מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

Final Report

דוח מסכם

By
Ilya Venger

מאת
איליה ונגר

עמידות שמר האפייה למוטציות מחיקה ותגובה לשינויים בסביבה דרך עדשת מטבולומיקה

Robustness of *Saccharomyces cerevisiae* to deletion mutations and response to environmental perturbations through the lens of large-scale metabolic profiling

Advisors:
Prof. Yitzhak Pilpel
Dr. Asaph Aharoni

מנחים:
פרופ' יצחק פלפל
ד"ר אסף אהרוני

Month and Year
April 2009

חודש ושנה עבריים
ניסן תשס"ט

Table of contents

Introduction

1.1 General

It was previously shown that ~80% of the genes in the yeast *Saccharomyces cerevisiae* are not essential for the survival of the organism when deleted[1, 2]. Three different mechanisms were suggested to be responsible for this observed resilience to deletion mutations. Namely, genetic redundancy, distributed robustness and conditional non-functionality. Where: *genetic redundancy* is a case in which a paralog or an analog of the deleted gene may provide functional compensation for the genetic deficiency and restore a viable phenotype[3-5], *distributed robustness* is a mode where the compensation for the lost function is carried out through the rearrangement of portions of the entire metabolic network[4, 6], and *conditional non-functionality* is a case when the deleted gene is essential only in a particular subset of conditions that were not tested in the deletion studies; thus, a gene's dispensability is observed only due to lack of an appropriate experiment[5-7].

In general, the complement of the biological molecules inside the cells may be divided to four levels. The genome, the transcriptome, the proteome and the metabolome[8]; the metabolome being the set of all the intra-cellular metabolites in a particular condition. The metabolome can be seen as the uppermost "omics" level since it is affected by the mRNA and protein concentrations, and is the most basic manifestation of the cellular phenotype.

The intracellular metabolites participate in almost all reactions in the living cell. The level of the metabolites is directly affected by deletion of the enzymes that carry out reactions with the metabolites as substrates and products. Therefore, one can expect a direct connection between the metabolome of the different *S. cerevisiae* strains and their resistance to deletion mutations.

In addition to being used as a model for genetic perturbations yeast has been employed in studies of the effect of various environmental stresses on living cells. Multiple experiments have been performed to assess the effect of external stresses on yeast transcriptome and proteome (ref…). Understanding the manifestation of the stress-response on a temporal scale on the level of metabolites is paramount in the understanding of development of basic cellular responses to the environment.

1.2 Research main goals and motivation

In my research I have set out to determine the cellular manifestation on the molecular level of yeast function when faced with either genetic stress (gene deletion) or environmental perturbation (ethanol stress).

I have chosen to analyze how *genetic redundancy* affects the cellular metabolome by analyzing yeast gene deletion strains. In addition, I have explored the relationship between duration and intensity of ethanol stress in terms of yeast intracellular metabolic response.

Yeast paralogs tend to be dispensable upon deletion more often than singleton genes (ref.). Prior work in our lab has shown that paralogous pairs regulate the expression of each other on the transcriptional level, i.e. transcriptional reprogramming occurs (ref.). Further studies (ref. to Kafri articles) have outlined the natural selection forces that acted on paralogous genes to retain a certain degree of redundant functionality.

As the metabolome is the most direct measurement of the cellular phenotype we expect that the reprogramming phenomenon as well as natural selection towards partial functional redundancy to be manifested when measuring the metabolome and allow us to better understand these phenomena.

Ethanol stress response is one of the most prominent cellular reactions to environmental perturbations in *S. cerevisiae*. Yeast cells produce ethanol during fermentation and have developed mechanisms to cope with elevated concentrations of ethanol in the surrounding environment. One of the major questions in the development of the stress response in yeast is the relationship between the duration of the stress and its intensity. The metabolome is both affected by the rising ethanol concentrations and is one of the major effectors in the protection of cells from stress. Observing metabolic response of cells to ethanol may shed light on the time-scale of cellular response and on the modality of stress; i.e. whether the response to stress is concentration-dependent or not.

1.3 Metabolomics in Saccharomyces cerevisiae

It was shown that there are about 550 metabolites in *S. cerevisiae*[9, 10]. Any single intracellular metabolomics method can identify and quantify 70-80 of these metabolites.

Most of the research up until now was focused on the development of methods for the metabolic profiling of yeast strains. There are several research groups that study the metabolomics of *S. cerevisiae*[11-13] both on the level of the intra-cellular metabolome and on the

level of the extracellular metabolome. Using metabolomics information it was shown that metabolic profiles discriminate among various yeast strains[14] and growth conditions[11, 13]. However no study has studied the response of yeast cells to gene deletion in a large scale manner specifically focusing on the response to paralogous gene deletions or to ethanol stress in an ordered fashion.

2 Summary of Results

2.1 Analytical Method Development

To analyze the yeast metabolome in a large scale manner we have set-up a method that would allow both fast and robust metabolomic analysis of multiple yeast strains. Several methods exist that allow the metabolic profiling of *S. cerevisiae*[11, 15-17], however, none of the published methods in their original form provided satisfactory results in our experimental settings. Existing protocols were either too laborious for carrying out a large scale study, were not comprehensive enough, i.e. did not cover enough metabolites, or resulted in high noise levels. The main parameters to control for in the experiments included robustness of the analytical method and repeatability between biological replicates.

The whole experimental flow for data acquisition and analysis is shown in Figure 1. The numbers 1-6 above the different experimental stages are described respectively in subparagraphs 2.1.1 to 2.1.6.

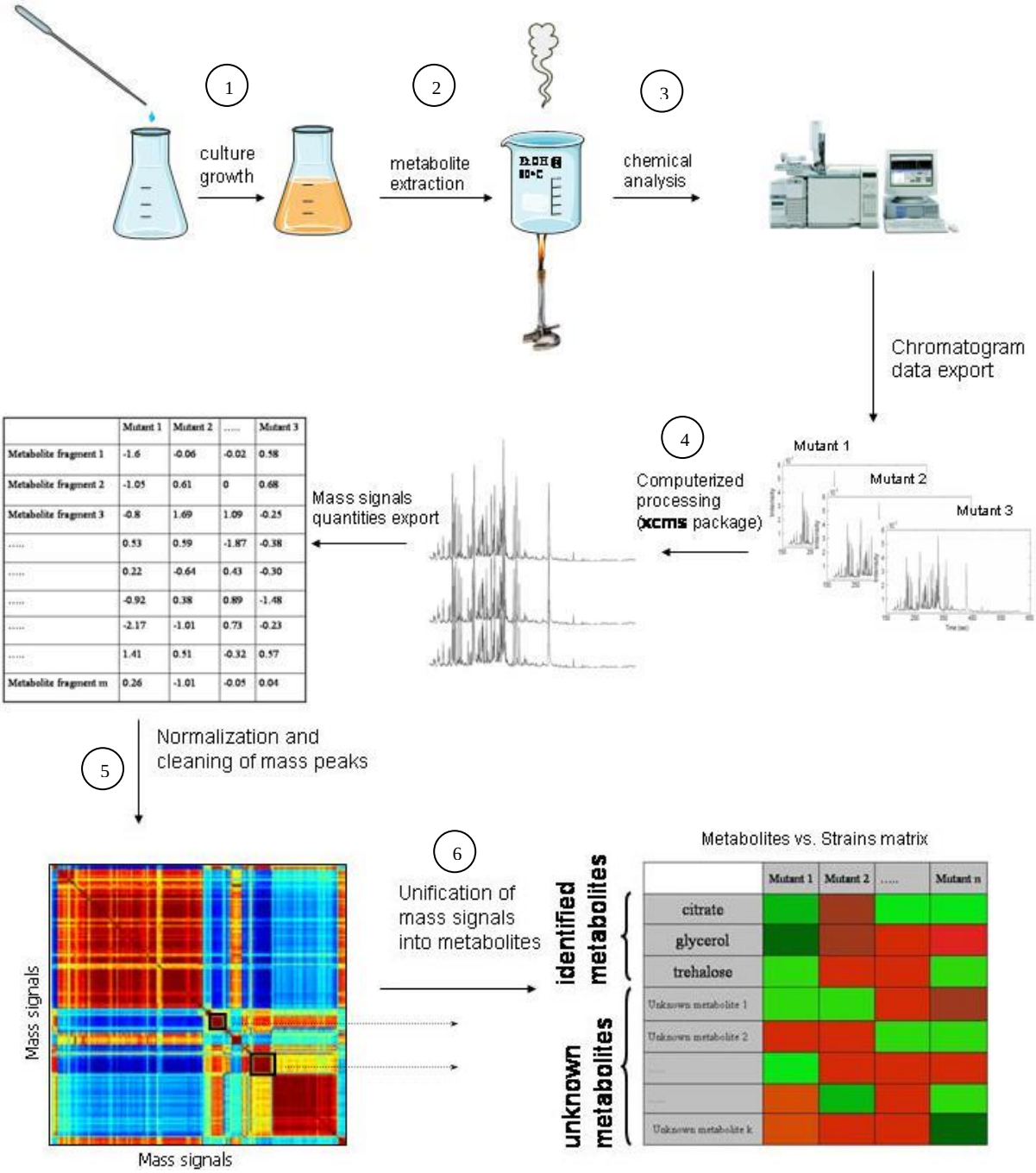

Figure 1 - The workflow for sample preparation, chemical analysis and raw data analysis. Stages marked with numbers 1-6 are described in detail in sub-paragraphs 2.1.1 to 2.1.6 respectively.

2.1.1 Controlled growth of biological samples

Many of the differences in the intracellular metabolome depend upon growth conditions and growth rate[18]. Therefore, uneven growth of cultures was one of the main caveats when analyzing yeast metabolome. To ensure robust results when analyzing multiple samples we developed a protocol to ensure controlled cell growth in all cultures.

Additionally, there was need to minimize the footprint of the culture medium on the results. We used minimal defined mineral medium[19] supplemented with necessary amino acids for culturing. Ammonium sulfate was used as the nitrogen source and glucose as the carbon source.

To grow the cultures to equal cell density and harvest them in similar growth stage the sampling procedure for each experiment was as follows:

a) 24 hours before sampling a 5 ml starter cultures were inoculated from frozen cultures.
b) After ~4 hours OD was measured for all starters and starters were diluted such that in ~18 hours OD = 0.7. Growth rate estimate of 2.35 hours/generation was used (wild type (WT) growth rate in mineral medium).
c) 18 hours later OD was measured and growth rate was calculated for each mutant. Samples were diluted such that in ~2 hours OD reached 0.75. This step ensured that all strains were harvested the same growth stage.
d) After 2 hours samples were harvested for metabolite extraction as described in section 2.1.2.

From Figure 2 we can observe that the final variation in optical density between the cultures was small (relative standard deviation of ~11%).

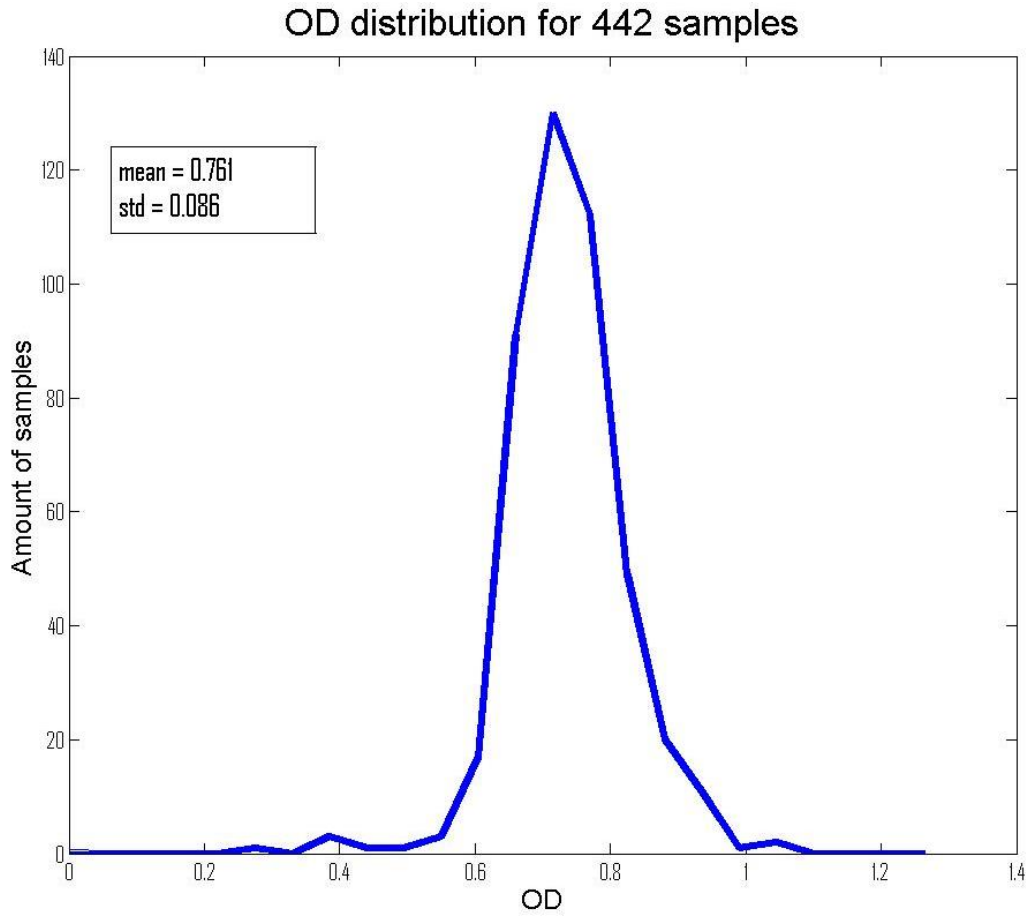

Figure 2 - Distribution of the final OD values before harvesting for 442 samples taken for the analysis of paralogous backup (described in paragraph Error! Reference source not found.**). The final mean optical density over 442 cultures was 0.761±0.086**

2.1.2 Extraction of metabolites from the cells

Cell quenching and extraction protocol was mainly based on the protocol developed by Castrillo *et al*[15] with the main modification being the pH buffering agent. The original pH buffering agent (tricine) was seen as an overloaded peak in chemical analysis as we carried it out. We checked several extraction and quenching conditions differing in pH buffering conditions (tricine buffered, ammonium acetate buffered and non-buffered). Conditions that had the least variability were those in which the quenching solution was buffered with 100mM ammonium acetate. Ammonium acetate buffer is volatile, and evaporates during sample drying leaving little impact on the samples. This protocol was recently independently shown[20] to be very effective for large-scale sampling of *S. cerevisiae*. The volume of quenching was reduced to 18ml of yeast culture to allow for higher throughput in the experiments. The final protocol allowed for 12 samples to be harvested at once. This was an improvement over the more laborious protocol for 6 samples employed by Castrillo *et al*[15].

Briefly, cells were quenched in buffered methanol @ -40ºC, centrifuged and washed with cold methanol to remove traces of medium. Subsequently, samples were extracted in boiling ethanol, followed by volume reduction in speed-vac dryer (Savant). Finally, samples were freeze-dried in lyophilizer over-night and kept @ -80ºC until chemical analysis.

2.1.3 Chemical analysis of metabolites

For chemical analysis of metabolites, gas chromatography – mass spectrometry (GC-MS) was carried out. This method was chosen due to the wide range of the metabolites it allows to detect, high degree of separation between metabolites, the high precision in quantification, the availability of the equipment and the relatively low costs per sample. Other methods for analysis of primary metabolites were considered (specifically liquid chromatography – mass spectrometry), however, the absence of readily-available equipment and protocols has stopped us from pursuing this direction further.

The GC-MS system was composed of a COMBI PAL autosampler (CTC analytics), a Trace GC Ultra gas chromatograph equipped with a PTV injector, and a DSQ™ II quadrupole mass spectrometer (ThermoElectron).

The protocol for the analysis followed established techniques in our lab[21-23]. Samples were injected into the GC-MS following methoxymation by Methoxy amine HCl and derivatization with N-Methyl trimethyl silyl trifluoro acetamid (MSTFA)[24]. Methoxymation causes the opening of sugar rings, and MSTFA is an effective trimethylsilyl donor which reacts to replace labile hydrogens on a wide range of polar compounds with a $-Si(CH3)_3$ group. MSTFA produces volatile and thermally stable derivatives.

2.1.4 Computerized analysis of chromatograms

After chemical analysis our goal was to automatically analyze the abundance of mass signals in the data. The goals of such an analysis are: robust peak detection in chromatograms, alignment of peaks in different chromatograms and peak integration. We have experimented with several software suits (MZmine[25], XCalibur (ThermoFinnigan) and xcms[26]) that performed the above actions. Xcms produced the best results for automatic analysis among the three, as it allows relatively fast processing times, has a very precise quantitation algorithm and highly sensitive detection of compounds.

Pre-processing of multiple chromatograms was performed using xcms package v.1.14 for R v.2.8 programming language. Xcms allows multiple chromatogram alignment, signal extraction and quantitation of mass signals. Mass quantitation of xcms corresponded to manual

quantitation (see Figure 3) very well with a fit of R^2=0.99. Manual quantitation for this test was performed in XCalibur v. 1.4.

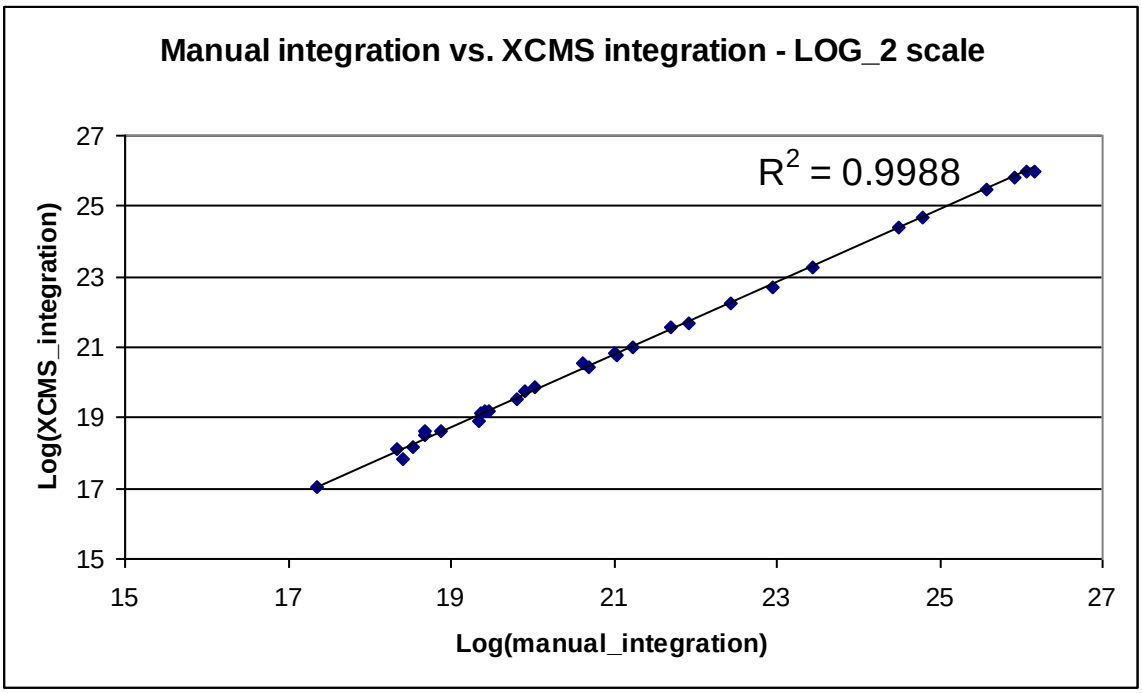

Figure 3 - Comparison of manual quantitation and automatic quantitation by xcms based on 30 different peaks from a standard mixture

Although xcms presents good results in peak quantitation, manual adjustments must be carried out for proper chromatogram alignment. Chromatograms were re-aligned by time after xcms quantitation and the minimal common set of mass signals identified in all chromatograms was kept for further analysis.

2.1.5 Data normalizations

Following our analysis we have encountered several sources of variability in the samples, due to sample loss, growth medium effects, different cell density in the cultures and intrinsic noise due to minute differences in growth conditions. We could compensate for sample loss by careful normalization of data to internal standards and for enhanced cell amounts by normalization to the measurements of cell amounts at sampling time. To cope with other effects we performed statistical analysis of replicates.

2.1.5.1 Normalizing for sample loss during extraction

To normalize for possible loss of samples in the extractions a defined amount of an internal standard[24] (ribitol) was added to each sample (30µl of 0.017mg/ml). During the data analysis,

for each sample the mass signals were divided by the amounts of ribitol as identified by injection of a standard into the GC-MS.

2.1.5.2 Normalizing the metabolome data for variability in cell density

Before quenching, optical density (OD) was measured for each sample at λ=600nm. The OD is proportional to cell density. To calculate the culture opacity at the moment of quenching, the OD for each sample was corrected by the growth rate of the sample and the time that passed from measuring the OD until sampling. $OD_{final} = OD_0 \bullet 2^{t/gr}$ where ***gr*** is the generation time for the sample and ***t*** is the time that passed from OD_0 measurement until quenching. The intensities of mass signals were divided by the OD_{final} value to account for variability in cell/ml quantities in samples.

2.1.6 Method for mass signals assignment to metabolites

During quadruple mass spectrometry analysis metabolites are ionized and broken up to fragments. The original molecule is screened by the abundance of fragments, however, the amounts of each of these fragments is proportional to the original amount of the metabolite. As we strove to research the nature of metabolic response to perturbations and deletion we deemed important to reassemble the mass signals to metabolites for all the samples. We have devised an automatic algorithm that associates mass-signal to metabolites and implemented it in Matlab v. 7.7 (MathWorks). Following the automatic assignment of mass signals to metabolites the data set was further manually validated.

The algorithm for mass signal to metabolite assignment utilizes the fact that metabolites vary in their levels across multiple experiments of different biological samples. This variability is due to slight differences in the extraction procedure of samples, as well as biological variation between samples. At the same time, the relative amounts of different mass fragments resulting from a single metabolite are stable across all experiments due to the robust hard ionization of substances in quadrupole mass spectrometer.

The method unifies mass signals based on the correlation between their intensities in multiple conditions as well as on the closeness of chromatographic retention times.

Since the ratio between the intensities of two mass fragments that belong to the same metabolite is almost constant across different samples, the correlation between them is high across multiple experiments. The correlation between a pair of fragments where each one belongs to another metabolite, on the other hand, is expected to be lower because of the variation in relative levels of the two metabolites in different biological samples.

To assign mass signals to metabolites a hierarchical clustering algorithm was used on the results of xcms output from multiple samples.

2.1.6.1 Distance calculation for clustering

The distance between each two mass signals (Equation 1) is defined as the correlation between them, if the retention time difference between the two mass signals is equal or lower than a user-defined cutoff. If the retention time difference is larger than the threshold, it means that the two mass signals originate from two different compounds, which were separated chromatographically. Therefore, the distance between them is set to be very large.

$$\begin{cases} \text{If} \quad D_{RTij} \le User\_Threshold, & D_{ij} = 1 - \rho_{ij} \\ \text{Otherwise:} & D_{ij} = 100 \bullet D_{RTij} \end{cases}$$

where:

D_{RTij} is the retention time difference between two mass signals *i* and *j*

User_Threshold is the threshold defined by user of the maximal retention time difference between two mass peaks that can belong to the same compounds. This distance is usually defined based on a set of compounds from a standard sample. In the test data set the retention time distance was set to be 1.5 seconds.

ρ_{ij} is the Spearman's rank correlation coefficient between the intensities of the mass signals across multiple conditions.

D_{ij} is the final distance between the two mass signals that will be used as the distance measure for the clustering procedure.

Equation 1 - Calculation of the distance for the clustering algorithm which unifies mass signals based on Spearman's correlation coefficient and distance in the retention time. If the retention time difference between two mass signals is larger than a user defined threshold, the distance between them is much bigger then between two peaks that are near each other in retention time.

We determined the clustering parameters and method by benchmarking clustering results vs. a test set of mass signals assigned to metabolites.

The benchmark set included 51 mass signals in 8 groups with overlapping retention times assigned to 17 metabolites. Different clustering methods and cutoff parameters were tested for the automatic assignment of these mass signals to metabolites (Figure 4).

The similarity of automatic clustering results to the manual assignment of mass signals to metabolites was assessed each time by the Jaccard similarity coefficient (Equation 2).

$$Jaccard = \frac{n11}{n11 + n10 + n01}$$

where for each pair of mass signals
n11 is the amount of pairs that were assigned to the same metabolite both automatically and manually
n10 is the amount of pairs that were assigned to the same metabolite manually, but not automatically
n01 is the amount of pair that were assigned to the same metabolite automatically, but do not belong to the same metabolite in the manual assignment.

Equation 2 - Calculation of the Jaccard similarity coefficient. The Jaccard coefficient penalizes both for splitting the same manually identified cluster to smaller clusters, and for combining too many mass signals into one metabolite.

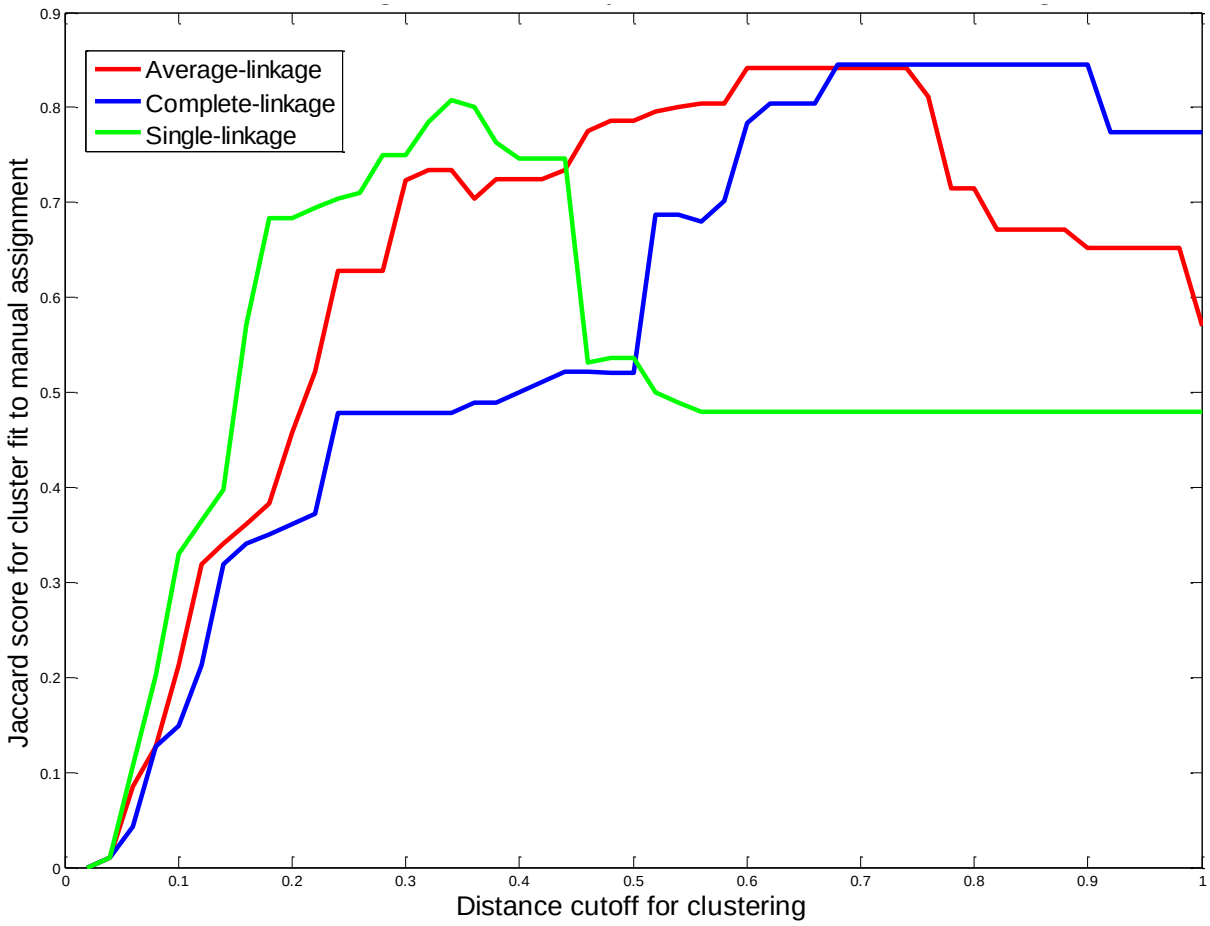

Figure 4 –Jaccard score for the fit of automatic assignment of metabolites to clusters based on a test set for three different clustering methods and different cutoffs for clustering.

Average linkage was the most robust method with respect to the range of parameters. A cutoff of 0.6 gave the highest assignment score for the average linkage method.

2.1.6.2 Manual curation of automatic mass-signal to metabolite assignment and data quality assessment

After clustering of data we have manually curated all resulting clusters by observing in the original chromatograms the masses that clustered together. If the masses did not overlap exactly in time in one or more of the chromatograms the masses were separated to different metabolites. Mass signals that appeared in two metabolites with somewhat overlapping retention times were removed from further analysis due to possible mistakes in quantitation by xcms (see example in Figure 5).

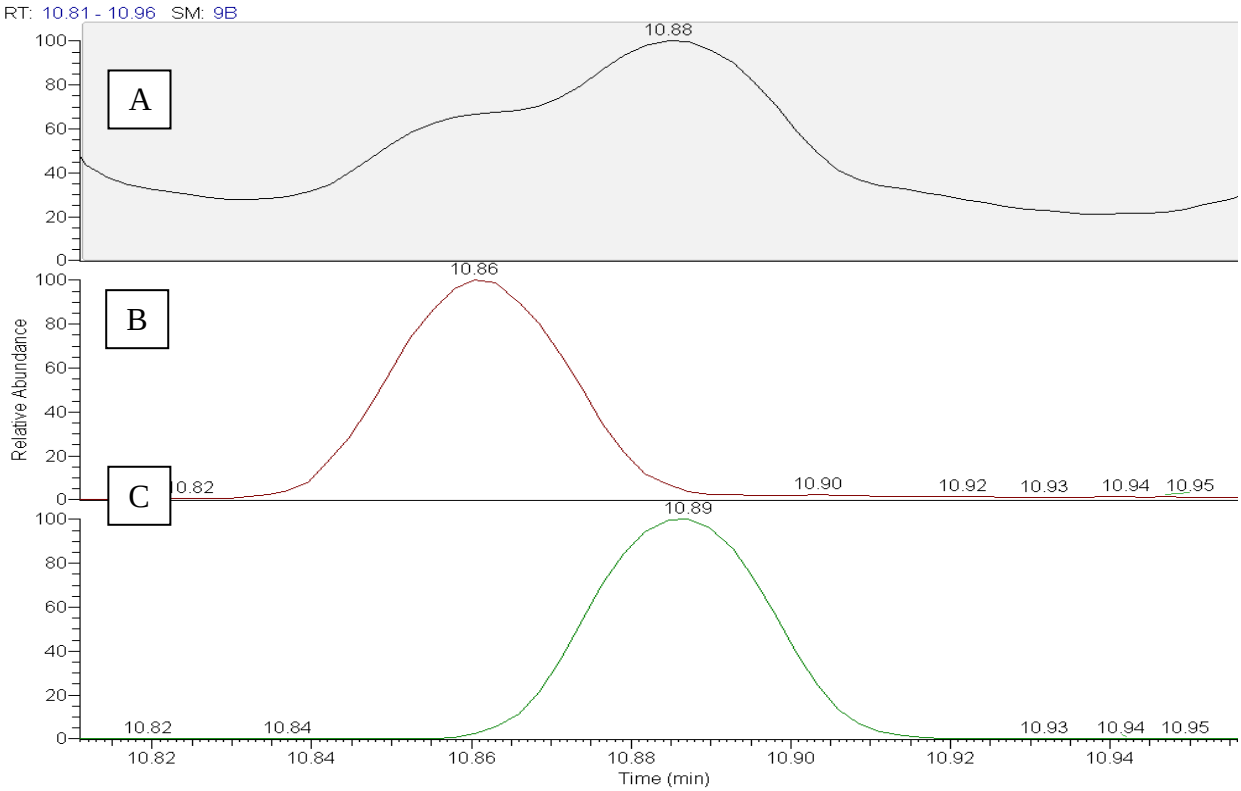

Figure 5 – an example of manual post-processing of chromatograms. A screenshot from a sample chromatogram in XCalibur with single ions 73, 245 and 263 (A, B and C respectively). The mass in section A. (mz=73), exists both at retention time 10.86 min. and at 10.89 min. xcms is prone to error in quantitation of such masses, therefore this mass signal and similar instances within other samples were removed from further analysis. The two compounds were quantified by their unique respective masses at mz=245 (B) and mz=263 (C).

The above pre-processing steps resulted in a list of masses, their quantities in each sample (peak areas) and their assignment to metabolites.

2.1.6.3 Identification of metabolites (mass-signal clusters)

All previous analysis allowed us to assign peaks to metabolites. Some insights can be derived from this non-targeted detection of metabolites. However, to get more biological understanding from metabolomics results we needed to identify as many metabolites as possible in the samples.

Compounds were putatively identified by comparison of their retention index and mass spectrum with those generated for authentic standards analyzed on our instrument[23]. When the corresponding standards were not available, compounds were putatively identified by comparison of their retention index and mass spectrum with those present in the mass spectra library of the Max-Planck-Institute for Plant Physiology (Q_MSRI_ID; http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/msri/gmd_msri.html) and the commercial mass spectra library NIST (www.nist.gov). MS-Search v.2.0d software (NIST) was used for matching spectra of metabolites to standards. Manual comparison of retention indices of standards to the detected metabolites was used to filter the hits from MS-Search software. Retention time indices calculations were based on a mix of alkane chains injected into GC-MS.

Our method allowed detection of 136 compounds. Forty three (43) out of them could be putatively identified. The list of identified compounds can be found in Supplementary 1.

2.1.6.3.1 Metabolic network coverage by identified metabolites.

In the next step we wanted to observe how well the metabolic network of *S. cerevisiae* was covered by the experimental data. Major classes of the identified metabolites included organic acids, amino acids, sugar-phosphates, sugar-alcohols and others.

We plot the metabolites that we could identify on the metabolic network representation of yeast (Figure 6– in red). Overall, we see that the coverage of the metabolic network is quite uniform apart from the ergosterol/lipid biosynthesis pathways and vitamins/co-factors pathways. This is mainly due to detection limits of the GC-MS apparatus we used.

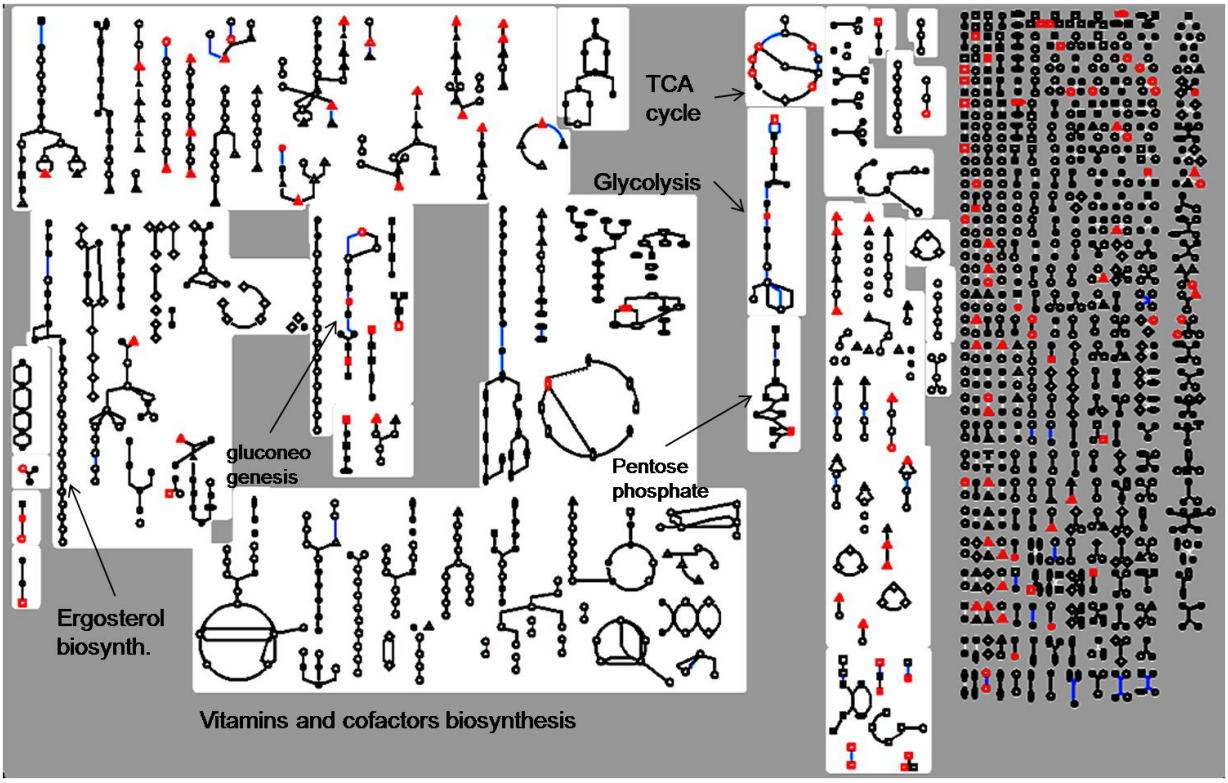

Figure 6 - Representation on the metabolic network of *S. cerevisiae* of reactions that were perturbed (by deletion mutations – see paragraph Error! Reference source not found.**) in cyan and metabolites that could be identified in red. The names near each section represent a cluster of similar pathways.**

2.1.7 Statistical analysis of metabolome measurements of different strains and treatments vs. control

2.1.7.1 Experimental design for sampling and subsequent replicates analysis

To allow for multiple replicates we analyzed 12 samples in every day of experiments. As a general rule, different strains or treatments were analyzed in four replicates along with four controls (wild type (WT) or untreated quadruplicate) within each day. A major goal was to reduce variation in sampling conditions between the replicates.

The standard daily setup included 3 experiment sets of 4 samples each containing:

a) 4 wild type/untreated replicates

b) 4 replicates of mutant_1 or treatment_1

c) 4 replicates of mutant_2 or treatment_2.

Time between quenching of first and last samples in every experiment day was 15 minutes at most. After extraction and drying, samples were kept up to one month at -80ºC. Minimal period of freezing was two weeks to ensure minimal variability between the samples.

The order of the injections into GC-MS was randomized within every day of chemical analysis. Samples were prepared every day, such that the time between injection into GC-MS of the first and last sample was not more than 24 hours.

2.1.7.2 Relative quantitation and analysis of significance of replicate measurements via randomization

The first stage in analysis of a large set of samples consisted of removal from further analysis of outlier samples detected by manual inspection of the data.

Due to the high variability in instrument sensitivity across days it was not possible to directly compare the signal intensity between different strains or treatments. Therefore, for each metabolite within each of the samples we needed to normalize the signal to the WT/untreated samples. The mean of ratios of experiments (***r*****h** in Equation 3) to control samples within the same day were treated as the metric for a change.

$$r_h = \frac{\left(1/n_h\right) \cdot \sum_{j=1}^{n_h} a_{h_j}}{\left(1/n_w\right) \cdot \sum_{i=1}^{n_w} W_i}$$

Equation 3 – For every metabolite, for every day of experiments for each experiment set *h* within that day, the ratio between the mean of the values of experiments was divided by the mean of the WT values within that day. n_h is the amount of replicates in an experimental set. n_w is the amount of wild type/untreated control samples within a day. a_{h_j} to $a_{h_{n_h}}$ are the measured metabolite values of each experimental set within a day. W_i to W_{n_w} are the measured metabolite values of wild type/untreated samples within a day.

To determine the significance of the ratio score we constructed a null distribution for each metabolite in the following manner:

For every set of replicates ***h*** out of a total of ***m*** sets of experiments (including WTs/non-treated sets), and every replicate $\boldsymbol{h_i}$ out of total $\boldsymbol{n_h}$ replicates within a set having metabolite levels a_{h_i} to $a_{h_{n_h}}$ corresponding values α_{h_i} to $\alpha_{h_{n_h}}$ were calculated for the null model according to Equation 4:

$$\alpha_{h_i} = \frac{a_{h_i}}{\left(\frac{1}{n_h - 1}\right) \cdot \left(\sum_{j=1}^{n_h} a_{h_j} - a_{h_i}\right)}$$

Equation 4 – Formula for calculation of metabolite-specific null values for significance analysis of sample ratios to control. *h* is a particular set of replicates, n_h is the amount of replicates for set *h*. a_{h_i} is the intensity of metabolite in an instance *i* of set *h*.

This means that every value was normalized by the mean of the other replicates, thus representing the variability within each sample group and creating a null distribution which is has the same variance or higher than variance in the original samples. We calculated the p-value for each mean of ratios of set *h* normalized to the mean of WT (r_h in Equation 3) by randomization of the null model set. To mimic the fact that we have averaged ratios across n_h replicates we have drawn at random and averaged sets of the same size from the null set. This procedure was carried out 1,000,000 times for each metabolite and for each ratio r_h.

We defined the p-value of a given ratio score as the fraction of random sets that had a similar or higher average than the ratio score r_h.

We corrected for multiple testing using the FDR method[27] with an FDR of 5%. All ratios of metabolites in all the mutant strains/treatments that did not pass the resulting cutoff of 5% were set to 1 to abstain from further analysis of non-significant results.

2.1.8 Reproducibility of data for deletion mutants

Metabolomics results are highly unstable and may vary greatly even between replicate samples within the same day[28]. To assess long-term reproducibility of our method we have carried out independent profiling of 8 deletion mutants with a difference of 10 months between samplings. First we analyzed changes that were labeled significant in both studies. The correlation coefficient between metabolites was 0.82 ($p < 10^{-8}$). Looking at the dot plot (Figure 7), we can observe that only three metabolites out of 33 change in opposite directions, while all the other metabolites change in the same direction both in the new study and in the old study (same-direction changes are 90% of all significant changes).

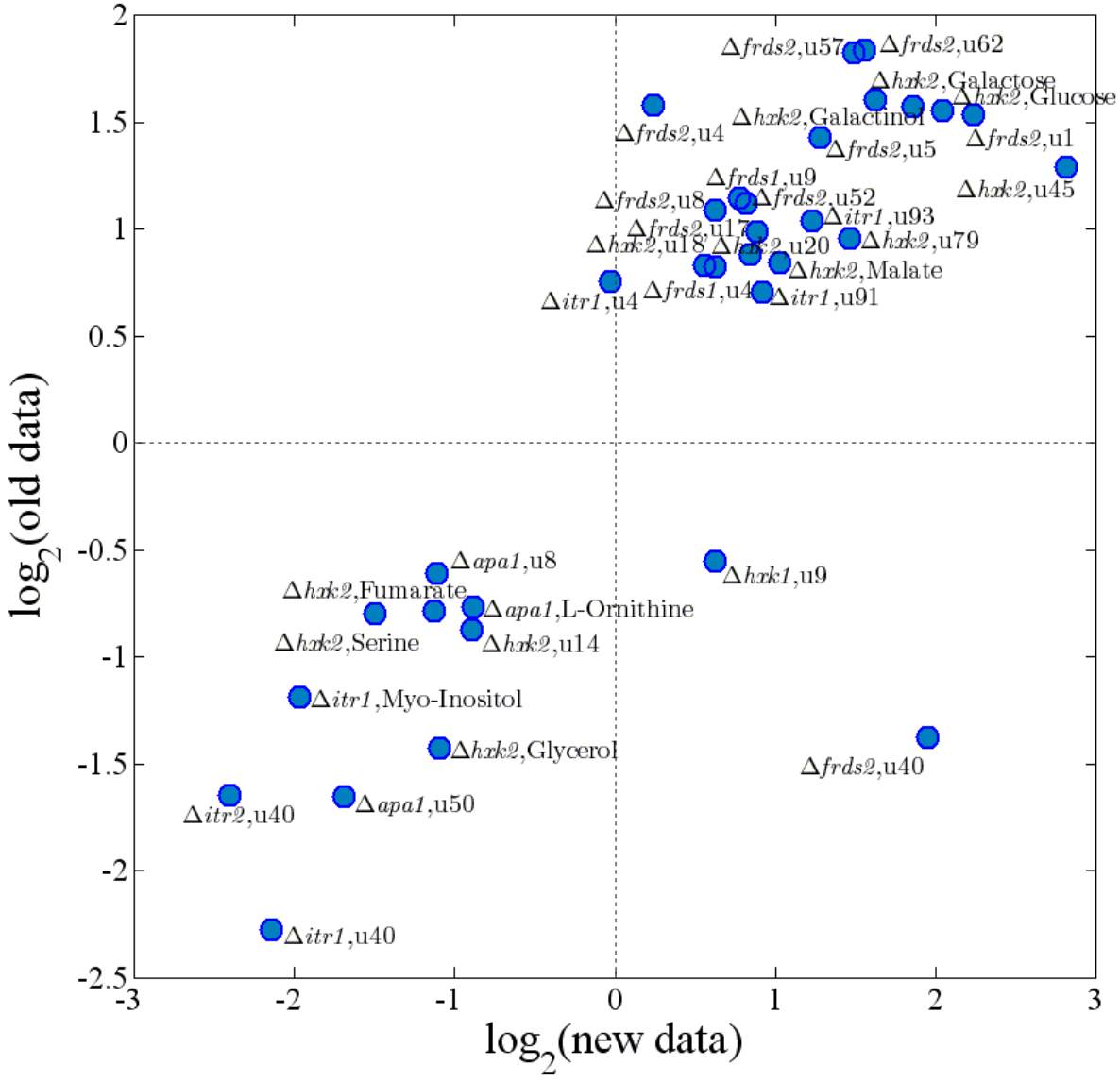

Figure 7 – Dot plot of (mutants/WT) on $\log_2$ scale for metabolites significantly changing in two replicate studies of 8 deletion mutants with 10 months difference between sampling dates. Each label designates the metabolite that significantly changed, and the mutant in which this change occurred.

However, when we performing such a comparison on the union of significant changes in the two studies the correlation dropped markedly to r=0.23 ($p<10^{-4}$). Nevertheless, the proportion of same-direction calls (i.e. instances in which a metabolite's concentration changed in the same direction in both studies) was higher than 63.8% that are expected by random chance (expectation is higher than 50% due to abundance of metabolites with increased levels in all mutants). Same-direction changes represented 73% (213/290) of overall significant changes in at least one of the experimental sets. To assess the significance of this non-randomness we have performed a shuffling permutation analysis and derived the null distribution of expected same-direction changes. With 100,000 permutations we have not even once reached 213 same-direction calls in shuffled data (Figure 8) ($p\text{-val}<10^{-5}$).

old data / new data	↑ (significant)	↑ (non-significant)	↓ (non – significant)	↓ (significant)
↑ (sig)	20	75	35	2
↑(non-significant)	91	359	153	11
↓(non-significant)	15	198	70	8
↓(significant)	1	14	11	10

Table 1 – Distribution of amounts of metabolites increase/decrease in two independent experiment sets (sampled 10 months apart) of 8 single gene deletion mutants. For each category of metabolites (significantly or non-significantly changing upwards/downwards as signified by arrows) in the old set the table contains distribution between the categories for the new set of experiments.

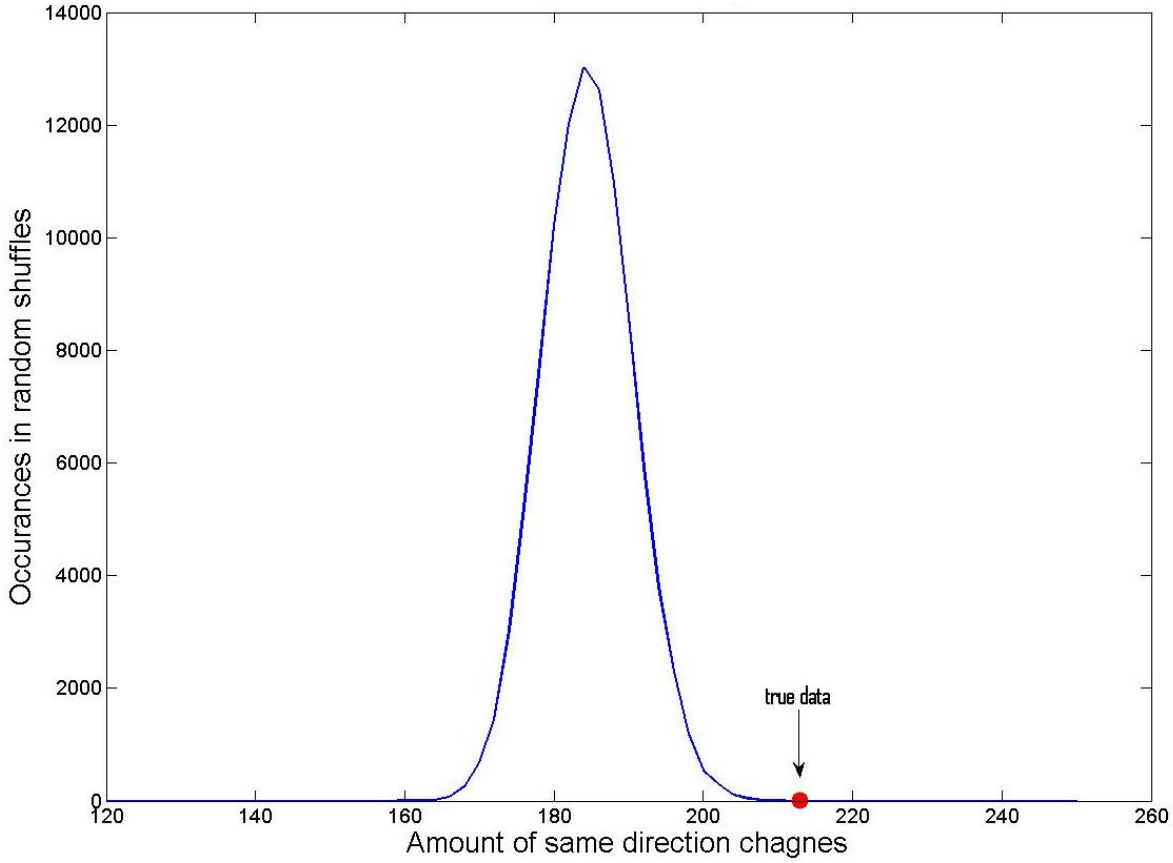

Figure 8 – Distribution of randomly shuffled same-direction changes for old and new extraction experiments of deletion mutants. At random ~185 same-direction hits are expected in the data, this is much less than the observed 213 in true data (marked with a red dot).

Therefore we can deduce that the direction of change is very significantly repeatable overall, but caution must be employed when observing changes that are significant in only one study.

2.2 Metabolic profiling of S. cerevisiae paralogous deletion mutants

After the development of the analytical method we have set out to explore the phenomenon of backup through genetic redundancy that is carried out by paralogs in the *S. cerevisiae* genome. By analyzing deletion mutants that have a paralog elsewhere in the genome we expected to uncover the metabolic component underlying genetic redundancy and transcriptional reprogramming.

2.2.1 Hypotheses

Several potential metabolic responses can be predicted in paralogous deletion mutants.
For mutants that provide perfect backup for each other (e.g. paralogs that diverged only recently in evolution) we would expect no or little difference in the metabolome compared to WT. The same lack of phenotype would be expected with respect to deletion of genes which are not active in tested conditions e.g. glucose-repressed genes[29, 30]. Deletion of such genes is not expected to have an effect on the metabolic composition of yeast cells.
There are two other interesting scenarios for paralogous deletion. One in case of backup with transcriptional reprogramming of the remaining paralog[31-33] and the other in case there is continued persistence of both paralogs to augment gene expression and flux through specific pathways[6, 34].
In the case of **backup with transcriptional reprogramming**, upon deletion of one paralog the mRNA levels of the other paralog might need to be elevated[33]. This effect might lead to only partial complementation and might have an effect on the metabolome. On the other hand, it was also shown that many genes have more than just one function[35]. In duplicate genes this multi-functionality of the common ancestral gene may be responsible for the process leading to specialization and functional divergence of paralogs[36, 37]. Therefore, in the case of deletion with reprogramming, cellular effects beyond those resulting from incomplete backup by the upregulated paralog may arise. The so-called “moonlighting” or minor function unique to the remaining paralog with elevated levels would tend to affect the metabolome. Therefore, this is another reason we would expect different metabolic profiles of the two deletion mutants.
In some instances, however, transcriptional reprogramming occurs only upon deletion of one of the paralogs, but not upon the deletion of the other[32]. In special cases of good backup between paralogs a one-sided metabolic response would occur upon the deletion of one paralog but not the other.

In the second case, paralogs which perform similar functions may **complement each other to increase the metabolic flux** through a reaction. One such example are genes duplicated during the whole genome duplication event in yeasts that increased the overall glycolytic flux within cells[38].

Upon deletion of each one of the paralogs in such case, the concentration of the precursor of the reaction would increase, and may further drive downstream effects. These downstream effects may be similar for both paralogs. Therefore, the metabolic profiles of the deletion mutants relative to WT samples would tend to be similar for both of the paralogs.

The summary of predicted relative metabolic profiles can be found in Table 2.

Hypotheses summary for paralogous backup	
redundancy scheme employed by paralogs	**predicted response to deletion**
full complementation	little effect on metabolic profile
backup with transcriptional response	one-sided or two-sided effect on metabolome
retention of paralogs for flux increase	similar effects for both paralogs

Table 2 – Hypotheses summary regarding behavior of mutant strains with deletions of paralogous genes.

To check the above hypotheses we have carried out metabolic profiling of 39 pairs of paralogous deletion mutants (mutant list can be found in Supplementary 2a) from diverse gene families in the metabolic network of *S. cerevisiae* (a total of 78 mutants). Results are found in section 2.2.2. Looking at these profiles we will gain additional understanding of the deletion phenotype of genes in *S. cerevisiae.*

Additionally, we wished to further validate the results and to gain additional insight into the mechanisms of backup and co-function of paralogous genes. For this we have repeated our experiments and acquired again metabolic profiles of four pairs of the deletion mutants along with their double mutants, in which both paralogs were deleted (section 2.2.3). The double mutants were chosen such that the resulting strain was viable. Mutant list can be found in Supplementary 2b.

2.2.2 Analysis of paralogous single mutants

The dataset constructed for the single mutants represented deletion mutants in 78 different paralogous enzymes related to the metabolic network of *S. cerevisiae.*

The procedures carried out for the acquisition of the data are described in section 2.1.

2.2.2.1 Observing metabolic response in the Δaco1 and Δaco2 mutants

As a primary check for our results we wanted to focus on the behavior of one specific pair of paralogous genes with well studied function. For this we chose *ACO1* and *ACO2* and analyzed the metabolic profile obtained for their deletion mutants. *ACO1* encodes the enzyme aconitase which is part of the tri-carboxylic acid cycle (TCA cycle - Figure 9) and converts citrate to iso-citrate via the intermediate aconitate. *ACO2* has a sequence similarity of ~55% to *ACO1* and has putative aconitase activity[39].

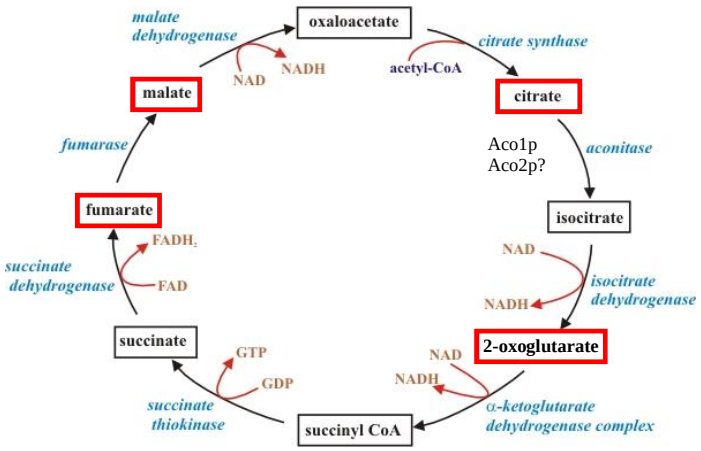

Figure 9 – TCA cycle, Aco1p and Aco2p enzymes are noted on the plot. Metabolites which changed significantly upon deletion of either *ACO1* or *ACO2* are marked in red.

In Figure 10 we plot the changes in both mutants relative to the WT. We can see that Δ*aco1* mutant has more metabolites significantly changing in response to the deletion than Δ*aco2* (27 in Δ*aco1* and 19 in Δ*aco2*, with mean fold change of 1.98). The *ACO1* gene is known to be the major isoenzyme that is responsible for the aconitase activity[40], it was also shown to be one of the factors for mitochondrial genome maintenance[41]. As could be expected, metabolites from the TCA cycle (marked with red color in Figure 9 and in Figure 10A) are up-regulated in Δ*aco1*, but less so in Δ*aco2* possibly due to lower flux through the cycle. Thirteen (13) metabolites exhibited significant changes in both mutants with similar directionality of the change. Interestingly, two of the identified common metabolites are sterols of unidentified structure. Sterols composition of yeast cells is known to be regulated by functioning mitochondria. Deficiency in mitochondrial function by deletion of aconitase may have lead to

accumulation of some sterols in the cells.

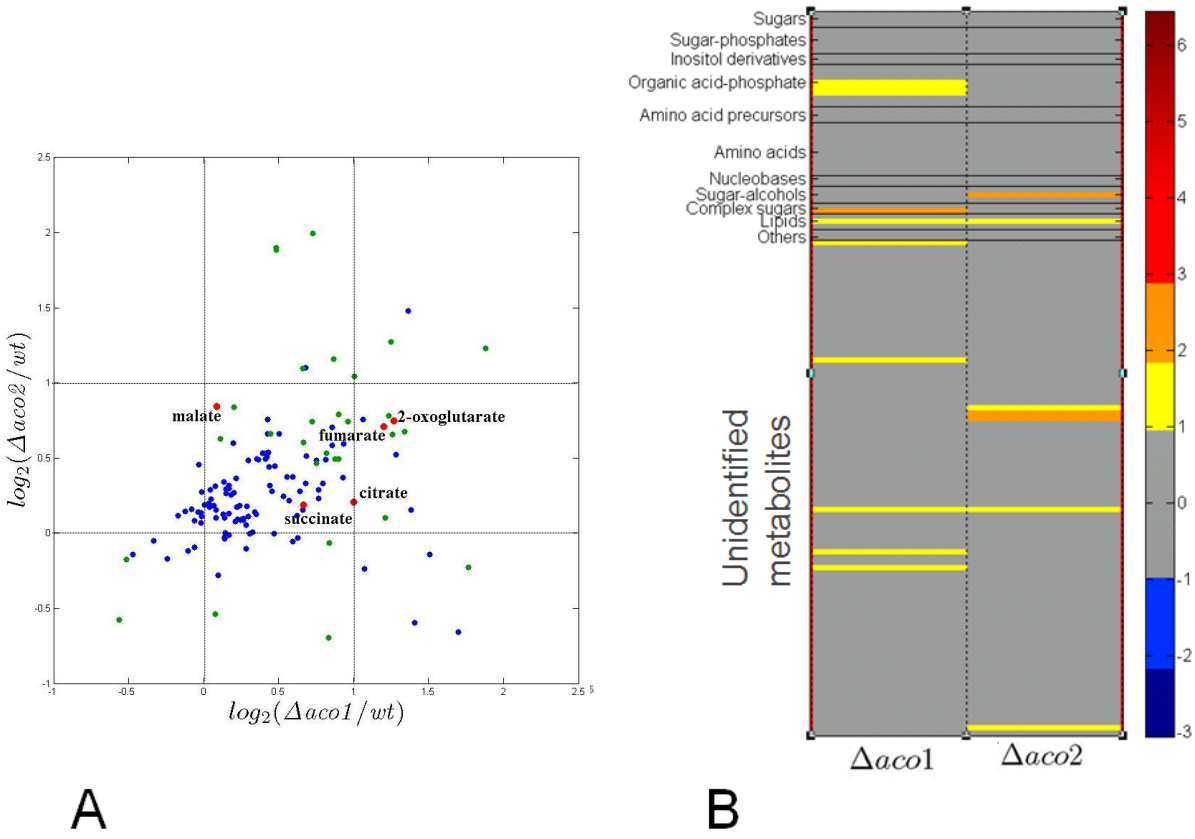

Figure 10 - A. Dot plot on a log_2 scale describing the ratios of metabolites in Δ*aco1* and Δ*aco2* mutants to WT. Metabolites changing significantly in one of the mutants are marked in green, metabolites from the TCA cycle are labeled marked in red. B. Relative metabolite changes in the two mutants on log_2 scale. Colorbar on the right presents colors associated with two-fold change steps. All metabolites not changing significantly or having less than two-fold change are grey.

2.2.2.2 Reciprocal metabolic response of paralogous gene deletion mutants

After the analysis of a particular pair of paralogs we have set out to get a large-scale view of paralogous backup. To notice to the most prominent effects we have limited our view at the first stage to changes higher than 2-fold in the mutants relative to the WT controls (Figure 11).

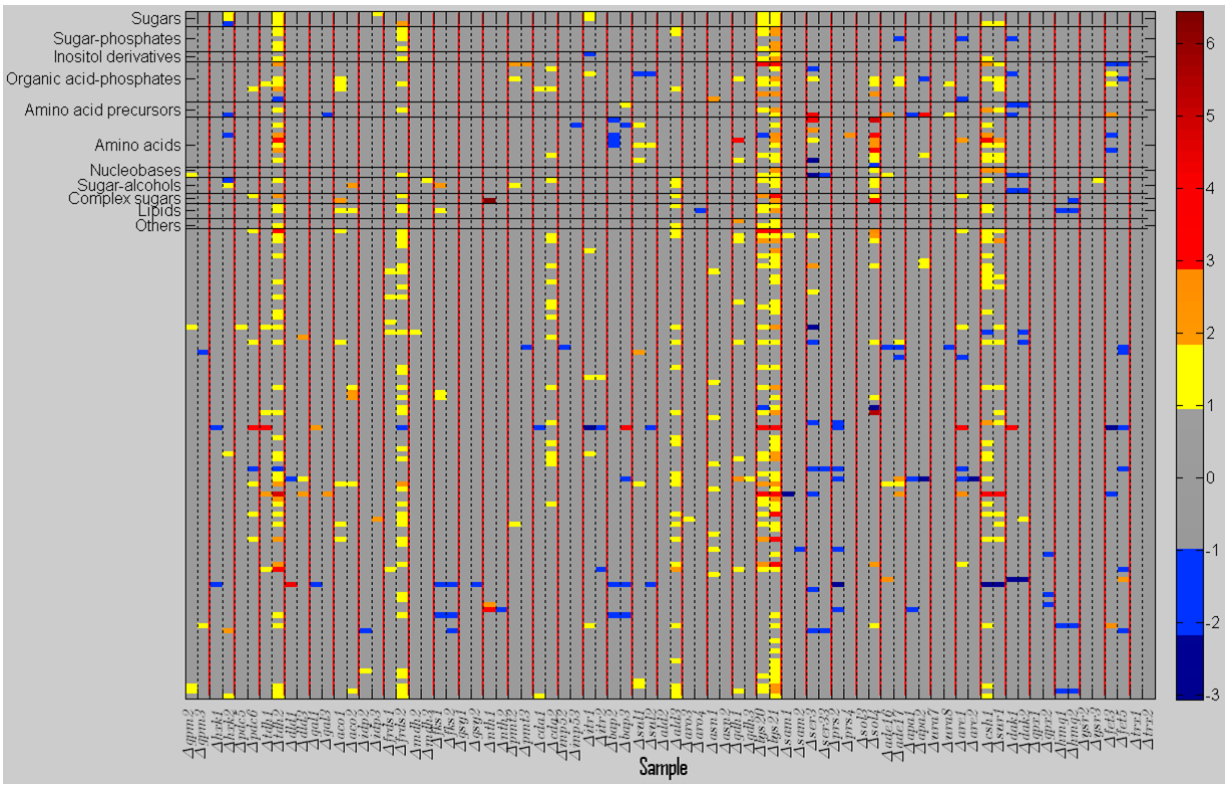

Figure 11 – Metabolite changes in paralogs relative to WT. Rows represent metabolites' relative intensity. columns mark different mutants . Both members of each paralogous pair are found near each other and the different pairs are separated by vertical red lines. The data is on a $\log_2$ scale with the color bar representing the fold change on the right. Grey metabolites are such that have either response of less than 2-fold, or that are not significantly changing. In the upper part of the figure appear metabolites that could be identified, different groups of metabolites are separated by black horizontal lines. The groups are according to the list found in Supplementary 1.

Three different types of responses could be identified when comparing reciprocal metabolic response within deletions of pairs of paralogs.

1) **Little or no metabolic response** upon the deletion of each of the paralogs.
2) **Concerted changes** in metabolites upon the deletions of the paralogs, i.e. the deletion mutants shared significantly up- or down-regulated metabolites.
3) **One-sided response**. We observe pairs, in which only of one of the deletants responds strongly to the deletion, while the metabolic response to the deletion of the paralogous counterpart is very mild.

Examples of the three types of reciprocal relationships can be found in Figure 12.

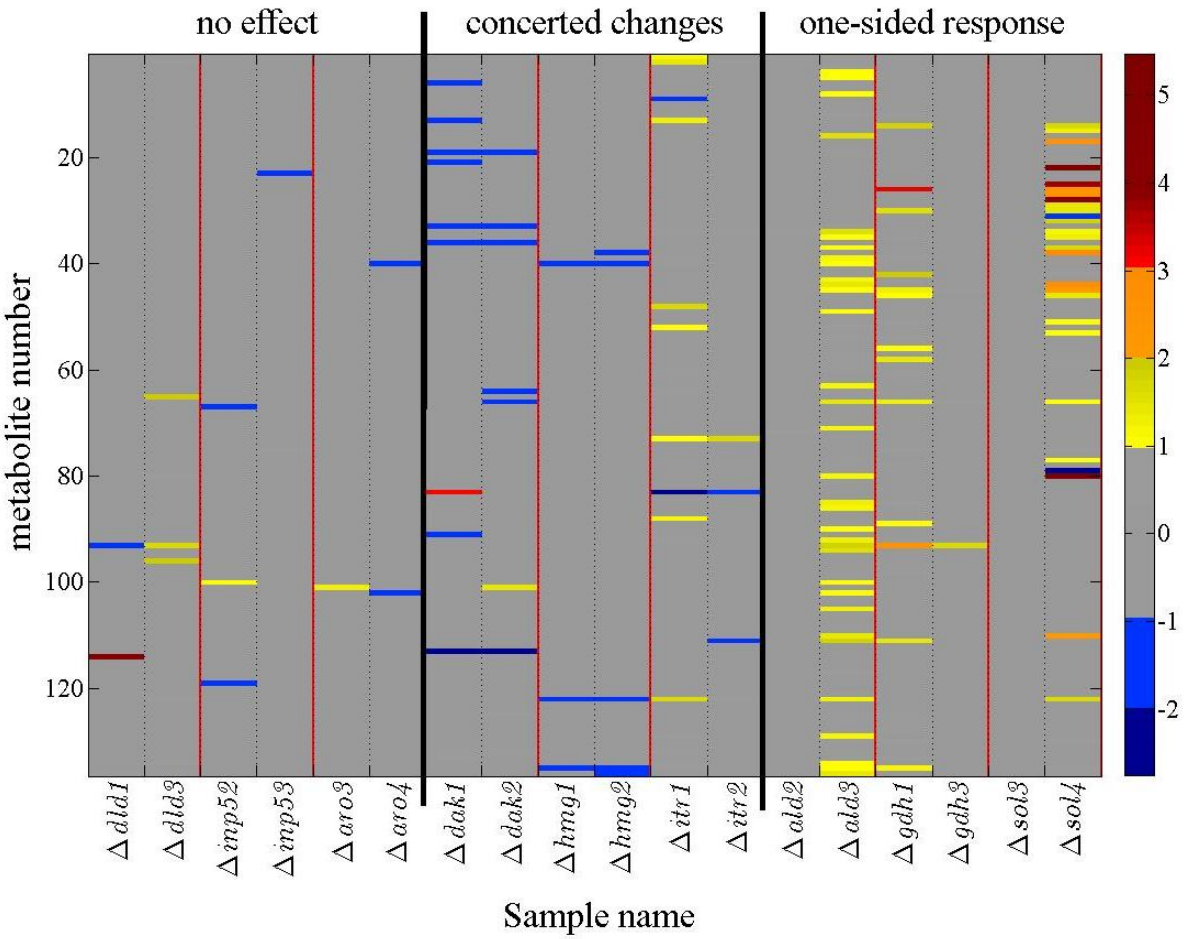

Figure 12 - Nine examples of different types of metabolic responses (from left to right – little metabolic effect, multiple concerted changes and one-sided response to deletion). The data is on a log2 scale with the color bar representing the fold change on the right. Grey metabolites either have a response to deletion of less than 2-fold, or are not significantly changing.

Several explanations may be offered for the observed phenomena. We will go over the different types one by one.

2.2.2.2.1 Pairs exhibiting little or no metabolic response

There are two possible explanations for pairs that exhibited little or no metabolic changes in response to the deletion of both paralogs. One is that the genes that were deleted have no important function in the checked experimental conditions. They can either be repressed, or even if transcribed their function might be of little importance in the checked conditions (e.g. DLD1 and DLD3 which both play a function in lactate metabolism, but are repressed upon growth on glucose). The second explanation for no apparent effect can be the fact that paralogs provide good backup to each other, i.e. that the genes function in a similar manner.

A possible key for choosing between the two hypotheses may be found in the rate of evolutionary divergence of paralogous gene pairs. Selective pressure toward similar

functionality of genes (purifying selection) that exhibit little metabolic change would favor the second hypothesis, as non-functional genes are expected to be evolving independently of the amount of metabolic changes.

To assess the rate of purifying selection on paralogs in question we have employed Ka/Ks analysis[42, 43]. The rate of mutation of sequences is not similar for all positions in genes. Mutations which do not change amino acid composition of a gene (synonymous mutations) are by large evolutionary neutral, while mutations which do result in amino acid substitution are usually much slower to get fixated. It is possible to assess the rate of synonymous (Ks) and non-synonimous (Ka) substitutions in a pairs of genes[31, 44]. Gene pairs with a lower Ka/Ks ratio are usually thought to be under purifying selection[42, 43], i.e. these genes are under selective pressure to weed out non-synonymous mutations which disrupt gene function.

We analyzed the Ka/Ks ratios in our gene sequences[31, 44] and compared it with the mean amount of changed metabolites in each pair (Figure 13).

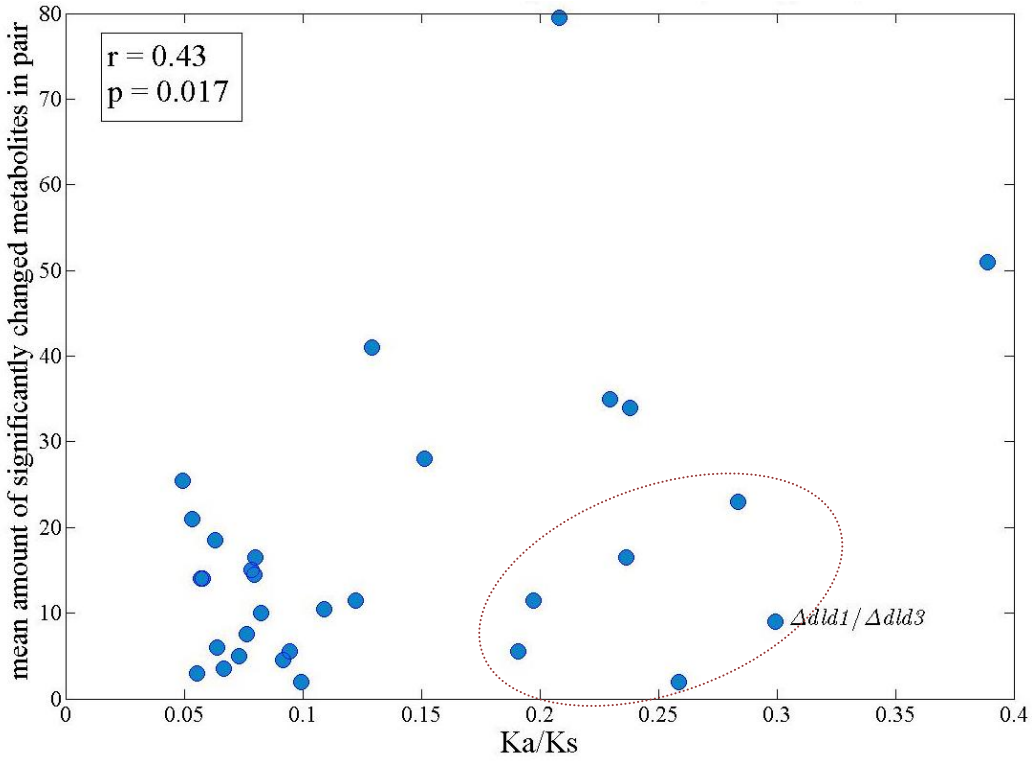

Figure 13 – Ka/Ks ratio vs. the mean amount of significantly changed metabolites in each paralogous deletion pair. Red ellipse marks gene pairs which are outliers to the main correlation trend.

We observed a significant positive correlation between the mean amount of changed metabolites in a pair and its Ka/Ks (r=0.43, p=0.017). However, we have additionally found

that there is a negative correlation between Ks and the mean amount of changes in metabolites (r=-0.45). To verify that the positive correlation with Ka/Ks is not solely due to the Ks values, but also due to the Ka, we perfomed a partial correlation analysis of Ka/Ks vs. the mean amount of metabolic changes **controlling** for Ks values. The resulting partial correlation was still significantly positive (r=0.34, p=0.03).

This positive correlation implies that the amount of changed metabolites is associated with purifying selection towards keeping paralogous pairs similar. Upon deletion of genes that are slower to diverge there are less metabolites changing. Therefore, we propose that the effect of little or no metabolic change is usually due to well carried out backup between paralogs, which were selected to similar functionality, rather than due to the occurrence of non-functional genes. The above analyses settle well with the suggestion by Kafri and colleagues[32, 33, 45] that paralogous gene pairs are selected to retain some degree of similarity in their activity to allow for better functioning of the cells e.g. with respect to molecular noise.

Looking beyond the correlations we can discern two groups on in Figure 13. We can identify specific groups of paralogous pairs which exhibit a deviation from the overall pattern (see red circled gene pairs in Figure 13). These pairs exhibit a relatively lower purifying selection, while having relatively low amounts of metabolic changes upon deletion. We can hypothesize, that genes in this part of plot are not strongly functional in the conditions we checked (high glucose). Indeed, when looking at their identity we notice that the *DLD1* and *DLD3* gene pair (which is repressed by glucose) is in that group.

2.2.2.2.2 Pairs exhibiting "concerted" changes

Another phenotype is the one of "concertedness", as exemplified by the pairs of paralogs appearing in Figure 14.

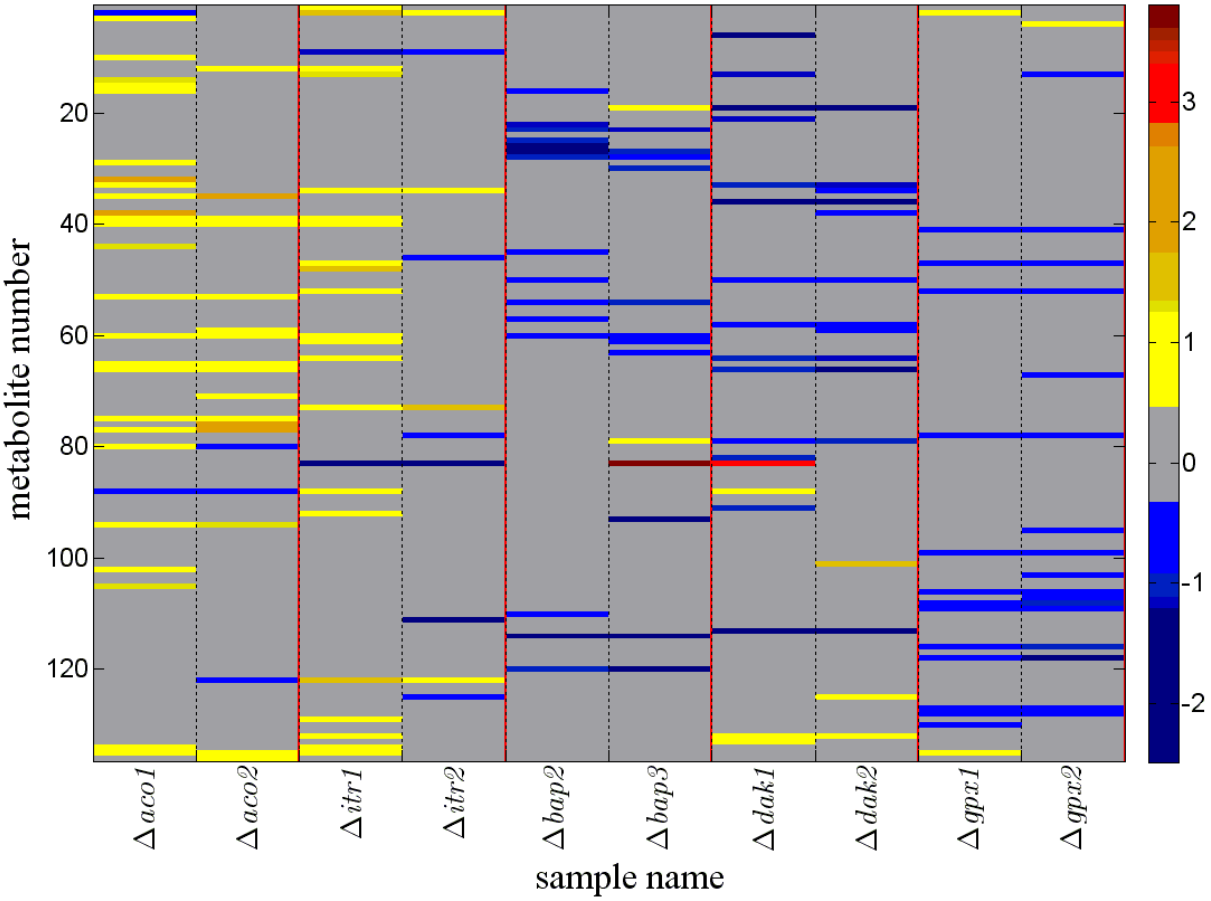

Figure 14 – Five mutant pairs with high concertedness scores as calculated by random permutation analysis. The data is on a $\log_2$ scale with the color bar representing the fold change on the right. Grey metabolites are not significantly changing.

Pairs with concerted changes are these in which common metabolic changes relative to the WT strain occur upon deletion of any of the two paralogs. Two possible reasons may contribute to such phenotypes.

One is the dosage effect that two paralogs; i.e. situation in which both paralogs perform a common reaction and both are required to provide the full flux through the reaction and synthesize the full amount of the downstream product. In case of deletion of any of these paralogs the flux through the reaction will be diminished and the downstream effects will be similar for both of the paralogs (mainly increase in upstream metabolites). Another possible explanation is that the two paralogous genes belong to the same protein complex. Without either of them the complex is destabilized or its formation happens with lower probability. Therefore the flux through the reaction is diminished and the downstream effects are similar.

2.2.2.2.2.1 Scoring “concerted” response

In order to have a better look at paralogous pairs with concerted response we first needed to devise a score for the “concertedness”. We have counted for each pair of deletion mutants the

amount of same-direction changes in significantly fluctuating metabolites in both mutants and subtracted it by the amount of opposite direction changes in significantly fluctuating metabolites. To assess the significance of this overlap we have performed $2 \cdot 10^7$ shuffles for each pair of paralogous deletants such that the labels of significantly changing metabolites were distributed between all measured metabolites for each of the deletion mutants then we calculated the overlap score for each of the shuffles. For the significance of metabolic response "concertedness" for each of the pairs we calculated the proportion of times shuffled data had overlap score ≥ true overlap score. The final "concertedness" measure was derived by taking $-\log_{10}$ of the p-value of overlap (Figure 15).

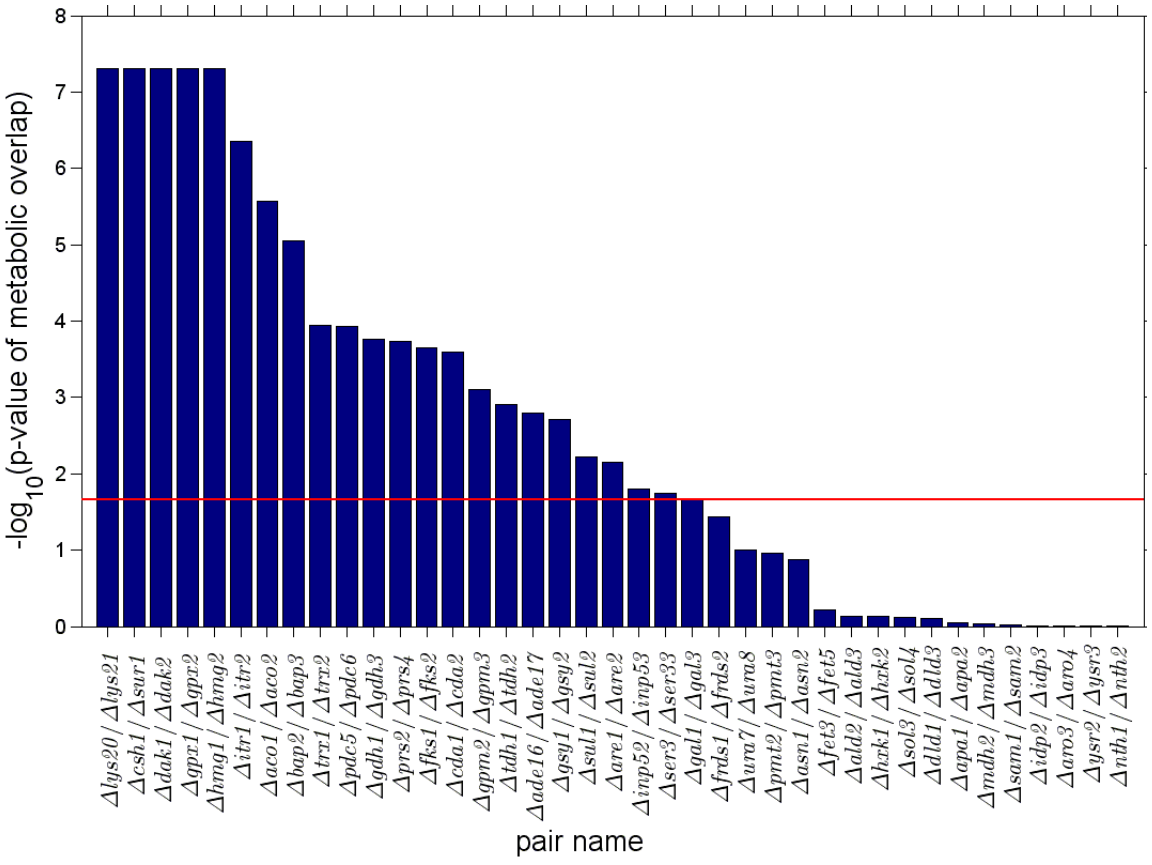

Figure 15 – "concertedness" levels of different paralogous gene deletion pairs. The score was constructed as noted above, y-axis is $-\log_{10}$(p-value of concertedness measure for each pair). Values above the red line (p-value≤0.02) have passed significance testing controlling for multiple hypotheses with FDR q-value of 0.05.

2.2.2.2.3 Pairs exhibiting one-sided response to gene deletion

Third phenotype that could be discerned in the relative metabolic profiles of paralogous deletion pairs was the phenotype of one-sided response; a situation in which deletion of one of the paralogs evokes a strong metabolic response, while the deletion of the second paralog causes little or no metabolic phenotype. Several examples of such a behavior can be found in

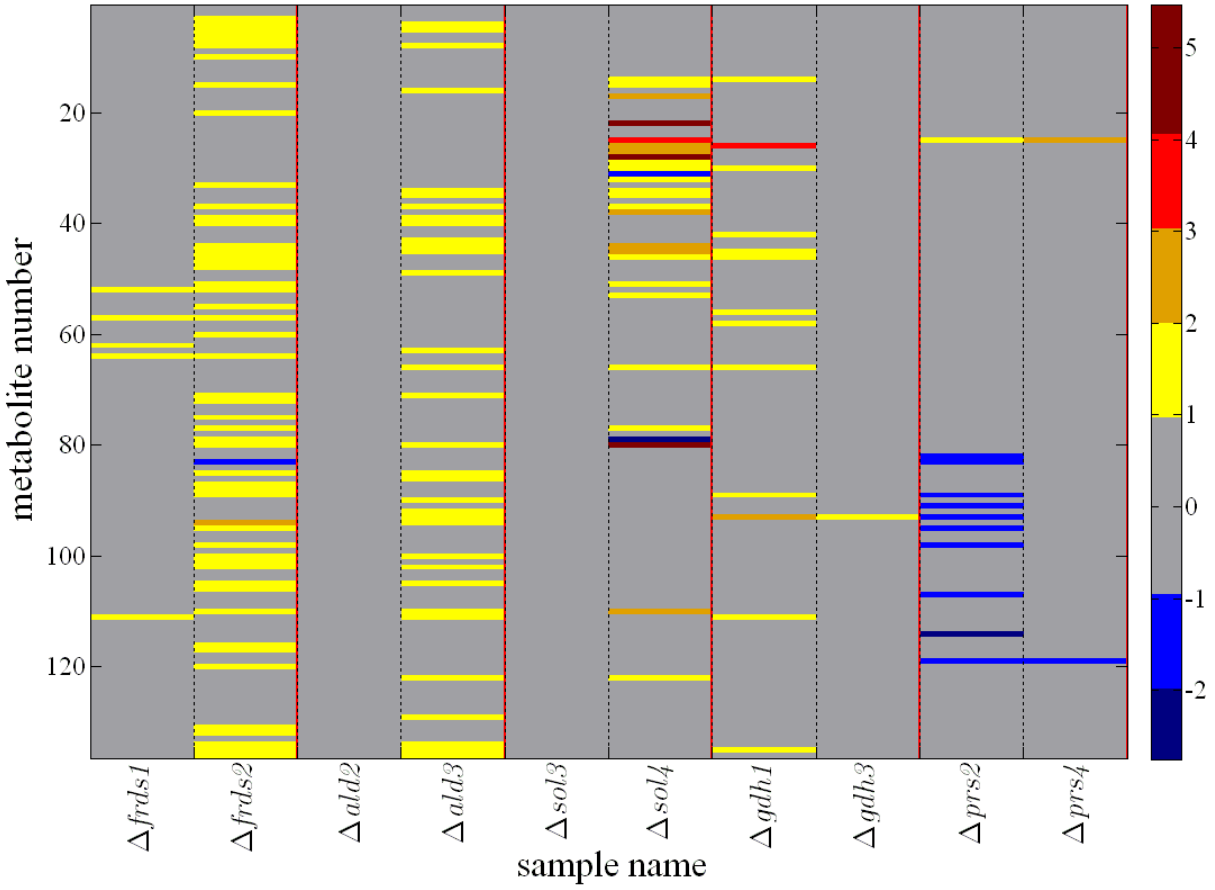

Figure 16 - Five mutant pairs with high one-sidedness scores as calculated by random permutation analysis. The data is on a $\log_2$ scale with the color bar representing the fold change on the right. Grey metabolites are not either not significantly changing or are have less than 2-fold change.

We have quantified the basic measure for one-sidedness of response of paralogous pairs to deletion by calculating the value given in Equation 5. We have measured the difference between the amounts of significantly changing metabolites in each of the samples, subtracted the overlap in changed metabolites and normalized by the total amount of changing metabolites.

$$O_k = \frac{\left|N_i - N_j\right| - N_{common_{ij}}}{N_i + N_j}$$

Where O_k is the one-sidedness measure for each pair of single gene deletion mutants. N_i and N_j are the amount of significant changes in each of the deletion mutants' metabolic profiles. $N_{common_{ij}}$ is the amount of metabolites changed to the same direction in both mutants.

Equation 5 – Concertedness measure calculation for each pair of gene deletion mutants.

To assess the significance of one-sidedness we have performed 10,000 shuffles for each pair of paralogous deletants such that the labels of significantly changing metabolites were randomly

distributed between all measured metabolites in both deletion mutants. For the p-value of "one-sidedness" of response for each of the pairs we calculated the proportion of times shuffled data had one-sidedness score ≥ true one-sidedness score. The final "concertedness" score was derived by taking -$\log_{10}$ of the p-value of one-sidedness measure.

We can see that only 8 pairs of mutants had a significant one-sided response to deletion mutation.

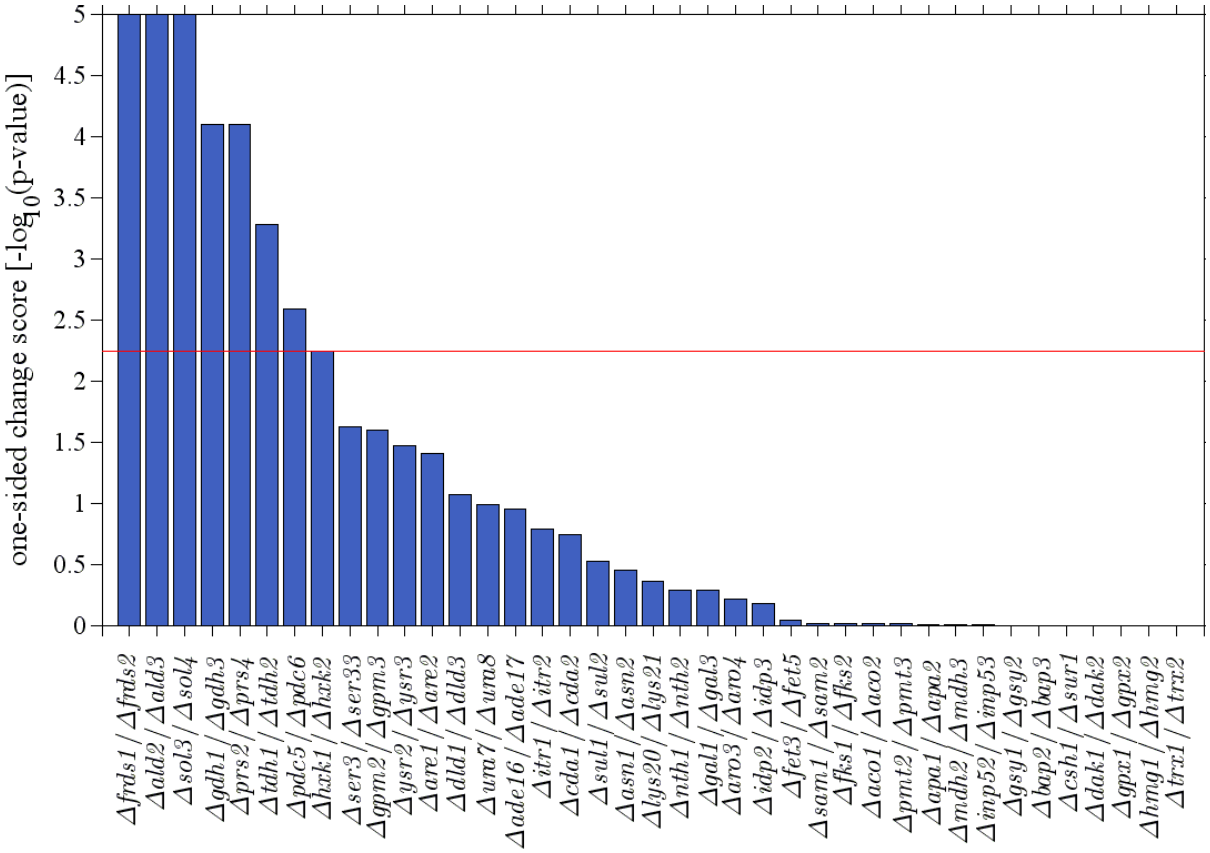

Figure 17 – "one-sidedness" levels of different paralogous gene deletion pairs. The score was constructed as noted above. Values above the red line (p-value≤0.0057) have passed significance testing control for multiple hypotheses with FDR q-value of 0.05.

We can offer two plausible hypotheses explaining the observed one-sided response of paralogs. One is existence of good backup by transcriptional reprogramming of one paralog, but not the another[32]. Another explanation may lie once again in non-functionality of one of the checked genes in our experimental conditions. For example, the glutamate dehydrogenase gene *GDH3* is known to be repressed by glucose rich conditions, while its paralog (*GDH1*) **is** active in glucose rich conditions[46]. This is in good correspondence with the absence of metabolic response we observed in the Δ*gdh3* mutant, and a strong response upon deletion of Δ*gdh1*. Moreover, it was experimentally shown that deletion of *GDH3* does not cause any change in the levels of *GDH1* (Michael Springer personal communication). Such a change could have

caused an increased flux through the reaction and might have resulted in some change in the metabolic phenotype.

2.2.2.3 Analysis of metabolites responsive to cell growth rate

Among the mutants we have profiled, some differed in their growth rate on minimal medium relative to the wild type strain. Metabolites both regulate the growth rate of yeast cells and are affected by it. Therefore, we have set out to check which metabolites are either positively or negatively correlated with the growth rate or fitness of cells. For each mutant we have measured the growth rate relative to the WT grown in the same day. As a proxy to fitness we used the inverse of the relative growth rate (1/relative growth rate).

First of all we wanted to examine whether the sheer amount of changed metabolites in mutants was indicative of deviation from the wild type fitness levels.

We can observe that there is no direct correlation between the count of changed metabolites and the fitness of mutants (Figure 18). However, when looking only at mutants with relative fitness lower than wild type we see a significant, albeit somewhat low negative correlation (r=-0.43, p-value = 0.02) between relative fitness and the amount of changed metabolites.

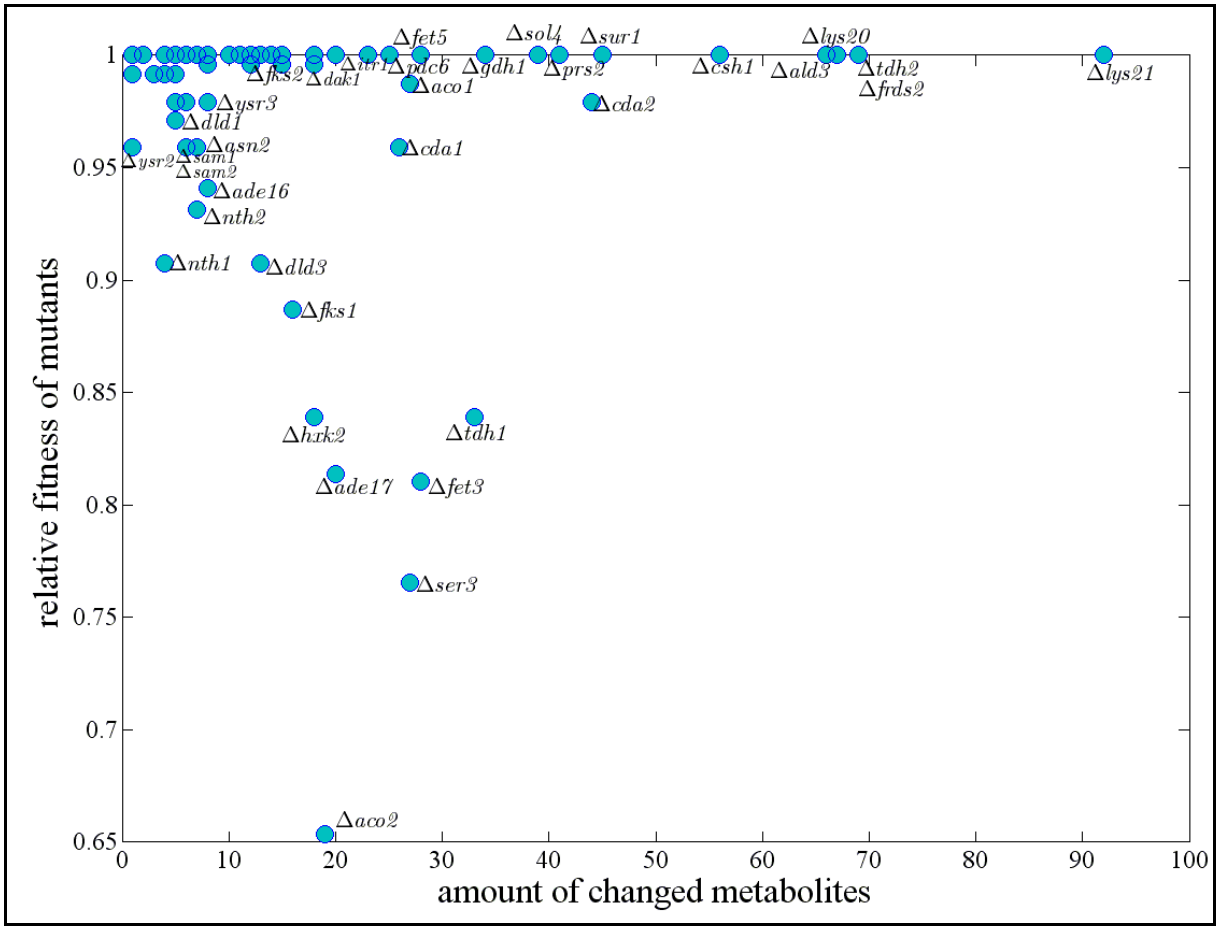

Figure 18 – Relative fitness of paralogous single mutants as a function of amount of significantly changed metabolites in a mutant. Labels of select mutants are given on the plot.

We further performed a similar analysis splitting between metabolites whose concentrations increased in mutants and those whose concentrations decreased (Figure 19).

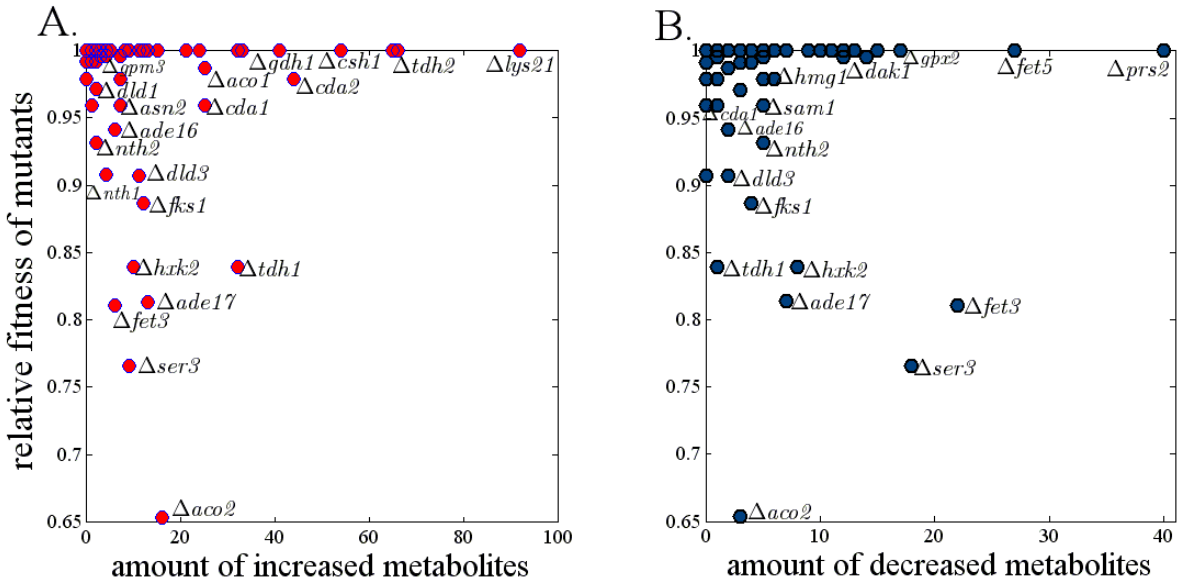

Figure 19 - A. Amount of metabolites <u>increased</u> after deletion of each mutant vs. relative fitness of the mutant (mutant growth rate/wt growth rate). B. Amount of metabolites <u>decreased</u> after deletion of each mutant vs. relative fitness of the mutant. Labels of select mutants are given on the plots.

The correlation between the amount of changes to each of the sides and the relative fitness of mutants different than WT is still negative (-0.1 and -0.3 for amounts of increased and decreased metabolites respectively) but not significant given the sample size. This suggests, then, that the association of fitness is with the overall metabolic change, rather than the directionality of the change.

To check which specific metabolite's levels are associated with fitness of mutant strains we have performed a correlation analysis between the two variables. Yet, no individual metabolite showed significant correlation with the mutant fitness. Perhaps, however, due to some higher level interaction between metabolites the combined information from metabolic profiles of mutants can be used to predict fitness defects?

To tackle this question we have used a simple artificial neural network which learned to predict the fitness of all mutants from metabolic profiles. The predictor network was constructed with just one perceptron neuron. The perceptron was trained on a randomly chosen set of metabolic profiles of mutants containing 70% of the data and each time validated on the remaining 30%. The proportion of training and the validation set contained the same amounts of data with growth defects. The target was a vector with designation whether a mutant had or did not have

a growth defect. As a control we trained a similar perceptron on the same data with a shuffled target vector. The process was repeated 500 times. Figure 20 depicts the distributions of the amount of correctly classified mutants in the validation set of shuffled and true data.

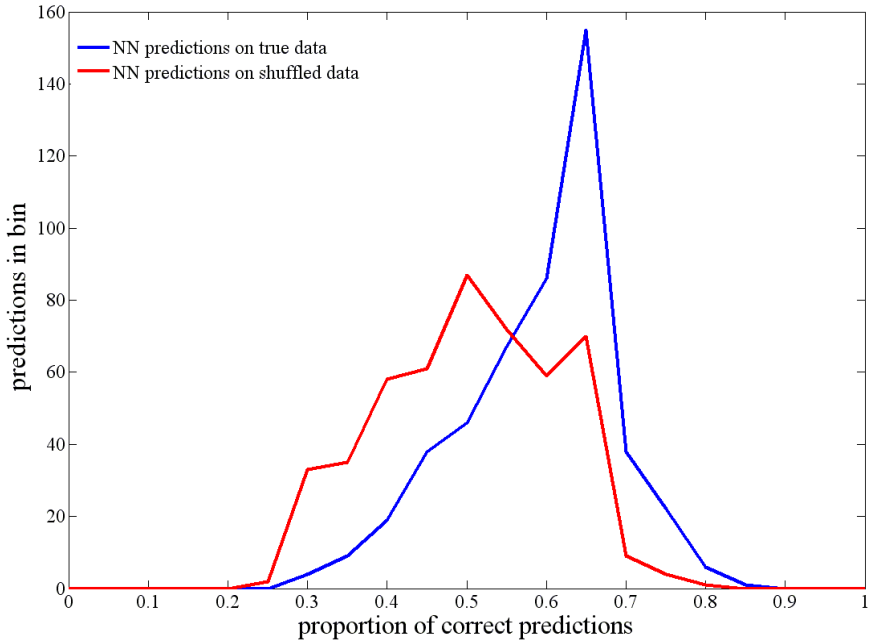

Figure 20 – Distribution of amount of correctly predicted growth defect phenotypes for mutants based on the metabolic profiles of deletants – 500 randomizations. In blue appears the distribution of correct predictions based on true designations of growth defects to mutants, in red appears the distribution of correct predictions based on shuffled designations of growth phenotype.

We can see a difference in the two distributions. The mean and the median proportion of correct predictions using the shuffled labels is as expected 50%. Based on the true data, however, correct predictions exist in 60% of the cases on average. This difference is highly significant (Wilcoxon rank sum test p-value<10^{-32}). From this result we can infer that metabolic profiles of deletion mutants can to some degree increase the chances to predict growth rate defects in different strains.

2.2.2.4 Prediction of subcellular localization of proteins based on the metabolic profile of deletion mutants

Deletion mutants that we analyzed differed in their subcellular localization. It is plausible to assume that deletion of a gene in a particular subcellular location e.g. mitochondria will cause accumulation or decrease in metabolite levels belonging to that specific location. We have anecdotally observed this behavior upon deletion of *ACO1* and *ACO2* genes (which are part of the TCA cycle genes localizing to the mitochondria). Metabolites that participate in mitochondrial respiration exhibited significant changes upon deletion of the genes (Figure 10, section 2.2.2.1).

We wanted to check further whether this behavior is a general property of deletion mutants. If so, then metabolomics data from deletion mutants should predict significantly better than random the subcellular localization of genes.

We have employed a strategy similar to that described in section 2.2.2.3. We constructed a dataset with the annotation of subcellular localization of genes that we profiled according to GO [47] (Saccharomyces Genome Database accessed June, 2009). Subsequently, we divided the genes into 4 groups; mitochondrial only, cytoplasmic and mitochondrial, cytoplasmic only and belonging to other subcellular localizations (peroxisome, endopasmic reticulum *etc.*). A simple two-layer probabilistic neural network was constructed using a built in Matlab program to predict subcellular localization of a mutant by its metabolic profile. The data was divided once again to 70% training set mutants and 30% validation set. The proportion of genes belonging to each of the 4 subcellular localization groups was kept constant in the training and the validation sets. As a control we have trained a similar neural network using shuffled labels of subcellular localization of the mutants. The process was repeated 5000 times. Figure 21 depicts the distributions of the amount of correctly classified mutants in the validation sets of shuffled and true data.

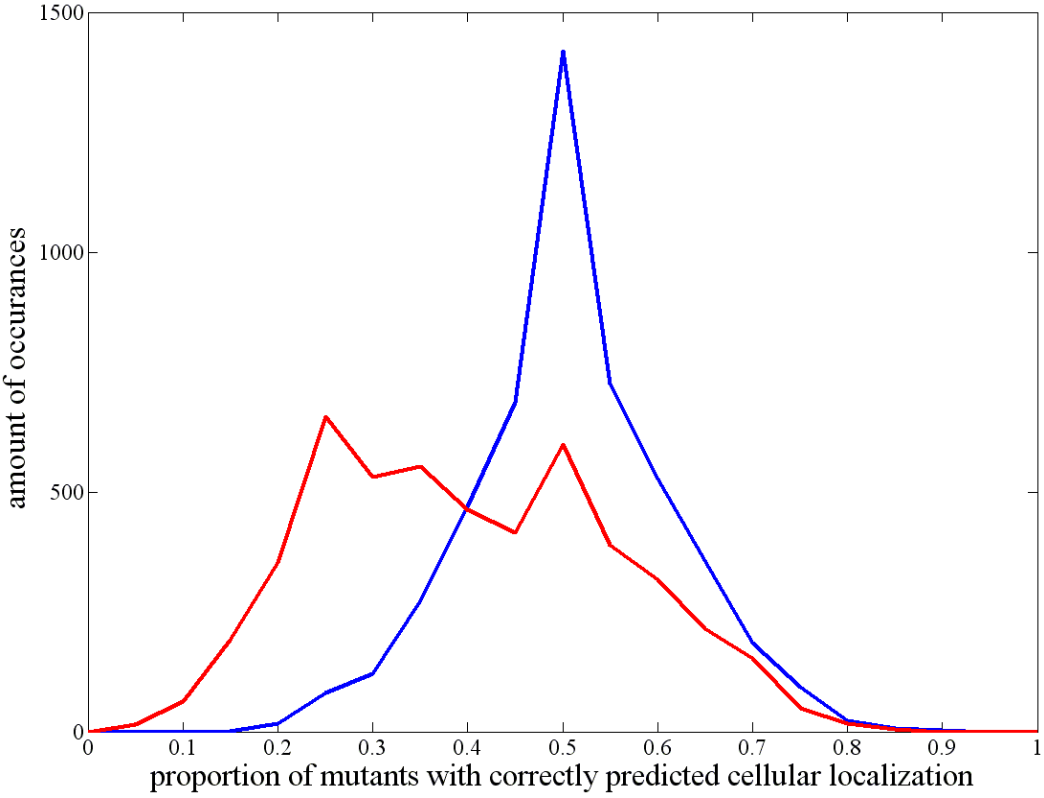

Figure 21 - Distribution of amount of correctly predicted cellular localizations of mutants based on the metabolic profiles of deletants – 5000 randomizations. In blue appears the distribution of correct predictions based on true designations of growth defects to mutants, in red appears the distribution of correct predictions based on shuffled designations of growth phenotype.

We can see a difference in the two distributions. The mean and the median proportion of correct predictions in the control is 43%. Whereas correct predictions ratio based on the **true** data rises to 53% of the cases on average. This difference is highly significant (Wilcoxon rank sum test p-value$<10^{-200}$).

While the increase in the quality of prediction is not very high, it nevertheless allows us to conclude that on a global scale metabolic profiles of deletion mutants contain information regarding the cellular localization of the deleted gene.

2.2.3 Metabolic profiling of S. cerevisiae double mutants in paralogous genes

In order to further elucidate the behavior of paralogous gene pairs in *S. cerevisiae* we have carried out metabolic profiling on selected double mutants in paralogous genes. We chose four different mutants whose double mutant was viable in the minimal medium used in the study. Since we mainly attempted to understand the relationship between single gene deletions and double deletions, we were especially interested in the metabolic manifestation of negative epistasis in yeast. Epistasis is a phenomenon in which a gene either masks or augments the phenotype of another gene. In the case of gene deletions and their effect on fitness one can define epistasis in cases when the fitness of a double mutant is lower than the randomly expected product of the fitnessess of the single mutants ($\varepsilon = W_{x,x''} - W_x W_{x''}$). We have obtained epistasis measurements from the study of DeLuna *et al.*[48]. In this study epistasis was experimentally determined for each pair in a set of multiple pairs of paralogous genes including the four genes used in our study.

The setup of the experiments in the current study slightly differed from that of the single gene deletion experiments. Each day of experiments contained triplicates of WT, mutant 1, mutant 2 and double mutant 1/2. Comparison of results for single mutants from this study with those obtained for the same strain in the earlier study of single gene deletions is presented in section 2.1.8.

Figure 22 depicts the metabolite changes in the single and double mutants following the standard normalization procedures.

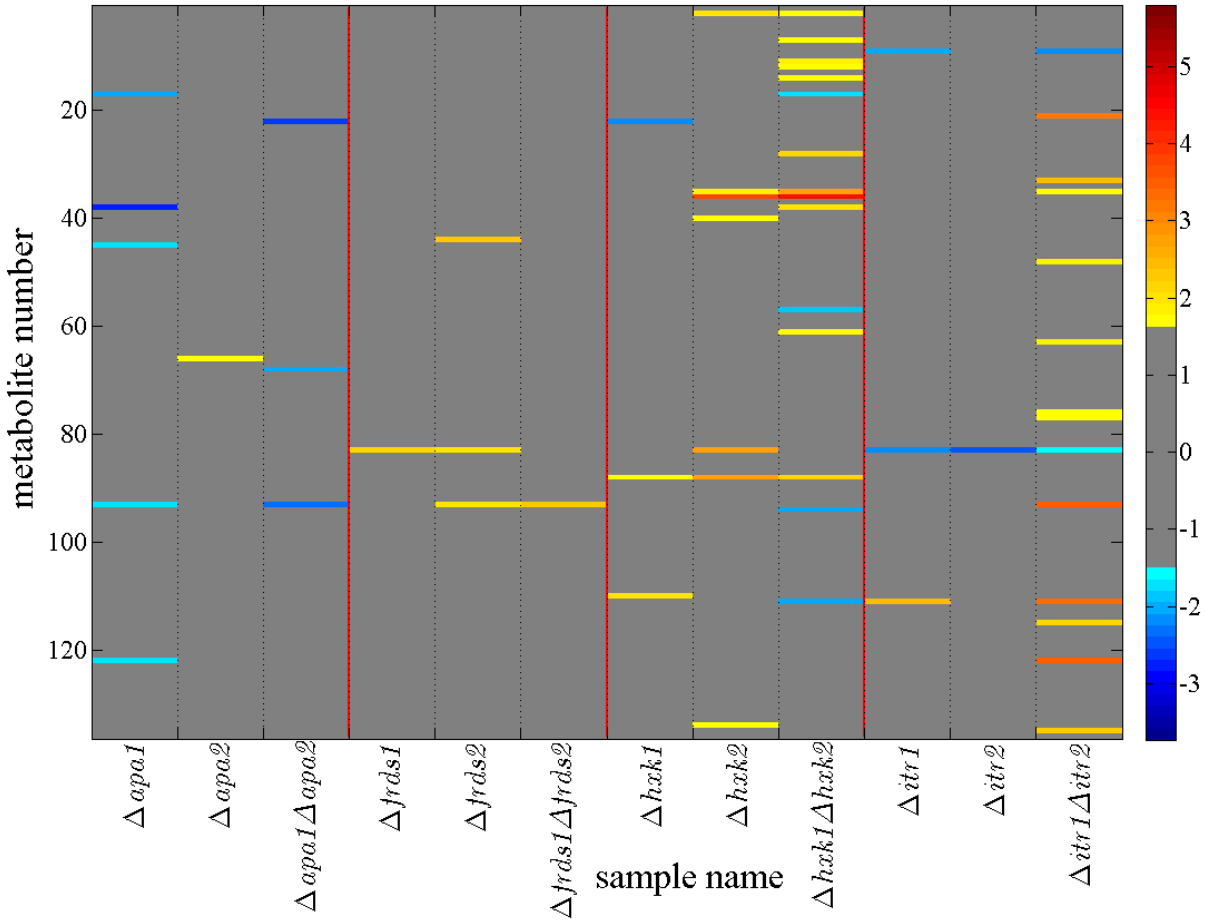

Figure 22 – Metabolite changes in single and double mutants in paralogs relative to WT. Rows represent metabolites' relative intensity. Columns mark different mutants. Metbolomics data for each single mutant and the double mutant from each paralogous pair are found near each other. Diifferent pairs are separated by vertical red lines. The data is on a $\log_2$ scale with the color bar representing the fold change on the right. To facilitate visual analysis metabolites that have either response of less than 3-fold, or that are not significantly changing were greyed out.

One can see that there are marked differences between the response of each pair of genes to deletion and their double mutant. Both double mutants Δ*apa1*Δ*apa2* and Δ*frds1*Δ*frds2* are quite similar to each of their single mutants in the magnitude of their metabolic response to deletion; in contrast, Δ*hxk1*Δ*hxk2* and Δ*itr1*Δ*itr2* are exhibiting very strong metabolic changes relative to their single mutants.

In the next step of analysis we looked at the relationship between each of the mutants in the mutant set and the double mutant with regard to the amounts and identity of all significantly changed metabolites (Figure 23).

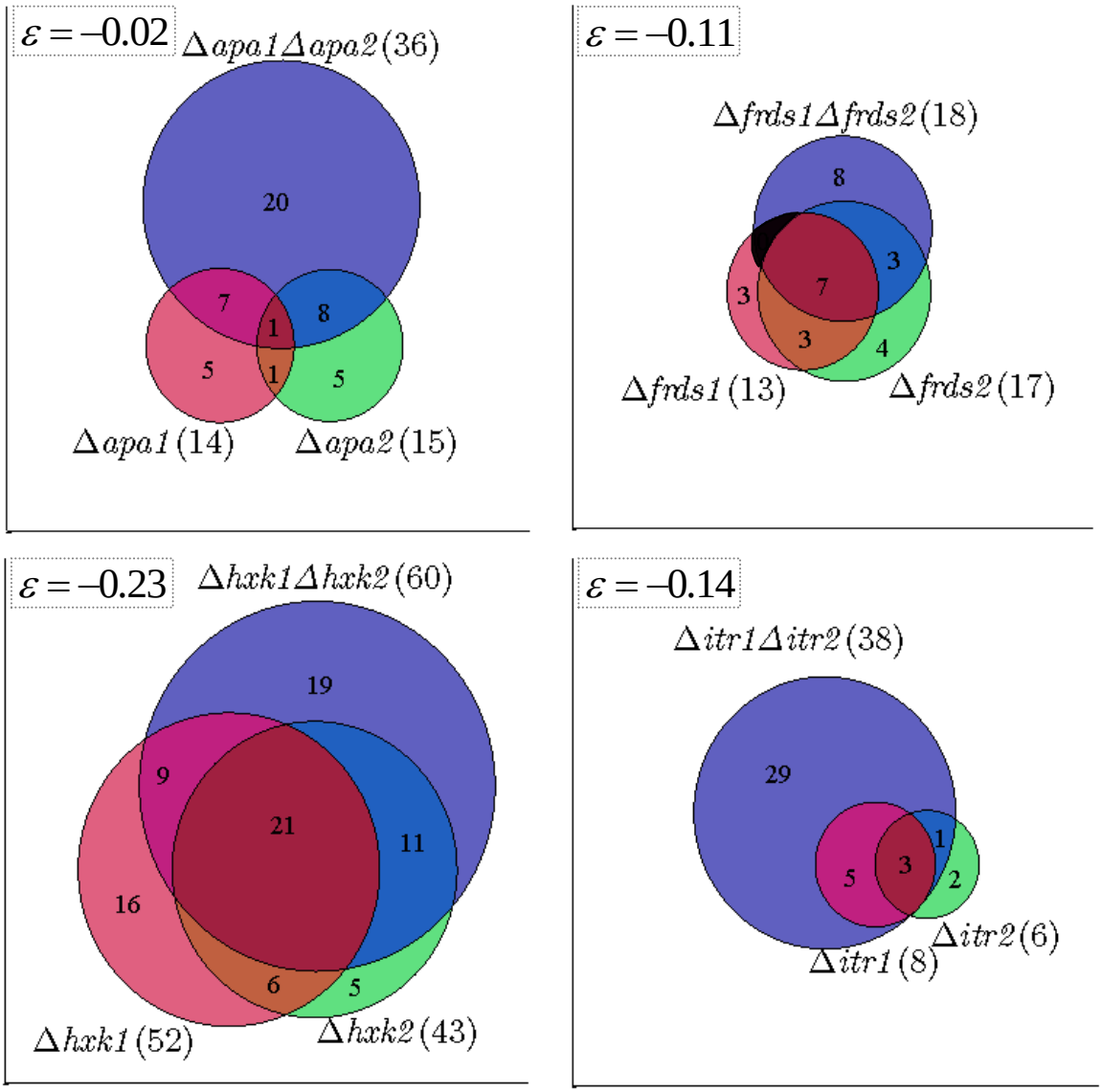

Figure 23 – A Venn diagram describing the amounts of significantly changed metabolites in each of the mutant sets. The labels of the mutants are adjacent to the circles. Circle size and intersections are approximately proportional to the amount of metabolites changed in each group – total amount of significantly varying metabolites in each mutant is given in parentheses. Epistasis (ε) from the DeLuna study for each gene pair is given in the upper left part of each subplot.

If there were no epistasis, the basic expectation for double mutants would be to have a response to the mutation contained only to the union of metabolic response of the mutants. We, however, observed an interesting phenomenon. The double mutants in all cases contained most of the union of the two single mutants and additionally had a relatively very large set of metabolites changing uniquely in these strains. The Δ*hxk1*, Δ*hxk2* and Δ*hxk1*Δ*hxk2* strains present a special case. In these mutants the overlap between the three mutants is very large. This can be explained by the fact that each of the single mutants by itself exhibits significantly slow growth, and the double mutant adds upon that phenotype a synergistic effect resulting in a markedly pronounced growth defect. We noticed a somewhat similar effect in the Δ*apa1*Δ*apa2*

mutant. The growth defect existed in only one of the single mutants, and was not very pronounced. However, the growth of the double mutant was still much lower than in the single mutants, hence the large amount of changed metabolies. Δ*frds1* and Δ*frds2* on the contrary, did not exhibit any growth defect in our medium, and that could be the reason for the absence of a strong metabolic response or a very marked prevalation of the double mutant phenotype. Overall we can see that metabolomic analysis of double mutants carries good potential for understanding the mechanisms underlying the epistasis of genes, although the full understanding and decryption of metabolic profiles requires more accurate results and large data sets.

2.3 Metabolic profiling of S. cerevisiae response to varying intensities and periods of ethanol stress

Another experiment set that we have carried out considered the metabolic response of *Saccharomyces cerevisiae* cells to ethanol stress.

During yeast fermentation ethanol is a major by-product of energy metabolism. Ethanol is toxic for most organisms, as it disrupts cellular membranes, exerts osmotic stress on cells and may inhibit enzyme function within cells. While, yeasts can better cope with ethanol stress, they too suffer from its adverse effects[49-51]. Yeast response to ethanol was investigated on the level of the transcriptome[52]. Several works were presented on the level of the metabolome[53, 54] in which the authors mainly investigated processes associated with ethanol fermentation in *S. cerevisiae*. The metabolomics studies investigated the multitude of effects arising from ethanol fermentation, rather than isolating only the component of stress imposed by ethanol on the cells. We were particularly interested in the specific metabolic effects of ethanol as a model for general stress in yeast. We wanted to address both the question of dosage and duration of ethanol response. The interplay between the two factors is not well studied or understood. Response to ethanol exposure develops over time but it is not known at what is the rate at which the metabolome responds. The response to ethanol may depend on the concentration. It can either be stepwise (absence or presence of ethanol) or depend on the severity of stress.

To get a glimpse at the answers to these questions we have performed a set of experiments involving different times of exposure to different concentrations of ethanol (Table 3).

Ethanol	5%	9%	13%

Duration of stress / concentration			
10 minutes	Sample 1	Sample 2	Sample 3
30 minutes	Sample 4	Sample 5	Sample 6
60 minutes	Sample 7	Sample 8	Sample 9

Table 3 – Durations and intensities of ethanol stress in metabolic profiling experiments

In Figure 24 we can see the distribution of metabolic responses in our experiments with varying stress duration and intensity.

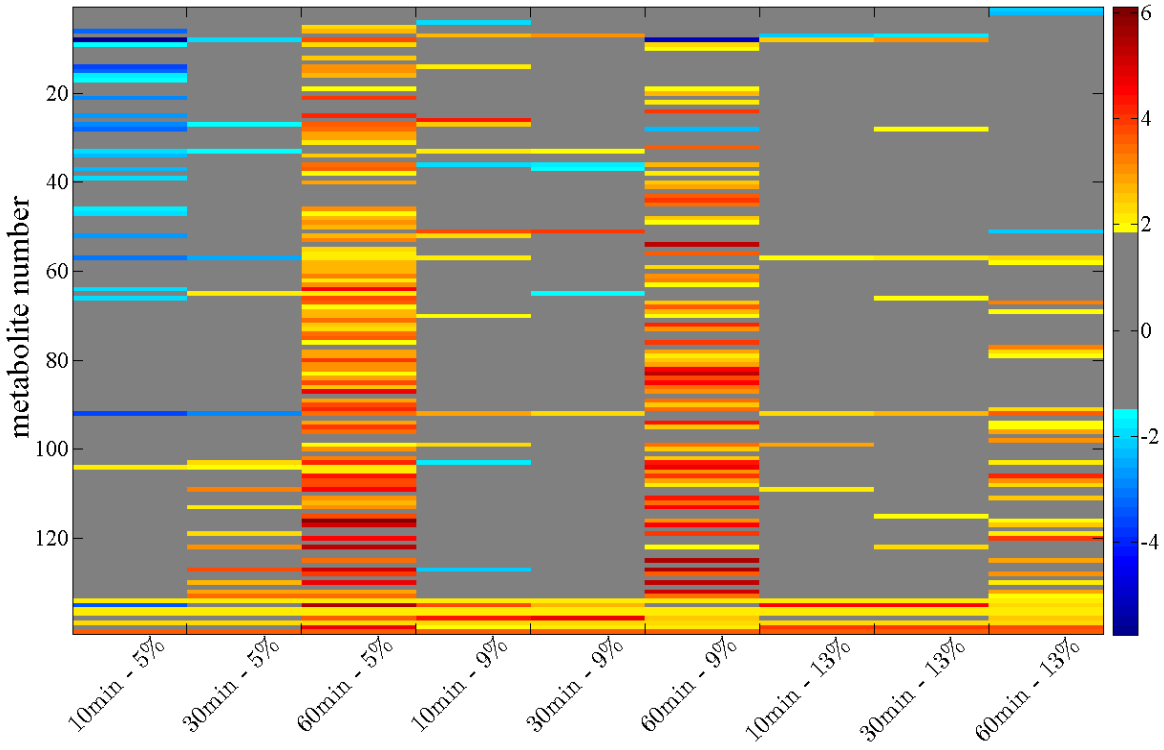

Figure 24 - Distribution of changed metabolites upon exposure to different durations and concentrations of ethanol (labels on x axis hold the different experiments). Grey metabolites exhibited less than 3-fold change.

2.3.1 Analysis of stress duration vs. stress intensity effect on cells

From the first glance we can see marked changes in many metabolites after 60 minutes of ethanol exposure to stress. This result is not dependent upon the intensity of stress. It has been shown that the transcriptional response to ethanol[52] and for other stresses[55, 56] is at its peak after ~30 minutes of exposure to stress. We observe that a metabolic response exists in all time points, but gets extremely high only after the transcriptional reprogramming occurs. Although cells have some capacity for metabolome remodeling even without any new protein synthesis, major changes require enzymes' production. Additionally, it was shown that ~7% of the yeast

genome is rearranged upon ethanol exposure, while the metabolome exhibits a much stronger effect.

Contrary to our expectations severe ethanol stress resulted in a decrease in the amounts of metabolites changed after 60 min. We can accredit this to massive cellular death as was measured by plating and counting colonies.

Specific metabolites that are changed in response to stress include ……..

To be continued…

3 Summary and discussion

We have carried out analysis of paralogous backup on the level of metabolites in *S. cerevisiae.* Several different types of metabolic response could be discerned when looking at the paralogous pairs. Some paralogs did not exhibit any metabolic phenotype, others exhibited one-sided metabolic response, while the third group exhibited a large amount of shared metabolic changes in both paralogs. The mode of metabolic backup might be linked to

4 References

1. Giaever, G. et al. Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387-391 (2002).
2. Steinmetz, L.M. et al. Systematic screen for human disease genes in yeast. *Nat Genet* **31**, 400-404 (2002).
3. Wagner, A. Robustness against mutations in genetic networks of yeast. *Nat Genet* **24**, 355-361 (2000).
4. Wagner, A. Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays* **27**, 176-188 (2005).
5. Gu, Z. et al. Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63-66 (2003).
6. Papp, B., Pal, C. & Hurst, L.D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**, 661-664 (2004).
7. Thatcher, J.W., Shaw, J.M. & Dickinson, W.J. Marginal fitness contributions of nonessential genes in yeast. *Proc Natl Acad Sci U S A* **95**, 253-257 (1998).
8. Nielsen, J. & Oliver, S. The next wave in metabolome analysis. *Trends Biotechnol* **23**, 544-546 (2005).
9. Duarte, N.C., Herrgard, M.J. & Palsson, B.O. Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. *Genome Res.*, 2250904 (2004).
10. Herrgard, M.J. et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* **26**, 1155-1160 (2008).
11. Villas-Boas, S.G., Moxley, J.F., Akesson, M., Stephanopoulos, G. & Nielsen, J. High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochem J* **388**, 669-677 (2005).
12. Allen, J. et al. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol* **21**, 692-696 (2003).
13. Martins, A.M. et al. A systems biology study of two distinct growth phases of Saccharomyces cerevisiae cultures. *Curr Genomics* **5**, 649-663 (2004).
14. Raamsdonk, L.M. et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* **19**, 45-50 (2001).

15. Castrillo, J.I., Hayes, A., Mohammed, S., Gaskell, S.J. & Oliver, S.G. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* **62**, 929-937 (2003).
16. Villas-Boas, S.G., Hojer-Pedersen, J., Akesson, M., Smedsgaard, J. & Nielsen, J. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **22**, 1155-1169 (2005).
17. Canelas, A.B. et al. Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics. *Anal Chem* (2009).
18. Castrillo, J.I. et al. Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol* **6**, 4 (2007).
19. Baganz, F., Hayes, A., Marren, D., Gardner, D.C. & Oliver, S.G. Suitability of replacement markers for functional analysis studies in Saccharomyces cerevisiae. *Yeast* **13**, 1563-1573 (1997).
20. Ewald, J.C., Heux, S. & Zamboni, N. High-throughput quantitative metabolomics: workflow for cultivation, quenching, and analysis of yeast in a multiwell format. *Anal Chem* **81**, 3623-3629 (2009).
21. Fernie, A.R., Trethewey, R.N., Krotzky, A.J. & Willmitzer, L. Metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Biol* **5**, 763-769 (2004).
22. Mintz-Oron, S. et al. Gene expression and metabolism in tomato fruit surface tissues. *Plant Physiol* **147**, 823-851 (2008).
23. Malitsky, S. et al. The transcript and metabolite networks affected by the two clades of Arabidopsis glucosinolate biosynthesis regulators. *Plant Physiol* **148**, 2021-2049 (2008).
24. Roessner, U. et al. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11-29 (2001).
25. Katajamaa, M. & Oresic, M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics* **6**, 179 (2005).
26. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779-787 (2006).
27. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).
28. Steuer, R. On the analysis and interpretation of correlations in metabolomic data. *Briefings in Bioinformatics* **7**, 151-158 (2006).
29. Westergaard, S.L., Oliveira, A.P., Bro, C., Olsson, L. & Nielsen, J. A systems biology approach to study glucose repression in the yeast Saccharomyces cerevisiae. *Biotechnol Bioeng* **96**, 134-145 (2007).
30. Macquillan, A.M. & Halvorson, H.O. Physiological changes occurring in yeast undergoing glucose repression. *J Bacteriol* **84**, 31-36 (1962).
31. Kafri, R., Bar-Even, A. & Pilpel, Y. Transcription control reprogramming in genetic backup circuits. *Nat Genet* **37**, 295-299 (2005).
32. Kafri, R., Levy, M. & Pilpel, Y. The regulatory utilization of genetic redundancy through responsive backup circuits. *Proc Natl Acad Sci U S A* **103**, 11653-11658 (2006).
33. Kafri, R., Springer, M. & Pilpel, Y. Genetic redundancy: new tricks for old genes. *Cell* **136**, 389-392 (2009).
34. Seoighe, C. & Wolfe, K.H. Yeast genome evolution in the post-genome era. *Curr Opin Microbiol* **2**, 548-554 (1999).
35. Khersonsky, O., Roodveldt, C. & Tawfik, D.S. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol* **10**, 498-508 (2006).

36. Nobeli, I., Favia, A.D. & Thornton, J.M. Protein promiscuity and its implications for biotechnology. *Nat Biotechnol* **27**, 157-167 (2009).
37. Jensen, R.A. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30**, 409-425 (1976).
38. Conant, G.C. & Wolfe, K.H. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol Syst Biol* **3**, 129 (2007).
39. SGD (SGD project. "Saccharomyces Genome Database", 2009).
40. Gangloff, S.P., Marguet, D. & Lauquin, G.J. Molecular cloning of the yeast mitochondrial aconitase gene (ACO1) and evidence of a synergistic regulation of expression by glucose plus glutamate. *Mol Cell Biol* **10**, 3551-3561 (1990).
41. Chen, X.J., Wang, X., Kaufman, B.A. & Butow, R.A. Aconitase couples metabolic regulation to mitochondrial DNA maintenance. *Science* **307**, 714-717 (2005).
42. Hughes, A.L. Adaptive evolution of genes and genomes. (Oxford University Press, 1999).
43. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275-276 (1977).
44. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11**, 725-736 (1994).
45. Kafri, R., Dahan, O., Levy, J. & Pilpel, Y. Preferential protection of protein interaction network hubs in yeast: evolved functionality of genetic redundancy. *Proc Natl Acad Sci U S A* **105**, 1243-1248 (2008).
46. DeLuna, A., Avendano, A., Riego, L. & Gonzalez, A. NADP-glutamate dehydrogenase isoenzymes of Saccharomyces cerevisiae. Purification, kinetic properties, and physiological roles. *J Biol Chem* **276**, 43775-43783 (2001).
47. Dwight, S.S. et al. Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Res* **30**, 69-72 (2002).
48. DeLuna, A. et al. Exposing the fitness contribution of duplicated genes. *Nat Genet* **40**, 676-681 (2008).
49. Mager, W.H. & Ferreira, P.M. Stress response of yeast. *Biochem J* **290 (Pt 1)**, 1-13 (1993).
50. Alexandre, H. & Charpentier, C. Biochemical aspects of stuck and sluggish fermentation in grape must. *Journal of Industrial Microbiology & Biotechnology* **20**, 20-27 (1998).
51. Bisson, L.F. Stuck and sluggish fermentations. *American Journal of Enology and Viticulture* **50**, 107-119 (1999).
52. Alexandre, H., Ansanay-Galeote, V., Dequin, S. & Blondin, B. Global gene expression during short-term ethanol stress in Saccharomyces cerevisiae. *FEBS Lett* **498**, 98-103 (2001).
53. Ding, M.Z., Cheng, J.S., Xiao, W.H., Qiao, B. & Yuan, Y.J. Comparative metabolomic analysis on industrial continuous and batch ethanol fermentation processes by GC-TOF-MS. *Metabolomics* **5**, 229-238 (2009).
54. Devantier, R., Scheithauer, B., Villas-Boas, S.G., Pedersen, S. & Olsson, L. Metabolite profiling for analysis of yeast stress response during very high gravity ethanol fermentations. *Biotechnology and Bioengineering* **90**, 703-714 (2005).
55. Causton, H.C. et al. Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell* **12**, 323-337 (2001).
56. Gasch, A.P. et al. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11**, 4241-4257 (2000).