Behind the competing pathway of vincristine and vinblastine

Introduction

The biochemical compounds vincristine and vinblastine are used in the clinic as anti-cancer agents. To date, this compounds are only found in the plant *Catharanthus roseus,* a member of the euasterids I clade (Gentianales order, Apocynaceae family). These are synthesized in very low proportions. Thus, the production of vincristine and vinblastine is desirable but very expensive. Knowledge about the genomic mechanisms of these two compounds may enable the enhancement in the production of these. For instance, biotechnology approaches may enable the creation of the vincristine and vinblastine genetic pathway (VVP) in a synthetic way.

Miettienen et al. (2014) raised the question of a possible competition between heteroyohimbine synthases for recruitment by SGD (Strictosidine β-glucosidase) when distinct enzymes are co-expressed in the same tissue/cells. They speculated that this could be causing the low proportions of vincristine and vinblastine. Kellner et al. (2015) provided high quality draft genome assemblies that represent genetic regions of the genome and revealed the pathway of the MIA (monoterpene-derived indole alkaloid) genes (including the VVP). Yet, the competing pathway of vincristine and vinblastine has never been investigated. The main objective of this study was to identify genes that are part of the hypothetical competing pathway of VVP.

Analysis design

In this study, three aspects were considered to determine which genes may be competing with VVP:

1) Some candidate genes have been proposed for the competing pathway (Figure 1 –Appendix I)

The heteroyohimbine synthase may require SGD, as well as SLS (secologanin synthase). SGD is a critical branch point of, where the chemical diversity of the different classes of MIAs emerges (Kellner et al., 2015), whereas SLS controls secologanin biosynthesis. Another compound whose synthesis may be recruiting SGD is Horhammercine. Five known genes in the pathways of horhammercine and heteroyohimbine were considered in this study (Table 1 –Appendix I). To enable contrast, the genes located in the same scaffolds were also considered. The compound Catharantihine (Figure 1 – Appendix I) has shown to be a precursor of vincristine (Que et al., 2015); so, it was discarded as a possible competitor.

2) Genes in any pathway are expected to be together

Studies have shown that genes encoding specialized metabolism in plants are often physically clustered in the genome (Frey *et al*., 1997; Amoutzias and Van de Peer, 2008; Field and Osbourn, 2008; Swaminathan *et al*., 2009; Winzer *et al*., 2012; Itkin *et al*., 2013; Mugford *et al*., 2013). Using Genome Viewer we observed that one of the suggested competing genes (CRO_005213) was located in the same scaffold that one of the essential genes for the synthesis of the Vincristine and Vinblastine (CRO_005215) (Figure 2 – Appendix I). Moreover, it is known that at least one gene encoding

heteroyohimbine biosynthesis is physically clustered with one of the distinct paralogs of SLS, linking the iridoid and alkaloid portions of the pathway (Kellner et al., 2015).

3) Genes of the competing VVP are expected to be up-regulated with a methyl jasmonate (MeJA) treatment.

It has been shown that most genes within the MIA pathway (VVP included) are more highly expressed when the plant is treated with MeJA (Vazques-Flota and De Luca, 1998). The competing pathway is most probably within the MIA pathway and it possibly recruits some of the compounds that are essential for the VVP synthesis. Since the vinblastine/vincristine biosynthesis is induced with the MeJA treatment (Gongora-Castillo *et al*., 2012), it is expected that the increase in expressions levels is higher in VVP than in the competing genes.

Implementation

Values of expression abundance in MeJA-treated seedlings was determined using RNA-seq data from a previous study (Gongora-Castillo *et al*., 2012), as explained in Kellner et al., (2015). For this, three different datasets were used:

- Expression values provided in Kellner et al., 2015.
- Expression values estimated in this study in an attempt to reproduce procedure explained in Kellner et al. (2015).
- Expression values estimated in this study using other suits of tools. (hisat_2&STringTie instead of TOPHAT&CUFFLINKS).

We used the assembly provided in Kellner et al. (2015). Data was downloaded from Medicinal Plant Genomics Consortium (http://medicinalplantgenomics.msu.edu/). The values of gene expression were given in fragments per kilobase of transcript per million mapped reads (FPKM).

Obtaining the expression values with 2 approaches

We performed the analysis using Hisat_2 and StringTie; and using TOPHAT, Bowtie and CUFFLINKS. This enabled us to compare two different approaches. In both cases, we used cutadapt (v.1.4.1) to remove adapters from the reads. We fixed a minimum quality value of 20 and a minimum trimmed read length of 30 bp, as explained in Kellner et al. (2015).

In the first approach, we used Bowtie (v.0.12.7), TOPHAT (v.2.0.10) and CUFFLINKS (v.2.2.1). We used cuffmerge to merge the assemblies of the trasncriptomes (seeds, seeds after 5 days of MeJA treatment, and seeds after 12 days of MeJA treatment), and cuffdiff to obtain the FPKM values under the 3 conditions. Details of this procedure are given somewhere else (Kellner et al., 2015).

Alternatively, we used Hisat_2 (v.2.0.4) to align the reads to long reference sequences, and StringTie (v.1.3.1c) to obtain the gene expression profiling. We used gffread to transform .gff3 into a .gtf file that can be used by Hisat_2 to create an index. This index is required to align the reads to the reference. To transform the .sam file produced by Hisat_2 into a .bam file that can be used by StringTie, we used samtools.

Differentially expressed genes

We analyzed the FPKM values of 8 genes located in the same scaffold as AF253415.1, which is directly related to the MAT compound, 6 genes in the same scaffold as HQ901597 (SH19), 8 genes that are located together with the genes responsible for RO, THAS and SLS4 (KM524259, KM524258 and SLS_4, respectively), and the 8 genes that are essential for the synthesis of VVP. These groups of genes can be found in Appendix I – Tables 2 to 5.

We calculated some statistics for each individual gene (Appendix II), as well as two samples independent t-test. For the test, one sample was the five suggested competing genes and the other sample was the group of genes known to be involved in the VVP. The response of the suggested genes to MeJA treatment was equivalent to the response of the VVP genes. This suggests that both pathways may be regulated by a similar compound; possibly SGD (Miettienen et al., 2014).

Results

The five genes were differentially expressed with respect to the genes involved in the VVP pathway under the condition of no treatment. A possible explanation for this is that the competing pathway is recruiting SGD. We did not observe a significant difference between competing genes and VVP genes, which is in agreement with Gongora-Castillo *et al*. (2012). They observed that Vincristine and Vinblastine was more highly expressed under the MeJA treatment. We concluded that the suggested genes may be part of a competing pathway that is responsible for the low production of VVP.

The relative increase of 5d compared to MIA 5d was lower in the suggested competing genes than in the other genes analyzed, indicating that the response to the treatment is stronger in the VVP genes (Appendix I – Table 6). This is in line with previous observations and with the assumption that the 5 genes proposed are part of a competing pathway.

The relative increase under the MeJA treatment was much larger for the MAT gene; then, for the T19H gene, for the MIA pathway genes and lastly, for RO, THAS and SLS4 (Figure 3 – Appendix I). We conclude that although the MeJA treatment increases the production of Vincristine and Vinblastine, certain genes in the competing pathway, such as AF253415.1 (MAT) and HQ901597.1 (T19H) are more highly expressed than the overall VVP genes. This could be regarded as an indicator that the competing pathway is still incomplete; as overall, the competing pathway responds less to the treatment than VVP (Gongora-Castillo *et al*., 2012).

Replication VS alternative approach

The Hisat_2&StringTie run much faster than TOPHAT&CUFFLINKS. The most striking difference between both approaches was that with TOPHAT&CUFFLINKS, one of the MIA genes (CRO_017448) appeared to be down-regulated, whereas with Hisat_2&StringTie, this was not the case (Figure 4 - Appendix I).

Overall, the differences between both approaches were very small with regard to the bias that arose from the reproduction. In other words, the differences between the two approaches were small in comparison to the differences observed between the paper and the reproduction (using the TOPHAT&CUFFLINKS) (Figure 7 - Appendix I).

Discussion

Our findings are consistent with those of Stavrinides et al. (2015) in that THAS is using compounds from the VVP pathway to generate a product different than vincristine/vinblastine. They proved that the alcohol dehydrogenase THAS is responsible for the conversion of strictosidine aglycone to the MIA tetrahydroalstonine, a monomeric MIA of the heteroyohimbine structural class. They also reported that THAS is a crucial example of a gene that acts at the critical branch point of SGD.

It is expected that the competition takes place at the level of SGD (Gongora-Castillo et al., 2012). Some arguments that support this are: the importance of SGD in VVP, the observed expression values and the physical distance between the genes. However, competition could also take place at the level of the ORCA2 or ORCA3 transcription factors. These regulate the expression of LAMT, SLS and several other genes in the MIA biosynthesis pathway. In any case, the hypothetical competing pathway is most likely incomplete. A more detailed study, taking into account the coordinates of the gene may reveal other MIA genes that play a competing role. A more scaffolded genome may also reveal more accurate information about competing pathway.

Biotechnology approaches can be used to knock out the proposed genes and obtain empirical prove of their role as competing genes. However, a complicating factor for gene discovery as well as biotechnological production is that the MIA pathway in *C. roseus* is organized in a complex manner, with the enzymes localized in different cell types and subcellular compartments (Miettienen et al., 2014).

One aspect that was not cleared with this study is the role that RO plays in the MIA pathway. It is known that RO is not involved in VVP, and in this study we have shown that RO is not part of the hypothetical competing pathway. In fact, the RO-like protein exhibits a different expression profile compared to its pathway neighbors THAS and SLS. Accounting for the physical distance between the genes may clarify this.

Replication VS alternative approach

Some possible explanations for the differences in heat map between both approaches are: in the case of the TOPHAT&CUFFLINKS it was possible to fix a minimum intron size of 5, whereas in hisat_2 the minimum values for this parameter is 20. Moreover, whereas with TOPHAT it is possible to specify segment length of 15 bp, with hisat_2 this parameter comes as a function of the length of the reads.
Another issue worthwhile mentioning is that the labeling of “5days” - ”12days” appeared to have swapped in one of the two approaches (TOPHAT&CUFFLINKS vs hisat_2&StringTie) (Figures 5 and 6 – Appendix I).

Conclusions

We have identified 5 genes that could be part of a competing pathway with VVP (Appendix I – Table 1). We speculate that these genes are recruiting SLS, which is a key compound for VVP and therefore, they may be responsible for the low concentration of VVP. The proposed competing pathway seems to be in any case incomplete. Future research considering the coordinates and physical distance could reveal the other genes that are within the suggested competing pathway. A more scaffolded genome

annotation will also throw some light on this. *In vitro* experiments could follow to knock out the suggested competing genes in order to obtain empirical proof of the role of these 5 genes. Through breeding practices or biotechnology it should be feasible to generate VVP in a synthetic way.

Appendix I – Supplementary material

VVP

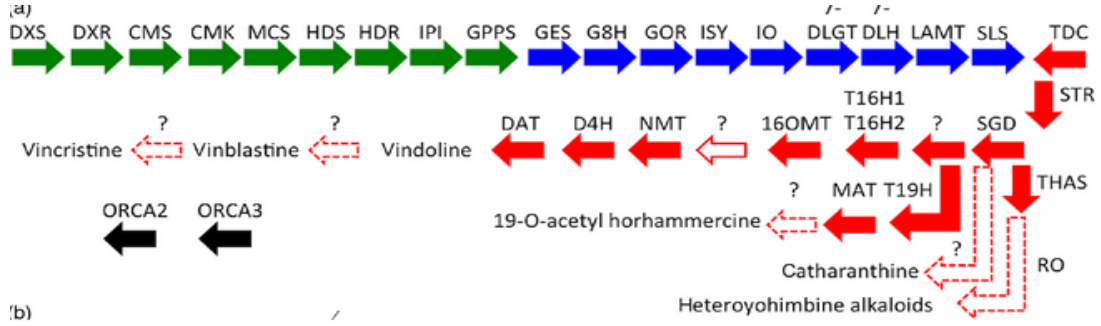

Figure 1. VAV genes as reported in Kellner et al. (2015) – Figure 2.

The compounds MAT, T19H, THAS and RO may be preventing Vincristine and Vinblastine from capturing enough SGD.

Suggested genes for the competing pathway of VVP

accesion no.	gene name	abbreviation
CRO_005213	AF253415.1	minovincinine 19-hydroxy-O-acetyltransferase MAT
CRO_021082	HQ901597.1	tabersonine/lochnericine 19-hydroxylase (CYP71BJ1) T19H
CRO_024553	KM524258	tetrahydroalstonine synthase THAS (nucleotide ID, is the same as THAS mentioned previously.
CRO_024552	KM524259	reticuline oxidase like-protein RO (not known before?)
CRO_024556	SLS_4	

Tabla 1. Genes that could be part of the hypothetical competing pathway of VVP.

The 5 proposed genes are part of the of the horhammercine and heteroyohimbine pathways. The expression values of these genes differ significantly from the expression values of the VVP genes under no MeJA treatment. However, after applying the treatment, the expression values become equal (2 sample independent t-test. See Excel file). This would explain why the production of vincristine and vinblastine increased after the treatment was applied.

Physical distance of genes in MIA. VVP vs competing pathway

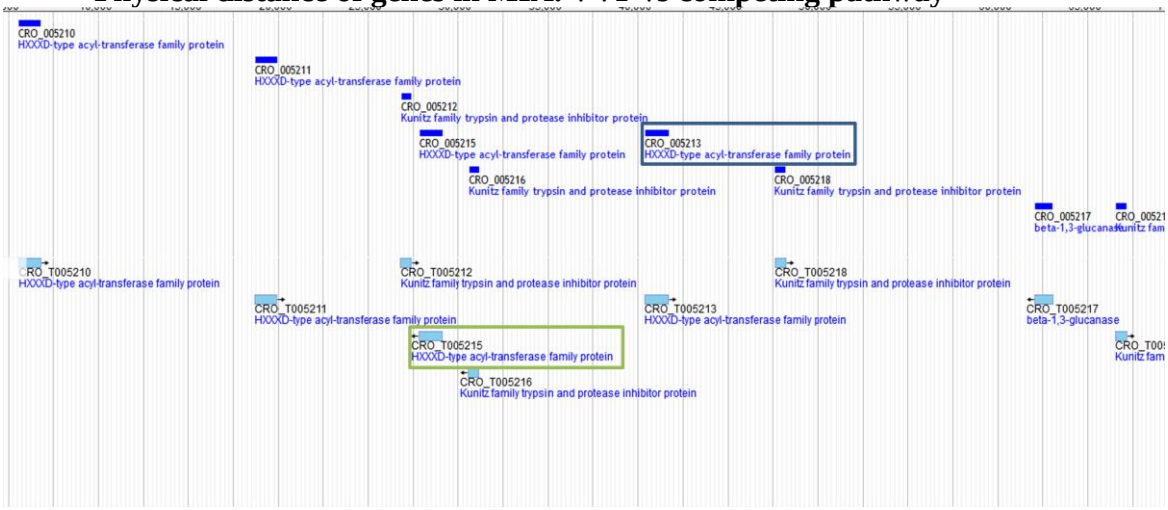

Figure 2. Genome viewer showing physical distance between some Vincristine Vinblastine genes and one of the suggested competing genes (AF253415 - MAT).

The suggested competing gene (AF253415 - MAT) is located in the same scaffold than one of the key genes for the synthesis of the Vincristine Vinblastine (CRO_005215). This could be regarded as an indicator that both genes are involved in a given process. Possibly, both genes are using some of the compounds that are essential for the syunthesis of Vincristine Vinblastine (i.e. both genes could be using SGD).

Genes considered in the expression analysis

Accession no.	Abbreviation
CRO_005213	**minovincinine 19-hydroxy-O-acetyltransferase**
CRO_005210	HXXXD-type acyl-transferase family protein
CRO_005211	HXXXD-type acyl-transferase family protein
CRO_005214	Kunitz family trypsin and protease inhibitor protein
CRO_005215	HXXXD-type acyl-transferase family protein
CRO_005216	Kunitz family trypsin and protease inhibitor protein
CRO_005217	beta-1,3-glucanase
CRO_005218	Kunitz family trypsin and protease inhibitor protein

Tabla 2. Genes located in the same scaffold as the gene of MAT. In bold, the gene responsible for MAT. All the genes in the table are located in scaffold no. 3067490.

Accession no.	Abbreviation
CRO_021082	**tabersonine/lochnericine 19-hydroxylase (CYP71BJ1)**
CRO_021078	DEAD box RNA helicase (PRH75)
CRO_021079	Peptidase family M48 family protein
CRO_021080	Tetratricopeptide repeat (TPR)-like superfamily protein
CRO_021081	cytochrome P450, family 704, subfamily A, polypeptide
CRO_021083	hypothetical protein

Tabla 3. Genes located in the same scaffold as the gene of T19H. In bold, the gene responsible for T19H. All the genes in the table are located in scaffold no. 2964965.

Accession no.	Abbreviation
CRO_024552	**reticuline oxidase like-protein**
CRO_024553	**tetrahydroalstonine synthase**
CRO_024554	hypothetical protein
CRO_024555	hypothetical protein
CRO_024556	**secologanin synthase-like protein 3**
CRO_024557	hypothetical protein
CRO_024558	hypothetical protein
CRO_024559	hypothetical protein

Table 4. Genes located in the same scaffold as the gene of RO(top), THAS(second) and SLS4. In bold, the gene responsible for the components. All the genes in the table are located in the scaffold no. 3063455.

MIA pathway genes:SGD,DAT, NMT, D4H, T16h1, T16H2, 16OMT

Accession no.	Abbreviation	Contig no.
CRO_020280	deacetylvindoline 4-O-acetyltransferase	3060125
CRO_033266	16-hydroxy-2,3-dihydro-3-hydroxytabersonine N-methyltransferase	3065019
CRO_012504	desacetoxyvindoline-4-hydroxylase	2969470
CRO_017447	T16H-like protein	3064268
CRO_017448	tabersonine 16-hydroxylase CYP71D12	3064268
CRO_004356	16-hydroxytabersonine O-methyltransferase	3051716

Table 5. Main VVP genes.

Relative increase in expression VVP vs competing pathway

	ID	relative increase 5d compared to MIA 5d
	MAT	**FPKM**
Contig no. 3063455	CRO_005210	-68,53256138
	CRO_005211	-98,23943398
	CRO_005213	**-8,954009342**
	CRO_005214	-75,58806805
	CRO_005215	-35,70292957
	CRO_005216	-91,38611869
	CRO_005217	-99,58741023
	CRO_005218	0
	T19H	
Contig no. 2964965	CRO_021078	-88,44368327
	CRO_021079	-98,74486397
	CRO_021080	-98,34819805
	CRO_021081	-11,57306448
	CRO_021082	**-99,95487171**
	CRO_021083	0
	RO(top), THAS(second),SLS4	
Contig no. 3063455	**CRO_024552**	**-47,55522082**
	CRO_024553	**-98,77859093**
	CRO_024554	63,07897143
	CRO_024555	-95,08230174
	CRO_024556	**-99,37882754**
	CRO_024557	0
	CRO_024558	-70,82942354

Tabla 6. Lower relative increase in the competing genes. In bold, the genes corresponding to the compounds.

FPQM values were calculated with the formula:

$$\text{Relative increase of 5d compared to MIA} = \frac{FPQM_{competinmg\ gene,5D}}{\overline{FPQM_{VAV\ 5D}}} - 1*100$$

The suggested competing genes (bold) have a low relative increase with respect to the increase of the MIA genes. However, the absolute increase was higher for two of the suggested competing genes (Figure 3 –Appendix I).

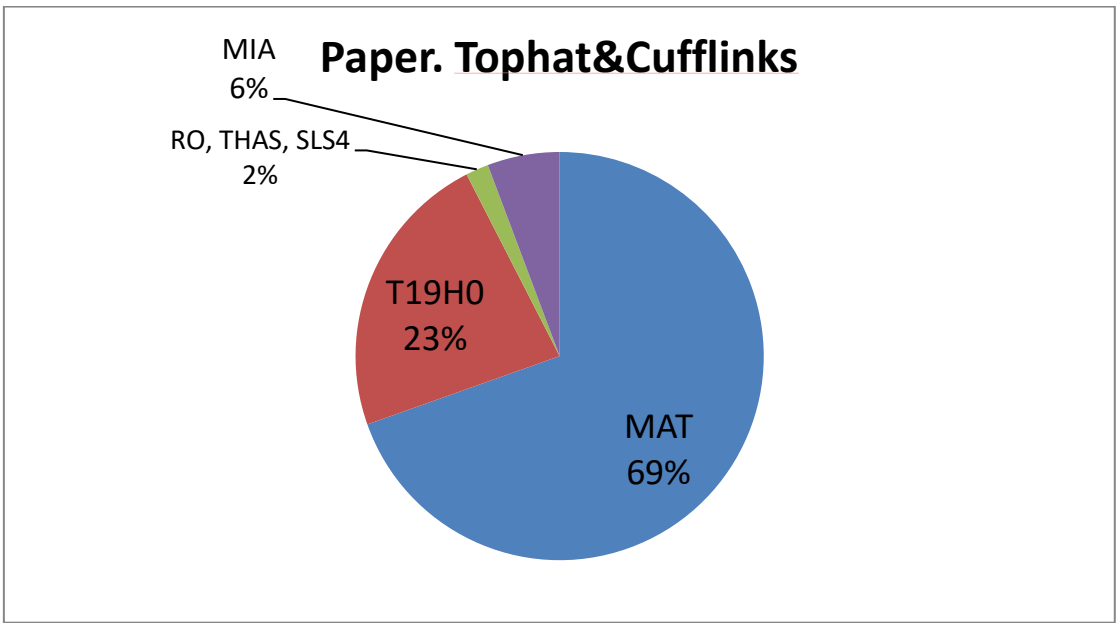

Figure 3. Proportional increase in FPQM

The highest increase is for two of the competing genes (MAS and T19H). However, it has been observed that overall VVP genes are more expressed than the overall genes of the hypothetical competiove pathway (Gongora-Castillo et al., 2012). From this, we deduce that the competing pathway is still incomplete. A similar pie plot can be found in the excel file for the Hisat_2&StringTie approach.

Tophat&cufflinks vs Hisat_2&StringTie

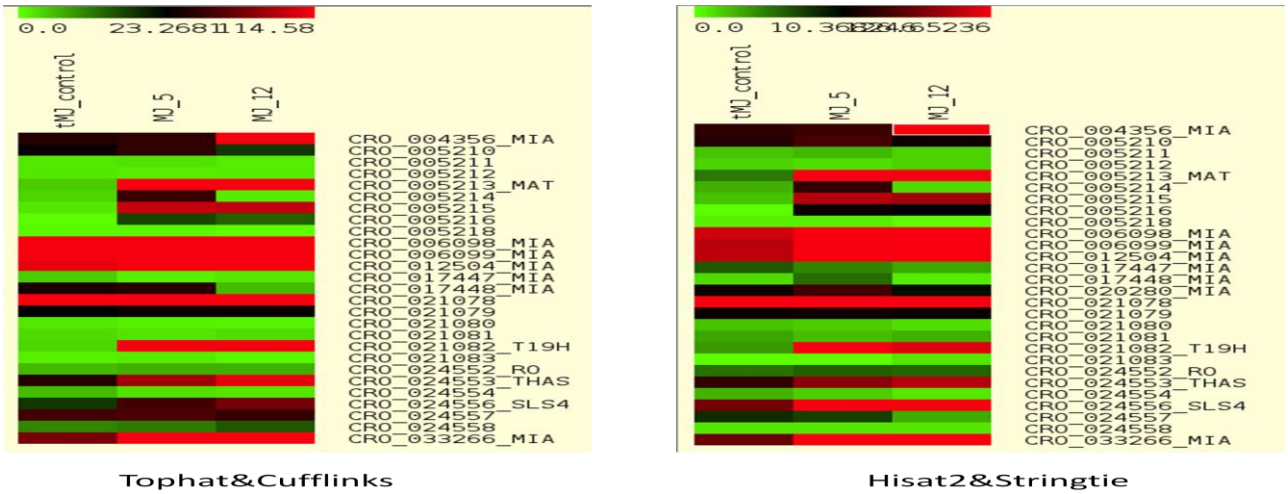

Figure 4. Heat maps to compare TOPHAT&CUFFLINKS and Hisat_2&StringTie

The gene with accession number CRO_017448 appeared to be down-regulated in the TOPHAT approach, whereas the same was not observed for Hisat_2.

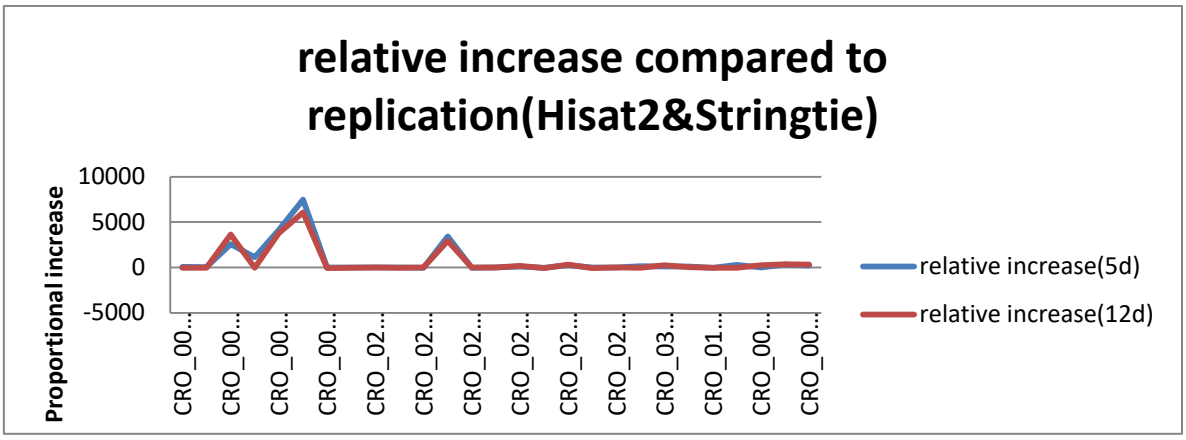

Figure 5. Figure 4. Relatib¡ve increase (FPKM) after MeJA treatment. Hoisat2&StringTie approach

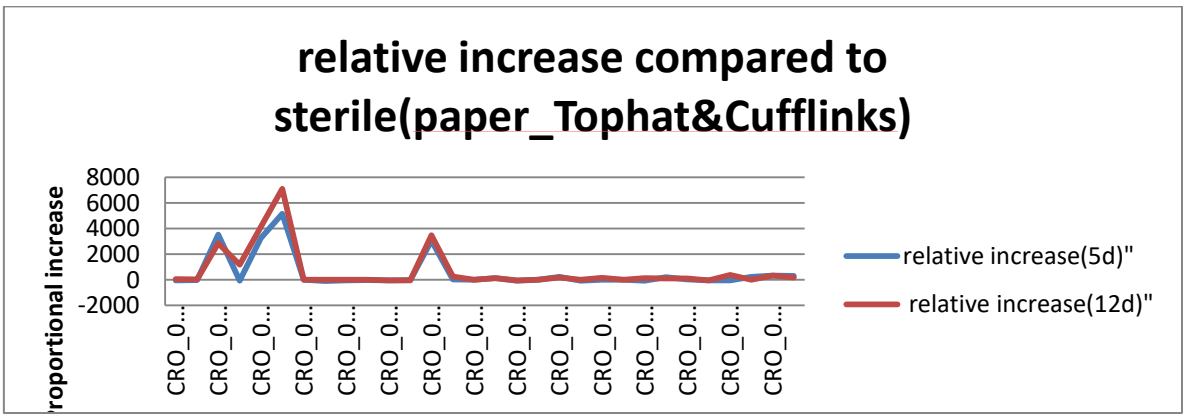

Figure 6. Relative increase (FPKM) after MeJA treatment. Tophat&Cufflinks approach

Form the Figure s it can be observed that both approaches lead to very similar expression values, but that the labeling (5 dats/12 days), may be swapped in one of the cases.

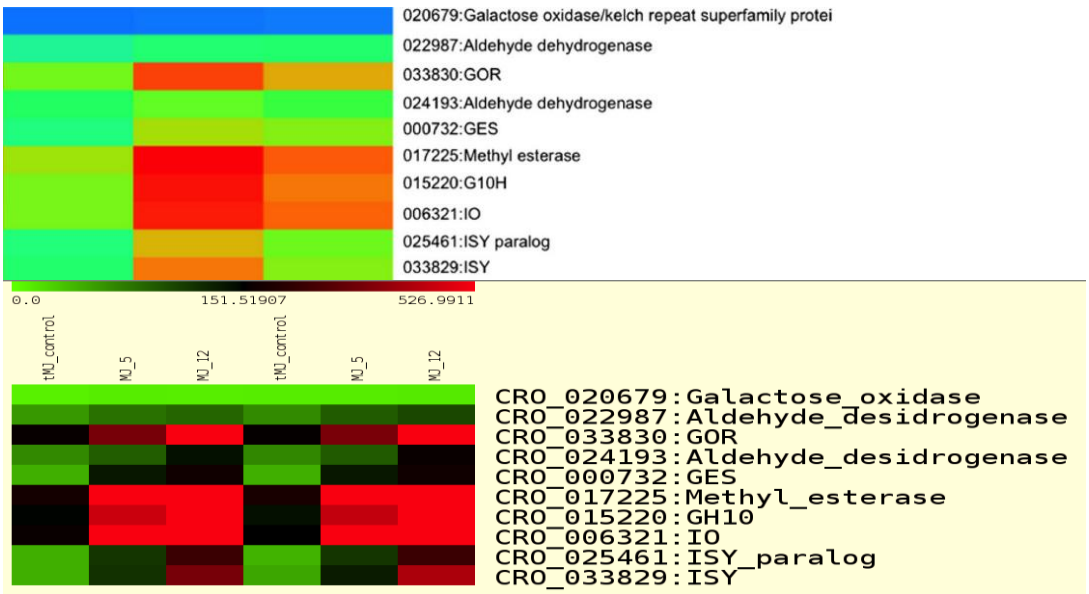

Figure 7. Heat map original (TOPHAT-above) vs replication (TOPHAT-bellow left) vs replication (Hisat_2 - bellow right).

The difference between both approaches was smaller than original vs reproduction.

Appendix II – Statistics

Relative increase with respect to the mean:

$$\text{Relative increase} = \frac{\frac{FPQM_{Control}}{FPQM_{Treatment}} - 1*100}{\frac{FPQM_{Control}}{FPQM_{Treatment}} - 1*100} - 1*100$$

Average of relative increase:

$$(\frac{FPQM_{Control}}{FPQM_{Treatment\ 5d}} - 1*100 + \frac{FPQM_{Control}}{FPQM_{Treatment\ 12d}} - 1*100)/2$$

References

Amoutzias, G. and Van de Peer, Y. (2008) Together we stand: genes cluster to coordinate regulation. Dev. Cell, 14, 640–642.

Field, B. and Osbourn, A.E.(2008) Metabolic diversification- independent assembly of operon-like gene clusters in different plants. Science, 320, 543–547.

Frey, M., Chomet, P., Glawischnig, E. et al. (1997) Analysis of a chemical plant defense mechanism in grasses. Science, 277, 696–699.

Gongora-Castillo, E., Childs, K.L., Fedewa, G.et al. (2012) Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. PLoS One, 7, e52506.

Itkin, M., Heinig, U., Tzfadia, O. et al. (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science, 341, 175–179.

Kellner, F., Kim, J., Clavijo, B. J., Hamilton, J. P., Childs, K. L., Vaillancourt, B., Cepela, J., Habermann, M., Steuernagel, B., Clissold, L., McLay, K., Buell, C. R. and O'Connor, S. E. (2015), Genome-guided investigation of plant natural product biosynthesis. Plant J, 82: 680–692. doi:10.1111/tpj.12827

Miettinen, Karel, Lemeng Dong, Nicolas Navrot, Thomas Schneider, Vincent Burlat, Jacob Pollier, Lotte Woittiez, et al. 2014. "The Seco-iridoid Pathway from Catharanthus Roseus." Nature Communications5.

Mugford, S.T., Louveau, T., Melton, R. et al. (2013) Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat. Plant Cell, 25, 1078–1092

O'Connor SE, Maresh JJ (2006) Chemistry and biology of monoterpene indole alkaloid biosynthesis. Natural Product Reports 23: 532–547.

Qu Y, Easson MLAE, Froese J, Simionescu R, Hudlicky T, De Luca V. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(19):6224-6229. doi:10.1073/pnas.1501821112.

Stavrinides A, Tatsis EC, Foureau E, et al. Unlocking the Diversity of Alkaloids in Catharanthus roseus: Nuclear Localization Suggests Metabolic Channeling in Secondary Metabolism. Chemistry & Biology. 2015;22(3):336-341. doi:10.1016/j.chembiol.2015.02.006.

Swaminathan, S., Morrone, D., Wang, Q., Fulton, D.B. and Peters, R.J.(2009) CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. Plant Cell, 21, 3315–3325.

Vazques-Flota, F.A. and De Luca, V. (1998) Jasmonate modulates development and light regulated alkaloid biosynthesis in Catharanthus roseus. Phytochemistry, 49, 395–402.

Winzer, T., Gazda, V., He, Z.et al. (2012) A Papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science, 336, 1704–1708.