

METABOLITE FOOTPRINTING AND FINGERPRINTING IN
SACCHAROMYCES CEREVISIAE

by

Ali Erdem Aygök

B. S. in Ch.E., Hacettepe University, 2002

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering

Boğaziçi University

2007

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my thesis supervisor, Prof. Kutlu Ülgen, who has given continuous encouragement, support and trust. I am grateful to her for her patience.

My sincere gratitude is due to my thesis committee members, Prof. Zeynep İlşen Önsan and Prof. Dilek Kazan who have read and commented on my thesis. I would like to thank to Prof. Alp Sayar, Dean of School of Engineering of Marmara University, for allowing me to use the mass spectrometer system in Marmara University.

Special thanks go to Pınar Pir, Selim Ceylan, Kazım Yalçın Arga, Tunahan Çakır, and Barış Akkaya. Thanks to Beste Kınıkoğlu, Betül Kavun, Duygu Dikicioğlu, Saliha Durmuş and all Genetik Geyikler. Thanks to EGE, Erdal, Özgür, and Salih for their friendship.

Finally, I wish to thank my family for their support throughout my entire education. This thesis would not have been possible without them.

ABSTRACT

The metabolic response of *S. cerevisiae* cells of galactose metabolic pathway genes, hexose transporter genes, transcription regulation and respiration related genes were investigated. Homozygous deletion mutants *rgt1Δ/ rgt1Δ*, *adh6Δ/ adh6Δ*, *gal4Δ/ gal4Δ*, *gal6Δ/ gal6Δ*, *gal7Δ/ gal7*, *gal80Δ/ gal80Δ*, *hap4Δ/ hap4Δ*, *hxt10Δ/hxt10Δ*, *hxt12Δ/ hxt12Δ*, *mba1Δ/ mba1Δ*, *mig1Δ/ mig1Δ*, *snf3Δ/ snf3Δ*, *yel057cΔ/yel057cΔ*, and *hoΔ/hoΔ* (reference strain) were grown in batch reactors which were containing 20 gL⁻¹ initial glucose in minimal medium. The samples were collected at stationary phase of fermentation analysed for extracellular (metabolic footprinting) and intracellular (metabolic fingerprinting) metabolome profiles by Q-TOF mass spectrometer. Then, metabolic data were processed by PCA and ICA methods to identify the clusters within the biologically or functionally related deletion mutants.

Pearson correlation analysis was applied to metabolic data, which shows the interaction between the deletion mutants. The results of correlation analysis confirmed the accuracy of PCA and ICA. Signatures of deletion mutants at certain m/z values were obtained for extracellular and intracellular metabolome data. Then, these m/z values were used for pair-wise mutant correlations.

In conclusion, extracellular metabolome data of *S. cerevisiae* at 20th hour of fermentation, and intracellular metabolome data at 25th hour of fermentation provide better discrimination of known genes. It was also observed that *yel057cΔ* deletion mutant (unknown function or relation) appeared in the same region with *snf3Δ*, *adh6Δ*, and *hxt12Δ*. It could be stated that *yel057cΔ* would play a role in glucose or hexose transport like *snf3Δ* and *hxt12Δ*.

ÖZET

Bu çalışmada *S. cerevisiae* hücrelerinin galaktoz metabolik yolundan sorumlu genlerinin, hexose taşıyıcı, transkripsiyon düzenleyici ve solunumdan sorumlu genlerinin metabolik tepkileri araştırılmıştır. *Rgt1Δ/ rgt1Δ*, *adh6Δ/ adh6Δ*, *gal4Δ/ gal4Δ*, *gal6Δ/ gal6Δ*, *gal7Δ/ gal7*, *gal80Δ/ gal80Δ*, *hap4Δ/ hap4Δ*, *hxt10Δ/hxt10Δ*, *hxt12Δ/ hxt12Δ*, *mba1Δ/ mba1Δ*, *mig1Δ/ mig1Δ*, *snf3Δ/ snf3Δ*, *yel057cΔ/yel057cΔ* homozigot delesyon mutantları *hoΔ/hoΔ* (referans şuş) 20 gL⁻¹ glukoz çözeltisi içeren batch reaktörlerde büyütülmüştür. Fermantasyonun durağan fazında toplanan örnekler, QTOF kütle spektrometresi yardımıyla hücre dışı ve hücre içi metabolom profillerinin analizi için kullanılmıştır. Daha sonra, PCA ve ICA istatistiksel metodlarının metabolik verilere uygulanması sonucunda grafiklerde aynı bölgede bulunan delesyon mutantlarının biyolojik veya fonksiyonel açıdan birbirleriyle yakın görevlere sahip genler olup olmadığı incelenmiştir.

Metabolik verilere delesyon mutantlarının birbirleriyle olan ilişkisini göstermek amacıyla Pearson korelasyon analizi uygulanmıştır. Korelasyon analizinin sonuçları PCA ve ICA yöntemlerinin geçerliliğini doğrulamaktadır. Delesyon mutantlarının hücre dışı ve hücre içi metabolom verileri için belirli m/z değerlerindeki imzaları elde edilmiştir. Elde edilen m/z değerleri çiftli mutant korelasyon analizi için kullanılmıştır.

Sonuç olarak, *S. cerevisiae* hücrelerinin 20. saatteki hücre dışı metabolom verileri ve 25. saatteki hücre içi metabolom verileri görevi bilinen gen gruplarının birbirlerinden daha iyi ayırt edilebilmesine imkan sağlamıştır. Ayrıca fonksiyonu ya da diğer genlerle ilişkisi bilinmeyen *yel057cΔ* delesyon mutantının *snf3Δ*, *adh6Δ* ve *hxt12Δ* ile aynı bölgede yer aldığı gözlemlenmiştir. Bu bulgudan yararlanarak *yel057cΔ* delesyon mutantının glukoz veya hexose taşınmasında *snf3Δ* ve *hxt12Δ* gibi bir rolü olduğu belirtilebilir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
ABSTRACT.....	iv
ÖZET.....	v
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xv
LIST OF SYMBOLS/ABBREVIATIONS.....	xvi
1. INTRODUCTION.....	1
2. LITERATURE SURVEY.....	3
2.1. Scope of Metabolomics.....	3
2.2. Yeast.....	4
2.3. Yeast Metabolism.....	5
2.4. Literature Search on the Genes Investigated in This Study.....	7
2.4.1. <i>HO</i>	7
2.4.2. <i>HXT Genes</i>	7
2.4.3. <i>YEL057c</i>	8
2.4.4. <i>SNF3</i>	8
2.4.5. <i>RGT1</i>	9
2.4.6. <i>MIG1</i>	9
2.4.7. <i>MBA1</i>	10
2.4.8. <i>HAP4</i>	10
2.4.9. <i>GAL4</i>	10
2.4.10. <i>GAL6</i>	11
2.4.11. <i>GAL7</i>	11
2.4.12. <i>GAL80</i>	11
2.4.13. <i>ADH6</i>	12
2.5. Metabolome Analysis.....	12
2.5.1. Metabolic Fingerprinting and Footprinting	14

3. MATERIALS AND METHODS.....	20
3.1. Materials.....	20
3.1.1. Microorganisms.....	20
3.1.2. Chemicals.....	20
3.1.3. Culture Media	20
3.1.4. Laboratory Equipment.....	22
3.2. Experimental Methods.....	23
3.2.1. Sterilization.....	23
3.2.2. Preparation of Agar Plates and Preculture.....	23
3.2.3. Batch Experiments	23
3.2.4. Sampling Procedure.....	24
3.2.4.1. Stability of Metabolites During the Extraction Procedure.....	24
3.2.5. Mass Spectrometry and Sample Preparation.....	25
3.3. Computational Work.....	25
3.3.1. Multivariate data analysis	25
3.3.2. Normalisation	27
3.3.3. PCA-Principal Component Analysis	29
3.3.4. ICA-Independent Component Analysis	30
3.3.4.1. Significant Components—Kurtosis	31
3.3.5. HCA-Hierarchical Cluster Analysis	32
3.3.6. GO Analysis	32
4. RESULTS.....	34
4.1. Metabolic Footprinting	34
4.1.1. PCA and ICA of binned extracellular Metabolome Data at 15 th hour	39
4.1.2. PCA and ICA of filtered extracellular Metabolome Data at 15 th hour	43
4.1.3. PCA and ICA of binned extracellular Metabolome Data at 20 th hour.....	46
4.1.4. PCA and ICA of filtered extracellular Metabolome Data at 20 th hour.....	51
4.1.5. PCA and ICA of binned extracellular Metabolome Data at 25 th hour	55
4.1.6. PCA and ICA of filtered extracellular Metabolome Data at 25 th hour	58
4.1.7. PCA and ICA of binned extracellular Metabolome Data at 31 st hour.....	61

4.1.8. PCA and ICA of filtered extracellular Metabolome Data at 31 st hour	67
4.1.9. PCA and ICA of all extracellular Metabolome Data at four time intervals (preprocessing by binning)	73
4.2. Metabolite Fingerprinting	76
4.2.1. PCA and ICA of binned intracellular Metabolome Data at 20 th hour	80
4.2.2. PCA and ICA of filtered intracellular Metabolome Data at 20 th hour	83
4.2.3. PCA and ICA of binned intracellular Metabolome Data at 25 th hour.....	87
4.2.4. PCA and ICA of filtered intracellular Metabolome Data at 25 th hour	90
4.2.5. PCA and ICA of binned intracellular Metabolome Data at 31 st hour	93
4.2.6. PCA and ICA of filtered intracellular Metabolome Data at 31 st hour	96
4.2.7. PCA and ICA of all intracellular Metabolome Data at three time intervals (preprocessing by binning).....	99
5. DISCUSSION	104
5.1. Comparison of the Two Preprocessing and Clustering Approaches	104
5.2. Significant m/z Values in PCA	107
5.3. m/z Signatures of Deletion Mutants	110
5.4. Correlation Analysis	114
5.4.1. Metabolic Footprinting Results	114
5.4.2. Metabolic Fingerprinting Results	118
5.5. GO Term Analysis	123
6. CONCLUSION AND RECOMMENDATIONS	125
6.1. Conclusions	125
6.2. Recommendations	127
APPENDIX: HCA RESULTS	128
A. HCA results of Preprocessing of Extracellular Metabolome Data by Binning	128
B. HCA results of Preprocessing of Intracellular Metabolome Data by Filtering	156
REFERENCES	184

LIST OF FIGURES

Figure 2.1.	Metabolism of yeast	6
Figure 2.2.	Metabolic footprinting of <i>S. cerevisiae</i>	19
Figure 3.1.	Procedure for multivariate data analysis	26
Figure 3.2.	Normalisation and analysis techniques	27
Figure 4.1.	MS spectral data of extracellular metabolites belonging to 17 mutants at 20th hour	36
Figure 4.2.	Kurtosis measure of ICA data	39
Figure 4.3.	Scores of binned metabolome data on first four Latent Variables at 15 th hour of extracellular samples	40
Figure 4.4	Scores of binned metabolome data on first three IC Values at 15 th hour of extracellular samples (PCA preprocessing number is 5)	41
Figure 4.5	Scores of binned metabolome data on first three IC Values at 15 th hour of extracellular samples (PCA preprocessing number is 6)	42
Figure 4.6.	Scores of filtered metabolome data on first four Latent Variables at 15 th hour of extracellular samples	43
Figure 4.7.	Scores of filtered metabolome data on first three IC Values at 15 th hour of extracellular samples (PCA preprocessing number is 5)	44
Figure 4.8.	Scores of filtered metabolome data on first three IC Values at 15 th hour of extracellular samples (PCA preprocessing number is 6)	45
Figure 4.9.	Galactose metabolic pathway	46

Figure 4.10. Scores of binned metabolome data on first four Latent Variables at 20 th hour of extracellular samples	47
Figure 4.11. Scores of binned metabolome data on first three IC Values at 20 th hour of extracellular samples (PCA preprocessing number is 5)	48
Figure 4.12. Scores of binned metabolome data on first three IC Values at 20 th hour of extracellular samples (PCA preprocessing number is 6)	49
Figure 4.13. HCA of PC1 and PC2	50
Figure 4.14. Scores of filtered metabolome data on first four Latent Variables at 20 th hour of extracellular samples.....	52
Figure 4.15. Scores of filtered metabolome data on first three IC Values at 20 th hour of extracellular samples (PCA preprocessing number is 5).....	53
Figure 4.16. Scores of filtered metabolome data on first three IC Values at 20 th hour of extracellular samples (PCA preprocessing number is 6).....	54
Figure 4.17. Scores of binned metabolome data on first four Latent Variables at 25 th hour of extracellular samples	55
Figure 4.18. Scores of binned metabolome data on first three IC Values at 25 th hour of extracellular samples (PCA preprocessing number is 5)	56
Figure 4.19. Scores of binned metabolome data on first three IC Values at 25 th hour of extracellular samples (PCA preprocessing number is 6).....	57
Figure 4.20. Scores of filtered metabolome data on first four Latent Variables at 25 th hour of extracellular samples.....	59
Figure 4.21. Scores of filtered metabolome data on first three IC Values at 25 th hour of extracellular samples (PCA preprocessing number is 5).....	60

Figure 4.22.	Scores of filtered metabolome data on first three IC Values at 25 th hour of extracellular samples (PCA preprocessing number is 6).....	61
Figure 4.23.	Scores of binned metabolome data on first four Latent Variables at 31 st hour of extracellular samples	62
Figure 4.24.	Scores of binned metabolome data on first three IC Values at 31 st hour of extracellular samples (PCA preprocessing number is 5)	63
Figure 4.25.	Scores of binned metabolome data on first three IC Values at 31 st hour of extracellular samples (PCA preprocessing number is 6).....	64
Figure 4.26.	Scores of binned replicate metabolome data on first four Latent Values at 31 st hour of extracellular samples	65
Figure 4.27.	Scores of binned replicate metabolome data on first three ICs at 31 st hour of extracellular samples (PCA preprocessing number is 5).....	66
Figure 4.28.	Scores of binned replicate metabolome data on first three ICs at 31 st hour of extracellular samples (PCA preprocessing number is 6).....	67
Figure 4.29.	Scores of filtered metabolome data on first four Latent Variables at 31 st hour of extracellular samples.....	68
Figure 4.30.	Scores of filtered metabolome data on first three IC Values at 31 st hour of extracellular samples (PCA preprocessing number is 5).....	69
Figure 4.31.	Scores of filtered metabolome data on first three IC Values at 31 st hour of extracellular samples (PCA preprocessing number is 6).....	70
Figure 4.32.	Scores of filtered replicate metabolome data on first four Latent Values at 31 st hour of extracellular samples	71
Figure 4.33.	Scores of filtered replicate metabolome data on first three ICs at 31 st hour of extracellular samples (PCA preprocessing number is 5).....	72

Figure 4.34. Scores of filtered replicate metabolome data on first three ICs at 31 st hour of extracellular samples (PCA preprocessing number is 6).....	73
Figure 4.35. Scores of binned all metabolome data on first four Latent Variables at four time intervals of extracellular samples	74
Figure 4.36. Kurtosis measure of ICA data	75
Figure 4.37. Scores of binned all metabolome data on first three Independent Component Values at four time intervals of extracellular samples (PCA preprocessing number is 3).....	76
Figure 4.38. Mass spectral data of intracellular metabolites belonging to 17 mutants at 20 th hour.....	77
Figure 4.39. Mass spectra of extraction solution(acetonitrile:water:formic acid).....	79
Figure 4.40. Mass spectra of hot ethanol solution for extraction of intracellular metabolites	80
Figure 4.41. Mass spectra of methanol solution for extraction of intracellular metabolites	80
Figure 4.42. Scores of binned metabolome data on first four Latent Variables at 20 th hour of intracellular samples	81
Figure 4.43. Scores of binned metabolome data on first three IC Values at 20 th hour of intracellular samples (PCA preprocessing number is 5)	82
Figure 4.44. Scores of binned metabolome data on first three IC Values at 20 th hour of extracellular samples (PCA preprocessing number is 6)	83
Figure 4.45. Scores of filtered metabolome data on first four Latent Variables at 20 th hour of intracellular samples.....	84

Figure 4.46.	Scores of filtered metabolome data on first three IC Values at 20 th hour of intracellular samples (PCA preprocessing number is 5).....	85
Figure 4.47.	Scores of filtered metabolome data on first three IC Values at 20 th hour of intracellular samples (PCA preprocessing number is 6).....	86
Figure 4.48.	Scores of binned metabolome data on first four Latent Variables at 25 th hour of intracellular samples	87
Figure 4.49.	Scores of binned metabolome data on first three IC Values at 25 th hour of intracellular samples (PCA preprocessing number is 5)	88
Figure 4.50.	Scores of binned metabolome data on first three IC Values at 25 th hour of extracellular samples (PCA preprocessing number is 6).....	89
Figure 4.51.	HCA of IC1 and IC2 by 5 times pre-processing of PCA	90
Figure 4.52.	Scores of filtered metabolome data on first four Latent Variables at 25 th hour of intracellular samples.....	91
Figure 4.53.	Scores of filtered metabolome data on first three IC Values at 25 th hour of extracellular samples (PCA preprocessing number is 5).....	92
Figure 4.54.	Scores of filtered metabolome data on first three IC Values at 25 th hour of extracellular samples (PCA preprocessing number is 6).....	93
Figure 4.55.	Scores of binned metabolome data on first four Latent Variables at 31 st hour of intracellular samples	94
Figure 4.56.	Scores of binned metabolome data on first three IC Values at 31 st hour of intracellular samples (PCA preprocessing number is 5)	95
Figure 4.57.	Scores of binned metabolome data on first three IC Values at 31 st hour of intracellular samples (PCA preprocessing number is 6).....	96

Figure 4.58.	Scores of filtered metabolome data on first four Latent Variables at 31 st hour of intracellular samples.....	97
Figure 4.59.	Scores of filtered metabolome data on first three IC Values at 31 st hour of intracellular samples (PCA preprocessing number is 5).....	98
Figure 4.60.	Scores of filtered metabolome data on first three IC Values at 31 st hour of intracellular samples (PCA preprocessing number is 6).....	99
Figure 4.61.	Scores of binned all metabolome data on first four Latent Variables at three time intervals of intracellular samples	100
Figure 4.62.	Kurtosis measure of ICA data	102
Figure 4.63.	Scores of binned all metabolome data on first three IC Values at three time intervals of intracellular samples (PCA preprocessing number is 5)	103
Figure 5.1.	Loadings of PC1 vs PC2 at 20 th hour of extracellular samples	107
Figure 5.2	Loadings of PC3 vs PC4 at 20 th hour of extracellular samples	108
Figure 5.3.	Loadings of PC1 vs PC2 at 25 th hour of intracellular samples	108
Figure 5.4.	Loadings of PC3 vs PC4 at 25 th hour of intracellular samples	109
Figure 5.5.	Mass spectrum of 8 deletion mutants (extracellular samples taken at 20th hour of fermentation)	110
Figure 5.6.	Pair-wise mutant correlations and intensity values at peak conditions.....	111
Figure 5.7.	Mass spectrum of 8 deletion mutants (intracellular samples taken at 25th hour of fermentation)	112
Figure 5.8.	Pair-wise mutant correlations and intensity values at peak conditions.....	113

Figure 5.9.	Pearson correlation maps of binned metabolome data at 15 th hour of extracellular samples	114
Figure 5.10.	Pearson correlation maps of filtered metabolome data at 15 th hour of extracellular samples.....	115
Figure 5.11.	Pearson correlation maps of binned metabolome data at 20 th hour of extracellular samples.....	115
Figure 5.12.	Pearson correlation maps of filtered metabolome data at 20 th hour of extracellular samples	116
Figure 5.13.	Pearson correlation maps of binned metabolome data at 25 th hour of extracellular samples	116
Figure 5.14.	Pearson correlation maps of filtered metabolome data at 25 th hour of extracellular samples.....	117
Figure 5.15.	Pearson correlation maps of binned metabolome data at 31 st hour of extracellular samples.....	117
Figure 5.16.	Pearson correlation maps of filtered metabolome data at 31 st hour of extracellular samples.....	118
Figure 5.17.	Pearson correlation maps of binned metabolome data at 20 th hour of intracellular samples.....	118
Figure 5.18.	Pearson correlation maps of filtered metabolome data at 20 th hour of intracellular samples.....	119
Figure 5.19.	Pearson correlation maps of binned metabolome data at 25 th hour of intracellular samples.....	119
Figure 5.20.	Pearson correlation maps of filtered metabolome data at 25 th hour of intracellular samples.....	120

Figure 5.21. Pearson correlation maps of binned metabolome data at 31 st hour of intracellular samples.....	120
Figure 5.22. Pearson correlation maps of filtered metabolome data at 31 st hour of intracellular samples.....	121
Figure 5.23. Color Code for Pearson Correlation Coefficients	121

LIST OF TABLES

Table 3.1.	F1 synthetic defined medium.....	20
Table 3.2.	Glucose and vitamin solution	21
Table 3.3.	Trace elements 1	21
Table 3.4.	Trace elements 2	21
Table 3.5.	Vitamin solution	21
Table 3.6.	Laboratory Equipment	22
Table 5.1.	Summary of the Discrimination Results Obtained by Binning Approach (Extracellular Samples).....	105
Table 5.2.	Summary of the Discrimination Results Obtained by Filtering Approach (Extracellular Samples)	106
Table 5.3.	Summary of the Discrimination Results Obtained by Binning Approach (Intracellular Samples)	106
Table 5.4.	Summary of the Discrimination Results Obtained by Filtering Approach (Intracellular Samples)	107
Table 5.5.	m/z values with the highest absolute loadings in the first four PCs (extracellular samples)	108
Table 5.6.	m/z values with the highest absolute loadings in the first four PCs (intracellular samples)	109
Table 5.7.	Go Terms from the Process Ontology	124
Table 5.8.	Go Terms from the Function Ontology	124

LIST OF SYMBOLS/ABBREVIATIONS

n	Number of replicates
p	P-value
X	Metabolic data matrix
Δ	Gene deletion
ρ	Pearson correlation coefficient
<i>AAA1</i>	Gene name
<i>aaa1Δ/aaa1Δ</i>	Homozygous deletion mutant of gene AAA1
ADH	Alcohol dehydrogenase
NADPH	Nicotinamide adenine dinucleotide phosphate-oxidase
GAL	Galactose
GO	Gene ontology
HEPES	4-2-hydroxyethyl-1-piperazineethanesulfonic acid
MS	Mass spectrometry
QTOF	Quadropole-time of flight
FWHM	Full-width half-height maximum
PC	Principal component
HPLC	High performance liquid chromatography
PCA	Principal component analysis
IC	Independent component
ICA	Independent component analysis
LC	Liquid chromatography
HCA	Hierarchial cluster analysis
LV	Latent variable
IR	Infrared spectroscopy
TLC	Thin-layer chromatography

MALDI-TOF

Matrix-assisted laser desorption/ionization – time of flight

YPD

Yeast extract – Peptone – D-glucose

1. INTRODUCTION

Advances in high-throughput technologies have led to a rapidly increased number of simultaneously measured expression levels of various genes as well as concentration levels of metabolites or proteins. Thus, there is a great demand for methods for analysing such profile data of many variables. Although determining the concentration of individual and targeted metabolites has a long tradition in biochemistry (apRees, 1980; Stitt et al., 1988), only recent advances in mass spectrometry allowed to establish methods that offer both high accuracy and sensitivity for the non-targeted measurement of metabolites in complex mixtures (Fernie et al., 2004).

In particular, for *metabolomics* – termed in analogy to other ‘omics’ approaches such as transcriptomics and proteomics – classical methodologies like gas chromatography coupled to mass spectrometry (GC–MS) and liquid chromatography coupled to MS (LC–MS), have developed into important methods (Weckwerth, 2003). Yet, a variety of alternatives exists which differ in various aspects such as metabolite coverage, sensitivity, accuracy and instrumentation (Goodacre et al., 2004; Fiehn, 2001, 2002; Sumner et al., 2003; Kell, 2004). It should be noted, though metabolomics ultimately aims at an unbiased and comprehensive monitoring of all metabolites in a cellular system, all current experimental techniques are still far from this. Nonetheless, recent advances using a time of flight mass analyzer resulted in the detection of several hundred distinct metabolic compounds in parallel with a reasonable throughput of samples (20 min chromatography per sample) (Weckwerth et al., 2001, 2004a,b; Morgenthal et al., 2005).

Clearly, and similar to other ‘omic’ technologies aiming to measure many variables simultaneously, this torrent of data necessitates the development of novel algorithms and approaches to analyze the data—and eventually turn it into knowledge. To this end, metabolomics greatly profits from the computational methods already developed in related fields, such as transcriptomics or proteomics.

Indeed, most popular types of analysis are based on clustering, principal component analysis (PCA) and various other unsupervised or supervised machine learning algorithms (Goodacre et al., 2004; Sumner et al., 2003). Though often perceived as ‘black box’ methods, their power has already been demonstrated (Kell, 2002; Taylor et al., 2002; Goodacre et al., 2004).

The incorporation of metabolomics into systems-biology initiatives has just begun: the unbiased metabolomics approach opens the way to investigate the organization, regulation and control of metabolic networks experimentally and will profoundly contribute to the understanding of metabolism.

Many modern techniques were developed for metabolome analysis. It is therefore not surprising that many of the machine-learning tools became important in bioinformatics. In this study, metabolome data was collected from *S. cerevisiae* deletion mutants and was processed by different statistical techniques such as PCA, ICA, and HCA. The experiments conducted in batch system and the list of the materials and methods used are given in “Materials and Methods” section.

The results of this study are explained under two main topics: Possible clusterings of galactose metabolic pathway genes; hexose transporter genes and regulation related genes of *S. cerevisiae* were investigated under metabolite footprinting and fingerprinting sections. 14 of the used deletion mutants were annotated with known functions. A GO Term Analysis was carried out to find evidence to support these clusterings in PCA and ICA.

In the “Conclusions and Recommendations” chapter, an overall summary of the results is given and future applications are discussed.

2. LITERATURE SURVEY

2.1. Scope of Metabolomics

Since last decade, much time has been invested into the study of the genetic code of organisms ('genomics'), as well as the transcription ('transcriptomics') and translation ('proteomics') of the information contained within the genetic code to yield the current protein content of cells. In the interest of understanding metabolism from a quantitative aspect, it is, however, necessary to take this study further, namely to include the analysis of small molecular weight metabolites ('metabolomics'). Metabolomics covers the quantification of (microbial) intracellular and extracellular metabolite concentrations with the help of sophisticated analytical devices. Since metabolites represent the reactants of the metabolic network, quantification of metabolite concentrations offers a direct access to the identification of internal metabolism kinetics. Being able to quantify changes of internal metabolite concentrations in appropriate time scales, reaction rates can be derived that provide a quantitative understanding of metabolite production and consumption rates. Hence, the estimation of 'in vivo' kinetics can be achieved, if suitable experimental conditions were realised that allow the interpretation of internal metabolite concentration changes only with respect to microbial metabolism and not due to experimental conditions.

The study of the metabolome in different physiological states will provide useful information regarding metabolism and regulation and will thus complete the data provided by the study of the genome, transcriptome and the proteome. Metabolomics may be practised by a number of different ways, depending mainly on the data of interest. This then determines the type of fermentation experiment, the method of sampling, and the mathematical interpretation of the data.

The quantification of intracellular metabolites may be applied to only a single metabolite or a small, isolated group of metabolites. In this case, a punctual investigation of one or two metabolites in different phases of a fermentation experiment (i.e. exponential stationary, etc.,

or fedbatch) may shed light on limiting or inhibiting effects in metabolism, the activity of specific metabolic pathways, or simply provide a metabolite concentration which could be optimised in an iterative strain improvement. Typically, this sort of experiment would have a very low sampling frequency (min^{-1}). The data gained could be easily interpreted without the use of extensive modelling.

Metabolomics also represent a useful tool for functional genomics programmes. In order to elucidate an unknown gene function, genetic alterations are introduced into the system. By analysing the phenotypic effect of such a mutation (i.e. by analysing the metabolome), functions may be assigned to the respective gene (Fiehn, 2001 and Raamsdonk *et al.*, 2001). Sampling would be performed once for every genotypic mutation and the analysis of the metabolome would optimally include an identification and quantification of all metabolites of a biological system (Fiehn *et al.*, 2000).

Another application of metabolomics is the quantification of (parts of) the cellular metabolism. This includes aspects of regulation and control in the pathway of interest and requires the comprehensive profiling of a large number of metabolites at a very high sampling frequency (correlated to the turnover times of the metabolites, usually on a subsecond scale in bacteria). Stimulus-response experiments are a typical example for this kind of metabolic investigation (i.e. metabolic reaction to a shift or pulse in C-source, temperature, inhibitory substances, etc). The interpretation of data gained from this kind of rapid sampling experiment is too complex to be interpreted empirically and requires extensive modelling. The mathematical analysis of metabolic pathways may allow an understanding of the dynamics involved in metabolic flux and permit a greater comprehension of metabolic control (Tweeddale *et al.*, 1998).

2.2. Yeast

Yeasts are basically recognized as being unicellular fungi that reproduce by budding or fission, and that form sexual states which are not enclosed in a fruiting body (Walker, 2000). Yeasts are easily grown unicellular eucaryotes, which portray the ultra structural features of

higher eucaryotic cell. Yeasts are also amenable to a variety of biochemical, genetic and molecular biological investigations of their cytology (Walker, 2000). Beside being useful in daily brewing and baking practice, yeast, (as a simple, unicellular eukaryote) developed into a powerful model system for biological research. Its prominent useful features are cheap and easy cultivation, short generation times, detailed genetic and biochemical knowledge accumulated in many years of research and ease of the application of molecular techniques for its genetic manipulation. Thus, this organism provides a highly suitable system to study biological processes that are relevant for many higher eukaryotes including man.

S. cerevisiae is a chemoorganotrophic organism obtaining carbon and energy from compounds in fixed, organic linkage. It can utilize the sugar types glucose, fructose, mannose, galactose, sucrose and maltose by fermentative or respirative metabolism. Other carbon sources (e.g. ethanol, acetate) can be utilized as respiratory substrates. Yeasts cannot fix molecular nitrogen, thus simple inorganic nitrogen sources such as ammonium salts are widely utilized. Ammonium sulphate is a commonly used nitrogen source in yeast growth media since it also provides a source of assimilable sulphur. Yeast require sulphur principally for the biosynthesis of sulphur-containing amino acids (Walker, 1998).

2.3. Yeast Metabolism

The sequence of enzyme-catalyzed reactions that oxidatively convert glucose to pyruvic acid in the yeast catabolism is known as glycolysis (Figure 1). Glycolysis provides yeast with energy, together with precursor molecules and reducing power for biosynthetic pathways. Yeast glycolytic intermediates are proved to be valuable industrial commodities.

There are a number of biotechnologically important yeast strains that use organic compounds anaerobically as electron donors, acceptor and as carbon source. In alcoholic fermentation NADH is reoxidized to NAD in terminal step reactions from pyruvate. First of these reactions is catalyzed by pyruvate decarboxylase. Pyruvate is decarboxylated before a final reduction, catalyzed by alcohol dehydrogenase to ethanol. An alternative way to generate NAD is the pathway from dihydroxyacetone phosphate to glycerol. Glycerol is quantitatively

one of the most important products of the yeast alcoholic fermentation, besides ethanol and carbon dioxide (Phaff *et al.*,1978).

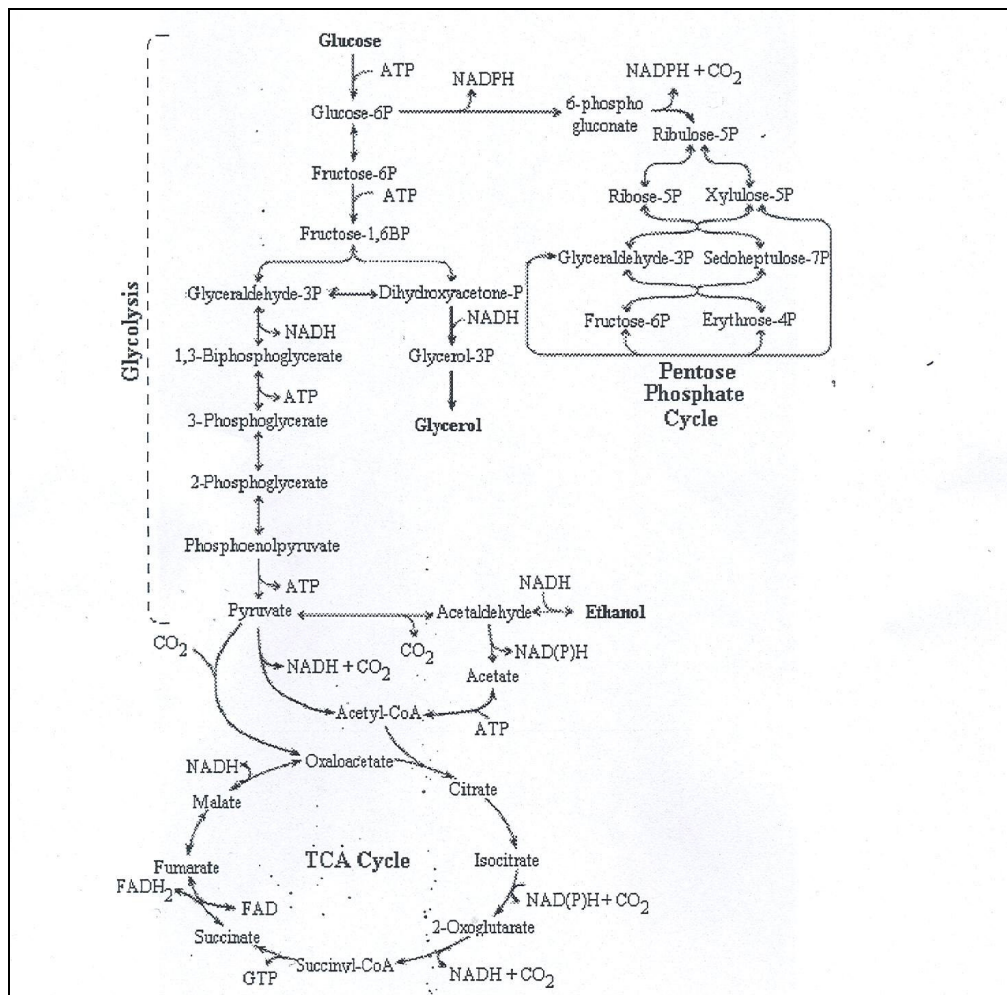


Figure 2.1. Metabolism of yeast

Under aerobic conditions, energy generation from glucose is oxidative decarboxylation of pyruvate to form acetyl coenzyme A (acetyl CoA). This activated acetyl group is then completely oxidized to CO₂ by the citric acid cycle, a series of reactions that is also known as the tricarboxylic acid cycle (TCA cycle) or the Krebs cycle (Figure 1).

The isolation and characterization of mitochondrial mutants of *S. cerevisiae* has provided fundamental biochemical, genetic and molecular biological information on the biogenesis and function of mitochondria. Moreover, respiratory-deficient mutants of yeasts have been studied with regard to their possible exploitation in fermentation biotechnology. For example, the use of petite mutants of *S. cerevisiae* in alcohol fermentation has been suggested since alcohol yields can be quite high due to the cells' insensitivity to oxygen and their minimal diversion of carbon to cell growth. However, respiratory-deficient petites grow more slowly than wild-type 'grande' strains of *S. cerevisiae* (Hutter and Oliver, 1998).

2.4. Literature Search on the Genes Investigated in This Study

Seventeen deletion mutants of *S. cerevisiae* are investigated for metabolic profiling. The following sections summarize the information in the literature about these genes and their products.

2.4.1. Ho

Baganz et al. (1997), constructed a different standard strain in which the *ho* gene is replaced by kanMX. They reported that the kanMX marker exhibited only a small effect on specific growth rate (<4 per cent). Therefore, *hoΔ* was chosen as the reference strain because it was found to be a selectively neutral site for gene replacement, it has no known role apart from mating-type switching (inactive in diploids) and it has been used as the site of insertion of heterologous genes in brewing yeasts without any perceptible effect on the fermentation characteristics of the organism or quality of the product (Oliver *et al.*, 1998).

2.4.2. Hxt Genes

Although these genes were named *HXT* because of their homology to the members of the *HXT* gene family, it remains to be determined whether they indeed encode glucose transporters. Some of these genes might be involved in the transport of sugars other than

glucose (though there are not many candidates since *S. cerevisiae* grows on few sugars), and some could have only a regulatory function. *Hxt10*, which encodes a protein approximately 80% identical to *hxt2*, does not complement the *hxt* null mutant, so it appears unable to transport a significant amount of glucose. It is a putative hexose transporter, expressed at low levels and expression is repressed by glucose.

Hxt12 was identified through the systematic sequencing project as having high similarity to other representatives of the 18-20 member *S. cerevisiae* hexose transporter family. *Hxt12* is represented in SGD as two separate ORFs, YIL170w and YIL171w, and may be a pseudogene; deletion of two nucleotides from the systematic sequence is required to yield the predicted Hxt12p sequence *hxt12*, linked to the telomere of chromosome IX-L, is most similar to subtelomeric genes *hxt11* (chromosome XV-L; 98% amino acid identity) and *hxt9* (chromosome X-L; 97% amino acid identity). The high similarity of *hxt12*'s sequence to other HXT genes suggests that either it is a recently formed pseudogene, or there is an error in the systematic sequence. When 18 members of the hexose transporter family were amplified from genomic DNA and overexpressed, *hxt12* was the only one unable to transport hexoses, supporting its designation as a pseudogene. In common with all the *HXT* genes, deletions of YIL170w and YIL171w are viable (Kruckeberg, 1996).

2.4.3. Yel057c

Yel057c is a protein of unknown function involved in telomere maintenance. Telomeres are nucleoprotein structures present at the ends of eukaryotic chromosomes that play a central role in guarding the integrity of the genome by protecting chromosome ends from degradation and fusion. Length regulation is central to telomere function (Askree *et al.*, 2004).

2.4.4. Snf3

The snf3 protein is composed of distinct cytoplasmic and integral-membrane domains and functions as a low glucose sensor required for the expression of hexose transporters (the *HXT* genes) in *S. cerevisiae*. The C-terminal domain, when expressed independently of the

integral membrane domain, leads to glucose-independent expression of *hxt2* on gluconeogenic carbon sources (Ozcan and Johnston, 1999).

2.4.5. Rgt1

The *rgt1* gene of *Saccharomyces cerevisiae* plays a central role in the glucose-induced expression of hexose transporter (*HXT*) genes. Genetic evidence suggests that it encodes a repressor of the *HXT* genes whose function is inhibited by glucose. *Rgt1* encodes a bifunctional transcription factor.

Rgt1p displays three different transcriptional modes in response to glucose: (i) in the absence of glucose, it functions as a transcriptional repressor; (ii) high concentrations of glucose cause it to function as a transcriptional activator; and (iii) in cells growing on low levels of glucose (Ozcan *et al.*, 1996).

2.4.6. Mig1

Mig1 is a transcriptional repressor responsible for glucose repression of many genes in the budding yeast *Saccharomyces cerevisiae*. Sucrose metabolism was alleviated of glucose control in the *mig1* disruptant, and even more so in the *mig1mig2* disruptant compared with their wild-type strain. Phosphorylation alters the subcellular localization of *mig1*, causing it to be nuclear when glucose is present, and cytoplasmic when glucose is absent. One mechanism represses transcription of some *HXT* genes when glucose levels are high and works through the *mig1* transcriptional repressor, whose function is regulated by the Snf1-Snf4 protein kinase and Reg1-Glc7 protein phosphatase.

Deletion of the *mig1*-coding gene reduced but did not eliminate glucose repression mediated by the *mig1* site. A major mediator of glucose repression in yeast is *mig1*, a zinc finger protein that binds to a GC-rich recognition sequence found upstream of many glucose-repressible genes. The *mig1* repressor is a zinc finger protein that mediates glucose repression in yeast (Carlson, 1999).

2.4.7. Mba1

Mba1p is thought to have a similar function as Oxa1p. Mba1p can act either in cooperation with Oxa1p or may accomplish a similar function independently in protein insertion machinery of *S. cerevisiae* mitochondria (Preuss *et al.*, 2001). Strains with the *mba1Δ* deletion suffer partial respiratory deficiency, thus its protein may be involved in assembly of respiratory chain (Rep and Grivell., 1996). Pir *et al.* (2006) reported that homozygous *mba1Δ* deletants suffer partial respiratory deficiency. However, the expression behaviour of *mba1Δ* deletion mutants is similar to *mig1Δ* deletion mutants. Moreover, *mba1Δ* deletion causes *mig1Δ* expression to be repressed. Thus, it may be concluded that Mba1p is an activator of *mig1* and deletion of these two genes cause the same expression profiles (Pir *et al.*, 2006).

2.4.8. Hap4

The function of the Hap4p protein was reported to be the regulation of the respiration related-genes (DeRisi *et al.*, 1997). The activation mechanism of *hap2/3/4/5* protein complex was reviewed by Gancedo (1998). Hap4 deletion mutants were reported to be respiratory deficient (Steinmetz *et al.*, 2001) and deletion of *hap4Δ* caused downregulation of the respiration-related genes (Buschlen *et al.*, 2003). On the contrary, these genes were expressed at higher levels in *hap4*-overexpressing strains growing under aerobic conditions and, increased levels of respiratory capacity were observed due to overexpression of *hap4* (Blom *et al.*, 2004).

2.4.9. Gal4

Gal4 is a DNA-binding transcription factor required for the activation of the *GAL* genes in response to galactose; repressed by Gal80p and activated by Gal3p. *Gal4* encodes one of the best studied transcription factors in the explored universe (Keegan *et al.*, 1986). In the presence of galactose, *Gal4p* activates transcription of the *GAL* structural genes, which encode

galactose metabolic proteins. Gal4p binds to a specific sequence, the UAS_G, in the promoters of *GAL* structural genes (Lohr *et al.*, 1995).

2.4.10. Gal6

Null mutant of *gal6Δ* is viable with no obvious growth defects but is leucine aminopeptidase deficient and hypersensitive to bleomycin; overexpression confers resistance to bleomycin. Deletion of this gene, *gal6Δ*, leads to higher expression of *GAL*-regulated genes. This implies that *gal6* is a newly defined negative regulator of the *GAL* system and part of an autoregulatory circuit (Zheng *et al.*, 1997).

2.4.11. Gal7

The Gal7p galactose-1-phosphate uridylyltransferase catalyzes the conversion of galactose-1-phosphate to UDP-galactose, a key step in galactose catabolism. The enzyme is found in the cytoplasm and functions as a dimer. When galactose is the sole carbon source, *gal7Δ* null mutants are unable to grow and accumulate high levels of galactose-1-phosphate (Christacos *et al.*, 2000).

All the galactose structural genes (*gal1*, *gal10*, *gal7*, *gal2*) are coordinately regulated at the level of transcription in response to galactose by Gal4p, Gal80p, and Gal3p. Regardless of carbon source, the Gal4p transcriptional activator is bound as a dimer to upstream activation sites found in the promoters of the *GAL* genes. In the presence of galactose, Gal3p sequesters the transcriptional repressor Gal80p in the cytoplasm, thereby relieving inhibition of Gal4p and resulting in *GAL* gene expression (Riehmman *et al.*, 2001).

2.4.12. Gal80

Gal80 encodes a transcriptional repressor involved in transcriptional regulation in response to galactose. In conjunction with Gal4p and Gal3p, Gal80p coordinately regulates the

expression of genes containing a *GAL* upstream activation site in their promoter (UAS-GAL), which includes the *GAL* metabolic genes *gal1*, *gal10*, *gal2*, and *gal7*. Cells null for *gal80Δ* constitutively express *GAL* genes, even in non-inducing media (Torchia *et al.*, 1984).

In the absence of galactose, Gal80p binds, as a dimer, to the transcription factor Gal4p, thus preventing Gal4p from recruiting other factors of the Pol II transcription machinery. In the presence of galactose, the transcriptional regulator Gal3p sequesters Gal80p in the cytoplasm which relieves inhibition of Gal4p and results in *GAL* gene expression. Gal80p interacts with Gal3p as a monomer and only in the presence of both galactose and ATP. In the absence of Gal3p, the Gal1p galactokinase can interact with and partially inhibit Gal80p function (Timson *et al.*, 2002). *Gal80* transcription is also regulated by the Gal3p/Gal80p/Gal4p feedback mechanism that regulates the other *GAL* genes. However, unlike expression of the *GAL* structural genes, which is tightly controlled because of the presence of multiple UAS-GALs in their promoters, *Gal80* has high basal levels of transcription due to the presence of only a single UAS-GAL in its promoter (Melcher *et al.*, 2001).

2.4.13. Adh6

Adh6 and *adh7* are non-essential genes that encode the only *S. cerevisiae* members of the cinnamyl alcohol dehydrogase (CAD) family. CADs are a subgroup of medium-chain zinc-containing alcohol dehydrogenases that were originally identified in plants where they are involved in the biosynthesis of lignin. However, since lignin biosynthesis does not occur in yeast, these enzymes are instead thought to participate in lignin degradation, the synthesis of fusel alcohols, and NADPH homeostasis (Larroy *et al.*, 2002).

2.5. Metabolome Analysis

Metabolites are produced at the last step of the cellular machinery and define the biochemical phenotype of a cell or tissue. Quantitative and qualitative measurements of large

numbers of metabolites thus provide a broad view of the biochemical status of an organism that can be used to monitor and assess gene function. Profiling of the transcriptome and proteome has received some criticism due to its inability to predict gene function in some cases. Although the transcriptome represents the delivery mechanism of a translational code to the cellular machinery for protein synthesis, increases in mRNA levels do not always correlate with increases in protein levels (Gygi *et al.*, 1999). Furthermore, once translated, a protein may or may not be enzymatically active. Due to these factors, changes in the transcriptome or in the proteome do not always correspond to alterations in biochemical (i.e. metabolic) phenotypes. Another consideration when profiling the transcriptome and proteome is that most modern techniques identify mRNA and proteins through sequence similarity or database matching; thus, identification is based primarily on the quality of the match and is therefore indirect. In the absence of existing database information, transcript or protein profiling yields only limited information. Based on the above limitations, profiling the metabolome may actually provide the most “functional” information of the “omics” technologies. There are, nevertheless, many instances in which transcriptome and proteome profiling have successfully pointed the observer to functional information, and therefore an integrated approach is preferred when resources permit (Sumner *et al.*, 2003).

For a number of metabolites, kits for concentration measurement using enzymatic methods are commercially available. Though the results obtained are reliable and concentrations are obtained readily without utilizing any software, these methods are time consuming when a large number of samples is to be handled, moreover kits are available for a limited number of metabolites like glucose, ethanol, and a few acids like succinate and pyruvate. The kits are suitable and satisfactory if a low number of samples is to be analysed with the aim of determining basic metabolome data like amount of glucose utilized or ethanol produced.

Since the techniques for obtaining mRNA and protein profiles allow thousands of species to be detected in one reaction, it is desired to have the corresponding metabolome data to follow the cell machinery until the last step. It is possible to screen for a high number of metabolites simultaneously using analytical tools such as chromatography and spectroscopy.

Methods based on infrared spectroscopy (IR), nuclear magnetic resonance (NMR), thin-layer chromatography (TLC), HPLC with ultraviolet and photodiode array detection (LC/UV/PDA), capillary electrophoresis coupled to ultraviolet absorbance detection (CE/UV), capillary electrophoresis coupled to laser-induced fluorescence detection (CE/LIF), capillary electrophoresis coupled to mass spectrometry (CE/MS), gas chromatography-mass spectrometry (GC/MS), liquid chromatography-mass spectrometry, (LC/MS), liquid chromatography tandem mass spectrometry (LC/MS/ MS), HPLC coupled with both mass spectrometry and nuclear magnetic resonance detection (LC/NMR/ MS) and LC/NMR/MS/MS have all been used to measure metabolites (Sumner *et al.*, 2003). The selection of the most suitable technology is generally a compromise between speed, selectivity and sensitivity. Tools such as NMR are rapid and selective, but have relatively low sensitivity. Other methods such as capillary electrophoresis coupled to laser induced fluorescence (CE/LIF) detection are highly sensitive, but lack selectivity. Hyphenated mass spectrometry methods such as GC/MS and LC/MS offer good sensitivity and selectivity, but relatively longer analysis times (Sumner *et al.*, 2003).

2.5.1. Metabolic Fingerprinting and Footprinting

Mutations, natural variation, e.g. ecotypes, and environmental conditions can all influence metabolic processes. A large number of individual analyses are required to understand the link between these factors and the overall characteristics of an organism. All of these analytical approaches cannot be done by full-composition metabolomic tests, but instead call for a cheaper and faster first-round screening method that groups data according to inherent biological characteristics and distinguishes these from inherent, unrelated background noise. Such strategies, without individually determining metabolite identities, have been termed metabolite fingerprinting and footprinting (Scholz *et al.*, 2004).

If rapid classification of sample types is needed, for example for diagnostic purposes, product quality control, or screening of mutant collections, even faster methods can be applied. The general idea behind metabolic fingerprinting and footprinting is that, even without any chromatographic separation the resolution power of nuclear magnetic resonance

(NMR), mass spectrometry (MS) or infrared (IR) spectroscopy is still good enough to yield valid information about the net result of metabolic regulation in the biological sample. For example, bacteria species could be classified by pyrolysis-MS (Taylor *et al.*, 1998), but it remained unclear if this classification could be achieved when mixtures of bacteria are analysed. Using a combination of pyrolysis-MS and Fourier transform IR spectrometry, bacteria species have been classified with novel programming tools, and biomarkers were identified that were directly distinguishing these species (Goodacre *et al.*, 2000).

Nuclear magnetic resonance (NMR) has also been used without prior separation of individual compounds, in order to classify yeasts that had been dosed with different amounts of various metabolites. Classification was achieved through principal component-analysis (PCA), and metabolite effects on yeasts could be detected at much lower levels compared to visual inspection of the yeast phenotypes. The data acquired were evaluated by means of PCA and hierarchical clustering.

A small number of extracellular and intracellular metabolites of *S. cerevisiae* have been quantified by many research groups using enzymatic methods with aim of defining the metabolic state of yeast under various conditions. These limited data were obtained by enzymatic methods, HPLC or NMR and frequently used for the kinetic modelling of central metabolism. However, in the post-genomic era, metabolome data are expected to be a powerful tool for functional analysis.

The term “metabolic fingerprinting” corresponds to determination of the intracellular metabolome. Comprehensive analysis of metabolites within yeast cell extracts was done using FTIR, ES-MS and NMR spectroscopy. Unfortunately, measuring intracellular metabolites is time-consuming and subject to technical difficulties caused by the rapid turnover of intracellular metabolites. Thus, this method is difficult to scale-up for high throughput screening. A functional analysis study was done by Raamsdonk *et al.* (2001) that used metabolic fingerprinting to reveal the phenotype of silent mutations of *S. cerevisiae*.

Metabolic fingerprinting describes the unbiased analysis of the metabolome by examination of metabolite patterns in different experimental groups and subsequent classification of the patterns. Samples can be classified if the metabolite fingerprints change between groups, resulting in sample clustering. In mass spectrometry-based investigation, metabolite fingerprints are described by m/z values and corresponding intensities of detected ions. If a separation step is performed, retention times are also used to index metabolites. Thus, m/z values, retention times, and intensities represent the metabolic fingerprint of the analyzed sample and are exported for sample classification using multivariate data analysis. The chemical structure of the detected metabolites typically remains unknown (Dettmer *et al.*, 2005).

Comprehensive metabolomic analyses cannot be achieved without pre-fractionation steps, chromatographic separation, and use of different analytical instruments. Therefore, each sample has to be portioned into a (limited) number of aliquots, reducing the total sample throughput. If a higher number of samples need to be analysed, for example for rapid classifications, even faster methods can be applied that completely refrain from sample clean-up steps or time consuming chromatography. This might be needed for diagnostic purposes in the clinical routine, for product quality controls, or for analysing large mutant collections in functional genomics programmes. The bottom line of metabolic fingerprinting is to obtain enough information to unravel (otherwise hidden) metabolic alterations, without aiming to get quantitative data for all biochemical pathways. Therefore, the resolution of the analytical devices must be high enough to handle critical information. Such devices as nuclear magnetic resonance, mass spectrometry, or Fourier transform infrared spectroscopy (FT-IR) provide this resolution. Using a combination of pyrolysis mass spectrometry and FT-IR, bacterial species have been classified using novel programming tools, resulting in potential biomarkers then used to rapidly distinguish among these species (Goodacre *et al.*, 2000). A similar approach was taken by Smedsgaard and Frisvad (1996), who used direct infusion of crude fungal extracts into MS/MS instruments in order to classify ten different fungal species. NMR was used to detect effects of toxins on rats via the direct analysis of dried urine samples, and principle components analysis for classification of metabolic alterations (Robertson *et al.*, 2000). However, metabolic fingerprinting can easily be over-interpreted, since signals suitable

for distinguishing among samples might not be biologically relevant, or might not be applicable when distinguishing among samples from other species (or situations). For example, Warne *et al.*, (2000) studied metabolic effects by NMR after dosing earthworms with toxins. By pattern recognition, they noted elevated levels of glucose, citrate, and succinate as potential biomarkers for toxicity. However, there are clearly a lot of situations where intermediates of the TCA cycle become elevated, and generalisations about the suitability of this method for detecting toxic effects should be avoided. For example, differences in the levels of TCA intermediates were also found by NMR analyses when investigating urine samples from mutant mice (Fiehn, 2001).

In the realm of functional genomics, NMR was used to detect metabolic phenotypes in yeast mutants that did not show obvious visible phenotypes. However, the informative power of NMR was not sufficient in this instance to quantify individual metabolite levels; enzymatic analysis had to be applied additionally (Raamsdonk *et al.*, 2001).

It has long been known that microorganisms and humans (Nicholson *et al.*, 2004) secrete a large number of metabolites into their external environment (especially under conditions of unbalanced growth). Fermented beverages provide an obvious example of what is now known to be microbial activity. Other examples include staling factors, uracil and xanthine, 1,5-anhydroglucitol, furanones, a vast array of intercellular signaling molecules, many charged but uncharacterized metabolites and the many biologically active products of both ‘primary’ and ‘secondary’ metabolism, which are often of applied interest (Devlin, 1997).

As such secretory activities clearly reflect cellular metabolic activity (and the level of transcription and translation of relevant genes), it occurred to researchers that the global measurement of secreted metabolites might be exploited to analytical advantage. In particular, the rapid quenching required to profile intracellular metabolites would be avoided, as the turnover time of metabolites diluted into the greater extracellular space is decreased in proportion to the relative ratios of extracellular and intracellular volumes. Although the characterization of metabolites in microbial culture broths has already been carried out using various detection methods (and in culture off-gas using an electronic ‘nose’), such analyses

were largely confined to specific metabolites and have not been done on a global scale. Consequently, metabolic footprinting was devised as a novel method for the functional analysis and characterization of cells using the metabolome. Metabolic footprinting as described relies not on the measurement of intracellular metabolites (a technique which is widely referred to as metabolic fingerprinting) but on the monitoring of metabolites consumed from, and secreted into, the growth medium by batch cultures of yeast using direct injection MS, in which samples of culture media are injected directly into an electrospray ionization mass spectrometer.

To maximize the excretion of metabolites, overflow metabolism is stimulated by adding to the fully defined medium various carbon compounds that ‘probe’ metabolically active networks in the same way that an engineer might probe an electrical circuit. Although such reasoning is not entirely watertight, it is reasonable to infer that those metabolites that do not appear in the medium in a particular gene knockout (or conversely those that increase in concentration when the encoding gene is cloned downstream of a high-expression promoter) are produced through flux- or information-controlling pathways that more or less directly involve the gene product in question.

Metabolic footprinting has proved useful for detecting the patterns of metabolites in single-gene-knockout strains in functional genomic analyses as well as in mode-of-action studies, in which the pattern of metabolites excreted when strains are challenged with sub-lethal concentrations of growth inhibitors makes it possible to discriminate the site or mode of action of those inhibitors. Genetic programming, statistical tool boxes allow the evolution of simple rules that can both discriminate the mode of action and explain how this discrimination works.

Metabolic footprinting focuses on direct mass spectrometric monitoring of extracellular metabolites in spent culture medium. Allen *et al.* (2003) used metabolic footprinting to distinguish between different physiological states of wild-type yeast and between yeast single-deletion mutants even from related areas of metabolism (Figure 2.2). Using principle components analysis (PCA) and genetic programming, Allen *et al.* (2003) showed that

metabolic footprinting is an effective method to classify unknown gene deletions. Footprinting was reviewed recently with all aspects (Kell *et al.*, 2005).

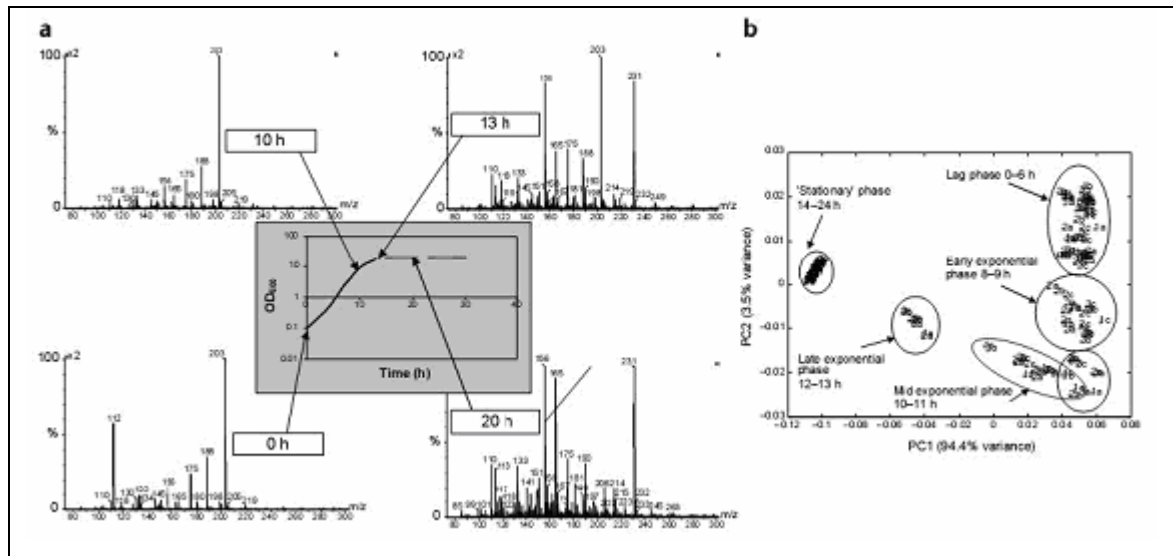


Figure 2.2. Metabolic footprinting of *S. cerevisiae*, a) Representative spectra, b) PCA of data (Allen *et al.*, 2003)

Most of the genes involved in fermentation and respiration by *S. cerevisiae* are known. However, there are unclear points in the regulation of the pathways that are part of fermentation and respiration. Metabolic footprinting and fingerprinting of fermenting or respiring yeast might be used to identify new genes involved in these pathways. Moreover, the extent of fermentation and respiration in an uncharacterized culture might be quantified by comparing its profile with profiles of previous cultivations to make conclusions about the physiology of the culture in question. It is evident that this kind of metabolic footprinting strategy could be useful in ('forward') chemical genetics and/or chemical genomics (Kell *et al.*, 2005).

3. MATERIALS AND METHODS

3.1. Materials

3.1.1. Microorganisms

Homozygous deletion strains of *Saccharomyces cerevisiae*, *rgt1Δ*, *adh6Δ*, *gal4Δ*, *gal6Δ*, *gal7Δ*, *gal80Δ*, *hap4Δ*, *hxt10Δ*, *hxt12Δ*, *mba1Δ*, *mig1Δ*, *snf3Δ*, *yel057cΔ* were used for metabolome analysis. The reference strain was *hoΔ*. These strains were provided by Prof. Stephen Oliver (School of Biological Sciences, Victoria University of Manchester, England).

3.1.2. Chemicals

The chemicals and solutions which were used in this study were purchased from MERCK (Germany), SIGMA (USA), and FLUKA (Switzerland).

3.1.3. Culture Media

The liquid media used in this study were F1 synthetic defined medium which was described by Baganz (1997) and the complex medium (YPG) for precultures. The composition per litre of the medium is given as follows:

Table 3.1. F1 synthetic defined medium

(NH ₄) ₂ SO ₄	3.13 g/l
KH ₂ PO ₄	2.00 g/l
MgSO ₄ .7H ₂ O	0.55 g/l
NaCl	0.10 g/l
CaCl ₂ .2H ₂ O	0.09 g/l
Uracil	0.02 g/l

Table 3.1. continued

L- Histidine	0.02 g/lt
L- Leucine	0.10 g/lt
Trace elements 1	0.10 ml
Trace elements 2	0.10 ml

Table 3.2. Glucose and vitamin solution

Glucose	6.25 ml
Vitamin solution	1.67 ml

Table 3.3. Trace elements 1 [x 10000] (g/L)

ZnSO ₄ .7H ₂ O	0.70
CuSO ₄ .7H ₂ O	0.10
H ₃ BO ₃	0.10
KI	0.10

Table 3.4. Trace elements 2 [x 10000] (g/L)

FeCl ₃ .6H ₂ O	0.50
--------------------------------------	------

Table 3.5. Vitamin solution [x 600] (g/L)

Inositol	37.2
Thiamine/HCl	8.4
Pyroxidine	2.4
Ca-pantothenate	2.4
Biotin	0.18

3.1.4. Laboratory Equipment

List of laboratory equipment and suppliers are given in Table 3.6.

Table 3.6. Laboratory Equipment

Equipment	Supplier
Autoclave	Eyela Model MAC-601 (Japan)
Balance	Precisa 80A-200M (Switzerland)
Centrifuges	SORVALL RC-5B Refrigerated Superspeed Centrifuge, DuPont (USA) Eppendorf 5415 C (Germany)
Deep freezers	-80°C, Heterofrig CL410HETO (Denmark) -20°C, Bosch (Germany)
Heating Magnetic Stirrers	MR 3001, Heidolph (Germany) Scientifica ARE, VELP (Italy)
Incubator	NÜVE EN500 (Turkey)
Laminar Flow Cabinet	HBB 2460 LaminAir, Holten (Denmark)
Mass Spectrometer	Micromass Q-tof Spectrometer (England)
Orbital Shakers	GFL 702 , GFL (Germany) INNOVA 4340 Illuminated Refrigerated Incubator Shaker, New Brunswick Scientific (USA)
Refrigerators	+4°C Ariston (Italy) +4°C Arçelik (Turkey)
Spectrophotometer	DU 640 Beckman (USA)
Water Purification Systems	Millipore, Milli Ro Plus (USA) Millipore, Milli-Q UF Plus (USA)

3.2. Experimental Methods

3.2.1. Sterilization

Sterilization was carried out by using steam. The growth media used in this study were sterilized in an autoclave. Autoclave sterilization has the advantages of short cycle time and good penetration. The media were prepared in Erlenmeyer flasks such that the total volume of the medium was always one fifth of the flask volume. The flask was sealed with cotton stopper and wrapped with aluminum foil. The media were sterilized by autoclaving at 15 psig and 121 °C for 15 min. Glucose was autoclaved separately from the salts and vitamin solution which was filter sterilized.

3.2.2. Preparation of Agar Plates and Preculture

YPG medium (2% agar, 2% yeast extract, 2% glucose, 1% peptone) was sterilized in an Erlenmeyer flask and transferred to sterile plates. They were kept at room temperature until the medium becomes stiff. Cultures from frozen stocks were spread on YPG agar plates by using inoculating needle. The needle was flamed, sterilized and cooled on the agar plate. The plates were kept at 30 °C for 24-36 h until the colonies appeared. Then master plates were kept at 4 °C until needed .

A single colony from YPG plates were seeded into 30 mL YPG medium (2% yeast extract, 2% glucose, 1% peptone) which was sterilized in an Erlenmeyer flask. Then the preculture was incubated in GFL 702 shaker overnight. The shaker was set to 30 °C and the rate of agitation was set to 180 rpm.

3.2.3. Batch Experiments

Innova 4340 shaker which provides agitation and constant temperature conditions was used for batch experiments. The temperature is set to 30° C and the rate of agitation was set to

180 rpm to obtain growth conditions. F1 defined medium containing 20 g L⁻¹ glucose was used in experiments.

3.2.4. Sampling Procedure

Samples were taken at regular time intervals during the batch cultivations. For determination of extracellular metabolites 1 ml of samples were transferred to eppendorf tubes and centrifuged at 6000 rpm for 5 minutes using an Eppendorf 5154c centrifuge. Thereafter the supernatants were taken and stored at -20° C till MS analysis . Supernatants were kept at -80° C for long term storage.

The method applied for intracellular metabolite extraction was adapted from Gonzales *et al.* (1997) and Castrillo *et al.* (2003). 5 ml cell cultures were quenched in 20 ml 60 per cent methanol, 70mM HEPES, 7.5 pH solution at -40° C. The mixture was centrifuged at 12000 rpm for 3 minutes. Metabolites from the pellet were extracted by adding 5 ml of boiling ethanol (75 per cent), and 0.25M HEPES solution at pH of 7.5. The samples were analysed for metabolome profile using MS.

3.2.4.1. Stability of Metabolites During the Extraction Procedure. A critical requirement when using ethanol as an extracting agent was to ascertain the lack of chemical interference of this solvent with the stability of metabolites during extraction. This problem was first investigated with solutions of commercial metabolites which were prepared in mineral medium pH 5.0 (Verduyn *et al.*, 1992), in order to be as close as possible to the extraction of metabolites from whole yeast cultures. Certain metabolites such as pyruvate, NAD⁺ and NADH were relatively unstable during extraction and a destruction rate ranging from 1.5 to 10% per min could be calculated. To circumvent this serious problem, the ethanol solution was buffered around pH 7.0. It was found that the loss of metabolites was reduced to less than 1% per min when Hepes buffer, pH 7.5, was added at a final concentration of 70 mM to the solution of boiling ethanol. Hence, we can assume that other intracellular metabolites which have not been tested here, would also be stable in boiling buffered ethanol.

3.2.5. Mass Spectrometry and Sample Preparation

Before mass spectral analysis, extracellular and intracellular samples are diluted 10-fold in 50:50:1 (vol/vol/vol) acetonitrile/water/formic acid. Formic acid is added to a final concentration of 0.1% (27 mM) to aid in the ionisation of amino acids. Mass spectra are obtained using a Micromass (Marmara University Faculty of Engineering, İstanbul) LCT electrospray ionization time-of-flight mass spectrometer. The percent intensities between 100-1000 m/z were recorded and preprocessed prior to data analysis. The optimized parameters from Castrillo *et al.*(1997) were used.

ESI/QTOF mass spectra were acquired under the following conditions: 0.09 psi nebulizer gas pressure, 1.43 kV capillary voltage, cone temperature 60 °C, 3 μ l air gap, 1 s aspiration delay, 2 s voltage delay, 3 s equalization delay, and 2 min mass spectra acquisition from 50–1500 amu with a scan rate of 1 amu/ms and a mass resolution of $R = 6000$ FWHM.

3.3. Computational Work: Multivariate Data Analysis

3.3.1. Multivariate Data Analysis

Before any multivariate analysis is carried out, each mass spectral vector is normalized to the total ion count for that vector, so that different spectra can be compared quantitatively. Once a set of N spectra (with mass range p) is concatenated into a single matrix (N objects $\times$ p variables), each column of the data set can be optionally normalized to unit vector norm. The vector norm is a useful column (sample) normalization, where the length (intensity or power) of each sample is set to one, and hence the samples might be more comparable. By using the vector norm it is assumed that only the ratios between m/z values are relevant and not their intensity. Thus vector norm is closely related to correlation analysis, highly correlated samples will get a small distance, and hence they are projected close to each other in PCA and ICA.

In the present study, the l_2 vector norm is implemented, also known as Euclidean norm. Each sample (column) vector $x = (x_1, x_2, \dots, x_d)$ is divided by its norm vector which is equal

$$\|x\| = \sqrt{\sum_i |x_i|^2}, \text{ such that } x_{\text{normed}} = \frac{x}{\|x\|}, \text{ d is the number of m/z variable.}$$

This is done to eliminate bias, in subsequent analysis, toward any column that contains either large absolute values or large variances, notably, however, normalization can sometimes be more detrimental than helpful. If there are a large number of redundant variables in the data, the noise affecting such variables is amplified to the same importance as relevant variables, and this can easily cloud any underlying statistical trends. Consequently, in this instance scaling can not be done before statistical analysis. To reduce the dimensionality of the mass spectral data, PCA will be used in the first stage of analysis.

Second stage involves ICA analysis. ICA (independent component analysis) was applied to MS data after reducing the high dimension of the MS data set, by using PCA. The number of principal components determines the quality of ICA significantly, therefore a criterion was proposed for estimating the optimal dimension automatically. The kurtosis measure is used to order the extracted components. As ICA directly applied to high dimensional data gives often bad results, so the dimension was reduced by PCA at first, before applying ICA. The extracted independent components are sorted by the measure of kurtosis, such that the component IC 1 will be the component with the most negative kurtosis. It is assumed that the components with the most negative kurtosis are the most interesting components in biological experiments. In the last step ICs are processed by HCA (Hierarchical Cluster Analysis) to obtain a clustering algorithm of the analysed samples.

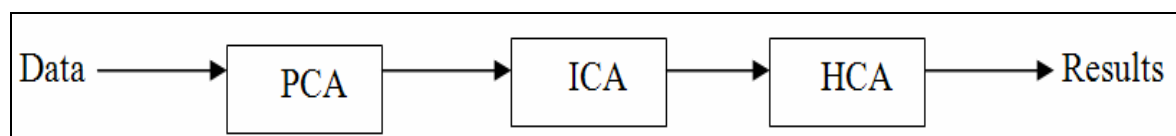


Figure 3.1. Procedure for multivariate data analysis

In the present work, MetaGeneAlyse which is a web-based service allowing the integrated analysis of datasets containing gene-expression data or metabolite data is used for

PCA and ICA. Tab-separated datasets containing genes, metabolites or m/z values in the rows and the experiments in the columns can be uploaded using an online form. Several normalization algorithms are implemented. Principal component analysis (PCA) and independent component analysis (ICA) are then applied to the data set (Dysvik *et al.*, 2001 and Sturn *et al.*, 2002).

3.3.2. Normalisation

Normalisation has a significant impact on the performance of the final analysis (Figure 3.2).

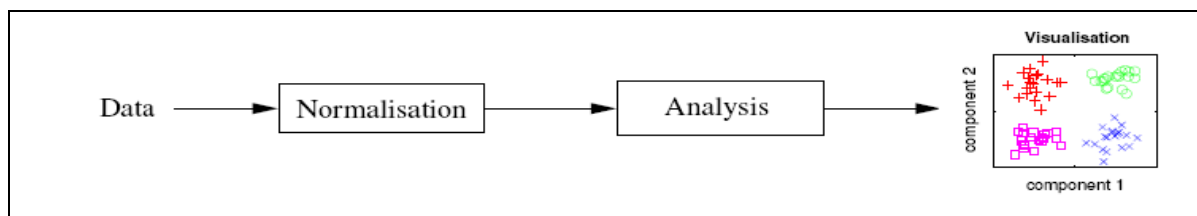


Figure 3.2. Normalisation and analysis techniques

The choice of normalisation is in general as crucial as the choice of the analysis technique itself. The final analysis technique is usually chosen to achieve a specific purpose of analysis with respect to the information one is searching for. The optimal normalisation, by contrast, depends more on the characteristics of the data. In general, due to technical and biological reasons, different variables (e.g., different genes) or different samples are not on the same scale and have to be rescaled in a suitable way. Normalisation, in principle, means to make different variables and different samples comparable to integrate them in a joint analysis.

Another important aspect of normalisation is to represent the relevant information in an optimal way and to remove non-biological contributions. Normalisation is a convenient way to include prior knowledge such as additional technical or biological knowledge and to take into account assumptions about the characteristics of the data.

Rescaling the data is identical to using a new metric in the data space. The aim is to find a metric that optimally represents the desired information. This is closely related to finding the optimal distance or similarity measure as used in cluster or network analysis. It is to some extent also related to finding the optimal kernel in a kernel technique such as support vector machine (SVM) (Vapnik, 1995), where a kernel can be interpreted as a similarity measure. Therefore, a good normalisation or metric can reduce the effort in subsequent analysis techniques.

Consider a set of n experimental samples, each characterised by d measurements (one for each variable), e.g., d different genes. The data can be arranged in a $d \times n$ matrix where rows represent variables (e.g., genes) and columns correspond to different samples. The data matrix can be normalised row-wise or column-wise, such that either the variables or the samples are normalised to make them more comparable. The variables can be rescaled or the total intensity amount of sample vectors can be set constant. One advantage of a sample normalisation is that in case of additional samples the new samples can be normalised individually, no renormalisation of all samples is required, which is important. Variable normalisation, however, is important when the exact contribution of each variable (gene) to a component or visualisation result is of interest. Both Log fold change as variable and unit vector norm as sample normalisation are convenient normalisation techniques for molecular data. Nevertheless, it is often useful to preselect the usually very large number of variables in advance, e.g., by variance or intensity. Although a small intensity or variance might have a large biological impact, small observed values are usually strongly corrupted by the relatively large amount of background noise and, therefore, are normally of no use.

The mass spectrometric methods produce vast amounts of potentially useful data. Direct-injection Mass Spectrometer produces a spectrochromatogram (an array of the mass spectra versus time) for each sample analyzed that can typically hold 10^6 values (depending upon the MS range and sampling rates). In their native form, such data are extremely difficult to interpret. To convert the data into information of chemical or biological interest, some sort of statistical analysis must be used. To simplify any subsequent statistical analysis, two simple processing algorithms are applied to the direct-injection MS spectrochromatograms. Each MS

array could be reduced into a single ‘aggregate’ MS vector by summing the ion counts of a given m/z ratio over the total scan cycle. Then each MS vector will be ‘binned’ to unit m/z ratio (that is, ion counts of fractional m/z ratios are added to the nearest integer m/z). In the **binning process**, the spectrum data set from all gene deletion experiments were binned from 100 to 1000 m/z values (eg. 100, 100.2, 100.4,.....999.8, 1000) .

As an alternative approach of data preprocessing, filtering is applied to MS data. For the **filtering process**, a number of mass values ($n = 1000$) of highest density (highest peaks of the density function) were calculated. These 1000 masses were substituted as variables. After removing variables which are not common for all the deletion mutants, approximately, 150 out of 1000 common variables were left in each group. Finally, the data evaluated by both of the algorithms were processed using Metagenealyse software.

3.3.3. PCA-Principal Component Analysis

PCA involves projecting the original $\mathbf{X}$ -matrix (N objects; p variables) onto a d -dimensional subspace using a projection (or 'loading') matrix, thus creating object coordinates (a 'scores' matrix) in a new coordinate system. This is achieved by the method known as singular value decomposition (SVD) of $\mathbf{X}$:

$$\mathbf{X}_{N \times p} = \mathbf{U}_{N \times d} \mathbf{\Lambda}_{d \times d} \mathbf{L}_{p \times d}^T = \mathbf{T}_{N \times d} \mathbf{L}_{p \times d}^T \quad (3.1)$$

where $\mathbf{U}$ is the unweighted (normalized) score matrix and $\mathbf{T}$ is the weighted (or biased) score matrix. $\mathbf{L}$ is the loading matrix where the columns of $\mathbf{L}$ are known as eigenvectors or loading principal components. $\mathbf{\Lambda}$ is a diagonal matrix (that is, all of the off-diagonal elements are equal to 0) containing the square roots of the first d eigenvalues of the covarianced matrix ($\mathbf{X}^T \mathbf{X}$), where $d < N$ and $d < p$.

The principal components can be considered as a basis set used to project the original data matrix, $\mathbf{X}$, onto the scores, $\mathbf{T}$. In other words, the new coordinates are linear combinations

of the original variables. For example, the elements of the first principal component can be represented as:

$$\begin{aligned} t_{11} &= x_{11}l_{11} + x_{12}l_{12} + \dots + x_{1p}l_{1p} \\ t_{21} &= x_{21}l_{11} + x_{22}l_{12} + \dots + x_{2p}l_{1p} \end{aligned} \quad (3.2)$$

The influence of each of the original variables on the new principal components (that is, the contents of the loading matrix) is determined on the basis of the maximum variance criterion. The first principal component is considered to lie in the direction describing maximum variance in the original data. Each subsequent principal component lies in an orthogonal direction of maximum variance that has not been considered by the former components. The number of principal components computed for a given data set is up to the analyst. However, usually as many principal components are calculated as are needed to explain a preset percentage of the total variance in the original data (the number of principal components is always less than or equal to the number of original variables).

By applying PCA for visualization it is assumed that the most interesting information is directly related to the highest variance in the data. The best projection or visualization is then given by the first two principal components (PCs) of highest variance. This assumption often does not hold for the other components (PC3, PC4, etc.) which may give interesting projections. It is still assumed that the interesting information is related to a significantly high amount of variance but not to the highest amount. PCA can still be used to reduce the high dimension of the data to a relatively low dimension, which covers all relevant variances. Such a pre-processing step should preserve all of the relevant information and reduce only the noise given by small variances.

3.3.4. ICA-Independent Component Analysis

Once the relevant variances are discovered by PCA, on this lower dimensional dataset, a technique can be applied which optimizes other criteria than the variance. A promising technique is ICA. Similar to PCA, ICA gives a set of components. In contrast to PCA these

components are constructed such as to minimize the dependence and are therefore termed independent components (ICs). Independence is a stronger condition than uncorrelation in PCA. This allows detection of more meaningful components than by PCA. These components are not restricted to be orthogonal in ICA.

To achieve ICs, different criteria (contrast functions) can be optimized: higher order dependencies, entropy or kurtosis. In this study, ICA was performed by MetaGeneAlyse software (Hyvärinen and Oja, 2000).

3.3.4.1. Significant Components – Kurtosis. ICA is able to extract as many components as the dataset has dimensions. These components have no order. A criterion was defined for sorting these components. One measurement which can match our interest very well, is the kurtosis.

Kurtosis is a classical measure of non-Gaussianity, and is computationally and theoretically, relatively simple. It indicates whether the data are peaked or flat relative to a Gaussian (normal) distribution. A Gaussian distribution has a kurtosis of zero. Positive kurtosis indicates a ‘peaked’ distribution (super-Gaussian) and negative kurtosis indicates a ‘flat’ distribution (sub-Gaussian).

$$kurtosis(z) = \frac{\sum_{i=1}^n (z_i - \mu)^4}{(n-1)\sigma^4} - 3 \quad (3.3)$$

where $z = (z_1, z_2, \dots, z_n)$ is representing a variable or component with mean μ and standard deviation σ , n is the number of samples. The kurtosis is the fourth auto-cumulant after mean (first), variance (second) and skewness (third).

From purely Gaussian distributed data no unique ICs can be extracted, therefore, ICA should only be applied to datasets where the components that have a non-Gaussian distribution. Examples of super-Gaussian distributions (highly positive kurtosis) are speech signals, because these are predominantly close to zero. However, for metabolite data sub-

Gaussian distributions (negative kurtosis) are more interesting. Negative kurtosis can indicate a cluster structure or at least an uniformly distributed factor. The former can resolve between two experimental conditions (high and low concentrations of metabolites), whereas the latter can represent a continuously changed experimental factor such as the temperature or the light intensity. Thus the components with most negative kurtosis could give the most relevant information.

3.3.5. HCA-Hierarchical Cluster Analysis

Hierarchical cluster analysis (HCA) is a method of grouping samples in a data set by their similarity. HCA involves a progressive pair-wise grouping of samples by distance. Several distance measures can be used in HCA, such as Euclidean distance, Manhattan distance, or correlation. Results vary according to which distance metric is used. The result of hierarchical clustering is usually visualized as a dendrogram or a tree. Branch lengths can be made proportional to the distances between groups. This can provide an easy visualization of the similarities of samples within data sets (Sumner *et al.*, 2003).

3.3.6. GO Analysis

The Gene Ontology (GO) project was established to provide a common language to describe aspects of a gene product's biology. The use of a consistent vocabulary allows genes from different species to be compared based on their GO annotations. The Gene Ontology (GO) project started as a collaboration between three model organism databases, the *Saccharomyces* Genome Database (SGD), FlyBase (for *Drosophila*), and Mouse Genome Informatics (MGI). The GO Consortium has expanded considerably to include many additional model organism databases and annotation groups, each of which contributes to the development of the ontologies, generation of GO annotation files, or development of software tools to utilize GO depending on the nature of its affiliation.

The objective of GO is to provide controlled vocabularies for the description of the molecular function, biological process, and cellular component of gene products. The name

and definition for each GO term and the parent-child relationships between terms are defined by the members of the GO Consortium. This combination of a controlled vocabulary of defined terms with a structure of relationships between items is referred to as an ontology.

4. RESULTS

Batch fermentation experiments were carried out using 14 different deletion mutants, namely *hoΔ*, *snf3Δ*, *rgt1Δ*, *hxt10Δ*, *hxt12Δ*, *gal4Δ*, *mig1Δ*, *gal6Δ gal7Δ*, *gal80Δ*, *adh6Δ*, *mba1Δ*, *hap4Δ*, *yel057cΔ*. *HoΔ* mutant was used as the reference strain in the experiments. *Snf3* is a plasma membrane glucose sensor that regulates glucose transport. *Rgt1* is a glucose-responsive transcription factor that regulates expression of several glucose transporter (*HXT*) genes in response to glucose. *Hxt10* and *hxt12* are putative hexose transporters, expressed at low levels and expression is repressed by glucose. *Mig1* is a transcription factor involved in glucose repression. *Gal4*, *gal6*, *gal7* and *gal80* genes are responsible for the galactose utilization and are members of the biochemical pathway that converts galactose into glucose-6-phosphate. *Adh6* is a NADPH-dependent cinnamyl alcohol dehydrogenase family member with broad substrate specificity; may be involved in fusel alcohol synthesis or in aldehyde tolerance. *Mba1p* is a protein which is involved in assembly of mitochondrial respiratory complexes; may act as a receptor for proteins destined for export from the mitochondrial matrix to the inner membrane. *Hap4* is a transcriptional activator and global regulator of respiratory gene expression. *Yel057c* is a protein of unknown function involved in telomere maintenance.

The effect of these gene deletions on the metabolome profile of yeast cells were investigated using metabolome analysis (metabolic fingerprinting and footprinting) and the bioinformatic tools of PCA, ICA, and HCA.

4.1. Metabolic Footprinting

Yeast single-gene deletants (having defects at different points of their metabolism) were compared in terms of their metabolome profiles. The steady state phase samples taken at 15th, 20th, 25th and 31st hours of fermentation were processed and analysed by Mass Spectrometer as described in Experimental Protocol. The reason for taking the samples during the stationary

state is that the most marked changes are known to occur during the transition phase from exponential to the stationary phase, where the spectra became increasingly complex with the appearance of numerous small peaks (Oliver *et al.*, 1998). This implies that the cells are secreting or excreting metabolites into the culture medium.

Mass spectra of different *S. cerevisiae* deletion strains are given in Figure 4.1. These spectrums show the extracellular metabolome profiles of the 14 deletion strains (17 deletion strains were investigated but only one of the *ho* deletion strains' spectrum out of 4 was shown) of *S. cerevisiae*. The samples were analyzed by direct injection mass spectrometer and the intensities are plotted against the mass-to-charge ratios, m/z . In m/z term, m indicates the mass of ion and z indicates the charge of that ion. The ion which has the highest intensity value is represented at 100%. The intensity values of other ions were changed according to that base peak ion.

To obtain the information on the classification of yeast mutants is the challenge of this metabolome analysis. The global structures of metabolome profiles of deletion mutants are investigated using two different preprocessing approaches. In the first approach, the metabolome data is compiled according to binning procedure. In the binning approach, each mutant has attained 4501 m/z values. In the second approach, the metabolome data in each sampling group (15th hour, 20th hour, 25th and 31st hour) are grouped by applying a filtering procedure. In other words, 1000 highest peaks of the density function belonging to all gene deletion experiments were determined. Then, the highest peaks common in all 17 mutants were used as the reference mass variables. Approximately, 100 out of 1000 common variables were left in each group.

Multivariate data visualization methods, PCA and ICA, are then applied to sort out the spectra obtained from MS analysis. PCA is able to rotate the data points into a new coordinate system such that the majority of the variance in the data is accounted for in the directions of a subset of these rotated axes. By plotting the points in this new coordinate system, the significant effects within the data can be more easily visualised.

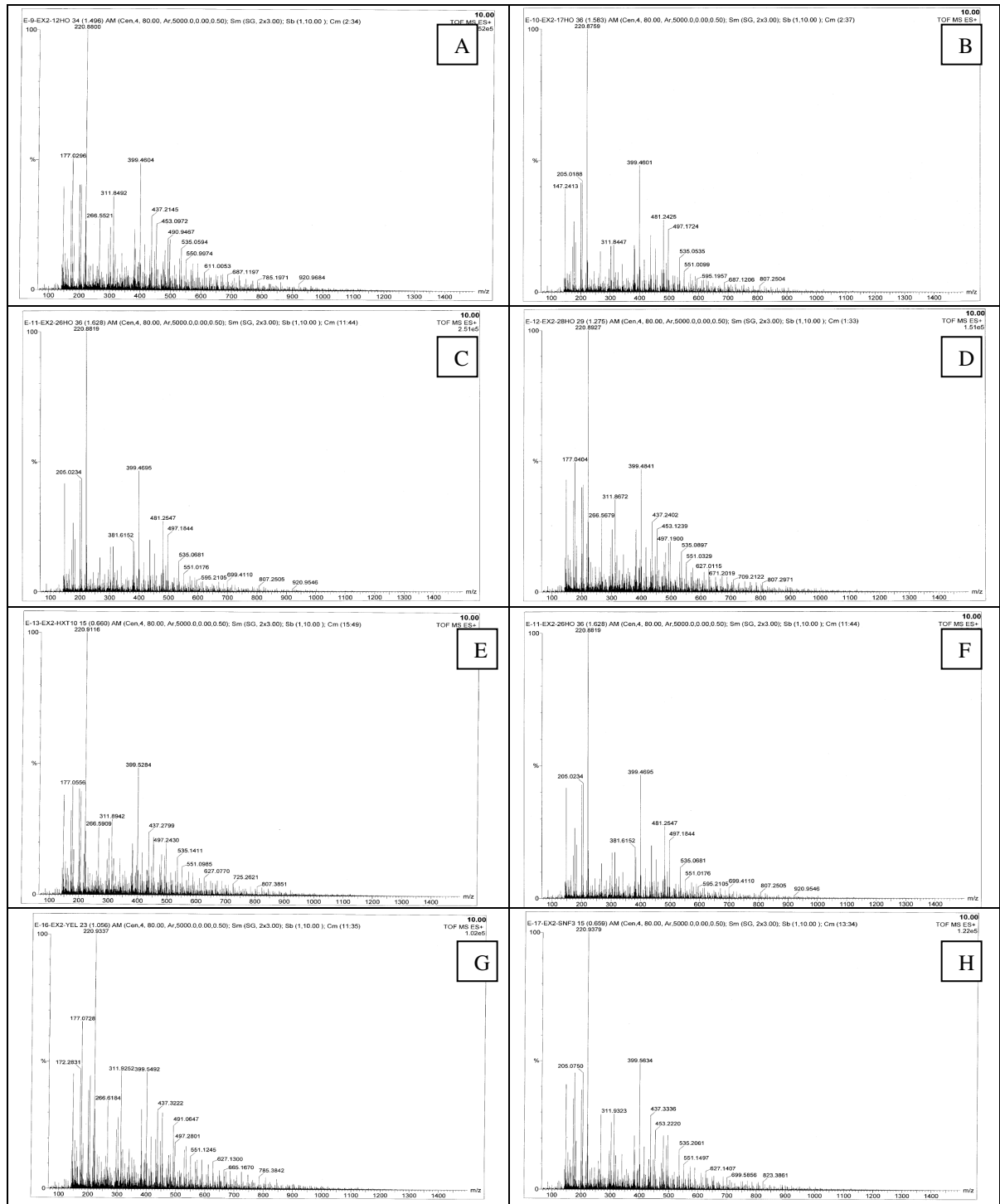


Figure 4.1. Mass spectral data of extracellular metabolites belonging to 17 mutants at 20th hour of fermentation (A: *ho12*, B: *ho17*, C: *ho26*, D: *ho28*, , E: *hxt10*, F: *hxt12*, G: *yel057c*, H: *snf3*)

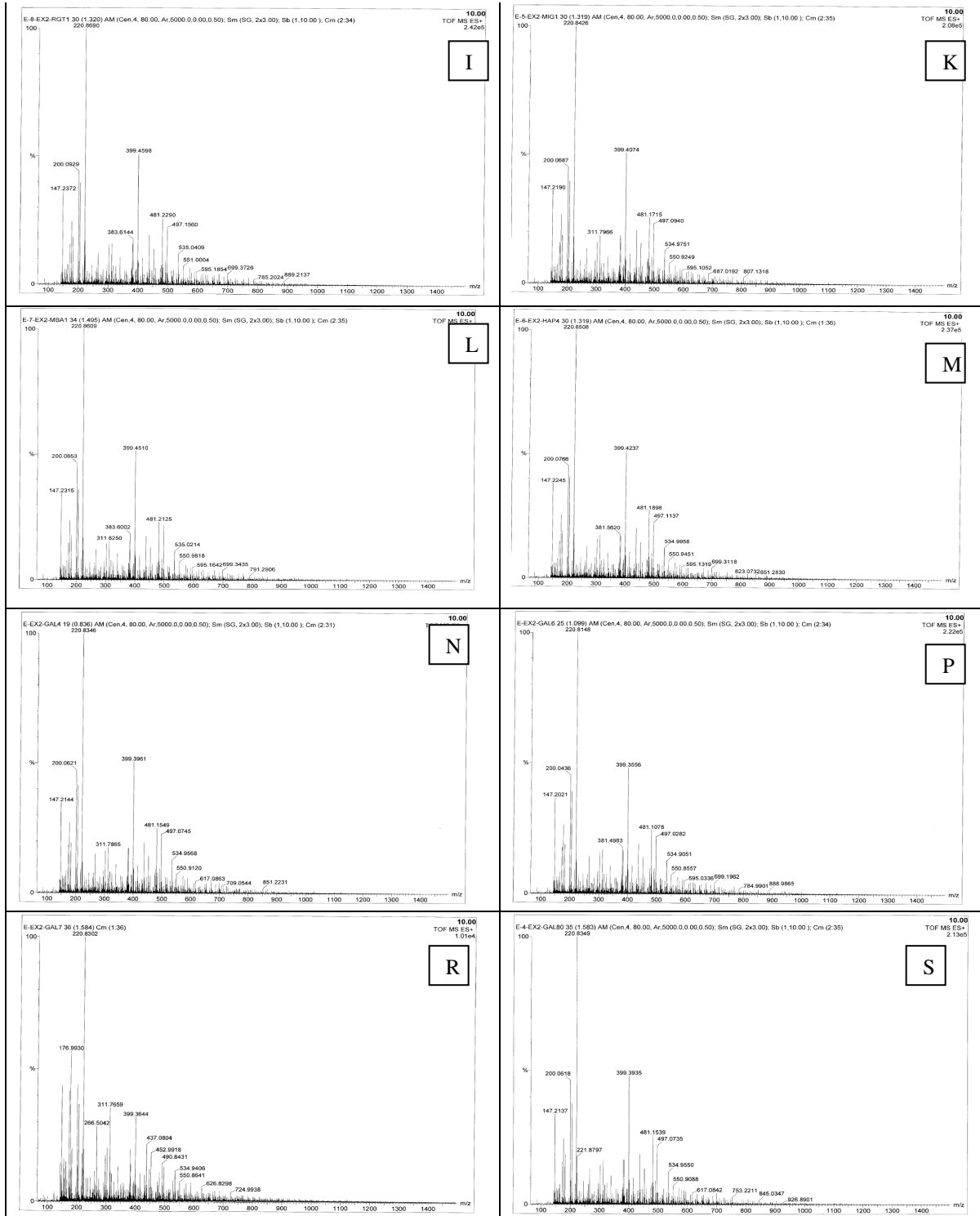


Figure 4.1. continued (I: *rgt1*, K: *mig1*, L: *mba1*, M: *hap4*, N: *gal4*, P: *gal6*, R: *gal7*, S: *gal80*)

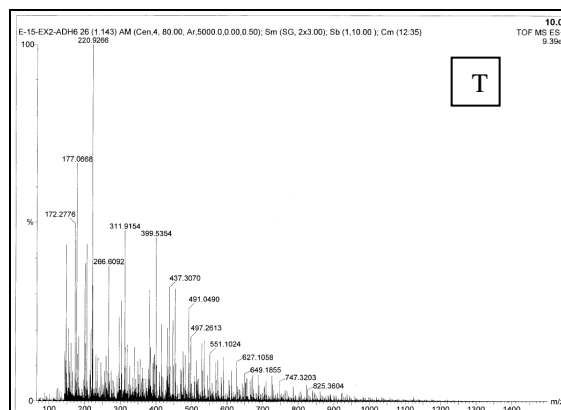


Figure 4.1. continued (T: adh6)

The results of PCA conducted on the preprocessed (binned or filtered) extracellular mass spectral metabolome profiles of *S. cerevisiae* mutants are shown in sections 4.1.1.-4.1.9. The PCA results give an information about the global pattern of deletants or possible clustering but it is not very informative, i.e. it is not possible to discriminate the mutants according to their function by using PCA alone. So, other statistical methods should be used to determine the possible discriminations among the mutants.

Independent component analysis, ICA, was applied as an alternative method in order to reveal some hidden data in PCA results and to confirm the possible clusterings in PCA. Due to its independence condition, ICA seems to be more suitable for the metabolome analyses than PCA. However, ICA has not been developed for a small number of high dimensional samples, therefore a strategy is needed to overcome this limitation.

To apply ICA successfully it is essential to reduce the high dimension of the data set by using PCA. The number of principal components determines the quality of ICA significantly, therefore a criterion should be proposed for estimating the optimal dimension automatically. The kurtosis measure is used to order the extracted components. Negative kurtosis can indicate a cluster structure or at least an uniformly distributed factor. Thus the components with most negative kurtosis could give us the most relevant information. The most negative kurtosis is obtained after pre-processing the PCA data for 5 times (Figure 4.2). The results of ICA are presented in sections 4.1.1 to 4.1.9.

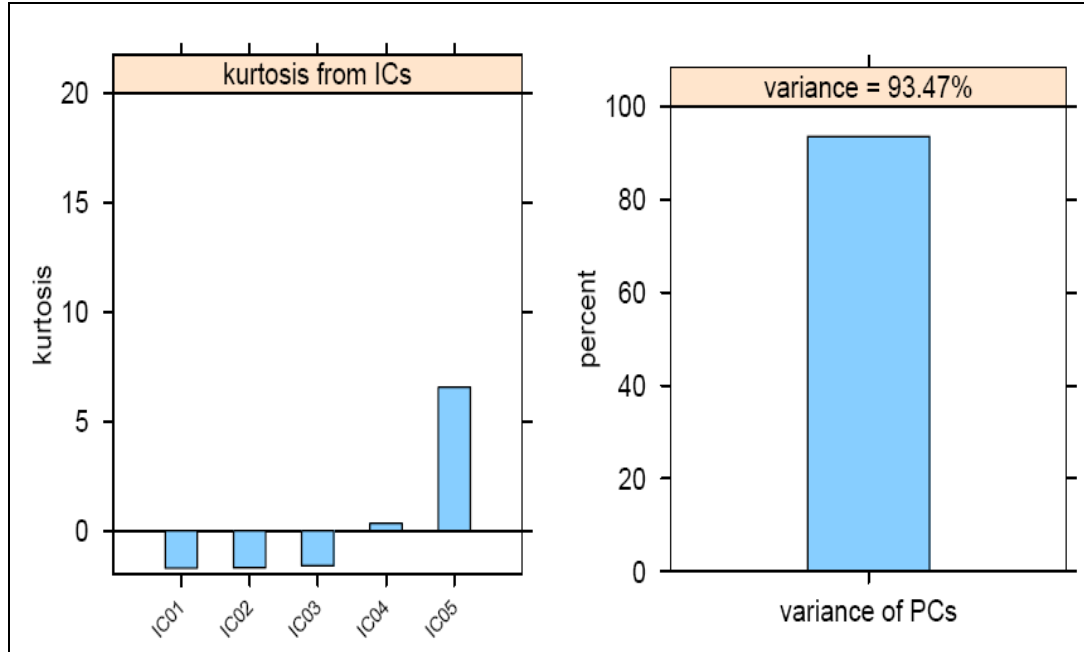


Figure 4.2. Kurtosis measure of ICA data

4.1.1. PCA and ICA of Binned Extracellular Metabolome Data at 15th hour

PCA was applied to normalized metabolome data of 17 deletion mutants. In Figure 4.3 (a); cluster 1 includes *ho12Δ*, *ho17Δ*, *ho26Δ*, *ho28Δ*, *mig1Δ*, *hxt10Δ*, *hxt12Δ*; cluster 2 includes *adh6Δ*, *gal80Δ*, *hap4Δ*; cluster 3 includes *snf3Δ*, *yel057cΔ*, *gal7Δ*; cluster 4 includes *gal4Δ* and *gal6Δ*. In Figure 4.3 (b); cluster 1 includes *snf3Δ*, *yel057cΔ*, *gal7Δ*; cluster 2 includes *ho17Δ*, *ho26Δ*, *ho28Δ*, *hxt10Δ*, *hxt12Δ*, *mig1Δ*; cluster 3 includes *mba1Δ*, *gal80Δ*, *hap4Δ*. In Figure 4.3 (c); cluster 1 includes *gal4Δ*, *gal6Δ*; cluster 2 includes *gal7Δ*, *snf3Δ*, *yel057cΔ*; cluster 3 includes *adh6Δ*, *hap4Δ*, *mba1Δ*, *rgt1Δ*, *gal80Δ*. In Figure 4.3 (d); cluster 1 includes *adh6Δ*, *hap4Δ*, *mba1Δ*, *rgt1Δ*, *gal80Δ*; cluster 2 includes *gal7Δ*, *snf3Δ*, *yel057cΔ*. The deletion mutants *gal4Δ*-*gal6Δ*; and *snf3Δ*-*yel057cΔ*-*gal7Δ* are clustered together in all projections given in Figure 4.3 while other homozygous deletion mutants are scattered randomly. *Gal4Δ* and *gal6Δ* are reported to have transcriptional regulator activity and a *p* value of 0.00046 in GO term analyses confirms the relation between them. However, there is no such correlation that could be stated among the other groups.

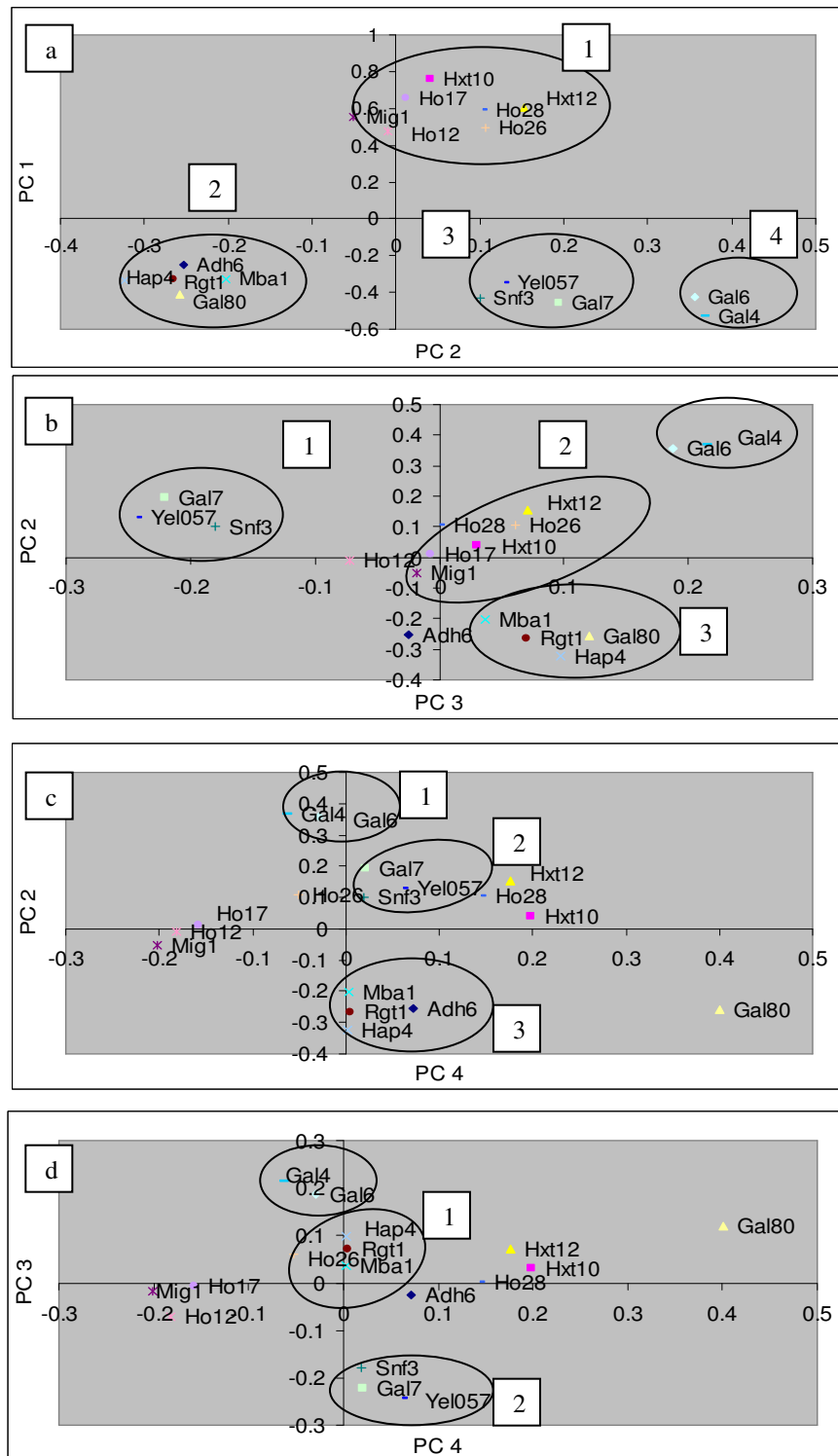


Figure 4.3. Scores of binned metabolome data on first four Latent Variables at 15th hour of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

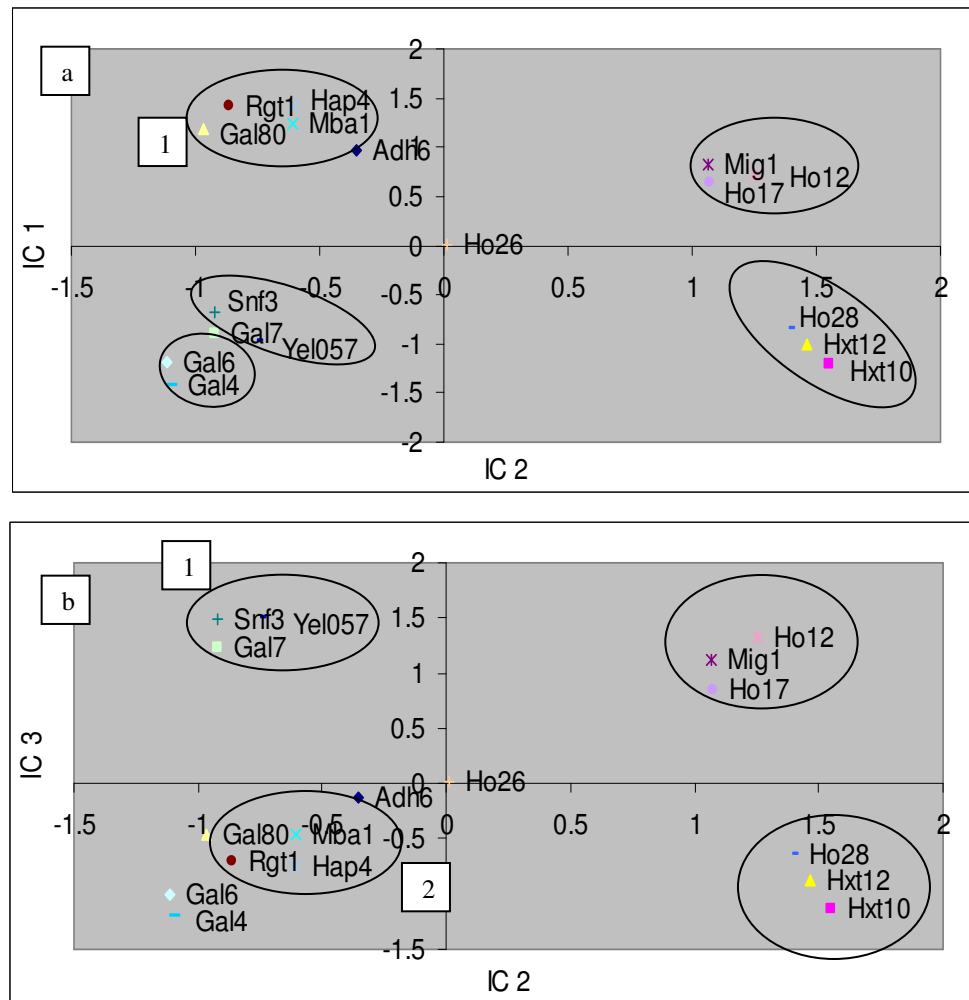


Figure 4.4. Scores of binned metabolome data on first three ICs at 15th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

ICA was then applied to PCA data and the ICs were plotted against each other (Figure 4.4). In Figure 4.4.a; cluster 1 includes *hap4Δ*, *gal80Δ*, *mba1Δ*. In Figure 4.4 (b); cluster 1 includes *gal7Δ*, *snf3Δ*, *yel057cΔ*; cluster 2 includes *hap4Δ*, *mba1Δ*, *rgt1Δ*, *gal80Δ*. The clusters in Figure 4.4 (a) and (b) do not change as in PCA. For example, *gal4Δ-gal6Δ*; *hxt10Δ-hxt12Δ-ho28Δ*; *snf3Δ-gal7Δ-yel057cΔ*; and *mig1Δ-ho12Δ-ho17Δ* groups remain in all components (IC1 vs IC2 and IC3 vs IC4). *HoΔ* reference strain appears to scatter among the groups. It is involved in HXT cluster although *hxt10Δ-hxt12Δ* pair is a member of hexose transport pathway. *HoΔ* deletion mutant also clusters with *mig1Δ* deletion mutant which has a

role in glucose repression. Nevertheless, one should note that IC2 discriminates *hoΔ* and *hxtΔ* mutants from other deletion strains.

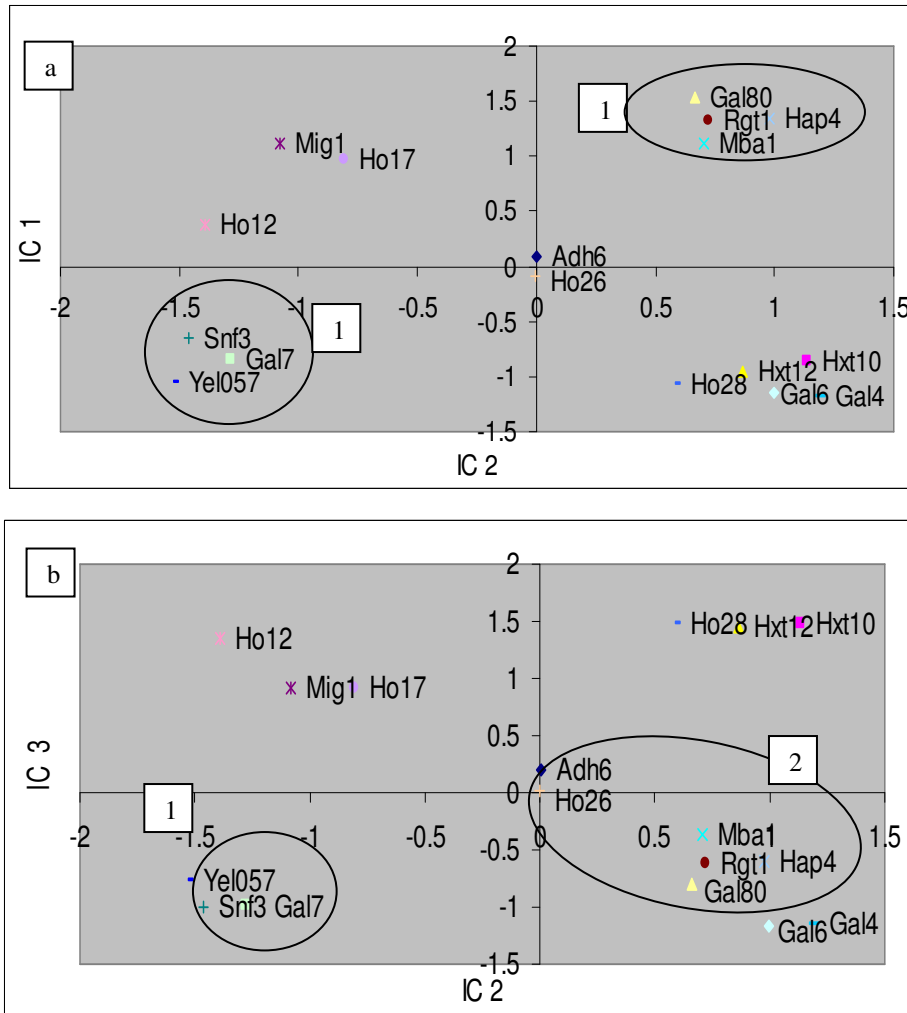


Figure 4.5. Scores of binned metabolome data on the first three ICs at 15th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

When the PCA preprocessing is increased from 5 to 6, the groups of the deletion mutants in Figure 4.5 are nearly the same as in Figure 4.4. PCA preprocessing does not affect the ICs. In Figure 4.5 (a); cluster 1 includes *hap4Δ*, *mba1Δ*, *rgt1Δ*, *gal80Δ*. In Figure 4.5 (b); cluster 1 includes *gal7Δ*, *snf3Δ*, *yel057cΔ*; cluster 2 includes *adh6Δ*, *ho26Δ*, *hap4Δ*, *mba1Δ*, *rgt1Δ*, *gal80Δ*.

4.1.2. PCA and ICA of Filtered Extracellular Metabolome Data at 15th hour

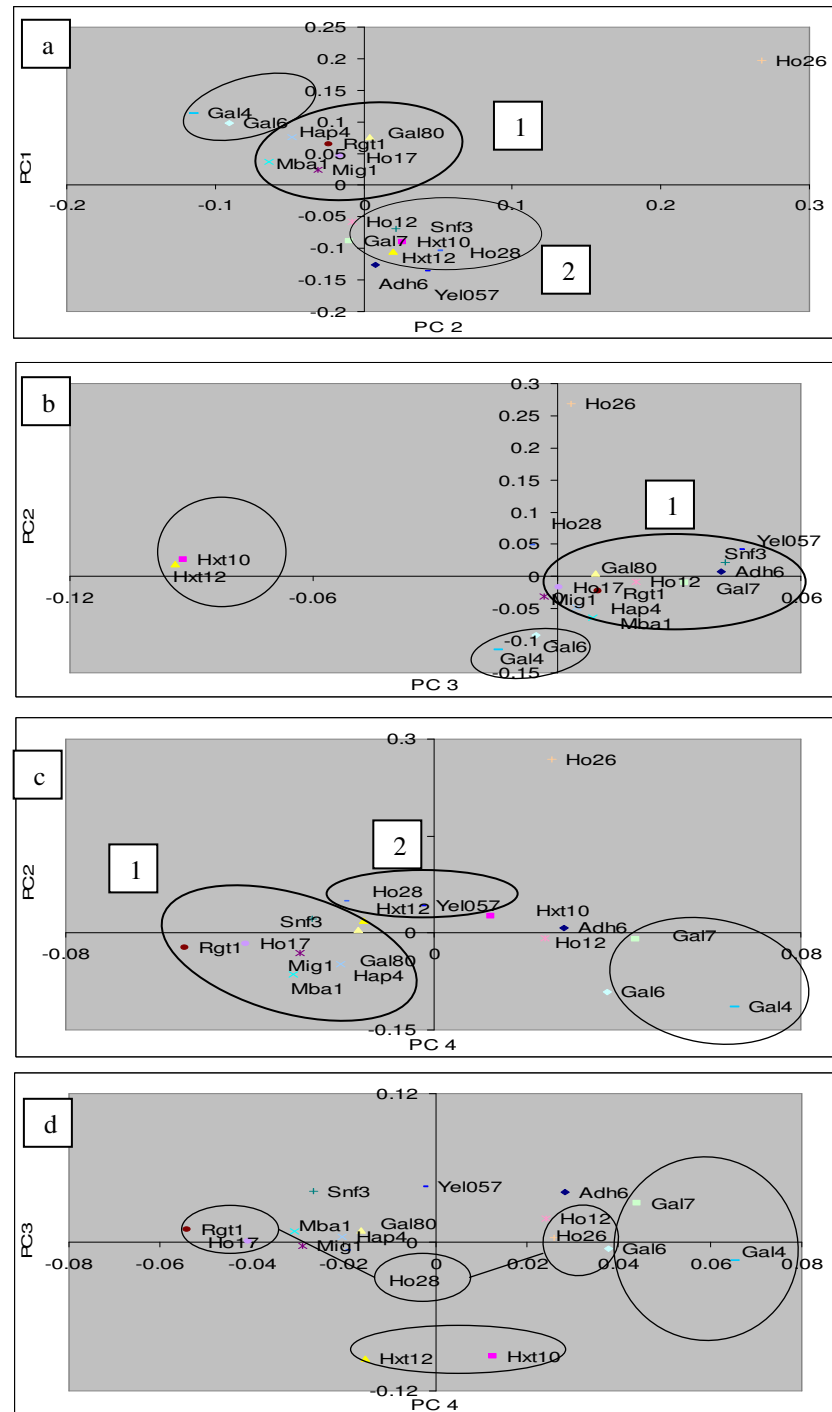


Figure 4.6. Scores of filtered metabolome data on first four LVs at 15th hour of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

In Figure 4.6 (a), cluster 1 includes *gal80Δ*, *hap4Δ*, *mba1Δ*, *ho17Δ*, *mig1Δ*, *rgt1Δ*; cluster 2 includes *gal7Δ*, *snf3Δ*, *ho12Δ*, *ho28Δ*, *hxt10Δ*, *hxt12Δ*. In Figure 4.6 (b); cluster 1 includes *adh6Δ*, *snf3Δ*, *gal7Δ*, *ho12Δ*, *mig1Δ*, *mba1Δ*, *hap4Δ*, *gal80Δ*, *rgt1Δ*. In Figure 4.6 (c); cluster 1 includes *rgt1Δ*, *ho17Δ*, *snf3Δ*, *mba1Δ*, *mig1Δ*, *hap4Δ*, *gal80Δ*; cluster 2 includes *ho28Δ*, *yel057cΔ*, *hxt12Δ*. PC1 distinguishes regulation related genes (cluster 1), and PC2 distinguishes gal genes and hxt genes from the others. Other mutants which are expected to be in the same biological process according to GO term analysis do not appear close to each other.

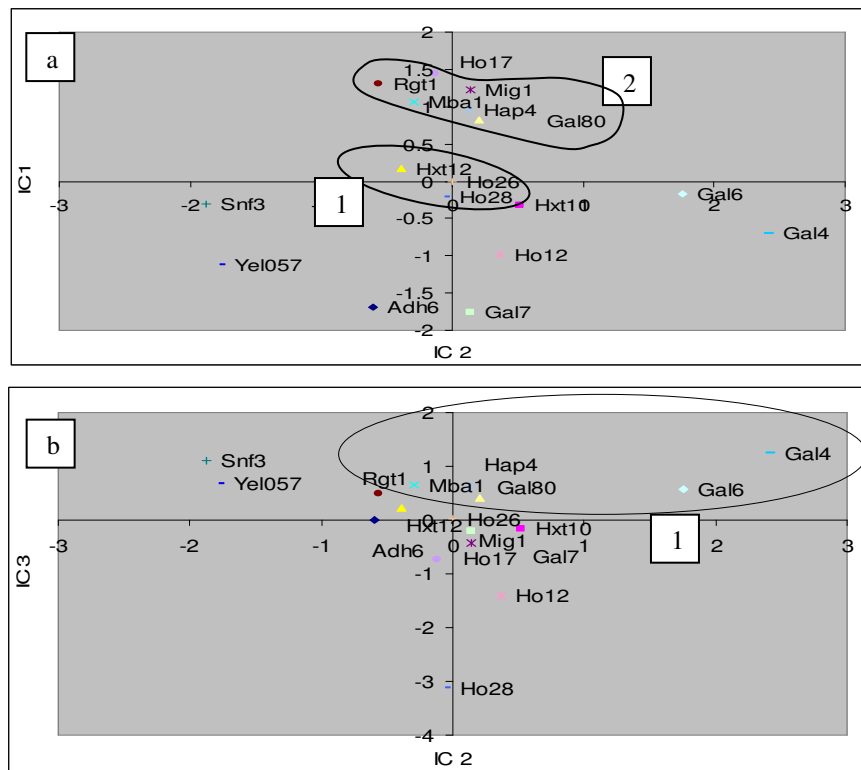


Figure 4.7. Scores of filtered metabolome data on first three ICs at 15th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

ICA of extracellular metabolome data obtained at 15th hour of fermentation was shown in Figure 4.7. The data was compiled by applying 5th pca-preprocessing step in the Metagenealyse software. However, the results were more satisfactory with respect to PCA results shown in Figure 4.6. In Figure 4.7 (a); cluster 1 includes *ho26Δ*, *ho28Δ*, *hxt12Δ* and

cluster 2 includes *gal80Δ*, *hap4Δ*, *mig1Δ*, *rgt1Δ*, *mba1Δ*. However, *hoΔ* reference strains are scattered widely in the first 3 ICs and there is no clear distinction around the origin. Actually, the kurtosis may affect the dispersion and another step was carried out at 6th pca-preprocessing in order to improve the outcomes.

6th step of pre-processing was more promising than the former one (5th). In Figure 4.8 (a); cluster 1 includes *ho26Δ*, *ho28Δ*, *snf3Δ*, *gal6Δ*, *gal4Δ*; cluster 2 includes *adh6Δ*, *gal7Δ*. In Figure 4.8 (b); cluster 1 includes *adh6Δ*, *gal7Δ*, *gal80Δ*, *snf3Δ*, *ho12Δ*, *ho26Δ*, *ho28Δ*, *mba1Δ*, *mig1Δ*, *hap4Δ*, *rgt1Δ*. *Hxt10Δ-hxt12Δ* deletion mutants were extracted from other groups but there was no clustering among the biologically interacting genes such as galactose transporters.

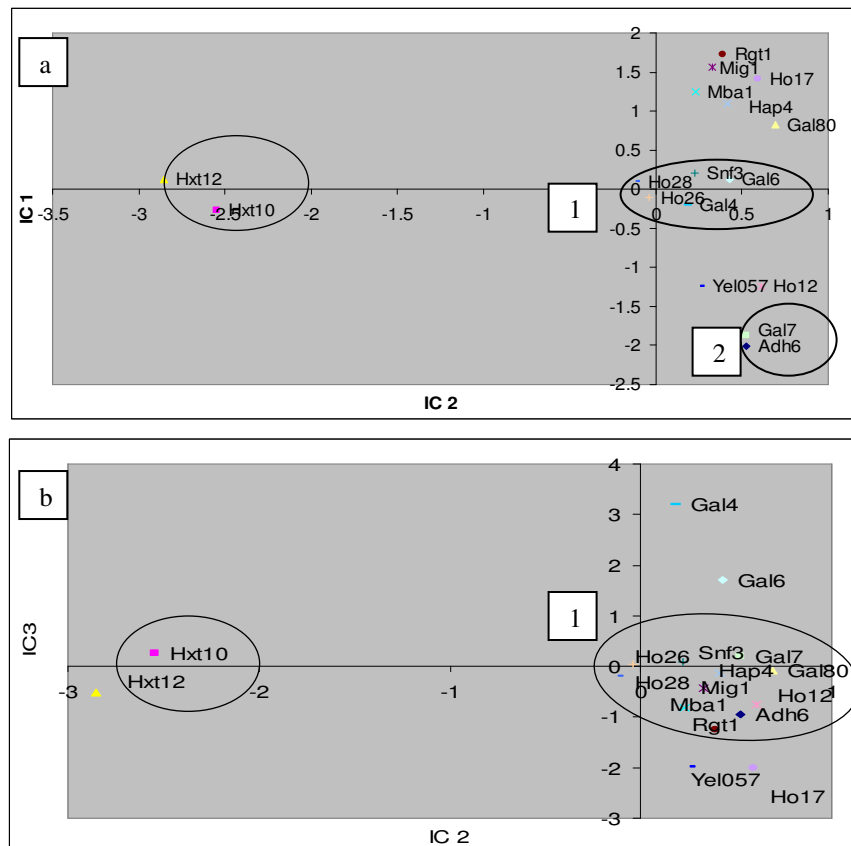


Figure 4.8. Scores of filtered metabolome data on the first three ICs at 15th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

4.1.3. PCA and ICA of Binned Extracellular Metabolome Data at 20th hour

The scores of 20th hour of extracellular metabolome data given in Figure 4.10 (a) show that PC1 gives 4 distinct clusters, namely cluster 1 (*hoΔ* mutants), cluster 2 (*galΔ* mutants), cluster 3 (respiration related mutants, *hap4Δ* and *mba1Δ*), and cluster 4 (transporters; *hxt10Δ*, *hxt12Δ* mutants, *snf3Δ* and *yel057cΔ*). *Mig1Δ* mutant is also close to *Gal80Δ* and *Gal4Δ* mutants as expected, because *Mig1* gene is a transcription factor involved in repression. In Figures 4.10 (a), 4.10 (b) and 4.10 (c) cluster 2 includes *gal4Δ*, *gal80Δ*, *mig1Δ*. In Figure 4.10 (d); *adh6Δ*, *gal4Δ*, *gal80Δ*, *ho26Δ*, *mig1Δ*, *gal6Δ*, *hxt12Δ* deletion mutants are all clustered around the origin.

The genes that have a function in galactose metabolic pathway are shown in Figure 4.9. *Gal4*, *gal6*, and *gal80* genes have a regulatory role controlling whether the pathway is on or off. *Gal4* is a DNA-binding transcription factor required for the activation of the *gal* genes in response to galactose; repressed by Gal80p and activated by Gal3p. *Gal6* is an aminopeptidase of cysteine protease family, has a DNA binding activity and acts as bleomycin hydrolase in vitro; its transcription is regulated by galactose via Gal4p. *Gal80* is a transcriptional regulator involved in the repression of *Gal* genes in the absence of galactose; inhibits transcriptional activation by Gal4p; inhibition relieved by Gal3p or Gal1p binding.

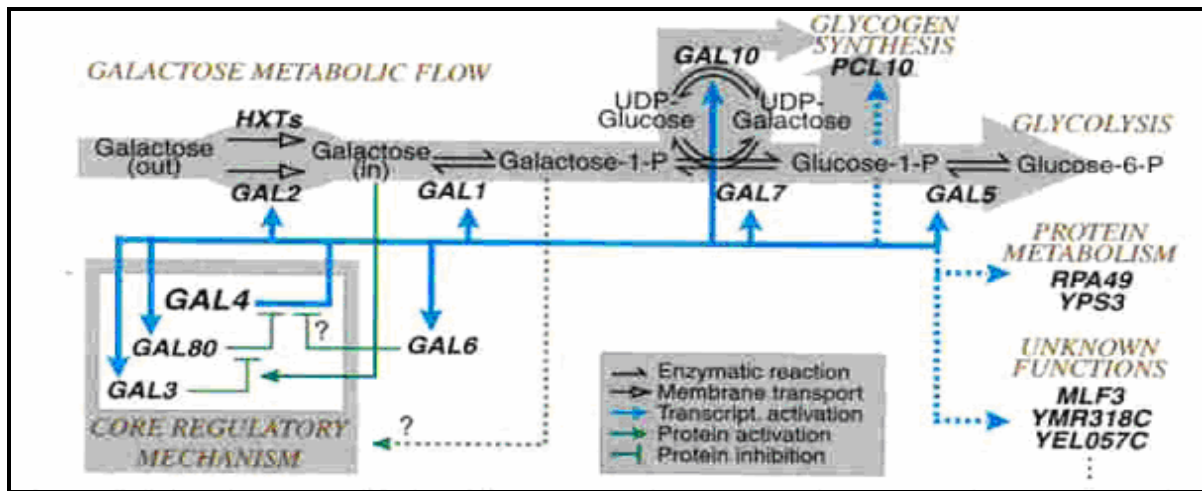


Figure 4.9. Galactose metabolic pathway

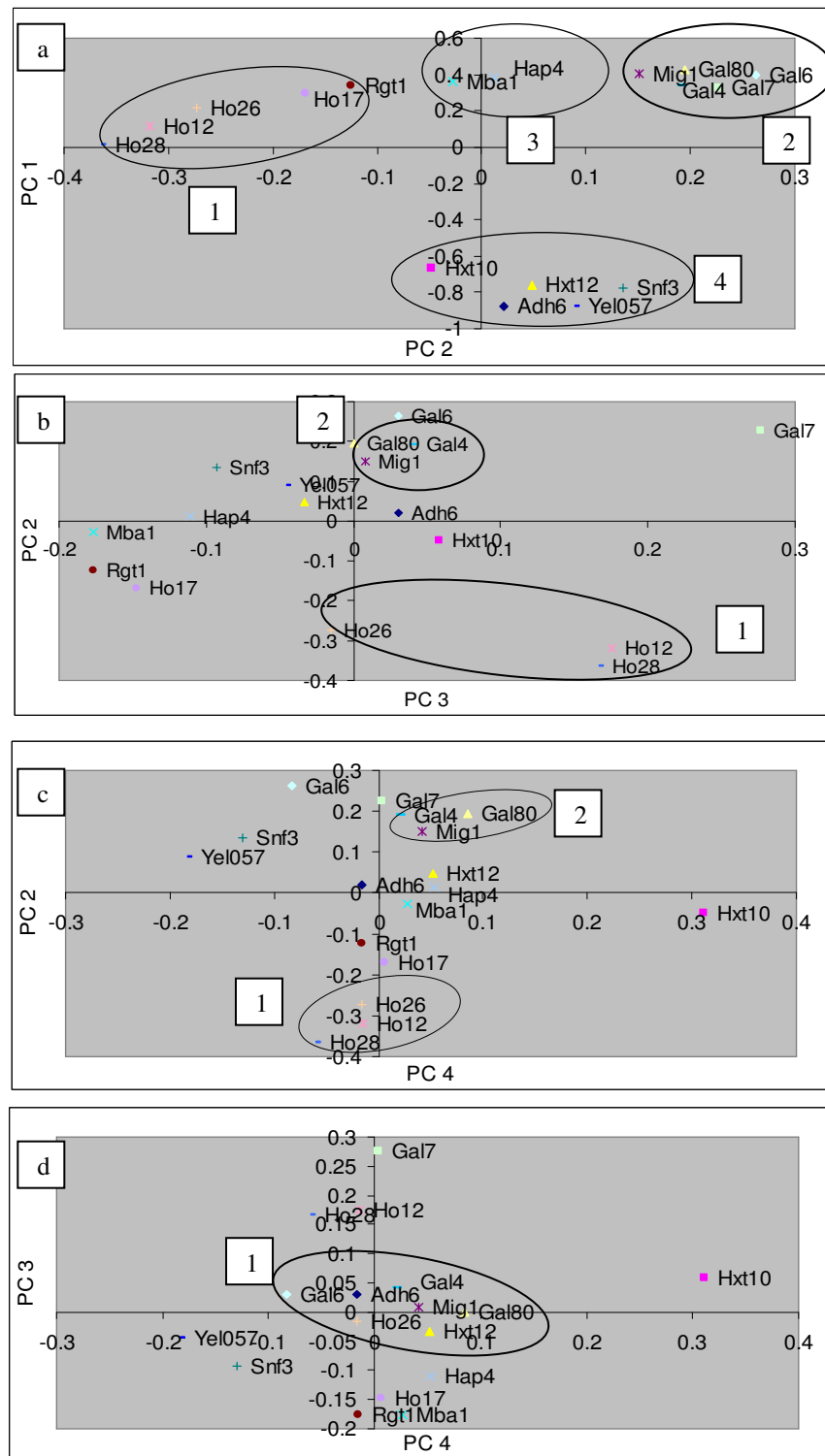


Figure 4.10. Scores of binned metabolome data on first four LVs at 20th hour of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

The results of ICA belonging to batch experiments using the deletion strains were shown in Figure 4.11. *Gal4* Δ , *gal6* Δ and *gal80* Δ mutants were grouped in the right bottom corner. *Gal4*, *gal6*, and *gal80* genes could be responsible in core regulatory mechanism of galactose pathway.

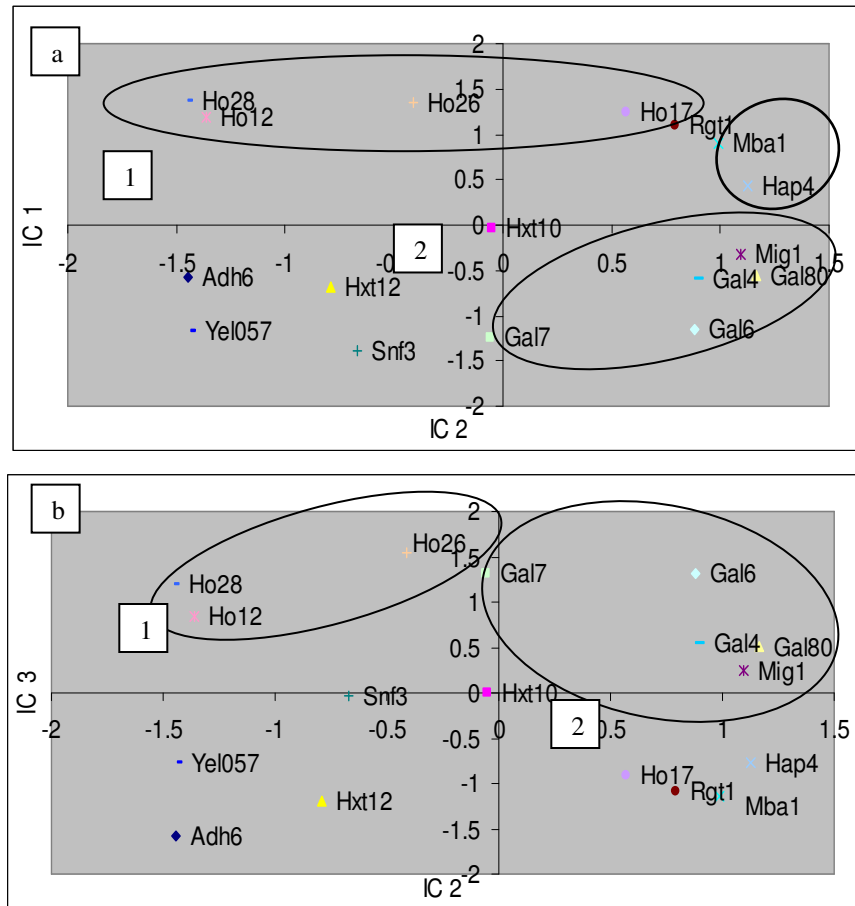


Figure 4.11. Scores of binned metabolome data on first three ICs at 20th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

The GO terms of *gal4*, *gal6*, *gal80* and *mig1* show that there should be a relationship among these genes in terms of the transcriptional regulator activities. The evaluated *p* value is rationally small (0.00037). *Hxt10* Δ and *hxt12* Δ mutants are similarly found in the lower region of Figure 4.11 (a). They have a role in galactose pathway but their function is different from the *gal* genes stated above. These genes take part in galactose transportation. Deletion mutants of *hap4* Δ and *mba1* Δ genes could be seen at the top of Figure 4.11 (a). These two genes have

very different roles compared to *Gal* genes or *HXT* genes, they are related to respiration. The two respiration related *hap4Δ* and *mba1Δ* mutants are clustered in the lower right quadrant of Figure 4.11 (b). In Figure 4.11 (a) and 4.11 (b) ; cluster 1 includes *ho12Δ*, *ho26Δ*, *ho28Δ* and cluster 2 includes *gal4Δ*, *gal6Δ*, *gal7Δ*, *gal80Δ* and *mig1Δ*.

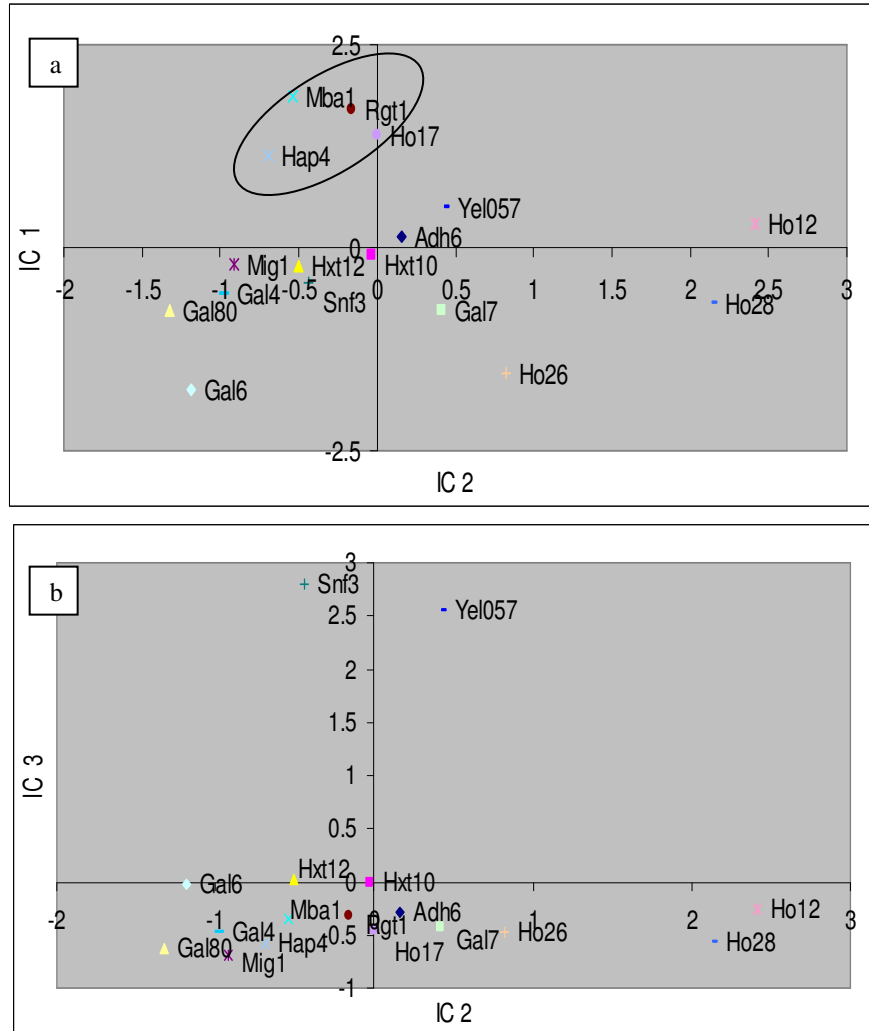


Figure 4.12. Scores of binned metabolome data on first three ICs at 20th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

When the PCA preprocessing is increased from 5 to 6, ICA results shown in Figure 4.12 do not reflect any significance and do not correspond with the former ICA results (Figure 4.11) which are preprocessed by 5 times. The deletion mutants are scattered around the origin and any discrimination could not be obtained.

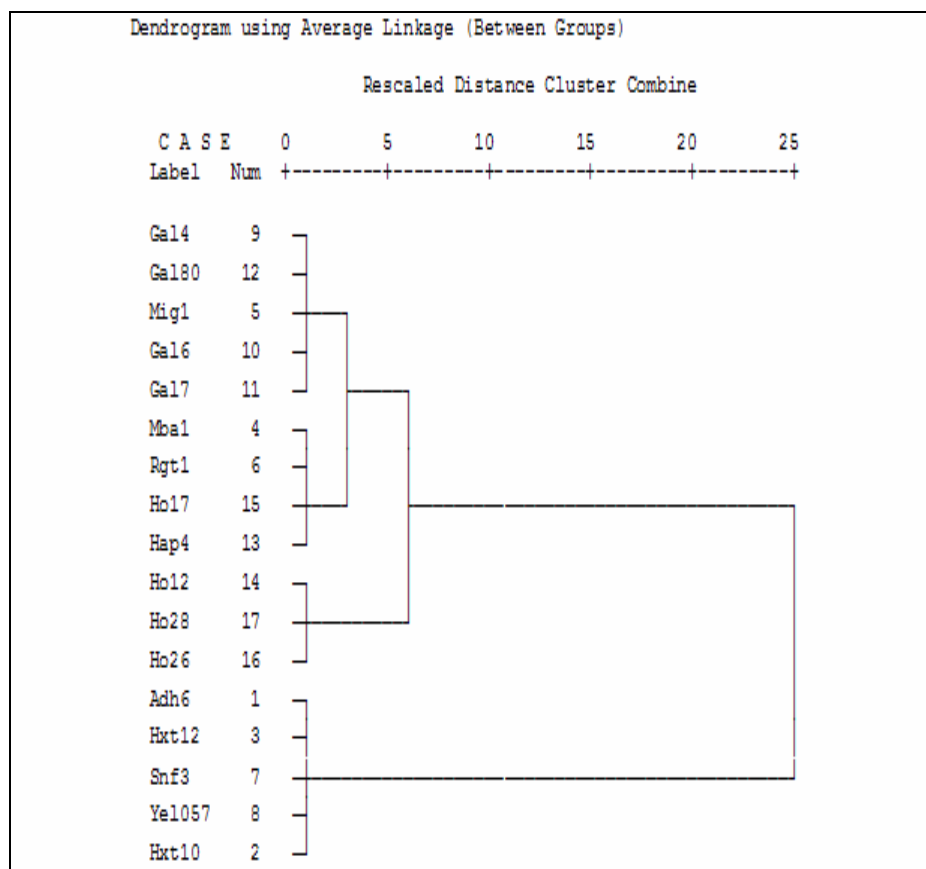


Figure 4.13. HCA of PC1 and PC2

To further investigate the differences in the metabolome data among deletion mutants HCA was carried out (this is also an unsupervised approach). Hierarchical cluster analysis of two PCs of binned extracellular metabolome data at 20th hour is given in Figure 4.13.

In Figure 4.13; the cluster including *gal4Δ-gal80Δ-mig1Δ-gal6Δ-gal7Δ* and the cluster including *adh6Δ-hxt12Δ-snf3Δ-yel057cΔ-hxt10Δ* are farthest apart as expected. It is obvious that HCA represents the PCA clustering correctly, and it can be used as an unsupervised method with PCA or ICA.

The majority of the clustering algorithms undertake an initial calculation to produce a similarity or distance matrix between entities in the data. For the calculation of the similarity matrix, same data which was used for the variance calculations were also used. Various distance (or *dissimilarity*) metrics were investigated (Duda *et al.*, 1973; Jain *et al.*, 1988) and

they all produced similar results. Therefore the most commonly used distance metric, that of Euclidean distance was chosen to illustrate the results. Objects are compared in a pairwise fashion to calculate the distance. The Euclidean distance describes a circular population boundary in two dimensions, and a sphere or hyper-sphere in 3 or more dimensions. If a particular population or cluster does not have a hyper-spherical boundary (it may for example have an elliptical shape along a particular axis) then this distance measure may fail to group objects correctly. Calculation of these distance metrics allows the plotting of a dendrogram, or hierarchical tree, which illustrates how far apart (or *dissimilar*) objects are from each other. Dendrograms may also be employed to cluster the data by terminating the tree hierarchy at a desired threshold. Figure 4.13 shows an example of a dendrogram produced by the calculation of the Euclidean distances.

Remaining HCA results of other metabolome data are given in Appendix A and B. These HCA outcomes verify the PCA and ICA results to some extent.

4.1.4. PCA and ICA of Filtered Extracellular Metabolome Data at 20th hour

The first four latent values of extracellular deletion mutants were shown in Figure 4.14. 116 out of 1000 common mass variables (m/z) were used to carry out PCA and ICA. In Figure 4.14 (a), cluster 1 includes *mba1Δ*, *hap4Δ*; cluster 2 includes *gal4Δ*, *gal6Δ*, *gal80Δ*, *mig1Δ*; cluster 3 includes *ho12Δ*, *ho28Δ*, *hxt10Δ*, *hxt12Δ*; cluster 4 includes *ho17Δ*, *ho26Δ*, *snf3Δ*, *rgt1Δ*. In Figure 4.14 (b), cluster 1 includes *gal4Δ*, *gal80Δ*, *mba1Δ*, *mig1Δ*, *hap4Δ*. In Figure 4.14 (c); cluster 1 includes *gal4Δ*, *gal80Δ*, *mig1Δ*, *hap4Δ*; cluster 2 includes *ho12Δ*, *ho28Δ*, *hxt10Δ*, *hxt12Δ*, *mba1Δ*. In Figure 4.14 (d); cluster 1 includes *gal4Δ*, *gal80Δ*, *mig1Δ*, *hap4Δ*.

The mutants that appeared in clusters 1 and 2 at Figure 4.14 (a) have regulator activity while the mutants in clusters 3 and 4 have roles in glucose signaling and transport. While PC1 distinguishes regulation related genes (upper part of Figure 4.14 (a)); PC2 distinguishes gal genes.

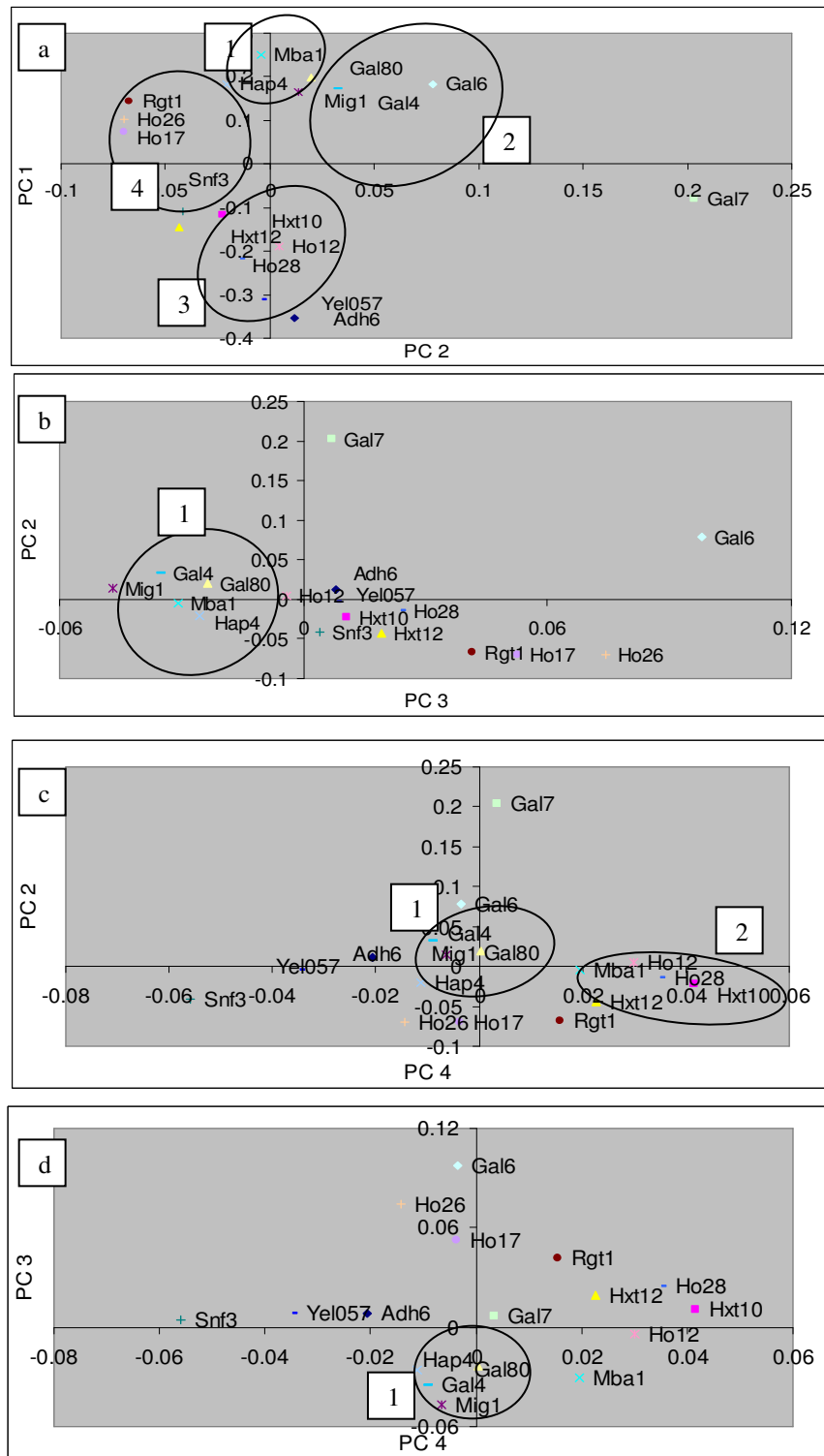


Figure 4.14. Scores of filtered metabolome data on first four LVs at 20th hour of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

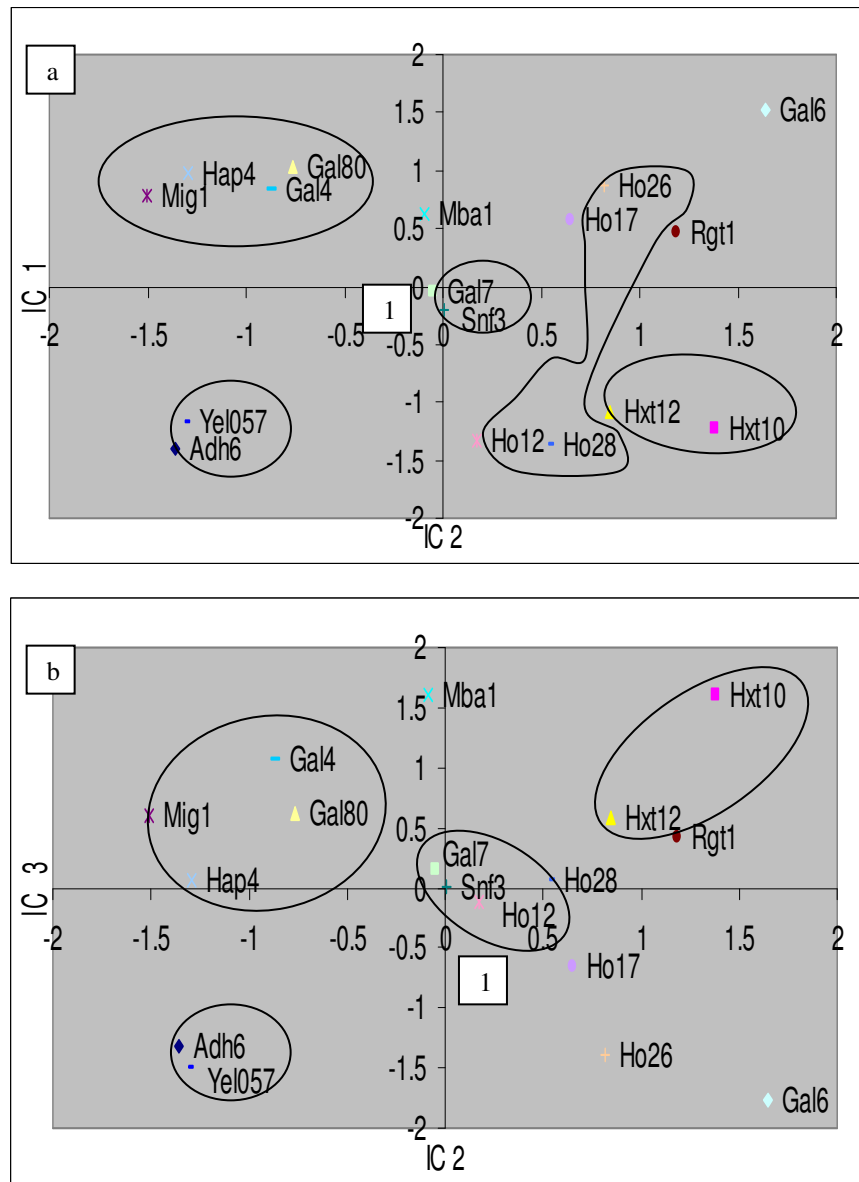


Figure 4.15. Scores of filtered metabolome data on first three ICs at 20th hour of extracellular samples, , a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

There are 3 negative ICs out of 5 after preprocessing PC data 5 times. The distribution of the extracellular metabolome data notably changed compared to Figure 4.14. This may help to distinguish the groups more easily. For example, the clusters of regulator genes (*gal4Δ-gal80Δ-mig1Δ-hap4Δ*), and hexose transporter genes (*hxt10Δ-hxt12Δ*) could be detected at a glance.

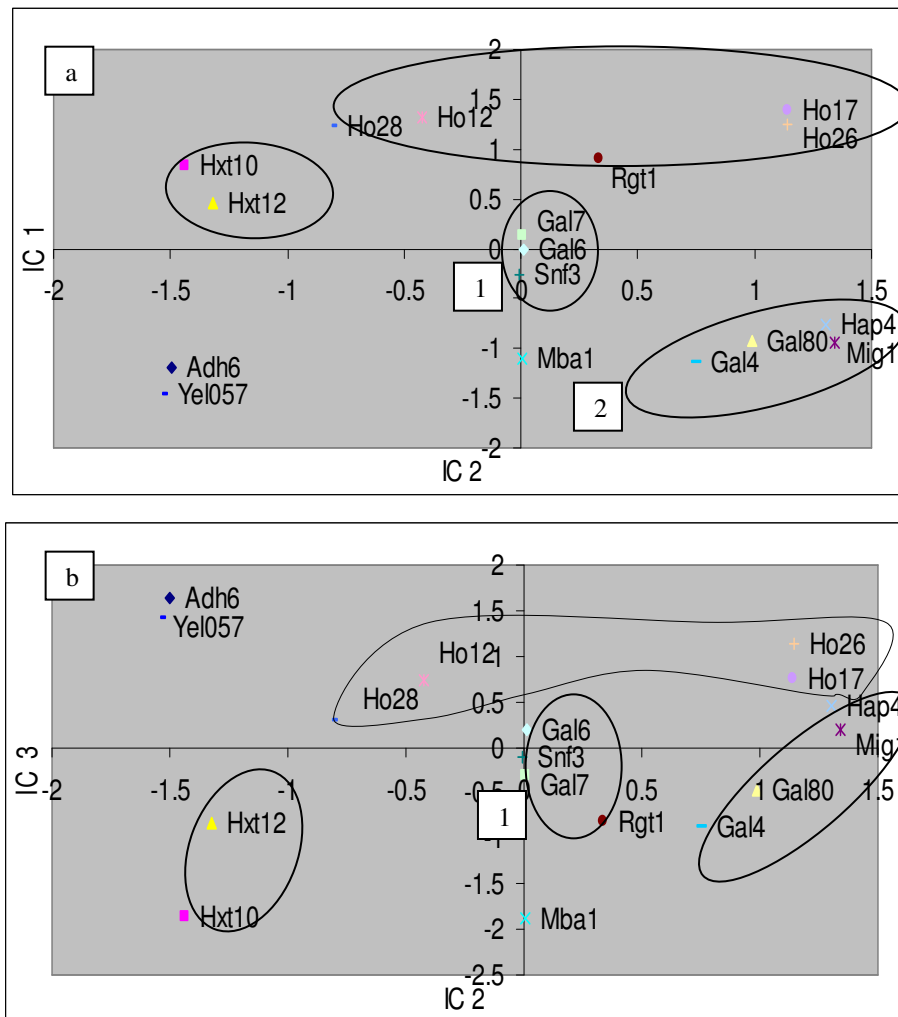


Figure 4.16. Scores of filtered metabolome data on first three ICs at 20th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

There are 4 negative ICs out of 6 after preprocessing PC data 6 times. The distribution of the extracellular metabolome data is better than shown in Figure 4.15. It could be stated that the PCA preprocessing procedure has a positive effect on the clustering power. The deletion mutants are easily identified and analysed according to their GO terms. In Figure 4.16 (a) and (b); cluster 1 includes *gal7Δ*, *snf3Δ*. In Figure 4.16 (a); cluster 2 includes *gal4Δ*, *gal80Δ*, *hap4Δ*, *mig1Δ*. The *p* value of *gal4Δ-gal80Δ-mig1Δ-hap4Δ* is 0.00046. While IC1 is helpful to distinguish *hoΔ* mutants and transporter genes, IC2 extracts respiration and regulation related genes.

4.1.5. PCA and ICA of Binned Extracellular Metabolome Data at 25th hour

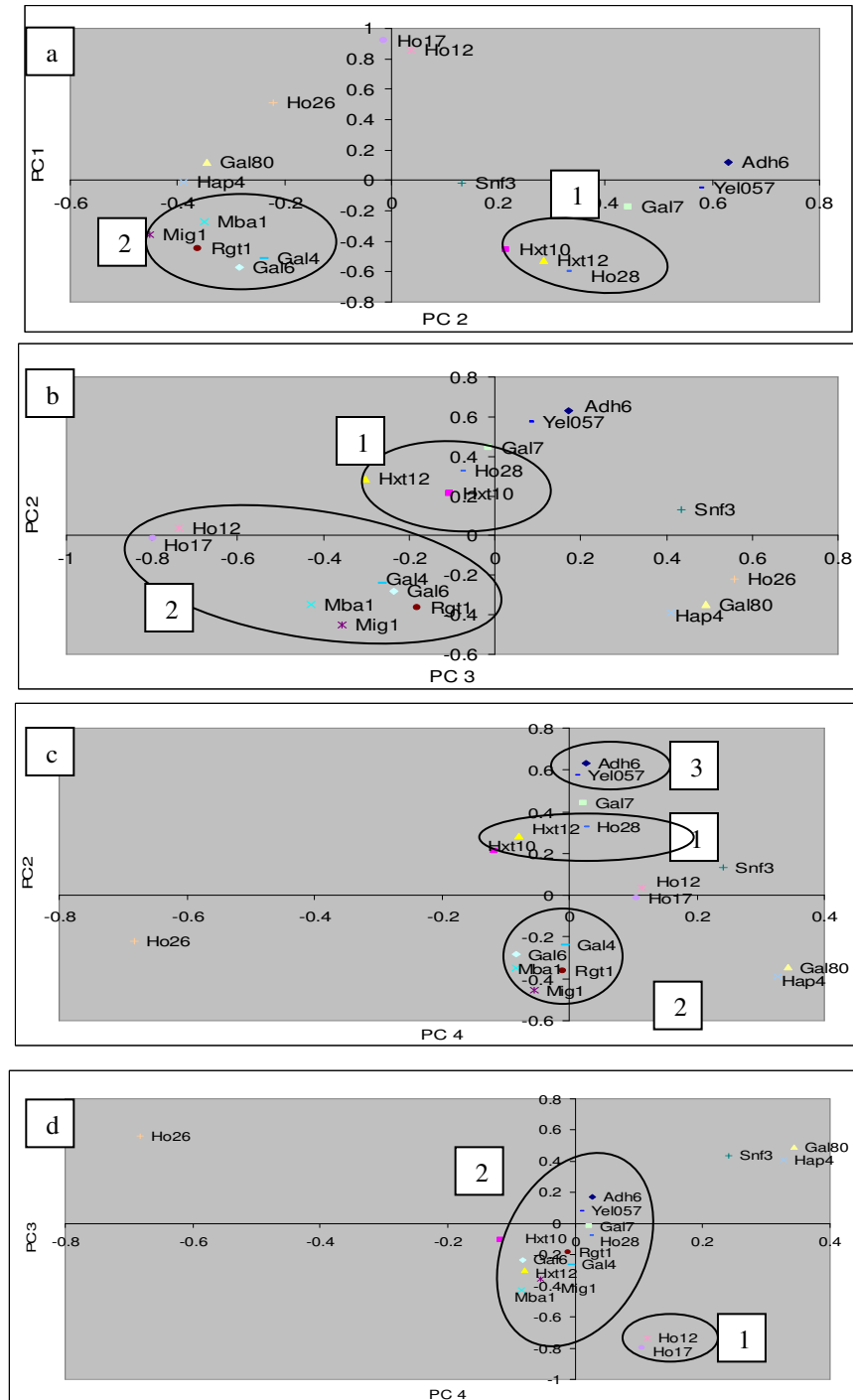


Figure 4.17. Scores of binned metabolome data on first four LVs at 25th hour of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

In Figure 4.17 (a), (b), and (c); cluster 1 includes *hxt10Δ*, *hxt12Δ*, *ho28Δ*. In Figure 4.17 (a); cluster 2 includes *mig1Δ*, *mba1Δ*, *gal4Δ*, *gal6Δ*, *rgt1Δ*. In Figure 4.17 (b); cluster 2 includes *ho12Δ*, *ho17Δ*, *mig1Δ*, *mba1Δ*, *gal4Δ*, *gal6Δ*, *rgt1Δ*. In Figure 4.17 (c); cluster 2 includes *gal4Δ*, *gal6Δ*, *rgt1Δ*, *mba1Δ*, *mig1Δ*; cluster 3 includes *ho12Δ* and *ho17Δ*. In Figure 4.17 (d); cluster 1 includes *ho12Δ* and *ho17Δ*; cluster 2 includes *adh6Δ*, *gal4Δ*, *gal6Δ*, *gal7Δ*, *rgt1Δ*, *mba1Δ*, *mig1Δ*, *yel057Δ*, *hxt10Δ*, *hxt12Δ*, *ho28Δ*. PCs of extracellular samples taken at 25th hour of fermentation could not be differentiated in Figure 4.17. Almost all of the deletion mutants overlap and the results that are implied from the graphs could not be correct. Another normalization technique may be investigated to discriminate these gene deletions.

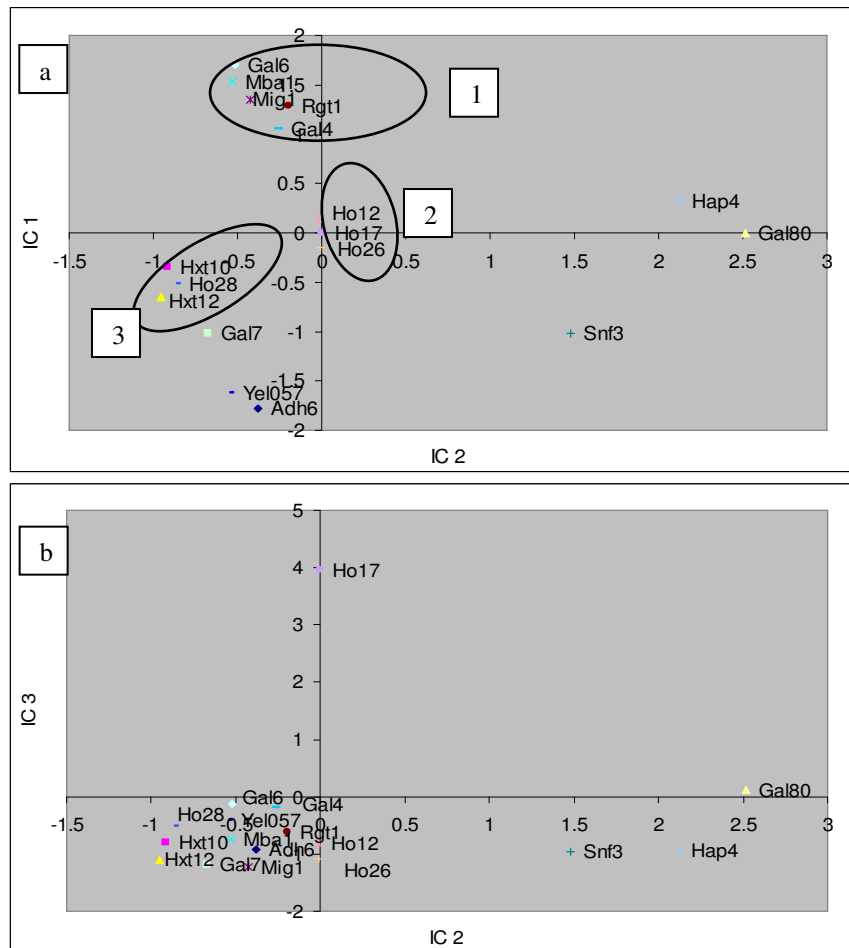


Figure 4.18. Scores of binned metabolome data on first three ICs at 25th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

In Figure 4.18 (a).; cluster 1 includes *rgt1Δ*, *gal4Δ*, *gal6Δ*, *mig1Δ*, *mba1Δ*; cluster 2 includes *ho12Δ*, *ho17Δ*, *ho26Δ*; cluster 3 includes *hxt10Δ*, *hxt12Δ*, *ho28Δ*. ICA results shown in Figure 4.18 give no evidence about the biological or functional relevance of mutants like PCA. There is only one group that is extracted from graph. *Gal80* and *hap4* may probably have an activity in the transcriptional regulation. The p value for this deletion group is 0.00991 and quite reasonable to accept the relationship between these mutants.

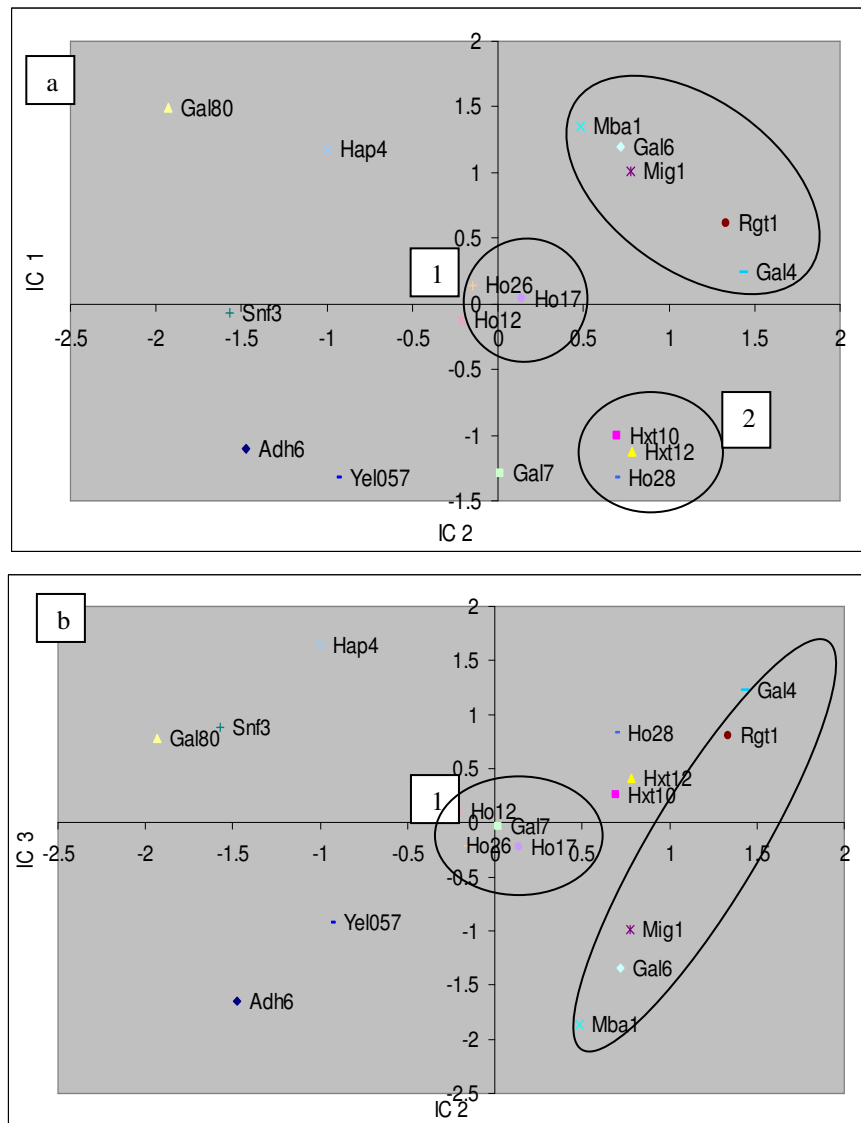


Figure 4.19. Scores of binned metabolome data on first three ICs at 25th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

When the PCA preprocessing is increased to 6, the distribution of deletion mutants is better compared to previous case, but it is still not satisfactory (Figure 4.19).

In Figure 4.19 (a); cluster 1 includes *ho12Δ*, *ho17Δ* and *ho26Δ*; cluster 2 includes *hxt10Δ*, *hxt12Δ*, and *ho28Δ*. In Figure 4.19 (b).; cluster 1 includes *ho12Δ*, *ho17Δ*, and *ho26Δ*.

In Figure 4.19 (a), IC2 gives a cluster of *gal6Δ*, *mig1Δ*, *mba1Δ*, *rgt1Δ* and *gal4Δ* deletion mutants. The *p* value of this set which consists of *gal6*, *mig1*, *rgt1* and *gal4* is 7.38e-05 and low enough to accept the possibility of biological relevance of being in the same function. On the other hand, remaining groups could not be correlated due to large *p* values or insufficient data as can be seen from the figure by the scattering of data.

4.1.6. PCA and ICA of Filtered Extracellular Metabolome Data at 25th hour

PCA was applied to column vector normalized data of 70 out of 1000 mass variables. Figure 4.20 shows the comparison of scores for extracellular metabolome data on the first 4 latent values.

In Figure 4.20 (a); cluster 1 includes *rgt1Δ*, *gal4Δ*, *gal6Δ*, *gal80Δ*, *mba1Δ*, *mig1Δ*, *hap4Δ*; cluster 2 includes *hxt10Δ*, *hxt12Δ*, *ho28Δ*, *gal7Δ*. In Figure 4.20 (b); cluster 1 includes *snf3Δ*, *ho28Δ*, *hxt10Δ*, *hxt12Δ*, *gal7Δ*; cluster 2 includes *adh6Δ*, *yel057cΔ*.

In Figure 4.20 (c); cluster 1 includes *adh6Δ*, *yel057cΔ*; cluster 2 includes *rgt1Δ*, *gal4Δ*, *gal6Δ*, *mig1Δ*, *mba1Δ*; cluster 3 includes *hxt10Δ*, *hxt12Δ*, *ho28Δ*.

In Figure 4.20 (d); cluster 1 includes *gal4Δ*, *rgt1Δ*; cluster 2 includes *gal6Δ*, *gal7Δ*, *snf3Δ*, *hxt10Δ*, *hxt12Δ*, *ho28Δ*. The modelling capability of LV2 and LV4 is better when compared to LV1 and LV3.

ICA was applied to PC data and the results are shown in Figure 4.21 and 4.22.

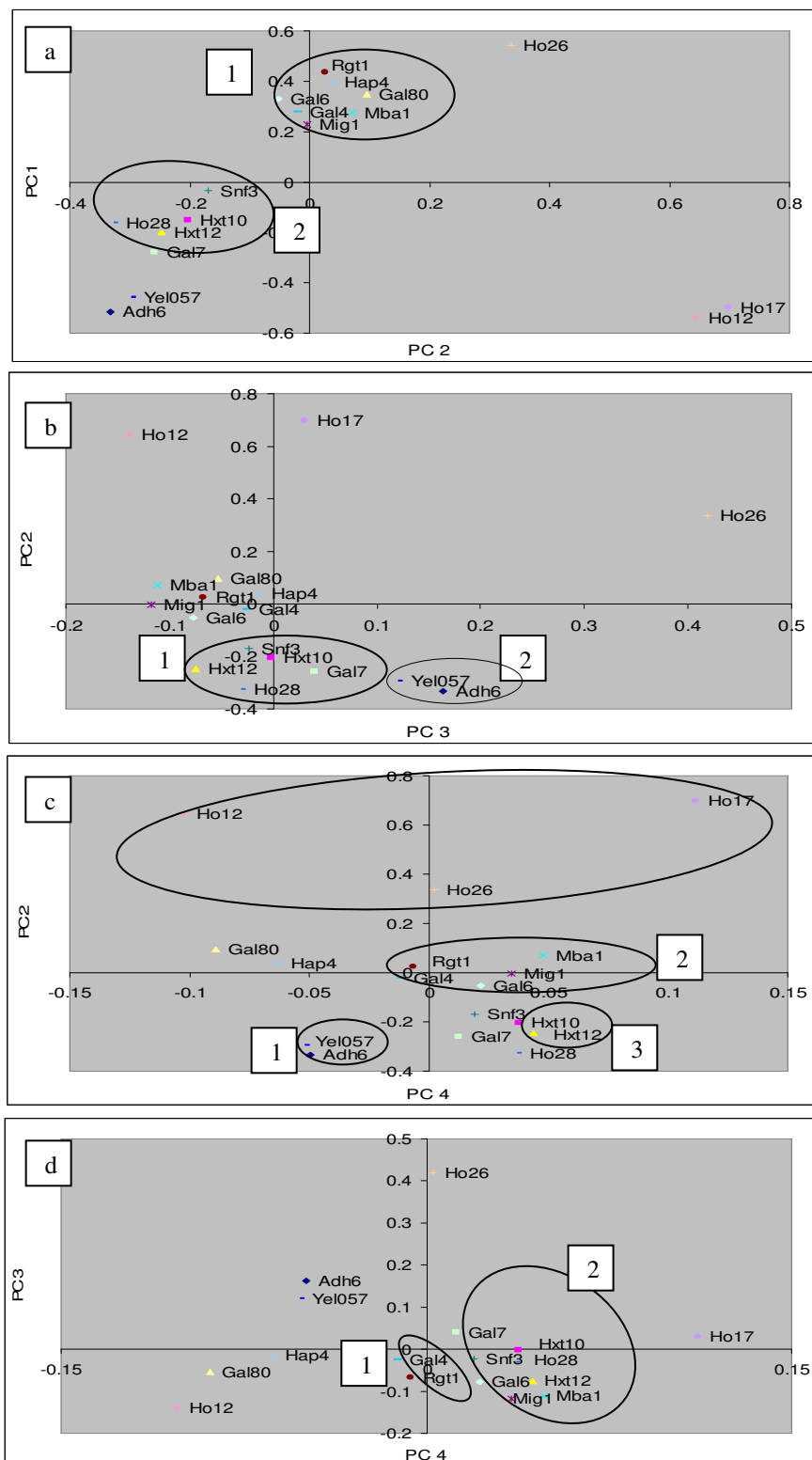


Figure 4.20. Scores of filtered metabolome data on first four LVs at 25th hour of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

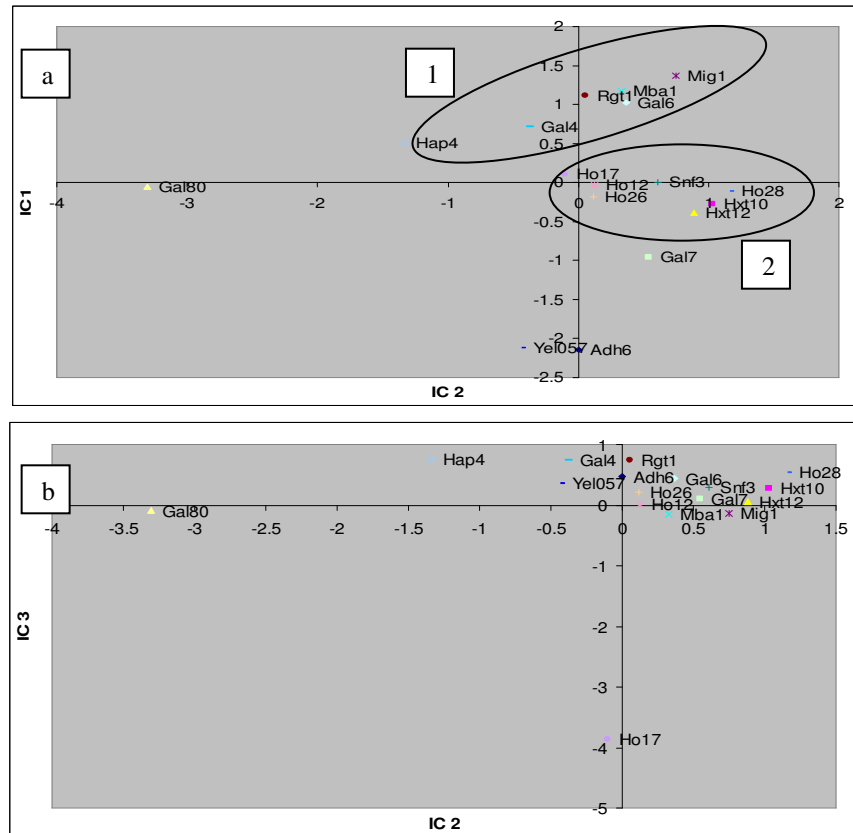


Figure 4.21. Scores of filtered metabolome data on first three ICs at 25th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

There is only one negative IC out of 5 after preprocessing PC data by 5 times. In other words, the power of clustering strongly decreased due to kurtosis or less negative IC effect. The more negative values the ICs in kurtosis reach, the better clustering the deletion mutants get.

In Figure 4.21 (a); cluster 1 includes *rgt1Δ*, *mig1Δ*, *mba1Δ*, *gal6Δ*, *gal4Δ* and *hap4Δ*; cluster 2 includes *ho12Δ*, *ho17Δ*, *ho26Δ*, *snf3Δ*, *ho28Δ*, *hxt10Δ*, *hxt12Δ*. There is also one negative IC out of 6 after preprocessing PC data by 6 times in Figure 4.22. The distribution of gene deletions are quite complex. In Figure 4.22 (a); cluster 1 includes *ho12Δ*, *ho17Δ*, *ho26Δ*, *gal80Δ*. In Figure 4.22 (b); cluster 1 includes *hxt10Δ*, *hxt12Δ*, *mig1Δ*, *mba1Δ*, *gal6Δ*, *ho28Δ*; cluster 2 includes *ho12Δ*, *ho17Δ*, *ho26Δ*, *gal7Δ*. One can say that IC1 separates the genes with regulation activity.

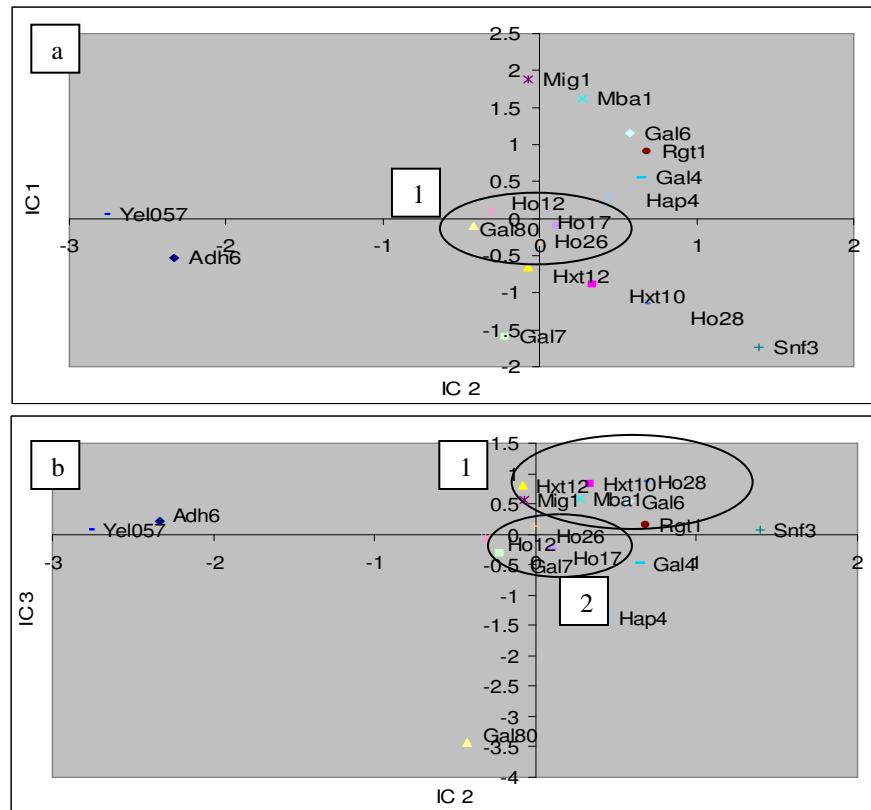


Figure 4.22. Scores of filtered metabolome data on first three ICs at 25th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

4.1.7. PCA and ICA of Binned Extracellular Metabolome Data at 31st hour

In Figure 4.23 (a); cluster 1 includes *adh6Δ*, *hxt10Δ*, *hxt12Δ*, *gal7Δ*, *snf3Δ*, *yel057cΔ*, *ho12Δ*. In Figure 4.23 (b); cluster 1 includes *gal80Δ*, *ho17Δ*, *ho26Δ*, *mba1Δ*; cluster 2 includes *adh6Δ*, *hxt10Δ*, *hxt12Δ*, *gal7Δ* and cluster 3 includes *ho12Δ*, *yel057Δ*. In Figure 4.23 (c) and (d); cluster 1 includes *yel057cΔ*, *ho12Δ*; cluster 2 includes *adh6Δ*, *hxt10Δ*, *hxt12Δ*, *gal7Δ*. In Figure 4.23 (d); cluster 3 includes *ho17Δ*, *ho26Δ*, *gal80Δ*, *mba1Δ*. The widely used assumption that the desired information is related to the highest variance in the dataset does not hold for this set of experiments. Thus satisfactory results cannot be obtained by solely optimizing the variance, as is done by PCA. For example, there is no evidence about any functional or biological relevance for the data found in Figure 4.23. Possible clusters such as *galΔ* strains and *hoΔ* reference strain remained far apart from each other which should have been closer.

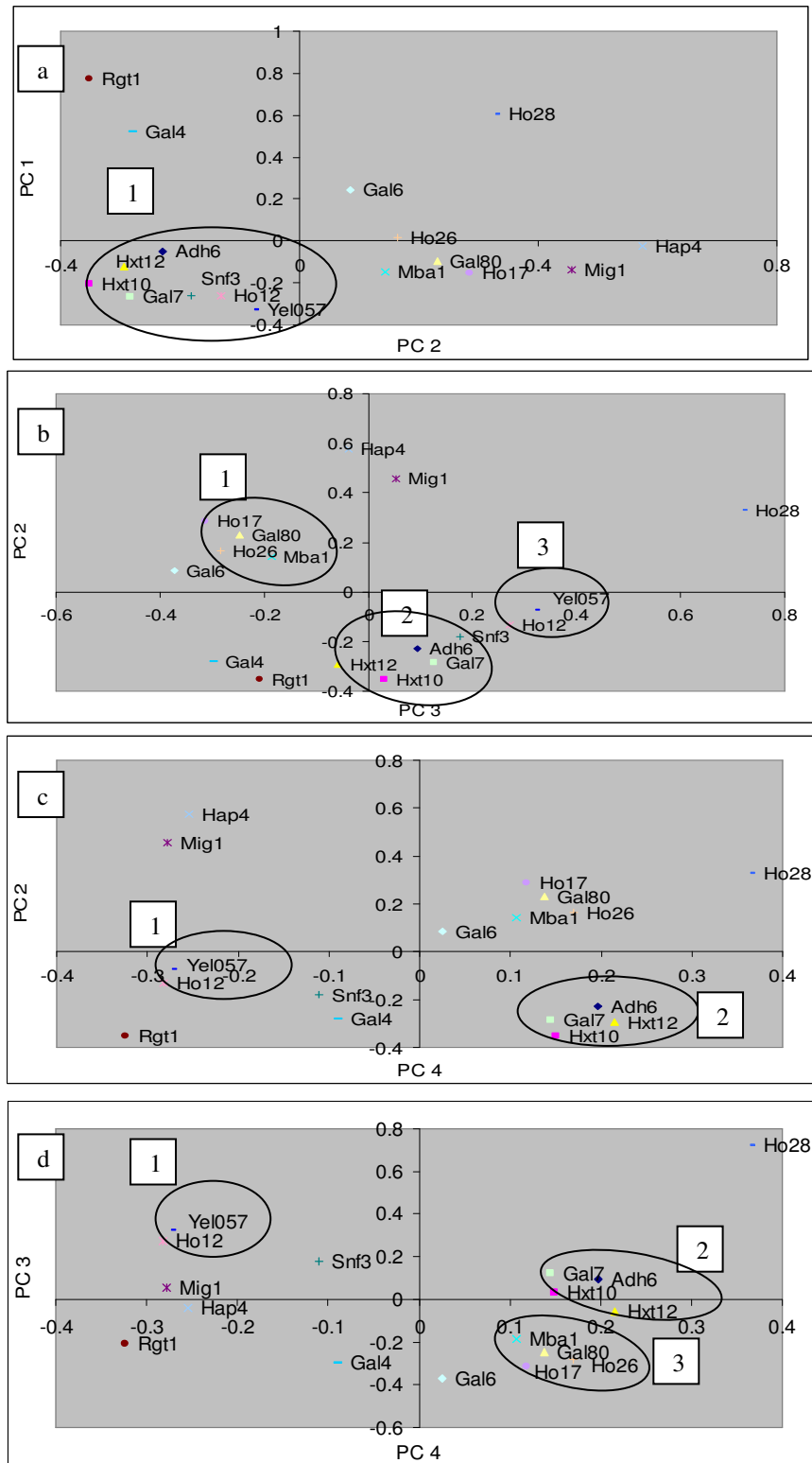


Figure 4.23. Scores of binned metabolome data on first four Latent Variables at 31st hour of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

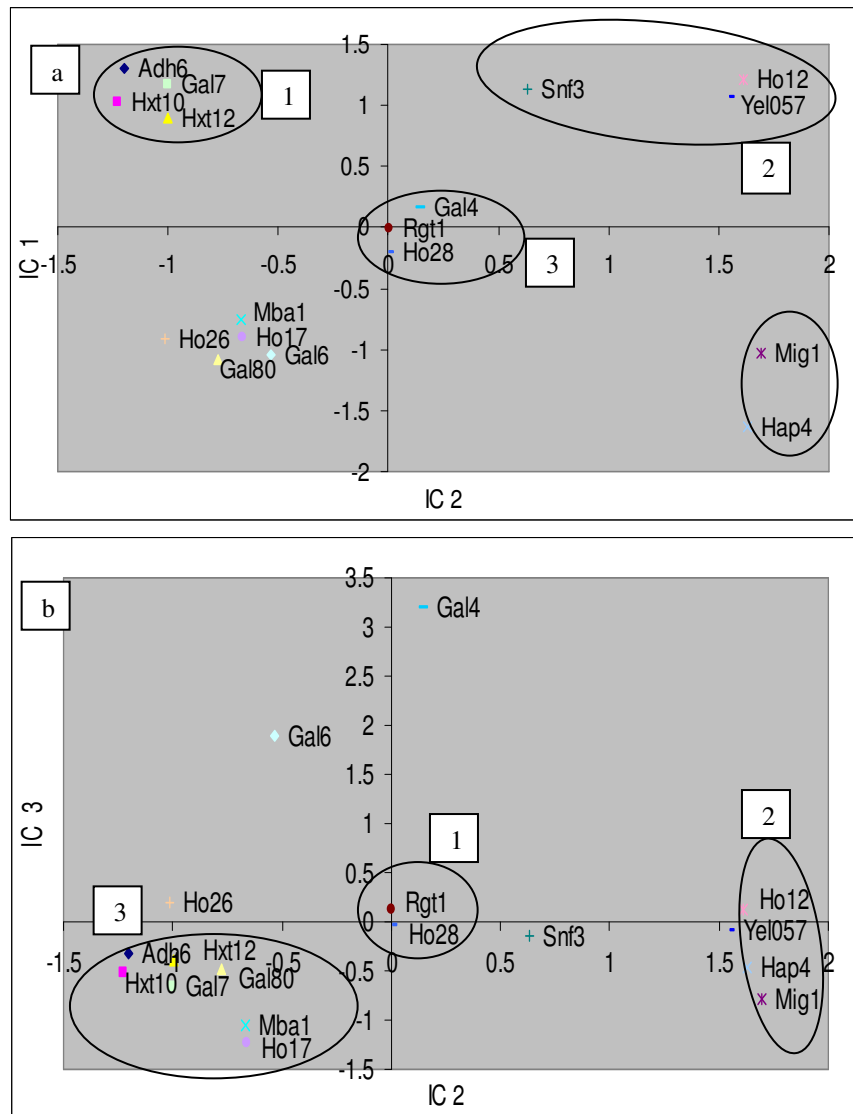


Figure 4.24. Scores of binned metabolome data on first three ICs at 31th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

In Figure 4.24 (a); cluster 1 includes *adh6Δ*, *hxt10Δ*, *hxt12Δ*, *gal7Δ*; cluster 2 includes *ho12Δ*, *snf3Δ* and *yel057cΔ*; cluster 3 includes *rgt1Δ*, *ho28Δ* and *gal4Δ*. In Figure 4.24 (b); cluster 1 includes *rgt1Δ*, *ho28Δ*; cluster 2 includes *hap4Δ*, *ho12Δ*, *mig1Δ* and *yel057cΔ*; cluster 3 consists of remaining 7 deletions. Any discrimination cannot be found from the ICA results shown in Figure 4.24. The deletion mutants are not dispersed effectively. In other words, ICA is not helpful to extract the clustering information from latent values of the metabolome profiles of these mutants.

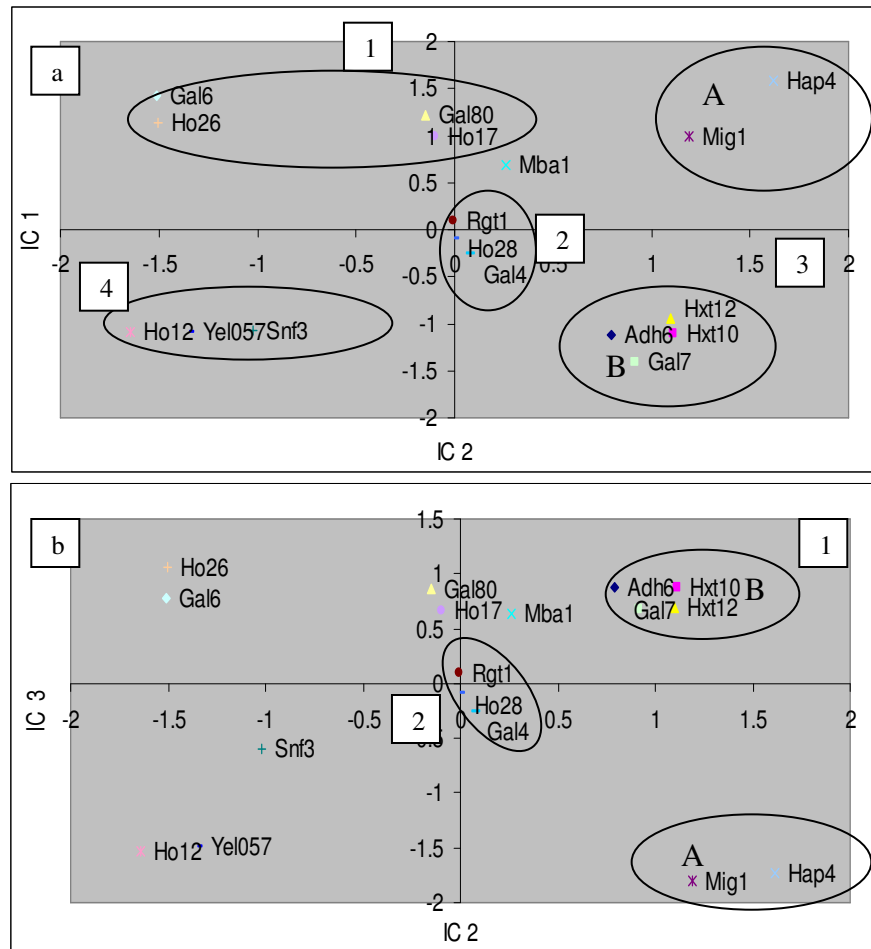
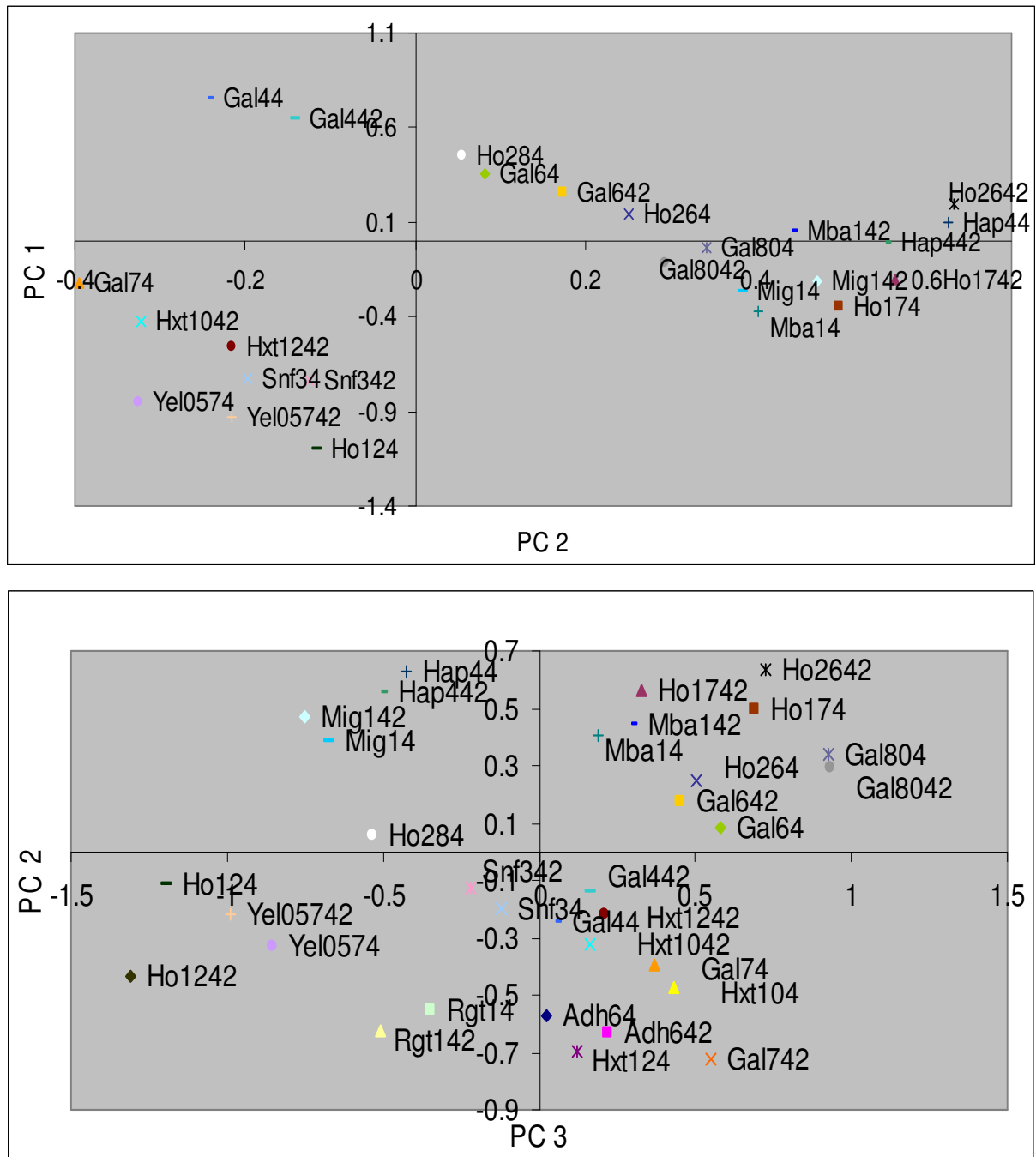


Figure 4.25. Scores of binned metabolome data on first three ICs at 31th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

Figure 4.25 shows the ICA results when the PCA preprocessing is increased to 6. In Figure 4.25 (a); cluster 1 includes *gal6Δ*, *gal80Δ*, *ho17Δ*, *ho26Δ*; cluster 2 includes *ho28Δ*, *gal4Δ*, *rgt1Δ*; cluster 3 includes *adh6Δ*, *hxt10Δ*, *hxt12Δ*, *gal7Δ*; cluster 4 includes *ho12Δ*, *yel057cΔ*, *snf3Δ*. In Figure 4.25 (b); cluster 1 includes *adh6Δ*, *hxt10Δ*, *hxt12Δ*, *gal7Δ*; cluster 2 includes *ho28Δ*, *gal4Δ*, *rgt1Δ*. The only difference between the results of Figure 4.25 (a) and (b) is the projection of some mutants around the coordinate axis. While *mig1Δ-hap4Δ* group denoted by circle A, appear in the upper right side in Figure 4.25 (a), it is projected around the x axis in Figure 4.25 (b). Similarly, the gene deletions denoted by circle B (*gal6Δ*, *gal7Δ*, *hxt10Δ*, *hxt12Δ*) in Figure 4.25 (a) are also projected around the x-axis in Figure 4.25 (b).



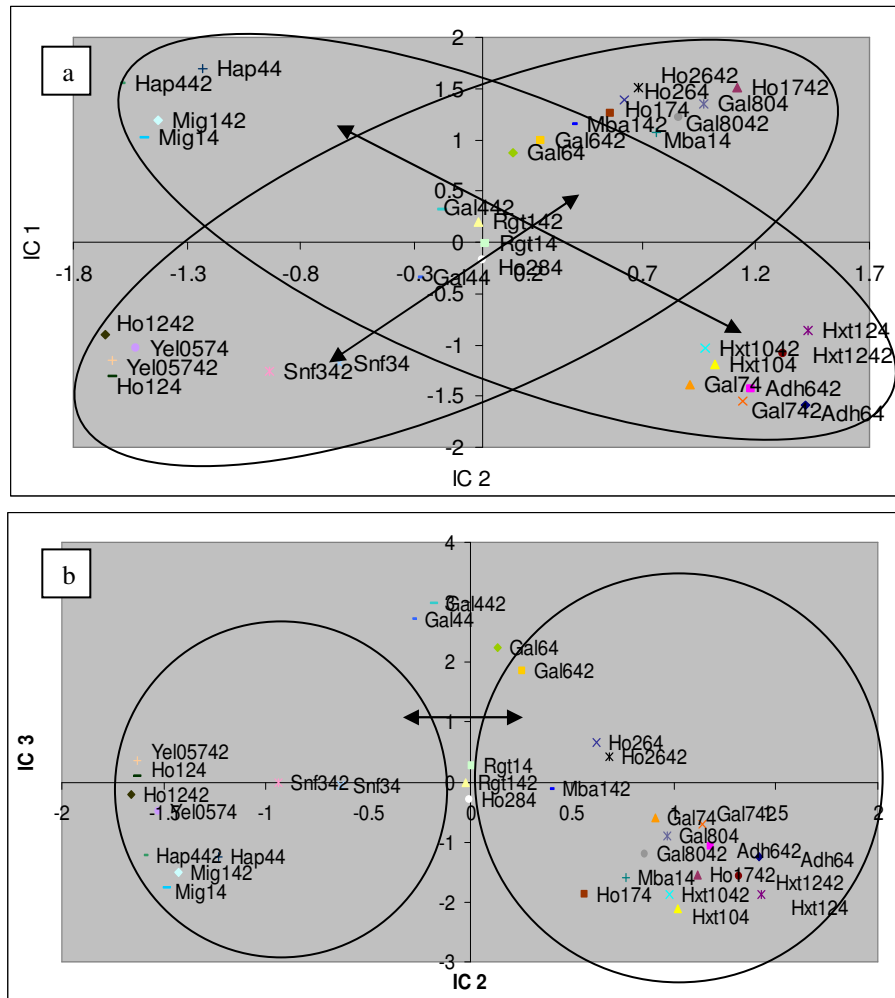


Figure 4.27. Scores of binned replicate metabolome data on first three ICs at 31st hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

ICs of replicate metabolome data in Figure 4.27 is similar to those in Figure 4.24. The gene nodes are projected around first ($y=x$) and second ($y=-x$) bisectors. While the deletion mutants in Figure 4.27 (a) are symmetrical around the bisectors, the deletion mutants in Figure 4.27 (b) are symmetrical around y-axis (like mirror image).

Figure 4.28 and 4.25 which show the ICs of metabolome data preprocessed by 6 times are identical. The only difference between these two figures is the projection of deletion mutants in IC2 vs IC3 plot around x-axis. The upper and lower regions of Figure 4.28 (b) reflects the shifts of mutants compared to their positions in Figure 4.25 (b).

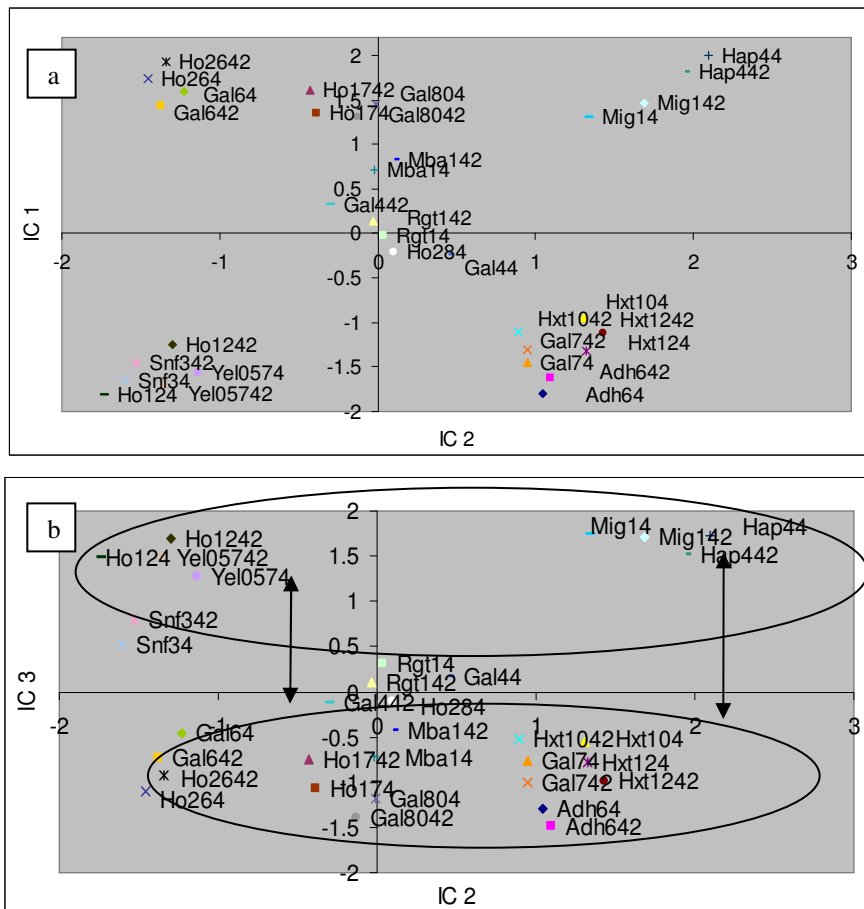


Figure 4.28. Scores of binned replicate metabolome data on first three ICs at 31st hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

4.1.8. PCA and ICA of Filtered Extracellular Metabolome Data at 31st hour

PCA was applied to column vector normalized data of 115 out of 1000 mass variables (Figure 4.29). In Figure 4.29 (a); cluster 1 includes *ho17Δ*, *ho26Δ*, *gal4Δ*, *gal6Δ*, *gal80Δ*, *mig1Δ*, *mba1Δ*; cluster 2 includes *gal7Δ*, *ho12Δ*, *ho28Δ*, *adh6Δ*, *hxt10Δ*, *hxt12Δ*, *snf3Δ*, *yel057cΔ*. In Figure 4.29 (b); cluster 1 includes *adh6Δ*, *gal6Δ*, *gal7Δ*, *gal80Δ*, *ho12Δ*, *ho26Δ*, *hxt10Δ*, *hxt12Δ*, *snf3Δ*, *mba1Δ*, *yel057cΔ* cluster 2 includes *hap4Δ*, *ho17Δ*, *mig1Δ*. In Figure 4.29 (c); cluster 1 includes *adh6Δ*, *gal6Δ*, *gal7Δ*, *gal80Δ*, *ho12Δ*, *hxt10Δ*, *hxt12Δ*, *snf3Δ*, *mba1Δ*, *rgt1Δ*, *yel057cΔ*; cluster 2 includes *ho17Δ*, *mig1Δ*. In Figure 4.29 (d); cluster 1 includes *adh6Δ*, *gal7Δ*, *gal80Δ*, *ho12Δ*, *ho28Δ*, *hxt10Δ*, *hxt12Δ*, *snf3Δ*, *mba1Δ*, *yel057cΔ*, and cluster 2 includes *gal6Δ*, *hxt10Δ*, *hxt12Δ*.

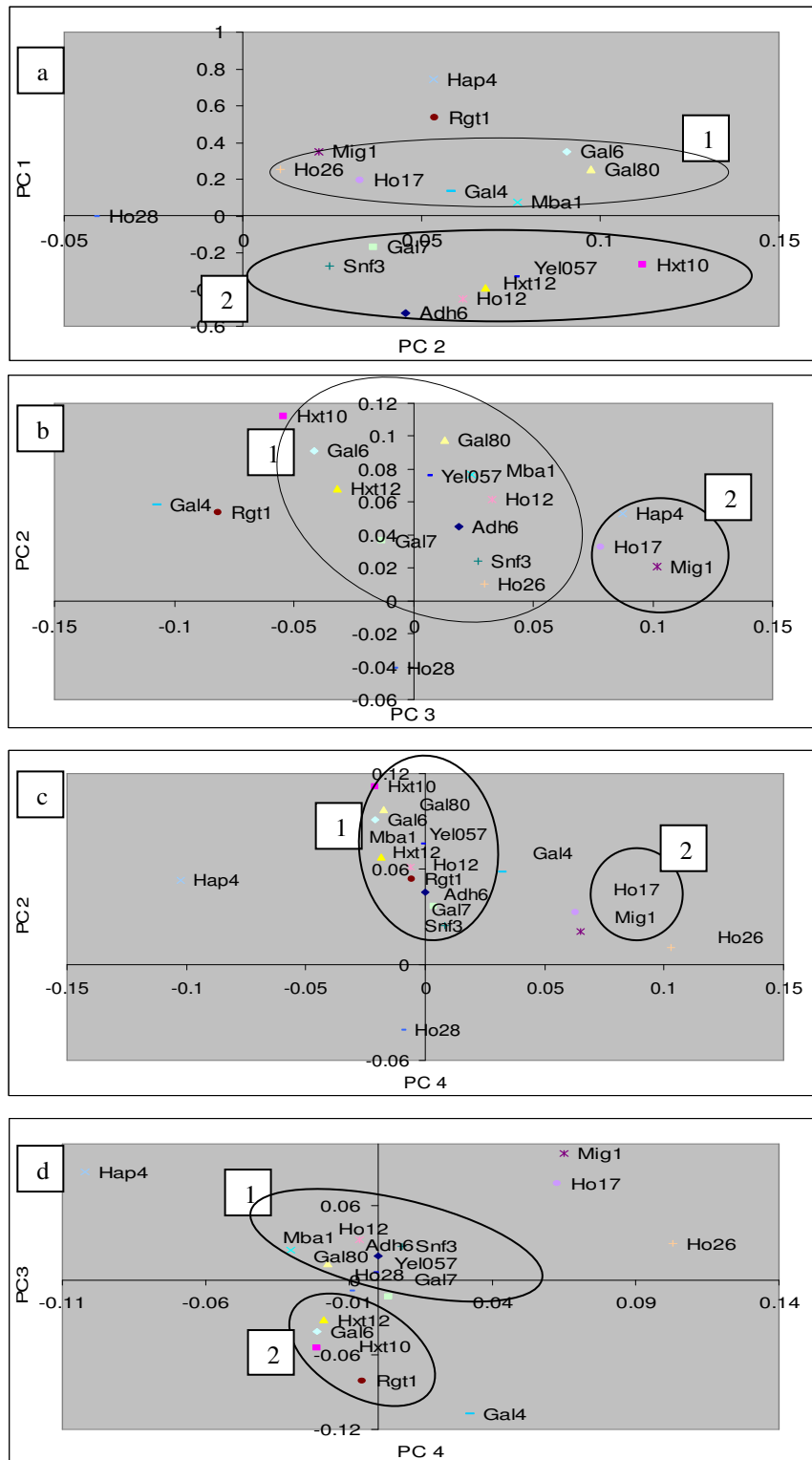


Figure 4.29. Scores of filtered metabolome data on first four LVs at 31st hour of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

The present filtering procedure is not useful for separating the deletion mutants in Figure 4.29. The hidden clustering underneath the selected metabolome data cannot be elucidated. One reason for that is the specific mass variables belonging to gene deletions could be omitted after filtering the whole metabolome data. Another reason is the incapability of PCA compared to ICA.

IC results of extracellular metabolome data obtained at 31st hour of fermentation are shown in Figure 4.30. In Fig. 4.30 (a); cluster 1 includes *rgt1* Δ , *hap4* Δ , *ho28* Δ . In Fig. 4.30 (b); cluster 1 includes *hap4* Δ , *ho28* Δ ; cluster 2 includes *adh6* Δ , *snf3* Δ , *ho12* Δ . There is one negative IC out of 5 after preprocessing PC data by 5 times. Nevertheless, IC2 is able to discriminate the gal and hxt groups in the left half of the Figures 4.30 (a) and (b).

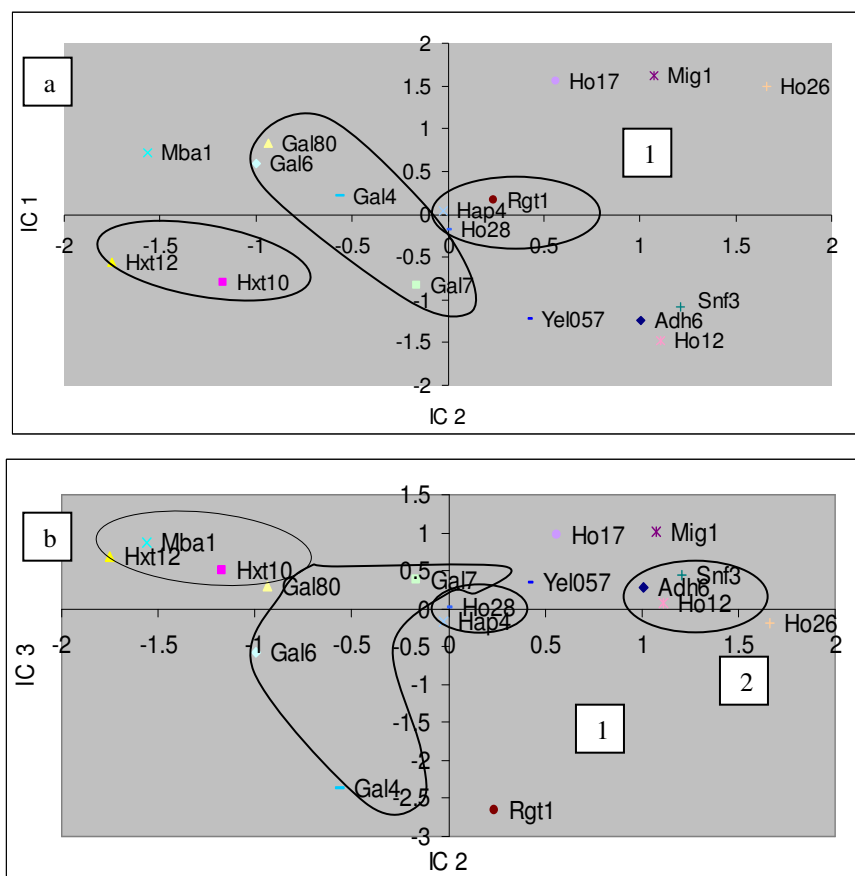


Figure 4.30. Scores of filtered metabolome data on first three ICs at 31th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

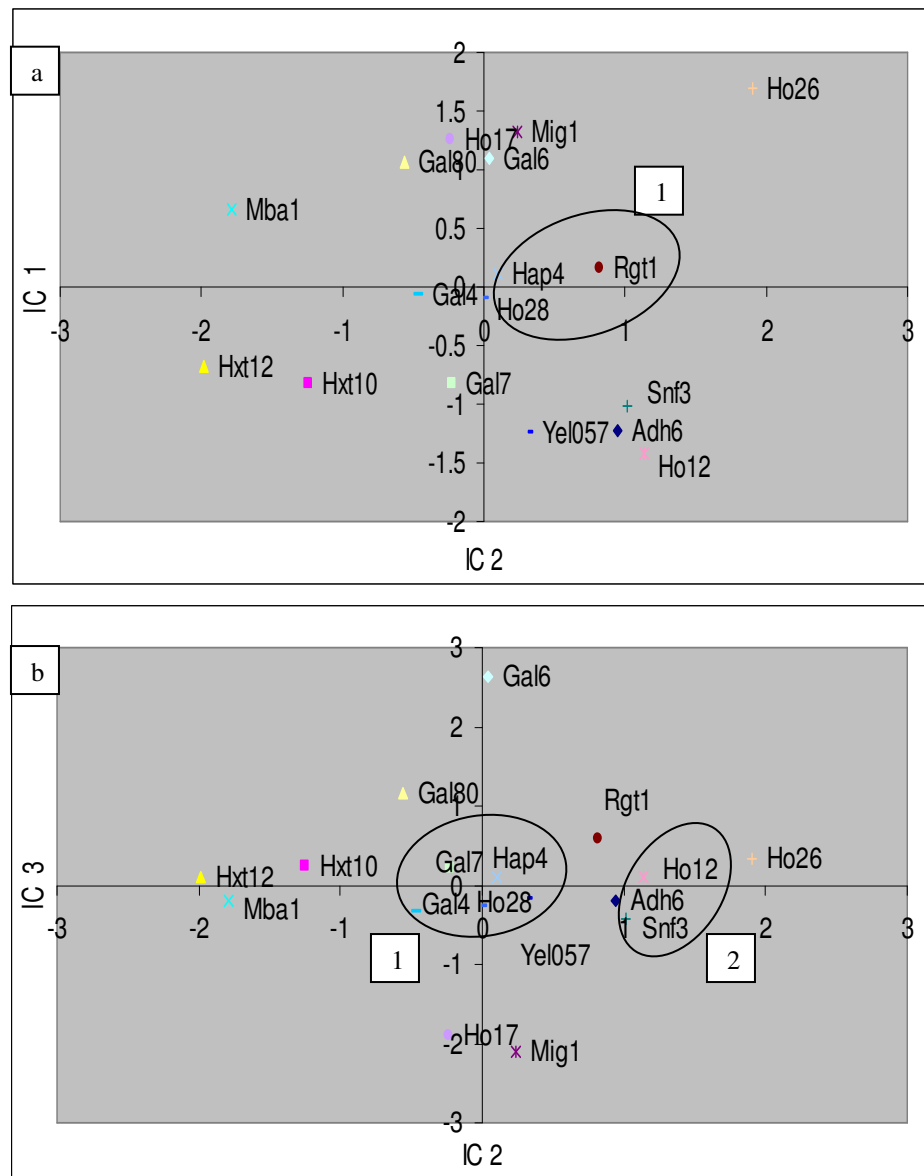


Figure 4.31. Scores of filtered metabolome data on first three ICs at 31th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

Preprocessing PCA data by 6 times is not helpful to find out satisfactory clusters. One reason could be the kurtosis measure. In other words, there should have been more negative ICs for efficient discriminations. In Fig. 4.31 (a); cluster 1 includes *rgt1* Δ , *hap4* Δ , *ho28* Δ . In Fig. 4.31 (b); cluster 1 includes *gal7* Δ , *hap4* Δ , *ho28* Δ ; cluster 2 includes *adh6* Δ , *snf3* Δ , *ho12* Δ . IC2 again discriminates the genes related to galactose metabolism and transport.

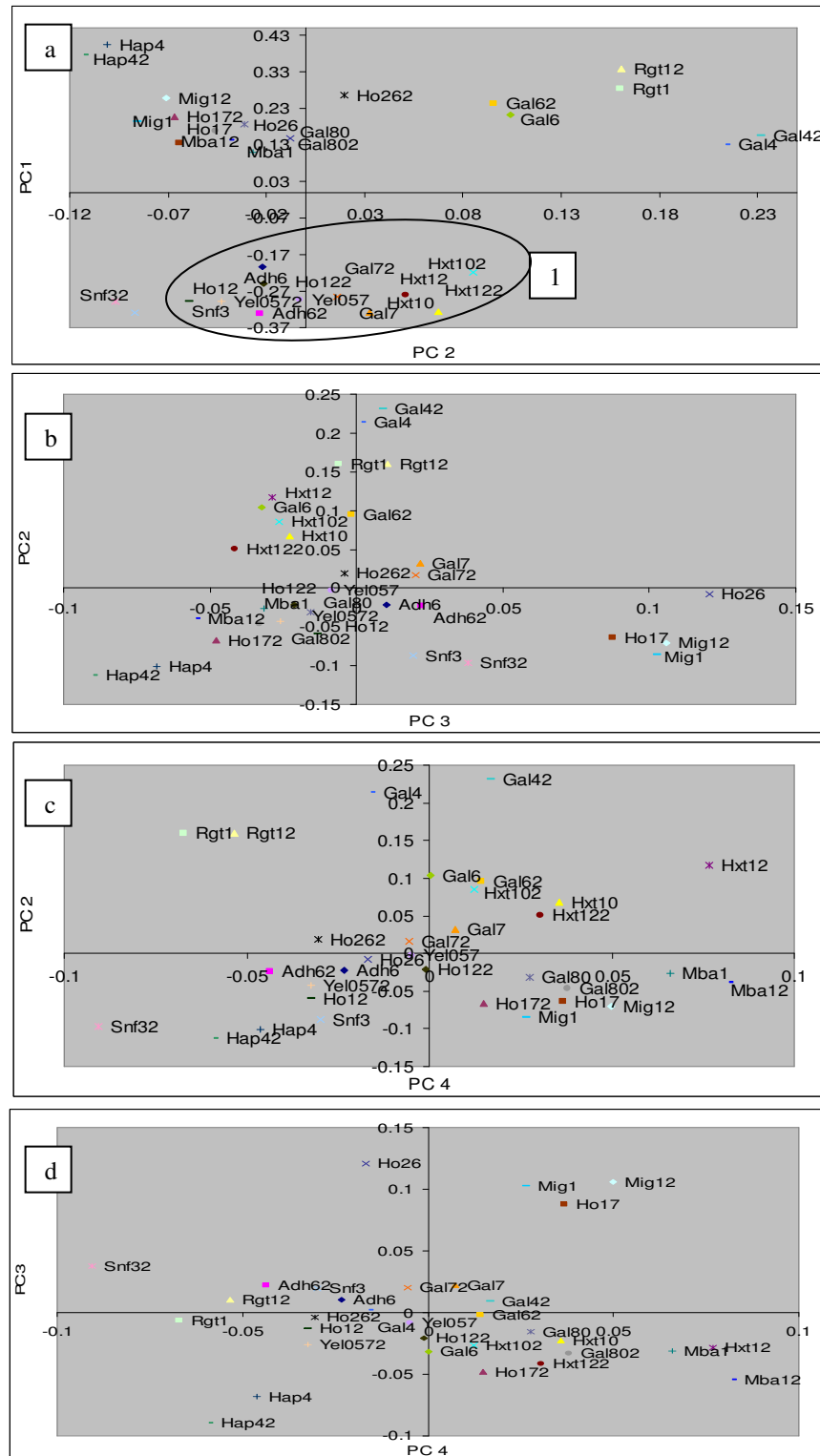


Figure 4.32. Scores of filtered replicate metabolome data on first four LVs at 31st hour of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

Scores of the replicate metabolome data on the first four LVs are plotted in Figure 4.32. The replicates overlap onto their original data set. So, the results confirm the validity of PCA.

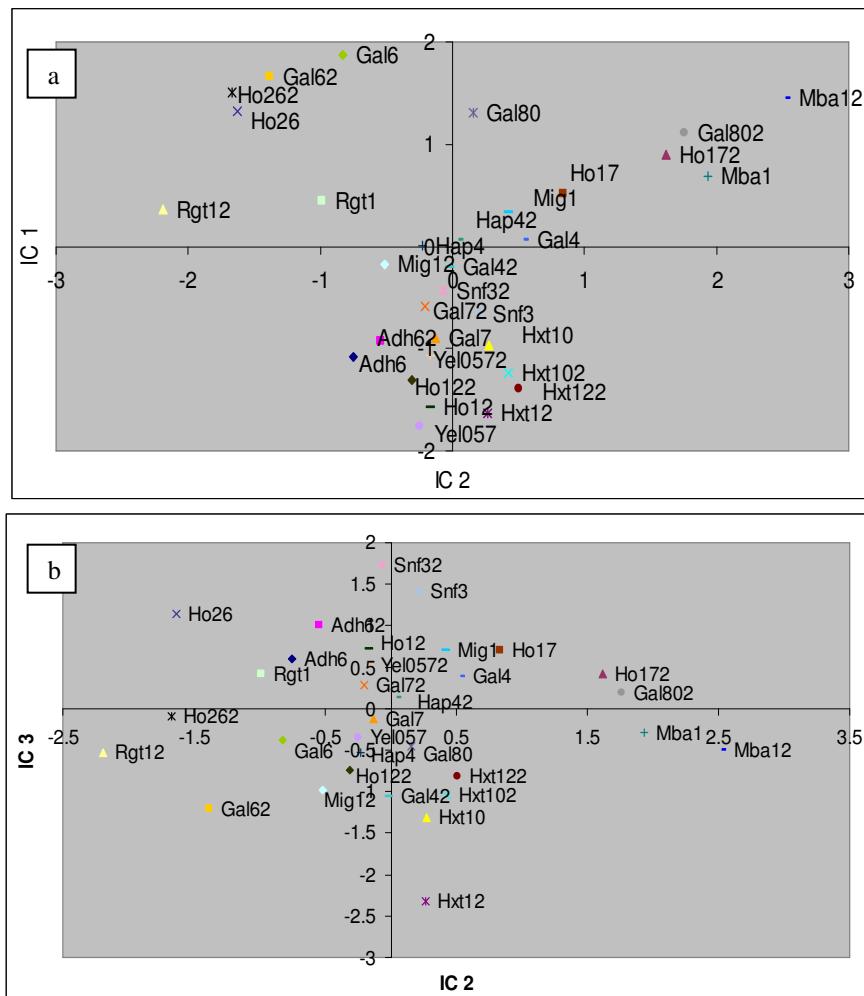


Figure 4.34. Scores of filtered replicate metabolome data on first three ICs at 31th hour of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

4.1.9. PCA and ICA of All Extracellular Metabolome Data at Four Time Intervals (Preprocessing by Binning)

In Figure 4.35 (a), 4.35 (b) and 4.35 (c); cluster 1 includes all extracellular metabolome data collected at 25th hour of fermentation. In Figure 4.35 (a); cluster 2 includes metabolome data belonging to deletion mutants at 31st hour of fermentation, and cluster 3 includes metabolome data of deletion mutants at 15th and 20th hour of fermentation. In Figure 4.35 (b); cluster 2 includes metabolome data at 20th hour of fermentation. There is not any available clustering in Figure 4.35 (d). Particularly, PC1 and PC2 give better results than the other PCs.

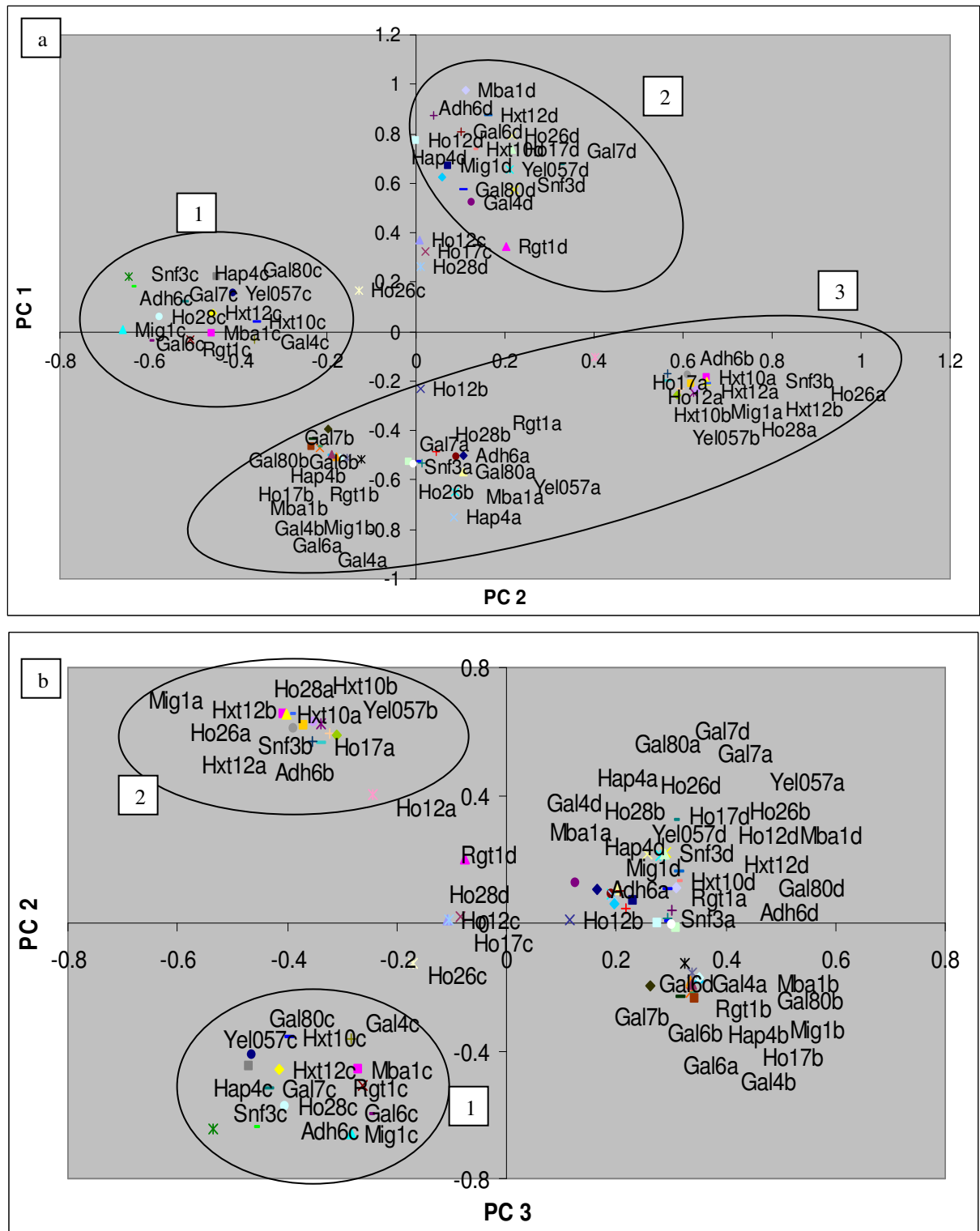


Figure 4.35. Scores of all metabolome data binned on first four LVs at 4 time intervals of extracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

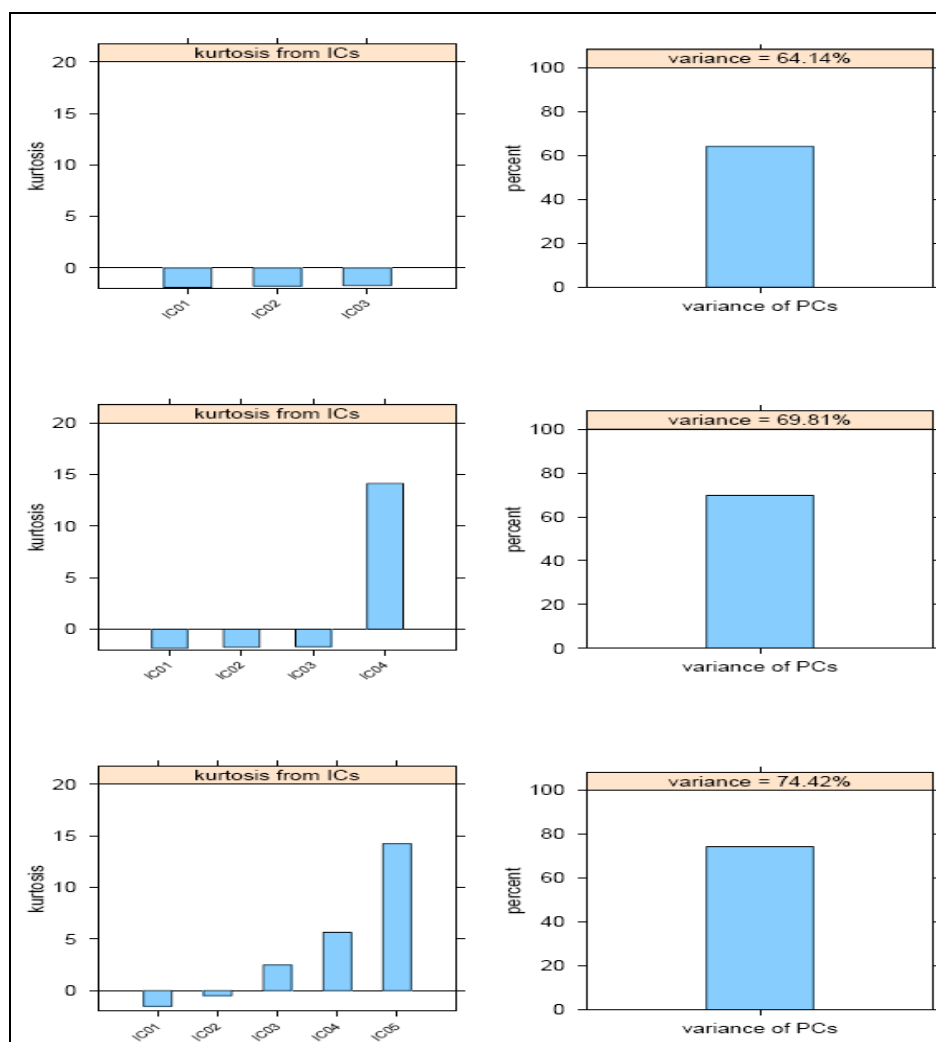


Figure 4.36. Kurtosis measure of ICA data

The kurtosis measure of ICA data is given in Figure 4.36. It could be implied that the best clustering will be obtained after preprocessing PCA data by 3 times. In other words, only the first three PCs should be used to perform ICA ideally.

In Figure 4.37 (a) and 4.37 (b), cluster 1 includes all extracellular metabolome data of deletion mutants at 31st hour of fermentation; cluster 2 includes metabolome data at 25th hour of fermentation. In Figures 4.37 (a) and (b); cluster 3 and 4 includes metabolome data at both 15th and 20th hour of fermentation. ICA is not helpful to discriminate these two groups (15th and 20th hour).

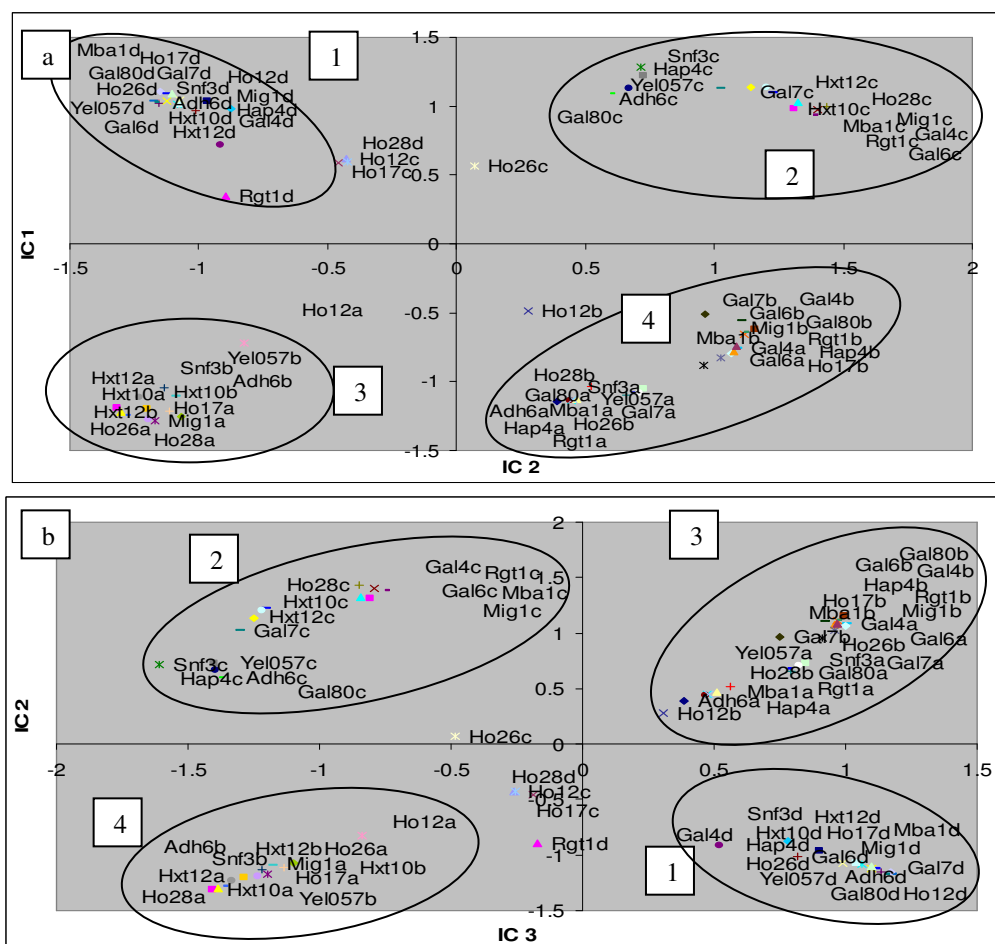


Figure 4.37. Scores of binned all metabolome data on first three ICs at 4 time intervals of extracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 3)

As an overall result, while PC1 and PC2 are able to discriminate 25th and 31st hour of extracellular metabolome data, PC3 extracts metabolome data at 20th hour of fermentation successfully with PC2. In ICA; the metabolome data at 25th and 31st hour of fermentation are always discriminated. But ICA is not as helpful as PCA in revealing other clusterings such as data at 15th or 20th hour of fermentation.

4.2. Metabolite Fingerprinting

Intracellular metabolome patterns were determined by Mass Spectrometer. Metabolic fingerprints of different yeast deletion mutants were compared using the bioinformatic tools of PCA, ICA and HCA.

The metabolome profiles in Figure 4.38 are almost the same. Visual inspection of intracellular metabolome profiles indicated that peaks with high intensities were obtained nearly at 260, 282, 4.459... m/z values. These peaks would probably come from the medium or extraction solutions. So, the metabolites which have low intensity values are the main source of statistical analysis as they may underlie significant information. Although the global structure of the spectra is very similar for each deletion mutant there are differences among masses of smaller intensities.

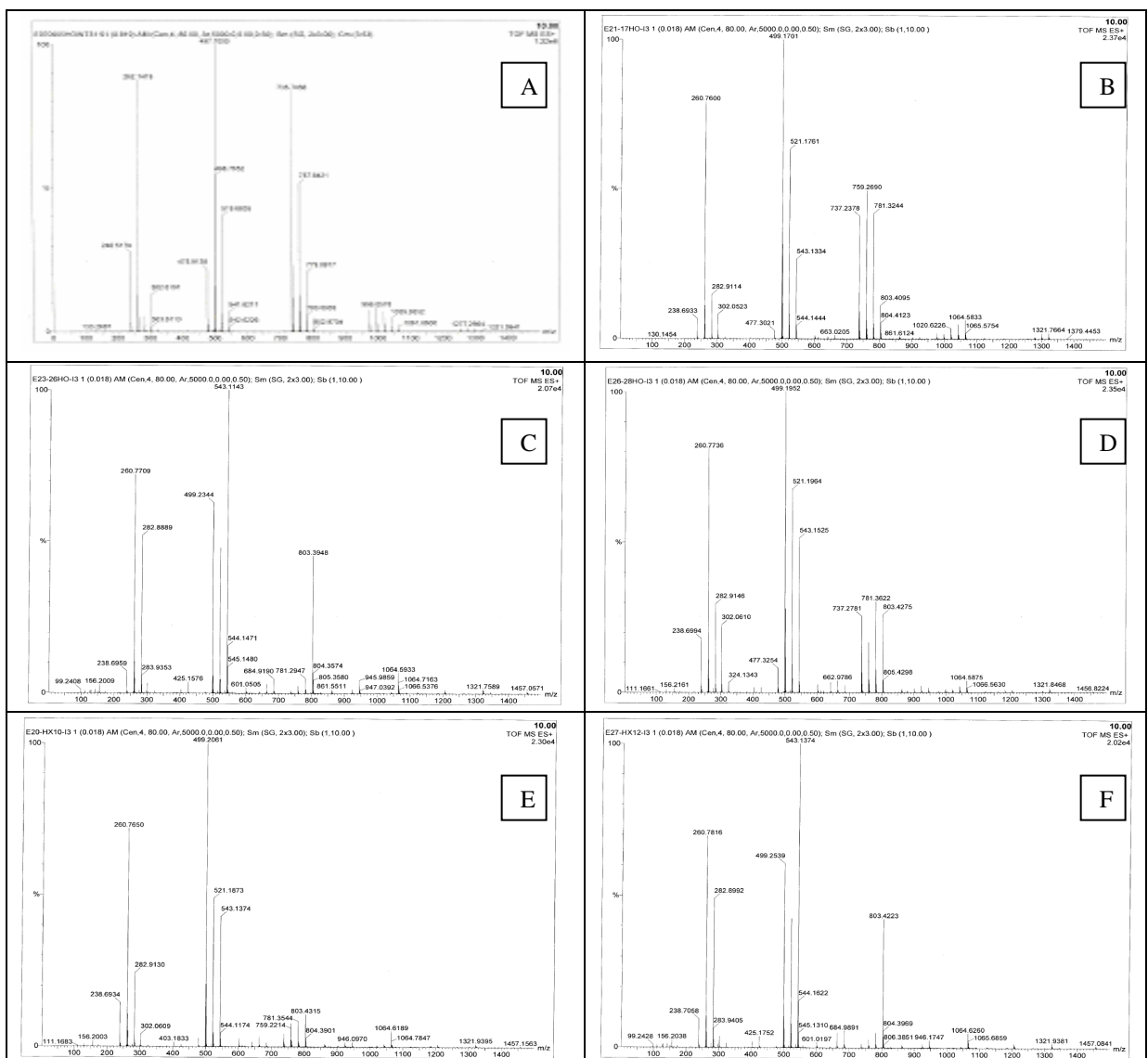


Figure 4.38. MS data of intracellular metabolites belonging to 17 mutants at 20th hour of fermentation (A: *ho12*, B: *ho17*, C: *ho26*, D: *ho28*, E: *hxt10*, F: *hxt12*)

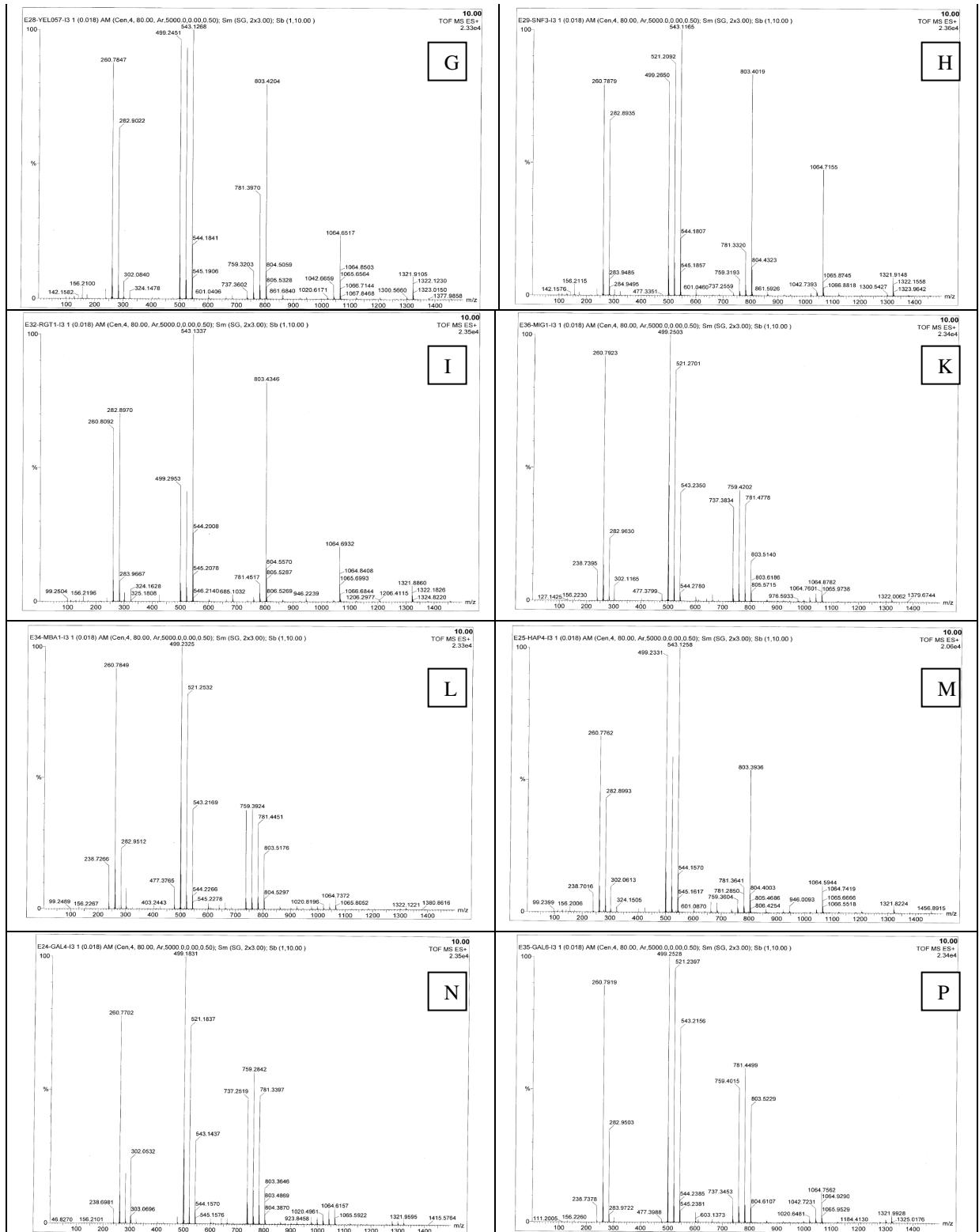


Figure 4.38. continued (G: *yel057c*, H: *snf3*, I: *rgt1*, K: *mig1*, L: *mba1*, M: *hap4*, N: *gal4*, P: *gal6*)

The MS profiles of dilution solution (acetonitrile:water:formic acid), hot ethanol and methanol solution, used for extraction of intracellular metabolites are given in Figures 4.39-4.41.

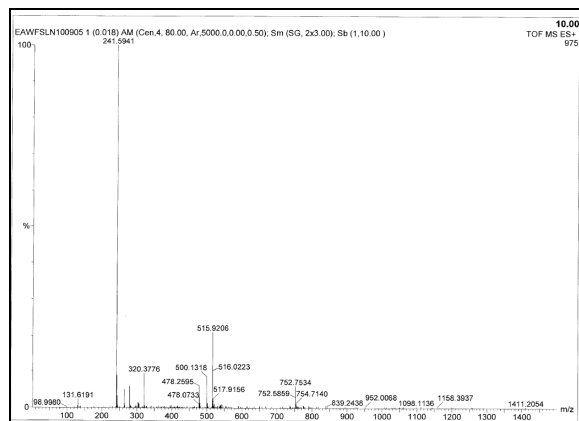


Figure 4.39. MS of extraction solution(acetonitrile:water:formic acid)

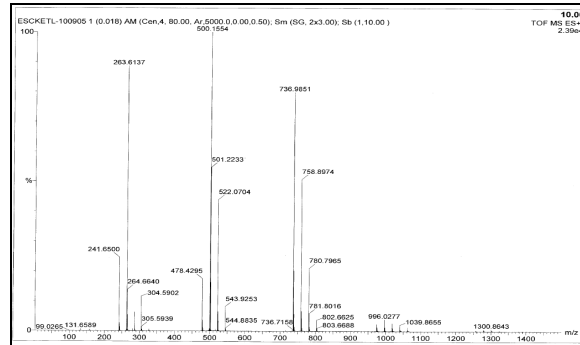


Figure 4.40. MS of hot ethanol solution for extraction of intracellular metabolites

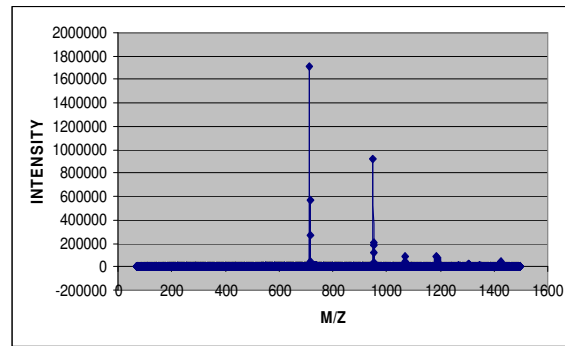


Figure 4.41. MS of methanol solution for extraction of intracellular metabolites

4.2.1. PCA and ICA of Binned Intracellular Metabolome Data at 20th hour

In order to see the distribution of the samples according to latent values, the scores for the intracellular metabolome data on first four LVs were plotted against each other in Figure 4.42. In Figure 4.42 (a); cluster 1 includes *gal4Δ*, *gal7Δ*, *gal80Δ*, *ho17Δ*, *mba1Δ*, *mig1Δ*; cluster 2 includes *adh6Δ*, *gal6Δ*, *hap4Δ*, *ho26Δ*, *ho28Δ*, *hxt12Δ*, *rgt1Δ*, *snf3Δ*, *yel057cΔ*. In Figure 4.42 (b), cluster 1 includes *gal4Δ*, *gal7Δ*, *ho17Δ*, *mba1Δ*, *mig1Δ*; cluster 2 includes *adh6Δ*, *gal6Δ*, *hap4Δ*, *ho26Δ*, *ho28Δ*, *hxt12Δ*, *rgt1Δ*, *snf3Δ*, *yel057cΔ*. In Figure 4.42 (c), cluster 1 includes *adh6Δ*, *snf3Δ*, *ho26Δ*; cluster 2 includes *gal4Δ*, *gal7Δ*, *gal80Δ*, *ho17Δ*, *mba1Δ*, *mig1Δ*. In Figure 4.42 (d), cluster 1 includes *adh6Δ*, *snf3Δ*, *ho26Δ*; cluster 2 includes *gal4Δ*, *gal7Δ*, *gal80Δ*, *ho17Δ*, *mba1Δ*, *mig1Δ*. Distribution of the scores for the PC1 through PC4 are rather unstable. Besides, the distribution of scores in Figures 4.42 (c) and (d) are close to axis, indicating a weakness of the prediction power of the PCA for the variation of the metabolic data represented.

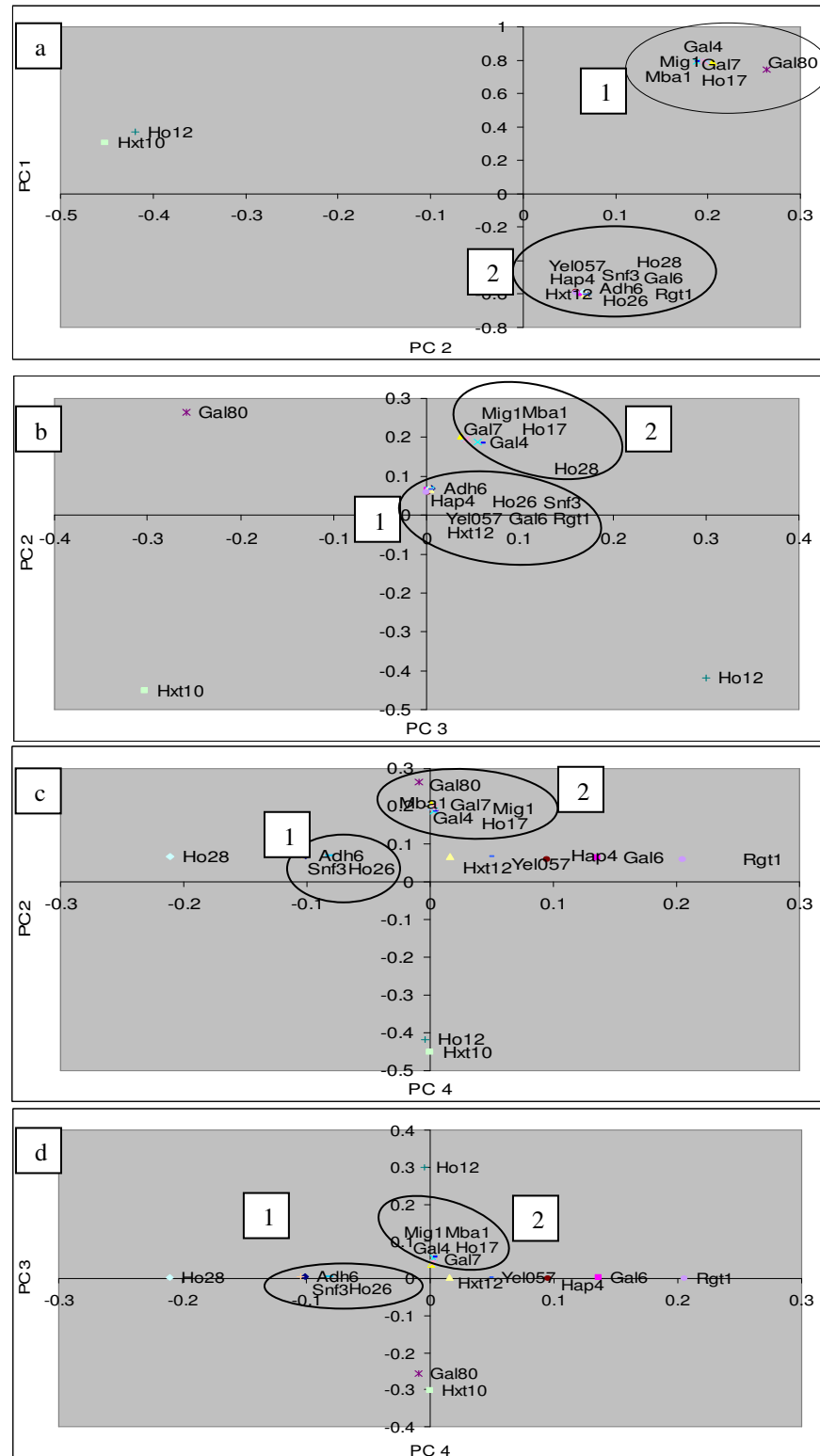


Figure 4.42. Scores of binned metabolome data on first four LVs at 20th hour of intracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

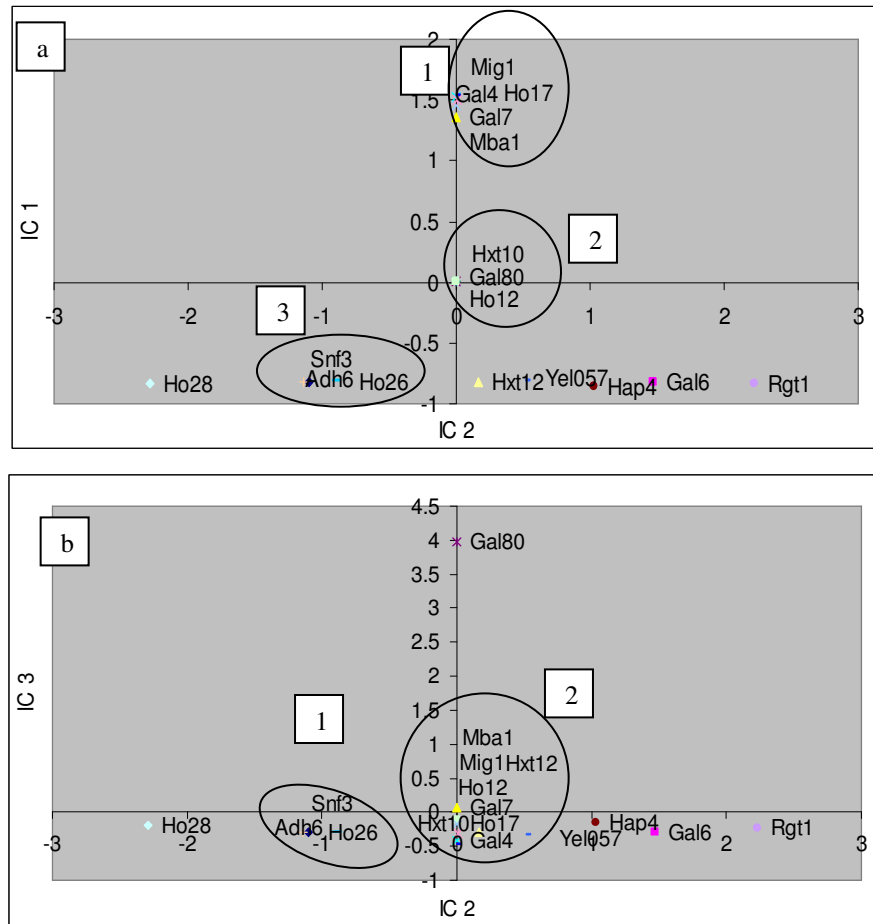


Figure 4.43. Scores of binned metabolome data on first three ICs at 20th hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

ICs of metabolite data of intracellular samples at 20th hour of fermentation give no information about the clustering of deletion mutants due to hidden data corresponding to PCA. In Figure 4.43 (a), cluster 1 includes *gal4* Δ , *gal7* Δ , *ho17* Δ , *mba1* Δ , *mig1* Δ ; cluster 2 includes *gal80* Δ , *ho12* Δ , *hxt10* Δ ; cluster 3 includes *adh6* Δ , *snf3* Δ , *ho26* Δ . In Figure 4.43 (c), cluster 1 includes *adh6* Δ , *snf3* Δ , *ho26* Δ ; cluster 2 includes *gal4* Δ , *gal7* Δ , *ho12* Δ , *ho17* Δ , *hxt10* Δ , *hxt12* Δ , *mba1* Δ , *mig1* Δ . When PCA preprocessing is increased to 6, there is only slight differences and a proper clustering information cannot be obtained (Figure 4.44). In Figure 4.44 (a), cluster 1 includes *gal4* Δ , *gal7* Δ , *ho17* Δ , *mba1* Δ , *mig1* Δ ; cluster 2 includes *gal80* Δ , *hap4* Δ , *ho12* Δ , *hxt10* Δ ; cluster 3 includes *adh6* Δ , *snf3* Δ , *ho26* Δ . In Figure 4.44 (b), cluster 1 includes *adh6* Δ , *snf3* Δ , *ho26* Δ ; cluster 2 includes *gal4* Δ , *gal7* Δ , *gal80* Δ , *ho12* Δ , *ho17* Δ , *mba1* Δ , *mig1* Δ .

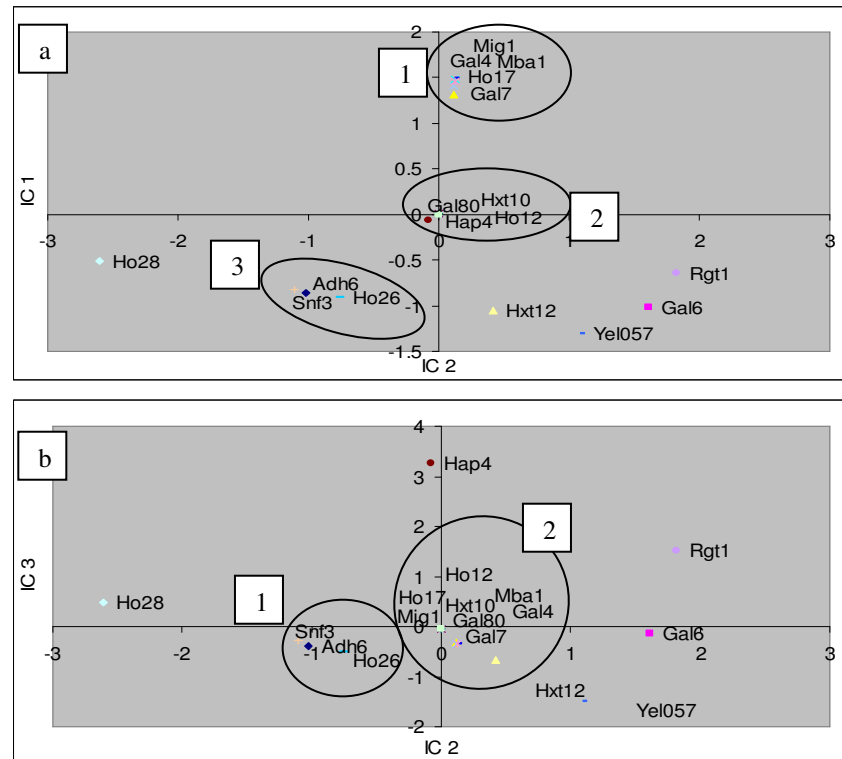


Figure 4.44. Scores of binned metabolome data on first three ICs at 20th hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

4.2.2. PCA and ICA of Filtered Intracellular Metabolome Data at 20th hour

PCA was applied to column vector normalized data of 94 out of 1000 mass variables (Figure 4.45). In Figure 4.45 (a) and (b); cluster 1 includes *gal80Δ*, *mig1Δ*, *mba1Δ*, *ho17Δ*, *gal4Δ*. In Figure 4.45 (a); cluster 2 includes *adh6Δ*, *gal6Δ*, *ho12Δ*, *ho26Δ*, *ho28Δ*, *hap4Δ*, *rgt1Δ*, *yel057cΔ*, *snf3Δ*, *hxt12Δ*. In Figure 4.45 (b); cluster 2 includes *rgt1Δ*, *gal6Δ*, *yel057cΔ*, *hxt12Δ*; cluster 3 includes *ho12Δ*, *ho26Δ*, *adh6Δ*, *snf3Δ*. In Figure 4.45 (c); cluster 1 includes *adh6Δ*, *gal4Δ*, *gal80Δ*, *ho12Δ*, *ho17Δ*, *ho26Δ*, *ho28Δ*, *mba1Δ*, *mig1Δ*, *snf3Δ*. Due to elimination of some useful data in filtering approach, the correlation between the mutants is not quite clear or it could be hidden. Another reason for the bad clustering could be outliers. For example, *hxt10Δ* and *gal7Δ* seem to be outliers in Figures 4.45 (a), 4.45 (b) and 4.45 (c). To check out whether these deletion mutants are outliers or not, statistical tools should be used. If they appear as outlier in any statistical test, they may affect the discrimination of other deletion mutants and should be omitted.

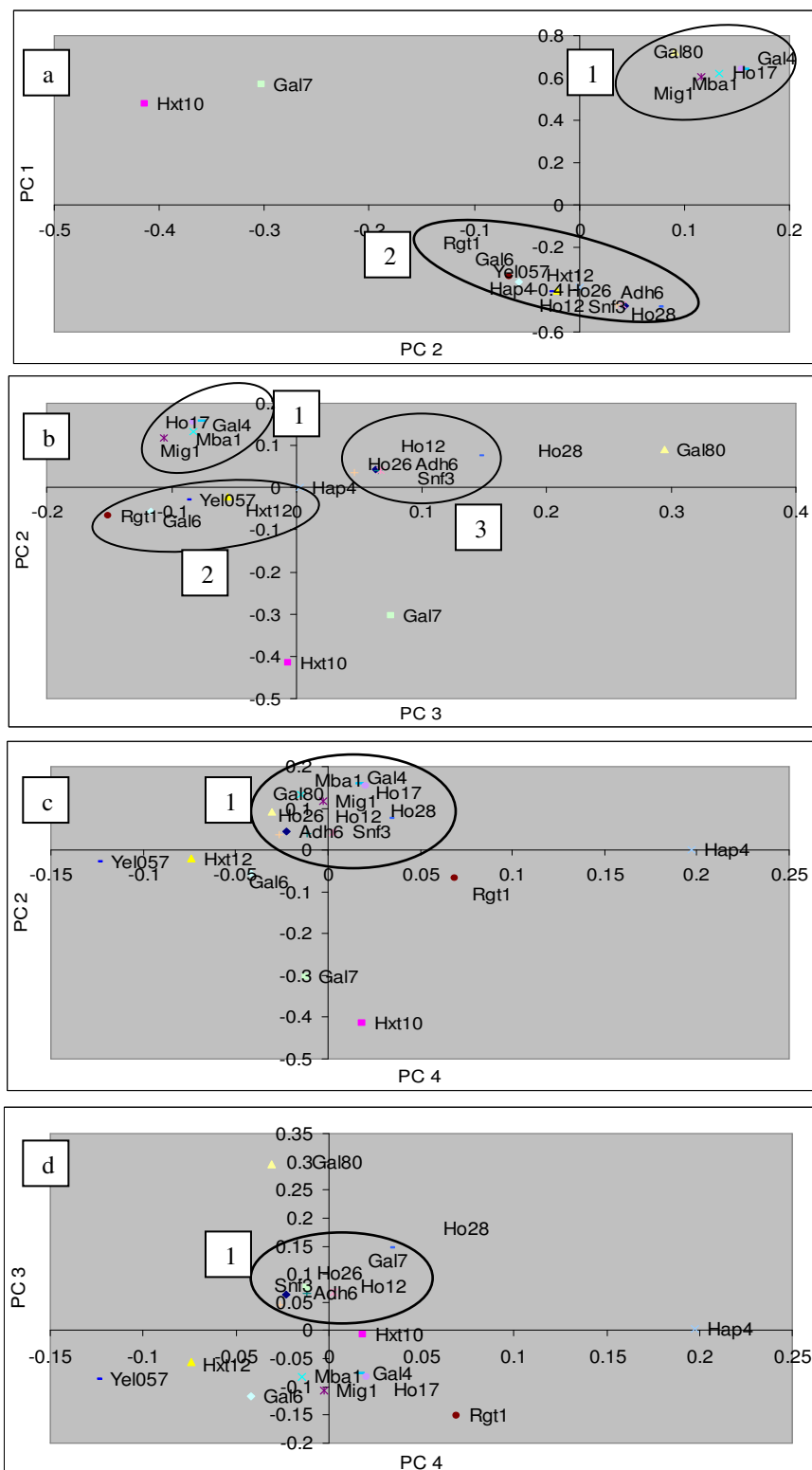


Figure 4.45. Scores of filtered metabolome data on first 4 LVs at 20th hour of intracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

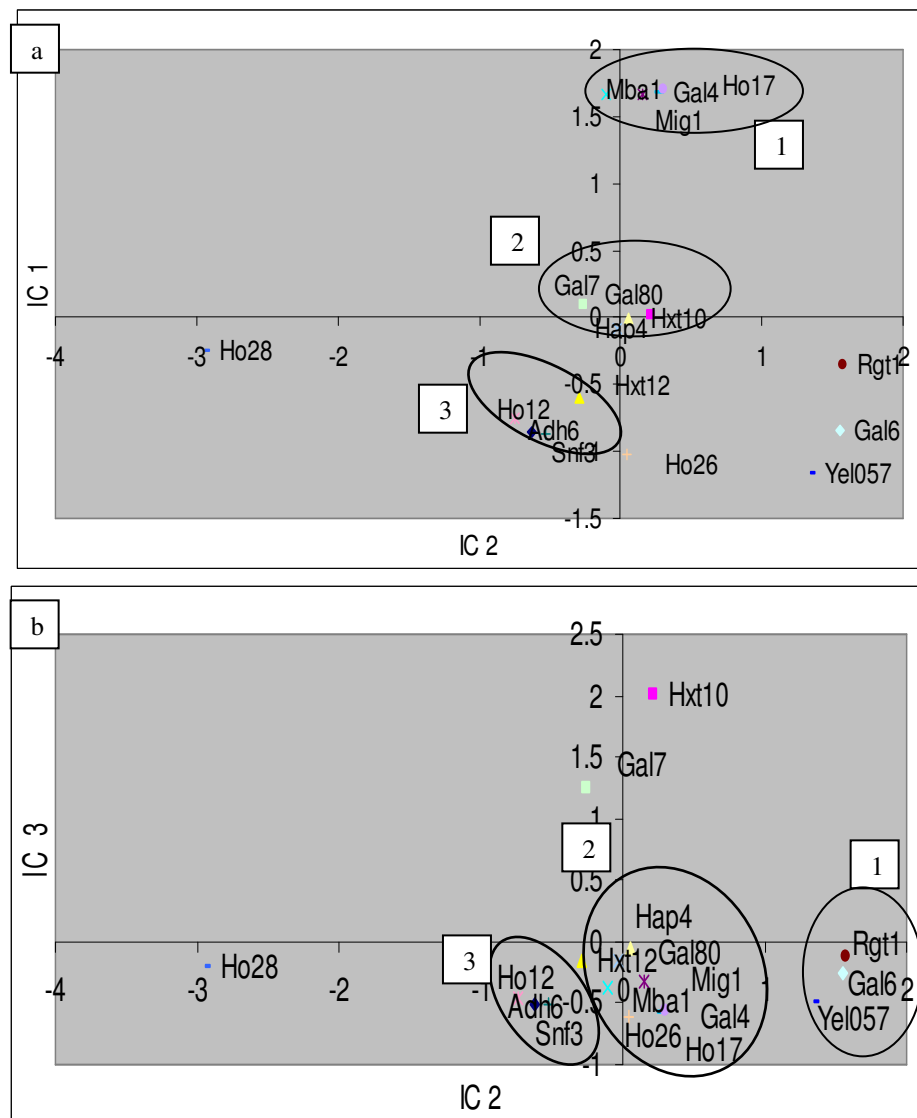


Figure 4.46. Scores of filtered metabolome data on first three ICs at 20th hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

There is only one negative IC out of 5 after preprocessing PC data by 5 times. As mentioned before, the more negative kurtosis evaluated, the better clustering could be obtained. In Figure 4.46 (a); cluster 1 includes *mba1*, *mig1*, *gal4*, *ho17*; cluster 2 includes *gal7*, *gal80*, *hap4*, *hxt10* and cluster 3 includes *adh6*, *snf3*, *ho12*. In Figure 4.46 (b); cluster 1 includes *rgt1*, *gal6*, *yel057c*; cluster 2 includes *gal4*, *gal80*, *hap4*, *ho17*, *ho26*, *hxt12*, *mba*, *mig1*; cluster 3 includes *adh6*, *ho12*, *snf3*. According to results shown in Figure 4.46 (a) and (b), filtering process is not helpful to identify the possible clusters.

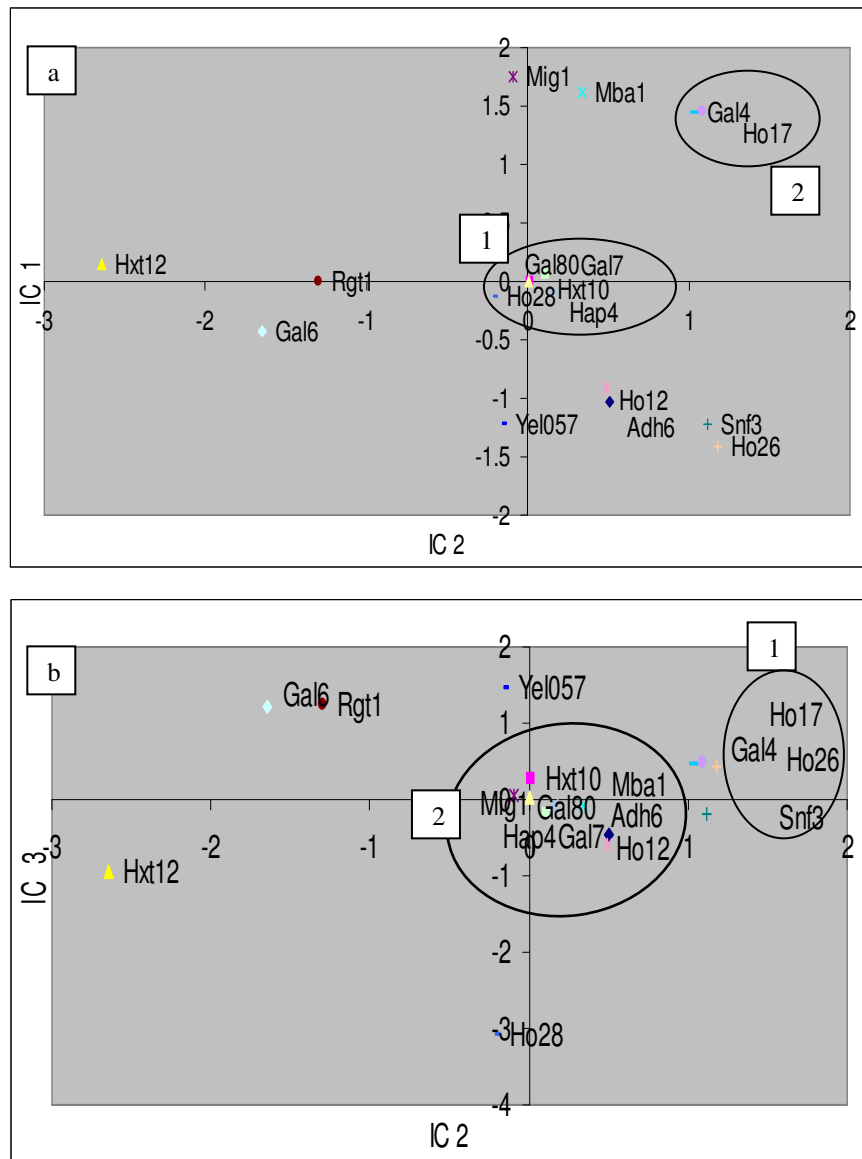


Figure 4.47. Scores of filtered metabolome data on first three ICs at 20th hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

There is again one negative IC out of 6 after preprocessing PC data by 6 times. The discrimination is not possible in Figures 4.47 (a) and (b). In Figure 4.47 (a); cluster 1 includes *hxt10*, *gal7*, *gal80*, *hap4*, *ho28*; cluster 2 includes *gal4*, *ho17*. In Figure 4.47 (b); cluster 1 includes *gal4*, *ho17*, *ho26*; cluster 2 includes *adh6*, *ho12*, *gal7*, *gal80*, *hxt10*, *mba1*, *mig1*. If the outlier problem in PCA is handled, the power of ICA to reveal the hidden relationships among the mutants may increase.

4.2.3. PCA and ICA of Binned Intracellular Metabolome Data at 25th hour

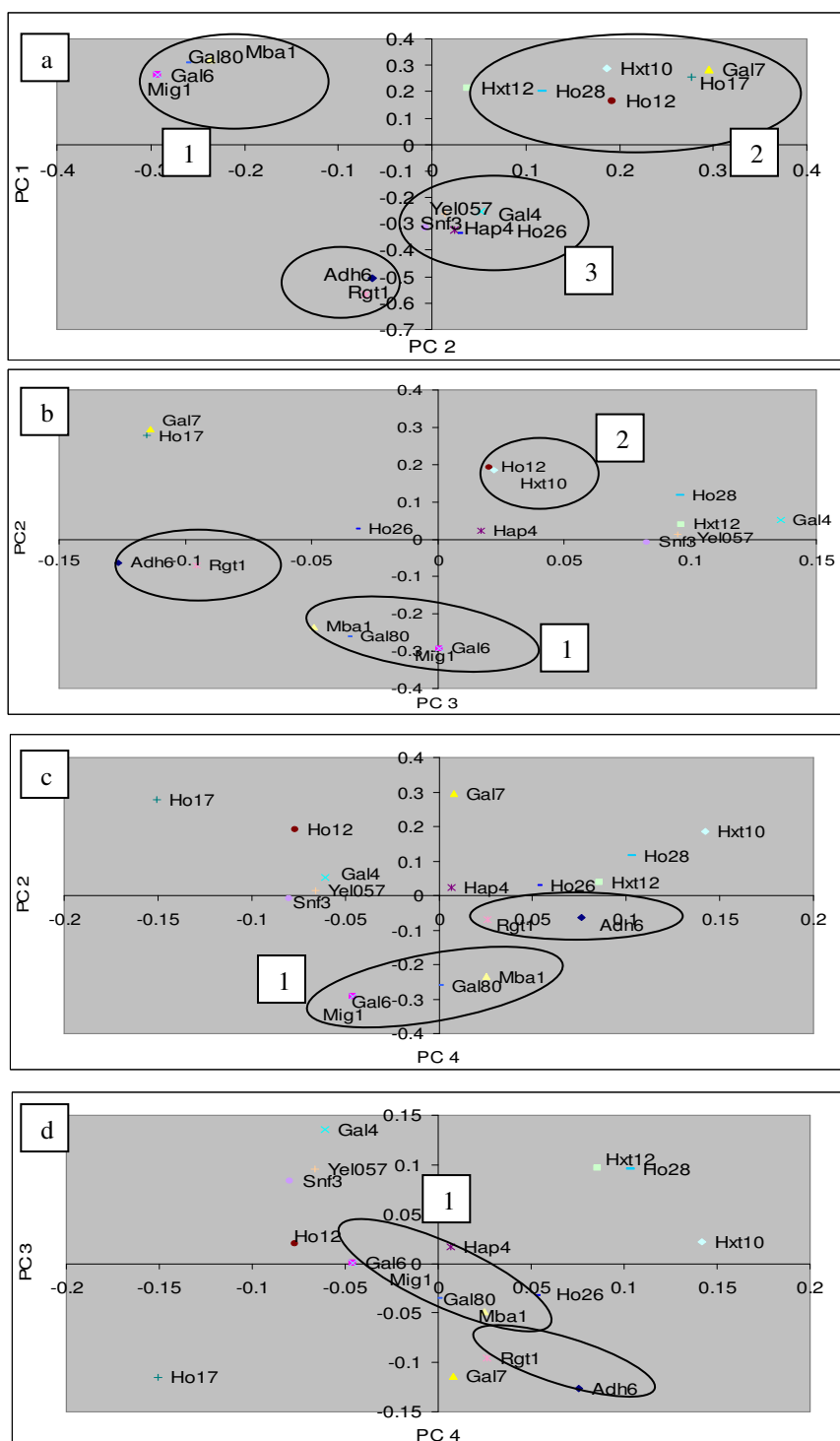


Figure 4.48. Scores of binned metabolome data on first four LVs at 25th hour of intracellular samples, a) PC1 vs PC2, b) PC2 vs PC3, c) PC2 vs PC4, d) PC3 vs PC4

PCA results of intracellular samples at 25th hour of fermentation is given in Figure 4.48. Cluster 1 includes *gal6Δ*, *gal80Δ*, *mig1Δ*, *mba1Δ* in all parts of Figure 4.48. In Figure 4.48 (a); cluster 2 includes *gal7Δ*, *ho12Δ*, *ho17Δ*, *ho28Δ*, *hxt10Δ*, *hxt12Δ*; cluster 3 includes *gal4Δ*, *hap4Δ*, *snf3Δ*, *ho26Δ*, *yel057cΔ*. In Figure 4.48 (b); cluster 2 includes *ho12Δ*, *hxt10Δ*. *HoΔ* reference strains and hexose transport members such as *hxt10Δ* and *hxt12Δ* have close scores on LV1 and LV2 (Figure 4.48 (a)). The results of GO term analysis have also confirmed that *adh6Δ-rgt1Δ* and *gal6Δ-gal80Δ-mig1Δ* deletion pairs remain closer in all conditions. While the *p* value of *adh6Δ-rgt1Δ* is 1.75e-05, the *p* value of *gal6Δ-gal80Δ-mig1Δ* is 0.00046. As *p* values are lower than 0.01 in both pairs, it could be verified that *adh6Δ-rgt1Δ* are member of alcohol metabolic process and *gal6Δ-gal80Δ-mig1Δ* have an activity in transcription regulation.

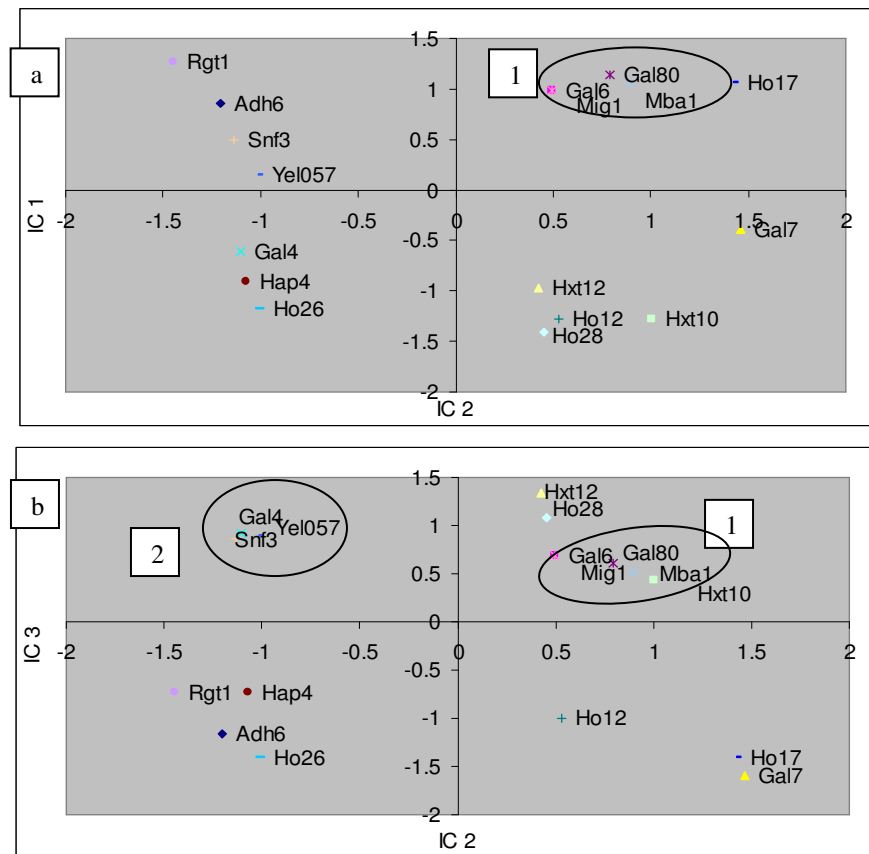


Figure 4.49. Scores of binned metabolome data on first three ICs at 25th hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

The first three ICs do not distinguish the gene deletion strains effectively compared to PCA. In Figure 4.49 (a) and (b); cluster 1 includes *gal6Δ*, *gal80Δ*, *mba1Δ*, *mig1Δ*. In Figure 4.49 (b); cluster 2 includes *snf3Δ*, *gal4Δ*, *yel057cΔ*. The gene deletion mutants scattered randomly in both figures and only one cluster, *gal80Δ-gal6Δ-mba1Δ-mig1Δ*, was obtained.

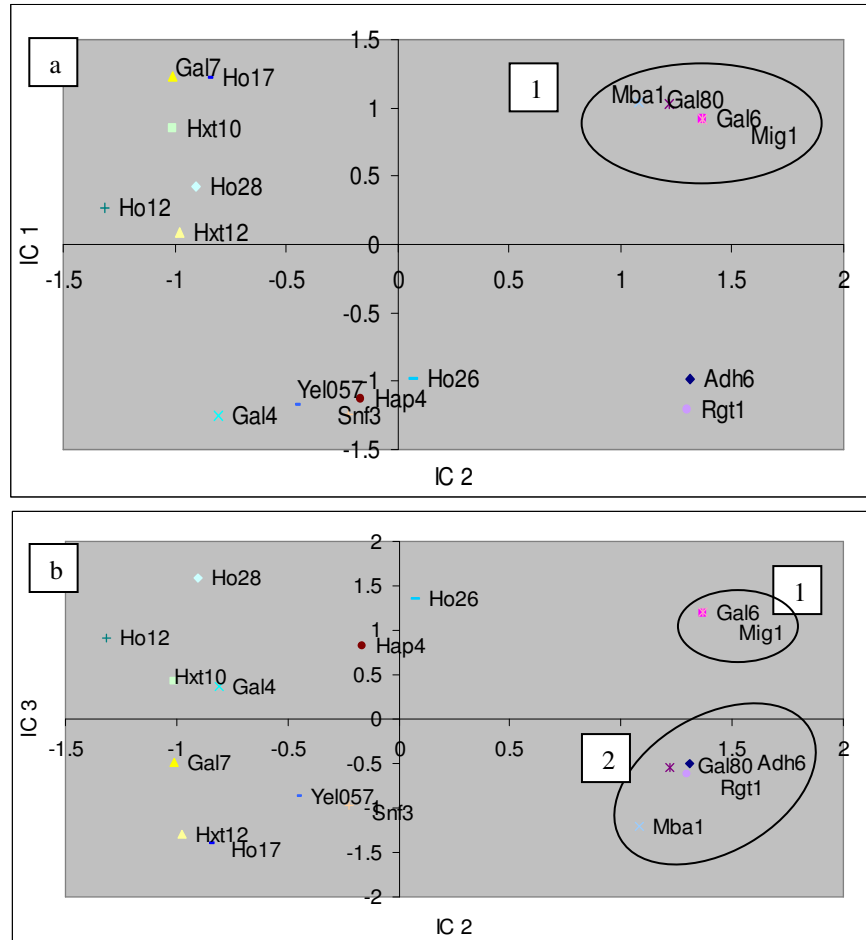


Figure 4.50. Scores of binned metabolome data on first three ICs at 25th hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

ICA results of intracellular samples at 25th hour preprocessed by PCA by six times are presented in Figure 4.50. In Figure 4.50 (a); cluster 1 includes *gal6Δ*, *gal80Δ*, *mba1Δ*, *mig1Δ*. In Figure 4.50 (b); cluster 1 includes *gal6Δ*, *mig1Δ*; cluster 2 includes *adh6Δ*, *gal80Δ*, *mba1Δ*, *rgt1Δ*. The cluster profiles of *gal80Δ-gal6Δ-mba1Δ-mig1Δ* mutants were approximately the same except the IC3 values attaining negative values that were encircled in Figure 4.50.

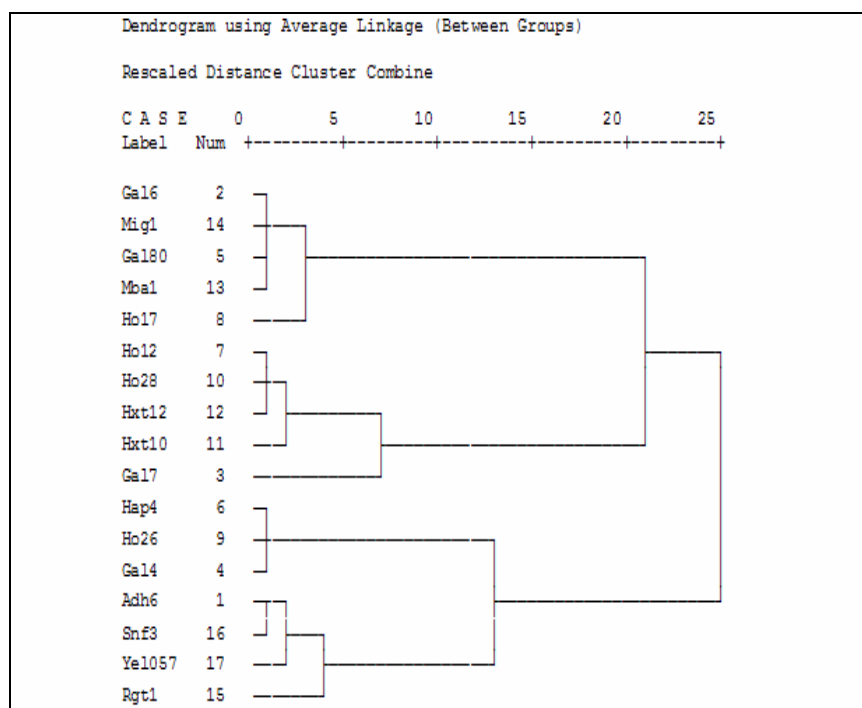


Figure 4.51. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Hierarchical cluster analysis of intracellular m/z signatures was performed to graphically depict how the deletion strains relate to each other. HCA of two ICs of binned intracellular metabolome data at 25th hour of fermentation is given in Figure 4.51. In Figure 4.51; first cluster includes *gal6Δ-gal80Δ-mig1Δ-mba1Δ* and the bottom cluster includes *adh6Δ-snf3Δ-yel057cΔ-rgt1Δ*. HCA results are in agreement with the ICA results.

4.2.4. PCA and ICA of Filtered Intracellular Metabolome Data at 25th hour

PCA was applied to column vector normalized data of 137 out of 1000 mass variables in Figure 4.52. The number of common mass variables is greater than that of the same mutants (In the previous sample set at 20th hour, the common mass variables were only 94). In Figure 4.52 (a), (b), (c) and (d); cluster 1 includes *gal7Δ-hxt10Δ-hxt12Δ*. Cluster 2 includes *gal6Δ-gal80Δ-mig1Δ-mba1Δ*; cluster 3 includes *gal4Δ-yel057cΔ-snf3Δ* in Figure 4.52 (a) and (c). In Figure 4.52 (b) and (d), cluster 2 includes *gal6Δ* and *mig1Δ* mutants. There is no clear discrimination of mutants by PCA.

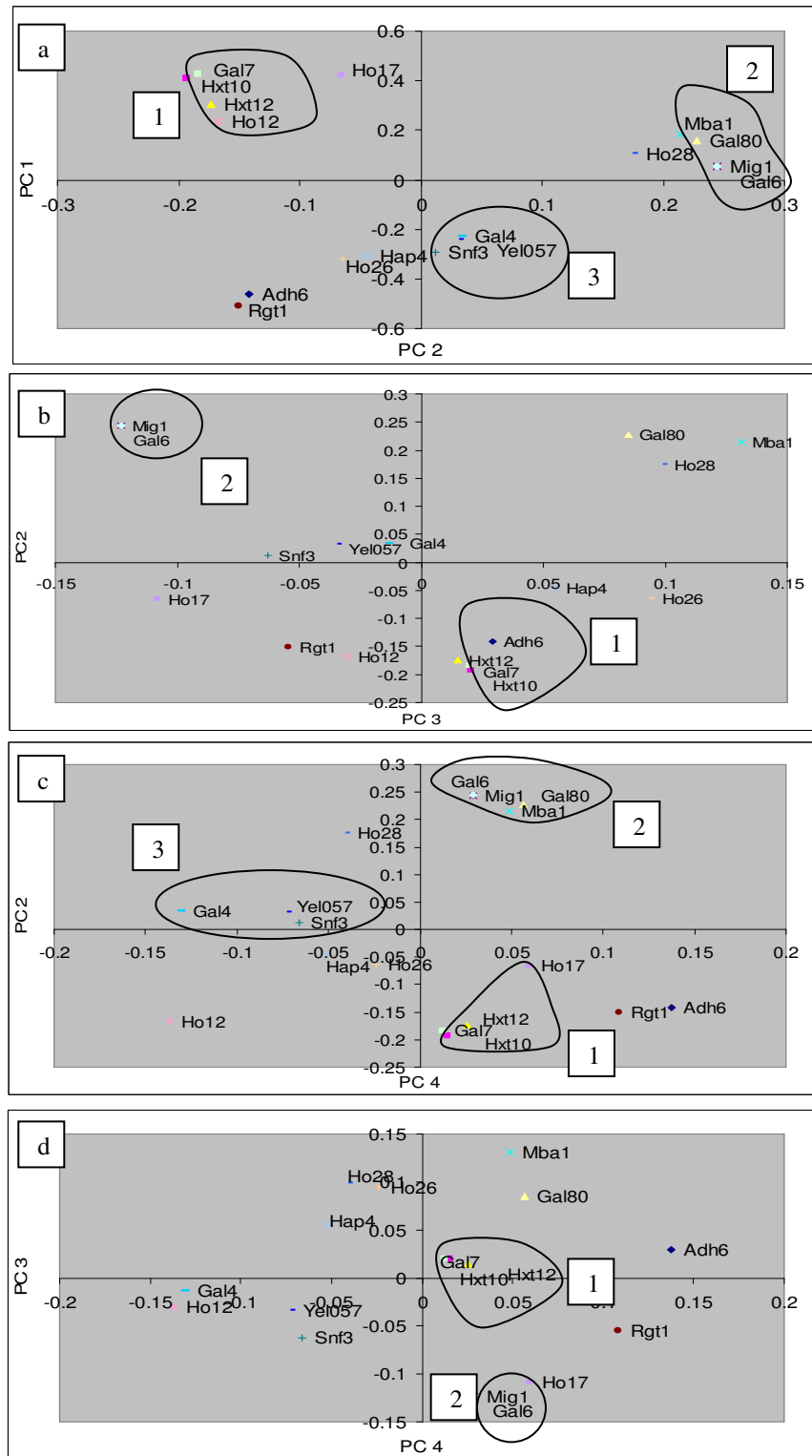


Figure 4.52. Scores of filtered metabolome data on first four LVs at 25th hour of intracellular samples

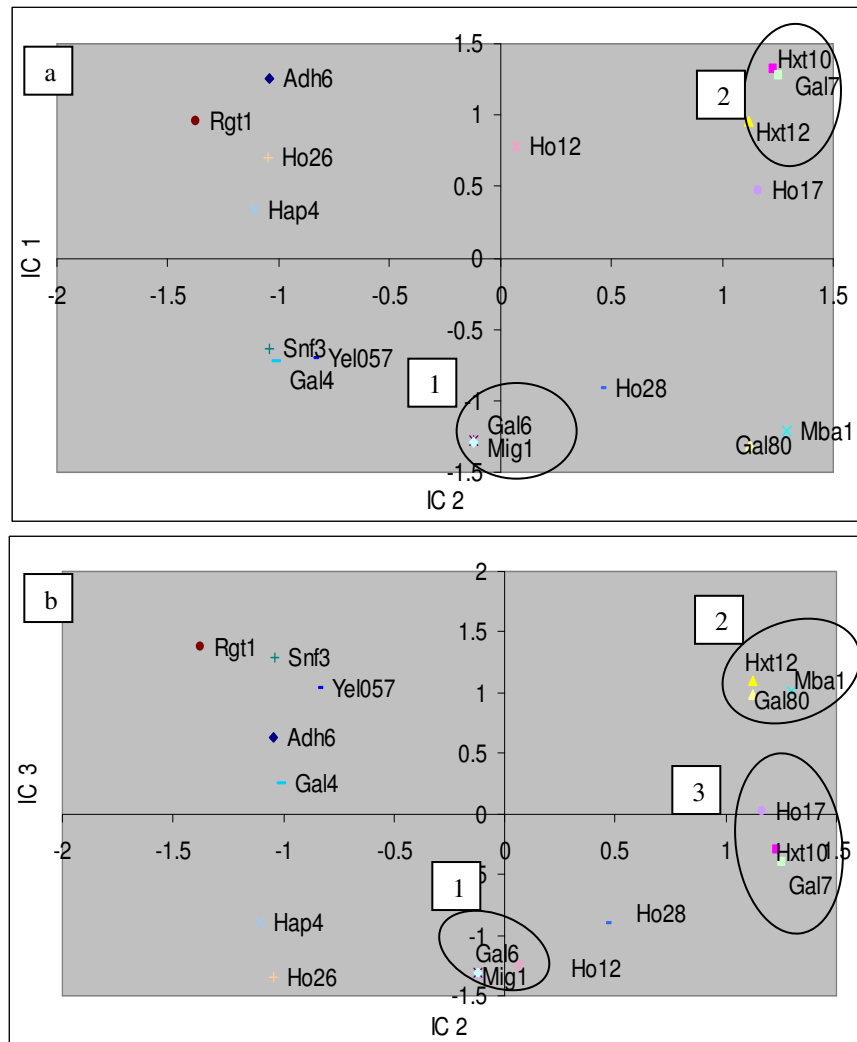


Figure 4.53. Scores of filtered metabolome data on first three ICs at 25th hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

There are 5 negative ICs out of 5 after preprocessing PC data by 5 times. Cluster 1 includes *gal6Δ* and *mig1Δ* in Figures 4.53 (a) and (b). In Figure 4.53 (a), cluster 2 includes *hxt10Δ*, *hxt12Δ* and *gal7Δ*. In Figure 4.53 (b), cluster 2 includes *gal80Δ*, *hxt12Δ*, *mba1Δ*; cluster 3 includes *gal7Δ*, *ho17Δ* and *hxt10Δ*. There is again no clear discrimination of mutants.

In Figure 4.54 (a) and (b); cluster 1 includes *gal6Δ-mig1Δ*. In Figure 4.54 (b); cluster 2 includes *gal7Δ-hxt10Δ*. There is no satisfactory separation among the mutants even though the dispersion of the ICs is better than PCs.

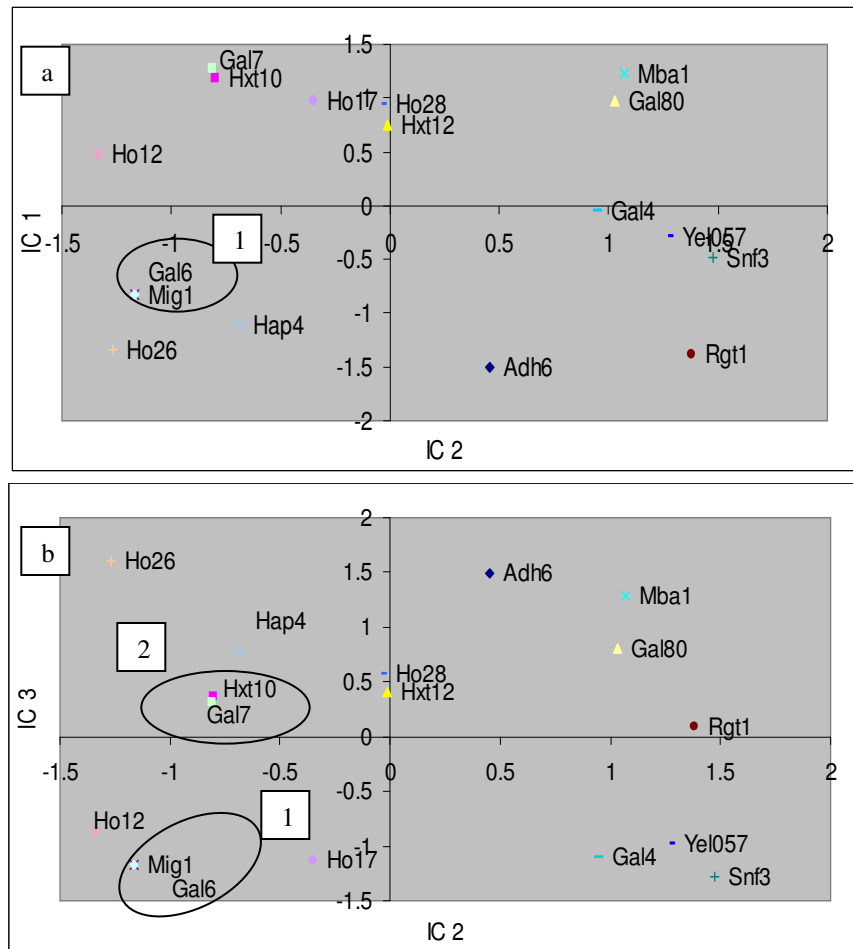


Figure 4.54. Scores of filtered metabolome data on first three ICs at 25th hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

4.2.5. PCA and ICA of Binned Intracellular Metabolome Data at 31st hour

In Figure 4.55 (a), cluster 1 includes *hap4Δ*, *ho28Δ*; cluster 2 includes *ho12Δ*, *ho26Δ*, *gal7Δ*, *snf3Δ*, *yel057cΔ*, cluster 3 includes *adh6Δ*, *mba1Δ*, *mig1Δ*, *gal6Δ*, *gal80Δ*, *rgt1Δ*. In Figure 4.55 (b), cluster 1 includes *adh6Δ*, *mba1Δ*, *hap4Δ*, *mig1Δ*, *gal6Δ*, *gal4Δ*, *gal80Δ*, *rgt1Δ*. PC2 discriminates the regulation related deletion mutants. The only satisfactory cluster contains *hap4Δ*, *mba1Δ*, *mig1Δ*, *gal6Δ*, *gal80Δ*, *gal4Δ*, *rgt1Δ*, *adh6Δ* (Figure 4.55 (b) and (c)). The *p* value *gal6Δ-gal80Δ-mig1Δ-adh6Δ* is less than 0.01 and GO term analyses confirms the PCA results.

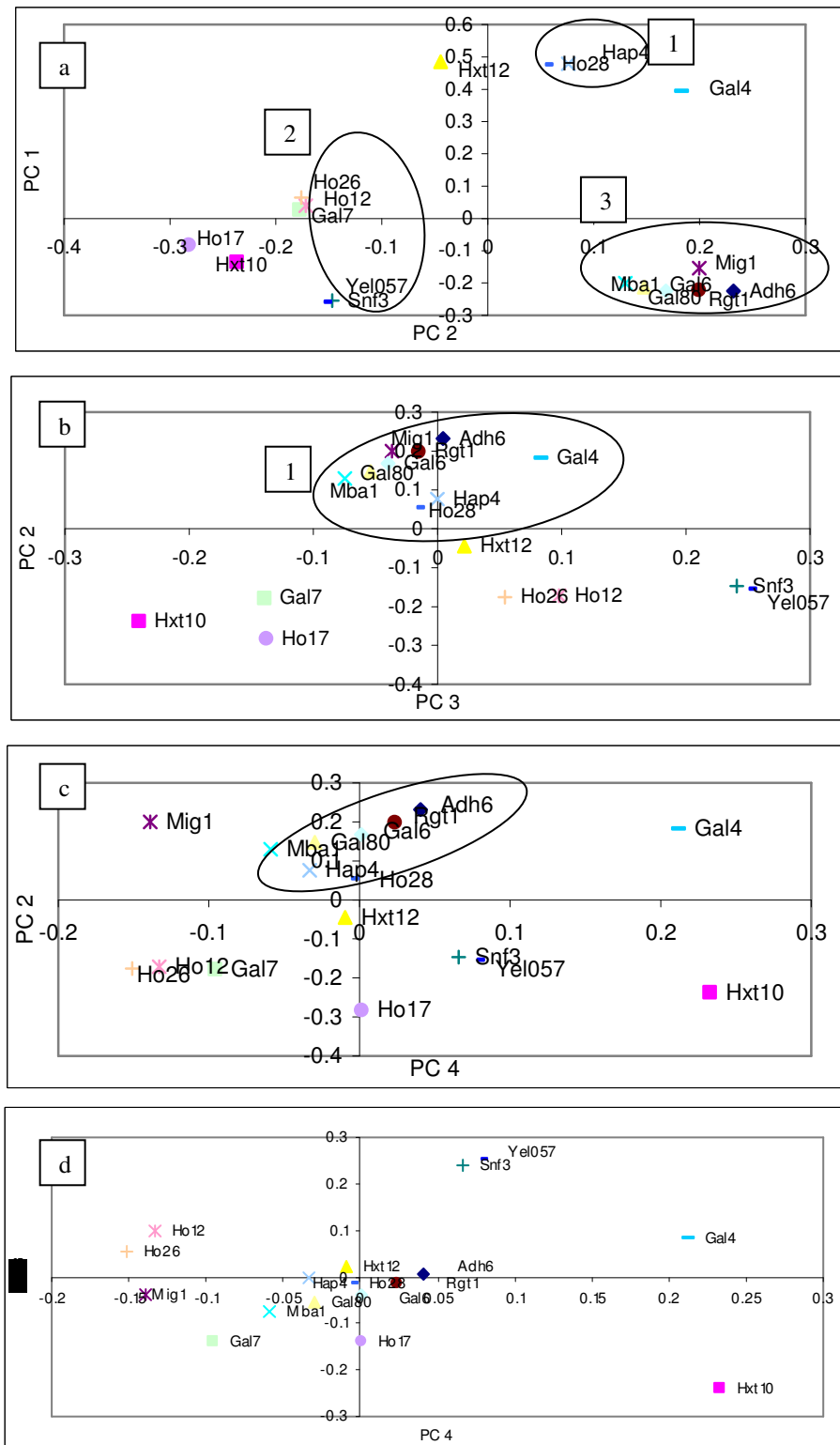


Figure 4.55. Scores of binned metabolome data on first four LVs at 31st hour of intracellular samples

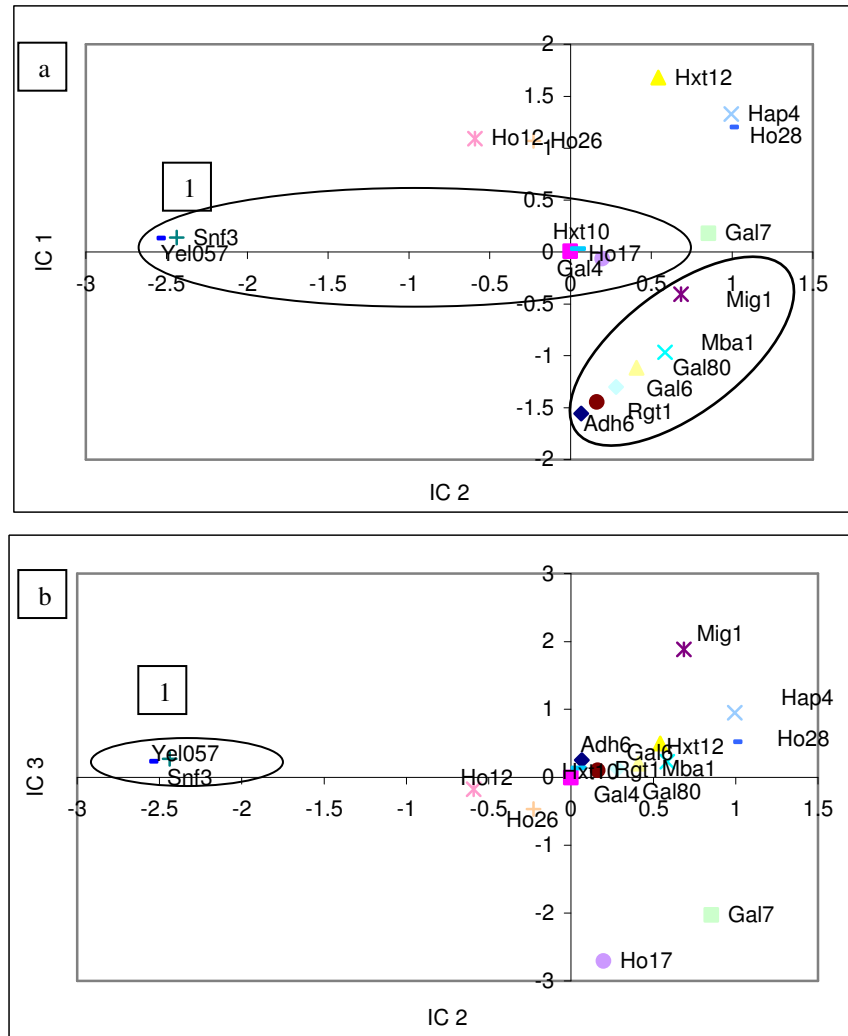


Figure 4.56. Scores of binned metabolome data on first three ICs at 31st hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

ICA results shown above in Figures 4.56 and 4.57 were quite similar. It could be stated that the PCA preprocessing procedure had minor effect on intracellular samples taken at 31th hour.

In Figure 4.56 (a) and 4.57 (a); cluster 1 includes *yel057cΔ*, *snf3Δ*, *gal4Δ*, *hxt10Δ*, *mig1Δ*, *ho17Δ*. In Figure 4.56 (b) and 4.57 (b); cluster 1 includes *yel057cΔ*, *snf3Δ*. IC2 again discriminates the regulation related deletion mutants (*mig1Δ*, *gal80Δ*, *gal6Δ*, *rgt1Δ*, *mba1Δ*, *adh6Δ* in Figure 4.57 (a)).

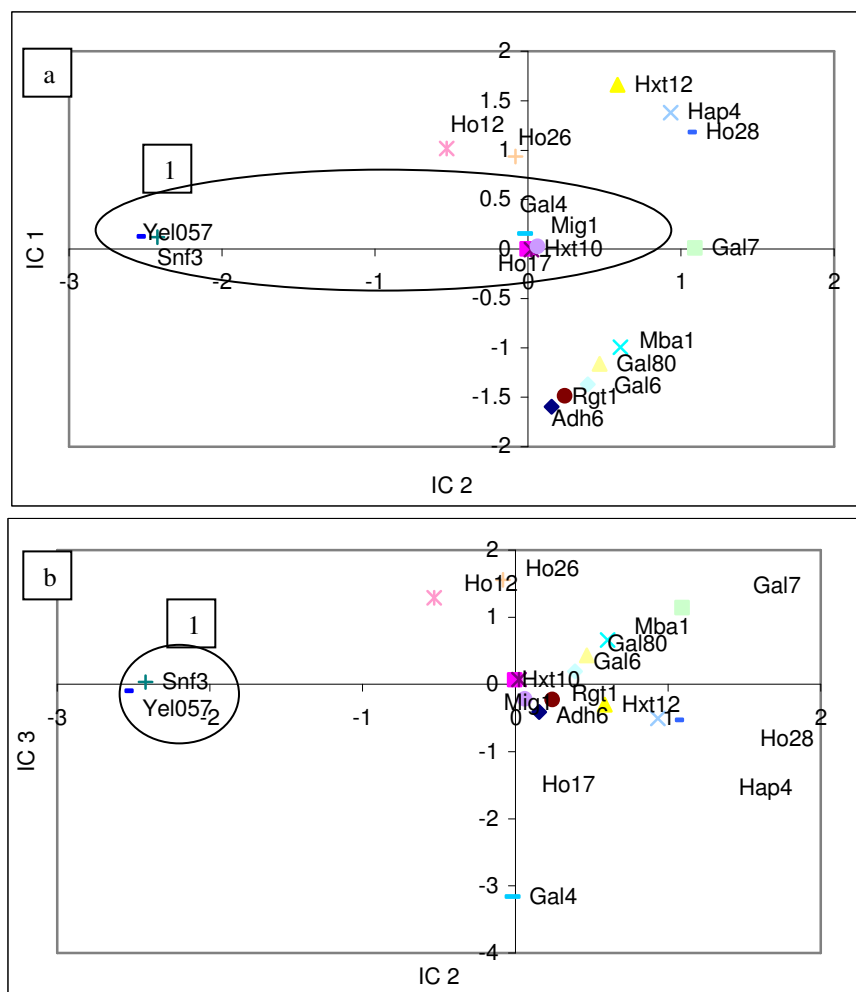


Figure 4.57. Scores of binned metabolome data on first three ICs at 31st hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

4.2.6. PCA and ICA of Filtered Intracellular Metabolome Data at 31st hour

PCA was applied to column vector normalized data of 75 out of 1000 mass variables (Figure 4.58). The common mass variables are quite lower than that obtained by 20th and 25th hour of sampling. Any discrimination could not be observed due to insufficient mass variables. In other words, the possible clusterings may be hidden in the dataset.

In Figure 4.58 (a); cluster 1 includes *adh6Δ*, *snf3Δ*, *yel057cΔ*. In Figure 4.58 (c); cluster 1 includes *rgt1Δ*, *ho26Δ* and in Figure 4.58 (d); cluster 1 includes *adh6Δ*, *ho26Δ*.

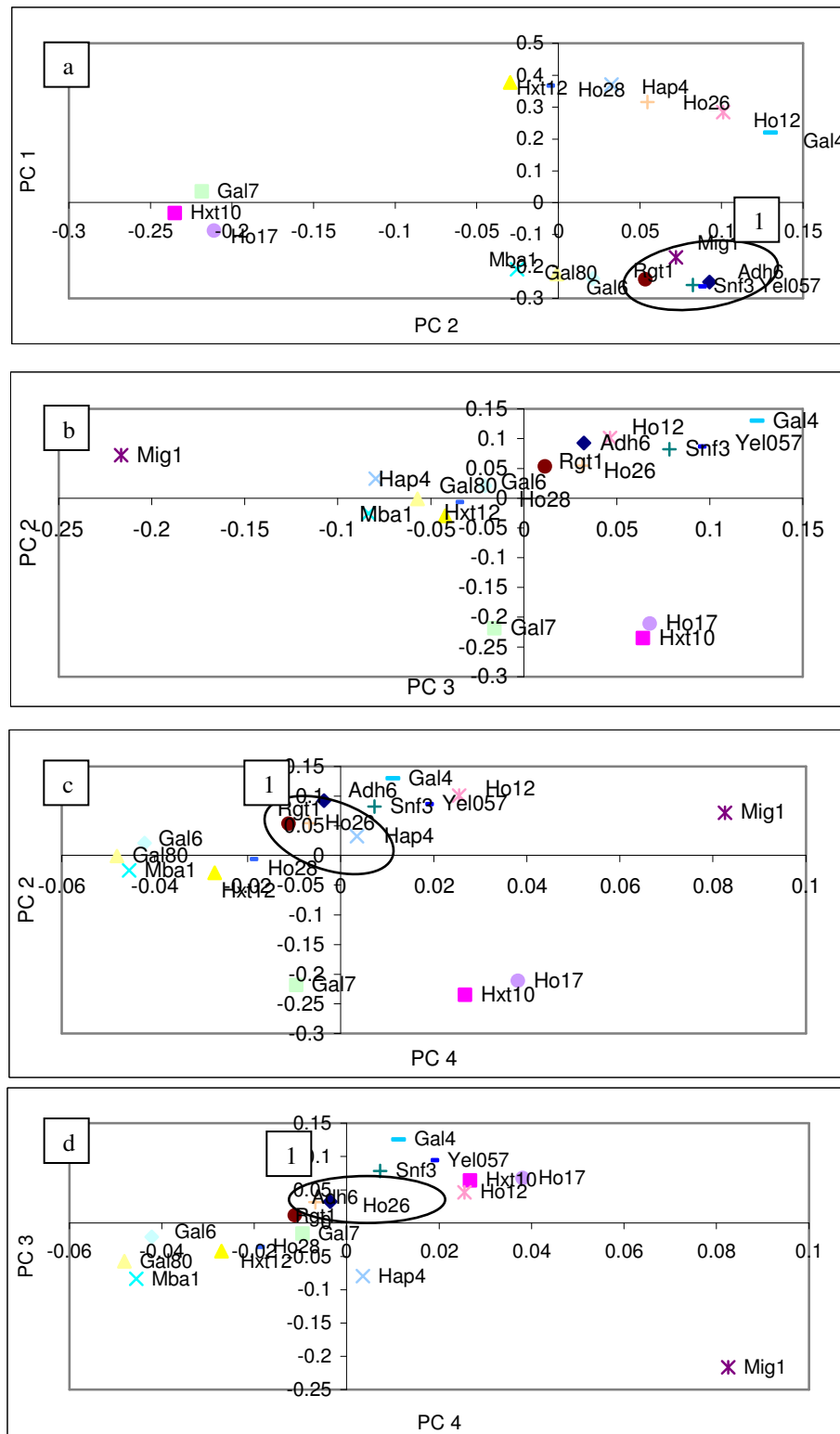


Figure 4.58. Scores of filtered metabolome data on first four LVs at 31st hour of intracellular samples

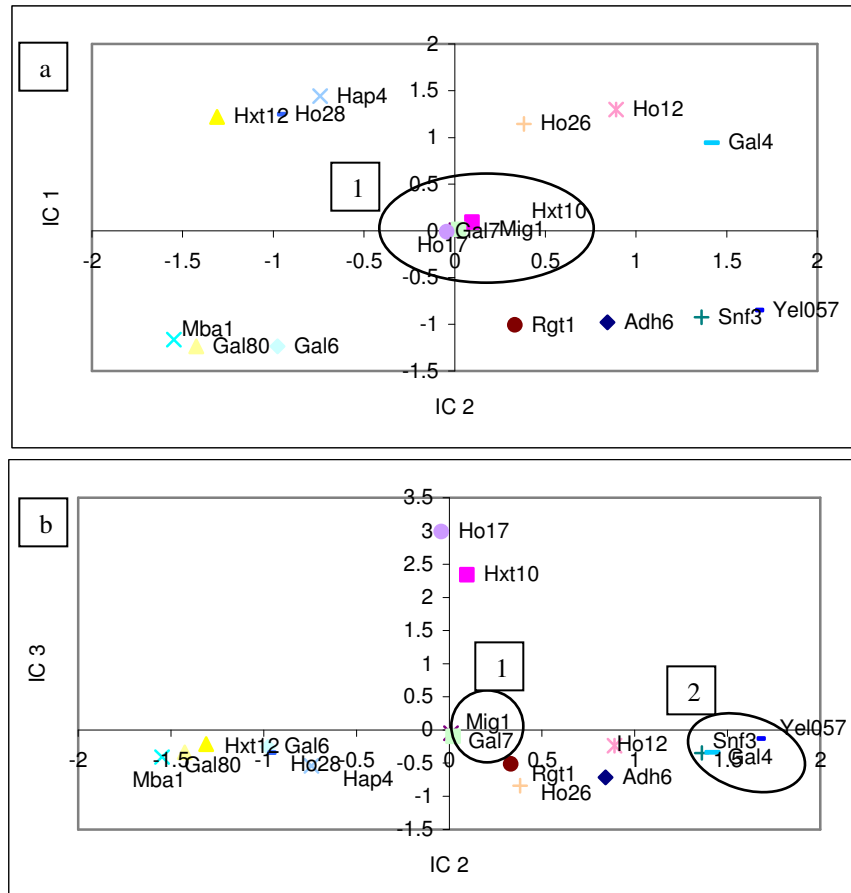


Figure 4.59. Scores of filtered metabolome data on first three ICs at 31st hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

In Figure 4.59 (a); cluster 1 includes *hxt10Δ*, *mig1Δ*, *gal7Δ*, *ho17Δ*, and in Figure 4.59 (b); cluster 1 includes *mig1Δ*, *gal7Δ*; cluster 2 includes *snf3Δ*, *gal4Δ*. The IC results reveal more information than the PC results due to discrimination power. In Figure 4.59 (a); while *rgt1* and *snf3* which have roles in glucose signaling pathway remain in the lower right quadrant together with *yel057cΔ*, the *hoΔ* deletion mutants appear in the upper side of Figure 4.59 (a).

Figure 4.60 (a) gives more appropriate results than those in Figure 4.60 (b). In Figure 4.60 (a), cluster 1 includes *gal4Δ*, *gal7Δ*, *mig1Δ*, and in Figure 4.60 (b); cluster 1 includes *gal4Δ*, *gal7Δ*, *mig1Δ*, *rgt1Δ*. The *hoΔ* deletion group is clearly identified and the biological relevance of *gal6Δ-gal80Δ-mba1Δ* is verified with the ICA results.

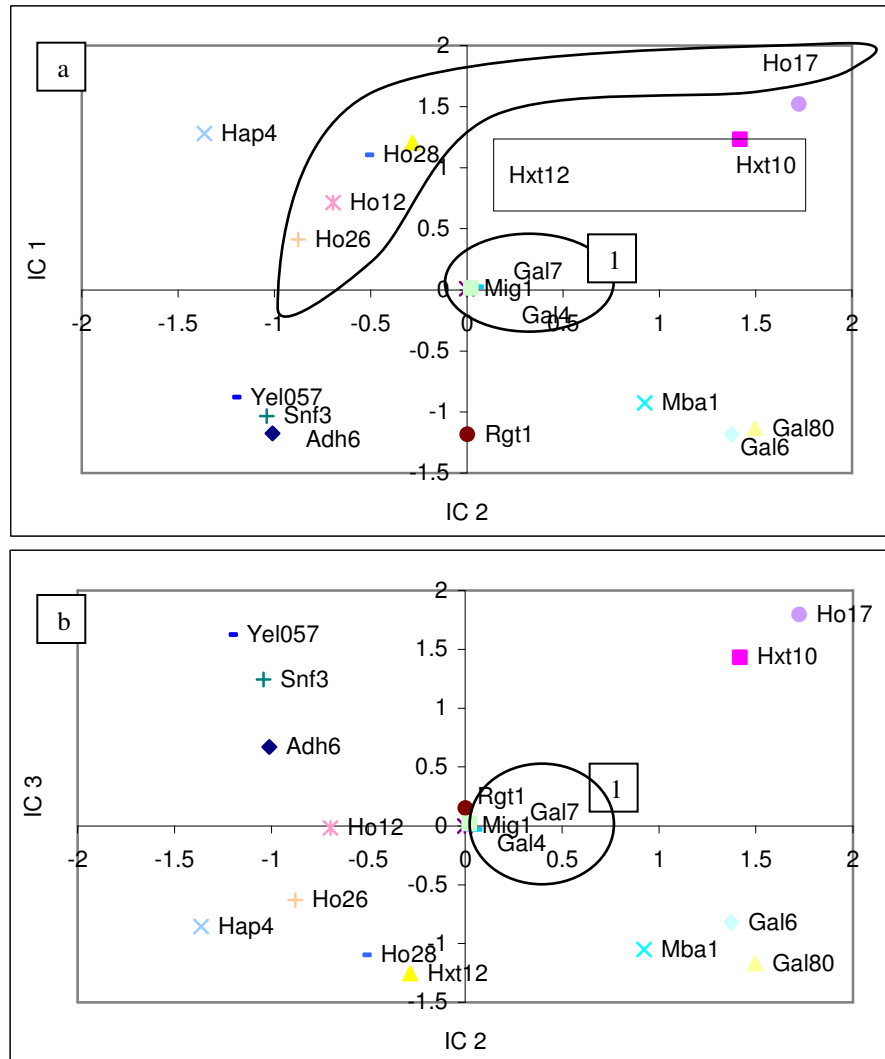


Figure 4.60. Scores of filtered metabolome data on first three ICs at 31st hour of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 6)

4.2.7. PCA and ICA of All Intracellular Metabolome Data (preprocess. by binning)

In Figure 4.61 (a), 4.61 (b), 4.61 (c), 4.61 (d); cluster 1 includes intracellular metabolome data at 20th hour of fermentation. In Figure 4.61 (a); cluster 2 is a combination of metabolome data at 20th and 25th hour of fermentation and cluster 3 represents only intracellular dataset at 31st hour of fermentation. In Figure 4.61 (b), cluster 2 also includes metabolome data at 20th hour of fermentation.

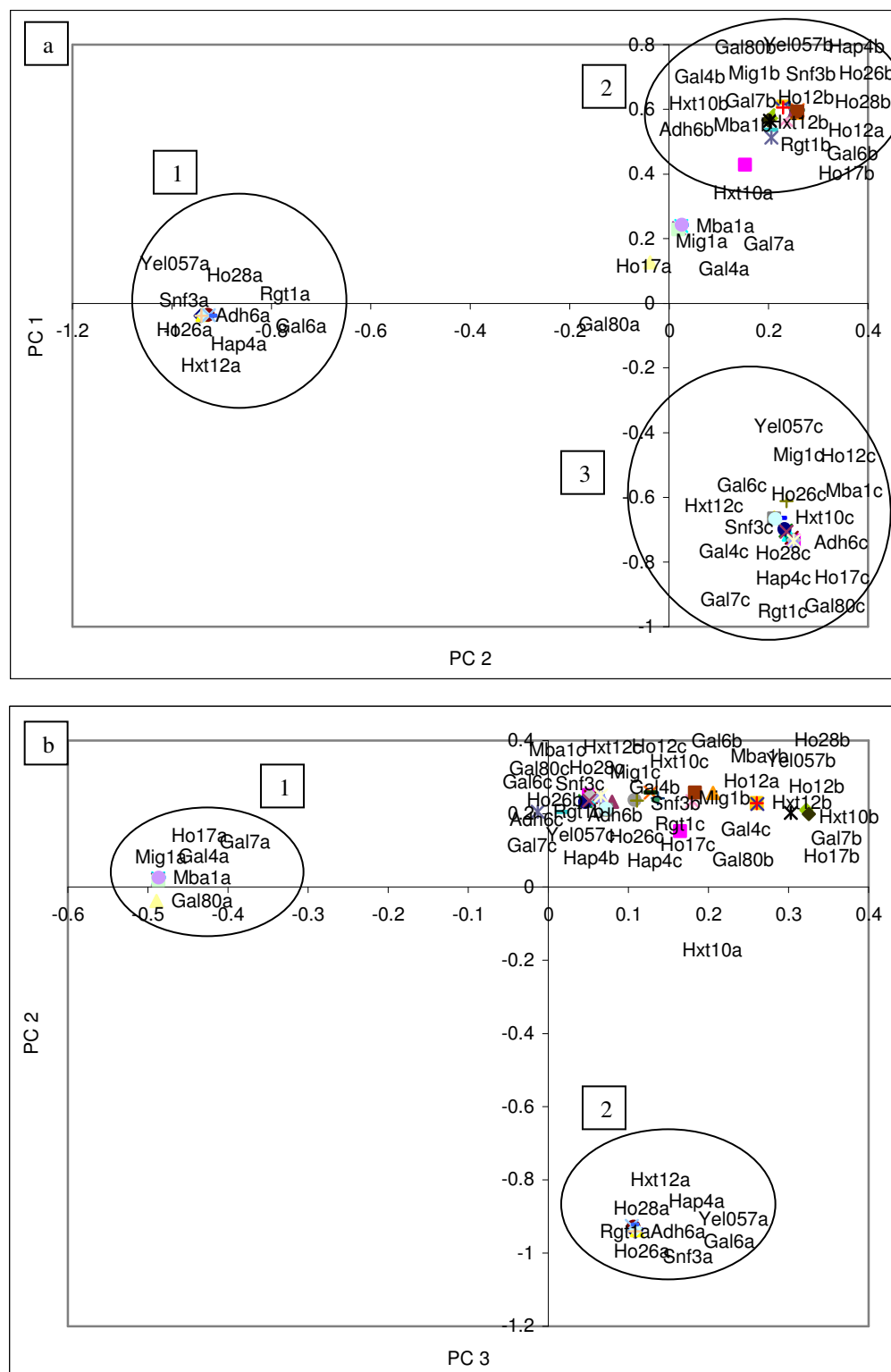


Figure 4.61. Scores of all metabolome data binned on first four LVs at three time intervals of intracellular samples

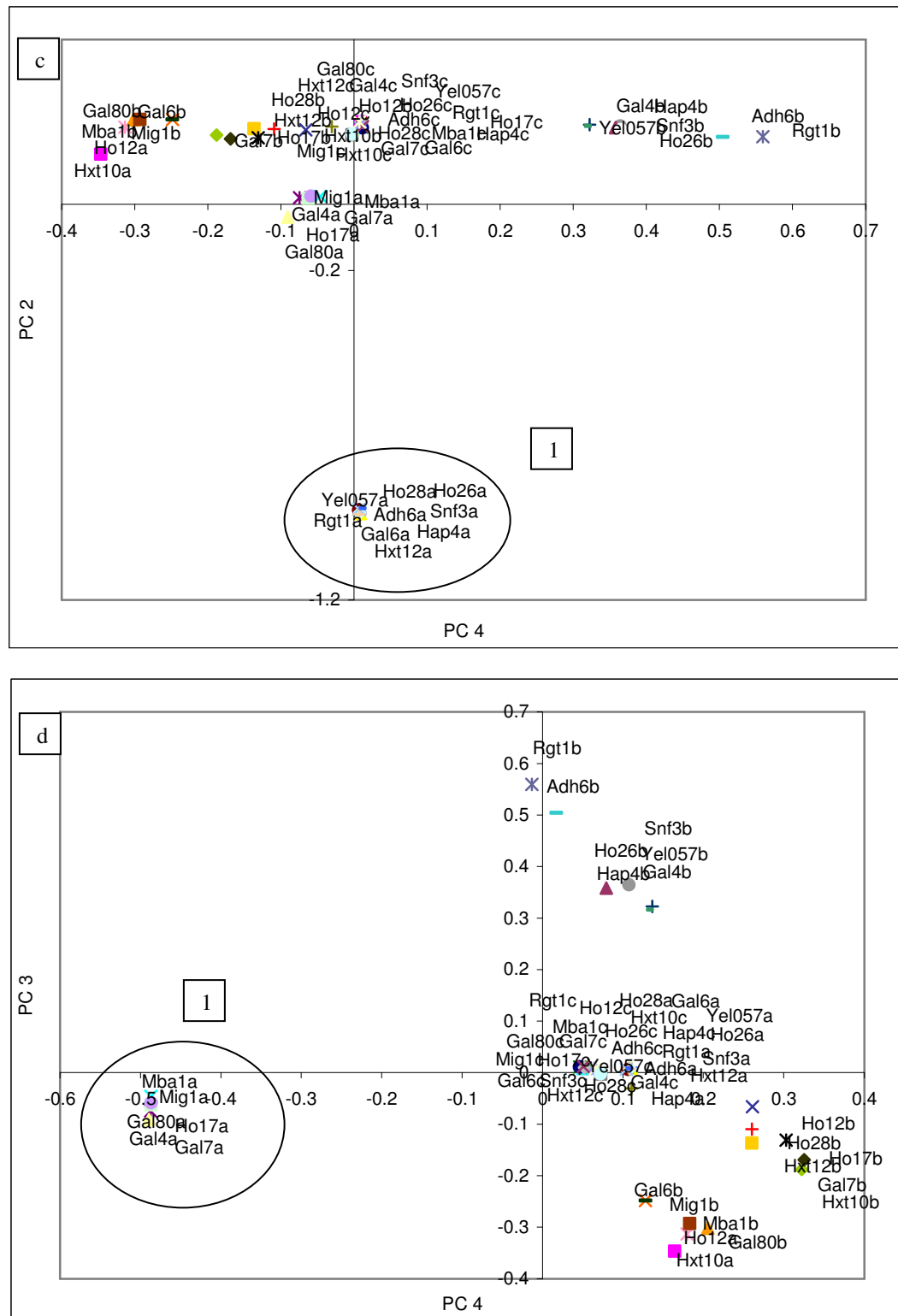
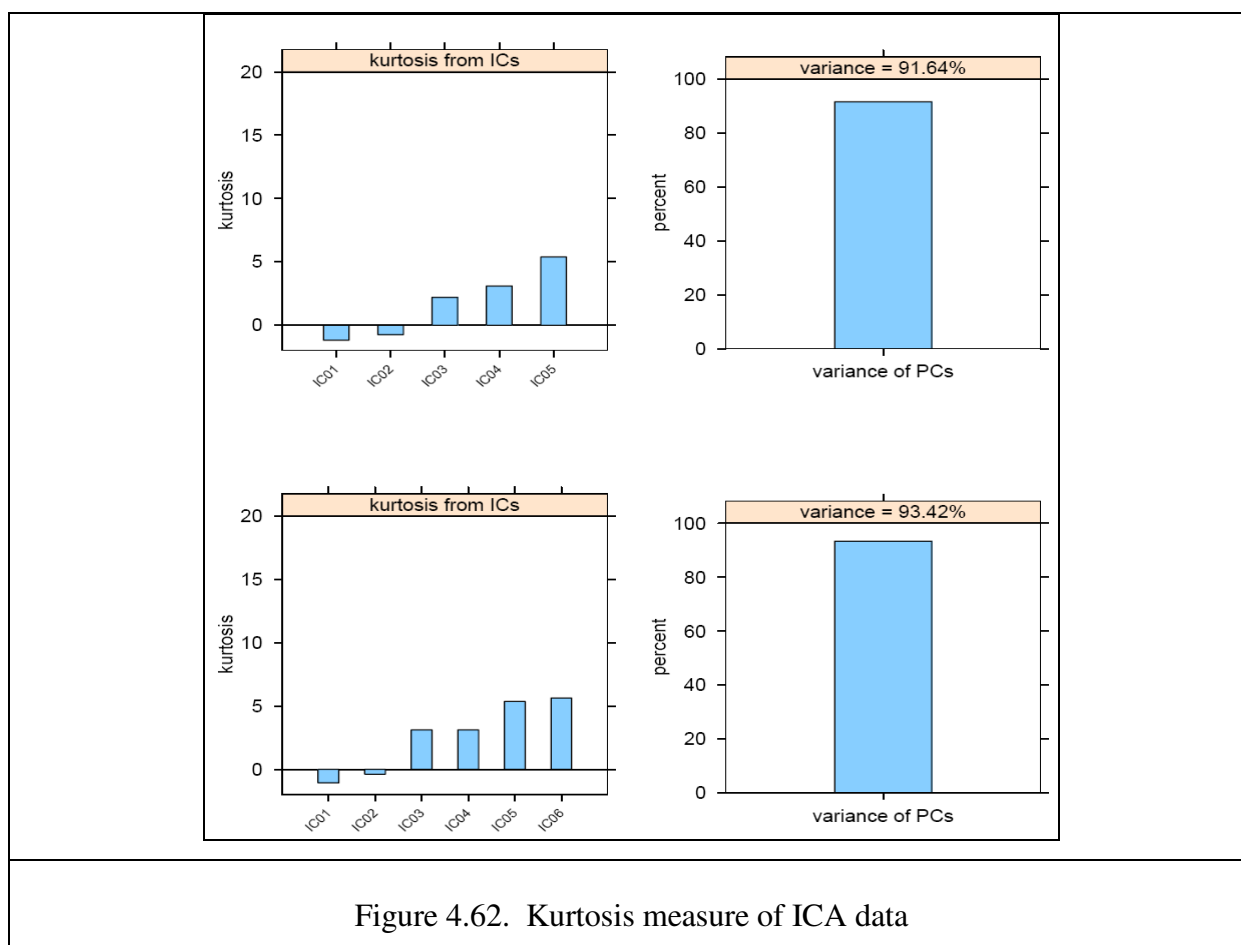


Figure 4.61. continued

It was stated in Materials and Methods section that the more negative values the kurtosis gets, the better clustering in ICA is obtained. In Figure 4.62, there are 2 ICs whose kurtosis are negative. However when PCA preprocessing number was 5, the degree of negativeness in IC2 is higher. Consequently, ICA preprocessing by 5 times was selected in analysis.



In Figure 4.63 (a) and (b); cluster 1 includes 20th hour; cluster 2 includes intracellular metabolome data at 25th hour, and cluster 3 includes intracellular metabolome data at 31st hour of fermentation. There are still groups which can not be separated according to time values.

As an overall result, ICs are more helpful in discriminating the 3 different clusters (20th, 25th, and 31st) than PCs. For example, in Figure 4.61 (a) (the best clustering in PCA), some of intracellular metabolome data collected at 20th and 25th hour of fermentation overlap (cluster 2).

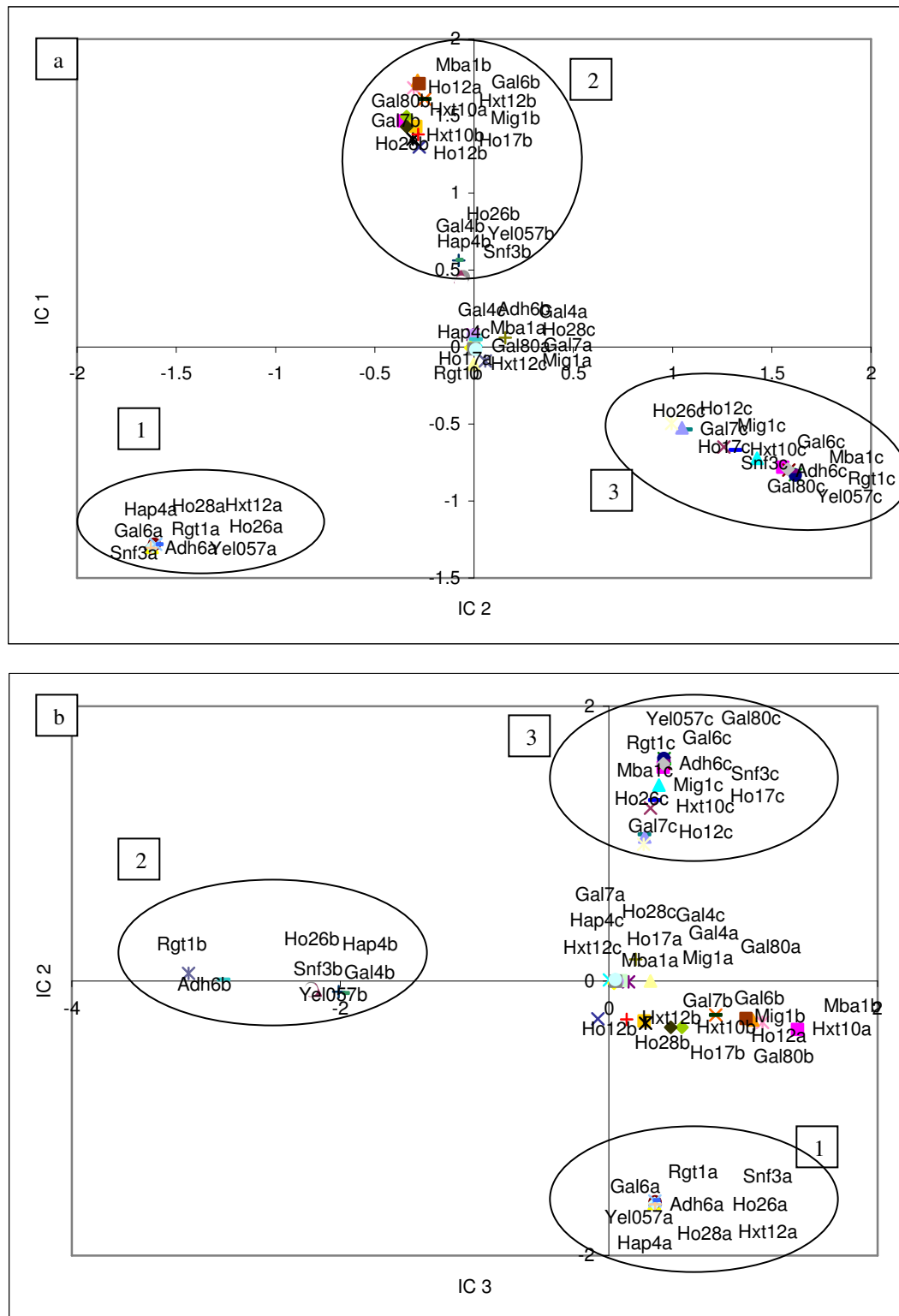


Figure 4.63. Scores of binned all metabolome data on first three ICs at 3 time intervals of intracellular samples, a) IC1 vs IC2, b) IC3 vs IC2 (PCA preprocessing number is 5)

5. DISCUSSION

In this study, the first objective was to analyze the deletion mutants of *S. cerevisiae* by metabolic footprinting and fingerprinting approaches, i.e. by m/z signatures. The second objective was to investigate the discrimination power of two approaches (binning vs filtering) together with PCA and ICA using the extracellular and intracellular metabolome profiles of yeast deletion mutants. Possible clusterings for 17 deletion mutants of *S. cerevisiae* are proposed. The third objective was to establish a library of main spectral footprints and fingerprints of gene deletion strains of *S. cerevisiae*. The fourth objective was to annotate a function to *yel057c* gene whose function is still not known or reported.

5.1. Comparison of the Two Preprocessing and Clustering Approaches

The genes which are deleted in *S. cerevisiae* mutants fall into the following biological process categories: galactose metabolism, hexose transport, transcription regulation (*rgt1*, *hap4*, *gal80*, *gal4*, *gal6*, *mig1*), and respiration (*mba1*, *hap4*). Both filtering and binning preprocessing approaches reflect the fact that the most proper clustering is obtained at the 20th hour of fermentation for extracellular samples. At this time of stationary phase, *galΔ*, *hoΔ* and *hxtΔ* deletions are discriminated successfully. Consequently, the success of the clustering is case dependent, i.e., changes in metabolite levels that affect possible clusterings would give completely different results.

The results of extracellular samples preprocessed by binning and filtering approaches are summarized in Table 5.1, and Table 5.2, respectively. These results are more promising than that of the intracellular ones. For example, ICA could reveal some clusters which are hidden in PCA. The discrimination power of ICA is more obvious in filtering approach (Table 5.2). Discrimination of the data belonging to extracellular samples at the 25th and 31st hour of fermentation was not possible by PCA. After the application of ICA to the same dataset, the discrimination became possible. ICA6, or in other words the processing of first 6 PCs in ICA

extracts the clusters and shows the importance of negative kurtosis. For the data set of extracellular samples, ICA6 counts higher than IC5 and gives more satisfactory results.

The results of intracellular samples preprocessed by binning and filtering approaches are summarized in Table 5.3, and Table 5.4, respectively. Both of the methods could not satisfactorily discriminate *gal* Δ , *ho* Δ , and *hxt* Δ clusterings. The only satisfactory result is obtained by the binning approach which gives two clusterings (*ho* Δ , *hxt* Δ) accurately at 25th hour of sampling. Neither PCA nor ICA is able to reveal known clusters. Although, many experiments should be conducted to elucidate the function of *Yel057c* Δ gene, it appears in the same region with *snf3* Δ , *adh6* Δ , and *hxt12* Δ in most figures. According to this result, it could be stated that *yel057c* might play a role in glucose or hexose transport like *snf3* Δ and *hxt12* Δ . As *snf3* Δ is required for the expression of hexose transporter genes and *hxt12* Δ is known to be linked to the telomere of chromosome IX-L, the clustering of *yel057c* Δ (involved in telomere maintenance) with *snf3* Δ and *hxt12* Δ is an expected outcome.

Table 5.1. Summary of the Discrimination Results Obtained by Binning Approach
(Extracellular Samples)

	<i>gal</i> deletions	<i>Ho</i> reference	<i>Hxt</i> deletions	Regulation related genes	Respiration related genes
15 th hour extra-pca	No	No	No	No	No
15 th hour extra-ica5	No	No	Yes	No	Yes
15 th hour extra-ica6	No	No	Yes	Yes	Yes
20 th hour extra-pca	Yes	Yes	Yes	Yes	Yes
20 th hour extra-ica5	Yes	Yes	Yes	Yes	Yes
20 th hour extra-ica6	Yes	Yes	Yes	Yes	Yes
25 th hour extra-pca	No	No	Yes	No	No
25 th hour extra-ica5	No	Yes	Yes	No	No
25 th hour extra-ica6	No	Yes	Yes	Yes	No
31 st hour extra-pca	Yes	No	No	No	No
31 st hour extra-ica5	Yes	No	No	No	No
31 st hour extra-ica6	Yes	No	No	No	No

Table 5.2. Summary of the Discrimination Results Obtained by Filtering Approach
(Extracellular Samples)

	<i>gal</i> deletions	<i>Ho</i> reference	<i>Hxt</i> deletions	Regulation related genes	Respiration related genes
15 th hour extra-pca	Yes	No	Yes	No	Yes
15 th hour extra-ica5	Yes	Yes	Yes	Yes	Yes
15 th hour extra-ica6	No	No	Yes	Yes	Yes
20 th hour extra-pca	Yes	No	Yes	Yes	Yes
20 th hour extra-ica5	Yes	Yes	Yes	Yes	No
20 th hour extra-ica6	Yes	Yes	Yes	No	No
25 th hour extra-pca	No	Yes	No	No	No
25 th hour extra-ica5	No	No	No	Yes	No
25 th hour extra-ica6	Yes	Yes	Yes	Yes	No
31 st hour extra-pca	No	No	No	No	No
31 st hour extra-ica5	Yes	No	Yes	Yes	No
31 st hour extra-ica6	Yes	No	Yes	Yes	No

Table 5.3. Summary of the Discrimination Results Obtained by Binning Approach
(Intracellular Samples)

	<i>gal</i> deletions	<i>Ho</i> reference	<i>Hxt</i> deletions	Regulation related genes	Respiration related genes
20 th hour intra-pca	No	No	No	No	No
20 th hour intra-ica5	No	No	No	No	No
20 th hour intra-ica6	No	No	No	No	No
25 th hour intra-pca	No	Yes	Yes	No	No
25 th hour intra-ica5	No	Yes	Yes	Yes	No
25 th hour intra-ica6	No	Yes	Yes	Yes	No
31 st hour intra-pca	No	No	No	Yes	No
31 st hour intra-ica5	No	No	No	Yes	No
31 st hour intra-ica6	No	No	No	Yes	No

Table 5.4. Summary of the Discrimination Results Obtained by Filtering Approach
(Intracellular Samples)

	<i>gal</i> deletions	<i>Ho</i> reference	<i>Hxt</i> deletions	Regulation related genes	Respiration related genes
20 th hour intra-pca	No	No	Yes	No	No
20 th hour intra-ica5	No	No	No	No	No
20 th hour intra-ica6	No	No	No	No	No
25 th hour intra-pca	No	No	Yes	No	No
25 th hour intra-ica5	No	No	No	No	No
25 th hour intra-ica6	No	No	No	No	No
31 st hour intra-pca	No	No	No	No	No
31 st hour intra-ica5	No	No	No	No	No
31 st hour intra-ica6	Yes	Yes	No	No	No

5.2. Significant m/z Values in PCA

Significant m/z values were obtained by taking the transpose of metabolome matrix at two retention times. These are 20th hour of fermentation for extracellular samples, and 25th hour of fermentation for intracellular samples. In these data matrix, the columns or objects become m/z values, and the rows or variables become deletion mutants.

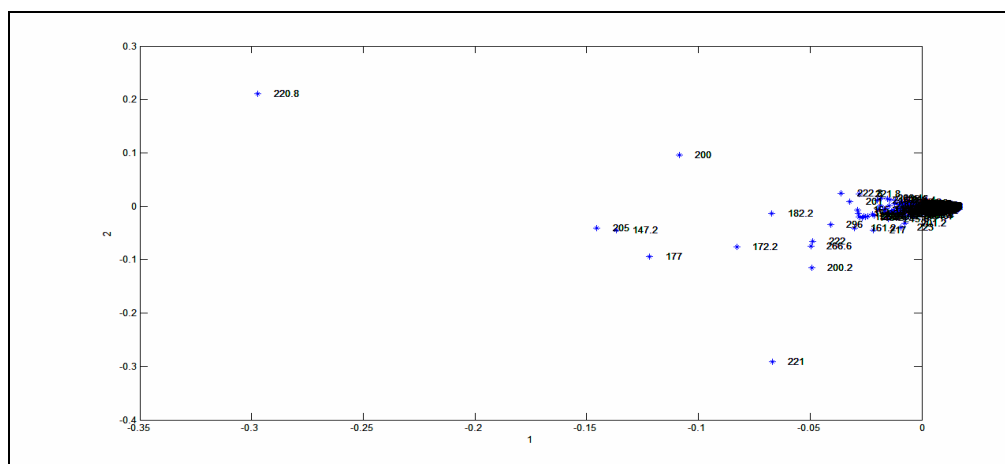


Figure 5.1. Loadings of PC1 vs PC2 of extracellular samples taken at 20th hour of fermentation

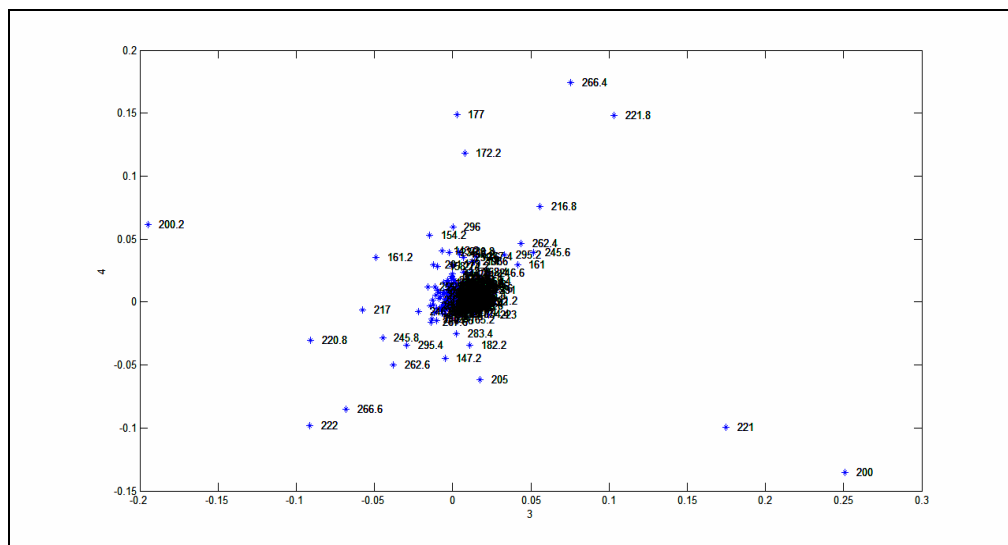


Figure 5.2. Loadings of PC3 vs PC4 of extracellular samples taken at 20th hour of fermentation

Table 5.5. m/z values with the highest absolute loadings in the first four PCs (ext. samples)

	PC1	PC2	PC3	PC4
Loadings	147.2, 172.2, 177.2, 182.2, 200, 205, 220.8, 222,	200.2, 221	161.2, 200, 200.2, 216.8, 217, 220.8, 221, 245.6,	147.2, 154.2, 172.2, 177, 205, 216.8, 221.8, 222,

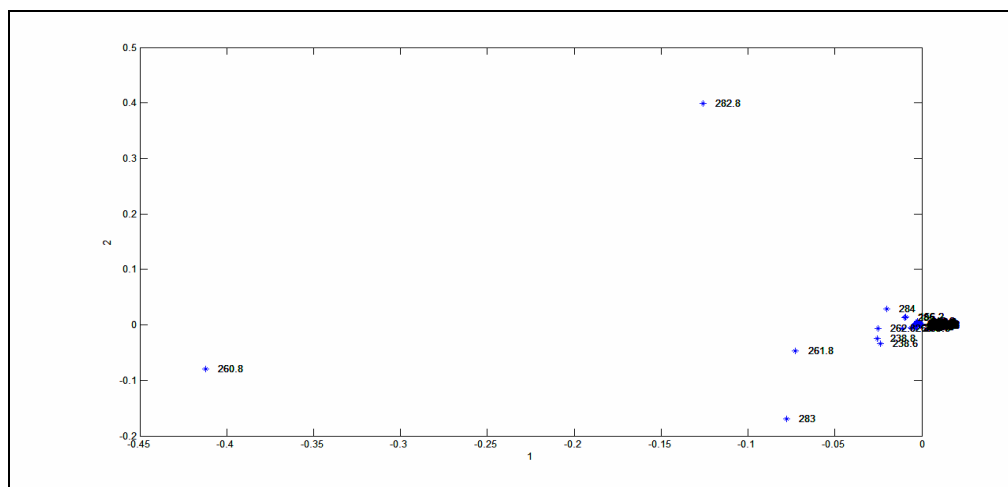


Figure 5.3. Loadings of PC1 vs PC2 of intracellular samples taken at 25th hour of fermentation

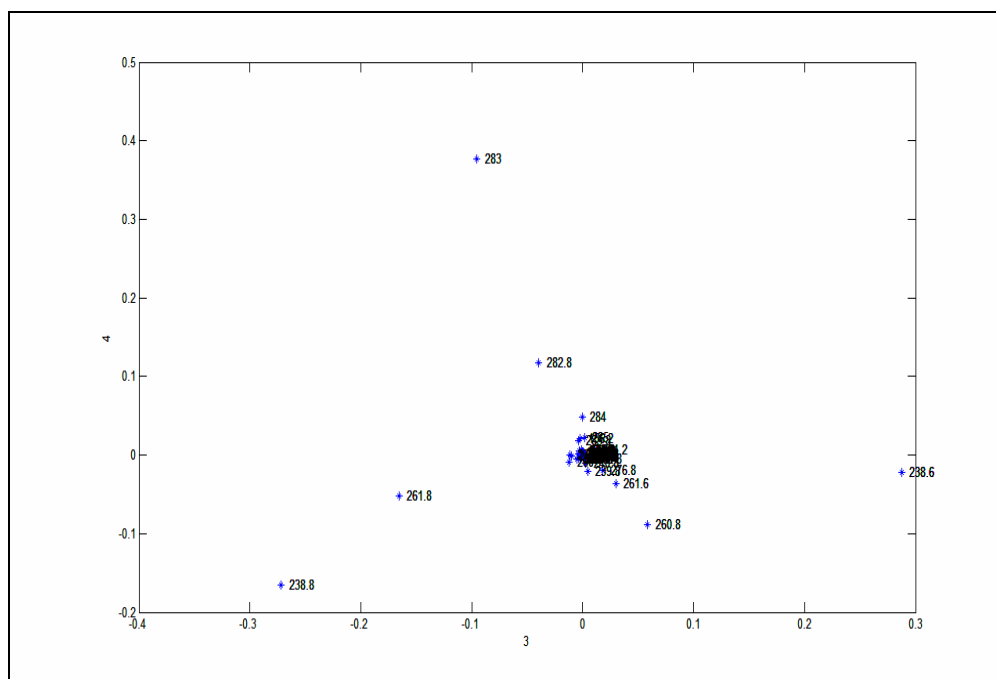


Figure 5.4. Loadings of PC3 vs PC4 of intracellular samples taken at 25th hour of fermentation

Table 5.6. m/z values with the highest absolute loadings in the first four PCs (int. samples)

	PC1	PC2	PC3	PC4
Loadings	238.6, 238.8, 260.8, 262.8	261.8, 282.8, 283	238.6, 238.8, 260.8, 261.6, 261.8, 276.8	282.8, 283, 284

Tables 5.5 and 5.6 list the m/z values with the highest loading values for each principal component. Inspection of the weight loadings in these tables and in Figures 5.1-5.4 reveals that the first principal component is determined more by m/z values of 200, 220, and 266. The third principal component is determined by m/z values of 220 and 245 for extracellular samples. For intracellular samples, the first PC is determined more by m/z values of 238, 260, and the second principal component is determined by m/z values of 262 and 283.

It is observed that similar m/z values appear significant in more than one PC, leading to 3 key m/z values that are identified as significant by the PCA. For metabolite footprinting, m/z values of 200, 220, 266 can be considered as the key variables. For metabolite fingerprinting,

m/z values of 238, 260, 262, and 283 are identified as the three key variables. They contribute most highly to the clustering of the data in the first four PCs.

5.3. m/z Signatures of Deletion Mutants

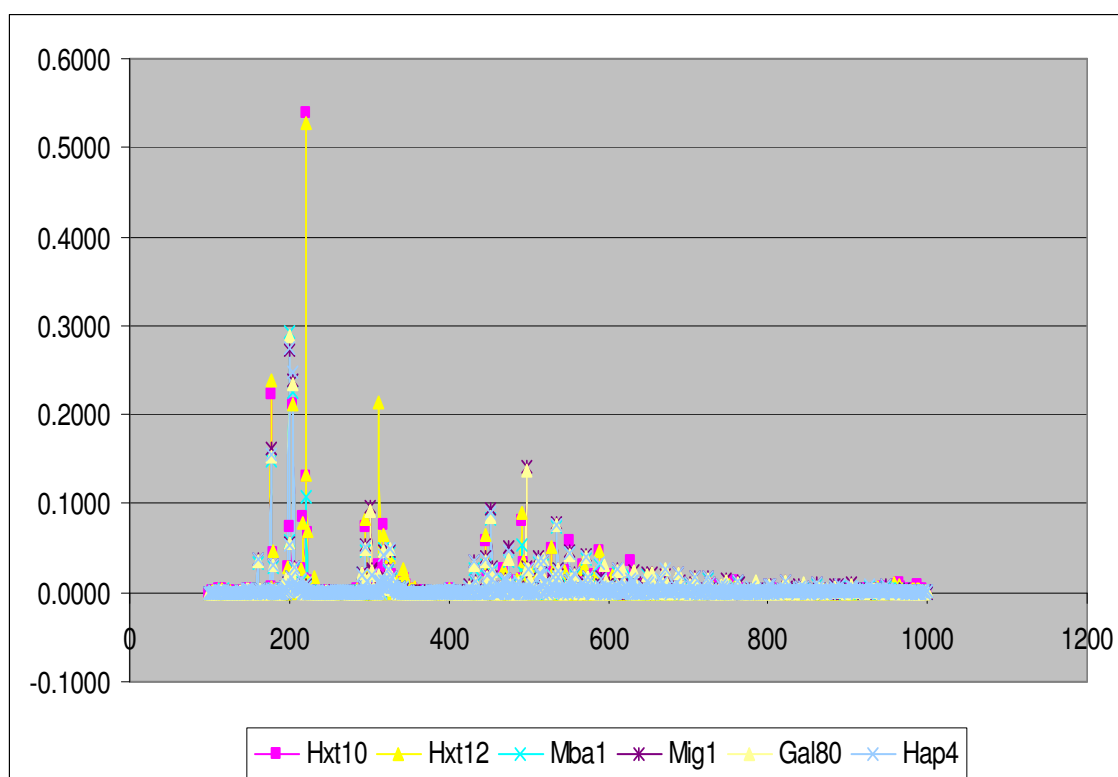


Figure 5.5. Mass spectrum of 8 deletion mutants (extracellular samples taken at 20th hour of fermentation)

The highest intensities are obtained at m/z values of 177, 200, 221, 312, 453, 497, and 535 in Figure 5.5. The metabolites at these locations can be found by the application of LC/MS/MS.

In Figure 5.6 (a), the footprints of the two transcriptional repressors (*mig1Δ* and *gal80Δ*) are highly correlated. The metabolome profiles of the two respiration related deletion mutants, *hap4Δ* and *mba1Δ*; reference strains *ho12Δ* and *ho17Δ* are also observed to be highly correlated (Figure 5.6 (b) and 5.6 (c)). The correlation between *ho12Δ* and *snf3Δ* is quite bad

with respect to others as expected (Figure 5.6 (d)) and this result compromises the scientific facts.

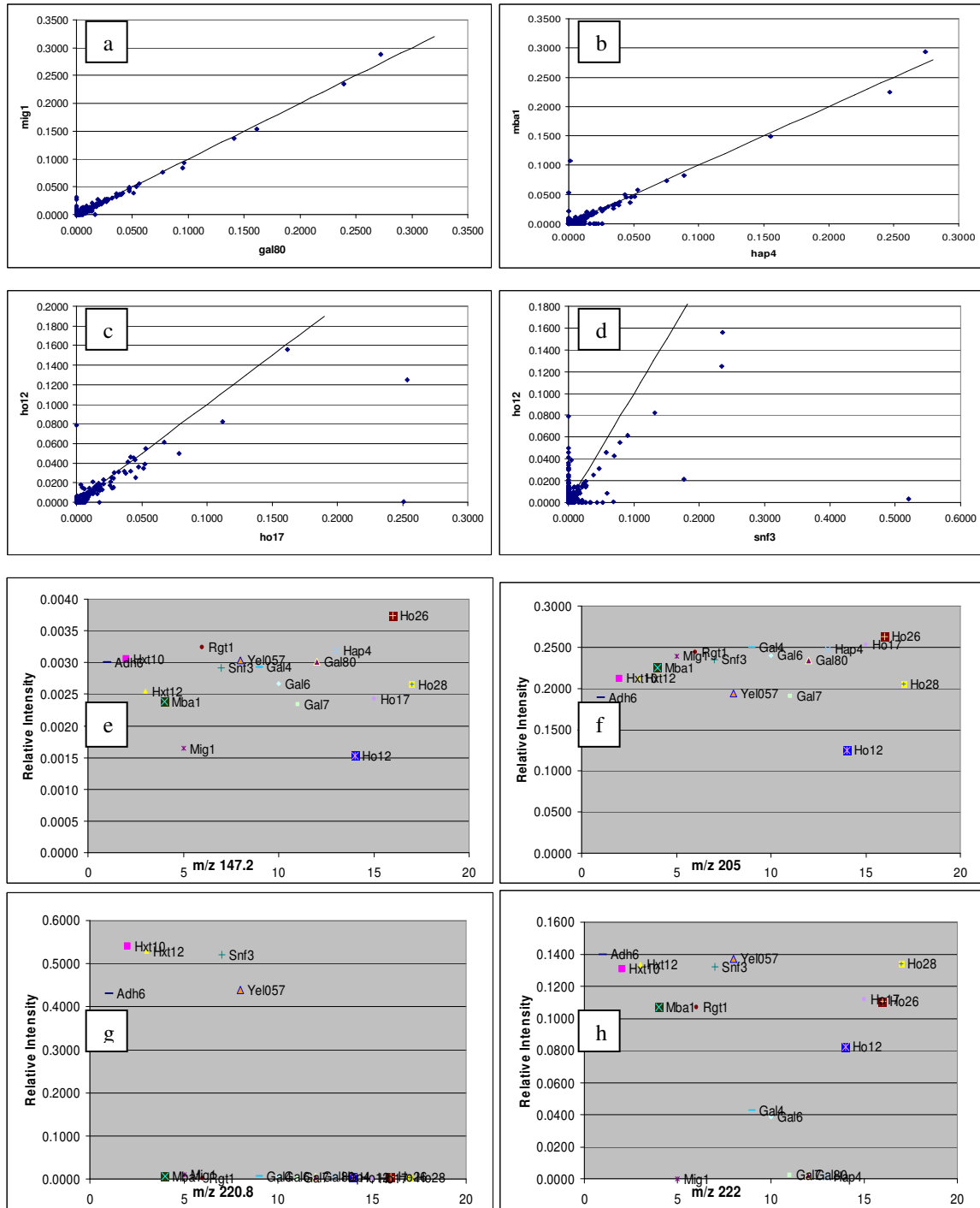


Figure 5.6. Pair-wise mutant correlations and intensity values at peak conditions

It is observed In Figure 5.6 (e-h) that *ho4* deletion mutants, Gal deletion mutants and hexose transporter mutants are close to each other within the group *snf3*, *RpH* and *yel057* are also found to be close confirming the PCA and ICA results.

In Figure 5.6 (e) and 5.6 (f); the deletion mutants have nearly the same relative intensity values and the discrimination is not possible. In Figure 5.6 (g); *hxt10Δ*, *hxt12Δ*, *snf3Δ*, and *adh6Δ*, *yel057cΔ* have close intensity values. In Figure 5.6 (h); *hxt10Δ*, *hxt12Δ*, *adh6Δ*, *snf3Δ*, *yel057cΔ*, *ho28Δ* and *mba1Δ*, *rgt1Δ*, *ho17Δ*, *ho26Δ* could be possible clusters.

Seven marker *m/z* values are identified from the extracellular metabolome data taken at the 20th hour of fermentation. These are 147.2, 205, 220.8, 222, 381.6, 399.4, and 437.2. In other words, these values can be used to discriminate the metabolome profiles of deletion mutants.

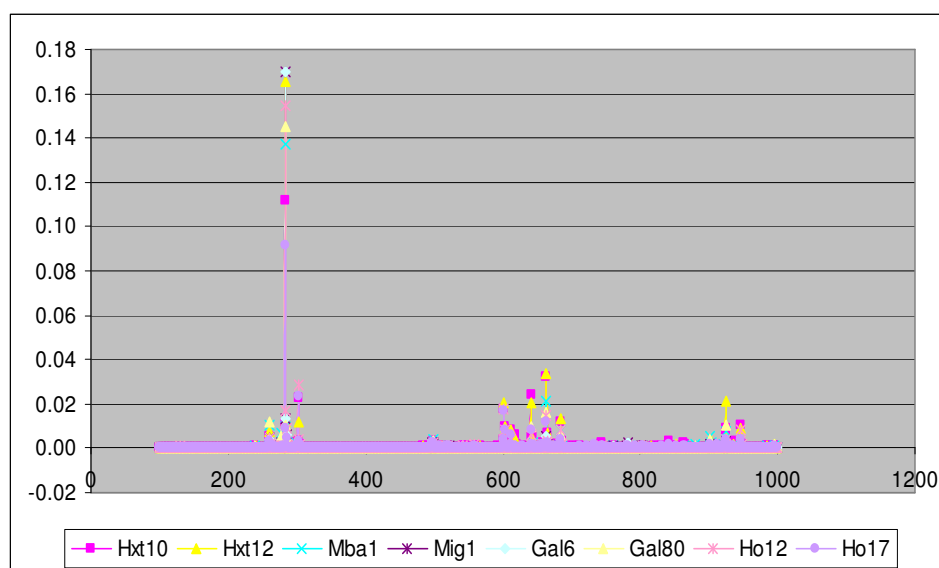


Figure 5.7. Mass spectrum of 8 deletion mutants (intracellular samples taken at 25th hour of fermentation)

The highest intensities are obtained at *m/z* values of 283, 601, 641, 663 in Figure 5.7. This mass spectrum is quite similar with any of the deletion mutants' metabolic fingerprinting spectrum.

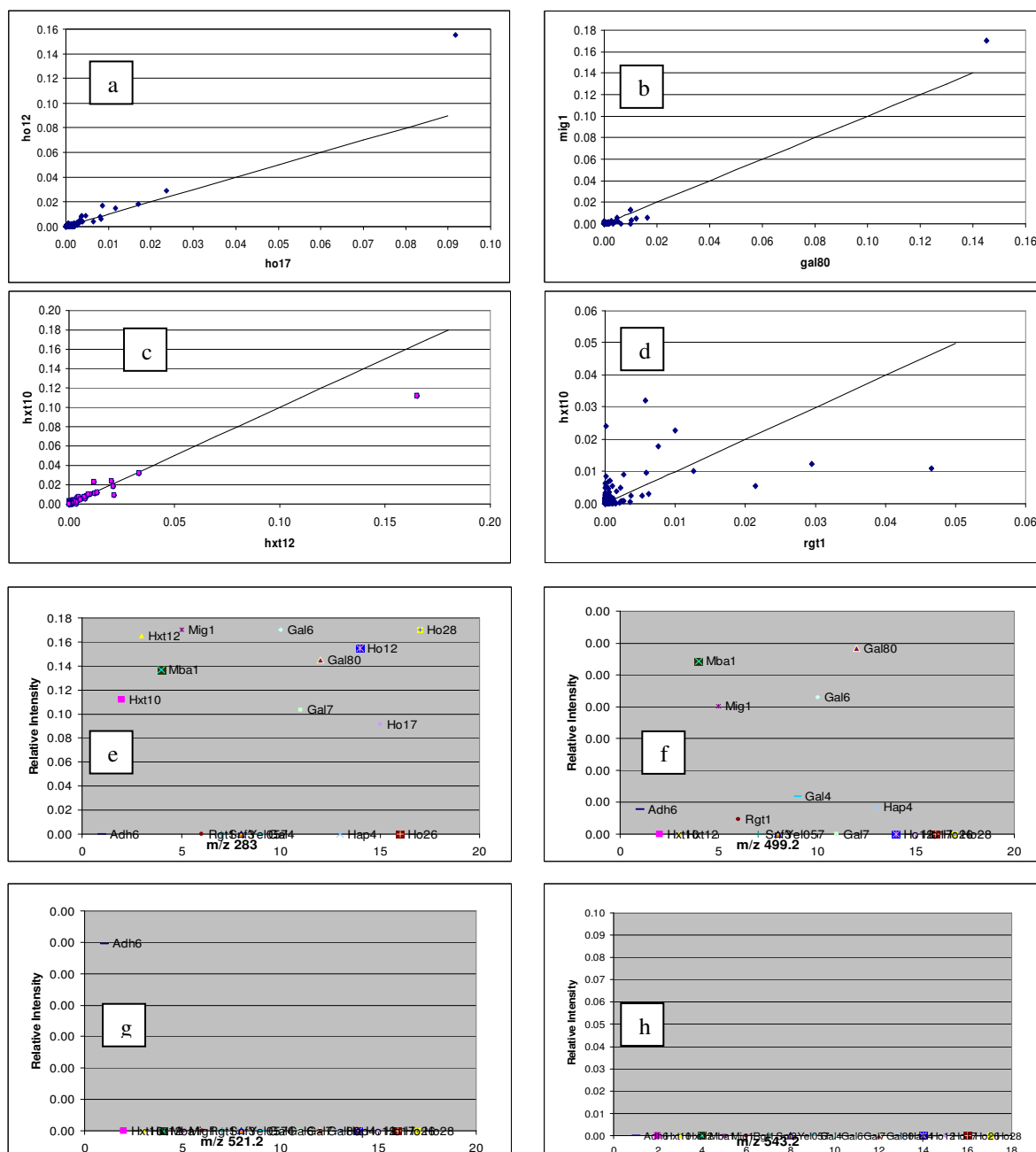


Figure 5.8. Pair-wise mutant correlations and intensity values at peak conditions

The pair-wise metabolite correlations for the reference strains (*ho12Δ*, *ho17Δ*) and transcriptional repressors (*mig1Δ*, *gal80Δ*) are shown in Figure 5.8 (a) and (b), respectively. These two pairs should have been highly correlated due to their functional relation in *mig1Δ*-*gal80Δ* pair, or being the same strains in *ho12Δ*-*ho17Δ* pair. In Figure 5.8 (c), *hxt10Δ*-*hxt12Δ*

pair also seems to be highly correlated and verifying the relation between the two deletion mutants. As it can be expected, the correlation between *hxt10Δ-rgt1Δ* is weak because these deletion mutants are functionally not related (Figure 5.8 (d)). In Figure 5.8 (e); the relative intensity values of 17 deletion mutants were analysed at 283 m/z value. Being in the same metabolic pathway *GalΔ* genes appeared close to each other. In Figure 5.8 (f); *mba1Δ*, *gal80Δ* and *mig1Δ*, *gal6Δ* have almost the same intensity values. In Figure 5.8 (g) and 5.8 (h); the relative intensity values are so low that there is no cluster to visualise. Eleven marker m/z values are identified from the intracellular metabolome data taken at the 25th hour of fermentation. These are 283, 499.2, 500.2, 521.2, 543.2, 544.2, 545.2, 737.2, 759.2, 760.2, and 803.4 m/z values. Due to the closeness of mass value pairs such as; 499.2-500.2; 543.2-544.2-545.2; 759.2-760.2; each pair could attain a single m/z value. For example, 499.2-500.2 ion pair could be substituted as 499.2.

5.4. Correlation Analysis

5.4.1. Metabolic Footprinting Results

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.471	0.484	0.971	0.531	0.970	0.872	0.876	0.731	0.745	0.862	0.965	0.970	0.570	0.522	0.495	0.504
Hxt10	0.471	1.000	0.980	0.456	0.904	0.458	0.365	0.359	0.218	0.231	0.340	0.448	0.452	0.900	0.915	0.882	0.967
Hxt12	0.484	0.980	1.000	0.469	0.917	0.471	0.386	0.381	0.234	0.248	0.359	0.459	0.463	0.913	0.930	0.898	0.986
Mba1	0.971	0.456	0.469	1.000	0.523	0.997	0.870	0.862	0.754	0.767	0.857	0.988	0.996	0.536	0.513	0.497	0.482
Mig1	0.531	0.904	0.917	0.523	1.000	0.525	0.435	0.427	0.279	0.294	0.407	0.516	0.519	0.984	0.985	0.901	0.923
Rgt1	0.970	0.458	0.471	0.997	0.525	1.000	0.870	0.863	0.755	0.768	0.856	0.987	0.995	0.536	0.516	0.504	0.484
snf3	0.872	0.365	0.386	0.870	0.435	0.870	1.000	0.990	0.854	0.871	0.988	0.873	0.864	0.454	0.423	0.422	0.400
Yel057	0.876	0.359	0.381	0.862	0.427	0.863	0.990	1.000	0.859	0.875	0.989	0.864	0.856	0.452	0.416	0.417	0.397
Gal4	0.731	0.218	0.234	0.754	0.279	0.755	0.854	0.859	1.000	0.997	0.871	0.757	0.751	0.281	0.268	0.287	0.244
Gal6	0.745	0.231	0.248	0.767	0.294	0.768	0.871	0.875	0.997	1.000	0.886	0.770	0.764	0.298	0.283	0.303	0.259
Gal7	0.862	0.340	0.359	0.857	0.407	0.856	0.988	0.989	0.871	0.886	1.000	0.863	0.853	0.429	0.395	0.392	0.372
Gal80	0.965	0.448	0.459	0.988	0.516	0.987	0.873	0.864	0.757	0.770	0.863	1.000	0.989	0.525	0.506	0.491	0.472
Hap4	0.970	0.452	0.463	0.996	0.519	0.995	0.864	0.856	0.751	0.764	0.853	0.989	1.000	0.531	0.508	0.492	0.476
Ho12	0.570	0.900	0.913	0.536	0.984	0.536	0.454	0.452	0.281	0.298	0.429	0.525	0.531	1.000	0.971	0.875	0.923
Ho17	0.522	0.915	0.930	0.513	0.985	0.516	0.423	0.416	0.268	0.283	0.395	0.506	0.508	0.971	1.000	0.910	0.936
Ho26	0.495	0.882	0.898	0.497	0.901	0.504	0.422	0.417	0.287	0.303	0.392	0.491	0.492	0.875	0.910	1.000	0.904
Ho28	0.504	0.967	0.986	0.482	0.923	0.484	0.400	0.397	0.244	0.259	0.372	0.472	0.476	0.923	0.936	0.904	1.000

Figure 5.9 Pearson correlation maps of binned metabolome data at 15th hour of extracellular samples

In Figure 5.9, *adh6Δ*, *mba1Δ*, *rgt1Δ*, *gal80Δ* and *hap4Δ* are highly correlated with each other. *Hxt10Δ*, *hxt12Δ*, *ho12Δ*, *ho17Δ*, *ho26Δ* and *ho28Δ* deletion mutants are also highly correlated within the group.

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.987	0.985	0.975	0.974	0.974	0.994	0.998	0.958	0.966	0.997	0.973	0.973	0.997	0.975	0.893	0.993
Hxt10	0.987	1.000	0.998	0.977	0.978	0.976	0.985	0.985	0.963	0.969	0.988	0.976	0.975	0.990	0.978	0.901	0.991
Hxt12	0.985	0.998	1.000	0.973	0.974	0.973	0.985	0.986	0.953	0.961	0.984	0.971	0.971	0.987	0.975	0.902	0.991
Mba1	0.975	0.977	0.973	1.000	0.999	1.000	0.984	0.974	0.989	0.994	0.982	0.999	0.999	0.985	0.999	0.928	0.979
Mig1	0.974	0.978	0.974	0.999	1.000	0.999	0.982	0.972	0.989	0.994	0.981	0.999	0.999	0.985	0.999	0.931	0.979
Rgt1	0.974	0.976	0.973	1.000	0.999	1.000	0.984	0.974	0.986	0.992	0.981	0.999	0.999	0.983	0.999	0.932	0.979
snf3	0.994	0.985	0.985	0.984	0.982	0.984	1.000	0.995	0.962	0.970	0.993	0.981	0.981	0.994	0.984	0.905	0.993
Yel057	0.998	0.985	0.986	0.974	0.972	0.974	0.995	1.000	0.949	0.958	0.994	0.970	0.970	0.993	0.973	0.898	0.994
Gal4	0.958	0.963	0.953	0.989	0.989	0.986	0.962	0.949	1.000	0.998	0.970	0.992	0.991	0.971	0.985	0.904	0.956
Gal6	0.966	0.969	0.961	0.994	0.994	0.992	0.970	0.958	0.998	1.000	0.976	0.996	0.996	0.978	0.992	0.913	0.965
Gal7	0.997	0.988	0.984	0.982	0.981	0.981	0.993	0.994	0.970	0.976	1.000	0.981	0.980	0.997	0.982	0.900	0.992
Gal80	0.973	0.976	0.971	0.999	0.999	0.999	0.981	0.970	0.992	0.996	0.981	1.000	1.000	0.983	0.999	0.929	0.976
Hap4	0.973	0.975	0.971	0.999	0.999	0.999	0.981	0.970	0.991	0.996	0.980	1.000	1.000	0.983	0.998	0.926	0.976
Ho12	0.997	0.990	0.987	0.985	0.985	0.983	0.994	0.993	0.971	0.978	0.997	0.983	0.983	1.000	0.985	0.903	0.994
Ho17	0.975	0.978	0.975	0.999	0.999	0.999	0.984	0.973	0.985	0.992	0.982	0.999	0.998	0.985	1.000	0.931	0.980
Ho26	0.893	0.901	0.902	0.928	0.931	0.932	0.905	0.898	0.904	0.913	0.900	0.929	0.926	0.903	0.931	1.000	0.909
Ho28	0.993	0.991	0.991	0.979	0.979	0.979	0.993	0.994	0.956	0.965	0.992	0.976	0.976	0.994	0.980	0.909	1.000

Figure 5.10. Pearson correlation maps of filtered metabolome data at 15th hour of extracellular samples

The filtering approach could not give significant information about deletion mutants correlation. All deletion mutants seem highly correlated but this result is unacceptable.

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.881	0.972	0.307	0.260	0.328	0.904	0.931	0.253	0.224	0.285	0.235	0.285	0.489	0.367	0.425	0.545
Hxt10	0.881	1.000	0.912	0.352	0.302	0.373	0.855	0.835	0.291	0.243	0.312	0.273	0.331	0.527	0.407	0.477	0.575
Hxt12	0.972	0.912	1.000	0.292	0.242	0.311	0.933	0.927	0.238	0.213	0.256	0.221	0.272	0.455	0.345	0.416	0.501
Mba1	0.307	0.352	0.292	1.000	0.929	0.986	0.253	0.275	0.924	0.876	0.823	0.921	0.967	0.854	0.970	0.885	0.827
Mig1	0.260	0.302	0.242	0.929	1.000	0.915	0.219	0.227	0.980	0.930	0.907	0.984	0.960	0.791	0.897	0.819	0.764
Rgt1	0.328	0.373	0.311	0.986	0.915	1.000	0.269	0.295	0.910	0.861	0.806	0.907	0.956	0.876	0.983	0.903	0.850
snf3	0.904	0.855	0.933	0.253	0.219	0.269	1.000	0.951	0.219	0.207	0.227	0.199	0.233	0.393	0.300	0.368	0.439
Yel057	0.931	0.835	0.927	0.275	0.227	0.295	0.951	1.000	0.228	0.216	0.252	0.207	0.252	0.434	0.328	0.384	0.489
Gal4	0.253	0.291	0.238	0.924	0.980	0.910	0.219	0.228	1.000	0.948	0.927	0.992	0.944	0.775	0.890	0.811	0.750
Gal6	0.224	0.243	0.213	0.876	0.930	0.861	0.207	0.216	0.948	1.000	0.909	0.945	0.895	0.718	0.839	0.764	0.696
Gal7	0.285	0.312	0.256	0.823	0.907	0.806	0.227	0.252	0.927	0.909	1.000	0.920	0.865	0.755	0.790	0.726	0.718
Gal80	0.235	0.273	0.221	0.921	0.984	0.907	0.199	0.207	0.992	0.945	0.920	1.000	0.950	0.765	0.887	0.808	0.740
Hap4	0.285	0.331	0.272	0.967	0.960	0.956	0.233	0.252	0.944	0.895	0.865	0.950	1.000	0.824	0.941	0.861	0.798
Ho12	0.489	0.527	0.455	0.854	0.791	0.876	0.393	0.434	0.775	0.718	0.755	0.765	0.824	1.000	0.883	0.926	0.951
Ho17	0.367	0.407	0.345	0.970	0.897	0.983	0.300	0.328	0.890	0.839	0.790	0.887	0.941	0.883	1.000	0.920	0.878
Ho26	0.425	0.477	0.416	0.885	0.819	0.903	0.368	0.384	0.811	0.764	0.726	0.808	0.861	0.926	0.920	1.000	0.937
Ho28	0.545	0.575	0.501	0.827	0.764	0.850	0.439	0.489	0.750	0.696	0.718	0.740	0.798	0.951	0.878	0.937	1.000

Figure 5.11. Pearson correlation maps of binned metabolome data at 20th hour of extracellular samples

In Figure 5.11, there is a quite good correlation among the galactose transporters (*gal4Δ*, *gal6Δ*, *gal7Δ*, *gal80Δ*), *hap4Δ* and *mig1Δ*. Other novel findings at the 20th hour of extracellular metabolome data are the correlations of *hxt10Δ-hxt12Δ*; *mba1Δ-mig1Δ-rgt1Δ*; and *hoΔ* reference strains.

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.950	0.966	0.790	0.808	0.829	0.948	0.997	0.810	0.813	0.909	0.797	0.800	0.975	0.850	0.844	0.974
Hxt10	0.950	1.000	0.996	0.932	0.938	0.954	0.988	0.954	0.939	0.936	0.949	0.932	0.934	0.989	0.962	0.957	0.991
Hxt12	0.966	0.996	1.000	0.915	0.923	0.941	0.989	0.969	0.924	0.921	0.946	0.917	0.919	0.994	0.951	0.947	0.995
Mba1		0.932	0.915	1.000	0.996	0.991	0.927	0.802	0.995	0.985	0.906	0.996	0.997	0.892	0.984	0.983	0.893
Mig1	0.808	0.938	0.923	0.996	1.000	0.990	0.936	0.818	0.997	0.986	0.921	0.998	0.998	0.903	0.987	0.986	0.904
Rgt1	0.829	0.954	0.941	0.991	0.990	1.000	0.952	0.840	0.989	0.980	0.904	0.989	0.992	0.922	0.997	0.996	0.923
snf3	0.948	0.988	0.989	0.927	0.936	0.952	1.000	0.955	0.937	0.931	0.940	0.930	0.933	0.983	0.963	0.959	0.982
Yel057	0.997	0.954	0.969	0.802	0.818	0.840	0.955	1.000	0.820	0.823	0.912	0.808	0.810	0.976	0.859	0.856	0.977
Gal4	0.810	0.939	0.924	0.995	0.997	0.989	0.937	0.820	1.000	0.988	0.925	0.998	0.995	0.903	0.985	0.984	0.905
Gal6	0.813	0.936	0.921	0.985	0.986	0.980	0.931	0.823	0.988	1.000	0.939	0.988	0.986	0.902	0.977	0.978	0.905
Gal7	0.909	0.949	0.946	0.906	0.921	0.904	0.940	0.912	0.925	0.939	1.000	0.917	0.910	0.943	0.911	0.908	0.946
Gal80	0.797	0.932	0.917	0.996	0.998	0.989	0.930	0.808	0.998	0.988	0.917	1.000	0.996	0.895	0.985	0.984	0.896
Hap4	0.800	0.934	0.919	0.997	0.998	0.992	0.933	0.810	0.995	0.986	0.910	0.996	1.000	0.898	0.989	0.988	0.898
Ho12	0.975	0.989	0.994	0.892	0.903	0.922	0.983	0.976	0.903	0.902	0.943	0.895	0.898	1.000	0.935	0.930	0.997
Ho17	0.850	0.962	0.951	0.984	0.987	0.997	0.963	0.859	0.985	0.977	0.911	0.985	0.989	0.935	1.000	0.998	0.935
Ho26	0.844	0.957	0.947	0.983	0.986	0.996	0.959	0.856	0.984	0.978	0.908	0.984	0.988	0.930	0.998	1.000	0.932
Ho28	0.974	0.991	0.995	0.893	0.904	0.923	0.982	0.977	0.905	0.905	0.946	0.896	0.898	0.997	0.935	0.932	1.000

Figure 5.12. Pearson correlation maps of filtered metabolome data at 20th hour of extracellular samples

As it can be shown in Figure 5.12, the pearson correlation of filtering approach is not helpful for 20th hour of extracellular metabolome data.

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.776	0.747	0.312	0.316	0.321	0.693	0.917	0.389	0.320	0.886	0.276	0.288	0.025	0.004	0.953	0.784
Hxt10	0.776	1.000	0.955	0.704	0.722	0.736	0.672	0.818	0.782	0.727	0.929	0.373	0.471	0.034	0.020	0.096	0.934
Hxt12	0.747	0.955	1.000	0.694	0.715	0.715	0.647	0.816	0.757	0.702	0.893	0.348	0.460	0.039	0.024	0.087	0.923
Mba1	0.312	0.704	0.694	1.000	0.964	0.918	0.477	0.360	0.901	0.972	0.563	0.476	0.611	0.046	0.031	0.114	0.683
Mig1	0.316	0.722	0.715	0.964	1.000	0.945	0.505	0.373	0.930	0.968	0.579	0.500	0.644	0.050	0.033	0.111	0.705
Rgt1	0.321	0.736	0.715	0.918	0.945	1.000	0.528	0.373	0.978	0.941	0.594	0.552	0.689	0.047	0.033	0.118	0.725
snf3	0.693	0.672	0.647	0.477	0.505	0.528	1.000	0.673	0.574	0.487	0.717	0.746	0.802	0.032	0.007	0.293	0.707
Yel057	0.917	0.818	0.816	0.360	0.373	0.373	0.673	1.000	0.443	0.363	0.886	0.271	0.324	0.027	0.005	0.049	0.817
Gal4	0.389	0.782	0.757	0.901	0.930	0.978	0.574	0.443	1.000	0.927	0.663	0.542	0.685	0.047	0.033	0.119	0.790
Gal6	0.320	0.727	0.702	0.972	0.968	0.941	0.487	0.363	0.927	1.000	0.581	0.490	0.619	0.042	0.030	0.119	0.697
Gal7	0.886	0.929	0.893	0.563	0.579	0.594	0.717	0.886	0.663	0.581	1.000	0.346	0.414	0.039	0.022	0.079	0.932
Gal80	0.276	0.373	0.348	0.476	0.500	0.552	0.746	0.271	0.542	0.490	0.346	1.000	0.878	0.052	0.049	0.322	0.401
Hap4	0.288	0.471	0.460	0.611	0.644	0.689	0.802	0.324	0.685	0.619	0.414	0.878	1.000	0.038	0.018	0.329	0.502
Ho12	0.025	0.034	0.039	0.046	0.050	0.047	0.032	0.027	0.047	0.042	0.039	0.052	0.038	1.000	0.784	0.023	0.042
Ho17	0.004	0.020	0.024	0.031	0.033	0.033	0.007	0.005	0.033	0.030	0.022	0.049	0.018	0.784	1.000	0.021	0.027
Ho26	0.953	0.096	0.087	0.114	0.111	0.118	0.293	0.049	0.119	0.119	0.079	0.322	0.329	0.023	0.021	1.000	0.092
Ho28	0.784	0.934	0.923	0.683	0.705	0.725	0.707	0.817	0.790	0.697	0.932	0.401	0.502	0.042	0.027	0.092	1.000

Figure 5.13. Pearson correlation maps of binned metabolome data at 25th hour of extracellular samples

In Figure 5.13; *hxt10Δ-hxt12Δ-ho28Δ*; *mba1Δ-mig1Δ-rgt1Δ-gal4Δ-gal6Δ* groups are highly correlated with respect to each other. On the other hand, *ho12Δ*, *ho17Δ*, and *ho26Δ* are not correlated with other deletion mutants as expected.

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.765	0.794	0.224	0.196	0.172	0.594	0.990	0.232	0.212	0.889	0.100	0.151	-0.083	-0.055	-0.008	0.755
Hxt10	0.765	1.000	0.990	0.790	0.771	0.754	0.962	0.784	0.794	0.784	0.968	0.693	0.731	-0.043	-0.001	0.498	0.985
Hxt12	0.794	0.990	1.000	0.755	0.737	0.710	0.946	0.821	0.753	0.741	0.974	0.644	0.685	-0.050	-0.014	0.436	0.994
Mba1	0.224	0.790	0.755	1.000	0.999	0.991	0.900	0.263	0.992	0.995	0.630	0.965	0.979	0.003	0.051	0.786	0.784
Mig1	0.196	0.771	0.737	0.999	1.000	0.991	0.888	0.236	0.990	0.994	0.606	0.966	0.979	0.006	0.052	0.789	0.768
Rgt1	0.172	0.754	0.710	0.991	0.991	1.000	0.879	0.210	0.995	0.996	0.585	0.982	0.993	-0.016	0.027	0.832	0.744
snf3	0.594	0.962	0.946	0.900	0.888	0.879	1.000	0.617	0.909	0.896	0.890	0.834	0.866	-0.036	0.005	0.641	0.962
Yel057	0.990	0.784	0.821	0.263	0.236	0.210	0.617	1.000	0.270	0.250	0.896	0.140	0.187	-0.105	-0.078	0.016	0.782
Gal4	0.232	0.794	0.753	0.992	0.990	0.995	0.909	0.270	1.000	0.996	0.638	0.977	0.989	-0.016	0.029	0.810	0.787
Gal6	0.212	0.784	0.741	0.995	0.994	0.996	0.896	0.250	0.996	1.000	0.619	0.974	0.987	-0.009	0.038	0.816	0.771
Gal7	0.889	0.968	0.974	0.630	0.606	0.585	0.890	0.896	0.638	0.619	1.000	0.522	0.565	-0.042	-0.003	0.338	0.958
Gal80	0.100	0.693	0.644	0.965	0.966	0.982	0.834	0.140	0.977	0.974	0.522	1.000	0.988	-0.025	0.017	0.834	0.682
Hap4	0.151	0.731	0.685	0.979	0.979	0.993	0.866	0.187	0.989	0.987	0.565	0.988	1.000	-0.027	0.013	0.839	0.725
Ho12	-0.083	-0.043	-0.050	0.003	0.006	-0.016	-0.036	-0.105	-0.016	-0.009	-0.042	-0.025	-0.027	1.000	0.921	-0.013	-0.055
Ho17	-0.055	-0.001	-0.014	0.051	0.052	0.027	0.005	-0.078	0.029	0.038	-0.003	0.017	0.013	0.921	1.000	0.156	-0.027
Ho26	-0.008	0.498	0.436	0.786	0.789	0.832	0.641	0.016	0.810	0.816	0.338	0.834	0.839	-0.013	0.156	1.000	0.465
Ho28	0.755	0.985	0.994	0.784	0.768	0.744	0.962	0.782	0.787	0.771	0.958	0.682	0.725	-0.055	-0.027	0.465	1.000

Figure 5.14. Pearson correlation maps of filtered metabolome data at 25th hour of extracellular samples

In Figure 5.14; *adh6Δ*, *mig1Δ*, *mba1Δ*, *rgt1Δ*, *gal4Δ*, *gal6Δ*, *gal80Δ*, *hap4Δ*, *ho12Δ*, *ho17Δ* are not correlated with each other. *Hxt10Δ-hxt12Δ-snf3Δ-gal7Δ-ho28Δ*; *mba1Δ-mig1Δ-rgt1Δ-gal4Δ-gal6Δ-gal80Δ-hap4Δ* are the correlated groups.

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.966	0.962	0.782	0.519	0.255	0.837	0.750	0.496	0.595	0.966	0.729	0.417	0.746	0.714	0.686	0.063
Hxt10	0.966	1.000	0.975	0.807	0.548	0.263	0.844	0.775	0.493	0.601	0.969	0.752	0.444	0.761	0.739	0.686	0.067
Hxt12	0.962	0.975	1.000	0.805	0.552	0.249	0.843	0.778	0.495	0.599	0.968	0.756	0.445	0.764	0.739	0.688	0.064
Mba1	0.782	0.807	0.805	1.000	0.756	0.224	0.745	0.693	0.467	0.744	0.806	0.956	0.713	0.681	0.950	0.844	0.108
Mig1	0.519	0.548	0.552	0.756	1.000	0.074	0.631	0.691	0.257	0.529	0.549	0.736	0.921	0.682	0.766	0.639	0.136
Rgt1	0.255	0.263	0.249	0.224	0.074	1.000	0.202	0.174	0.619	0.400	0.245	0.216	0.067	0.178	0.177	0.264	0.014
snf3	0.837	0.844	0.843	0.745	0.631	0.202	1.000	0.899	0.399	0.524	0.856	0.690	0.502	0.901	0.696	0.628	0.064
Yel057	0.750	0.775	0.778	0.693	0.691	0.174	0.899	1.000	0.349	0.449	0.783	0.631	0.543	0.980	0.649	0.558	0.064
Gal4	0.496	0.493	0.495	0.467	0.257	0.619	0.399	0.349	1.000	0.680	0.481	0.474	0.268	0.349	0.404	0.491	0.048
Gal6	0.595	0.601	0.599	0.744	0.529	0.400	0.524	0.449	0.680	1.000	0.578	0.781	0.564	0.451	0.718	0.831	0.098
Gal7	0.966	0.969	0.968	0.806	0.549	0.245	0.856	0.783	0.481	0.578	1.000	0.741	0.434	0.772	0.740	0.679	0.066
Gal80	0.729	0.752	0.756	0.956	0.736	0.216	0.690	0.631	0.474	0.781	0.741	1.000	0.719	0.620	0.949	0.883	0.113
Hap4	0.417	0.444	0.445	0.713	0.921	0.067	0.502	0.543	0.268	0.564	0.434	0.719	1.000	0.538	0.718	0.629	0.148
Ho12	0.746	0.761	0.764	0.681	0.682	0.178	0.901	0.980	0.349	0.451	0.772	0.620	0.538	1.000	0.636	0.557	0.065
Ho17	0.714	0.739	0.739	0.950	0.766	0.177	0.696	0.649	0.404	0.718	0.740	0.949	0.718	0.636	1.000	0.862	0.117
Ho26	0.686	0.686	0.688	0.844	0.639	0.264	0.628	0.558	0.491	0.831	0.679	0.883	0.629	0.557	0.862	1.000	0.111
Ho28	0.063	0.067	0.064	0.108	0.136	0.014	0.064	0.064	0.048	0.098	0.066	0.113	0.148	0.065	0.117	0.111	1.000

Figure 5.15. Pearson correlation maps of binned metabolome data at 31st hour of extracellular samples

Adh6Δ-hxt10Δ-hxt12Δ-gal7Δ-snf3Δ-mba1Δ deletion mutants are highly correlated in Figure 5.15. According to Figure 5.15; *rgt1Δ* deletion mutant seems to be not correlated with *mig1Δ*, *hap4Δ* and *yel057cΔ*.

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.983	0.974	0.747	0.640	0.478	0.995	0.996	0.661	0.630	0.991	0.672	0.430	0.996	0.684	0.643	0.126
Hxt10	0.983	1.000	0.995	0.784	0.655	0.525	0.981	0.989	0.716	0.674	0.994	0.708	0.460	0.980	0.705	0.659	0.130
Hxt12	0.974	0.995	1.000	0.785	0.657	0.531	0.969	0.982	0.722	0.674	0.987	0.709	0.459	0.968	0.705	0.660	0.122
Mba1	0.747	0.784	0.785	1.000	0.957	0.900	0.763	0.760	0.930	0.965	0.768	0.984	0.898	0.748	0.972	0.945	0.176
Mig1	0.640	0.655	0.657	0.957	1.000	0.905	0.660	0.647	0.887	0.941	0.651	0.966	0.929	0.638	0.994	0.978	0.242
Rgt1	0.478	0.525	0.531	0.900	0.905	1.000	0.495	0.493	0.952	0.957	0.505	0.941	0.936	0.478	0.902	0.928	0.146
snf3	0.995	0.981	0.969	0.763	0.660	0.495	1.000	0.990	0.674	0.647	0.991	0.688	0.452	0.993	0.704	0.661	0.156
Yel057	0.996	0.989	0.982	0.760	0.647	0.493	0.990	1.000	0.678	0.644	0.991	0.684	0.444	0.997	0.691	0.647	0.120
Gal4	0.661	0.716	0.722	0.930	0.887	0.952	0.674	0.678	1.000	0.946	0.694	0.935	0.840	0.659	0.904	0.913	0.163
Gal6	0.630	0.674	0.674	0.965	0.941	0.957	0.647	0.644	0.946	1.000	0.655	0.986	0.929	0.632	0.949	0.963	0.150
Gal7	0.991	0.994	0.987	0.768	0.651	0.505	0.991	0.991	0.694	0.655	1.000	0.691	0.443	0.985	0.699	0.658	0.166
Gal80	0.672	0.708	0.709	0.984	0.966	0.941	0.688	0.684	0.935	0.986	0.691	1.000	0.934	0.672	0.972	0.971	0.171
Hap4	0.430	0.460	0.459	0.898	0.929	0.936	0.452	0.444	0.840	0.929	0.443	0.934	1.000	0.436	0.914	0.906	0.140
Ho12	0.996	0.980	0.968	0.748	0.638	0.478	0.993	0.997	0.659	0.632	0.985	0.672	0.436	1.000	0.681	0.638	0.119
Ho17	0.684	0.705	0.705	0.972	0.994	0.902	0.704	0.691	0.904	0.949	0.699	0.972	0.914	0.681	1.000	0.977	0.230
Ho26	0.643	0.659	0.660	0.945	0.978	0.928	0.661	0.647	0.913	0.963	0.658	0.971	0.906	0.638	0.977	1.000	0.233
Ho28	0.126	0.130	0.122	0.176	0.242	0.146	0.156	0.120	0.163	0.150	0.166	0.171	0.140	0.119	0.230	0.233	1.000

Figure 5.16. Pearson correlation maps of filtered metabolome data at 31st hour of extracellular samples

In Figure 5.16; *adh6Δ-hxt10Δ-hxt12Δ-gal7Δ-snf3Δ-yel057cΔ* deletion mutants are also highly correlated as in Figure 5.16.

5.4.2. Metabolic Fingerprinting Results

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.010	0.981	0.000	0.001	0.948	0.995	0.979	0.000	0.969	0.001	-0.001	0.955	0.007	0.001	0.994	0.981
Hxt10	0.010	1.000	0.012	0.167	0.169	0.019	0.010	0.013	0.174	0.016	0.154	0.120	0.017	0.786	0.174	0.011	0.006
Hxt12	0.981	0.012	1.000	0.000	0.001	0.968	0.973	0.985	0.000	0.988	0.000	-0.001	0.948	0.008	0.001	0.973	0.961
Mba1	0.000	0.167	0.000	1.000	0.996	0.001	0.000	0.000	0.995	0.001	0.990	0.914	0.001	0.277	0.995	0.000	-0.001
Mig1	0.001	0.169	0.001	0.996	1.000	0.002	0.001	0.000	0.996	0.002	0.986	0.910	0.002	0.278	0.996	0.000	-0.001
Rgt1	0.948	0.019	0.968	0.001	0.002	1.000	0.946	0.969	0.002	0.990	0.002	0.000	0.978	0.014	0.002	0.949	0.908
snf3	0.995	0.010	0.973	0.000	0.001	0.946	1.000	0.975	0.000	0.964	0.000	-0.001	0.955	0.007	0.001	0.998	0.976
Yel057	0.979	0.013	0.985	0.000	0.000	0.969	0.975	1.000	0.000	0.989	0.000	-0.001	0.937	0.009	0.000	0.980	0.935
Gal4	0.000	0.174	0.000	0.995	0.996	0.002	0.000	0.000	1.000	0.001	0.975	0.905	0.002	0.288	1.000	0.000	-0.001
Gal6	0.969	0.016	0.988	0.001	0.002	0.990	0.964	0.989	0.001	1.000	0.001	0.000	0.962	0.011	0.002	0.968	0.925
Gal7	0.001	0.154	0.000	0.990	0.986	0.002	0.000	0.000	0.975	0.001	1.000	0.910	0.001	0.265	0.976	0.000	-0.001
Gal80	-0.001	0.120	-0.001	0.914	0.910	0.000	-0.001	-0.001	0.905	0.000	0.910	1.000	0.000	0.112	0.903	-0.001	-0.001
Hap4	0.955	0.017	0.948	0.001	0.002	0.978	0.955	0.937	0.002	0.962	0.001	0.000	1.000	0.013	0.002	0.953	0.937
Ho12	0.007	0.786	0.008	0.277	0.278	0.014	0.007	0.009	0.288	0.011	0.265	0.112	0.013	1.000	0.290	0.007	0.004
Ho17	0.001	0.174	0.001	0.995	0.996	0.002	0.001	0.000	1.000	0.002	0.976	0.903	0.002	0.290	1.000	0.001	0.000
Ho26	0.994	0.011	0.973	0.000	0.000	0.949	0.998	0.980	0.000	0.968	0.000	-0.001	0.953	0.007	0.001	1.000	0.966
Ho28	0.981	0.006	0.961	-0.001	-0.001	0.908	0.976	0.935	-0.001	0.925	-0.001	-0.001	0.937	0.004	0.000	0.966	1.000

Figure 5.17. Pearson correlation maps of binned metabolome data at 20th hour of intracellular samples

The results in Figure 5.17 are questionable in some aspects. For example, *hxt10Δ-hxt12Δ* deletion mutant pair should have been correlated but the correlation coefficient is too low. In addition, *gal4Δ-gal6Δ* and *hoΔ* deletion mutants should also have been correlated because they have a similar function in galactose pathway (*gal4v-gal6Δ*), or *ho12Δ-ho17Δ-ho26Δ* are same strains.

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.340	0.978	0.287	0.296	0.945	0.999	0.972	0.261	0.964	0.292	0.141	0.962	0.999	0.262	0.998	0.985
Hxt10	0.340	1.000	0.433	0.807	0.816	0.513	0.345	0.428	0.790	0.485	0.979	0.755	0.437	0.349	0.792	0.346	0.305
Hxt12	0.978	0.433	1.000	0.367	0.379	0.969	0.974	0.985	0.335	0.988	0.377	0.179	0.948	0.977	0.337	0.975	0.956
Mba1	0.287	0.807	0.367	1.000	0.997	0.427	0.291	0.358	0.994	0.403	0.872	0.899	0.368	0.296	0.994	0.291	0.266
Mig1	0.296	0.816	0.379	0.997	1.000	0.445	0.300	0.370	0.991	0.418	0.872	0.888	0.382	0.305	0.990	0.300	0.272
Rgt1	0.945	0.513	0.969	0.427	0.445	1.000	0.947	0.970	0.403	0.990	0.441	0.219	0.972	0.948	0.406	0.951	0.906
snf3	0.999	0.345	0.974	0.291	0.300	0.947	1.000	0.973	0.265	0.963	0.297	0.146	0.965	0.999	0.266	0.999	0.983
Yel057	0.972	0.428	0.985	0.358	0.370	0.970	0.973	1.000	0.327	0.990	0.367	0.166	0.934	0.970	0.329	0.978	0.928
Gal4	0.261	0.790	0.335	0.994	0.991	0.403	0.265	0.327	1.000	0.376	0.859	0.905	0.350	0.270	0.999	0.265	0.238
Gal6	0.964	0.485	0.988	0.403	0.418	0.990	0.963	0.990	0.376	1.000	0.421	0.207	0.956	0.963	0.378	0.968	0.920
Gal7	0.292	0.979	0.377	0.872	0.872	0.441	0.297	0.367	0.859	0.421	1.000	0.859	0.380	0.302	0.859	0.296	0.268
Gal80	0.141	0.755	0.179	0.899	0.888	0.219	0.146	0.166	0.905	0.207	0.859	1.000	0.211	0.151	0.902	0.142	0.138
Hap4	0.962	0.437	0.948	0.368	0.382	0.972	0.965	0.934	0.350	0.956	0.380	0.211	1.000	0.969	0.352	0.963	0.950
Ho12	0.999	0.349	0.977	0.296	0.305	0.948	0.999	0.970	0.270	0.963	0.302	0.151	0.969	1.000	0.271	0.998	0.987
Ho17	0.262	0.792	0.337	0.994	0.990	0.406	0.266	0.329	0.999	0.378	0.859	0.902	0.352	0.271	1.000	0.266	0.239
Ho26	0.998	0.346	0.975	0.291	0.300	0.951	0.999	0.978	0.265	0.968	0.296	0.142	0.963	0.998	0.266	1.000	0.975
Ho28	0.985	0.305	0.956	0.266	0.272	0.906	0.983	0.928	0.238	0.920	0.268	0.138	0.950	0.987	0.239	0.975	1.000

Figure 5.18. Pearson correlation maps of filtered metabolome data at 20th hour of intracellular samples

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.630	0.694	0.626	0.653	0.989	0.941	0.928	0.911	0.653	0.615	0.633	0.956	0.700	0.621	0.958	0.696
Hxt10	0.630	1.000	0.951	0.806	0.859	0.578	0.768	0.803	0.809	0.859	0.970	0.879	0.774	0.944	0.932	0.762	0.974
Hxt12	0.694	0.951	1.000	0.927	0.905	0.646	0.830	0.857	0.852	0.905	0.920	0.921	0.828	0.941	0.906	0.807	0.959
Mba1	0.626	0.806	0.927	1.000	0.970	0.573	0.749	0.776	0.763	0.970	0.848	0.997	0.741	0.867	0.834	0.725	0.890
Mig1	0.653	0.859	0.905	0.970	1.000	0.612	0.777	0.797	0.787	1.000	0.811	0.985	0.767	0.860	0.816	0.748	0.891
Rgt1	0.989	0.578	0.646	0.573	0.612	1.000	0.937	0.917	0.904	0.612	0.560	0.583	0.938	0.663	0.576	0.934	0.654
snf3	0.941	0.768	0.830	0.749	0.777	0.937	1.000	0.995	0.986	0.777	0.748	0.757	0.976	0.838	0.769	0.959	0.826
Yel057	0.928	0.803	0.857	0.776	0.797	0.917	0.995	1.000	0.991	0.797	0.782	0.782	0.978	0.864	0.799	0.961	0.853
Gal4	0.911	0.809	0.852	0.763	0.787	0.904	0.986	0.991	1.000	0.787	0.788	0.770	0.977	0.877	0.801	0.959	0.869
Gal6	0.653	0.859	0.905	0.970	1.000	0.612	0.777	0.797	0.787	1.000	0.811	0.985	0.767	0.860	0.816	0.748	0.891
Gal7	0.615	0.970	0.920	0.848	0.811	0.560	0.748	0.782	0.788	0.811	1.000	0.838	0.763	0.959	0.978	0.754	0.940
Gal80	0.633	0.879	0.921	0.997	0.985	0.583	0.757	0.782	0.770	0.985	0.838	1.000	0.747	0.865	0.830	0.731	0.891
Hap4	0.956	0.774	0.828	0.741	0.767	0.938	0.976	0.978	0.977	0.767	0.763	0.747	1.000	0.852	0.764	0.993	0.829
Ho12	0.700	0.944	0.941	0.867	0.860	0.663	0.838	0.864	0.877	0.860	0.959	0.865	0.852	1.000	0.950	0.834	0.949
Ho17	0.621	0.932	0.906	0.834	0.816	0.576	0.769	0.799	0.801	0.816	0.978	0.830	0.764	0.950	1.000	0.746	0.922
Ho26	0.958	0.762	0.807	0.725	0.748	0.934	0.959	0.961	0.959	0.748	0.754	0.731	0.993	0.834	0.746	1.000	0.814
Ho28	0.696	0.974	0.959	0.890	0.891	0.654	0.826	0.853	0.869	0.891	0.940	0.891	0.829	0.949	0.922	0.814	1.000

Figure 5.19. Pearson correlation maps of binned metabolome data at 25th hour of intracellular samples

There is no significant correlation observed in Figure 5.18. Even the possible clusterings like *hxt10Δ-hxt12Δ* appear uncorrelated. So, it is not possible to discriminate any of the possible clusters at 20th hour of intracellular metabolome data.

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.554	0.639	0.673	0.739	0.988	0.933	0.920	0.902	0.739	0.539	0.688	0.955	0.672	0.529	0.960	0.729
Hxt10	0.554	1.000	0.978	0.864	0.803	0.507	0.681	0.718	0.717	0.803	1.000	0.853	0.684	0.965	0.975	0.675	0.857
Hxt12	0.639	0.978	1.000	0.887	0.838	0.596	0.759	0.789	0.780	0.838	0.974	0.881	0.754	0.964	0.958	0.745	0.870
Mba1	0.673	0.864	0.887	1.000	0.944	0.626	0.818	0.853	0.848	0.944	0.863	0.998	0.805	0.874	0.882	0.792	0.975
Mig1	0.739	0.803	0.838	0.944	1.000	0.714	0.880	0.903	0.895	1.000	0.801	0.959	0.853	0.861	0.857	0.835	0.954
Rgt1	0.988	0.507	0.596	0.626	0.714	1.000	0.934	0.914	0.900	0.714	0.492	0.647	0.939	0.643	0.494	0.935	0.687
snf3	0.933	0.681	0.759	0.818	0.880	0.934	1.000	0.997	0.986	0.880	0.670	0.833	0.976	0.809	0.687	0.960	0.861
Yel057	0.920	0.718	0.789	0.853	0.903	0.914	0.997	1.000	0.990	0.903	0.709	0.866	0.976	0.836	0.723	0.961	0.893
Gal4	0.902	0.717	0.780	0.848	0.895	0.900	0.986	0.990	1.000	0.895	0.709	0.861	0.975	0.842	0.720	0.958	0.902
Gal6	0.739	0.803	0.838	0.944	1.000	0.714	0.880	0.903	0.895	1.000	0.801	0.959	0.853	0.861	0.857	0.835	0.954
Gal7	0.539	1.000	0.974	0.863	0.801	0.492	0.670	0.709	0.709	0.801	1.000	0.852	0.673	0.963	0.976	0.663	0.856
Gal80	0.688	0.853	0.881	0.998	0.959	0.647	0.833	0.866	0.861	0.959	0.852	1.000	0.816	0.872	0.883	0.801	0.975
Hap4	0.955	0.684	0.754	0.805	0.853	0.939	0.976	0.976	0.975	0.853	0.673	0.816	1.000	0.810	0.663	0.996	0.858
Ho12	0.672	0.965	0.964	0.874	0.861	0.643	0.809	0.836	0.842	0.861	0.963	0.872	0.810	1.000	0.941	0.796	0.896
Ho17	0.529	0.975	0.958	0.882	0.857	0.494	0.687	0.723	0.720	0.857	0.976	0.883	0.663	0.941	1.000	0.644	0.871
Ho26	0.960	0.675	0.745	0.792	0.835	0.935	0.960	0.961	0.958	0.835	0.663	0.801	0.996	0.796	0.644	1.000	0.847
Ho28	0.729	0.857	0.870	0.975	0.954	0.687	0.861	0.893	0.902	0.954	0.856	0.975	0.858	0.896	0.871	0.847	1.000

Figure 5.20. Pearson correlation maps of filtered metabolome data at 25th hour of intracellular samples

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.817	0.697	0.979	0.950	0.998	0.892	0.887	0.779	0.992	0.858	0.987	0.730	0.858	0.832	0.850	0.731
Hxt10	0.817	1.000	0.714	0.861	0.802	0.835	0.848	0.845	0.685	0.850	0.893	0.856	0.690	0.848	0.916	0.850	0.704
Hxt12	0.697	0.714	1.000	0.741	0.737	0.711	0.692	0.685	0.922	0.718	0.855	0.731	0.988	0.869	0.774	0.881	0.992
Mba1	0.979	0.861	0.741	1.000	0.971	0.988	0.901	0.893	0.757	0.994	0.915	0.998	0.759	0.899	0.884	0.899	0.762
Mig1	0.950	0.802	0.737	0.971	1.000	0.956	0.860	0.852	0.748	0.960	0.865	0.967	0.771	0.882	0.825	0.877	0.761
Rgt1	0.998	0.835	0.711	0.988	0.956	1.000	0.901	0.895	0.779	0.997	0.878	0.993	0.739	0.872	0.853	0.867	0.740
snf3	0.892	0.848	0.692	0.901	0.860	0.901	1.000	1.000	0.696	0.905	0.863	0.906	0.669	0.917	0.883	0.900	0.675
Yel057	0.887	0.845	0.685	0.893	0.852	0.895	1.000	1.000	0.691	0.899	0.857	0.899	0.661	0.912	0.880	0.894	0.667
Gal4	0.779	0.685	0.922	0.757	0.748	0.779	0.696	0.691	1.000	0.766	0.787	0.762	0.938	0.803	0.720	0.805	0.942
Gal6	0.992	0.850	0.718	0.994	0.960	0.997	0.905	0.899	0.766	1.000	0.892	0.998	0.741	0.881	0.867	0.877	0.744
Gal7	0.858	0.893	0.855	0.915	0.865	0.878	0.863	0.857	0.787	0.892	1.000	0.904	0.837	0.956	0.970	0.969	0.849
Gal80	0.987	0.856	0.731	0.998	0.967	0.993	0.906	0.899	0.762	0.998	0.904	1.000	0.751	0.892	0.877	0.890	0.754
Hap4	0.730	0.690	0.988	0.759	0.771	0.739	0.669	0.661	0.938	0.741	0.837	0.751	1.000	0.855	0.746	0.866	0.996
Ho12	0.858	0.848	0.869	0.899	0.882	0.872	0.917	0.912	0.803	0.881	0.956	0.892	0.855	1.000	0.923	0.996	0.854
Ho17	0.832	0.916	0.774	0.884	0.825	0.853	0.883	0.880	0.720	0.867	0.970	0.877	0.746	0.923	1.000	0.928	0.760
Ho26	0.850	0.850	0.881	0.899	0.877	0.867	0.900	0.894	0.805	0.877	0.969	0.890	0.866	0.996	0.928	1.000	0.867
Ho28	0.731	0.704	0.992	0.762	0.761	0.740	0.675	0.667	0.942	0.744	0.849	0.754	0.996	0.854	0.760	0.867	1.000

Figure 5.21. Pearson correlation maps of binned metabolome data at 31st hour of intracellular samples

	Adh6	Hxt10	Hxt12	Mba1	Mig1	Rgt1	snf3	Yel057	Gal4	Gal6	Gal7	Gal80	Hap4	Ho12	Ho17	Ho26	Ho28
Adh6	1.000	0.891	0.726	0.977	0.944	0.998	0.998	0.996	0.836	0.992	0.875	0.985	0.736	0.808	0.913	0.787	0.739
Hxt10	0.891	1.000	0.854	0.929	0.863	0.909	0.894	0.891	0.862	0.916	0.989	0.922	0.834	0.860	0.997	0.864	0.855
Hxt12	0.726	0.854	1.000	0.768	0.765	0.739	0.714	0.707	0.948	0.745	0.899	0.756	0.995	0.976	0.826	0.989	0.999
Mba1	0.977	0.929	0.768	1.000	0.966	0.987	0.968	0.962	0.817	0.994	0.927	0.999	0.771	0.808	0.943	0.799	0.775
Mig1	0.944	0.863	0.765	0.966	1.000	0.951	0.927	0.920	0.802	0.955	0.877	0.962	0.787	0.810	0.878	0.794	0.774
Rgt1	0.998	0.909	0.739	0.987	0.951	1.000	0.995	0.993	0.836	0.997	0.895	0.993	0.746	0.811	0.929	0.793	0.750
snf3	0.998	0.894	0.714	0.968	0.927	0.995	1.000	1.000	0.835	0.987	0.871	0.977	0.721	0.801	0.917	0.778	0.727
Yel057	0.996	0.891	0.707	0.962	0.920	0.993	1.000	1.000	0.833	0.983	0.865	0.972	0.713	0.797	0.914	0.773	0.720
Gal4	0.836	0.862	0.948	0.817	0.802	0.836	0.835	0.833	1.000	0.825	0.879	0.820	0.951	0.991	0.849	0.984	0.956
Gal6	0.992	0.916	0.745	0.994	0.955	0.997	0.987	0.983	0.825	1.000	0.906	0.998	0.749	0.805	0.935	0.790	0.754
Gal7	0.875	0.989	0.899	0.927	0.877	0.895	0.871	0.865	0.879	0.906	1.000	0.916	0.882	0.888	0.979	0.898	0.898
Gal80	0.985	0.922	0.756	0.999	0.962	0.993	0.977	0.972	0.820	0.998	0.916	1.000	0.760	0.807	0.938	0.795	0.765
Hap4	0.736	0.834	0.995	0.771	0.787	0.746	0.721	0.713	0.951	0.749	0.882	0.760	1.000	0.981	0.808	0.989	0.997
Ho12	0.808	0.860	0.976	0.808	0.810	0.811	0.801	0.797	0.991	0.805	0.888	0.807	0.981	1.000	0.842	0.997	0.982
Ho17	0.913	0.997	0.826	0.943	0.878	0.929	0.917	0.914	0.849	0.935	0.979	0.938	0.808	0.842	1.000	0.842	0.828
Ho26	0.787	0.864	0.989	0.799	0.794	0.793	0.778	0.773	0.984	0.790	0.898	0.795	0.989	0.997	0.842	1.000	0.992
Ho28	0.739	0.855	0.999	0.775	0.774	0.750	0.727	0.720	0.956	0.754	0.898	0.765	0.997	0.982	0.828	0.992	1.000

Figure 5.22. Pearson correlation maps of filtered metabolome data at 31st hour of intracellular samples

Mig1Δ-rgt1-snf3Δ-yel057cΔ-gal4Δ-gal6Δ-gal80Δ deletion mutants are correlated with respect to each other in Figures 5.19-5.22. However, one can not find any uncorrelated deletion mutant this could not be possible in any circumstances. Nevertheless, these figures give more satisfactorily results than the other metabolic fingerprinting results given in Figures 5.17-5.22.

COLOR CODE	
0.9-1.0	
0.8-0.899	
0.7-0.799	
0.6-0.699	
0.5-0.599	
0.4-0.499	
0.3-0.399	
0.2-0.299	
0.1-0.199	
0 -0.099	

Figure 5.23. Color Code for Pearson Correlation Coefficients

Metabolites show a remarkable biological variability in their abundance, even when repeatedly measured for identical genotypes under identical experimental conditions.

Importantly, these variations are not independent, but metabolites often tend to vary concertedly with other metabolites. This variability is a result of biological causes, reflecting the intrinsic flexibility of metabolic networks. Even under approximately constant experimental conditions, yeast metabolism is a highly dynamic system, responding to (small) stimuli within seconds or milliseconds. For example, slight differences in temperature or glucose uptake will induce differences in certain metabolite levels, which in turn affect other metabolites. Similarly slight differences in enzyme concentrations or metabolic oscillations, among many other factors, contribute to the variability in metabolite levels. The results are metabolic fluctuations, which propagate through metabolic reaction chains and ultimately create an emergent and experimentally observable pattern of correlations (Weckwerth, 2003; Weckwerth *et al.*, 2004 a,b; Morgenthal *et al.*, 2005).

Figures 5.9 through 5.22 visualize the correlation profile for metabolome data of different yeast mutants at different time intervals of sampling and at different cellular location (intracellular or extracellular). The correlation between two strains A and B was quantified using the Pearson correlation $C(A, B)$

$$C(A, B) = \frac{\Gamma_{AB}}{\sqrt{\Gamma_{AA}\Gamma_{BB}}} \quad C(A, B) = \frac{\Gamma_{AB}}{\sqrt{\Gamma_{AA}\Gamma_{BB}}} \quad (5.1)$$

where $\Gamma_{AB} = \langle ab \rangle - \langle a \rangle \langle b \rangle$ denotes the respective element of the covariance matrix. The observed patterns of correlations, as depicted in Figures 5.9 through 5.22, represent the interconnectivity of metabolome profiles with respect to their pair-wise correlations (according to time). However, correlations per se contain no information about direction or causality within the underlying biochemical network of reactions. Nonetheless, the observed correlations are a direct consequence of the reactions and enzymatic regulations that underpin metabolism. In this sense, the pattern of correlations can be interpreted as a ‘footprint’ or ‘fingerprint’ of the underlying system and provide additional information about the specific physiological state of the tissue or organism at a given point in time.

It is emphasized that the covariance and correlation matrices are the basis for many, if not most, current methods of data analysis, including various clustering algorithms, PCA and ICA. Therefore, a deeper understanding of how different physiological states or experimental conditions manifest itself as altered patterns of correlations is of paramount interest beyond the comparative approach presented here.

Experimental results demonstrated that differences between the deletion mutants of the *S. cerevisiae* are indeed reflected on the correlation network level. While most correlations of extracellular metabolome data are preserved across the strains, intracellular metabolites do not show pronounced differential correlations.

In conclusion, most remarkable, in some cases correlation coefficients do not only slightly change, but the correlation may also be completely reversed, as, e.g. observed in the intracellular metabolome data at 31st hour of fermentation. While the correlation coefficient values of intracellular samples at 20th hour remain under 0.1 which means that they are not correlated, the values at 31st hour are close to 1 or highly correlated. This must be attributed to marked changes in the regulation of the underlying metabolic pathways. The comparative analysis of correlations allows to pinpoint key-points at which metabolic regulation is changed and supplements the more commonly used approach to look for changes in the average levels only. Importantly, this approach can be readily extended to include data on transcript levels and protein expression. A similar analysis of these integrative data matrices will further contribute to an understanding of complex biochemical systems, reaching beyond the view of metabolism as a sum of linear metabolic pathways.

5.5. GO Term Analysis

Table 5.7 and 5.8 summarize the number of genes that have been directly annotated to the process and function term ontologies and their *p* values obtained from high-throughput experiments. The gene groups in the tables help us to discover what the genes may have in common. For example, *gal4Δ-gal6Δ-gal80Δ-mig1Δ* cluster obtained by ICA for extracellular metabolome data at 20th hour of fermentation gives a *p* value of 0.00037. This value is

relatively small to accept the biological relevance of these genes being in the same process. Other possible clusters which were obtained by PCA and ICA are *gal4Δ-gal6Δ*; *gal80Δ-hap4Δ*; *gal4Δ-gal6Δ-mig1Δ-rgt1Δ*; and *gal6Δ-gal80Δ-adh6Δ-mig1Δ*, respectively

Table 5.7. Go Terms from the Process Ontology

Gene Ontology term	P-value	Genes annotated to the term
alcohol metabolic process	1.75e-05	ADH6, GAL4, GAL7, GAL80, MIG1, RGT1
galactose metabolic process	2.34e-05	GAL4, GAL7, GAL80
hexose metabolic process	2.44e-05	GAL4, GAL7, GAL80, MIG1, RGT1
monosaccharide metabolic process	3.61e-05	GAL4, GAL7, GAL80, MIG1, RGT1
carbohydrate metabolic process	0.00012	GAL4, GAL7, GAL80, HAP4, MIG1, RGT1
cellular carbohydrate metabolic process	0.00196	GAL4, GAL7, GAL80, MIG1, RGT1
regulation of transcription by galactose	0.00278	GAL4, GAL80
response to chemical stimulus	0.00037	GAL4, GAL80, GAL6, MIG1

Table 5.8. Go Terms from the Function Ontology

Gene Ontology term	P-value	Genes annotated to the term
transcription regulator activity	0.00046	GAL4, GAL6, GAL80, HAP4, MIG1, RGT1
transcriptional activator activity	0.00400	GAL4, HAP4, RGT1
transcription regulator activity	0.00991	GAL80, HAP4
transcription regulator activity	7.38e-05	GAL4, GAL6, MIG1, RGT1

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

In the present study two novel methods, metabolic footprinting and fingerprinting, were used for the characterisation of yeast deletion mutants as well as for the function annotation of *yel057cΔ* gene. The global measurement of metabolites consumed from or secreted into the growth medium by cultures of *S. cerevisiae* was performed by using direct injection MS. The samples of culture media were collected at 4 different time intervals for footprinting study, and at 3 different time intervals for fingerprinting study. These stationary phase samples were then injected directly into an electrospray ionisation mass spectrometer. Then, Mass Spectrometer data of each deletion mutant were used to estimate the possible clusterings. The major conclusions that can be drawn from this study are given as follows:

1) The major discriminations of deletion mutants were identified at 20th hour of fermentation using extracellular metabolites and using the binning approach for the preprocessing of MS data. At this time interval, the clusterings of *galΔ* mutants, *hxtΔ* and *hoΔ* mutants are satisfactorily obtained by PCA and ICA. In order to confirm the clusterings, GO Term analysis is performed for the mutants which are in close region with respect to each other.

2) Metabolite fingerprinting results at 25th hour of fermentation using the binning approach are slightly better than those obtained at other sampling hours. The clusterings of *hoΔ*, *hxtΔ*, and regulation related mutants are identified by ICA. In addition, neither PCA nor filtering approach is able to extract other data sets.

3) The filtering approach did not reveal the hidden data as good as the binning approach. Possible explanation for this situation could be the lack of sufficient data. In all circumstances, the common mass variables did not exceed 150 out of 1000. In other words, specific mass variables of each deletion mutant could have been eliminated in the filtering approach.

4) ICA was helpful to identify the complexities encountered in PCA results. In general, it could be quite useful to confirm the PCA results and it widens the possible clusterings from each other.

5) HCA was additionally performed on the PCA and ICA scores so that grouping relationships might be better visualized. Clearly, the greatest separation is between galactose metabolism and hexose transporter genes at 20th hour of extracellular metabolome data (Figure 17), which form two separate branches in the resulting dendrogram. *Gal*Δ mutants, respiration related deletion mutants (*hap4*Δ, *mba1*Δ) and *snf3*Δ, *hxt12*Δ, *yel057c*Δ are found in close proximity in agreement with PCA and ICA results.

6) *Yel057c*Δ deletion mutant appears in the same region with *snf3*Δ, *adh6*Δ, and *hxt12*Δ in Figures 13-15. This outcome should be verified with different experiments but according to this result, it could be stated that *yel057c*Δ would play a role in glucose or hexose transport like *snf3*Δ and *hxt12*Δ.

“Metabolic footprinting” approach for assessing the metabolic status of cells is simpler than monitoring intracellular metabolites because it avoids the problems of metabolite turnover, the need to quench metabolism, and the requirement to separate intracellular material from the extracellular space. Similar to footprinting, in the metabolic fingerprinting approach the intention is not to identify each observed metabolite, but to compare patterns or “fingerprints” of metabolites that change in response to environmental or genetic alterations.

In conclusion, these two methods (metabolic footprinting and fingerprinting) are successfully applied to functional analysis and characterisation of yeast deletion strains used in the present study. The two preprocessing approaches of filtering and binning of metabolome data show promise but also have limitations. The clusterings obtained cannot be interpreted as direct biochemical or regulatory interactions, but rather as general influences. However, the clusterings and directions of the influences generally agree with those known for galactose utilization and transport pathway mechanisms. These PCA and ICA methods do not produce a

physiological model of gene regulation, but serve as data mining tools for finding tentative, hitherto unknown, relations.

6.2.Recommendations

In metabolite fingerprinting the discrimination is not satisfactory as it was in metabolic footprinting. The reason for this could be the difficulties of intracellular metabolite extraction procedures, and some problems in MS-TOF equipment. Alternative way in extraction procedure could be searched to avoid the loss of specific metabolite data of each single gene deletions.

Discriminant function analysis which is a supervised method could be used instead of ICA (unsupervised method), and the discrimination power of both supervised and unsupervised methods could be compared. However, in DFA a preprocessing approach should be used like in PCA to decrease the dimension of metabolome data matrix.

Alternatively, sampling rate across 20th hour of fermentation could be increased to provide data at the early times of stationary phase. These data may be helpful to unravel the possible clusterings.

APPENDIX A: HCA RESULTS OF PREPROCESSING OF METABOLOME DATA BY BINNING

A.1. PCA of Binned Extracellular Metabolome Data at 15th hour

Dendrogram using Average Linkage (Between Groups)

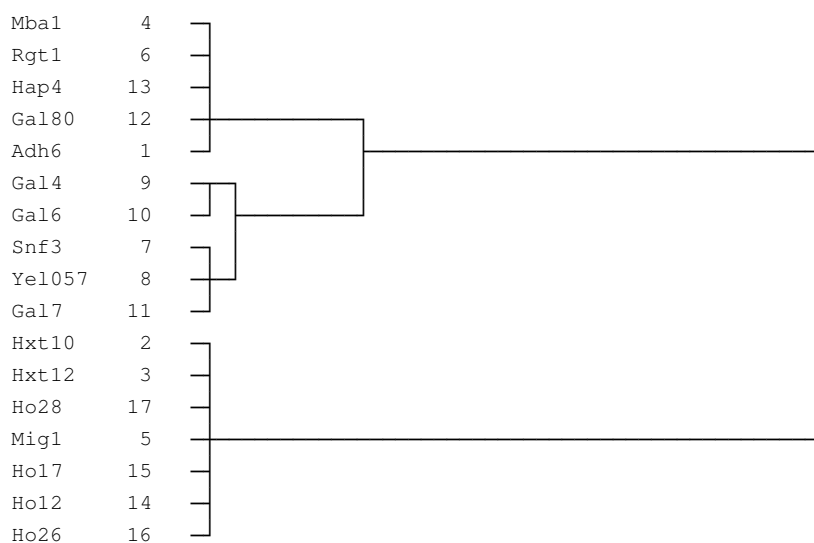


Figure A.1. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

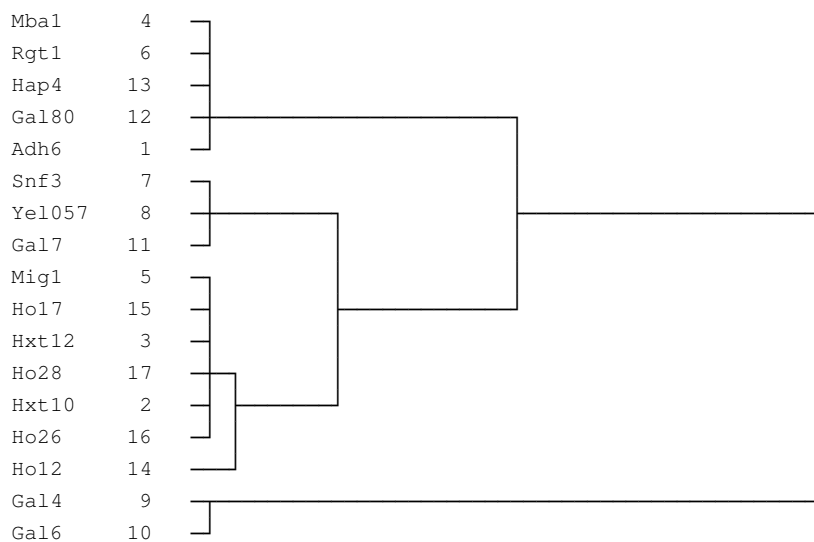


Figure A.2. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

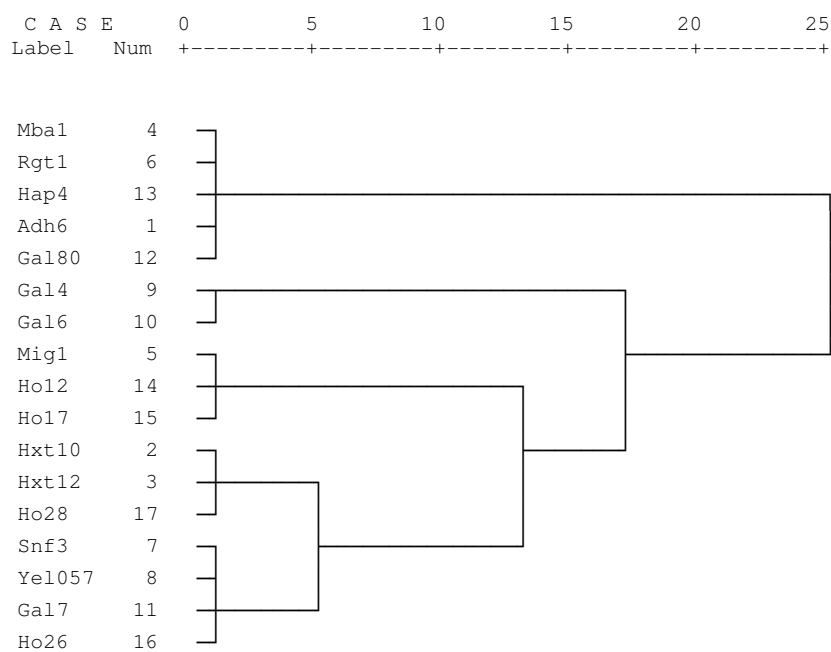


Figure A.3. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

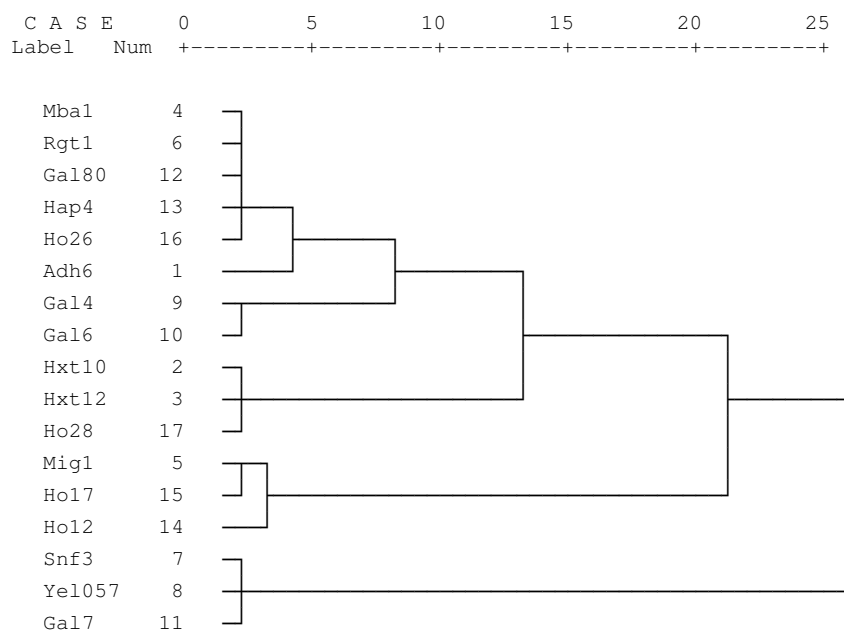


Figure A.4. HCA of PC3 and PC4

A.2. ICA of Binned Extracellular Metabolome Data at 15th hour

Dendrogram using Average Linkage (Between Groups)

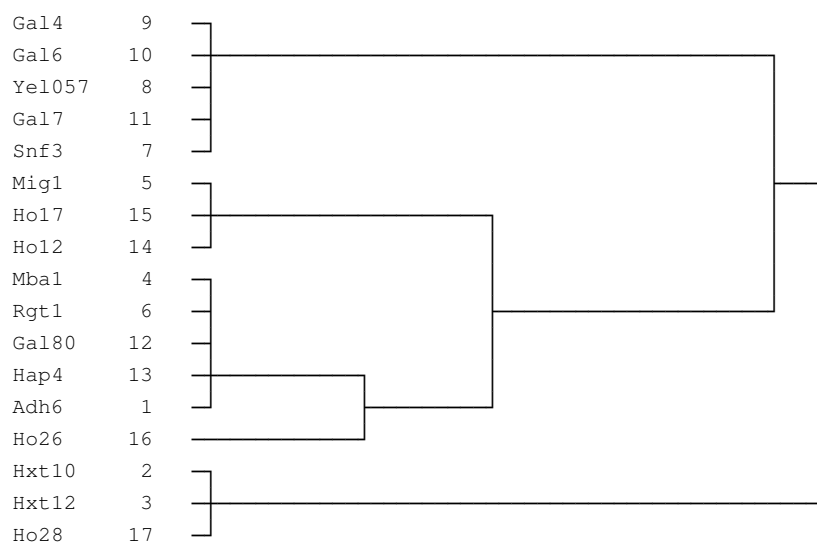


Figure A.5. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

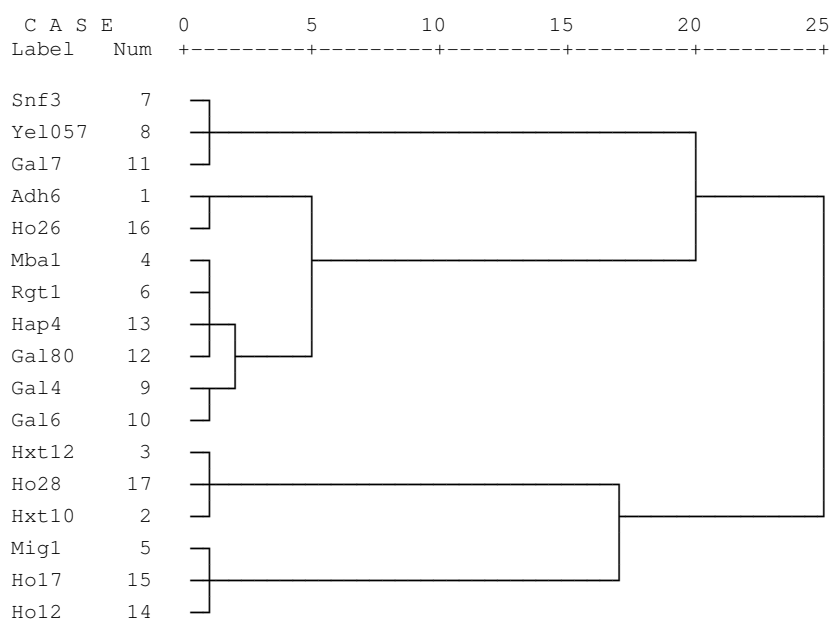


Figure A.6. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

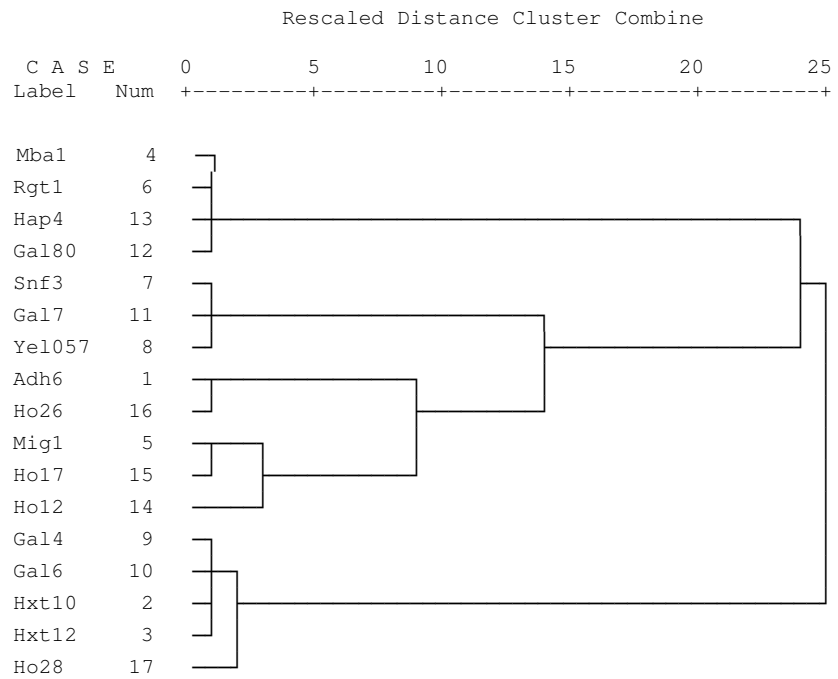


Figure A.7. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

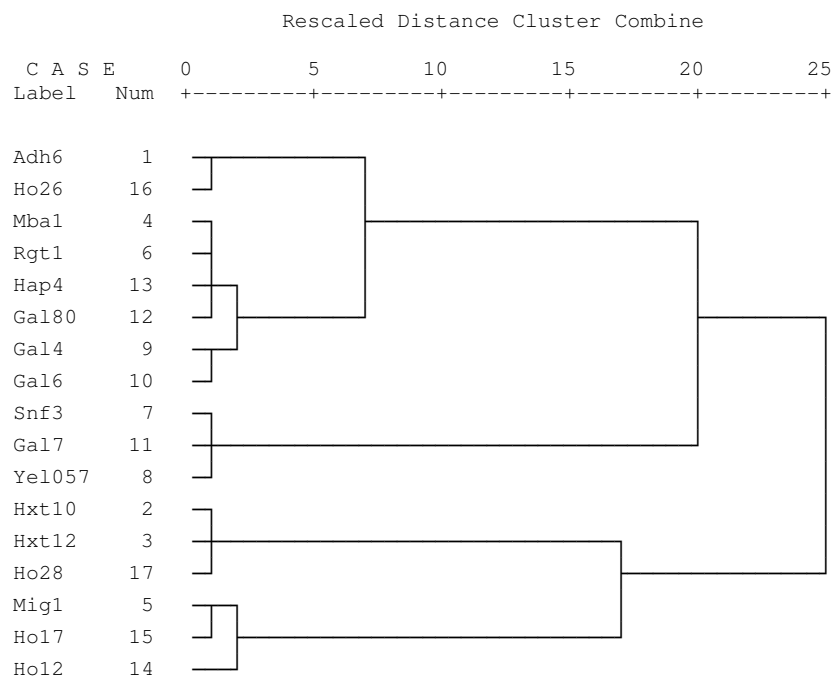


Figure A.8. HCA of IC2 and IC3 by 6 times pre-processing of PCA

A.3. PCA of Binned Extracellular Metabolome Data at 20th hour

Dendrogram using Average Linkage (Between Groups)

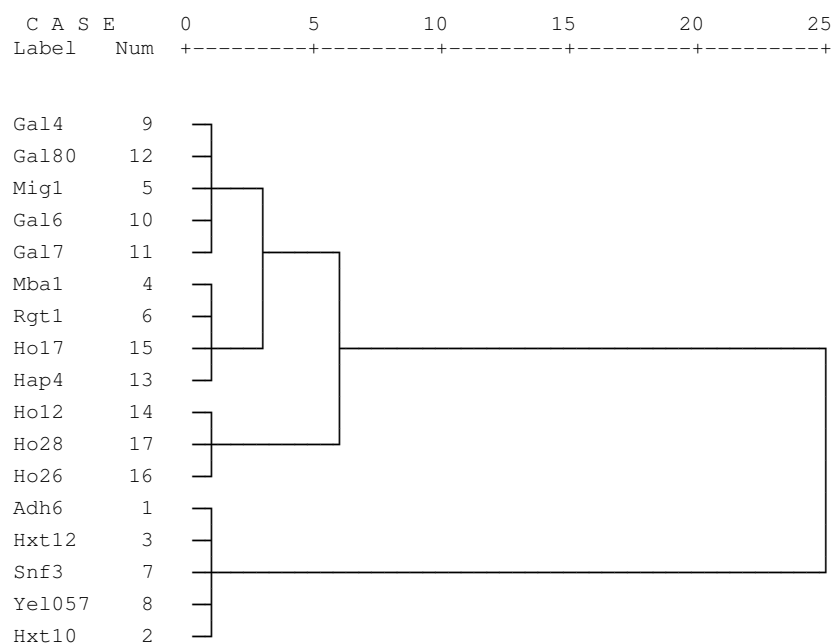


Figure A.9. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

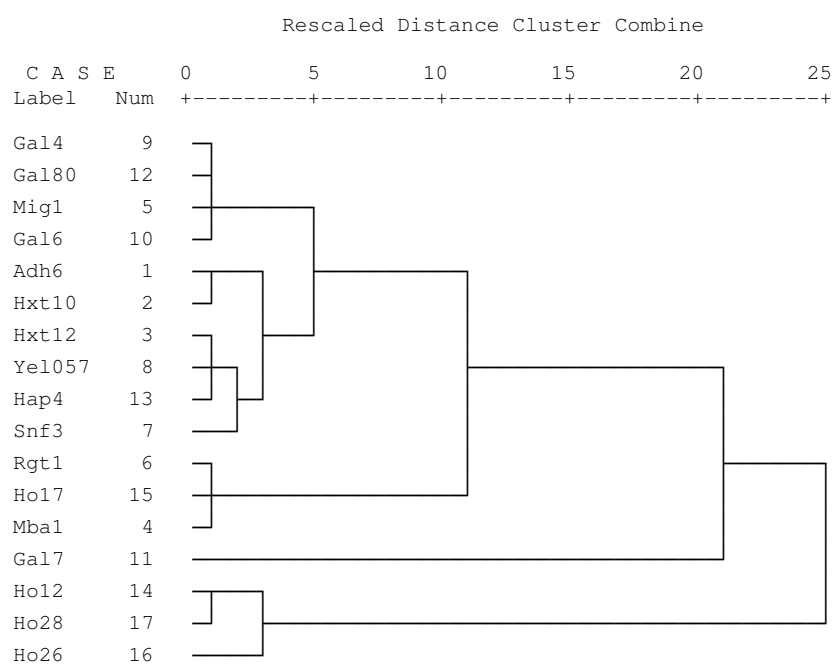


Figure A.10. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

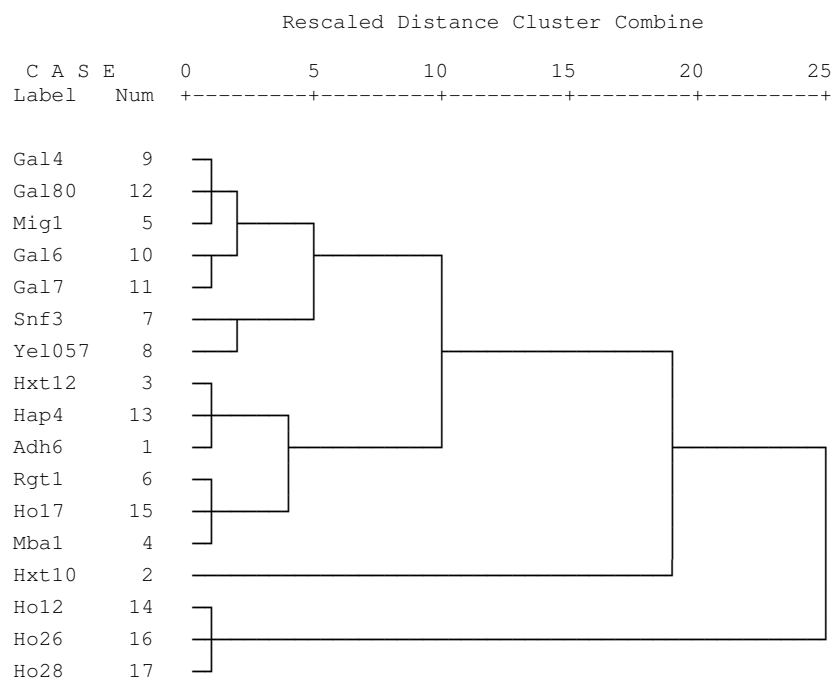


Figure A.11. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

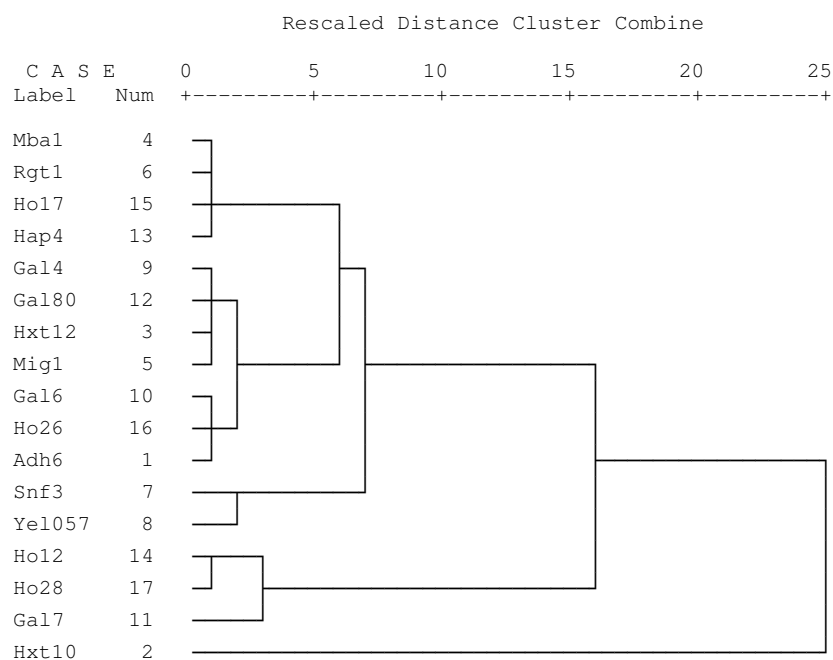


Figure A.12. HCA of PC3 and PC4

A.4. ICA of Binned Extracellular Metabolome Data at 20th hour

Dendrogram using Average Linkage (Between Groups)

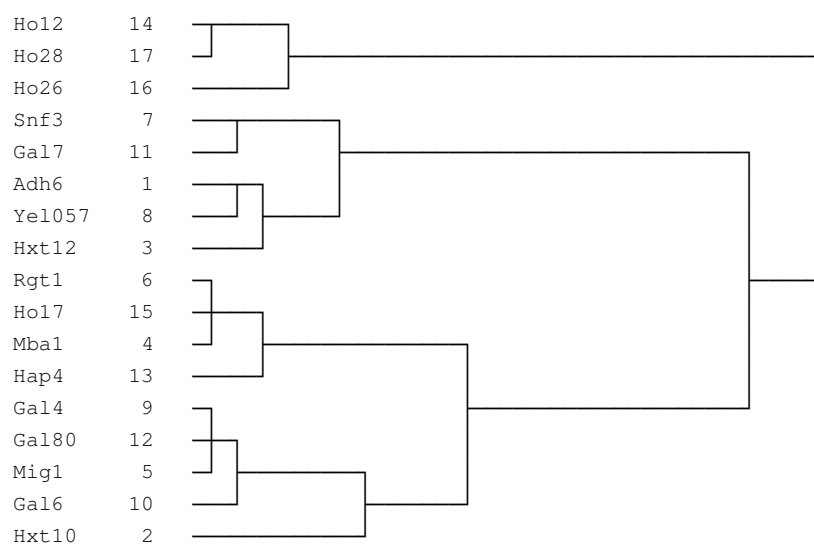


Figure A.13. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

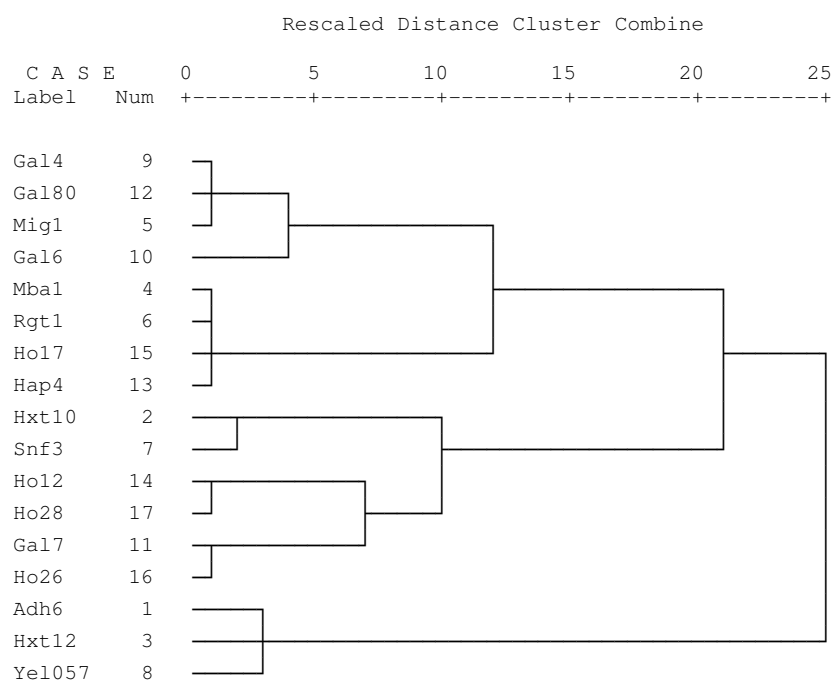


Figure A.14. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

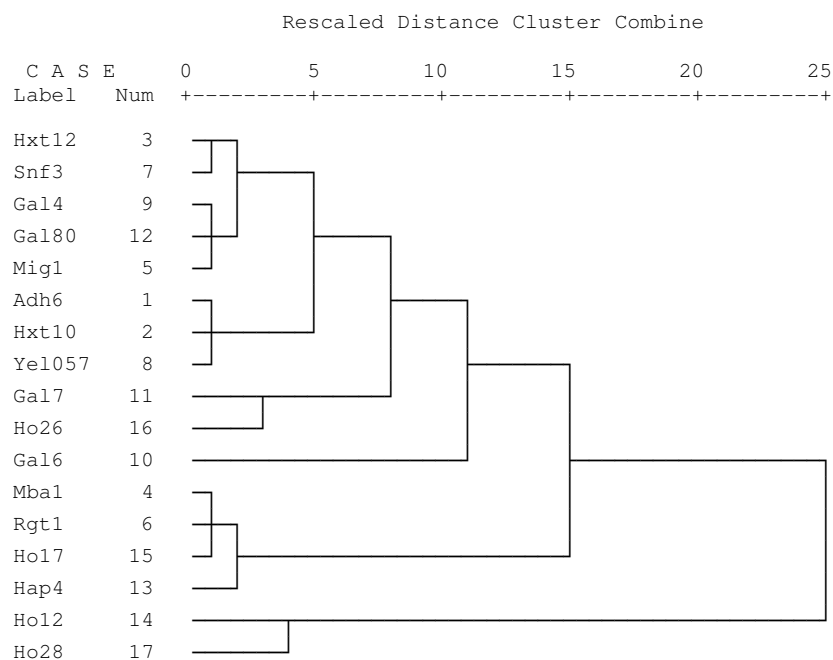


Figure A.15. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

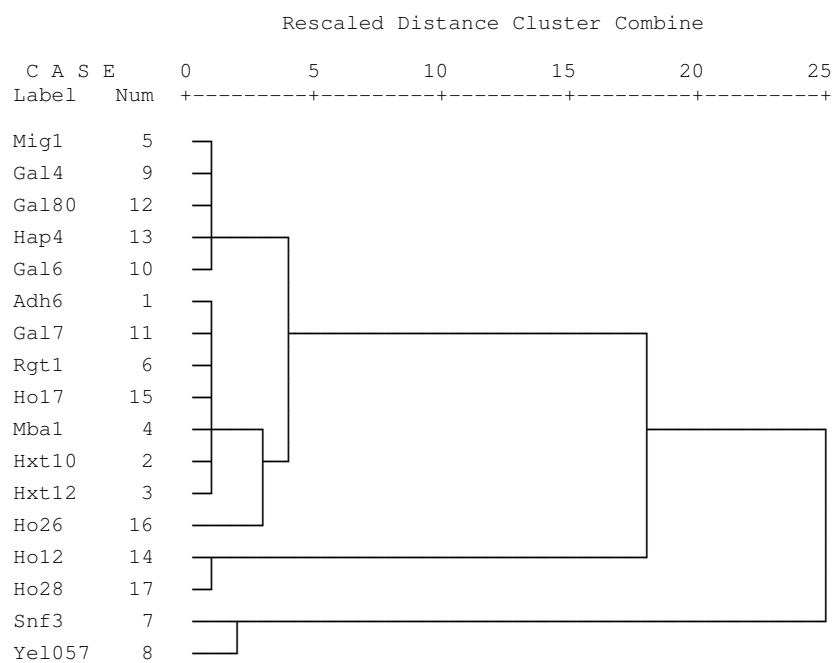


Figure A.16. HCA of IC2 and IC3 by 6 times pre-processing of PCA

A.5. PCA of Binned Extracellular Metabolome Data at 25th hour

Dendrogram using Average Linkage (Between Groups)

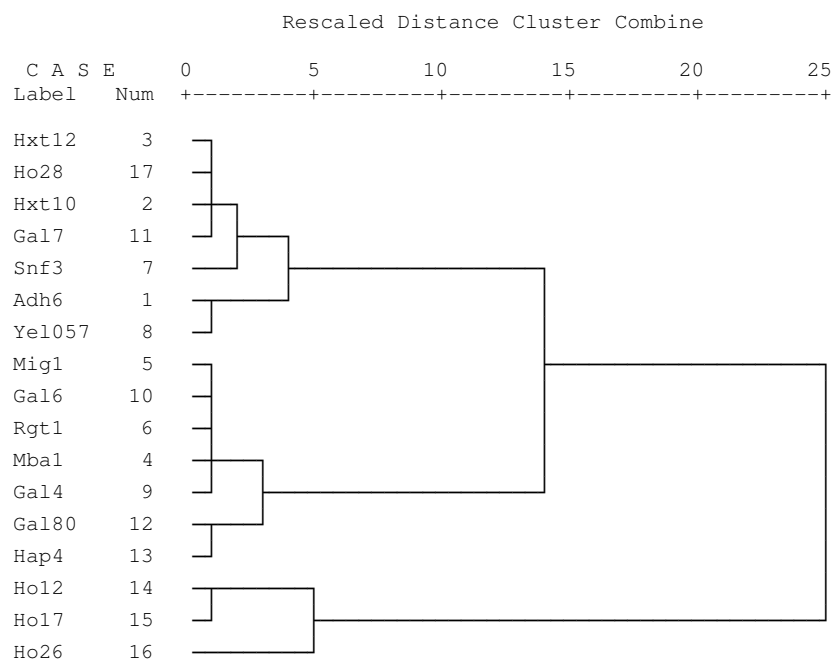


Figure A.17. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

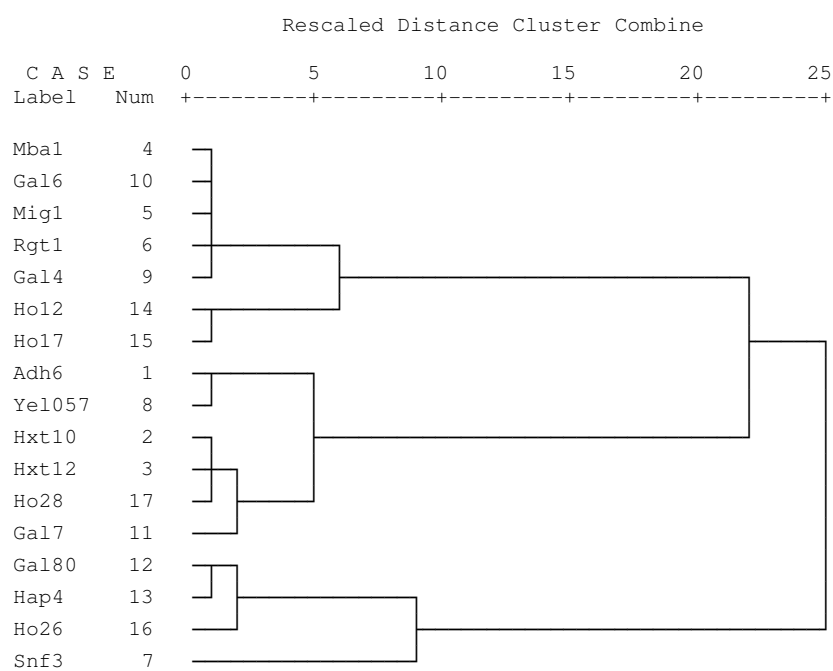


Figure A.18. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

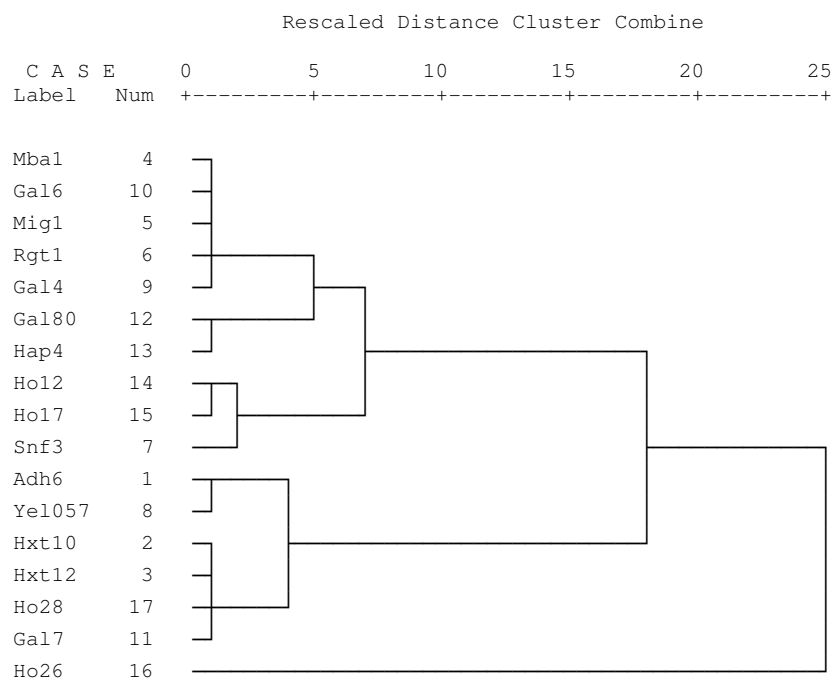


Figure A.19. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

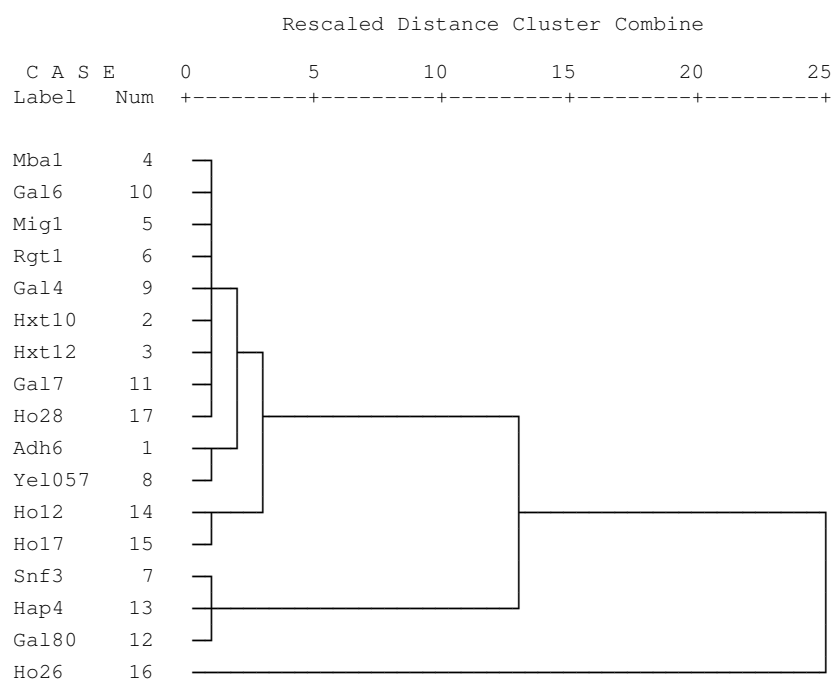


Figure A.20. HCA of PC3 and PC4

A.6. ICA of Binned Extracellular Metabolome Data at 25th hour

Dendrogram using Average Linkage (Between Groups)

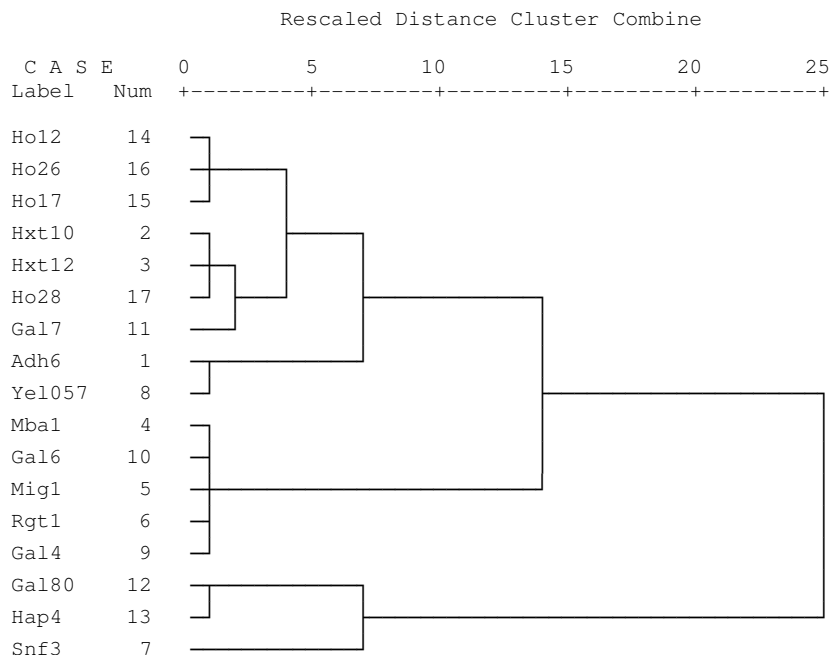


Figure A.21. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

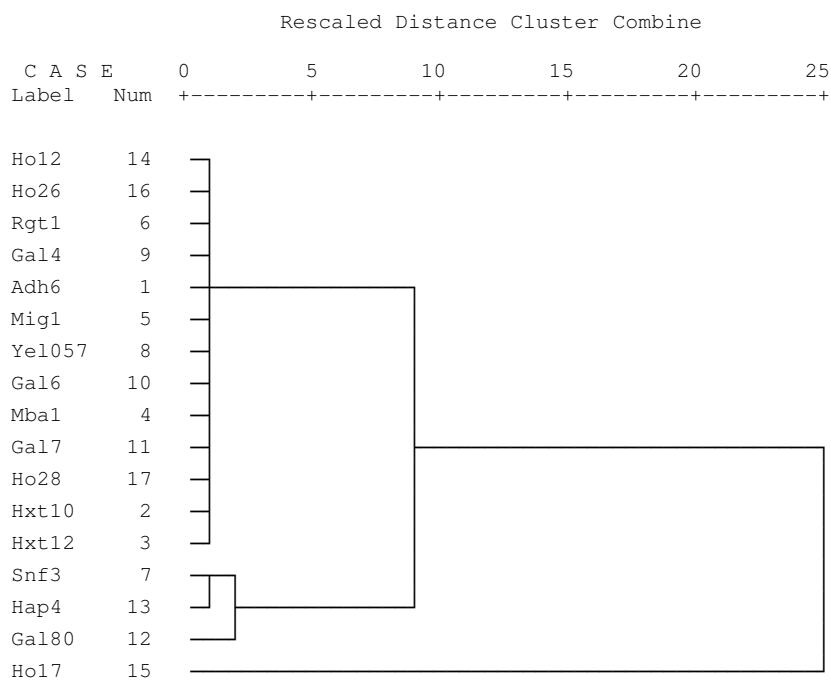


Figure A.22. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

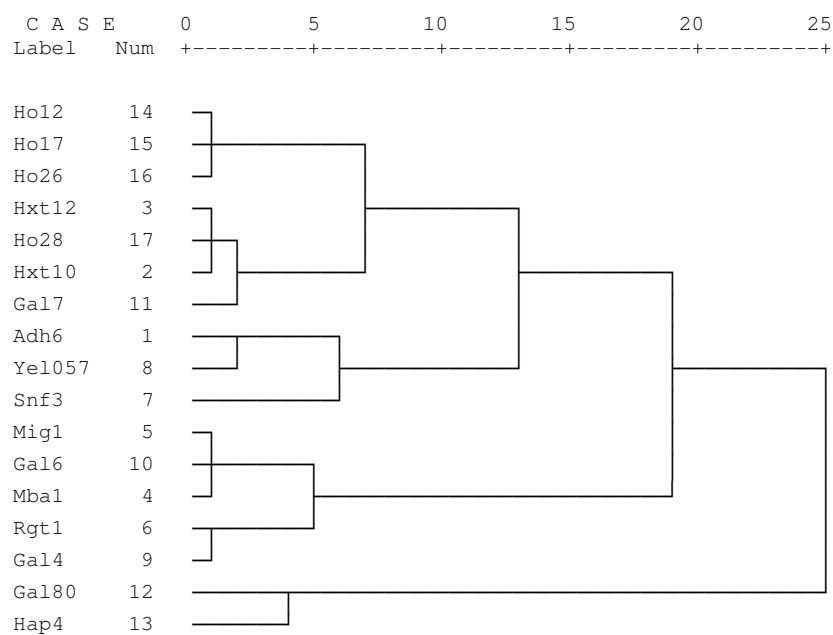


Figure A.23. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

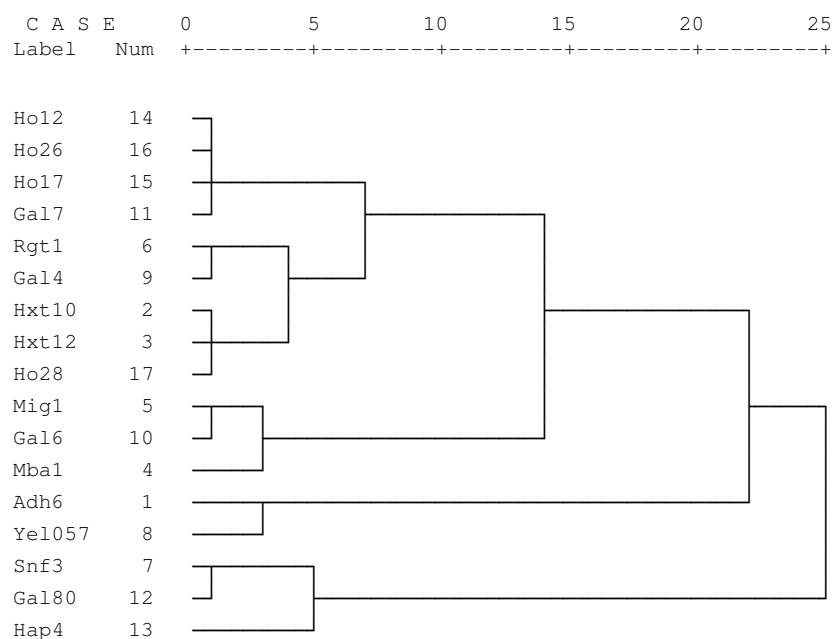


Figure A.24. HCA of IC2 2and IC3 by 6 times pre-processing of PCA

A.7. PCA of Binned Extracellular Metabolome Data at 31st hour

Dendrogram using Average Linkage (Between Groups)

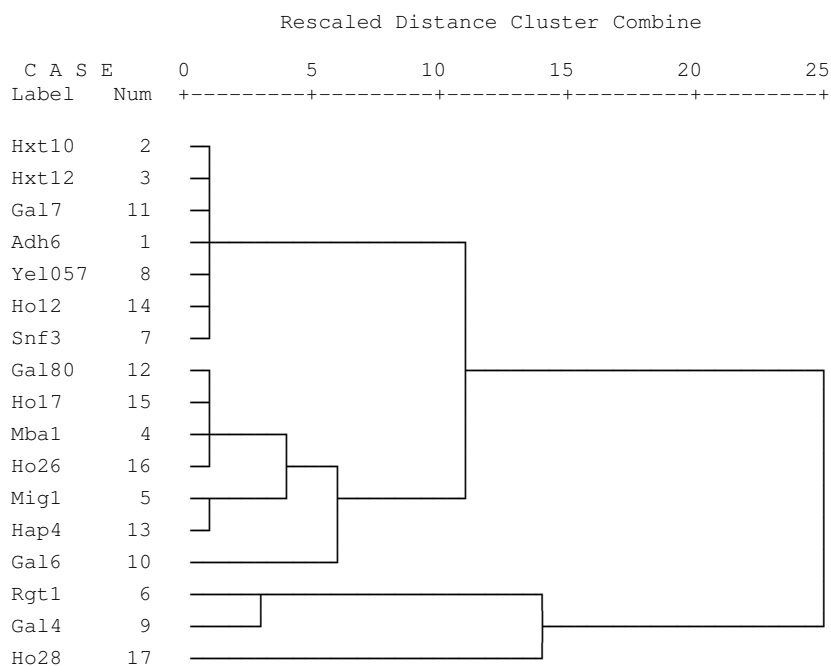


Figure A.25. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

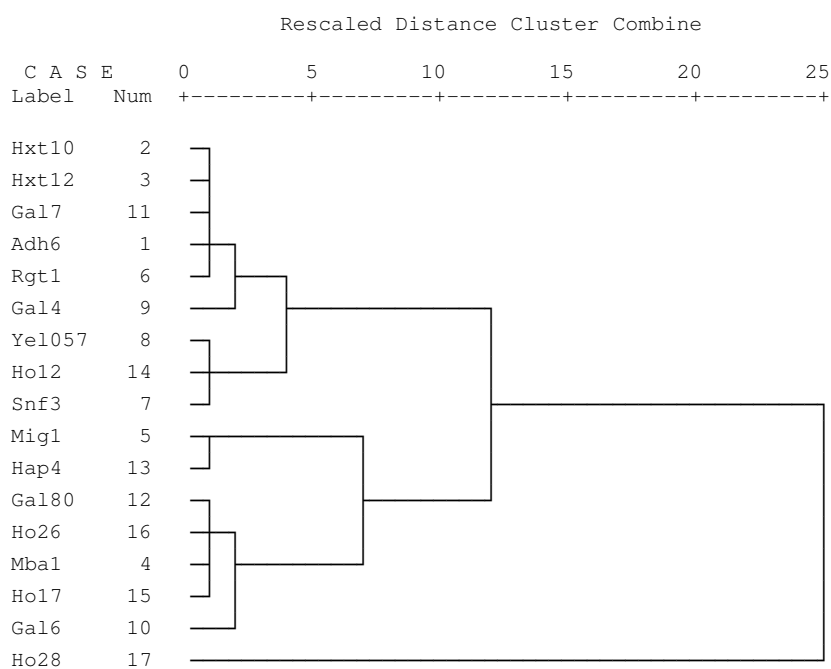


Figure A.26. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

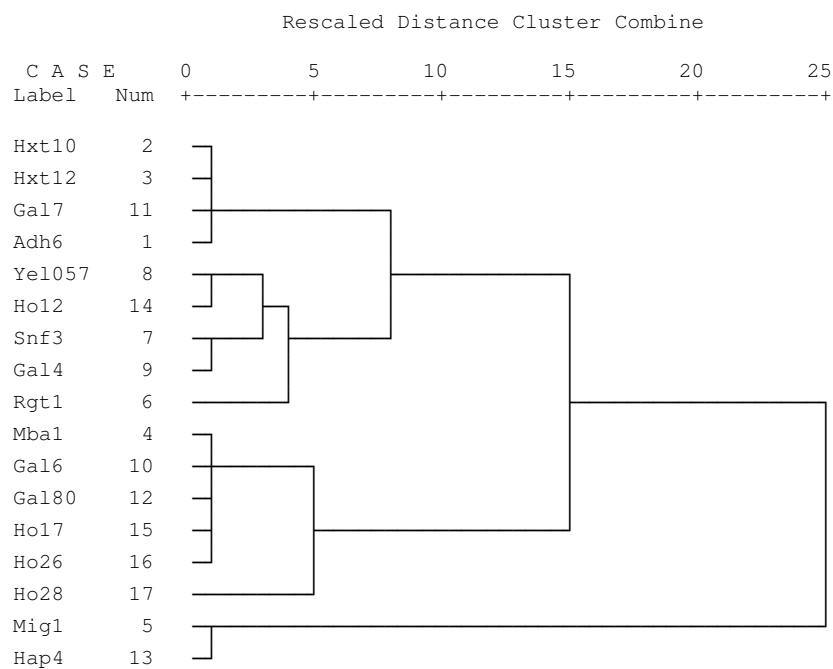


Figure A.27. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

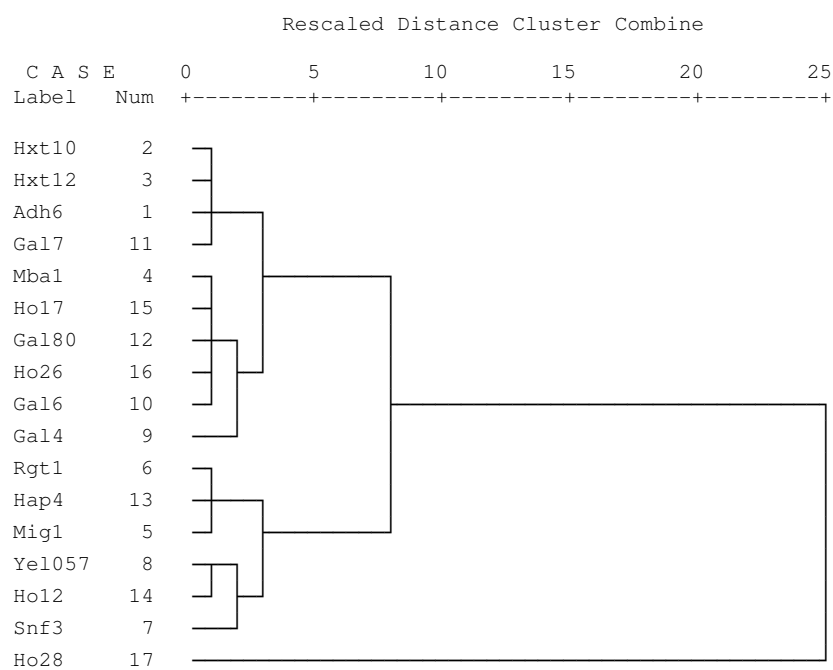


Figure A.28. HCA of PC3 and PC4

A.8. ICA of Binned Extracellular Metabolome Data at 31st hour

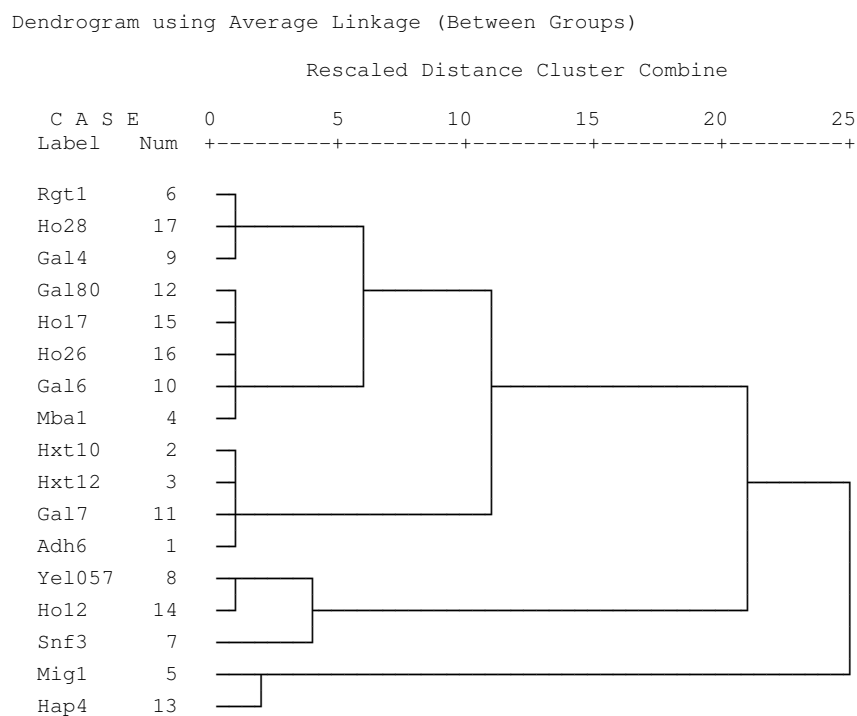


Figure A.29. HCA of IC1 and IC2 by 5 times pre-processing of PCA

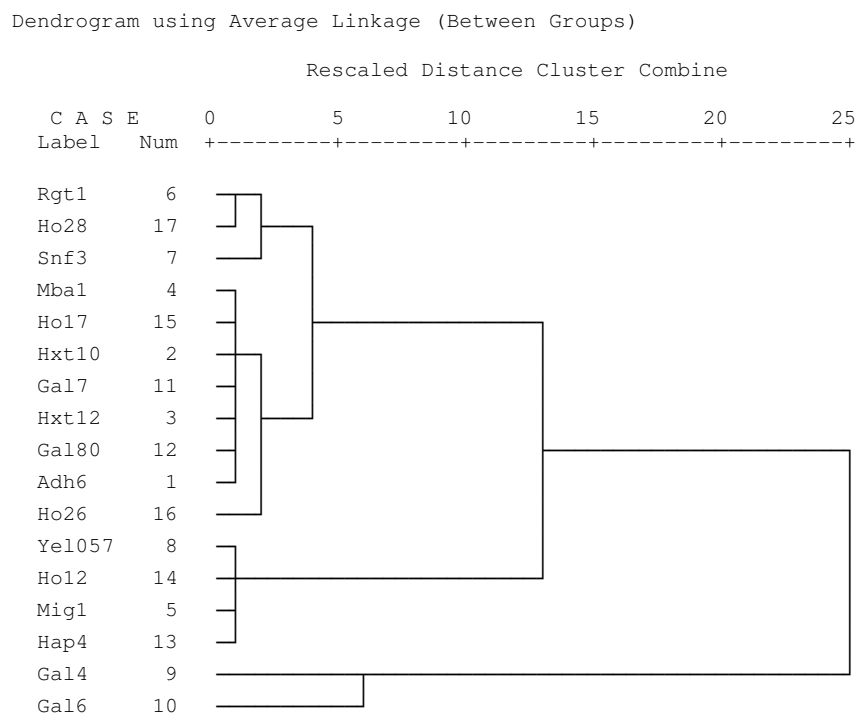


Figure A.30. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

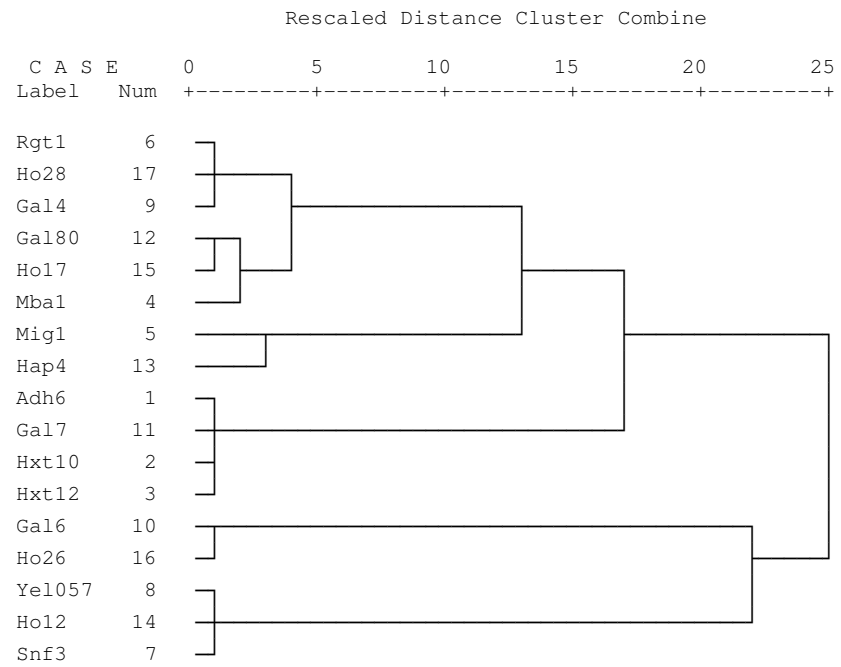


Figure A.31. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

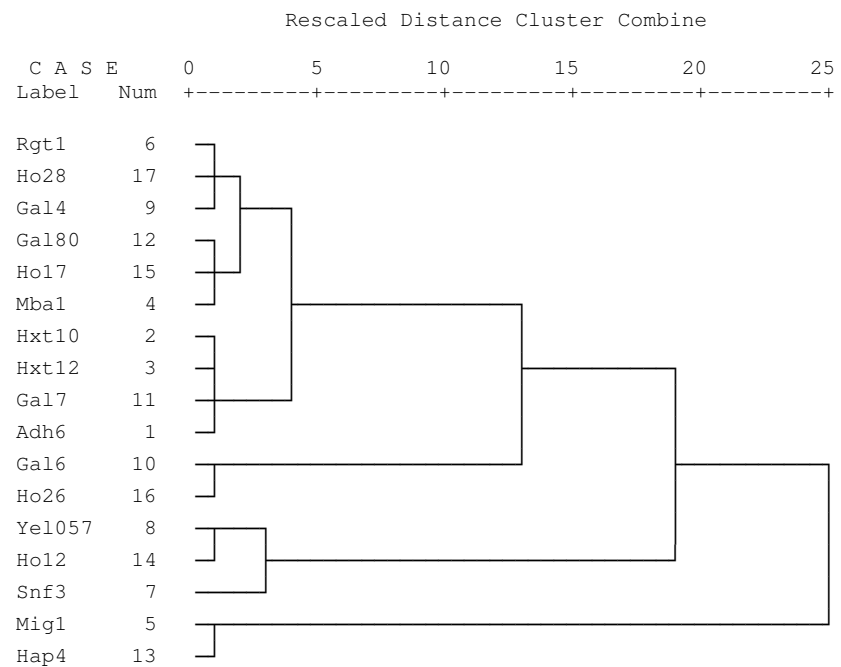


Figure A.32. HCA of IC2 and IC3 by 6 times pre-processing of PCA

A.9. PCA of Binned Intracellular Metabolome Data at 20th hour

Dendrogram using Average Linkage (Between Groups)

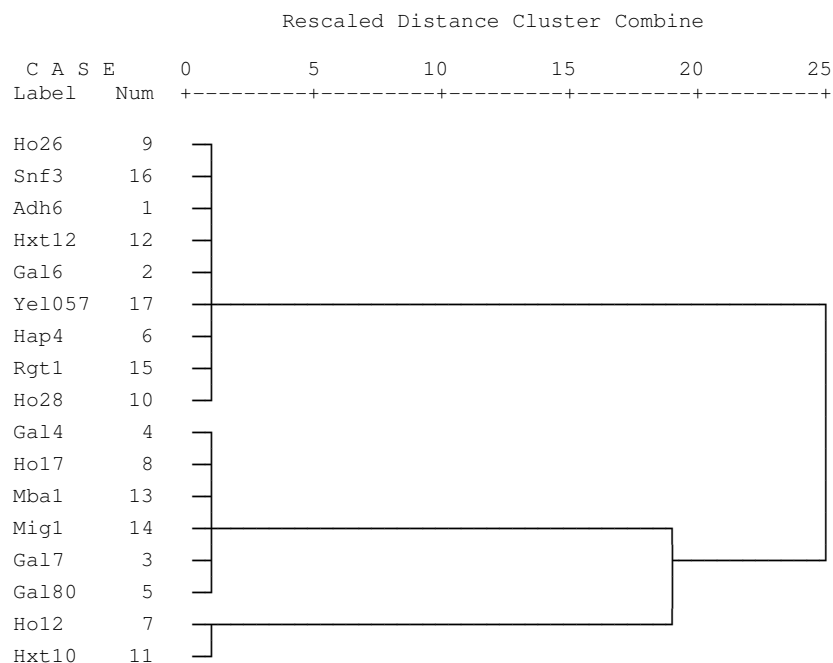


Figure A.33. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

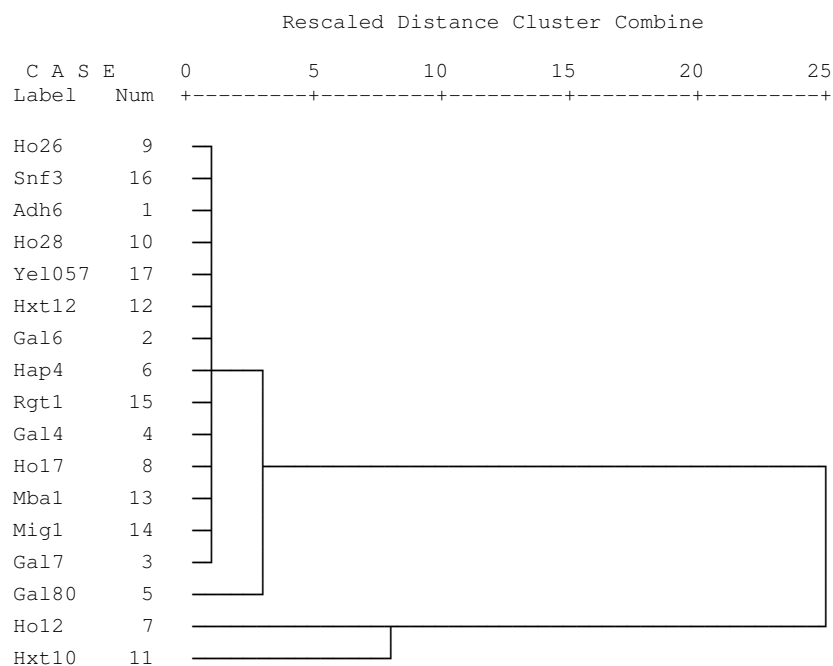


Figure A.34. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

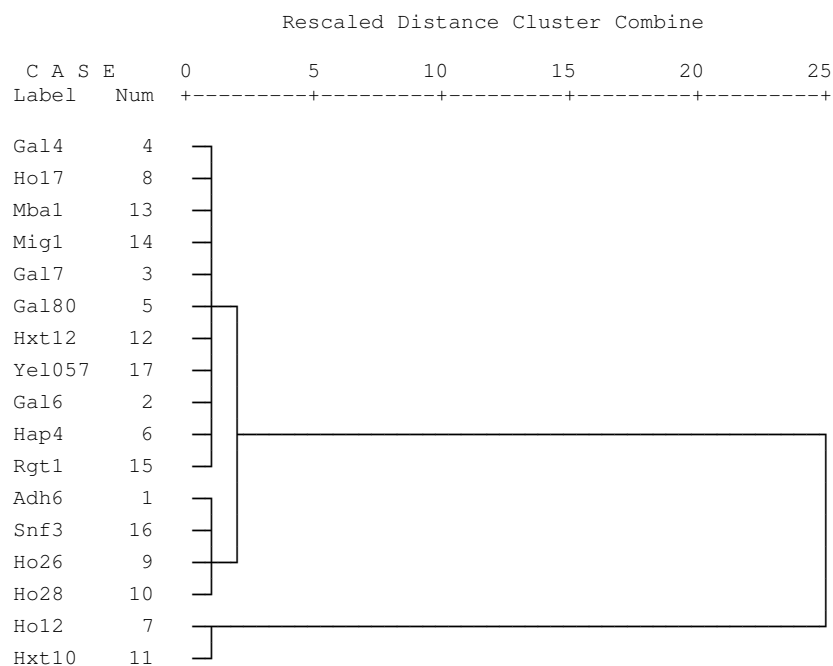


Figure A.35. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

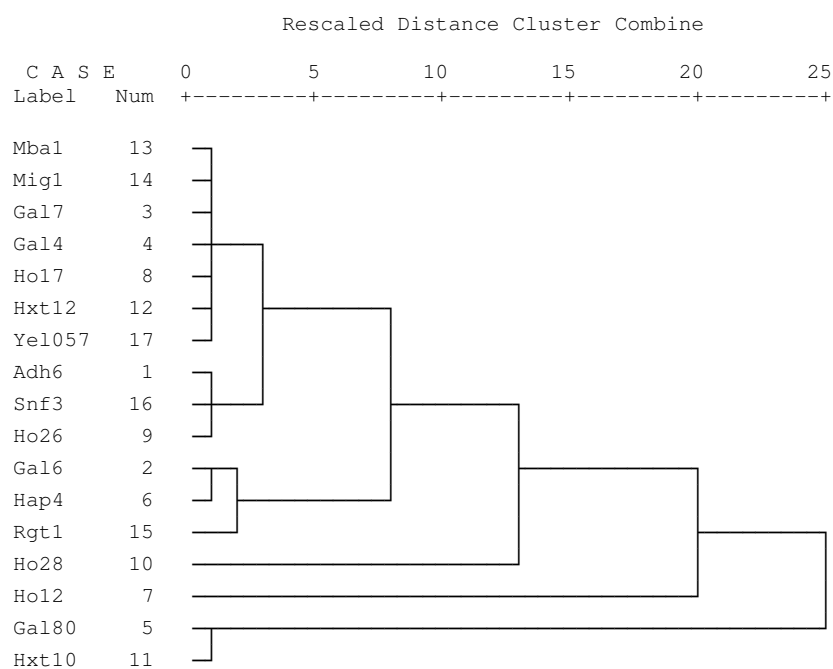


Figure A.36. HCA of PC3 and PC4

A.10. ICA of Binned Intracellular Metabolome Data at 20th hour

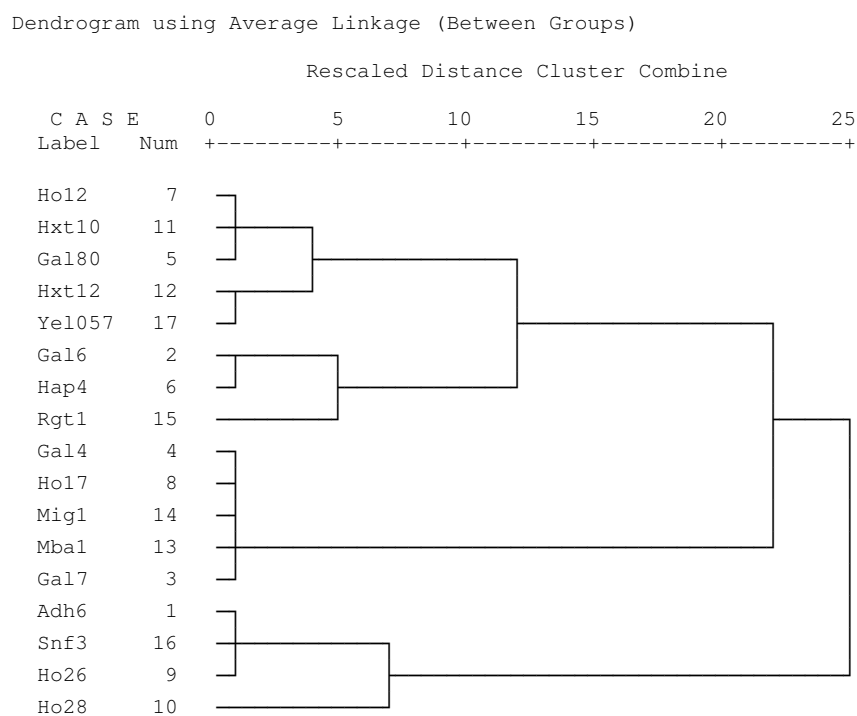


Figure A.37. HCA of IC1 and IC2 by 5 times pre-processing of PCA

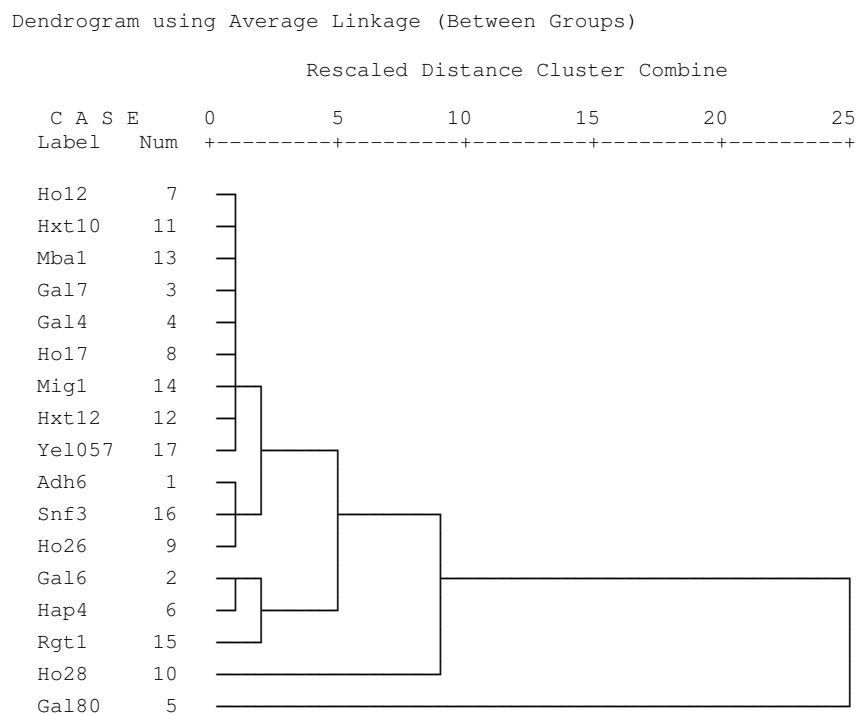


Figure A.38. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

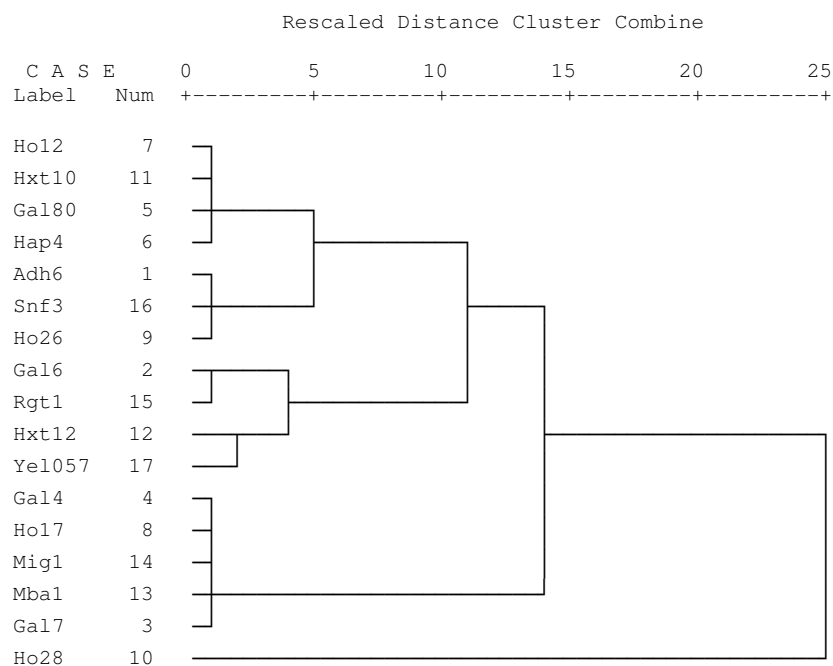


Figure A.39. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

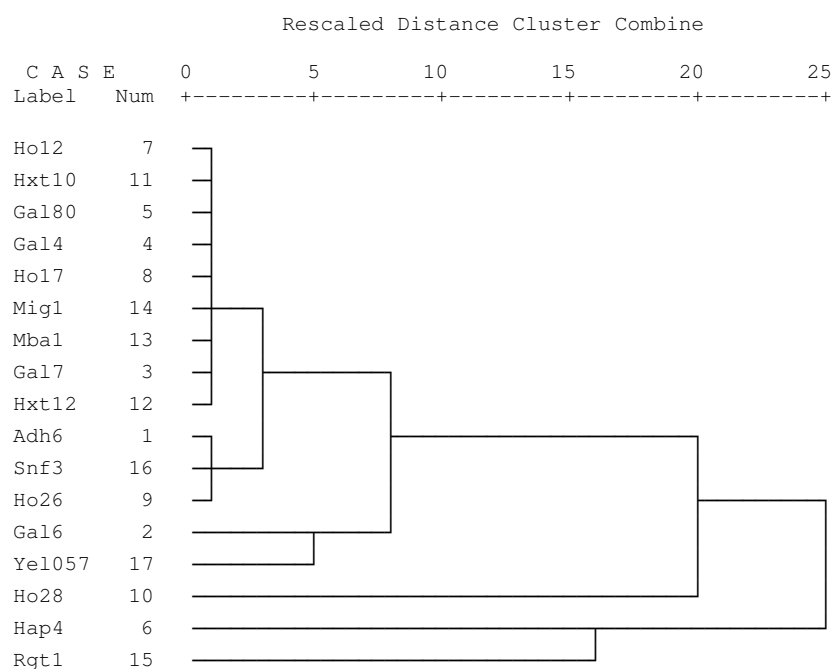


Figure A.40. HCA of IC2 and IC3 by 6 times pre-processing of PCA

A.11. PCA of Binned Intracellular Metabolome Data at 25th hour

Dendrogram using Average Linkage (Between Groups)

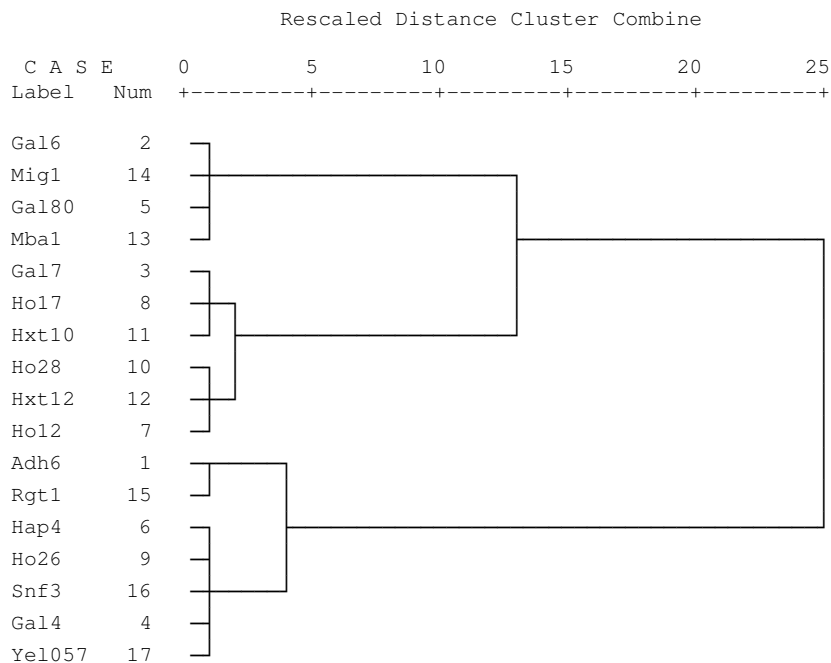


Figure A.41. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

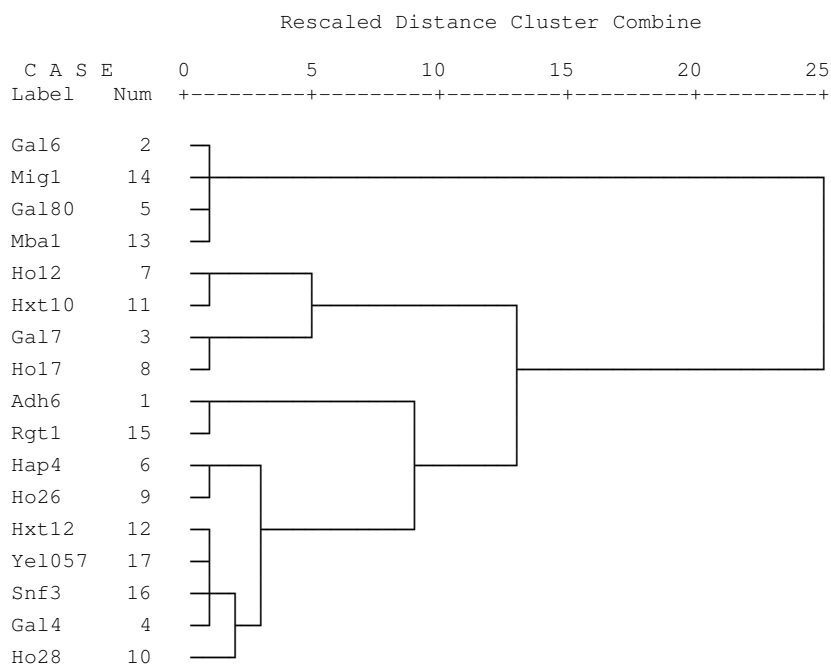


Figure A.42. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

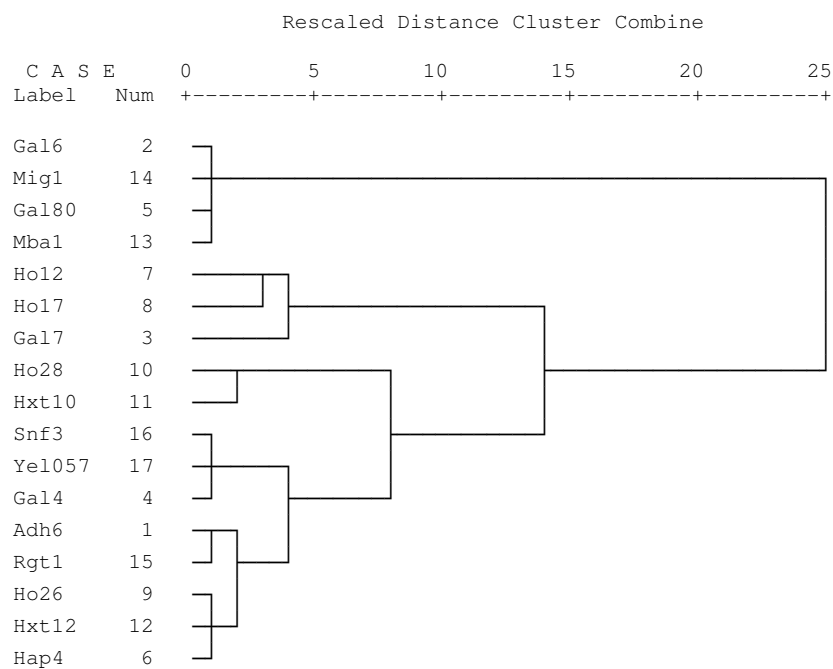


Figure A.43. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

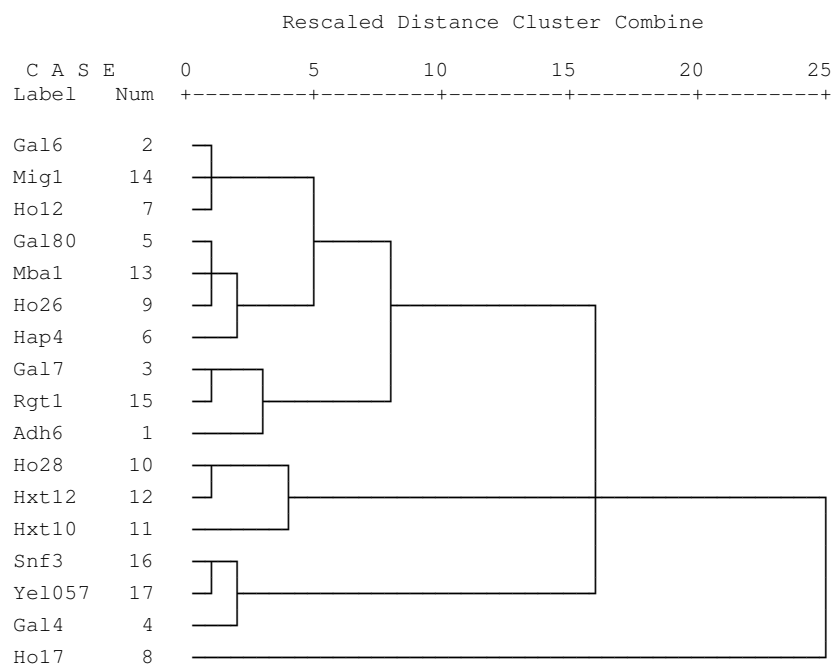


Figure A.44. HCA of PC3 and PC4

A.12. ICA of Binned Intracellular Metabolome Data at 25th hour

Dendrogram using Average Linkage (Between Groups)

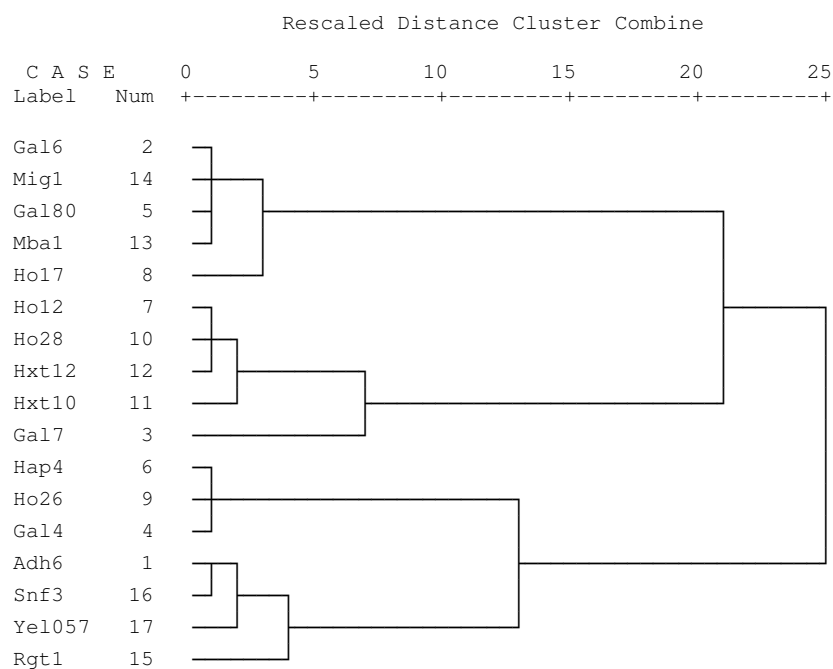


Figure A.45. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

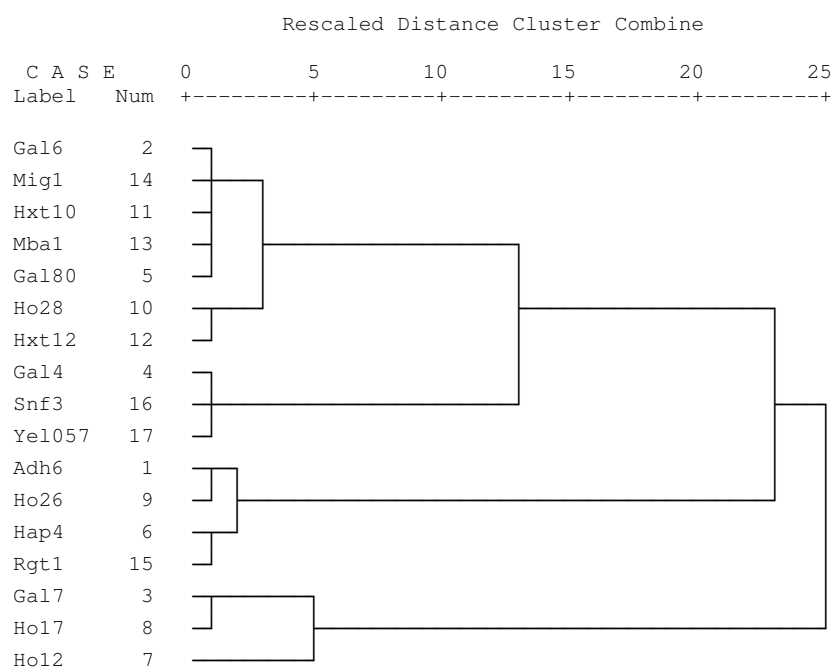


Figure A.46. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

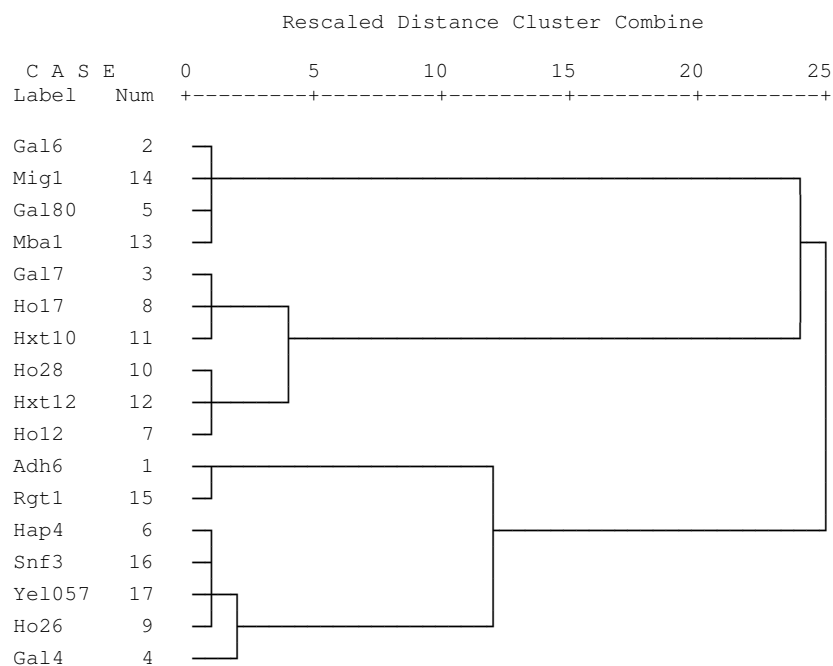


Figure A.47. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

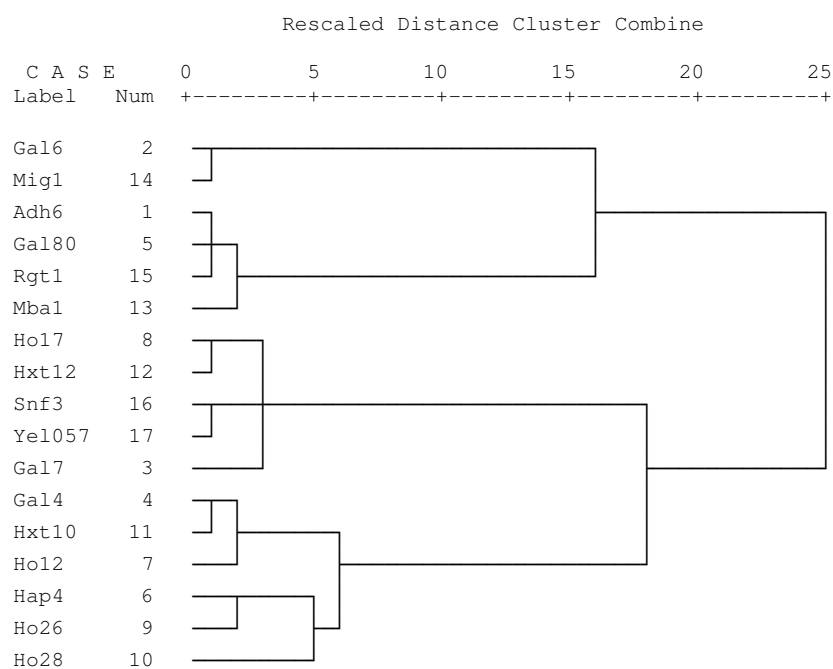


Figure A.48. HCA of IC2 and IC3 by 6 times pre-processing of PCA

A.14. PCA of Binned Intracellular Metabolome Data at 31st hour

Dendrogram using Average Linkage (Between Groups)

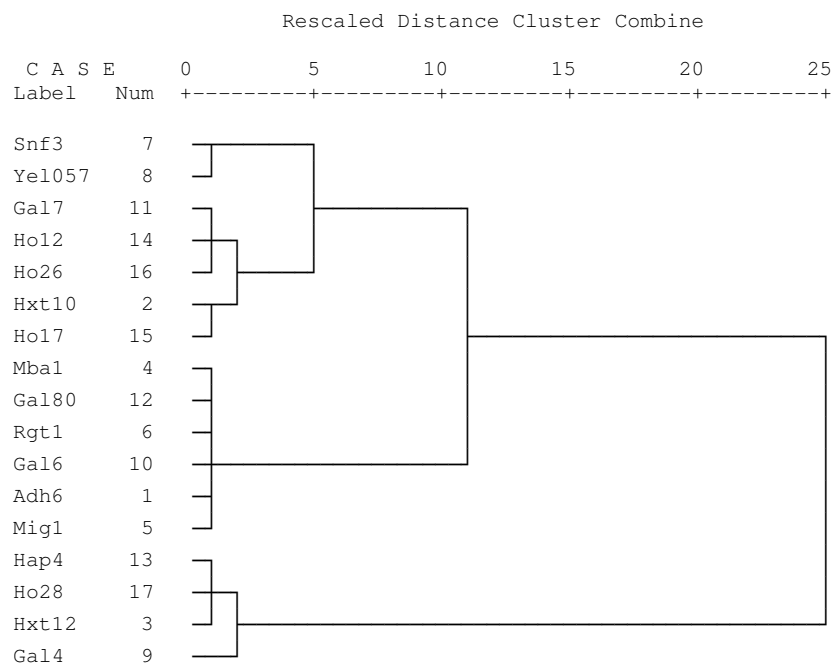


Figure A.49. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

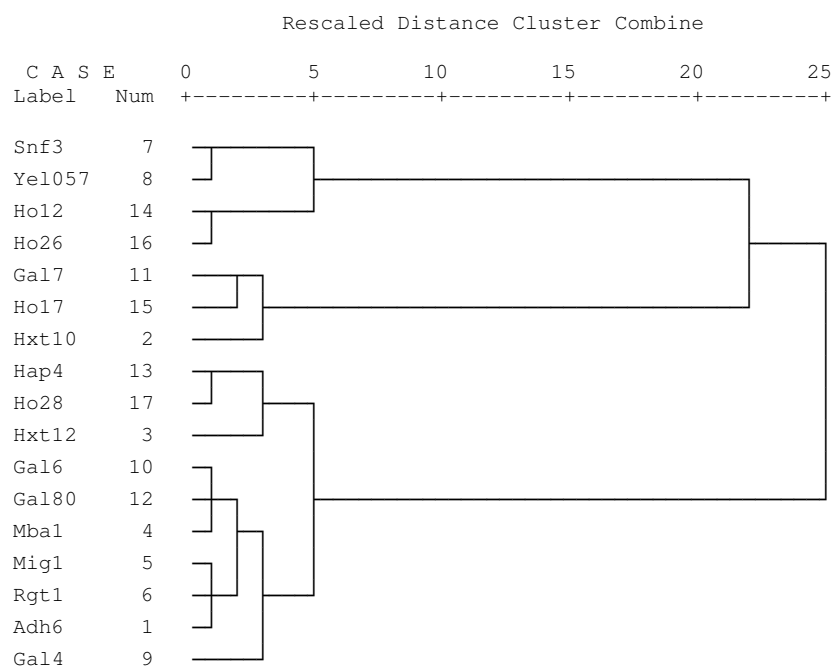


Figure A.50. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

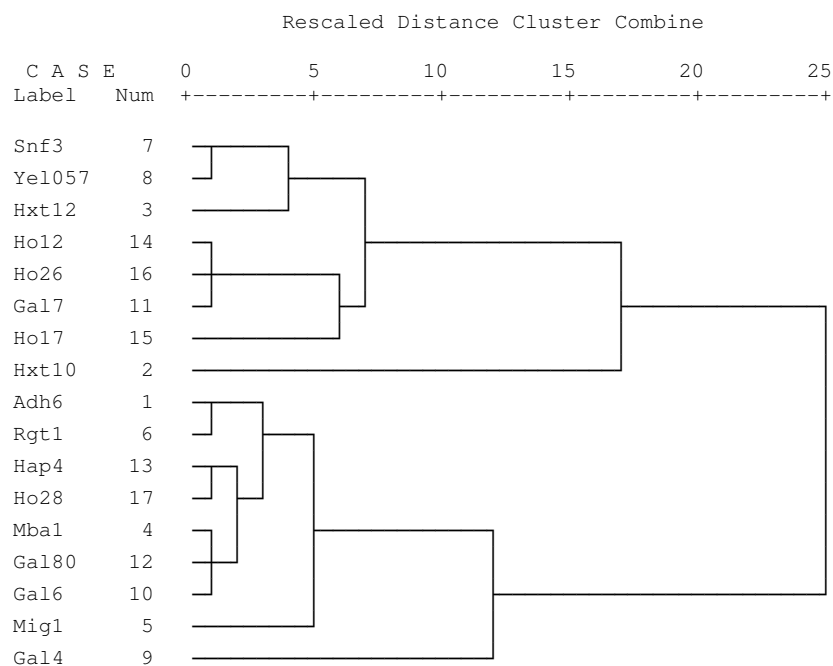


Figure A.51. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

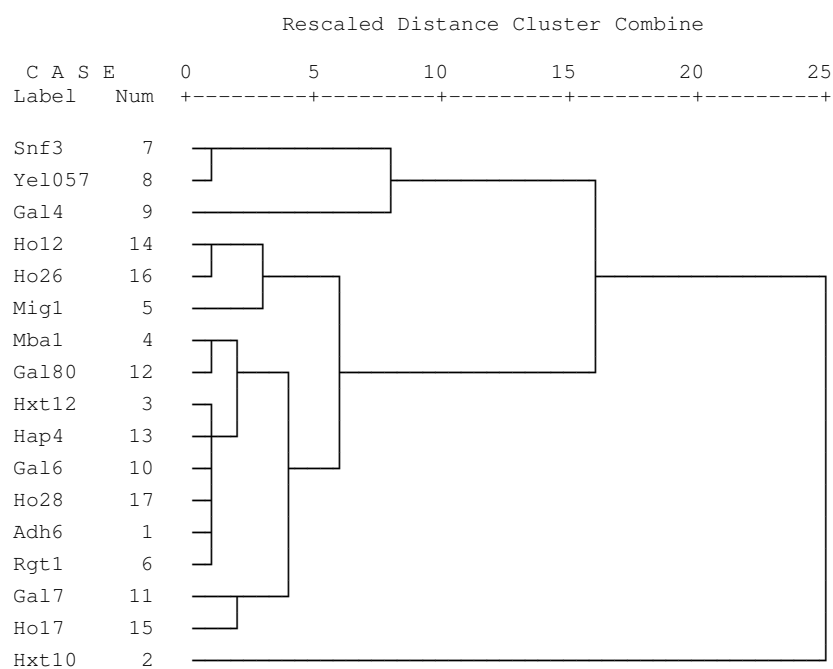


Figure A.52. HCA of PC3 and PC4

A.15. ICA of Binned Intracellular Metabolome Data at 31st hour

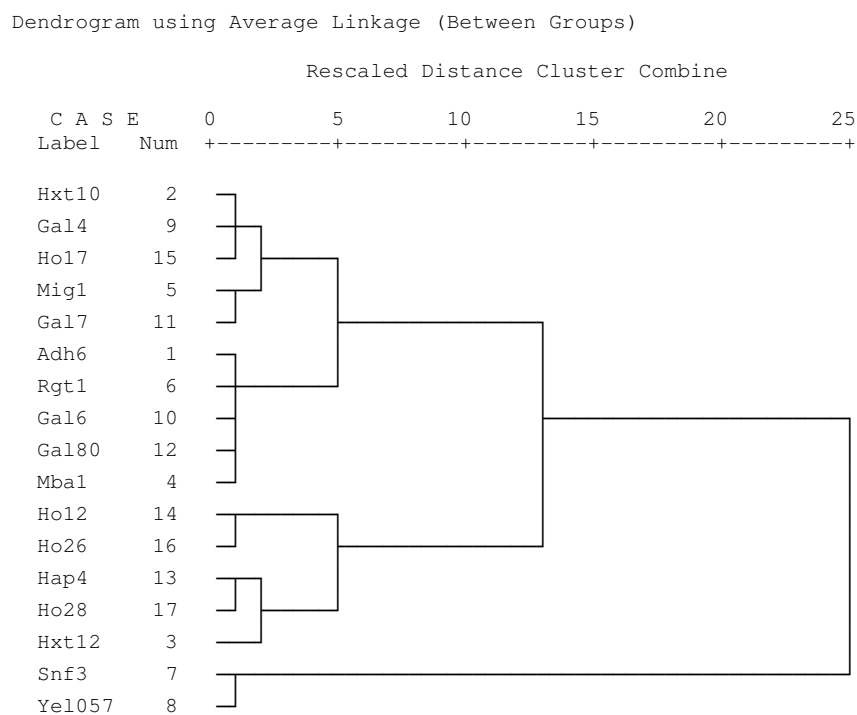


Figure A.53. HCA of IC1 and IC2 by 5 times pre-processing of PCA

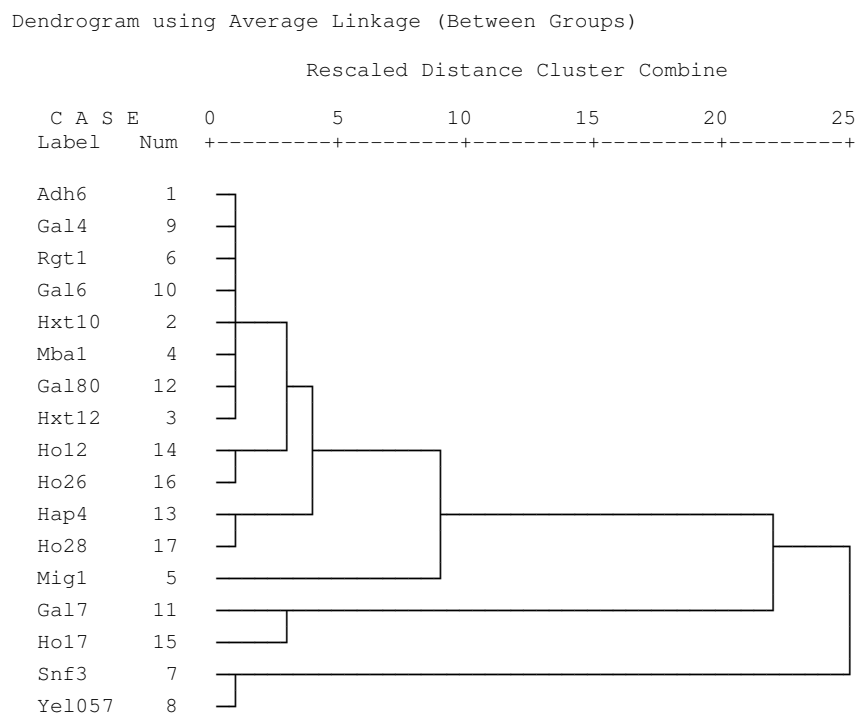


Figure A.54. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

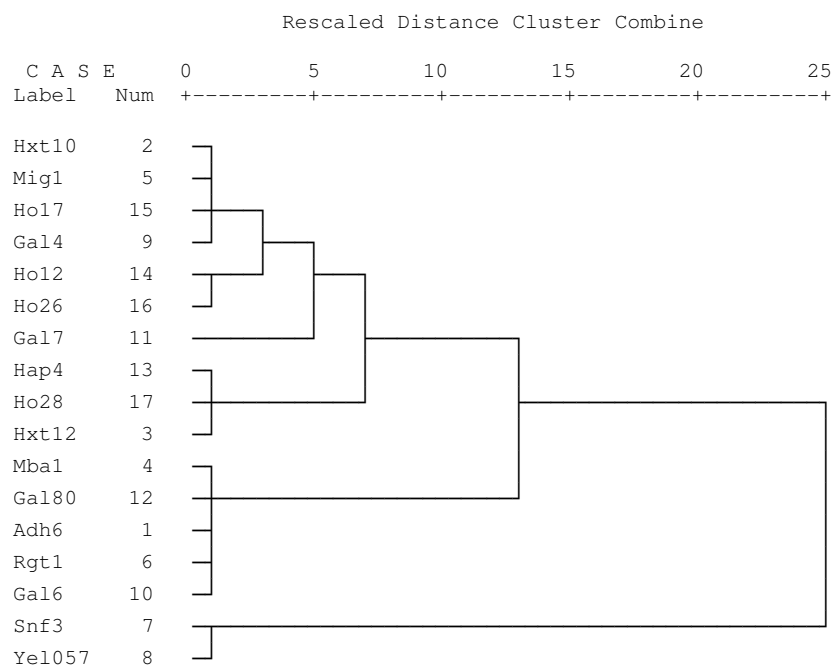


Figure A.55. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

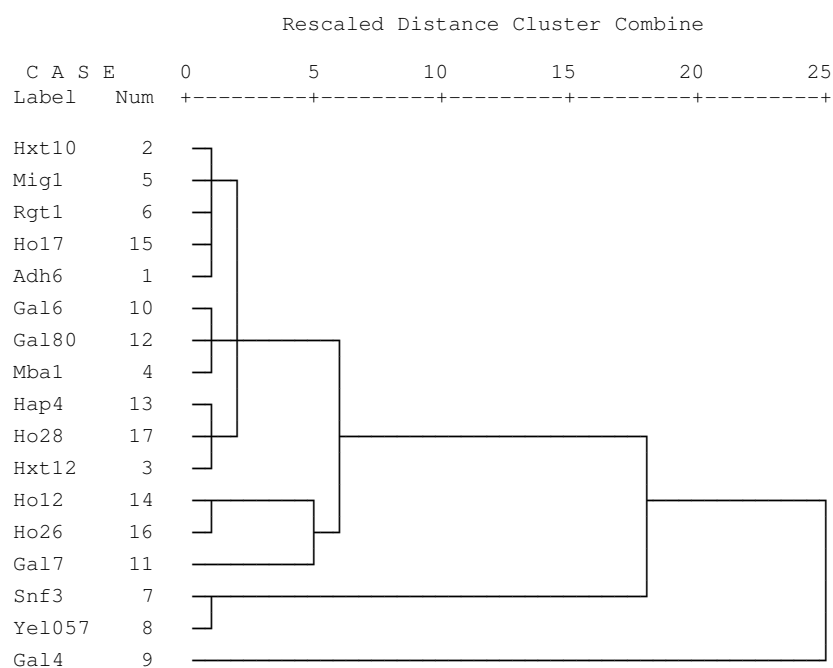


Figure A.56. HCA of IC2 and IC3 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

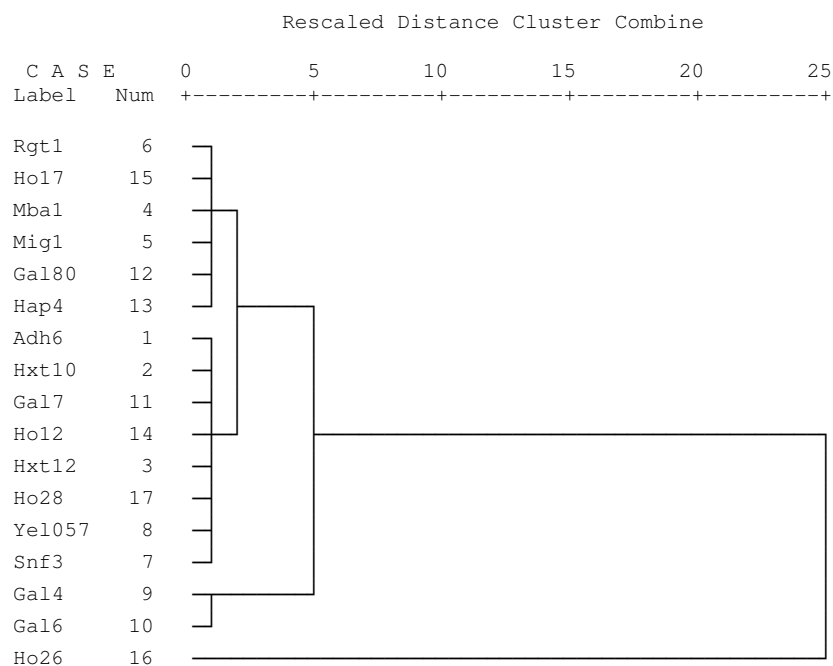


Figure B.3. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

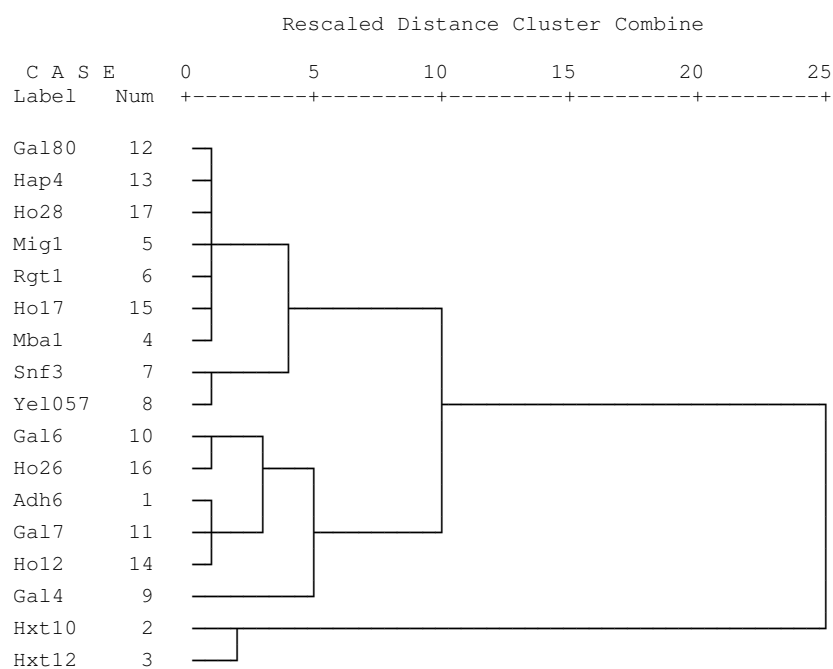


Figure B.4. HCA of PC3 and PC4

B.2. ICA of Filtered Extracellular Metabolome Data at 15th hour

Dendrogram using Average Linkage (Between Groups)

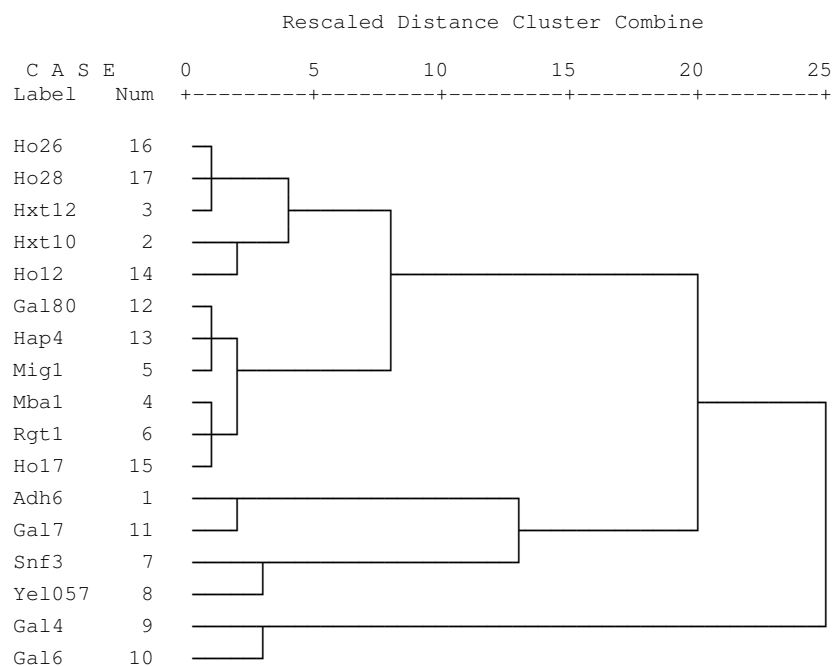


Figure B.5. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

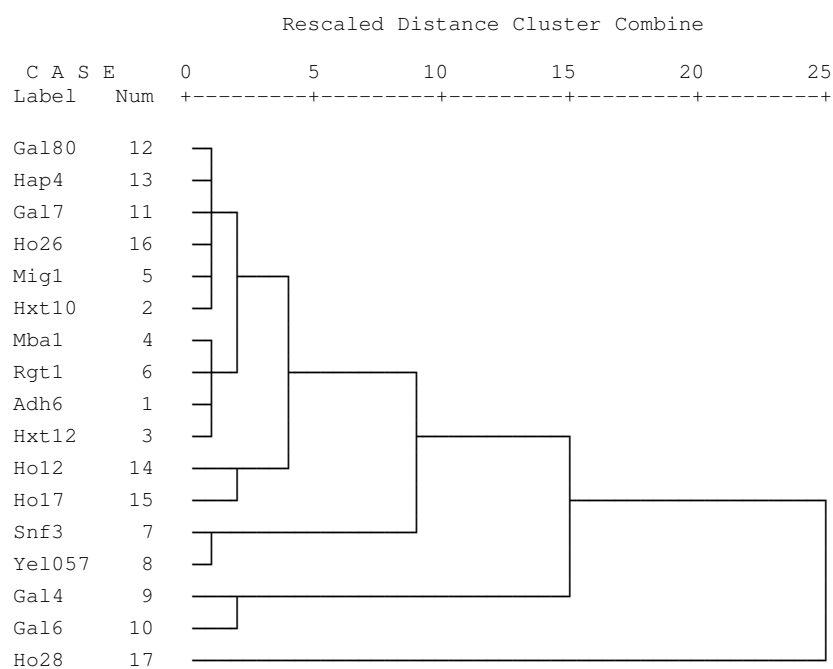


Figure B.6. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

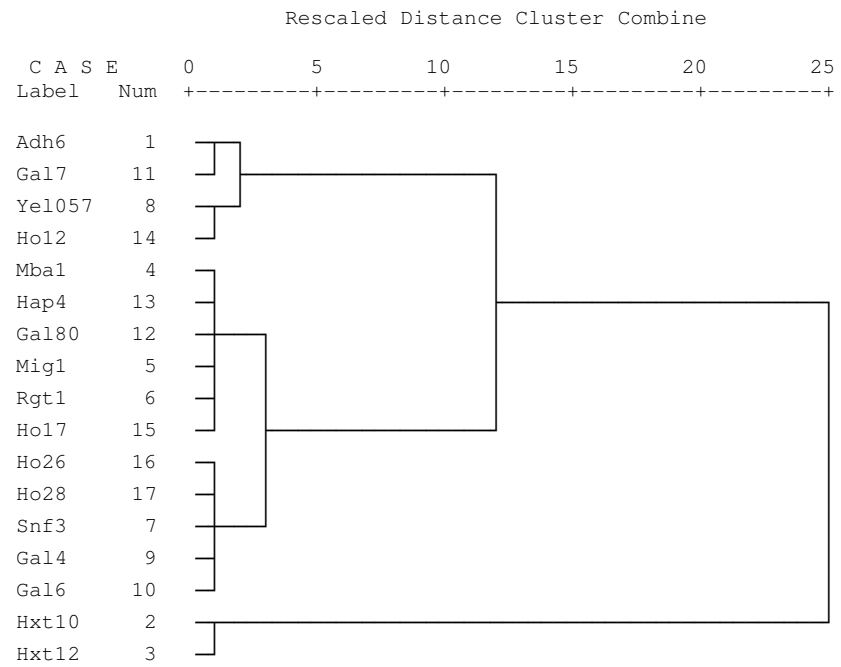


Figure B.7. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

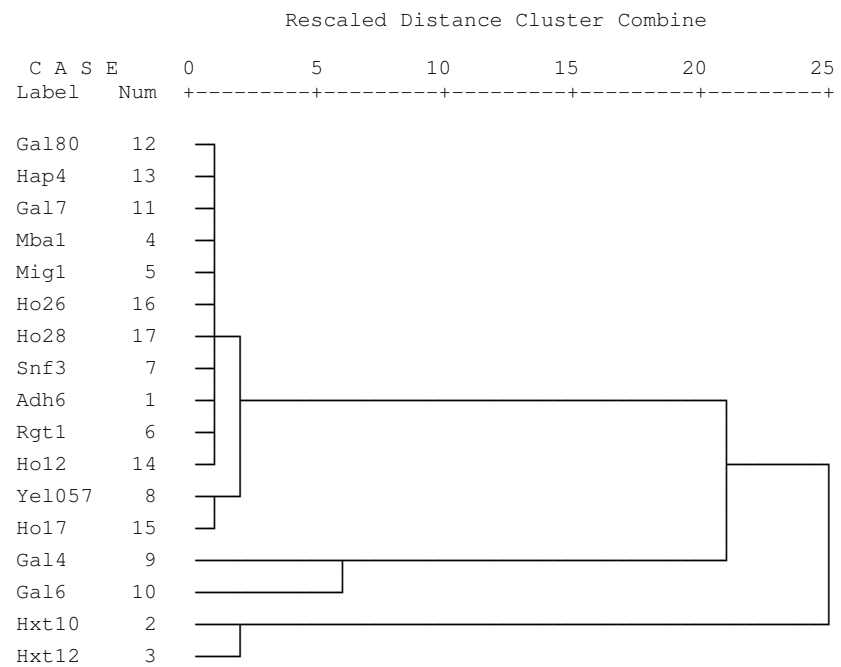


Figure B.8. HCA of IC2 and IC3 by 6 times pre-processing of PCA

B.3. PCA of Filtered Extracellular Metabolome Data at 20th hour

Dendrogram using Average Linkage (Between Groups)

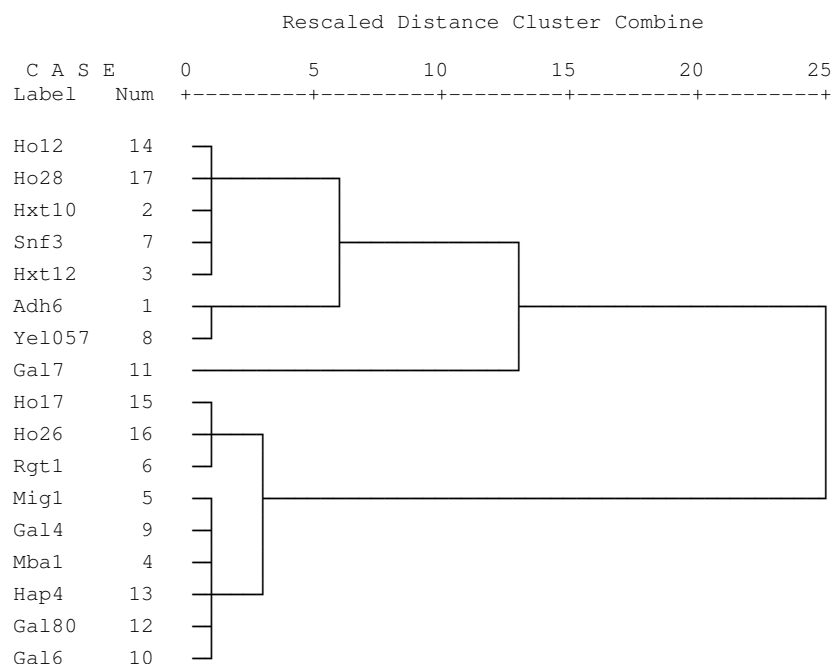


Figure B.9. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

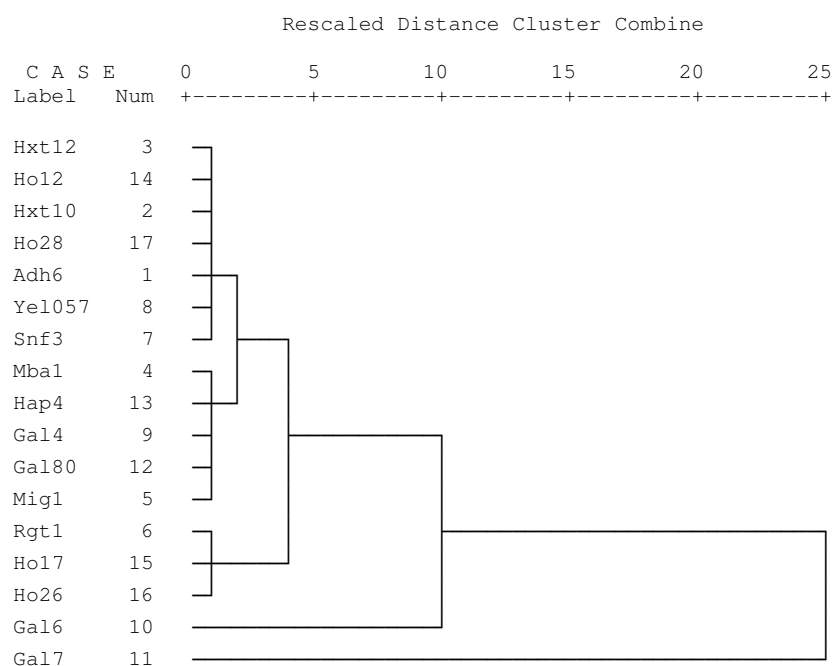


Figure B.10. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

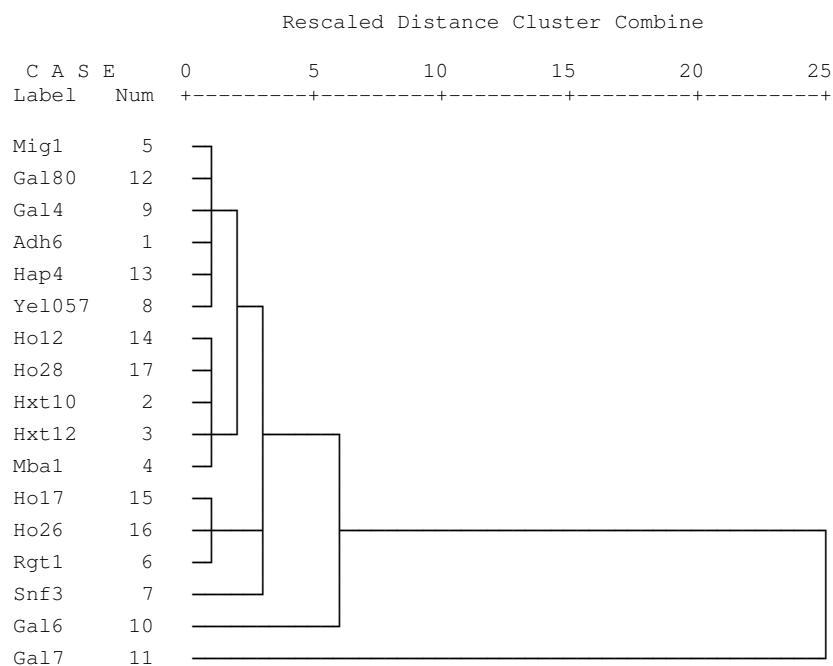


Figure B.11. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

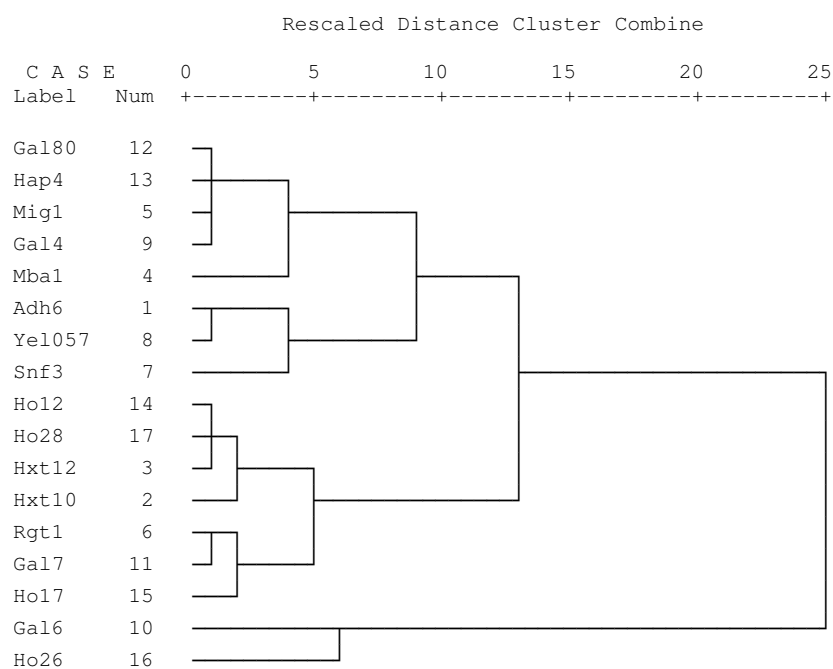


Figure B.12. HCA of PC3 and PC4

B.4. ICA of Filtered Extracellular Metabolome Data at 20th hour

Dendrogram using Average Linkage (Between Groups)

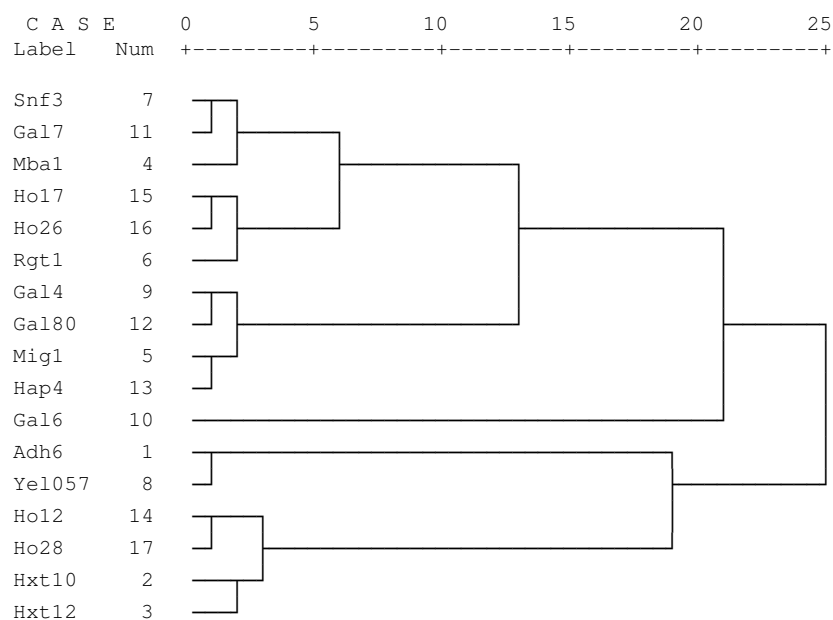


Figure B.13. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

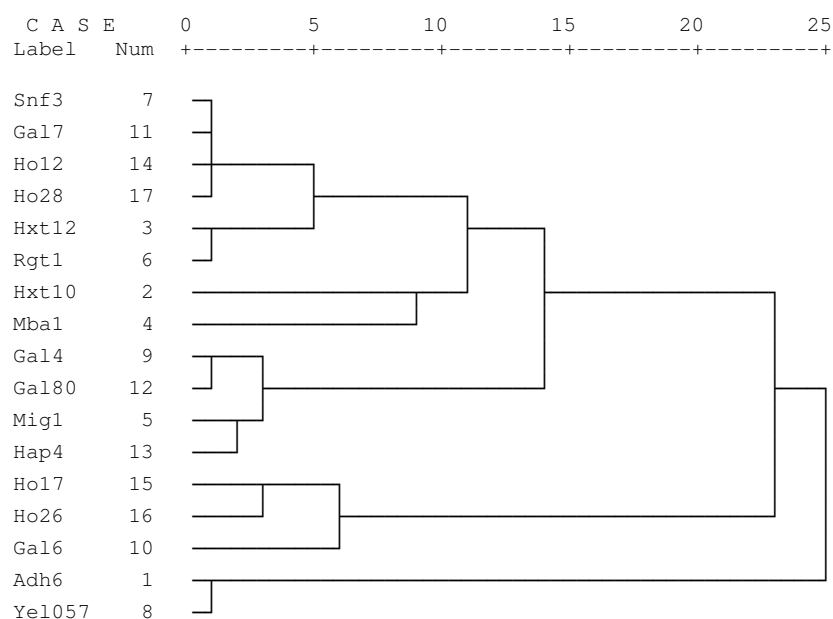


Figure B.14. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

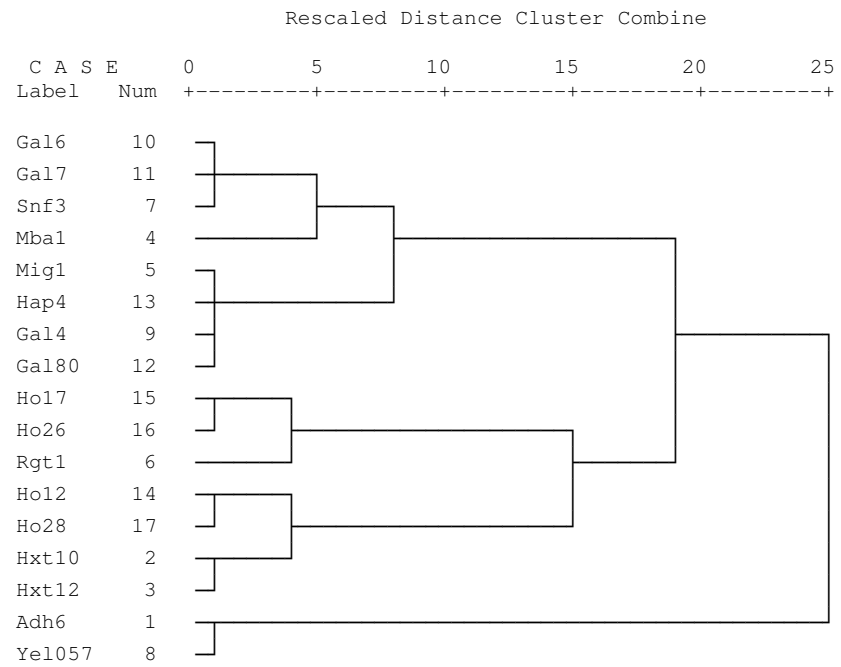


Figure B.15. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

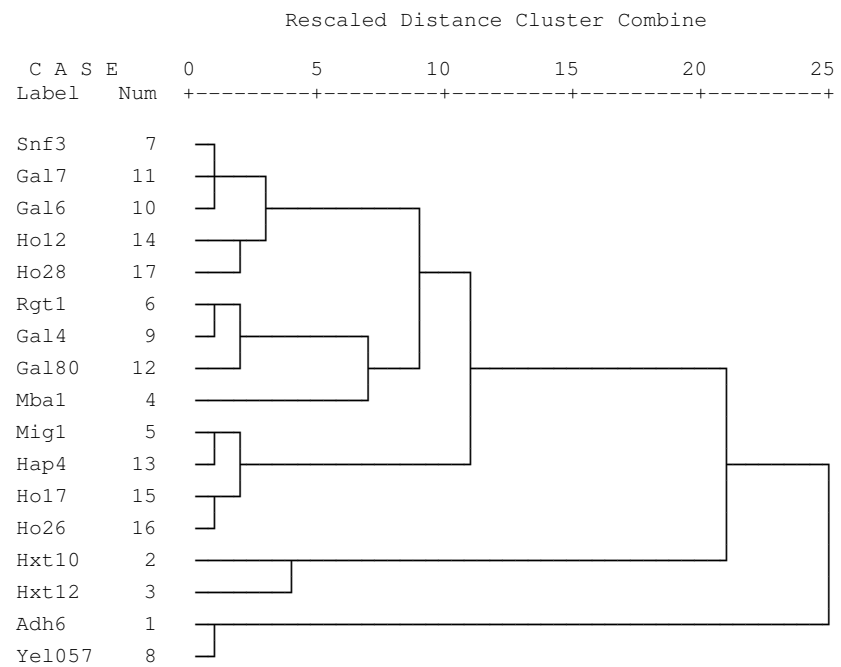


Figure B.16. HCA of IC2 and IC3 by 6 times pre-processing of PCA

B.5. PCA of Filtered Extracellular Metabolome Data at 25th hour

Dendrogram using Average Linkage (Between Groups)

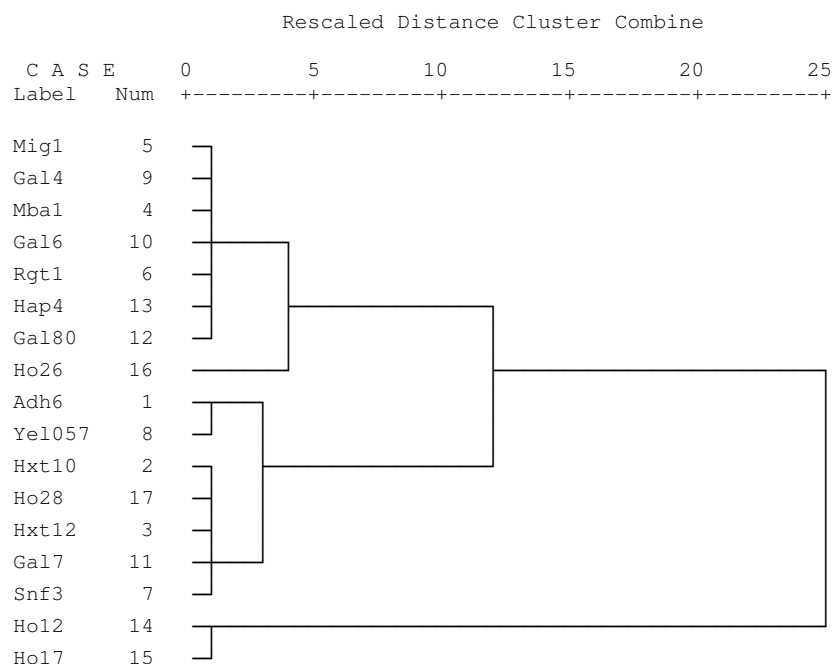


Figure B.17. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

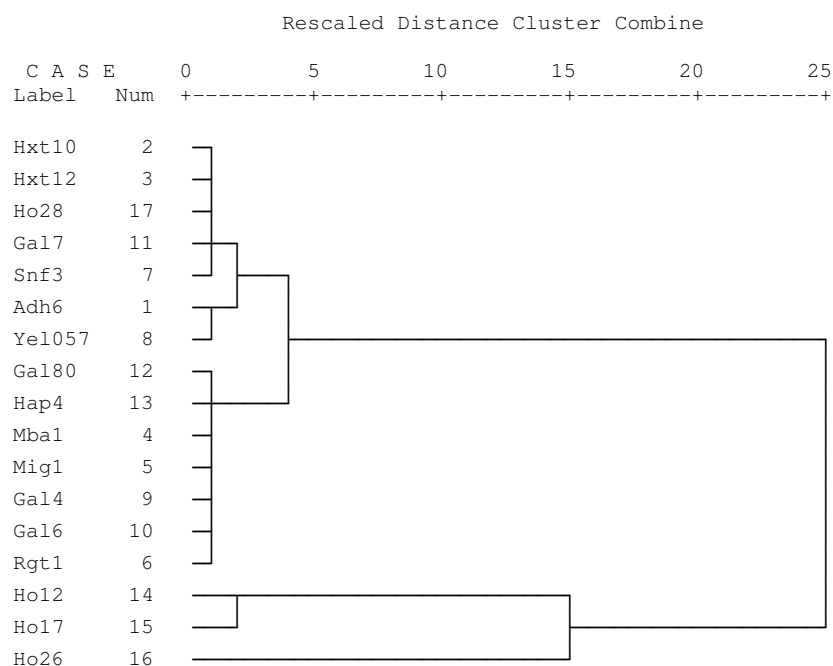


Figure B.18. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

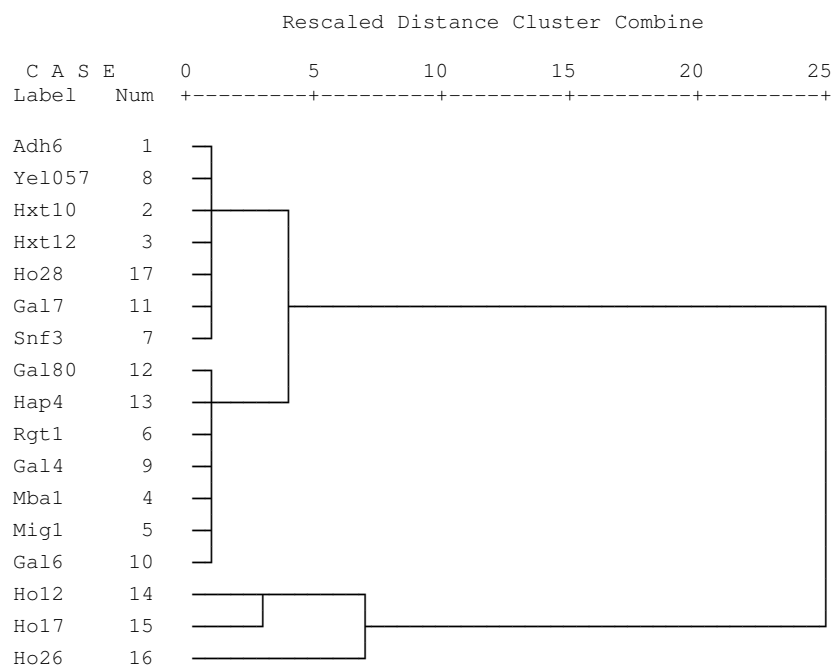


Figure B.19. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

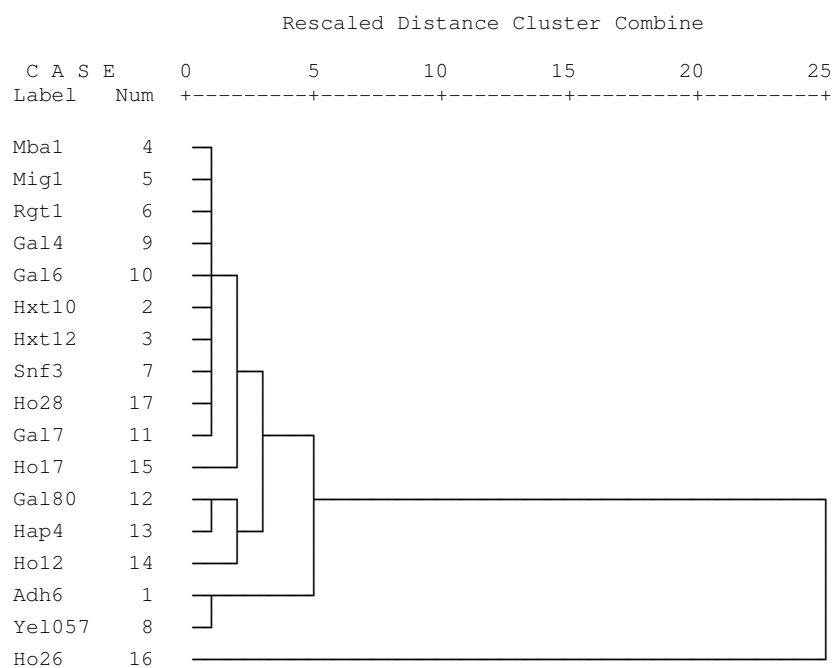


Figure B.20. HCA of PC3 and PC4

B.6. ICA of Filtered Extracellular Metabolome Data at 25th hour

Dendrogram using Average Linkage (Between Groups)

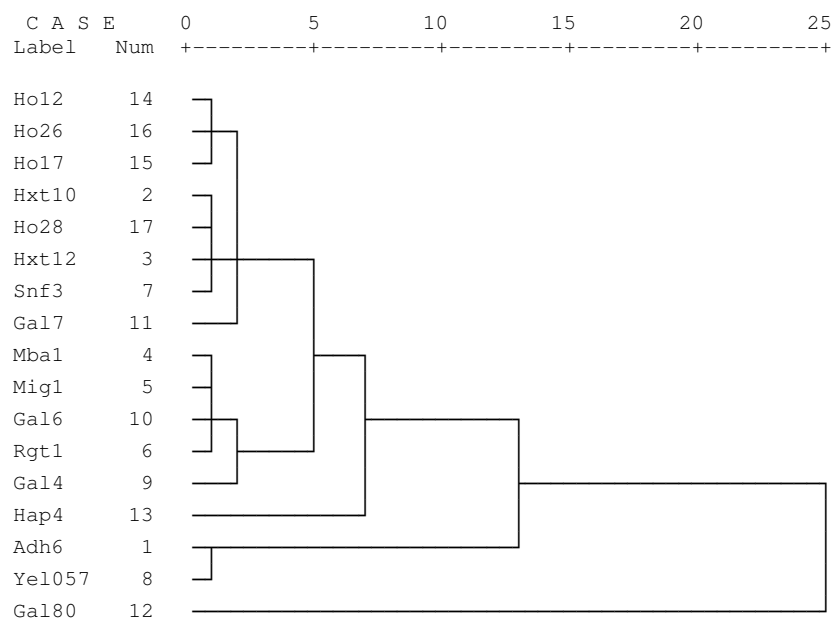


Figure B.21. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

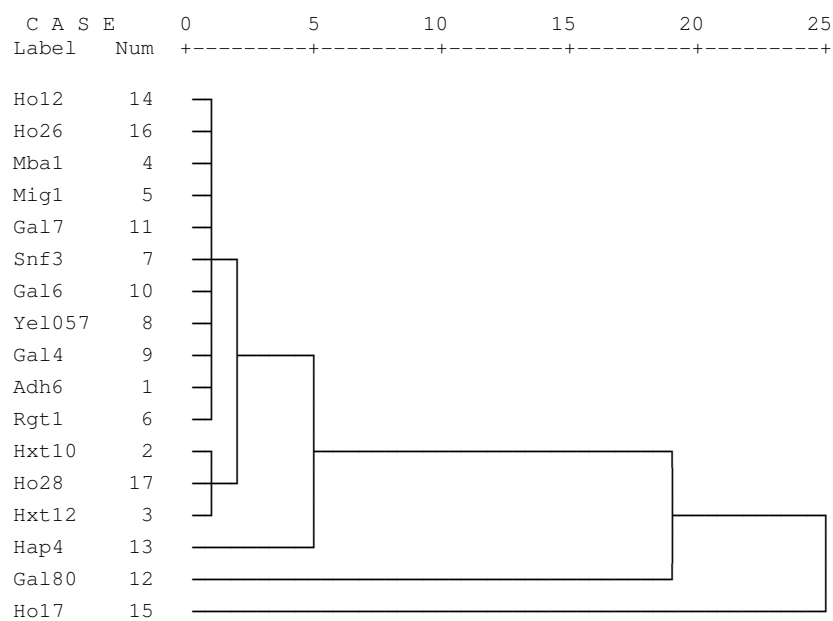


Figure B.22. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

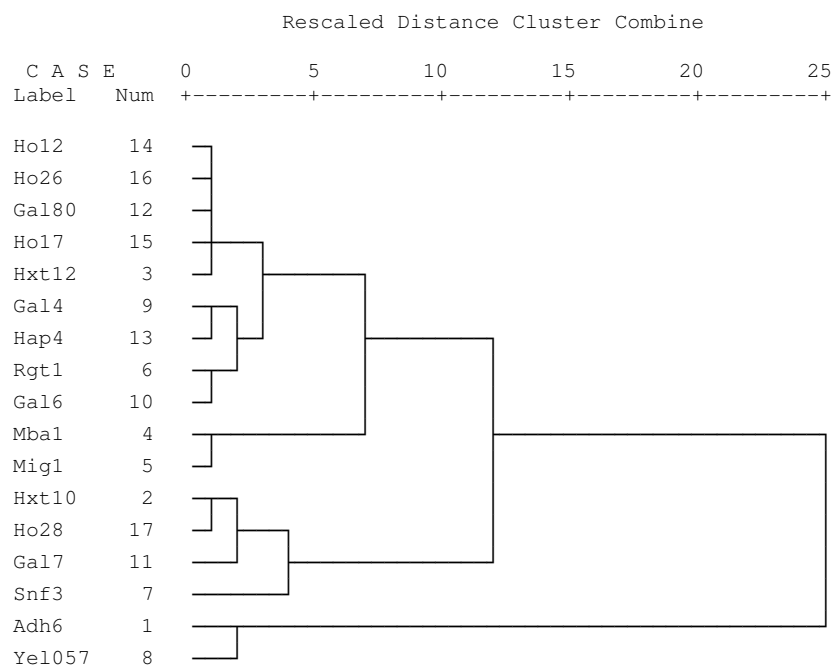


Figure B.23. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

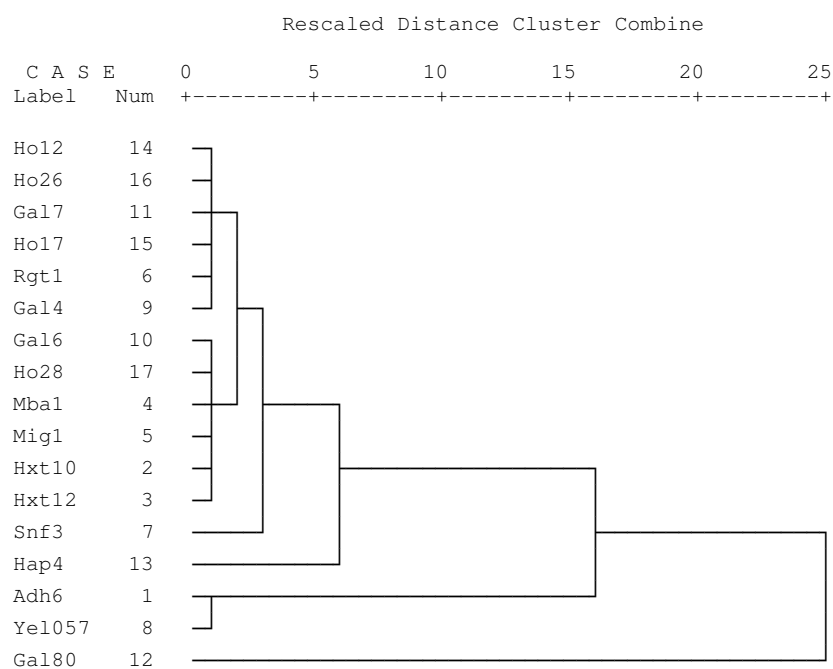


Figure B.24. HCA of IC2 and IC3 by 6 times pre-processing of PCA

B.7. PCA of Filtered Extracellular Metabolome Data at 31st hour

Dendrogram using Average Linkage (Between Groups)

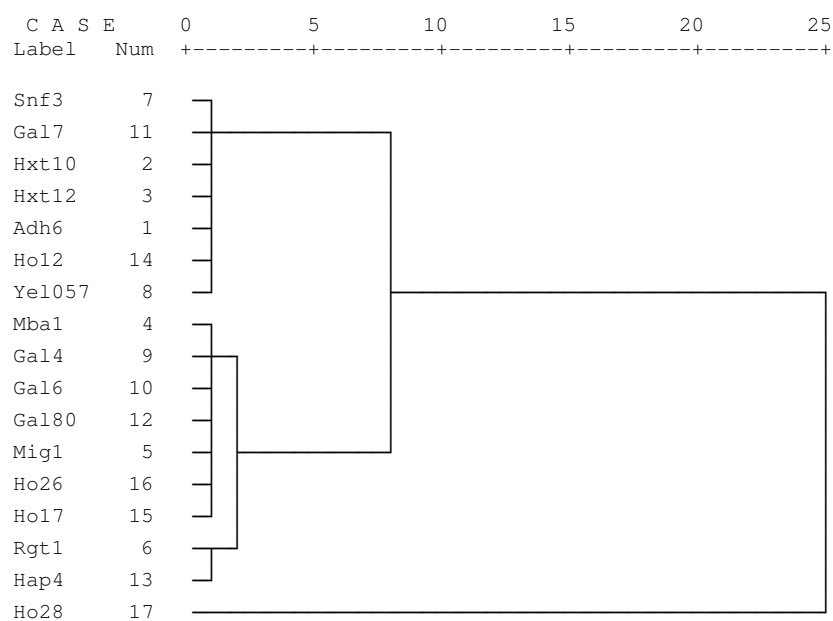


Figure B.25. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

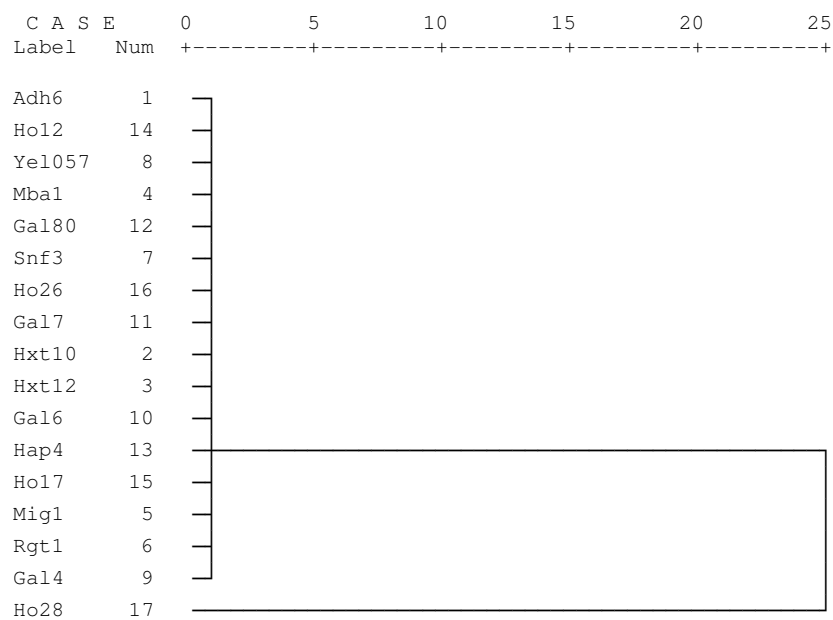


Figure B.26. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

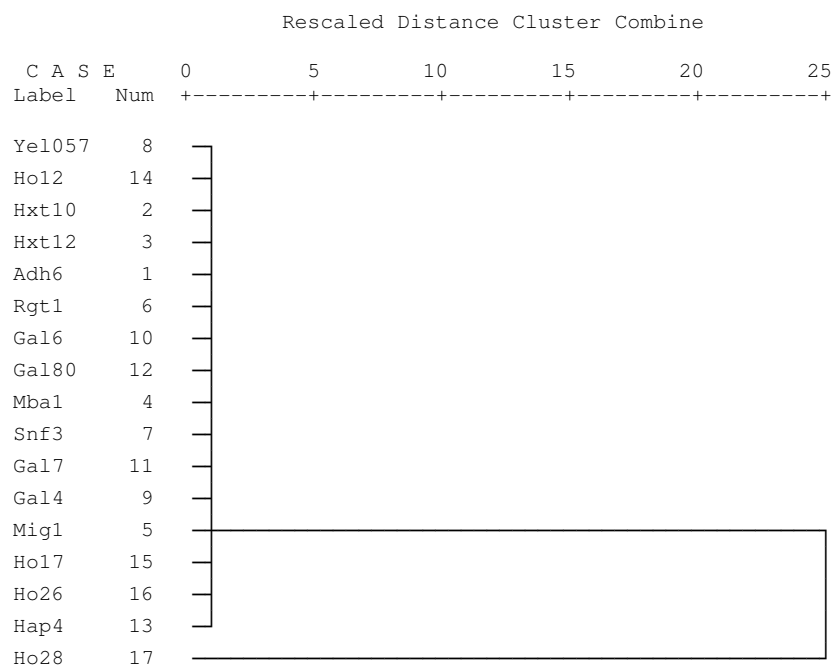


Figure B.27. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

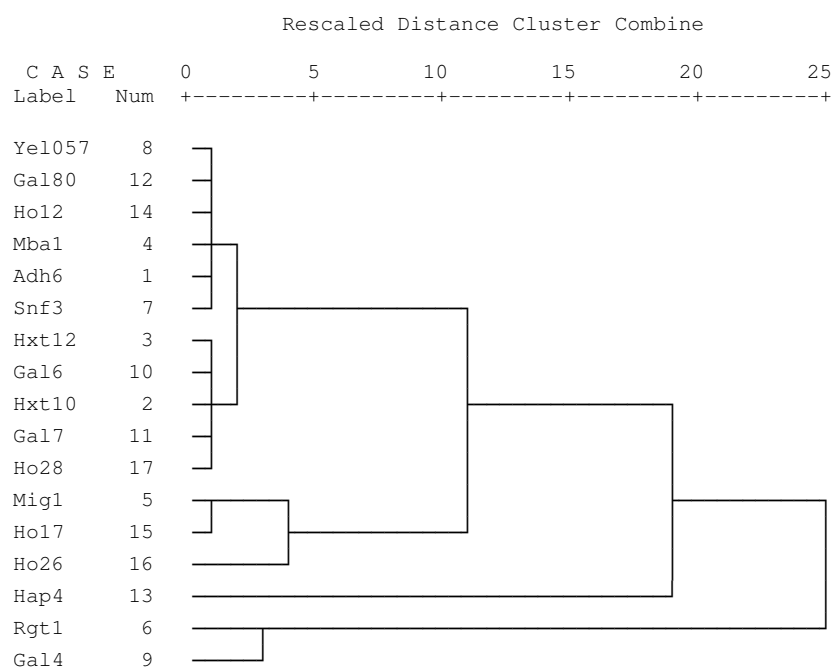


Figure B.28. HCA of PC3 and PC4

B.8. ICA of Filtered Extracellular Metabolome Data at 31st hour

Dendrogram using Average Linkage (Between Groups)

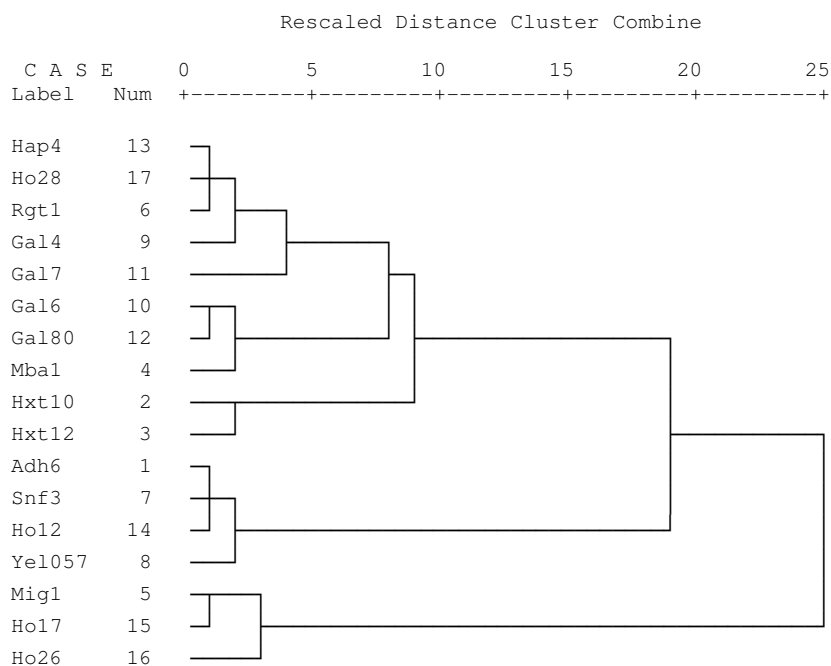


Figure B.29. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

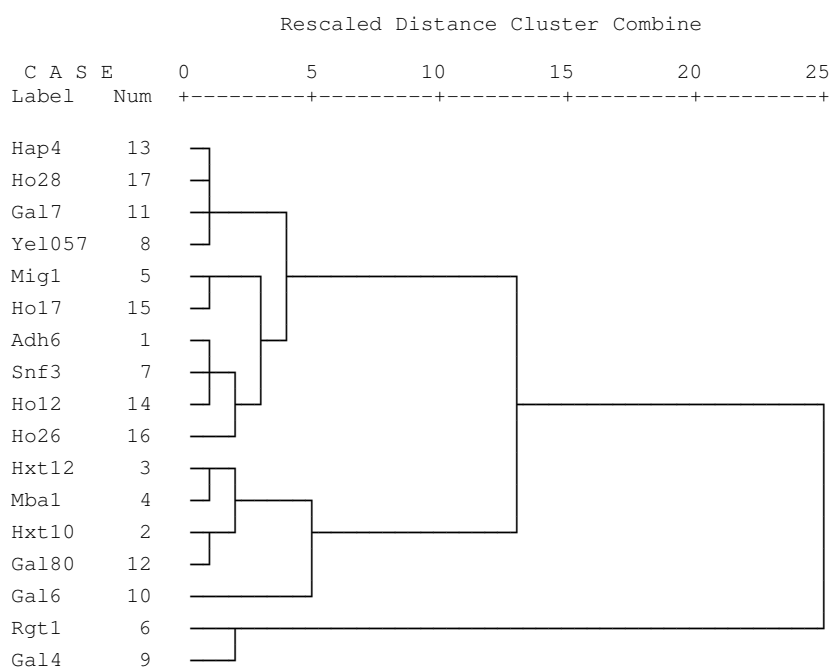


Figure B.30. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

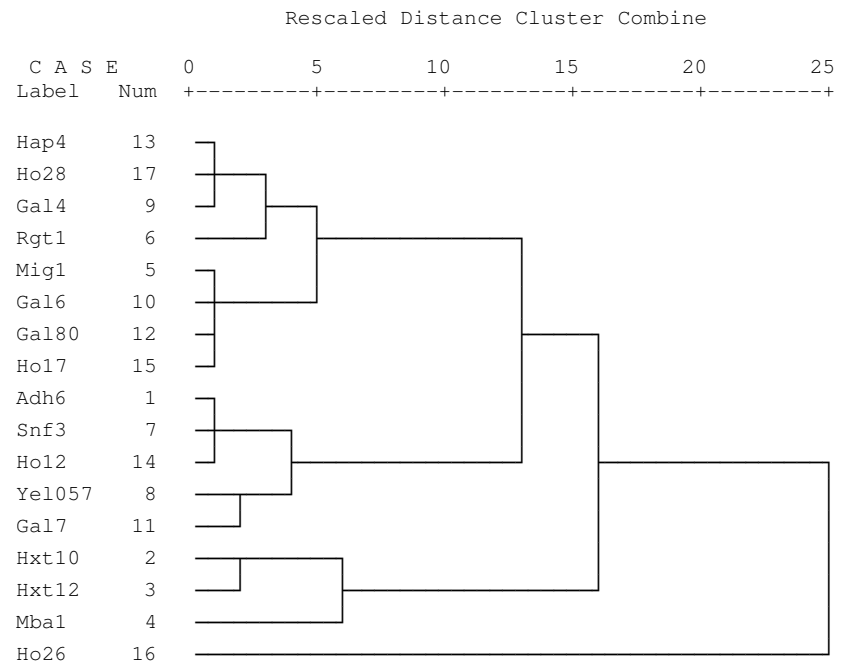


Figure B.31. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

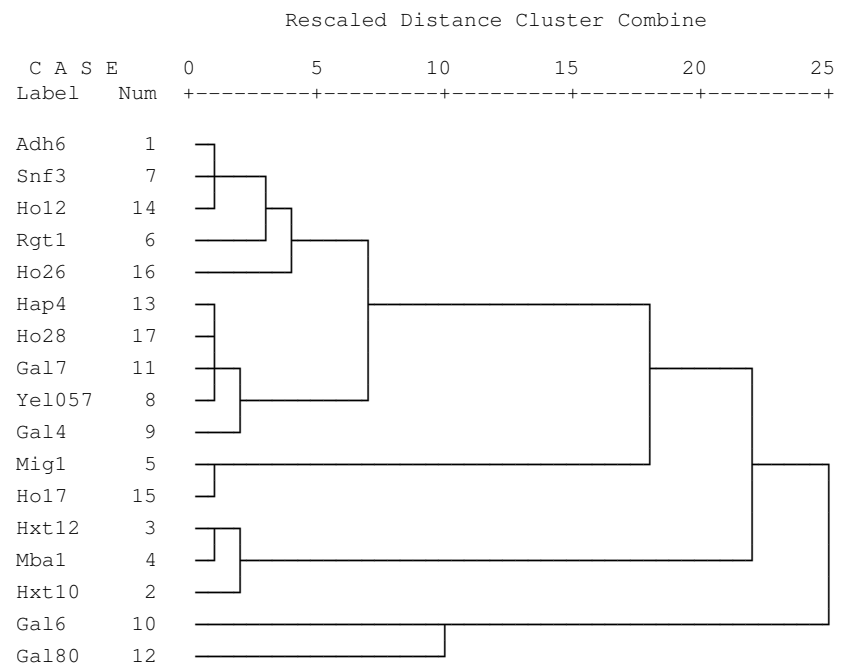


Figure B.32. HCA of IC2 and IC3 by 6 times pre-processing of PCA

B.9. PCA of Filtered Intracellular Metabolome Data at 20th hour

Dendrogram using Average Linkage (Between Groups)

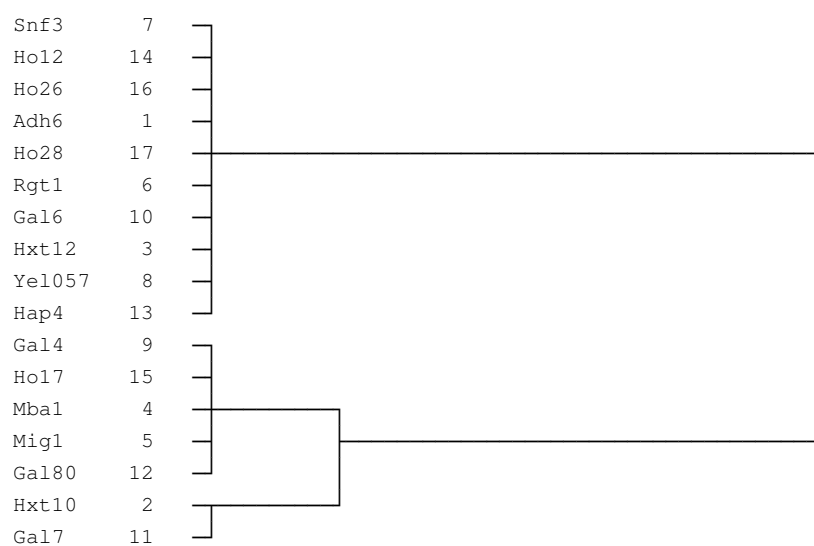


Figure B.33. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

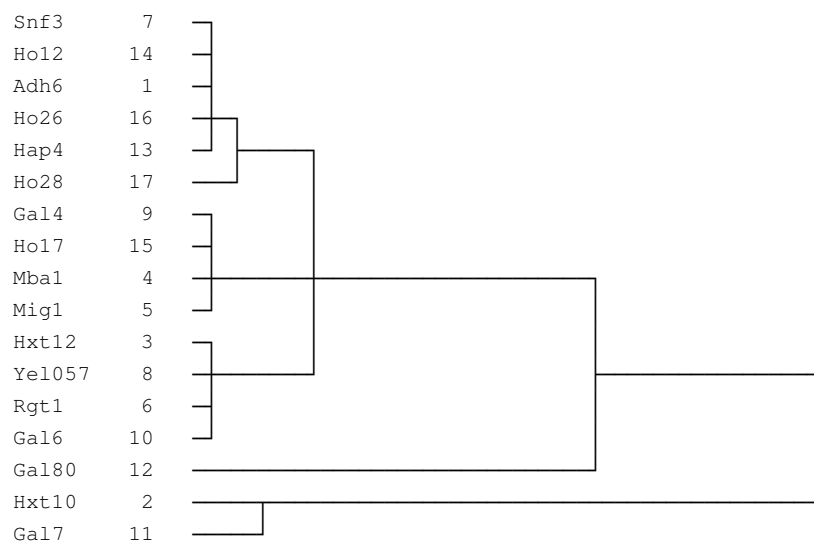


Figure B.34. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

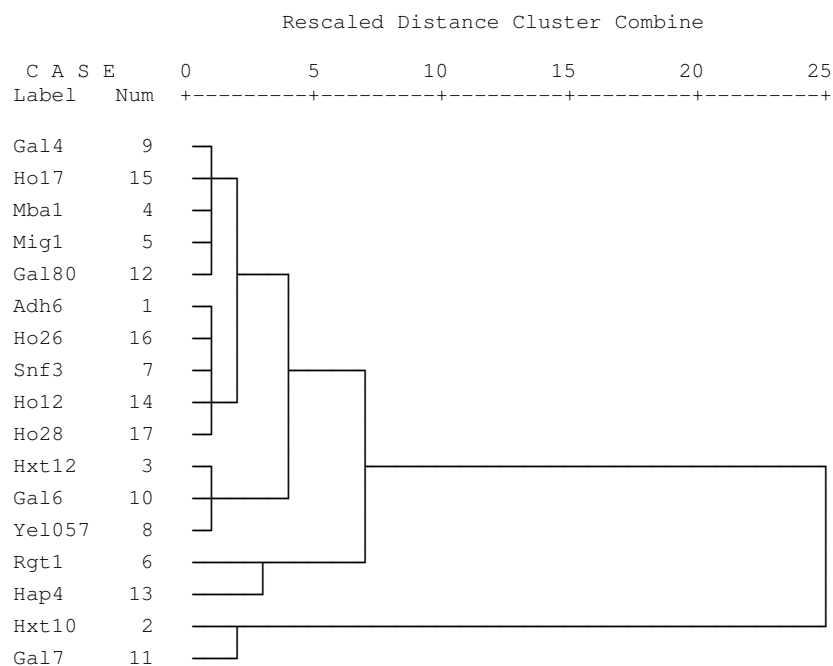


Figure B.35. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

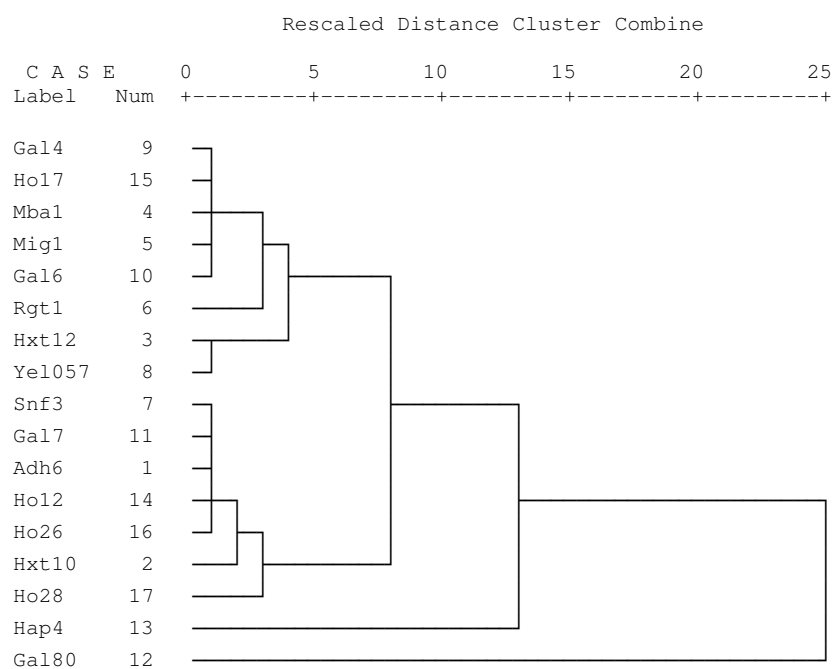


Figure B.36. HCA of PC3 and PC4

B.10. ICA of Filtered Intracellular Metabolome Data at 20th hour

Dendrogram using Average Linkage (Between Groups)

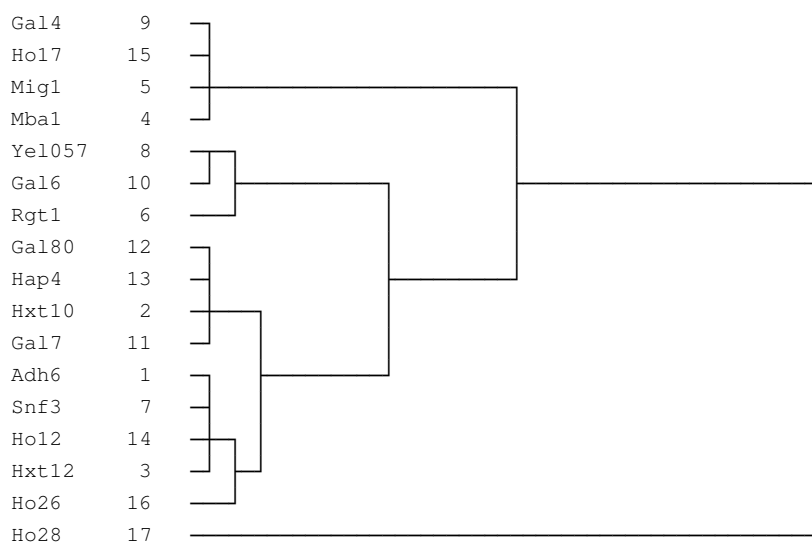


Figure B.37. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

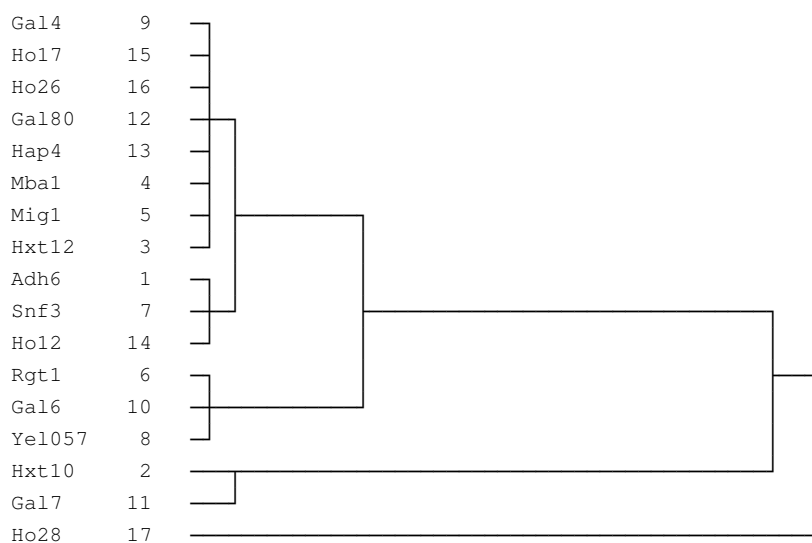


Figure B.38. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

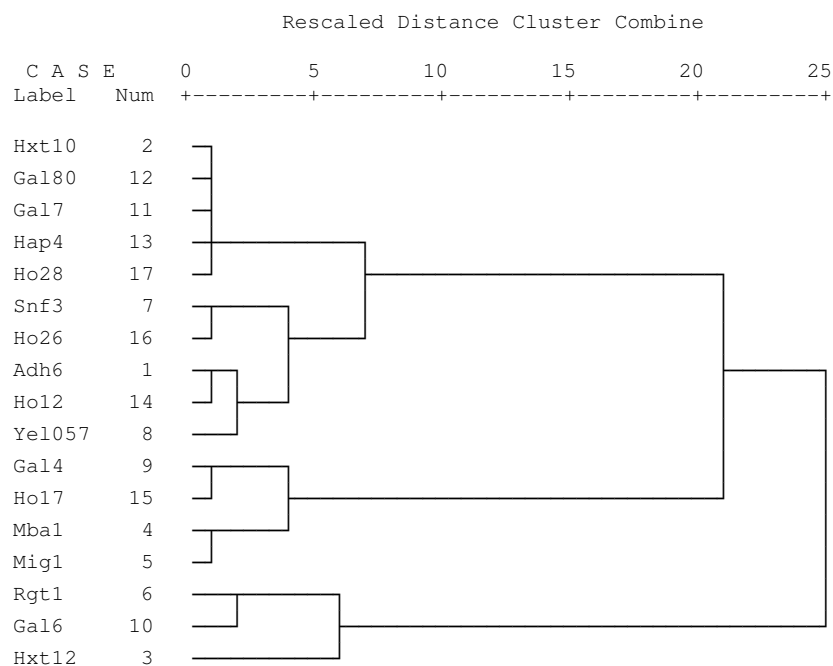


Figure B.39. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

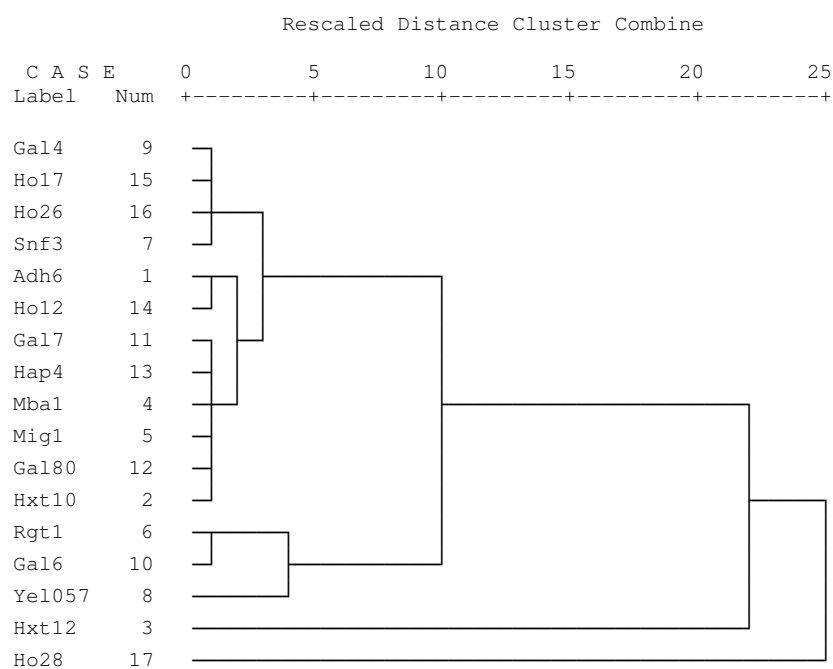


Figure B.40. HCA of IC2 and IC3 by 6 times pre-processing of PCA

B.11. PCA of Filtered Intracellular Metabolome Data at 25th hour

Dendrogram using Average Linkage (Between Groups)

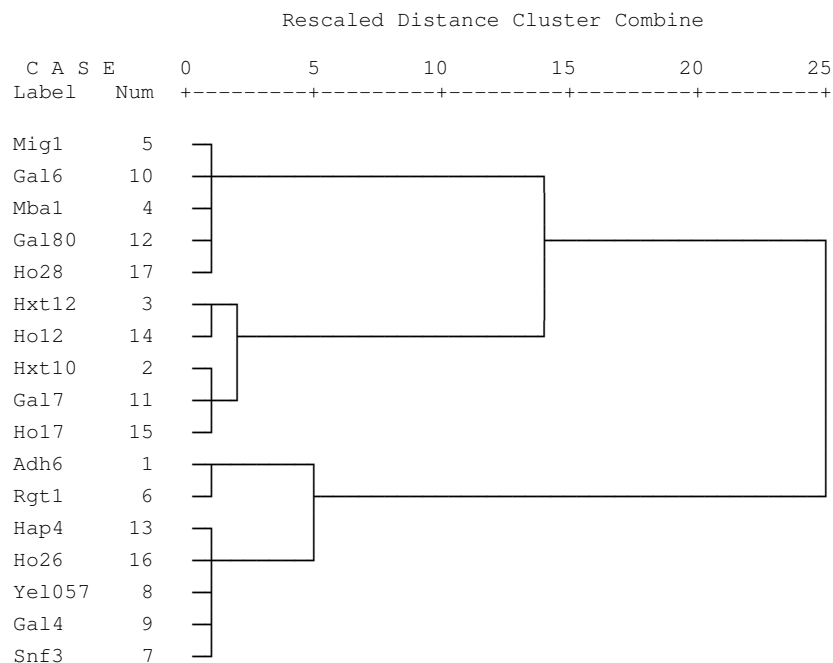


Figure B.41. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

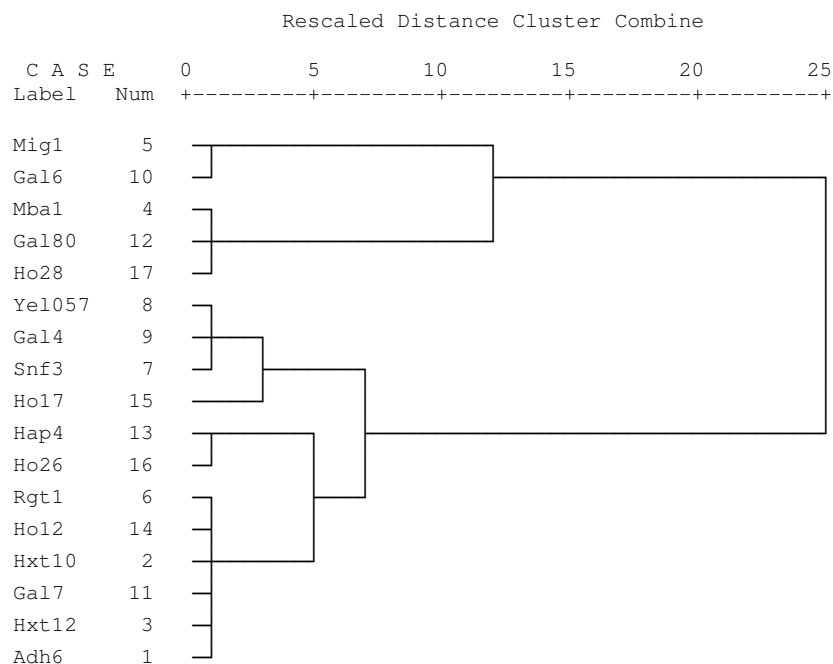


Figure B.42. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

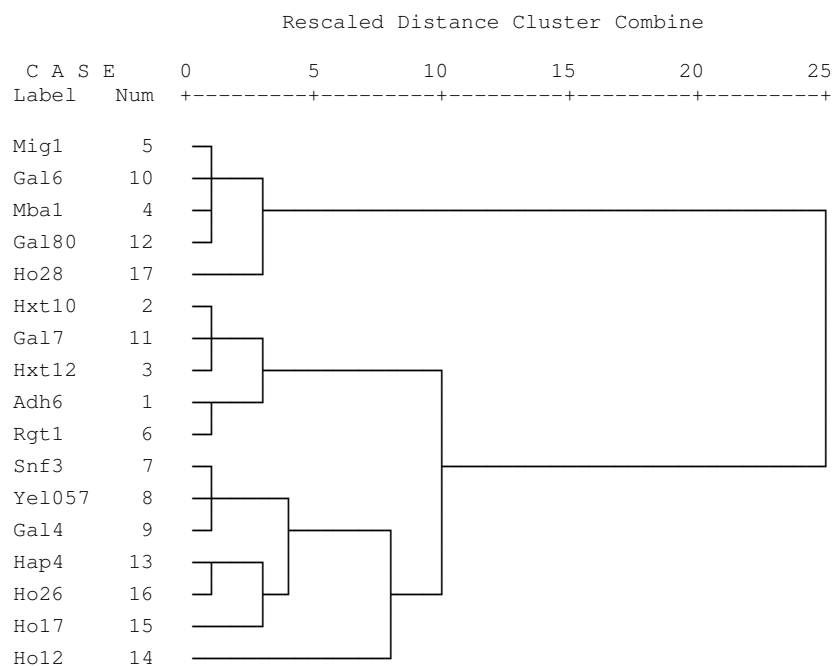


Figure B.43 HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

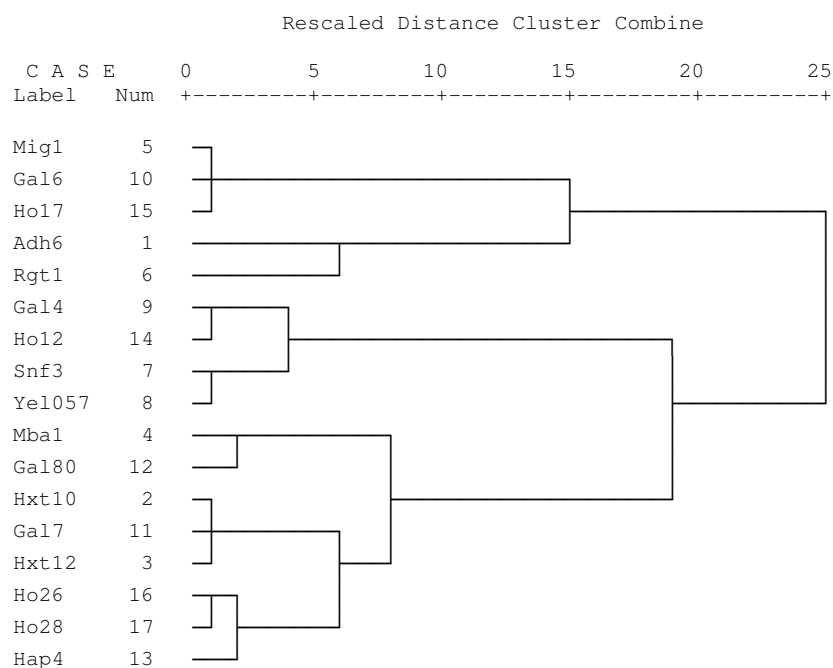


Figure B.44. HCA of PC3 and PC4

B.12. ICA of Filtered Intracellular Metabolome Data at 25th hour

Dendrogram using Average Linkage (Between Groups)

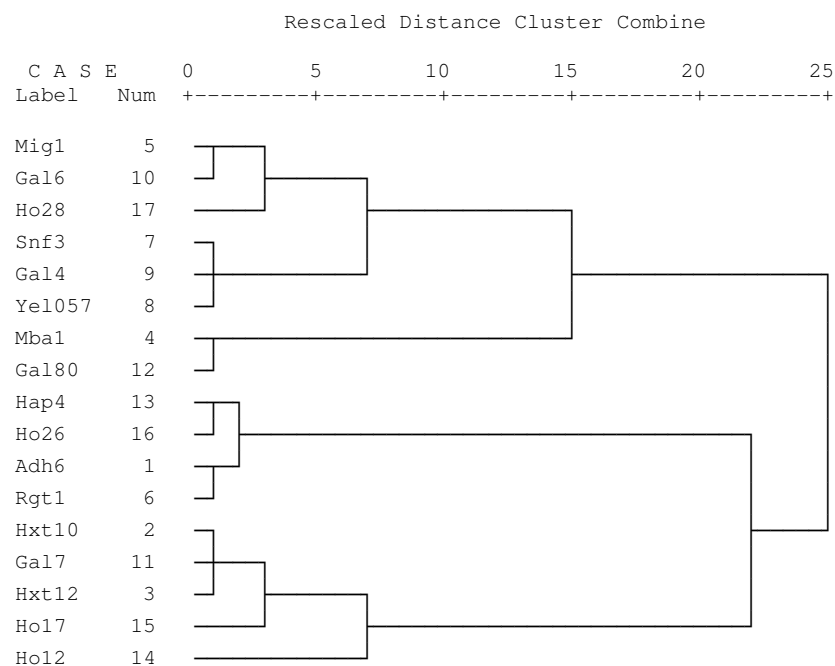


Figure B.45. HCA of IC1 and IC2 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

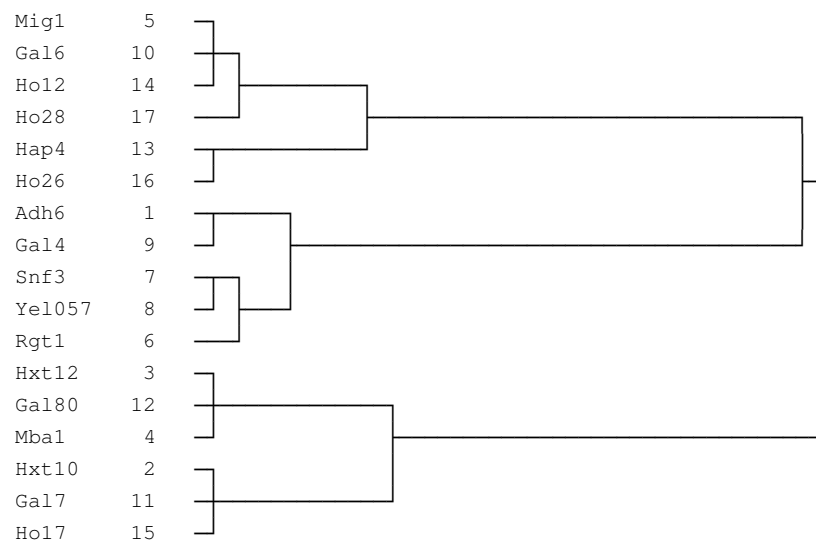


Figure B.46. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

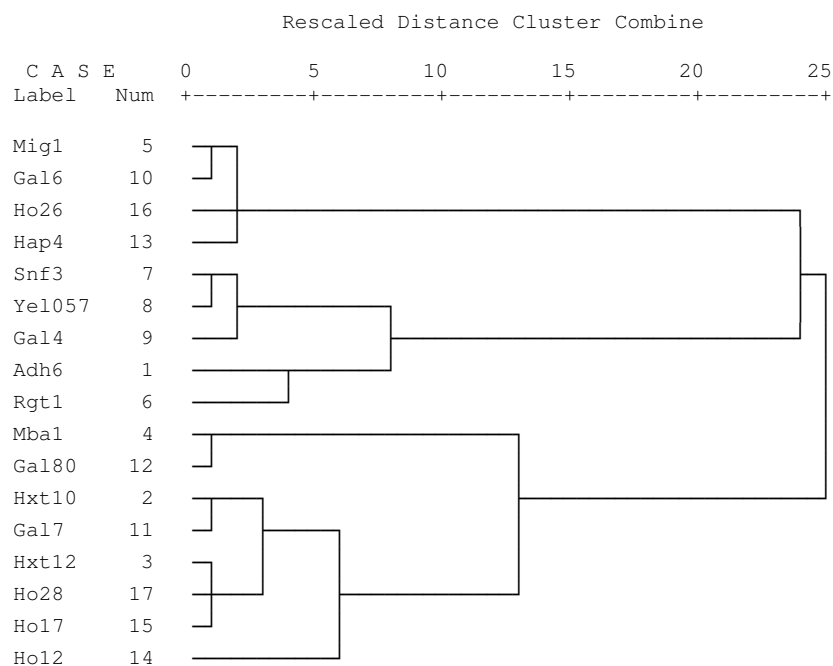


Figure B.47. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

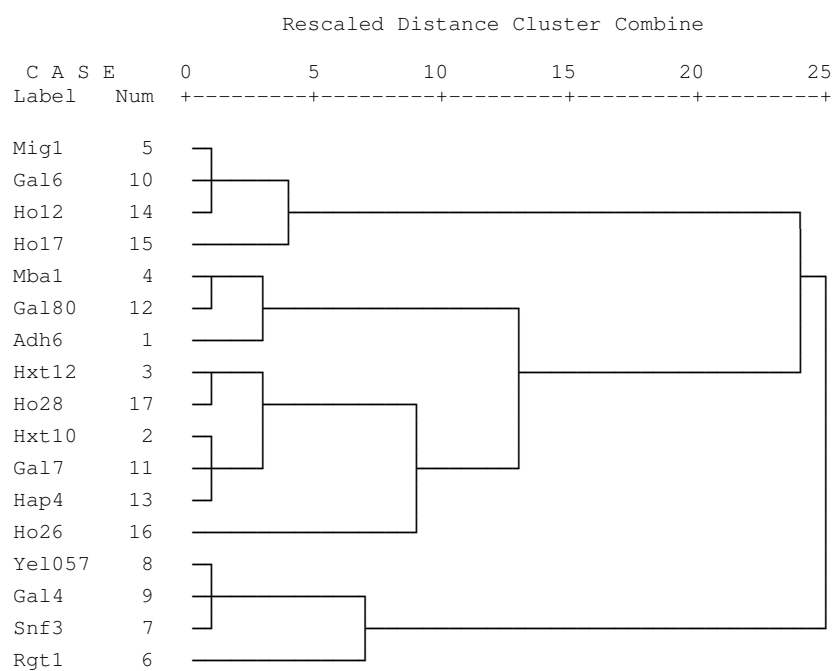


Figure B.48. HCA of IC2 and IC3 by 6 times pre-processing of PCA

B.13. PCA of Filtered Intracellular Metabolome Data at 31st hour

Dendrogram using Average Linkage (Between Groups)

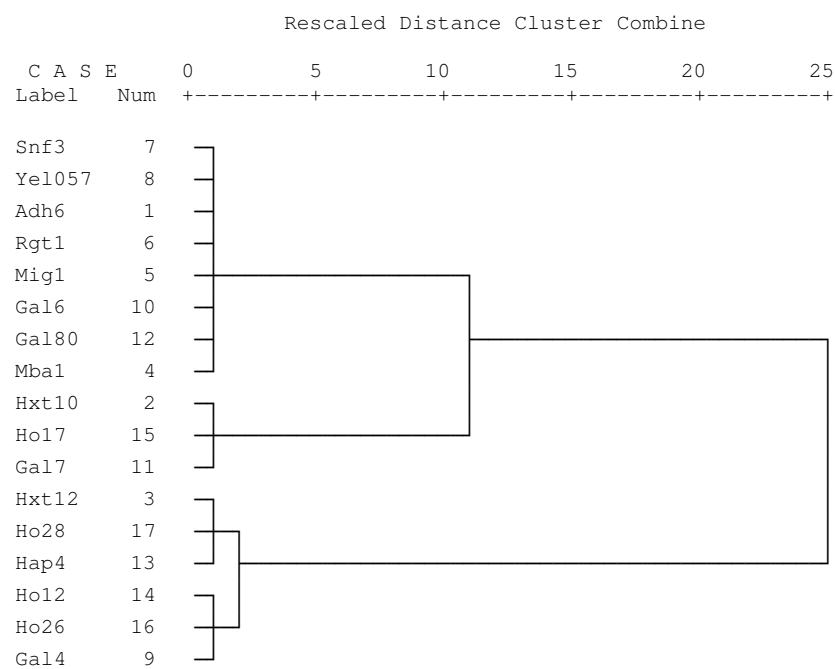


Figure B.49. HCA of PC1 and PC2

Dendrogram using Average Linkage (Between Groups)

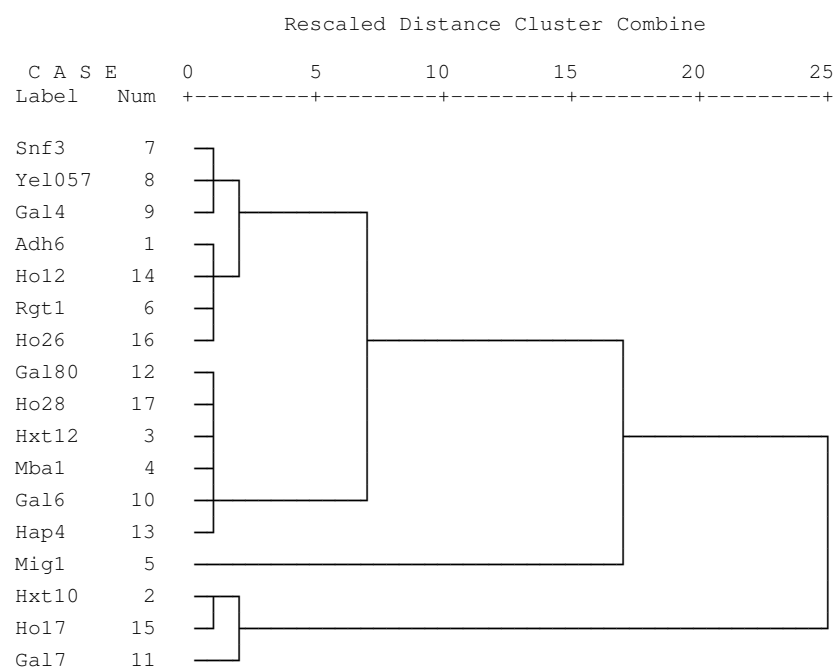


Figure B.50. HCA of PC2 and PC3

Dendrogram using Average Linkage (Between Groups)

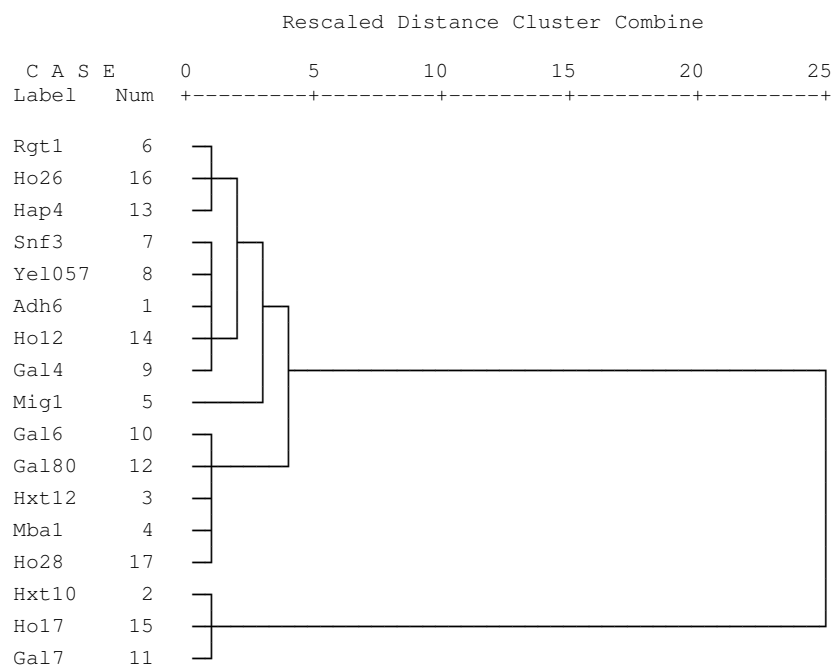


Figure B.51. HCA of PC2 and PC4

Dendrogram using Average Linkage (Between Groups)

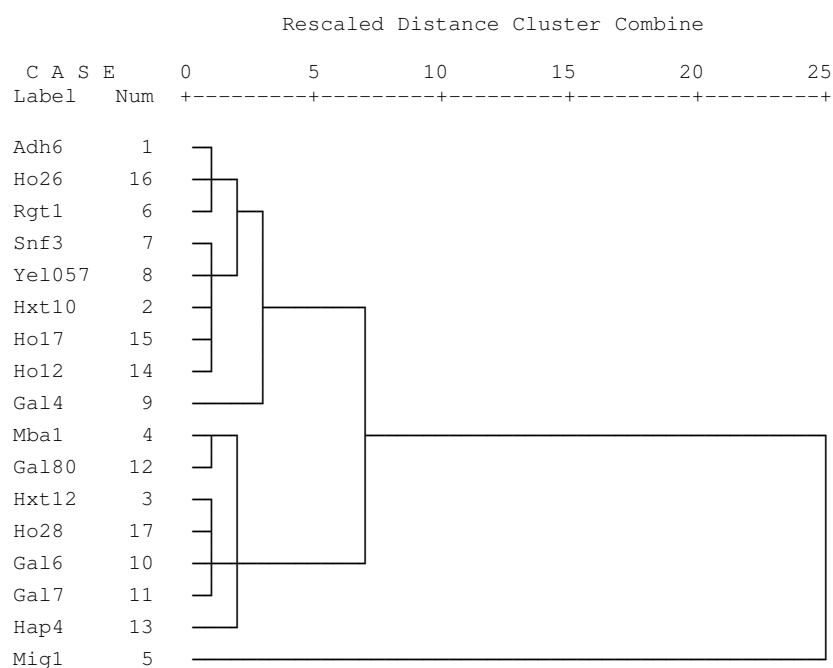


Figure B.52. HCA of PC3 and PC4

B.14. ICA of Filtered Intracellular Metabolome Data at 31st hour

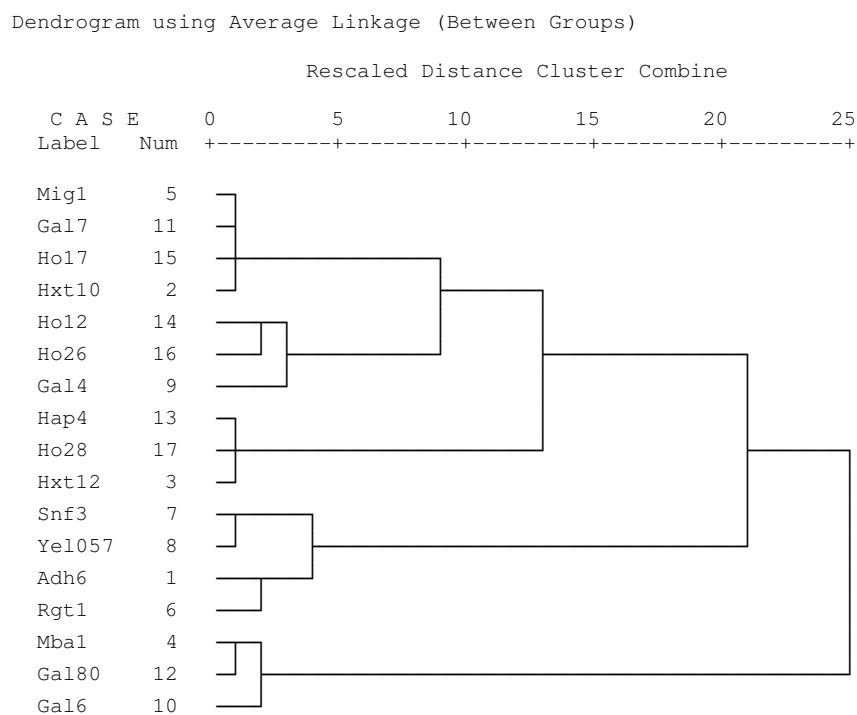


Figure B.53. HCA of IC1 and IC2 by 5 times pre-processing of PCA

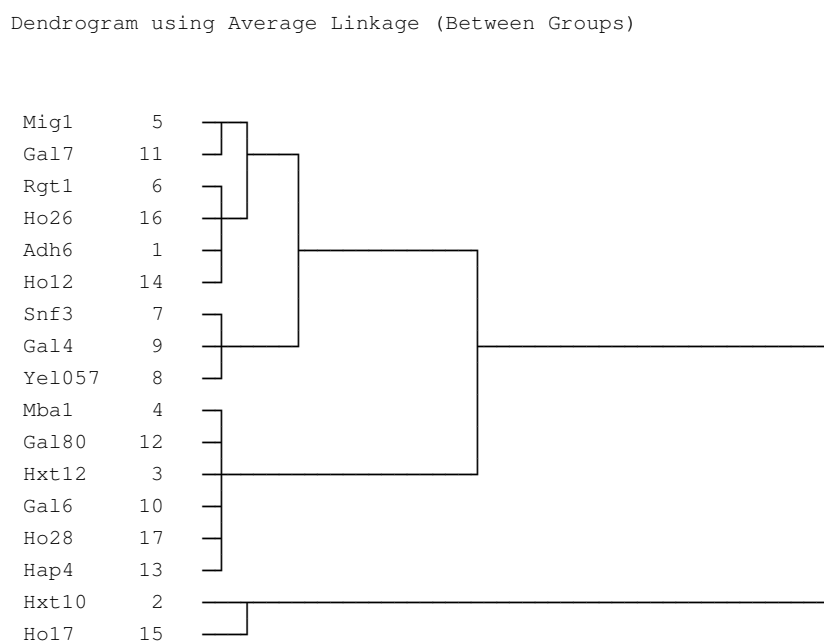


Figure B.54. HCA of IC2 and IC3 by 5 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

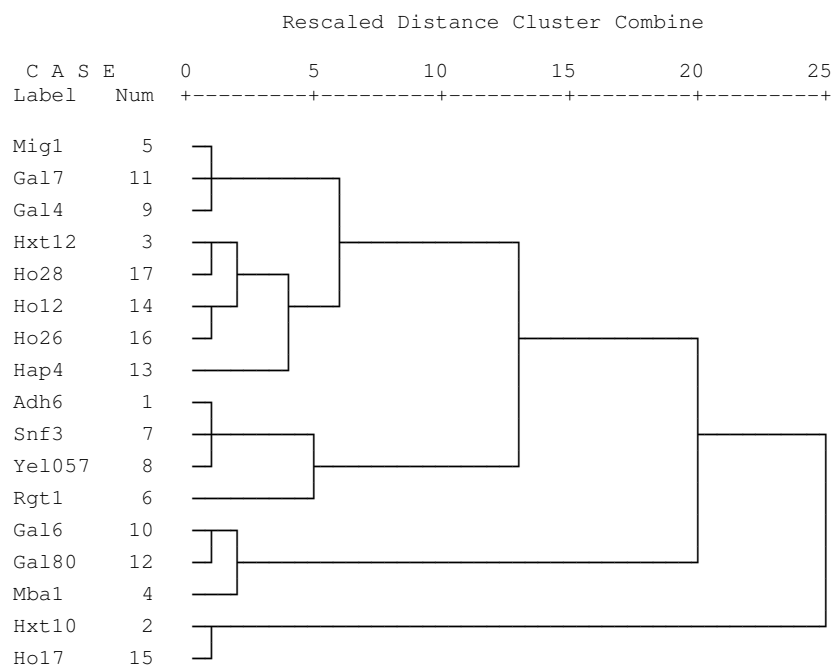


Figure B.55. HCA of IC1 and IC2 by 6 times pre-processing of PCA

Dendrogram using Average Linkage (Between Groups)

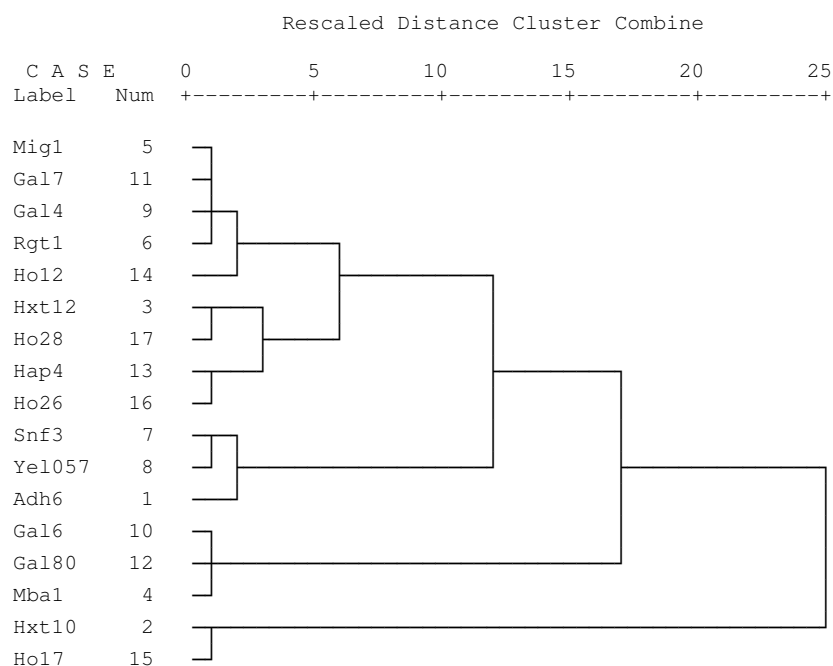


Figure B.56. HCA of IC2 and IC3 by 6 times pre-processing of PCA

REFERENCES

- Askree, S., 2004, “A genome-wide screen for *Saccharomyces cerevisiae* deletion mutants that affect telomere length”, *Proc. Natl. Acad. Sci. U S A*, Vol. 101, pp. 8658-8663.
- Allen, J., H. M. Davey, D. Broadhurst, J. K. Heald, J. J. Rowland, S. G. Oliver and D. B. Kell, 2003, “High-throughput classification of yeast mutants for functional genomics using metabolic footprinting”, *Nature Biotechnology*, Vol. 21, No. 6, pp. 692-696.
- ApRees, T., 1980, “Integration of pathways of synthesis and degradation of hexose phosphates”, pp. 1–29 in *The Biochemistry of Plants*, Academic Press, New York.
- Baganz, F., A. Hayes, D. Marren, D. C. Gardner, S. G. Oliver, 1997, “Suitability of replacement markers for functional analysis studies in *Saccharomyces cerevisiae*”, *Yeast*, Vol. 16, No. 13, pp. 1563-1573.
- Berry, Dr., G. Steward, I. Russell, *Yeast Biotechnology*, Allen Unwin, London.
- Blom, J., J. T. de Mattos, M. J. Teixeira, L. A. Grivell, 2000, “Redirection of the Respiratory Fermentative Flux Distribution in *Saccharomyces cerevisiae* by Overexpression of the Transcription Factor HAP4p”, *Applied and Environmental Microbiology*, Vol. 66, No. 5, pp. 1970–1973.
- Buchholtz, A., J. Hurlbaush, C. Wandrey, R. Takors, 2002, “Metabolomics: quantification of intracellular metabolite dynamics”, *Biomolecular Engineering*, Vol. 19, No. 1, pp. 5-15.
- Buschlen S., J. M. Amillet, B. Guiard, A. Fournier, C. Marcireau, M. Bolotin, 2003, “The *S. cerevisiae* HAP complex, a key regulator of mitochondrial function, coordinates nuclear and mitochondrial gene expression”, *Comparative and Functional Genomics*, Vol. 4, pp. 37–46.

- Carlson, M., 1999, "Glucose repression in yeast", *Curr Opin Microbiol*, Vol. 2, pp. 202-207.
- Christacos, N., 2000, "Subcellular localization of galactose-1-phosphate uridylyltransferase in the yeast *Saccharomyces cerevisiae*", *Mol. Genet. Metab.*, Vol. 70, No. 4, pp. 272-280.
- DeRisi, J. L., V. R. Iyer, P. O. Brown, 1997, "Exploring the metabolic and genetic control of gene expression on a genomic scale", *Science*, Vol. 278, No. 5338, pp. 680-686.
- Dysvik, B. and I. Jonassen, 2001, "J-express: exploring gene expression data using Java", *Bioinformatics*, Vol. 2, pp. 96-103.
- Fernie, A. R., R. N. Trethewey, A. J. Krotzky, L. Willmitzer, 2004, "Metabolite profiling: from diagnostics to systems biology", *Nat. Rev. Mol. Cell Biol.*, Vol. 5, pp. 1-7.
- Fiehn, O., J. Kopka, P. Dormann, T. Altmann, R. Trethewey, L. Willmitzer, *Nat Biotechnol* 2000, Vol. 18, pp. 1157.
- Fiehn, O., *Comparative Functional Genomics* 2001, Vol. 2, pp. 155.
- Fiehn, O., *Plant Molecular Biology* 2002, Vol. 48, No. 1, pp. 155-71.
- Gancedo, J. M., 1998, "Yeast Carbon Catabolite Repression", *Microbiology and Molecular Biology Review*, Vol. 62, No. 2, pp. 334-361.
- Goodacre, R., B. Shann, R. J. Gilbert, 2000, "Detection of the dipicolinic acid biomarker in *Bacillus* spores using curiepoint pyrolysis mass spectrometry and fourier transform infrared spectroscopy", *Anal. Chem.*, Vol. 72, pp. 119-127.

- Goodacre, R., S. Vaidyanathan, W. B. Dunn, G. G. Harrigan, D. B. Kell, 2004, "Metabolomics by numbers: acquiring and understanding global metabolite data", *Trends Biotechnol.*, Vol. 22, pp. 245–252.
- Gygi, S. P., Y. Rochon, B. R. Franza, R. Aebersold, 1999, "Correlation between Protein and mRNA Abundance in Yeast", *Molecular and Cellular Biology*, Vol. 19, pp. 1720-1730.
- Hutter, A. and S. G. Oliver, 1998, "Ethanol production using nuclear petite yeast mutants", *Applied Microbiology and Biotechnology*, Vol. 49, pp. 511-516.
- Hyvarinen, A. and E. Oja, 2000, "Independent component analysis: Algorithms and applications", *Neural Networks*, Vol. 4–5, No. 13, pp. 411–430.
- Keegan, L., 1986, "Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein", *Science*, Vol. 231, No. 4739, pp. 699-704.
- Kell, D. B., 2004, "Metabolomics and systems biology: making sense of the soup", *Current Opinion in Microbiology*, Vol. 7, pp. 296-307.
- Kell, D.B., 2004, "Metabolomics and systems biology: making sense of the soup", *Curr. Opin. Microbiol.*, Vol. 7, No. 3, pp. 296–307.
- Kell, D. B., 2002. Metabolomics and machine learning: explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules. *Mol. Biol. Rep.* 29, 237–241.
- Kruckeberg, A., 1996, "The hexose transporter family of *Saccharomyces cerevisiae*", *Arch. Microbiol.*, Vol. 166, No. 5, pp. 283-292.

- Larroy, C., 2002, "Characterization of the *Saccharomyces cerevisiae* YMR318C (ADH6) gene product as a broad specificity NADPH-dependent alcohol dehydrogenase: relevance in aldehyde reduction", *Biochem. J.*, Vol. 361, pp. 163-172.
- Lohr, D., 1995, "Transcriptional regulation in the yeast GAL gene family: a complex genetic network", *FASEB J.*, Vol. 9, No. 9, pp. 777-787.
- Melcher, K. and H. Xu, 2001, "Gal80-Gal80 interaction on adjacent Gal4p binding sites is required for complete GAL gene repression", *EMBO J.*, Vol. 20, No. 4, pp. 841-851.
- Morgenthal, K., S. Wienkoop, M. Scholz, J. Selbig, W. Weckwerth, 2005, "Correlative GC-TOF-MS based metabolite profiling and LC-MS based protein profiling reveal time-related systemic regulation of metabolite-protein networks and improve pattern recognition for multiple biomarker selection", *Metabolomics*, Vol. 1, pp. 109-121.
- Oliver, S. G., M. K. Winson, D. B. Kell, F. Baganz., 1998, "Systematic functional analysis of the yeast genome", *TIBTECH*, Vol. 16, pp. 373-378.
- Ozcan, S. and M. Johnston, 1999, "Function and regulation of yeast hexose transporters", *Microbiol. Mol. Biol. Rev.*, Vol. 63, No. 3, pp. 554-69.
- Ozcan, S., 1996, "Rgt1p of *Saccharomyces cerevisiae*, a key regulator of glucose-induced genes, is both an activator and a repressor of transcription", *Mol. Cell. Biol.*, Vol. 16, No. 11, pp. 6419-6426.
- Phaff, H. J., M. W. Miller, E. M. Mrak, 1978, *The Life of Yeast*, 2nd edition, Harvard University Press, Cambridge (USA).

- Pir, P., 2005, *Integrated Analysis of Metabolome Profiles and Gene Expression in Respiratory Deficient Deletion Mutants of Saccharomyces Cerevisiae*, Ph.D. Thesis, Bogazici University.
- Preuss, M., K. Leonhard, K. Hell, R. A. Stuart, W. Neupert, J. M. Herrmann, 2001, “Mba1, a novel component of the mitochondrial protein export machinery of the yeast *Saccharomyces cerevisiae*”, *Journal of Cell Biology*, 2001, Vol. 153, No. 5, pp. 1085-1096.
- Raamsdonk, L. M., B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell, J. J. Rowland, H. V. Westerhoff, K. van Dam, S. G. Oliver, 2001, “A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations”, *Nature Biotechnology*, Vol. 19, No. 1, pp. 45-50.
- Rep, M. and L. A. Grivell, 1996, “MBA1 encodes a mitochondrial membrane-associated protein required for biogenesis of the respiratory chain”, *FEBS Letters*, Vol. 388, No. 2-3, pp. 185-188.
- Steinmetz, L. M., C. Scharfe, A. M. Deutschbauer, D. Mokranjac, Z. S. Herman, T. Jones, A. M. Chu, G. Giaever, H. Prokisch, P. J. Oefner, R. W. Davis, 2002, “Systematic screen for human disease genes in yeast”, *Nature Genetics*, Vol. 31, pp. 400-404.
- Steuer, R., J. Kurths, O. Fiehn, W. Weckwerth, 2003 a, “Interpreting correlations in metabolic networks”, *Biochem. Soc. Trans.*, Vol. 31, No. 6, pp. 1476–1478.
- Steuer, R., J. Kurths, O. Fiehn, W. Weckwerth, 2003 b, “Observing and interpreting correlations in metabolomic networks”, *Bioinformatics*, Vol. 19, No. 8, pp. 1019–1026.
- Stitt, M., I. Wilke, R. Feil, H. W. Heldt, 1988, “Coarse control of sucrose-phosphate synthase in leaves—alterations of the kinetic properties in response to the rate of photosynthesis and the accumulation of sucrose”, *Planta*, Vol. 174, pp. 217–230.

- Sturn, A., J. Quackenbush, Z. Trajanoski, 2002, "Genesis: cluster analysis of microarray data", *Bioinformatics*, Vol. 18, pp. 207–208.
- Sumner, L. W., P. Mendes, R. A. Dixon, 2003, "Plant metabolomics: large-scale phytochemistry in the functional genomics era", *Phytochemistry*, Vol. 62, pp. 817–836.
- Taylor, J., R. Goodacre, W. G. Wade, J. J. Rowland, D. B. Kell, 1998, "The deconvolution of pyrolysis mass spectra using genetic programming: application to the identification of some *Eubacterium* species", *FEMS Microbiol. Lett.*, Vol. 160, pp. 237–246.
- Taylor, J., R. D. King, T. Altmann, O. Fiehn, 2002, "Application of metabolomics to plant genotype discrimination using statistics and machine learning", *Bioinformatics*, Vol. 18, pp. 241–248.
- Timson, D., 2002, "Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry", *Biochem. J.*, Vol. 363, pp. 15–20.
- Torchia, T., 1984, "Disruption of regulatory gene GAL80 in *Saccharomyces cerevisiae*: effects on carbon-controlled regulation of the galactose/melibiose pathway genes", *Mol. Cell. Biol.*, Vol. 4, No. 8, pp. 1521–1527.
- Tweeddale, H., L. Notley-McRobb, T. Ferenci, 1998, *J. Bacteriology*, Vol. 180, pp. 5109–5116.
- Vapnik, V., 1995, *The nature of statistical learning theory*, Springer Verlag, New York.
- Verduyn, C., E. Postma, W. A. Scheffers, J. P. Van Dijken, 1992, "Effect of benzoic acid on metabolite fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation", *Yeast*, Vol. 8, pp. 501–517.

- Walker, G. M., 1998, *Yeast Physiology and Biotechnology*, John Wiley and Sons, New York.
- Walker, G. M., 2000, *Yeast Physiology and Biotechnology*, John Wiley and Sons, Chichester.
- Weckwerth, W., 2003, “Metabolomics in systems biology”, *Annu. Rev. Plant Biol.*, Vol. 54, pp. 669–689.
- Weckwerth, W., M. Loureiro, K. Wenzel, O. Fiehn, 2004 a, “Differential metabolic networks unravel the effects of silent plant phenotypes”, *Proc. Natl. Acad. Sci. USA*, Vol. 101, pp. 7809–7814.
- Weckwerth, W., K. Wenzel, O. Fiehn, 2004 b, “Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks”, *Proteomics*, Vol. 4, No. 1, pp. 83.
- Weckwerth, W., V. Tolstikov, O. Fiehn, 2001, “Metabolomic characterization of transgenic potato plants using GC/TOF and LC/MS”, in Proceedings of the 49th ASMS Conference on Mass spectrometry and Allied Topics, pp. 1–2.
- Zheng, W., 1997, “The cysteine-peptidase bleomycin hydrolase is a member of the galactose regulon in yeast”, *J. Biol. Chem.*, Vol. 272, No. 48, pp. 7050-7055.