Doctoral Dissertation

Metabolomics-based analysis of gene-metabolite correlations in yeast transcription factor knockouts

Zanariah Binti Hashim

July 2014

Graduate School of Engineering
Osaka University

Table of Contents

List of Abbreviations

(in alphabetical order)

ChIP	chromatin immunoprecipitation
CLS	chronological lifespan
DBD	DNA-binding domain
DNA	deoxyribonucleic acid
GC/MS	gas chromatography/mass spectrometry
HCA	hierarchical clustering analysis
IS	internal standard
LC/MS	liquid chromatography/mass spectrometry
MRM	multiple reaction monitoring
mRNA	messenger RNA
mtDNA	mitochondrial DNA
MVA	multivariate (data) analysis
NMR	nuclear magnetic resonance
ORF	open reading frame
PC	principal component
PCA	principal component analysis
PIPES	1,4-piperazinediethanesulfonic acid
PPP	pentose phosphate pathway
QC	quality control
RLS	replicative lifespan
RNA	ribonucleic acid
rRNA	ribosomal RNA
RSD	relative standard deviation
SC	synthetic complete (medium)
SD	synthetic defined (medium)
TBA	tributylamine
TBP	TATA-box binding protein
TCA	tricarboxylic acid
TF	transcription factor
tRNA	transfer RNA
UV	unit variance (a scaling method)
YPD	yeast peptone dextrose (medium)

Chapter 1

General introduction

Cellular functions are determined by integrative interactions between various constituents, i.e., genes, transcripts, proteins, and metabolites. Thus, it is important to study these interactions to understand the whole biological system. Genetic perturbations are often used to investigate the contribution of individual components. One of such components is a transcription factor. Transcription factors are the regulatory proteins that interact with DNA to either promote or suppress gene expression. Due to the importance of transcription factors in gene regulation, they have been widely studied and much attention has been paid regarding the roles of transcription factors. In this thesis, the effects of transcription factor-related gene deletion towards metabolic levels were studied, using the budding yeast *Saccharomyces cerevisiae* as a model organism.

1.1. Yeast as a model organism

1.1.1. Yeast transcription factors

Since its completion of genome sequencing in 1996 [1], research on yeast *Saccharomyces cerevisiae* has shifted from merely decoding the DNAs to understanding the function of the genes, i.e., functional genomics. *S. cerevisiae* has been used for a model eukaryote for its convenience to handle and manipulate genetically, fast growth and short generation time. Although this single-celled eukaryote is much simpler than multicellular organisms, the cell cycle is very similar to the cell cycle in humans. Up to 30% of genes implicated in human disease may have orthologs in the yeast proteome [2], and many studies regarding aging, apoptosis, metabolism and gene expression have been performed using yeasts [3]. Additionally, *S. cerevisiae* is an industrially important microorganism, used in many fields, from food industry to the production of chemicals. The use of *S. cerevisiae* has been assisted by vast literatures and curated databases. Examples are SGD (Saccharomyces Genome Database, http://www.yeastgenome.org/), YEASTRACT (Yeast Search for Transcriptional Regulators

And Consensus Tracking, http://www.yeastract.com/) and YMDB (Yeast Metabolome Database, http://www.ymdb.ca/).

Cells employ an elaborate and complex gene expression system that allows them to reprogram their genetic makeup in response to different environments and growth demands. At the forefront of this control system lies transcription factors (TFs). TFs are the regulatory proteins that initiate or suppress gene expression, directly by binding to the promoter regions in the DNA, or indirectly by forming complex with other TFs. For decades, researchers have been interested in the function and regulation of TFs; biotechnologically this knowledge can help in the improvement of industrially important microbial strains as demonstrated by global transcriptional machinery engineering (gTME) technique [4,5], and clinically TFs themselves can serve as potential drug targets such as estrogen receptors and c-Myc for cancer therapies [6,7], and generally proposed for new drug discovery [8,9].

Yeast has been used to study eukaryotic transcriptional regulatory mechanisms as well. Transcriptional regulatory mechanisms are fundamentally similar in eukaryotes, in which complex promoters with multiple protein binding sites are typical [10]. Components of the basal RNA polymerase II machinery and several general transcription factors have been determined, and the yeast system has been the leading model for these discoveries [11].

Generally, TFs can be categorized based on their transcription modes or protein structures (DNA binding domain motifs). In terms of transcription modes, TFs can be divided into three classes; 1) core machinery transcriptional component (i.e., basal transcription factors) that binds to the DNA promoter region, examples are the TATA-binding proteins; 2) activator or repressor proteins (i.e., sequence-specific binding proteins) that recognize specific DNA sequences and directly bind to the UAS (upstream activation sequence) or URS (upstream repression sequence); and 3) co-activator proteins that do not by themselves bind to the DNA but instead interact with other TFs to activate gene expression machinery.

On the basis of DNA-binding domain (DBD), TFs are categorized into three general classes: zinc-stabilized, zipper type and helix-turn-helix (**Fig. 1-1**, reviewed by Hahn and Young, 2011[11]). In the zinc-stabilized class, the TFs can be further classified into three sub-classes: C2H2 zinc fingers, C6 (zinc knuckle or Zn_2Cys_6 binuclear zinc cluster) and C4 (or GATA fingers). C2H2 and C4 are ubiquitous while C6 is unique to fungi. The zinc-stabilized DBD is

the most abundant in all organisms, and as the name indicates, requires Zn^{2+} to stabilize. At least one TF in yeast, i.e., Ace1/Cup2 is stabilized by Cu^{2+}. There are 53 members in C2H2 (e.g., Adr1, Mig1, Zap1), 55 members in C6 (e.g., Gal4, Hap1, Leu3) and 5 members in C4 (Gln3, Gat1, Nil1, Dal80, Ash1) proteins.

The second most abundant TF class is the zipper type. DBD of this class is characterized by a dimerization motif and a basic region. There are two sub-classes of zipper type: bZIP (basic leucine zippers; 14 members, e.g., Gcn4, Yap1, Sko1) and bHLH (basic helix-loop-helix; 9 members, e.g., Ino2, Ino4, Rtg1, Rtg3). Zipper type TFs form homo- or heterodimers, a feature that enables multi-regulatory control of transcription. Consequently, zipper type TFs involve in many processes including cell development and stress responses.

The third class, HTH (helix-turn-helix; 8 members, e.g., *Mat*α1, *Mat*α2, *Mat***a**1), also forms homo- and heterodimers. A classical HTH protein in yeast is *Mat*α2, which, together with Mcm1, represses **a**-specific genes in *Mat*α haploids. The forkhead (Fkh) transcription factors (Mcm1, Fkh1, Fkh2, Mcm1) and the heat shock factor (HSF) are related to the HTH proteins.

However, there are also TFs that lack a DBD motif, such as Met4 and Swi6, while other TFs such as Gcr1 and Dal81 have a DBD that is dispensable. These proteins form a heterodimer and interact with DNA through their binding partner.

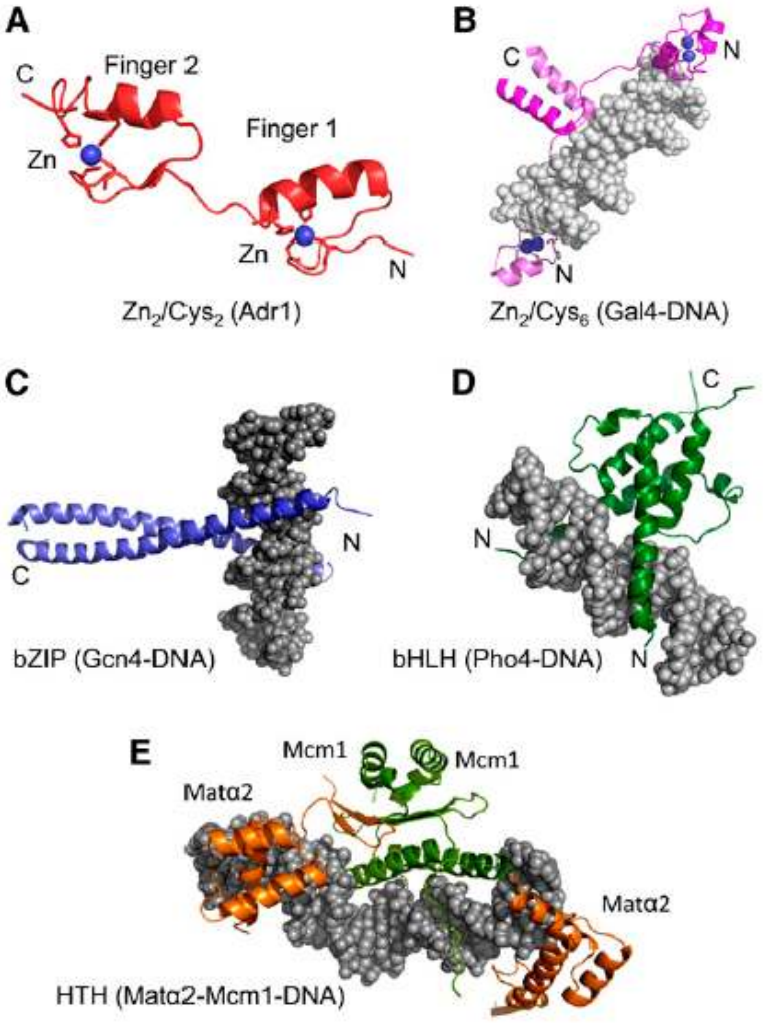


Fig. 1-1. Classification of yeast transcription factors according to DNA binding domain motifs. (A) C2H2 zinc fingers of Adr1; (B) C6 (zinc knuckle) of Gal4; (C) bZIP structure of Gcn4; (D) bHLH of Pho4; (E) helix-turn-helix of *Mat*α2 and winged helix of Mcm1. (Hahn and Young, 2011[11])

1.1.2. Transcription factors and gene regulatory studies

Because of the important role of TFs in gene expression, various studies have been undertaken to find which TFs are involved in the expression of a certain set of genes, when and why the genes are expressed, and more importantly the consequences of such gene expression. Large scale gene-protein and protein-protein interactions and transcript analysis have contributed to the vast knowledge of TFs, following current advancement in microarray and ChIP techniques. Several strategies commonly employed are; computer-based approach such

as development of algorithms that seek conserved promoter elements or common sequence elements in the promoters of co-expressed genes [12,13], microarray-based biochemical approaches that identifies binding and the sequence of individual TFs [14–16], and ChIP-chip [17–19] that identifies sequences bound by a TF *in vivo*.

Despite the large amount of studies, the understanding of TFs is far from complete due to the complex nature of gene regulation; multiple-stage control and modularity element of TFs, post-transcriptional and post-translational modifications, and lack of information regarding the signaling molecules itself and intercellular communication that all lead to gene expression. Particularly, the connection between the transcript and protein to final phenotypic change is lacking, and thus an alternative approach to studying TF regulation is necessary. Although transcriptomics can analyze genome-wide gene expression levels and many large-scale genetic perturbations using microarrays have been performed [16,20,21], transcript levels are known to be unstable, and it is difficult to compare such large data generated across different technology platforms, genetic backgrounds and degrees of replication [22]. Moreover, while the expression levels of genes encoding an enzyme can be relatively easy to infer (for example, the upregulation of an enzyme catalyzing a biochemical reaction can be interpreted as increased products and decreased substrates), interpretation of genes encoding e.g., a transporter or a permease may not be as simple. Therefore, additional parameters (in this case, metabolite levels) can help in the interpretation of gene transcription process and its effects to the cells.

1.1.3. S. cerevisiae central carbon metabolism

Yeast has been the subject of study since the 17th century, when it was identified by a Dutch lens maker, Antonie van Leeuwenhoek who first developed the microscope. Today, it is still one of the important microorganisms, due to its industrial relevance in many applications, from bread to winemaking, to the production of chemicals. *S. cerevisiae* is a Crabtree positive yeast, a facultative anaerobe, able to perform alcoholic fermentation of glucose under fully aerobic conditions. The adaptation of *S. cerevisiae*'s metabolism under different conditions represents an excellent model for studying metabolic regulation.

At the center of metabolic network lies the central carbon metabolism. Central carbon metabolism in *S. cerevisiae* refers to the breakdown of carbon source (typically glucose), to produce biosynthetic precursors for biomass formation and generate energy required for growth. It comprises both catabolic (breakdown of large molecules, energy-producing) and anabolic (building up of large molecules from smaller units, energy-requiring) pathways. Central carbon metabolism includes; a) glycolysis: the breakdown of sugar to pyruvate, b) pentose phosphate pathway, PPP: used for the generation of NADPH, c) tricarboxylic acid cycle, TCA or Krebs cycle: generating $FADH_2$ and NADH, which are then used for ATP production under oxidative phosphorylation, and d) glyoxylate cycle: an anaplerotic pathway that ensures continuous supply of intermediates when TCA cycle is compromised, and of growth in 2- and 3-carbon molecules. Related to glycolysis is gluconeogenesis, the anabolic pathway to produce glucose from non-carbohydrate carbon substrates such as pyruvate, lactate, glycerol, glucogenic amino acids, and fatty acids. **Fig. 1-2** illustrates the central carbon metabolism of *S. cerevisiae*.

The central carbon metabolism is highly conserved among various organisms and holds the key to understanding cell regulation under different metabolic states, either caused by genetic or environmental perturbations. Thus, examining alterations at the central metabolic level signifies a fundamental step in functional genomics studies.

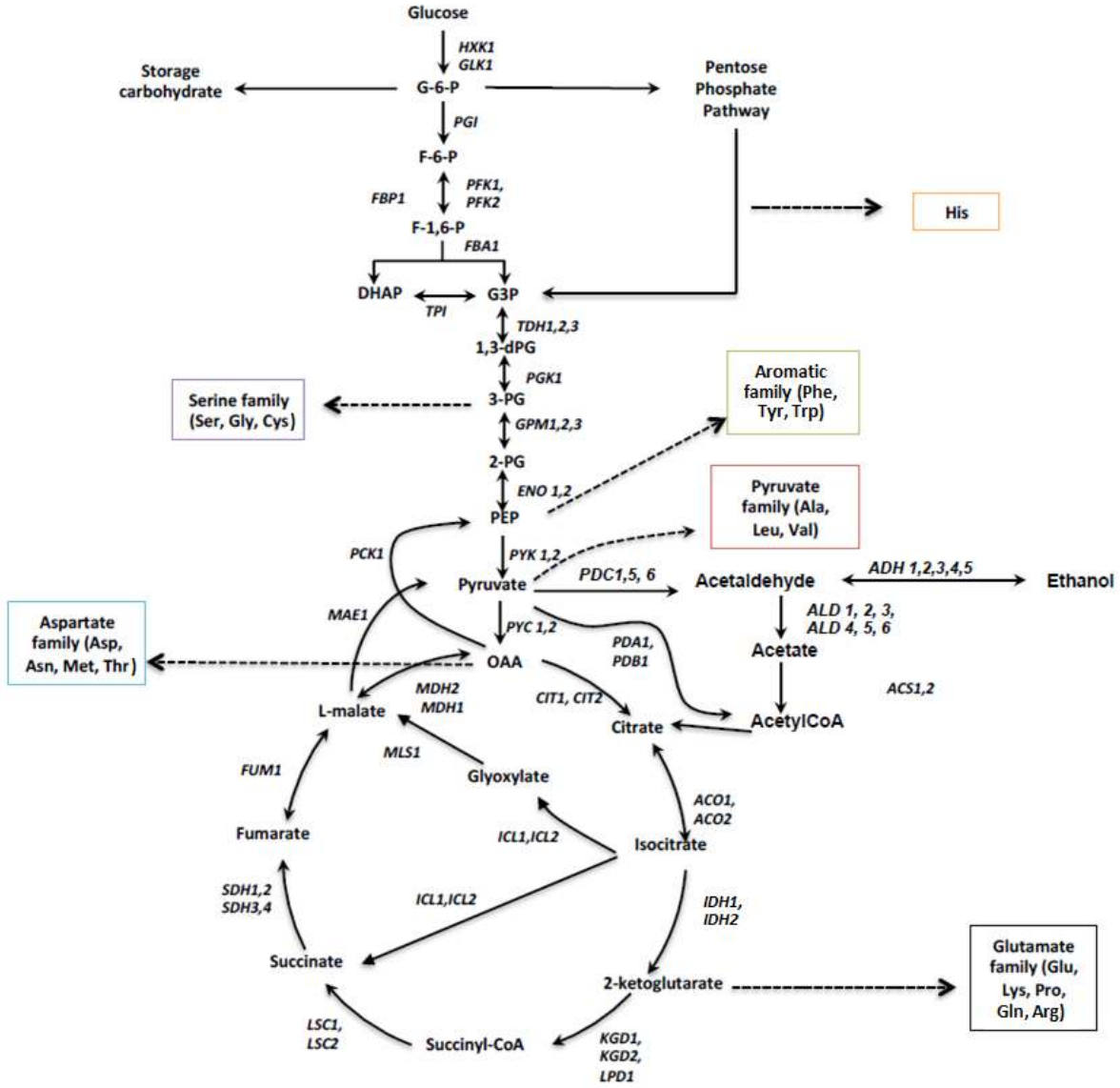


Fig. 1-2. Central carbon metabolism of *S. cerevisiae*

1.2. Metabolomics

1.2.1. General concept

Biological systems are comprised of four main biochemical components, i.e., genes, transcripts, proteins and metabolites. The complete collection of each component is referred to as genome, transcriptome, proteome and metabolome (**Fig. 1-3**). These components interact with each other in an integrative manner to determine cellular phenotypes. Systems level studies

of these biological components on a global scale has been driven by various 'omics' technologies, each built based on the individual component properties. These techniques are driven by high-throughput approaches that yield a large set of data that are often challenging to analyze, but present a holistic view of cellular functions. **Table 1-1** summarizes 'omics' technologies. Besides these four main techniques, several other branched 'omics' such as fluxomics (measurement of the ensemble of metabolic fluxes) and lipidomics (comprehensive profiling of lipid molecules) have also been described.

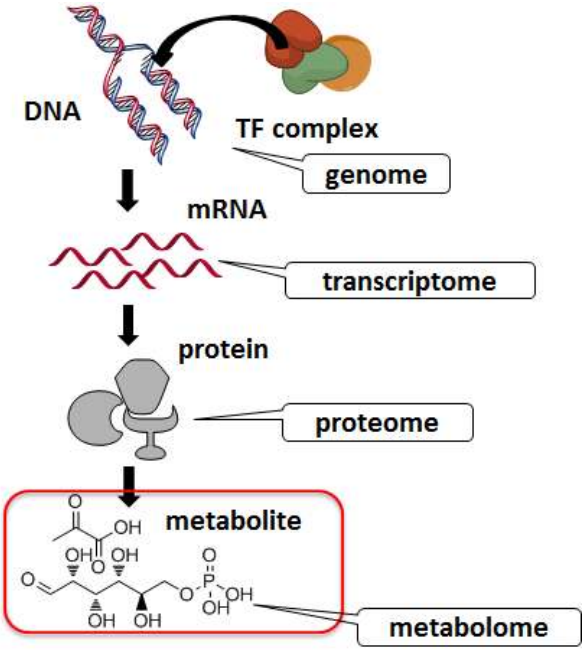


Fig. 1-3. The central dogma of biology, where genes are transcribed into mRNA, which is further translated into protein, which then participates in a metabolic pathway to give rise to a certain metabolite. This schematic diagram also captures the role of transcription factors as the forefront molecule in gene transcription process. By measuring the metabolic alteration following a transcription factor perturbation, elucidation of the TF function can be achieved.

Table 1-1. The 'omics' technologies

Omics	Description	Methods	Applications
Genomics	Comprehensive study of a genome, including protein coding genes, regulatory elements and non-coding sequences	• Gene sequencer	Genome sequence information
Transcriptomics	Quantitative study of mRNA (transcript) expression levels	• Hybridization arrays (microarrays) • RNA sequencer	Analysis of differential gene expression, gene-gene network
Proteomics	Analysis of protein content and abundances	• 2D-PAGE gels • Protein arrays • MALDI-TOF, LC/MS	Identification of protein function, protein-protein interactions
Metabolomics	Comprehensive study of metabolites and metabolic network	• GC/MS, LC/MS • NMR	Identification and quantification of key metabolites, elucidation of metabolic behavior

Metabolites hold a special position in systems biology since they are most downstream products of gene expression process. Transcripts or proteins can undergo various post-transcriptional and post-translational modifications, and thus the changes in transcript or protein abundances do not necessarily lead to an equal change in phenotype. In contrast, metabolites represent the final outcome of gene expression, and thus are the ultimate readouts of a phenotype. Moreover, metabolites also serve as the building block for genes and transcripts (nucleotides), proteins (amino acids) and organelles, and well-preserved among different organisms. In yeast, it is estimated that there are approximately 600-1000 metabolites [23], a number far less than the number of genes or proteins. However, this also means that there is a higher complexity since metabolites have no direct one-to-one relation with genes/proteins and involve in various biochemical reactions simultaneously.

Generally, metabolomics strategies can be largely divided into; targeted analysis, metabolite profiling, metabolomics and metabolic fingerprinting [24,25]. Targeted analysis approach is used when the metabolites of interest are known, and involves quantification of the metabolites. Metabolite profiling, also called semi-targeted approach, is the quantitative or

qualitative determination of a group of related compounds or of specific metabolic pathways. Metabolomics (qualitative and quantitative analysis of all metabolites) and metabolic fingerprinting (sample classification by rapid, global analysis), are related to non-targeted analysis, which typically aims to profile all metabolites, so-called 'measure everything' approach [26]. In this study, a semi-targeted metabolomics approach is employed. Targeted and semi-targeted analyses deal with a defined set of metabolites; the difference is mainly in terms of the number of metabolites measured, typically around 20 for targeted and a hundred to a few hundreds for semi-targeted [26]. The number of metabolites that can be measured is often limited by the number of commercially available authentic metabolite standards. Semi-targeted approach allows for a wider coverage of metabolites than targeted approach, consequently higher chance of finding significant metabolites, with higher accuracy and quantification ability than non-targeted approach. Thus, a good compromise between metabolite numbers and quantification ability is obtained.

1.2.2. Metabolomics approach in this study

To elucidate the complex metabolic alteration following the deletion of a transcription factor, ideally all of the metabolites are measured. However, the diverse chemical properties of metabolites, including molecular weight, polarity, hydrophobicity, volatility, and chemical structures, make simultaneous measurement technically demanding [26]. Recently, ion pairing liquid chromatography/mass spectrometry (LC/MS) has been developed as a widely-targeted metabolome analysis platform that covers a wide range of metabolites [27–31]. Particularly, highly polar intermediates from central metabolism, such as sugar phosphates and nucleotide triphosphates, can be measured with good reproducibility using this platform. By adding an ion pairing reagent in the mobile phase, the ability to retain highly polar metabolites that otherwise are eluted near the void volume in regular reversed-phase LC is improved. Moreover, ion pairing LC has a better separation and higher signals compared to hydrophilic interaction chromatography (HILIC) in the analysis of central metabolites [32]. Therefore, ion pairing LC/MS fits as an analysis platform for metabolic profiling of transcription factor deletion mutants in this study. The use of triple quadrupole mass spectrometry (MS/MS) helps to separate isomers with an additional filter at the third quadrupole and improve selectivity.

In this study, ion pairing LC/MS is used as the main analysis platform for metabolite profiling. Tributyl amine (TBA) is added to the mobile phase as an ion pairing reagent for the separation of anionic metabolites. In addition, regular reversed-phase LC/MS and GC/MS were also used as complementary platforms for the measurement of important metabolites undetected or having poor performance in ion pairing LC/MS.

The overall workflow for this study is depicted in **Fig. 1-4**. First, analytical platform dedicated for the profiling of yeast samples was developed. This step includes analysis of standard metabolites, analysis of a reference strain and construction of an in-house yeast metabolite library. Next, metabolic profiling of selected yeast strains was carried out, using the established analysis platform. In this step, yeast samples were obtained after culture and metabolite extraction, then subjected to metabolite measurement. In the subsequent step, peak identification was performed, after which a peak list table (metabolome dataset) was obtained. Finally, after suitable data pre-processing (normalization and scaling), multivariate data analysis was conducted.

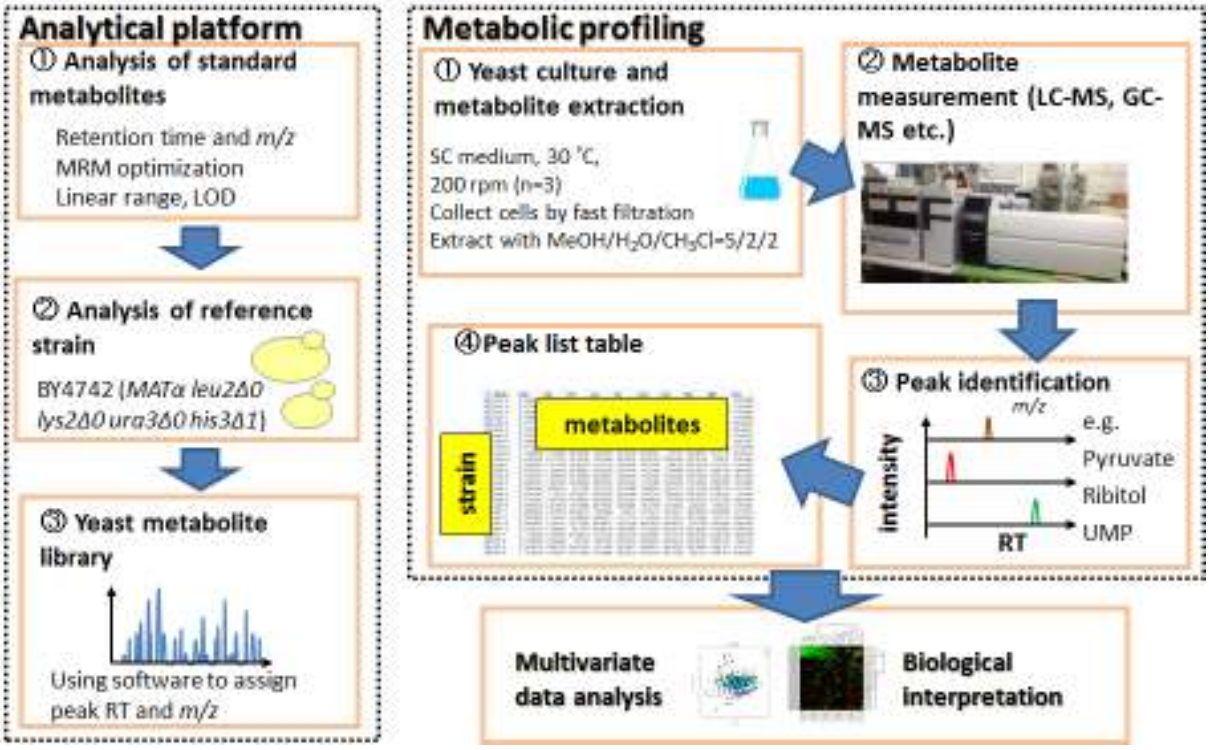


Fig. 1-4. Schematic diagram of experimental workflow and strategies employed in this study.

1.3. Research objective

As described in *Section 1.1.2*, the knowledge of transcription factors still has a long way to go. Particularly, the missing link between TFs to final phenotypic change needs to be addressed. To fill the gap of the connection between TFs and phenotype (cellular composition, physiological appearances, growth competency etc.), the most downstream product of gene expression i.e., metabolites should be characterized. Metabolites can serve as an indicator of the effects of such gene up- or down- regulation caused by TF deletion. Therefore, metabolite profiling fits as an excellent approach for studying metabolic alteration in TF deficient strains.

In this study, a dispensable activator/repressor group of TFs (non-essential for growth) is dealt with. Intermediates from the central carbon metabolism and related amino acid biosynthetic pathways are measured following TF-gene perturbation, and the correlations between TF-metabolite are derived. The overall goal of this study is to deepen the knowledge of TFs and transcriptional regulation by examining metabolic alteration levels. To achieve this goal, the following strategies were set;

1. Demonstrate the utility of metabolomics in finding novel TF-metabolite correlations using a model transcription factor complex

2. Perform a global metabolome analysis for a comprehensive set of TF deletion strains

3. Use the obtained TF-metabolite correlation data to suggest possible new associations and/or further enhance existing knowledge

1.4. Outline of the thesis

This thesis regards the effects of transcription factor deletion towards metabolic alteration. Specifically, the correlations between a TF gene and metabolites were investigated. In Chapter 1, general introduction and research background are presented. In particular, yeast transcription factors and metabolomics techniques are discussed. In Chapter 2, metabolic profiling of two representative basic helix-loop-helix (bHLH) transcription factors Rtg1 and Rtg3 yeast mutant is demonstrated as a proof-of-principle of the utility of metabolomics approach in finding TF-metabolite correlations. Using exploratory data analysis, information regarding important metabolites that discriminate between mutant and wild-type strain was obtained. Of note was the identification of metabolites/metabolic pathways previously unidentified from other approaches, and metabolic changes in the early growth phase. In Chapter 3, a global metabolome analysis was performed for 154 TF deletion strains. Characterization using hierarchical clustering analysis (HCA) and differential analysis showed that the strains can be categorized according to their metabolic phenotype; some clustering shared similar known function or the same gene annotation, in agreement with previous findings, whereas some demonstrated new associations. These results illustrate that metabolomics can assist in the generation of new working hypotheses of TF functional analysis based on TF-metabolite correlations, which were not necessarily evident from transcript data. Also discussed are issues regarding data normalization and correction of batch-to-batch variation, a prevalent problem in mid- to large-scale metabolomics studies. Finally, in Chapter 4, general conclusions and future perspectives are presented.

Chapter 2

Metabolic profiling of retrograde pathway transcription factors Rtg1 and Rtg3 knockout yeast

2.1. Introduction

In this chapter, as a proof-of-principle, two representative transcription factors were chosen for metabolic profiling. To ensure that the TF deletion will result in a substantial alteration in metabolite levels (so that the difference between wild-type and disruptant strains can be clearly seen in terms of metabolic profile), TFs that are known to affect metabolic pathways were selected. In this regard, Rtg1 and Rtg3, two mitochondrial retrograde pathway regulators that have several target genes in the tricarboxylic acid (TCA) cycle, fit this purpose. Furthermore, mitochondrial retrograde pathway is conserved in many organisms including humans [33,34].

Rtg1 and Rtg3 are two basic helix-loop-helix (bHLH) transcription factors found in yeast *Saccharomyces cerevisiae* and are known regulators of mitochondrial retrograde (RTG) response [47]. bHLH proteins were chosen as they represent a large family of regulators, conserved in all eukaryotes [35–37] and have been widely studied. They engage in diverse metabolic pathways including phosphate utilization, amino acid biosynthesis, and glycolysis. Owing to the characteristic dimer formation of bHLH proteins, it is expected that they involve in various metabolic pathways and are inter-connected with each other as well as other transcription factors [38,39].

Mitochondrial RTG response is the signaling pathway from mitochondria to the nucleus triggered by the functional states of mitochondria [40–42]. **Fig. 2-1** summarizes the regulatory mechanism of retrograde response. This pathway maintains a continuous supply of 2-oxoglutarate, a precursor of glutamate and glutamine biosynthesis, by activating anaplerotic metabolism of citrate and oxaloacetate via glyoxylate cycle when respiratory metabolism through the TCA cycle is compromised in the event of reduced mitochondrial functions. It is

thus one of the important signaling pathways that ensure a continuous supply of precursors, such as 2-oxoglutarate for biosynthetic reactions through alternative metabolic pathways.

Rtg1 and Rtg3 form heterodimers and translocate from the cytoplasm to the nucleus when RTG response is activated [43]. This translocation depends on the phosphorylation state of Rtg3 and the transcriptional activation domain is contained within Rtg3. Rtg1/Rtg3 complex binds to an R-box (GTCAC) which differs from the canonical E-box site (CANNTG) to which most other bHLH proteins bind [44]. Among Rtg1/Rtg3 target genes are several TCA cycle genes, but the prototypical target is *CIT2* [42] which encodes a peroxisomal citrate synthase in *S. cerevisiae*. In petite cells (cells that contain nonfunctional, mutated mtDNA (ρ^-) or have completely lost their mtDNA (ρ^0)), the transcripts encoding TCA cycle and glycolytic enzymes were found to be increased under repressing (i.e., glucose) and derepressing (i.e., raffinose) growth conditions [45], while stimulation of glycolysis was also observed in ρ^- cell when grown under glucose condition, with increased glycerol synthesis and decreased trehalose production [46].

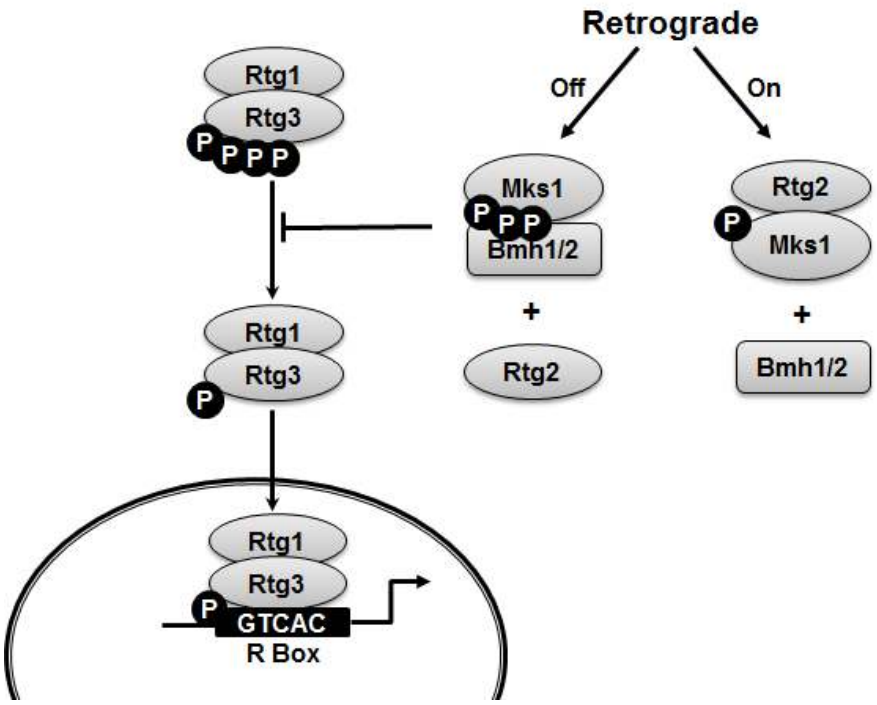


Fig. 2-1. Retrograde regulation in yeast, induced by Rtg1 and Rtg3 complex. Rtg1/Rtg3 is, in turn, regulated by another repressor protein Mks1, whose association with Bmh1/Bmh2 prevents translocation of Rtg1/Rtg3 from the cytoplasm to the nucleus. A phosphatase Rtg2, acts upstream, dephosphorylates and binds to Mks1 to allow retrograde response (adapted from Butow and Avadhani, 2004 [41]).

In addition, the target of rapamycin (TOR) kinase pathway, an essential pathway that controls multiple cellular processes in response to nutritional signals, was shown to negatively regulate RTG target genes [47–49]. Glutamate, glutamine and proline were reported as signaling molecules and both TOR-dependent and TOR-independent modes for RTG target gene expression have been demonstrated [47,50,51]. The interplay between TOR and RTG pathways, as well as the heterodimeric nature of Rtg1/Rtg3 regulators suggest that more complex metabolic regulations exist corresponding to various nutrition and growth conditions. In particular, metabolic signals that regulate TOR and RTG target genes are only partly understood, and it is unclear if the metabolites themselves are regulated by these pathways. In a recent study by Zhang *et al.* (2013) [52], adenosine triphosphate (ATP) was identified as a candidate signaling molecule in the mitochondrial retrograde pathway. However, the association of other metabolites are unknown, and there is a possibility of other metabolic signals since RTG pathway functions to recover the metabolic balance when TCA cycle is repressed. Therefore, characterization of metabolite pools would represent the first screening step to identify these metabolic signals.

In this chapter, metabolomics approach is applied to find metabolic regulations possibly mediated by Rtg1 and Rtg3. While RTG gene deletion exhibited no difference in growth rates when grown in synthetic complete media, a significant alteration in metabolic pathways, especially those involving polyamine biosynthesis, as well as TCA and glyoxylate cycles was observed. It was found that metabolic alterations occur at various metabolic sites, and that these changes relate to different growth phases, but the difference can be detected even at mid-exponential phase. This study illustrates a broader assessment of metabolic change following RTG-gene deletion than previously described.

2.2. Experimental section

2.2.1. Strain growth conditions and sample preparation

Yeast BY4742 (*MATα leu2Δ0 lys2Δ0 ura3Δ0 his3Δ1*) was used as the parental/wild-type strain for all experiments. BY4742 isogenic derivatives, *rtg1*Δ disruptant (*MATα leu2Δ0 lys2Δ0 ura3Δ0 his3Δ1 rtg1*::*kanMX*) and *rtg3*Δ disruptant (*MATα leu2Δ0 lys2Δ0 ura3Δ0 his3Δ1 rtg3*::*kanMX*), were purchased from Open Biosystems (Huntsville, AL, USA). All cultures were grown in synthetic media composed of 0.67% Difco™ yeast nitrogen base without amino acids (BD, MD, USA), 2% glucose (Nacalai Tesque, Kyoto, Japan) and standard concentrations of amino acids and bases [53]. The cultivation and sampling were performed as follows. Yeast cells from frozen glycerol stock were plated onto YPD agar plates (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose, 20 g/L agar (all from BD, MD, USA except glucose and agar from Nacalai Tesque), with added geneticin G418 (Wako, Osaka, Japan) 200 µg/mL for disruptant strains) and grown at 30 °C for 2 days. After two days, a single colony was obtained, transferred to liquid media and let to grow overnight (pre-culture) at 30 °C in a rotary shaker (200 rpm). A portion of the pre-cultured yeast cells were then transferred to fresh media, starting optical density at 600 nm (OD_{600}) of 0.1, and continued to grow at 30 °C to desired OD_{600} and harvested using a rapid filtration system. The harvested cell amount was kept at 5 OD units, equivalent to 1 mg of cells by dry weight at each sampling point. After washing with 5 mL water, the filter-bound cells were inserted into 1 mL of −30 °C precooled single-phase extraction solvent (methanol/chloroform/water = 5/2/2 *v/v/v* %) with added 1.2 µg/mL of 1,4-piperazinediethanesulfonic acid, PIPES (Dojindo, Kumamoto, Japan) as an internal standard and immediately quenched in liquid nitrogen. The samples were kept at −80 °C until extraction. Extraction was carried out at 4 °C, 1200 rpm for 30 min. After that, all liquid extract was transferred to a new tube, 400 µL water added, vortexed and centrifuged at 4 °C, 16100 rcf (relative centrifugal force) for 3 min to separate polar and non-polar phases. The upper polar phase was collected, concentrated five times from the initial volume and ready for LC-MS analysis. Extracted samples were analyzed within 24 h after extraction.

For extracellular metabolome, ~1 mL of the medium filtrate was collected at the same time during cell filtration and diluted four times with water prior to LC-MS analysis.

2.2.2. Metabolite profiling and quantification

The analysis platform consists of a Shimadzu Nexera series UHPLC system (Shimadzu, Kyoto, Japan) coupled to a triple quadrupole mass spectrometer, LCMS-8030, with a modification to improve the sensitivity (Shimadzu, Kyoto, Japan). Two LC/MS methods were employed; (1) ion-pairing reversed phase ultrahigh performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) to detect mainly anionic metabolites, such as sugar phosphates and nucleotides from primary metabolism; and (2) regular reversed phase UHPLC-MS/MS for the analysis of other metabolites undetected in ESI negative mode. For ion-pairing UHPLC-MS/MS, the method was developed based on Luo *et al.* (2007) [29] and Buescher *et al.* (2010) [31], modified to match in-house LC and MS system and tested with several types of columns and analytical parameters to optimize peak shape and separation profile. The MS/MS fragment for each analyte was determined using authentic standards. In addition, the analysis time was successfully accelerated from 36 min [31] to 15 min. The final analytical conditions were as follows; column: L-Column2 ODS (150 mm x 2.1 mm, 3 μm, Chemicals Evaluation and Research Institute Japan); flow rate: 0.3 mL/min; column temperature: 35 °C; mobile phase A: water containing 10 mM tributyl amine and 15 mM acetic acid; mobile phase B: methanol; gradient program: 0% B (0-0.5 min) - 25% B (7.5 min) - 90% B (11-11.5 min) - 0% B (11.6-15 min); sample cycle time: 15 min; injection volume: 3 μL. The mass spectrometric parameters were: ESI negative mode; desolvation line (DL) temperature: 250 °C; nebulizer gas flow: 2 L/min; heat block temperature: 400 °C; other parameters were optimized automatically by flow injection analysis and auto-tuning.

For regular reversed phase UHPLC-MS/MS, the parameters were as follows: column: Discovery HS F5-3 (150 mm × 2.1 mm, 3 μm, Supelco Analytical, PA, USA); flow rate: 0.3 mL/min; column temperature: 40 °C; mobile phase A: water with 0.1% formic acid; mobile phase B: acetonitrile with 0.1% formic acid; gradient program: 0% B (0-1 min) - 20% B (11 min) - 100% B (11.5-13 min) - 0% B (13.1-15 min); sample cycle time: 15 min; injection volume: 3 μL. The mass spectrometric parameters were: ESI positive mode; desolvation line (DL) temperature: 250 °C; nebulizer gas flow: 2 L/min; heat block temperature: 400 °C; other parameters were optimized automatically by flow injection analysis and auto-tuning. The optimized multiple reaction monitoring (MRM) parameters and retention time for each metabolite are listed in Supplementary Table S1.

All samples were kept in a 4 °C autosampler during analysis. Standard mixtures of authentic metabolites and the pooled QC sample [54,55] were injected periodically throughout the analysis run for evaluating the stability and reproducibility of the analytical system. All reagents were of LC-MS grades (Wako, Osaka, Japan).

Peak picking was conducted by LabSolutions (Shimadzu, Kyoto, Japan) followed by manual inspection. The parameters were set as follows: integration: auto, max peak: 3, width: 5 s; smoothing: standard, counts: 5, width: 1 s; identification: absolute RT and closest peak, target window: 5%, reference window: 5%, process time: ±1 min. Obtained peaks were identified based on an in-house metabolite library. The identity was checked by spiking authentic standards to yeast extract and confirming that the particular metabolite peak intensity increases with added concentration. Pooled yeast aliquots were used as a quality control for reproducibility monitoring [54,56]. Peaks with poor reproducibility (relative standard deviation, RSD of peak intensity >30% [55]) were omitted from the list.

2.2.3. Multivariate data analysis

The amount of each metabolite (peak intensity) was normalized to internal standard 1,4-piperazinediethanesulfonic acid (PIPES), mean-centered and scaled to unit variance. Principal component analysis (PCA) was performed using SIMCA-P+ ver13 (Umetrics, Umeå, Sweden). Pathway analysis was performed using MetaboAnalyst 2.0 [57]. Heat map and hierarchical clustering of fold-change normalized intensities were performed on Cluster 3.0 [58] and viewed on Java Treeview [59]. The statistical difference (two-tailed heteroscedastic *t*-test) was calculated using MS Excel. Pathway mapping was performed by VANTED V2.1.0 [60].

2.2.4. Yeast chronological lifespan measurement

The chronological lifespan (CLS) measurement was based on Parella and Longo (2008) [61]. Briefly, aliquots of yeast culture grown to the stationary phase were diluted to approximately 10^3–10^4 cells/mL, and 100 µL were spread onto YPD (1% yeast extract, 2% peptone, 2% dextrose, 2% agar (% *w/v*)) plates. Yeast colonies were counted after 2–4 days of incubation at

30 °C. CLS (at Day X) is defined as the percentage of the number of colonies at Day X divided by the number of colonies at Day 3.

2.3. Results and discussion

2.3.1. Time-course metabolic profiling of RTG-deleted strains

Wild-type BY4742 and RTG-deletion strains showed a comparable growth under SC medium (maximum specific growth rate, $\mu = 0.456 \pm 0.011\ h^{-1}$, $0.461 \pm 0.015\ h^{-1}$ and $0.462 \pm 0.013\ h^{-1}$ for BY4742, *rtg1*Δ disruptant and *rtg3*Δ disruptant respectively). Since RTG response is initiated under decreased mitochondrial and respiratory function, it is expected that declined growth resulting from nutrient cessation and stress accumulation in the stationary growth phase would yield sufficient metabolomics pattern which can distinguish between wild-type and strains lacking RTG response. Therefore, it is appropriate that the metabolic profiling is performed at stationary phase. However, such response against stress or growth adjustment can also be sensed at metabolite levels sooner before there is a detectable change in phenotype, such as demonstrated previously in yeast replicative lifespan study [62]. Therefore, a time-course metabolic profiling should be designed so that the difference between wild-type and *rtg1*Δ disruptant, or *rtg3*Δ disruptant can be captured as early as possible.

The yeast strains grown to stationary phase were sampled at four sampling points. The culture was started at OD_{600}=0.1 (0 h). Each sampling point was taken at various times with different optical density values, OD_{600}, corresponding to different growth phases (OD_{600}=1 at 5 h for mid-exponential, OD_{600}=5 at 9 h for late-exponential, OD_{600}=10 at 26 h for post-diauxic and at 76 h for stationary phases). The collected culture volume was adjusted according to the OD value so that the total cell number for metabolomics profiling is kept constant (approximately 5 x 10^7 cells for each sample). Under high glucose condition, initially, *S. cerevisiae* operates mainly in the glycolytic mode to ferment glucose to ethanol independent of the presence of oxygen. During this stage, the expression of the genes encoding TCA cycle enzymes and other genes required for growth under non-fermentable carbon sources is repressed, a phenomenon known as glucose repression. Mitochondrial function is also repressed. Along with decreased glucose concentration, cells switch to gluconeogenesis and increase their

respiratory rate and, finally, enter the stationary phase, where they accumulate storage carbohydrates. Therefore, the sampling points cover different metabolic states of the cells. Additionally, extracellular metabolites from the growth medium were also measured.

As a result, 96 intracellular metabolites from yeast cell extracts and 53 extracellular metabolites from the growth medium were identified. The metabolite peaks were normalized to an internal standard 1,4-piperazinediethanesulfonic acid (PIPES), mean-centered and scaled to unit variance before subjected to data analysis.

First, to reveal metabolic alteration patterns between RTG disruptants and wild-type strains, principal component analysis (PCA) was performed. PCA is an unsupervised multivariate analysis that seeks the variance among different groups of samples and plots them so that the largest variance is contained in the first principal component, the second largest variance in the second principal component, and so on. Metabolome data (intracellular metabolites) were fitted into PCA with five significant components (Supplementary Table S2). PCA score plot (**Fig. 2-2 (A)**) shows that the first principal component (PC1, accounting for 53.7% of the total variance), separates between different growth phases, while principal component 2 (PC2, accounting for 13.5% of the total variance), separates between wild-type and mutant strains. This result indicated that gene deletion effects can be observed at metabolite levels with high resolution, even when there is no observable change in growth rate.

Next, PCA loading plot (**Fig. 2-2 (B)**) was examined, which shows metabolites that contribute to the separation observed on the score plot (for a complete list of loading values, see Supplementary Table S3). Along PC1, nucleotide monophosphates and ribonucleosides were seen as major contributors to discrimination of samples at late growth phases (26 and 76 h), while proteinogenic amino acids except for proline and cysteine, and glycolysis intermediates were generally abundant in samples at early growth phases. Along PC2, increased level of 2-oxoglutarate and glyoxylate was distinctive in wild-type at 76 h, while putrescine, cAMP, threonine and ornithine were high in RTG-deficient strains.

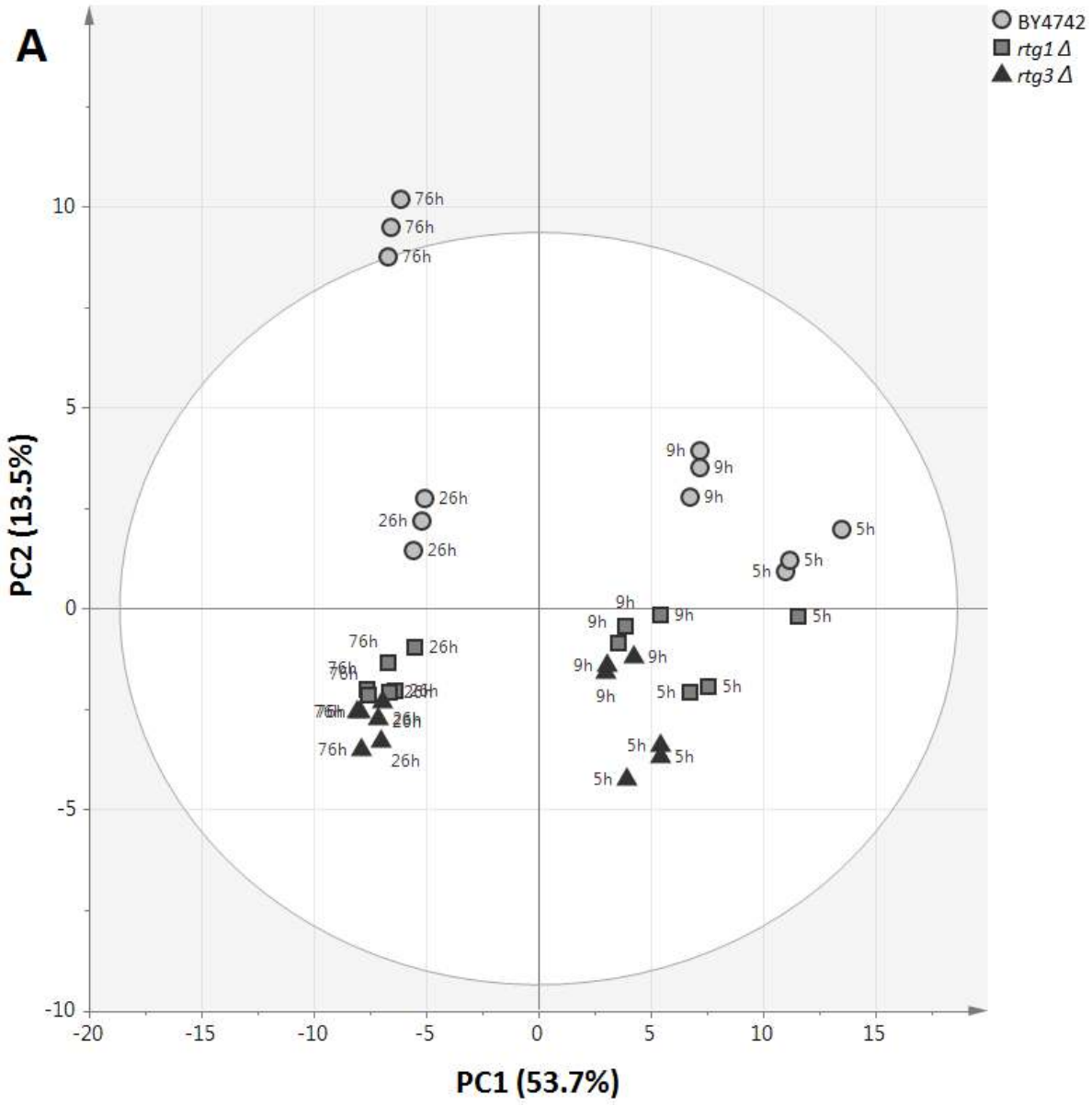


Fig. 2-2. (A) PCA score plot for time-course metabolic profiling at 5 h, 9 h, 26 h and 76 h of control strain BY4742, and *rtg1*Δ and *rtg3*Δ disruptants (n=3). The metabolites were normalized to an internal standard (PIPES) and auto-scaled. Ellipse indicates 95% confidence border based on Hotelling's T^2. Separation among different sampling points (different growth phases) can be seen along PC1, while the separation between control and disruptant strains was observable on PC2.

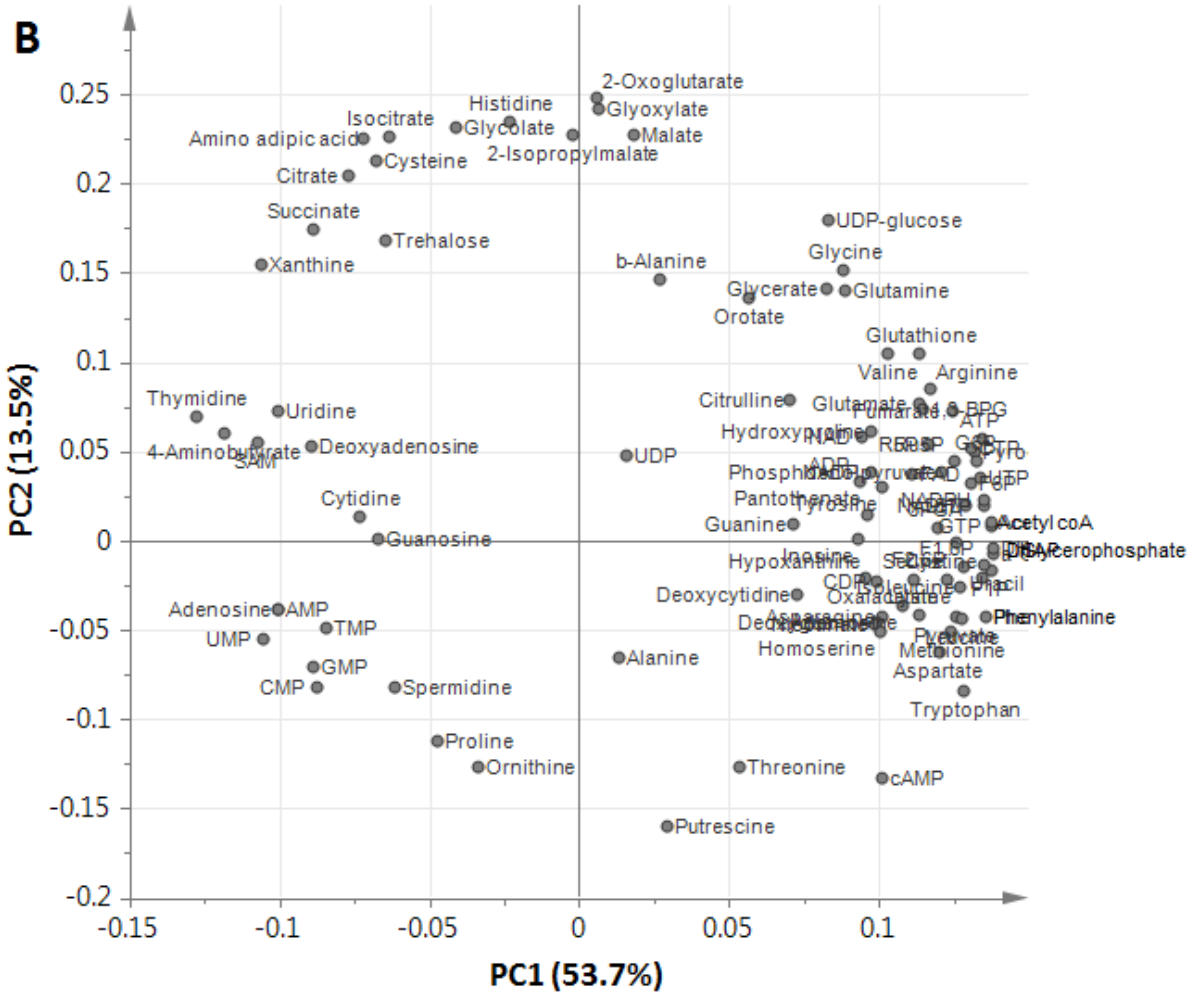


Fig. 2-2. (B) The corresponding loading plot illustrating metabolites that contribute to the separation on PC1 and PC2. Along PC1, nucleotide monophosphates and ribonucleosides were seen as major contributors to discrimination of samples at late growth phases (26 and 76 h), while proteinogenic amino acids except for proline and cysteine, and glycolysis intermediates were generally abundant in samples at early growth phases. Along PC2, increased level of 2-oxoglutarate and glyoxylate was distinctive in wild-type BY4742 at 76 h, while putrescine, cAMP, threonine and ornithine were high in RTG-deficient strains. For the complete list of loading values, see Supplementary Table S3.

2.3.2. Metabolites and metabolic pathways associated with RTG1 and RTG3

The purpose of PCA is to observe the separation pattern between wild-type and RTG-deletion strains. As the separation was successfully observed on the second principal component, the metabolites that showed a large loading value on PC2 were further analyzed. The loadings describe the multivariate makeup as a vector in multivariate space, and thus determine the underlying variables that are important to each PC. The 50 most important metabolites, with absolute loading values ≥ 0.05 are shown in **Fig. 2-3 (A)**. High levels of TCA and glyoxylate cycle intermediates (2-oxoglutarate, glyoxylate, malate, isocitrate, citrate, succinate) positively correlate with RTG-genes (increased in BY4742 and decreased when RTG-genes were deleted), while high levels of polyamine biosynthetic intermediates (putrescine, ornithine, spermidine) negatively correlate with RTG-genes (increased when RTG-genes were deleted).

To get the overall view of the contribution of these metabolites into different metabolic pathways, the 50 most influential metabolites were subjected into pathway enrichment analysis using MetaboAnalyst 2.0 [57]. The result is shown in **Fig. 2-3 (B)** and **Table 2-1**. Besides TCA and glyoxylate cycles, amino acid metabolism makes up the majority of the affected pathways. TOR activity is closely related to amino acid signaling, thus the result in part reflects the involvement of TOR in RTG pathway. This result also suggests that Rtg1 and Rtg3 may also hold regulatory effects on amino acid metabolisms other than glutamate and glutamine.

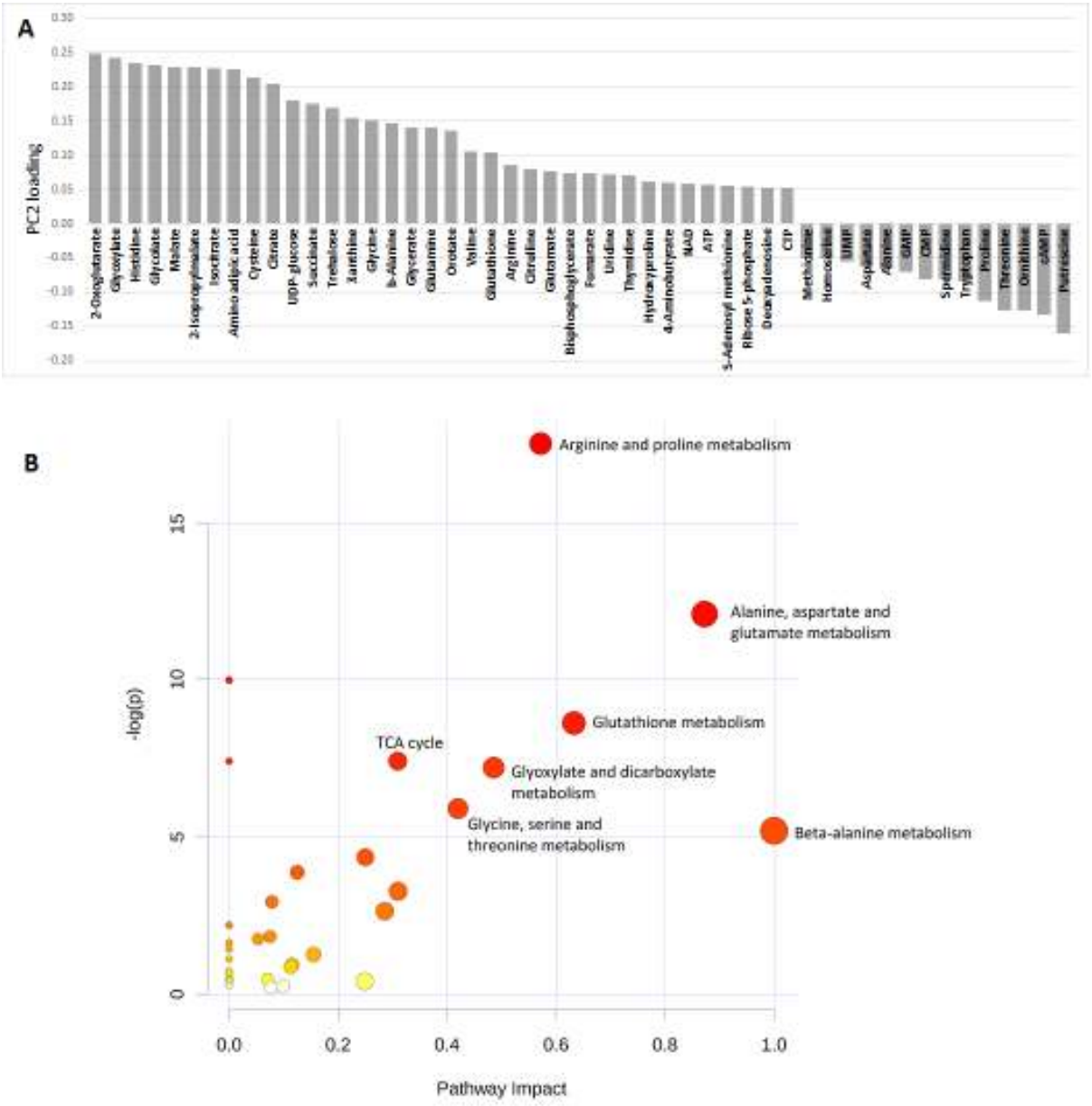


Fig. 2-3. (A) Loading values for 50 most influential metabolites along principal component 2, PC2 that distinguishes RTG-deficient strains from control BY4742. Positive loadings indicate a positive correlation with RTG regulation (decreased when the RTG regulatory gene was deleted), while negative loadings indicate a negative correlation with RTG regulation (increased when the RTG regulatory gene was deleted). **(B)** Overview of pathway analysis, showing matched pathways according to pathway enrichment analysis and pathway impact values from pathway topology analysis. Circles represent the metabolite-matched pathways of *S. cerevisiae* retrieved from Kyoto Encyclopedia of Genes and Genomes (KEGG). Color intensity indicates the significance of the pathway (the darker the more significant), while size indicates pathway impact score (the centrality of its involved metabolites).

Table 2-1. Pathway analysis using MetaboAnalyst 2.0 (hits ≥ 2, arranged according to *p*-values).

Pathway name	Total	Hits	p	-log (p)	FDR	Impact
Arginine and proline metabolism	37	13	2.37E-08	17.556	1.54E-06	0.57168
Alanine, aspartate and glutamate metabolism	20	8	5.69E-06	12.077	0.000185	0.87254
Aminoacyl-tRNA biosynthesis	67	13	4.65E-05	9.9767	0.001007	0
Glutathione metabolism	23	7	0.000182	8.6137	0.002951	0.63277
Citrate cycle (TCA cycle)	20	6	0.000608	7.4058	0.006591	0.30939
Nitrogen metabolism	8	4	0.000608	7.4047	0.006591	0
Glyoxylate and dicarboxylate metabolism	14	5	0.000747	7.1989	0.00694	0.48551
Glycine, serine and threonine metabolism	26	6	0.00275	5.8963	0.022342	0.41988
beta-Alanine metabolism	7	3	0.005565	5.1912	0.040194	1
Pyrimidine metabolism	35	6	0.012898	4.3507	0.083836	0.25014
Lysine biosynthesis	19	4	0.020822	3.8717	0.12304	0.125
Cysteine and methionine metabolism	33	5	0.037927	3.2721	0.20544	0.31009
Purine metabolism	60	7	0.053069	2.9362	0.26535	0.07859
Butanoate metabolism	17	3	0.071451	2.6387	0.33174	0.28571
Cyanoamino acid metabolism	10	2	0.11148	2.1939	0.48308	0
Valine, leucine and isoleucine biosynthesis	24	3	0.16013	1.8318	0.65052	0.07519
Sulfur metabolism	13	2	0.17308	1.754	0.66178	0.05319
Propanoate metabolism	14	2	0.19472	1.6362	0.70314	0
Pantothenate and CoA biosynthesis	16	2	0.23889	1.4318	0.81724	0
Starch and sucrose metabolism	18	2	0.28357	1.2603	0.92162	0.15497
Porphyrin and chlorophyll metabolism	20	2	0.32811	1.1144	1	0
Pyruvate metabolism	23	2	0.39349	0.93271	1	0.1159

Since the highest positive and negative loading was observed in 2-oxoglutarate and putrescine respectively, the regulatory effects of Rtg1/Rtg3 on TCA/glyoxylate cycle and superpathway of polyamine biosynthesis were further investigated. Time-course profiles of metabolites from TCA/glyoxylate cycle and polyamine biosynthesis are shown in **Fig. 2-4**. As expected, citrate levels were reduced significantly in RTG disruptant strains, consistent with previous studies that reported the transcriptional regulation of *CIT2* by the Rtg1/Rtg3 complex [42]. Other metabolic intermediates shared in TCA and glyoxylate cycles (2-oxoglutarate, isocitrate, glyoxylate, malate, succinate) were also decreased in the disruptants, especially during deceleration/post-diauxic and stationary phases (**Fig. 2-4 (A)**). Fumarate, which is exclusive to the TCA cycle, showed no significant difference. These observations can be explained according to the different growth phases. Initially, yeast cells were under a fermentative (glucose repressing) condition, during which the TCA cycle and mitochondrial biogenesis are repressed [63]. As glucose concentration decreases, the cells prepare for the reversion of metabolic fluxes; reducing glycolytic activity and increasing the flux thorough the glyoxylate cycle and gluconeogenesis [64]. Glucose exhaustion leads to a transient diauxic phase, which induces gene transcription for mitochondrial proteins and adaptation to respiratory metabolism [65]. The low levels of TCA/glyoxylate cycle intermediates in RTG deletion mutants after the post-diauxic phase thus reflect the inability of the cells to supply anaplerotic citrate from the glyoxylate cycle, since the expression of *CIT2* requires Rtg1/Rtg3. Interestingly, 2-oxoglutarate readily showed a significant decrease from the mid-exponential phase (OD_{600} = 1), in contrast to other TCA/glyoxylate cycle intermediates which only showed clear differences after the post-diauxic phase (**Fig. 2-4 (A)**).

Another interesting and unexpected observation was the elevated levels of polyamines putrescine and spermidine in *rtg1*Δ and *rtg3*Δ disruptant strains at stationary phase (the intensities of spermine from yeast extract were too low and could not be measured reliably, **Fig. 2-4 (B)**). Polyamine compounds have been associated with cytoprotective effects against oxidative and inflammatory stresses and its depletion has been linked to yeast aging and necrosis [66,67]. However, other stress response-related metabolites such as glutathione and trehalose showed an opposite trend (Loading plot **Fig. 2-2 (B)** and **Fig. 2-3 (A)**). It is possible that polyamines might serve as defense metabolites against stresses when RTG pathway is inactivated.

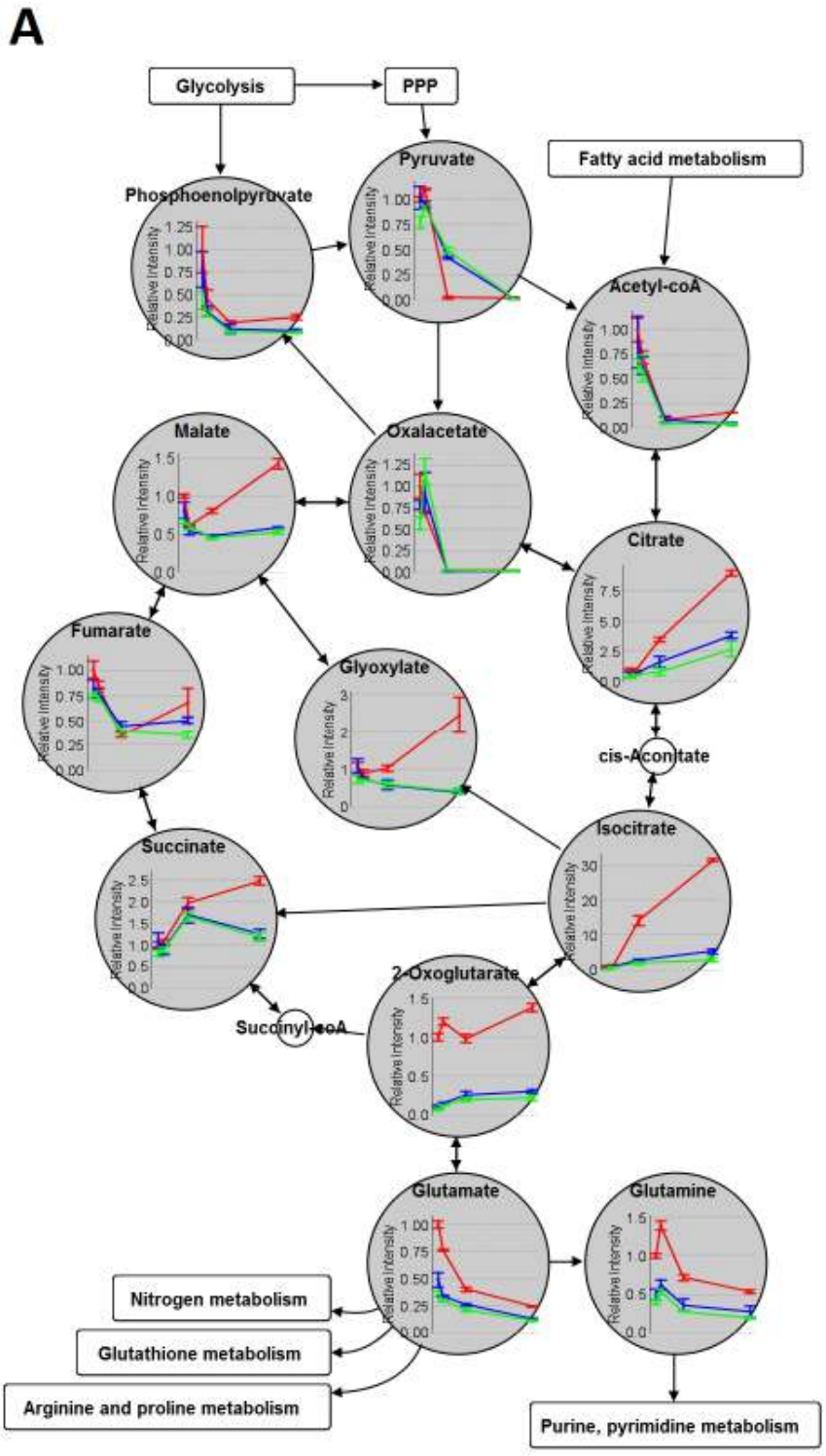


Fig. 2-4. *Legend in the following page.*

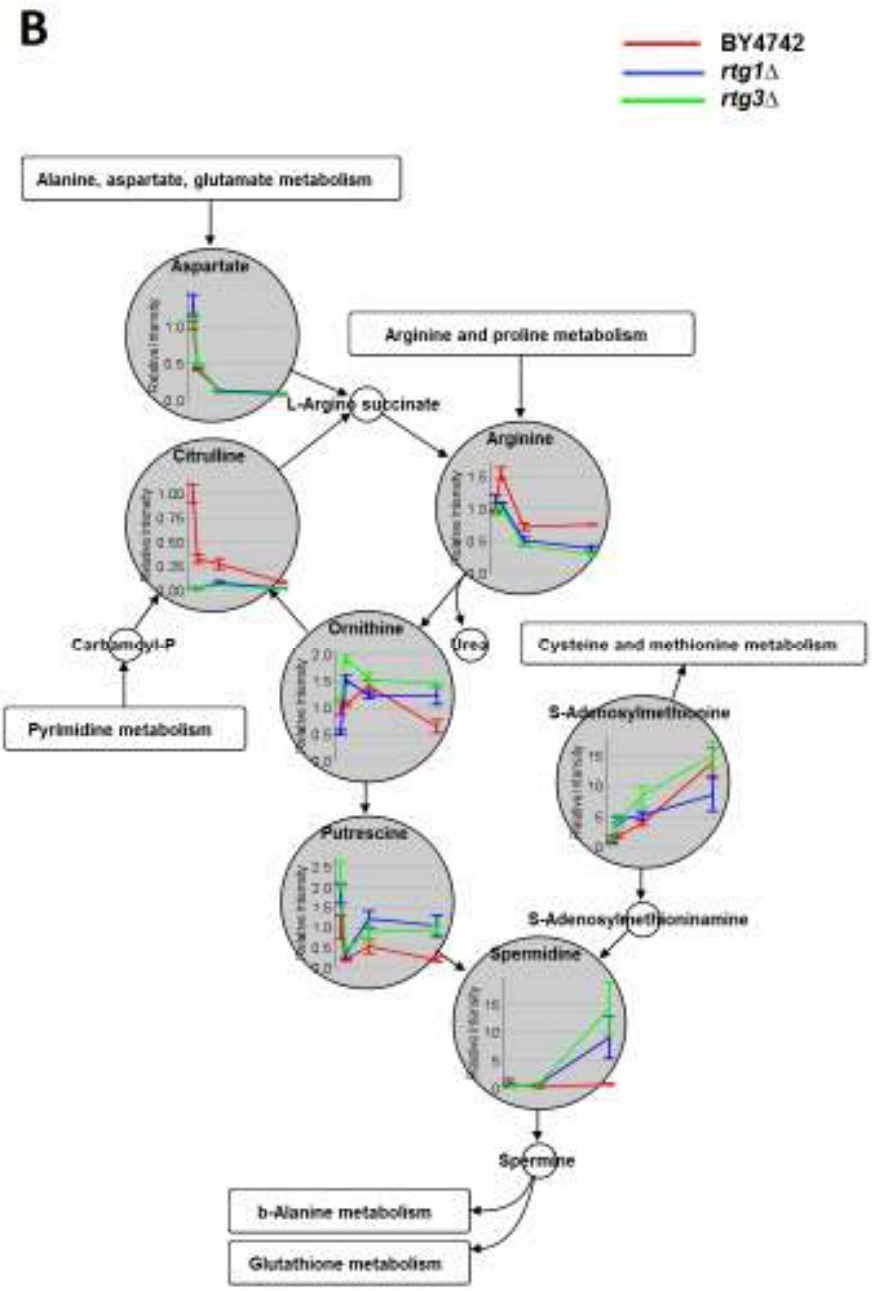


Fig. 2-4. Time-course metabolic profiles of wild-type BY4742, *rtg1*Δ disruptant and *rtg3*Δ disruptant (n=3) in **(A)** TCA/glyoxylate cycle and **(B)** superpathway of polyamine biosynthesis, shown together with the neighboring metabolic pathways (PPP: pentose phosphate pathway). Metabolite intensities were normalized to an internal standard and relative to those of control (BY4742) at time 5 h (OD_{600}=1). In *S. cerevisiae*, TCA cycle occurs in the mitochondria, while glyoxylate cycle in the peroxisome, however both are drawn combined in this figure since only bulk metabolites were measured. Note that in the event of fermentative metabolism and glyoxylate cycle activation, the flow from succinate to oxaloacetate is blocked (explaining the decreased levels of fumarate which cannot be supplemented through the anaplerotic pathway).

2.3.3. Metabolic alteration levels in rtg1Δ and rtg3Δ disruptants

Since Rtg1 and Rtg3 act in a heterodimer complex, and neither protein alone is able to bind to R-box sites [44], it is anticipated that the deletion of either gene would result in a similar metabolic alteration. **Fig. 2-5** depicts a heat map of metabolite changes of BY4742, *rtg1*Δ and *rtg3*Δ disruptants at four different culture time, corresponding to different growth phases. Both *rtg1*Δ and *rtg3*Δ disruptants displayed a strikingly similar metabolic alteration pattern. In addition, fold-change values were calculated and the statistical difference between the two deletion strains was compared (**Table 2-2)**. The fold-change values range from approximately -37 times (citrulline in *rtg3*Δ disruptant at 5 h) to 21 times (spermidine in *rtg3*Δ disruptants at 76 h). Only ornithine showed a significant difference between *rtg1*Δ and *rtg3*Δ disruptants at 5 h, while for the rest of the metabolites, *rtg1*Δ and *rtg3*Δ disruptants did not differ statistically across all time points. Therefore, it is concluded that the deletion of either *RTG1* or *RTG3* yields the same metabolic rearrangements, and the absence of either one component is sufficient for a shortfall of RTG response. However, for a majority of metabolites, *RTG3* appears to have more profound effects on metabolomics parameters (larger fold-change) upon deletion than *RTG1*. This result was reflected in the PCA score plot (**Fig. 2-2 (A)**) where *rtg1*Δ disruptant was positioned closer to BY4742 than *rtg3*Δ disruptant.

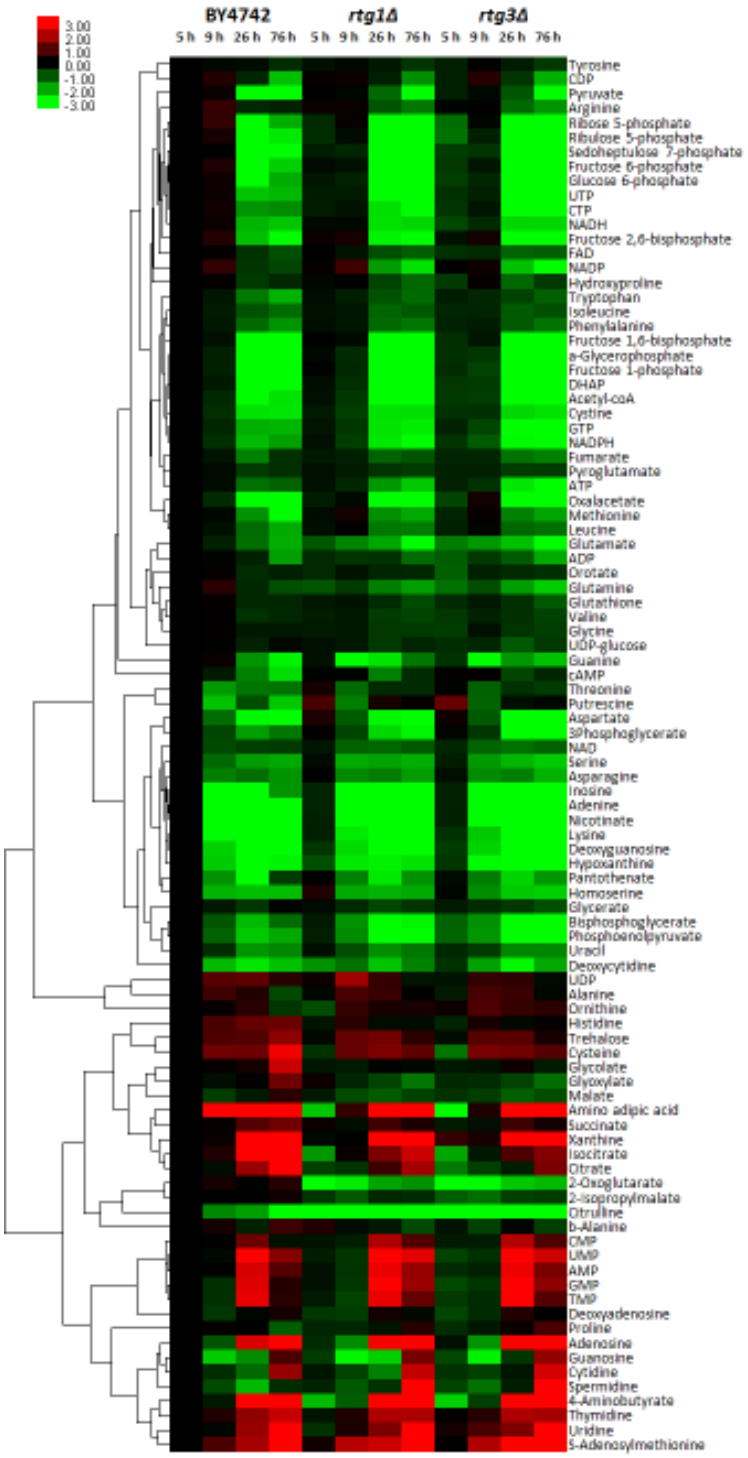


Fig. 2-5. Heat map showing the differential expression in BY4742, *rtg1*Δ and *rtg3*Δ disruptants at four different time points. Metabolite intensities were normalized to internal standard and relative to those of wild-type BY4742 at time 5 h (OD_{600}=1), averaged and $\log_2$ transformed. Metabolite clustering was based on Pearson's correlation and average linkage.

Table 2-2. Metabolite fold-change for 50 most important metabolites for *rtg1*Δ and *rtg3*Δ disruptants relative to wild-type strain BY4742 at each sampling time (-inversed in the case of down-regulation). Bold values indicate statistically significant difference between *rtg1*Δ and *rtg3*Δ disruptants ($p < 0.05$, determined by two-tailed heteroscedastic *t*-test and corrected for multiple testing (Benjamini & Hochberg false discovery rate, FDR)).

Metabolites	5 h			9 h			26 h			76 h		
	*rtg1*Δ	*rtg3*Δ	*p*	*rtg1*Δ	*rtg3*Δ	*p*	*rtg1*Δ	*rtg3*Δ	*p*	*rtg1*Δ	*rtg3*Δ	*p*
TCA/glyoxylate cycle												
2-Oxoglutarate	-9.26	-15.92	0.179	-8.32	-9.48	0.682	-3.79	-5.07	0.205	-4.54	-6.28	0.356
Malate	-1.23	-1.51	0.308	-1.14	-1.01	0.647	-1.69	-1.76	0.434	-2.45	-2.68	0.273
Isocitrate	-3.47	-3.89	0.679	-1.17	-1.55	0.721	-5.15	-7.50	0.200	-5.94	-10.86	0.264
Citrate	-1.69	-2.73	0.363	-1.37	-1.51	0.686	-2.11	-4.42	0.206	-2.35	-3.31	0.306
Succinate	1.09	-1.25	0.358	-1.20	-1.21	0.980	-1.16	-1.19	0.830	-1.97	-2.11	0.530
Fumarate	-1.23	-1.35	0.377	-1.10	-1.16	0.699	1.29	1.12	0.351	-1.34	-1.94	0.304
Glyoxylate	1.09	-1.39	0.352	-1.21	-1.26	0.620	-1.77	-1.75	0.981	-6.40	-6.00	0.792
Glycolate	-1.20	-1.22	0.885	-1.31	-1.23	0.665	-1.23	-1.06	0.236	-5.50	-6.52	0.376
Starch and sucrose metabolism												
UDP-glucose	-1.15	-1.46	0.377	-1.20	-1.34	0.719	-1.21	-1.56	0.332	-1.25	-1.50	0.330
Trehalose	-1.08	-1.11	0.680	1.17	1.07	0.537	1.29	1.04	0.303	-2.23	-2.70	0.362
Pyrimidine metabolism												
Orotate	-1.40	-2.21	0.341	-1.25	-1.39	0.668	1.05	1.01	0.622	-1.08	-1.15	0.733
Uridine	-1.48	1.18	0.385	1.31	1.96	0.659	-1.24	-1.26	0.987	-1.03	-1.05	0.907
Thymidine	-1.19	1.03	0.376	-1.02	1.03	0.954	1.26	1.26	0.982	-1.26	-1.32	0.802
CTP	-1.27	-1.55	0.363	-1.34	-1.36	0.953	-1.84	-3.12	0.214	-3.51	-4.69	0.341
UMP	-1.28	-1.72	0.398	-1.44	-1.37	0.819	1.26	1.60	0.333	2.12	1.78	0.336

CMP	-1.17	-1.23	0.867	-1.25	-1.25	0.999	1.71	1.77	0.831	2.25	2.16	0.808
Purine metabolism												
Xanthine	-1.08	1.58	0.374	-1.08	1.14	0.646	-1.24	-1.51	0.217	-1.92	-2.65	0.376
ATP	-1.14	-1.33	0.385	-1.24	-1.37	0.660	-1.38	-2.64	0.196	-2.31	-3.81	0.298
Deoxyadenosine	-1.57	-1.77	0.589	-1.04	1.10	0.675	1.24	1.36	0.325	-1.28	-1.17	0.434
GMP	-1.15	-1.82	0.390	-1.06	-1.05	0.938	1.41	1.95	0.232	2.64	2.34	0.522
cAMP	-1.02	1.03	0.864	1.20	1.29	0.787	1.01	1.63	0.287	4.27	4.43	0.800
Amino acid metabolism												
Histidine metabolism												
Histidine	-1.18	-1.34	0.339	-1.30	-1.50	0.586	-2.52	-2.46	0.851	-2.62	-2.24	0.264
Cysteine and methionine metabolism												
Cysteine	-1.50	-2.62	0.210	-1.11	1.07	0.570	1.17	1.03	0.505	-3.34	-3.58	0.800
Methionine	-1.11	-1.34	0.312	1.26	1.14	0.651	1.03	1.16	0.228	3.87	4.03	0.539
S-Adenosylmethionine	1.15	1.02	0.840	2.40	2.29	0.924	1.33	2.12	0.212	-1.62	1.08	0.352
Valine, leucine and isoleucine metabolism												
2-Isopropylmalate	-1.54	-2.13	0.377	-2.07	-2.48	0.735	-1.37	-1.66	0.380	-1.64	-1.87	0.434
Valine	-1.28	-1.58	0.235	-1.24	-1.36	0.795	-1.09	1.00	0.281	-1.26	-1.24	0.797
Lysine metabolism												
Amino adipic acid	-5.14	-11.67	0.369	-12.36	-14.42	0.644	-2.47	-3.68	0.190	-5.15	-7.05	0.339
Glycine, serine and threonine metabolism												
Glycine	-1.24	-1.64	0.354	-1.26	-1.18	0.812	-1.31	-1.25	0.835	-1.32	-1.36	0.588
Glycerate	-1.09	-1.23	0.547	-1.15	-1.19	0.936	-1.13	-1.01	0.340	-1.50	-1.85	0.288
Homoserine	1.31	-1.05	0.342	1.08	1.36	0.687	1.09	-1.10	0.323	1.17	-1.16	0.320

Threonine	1.18	-1.02	0.345	1.49	1.60	0.740	1.85	1.77	0.248	1.79	1.56	0.285
Beta-alanine metabolism												
b-Alanine	1.34	-1.17	0.384	-1.30	-2.01	0.673	1.07	1.26	0.836	-2.98	-2.72	0.798
Arginine and proline metabolism												
Arginine	1.09	-1.04	0.395	-1.41	-1.51	0.868	-1.43	-1.69	0.307	-2.00	-2.53	0.330
Citrulline	-35.03	-37.33	0.842	-12.89	-10.94	0.669	-3.39	-5.10	0.209	-3.48	-4.42	0.595
Hydroxyproline	-1.06	-1.61	0.090	-1.05	1.02	0.938	-1.08	-1.46	0.325	-1.69	-1.15	0.264
4-Aminobutyrate	-4.65	-5.55	0.689	-1.76	-1.37	0.678	-2.28	-1.48	0.232	1.14	-1.05	0.636
Proline	-1.34	-1.45	0.817	-1.42	-1.25	0.682	-1.17	1.08	0.215	2.91	3.78	0.302
Alanine, aspartate and glutamate metabolism												
Glutamine	-2.01	-2.53	0.306	-2.25	-2.62	0.718	-2.02	-2.61	0.328	-1.93	-2.84	0.340
Glutamate	-2.04	-2.76	0.369	-2.27	-2.63	0.189	-1.57	-1.93	0.237	-1.94	-2.19	0.266
Aspartate	1.27	1.13	0.387	1.13	1.16	0.678	1.01	-1.20	0.179	1.41	1.16	0.286
Alanine	-1.05	-1.23	0.355	1.08	1.07	0.989	1.21	1.22	0.967	1.86	1.71	0.286
Phenylalanine, tyrosine and tryptophan metabolism												
Tryptophan	-1.11	-1.26	0.336	-1.08	-1.14	0.714	1.38	1.56	0.218	1.79	1.97	0.275
Glutathione metabolism, polyamine biosynthesis												
Glutathione	-1.21	-1.43	0.344	-1.21	-1.27	0.640	-1.03	-1.07	0.272	-1.36	-1.51	0.332
Spermidine	-1.33	-1.74	0.674	-1.12	-1.37	0.675	3.01	3.56	0.641	13.24	20.61	0.383
Ornithine	**-1.87**	**1.13**	**0.002**	1.45	1.81	0.110	-1.15	1.09	0.251	1.90	2.21	0.334
Putrescine	1.83	2.32	0.343	1.75	2.13	0.723	2.28	1.68	0.324	5.33	4.85	0.732
Pentose phosphate pathway												
Ribose 5-phosphate	-1.41	-2.46	0.341	-1.41	-1.58	0.616	-1.98	-2.06	0.934	-2.67	-2.47	0.651

Glycolysis												
Bisphosphoglycerate	-1.23	-2.25	0.304	-1.64	-1.62	0.985	-1.92	-2.48	0.522	-3.12	-3.99	0.446
Others (co-factors)												
NAD	-1.20	-1.37	0.328	-1.23	-1.24	0.932	-1.39	-1.51	0.294	-1.29	-1.39	0.343

2.3.4. Yeast chronological lifespan and its relation with RTG1 and RTG3

As described previously, mitochondrial function closely relates to the aging process, and therefore, the association between RTG response with aging (lifespan) was further investigated. The effect of metabolic rearrangements following RTG response activation/deactivation towards aging can be evaluated in terms of yeast chronological lifespan. Activation of the mitochondrial RTG pathway has been reported to contribute to genome stability [68] and increase the yeast chronological lifespan, CLS [69]. In a separate study, decreased TOR signaling was also shown to extend CLS [70]. CLS is the period of time in which cells remain viable in a non-dividing state after nutrients cease in stationary phase [71], often expressed as the number of colonies recovered when the yeast cells are transferred back to growth allowing environment.

In this study, the CLS of *rtg1*Δ and *rtg3*Δ disruptants was measured (**Fig. 2-6**) and indeed, CLS was shortened in these strains. Although metabolic parameters showed that *RTG3* deletion imposed a greater effect than *RTG1*, CLS between *rtg1*Δ and *rtg3*Δ disruptants did not seem to differ. Together with metabolome data (**Table 2-2**), several observations can be made. Trehalose and glutathione, two metabolites that have been positively related to stress response, were accumulated in BY4742 at stationary phase. Meanwhile, amino acids such as histidine, glycine, glutamine, valine, arginine and glutamate were low, while methionine, aspartate, alanine, tryptophan, proline and threonine were higher at stationary phase in deletion strains. Analysis of growth media (PCA plots of extracellular metabolite from the growth media, Supplementary Fig. S1) revealed that extracellular threonine and valine were high in BY4742. While the addition of isoleucine, threonine, and valine to growth media was reported to extend CLS [72], the effects of intracellular amino acid level were not clarified.

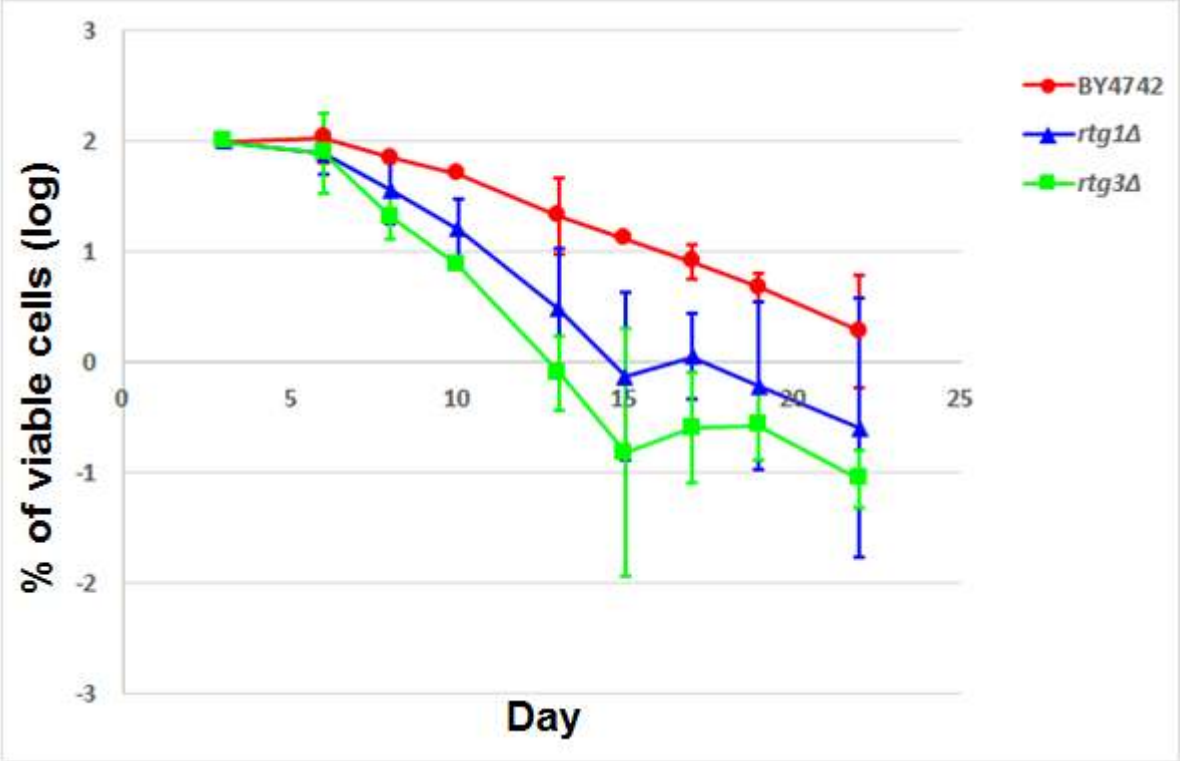


Fig. 2-6. The chronological lifespan (CLS) of BY4742, *rtg1*Δ and *rtg3*Δ disruptants, measured as the number (log %) of viable cells in exhausted growth media after revival on YPD (1% yeast extract, 2% peptone, 2% dextrose, 2% agar (% *w/v*)) plates. CLS at Day 3 (start of the stationary phase) was defined as 100%.

2.3.5. Comparison with previous literatures

The enzymatic and metabolic activities in *rtg1*Δ and *rtg2*Δ disruptants have been described previously by Small *et al.* (1995) [73]. The authors reported that when compared to levels in the parental strain, the only changes seen, aside from the absence of peroxisomal citrate synthase, were a reduction in mitochondrial citrate synthase activity (~30%–50%), a reduction in acetyl-coA synthetase activity (~50%), a reduction in cytosolic (NAD) isocitrate dehydrogenase activity (~50%) and a reduction in pyruvate carboxylase activity (~50%). These enzymes are encoded by the genes, *CIT2*, *CIT1*, *ACS1/ACS2*, *IDP2* and *PYC1/PYC2*, respectively. While the reduced citrate and 2-oxoglutarate levels observed in this study might be explained by the reduced activity of citrate synthases and isocitrate dehydrogenase, a difference in acetyl-coA and oxaloacetate levels between wild-type and RTG disruptants was not observed. However, the experiment by Small *et al.* (1995) [73] was conducted under a non-repressive condition, i.e., using raffinose as a carbon source. Moreover, they indicated that the disruptant cells have normally respiring mitochondria. In this study, mitochondrial function was confirmed to be intact in *rtg1*Δ and *rtg3*Δ disruptants, as there were no difference in the cell counts of these strains, when grown on YPD (1% yeast extract, 2% peptone, 2% dextrose, 2% agar (% *w/v*)) *vs.* YPG (1% yeast extract, 2% peptone, 2% glycerol, 2% agar (% *w/v*)) plates (Supplementary Table S4).

In a recent large-scale microarray transcript profiling by Kemmeren *et al.* (2014) [22], they indicated that there is a marked decrease in *CIT1*, *CIT2*, *ACO1*, *IDH1* and *IDH2* expressions in *rtg1*Δ and *rtg3*Δ disruptants. Similar to this experiment, they used BY4742 derivatives and grew the yeast strains in synthetic complete (SC) medium with 2% glucose and sampled the cultures at the mid-exponential phase. Therefore, the reduced levels of citrate, isocitrate and 2-oxoglutarate observed in this study might be attributed to the reduced expression of *CIT1/CIT2* (citrate synthases), *ACO1* (aconitase) and *IDH1/IDH2* (isocitrate dehydrogenases), respectively. Moreover, the use of a rich medium does not seem to overcome the lack of 2-oxogluratarate production in RTG disruptants. Kemmeren *et al.* (2014) [22] also showed that *GAP1* and *AGP1* were upregulated in *rtg1*Δ and *rtg3*Δ disruptants. Gap1 is a general amino acid permease that directs the uptake of all naturally occurring amino acids [74,75] and has been reported to be regulated by the nitrogen source and amino acid levels [76], while Agp1 is an amino acid permease, which transports asparagine and glutamine [77]. Chen and Kaiser (2002) [76] showed

that high levels of endogenous glutamine and glutamate induced by the deletion of Mks1 (a negative regulator of RTG pathway) caused extremely low Gap1 activity. In this study, while the endogenous glutamate level was significantly lower in RTG disruptants, the exogenous glutamate concentration in the medium for all strains did not differ (Supplementary Figure S2). However, reduced levels of both intracellular and extracellular glutamine in RTG disruptants can be seen (Supplementary Figure S2). Why increased Gap1 and Agp1 induced by *RTG1/RTG3* deletion seemed to result in an increased uptake of extracellular glutamine, but not glutamate, is not clear. Interestingly, two genes that encode glutamate dehydrogenases for the synthesis (*GDH1*) and degradation (*GDH2*) of glutamate were also increased in RTG disruptants. Gdh1 synthesizes glutamate from ammonia and 2-oxoglutarate, while Gdh2 degrades glutamate to ammonia and 2-oxoglutarate.Taken together, several explanations may underlie these observations; (1) there's a limit on the uptake level of glutamate when it is abundantly present in the growth medium; (2) glutamine is preferred over glutamate for an uptake into the cells; and (3) 2-oxoglutarate accumulation is primarily governed by *de novo* synthesis from isocitrate by *IDH1/IDH2* and not much from glutamate degradation by *GDH2*.

Moreover, the characteristic decrease in 2-oxogutarate concentrations in *rtg1*Δ and *rtg3*Δ disruptants preceding the decrease of other TCA cycle intermediates (citrate, isocitrate, succinate, malate) during the mid-exponential growth phase (**Fig. 2-4 (A)**) suggests that this metabolite might play a critical role in controlling the flow and balance of TCA/glyoxylate cycles. Further experiments, e.g., a flux analysis using labeled substrates, should be performed to confirm the origin of 2-oxoglutarate under sufficient glutamate/glutamine concentrations in the growth medium and to investigate the physiological attributions of this metabolite to the metabolic reprogramming under RTG deletion.

2.4. Conclusions

In this chapter, metabolic profiling of *rtg1*Δ and *rtg3*Δ disruptant strains was performed. By relative comparison of metabolic alteration in these deletion strains with wild-type BY4742, metabolites and metabolic pathways associated with *RTG1*/*RTG3* genes and possibly related to mitochondrial RTG response were identified. Besides TCA and glyoxylate cycles which have been identified previously, other pathways including amino acid metabolism were affected, and

thus supporting the idea of multi-regulatory coordination of bHLH proteins in different transcription programs. The new insights obtained from this chapter are the markedly reduced 2-oxoglutarate level which precedes other TCA cycle intermediates, suggesting a key role of 2-oxoglutarate in balancing TCA/glyoxylate cycle, as well as the elevated levels of polyamines at stationary phase in *rtg1*Δ and *rtg3*Δ disruptants. In addition, the relationship between RTG-gene deletion and chronological lifespan (CLS) was confirmed. This study illustrates the power of metabolomics in finding gene/transcription factor-metabolite correlations and provides a broader assessment of metabolic change following RTG-gene deletion. The outcome of this study is expected to lead to deeper investigations into RTG response and bHLH proteins in general.

Chapter 3

Global analysis of gene-metabolite correlations in154 transcription factor deletion strains

3.1. Introduction

In Chapter 2, the utility of metabolomics approach in TF related study using a representative TF complex was demonstrated. In this chapter, I proceeded to perform a global analysis of metabolome that covers 154 TF-gene deletion strains.

As described previously, research in transcriptional regulation involves, among others, the determination of DNA-binding domain (DBD) motifs and protein-protein interaction, identification of downstream effector genes, quantification of transcript and protein abundances, as well as network construction from genome-wide expression data using computational methods [16,21,78–80]. Despite the huge amount of research, the understanding of global gene regulation by transcription factors is not yet complete; in yeast, for about half of the apparent sequence-specific DNA-binding TFs, physiological functions and/or DNA-binding sites remain unknown [17,81,82]. There are also inductions or repressions of pathways that do not seem to be the direct target of the TF, which are probably due to transcriptional cascades. For example, Hms1, appears to positively regulate genes involved in several diverse pathways, including several that have dedicated TFs, and some genes that do not appear to contain Hms1-binding sites in their promoters [82]. Moreover, while TFs essentially bind to DNA promoter regions to initiate their action, transcriptional regulation is not a simple binary on/off control. The number of TF molecules also plays a part in determining the level of transcription [83]. In addition, one TF may be involved in various genes and requires a precise set of protein complex and co-activators before transcription can be initiated. Furthermore, many regulatory events that link triggering cues to final phenotypic reprograming remain poorly characterized, making it difficult to predict cellular behavior even when the transcriptional machinery is known. The missing link between the change in transcript or protein levels and phenotype (e.g., growth rate,

chemical resistance, production of secondary metabolites) has to be investigated. Clearly, an alternative approach to studying transcription factors besides gene expression profiling or DNA and protein-protein binding is necessary in order to gain an overall picture of gene regulatory mechanism. In particular, how perturbation in transcription factors affects metabolite levels and ultimately cellular function needs to be addressed.

So far, studies regarding global transcriptional regulation and/or network using metabolomics approach have been limited. Analysis of condition-dependent TF network using metabolic flux distribution of 119 yeast TF deletion stains was reported previously [84], which reveals that metabolic flux alteration caused by TF deletion occurs almost exclusively at TCA cycle, and only 23 strains exhibited differential flux ratio change. Amino acid profiling for ~5000 yeast single gene deletion strains was also conducted [85], which shows that clustering of functionally related genes can be found for arginine biosynthesis and urea cycle pathways but not other pathways. However, there are no comprehensive reports on metabolites other than amino acids nor a dedicated metabolite profiling for transcription factor deletion strains. Although Yeast Metabolome Database (YMDB) [86] serves as a database that lists all the metabolites contained in yeast, there are no reports on metabolite levels of specific strain types.

Metabolomics has been regarded as a high resolution approach, due to its ability to capture subtle change in metabolite levels, which often does not manifest in a change in phenotype until at a much later stage. For example, Yoshida and colleagues demonstrated that the difference in yeast replicative lifespan can be detected from metabolic fingerprints of exponentially growing yeast cells (~4 doublings) whereas a conventional method by counting the number of daughter cells from a single mother cell requires at least 20 generations before a comparison can be made. This finding shows that metabolomics is a powerful tool to uncover a complicated phenotype (i.e., lifespan) at earlier stage compared to conventional method. Metabolomics has also been used to reveal silent genes, i.e., genes that produce no overt phenotype when deleted from the genome, but have distinct metabolic concentrations [87,88]. Furthermore, the use of single gene knockouts of yeast *S. cerevisiae* has been proven to be useful in functional genomics studies, using transcriptomics and metabolomics approaches [16,84,85,89]. Therefore, metabolomics serves as an excellent tool to study metabolic phenotype of non-essential gene knockouts of transcription factors due to; 1) unlike enzymes, TFs have no one-to-one relation with metabolic pathways, and thus hold much more complex regulatory network, and 2) most of these genes

are purportedly silent, i.e., the gene deletion does not affect the growth rate of the organism, leaving metabolic change as the sole indicator of strain condition.

In this chapter, the commercially available yeast single gene deletion library was used, and metabolite profiling of 154 strains each lacking a non-essential gene putatively encoding transcription factor was performed. The strains were then characterized according to their metabolic profiles. Core metabolites and co-factors deriving from central metabolic pathways such as glycolysis and TCA cycle, as well as amino acids and organic acids that are commonly conserved in most organisms were identified. Metabolome dataset can serve as invaluable inputs to assist researchers working on transcription factors and yeast biology in general.

3.2. Experimental section

3.2.1. Strain and culture condition

All strains used in this study were single gene knockouts from the European Saccharomyces Cerevisiae Archive for Functional Analysis (EUROSCARF) [90] collection, with BY4742 (*MATα leu2Δ0 lys2Δ0 ura3Δ0 his3Δ1*) as the parental/wild-type strain. The knockout strains were purchased from Open Biosystems (Huntsville, AL, USA) which were constructed by replacing the target genes with *kanMX* cassette that confers resistance to geneticin [91]. Each experiment was conducted with at least three replicates, and BY4742 was used as a control.

Yeast cells from frozen glycerol stock were plated onto YPD agar plates (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose, 20 g/L agar (all from BD, MD, USA except glucose and agar from Nacalai Tesque), with added geneticin G418 (Wako, Osaka, Japan) 200 µg/mL for knockout strains) and grown at 30 °C for 2 days. After two days, a single colony was obtained and re-streaked on a new YPD plate to increase cell number. This plate was used as a master plate for all cultivations. Cultivation was performed in three steps to reduce variation in cell growth; pre-pre culture, pre-culture and main culture. All liquid cultures were grown in synthetic complete (SC) medium [53]. SC was chosen as the growth medium to enable all deletion strains to grow comparably without severe growth defect, while still allowing controlled and known nutrient composition [85]. Preliminary experiment revealed that some of the knockouts were auxotroph for certain amino acids. Moreover, minimal medium such as synthetic defined

(SD) medium causes severe growth delay in some strains, thus the use of SD medium was not feasible. SC medium was prepared as follows: 20 g/L glucose, 6.7 g/L Difco™ yeast nitrogen base without amino acids (BD, MD, USA), 1.92 g/L yeast synthetic drop-out media supplement without uracil (Sigma-Aldrich, MO, USA), 76 mg/L uracil (Sigma-Aldrich), with geneticin G418 added to a final concentration of 200 µg/mL for knockout strains.

For pre-pre culture, cells were inoculated from the master plate into 3 mL of culture medium, followed by incubation at 30 °C in a rotary shaker (200 rpm) for 18 h. Next, for pre-culture, a portion of pre-pre-culture broth was diluted into 15 mL fresh culture medium so that the starting optical density at 600 nm (OD_{600}) = 0.01, and incubation with shaking was continued for 18 h. For main culture, the pre-culture broth was diluted in 15 mL culture medium so that the starting OD_{600} = 0.1, and incubation with shaking was continued until desired optical density values are reached. An iMark microplate reader (Bio-Rad, CA, USA) was used to monitor optical density.

For metabolome sampling, a fast filtration method based on Crutchfield *et al.* (2010) [92] was applied with slight modifications. 5 mL of culture broth at OD_{600} = 1 (approximately equivalent to 5 x 10^7 cells) were rapidly filtered using a 0.45 µm-pore size, 25 mm-diameter nylon membrane (Millipore, MA, USA) under a vacuum filtration. The filter membrane was folded and inserted into a 2-mL sampling tube filled with 1 mL single-phase extraction solvent (methanol/chloroform/water = 5/2/2 *v/v/v* %, with 1.2 µg/mL each of 1,4-piperazine diethanesulfonic acid (PIPES) and ribitol as internal standards [62]) pre-cooled at -30 °C, after which the tube was flash frozen in liquid nitrogen and stored at -80 °C until extraction.

3.2.2. Metabolite extraction and sample preparation

For extraction, the tubes filled with membrane-bound cells and extraction solvent were placed in a thermomixer (Eppendorf, Hamburg, Germany) at 4 °C, 1200 rpm for 30 min. After that, all liquid extract (900 µL) was transferred to a new tube filled with 400 µL water, vortexed and centrifuged at 4 °C, 16100 rcf (relative centrifugal force) for 3 min to separate polar and non-polar phases. Next, the upper polar phase was transferred to a new tube via syringe filtration (0.2 nm PTFE hydrophilic membrane, Millipore, MA, USA) and divided into two, 300 µL for LC-MS and 600 µL for GC-MS. The extracts were concentrated five times from the

initial volume under a vacuum centrifugation system (VC-96R, Taitec, Japan), transferred to glass vials (Chromacol, Hertfordshire, UK) and ready for UHPLC-MS/MS analysis. Samples were analyzed within 24 h after extraction.

For GC/MS samples, concentrated extracts were lyophilized overnight, followed by derivatization by oximation and silylation [93]. The oximation reagent, methoxyamine hydrochloride (Sigma-Aldrich, MO, USA) was first dissolved in pyridine (Wako, Osaka, Japan) to a concentration of 20 mg/mL and 75 µL added to each sample tube containing the lyophilized extracts. After reaction at 30 °C, 1200 rpm for 90 min, 50 µL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) (GL Sciences, Tokyo, Japan) was added and the silylation reaction was performed at 37 °C, 1200 rpm for 30 min. The derivatized samples were transferred to glass vials (Chromacol, Hertfordshire, UK) and analyzed within 24 h.

3.2.3. LC/MS analysis

The analysis platform consists of Shimadzu Nexera series UHPLC system (Shimadzu, Kyoto, Japan) coupled to a triple quadrupole mass spectrometer, LCMS-8030 with modification to improve sensitivity (Shimadzu, Kyoto, Japan). The analytical conditions were as follows; column: L-Column2 ODS (150 mm x 2.1 mm, 3 µm, Chemicals Evaluation and Research Institute Japan); flow rate: 0.3 mL/min; column temperature: 35 °C; mobile phase A: water containing 10 mM tributylamine and 15 mM acetic acid; mobile phase B: methanol; gradient program: 0% B (0-0.5 min) - 25% B (7.5 min) - 90% B (11-11.5 min) - 0% B (11.6-15 min); sample cycle time: 15 min; injection volume: 3 µL. The mass spectrometric parameters were: ESI negative mode; desolvation line (DL) temperature: 250 °C; nebulizer gas flow: 2 L/min; heat block temperature: 400 °C; other parameters were optimized automatically by flow injection analysis and auto-tuning. The MS/MS fragment for each analyte was determined using authentic standards. Multiple reaction monitoring (MRM) transition parameters and retention time for each metabolite are listed in Supplementary Table S1. All samples were kept in a 4 °C autosampler during analysis. All reagents were of LC-MS grades (Wako, Osaka, Japan).

3.2.4. GC/MS analysis

GC/MS was performed on a GCMS-QP2010 Ultra (Shimadzu, Kyoto, Japan) gas chromatograph coupled with a quadrupole mass spectrometer equipped with an AOC-20i/s autoinjector (Shimadzu, Kyoto, Japan). A CP-SIL 8 CB Low Bleed/MS column (Varian, CA, USA) 30 m × 0.25 mm (0.25 µm) was used for the GC separation. The mass spectrometer was auto-tuned and calibrated prior to analysis. 1 µL of sample was injected in split mode with a split ratio of 1:25. The inlet temperature was set at 230 °C and the column flow rate was 1.12 mL/min (linear velocity 39 cm/s). The column temperature was held at 80 °C for 2 min, raised by 15 °C/min to 330 °C, and held at 330 °C for 6 min. The transfer line and ion source temperatures were 250 °C and 200 °C respectively. Electron ionization (EI) was performed at 70 eV. The mass range of the detector was set to *m/z* 85 to 500 and the detector voltage (set by auto-tuning) was 0.93 kV.

An alkane standard mix was prepared from 25 µL each of C8-C20 and C21-C40 alkane standard solutions, diluted with an addition of 25 µL pyridine and injected at the start of each analytical run for calculating retention indices. In addition, a blank pyridine sample was injected every 8 samples for diagnostic purposes (to check for column bleed and carryover).

3.2.5. Metabolite identification and validation procedure

The stability and reproducibility of the method were evaluated using pooled quality control (QC) samples [55,56,94]. QC samples were prepared by pooling an equal volume of yeast extracts from each sample within the same analytical batch. The same QC aliquot from one injection vial was used for each analytical batch, injected at least three times at the start of the analytical batch after the system has stabilized, then at every sixth injection throughout the entire analytical workflow. Additionally a standard mixture of 1 µM was routinely injected at the beginning, middle and end of each analysis run.

3.2.6. Dataset construction

For LC/MS, peak picking was conducted by LabSolutions (Shimadzu, Kyoto, Japan) followed by manual inspection. The parameters were set as follows: integration: auto, max peak: 3, width: 5 sec; smoothing: standard, counts: 5, width: 1 sec; identification: absolute RT & closest peak, target window: 5%, reference window: 5%, process time: ± 1 min. Obtained peaks were identified as metabolites contained in yeast extracts by matching the extracted ion chromatograms with in-house metabolite library (MS/MS fragment and retention time). The identity was checked by spiking authentic standards to yeast extract and confirming that the particular metabolite peak intensity increases with an added concentration. To correct for matrix effect commonly observed in ESI-based LC/MS, the raw peaks were calibrated using the external calibration method. Initially, primary stock solutions from authentic standards were prepared in water at a concentration of 10-100 mM for each metabolite, from which standard mixtures of various concentrations were made. Standard mixtures were spiked into yeast extracts and used for making calibration curves. For analysis from different batches, the calibrated peaks were then multiplied by a correction factor (peak intensity of a standard mixture during calibration / peak intensity of a standard mixture during actual run), before integrated into one dataset.

For GC/MS, raw data files were converted into netCDF (*.cdf) format according to the ANDI (Analytical Data Interchange Protocol) specification using the proprietary software GCMSsolution (Shimadzu, Kyoto, Japan) before peak detection, baseline correction and retention time alignment using the freely available data processing tool MetAlign [95]. Data matrices from the alignment were then imported into AIoutput2 ver.1.29 [96] for an automated retention indices (RI)-based target compound identification and quantification.

3.2.7. Multivariate data analysis

Principal component analysis (PCA) was performed using SIMCA-P+ ver13 (Umetrics, Umeå, Sweden). Heat map and hierarchical clustering of fold-change normalized intensities were performed on Cluster 3.0 [58] and viewed on Java Treeview [59]. Statistical difference (two-tailed heteroscedastic *t*-test) was calculated using MS Excel.

3.3. Results and discussion

3.3.1. Metabolites identification and quantification

In this study, due to the large number of strains, cultivation and analysis steps were performed in four batches and included wild-type strain BY4742 as control in each batch. Supplementary Table S5 lists all the strains used in this study, the maximum specific growth rate, and the adenylate energy charge (EC). EC indicates the energy status of the cells, where exponentially growing cells have an EC of ~0.8 while an EC of <0.5 is indicative of dead cells [97]. Growth rates serve as a general measure of the effects of gene deletion. Wild-type strain BY4742 grew at a maximum specific growth rate of 0.45-0.48 h^{-1}. All deletion strains grew well in SC medium except for six knockout strains (*ino2*Δ disruptant, *ino4*Δ disruptant, *opi1*Δ disruptant, *gcr2*Δ disruptant, *aft1*Δ disruptant, *ada2*Δ disruptant) that exhibited a growth defect of > 20% compared to wild-type.

In microbial metabolome experiments, it is important to ensure that metabolites are rapidly quenched at the time of sampling. A fast filtration method followed by subsequently dipping the cells into cold extraction solvent [92] was used, which usually takes ~30 s from taking out samples from liquid culture to quenching. Adenylate energy charge, EC, calculated as ([ATP] + 0.5[ADP])/([AMP] + [ADP] + [ATP]), was in the range of 0.72 to 0.88 (Supplementary Table S5), which is typical of exponentially growing cells [28,97,98], suggesting that quenching was sufficient.

A total of 84 metabolites were successfully identified and quantified from LC/MS and GC/MS (**Table 3-1**). Similar to the previous chapter, selection of metabolites was based on < 30% of RSD of QC samples [55]. The median RSD for Batch 1-4 were 6.6%, 8.1%, 10.9%, 9.5%, respectively.

Table 3-1. List of metabolites obtained from the metabolic profiling of 154 TF-deletion strains.

	Amino acids	Bases, nucleotides	Sugar and derivatives	Organic acids	Others, co-factors
LC-MS (67)	Arginine Histidine Serine Asparagine Glutamine Homoserine Threonine Trehalose Proline Valine Methionine Isoleucine Tyrosine Amino adipic acid Glutamate Aspartate Phenylalanine Pyroglutamate Tryptophan Glutathione	Guanine Inosine Guanosine CMP UMP GMP AMP cAMP CDP GDP ADP CTP GTP UTP ATP	G6P R5P S7P F6P DHAP GAP Ru5P F1P UDP-Glu F2,6P F1,6P 1,3-BPG α-Glycerophosphate (Glycerol 3P)	Nicotinate Pantothenate Succinate Fumarate Oxalacetate Malate 2-Oxoglutarate Isocitrate Citrate PEP 2-Isopropylmalate Orotate Pyruvate	Acetyl CoA NAD NADP FMN FAD NADPH
GC-MS (17)	Alanine Glycine Leucine Lysine Cysteine+Cystine Citrulline	Uracil Adenine	Inositol Glycerol Glucose β-Lactose Melibiose		2-Aminoethanol Urea Phosphate Octadecanoate (Stearic acid)

Abbreviations: G6P: glucose 6-phosphate; R5P: ribose 5-phosphate; S7P: sedoheptulose 7-phosphate; F6P: fructose 6-phosphate; DHAP: dihydroxyacetone phosphate; GAP: glyceraldehyde 3-phosphate; Ru5P: ribulose 5-phosphate; F1P: fructose 1-phosphate; UDP-Glu: uridine diphosphate-glucose; F2,6P: fructose 2,6-bisphosphate; F1,6P: fructose 1,6-bisphosphate; 1,3-BPG: 1,3 bisphosphoglycerate; PEP: phosphoenolpyruvate; NAD: nicotinamide adenine dinucleotide; NADP: nicotinamide adenine dinucleotide phosphate; FMN: flavin mononucleotide; FAD: flavin adenine dinucleotide; NADPH: reduced nicotinamide adenine dinucleotide phosphate; ATP: adenosine 5'-triphosphate; ADP: adenosine 5'-diphosphate; AMP: adenosine 5'-monophosphate; GTP: guanosine 5'-triphosphate; GDP: guanosine 5'-diphosphate; GMP: guanosine 5'-monophosphate; CTP: cytidine 5'-triphosphate; CDP: cytidine 5'-diphosphate; UMP: uridine 5'-monophosphate; cAMP: adenosine 3',5'-cyclic monophosphate.

3.3.2. Validation of analytical performance and data normalization

Medium- to large- scale metabolomics studies often suffer from batch-to-batch reproducibility problem. To demonstrate the degree of batch-to-batch variation, first only wild-type BY4742 strains were plotted in PCA (**Fig. 3-1**). It was found that normalization to an internal standard (calculated as the ratio of the peak intensity of each metabolite to the peak intensity of the internal standard) was not sufficient for eliminating batch-to-batch variation. The same wild-type strain (BY4742) was used in each batch, thus technically the wild-type samples should all be clustered together on PCA. However, from **Fig. 3-1**, while Batch 1 and Batch 3 were clustered together, Batch 2 and 4 were separated, showing that batch-to-batch variation is inevitable.

To further examine this problem, QC samples were used as a benchmark. The use of QC samples from pooled test extracts to monitor analytical performance has been demonstrated in metabolome studies for urine [56,99] and plasma or serum [100]. QC samples prepared from aliquots of test samples provide 'mean' representative of all the metabolites contained and thus considered appropriate for the evaluation of reproducibility and sample stability. It was observed that QC samples were clustered together on the PCA plot (**Fig. 3-2 (A)**), demonstrating that the analysis platform is sufficiently stable throughout the run.

However, in terms of between-batch reproducibility, a clear separation between different batches can be seen (**Fig. 3-2 (B) and Fig. 3-3 (A)**). While ion pairing LC-MS has the advantage of wide coverage of metabolites of various species, including polar metabolites from central metabolism, with relatively stable retention time [29,31,101], it lacks reproducibility of different batch analysis. Peak intensities tend to deteriorate over time, while there is a need for regular cleaning due to accumulation of residual ion pairing reagent in the analysis line. As reported previously, day-to-day analytical variation was inevitable, accounting for the major portion of data variability [55]. Therefore, integration of data from different analytical runs and different batches needs a thorough consideration and a proper normalization method. It is important to minimize these differences so that the true interpretation of biological phenomena can be derived, which reflects strain differences instead of batch differences.

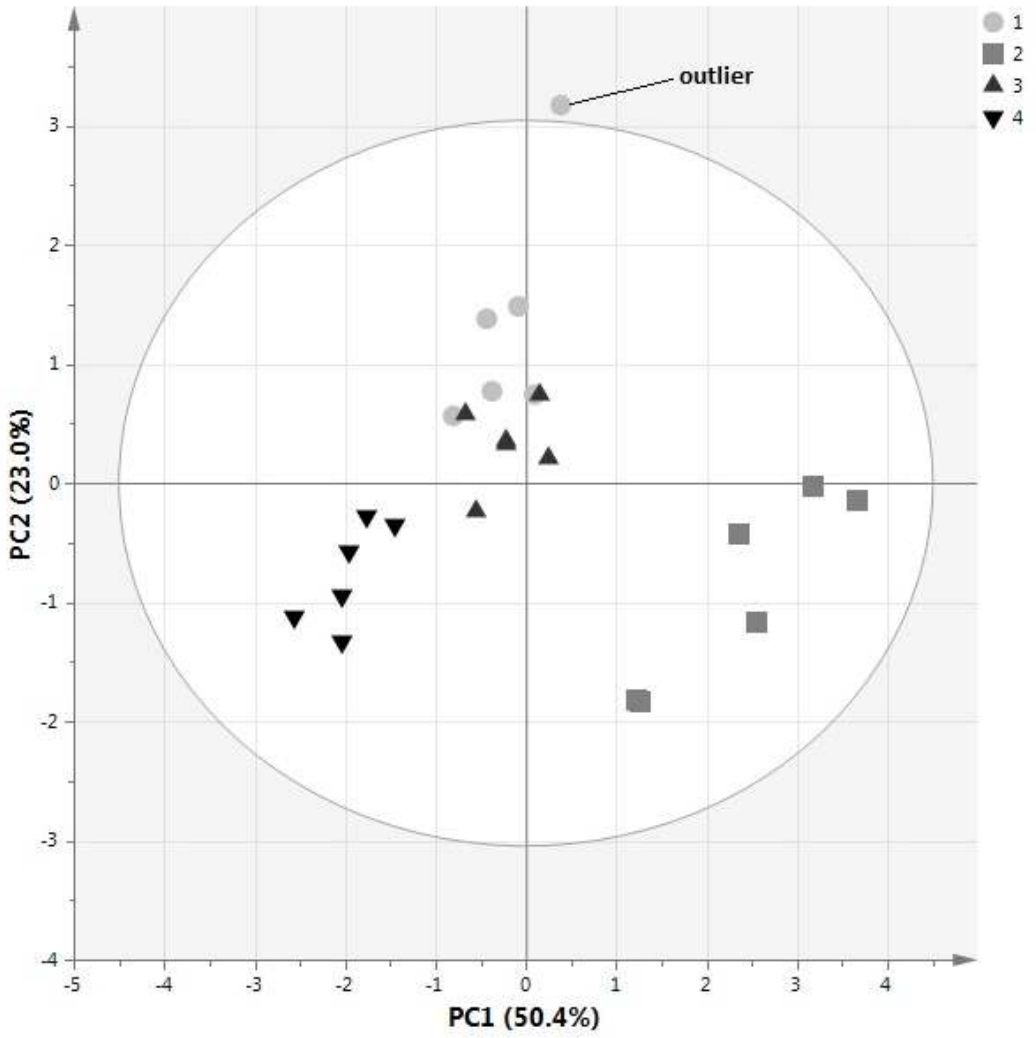


Fig. 3-1. Batch-to-batch variation observed in wild-type samples BY4742. Data was normalized to internal standard and Pareto-scaled (mean-centered and divided by the square root of standard deviation). Numbers 1-4 indicate batch number. Ellipse indicates 95% confidence border based on Hotelling's T^2 statistics. One control sample was identified as an outlier and removed from the dataset.

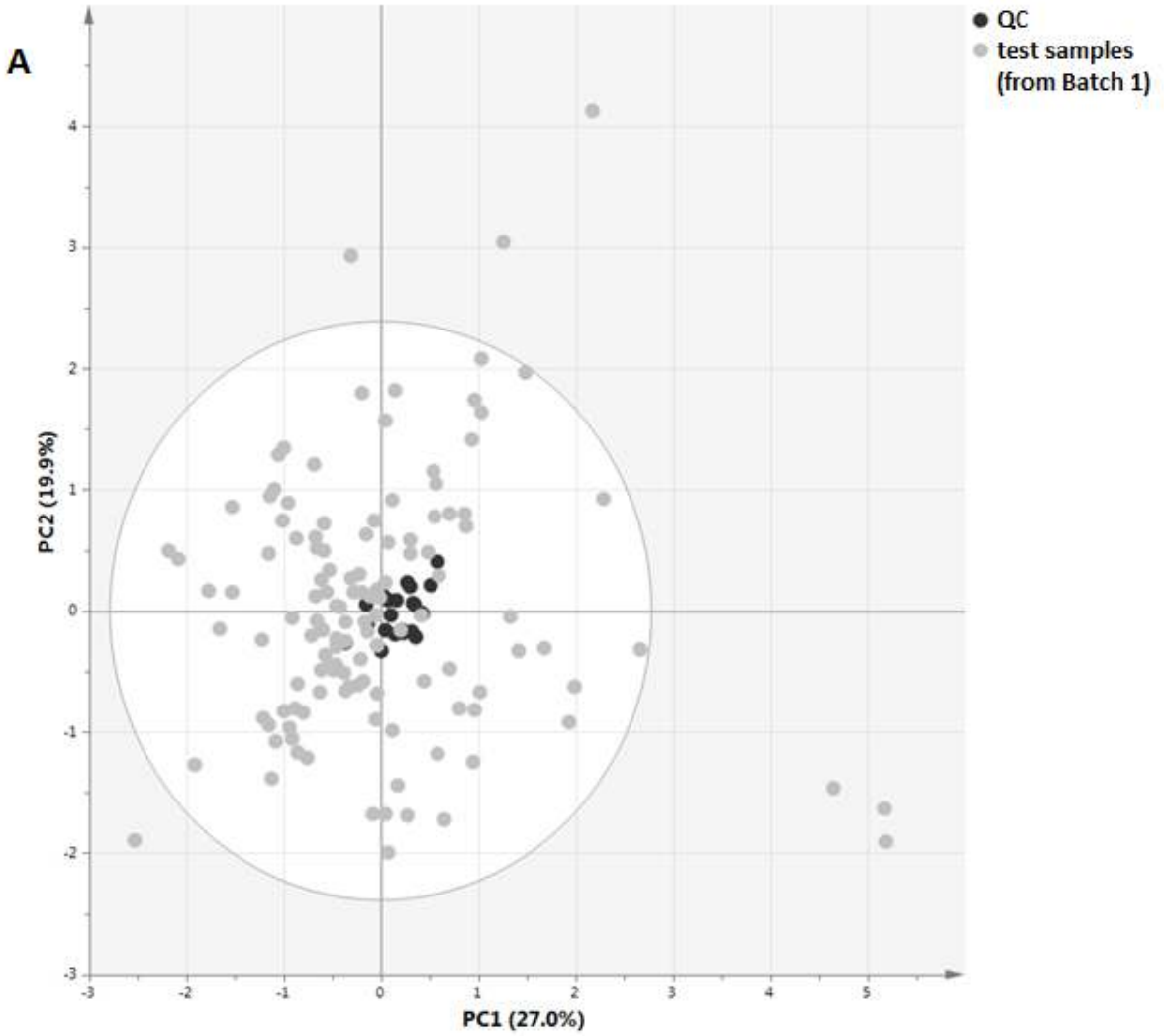


Fig. 3-2. (A) Stability and reproducibility of the analysis method **within the same analytical run**. The peaks were normalized to internal standard and Pareto-scaled. QC samples injected periodically were clustered together, showing that the method is stable and reproducible within the same analysis run. Data were taken from Batch 1 from ion pairing-LC-MS/MS data. Ellipse indicates 95% confidence border based on Hotelling's T^2 statistics.

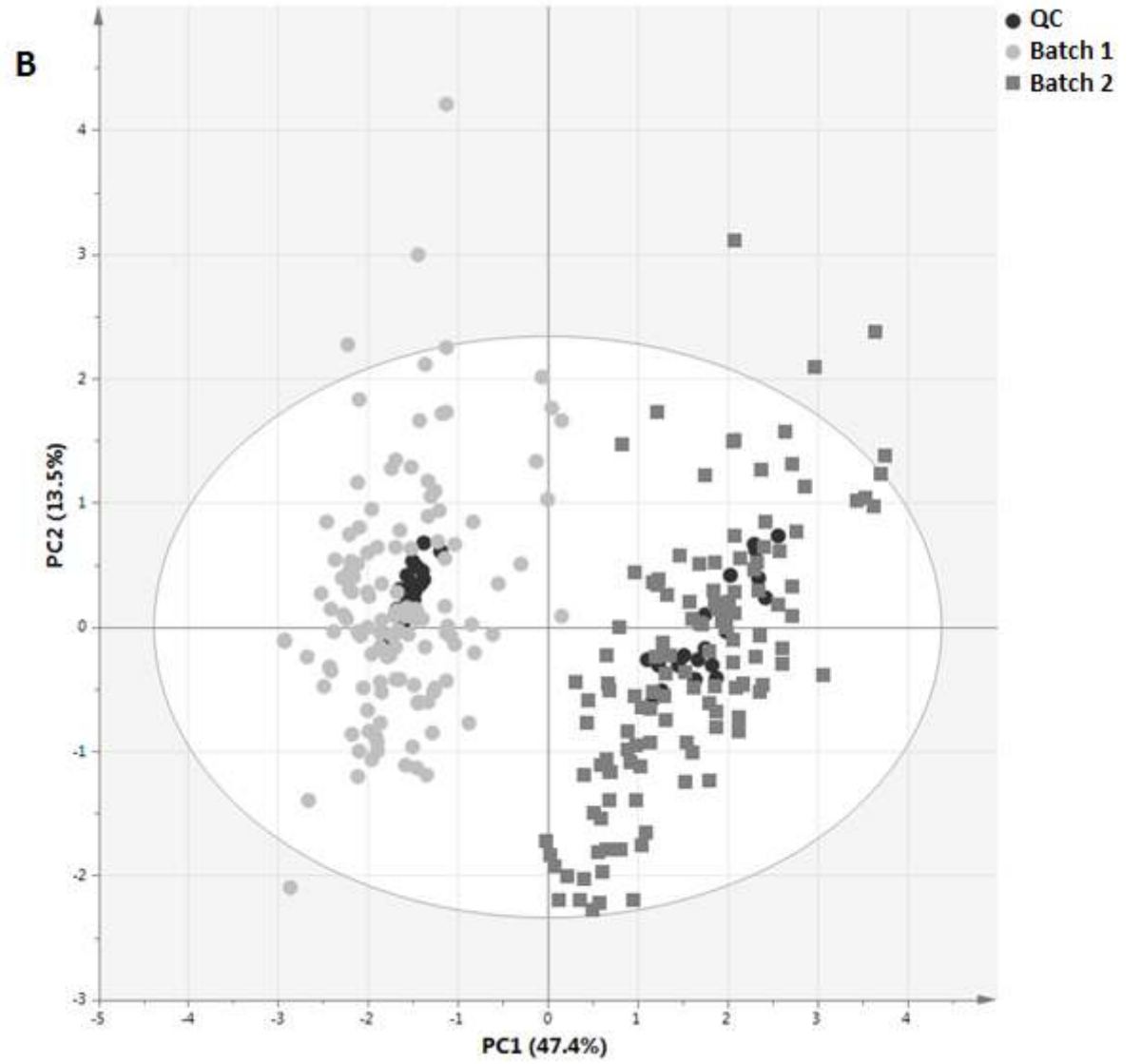


Fig. 3-2. (B) Stability and reproducibility of the analysis method **between two separate analytical runs**. The peaks were normalized to internal standard and Pareto-scaled. When two separate runs were combined, batch separation can be observed along PC1, indicating that an alternative normalization procedure is necessary. Data were taken from Batch 1 and 2 from ion pairing-LC-MS/MS data. Ellipse indicates 95% confidence border based on Hotelling's T^2 statistics.

These observations are consistent with a previous study of long-term human serum metabolomics which reported that longitudinal variations cannot be easily compensated for with internal standards [54]. As a solution, a normalization method using wild-type strain as a reference was employed. This method is frequently used in microarray analysis (relative transcript expression). As a result, the variation was remarkably reduced when normalization to wild-type strain (calculated as $\log_2$-transformed fold-change, i.e., ratio of metabolite in the disruptant strain vs. wild-type) was performed (**Fig. 3-3 (B)**). For LC/MS, because of a narrow linear range possibly caused by ion pairing reagent, the peaks were first calibrated using external calibration curves and corrected by a correction factor (see *Section 3.2.6*) before calculation of fold-change, while for GC/MS, peak areas normalized to ribitol (the internal standard) were used directly. Peaks from LC/MS and GC/MS were integrated after normalization.

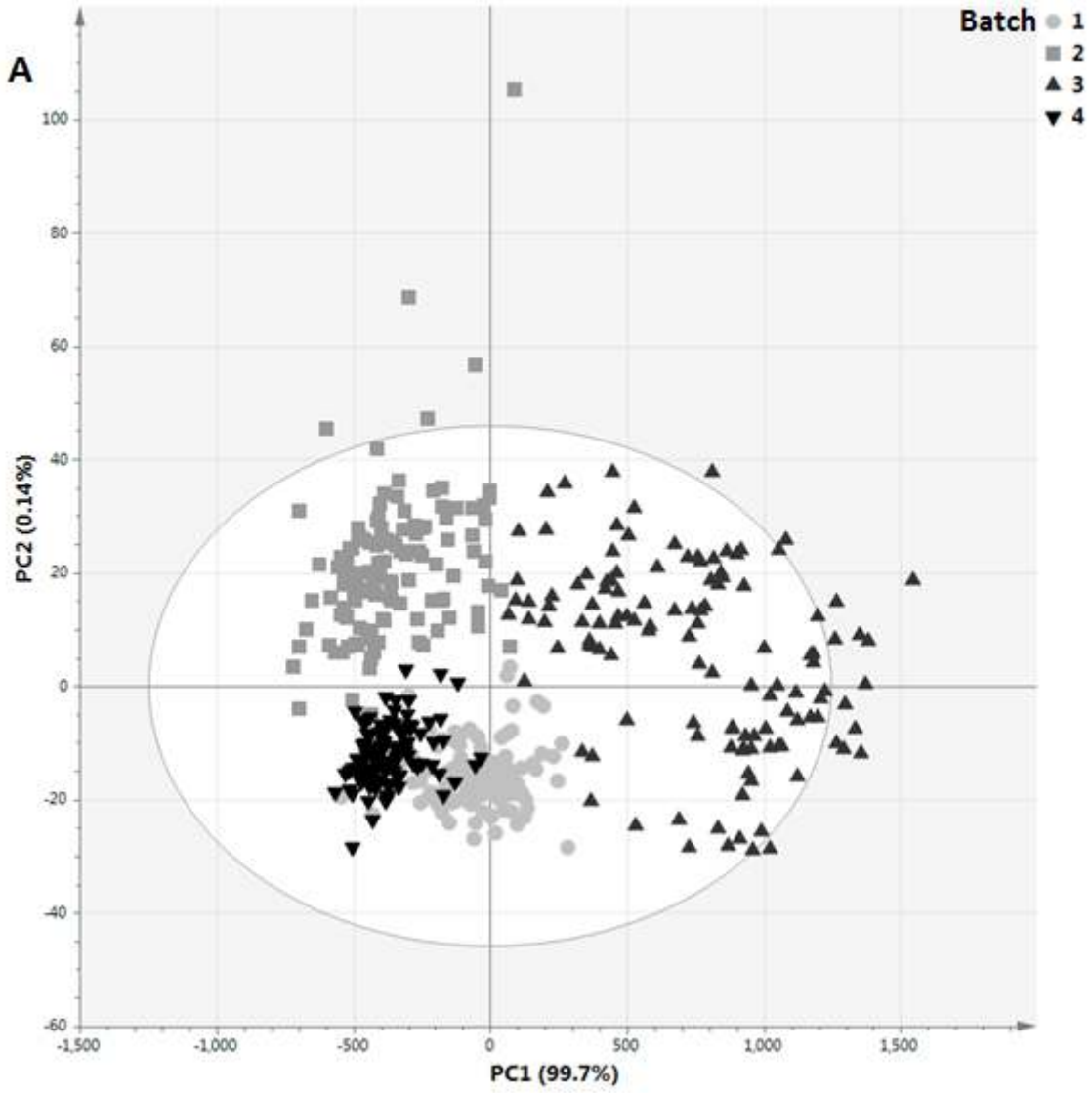


Fig. 3-3. (A) Normalization to internal standard, followed by Pareto-scaling. Numbers 1-4 indicate batch number. Batch-to-batch variation cannot be eliminated using this normalization method. Ellipse indicates 95% confidence border based on Hotelling's T^2 statistics.

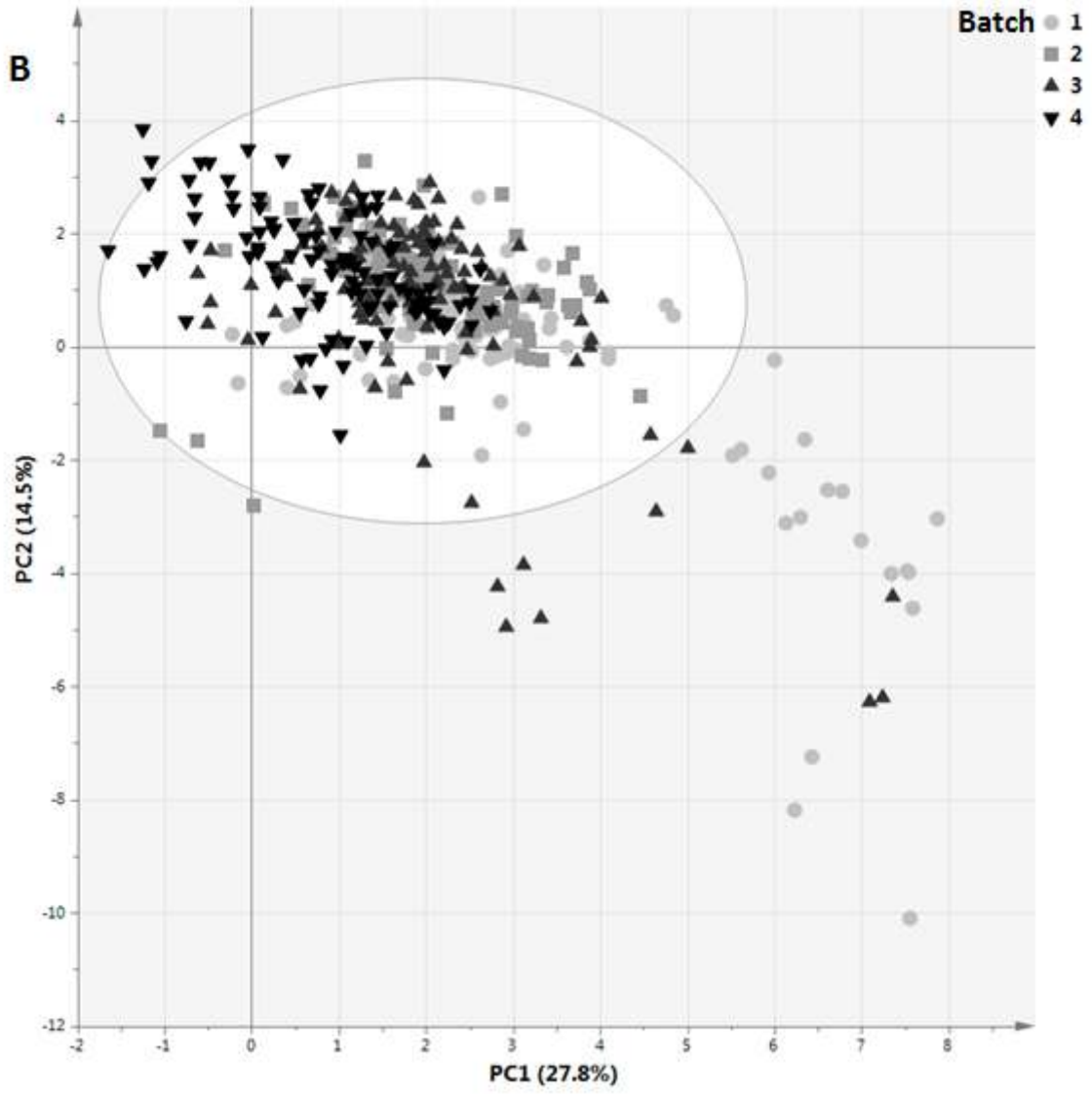


Fig. 3-3. (B) Normalization to wild-type strain (log_2-transformed fold-change of the metabolites relative to wild-type). Numbers 1-4 indicate batch number. Batch-to-batch variation was greatly reduced and data were evenly distributed. Ellipse indicates 95% confidence border based on Hotelling's T^2 statistics.

3.3.3. Clustering analysis of TF deletion strains

The objectives of this chapter are to characterize transcription factors according to their metabolic profiles and derive possible metabolite-TF and TF-TF correlations, which can deepen our knowledge regarding transcriptional regulation. To classify TF deletion strains according to their metabolic profiles, hierarchical clustering analysis (HCA) was performed. HCA is a method which finds a hierarchy of clusters that share similar characteristics based on distance or similarity measure. Compared to other methods such as K-means clustering and self-organizing map (SOM), HCA does not require *a priori* information about the data structure such as cluster number, thus it serves as a convenient unsupervised tool for interpreting complex experimental data [102].

Several algorithms are available for HCA. However, selection of a proper algorithm is subjective and depends on the purpose of the classification and whether or not the classification result satisfies a pre-determined criterion. In this chapter, the clustering algorithm was chosen based on two criteria; 1) ability to keep the original structure of the data matrix as close as possible, in which differential strains with strong characteristics (large fold change values) and outliers can be distinguished from the rest, and 2) having reasonably distributed clusters when the hierarchical tree is cut at a certain cut-point. Euclidean distance is appropriate for this purpose, as it gives a direct measure of magnitude and thus was able to separate differential strains. In contrast, Pearson's correlation provides a relative distance measure independent with magnitude, which made it fail to isolate differential strains. When comparing different linkage methods, single linkage could not generate an appropriate cut interpretation, with many small distorted clusters having few members. Average and complete linkages performed comparably, but the former had an overall structure closer to the original data matrix. Ward's minimum linkage was efficient in finding compact, homogenized clusters, but incapable of filtering outliers. Based on these findings, Euclidean distance with average linkage was chosen as the clustering algorithm in this study.

The main purpose for performing HCA is to identify differential strains and clusters. The procedures to identify differential strains and clusters are depicted in **Fig. 3-4**. First, the cut-off value was determined by taking into account the average fold-change value of ≥ 1.3 with $p < 0.05$, to separate between "differential" and "non-differential" strains. Then, strain clusters with

correlation values of 0.85 or more were identified. Differential strains here refer to the strains that have large differences in metabolic profiles compared with wild-type.

Using the metabolome dataset of $\log_2$ fold-change values, hierarchical clustering analysis was performed (**Fig. 3-5**). Here, the farther a strain is located on the outer hierarchy, the more differential it is relative to the control. **Table 3-2** summarizes cluster sets obtained from HCA. Here, the 154 transcription factor deletion strains can mainly be categorized into four groups; 1) differential, no clusters, 2) differential and formed clusters, 3) not differential and formed clusters, and 4) not differential, no clusters. A total of 27 strain clusters and two sets of no-cluster were obtained.

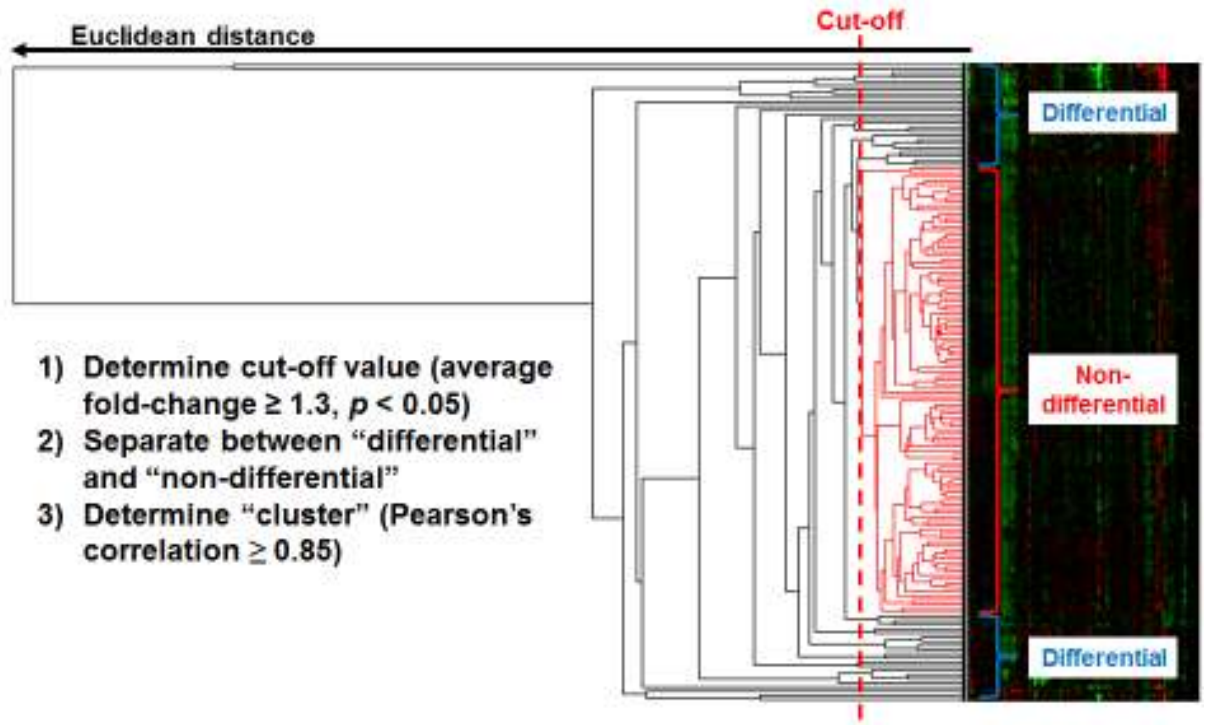


Fig. 3-4. Schematic diagram showing the procedures to determine differential strains as well as clusters.

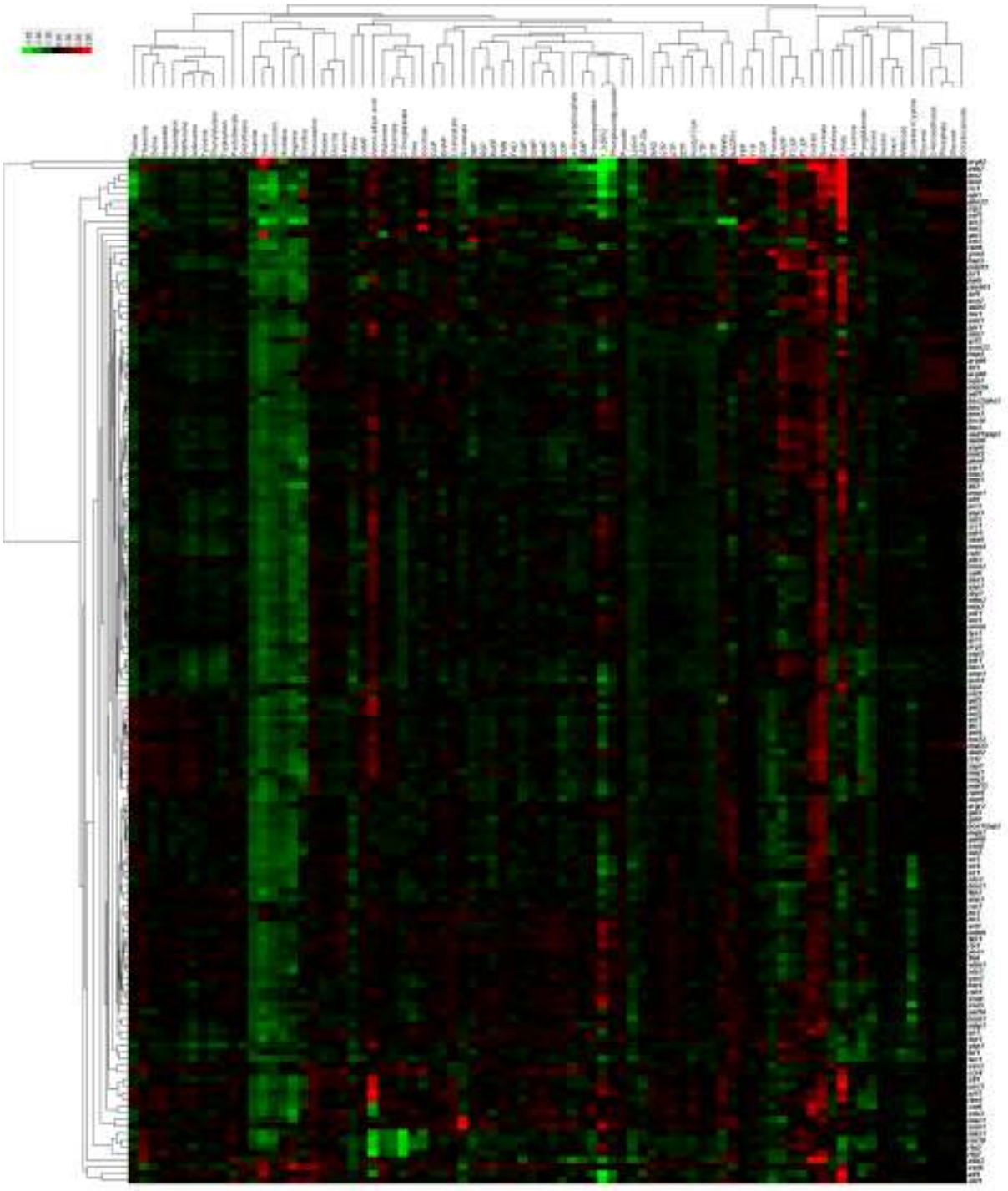


Fig. 3-5. Heat map showing metabolite changes in 154 transcription factor-related mutant strains analyzed in this study. Clustering was based on hierarchical clustering analysis, HCA using $\log_2$ normalized to the wild-type strain (fold-change) dataset. The clustering parameters were as follows: Euclidean distance and average linkage for strain clustering; Pearson's *r* and average linkage for metabolite clustering. The more differential a strain is, the farther it is located from the center and closer to the outer hierarchy.

Table 3-2 List of differential strain clusters obtained from hierarchical clustering. Strains were listed in hierarchical importance (i.e., from outer to inner hierarchy).

Group	Cluster	Members	Correlation
1: Differential, no clusters	-	*arg82*Δ, *ada2*Δ, *sef1*Δ, *swi6*Δ, *aft1*Δ, *cbf1*Δ, *gcr2*Δ, *ada3*Δ, *sin3*Δ, *gln3*Δ, *stb5*Δ, *rpn4*Δ	-
2: Differential and formed clusters	1	*ino2*Δ, *ino4*Δ, *opi1*Δ, *ric1*Δ	0.85
	2	*pho23*Δ, *stp2*Δ	0.86
	3	*mks1*Δ, *rrn10*Δ, *rtg3*Δ, *rtg1*Δ	0.87
	4	*mac1*Δ, *sum1*Δ	0.89
	5	*cst6*Δ, *sds3*Δ	0.92
	6	*sfl1*Δ, *usv1*Δ, *azf1*Δ, *rlm1*Δ	0.92
	7	*sas2*Δ, *ccr4*Δ	0.91
	8	*gcn4*Δ, *hap3*Δ, *met31*Δ	0.89
	9	*yap1*Δ, *hir1*Δ, *tec1*Δ	0.91
	10	*ixr1*Δ, *hal9*Δ, *rim101*Δ, *tuf1*Δ	0.91
	11	*ace2*Δ, *dal81*Δ, *tea1*Δ, *ash1*Δ	0.92
	12	*ppr1*Δ, *skn7*Δ	0.95
3: Not differential and formed clusters	13	*ecm22*Δ, *hap4*Δ, *aro80*Δ, *fzf1*Δ, *arg80*Δ, *uga3*Δ	0.95
	14	*met28*Δ, *oaf1*Δ	0.95
	15	*bas2*Δ, *bas1*Δ, *ime1*Δ, *lys14*Δ, *leu3*Δ, *cad1*Δ/*yap2*Δ, *dal80*Δ, *yap6*Δ, *mot3*Δ, *pho4*Δ, *stp1*Δ	0.96
	16	*hap2*Δ, *hap5*Δ, *thi2*Δ	0.95
	17	*msn1*Δ, *aft2*Δ, *arr1*Δ, *yap5*Δ, *cin5*Δ, *crz1*Δ, *pdr8*Δ, *cha4*Δ, *msn4*Δ, *rgt1*Δ, *pdr3*Δ, *msn2*Δ, *cat8*Δ, *sko1*Δ, *yap7*Δ, *xbp1*Δ, *wtm2*Δ, *mig2*Δ, *pdr1*Δ, *sut1*Δ, *ume6*Δ, *tye7*Δ, *yrr1*Δ	0.96
	18	*nrg1*Δ, *yap3*Δ	0.95
	19	*adr1*Δ, *hac1*Δ, *smp1*Δ, *yrm1*Δ	0.95
	20	*gat3*Δ, *gat2*Δ, *put3*Δ, *gat1*Δ, *gis1*Δ, *gat4*Δ, *spt23*Δ, *mal33*Δ, *dal82*Δ, *rsf2*Δ, *zap1*Δ, *mig1*Δ, *mig3*Δ, *mal13*Δ, *rgm1*Δ	0.94

	21	*cup9*Δ, *argr2*Δ, *gal3*Δ, *gal4*Δ, *ace1*Δ/*cup2*Δ, *mga1*Δ, *gal80*Δ, *ssn6*Δ, *sut2*Δ	0.96
	22	*sir3*Δ, *sir4*Δ, *sir1*Δ, *zds2*Δ	0.96
	23	*hms1*Δ, *fkh2*Δ, *yhp1*Δ	0.95
	24	*rox1*Δ, *hir2*Δ, *hir3*Δ	0.96
	25	*ezl1*Δ, *ndt80*Δ, *fkh1*Δ, *rfx1*Δ, *phd1*Δ, *flo8*Δ, *wtm1*Δ, *zds1*Δ, *yox1*Δ, *kar4*Δ, *rph1*Δ, *swi4*Δ, *swi5*Δ	0.94
	26	*hcm1*Δ, *mbp1*Δ	0.95
	27	*sir2*Δ, *tsp1*Δ	0.94
4: Not differential, no clusters	-	*gzf3*Δ, *sip4*Δ, *not3*Δ, *uaf30*Δ	-

The cut-point for "differential" in HCA is justified by average deviation from the mutant median and Hotelling's T^2 values (**Table 3-3**). A strain is defined as differential if it is located outside the confidence border of 95% based on T^2 statistics on PCA plot or having an average deviation from the mutant median of ≥ 1.3. The Hotelling's T^2 is the multivariate extension of the common two group Student's *t*-test. In a *t*-test, differences in the mean response between two populations are studied. T^2 is used when the number of response variables is two or more, although it can be used when there is only one response variable. The null hypothesis is that the group means for all response variables are equal.

Average deviation from the mutant median specifically denotes the average deviation from the median of mutant measurement, and was employed as an alternative measure of difference, independent from the wild-type strain BY4742, since there is a possibility that the wild-type profile is distorted in some metabolites. Average deviation from the mutant median was calculated as follows: the concentration of each metabolite in individual sample was divided by the median value of the metabolite concentration across all mutants, summed over all metabolites, and averaged over the number of metabolites. Division, instead of subtraction was used considering the different magnitudes of metabolite intensity values, to avoid overrepresentation of metabolites with large magnitude/highly abundant metabolites.

Table 3-3. List of differential strains based on average deviation from mutant values and Hotelling's T^2 statistics.

No.	Strain	Average deviation from the mutant median[a]	Hotelling's T^2 Range[b]	Group[c]
1	*arg82*Δ	3.16	96.77	1
2	*ada2*Δ	1.85	69.95	1
3	*aft1*Δ	1.50	40.62	1
4	*gcn4*Δ	1.44	*11.31*	2
5	*azf1*Δ	1.43	22.83	2
6	*swi6*Δ	1.37	54.94	1
7	*usv1*Δ	1.37	*19.70*	2
8	*ino2*Δ	1.37	35.02	2
9	*sum1*Δ	1.37	25.15	2
10	*mac1*Δ	1.32	*13.86*	2
11	*ada3*Δ	1.32	50.20	1
12	*gln3*Δ	1.31	*15.62*	1
13	*rlm1*Δ	1.30	*15.04*	2
14	*tea1*Δ	1.30	*5.19*	2
15	*cbf1*Δ	1.30	*12.82*	1
16	*sin3*Δ	1.30	*13.68*	1
17	*sas2*Δ	1.30	*6.96*	2
18	*ace2*Δ	1.29	*13.76*	2
19	*cst6*Δ	1.28	30.21	2
20	*dal81*Δ	1.26	*7.75*	2

[a] $\frac{\sum_{i=1}^{n} Xi/Mi}{n}$, *X*: metabolite concentration in individual mutant, *M*: median of metabolite concentration across all mutants, *i*, …, *n*: metabolite ID. This value was averaged over replicate number (3) for each strain, and ≥ 1.3 is defined as differential.

[b] based on PCA with eleven significant principal components (SIMCA-P+ ver13, Umetrics, Umeå, Sweden) (Supplementary Table S6). *Italicized values* indicate less than 95% confidence range (< 21.85).

[c] refers to the group designation in HCA (**Table 3-2**).

Moreover, clustering of metabolites based on correlation on **Fig. 3-5** revealed that metabolites that share similar pathway (e.g., amino acids biosynthesis) or having similar chemical structure (e.g., nucleotides, sugar phosphates) tended to be grouped together or closely positioned with each other. Five metabolites, namely arginine, histidine, guanine, inosine and guanosine indicated significant changes ($p < 0.05$) in more than 70% of the mutant strains, which could possibly be a unique characteristic of BY4742 derivatives, whereas pyroglutamate, fructose 6-phosphate and melibiose were altered in only 2 out of 154 mutants (Supplementary Table S7). In terms of between-metabolite correlations, the highest correlation (Pearson's *r* 0.922) was observed between 1,3-bisphosphoglycerate and phosphoenolpyruvate which are two intermediates of glycolysis. Other strong correlations were mainly exhibited by amino acids and nucleotides (**Table 3-4**).

Table 3-4. Between-metabolite correlations (≥0.6).

Positive correlation		**Negative correlation**	
Metabolites	**Pearson's *r***	**Metabolites**	**Pearson's *r***
1,3-BPG-PEP	0.922	Phenylalanine-NADP	-0.653
Phosphate-glucose	0.902	Methionine-NADP	-0.643
F1,6-F2,6P	0.858	Phenylalanine-F2,6P	-0.629
Methionine-tyrosine	0.838	Methionine-F2,6P	-0.608
Tyrosine-phenylalanine	0.837	Citrate-PEP	-0.601
Alanine-glycine	0.813	Phenylalanine-GDP	-0.601
Glutamate-2-oxoglutarate	0.802		
Isoleucine-phenylalanine	0.794		
Methionine-isoleucine	0.776		
NADP-F2,6P	0.776		
Isoleucine-tyrosine	0.775		
Phenylalanine-tryptophan	0.775		
CTP-GTP	0.770		
CTP-UTP	0.768		
Glycine-leucine	0.761		
Uracil-inositol	0.753		
UMP-2-isopropylmalate	0.752		
Methionine-phenylalanine	0.751		
AMP-ADP	0.745		
3-aminoethanol-glucose	0.744		

Inosine-guanosine	0.731
GTP-ATP	0.717
R5P-Ru5P	0.713
Uracil-melibiose	0.701
Asparagine-methionine	0.698
Proline-methionine	0.695
Alanine-leucine	0.686
2-aminoethanol-phosphate	0.679
F6P-F1P	0.678
CDP-ADP	0.674
Asparagine-tyrosine	0.673
Tyrosine-asparagine	0.673
Serine-aspartate	0.648
Serine-asparagine	0.636
Pyruvate-UMP	0.634
NAD-ATP	0.633
GMP-ADP	0.615
Asparagine-aspartate	0.614
Tyrosine-tryptophan	0.611
Arginine-histidine	0.610
GAP-DHAP	0.604
UMP-PEP	0.602
GMP-AMP	0.602
Guanine-guanosine	0.601
Methionine-aspartate	0.601

Abbreviations: 1,3BPG: 1,3 bisphosphoglycerate; PEP: phosphoenolpyruvate; F1,6P: fructose 1,6-bisphosphate; F2,6P: fructose 2,6-bisphosphate; R5P: ribose 5-phosphate; Ru5P: ribulose 5-phosphate; F6P: fructose 6-phosphate; F1P: fructose 1-phosphate; GAP: glyceraldehyde 3-phosphate; DHAP: dihydroxyacetone phosphate; NAD: nicotinamide adenine dinucleotide; NADP: nicotinamide adenine dinucleotide phosphate; ATP: adenosine 5'-triphosphate; ADP: adenosine 5'-diphosphate; AMP: adenosine 5'-monophosphate; GTP: guanosine 5'-triphosphate; GDP: guanosine 5'-diphosphate; GMP: guanosine 5'-monophosphate; CTP: cytidine 5'-triphosphate; CDP: cytidine 5'-diphosphate; UMP: uridine 5'-monophosphate.

3.3.4. Analysis of differential strains

Characterization of deletion strains based on HCA (**Fig. 3-5** and **Table 3-2**) revealed that twelve strains have differential and unique metabolic profiles. 36 more strains showed differential profile and formed clusters. HCA result is in accordance with differential analysis as depicted in **Table 3-3**, in which strains with high average deviation from the mutant median and T^2 values were located at the outer hierarchy of HCA, and thus belong to Group 1 or 2. Within these two groups, six disruptants strains (*gcr2*Δ disruptant, *aft1*Δ disruptant, *ada2*Δ disruptant, *ino2*Δ disruptant, *ino4*Δ disruptant, *opi1*Δ disruptant) showed growth defect of > 20% and six (*rpn4*Δ disruptant, *arg82*Δ disruptant, *sin3*Δ disruptant, *swi6*Δ disruptant, *cbf1*Δ disruptant, *ada3*Δ disruptant) had somewhat lower maximum specific growth rate compared to control, suggesting that altered metabolism is likely to be related with poor growth. To test whether the metabolic phenotype is a function of growth rate, Pearson's correlation coefficient was calculated between maximum specific growth rate (μ) and average deviation from the mutant median and Hotelling's T^2 range, and between μ and each metabolite in these twelve strains. If the transcription factors have an indirect effect on metabolite distributions via reduced growth rates in the deletion mutants, a correlation between mutant growth rates and the metabolic profile (expressed as the average deviation from the mutant median and Hotelling's T^2) is expected [84]. As a result, average deviation from the mutant median and Hotelling's T^2 were not correlated with growth rate (correlation coefficients were 0.17 and -0.08 respectively), indicating that metabolic profile alteration was directly due to the gene deletion and not indirectly influenced by poor growth. Almost all metabolites also showed no correlation with the maximum specific growth rate, with the exception of four metabolites i.e., trehalose, glyceraldehyde 3-phosphate (GAP) and its isomer dihydroxyacetone phosphate (DHAP), and 2-oxoglutarate, in which the correlation coefficients were -0.72, 0.72, 0.64, and 0.68 respectively. It was reported that trehalose and glycogen accumulate in *S. cerevisiae* when growth condition deteriorates, suggesting that these carbohydrates may be required for cell cycle progression at low growth rates [103]. In a separate study, intracellular DHAP+GAP was found to increase in response to an increase in growth rate in *E. coli* [104]. However the association between 2-oxoglutarate with growth rate is presently unclear.

In this study, *arg82*Δ disruptant was identified as the most differential strain with the highest values of the average deviation from the mutant median and Hotelling's T^2 (**Table 3-3**). Arg82

was originally identified as a regulator of arginine biosynthesis [105]. Arg82 is also an inositol polyphosphate multikinase (IPMK), a global regulator involved in the regulation of arginine-, phosphate-, and nitrogen-responsive genes [106]. Arginine level was increased (fold-change=1.5) in *arg82*Δ disruptant, in agreement with a previous study that reported mutation in *ARG82* leads to constitutive production of the arginine biosynthetic enzymes encoded by the *ARG1*, *ARG3*, *ARG5,6* and *ARG8* genes [107]. Deletion of *ARG82* also caused altered levels of various metabolites, mainly amino acids but includes TCA cycle intermediates and co-factors such as UTP and ATP. The most apparent was 370-fold change in citrate level. TF association search using YEASTRACT [108] (Yeast Search for Transcriptional Regulators And Consensus Tracking) showed no association between Arg82 and the genes encoding citrate synthase. However, two regulatory proteins, Arg80 and Mcm1, that have been reported to be stabilized by Arg82 [109], showed positive association with *CIT1* [21]. This observation suggests that Arg82 might regulate citrate metabolism indirectly through the interaction with other TFs. More importantly, *arg82*Δ disruptant showed a great level of metabolic alteration, with an average deviation from the mutant median of 3.16 vs. 1.85 in the second most differential strain, *ada2*Δ disruptant. Arg82 has been described as a global regulator [107], thus this finding suggests that *ARG82* may take part in more metabolic regulations than previously reported.

In addition, differential profile in *ada2*Δ disruptant and a*da3*Δ disruptant was also observed. Ada2 and Ada3 are dual function regulators involved in the regulation of many other transcription factors, and component of three chromatin modifying histone acetyltransferase (HAT) complexes: SAGA, SLIK and ADA complexes (reviewed in Sterner and Berger (2000) [110]). Between these two strains, *ada2*Δ disruptant exhibited a larger variation in metabolic profile compared to *ada3*Δ disruptant (average deviation from mutant median of 1.85 and 1.32 respectively). While both shared similar metabolic pattern in some metabolites such as amino adipic acid, inosine, orotate, succinate, malate (increased) and alpha-glycerophosphate, glutathione, CMP, pyruvate, nicotinate (decreased), only *ada2*Δ disruptant showed a marked change in citrate, 1,3-bisphosphoglycerate and phosphoenolpyruvate levels. Besides, there was no strong correlation in metabolic profile between *ada2*Δ and *ada3*Δ disruptants (correlation coefficient was 0.59). The different degree and configuration of metabolic alteration between *ada2*Δ disruptant and *ada3*Δ disruptant suggests that, while Ada2 and Ada3 share a common function in some regulations (i.e., in the histone modification [110]), there might be additional pathways exclusively affected by only one of the TFs.

3.3.5. Analysis of functionally related strains

Another interesting feature observed from HCA based on metabolome dataset was the ability to screen a specific group of strains with related functions. From **Table 3-2**, *ino2*Δ*, ino4*Δ*, opi1*Δ*, ric1*Δ cluster (correlation coefficient 0.85) can be found. This cluster is characterized by a marked increase in trehalose, succinate and citrate, as well as decrease in proline, guanosine and 1,3-bisphosphoglycerate (**Fig. 3-6 (A)**). Ino2, Ino4 and Opi1 engage in inositol metabolism and glycerolipids regulation [111–113]. Interestingly, *ino2*Δ, *ino4*Δ and *opi1*Δ disruptants all showed growth defect of >20% compared to control, with remarkably high correlation in metabolic profile, while the association of Ric1 with these three regulators has not been described yet. Similarly, *mks1*Δ*, rtg3*Δ*, rtg1*Δ and *rrn10*Δ disruptants formed a cluster with high correlation (0.87). This cluster share a similar feature of decreased amino adipic acid, glutamate and 2-oxoglutarate (**Fig. 3-6 (B)**). While Mks1, Rtg1 and Rtg3 have been previously demonstrated to be involved in mitochondrial retrograde response (RTG) in yeast [41,49,51], there was no report about the involvement of Rrn10. Here, *mks1*Δ disruptant showed the same metabolic profile as *rtg1*Δ and *rtg3*Δ disruptants, contradicted to its purported role as a negative regulator of RTG pathway [51]. Interestingly, large-scale transcriptomics data from Kemmeren *et al.* (2014) [76] also revealed a similar transcript pattern between *mks1*Δ and *rtg1*Δ or *rtg3*Δ disruptants. The discrepancy in *MKS1* deletion was probably due to the difference in the background strain as reported previously [114].

The findings in this chapter demonstrated the possibility of novel gene function exploration based on metabolic phenotype that can be unraveled by metabolomics. Further studies regarding these uncharacterized roles of transcription factors would be an interesting topic for future studies. Overall, more metabolic change was found in terms of the number of correlations and significantly altered metabolites compared to previous reports [84,85], and thus verified the high resolution approach of metabolomics employed in this study. However, it is important to note that transcriptional regulation is condition-specific and tightly controlled. In this chapter, only a 'standard' growth condition is examined. Therefore, investigation of regulations that are activated or repressed only during a specific condition should be performed under a defined experimental set-up and might involve time-series profiling. Moreover, some transcription

factors can take part in multiple pathways, resulting in compounding effects in the final metabolic profile. Un-assembling of these pathways requires thorough consideration and further experiments, such as flux measurement and kinetic modeling to investigate allosteric protein-metabolite interactions [115].

Additionally, BY4742 derivatives which carry multiple gene deletions were used here. Use of a prototrophic strain collection [116] may yield less bias caused by auxotrophic markers and feasible with minimal media, thereby reducing the compounding factors posed by additional nutrient supplementation. Nonetheless, useful TF-metabolites correlations were obtained which can be used to predict or generate new working hypotheses regarding the function of the TFs. For future studies, researchers can select only cluster of interest and conduct further experiments under a more defined condition based on the hints provided by metabolic profile similarity. Further examination of less differential clusters but share highly similar metabolic pattern (i.e., clusters in Group 3 in **Table 3-2**) might also reveal other previously unknown correlations and lead to better understanding of transcriptional regulations.

3.3.6. Comparison with transcriptomics analysis

Transcriptomics analysis using microarrays has been the leading approach for functional characterization of TFs, by which the gene expression levels are examined (up-regulated or down-regulated), usually under the deletion of the TF. Currently, the YEASTRACT [108] repository provides a convenient platform for researchers to find TF-gene associations, based on more than 1300 bibliographic references. Many researchers also deposit their raw microarray data into public repositories such as the Gene Expression Omnibus (GEO) [117] and ArrayExpress [118]. However, the most widely recognized limitation of microarrays is the low reproducibility observed when using different array platforms [119]. This is usually the result of probe-specific effects such as oligonucleotide probes versus PCR product probes, and array-synthesis effects such as on-slide synthesis versus robotic spotting.

When studying the effects of TF deletion towards metabolism, transcriptomics data alone may not be sufficient. While the interpretation of genes encoding an enzyme that catalyzes specific metabolic pathways is rather straightforward, for genes encoding a permease, for

example, it might be difficult to judge what happens to the cell by only looking at the transcript levels. For instance, in *Section 2.3.5.*, the deletion of Rtg1 and Rtg3 resulted in the decreased concentrations of TCA cycle intermediates, in agreement with the decrease in transcript levels of the genes encoding TCA cycle enzymes. However, for amino acids, it is difficult to predict the intracellular concentrations based on transcript levels only, because apart from *de novo* synthesis and degradation, they may also be up taken from the medium by a general (non-specific) transporter. Moreover, metabolic alteration may also occur even when the TF does not seem to have gene targets in that pathway through transcriptional cascade, i.e., via the interaction with other TFs. For example, in *Section 3.3.4.*, the change in citrate in *arg82*Δ disruptant was not seen at the transcript level and is likely due to the interaction of Arg82 with Arg80 and Mcm1 proteins.

In transcriptomics analysis, a standard RNA extraction protocol can practically extract all the RNAs at one time. In contrast, metabolites are composed of molecules with diverse chemical properties (polarity, water-solubility, volatility, etc.) that necessitate different extraction techniques, making comprehensive metabolite profiling very challenging. In addition, raw mass spectral data are huge in size and difficult to manage in a repository, and require large funding and trained specialists to extract the data. However, some research groups (e.g., Fiehn group from UC Davis, USA [120]) already started the initiative of making a metabolomics data repository, so that datasets from various researchers under various conditions and extraction methods can be combined and analyzed simultaneously. The availability of public cumulated data is expected to lead to new discoveries that are only possible with large diverse datasets.

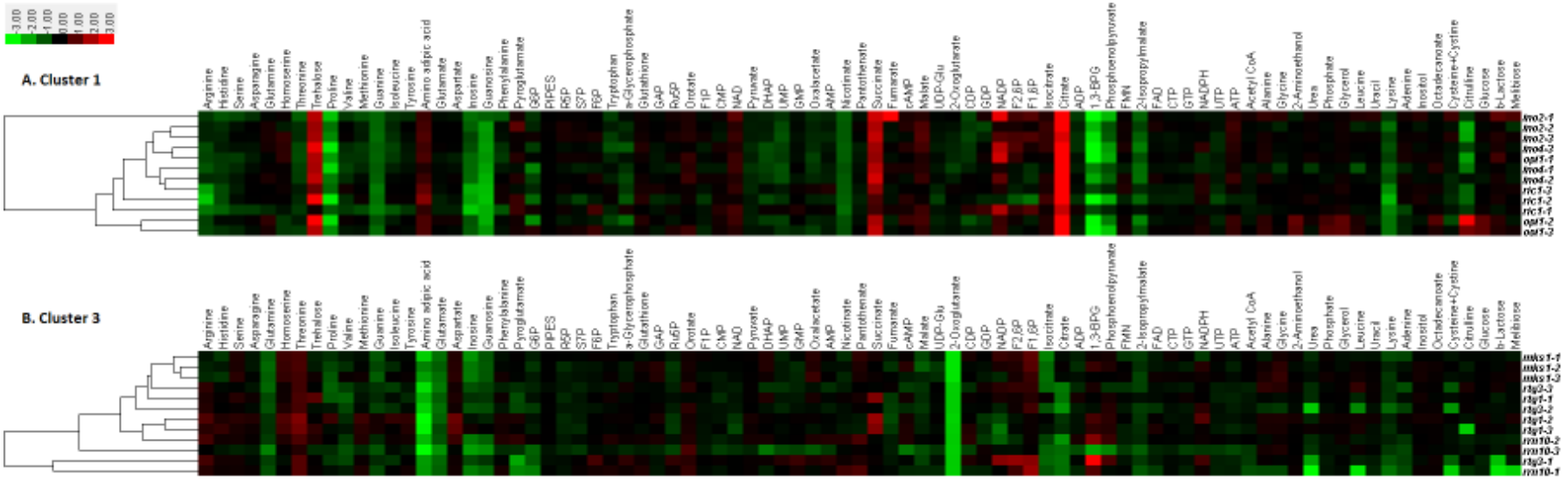


Fig. 3-6. Heat map of metabolite fold-change in two representative differential clusters **(A)** Cluster 1 (*ino2*Δ, *ino4*Δ, *opi1*Δ, *ric1*Δ) and **(B)** Cluster 3 (*mks1*Δ, *rtg1*Δ, *rtg3*Δ, *rrn10*Δ). For example, in Cluster 1, although Ino2, Ino4 and Opi1 genes have been associated with inositol and phospholipid regulation, the role of Ric1 in these regulatory pathways has not been reported. Clustering of strains that share similar metabolic profiles proves to be useful in finding both known and new gene/TF-metabolite correlations.

3.4. Conclusions

A global analysis of gene-metabolite correlations in 154 transcription factor deletion strains was conducted. Metabolome dataset provides useful insights into the effects of transcription factor deletion towards metabolic pathway rearrangement. Metabolites as the final readouts of gene transcription process can help delineate the complex rearrangement of metabolism under TF deletion which may not be always evident in transcript levels. Characterization of deletion strains using principal component analysis and hierarchical clustering analysis proved to be useful for the screening of differential and functionally related strains/genes. Both previously reported and possibly new correlations were obtained. This information can be used to open the doors to deeper investigations. The metabolome dataset presented in this chapter does not only provide information about key metabolites but also represents a useful resource for future transcriptional regulation studies.

Chapter 4 Conclusions and perspectives

The complete elucidation of cellular functions is an enormous effort and requires various strategies to capture the entire system. Metabolomics-based gene-metabolite correlation analysis is a practical and useful method to unravel new working hypotheses on the basis of metabolic phenotype. In this study, the application of metabolomics in studying metabolic alteration caused by TF deletion was investigated.

By comprehensive metabolic profiling, the differences between wild type and mutant strains defective in specific transcription factor-encoding genes were observed. Relative comparison of metabolic profiles using wild-type and knock-out strains proved to be useful in deriving possible regulatory pathways controlled by the transcription factors. In the first part of the thesis, I demonstrated the application of metabolic profiling in understanding retrograde regulation in yeast, by two bHLH regulators Rtg1 and Rtg3. The remarkable decrease in 2-oxogluratarate was reported to be the hallmark of RTG-gene perturbation. Additionally, a change in polyamine biosynthesis was also observed.

In the second part, I applied metabolomics-based screening for the characterization of 154 disruptant strains each defective in a gene encoding a transcription factor. Using two multivariate data analysis methods, principal component analysis and hierarchical clustering analysis, I assigned the deletion strains into several clusters according to their metabolic signatures. Several TF-gene and TF-TF correlations were discussed, covering both known and previously unreported observations.

At present, functional assignment of TFs is not yet complete [17,81,82]. There are also inductions or repressions of pathways that do not seem to be the direct target of the TF, which are probably due to transcriptional cascades [82]. *In silico* sequence homology analysis using computational methods has been the main tool for annotation and contributes vastly to our understanding of TF/gene regulation, but this approach cannot assign orphan genes with little or no homology to existing databases, and misannotation may occur since two genes can have very similar sequences but function differently [121]. For example, a gene sequence with a gene identifier, gi: 71915096 (GenBank:AAZ54998) was annotated as an *o*-succinylbenzoate

synthase (OSBS) in GenBank database [122] based on its high score against OSBS family, although this sequence contains a number of additional substitutions in sequence motifs conserved in authentic members of the OSBS family, an error known as 'MRF' (Missing Functionally important Residue(s)) [123]. This gene sequence was later shown to likely represent a new and unknown function in the enolase superfamily, rather than an OSBS [124]. In this regard, a global approach such as metabolomics is a more advanced option since not only it directly measures the effect of a gene deletion and allows characterization of genes according to metabolic profile similarity, but may also lead to novel discovery of biochemical pathways, such as demonstrated previously by the discovery of riboneogenesis in yeast [125].

To enable genome-wide metabolic profiling, specifically these factors must be taken into account; 1) a reliable and reproducible high-throughput analysis platform which covers as many metabolites as possible, 2) a reproducible and stable sample extraction protocol that ensures efficient recovery of various metabolites, 3) a robust peak-picking and alignment algorithm, and 4) a sophisticated data analysis software and curated database that allows cross-referencing with up-to-date research finding.

Ultimately, functional assignment of all genes is desirable, but this task requires huge and concerted effort from various researchers, as validation experiments ('omics' and other systems biology approaches) are laborious, highly sophisticated and technologically demanding. Moreover, the massive data from omics approaches require careful selection of candidate targets, and an appropriate statistical analysis must be performed to minimize false negatives and false positives. This study represents a small, but nonetheless, a significant portion of this effort. Undoubtedly, metabolomics, together with other omics, can aid the identification of important target genes and/or proteins, to be applied for example in the engineering of strains with improved phenotype (by overexpression or knockout of the identified target genes) or screening of target molecules for drug development. It is expected that metabolomics will be routinely performed, whether as a primary or complementary means in many gene regulation studies.

References

1. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S. G. Life with 6000 genes. *Science* **274,** 546, 563–567 (1996).
2. Foury, F. Human genetic diseases: a cross-talk between man and yeast. *Gene* **195,** 1–10 (1997).
3. Karathia, H., Vilaprinyo, E., Sorribas, A. & Alves, R. *Saccharomyces cerevisiae* as a model organism: a comparative study. *PLoS One* **6,** e16015 (2011).
4. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314,** 1565–1568 (2006).
5. Alper, H. & Stephanopoulos, G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab. Eng.* **9,** 258–267 (2007).
6. Sommer, S. & Fuqua, S. A. Estrogen receptor and breast cancer. *Semin. Cancer Biol.* **11,** 339–352 (2001).
7. Delmore, J. E., Issa, G. C., Lemieux, M. E., Rahl, P. B., Shi, J., Jacobs, H. M., Kastritis, E., Gilpatrick, T., Paranal, R. M., Qi, J., Chesi, M., Schinzel, A. C., McKeown, M. R., Heffernan, T. P., Vakoc, C. R., Bergsagel, P. L., Ghobrial, I. M., Richardson, P. G., Young, R. A., Hahn, W. C., Anderson, K. C., Kung, A. L., Bradner, J. E. & Mitsiades, C. S. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146,** 904–917 (2011).
8. Blancafort, P., Segal, D. J. & Barbas, C. F. Designing transcription factor architectures for drug discovery. *Mol. Pharmacol.* **66,** 1361–1371 (2004).
9. Shi, Y. Orphan nuclear receptors in drug discovery. *Drug Discov. Today* **12,** 440–445 (2007).
10. Struhl, K. Yeast transcriptional regulatory mechanisms. *Annu. Rev. Genet.* **29,** 651–674 (1995).
11. Hahn, S. & Young, E. T. Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189,** 705–736 (2011).
12. Roth, F. P., Hughes, J. D., Estep, P. W. & Church, G. M. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.* **16,** 939–945 (1998).
13. Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. & Church, G. M. Systematic determination of genetic network architecture. *Nat. Genet.* **22,** 281–285 (1999).

14. Mukherjee, S., Berger, M. F., Jona, G., Wang, X. S., Muzzey, D., Snyder, M., Young, R. A. & Bulyk, M. L. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.* **36,** 1331–1339 (2004).
15. Liu, X., Noll, D. M., Lieb, J. D. & Clarke, N. D. DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome Res.* **15,** 421–427 (2005).
16. Hu, Z., Killion, P. J. & Iyer, V. R. Genetic reconstruction of a functional transcriptional regulatory network. *Nat. Genet.* **39,** 683–687 (2007).
17. Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J.-B., Reynolds, D. B., Yoo, J., Jennings, E. G., Zeitlinger, J., Pokholok, D. K., Kellis, M., Rolfe, P. A., Takusagawa, K. T., Lander, E. S., Gifford, D. K., Fraenkel, E. & Young, R. A. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431,** 99–104 (2004).
18. Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P. & Young, R. A. Genome-wide location and function of DNA binding proteins. *Science* **290,** 2306–2309 (2000).
19. Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M. & Brown, P. O. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409,** 533–538 (2001).
20. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H. & Marton, M. J. Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.* **25,** 333–337 (2000).
21. Reimand, J., Vaquerizas, J. M., Todd, A. E., Vilo, J. & Luscombe, N. M. Comprehensive reanalysis of transcription factor knockout expression data in *Saccharomyces cerevisiae* reveals many new targets. *Nucleic Acids Res.* **38,** 4768–4777 (2010).
22. Kemmeren, P., Sameith, K., van de Pasch, L. A. L., Benschop, J. J., Lenstra, T. L., Margaritis, T., O'Duibhir, E., Apweiler, E., van Wageningen, S., Ko, C. W., van Heesch, S., Kashani, M. M., Ampatziadis-Michailidis, G., Brok, M. O., Brabers, N. A. C. H., Miles, A. J., Bouwmeester, D., van Hooff, S. R., van Bakel, H., Sluiters, E., Bakker, L. V., Snel, B., Lijnzaad, P., van Leenen, D., Groot Koerkamp, M. J. A. & Holstege, F. C. P. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* **157,** 740–752 (2014).
23. Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M. L., Oliveira, A. P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasić, I., Weichart, D., Brent, R., Broomhead, D. S., Westerhoff, H.

V, Kirdar, B., Penttilä, M., Klipp, E., Palsson, B. Ø., Sauer, U., Oliver, S. G., Mendes, P., Nielsen, J. & Kell, D. B. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* **26,** 1155–1160 (2008).

24. Fiehn, O. Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol.* **48,** 155–171 (2002).
25. Fukusaki, E. & Kobayashi, A. Plant metabolomics: potential for practical operation. *J. Biosci. Bioeng.* **100,** 347–354 (2005).
26. Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R. & Griffin, J. L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* **40,** 387–426 (2011).
27. Zheng, J., Ohata, M. & Furuta, N. Reversed-phase liquid chromatography with mixed ion-pair reagents coupled with ICP-MS for the direct speciation analysis of selenium compounds in human urine. *J. Anal. At. Spectrom.* **17,** 730–735 (2002).
28. Coulier, L., Bas, R., Jespersen, S., Verheij, E., van der Werf, M. J. & Hankemeier, T. Simultaneous quantitative analysis of metabolites using ion-pair liquid chromatography-electrospray ionization mass spectrometry. *Anal. Chem.* **78,** 6573–6582 (2006).
29. Luo, B., Groenke, K., Takors, R., Wandrey, C. & Oldiges, M. Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry. *J. Chromatogr. A* **1147,** 153–164 (2007).
30. Lu, W., Clasquin, M. F., Melamud, E., Amador-Noguez, D., Caudy, A. a & Rabinowitz, J. D. Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. *Anal. Chem.* **82,** 3212–3221 (2010).
31. Buescher, J. M., Moco, S., Sauer, U. & Zamboni, N. Ultrahigh performance liquid chromatography-tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Anal. Chem.* **82,** 4403–4412 (2010).
32. Lu, W., Bennett, B. D. & Rabinowitz, J. D. Analytical strategies for LC-MS-based targeted metabolomics. *J. Chromatogr. B* **871,** 236–242 (2008).
33. Lerner, C., Bitto, A., Pulliam, D., Nacarelli, T., Konigsberg, M., Van Remmen, H., Torres, C. & Sell, C. Reduced mammalian target of rapamycin activity facilitates mitochondrial retrograde signaling and increases life span in normal human fibroblasts. *Aging Cell* **12,** 966–977 (2013).
34. Guha, M., Srinivasan, S., Ruthel, G., Kashina, A. K., Carstens, R. P., Mendoza, A., Khanna, C., Van Winkle, T. & Avadhani, N. G. Mitochondrial retrograde signaling induces epithelial-mesenchymal transition and generates breast cancer stem cells. *Oncogene* (2013). doi:10.1038/onc.2013.467

35. Massari, M. E. & Murre, C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* **20,** 429–440 (2000).
36. Ledent, V. & Vervoort, M. The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. *Genome Res.* **11,** 754–770 (2001).
37. Jones, S. An overview of the basic helix-loop-helix proteins. *Genome Biol.* **5,** 1–6 (2004).
38. Chen, M. & Lopes, J. M. Multiple basic helix-loop-helix proteins regulate expression of the *ENO1* gene of *Saccharomyces cerevisiae*. *Eukaryot. Cell* **6,** 786–796 (2007).
39. Chen, L. & Lopes, J. M. Multiple bHLH proteins regulate *CIT2* expression in *Saccharomyces cerevisiae*. *Yeast* **27,** 345–359 (2010).
40. Liao, X. & Butow, R. A. *RTG1* and *RTG2*: Two yeast genes required for a novel path of communication from mitochondria to the nucleus. *Cell* **72,** 61–71 (1993).
41. Butow, R. A. & Avadhani, N. G. Mitochondrial signaling: the retrograde response. *Mol. Cell* **14,** 1–15 (2004).
42. Liu, Z. & Butow, R. A. Mitochondrial retrograde signaling. *Annu. Rev. Genet.* **40,** 159–185 (2006).
43. Sekito, T., Thornton, J. & Butow, R. A. Mitochondria-to-nuclear signaling is regulated by the subcellular localization of the transcription factors Rtg1p and Rtg3p. *Mol. Biol. Cell* **11,** 2103–2115 (2000).
44. Jia, Y., Rothermel, B., Thornton, J. & Butow, R. A. A basic helix-loop-helix-leucine zipper transcription complex in yeast functions in a signaling pathway from mitochondria to the nucleus. *Mol. Cell. Biol.* **17,** 1110–1117 (1997).
45. Traven, A., Wong, J. M., Xu, D., Sopta, M. & Ingles, C. J. Interorganellar communication. Altered nuclear gene expression profiles in a yeast mitochondrial DNA mutant. *J. Biol. Chem.* **276,** 4020–4027 (2001).
46. Herve, M., Buffin-Meyer, B., Bouet, F. & Tran-Dinh, S. Detection of modifications in the glucose metabolism induced by genetic mutations in *Saccharomyces cerevisiae* by 13C- and 1H-NMR spectroscopy. *Eur. J. Biochem.* **267,** 3337–3344 (2000).
47. Crespo, J. L., Powers, T., Fowler, B. & Hall, M. N. The TOR-controlled transcription activators *GLN3*, *RTG1*, and *RTG3* are regulated in response to intracellular levels of glutamine. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 6784–6789 (2002).
48. Dilova, I., Chen, C.-Y. & Powers, T. Mks1 in concert with TOR signaling negatively regulates RTG target gene expression in *S. cerevisiae*. *Curr. Biol.* **12,** 389–395 (2002).
49. Dilova, I., Aronova, S., Chen, J. C.-Y. & Powers, T. Tor signaling and nutrient-based signals converge on Mks1p phosphorylation to regulate expression of Rtg1.Rtg3p-dependent target genes. *J. Biol. Chem.* **279,** 46527–46535 (2004).

50. Liu, Z. & Butow, R. A. A transcriptional switch in the expression of yeast tricarboxylic acid cycle genes in response to a reduction or loss of respiratory function. *Mol. Cell. Biol.* **19,** 6720–6728 (1999).
51. Tate, J. J., Cox, K. H., Rai, R. & Cooper, T. G. Mks1p is required for negative regulation of retrograde gene expression in *Saccharomyces cerevisiae* but does not affect nitrogen catabolite repression-sensitive gene expression. *J. Biol. Chem.* **277,** 20477–20482 (2002).
52. Zhang, F., Pracheil, T., Thornton, J. & Liu, Z. Adenosine triphosphate (ATP) is a candidate signaling molecule in the mitochondria-to-nucleus retrograde response pathway. *Genes (Basel).* **4,** 86–100 (2013).
53. Kaiser, C., Michaelis, S. & Mitchell, A. *Methods in yeast genetics: A Cold Spring Harbor Laboratory course manual*. *Cold Spring Harb. NY* 234 (Cold Spring Harbor, 1994).
54. Zelena, E., Dunn, W. B., Broadhurst, D., Francis-McIntyre, S., Carroll, K. M., Begley, P., O'Hagan, S., Knowles, J. D., Halsall, A., Wilson, I. D. & Kell, D. B. Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal. Chem.* **81,** 1357–1364 (2009).
55. Gika, H. G., Theodoridis, G. A., Earll, M. & Wilson, I. D. A QC approach to the determination of day-to-day reproducibility and robustness of LC-MS methods for global metabolite profiling in metabonomics/metabolomics. *Bioanalysis* **4,** 2239–2247 (2012).
56. Gika, H. G., Theodoridis, G. A., Wingate, J. E. & Wilson, I. D. Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. *J. Proteome Res.* **6,** 3291–3303 (2007).
57. Xia, J., Mandal, R., Sinelnikov, I. V, Broadhurst, D. & Wishart, D. S. MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis. *Nucleic Acids Res.* **40,** W127–133 (2012).
58. De Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20,** 1453–1454 (2004).
59. Saldanha, A. J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20,** 3246–3248 (2004).
60. Klukas, C. & Schreiber, F. Integration of -omics data and networks for biomedical research with VANTED. *J. Integr. Bioinform.* **7,** 112–118 (2010).
61. Parrella, E. & Longo, V. D. The chronological life span of *Saccharomyces cerevisiae* to study mitochondrial dysfunction and disease. *Methods* **46,** 256–262 (2008).
62. Yoshida, R., Tamura, T., Takaoka, C., Harada, K., Kobayashi, A., Mukai, Y. & Fukusaki, E. Metabolomics-based systematic prediction of yeast lifespan and its application for semi-rational screening of ageing-related mutants. *Aging Cell* **9,** 616–625 (2010).

63. Poyton, R. O. & McEwen, J. E. Crosstalk between nuclear and mitochondrial genomes. *Annu. Rev. Biochem.* **65,** 563–607 (1996).
64. Zampar, G. G., Kümmel, A., Ewald, J., Jol, S., Niebel, B., Picotti, P., Aebersold, R., Sauer, U., Zamboni, N. & Heinemann, M. Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. *Mol. Syst. Biol.* **9,** 1–13 (2013).
65. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278,** 680–686 (1997).
66. Eisenberg, T., Carmona-Gutierrez, D., Büttner, S., Tavernarakis, N. & Madeo, F. Necrosis in yeast. *Apoptosis* **15,** 257–268 (2010).
67. Farrugia, G. & Balzan, R. Oxidative stress and programmed cell death in yeast. *Front. Oncol.* **2,** 2–21 (2012).
68. Borghouts, C., Benguria, A., Wawryn, J. & Jazwinski, S. M. Rtg2 protein links metabolism and genome stability in yeast longevity. *Genetics* **166,** 765–777 (2004).
69. Barros, M. H., Bandy, B., Tahara, E. B. & Kowaltowski, A. J. Higher respiratory activity decreases mitochondrial reactive oxygen release and increases life span in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279,** 49883–49888 (2004).
70. Powers, R. W., Kaeberlein, M., Caldwell, S. D., Kennedy, B. K. & Fields, S. Extension of chronological life span in yeast by decreased TOR pathway signaling. *Genes Dev.* **20,** 174–184 (2006).
71. Fabrizio, P. & Longo, V. D. The chronological life span of *Saccharomyces cerevisiae*. *Aging Cell* **2,** 73–81 (2003).
72. Alvers, A. L., Fishwick, L. K., Wood, M. S., Hu, D., Chung, H. S., Dunn, W. A. & Aris, J. P. Autophagy and amino acid homeostasis are required for chronological longevity in *Saccharomyces cerevisiae*. *Aging Cell* **8,** 353–369 (2009).
73. Small, W. C., Brodeur, R. D., Sandor, A., Fedorova, N., Li, G., Butow, R. A. & Srere, P. A. Enzymic and metabolic studies on retrograde regulation mutants of yeast. *Biochemistry* **34,** 5569–5576 (1995).
74. Grenson, M., Hou, C. & Crabeel, M. Multiplicity of the amino acid permeases in *Saccharomyces cerevisiae* IV. Evidence for a general amino acid permease. *J. Bacteriol.* **103,** 770–777 (1970).
75. Jauniaux, J.-C. & Grenson, M. *GAP1*, the general amino acid permease gene of *Saccharomyces cerevisiae*. Nucleotide sequence, protein similarity with the other bakers yeast amino acid permeases, and nitrogen catabolite repression. *Eur. J. Biochem.* **190,** 39–44 (1990).

76. Chen, E. J. & Kaiser, C. A. Amino acids regulate the intracellular trafficking of the general amino acid permease of *Saccharomycescerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 14837–14842 (2002).
77. Schreve, J. L., Sin, J. K. & Garrett, J. M. The *Saccharomyces cerevisiae YCC5* (YCL025c) gene encodes an amino acid permease, Agp1, which transports asparagine and glutamine. *J. Bacteriol.* **180,** 2556–2559 (1998).
78. Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J.-B., Volkert, T. L., Fraenkel, E., Gifford, D. K. & Young, R. A. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298,** 799–804 (2002).
79. Zheng, W., Zhao, H., Mancera, E., Steinmetz, L. M. & Snyder, M. Genetic analysis of variation in transcription factor binding in yeast. *Nature* **464,** 1187–1191 (2010).
80. Dikicioglu, D., Karabekmez, E., Rash, B., Pir, P., Kirdar, B. & Oliver, S. G. How yeast re-programmes its transcriptional profile in response to different nutrient impulses. *BMC Syst. Biol.* **5,** 1–16 (2011).
81. Chua, G., Robinson, M. D., Morris, Q. & Hughes, T. R. Transcriptional networks: reverse-engineering gene regulation on a global scale. *Curr. Opin. Microbiol.* **7,** 638–646 (2004).
82. Chua, G., Morris, Q. D., Sopko, R., Robinson, M. D., Ryan, O., Chan, E. T., Frey, B. J., Andrews, B. J., Boone, C. & Hughes, T. R. Identifying transcription factor functions and targets by phenotypic activation. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 12045–12050 (2006).
83. MacArthur, S., Li, X.-Y., Li, J., Brown, J. B., Chu, H. C., Zeng, L., Grondona, B. P., Hechmer, A., Simirenko, L., Keränen, S. V. E., Knowles, D. W., Stapleton, M., Bickel, P., Biggin, M. D. & Eisen, M. B. Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* **10,** R80 (2009).
84. Fendt, S.-M., Oliveira, A. P., Christen, S., Picotti, P., Dechant, R. C. & Sauer, U. Unraveling condition-dependent networks of transcription factors that control metabolic pathway activity in yeast. *Mol. Syst. Biol.* **6,** 1–11 (2010).
85. Cooper, S. J., Finney, G. L., Brown, S. L., Nelson, S. K., Hesselberth, J., MacCoss, M. J. & Fields, S. High-throughput profiling of amino acids in strains of the *Saccharomyces cerevisiae* deletion collection. *Genome Res.* **20,** 1288–1296 (2010).
86. Jewison, T., Knox, C., Neveu, V., Djoumbou, Y., Guo, A. C., Lee, J., Liu, P., Mandal, R., Krishnamurthy, R., Sinelnikov, I., Wilson, M. & Wishart, D. S. YMDB: the Yeast Metabolome Database. *Nucleic Acids Res.* **40,** D815–820 (2012).

87. Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., Berden, J. a, Brindle, K. M., Kell, D. B., Rowland, J. J., Westerhoff, H. V, van Dam, K. & Oliver, S. G. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* **19,** 45–50 (2001).
88. Allen, J., Davey, H. M., Broadhurst, D., Heald, J. K., Rowland, J. J., Oliver, S. G. & Kell, D. B. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.* **21,** 692–696 (2003).
89. Shirai, T., Matsuda, F., Okamoto, M. & Kondo, A. Evaluation of control mechanisms for *Saccharomyces cerevisiae* central metabolic reactions using metabolome data of eight single-gene deletion mutants. *Appl. Microbiol. Biotechnol.* **97,** 3569–3577 (2013).
90. EUROSCARF. at <http://web.uni-frankfurt.de/fb15/mikro/euroscarf/index.html>
91. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A. P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.-D., Flaherty, P., Foury, F., Garfinkel, D. J., Gerstein, M., Gotte, D., Güldener, U., Hegemann, J. H., Hempel, S., Herman, Z., Jaramillo, D. F., Kelly, D. E., Kelly, S. L., Kötter, P., LaBonte, D., Lamb, D. C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S. L., Revuelta, J. L., Roberts, C. J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D. D., Sookhai-Mahadeo, S., Storms, R. K., Strathern, J. N., Valle, G., Voet, M., Volckaert, G., Wang, C., Ward, T. R., Wilhelmy, J., Winzeler, E. A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J. D., Snyder, M., Philippsen, P., Davis, R. W. & Johnston, M. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418,** 387–391 (2002).
92. Crutchfield, C. A., Lu, W., Melamud, E. & Rabinowitz, J. D. Mass spectrometry-based metabolomics of yeast. *Methods Enzymol.* **470,** 393–426 (2010).
93. Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N. & Willmitzer, L. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J.* **23,** 131–142 (2000).
94. Gika, H. G., Macpherson, E., Theodoridis, G. A. & Wilson, I. D. Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. *J. Chromatogr. B* **871,** 299–305 (2008).
95. Lommen, A. MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal. Chem.* **81,** 3079–3086 (2009).
96. Tsugawa, H., Tsujimoto, Y., Arita, M., Bamba, T. & Fukusaki, E. GC/MS based metabolomics: development of a data mining system for metabolite identification by

using soft independent modeling of class analogy (SIMCA). *BMC Bioinformatics* **12,** 131 (2011).

97. Chapman, A. G., Fall, L. & Atkinson, D. E. Adenylate energy charge in *Escherichia coli* during growth and starvation. *J. Bacteriol.* **108,** 1072–1086 (1971).
98. Mashego, M. R., Jansen, M. L. a, Vinke, J. L., van Gulik, W. M. & Heijnen, J. J. Changes in the metabolome of *Saccharomyces cerevisiae* associated with evolution in aerobic glucose-limited chemostats. *FEMS Yeast Res.* **5,** 419–430 (2005).
99. Kamleh, M. A., Ebbels, T. M. D., Spagou, K., Masson, P. & Want, E. J. Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies. *Anal. Chem.* **84,** 2670–2677 (2012).
100. Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B. & Goodacre, R. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **6,** 1060–1083 (2011).
101. Lu, W., Clasquin, M. F., Melamud, E., Amador-Noguez, D., Caudy, A. a & Rabinowitz, J. D. Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. *Anal. Chem.* **82,** 3212–3221 (2010).
102. Sharma, S. *Applied multivariate techniques*. 493 (John Wiley & Sons, Inc. New York, NY, USA, 1996).
103. Sillje, H. H. W., Paalman, J. W. G., ter Schure, E. G., Olsthoorn, S. Q. B., Verkleij, A. J., Boonstra, J. & Verrips, C. T. Function of trehalose and glycogen in cell cycle progression and cell viability in *Saccharomyces cerevisiae*. *J. Bacteriol.* **181,** 396–400 (1999).
104. Schaub, J. & Reuss, M. In vivo dynamics of glycolysis in *Escherichia coli* shows need for growth-rate dependent metabolome analysis. *Biotechnol. Prog.* **24,** 1402–1407 (2008).
105. Bechet, J., Grenson, M. & Wiame, J. M. Mutations affecting the repressibility of arginine biosynthetic enzymes in *Sacchromyces cerevisiae*. *Eur. J. Biochem.* **12,** 31–39 (1970).
106. Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., Simison, M., Weng, S. & Wong, E. D. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40,** D700–705 (2012).
107. Messenguy, F. & Dubois, E. Genetic evidence for a role for *MCM1* in the regulation of arginine metabolism in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13,** 2586–2592 (1993).

108. Teixeira, M. C., Monteiro, P. T., Guerreiro, J. F., Gonçalves, J. P., Mira, N. P., dos Santos, S. C., Cabrito, T. R., Palma, M., Costa, C., Francisco, A. P., Madeira, S. C., Oliveira, A. L., Freitas, A. T. & Sá-Correia, I. The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **42,** D161–166 (2014).

109. El Bakkoury, M., Dubois, E. & Messenguy, F. Recruitment of the yeast MADS-box proteins, ArgRI and Mcm1 by the pleiotropic factor ArgRIII is required for their stability. *Mol. Microbiol.* **35,** 15–31 (2000).

110. Sterner, D. E. & Berger, S. L. Acetylation of histones and transcription-related factors. *Microbiol. Mol. Biol. Rev.* **64,** 435–459 (2000).

111. Graves, J. A. & Henry, S. A. Regulation of the yeast *INO1* gene: the products of the *INO2*, *INO4* and *OPI1* regulatory genes are not required for repression in response to inositol. *Genetics* **154,** 1485–1495 (2000).

112. Henry, S. A., Kohlwein, S. D. & Carman, G. M. Metabolism and regulation of glycerolipids in the yeast *Saccharomyces cerevisiae*. *Genetics* **190,** 317–349 (2012).

113. Hoppen, J., Repenning, A., Albrecht, A., Geburtig, S. & Schüller, H.-J. Comparative analysis of promoter regions containing binding sites of the heterodimeric transcription factor Ino2/Ino4 involved in yeast phospholipid biosynthesis. *Yeast* **22,** 601–613 (2005).

114. Dilova, I. & Powers, T. Accounting for strain-specific differences during RTG target gene regulation in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **6,** 112–119 (2006).

115. Link, H., Kochanowski, K. & Sauer, U. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nat. Biotechnol.* **31,** 357–61 (2013).

116. Mülleder, M., Capuano, F., Pir, P., Christen, S., Sauer, U., Oliver, S. G. & Ralser, M. A prototrophic deletion mutant collection for yeast metabolomics and systems biology. *Nat. Biotechnol.* **30,** 1176–1178 (2012).

117. Barrett, T., Suzek, T. O., Troup, D. B., Wilhite, S. E., Ngau, W.-C., Ledoux, P., Rudnev, D., Lash, A. E., Fujibuchi, W. & Edgar, R. NCBI GEO: mining millions of expression profiles--database and tools. *Nucleic Acids Res.* **33,** D562–566 (2005).

118. Parkinson, H., Sarkans, U., Shojatalab, M., Abeygunawardena, N., Contrino, S., Coulson, R., Farne, A., Lara, G. G., Holloway, E., Kapushesky, M., Lilja, P., Mukherjee, G., Oezcimen, A., Rayner, T., Rocca-Serra, P., Sharma, A., Sansone, S. & Brazma, A. ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **33,** D553–555 (2005).

119. Kuo, W. P., Jenssen, T.-K., Butte, A. J., Ohno-Machado, L. & Kohane, I. S. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* **18,** 405–412 (2002).

120. Metabolomics Repository and LIMS — Metabolomics Fiehn Lab. at <http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/LIMS>
121. Prosser, G. A., Larrouy-Maumus, G. & de Carvalho, L. P. S. Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. *EMBO Rep.* (2014). doi:10.15252/embr.201338283
122. Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **41,** D36–42 (2013).
123. Schnoes, A. M., Brown, S. D., Dodevski, I. & Babbitt, P. C. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.* **5,** 1–13 (2009).
124. Glasner, M. E., Fayazmanesh, N., Chiang, R. A., Sakai, A., Jacobson, M. P., Gerlt, J. A. & Babbitt, P. C. Evolution of structure and function in the o-succinylbenzoate synthase/N-acylamino acid racemase family of the enolase superfamily. *J. Mol. Biol.* **360,** 228–250 (2006).
125. Clasquin, M. F., Melamud, E., Singer, A., Gooding, J. R., Xu, X., Dong, A., Cui, H., Campagna, S. R., Savchenko, A., Yakunin, A. F., Rabinowitz, J. D. & Caudy, A. a. Riboneogenesis in yeast. *Cell* **145,** 969–980 (2011).

Appendices

Supplementary Table S1. Optimized multiple reaction monitoring (MRM) parameters and retention time for each metabolite, measured by LC-MS/MS.

No	Metabolite	LC-MS Method*/mode	Precursor ion *m/z*	Product ion *m/z*	Retention Time (min)	Target Q1 Pre Bias (V)	Target Collision Energy (V)	Target Q3 Pre Bias (V)
1	Arginine	Ion-pair RP/ESI negative	173.10	131.20	1.039	13	15	24
2	Histidine	Ion-pair RP/ESI negative	154.00	93.15	1.038	12	21	16
3	Serine	Ion-pair RP/ESI negative	104.00	74.15	1.138	12	16	13
4	Asparagine	Ion-pair RP/ESI negative	131.00	113.15	1.151	10	15	21
5	Glutamine	Ion-pair RP/ESI negative	145.00	127.05	1.160	12	18	18
6	Threonine	Ion-pair RP/ESI negative	118.00	74.05	1.186	11	15	13
7	Trehalose	Ion-pair RP/ESI negative	341.00	89.00	1.302	15	23	16
8	Proline	Ion-pair RP/ESI negative	114.00	68.10	1.331	10	15	11
9	Methionine	Ion-pair RP/ESI negative	148.00	47.05	1.987	11	14	16
10	Isoleucine	Ion-pair RP/ESI negative	130.10	45.00	2.578	11	15	15
11	Adenine	Ion-pair RP/ESI negative	134.05	107.35	2.829	28	20	20
12	Tyrosine	Ion-pair RP/ESI negative	180.00	163.05	2.834	12	18	18
13	Xanthine	Ion-pair RP/ESI negative	151.20	108.15	3.000	16	20	19
14	Amino adipic acid	Ion-pair RP/ESI negative	160.00	116.20	3.171	12	17	21
15	Glutamate	Ion-pair RP/ESI negative	146.00	102.20	3.273	11	15	18
16	Aspartate	Ion-pair RP/ESI negative	132.00	88.05	3.467	10	14	15
17	Inosine	Ion-pair RP/ESI negative	267.00	135.15	4.559	21	23	24
18	Guanosine	Ion-pair RP/ESI negative	282.10	150.20	4.706	22	21	28
19	Phenylalanine	Ion-pair RP/ESI negative	164.00	103.15	4.854	13	18	19
20	Glycolate	Ion-pair RP/ESI negative	75.00	75.00	5.001	16	15	15
21	Glycerate	Ion-pair RP/ESI negative	105.00	75.15	5.088	12	15	26
22	Adenosine	Ion-pair RP/ESI negative	266.10	134.15	5.385	18	20	26

23	Glyoxylate	Ion-pair RP/ESI negative	73.00	73.00	5.577	14	13	15
24	Pyroglutamate	Ion-pair RP/ESI negative	128.00	84.10	6.438	10	14	15
25	Glucose 6-phosphate	Ion-pair RP/ESI negative	258.90	97.05	6.734	20	21	17
26	PIPES (Internal standard)	Ion-pair RP/ESI negative	301.00	193.25	6.949	12	28	21
27	Sedoheptulose 7-phosphate	Ion-pair RP/ESI negative	288.90	97.10	7.001	23	23	17
28	Fructose 6-phosphate	Ion-pair RP/ESI negative	258.90	97.10	7.058	20	15	17
29	Ribose 5-phosphate	Ion-pair RP/ESI negative	229.10	96.95	7.067	18	13	18
30	Tryptophan	Ion-pair RP/ESI negative	203.10	116.15	7.070	16	18	21
31	a-Glycerophosphate	Ion-pair RP/ESI negative	171.10	79.10	7.226	13	16	13
32	Glutathione	Ion-pair RP/ESI negative	305.90	143.20	7.536	21	19	26
33	Ribulose 5-phosphate	Ion-pair RP/ESI negative	229.00	97.10	7.767	17	13	17
34	Orotate	Ion-pair RP/ESI negative	155.00	111.15	7.928	12	14	20
35	CMP	Ion-pair RP/ESI negative	322.00	79.10	8.099	25	28	13
36	Fructose 1-phosphate	Ion-pair RP/ESI negative	258.90	97.05	8.102	20	21	17
37	NAD	Ion-pair RP/ESI negative	662.10	540.10	8.281	26	18	26
38	Pyruvate	Ion-pair RP/ESI negative	87.00	43.05	8.318	10	11	14
39	DHAP	Ion-pair RP/ESI negative	168.90	97.05	8.608	13	12	17
40	UMP	Ion-pair RP/ESI negative	322.90	97.10	8.815	25	24	17
41	GMP	Ion-pair RP/ESI negative	362.00	79.10	8.995	29	28	13
42	Oxalacetate	Ion-pair RP/ESI negative	131.00	87.00	9.310	25	11	27
43	TMP	Ion-pair RP/ESI negative	321.00	79.10	9.719	25	38	14
44	AMP	Ion-pair RP/ESI negative	346.00	79.05	9.811	14	32	14
45	Nicotinate	Ion-pair RP/ESI negative	122.00	78.15	9.983	14	15	13
46	Pantothenate	Ion-pair RP/ESI negative	218.00	88.00	10.022	21	14	16
47	Succinate	Ion-pair RP/ESI negative	117.00	73.20	10.155	13	15	12
48	Fumarate	Ion-pair RP/ESI negative	115.00	71.10	10.278	13	10	12
49	cAMP	Ion-pair RP/ESI negative	328.00	134.10	10.465	15	27	24

50	Malate	Ion-pair RP/ESI negative	132.90	115.20	10.578	10	17	21
51	UDP-glucose	Ion-pair RP/ESI negative	564.80	323.10	10.712	24	26	15
52	2-Oxoglutarate	Ion-pair RP/ESI negative	145.00	101.20	10.745	11	13	18
53	CDP	Ion-pair RP/ESI negative	401.80	79.05	10.753	16	43	13
54	UDP	Ion-pair RP/ESI negative	402.90	79.05	10.807	16	48	13
55	NADP	Ion-pair RP/ESI negative	741.80	620.10	10.811	26	18	30
56	3Phosphoglycerate	Ion-pair RP/ESI negative	184.90	97.05	10.829	14	16	17
57	Fructose 2,6-bisphosphate	Ion-pair RP/ESI negative	338.90	241.15	10.834	26	19	27
58	Fructose 1,6-bisphosphate	Ion-pair RP/ESI negative	338.90	97.10	10.838	26	22	17
59	NADH	Ion-pair RP/ESI negative	664.00	78.95	10.876	24	57	13
60	Isocitrate	Ion-pair RP/ESI negative	190.90	73.20	10.891	13	22	26
61	Citrate	Ion-pair RP/ESI negative	190.90	87.00	10.892	13	18	14
62	ADP	Ion-pair RP/ESI negative	425.90	79.10	10.913	17	47	13
63	Bisphosphoglycerate	Ion-pair RP/ESI negative	265.00	167.15	10.919	11	18	29
64	Phosphoenolpyruvate	Ion-pair RP/ESI negative	167.00	78.95	10.928	15	13	13
65	2-Isopropylmalate	Ion-pair RP/ESI negative	175.00	115.20	10.998	13	16	21
66	FAD	Ion-pair RP/ESI negative	783.90	97.10	11.155	20	51	17
67	CTP	Ion-pair RP/ESI negative	481.90	159.10	11.171	19	36	29
68	GTP	Ion-pair RP/ESI negative	521.90	159.05	11.185	20	32	29
69	NADPH	Ion-pair RP/ESI negative	744.00	159.00	11.201	26	60	30
70	UTP	Ion-pair RP/ESI negative	482.90	159.10	11.206	19	36	29
71	ATP	Ion-pair RP/ESI negative	505.90	159.10	11.226	20	35	29
72	Acetyl-coA	Ion-pair RP/ESI negative	808.00	408.00	11.382	20	37	28
73	Cystine	RP/ESI positive	241.05	74.00	1.349	-25	-40	-14
74	Hydroxyproline	RP/ESI positive	131.70	85.95	1.437	-30	-19	-14
75	Cysteine	RP/ESI positive	122.00	59.00	1.482	-29	-40	-13
76	Homoserine	RP/ESI positive	119.70	74.15	1.514	-30	-15	-12
77	Alanine	RP/ESI positive	90.05	44.05	1.563	-16	-20	-19
78	Citrulline	RP/ESI positive	175.60	70.00	1.572	-30	-30	-30
79	Ornithine	RP/ESI positive	132.70	69.75	1.721	-30	-30	-30
80	Lysine	RP/ESI positive	146.70	83.95	1.783	-30	-25	-13

81	b-Alanine	RP/ESI positive	89.90	30.15	1.850	-14	-15	-30
82	Uracil	RP/ESI positive	113.15	70.05	2.278	-19	-43	-29
83	4-Aminobutyrate	RP/ESI positive	103.70	87.05	2.302	-30	-16	-15
84	Putrescine	RP/ESI positive	88.80	71.70	2.375	-30	-20	-30
85	Glycine	RP/ESI positive	118.05	43.05	2.829	-17	-40	-13
86	Valine	RP/ESI positive	118.10	72.10	2.857	-19	-10	-29
87	Spermidine	RP/ESI positive	145.70	72.20	3.130	-30	-20	-30
88	Hypoxanthine	RP/ESI positive	137.05	55.05	3.236	-21	-40	-28
89	Uridine	RP/ESI positive	244.90	113.05	3.322	-27	-10	-15
90	Guanine	RP/ESI positive	151.95	135.05	3.619	-29	-20	-16
91	S-Adenosylmethionine	RP/ESI positive	398.50	250.20	3.632	-30	-17	-25
92	Cytidine	RP/ESI positive	244.00	112.05	4.224	-16	-20	-16
93	Deoxycytidine	RP/ESI positive	228.10	112.10	5.629	-25	-10	-24
94	Leucine	RP/ESI positive	131.70	43.05	6.203	-30	-25	-17
95	Deoxyguanosine	RP/ESI positive	268.00	152.00	6.222	-18	-10	-15
96	Thymidine	RP/ESI positive	243.10	127.05	6.436	-27	-10	-29
97	Deoxyadenosine	RP/ESI positive	252.10	136.10	6.890	-17	-20	-27

*RP: reversed phase

Supplementary Table S2. Summary of PCA with five significant components performed for metabolome dataset from metabolic profiling of BY4742 and RTG deletion strains.

Component	R2X	R2X (cum)	Eigen value	Q2	Limit	Q2 (cum)	Significance	Iterations
0	Cent.							
1	0.537	0.537	19.3	0.507	0.0378	0.507	R1	9
2	0.135	0.672	4.87	0.182	0.0387	0.597	R1	75
3	0.122	0.794	4.39	0.324	0.0396	0.728	R1	13
4	0.0618	0.856	2.23	0.237	0.0406	0.792	R1	30
5	0.0417	0.897	1.5	0.209	0.0417	0.835	R1	14

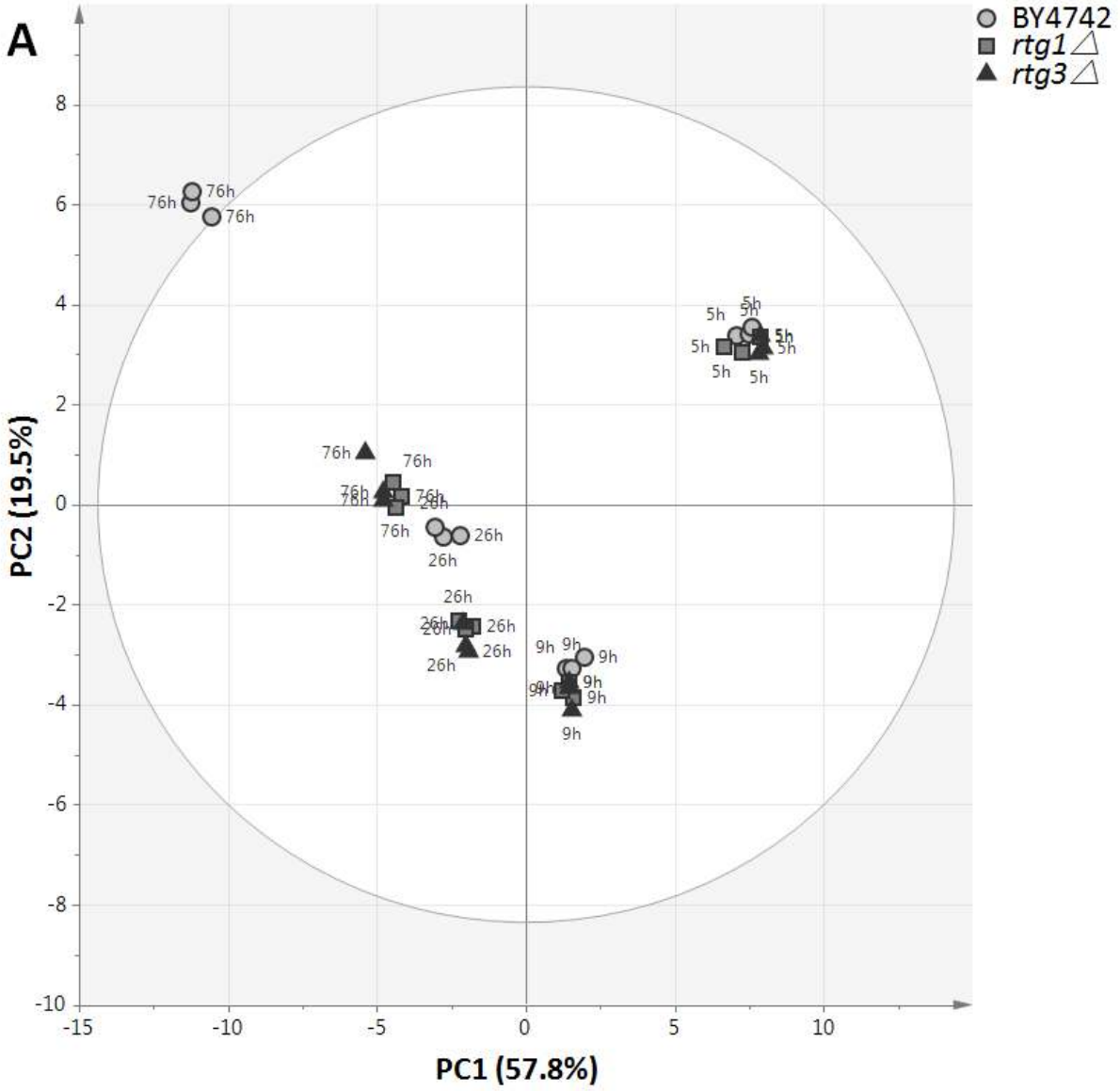
Supplementary Table S3. Loading values on principal component 1 (PC1) and principal component 2 (PC2) for each metabolite (intracellular: from the yeast extract and extracellular: from the growth medium).

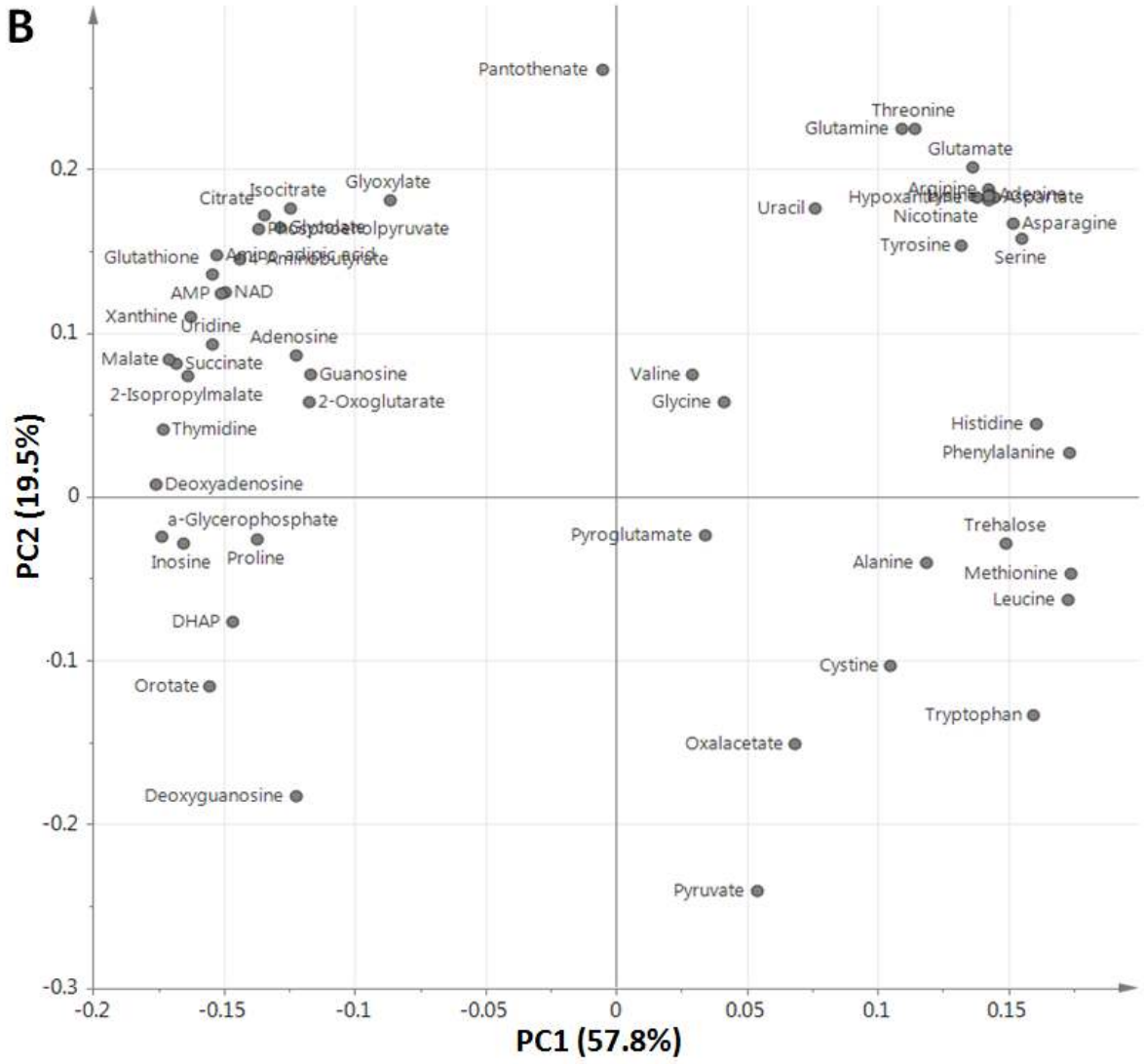
Metabolite (intracellular)	PC1 Loading	PC2 Loading	Metabolite (extracellular)	PC1 Loading	PC2 Loading
2-Isopropylmalate	-0.00209	0.22745	2-Isopropylmalate	-0.16400	0.07362
2-Oxoglutarate	0.00588	0.24828	2-Oxoglutarate	-0.11752	0.05733
3Phosphoglycerate	0.11976	0.00749	4-Aminobutyrate	-0.14398	0.14484
4-Aminobutyrate	-0.11829	0.05978	Adenine	0.14294	0.18411
Acetyl-coA	0.13741	0.01074	Adenosine	-0.12229	0.08622
Adenine	0.09936	-0.04542	a-Glycerophosphate	-0.17368	-0.02441
Adenosine	-0.10064	-0.03844	Alanine	0.11893	-0.04041
ADP	0.09370	0.03354	Amino adipic acid	-0.15262	0.14734
a-Glycerophosphate	0.13798	-0.00768	AMP	-0.15113	0.12424
Alanine	0.01347	-0.06582	Arginine	0.14229	0.18760
Amino adipic acid	-0.07215	0.22560	Asparagine	0.15189	0.16691
AMP	-0.10022	-0.03836	Aspartate	0.14460	0.18242
Arginine	0.11712	0.08488	Citrate	-0.13431	0.17177
Asparagine	0.10135	-0.04259	Cystine	0.10505	-0.10305
Aspartate	0.12016	-0.06230	Deoxyadenosine	-0.17569	0.00762
ATP	0.13473	0.05739	Deoxyguanosine	-0.12206	-0.18288
b-Alanine	0.02704	0.14669	DHAP	-0.14678	-0.07630
Bisphosphoglycerate	0.11485	0.07410	Glutamate	0.13630	0.20084
cAMP	0.10106	-0.13327	Glutamine	0.10919	0.22477
CDP	0.09910	-0.02309	Glutathione	-0.15461	0.13571
Citrate	-0.07668	0.20449	Glycine	0.04117	0.05761
Citrulline	0.07007	0.07924	Glycolate	-0.12834	0.16408
CMP	-0.08741	-0.08178	Glyoxylate	-0.08650	0.18133
CTP	0.13086	0.05169	Guanosine	-0.11670	0.07474
Cysteine	-0.06774	0.21267	Histidine	0.16084	0.04404

Cystine	0.13480	-0.02100	Hypoxanthine	0.13818	0.18256
Cytidine	-0.07314	0.01314	Inosine	-0.16559	-0.02855
Deoxyadenosine	-0.08935	0.05245	Isocitrate	-0.12434	0.17564
Deoxycytidine	0.07273	-0.02977	Leucine	0.17279	-0.06298
Deoxyguanosine	0.10825	-0.03625	Lysine	0.14266	0.18386
DHAP	0.13796	-0.00388	Malate	-0.17103	0.08345
FAD	0.11104	0.03701	Methionine	0.17403	-0.04695
Fructose 1,6-bisphosphate	0.13499	-0.01370	NAD	-0.14924	0.12477
Fructose 1-phosphate	0.13737	-0.01639	Nicotinate	0.14247	0.18098
Fructose 2,6-bisphosphate	0.12608	-0.00098	Orotate	-0.15549	-0.11610
Fructose 6-phosphate	0.13058	0.03167	Oxalacetate	0.06816	-0.15088
Fumarate	0.12440	0.07301	Pantothenate	-0.00519	0.26084
Glucose 6-phosphate	0.13294	0.04489	Phenylalanine	0.17340	0.02699
Glutamate	0.11327	0.07720	Phosphoenolpyruvate	-0.13665	0.16323
Glutamine	0.08882	0.13960	Proline	-0.13749	-0.02614
Glutathione	0.11373	0.10429	Pyroglutamate	0.03414	-0.02368
Glycerate	0.08263	0.14066	Pyruvate	0.05378	-0.24100
Glycine	0.08802	0.15103	Serine	0.15543	0.15707
Glycolate	-0.04132	0.23183	Succinate	-0.16794	0.08129
Glyoxylate	0.00639	0.24135	Threonine	0.11451	0.22466
GMP	-0.08884	-0.07027	Thymidine	-0.17320	0.04120
GTP	0.13770	0.00852	Trehalose	0.14937	-0.02877
Guanine	0.07134	0.00928	Tryptophan	0.15976	-0.13349
Guanosine	-0.06730	0.00043	Tyrosine	0.13203	0.15300
Histidine	-0.02338	0.23441	Uracil	0.07631	0.17593
Homoserine	0.10027	-0.05115	Uridine	-0.15427	0.09296
Hydroxyproline	0.09716	0.06096	Valine	0.02936	0.07420
Hypoxanthine	0.09569	-0.02078	Xanthine	-0.16286	0.10989
Inosine	0.09335	0.00044			
Isocitrate	-0.06356	0.22648			
Isoleucine	0.12724	-0.02595			
Leucine	0.12801	-0.04393			
Lysine	0.11374	-0.04160			
Malate	0.01816	0.22764			
Methionine	0.12383	-0.05087			
NAD	0.09416	0.05793			
NADH	0.12883	0.01969			
NADP	0.09754	0.03847			
NADPH	0.13540	0.02288			
Nicotinate	0.09883	-0.04632			
Ornithine	-0.03392	-0.12659			
Orotate	0.05678	0.13618			
Oxalacetate	0.12304	-0.02171			
Pantothenate	0.09640	0.01449			
Phenylalanine	0.13560	-0.04245			
Phosphoenolpyruvate	0.12118	0.03853			
Proline	-0.04735	-0.11264			
Putrescine	0.02963	-0.16010			

Pyroglutamate	0.13187	0.04943
Pyruvate	0.12589	-0.04248
Ribose 5-phosphate	0.11655	0.05339
Ribulose 5-phosphate	0.12525	0.04431
S-Adenosylmethionine	-0.10702	0.05528
Sedoheptulose 7-phosphate	0.13532	0.02000
Serine	0.11174	-0.02229
Spermidine	-0.06151	-0.08182
Succinate	-0.08868	0.17453
Threonine	0.05368	-0.12654
Thymidine	-0.12763	0.06990
TMP	-0.08416	-0.04919
Trehalose	-0.06462	0.16862
Tryptophan	0.12819	-0.08465
Tyrosine	0.10120	0.02960
UDP	0.01583	0.04811
UDP-glucose	0.08317	0.17915
UMP	-0.10538	-0.05556
Uracil	0.12805	-0.01441
Uridine	-0.10055	0.07235
UTP	0.13373	0.03512
Valine	0.10324	0.10526
Xanthine	-0.10614	0.15427

Supplementary Fig. S1. (A) PCA score plot for time-course extracellular metabolic profiling (from growth media) of wild-type strain BY4742, and *rtg1*Δ and *rtg3*Δ disruptants (n=3). The metabolites were scaled to unit variance. Ellipse indicate 95% confidence border based on Hotelling's T^2. **(B)** The corresponding loading plot illustrating metabolites that contribute to the separation on PC1 and PC2 (see Supplementary Table S3 for the loading values).

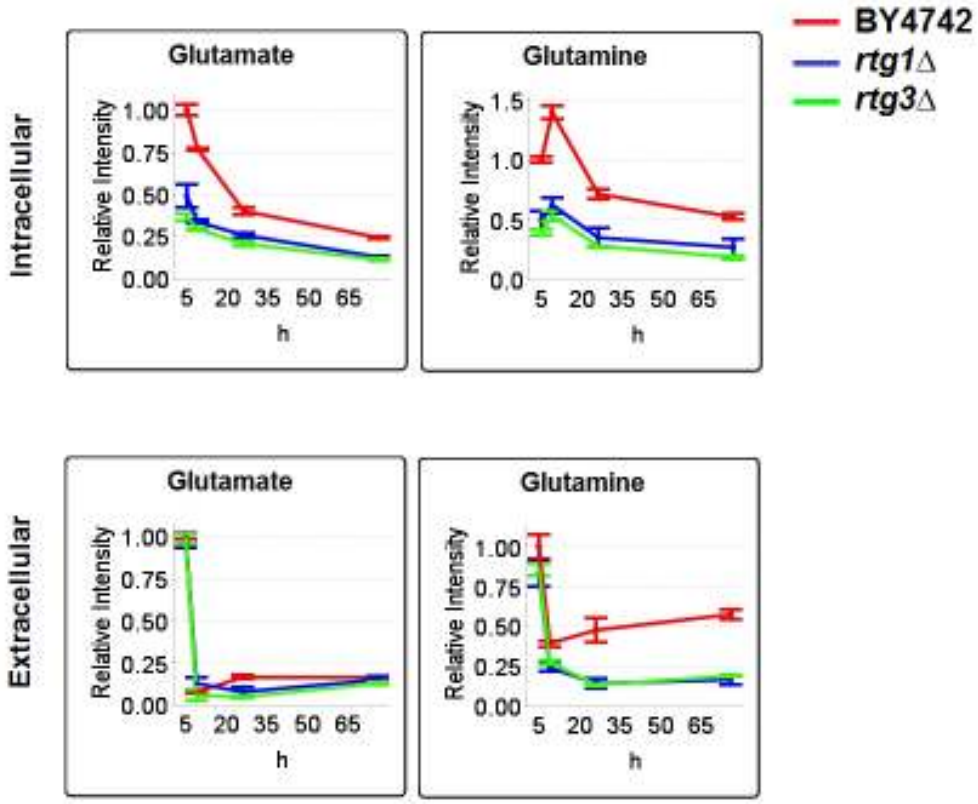




Supplementary Table S4. Cell growth on YPD (1% yeast extract, 2% peptone, 2% dextrose, 2% agar (% *w/v*)) and YPG (1% yeast extract, 2% peptone, 2% glycerol, 2% agar (% *w/v*)) plates, expressed as the number of colonies. Cell cultures were diluted to approximately 10^3 cells/ mL, 100 µL were spread on YPD or YPG plates, and the colony number was counted after 2-4 days. Measurement was done in duplicate (separated by a comma) for each sampling point.

Strain	Plate	Day 1	Day 3	Day 5
BY4742	YPD	86, 115	100, 100	68, 53
	YPG	81, 151	96, 112	66, 54
*rtg1*Δ disruptant	YPD	53, 52	31, 34	25, 28
	YPG	61, 62	39, 43	29, 28
*rtg3*Δ disruptant	YPD	23, 16	27, 45	6, 8
	YPG	28, 15	25, 40	10, 11

Supplementary Fig. S2. Intracellular (from cell extracts) and extracellular (from the growth medium) concentrations of glutamate and glutamine in BY4742, and *rtg1*Δ and *rtg3*Δ mutants (n = 3). *Y*-axis indicates relative intensity while *x*-axis indicates time. The metabolite intensities were relative to that of BY4742 at time 5 h.



Supplementary Table S5. List of disruptant strains used in this study, maximum specific growth rate μ (h^{-1}) in synthetic complete medium and adenylate energy charge (EC). The mutants are isogenic derivatives of wild-type BY4742 (*MATα leu2Δ0 lys2Δ0 ura3Δ0 his3Δ1*).

Strain	Batch	μ (/h), average	μ (/h), stdev	EC (-), average	EC (-), stdev
BY4742(1)	1	0.47	0.04	0.75	0.02
*ace2*Δ	1	0.48	0.04	0.81	0.02
*arg80*Δ	1	0.49	0.05	0.72	0.06
*arg82*Δ	1	0.39	0.05	0.84	0.04
*aro80*Δ	1	0.52	0.03	0.76	0.02
*bas1*Δ	1	0.55	0.05	0.72	0.05
*bas2/grf10/pho2*Δ	1	0.51	0.02	0.76	0.03
*cad1/yap2*Δ	1	0.50	0.05	0.75	0.04
*dal80*Δ	1	0.52	0.04	0.78	0.03
*dal81*Δ	1	0.52	0.04	0.76	0.03
*ecm22*Δ	1	0.49	0.04	0.78	0.02
*fzf1*Δ	1	0.48	0.06	0.76	0.04
*gcn4*Δ	1	0.51	0.02	0.74	0.06
*gln3*Δ	1	0.50	0.06	0.77	0.01
*gzf3*Δ	1	0.47	0.04	0.74	0.04
*hap2*Δ	1	0.47	0.04	0.78	0.02
*hap3*Δ	1	0.50	0.05	0.76	0.04
*hap4*Δ	1	0.50	0.01	0.74	0.05
*hap5*Δ	1	0.52	0.08	0.79	0.00
*ime1*Δ	1	0.50	0.05	0.76	0.04
*ino2*Δ	1	0.30	0.02	0.86	0.00
*ino4*Δ	1	0.28	0.03	0.83	0.02
*ixr1*Δ	1	0.45	0.01	0.74	0.02
*leu3*Δ	1	0.53	0.01	0.74	0.05
*lys14*Δ	1	0.53	0.02	0.74	0.06
*met28*Δ	1	0.53	0.06	0.77	0.05
*met31*Δ	1	0.52	0.06	0.79	0.01
*mks1*Δ	1	0.46	0.04	0.74	0.00
*mot3*Δ	1	0.56	0.04	0.79	0.03
*oaf1*Δ	1	0.48	0.07	0.76	0.04
*opi1*Δ	1	0.24	0.01	0.83	0.02
*pho23*Δ	1	0.43	0.01	0.76	0.01
*pho4*Δ	1	0.51	0.03	0.79	0.02
*ric1*Δ	1	0.37	0.03	0.75	0.01
*rpn4*Δ	1	0.41	0.02	0.83	0.02
*sin3*Δ	1	0.41	0.02	0.72	0.04
*stp1*Δ	1	0.52	0.02	0.74	0.06
*stp2*Δ	1	0.45	0.09	0.78	0.02
*thi2*Δ	1	0.54	0.06	0.77	0.02

*uga3*Δ	1	0.52	0.02	0.75	0.03
*yap6*Δ	1	0.56	0.02	0.78	0.02
BY4742(2)	2	0.45	0.05	0.84	0.01
*adr1*Δ	2	0.43	0.06	0.85	0.02
*aft2*Δ	2	0.46	0.03	0.86	0.01
*arr1*Δ	2	0.45	0.03	0.83	0.02
*cat8*Δ	2	0.47	0.02	0.85	0.01
*cha4*Δ	2	0.49	0.02	0.84	0.02
*cin5*Δ	2	0.49	0.02	0.82	0.02
*crz1*Δ	2	0.49	0.02	0.84	0.02
*gcr2*Δ	2	0.35	0.05	0.77	0.02
*hac1*Δ	2	0.45	0.04	0.86	0.02
*hal9*Δ	2	0.50	0.02	0.84	0.01
*mig2*Δ	2	0.49	0.02	0.86	0.01
*msn1*Δ	2	0.49	0.07	0.84	0.02
*msn2*Δ	2	0.51	0.07	0.80	0.04
*msn4*Δ	2	0.48	0.04	0.84	0.01
*nrg1*Δ	2	0.50	0.01	0.84	0.02
*pdr1*Δ	2	0.47	0.04	0.85	0.02
*pdr3*Δ	2	0.42	0.05	0.85	0.02
*pdr8*Δ	2	0.47	0.03	0.80	0.03
*ppr1*Δ	2	0.48	0.04	0.85	0.01
*rgt1*Δ	2	0.49	0.05	0.82	0.05
*rim101*Δ	2	0.44	0.04	0.84	0.02
*sfl1*Δ	2	0.49	0.05	0.84	0.01
*sip4*Δ	2	0.49	0.02	0.83	0.02
*skn7*Δ	2	0.49	0.03	0.84	0.02
*sko1*Δ	2	0.48	0.04	0.85	0.01
*smp1*Δ	2	0.49	0.02	0.86	0.03
*stb5*Δ	2	0.45	0.06	0.83	0.01
*sut1*Δ	2	0.45	0.04	0.84	0.04
*swi6*Δ	2	0.40	0.01	0.84	0.02
*tye7*Δ	2	0.45	0.03	0.84	0.02
*ume6*Δ	2	0.45	0.03	0.85	0.02
*usv1*Δ	2	0.46	0.02	0.83	0.03
*wtm2*Δ	2	0.48	0.06	0.83	0.04
*xbp1*Δ	2	0.46	0.05	0.84	0.03
*yap1*Δ	2	0.49	0.03	0.80	0.03
*yap3*Δ	2	0.47	0.05	0.81	0.03
*yap5*Δ	2	0.43	0.02	0.85	0.02
*yap7*Δ	2	0.49	0.09	0.82	0.07
*yrm1*Δ	2	0.49	0.02	0.87	0.01
*yrr1*Δ	2	0.47	0.01	0.83	0.01
BY4742(3)	3	0.48	0.03	0.83	0.01

*ace1/cup2*Δ	3	0.52	0.01	0.84	0.01
*ada2*Δ	3	0.34	0.01	0.88	0.01
*ada3*Δ	3	0.41	0.01	0.87	0.01
*aft1*Δ	3	0.28	0.05	0.82	0.01
*argr2*Δ	3	0.51	0.02	0.83	0.01
*azf1*Δ	3	0.57	0.04	0.82	0.01
*cbf1*Δ	3	0.37	0.04	0.84	0.01
*cst6*Δ	3	0.48	0.03	0.83	0.02
*cup9*Δ	3	0.51	0.02	0.83	0.02
*dal82*Δ	3	0.50	0.02	0.82	0.01
*gal3*Δ	3	0.52	0.01	0.83	0.01
*gal4*Δ	3	0.49	0.02	0.85	0.01
*gal80*Δ	3	0.52	0.02	0.83	0.02
*gat1*Δ	3	0.47	0.09	0.82	0.03
*gat2*Δ	3	0.53	0.06	0.82	0.02
*gat3*Δ	3	0.44	0.02	0.82	0.02
*gat4*Δ	3	0.43	0.04	0.81	0.03
*gis1*Δ	3	0.48	0.06	0.82	0.01
*mac1*Δ	3	0.46	0.03	0.81	0.01
*mal13*Δ	3	0.47	0.07	0.81	0.02
*mal33*Δ	3	0.53	0.04	0.82	0.00
*mga1*Δ	3	0.54	0.05	0.83	0.02
*mig1*Δ	3	0.43	0.06	0.81	0.02
*mig3*Δ	3	0.47	0.06	0.83	0.01
*not3*Δ	3	0.45	0.04	0.84	0.02
*put3*Δ	3	0.44	0.03	0.82	0.01
*rgm1*Δ	3	0.51	0.03	0.81	0.01
*rlm1*Δ	3	0.53	0.01	0.82	0.02
*rox1*Δ	3	0.50	0.03	0.82	0.01
*rsf2*Δ	3	0.45	0.03	0.82	0.01
*rtg1*Δ	3	0.45	0.02	0.80	0.02
*sas2*Δ	3	0.46	0.01	0.81	0.02
*sef1*Δ	3	0.44	0.01	0.84	0.02
*spt23*Δ	3	0.50	0.04	0.83	0.01
*ssn6*Δ	3	0.49	0.01	0.83	0.02
*sut2*Δ	3	0.50	0.01	0.84	0.00
*tea1*Δ	3	0.49	0.04	0.84	0.02
*tuf1*Δ	3	0.46	0.03	0.84	0.01
*zap1*Δ	3	0.48	0.06	0.81	0.00
BY4742(4)	4	0.45	0.03	0.76	0.04
*ash1*Δ	4	0.45	0.05	0.78	0.02
*ccr4*Δ	4	0.47	0.03	0.77	0.03
*ezl1*Δ	4	0.47	0.05	0.77	0.03
*fkh1*Δ	4	0.50	0.02	0.78	0.03

*fkh2*Δ	4	0.57	0.03	0.79	0.01
*flo8*Δ	4	0.49	0.01	0.80	0.00
*hcm1*Δ	4	0.48	0.09	0.79	0.04
*hir1*Δ	4	0.49	0.07	0.76	0.04
*hir2*Δ	4	0.48	0.02	0.81	0.01
*hir3*Δ	4	0.52	0.02	0.77	0.02
*hms1*Δ	4	0.61	0.02	0.79	0.02
*kar4*Δ	4	0.52	0.05	0.78	0.02
*mbp1*Δ	4	0.42	0.02	0.79	0.01
*ndt80*Δ	4	0.53	0.02	0.74	0.04
*phd1*Δ	4	0.54	0.03	0.79	0.01
*rfx1*Δ	4	0.54	0.03	0.75	0.08
*rph1*Δ	4	0.49	0.04	0.76	0.07
*rrn10*Δ	4	0.51	0.09	0.73	0.02
*rtg3*Δ	4	0.47	0.07	0.74	0.06
*sds3*Δ	4	0.56	0.07	0.77	0.05
*sir1*Δ	4	0.49	0.03	0.81	0.01
*sir2*Δ	4	0.55	0.07	0.78	0.02
*sir3*Δ	4	0.52	0.03	0.81	0.01
*sir4*Δ	4	0.54	0.06	0.80	0.02
*sum1*Δ	4	0.46	0.11	0.74	0.08
*swi4*Δ	4	0.47	0.05	0.77	0.02
*swi5*Δ	4	0.50	0.02	0.77	0.02
*tec1*Δ	4	0.48	0.07	0.80	0.02
*tsp1*Δ	4	0.50	0.05	0.77	0.03
*uaf30*Δ	4	0.49	0.04	0.78	0.01
*wtm1*Δ	4	0.54	0.07	0.77	0.02
*yhp1*Δ	4	0.52	0.03	0.78	0.04
*yox1*Δ	4	0.52	0.03	0.79	0.01
*zds1*Δ	4	0.50	0.03	0.78	0.03
*zds2*Δ	4	0.51	0.01	0.79	0.02

Supplementary Table S6. Summary of PCA with eleven significant components performed for metabolome dataset from metabolic profiling of 154 deletion strains.

Component	R2X	R2X (cum)	Eigen value	Q2	Limit	Q2 (cum)	Significance	Iterations
0	Non-Cent.							
1	0.366	0.366	11.5	0.252	0.0118	0.252	R1	16
2	0.173	0.539	18.1	0.155	0.0119	0.368	R1	12
3	0.0696	0.608	8.25	0.00218	0.012	0.37	R2	52
4	0.0576	0.666	6.8	0.07	0.0122	0.414	R1	37
5	0.0425	0.708	4.92	-0.035	0.0123	0.393	R2	52
6	0.036	0.744	4.24	0.0349	0.0125	0.414	R1	40
7	0.0303	0.775	3.58	0.0432	0.0127	0.44	R1	60
8	0.0241	0.799	2.83	-8.74×10^{-5}	0.0128	0.44	R2	46
9	0.0203	0.819	2.39	-0.00329	0.013	0.438	R2	117
10	0.0186	0.838	2.21	0.0157	0.0132	0.447	R1	70
11	0.0158	0.853	1.87	-0.0157	0.0133	0.438	R2	39

Supplementary Table S7. List of metabolites and the number of deletion strains (out of 154) that were significantly changed or unchanged compared with wild-type BY4742 (p-value < 0.05).

	No. of strains			No. of strains	
Metabolite	Changed	Unchanged	**Metabolite**	Changed	Unchanged
Arginine	111	43	Succinate	63	91
Histidine	115	39	Fumarate	4	150
Serine	15	139	cAMP	7	147
Asparagine	5	149	Malate	44	110
Glutamine	21	133	UDP-Glucose	59	95
Homoserine	25	129	2-Oxoglutarate	49	105
Threonine	41	113	CDP	16	138
Trehalose	3	151	GDP	3	151
Proline	23	131	NADP	7	147
Valine	8	146	F2,6P	9	145
Methionine	5	149	F1,6P	8	146
Guanine	126	28	Isocitrate	16	138
Isoleucine	12	142	Citrate	19	135
Tyrosine	5	149	ADP	9	145
Amino adipic acid	22	132	1,3-BPG	15	139
Glutamate	16	138	Phosphoenolpyruvate	12	142
Aspartate	16	138	FMN	10	144
Inosine	135	19	2-Isopropylmalate	32	122
Guanosine	132	22	FAD	3	151
Phenylalanine	3	151	CTP	45	109
Pyroglutamate	2	152	GTP	14	140
Glucose 6-phosphate	11	143	NADPH	5	149
Ribose 5-phosphate	19	135	UTP	54	100
Sedoheptulose 7P	21	133	ATP	14	140
Fructose 6-phosphate	2	152	Acetyl CoA	20	134
Tryptophan	14	140	Alanine	12	142
α-Glycerophosphate	39	115	Glycine	7	147
Glutathione	7	147	2-Aminoethanol	5	149
GAP	11	143	Urea	3	151
Ribulose 5-phosphate	18	136	Phosphate	13	141
Orotate	32	122	Glycerol	27	127
Fructose 1-phosphate	19	135	Leucine	6	148
CMP	10	144	Uracil	6	148
NAD	8	146	Lysine	92	62
Pyruvate	27	127	Adenine	45	109
DHAP	14	140	Inositol	9	145
UMP	24	130	Octadecanoate	4	150
GMP	11	143	Cysteine+Cystine	6	148
Oxalacetate	29	125	Citrulline	48	106
AMP	14	140	Glucose	15	139
Nicotinate	40	114	β-Lactose	14	140
Pantothenate	26	128	Melibiose	2	152

List of publications

Conferences

1) Zanariah Hashim, Takeshi Bamba and Eiichiro Fukusaki. "Metabolome on stress-free condition involves stress tolerance of *S. cerevisiae*" (Oral). Society for Biotechnology Japan (SBJ) 63rd Annual Meeting, Tokyo, Japan. 26-28 September 2011.

2) Zanariah Hashim, Yudai Dempo, Tairo Ogura, Ichiro Hirano, Junko Iida, Takeshi Bamba and Eiichiro Fukusaki. "Quantitative analysis of hydrophilic metabolite using ionpairing chromatography with a high-speed triple quadrupole mass spectrometer" (Poster). 60th American Society for Mass Spectrometry (ASMS) Conference on Mass Spectrometry and Allied Topics, Vancouver, Canada. 20-24 May 2012.

3) Zanariah Hashim, Yudai Dempo, Tairo Ogura, Ichiro Hirano, Takeshi Bamba and Eiichiro Fukusaki. "Development of accelerate quantification analysis for hydrophilic metabolites using ionpairing chromatography with a high-speed triple quadrupole mass spectrometer" (Poster). 19th International Mass Spectrometry Conference (IMSC 2012), Kyoto, Japan. 15-21 September 2012.

4) Zanariah Hashim, Shao Thing Teoh, Takeshi Bamba and Eiichiro Fukusaki. "Metabolic profiling of *Saccharomyces cerevisiae* transcription factor-related single gene mutants" (Oral). Biological and Chemical Methods for Selective Catalysis Symposium (SeleCa 2012), RWTH Aachen, Germany. 2-8 December 2012.

5) Zanariah Hashim, Shao Thing Teoh, Takeshi Bamba and Eiichiro Fukusaki. "Metabolomics-based screening of transcription factor deletion strains reveals discrimination of RTG-related genes in yeast" (Poster). Society for Biotechnology Japan (SBJ) 65th Annual Meeting, Hiroshima, Japan. 18-20 September 2013.

6) Zanariah Hashim, Shao Thing Teoh, Takeshi Bamba and Eiichiro Fukusaki. "Construction of a metabolome library for transcription factor-related single gene mutants of *Saccharomyces cerevisiae*" (Oral -**selected for 2014 Metabolomics Society Student Travel Award**). 10th International Conference of the Metabolomics Society, Tsuruoka, Japan. 23-26 June 2014.

Original Articles

1) Zanariah Hashim, Yukio Mukai, Takeshi Bamba and Eiichiro Fukusaki. Metabolic profiling of retrograde pathway transcription factors Rtg1 and Rtg3 knockout yeast. *Metabolites* **4**, 580-598 (2014).

2) Zanariah Hashim, Shao Thing Teoh, Takeshi Bamba and Eiichiro Fukusaki. Construction of a metabolome library for transcription factor-related single gene mutants of *Saccharomyces cerevisiae*. *J. Chromatogr. B* **966**, 83-92 (2014).

Acknowledgment

I sincerely thank Prof. Eiichiro Fukusaki for introducing me to the wonderful world of metabolomics and for his guidance and supervision. I thank my thesis committee Prof. Satoshi Harashima and Prof. Kazuhito Fujiyama for their critical comments and valuable advice for my thesis. I also thank Assoc. Prof. Yukio Mukai from Nagahama Institute of Bio-Science and Technology for helpful discussions and teaching me about yeasts. I also acknowledge Assoc. Prof. Takeshi Bamba and Assist. Prof. Hisayo Ono for their assistance in the laboratory.

Special thanks to the Hitachi Scholarship Foundation (HSF), which provided me financial support for five years for my Masters and Doctorate degrees. I thank previous and current members of HSF, Mr. Sasamori, Mr. Miyanaga, Dr. Homma, Mr. Miyake, Mr. Kawamoto, Ms. Nunokami, Ms. Masuda, Ms. Namiki and Ms. Kimura, for their tireless support and hospitality. I also appreciate friendship and moral support from fellow HSF recipients from Malaysia, Thailand, Indonesia, Philippines and Vietnam.

To my employer Universiti Teknologi Malaysia, especially the staff at the Faculty of Chemical Engineering and Human Capital Development for assistance during my study leave and financial support.

To Japanese-German Graduate Externship Program and the members from Osaka University and RWTH Aachen, for the chance to visit Germany and participate in a graduate exchange program.

To all my lab mates, especially Shao, Udi, An, Hanghang, Jae, Yang, Nagasawa-san, Dempo, Mitsunaga, Noguchi, Teruko, Risa and Dr. Tsugawa for making my lab life cheerful and keeping me positive. I am very grateful to be able to work at a laboratory with state-of-the-art equipment and colleagues who are always eager to help, both in studies and personal matters. Also, thanks to my Malaysian friends in Japan for sharing food and ‘home’ stories, visiting me when I feel sick and keeping me company.

Finally, my heartfelt appreciation goes to my family: my parents, husband and son, who always believe in me, even when I myself have little faith. You taught me patience and the most important thing in life: family. I am forever thankful to God for His guidance and blessings.