LC-MS and GC-MS metabolite data processing using Mass Profiler Professional, a chemometric data analysis and visualization tool to determine metabolomic pathways

Syed Salman Lateef; Siji Joseph; Nilanjan Guha; Yugandhar Reddy; Sudha Rajagopalan
Agilent Technologies, Bangalore, Karnataka India.

HUPO 2010
Poster # PO 244

Introduction

Metabolomics in combination with proteomics and genomics helps to better understand biological processes. In this study, the effect of rapamycin drug on HEK cell line is looked at the view point of all three Omics. Rapamycin is an immunosuppresant drug which was recently shown to inhibit cancer growth. In this work, metabolites mass spectral data acquisition is obtained from a orthogonal studies using Liquid chromatographic Mass Spectrometry (LCMS) in both positive and negative mode and Gas chromatographic Mass Spectrometry (GCMS). Such multiplatform data acquisition helps in increase the coverage of metabolites. Large amount of data is generated from metabolobics study and effective tools are needed to process, analyze and interpret the data. Mass Profiler Professional is a chemometric data analysis and visualization tool that is used in this study to process metabolomics data. MPP is used to compare and analyze sample groups. A differential list of metabolites were obtained following statistical analysis and fold change. This list is used to arrive at a significant biological pathways that are contributed by the metabolites. This data is compared against data obtained from genomics and proteomics parallel studies to narrow down potential biological processes affected on rapamycin treatment.

Experimental

Study Design and cell culture conditions:
Please see poster number PO034, Tuesday session HUPO2010.

Metabolomics Workflow: **Figure 1**

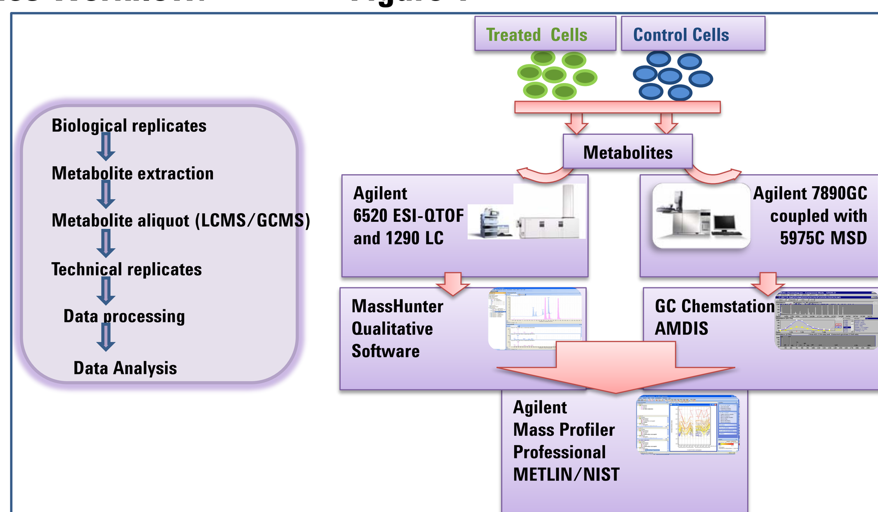


Figure 1. Shows the untargeted metabolomics workflow used in this study. Both control and treated cells are subjected to metabolite extraction followed by LCMS and GCMS analysis. The data for LCMS is processed using Agilent MassHunter Qualitative software while GCMS data is processed using Agilent GC Chemstation followed by deconvolution and interpretation using AMDIS. Both data are fed into MPP to obtain statistically differential list of metabolites. The metabolites are annotated with CAS number using METLIN and NIST databases linked to MPP. The annotated metabolites are searched for relevant biological processes and pathways.

Extraction of metabolites and data acquisition:
Metabolites were extracted from one million cells with 2:1:2 methanol-water/chloroform mixture. The aqueous portion was dried and used for LCMS experiment using Agilent 6520 QTOF coupled to Agilent 1290 Liquid Chromatographic system. LCMS separation was carried out on a 2.1x50 mm, 1.8 um SB-Aq column having a guard column (2.1X30 mm, 3.5 um SB-C8). A sub2 micron particle size column helps increase the peak capacity. A linear gradient from 2% methanol to 98% methanol is performed in 13 min, held at 98% methanol for 6 min. Both mobile phase having 0.2% acetic acid. Flow rate at 0.6 ml/min. Data acquisition was performed using ESI source in both positive and negative mode. 5 uL of Epicatechin (0.02 ppm) was added into the extraction vial to normalize the intensity of the technical replicates. For each biological replicate five technical replicates were performed. For 16 hrs data presented here, five biological replicates were pooled.
The organic phase after drying is derivatized in two steps for GCMS analysis. First, using methoxyamine hydrochloride in pyridine to derivatize carbonyl groups, then using N-methyl- N trimethylsilyltrifluoroacetamide containing 1% trimethylchlorosilane to derivatize carboxyl, hydroxyl, amino, imino or sulphonyl groups. d27-myristic acid, was added as internal standard for normalization of technical replicates. GC-MS analysis is performed using Agilent 7890 GC coupled to Agilent 5975C inert MSD. For each biological replicate five technical replicate were performed. For GCMS a 30 m long Agilent DB-5MS column with 10 m Duragard precolumn (Part Number: 122-5532G) with 0.25 µm film thickness and 250 µm diameter was used for the separation. All GC-MS experiments were performed as described in the user manual (G1676-80000). See Ref (1 and 2).
Data analysis:
Statistical comparison between control and treated samples from both LCMS and GCMS were performed using Agilent Mass Profiler Professional (MPP). Differentially expressed metabolites were identified using Agilent Fiehn GC/MS Metabolomics RTL Library (version June 2008) in case of GCMS data and using METLIN database for LCMS data. LCMS and GCMS differential list were pooled and added into MPP as a single generic experiment. This pooled list was used to arrive at significant pathways ($p<0.05$) that show metabolite enrichment. The significant pathways generated were compared against significant pathway study performed separately on genomics and proteomics on the same samples.

Results and Discussion

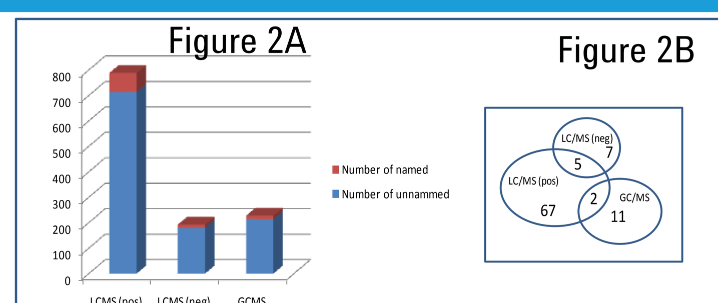


Figure 2A shows the number of entities that pass the quality control of samples in data analysis (see workflow 2). >700 compounds are detected by LCMS positive more while close to 200 are detected by GCMS method. The number of entities that are identified are smaller in number. Their total count and the overlap is shown in **figure 2B**. All entities detected by LCMS and GCMS play an important role in determining the significant pathways.

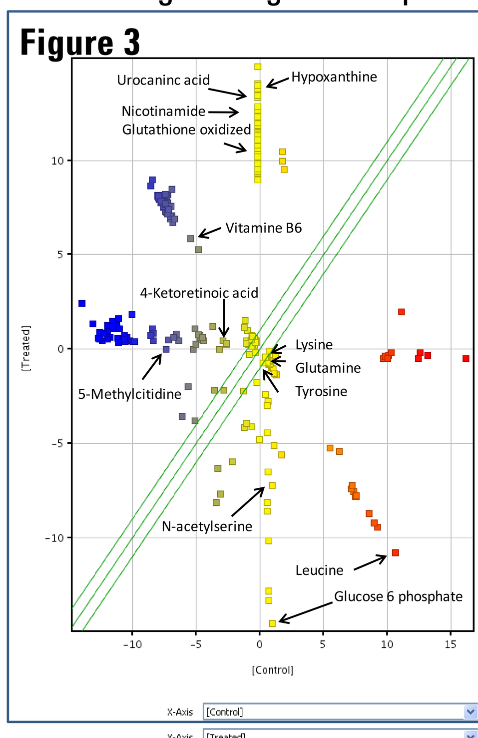
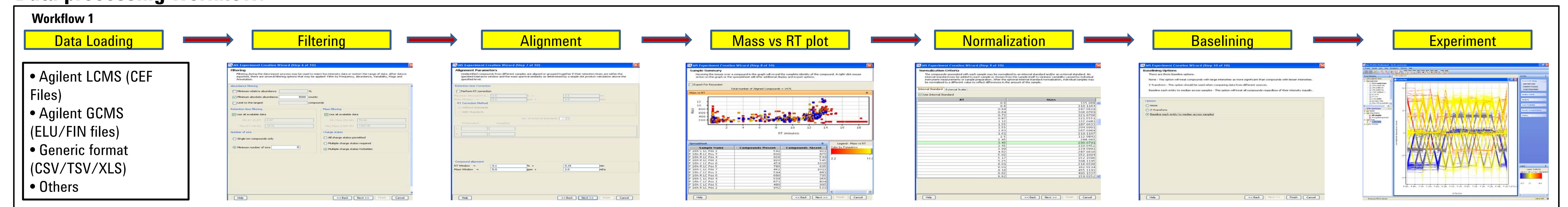


Figure 4 shows the PCA plot of samples for 16 hrs LCMS positive runs after the statistical analysis. The blue dots represent the treated samples while red dots are the untreated. PCA plot show that treated and control samples cluster in separate group. The x axis component 1 is 81%, y axis component 2 is 9% and Z axis component 3 is 7%

Figure 3 shows the scatter plot of rapamycin treated samples vs the control samples. This scatter plot represent the compounds found in 16 hrs pos LCMS experiment. Vitamin B6, Nicotinamide, and oxidized glutathione are some of the compounds found to be significantly higher in treated samples and not detected in control samples

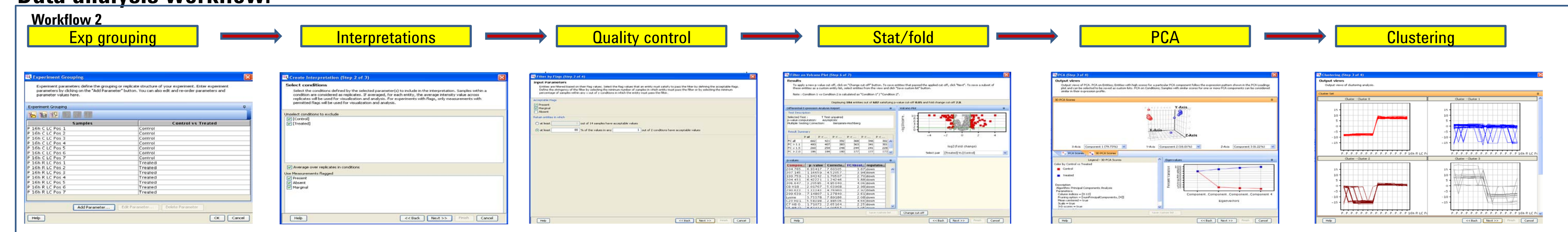
Workflows

Data processing workflow:



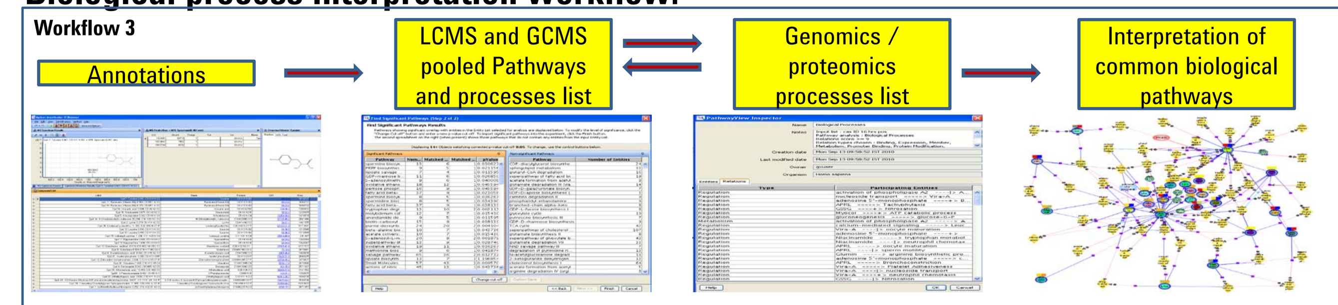
Workflow 1. The data processing workflow includes data loading where data is uploaded into MPP using any of the file formats based on the data acquisition instrument. The filtering is performed to reject low intensity data and restrict the range of data. The alignment aligns the compounds from different samples based on retention time and mass . The mass vs RT plot shows the total number of aligned compounds from all samples in a given experiment. The intensity of the compounds are normalized based on the internal standard (Epicatechin for LCMS experiment and d27-myristic acid for GCMS experiment). The baseline is chosen based on how the compounds are to be treated. Following the data processing, the experiment is set up. For metabolomics data processing information see ref 3.

Data analysis workflow:



Workflow 2. The data analysis workflow includes grouping of the data to perform the analysis. The groupings include the segregation of the samples into groups of samples treated with rapamycin and not treated. The data interpretations allow samples to be averaged or non-averaged depending on specific type of analysis. The quality control filters out those compounds that do not occur consistently in a groups and therefore corrects for variations in technical replicates. To find entities that are differentially abundant between two sets of control and treated samples, t test is performed followed by fold change which tells by how much the compounds are abundant in the differential list. The PCA plots is another tool available in the workflow. The clustering feature of MPP such as K means clustering is used to cluster compounds into groups that behave similarly.

Biological process interpretation workflow:



Workflow 3. The biological process interpretation workflow uses ID browzer feature of MPP to annotate compounds from METLIN/NIST/FIEHN libraries using accurate mass and formula generator feature. The determine pathways that are significantly enriched by the metabolites, LCMS and GCMS data are pooled into one experiment to determine the significant pathway list. The next step is to compare pathways or biological processes list obtained from GeneSpring GX for genomics and proteomics data with that from metabolomics study to arrive at common pathways. The common pathways were arrived by looking for differential metabolites in the pathways from genomics and proteomics.

Results and discussion

Biological interpretation:
Table 1

Compound name	FCAbsolute	Regulation	Mass	Retention time	CAS Number
N-Acetylserine	3.4	DOWN	147.0536	0.342	16354-58-8
3-hydroxy-3-methyl-Glutaric acid	2.9	DOWN	162.0514	0.35	503-49-1
Glutathione, oxidized	16	UP	612.1522	0.619	27025-41-8
Nicotinamide N-oxide	16	UP	138.0428	0.604	1986-81-8
Niacinamide	16	UP	122.0478	0.875	98-92-0
Isobutyryl carnitine	16	UP	231.1468	1.236	25518-49-4
Adenosine	2.1	UP	267.0964	1.605	58-61-7
Palmitoleic acid	16	UP	254.2242	11.69	373-49-9
glucose 6-phosphate	3.3	DOWN	260.0274	0.36	56-73-5
D-Ribose	3.0	DOWN	73	24.26	50-69-1
D-(-)-Lactic acid	16	UP	147	7.769	10326-41-7
Fumaric acid	1.8	UP	155	11.57	110-17-8

Table 1. Shows the partial list of differential metabolites from the LCMS and GCMS experiment. The fold change the amount of fold change is provided. The CAS number is annotated using METLIN or NIST/FIEHN databases for LCMS and GCMS experiments respectively.

Table 2

Regulation	Gene Symbol	Name and comments	Metabolites found	Fold change of Metabolite
up	LMNA	Lamin A/C. Found in Cardiovascular Diseases; Hypertrophic cardiomyopathy	Adenylic acid (adenosine monophosphate)	UP
up	PGK1	Phosphoglycerate kinase 1. Found in glycolysis and gluconeogenesis	Lactic acid, glucose 6 phosphate	UP DOWN
up	OR5L2	olfactory receptor, family 5, subfamily L, member 2. Found in olfactory transduction	Adenylic acid (adenosine monophosphate)	UP
up	ABCA7	ATP-binding cassette, sub-family A (ABC1), member 7. Found in ABC transporters	Mannitol, Lactose and Ribose Spermidine Valine	DOWN SAME SAME
up	PLD3	phospholipase D family, member 3	Nicotinamide N-oxide;	UP
up	LPIN3	Lipin 3. Found in glycerophospholipid metabolism	1-alkyl-2-acyl-glycerophosphoethanolamine	UP
down	AHSA1	AHA1, activator of heat shock 90kDa protein ATPase homolog 1 (yeast)	Adenylic acid (adenosine monophosphate)	UP
down	CAMK2D	calcium/calmodulin-dependent protein kinase II delta	Nicotinamide N-oxide	UP
down	BCCIP	BRCA2 and CDKN1A interacting protein. Found in pathways of cancer	Fumaric acid	UP
down	PPAT	phosphoribosyl pyrophosphate amidotransferase. Found in purine metabolism and also in alanine, asparte and glutamine metabolism	l-glutamic acid	DOWN

Table 2. Shows the common list of expressed proteins from the combined genomics and proteomics data for 16 hrs time point. The metabolite found are the metabolites found in those pathways from the LCMS and GCMS pooled metabolomics data.

Results and discussion

Figure 5A

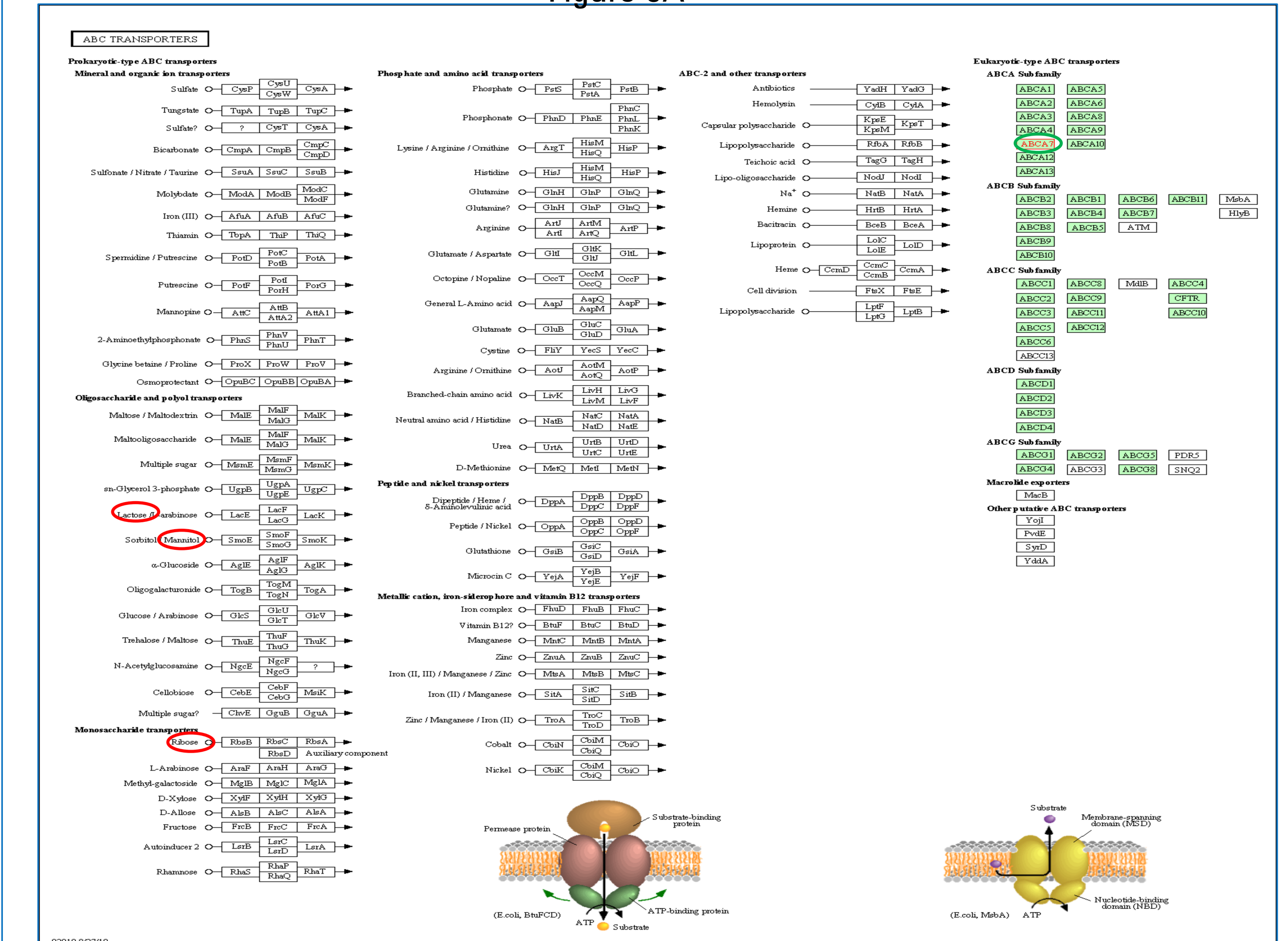


Figure 5 shows the KEGG pathways from the combined list of genomics and proteomics data as shown in Table 2. Figure 5A represent the ABC transporters pathway, figure 5B represent the pathways found in cancer. The red and green oval represent the down and up regulations of the proteins/enzymes/metabolites. The results of this experiments are similar to results presented earlier on rapamycin treatment (Ref 4) on Human BJAB lymphoma cells. Similar to this results, up regulation of genes involved in transport of amino acids were shown.

Conclusion

1. The untargeted metabolomics study for the treatment of rapamycin on HEK 293 cells was effectively performed using the combination of LCMS and GCMS experiments. The results show that both LCMS and GCMS experiments were needed to obtain comprehensive lists of metabolites.
2. Mass Profiler Professional was effectively used to produce statistically significant differential list of metabolites which was then used to determine significant pathways.
3. The metabolomics results were compared with genomics and proteomics results to arrive at possible pathways that maybe affected by rapamycin treatment.
4. The results obtained on repamycin treatment were similar to one reported earlier.
5. Such a study will help understand cellular dynamics and help identify potential biomarkers for diagnostics purposes
6. The combinations of Omics is a one of the important steps in a system biology approach to a problem.

Reference

1. S. Baumann, H. Prest, M. Jensen, M. Muehlbauer, J. Bain, "Non-targeted GC/MS metabolomics using a large-volume inlet, mid-column backflushing, and a retention time locked spectral library," *ASMS* poster. 2010 (searchable at Agilent website).
2. T. R. Sana, S. Fischer, G. Wohlgemuth, A. Katrekar, K. Jung, P.C. Ronald, O. Fiehn, " Metabolomic and transcriptomic analysis of the rice response to the bacterial blight pathogen Xanthomonas oryzae pv. Oryzae," *Metabolomics* 6:451–465. 2010
3. J. Boccard, J. Veuthey, S. Rudaz, "Knowledge discovery in metabolomics: An overview of MS data handling," *J. Sep. Sci* 33: 290-304. 2010
4. T. Peng, T.R. Golub, and D. M. Sabatini, "The Immunosuppressant Rapamycin Mimics a Starvation-Like Signal Distinct from Amino Acid and Glucose Deprivation," *Molecular and Cellular Biology*, August: 5575–5584. 2002