

ChIP-seq 检测报告

2018-08-21

Pan

一、 样本信息

本次分析使用数据集为 NCBI SRA 编号 [SRP065184](#)。实验目的：在 T-ALL 细胞系 LOUCY 中产生针对 H3K27ac 的 ChIP 测序数据。

表 1-1 样本信息

SRA 编号	类型	上传时间	其他信息	检测平台
SRR2774675	H3K27ac ChIP-seq	2016-05-17	H3K27ac	Illumina
SRR2774676	input DNA	2016-05-17	input	Illumina

二、 数据质量统计

2.1 测序数据情况汇总

使用 fastp (S Chen,et al.) 进行统计。

表 2-1 测序情况汇总

SRA ID	Sample Name	Row reads	Raw bases	GC%	Q20%	Q30%
SRR2774675	H3K27ac	66830848	5613791232	45.18	94.76	92.70
SRR2774676	input	81856618	6875955912	38.62	94.80	92.72

注：

- (1) SRA ID: 数据在 SRA 数据库中的 ID;
- (2) Sample Name: 数据在实验中的命名;
- (3) Row reads: 原始数据 reads 数;
- (4) Raw bases: 原始数据的碱基数;
- (5) GC%: GC 碱基占总碱基数的比例;
- (6) Q20%: phred 得分达到 20 的碱基数的比例;
- (7) Q30%: phred 得分达到 30 的碱基数的比例。

2.2 测序质量分布图

使用 FastQC (Andrews S, et al.) 以及 MultiQC (Ewels P, et al.) 进行统计。



图 2-1 序列测序质量分布图

可以看出，大多数 reads 的 phred 得分在 20 以上。

2.3 碱基比例

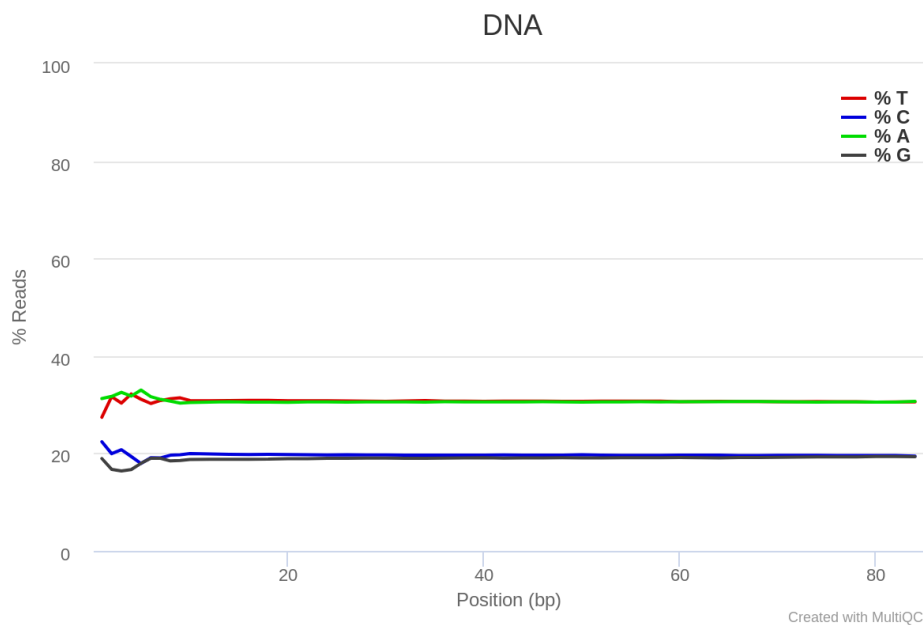


图 2-2 input DNA 各碱基比例图

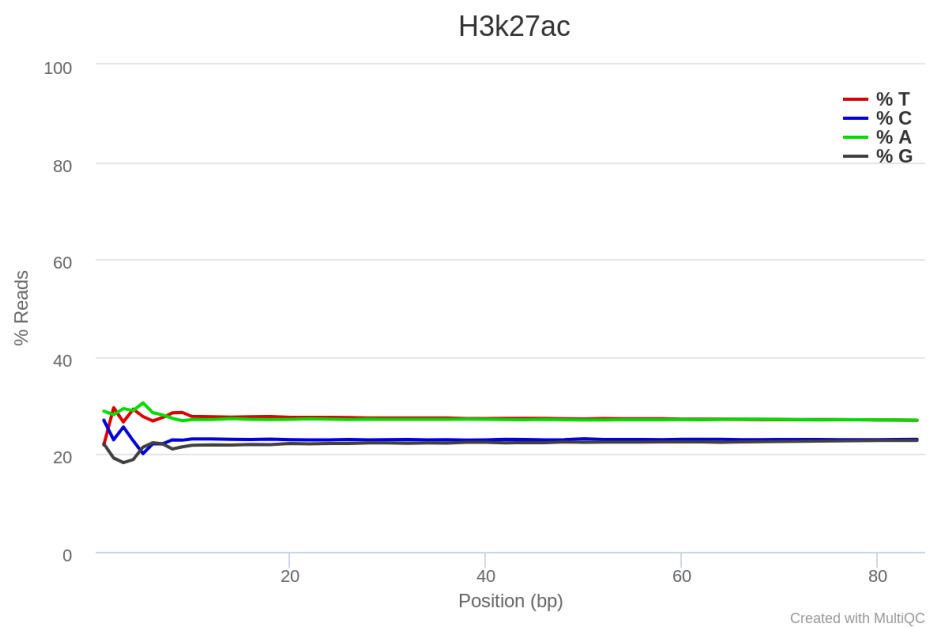


图 2-3 H3k27ac 各碱基比例图

三、ChIP 分析

使用 Bowtie2 (Langmead B and Salzberg SL) 将原始数据比对到参考基因组 (ucsc.hg19.fa) 后, 使用 MACS2 (Zhang Y, et al.) 进行 peak calling。其中 input 为 control 组, H3k27ac 为 treated 组。再使用 ChIPseeker (Yu G, et al.) 进行分析。

3.1 查看 peaks 在基因组中的位置

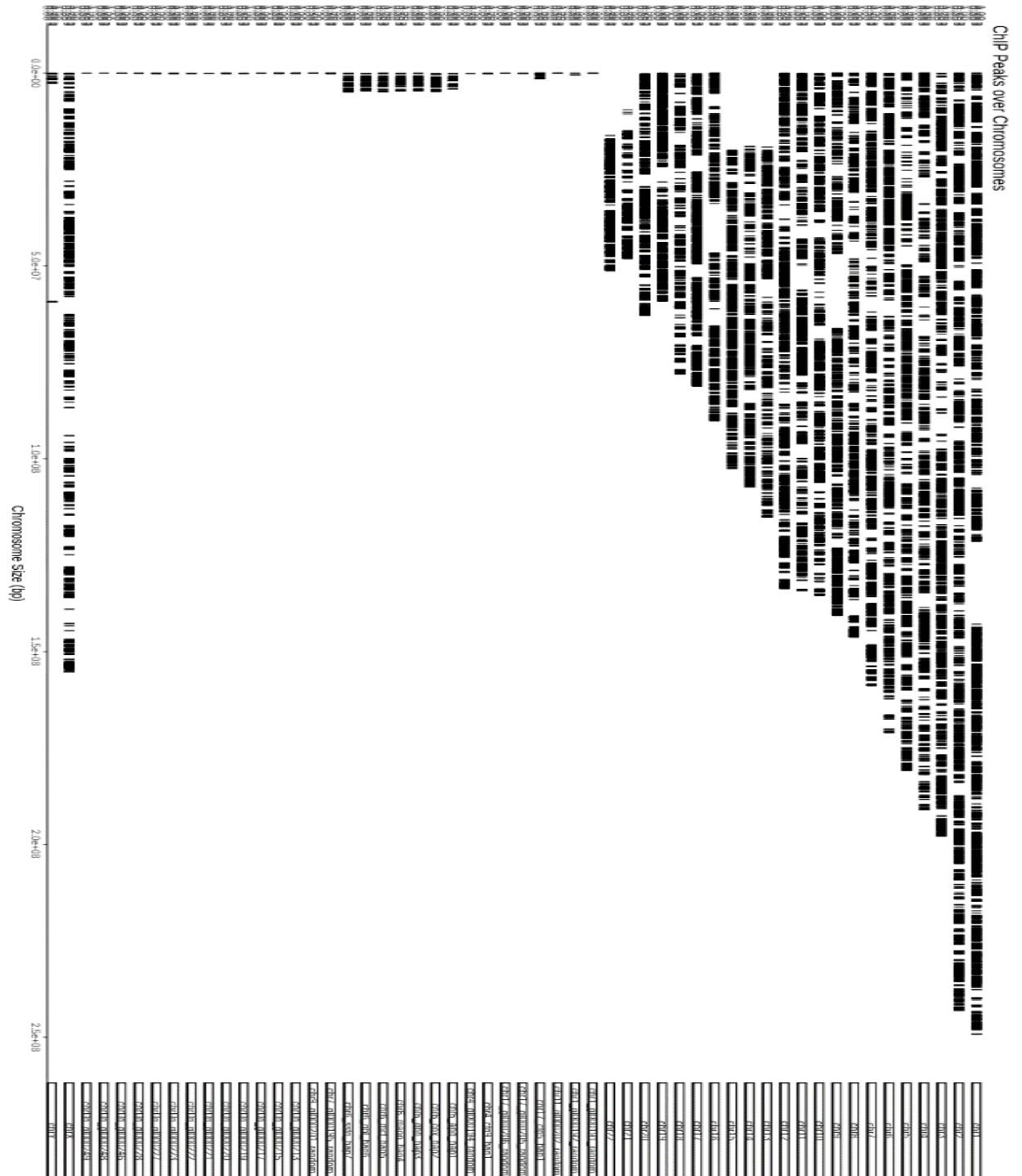


图 3-1 peaks 在基因组中的位置 (需横向查看)

3.2 与 TSS 区域结合的 peaks 的概况

TSS 为转录起始位点，下图为落在转录起始位点前后的 peaks 数目的统计图。

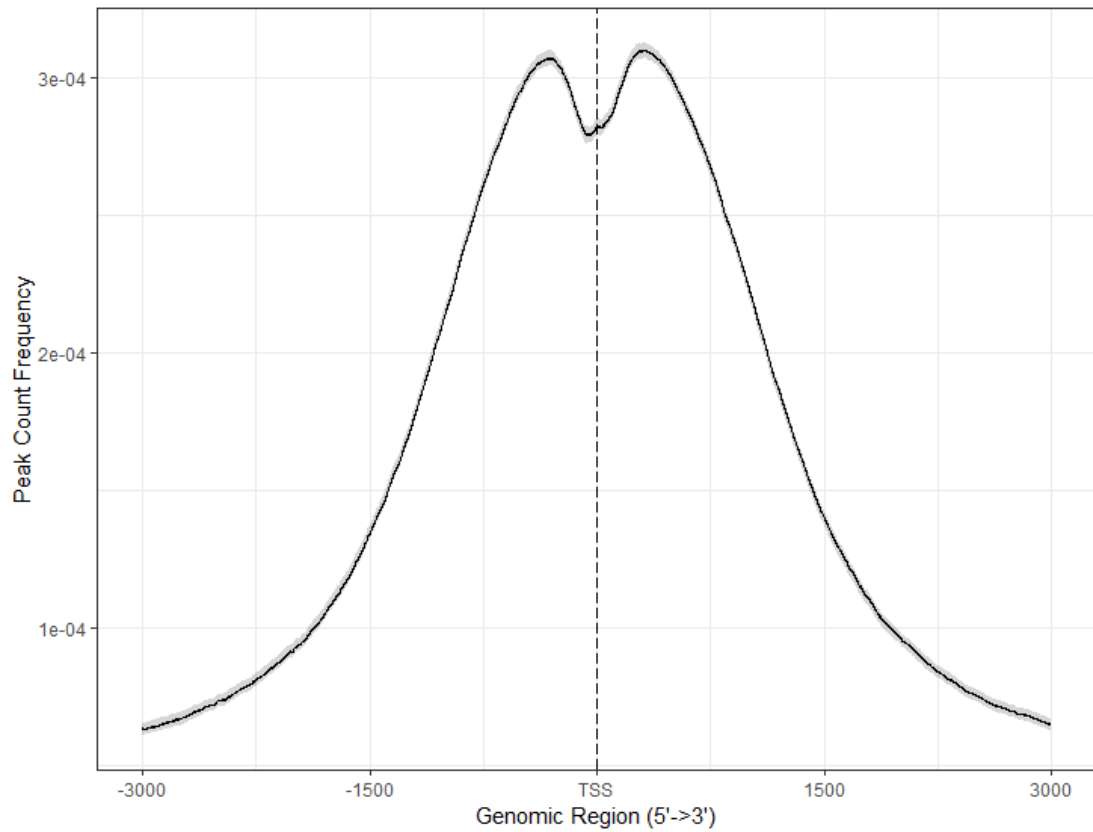


图 3-2 TSS 距离图

3.3 与 TSS 区域结合的 peaks 的热图

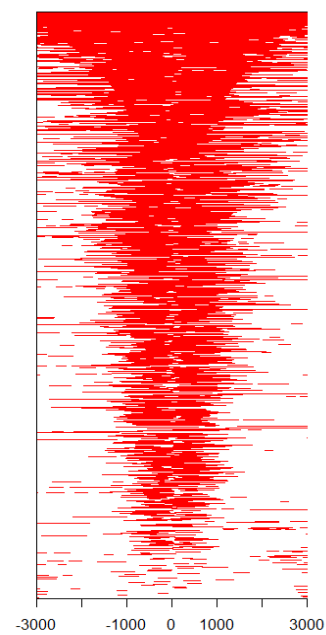


图 3-3 TSS 热图

四、 ChIP 注释

4.1 ChIP 注释结果

使用 ChIPseeker 进行 ChIP 注释。得到以下结果（只显示前 10 行。）

表 4-1 ChIP 注释结果

seqnames	start	end	width	V5	annotation	geneChr	geneStart	geneEnd	geneId	SYMBOL
chr1	20591	20958	368	18	Promoter (≤1kb)	1	16858	19759	653635	WASH7P
chr1	21067	22121	1055	90	Promoter (1-2kb)	1	16858	19759	653635	WASH7P
chr1	22220	22815	596	45	Promoter (2-3kb)	1	16858	19759	653635	WASH7P
chr1	28149	30442	2294	2044	Promoter (≤1kb)	1	14362	29370	653635	WASH7P
chr1	383656	384290	635	498	Distal Intergenic	1	367659	368597	729759	OR4F29
chr1	416732	417294	563	177	Distal Intergenic	1	367659	368597	729759	OR4F29
chr1	440422	443000	2579	1196	Distal Intergenic	1	367659	368597	729759	OR4F29
chr1	539841	542733	2893	1156	Distal Intergenic	1	621096	622034	729759	OR4F29
chr1	605225	606042	818	379	Distal Intergenic	1	621096	622034	729759	OR4F29
chr1	712917	715476	2560	2010	Promoter (≤1kb)	1	700245	714068	1E+08	LOC100288069

4.2 注释结果可视化

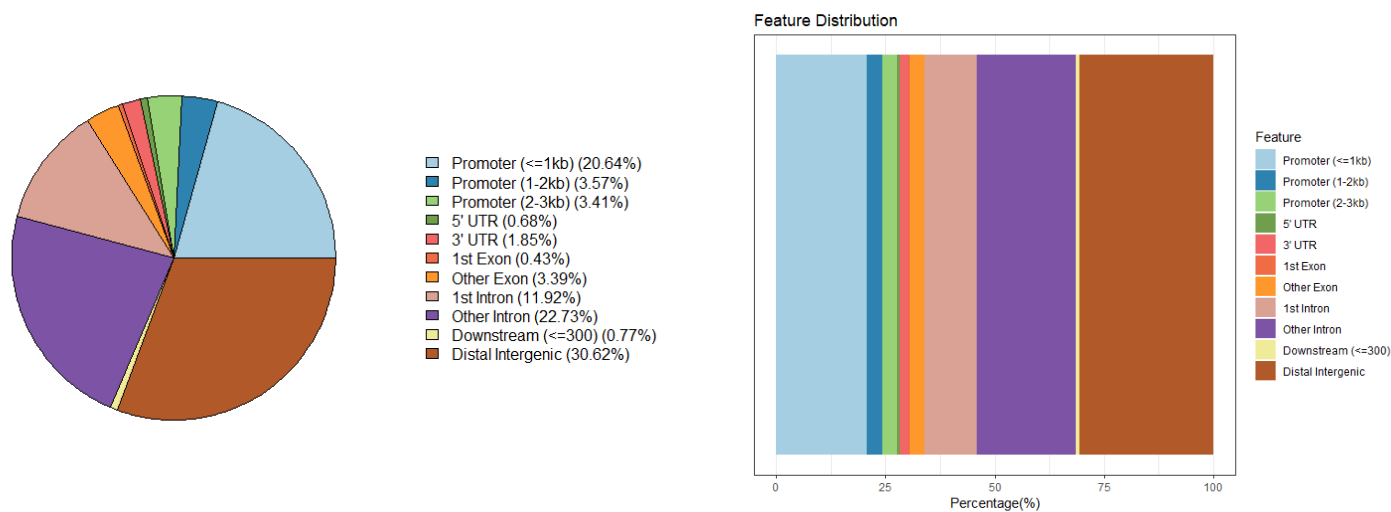


图 4-1 注释结果比例图

由于 UpSet 适用于多于 5 个集合的情况，而本次实验只有单个 ChIP 样本，因此下图仅供参考。

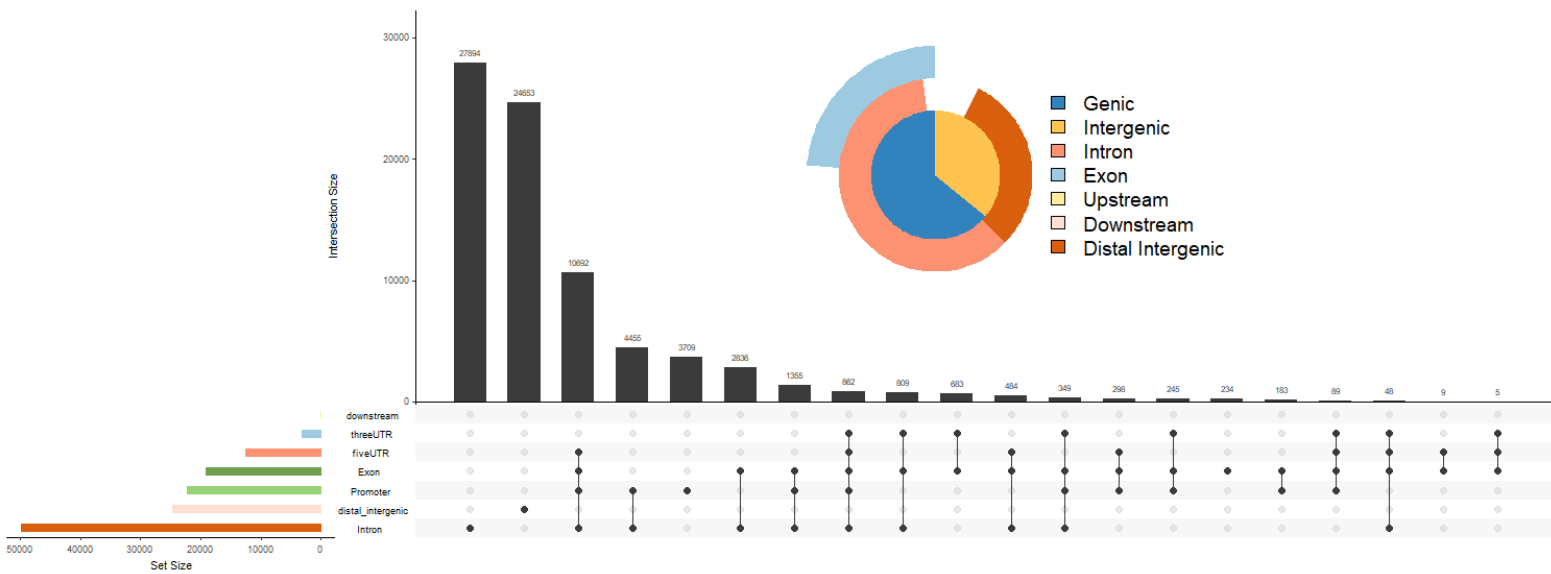


图 4-2 UpSet 图

4.3 富集分析

使用 clusterProfiler (Yu G , et al.) 对 peaks 所在基因进行富集分析，得到以下结果。

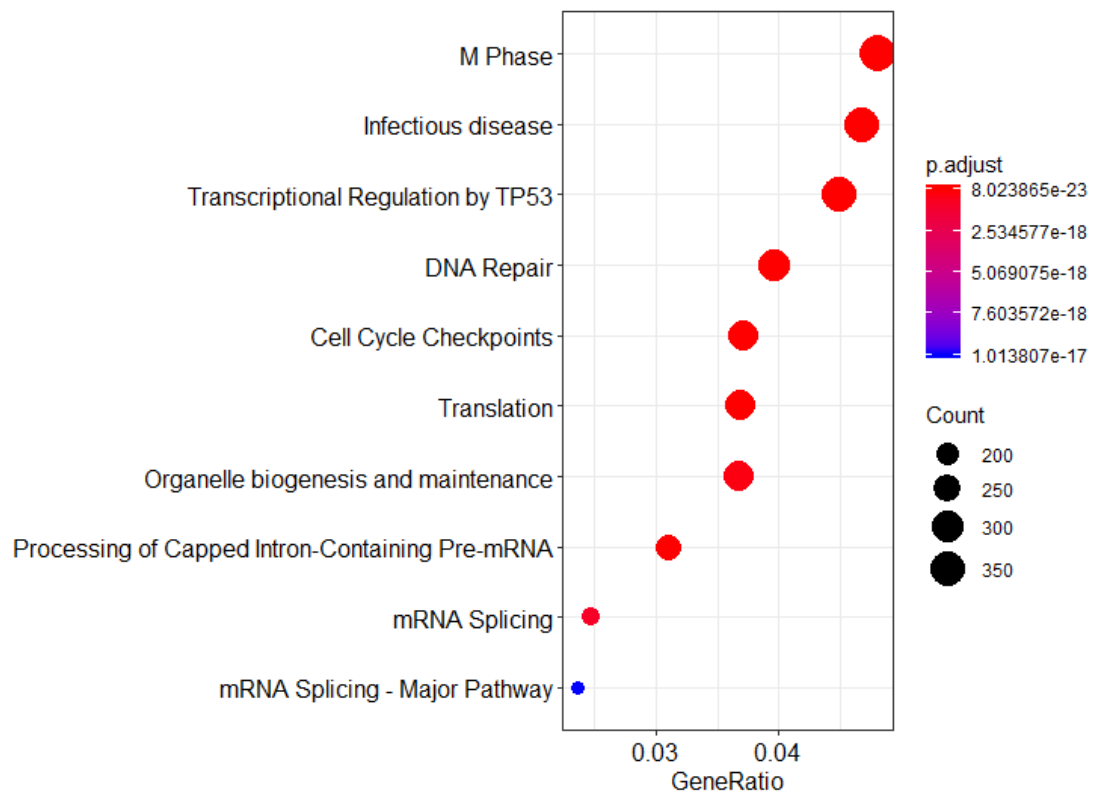


图 4-3 富集分析

五、 参考文献

1. Long noncoding RNA signatures define oncogenic subtypes in T-cell acute lymphoblastic leukemia. Wallaert A , et al. Leukemia. 2016 Sep;30(9):1927-30. doi: 10.1038/leu.2016.82. Epub 2016 Apr 22.
2. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. Leggett RM , et al. Front Genet. 2013 Dec 17;4:288. doi: 10.3389/fgene.2013.00288.
3. Quality control of next-generation sequencing data without a reference. Trivedi UH , et al. Front Genet. 2014 May 6;5:111. doi: 10.3389/fgene.2014.00111. eCollection 2014.
4. MultiQC: summarize analysis results for multiple tools and samples in a single report. Ewels P , et al. Bioinformatics. 2016 Oct 1;32(19):3047-8. doi: 10.1093/bioinformatics/btw354. Epub 2016 Jun 16.
5. Fast gapped-read alignment with Bowtie 2. Langmead B and Salzberg SL Nat Methods. 2012 Mar 4;9(4):357-9. doi: 10.1038/nmeth.1923.
6. Model-based analysis of ChIP-Seq (MACS). Zhang Y , et al. Genome Biol. 2008;9(9):R137. doi: 10.1186/gb-2008-9-9-r137. Epub 2008 Sep 17.
7. Identifying ChIP-seq enrichment using MACS. Feng J , et al. Nat Protoc. 2012 Sep;7(9):1728-40. doi: 10.1038/nprot.2012.101. Epub 2012 Aug 30.
8. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Yu G , et al. Bioinformatics. 2015 Jul 15;31(14):2382-3. doi: 10.1093/bioinformatics/btv145. Epub 2015 Mar 11.
9. clusterProfiler: an R package for comparing biological themes among gene clusters. Yu G , et al. OMICS. 2012 May;16(5):284-7. doi: 10.1089/omi.2011.0118. Epub 2012 Mar 28.