RNA-Seq简单分析流程及网络资源介绍

组别：G10
成员：吴小园 尚丽 陈星 何兴波
日期：2013-06-21

主要内容

- 简要介绍
- 分析流程及网络资源介绍
- 小结

RNA-seq简介

转录组（Transcriptome）

广义：所有 RNA 的总和，mRNA、rRNA、tRNA及ncRNA；

狭义：细胞所能转录出的所有 mRNA 。

基因组层面：碱基数量和组成基本上是不变的，遗传信息也是相对稳定的；

转录组层面：可随着外部环境条件的变化而有所改变。转录组学或转录组研究，则是在某一个或某一种特定的细胞群在特定的环境条件和发育阶段上所有转录物的变化情况。

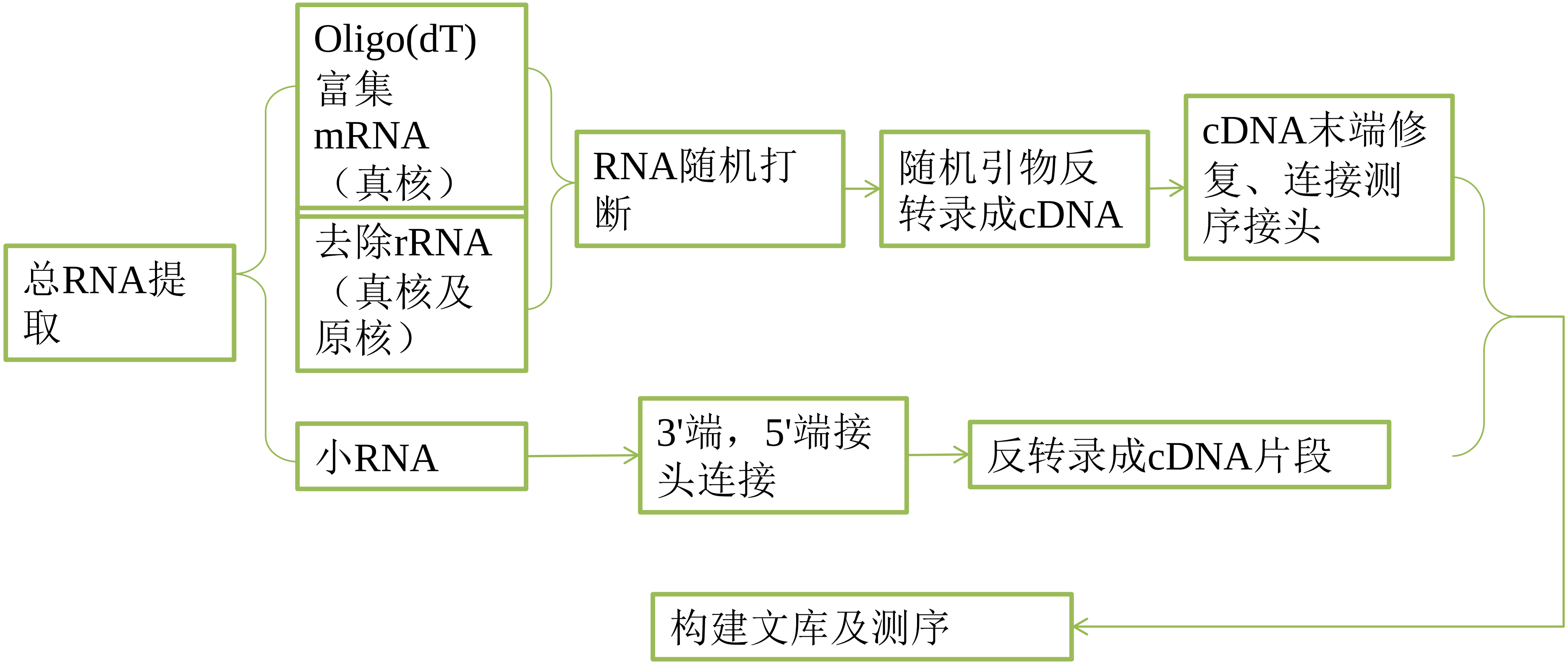
从RNA 水平上研究基因表达变化的方法：

如半定量PCR，定量 PCR，基因芯片

RNA-Seq

是基于新一代测序技术的转录组学研究方法，可以形成对该生物样品当前发育状态的基因表达状况的全局了解。将不同阶段或部位的生物样品的 RNA-Seq转录组进行比较，则可以在转录层面得到基因表达水平的变化——这就形成了表达谱（Expression Pattern），针对关键基因则可以构建代谢通路以及相互作用网络。

RNA-seq技术路线



总RNA提取

Oligo(dT)富集mRNA（真核）

去除rRNA（真核及原核）

RNA随机打断

随机引物反转录成cDNA

cDNA末端修复、连接测序接头

小RNA

3'端，5'端接头连接

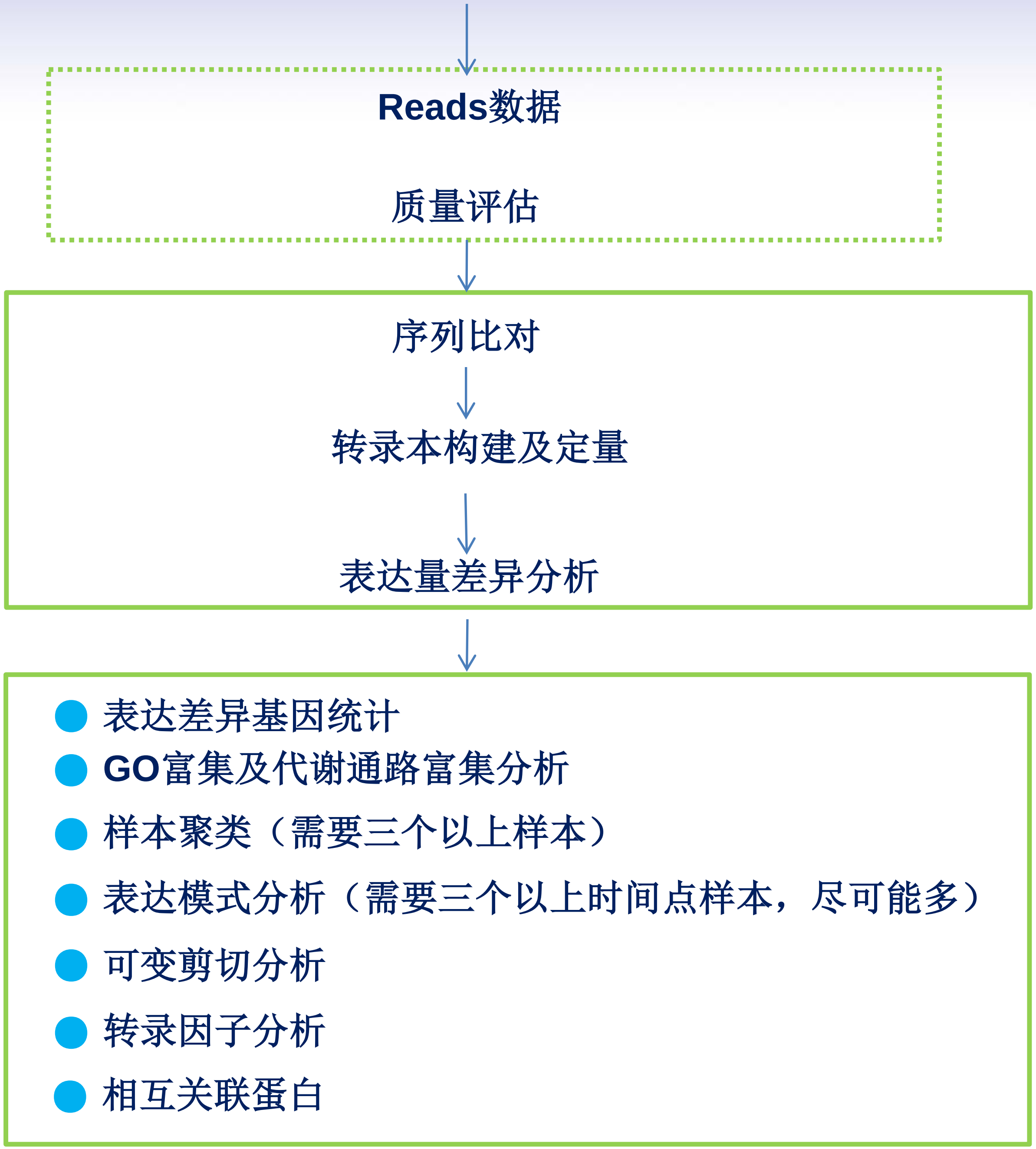
反转录成cDNA片段

构建文库及测序

原始数据的质量控制：去除接头、低质量reads

http://www.ebioservice.com/show_product.asp?id=471

分析流程



Reads数据

质量评估

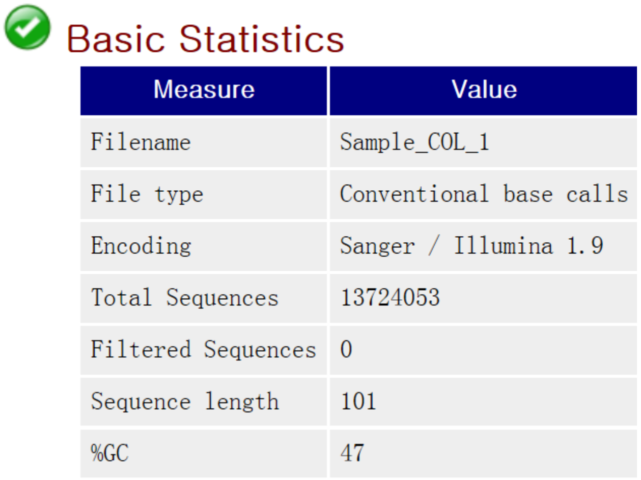
序列比对

转录本构建及定量

表达量差异分析

- 表达差异基因统计
- GO富集及代谢通路富集分析
- 样本聚类（需要三个以上样本）
- 表达模式分析（需要三个以上时间点样本，尽可能多）
- 可变剪切分析
- 转录因子分析
- 相互关联蛋白

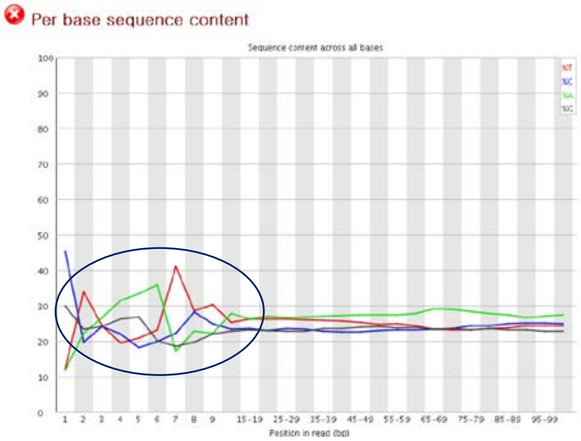
质量评估



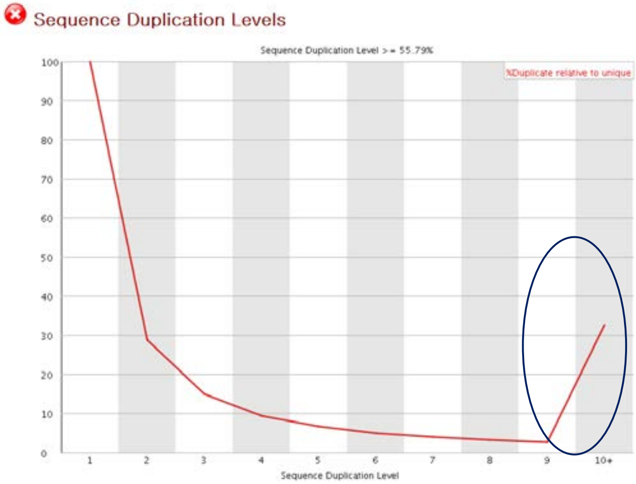
✓ **Basic Statistics**

Measure	Value
Filename	Sample_COL_1
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	13724053
Filtered Sequences	0
Sequence length	101
%GC	47

✓ Per sequence quality scores



✗ Per base sequence content



✗ Sequence Duplication Levels

命令行：`fastqc sampl1.fa sample2.fa …`

RNA-seq分析软件

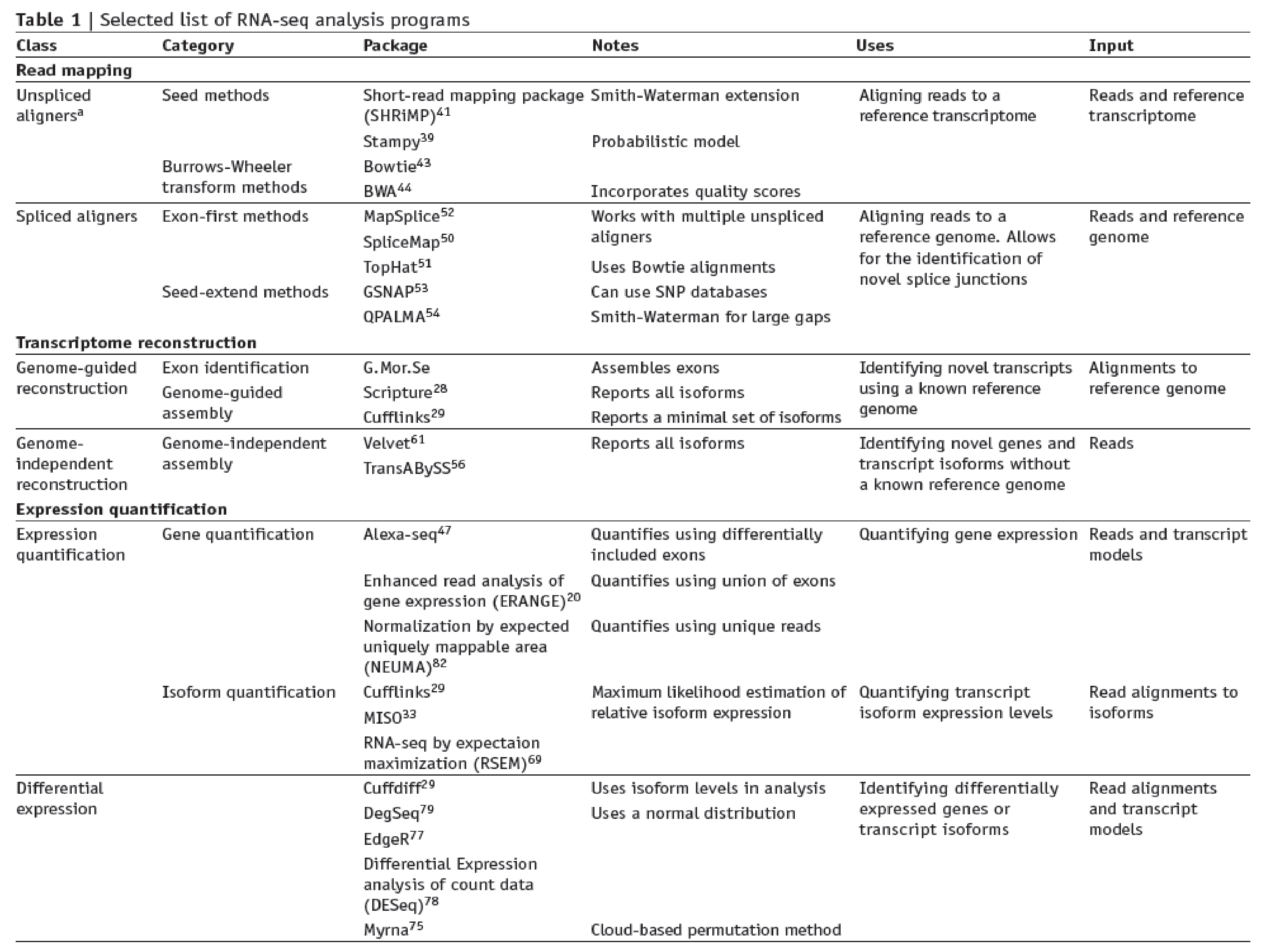


Table 1 | Selected list of RNA-seq analysis programs

Class	Category	Package	Notes	Uses	Input
Read mapping					
Unspliced aligners[a]	Seed methods	Short-read mapping package (SHRiMP)[41]	Smith-Waterman extension	Aligning reads to a reference transcriptome	Reads and reference transcriptome
		Stampy[39]	Probabilistic model		
	Burrows-Wheeler transform methods	Bowtie[43]			
		BWA[44]	Incorporates quality scores		
Spliced aligners	Exon-first methods	MapSplice[52]	Works with multiple unspliced aligners	Aligning reads to a reference genome. Allows for the identification of novel splice junctions	Reads and reference genome
		SpliceMap[50]			
		TopHat[51]	Uses Bowtie alignments		
	Seed-extend methods	GSNAP[53]	Can use SNP databases		
		QPALMA[54]	Smith-Waterman for large gaps		
Transcriptome reconstruction					
Genome-guided reconstruction	Exon identification	G.Mor.Se	Assembles exons	Identifying novel transcripts using a known reference genome	Alignments to reference genome
	Genome-guided assembly	Scripture[28]	Reports all isoforms		
		Cufflinks[29]	Reports a minimal set of isoforms		
Genome-independent reconstruction	Genome-independent assembly	Velvet[61]	Reports all isoforms	Identifying novel genes and transcript isoforms without a known reference genome	Reads
		TransABySS[56]			
Expression quantification					
Expression quantification	Gene quantification	Alexa-seq[47]	Quantifies using differentially included exons	Quantifying gene expression	Reads and transcript models
		Enhanced read analysis of gene expression (ERANGE)[20]	Quantifies using union of exons		
		Normalization by expected uniquely mappable area (NEUMA)[82]	Quantifies using unique reads		
	Isoform quantification	Cufflinks[29]	Maximum likelihood estimation of relative isoform expression	Quantifying transcript isoform expression levels	Read alignments to isoforms
		MISO[33]			
		RNA-seq by expectaion maximization (RSEM)[69]			
Differential expression		Cuffdiff[29]	Uses isoform levels in analysis	Identifying differentially expressed genes or transcript isoforms	Read alignments and transcript models
		DegSeq[79]	Uses a normal distribution		
		EdgeR[77]			
		Differential Expression analysis of count data (DESeq)[78]			
		Myrna[75]	Cloud-based permutation method		

Garber *et al*. (2011) *Nat. Meth.*

选择的流程：
Tophat----Cufflinks----Cuffdiff

Step 1：Tophat 将reads比对到基因组。

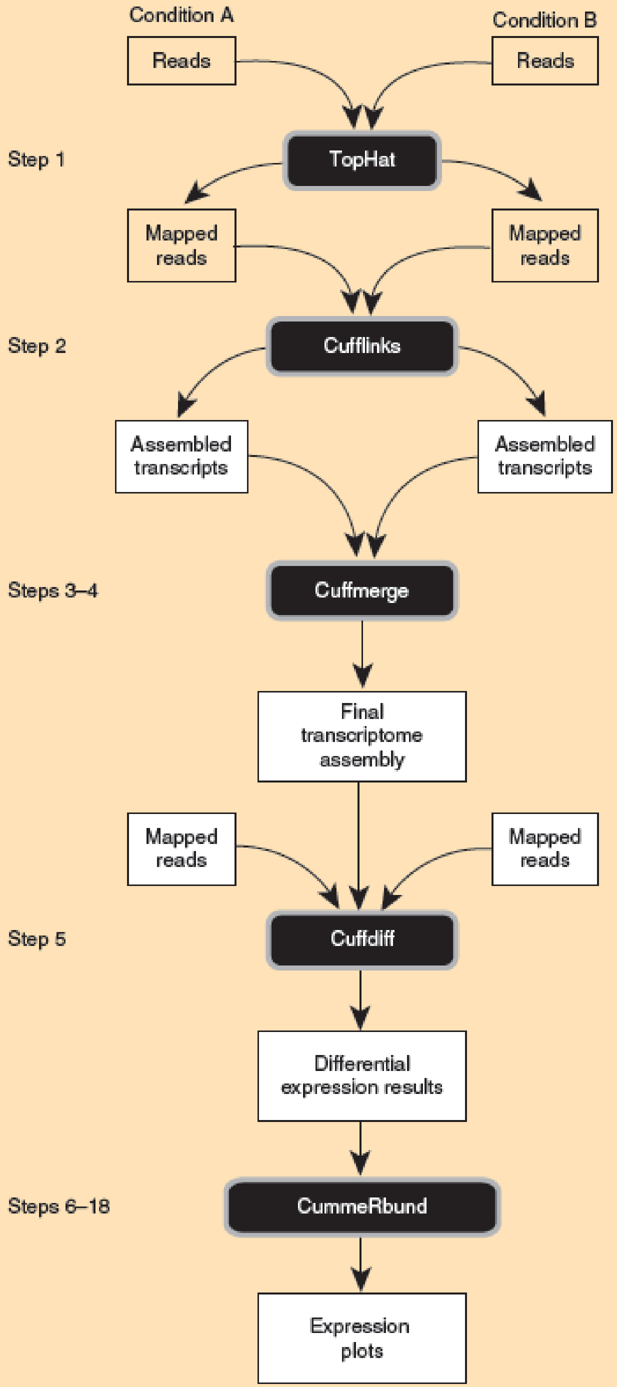
Step 2：Cufflinks对每个样本组装成一个转录本。

Step 3-4：Cuffmerge将所有转录本合并成一个新的转录本。

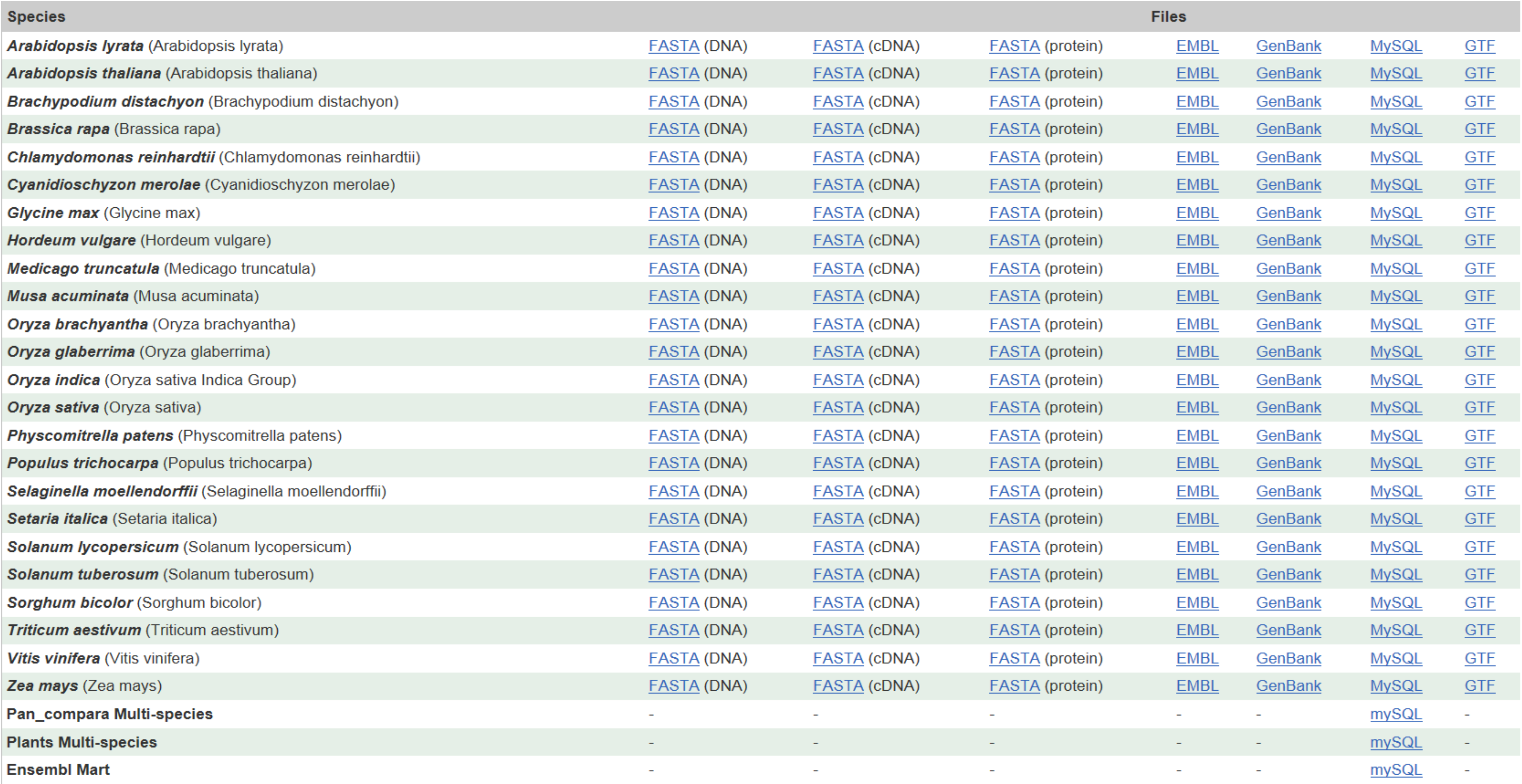
Step 5：Cuffdiff筛选差异表达基因

Step 6-18：CummeRbund图形展示

Trapnell *et al*. (2012). *Nat. Protocol*



下载基因组序列、基因组序列索引以及基因注释文件



Species	Files						
Arabidopsis lyrata (Arabidopsis lyrata)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Arabidopsis thaliana (Arabidopsis thaliana)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Brachypodium distachyon (Brachypodium distachyon)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Brassica rapa (Brassica rapa)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Chlamydomonas reinhardtii (Chlamydomonas reinhardtii)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Cyanidioschyzon merolae (Cyanidioschyzon merolae)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Glycine max (Glycine max)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Hordeum vulgare (Hordeum vulgare)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Medicago truncatula (Medicago truncatula)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Musa acuminata (Musa acuminata)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Oryza brachyantha (Oryza brachyantha)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Oryza glaberrima (Oryza glaberrima)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Oryza indica (Oryza sativa Indica Group)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Oryza sativa (Oryza sativa)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Physcomitrella patens (Physcomitrella patens)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Populus trichocarpa (Populus trichocarpa)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Selaginella moellendorffii (Selaginella moellendorffii)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Setaria italica (Setaria italica)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Solanum lycopersicum (Solanum lycopersicum)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Solanum tuberosum (Solanum tuberosum)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Sorghum bicolor (Sorghum bicolor)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Triticum aestivum (Triticum aestivum)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Vitis vinifera (Vitis vinifera)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Zea mays (Zea mays)	FASTA (DNA)	FASTA (cDNA)	FASTA (protein)	EMBL	GenBank	MySQL	GTF
Pan_compara Multi-species	-	-	-	-	-	mySQL	-
Plants Multi-species	-	-	-	-	-	mySQL	-
Ensembl Mart	-	-	-	-	-	mySQL	-

http://plants.ensembl.org/info/data/ftp/index.html

Reads 的map率

Samtools 对Tophat 运行后输出文件accepted_hits.bam进行统计，也可建索引之后用IGV等基因组浏览器查看map的情况。

Sample	Total_reads	Total_mapped_reads	Percentage
W	27448106	14575338	53.1014%
M	44379092	26059195	58.7195%

Li *et al*. 2009 Bioinformatics
http://www.broadinstitute.org/igv/
James *et al*. 2011 *Nat. Biotech*

差异表达基因统计

Cuffdiff 输出文件gene_exp.diff，选择significant项为“yes”的基因

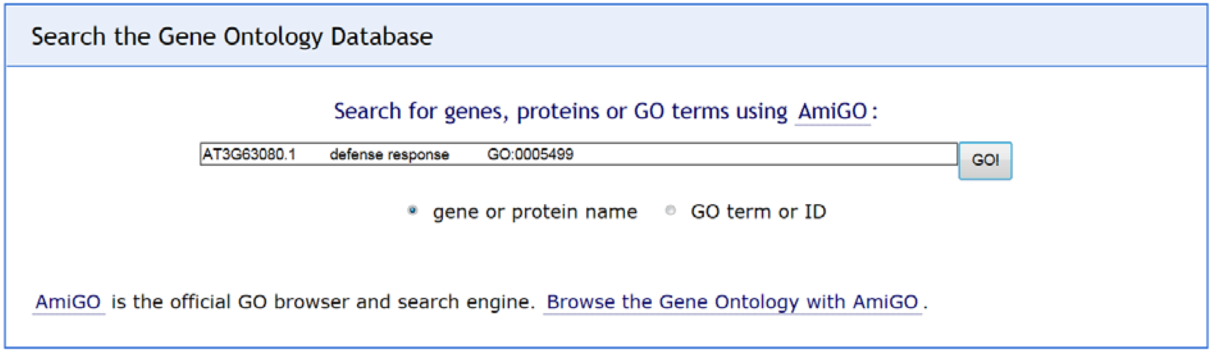
DEGs Mutant vs. WT	
Up-regulated	Down-regulated
144	332

http://cufflinks.cbcb.umd.edu/
Trapnell et al. 2013 *Nat. Biotech.*

GO 注释及其富集分析

GO
(Gene Ontology)

- 分子功能（催化活性、结合活性以及转移酶活性……）
- 生物学过程（调控、跨膜转运、代谢及应激……）
- 细胞组分（胞膜、胞质、核、细胞器……）



Search the Gene Ontology Database

Search for genes, proteins or GO terms using AmiGO:

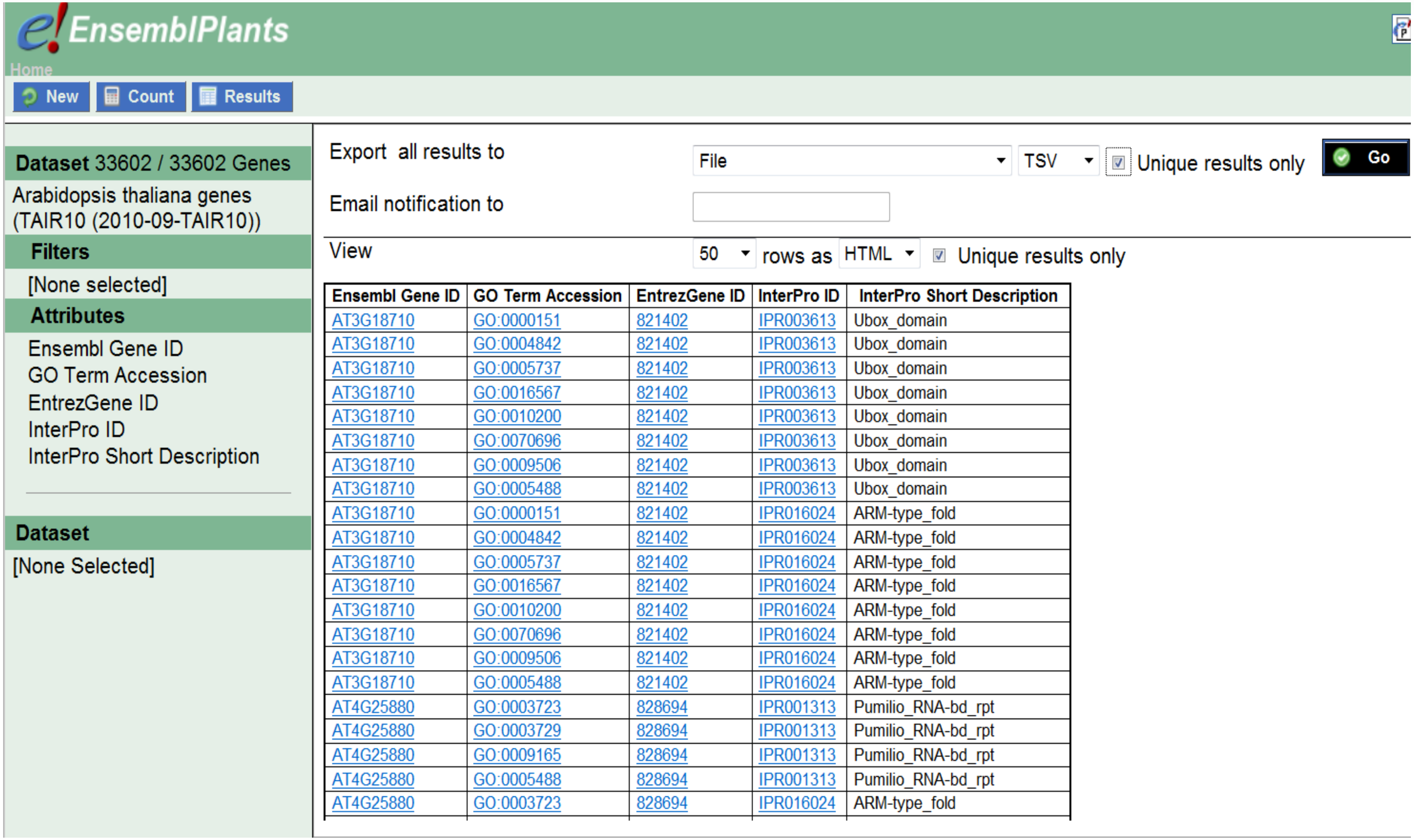
AT3G63080.1 defense response GO:0005499

● gene or protein name ○ GO term or ID

AmiGO is the official GO browser and search engine. Browse the Gene Ontology with AmiGO.

http://www.geneontology.org/

GO 注释下载



Dataset 33602 / 33602 Genes

Arabidopsis thaliana genes
(TAIR10 (2010-09-TAIR10))

Filters

[None selected]

Attributes

Ensembl Gene ID
GO Term Accession
EntrezGene ID
InterPro ID
InterPro Short Description

Dataset

[None Selected]

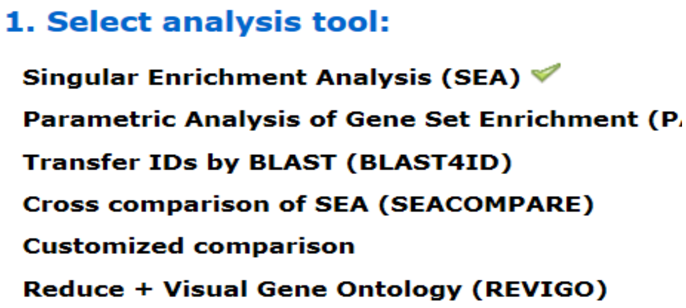
Export all results to: File, TSV, ☑ Unique results only

Email notification to

View 50 rows as HTML ☑ Unique results only

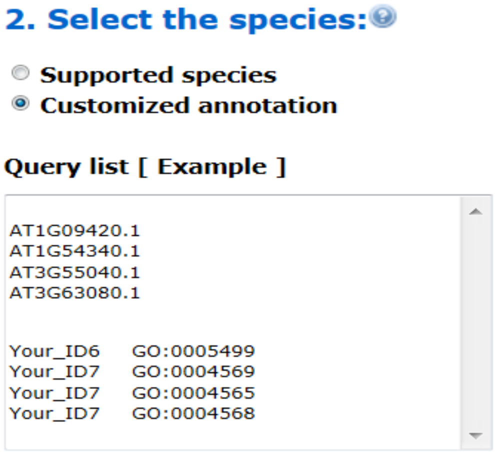
Ensembl Gene ID	GO Term Accession	EntrezGene ID	InterPro ID	InterPro Short Description
AT3G18710	GO:0000151	821402	IPR003613	Ubox_domain
AT3G18710	GO:0004842	821402	IPR003613	Ubox_domain
AT3G18710	GO:0005737	821402	IPR003613	Ubox_domain
AT3G18710	GO:0016567	821402	IPR003613	Ubox_domain
AT3G18710	GO:0010200	821402	IPR003613	Ubox_domain
AT3G18710	GO:0070696	821402	IPR003613	Ubox_domain
AT3G18710	GO:0009506	821402	IPR003613	Ubox_domain
AT3G18710	GO:0005488	821402	IPR003613	Ubox_domain
AT3G18710	GO:0000151	821402	IPR016024	ARM-type_fold
AT3G18710	GO:0004842	821402	IPR016024	ARM-type_fold
AT3G18710	GO:0005737	821402	IPR016024	ARM-type_fold
AT3G18710	GO:0016567	821402	IPR016024	ARM-type_fold
AT3G18710	GO:0010200	821402	IPR016024	ARM-type_fold
AT3G18710	GO:0070696	821402	IPR016024	ARM-type_fold
AT3G18710	GO:0009506	821402	IPR016024	ARM-type_fold
AT3G18710	GO:0005488	821402	IPR016024	ARM-type_fold
AT4G25880	GO:0003723	828694	IPR001313	Pumilio_RNA-bd_rpt
AT4G25880	GO:0003729	828694	IPR001313	Pumilio_RNA-bd_rpt
AT4G25880	GO:0009165	828694	IPR001313	Pumilio_RNA-bd_rpt
AT4G25880	GO:0005488	828694	IPR001313	Pumilio_RNA-bd_rpt
AT4G25880	GO:0003723	828694	IPR016024	ARM-type_fold

GO 富集分析网站



1. Select analysis tool:

Singular Enrichment Analysis (SEA) ✓
Parametric Analysis of Gene Set Enrichment (P
Transfer IDs by BLAST (BLAST4ID)
Cross comparison of SEA (SEACOMPARE)
Customized comparison
Reduce + Visual Gene Ontology (REVIGO)



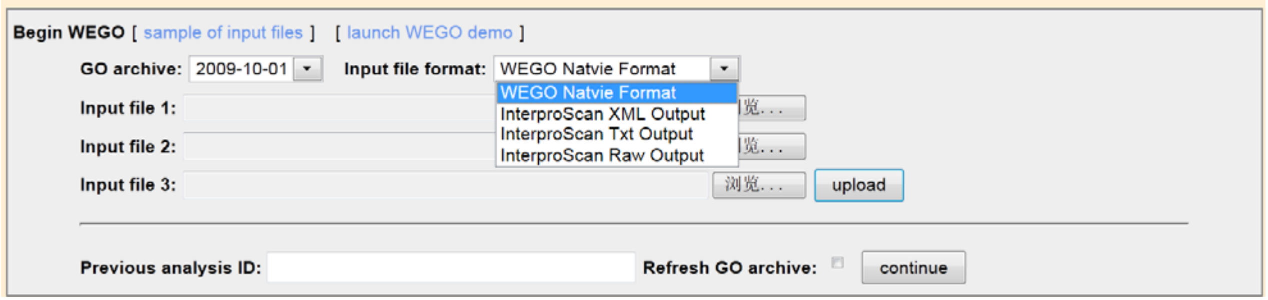
2. Select the species:

○ Supported species
● Customized annotation

Query list [Example]

```
AT1G09420.1
AT1G54340.1
AT3G55040.1
AT3G63080.1


Your_ID6     GO:0005499
Your_ID7     GO:0004569
Your_ID7     GO:0004565
Your_ID7     GO:0004568
```



Begin WEGO [sample of input files] [launch WEGO demo]

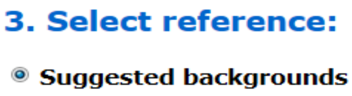
GO archive: 2009-10-01 Input file format: WEGO Natvie Format

WEGO Natvie Format
InterproScan XML Output
InterproScan Txt Output
InterproScan Raw Output

Input file 1:
Input file 2:
Input file 3:

Previous analysis ID:

Refresh GO archive: ☐

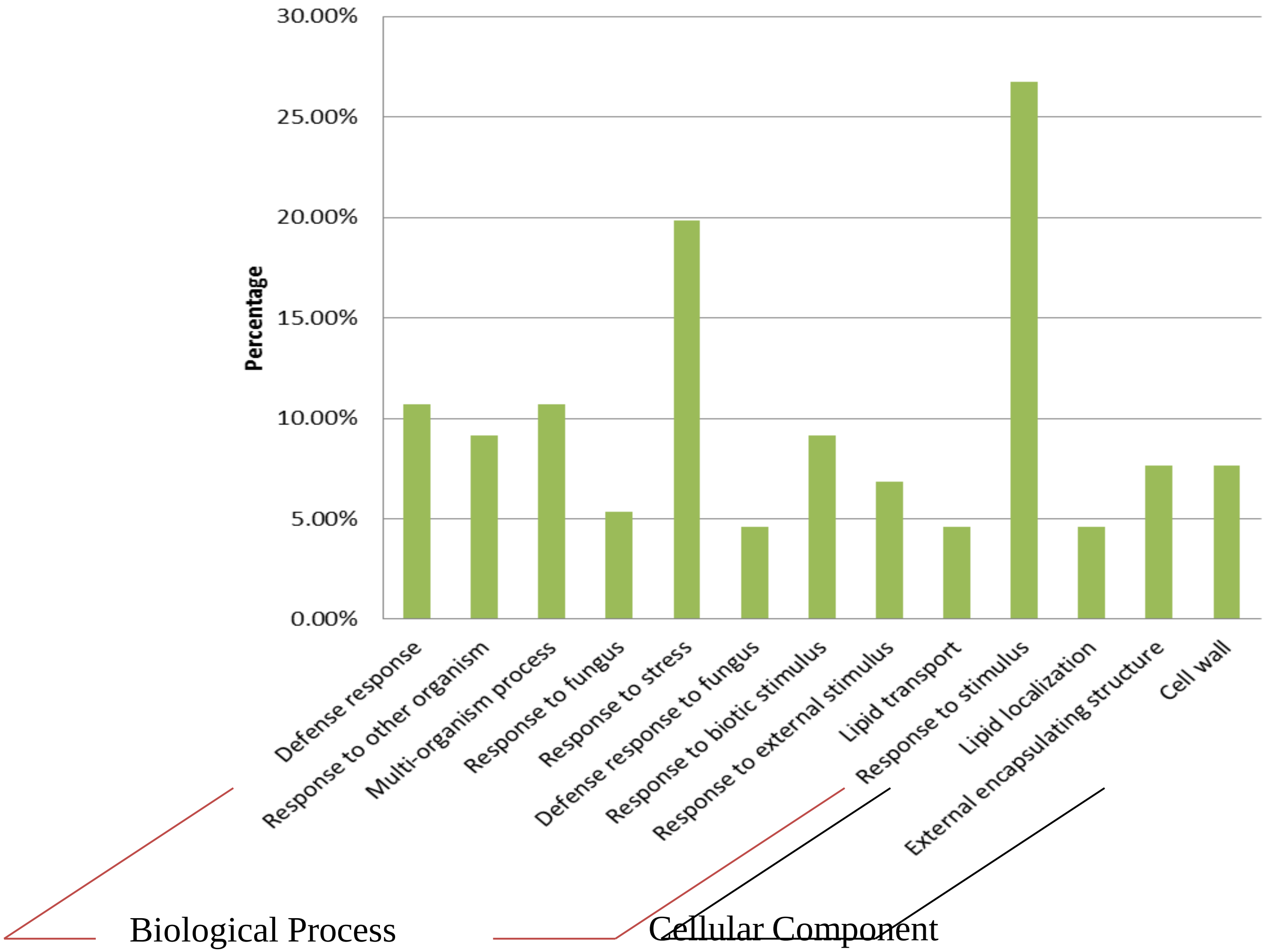


3. Select reference:

● Suggested backgrounds

http://wego.genomics.org.cn/cgi-bin/wego/index.pl
http://bioinfo.cau.edu.cn/agriGO/analysis.php
http://omicslab.genetics.ac.cn/GOEAST/

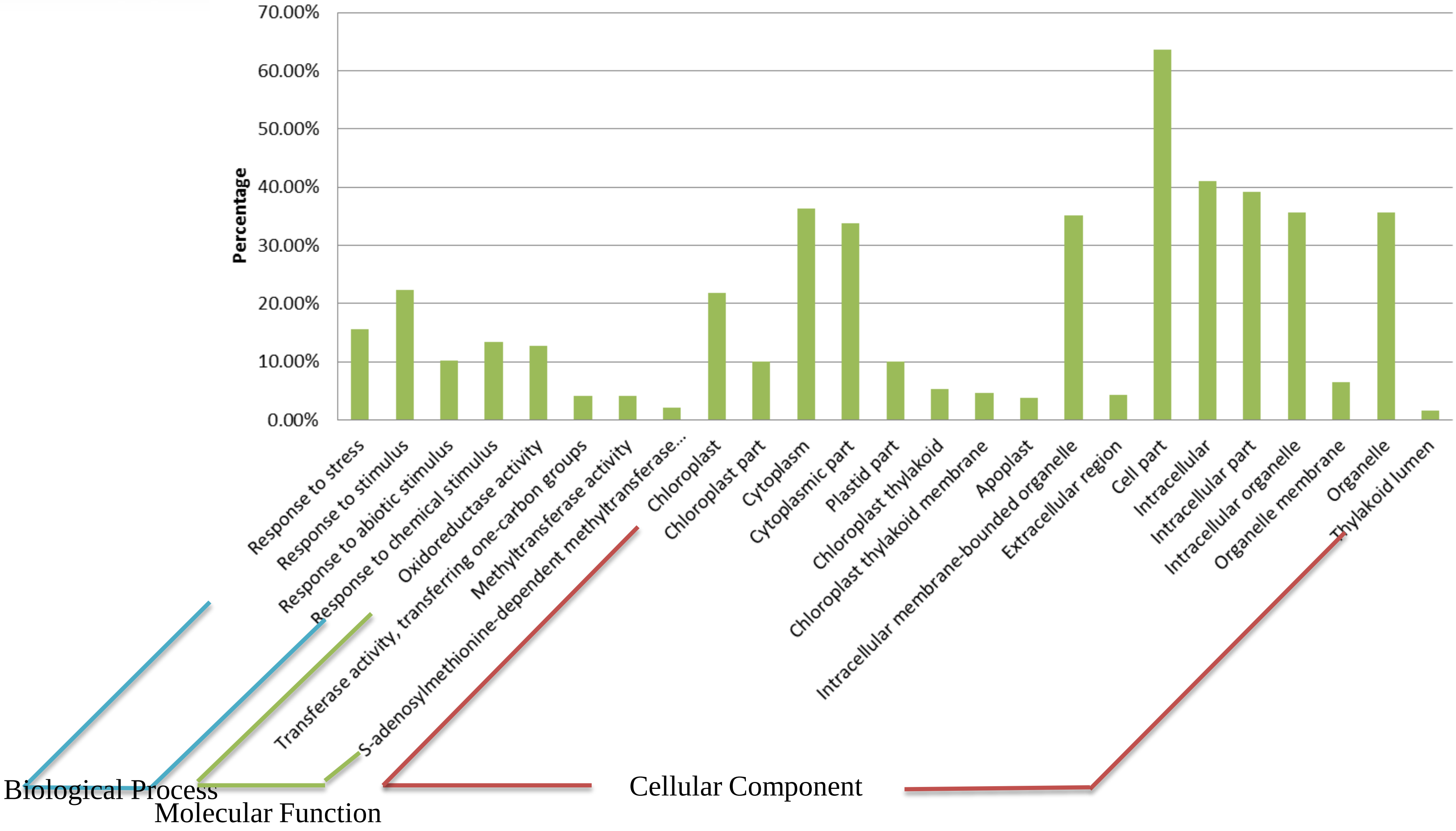
Enriched GO terms in up-regulated genes



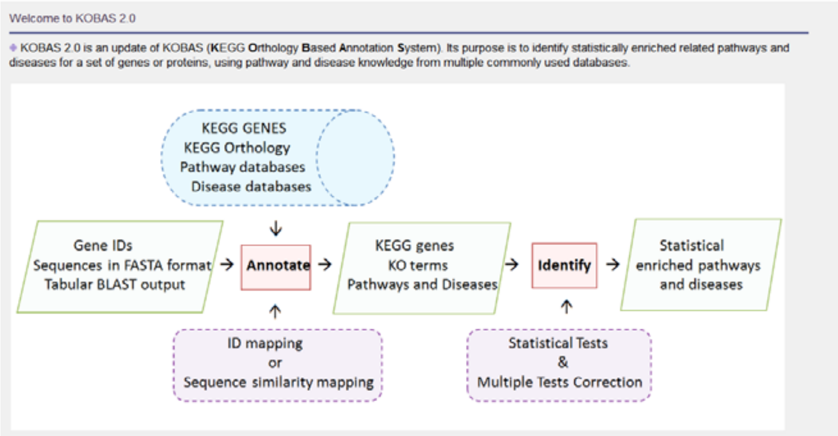
Biological Process

Cellular Component

Enriched GO terms in down-regulated genes



代谢通路注释与富集分析



Welcome to KOBAS 2.0

◆ KOBAS 2.0 is an update of KOBAS (**K**EGG **O**rthology **B**ased **A**nnotation **S**ystem). Its purpose is to identify statistically enriched related pathways and diseases for a set of genes or proteins, using pathway and disease knowledge from multiple commonly used databases.

http://kobas.cbi.pku.edu.cn/home.do

Help and Tool Manual

139 DAVID IDs

Check Defaults ☑ Clear All

Step 3: List Type

Gene List ●

Background ○

Step 4: Submit List

Submit List

		19	Chart	
		81	Chart	
☐ EC_NUMBER	10.8%	15	Chart	
☑ KEGG_PATHWAY	7.9%	11	Chart	
☐ PANTHER_PATHWAY	2.2%	3	Chart	

⊞ Protein_Domains (3 selected)

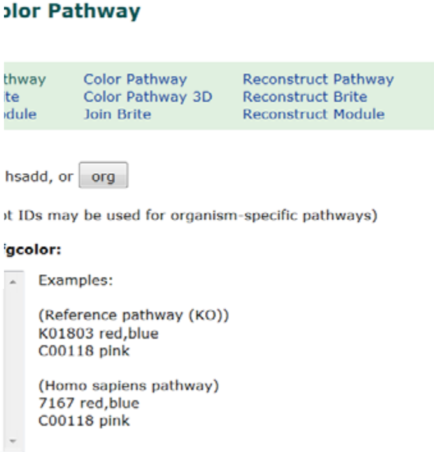
⊟ Protein_Interactions (0 selected)

☑ BIND	3.6%	5	Chart	

⊟ Tissue_Expression (0 selected)

☐ PIR_TISSUE_SPECIFICITY	10.1%	14	Chart	
☐ UP_TISSUE	6.5%	9	Chart	

http://david.abcc.ncifcrf.gov/home.jsp



olor Pathway

thway	Color Pathway	Reconstruct Pathway
te	Color Pathway 3D	Reconstruct Brite
dule	Join Brite	Reconstruct Module

hsadd, or org

t IDs may be used for organism-specific pathways)

gcolor:

```
ath:AT1G05894    magenta,blue
ath:AT1G09530    magenta,blue
```

Examples:

(Reference pathway (KO))
K01803 red,blue
C00118 pink

(Homo sapiens pathway)
7167 red,blue
C00118 pink

http://www.genome.jp/kegg/tool/map_pathway2.html

Enriched Pathway in up-regulated genes

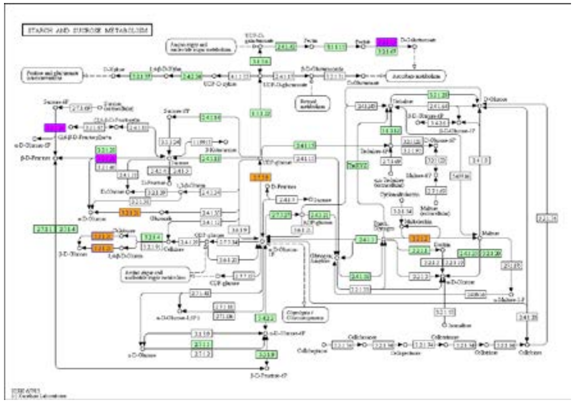
ath04145 Phagosome - Arabidopsis thaliana (thale cress) (4)			
AT1G04820	TUA4	tubulin alpha-2/alpha-4 chain	up
AT1G22740	RABG3B	Ras-related protein Rab7	down
AT1G50010	TUA2	tubulin alpha-2/alpha-4 chain	up
AT4G14960	TUA6	tubulin alpha-6 chain	up
ath04626 Plant-pathogen interaction - Arabidopsis thaliana (thale cress) (5)			
AT2G14610	PR1	pathogenesis-related protein 1	up
AT2G19190	FRK1	senescence-induced receptor-like serine/threonine-protein kinase	up
AT4G02930		Elongation factor Tu	up
AT5G20480	EFR	LRR receptor-like serine/threonine-protein kinase EFR	up
AT5G57940	CNGC5	cyclic nucleotide gated channel	down
ath00592 alpha-Linolenic acid metabolism - Arabidopsis thaliana (thale cress) (2)			
AT3G25770	AOC2	allene oxide cyclase 2	up
AT3G45140	LOX2	lipoxygenase 2	up

Enriched Pathway in down-regulated genes

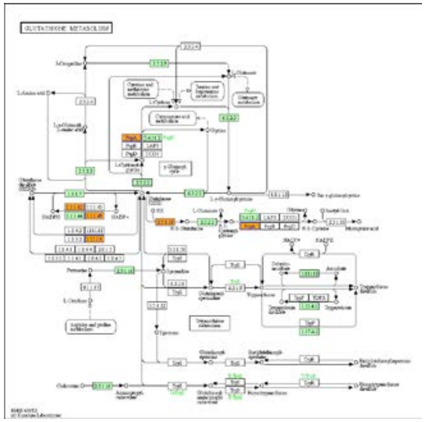
ath00902 Monoterpenoid biosynthesis - Arabidopsis thaliana (thale cress) (3)			
AT3G25820	TPS-CIN	1,8-cineole synthase	down
AT3G25830	TPS-CIN	1,8-cineole synthase	down
AT3G61220		(+)-neomenthol dehydrogenase	down
ath00480 Glutathione metabolism - Arabidopsis thaliana (thale cress) (5)			
AT1G09420	G6PD4	glucose-6-phosphate dehydrogenase 4	down
AT1G54340	ICDH	isocitrate dehydrogenase	down
AT3G55040	GSTL2	glutathione transferase lambda 2	down
AT3G63080	GPX5	glutathione peroxidase	down
AT4G30910		leucine aminopeptidase 2	down
ath00280 Valine, leucine and isoleucine degradation - Arabidopsis thaliana (thale cress) (3)			
AT1G48030	mtLPD1	dihydrolipoyl dehydrogenase 1	down
AT3G13450	DIN4	2-oxoisovalerate dehydrogenase E1 component, beta subunit	down
AT3G45300	IVD	isovaleryl-CoA-dehydrogenase	down

Enriched Pathway in all DEGs (combined up/down-regulated genes)

ath00500 Starch and sucrose metabolism - Arabidopsis thaliana (thale cress) (8)			
AT1G02640	BXL2	beta-glucosidase	down
AT2G44460	BGLU28	beta glucosidase 28	up
AT3G13790	ATBFRUCT1	beta-fructofuranosidase	up
AT3G60140	DIN2	beta-glucosidase 30	up
AT4G17090	CT-BMY	beta-amylase 3	down
AT5G14650	polygalacturonase		up
AT5G17310	UGP2	UTP--glucose-1-phosphate uridylyltransferase 1	down
AT5G20950		Glycosyl hydrolase family protein	up
ath00480 Glutathione metabolism - Arabidopsis thaliana (thale cress) (5)			
AT1G09420	G6PD4	glucose-6-phosphate dehydrogenase 4	down
AT1G54340	ICDH	isocitrate dehydrogenase	down
AT3G55040	GSTL2	glutathione transferase lambda 2	down
AT3G63080	GPX5	glutathione peroxidase	down
AT4G30910		leucine aminopeptidase 2	down

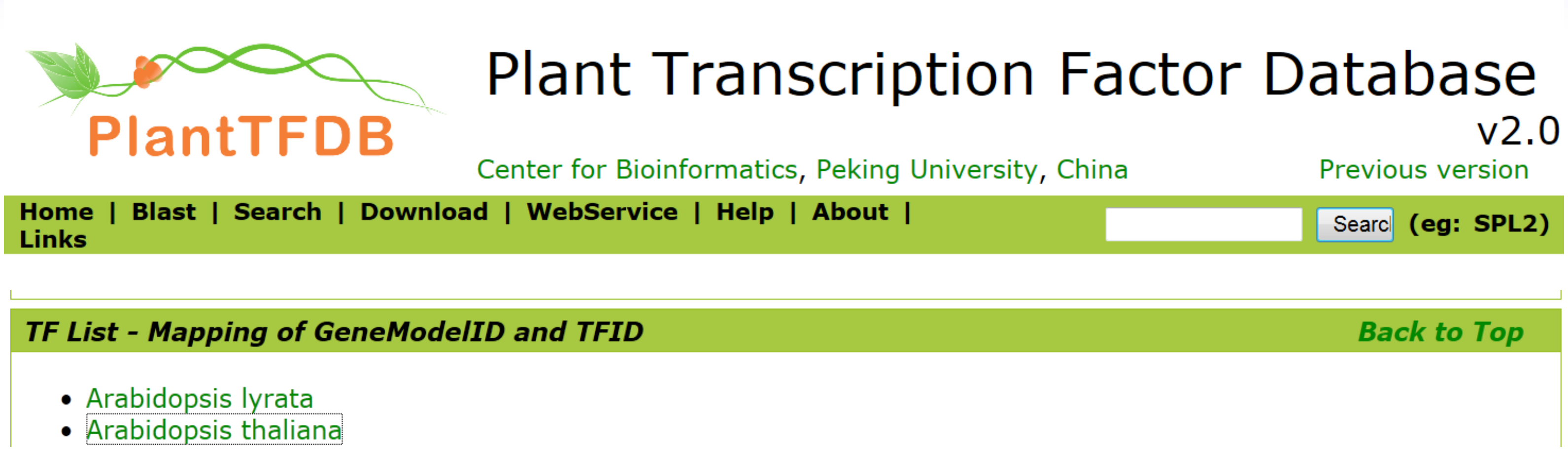


Starch and sucrose metabolism



Glutathione metabolism

差异表达基因中转录因子的分析



Plant Transcription Factor Database
v2.0
Center for Bioinformatics, Peking University, China
Previous version

TF List - Mapping of GeneModelID and TFID

- Arabidopsis lyrata
- Arabidopsis thaliana

PLACE

A Database of Plant Cis-acting Regulatory DNA Elements

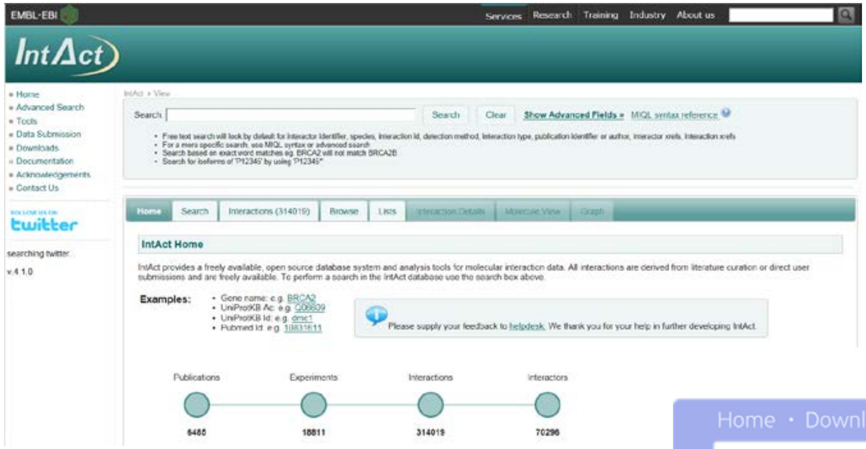
TRANSFAC® Transcription Factor Binding Sites

http://planttfdb.cbi.edu.cn/
http://www.dna.affrc.go.jp/PLACE/signalscan.html
http://www.biobase-international.com/product/transcription-factor-binding-sites

差异表达基因中转录因子的分析

Family name	**Gene NO.**	**Mutant vs. WT**	**Gene name**	**Description**
bHLH	AT1G09530	up	POC1	transcription factor PIF3 [Source:EMBL
bHLH	AT1G10610	up	AT1G10610	transcription factor bHLH90 [Source:EMBL
bHLH	AT5G04150	down	BHLH101	transcription factor bHLH101 [Source:EMBL
C3H	AT1G29560	up	AT1G29560	Zinc finger C-x8-C-x5-C-x3-H type family protein [Source:EMBL
C3H	AT2G40140	up	SZF2	zinc finger CCCH domain-containing protein 29 [Source:EMBL
C3H	AT4G29190	up	AT4G29190	zinc finger CCCH domain-containing protein 49 [Source:EMBL
CO-like	AT3G02380	down	COL2	zinc finger protein CONSTANS-LIKE 2 [Source:EMBL
CO-like	AT5G15850	down	COL1	zinc finger protein CONSTANS-LIKE 1 [Source:EMBL
ERF	AT1G21910	down	DREB26	ethylene-responsive transcription factor ERF012 [Source:EMBL
FAR1	AT1G80010	down	FRS8	FAR1-related sequence 8 [Source:EMBL
GRF	AT5G53660	up	GRF7	growth-regulating factor 7 [Source:EMBL
MIKC	AT5G10140	down	FLF	MADS-box protein FLOWERING LOCUS C [Source:EMBL
MYB_related	AT1G01060	down	LHY	protein late elongated hypocotyl [Source:EMBL
MYB_related	AT3G09600	down	AT3G09600	myb family transcription factor [Source:EMBL
MYB_related	AT5G17300	down	RVE1	myb family transcription factor [Source:EMBL
MYB_related	AT5G37260	down	RVE2	protein REVEILLE 2 / DNA binding / transcription factor [Source:EMBL
RAV	AT1G13260	down	RAV1	AP2/ERF and B3 domain-containing transcription factor RAV1 [Source:EMBL
WRKY	AT5G49520	up	WRKY48	putative WRKY transcription factor 48 [Source:EMBL

相互关联基因/蛋白分析



- Free text search will look by default for Interactor Identifier, species, Interaction Id, detection method, Interaction type, publication identifier or author, interactor xrefs, Interaction xrefs
- For a more specific search, use MIQL syntax or advanced search
- Search based on exact word matches eg. BRCA2 will not match BRCA2B
- Search for isoforms of 'P12345' by using 'P12345*'

IntAct Home

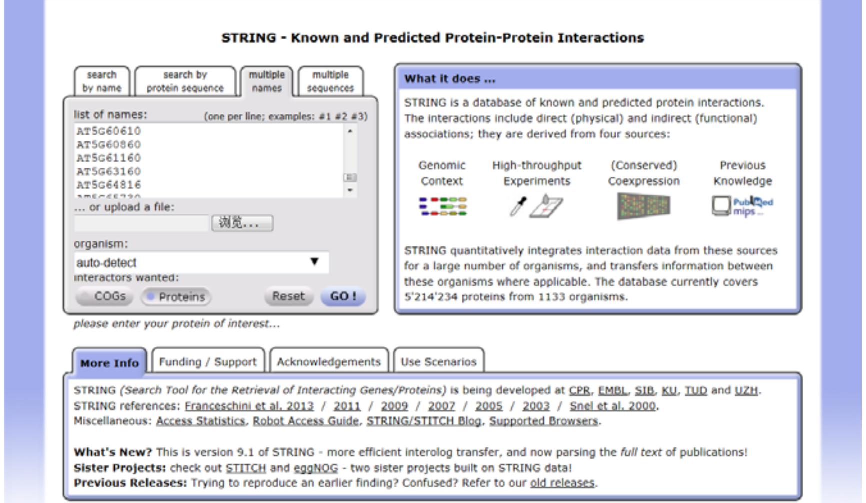
IntAct provides a freely available, open source database system and analysis tools for molecular interaction data. All interactions are derived from literature curation or direct user submissions and are freely available. To perform a search in the IntAct database use the search box above.

Examples:
- Gene name: e.g. BRCA2
- UniProtKB Ac: e.g. Q06609
- UniProtKB Id: e.g. dmc1
- Pubmed Id: e.g. 10831611

Please supply your feedback to helpdesk. We thank you for your help in further developing IntAct.

Publications	Experiments	Interactions	Interactors
6485	18811	314019	70296

http://



STRING - Known and Predicted Protein-Protein Interactions

list of names: (one per line; examples: #1 #2 #3)

```
AT5G60610
AT5G60860
AT5G61160
AT5G63160
AT5G64816
```

... or upload a file:

organism: auto-detect

interactors wanted: Proteins

please enter your protein of interest...

What it does ...

STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5'214'234 proteins from 1133 organisms.

STRING *(Search Tool for the Retrieval of Interacting Genes/Proteins)* is being developed at CPR, EMBL, SIB, KU, TUD and UZH.
STRING references: Franceschini et al. 2013 / 2011 / 2009 / 2007 / 2005 / 2003 / Snel et al. 2000.
Miscellaneous: Access Statistics, Robot Access Guide, STRING/STITCH Blog, Supported Browsers.

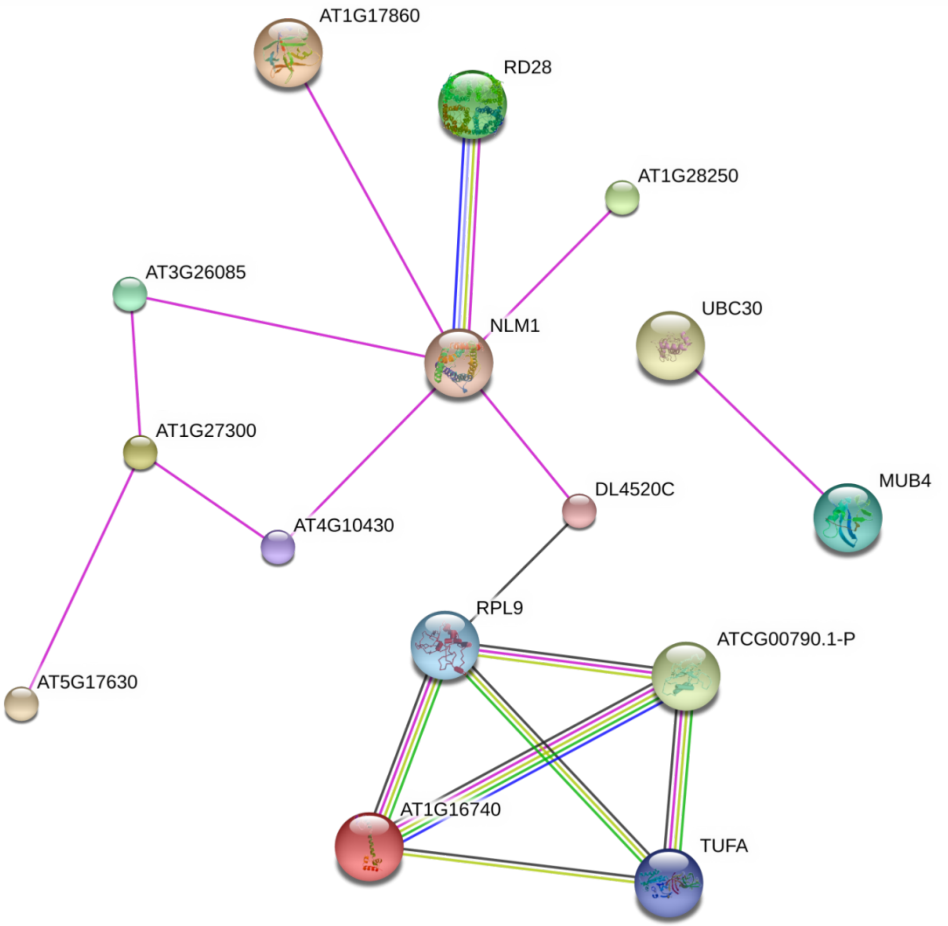
What's New? This is version 9.1 of STRING - more efficient interolog transfer, and now parsing the *full text* of publications!
Sister Projects: check out STITCH and eggNOG - two sister projects built on STRING data!
Previous Releases: Trying to reproduce an earlier finding? Confused? Refer to our old releases.

http://string-db.org/newstring_cgi/show_input_page.pl?UserId=0PM0sh3Lgc9H&sessionId=Rp5eTfli64fj&input_page_type=multiple_identifiers

相互关联蛋白总结（部分）

node1_external_id	node2_external_id	neighborhood	fusion	cooccurence	homology	coexpression	experimental	knowledge	textmining	combined_score
AT5G17630.1-P	AT1G27300.1-P	0	0	0	0	0	0.923	0	0	0.923
AT4G10430.3-P	AT1G27300.1-P	0	0	0	0	0	0.923	0	0	0.923
AT3G26085.2-P	AT1G27300.1-P	0	0	0	0	0	0.923	0	0	0.923
AT4G19030.1-P	AT2G37180.1-P	0	0	0.523	0.759	0	0.923	0	0.68	0.942
AT4G19030.1-P	AT1G28250.1-P	0	0	0	0	0	0.923	0	0	0.923
ATCG00790.1-P	AT3G44890.1-P	0	0	0	0	0.62	0.718	0	0.773	0.972
AT3G44890.1-P	AT1G16740.1-P	0.193	0	0	0	0.351	0.633	0	0.795	0.952
ATCG00790.1-P	AT1G16740.1-P	0.16	0	0.291	0	0.286	0.612	0	0.8	0.957
AT5G56150.1-P	AT3G26980.1-P	0	0	0	0	0	0.612	0	0	0.611
AT4G19030.1-P	AT4G10430.3-P	0	0	0	0	0	0.462	0	0	0.462
AT4G19030.1-P	AT3G26085.2-P	0	0	0	0	0	0.462	0	0	0.462
AT4G19030.1-P	AT1G17860.1-P	0	0	0	0	0	0.462	0	0	0.462
AT4G19030.1-P	AT4G16980.1-P	0	0	0	0	0	0.462	0	0	0.462
ATCG00790.1-P	AT4G02930.1-P	0.784	0	0	0	0.341	0.108	0	0.722	0.957

有实验证据的关联蛋白的关联图（evidence view）

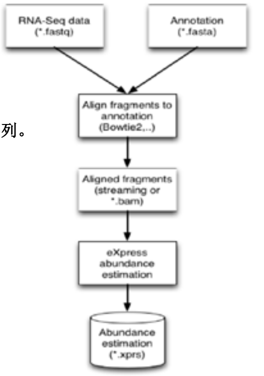


不同颜色的线条代表不同（级别）的证据

分析流程不是唯一的

- 软件不唯一

Bowtie2 ---- eXpress ---- EdgeR/DEGseq



Step 1： Bowtie2将reads比对到注释的转录组序列。
产生比对文件（.bam格式）

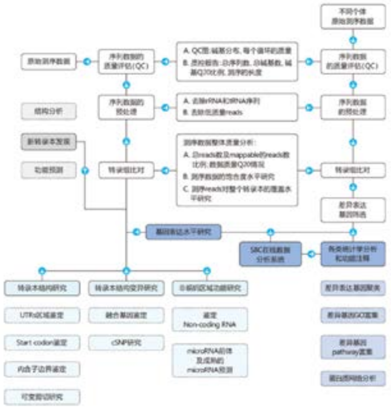
Step 2： eXpress对比对文件进行定量估计。
产生包含counts数及FPKM值的文件
（.xprs格式）

Step 3： 表达差异分析软件进行表达差异分析

EdgeR, DEGseq …

Roberts and Pachter 2012, Nat. Meth.
Langmead and Salzberg 2012, Nat. Meth.
Robinson et al. 2010, Bioinformatics

- 生物学问题不唯一



http://www.ebioservice.com/show_product.asp?id=471

路还很长 ……

谢谢！