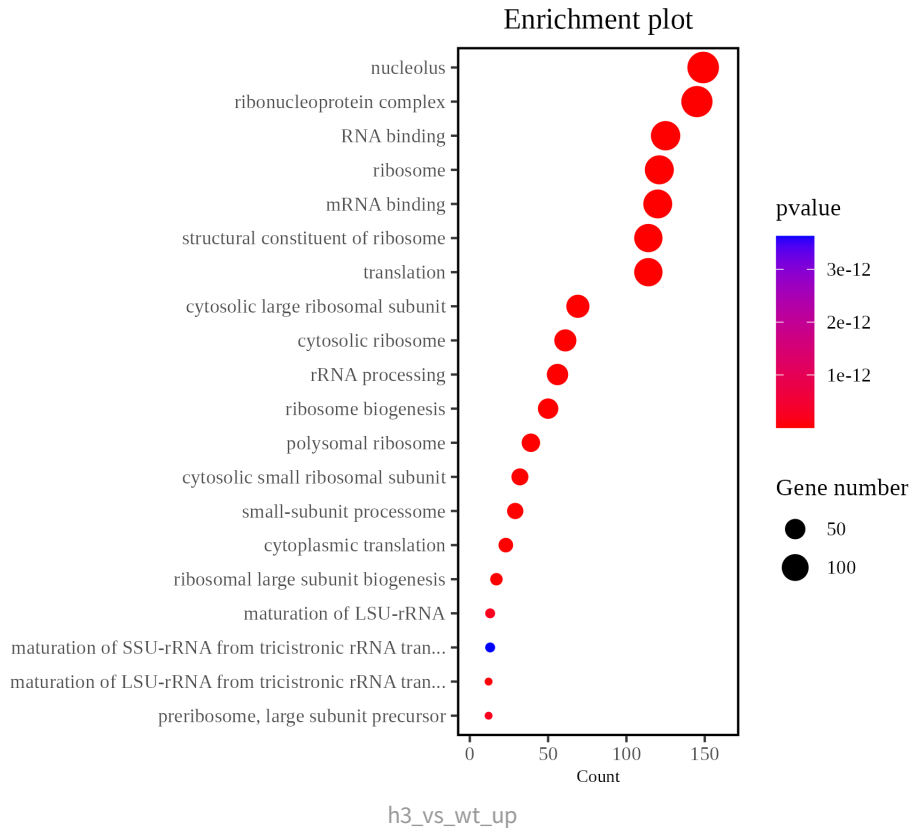
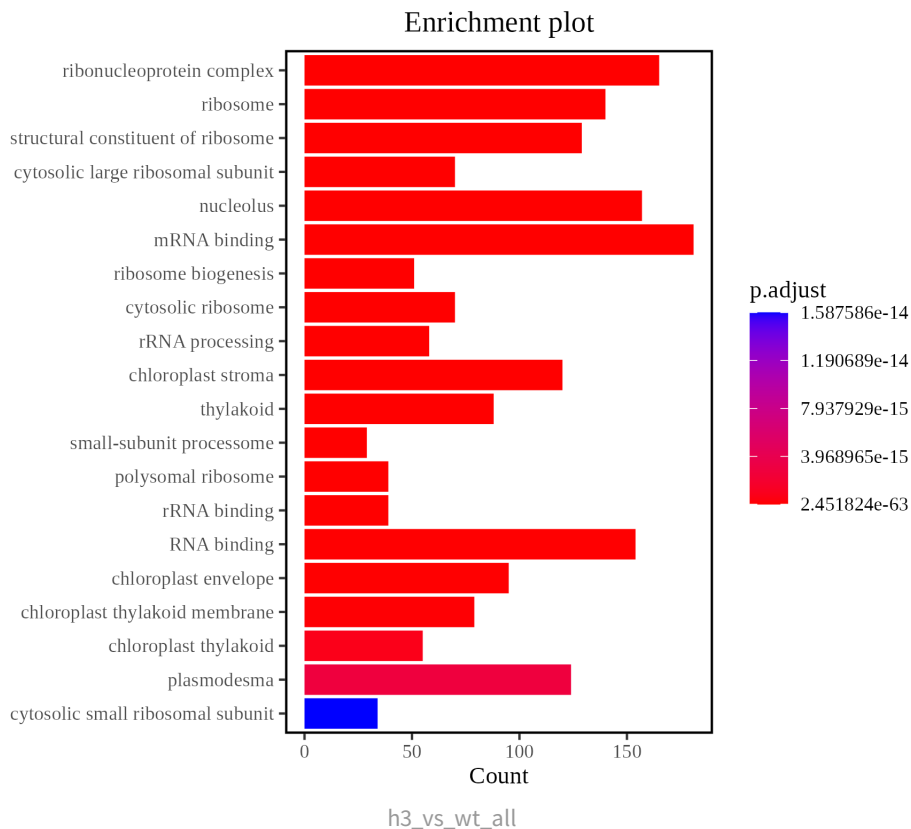


图10.1 GO功能聚类气泡图

GO 功能聚类bar图展示。纵坐标是GO Term 名称，横坐标是对应GO Term 中检出的基因个数，颜色代表显著性p.adjust。



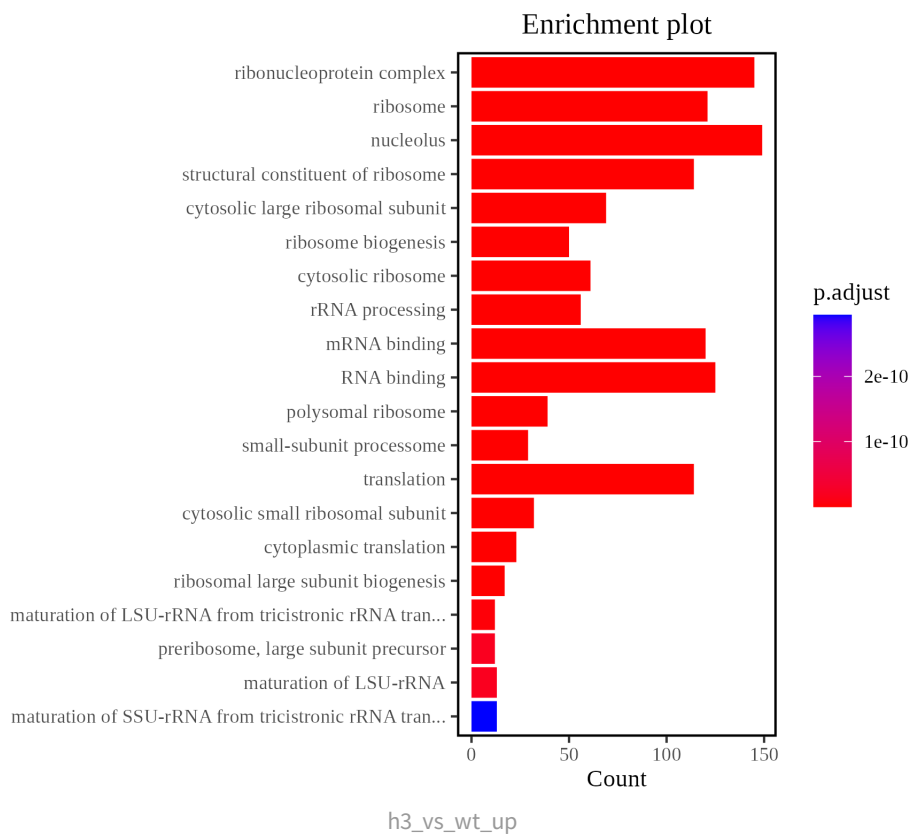
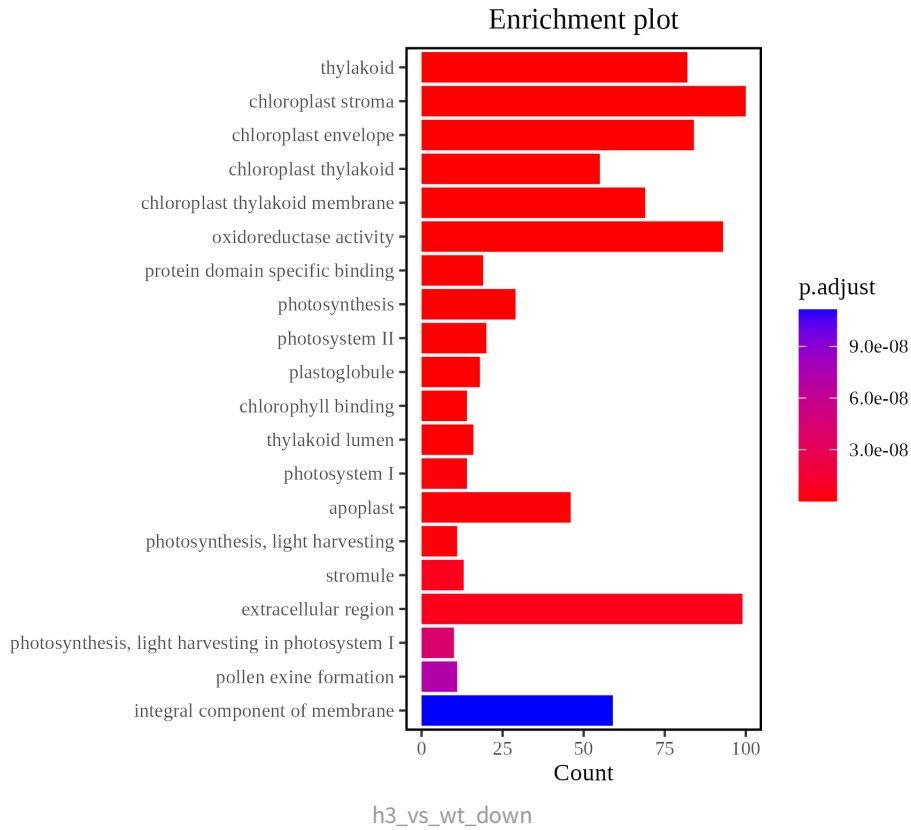
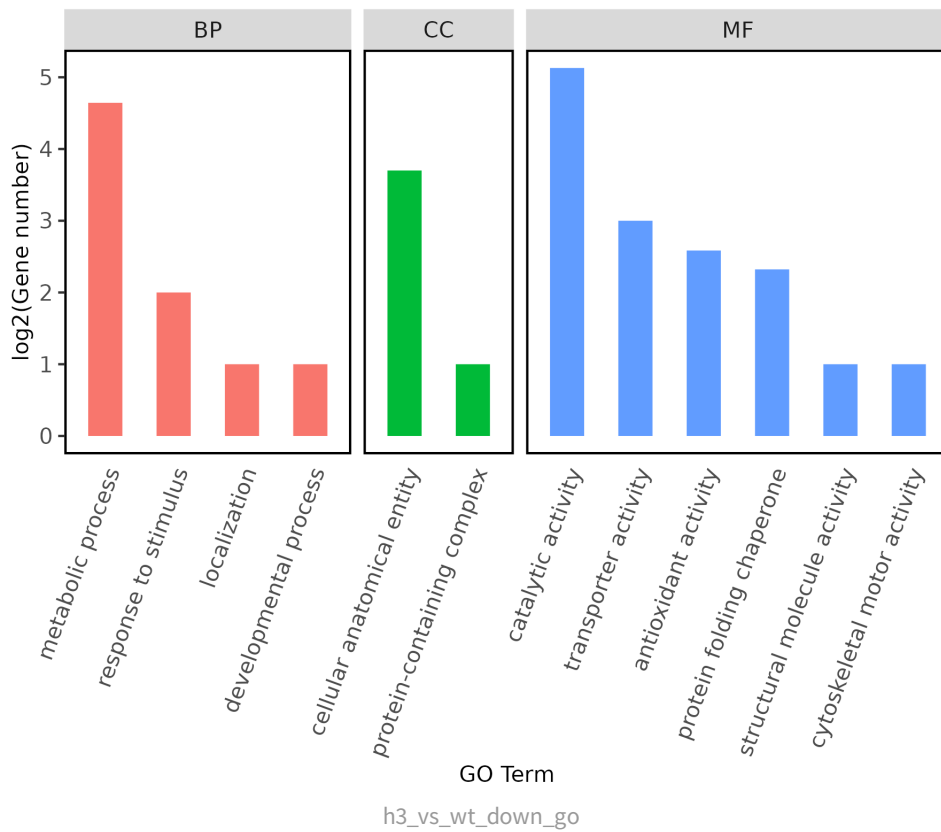
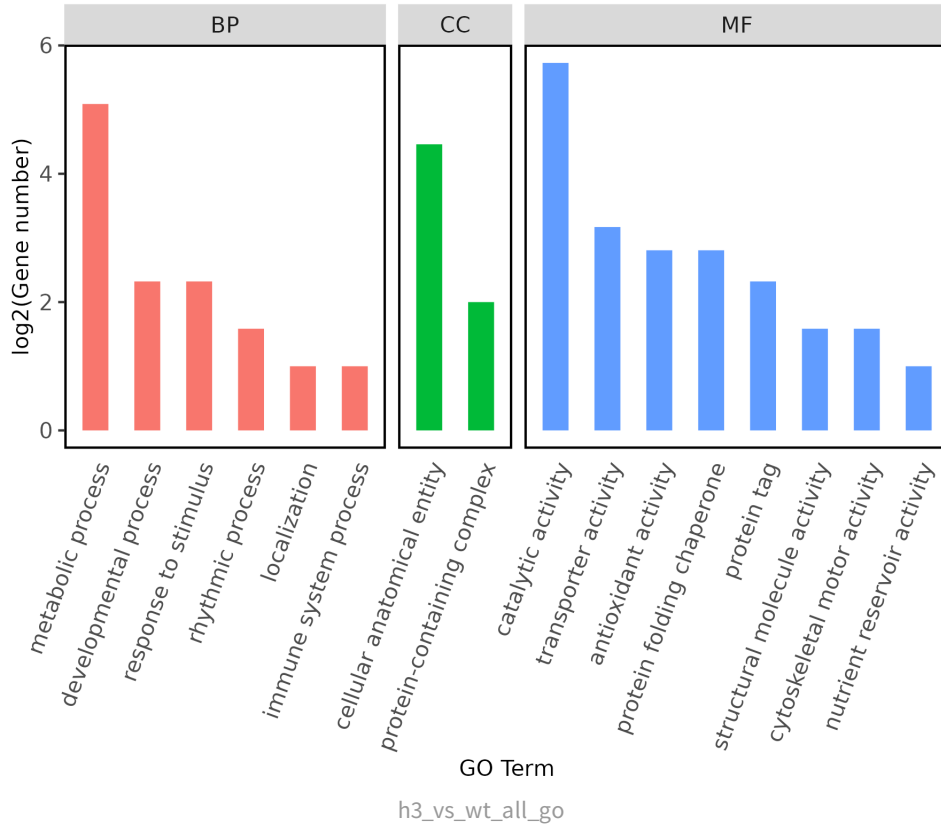


图10.2 GO功能聚类bar图

GO 功能聚类wego图展示。横坐标坐标是GO Term 名称，纵坐标是对应GO Term 中检出的基因个数，取对数。根据GO分成了3大类，分别是分子功能（Molecular Function, MF）、细胞组分（Cellular Component, CC）、生物过程（Biological Process, BP）。



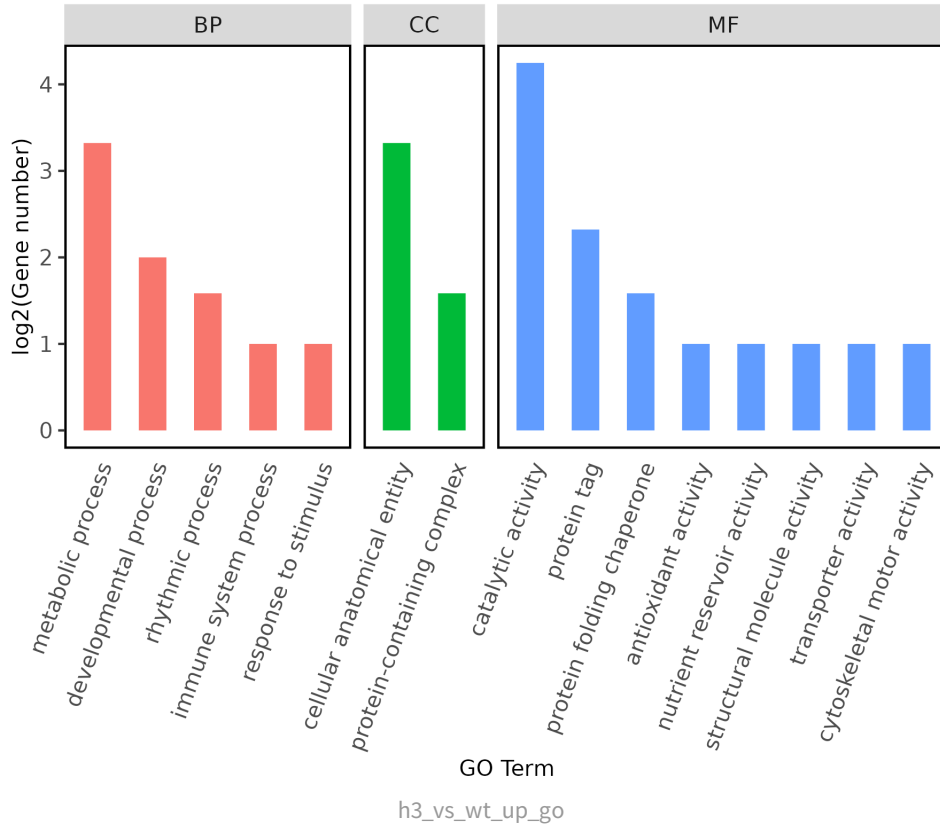


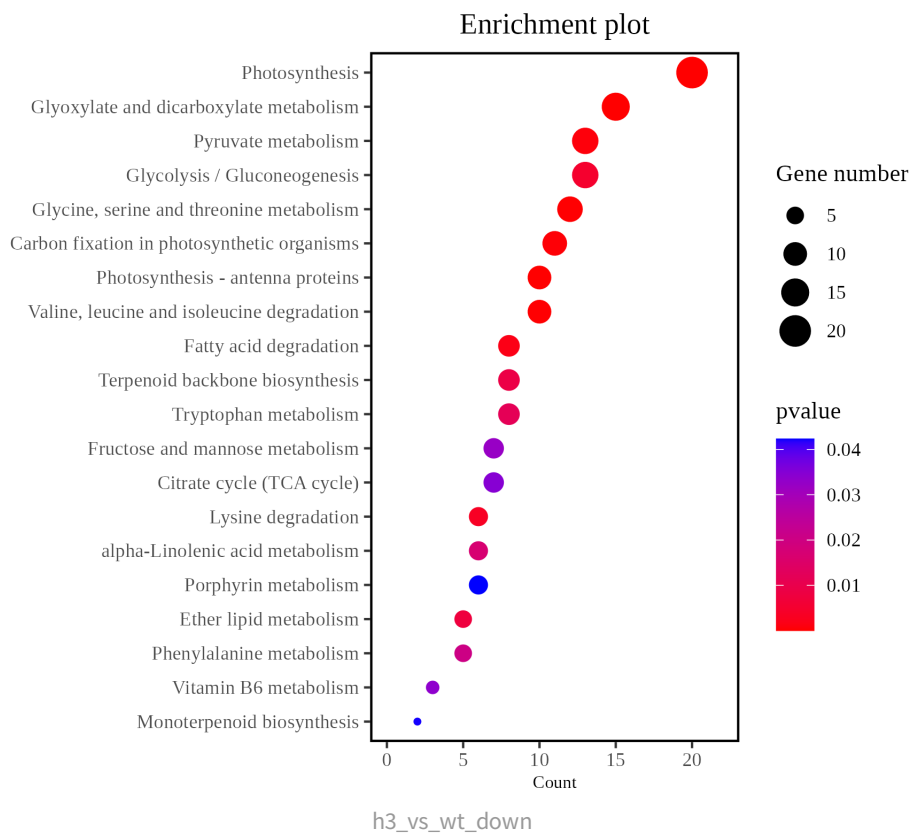
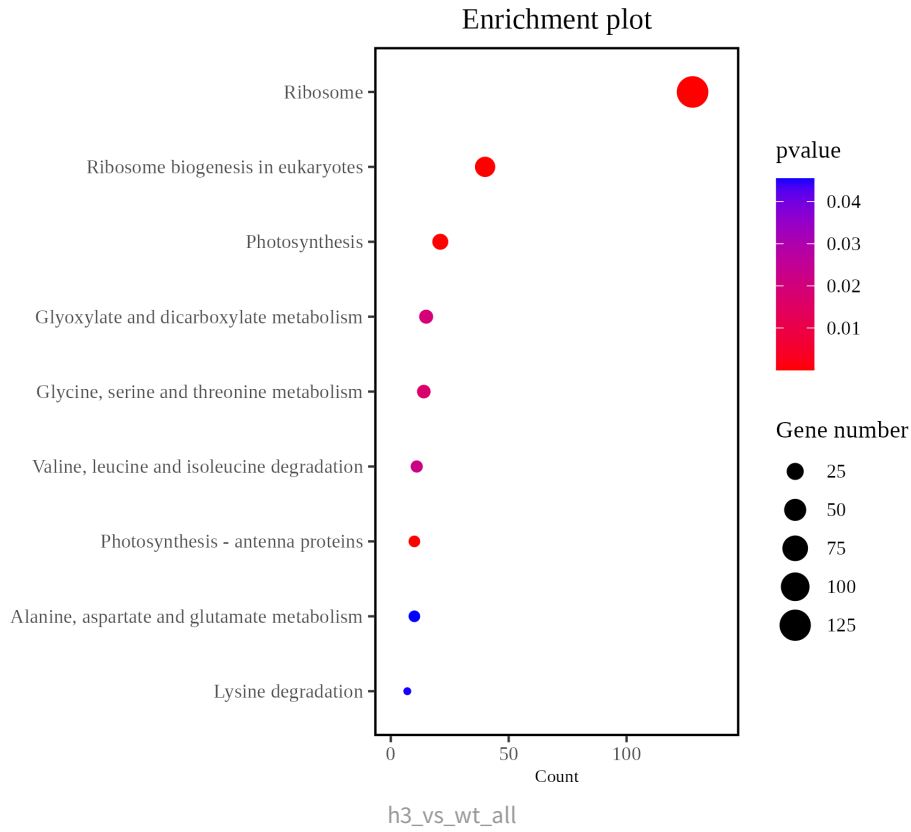
图10.3 GO功能聚类wego图

### 10.1.2 KEGG 功能分析

对于RNA项目，根据基因上调 (up)，下调(down)和上下调并集(all)分别进行分析。详细结果见./sup/10.gokegg/\*\_kegg。

KEGG功能聚类气泡图展示。纵坐标是KEGG pathway名称，横坐标是对应KEGG pathway中检出的基因占背景基因的个数，颜色代表显著性pvalue。

•



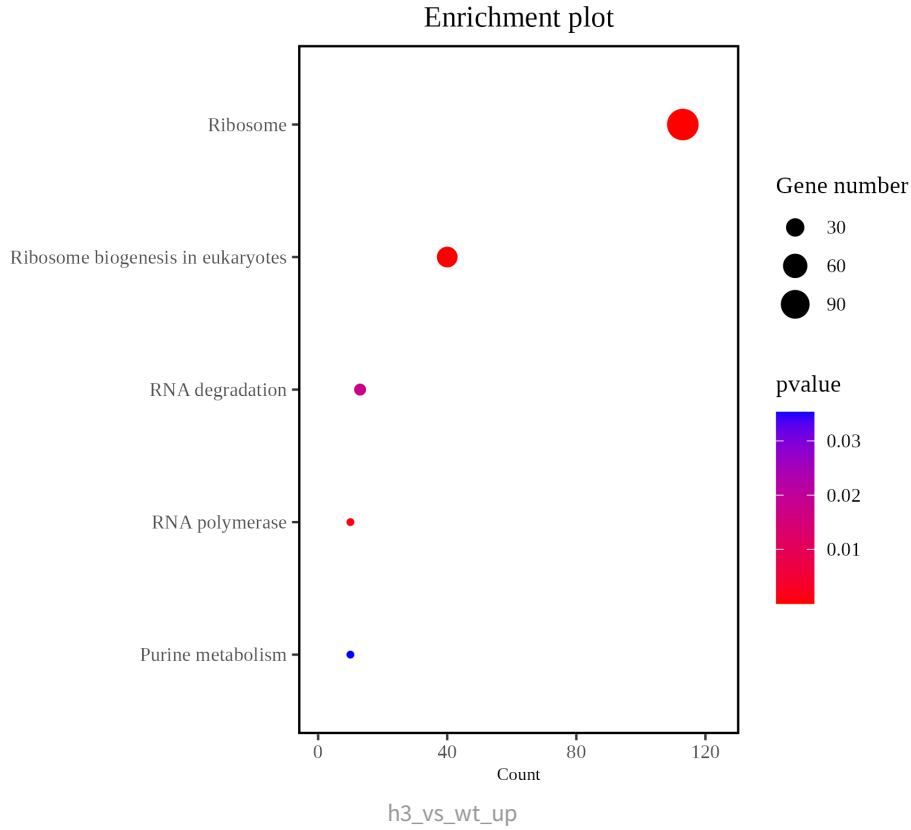
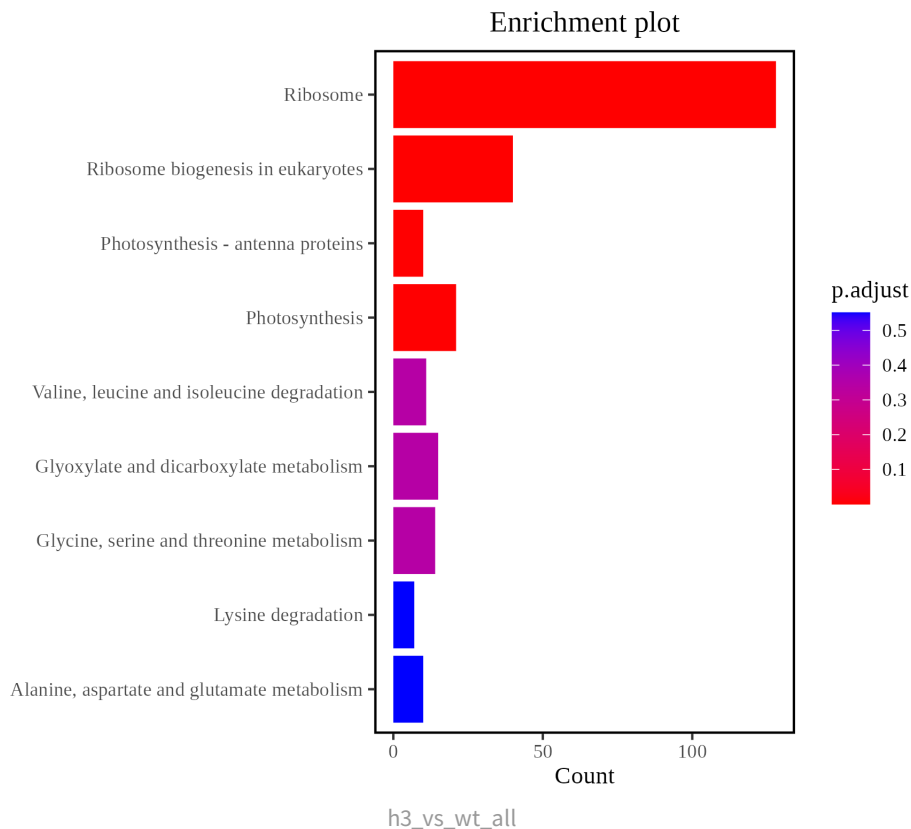


图10.4 功能聚类气泡图

KEGG功能聚类bar图展示。纵坐标是KEGG pathway名称，横坐标是对应KEGG pathway中检出的基因个数，颜色代表修正后的p.adjust。



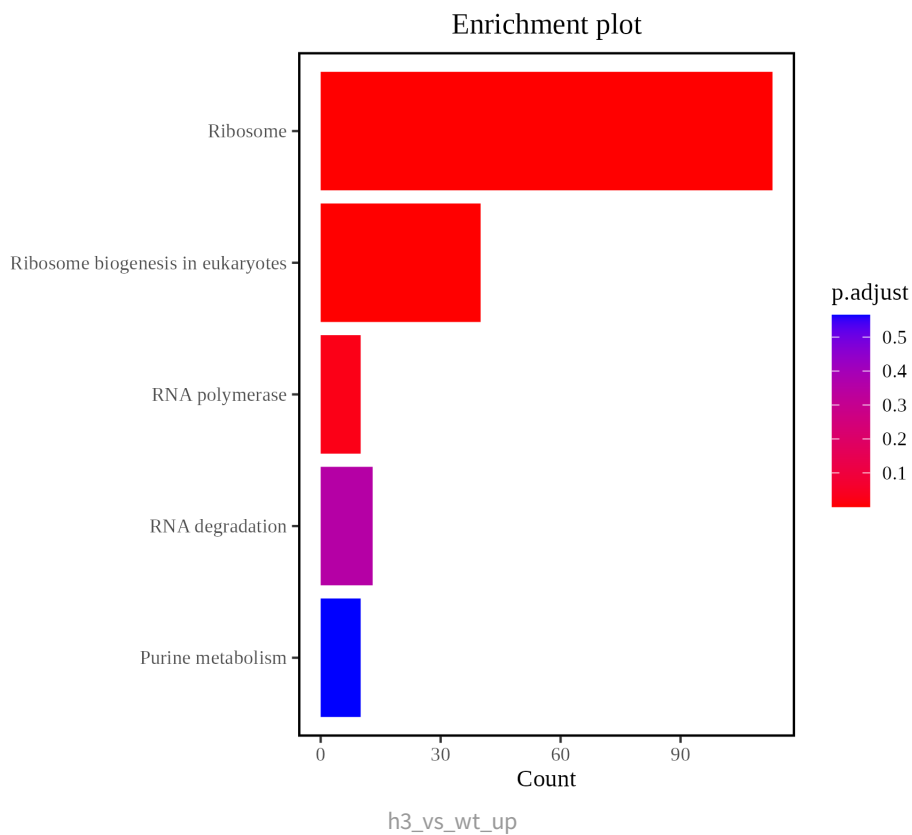
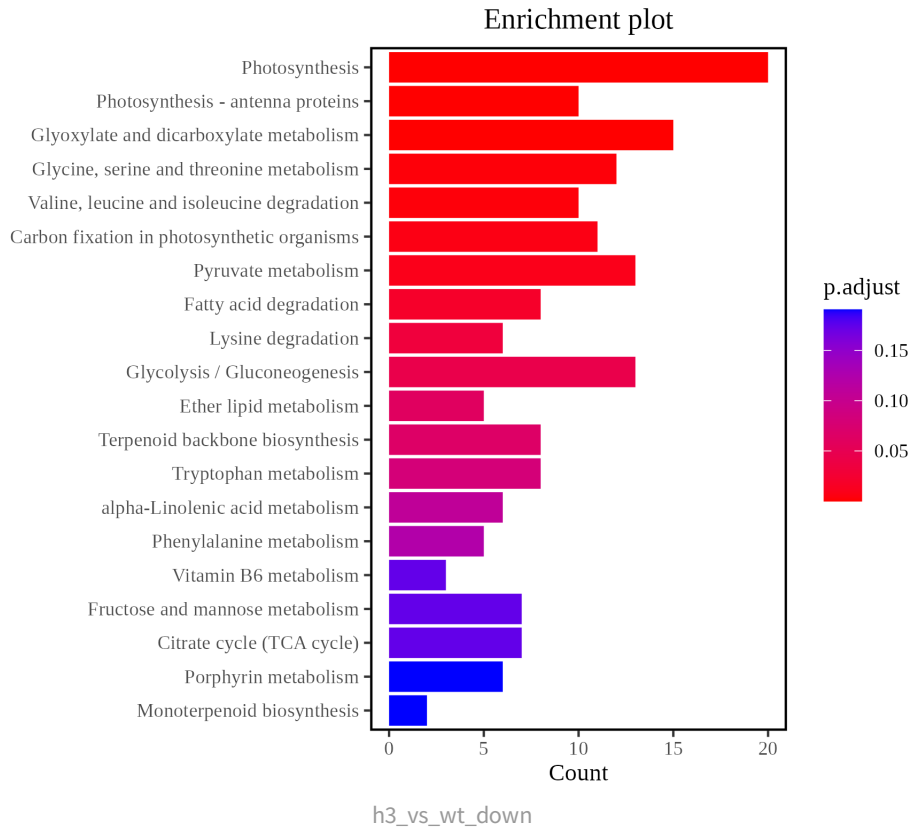
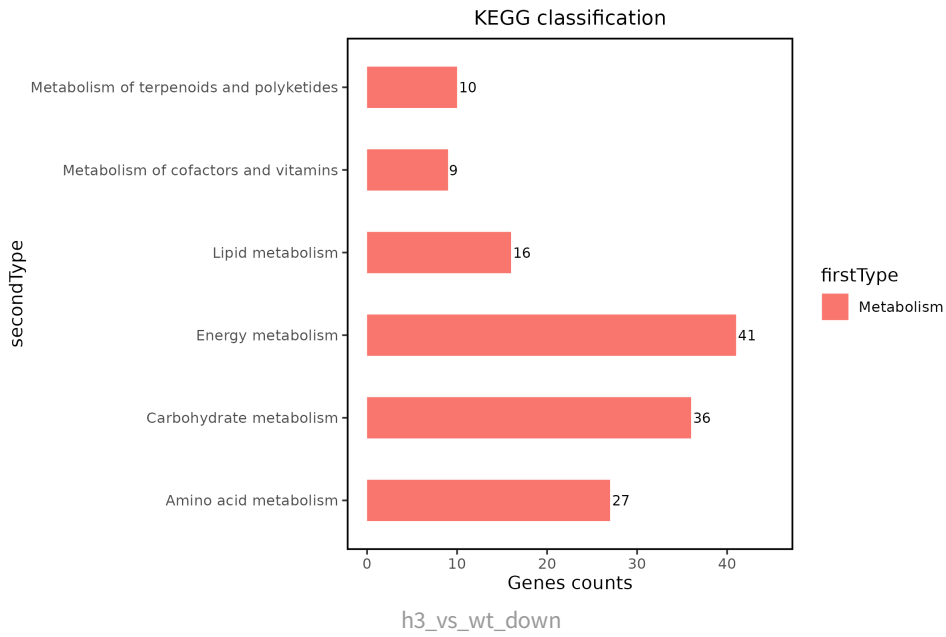
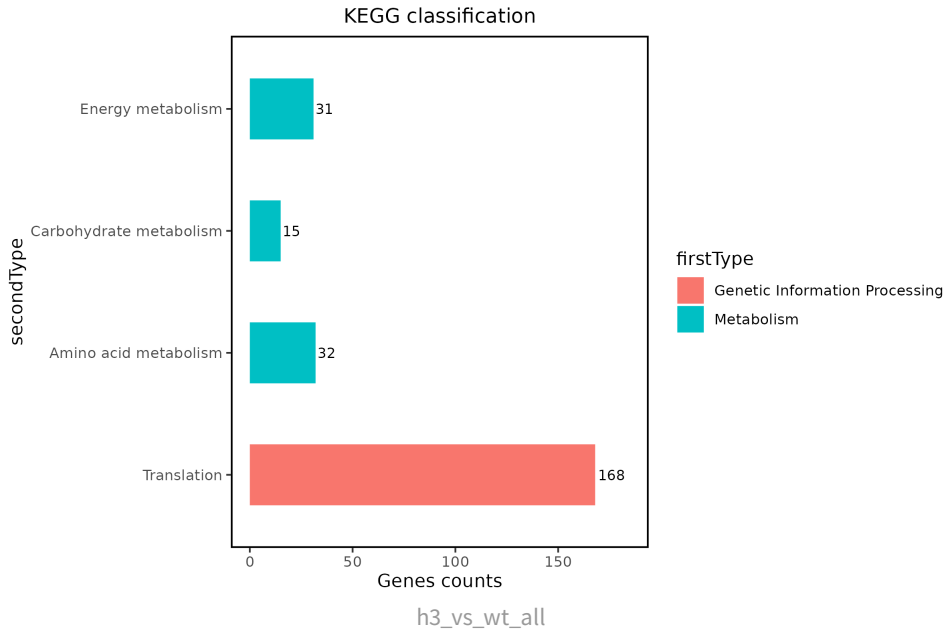
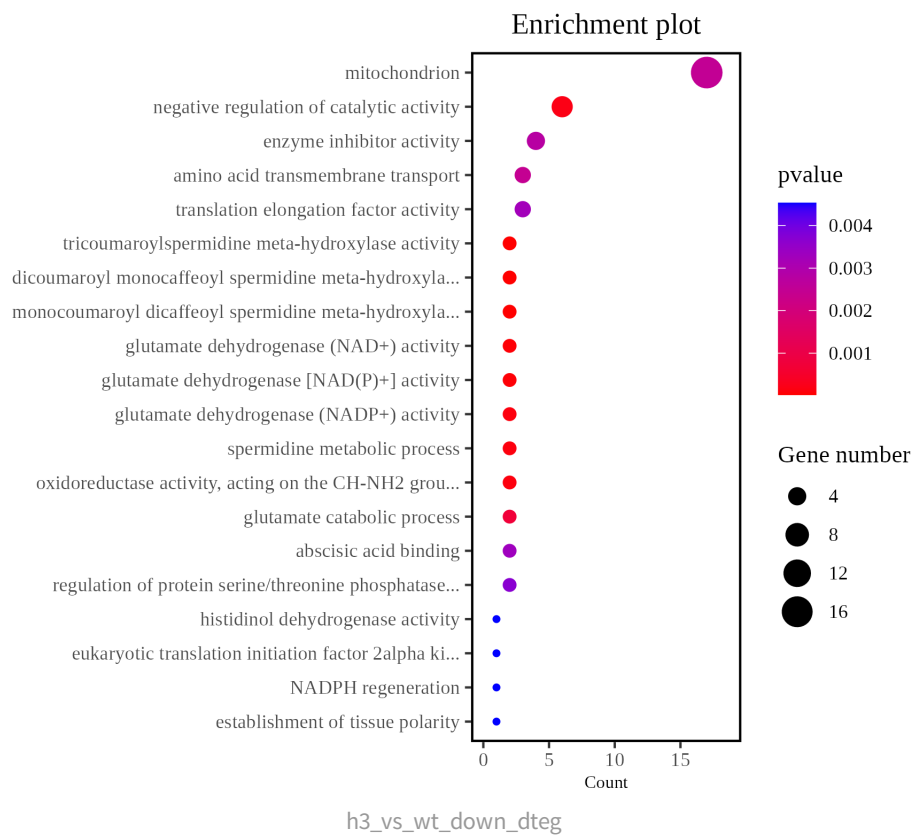
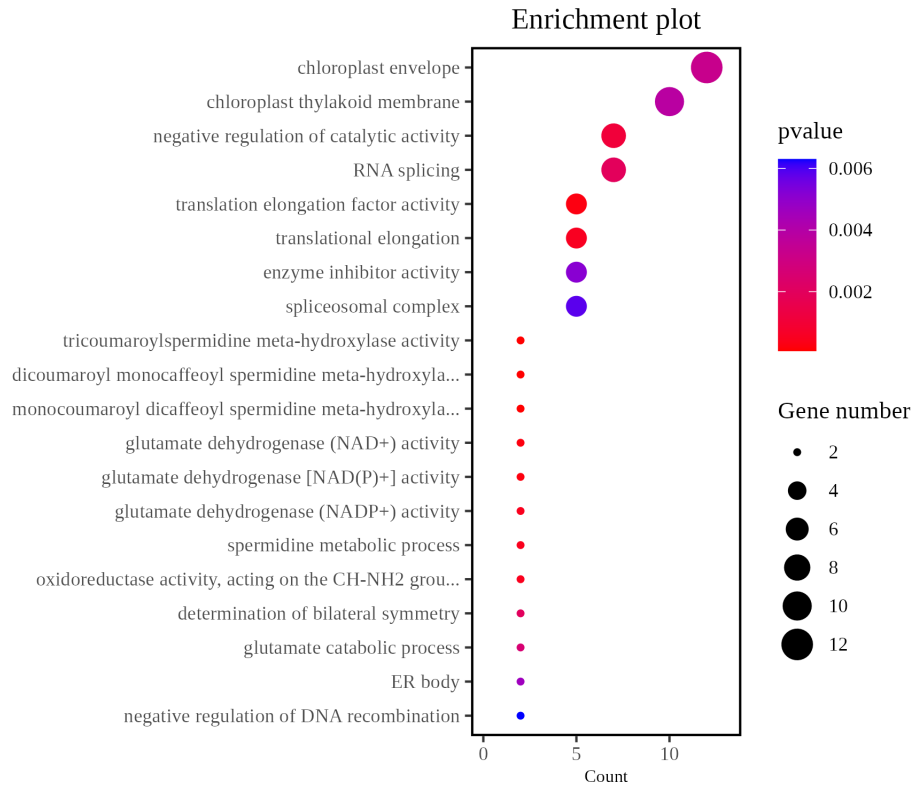


图10.5 功能聚类气泡图

KEGG Pathway主要划分为7类：分别为Metabolism， Genetic information Processing等。其中每类又分为二、三、四级子条目。功能聚类分类图展示pathway所属一级条目和二级条目的情况。







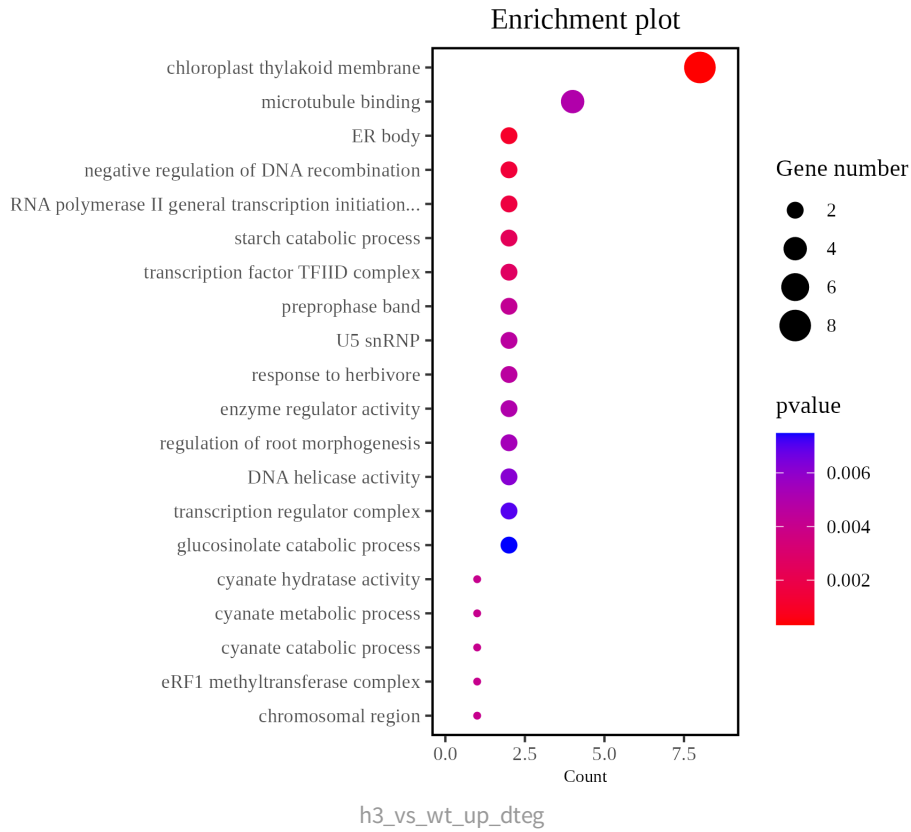
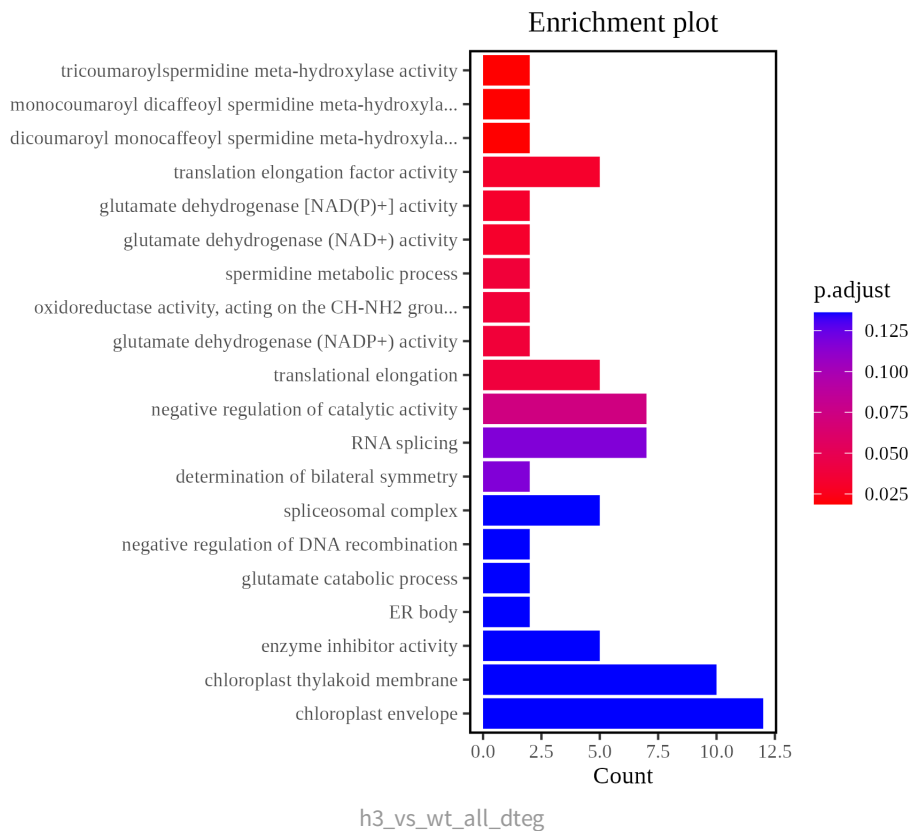


图10.8 GO功能聚类气泡图

GO 功能聚类bar图展示。纵坐标是GO Term 名称，横坐标是对应GO Term 中检出的基因个数，颜色代表显著性p.adjust。



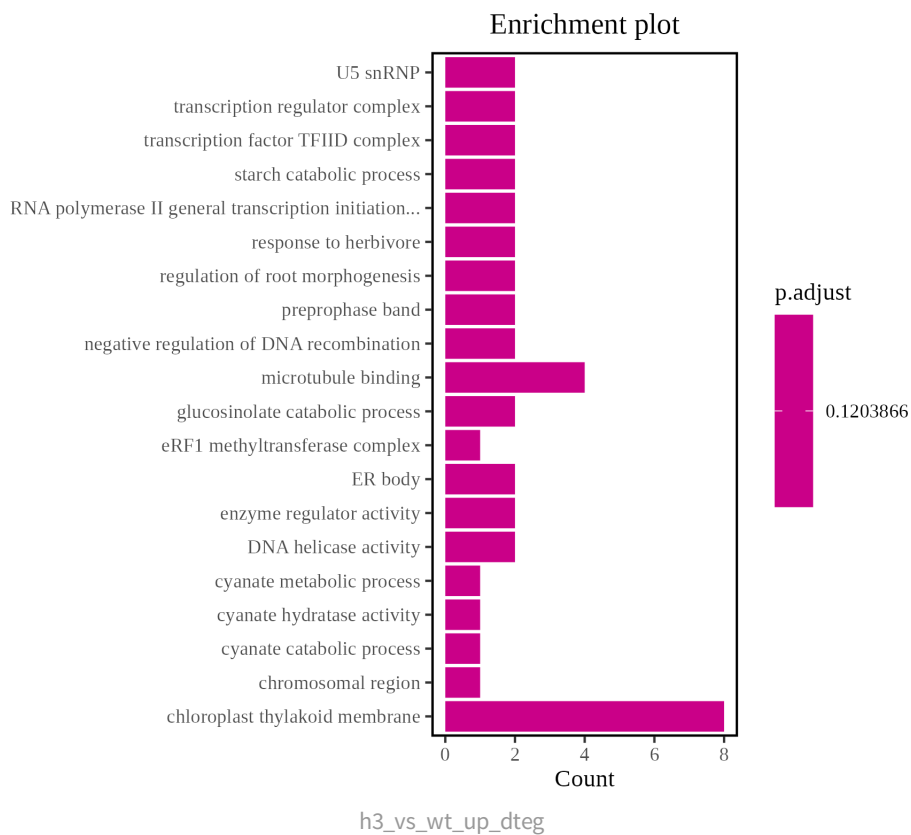
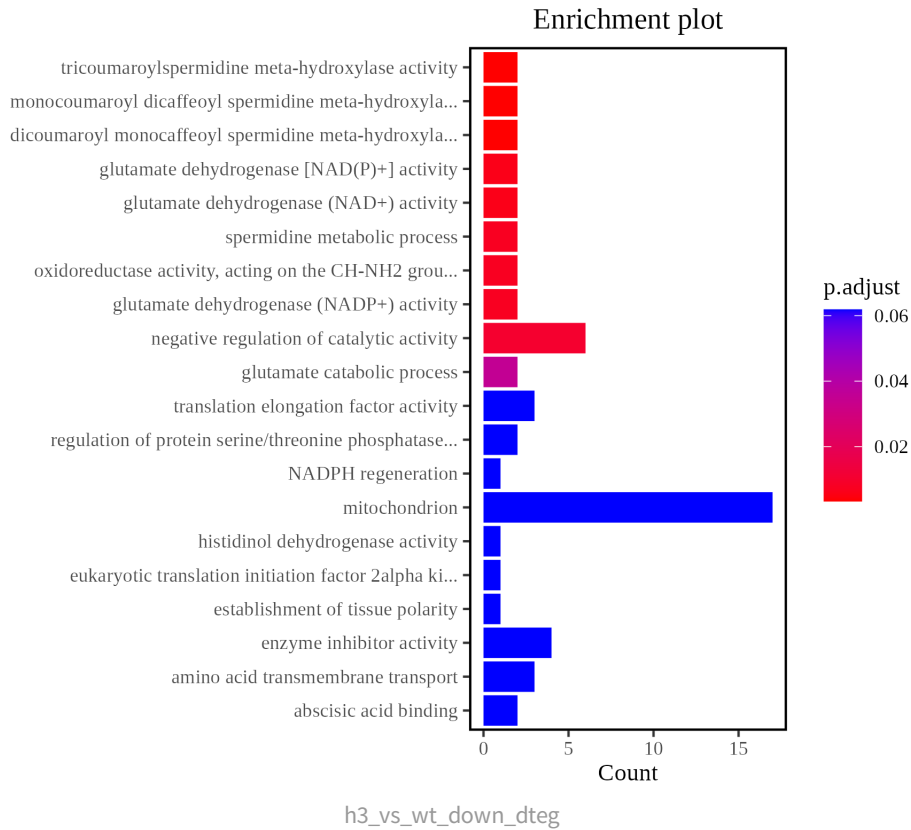
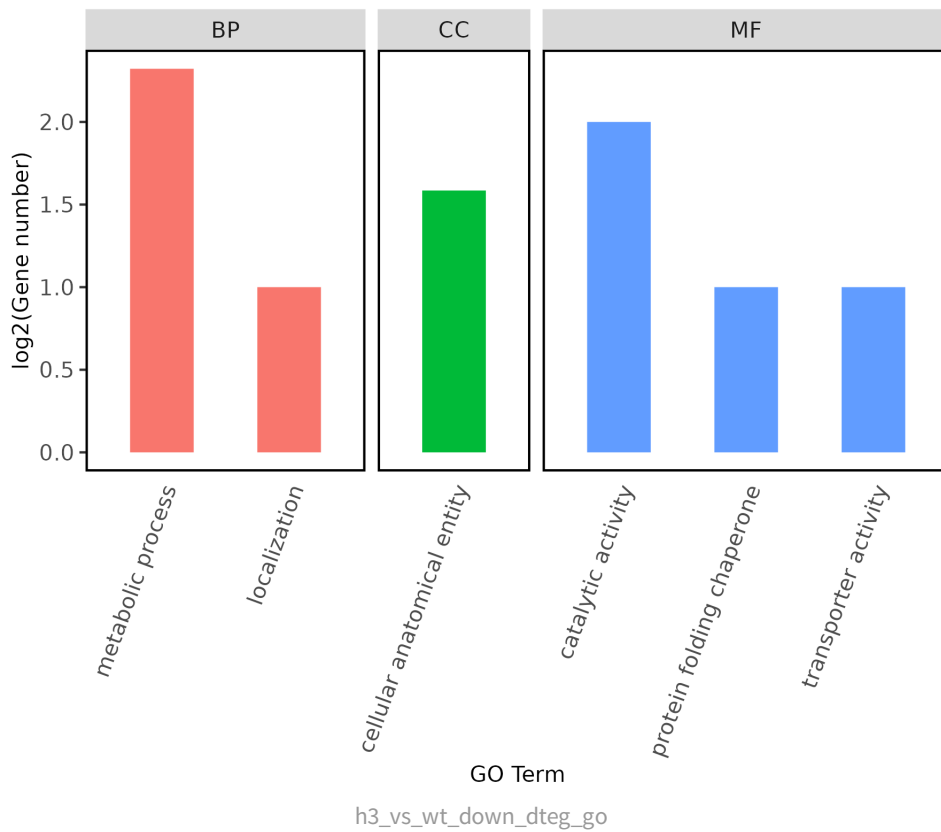
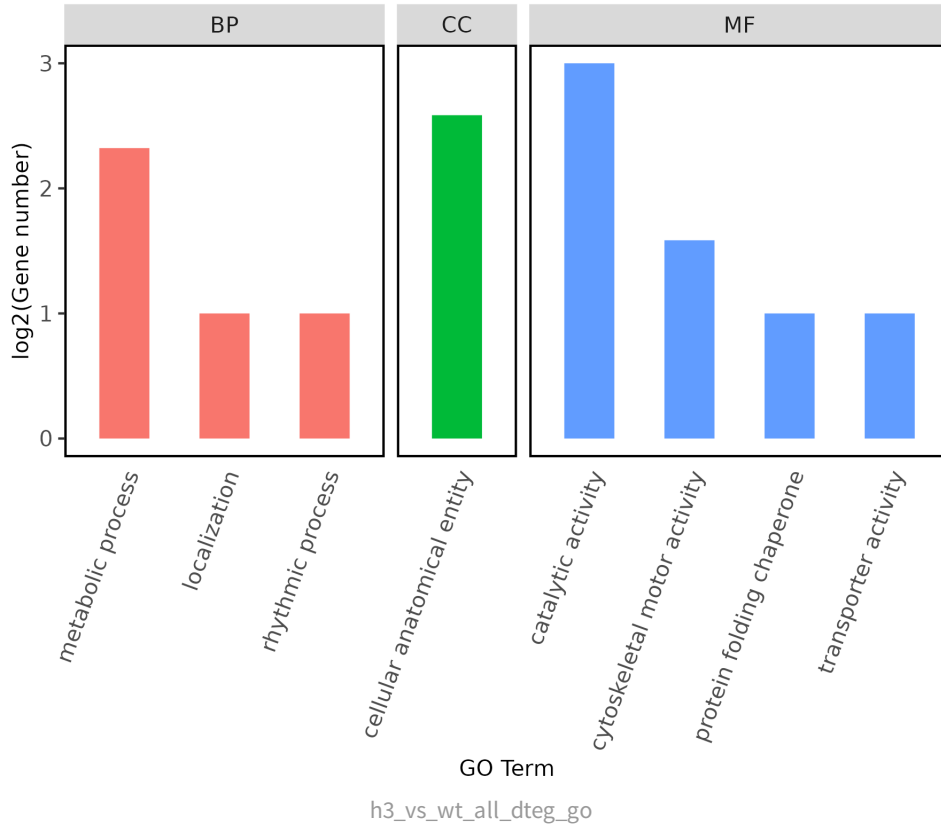


图10.9 GO功能聚类bar图

GO 功能聚类wego图展示。横坐标坐标是GO Term 名称，纵坐标是对应GO Term 中检出的基因个数，取对数。根据GO分成了3大类，分别是分子功能（Molecular Function, MF）、细胞组分（Cellular Component, CC）、生物过程（Biological Process, BP）。





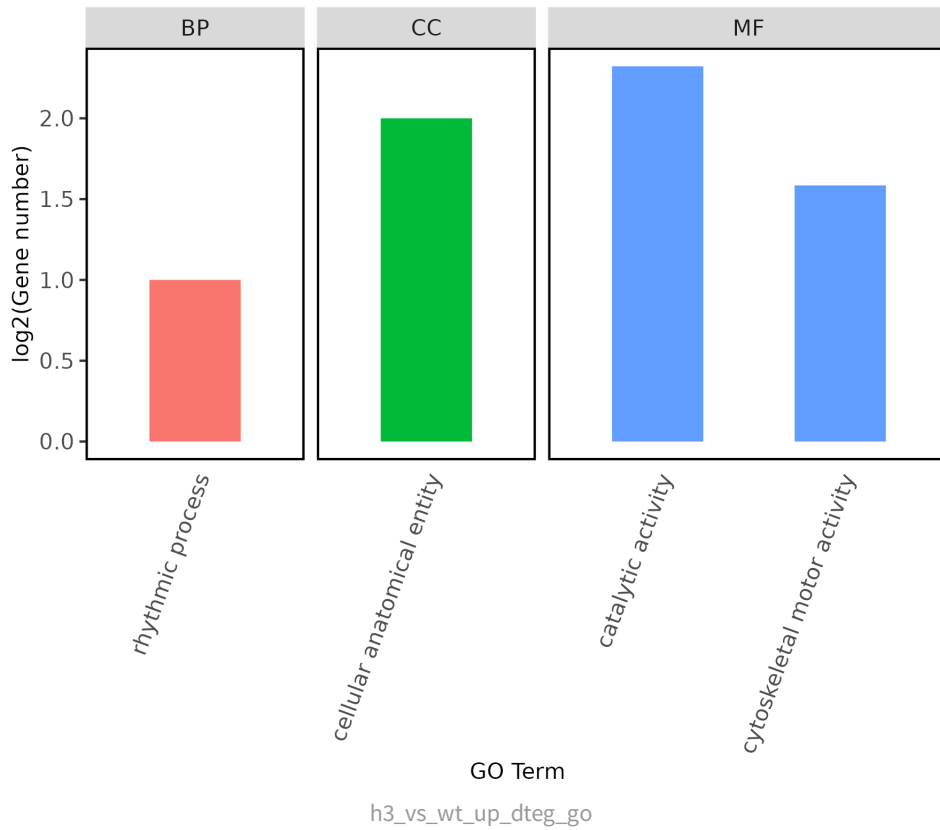


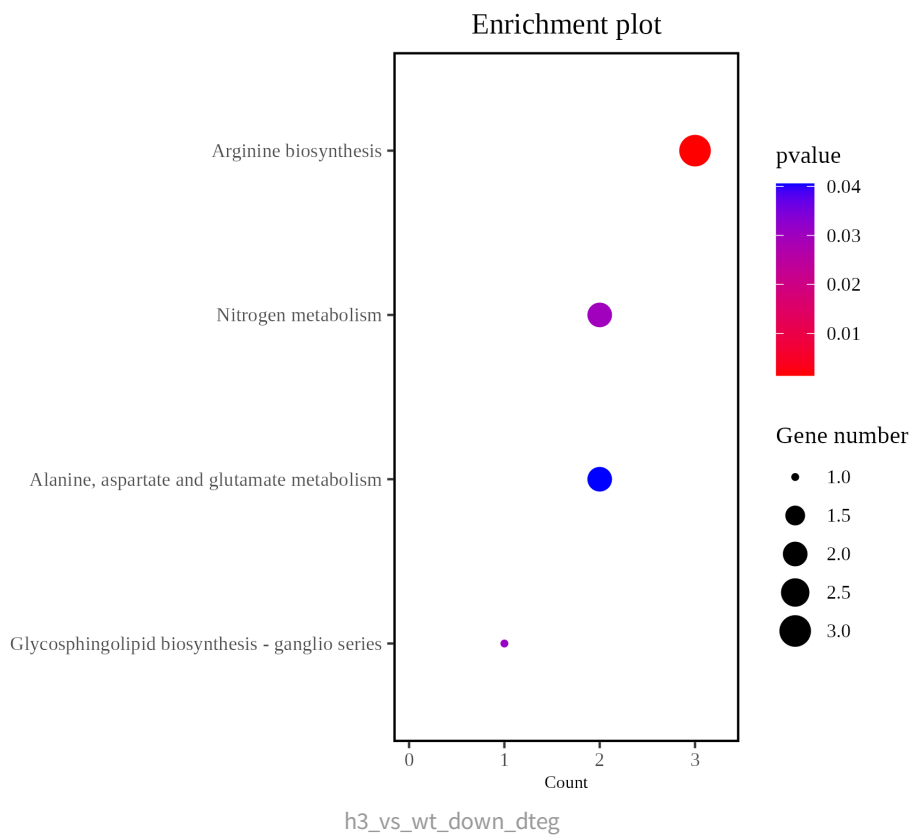
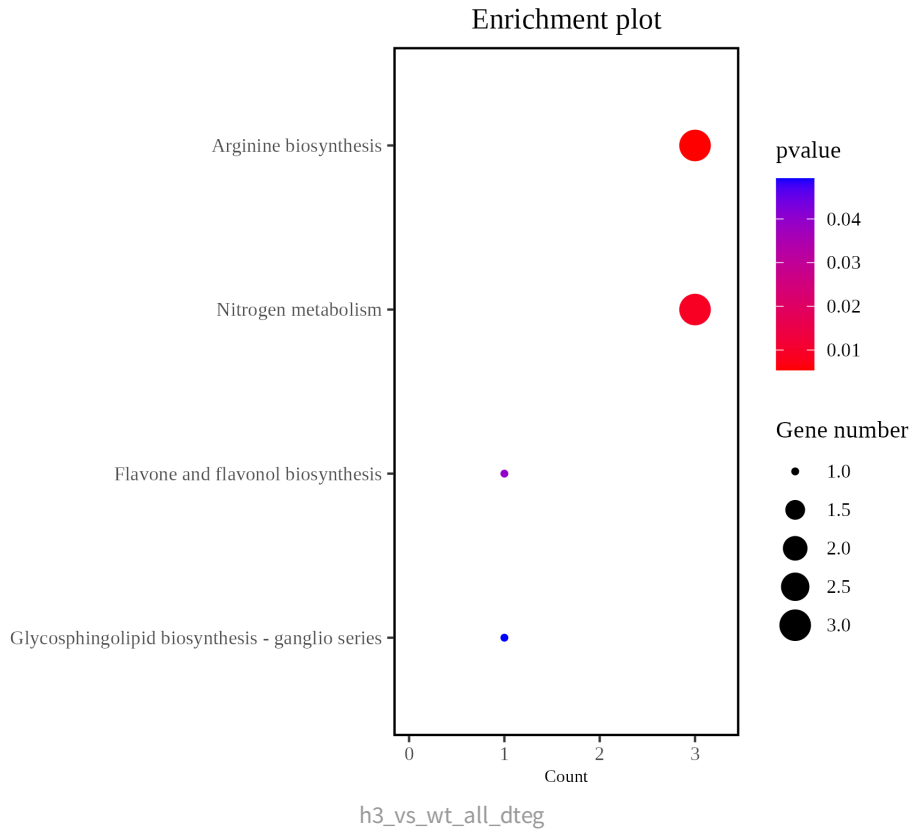
图10.10 GO功能聚类wego图

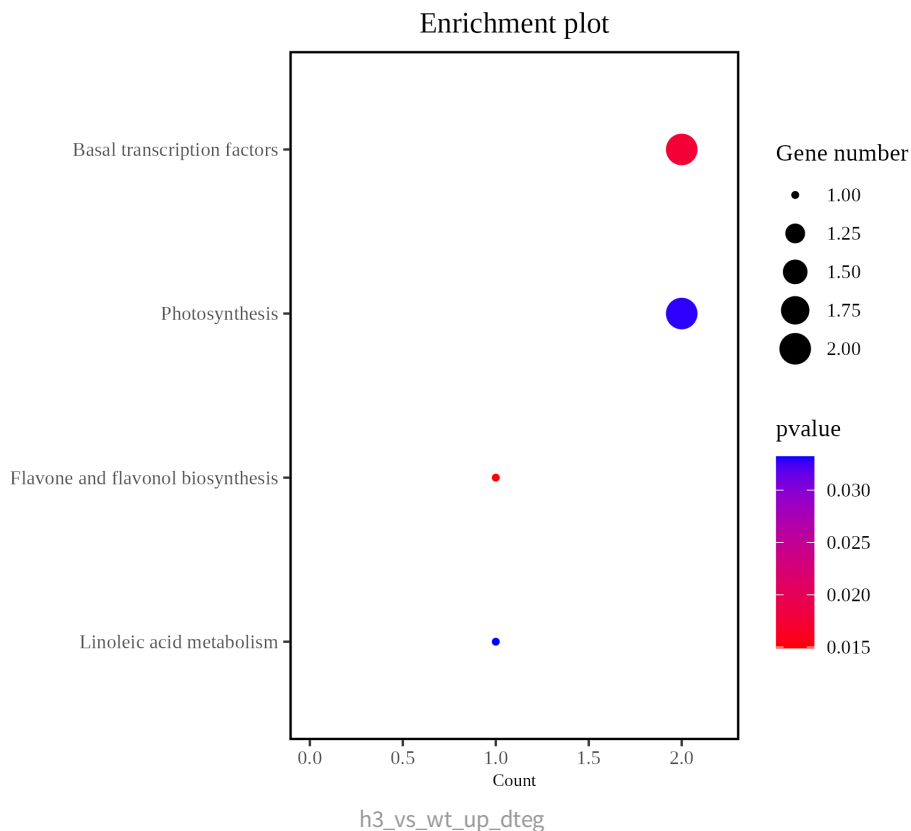
## 10.2.2 KEGG 功能分析

对于RNA项目，根据基因上调 (up)，下调(down)和上下调并集(all)分别进行分析。详细结果见./sup/10.gokegg/dteg/\*\_kegg。

KEGG功能聚类气泡图展示。纵坐标是KEGG pathway名称，横坐标是对应KEGG pathway中检出的基因占背景基因的个数，颜色代表显著性pvalue。

•

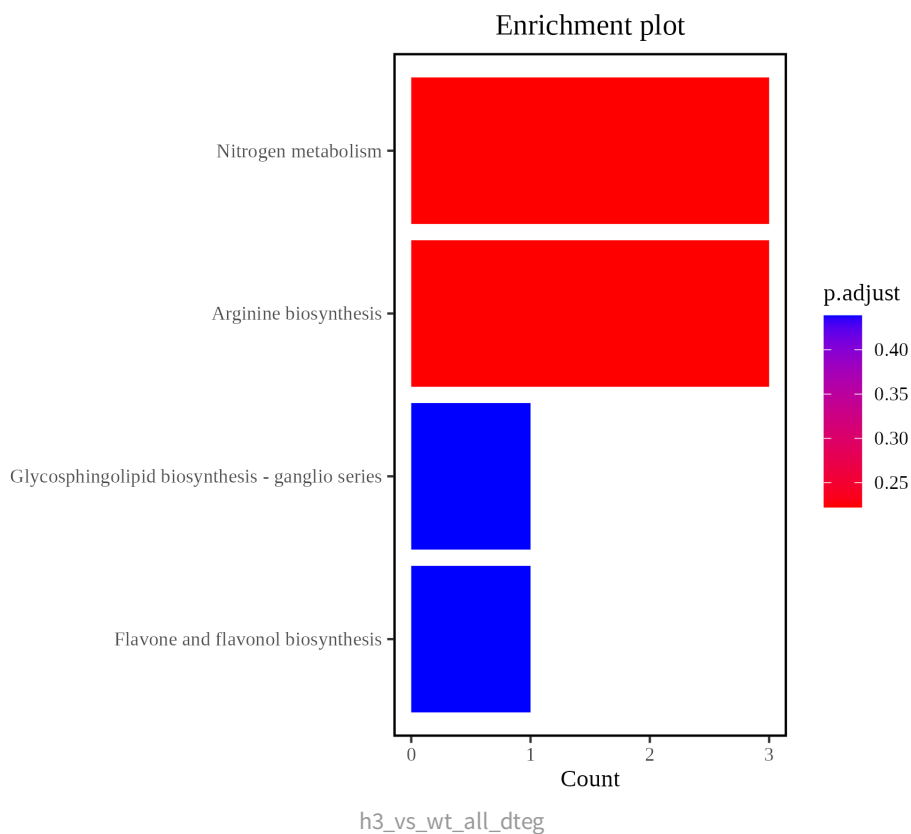




pdf版报告

图10.11 功能聚类气泡图

KEGG功能聚类bar图展示。纵坐标是KEGG pathway名称，横坐标是对应KEGG pathway中检出的基因个数，颜色代表修正后的p.adjust。



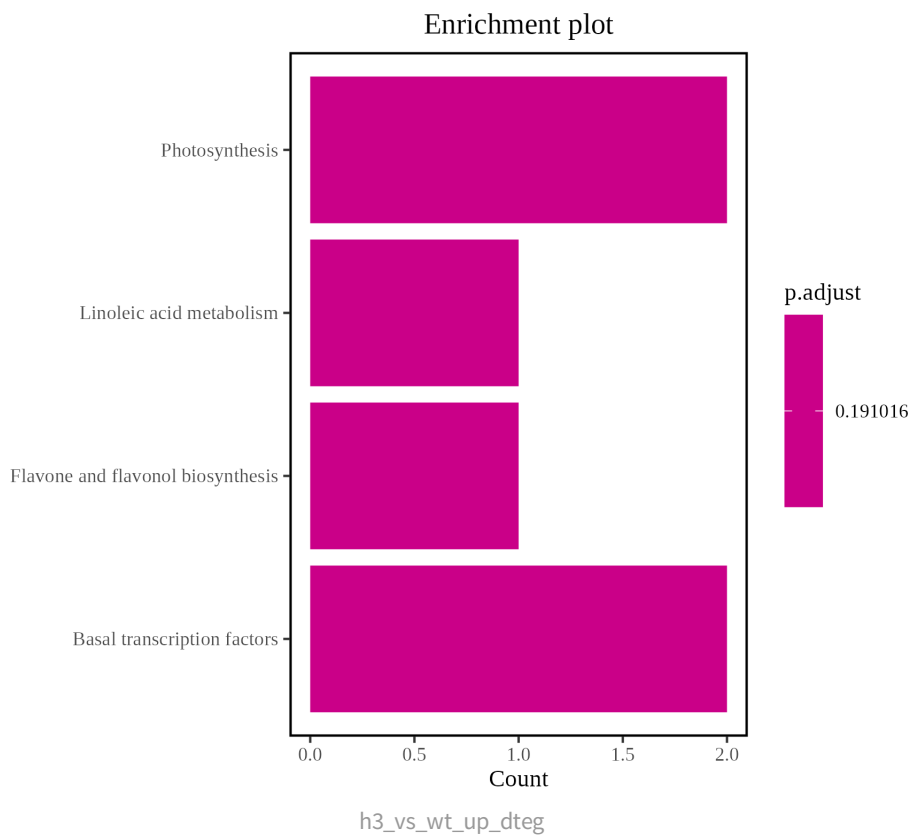
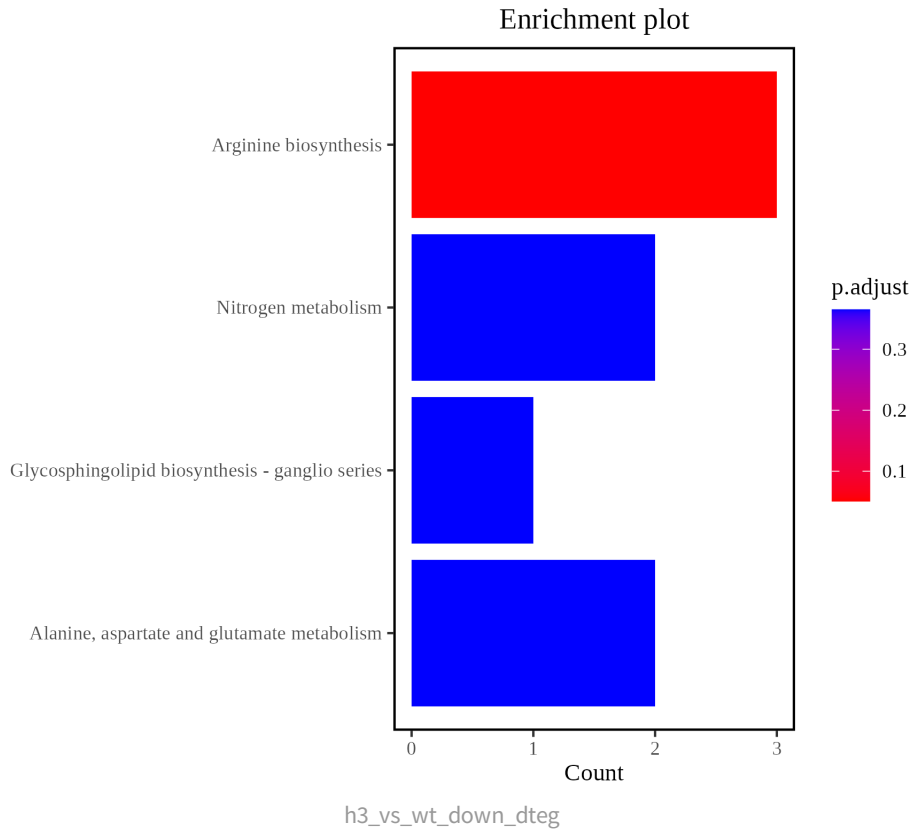


图10.12 功能聚类气泡图

KEGG Pathway主要划分为7类：分别为Metabolism， Genetic information Processing等。其中每类又分为二、三、四级子条目。功能聚类分类图展示pathway所属一级条目和二级条目的情况。

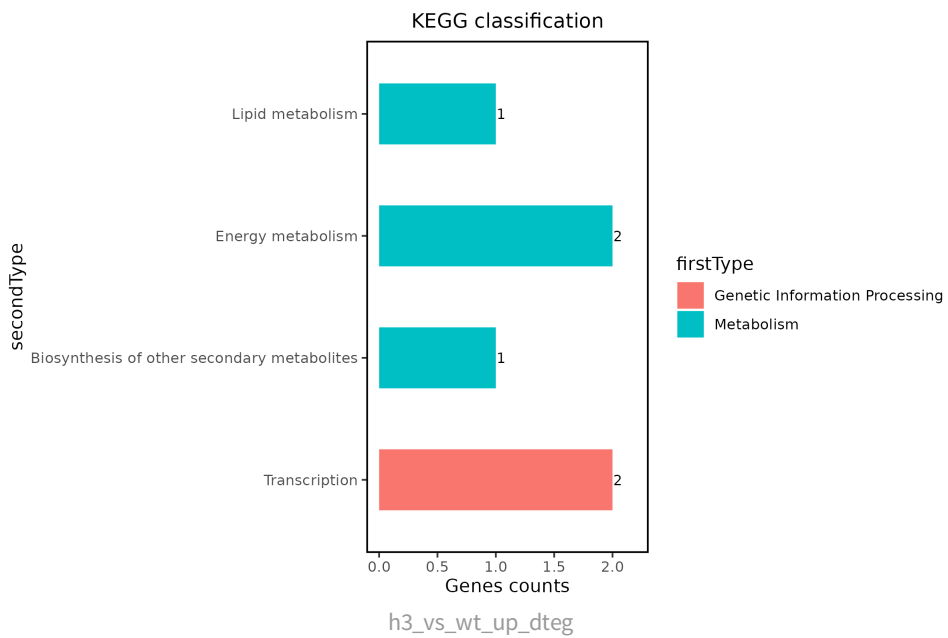
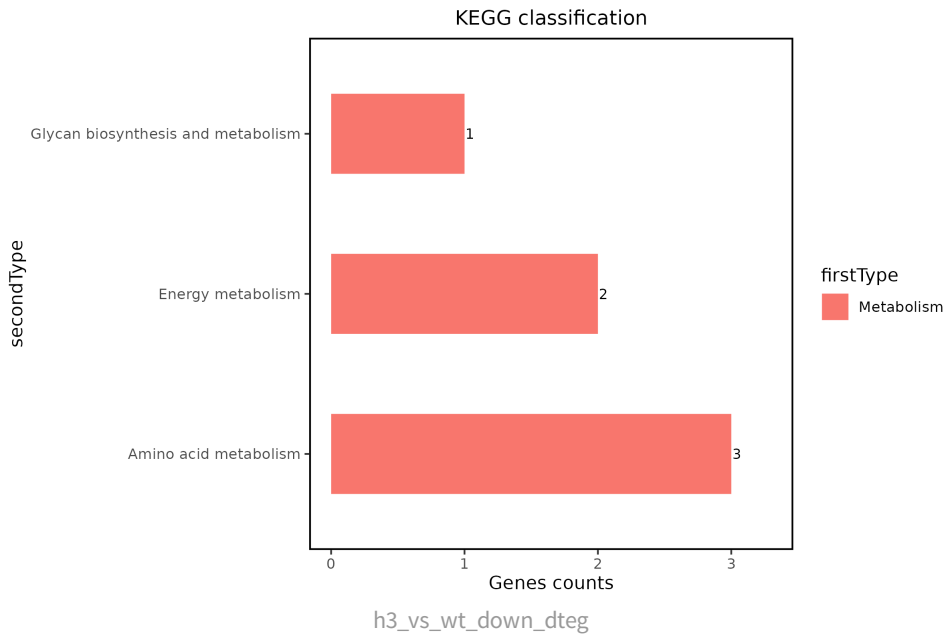
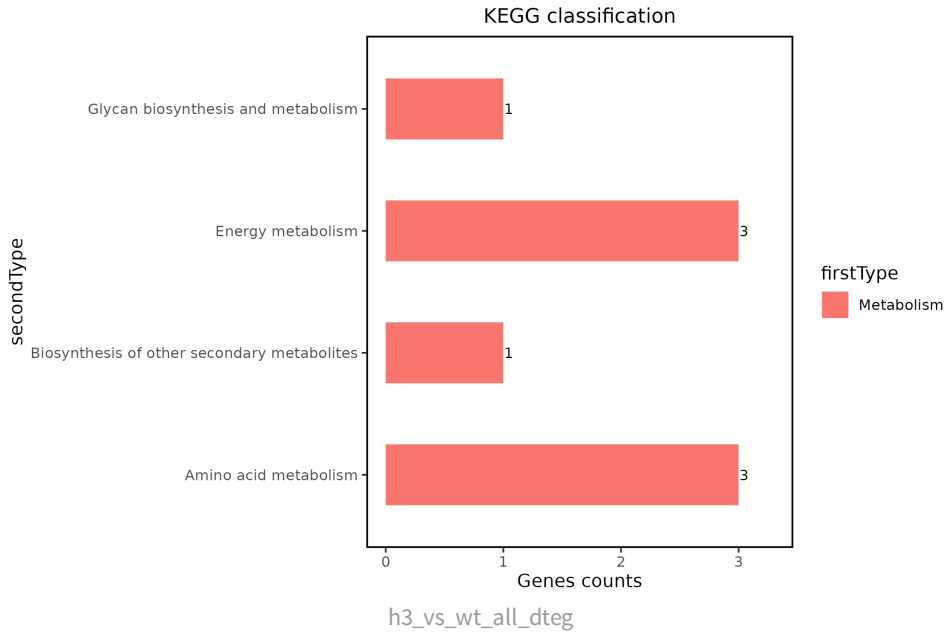


图10.13 功能聚类分类图

KEGG通路图具有强大的图形功能，将代谢、调控、通路、生化、疾病、药物等相关的分子相互作用整合到一起并形成关系网络，通过功能富集分析得到的KEGG通路示例图如下，部分富集结果中可能没有KEGG通路图。具体文件见./sup/10.gokegg/dteg/\*\_kegg 文件夹各组下面的map文件夹。

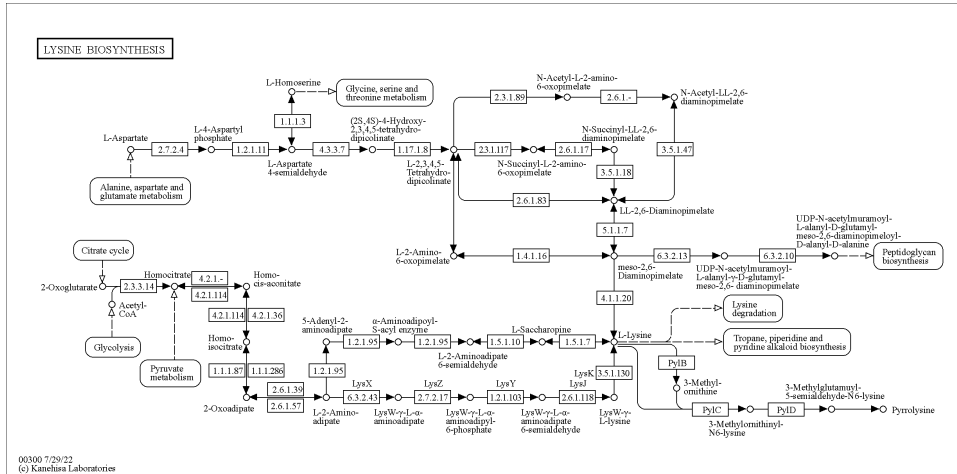


图10.14 KEGG通路示意图

## □ 10.3 转录翻译变化基因

我们对转录翻译变化基因进行GO与KEGG富集分析。基因注释结果请见./sup/10.gokegg/with\_mrna/。其中“homo”代表转录和翻译同向变化，“opposite”表达转录和翻译反向变化。

### ■ 10.3.1 GO 功能分析

GO 功能聚类气泡图展示。纵坐标是GO Term 名称，横坐标是对应GO Term 中检出的基因占背景基因的个数，颜色代表显著性pvalue。

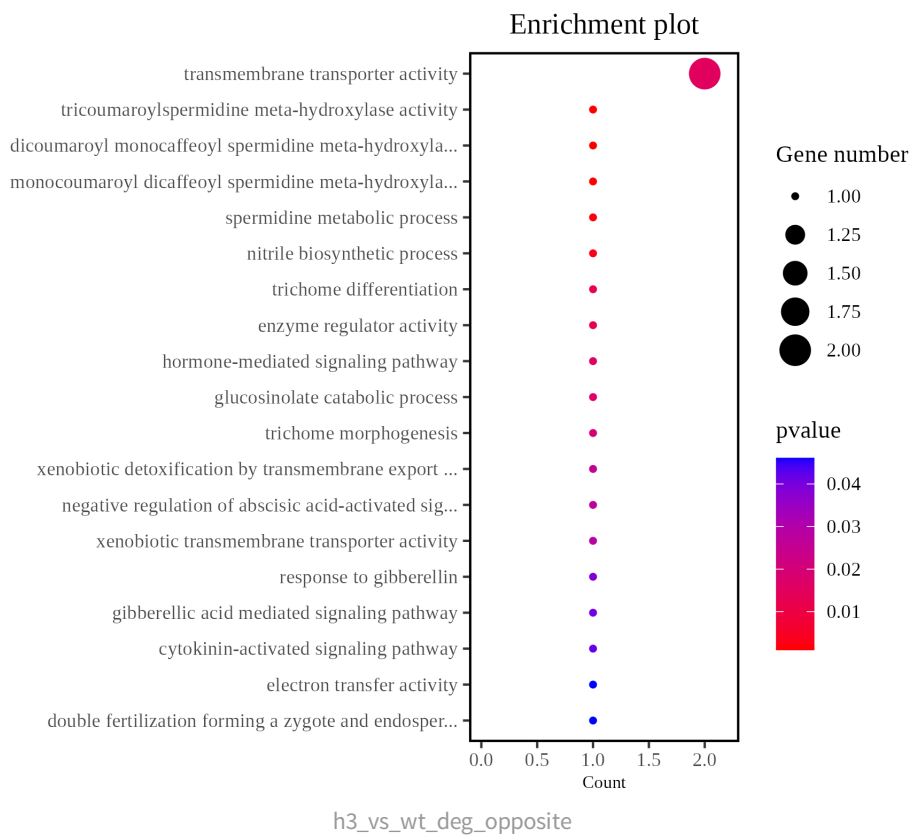
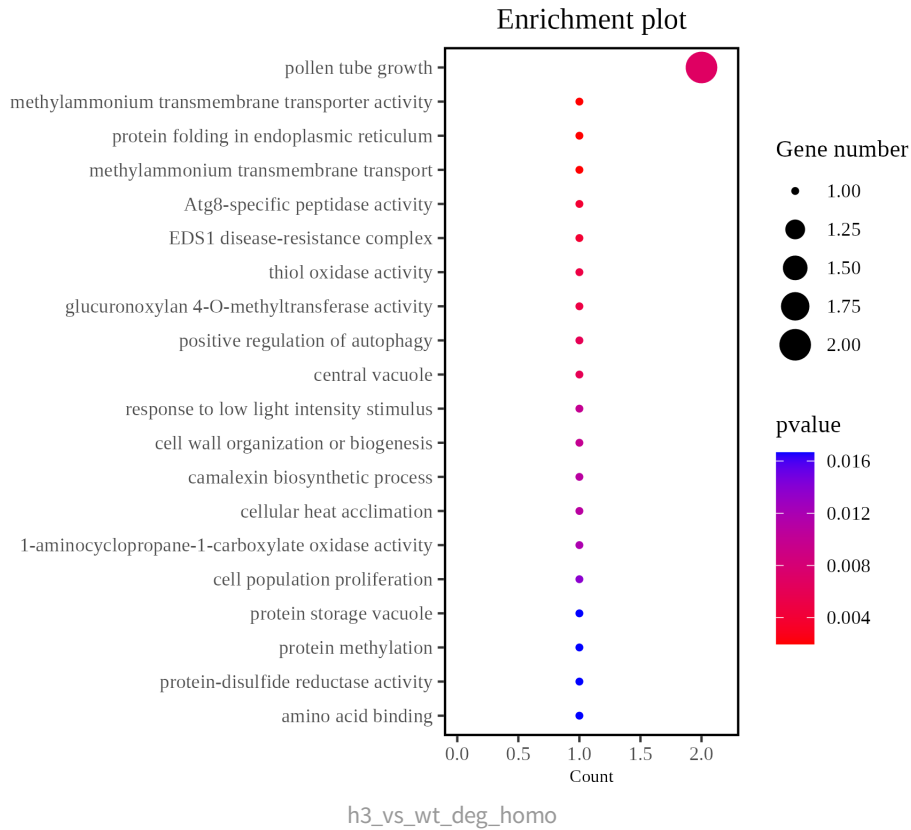


图10.15 GO功能聚类气泡图

GO 功能聚类bar图展示。纵坐标是GO Term 名称，横坐标是对应GO Term 中检出的基因个数，颜色代表显著性p.adjust。

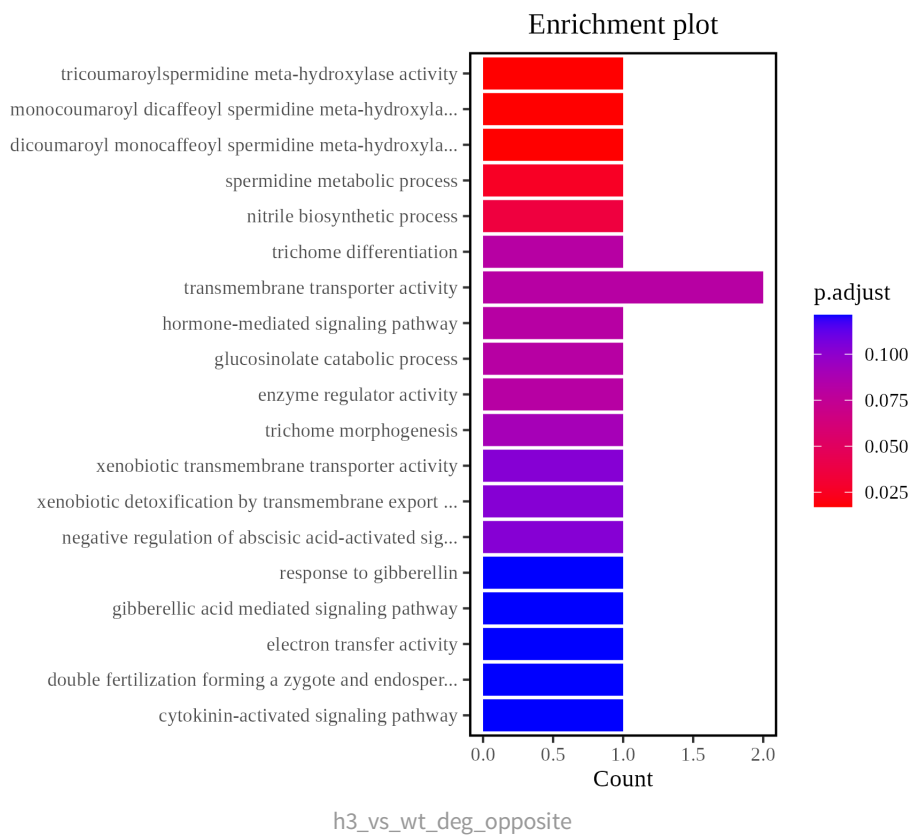
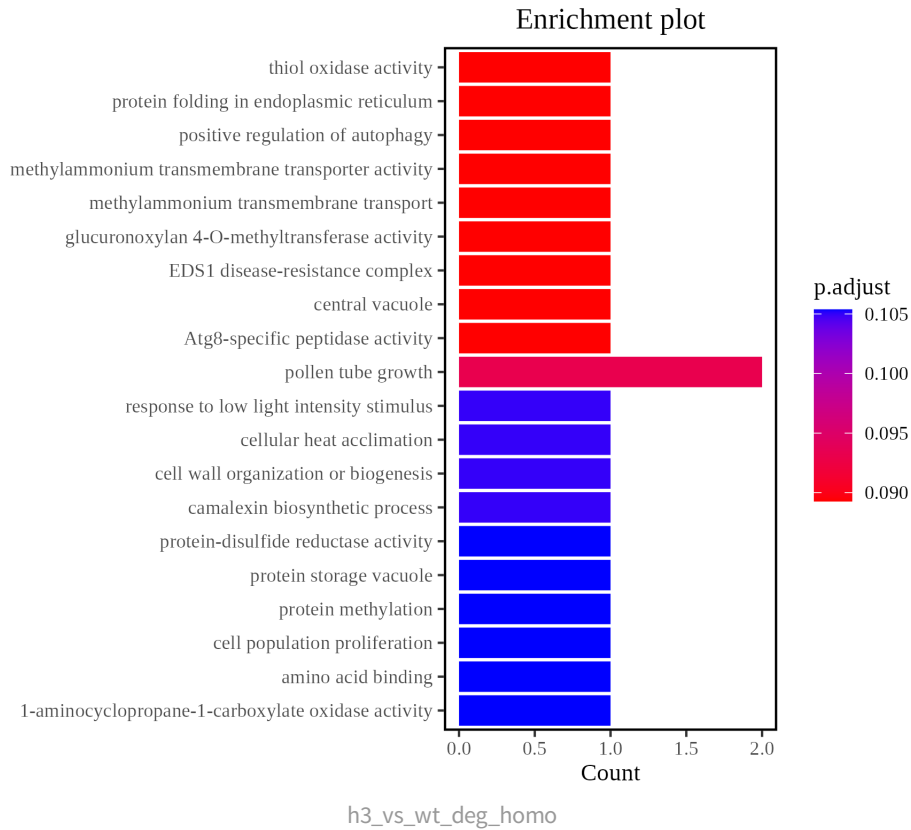


图10.16 GO功能聚类bar图

GO 功能聚类wego图展示。横坐标坐标是GO Term 名称，纵坐标是对应GO Term 中检出的基因个数，取对数。根据GO分成了3大类，分别是分子功能（Molecular Function, MF）、细胞组分（Cellular Component, CC）、生物过程（Biological Process, BP）。

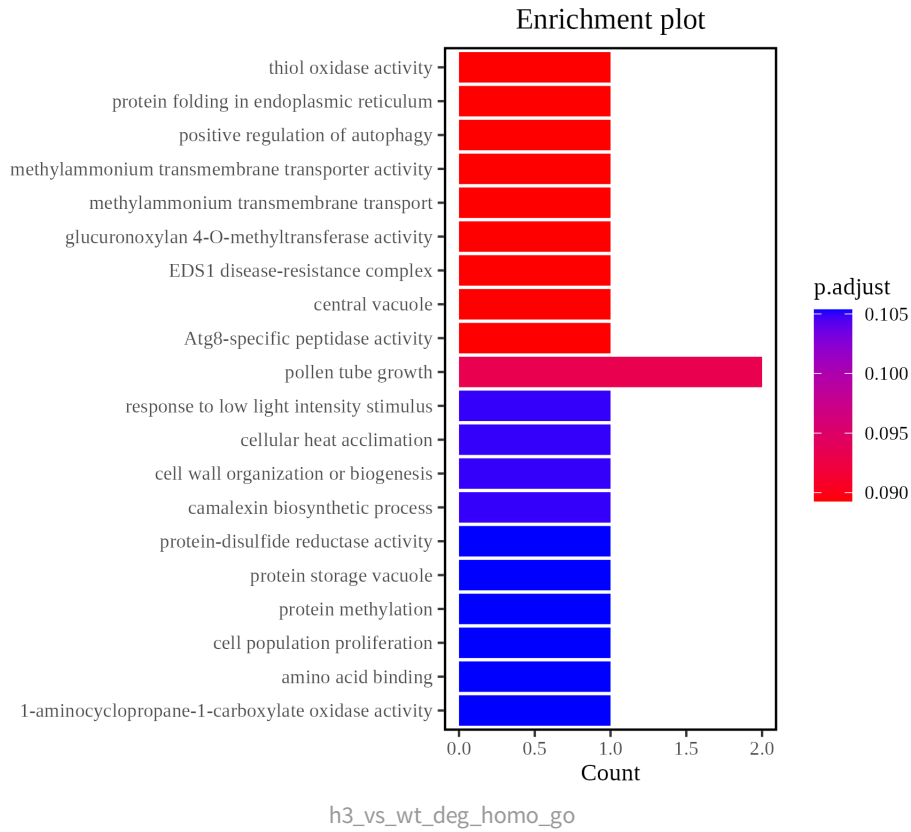
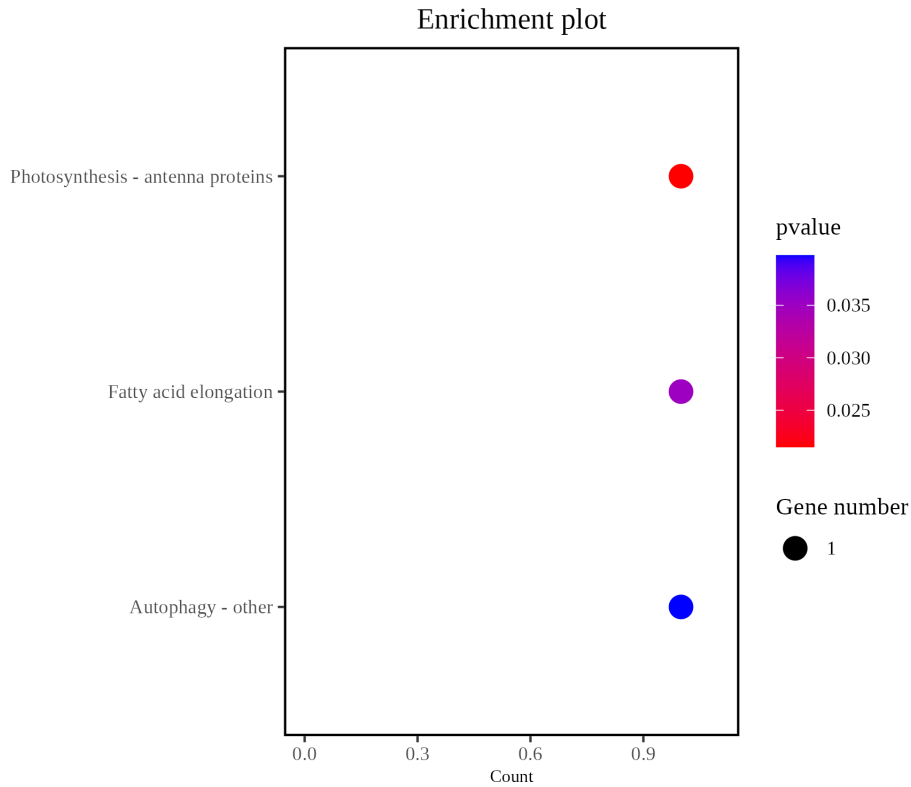


图10.17 GO功能聚类wego图

### 10.3.2 KEGG 功能分析

KEGG详细结果见./sup/10.gokegg/with\_mrna/\*\_kegg。

KEGG功能聚类气泡图展示。纵坐标是KEGG pathway名称，横坐标是对应KEGG pathway中检出的基因占背景基因的个数，颜色代表显著性pvalue。

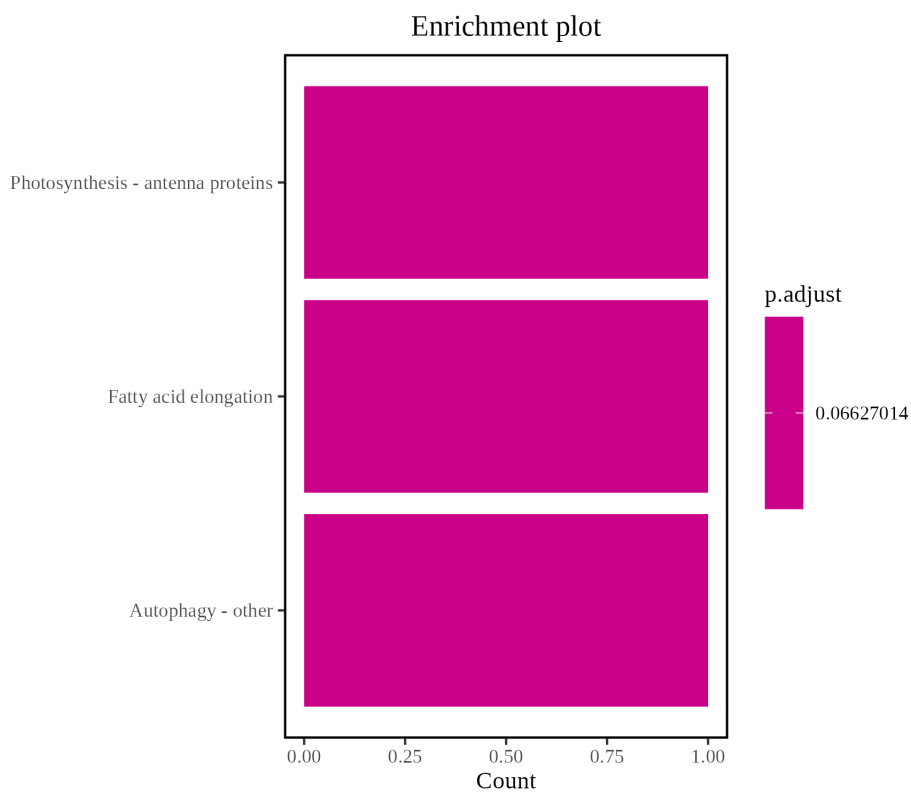


h3\_vs\_wt\_deg\_homo

h3\_vs\_wt\_deg\_opposite

图10.18 功能聚类气泡图

KEGG功能聚类bar图展示。纵坐标是KEGG pathway名称，横坐标是对应KEGG pathway中检出的基因个数，颜色代表修正后的p.adjust。

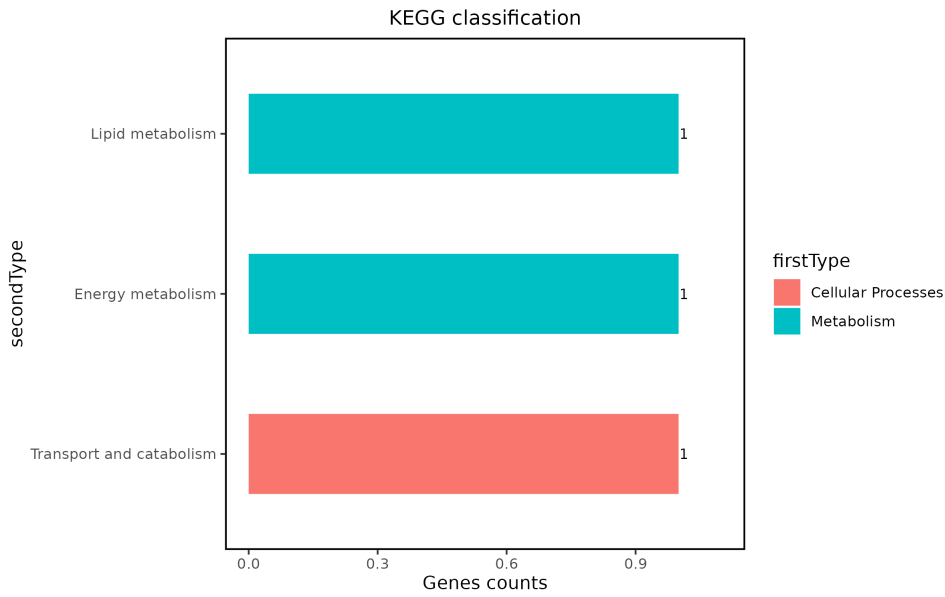


h3\_vs\_wt\_deg\_homo

h3\_vs\_wt\_deg\_opposite

图10.19 功能聚类气泡图

KEGG Pathway主要划分为7类：分别为Metabolism， Genetic information Processing等。其中每类又分为二、三、四级子条目。功能聚类分类图展示pathway所属一级条目和二级条目的情况。



h3\_vs\_wt\_deg\_homo

h3\_vs\_wt\_deg\_opposite

图10.20 功能聚类分类图

KEGG通路图具有强大的图形功能，将代谢、调控、通路、生化、疾病、药物等相关的分子相互作用整合到一起并形成关系网络，通过功能富集分析得到的KEGG通路示例图如下，部分富集结果中可能没有KEGG通路图。具体文件见./sup/10.gokegg/with\_mrna/\*\_kegg 文件夹各组下面的map文件夹。

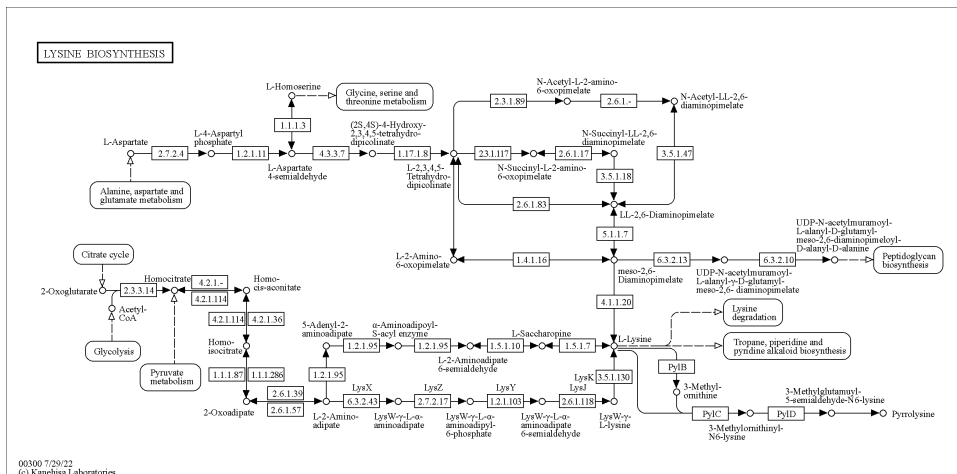


图10.21 KEGG通路示意图

Q:进行富集分析意义是什么？

A:富集分析是一种常用的生物信息学分析方法，用于识别基因集合中富含的生物学意义和功能。它们的作用主要包括以下几个方面：1. 生物学解释：通过富集分析，可以将基因集合中富含的生物学意义和功能与相关文献进行比对，从而对基因集合的生物学含义进行解释。2. 基因集合筛选：通过富集分析，可以筛选出与研究对象相关的基因集合，从而缩小研究范围，提高研究效率。3. 基因表达分析：通过富集分析，可以对基因的表达模式和调控机制进行分析和预测，从而更好地理解基因在生物学过程中的作用。

## 11 备注

### □ 11.1 软件列表

此报告中涉及到的分析所用的软件及相关信息如下。

表11.1 软件说明

分析	软件	版本	参数
clean	fastp	v0.23.2	--length_required 15 -q 20
去除rRNA	Bowtie	v1.3.1	-n 0 -y -a --norc --best --strata -S -l 15
比对参考基因组	STAR	2.7.11a	--outFilterMultimapNmax 1 --outFilterMatchNmi
覆盖度统计	bamCoverage	3.5.1	-bs 200000
reads在染色体上的分布	ggplot2	3.5.0	-
RPF分析	RiboCode	1.2.13	默认参数
元基因分析	RiboMiner	0.2	默认参数
表达量分析	featureCounts	v2.0.1	默认参数
翻译效率分析	Xtail	1.1.5	默认参数
差异表达基因	DESeq2	v1.34.0	fc >=2, p <=0.05
差异表达基因	edgeR	v3.36.0	fc >=2, p <=0.05
GO和KEGG	clusterProfiler	v4.2.2	默认参数

### □ 11.2 文件目录

报告的整体文件结构如下表，部分文件可能由于分析策略的不同和分析结果的差异，可能会没有。

### 结果目录结构

[全部展开](#) [全部折叠](#)

目录	类型	说明
index.html	文件	报告网页版入口
index.pdf	文件	报告PDF版
libs	目录	网页配置文件
sup	目录	报告结果, 图表存放目录
*_go	文件	GO分析结果, opposite代表反向变化, homo代表同向变化
*_kegg	目录	KEGG分析, opposite代表反向变化, homo代表同向变化

## 12 参考文献

- [1]. Wang, Q., Mao, Y. Principles, challenges, and advances in ribosome profiling: from bulk to low-input and single-cell analysis. *Adv. Biotechnol.* 1, 6 (2023).
- [2]. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009 Apr 10;324(5924):218-23.
- [3]. Ingolia NT, Hussmann JA, Weissman JS. Ribosome Profiling: Global Views of Translation. *Cold Spring Harb Perspect Biol*. 2019 May 1;11(5):a032698.
- [4]. Ingolia NT, Hussmann JA, Weissman JS. Ribosome Profiling: Global Views of Translation. *Cold Spring Harb Perspect Biol*. 2019 May 1;11(5):a032698.
- [5]. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884-i890 (2018).
- [6]. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884-i890 (2018).
- [7]. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
- [8]. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15-21.
- [9]. Xiao Z, Huang R, Xing X, Chen Y, Deng H, Yang X. De novo annotation and characterization of the translome with ribosome profiling data. *Nucleic Acids Res*. 2018 Jun 1;46(10):e61.

- [10]. Li F, Xing X, Xiao Z, Xu G, Yang X. RiboMiner: a toolset for mining multi-dimensional features of the translome with ribosome profiling data. BMC Bioinformatics. 2020 Aug 1;21(1):340.
- [11]. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923-930 (2014).
- [12]. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014).
- [13]. Xiao Z, Zou Q, Liu Y, Yang X. Genome-wide assessment of differential translations with ribosome profiling data. Nat Commun. 2016 Apr 4;7:11194.

联系电话: 400-027-0273

公司邮箱: [market@biorun.com](mailto:market@biorun.com)

公司地址: 湖北省武汉市东湖高新区高新大道888号高农生物园总部三期3栋伯远生物

网站备案号: Copyright © 2011-2024 Company Biorun, Inc.