# Myb5 Report

CiCi Zheng

July, 11th, 2019

This is the report for process we find the luteus gene that correspond to "myb5", which is a candidate gene that activate the anthocyanin pathway in M. variegatus. After we solved the sample assignment mistake, I re-did the differential expression(DE) analysis in DESeq2. We get the same set of 632 genes that are differentially expressed. Information of that part can be access on the GitHub site: github link

In this report the workflow we find the "myb5" after obtaining the set of DE genes.

First, we subset the luteus cds to individual gene by gene fasta. file. Then we PSI-BLAST each gene with Arabidopsis genome database, and we grab the top hit and add discription for each corresponding M. luteus gene. From those, we subset the ones that belongs to the myb transcriptional factors or the bHLH gene binding proteins (mybs function in complex with bHLH to activate downstream genes).

genes	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	blast	annotation
Mlu_39530	868.61660	1.5150286	0.2112962	7.169012	7.55000e-13	2.190000e-09	AT4G09820	y_bHLH
Mlu_17841	102.36359	2.3505744	0.3719788	6.291004	3.15000e-10	5.010000e-07	AT3G61250	x_myb

We plotted these mybs and bHLHs on the transcriptome vocano plot, visualizing the fold change and p-value of each gene:
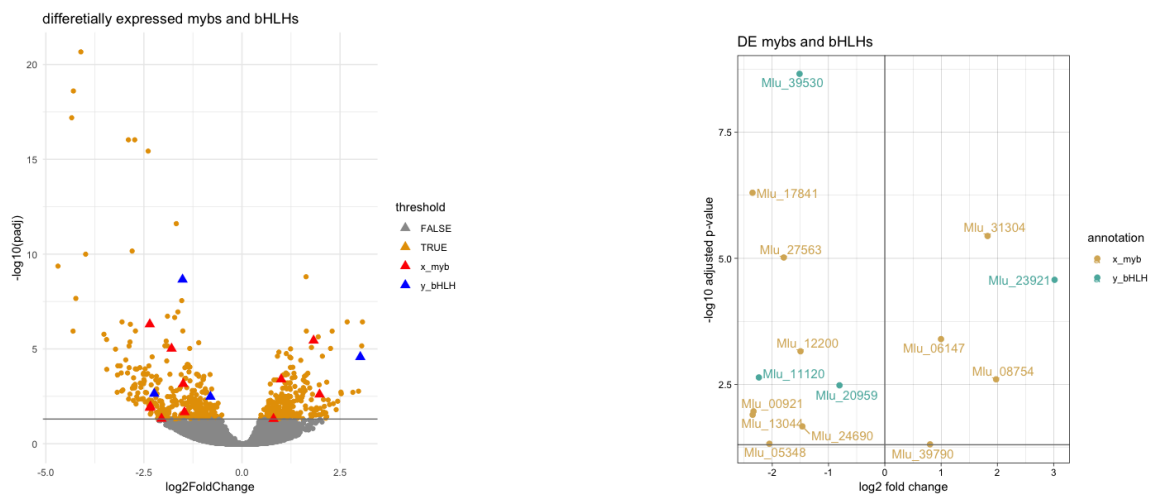

Then I obtained the myb5_consensus_sequence from Dr. Cooley, which blasts to 3 genes in luteus cds:

```
Query= MLV_Myb5_gDNA_consensus 9-14-15

Length=2506
                                                               Score        E
Sequences producing significant alignments:                   (Bits)     Value

Mlu_42095_MLU                                                   854        0.0
Mlu_12200_MLU                                                   680        0.0
Mlu_12207_MLU                                                   595        7e-169
```

Among them, only Mlu_12200 is one of the DE genes, and shows signs of down regulation in the RNAi experiment.
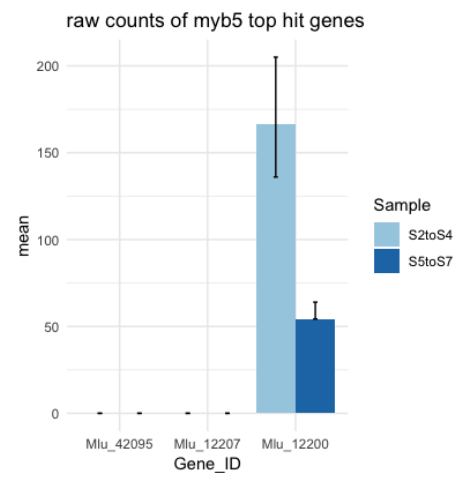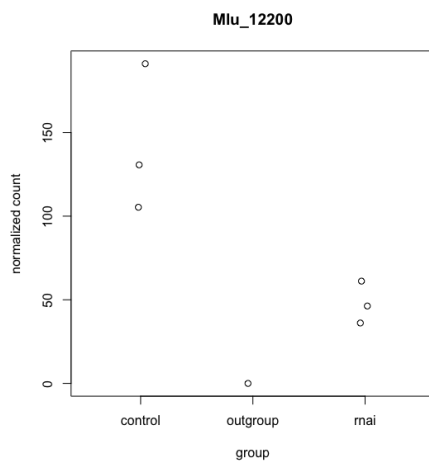
Other Evidence:

- From Dr. Cooley, Myb5 is one of the predicted genes on M. l. luteus Scaffold 1235; it also has a homeolog, Myb5b, on Scaffold 749 but never got expressed;

- From our latest coordinate: start codon of Myb5 is 39,551 and the end codon of exon 4 is 35,849 (negative strand) which aligns well with Mlut_12200's coordinate (35,742- 39,447);

- Another top hit Mlut_42095 is located on Scaffold 749 and has no expression, which corresponds well to the Myb5b gene;

- We replace all 3 of the top hit genes with myb5 consensus sequence in luteus cds, and do the alignment of the fastq files again with Bowtie2 again just to see what result we would get if we use a direct mapping, to have some more confidence in our aligment and read count pipeline. Then we calculate the RPKM directly for myb5, and it shows the very similar fold (3.2 fold in the regular analysis, and 3.4 fold here) down regulation in the RNAi experiment.

|	sum	per million	MLV_Myb5_Count	RPKM	Scaled&Rounded
s1	7265681	7.265681	55	0.003020684	3
s2	14324854	14.324854	3080	0.085798451	86
s3	16216062	16.216062	3066	0.075447645	75
s4	16160738	16.160738	1760	0.043458068	43
s5	10290664	10.290664	463	0.017953806	18
s6	17053029	17.053029	916	0.021434476	21
s7	24145720	24.14572	1177	0.019451596	19
			RPKM = MLV_Myb5_count/per_million/length(2506)		

Thus we conclude that Mlu_12200 is the Myb5 we want.
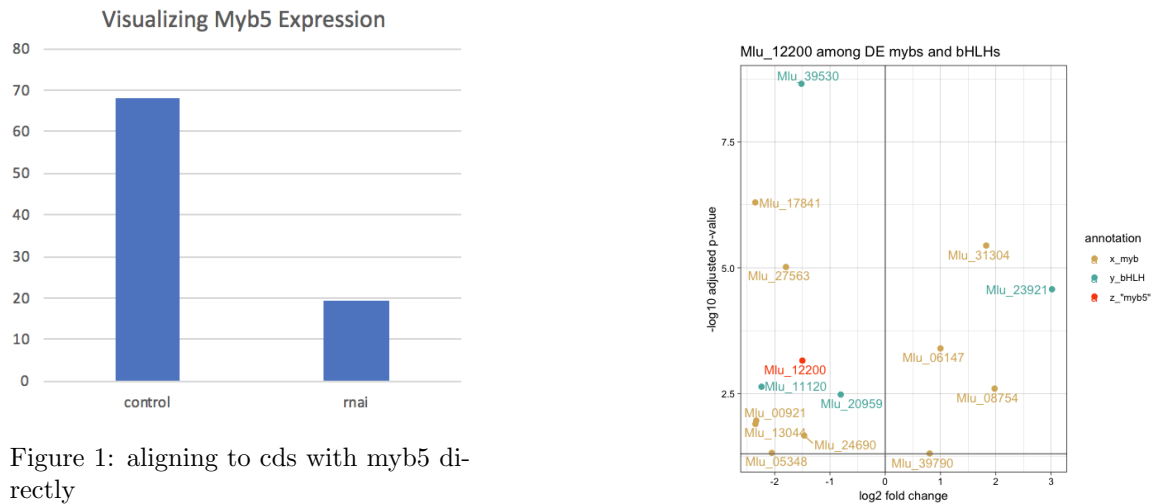
Figure 1: aligning to cds with myb5 directly