

Transcriptome analysis using RNA-Seq

유전체 빅데이터 분석 예비 전문가 과정
실무 프로젝트 리포트

8기 이택규

2020.09.02 ~ 04

Reference Information

- 종 : Arabidopsis thaliana
- 데이터베이스 : Ensembl (release 44)
- Genome size : ~ 135 Mbp
- Gene model
 - No. All genes : 32833 (제공된 GTF 기준)
 - No. Masked genes : 5205 (제공된 GTF 기준)



RNA-Seq Information

Tissue	Treatment	Time point	Sample Name	No. Raw Reads
Shoot	Control	6h	Shoot-Control-6h	8135267
Shoot	Control	12h	Shoot-Control-12h	7387742
Shoot	Control	24h	Shoot-Control-24h	7443630
Shoot	PA01	6h	Shoot-PA01-6h	7245643
Shoot	PA01	12h	Shoot-PA01-12h	7763482
Shoot	PA01	24h	Shoot-PA01-24h	7213743

- Inner distance : 250 bp (standard deviation : 50 bp)
- Library type : fr-unstranded
- Library Prep Kit : TruSeq stranded mRNA Library Prep kit
 - * TruSeq Universal Adapter: 5 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT 3
 - * TruSeq Indexed Adapter: 5 GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-NNNNNN-ATCTCGTATGCCGTCTTCTGCTTG 3

Sample Info.

About Arabidopsis thaliana (PRJNA361001)

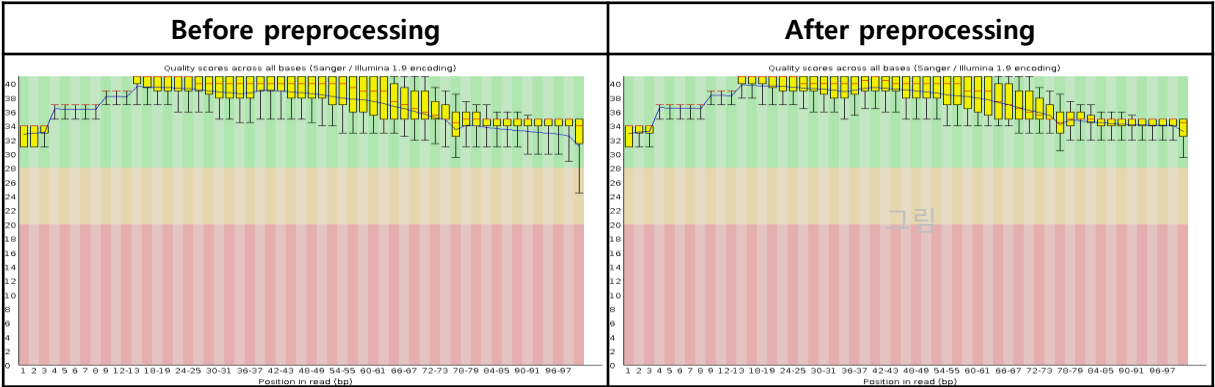
Arabidopsis thaliana is a small flowering plant that is widely used as a model organism in plant biology. Arabidopsis is a member of the mustard (Brassicaceae) family, which includes cultivated species such as cabbage and radish. Arabidopsis is not of major agronomic significance, but it offers important advantages for basic research in genetics and molecular biology. Arabidopsis thaliana has a genome size of ~135 Mbp, and a haploid chromosome number of 5.

Reads count & Expressed gene count Table

Sample Name	Read Preprocessing			Reference mapping (Alignment)		Expressed Genes
	Raw Reads	Clean Reads	Rate, % (per Raw)	Mapped Reads	Rate, % (per Clean)	RPKM >= 0.3
Shoot-Control-6h	8135267	7770134	95.51	7622651	97.8	19151
Shoot-Control-12h	7387742	7068967	95.69	6922417	97.7	19122
Shoot-Control-24h	7443630	7117908	95.62	6985858	98.0	19218
Shoot-PA01-6h	7245643	6917874	95.48	6783048	97.9	19398
Shoot-PA01-12h	7763482	7439003	95.82	7295714	97.9	19518
Shoot-PA01-24h	7213743	6901930	95.68	6765528	97.9	19645

- Raw Reads를 cutadapt 실행. TruSeq Universal Adapter, TruSeq Indexed Adapter 를 제거.
- 이후, SLIDINGWINDOW size는 4bp, average quality는 20, MINLEN 은 50bp로 설정하여 trimming을 진행.
- Library type이 fr-unstranded이기 때문에, Expressed Genes이 RPKM >= 0.3 인 것들을 in vivo 상에서 발현하고있다고 정의.
- Align 정보와 trimming정보는 추가 page에 기재.(Fig.01)

Per base sequence quality (Shoot-Control-6h R1)



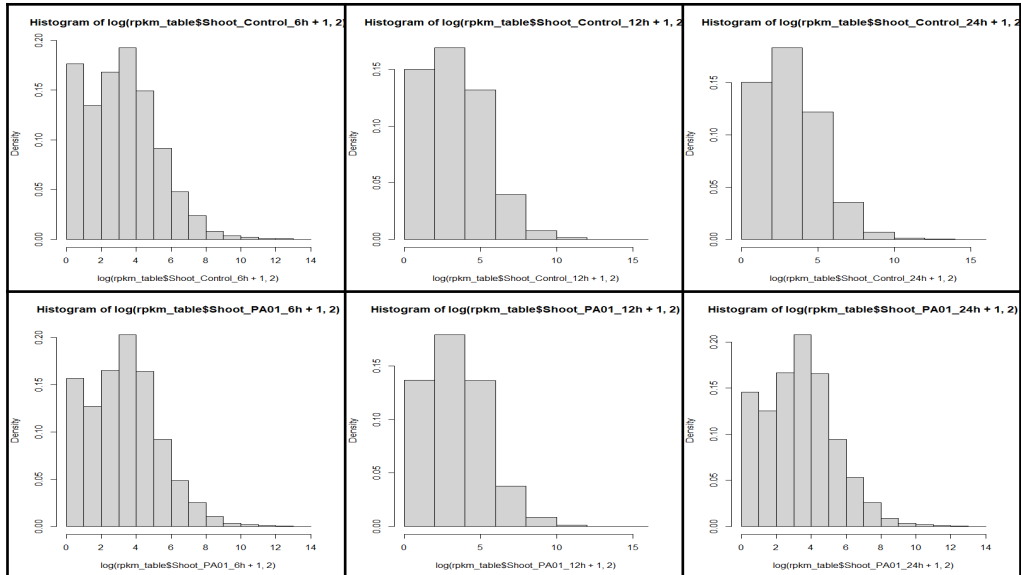
- trimming 후 Shoot-Control-6h R1 에대한 quality 정보.
- 서열 끝부분에서 base quality 의 변화.
- GC contents는 추가 page에 기재.(Fig.02)

Expressed Gene Selection

- No. Total genes : 32823 (cuffnorm 결과 기준)
- No. Selected genes : 20863 (적어도 한 샘플에서라도 RPKM ≥ 0.3 인 유전자만 선택)

Histogram

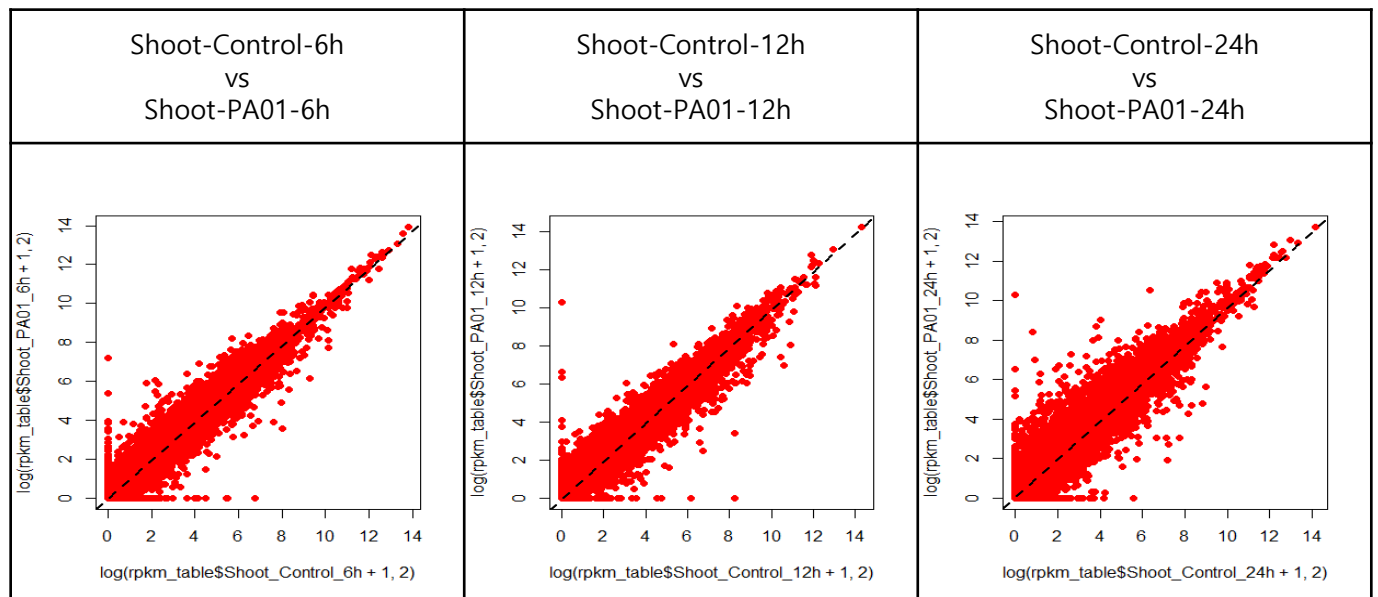
- Input genes : Selected genes
- Value : $\log_2(\text{RPKM} + 1.0)$



- 각 sample에서 발현량을 확인하기 위한 histogram.
- 각 sample에서 RPKM ≥ 0.3 인 gene을 채택.
- 각 sample에서 발현량 확인.
- density 정보는 추가 page에 기재.(Fig.03)
- Control, Experiment의 6h와 12h sample에서 발현량이 비슷한 양상을 가지고 있음.
- 24h sample에서는 발현량의 양상에서 차이 발생.
- PA01처리 이후 24h에서 발현량의 차이 발생.

Scatter plot

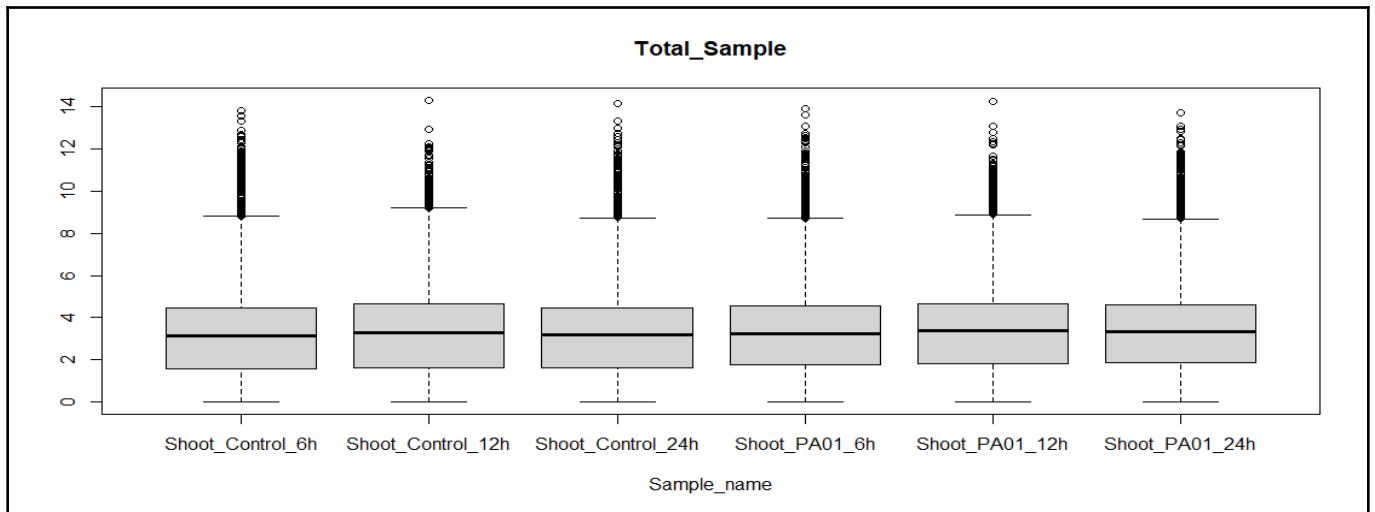
- Input genes : Selected genes
- Value : $\log_2(\text{RPKM} + 1.0)$



- 동일 시간에서 sample간 scatter plot의 결과.
- $\text{RPKM} \geq 0.3$ 인 gene을 채택.
- sample간 발현양상을 대략적으로 확인가능.
- 6h sample 간 비교시, 발현량의 차이가 있지만, 발현 양상이 유사하기 때문에 얇게(narrow)하게 분포하는 양상을 가짐.
- 12h sample간 비교시, 6h sample과는 약간의 차이 존재, 발현 양상이 유사하지 않은 gene이 약간 증가.
- 24h sample간 비교시, 발현량의 차이가 6h,12h sample과는 달리 분포가 넓게(wide)하게 나타남.
- 결론적으로 scatter plot만 분석해보았을때, 각 gene의 발현이 6h,12h에는 많은 차이가 존재하지 않는다는 것을 확인. 하지만 24h에서는 발현량이 달라졌다는 것을 대략적으로 확인가능.
- 동일 group 다른 시간대 추가 분석자료 추가 page기재.(Fig.04)

Box plot

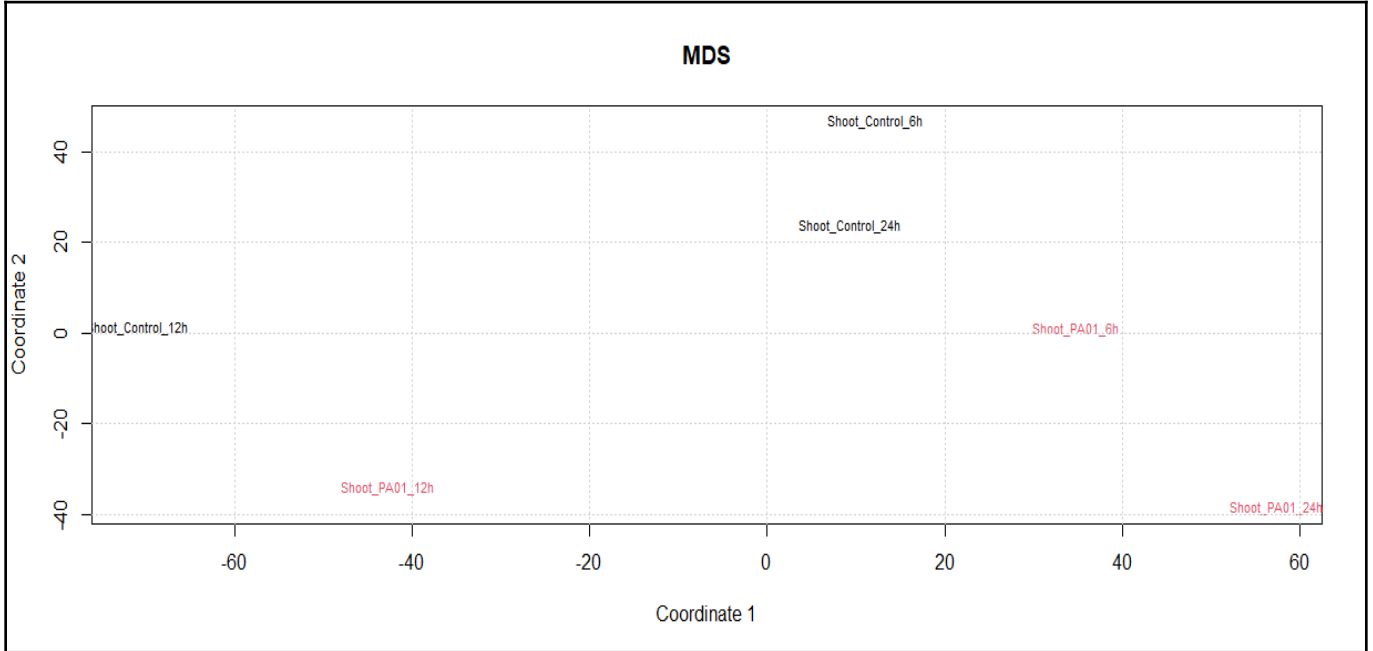
- Input genes : Selected genes
- Value : $\log_2(\text{RPKM} + 1.0)$



- boxplot으로 전체 sample에 대한 시각화.
- $\text{RPKM} \geq 0.3$ 인 gene을 채택.

MDS

- Input genes : Selected genes
- Value : $\log_2(\text{RPKM} + 1.0)$
- Method for distance calculation : "euclidean"

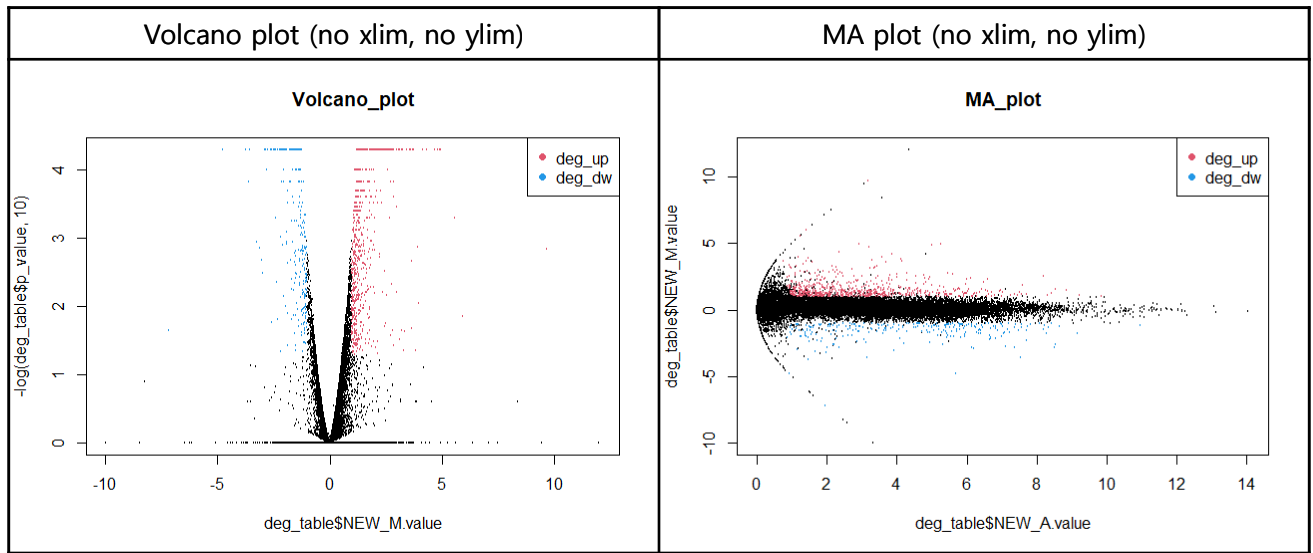


- MDS로 전체 sample에 대한 시각화.
- $\text{RPKM} \geq 0.3$ 인 gene을 채택.
- Euclidean distance를 통해 거리 계산.
- MDS결과를 통해, 각 sample간의 거리 확인.
- Control과 Experiment group에서 group안의 sample의 위치를 확인한 결과, 공통적으로 group 안에서 12h sample의 거리가 다른 두 sample 보다 먼 것을 확인.
- 이를 통해 12h에서 발현양의 차이 유추가능.

DEG Selection

- No. Total genes : 32823 (Cuffdiff 결과 기준)
- Average New A-value : 2.1(2.115738) (소수점 2자리에서 반올림)
- Average New M-value : 0.1(0.1102173) (소수점 2자리에서 반올림)
- No. Selected DEG : 876
 - Up regulated DEG : 628
 - Down regulated DEG : 248

Volcano plot & MA plot - Cuffdiff



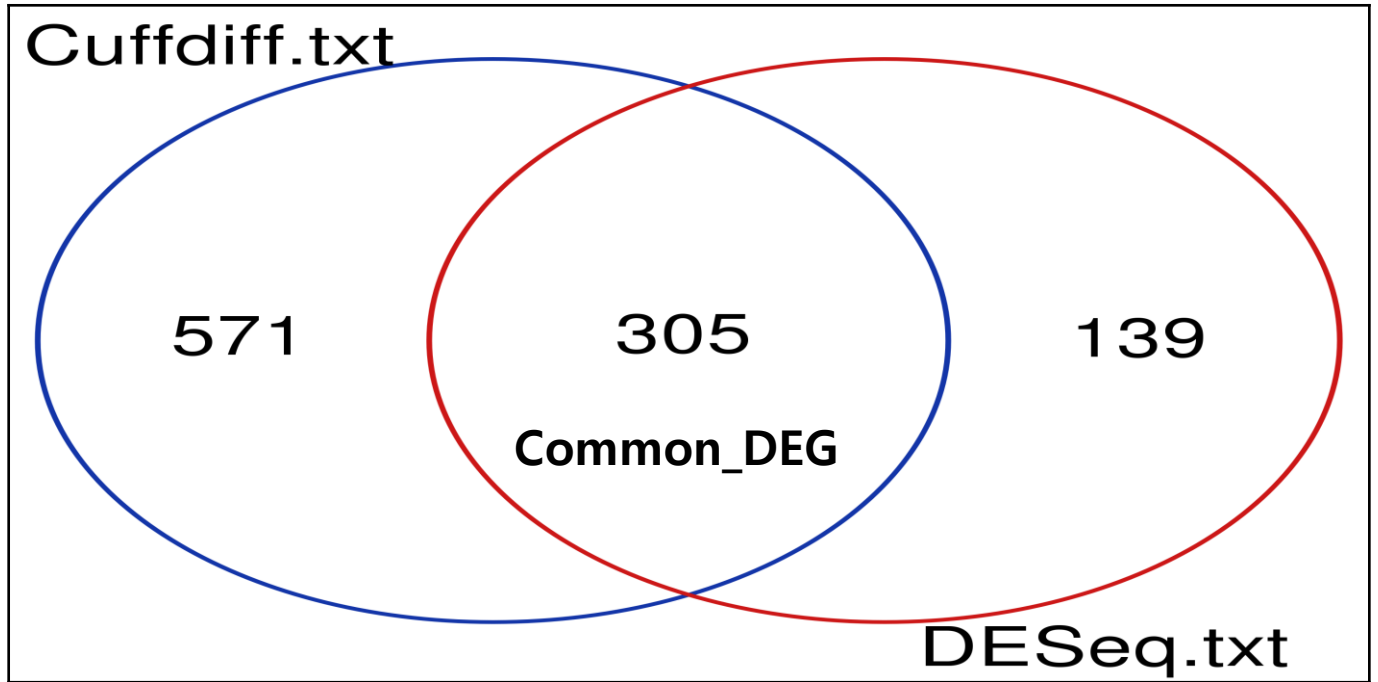
- Cuffdiff의 결과인 DEG에 대한 Volcano plot & MA plot.
- $\log_2(\text{value}_2 + 0.1) - \log_2(\text{value}_1 + 0.1)$ 을 통해 New M-value 계산.
- $1/2(\log_2(\text{value}_2 + 1) + \log_2(\text{value}_1 + 1))$ 을 통해 NEW A-value 계산.
- DEG selection은 $\text{abs}(\text{New } \log_2(\text{FC})) \geq 1$ & $p\text{-value} < 0.05$.
- DEG selection후, Control대비 case에서 up regulate 된 DEG는 628개.
- DEG selection후, Control대비 case에서 down regulate 된 DEG는 248개.
- TCC결과 추가 page기재.(Fig.05)

Common DEG

- No. Selected DEG from Cuffdiff : 876
- No. Selected DEG from TCC : 444

Venn diagram

참고 : <http://bioinformatics.psb.ugent.be/webtools/Venn>



- DEG가 서로 다르게 선별됨.
- 따라서 서로 다른 방법으로 선별된 DEG중에서 Common DEG를 선별하여 사용해야함.
- Common DEG에 대한 정보 추가 page 기재.(Fig06)
- 추가적인 Common DEG 선별위해 edgeR 사용. 추가 page 기재(Fig.07)

DAVID (Optional)

참고 : <https://david.ncifcrf.gov/summary.jsp>

- Input genes : Common DEG

Upload

List

Background

Upload Gene List

[Demolist 1](#) [Demolist 2](#)

Upload **Common DEG**

Step 1: Enter Gene List

A: Paste a list

AT

AT

AT

AT

AT

Clear

Or

B: Choose From a File

파일 선택

선택된 파일 없음

☐ Multi-List File

Step 2: Select Identifier

TAIR_ID

Step 3: List Type

Gene List

Background

Step 4: Submit List

Submit List

Upload

List

Background

Gene List Manager

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

- Use All Species -

Arabidopsis thaliana

Unknown

Select Species

List Manager [Help](#)

List_1

Select List to:

Use

Rename

Remove

Combine

Show Gene List

[View Unmapped Ids](#)

Annotation Summary Results

Current Gene List: List_1

Current Background: Arabidopsis thaliana

DAVID IDs

Check Defaults

Functional_Categories

Gene_Ontology

General_Annotations

Literature

Main_Accessions

Pathways

Protein_Domains

Protein_Interactions

Tissue_Expression

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Functional Annotation Clustering

Functional Annotation Chart

Functional Annotation Table

Enrichment score : 2.78		Gene_ID			
		TAIR_ID	GENE NAME	Related Genes	Species
wall-associated kinase 2 (WAK2)		AT5G07475	Cupredoxin superfamily protein(AT5G07475)	RG	Arabidopsis thaliana
PLC-like phosphodiesterases superfamily protein(GDPD2)		AT1G26410	FAD-binding Berberine family protein(AT1G26410)	RG	Arabidopsis thaliana
glutathione S-transferase 7(GSTF7)		AT1G26420	FAD-binding Berberine family protein(AT1G26420)	RG	Arabidopsis thaliana
glutathione peroxidase 1(GPX1)		AT5G41080	PLC-like phosphodiesterases superfamily protein(GDPD2)	RG	Arabidopsis thaliana
glutathione S-transferase tau 9(GSTU9)		AT3G62950	Thioredoxin superfamily protein(AT3G62950)	RG	Arabidopsis thaliana
FAD-binding Berberine family protein(AT1G26410)		AT4G15660	Thioredoxin superfamily protein(AT4G15660)	RG	Arabidopsis thaliana
Cupredoxin superfamily protein(AT5G07475)		AT4G15680	Thioredoxin superfamily protein(AT4G15680)	RG	Arabidopsis thaliana
acyl-CoA oxidase 4(ACX4)		AT4G15700	Thioredoxin superfamily protein(AT4G15700)	RG	Arabidopsis thaliana
early nodulin-like protein 19(ENODL19)		AT5G18600	Thioredoxin superfamily protein(AT5G18600)	RG	Arabidopsis thaliana
isovaleryl-CoA-dehydrogenase(IVD)		AT5G14070	Thioredoxin superfamily protein(ROXY2)	RG	Arabidopsis thaliana
FAD-binding Berberine family protein(AT1G26420)		AT3G51840	acyl-CoA oxidase 4(ACX4)	RG	Arabidopsis thaliana
Thioredoxin superfamily protein(ROXY2)		AT4G12880	early nodulin-like protein 19(ENODL19)	RG	Arabidopsis thaliana
Thioredoxin superfamily protein(AT3G62950)		AT1G02920	glutathione S-transferase 7(GSTF7)	RG	Arabidopsis thaliana
Thioredoxin superfamily protein(AT5G18600)		AT3G62480	glutathione S-transferase tau 9(GSTU9)	RG	Arabidopsis thaliana
Thioredoxin superfamily protein(AT4G15700)		AT2G25080	glutathione peroxidase 1(GPX1)	RG	Arabidopsis thaliana
Thioredoxin superfamily protein(AT4G15660)		AT3G45300	isovaleryl-CoA-dehydrogenase(IVD)	RG	Arabidopsis thaliana
Thioredoxin superfamily protein(AT4G15680)		AT1G21270	wall-associated kinase 2(WAK2)	RG	Arabidopsis thaliana

결론

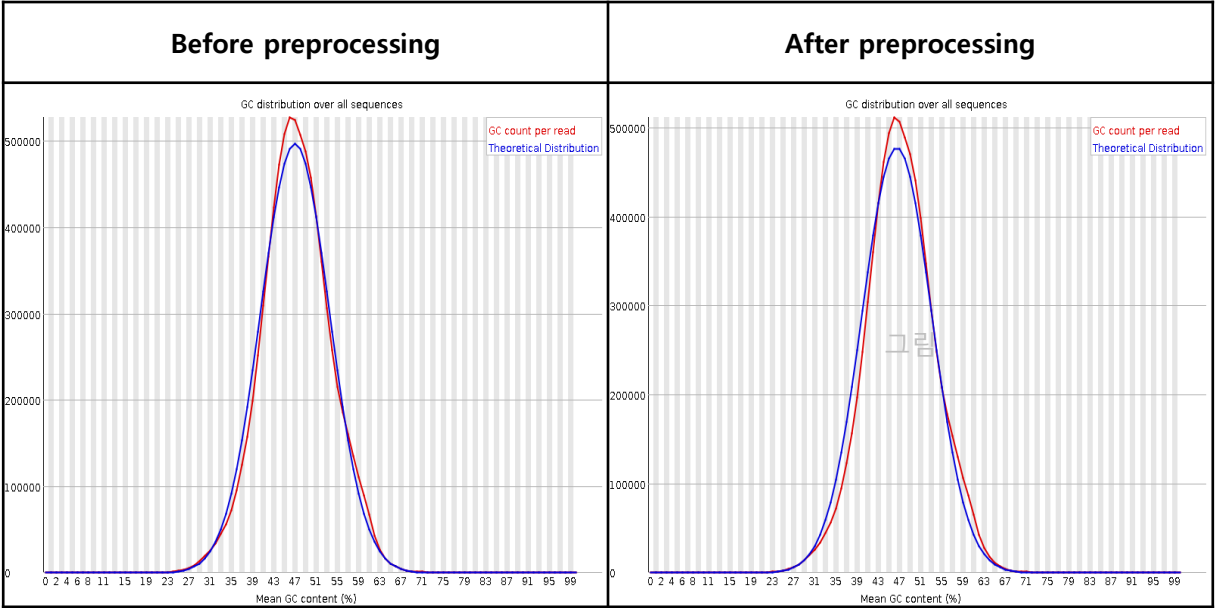
- *Arabidopsis thaliana*의 Shoot tissue를 각각 6h, 12h, 24h 로 Control group과 Experiment(PA01) group으로 나누었다. 분석에 필요한 data는 Ensembl data를 사용하였다. 분석은 rawdata로부터 quality trimmin과정을 거쳐 quantification mode로 분석을 진행하였다. 그리고 HTSeq-count에서 TCC를 deseq으로 진행하였다.
- 결과는 Cuffnorm의 결과로 생성된 genes.fpkm_table을 사용하여 발현량을 통해 비교하였다. 각 sample에 대한 발현량 확인은 histogram, sample간 발현량 비교는 scatter plot, 전체 sample의 대략적인 발현량 정보는 boxplot과 MDS를 통해 확인하였다. control group과 experiment group간의 비교를 확인하기 위하여 scatter plot을 통해 결과를 관찰 하였다. 우선 동일 시간끼리 비교한 결과를 통해서는, Shoot-Control-24h와 Shoot-PA01-24h에서 발현량의 차이가 보였다. 하지만 MDS로 판단했을 경우에는, Shoot-control-12h와 Shoot-PA01-12h가 각 group에서 거리가 가장 멀리 떨어져 있는 것으로 확인되었다. 따라서 결과 해석이 모호 하므로, Shoot-control-12h와 Shoot-PA01-12h를 group내에서 다른 sample들과 비교하여 추가적인 분석을 진행하였다. 추가 분석은 Control과 Experiment group내에서 Shoot-Control-12h와 Shoot-Control-6h , Shoot-Control-12h와 , Shoot-Control-24h를 비교하였고, Shoot-PA01-12h와 Shoot-PA01-6h, Shoot-PA01-12h와 Shoot-PA01-24h를 비교하였다. 또한 추가적으로 Shoot-Control-6h와 Shoot-Control-24h를 비교하였고 Shoot-PA01-6h와 Shoot-PA01-24h를 비교하였다. 결과적으로 각 group에서 12h sample의 발현량이 다른 sample과 차이가 남을 대략적으로 확인하였으며, MDS분석 결과를 증명하였다. 발현량의 변화가 특정 gene의 발현 유무에 따라 변했다는 것을 확인하기 위해 DEG(differentially expressed gene)을 확인하였다. Cuffdiff를 통해 나온 결과로 DEG selectio을 진행하였고, New A-value와 New M-value를 통하여 발현되고 있는 DEG가 Control대비 experiment group에서 얼마정도 up regulate되는지, 혹은 down regulate 되고 있는지 확인해보았다. 분석 결과로 총 876개의 DEG중 628개의 DEG가 experiment상에서 up regulate되고 있었고, 248개의 DEG가 experiment상에서 down regulate되고 있음을 확인하였다. 결과적으로 발현량의 차이가 DEG 조절에 의해 나타났다고 말할 수 있다. 하지만 deseq으로 분석한 결과 DEG는 444개로 Cuffdiff의 결과와 차이가 있었다. 추가 분석 결과로 총 444개의 DEG중 204개의 DEG가 experimen상에서 up regulate되고 있었고, 240개의 DEG가 experimen상에서 down regulate되고 있음을 확인하였다. 따라서 여러가지 분석 방법을 이용해 공통되는 DEG를 발견하고, 발견한 DEG를 대상으로 sample내에서 common DEG가 발현이 되는지 안되는지, 추가적인 분석을 통해서 확인을 해 보아야 한다. 또한 scatter plot 추가분석 결과를 통해 발현량의 차이가 PA01처리 후 나타난다 라고 단정 짓기가 제한된다고 할 수 있다. 이유는 같은 group에서 시간의 차이에 따라 발현량의 차이를 보였으므로, 발현량의 차이가 진정 다른 조건을 가진 group사이에서 발현되는 DEG의 차이인지, 아니면 같은 group내에서 시간에 따른 변화인지 확실하지 않기 때문이다.
- 본 분석에서 Control group과 Experiment group간의 차이를 RPKM값과 DEG를 통해 확인하였다. 분석 결과로 발현량과 DEG의 차이가 나타남을 확인하였으나, 이번 실험에서 정말로 약물처리에 의해서 나타난 결과인지 확신하기는 제한된다. 분석을 통해 약물이라는 조건 외에 발현량에 영향을 미치는 요인을 확인 했기 때문이다. 따라서 추가적인 분석시, 약물외 다른 조건을 명확하게 규명한 상태에서 분석을 진행한다면, 약물처리와 DEG, 발현량의 관계를 좀더 깊게 알 수 있을 것이다.

Fig.01

Trimming Info		
Shoot-Control-6h	Shoot-Control-12h	Shoot-Control-24h
Input Read Pairs: 8135267 Both Surviving Reads: 7770134 Both Surviving Read Percent: 95.51 Forward Only Surviving Reads: 155803 Forward Only Surviving Read Percent: 1.92 Reverse Only Surviving Reads: 148125 Reverse Only Surviving Read Percent: 1.82 Dropped Reads: 61205 Dropped Read Percent: 0.75	Input Read Pairs: 7387742 Both Surviving Reads: 7068967 Both Surviving Read Percent: 95.69 Forward Only Surviving Reads: 140907 Forward Only Surviving Read Percent: 1.91 Reverse Only Surviving Reads: 126159 Reverse Only Surviving Read Percent: 1.71 Dropped Reads: 51709 Dropped Read Percent: 0.70	Input Read Pairs: 7443630 Both Surviving Reads: 7117908 Both Surviving Read Percent: 95.62 Forward Only Surviving Reads: 141664 Forward Only Surviving Read Percent: 1.90 Reverse Only Surviving Reads: 129002 Reverse Only Surviving Read Percent: 1.73 Dropped Reads: 55056 Dropped Read Percent: 0.74
Shoot-PA01-6h	Shoot-PA01-12h	Shoot-PA01-24h
Input Read Pairs: 7245643 Both Surviving Reads: 6917874 Both Surviving Read Percent: 95.48 Forward Only Surviving Reads: 145991 Forward Only Surviving Read Percent: 2.01 Reverse Only Surviving Reads: 128037 Reverse Only Surviving Read Percent: 1.77 Dropped Reads: 53741 Dropped Read Percent: 0.74	Input Read Pairs: 7763482 Both Surviving Reads: 7439003 Both Surviving Read Percent: 95.82 Forward Only Surviving Reads: 140077 Forward Only Surviving Read Percent: 1.80 Reverse Only Surviving Reads: 128388 Reverse Only Surviving Read Percent: 1.65 Dropped Reads: 56014 Dropped Read Percent: 0.72	Input Read Pairs: 7213743 Both Surviving Reads: 6901930 Both Surviving Read Percent: 95.68 Forward Only Surviving Reads: 126924 Forward Only Surviving Read Percent: 1.76 Reverse Only Surviving Reads: 134237 Reverse Only Surviving Read Percent: 1.86 Dropped Reads: 50652 Dropped Read Percent: 0.70

Align Info		
Shoot-Control-6h	Shoot-Control-12h	Shoot-Control-24h
Left reads: Input : 7770134 Mapped : 7683225 (98.9% of input) of these: 184736 (2.4% have multiple alignments (32 have >20) Right reads: Input : 7770134 Mapped : 7678096 (98.8% of input) of these: 184565 (2.4% have multiple alignments (32 have >20) 98.8% overall read mapping rate. Aligned pairs: 7622651 of these: 183142 (2.4% have multiple alignments 23965 (0.3% are discordant alignments 97.8% concordant pair alignment rate.	Left reads: Input : 7068967 Mapped : 6983740 (98.8% of input) of these: 128110 (1.8% have multiple alignments (100 have >20) Right reads: Input : 7068967 Mapped : 6977577 (98.7% of input) of these: 128190 (1.8% have multiple alignments (99 have >20) 98.8% overall read mapping rate. Aligned pairs: 6922417 of these: 127055 (1.8% have multiple alignments 17399 (0.3% are discordant alignments 97.7% concordant pair alignment rate.	Left reads: Input : 7117908 Mapped : 7039360 (98.9% of input) of these: 124240 (1.8% have multiple alignments (25 have >20) Right reads: Input : 7117908 Mapped : 7036821 (98.9% of input) of these: 124125 (1.8% have multiple alignments (27 have >20) 98.9% overall read mapping rate. Aligned pairs: 6985858 of these: 123165 (1.8% have multiple alignments 9986 (0.1% are discordant alignments 98.0% concordant pair alignment rate.
Shoot-PA01-6h	Shoot-PA01-12h	Shoot-PA01-24h
Left reads: Input : 6917874 Mapped : 6837226 (98.8% of input) of these: 116887 (1.7% have multiple alignments (29 have >20) Right reads: Input : 6917874 Mapped : 6832838 (98.8% of input) of these: 116772 (1.7% have multiple alignments (28 have >20) 98.8% overall read mapping rate. Aligned pairs: 6783048 of these: 115823 (1.7% have multiple alignments 13420 (0.2% are discordant alignments 97.9% concordant pair alignment rate.	Left reads: Input : 7439003 Mapped : 7354145 (98.9% of input) of these: 117982 (1.6% have multiple alignments (67 have >20) Right reads: Input : 7439003 Mapped : 7350113 (98.8% of input) of these: 117981 (1.6% have multiple alignments (69 have >20) 98.8% overall read mapping rate. Aligned pairs: 7295714 of these: 116985 (1.6% have multiple alignments 10962 (0.2% are discordant alignments 97.9% concordant pair alignment rate.	Left reads: Input : 6901930 Mapped : 6818424 (98.8% of input) of these: 98818 (1.4% have multiple alignments (32 have >20) Right reads: Input : 6901930 Mapped : 6814215 (98.7% of input) of these: 98752 (1.4% have multiple alignments (33 have >20) 98.8% overall read mapping rate. Aligned pairs: 6765528 of these: 97943 (1.4% have multiple alignments 11583 (0.2% are discordant alignments 97.9% concordant pair alignment rate.

Fig.02

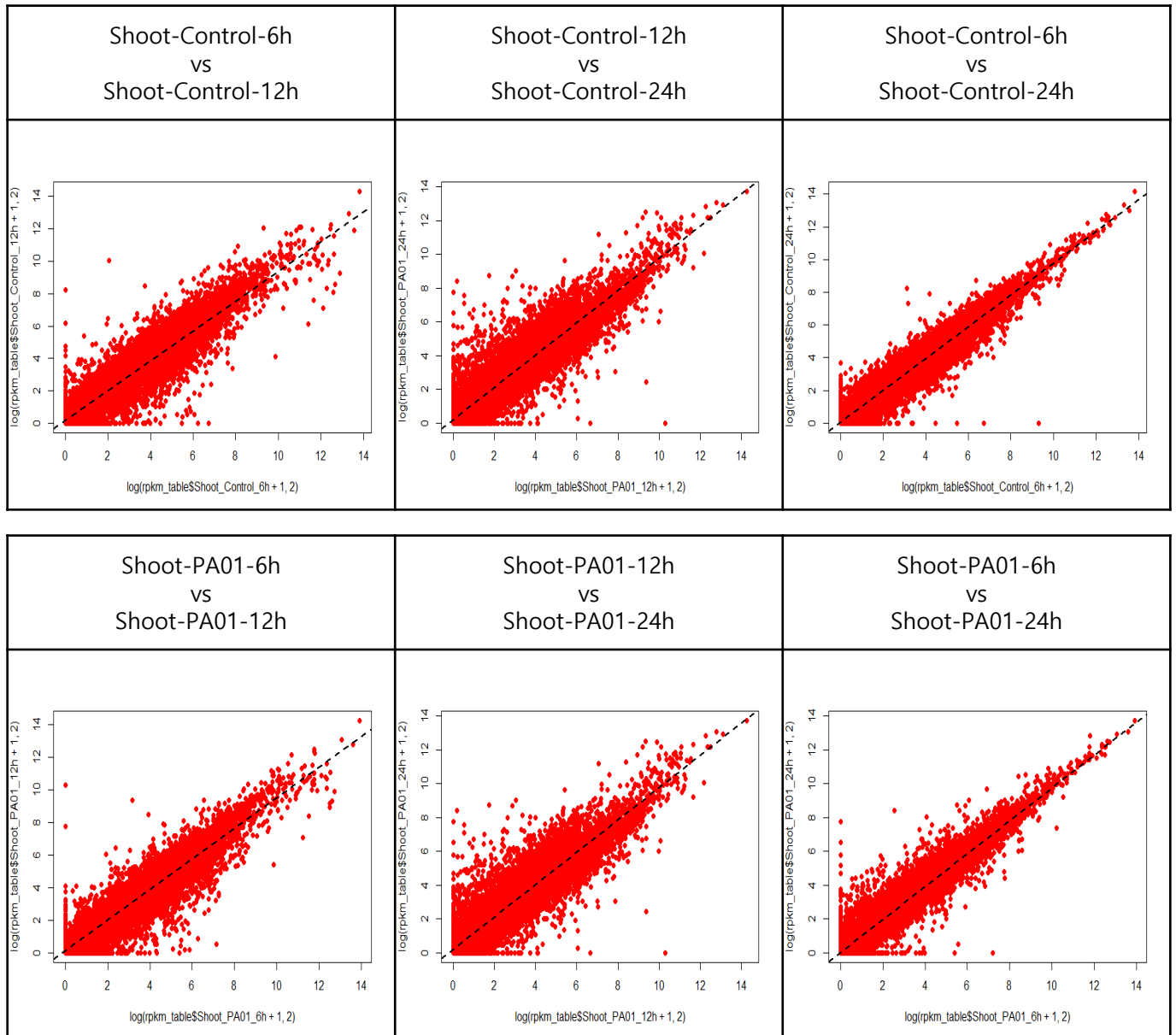


GC contents를 통해 contamination 되지 않은 것을 확인.

Fig.03

Histogram-count, density		
Shoot-Control-6h	Shoot-Control-12h	Shoot-Control-24h
<div>counts</div> <div>[1] 3686 2806 3508 4019 3111 1914 997 498 170 73 44 20 14 3</div> <div>density</div> <div>[1] 0.1766764128 0.1344964770 0.1681445621 0.1926376839 0.1491156593</div> <div>[6] 0.0917413603 0.0477879500 0.0238700091 0.0081483967 0.0034990174</div> <div>[11] 0.0021089968 0.0009586349 0.0006710444 0.0001437952</div>	<div>counts</div> <div>[1] 6256 7059 5500 1657 327 58 5 1</div> <div>density</div> <div>[1] 1.499305e-01 1.691751e-01 1.318123e-01 3.971145e-02 7.836840e-03 1.390021e-03</div> <div>[7] 1.198294e-04 2.396587e-05</div>	<div>counts</div> <div>[1] 6282 7649 5090 1480 289 61 11 1</div> <div>density</div> <div>[1] 1.505536e-01 1.833150e-01 1.219863e-01 3.546949e-02 6.926137e-03 1.461918e-03</div> <div>[7] 2.636246e-04 2.396587e-05</div>
Shoot-PA01-6h	Shoot-PA01-12h	Shoot-PA01-24h
<div>counts</div> <div>[1] 3273 2657 3451 4239 3422 1926 1014 524 219 65 37 23 10 3</div> <div>density</div> <div>[1] 0.1568806020 0.1273546470 0.1654124527 0.2031826679 0.1640224321</div> <div>[6] 0.0923165412 0.0486027896 0.0251162345 0.0104970522 0.0031155634</div> <div>[11] 0.0017734746 0.0011024301 0.0004793175 0.0001437952</div>	<div>counts</div> <div>[1] 5698 7476 5686 1374 363 59 6 1</div> <div>density</div> <div>[1] 1.365575e-01 1.791689e-01 1.362700e-01 3.772226e-02 8.699612e-03 1.413986e-03</div> <div>[7] 1.437952e-04 2.396587e-05</div>	<div>counts</div> <div>[1] 3035 2611 3477 4334 3452 1972 1108 543 184 73 38 24 10 2</div> <div>density</div> <div>[1] 1.454728e-01 1.251498e-01 1.666587e-01 2.077362e-01 1.654604e-01</div> <div>[6] 9.452140e-02 5.310837e-02 2.602694e-02 8.819441e-03 3.499017e-03</div> <div>[11] 1.821406e-03 1.150362e-03 4.793175e-04 9.586349e-05</div>

Fig.04



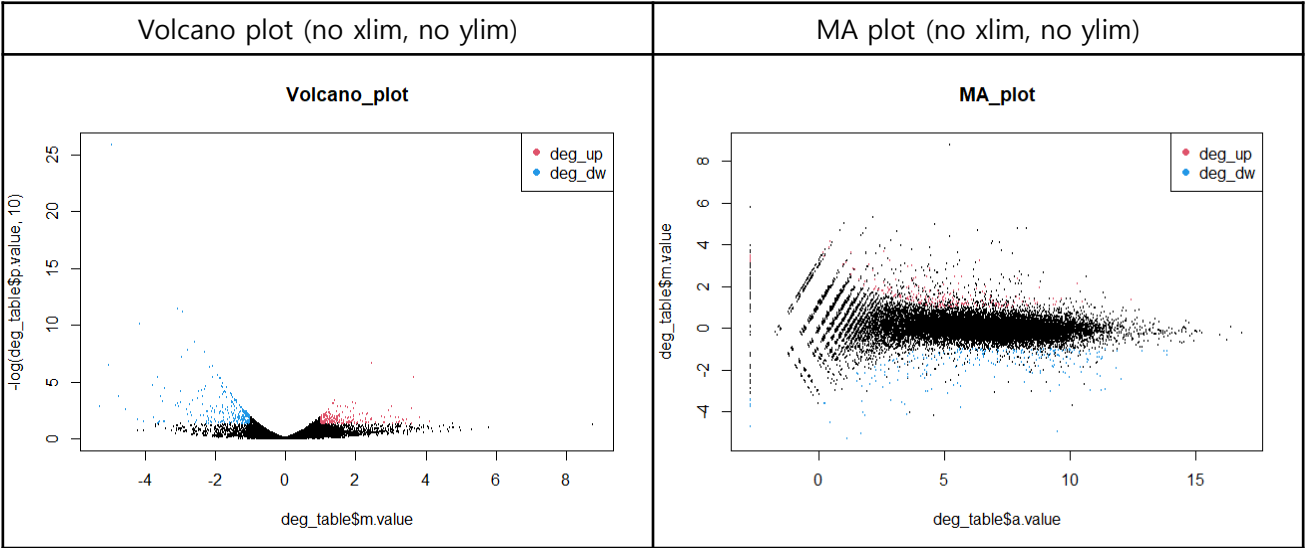
- 같은 group에서 6h와 12h, 12h와 24h, 6h와 24h를 분석한 scatter plot.
- 결과를 분석해보니, 12h와 비교를 하면 그래프가 wide해짐.
- 따라서 12h에서는 6h와 24h와는 별개로 발현되는 gene이 있다고 판단.
- Experiment group에서도 결과는 마찬가지.
- 결과를 분석할때 DEG가 PA01로 인한 DEG차이인지, 시간의 흐름에 따라 다르게 발현되는 DEG인지 구별을 해야함. -> 추가적인 연구 필요.

Fig.05

연구개역자

- No. Total genes : 32833 (TCC 결과 기준)
- No. Selected DEG : 444
 - Up regulated DEG : 204
 - Down regulated DEG : 240

Volcano plot & MA plot – TCC(deseq)

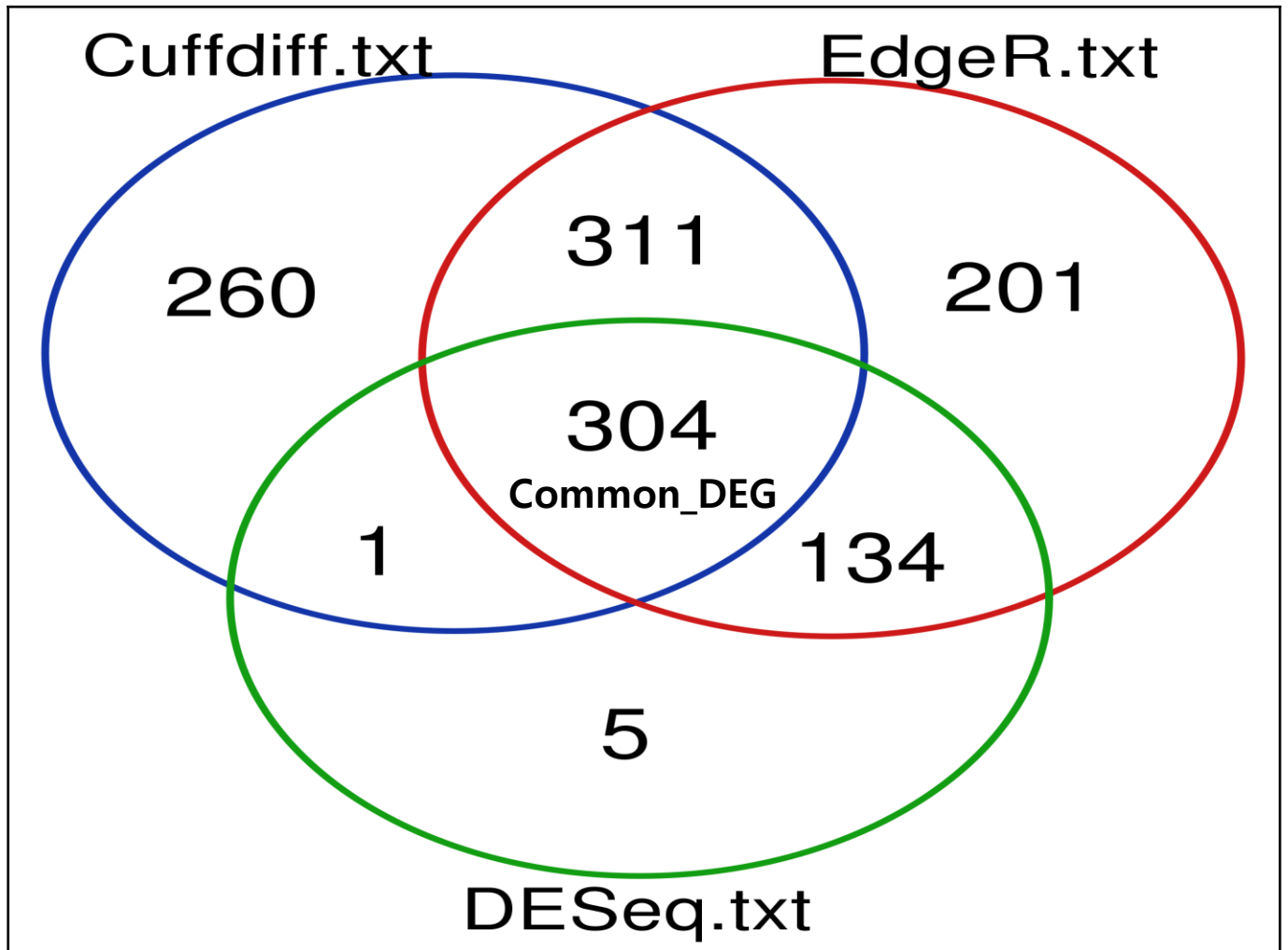


- Cuffdiff 외 TCC실행 결과로 인한 volcano plot과 MA plot.
- TCC실행 결과 같은 정보임에도 불구하고, DEG의 개수가 다름
- 또한 UP regulated DEG와 DOWN regulated DEG개수 또한 차이 발생.
- 따라서 연구목적에 따라서 DEG를 알맞게 선별해서 사용해야함.

Fig.06

DEG	Gene_ID
Common_DEG	AT2G42900 AT5G48290 AT4G29905 AT1G65690 AT3G57520 AT2G41240 AT5G37600 AT4G29950 AT4G37445 AT4G09500 AT1G66700 AT3G47340 AT4G17680 AT1G53070 AT5G41080 AT4G16680 AT2G02680 AT1G77640 AT4G08685 AT4G13990 AT3G56620 AT5G10580 AT2G38230 AT5G14120 AT3G16360 AT4G24450 AT5G59590 AT4G13195 AT5G21170 AT1G79700 AT3G23510 AT3G18610 AT4G15990 AT1G02640 AT1G11380 AT2G46430 AT4G23200 AT1G11700 AT5G51970 AT3G15720 AT3G13650 AT1G64390 AT4G37220 AT1G44800 AT3G54880 AT2G17880 AT1G25510 AT1G35260 AT1G66760 AT3G48360 AT1G24575 AT1G73600 AT2G32880 AT5G49360 AT2G32990 AT1G75780 AT1G65900 AT5G65730 AT2G38860 AT3G29030 AT2G47010 AT3G52370 AT3G43960 AT5G11590 AT1G06520 AT2G35700 AT3G45650 AT2G31945 AT1G35580 AT4G15500 AT2G37025 AT5G56100 AT1G02650 AT3G26830 AT1G50040 AT5G20250 AT4G11990 AT2G30600 AT1G76410 AT1G42560 AT1G53625 AT2G01530 AT2G22190 AT3G48100 AT2G18700 AT2G01505 AT1G45010 AT4G39510 AT3G28930 AT2G25770 AT2G40130 AT3G61820 AT5G05340 AT1G05170 AT5G49450 AT3G50560 AT2G46440 AT2G17630 AT4G20000 AT2G30750 AT1G11300 AT2G44940 AT1G18980 AT1G55760 AT5G14690 AT4G21960 AT5G22580 AT1G76620 AT1G19960 AT2G25090 AT5G11790 AT3G55710 AT2G25780 AT3G21700 AT2G16890 AT5G59750 AT4G19030 AT1G75250 AT2G34150 AT1G59970 AT1G13700 AT4G23020 AT2G45170 AT4G26950 AT1G15380 AT1G67940 AT4G12690 AT2G15760 AT2G47270 AT5G49730 AT5G22920 AT4G35030 AT1G14540 AT1G35140 AT4G36850 AT2G22990 AT4G37390 AT4G15680 AT1G69440 AT2G17280 AT5G19120 AT5G56870 AT5G14470 AT3G56260 AT4G12880 AT1G26420 AT5G01790 AT1G72150 AT4G34590 AT5G47040 AT1G21520 AT5G41410 AT1G80180 AT2G37030 AT3G56980 AT2G05510 AT5G67280 AT4G36640 AT1G12780 AT5G59490 AT1G24020 AT3G61260 AT3G10740 AT5G14740 AT5G20700 AT4G36730 AT4G35630 AT2G21550 AT3G20860 AT1G70290 AT2G25080 AT4G29110 AT1G58180 AT2G42870 AT5G42830 AT1G33030 AT2G28900 AT4G19160 AT5G23020 AT4G10955 AT4G28040 AT4G24265 AT2G23270 AT1G11260 AT3G51840 AT1G08430 AT1G26410 AT2G39510 AT1G77920 AT5G33290 AT4G16690 AT3G13610 AT3G09440 AT3G16240 AT1G21270 AT3G52180 AT2G18260 AT2G15440 AT3G45300 AT5G51790 AT5G04770 AT3G49790 AT5G63590 AT1G15520 AT2G22980 AT2G12462 AT5G17700 AT3G62950 AT5G44020 AT1G08650 AT4G38470 AT4G15700 AT5G38710 AT1G30530 AT1G13480 AT5G07475 AT3G29670 AT4G16000 AT5G14070 AT1G02620 AT3G61060 AT1G12200 AT4G16008 AT5G48110 AT1G33340 AT3G29240 AT4G07995 AT1G02610 AT2G02710 AT1G55210 AT1G77280 AT5G46871 AT4G30270 AT3G59350 AT1G66590 AT2G35170 AT1G02920 AT3G45730 AT3G13062 AT1G01620 AT1G06570 AT2G43000 AT1G55200 AT1G35720 AT1G27030 AT2G02010 AT3G29370 AT1G23870 AT1G75220 AT1G12240 AT3G51230 AT1G67070 AT2G17740 AT1G65500 AT5G42860 AT1G34380 AT1G18270 AT2G01850 AT3G25250 AT5G16570 AT1G67810 AT5G63160 AT5G24490 AT4G18360 AT5G14180 AT5G48400 AT5G57550 AT5G56550 AT2G19800 AT5G28770 AT5G02190 AT1G74010 AT4G08770 AT5G52250 AT2G01300 AT1G66646 AT2G19970 AT4G15660 AT1G75490 AT3G48390 AT1G13300 AT5G26740 AT2G42200 AT5G07100 AT5G49740 AT5G64550 AT4G19970 AT3G13750 AT4G21903 AT2G32487 AT3G14210 AT5G18600 AT2G28120 AT3G12700 AT3G06070 AT3G15450 AT5G08640 AT1G21400 AT5G52390 AT1G13430 AT5G44260 AT3G20395 AT1G30820 AT5G62480 AT2G20670

Fig.07



- 연구목적에 따라서 DEG를 알맞게 선별해서 사용해야함.