

Differential expression analysis of RNA-seq data

Ornela Maloku

February 2021

1 Introduction

In this work some preliminary results obtained from differential expression analysis of 36 RNA-Seq samples of *Arabidopsis Thaliana* are reported in order to evaluate which are the most differential expressed genes among the different Time Points (TP). As for the Ribo-Seq data, we have 36 samples, 12 TP and for each TP 3 replicates. Among the numerous packages developed for this kind of analysis, the Bioconductor package edgeR was used for differential expression analyses of read counts from RNA-Seq (<https://www.bioconductor.org/packages/devel/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf>). Another important tool used is featureCounts to get the gene counts. This tool is accurate, easy to use and fast (<http://subread.sourceforge.net/>).

2 Preliminary steps

In this step we skipped the trimming process (Since the reads are 200-500 nt and were sequenced only 100 nt). Differently from the previous work of Ribo-Seq (in which reads were aligned onto the transcriptome reference), here, RNA-Seq reads were aligned onto the reference genome. We used HISAT2 2.20 (<http://daehwankimlab.github.io/hisat2/>) as an aligner. In addition also the *Arabidopsis thaliana*.TAIR10.45.gtf.gtf (<https://www.ensembl.org/info/website/upload/gff.html>). The genome was used as a reference (*Arabidopsis thaliana*.TAIR10.dna.toplevel.fa.gz), the FASTA file and the gtf file were downloaded from (<https://plants.ensembl.org/index.html>).

To map the reads and to assemble and quantify the transcripts represented by those reads, the hisat-build function was used. The function takes in input the FASTA file *Arabidopsis thaliana*.TAIR10.dna.all.fa.gz and generates a set of files (1.ht2, 2.ht2, 3.ht2, 4.ht2, 5.ht2, 6.ht2, 7.ht2, 8.ht2) which build the index file for the alignment step.

The hisat2 function takes in input the index file generated by hisat2-build and the fastq sample to be analyzed. For each sample it gives in output a SAM file (Sequence Alignment Map). In order to obtain sorted BAM files we used the function samtools sort from SAMtools (<https://en.wikipedia.org/wiki/SAMtools>) to convert SAM files in BAM files.

Summarizing all the steps: from 36 raw fastq samples we obtained 36 SAM files and 36 sorted BAM files.

3 DGE analysis

Differential expression analysis takes in consideration the normalised read count data to find out quantitative changes in expression gene levels between experimental groups by the use of statistical methods. Different packages are used for this purpose such as limma, edgeR and DESeq based on negative binomial (NB) distributions or others based on Bayesian approaches based on a negative binomial model. In this study edgeR was used to compute DGE. Before performing the DGE, the BAM files were processed by featureCounts (function of the Rsubread package) in order to obtain a count matrix with raw counts where genes are rows and samples are columns as shown in the picture below Figure 1., where how many the reads “assigned” or counted are tabulated for each sample. This “raw” counts will be used in statistical programs downstream for differential gene expression. The counts represent the total number of reads aligning to each gene. edgeR stores the count matrix data in a list-based data object called a DGEList. In order to circumvent some biological requisitions genes with very low counts should be filtered out prior further analysis. To be considered express a gene should have a count of reads at least 5-10 in a library. The filtering step was computed with count-per-million (CPM), we considered genes that have count-per-million (CPM) values above 1 in at least two libraries. Additional steps were applied to before computing the normalization, such as adding gene annotation using org.At.tair.db and filtering for those genes with defined annotation, and recompute library sizes.

In order to compute the normalization step a calcNormFactors function is used to normalize the library sizes by finding a set of scaling factors for the library sizes that minimizes the log-fold changes between the samples for most genes. The default method used is trimmed mean of M-values (TMM) between each pair of samples in order to eliminate biases between libraries. In Figure 2. the library size is shown how it appears after the normalization, with the DGEList argument with only the norm.factors changed.

4 Differential Expression Analysis

This step requires a designed matrix such that in Figure 3.

The log2-fold-changes(logFC) and correlations among samples were computed as shown in Table 1. and Figures 4.

As it's possible to see, TP23 has a positive corr with T15 and T7 and a negative corr with T3,T11,T19. TPs such T9, T19 and T21, has no correlation with other TPs.

edgeR uses the negative binomial (NB) distribution to model the read counts for each gene in each sample. The dispersion estimation was calculated by using

	mpg_L20845_Pool1- RNA-A1- 1_S1_R1_001.bam	mpg_L20845_Pool1- RNA-A11- 1_S6_R1_001.bam	mpg_L20845_Pool1- RNA-A13- 1_S7_R1_001.bam	mpg_L20845_Pool1- RNA-A15- 1_S8_R1_001.bam	mpg_L20845_Pool1- RNA-A17- 1_S9_R1_001.bam
AT1G01010	4	5	3	4	4
AT1G01010.1	20	16	18	9	32
AT1G01010.2	3	5	0	2	0
AT1G01010.3	32	6	4	4	16
AT1G01010.4	5	5	3	2	8
AT1G01010.5	37	16	17	17	27
AT1G01020	0	0	0	0	0
AT1G01020.1	0	0	0	0	0
AT1G01020.2	0	0	0	0	0
AT1G01020.3	0	0	0	0	0
AT1G01020.4	0	0	0	0	0
AT1G01020.5	0	0	0	0	0
AT1G01020.6	0	0	0	0	0
AT1G01020.7	0	0	0	0	0
AT1G01020.8	0	0	0	0	0
AT1G01020.9	0	0	0	0	0
AT1G01020.10	0	0	0	0	0
AT1G01020.11	0	0	0	0	0
AT1G01020.12	0	0	0	0	0
AT1G01020.13	0	0	0	0	0
AT1G01020.14	0	0	0	0	0

Figure 1: a count matrix generated from featureCounts

the estimateGLMCommonDisp, estimateGLMTrendedDisp and estimateGLM-TagwiseDisp. The dispersion estimation was plotted with the plotBVC as shown in Figure 5. Biological CV (BCV) is the coefficient of variation with which the (unknown) true abundance of the gene varies between replicate RNA samples. The NB dispersions tend to be higher for genes with very low counts. In our case the BCV tends to be in range from 0.2 to 0.5

We tested the DE between groups with the QLF-tests function. In the tables below we can see the top DE with the topTags function, the (FDR), multiple testing correction is performed using the Benjamini-Hochberg method for each contrast pairwise (Table 2 - Table 10).

Table 2 shows the Top DE genes for the contrast TP1 vs TP3, in particular, ATSPX1 gene has the highest logFC, showing that it is highly expressed in TP1 than TP3.

The total number of DE genes identified at an FDR of 5% can be shown with decideTestsDGE function.

As a result TP1vsTP3, TP1vsTP5, TP1vsTP9, TP1vsTP17 and TP1vsTP19 show (Up=0, Down=0, NotSig=12996), TP1vsTP7 (Up=4, Down=9, NotSig=12983), TP1vsTP11 (Up=0, Down=1, NotSig=12995) TP1vsTP13 (Up=3, Down=2, NotSig=12991), TP1vsTP15 (Up=4, Down=12, NotSig=12980), TP1vsTP21 (Up=264 Down=154, NotSig= 12578) and TP1vsTP23 (Up=1, Down=4, NotSig=12991).

The magnitude of the differential expression changes can be visualized with a fitted model MD plot for TP1vsTP3 and TP1vsTP21 as an example Figure 6 and Figure 7. The black dots stands for NotSig genes, the blue ones for Down regulated and Up for overexpressed.

```
> dgList$samples
```

	group	lib.size	norm.factors
mpg_L20845_Pool1-RNA-A1-1_S1_R1_001.bam	1	15884290	1.0737535
mpg_L20845_Pool1-RNA-A11-1_S6_R1_001.bam	1	14811450	1.2529985
mpg_L20845_Pool1-RNA-A13-1_S7_R1_001.bam	1	17225855	0.9551689
mpg_L20845_Pool1-RNA-A15-1_S8_R1_001.bam	1	18926295	0.9930667
mpg_L20845_Pool1-RNA-A17-1_S9_R1_001.bam	1	18353588	1.1013427
mpg_L20845_Pool1-RNA-A19-1_S10_R1_001.bam	1	16276238	1.0076328
mpg_L20845_Pool1-RNA-A21-1_S11_R1_001.bam	1	17324452	0.9575404
mpg_L20845_Pool1-RNA-A23-1_S12_R1_001.bam	1	17060373	0.8778263
mpg_L20845_Pool1-RNA-A3-1_S2_R1_001.bam	1	16428002	0.9934545
mpg_L20845_Pool1-RNA-A5-1_S3_R1_001.bam	1	14384752	0.9181337
mpg_L20845_Pool1-RNA-A7-1_S4_R1_001.bam	1	16102278	0.9868571
mpg_L20845_Pool1-RNA-A9-1_S5_R1_001.bam	1	16735261	0.9702839
mpg_L20846_Pool2-RNA-A1-2_S13_R1_001.bam	1	19445040	1.0104706
mpg_L20846_Pool2-RNA-A11-2_S18_R1_001.bam	1	16986567	1.2919555
mpg_L20846_Pool2-RNA-A13-2_S19_R1_001.bam	1	17632921	1.2154652
mpg_L20846_Pool2-RNA-A15-2_S20_R1_001.bam	1	17140672	1.0953377
mpg_L20846_Pool2-RNA-A17-2_S21_R1_001.bam	1	19221092	1.2361560
mpg_L20846_Pool2-RNA-A19-2_S22_R1_001.bam	1	15383414	0.9185436
mpg_L20846_Pool2-RNA-A21-2_S23_R1_001.bam	1	16662846	0.9022014
mpg_L20846_Pool2-RNA-A23-2_S24_R1_001.bam	1	15132236	0.8686190
mpg_L20846_Pool2-RNA-A3-2_S14_R1_001.bam	1	15879287	1.0263717
mpg_L20846_Pool2-RNA-A5-2_S15_R1_001.bam	1	16600071	1.0703930
mpg_L20846_Pool2-RNA-A7-2_S16_R1_001.bam	1	18483358	1.0780040
mpg_L20846_Pool2-RNA-A9-2_S17_R1_001.bam	1	17948554	1.0756313
mpg_L20847_Pool3-RNA-A1-3_S25_R1_001.bam	1	15648892	1.0733293
mpg_L20847_Pool3-RNA-A11-3_S30_R1_001.bam	1	16244540	0.9847270
mpg_L20847_Pool3-RNA-A13-3_S31_R1_001.bam	1	14202491	1.2009848
mpg_L20847_Pool3-RNA-A15-3_S32_R1_001.bam	1	15974743	0.9067382
mpg_L20847_Pool3-RNA-A17-3_S33_R1_001.bam	1	14645628	1.0475735
mpg_L20847_Pool3-RNA-A19-3_S34_R1_001.bam	1	15730326	1.0020615
mpg_L20847_Pool3-RNA-A21-3_S35_R1_001.bam	1	14961199	0.7819453
mpg_L20847_Pool3-RNA-A23-3_S36_R1_001.bam	1	14470845	0.8984125
mpg_L20847_Pool3-RNA-A3-3_S26_R1_001.bam	1	16132468	0.7819131
mpg_L20847_Pool3-RNA-A5-3_S27_R1_001.bam	1	15418668	0.8798353
mpg_L20847_Pool3-RNA-A7-3_S28_R1_001.bam	1	14891755	0.8067525
mpg_L20847_Pool3-RNA-A9-3_S29_R1_001.bam	1	15641841	1.0363107

Figure 2: DGEList object list after the normalization step

An heatmap plot was created to display the top 100 Genes among samples. First we converted the read counts into log2-counts-per-million (logCPM) values with the `cmp` function. Then we scale each row (each gene) to have mean zero and standard deviation one. Heatmap clusters genes and samples based on the Euclidean distance between the expression values (the heatmap will cluster together genes that have positively correlated logCPM values).

The top 100 genes were used for functional annotation analysis in DAVID database (<https://david.ncifcrf.gov/tools.jsp>). Look at Figure 10.

5 Gene Ontology with Kegg

Here are shown the Kegg pathway for TP1vsTP7 (Up=4, Down=9, NotSig=12983), TP1vsTP13 (Up=3, Down=2, NotSig=12991), TP1vsTP15 (Up=4, Down=12, NotSig=12980), TP1vsTP21 (Up=264 Down=154, NotSig= 12578) and TP1vsTP23 (Up=1, Down=4, NotSig=12991).

```

> mm
      timeT1 timeT11 timeT13 timeT15 timeT17 timeT19 timeT21 timeT23 timeT3 timeT5 timeT7 timeT9
RNAseq_A1_1      1      0      0      0      0      0      0      0      0      0      0      0
RNAseq_A1_2      1      0      0      0      0      0      0      0      0      0      0      0
RNAseq_A1_3      1      0      0      0      0      0      0      0      0      0      0      0
RNAseq_A3_1      0      0      0      0      0      0      0      0      1      0      0      0
RNAseq_A3_2      0      0      0      0      0      0      0      0      1      0      0      0
RNAseq_A3_3      0      0      0      0      0      0      0      0      1      0      0      0
RNAseq_A5_1      0      0      0      0      0      0      0      0      0      1      0      0
RNAseq_A5_2      0      0      0      0      0      0      0      0      0      1      0      0
RNAseq_A5_3      0      0      0      0      0      0      0      0      0      1      0      0
RNAseq_A7_1      0      0      0      0      0      0      0      0      0      0      1      0
RNAseq_A7_2      0      0      0      0      0      0      0      0      0      0      1      0
RNAseq_A7_3      0      0      0      0      0      0      0      0      0      0      1      0
RNAseq_A9_1      0      0      0      0      0      0      0      0      0      0      0      1
RNAseq_A9_2      0      0      0      0      0      0      0      0      0      0      0      1
RNAseq_A9_3      0      0      0      0      0      0      0      0      0      0      0      1
RNAseq_A11_1     0      1      0      0      0      0      0      0      0      0      0      0
RNAseq_A11_2     0      1      0      0      0      0      0      0      0      0      0      0
RNAseq_A11_3     0      1      0      0      0      0      0      0      0      0      0      0
RNAseq_A13_1     0      0      1      0      0      0      0      0      0      0      0      0
RNAseq_A13_2     0      0      1      0      0      0      0      0      0      0      0      0
RNAseq_A13_3     0      0      1      0      0      0      0      0      0      0      0      0
RNAseq_A15_1     0      0      0      1      0      0      0      0      0      0      0      0
RNAseq_A15_2     0      0      0      1      0      0      0      0      0      0      0      0
RNAseq_A15_3     0      0      0      1      0      0      0      0      0      0      0      0
RNAseq_A17_1     0      0      0      0      1      0      0      0      0      0      0      0
RNAseq_A17_2     0      0      0      0      1      0      0      0      0      0      0      0
RNAseq_A17_3     0      0      0      0      1      0      0      0      0      0      0      0
RNAseq_A19_1     0      0      0      0      0      1      0      0      0      0      0      0
RNAseq_A19_2     0      0      0      0      0      1      0      0      0      0      0      0
RNAseq_A19_3     0      0      0      0      0      1      0      0      0      0      0      0
RNAseq_A21_1     0      0      0      0      0      0      1      0      0      0      0      0
RNAseq_A21_2     0      0      0      0      0      0      1      0      0      0      0      0
RNAseq_A21_3     0      0      0      0      0      0      1      0      0      0      0      0
RNAseq_A23_1     0      0      0      0      0      0      0      1      0      0      0      0
RNAseq_A23_2     0      0      0      0      0      0      0      1      0      0      0      0
RNAseq_A23_3     0      0      0      0      0      0      0      1      0      0      0      0
attr("assign")
[1] 1 1 1 1 1 1 1 1 1 1 1 1
attr("contrasts")
attr("contrasts")$time
[1] "contr.treatment"

```

Figure 3: designed matrix for DGE

	T1	T11	T13	T15	T17	T19	T21	T23	T3	T5	T7	T9
T1	1.00	0.94	0.95	0.97	0.99	0.94	0.95	0.97	0.95	0.96	0.97	0.99
T11	0.94	1.00	0.95	0.92	0.95	0.99	0.94	0.92	0.98	0.94	0.92	0.95
T13	0.95	0.95	1.00	0.94	0.96	0.95	0.99	0.95	0.94	0.98	0.94	0.96
T15	0.97	0.92	0.94	1.00	0.96	0.91	0.93	0.99	0.92	0.94	0.99	0.97
T17	0.99	0.95	0.96	0.96	1.00	0.95	0.96	0.97	0.95	0.96	0.96	0.99
T19	0.94	0.99	0.95	0.91	0.95	1.00	0.95	0.92	0.98	0.95	0.92	0.94
T21	0.95	0.94	0.99	0.93	0.96	0.95	1.00	0.94	0.93	0.98	0.93	0.95
T23	0.97	0.92	0.95	0.99	0.97	0.92	0.94	1.00	0.92	0.95	0.99	0.97
T3	0.95	0.98	0.94	0.92	0.95	0.98	0.93	0.92	1.00	0.95	0.92	0.94
T5	0.96	0.94	0.98	0.94	0.96	0.95	0.98	0.95	0.95	1.00	0.95	0.96
T7	0.97	0.92	0.94	0.99	0.96	0.92	0.93	0.99	0.92	0.95	1.00	0.96
T9	0.99	0.95	0.96	0.97	0.99	0.94	0.95	0.97	0.94	0.96	0.96	1.00

Table 1: correlation of logFC among samples

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
AT5G20150.1	AT5G20150	662	ATSPX1	2.81	2.81	26.85	0.00	0.13
AT2G33550	AT2G33550	795	ASR3	-0.96	2.63	24.73	0.00	0.13
AT2G37710	AT2G37710	2558	LecRK-IV.1	1.04	6.21	24.68	0.00	0.13
AT5G35220.2	AT5G35220	212	AMOS1	1.34	1.60	21.43	0.00	0.16
AT2G40490.4	AT2G40490	730	HEME2	0.76	5.98	21.29	0.00	0.16
AT4G04210.2	AT4G04210	665	PUX4	1.04	4.13	21.18	0.00	0.16
AT4G38730	AT4G38730	499	AVI2H	1.74	0.34	20.25	0.00	0.16
AT1G22770	AT1G22770	372	FB	5.79	3.36	20.03	0.00	0.16
AT2G39570.2	AT2G39570	325	ACR9	-2.33	4.23	19.68	0.00	0.16
AT1G60710	AT1G60710	470	ATB2	-1.33	1.89	19.10	0.00	0.16

Table 2: top DE genes TP1 vs TP3

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
AT2G22500	AT2G22500	2077	ATPUMP5	1.78	4.09	29.43	0.00	0.07
AT1G20650.1	AT1G20650	155	ASG5	-1.27	1.83	28.69	0.00	0.07
AT1G20650	AT1G20650	668	ASG5	-1.25	2.05	26.09	0.00	0.09
AT2G33550	AT2G33550	795	ASR3	-0.93	2.63	22.93	0.00	0.14
AT4G25100	AT4G25100	55	ATFSD1	-1.20	1.55	22.14	0.00	0.14
AT1G20650.3	AT1G20650	392	ASG5	-1.03	2.87	21.99	0.00	0.14
AT2G40610	AT2G40610	375	ATEXP8	-1.55	2.36	20.47	0.00	0.19
AT4G08920	AT4G08920	943	ATCRY1	-0.87	4.59	19.29	0.00	0.19
AT1G77510.2	AT1G77510	186	ATPDI6	1.55	0.01	19.18	0.00	0.19
AT4G08920.2	AT4G08920	1489	ATCRY1	-0.98	7.28	19.09	0.00	0.19

Table 3: top DE genes TP1 vs TP5

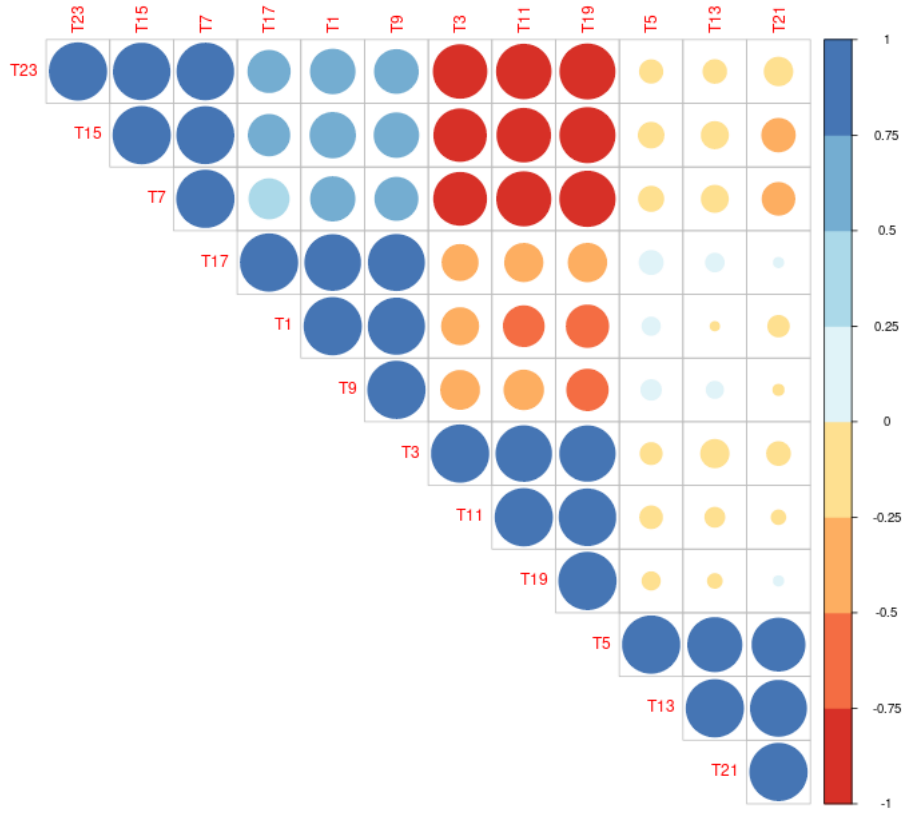


Figure 4: correlation plots for the logFC

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
	AT3G21750	1961	UGT71B1	1.27	5.57	68.10	0.00	0.00
	AT2G40610	375	ATEXP8	-2.21	2.36	40.62	0.00	0.00
	AT5G65480.1	530	CCl1	-1.31	4.21	33.71	0.00	0.01
	AT2G40610.2	637	ATEXP8	-2.14	3.71	29.77	0.00	0.03
	AT2G40610.1	313	ATEXP8	-2.14	2.49	28.55	0.00	0.03
	AT1G20650	668	ASG5	-1.28	2.05	27.05	0.00	0.03
	AT4G08920	943	ATCRY1	-1.02	4.59	26.07	0.00	0.03
	AT1G78020.1	429	FLZ6	1.59	4.05	25.96	0.00	0.03
	AT1G54130	1092	AT-RSH3	-1.13	4.26	25.46	0.00	0.03
	AT3G56480.2	129	SCAB3	1.67	0.34	24.76	0.00	0.03

Table 4: top DE genes TP1 vs TP7

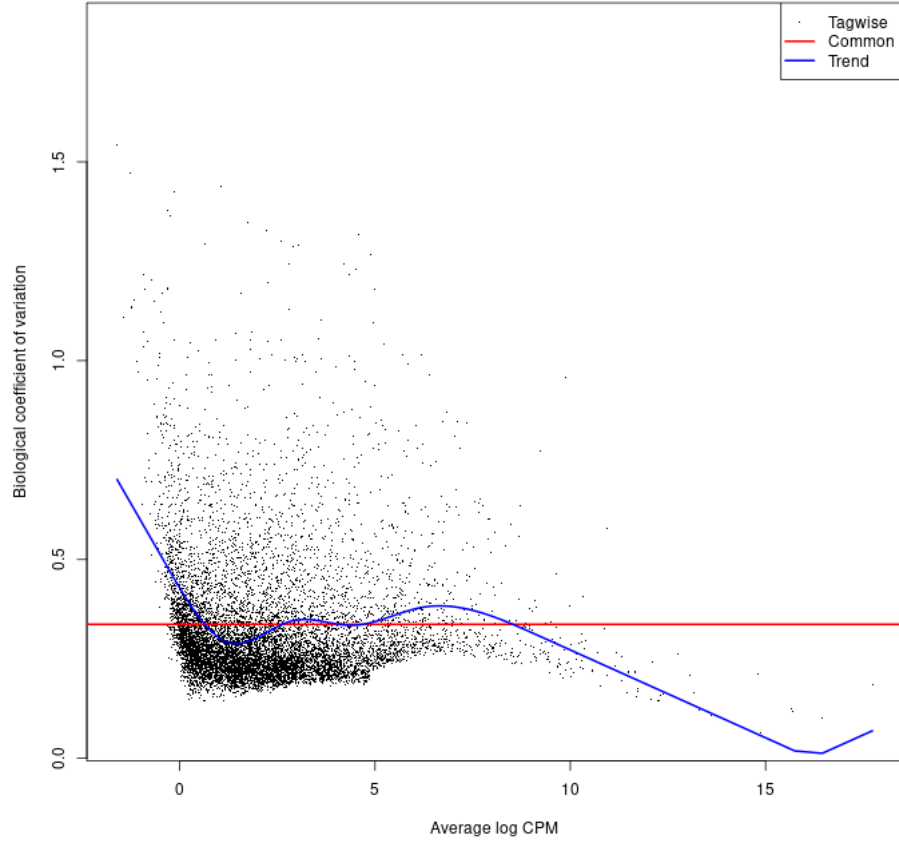


Figure 5: designed matrix for DGE

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
	AT1G49890	1700	QWRF2	-16.13	3.87	6119.21	0.00	0.00
	AT5G24020	1423	ARC11	-15.16	4.71	5997.91	0.00	0.00
	AT2G20190.6	1793	ATCLASP	-14.41	5.60	5916.11	0.00	0.00
	AT5G04920	748	VPS36	-15.78	4.18	5912.72	0.00	0.00
	AT2G17980.1	2072	ATSLY1	-14.18	5.83	5616.39	0.00	0.00
	AT4G32210	752	SDH3-2	-14.95	4.94	5597.68	0.00	0.00
	AT1G80070.11	254	EMB14	-15.66	4.23	5481.68	0.00	0.00
	AT3G11130.13	588	AtCHC1	-14.71	5.13	5278.61	0.00	0.00
	AT5G04990	1348	ATSUN1	-14.95	5.06	5244.69	0.00	0.00
	AT1G71090	2018	PILS2	-14.78	5.24	5187.58	0.00	0.00

Table 5: top DE genes TP1 vs TP9

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
AT5G67590.1	AT5G67590	130	FRO1	-1.31	2.59	45.34	0.00	0.00
AT2G39570.1	AT2G39570	219	ACR9	-2.55	3.35	27.93	0.00	0.06
AT1G76160.3	AT1G76160	346	sks5	-1.41	3.50	27.70	0.00	0.06
AT1G64940	AT1G64940	1726	CYP89A6	1.57	1.60	26.17	0.00	0.06
AT3G20630.2	AT3G20630	134	ATUBP14	-1.11	1.32	23.43	0.00	0.09
AT3G42790.2	AT3G42790	130	AL3	-1.51	0.61	22.92	0.00	0.09
AT5G18650.3	AT5G18650	454	MIEL1	-1.07	2.94	22.13	0.00	0.09
AT5G09920.1	AT5G09920	545	ATRPB15.9	-1.21	1.84	21.68	0.00	0.09
AT2G37710	AT2G37710	2558	LecRK-IV.1	0.96	6.21	20.99	0.00	0.09
AT2G39570.2	AT2G39570	325	ACR9	-2.43	4.23	21.17	0.00	0.09

Table 6: top DE genes TP1 vs TP11

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
AT3G06860.6	AT3G06860	122	ATMFP2	1.77	0.55	37.73	0.00	0.02
AT5G48930	AT5G48930	551	HCT	-1.45	4.35	30.47	0.00	0.03
AT4G22910.1	AT4G22910	243	CCS52A1	1.92	0.13	29.69	0.00	0.03
AT2G22500	AT2G22500	2077	ATPUMP5	1.70	4.09	26.97	0.00	0.05
AT5G67590.1	AT5G67590	130	FRO1	-1.00	2.59	26.49	0.00	0.05
AT4G08920	AT4G08920	943	ATCRY1	-0.98	4.59	24.17	0.00	0.05
AT4G05460.1	AT4G05460	1175	AtSKIP19	-1.50	2.33	24.14	0.00	0.05
AT2G39570.1	AT2G39570	219	ACR9	-2.36	3.35	24.07	0.00	0.05
AT1G03330.1	AT1G03330	347	LSM2	-1.13	2.97	23.40	0.00	0.05
AT1G20650.3	AT1G20650	392	ASG5	-1.06	2.87	23.32	0.00	0.05

Table 7: top DE genes TP1 vs TP13

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
AT3G21750	AT3G21750	1961	UGT71B1	1.12	5.57	53.22	0.00	0.00
AT2G40610	AT2G40610	375	ATEXP8	-2.34	2.36	45.83	0.00	0.00
AT2G40610.2	AT2G40610	637	ATEXP8	-2.09	3.71	28.45	0.00	0.04
AT5G67590.1	AT5G67590	130	FRO1	-1.02	2.59	27.59	0.00	0.04
AT5G05170.6	AT5G05170	138	ATCESA3	-1.03	3.74	26.50	0.00	0.04
AT1G60260	AT1G60260	253	BGLU5	1.62	1.51	26.50	0.00	0.04
AT2G40610.1	AT2G40610	313	ATEXP8	-2.00	2.49	25.21	0.00	0.04
AT5G16400	AT5G16400	465	ATF2	-0.99	4.14	24.91	0.00	0.04
AT4G23890	AT4G23890	1051	CRR31	-0.91	4.69	24.79	0.00	0.04
AT1G03130	AT1G03130	956	PSAD-2	-1.79	8.16	24.03	0.00	0.04

Table 8: top DE genes TP1 vs TP15

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
AT5G27150	AT5G27150	468	AT-NHX1	0.92	1.17	15.11	0.00	1.00
AT3G52190.2	AT3G52190	444	AtPHF1	-1.18	0.43	14.77	0.00	1.00
AT1G70560.1	AT1G70560	290	CKRC1	-1.11	0.08	14.37	0.00	1.00
AT3G57080.1	AT3G57080	731	DMS12	-0.99	0.88	12.69	0.00	1.00
AT1G09020.2	AT1G09020	145	ATSNF4	-1.17	0.88	12.60	0.00	1.00
AT3G45890	AT3G45890	821	RUS1	-0.64	2.44	12.43	0.00	1.00
AT5G40950.1	AT5G40950	102	RPL27	1.43	0.28	12.36	0.00	1.00
AT1G68790.2	AT1G68790	254	CRWN3	0.87	0.92	11.98	0.00	1.00
AT1G22020	AT1G22020	1156	SHM6	1.31	0.16	11.92	0.00	1.00
AT5G67590.1	AT5G67590	130	FRO1	-0.67	2.59	11.89	0.00	1.00

Table 9: top DE genes TP1 vs TP17

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
AT1G33990.1	AT1G33990	240	ATMES14	-1.02	3.16	26.33	0.00	0.09
AT1G03330.1	AT1G03330	347	LSM2	-1.18	2.97	25.57	0.00	0.09
AT1G78650.1	AT1G78650	459	POLD3	1.76	0.28	24.27	0.00	0.09
AT1G76160.3	AT1G76160	346	sks5	-1.27	3.50	22.61	0.00	0.09
AT2G33550	AT2G33550	795	ASR3	-0.90	2.63	21.83	0.00	0.09
AT2G39570.1	AT2G39570	219	ACR9	-2.21	3.35	21.35	0.00	0.09
AT1G16510	AT1G16510	872	SAUR41	-1.91	1.92	20.94	0.00	0.09
AT5G09350	AT5G09350	2104	PI-4KBETA2	1.52	2.41	20.66	0.00	0.09
AT4G11660	AT4G11660	1006	AT-HSFB2B	1.59	1.75	20.47	0.00	0.09
AT2G39570.2	AT2G39570	325	ACR9	-2.39	4.23	20.59	0.00	0.09

Table 10: top DE genes TP1 vs TP19

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
AT5G39980	AT5G39980	2361	EMB3140	1.48	3.31	35.49	0.00	0.02
AT1G33990.1	AT1G33990	240	ATMES14	-1.16	3.16	33.83	0.00	0.02
AT1G70200.3	AT1G70200	305	HPE1	1.81	2.53	31.63	0.00	0.02
AT3G05330.1	AT3G05330	1348	ATN	1.98	2.11	30.25	0.00	0.02
ATCG00210	ATCG00210	90	YCF6	-1.43	10.91	29.24	0.00	0.02
AT4G37120.4	AT4G37120	378	SMP2	1.39	1.48	28.73	0.00	0.02
AT4G08920	AT4G08920	943	ATCRY1	-1.07	4.59	28.40	0.00	0.02
AT1G03330.1	AT1G03330	347	LSM2	-1.23	2.97	27.33	0.00	0.02
ATMG00280	ATMG00280	333	ORF110A	-1.23	11.60	26.77	0.00	0.02
AT1G69200.1	AT1G69200	653	FLN2	1.55	3.25	25.87	0.00	0.02

Table 11: top DE genes TP1 vs TP21

	GeneID	Length	Symbol	logFC	logCPM	F	PValue	FDR
	AT3G21750	1961	UGT71B1	1.05	5.57	47.13	0.00	0.00
	AT2G40610	375	ATEXP8	-2.18	2.36	39.70	0.00	0.01
	AT5G65480.1	530	CCI1	-1.28	4.21	31.80	0.00	0.02
	AT2G40610.2	637	ATEXP8	-2.19	3.71	30.99	0.00	0.02
	AT2G40610.1	313	ATEXP8	-2.17	2.49	29.22	0.00	0.02
	AT1G03130	956	PSAD-2	-1.78	8.16	23.67	0.00	0.08
	AT1G54130	1092	AT-RSH3	-1.05	4.26	22.21	0.00	0.09
	AT1G47370.1	149	RBA1	-1.40	1.23	21.57	0.00	0.09
	ATCG00270	1062	PSBD	-0.31	14.86	21.46	0.00	0.09
	AT3G19100	931	TAGK2	1.35	3.22	21.36	0.00	0.09

Table 12: top DE genes TP1 vs TP23

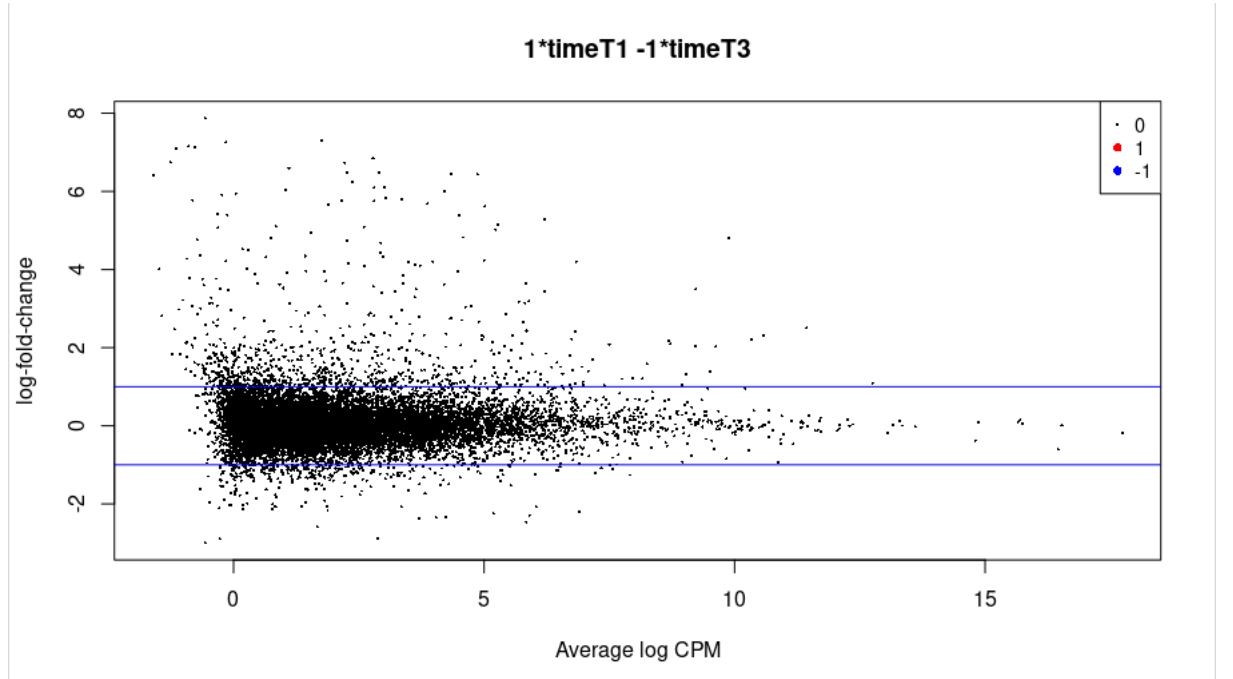


Figure 6: MD plot for TP1vsTP3

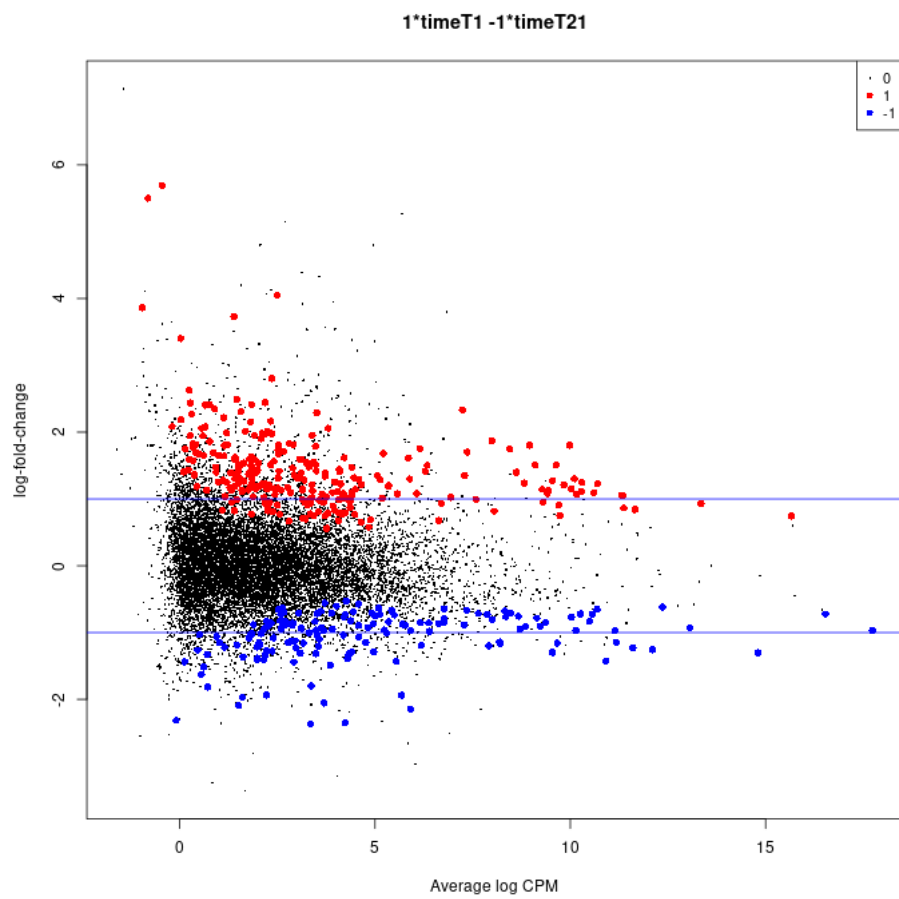


Figure 7: MD plot for TP1vsTP21

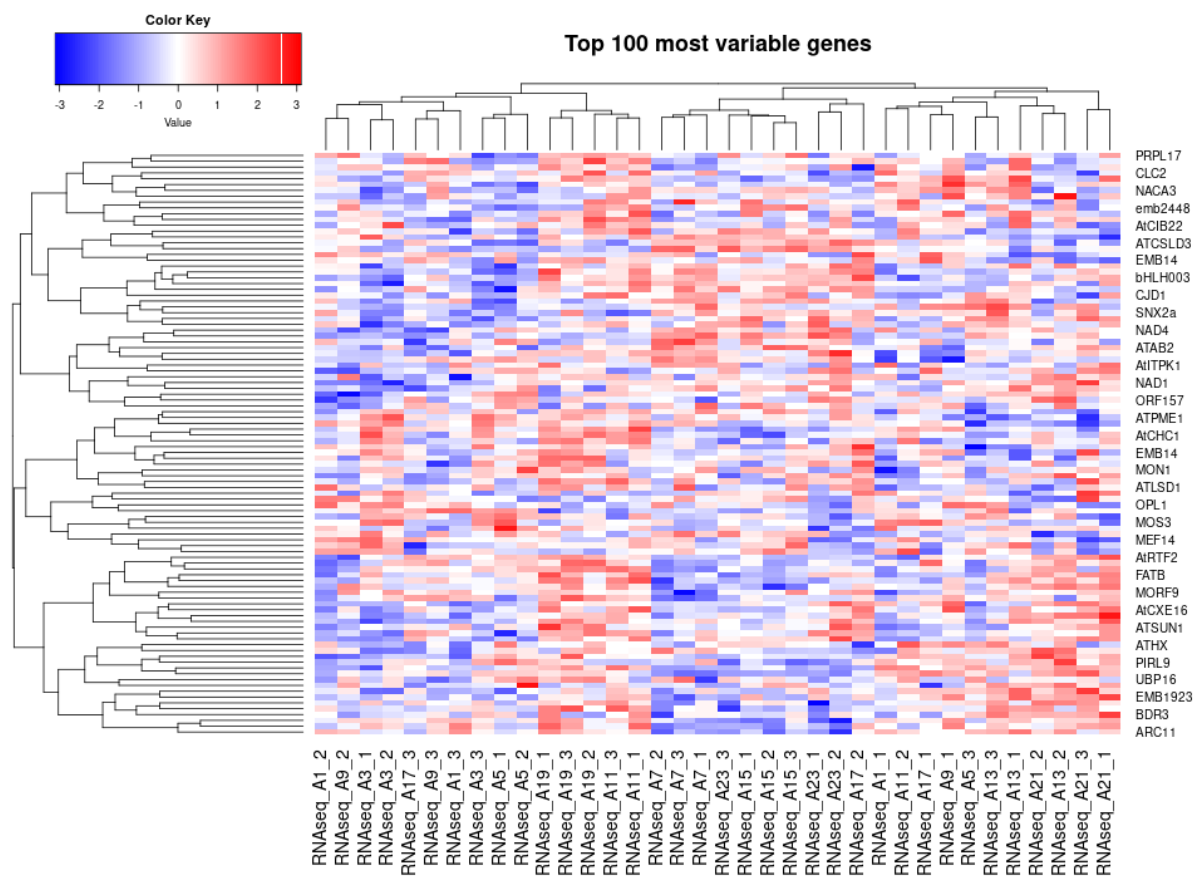
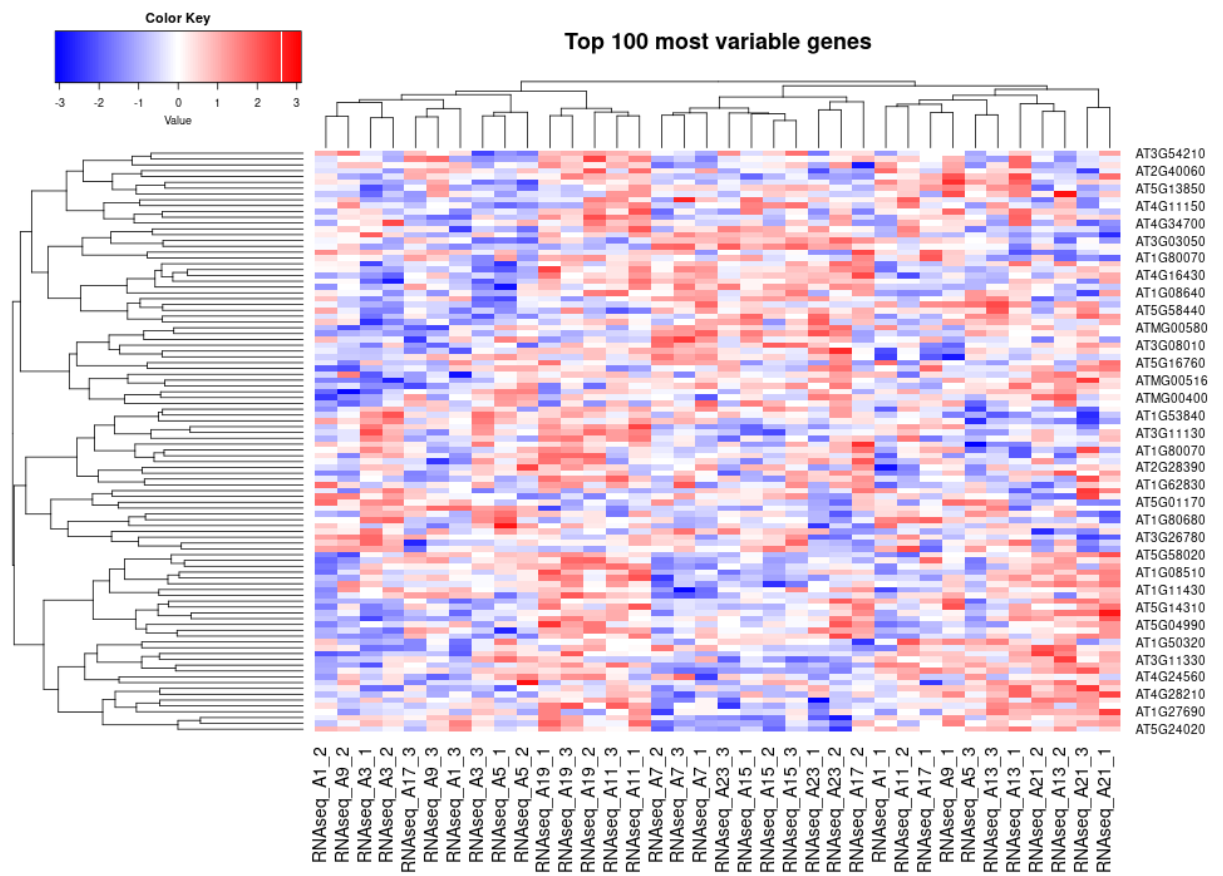



Figure 8: heatmap of the top 100 variable Symbols





Analysis Wizard
 DAVID Bioinformatics Resources 6.8, NIAID/NIH

[Home](#)
[Start Analysis](#)
[Shortcut to DAVID Tools](#)
[Technical Center](#)
[Downloads & APIs](#)
[Term of Service](#)
[About DAVID](#)
[About LHRI](#)

[Upload](#)
[List](#)
[Background](#)

Gene List Manager
 Select to limit annotations by one or more species [Help](#)

- Use All Species -
 Arabidopsis thaliana(8)
 Unknown(5)

Select Species

[List Manager](#)
[Help](#)
 List_1

Select List to:

Use

Rename

Remove

Combine

Show Gene List

[View Unmapped Ids](#)

Analysis Wizard

[Tell us how you like the tool](#)
[Contact us for questions](#)

☒ Step 1. Successfully submitted gene list
 Current Gene List: List_1
 Current Background: Arabidopsis thaliana

Step 2. Analyze above gene list with one of DAVID tools
 ↓ [Which DAVID tools to use?](#)

↻ Functional Annotation Tool

- Functional Annotation Clustering
- Functional Annotation Chart
- Functional Annotation Table

↻ Gene Functional Classification Tool

↻ Gene ID Conversion Tool

↻ Gene Name Batch Viewer

Figure 10: DAVID tool for functional annotation

	Pathway	N	Up	Down	P.Up
path:ath00750	Vitamin B6 metabolism	9.00	1.00	0.00	0.00
path:ath04712	Circadian rhythm - plant	19.00	0.00	1.00	1.00
path:ath00230	Purine metabolism	34.00	0.00	1.00	1.00
path:ath01240	Biosynthesis of cofactors	88.00	1.00	0.00	0.03
path:ath01100	Metabolic pathways	747.00	1.00	1.00	0.21
path:ath00942	Anthocyanin biosynthesis	1.00	0.00	0.00	1.00
path:ath00965	Betalain biosynthesis	1.00	0.00	0.00	1.00
path:ath00998	Biosynthesis of various secondary metabolites - part 2	1.00	0.00	0.00	1.00
path:ath00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	1.00	0.00	0.00	1.00
path:ath00901	Indole alkaloid biosynthesis	1.00	0.00	0.00	1.00
path:ath00902	Monoterpenoid biosynthesis	1.00	0.00	0.00	1.00
path:ath00232	Caffeine metabolism	2.00	0.00	0.00	1.00
path:ath00904	Diterpenoid biosynthesis	2.00	0.00	0.00	1.00
path:ath00604	Glycosphingolipid biosynthesis - ganglio series	2.00	0.00	0.00	1.00
path:ath00603	Glycosphingolipid biosynthesis - globo and isoglobo series	2.00	0.00	0.00	1.00
path:ath00785	Lipoic acid metabolism	2.00	0.00	0.00	1.00
path:ath00440	Phosphonate and phosphinate metabolism	2.00	0.00	0.00	1.00
path:ath00430	Taurine and hypotaurine metabolism	2.00	0.00	0.00	1.00
path:ath00780	Biotin metabolism	3.00	0.00	0.00	1.00
path:ath00905	Brassinosteroid biosynthesis	3.00	0.00	0.00	1.00
path:ath00660	C5-Branched dibasic acid metabolism	3.00	0.00	0.00	1.00
path:ath00944	Flavone and flavonol biosynthesis	3.00	0.00	0.00	1.00
path:ath00903	Limonene and pinene degradation	3.00	0.00	0.00	1.00
path:ath03450	Non-homologous end-joining	3.00	0.00	0.00	1.00
path:ath00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	3.00	0.00	0.00	1.00
path:ath00590	Arachidonic acid metabolism	4.00	0.00	0.00	1.00
path:ath00531	Glycosaminoglycan degradation	4.00	0.00	0.00	1.00
path:ath00909	Sesquiterpenoid and triterpenoid biosynthesis	4.00	0.00	0.00	1.00
path:ath04122	Sulfur relay system	4.00	0.00	0.00	1.00
path:ath00591	Linoleic acid metabolism	5.00	0.00	0.00	1.00

Table 13: top 30 pathways for T1vsTP7

	Pathway	N	Up	Down	P.Up	P
path:ath00942	Anthocyanin biosynthesis	1.00	0.00	0.00	1.00	
path:ath00965	Betalain biosynthesis	1.00	0.00	0.00	1.00	
path:ath00998	Biosynthesis of various secondary metabolites - part 2	1.00	0.00	0.00	1.00	
path:ath00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	1.00	0.00	0.00	1.00	
path:ath00901	Indole alkaloid biosynthesis	1.00	0.00	0.00	1.00	
path:ath00902	Monoterpenoid biosynthesis	1.00	0.00	0.00	1.00	
path:ath00232	Caffeine metabolism	2.00	0.00	0.00	1.00	
path:ath00904	Diterpenoid biosynthesis	2.00	0.00	0.00	1.00	
path:ath00604	Glycosphingolipid biosynthesis - ganglio series	2.00	0.00	0.00	1.00	
path:ath00603	Glycosphingolipid biosynthesis - globo and isoglobo series	2.00	0.00	0.00	1.00	
path:ath00785	Lipoic acid metabolism	2.00	0.00	0.00	1.00	
path:ath00440	Phosphonate and phosphinate metabolism	2.00	0.00	0.00	1.00	
path:ath00430	Taurine and hypotaurine metabolism	2.00	0.00	0.00	1.00	
path:ath00780	Biotin metabolism	3.00	0.00	0.00	1.00	
path:ath00905	Brassinosteroid biosynthesis	3.00	0.00	0.00	1.00	
path:ath00660	C5-Branched dibasic acid metabolism	3.00	0.00	0.00	1.00	
path:ath00944	Flavone and flavonol biosynthesis	3.00	0.00	0.00	1.00	
path:ath00903	Limonene and pinene degradation	3.00	0.00	0.00	1.00	
path:ath03450	Non-homologous end-joining	3.00	0.00	0.00	1.00	
path:ath00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	3.00	0.00	0.00	1.00	
path:ath00590	Arachidonic acid metabolism	4.00	0.00	0.00	1.00	
path:ath00531	Glycosaminoglycan degradation	4.00	0.00	0.00	1.00	
path:ath00909	Sesquiterpenoid and triterpenoid biosynthesis	4.00	0.00	0.00	1.00	
path:ath04122	Sulfur relay system	4.00	0.00	0.00	1.00	
path:ath00591	Linoleic acid metabolism	5.00	0.00	0.00	1.00	
path:ath00300	Lysine biosynthesis	5.00	0.00	0.00	1.00	
path:ath00600	Sphingolipid metabolism	5.00	0.00	0.00	1.00	
path:ath00730	Thiamine metabolism	5.00	0.00	0.00	1.00	
path:ath00650	Butanoate metabolism	6.00	0.00	0.00	1.00	
path:ath00950	Isoquinoline alkaloid biosynthesis	6.00	0.00	0.00	1.00	

Table 14: top 30 pathways for T1vsTP11

	Pathway	N	Up	Down	P.Up
path:ath00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	3.00	0.00	1.00	1.00
path:ath00941	Flavonoid biosynthesis	8.00	0.00	1.00	1.00
path:ath00940	Phenylpropanoid biosynthesis	35.00	0.00	1.00	1.00
path:ath01110	Biosynthesis of secondary metabolites	414.00	0.00	1.00	1.00
path:ath01100	Metabolic pathways	747.00	0.00	1.00	1.00
path:ath00942	Anthocyanin biosynthesis	1.00	0.00	0.00	1.00
path:ath00965	Betalain biosynthesis	1.00	0.00	0.00	1.00
path:ath00998	Biosynthesis of various secondary metabolites - part 2	1.00	0.00	0.00	1.00
path:ath00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	1.00	0.00	0.00	1.00
path:ath00901	Indole alkaloid biosynthesis	1.00	0.00	0.00	1.00
path:ath00902	Monoterpenoid biosynthesis	1.00	0.00	0.00	1.00
path:ath00232	Caffeine metabolism	2.00	0.00	0.00	1.00
path:ath00904	Diterpenoid biosynthesis	2.00	0.00	0.00	1.00
path:ath00604	Glycosphingolipid biosynthesis - ganglio series	2.00	0.00	0.00	1.00
path:ath00603	Glycosphingolipid biosynthesis - globo and isoglobo series	2.00	0.00	0.00	1.00
path:ath00785	Lipoic acid metabolism	2.00	0.00	0.00	1.00
path:ath00440	Phosphonate and phosphinate metabolism	2.00	0.00	0.00	1.00
path:ath00430	Taurine and hypotaurine metabolism	2.00	0.00	0.00	1.00
path:ath00780	Biotin metabolism	3.00	0.00	0.00	1.00
path:ath00905	Brassinosteroid biosynthesis	3.00	0.00	0.00	1.00
path:ath00660	C5-Branched dibasic acid metabolism	3.00	0.00	0.00	1.00
path:ath00944	Flavone and flavonol biosynthesis	3.00	0.00	0.00	1.00
path:ath00903	Limonene and pinene degradation	3.00	0.00	0.00	1.00
path:ath03450	Non-homologous end-joining	3.00	0.00	0.00	1.00
path:ath00590	Arachidonic acid metabolism	4.00	0.00	0.00	1.00
path:ath00531	Glycosaminoglycan degradation	4.00	0.00	0.00	1.00
path:ath00909	Sesquiterpenoid and triterpenoid biosynthesis	4.00	0.00	0.00	1.00
path:ath04122	Sulfur relay system	4.00	0.00	0.00	1.00
path:ath00591	Linoleic acid metabolism	5.00	0.00	0.00	1.00
path:ath00300	Lysine biosynthesis	5.00	0.00	0.00	1.00

Table 15: top 30 pathways for T1vsTP13

	Pathway	N	Up	Down	P.Up
path:ath00195	Photosynthesis	28.00	0.00	2.00	1.00
path:ath00460	Cyanoamino acid metabolism	13.00	1.00	0.00	0.00
path:ath00940	Phenylpropanoid biosynthesis	35.00	1.00	0.00	0.01
path:ath00500	Starch and sucrose metabolism	47.00	1.00	0.00	0.01
path:ath00230	Purine metabolism	34.00	0.00	1.00	1.00
path:ath01110	Biosynthesis of secondary metabolites	414.00	1.00	0.00	0.12
path:ath01100	Metabolic pathways	747.00	1.00	2.00	0.21
path:ath00942	Anthocyanin biosynthesis	1.00	0.00	0.00	1.00
path:ath00965	Betalain biosynthesis	1.00	0.00	0.00	1.00
path:ath00998	Biosynthesis of various secondary metabolites - part 2	1.00	0.00	0.00	1.00
path:ath00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	1.00	0.00	0.00	1.00
path:ath00901	Indole alkaloid biosynthesis	1.00	0.00	0.00	1.00
path:ath00902	Monoterpenoid biosynthesis	1.00	0.00	0.00	1.00
path:ath00232	Caffeine metabolism	2.00	0.00	0.00	1.00
path:ath00904	Diterpenoid biosynthesis	2.00	0.00	0.00	1.00
path:ath00604	Glycosphingolipid biosynthesis - ganglio series	2.00	0.00	0.00	1.00
path:ath00603	Glycosphingolipid biosynthesis - globo and isoglobo series	2.00	0.00	0.00	1.00
path:ath00785	Lipoic acid metabolism	2.00	0.00	0.00	1.00
path:ath00440	Phosphonate and phosphinate metabolism	2.00	0.00	0.00	1.00
path:ath00430	Taurine and hypotaurine metabolism	2.00	0.00	0.00	1.00
path:ath00780	Biotin metabolism	3.00	0.00	0.00	1.00
path:ath00905	Brassinosteroid biosynthesis	3.00	0.00	0.00	1.00
path:ath00660	C5-Branched dibasic acid metabolism	3.00	0.00	0.00	1.00
path:ath00944	Flavone and flavonol biosynthesis	3.00	0.00	0.00	1.00
path:ath00903	Limonene and pinene degradation	3.00	0.00	0.00	1.00
path:ath03450	Non-homologous end-joining	3.00	0.00	0.00	1.00
path:ath00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	3.00	0.00	0.00	1.00
path:ath00590	Arachidonic acid metabolism	4.00	0.00	0.00	1.00
path:ath00531	Glycosaminoglycan degradation	4.00	0.00	0.00	1.00
path:ath00909	Sesquiterpenoid and triterpenoid biosynthesis	4.00	0.00	0.00	1.00

Table 16: top 30 pathways for T1vsTP15

	Pathway	N	Up	Down	P.Up
path:ath00604	Glycosphingolipid biosynthesis - ganglio series	2.00	0.00	1.00	1.00
path:ath00603	Glycosphingolipid biosynthesis - globo and isoglobo series	2.00	0.00	1.00	1.00
path:ath00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	3.00	0.00	1.00	1.00
path:ath00531	Glycosaminoglycan degradation	4.00	0.00	1.00	1.00
path:ath04141	Protein processing in endoplasmic reticulum	74.00	1.00	3.00	0.78
path:ath00511	Other glycan degradation	6.00	0.00	1.00	1.00
path:ath00941	Flavonoid biosynthesis	8.00	0.00	1.00	1.00
path:ath00591	Linoleic acid metabolism	5.00	1.00	0.00	0.10
path:ath00130	Ubiquinone and other terpenoid-quinone biosynthesis	9.00	0.00	1.00	1.00
path:ath01100	Metabolic pathways	747.00	3.00	13.00	1.00
path:ath00450	Selenocompound metabolism	10.00	0.00	1.00	1.00
path:ath00270	Cysteine and methionine metabolism	49.00	1.00	2.00	0.63
path:ath00520	Amino sugar and nucleotide sugar metabolism	52.00	0.00	2.00	1.00
path:ath01200	Carbon metabolism	105.00	0.00	3.00	1.00
path:ath02010	ABC transporters	7.00	1.00	0.00	0.13
path:ath00460	Cyanoamino acid metabolism	13.00	0.00	1.00	1.00
path:ath03410	Base excision repair	8.00	1.00	0.00	0.15
path:ath00280	Valine, leucine and isoleucine degradation	14.00	0.00	1.00	1.00
path:ath00513	Various types of N-glycan biosynthesis	15.00	0.00	1.00	1.00
path:ath00920	Sulfur metabolism	16.00	0.00	1.00	1.00
path:ath00061	Fatty acid biosynthesis	18.00	0.00	1.00	1.00
path:ath00400	Phenylalanine, tyrosine and tryptophan biosynthesis	18.00	0.00	1.00	1.00
path:ath04712	Circadian rhythm - plant	19.00	0.00	1.00	1.00
path:ath04075	Plant hormone signal transduction	131.00	0.00	3.00	1.00
path:ath01110	Biosynthesis of secondary metabolites	414.00	0.00	7.00	1.00
path:ath00030	Pentose phosphate pathway	23.00	0.00	1.00	1.00
path:ath00561	Glycerolipid metabolism	25.00	0.00	1.00	1.00
path:ath03030	DNA replication	15.00	1.00	0.00	0.27
path:ath04130	SNARE interactions in vesicular transport	15.00	1.00	0.00	0.27
path:ath01212	Fatty acid metabolism	28.00	0.00	1.00	1.00

Table 17: top 30 pathways for T1vsTP21

	Pathway	N	Up	Down	P.Up	P
path:ath00942	Anthocyanin biosynthesis	1.00	0.00	0.00	1.00	
path:ath00965	Betalain biosynthesis	1.00	0.00	0.00	1.00	
path:ath00998	Biosynthesis of various secondary metabolites - part 2	1.00	0.00	0.00	1.00	
path:ath00563	Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	1.00	0.00	0.00	1.00	
path:ath00901	Indole alkaloid biosynthesis	1.00	0.00	0.00	1.00	
path:ath00902	Monoterpenoid biosynthesis	1.00	0.00	0.00	1.00	
path:ath00232	Caffeine metabolism	2.00	0.00	0.00	1.00	
path:ath00904	Diterpenoid biosynthesis	2.00	0.00	0.00	1.00	
path:ath00604	Glycosphingolipid biosynthesis - ganglio series	2.00	0.00	0.00	1.00	
path:ath00603	Glycosphingolipid biosynthesis - globo and isoglobo series	2.00	0.00	0.00	1.00	
path:ath00785	Lipoic acid metabolism	2.00	0.00	0.00	1.00	
path:ath00440	Phosphonate and phosphinate metabolism	2.00	0.00	0.00	1.00	
path:ath00430	Taurine and hypotaurine metabolism	2.00	0.00	0.00	1.00	
path:ath00780	Biotin metabolism	3.00	0.00	0.00	1.00	
path:ath00905	Brassinosteroid biosynthesis	3.00	0.00	0.00	1.00	
path:ath00660	C5-Branched dibasic acid metabolism	3.00	0.00	0.00	1.00	
path:ath00944	Flavone and flavonol biosynthesis	3.00	0.00	0.00	1.00	
path:ath00903	Limonene and pinene degradation	3.00	0.00	0.00	1.00	
path:ath03450	Non-homologous end-joining	3.00	0.00	0.00	1.00	
path:ath00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	3.00	0.00	0.00	1.00	
path:ath00590	Arachidonic acid metabolism	4.00	0.00	0.00	1.00	
path:ath00531	Glycosaminoglycan degradation	4.00	0.00	0.00	1.00	
path:ath00909	Sesquiterpenoid and triterpenoid biosynthesis	4.00	0.00	0.00	1.00	
path:ath04122	Sulfur relay system	4.00	0.00	0.00	1.00	
path:ath00591	Linoleic acid metabolism	5.00	0.00	0.00	1.00	
path:ath00300	Lysine biosynthesis	5.00	0.00	0.00	1.00	
path:ath00600	Sphingolipid metabolism	5.00	0.00	0.00	1.00	
path:ath00730	Thiamine metabolism	5.00	0.00	0.00	1.00	
path:ath00650	Butanoate metabolism	6.00	0.00	0.00	1.00	
path:ath00950	Isoquinoline alkaloid biosynthesis	6.00	0.00	0.00	1.00	

Table 18: top 30 pathways for T1vsTP23