Standardizing

Gas Chromatography-Mass Spectrometry

Metabolomics

Maria I. Klapa

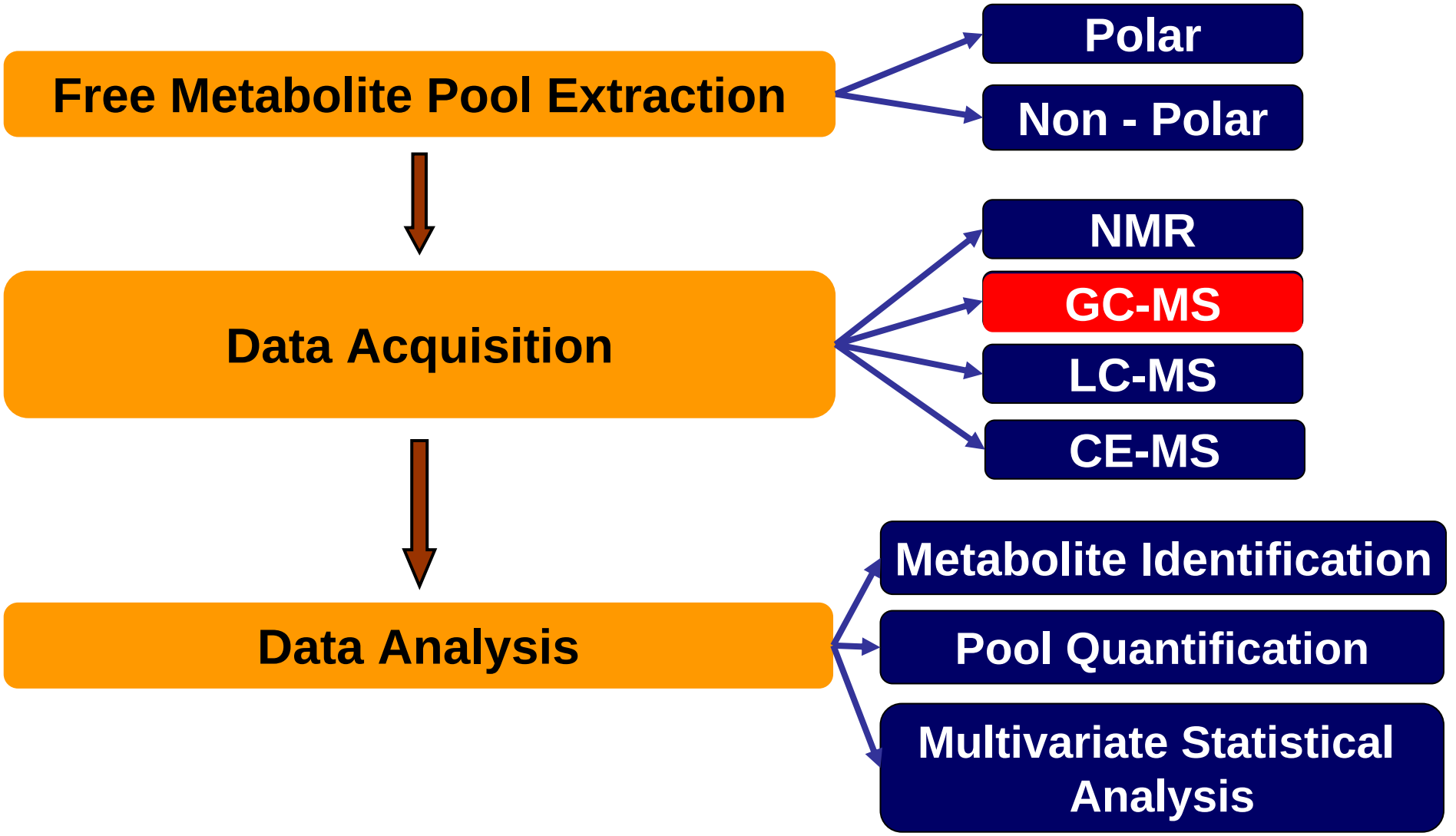
Metabolic Engineering and Systems Biology Laboratory

Institute of Chemical Engineering and High Temperature Chemical Processes,

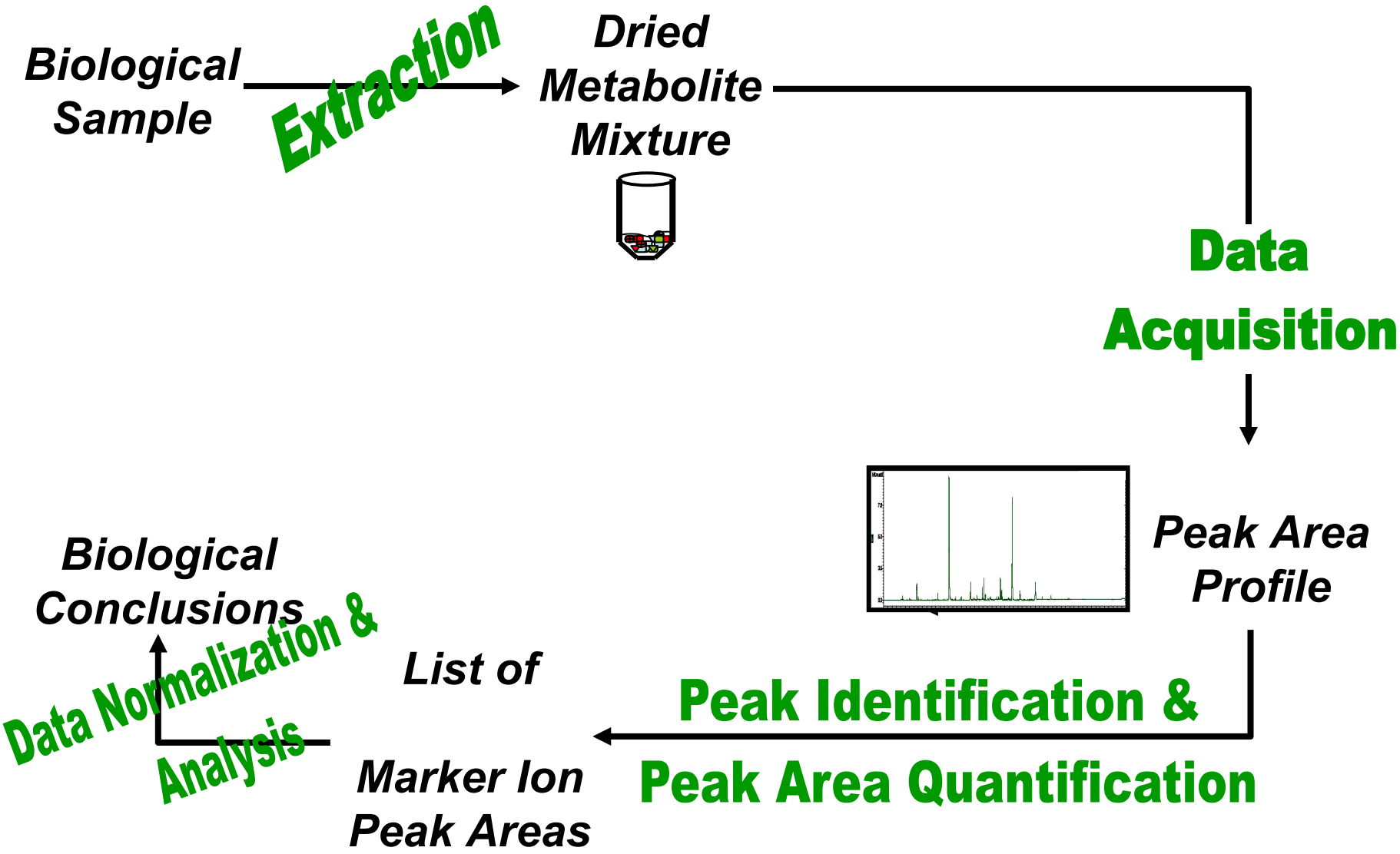
Foundation for Research and Technology-Hellas (FORTH),

Patras, GREECE

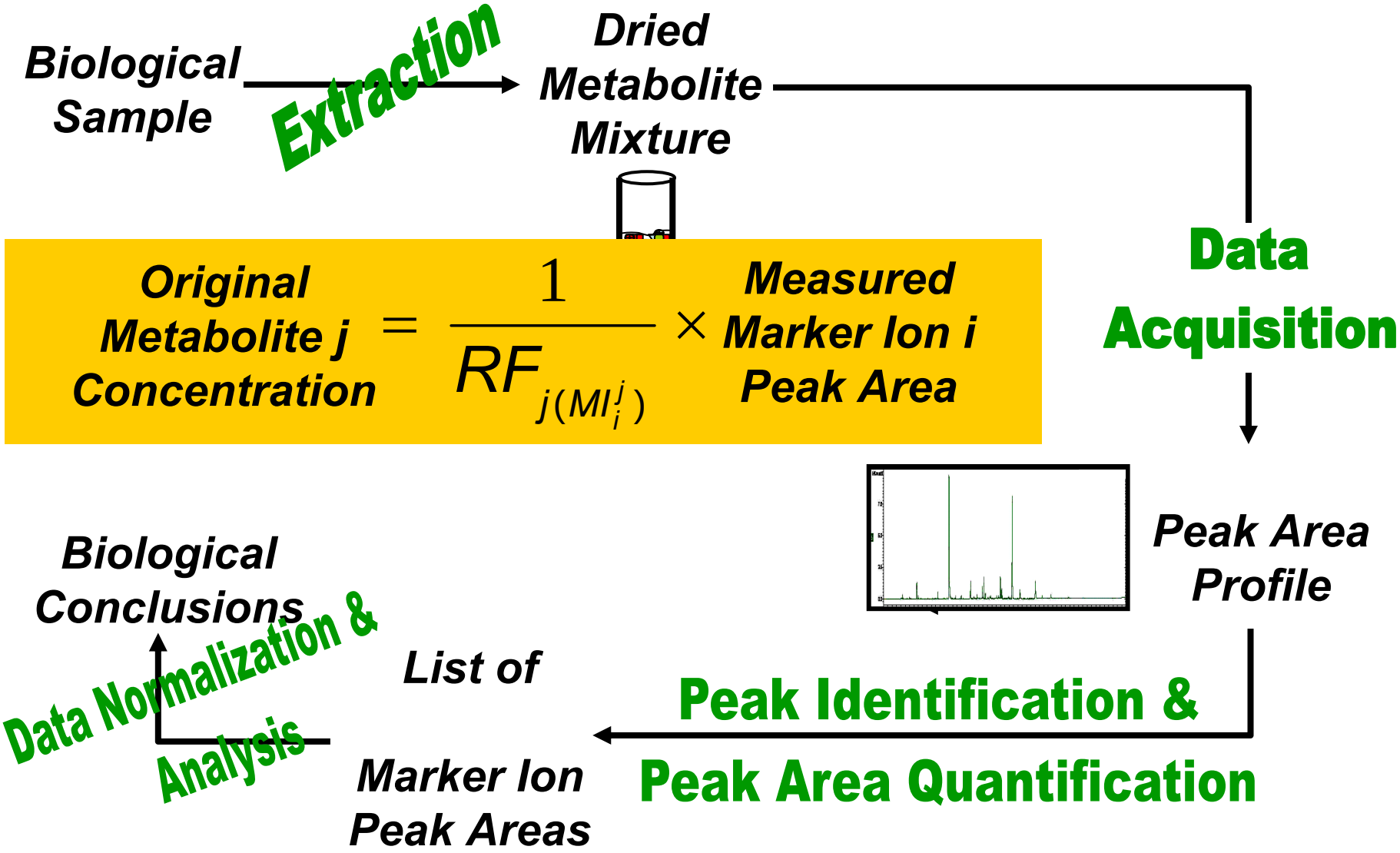
Metabolomic Profiling: a multi-step procedure



Schematic Diagram of Metabolomic Analysis



Schematic Diagram of Metabolomic Analysis



Internal Standard Normalization

- only biases that change RF **to the same extent for all metabolites** (Type A) might be present

  e.g. variation in the injected volumes, variation in drying, variation in replicate division, &

- Equipment's operating conditions remain **constant** among runs

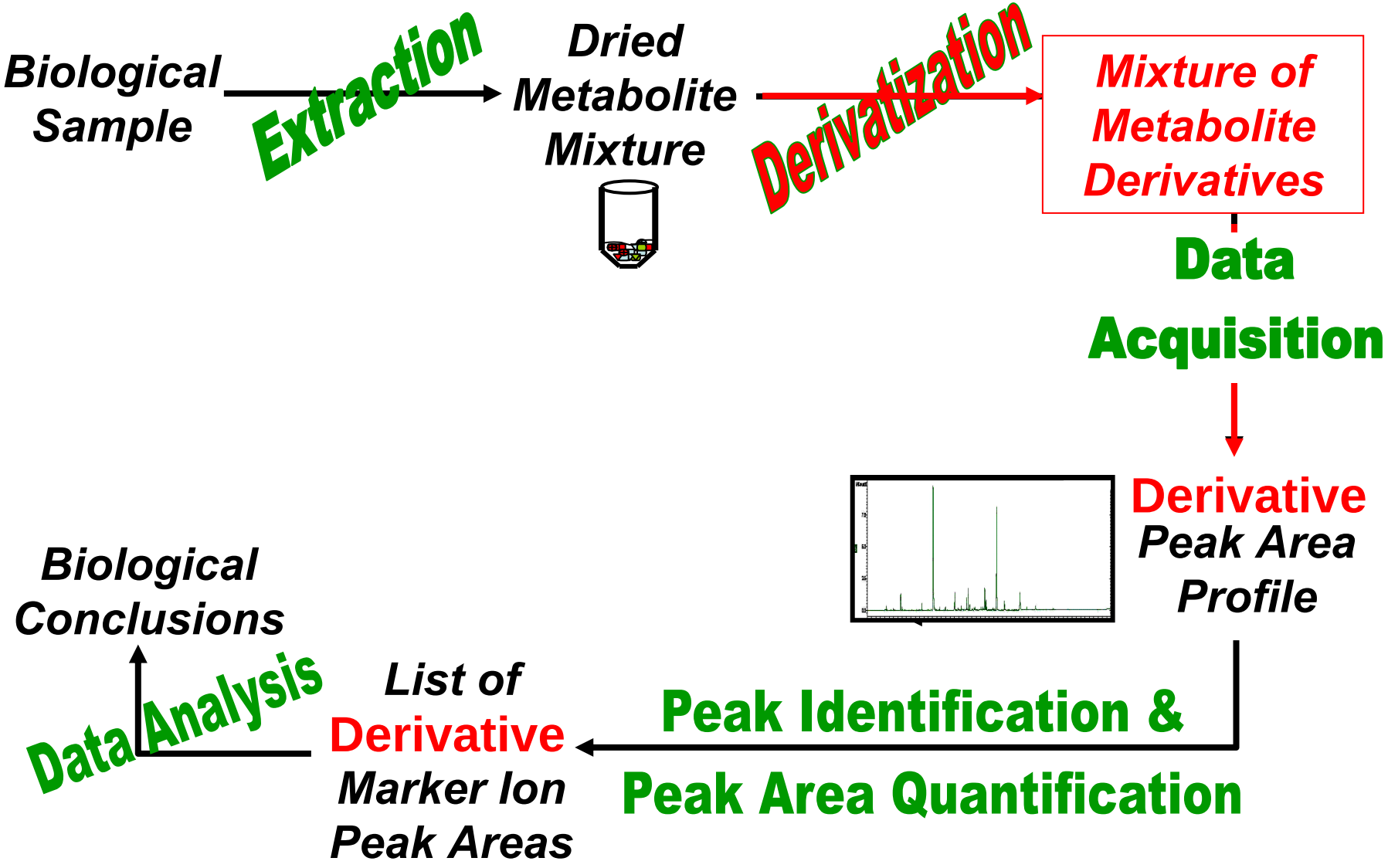
$$\textit{Original Metabolite j Concentration} = \frac{1}{RF_{j(MI_i^j)}} \times \textit{Measured Marker Ion i Peak Area}$$

$$\textit{Internal Standard (IS) Original Concentration} = \frac{1}{RF_{IS(MI_k^{IS})}} \times \textit{Measured IS Marker Ion k Peak Area}$$

$\Big\rangle$ **RPA $_j$**

ratio $_{\text{between 2 states}}$ (Metabolite's j Concentration) = ratio (RPA$_j$)

Schematic Diagram of GC-MS Metabolomic Analysis



From Original Metabolite to Derivative Peak Area

- concentration of the original metabolite
- concentration of a derivative of the original metabolite
- Measured peak area of the derivative's marker ion(s)

$$\textit{Derivative's I concentration} = \frac{1}{RF_{I(MI_h^I)}} \times \textit{Measured Marker Ion h Peak Area}$$

$$\textit{Internal Standard (IS) Original Concentration} = \frac{1}{RF_{IS(MI_k^{IS})}} \times \textit{Measured IS Marker Ion k Peak Area}$$

$$\Big\rangle \; RPA_{\text{deriv. I of Mj}}$$

ratio $_{\text{between 2 states}}$ **(Mj Concentration) ? ratio(**$RPA_{\text{deriv. I of Mj}}$**)**

Type B Biases

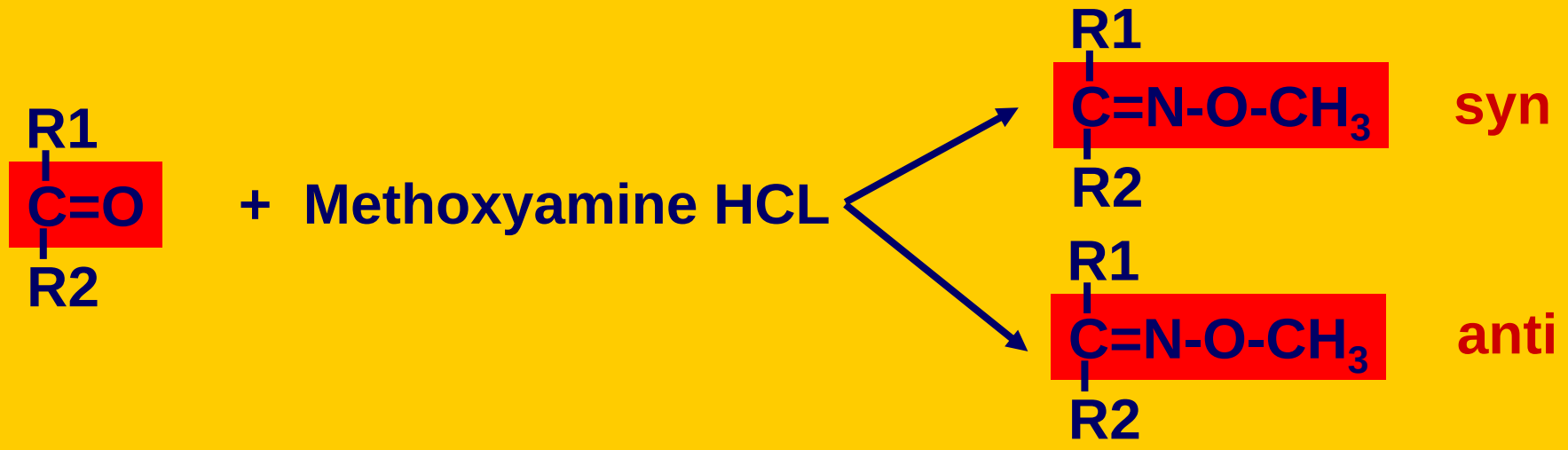
- Incomplete derivatization
- Multiple Derivatives for some Metabolites
- Potential Change in Equipment's Conditions between Runs

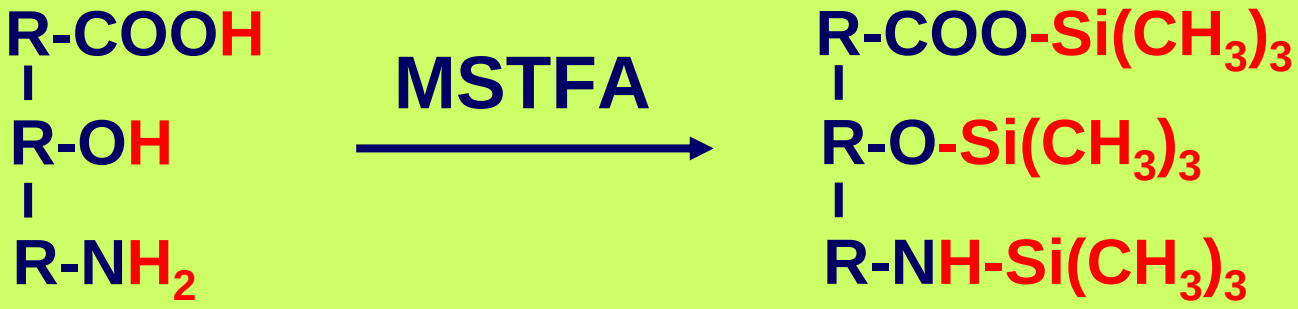
Need for a NEW Data Normalization, Correction and Validation Strategy

not jeopardizing the high-throughput nature of metabolomic analysis

H. Kanani and M.I. Klapa #. 2007. Data Correction Strategy for Metabolomics Analysis using Gas Chromatography-Mass Spectrometry, Metabolic Engineering Vol.9:39-51

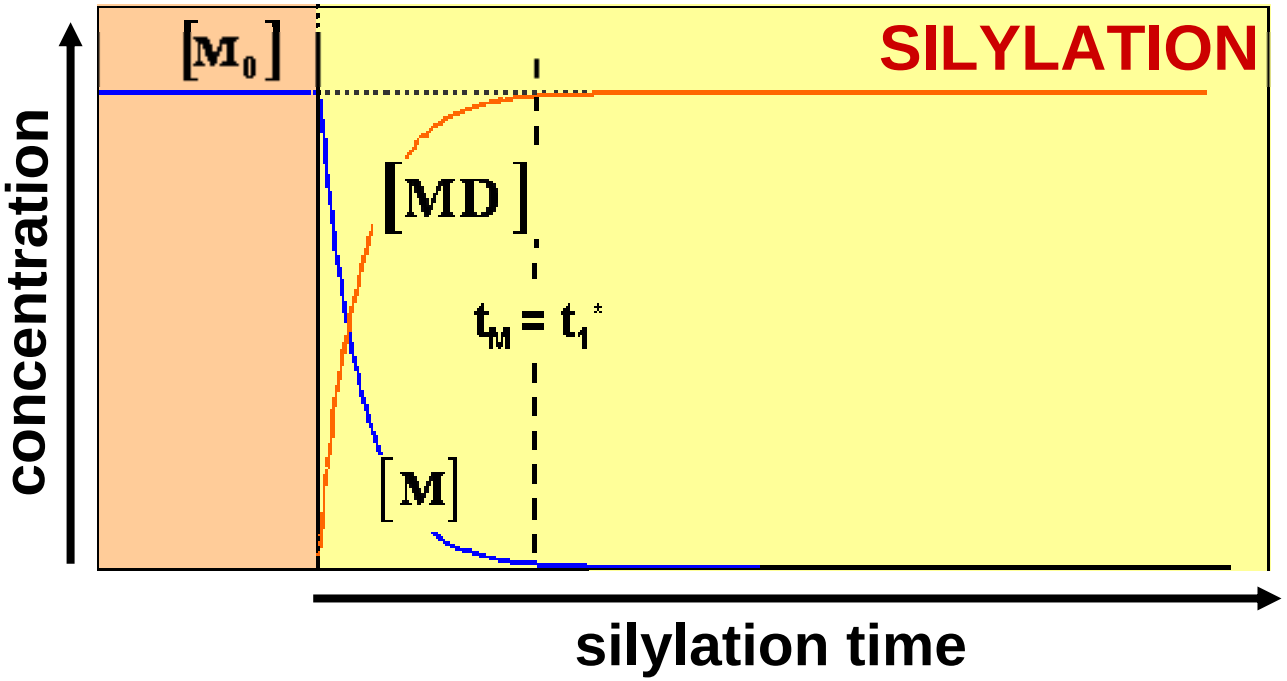
TMS and MeOX Derivatization





Metabolite Category 1

$$M + MSTFA \xrightarrow{k} MD$$

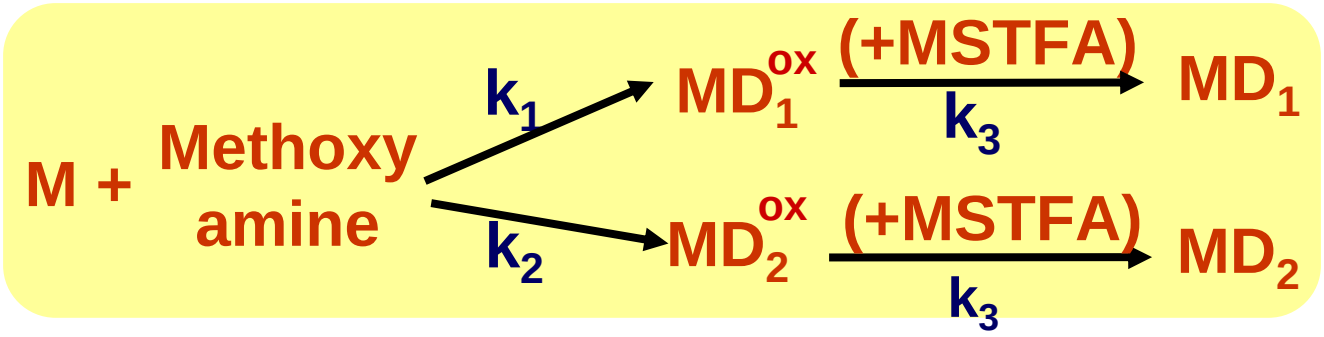


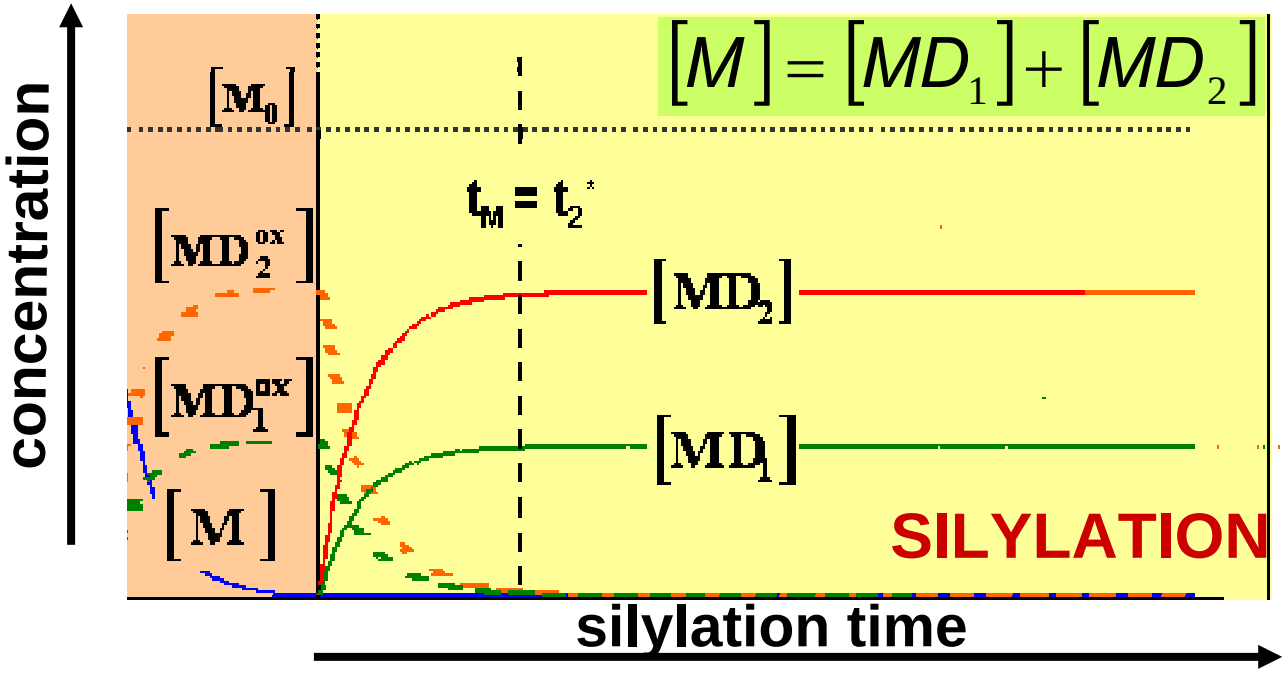
for $t > t_M$

$$[M] = [MD] = w_{MD} * RPA_{MD}$$

$$w_{MD} = \frac{RF_I^{IS}}{RF_j^{MD}}$$

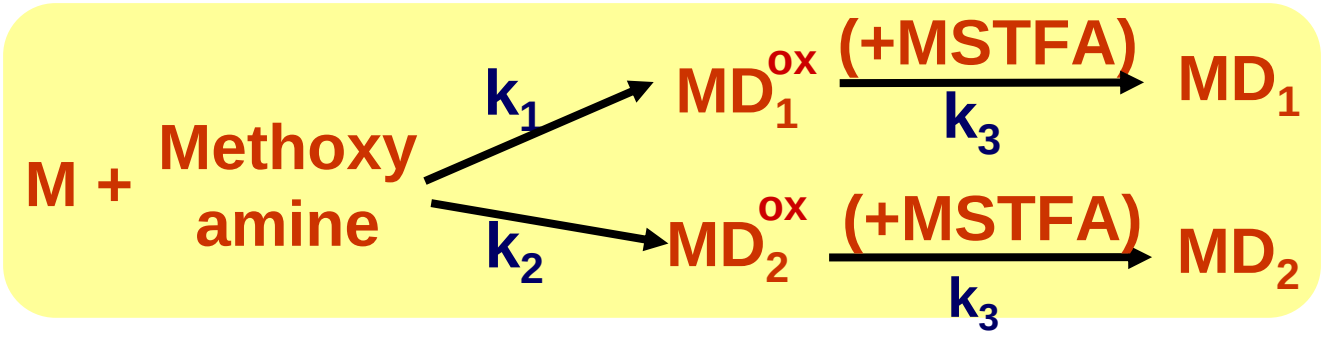
Metabolite Category 2

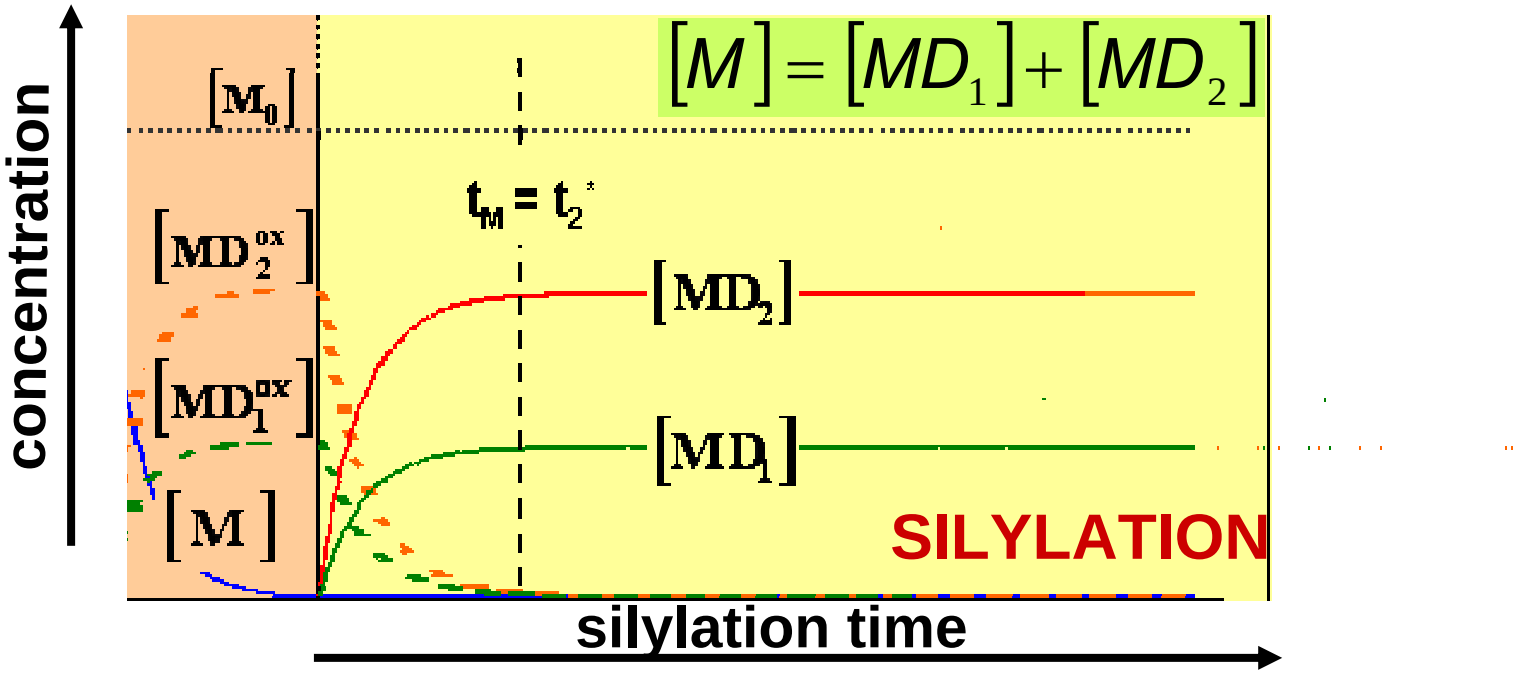




$$\frac{[MD_1]}{[MD_2]} = \frac{k_1}{k_2} = k_0 = \frac{w_{MD_1} * RPA_{MD_1}}{w_{MD_2} * RPA_{MD_2}}$$

Metabolite Category 2

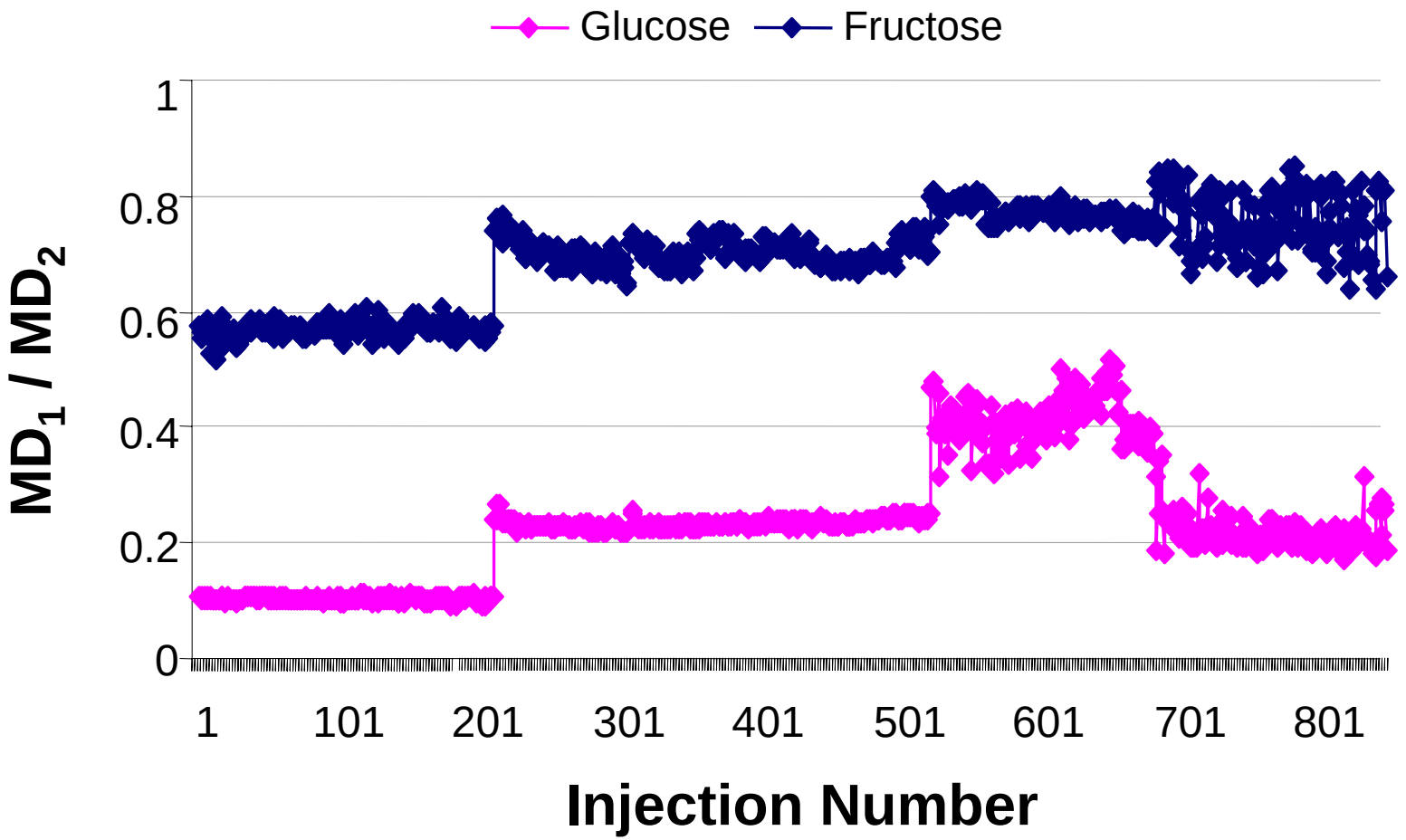




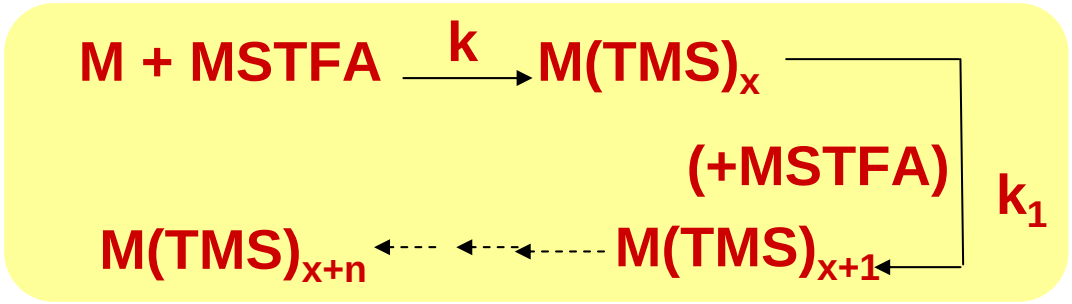
$$[MD_2] \quad k_2 \quad w_{MD_2} \quad RPA_{MD_2}$$

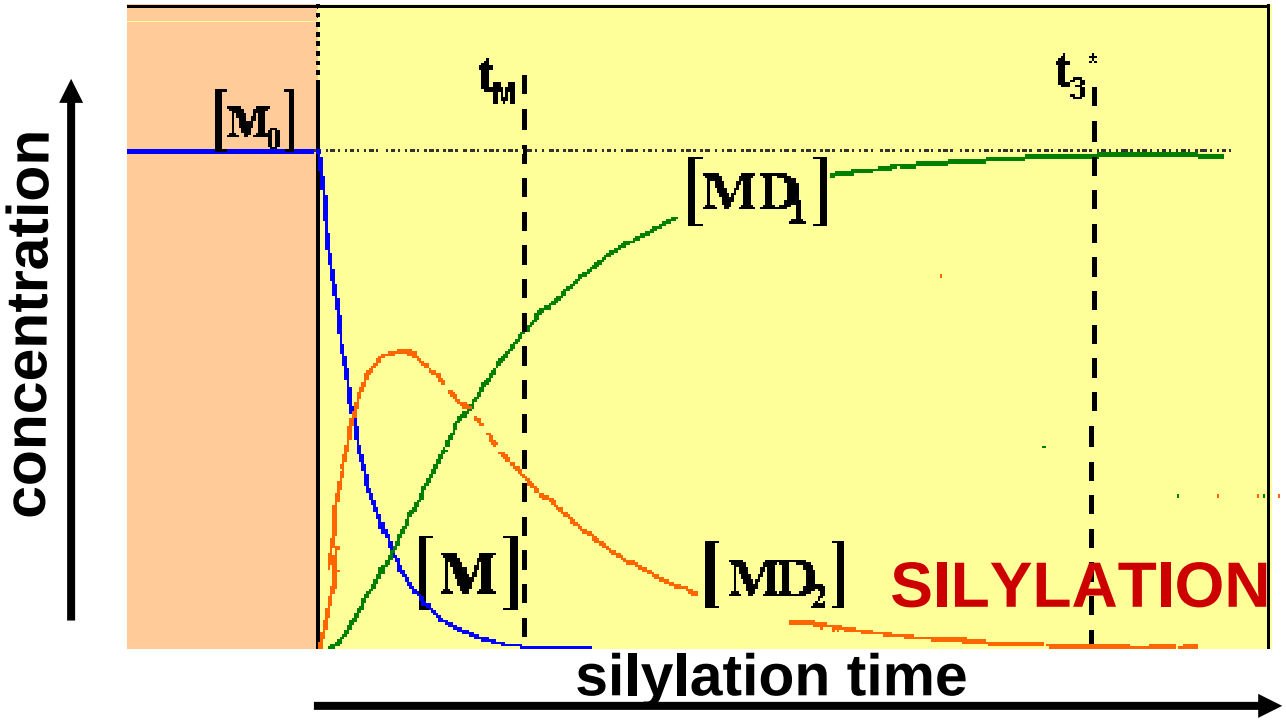
Data Validation Criterion!

Published Metabolomic Analysis based on Metabolomic Data Acquired at Different Equipment conditions



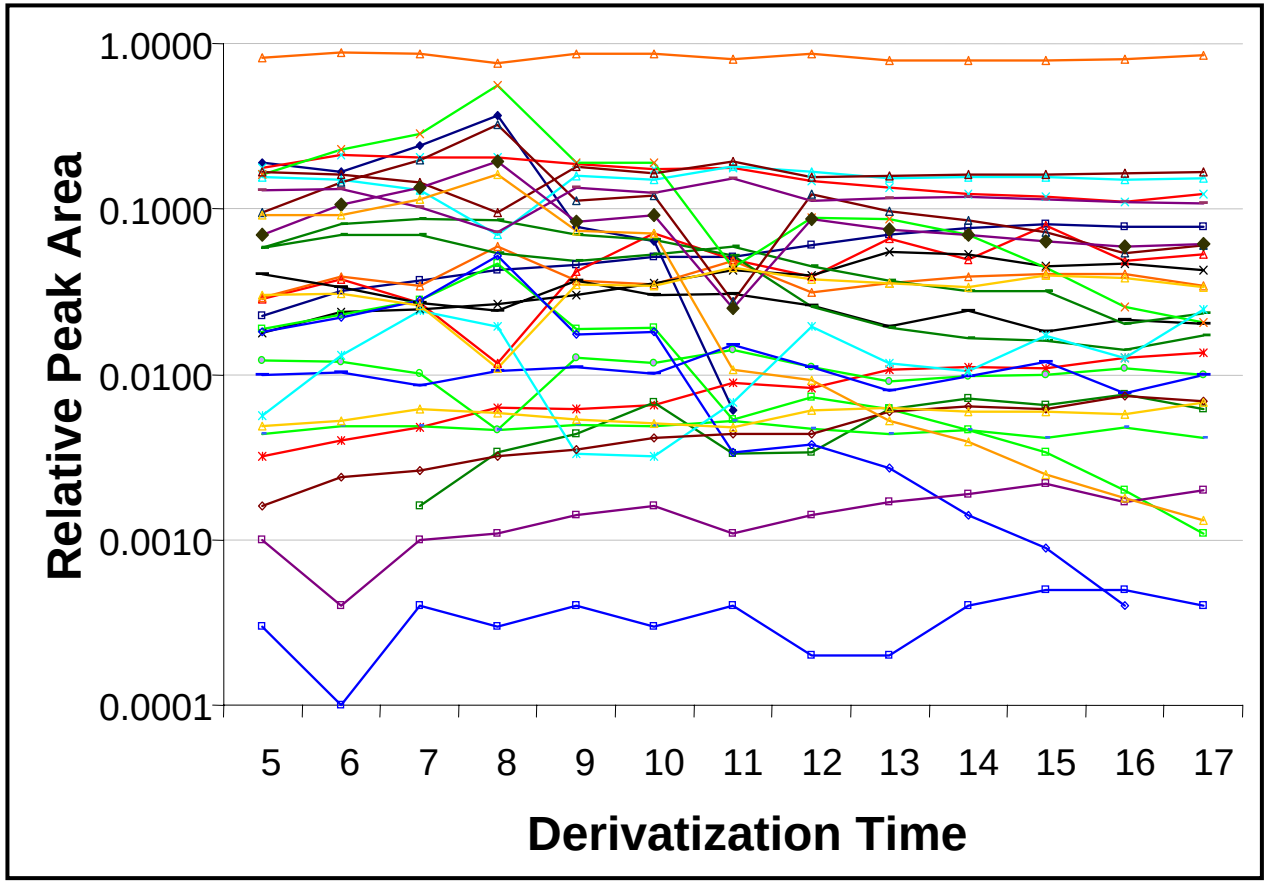
Metabolite Category 3





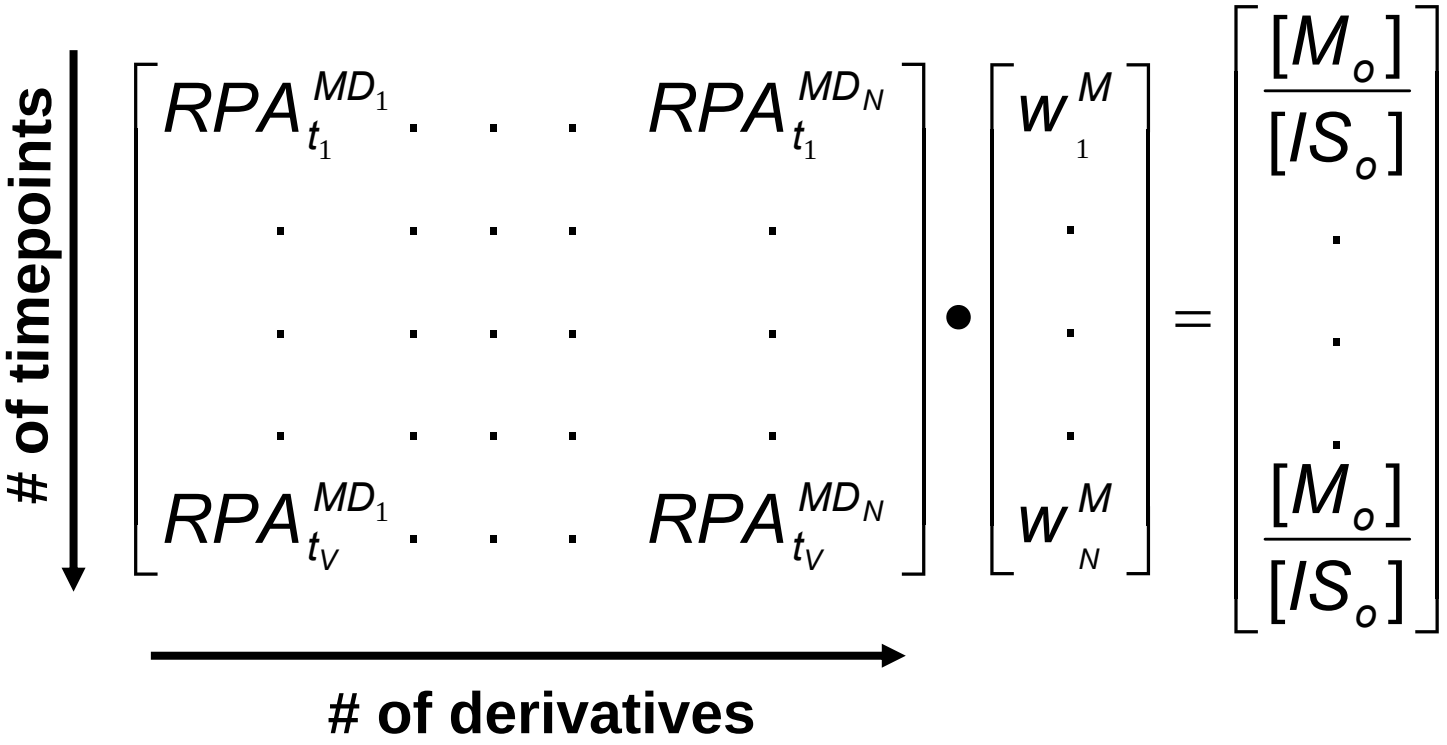
$$[M] = \sum_{i=1}^{n} [MD_i] = \sum_{i=1}^{n} w_{MD_i} * RPA_{MD_i}$$

Raw Data from Standard Amino Acid Mixture



Peak Area Variation with derivatization time among replicates of the same sample 15-100%

New Normalization Algorithm



$$\begin{bmatrix} RPA_{t_1}^{MD_1} & \cdots & RPA_{t_1}^{MD_N} \\ \vdots & \cdots & \vdots \\ RPA_{t_V}^{MD_1} & \cdots & RPA_{t_V}^{MD_N} \end{bmatrix} \bullet \begin{bmatrix} w_1^M \\ \vdots \\ w_N^M \end{bmatrix} = \begin{bmatrix} \frac{[M_o]}{[IS_o]} \\ \vdots \\ \frac{[M_o]}{[IS_o]} \end{bmatrix}$$

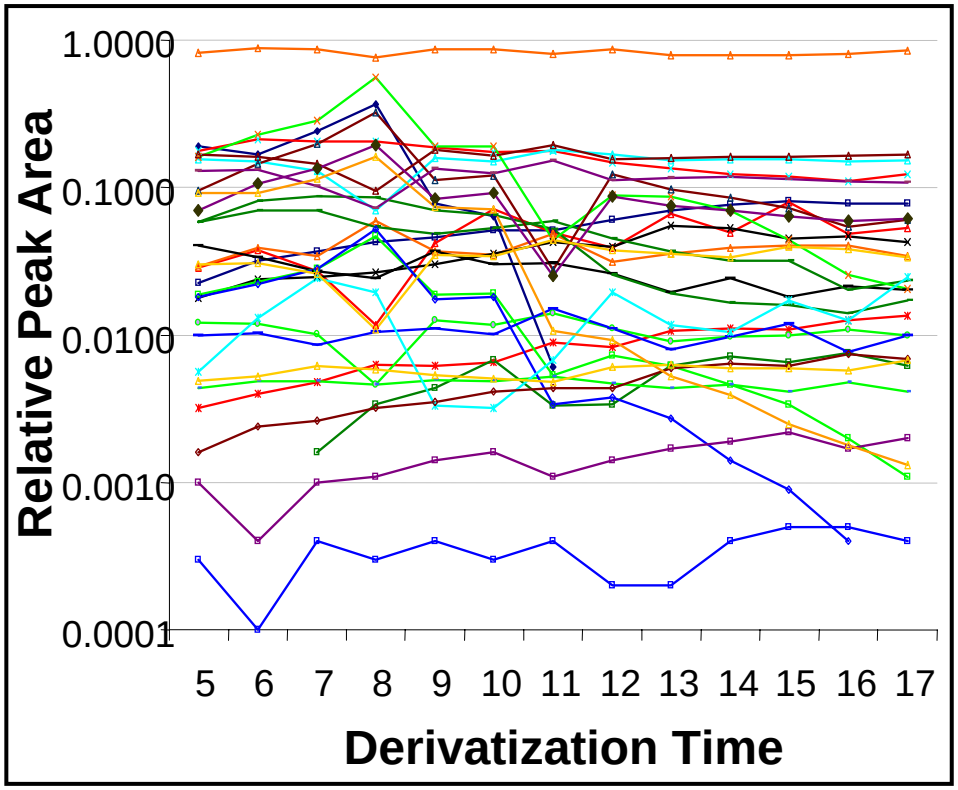
$RPA_{t_1}^{MD_1}$: relative (with respect to the peak area of the internal standard) peak area corresponding to the i-th derivative of M metabolite at derivatization time t_j

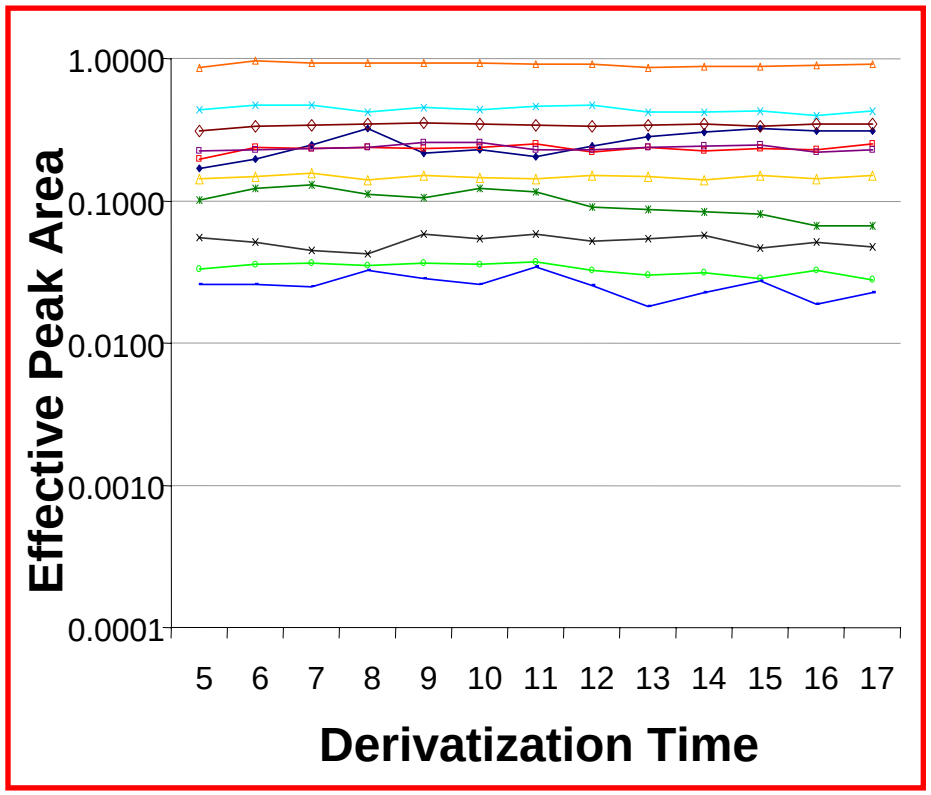
• U.S. Patent Application No. 11/362,717

• Best University of Maryland Invention of the Year 2005 in Information Sciences

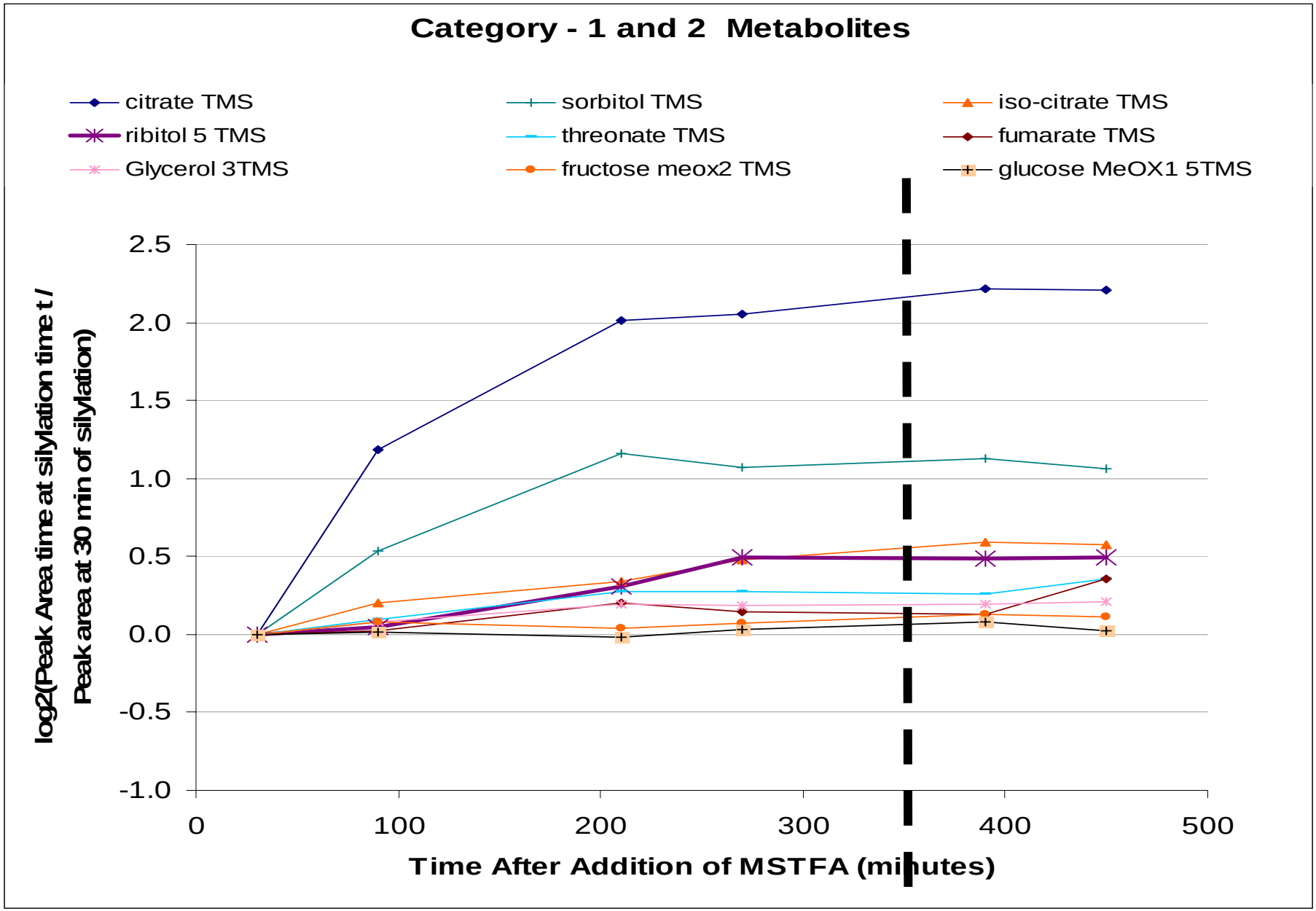
• H. Kanani and M.I. Klapa #. 2007. Data Correction Strategy for Metabolomics Analysis using Gas Chromatography-Mass Spectrometry, Metabolic Engineering Vol.9:39-51

Normalized Data from Standard Amino Acid Mixture



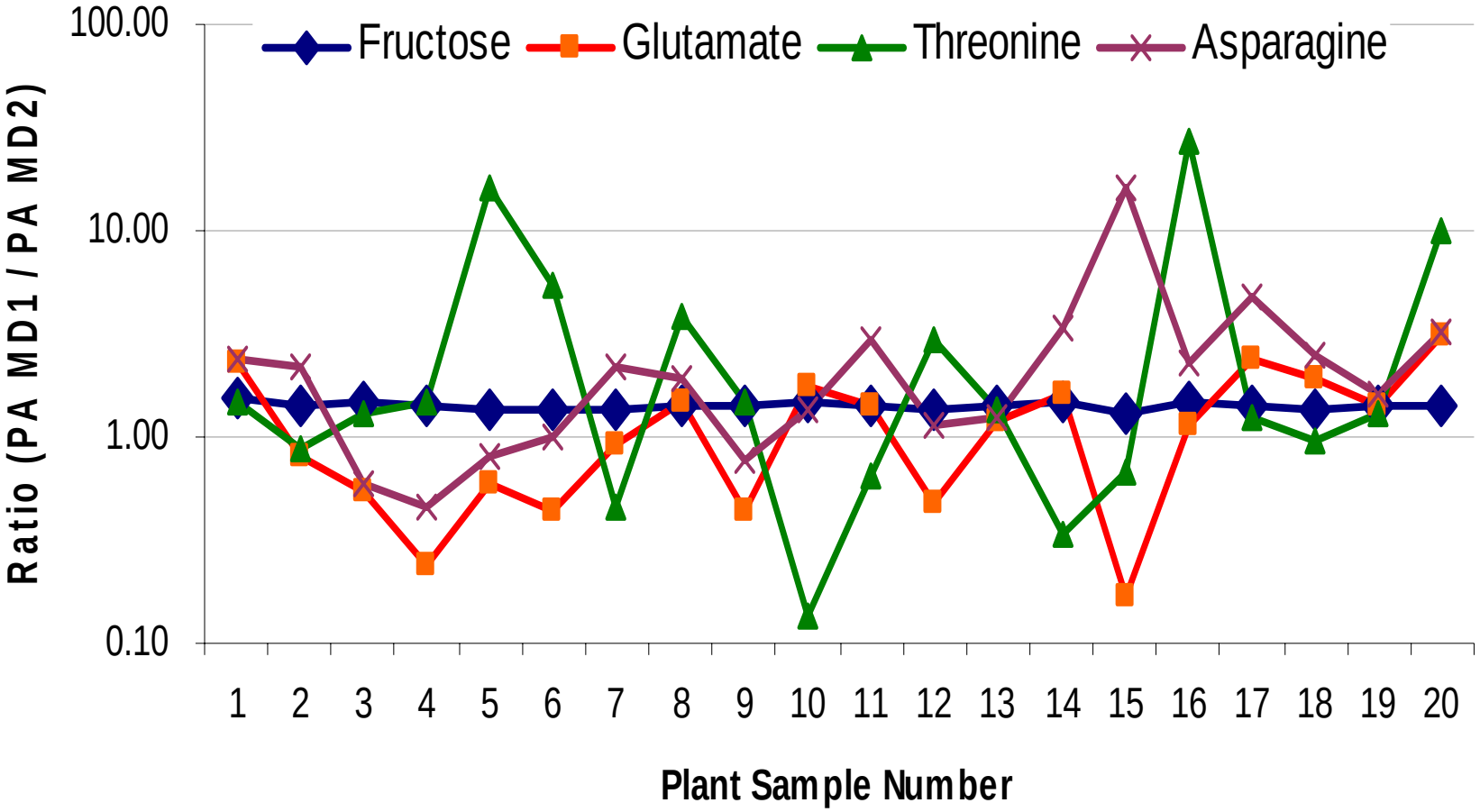


Peak Area Variation with derivatization time among replicates of the same sample dropped from 15-100% to 2-8%



Kanani HH, Chrysanthopoulos P, Klapa MI. 2008. Standardizing GC-MS Metabolomics. J. Chromatogr. B Analyt Technol Biomed Life Sci. 871: 191-201

Matrix Effects Limit the Accuracy of the Measurements even in the presence of an automated derivatization scheme



Kanani HH, Chrysanthopoulos P, Klapa MI. 2008. Standardizing GC-MS Metabolomics. J. Chromatogr. B Analyt Technol Biomed Life Sci. 871: 191-201

Identification of Unknown Peaks

	Amino acid	Derivative 1		Derivative 2		Derivative 3	
	(M)	MD_1	w_1	MD_2	w_2	MD_3	w_3
1	Alanine*	Alanine N O	1.025	Alanine N N O	0.774		
2	Arginine	Ornithine N N N O	1.10	Ornithine N N N O^2	0.48	Ornithine N N N N O	n/d
3	Asparagine	Asparagine N N O	0.726	Asparagine N N N O	1.904	**Asparagine N N N N $O^{2,3}$ (putative)**	1.595
4	Aspartate*	**Aspartate O $O^{2,3}$**	3.824	Aspartate N O O	0.224		
5	Cysteine	Cysteine N O^2	n/d	Cysteine N S O	12.67	Cysteine N N O	0.37
6	Glutamate	Glutamate N O O	1.014	Pyroglutamate N O^1	0.988		
7	Glutamine	Glutamine N N O	0.667	Glutamine N N N O	10.3	**Pyroglutamine NNO ^{1,2,3} (putative)**	9.000
8	Glycine*	Glycine N O	9.397	Glycine N N O	0.773		
9	Histidine	Histidine O^2 (putative)	n/d	Histidine N O	n/d	Histidine N N O	1.00
10	iso-Leucine*	iso-Leucine O	2.55	iso-Leucine N O	0.92	iso-Leucine N N O^2	n/d
11	Leucine	Leucine O	n/d	Lecine N O	1.000	Leucine N N O^2	n/d
12	Lysine*	Lysine N N O	n/d	Lysine N N N O	1.005	Lysine NNNNO^2	2.124
13	Methionine*	Methionine N O	1.42	Methinonine N N O^2	0.369		
14	Phenylalanine	Phenylalanine O	1.30	Phenylalanine N O	0.48		
15	Serine	Serine O O	2.97	Serine N O O	0.299	Serine NNOO^2	7.87
16	Threonine	Threonine O O	3.30	Threonine N O O	0.321	Threonine NNOO^2	33.5
17	Tryptophan	Tryptophan O^2 (putative)	n/d	Tryptophan N O	1.0	Tryptophan N N O	n/d
18	Tyrosine	Tyrosine O^2 (putative)	1.18	Tyrosine O O	0.94	Tyrosine N O O	0.26
19	Valine*	Valine O	1.638	Valine N O	0.842	**Valine N N $O^{2,3}$**	n/d
20	Allantoin	Allantoin N N N	25.3	Allantoin N N N N	0.530	Allantoin N N N N N	2.12
21	Beta-Alanine	B-Alanine O	8.88	B- Alanine N O	n/d	b–Alanine N N O	0.80
22	Gaba	Gaba N O	n/d	Gaba N N O	1.0		
23	Dopamine	Dopamine N O O	4.16	Dopamine N N O O	0.73		
24	Homoserine	Homoserine O O	6.51	Homoserine N O O	0.231	Homoserine N N O	
25	Ornithine	Ornithine N N N O	1.10	Ornithine N N N O^2	0.48	Ornithine N N N N	

n/d: Not detected consistently in all the samples

Derivatives not present in major public databases

Derivatives formed from chemical transformations

Derivatives treated as unknowns in public databases

Conclusions

We developed a GC-MS metabolomic data validation, normalization and correction strategy that does NOT jeopardize the high-throughput nature of the analysis

The method is easy to implement and increases the accuracy of measurements by an order of magnitude for some metabolites (NH_2 containing compounds)

In light of the importance of metabolomics research, this method is expected to provide a valuable tool for the acquisition of accurate metabolomic data

Objective

To analyze stress-induced molecular interaction networks

in the context of plant primary metabolism

during the first (30) hours of the stress treatment

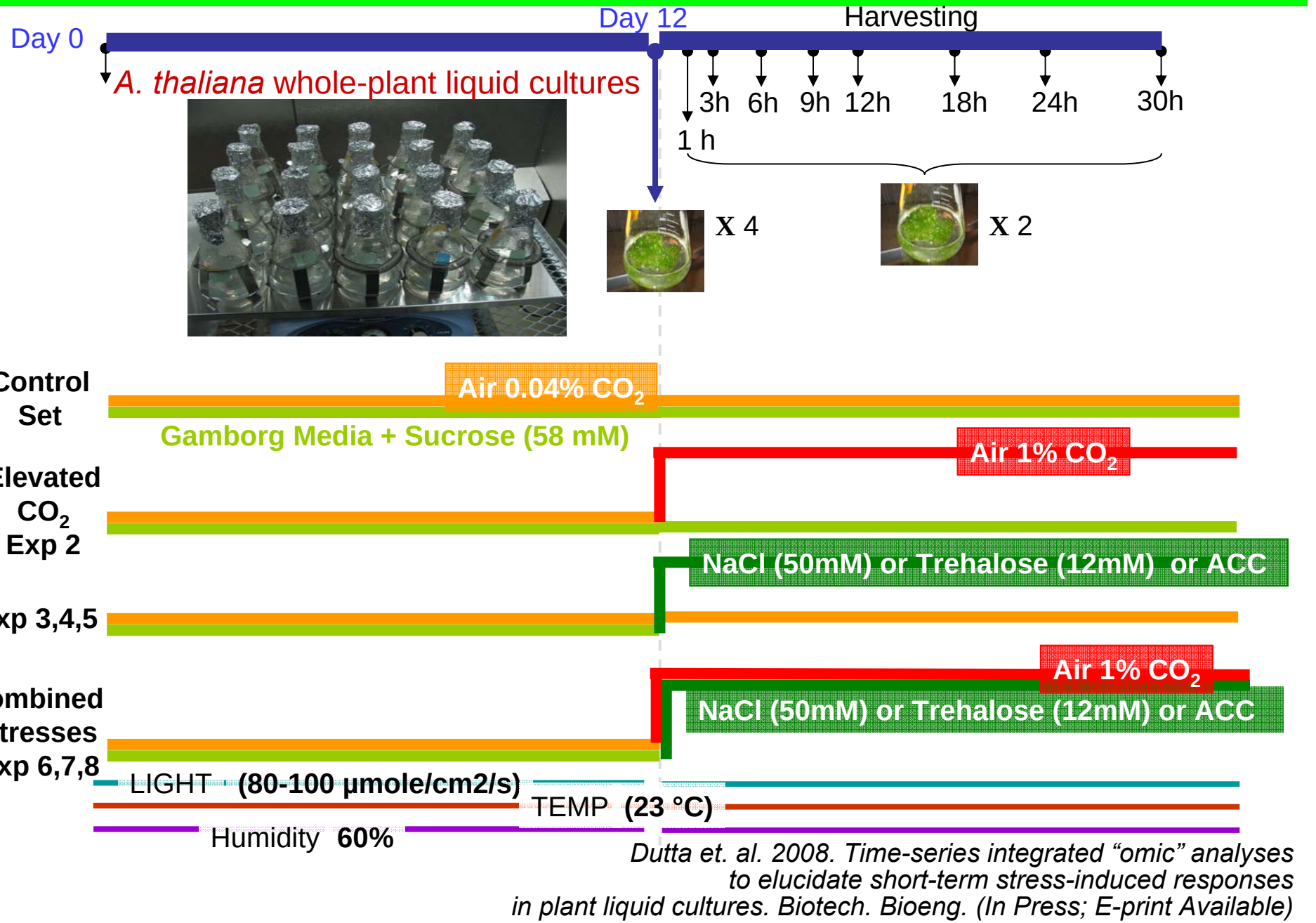
under a variety of individual or combined perturbations

using integrated time-series transcriptomic & metabolomic analyses

Model System: *Arabidopsis thaliana* Whole Plant Liquid Cultures

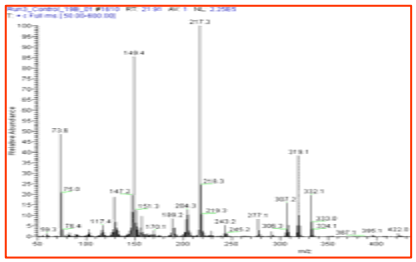
Well-controlled growth environment

Experimental Design & Setup

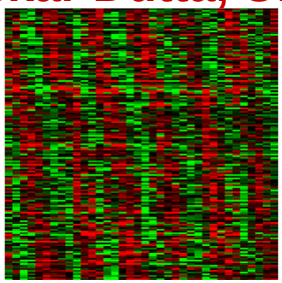


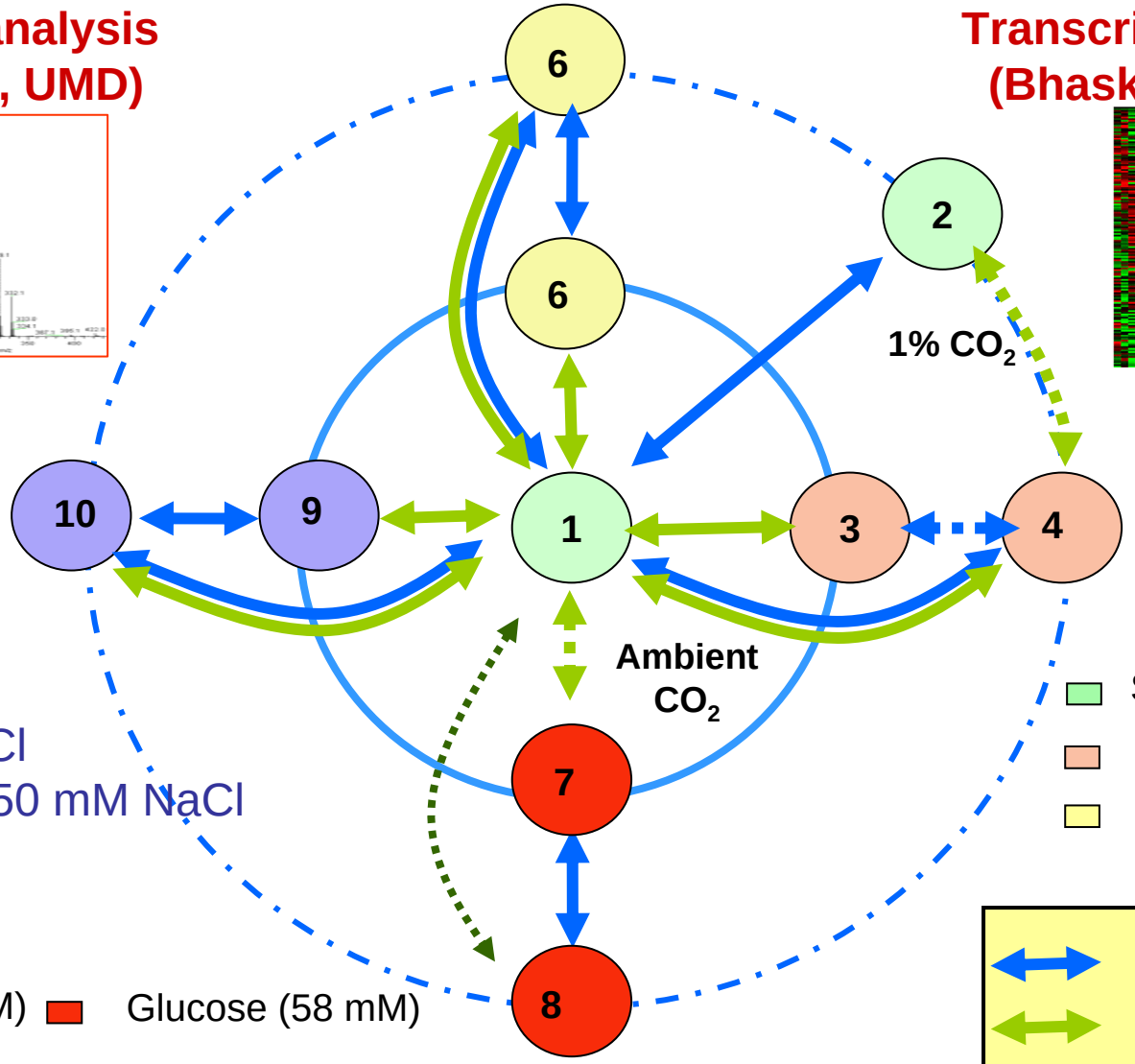
Dutta et. al. 2008. Time-series integrated "omic" analyses to elucidate short-term stress-induced responses in plant liquid cultures. Biotech. Bioeng. (In Press; E-print Available)

GC-MS
Metabolomic analysis
(Harin Kanani, UMD)



cDNA Microarray
Transcriptomic analysis
(Bhaskar Dutta, UMD)





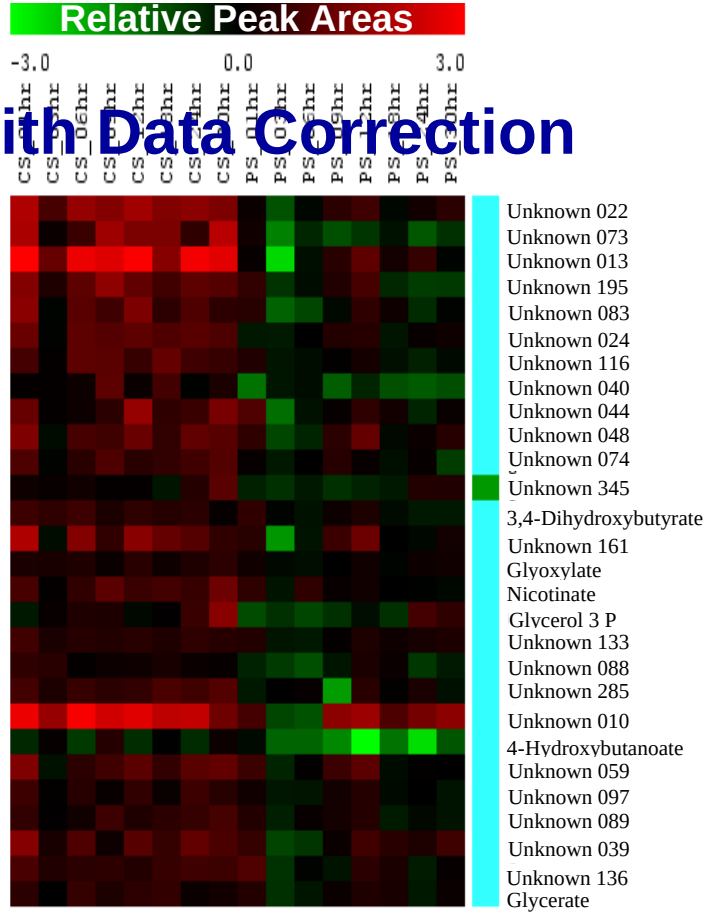
1. Control
2. 1% CO_2
3. 50 mM NaCl
4. 1% CO_2 + 50 mM NaCl

10 Exp * 20 samples * 2 Injections * 550 Peaks = 220,000 Total Measurements

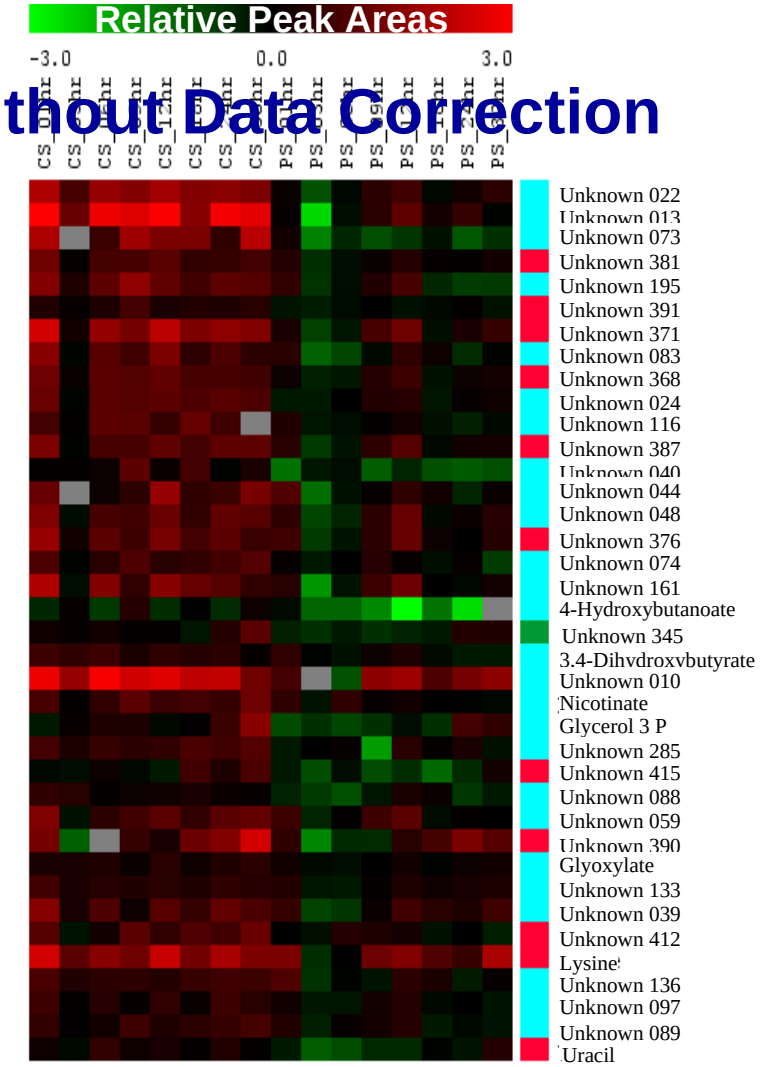
(8 Exp * 20 samples) Trizol extractions → 160 mRNA amplifications → 640 cDNA syntheses → 320 Dye Injections → 320 Micro-array hybridizations (flip-dye)

GC-MS Metabolomic Data Correction Methodology

Paired-SAM analysis (TIGR MeV v.3.1)
delta = 1.2, FDR (median)= 0%



27 + 1 significant



26+1+11 significant

Cat-1 **Cat-2**

Cat-3

*H. Kanani and M.I. Klapa #. **2007**. Data Correction Strategy for Metabolomics Analysis using Gas Chromatography-Mass Spectrometry, Metabolic Engineering 9:39-51*

*Kanani HH, Chrysanthopoulos P, Klapa MI. **2008**. Standardizing GC-MS Metabolomics. J. Chromatogr. B Analyt Technol Biomed Life Sci. 871: 191-201*

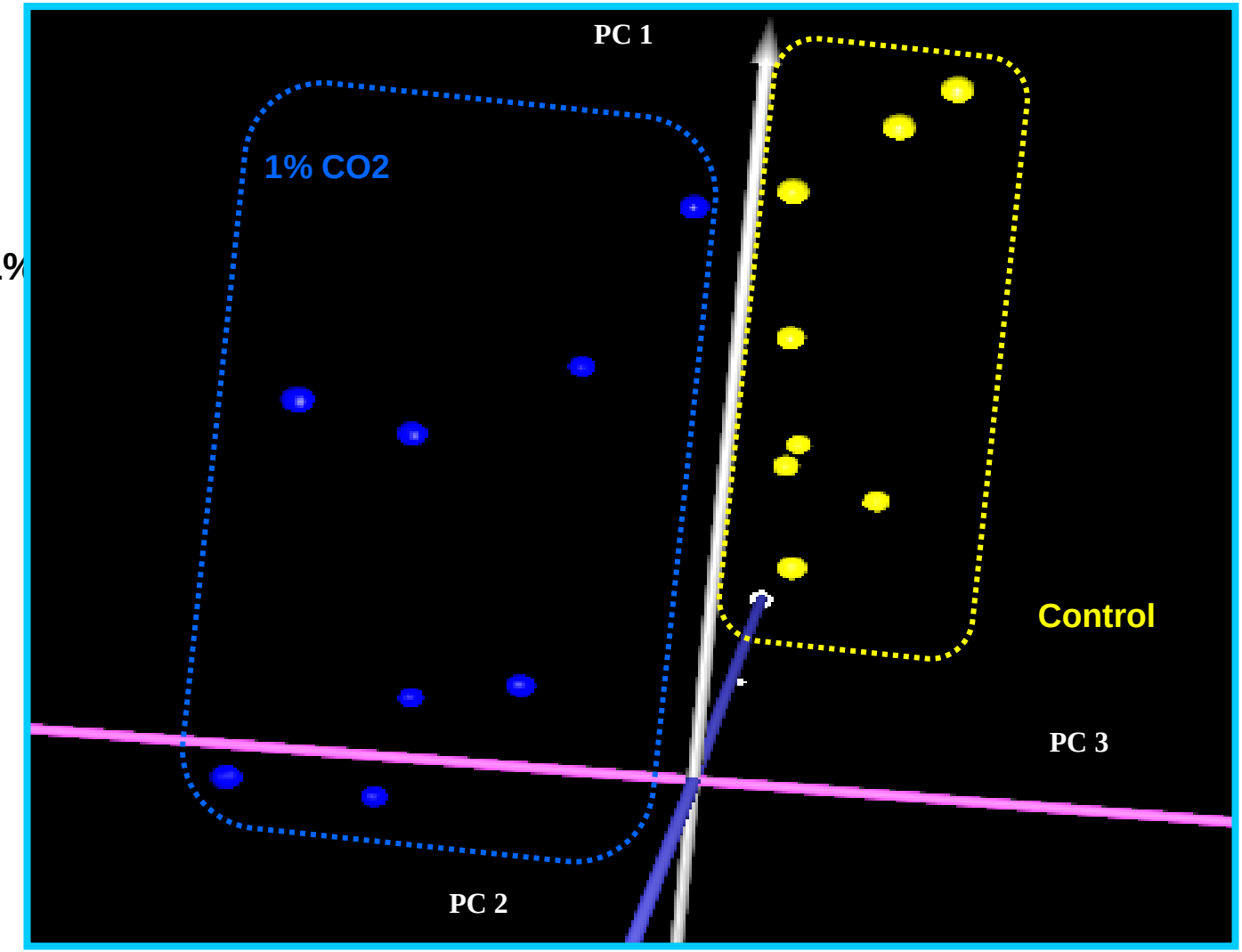
PCA- Metabolomic Data: Individual Stress Response

PC1: 48%

PC2: 14%

PC3: 9%

Total: 71%



Microarray Time-series Significance Analysis

Identification of Significant Genes at each time-point

Consists of 4 modules

Analysis of Gene Variability in Significance Level Among Time Points

MiTimeS

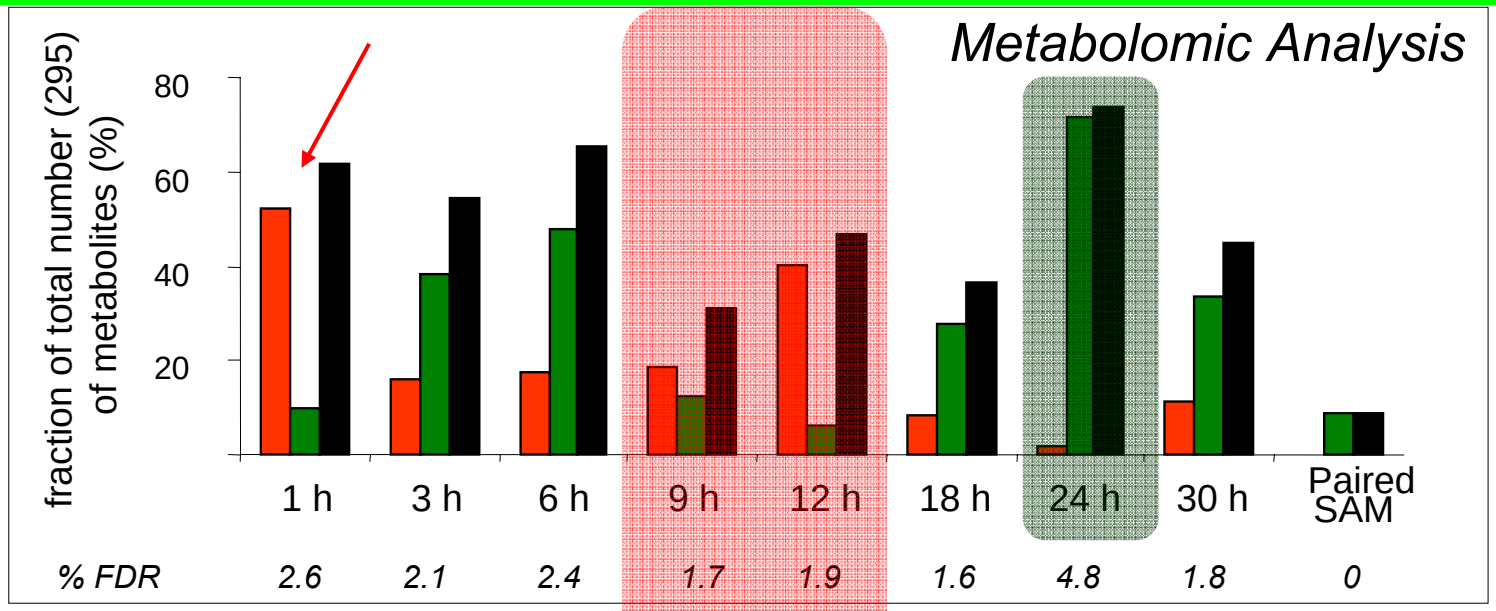
Correlation Analysis between timepoints with respect to their common significant genes

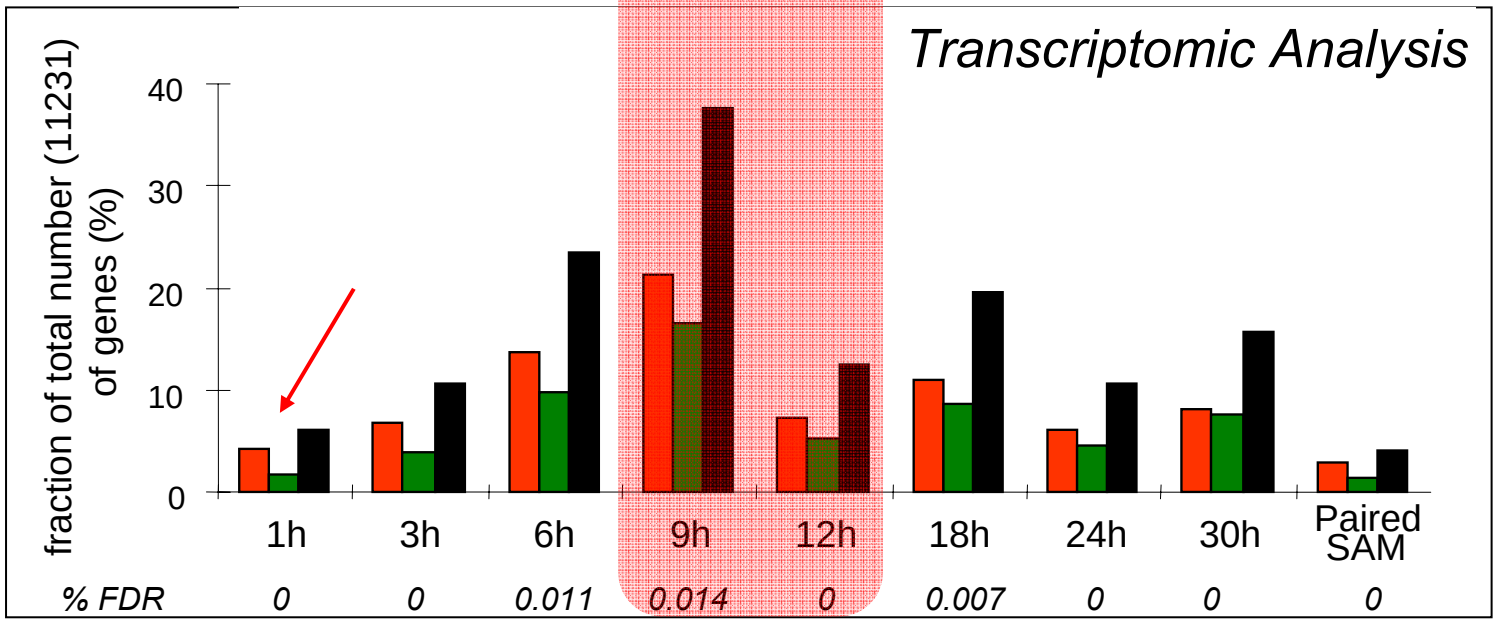
Comparison of significant genes' GO analysis results between time-points

US Patent Application, 2006

*Dutta B, Snyder R, Klapa MI. 2007. Significance Analysis of Time-Series Transcriptomic Data: A methodology that enables the identification and further exploration of the differentially expressed genes at each time-point. Biotech. Bioeng. **98**: 668-678.*

Elevated CO_2 stress : Time-profile of No of Significant Genes





Positively Significant

Negatively Significant

Total Significant

Dutta et. al. 2008. Time-series integrated "omic" analyses to elucidate short-term stress-induced responses in plant liquid cultures. Biotech. Bioeng. (In Press; E-print Available)

Acknowledgements

Funding

US NSF Grant: QSB-0331312,
UMD Minta Martin Foundation,
UMD Department of Chemical & Biomolecular Engineering,
FORTH/ICE-HT
Bayer HealthCare LLC

Thank you!