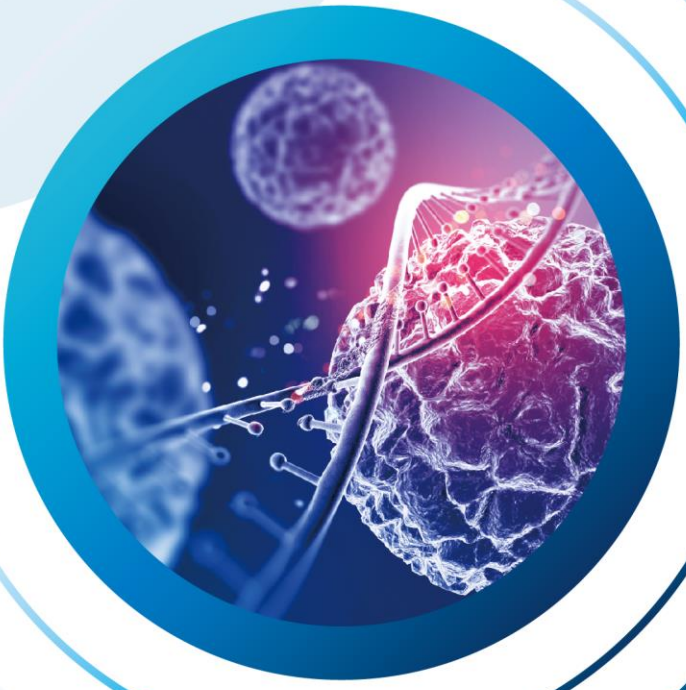
Bluescape Scientific

蓝景科信河北生物科技有限公司



DAP-seq项目分析报告

客户单位：北京农学院

联 系 人：陈雪雪

物　　种：森林草莓

样本名称：TCP7

www.bluescape.com.cn
info@bluescape.cc
河北省保定市惠阳街369号保定中关村创新基地12号楼

400-618-7099

目录

一、DAP-seq 技术简介 3
二、DAP-seq 和 ChIP-seq 的技术特点 3
三、本公司已做物种和转录因子 3
四、DAP-seq 实验技术流程 4
五、实验材料与方法 5
1. 蛋白表达载体构建 5
（1）蛋白表达载体构建试剂 5
（2）蛋白表达载体构建方法 5
2. 基因组 DNA 的提取 6
（1）基因组 DNA 提取试剂 6
（2）基因组 DNA 提取方法 6
3. 亲和纯化文库构建 6
（1）亲和纯化文库构建试剂 6
（2）亲和纯化文库构建方法 6
4. 体外无细胞蛋白表达与检测 6
（1）体外无细胞蛋白表达的主要试剂 6
（2）体外无细胞蛋白表达反应方法 6
（3）体外蛋白表达检测 7
5. 蛋白质与文库的亲和纯化 7
（1）亲和纯化所需要的主要试剂和设备 7
（2）亲和纯化方法 7
6. 洗脱文库扩增质检与测序 7
7. 生物信息学分析方法 7
六、实验结果 8
1. PCR 扩增获得目的 CDS 序列 8
2. 同源重组构建蛋白表达载体 8
3. 体外蛋白表达与检测 8
4. 基因组 DNA 提取 8
5. 基因组 DNA 片段化 8
6. 文库电泳结果 9
7. 亲和纯化后的文库电泳结果 9
8. 亲和纯化之后的文库进行测序 9
七、分析结果 9
1. Clean Data 数据统计 9
2. 参考序列比对结果 10
3. Callpeak 以及 peak 合并 10
4. Peaks 合并的韦恩图 12
5. Peak 在 TSS 附近的分布 12
6. Peak 在染色体上的分布 13
7. Peak 在基因组功能区的分布 13
8. Peak 相关基因的 Gene Ontology 富集分析 14
9. Peak 相关基因的 KEGG 富集分析 16

10. 富集区间 Motif 分析.....19
11. Motif 对应基因的查找.....20
参考文献.....20
与 DAP-seq 相关的技术服务.....21
染色质免疫沉淀测序（ChIP-Seq）.....21
凝胶阻滞或电泳迁移率实验（EMSA）.....21
酵母单杂交（Yeast One-Hybrid）.....21
关于蓝景科信河北生物科技有限公司.....22

一、DAP-seq 技术简介

在功能基因组学和表观遗传学研究中，发掘转录因子结合位点（Transcription Factor Binding Sites）一直是研究热点。传统的 ChIP-seq（染色质免疫共沉淀测序）方法，在抗体质量很好的情况下能够有效检测到转录因子结合位点。然而，要获得 ChIP-seq 级别的高质量抗体难度很大，所以高质量抗体的缺乏限制了 ChIP-seq 的大规模扩展。2016 年，Cell 期刊发表了使用 DNA 亲和纯化测序技术（DNA Affinity Purification Sequencing, DAP-seq）在基因组水平上，快速精准鉴定转录因子结合位点的论文（O'Malley et al., 2016）。2017 年，Nature Protocol 发表了 DAP-seq 的实验方法（Bartlett et al., 2017）。DAP-seq 技术的出现，使转录因子结合位点的研究不再局限于任何物种，不再受抗体质量的限制，进一步拓展和加深了生命科学和医学领域的研究广度和深度。

DAP-seq 是通过体外蛋白表达技术，表达出带有标签的转录因子，然后和基因组 DNA 文库在体外进行结合，进一步分离出所有与转录因子结合的 DNA，再使用高通量测序技术，找到转录因子的结合位点。

二、DAP-seq 和 ChIP-seq 的技术特点

表 1. DAP-seq 与 ChIP-seq 技术的特点对比

技术名称	DAP-seq	ChIP-seq
实验模式	体外	体内
是否需要目的蛋白特异性抗体	否	是
时间成本	低	高
是否高通量	是	否

三、本公司已做物种和转录因子

已经做过的材料包括：拟南芥，烟草，水稻，小麦，玉米，大豆，棉花，苜蓿，高粱，大麦草，百脉根，黄瓜，菜心，荔枝，香蕉，葡萄，苹果，柑橘，甜橙，桃，甜瓜，甘蔗，樱桃，芍药，枣，核桃，毛果杨，胡杨，油松，毛白杨，大青杨，白桦，光皮桦，欧洲云杉，柳树，麻疯树，金银花，丹参、木薯，地钱，腐霉，飞蝗等。

已经做过的转录因子家族：bHLH，BES1，GATA，ARF，ERF，bZIP，C2H2，GRAS，DREB，EIN3，WRKY，AP2，NAC，TCP，G2-like，HD-Zip，MADS，MYB，BZP，NFY，RAV，SPL，WOX，CAMTA，BZR，SBP，Co-like，Dof。

四、DAP-seq 实验技术流程

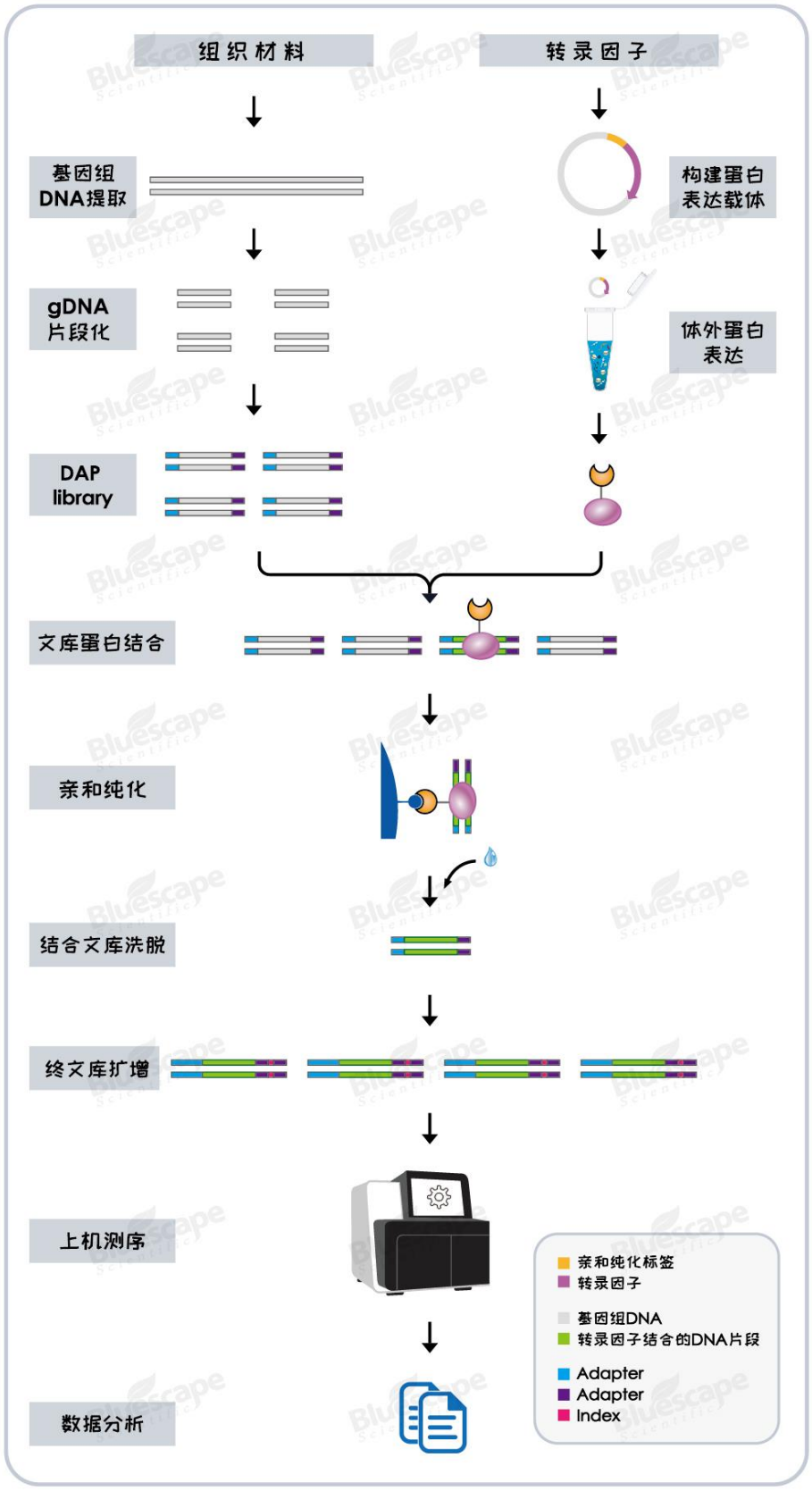


图 1. DAP-seq 技术流程概要

五、实验材料与方法

1. 蛋白表达载体构建

（1）蛋白表达载体构建试剂

高保真 PCR 预混液，琼脂糖凝胶回收试剂盒，同源重组克隆试剂盒，DH5α 感受态细胞，高纯度质粒小提试剂盒。

（2）蛋白表达载体构建方法

- 根据基因 CDS 序列，设计上游和下游引物并进行引物合成。
- 使用高纯度质粒小提试剂盒提取含有 CDS 序列的质粒模板，或者使用客户提供的质粒模板，对目的 CDS 进行 PCR 扩增。

表 2. 构建载体的 PCR 反应体系

PCR 组分名称	体积
PCR 预混液	10 μL
上游引物	0.5 μL
下游引物	0.5 μL
模板	X μL
总体积	加 ddH_2O 补齐至 20 μL

- PCR 反应程序

表 3. 构建载体的 PCR 反应程序

反应步骤	温度	时间	循环数
热盖	105 ℃		
预变性	94 ℃	2 min	1
变性	98 ℃	10 s	
退火	60 ℃	30 s	30
延伸	68 ℃	1 kb/min	
延伸	68 ℃	5 min	1
保存	4 ℃		

- PCR 产物回收

PCR 结束后，进行琼脂糖凝胶电泳。使用琼脂糖凝胶回收试剂盒，对目的条带进行胶回收。PCR 产物回收后，测定回收产物的浓度，并进行琼脂糖凝胶电泳检测。

- PCR 产物与蛋白表达载体的连接转化与鉴定

使用同源重组方法，将目的片段与蛋白表达载体进行连接。将连接产物转化到 DH5α 感受态细胞中，涂布于含有抗生素的 LB 培养基上，倒置培养。挑取单菌落进行菌落 PCR 鉴定，对扩增出目的条带的菌落进行质粒提取与测序。将测序结果与 CDS 进行序列比对。

- 蛋白表达载体的质粒提取

对构建正确的蛋白表达载体，进行质粒的提取，为后续蛋白表达做准备。

2. 基因组 DNA 的提取

（1）基因组 DNA 提取试剂

根据不同物种、组织和器官，使用相应的提取试剂或者试剂盒。例如：CTAB 提取试剂盒、植物通用型提取试剂盒、磁珠法提取试剂盒、含多糖多酚样本的基因组 DNA 提取试剂盒。

（2）基因组 DNA 提取方法

- 称取适量样本材料，液氮速冻，研磨成粉末，使用相应的试剂盒进行基因组 DNA 提取。
- 对提取的基因组 DNA 的质量进行检测，包括基因组 DNA 浓度，纯度。

3. 亲和纯化文库构建

（1）亲和纯化文库构建试剂

- DNA 建库试剂盒（MICH，NGS0602- MICH TLX DNA-Seq Kit）。
- DNA Clean Beads（MICH）

（2）亲和纯化文库构建方法

- 取适量基因组 DNA 进行片段化。
- 对片段化后的基因组 DNA 进行筛选。
- 取适量片段筛选后的基因组 DNA 进行亲和纯化文库的构建，具体方法详见 NGS0602-MICH TLX DNA-Seq Kit 建库试剂盒说明书。
- 文库构建结束后进行质检。

4. 体外无细胞蛋白表达与检测

（1）体外无细胞蛋白表达的主要试剂

- 无细胞蛋白表达试剂盒
- 用于进行 Western Blot 检测的一抗、二抗和化学发光试剂盒。

（2）体外无细胞蛋白表达反应方法

根据体外无细胞蛋白表达试剂盒的使用说明，建立以下反应体系，然后 25°C 孵育 2h。

表 4. 无细胞蛋白表达反应体系

反应组分	体积
无细胞蛋白表达预混液	30 μL
蛋白表达载体	4-10 μg
用 Nuclease-Free Water 补齐至 50 μL	

（3）体外蛋白表达检测

- 体外蛋白表达结束后，取适量反应液，进行 SDS PAGE 和 Western Blot，检测有无目标蛋白的特异性条带。

5. 蛋白质与文库的亲和纯化

（1）亲和纯化所需要的主要试剂和设备

- 亲和纯化磁珠
- 平衡缓冲液
- 磁力架（MICH）

（2）亲和纯化方法

- 将平衡好的磁珠与平衡缓冲液混匀。
- 将蛋白表达预混液与磁珠充分混匀后进行孵育。
- 向上述混合液中加入适量文库后充分混匀后进行孵育。
- 洗涤非特异性结合 DNA，洗脱特异性结合的 DNA。

6. 洗脱文库扩增质检与测序

7. 生物信息学分析方法

（1）对两个技术重复样本，以及 input 对照的 DAP-seq 文库，使用 NovaSeq，经过高通量测序，获得 PE150 bp 的 fq 格式的原始数据。

（2）使用 BWA-MEM（Vasimuddin et al., 2019），比对到物种基因组上。

（3）使用 MACS（Zhang et al., 2008）callpeak。

（4）使用 IDR 软件（Li et al., 2011），合并两个技术重复样本的 peaks，并且对重复 peak 的可靠性打分。

（5）使用 MEME 软件（Machanick & Bailey, 2011），分析 peaks 区域的典型 motif。

（6）使用 HOMER 软件（Heinz et al., 2010），对 peaks 进行注释。

（7）使用 depTools2 软件（Fidel et al., 2016），分析 reads 在 TSS 附近的分布频率，绘制密度曲线图和对应的热图。统计 peaks 在 TSS（Transcription Start Site）附近的分布频率，可视化 peaks 在染色体上的分布。

六、实验结果

1. PCR 扩增获得目的 CDS 序列

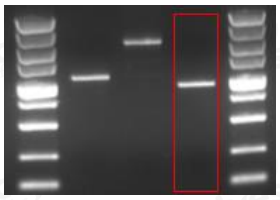


图 2. 以客户提供的质粒为模板，使用高保真 PCR 预混液进行 PCR 扩增，得到编码转录因子的 CDS 序列，琼脂糖凝胶电泳检测扩增特异性，并对目的条带进行回收。

2. 同源重组构建蛋白表达载体

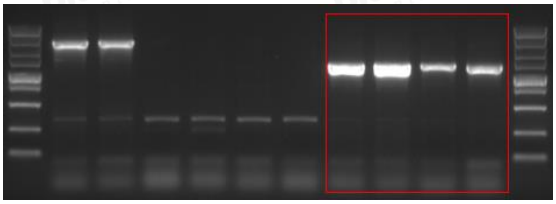


图 3. 将编码转录因子的 CDS 构建到蛋白表达载体上，转化感受态细胞后，进行菌落 PCR 鉴定阳性克隆。

3. 体外蛋白表达与检测

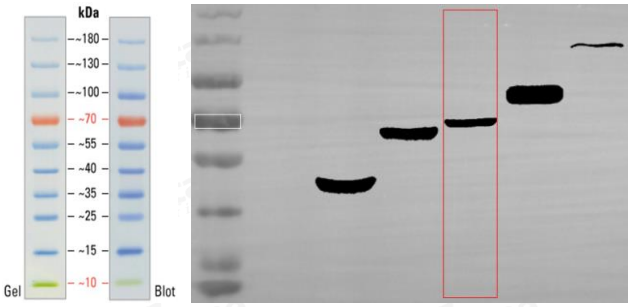


图 4. 将构建成功的蛋白表达载体置于体外蛋白表达反应液中，进行蛋白的表达。对表达产物进行 Western Blot 检测（图中白色框内的蛋白 marker 大小为 70kDa）。

4. 基因组 DNA 提取

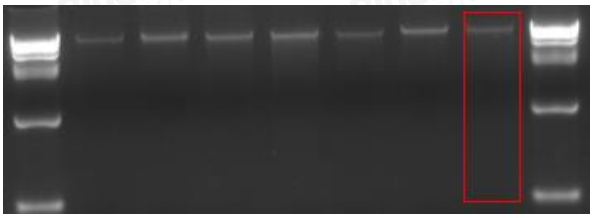


图 5. 提取物种的基因组 DNA，琼脂糖凝胶电泳检测提取质量。

5. 基因组 DNA 片段化

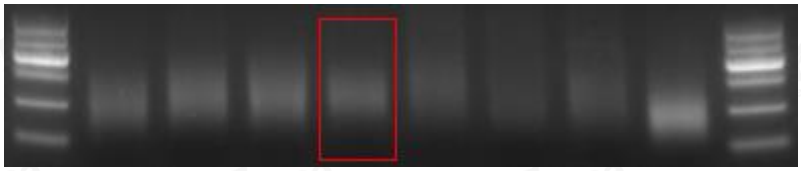


图 6. 将提取合格的基因组 DNA 进行打断处理，琼脂糖凝胶电泳检测打断质量。

6. 文库电泳结果

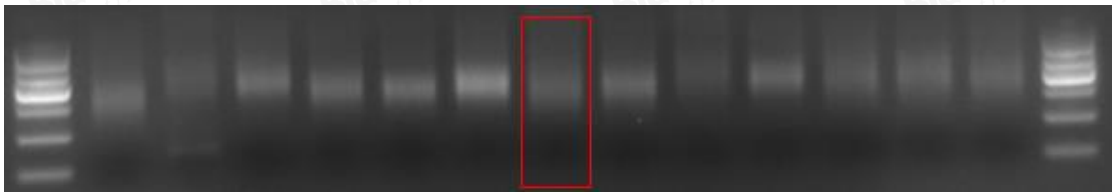


图 7. 对构建好的 DNA 亲和纯化文库进行电泳检测。

7. 亲和纯化后的文库电泳结果

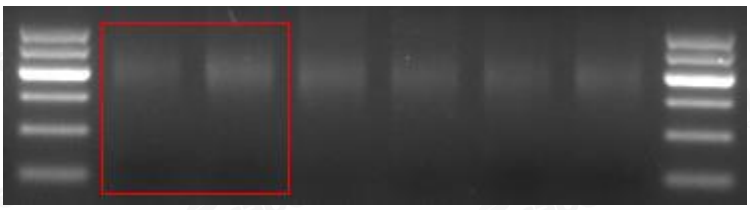


图 8. 蛋白与 DNA 文库进行亲和纯化之后，对文库进行电泳检测。

8. 亲和纯化之后的文库进行测序

测序仪器 NovaSeq，测序策略 PE150。

七、分析结果

1. Clean Data 数据统计

对 TCP7-1 和 TCP7-2 两个技术重复样本，以及对照 Input 的 DAP-seq 文库，使用 NovaSeq 测序，获得 PE150 bp 的 fq 格式的原始数据。使用 fastp 默认参数过滤原始数据，获得高质量的 reads，对 clean data 的质量信息进行统计，统计结果见下表。

表 5. Clean data 基本信息

samples	TCP7_1	TCP7_2	TCP7_input
raw reads	19.475234 M	20.364788 M	19.564894 M
raw bases	2.921285 G	3.054718 G	2.934734 G
raw bases Q20	2.823061 G (96.637626%)	2.947845 G (96.501368%)	2.842054 G (96.841951%)
raw bases Q30	2.677907 G (91.668792%)	2.794386 G (91.477695%)	2.694419 G (91.811347%)
clean reads	18.867706 M	19.712078 M	19.242822 M
clean bases	2.704368 G	2.816564 G	2.741882 G
clean bases Q20	2.636914 G (97.505736%)	2.745888 G (97.490694%)	2.671423 G (97.430269%)
clean bases Q30	2.509404 G (92.790763%)	2.612618 G (92.759053%)	2.540135 G (92.642048%)
filter ratio	0.968805	0.967949	0.983538

- raw reads：原始数据的总 Reads 数目。
- raw bases：原始数据的总总碱基数。
- raw bases Q20：原始数据中测序错误率小于 1%的碱基数目占总碱基数比例。
- raw bases Q30：原始数据中测序错误率小于 0.1%的碱基数目占总碱基数比例。

- clean reads：原始数据过滤后的总 reads 数目。
- clean bases：原始数据过滤后的总碱基数。
- clean bases Q20：原始数据过滤后测序错误率小于 1%的碱基数目占总碱基数比例。
- clean bases Q30：原始数据过滤后测序错误率小于 0.1%的碱基数目占总碱基数比例。
- filter ratio：数据过滤效率，Clean Data/Raw Data。

2. 参考序列比对结果

将 Clean reads 数据使用 BWA-MEM（Vasimuddin et al., 2019）软件，比对到物种基因组上，比对后的统计结果见下表。

表 6. 数据统计表

samples	total reads	reads mapped	reads mapped and paired	reads unpaired	map ratio
TCP7_1	18867706	18496436	18427792	68644	0.980322
TCP7_2	19712078	19324114	19254302	69812	0.980318
TCP7_input	19242822	18845373	18749610	95763	0.979346

- samples：样本名称。
- total reads：原始数据过滤后的总 reads 数。
- reads mapped：比对到参考基因组上的 reads 数。
- reads mapped and paired：双端比对到参考基因组上的 reads。
- reads unpaired：只有单端 reads 比对到参考基因组上的 reads 数目。
- map ratio：比对上 reads 与总 reads 的比值。

3. Callpeak 以及 peak 合并

使用 MACS（Zhang et al., 2008）软件得到 Peak，使用 IDR（Li et al., 2011）软件合并两个技术重复样本的 peaks，并且对重复 peak 的可靠性打分。使用 HOMER（Heinz et al., 2010）对 peaks 进行注释，全基因组功能性区域分为 Promoter, Exon, Intron, Downstream, Distal Intergenic, 5’ UTR 和 3’ UTR 区域。转录因子的结合位点，一般富集在基因转录起始位置附近，即 TSS（Transcription Terminate Site）区域，所以我们需要关注注释为 Promoter 的 Peak。找到距离 Peak 的 summit 位置（无 summit 则取中点位置）最近的转录起始位点对应的基因，即认为此 Peak 与该基因具有相互关联的特性。Peak 的详细结果见“4_Peak 分析/4.1_Peak 注释/TCP7_Peaks.anno.xls”。

表 7. TCP7_Peaks.anno.xls 表头注释

表头名称	表头含义
merge_peakname	两个重复合并后的 Peak 名称
Chr:Start-End	Peak 所在染色体：Peak 开始位置与结束位置
Annotation	Peak 所在位置的注释，定义转录起始位点-2k 到+500 为 Promoter 区域
Distance to TSS	Peak 距离 TSS（转录起始位点）的碱基数，－为 TSS 上游，+为 TSS 下游
transcriptid	距离 Peak 最近的 TSS 对应的转录本 ID
Venn	Peak 是否出现在两个技术重复中
TCP7_1_peakname	重复 1 的 Peak 名称
TCP7 -1_peak_Chr:Start-End	重复 1 的 Peak 所在染色体：Peak 的开始位置到结束位置
TCP7 -1_length	重复 1 的 Peak 的长度
TCP7 -1_abs_summit	重复 1 的 Peak 峰顶点在染色体上的位置
TCP7 -1_-log10 (P-value)	重复 1 P-value 值以 10 为底的负对数，数值越大，Peak 越可靠
TCP7 -1_fold_enrichment	重复 1 相对于 Input 的 Peak 峰高度的差异，数值越大，富集效果越明显
TCP7 -1_-log10 (Q-value)	重复 1 的 Q-value 值以 10 为底的负对数，数值越大，Peak 越可靠
TCP7 _2_peakname	重复 2 的 Peak 名称
TCP7 -2_chr:start-end	重复 2 的 Peak 所在染色体：Peak 的开始位置到结束位置
TCP7 -2_length	重复 2 的 Peak 的长度
TCP7 -2_abs_summit	重复 2 的 Peak 峰顶点在染色体上的位置
TCP7 -2_-log10 (P-value)	重复 2 的 P-value 值以 10 为底的负对数，数值越大，Peak 越可靠
TCP7 -2_fold_enrichment	重复 2 相对于 Input 的 Peak 峰高度的差异，数值越大，富集效果越明显
TCP7 -2_-log10 (Q-value)	重复 2 的 Q-value 值以 10 为底的负对数，数值越大，Peak 越可靠
Gene ID	距离 Peak 最近的 TSS 所对应的基因
Description	基因功能的描述

4. Peaks 合并的韦恩图

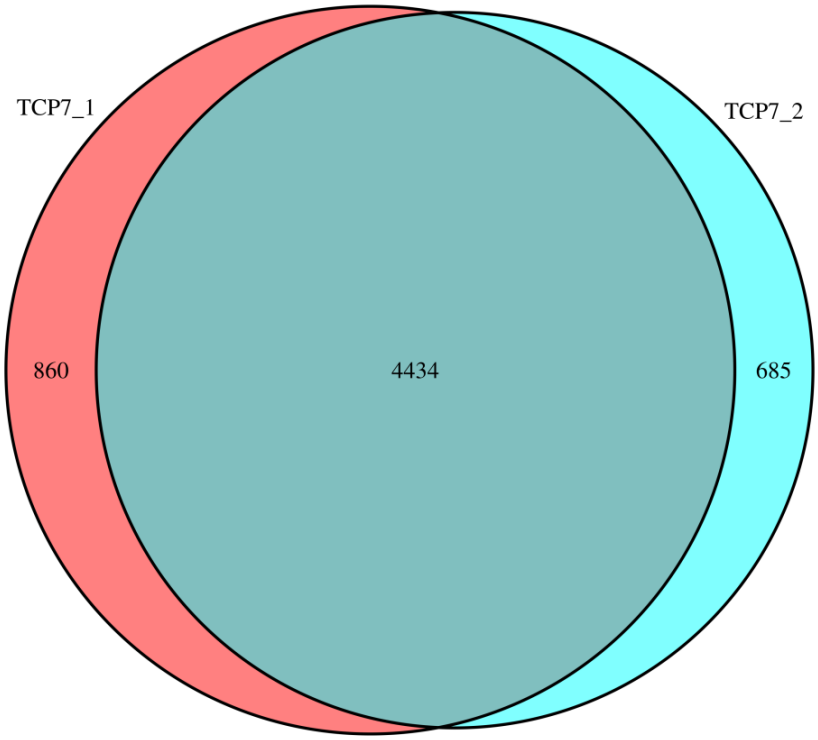


图 9. 分析两个技术重复 Peak 之间的一致性。

原图见“4_Peak 分析/ 4.2_Peak_venn”文件夹。

5. Peak 在 TSS 附近的分布

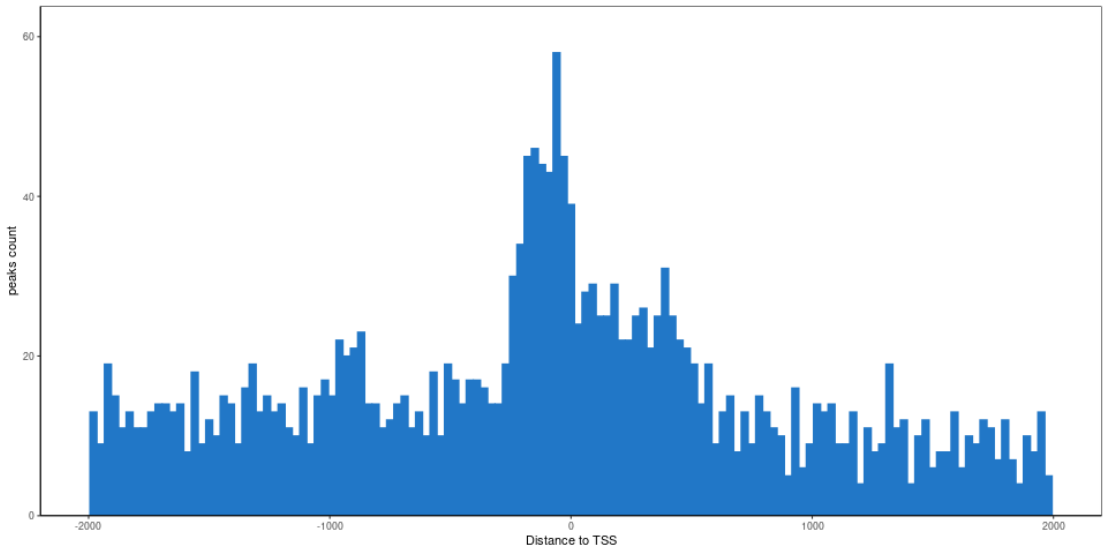


图 10. 分析富集的 Peak 在 TSS 附近的分布情况。

原图见“4_Peak 分析/ 4.3_Peak_plot”文件夹。

6. Peak 在染色体上的分布

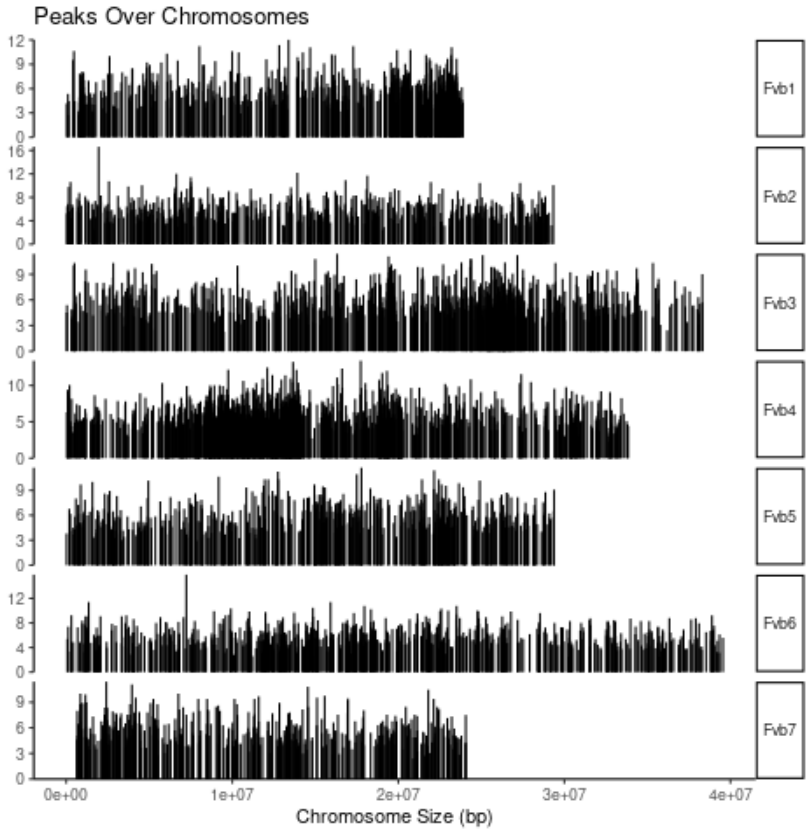


图 11. 分析 Peak 在不同染色体上的分布。

原图见“4_Peak 分析/ 4.3_Peak_plot”文件夹。

7. Peak 在基因组功能区的分布

全基因组功能性区域分为 Promoter, Exon, Intron, Downstream, Distal Intergenic, 5’ UTR 和 3’ UTR。对 Peak 进行注释后将注释的结果和 Peak 的分布进行统计。

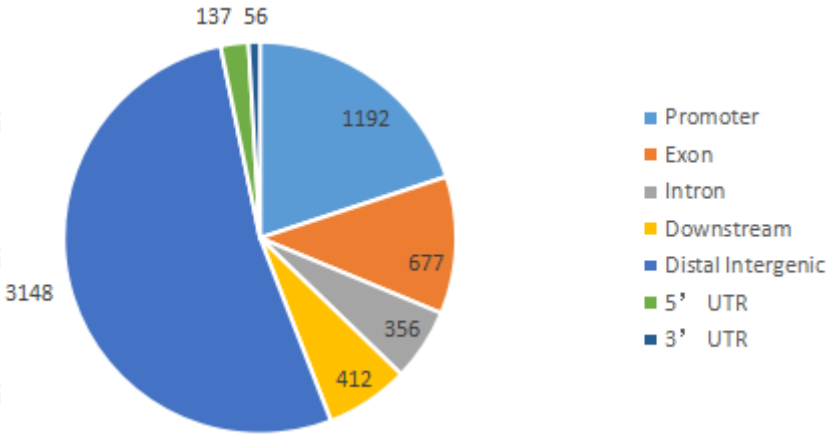


图 12. Peak 在基因功能区的注释统计。

原表格见“4_Peak 分析/ 4.4_Peak 在基因功能区的分布”文件夹。

8. Peak 相关基因的 Gene Ontology 富集分析

Gene Ontology 可分为分子功能（Molecular Function），生物过程（Biological Process）和细胞组成（Cellular Component）三个部分。它常用于提供基因功能分类标签和基因功能研究的背景知识。通过物种和基因信息，用 Gene Ontology 数据库进行查找，从而得到基因的 GO 注释信息（功能信息）。根据基因的 GO 注释，选择本物种的所有基因作为背景基因，使用统计方法计算 P 值，以 P<0.05 为显著性阈值分别得到相对于背景具有统计意义的高频率注释，从而得到基因集合在 GO 类别上的分布信息和显著性情况。

所有 Peak 相关基因的 GO 结果见文件:“5_GO 和 KEGG 分析/ gokegg_all/go_all.go.xls”。注释为 Promoter 的 Peak 相关基因的 GO 结果见“5_GO 和 KEGG 分析/ gokegg_promoter/go_promoter.go.xls”。

表 8. go_promoter.go.xls 表头注释

Term	靶基因富集到的 GO 类别名称
Class	分类：分子功能（Molecular Function），生物过程（Biological Process）和细胞组成（Cellular Component）
ID	在 GO 数据库中的通路 ID
Input number	该 GO 类别中靶基因数
Background number	该 GO 类别中背景基因数
P-Value	GO 类别富集的错误概率 P 值
fdr	矫正后的 P 值
Input gene	该 GO 类别中的靶基因名字

将所有 GO 富集结果，按照 P-Value 从小到大排序，取前 25 个 GO 富集功能绘制气泡图（图 13），原图见“5_GO 和 KEGG 分析/ gokegg_promoter”文件夹。

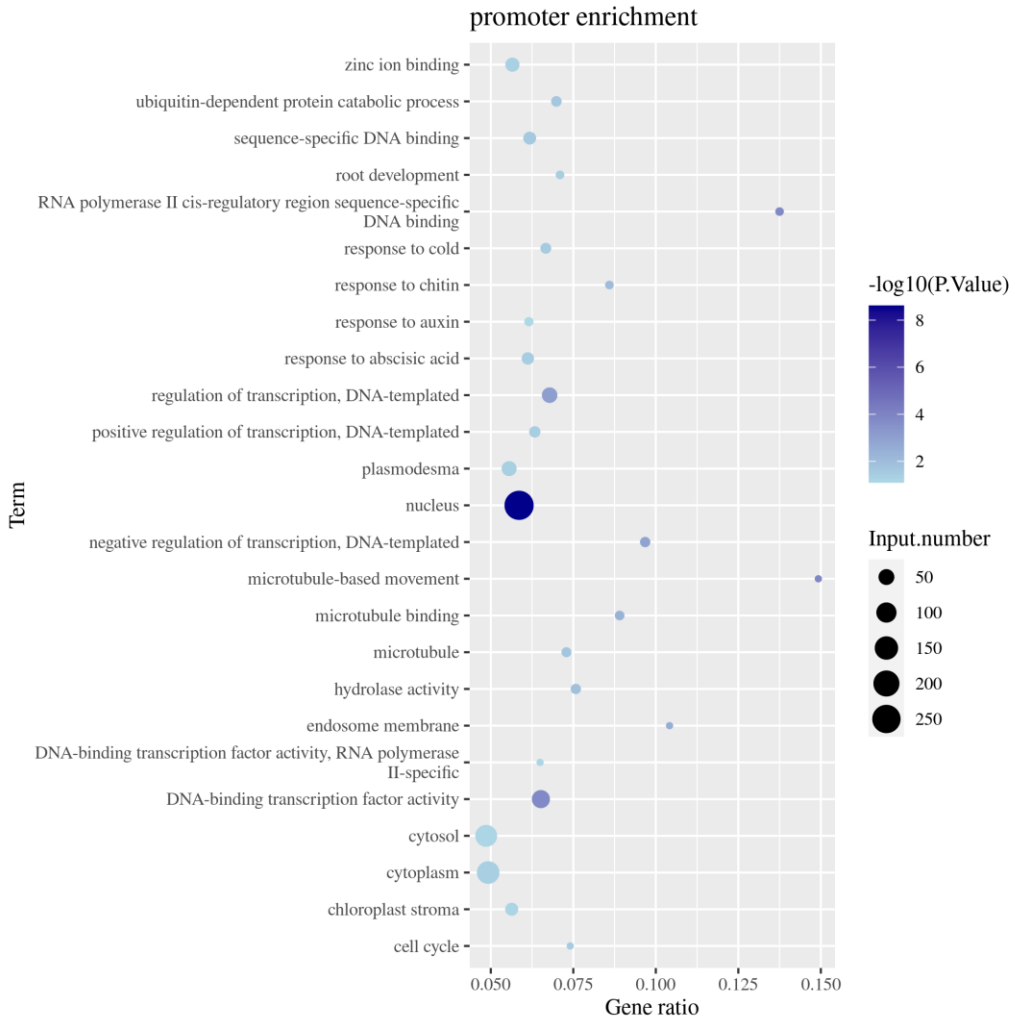


图 13. Promoter 对应的相关基因 Top25GO 富集的气泡图。

Term：靶基因富集到的 GO 类别名称。气泡颜色：－lg(P.Value)，数值越大颜色越深。气泡大小：Input number，即该 GO 类别中靶基因数。Gene ratio：Input number/ Background number，即该 GO 类别中靶基因数/该 GO 类别中背景基因数。

将所有 GO 富集结果，按照 P-Value 从小到大排序，取前 25 个 GO 富集功能绘制柱状图（图 14），原图见“5_GO 和 KEGG 分析/go_promoter”文件夹。

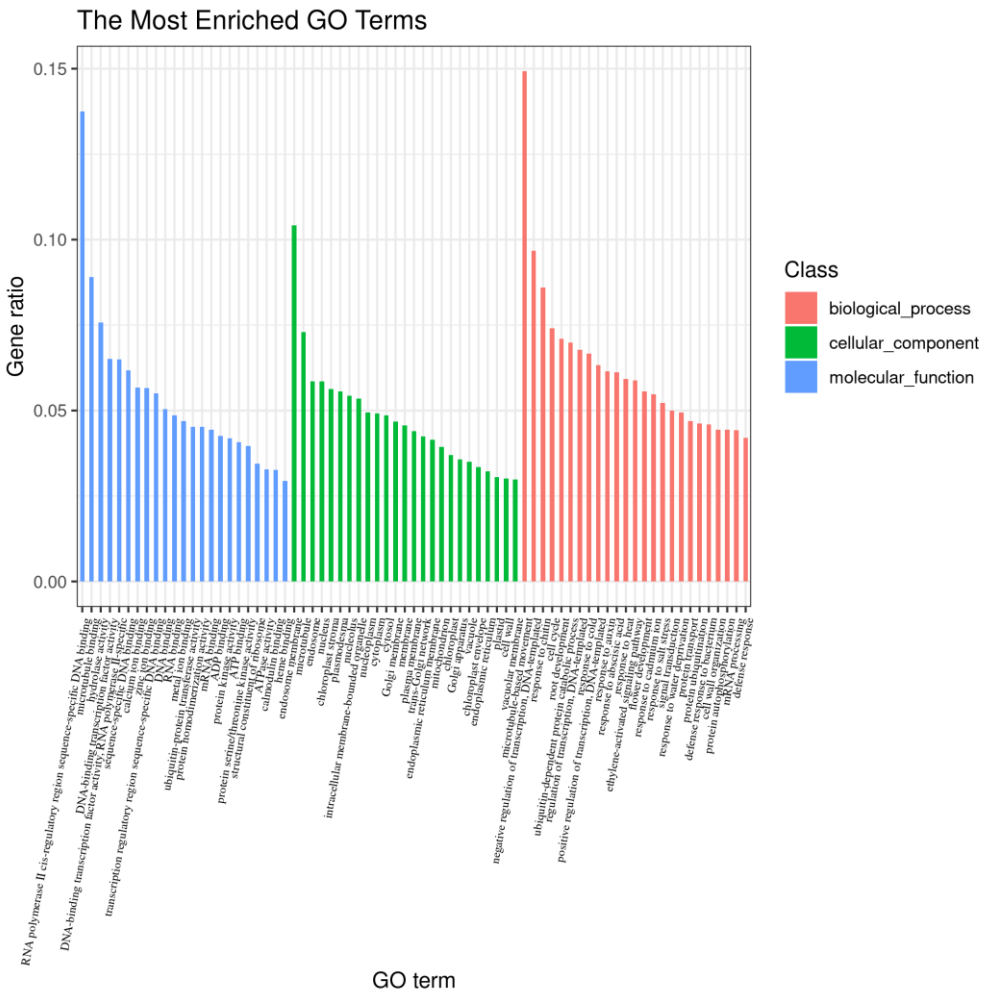


图 14. Peak 关联基因的 GO 富集柱状图。

横坐标：GO Term，靶基因富集到的 GO 类别名称。纵坐标：Gene ratio，Input number/ Background number，即该 GO 类别中靶基因数/该 GO 类别中背景基因数。Class：分子功能（Molecular Function），生物过程（Biological Process）和细胞组成（Cellular Component）。

9. Peak 相关基因的 KEGG 富集分析

Kyoto encyclopedia of genes and genomes（KEGG）（http://www.genome.jp/）是一个包含生化反应、信号通路、代谢通路和生物学过程的数据库。对差异表达基因集进行基于 KEGG 数据库的生物通路富集分析，提取出差异表达基因显著富集的通路（pathway），有利于下游实验的开展。根据基因的 KEGG 注释，选择本物种的所有基因作为背景基因，使用超几何分布计算 P 值，以 P<0.05 为显著性阈值分别得到相对于背景具有统计意义的高频率注释，从而得到基因在 KEGG Pathway 上的分布信息和富集情况。

所有 Peak 相关基因的 KEGG 结果见文件：“5_GO 和 KEGG 分析/kegg_all/TCP7.kegg_all.xls”。注释为 Promoter 的 Peak 相关基因的 KEGG 结果见“5_GO 和 KEGG

分析/kegg_promoter/ TCP7.kegg_promoter.xls”。

表 9. TCP7.kegg_promoter.xls 表头注释

Term	靶基因富集到的通路名称
ID	在 KEGG 数据库中的通路 ID
Input number	注释到该 Term 中的靶基因数
Background number	该物种中注释到该 Term 中的基因数
All annotated targert gene	靶基因中有 KEGG 注释信息的基因数
All annotated gene	该物种中有 KEGG 注释信息的基因数
P-Value	通路富集的错误概率 P 值
Corrected P-Value	矫正后的 P 值
Input	该通路中的靶基因名字，括号前面的名字为该靶基因组在该通路中显示的名称，对应的靶基因为括号里面的基因。
Hyperlink	KEGG 数据库中通路的超链接，可以在表格中直接打开或者复制到网页浏览器中打开，标红的为靶基因，绿色的为背景基因

将所有 KEGG 富集结果，按照 P-Value 从小到大排序，取前 20 个富集通路绘制气泡图（图 15），原图见“5_GO 和 KEGG 分析/gokegg_promoter”文件夹。

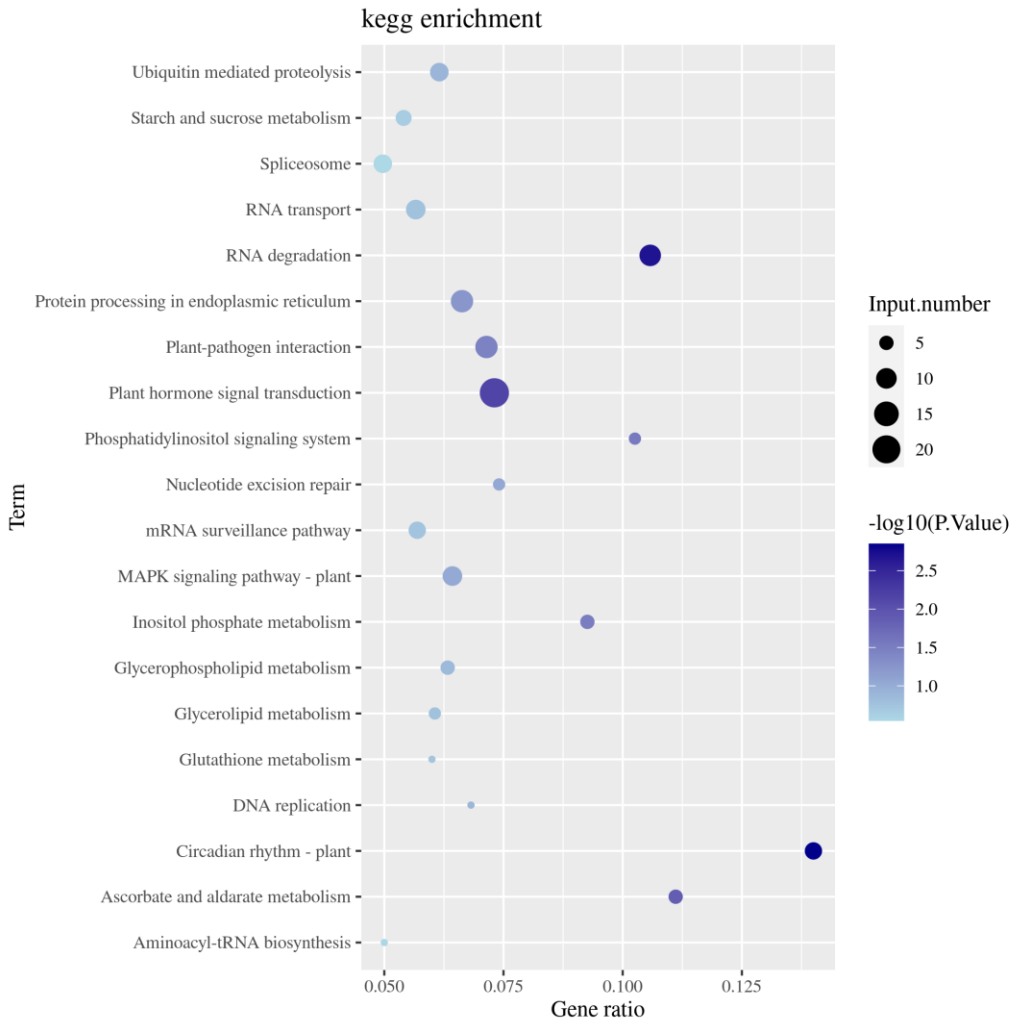


图 15. Promoter 相关基因 Top20 KEGG 富集的气泡图。

Term：靶基因富集到的通路名称；气泡颜色：－lg(P.Value)，数值越大颜色越深；气泡大小：Input number，即该通路中靶基因数目；Gene ratio：Input number/ Background number，即该通路中靶基因数目/该通路中背景基因数目。

将所有 KEGG 富集结果，按照 P-Value 从小到大排序，取前 60 个富集通路绘制柱状图（图 16），详细结果见“5_GO 和 KEGG 分析/gokegg_promoter”文件夹。

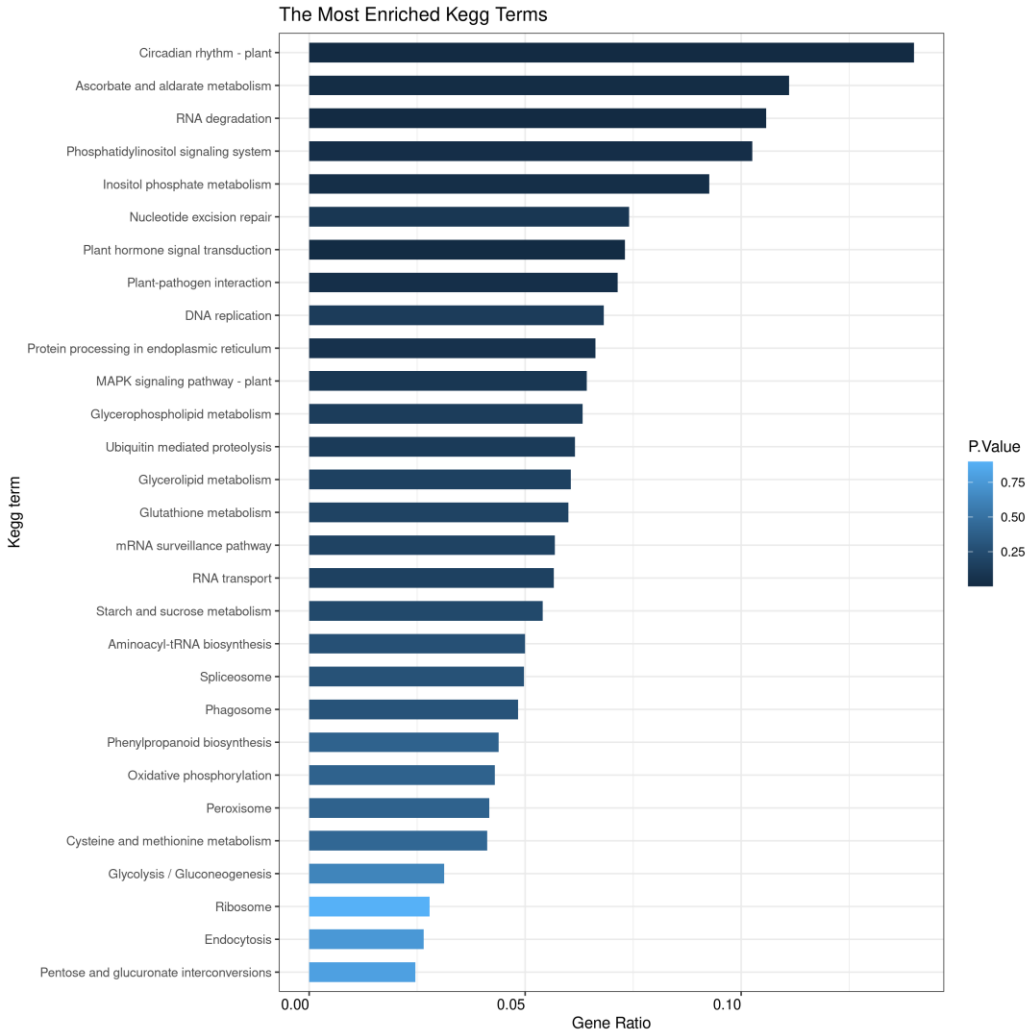


图 16. Peak 关联基因的 KEGG 富集柱状图。

横坐标：Gene ratio，Input number/ Background number，即该通路中靶基因数/该通路中背景基因数。纵坐标：KEGG Term，靶基因富集到的通路名称。

10. 富集区间 Motif 分析

Motif 是 DNA 与蛋白结合的重要桥梁，不同的 motif 具有不同结合特性与能力，每种蛋白通过不同的 motif 识别 DNA 结合位点。我们使用 MEME 软件（Machanick & Bailey, 2011）来提取 peak 所在区间的序列，对 peak 之间共有的 motif 进行扫描，查找其共有的 motif 区域，并绘制 motif 图。详细结果见“6_motif 分析/meme-chip_merge/meme-chip.html”。原图见“6_motif 分析/meme-chip_merge/meme-out”文件中。

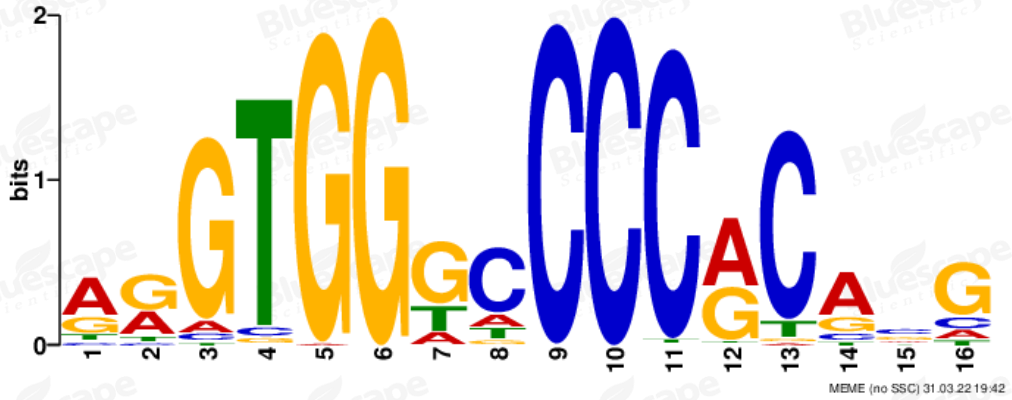


图 17. 对蛋白结合的 motif 序列进行分析，使用位置权重矩阵（Position Weight Matrix, PWM）的形式展示。

11. Motif 对应基因的查找

通过分析结果中的“4_Peak 分析/4.1_Peak 注释/TCP7_Peaks.anno.xls”文件，可以查看 Peak 的相关联基因，需要筛选 Annotation 为 Promoter（对 Promoter 的定义是 Peak 位于转录起始位点上游 2K 以内）的 Peak 来查看注释在启动子区的 Peak 所关联的基因。

通过“6_motif 分析/meme-chip_merge/meme-chip.html”文件，可以查看转录因子结合的 motif。

Motif 对应 Peak 和基因的查找可以查看“7_motif_in_mergepeaks”文件夹，该文件夹中的“7.1_Peak 序列/merge_peak.fa”文件是两个重复合并后的 Peak 的序列信息，文件夹中的“7.2_motif 对应的 Peak/merge/fimo_out.xls”文件是 motif 对应的 Peak 信息,“fimo_out.xls”和“6_motif 分析/meme-chip_merge/meme-chip.html”中的 motif 是一一对应的，在“fimo_out.xls”文件中可以查看 motif 对应的 Peak 和关联的基因。

参考文献

Bartlett A, O'Malley RC, Huang SC, et al., Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. 2017, 12: 1659–1672.

Fidel R, Devon PR, Björn G, et al., deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Research, 2016, 44: W160–W165.

Heinz S, Benner C, Spann N, et al., Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and b cell identities. Mol Cell, 2010, 38: 576–589.

Kidder BL, Hu GQ, Zhao KJ. ChIP-Seq: technical considerations for obtaining high-quality data.

Nature Immunology, 2011, 12: 918–922.

Vasimuddin Md, Sanchit Misra, Heng Li, Srinivas Aluru. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. IEEE Parallel and Distributed Processing Symposium (IPDPS), 2019.

Li QH, Brown JB, Huang HY, et al., Measuring reproducibility of high-throughput experiments. Annals of Applied Statistics, 2011, 5: 1752–1779.

Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics, 2011, 27: 1696–1697.

O'Malley RC, Huang SC, Song L, et al., Cistrome and epicistrome features shape the regulatory DNA landscape. Cell, 2016, 165: 1280–1292.

Zhang Y, Liu T, Meyer CA, et al., Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008, 9:R137.

与 DAP-seq 相关的技术服务

染色质免疫沉淀测序（ChIP-Seq）

在全基因组水平上检测与组蛋白、转录因子相互作用的 DNA 片段信息。

http://www.bluescape.cc/index.php/Index/jishu_details/id/11.html?pin_id=&fj_id1=12&fj_id2=&fj_id3=

凝胶阻滞或电泳迁移率实验（EMSA）

鉴定或验证 DNA 结合蛋白和 DNA 序列的相互作用。

http://www.bluescape.cc/index.php/Index/jishu_details/id/12.html?pin_id=&fj_id1=11&fj_id2=&fj_id3=

酵母单杂交（Yeast One-Hybrid）

确定已知 DNA 和蛋白质之间是否存在相互作用。

筛选结合于目标顺式作用元件或其他短 DNA 序列的新蛋白。

鉴定蛋白结合 DNA 的结构域，精准定位与 DNA 结合的蛋白序列。

http://www.bluescape.cc/index.php/Index/jishu_details/id/13.html?pin_id=&fj_id1=16&fj_id2=&fj_id3=

关于蓝景科信河北生物科技有限公司

本公司提供 DAP-seq 全流程技术服务与个性化数据分析，与中国科学院、中国农业科学院、中国林业科学研究院、浙江大学、中国农业大学、华中农业大学、北京林业大学、河北农业大学等多个科研院所与高校合作，具有丰富的技术服务经验。已经做过的材料包括：拟南芥，烟草，水稻，小麦，玉米，大豆，棉花，苜蓿，高粱，大麦草，百脉根，黄瓜，菜心，荔枝，香蕉，葡萄，苹果，柑橘，甜橙，桃，甜瓜，甘蔗，樱桃，芍药，枣，核桃，毛果杨，胡杨，油松，毛白杨，大青杨，白桦，光皮桦，欧洲云杉，柳树，麻疯树，金银花，丹参、木薯，地钱，腐霉，飞蝗等。

电话：400-618-7099

QQ：2522026793

地址：河北省保定市惠阳街 369 号保定中关村创新基地 12 号楼

官网：www.bluescape.com.cn

邮箱：info@bluescape.cc