Environmental regulation of male fertility mediated through Arabidopsis bHLH89 and 91 transcription factors

Date: 13-01-2024

Introduction

Formation of functional pollen and subsequent successful fertilization rely upon the spatial and temporal regulation of anther and pollen development. The tapetum plays an important role in this through the regulation of meiotic progression and the coordination, synthesis, and secretion of the pollen wall. Genetic control of Arabidopsis tapetum development involves a series of complex interacting regulatory networks containing feedback and feedforward loops in which several transcription factors are known to play key roles including the pivotal basic helix–loop–helix (bHLH) transcription factors, DYSFUNCTIONAL TAPETUM1 (DYT1) and ABORTED MICROSPORES (AMS). DYT1 and AMS have also been shown to interact with three bHLH proteins: bHLH10, bHLH89, and bHLH91.

Previous studies demonstrated that bHLH 10, 89 and 91 are functionally redundant in anther development, regulating control of callose deposition and pollen wall formation. The bhlh single mutants have been shown to be fully fertile and morphologically indistinguishable from wild-type (WT) Col-0, while the double and triple bhlh mutants have increasingly defective anther development. The double mutants bhlh089 bhlh091 and bhlh089 bhlh010 have been reported as having reduced numbers of mature pollen grains, whereas the triple bhlh089 bhlh010 amiR-bHLH091 mutant is completely male sterile, with smaller anthers and shorter filaments.

In bhlh089 bhlh091 anthers, the tapetal cell size was observed to be uneven and the tapetum layer less organized than those of the WT. Despite the high sequence similarity and functional redundancy indicated by the bhlh89, 91 mutant phenotypes, these bHLHs have been shown to play distinct roles in activating downstream transcription.

Recent studies have shown that bhlh089 bhlh091 double mutant exhibit conditional sterility in response to high temperatures. In this study we characterize new alleles of bhlh89,91 double mutant and show that they are not fully redundant but have distinct impacts on fertility. They present novel responses to environmental conditions with a distinct light-sensitive male sterility phenotype, manifesting in an extended sterility phenotype in early flowers under low light conditions. Light is known to influence male reproductive development in terms of daylength and the duration of light exposure, but additionally light intensity has been shown to affect the initiation of flowering via phytochrome-mediated pathways involving PHYTOCHROME INTERACTING FACTORS (PIFs), a class of bHLH proteins. Here we show that bhlh89,91 mutant exhibits increased sterility in response to not only high temperatures but also low light.

Experiment conditions

Plants were grown at 21 °C (16 h days under fluorescent lighting, with a 4:4:1 ratio of red:green:blue light). At the onset of flowering, just before bolting, plants were moved to different environmental treatments. In normal light conditions, plants were illuminated with a photosynthetic photon flux density (PPFD) of 205 (±8.6 SD) µmol m–2 s–1, while the low light plants received 53 (±5) µmol m–2 s–1 PPFD through use of net shading. For high temperature treatment, plants were transferred to 28 °C.

After quantitative reverse transcription-PCR (qRT-PCR), RNA isolation and sequencing were performed. The details of each sample are shown in the following table:

Run	Bases	Experiment	Genotype	Sample Name	Assay Type	Library Layout	Library Source	SRA Study	Treatment
SRR2696 1623	455176 5400	SRX2265 5351	Col-0	GSM791 9278	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	normal light (200umol/m2/s)
SRR2696 1624	477568 2600	SRX2265 5350	Col-0	GSM791 9277	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	normal light (200umol/m2/s)
SRR2696 1625	478717 5000	SRX2265 5348	Col-0	GSM791 9275	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	normal light (200umol/m2/s)
SRR2696 1626	456538 1200	SRX2265 5346	Col-0	GSM791 9273	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	low light (50umol/m2/s)
SRR2696 1627	480422 8200	SRX2265 5345	Col-0	GSM791 9272	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	low light (50umol/m2/s)
SRR2696 1628	456433 4800	SRX2265 5347	Col-0	GSM791 9274	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	low light (50umol/m2/s)
SRR2696 1629	455058 2800	SRX2265 5349	Col-0	GSM791 9276	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	normal light (200umol/m2/s)
SRR2696 1630	455535 0000	SRX2265 5344	Col-0	GSM791 9271	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	low light (50umol/m2/s)
SRR2696 1631	420671 1800	SRX2265 5343	bhlh89 \,91	GSM791 9270	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	normal light (200umol/m2/s)
SRR2696 1632	455789 4600	SRX2265 5341	bhlh89 \,91	GSM791 9268	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	normal light (200umol/m2/s)
SRR2696 1633	455617 5800	SRX2265 5339	bhlh89 \,91	GSM791 9266	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	low light (50umol/m2/s)
SRR2696 1634	439810 3200	SRX2265 5338	bhlh89 \,91	GSM791 9265	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	low light (50umol/m2/s)
SRR2696 1635	456024 0800	SRX2265 5340	bhlh89 \,91	GSM791 9267	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	normal light (200umol/m2/s)
SRR2696 1636	455477 6200	SRX2265 5342	bhlh89 \,91	GSM791 9269	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	normal light (200umol/m2/s)
SRR2696 1637	456556 7400	SRX2265 5337	bhlh89 \,91	GSM791 9264	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	low light (50umol/m2/s)
SRR2696 1638	456249 7000	SRX2265 5336	bhlh89 \,91	GSM791 9263	RNA-Seq	PAIRED	TRANSCRIPTOMIC	SRP47 4470	low light (50umol/m2/s)

1. Download raw reads

In order to download the reads of each sample to start the project, we make use of the script bit09-download-srr.sh (modified to download all the reads, not only 100) located on /data/bit09/scripts on server 172.21.22.201. It asks for three parameters: the file with accession numbers of each sample (SRR.txt located on /home/gemam), the path of the output folder (/home/gemam/1-sra-data) and if the data is single end or paired end (PE in this case).

Once we get the fastq.gz file in the folder 1-sra-data, we proceed to the next step.

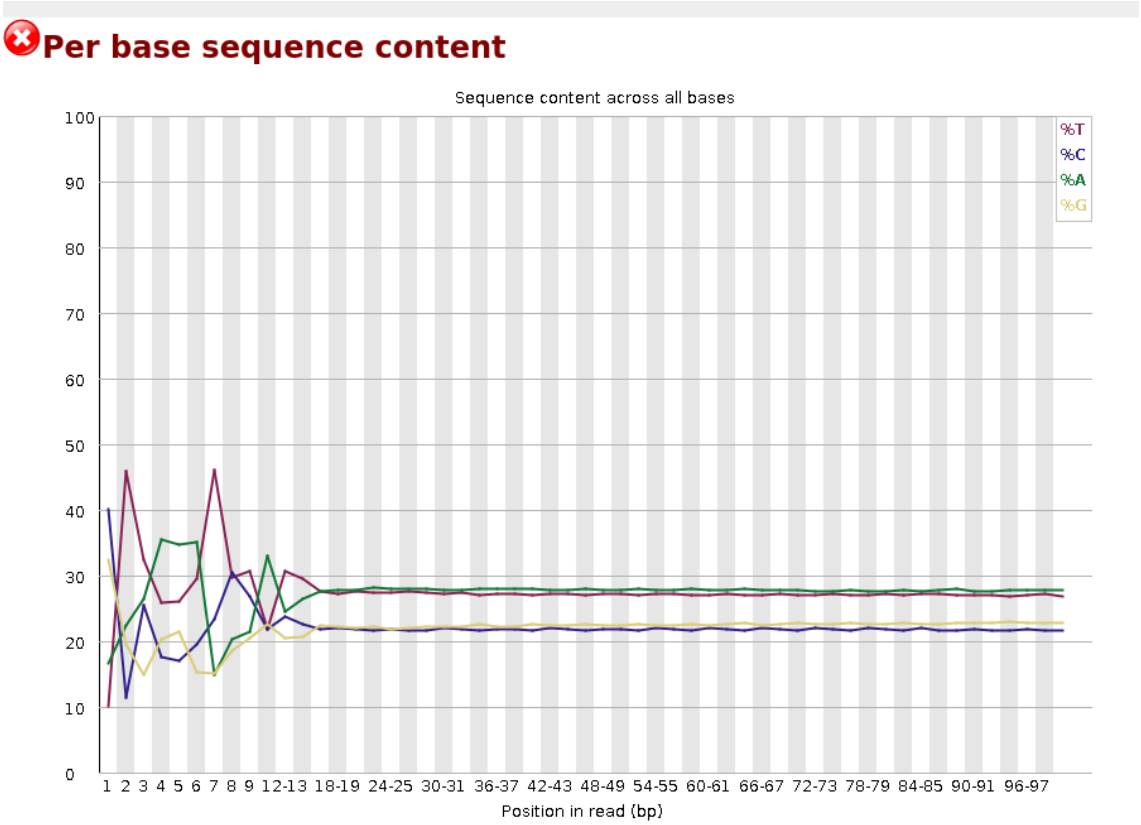
2. Quality Control with FastQC and trimming

To do this step, we use fastQC (version v0.12.1) and the script bit09-fastqc.sh located on the same folder as the script used above. The path of the input folder is /home/gemam/1-sra-data, the path of the output folder is /home/gemam/2-fastQC and the number of threads to use is 4. By doing this, we get a report about the quality of the reads for each sample, in html format. As we have 16 samples, it is easier to transfer the html files to the local machine using filezilla (version 3.66.4) and use multiQC (version 1.19) to get an overview about the quality of the reads on each sample, all in one report.

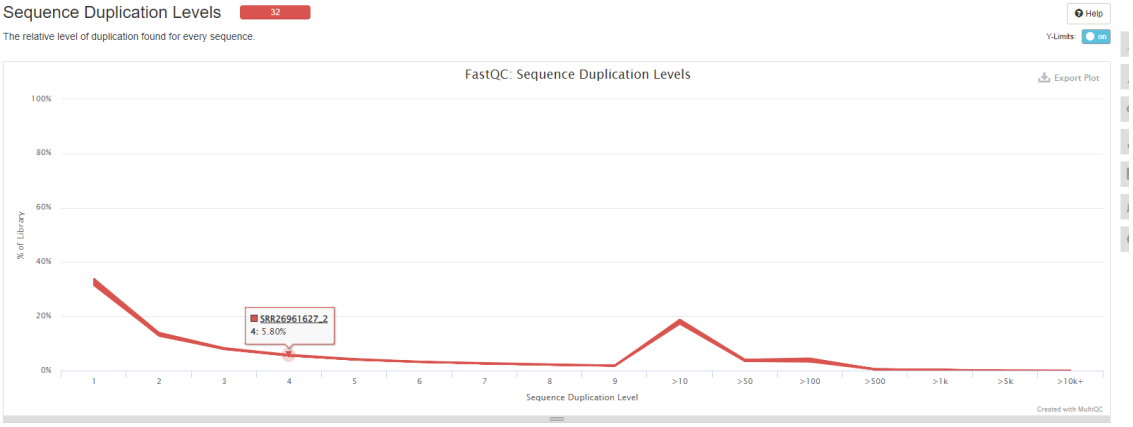
After this, we proceed to do trimmomatic (version 0.39) to remove adapters and low-quality bases using the script bit09-trimmomatic-PE.sh, the folder 1-sra-data as input and the folder 3-trimmomatic as the output folder were the results are stored. Then, we run bit09-fastqc.sh script and fastqc and multiQC later, to get a report about the quality of the samples after the trimming process, so we can compare the quality before and after trimming.

FastQC uses the fastq.gz files stored in the folder 3-trimmomatic as input and the folder 4-fastqc_trimmed as output. The multiQC report before trimming is stored on 2-fastQC and the report after trimming is stored on 4-fastqc_trimmed.

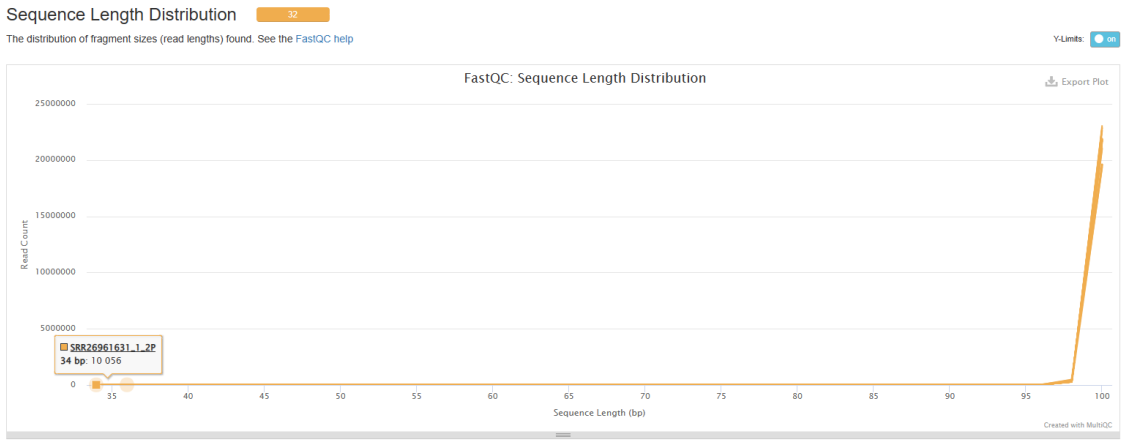
Before trimming, we only had problems in per base sequence content and sequence duplication levels, as you can see in the images below.



The deviations we can see at the beginning of the reads happen in all samples and are quite normal due to the preparation of the genetic material. This does not impact further analysis.



When sequencing RNA, both highly represented and less frequent transcripts are present. It is expected to observe duplicate reads for highly represented transcripts. Despite the anticipated presence of duplicates in the data, the RNA-Seq data was still marked as fail by FastQC, but this do not impact further analysis either.



After trimming, we got the same two failures as before, plus a sequence length warning. This warning is expected, as the process trims low-quality bases, resulting in varied read lengths, which is not problematic for the rest of the analysis.

3. Mapping

Once the quality tests are done, we are ready to map our RNA-seq samples against the transcriptome of a reference genome, that in our case is the Ref-Seq of *Arabidopsis thaliana*. The index needed is already built on the server (/data/igenomes/hisat2-index-TAIR10.1_genome_tran), so we use hisat2 (version 2.2.1) and the script bit09-mapping-hisat2-PE.sh, using the folder 4-fastqc_trimmed as input, the path to the transcriptome index and prefix of the files and the folder 5-hisat2-trimmed-genome-tran as output.

Additionally, we specified to redirect the standard output and standard error to different files, located on /home/gemam/5-hisat2-trimmed-genome-tran. If we inspect the file hisat2_stderr.txt, we can see the following alignment rates:

Samples	Overall alignment rates
SRR26961623	98.94%
SRR26961624	98.91%
SRR26961625	98.88%
SRR26961626	99.01%
SRR26961627	98.88%
SRR26961628	99.02%
SRR26961629	98.92%
SRR26961630	98.92%
SRR26961631	98.55%
SRR26961632	98.95%
SRR26961633	98.94%
SRR26961634	98.95%
SRR26961635	99.02%
SRR26961636	98.95%
SRR26961637	98.98%
SRR26961638	98.98%

As any of the alignment rates is lower than 95%, this means that everything is fine, there is no contamination and the percentage of mismatches is adequate to our circumstances.

4. Samtools

After mapping it is crucial to filter on mapping quality, sort and index the .bam files using Samtools (version 1.18), so we are able to inspect the results on IGV. In order to do that, we use the script bit09-filtering-samtools.sh, providing 5-hisat2-trimmed-genome-tran as input folder, 6-hisat2_filtered as output folder and 20 as MAPQ quality score.

5. IGV

After Samtools, the correspond gene of each of the bHLH proteins was inspected in IGV (web app version 1.13.11), were we loaded the reference genome and the sorted and indexed .bam files to compare to each other. The bHLH proteins analyzed in this project correspond to the following TAIR number:

Protein	TAIR number
bhlh89	At1g06170
bhlh91	At2g31210
Bhlh10	At2g31220

The reference genome was loaded against the sample SRR26961623, which corresponds to the wild type, and the reads corresponding to both bHLH transcription factors are found in this sample, so that means everything is correct. The same has been done to the rest of samples and they are all correct.

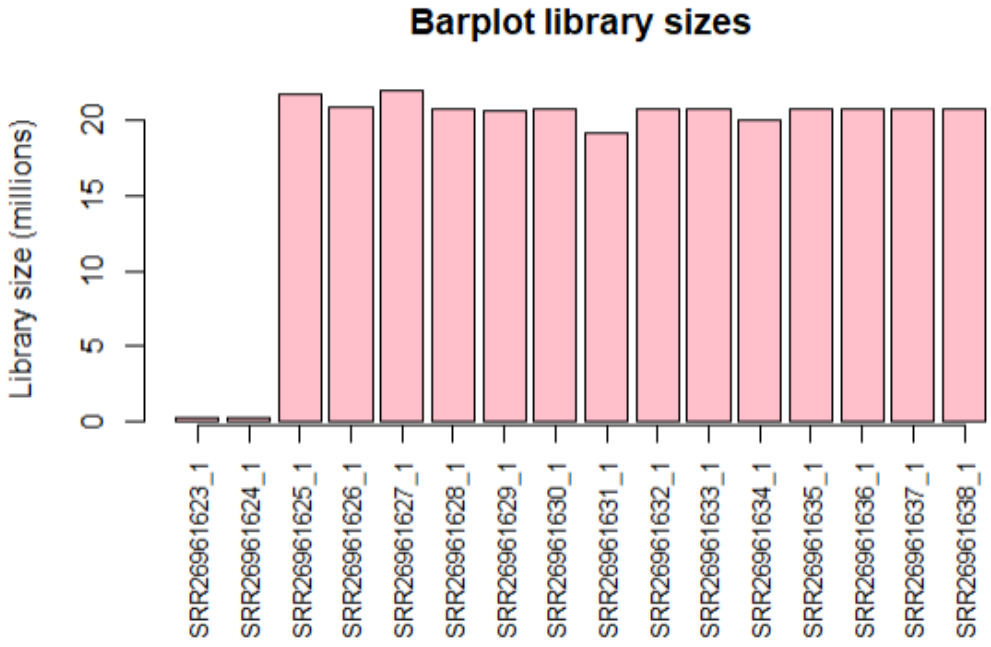
6. Counting

The last step after DE analysis is the counting step, that means we are going to estimate the transcript abundances using HTSeq-count (version 0.12.3). For this, we used bit09-htseqcount.sh script, using 6-hisat2_filtered as the input folder, /data/igenomes/Arabidopsis_thaliana/GCF_000001735.4_TAIR10.1_genomic.gtf as the annotation file, 7-hisat2_trimmed_genome_tran_htseqcount as the output folder, exon as the feature type to use from GTF file, gene_id as the attribute to use as label for counting, PE as library layout, NO strandedness, intersection-strict mode and 4 threads to use. Doing this I encountered a problem in the gtf file used, as the strands needed to be specified as '+', '-' or '.', and some of them had an ? symbol (strandedness is relevant but unknown) instead of one of those three. In order to solve this problem, the ? symbols were substituted for a '.' symbol (feature is not stranded) in the gtf file, that was renamed to Adjusted_GCF_000001735.4_TAIR10.1_genomic.gtf. After this, the problem was solved and could get the counting files for DE analysis.

7. Differential expression analysis

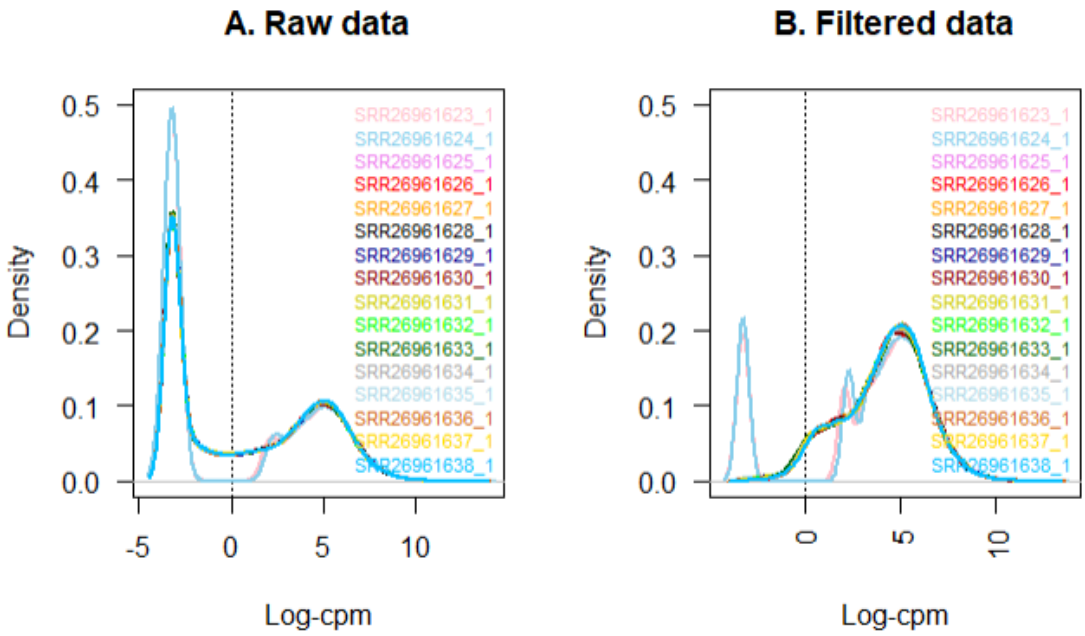
The last step in the analysis is the differential expression analysis using the packages edgeR (version 4.0.9) and Limma (version 3.58.1) in R (version 4.3.1). As we have the counts of each sample in separated files, we combine them in one data frame using the script bit09-merge-htseqcounts.R.

❖ Library sizes



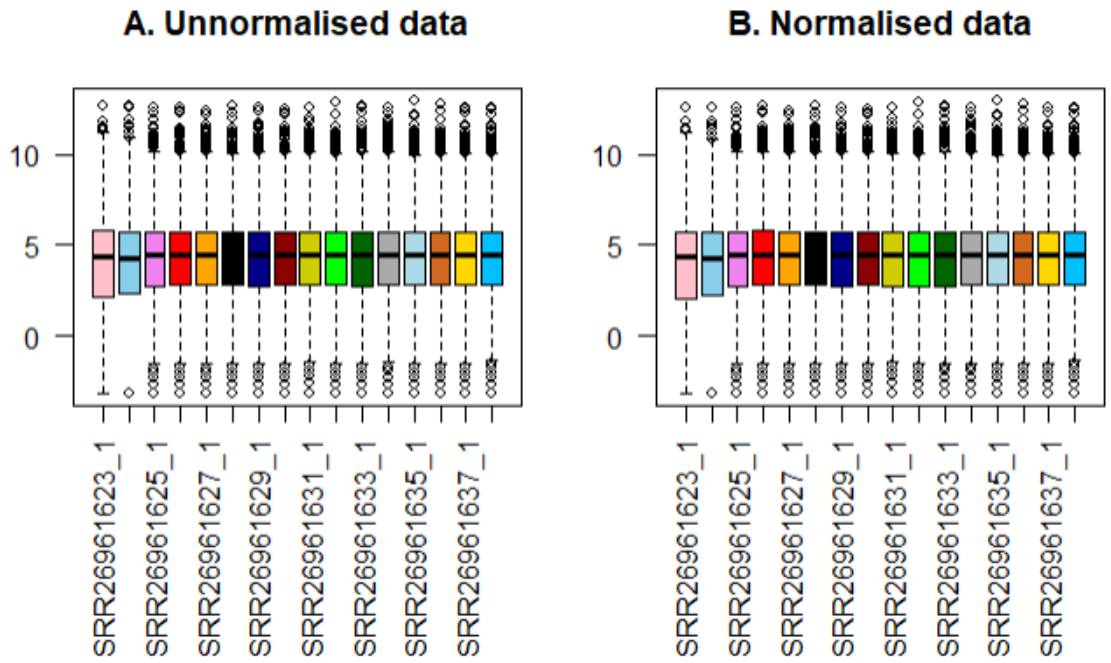
As we can see, the library sizes of the samples, is more or less the same, except for the first two samples, that as we have inspected, they have low read counts compared to the rest of the samples. Libraries that have been sequenced at a greater depth will result in higher counts, so maybe this is the case in here. For this reason, a transformation of the raw counts onto a scale that accounts for library size differences is needed. Now we proceed to transform the count data into counts per million and remove the lowly expressed genes, using CPM>1 as threshold. Also, genes must be expressed in at least three samples to be kept for downstream analysis.

❖ Data pre-processing: transformations from the raw scale and reduction of genes



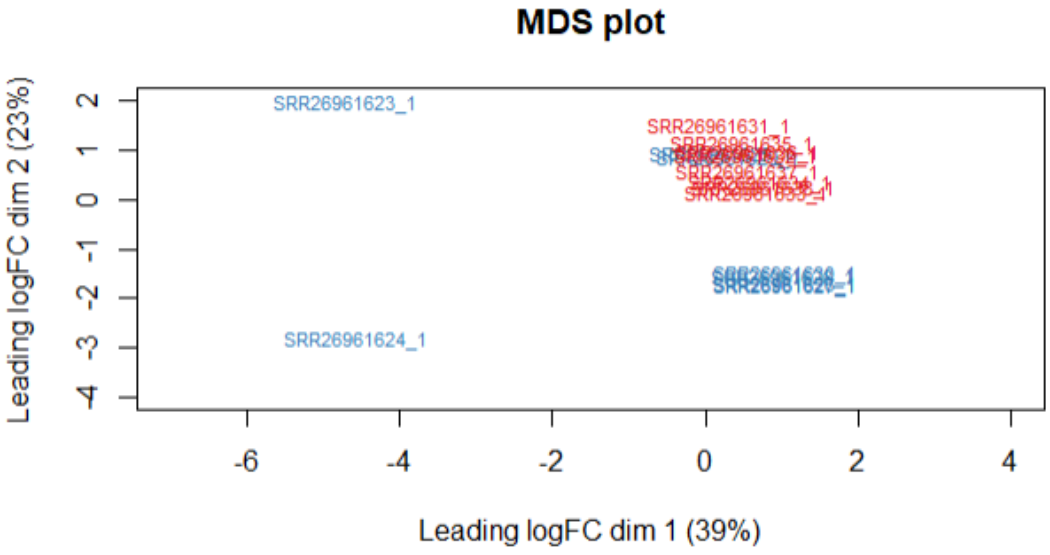
The density plot of raw data confirms that the first two samples have a weird distribution, whereas the rest of the samples share the same distribution so this means that something is going on in the first two samples. After eliminating the low expressed genes, we can see that the result is the same, again the first two samples (SRR26961623 and SRR26961624, both wild types), have a different distribution compared to the rest of the samples. Maybe this is happening because CPM only normalize for sequencing depth, but not for gene length.

❖ Data pre-processing: Normalizing gene expression distributions

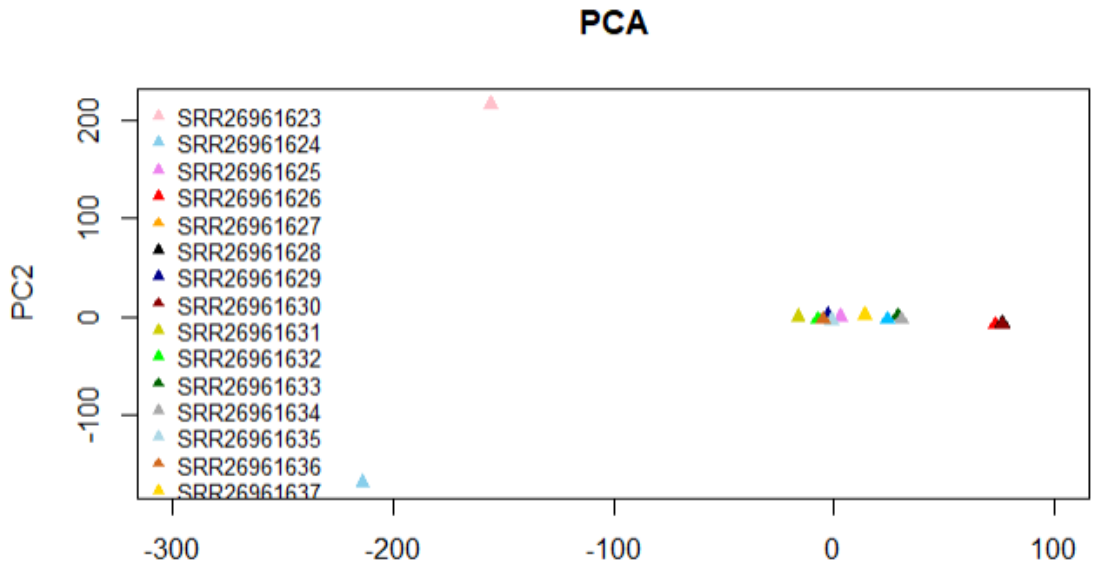


Both unnormalized and normalized boxplots are quite similar, the first two samples are slightly different from the others, so we can perform multidimensional scaling (MDS) to see how the samples cluster. Ideally samples should cluster within the conditions of interest.

❖ Data pre-processing: Unsupervised clustering of samples



As we can see in the MDS plot, the same two samples that showed a different density distribution, seem to be different from each other and quite different from the rest of the samples. Al the mutant samples (in red) cluster together whereas the wild types have form two clusters, excluding the two samples mentioned before. In the paper, they have checked the batch effects through PCA, so comparison between both techniques is necessary, as in the paper, scientists do not report any differences between the first two samples and the rest.



PCA results show the same as the MDS plot, so it would be useful to perform a t-student test to see if the differences are significant between the samples.

```
        One Sample t-test

data:  x
t = 205.64, df = 612719, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 470.3501 479.4021
sample estimates:
mean of x
 474.8761
```

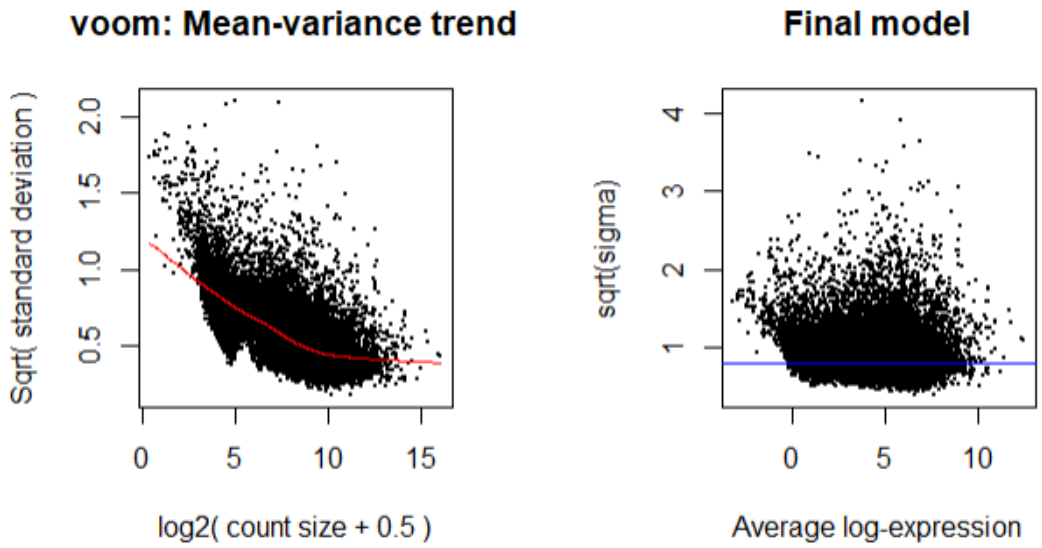
As we can see, p value is lower than 0.05, so this means that the different samples are statistically different in terms of counts. This proves that the first two samples, are statistically significant from the rest. As we have 8 wildtype samples, and the first two have low counts, maybe it is better to ignore these two and keep going with the rest of them.

❖ Differential expression analysis: Design matrix and contrasts

To start a DE analysis we need a design matrix based in a model. In our dataset we want to know which genes are expressed differently between the wildtype and the bHLH89, 91 doble mutant. We use makeContrasts function from lima package to create the matrix.

```
      Contrasts
Levels KOvsWT
    KO       1
    WT      -1
```

❖ Differential expression analysis: Removing heteroscedasticity



Distributions of counts are naturally heteroscedastic, meaning that the variance of the error is not constant through the different samples, having larger variances on larger counts. Voom converts raw counts into log-CPM values, and the plot typically shows a decreasing trend, as we can see on the left plot. After a linear modelling process, we get the final model where the heteroscedasticity is adjusted. The model's residual variances are plotted against average expression values (in Final model) and variance is no longer dependent on mean expression level.

❖ Differential expression analysis: Examining the DE genes

```
       KOvsWT
Down     2628
NotSig  14644
Up       2784
```

```
       KOvsWT
Down       21
NotSig  20026
Up          9
```

As we can see in the first table, we have 2628 genes downregulated and 2784 genes upregulated in KO (mutant) relative to the wildtype. Using the log-fold changes, we reduce the number of genes downregulated to 21 and the number of genes upregulated to 9.

	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>
AT1G55560	-3.639174	3.6809272	-15.97371	1.423654e-11	2.855281e-07
AT5G23700	-1.360468	3.7138526	-14.82160	4.615455e-11	3.085586e-07
AT1G32450	-2.267911	2.6265478	-14.03325	1.081409e-10	4.337746e-07
AT5G50120	-3.671797	0.8268515	-14.93512	4.096117e-11	3.085586e-07
AT1G08065	-6.049169	2.1594798	-14.28223	8.228492e-11	4.125766e-07
AT3G57620	-2.363846	5.4447099	-12.61780	5.552672e-10	1.856073e-06
AT1G65620	-1.807261	4.7313235	-12.15688	9.775204e-10	2.178350e-06
AT3G29644	2.773790	2.4432922	12.29824	8.203544e-10	2.178350e-06
AT2G21140	-2.227055	5.9223318	-11.82069	1.493066e-09	2.722266e-06
AT2G02990	2.441818	3.4419306	11.67140	1.807635e-09	3.021160e-06

1-10 of 10 rows | 1-6 of 6 columns

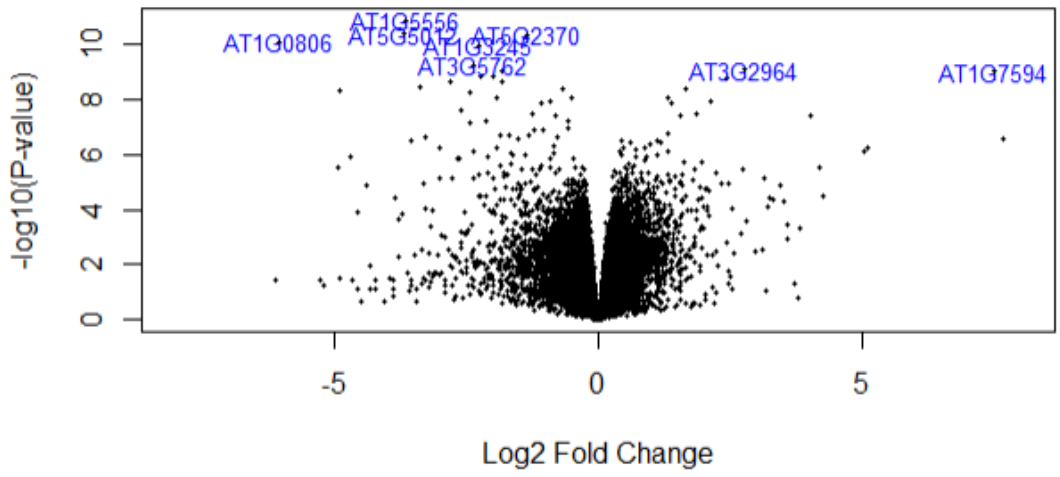
In this table we can see the top 10 differential expressed genes using e-Bayes moderated t-statistics.

	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>
AT1G08065	-6.049169	2.1594798	-14.28223	8.228492e-11	4.125766e-07
AT5G50120	-3.671797	0.8268515	-14.93512	4.096117e-11	3.085586e-07
AT1G55560	-3.639174	3.6809272	-15.97371	1.423654e-11	2.855281e-07
AT3G57620	-2.363846	5.4447099	-12.61780	5.552672e-10	1.856073e-06
AT1G32450	-2.267911	2.6265478	-14.03325	1.081409e-10	4.337746e-07
AT2G21140	-2.227055	5.9223318	-11.82069	1.493066e-09	2.722266e-06
AT1G65620	-1.807261	4.7313235	-12.15688	9.775204e-10	2.178350e-06
AT5G23700	-1.360468	3.7138526	-14.82160	4.615455e-11	3.085586e-07
AT2G02990	2.441818	3.4419306	11.67140	1.807635e-09	3.021160e-06
AT3G29644	2.773790	2.4432922	12.29824	8.203544e-10	2.178350e-06

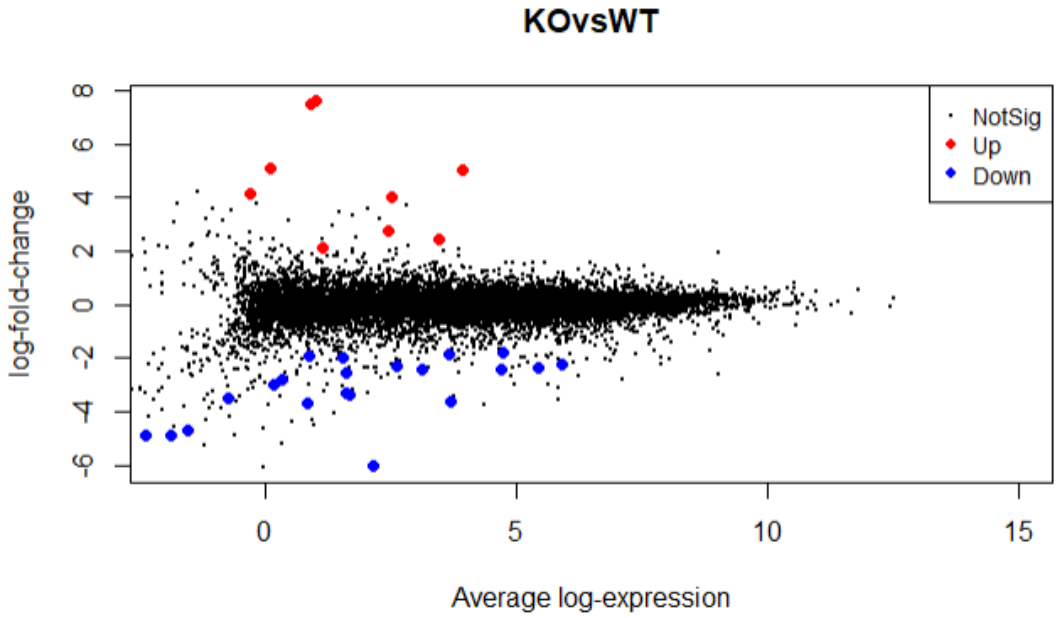
Here we can see the top 10 differential expressed genes sorted by logFC.

	logFC <dbl>	AveExpr <dbl>	t <dbl>	P.Value <dbl>	adj.P.Val <dbl>
AT1G08065	-6.049169	2.1594798	-11.921206	6.607920e-10	1.014313e-05
AT1G61230	-4.886357	-1.8673844	-8.678461	6.626027e-08	2.657832e-04
AT5G50120	-3.671797	0.8268515	-10.867600	2.619016e-09	1.750900e-05
AT1G55560	-3.639174	3.6809272	-11.584333	1.011481e-09	1.014313e-05
AT3G45060	-3.362458	1.6831728	-7.860683	2.546920e-07	6.551402e-04
AT3G16460	-2.806244	0.3200277	-7.373331	5.926345e-07	1.320653e-03
AT3G57620	-2.363846	5.4447099	-7.279974	6.992169e-07	1.402349e-03
AT1G32450	-2.267911	2.6265478	-7.845506	2.613244e-07	6.551402e-04
AT3G29644	2.773790	2.4432922	7.864508	2.530121e-07	6.551402e-04
AT1G75945	7.497689	0.9192325	10.552818	4.108219e-09	2.059861e-05

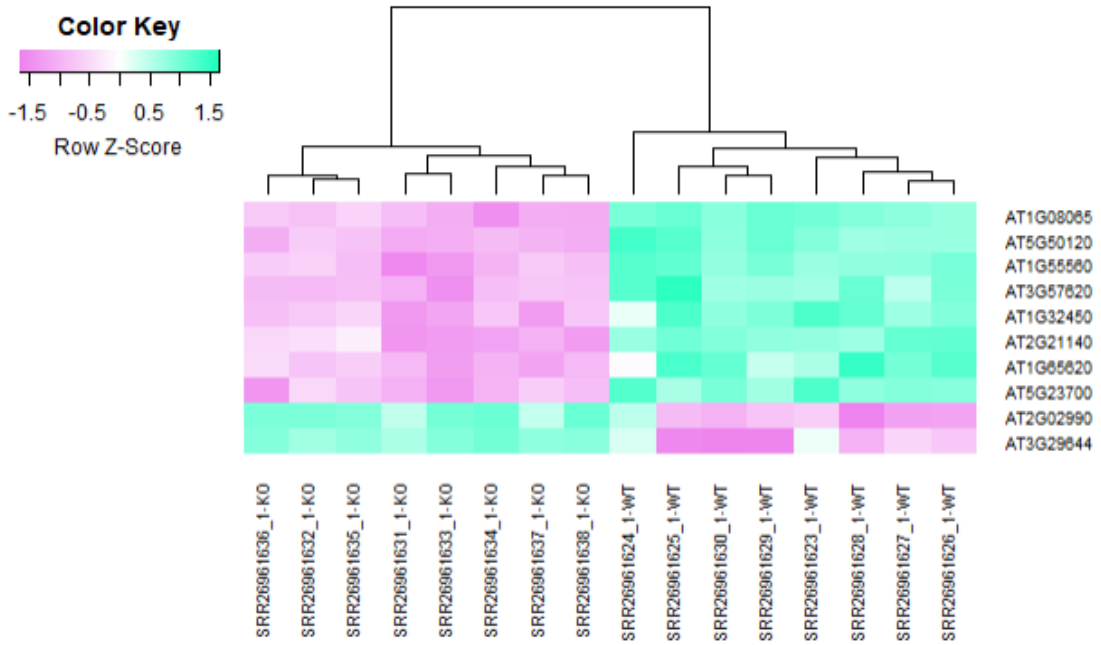
Here we have used treat instead of eBayes to get the top 10 differential expressed genes. We can see that the first gene is the same as in eBayes sorted by logFC.



As we can see in the volcano plot, the genes that do not change their expression, will have a log2 Fold Change close to 0, whereas the genes that are on the top of the plot, will be the most variable in terms of expression (codes in blue) and thereby, the most relevant for our analysis.



With the MD plot we can check the genes that have an expression significantly different from the mean. The ones in red are up regulated whereas the ones in blue are down regulated. The number of genes(dots) of each correspond to the up and down regulated gene table we have seen above in the analysis.



If we take a closer look to the heatmap created, we can see how the top 10 DE genes are expressed through the different samples. The most DE gene, AT1G08065, seems to be upregulated in all of the wildtype (WT) samples, whereas it is downregulated in all of the knock-out (KO) samples, so maybe this gene is related to the bHLH89 and bHLH91 genes, that have been eliminated in the knock-out samples. Same happens for the genes AT5G50120, AT1G55560, AT3G57620, AT1G32450, AT2G21140, AT1G65620 and AT5G23700.

On the opposite side we have the genes AT2G02990 and AT3G29644, that are downregulated in the wildtype samples, whereas they are upregulated in the knock-out samples.

We can also see associations between different samples. If we pay attention to the metadata table at the beginning of the report, we can see the light conditions for the different samples. Here, we can see that the dendrogram is separated into two big groups, one belongs to the KO samples (on the left), and the other one belongs to the wildtypes (on the right). Now if we check the most related samples at a lower level, we see that, for example, SRR26961637_1_KO and SRR26961638_1_KO, have a similar expression profile of the top 10 DE genes. These two samples are bHLH89, 91 mutants and also were treated with low light conditions. This pattern happens in all the groups, the samples with the same light conditions are grouped together, which is logical, but there are a few samples that need to be pointed out. SRR26961631_1_KO and SRR26961633_1_KO are grouped together at the lowest level, showing some genes highly downregulated (especially AT1G55560), but the first one had a normal light treatment whereas the second one has a low light treatment. This could mean that this gene is not affected by the light treatment but it is affected by the knock-out of bHLH89 and 91, which are transcription factors involved in the development of the pollen tapetum in *Arabidopsis thaliana*. Additionally, AT1G55560 is a gene that belongs to the SKS family, than seems to be involved in pollen tube formation, so the knock-out of bHLH89, 91 transcription factors could affect the expression of this gene and ultimately, the fertility of the plant.

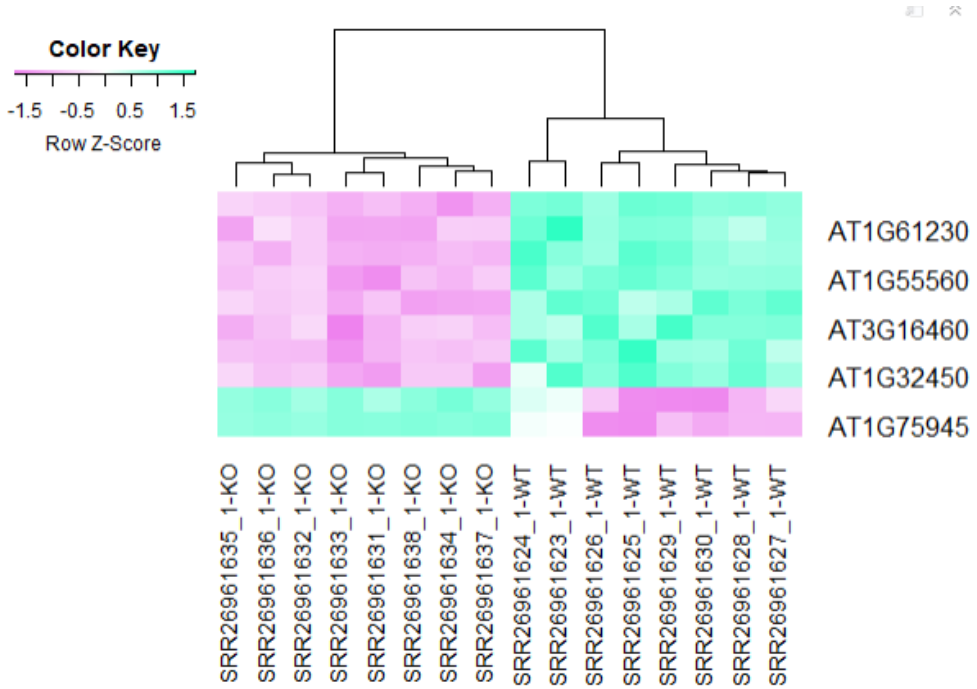
Similar thing happens with SRR26961629_1_WT and SRR26961630_1_WT, the first one was treated with normal light conditions but the second one was treated with low light conditions.

They both have similar expression profiles, especially for the gene AT32G29644, that corresponds to non-coding RNA, which has an unknown function.

As we pointed out previously, sample SRR26961624_1_WT had low read counts, and in the heatmap we can see this is the sample with the least changes in gene expression if we compared it to the rest of the samples. This could be explained because it is wildtype and also had normal light conditions. The weird thing here is that SRR26961625_1_WT and SRR26961629_1_WT are also wildtypes with normal light conditions but the last two genes, corresponding to a ribonuclease and non-coding RNA, respectively, are highly downregulated. This event is something that definitely needs further investigation.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using the online Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/home.jsp). We searched for the top 10 most different expressed genes and we got the following descriptions, used in the explanation of the heatmap above:

Loci	Gene description
AT1G08065	Alpha carbonic anhydrase 5(ACA5)
AT1G55560	SKU5 similar 14(sks14)
AT1G32450	Nitrate transporter 1.5(NRT1.5)
AT1G65620	Lateral organ boundaries (LOB) domain family protein (AS2)
AT2G02990	Ribonuclease 1(RNS1)
AT5G50120	Transduction/WD40 repeat-like superfamily protein
AT3G57620	Glyoxal oxidase-related protein
AT2G21140	Proline-rich protein 2(PRP2)
AT5G23700	Coiled-coil protein
AT3G29644	ncRNA



One of the main differences with the other heatmap is that four genes have been replaced by others. Also, the gene expression for SSR26961623 and SRR26961624 is quite similar, which makes sense, as they are both wildtypes and treated with normal light but it is still very different to SRR26961625 and SRR26961629, which have two last genes on the map downregulated, corresponding to non-coding RNA and a uncharacterized protein, respectively, so we cannot explain why this is happening. Additionally, SRR26961626 and SRR26961625 have been grouped but they have different light conditions, so this could mean that these 10 genes are not very affected by the light conditions in the wildtype.

Here we can see the gene description of the newly genes added to the top DE genes.

Loci	Gene Description
AT1G61230	Manose-binding lectin superfamily protein (AT1G61230)
AT3G16460	Manose-binding lectin superfamily protein (JAL34)
AT3G45060	High affinity nitrate transporter 2.6 (NRT2.6)
AT1G75945	Uncharacterized protein (AT1G75945)

If we pay attention to the knock-outs, we can see that the samples SRR26961635, SRR26961636 and SRR26961632(all KO and normal light treatment) have a similar expression pattern between them, but different to the knock-outs that have been exposed to low light. This is not so clear in the wild type, so the theory here is that the low light conditions in the bHLH89,91 mutant could affect the transcription of some genes. The authors of the experiment conclude that the bhlh89,91 mutant showed a lower transcriptional response to light than the WT (Col-0). The also performed a GO enrichment analysis showing significant enrichment of GO terms such as sexual reproduction, lipid storage, glucose import, oxidation–reduction process, and SCF-dependent proteasomal ubiquitin-dependent catabolic process, suggesting that these responses are impaired in the bhlh89,91 mutant light response.