

## SHORT TERM SCIENTIFIC MISSION (STSM) SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

**Action number: CA15110**

**STSM title: Integration of GoMapMan and PaintOmics**

**STSM start and end date: 01/10/2017 to 30/10/2017**

**Grantee name: Dr Marko Petek**

### PURPOSE OF THE STSM:

The aim of the STSM was to investigate the possibility of integrating GoMapMan plant ontologies and MapMan biological pathways into PaintOmics, a pathway analysis and visualization web application developed by the host group.

### DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

Following the proposed work plan, I had a meeting first with the development team and we discussed the structure of the GoMapMan ontology, MapMan pathway files and Paintomics database architecture. At the meeting, we have defined the procedure and file formats for the integration.

The first task was to expand the database if gene identifiers that PaintOmics recognises to incorporate PGSC and iTAG potato gene models. After completing this task, I've worked closely with one of PaintOmics developers to integrate the GoMapMan ontology for *Arabidopsis thaliana* and potato (*Solanum tuberosum*) into their databases. Since GoMapMan does not have an API, we prepared specifications for GoMapMan export files to feed PaintOmics databases and requested the GoMapMan developers at NIB to produce these files and make them available online. We have tested the integration of GoMapMan ontologies by importing Arabidopsis and potato transcriptomics datasets into PaintOmics.

With the identifier and ontology in place, we proceeded to integrating pathways. Because visualization of pathways in PaintOmics and MapMan applications is inherently different i.e. MapMan visualizes groups of genes in a similar manner that PaintOmics visualizes conditions, we decided to integrate only pathways that use the most specific BINs in the ontology and thus include only few genes. The visualization was tested using a dummy Arabidopsis and potato transcriptomics dataset.

I have also inspected KEGG pathways already integrated in PaintOmics as well as MapMan pathways for overlap and missing or incomplete plant specific pathways.

### DESCRIPTION OF THE MAIN RESULTS OBTAINED

We have integrated potato gene model identifiers, GoMapMan ontology and selected MapMan pathways into PaintOmics web application (see screenshots below). We established interoperability of GoMapMan

and PaintOmics by producing GoMapMan web application export files that are directly used to feed the PaintOmics database.

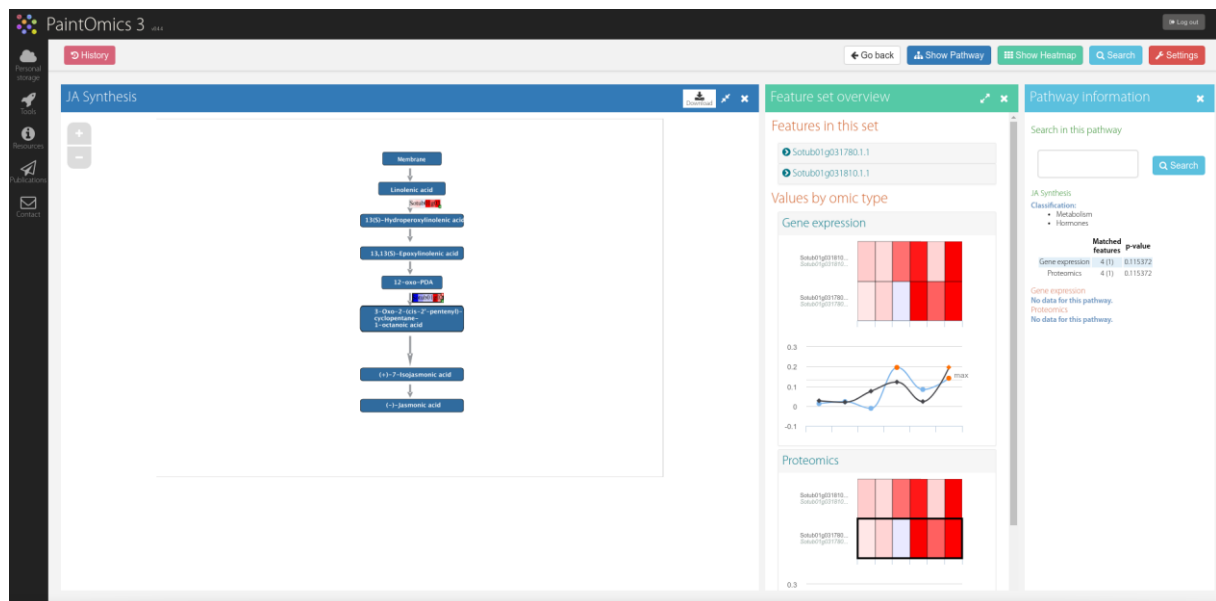
Currently the application is still at the development stage and therefore not yet publicly available. When finalized and complemented with additional pathways, the application will provide users the option to visualize their high-throughput omics data in the context of KEGG and MapMan pathways. By identifying gaps and integration challenges, we also contribute to standardization of pathway-based omics data analysis.



**Matched Pathways**

Search:   Regular expression  Case sensitive Databases to view:  KEGG  MapMan [Download as XLS](#)

Paint	Pathway name	Features		Significance tests			External links
		Un... ge...	Un... me...	Gene expression	Expression 2	Combined P-values	
<input checked="" type="checkbox"/> (K)	Arginine and proline metabolism	5	0	0.00854	1.5176e-4	1.8808e-3	<a href="#">KEGG</a> <a href="#">Q</a> <a href="#">PubMed</a>
<input checked="" type="checkbox"/> (K)	Isoquinoline alkaloid biosynthesis	1	0	0.03018	9.1785e-4	3.1836e-4	<a href="#">KEGG</a> <a href="#">Q</a> <a href="#">PubMed</a>
<input checked="" type="checkbox"/> (K)	One carbon pool by folate	2	0	0.05945	0.00529	0.00285	<a href="#">KEGG</a> <a href="#">Q</a> <a href="#">PubMed</a>
<input checked="" type="checkbox"/> (K)	Glutathione metabolism	15	0	0.07358	0.01228	0.00723	<a href="#">KEGG</a> <a href="#">Q</a> <a href="#">PubMed</a>
<input checked="" type="checkbox"/> (K)	Phenylalanine metabolism	3	0	0.08784	0.01270	0.00870	<a href="#">KEGG</a> <a href="#">Q</a> <a href="#">PubMed</a>
<input checked="" type="checkbox"/> (K)	Cyanoamino acid metabolism	3	0	0.08784	0.01270	0.00870	<a href="#">KEGG</a> <a href="#">Q</a> <a href="#">PubMed</a>
<input checked="" type="checkbox"/> (K)	Steroid biosynthesis	3	0	0.08784	0.01270	0.00870	<a href="#">KEGG</a> <a href="#">Q</a> <a href="#">PubMed</a>
<input checked="" type="checkbox"/> (M)	JA Synthesis	4	0	0.11537	0.02276	0.01623	<a href="#">KEGG</a> <a href="#">Q</a> <a href="#">PubMed</a>



**PaintOmics 3**

History [Go back](#) [Show Pathway](#) [Show Heatmap](#) [Search](#) [Settings](#)

**JA Synthesis**

**Feature set overview**

Features in this set

- Scotub1g31780.1.1
- Scotub1g31810.1.1

Values by omic type

Gene expression

Proteomics

**Pathway information**

Search in this pathway

JA Synthesis  
Classification:  
• Metabolism  
• Hormones

Matched features P-value  
Gene expression: 4 (3) 0.11537  
Proteomics: 4 (3) 0.11537

No data for this pathway.  
Proteomics:  
No data for this pathway.

### **FUTURE COLLABORATIONS**

While integrating MapMan pathways we realized that for more informative visualization in PaintOmics the pathways will have to be populated with more specific BINs, preferably such including only paralogs/homeologs of one gene or a small gene family. Additionally, pathways not present in KEGG or MapMan but valuable for the plant omics community were identified and wheels were set in motion to complement existing or produce new MapMan pathways. This requires a coordinated effort by the developers of MapMan/GoMapMan and PaintOmics ensuring future collaborations.

Additionally, we collaborate with prof. Conesa on PacBio IsoSeq data analysis where we will also use the bioinformatics tools developed by her group to get insight into potato plant's regulation at the transcript isoform level.